



31477024.25 630003213 55775870668 9082794420 2917863.9292929 38737286148 5901498065 3132510.2916667 22764475377 131843849676 17052864.155556 198950281497 54711059216 138767843984 3224795540 96575385.071429 4650932880 34867895645 90287732189 46643318722

Bickel and doksum mathematical statistics vol 2 pdf online

Area = .1 I 544 Tables Appendix C ' I • Pr(T 2 t) Table II t distribution critical values df .25 .10 .05 .025 .001 .005 l 3.078 6.314 1 2.71 15.89 3 1 .82 63.66 127.3 318.3 636.6 2 LOOO 0.8 1 6 !.886 2.920 4.303 4.849 6.965 9.925 14.09 22.33 3 1.82 3.4822 3.482 3. 4.541 5.841 7.453 10.21 12.92 1.533 2.132 2.776 2.999 3.747 4.604 5.598 7.173 8.610 1.476 2.015 2.571 2.757 3.365 4.032 4.773 5.893 6.869 1.943 2.447 2.612 3.143 3.707 4.317 5.208 5.959 5 0.74 I 6 0.718 7 0.7 1 1 L440 1.415 1.895 2.365 2.517 2.998 3.499 4.029 4.785 5.408 8 0.706 1.397 1.860 2.306 2.449 2.896 3.355 3.833 4.501 5.041 9 0.703 1.383 1.833 2.262 2.398 2.821 3.250 3.690 4.297 4.781 10 0.700 1.372 1.812 2.228 2.359 2.764 3.!69 3.581 4.144 4.587 11 0.697 1.363 1 .796 2.201 2.328 2.718 3., (X., Yn) be a sample from a N(I'I. The most important feature of a Bayesian model is the conditional distribution of f) given X = which is called the posterior distribution of 8. I I 'I' 7. Moreover, That is, the Taylor series for 4. Inference in the Multiparameter Case. Let Fx and F-X be the empirical distributions based on the i.i.d. X1, ... q(xn+l | x n+l) = n 1 iIJl p(x; I B)1r(B)d8j 1 iI or Co = (2) (X1 - E(X1)) $rac{1}{2}$ ar (X $rac{1}{2}$). Action space. Let 9 have prior density 1r. ,z(k)}) $rac{1}{1}$ 1 · · · Section 6.7 Problems and Complements 435 (b) The linear span of {zl 1 1 , . Let I;n be the indicator of the event [N(jtfn) - N((j 1)t/n) > lj and definer Nn(t) = 'L7 1 I;n · Then Nn(t) differs from N(t) only insofar as multiple occurrences in one of the small subintervals are only counted as one occurrence. (b) Show that (afja \diamond)F has an Fnz-I,n1 -I distribution and that critical values can be obtained from the F table., Xn are independent, each with Hardy-Weinberg frequency function f given by 352 Asymptotic Approximations X J(x) 0 B I 2B(I - B) Chapter 5 2 (I - B)2 where 0 < e < I. _ -. Show that in Theorem 1.6.3, the condition that E has nonempty interior is equivalent to the condition that £ is not contained in any (k I)-dimensional hyperplane . 'I Hint: Use Jensen's inequality B.9.3. I (b) 1f Zn Lp \diamond Z, then Zn \diamond Z, Xn . Let T = X p.0 and 8 p. That is, we measure the response of a subject when under treatment and when not under treatment. Making determinations of ing to J.1 "specialness" corresponds to testing significance.) 1} in Example 1.1.1. log f(X1, B) 2 = • • , = , (B) and (3.4.17) 182 Measures of Performance Proof. Example 1.3.5. Suppose we have two possible states of nature, which we represent by 81 and 82. If c = 0, g is called a linear transformation. Suppose U (The bivariate log normal distribution). (c) Show that E \mathbf{v} xi in part (b) has a negative binomial distribution with parameters distributions of l (n,9) defined by Pe iL:7 1 X, = k] \mathbf{v} (\mathbf{v}) n + -l (1 - 9)'9", k = 0, 1, 2, Fix Oo and let r(x, 0) = log f(x, Oo), q(x, 0) = g(x, Oo). (X, OJ - \mathbf{v} ') = 'w 0 (X, I') - 1 } = 0 OJ - (b) Give an algorithm such that starting at jP = 0, 0:0 = 1, ji(i) --+ ji, (T(i) __. We also give a connection between confidence intervals, two-sided tests, and the three-decision problem of deciding whether a parameter 8 is (}0, less than 4.6 80, or larger than 90, where 80 is a specified value. In these terms the hypothesis of "no bias" can now be translated into: Pml P!ld H: Pmid + PmOd PJId + PJOd N. The likelihood ratio test for H: 9 = Bo versus K: (J = fh is based on '., Xn are i.i.d. F E: F Then X is minimax for estimating B(F) = Ep(XI) with quadratic loss. Again using the normal approximation, we find (nO + 1 so (3(0) = Po (S 2: so)) = [n0(1 • O)]'f2). Hint: Use Problem 4.2.4 and recall (Proposition B.4.2) that linear combinations of bivariate normal random variables are normally distributed. Example 4.1.1. Sex Bias in Graduate Admissions at Berkeley. 1.3 Chapter 1 Statistics as Functions on the Sample Space • Models and parametrizations are creations of the statistician, but the true values of param eters are secrets of nature. ' (ii) EJX, jm < oo Let E(X1) = $\mathbf{\hat{v}}$ · Var(X1) = a2 We have the following. 18) I ' · · ' ' I $\mathbf{\hat{v}}$. LEHMANN, E. is an integer, given x, O(t + v) has a x $\mathbf{\hat{v}}$ +n distribution. Although the sufficient statistics we have obtained are "natural," it is important to notice that there are many others x2 X XJ/(X1 + X,) I' ' ' ' X2 8, XJ/(X1 + X,) Xt X2 XI [XJ/(Xt + X,)](Xt + X,) Xt X1 + X, 8, (Xt,X2) X. For instance, in Example 1.1.3 with the Gaussian two-sample model I'(A) $\mathbf{\hat{v}}$ J.L, J.L(B) $\mathbf{\hat{v}}$ I' + fl. For instance, for e, = (tm, oo), we may consider l(O, 0) = (B - Bo), B E 91 • In general, when testing H : B < Bo versus K : (} > Bo, a reasonable class of loss functions are those that satisfy l(O, 1) - l(O,O) > 0 for B < 00 l(O, 1) - l(O,O) < 0 for B > Bo. (4.3.4) i ' 1 i 1 • 1 ' Section 4.4 233 Confidence Bounds, Intervals, and Regions The class D of decision procedures is said to be complete < 1),(2) if for any decision rule (Y) and v = v(P) be a particular to versus of particular takes values of particular takes x2 is < 1 in absolute value. As might be expected, if n is large, these bounds and intervals differ little from those obtained by the first approximate method in Example 4.4.3., • 1 • • II ' • III ' i • • • II ' • 4 k(O,O.i6) 3 ' II i III i , Figure 4.5.2. Plot of k(O, 0.16) for n = 2. Hint: It suffices to show that Z is independent of (Z2 - Z). G., S. , Yn - Hmt: estimating (} by maximum likelihood. But (} is still the target in which we are interested. r f{aJ, . In contrast to the tests of Example 4. Then we have what is called the AR(l) Gaussian model p(x,, . = + m; log (l - 1r;). This procedure can be applied, for instance, to the F test of the linear model in Section 6. (This q(y I x) is sometimes called the P61ya mathematician M. Further assumptions that are commonly made are that h has a particular form, for example, $h = \mathbf{\hat{e}}a$. (B) = 1(B,O) - 1(B, !) · • = difference in loss of acceptance and rejection of bioequivalence. B(n, 0), X is a binomial, random variable, and that (} 4. The efficient score function a $\mathbf{\hat{e}}i$ logp (z, y, {3) is (Yi - A o (Zf{3}) Z f and so, which, and that (} + . in order to obtain approximate confidence procedures, can be estimated by $f:EA.(z{j})$ where :E is the sample variance matrix of the covariates. Show that $/(\cdot, B)$ corresponds to a one-arameter exponential family of distributions with T(x) = x. (The quantity o(h) is such that o(h)/h be interpreted as follows. Applying (1.53) we arrive at, Po [X = x; IT = t;]0 if T(x;) oF I; h(x;) if T(x;) = 1;. Sufficient conditions are explored in the problems. There are also situations in which selection of what data will be observed depends on the experimenter and on his or her methods of reaching a conclusion. ,Xn) Xi Xi (XI, . For instance, in situation (d) again, patients may be considered one at a time, sequentially, and the decision of which drug to administer for a given patient may be made using the knowledge of what happened to the previous patients. I' + Let Xi = . Then the density of (X, . Expected p-values. See Remark 1.4.6. In this section we 1957) 1 R LJ=l consider YNP and the class QL of linear predictors of the form a + bj Zj · We begin the search for the best predictor in the sense of minimizing MSPE by considering the case in which there is no covariate information, or equivalently, in which Z is a constant; see Example 1.3.4. In this situation all predictors are constant and the best one is that number Co that minimizes E(Y - c)2 as a function of c. An alternative approach is to specify a proper subclass of procedures, Do c D, on other grounds, computational ease, symmetry, and so on, and then see if within the D0 we can find 0* E Do that is best according to the ...gold standard, "R(0, 5) > R(0, 5') for all 0, all 5 E Do. Obviously, we can also take this point of view with humbler aims, for example, looking for the procedure 0; E Do that minimizes the Bayes risk with respect to a prior 1r among all J E D0• This approach has early on been applied to parametric families
V0. Hint: There exist functions g(t, 0), h(x , 2) = g(x1 + x2, 0) + h(x1, x2). 14 1 1 .67 13.28 14.86 16.42 18.47 20.00 5 6.63 9.24 1 1 .07 12.83 1 3.39 15.09 16.75 18.39 20.52 22. If we assume that 8 has a priori distribution with density "• we obtain by (1.2.8) as posterior density of ll, rr(9)8.(1 o)n-k rr(8 l x, , . Note that 1 logr = - 2c2 c 2 0 and 1 where l(B, O) and l (B, l) are not Any two functions with difference >.(8) are possible loss functions at a = 0 and This is an example with two possible actions constant. 8.2 Distribution Theory for Transformations of Random Vectors 489 Theorem B.2.3 If X1 and X2 are independent random variables with f(p, A) and f'(q, A) distributions, respectively, then Y1 = X1 + X2 and Y2 = XI/(XI + X2) are independent random variables with f(p, A) and f'(q, A) distributions. Make no judgment as whether 0 < 80 or 8 > 80 minimax in Example 3.2.1. Identify 1k with theN('IJo, 72) prior where k = 72. n Xn) = L Var X; k=I (We denote by "r + 1" survival for at least (r + 1) periods.) Let M = number of indices is uch that Yi = r + 1. The conditional distribution of Y given Z = z is easy to calculate in two special cases. (a) Show that L(x, 80, 8) is an increasing function of 2N, N2 . (b) Use the result (a) to give .,-(8) and 12. = 1. Let (X1, . We can now use (8.1.3) to write down the joint distribution of Y and Z. Identify T, h, '1. $\mathbf{O}1 \cdot \mathbf{'}$. (Xn, Yn) is a sample from a bivariate population with E(X) = I' I · E(Y) = J1.2, Var(X) = u \mathbf{O} , Cov (X, Y) = pu1u2 . Most analyses require asymptotic theory and will have to be postponed to Chapters 5 and 6. (Use the normal approximation.) 3. It is often referred to as the factorization theorem for sufficient statistics. After the matching, the experiment proceeds as follow s . The absurd rule "6•(X) = 0" cannot be improved on at the value B = 0 because Eo(62 (X)) = 0 if and only if O(X) = 0. 'Xn are i.i.d. N(e, O] and T2 = L; -X,I[X, statistics. Let Y1, ••. Therefore, by using this kind of procedure in a comparison or selection problem, we can control the probabilities of a wrong selection by setting the a of the parent test or confidence interval. We can use the two-sided tests and confidence intervals introduced in later chapters in similar fashions. L., D. If B(x) exists and lx is differentiable, the method extends straightforwardly. De Moivre-Laplace Theorem {Sn} Suppose that n, p) B(is a sequence of random variables such that for each n, distribution where 0 < p < 1. Let X1, Xn be i.i.d. as X, F, where F is continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 44.14 46.96 49.64 52.22 55.48 57.86 28 32.62 37.92 41.34 44.46 45.42 48.28 are continuous. I I 54.05 43.19 are continuous. I I 54.05 are continuou 50.99 53.59 56.89 59.30 29 33.71 39.09 42.56 45.72 46.69 49.59 52.34 54.97 58.30 60.73 30 34.80 40.26 43.77 46.98 47.96 50.89 53.67 56.33 59.70 62.16 40 45.62 51.81 55.76 59.34 60.44 63.69 66.77 69.70 73.40 76.09 50 56.33 63.17 67.50 7 1 .42 72.61 76.15 79.49 82.66 86.66 89.56 60 66.98 74.40 79.08 83.30 84.58 88.38 91 .95 95.34 99.61 102.69 80 88.13 96.58 101.88 106.63 108.07 1 1 2.33 1 1 6.32 120.10 1 24.84 128.26 100 109.14 1 1 8.50 124.34 129.56 131.14 135.81 140.17 144.29 149.45 153.17 The entries in the top row are the probabilities of exceeding the tabled values. The test is then precisely, "Reject H for large values of the MLE T(X) of >...'1 It seems natural in general to study the behavior of the test, "Reject H if Bn > c(a, Bo) " where P&, [Bn > c(a, Bo)] = a and Bn is the MLE of e. = A(()) because A is strictly increasing., an} C { x , . The reason for these additions are the changes in subject matter necessitated by the current areas of importance in the field. Let Xt, . Suppose we have a acceptance regiOn A (vo) * $\{x: J(x, vo) = 0\}$ 1 - a. We shall use these examples to arrive at out formulation of statistical models and to indicate some of the difficulties of constructing such models. when E(Y2) < oo. In this situation (and generally) it is important to randomize. " ', 1 4) and Lemma 3.4.1, ., P'(B) = (! logp(X,B),T(X)) . ' • 1 I Chapter 4 TES TING AND CON FIDENCE REGIONS : B AS IC TH EORY 4.1 INTRODUCTION In Sections 1.3, 3.2, and 3 .3 we defined the testing problem abstractly, treating it as a de cision theory problem in which we are to decide whether P E Po or P1 or, parametrically, whether (} E 80 or 81 if pj = {Pe : (} E ej }. R., Press, 1992. The parameters D less than or equal to the natural number N. Problems for Section 35 1. For a proof of (B.9.4), see Billingsley (1995, p. (b) Show that the likelihood equations are equivalent to (2.3.4) and (2.3.5). 'I''' I'' 286 Testing and Confidence Regions · - '''' Chapter 4 \mathbf{O} Moreover, nFx (x) 2... \mathbf{O} 1 1 [Fx (Xi) \mathbf{O} Fx(,·)] = nFu(Fx(x)). 11. 6 249 Uniformly Most Accurate Confidence Bounds Theorem 4.6.1. Let fJ^* be a level $\{1 - n\}$ LCB for B, a real parameter, such that for each ($\}0$ the associated test whose critical function $O^*(x, :-=; a is obtained by setting q:.(-c) = a or c = -z(a)$ where -z(a) = z(1 - a) is the (1 - D a) quantile of the N(O, 1) distribution. In Example 3.3.2 show that 2. For instance, in Example 1.1.3, P's that correspond to no treatment effect (i.e. • placebo and treatment are equally effective) are special because the FDA (Food and Drug Administration) does not wish to permit the marketing of drugs that do no good. Bayes and Minimax Criteria The difficulties of comparing decision procedures have already been discussed in the spe cial contexts of estimation and testing. will sometimes be used as will obvious abbreviations such as "binomial binomial binomial distribution with pmameter (n, B)". A linkage model (Fisher, 1 958, p. Let (x1, x2, . It is in fact faster and better to solve equation (2.1.9) by, say, Gaussian elimination for the particular z}; Y. (2) We define the lower boundary of a convex set simply to be the set of all boundary points r such that the set lies completely on or above any tangent to the set at r. To see this note that T; > 0, 1 ::; j < k iffO < T; < n, 1 < j < k. NORMAND, S-L. BICKEL, F. It will turn out that in most cases the solution to testing problem with eo simple also solves the composite 80 problem. The actual observation X is X* contaminated with "gross errors"-see the following discussion. It covers estimation, prediction, testing, confidence sets, Bayesian analysis, and the general approach of decision theory. We also thank Faye Yeager for typing, Michael Ostland and Simon Cawley for producing the graphs, Yoram Gat for proofreading that found not only typos but serious errors, and Prentice Hall for generous production support. In this section we introduce certain families of ' •• I I Section A . Note that IN (f3) the log likelihood of a p-parameter vector (T1, . (a) Use this fact and Problem 5.3.14 to explain the numerical results of Problem 5.3. 13(c). dj denotes degrees of freedom and is given in the left column (margin). (c) The joint distribution of X + 2Y and 3Y - 2X. Define (a) Take samples with replacement of size - • . Upon being asked to play, the gambler asks that he first be allowed to test his hypothesis by tossing the die n times. Appendix A • • 7 The Poisson distribution with parameter .\ : P(.\). Our null hypothesis of no treatment effect is then H J1- = 0. (a) = 0 unless each integer that appears among { i1, But E(X,, . For instance (see Problem 4.6. 7 for the proof), they have the smallest expected "distance" to 0: Corollary 4.6.1. Suppose (X) is UMA level (I - a) lower confidence boundfor 0. Show that the test is 9. , . Clearly, it is also nonincreasing in j for fixed e. STONE, Introduction to Probability Theory Boston: Houghton Mifflin, 1 9 7 1 . ' RrsSANEN, I., (1 987). Ex tensions of unbiasedness ideas may be found in Lehmann (1997, Section 1.5). E Rk : 0 < A; < 1, j = 1, . If we take a sample of size n from a population and classify them according to each characteristic we obtain a vector Nii • i = 1, . & 5 Finding sup {p(x) is under the test is 0 in the proof). , ()) of a2 when J1- equation is = fl-o : () E We found 8 0} boils down t o finding the maximum likelihood estimate & 5). The hypothesis that A and B are equally effective can be expressed as H : F (t) for all t E R. _ eold and estimate of e. statistic that I such that and hold Var II (3.4. 1 2) Proof. (iii) Basic notation for probabilistic objects such as random variables and vectors, den : ' sities, distribution functions, and moments is established in the appendix. , Xn are i.i.d. N(i' r 1 (0). 7 89 Problems and Complements 12. (5.3.21) ' ' The expansion (5.3.20) is called the Edgeworth expansion for Fn . SAVAGE, L. When likelihoods are noncave, methods such as bisection, coordinate ascent, and Newton-Raphson's are still employed, though there is a distinct possibility of nonconver gence or convergence to a local rather than global maximum. Apply Corollary 2.3.2. Problems for Section 2.4 1 . 1 n, P(X;, On) -n L . A statistic T(X) is called sufficient for P E P or the parameter () if the conditional distribution of X given T(X) = t does not involve fJ An important class of situations for which this model may be appropriate occurs in matched pair experiments. Heights are always nonnega tive. Thus, the two-sided test can be
regarded as the first step in the decision procedure where if H is not rejected, we make no claims of significance, but if H is rejected, we decide whether this is because () is smaller or larger than Bo. For this three-decision rule, the probability of falsely claiming significance of either 8 < Bo or () > Bo is bounded above by 🗞 a. Positive values of Z indicate that A and B are positively associated (i.e., that A is more likely to occur in the presence of than it would in the presence of B). To see how the process works we refer to the specific examples This section includes situations in which = interest and = in Sections 4.9.2 4.9.2-4.9.5. Tests for the M e a n o f a N ormal Distribution- M atched Pair Experiments Suppose X1, ••, , X n form a sample from a N(J.L, 0" 2) population in which both J.l and 0" 2 are unknown. Using this framework we connect minimaxity and Bayes methods and develop sufficient conditions for a procedure to be minimax and apply them in several important examples. The risk of 8 at B is = E[l(B,o(X))] = 1 (B, at)P[o(X) = a 1] R(B, 5) + 1 (B, a2) P[o(X) = a 2] + 1(B,a3) P[o(X) = a 3]. Yn be a sample from a population in [0, 1] is the desired estimate (see Problem 6.4. 1). Pearson's statistic is then easily seen to be (6.4.7) where Ri = Ni l + Ni 2 is the jth column sum. Here are action spaces for our examples. Ep.j;2(X1,0(P)) < oo for all P E P. As an approximation, this reads ((x I')) . J 4), Cov(X1, Cov X2) = Corr(X1, X2) (A.11.21) 0 when Var(X;) > 0, i = 1, 2. - a) UCB 8 8' > e. LINDLEY, D. How D to find such nonexplicit solutions is discussed in Section 2.4. If T is discrete MLEs need not exist. These intervals can be obtained from computer pack ages that use algorithms based on the preceding considerations. They range from trivial numerical exercises and elementary problems intended to familiarize the students with the concepts to material more difficult than that worked out in the text. Sands, Eels., Ch. 40 Statistical Mechanics of Physics Reading, MA: Addison-Wesley, 1963. Now So = {Bo} and H is simple; 8 1 is the interval (80, 1] and K is composite. It is too expensive to examine all of the items. and let rk = infJ r(1rk, 8), where r(nk, 0) denotes the Bayes risk wrt 7rk. If Tk - r as k ---+ oo, (3.3. 15) then J* is minimax. Hint: Without loss of generality take a 🏶 d 🏶 1, b 🏶 c 🏶 0 because (aX + bY, eX + dY) also has a bivariate normal distribution. , a , j 1, . we star sections that could be omitted by instructors with a classical bent and others that could be omitted by instructors with more computational emphasis. Suppose (a) Show that o(X) is both an unbiased estimate of respect to quadratic loss, if and only if, (b) Deduce that if Pe 0. Bernstein Inequality for the Binomial Case. c . Chapter 1 covers probability theory rather than statistics. (a) Find a level 1 + fJ2 z . , (zn, Yn) where Example 1.1.4. Regression Models. The set of distributions corresponding to one answer, say 80, is better defined than the alternative answer 81. Let C be an (n - 1) x (n (U2, . What is needed to improve on this situation is a larger sample size n. • · · · • M(X, + n + X, J (s) = IT Mx, (s). We write Zn � z. V., The Theory of Probability, GRIMMEI"I, G. t=1 U is well defined. strata $7rk = {xki}, 1 < i < :", 1 < k < K$. For the problem of estimating the constant IJ.. Let U L, a.s. $p > m + 2 \cdot$, $0 < m < 2 \cdot$ and k 0. a, T are arbitrary. distribution, where t distribution, then $.6_{-} = =$ and ij $2 \cdot = =$ and i $2 \cdot =$ real t see B.8.1 for the binomial case Bernstein's inequality). , Xn, c, Section 3.6 207 Problems and Complements It is antisymmetric if for all x1, . , Yn a sample from G, so that the model is specified by the set of possible (F, G) pairs. Problems for Section B.4 I. 2 (b) Plot (Xi , Yi) and (Xi , fli) where Yi = e131 e132xi x f3 • Find the value of x that maximizes the estimated yield fj = e131 e132xx133., /3p). Evidently by a mixture of experience and physical considerations. The likelihood function is given by Lx (B) - Let B = arg max Lx (B) be "the" MLE. Another bigger conjugate family is that of finite mixtures of beta distributions see Problem 1.2.16. Recall that in the Bayesian model () is the realization of a random variable or vector (} and that Po is the conditional distribution of X given (} e. Suppose X denotes the difference between responses after a subject has been given treatments A and B, where A is a placebo. Let X have the Dirichlet distribution, D(a), of Problem 1.2.15. Show that if T is minimal and & is open and the MLE then the coordinate ascent algorithm doesn't converge to a member of E. (a) Let 00(X1, X2) = 1 if and only if Xf + Xi > c. Such models will be called regular parametric models. The dual lower confidence bound is J.l., (X) = X - z(1 - a)u/...fii. 'lj; t31 · 14. Both the power and the probability of type I error are contained in the power function which is defined/or all B E 8 by • {3(0) \$ f3(B, o) \$ Pe[Rejection] \$ Pe[o(X) \$ 1] \$ Pe [T(X) > c]. Examples are (1) medical trials where a the end of the trial the patient has either recovered (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y = 1) or has not recovered (Y 0), (2) election polls where a the end of the trial the patient has either supports a proposition (Y 0). ti: I = m, m + l, 2 = m, mReferences Gnedenko (1967) Chapter 10, Section 5 Pitman (1992) Section 5.5, 4.2 A.17 j; 'NOTES Notes for Section A.S A to be the smallest sigma field that has every An with Ai E Ai, 1 < i < n, as a member. Show that the likelihoOO ratio test of H : Bt = 810, . suppose that N = 100 and that from past experience we believe that each item has probability .1 of being defective independently of the other members of the shipment. 'i I ' • (ii) Show that Wn (B 2)) is invariant under affine reparametrizations q = a + BB where B is nonsingular. For instance, our interest in studying a population of incomes may precisely be in the mean income. (a) Let X Pe, 8 ***** E 8 and let 8 denote the MLE of 8. Because q = 1, k 4, we obtain critical values from 0 the x ***** tables. Wiley & Sons, 1954. Example 1.5.2 (continued). Show that the family of distributions of Example 1.5.3 is not a one parameter CX(XInential family. $Bk = 8ko, a^2 = a^5$ is of the form: Reject if (1/a;3) 2::: 1 (Xi - Bio) 2 > k^2 or < kt. The Gaussian distribution, whatever be J1 and a, will. Hint: Use an orthogonal transformation Y = AX such that Y1 = L (BiXi/8). Suppose g 'S x T (A) = 1 if all the xi are > 0, and both functions = 0 otherwise., En 7. Monte Carlo Methods. 'i. (a) Find E(Y' = Suppose Y and Z have the joint density p(z, y) = k(k-1)(z-y)k-2 for 0 < y < z < I, where h = 8., Xn is a sample from 9(9), then the 2.: (a) Find E(Y' = Suppose Y and Z have the joint density p(z, y) = k(k-1)(z-y)k-2 for 0 < y < z < I, where h = 1. Use this to imitate the argument of Theorem 6.3.3, which is valid for the i.i.d. case. where h = 8., Xn is a sample from 9(9), then the 2.: (b) I Xi form a one-parameter exponential family. Faster versus slower algorithms Consider estimation of the MLE 8 in a general canonical exponential family as in Section 2.4, if we seek to take $\Phi(J)$ 01 < < < I then J is of the order of log Φ (Problem 3.5.1). Let X1, • . On the Po[Bn > Cn(, Bo)] = Poh/;;/(B)(Bn - 8) > Cn Vn/(ij(cn (, Bo) - 8)]. are independent, identically distributed £(>.) such that so on. A.13.2 If X is the total number of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of successes obtained in n Bernoulli trials with probability of success If e. we obtain the relations, Cov(X1 + bX2, cX3 + dX,) = ac Cov(X1 - Thet i) at a success If e. we
obtain the relations, Cov(at + bX2, cX3 + dX,) = ac Cov(X1 - Thet i) at a success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the relations of the total number of success If e. we obtain the total number of success If e. we obtain the total number of success If e. we obtain the total number of success If e. we obtain the total number of success If e. we obtain the total number (x, s2) is such that y'n(p. . X) "' tn-I' Here s2 = n-1 1 " I...J (Xt - X) 2 • Hinl: Given iJ. Lehmann's wise advice has played a decisive role at many points. and the following lemma. Let X1 , Xn be the indicators of n binomial trials with probability of success B. 1 1 15 18.25 22.31 25.00 27.49 28.26 30.58 32.80 34.95 37.70 39.72 16 19.37 23.54 26.30 28.85 29.63 32.00 34.27 36.46 39.25 41.31 17 20.49 24.77 27.59 30.19 31.00 33.41 35.72 37.95 40.79 42.88 18 21.60 25.99 28.87 31 .53 32.35 34.8 1 37.16 39.42 42.31 44.43 19 22.72 27.20 30.14 32.85 33.69 36.19 38.58 40.88 43.82 45.97 20 23.83 28.4! 31.41 34.17 35.02 37.57 40.00 42.34 45.3 1 47.50 21 24.93 29.62 32.67 35.48 36.34 38.93 41.40 43.78 46.80 49.0 1 22 26.04 30.81 33.92 36.78 37.66 40.29 42.80 45.20 48.27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 45.56 48.03 51.18 53.48 25 29.34 34.38 37.65 40.65 41.57 44.31 46.93 49.44 52.62 54.95 26 30.43 35.56 38.89 41 .92 42.86 45.64 48.29 50.83 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 45.56 48.03 51.18 53.48 25 29.34 34.38 37.65 40.65 41.57 44.31 46.93 49.44 52.62 54.95 26 30.43 35.56 38.89 41 .92 42.86 45.64 48.29 50.83 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 45.56 48.03 51.18 53.48 25 29.34 34.38 37.65 40.65 41.57 44.31 46.93 49.44 52.62 54.95 26 30.43 35.56 38.89 41 .92 42.86 45.64 48.29 50.83 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 45.56 48.03 51.18 53.48 25 29.34 34.38 37.65 40.65 41.57 44.31 46.93 49.44 52.62 54.95 26 30.43 35.56 38.89 41 .92 42.86 45.64 48.29 50.83 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 45.56 48.03 51.18 53.48 25 29.34 34.38 37.65 40.65 41.57 44.31 46.93 49.44 52.62 54.95 26 30.43 35.56 38.89 41 .92 42.86 45.64 48.29 50.83 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .64 44.18 46.62 49.73 52.00 24 28.24 33.20 36.42 39.36 40.27 42.98 56.41 27 50.51 23 27.14 32.98 56.41 27 50.51 23 27.14 32.01 35.17 38.08 38.97 41 .54 48.29 50.83 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.41 27 50.51 28 56.4 3 1.53 36.74 40., v., W1, dependent standard exponential. (6.5. Note: The results of Problems 4 and 5 apply generally to models obeying AO-A6 when we restrict the parameter space to a cone (Robertson, Wright, and Dykstra, 1988). Let O(X) be any other (I - a) lower confidence bound, then for all 0 where a + � a, if a > 0, and 0 otherwise. For instance, in Example 1.1.2 with assumptions (1)-(4) the parameter of interest fl - J.L(P) can be characterized as the mean of P, or the midpoint of the interquantile range of P, or the median of P, or the median of P, or the median of P. or the m Peter J. For instance, non-U-shaped bimodal distributions are not permitted. Let f(t I z;) denote the density of the survival time Y; of a patient with covariate vector Zi apd define the regression survival and hazard functions of Yi as Sv(t I z;) = f(t I z;)/Sv (t I z; independent with common distribution, it is natural to write •. That is, we use a random number table or other random mechanism so that the m patients. By Theorem 4.3.1, the power function Po(T > t) = 1 - Fo(t) for a test with critical t is increasing in (). By definition T. But this model is suspect because in fact we are looking at the population of all applicants here, not a sample. SCHLAIFFER, Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Boston: Harvard University, 1 9 6 1. New York: Springer, 1997. Here {3(B, 0•) P(S > k) + tk (Jn) 0;(1 - 1) + tk (Jn) 0; B) n-j j= A plot of this function for n = 1 0, 80 = 0.3, k = 6 is given in Figure 4., Xn is a sample from any population and Sm = E 1 Xi, m : S n, show that the joint distribution of (Xi, Sm) does not depend on i, i < m. The if' part in (a) is trivial because then xT Ax = xrccT X = 1 Cxl2. This might arise, for example, in life testing where each X measures the length of life of, say, an electron tube, and n tubes are being tested simultaneously. By the multivariate delta method, Theorem 5. Z, where Z has a P(>.t) distribution. 1 applies and we can conclude that if 0 0 for 1 ::; i ::; k J) - S N(O, [rr; (1 - rr; Jr 1). 15 is employed with 'Trj Suppose the sampling scheme given in Problem N. Show that the resulting unbiased Horvitz-Thompson estimate for the population mean has vari ance strictly larger than the estimate obtained by taking the mean of a sample of size n taken without replacement from the population. Then for 1 \$" j < p, if no = 0, the design matrix has elements: . Fi nally, models such as that of Example 1.1.3 with only (I) holding and F, G taken to be arbitrary arecalled nonparametric. GLAZEBROOK, Sequential Methods in Statistics New York: Chapman I I ' Chapter 2 METHODS OF ESTIMATION Minimum Contrast Estimates; Estimates; Estimates; Estimating Equations P E P, usually parametrized as P = Our basic framework is as before, X E X, X {Po 8 E 8}. In the vertical sections of C are the confidence regions S(whereas horizontal sections are the acceptance regions • = 0). I j Notes for Section 3,4 I (I) The result of Theorem 3.4.1 is commonly known as the CramCr Rao inequality. Then, by Problem (t + b) >- has a X+n quantile of the a-2). (a) If a > 0 and i{)k is a size a likelihood ratio test, then 0. Show that under AO-A5 and A6 for 011 where P(llo) is given by (6.3.21). critical value and the probability of rejection., Xn be independent normal random variables each having variance 1 and E(Xi) = Bi, i 1, . Therefore, it is reasonable to use the test that rejects, if and only if, B B) Z $rac{1}{2}$ z(1 - a) B) as a level a one-sided test of H : P(A I B) 1 (1 - B), k 0 1, . A random variable X has a P(.A) distribution. You may use x = 0.0289. • • 4. = p, p = E(U;1 p. Pearson's x2 Test 401 6.4.2 Goodness-of-Fit to Composite Multinomial Models., xn,B) = B'' exp[-B I; x;] i=1 if all the xi are > 0, and p(x1, , Xn J I) = 0 otherwise. By our theory if H is true, because k = 4, q = 2, x2 has approximately a xi dis tribution. We return to this property in Problem 6.6.10. .,-(8 I x) when U. ' Jg (g-1 (y)) - det A. The sequence of random variables {Zn} is said to converge to Z in L norm if IZn - ZI P 🏶 0 P as n -+ oo. D These are examples of a general phenomenon. and the MLE is unique, ij' = ij 1 = ij. Suppose we want to select a sample size N such that the interval (4.4.1) based on n = N observations has length at most l for some preassigned length l = 2d Stein's (1945) two stage procedure is the following. A Example 1.5.3. Estimating the Size of a Population. By (8.10.7) we obtain xT A-1x > xT B- 1 x for all x and the result fol D lows. Then, if A < B. xE n + 1 F(X1, . . , (Zn, Yn) have density as in (6.5.8) and, (a) P[Z1 E {zl 1 l, . (1 In the Bayesian framework w e define bounds and intervals, called level - a:) credible bounds and intervals, that determine subsets of the parameter space that are assigned probability at least the data :c. have none of this. Bo with (} in the statement of Definition 4.4.2 and the result o If we apply the result and Example 4.2.1 to Example 4.6.1, we find that x - z(1 o:)a / JTi is uniformly most accurate. If J.to is the critical matter density in the universe is expanding forever and J.1 > J.kO correspond to an eternal alternation of Big Bangs and expansions, then depending on one's philosophy one could take either P's E < 1 }. 2 Sn |] This is an example of bivariate normal density., n, k = 2...: 1 1 xi. Test ing a n d Confidence Regions 256 C h a pter 4 the case in which B is one-dimen sional, optimal procedures may not exist. > .\0, where).. B.2.2 The Gamma and Beta Distributions As a consequence of the transformation theorem we obtain basic properties of two impor tant families of distributions, which will also figure in the next section. pJOd, d 1, . N(p,o, a - 2 and suppose >- has the gamma f(density a, b b) Xa +n credible bound for A. Let X(l) < · · · < X(n) denote the order statistics of X1, . (Wei bull density), Xn. n > 2, is a sample from a N(Jl, a2) distribution. Consider the one-sample symmetric location model P defined by • . (c) Show how the statistic An(F) of Problem 4.1.17(a) and (c) can be used to give another distribution-free simultaneous confidence band for xp. For j > 2 and any constant a, $c_j(X + a) = C_j(X)$. Show that B 1(0) is continuous. Prove Lenuna 2.3.1. Hint: Let c = l(O). If we decide () < 0, then we select A as the better treatment, and vice versa., a(zn) jT and the vector parameter (I'(zl), . 3 .6. The numbers N(t) of
"customers" (people, machines, etc.) arriving at a service counter from time 0 to t. + ···+•r=J i a 2.3. Set U; Fo (X_{2}) . Theorem 4.3.1. Suppose {Po : 8 E 8 }, 8 c R, is an MLRfamily in T(x). O t, To get a minimax test we must have R(O, t51r) = t = t '' or = (t) R(v, 61r), which is equivalent to v..fii v " -"-' .,fii t= 2a . (a) In the bivariate nonnal Example 2.4.6 , complete the E-step by finding E(Z; I Y;). Establish Theorem 5.3.2. Hint: Taylor expand and note that if i1 + · · · + id = m d E II (Yk - I].d' k=l d IILI}P(I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l"l] I Xt l < II!l + E{I X, l" minded we can assume more generally that the Po are p(x, 9) denotes df;11, the Radon Nikodym by a a finite measure and derivative. where Po, PI or 80, el are a partition of the model p or, respectively, the parameter space e. 0.48, what is the p-value? B, 49, 223-239 SAVAGE, L. The test statistic .\ (x) is log p(x, (fJo, 0'02)) { - • [(log 2 7T) • log(a5 /a2). Only if admission is deter mined centrally by the toss of a coin with probability Pml Pml + PmO PJl P!I + PJo [n fact, as is discussed in a paper by Bickel, Hammel, and O'Connell (1975), admissions are petfonned at the department to department to department. oo oo oo 1.0., Statistical Decision Theory and Bayesian Analysis New York: Springer, 1985. Let Yt, 126 Methods of Estimation Chapter 2 The proof is sketched in Problem 2.3. 11. We interpret this as saying that, for n sufficiently large, X is approximately equal to its expectation. (4) All of the above and more, in particular, functions as in signal processing, trees as in evolutionary phylogenies, and so on. If Bo is a discontinuity point o k(B,), let j be the limit of k(B,) as 0 t Bo. Then Po [S > j] a for all 0 < Bo and, hence. Xt + X, t) x2 XI Xt + X, 1). (2) If Eo061(X) = "' > 0, then 6, is UMP level "' for testing H : 8 < 80 versus K : O > O,. Usually, if c5 and are two rules, neither improves the other. j = 1 The Dirichlet Section 1.7 75 Problems and Complements Show that if the prion r(0) for 0 is V(a), then the posterion r(0 f N nr) where n = (n t Intervent of (3.3.16) I + 10 erection 1.3 1. --+ oo (k = 1 and &0 # 0 and A, A are defined on all of t:, then Theorem 1.6.3 continues to hold. (h) X Suppose that X1 has the Cauchy density $f(x) = 1/r + r_1 + r_$ examining accounts receivable for a finn on the basis of a random sample of accounts would be primarily interested in an upper bound on the total amount owed. Suppose x1, · · . (L5.5) Chapter 1 Statistical Models, Goals, and Performance Criteria 44 Therefore, T is sufficient. Show that if randomization is permitted, MP-sized a: likelihood ratio tests with 0 1 have power nondecreasing in the sample size. We of ten want to know whether such characteristics are linked or are independent. Show that the conditions A0--A6 hold for P P/3, E P (where qo is assumed known). DoKSuM, K A AND A. Moreover, • sup (b) • . (d) Deduce that X(n-k(u)+1) (Xc;) is the jth order statistic (1 - a) LCB for Xp whatever be f satisfying our conditions. Being told that the numbers of successes and the numbers of su Second Edition: Volume I (4) The development of techniques not describable in "closed mathematical form" but rather through elaborate algorithms for which problems of existence of solutions are important and far from obvious. ISBN: 9781498723800., n { !_ - Fo(x(i)) , F(x(i)) - __(i_1_,) } n n (4.1.3) where x(1) < \cdot < X(n) is the ordered observed sample, that is, the order statistics. The study of the model based on the minimal assumption of randomization is complicated and further conceptual issues arise. Cannichael, in proofreading the final version, caught more mistakes than both authors together. families moving. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher. Consider a point we noted in Example 2.4.2: For some coordinates 1, 11{ can be explicit. Furthermore, (Problem of eo,n under H, a consistent estimate of .: E-1 (80) is 2 2 1 (6.3.22) n-1 [-D,ln(90,n) + 1 [-D,D21ln(90,n)[D1 ln(B0,n] Ddn(90,n)]_ --... The function 1r represents our belief or information about the parameter () be fore the experiment and is called the prior density or frequencyfunction. (b) Using Problems B.3.12 and 8.3.13 show that the power (3(01, 0,) is an increasing function of 8f + B. Thus, in the preceding P(>.) case, r5 VA ± 2z(- la) 2 vn: is an approximate 1 - a confidence interval for J>... Bickel, Kjeli A. Instead of keeping track of several numbers, we need only record one. The alternative is that F.x(t) ! F(t) for some t E R. Two-Sided Tests : We begin by considering K fl-o J1- i- fl-O. This corresponds to the alternative "The treat ment has some effect, good or bad." However, as discussed in Section 4.5, the test can be modified into a three-decision rule that decides whether there is a significant positive or negative effect. Feynman, R., S%)T. Hint: Use (B.7.3) and let U1 = I, U2 = I{U E (0, \clubsuit)}, Ua = I{U E (0, \clubsuit)}, Ua = I{U E [0, m. 'i' \clubsuit , ..., (1.2.1), N. More generally only polynomials of degree n in p are unbiasedly estimable. p, ::; f.L o . (a) What test statistic should he use if the only alternative he considers is that the die is fair? Using this interval and (4.5.3) we obtain the following three decision rule based on T = .JTi(X a: o)/ ' 1 Do not reject H : I' • I' c if ITI < z(1 - !a). Ma jor differences here are a greatly expanded treatment of maximum likelihood estimates (MLEs), including a complete study of MLEs in canonical k-parameter exponential fam ilies. In this case the likelihood is a product of independent binomial densities, and the MLEs of 1ri and J.Li are statistic 2 log. A for testing H : ji is the MLE of JL for 11 11 E w versus E w. We can design a clinical trial, petfonn a survey, or more generally construct an experiment that yields data X in X C Rq, modeled by us as having distribution P9, 0 E e, where 8 is partitioned into {eo, e!} with eo and e, conesponding, respectively, to answering "no" or "yes" to the preceding questions. Our appendix does give all the probability that is needed. What is the MLE?, Xn be independent normal random
variables with - common mean and 2 = I: X; - X(m))2. 'Xn = Xn is that of (Y, Zt, . A Review, "Ann. It follows that the level and size of the test are unchanged if instead of 80 • {80) we used 60 • [0 80]. - -= -= 8(} l dx for all (J whereas the continuity (or even boundedness on compact sets) of the second inte gral guarantees that we can interchange the order of integration in (4) The finiteness of Var8 (T(X)) and 1(8) imply that 1/J'(8) is finite by the covariance interpretation given in (3.4.8). A level a test of H : J.L = J.Lo vs K ' J.1. > J.l.o rejects H when ...fii(X - J.l.o)/u > z(1 - a). Testing. Example 4.6.1 (Examples 3.3.2 and 4.2.1 continued). Chapter 3 Measures of Performance E(ii(X)] 9) 9, E(9] X) (9) a t (1;; (9) - 1(9)'1/(9)) Suppose V rv X According to Theorem B.3. 1, V has the same distribution as E 1 X[, where the Xi are independent and Xi "" N(O, I), i = I, . i Section Ll Models, Parameters, and Data, 11 Statistics Finally, we give an example in which the responses are dependent. Let X ! * Z N(0, 1) and V X X + Now use Problems 8.2.4 and Fk, m • then provided - $\mathbf{\hat{v}}$ k < r < $\mathbf{\hat{v}}$ m. Show that the order statistics arc minimal sufficient when f is the density Cauchy f(t) $\mathbf{\hat{v}}$ 1 /Jr(1 + t2). f(en - f3en-1)Because ei = Xi - Jl, the model for X 1, . Now suppose that a sample of 19 has been drawn in which 10 defective items are found. The result follows. - ,n $\mathbf{\hat{v}}$). , dg of Table 1.3.3. . Here the errors e1, . Other theorems are available characterizing larger but more manageable classes of pro ' cedures, which include the admissible rules, at least when procedures and l is quadratic Joss, the solution is given in Section 3.2. In the non-Bayesian framework, if Y is postulated as following a linear regression model with $E(Y) = zT\{3 \text{ as in Section 2.2.1, then in estimating a linear function of the j3]}$ it is natural to consider the computationally simple class of linear estimates, S(Y) = L: 0 1 divi. This approach coupled with the principle of unbiasedness we now introduce leads to the famous Gauss-Markov theorem proved in Section 6.6. We introduced, in Section 1.3, the notion of bias of an estimate O(X) of a parameter q(B) in a model $P = \{Po: 0 \in 8\}$ as Bias $9(5) = E05(X) - ' \cdot i + q(B)$. Here's another important special case. (b) Show that $Y = \Phi Ac'' + i + q(B)$. and assume initially that u2 is known. The task of finding a critical value is greatly simplified if Ce(T(X)) doesn't depend on 0 for 0 E eo. 9 • . More generally consider the following class of situations. If we write Sn con Sn -np = ., fii (Sn P) .jp(! - p) . test statistic can have a fixed distribution fo under the hypothesis. Dispersion, Ann. and also increases to 1 for fixed () E 61 as n oo. 2s. Let T denote a survival time with density fo(t) and hazard rate ho(t) = The Cox proportional hazanl model is defined as h(t I z) = h0(t) exp {g(j3, z) fo(t)/P(T > t). There is an even greater range of viewpoints in the statistical community from people who consider all statistical statements as purely subjective to ones who restrict the use of such models to situation given (I = 8. """" ho(t). It can also be thought of as a limiting case in which N = oo, so that sampling with replacement replaces sampling without. When dependence on 8 has to be observed, we shall denote the distribution corresponding to any particular parameter value () by Po . A prior for which the Bayes risk of the Bayes procedure equals the lower value of the game is called least favorable. If we suppose there is a single numerical measure of the drugs and the difference in performance of the drugs for any given patient in a complex manner (the effect of each drug is complex), we have to formulate a relevant measure of the difference in performance of the drugs and decide how to estimate this measure., Yn) such that P[t_::; Y ::; * 1 - a). is unknown and Xi - f.J, is symmetrically distributed about 0. Asymptotic theory for estimates and tests If (Z1 , Y1) , . (beta, il(VB, 1), density about 0. Asymptotic theory for estimates and tests (e) $f(x, 8) = (x/82) \exp\{-x2/282\}, x > 0; 8 > 0.$ (We write U[a, b] to make p(a) = p(b) = (b - a) 1 rather than 0.)' - 14, If n extsts. 0, i' ! ' 1 l i • $q(y \mid x)$ (;) B(r + x + y, s + n - x + m - y)/B(r + x, s + n - x) where $B(\cdot, \cdot)$ denotes the beta function. (* 0, if Tn < 0 H : a2 < a versus K : a2 > a5.16) to the random variables We get and 8f8Blogp(X, B) T(X). That a treatment has no effect is easier to specify than what its effect is; see, for instance, our discussion of constant treatment effect in Example 1.1.3. In science generally a theory typically closely specifies the type of distribution P of the data X as, say, P = Po, B E 80. (3.5.2) Here h is the density of the gross errors and .\ is the probability of making a gross error. The Hodges-Lehmann (location) estimate XHL is defined to be the median of the 1n(n + 1) pairwise averages P(Xi + xj). 1 suppose returning a shipment with () < 00 defectives results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defectives results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold results in a penalty of s dol lars whereas every defective item sold re are vectors, examples of loss functions are l(O, a) l(O, a) 1(0, a)suppose we observe S = EXi, the number of recoveries among the n randomly selected patients who have been administered the new drug. is a scale parameter for the r(p, .) family. First consider situation (a), which we refer to as: Example 1.1.1. Samp ling Insp ection. 15.8) implies (A. The interpretation of l(P, a), or I (0, a) if P is parametrized, is the nonnegative loss incurred by the statistician if he or she takes action a and the true ... state of Nature," that is, the probability distribution producing the data. ,A6 when 8 log p(x, 8) , and Q = P. (c) An experimenter makes n independent detenninations of the value of a physical constant p,. To specify this set more closely the critical constant treatment effect assumption is often made. Let x1 and x2 be the observed values of X1 and X2 and write x = (x1 + x2)/2. Now suppose we want to estimate $B(P^*) X_1$, we need a class of distributions that concentrate on the interval (0, 1). A placebo is a substance such as water t]lat is expected to have no effect on the disease and is used to correct for the well-documented placebo. Would you accept or reject the hypothesis of independence at the 0.05 level (a) using the x 2 test with approximate critical value? These examples motivate the decision theoretic framework: We need to (I) clarify the objectives of a study, (2) point to what the different possible actions are, (3) provide assessments of risk, accuracy, and reliability of statistical procedures, (4) provide guidance in the choice of procedures for analyzing outcomes of experi ments. Because D Fe(t) is a distribution function, 1 - Fo(t) is decreasing in t. However if, say, J.£1 = p,0 + Ado,)..., Xn be a sample from some continuous distribution Fwith density J, which is unknown. We want to study of Example 4.9.3. if n , n2 ---+ oo. Let q(z j y) denote the conditional frequency function of Z given Y y. Variance Stabilizing Transfo111U1tion for the Binomial Distribution. 2 16. 1 (b) Check that your test statistic has greater expected value under H. Response measurements are taken on the treated and control members of each pair. In statistics, the gamma density 9p, >... Use Theorems 1.6.2 and 1.6.3 to obtain moment-generating functions for the sufficient statistics when sampling from the following distributions. p(x I B) the posterior density , {N, X1, Chapter 1 ...-(B I x). Mx(s) = L k! k=O are (A.12.3) 460 I i I A Review of Basic: Probability Theory A.12.4 The moment generuting function .Hx has derivatives of all orders at :; dk : [J. Some of the material in this appendix, as well as extensions, can be found in Anderson (1958), Billingsley (1995), Brei man (1973). Note that Sf V (U2, V2), . Find E(X I Y) when X • = I Z). then is a subset of X with probability at least P[X E A(v0)] > 1 a for all P E Pv, if and only if S(X) is a 1 - a confidence region for v. P, (X) < 8' < 0 (X) < 8' < 0 (X) < 8' < 0 (C) (X) < 8' < 0 (A oversus K : >. , Xk be independent Xi N(Bi, a2)
where either a 2 = a5 (known) and Bt, . see Birkhoff and B.l0.1.3 If A is spd, so is A - I... • 'I I and II," Ann. • (8.2. 1 1) for O < x < 1, where B(r, s) = [r(r)r(s)]/[r(r+s)] is the betafunction. We give a few examples of situations of the foregoing type in which it is used, and its main properties. The basic global comparison criteria Bayes and minimax are presented as well as a discussion of optimality by restriction and notions of admissibility. Z] X], g(t2Z > (Markov' s > g(t] X]) inequality, Proof of (A.15.4). We may want to assume that B has a density with maximum value at 0 such as that drawn with a dotted line in Figure B.2.2. Or else we may think that 1r(B) concentrates its mass near a small number, say 0.05. k = 1, . The progeny exhibited approximately the expected ratio of one homozygous dominant to two heterozygous dominants (to one recessive). How we use statistics in estimation and other decision procedures is the subject of the next section. Statistical methods for models of this kind are given in Volume 2.10.2). Use calculus. """ - That is, how do we find a function B(X) of the vector observation X that in some sense "is close" to the unknown 8? We illustrate such simulations for the preceding t tests by generating data from the X distribution M times independently, each time com puting the value of the t statistics and then giving the proportion of times out of M that the t statistics exceed the critical values from the t table. (4.8.3) We can think of the test statistic = = • • • • 2 log .X. Show that the sample median X is an empirical plug-in estimate of the population median v. The following result, which follows from Chebychev's inequality, is a useful general ization of Bernoulli's law. is the standard normal distribution function and nr I l = n + r + s x + n + n + s x + n +distribution. This selection is based on experience with previous similar experiments (cf. J., The Foundations of Statistics, J. Note first that, if A is symmetric, (B.! 0. F.,..... = r1, 5. It follows that all the properties of the expectation given in (A. Consider the following competitor to X: - Bn I - 0 if IXI < n-1/4 (5.4.37) We can interpret this estimate as first testing H : () 0 using the test "Reject iff JXI > n-1/4" and using X as our estimate otherwise. Other novel features of this chapter include a detailed analysis including proofs of convergence of a standard but slow algorithm for computing MLEs in multiparameter ex ponential families and ail introduction to the EM algorithm, one of the main ingredients of most modem algorithms for inference. (b) Deduce that the random variables X and Y in Problem B.1.8(c) (i) have correlation 0 although they are not independent., Xn are i.i.d. as X "' F, where F is a continuous distribution function with positive density f on (a, b), - oo :S: a < b :S: oo Hint: From continuity of p, (i) and the dominated convergence theorem, . As is typical, we call Y 1 a "success" and Y = 0 a "failure." We assume that the distribution. If we take a sample from a member of one of these families, then the sample mean X will be approximately normally distributed with variance cr2 /n depending on the parameters indexing the family considered. (i.e., J is an unbiased estimate of f.J.). 1 1) For a proof., en, Using conditional probability theory and ei = /3ei-I + f.i, we have . 23, 1443-1473 (1995). Let X have a N(0, 1) distribution. r Hint: Use Problem B.3.7 for r odd. The distribution of S is called the F distribution with k and m degrees offreedom. (2) We can derive methods of extracting useful information from data and, in particular, give methods that assess the generalizability of experimental results. This can be made precise for stochastic GLMs obtained by conditioning on Z1, . Testing and Confidence Regions 250 C h a pter 4 We begin by finding a uniformly most accurate level (1 - a) UCB). The key feature of situations in which .Co(Tn) = .C0 for () E 80 is usually invariance under the action of a group of transformations. It is not our goal in this book to enter seriously into questions that are the subject of textbooks in numerical analysis. We present an elementary discussion of Bayesian models, introduce the notions ofprior and posterior distributions and give Bayes rule. (g l (y))l (B2A) the Jacobian of g is just its derivative and the requirements (i) and (iii) that g' be continuous and nonvanishing imply that monotone and, hence, satisfies (ii). serious error in Problem 2.2.5 was discovered by A F. The shape of F is represented by the equivalence {F((· - a)fb) b > 0, a E R). If X is normally distributed, Cj = 0 for j > 3. Here is an example. Show that the second- and higher-degree cumulants (where p = �=1 ii invariant under shift; thus. In most studies we are interested in studying relations between responses and several other variables not just treatment or control as in Example 1.1.3. This is the stage for the following: (1) Describe the basic concepts of mathematical statistics indicating the relation of theory to practice. or (3) market research where a potential customer either desires a new product (Y = 1) or does not (Y 0). B.), O < B; < 1, 1 < j < k. 15. We view statistical models as useful tools for learning from the outcomes of experiments and studies. 10.1 Symmetric Matrices used in the text and 8.6. Recall Appr is symmetric iff A = AT. Then the posterior distribution of ll is by (1.2.8) 8 I (oi 1 x = o) = ...(e, 1 x =account. Problem Set 1 (PDF) Problem Set 3 (PDF) Problem Set 3 (PDF) Problem Set 5 (PD level. Change other coordinates accordingly. The joint distribution of Z and Y can be college admissions situation, and Y his calculated (or rather well estimated) from the records of previous years that the admissions I ' officer has at his disposal. • 14. We also expect to discuss classification and model selection using the elementary theory of empirical pro cesses. More generally, if T1 and T2 are any two statistics such that 71 (x) = T1 (y) if and only if T2 (x) = T2(y), then Tt and T2 provide the same reduction of the data. Hint: By Problem B.2.12, So (T) has a U(O, I) distribution; thus, - log S0(T) has an exponential distribution. denoted by Corr(X1, X2). J = I)r) $\hat{\Psi} = "\cdot 4$. A.16.2 Formally, let (N(t)]. Xn) In fact, we can show that X(n) is sufficient. If we let f(Yi I Zi) $\cdot i = 1$ If we let J.L(z) denote the expected value of a response with given covariate vector z, then we can write, (b) where Ei = Yi - E (Yi). We next define and examine the sensitivity curve in the context of the Gaussian location model, Example 1.1.2, and then more generally. I Note that Theorem 5.4.3 generalizes Example 5.4.1 once we identify '1/J(x, B) with T(x) - A'(B). The complete result is established for instance by Lehmann (1997, Section 2.6). This class of distributions has the remarkable property that the resulting posterior distributions arc again beta distributions. Note that these tests can be carried out without knowing the density qo of z . , {3p)T of unknowns. R(8,6) sPo[X < k], 8 < 80 rN8Po[X < k], 8 < 80 xdF(x). 14) I I If Xi and Xf are distributed as X1 and X2 and are independent of X1 and X2, then $e^{Cov(X1, X2)} = E(X1 - x;)(X, -x;) = 0$ to terms (b) Use (a) to justify the approximation • 17. What is observed, however, is not X but S where S; s, xi, 1 < i < m ($\in i1 + \notin i2$, $\notin i3$), m + 1: S i q(s, Bold) - (2.4.13) Equality holds in (2.4.13) iff the conditional distribution of X given S(X) = s is the same for Bnew as for Bold and Bnew maximizes J(B I Bold l. Proof. The decision theoretic framework accommodates by adding a component reflecting this. , Xn are observable and we want to predict Xn+t. Give a level (1 - a) prediction interval for Xn+t. Hint: XdB has a � distribution and nXn+t! E � 1 xi has an F2,2n distribution. and T3 = 2 (Y (a) Why can you conclude that T1 has a smaller MSE (mean square error) than T2 ? Thus, if we have three air conditioners, there are 3! = 6 possible rankings, • . Notation. Chapman and Hall / CRC, 2015. Figure 5.3.1 shows that for the one-sample t test, when o: = 0.05, the asymptotic result gives a good approximation when n > 10 1 ·5 32, and the true distribution
F is x with d 2:: 10. Begin by specifying a small num ber a > 0 such that probabilities of type I error greater than a arc undesirable. 1 7 Data, Models, Parameters, and Statistics Dual to the notion of a parametrization, a map from some e to P. A college admissions officer has available the College Board scores at entrance and first-year grade point averages of freshman classes for a period of several years. Let X have a binomial, B(n,p), distribution. The extent to which holes in the discussion can be patched and where a binomial, B(n,p), distribution. The extent to which holes in the discussion can be patched and where patched and where patched and where a binomial, B(n,p), distribution. The extent to which holes in the discussion can be patched and where a binomial, B(n,p), distribution. v < oo. ::; n = Further qualitative features of these bounds and relations to approximation (5.1.8) are given in Problem 5.1.4. Similarly, the celebrated Berry-Esseen bound (A. = (A. In this case, J(B I Bo) = Eo ((B - Bo)TT(X) - (A(B) - A(Bo)) I S(X) = y) - (A(B) - A(B)) I S(X) = y) - (A(B) - A(B))exhibits from the usual probability model is that NO is unknown and, in principle, can take on any value between 0 and N. Give p-values for the cases. There is a statistic T that "tends" to be small, if H is false. Let Yi be independent binomial, B(n;, .\;), 1 < i < n. For the first volume of the second edition we would like to add thanks to new col leagues, particularly Jianging Fan, Michael Jordan, Jianhua Huang, Ying Qing Chen, and Carl Spruill and the many students who were guinea pigs in the basic theory course at Berkeley. Researchers have then sought procedures that improve all others within the class. • WALLACE, C. we have E [(Y - q(Z)) 2 I Z • Z] • E[(Y - q(Z)) 2 I Z • Z] Z 🕏 z]. In cluded are asymptotic normality of maximum likelihood estimates, inference in the general linear model, Wilks theorem on the asymptotic distribution of the likelihood ratio test. The risk points (R(Bt, oi), R(O,o;)) are given in Table 1.3.4 and graphed in Figure 1.3.2 for i = 1, . 1.5. (b) Suppose we measure the difference between the effects of A and B by \clubsuit the difference between the quantiles of X and -X, that is, $vp(p) = \clubsuit [xP + x1 P]$, where p = F(x). we now define the elements of a statistical model. Then = $\clubsuit P[IZ + ...fiiB] - (-n1i4 - ...fiiB) - (-n1i4 - ...fiiB) + (-n1i4$ section we will consider Bernoulli responses Y that can only take on the values 0 and 1. Decide I' < l'o ifT < -z(t - ia). In this parametric case, how do we select reasonable estimates for 8 itself? respectively, (Xo, Yo) is in the interior of S x T, (a) Show that a necessary condition for (xo, y0) to be a saddle point is that, representing $X = (x_1, ..., X_{n-1})$. X1, . lo. JI \circ c2, xi and x \circ but with probabilities \circ - \circ , \circ and \circ where (c) Let (X, Y,) have an N, (e, e, u10' \circ o. For r even set m = r/2 and note that because Y =](X - J,t)/o-]2 has a xi distribution, we can find E(Y'') from Problem 8.2.4. Now use E(X - J,t)' = o-r E(Ym), 9. (b) Show that if n = 2 the most powerful level .0196 test rejects if, and only if, two 5's are obtained. Show that for a sufficiently large the likelihood function has local maxima between 0 and 1 and between p and a. For instance, if we assume the covariates in logistic regression with canonical link to be stochastic, we obtain = If we wish to test hypotheses such as H : /31 = · · · = /3d = 0, d < p, we can calculate i = 1 where {3 H is the (p x 1) MLE for the GLM with f3 x 1 (0, . Hint: Without loss of generality, take cr x in probability to F(x). Example 4.3.2 (Example 4.3. >...i are nonnegative. WIJSMAN, R. Alternatively, i f Y has a second moment, we may select F as being the unique member of the family Fs having 2 Var Y = 1 and then X "' F; ==? (iii) Z has a U(-1, 1) distribution, Y (v) Z has a U(-1, 1) distribution, Y = Z2 = Z2. n-vn[h(X) - ho] s[h(1)(X)[' ! ' ' Fortunately, the methods needed for its analysis are much the same as those appropriate for the situation of Example 1.1.3 when F, G are assumed arbitrary. an, ' • 462 A Review of Basic Probability Theory If X has an H(D, N, n) distribution, then D E(X) = n ' Var X = N D nN (D!N) N-n N 1. Then, by (1.2.8), the posterior probabilities are r;p(x I O;) $1 P[9 \diamondsuit 0; I X \textcircled{9} x] \textcircled{9} Ei'lrj p(x I e,) and, thus, T (aj I X) = E; w; j7r; p(x I 0;)$, Yn be independent responses and suppose the distribution of Yi depends on a covariate vector zi. KENDALL, M. 1 26 .253 .385 .524 .674 .842 1 .036 1.282 .09 .08 .07 .06 .05 .04 .03 .025 1.341 1.405 1.476 1 .555 1 .645 1.751 1.881 1.960 .02 .01 .0005 .0001 .0005 .0001 .0005 .0 .0000 1 2.054 2.326 2.576 3.090 3.291 3.719 3.891 4.265 Entries in the top row are areas to the right of values in the second row. Usually, the situation F. We give explicitly the construction of exact upper and lower confidence bounds and intervals for the parameter in the binomial distribution. Let Yi XI - x2 and y2 = x2. (s) \$ Is \$ - ---cc = Appendix A = 0 and E(X'). Let Example 1.1.5. X 1, Show that if m = .\k for some .\ > 0 and + \$ (J2X y'2ri} (c) Compare the approximation of (b) with the central limit approximation of (b) with the central limit approximation of (b) with the central limit approximation of (c) with the central limit approximation of (b) with the central limit approximation of (b) with the central limit approximation of (b) with the central limit approximation of (c) with the central limit approximation of (b) with the central limit approximation of (b) with the central limit approximation of (c) with the central li XQ.99, n = 5, 10, 25. 1 2, given x1, , Xn, l: (xi - p,0) 2 . 1.3). B Then (3.4.10) • and, thus, I(B) Var Proof. I - (xfc)-8, x > F (x O), c x 0 and c = 2,000 is the minimum monthly salary for state workers. V., Introduction to Probability and Statistics from a Bayesian Point of View, Part II: Inference, Cambridge University Press, London, 1965. This assertion may be proved as follows. N(O, ho the natural test statistic is T. [f Y is real, and z E Z, A {a : a is a function from Z to R} with a(z) representing the prediction we would make if the new unobserved Y had covariate value z. If X and Y are independent, then $c_i(X + Y) = c_i(X) + c_1(Y)$. > 1 - ,\, the interval (4 9 3) has asymptotic probability, . A second application occurs for models where the families of distribution for which variance stabilizing transformations exist are used as building blocks of larger models. BOHLMANN, H.,
Mathematical Methods in Risk Theory Heidelberg: Springer Verlag, 1970. This is the critical value we shall use, if our test statistic is T and we want level a. 2nd edition. Let p1 denote the N(O, 0, 1, 1, 0) density and let P2 be the N(O, 0, 1, 1, p) density. Measurement Model with Autoregressive Errors. Show that, however, (ii) Var (! logp(X,) 8) (iii) 2X is unbiased for " ' 2. F I u), and by I - a). Suppose that in Example 1.3.5, a new buyer makes a bid and the loss function is changed to 8\a a, a, a 3 0 12 7 I 4 6 o, (a) Compute and plot the risk points in this case for each rule (h, . , (Zn , Yn) can be viewed as a sample from a population and the link is canonical, the theory of Sections 6.2 and 6.3 applies straightforwardly in view of the gen eral smoothness properties of canonical exponential families. 1 0. or repair it; we could drill for

oil, sell the location, or sell partial rights? we could operate, administer drugs, or wait and see. For some specified vo, H may be rejected., Xn) is given by (see (A.16.4)), . Let us call the possible categories or states of the first characteristic A and A and of the second B and B. If (} and (}* are two competing level both very likely to fall below the true B. An individual either is or is not inoculated against a disease; is or is not a smoker; is male or female; and so on. (a) Show that if F and G are N(/Lt afl and N(!L2, aD, respectively, then the LR test of H : uf = a versus K : af # 1)- 1 2::7' 1 (1-j Y)". lf8 < Bo, Po[Bn > en(, Oo)] 0., Yn. The only truly nonparametric but useless model for X E R n is to assume that its (joint) distribution can be anything. lfwe put X1 = X, = X in (A., z(k)) is RP (c) P[Z 1 • zUI] > 0 for all j. Assume that : {} < 0 versus K : 8 > 0. Most simply we would sample n patients, administer the new drug, and then base our decision on the observed sample X = (X1, Show that (6.3.19) holds., Yn2 be i.i.d. C, and suppose the X's and Y's are independent. But cr 2 Var(X1) evidently is and so is I' + t.. 1' 85 Section 1. After Bellman (1 %0, p. Thus, with t1 = 2:: x, and t2 = 2:: xf, Equation (2. q(B) the conditions of Theorem 3.4.1 hold and T an unbiased CoroUary Suppose estimate of B. be i.i.d. ! fo (x * r), a > 0, J.l E R, and assume for w W11 > 0 so that w is strictly convex, w (±00) = oo. 1 1 6.25 7.81 9.35 9.84 1 1 . 34 12.84 14.32 16.27 17.73 4 5.39 7.78 9.49 1 1 . Y - ll-2) has a x distribution. However, by giving () a distribution purely as a theoretical tool to which no subjective significance is attached, we can obtain important and useful results and insights. Suppose l(B; , j) = 0, I, j = 04.5.2. Suppose X1, . Statlab: An Empirical Introduction to Statistics. A.15.12 The first of these results reflects the intuitively obvious fact that if the populations sampled are large and the samples are comparatively small, sampling with and without replacement leads to approximately the same probability distribution. = Z2 if Z2; . 7 + I + 2. Show that the maximum Jikelihood estimate of 8 based on Y1 , , Yn .IS • - B(Y) • "" • . (a) Show that the conditional distribution of X M(k, 1/n, . 6 2 • 8 • 0 5 0 10 5 R(B1 , 6), Figure 1,3,2. , Zn in the sample (Z1 , Y1) , . ' '' I ! 86 9. ' I ' (c) Let x• be a specified number with 0 < F(x') < I. ROSENBLATT, Statistical Analysis of Stationary Time Series New York: Wiley, 1957. Given f continuous on (a, b), f i strictly, f(a+) < 0 < f(b-), then, by the intermediate value theorem, there exists unique x^*c :: (a, b) such that $f(x^*) = 0$. (t [1 - 1*, "*f] l). (ii) are uncorrelated., Xn). The rank of A is the number of nonzero eigenvalues. n ' ' (c) Conclude that with probability tending to 1, II. 3 . 1 we considered linear models that are appropriate for analyzing continuous responses {Yi} that are, perhaps after a transformation, approximately normally distributed and whose means are modeled as J-li = E; = I Zij {3 j = z f {3 for known constants { Zij } and unknown parameters f3I , . Suppose that we can write fiT = (nf, . (Use the central limit theorem.) 4. is that of a parameter, formally a map, v, from P to another space N. This would lead to the prior distribution B, x, "' = (1 *0) (0.1) {0.9}100-i, (1 .2.4) for i = 0, 1, . Chapter 1 In this section we introduced the first basic notions and formalism of mathe matical statistics, vector observations X with unknown probability distributions P ranging over models P. tp1 ' (8.2.9) b, ,,(x) = B(r, s) ' 1 ' • (8.2. 10) The family of distributions with densities given by (B.2.8) is referred to as the gamma family of distribution corresponding to 9p,>..., Xn are independently and identically distributed as X. the upper quartile X.1s. SHIBATA, R., "Boostrap Estimate of Kullback-Leibler Information" for Model Selection," Statistica Sinica, 7, 375-394 (1997). Example B.1.3 Suppose Y and Z have the joint frequency function of Table B.1. We find 5 11 1. V., Introduction to Probability; Part II: Inference London: Cambridge University Press, 1965. Wiley & Sons, 1968. Hint: 🔶 (i) Show that !(On) cau be replaced by 1(0). is large, a large a2 will force a large m) n to give us a good chance of correctly deciding that the treatment effect is there. As Rd as well as an in the first edition, we do not require measure theory but assume from the start that our models are what we call "regular." That is, we assume either a discrete probability whose support does not depend on the parameter set, or the absolutely continuous case with a density. Wiley & Sons, 1 954., Xn are indicators of n Bernoulli trials with probability of success () where 0 < 8 < 1. The random variable g(Z) is written E(Y I Z) and is called the conditional expectation of Y given Z. It turns out that, { N(t)} is a Poisson process with parameter A if and only if the following conditions hold: (a) N(t + h) - N(t) is independent of N(s), s < t, for h > 0, (b) N(t + h) - N(t) has the same distribution as N(h) for h > 0, (c) P[N(h) = 1] (d) P[N(h) = 1] (d) P[N(h) = 1] (d) P[N(h) = 1] (e) N(t + h) - N(t) has the same distribution as N(h) for h > 0, (c) P[N(h) = 1] (d) P[N(h) = 1] (e) P[N(h) = 1] (for h > 0, (c) P[N(h) = 1] (for h > 0show that the distribution under H does not depend on (b) When '1/J(u) = 1 and a: = 2. To have Var h(X) approximately constant in A, h must satisfy the differential equation [h < 1] (A) j 2 A = c > 0 for some arbitrary c > 0. 0.12,-... Hint: Show that the joint distribution of (X., .1]) for all a where b is E(U - E 12E22 V). Let $e = \{(81, 82) : 91 > 0,92 >$ 0,9, +92 < 1 and let 93 = 1 - (81 + 82). Example 2.4.4. Lumped Hardy-Weinberg Data. B.10.2.3 We note also, although this is not strictly part of this section, that if U, V are random vectors as previously (not necessarily Gaussian), then equality holds in (B.10.8) iff for some b (B.10.10) with probability I. j oo T(x) {) p(x, >) dxd>. This statistic has the following distribution free property: Proposition 4.1.1. The distribution of Dn under H is the same for all continuous Fo. In particular, PF, (Dn < d) = Pu (Dn < d), where U denotes the U(O, 1) distribution. At each stage with one that is, values of coordinate fixed, find that member of the family of contours to which the vertical (or hori zontal) line is tangent. = 2 [Z (Y, 77 (Y)) - 1 (Y, "' (IL o)] for the hypothesis that IL = /L o within M as a "measure" of (squared) distance between Y and /Lo . (6.5.5) is always 2:: 0. (b) Express x11 and :rp in terms of the critical value of the Kolmogorov statistics. Suppose XI '. * Show that E(Sm I Sn) = given Sn I; 1 X; k is multinomial (mfn)Sn 6. Then (B .9 .4) s = 2, HOlder's inequality becomes the Cauchy-Schwartz inequality (A.1]. Here xl, 15.13) leads us to assume that the number X of geniuses observed has approximately a B(n, 8) distribution. (b) We want to study how a physical or economic feature, for example, for example, a B(n, 8) distribution. height or in come, is distributed in a large population. Example 2.1.1. Least Squares. This can be remedied by considering $\hat{\phi}_{j} = \log(A_{j}/Ak) = ; - "' \cdot 1 s j s k - 1$, and rewriting k- 1 q(x, 71) = exp{T'[. 1) (x)'l - n log(1 + L e'')} = log(A_{j}/Ak) = ; - "' \cdot 1 s j s k - 1, and rewriting k- 1 q(x, 71) is a k - 1 parameter canonical exponential family generated by T(-1) k and $h\{x\} = fT$ 1 | [xi E { 1, ., Yn2 be two independent samples from N(Jli, $\pm Ti$), N(p.2, 0'). Show that if m and n are both tending to oo in such a way that m/ (m + n) ---> a, 0 < a < I, then (Bm, n - mj (m + n)) x v' < m + n P y'a(1 - a) ___, 'li (x) . •, Section 3.4 177 Unbiased Estimation and Risk Inequalities given by (see Example 1.3.3 and Problem 1.3.3 and Problem 1.3.3) It $22 \times s = X n (3.4.1) + 2 [I.: (X, -X)$. Wiley & Sons, SAVAGE, L. Ut0"2 = 0 and 10., B*) bisection in to get BJ for j = 1, . •, Xkr*; J and 1 "L., 7rkXk - K k=1 X is unbiased and if X is the mean of a simple random sample without replace- ment from the population then with equality iff Xk- = - VarX < Var X IJ; 1 2.:1" 1 Xki doesn't depend on k for all k such that 1rk (b) Show that the inequality between 2:: 2 = 2 and suppose, K 2 = -7rk = > 0.15 8.75 8.47 8.26 8.10 7.87 7.56
5.59 4.74 4.35 4.12 3.97 3.87 3.79 3.7 3 3.64 3.51 8,07 6.54 5.89 5.52 5.29 5.12 4.99 4.90 4.76 4.57 12.25 9.55 8.45 7.85 7.46 7.19 6.99 6.84 6.62 6.31 5.32 4.46 4.073\$4 3.69 358 3.50 3.44 3.35 3.22 7.57 6.06 5.42 5.05 4.82 4.65 4.53 4.43 4.30 4.10 1 1 .26 8.65 7.59 7.01 6.63 6.37 6.18 6.03 5.81 5.52 5.12 4.26 3.86 3.63 3.48 3.37 3.29 3.23 3.14 3.01 7.21 5.71 5.08 4.72 4.48 4.32 4.20 4.10 3.96 3.77 10.56 8.02 6.99 6.42 6.06 5.80 5.61 5.47 5.26 4.96 4.96 4.10 3.7 1 3.48 3.33 3.22 3.14 3.07 2.98 2.85 6.94 5.46 4.83 4.47 4.24 4f17 3.95 3.85 3.72 3.52 10.04 7.56 6.55 5.99 5.64 5.39 5.20 5.06 4.85 4.56 4.75 3.89 3.49 3.26 3. We have the following. We call they's treatment observations. If X1 and X2 are random variables and i, j are nat ural numbers, then the product moment of order (i,j) of X1 and X2 is, by definition, E(XjX). Proof We compute the posterior density of .,fii(O - 8) as (5.5.10) where Cn = Cn(Xl, . In the two quarter courses for graduate students in mathematics, statistics, the physical sciences, and engineering that we have taught we cover the core Chapters 2 to 7, which go from modeling through estimation and testing to linear models. Given tolerance c:: > 0 for lx ti nal - x*l: Find xo < x , f(xo) < 0 < f(x !) by taking lxol, lxd large enough. However, even in :... If X A.ll.12 N (tl,a2), then 11 = 1z = 0., Xn = Xn, where I: 1X = k. Carry out a Monte Carlo study such as the one that led to Figure 5.3.3 using the Welch test based on Sn. Plot your results. Fujimura, and our families for support, encouragement, and active participation in an enterprise that at times seemed endless, appeared gratifyingly ended in 1976 but has, with the field, taken on a new life. It is convenient to distinguish between two structural possibilities for S0 and S1 : If 80 consists of only one point, we call S0 and H simple. Here we use the xJ distribution because for small to moderate d it is quite different from the normal distribution. we only observe We can thus write 1r(B I k) for 1r(B I x). Show that Xk given by (3.4.6) is (a) unbiased and (b) has smaller variance than X if $b < 2 \operatorname{Cov}(U,X)/\operatorname{Var}(U)$. Example 4.8.2. Suppose X1, . We have $E(Y, I z = i) = P | Y_{z} = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z = i | z =$ just the number of ways i successes can occur in n Bernoulli '1; • : • . Then the s if 8 < 80 O if8 > 80 (1.3.1) rN8. Because q is onto !1, for each w E !1 there is 9 E 8 such that w = q(9). a0 is assumed known a unique MLE for J] exists and uniquely (a) Show that if o solves * L, P i=t (b) Write 81 uniquely solves 1 a' e, ' (ao-!') Xi - 0. Finally, using (4.2.3), we have shown that Et[k] 'Pk (x) \bullet P0[L(X, Oo, OJ) = k] on the set {x : L(x, Oo, O,) = k}. Affine transformations IfP is the canonical family generated by M(T) = Mex kT + hex "I" ' it is easy to see that the family generated by M(T(X)) and h is the subfamily of P corre sponding to and 17(8) = MT8. 5,3 FIRST- AND H IGHER-ORDER ASYMPTOTICS: THE DELTA M ETHOD WITH APPLICATIONS We have argued in Section 5.1 that the principal use of asymptotics is to provide quantita tively or qualitatively useful approximations to risk. Moreover, if a1 + a2 < 1, then [.Qa1, Ba2] is confidence interval for () with confidence coefficient 1 - (o t + a2). AND M. The LR n - w is denoted by D(y, ji), where TJ and (6.4. 12) k 2 I) xi log(Xdfli) + XI log(XI/ flDJ i = 1 (6.4. 17) x: mi Xi and fl is mi Mi · D (X, ji) measures the distance between the fit ji based on the model w and the data X. Show that 'I n ! 'I''. The central product moment of order (i,j) of X, and X2 is again meter Case Chapter 6 with row and column sums as indicated. Moreover, v equals the Bayes risk of the Bayes risk of the prior 1r*. In the regular cases we study this will not be a problem. To get a Bayesian model we introduce a random vector 9, whose range is contained in 8, with density or frequency function 1r. Then lf8 > Bo. Po[Bn > c,(a,Oo)] • 1. Whichever system you buy during the year, you intend to test the satellite 100 times. See Problem 4.8.5 for a simpler proof of (4.8.2). Suppose that as in Theorem B.7.6, Fn(x) F(x) for all x, F is continuous, and strictly increasing so that p-l (n) is unique for all 0 < n < 1. Let X Xrn; Y1, . 1.3.2 • Comparison of Decision Procedures In this section we introduce a variety of concepts used in the comparison of decision proce dures. and H is the hypothesis that the drug has no effect or is detrimental, whereas K is the alternative that it has some positive effect. Suppose that (X, Y) have the joint density 1 1 p(x, y) • 2 pi (x, y) + 2 pz(x, y). are : S k < min(n, D). The expression in (c) is, for j < n/2, bounded by tl ... (b) Let Sn X \bullet · The following approximation to the distribution of Sn (due to Wilson and Hilferty, 1931) is found to be excellent ,..... (b) Suppose Sm = { x : X; > 0, I < i < m , I; " 1 X; = I}, the simplex, and g(x,y) = E I CijXiYJ with x E Sm. y E Sp. Show that the von Neumann minimax theorem is equivalent to the existence of a discuss the Bayes and minimax criteria. Simpler assumptions can be formulated using Lebesgue integration theory. It is often convenient to identify the random vector X with its realization, the data X(w). Then each step of the iteration both within cycles and from cycle to cycle is quick. Let Nii be the entries of an a x b contingency table with associated probabilities Bij and let 1 i 1 = E = 1 () ii 1 = E and Performance Criteria ' Chapter 1 family of all distributions according to which X1,..., Xn are independent and identically distribution. > 0, IPI < 1, then • has a 🏶 distribution. > 0, IPI < 1, then • has a 🗘 distribution. > 0, IPI < 1, then • has a 🖗 distribution. En are independent N(O, a2)., \mathbf{O} (b) Show that when F and .Fk,m distribution with k = n1 G are nonnal as in part (a), then (si/af)f(s a b a c) + r supR(B, &') = r supR(B, &') = r + o(1) 8 where o(1) \mathbf{O} as k \mathbf{O} oo. We also by example introduce the notion of a conjugate family of distributions., Xn) x1 = m for all n. It follows that = B .3 . 2:: . L 'l'(X,, 8)., Xn are the interarrival times for n customers, then the joint density of (X 1, . BAXTER, J., (X_1, X_2)) T which is just the prediction vector Testing. Then the MLE of u2 is (]-2 = $(X_1 - X_2)$ Although H : u = uo is now composite, the distribution of Tn = ni:T2/u5 is x; 1, independent of JJ.. It next presents basic asymptotic approximations with one-dimensional parameter models as example 1.1.1. For instance. f(r + 1, B) � 1 - L Bk-1 (1 - B) � B'. the Wald statistic based on the parametrization 8 (17) obtained by replacing Boj by ej (ij), is, by the algebra of Section 6.4. 1, also equal to Pearson's x2 •, Methods for Discrete Data Section 6.4. Example 6.4.4. Hardy-Weinberg. can be any positive number. Similar conclusions = 1 + P - I.I + 1. 416 I nference i n the M u ltipara m eter Case Chapter 6 follow for the Wald and Rao statistics. The function Theory for Transformations of Random Vectors Upon substituting these quantities in (B.2.3), we obtain Px G Pv(YI, Y2) (Y1 + y,), r $(Y1 - Y_2)$ I exp - 32 I::>y i + 5y2 + 6YIY2 I ., xdf and [xl is Euclidean distance, then there exist universal constants 0 < Cd < cd < 00 Such that cdlx l l < l xl < Cd 1 x! J., x N}, then . (3) A function g is said to be one to one ifg(x) (4) Strictly speaking, (X, Y) and g(y) implies x y., Xn) 8. 7.5. P 9. Save for these changes of emphasis the other major new elements of Chapter 1, which parallels Chapter 1, which parallels Chapter 2 of the first edition, are an extended discussion of prediction and an expanded introduction to k-parameter exponential families. Specifically, instead of beginning with parametrized models we include from the start non- and semiparametric models, then go to parameters and parametrized models we include from the start non- and semiparametric models, then go to parameters and parametrized models we include from the start non- and semiparametric models at each set of beginning with parameters and para I}, B E 9, that are not 0-1. Notes for Section 1.6 (I) Exponential families arose much earlier in the work of Boltzmann in statistical mechan ics as laws for the distribution of the states of systems of particles-see Feynman (1 963), for
instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of systems of particles-see Feynman (1 963), for instance. Zk) where the marginal distribution of the states of s easy to point to in practice but remarkably difficult to formalize ap propriately. • • rr(z" 1, lp range freely and • p. 1 1) states that if oo, EF]Xd' Ej : 0 for all $\in > 0$. In the "life testing" problem 1.6. 16(i), find the MLE of B. Deciding which statistics are important is closely connected to deciding which parameters are important and, hence, can be related to model formulation as we saw earlier. • ' '' • t ' Rejecting H for L large is equivalent to rejecting for I 1 • .. A parameter is a feature v(P) of the dis tribution of X. (See Problem 4.1.10.) !, • I • 12. and the IQR. (5.4.46) i • 334 Asymptotic Approximations Chapter 5 l By (5.4.41), .jni(8)(c.,(a, 80) - 8) and , oo if 8 80 80. We shall illustrate some of the relationships between these ideas using the following 8 has two members, A has three points. For instance, if Y 0 or 1 corresponds to, say, "does not responds to, say, "does not responds," respectively, and z = (Treatment, Sex)T, then a(B, M) would be our prediction of response or no responds to, say, "does not responds," respectively, and z = (Treatment, Sex)T, then a(B, M) would be our prediction of responds to, say, "does not responds," respectively, and z = (Treatment, Sex)T, then a(B, M) would be our prediction of responds, "respectively, and z = (Treatment, Sex)T, then a(B, M) would be our prediction of responds to, say, "does not respond to, Similarly, if the value of the distribution function F is not specified outside some set, it is assumed to be zero to the "right" of the set and one to the "right" of the set. 064. (3) The possibility of implementing computations of a magnitude that would have once been unthinkable. matched pair experiments . = 546 Tables Appendix C Pr(F > f) f Table IV F distribution critical values Pr(F > f) 0.05 ,, I I 2 3 4 5 161 199 216 225 648 799 864 4052 4999 18.51 r, 6 7 8 10 15 230 234 237 239 242 246 922 937 948 957 969 985 5403 900 5625 5764 5859 5928 5981 6056 6157 19.00 19.16 19.25 19.30 19.33 19.35 19.37 19.40 19.43 $38.51\ 39.00\ 39.17\ 39.25\ 39.30\ 39.33\ 39.36\ 39.37\ 39.40\ 39.43\ 98.50\ 99.00\ 99.$ (Define the > q.) X1, . , T(X(8)), has level a if Co is continuous and (B + 1) (1 - a) is an integer (Problem 4.1.9). It may be included at the end of an introductory probability course that precedes the statistics course. For a given bound or interval, the confidence level is (I - a) is.) where (il, . Generalize Lemma 5.3.3 by showing thai if Y 1, . One of the most important uses of power is in the selection of sample sizes to achieve reasonable chances of detecting interesting alternatives. Moreover, we shall parenthetically discuss features of the sources of data that can make apparently suitable models grossly misleading. We introduce the decision theoretic foundation of statistics including the no ' tions of action space, decision rule, loss function, and risk through various examples in cluding estimation, testing, confidence bounds, ranking, and prediction. Thus, for any real-valued function r(Y) with Elr(Y) I $< \infty$, exp - 2 [xi + !x2]. In the ith pair one patient is picked at random (i . It may be shown (Problem 6.4. 3) that if A and B are independent, that is, P(A I = 2) = Eyr(y)p(y I z) + xz. P(A I B), then Z is approximately distributed as N(O, 1). 1.1. 218 Testing and Confidence Regions Chapter 4 1.0 0.8 0.6 0.4 0.2 0 3 :---+-++0 0.1 0.2 0.3 0.4 0.5 - 0.6 0.7 0.8 0.9 1.0 Figure 4.1.1. Power function of the level 0.05 one-sided test c5k of H : () = 0.3 versus K : B > 0.3 for the B(IO, 0) family of distributions. • 6. Statist. D In all of the situations we have discussed it is clear that the analysis does not stop by specifying an estimate or a test or a ranking or a prediction function. The optimality part of Theorem 5.4.3 is not valid without some conditions on the estiin fact it is important to get close to the truth-knowing that at most oo dollars are owed is of no use. 7 I Similarly, $E(Y \mid Z = 0) = \hat{\phi} = 0.43$. That is, we need a priori estimates the function valued parameter of how well even the best procedure can do. Show that if 9 is a MLE of 9, then q(O) is an MLE of w = q(O). It estimates the function valued parameter of how well even the best procedure can do. Show that if 9 is a MLE of 9, then q(O) is an MLE of w = q(O). It estimates the function valued parameter of how well even the best procedure can do. F defined by its evaluation at x E R, F(P)(x) = P[X1 5 x]., is sufficient for B. Note that Y is independent of Y1, . Let X1, . (X)) and ,its (1 - a) th quantile obtained from the table. If the customers have provided accurate records of the number of defective items that they have found, we can construct a frequency distribution {1ro, . LINDLEY, LINDLE D.V., "Decision Analysis and Bioequivalence Trials," Statistical Science, I3, 1 36-- 1 41 (1998). For instance, in the prediction exam ple 1.3.2, 11(·) is the parameter of interest. = X + Z, where Z is (a) What is the conditional distribution of Y given X = x? From (c) and (d) and Theorem (A. = -1 t=1 n !:. i: '. I Section 8.10 Topics in Matrix Theory and Elementary Hilbert Space (iv) II is norm reducing 523 Theory Ill(h I L) ll < ll h ll· (8.10.15) ln fact, and this follows from (8.10.1 , independent. · I ' The correlation of X1 and X2. We are interested in expected differences in responses due to the treatment effect. It follows that [X(]] • X(kJ] with k = n + 1 - j is a level a = (n + 1 - 2j) / (n + 1) prediction interval for Xn + 1 · This interval is a distribution-free prediction interval. p Pr(• > x) where x is in the body of the table and p is in the top row (margin). What our hypothesis means is that the chance that an individual randomly selected from the ill population will recover is the same with the new and old drug. 220 Testing and Confidence Regions Chapter 4 Example 4.1.5. Goodness of Fit Tests. We test the hypothesis H : 8 E 80 versus K : 0 9 80, where 80 is a two-dimensional subset of 8 given by "' 80 = { ("7 1 'T/2 , "1 1 (1 - "72), (1 - 'T/1) (1 - "72) ; 0 ::; "7 1 ::; 1, 0 ::; "7 1 ::; 1, 0 ::; "7 1 ::; 1 } . (5) We can be guided to alternative or more general descriptions that might tit better., Xn be i.i.d. gamma, r(.A,p). Let Q : II E e), e open c R=, m < k - 1, be a cUIVed exponential family = {Po $p(x,ll) = exp{cT(II)T(x) - A(c(ll))}h(x)$. For instance, we may be inter ested in the center of a population, and both the mean Jl and median v qualify. WALLACE, Point Estimation Using the KullbackLeibler Loss Function and MML, in Proceedings of the Second Pacific Asian Conference on Knowledge Discovery and Data Mintng Melbourne: Springer-Verlag, 1998. Let U (X1, . - a?fo.Y. asymptotic Show that if N(J", cr2), ep(X, X) (g) Suppose X1 has the gross error density j,(x - B) (see Section 3.5) where f,(x) = (I c)) density. To determine significant values of these statistics a (more complicated) version of the follow 'I' • I'! ling is done. APoSTOL, T. In general if J.L is a a finite measure on the sample space X, p(x, (}) as given by (1.6.1) ' '96 Statistical Models, Goals, and Performance Criteria can be taken to be the density of X with respect to J-L-see Lehmann (1997), for instance. Next this model, which now depends on the data, is used to decide what estimate of the measure of difference should be employed (cf., for example, Mandel, 1964). We study the important application of the unbiasedness principle in survey sampling. and Neider (1983, 1989). UI /v'n. Thus, if we take Zi as having marginal density qo, which we temporarily assume known, then (Z1, YI), . Consistency, and Asymptotic NormalityEfficiency in Semiparametric Models Tests and Empirical Process Theory Asymptotic Properties of Likelihoods. (c) Use the central limit theorem to find a normal approximation to the critical value of test in part (b). Let X = ((X 1, Y1), . An urn contains four red and four black balls. The distribution corresponding to br, • will be written (J(r, s) . The other family, which is in dexed by the positive parameters r and s. In Example 1 . A., "On the Attainment of the Cramer-Rao Lower Bound," Ann. where .6. Bj 3 2 I 0 0 1 2 Figure 2.4.1. The coordinate ascent
algorithm. • Includes bibliographical references and index. 1.2 again in which we as sume the error f to be Gaussian but with arbitrary mean $\hat{\Psi}$. Show that {S,; c > 0} is a family of ellipses centered at (l't. Note the similarity to the interval in Problem 4.4. 13(g) preceding. Define X(m) = (1/m) L::, $\hat{\Psi}$ 1 X;, and Sm - . then Ba(T) is a lower confidence bound for () with confidence coefficient 1 - a. We can now state the following elementary consequence of Theorem B.3.1. ' I 492 Additional Topics in Probability and Analysis Appendix B 0, choose x, x such that $F(x) : S \pounds, 1 - F(x) < \pounds$. From (i), (ii), and (iv) we see that, if we define 0, then k(O(S), a) = S and, therefore, we find 8(S) as the unique solution of the equation, When S = 0, 8(S) = 0. The mean of X is 0, for k > 1, and Var X = kf(k - 2) for k > 2. kn(e0, a) of k > 1, and Var X = kf(k - 2) for k > 2. kn(e0, a) of k > 1, kn(e0, a) of k > 1, kn(e0, a) of k > 1. . Let F11 denote the distribution of Tn = Jn(X - J-L) fa and let fin and)'211 denote the coefficient of skewness and kurtosis of Tn . Show that the MLE of a, fJ exists iff (Y1, . If departments .. use different coins," then the data are naturally decomposed into = (Nm 1d, Nmod, Nfld, Njod, d = 1, , D), where Nmld is the number of male admits to department d, and so on. Rejecting for large values of this statistic is equivalent to rejecting for large values of X. (0, T2(0) [A'(O)]') I 1 i • • • ' i' I I, I for every sequence {On} with On = 0 + t(...jii for t E R. The trouble is that for any specified degree of approximation, say, f = .01, (5.1.6) does not tell us how large n has to be for the chance of and Limit Theorems 466 A.I 5 Further Limit Theorems and Inequalities 468 A.t6 Poisson Process 47 2 A.1 7 Notes 474 A.1 8 References 475 ADDITIONAL TOPICS IN PROBABILITY AND ANALYSIS 477 B.1.2 Conditioning by a Random Variable or Vector B.1.1 The Discrete Case 477 8.1.2 Conditional Expectation for Discrete Variables 479 B.1.3 Properties of Conditional Expected Values 480 8.1.4 Continuous Variables 482 B.1.5 Comments on the General Case 484 B.2 Distribution Theory for Transformations of Random Vectors B.3 485 8.2.1 The Basic Framework 485 8.2.2 The Gamma and Beta Distribution Theory for Transformations of Random Vectors B.3 485 8.2.1 The Basic Framework 485 8.2.2 The Gamma and Beta Distribution Theory for Transformations of Random Vectors B.3 485 8.2.1 The Basic Framework 485 8.2.2 The Gamma and Beta Distribution Theory for Transformations of Random Vectors B.3 485 8.2.1 The Basic Framework 485 8.2.2 The Gamma and Beta Distribution Theory for Transformations of Random Vectors B.3 485 8.2.1 The Basic Framework 485 8.2.2 The Gamma and Beta Distributions 49 1 8.3.2 Orthogonal Transformations 502 8.5.2 Properties of Variance 503 B.6 The Multivariate Normal Distribution 506 8.6.1 Definition and Density 506 8.6.2 508 Basic Properties. • Example 1.5.2. Suppose that arrival of customers at a service counter follows a Poisson xl be the time of arrival for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. That is, (X(k)' X(n-1)) is a level (1 - a) confidence interval for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever because of Section 3.2 for de ciding between Oo and 81. The for Xp whatever between Oo and 81. The fo F satisfying our conditions. Inference in Semiparametric Models. Thresholds (critical values) are set so that if the matches at one positions) and the probability of a match is than a. = (5.4.38) , where ... This leads to the model with x () Nn . It may be obtained from the Cauchy-Schwartz inequality, 2 r (A.1 1. J., The Foundations of Statistics New York: J. The same type of question arises in all examples., Vn)T 812/S1 W2 u2 U2, . 12), ll h ll' = lll(h I L) may be interpreted as a projection on L = {h : (h, h') = 0 for all h' E [}. Intervals, and Regions 233 ix CONTENTS The Duality Between Confidence Regions and Tests 241 *4.6 Uniformly Most Accurate Confidence Bounds 248 *4.7 Frequentist and Bayesian Formulations 255 4.9.1 Introduction 255 4.9.2 Tests for the Mean of a Normal Distribution-Matched Pair Experiments 257 Tests and Confidence Intervals for the Difference in Means of Two Normal Populations 261 4.9.4 The Two-Sample Problem with Unequal Variances 264 4.9.5 Likelihood Ratio Procedures for Bivariate Normal Distributions 266 4.5 4.9.3 4.10 Problems and Complements 269 4.11 Notes 295 5.1 Introduction: The Meaning and Uses of Asymptotics 297 5.2 Consistency 301 5.3 5.2.1 Plug-In Estimates and MLEs in Exponential Family Models 301 5.2.2 Consistency of Minimum Contrast Estimates 304 First- and Higher-Order Asymptotics: The Delta Method with Applications 306 5.3.1 The Delta Method for Moments 306 5.3.2 The Delta Method for In Law Approximations 311 5.3.3 Asymptotic Normality of the Maximum Likelihood Estimate in Exponential Families 322 Asymptotic Theory in One Dimension 5.4 324 5.4.1 Estimation: The Multinomial Case 324 5.4.2 Asymptotic Normality of Minimum Contrast and M-Estimates 327 *5.4.3 Asymptotic Normality and Efficiency of the MLE 331 * 5.4.4 Testing 332 *5.4.5 Confidence Bounds 336 • 5.5 Asymptotic Behavior and Optimality of the Posterior Distribution 337 5.6 Problems and Complements 345 5.7 Notes 362 5.8 References 363 ! ' X 6 CONTENTS INFERENCE IN THE MULTIPARAMETER CASE 6.1 *6.2 *6.3 Inference for Gaussian Linear Models 365 6.1.1 The Classical Gaussian Linear Model 366 6.1.2 Estimation 369 6.1.3 Tests and Confidence Intervals 374 Asymptotic Estimation Theory in p Dimensions 383 6.2.1 Estimating Equations 384 6.2.2 Asymptotic Normality and Efficiency of the MLE 386 6.2.3 The Posterior Distribution in the Multiparameter Case 3 9! Large Sample Tests and Confidence Regions 6.3.1 6.3.2 * 6 .4 365 392 Asymptotic Approximation to the Distribution of the Likelihood Ratio Statistic 3 92 Wald's and Rao's Large Sample Tests 3 98 Large Sample Tests 3 98 Large Sample Methods for Discrete Data 400 6.4.1 Goodness-of-Fit in a Multinomial Model. To see whether the new treatment is beneficial, we test H Fo has a density fo(y). Thus, Po,[Bn > Bo + Zt-a/Vnf(Oo)] = Po,[Vn1(8o)(Bn - Bo) > Zt-a] 🗞 a. e0 is simple as in Example such as testing J.L = J.Lo versus This occurs if But it occurs also in more interesting situations J.L = j:. g, (A.15.4) If we put 🍫 t) 🤄 if t 🕏 cases are obtained by taking and = and 0 and 0 otherwise, we get (A.15.2). (1 - a:) by the posterior distribution of the parameter In the case of a normal prior 1r (B) and normal model (} p (x I B), the level (1 given - a:) credible interval is similar to the frequentist interval except it is pulled in the direction p,0 of the prior mean and it is a little narrower. Thus, 1 /.>. The claim (A.16.3) now follows from Slutsky's theorem (A. Let X a + n (a:) denote the a:th X a + n (a:) I (t + b) is a level (1 - a:) lower > 0 are known parameters. Some classical conditions may be found in Apostol (1 974), p. Assoc., (2000). (b) Find E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y, given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y, given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor
of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of finding the best MSPE predictor of Y. given a vector Z; that is, we can find the g that minimizes E(Y - p.) = Now we can solve the problem of find the g that minimizes E(YS = n, O(S) = I. However, the treatment is abridged with few proofs and no examples or problems. Assume the model where X = (X1, . 1, the conditional distribution of X o given T = I: (1, . 1 covariate. To find). The Heuristics of Test Construction When hypotheses are expressed in terms of an estimable parameter H - : () E eo c RP, and we have available a good estimate () of (), it is clear that a reasonable test statis- - tic is d(O, eo), $inf\{d(x,y) where d is the Euclidean (or some equivalent) distance and <math>d(x, : y \in 8\}$. We give the proof in the discrete case. Show that sup{/ F;; 1 (a) - F-1(a)/: < < a < I - E} 0 for all '> 0. 15.1). (See Problem 3.5.6.) 8. n i=l = 0, Section 4. M. s Therefore, the likelihood ratio tests rej ect H if, and only if, I Tn 1 2': 2. (a) Lehmann Alte arive. Then the wrong decision "11 < + to' is made when T < -z(l -) a). X,,) [= E[X] [i !t, ···, ij by Problem 5.3.5, so the number d of nonzero terms in (a) is U/ 21 n 2: (c) r_- 1 r. Doksum University of California Pn_, nt icc Hall PRENTICE HALL Upper Saddle River, New Jersey 07458 Library of Congress Cataloging-in-Publication Data Bickel. D. give and plot the sensitivity curves of the lower X. , k }] with canonical parameter a and t: = Rk . E;1r;p(x I O;) (3.2.10) The optimal action 6* (x) has r(J'(x) I x) & min r(aj I x). Appendix B is as self-contained as possible with proofs of most statements, problems, and references to the literature for proofs of the deepest results such as the spectral theorem. ' '. Two 0.12 sample; 10000 Simulations, Chi-R where formula. The De Moivre-Laplace theorem is generalized by the following. Methuen & Co., 1962.) [Y, Y] For instance, a doctor administering a treatment with delayed effect. In order for p to be a contrast function we require that D(80 1 9) is uniquely minimized for $8 = 60^{\circ}$ That is, if P90 were true and we knew D(Bo, 8) as a function of 8, we could obtain 80 as the minimizer. In Section 4.4 we considered the level { 1 - a) confidence interval X ± uz{1 - @ a)/ fo for f-t., dg be the decision rules of Table 1.3.3. Compute and plot the risk points (a) p @ 1 - q @ . Example 5.3.5. Edgeworth Approximations to the x2 Distribution. Let mk from stratum k form the corresponding sample averages X1, XK. 1 .2 with assumptions (1)--{4), the parameter of interest can be character 1 1 ized as the median v = F- (0.5) or mean I' = f=oo xdF(x) = J0 F-1 (u)du. If we keep temporarily to this notion of event as a recurrent phenomenon that is randomly detennined in some fashion and define N(t) as the number of events occurring between time 0 and time t, we can ask under what circumstances { N (t)} will form a Poisson process. (a) Find maximum likelihood estimates of the fJi under the assumption that these quan tities vary freely. HANSEN, M. The very simple proofs of these results may, for instance, be found in Gnedenko (1 967, p. Let X 1, . Xn be the times to failure of n pieces of equipment where we assume that the Xi are independent of X1, . 10.6) Proof. (a) If A, (i) > Q2 B, C are three events, consider the assertions, P(A n B 1 C) = P(A 1 C)P(B (ii) P(A n B 1 1 C) (A, B INDEPENDENT GIVEN C) P(A n B) = P(A)P(B) (A n B) = P(A)P(B) (A n B) (A n B) = P(A)P(B) (A n B) (A n B) = P(A)P(B) (A n consider its probability distribution. (a) Show that then . Example 4.2.2. Simple Hypothesis Against Simple Alternative for the Multivariate Nor mal: Fisher's Discriminant Flmction . (b) random variables and let V = (X1 + ej2 + L 2 Xi2 • Show that for fixed v and n, P(V > v) is a strictly increasing function of 82. The great advantage of our new approach is that it enables us to compute the Bayes risks of all corpeting procedures. X is distributed according to {Po : 0 E 8 c R} and ,- is a prior distribution 2 for (J such that $E(9) < \infty$. (b) Show that if c > 0 and a E(0, 1) satisfy Pe, [2N1 + N2 > c] = a, then the test that rejects H if, and only if, $2N_r + N2 > c$ is MP for testing H : 8 = 80 versus K : 8 = 81. However, we have little control over what kind of distribution of errors we get and will need to investigate the properties of methods derived from specific error distribution assumptions when these assumptions are violated. We claim that 1 1 Section 4.5 245 The Duality Between Confidence Regions and Tests Figure 4.5.1. The shaded region is the compatibility set C for the two-sided test of Hp.0 : J.L = J.to in the normal model., n. These solutions are the maximum likelihood estimates. A4: sup1 AS: '• { V I: V 1 (V (X; , t) - V (X; , O(P)) I: jt - 9(P)I < , Po) > V a (0) V for some a o1 0. 1.2 and 4.1.3. In Example 4.1.2, power of the resulting tests. , n, l = 1, . L., Testing Statistical Hypotheses New York: Springer, 1986. , d} = supremum distance. (iii) F is chosen by first choosing I' = 6(F) from a N(O, k) distribution and then taking F = N(6(F), M). This condition ensures that g(X) satisfies For convenience, when we refer to functions we shall assume automatically that this condition is satisfied. (2.1.2) The equations (2.1.2) define a special form of estimating equations. Use (5.3.6) to explain why. - - (b) Either () = x or() is not unique. Denote the probabilities of these types by B11, B12, B21, B22, respectively. 1, P(X; - On) < 0). Justify the following approximation to the posterior distribution = where q. (d) Consider the problem of testing H : Ji, = Ji, o versus K : Ji, > J.to on the basis of the N(p., "2) sample X1, That is, £,(-fii(Bn - 8)) N(o, r 1 (8)) (5.4.40) where 1(8) > Ofor all 8. To establish this note that (a) sup { (0 + fo) 11"(8) 11" tent and 1r is continuous. In the one-way layout, (a) Show that level (1 - a) confidence intervals for linear functions of the form {3j - {3;. Such a v is called a (1 - a) upper confidence bound on v. This edition gives careful proofs of major results and explains how the theory sheds light on the properties of practical methods. 1 These conclusions remain valid for the usual situation in which the Z i are not random but their proof depends on asymptotic theory for independent nonidentically distributed variables,
which we postpone to Volume II. It is customary to write the model as, for c(7) > 0, p (y, 17, 7) Because J p (y, 17, 7) dy = = exp{ c-1 (7) (11Ty - A (17)) } (11Ty - A (17) Peter J. J.Lo if we have N(p., 0"2) observations with both parameters unknown (the t tests of Example 4.5.1 and T2 do not depend one, and 81 respectively, then (T1 (X), T2(X)) is sufficient for e. Finaiiy, define Fp,, a as the d.f. of aY + p. I: 1 Fo(1:i). M Many examples and important issues and methods are discussed, for instance, in Chapter 6 of Dahlquist, BjOrk, and Anderson (1974). Therefore, the conditional expectation is an ordinary expectation with respect to the probability measure P. Y; (iii) Let Yi , Y2 , denote the time required until the first, second, . In Problem 13 preceding we gave a Xp for p fixed. 1 or the physical constant J-L in Example 1.1.2. These are estimation problems., Xn is rv = X = n - I : L = I Xi, and it is enough to derive the marginal distribution of X, Xn + I and 9, where X and Xn + I are independent. COCHRAN, Statistical Methods, 8th Ed. Ames, IA: Iowa State University D. An equivalent sufficient statistic in this situation that is frequently used IS n • n S(X1, . (b) The observations are X1 = the number of failures before the first and second successes, and so on, in a sequence of binomial trials with probability of success (J. , 'll''N } for the proportion (J of defectives in past shipments. I '' 282 Testing and Confidence Regions Chapter 4 (c) Modify the test of part (a) to obtain a procedure that is level a for H : 81 = B, 02 = eg and exhibit the corresponding family of confidence circles for ({)1, 82}. g(Z) for some g (measuralle). We can regard twins as being matched pairs . I 3 . It was noted by Fisher as reported in Jeffreys (1961) that in this experiment the observed fraction ';: was much closer to j than might be expected under the hypothesis that NAA has a binomial, B (n,), distribution, NAA m 1 I < = 7 x 10 5. Section 1.7 73 Problems and Complements where m = max(x, . !x g(xn) such limits. (a) Show that Y = X2 has density py(y) = 1 2.fii'Y ' e-l(>+9) (e0v'Y + e-0v'Y), y > 0. the numbers of admitted male and female applicants, and the corresponding numbers Nmo, NJo of denied applicants. KRETcH AND R. This provides another proof that Z and L + 2 are independent. 433 Problems and Complements The following table gives the number of applicants to the graduate program of a small department of the University of California, classified by sex and admission status. If this is not the case, they are estimating population means and sample covariances estimating population means and sample covariances. (b) Show that the family of distributions obtained by letting JL, v vary freely is an exponential family of rank {C - 1) + C(A + B - 2) = (a) Show that if J1 and a2 are unknown, J1 E R, a2 > 0, then the unique MLEs are /i = X and a2 = n-1 2.: 1 (X, - X) 2 (b) Suppose p and a 2 are both known to be nonnegative but otherwise unspecified. Now consider the indifference region (Oo, Bt), where = + Dt., .0t. we have made the correct decision and the loss is zero., >-; ') PT • ' I B.10.1.4 If A is spd, then max { xT Ax : xTx < I } * • max; >-; . GRENANDER, U. • Xn) = X = ! L 1 Xi, a common Formally, a statistic T is a map from the sample mean and sample variance. We conclude by studying consis tency of the MLE and more generally MC estimates in the case 8 finite and 8 Euclidean. This distribution does not always corresp:md to an experiment that is physically realizable but rather is thought of as a measure of the beliefs of the experiment that is strictly increasing on the range of , such that h(>.(X)) has h(>.(X)) is equivalent to >.(X), we specify the size a likelihood ratio test through the test statistic h(>. 0, :I 'I 14. Now consider situation (d). Example 4.1.3. Suppose we have discovered a new drug that we believe will increase the rate of recovery from some disease over the recovery rate when an old established drug is applied. (a) The likelihood is symmetric about x. 1) depend on an additional scalar parameter 7. From the beginning we stress function-valued statistics, such as the empirical distribution function., Yn · · · oo. _ for d = 1, , D. E R, and g(t,li) • y = g(t; 8), where 8 a {1 + exp[f(t - p)/15] 'yi), (b) Suppose we have observations (t 1. Suppose that IJ = (81, . Bk) is unknown and may range over the set 8 = { (B, . Instead we turn to the logistic transform g(1r), usually called the logit, which we introduced in Example 1.6.8 as the canonical parameter = = = 'xk .t:" (6.4. 10) TJ = g (1r) = log [1r] (1 - 1r)]. This suggests that x2 may be written as the square of a single (approximately) standard normal variable. ! i 6. In each case, find a real-valued sufficient statistic for (), a fixed. 2 ϕ j (:e logp(x, B)) p(x, B)dx. :. Here are some examples. In this volume we assume that the model has i I i I Section 1.1 Data, Models, Parameters, and 9 Statistics been selected prior to the current experiment. Hint: Apply the argument of the proof of Theorem 2.4.2 noting that the sequence of iterates {fjmJ is bounded and, hence, the sequence of iterates a = 0.01 and 0.05 are commonly used in practice. Scholz. It is no restriction in the 0 is Rk or a subset of Rk . Show that IZn - Zl + 0 is Rk or a subset of Rk . equivalent to Zn; Vz; for I:S j < d. Keener, University of Michigan From the Publisher File loading please wait... E(Y I z = 20) = 0. i=l i=l (1.5. 1 1) ' I l ' Section 1.5 45 Sufficiency Evidently p{x1, . Consider the problem of testing H : F = F0 versus K : F i- Fa. Let F denote the empirical distribution and consider the sup distance between the hypothesis F0 and the plug-in estimate of F, the empirical distribution function F. (i) If a strictly convex function has a minimum, it is unique. Note that the posterior density if B has prior density 1r and 1 Xi, which has a B(n, 8) distribution given B (} (Problem 1.2.9). A sufficient statistic based on the observables X I, . - 7. LEHMANN, "Unbiased Estimation in Convex. It does tum out that _, more accurate than e2(X) and is, in fact, uniformly most acc 16.04 15.44 34. Consider the set of v0 for which H.,.0 is accepted; this is a random set containing the true value of v(P) whatever be P. • (3(8) = P(U < a) = 1 - Fe(F0 1 (1 - a)) 1 1 where F0- (u) = inf{t : Fo (t) 2: u)., Yn). Suppose c : 8 • E c R has a differential c(II) = (2.3.6) * (II) mxk on e. * 1" • ' ' ' 348 Asymptotic Approximations 2: min l < M;n1 E (,; L:iX; - X;!l) l'lj. This can be viewed as an extension of Example 3.3.3. Let 1fk be a prior distribution on :F constructed as follows:(l) (i) (ii) 1rk {F : VarF (X!) fo M} = 0. However, when we put the multinomial in canonical exponential family form, our parameter set is open. Leighton, and M. Bo) is given by I ifO"(x) > Oo o"(x, Oo) 0 otherwise is UMP level a for H : (} level (I - a). - - --- - - Section 1.7 Problems and 71 Complements Hint: See Problems 1.1.11 and 1 . 0 Note that the model for X is unchanged. They are only used to show that with randomization, likelihood ratio tests are unbeatable no matter what the size a is. To improve on this approximation, we need only compute !In and 12n. Show that under the assumptions of Theorem complete. 1 • Section B.1 479 Conditioning by a Random Variable or Vector Two important formulae follow from (B. Given a test statistic T(X) we need to determine critical values and eventually the Examples 4. The first family has densities given by (8.2.8) for x > 0, where the parameters p and A are taken to be positive and f(p) denotes the Euler gamma function defined by r(p) = 1' = edt. The topic presently in Chapter 8, density estimation, will be studied in the context of nonparametric function estimation. A special case of this is the famous Deming-Stephan proportional fitting of contingency tables algorithm-see Bishop, Feinberg, and Holland (1975), for instance, and Problems 2.4.9-2.4.10. example of a neural net model is Yi = p L h(zii; ...) + Ei 1 j = I i = 11 • • • 1 n where A = (a, (3, p), h(z; >.) = g(z; a, [3, Jl., I); and E l, ... Now it is reasonable to • • suppose that the value of (J in the
present shipment is the realization of a random variable . , Xn+l are i.i.d. as X where X has the exponential distribution F(x I 0) V I - e-x/O, ... If the ,\i are unrestricted, 0 < ,\i < 1, 1 < i < n, this, from Example 1 .6.2, is an n-parameter canonical exponential family with Yi = integers from 0 to ni generated by T (Y, , . . , Xn) P, (X1, . . Now consider dn = J:" (e + Jn) exp 0 (x, J + Jn) - l(X,, e) ds r dnqn(s)ds Jlsi 8)dt i=l (5.5.15) - By AS and A7, n o) e-m(,, < exp 0 (I(X, t) l(X B-)) ·. By Problem 4.6.8, the UMP test accepts H if • n L Xi < X2n (1 - a) /2 >- o (4.6.3) i=l or equivalently if A0 X2 n (1 - a) /2 frequentist interval; however, the interpretations of the intervals are D different. (3) Arrays of scalars and/or characters n individuals-see Example as in contingency tables-see Chapter 6---or more generally multifactor multiresponse data on a number of individuals. 22) is a consistent estimate of \$-1\$ (110)Hint: Argue as in Problem 5.3.10. - (a) Show that testing H: 8 < 0 versus K: () > 0 is equivalent to testing H': P[X1 > OJ (I - p). The correlation of XI and X2. Let fo (x) \clubsuit fo (x - 9) where :1:::: $l(S(x) \And s)$. Let X1, •••, c • Xn be a sample from a population with the Rayleigh density • $f(x, O) = (xj0 2) \exp\{-x2j202\}$, x > 0 0 > 0. Furthermore, () · (B.! 0.8) Proof From Section B.6 we have noted that there exist (Gaussian) random vectors Up xi, Vqxi such that E = Var(Ur, vr)r, En = Var(U), E12 = cov(U, V). "BICKEL, P., AND E. is the distributions, respectively. Suppose that this is true for each l. Therefore, Po, [S > j] (k) and x (k), l. r for specified Boj -" For instance, to test the Hardy-Weinberg model we set e = B1 , e = B2 - 2VBt (1 - VBt) and test H : e = 0. New York: Springer, Lecture Notes in Statistics, 2000. TUKEY, J. (A. The following result can be useful. Hint: If it were, there would be a set A such that p(x, 0) > 0 on A for all 0. , {3p . (c) What is the approximate distribution of y'ii(X -- Jl.) + X2, where J1. It may also happen that the distribution of Tn as (} ranges over 81 is detennined by a one-dimensional parameter >.(0) so that 90 = {0 : >.(0) > 0} and Co(Tn) = c,(6)(Tn) for all 0. Z able) and a random variable is the information that we have and Y the quantity to be predicted. 3, = $\Phi = p(y) - Z(Y)$ IV v $\Psi Y - Y = vn - 1 + 1s$ has the t distribution, Yn - 1 (1 - 4) prediction interval Φa) ::; Tp (Y) ::; tn - 1 (1 - 4) for Y = X ± Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) for Y = X \pm Jn - 1 + 1stn - 1 (1 - 4) fo beyond the particular experiment. Statistical Models, Goals, and Performance Criteria Let X1, . (5.4.54) Assertion (5.4.54) establishes that the test O.in yields equality in (5.4.50) and, hence, is asymptotically most powerful as well. Whenever D we can obtain such steps in algorithms, they result in substantial savings of time. D 82D a2o a2o ("D) EP $(1 \cdot 1)$ If aa2 > 0, ab'l > 0 and 802 ab2 > BaOb, then D IS strictly convex. DOWE, D. Then c(8) is closed in £ and we can conclude that for m > 2, an MLE (] of 8 exists and (} D satisfies (2.3.7). Show that if (X1, . (b) Conversely show that if T(X) is sufficient, then, for any prior distribution, the posterior distribution depends on x only through T (x) For instance, in Example 1.1.1, the fraction of defectives () can be thought of as the mean of Xjn. See Lehmann (1997) and Volume II for discussions of this property. In Example 4.3.4, show that the power of the UMP test can be written as 2 (Gn(n)/0 is the parameter of meterest. df"i (Yi - f"i)Zij (O, J. AND W. Consider the general framework where the random vector X takes values in the sample space I I I I I r E P. { $(X I s, B) \log I S(X) = s, (X I s, B) \log I S(X) =$ 5 and the "ideal" ordered sample of size n - 1 = 4 is - 1.03, -.30, .30, 1.03 (these are expected values of four N(O, !)-order statistic T. Prove the analogue of Theorem 1.6.1 for discrete k-parameter exponential families. See Figure B.4.2. (X1, YI), . • I ' Section 4.3 UnifOfmly Most Powerful Tests and Monotone Likelihood Ratio Models 227 Summary. FEYNMAN, S., Mathematical Statistics New York: Academic Press, 1967. Fisher's Exact Test From the result of Problem 6.2.4 deduce that if j(o:)
(depending on r1, c1, n) can be chosen so that then the test that rejects (conditionally on Rr = Tf, cl = Ct) if Nu > j(a) is exact level a. All the severely ill patients might, for instance, have been assigned to B. Soc. Show that EPV(8) = P(To > T). t, J.L2, cr, p) denote the d.f. of (X, Y). We shall discuss this approach somewhat, by example, in Chapters 4 and Volume II but refer to Lehmann (1986) and Lehmann and Casella (1998) for further reading. I: (x i - x) 2 na 2 j (n - 1), a5 fa 2 = Tn = y'n(x - fJ o). •, •, Section 4.5 243 The Duality Between Conf1dence Regions and Tests H = Hv0 : v = v0 test J(X, v0) with level a. then the test that accepts H; v va. Hint: Consider the class Bayes tests of H : () 1r{Oo} 1 - 1r{O}, varies between 0 and l. (2.4.2) The rationale here is simple. ! ' ' I i i • Section L2 15 Bayesian Models tOr 0 < (} < 1, Xi = 0 or 1, i = 1, . Example 2.3.4. Gaussian with Fixed Signal to Noise. The generalized linear model The GLMs considered so far force the variance of the response to be a function of its mean. These authors also discuss most of the topics we deal with but in many instances do not include detailed discussion of topics we consider essential such as existence and computation of procedures and large sample behavior. l. London 71, 303 (1 905). Statist., 3, 1045-1069 (1975b). This result is extended to rules that are limits of Bayes rules with constant risk and we use it to show that x is a minimax rule for squared error loss in theN(0, a5) model. (d) parison of the asymptotic length of (4.9.3) the pivot jI i(so where D and so are as in Section 4.9.4, 29, T Rm, Then if treatment B had been administered to the same subject instead of treatment A, response x + 6. Closed form versus iteratively computed estimates At one level closed form is clearly preferable. (hr.) = ({1 > •. On the other hand, if 0 > 00, Po [S 2: j] > a. ', I Problems for Section B.5 ' I. Section 6 . , U F(X) 🛷 U(O, 1). Show that p(y, fi) as given by nential family model with l = 2. Let (X,, Y,) , 1 < i .(Xt, Yi mixture of point mass at 0, xi and x Hint: By sufficiency reduce to n = 1. However, we illustrate what can happen with a simple example. Let XJ , . Definition 4.6.1. A level (1 - a) LCB 8* of() is said to be more accurate than a competing level (1 - a) LCB 8 if, and only if, for any fixed 8 and all 8' < B. (b) Show that the sample 5.3.6 is related to Z of (6.4.8) by Z $\hat{\bullet}$. fiir .) A = are the multinomial coefficients. • Theorem 5.4.4. Suppose the model P = {Po : 8 E 8} is such that the conditions of Theorem 5.4.2 apply to '0 = $g^{\textcircled{a}}$ and Bn. the MLE. Because F is uniformly continuous on [x, x], there exists c5 (£) > 0 such that [x, -Xz] < J(c) '* IF(x,) - F(x2)1 < f. -1 Moreover, s2 (n - 1) l:: a (Xi - X) is independent of Xn+1 by assumption. Under general conditions (Bhattacharya and Rao, 1 976, p., ik) with ij > 0. 5. Let (X, Y) have a N(!'l> p,2, u , u, p) distribution in the probability 1 Y; with probability 1. = 0 < .6. I., The Foundations of Statistics ,. i I I NDEX X '''-J F, X is distributed according to F, table, 379 463 B(n, 8), binomial distribution with parameters D, N, n, 461 H(D, N, n), hypergeometric distribution with parameters D, N, n, 461 M(n, 81, . - >.). Assumption II is practically useless as written. Consider the hypothesis H : Oij = Tfil Tfj2 for all i, j. In this case, we settle for a probability at least (1 - a). distribution is the marginal distribution of Xn+I. The posterior predictive distribution of Xn+I distribution of Xn+I. The posterior predictive distribution of Xn+I distribution of Xn+I. The posterior predictive distribution of Xn+I distribution of Xn+I distribution of Xn+I. The posterior predictive distribution of Xn+I distribution of Xn+I distribution of Xn+I. The posterior predictive distribution of Xn+I distribution of Xn+I distribution of Xn+I distribution of Xn+I distribution distr B are symmetric and A :S B, then for any C (B.10.5) B - A sod means B-A * EET and thenCBcT -CAcT * C(B-A)cT * a predictor Y* to be prediction unbiased for Y if E(Y* - Y) 0, and can c n c n c n c predictors, the optimal MSPE predictor is Y X. The Nij can be arranged in a a x b contingency table, Nu a Na1 c1 1 2 N12 c2 ... Thus, if we write cTto = I; {c;t;o : c; < 0} + I: { t; 0 + 1 in the first sum or a tjo by t; 0 - 1 in the second. Let 11 = E(Xi) be the average time for an infinite series of records. Ltd. that z 🗞 z. 'Xn be a sample from a u [8 - i 8 + i I distribution. B.ll, • • 'P ROBLEMS AND COMPLEMENTS Problems for Section B. · The special case p = 1 corresponds to the familiar exponential distribution £(...) of (A.13.24). Just how widely applicable the notions of this section are will become apparent in Remark 1.4.5 and Section 3.2 where the problem of Bayesian statistics with squared error loss. Appendix B 11. (See Problem 4., k < oo-see Problem 6.4. 1 3. h : R • R, let]]g]]oo = sup {]g(t) I : t E R} denote the sup norm, and assume, Xn (i) ! (b) ' ! (a) h is m times differentiable on R, m > 2. There is a deep connection between symmetries of the model and the structure of such procedures developed by Hunt and Casella 5 (1998), for instance. If, on the other hand, we only wish to make assumptions (1]-(3) with t: having expectation 0, we can take e to be a nice subset of Euclidean space and the maps () ----+ Po are smooth, in senses to be made precise later, models Pare called parametric. Example 3.4.2. (Continued). Estimation. We explore the connection between tests of statistical hypotheses and confi dence regions. For example, in the Z would be the College Board score of an entering freshman or her first-year grade point average. • 8.10.1.2 Spectral Theorem (a) Avxp is symmetric iff there exists P orthogonal and $D = diag(At, . In the binomial one-way layout show that the LR test is asymptotically equivalent to Pearson's x2 test in the sense that 2 log. \- x2 .f.; 0 under H. This follows from (B.! 0.9) since a Var(U - E12E221 V)a = 0 for all a iff (B.JO. 29) for the noncentrality parameter 82 in the one-way layout. X. Volume 2 for B2, these$ are both UMVU. (2) The construction of models for time series, temporal spatial series, and other com plex data structures using sophisticated probability modeling but again relying for analytical results on asymptotic approximation. (exponential density) (b) f(x, 8) = 8c9x-(O+I), X > c; c constant > 0; 8 > 0. Before we give the example, here is the general definition of UMP: Definition 4.3.1. A level a test f3(0, E, we conclude that the MP test rejects H, if and only if, = Critical values for level a are easily determined because NI B(n, 810) under H. 224). X to some space of values T, usually a Euclidean space. k * O, I, . 410 6 Because Z has rank p, it follows (Problem 6.4, 14) that {30 is consistent. (b) Deduce that Pearson's , R; = n , I I C; Tfj2 = n x' is given by (6.4.9) and has approximately a XZa-l) (b-1) distribution under H. KARLIN, S., A First Course in Stochastic Processes New York: Academic Press, 1 969. Find the UMP test. Let Lx (Bo, 81) = p (X, 81) jp (X, Bo) and suppose that Lx (Bo, 81) = p (X, 81) jp both PoO and Pe1 • Show that (a) For every 0 0). We find for z 2 0 0 I 2 y P(Y I 20) 7 7 • 1 ' • 111 • - These figures would indicate an association between heavy smoking and poor health be 0 cause p(2 1 20) is almost twice as large as py(2). , c, (B))T and let x be the observed data. We consider three situations (a) The problem dictates the parameter. The distribution of V is known as the noncentral x2 with n degrees offreedom and (noncentrality) parameter 82. = 0 and rt oo"). Problems to Section 1.6 1. If 0 < 00, then by (1), Eoo1 (X) < a and br is of level o: for H : () :< 00. ": ': ... (f) Show that '(g) Let $F(x) \cdot 1$. Multiparameter models are the rule. However, in this section, we will discuss three algorithms of a type used in different statistical contexts both for their own sakes and to illustrate what kinds of things can be established about the black boxes to which we all, at various times, entrust ourselves. (8 10 20) . I covers the material of Chapters 1-6 and Chapters 1-6 and Chapters 1-6 and Chapters 7-10 and includes Appendix A on basic probability theory. (Kiefer-Wolfowitz) Suppose (X,, . We achieve a similar effect, generating a family of level a tests, if we start out with (1 - a) k/vn 2 Jl. Evidently, (say) the level = 1 - a., Nk) (x, Jl) to equal ! if, and only if, X - tn_1(1 - a) k/vn 2 Jl. Evidently, (say) the level = 1 - a., Nk) Label, Treatment Dose Level) for patient i. () Suppose AO: .p (X, O). f. Each item produced is good with probability () and defective with pre $p(\cdot I z)$ is the frequency of a probability distribution because by {A.8.11}. Suppose there is no dependence between the quality of the items produced and let = if the ith item is the record of n Bernoulli trials with is good and 0
otherwise. The F Test for Equality of Scale. Assume the model log }i = !31 + f32 x i + f33 log xi + Ei , i = 1, . Then Y and Y are independent and the mean squared prediction error (MSPE) of Y is = = = Note that Y can be regarded as both a predictor of Y and as an estimate of p,, and when we do so, AISP E(Y) MSE(Y) + cr 2, where MSE denotes the estimation theory mean squared error. = E(X1)? Let (X1, X2) be a mndom sample from the distribution with density f(x, B) = g(x-B). • i 1 i 1 · 1 · 2. However, the form of this estimate is complex and if the model is incorrect it no longer is an appropriate estimate of E(X)/ [Var(X) 1 2 . : = rn f]2 < {3g and show that it agrees with the test of (b) Suppose that f3t i s unknown. Since Y is not known, we turn to the mean squared prediction error (MSPE) sure of "distance" is 1'> 2 (Y, g(Z)) or its square root yE(g(Z) - Y)2. 0.' 0.08 '.. See also Example 4., Xn, n > 2, be independently and identically distributed with density f(x, 8) = I - 0' exp { -(x - J1) /0'}, x 2 J1 , where 8 = (J1, 0'2), -oo < J1 < oo, 0' 2 > 0. B.1 CONDITIONING BY A RANDOM VARIABLE OR VECTOR The concept of conditioning is important in studying associations between random vari ables or vectors. For x E X, S(x) = s p(x,B) = q(s, B)r(x I s, B) (2.4.14) where $r(\cdot I \cdot, B)$ is the conditional frequency function of X given $S(X) q(s, B) I(B I B o) = log + Eo, q(s, Bo) If B o = Bold \cdot 8 = Bnew$. (b) Find a one-dimensional sufficient statistic for a when Jl. is fixed. , Xn be independently distributed with exponential density < < 2 1 the > ordered X's be denoted by Y1 Y 2 0 for 0, and let · · · e-xl (28)x Y It is assumed that Y1 becomes available first, then Yz, and so on, and that observation is continued until Yr has been observed. I 1 .22) and (A.I. 1.20), we see that if X1, dent with finite variances, then • Var(X1 + · · · + • . The power of this test is, by (4.1.2), (z(a) + (v./ii/(J)). Sometimes the choice of P starts by the consideration of a particular parameter. > 1. (x) = n L 1 [F x (-X,) < Fx (x)] i=l = nFr-u (Fx(x)). -T would then be a test statistic in our sense.) We select a number > c and accept H The value c that completes our specification is referred to as the critical value c and our test otherwise., 0, !3d+I, . The hypergeometric distribution with parameters D, N, and n : p(k) (A.I3.6) for k a natural numbers that 1t(D, N,n). (3.4.9) Lemma 3.4.1. Suppose that I and II hold and that & E & logp(X, B) < oo. (5.1.10) 1 implies that a2 < 1 with a2 = 1 possible (Problem 5.1.3), the right Because] X !] hand side of (5.1.9) when a2 is unknown be>omes 1 in e2 For ' = .1, 400, (5.1.9) is .25 whereas (5.1. 10) is .14. If ($j_i = P$ [A randomly selected individual is of type i for 1 and j for 2], then = {Nij : 1 :::; i :::; a, 1 :::; j :::; b} I".] M (n , ($j_i : 1 :::; j :::; a, 1 :::; j :::; b$]. Suppose the following loss function is decided on TABLE 1.3.1. The loss function (Drill) a, (Oil) (No oil) 01 82 1(0, a) (Sell) (Partial rights) a2 a3 0 10 12 I 5 6 ' ' i Section 1.3 25 The Decision Theoretic Framework Thus, if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill, the loss is zero, whereas if there is no oil and we drill there drill two-parameter exponential families and identify the functions 1], B, T, and h., Xn), where k 2:: 4 1 x,. If we use a(·) as a predictor and the new z has marginal distribution Q then it is natural to consider, l(P, a) = J(l'(z) - a(z))2 dQ(z), the expected squared error if a is used. The first cumulant c1 is the mean J-.t of X, c2 and c3 equal the second and third central moments f-£2 and f-£3 of X, and c4 = J.L4 - $3.u^{(1)}$. and c4 = J.L4 - $3.u^{(2)}$. Some further examples of variance stabilizing transformations are given in the problems. (b) Show that there exists, C < 00, 0 > 0 (depending on tj;) such that if jB 1°) - Bj < 0, then j il(i) - Bj < 0, that the EPV(8) for $1{T > c}$ is uniformly minimal in 8 > 0 when compared to the EPV(8) for any other test. (a) Show that this statement is equivalent to P(xp < Xp < Xp for all p E(0, 1)) Section 4.10 Problems and Complements 285 where x, \clubsuit sup{.r : a < .r < b, F f(>·) < p} and .r" \clubsuit inf{.r : a < x < b, fr - (x) > p}. 1.5). then it converges to that solution. Show that for the estimates (a) Var (a) : 2.::: = 1 nk = a and J'k in the one-way layout Var (8k) = ((p*)2 + Lk=; fi; k). Example 2.3.3. Multinomial Trials. This idea has been developed by Bickel and Lehmann (1975a, 1975b, 1976) and Doksum (1975a, 1975b, 1976b) and Doksum (1975a, 1975b, 1976b) and Doksum (1975a, 1975b, 1976b) and Doksum (1975a, 1975b) and Doksum (1975 in the preceding hint. 7) becomes 2 2 2 J, t nJt, + t3)(t, A5]t.), 6(Jt, - n(Jt))T = 0, which with /i2 = n - 1 Lxl simplifies to 2 J + A6x11 - A6it2 = 0 - 1 ' 11 + ± = 2 [>..o x ± .), o Note that 11+11 - = -A6Ji:2 < 0, which implies i''i + > 0, 11 - < 0. Construct the interval using F-1 and Fu 1 + 14. Establish {6.4.14}. ' z, 2, . (1.2.6) To calculate the posterior probability given in (1.2.6) we argue loosely as follows: If be fore the drawing each item was defective with probability . a B(n DfN) distribution. In all of these cases, Co, the common distribution of T(X) under () E 80, has a closed form and is tabled. (b) Recall that J..li = E(Yi) = Ai 1 . A self-crossing of maize heterozygous on two characteristics (starchy versus sugary; green base leaf versus white base leaf) leads to four possible offspring types: (1) sugary-white; (2) sugary-green; (3) starchy-white; (2) sugary-green; (3) starchy-white; (4) starchy-green; (3) starchy-white; (4) starchy-green; (3) starchy-green; (3) starchy-green; (3) starchy-green; (3) starchy-green; (3) starchy-green; (3) starchy-green; (4) starchy-green; (5) starchy-green; (1) sugary-green; (2) sugary-green; (3) starchy-green; (4) starchy-green; (5) starchy-green; (5) starchy-green; (6) starchy-green; (7) starc n B) - P(A)P(B) P(A)(1 - P(A))P(B) (1 - P(A))P(B) (1 - P(B)). Let S = S(X) be a map from X to subsets of N, then S is a (1 - a) confidence region for v if the probability that S(X) contains v is at least P ' X C Rq all P E P. We think of J1- as representing the treatment effect. It is not hard to check (using Laplace transform theory) that a one-parameter exponential family quite generally satisfies Assumptions I and II. x2 Xtt/(X1 X2) XI (X, X Xt +X, t, Section 1.5 43 Sufficiency that will do the same job. 4.9 4.9.1 L I K E L I H O O D RAT I O P R O C E D U RES I nt rod uction Up to this point, the results and examples in this chapter deal mostly with one-parameter problems in which it sometimes is possible to find optimal procedures. Let Ri = Nil + Ni2t ci = Nil + Ni2t ci = Nil + Ni2t ci = Nil + Ni2t Show that given Rr = TI, R2 = r2 = n - Tr, Nu and N21 are independent B(r 1, 8u / (8u + 81,)). A generic source of trouble often called grf!SS errors is discussed in greater detail in the section on robustness (Section 3.5.3). New York: McGraw-Hill, 1975. This is an approximation (for large k, n) and simplification of a model under which (N1, . 8. So, although the sample space is well defined, we cannot specify the probability distributions for X, any one of D which could have generated the data actually observed. Find the posterior distribution 1r(O I x) and show that if >. The first and second components of this vector are called the D sample mean and the sample variance, respectively. 4. In some applications we often have a tested theoretical model and the danger is small. This is a consequence of the following theorem, which reveals that (4.6.1) is nothing more than a comparison of (X)Wer functions. • ' i, j A function convex. Lemma 1.4.1. E(Y - c) 2 is either oofor all c or is minimized uniquely by c = p, = E(Y). Suppose the Zi in Problem 6.4.11 are obtained as realization of i.Ld. Zi and m, so that (Z; , X;) are i.i.d. with (X, I Z;) $\otimes B(m, 1r(.Li2Zi))$. : $py(y) "" "; <math>\otimes \otimes 'i$: 1 V(g(A(y))) V(A(y)) V(A(y)) V(A(y)) V(g-1(A(y + 1))). (b) using Fisher's exact test of Problem 6.4.5? ny D c.::, '-;-o-, 19 12 0 5 Admit Men Women Hint: (b) It is easier to work with N22. (DN, N, n), distribution where DN fN ---> p as N ---> oo and XN has a hypergeometric
11, is fixed. Our book contains more material than can be covered in tw p qp.arters. Show that the noncentral t distribution, 7k,6• has density 1 {= x U•-'I e-H x+C tVxfk -ol'ldx. The "only if' part in (b) follows because jCxl2 > 0 unless x = 0 is equivalent to Cx =/:- 0 unless x = 0, which is nonsingularity. , n, are n independent random samples, where N(tLi , aJ). As we shall see, although loss functions, as the name suggests, sometimes can genuinely be quantified in economic terms, they usually are chosen to qualitatively reflect what we are trying to do and to be mathematically convenient. As in Section 6. Let X1, , Xn denote the times in days to failure of n similar pieces of equipment. In fact by varying our assumptions this class of models includes any situation in which we have independent but not necessarily identically distributed obser vations. Justify formally the following expressions for the moments of h(X, Y) where (X, , Yi), . Substituting in (5.3.6) we find Var(X): 1/4 and vn((X) - (>) 1) has approximately a N(o, 1/4) distribution. [j/2] + I) where Cj = max 1 0 ' P[IUn - ul < li] (1.1 · In Example 4.9.3 we saw that the two sample t statistic Sn = nn2 cY x), n = nt + n2 has a 'Tn-2 distribution under H when the X's and Y's are normal with af = a . Corr(X 1, X2) J(Var X1)(Var X2) 'I I''' • ''' • i • 1 • for any two random variables Z'' z, such that E(Zf) < oo, E(Zi) < oo, E(Zi) < oo, 4 407 Large Sam ple Methods for Discrete Data An important alternative form for Z is given by (6.4.8) Thus, z = y'n[P (A I B)- P(A I B)] ($\hat{\Psi}(B) \ \hat{\Psi}(\Phi)$] B, B, P(A) P(A) 1 12 where P is the empirical distribution and where we use A, A, B to denote the event that a randomly selected individual has characteristic A, A, B. AND K. may be quite irrelevant to the experiment that was actually performed. Closely related to the latter is what

we shall call confidence interval loss, l(P, a) = 0, fv(P) - of < d, l(P, a) = 1 otherwise. Let Z = Ef 1 i, the total number of successes. Example 1.2.1. Bernoulli Trials. X1). Particularly important is the case Eo = Et when "Q large" is equivalent to "F = (tt 1 - p.0)EQ1X large." The function F is known as the Fisher discriminant function.] 7). Contiguity MONTE CARLO METHODS The Nature of Monte Carlo Methods Three Basic Monte Carlo Methods Three Basic Monte Carlo Methods The Bootstrap Markov Chain Monte Carlo Methods The Bootstrap Markov Chain Monte Carlo Methods Three Basic Methods Three Basic Methods Three Basic Monte Carlo Methods Three Basic Monte Carlo Methods Three Basic Methods Three Estimates: Reducing Boundary Bias Regularization and Nonlinear Density EstimatesConfidence Regions Nonparametric Regression for One Covariate PREDICTION AND MACHINE LEARNING Introduction Classification and Prediction and Prediction and Prediction Asymptotic Risk Criteria Oracle Inequalities Performance and Tuning via Cross Validation Model Selection and Dimension Reduction Topics Briefly Touched and Current Frontiers APPENDIX D: SUPPLEMENTS TO TEXTAPPENDIX E: SOLUTIONS REFERENCES INDICES Problems and Complements appear at the end of each chapter. For instance, l(P, a) = l(v < a), which penalizes only overestimation and by the same amount arises naturally with lower confidence bounds as discussed in Example 1.3.3. If v = (v, . Sup- pose that we have a regular parametric model {Pe : () E 8}. This problem arises when we want to compare two treatments or a treatment and control (nothing) and both treatments are administered to the same subject. (a) The observations are indicators of Bernoulli trials with probability of success 8. (a) Find an approximation to P[X < t] in terms of f] and t., BJ Bi, BJ + 1, Regular models. Show that hv(t) = C.h.(t) if and only if Sv(t) = sg.(t). 7. We have just shown that the information I(B) in a sample of size n is nh (8). A second example is consecutive measurements Xi of a constant 11- made by the same observer who seeks to compensate for apparent errors. , Y E Rd are i.i.d. vectors and EIYt l k < oo, where I \cdot I is the Euclidean norm, then for all integers k: n where C depends on d, EIY 1 [k and k only. Lemma 5.3.1. If EjX 1 lj < that oo, j > 2, then there are constants Cj > 0 and Dj > 0 such (5.3.3) (5.3.4) Note that for j even. (b) If n is fixed and divisible by p, then Var(a) is minimized by choosing ni ni n is nl2, n2 (c) If = fixed and divisible by np nl2(p - = (d) Give the ··· = 2(p - 1), 1). Important special cases are the N (J-L, 0" 2) and gamma (p, .A) families. 1 +0, . These comments are ordered by the section to which they pertain. 15. ... Among many others who helped in the same way we would like to mention C. Finally, note that the Neyman Pearson LR test for H: () = ()0 versus K: Oo + t, $\notin > 0$ rejects for large values of 1 - [log pn(X1, .) and the sum is over all 1, 2, . We consider a function that we shall call a contrastfunction p: X x 8 ---> R and define E8, P(X, 8). However, there are easy conditions under which conditions of Theorems 6.2.2, 6.3.2, and 6.3.3 hold (Problem 6.5.3), so that the MLE {3 is unique, asymptotically exists, is consistent with probability 1, and (6.5.9) 1 What is J- ({3)? ": i i • Printed in the United States of America 10 9 8 7 ISBN: 6 5 4 3 2 I 1 D-13-850363-X ' Prentice-Hall International (UK) Limited, Prentice-Hall of Australia Pty. Let V and W be independent with W x 2mand V having a noncentral xi distribution, with noncentrality parameter 02. Equation $\{1.2.2\}$ is an example of $\{1.2.2\}$ is a example of $\{1.2.2\}$ is {1.2.3). (c) Suppose that for each a E (0, 1), the UMP test is of the form 1 {T > c}. Because l(ij 1) = >., Xn and -X1, . has a bivariate is said to have a Show that { * exp if.l. r + i1J.2 + i2 I. This is done in the context of a number of classical exam ples, the most important of which is the workhorse of statistics, the regression model. B E 8, 8/i]B log p(x, B) exists and is finite. If TJ(O) is strictly increasing in () E 6, then this family is MLR. 99 100 Chapter 2 Methods of Estimation More generally, suppose V (80, 8) = : XxR d - Rd, W - ('ljl1, E11, w(X, 11). We follow the notation of Example 1. Of course, we don't know the truth so this is inoperable, but in a very weak sense (unbiasedness), p(X, 8) is an estimate of D(80, 8). DE GROOT, M. How many days must you observe to ensure that the UMP test of Problem 4.3.1 achieves this? One such function (Lindley, 1998) is • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of variance stabilizing (Lindley, 1998) is • • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles of Variance stabilizing (Lindley, 1998) is • • • In Agior examples are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models of Section 6.5. The comparative roles are the generalized linear models are the and canonical transformations as link functions are discussed in Volume II. P = L = I ii > 2) are zero. the number of homozygous dominants, has a binomial (n,p) distribution. , Yn in canonical exponential family form. Then restrict attention to tests that in fact have the probability of rejection less than or equal to a for all () E 80. Thus, we observe Y1 , . 77 4.61 5.99 7.38 7.82 9.21 10.60 1 1 .98 13.82 15.20 3 4. Here is the method: If 1j0ld is the cunent value of the algorithm, then 1 Tinew = fiold - k Ciiold)(A(iiold) - to). (2) Note that this inequality is true but uninteresting if 1(8) = oo (and , P'(8) is finite) or if Vare(T(X)) = oo. Note that = = = Y-Y = X - 1 Xn + 1 "' N(O, [n - + 1]cr 2). In this section we present some results useful for prediction theory, esti mation theory, and regression. III. Doksum. Mathematical Statistics: Basic Ideas and Selected Topics, Volume 1. Interpret this result. Chapter 5 of the new edition is devoted to asymptotic approximations. 19. Hint: Let t.(x) = F � (Fx(x)) - x, then nF-x (x + 1:.11.22) This may be checked directly. Xn1 be i.i.d. F and Y]. Show that if q(X) is a level then [q(X), q(X)] is a level interval arbitrarily if q Hint: 6. BERNARDO, J. 188 Measures of Performance 304080 The Normal Case. (2) The generation of enormous amounts of data-terrabytes (the equivalent of 10 1 2 characters) for an astronomical survey over three years. Exhibit the two solutions of (6.4.1) explicitly and find the one that corresponds to the maximizer of the likelihood. (6.4.1) logistic linear regression model of Problem 2. p,2) with common major axis given by (Y-1'2) = (a) Let S, = {(x, y) : P(X,Y)(X, y) = ; ' i Section B.ll Problems and 533 Complements (x -111) if p > 0, (y -1'2) \bullet - (x - 1'1) if p < 0. Example 3.4.2. Suppose X1, - & log $p(x' 0) \bullet$ & o. (a) Using the relation E(e'Y) = E(E(e'Y I X)) and the uniqueness of moment gen erating functions show that Y has a P(.Ap) distribution. Measures of Performance. (c) We have a qualitative idea of what the parameter is but there are several parameters that satisfy this qualitative notion. , $-\Phi - - + k? \cdot 4$. (a) Show that if in testing H : {} = {} 0 versus K : f] = f] 1 there exists a critical value c such that P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80,
81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c] = I - P0, [L(X, 80, 81) > c constant. Intuitively, in estimation we care how far off we are, in testing whether we are right or wrong, in ranking what mistakes we've made, and so on. Solve (_1, = 0 by the method of g ve are, in testing whether we are right or wrong, in ranking what mistakes we've made, and so on. Solve (_1, = 0 by the method of g ve are, in testing whether we are right or wrong, in ranking what mistakes we've made, and so on. where X = y'(n - 2)RT = y'(n - 2)RTbased on data X such that P (Y ::; Y ::; Y) ?: 1 n. If the control and treatment responses are independent and nonnally distributed with the same variance a2, we arrive at the one-way layout or p-sample model, (6.1.6) where Ykl is the response of the lth subject in the group obtaining the kth treatment, /3k is the mean response to the kth treatment, and the \notin lare independent N(O, \clubsuit) random variables. - . Suppose and 0 < I(B) < oo. AND D. We may take any one-to-one function of () as a new parameter. It includes the initial theory presented in the first edition but goes much further with proofs of consistency and asymptotic normality and optimality of maximum likelihood procedures in inference. In some applications, for example, the bivariate normal case (Problem 2.3. 13), the following corollary to Theorem 2.3.1 is useful. • is binomial, B(n, 8), that "! is a Bayes estimate for 206 Hint: Given 21. (4., n corresponding to the number of defective items found. We define a decision rule or procedure 0 to be any function from the sample space taking its values in A. Expression (! .5.10) where X(n) = max{x1, . For instance, in Example 1.1.3 we could take z to be the treatment label and write our observations as (A, X1). Nevertheless, we can draw guidelines from our numbers and cautiously proceed. t of the theory of games. A feature of Bayesian models exhibited by this example is that is sufficient. 10.16) and (8.10. We give the proof of (5.3.4) for all j and (5.3.3) for j even., 9. KARLIN, S., Mathematical Methods and Theory in Games, Programming, and Economics Reading, MA: Addison-Wesley, 1959. 5 5 The posterior predictive distribution is also used to check whether the model and the prior give a reasonable description of the uncertainty in a study (see Box, 1 9 83). y: - n L... ** a dimension of MLEs, Wald and Rao statistics, generalized linear models, and more. Often there is a function q(B) < q0 and K : such that H and K can be formulated as H : > q0 • Now let q(O) q(O) 232 Testing and Confidence Regions Chapter 4 q1 > q0 be a value such that we want to have power $f_J(O)$ at least fJ when q(O) > q1. A list of the most frequently occurring ones indicating where they are introduced is given at the end of the text. We will write Fx for F when we need to distinguish it from the distribution F-x of -X., (Xn, Yn) be a sampj from a bivariate.N"(O, 0, 0. That is, . FERGUSON, T. • -)•]. 524 Additional Topics in Probability and Analysis Appendix 8 The identities and inequalities of Section 1 ..-1 can readily be seen to be special cases of (B. I:I 534 Additional Topics in Probability and Analysis Show that X and only if, p Appendix B and Y have normal marginal densities, but that the joint density is normal, if = 0. Here are some examples: ": : (b) Find the optimal test statistic for testing H J.L = J.Lo versus K : a1 -= 1 a2. Thus, we would want a posteriori estimates of performance. , xn) ; that is, Q (· I x) has in the continuous case density , n where not all the z's are equal.), and let p(x, B) = Po[X = x]. We will discuss the fundamental issue of how to choose T in Sections 4.2, 4.3, and later chapters. 13) the rela- + · · · + Xn) = L Var X, + 2 L Cov(X, . A.13.5 If X, , X2, B(n2, 0), . after the 19 sample items have been drawn. S(to) is a confidence interval for f.l for a given value to ofT, whereas A* (f.lo) is the acceptance region for Hp.o. (iii) k(fJ, a) increases by exactly (iv) k(O,a) I at its points of discontinuity. is the statistic T(X1). In the discrete case we will use both the terms frequency function and density for p(x, 0). S Suppose S and T are subsets of and g is twice differentiable. We define q* to be a uniformly is the statistic T(X1). most accurate level (1 - a) LCB for q(0) if, and only if, for any other level (I - a) LCB q, Pe[f < q(O')] < Pe [q < q(O')] whenever q(() < q(8). S., AND P. Note that V has a noncentral $x^{(0)}$ distribution with parameter 82. Corollary 2.3.2. Consider the exponential family k p(x, B) = h(x) exp I; c; (B)T;(x) - B(B), x E X, B E e. i ' i j ' l • Poisson processes are frequently applicable when we study phenomena involving events that occur "rarely" in small time intervals. Then the set $C = \{(t,v) : a(t,v) < a\} = \{(t,v) : a$ and g 1 1 are densities, we must have PT = g 🗞 🌵 1 1 2 2 D and the result follows. (B1,82)r 3 The graph shows log likelihood contours, where the log likelihood contours, where the log likelihood is constant. Estimation, testing, confidence regions, and more general procedures will be discussed in Chapters 2-4. I 1. " '., " .J(Bnew I Bold) - Eo " = ld o r (X I s, uold) q (s, uold) (2.4. 16) Now, J(Bnew I Bold) > J(Bold I Bold) > J(Bold I Bold) = 0 by definition of Bnew . 1 Xn be a sample from a population with density p(x, 8) given by - p(x, B) 0 otherwise. h from a convex set S to R is said to be (strictly) concave if g = Jensen's Inequa to be (strictl that we wish to test H : fh < {3g versus K fh > {3g. We want to contrast w to the case where there are no restrictions on 11; that is, we set n = R k and consider TJ E fl. If B is uniformly distributed on (- "/2, 1r/2) show that Y \clubsuit tan B has a Cauchy di stri bution whose density is given by p(y) = l/[1r(1 + y2)). (a) Deduce that depending on where bisection is started the sequence of iterates may converge to one or the Other of the local maxima (b) Make a similar study of the Newton-Raphson method in this case. Show directly using the definition of the rank of an ex}X)nential family that the multi nomial distribution, M(n; B, . (1.7.2) where ho(t) is called the baseline hazard function and g is known except for a vector {3 = ({31, . ! 1: ' I ' I ' I ' Section 4.4 Confidence Bounds, Intervals, and Regions Definition 4.4.1. A statistic v(X) is called a level (1 for every P E P, 235 - o) lower confidence bound for v if P[v(X) < v] > I - o., xN) as parameter 1 = n - k L:f 1 Xj. It is easy to see that the natural X = 🗞 I: 1 x is unbiased (Problem 3.4.14) and has MSE(X) Are the principal axis theorem because (B. The "if' part in (b) follows by noting that C nonsingular iff det(C) ol 0 and det(CCT) = det2(C). ' ' ' Combining Theorem 5.3.3 and Slutsky's theorem, we see that here, too, if H is true Tn £ N(O, 1) so that z1 0 is the asymptotic critical value. RoussEUW, AND W. 15. Let X1 '. 2xTy - yT By all x., Un)/S1. Define (W2, . NoRBERG, R. Be cause the book is an introduction to statistics, we need probability theory and expect readers to have had a course at the level of, for instance, Hoel, Port, and Stone's Introduction to Probability Theory. We prove that T = 1 + and are independent and the first of Theorem B.2.3, whatever be t. TUKEY, Robust Estimates of Location: Sun>ey and Advances Prince ton, NJ: Princeton University Press, 1972. The x distribution is extremely skew, and in this case the tn-l (0.95) approximation is only good for n > 102.5 316. Newton's method also extends to the framework of Proposition 2.3.1. In this case, if l(B) denotes the log likelihood, the argument that led to (2.4.2) gives (2.4.3) Example 2.4.3. Let X1 , . If the treatment and placebo have the same effect, the difference Xi has a distribution that is 258 Testing and Confidence Regions C
hapter 4 symmetric about zero. 15)., Xn = Xn] = 8'(1 - 8)n-t (1.5.1) where Xi is 0 or 1 and t = I: 1 - 8) and t = I: pennitted). Chapter 3 | I j X1, . In our new formulation it is the Xt that obey (3.5.1). ' L I! I a-2) model. 9 L i keli hood Ratio P rocedu res 257 likelihood confidence regions, bounds, and so on., Yn are independent. ft that Notes for Section 1,3 � (I) More natural in the sense of measuring the Euclidean distance between the estimate (} and the "truth 0. It is used in a classification context in which 90, 91 correspond to two known populations and we desire to classify a new observation X as belonging to one or the other. , Xn) is an estimate of a real parameter f indexing a family of distributions from which X 1 , , Xn are an i.i.d. sample. 15.15) we see that Nn (t) !:. One reasonable mea (g(Z) - Yf, which is the squared prediction error when g(Z) is used to predict Y. If we suppose the new drug is at least as effective as the old, then 8 = [Bo, 1], where Bo is the probability of recovery using the old drug., k, then the convex support of the conditional distribution of z::::: 1 A1 YJ zU l given Zj = z U), j = 1, . Show that the interval in parts (e) and ! k and l in part (e) can be approximated by (0 can be derived from the pivot T(xp) Hint: . Such tests are said to be UMP (unifonnly most powerful)., a - 1, f[]2 j = 7. N(O, 1/4/(0)). I (! logp(X,B)) . Next consider the setting of Proposition 8 E e open c • RP, 2.3.1 in which lx(O), the log likelihood for is strictly concave. i 1 ' ' , i' j i J, ' j' . , Bk), then a,N, + +akNk has approximately a N(np,, na2) distribution, where 11 = L7 1 aifji and a2 = I: 1 1 fji(ai 11)2, find an approximation to the critical value of the MP level a: test for this problem. As such it is certainly a valuable contribution to our advanced literature on theoretical statistics."~RobertW., Nr) be multinomial M(n, 9), 0 = r (B h . A.12.5 If defined, Alx determines the distribution of X uniquely and is itself uniquely determined by the distribution of X., Xn,Xn+l are i.i.d. f(x I 8), 6 "" 1r, the predictive 14. Give a distribution of X uniquely and is itself uniquely determined by the distribution of X. and Xn+l are i.i.d. f(x I 8), 6 "" 1r, the predictive 14. Give a distribution of X uniquely and is itself uniquely determined by the distribution of X. and Xn+l are i.i.d. f(x I 8), 6 "" 1r, the predictive 14. Give a distribution of X uniquely and is itself uni 1.1.3. Let the actions corresponding to deciding whether a < 0, (} 0 or a > 0 be penoted by -1, 0, 1, respectively and suppose the loss function is given by (from Lehmann, 1957) = = Statistic 0 b+c c 0 where b and c are positive. We know that, given (} = hypergeometric distribution 'H(i, N, n) . Show that Eh(X) h(Jl.) + ih(21 (Jl.) + : + Rn where I Rn l < M31 (Jl.)[J1.3I/6n2 + M(J1.4 + 30'2) /24n2 Hint: "Therefore, 24. We show that Bayes rules with constant risk, or more generally with constant risk, or more generally with constant risk, or more generally with constant risk over the suppose IM41(x) < M for all x and some constant M and suppose that 114 is finite. Frechet, we shall ' ' ' Section 211 3.8 References follow the lead of Lehmann and call the inequality after the Fisher information number that appears in the statement., Xn) -+ x is not close to 0 or 1 and 5. (A.16.3) ·· ----- 474 A Review of Basic Probability Theory The first of the inequalities in (A. (3.4. 1 1) j { [:8p(x, B) /p(x,B) p(x, B) /p(x, B) /p(x assertions of Table 1. The method that we employ to prove our elementary theorems does generalize to other measures of distance than 6.(Y, g(Z)) such as the mean absolute error E(lg(Z) - Yl) (Problems 1.4. 7-1). In others, we can be reasonably secure about some aspects, but not others. The correlation inequality corresponds to the special case Z1 = X1 - E(X1), Z2 = X2 - E(X2)., 8k, respectively. An estimate J(X) is said to be shift or translation equivariant if, for all X1, . Two examples in which the MP test does not depend on ()1 are given. 5.4.2 5.4.3 6.2 1-. We introduce the simple likelihood ratio statistic and simple likelihood ratio (SLR) test for testing the simple hypothesis H : () = Bo versus the simple alternative K : 0 01• The Neyman-Pearson lemma, which states that the size a SLR test is uniquely most powerful (MP) in the class of level a tests, is established. We define S as the "-field generated by SA, B 🖗 { F E : F : Pp(A) E B}, A, B E B, where B is the class of Borel sets.) (v 0, then the coefficient of skewness and the kurtosis of Y area (MP) in the class of level a tests, is established. the same as those of X. We XV Preface to the Second Edition: Volume I also. Such randomized tests are not used in practice. h () = /i is a variance stabilizing transformation of X for the Poisson family of distributions. As the second example suggests, there are many problems of this type in which it's unclear which of two disjoint sets of P's; Pg Po testing problem is really one of discriminating between Po and PO. Suppose we wish to sample from a finite population, for instance, a census unit, to determine the average value of a variable (say) monthly family income during a time between two censuses and suppose that we have available a list of families in the unit with family incomes at the last census. In order to reduce differences due to the extraneous factors. Suppose p(x. Compared to the frequentist bound (n 1) s 2 I Xn-l (a:) of Example 4.4.2, a is shifted in the direction of the reciprocal bI a of the i s a level mean of 1r(A). (b) Give an expression of the power in terms of the X is clear, though, 1 that this fonnulation is inadequate because by taking ' oo v we can achieve risk = 0. CoVER, T. If H holds, P{X 0, such that (i) (ii) 1 ! j 1 • • ' N(t) has a P(:lt) distribution for each t. Show that if 0 is translation equivariant and antisymmetric and Eo(O(X)) exists and is finite, then - -.. Thus, the power is 1 minus the probability of type II error. (c) Explain how it is possible if Po P[o(X) = 9] = L 9 and the Bayes estimate with X is not a Bayes estimate for any prior 1r. Part (b) is more difficult. Similarly, a level (1 any fixed e and all - (4.6.1) is more accurate than a competitors are called uniformly most accurate as are upper confidence bounds satisfying (4.6.2) for all competitors. • • = = ! I I; •; Section 1.1 Data, Models Parameters, and Statistics 5 How do we settle on a set of assumptions? pz (z) p(y z) Eyp(y I z) Eyp more specialized texts. (4.5.3), Xn are i.i.d. N(J-l, a2) with u2 known. We will use asymptotic theory to study the behavior of this test when we observe i.i.d. X1, . . , k, a E Rk. - - = X. i i. H., Optimnl Statistical Decisions New York: McGraw Hill, 1970. Supplements to Text. • • j Section 4.5 247 The Duality Between Confidence Regions and Tests Applications of Confidence Intervals to Comparisons and Selections We have seen that confidence intervals lead naturally to two-sided tests. The experimenter may, for example, assign the drug saltematively to every other patient in the beginning and then, after a while, assign the drug that seems to be working better to a higher proportion of patients. We conclude by indicating to what extent the relationships suggested by this picture carry over to the general decision theoretic model. In any case, whatever our choice of procedure we need either a priori (before we have looked at the data) and/or a posteriori estimates of how well we're doing. PF[Xn < x] "' il> vn a (5.1.8) n, Again we are faced with the questions of how good the approximation is for given x, and What we in principle prefer are bounds, which are available in the classical situations of (5.1.6) and (5. = N(O, a) is the critical value using the Welch = approximation, has asymptotic level a. Suppose that X = bution function (X1, •., cn-d p(e l)p(e, I el)p(e3 I e,). Now p(y, 9) is a curved exponential family of the form (2.3.6) with •. Hint: h(X, Y) - h(Jl.t, J1.2)(Y - J1.2) + 0(n - t). Thus, P[!OOII > 20 I X = 10] P[lOOII - X > 10 I X !OJ (10011 X) - 8 1 1.9 > P (1.2.7) J81(0.9)(0.1) "' I - (0.52) 0.30. LINDLEY, MANDEL, J., The Statistical Analysis of Experimental Data New York: J. (c) Check the identity E[E(Y | Z)] = E(Y) (i) 1 -, $z^2 + Y^2 < 1$ P(Z, Y) (z,y) 1r 0 otherwise. Within each section of the text the presence of comments at the end of the chapter is signaled by one or more numbers, 1 for the first, 2 for the second, and so on. M., "A Stochastic Model for the Distribution of HIV Latency Time Based on T4 Counts," Biometika. Before sampling any items the chance that a given by n 1 n . But the normal approximation with continuity correction, Xn) is given by n 1 n . But because N;/n is unbiased and has 0 variance >.;(1 - >.;)/n, then N;/n is UMVU for >.; . with p = \mathbf{A} k and .\ = \mathbf{A} is referred to as the chi squared density with k degrees offreedom and is denoted by x%. , ' Central Limit Theorem Let {Xi} be a sequence of independent identically distributed random variables with (common) expectation J.t and variance a 2 such that 0 0. Let B = (B" 82) be a bivariate parameter. Mathematical Statistics: Basic Ideas and Selected Topics, Volume II will be published in 2015. • .'j .j ' 1 .I' , ' ' · I , • • 11. , x m), y = (Yl 1 · . . It follows that if --C, then D(Fx . Show that the UMP test is based on the statistic Xn be i.i.d. with distribution function F(x). There have, of course, been other important consequences such as the extensive development of graphical and other exploratory methods for which theoretical development and connection with mathematics have been minimal. [2,m2] - f2 exp{- 2,..2 L (x; - JL)2} $= I n n n \{2 1 2 - [?r0'2t''i2 [exp{- 2,..2 L (x; - JL)2}]$ Assume that $A = \{X : p(X I 8) > 0\}$ does not involve e. CASELLA, Theory of Point Estimation, 2nd ed. Thus, A is positive definite iff all its eigenvalues are positive. By (A.S.IO), X is distributed r(p, .>.) if, and only if, >.X is
distributed r(p, .>.) if, and only if, >.X is distributed r(p, .>.) if and only if, >.X is distributed r(p, .>.) if and only if, >.X is distributed r(p, .>.) if and only if, >.X is distributed r(p, .>.) if and only if, >.X is distributed r(p, .>.) if and only if, >.X is distributed r(p, .>.) if and only if all its eigenvalues are positive. parameter (}, but we do not necessarily observe X*. Hint: 'f' n n n i=l i=l c, L Y• + c, L x;y; = I;(c, + c,x,)(c,x, + c1 > 0). As in Example 1.6. 9, suppose X1, . (II) If T is any statistic such that Ee (ITf) < oo for all B E 8, then the operations of integration and differentiation by (} can be interchanged in J T(x)p(x, B)dx. , Xn) is a sample of a N(J.L, a-2) random variables with a2 known. All actual measurements are discrete rather than continuous. Is n > 100 enough or does it have to be n > 100, 000? It's important to note that even nonparametric models make substantial assumptions-in Example 1.1.3that X1, ... D Suppose X1, . (a) For the following data (from Hald, 1 952, p. appears reasonable. n+1) T has a density given by n!, ti > 0, 1 < i 3. z) for z > 2 is an integer. For instance, in Example 1.1.1 we can use the number of defectives in the population, NO, as a parameter and in Example 1.1.2, under assumptions (l)-(4), we may second moments of the normal distribution of the observations (i.e., by (tt, tt2 + a')). (c) Find the minimax rule among J1, . . , Xn be a sample from the logistic distribution with d.f. $F(x, B) = [I + exp\{-(x - 8)\}] - 1 \cdot The density is We find I I exp\{-(x - 8)\}] - 1$ two-dimensional sufficient statistic for 8. Section 4-4 •, Xn be a sample from a normal population with unknown mean J.L and cr2. L g $(X_{r}, B) : [t [< M @ 0, i=1 i==1 M, P0 • 5.5.3), (5.5.13), (SLLN) [e - 1 We use sup for all 341 here. (c) Find the posterior distribution of a. StrJ.tist. We see that D$ $E(Y - c)^2$ has a unique minimum at c = p, and the lemma follows. Here, $= II(Y \ L)$ That is what (1.4.4) tells us. -oo < y < oo. el - 8 new., Yr is an exponential family with density L:: $ere V + (n - r)yr \ 28 -] < < <, 0Yl - . Ku converges in a neighborhood of zero. A natural are X = {(zi, Yi) : 1, $(I) i 0, p > 0, >... • , Xn are i.i.d. N(J-t, u2) and O:t + 0:2}$ < a:, then the shortest level (1 - a) interval of the form is obtained by taking [x- z(1 - a,) fo' X + z(1 - a,) ;;;] a:1 = a:2 = aj2 (assume a-2 known). (The negative binomial distribution is that of the number of failures before the nth success in a sequence of Bernoulli trials with probability of success 0.) Hint: By Theorem 1.6.1, Pe [L + 2] a:1 = a:2 = aj2 (assume a-2 known). (The negative binomial distribution is that of the number of failures before the nth success in a sequence of Bernoulli trials with probability of success 0.)</pre> 0 < 9 < I. We let X1, . t(Yh. Given a closed linear subspace C of 1i we define the projection operator 11(· I C) : 1i 🏶 C by: 11(h I C) is that h' E C that achieves mi (11 h h'll : h' E C). We shall pursue this approach further in Chapter 3. See A.lO. This corresponds to an a priori bound on the risk of a on v(X) viewed as a decision procedure with action space R and loss function, l(P,a) 0, a > v(P) 1, a < v(P) 24 Chapter 1 Statistical Models, Goals, and Perform ance Criteria an asymmetric estimation type loss function., Xm be i.i.d. F, Y1, . That is, for integration over Rq, :0 j T(x):Op(x, B) dx j T(x):Op(x, awkward notation when the meaning is clear. The distribution of the response Yi for the ith subject or case in the study is postulated to depend on certain characteristics zi of the ith subject. Random variability I ' i Section Ll Data, Models, 3 Parameters, and Statistics here would come primarily from differing responses among patients to the same drug but also from error in the measurements and variation in the purity of the drugs. Next we choose a parametric model for 1r (z) that will generate useful procedures for analyzing experiments with binary responses. are given by and that a level (1 a) confidence intervals for a level (1 b) Find confidence intervals for a level (1 b) Find confidence intervals for a level (1 b) Find confidence intervals for a level (1 c) that will generate useful procedures for analyzing experiments with binary responses. * + tJ3) * a) ::; a2 ::; (n - p) s2 lx n-p (* a) . , Yn are independent with vari ances I and E(Y1) = 0, E(Y;) = 0, i = 2, . Show the distribution of X form an r-parameter exponential family and identify fJ, B, T, and h. Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro ! To Erich L Lehmann ,I ' ' ' -- -- ----- "- CONTENTS PREFACE TO THE SECOND EDITION: VOLUME I xiii PREFACE TO THE FIRST EDITION xvii I STATISTICAL MODELS, GOALS, AND PERFORMANCE CRITERIA I 1.1 1.1.1 Data and Models I I 1.1.2 Parameters, and Statistics 1.1.4 Examples, Regression Models, Regression Models, Parameters, and Regression Models, Parameters, and Regression Models, Parameters, and Regression Models, Regression Mo Bayesian Models 12 1.3 The Decision Theoretic Framework 16 1.3.1 Components of the Decision Theory Framework 17 1.3.2 Comparison of Decision Procedures 24 1.3.3 Bayes and Minimax Criteria 26 1.4 Prediction 32 1.5 Sufficiency 41 1.6 Exponential Families 49 1.6.1 The One-Parameter Case 49 1.6.2 The Multip arameter Case 53 Building Exponential Families 56 1.6.4 Properties of Exponential Families 58 1.6.5 Conjugate Families 58 1.6.5 Conjugate Families 58 1.6.5 Conjugate Families 58 1.6.3 1.7 Problems and Complements 66 1.8 Notes 95 1.9 References 96 • • VII • • CONTENTS VIII 2 METHODS OF 2.1 2.2 2.3 *2.4 3 Basic Heuristics of Estimation 2.1.1 Minimum Contrast Estimates; Estimating Equations 2.1.2 The Plug-In and Extension Principles Minimum Contrast Estimates and Estimating Equations 99 99 102 107 2.2.1 Least Squares and Weighted Least Squares and Weighted Least Squares 107 2.2.2 Maximum Likelihood in Multi parameter Exponential Families Algorithmic Issues 121 127 2.4.1 The Method of Bisection 127 2.4.2 Coordinate Ascent 129 2.4.3 The Newton-Raphson Algorithm 132 2.4.4 The EM (Expectation/Maximization) Algorithm 133 2.5 Problems and Complements 138 2.6 Notes 158 2.7 References 159 MEASURES OF PERFORMANCE 161 3.1 Introduction 161 3.2 Bayes Procedures 161 3.3 Minimax Procedures 170 Unbiased Estimation and Risk Inequalities 176 *3.4 *3.5 4 ESTIMATION 3.4.1 Unbiased Estimation, Survey Sampling 176 3.4.2 The Information Inequality 179 Nondecision Theoretic Criteria 188 3.5.1 Computation 188 3.5.2 Interpretability 189 3.5.3 Robustness 190 3.6 Problems and Complements 197 3.7 Notes 210 3.8 References 211 TESTING AND CONFIDENCE REGIONS 213 4.1 Introduction 213 4.2.3 Computation 188 3.5.1 Computation 188 3.5.2 Interpretability 189 3.5.3 Robustness 190 3.6 Problems and Complements 197 3.7 Notes 210 3.8 References 211 TESTING AND CONFIDENCE REGIONS 213 4.1 Introduction 213 4.2.3 Computation 188 3.5.1 Computation 188 Choosing a Test Statistic: The Neyman-Pearson Lemma 223 4.3 Uniformly Most Powerful Tests and Monotone Likelihood Ratio 4.4 Models 227 Confidence Bounds. If x is a vector in RP, Il(x I L) is the point of L at which the perpendicular to L from x meets L. Then under some conditions, O l ' I where r11 tends to zero at a rate faster than 1/n and H2. Xn denote the incomes of a sample of replacement. This is a curved exponential 2, z, z, p) + 2::: ..0z, 1 + 0, j = 1, 2. Quang, and A Samulon., Xn be a sample from a population with density f (t 8) where () and f are unknown, but f (t) = f(-t) for all t, and f is continuous and positive. (iii) p, = 17. Here are two points to note: (1) A parameter can have many representations. 1-'2 • cr (1, 1), dis2 tribution. Next use Bayes rule. R. The example illustrates both the difficulty of speci fying a stochastic model and translating the question one wants to answer into a statistical 0 hypothesis. Of course, with ever faster computers a difference at this level is irrelevant. AND Z. If a and b are integers, then /3 (a,
b), W, are in is the distribution of (aV fbW)[1 + (aVfbW)]-1, where V1, 1.1 Data and Models Most studies and experiments, scientific or industrial, large scale or small, produce data whose analysis is the ultimate object of the endeavor. There is more material on Bayesian models and analysis. The frame we shall fit them into is the ultimate object of the endeavor. the following. GNEDEN KO, B. However, two-sided tests seem incomplete in the sense that if H B = 80 is rejected in favor of H: () -1- Bo, we usually want to know whether H: () > Bo or H: B < Bo. For instance, suppose B is the expected difference in blood pressure when two treat ments, A and B, are given to high blood pressure patients. (5) The study of the interplay between numerical and statistical considerations. and the risk of all possi simple example in which ble decision procedures can be computed and plotted. Hint: (b) Because (X1, Y1) has a density you may assume that 'if/ > 0, (f > 0, IPl < 1. When So contains more than one point, 80 and H are called composite. (a) lilt II = 0 iii It = 0 ((cb) llah1hll = lall I is such that !]hm - hn II ---> 0 as there exists h E 1i such that ll hn - h ll 🕏 0. :::::> n - 1 , , , I • • ' 1 , j ' ' '. SIDAK, Theory of Rank Tests New York: Academic Press, 1 967. The coefficients of skewness and kurtosis (see (A. This can be thought of as a problem of comparing the efficacy of two methods applied to the members of a certain population. New York: Chelsea, 1967. A. If h : Rd 🅏 for izl < z) z Pr(Z :: 0: z) z . 50. .45 .40 .35 .30 .25 .20 .1 5 .10 0 . is point mass at ' � 3" n 0 What the second of these examples suggests is often the case. The hypothesis of dominant inheritance corresponds to H : p = � with the alternative K : p f. the integrals j T(x) tep(x, e)] dx and j T(x) [:ep(x,B) Chapter 3 dx are continuous functions(3) of(). (b) lf01 = 02 = 0 where U • N(O, 1) with probability U with the same distribution. The argument given in B.6 establishes that (B.10.9) 0 and the result follows. For the normal measurement problem we have just discussed the probability of coverage is independent of P and equals the confidence coefficient. (L52) We shall give the proof in the discrete case. We then have h1 1 hz. ; ll 'l I I I J 'l I i T(x), the denominator decreasing. A'(O) = Cov(,P(X, - 0), f(X,)). The answer will depend on the model and, most significantly, on what criteria of performance we use. (4) 6.6 6.3 (4) 6.5 Volume II is expected to be forthcoming in 2003. (b) Using Bayes rule find the conditional distribution of X given Y = y. Show that for the Gaussian linear model with known variance D(Y, f.'o) Jy - P.ol2/a, l. What value of c gives sizec>? 1.3., Yn) • Y, h(y) = A(17) = I; • 1 (0, 0, k, L7-I) = I; • 1 (1, 0, 1) = I; • 1 (1, 0, 1distribution. 7, Show rigorously using (1.2.8) that if in Example 1.1.1, D = NO has a B(N - n, ?To) distribution. (8.10.16) is Pythagoras's theorem again. (a) Show that the only transformations h that make E[h(X) - E(h(X))]' up to order 1/n2 for all , > 0 are of the form h(t) = ct2i3 + d. 449) or Hoeffding (1963)., Xn is a sample from a population with d.f. F(x - J1.) where J1. For instance, suppose i s monotone increasing t distribution, or Table III, to find a = 0.05. Royal Statist. Bickel Berkeley ! Mathematical Statistics Basic Ideas and Selected Topics Volume I Second Edition ; 'j' 1 j 1 '' I ' i I i \ Chapter 1 STATISTICAL MODELS, GOALS, AND PERFORMANCE CRITERIA 1.1 DATA, MODELS, PARA METERS AND STATISTICS 1. Hint: il ' p . The results are assembled in what is called a 2 x 2 contingency table such as the one shown. The goals of science and society, which statisticians share, are to draw useful infor mation from data using everything that we know., Xn be i.i.d. as X \clubsuit F and let I' = E(X), jJ(Xt - 0) and T2(0) Varp¢(X 1 - 0). L., .. L., R. Consider testing H : 81 = 82 = 0 versus K : IJ, > 0 or IJ2 > 0. An unknown number of a sample of patients receiving an experimental treatment. If n is the total number of applicants, it might be tempting to model (Nmt, Nmo, Nft, Njo) by a multinomial, M(n,Pmt,Pmo,Pjt,PJO), distribution., Xn) = I 2a2 (x, - JL)2 + n (x, - f3x_, - (I - /3)1') 2 L i = , • 2 We include this example to illustrate that we need not be limited by independence. Then, q(x 1 + x2,0) = r(x, 0) + r(x, 0), and hence, [r(x • 1 + x2,0) + r(x, 0) +0] = r(x1 + x2, 0) - r(O, 0). Suppose that we know from past experience that a fixed proportion Bo = 0.3 recover from the disease with the old drug. Po) distribution and (x, Y;), 1 < i < n, be i.i.d. Let 81, 82 > 0 and H be as above. Example 4.7.2. known. Minimum Distance Estimates Permutation of Irregular Parameters Stein and Empirical Bayes EstimationModel Selection TOOLS FOR ASYMPTOTIC ANALYSISWeak Convergence in Function Spaces The Delta Method in Infinite Dimensional SpaceFurther Expansions DISTRIBUTION-FREE, UNBIASED, AND EQUIVARIANT PROCEDURES Introduction Similarity and Completeness Invariance, and Minimax Procedures INFERENCE IN SEMIPARAMETRIC MODELS Estimation in equality can be extended to the multiparameter case. j • This is easily seen to be an ordering. Limited, Prentice-Hall of Canada Inc., London Sydney Toronto Prentice Hall Hispanoamericana, S.A., Mexico Prentice-Hall of India Private Limited, New Delhi Prentice-Hall of Japan, Inc., Tokyo Pearson Education Asia Pte. STAHEL, Robust Statistics: The Approach Based on Influence Functions New York: J. ET AL., The Foundation of Statistical Inference London: SNEDECOR, G. Show that, in the logistic regression model, if the design matrix has rank p, then {30 as defined by (6.4.15) is consistent. 2.4.4 The EM (Expectation/Maximization) Algorithm There are many models that ep(X,) = O. (a) X is an unbiased estimate of $X = \mathcal{O} L \mathcal{O} 1 Xi + 1$. Froblems Section 3.6 and 205 Complements (b) The variance of X is given by 15. Show that the conditional distribution of (6, Xn + 1, . In each of the preceding examples of a Poisson process N(t) represents the number of times an ..event" (radioactive disintegration, arrival of a customer) has occurred in the time from 0 to t. Suppose further that . B y (A.1 4.20) we conclude that . In Example with X and Y distributed as N(l' + *, k," in Example 1.1.1 lead to (Problem 1.3.18). For a review of the ele mentary properties of matrices needed in its formulation, we refer the reader to Section BJO. , Xn) is a sample from a population with density f(x, B) * 9 lOa (X - p) 0 for some nonzero y. By the Neyman-Pearson lemma this is the largest power available with a level et test. The Graduate Division of the University of California at Berkeley attempted to study the possibility that sex bias operated in graduate admissions in 1973 by examining admissions data. .'. i' 1 ' i • PjY = y, Z = z] ($\mathbf{\Phi}$) p'(l - p)"-' ($\mathbf{\Phi}$) P'(l - p)"-' I ($\mathbf{\Phi}$) - ,-c $\mathbf{\Phi}$; . For a fuller treatment of these introductory aspects of Hilbert space theory, see Halmos (1951), Royden (1968), Rudin (1991), or more extensive works on functional analysis such as Dunford and Schwartz (1964). Nonparametric Inference for Functional analysis such as Dunford and Schwartz (1964). X(2) : S · · · : S X(n denotes) 2 the ordered X,, . Using the cen tral limit theorem, Slutsky's theorem, and the foregoing arguments, we find (Problem 5.3.28) that if n t fn A, 0 < A < 1, then - 8" ", N (o (1 - A)u) · 314 Asymptotic Approximations It follows that if 111 = n2 or af a , then the critical value t, 1-2(1 approximately correct if H is true and the X's are not normal. 1) where V denotes the gradient, Arguing heuristically again we are led to estimates B that solve $v \otimes p(X, e) = o$. (b) P(X < 2 I y = 1). We find • $c(B) = \bigoplus (-Jt-2, Jt-3)$ T, and from Example 1.6.5 IT 2 2 · A (1/) = ;zn (-ryl/12, '1, /2 '12 - 1/'12) . = 2: 2) are 1 I ' 5. The LR statistic for H : IL with w 1 : wo is wo } E wo versus K : IL E w 1 - wo D(Y, Ji 0) - D(Y, Ji 1) where j1,1 is the MLE under w 1. Volume II covers a number of topics that are important in current measure theory and practice. As in Example 2.2.6, letXi, i = 1, . (I) For each t E (0, oo), the powerfunction (3(1:1) = Eo 61 (X) is increasing in 0. For 8 E 80, the likelihood equations (6.4.3) become + n12) rh (nu + n2!) Tf2 (nu whose solutions are Tf1 'T/2 = (n21 + n22) (1 - ifl) (n12 + n22) (1 - ifl) (1P., "Robust Statistics: JAECKEL, L. we start by finding the density of c 1, . 5 415 General ized Linear Models Formally if w0 is a GLM of dimension p and w1 of dimension p and the values I, 3, 6, and 9, as indicated in the plot. Figure 2.4.1 illustrates the process. If Ni is the number of offspring of type i among a total of n offspring, then (Nb . information, in the context of a model P = For instance, suppose that in Example 1.1.1 we had sampled the manufactured items in order, recording at each stage whether the examined item was defective or not. DOI link for Mathematical Statistics Mathematical Statistics book Most of the problems are assigned in the required textbook: Bickel, Peter J., and Kjell A. The fundamental heuristic is typically the following. Then we say 8 solving v w(X,11) v (X,11) to an estimating equation estimate. I Section 1.2 13 Bayesian Models Thus, the resulting statistical inference becomes subjective. Establish (5.3.11). Although estimation loss functions are typically symmetric loss functions are typically in situation (c). (a) Let (Nu' Nl2· N21 ' N22) rv M (n, Bu (}12, (}2 1 ' 022) as in the contingency table. Testing Independence of Classifications in Contingency Tables Many important characteristics have only two categories. - J' !! (1) Narrow classes of procedures have been proposed using criteria such as con siderations of symmetry, unbiasedness (for estimates and tests), or level of sig nificance (for tests). "'' error model (3.5.2), show that (a) If h is a density that is symmetric about zero, then J-.t is identifiable. Life
testing. The argument is left to the problems as are some numerical applications. , Xn, where " is known. k when the prior distribution is beta, i Let X1, . Suppose X1 and X2 are independent exponential £(A) random variables. , Xn be i.i.d. where X = (U, V, W), P[U = a, V = b, W = c] 9. We next apply the duality theorem to MLR families: Theorem 4.5.1. Suppose X I have read of the variables t and B when the other is fixed. approximately has a x;- q distribution for large n, we define Bj = 9j (8), j = 1, 105). (B.I.8) = pz z pz z Thus, when Pz (z) > 0, the conditional expected value of A student or a future value of A student or a at least constant risk on the "most difficult" B., $T(X(B) \ Co, T(X) \ T((B+1) \ (t-a))$ where $T(t) < \cdots < T(B+1)$ are the ordered T(X), T(X(1)), . Let 0 < 80 < 8, < 1. By varying the assumptions we obtain parametric if we drop (I) and simply .treat the Zi as a label of the completely unknown distributions of Yf. Iden tifiability of these parametrizations and the status of their components as parameters are D discussed in the problems. THOMAS, Elements of Information Theory New York: Wiley, 1991. = I, 2, the I .4) Bayes' Rule whenever the denominator of the right-hand side is positive. The (f11 G) parametrization of Example 1.1.2 is now well defined and identifiable by (1. (b) Show that XHL is translation equivariant and antisymmetric. Often the final simplification is made. Derive the formula (6. 3.5.2 Interpretability Suppose that in the normal N(p, .')-11 2 = z,I/2 We can now use the MLE f;1/2, which as we shall see later (Section 5.4) is for n large a more precise estimate than Xf& if this model is correct. Deduce from (a) that E(X, I Sn) = · · · = E(Xn I Sn) and, hence, that E(Sm I Sn) = (m/n)Sn. 526 Additional Topics in Probability and Analysis N. = = Loss function. An important example is the class of procedures that depend only on knowledge of a sufficient statistic (see Ferguson 1 967; Section 3.4). N(TJ, r2), respectively., ((n - 1)t/n, t]. The x test is equivalent to rejecting (two-sidedly) if, versus K : P(A I and only if, Next we consider contingency tables for two nonnumerical characteristics having a and b states, respectively, a , b � 2 (e.g., eye color, hair color). - Yr · 2 + (n r)Yrl/B is x with 2r degrees of freedom. Bickel stat.berkeley.edu Kjell Doksum I • I I I I PREFACE TO THE FIRST EDITION This book presents our view of what an introduction to mathematical statistics for students with a good mathematics background should be. Hint: Use the central limit theorem for the critical value. As an example, suppose that X1, . $Dn = \sup IF(x) - Fo(x)l - x It can be shown (Problem 4.1.7)$ that Dn. which is called the Kolmogorov statistic, can be written as $Dn = \cdots$ max max >-l, . In Example 1.1.2 with assumptions (l)-(4) we have implicitly taken e = R X R + and, if ($l = (p_{r,1}, a_{2})$, Pe the distribution on R" with density x, jL) where cpis the standard normal density. n Var(X1 (A. We have seen that smooth transfor mations h(X) are also approximately normally distributed. (b) Assume that if X and Y are any two random variables, then the family of condi tional distributions of X given Y depends only on the joint distribution of (X, Y). In addition we feel Chapter 10 on decision theory is essential and cover at least the first two sections. Let F, F-(x), and F+ (x) be as in Examples 4.4.6 and 4.4.7. Then disstribution-free confidence region for Xp valid for all - P(F -(x) < F'+(x)) for all x E (a, b) = 1 - a. given l:: X, As Figure B.2.2 indicates, the beta family provides a wide variety of shapes that can approximate many reasonable prior distributions though by no means all. D Example 1.5.4. Let X1, 1 Xn be independent and identically distributed random vari ables each having a normal distribution from 8 to ')' = vn(8 - 80). S imilarly, we may want an interval for the Section 4 . 10. 3 .77 x (nitrogen) y (yield) 1 67.5 Hint: Do the regression for /-Li = (31 + f32zil + f- < Zn where z1, Zn are given constants., B4) distribution. Here J xdF(x) denotes J xp(x)dx in the continuous case and L: xp(x) in the discrete case. (i) The distribution of II: : 1. Set *** *** q(f") and v(!") ***** (a) Show that the likelihood equations are ...;.... Note that we really want to estimate the function Jl(•) our results will guide the selection can estimate (3 from our observations Y; of g((j, z;) of doses of drug for future patients. Let X • U(O, 8) be the uniform distribution on (0, 8). Suppose X1, . (b) Calculate the smallest n needed for by 0.02. Assume the linear regression model with p future observation Y to be taken at the pont z. F., P. < a: 15, the probability of your deciding to close is also < 0.01., Xn) = c(n + a, m). Finally, Chapter 6 is devoted to inference in multivariate (multi parameter) models. Next use the central limit theorem and Slutsky's theorem. 1 2.7) is called the cumulant generating function of X., Xn 12. I ' I Section B.Il 537 Problems and Complements Problems for Section B.S 1., Xn), where Xi is 1 if the ith patient recovers and 0 otherwise. (b) L is the space of all X + (Ez:I, Ezy) T(Z - E..'(Z)). Suppose (A4') holds as well as (A6) and 1(8) < oofor all 8. • I ., 7. 1.5) Yi B + ei, i where ei = cei-l + Ei, i = 1, . I, 4th ed See Problem 8.3.8., I I • References Hoe!, Port, and Stone (1971) Chapter 8, Section 8.1 Parzen (1960) Chapter 5, Section 3; Chapter 8, Section 4, Section 3; Chapter 8, Section a:) lower and upper prediction (c) IfF is continuous with a positive density f on (a, b), -oo < a < b < oo, give level (I - a) distribution free lower and upper prediction bounds for Xn + I · 3. + o(l) and that if own (X1, . . , k, we get radically different best tests depending on which (} is continuous with a positive density of California). Press, 197-206 (1 956). SAVAGE., L. L Doksum, Kjell A., Xn -.. In this edition we pursue our philosophy of describing the basic concepts of mathemat ical statistics relating theory to practice. Methods of Estimation. 1) n can be any integer > 0 whereas 8 may be any number in [0, 1]. Suppose X is aN(B, 1) sample and consider the decision rule J,,,(X) = -1 if X < r 0 if r < X < s 1 if X > s. If n is small compared to the size of the city, (A. ' • III'! i • I i I! • ' ' ' I I I I I I Section 4.1 X Because values of 219 Introduction X. (n, B)" for "the The symbol p as usual stands for a frequency If anywhere below p is not specified explicitly for some value of x it shall be assumed that p vanishes at that point. Similarly, we define $O(S) = \sup(O : j(O,a) = S - 1)$ where j(8, a) is given by, • Then 0(S) is a level (I - a) UCB for 0 and when S < n, 0(S) is the unique solution of $P(\Phi)$ s or (! - o)n - r = a. and sufficient statistic T Tj, j = 1, Bayesian Predictive Distributions Suppose that () is random with () "' 1r and that given () = e, X1, Xn + 1 are i.i.d. p(X I e). Let p : X X e 🖗 R where . It will often be more convenient to work with unrestricted parameters. are independently distributed as x2 with 2 degrees , Yr is an exponential family with density of freedom, and the joint density of Y1 , • . Let and where).* maximizes iJnew = iJ(>-•) t: fij (>-) - A (i)(>.)) . , Xn) is a sample drawn without replacement from an unknown finite population { x1, . If A(y) is a ..small" cube surrounding y and we let V(B) denote the volume of a set B, then V (g 1 (A (Y)'-')-) P[g(X] E A(y)] P[X E g- 1 (A(y)_)] . , Nk) � M(n, 8,, . Suppose u1, . Let Xi denote the difference between the treated and control responses for the ith pair. More generally, a function q : 8 ----+ N can be identified with a parameter v(P) iff Po, Po, implies q(B1) QA276.B47 2001 00-031377 519.5-dc21 Acquisition Editor: Kathleen Boothby Sestak Editor in Chief: Sally Yagan Assistant Vice President of Production and Manufacturing: David W. (d) We want to compare the efficacy of two ways of doing something under similar conditions such as brewing coffee, reducing pollution, treating a maze, and so on. But given a parametrization (} ----+ Po, (} is a parametrization is identifiable. Next we note how we can apply the information inequality to the problem of unbiased estimation. , (Zn, Yn), this leads to the commonly considered l(P, a)] = - n L($r(z_1)$), a (z,))2, n .= l J n-1 times the squared Euclidean distance between (a(zl), . = 1.3.1 Components of the Decision Theory Framework As in Section 1.1, we begin with a statistical model with an observation vector X whose distribution P ranges over a set P., Xn+k) given xl = X J, . where 0 < 8 < 1.3 and Z 1 +y1cr p (Y) - vny - has a N(O, 1) distribution and is independent of V (n - 1)s2 /cr2, which has a x;, 1 distribution. I I 1 3) I ' and ' (A. Wiley & Sons, 1986. = tt. To study the relation between the two characteristics we take a random sample of size n from the population. For each simulation the two characteristics we take a random sample of size n from the population. x-axis), af = a . Proof. 0.0 tn 2 (1 - cr)} do not have approximate level o:. The correlation inequality is equivalent to the statement (A.1.1.9) Equality holds if and only if X2 is linear function (X2 = a + bX1 , b # 0) of X1 · 1 j Section A.12 tion 459 Moment and Cumulant Generating Functions If X1 , . The result might be a density such as the one marked with a solid line in Figure B.2.2. If we were interested in some proportion about which we have no information or belief, we might take B to be uniformly distributed on (0, 1), where u is known. (c) Under the ons of(b) above, show that there is an increasing function Q • (t) such that if $Y_i = Q \cdot (Y_i)$, then f'i* = 9 ({31 Zi} + £i for some appropriate \in i. Then (• 1 X; - n and I(B) • Var • 1 X; • • o 0 • - o) n = -nO • B' o I 0 Section 3.4 181 Unbiased Estimation and Risk Inequality alities Here is the main result of this section. Math Statist, The poste rioi predictive distribution $Q(\cdot I x)$ of
Xn + I is defined as the conditional distribution of Xn + I given x = (x i . IPI) Deal directly with the cases 1. (Pareto density) (d) f(x, 8) = v'ex v'0-1, 0 < x < 1, 8 > 0. This is a rank 2 canonical exponential family generated by $T = (I; \log X_{,, I}; X_{,, I})$, h(x) = x-1, with 'I, 'I' by Problem 2.3.2(a). Evidently, the Bayes risk is now the same as in Example 3.3.3 with a2 = M. For a generalization of the notion of moment generating function to random vectors, see Section B.S. The function Kx(s) = log Mx (s) (A. Thus, if there are k different brands, there are k different brands, there are k different brands are consistent with the data than others. Hint: P(To > T) = J P(To 2: t I T = t)J.(t)dt where fe(t) is the density of Fe(t). t yV/k] is increasing 12. (See also Problem h. Deduce that X, Xu, X are unbiased estimates of the center of symmetry of a symmetry o two tables into one, and petform the same test on the resulting table. In this section we derive necessary and sufficient conditions for existence of MLEs in canonical exponential families of full rank with £ open (Theorem 2.3.1 and Corol lary 2.3.1). (d) Use (c) to show that if t, E (a, b), then B is not unique. Let $p(x | B) = exp\{-(x - 8)\}, 0 < B 0, B > 0 0$ otherwise. Show that if f(x) (e) Suppose that the d.f. F(x) of X, is continuous and that $f(0) \mid F'(x)$ exists, then $\cdot \mid \cdot = F'(0)$ exists. Show that under the assumptions of Theorem 4.3.2, 1 - 6t is UMP for testing H : 8 > Bo versus K : 8 < Bo. . When m m/(m - 2), and when m > 4, Var X = 2m2 (k + m - 2)k(m - 2)2(m - 4) > 2, E(X) = . However, in any case, critical values yielding correct type I probabilities are easily obtained by Monte Carlo methods., (a) Construct a test of H : B = 1 versus K : B > with approximate size o: using a complete sufficient statistic for this model. Hint: Argue by contradiction. This is part of a general phenomena we now describe. Similarly, the central limit theorem tells us that oo, , is as above and then if n • Ep]Xf] (z) (5.1.7) where 4:1 is the standard normal d. Expectations Po will be written Eo. Distribution functions by p(·, 0). In this example, we can achieve this by the reparametrization k. j = e0' / L e0', j = l = 1, . (1 + 1).2.8) 8 and X are both continuous or both discrete this is precisely Bayes' rule applied to the joint distribution of (II, X) given by (1.2.3). If the error distribution is normal, X is the best estimate in a variety of senses. Therefore, a2 has the smallest posterior risk and, if 0" is the Bayes rule, o'(O) = a, 13. Set U; = F(Xi), i = 1, . Uniformly most accurate (UMA) bounds turn out to have related nice properties., Tp) T. where Y; has density h(y I') and (X,', Y;) are i.i.d. Note that this implies the possibly unreasonable assumption that committing a gross error is independent of the value of x•. (a) Show that the Bayes rule is equivalence if E(>. 16 Statistical Models, Goals, and Performance Criteria Chapter 1 I THE DECISION THEORETIC FRAMEWORK 1.3 ' Given a statistical model, the information we want to draw from data can be put in various forms depending on the purposes of our analysis. ' It has of 90, 91, is much better defined than its complement and/or the distribution of statistics T under 80 is easy to compute. Then o(X,Oo) is a level a test for H: 0 oversus K: 0 > Oo. Because J''(X, Oo) or Pe, [O''(X) > Oo] < Pe, [O''only through q(0) and is a continuous increasing function of q(0). The correlation coefficient roughly measures the amount and sign of linear relationship between X1 and X2., (tn, Yn), n > 4, on a population of a large number of organisms. Tt> test the validity of the linkage model we would take 80 { G(2 + ry), i (1 ry), i ("one dimensional curve" of the three-dimensional parameter space e. DoKsuM, K. . We shall denote it by Fk ,m. Next, we introduce the t distribution with k degrees offreedom, which we shall denote by 7,. Therefore, by the basic property of maximum contrast estimates, for each B f> B0, and < > 0 there is o(B) > 0 such that Eo, inf{p(X, B') - p(X, B0)} B' E S(B, i (B))) > .} c T U S(B, J (B,)) j = 1 Now inf 1 n - n L {p(X. where x E R and (} E R. • 1 19. A final major topic in Volume II will be Monte Carlo. 5. Once the level or critical value is fixed, the probabilities of type II error as 8 ranges over 81 are determined. (I) We define set of the form A1 x ··· x Notes for Section A.7 ' f ' 1 (I) Strictly speaking, the density is only defined up to a set of Lebesgue measure 0. 5.3.1 The Delta Method for Moments • • We begin this section by deriving approximations to moments of scalar means and even provide crude bounds on the remainders. 8 .2) n+ 1. 12 30.82 7.7 1 0.025 0.01 0.025 0.0! 0.05 2 0.025 0.0] 0.05 2 0.025 0.0] 0.05 3 14.88 14.73 14.62 14.25 29.46 ISJO 28.71 28.24 27.91 27.67 27.49 27.23 26.87 6.94 6.59 6.39 6.26 6.16 6.09 6.04 5.96 5.86 12.22 10.65 9.98 9.60 9.36 9.20 9.07 8.98 8.84 8.66 21 .20 18.00 16.69 15.98 15.52 15.21 14.98 14.80 14.55 14.20 6.61 5.79 5.41 5.1 9 5.05 4.95 4.88 4.82 4.74 4.62 0.025 10.01 8.43 7.76 7.39 7.15 6.98 6.85 6.76 6.62 6.43 0.01 16.26 13.27 12.06 1 1.39 10.97 10.67 10.46 10.29 10.05 9.72 5.99 5.14 4.76 4.53 4.39 4.28 4.21 4.15 4.06 3.94 8.81 7.26 6.60 6.23 5.99 5.82 5.70 5.60 5.46 527 13.75 10.92 9.78 9. N1 c} is UMP. we can start by making the difference of the means, 6 = f1Y - f1X, the focus of the study. (c) Show that under independence the conditional distribution of Nii given R; Ci = Ci, i = 1, 2 is Jt(ci, n, ri) (the hypergeometric distribution). In designing a study to compare treatments A and B we need to determine sample sizes that will be large enough to enable us to detect differences that matter. In addition to the parameters of interest, there are also usually nuisance parameters, which correspond to other unknown features of the distribution of X., xi, 1) and (A8.2). P. Show that any T such that X(n) i < T < X(l) + i is a maximum likelihood estimate of 8. Then P is parametrized by B = (f11 \clubsuit). The set {0 : qo < q(O) < q1} is our indifference region. Thus, (5.3. 13) we see that a first approximation to the variance of h(X) is finding a function h such that for all J-l and a appropriate to our family. 3.7 ' • NOTES Note for Section 3.3 (1) A technical problem is to give the class S of subsets of :F for which we can assign probability (the measurable sets). Po, [S > j] a. 1.1.2 I i Parametrizations and Parameters ----t To describe Pwe use a parametrization, that is, a map, (} Po from a space of labels, the parameter space 8,toP; or equivalently write P = {Po : BE 8}. Hint: Note that /io V X if X < Jl.o and Jl.o otherwise. Thus, (} is more accurate. (b) Var(h(X, Y)) " V{[h, (Jl.t, 1'2)) 20'7 n 2 2 (n + +2h, (Jt" 1'2 lh2 (J1.,, J1.2)pu, I'2 lh2 (J1.,, J1.2)pu]} "2 + [h2 (J1." 1'2ll ,.n o -) where a a h1(x, y) = &,h(x,y), h2(x,y) = ayh(x,y) . Then Z has a binomial, B(n,p), distribution and p(y I z) * .. II. If we assume X1 , . The most important H1lbert space other than RP is L2(P) = {All random variables X on a (separable) probability space such that EX2 < oo}. known, and () = Note that, unconditionally, v(} has a xX distribution. Claims (5.4.42) and (5.4.43) follow. L., AND G. • Then again, a variance stabilizing transformation h is such that vn(h('/) - h('y)) -+ N(o, c) (5.3.19) for all f. For instance, suppose that the number Z of defectives in a lot of N produced by a manufacturing process has a B(N, 0) distribution. Similarly, v(X) is called a level (1 - a) upper confidence bound for v if for every P E P, P[v(X) = v] > I - o. Xlll ... We do not find this persuasive, but j if this view is accepted, it again reason by leads to a Neyman Pearson formulation. EIX1 - 3. So it is natural to consider 8(X) minimizing p(X, 8). LatMANN, E. j=2 The default assumption, at best an approximation for the wave example, is that f is the N(O, 0'2) density. 1. Hint: See part (a) of the proof of Lemma 5.3.1. 4. ji. (O and then). G. (8.10.14) Furthermore, (i) 11(h I C) exists and is uniquely defined. • I • O, we can conclude that X is minimax. By (A.16.4), and are independent and identically distributed exponential random variables with parameter is sufficient for 0., T.-t) T, where T; (X) = I: (X) = 1: size of the smaller of the two samples. D We return to this question in Section 4.3. Remark 4.1. From Figure 4.1.1 it appears that the power function 6.5 1. J., The Foundation of Statistical Inference London: Methuen & Co., 1962. Hint: Let U1 and U2 be as defined by (B.4.19) and B.4.22, then P[X < 0, Y < O] P[U1 < O, pU1 + y'1 - p2U2 < O] p - p u, < 0, u2, > r. = 7; fn, 1 < j : S k. There are situations in which most statisticians would agree that more can be said For instance, in the inspection Example size N 1.1.1, it is possible that, in the past, we havehad many shipments of that have subsequently been distributed. Note that this density coincides with the Student t density with one degree of freedom obtainable from (B.3.10). Problem 2.2.1). Again it is natural to translate this into P[Admit I Male] = Pml Pml + Pmo = P[Admit I Female] = p!I PJ1 + PfO But is this a correct translation of what absence of bias means? K than under H, it is natural to reject H for large is convenient to replace X by the test statistic T(X) = foX/a, which tends to be larger under It generates the same family of critical regions. Because 1 1 Section 4 . I. Problems such as these lie in the fields of sequential analysis and experimental design. We next give a prediction interval that is valid from samples from any population with a continuous distribution. (6) The study of the interplay between the number of observations and the number of observations with one parameters fixed. 30. Kx (s)l, \mathbf{O} o ds] is called the jth cumulant of X, j > 1.) test exists and is given by: Reject if ' i • I', ' (4.2.7) where c = z(l - a)[\mathbf{O} E01 (Problem 4.2.8). STUART, The Advanced Theory of Statistics, Vols. 9 L i kelihood
Ratio P roced u res 255 5 5 Example 4.8.3. Consider Example 3 .: \mathbf{O} (' I ' tum determines what we call H and what we call K. Then $h_t(t) = f0(t)/S0(t)$ and hy(t) = g(t)Sy(t) are called the hazard rates of To and Y. A., "Measures of Location and Asymmetry," Scand. We conclude from Theorem 2.3.2 that (2.3.4) and (2.3.5) have a unique solution with probability 1., Nk)? Kolmogorov's Theorem. If the family {Pf] } satisfies I and II and if there exists an unbiased estimate T* of 1/(B) such that Varo[T'(X)] = [c'(8)] 2/I(B) for all B E 8, then T' is UMVU as an estimate of 1}; OT and Eo = I, then this test rule is to reject H if X1 is large; however, if E0 = j: F(x). If the theory is false, it's not clear what P should be as in the preceding Mendel example. Xn are i.i.d. N(p, a2) with J-l/a Ao > 0 known. EY2 < oo (Y - p) + (I' - c) 0 makes the cross product term vanish. Show directly using Problem B.2.5 that under the conditions of the previous problem, if m/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero at the rate 1/(m + n) =, Var Bm 'n = + Rm 'n n + m m+n where Rm, n tends to zero rule (B.1.4) can be used. = (a) Find the joint density of Y1 and Y2. Specify the distribution of fi. (b) Show that the power function of your test is increasing in 8. We need only that 0(P) is a parameter as defined in Section 1.1. As we saw in Section 2.1, parameters and their estimates can often be extended to larger classes of distributions than they originally were defined for. n Hint: nFx (x) = 2; 1 [Fx (X,) < Fx (x)] n F x(x) = n L 1 [-X, < x] = i = 1 nF1 - u{F(x)} appear in basic probability texts but we need to draw on it for the rest of the book. 0 I 12 Statistical Models, Goals, and Performance Criteria Summary., Section 2.4 Algorithmic Issues 127 If m > 2, then the full 2n-parameter model satisfies the conditions of Theorem 2.3.1. Let £ be the canonical parameter set for this full model and let 8 * {II : II , E R, 82 E R, 83 > 0 }. Hint: \langle ; • Lx(T] = T; (X) - E'IT;(X). In that case, e who se likelihood is on or above some fixed value dependent on It is often approximately true (see Chapter C(x) is just the set of all the data. Thus, in Example 1 and 1.1.1 we take (} to be the fraction ofdefectives in the shipment, e = { 0, k Po the 'H.(NB, N, n) distribution. Hint: Reduce to a: 1 + a: 2 = a: by showing that if 0: 1 + 0: 2 < a:, there is a shorter interval with a: 1 + a: 2 = a:. For sim k 2. Statist., 42, I 020-- 1034 (1971). t! $1 \cdots U$, 1 + 0: 2 < a:, there is a shorter interval with a: 1 + a: 2 = a:. For sim k 2. Statist., 42, I 020-- 1034 (1971). t! $1 \cdots U$, 1 + 0: 2 < a:, there is a shorter interval with a: 1 + a: 2 = a:. For sim k 2. Statist., 42, I 020-- 1034 (1971). t! $1 \cdots U$, 1 + 0: 2 < a:. For sim k 2. Statist., 42, I 020-- 1034 (1971). rule when (i) 3. Let the time between the arrival of the first and second customers. This method requires computation of the inverse of the Hessian, which may counterbalance its advantage in speed of convergence when it does converge. 0 I, ' 522 Additional Topics in Probability and Analysis Appendix B It follows that if l ilt II' = (/L It). In the goodness-of-fit Example 4.1.5, suppose that F0 (x) has a nonzero density on some interval (a, b), -oo < a < b < oo, and consider the alternative with distribution function F(x, B) = Fi!(x), 0 < B < 1. Suppose that cri = u ? [:o logp(X, B]] and Chapter 3 !) L ao logf(X; , B) and Z have the joint frequency function given by the table For instance, suppose Z is the number of cigarettes that a person picked at random from a certain population smokes per day (to the nearest 10), and Y is a general health rating for the same person with 0 corresponding to good, 2 to poor, and 1 to neither. I, , 8.10.3.2 Projections on Linear Spaces We naturally define that a sequence hn E H converges to h iff Jl hn - hi I --- t 0.] t, (;;) + log When we use the logit transform g (1r), we obtain what is called the where. It is referred to as the level " critical value. A3, A6 11 • !(II) is continuous. (b) Suppose Z1 and z; have a N(O, 1 and c(a) [2:;" r"]- Show that 1 1T (). L 1 Xf is an optimal test statistic for testing H : 1/. < 1/. \o versus (b) Show that the critical region [L - 1 Xf > k] is k = X2 n (1 -) /2>.0 where X2n (1 -) is the (1) th quantile of the X n distribution and that the power function of the UMP level a test is given by - 1 - G,n(>.x,n(1 -)/ >.0) where G2 n denotes the XI addenotes the XI addenotes the Level a test is given by - 1 - G,n(>.x,n(1 -)/ >.0) where G2 n denotes the XI addenotes the XI adde from the population with density l = f(x I t). (ii) Show that if we model the distribution of Y as C(min{Xt, . (3.4. 14) Cov Now let us apply the correlation (Cauchy-Schwarz) inequality (A. An interesting discussion of a variety of points of view on these questions may be found in Savage et a!. However, when it is easy to see that E(Y) = A (17) Var(Y) = c(7) A(17) (6.5, 13) (6.5, 14) so that the variance can be written as the product of a function of the longer tail of the statistic T. When F is not symmetric, /1 may be very much pulled in the direction of the longer tail of the density, and for this reason, the median is preferred in this case. (1) Our own point of view is that subjective elements including. If J{x, vo} is a level a test of H : v = v0, then the set S(x) of v0 where 248 Testing and Confidence Regions o ("" v0) * 0 is a level (1 - o) confidence region for fidence region for v. • = . of oY, "" rr "" F;(x) • F; If Y generates Fs C:) . In this case (B.2.4) g is strictly reduces to the familiar formula (A.8.9). Examples of such measurements are hours of sleep when receiving a drug and when receiving a drug and when receiving a bacebo, sales performance before and after a course in salesmanship, mileage of cars

with and without a certain ingredient or adjustment, and so on. Let Z be the number of red balls obtained in the first two draws and Y the total number of red balls drawn. 275-277, 3 1 4), for instance. Consider the problem of testing H: 0 = 00 versus K: 0 = 0, with Oo < 8, ... • i, . Similarly, in Example 1.1.3, instead of postulating a constant treatment effect . Hilbert space theory is not needed, but for those who know this topic Appendix B points out interesting connections to prediction and linear regression analysis. 1 5) (A. Figures 8.2.1 and 8.2.2 show some typical members of the two families. The power function of the test with critical value c is (4.1.2) because (z) = 1 - (-z). :1;1 For example, if N (t) is the number of disinte grations of a fixed amount of some radioactive substance in the period from time 0 to time t, then {N(t)} is a Poisson process. : JI I • E [0, LJ.] versus (a) Show that the test that rejects H for large values of y'n(X - LJ.) has p-value p = 2 treat ments on a population and that we administer only one treatment to each subject and a sample of nk subjects get treatment k, 1 < k < p, $n1 + \cdots + np = n$. Others do not and some though theoretically attractive cannot be implemented in a human lifetime. (1962). Because this value oft corresponds to $n = \clubsuit$, the intuitive test, which decides J = v if and 0 only if $T > \clubsuit$ [Eo (T) + Ev (T)], is indeed minimax. '0 3. I (a) We are faced with a population of N elements, for instance, a shipment of manufac tured items. This is in fact true for sections by any plane perpendicular to the y) plane. RA.IFFA, H., AND R. Let X1, X2,X3 be independent observations from the Cauchy distribution about (}, f(x, 9) = ,.- (1 + (x - 9)2) - 1 • Suppose X, = 0, X, = 1, X3 = a. In fact, it can be shown using the methods of 254 Testi n g a n d Confidence Re g ions Cha pter 4 Chapter 5 that the width of the confidence interval (4.4.1) tends to zero in probability at the aa). 357 ..-. The same conventions apply to S1 and K. Show that I and •• 3. Here are two examples that illustrate these issues. k(B,a) k(Bo, a) if B i Bo. k(Oo, a) (1 - a) Oo B k(8, a). Examples are the distribution of income and the distribution of wealth. Begin by noting that according to e. = z is discrete, continuous, (b) Give the conditional frequency, density, or distribution function in each case. Ll) Now a(c) is nonincreasing in c and typically (c) T 1 as c l -oo and a (c) l 0 as c T oo. Statistical Models, Goals, and 18 Performance Criteria Chapter 1 Prediction. Suppose that f(x, B) is a positive density on the real line, which is continuous in x for each 0 and such that if (X1, X2) is a sample of size 2 from f (·, 0), then X1 + X2 is sufficient for B. X11 are independent random variables with moment generating function given by ll x, ll x, then X1 + . I 0 is the density of i.i.d. x and let .,-(8) = 2 exp{-2B}, B > 0. (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972)., Ynf Show that) form a sample from aN(O, 0, o-i, a-), population. Definition 4.3.2. The family of i.i.d. x and let .,-(8) = 2 exp{-2B}, B > 0. (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972)., Ynf Show that) form a sample from aN(O, 0, o-i, a-), population. Definition 4.3.2. The family of i.i.d. x and let .,-(8) = 2 exp{-2B}, B > 0. (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972)., Ynf Show that) form a sample from aN(O, 0, o-i, a-), population. Definition 4.3.2. The family of i.i.d. x and let .,-(8) = 2 exp{-2B}, B > 0. (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972)., Ynf Show that) form a sample from aN(O, 0, o-i, a-), a-) (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972). (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972). (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972). (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = probability By (A.9.5), 1- xi (X1, . Stal! sr., 43, 1041-1067 (1972). (1.4.2) Let p(z) Because g(z) is a constant, Lemma E[(Y - g(z)) 2 I Z = z] = E(Y I Z = z). Then X = p models { P, : 0 E 8} with 8 c R is said to be a monotone likelihood ratio (MLR) family if for fit < 02 the distributions Po1 and Po2 are distributions Po1 and Po2 are distributions from Example 4.2. 1 that if f-1 1 > f-lo, the MP level a test z (1 - 2 is an increasing function of T(x). f-lo · To see this, note that it follows from Example 4.2. 1 that if f-1 1 > f-lo, the MP level a test z (1 - 2 is an increasing function of T(x). f-lo · To see this, note that it follows from Example 4.2. 1 that if f-1 1 > f-lo, the MP level a test z (1 - 2 is an increasing function of T(x). we want to study the effect of a treatment on a population of patients whose responses are quite variable because the patients differ with respect to age, diet, and 5.3.15 and 5.3.16. Generalized linear models are introduced as examples. Implicit in this description is the assumption that () is a parameter in the sense we have just defined. It then offers a detailed treatment of maximum likelihood estimates (MLEs) and confidence regions, including optimality theory for estimation and elementary robustness considerations. That is, in the 216 Testing and Confidence Regions, including optimality theory for estimation and elementary robustness considerations. is rule is $ok(X) = \{S > k\}$ with P1 Pu $\{X : S > k\}$ = probability of type I error = probability of type II error = Probab close to the root 'ij of A(ij) = t0 • then by expanding A(ij) around 17oid • we obtain f1new is the solution for ij to the approximation equation given by I(B) and given by I(B) and given by I(B) & E9 (! logp(X, B)) 2 Note that 0 < I(B) < oo. is 3.4.1. (T X)) > information Cramir-Rao lower bound Var9 (I(B) 1 (3.4.16) I(B) The number 1/ is often referred to as the for the variance of an unbiased estimate of 1/J(B). There are absolute bounds on most quantities-100 ft high men are impossible. C. J. , (X., Yn)) be a sample from a bivariate normal population. distribution, D(a:), a = (o:1, , ar)r, $O'_j > 0, 1::; j < r$, has density • Let $N = (N1, \bullet \bullet . !$ Section 3.2 165 Bayes Procedures Let 1r(O) be a prior distribution assigning mass 11' to Oi, so that 1fr > 0, i = 0, . It follows from the central limit theorem that Tn (2::7 1 Xf - n) /ffn \clubsuit (V - n) ffn \clubsuit (V - n) ffn \clubsuit (V - n) present in 2 000, covers material we now view as important for all beginning graduate students in statistics and science and engin eering graduate students whose research will involve statistics intrinsically rather than as an aid in drawing conclu• SIOnS. BICKEL, P. Let X1, I < a S A, I < b < B, 1 < c < C and La ,b,,Pabo = 1. n x /[1 Ox(l - 0)" (1 - 0)] (1 - 0) X = 1, . Example 2.4.2. The Two-Parameter Gamma Family (continued). trials with the first trial being a success. If Y and Z are discrete random vectors possibly of different dimensions, we want to study the conditional probability structure of Y given that Z has taken on a particular value z. G, where the model {(F, G)} is where 1/J is an unknown strictly increasing differentiable map from R to R, 1/J > 0, $1/J(\pm 00) = \pm 00$, and Z1 and Z{ are independent random variables . " 1 139-1158 (1976). For the model defined by (3.2.16) and (3.2.17), find (a) the linear Bayes estimate of 1 (b) the linear Bayes estimate of 1 (b) the linear Bayes estimate of 1 (b) the linear Bayes estimate of 1 (c) the linear Bayes estimate of 1 (b) the linear Bayes estimate of 1 (c) the linear Bayes estim E(g(n)IrpaJ) < E(g(Z)), (A.15.6) 0 which is equivalent to (A.15.4). i=O 13. = a5, t > 0. The foundation of oUr statistical knowledge was obtained in the lucid, enthusiastic, and stimulating lectures of Joe Hodges and Chuck Bell, respectively. (a) Show that the MLEs of a'f, ag, and p when f1, 1 and J.t2 are assumed to be known are 'ifi = (1/n) L: (3/2) 1 (X; P.tf. Define the conditional frequency function $p(\cdot I z)$ of Y given Z = z by $p(\cdot) y$, z p(y I z) = P[Y = y I Z = z] = p(z) z (B. A P diag(>-1, Establish DYi P(i = 1 where Y; P(i = 14.2.1. You want to buy one of two systems. approximately has a x r distribution under H. Then X is the MLE of ry is h (8) (i.e., MLEs are unaffected by reparametrization, they are equivariant under one-to-one transformations)., r. (and show that (3.6.1) is equivalent to "Accept bioequivalence if [E(O I B) I X = x) < 0" (3.6.1)]2 < (Tif (n) + c'){log(rg(v;+,') + x + v}" where Hint: (b) large See Example 3.2.1. It is proposed that the preceding prior is "uninformative" if it has ("76 ---+ TJo oo"). 'STEIN, C., "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Distribution," Proc. However a is not identifiable because $q_0(x, a + cl)$ of $q_0(x, a)$ for $1 = (1, . STIRZAKER, Probability and Random Processes Oxford: Clarendon HAJEK, J. 6) that c (e) is independent of e. Suppose X <math>(p_1, E_1)$, j = 0, 1. Note that >.(0) should be negative when () (e, E, t) and positive when () (e, E, t). 2k is even give and plot the sensitivity curve of the median. in the measurement model, we implicitly discussed two estimates or decision rules: 61 (x) =
sample median. (b) Show that (X, Y) and (8, SJ, 812) are independent. It emphasizes nonparametric methods which can really only be implemented with modern computing power on large and complex data sets. 2 coincides with the likelihood ratio statistic A (Y) for the o-2 known case with o-2 replaced by 0:2. It follows that Fo (t) < 1 - a iff B > lia(t) and S(t) = [lia, oo). Topics to be covered include per mutation and rank tests and their basis in completeness and equivariance. The extra factor in the continuous case appears roughly as follows. Suppose that given () ϕ has a beta, {J(r, s), distribution.} is a conjugate family of prior distributions for p(x I 8) and that the posterior distribution of() given X = x is 1r (8 I x) = N' IT vW I 8) i=1 I ! 74 Statistical Models, Goals, and Performance Criteria where N' = N + n and ({I, . We illustrate these ideas using the example, Xn are i.i.d. N(f-L, cr2) with a 2 known. > * ' 332 Asymptotic Approximations Chapter 5 Proof Claims (5.4.33) and (5.4.34) follow directly by Theorem 5.4.2. By (5.4.30) and (5.4.35). 1.4.1 assures us that $E[(Y - p(z)) 2 I Z = z] + [g(z) - p(z)] 2 \cdot (1.4.3)$ If we now take expectations of both sides and employ the double expectation theorem (B.1.20), we can conclude that Theorem 1.4.1. /f Z is any random variable, then either E(Y g(Z)) 2 = oofor every function g or E(Y - p(Z)) 2 < E(Y - g(Z)) 2 < E(Y and Performance Criteria for every g with strict inequality holding unless g(Z) best MSPE predictor., 100. Often a treatment that is beneficial in small doses is harmful in large doses. (iii) are not independent. az, the intervals (4.9.3) have correct asymptotic probability of coverage. In Example 1.5.4, express t1 as a function of Lx(O, I) and Lx(l, 1). An example is discussed in Section 4.9.2. e (el > e2) where e1 is the parameter of e 2 is a nuisance parameter. , Xn are normally distributed with mean p. The Wald statistic is only asymptotically invariant under reparameter. , Xn are normally distributed with mean p. The Wald statistic is only asymptotically invariant under reparameter. be beat for 8 = 80 but can obviously be arbitrarily terrible. Our multinomial assumption now becomes N M(pmid. Here is an example of affine transformations of 6 and T. (b) Assume (1.7.2) and that Fo(t) = P(T:S t) is known and strictly increasing. where Fu and F1-u are the empirical distributions of U and 1 - U with & Suppose that T1 (X) is sufficient for 81 whenever 82 is fixed and known. p [- - - -] I • • ' I ' ' I n 3 3 Fisher conjectured that rather than believing that such a very extraordinary event oc curred it is more likely that the numbers were made to "agree with theory" by an overzeal ous assistant. Science 5, 160-168 (1990). # 0. they depend only on the nioments about the mean. Consider the parametric version of the regression model of Example 1.1.4 with l'(z) = g({3, z}), {3 E Rd, where the function g is known. We make the blanket assumption that all sets and functions considered are measurable. 400 Inference in the Multiparameter Case Chapter 6 where D is the d x d matrix of second partials of 11 with respect to e., (c) Find the minimax rule among the randomized rules. (c) Conclude that if A and B are independent, 0 < P(A) < 1, 0 < P(B) < 1, then Z has a limiting N(01 l) distribution = E(Y I Z). One such class is the two-parameter beta family. Argue that given are, T has a Yn 2 distribution. Independent or in terms of the eij · eij = (Bil + ei2) (B11 + B21). However, these and other subscripts and arguments will be omitted where no confusion can arise. 'Xn are observable and Xn + l is to be predicted. On the other hand, P[[Nn (t) - N(t)] > 1] l)t j p [(N (\diamond - N (< \diamond)) > 1 = nP [N (\diamond) > 1] no (;) ---t 0 as n ---t 0. rejecting H if N22 N22 9. o One application of variance stabilizing transformations, by their definition, is to exhibit monotone functions of parameters of interest for which we can give fixed length (independent of the data) confidence intervals. j=l j=l This is a k-parameter canonical exponential family generated by T1, . For a function of Xi. Here an example of N is the class of all continuous distribution functions. (a) Find maximum likelihood estimates of J.t and a2 . j . Otherwise, the decision is wrong and the loss is taken to equal one . Intuitively, and as we shall see fonnally later, a reasonable prediction rule for an unseen Y (response of a new patient) is the function P(z), the expected value of Y given z. Unfortunately P(z) is unknown. Also F(z) n -1 $E1{X; < x} = n$ -1E1{Fo(X,) < Fo(x)} = U (Fo(x)) where U denotes the empirical distribution function of U1, In Example 1.1.2 in Example 1.1.2 in Example 1.1.1, the fraction defective in the sample 1.1.2 in Example 4.6.2. Boundsfor the Probability of Early Failure of L1.2 in Example 1.1.2 in Example 4.6.2. Boundsfor the Probability of Early Failure of L1.2 in Example 1.1.2 in Example 1.1.2 in Example 1.1.2 in Example 1.1.2 in Example 4.6.2. Boundsfor the Probability of Early Failure of Equipment. binomial distribution indicated by this theorem is rather good. Suppose X1 , . ' I' Yet show e and has finite variance. Actuarial J., 204-222 (1986). Robustness from an asymptotic theory point of view appears also. Show that this holds iff P[U 🏶 a , V = b I W = c] = PIU = a I W = c]P[V = b I W = c]. Press, 1989. (c) Let x E {x : p(x, OJ) > 0}, then to have equality in (4.2.4) we need to have 0} D and 0 0 Co. It follows from the Neyman-Pearson lemma that an MP test has power at least as large as its level; that is, Corollary 4.2.1. /f z(1 - a) (4.2.6) has probability of type I error o:. Which lower bound is more accurate? Solutions for Vol. Thus, by the definition of the (Student) t distribution in Section B Note for Section 1.7 (I) uT Mu > 0 for all $p \ge 1.91$ vectors u # 0. Typically, rather than this Lagrangian form, it is customary to first fix a in (1.3.8) and then see what one can do to control (say) R(P, v) = x. Further applications appear in (1.3.8) and then see what one can do to control (say) R(P, v) = x. Further applications appear in (1.3.8) and then see what one can do to control (say) R(P, v) = x. Further applications appear in (1.3.8) and then see what one can do to control (say) R(P, v) = x. Further applications appear in (1.3.8) and then see what one can do to control (say) R(P, v) = x. Further applications appear in (1.3.8) and then see what one can do to control (say) R(P, v) = x. the next D section., Xrnare independent of each other and Y1, Yn; moreover, X1, ... Then use py, (y1) = f py, y, (y2) dy. The most famous unbiased estimates of f.L and a-2 when X1, Xn are i.i.d. $N([l, a-2) \cdot Let Xh \cdot Justify fornally 3 E [h(X) - E(h(X))]' = !, [h'(I')] 311a + h''(I'Jih'(I')] 204 + O(n-3).$ (d) Suppose that () has prior 1r(BI) 0.5, 1r(B2) * 0.5. Find the Bayes rule for case * (a). E., Stochastic Processes San Francisco: Holden Day, 1962. Remark 2.3.1. In Example 2.2.8 we saw that in the multinomial case with the clQsed parameter set (.>.; :-'; > 0, 2:7*, -'; = 1}, n > k - 1, the MLEs ofA3 , j = 1 , . See Example 5.3.6. Also closely related but different are socalled normalizing transformations., Xn) be a sample from a Poisson P (.A) distribution and let Sm = m < n., Yn) T by least squares in a linear regression Y = A[3 + < and (8.10.16) is the population size and]l is the expected consumption of a randomly drawn individual. Then the map sending B = (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) into the distribution of (X1, , Xn) remains the same but 8 = { (!',G) i Moreover, hv(t) = C.ho(t) is called the Cox proportional hazard model. Conversely, if T is sufficient, let g(t; B) = PB[T = t], h(x) = P[X x] T(X) = t; (1.5.6) p(x, B) = Po[X = x, T = T(x)] = g(T(x), B)h(x)(1.5.7), = Then 0 by (B. Again informally we shall call such procedures robust. The assertion (ii) is a consequence of the following remarks. Show that $P(X(k) < Xp < X(n-1)) = 1 - . \bullet = E(U), u_{;;} - E(U, - M_{,})'(U2 - !-'2)i$, ul $\bullet - 0$. If a IQR. Thus, the number of patients in the study (the sample
size) is random. 21. However, two notions, the sense for fixed n. t 3) can be used for model construction. Examples of this process may be found in Chapter 5 For instance, in comparative experiments such as those of Example 1.1.3 the group of patients to whom drugs A and B are to be administered may be haphazard rather than a random sample from the population of sufferers from a disease. BERMAN, S. - i i I Section 2.3 Maximum likelihood in Multiparameter Exponential Families 125 On the other hand, if any TJ = 0 or n, 0 < j < k - 1 we can obtain a contradiction to (2.3.2) by taking c; = -1 (i = j), 1 < i < k - 1. We call T a test statistic. Moreover, recall that a decision procedure in the case of a test statistic. Moreover, recall that a decision procedure in the case of a test statistic. Theorem 6.2.2 to On . Hint: Consider T(X) = X in Theorem 3.4.1. = 14. We discuss some of the issues and the subtleties that arise in the context of some of our examples in estimation theory. Theorem B.2.2 provides one of the instances in which frequency and density functions q is one-to-one, and Y q(X), then $p_y(y) = p_y(q-1(y))$. (i 1, . • . 0, 1, . • . 0, 1) Suppose Z1 and Z2 are independent with exponential $\pounds(.A)$ distributions. Suppose that X \clubsuit (X1, ..., ', For instance $l(P,a) - - - \And \bigstar$ (P), a > v(P), for some constant c > 0. We have seen in the proof of Theorem 4.3.1 that 1 - Fo(t) is increasing in 0. If the distribution function Fa of T(X) under X,...., Pe0 is continuous and if (1 - a) is a solution of Fo (t) = 1 - a, then the test that rejects H if and only if T(r) > t(I ex) is /JMP level a for testing H : () < Oo versus K : () > Oo. - Example 4.3.4. Testing Precision. (i) Various notational conventions are used in the text., aq } let w13 > 0 be given constants, and let the loss incurred when 0; is true and action a3 is taken be given by = , . Because the resulting value oft is possible if 0 < tjo < n, 1 < j < k, and one of the two sums is nonempty because c i' 0, we see that (2.3.2) holds. Argue that the Fisher test is equivalent to + n - (r1 + c!), and that under H, is conditionally distributed 1t(r2, n, c2). 'I.' New York: J. • Section B.ll 525 Problems and Complements 7. - - 1 where X, . , n, and let 02 = L: 13. 7, whereas if there is no oil, formations 0 and 1 occur with frequencies 0.6 and 0.4. We list all possible decision rules in the following table. (T1 (X), T,(X)) is sufficient for 8, T1 (X) is su We run m + n independent experiments as follows: m + n members of the population are picked at random and m of these are assigned to the first method and the remaining n are assigned to the second method. J 'I , I, this is no longer the case (Problem 4.2.9). = Jl > 0, corresponding to 771 = --jfo, ry2 = -2!2 family with c1 (Jt) = ¥, c2 (Jt) = -2!2 family with c1 (Jt) = ¥, c2 (Jt) = -2!2 family with c1 (Jt) = ¥, c2 (Jt) = -2!2 family with c1 (Jt) = 4.2.9 Evidently $c(8) = \{(1)I, ry2\} : ry2 = -i | 1f \rangle 2$, $ry_2 < 0\}$, which is closed in $E = ((ry|>ry2) : ry1 \in R, ry2 < 0]$. The conventions established on footnotes and notation in the first edition remain, if somewhat augmented. • Theorem 1.S.I. Jn a regular model, a statistic T(X) with range T is sufficient/ore if, and only if, there exists afunction g(t, B)defined/or t in T and e in 8 and a function h defined on X such that for al/ x p(x, 0) = g(T(x), O)h(x) E X, 0 E 8. = Y. Note: You may use without proof (see Appendix B.9). A review of the mean, median, trimmed mean with a = Ij4, and the Hodges Lehmann estimate., b"g}. y. For a sample from -a P(B) distribution, the MLE is B Because X is unbiased and Var(X) = Bfn, then X is UMVU., Un = I{U E (m2 -•, (m + 1)T•)}, 7. exists, and by expanding Y- c • oo (1.4.1) for all c; see Problem 1.4.25. Let X 1, • . 653), compute {i1, {i, {i3 and level 0.95 2 confidence intervals for (31, !3, (33., Xn are i.i.d.. If = "\" c,w• = k=O I (1 - w)n, 0 < w < I, then Ck I d' n w) (I k! fiw k w=O Which of the following families? The advantage of piling on assumptions such as (1)-(4) of Exam ple 1.1.2 is that, if they are true, we know how to combine our measurements to estimate 1-L in a highly a support of the following families? efficient way and also assess the accuracy of our estimation procedure (Exam ple 4.4.1). A = {0, 1} with 1 corresponding to rejection of H. Hint: (c) x, -e?, x, -eg are independentN(01-B?, a-2), N(82 -eg, a-2), respectively., Xn)T where the X; are i.i.d. as X, then all models for X are exponential families because they are submodels of the multinomial trials model. 2:7 1 '!'(X, , 8) has a unique 0 in S(IJ0, 0 such that gn are 1-1 Chapter 6 on 8(90, 6) rithm starting at e;, converges to the unique root (6.2.3). 1 . (b) Expanding, (5.5.13) i 1 • ! ' I , Section 5.5 Asymptotic Behavior and Optimality of the Posterior Distribution where Bi, If <)n. (0, and Y = t In both of the foregoing examples considerable reduction has been achieved. I, R. Show your result to the U[81, that fxx(��) is minimial sufficient. BERGER, J. Four balls are drawn at random without replacement. defectives found in the sample. Give details of the proof or Corollary 2.3.1. 5. Then : (b) Suppose Xi, Yi are as above with the same hypothesis but $8 = \{ (81, fh) : 0 < e, < cl)1, 01 > 0 \}$. '(a) Compute the Rao test for H Problem 6.4.11. We next turn to the final topic of this section, general criteria for selecting "optimal" procedures. 02 and yet Pe1 = Pe2• Such parametrizations are called unidentifiable. We shall obtain likelihood ratio tests for hypotheses of the form H : e1 ew, which are composite because e 2 can vary freely. Here e = (!",a) with -oo 0. L, (x, - x) 0} docs not depend on B. (a) Show that for any 2 x 2 contingency table the table obtained by subtracting (esti mated) expectations from each entry has all rows and columns summing to zero, hence, is of the form Z2 where Z is given by (6.4.8) (c) Derive the alternative form (6.4.8) for Z. Examples of application such as the Cox model in survival analysis, other transformation models, and the classical nonparametric k sample and independence problems will be included. Decision procedures. Then, if N = (N1 1, N12, N21, N22) r, we have N M (n, Ou, 8 12, 82 1, 022). Duality Theorem. 1.3) 🕏 0 has 80 as its unique solution for the property that for any prior distribution of 9 depends on x only through T(x). (c) Give an approximate expression for the expression for the posterior distribution of 9 depends on x only through T(x). critical value if n is large and 8 not too close to 0 or oo., n, (3.5.1). X1, AND J. Suppose that X1, • . 11) with scale parameter 5. The power is plotted as a function of 8, k * 6 and the size is 0.0473. Note that because l can be any of the integers 1, . - * - , The interplay between estimated variance and computation As we have seen in special cases in Examples 3.4.3 and 3.4.4, estimates of parameters based on samples of size n have standard deviations of order n - 1 12. These will not be dealt with in our work., Xn be independently distributed with Xi having a N(ei, 1) distribution, 1 < i < n. If A = 1, the joint density of X1 and X2 is (B.2.I2) for x1 > 0, X2 > 0.11 3.00 2.91 2.85 2.75 2.62 0.025 6.55 5.10 4.47 4.12 3.89 3.73 3.61 3.51 3.37 3.18 0.01 9.33 6.93 5.95 5.41 5.06 4.82 4.64 4.50 4.30 4.01 4.54 3.68 3.29 3.06 2.80 0.01 8.68 6.36 5.42 4.89 4.56 432 4.14 4.00 3.80 3.52 0.025 0.01 0.05 0 fl5 0.05 4 5 6 0.025 0.01 0.05 9 fl5 0.05 4 5 6 0.025 0.01 0.0 0.025 0.01 0.05 10 0.025 O.Di 0.05 10 0.025 O.Di 0.05 12 15 r1 = numerator degrees of freedom. j "->. JEFFREYS, H., Theory of Probability, 2nd ed. We begin this discussion with decision theory in Section 1.3 and continue with optimality principles in Chapters 3 and 4. For instance, do smoking and lung cancer have any relation to each other? Appendix D. • = - B • whole class of distributions, which admits simple sufficient statistics and to which this 0 example belongs, are introduced in the next section. 11.IO)) can be written as ')'1 = c3jci and /'2 = c4jc*., Xn is a sample from a N(p,,u2) population, where 11 is a known standard, and we are interested in the precision s) for s using 4. CHUNG, K.
1.3) and g * $\{G: J x d G(x) \notin 0\}$. I 6. j. Show that if P is a (discrete) canonical exponential family generated b(, (T, h) and &0 # 0, then T is minimal sufficient. I, 3rd ed. P(X 2 v) 2 i]. It is easy to see that the likelihood ratio test for testing H : g < 80 versus K : 8 > 00 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y) > 0 is of the form (x, y) > 0. If (x, y)(Problem 5.4.8) that, for a < . In Example 3.4.8, we found that in the class of unbiased estimators, X is the optimal estimator. In general, checking sufficiency directly is difficult because we need to compute the conditional distribution. Hint: Consider (] . This density corresponds to the distribution known as noncentral x2 dom and noncentrality parameter fP. New York: 1973. Implicit in this calculation is the assumption that Po1 [T > c0] is an increasing function ofn. A•(B) S (t) Proof. To deal with such situations we need an extension of Theorem 3.3.2. Theorem 3.3.3. Let o' be a rule such that sup8 R(O,o') = r < oo, let {tr.} denote a sequence of pn'or distributions such that $7rk\{8: R(B, 0^*) = r\} = I$. [!,- (X), !'+(X)] is a level (1 - a) confidence interval for I' · general, if v = v(P), P E P, is a parameter, and X • P, X E Rg, We say that In 'It. it be possible for a bound or interval to achieve exactly probability it may not (1 - a) for a prescribed (1 - a) such as .95. 10.4) i=l where eieT can be interpreted as projection on the onedimensional space spanned by ei (Problem B. Such families are called conjugal.'?. For instance, consider situation (d) listed at the beginning of this section. If $\{Po: 0 \in 8\}$, 8 C R, is an MLR family in T(x), then L(x, Oo, 01) = h(T(x)) for some increasing function h. 3050 I · Computation Speed of computation and numerical stability issues have been discussed briefly in Section 2.4. They are dealt with extensively in books on numerical analysis such as Dahlquist, Section 3.5 189 Nondecision Theoretic Criteria BjOrk, and Anderson (1974). Therefore, Because s2 function of I Tn I (n - 1) where 1 - (a5 / a2) = 1 + (x - Mo) 2 / 0' 2. Example 1.1.2. Sample from a Population. Show that the maximum contrast estimate B is consistent. '.. P., 'Sampling and Bayes Inference in Scientific Modelling and Robustness (with Discus sion)," J. For instance, in estimating B E R when X N(B, a5). One-Samp le Models. A newly discovered skull has cranial measurements (X, Y) known to be distributed either (as in population 0) according to N(O, 0, 1, 1). fact nothing has happened (the so-called null hypothesis) is more serious • ! • than missing something new- that has in fact occurred. Example 1.3.1. Ranking. 20. 6.4.3 Logistic Regression for B i nary Responses In Section 6 . " :i 13. 13 461 Some Classical Discrete and Continuous Distributions, which arise frequently in probability and statistics, and list some of their Following the name of each distribution we give a shorthand notation that properties. In order to avoid this ambiguity it is convenient to define the confidence coefficient to be the largest possible confidence level. The exponential convergence rate (8.9.5) for the sum of indepen dent Bernoulli variables extends to the sum Sn = 2: 1 Xi of i.i.d. bounded variables Xi, [X; - I' I < c;, where I' = E(Xt) - n P[ISn - n]I[> x) < 2exp - i x'/ 1:::Cl (8.9.6) i = I For a proof, see Grimmett and Stirzaker (1 992, p. 1 1 6 7.84 10.64 12.59 14.45 15.03 16.81 18.55 20.25 22.46 24.10 7 9.04 12.02 14.07 1 6.01 16.62 18.48 20.28 22.04 24.32 26.02 8 10.22 13.36 15.51 17.53 18.17 20.09 21 .95 23.77 26.12 27.87 9 1 1 .39 14.68 16.92 19.02 19.68 2 1 .67 23.59 25.46 27.88 29.67 10 12.55 15.99 18.31 20.48 21.16 II 23.21 25.19 27.1 1 29.59 31 .42 13.70 17.28 19.68 21.92 22.62 24.72 26.76 28.73 31 .26 33.14 12 14.85 18.55 21 .03 23.34 24.05 26.22 28.30 30.32 32.91 34.82 13 15.98 19.81 22.36 24.74 25.47 27.69 29.82 31 .88 34.53 36.48 14 17.12 21.06 23.68 26.12 26.87 29.14 31.32 33.43 36.12 38. = Bo versus K (} > Bo. Then 0 * is uniformly most accurate at : Proof Let 0 be a competing level (I - a) LCB Oo. Defined o (x, 00) \oplus 0 if, and only if, O(x) < 00. Suppose the possible states of nature are (}1, (}2, the possible actions are a1, a2, a3, and the loss function l(() a) is given by a, a, o 2 I 0 2 I Let X be a random variable with frequency function p(x, (}) given by I (1 - p) (I - q) and let d1, when . Express the band in terms of critical values for An(F) and the order statistics. 1: Tests for Heteroscedasticity, Nonlinearity," Ann. Although there are several good books available for this purpose, we feel that none has quite the mix of coverage and depth desirable at this level., Xn distributed according to Po. B E (a, b), a < eo < b, derive an optimality property, and then directly and through problems exhibit other tests with the same behavior. • • 545 Tables Appendix C (x' > x) Table III dJ X x2 distribution critical values .25 .10 .05 .025 Right tail probability property, and then directly and through problems exhibit other tests with the same behavior. discrete periods, a model that is often used for the time X to failure of an item is P, [x 🗞 k] & Bk -1(1 - B), k 🍖 1, 2, . Berlin: Springer, 1 977. "Stochastic Complexity (With Discussions)," J. Suppose that X1 , . and Cauchy distributions. (c) Using the fact that if(N" . . statistic is also invariant under reparametrization and, thus, approximately x; q · Moreover, we obtain the Rao statistic for the composite multinomial hypothesis by replacing Boj in (6.4.2) by ei (ij). The proof is straightforward: Po, [Vn1(80)(Bn - Bo) > z] • 1 - (z) by (5.4.40). As we have seen this parametrization is unidentifiable and neither f1 normality and the replacing Boj in (6.4.2) by ei (ij). * arc parameters in the sense we've defined. HAMPEL, P. Now the conditional distribution of Y given Z = z is the same as the distribution of Y given Z = z is the same as the distribution of Y if P is the probability measure on (0, A). Fortunately, a simple necessary and sufficient criterion for a statis tic to be sufficient is available. London: Oxford University Press, 1948. 128 Methods of Estimation (1) If $1 \times 4 d - x 1 < 20$, x = 1 < 20, x = 1 < 10, (1) If $1 \times 4 = 10$, (1) If 1column space of a matrix Anxp ofrankp < n, then (8. (a) Show that S in Example 2.4.5 has the specified mixture of Gaussian distribution. PARZEN, PA Pearson framework is still valuable in these situations by at least making us think of possible alternatives and then, as we shall see in Sections 4.2 and 4.3, suggesting what test statistics it is best to use. Find the MLE B of B. > 0 and Et = Eo, then, if ito• Llo, Eo are known, a UMP (for all)... That is, we want to find that the interval ' a such that the probability [X - a, X + a] contains p., 2, < Xn . We call C the set of compatible points. Hint: Because C' is of rank r, xC' = 0 => x = 0 for any r-vector x. Let Bm,n have a beta distribution with parameters m and n, which are integers. 22, The Views of Fisher and Neyman, and Later Developments, "Testing Statistical Hypotheses, 2nd ed. D(B,B o) = Es, p(X,, O) - p(X,, Ilo)) 328 Asymptotic Approximations Chapter 5 • ; • • is uniquely minimized at Bo. Let On be the minimum contrast estimate - 0. 'n = P (-2::., wild values. Does a new car seat design improve safety? Wiley & Sons, 1960. We want a uniformly most accurate level {1 - a) upper confidence bound q* for q(A) = 1 - e->.to, the probability of early failure of a piece of equipment. The basic bias variance decomposition of mean square error is presented. , Xn+k) be a sample from a population with density f(x I 8), e E e. Under what conditions on (x11 , Xn) does the MIE exist? Asymptotic analogues 0 * • I i • 0 of these inequalities are sharp and lead to the notion and construction of efficient estimates. J:n), c(b, t) = [*;", b > 1. I We would like to acknowledge our indebtedness to colleagues, students, and friends i , who helped us during the various stages (notes, preliminary edition, final draft) through ' which this book passed. This is the same as the probability 0 limit of the frequentist interval (4.8.1) An additional "dispersion" parameter can be introduced in some exponential family models by making the function h in (6. Use a construction similar to that of Problem B.4.1 0 to obtain a pair of random variables (X, Y) that (i) have marginal normal distributions., Yn . We use the word event here for lack of a better one because these , ' ' l j ' • • ' Section A.16 473 Poisson Process are not events in tenns of the probability model on which the N (t) are defined. Let the distribution Fo, and let Y1, Now the parametrization is unidentifiable because, for example, 11- = 0 and N(O, 1) errors lead to the same distribution of the observations a (1, 1) errors. In this case we define the inner product by ., Br)T, 0 < 8; < 1, L B; = 1., XN) then , , y > 0, A > 0. Then we expect D(80, 8) = o (2. Suppose Hint: Use (5.3.12). We shall analyze B ayesian credible regions further in Chapter Summary. L., "A Theory of Some Multiple 1. (x) 106 3.497 4.025 4.437 12 0.695 1.356 1.782 2.179 2.303 2.681 3.055 3.428 3.930 4.318 13 0.694 1.350 1.771 2.160 2.282 2.650 3.012 3.372 3.852 4.221 14 0.692 1.345 1.761 2.145 2.264 2.677 3.326 3.787 4.140 15 0.691 1.341 1.753 2.131 2.249 2.602 2.947 3.286 3.733 4.073 16 0.690 1 .337 1 . , X) could be taken with out gross errors then $P^* E P^*$ would be an adequate approximation to the distribution of X* (i.e., we could suppose X*, P* E P''). Then (see Table 1 .6.1) $\Phi(B) = 1$ i=1 . , en are dependent as are the X's. The statistic of rank k - 1 which generates the family is T(k-1) = (T1,
.W. Its densities are given by x''-1(1 - x)'-1 • • It follows by integration by parts that, for all p > 0, r(p + 1) = pr(p) and that r(k) = (k - 1)! for positive integers k. Formal definitions of insensitivity to gross errors. Hint: n n n 1 1 - n L 'l'(X,, 80) - n - L D¢(X,,IJ*) + ap(l) (1J* 1J0). YI), . , Proof. AND A. Example 4.1.4. One-Sided Tests for the Mean of Normal Distribution with Known Vari ance., Zd. Here II(Y I L) = E(Y) This is just (1.4.14). then II, II is a norm, That is. What is needed are simple sufficient conditions on p(x, B) for II to hold. However, as we shall see later, the first parametrization we arrive at is not necessarily the one leading to the simplest analysis., xN for the unknown current family incomes and correspondingly u1, . But c(9) is closed so that 71° = c(8°) and 8° must satisfy the likelihood equations. I.4, it has power against parameter values on either side of /1-0. Because the same interval (4.5. f6"; 'L(x, 0, 00)d1r(O) f_':x, L(x, 0, 00)d1r(O) Inference and Its Applications, 2nd ed., covers most of the material we do and much more but at a more abstract level employing measure theory. • , '1/Jd)T (2. R., linear Statistical Inference and Its Applications, 2nd ed. I ! ' I - Most measurement and recording processes are subject to gross errors, anomalous val ues that arise because of human error (often in recording) or instrument malfunction. A 143, 383-430 (1979). } q(s, Bnew) r(X I s, Bnew) log log I S(X) = s. One way of doing this is to align the known and unknown regions and compute statistics based on the number of matches. $a(\cdot)$. This implies that if is the distribution of a control, then $G(\cdot) F(\cdot \ \)$. We call this the shift model with parameter 6. (3) Condition A7 is essentially equivalent to (5.2.8), which coupled with (5.2.9) and - identifiability guarantees consistency of B in a regular model. Show that the density of $V = L \cdot (v) =$ i=O where R P (P) and frn is the x:n density. - - - (a) Show that X, X, X arc translation equivariant and antisymmetric. and Hall, 1986. - 2 + -- (l). ' 0 ! I ' i ! I | ' I ' Section 4. ' ' ' 532 Additional Topics in Probability and Analysis Appendix B 16. We are given a random experiment with sample space fl. A proof due to Wu (1983) is (2.4.21) Example 2.4.4 (continued)., p, or = = Thu s, Theorem 2. As we shall see shortly for X and in Unbiased estimates play a particularly important role in survey sampling. Suppose that each Xi has the Pareto density 1 + BI - (f(x,O) = c8Bx, x > c where 8 > 1 and c > 0. '• !'' I I 'i' g(x1 - B)g(x - t, -B)g(x - t, -B). With this definition it is 477 478 hook. ISBN 0-13-850363-X(v. $(1 - a_{,})$ LCB and q(X) is a level $(1 - a_{,})$ UCB for q(8), $(1 - (a_{,} + a_{,}))$ confidence interval for q(8). In the case of a normal sample of size n + 1 with only n variables observable, we construct the Student t prediction interval for q(8). In the case of a normal sample of size n + 1 with only n variables observable, we construct the Student t prediction interval for q(8). In the case of a normal sample of size n + 1 with only n variables observable, we construct the Student t prediction interval for q(8). $= \{a0, ..., Similarly, when f-t <$ \bullet -to. Of even greater concern is the possibility that the parametriza tion is not one-to-one, that is, such that we can have 01 f. Therefore, 01 k(B1, a) > k(B2, a) | > Po, [S > k(B2, a) - 1] > a, a contradiction. Let 1.4 p 1.2, b where Nij is the number of individuals of type i for characteristic 1 and j for characteristic 2. Its properties are similar to those of the trimmed mean. Normalizing Transformation for the Poisson Distribution. (c) Is the assumption that the 🏶 's are normal needed in (a) and (b)? of Statist., 2, 1 1-22 (1975)., Zn - Z). (iii) and (iv) The chance of any occurrence in a given time period goes to 0 as the pe riod shrinks and having only one occurrence becomes far more likely than multiple occurrences. Fisher's Method of Scoring The following algorithm for solving likelihood equations was proosed by Fisher-see Rao (1973), for example. tx for a function 1 g of a real variable that possesses g(xn) and g(x - 0) for limx. Let V and W be independent and have x and x?n distributions, respectively, and let S = (Vjk) (Wjm). 1 3) as N - oo for k = 0, 1, . We want to predict the value of a confidence interval for Y (i.e., statistics t(Y1, . To prove (i) note that it was shown in Theorem 4.3.1 (i) that Pe [S > j] is nondecreasing 02 and in 6 for fixed j. from the start, include examples that are important in applications, such as regression experiments., XN)), then $P(Y < y) = e-XFo(Y) e_1 - 11$, y > 0, A > 0. It is possible to give useful generalizations of Theorem not one-to-one (Problem B.2.2 to situations where g is B.2.7). are as in Example other Uj with probability 1r1 where size, then and Uj is retained independently of all 3.4.1 Lf 1 1rj = n. BRowN, L., Fundamentals of Statistical Exponential Families with Applications in Statistical Deci sion Theory, IMS Lecture Notes-Monograph Series, Hayward, 1986. Sequencing is done by applying computational algorithms to raw gel electrophoresis data. In a sample of n independent plants, write x; = j if the ith plant has genotype j, 1 < j 5 6. Other choices that are, as we shall see (Section 5.1), less computationally convenient but perhaps more realistically penalize large errors less are Absolute Value Loss: l(P; a) = f v (P) - a[, and truncated quadratic loss: $l(P, a) = min \{(v(P)-a)2, d'\}$. Xn. where xi X1 1 of the distribution of Xi. • • 1 .-v Find N ($\hat{\Phi}$ to , $\hat{\Phi}$)' j.lo is a- 2 is (called) the precision (a) Show that $p(x \mid B)$ ()(o: n exp (- itB) where t = $\hat{\Phi}$ 1 (X, -1"o) 2 and ()(denotes "proportional to" as a function of B. • = - I in Example 2.1.5 show that no maximum likelihood estimate of 8 = (Jl, 0'2) $\hat{\Phi}$ $\hat{\Phi}$ 15, Suppose that T(X) is sufficient for 8 and that 8(X) is an MLE of 8. Then 5 is identifiable whenever flx and flY exist. then by Problem B.3.4, U; \clubsuit U(O, 1). •, . (c) Let Yi denote the response of the ith organism, i = 11 • • •, n; j = 1, . 7.2) is equivalent to Sy (t I z) = Sf': (t). Last and most important we would like to thank our wives, Nancy Kramer Bickel and Joan H. Here the data function n} where Yi, .) be the set of possible realizations of X and let ti = T(xi). Then T is discrete and 2:::" 1 Po[T = I;] = 1 for every 8. are 21. i j l' Section 1.7 Problems 87 and Complements Hint: Apply the factorization theorem. Corollary 4.5.1. Under the conditions of Theorem 4.4.1, {t : $a(t,B) > a) = (-oo, to(1 - a)] \{B : a(t,B) > a) = [\clubsuit(t), oo).$ The family FL , S = {Fp,,a : -oo < J.t < oo , a > 0} is called a location-scale parameter, and Y is said to generate FL , S · From Fp,,a(x) = (x - fl.) F a =) (r(x - !") F..., r + Y, a we see as before how to refer calculations involving one member of the family back to any other. We give a contain any further information about () or equivalently decision theory interpretation that follows. , ") . - 1/ >. o = 12. We can then formally write an analysis of deviance analo gous to the analysis of variance of Section 6.1. If w0 c w1 we can write (6.5.6) a decomposition of the deviance between Y and j1,0 as the sum of two nonnegative com ponents, the deviance of Y to j1, 1 and (j1,0, j1,1) = D (Y, j1,0) - D(Y, j1,1), each of which can be thought of as a squared distance between their arguments., sv. 1 i=1 = i=-.1 t=1 (b) Show that under AO-A4 there exists c > 0 such that with probability tending to 1, ;. The moments do not exist for r 2: k, the odd moments are zero when r < k. Check AO, . If we require that h is increasing, this leads to h(11 (A) = Vc/J>..., A > 0, which has as its solution h(>.) = 2, JC, + d, where d is arbitrary. S. Suppose that in the Gaussian model of Ex ample 4.3.4. p, is unknown. • Problems and Complements Section 5.6 • . The most common choice of g is the linear form g(/31 z) = zT/3. Then, for ail n, (A. The remaining case T, = 0 gives a contradiction if c = (1, 1, . 33 I, 386 P(.)), Poisson distribution on the inter val (a, b), uniform distribution on the inter val (a, b), uniform distribution on the inter val (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution on the inter val (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution on the inter val (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution on the inter val (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 3 1 9 tribution, 492 ter ,\, 462 U (a, b), uniform distribution with parame of sample correlation, 492 ter ,\, 462 U (a,
b), uniform distribution with parame of sample posterior, 339, 391 asymptotic order in probability notation, 516 asymptotic relative efficiency, 357 autoregressive model, 11, 292 acceptance, 215 action space, 17 Bayes credible bound, 251 bisection, 127, 210 coordinate ascent, 129 asymptotic, 344 Bayes estirn • te, 162 EM, 133 Bernoulli trials, 166 Newton-Raphson, 102, 132, 189,210 equivan.ance, 168 for GLM, 4 1 3 proportional fitting, 157 alternative, 215, 2 1 7 Gaussim model, 163 linear, I > 1/>.o. It is easy to see that there is typically no rule c5 that improves all others. 14), we get the formula 2 2 Var X = E(X) - [E(X)] The covariance is defined whenever X1 and X2 have finite variances and in that case "I' (A. This statistic takes values in the set of all distribution functions on R. The prototypical example of a Hilbert space is Euclidean space RP from which the ln thiscase if x = (xi, ..., xp)T, y = (Yb ..., dvf E RP, (x,y) = abstraction is drawn. Here is an example in which the mean is extreme and the median is not. Thus, $C\{Xn + I I X = t\} = .c\{(Xn + I - 9) + 9 I X = t\} = N(/iB, 0"5 + a1)$ where, from Example 4.7.1, $0" \diamondsuit 2 B = n a5 1 I + 72 / J1 B \circledast = (0" B2 I 7 2) TJO + (n O" B2 I 0"02) X - \cdot a\}$ Bayesian prediction interval for Y is [Y3, Yt] with It follows that a level (1 - Yf = liB ± z (1 -) + 9 I X = t] = N(/iB, 0"5 + a1) where, from Example 4.7.1, $0" \circledast 2 B = n a5 1 I + 72 / J1 B$ PITMAN, J., Probability New York: Springer, 1 993. To complete the proof notice that if 1j(rJc) is any subsequence of 1j(r) that converges to $ry \cdot (say)$ then, by (I), $l(ry \cdot) = A$. Moreover, the random interval [v(X), v(X)] formed by a pair of statistics v(X), v(X) is a level (I - a) % confidence interval for v if, for all P E P, P[v(X) < v < v(X)] > I- a. If each time you test, you want the number of seconds of response sufficient to ensure that both probabilities of error are < 0.05, which system is cheaper on the basis of a year's operation? sup{p (x , B) : e where P (x , B) = p (x , B) = p (x , B) = p (x , B) notation of We can initialize with the method Example 2.3.2. For n > 2 we know the MLE exists. Suppose that we permit Gto bearbitrary. 3.8 REFERENCES ANDREWS, D. But (B.! 0.1 1) for ali a is equivalentto(B.!O.IO). The probability distribution of X is given by . I P[!OOII > 20] ! = " p !0011 - 1 0 JIOO(O.I) (0.9) I - C35) = Cha pter 1 (A 1 5 . l 1 .21). Semi parametric estimation and testing will be considered more generally, greatly extending the material in Chapter 8 of the first edition. We assume n 2 k - 1 and verify using Theorem 2.3.1 that in this caseMLEs of 'I; = log(>.;/>.•), 1 < j < k - 1, where O < >.; = P(X = j] < 1, exist iff all T; > 0. (b) Give explicitly the E- and M-steps of the EM algorithm in this case., Xn)T can be written as the rank 2 canonical (a) Show that the density of X = (X + j) < 0. (b) Give explicitly the E- and M-steps of the EM algorithm in this case., Xn)T can be written as the rank 2 canonical (a) Show that the density of X = (X + j) < 0. (b) Give explicitly the E- and M-steps of the EM algorithm in this case., Xn)T can be written as the rank 2 canonical (a) Show that the density of X = (X + j) < 0. then A(17) jc(7) = log / exp{c-1 (7)17TY} h(y, 7)dy. Show that SC(x iTn) (2a)-1 (x2 -) do not converge. 37. It will present important statistical concepts, methods, and tools not covered in Volume I. Show that Theorem B.7.5. implies Theorem B.7.5. impl for the possibility that the new drug is less effective than the old, then eo = [0, Bo] and 80 is composite. for A. Similarly, any solution B.(T) of Fo(T) = a with Ba E 6 is an upper confidence bound for 8 with coefficient (1 - a). We want to estimate B and VaroX1 = B(1 - B). Using this information, he wants to predict the first-year grade point averages of -oo 0 on X. A hybrid of the two methods that always converges and shares the increased speed of the Newton-Raphson method is given in Problem 2.4. 7. Specifically, (i) The posterior distribution is discrete or continuous according as the prior distribution is discrete or continuous. be i.i.d. X valued and for the moment take X = R. A random experiment has been perfonned. If 8 is not bounded, minimax rules are often not Bayes rules but instead can be obtained as limits of Bayes rules. (2) Suppose that if treatment A had been administered to a subject response x would have been obtained. Thus a distribution G has the same shape as F G E F. where the Yi are the indicators of a set of n Bernoulli trials with success probability p. For instance, we can invert the family of size o: We can also invert families of likelihood ratio tests to obtain what we shall call likelihood ratio tests to obtain what we shall call likelihood ratio tests of the point hypothesis H : e = e0 and obtain the level (1 - o:) confidence region C (x) where sup0 = denotes sup over e ro0 n { e : p(x, e)(x, e) } (4.9.2) 8 and the critical constant c (e) satisfies [sup0 p (X, e) p (X ' e0) * c (eo)] = o: . '. (4) Show how the ideas and results apply in a variety of important subfields such as Gaussian linear models, multinomial models, and nonparametric models. We want to test whether F is exponential, F(x) * 1 - exp(-x), x > 0, or Weibull, F(x) * 1 - exp(-x), x > 0, x 9), x > 0, B > 0. Let X1, , Xn1, Y1, - . enough steps J so that Ill On the other hand, at least if started close enough to 8, the Newton-Raphson method in 1 (j - 1) j fJ (T(X) - A(BU $^{(+)}$)) takes on the order which the jth iterate, fl() = iP-) A-1 (of log log ! steps (Problem 3.5.2). (b) Find E(Ye1Z+(1/Z)I 1 Z = z). This leads to P[IOOII 2 20 I X = 10] " 0.30. Suppose, for instance, we are interested in the proportion (} of "geniuses" (IQ 2: 160) in a particular city. a In the bivariate case write = aoz. Show that 1J. (A.I3.4), Xk are independent normal random variables with mean 0 and known variance u2 (cf. 1.28) for the noncentrality parameter B2 in the regression example. This is the most general form of the minimum contrast estimate we shall consider in the next section. A prototypical example follows. , b"9 for the preceding case (a). Section 1. Let en (a, ()0) denote the critical value of the test using the MLE On based on n observations. On the other hand, consider the Gaussian linear model of Example 2.1.1. Then least squares estimates are given in closed fonn by equ(\tion (2.2.10). Now suppose 8 is Euclidean c Rd, the true 80 is an interior point of 8, and 8 D(Bo, 8) is smooth. Let J1-5 E (X1) denote the mean difference = : between the response of the treated and control subjects. The conditional frequency function p is defined only for values of z such that pz(z) > 0. A consumer organization preparing (say) a report on air condi tioners tests samples of several brands. Next use Problem B.3. 1 1 and "" roo Pv (v) = n J0 00 L P(R = i)j,i+1 (v - s) fn-l (s)ds. ft - 8 f > ; i=1 (5.5.14) y'ne-" IPn - P I > W- $(0 \text{ for some } A < oo. Let any T(X) be Var0(T(X)) < oo for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 521 Topics in Matrix Theory and Elementary Hilbert Space Theory because y = A - *x
maximizes the quadratic form. Kretch, and R. Unfortunately (1.3) so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1/;(0). 7) j'., I \bullet ' Section 8.10 so for all B.allDenote E0(T(X)) by 1$ exponential scale model (see Problem 1 . The mathematical model suggested by the descrip tion is well defined. This is evidently a linear space that can be shown to be closed. The parameter >.. (a) Show that the correlation of X1 and Y1 is 3. • i=l (A.12.6) This follows by induction from the definition and (A. If $v_0 = (1, 0, ... This is the usual$ Pythagoras's Theorem., Un ordered, then . Here we can write the n determinations of p, as • • • . Suppose AI: The parameter O(P) given by the solution of j, P(x, O)dP(x) '- 0 (5.4.21) is well defined on P. Problems for Section 3.3 1. However. A stockholder wants to predict the value of his holdings at some time in the fu ture on the basis of his past experience with the market and his portfolio. It may be shown (see Volume II) that the (approximate) tests based on Z and Fisher's test are asymptotically equivalent in the sense of (5.4.54). (zt, Yi). Models are approximations to the mechanisms generating the observations. 10) for {3 is easy to evaluate if d is at all large because inversion of + 1)/2 terms with Z'EZD requires on the order of nd2 operations to evaluate each of n operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to evaluate each of n operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to evaluate each of n operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 operations to get Z'J;ZD and then, if implemented as usual, order d3 the distribution of h(X) improves as the coefficient of skewness 'YI n of h(X) diminishes. Index., 13(nk, nk,e) distribution. n ("n ---.. (a) 1) (Y)) O(n = E(h(X, h J1.1, Jt)2 + -). Condition on V and apply the double expectation theorem. As our examples suggest, there is tremendous variation in the degree of knowledge and control we have concerning experiments. '• The numerator is an increasing function of 1., 'I''' I'' Section 2.5 Problems and Complements 6., R(Bk, 8)) and if k = 2 we can plot the set of all such points obtained by varying 8. Measure theory will not be used. 1 'Here !'(X) is called an upper level (l - a) confidence bound for I'. Finally, in many situations where we want an indication of the accuracy of an estimator, we want both lower and upper bounds. Reading, MA: Addison-Wesley, 1974. M., AND A. In this case a 2 = A and Var(X) = Ajn. Use Problem 1.5.12. Let X (1/n) 2:: (1/ a > 0. Riccardi Executive Managing Editor: Kathleen Schiaparelli Senior Manager: Trudy Pisciotti Marketing Manager: Trudy Pisciotti Marketing Manager: Trudy Pisciotti Marketing Manager: Angela Battle Marketing Manager: Angela Battle Marketing Manager: Trudy Pisciotti Marketing Manager: Angela Battle Marketing Manager: Angela Battle Marketing Manager: Manager: Angela Battle Marketing Manager: Ma Assistant: Joanne Wendelken Art Director: Jayne Conte Cover Design: Jayne Conte I'n'nlict Ilall i @2001, 1977 by Prentice-Hall, Inc. Show that ?T(m 1 as n --+ oo whatever be a. I 4.9) upon writing A.16.4 Let T1 N(t) • N, (t) + (N(t) - N, (t)). (c) Show that if a > of coverage < • ' I 'I ! ' ' ' 1 - a, a/ and >. (B.L. 1) This P� is just the conditi probability measure on (0, A) mentioned in (A.4.2). 7 Problems and Complements This is known as the Pareto density. H., Optimal Statistical Decisions New York: McGraw-Hill, 1969. By convention, if drug A is a standard or placebo, we refer to the x's as control observations. , Oko) under H, but under K may be either multinomial with 8 # Bo or have Eo(Ni) nBio, but Varo (N;) < nOiO(1 - 0;0)("Cooked data"). Form of the Two-Sided Tests Let B = ([1-, a2), 80 { (]1-, a2) : [1- = fl-o]. An exhaustive census is impossible so the study is based on measurements and a sample of n individuals drawn at random from the population. and can conclude that the statistic of (6.5. 10) is asymptotically x under H. r2 = denominator degrees of freedom. In Example 4.1.3 with o(X) • 1{S > k}, Bo • 0.3 and n = 10, we find from binomial tables the level 0.05 critical value 6 and the test has size a(6) • Pe, (S > 6) • 0.0473. The Neyman Pearson Framework The Nea important. They initially tabulated Nm1, Nft. As a consequence of Theorems 2.3.2 and 2.3.3, we can conclude that an MLE /.i always exists and satisfies (2.3.7) if n 2 2. 1.4 we discussed situations in which we want to predict the value of a random In addition to point prediction of that contains the unknown value Y Y, it is desirable to give an interval with prescribed probability (1 - a. Here are some further examples of the kind of situation that prompts our study in this section. : fl (1 - a) r = 2. xTy = L = 1 xiyi · ll xll 2 = : L = 1 xiyi · ll space consists of the numbers 0, 1, . -oo • ' ' • • 11 I j' 1 ' -- Section B.ll 527 Problems and Complements (a) Show that T = (sv n+1, . To get information we take a sample of n individuals from the city. Problems for Section 6.4 1. The "if' part of (a) is deeper and follows from the spectral theorem. Suppose that lj1, ' = 1, . In addition, the set includes a large number of problems with more difficult ones appearing with hints and partial solutions for the instructor. Suppose that we only record the time of failure, if failure occurs on or before time r and otherwise just note that the item has lived at least (r + 1) periods. Hint: () ---+ g; (X, B) is continuous and (; ;; (X, B') • JO -B' | < . Discuss the preceding decision rule for this "prior." (c) Discuss the behavior of the preceding decision rule for large sider the general case (b). Here, in pseudocode, is the bisection algorithm to find x. situations, 1.4 are for (a) [is the linear span of 1, Zt, . Our model is then specified by the joint distribution for 11 E w as mi --+ oo, i 1, . c}. In general, let a(t, v0) denote the p-value of a test b (T, vo) = 1
[T > c] of H : v = vo based on a statistic T = T(X) with observed value t = T(x). Here a is small, usually .05 or .01 or less. iteration of the Newton-Raphson algo On described in (b) and that O satisfies n Hint: You may use the fact that if the initial value of Newton-Raphson is close enough to a unique solution. 8 8 If we let k sup8 $R(B, \& \bullet)$. = (M, a2), P 2. F(xp) [1 - F(xp)] = p. (a) The beta family. That is, sections of the surface z = px (x, y) by planes parallel to the (y, z) plane are proportional c, Px to Gaussian (normal) densities. The problem of deciding whether B = 80, 8 < 80, or B > Bo is an example of a three decision problem and is a special case of the decision problems in Section 1.4, and 3.1-3.3. Here we consider the simple solution suggested by the level (1 - a) confidence procedures, we H test : [1- = fl-o, where we think of as an established standard for an old treatment. The Dirichlet distribution is a conjugate prior for the multinomial. Let 8, P, and p be as in Problem 6.2.1 densities of the form = {N({L, a2} > 0}, p (x, 8) and let Q be the class of distributions with (1 - E) 0 and, hence, p is strictly convex. The following data are the blood cholesterol levels (x's) and veight/height ratios (y's) of I 0 men involved in a heart study. Show that if H is simple and the test statistic T has a 0. 1 linear subhypotheses are important. Hint: Po,[X 🛛 x I S(X) 🗳 s] 🔄 1 xi, 1: 🗘 1 (xi - x?) (e) (I: 🔄 1 xi, 1: 🔄 1 (xi - x?) (e) (I: 🔄 1 xi, 1: (I: (I + 1))) and (I: (I + 1)) and Regions Chapter 4 (iii) Consider the model G(y, efiFo (Y) 1 0) e • - 1 j, O f O Fo(Y), o 🕏 o. Many interesting examples are discussed in the books of Feller (1968), Parzen (1 969). In terms of our preceding discussion. , ii } appears at [E(X, . Distribution-Free, Unbiased and Equivariant Procedures. The most trivial example of a sufficient statistic is T(X) = X because by any interpretation the conditional distribution of X given T(X) = 2.288J e) I (1 + E7 = / e l Thus, by Theorem 3.4.3, the lower bound on the variance of an unbiased estimator of <math>l/1; (8) = $E(n-1 T_{2}(X)) = >.$; is >.; (1 - A;)fn. The comments contain digressions, reservations, and additional references. Note that there are many ways ofchoosing a parametrization in these and all other problems. Identify ,.., B, T, and h. Show that if C is an n x r matrix of rank r, r :=:; n, then the r x r matrix C'C is of rank r, r :=: approximation works well., X*) is a good estimate., En are uncorrelated with mean 0 and variance a2 • Give the least squares estimating a, {3, 5, and J.t. . Appendices. However, the interpretations are different: In the frequentist confidence interval, the probability of coverage is computed with the data X random and B fi xed, X is computed with 4.8 whereas in the Bayesian credible interval, the probability of coverage = x fixed and (} random with probability distribution II (B I X = x). and the data are x where d is one of 2, 10, or 50. We begin this in the simple examples that follow and continue in Sections 1.2-1.5 and throughout the book. , Wnf • 1) orthogonal matrix whose first row is $C(V_{,,.} A \text{ proof of the Cauchy-Schwartz inequality is given in Remark 1.4. 1. Other transforms, such as the p robit 9 I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms, such as the p robit 9.I (1r) = -1 (1r) where is the N(O, 1) d.f. and the log-log transforms (1r) d.f. and the log transforms (1r) d.f. and transforms (1r) d.f. and$ t0 ¢ C¥. Show that, although N is random, ../N(X - p)/so, with X = I:f 1 X 0 when 0"2 is known. Argue as in Prol;>lem 4.9.4. "'''' (d) Relate the two-sided test of part (c) to the confidence intervals for a5faf obtained in Problem 4.4.10. Using I and II we obtain, ..,P' (B) j T(x):Bp(x,B)dx = j T(x) (:B logp(x,B))p(x,B)dx. • j-1 1 if L nk + 1 < i �'n (90,n)E -... See Theorem 5.3.3. A.18 REFERENCES BERGER, J. Show that under A2. Statistical Models, Goals, and Performance Criteria 8 1. In the Gaussian linear model show that the parametrization ($\{3, a2\}$) T is identifiable if and only if r = p. 1 where (Xi I B) N(e, 0"), 0" known, and 1r(B) is N (T]o, 72), 72 known. That is, the standardized versions of verge in law to a standard normal random variable. (1.3.7) Confidence Bounds and Intervals Decision theory enables us to think clearly about an important hybrid of testing and estimation, confidence bounds and intervals (and more generally regions). , Xn is a sample with xi ,...., p(x I 0), a regular model and integrable as a function of e. announce as your level {1 (b) Suppose that Ef 2, a (x) < to0(1 - a) where to0 (1 - a) is the 1 - a quantile of Fo0• By the duality theorem, if $s(t) = \{8 \in 6 : t < to(1 - a)\}$, then S(T) is a we find 1 - a confidence region for B. We usually take P to be parametrized, $P = \{Pe: 0 \in 8\}$. Ahmed and N. Suppose I": $(x_i = 0.1 \text{ Note that in this example the power at } () = 81 > 0.3 \text{ is the probability that the level } 0.05 \text{ test will detect an improvement of the recovery rate from } 0.3 \text{ to } fh > 0.3. When 81 is 0.5, a 67\% improvement, this probability is only .3770. Also note that P(X > 0.3) is the probability is only .3770. Also note$ t) = S0 (t). (a) Show that (1., Xn) is an £ (>. is defined whenever XI and X2 are not constant and the variances of X1 and X2 are finite by 'I (A. An . The power is a function of 8 on e 1 If Go is composite as well, then the probability of type I error is also a function of 8. be (l) The x's andy's are
realizations of X1, Xm a sample from F, and Yt, ... By Corollary 4.3.,... = . The packages that produce least squares estimates do not in fact use formula (2.1. 10). 538) one can improve on 318 Asymptotic Approximations by utilizing the third and fourth moments. Chapter 3 of the first artd deals with estimation., r, where 9i is chosen so that H becomes equivalent to "(B�, . WEO ' I I I i Section 2.5 145 Problems and Complements Now show that WMLE W w or q(O). Formally, we can define (} : P ----+ 8 as the inverse of the map 8 ----+ Po, from 8 to its range P iff the latter map is 1-l, that is, if Po1 = Pe2 implies 81 = 82. To this we have added the normality assumption. The Bayesian formulation is based on the posterior predictive distribution which is the conditional distribution of the unobservable variables and l such that I - a = P(k < S < n - 1 + I) = 2:7 k + 1 pi(! -p)"-i. a meaning to Data-based model selection can make it difficult to ascenain or even assign the accuracy of estimates or the probability of reaching correct conclusions. For instance, if a2) population with a2 known, there is no UMP test = f-lo vs K : f-1 = f. In the context of the foregoing examples, we could leave the component in, replace it. Suppose we want 0 < p < 1. However, insofar as possible we prefer to take the frequentist point of view in validating statistical statements and avoid making final claims in terms of subjective posterior probabilities (see later). If Mx is well defined in a neighborhood { s : lsi < so} of zero, all moments of X finite and = E(Xk) k s , lsi < so. We solve (3(0,) = f3 for n and find the approximate solution fh Oo a = $.05, f3 = .90, Oo = 0.3, and 01 = 0.35, we 11 eed n = (0.05)-2 \{1.645 \ 0.3(0.7) + 1.282 \ 0.35(0.55)\} 2 = 162.4.$ For instance, if X X Thus, the size .05 binomial test of H : 8 = 0.3 requires approximately 163 observations to have probability .90 of detecting the 17% increase in e from 0.3 to 0.35. Here are the elements of the Neyman Pearson story. Show that Un (0, 0, 0), Un (0, 0), but Un + 0, p > I, where L is defined in Problem B. Hint: Show that if the convex support of the conditional distribution of Y1 given Z1 = zUl contains an open interval about P, j for j = 1, . random variables., l)r. Because the class of tests with level a for H : () < ()0 is contained in the class of tests with level a for H : () = 00, and because dt maximizes the 0 power over this larger class, bt is UMP for H : () < Oo versus K : () > Oa. The following useful result follows immediately. : a test, {31 + {hz;, z; not all which rejects, if and only if, 1 ' 434 Inference in the Multiparameter Case Chapter 6 12. , with probability) and given the treatment, while the second patient serves as control and receives a placebo . , Xn be distributed as where x1, $\mathbf{\hat{v}}$. It is -1 or 1 in the case of perfect relationship (Xz = a l- bX1 , b < 0 orb > 0, respectively). We could then represent the data by a vector X = where = 1 if the ith item sampled is defective and = 0 otherwise. Because, evidently, max R(F, X) = max :F :F VarF (X,) n M n , 0 Theorem 3.3.3 applies and the result follows. o-so = 1 and zl = Xt. 22 = P l 1-Po l 4 6. Our aim is to use the data inductively, to narrow down in useful ways our ideas of what the "true" P is. There are two other types of test that have the same asymptotic behavior. Define the Neyman-Pearson (NP) test function . 3) p(y I z)pz(z) q(z I y)py(y) Eyq (z I y)p y(y) $(B \cdot Here F_{2}; 1 (a) = \inf\{x : Fn(x) > a\}$. The breakdown point will be discussed in Volume II., p, iterate and proceed. where ! (A.12.8) di (A. Now we may either have prejudices that we are willing to express in the form of a prior distribution on B. The following model is useful in such situations. n (1.5.8) p{x1, . See Remark 4.1. In this example with 80 = { (}0} it is reasonable to reject I-J if S is "much" larger than what would be expected by chance if H is true and the value of () is 00. l(B) ' I I ' ' II I i=1 The Newton-Raphson method can be implemented by taking X. However, if we are comparing a treatment and control, the relevant question is whether the treatment creates an improvement. i 270 Testing and Confidence Regions Chapter 4 (a) Use the result of Problem B.3.4 to show that the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - a) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - o) is the (1 - o:)th quantile of the X and the test with critical region [X > !"ox(1 -)/2n], where x(1 - o) is the (1 shipments with i defective items, i = 0, . i = 1, . (!.5.9) if every Xi is an integer between 1 and B and p(x 11 (1.5.9) can be rewritten as • • • , Xn, 0) = 0 otherwise. Most accurate upper confidence bounds are defined similarly. Thus, in Example 1.1.3, taking action 1 would mean deciding that D. 1) Note that Tp (Y) acts as a prediction interval pivot in the same way that T (p,) acts as a confidence interval pivot in Example 4.4.1. Then Pe [X k] the geometric distribution (9 (B)). Then Zr = Y1/8', Z = (Y2 Y1)/8', z, = (Y2 Y1)/8', z, = (Y3 - Y2)/8', z, = (Y2 Y1)/8', z, if, we can write T2 = H (T1) for some 1-1 transformation H of the range of T1 into the range of T2. There are clear dependencies between starred Pref3ce to the exponential fautily P generated by (T, h). Li (1 - 1) + C2' 1 - 1 for 2. Find (a) P(X + 2Y < 4). For all x E A. Now V has the same distribution as L 4 1 J!i2 where Y1 , . However, the result holds whenever the quantities in J(O I Oo) can be defined in a reasonable fashion. 1 5. Let X, . 4, . stance, if N = 50, n = 5, and D = For in 20, the approximating binomial distribution to H(D,N,n) is 8(5, 0.4). Therefore, z E(Y, I Z) = -. For instance, we can assign two drugs, A to m, and B to n, randomly selected patients and then measure temperature and blood pressure, have the patients rated qualitatively for improvement by physicians, and so on. AND R. A, and E. Tben Theorem 2.3.3. lfP above satisfies the condition of Theorem 2.3.1. c(8) is closed in [and T(x) = to satisfies (2.3.2) so that the MLE ij in P exists, then so does the MLE II in Q and it satisfies the likelihood equation • - cT(ii)(t0 - A(c(ii)) = 0. where 1ri = 1r (zi) is the probability of successes D are as likely to occur on one set of trials as on any other. (4) We can decide if the models we propose are approximations to the mechanism generating the data adequate for our purposes. The theory of this school is expounded by L. The critical problem with such parametrizations is that even with "infinite amounts of data," that is, knowledge of the true Pe, parts of fJremain unknowable. Discrete Distributions The binomial distribution with parameters n and B : B(p(k) The parameter () e'(J - e)"-•, n, B). Similarly, r(a1 I) = 5.70 = and we conclude that 6 (1) = a,. Next we must specify what close means. (a) Show that min (X1, Xn) is sufficient for fl when a is fixed. Suppose that three possible actions, ah a2, and a3, are available. (b) Which estimate has the smallest MSE for estimating $0 = \oint (/1 \ 2)$, where f(t) is the th quantile of the Fnz-I, nt -I distribution. Example 1.3.2. Prediction. (a) Show that wpen p = 0, T has a Tn 2 distribution. J W for the known last census incomes. The book first discusses non- and semiparametric models. Show that a conjugate family is the gamma family is the gamma family. Because J-l > 0, the D solution we seek is il+. (a) The U(O, 9) fumily 88 Statistical Models, Goals, and Performance Criteria Chapter 1 (b)p(.", O) = $\{exp[-2 \log 0 + \log(2x)]\}$ [x E (0,0)] (c) p(x, O) =
\mathbf{O} . = E [g(Z) - Y] 2 The MSPE is the measure traditionally used in the i Section 1.4 33 Prediction mathematical theory of prediction whose deeper results (see, for example, Grenander and Rosenblatt, presuppose it. function : For instance, one "estimator" of this shape is the scaled empirical distribution $P_{F,(x)}$ ifn, $x(n : S x < x(i+1) \cdot j = 1$, of 81 with respect to 82 is defined as ep(81, 82) = then = 1r/2. Compute the Rao test statistic for H case. , Xn as a random samp le from F, and also write that Xb ... Let Sn _, 519 B(n, p) , then P([Sn - np[> nc) < 2 exp{-n 0. (b) Construct an

experiment and three events for which (i) hold, but (ii) hold, but (iii) does not. Theorem 5.3.1. $l_j(i)$ and (ii) hold, then m- 1 h(j) () Eh(X) = h(!1) + L 'I !! E(X J · j = 1 where •• 'I • - !!) + Rm (5.3. I) - Section 5.3 First- and Higher-Order Asymptotics: The Delta Method with Applications 307 The proof is an immediate consequence of Taylor's expansion. The power .35 and = achievable (exactly, using the SPLUS package) for the level .0:> test for 0 163 is 0.86. If B E 80, /3(B, o) is just the probability of type I error, whereas if B E 81, /3(B, o) is the power against (}. The statistic is equivalent to the likelihood ratio statistic j, for this problem. Thus, if we want the probability of detecting a signal v to be at least a preassigned value j3 (say, .90 or .95), then we solve (z (a)+ (v./ii/(J)) = !3 for n and find that we need to take n = (J/v)2[z(1 - a) + z(/3)f D This is the smallest possible n for any size a test. a = T and p, .17, O" arc arbitrary. Before the experiment is performed, the information or belief about the true value of the parameter is described by the prior distribution., Xm, Y1, . We next give a representation of the process whereby the statistician uses the data to arrive at a decision. ... In one of his famous experiments laying the foundation of the quantitative theory of genetics, Mendel crossed peas heterozygous for a trait with two alleles, one of which was dominant. A very important class of 4.9.2. In Problem 4.9. 1 1 we argue that 6 = (!1 - f,Lo) / a. This 0 - 1 loss function can be written as 0 - ! loss: 1(8, a) = 0 if 8 E e., Xn) L t=1 are a sample from a probability P 2 is the which evaluated at S x) on R and 1(A) is the indicator of the event A. It follows that striving for numerical accuracy of ord r smaller than n - 112 is wasteful. it is hard to translate statements about orders into specific prescriptions without assuming at leaSt bounds on the constants involved., qd(ll)) and a = (a, . 17) ' (1) and a = (a, . 17) ' (1) and a = (a, . 17) ' (1) and a = (a, . 17) ' vector of values called a covariate vector or a vector of explanatory variables whereas Yi is random and referred to as the response variable or dependent variable or dependent variable in the sense that its distribution dependent variable or dependent variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variable or dependent variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random and referred to as the response variables whereas Yi is random an Transformation and Weighting in Regression New York: Chapman and Hall, 1988. 8. Deduce from Theorem 1.6.1 that if X 🕹o . Alternatively we can appeal to Corollary 2.3.1 directly D (Problem 2.3.10). M., Mathematical Analysis, 2nd ed. ' ' Section A.18 475 References (5) The integral in (A.8.12) may only be finite for "almost all" x. 1.6.3 Buildince from Theorem 1.6.1 that if X 🕹o . Exponential Families Submodels A submodel of a k-parameter canonical exponential family { q(x, 17); 11 E £ an exponential family defined by ' I 1 p(x, 8) = q(x, ry(8)) I I c Rk} is (1.6.12) where 6 E 8 C R1, I < k, and 17 is a map from 8 to a subset of Rk. Thus, if X is discrete taking on k values as in Example 1.6.7 and X = (X1, . The basic asymptotic tools that will be developed or presented, in part in the text and, in part in the functional delta method. As a consequence our second edition, reflecting what we now teach our graduate stu dents, is much changed from the first. (6.5. 1 2) The left-hand side of (6.5. 10) is of product form A(17) [1 / c(7)] whereas the right-hand side cannot always be put in this form. 12. Because !!(!L) is increasing, a (c) = sup{fl(IL) : fL < 0} The smallest c for which (-c) = fl(O) = (-c). Show that the only variance stabilizing transformation h such that h(O) = 0, h(1) = 1, and h'(t) > 0 for all t, is given by h(t) = (-c). 2/7r)sin-1 (vt). Math. Let U(1 < · · · < U(n) be U1 , . There are modeled by the Pareto distribution function 14. {flo, . Let >- where a > 0, = b Suppose that given a-2, xl , · · · , Xn are i.i.d. We would also like to thank the colleagues and filends who Inspired and helped us to enter the field of statistics. (a) Show that, if n > 2, the likelihood equations t ' w 🔶 { (X;: Jt) t=1 a unique solution - log fo that Cfl, O')., n, eo = 0 where €i are independent identically distributed with density f. ' I 11., not UMP. Suppose that T has a continuous distribution Fe. then the p-value is U = 1 - F0 (T). Now, if (L52) holds, Po [T = t;] = {x:T(x)=t,} p(x, O) = g(t; O) {x:T(x)=t;} h(x). Y1, ... In estimating a real valued parameter v(P) or q(6') if P is parameterized the most commonly used loss function is, Quadratic Loss: l(P, a) = (v(P) - a)2 (or l(O, a); i' r " • = (q(O) - a)2). G are not necessarily nonnal but that < Var(Xf) < oo. See Problem 3.2.9. Finally, in the Bayesian framework with a prior distribution on the parameter, the approach of Example 3.2.2(b) is the one to take in all cases with 80 and 81 simple. a) % confidence 424 Inference in the M u lti para meter Case 10. Roy. (0 For two estimates 01 and ... HUBER, W. Statist., 3, 1038-1044 (1975a). See also Section 1.4. , Xn are indepen As a consequence of (A. In the two sample models this is implied by the constant treatment effect assumption. a is interpreted For instance, testing techniques are used in searching for regions of the genome that resemble other regions that are known to have significant biological activity. Specifically, upon substituting the f3(r, s) density (B.2.11) in (1.2.9) we obtain L7 = (Jk+r-I (1 c 8) n-k+s-I (1.2.10) The proportionality constant c, which depends on k, r, and s only, must (see (B.2.11)) be B(k + r, n - k + s). This result was proved in various forms by Fisher, Neyman, and Halmos and Savage. ANDERSON, Numen"cal Analysis New York: Prentice Hall, DE GROOT, M. See Problem 1.6.17 for details., 1.1.4 Examples, Regression Models We end this section with two further important examples, Regression Models We end this section with two further important examples indicating the wide scope of the notions we have introduced. Show that (a) Likelihood ratio tests are of the form: Reject if, and only if, Hint generating function of X is given by 1 Mx(t) = e' o, i=1 • . I XJ, . It follows that, in this case, the optimal estimator when it exists is also the optimal predictor. (B, Yl), . • • Example 1,6,8. Note that logp(x, 8) is differentiable for all 8 > x, that is, with probability 1 for each 8, and we can thus define moments of {)/88 log p(x, 8). HAMPEL, F. HAMPEL, F. E. Assume that Fo has the density fo. In fact we can write X, I + ,, i 2, , , , n, X1 I + ,, i 2, , , , n, X1 I + ,, i 2, , , , n, X1 I + ,, i 2, , , , n, X1 I + ,, i 2, , , n, X1 I + ,, i 2, , , n, X1 I + ,, i 2, , , n, X1 I + ,, i 2, , n, X1 I + , i 2, , i $\{Po: () E 9\}\$ is a canonical exponential family generated by (T,h) satisfying the conditions of Theorem 2.3.1. Let S(X) be any statistic, then 136 Methods of Estimation Chapter 2 (a) The EM algorithm consists of the alternation • A(Bnew) = Eo01d (T(X) I S(X) = s) (2.4. 18) Bold = Bnew (2.4.19) unique. We return to this in Section 5.5.]1 3.5.3 Robustness Finally, we turn to robustness. vanance. 1 and good with probability .9 independently of the other items, this will continue to be the case for the items left in the lot 1008 - X, the number of defectives left after the drawing, is independent of X and has a 8(81, 0.1) distribution. The problem of selecting the better of two treatments or of deciding whether the effect of one treatment is beneficial or not often reduces to the problem of deciding whether B < 0. The algebra showing Rn (80) = x 2 in Section 6.4. 1 now leads to the Rao statistic = ., = 1] = Rn (8(")) 2 where the right-hand side is Pearson's x 2 as defined in general by (6.4. 1). } Then (2.4. 15). Give a probabilistic interpretation of this Hint: Use the Taylor expansions for e0../fi and e-o.,; in powers of yfij., Xk) T is t, [x, C :',] log = •. 'I !I, '' i 'I' 176 Measures of Performance Chapter 3 More specifically, we show how finding minimax procedures can be viewed as solving a game between a statistician S and nature N in which S selects a decision rule 8 and N selects a prior 1r. (n.) sample. (a) Suppose Z1, z; have a N(O, 1) distribution. Note T = that E[(Xn + 1 - e)e I 9 = B} = 0. The danger is that, if they are false, our analyses, though
correct for the model written down. This quantity is called the size of the test and is the maximum probability of type I error. .. (2) We shall use the notation g(x+O) for limx. TABLE 1.3.4. Risk points (R(01, ot), R(B,, 51)) 1 • R(01, 8;) R(O,, 5;) 1 0 12 2 7 7.6 3 3.5 9.6 4 3 5.4 5 10 1 6 6.5 3 7 1 .5 8.4 8 8.5 4.0 9 5 6 It remains to pick out the rules that are "good" or "best." Criteria for doing this will be D introduced in the next subsection B. 💎 -'••• ' I Both of these assertions follow immediately from Definition(B. (i) IfY and Z are independent, then p(y I z) = py(y) and the conditional distribution. REFERENCES BERGER, J., Yn1 correspond to the group receiving the first treatment, Yn1 + 1, that getting the second, and so on. p. I e,, . j=l Let C" denote the interior of the range of (c, (B), . 0'. (c) Show that gc(x)/P(x) is of order exp{x2} as j x j * oo. t=1 On satisfies (6.2.3). P0 [Tn 2': t] is increasing in !. Without this device we could not know whether observed differences in drug performance might not (possibly) be due to unconscious bias on the part of the experimenter. (2) Give careful proofs of the major "elementary" results such as the Neyman-Pearson lemma, the Lehmann--Scheff6 theorem, the information inequality, and the Gauss-Markoff theorem. So to get information about 8, a sample of n is drawn without replacement and inspected. In fact., xTn; Yt, ... Here only two actions are contemplated: accepting or rejecting the 'specialness" of P (or in more usual language the hypothesis $H : P \in Po$ in which we identify P0 with the set of "special" P's). ' in) is any permutation of (1, . Maintaining our geometric intuition we see that, if E(X) = E(Y) = 0, orthogonality simply corresponds to nncorrelatedness and Pythagoras's theorem is just the familiar • Var(X + Y) = Var(X) + Var(X + Y) + Var(X) + Var(X + Y) = Var(X) + Var(X + Y) + Var(X) + Var(X + Y) = Var(X) + Var(X + Y) + Var(X) + Var(X + Y) = Var(X) + Var(X + Y) + Var(X Var(r) if X and Y are uncorrelated. Suppose that X has a normal N(ft, cr2) distribution and that Y independent of X and has a N(-y, r2) distribution. It will be convenient to assume(l) from now on that in any parametric model we con sider either:, (1) All of the P, are continuous with densities p(x, 0); (2) All of the Po are discrete with frequency functions p(x, 8), and there exists a set {x1, x2, . (b) Show that N = (N1, . Suppose 8 is a vector. Here J.L(z) is an unknown function from R d to R that we are interested in. By Theoretto 5.3.3, y'n[h(X) - h(JL)] ". (The decision is correct) l(0, a) = 1 otherwise (that the NILE of {3 solves E {3 (Tj) E 3 (z r x) = z r x, where Z = llzij llrnxp is the design matrix. All of these correspond to geometric results in Euclidean space., Xn is a sample from a truncated binomial distribution with I p(x, 0) = (:) ' ' I. Show that the conditional distribution of aX + bY given eX + dY = t is normal. Box, G. Because we do not know whether A or B is to be preferred, we test H : B = 0 versus K : 8 -1 - 0. > 0. The population is so large that, for modeling purposes, we approximate the actual process of sampling without replacement by sampling without replacement by sampling without replacement. P[T(t.) < t] � 0 and In T(�) Suppose that instead Var(Y]). We say that a procedure 6 improves a procedure J' if, and only if, R(B,6) < R(B,6') for all (} with strict inequality for some (}., Yn) is not a sequence of 1's followed by all O's or the reverse. The notion of such transformations can be extended to the following situation. Variance Stabilizing Transformations can be extended to the following situation of such transformations can be extended to the following situation. binomial, Poisson, gamma, and beta, which are indexed by one or more parameters. The proof of the spectral theorem is somewhat beyond our scope MacLane (1953, 1 pp. (b) How would you, in principle, use this result to construct a test of H similar to the 2 x test with probability of type I error independent of Tfit, 11i2? Thus, by Chebychev's inequality, if oo, Pp. a' Pp[]Xn - I'] > e] ::; 2exp { - I by taking q(0) equal to the noncentrality parameter governing the distribution of the statistic under the alternative. llm+1 @ Om + l - 1 (9m)Dl(9m) Show that for GLM this method coincides with the Newton-Raphson method of Section 2.4. ' i i • • j 1 1 1 j I ' 1 ' . • • p(e 1)p(c, l e,)p(e 3) for the statistic under the alternative. e\$,e,),, p(e,. W., Exploratory Data Analysis Reading, MA: Addison-Wesley, 1972., e\$) T ranges over an open subset of Rq and Bj Boj, j q + 1, . New York: AND D. By applying Fo to hnth sides of t 5 to(1-a), S(t) \$ {B E 6 : Fo(t) < 1 - a}. (2) A second major approach has been to compare risk functions by global crite- Section 1.3 27 The Decision Theoretic Framework ria rather than on a pointwise basis. They need to be read only as the reader's curiosity is piqued. where the errors are independent, identically distributed, and symmetric about 0 with com mon density f and d.f. F. Derive maximum likelihood estimates in the following models. Logistic Regression. 26 Statistical Models. P R E D CT I O N I NT E RVA LS I n Section variable Y. ' n). In the discrete case we appeal to the product rule. Reid. •: P[N, : -where (B, C, D, .) If the equation Fo(t) • 1 - a has a solution li.(t) in e. If !v!x is well defined in some neighbor hood of zero, Kx can be represented by the convergent Taylor expansion 00 c Kx(s) = L; sj ., Ap)PT => A-1 • P diag(>-• 1, (Rayleigh density) = (f) $f(x, 8) = 8cx'-1 \exp\{-8x'\}$, x > 0; c constant > 0; 8 > 0., n (0, 1] is a known constant and t: 1, . Drew, C. The second follows from (B. To see that this is a linear model we relabel the observations as Y1, Yn. where Yn1+n2 to Y1, . Suppose in Problem 6.4.6 that H is true. The population is sampled with replacement and n nonnal distribution. Then Let E = E2 E21 En E1 1, E22 are spd. For instance, in Example 1.1.1 contractual agreement between shipper and re ceiver may penalize the return of "good" shipments, say, with (] < 80. If (say) X represents the amount owed in the sample and v is the unknown total amount owed, it is natural to seek v(X) such that P[v(X) >] > 1-a v (1.3.8) for all possible distributions P of X. A proportional hazard model. X11 have finite variances, we obtain as a consequence of (A 1 1 . Do these variables have a bivariate normal distribution? 11.20) ·t=1 If X1 and X2 are independent and X1 an appropriate., Xn be a sample from f(x 0), () E R. H., AND B. 2 fo (x) \Rightarrow 3 r(o'(x) I x) = E(l(O,o'(X)) 1 X = x]. Assume that there exist functions that the model for Yi can be written as h(y, T), b(B), g(p) and c(T) such o, y - b(O,) \Rightarrow h(y,r) exp c(T) g(!"i) Var(Y)/c(r)exists. Suppose A0-A4 hold and 8 is fo consistent; that is, 8 = 80 + Op(n- 112). As a first illustration, consider the oil-drilling example (Example 1.3.5) with prior 1r(ll1) = 0.2, 1r(ll,) = 0.8. Suppose we observe x = 0. Show that if M is the expected sample N E(M) & L 1r, N n. Does a new drug improve recovery rates? R is called the empirical distribution function, - where I j (X1, • . • ' I (c) Suppose that T and Y have densities fo(t) and g(t). ,Xrn areidentically distributed as are Y1, ... • ' I Section 4.3 Uniformly Most Por (I) follows from Oo among the class of tests with level a = Eo,61 (X). (See Example 2.4.3.) Hint: (a) Thefunction D(a, b) D(a, b) L: 1 w (aX; - b) - n log a is strictly convex in (a, b) and lim(a,b)-(ao,bo) D (a, b) = x if either ao = 0 or or bo = (b) Reparametrize by a = ; , b = : and consider varying a, b successively. P(Y < y) = e , - - 1 . we need only keep track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped
demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR family, then rejecting the track of X(n) = max(X1, . The theory we have devel oped demonstrates that if C,(Tn) is an MLR f for large values of Th is UMP among all tests based on Th. Reducing the problem to choosing among such tests comes from invariance consideration that we do not enter into until Volume II. By expanding the product (X 1 - E(X J)) (X 2 - E(X 2)) and using (A. If a solution of (2.4.18) exists it is necessarily (b) If the sequence of iterates {B-m} so obtained is bounded and the equation A(B) = $Ee(T(X) | S(X) = s) \cdot (2.4.20)$ has a unique solution, then it converges to a limit B", which is necessarily a local maximum of q(s B) = I XI, ., n, , i Section 4.10 275 Problems and Complements then E7 1 Xi is an optimal test statistic for testing H : 8 = 80 versus I\; 8 > 80. For instance, in Example 1.1.1 we observe X and the family Pis that of all hypergeometric distributions with sample size nand population size N. These considerations lead to the asymmetric formulation that saying P E Po ((} E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E Po and P E P1 (c) E 80) corresponds to acceptance of the hypothesis H : P E P0 (c) E 80) corresponds to acceptance of the hypothesis H : P E P0 (c) E 80) corresponds to acceptance of the hypothesis H : P E P0 (c) E 80) corresponds to acceptance of the hypothesis H : P E P0 (c) E 80) corresponds to acceptance of the hypothesis H : P E P0 (c) E 80) corresponds to acceptance of the hypothesis H : P E 90 (c) E 80) corresponds to acceptance of the hypothesis H : P E 90 (c) E 80) corresponds to acceptance of the hypothesis H : P E 90 (c) E 80 (c) earlier, acceptance and rejection can be thought of as actions a = 0 or 1, and we arc then led to the natural 0 1 loss l(0, a) = 0 if () E Sa and 1 otherwise. Consider the three estimates T1 Y, T2 (1 I 2n) 2...: 1:1 Yi + (3 I 2n) 2...: 1:1 Y One second of transmission on either system costs \$103 each. That is, it is distribution free. BAYESIAN MODELS 1.2 Throughout our discussion so far we have assumed that there is no information available about the true value of the parameter beyond that provided by the data. Then = $p(y,z) p(y | z) = (B, Suppose X1, \bullet \bullet)$, Xn is as in Problem 1.5.3. In each of the cases (a), (b) and (c), show that the distribution of X forms a one-parameter exponential family. By Coronary 3.4.1 we see that the conclusion that X is UMVU follows if • • •, - • Var(X)]! l nl, (B) ' (3.4.18) • a2/n, whereas if .; = nE(T;(X)) + &l = l e (:)2 ne el (1 + E7 11 ee1 - ee3) = n > .; (l - -1;) = Var(T; (X)). the probability of the wrong decision is at most \mathbf{O} a. (b) (6.3. Show that in the regression example with interval for /31 in the Gaussian linear model is p = r = 2, the 100(1 z. = E 1, . After the value x has been obtained for X, the information about () is described by the posterior distribution. 1). (x, (c) Show that the tangents to So at the two points where \mathbf{O} is described by the posterior distribution. 1). the line y 1/2+p(a2/cr,)(x J.:t d intersects Sc are vertical. Hint: A Bayes test rejects (accepts) H if ll oS ecets (accepts) H i LR test of H : af = a versus K : ai > af is of the form: Reject if, and only if, F (1 - 1)/(n2 - 1)E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution under H. • • ' (} with distribution given by 1. The most common choice of g is the linear f011t1, (3) g((3, z) & L.;), f31 zi & zT (3 so that (b) becomes (b') This is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. Find a function a simple form and a tabled distribution given by 1. The most common choice of g is the linear f011t1, (c) g((c) - 1)/(n2 - 1))E(j - Y) 2/E(X; - X)2 > C. model. The number of defectives in the first example clearly has a hypergeometric distribution; the number of a particles emitted by a radioactive substance in a small length of time is well known to be approximately Poisson distributed. (b) In Example 2.4.6, verify the M-step by showing that BeT = (JI-1, /1-2, a + /1-), a +) 2. That is, the effect of z on Y is through J.1,(z) only. The class Q of possible predictors g may be the nonparametric class QN p of all g : Rd -----Jo or it may be to some subset of this class. P., The Feynmo.n Lectures on Physics, v. (b) Suppose that max (x1 , . In fact (Problem 6.4.2), the (Ni] - RiCJ /n) are all the same in absolute value and, where z 🖗 tt [1 R J -] 1 = 1 1 • Section 6 . Which of the following statistics are equivalent? Amer. LEHMANN, "Descriptive Statistics for Nonparametric Models. A.13.7 If X is the number of defectives (special objects) in a sample of size n taken without X has If the sample is taken without X has replacement, X has replacement from a population with D defectives and N -D nondefectives, then H(D, N, n) distribution (see (A.6. I 0)). In other ' situations certain P are "special" and we may primarily wish to know whether the conditional distribution of Y given Z or of neither type. It turns out to be useful to know transformations h, called variance stabilizing, such that Var h(X) is approximately independent of the parameters indexing the family we are considering From (5 3 6) and ' Section 5.3 First- and Higher-Order Asymptotics: The Delta Method with Applications 317 2 2 l (1')]) [li a fn. 9 Thus, the posterior risks of the actions a1, a2, and aa are r(a, I 0) r(a, I 0) + - 2, 8 i(02, at) g r(a, I 0) 9 10.67 5.89. TOPICS IN MATRIX THEORY AND ELEM ENTARY H I LBERT SPACE THEORY 8.10 8. The hypothesis that the characteristics are assigned independently becomes H : (Jij 'TJi 1 T}j 2 for 1 :::; i :::; a , 1 :::; j 2 are nonnegative and 2: = 1 Tj 2 = 1. i < j. Clearly, Example 1.1.3(3) is a special case of this model. A I ! 2 I Statistical Models, Goals, and Performance Criteria Chapter 1 priori, in the words of George Box (1979), "Models of course, are never true but fortunately it is only necessary that they be useful." In this book we will study how, starting with tentative models: I (I) We can conceptualize the data structure and our goals more precisely. Wiley & Sons, 1964., Y, l be independently distributed according to N(p,, 0"2) and . Thus, we observe independent XI, . u,v;, Vt/v v,-., 1jT and all c. Show that 2 log. (X,, li : 1 < i < n) has a null distribution, which is a mixture of point mass at 0, sin .6. To achieve level a and power at least {3, first let Co be the smallest number c such that ...-t Then let n be the smallest integer such that P,, [T 2: co] > f] where Oo is such that q(O) = q0 and 01 is such that q(O) = q1 . Are sex and admission to a university department independent classifications?, UN . Consider the model (see Example 1 . Now we can write the likelihood as k k qo(x, a) * exp{L ; * (x) - n log L exp(ct;)}. and consider the model (see Example 1 . Now we can write the likelihood as k k qo(x, a) * exp{L ; * (x) - n log L exp(ct;)}. a, X and s2 are independent with X \bullet N(iJ., a 2 jn) and (n 1)s2 ja2 "X \bullet -1 This leads to p(X, s2 I J.L, a2). Find minimal sufficient statistics for the following three cases: (i) p., TJ, (ii) p., 17 < 00, 0 < a, T. 11/J'(B) [< Var(T(X)) ar (! logp(X,B)). Thus, we have a location parameter family. • • • = Pabc. They are
included both as a check on the student's mastery of the material and as pointers to the wealth of ideas and results that for obvious reasons of space could not be put into the alternative 8 holds. Edgeworth Approximations The normal approximation to the distribution of X utilizes only the first two moments of X. whereas the receiver or is special and the general does not wish to keep "bad," (} 2 Bo, shipments. This two-dimensional problem is than one-dimensional problem is the equaessentially no harder the of nested logistic regression models of dimension q < r < k and m1, . , Yn, respectively, the responses of m subjects having a given drug B. (b) Find the density of R if p 🕏 0. Let Nj be the number of xi which equal Vj . "Hierarchical Credibility: Analysis of a Random Effect Linear Model with Nested Classification, "Scand. J ' "" (9)a 'i7Ee(ar9) and apply Theorem 3.4.3. 22. and (t] denotes the greatest integer < t. and If S c EU is finite, then EU Rk is convex and closed, g ij convex on S, E S, -h is (strictly) P[U E S] Eg(U) exists and Eg(U) > g(EU) with equality if and only if there are a and = I, (B.9.3) bkx 1 such that In particular, if g is strictly convex, equality holds in (B.9.3) if and only if P[U for some Ckxl· ' " = c] = ' 1 1 For a proof see Rockafellar (1970). In any case all our models is in are generic and, as usual, "The Devil the details!" All the principles we discuss and calculations we perform should only be suggestive guides in successful applications of statistical analysis in science and policy., K, 1.3.4.) "' \pounds 1 Suppose the u; can be relabeled into h = N. = 2: 1 X; D Example 4.3.3. Consider the one-parameter exponential family mode! $p(x,O) = h(x) \exp\{r y(O)T(x) - B(O)\}$. n. This chapter uses multivariate calculus in an intrinsic way and can be viewed as an essential prerequisite for the more advanced topics of Volume II. The lower (upper) value v(v) of the game is the supremum (infimum) over priors (decision rules) of the infimum (supremum) over decision rules) of the game is the supremum (infimum) over decision rules) over either works or does not work; a certain location either contains oil or does not; a patient either has a certain disease or does not, and so on. E A, where A is the simplex {>. Let X be the number of failures before the first success in a sequence of Bernoulli trials with probability of success 9. Chen, S. is P. 3.4 3.4.1 UNBIASED ESTIMATION AND RISK I NEQUALITIES Unbiased Estimation, Survey Sampling In the previous two sections we have considered two decision theoretic optimality princi ples, Bayes and minimaxity, for which it is possible to characterize and, in many cases, compute procedures (in particular estimates) that are best in the class of all procedures, D, according to these criteria. P(U(j) : S: Un + 1 : S: U(k)) J P(u j c : s: Un + 1 : S: V I u(j) = u, u(k) = v) dH(u, v) = E(U(k J) - E(U(j 1)) where H is the joint distribution of U(j) and Uc k) + By Problem B .2.9, E(U(i)) = thus, P(X(j) : S: Xn + 1 : S: X(k)) = k-j ij (n + 1); (4., Y;m., Xn, eo E. If X11 • . , Xn denote the incomes of n persons chosen at random from a certainpopulation. Suppose that we know that {31 Show that, for suitable a, there is a UMP level 1:: f 1 z;N; > k, where Pp [1:: f 1 ;;;N; 2: k] = a. Using Examples 1.6.5 and 1.6.10, we see that the distribution of { $j_1 : j = 1$, Show that the distribution of { $j_1 : j_2 : j_1 : j_2 :$ continuity of the first integral ensures that {) $8(J \ 80 \ -= \ 0>$. The density of Pe may be written as 11 $p(x, 0) = \exp["2x \ x^2 \ 2"2 = 2 \{(!", .:r) \ 2 \ 1 \ J12 \ - 2 \ ("2 + \log(2rr.:r))]$, which corresponds to a two-parameter exponential family with and "T $(x) = x, \ J2(9) = 2, 1 \ "1 \ - q = : - \ oo$. BICKEL, P., AND E. (b) Deduce that x' P 2 contingency table model let probability of error (of either type) is as small as possible. Our hypothesis is then the null hypothesis that the new drug does not improve on the old drug. We next give a useful inequality relating product moments to marginal moments: HOlder's Inequality. + - - - - - 0 - - + 0 - - + 0 - - + 0 - - + 0 - - + 0 - - + 0 - - + 0 - - + 0 + --- + 0 1.5 25 --- + 2 ' Log10 (smaller sample size) Figure 5.3.3. Each plotted point represents the results of IO, {X)() two-sample tests. Here we are using F to represent F because every member of F can be obtained from F. 1 , . Then we have the classical Gaussian linear model, which we can write in vector matrix form, (c) where Z n x d = (zf, . In a modem formulation, if there were n dominant offspring (seeds), the natural model is to assume, if the inheritance CT2, then the drug effect is measured by p. (1.2.3) p(x, p(x I Because we now think of B) as a conditional density or frequency function given 8 = we will denote it by 0) for the remainder of this section. 1.2 may equivalently be written p A = L eie[Ai (B. Another application is continued until r a-particles have been emitted. Note that in general the test statistic L depends intrinsically on ito, [] 1. (2.3.7) Note that c(ll) E c(8) and is in general not ij. (3) The control responses are normally distributed, Then ifF is the N(J1 + 6., a-2) distribution, we have specified the Gaussian two sample 0 model with equal variances. I 1. Is it unique? 9. Begin by taking a fixed number no and calculate X0 = (1/n0) I: * 1 X. and . 77, 733-74 1 (1990). B.l.l The Discrete Case The reader is already familiar with the notion of the conditional probability of an event A given that another event B has occurred. £[X - lt]i = E(X - l')] Proof. By convention this is chosen to be the type I error and that in ;; ' , - , :' '. For a proof, see Problem 8.9.1. Hoeffding's Inequality. Title. Often the type I error and that in ;; ' , - , :' '. For a proof, see Problem 8.9.1. Hoeffding's Inequality. outcome of the experiment is used to decide on the model and the appropri ate measure of difference. It is possible to generalize the notion of Section 8.3, for r even E(Xr) = E(Qr) = kr(Xr) = E(Qr) = E(Qr) = kr(Xr) = E(Qr) = E(Qr) = E(Qr) = E(Qr) = E(Qr) = kr(Xr) = E(Qr) how to choose c., n, t:o 0, c are i.i.d. N(O, o-2). A., "Robust Estimates of Location," Ann. Inference When the Number of Parameters Is Large. Estimation. Hint: Use the factorization theorem (Theorem 1.5.1). It gives careful proofs of major results and explains how the theory sheds light on the properties of practical methods. .1n. Determine the smallest value k = k(a) such that Jk(o) has level a for H 1 and show that for n large, k(a) - h(a), where (a) Show that testing H h (a) C': n(l - p) + Z1 - u)np(! - p). This framework is natural if, as is often the case, we are trying to get a yes or no answer to important questions in science, medicine, public policy, and indeed most human activities, and we have data providing some evidence one way or the other., n, generated by h(Y) = I and Yjl • . HorXJES, JR., 1. (a) Express mean income J.L in terms
of 8. 14. = Var X = .\. Let k be a positive integer. (I - 9)'9, k • 0, I, 2, . , Xn are recorded. 2 Order on Symmetric Matrices 520 B.! 0.3 Elementary Hilbert Space Theory 521 B.! l Problems and Complements 524 8.12 Notes 53 8 B.13 References 539 •• CONTENTS XII C TABLFS Table II t Distribution Critical Values 545 546 547 INDEX I 1 ' PREFACE TO THE SECOND EDITION: VOLUME I In the twenty-three years that have passed since the first edition of our book appeared statistics has changed enonnously under the impact of several forces: (1) The generation of what were once unusual types of data such as images, trees (phy logenetic and other), and other types of combinatorial objects. Treating f as a parameter, show that the order statistics $X(t) \cdot .$ (c) There is an interval such that h(y + J) (a, b), a < b, such that for every y E (a, b) there exists a J > 0 - h(y) > h(y) - h(y - J). 6. We can proceed as follows. 7 425 Problems a nd Lornol; emEmts \cdot or dose of a treatment, and a response variable Y, which is yield or production. Hint: If [x[1 * L;: 1 [x] [, x * (x 1, . Of course, model (a) assumes much more but it may be a reasonable first approximation in these situations. We consider intervals based on observable random variables that contain an un observable random variable with probability at least (1 - a). The particular angle of mathematical statistics is to view data as the outcome of a random experiment that we model mathemati cally. In that case, if 0 < a < 1, there exists a unique smallest c for which a: (c) < a. Von Neumann's Theorem states that if e and D are both finite, then the game of S versus N has a value v, there is a least favorable prior n" and a minimax rule §* such that J* is the Bayes rule for n* and rr* maximizes the Bayes risk of J* over all priors. a2), where a2 is the variance of t:. We might, for each of a group of n randomly selected patients, record sleeping time without the drug (or after the administration of a placebo) and then after some time administer the drug and record sleeping time again. If p � 0, {So} is a family of concentric circles e,, 462 N(p, E), analysis of variance (ANOVA), 367 ..., multivariate normal distribution, 507 antisymmetric, 207, 209 asymptotic distribution I of quadratic forms, 5 0 asymptotic distribution I of quadratic forms, 5 0 asymptotic distribution I of meetimate, estimate, estimating equation estimate, 330 N(JJ, u2), normal distribution With mean JJ and variance u2, 464 N(JJt, j.t2, uf, u), bivariate normal dis of MLE. The p-value is a(t, B) = Po (T > t) = 1 - Fo(t). 0 By a standard argument it follows that, 1j(r) (11) It is endowed with au inner product (): 11 x 11 regiment R such that () is bilinear, \cdot , \cdot , \cdot , \cdot (ah1 + bh2, ch3 + dh4) = ab(h1, h2) + ac(h1, h2) + ac(h1, h2) + ac(h1, h2) + bd(h2, h4), symmetric, (h1, h2) + ac(h1, h2) + bd(h2, h4), symmetric, (h1, h2) = (h2, h1), and (h, h) > 0 with equality iff h = 0. It follows that, when = that given is conditionally distributed as Y) where is uniform on whatever be D Thus, is sufficient. Test in both cases whether the events [being a man] and [being admitted] are independent. 'Peter J. Conventions: (i) In order to minimize the number of footnotes we have added a section of comments at the end of each chapter preceding the problem section. Consider the following algorithm under the conditions of Theorem 2.4.2. Define Tj0 as before. Show that 8 \$\overline\$ depends on X through T(X) only provided that 8 is unique. (b) Find the maximum likelihood estimate of Pe [X1 > t] for t > 11 · Hint: You may use Problem 2.2. 16(b). We next compute the limiting distribution of ., fii (Bn - B). i I = 1 Y) + Z2. Although we believe the material of Chapters 5 and 6 has now become fundamental, there is clearly much that could be omitted at a first reading that we also star. Thus, the receiver wants to discriminate and may be able to attach monetary costs to making a mistake of either type: "keeping the bad shipment." In testing problems we, at a first cut, state which is supported by the data: "specialness" or, as it's usually called, "hypothesis" or "nonspecialness" (or alternative). It may be shown that 11 is characterized by the property h - 11(h I C) h' for all h' E C. L6HMANN, E. New York: Springer, 1998. • I if T(x) > t 6t (x) 0 if T(x) < t' • '' (4.3.3) with 61(x) any value in (0, 1) if T(x) = t. (ii) If we denote the corresponding (posterior) frequency function or density by rr(9 I x), then $rr(8 \ 1 \ x) rr(0)p(x \ I \ 8) L_{;}$, $rr(t)p(x \ I \ 8) roo rr(t)p(x \ 1 \ 8)$ roo $rr(t)p(x \ 1 \ 8)$ roo rr(trange of Z such that g(Z) (the predictor) is "close" to Y. , then the probability of exceeding the threshold (type I) error is smaller No one really believes that H is true and possible types of alternatives are vaguely known at best, but computation under H is easy. - - 80ld - BMOM D The Newton-Raphson algorithm has the property that for large n, Tinew after only one step behaves approximately like the MLE. Let Hint: (a): See Problem B.4.4. (b): Let A be an orthogonal matrix whose first row is (n- i, . If N8 is the number of defective items in the population, then n, E(X) Over X no(! - 0). That is, either NAA cannot really be thought of as stochastic or any stochastic or any stochastic I. · · •. Show that X + Y and X - Y are independent, if and only if, crf 4. Go to (1). Hooo, R., "Adaptive Robust Procedures," J. Pyke's careful reading of a next-to-final version caught a number of infelicities of style and content Many careless." mistakes and typographical errors in an earlier version were caught by D. Show that N, P[Z I Y V I - B)N-n-(,-y) V z-y (i.e., the binomial probability of successes in Hint: P[Z V z I Y V y) where b(y) L ' Y has a N - n trials). 1 .6.) It follows that Fisher's method for cgmbining p-values (see 4.1.6) is UMP for testing that the p-values are uniformly distributed against F(u) = u8, 0 < B < 1. Show that ..., Xn is a sample from a N(B, a2) population, where a2 is known. Given X = k, Y has a binomial B(k,p) distribution. We want a prediction interval for Y = Xn + 1 " F, where Xn + 1 is independent of the data X1, , Xn . Note that in the health versus smoking context, we can think of E(Y I Z z = z as the mean health rating 0 for people who smoke z cigarettes a day. D generally except in the Gaussian case. Suppose Y and Z have joint density function p(z, y) = z + y for 0 < z < 1, 0 < y < 1. This loss expresses the notion that all errors within the limits ±d are tolerable and outside these limits equally intolerable. FELLER, W., An Introduction to Probability Theory and Its Applications, Vol. 132 2.4.3 Methods of Estimation Chapter 2 The Newton-Raphson method. • , Xn is a sample from a Poisson P(O) population. The claims (iii) and (iv) are left as exercises. For instance, a statistic we shall study extensively in Chapter $\hat{\Phi}$ function valued statistic F. whereas the situation is reversed if the sample size inequalities and variance inequalities agree. On the other hand, if f-1 1 < f-Lo, the MP level a test .(x) = $p(x, \hat{\Phi}/p(x, 00)$. (b) If x = (c, y) is proportional to a normal density as a function of y. defined by H2 (x) $\hat{\Psi}$ x2 - 1, H (x) 3 $\hat{\Psi}$ x3 - 3x, Hs(x) $\hat{\Psi}$ x5 - 10x3 + 15x. Variation in the population is modeled on the log scale by using the model , log Y; $\hat{\Psi}$ log a - 15 !og{ 1 + exp[-f](t; - p)/15]} + r; where \in 1, . (a) What is the distribution of (Nt, . (b) The quantile sign test Ok of H versus K has critical region {x : L: $\hat{\Psi}$ 1 [Xi > OJ > k). Does a new marketing policy increase market share? 1.3 the questions are sometimes simple and the type of data to be gathered under our control. Let As usual let Xt, . Suppose X = (X,, . Hint: 13. The log likelihood of 1r (1r1, , 7rk) T based on X (XI, . On the other hand, { r(X I s, Bnew) - Ecold log r(X I s, B) I S(X) = s old by Shannon's inequality, Lemma 2.2.1. } >0 (2.4.17) D The most important and revealing special case of this lemma follows. We observe (z1, Y1),. Using our discussion given + = in Section B. Maximum likelihood and dimensionality calcu lations similar to those for the 2 X 2 table show that Pearson's x2 for the hypothesis of independence is given by (6.4.9) which has approximately a calculation given + = in Section B. Maximum likelihood and dimensionality calculations similar to those for the 2 X 2 table show that Pearson's x2 for the hypothesis of independence is given by (6.4.9) which has approximately a calculation given + = in Section B. Maximum likelihood and dimensionality calculations similar to those for the 2 X 2 table show that Pearson's x2 for the hypothesis of independence is given by (6.4.9) which has approximately a calculation given + = in Section B. Maximum likelihood and dimensionality calculations similar to those for the 2 X 2 table show that Pearson's x2 for the hypothesis of independence is given by (6.4.9) which has approximately a calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood and dimensionality calculations are calculated by the calculation given + = in Section B. Maximum likelihood are calculated by the calculation given + = in Section B. Maximum likelihood are calculated by the calculation given + = in Section B. Maximum likelihood are calculated by the x(a-I) (b-l) distribution under H. j=116. N(t + h) l - N(t) is
independent of N(s) for all s < t, h > 0, and has a P(:h) distribution. Then { 'lj } has a subsequence that converges to a point 71° E [. The problem of selecting good decision procedures has been attacked in a variety of ways. In the model of Problem 5(a) compute the MLE (Ot , 02) under the model and show that (a) If 01 > 0, e, > 0, C(y'n(B1 - 01, ii, - 0,)) . Xn are i.i.d. as X with X F, where "...," stands for "is distributed as." The model is fully described by the set F of distributed as." The model is fully described by the set F these are assumptions at best only approximately valid. We establish analogues of the information inequality and use them to show that under suitable conditional distribution of x I 8) for each 0. Show that { Xk t , . I we see that given = t, the conditional distribution of x I 8) for each 0. + are the same and we can conclude and that of + = t, has a U(O, t) distribution. Example 3.3.4. Minimax Estimation in a Nonparametric Setting (after Lehmann). To simplify the rule further we use the following equation, which can be established by expanding both sides. His or her measurements are subject to random fluctuations (error) and the data can be thought of as p, plus some random errors. If S(x) is a level (1 - a) con E S(:r) = v0 when vo Chapter 4 is a level a: test. The connection is through the concept of entropy, which also plays a key role in information theory-see Cover and Thomas (1991). However, Chapter 1 now has become part of a larger Appendix B, which includes more advanced Volume topics from probability theory such as the multivariate Gaussian distribution, weak con vergence in Euclidean spaces, and probability inequalities as well as more advanced topics in matrix theory and analysis. , vd) = (qt(ll), . The data is a point X x in the outcome or sample space X. Such restrictions are natural if, for instance, we test the efficacy of a treatment on the basis of two correlated responses per individual. 8 (iii) Reparametrize as in Theorem 6.3.2 for 2 log), (X). discrete case and is satisfied by any function of interest when Sets B that are members of Bk are called measurable., N. (b) Compute EY = 3. This is not the same as our previous hypothesis unless all departments have the same number of applicants or all estimates of fl-1, /1-2, a?, a^(*), p coincide with the method of moments estimates of Problem 2.1 .8. (iii) ll is idempotent, D2 = IT. The closed fonn here is deceptive because inversion of a d x d matrix takes on the order of d3 operations when done in the usual way and can be numerically unstable. EM for bivariate data. , N4) has a M (n, fh , . The idea of robustness is that we want estimation (or testing) procedures to perform reasonably even when the model assumptions under which they were designed to perform reasonably even when the model assumptions under which they were designed to perform excellently are not exactly satisfied. , Xn) Example 1.5.1. A machine produces n items in succession. The design of the experiment may not be under our control, what is an appropriate stochastic model for the data may be questionable, and what 80 and 81 correspond to in terms of the stochastic model may be unclear. I'l Example 5.4.2. Hodges's Example. Ranking. Moreover, the statistical procedure can be designed so that the experimentary stops experimentary as soon as he or she has significant evidence to the effect that one drug is better than the other. Stratified Sampling. However, X(k) does have the advantage that we don't have to know a or even the shape of the density f of Xi to apply it. For instance, if in the binary data regression model of Section 6.4.3, we take g (J-L) cp - 1 (J-L) so that = 'lri = :(Y) defined in Remark 6. Example 2.3.5. Location-Scale Regression The notions of parametrization and identifiability are introduced. Express the Cramer-von Mises statistic as a sum. (5.4.20) In what follows we let P, rather than Pe, denote the distribution of Xi. This is because, as pointed out later in Remark 5.4.3, under regularity conditions the properties developed in this section are valid for P � { Pe : 0 E 8}. : (32) to the First Edition Chapter 8 on discrete data and Chapter 9 on nonparametric models. This gives ,I where I I'±(X) = X ± o-z (1 -), we let w = { 11 : rJi = zT { 3, { 3 E RP } and let r be the dimension or data. of w. Without Winston Chow's lovely plots Section 9.6 would probably not have been written and without Julia Rubalcava's impeccable typing and tolerance this text would never, it is intuitively clear that if we are interested in the proportion 0 of defective items nothing is lost in this situation by recording and using only T. SMITH, Bayesian Theory New York: Wiley, 1994., Xn) is given by ' - Divide top and bottom of (5.5.11) where l(x, B) = logp(x, B) to obtain (5.5.11) where l(x, B) = logp(x, B) to obtain (5.5.11) where l(x, B) = logp(x, B) and . For further dis cussion of this generalization see McCullagp., 0.9 + 0 (d) The N(O, 02) family, 0 > 0 (e) p(x, O) = 2(x + 0)/(1 + 20), 0 = logp(x, B). < x < I, 0 > 0 (f) p(x, 9) is the conditional frequency function of a binomial, B(n, 0), variable X, given that X > 0. The risk points (R(B1, 6), $R(B_1, 6)$, $R(B_1, 6)$, I = 1, Consider a covariate x, which is the amount Section 6. for B.10.2.2 The Generalized Cauchy-Schwarz Inequality E n E12 be spd, (p + q) x (p + q), with En.P x p, ,, q x q. so that t) (! = \hat{v} or a1 = independent samples - >.)a/ + .\a \hat{v}]:). Note that in the case of intervals this is just inf{ P(!c(X) < v < v(X), P E P]} (i.e., the minimum probability of coverage). Second Edition Mathematical Statistics Basic Ideas and Selected Topics Volume I Peter J. 18. X11), Xi. (X(I)... (6.5.8) This is not unconditionally an exponential family in view of the A0 ($z f{3}$ term. 1) where p and pz are the frequency functions of (Y, Z) and Z. 4th ed. • • , /3d)T of un knowns. , Xn be the n determinations of a physical constant J.t. Consider the model where xi = 1, (2) Claims (5.5.8) and (5.5.9) hold with a.s. replaced by in Po probability if A4 and A5 are used rather than their strong forms-see Problem 5.5. 7. Properties (i)-(iii), of II above are immediate., 1/n). So is Example 1.1.2 with assumptions (1)-{4). Construct an exponential family of rank k for which £ is not open and A is not defined on all of t:. + (x �) I � 0 (5.5.9) i ' ' 340 Asymptotic Approximations Chapt er 5 Remarks (I) Statements (5.5.4) and (5.5.7)-(5.5.9) are, in fact, frequentist statements about the asymptotic behavior of certain function-valued statistics. For instance, the Hardy-Weinberg parameter (} has a clear biological interpretation and is the parameter for the experiment described in Example 2.1.4. Similarly, economists often work with median housing prices, that is, the parameter v that has half of the population prices on either side (formally, v is any value such that P(X < v) > . Suppose our primary interest in an estimation type of problem is to give an upper bound for the param eter v. Gray, U. See also Problem 2.4.7. The coordinate ascent algorithm can be slow if the contours in Figure 2.4.1 are not close to spherical It can be speeded up at the cost of further computation by Newton's method, which we now sketch. The improvement in speed may however be spurious since A -l is costly to compute if d is large-though the same trick as in computing least squares estimates can be used. 'III. For instance, R(B o o,) - 0(0.3) + 10(0.7) = 7 - 12(0.6) + 1 (0.4) = 7.6. If 8 is finite and has k members, we can represent the whole risk function of a procedure 0 by a point in k-dimensional Euclidean space, (R(01, 8), . A linear subspace £ of H is closed iff hn E l for all n, hn -++ h h E C., Bk vary freely, or Bi = Bio (known) i = 1, . Such densities will be considered further in Section Upon combining (B,22) and (B,23) we see that for y E g(S), py (y) If o B.4. X is a random variable (k = 1), Px(g - 1 (Y) 'JJ. (1.53) By (B.Ll) and (152), for O E S;, Po[X = x; IT = I;] P8[X = x; , T = t;] /P8(T = 1;) p(x; , B) Po[T = I;] g(t;, O)h(x;) if T(x;) = t; Po[T = t;] (L5.4) 0 if T(x;) = t; Po[T = t;] (L5.4) 0 if T(x;) = t; Po[T = t;]
(L5.4) 0 if T(x;) = t; Po[T = t;] (L5.4) 0 if T(x;) = t; Po[T = apply to a broader class of models. 9 Likelihood Ratio Proced u res 259 a5 gives the maximum of p(x, B) log .\(x) , which thus equals By Theorem 2.3. 1 , equivalent to 🕏 log .\(x) l og p(x, B) - for B E 80 . I. These results lead to a necessary condition for existence of the MLE in curved exponential families but without a guarantee of unicity or sufficiency If X1 . E(Y, Yi) where = = (Yi , f2)T = (eu\eu2)T 9. , yp), a9 a9 (xo,Yo) = 0, {) (xo , Yo) = aYj Xi 200 Measures of Performance Chapter 3 and for all l < i, a , b 0 0, (o, Yo) < aXaaXb a YcaYd 1 < j,c, d < p. 11 Var i = I Var 0 !1 (B) is often referred to as the information contained in one observation. 80) or Problem B.9.3. 'Section 8.10 Topics in Matrix Theory and Elementary Hilbert Space Theory We conclude with bounds for tails of distributions. Evidently, there is a substantial overlap between the two classes of estimates. We want to estimate 8. 1 0.3 by taking .! C • P diag(>-1, . Thus, the : f1 > fio (with fio = 0) is suggested. Consider the Hardy-Weinberg model with the six genotypes given in Problem 2.1.15. These techniques often have a strong intrinsic computational component. J., "Using Residuals Robustly 6, 266--291 (1978). PORT, AND C. Thus, log A(x) and (nl2) log(! + r;l(n - !)) for Tn > 0, where Tn is the statistic. the Wald and Rao statistics and associated confidence regions, and some parailels to the optimality theory and comparisons of Bayes and frequentist procedures given in the univariate case in Chapter 5. See Problem 3.5.13. S) = This minimum distance principle is essentially what underlies p = P[AA], N (y > 0). Goals, and Performance Criteria R(B2,6,) I 10.7 • 3.45. Thus, the MLE of w is by definition (b) Let 1' = {Po : WMLE = arg sup sup {Lx(O) : 9 E 8(w)}. does not depend on x. 1 Section 6.7 8. Assoc., 69, 909-927 (1 974). Now use Problem B.2.4. 11., X(n1), where X(, 1 = (X(i) - X)ju, is "irrel evant" (ancillary) for (J.L, a2). (J.J.l) if max(n-N(l-8), 0) < k < (J.L, a) = (X(i) - X)ju, is "irrel evant" (ancillary) for (J.L min(N8, n). L., A Course in Probability Theory, Vol., m} is a 2n-parameter canonical exponential family with 'lji = Jti/al, 1Jn+i = -1 /2af, i = 1, . The 0 -1 nature makes it resemble a testing loss function and, as we shall see in Chapter 4, the connection is close. Using our first three examples for illustrative purposes. T(X), P(B), n + 1, then, by Problem B.2. Thus, 6, equals the likelihood ratio test Bo, in fact. (3) Theorem 5.4.5. Suppose the conditions of Theorem 5.4.5. Suppose the co , eo + Jn) > dn (a,eo) < Pe, +)n I; log p(X e0) i=1 1, -l- en (a, eo)] of Theorems 5.4.4 and 5.4.5 in the future will be referred to as a Wald test. Let T = X, of testing H : fL = fLO when X1 , then C = { (, p : [t - p[S I, ,], i+I(Y) 't=1 x + I(Y) 't Distributions B.7 Convergence for Random Vectors: 0p and ap Notation 511 B.8 Multivariate Calculus 516 B.9 518 Convexity and Inequalities B.IO Topics in Matrix Theory and Elementary Hilbert Space Theory 519 B.IO.I Symmetric Matrices 519 B.IO. A government expert wants to predict the amount of heating oil needed next winter. , p. ' • " • (3.4.4). is 1 - a. Suppose $e = \{80, B\}$, $A = \{0, I\}$, and that the model is regular. We usually try to combine parameters of interest and nuisance parameters of interest (B.S. II). For detailed discussion we refer to Little and Rubin (1987) and MacLachlan and Krishnan (1997). ' Example 3.4.1. Unbiased Estimates in Survey Sampling. • 9, We are given a regular model with 9 finite. 1 1 I i ' Section 4.1 217 Introduction There is an important class of situations in which, even though there are just two actions, we can attach, even nominally, numbers to the two losses that are not equal and/or depend on 0. To test whether the new treatment is beneficial we test K : .6. As in Example 4.4. 1, let X1, . (4.5.2) These tests correspond to different hypotheses, O{X, J.lo} being of size pothesis H : J] = flo- a only for the hy Conversely, by starting with the test (4.5.2) we obtain the confidence interval (4.5.1) by finding the set of J1 where J(X, J1) * 0. INTRODUCTION AND EXAMPLES Tests of Goodness of Fit to Parameteric Hypotheses Regular Parameters. See Problem 1.1.8. On the basis of subject matter knowledge and/or convenience it is usually postulated that (2) I' (z) I' (z) (3, z) where g is known except for a vector (3, (3, 1, - - 19. It has the advantage that there is no trimming proportion a that needs to be subjectively specified. - p.0. Show that EPV(8) = if!(-.Jii8/...;2"), where if! denotes the standard normal distribution., k, contains an open !. or density function. In this framework R(B, J) is just E[I(O, J(X)) I (} 8], the expected loss, if we use ea-;ier to analyze than admissibility. Consider a population with () mem bers labeled consecutively from I to B. Use (A.2.7). Volume I presents fundamental, classical statistical concepts at the doctorate level without using measure theory. New York: J. DOKSUM, "Empirical Bayesibility." Procedures for a Change Point Problem with Application to HIV/AIDS Data," Empirical Bayes and Likelihood Inference, 67-79, Editors: S. Using a pivot based on Er 1 (Xi - X) 2 • 16.52, n • a) UCB for u2? A gambler observing a 🗇 🗣 -= c==-===='--- Section 4.10 273 Problems and Complements · - - four numbers are equally likely to occur (i.e., with probability .17). For in stance, jame in which a single die is tossed repeatedly gets the impres sion that 6 comes up about 18% of the time, 5 about 14% of the time, whereas the other , I i ! , ' j'1iI1i•I!I1�:-suppose we want to see if a drug induces sleep. Find a statistic T(X, Y) and a critical value c such that if we use the classification rule, (X, Y) belongs to population 1 if T > c, then the maximum of the two probabilities of misclassification rule, (X, Y) belongs to population 1 if T > c, then the maximum of the two probability that Sn exceeds its expected value np by more than a multiple nc of n tends to zero exponentially fast as n OCI. 98 for f) 20. SAMAROV, "Nonparametric Estimation of Global Functionals and a Measure of the Explanatory Power of Covariates in Regression." Ann. In fact, even in the classical regression model with design matrix ZD of full rank the formula (2. Under conditions AO-A6 for (a) and A0-A6 with A6 for i! Φ 'l for (b) establish that (a) [- Φ D2 ln(8n)]-l is a consistent estimate of J-1 (00). A = {{1, 2, 3}, (1, 3, 2), {2, 1, 3}, (2, 3, 1), {3, 1, 2}, (3, 2, !)}. Using 0 means that if X = x is observed, the statistician takes action o(x). (ii) The distribution of the number of occurrences of the "event" depends only on the length
of the time for which we observe the process., B) - p(X;, Bo)} : B E K n {I : 18 - Bol > . . HUBER, P., Robust Statistics New York: Wiley, 1981. as a test statistic . The test we derive will still have desirable properties in an approximate sense to be discussed in Chapter if the normality as sumption is not satisfied., Tk and h(x) = II + 1 1 [xi E { 1, . n (A.15.7) 470 A Review of Basic Probability Theory Upon taking the Xi to Appendix A be indicators of binomial trials, we obtain (A. As a consequence the emphasis of statistical theory has shifted away from the small sample optimality results that were a major theme of our book in a number of directions: (1) Methods for inference based on larger numbers of observations and minimal assumptions-asymptotic methods in non- and semiparametric models, models with "infinite" number of parameterfamily, a is a scale parameterfamily, a is a scale parameter, and Y is said to generate Fs. By definition, for any a > 0, X F; {...} Xja F. By convention. We usually need to postulate more. Chapters 1-4 develop the basic principles and examples of statistics. • ((n - 1)s 2 + b)lxa+n (a:) (1 - a:) upper credible bound for a-2. By convention 1 - P [type II error] is usually considered. We illustrate these ideas in the following example. However, our focus and order of presentation have changed. Bickel University of California Kjell A. It is not true in general that X1 and X2 that satisfy (A. We assume that we know the joint probability distribution of a random vector (or vari Z Y. B 0 or 8 > 0 for some parameter 8. Also new is a section relating Bayesian and frequentist inference via the Bernstein von Mises theorem. One-Sided Testsfor Scale. 1 0.17) • • the fonnula for obtaining the fitted value vector Y = (Y1, . J0.3HA. X,j) least twice. ZN (f3) of {3 (/31, . Show that Un * 0 but Un + 0. The link for us are things we can compute, statistics. (2) Such functions (A.8. Show that Un * 0 but Un + 0. The link for us are things we can compute, statistics. for H D(Y, j1,0) =: : IL E wo is just i nf { D(Y, IL) : 1L E where j1,0 is the MLE of IL in wo ., X11 • Let Y Y (X) denote a predictor based on X (X1, . N(Bo, r.f) densities, compute the where $\hat{\Psi}$ E A and N E {1,2, . Problems for Section 6.2 1. Let g(z) = E(Y | Z = z). For instance, in Example 1.1.2, if the errors are normally distributed with known variance a-5 (Problem 6.5.4). P is - le $\hat{\Psi}$ 5.4.5 continue to hold if ''' • ' I j'; is replaced by the likelihood ratio statistic 'I 1 p(X;, Oo + -;)n) $(X_{2}, Oo + -;)$, U_{2}, V_{2}, V Probability and Analysis Aj are real, unique up to labeling, and are the eigenvalues of A. Testing and Confidence Regions. Then for B, ., P(B) is differentiable and B ()] '> [1/]' (T(X)) ' - I(B) Theorem 3.4.1. (Information Inequality). 15. Suppose X⁽²⁾, . Hint: Using the notation of Section B.3, E(Xr) = E(Q'') = (m/kY E(Vr)E(W-r), where V rv x⁽²⁾ and W t''>.] x!. 2.4 ALGORITH MIC ISSUES As we have seen, even in the context of canonical multiparameter exponential families, such as the two-parameter gamma, MLEs may not be given explicitly by formulae but only implicitly as the solutions of systems of nonlinear equations. that ' 8* is a uniformly most accurate level LCB for B, if and only if -8* is a • Here we have relabeled 0.11 + 0.12, 0.11 + 0.21 as ry 1, ry 2 to indicate that these are parameters, which vary freely. Let J,(B) then [) (J', P.,P')) (X > n T) e((B Var I = 11 nl, e) Proposition 3.4.2. f(, B E 8, E (f. X is distributed according to the exponential family p(x, B) = exp{1/(2Ntn(x) + N2n(x)) - A{7J}}h(x) where 1) = log (P B), h(x) 1 = 2N, (x), (2.4.22) A(1] = 2nlog(1 + e'') and N; n = L: (3.10, 1) (b) A is symmetric positive definite (spd) iff C above is nonsingular. Next, an experiment is conducted to obtain information about B resulting in the randomvariable X with possible values coded as 0, 1, and frequency function p(x, 8), i = 1, 2 Rock formation X e, 02 (Oil) (No oil) 0 I o.3 o.7 0.4 0.6 Thus, X may represent a certain geological formation, and when there is oil, it is known that formation 0 occurs with frequency 0.3 and formation 1 with frequency 0. The problem of finding prediction intervals is similar to finding confidence intervals using a pivot: Example 4.8.1. The (Student) t Prediction Interval. In Example 1.1.2 IO estimate J1 do we use the mean of 🏶 L: 🗣 1 Xi, or the median, defined as any value such that half the measurements, X the Xi are at least as large and half no bigger? We may have other goals as illustrated by the next two examples. Ina Bayesian model where X 1, 13., k and a2 is unknown. Find the sample size needed for a level 0.01 test to have power at least 0.95 at the alternative value 1/>.1 = 1 5 Use the normal approximation to the (c) Suppose "•! j J! i i i. Show that Lq Lp (a) if p < q, then Zn � Z => Zn � Z. (1) As an example we calculate E(Y1 I Z) where Y1 and Z are given in Example B. Accepting this model provisionally, what does the 213 214 Testing and Confidence Regions Chapter 4 hypothesis of no sex bias correspond to? For instance, consider Example 1. 1 < i :S: n, if we believe l'(z) = g((3, z) we However, if we have observations we can try to estimate the function 1'0. For instance, and then plug our estimate of (3 into g. However, save for a brief discussion in Volume 2, the conceptual issues of stationarity, ergodicity, and the associated probability theory models and inference for dependent data are beyond the scope of this book. If g is a one-to-one affine transformatioll as defined earlier, then Y = q(X) has density Corollary (8.2.6) ' for $y \in q(S)$, where det A is the determinant of A. Here $R(B1, 64) > R(8_1, 65)$ but $R(B_1, 64) > R(8_1, 64)$ but $R(B_1, 64) >$ asymptotic results should be checked by Monte Carlo simulations. Pis referred to as the model. p(en I 'n-d f(e 1) f(c, - f3c 1). L: V = Cov(X, E(Y I X)). Identify follows. V(A(y)) differ., Xn) is the same as that of (Xip. This result may be derived by using (A.12.5) and (A.12.6) in conjunction with (A.13.4). EY2 < oo if and only if E(Y - c)2 < implies that p. For instance, in Example 1.1.2, we can ensure independence and identical distribution of the observa tions by using different, equally trained observers with no knowledge of each other's find ings. what choice of c would make be have size exactly (c) Draw a rough graph of the power function of be (d) How large should n be so that the be (e) If in a sample of size n = 20, Mn = specified in (b) when specified in (b) when specified in (b) has power n = 0. "While other general statistics texts at a similar level touch on some of the topics covered in this book, none of them cover the modern material in this book with comparable depth., Xn) is a sample from N(J 0. TABLE 1.3.3. Possible decision rules oi(x) • ' I x=O x=1 a, a, 2 a, a, 3 a, a, 4 a, a, 5 a, a, 8 a, a, 7 a, a, 6 a, a, 9 a, a, 7 a, a a1. and show that T can be written in the form T 🕏
L/M where L 🕏 2 f 🕏 and M2 🅏 2). Then c,(a,Bo) = Bo + Zt-a/Vn1(8o) + o(n-112) where Zt-a is the 1 - a quantile of the N(O, 1) distribution. If h1 j hz, then (8.10.12) I An interesting consequence is the inequality valid for all h1, hz, i (8.10.13) In R2 (8. We conclude this section with two simple limit theorems that lead to approximations of one classical distribution by another. Let $P = \{Po : B \in 8\}$ where Po is discrete and concepts developed in this second volume students will be ready for advanced research in modem statistics. Thus, X has an hypergeometric, 1t(N8, N, n) distribution. Un Un, 9. Show that if (4.4.3). ' E!Zn - ZIP 2: EfiZn - ZIP 2: EfiZn - ZIP 2: EfiZn - ZIP 2: EfiZn - ZI 2: E). 11.22) (i.e., are uncorrelated) need be independent. n-1 🕏 ooc (3.4.2) 1 Because for unbiased estimates for all 0, UMVU (uniformly minimum variance unbiased). 6. One way of making the notion "a statistic whose use involves no loss of information., 17'[:) where flj has dimension dJ and L;=I dJ = k and the problem of obtaining ij1(to, 71,; j ¥ l) can be solved in closed form. would be obtained where 6. To get expressions for the MLEs of 7r and JL, recall from Example 1 .6.8 that the inverse of the logit transform g is the logistic distribution function Thus, the MLE of 1ri is Jfi = g -1 (L: $\mathbf{\Phi}$ =I Xij(jj). (A.I3.3), Higher-order moments may be computed from the moment generating function Mx (t) (1 - (1 - 0)]". n = 25 I Tn 1 - Because Tn has a T a critical value is the conditional distribution of X given (J 8. However, S(X) is exactly what is needed to estimate the . We begin with the bisection and coordinate ascent methods, which give a complete though slow solution to finding MLEs in the canonical exponential families covered by Theorem 2.3.1. d, d(d 2.4.1 The Method of Bisection The bisection method is the essential ingredient in the coordinate ascent algorithm that yields MLEs in k-parameter exponential families. (For a recent review of expected p values see Sackrowitz and Samuel-Cabo, 1999.) - ': I = Problems for Section 4.2 1. Example 4.1.3 (continued). VZ I I 2 3 (x) -). Thus, the critical value for testing H : u = uo versus K : u < uo and rejecting H if Tn is small, is the a percentile of X+. It is evident from the argument of Example 4.3.3 that this test is UMP for H : u > uo versus K : u < uo among all tests 0 depending on U2 only. The likelihood equation (6.4.3) becomes n1 (2 + ry) (n2 + n3) n4 + (1 - ry) -: ry 0, (6.4.4) which reduces to a quadratic equation in if. We then sketch the extension to functions of vector means. In this section we assume that the data are grouped or replicated so that for each fixed i, we observe the number of successes Xi = E";;; I Yij where Yij is the response on the jth of the ffii trials in block i, 1 :::; i :::; k. Fonnally, if the measurements are scalar, we observe x 1, , Xn, which are modeled as realizations of X1, ... These are the likelihood ratio test and the score or Rao test. This process of combining a limit theorem with empirical investigations is applicable in many statistical situations where the distributions of transformations g(x) (see A.8.6) of interest become progressively more difficult to compute as the sample size increases and yet tend to stabilize. distribution of Y given Z and Z given Y. (8 10.2 1).) Xn) is the critical function of the Wald test and OL n (X1 '. P[i = N] = rr,, i = 0, . ' 464 Then if A Review of Basic Probability Theory Appendix A X "' F1, ,c-. Families," Ann. As we have seen, in examples such as 1. In this manner, we obtain one or more quantitative or qualitative measures of efficacy from each experiment. + 3. (a) Show that n(X - 1'1) as the first approximation to the maximum likelihood 156 Methods of Estimation (c) If n = 5, x - = Chapter 2 - 2, find (} a construction to the maximum likelihood 156 Methods of Estimation (c) If n = 5, x - = Chapter 2 - 2, find (} a construction (c) If n = 5, x - = Chapter 2 - 2, find (} a cons have a test statistic T and use critical value c, our test has size a(c) given by $a(c) \otimes sup(Pe[T(X) > c] : 8 E 80) \cdot (4. 14)$ where X has a B(n,p) distribution. When v = v, the game is said to have a value v. Fortunately in these cases the algorithm, as it should, refuses to converge (in 11 space!)-see Problem 2.4.2. We note some important generalizations. mk \bullet oo and H : f] E w0 is true then the law of the statistic of (6.4.18) tends to X2r-q. Hint: (X; - J',) / /m;K;(l - 1r;), I < i < k are independent, asymptotically N(O, 1). Introduction, "Ann. Models such as that of Example 1.1.2 with assumptions (1) -(3) are called semiparametric. X(n)), {Pe Xi, (X1, . Show that if X1, . Such tests are said to have level (of significance) a, and we speak of rejecting H at level o:. N(iJ., a2) and we formally put 7C(iJ., a) = ,; , then the posterior density !f"(J-t I x, s2) of J1. Similarly, if 8 c R' and 71(8) = Bk xe8 c Rk, then the resulting submodel of P above is a submodel of P above is Theorem is orthogonal to h2 iff (h1, h2) = 0. End Lemma 2.4.1. The bisection algorithm stops at a solution Xfinal such that l xfinal - x' l: S 17 1 r?. 305 NONDECISION THEORETIC CRITERIA • In practice, even if the loss function and model are well specified, features other than the risk function are also of importance in selection of a procedure. l6). A reasonable formulation of a model in which the possibility of gross errors is acknowledged is to make the ci still i.i.d. but with common distribution function F and density f of the form f(x) () P + >.h(x). For a sample of size n + 1 from a continuous distribution we show how the order statistics can be used to give a distribution-free prediction inter val. Regularity Conditions are Needed for the Information Inequality. 31. Initialize x ;; ld (x, , xold = xo. (b) Define the expected p-value as EPV(8) = EeU. Similar problems abound in every field. I Xn). By Theorem 1.5.1. X(n) is a sufficient statistic for 0. Asymptotic Approximations. CRUTCHFIELD, Statlab: An Empirical introduction to Statis tics New York: McGraw-Hill, 1975. The same issue arises when we are interested in a confidence interval (X) v(X) for v defined by the requirement that = - P[v(X) < v(P) < v(X)] > 1 - a • · I for all P E P. For instance, suppose X1 , . In the "mixed" cases such as (} continuous X discrete, the joint distribution is neither continuous nor discrete. (2) Matrices of scalars and/or characters, for example, digitized pictures or more rou tinely measurements of covariates and response on a set of 1.1.4 and Sections 2.2.1 and 6 .1. Proposition 3.4.1. ljp(x, B) • h(x) exp{ ($\hat{\Psi}(B)T(x) - B(B)$ } is an exponentialfamily and TJ(B) has a nonvanishing continuous derivative on e, then I and II hold. and Y has a firSt moment different from 0, we may without Joss of = 1 and, hence, if X F;;, then E(X) = rr. i • i and 0 with probability i and V is independent of (c) Obtain the limit distribution, then . We know that given Z = z, Y has a hypergeometric, 'H(z, N, n), distribution. Now 'Pk is MP size a. 1 9 315 2.68 250 2.64 298 2.37 384 2.61 310 2.12 337 1.94 Using the likelihood ratio test for the bivariate normal model, can you conclude at the I 0% level of significance that blood cholest or the bivariate normal model, can you conclude at the I 0% level of significance that blood ratio test for the bivariate normal model, can you conclude at the I 0% level of significance that blood cholest or the bivariate normal model, can you conclude at the I 0% level of significance that blood ratio test for the bivariate normal model, can you conclude at the I 0% level of significance that blood cholest or the bivariate normal model of the bivariate normal model of the bivariate normal model of test for the bivariate normal model of the bi Admit Men Women Admit Deny 235 1 " 3,8 -- * 35 270 45 7 Men Women 273 42 n 315 = three Deny 122 103 225 1 62 n = 387 93 69 215 172 (d) Relate your results to the phenomenon discussed in (a), (b). In the preceding discussion we used the fact that Z (!') = ., fii(X - I')/ tn-1 (1 - *a). Consider a population with three kinds of individuals labeled

S(B, o) is the o ball about B. Hint: Consider (U1, U2) defined by (B.4.19) and (B.4.22). (ii) If Y is a function of Z, h(Z), then the conditional distribution of Y, h(Z), then the conditional distribution of Y, h(Z), then the conditional distribution of Y is degenerate, Y \clubsuit h(Z) with probability I. X3) + be Cov(X2, X,) + ad Cov(X1, X4) (A. Thus, we will need to ensure that our parametrizations are identifiable, that is, lh i 02 ==> Po1 i= Pe2•I1''1 •••'••••, $\bullet • I''', j'j$ i 1 Section 1.) Xn) is the critical function of the LR test then, for all "(, . (1.7.1) Equivalently, Sv(t) = sg.(t) with C. For instance, a method of moments estimate of (.\,p) in Example 2.3.2 is given by where Q-2 is the empirical variance (Problem 2.2.11). i I I This default model is also frequently postulated for measurements taken on units ob tained by random sampling from populations, for instance, heights of individuals or log incomes. if X 0; take action a2, if X = 1," and so on. By A4(a.s.), \clubsuit n a.s. Using the strong law of large numbers we obtain (Problem Po [dnqn(t) \$1(B)exp {Eo :;:(Xr,B)\$} (5.5. 2). The geometry of the bivariate normal surface. If H is rejected, it is natural to carry the comparison of A and B further by asking whether 8 < 0 orB > 0. The family of such level o: likelihood ratio tests obtained by varying e10 can also be inverted and yield confidence regions for e1 . = 1 - - 2). Subject Index. This is written h 1 notion of orthogonality in Euclidean space. 1 is used for every [Lo we see that we have, in fact, generated a family of level a tests { [(X, 11]] where 1 if vnY, vo = up, then {8(w) : w E !1} is a partition of 8, and 8(w) Let Hint: 9 belongs to only one member of this partition, say 8(w). (x0, yo) is a saddle point of g if q(xo, Yo) = up. g(x,yo) = infg(xo, y). Here are two examples of testing hypotheses in a nonparametric context in which the minimum distance principle is applied and calculation of a critical value is straightforward. = /.00 e, p(x, O)d1r(O)j The left-hand side equals j'' -oo of all Bayes tests is = lh where loss, every Bayes test for > . One reasonable choice for k is k 2), P(IXI > 3) and P(IXI + 4) for the normal, Laplace. (5.4.44) But Polya's theorem (A.14.22) guarantees that su p which implies that other hand, IPo,[.fii(Bn - Bo) > z] - (1 - 1>(z)) + 0 (5.4.45) + 0 (5. note that if A is positive definite, A is nonsingular (Problem 8.10.1). Us = I{U E (hm, . For future reference we note that a statistic T is known, the sample X = given that P is valid. we obtain E(X) = JJ. Var X = a] < >-- aI - g(a) . The gen erable X definition of parameters and statistics is given and the connection between parameters and parameters denote the observed value t = T(x) of T(X) for the datum x, let 244 Testing and Confidence Regions Chapter 4 o: (t, Oo) denote the p-value for the UMP size a test of H : () > Oo, and let $A \bullet (B) = T(A(B)) = \{T(x) : x \in A(B)\}$. allele locus, Xi = (E; Jo ξ ; 2, ξ ; 3), where P9[X = (1, 0, 0)] = B2, Po [X = (0, 1, 0)] = 28(1 - B), Po[X (0, 0, 1)] = (I - B)2, 0 < B < 1. Summary. 'I • 8.10.2 Order on Symmetric A < B iff B - A is nonnegative definite. Therefore, E(l(O, o'(X)) 1 x] > E(l(O, o'(X)) 1 x], 0 and the result follows from (3.2.9). To investigate this question we would have to perform a random experiment. Far more important than the choice of loss function l : P x A R+. I' i ! Section 1.3 17 The Decision Theoretic Framework Xin as our estimate or ignore the data and usc hiswrical infonnation on past shipments, or combine them in some way? When P is not an exponential family both existence and unicity of MLEs become more problematic. We shall go into this further in Chapter 10 of the first edition at this stage. In this example we bave assumed that IJo and IJ, for the two populations are known. However, what reasonable means is connected to the choice of the parameter we are estimating (or testing hypotheses about). Nevertheless. Let X1, . 1 2, U1, . Un+1 are i.i.d. uniform, U(O, 1). The case we have I Section 2.4 131 Algorithmic Issues just discussed has d1 = \cdots = dr = 1, r = k. To be a bit formal, suppose that if n measurements X* = (Xi, . • ". Because 0"1 ----) 0, (n 0" 1 I 0") ----) 1, and X. !:.., e as n ----) oo , l; j • ' (B.I.2) if the Yi are all 0 or 1 and Eyi = z. SCHLAIFFER, Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961. If n = i � 0 and the information bound is infinite. Show that, under the conditions of (c), ... jii(X - 0) !:. An interesting feature of the preceding example is that the test defined by (4.2.6) that is MP for a specified signal v does not depend on v: The same test maximizes the power for all possible signals v > 0. Now [Y B • YB] is said to be a level (1 - a) Bayesian prediction interval for Y Xn + 1 if = Section 4 . The three principal issues we discuss are the speed and numerical stability of the method of computation used to obtain the procedure, interpretability of the procedure, and robustness to model departures. To obtain the predictive distribution, note that given X = t, Xn+1 - 9 and 9 are still uncorrelated and independent. The argument of Example 2.3.3 can be applied to determine existence in cases for which (2.3.3) does not have a closed-form solution as in Example 1.6.8-see Problem 2.3.1 and Haberman (1974). Describe in detail what the coordinate ascent algorithm does in estimation of the regres sion coefficients in the Gaussian linear model Y 2DJ3 + <, rank(ZD) = k, 1) = I \ B(1 0) [1 - (1 - B)"] (x - nB - x(1 - B)"] - [1 - (1 - B)"] - [1 - (1 - B)"] (x - nB - x(1 - B)"] - [1 - (1 - B)" theory of maximum likelihood estimates, and the structure of both Bayes and admissible solutions in decision theory. Let Z .N(O, 1). Say there is a closed-form MLE or at least lp,x(B) is concave in B. The following important example illustrates, among other things, that the UMP test phenomenon is largely a feature of one-dimensional parameter problems. By a good mathematics background we mean linear algebra and matrix theory and advanced calculus (but no measure theory). 11.23) i=l References Gnedenko (1967) Chapter 5; Chapter 5; Chapter 5, Sections 1-4 Pitman (1993) Section 6.4 A.12 MOMENT AND CUMULANT GEN ERATING FUNCTIONS A.12.1 If E(e" I X I) < oo for some s0 > 0, Mx (s) = E(e'X) is well defined for lsi S so and is called the moment generating function of X. This will be done in Sections 3.5.3 and 6.6. Experiments in medicine and the social sciences often pose particular difficulties. GIRSHICK, Theory of Games and Statistical Decisions New York: Wiley, 1954. Chou, G. Data can consist of: (1) Vectors of scalars, measurements, and/or characters, for example, a single time series of measurements. Fix t and divide [0, t] into n intervals [0, tfn], (t/n, 2tfn], . 1969. Suppose that (Z1, Yi), . . , /3p)T is, if N 2:::: 7=1 m i, The log likelihood l (1r({3)) p k k (6.4. 1 2) g((f3)) $\{3 \}$ N m; lo l + exp{z; } + log ';; f]; T; The special case p Zi = = = where Tj =) (= 0 (= 0 RP, p > I, be a family regarding these hypotheses provide conclusions regarding the set of the same data con of models for X E X c R v k Let q be a map from 8 onto 1, 1 c R, I k k < p. E. 7). 0.25 and net = 0.25 and (n - k is an integer, use (3.5.5) to plot the sensitivity curve of the l) a is an integer. Such a function can usually be found if a depends only on J-l, which varies freely. , k] with canonical parameter 1 and £ = Rk-l More over, the parameters 'I] = log(Pry [X = j]/Pry[X = k]), 1 < j < k - 1, are identifiable. a (b) Suppose X1, It can be shown that (Problem 6.3.8) E(Oo) = I, (Oo) - !2 1 (90)I!1 (90)I!2 (90) (6.3.21) where I11 is the upper right 6.3.9) under A0-A6 and consistency (9 x d block, and so on. 7 = 7 = 1.57). We derive thetest for H.,.0 • Formally, let Pv0 = {P : v(P) • v0 : vo E V}. Show in Example 1.2.1 that the conditional distribution of 6 given I:; 1 Xi = k agrees with the posterior distribution of 6 given X1 = X t, . The algorithm was fonnalized with many examples in Dempster, Laird, and Rubin (1977), though an earlier general form goes back to Baum, Petrie, Soules, and Weiss (1970). Third Berkeley Symposium on Math. Var X and Var X continues to hold if ";.,5 11 > k, even for sampling without replacement in each stratum. Situation (b) can be thought of as a generalization of (a) in that a quantitative measure is taken rather than simply recording "defective" or not. Hint: (a) Consider the likelihood as a function of Tfit. 3. • ' (b) If no assumptions are made about h, then p, is not identifiable. • • • • large. D Next, in the one-sample situation, let h(X) be an estimate of h(J.L) where h is con tinuously differentiable at I' hl 1 > (!') i 0. Riesz. , Zn - Z) are independent. (5.3.2) where [X' - I'f < [X - J'f. There is a substantial number of statisticians who feel that it is always reasonable, and indeed necessary, to think of the true value of the parameter (J as being the realization of a random variable 8 with a known distribution. (c) The following 2 x 2 tables classify applicants for graduate study in different depart ments of the university according to admission status and sex. (d) Find or write a computer program that carries out the Welch test. , 0 Theorem 5.4.4 tells us that the test under discussion is consistent and that for n large the power function of the test rises steeply to 1 to the right of 80. In practice, if we need the distribution of Sn we try to calculate it exactly for small values of n and then observe empirically when the approx imation can be used with safety. I (b) Show that 812/(811 + 81 2) • 82 1/(821 + B,) iff R1 and C, are independent. = 4.3 U N I FORMLY MOST POWERFUL TESTS ANO MONOTONE L I K E LIHOOD RATIO MODELS We saw in the two Gaussian examples of Section 4.2 that UMP tests for one-dimensional parameter problems exist. - 15. We return to conjugate families in Section 1.6. Summary. , XN } (3.4.3) 0 otherwise. The likelihood equations are equivalent to (Problem 2.3.2(b)) r' - r(j)) - log A = log X - p + X A (2.3.4) (2.3.5) where log X = 2.2.6, if 2n1 + n, = 0, the MLE does not exist if 8 = (0, 1), whereas if B = [0, 1] it does exist D It does exist D and A6 hold for .p A and A6 hold for .p (}. Common sense indicates that to get information about B. the second says that if Appendix A Nn (t) f N (t) 1 there must have been a multiple occurrence in a small subinterval, the third is just (A.2.5), and the remaining identities follow from (b) and (d). In Example 1.1.1 do we use the observed fraction of defectives '' • I '' i' i. FREEMAN "Estimation and Inference by Compact Coding (With Discussions)," J. They are not covered under our general model and will not be treated in this book. For the purpose of modeling, imagine a sequence X1, X2, of i.i.d. survival times with distribution F0 • Let N be a zero-truncated Poisson, P()..), random variable, which is independent of X1 J X2, . If a = quartile 4. Then Snj - np 🖗 Z, .np(! - p) Z has has a (A. Wiley & Sons, New York, 1954. Exhibit the null distribution of 2 log .\(Xi, Yi : 1 < i < n)... + (log a2)] - 🌾 { (log 2 7T) + (log a5)] 🔅 } - { (log a2)] - (lo ...!... This is obviously overkill but suppose that, in the study, drugs A and B are given at several 10 Statistical Models, Goals, and Performance Criteria Chapter 1 dose levels. ' 1.4 • PREDICTION The prediction Example 1.3.2 presented important situations in which a vector z of co variates can be used to predict an unseen response Y. H., Optimal Statistical Decisions New York: McGraw-Hill, 1974., Xn from this population, let N1 N2, and N3 denote the number of Xj equal to I, 2, and 3, respectively., 8v}, Example 3.2.2. Bayes Procedures Whfn 8 and A Are Finite. If X is discrete, "' P = X Using the fact that g-1 is approximately linear on A(y), it is not hard to show that V(g-l(A(y))) ! J•-' (y) 1. ' " i (8.2.7) Example B.2.1 is a special case of the corollary. - a 2 + n2) (2n3 + n2)) 2, (2n1 + n2) (2n3 2, 2n T 2n 0 Example 6.4.5. The Fisher Linkage Model. Unfortunately strict concavity of lx is not inherited by curved exponential families, and unicity can be lost-take c not one-to-one for instance. , ' We claim that (5.5. 12) for all 8. We next show that for a certain notion of accuracy of the bounds, which is connected to the power of the associated one-sided tests, optimality of the tests translates into accuracy of the bounds. Thus, from D (X, ji) K: (6.4. 1 1) Xdmi and Xi. We find such an interval by noting ,! and solving the inequality inside the probability for p.. That is, , a(k) ? sup{Pe[T(X) > k]: 8 E eo} Pe, [T(X) > k]. Subject matter specialists usually have to be principal guides in model formulation. 1Ji2 = z::: 1 Oii. claim (5.4.36) is just the correlation Because Eo'!j; (X1, B) inequality and the theorem follows because equality holds iff 'ljJ is a nonzero multiple a (B) o of 3i; (X1, B). Establish 5.3.28. Using A6 we obtain] for all t 1 and AS, (5.5.14) \diamond 1. P \diamond (Zit \diamond Z,)6t \diamond " \diamond [J, Z, j=2 \diamond (!) 2 ! " • I: \diamond 1 (Yi \diamond ' 2 • Similarly compute the information matrix when the model is written as • y; = {J, (Z" where {31, 12, I' j=1 and the c; do not depend on {3. Find the MP test for testing H : {} = 1 versus K: B = 81 > 1. At the other end of the scale of difficulty for books at this level is the work of Hogg and Craig, Introduction to Mathematical Statistics, 3rd ed. 53 and p. n2 n 14. • 1 (28') r) (Y' exp 28' • . We begin by fonnalizing what we mean by "a reduction of the data" Chapter 1 Statistical Models, Goals, and Performance Criteria 42 loses information about the labels of the Even T I Transferred by the idea of sufficiency is to reduce the data with statistics whose use involves no loss of : (} E 8 }. (c) Show that the MLEs exist iff 0 < Na+c • N+t c/A, N+bc - N++c/B, Nt-t !b!£ = =a',b',c' obtained by fixing the "b, c" and "a, c" parameters. There exists a compact set K c e such that l(ll) < c for all ll not in K. A one-dimensional problem 2.4.13. < I. 12.9) c 7 = c;(X) = . B(B), distribution and that given Z = z, hypergeometric, 'H(z, n), distribution. Also note that the prediction interval is much wider than the confidence interval (4.4.1) . , k, exist and are unique. 746 2.120 2.235 2.583 2.921 3.252 3.686 4.015 17 0.689 1.333 1.740 2.110 2.224 2.567 2.898 3.222 3.646 3.965 18 0.688 1 .330 1.734 2.101 2.214 2.552 2.878 3.197 3.610 3.922 19 0.688 1 .328 1.729 2.093 2.205 2.539 2.861 3.174 3.579 3.883 20 0.687 1.325 1.725 2.086 2.197 2.528 2.845 3.153 3.552 3.850 21 0.686 1.323 1.721 2.080 2.189 2.518 2.819 3.1 19 3.505 3.792 4 0.727 23 0.685 1.319 1.714 2.069 2.177 2.500 2.807 3.104 3.485 3.768 24 0.685 1.31 8 1.7 1 1 2.064 2.172 2.492 2.797 3.09 3.467 3.745 25 0.684 1.316 1.708 2.060 2.167 2.485 2.787 3.078 3.450 3.725 30 0.683 1.310 1.697 2.042 2.147 2.457 2.750 3.030 3.385 3.646 40 0.681 1.303 1.684 2.02 1 2.123 2.423 2.704 2.971 3.307 3.551 50 0.679 1.299 1 .676 2.009 2.109 2.403 2.678 2.937 3.261 3.496 60 0.679 1 296 1.671 2.000 2.099 2.390 2.660 2.915 3.232 3.460 100 0.677 1 .290 1.660 1.984 2.08 1 2.364 2.626 2.871 3.174 3390 1 000 0.675 1.282 1.646 1.%2 2.056 2.330 2.581 2.813 3.098 3.300 0.674 1.282 1 .645 1 .960 2.054 2326 2.576 2.807 3.090 3.291 50% 80% 99% 99.5% 99.8% 99.9% 00 Confidence level C The entries in the top row are the probabilities of exceeding the tabled values. We stress that looking at randomized procedures is essential for these conclusions, although it usually turns out that all admissible procedures of interest are indeed nonrandomized. L(x, O,v) A point > rr/ (1 - rr) is equivalent to T > t., (Zn, Yn) from the family with density p (z, y, {3) = h(y) qo (z) exp { (zT{3)y - A0 (zT{3)} }. Here is an example to be pursued later.) that is independent of O such that "'E';' 1 p(x; O) = I for all 0., xn) = ' (1.2.9) for r(t)tk(1 - t)n - kdt In the efficiency for £ = .05, 0.10, 0.15 and 0.20 and note that X is more efficient than X for these gross error cases. J - 1, .' Therefore, $6^* = 85$ as we found previously. (Pareto density) 144 Methods of Estimation (c) f(3c,8) Chapter 2 c8".r-lc+l), :r > 8; c constant > 0: 8 > 0. (X, Y) - E(XY) (8.10.18) II XII = E l (X') . , b - 1 only. 'Un (S S J - 8f2)/(n -)S L S 3 Ut/(n u2, . If the equations - have a solution B (x) E CO, then it is the unique MLE of B. For each simulation the two samples differ in size: The second sample is two times the size of the first. ! I · 1'' 1 ' i Section B.ll 531 Problems and Complements 12. Let (X, Y) 🗞 N (1, 1, 4, 1, 4). respectively, with all parameters assumed unknown. These issues will be discussed further in Volume 2. l 2.2) esxpx (x)dx if X is continuous. Show that both , P and 🏶 are identifiable. Khintchin's (Weak) Law of Large Numbers Let {Xi}, i > 1, be a sequence of independent identically distributed random variables with finite mean J-L and define Sn = L late las a measure of the weight of evidence we attach to the falsity of H. Natural initial assumptions here are: . (b) Find the test that is best in this sense for Example 4.2.1. 5. Hint: Use the results of Theorem 1.6.2 to find the mean and variance of the optimal test statistic. (ii) 4zy, 0 < z < 1, 0 < y < 1 P(z,Y) (z, y) 0 otherwise. Y - I'2)T !:.-1 (X - I' I. (2) A vector parameters (identifiable). Then we can conclude from Theorem 6.3.3 that 2 results of Theorem 6.3.3 that 2 results of Theorem 6.3.4 that are parameters (identifiable). log >. Also, the robustness considerations of Section 3.5 favor X{k} (see Example 3.5.2). Symmetry (or invariance) restrictions are discussed in Ferguson (1967). ! i· '. The Bayesian prediction interval is derived for the normal model with a normal prior. then X has a !3(n, B) distribution (see (A.6.3)). Thu s, Zi is a d dimensional vector that gives characteristics such as sex, age, height, weight, and so on of the ith subject in a study. By our definition of conditional probability in the discrete case, it is enough to show that Po [X = x; IT = I;) is independent of 8 on each of the sets S, = $\{0 : Po[T = t_i] > 0\}$, i = 1, 2, . Contingency Thbles 403 Logistic Regression for Binary Responses 408 6.4.3 *6.5 Generalized Linear Models 411 *6.6 Robustness Properties and Semiparametric Models 417 6.7 Problems and Complements 422 6.8 Notes 438 A A REVIEW OF BASIC PROBABILITY THEORY 441 A.1 The Basic Model 441 A.2 Elementary Properties of Probability Models 443 A.3 443 Discrete Probability Models A.4 Conditional Probability and Independence 444 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Vectors: Transformations 451 A.9 Independence of Random Variables and Vectors: Transformations 451 A.9 Independence of Random Variables and Vectors: Transformations 451 A.9 Independence of Random Variables and Vectors: Transformations 451 A.9 Independence 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities on Euclidean Space 448 A.5 Compound Experiments 446 A.6 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Probabilities and Vectors 458 A.8 Bernoulli and Multinomial Trials, Sampling With and Without Replacement 447 A.7 Bernoulli and Without Replacement 447 B.7 Bernoulli A.8 Bernoulli and Without Replacement 4 454 A. In the weed only show that Po [X = XjiT = ti] is independent of e for every i and j., Xn and k is defined by P(S > k) 1 - a for a binomial, B(n, 1.52), we need only show that Po [X = XjiT = ti] is independent of e for every i and j., Xn and k is defined by P(S > k) 1 - a for a binomial, B(n, 1.52), we need only show that Po [X = XjiT = ti] is independent of e for every i and j., Xn and k is defined by P(S > k) 1 - a for a binomial, B(n, 1.52), we need only show that Po [X = XjiT = ti] is independent of e for every i and j., Xn and k is defined by P(S > k) 2 - a for a binomial, B(n, 2.52), we need only show that Po [X = XjiT = ti] is independent of e for every i and j. sources of these changes have been the exponential change in com puting speed (Moore's "law") and the development of devices (computer controlled) using novel instruments and scientific techniques (e.g., NMR tomography, gene sequencing). , -Xn. I are the test statistic D(Fx , F x) and the development of devices (computer controlled) using novel instruments and scientific techniques (e.g., NMR tomography, gene sequencing). , -Xn. situations (b) and (c). F(O, 0, 0, 0, 1, 1, p) = ! + (1/21T) sin -l p. ! ' shape'' of class F = iff F(x) when F(x) is unknown. A detailed discussion of the appropriateness of the models we shall discuss in particular situations is beyond the scope of this book, but we will introduce general model diagnostic tools in Volume 2, Chapter 1. The latter include theprincipal axis and spectral theorems for Euclidean space and the elementary theory of convex functions on elementary introduction is beta, f3(r, s). Statist., I, 538-542 (1973). Thus, we reject H if S exceeds or equals some integer, say k, and accept H otherwise. How useful a particular model is is a complex mix of how good the approximation is and how much insight it gives into drawing inferences. Because a test of level a is also of level a '> a, it is convenient to give a name to the smallest level of significance of a test. Pmod, pfld. 10.8) hold for the conditional expectation given Z = z. • x > 0, 0 > 0. A general solution of this and related problems may be found in the book by Barlow, Bartholomew, Bremner, and Brunk (1972). Show that if X (7,., then for r even and r < k. Example 3.4.3. Suppose X1, Xn is a sample from a normal distribution with unknown mean B and known variance a 2. close enough to fj, this method is known to converge to Tj at a faster rate than coordinate ascent see Dahlquist, BjOrk, and Anderson (1974). Let X denote the sample median. Thus, if x2 measures devia tions from independence, Z indicates what directions these deviations take. (b) Show that Y and X - Y are independent and find the conditional distribution of X given Y = y. From the joint distribution of Z and V, get the joint distribution of Y1 = Z/ jV7k and Y2 = V. = (3.4. 13) By (A. -1 ' Note that F(xp) • ' ' ' • = vn[F(xp) - F(xp)]. em . We denote the jth derivative of h by h(i) and assume]jh(ml]]oo = SUPx]M ml (x)] < M < oo 11 • ' . 272 Testing and Confidence Regions Chapter 4 (c) Are any of the four statistics in (a) invariant under location and scale. = Under our assumptions, The problem of finding the supremum of that p (x, B) was solved in Example 3. • 0 as h • 0.) Physically, these a "umptions may (i) The time of recurrence of the "event" is unaffected by past occurrences. Here Xq denotes the qth quantile of the X; distribution. If, say, Q is the empirical distribution of the Zj in Section 1.3 19 The Decision Theoretic Framework the training set (z 1, Y), . Find maximum likelihood estimates of J1 and a2. f3(0o) (a, (L4) 1 so = nOo + 2 + z(1 - a) InOo(J - Oo)] 1;2. Thus, i/N, X has the '''' • P[X = k, II = N l - (1 .2.2) This is an example of a Bayesian model. rate n - !, whereas the width of the prediction interval tends to 2az (1 Moreover, the confidence level of (4.4.1) is approximately correct for large n even if the sample comes from a nonnormal distributions. Thus, minimizing n 1 ' p z; /3) $2 = 2::: \bullet \bullet \bullet \bullet$ Hint: Write n 1 l (6.2.26) and (6.2.27). Show that I • (X - p, 1 Y - p, 2), (Tl 0"2 has a N(O, 0, 1, 1, p) distribution and, hence, express F(\cdot, \cdot, \bullet , ttl , J.12, af, cr \mathbf{O} , p) in terms of F(\cdot, \cdot, \bullet , 0, 0, 1, 1, p). For instance, in Example 1.1.3 even with the simplest Gaussian model it is intuitively clear and will be made precise later that, even if .6. RP. n N (-) B'(1 - B)N-'/b(y) (😻;) z-y 8' (1 - B) N- , It is plicitly we consider the case = - I ' ' ' 458 A Review of Basic Probability Theory Appendix A called the comriance of X1 and .\:2 and is written Cov(. \1-X2)., p, and Lf 0 1r i = 1. •, Xn be a sample from a population with mean J1 and variance a2 Suppose h has a second derivative h 2(Jar + a * - 2 pawz)z(1 * (c) Show that if I. If c2 > 0, the bound is sharp and is attained only if Yi X·1 > fl.. LEHMANN, Descriptive Statistics for Nonparametric Models. Let $x = x_0 < x_1]x_i - x_i - y_i = 0$ or p E PI). If we take action a when the parameter is in ea. Theorem 4.2.1. (Neyman 4.2.1) (Neyma 4.2.1) (N Pearson Lemma). t We want to test 3. zlil in RP ball about "k L..... Such statistics are called equivalent. CARROLL, R. Complete Families of Tests The Neyman-Pearson framework is based on using the 0-1 loss function. ,...., xi = 1-t + l':i, 1 0} where tP is the standard normal distribution. Note for Section 1.4 (I) Source: Hodges, Jr., J. For instance, in Example 1. Corollary 4.3.1. Suppose {Po : 0 E 6}, e c R, is an MLRfamily in T(r). 1 1. For instance, if we observe an effect in our data, to what extent can we expect the same effect more generally? oo ±oo. Thus, for example N12 is the number of sampled individuals who fall in category A of the first characteristic and category B of the second characteristic. Then it is easy to see that Theorem 2.4.2 has a generalization with cycles of length r, each of whose members can be evaluated easily. 4. To find the density p(x 1, xn). rn \cdot rn (a) Show that the matrix A defined by Y = AX is orthogonal and, thus, satisfies the require ments of Theorem B.3.2. (c) Give the joint density of (X(n), $\hat{\bullet}$, . [L., dt. Putting the bounds f!(S), O(S) together we get the confidence interval (O(S), O(S)] of level (1 - 2a). By (a) and (b), Nn (t) has a B(n, P [N(tfn) > 1]) distribution. l ' • i Section 6.7 Problems and Complements 431 $\hat{\bullet}$ (1) 8. , >-\$) in (B.10.1). Example 4.4.1, The (Student) t Interval and Bounds. Decide I' > l'c ifT > z(l - ia). We want to estimate the parameter X = estimate families drawn at random without = $(x_1, 1, WErHERILL, G, E(E(Y | Z))$ using (a). Examples are f Gaussian, and f(x) = e-x(1 + ex)-', (logistic). Show that if 02 = BB a unique MLE for Bt exists and 10. 0; a y at time t is sometimes modeled by > 0, f3 > 0, is a sample from a population with density x,B) and that the conditions of Theorem 3.4.1 hold. In Problem 1.1.1 2, we derived the model $G(y, Ll.) = 1 - [1 - F0(y)]^{"}$, y > 0, Ll. > 0. We will discuss the phenomenon further in the next section. II, ffi New York: Hafner Publishing Co., 1961, 1966. The left column gives the degrees of freedom. (a) Show that the maximum likelihood estimates of Tfit, 1Ji 2 are given by Tfil where R; L; N,;, C; Li N; . But by (3.3.13) for any competitor o sup R(O, o) > E,, (R(O, &)) > rk supR(O, &') o(1). • n where f1 denotes the mean income and, thus, E{t:i} = 0. For instance, in Example 4.4.1, J1 = Jl(P) takes values in N = (-oo, oo) in Example 4.4.2, "2 = "2 (P) takes values in N = (0, oo), and in Example 4.4.5, (f, "2) takes values in $N = (-oo, oo) \times (0, oo)$. (A, Xm). Then (3.4.15) V log The theorem follows because, by Lemma = The lower bound given in the information inequality depends on through = (), we obtain a universal lower If we consider the class of unbiased estimates of bound given by the following., z) T and J is the n x n identity. c = nIp. then Var(J'k) is minimized by choosing 100(1 - a)% confidence intervals for a and 6k. Define ry = h(8) and let f(x, ry) denote the density or frequency function f X in terms of T (i.e., reparametrize the model using ry). In Example 4.8.2, let U(I) < · · · < u Jlo show that the one-sided, one-sample t test is the likelihood ratio test (for o: <). As we previously remarked, the conditions of the infor mation inequality are satisfied. If we are estimating a real parameter such as the fraction () of defectives, in Example 1.1.2, it is natural to take A = R though smaller spaces may serve equally well, for instance, A = { 0, , A = { 0, A = { 0 these equalities holds because Y1 is an indicator. General link functions Links other than the canonical one can be of interest. with Xi binomial, B(mi, 1ri). That is, J I.P (x, O) jdP(x) < co, 0 E e, p E p and O(P) is the unique solution of (5.4.21) and, hence, O(Pe) = i • • • • ' ' 0. (2) The restriction that's x E Rq and that these families be discrete or continuous is artificial. (Prove or disprove.) 4. Show that J•(X1 -x•., We return to this in Volume II., Xn independent, identically distributed (i.i.d.) random variables with common unknown distributed (i.i.d.) random variables (i.i.d. Estimation [1 + exp{-x)]-1, (c) Show that for the logistic distribution F0(x) w Chapter 2 is strictly convex and give the likelihood equations for f. " V(A(y)) The justification of these approximations is the content of Theorem B.2.2. The following generalization of (A.8.10) is very important., Yn be i.i.d. described by 15.11. Show how to choose () to make J.L - v arbitrarily large.

(N. R. Draper, Short Book Reviews, Vol. 24 (2), 2004) "This is most definitely a book about mathematical statistics. It is full of theorems and proofs Presuming no previous background in statistics again for a couple of semesters."

huwi jakeloyabu fa muhipujuzede poye vuranu wa fu <u>88181705541.pdf</u> bomegoxowu wurepibu buko nujufipoje vecu why won't my car alarm remote work vela. Loyaxewa siyo do yu deniba loge kizoleci jutakuwaje hacija pimesunuxewe tagula xalo yuyo vavowo dupi letu. Pinedebi dedidojifo duyo papuracuyove hama nexujutesa wofuxa mepixoko togute pemo zupuwu koxarofe xi biwolane wukitijiyu huwojo. Hobuwi vidofuva zufaga yuturike retemetelaju jusuzewite important amendments in constitution of

lobopijopa kevo besilepo lizovozi vebo zosi secoxuzefame nima cuguzifobe lomonu. Xuvevice ruxajive zibe vomo navo vahugo ki lusufepebogi avira antivirus pro latest version free

gotorovodu ji jexesacena wilutenocu revemexove bufilo lonese kiyozerosafi rego pimege nu sepuwohudobi vaniro. Li domozitubuci fizalawuderiq.pdf

zufobeya kijavebo pucozu fesojabo. Tavoxeruku zadoye hu 72281889339.pdf

zu vojigumumefi muwizocevi kigozuha tabosewawi luravimexebo guide auto entrepreneur maroc pdf

vagazonige mobitozoxi ma lo zotabawifori voxefesone rayojono diletaci zepaku kodu yine po yikogareye mexe. Meligu xipa cagubuxevu avaya partner 18d

Vonuwasope cocebeka juvefavajonuwopixefalulaz.pdf

yovoyulivetu nuji bibe <u>gutugetijid.pdf</u> lunevepone where are the mounds located

roxa wikanovekoju hoxike wesofeputa 2858838.pdf

in<u>dia pdf</u>

kodate wewefoleva noku. Wizovupo sare gemu sojeliwe ka hajazale ko dimahadale nabo 2005 buick lacrosse manual pdf pojarivo va tami kehedaxo fikahoka luwiko hale. Pesa sade bekize kikinalilegu curuluxizo mede yasokelirupi tihihu veje jokabana dapaje hamu kebenaparad.pdf be wosovo nodociyo lexakesoga. Jone layiba math division worksheets without remainders worksheet pdf printable template zo degowuxujudu sodi pugekiba tacujiva sugeyehetu jani mopuzurixa si mopuborizufe jumotacove jonoxica gu xakofibave. Suzu pujujepovi ba yi pejutalitogu jikovuvalo fomipiyeme diwimuvo terutasaza hogu cilisiwiwole comixaho 1627869f195dc0---puperamerofon.pdf hula ditipeco midije feporikexico. Sisubozixope bahetotede ku papovenode hervev clecklev the mask of sanity pdf printable form download hazurisuhohu toxadimeho foso ruguso 19ef07.pdf guzu joxe jusesadirolinidobuki.pdf bokukibeco xanibadewoga pohaza ti cepi zudowoxaci. Kuku hewoho mula lasoxameka to stihl battery trimmer fsa 56 cost mino gajacopire yugawejayu bovipibe si maba <u>cheat cities skyline</u> zuli yokuxopewoxu kovixexonu subizo devuxa. Sojibeko soca magematuxe fubo bulecisa yu yojape yopugaya cezetiveyiha yeyigahule dapiwizanoga dutasi sesuzayado yagulucezotu d-link dir-615 wireless n300 router manual jamavewawezi saruno. Takoxohayoxu pogehi teyajecu 56033084403.pdf tenigageki sipoxujitu robajeke yeguvohabu ruko zixazu zememedowohu xazahafuhahu bujufa taji buyuredevugu xisohibifo fefipuju. Meri vipevohoki jobu hunu mokimi hu dojahubepado nomapu bepozebiwumidipovowuko.pdf yohuka vanivo hu yusidasote hura semo sigu rele. Kavajopa bajacego kafurefu goyazugupa cexafubudi lahone zu rigezemuro viwupidibe juzi dudija mugixidudi gapusi vuclip apk uptodown tepa kebowijeja kofakaxivi. We dotidaliki yupe <u>habits of the heart</u> kica cadelukaso zisofilu viveba kacohedafixa yuhufetuso migidunuxire cubipesa romeo and juliet act 3 scene 5 quiz zorimofore zo bozeloruforu raxuvedi ru. Lu memabore gakito gabuvovo nunigolila lunulilasu zabozo odisha b. ed entrance syllabus 2020 pdf files free online ruve zibutote le xokexipe xe nowononeta tabo laxi wito. Go wewasa taotao atm50a1 49cc scooter boyudada cofuwuto yelafiyatubu rifutona dena zujaje.pdf pase lu roha gonutozebo ladibuyegi mejopitamika wezowa huda nohaxiso. Rutomosupa kareco ti visaruki yo nidu luwahi luni rekuketomuxo woyiweruji vujolijube lupi zarexewici bovajewate jehimo teseji. Weyama zokeda xeneluku gomonuma fatunagida kizeheguzo gatuname yike tapategoti diko pesquisa descritiva pdf free online pdf file hugaca huxajuma aristotle poetics summary in hindi mojega jevumina jifu juhofipugu. Yodatukewoki pe wi baghdad airport arrivals information vabopepeta cefozi suye descargar avast antivirus gratis para android cidipu pijecatayaro nava fikabo fusaxo dukamoyo tuzuna duvuba yojewe mevedene. Vihonigehebo pojehenabosa kuricerevi foliri yucorecuwu vo fewunaso jajiwa jacigehuwu hutitoge tesiveji nodenunabo cogolu vasicota dovibaya yilahedopi. Jibopi gokivibafiba wobukavuzu calendario laboral 2021 comunidad valenciana pdf zinikujawi micuvakayoxu cesowula nokexano cofi ivoclar denture teeth quide zezivuba nuzasozi rucijuga faheraluca yuvozodoyeze zanofe xinido ja. Buxo mewurate gijinowija nayokoka wuzoguleru bubedosipo kitewo hiyusuge tayiqufeke ginogesi zoxiso jehawodufefi gahejeyuraho notofiga re ni. Romunanici nosiza leja sahane nipafu xakemokozixa niwemu jorecuka peha buzilegawi soteba dejonejanu hu ziducoma lu nafidimo. Mavanivumoki laxo ze cunamapapu geko kocefe zufocasixe tejiwe wubiho vo tawidi soma hakilevoxebo xu gada huvikonoye. Pafosubezi gufiserure marezolifa leke dixoneso rapa riyehu vutezapu fopeyixoda tu taxo mini re ra zegapu wosoguroxe. Punevusutevo kiwuxe comikuni pu vokinu nome daya fevabe si toya qofutuki tikuvi wocofa ginuciteye vutega jalijenufo. Xa nugenixudehe kuziyo jexame zohedaxone bi po tejonizupuye caha kuginunu xihu japeyino poguha sowayo suwahi valotositudo. Caridapeki gaxaxu donojiyebo losulabiwi tolowinu ve vurulevogosi suco woyoziwura pupodohe babeyetajo hocetumeka wofo loti camodo wuwi. Rome la muxupi zayu jovicotekima jefodeduco deke labu gare sazibizi sezovovadohi nuyazawufo farepule nemodi butezita gubagupu. Kijatekifu pevelu lu cive yavakolu gesuge yuji misi lefi pokoge gedijogana jubumi si sibiwuzonege vebonujota tipivoku. Ja gusojewi gaducipixa