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 • • • Area= .8 Area = .1 1 544 Tables Appendix C ' I • Pr(T 2 1) Table II t distribution critical values of df .25 .10 .05 .025 Right tail probability p .02 .01 .005 .0025 .001 .0005 1 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 1.886 2.920 4.303 4.849 6.965 9.925 14.09 22.33 3 1 .60 3 0.765 1 .638 2.353 3.182 3.482 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 ...

oil, sell the location, or sell partial rights we could operate, administer drugs, or wait and see. For some specified vo, H may be accepted, for other specified vo, H may be rejected. , Xn) is given by (see (A.I6.4)), . Let us call the possible categories or states of the first characteristic A and A and of the second B and B. If () 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) and and N(L2, a2), respectively, then the LR test of H : uf = u versus K : af # ab is based on the statistic sT f.S, where sT = (n1 1)-1 Σ(n1 1)(X X), ... and then base our decision on the observed sample X = (X1 ... Show that (6.3.19) holds. , Yn2 be i.i.d. G, and suppose the X's and Y's are independent. But cr 2 Var(X1) evidently is and so is f = t. ! f. 85 Section 1. After Bellman ( 1 ... by a Consider the problem of estimating the given the hypothesis, model; so the sample and ... then H ... of the ... CoroUary Suppose estimate of B. be i.i.d. f (x ( ) ... , a ... n [ ( E 9 ... the null ...

Likelihood Ratio Procedures; Statistical Theory; Maximum Likelihood Estimation; Confidence Intervals and Regions; Tests and Confidence Regions; Bayesian Procedures; Asymptotic Approximations; Inference in the Multiparameter Case; Appendix: A Review of Basic Probability Theory; Appendix B: Additional Topics in Probability and Analysis ...

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. 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 devel opment of graphical and other exploratory methods for which theoretical development and connection with mathematics have been minimal. 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 distribnted r(p, .>.) if, and only if, >.X is distribnted r(p, 1). is to calculate T(x) and then reject H if T(x) of the test. For instance, suppose I holds. (a) Let IJn be the first iterate of the Newton-Raphson algorithm for solving (6.2.1) starting at o;p Show that n IJ�) Xi, D,P( IJn IJ� - !:.n L . This event has probability ( ) ( l' - l'o) < P[T < -z(l - la)] � -z(l - la) - .Jii a ( -z(l - la)) � la. 2. (b) We imagine that the random variable X* produced by the random experiment we are interested in has a distribution that follows a "true" parametric model with an inter pretable parameter (}, but we do not necessarily observe X*. 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. In the medical setting of Example 4.1.3 this asymmetry appears reasonable. For instance, in Example 1.1.1 we can use the number of defectives in the population, NO, asa parameter and in Example 1.1.2, under assumptions (l)-(4), we may parametrize the model by the first and 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 • The density is ... (c) Exhibit a two-dimensional sufficient statistic for 8. Section 4-4 • , Xn be a sample from a normal population with unknown mean J.L and cr2. (SLLN) We use sup for all 341 here. (c) Find the posterior distribution of a. We see that D E(Y - c)2 has a unique minimum at c = p, and the lemma follows. That is what (1.4.4) tells us. -oo < y < oo. , Yr is an exponential family with density ... Ku converges in a neighborhood of zero. , 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(l - a,) fo' X + z(l - a,) ;,;] a:l = 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, ... 0 < 9 < I. We let X1, . 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 ( l l 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, ... (3.4.8) whenever the right-hand side of (3.4.8) is finite. However, we avoid this 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 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. , Yn are independent with vari ances I and E(Y1 ) = 0, E(Y;) = 0, i = 2 , . Show the distribution of X forms 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 CONTENTS PREFACE TO THE SECOND EDITION: VOLUME I xiii PREFACE TO THE FIRST EDITION xvii l STATISTICAL MODELS, GOALS, AND PERFORMANCE CRITERIA l 1.1 1.1.1 Data and Models I I 1.1.2 Pararnetrizations and Parameters 6 1.1.3 Statistics 1.1.4 Examples, Regression Models Data, Models, Parameters, and Statistics as Functions on the Sample Space 8 9 1.2 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 of Prior Distributions 62 1.6.3 1.7 Problems and Complements 66 1.8 Notes 95 1.9 References 96 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 99 102 107 2.2.1 Least Squares and Weighted Least Squares 107 2.2.2 Maximum Likelihood 114 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 !58 2.7 References !59 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 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, 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 ... 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. 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)[l + (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 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. 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. 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 argument given in B.6 establishes that (B.l0.9) 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. The answer will depend on the model and, most significantly, on what criteria of performance we use. 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 ... What value of c gives sizec>? (a) If f and ,- are the N(B, aJ) and posterior predictive distribution. 7, Show rigorously using (1.2.8) that if in Example 1.1.1, D = NO has a B(N, ?To ) distri bution, then the posterior distribution of D given X = k is that of k + Z where Z 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 Chapter 1 STATISTICAL MODELS , GOALS , AND PERFORMANCE CRITERIA 1.1 DATA, MODELS, PARA METERS AND STATISTICS 1. 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. Consider testing H : 81 = 82 = 0 versus K : IJ, > 0 or IJ2 > 0. An unknown number N8 of these elements are defective. , Yn be the i.i.d. survival times 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. We include this example to illustrate that we need not be limited by independence. Suppose that we know from past experience that a fixed proportion Bo = 0.3 recover from the disease with the old drug. Let 81, 82 > 0 and H be as above. Example 4.7.2. known. Minimum Distance EstimatesPermutation TestsEstimation of Irregular ParametersStein 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, Equivariance, and Minimax Procedures INFERENCE IN SEMIPARAMETRIC MODELS Estimation in Semiparametric Models Asymptotics. Using inequalities from prediction theory, we show how the infom1ation inequality can be extended to the multiparameter case. 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 have the following structure. Example 1.5.5. Show that ep(X , ) = O. (a) X is an unbiased estimate of X. Problems 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.l 4.20) we conclude that . For a review of the ele mentary properties of matrices needed in its formulation, we refer the reader to Section BJO. 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. 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 Functions of One Variable. Using Problem 4.5.6, we find that a competing lower confidence bound is J.1. Using the cen tral limit theorem, Slutsky's theorem, and the foregoing arguments, we find (Problem 5.3.28) ... 314 Asymptotic Approximations ... approximately correct if H is true and the X's and Y's are not normal. Arguing heuristically again we are led to estimates B that solve v8p(X, e) = o. The LR statistic for H ... 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, . the proportions of individuals of type A and type B, respectively. As a function of 8, D(80) 9) measures the (population) discrepancy between 8 and the true value 80 of the parameter. HUBER, P., "Robust Statistics: JAECKEL, L. We start by finding the density of c 1 , . 5 415 General ized Linear Models ... The data in the first sample are N(O, 1) and in the second they are N(O)a2) where a2 takes on the values I, 3, 6, and 9, as indicated in the plot. Figure 2.4.1 illustrates the process. 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 StatisticsMathematical 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. Section 1.2 13 Bayesian Models Thus, the resulting statistical inference becomes subjective. Establish (5.3.11). Although estimation loss functions are typically symmetric in v and a, asymmetric loss functions can also be of importance. Testing Independence of Classifications in Contingency Tables Many important characteristics have only two categories. (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). Life testing. The argument is left to the problems as are some numerical applications. Suppose X1 and X2 are independent exponential £(A) random variables. We next apply the duality theorem to MLR families: Theorem 4.5.1. Thus, when Pz ( z) > 0, the conditional expected value of Y is finite whenever the expected value is finite. 8 P rediction Interva ls 253 future GPA of a student or a future value of a portfolio. Minimax procedures and symmetry As we have seen, minimax procedures have constant risk or at least constant risk on the "most difficult" B. By varying the assumptions we obtain parametric models as with (l), (3) and (4) above, semiparametric as with (l) and (2) with F arbitrary, and nonparametric 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. Bayes' Rule whenever the denominator of the right-hand side is positive. (b) Show that XHL is translation equivariant and antisymmetric. Often the final simplification is made. 3.5.2 Interpretability ... 526 Additional Topics in Probability and Analysis ... An important example is the class of procedures that depend only on knowledge of a sufficient statistic (see Ferguson, 1 967; Section 3.4). Next we consider contingency tables for two nonnumerical characteristics having a and b states, respectively (e.g., eye color, hair color). Bickel [email protected] stat.berkeley.edu Kjell Doksum [email protected] 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. ... which is called the Kolmogorov statistic ... We have seen that smooth transfor mations h(X) are also approximately normally distributed. In addition we feel Chapter 10 on decision theory is essential and cover at least the first two sections. 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 with mean fL and variance u2, both of which are unknown. Optimality claims rest on a more refined analysis involving a reparametrization. ... Here J xdF(x) denotes J xp(x)dx in the continuous case and L: xp(x) in the discrete case. ... Let X � U(O, 8) be the uniform distribution on (0, 8). (b) Calculate the smallest n needed to bound the length of the 95% interval of part (a). Finaliy, Chapter 6 is devoted to inference in multivariate (multi parameter) models. Next use the central limit theorem and Slutsky's theorem. ... References Hoe!, Port, and Stone (1971) Chapter 8, Section 8.1 Parzen (1960) Chapter 5, Section 3; Chapter 8, Sections 2-3 Rao (1973) Section 2b.4 A.13 SOME CLASSICAL DISCRETE AND CONTINUOUS DISTRIBUTIONS By definition, the probability distribution of a random variable or vector is just a probability measure on a suitable Euclidean space. ... In this edition we pursue our philosophy of describing the basic concepts of mathemat ical statistics relating theory to practice. Methods of Estimation. ... Bayesian Predictive Distributions ... It will often be more convenient to work with unrestricted parameters. ... Let Xi denote the difference between the treated and control responses for the ith pair. 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 disease, producing energy, learning a maze, and so on. But given a parametrization (} -----+ Po, (} is a parameter if and only if the parametrization is identifiable. Next we note how we can apply the information inequality to the problem of unbiased estimation. 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. ... Wiley & Sons, 1986. To study the relation between the two characteristics w e take a random sample o f size n from the population. For each simulation the two samples are the same size (the size indicated on the x-axis). The correlation inequality is equivalent to the statement (A.l l.l9) Equality holds if and only if X2 is linear function of X1. Section A.12 459 Moment and Cumulant Generating Functions If X1 , . The result might be a density such as the one marked 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 = O, with Oo < 8,. 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. Suppose that 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). For example, if N fills the number of disintegrations of a fixed amount of some radioactive substance in the period from time to time t, then (N(t)) is a Poisson process. Versus (a) Show that the test that rejects H for large values of y"n(X - L) has p-value p = 2 test meets on a population and that we administer only one treatment to each subject and a sample of nk subjects get treatment k. In (1.6.7), We are faced with a population of N elements, for instance, a shipment of manufactured items. This is in fact true for sections by any plane perpendicular to the (y) plane. We shall explore this matter further in subsequent chapters. Show that the Cauchy distribution about (), f(x, 9) = ... Does X have a pdf? Example 1.1.3 Regression. In fact, it can be shown that the conditional distribution is of the form (4.2.4). Examples are the distribution of income and the distribution of wealth. Begin by noting that according to (b), Give the conditional frequency, density, or distribution function in each case. A subject's lifetime... If we then have a set of observations with this distribution, the MLE is the method of moments estimate. It is not hard to verify that the power function is nonincreasing. Suppose that the distribution of X is symmetric about... The failure of the sample mean as an estimate of the center is a clear example of a problem of robustness. Chapter 4 and the test function for the sufficient statistic... Consider the model in which the outcomes are independent and identically distributed. Consider the problem of estimating the parameter of a Poisson distribution from a sample; the Bayes estimate with respect to a gamma prior is a weighted average of the prior mean and the sample mean. The likelihood ratio test, the Neyman-Pearson lemma, uniformly most powerful tests, confidence bounds and intervals, the Bayesian formulation, prediction, sufficiency, exponential families, the information inequality, asymptotic approximations, the delta method, large sample tests and confidence regions, and nonparametric models are treated in the chapters that follow. STIRZAKER, Probability and Random Processes, Oxford: Clarendon Press. BICKEL, P. and DOKSUM, K., Mathematical Statistics, Basic Ideas and Selected Topics. HAJEK, J. and SIDAK, Z., Theory of Rank Tests. LEHMANN, E. L., Testing Statistical Hypotheses. LEHMANN, E. L. and CASELLA, G., Theory of Point Estimation. Show that the estimate is unbiased and consistent. Problems and Complements. Suppose X1, ..., Xn is a sample from a population with density f(x, 0). Find the maximum likelihood estimate of 0. Show that the MLE exists and is unique. Hint: Use the fact that the log likelihood is strictly concave. Find the UMVU estimate. Compute the Fisher information. Show that the variance of the estimate attains the information bound. Let X1, ..., Xn be i.i.d. N(u, σ^2). Find the likelihood ratio test for testing H : u = u0 versus K : u ≠ u0 and show that it is equivalent to the two-sided t test. Construct a level (1 - α) confidence interval for u. Suppose that Y1, ..., Yn are independent with Yi ~ N(α + βzi, σ^2). Find the least squares estimates of α and β and show that they are the MLEs. Compute the Gauss-Markov theorem ... The Bayes risk of the estimate is ... Show that the posterior distribution of 0 given X = x is ... Compute the posterior mean and variance. The Bernoulli, binomial, Poisson, normal and gamma families are exponential families. Let X have a canonical exponential family distribution. Show that the moment generating function of T(X) exists in a neighborhood of zero. Use the central limit theorem and Slutsky's theorem to show that the statistic is asymptotically normal. Find the asymptotic variance. Compare the asymptotic relative efficiency of the median and the mean. Consider the Hardy-Weinberg model with three genotypes. Find the MLE of 0 and its asymptotic distribution. The Newton-Raphson algorithm and the EM algorithm can be used to compute the MLE when no closed form is available. Show that the Wald, Rao score and likelihood ratio tests are asymptotically equivalent under the null hypothesis. Chi-square tests for goodness of fit and contingency tables are treated in Section 6.4. Generalized linear models and the analysis of deviance are treated in Section 6.5. Robustness properties and semiparametric models are considered in Section 6.6. References: ANDERSON, T. W., An Introduction to Multivariate Statistical Analysis. BERGER, J. O., Statistical Decision Theory and Bayesian Analysis. BOX, G. E. P. and TIAO, G. C., Bayesian Inference in Statistical Analysis. CRAMER, H., Mathematical Methods of Statistics. FERGUSON, T. S., Mathematical Statistics: A Decision Theoretic Approach. RAO, C. R., Linear Statistical Inference and Its Applications. SERFLING, R. J., Approximation Theorems of Mathematical Statistics. WILKS, S. S., Mathematical Statistics. Construct an estimate of the parameter and compute its risk. Let X1, ..., Xn be a sample from a uniform distribution on (0, 0). Show that the maximum is sufficient and complete. Find the UMVU estimate of 0. Compute the mean squared error of the MLE and the UMVU estimate and compare them.

experiment and three events for which (i) and (ii) hold, but (iii) does not. Theorem 5.3.1. l(i) and (ii) hold, then m · 1 h(j) ( ) Eh(X) = h(!!) + L'!! l( EX) J· j= l where • ' ' I · · !!|j + 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' B0, /1(8, o) is just the probability of type I error, whereas if B E B1, /3( B, o) is the power against (). The statistic is equivalent to the likelihood ratio statistic 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 (a + v/(J) = 13 for n and find that we need to take n = ((J/v) 2(z1 - a) + z(1/3)f This is the analogue of (4.2. 7) with a + z(1-r) playing the role of the. - From the point of view of decision theory, the minimax risk is the Bayes risk of the least favorable prior. This is equivalent to the effective sample size. If we want the error probabilities to be equal, we solve. A proof of the simplest case is given in the problems. As mentioned earlier, we discussed the asymptotic distribution of MLEs in regular exponential families. Example 5.4.1. ...

Section 5.4 Asymptotic Theory in One Dimension. ... the Bayes procedure with respect to the prior π is given by the posterior mean ... Theorem 5.2.1 (consistency of the MLE in exponential families) ... Problems and Complements ... Example 4.4.1 ... Confidence Bounds, Intervals, and Regions ... the Neyman–Pearson Lemma ... Likelihood Ratio Procedures ... the Wald test ... Rao score test ... Bayesian and frequentist inference ... Bernstein–von Mises theorem ... Hoeffding's inequality ... the Delta method ... Fisher information ... asymptotic efficiency ... Gaussian linear models ... the Gauss–Markov theorem ... robustness ... Notes and References ... Bickel, P. J., and K. A. Doksum, Mathematical Statistics ... Lehmann, E. L., Testing Statistical Hypotheses ... Lehmann, E. L., and G. Casella, Theory of Point Estimation ... Rao, C. R., Linear Statistical Inference and Its Applications ... Ferguson, T. S., A Course in Large Sample Theory ... Huber, P. J., Robust Statistics ... Cramér, H., Mathematical Methods of Statistics ... Wilks, S. S., Mathematical Statistics ... Berger, J. O., Statistical Decision Theory and Bayesian Analysis ... Savage, L. J., The Foundations of Statistics ... Raiffa, H., and R. Schlaifer, Applied Statistical Decision Theory ... DeGroot, M. H., Optimal Statistical Decisions ... Pearson, E. S., and H. O. Hartley ... (Draper's Research Memoirs, Dulan & Co, Biometrics Series II.) PEARSON, RAIFFA, H. HOEL, P. This set K will have a power when the max is attained. i m Eo, sup{ ipX(B') - pX(B),) · B' E S'(B, o)} ◆ o l5-o where

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) with probability I. X3) + be Cov(X2, X,) + ad Cov(X] , X,) + bd Cov(X,, X4) (A. Thus, we will need to ensure that our parametrizations are identifiable,that is, lh i 02 ==> Po1 i= Pe2• I 1 ' ' 1 • • • ' • • • • , � • • 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.), A5(a.s.), � n a.s. Using the strong law of large numbers we obtain (Problem Po [dnqn(t) �1f(B)exp {Eo ::(Xr,B)�} (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 J.Lo we see that we have, in fact, generated a family of level a tests {J(X, 11]l where 1 if vnY�; � 1 > tn -1 (1 - �a) 0 otherwise. 3 ) is obvious. = {9 E 8 : q(O) = w }, 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 g(xo , Yo) = sup 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.l4.22) guarantees that su p which implies that under H0, IPo,[.fii(Bn - Bo) > z] - (1 - 1>(z))l � 0, Jn1(80)(c,(a,Oo) - Bo) - z1_0 (5.4.45) � 0, and (5.4.41) follows. Consider, for instance, 64 and 66 in our example. H, ROGERS, AND J. Lehmann, 1990). Parenthetically we 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 just as a parameter need not be real or Euclidean valued. Thus, once does not the value of a sufficient statistic T is known, the sample X = given that P is valid. we obtain E(X) = JJ., Var X = a) < >- - al - g(a) . The gen eral definition ofparameters and statistics is given and the connection between parameters and pararnetrizations elucidated. , X(n) (cf. The proofs for the upper confid�nce hnund and interval constant fotlow by the same type of argument, 0 We next give connections between confidence bounds, acceptance regions, and p-values for MLR families: Let t denote the observed value t = T(x) ofT(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• (B) = T(A(B)) = {T(x) : x E A(B)}. , Xn be i.i.d. as X "' N(p,, cr 2 ) . , D). , n, be a sample from a population in Hardy-Weinberg equilibrium for a two-allele locus, Xi = (E;Jo €;2, €;3), where P9[X = ( 1 , 0, 0)] = B2 , Po [X = (0, 1, 0)] = 28(1 - B), Po[X (0, 0, 1)] = (1 - B)2, 0 < B < 1. Summary. ' I • 8.10.2 Order on Symmetric Matrices As we defined in the text for A, B 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 action space is the choice of loss function defined as a function l : P x A R+. I' ' i ! Section 1.3 17 The Decision Theoretic Framework Xjn 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 Chapter4. The main difference in our new treatment is the downplaying of unbiasedness both in es timation and testing and the presentation of the decision theory of Chapter 10 of the first edition at this stage. In this example we have 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+ l are i.i.d. uniform, U(O, 1 ) . The case we have I Section 2.4 131 Algorithmic Issues just discussed has d1 = · · · = 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 , we find that the interval ( 4.8.3) converges in probability to e ± z ( 1 - � a) O"o as n ----) oo . 1 ; 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+ l 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+ l - 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 � ZD]3 + < , rank(ZD) = k, 1) = I \ B(1 0) [1 - (1 - B)"]{ x - nB- x(1 - B)"} , nB2(1 - B)"[n - 1 + (1 - B)"] - [1 - (1 - B)"]'(1 - 2B)x + nB2 ] - - _..... (3) Give heuristic discussions of more advanced results such as the large sample 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). I I .23) ==I References Gnedenko (1967) Chapter 5, Sections 27, 28, 30 Hoe!, Port, and Stone (1971) Sections 4.2-4.5, 7.3 Parzen ( I 960) Chapter 5; Chapter 8, Sections I-4 Pitman (1993) Section 6.4 A.l2 MOMENT AND CUMULANT GEN ERATING FUNCTIONS A.I2.1 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 ) , . , /)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 ) lN m; lo l + exp{z; } + log ';: f]; T; The special case p Zi = = = where Tj = ) ( � = � � = = 2:::7=1 ZijXi and we make the dependence on N explicit. Pml + PJt Pml + P/1 + Pmo + PJO In fact, the same data can lead to opposite conclusions regarding these hypotheses-a phe nomenon called Simpson's paradox. - 9 E 8}, 8 c RP, p > I , be a family of models for X E X c R"� k Let q be a map from 8 onto !1, !1 c R , I � 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 1J and £ = Rk-l_ More over, the parameters 'IJ = 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) = L,(Oo) - !2 1 ( 9o)I!1(9o)I12(9o) (6.3.21) where I11 is the upper left q x q block of the r x r infonnation matrix I( 80), I12 is the upper right 6.3.9) under A0--A6 and consistency � q x d block, and so on. 7 = 7 = 1.57. We derive the information inequa!ity in one-parameter models and show how it can be used to establish that i n a canonical exponential family, T(X) is the UMVU estimate of its expectation. and -' • • • • where � denotes the gamma function. Conversely, if S(X) is a level l - a confidence region for v, then the test that accepts H,,.0 if and only if v0 is in S(X), is a level o test 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' - l'f < [X - l'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 a from the left at 00 and continues rising 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 AND 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 = � L;� 1 logXi. Itiseasy to seethat ifn 2 2, T has a density. Similarly, note that in the Hardy-Weinberg Example 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 �nd is unique. Suppose A4', A2, and A6 hold for .p � &l(&B so that I(B) < oo. K., "On the General Theory of Skew Correlation and Nonlinear Regression," Proc. Squared error gives much more weight to those {} that are far away from {} than those close to {}. 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 tjmes 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). .. + ( lo g a2 ) ] - �} - { - � [ (log 2 7T) + (log a5 ) J � } - Our test ruie, therefore, rejects H for large values of (a5 fa 2 ) . 16. Let X1 , i p (x, O)d1r(O) ( 81 is of the form Ot for some t. Suppose that Z has a binomial. ..!.!... 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) l . ' " i (8.2.7) Example B.2.1 is a special case of the corollary. - a 2 + n2 ) ( 2n3 + n2 ) ) ) 2 , (2nl + 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 confidence 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 (5.4.35) is equivalent to (5.4.36) = 0, cross multiplication shows that (5.4.36) is just the correlation Because Eo'!j;(X1 , B) inequality and the theorem follows because equality holds iff 'lj 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) � 1. P � (Zit � Z, )6t � "' � {J, Z,, j=2 �(! ) 2 l' '' • I:� 1 (Yi � ' 2 • Similarly compute the information matrix when the model is written as • y; = {J, (Z" where {31, 12 , I ' j=l 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 � 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+be < N++c for all a, b, c and then are given by i • Section 2.5 Problems and Complements !57 Hint: (b) Consider 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 I Section 2.4 Algorithmic Issues 133 in which such difficuJties arise is given in Problem 2.4.13. < I. l2.9) 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 2.831 3.135 3.527 3.819 22 0.686 1.321 1.717 2.074 2.183 2.508 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 .3 1 8 1 .7 1 1 2.064 2.172 2.492 2.797 3.091 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% 90% 95% 96% 98% 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 ofO such that "'£';' 1 p(x; O) = I for all 0. , xn) = ' ( 1 .2.9) fo rr(t)tk(1 - t)n -kdt In the cases where . Find the efficiency ep( X, X) as defined in (f).- If and a = 1, = 4, evaluate 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�SJ - 8f2)/(n - )S L��3 Wt/(n u2, . Ifthe equations - have a solution B (x) E C0, 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 , � ). respectively, with all parameters assumed unknown. These issues will be discussed further in Volume 2. l 2.2) essxp (x)dx if X is continuous. Show that both ,P and � are identifi able. 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� l xi. In that case rejecting the hypothesis at level as 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 parametrization that is unidentifiable may still have components that are parameters (identifiable). Then we can conclude from Theorem 6.3.3 that 2 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. B y 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,] > 0}, i = 1 , 2 , . Contingency Thbles 403 Logistic Regression for Binary Responses 408 6.4.3 *6.5 Generalized Linear Models 4!1 *6.6 Robustness Properties and Semiparametric Models 417 6.7 Problems and Complements 422 6.8 Notes 438 6.9 References 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.S Random Variables and Vectors: Transformations 451 A.9 Independence of Random Variables and Vectors 453 A.!0 The Expectation of a Random Variable 454 A. � � (5.4.41) (5.4.42) (5.4.43) Property (5.4.42) is sometimes called consistency of the test against a fixed alternative. To prove the sufficiency of (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, � ), J.L (X) is random variable S. The underlying 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. � � (a) Consider the test statistic D(Fx , F x) � max{ [Fx(t) - F_ x(t)[ : t E R}. , We will consider 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 the principal axis and spectral theorems for Euclidean space and the elementary theory of convex functions on elementary introduction to Hilbert space theory. In Example 1 .2.1 suppose n is large and (1 jn) L� 1 Xi the prior distribution 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 a2. � U(O, I) and set Un = 2n 1{U E [0, �)}. If f1old is 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. - l ' Note that F(xp) • ' ' ' ' ' • = vn[F(xp) - F(xp)J . em . We denote the jth derivative of h by h(i) and assume ]jh(ml]Joo = SUPx ]M ml ( x)J < M < oo l l • ' . 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 distribution, whereas the level of the prediction interval ( 4.8.1) D i s not ( 1 - a ) i n the limit a s n --+ oo for samples from non-Gaussian distributions. Thus, minimizing n 1 ' p z; /3) 2 = 2::: ' • • • ' 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( ·, · , ttl , J.l2, af, cr�, p) in terms of F(·, · , 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'(l - B)N-'/b(y) ( ��; ) z-y 8' (1 - B) N- , � BY (l - B) N-y Problems for Section B.2 l. There are ideal observations, X � Po with density p(x, 8), 8 E 6 c Rd Their log likelihood lp,x(8) is "easy" to max imize. The central product moment of order (1, I) is plicity we consider the case = - I ' ' ' ' 458 A Review of Basic Prob.ability 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� - 2pawz)z(1 � (c) Show that if I . If c2 > 0, the bound is sharp and is attained only if Yi X·1 > _ £l.. LEHMANN, Descriptive Statistics for Nonparametric Models. Let x = xo < x1 ]x; - x; - , J :S J( I, if 8g2(x)j8xi8xi exists, convexity is equivalent to L u,u;82g (x)f8x,8x; > 0, all u E Rk and x E S. We ask whether the parameter B is in the subset 60 or subset 81 of e, where {So, 81}, is a partition of e (or equivalently if p E Po or p E PI)· If we take action a when the parameter is in ea. Theorem 4.2.1. (Neyman-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). l 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 · rn (a) Show that (b ) Let n-1 n 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), �, . [ L., dt. Putting the bounds f!(S), O(S) together we get the confidence interval (O(S), O(S)] oflevel (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 �(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 = (x1, . 1. WErHERILL, G. E(E(Y I Z)) using (a). Examples are f Gaussian, and f(x) = e-x(l + 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, 15 > 0. 15, Let ( X1 , Y1 ), . or Suppose that X (X1, ,Xn) 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( l). • 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) x (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 i s 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�of 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 1.1.1, or p, in Example 1.1.2, it is natural to take A = R though smaller spaces may serve equally well, for instance, A = { 0, �, . RAO, C. However, if gross errors occur, we ob serve not X* but X = (X1, Xn) where most of the Xi = x:, but there are a few • • • • ' ' ' ! , ! ' " ' • Section 3.5 191 Nondecision Theoretic Criteria � and use B(X1 , . Because 1r (z) varies between 0 and 1, a simple linear representation zT {3 for 1r ( ) over the whole range of z is impossible. n B, 1.3 ' ! The first of 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 artifi cial. (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 distribution function F. Establish (B.5.!6}-{B.5.19). where T is known, � zT{3, and b' and g are monotone. Show that T(X) is sufficient. 154 Methods of Estimation [1 + exp{ -x)J- 1 , (c) Show that for the logistic distribution F0(x) w Chapter 2 is strictly convex and give the likelihood equations for f.l and cr. Problem B.2.8) are sufficient 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. l l . 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 … this certainly would be my choice of textbook if I was required to learn mathematical statistics again for a couple of semesters."