Statistical Inference

Paul Garthwaite

Professor of Statistics
The Open University
Milton Keynes, UK

Ian Jolliffe

Professor of Statistics
University of Aberdeen
Aberdeen, UK

and

Byron Jones

Director, Research Statistics Unit
GlaxoSmithKline
Harlow, UK

OXFORD
UNIVERSITY PRESS

associated expectations and variances, while in statistics one or two courses on statistical methods, including an introduction to linear models, should be adequate. The mathematical background required is standard calculus and a passing acquaintance with matrix algebra, but little more.

Most of the material in the book can be covered in a course of about sixty one-hour lectures. Shorter courses could be based on Chapters 2 to 6, which contain what we regard as core material, although Sections 3.3, 4.7, and 6.7 might be omitted. The instructor could then select topics from subsequent chapters. These later chapters can largely be taught independently of each other, although Section 9.3 assumes Section 8.2 has previously been covered and Section 9.6 requires familiarity with much of the material in Chapters 6 and 7. Each chapter, except the first, includes a selection of exercises of varying difficulty. Some of these augment material in the text, but the purpose of most is to help the student consolidate ideas and results.

Acknowledgements

We are grateful to a number of friends and colleagues who have commented on earlier versions of parts of the book, and whose comments have led to improvements. In Chapter 10, Dave Collett, John Hinde, Peter Lane, Jim Lindsey, and Martin Ridout provided us with useful advice, and/or pieces of Genstat or GLIM code, and Chris Gilligan supplied one of the data sets. Jim Dickey made many useful comments on Chapters 6 and 7 and Anthony Davison gave helpful suggestions that improved Section 9.5. An anonymous reviewer made many helpful points throughout the book. Kay Penny was responsible for the excellent figures in the first eight chapters.

Most of the text, in various drafts, was expertly typed by Liz Ostrowski, Mavis Swain, and Louise Thomson. We are extremely grateful to them for their efficiency and patience.

Large parts of the book have evolved from courses that we have taught over a number of years. Some of the Exercises and Examples have been in our notes for so long that we have lost track of their original sources. We therefore acknowledge the debt we owe to the anonymous originators of these problems.

Finally, we are each grateful to our families for their patience and understanding throughout the gestation period of the book.

P.G.
I.J.
B.J.

Contents

Our single observation, X, is certainly complete since the Poisson distribution is a member of the one-parameter exponential family, and sufficient, so if we can find a function of X which is unbiased, then it must be an MVUE.

Consider

$$Y = (-1)^X,$$
$$E[Y] = \sum_{x=0}^{\infty}(-1)^x \frac{\theta^x e^{-\theta}}{x!} = e^{-\theta}\sum_{x=0}^{\infty}\frac{(-\theta)^x}{x!}$$
$$= e^{-\theta}e^{-\theta} = e^{-2\theta}.$$

Hence $(-1)^X$ is unbiased, and is, in fact, an MVUE. However, it does not seem sensible since $(-1)^X = \pm 1$ and $0 < e^{-2\theta} < 1$ for all $0 < \theta < \infty$. □

This example has been discussed at length in the literature. Meeden (1987), for instance, argues that it is the concept of unbiasedness itself which causes the problem, although Lehmann (1983b) suggests that it is 'inadequacy of information' in a certain sense which is to blame.

Although the example, and other aspects of unbiasedness, gives cause for concern, it does not detract from the general usefulness of MVUEs. It does, however, show that some caution may be necessary in their use.

2.9 Summary

In this chapter we have looked at a number of desirable properties of estimators. A dominant theme is the search for MVUEs. We have defined unbiasedness, and have seen how efficiency can be defined in terms of a lower bound to the variance of unbiased estimators. A condition for attainability of the lower bound was given.

The idea of sufficiency was then introduced. This concept is important throughout statistics and will appear prominently in several of the subsequent chapters. We gave results which enable us to find sufficient, and more particularly minimal sufficient, statistics. The role of sufficient statistics in finding MVUEs was then discussed.

Another important idea, which will reappear frequently in later chapters, is that of exponential families of distributions. These distributions were defined, and their relationship to the existence of sufficient statistics was discussed. Completeness was then defined, leading us finally to a strategy for finding MVUEs based on functions of complete sufficient statistics.

In the last section, two examples illustrated the fact that searching for MVUEs, although often useful, can be problematical.

2.10 Exercises

Exercise 2.1: Given that $X_1, X_2, \ldots, X_n$ is a random sample from $U[0, \theta]$, find the p.d.f. of $X_{(n)}$, the largest of the X_i.

Show that $2\bar{X}$ and $(n+1)X_{(n)}/n$ are both consistent estimators of θ and compare their variances.

Exercise 2.2: Suppose that $\hat{\theta}_1, \hat{\theta}_2$ are independent unbiased estimators of a parameter θ, with variances σ_1^2, σ_2^2, respectively, and $\tilde{\theta} = k_1\hat{\theta}_1 + k_2\hat{\theta}_2$, where k_1, k_2 are constants. Find the values of k_1, k_2 for which $\tilde{\theta}$ is unbiased and has the smallest possible variance.

Exercise 2.3: For a random sample $X_1, X_2, \ldots, X_n$ from a gamma distribution with p.d.f. $f(x;\, \theta) = \left(\frac{x^3}{\theta^4 3!}\right) e^{-x/\theta}$, $x > 0$, investigate whether $\hat{\theta} = \bar{X}/4$ is unbiased and consistent. [You may quote expressions for the mean and variance of this distribution, rather than derive them.]

Exercise 2.4: Suppose $\hat{\theta}$ is an estimator for θ with probability function $\Pr[\hat{\theta} = \theta] = (n-1)/n$ and $\Pr[\hat{\theta} = \theta + n] = 1/n$ (and no other values of $\hat{\theta}$ are possible); show that $\hat{\theta}$ is consistent but that bias $(\hat{\theta}) \not\to 0$ as $n \to \infty$.

Exercise 2.5: Find the Cramér–Rao lower bound to the variance of unbiased estimators of θ, given a random sample $X_1, X_2, \ldots, X_n$ from the distribution with density

$$f(x;\, \theta) = e^{-(x-\theta)}\exp(-e^{-(x-\theta)}), \quad -\infty < x < \infty.$$

Exercise 2.6: In the case of a random sample $X_1, X_2, \ldots, X_n$ from the Bernouilli distribution with probability function

$$f(x;\, \theta) = \theta^x(1-\theta)^{1-x}, \quad x = 0, 1, \ 0 \le \theta \le 1,$$

find, by differentiating the log of the likelihood or otherwise, whether the Cramér–Rao bound is attained for

(a) estimators of θ;
(b) estimators of θ^2.

Exercise 2.7: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from the Rayleigh distribution with p.d.f.

$$f(x;\, \sigma^2) = \frac{x}{\sigma^2}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad x \ge 0.$$

Determine the Cramér–Rao lower bound for θ when θ is (a) σ^2 (b) σ, and demonstrate whether or not the lower bound is attainable in the two cases.

Exercise 2.8: Suppose $X_1, X_2, \ldots, X_n$ form a random sample from the normal distribution with unknown variance σ^2. Show that the sample variance

$$S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$$

does not attain the Cramér–Rao lower bound for finite n, but does so as n tends to infinity. For what value of c does the estimator

$$c\sum_{i=1}^{n}(X_i - \bar{X})^2$$

of σ^2 have the smallest mean square error?

Exercise 2.9: Suppose that $\hat{\theta}$ is an estimator for the parameter θ, and that $E(\hat{\theta}) - \theta = b(\theta)$. Show that $\text{Var}(\hat{\theta}) \geq \left(1 + \partial b(\theta)/\partial\theta\right)^2 I_\theta^{-1}$, where I_θ is Fisher's information.

Exercise 2.10: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from $N(\theta, \sigma^2)$ where σ^2 is known. Use Lemma 2.1 to show that $I_\theta = n/\sigma^2$.

Exercise 2.11: A random sample of size n is available from the distribution with density

$$f(x;\lambda) = \begin{cases} \dfrac{\lambda^3 x(x+1)}{\lambda+2} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

For what functions of λ do there exist unbiased estimators which attain the Cramér–Rao lower bound? Show that Fisher's information is given by

$$I(\lambda) = \frac{2n(\lambda^2 + 6\lambda + 6)}{\lambda^2(\lambda+2)^2},$$

and verify, for one of the above functions, that it possesses an unbiased estimator attaining the Cramér–Rao lower bound.

Exercise 2.12: A random sample, $X_1, X_2, \ldots, X_n$, of n observations is taken from the two-parameter Weibull distribution with p.d.f.

$$f(x;\alpha,\beta) = \frac{\beta}{\alpha^\beta} x^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\}, \quad x > 0,\ \alpha > 0,\ \beta > 0.$$

Assuming β is known, find a single function of $X_1, X_2, \ldots, X_n$ which is sufficient for α.

Show, however, that if α is known there is no single function of $X_1, X_2, \ldots, X_n$ which is sufficient for β.

Exercise 2.13: Find minimal sufficient statistics for samples of size n from

(a) the uniform distribution on $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$,

(b) the uniform distribution on $[-\theta, \theta]$.

Exercise 2.14: Observations made on r.vs $X_1, X_2, \ldots, X_n$ are independent and identically distributed, each with a beta distribution whose p.d.f. is

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

where α, β are positive parameters, and $B(\alpha,\beta)$ is a beta function. Write down minimal sufficient statistics for (α,β).

Exercise 2.15: Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution whose p.d.f is

$$\begin{cases} f(x;\theta), & a \leq x \leq b(\theta), \\ 0 & x < a \text{ or } x > b(\theta) \end{cases}$$

where a is a constant, $b(\theta)$ is a fixed function of θ, and θ is a parameter to be estimated. Show that if a single sufficient statistic exists then it must be $X_{(n)}$, and that a necessary condition for $X_{(n)}$ to be a sufficient statistic is that $f(x;\theta) = g(x)h(\theta)$, where $g(x)$ does not depend on θ, and $h(\theta)$ does not depend on x.

Exercise 2.16: Show that the Rayleigh distribution with p.d.f.

$$f(x;\theta) = \frac{x}{\theta}\exp(-x^2/2\theta), \quad x > 0,\ \theta > 0,$$

belongs to the one-parameter exponential family of distributions, and deduce a minimal sufficient statistic, T, for θ, based on a random sample $X_1, X_2, \ldots, X_n$.

Exercise 2.17: Suppose that $X_1, X_2, \ldots, X_n$ are independent r.vs, each with the inverse Gaussian distribution whose p.d.f. is

$$f(x;\theta_1,\theta_2) = \sqrt{\frac{\theta_1}{2\pi x^3}} \exp\left\{\frac{-\theta_1(x-\theta_2)^2}{2\theta_2^2 x}\right\}.$$

Show that this distribution is a member of the two-parameter exponential family of distributions. Hence, or otherwise, find a minimal sufficient statistic for (θ_1, θ_2). Find also a minimal sufficient statistic

(a) for θ_2 when θ_1 is known;

(b) for θ_1 when θ_2 is known.

Exercise 2.18: Random variables X_1 and X_2 are independently and identically distributed with p.d.f. $\lambda e^{-\lambda x}$, $x \geq 0$. Given that the p.d.f. of $Z = X_1 + X_2$ is $\lambda^2 z e^{-\lambda z}$, show that the distribution of X_1 conditional upon $Z = z$ is the uniform distribution on $(0, z)$. Prove that the MVUE of

$$\Pr[X_i > 1] = e^{-\lambda}$$

based on observations X_1 and X_2 is given by

$$T = \begin{cases} 0 & \text{if } z \leq 1, \\ \dfrac{z-1}{z} & \text{if } z > 1. \end{cases}$$

Exercise 2.19: The r.vs $X_1, X_2, \ldots, X_n$ are independent with common probability density $\theta x^{\theta-1}$ $(0 < x < 1)$, where the parameter $\theta > 0$ is unknown.

(i) Find a sufficient statistic for θ.

(ii) Given that $-\log X_1$ is an unbiased estimator of θ^{-1}, find another unbiased estimator with smaller variance.

Exercise 2.20: Consider a binomial experiment with probability of success p in which m (fixed) trials are conducted, resulting in R successes; a further set of trials is then conducted until s (fixed) further successes have occurred. The number of trials necessary in the second set is a r.v., N. By considering the function $U(R,N) = R/m - (s-1)/(N-1)$ show that (R,N) are jointly sufficient for p, but not complete.

Exercise 2.21: Let $X_1, X_2, \ldots, X_n$ be n independent observations from an exponential distribution, with density

$$f(x;\theta) = \theta^{-1}e^{-x/\theta}, \quad x \geq 0,\ \theta > 0.$$

Consider the following estimators of θ; $T_1 = \sum X_i/n$, $T_2 = \sum X_i/(n+1)$, and $T_3 = nY$, where $Y = \text{Min}(X_1, X_2, \ldots, X_n)$. Which of these estimators are unbiased, which are functions of sufficient statistics and which are consistent?

Discuss the relative merits of T_1, T_2, and T_3 with regard to the above and any other relevant criteria.

Exercise 2.22: A random sample $X_1, X_2, \ldots, X_n$ is selected from $N(\mu, 1)$. Write down the joint distribution of X_1 and $Y = X_2 + \cdots + X_n$; hence or otherwise obtain the distribution of X_1 conditional upon the observed value $\bar{x}$ of the sample mean $\bar{X}$.

Let a r.v. W take the value 1 if X_1 is less than 0, and the value 0 if X_1 exceeds 0. If the parameter μ is unknown, show, by using W or otherwise, that $\Phi(-\bar{X}\sqrt{n}/\sqrt{n-1})$ is an efficient estimator of $\Phi(-\mu)$, where the function Φ is the c.d.f. of a r.v. distributed as $N(0,1)$.

Exercise 2.23: A random sample $X_1, X_2, \ldots, X_n$ is obtained from a truncated Poisson distribution with probability function

$$f(x; \lambda) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \cdot \frac{\lambda^x}{x!}, \quad x = 1, 2, \ldots \quad \lambda > 0.$$

For $i = 1, 2, \ldots, n$, a r.v. Z_i is defined by

$$Z_i = \begin{cases} X_i, & X_i \geq 2, \\ 0, & X_i = 1. \end{cases}$$

Show that $\sum_{i=l}^{n} Z_i/n$ is an unbiased estimator of λ with efficiency

$$(1 - e^{-\lambda}) \Big/ \left[1 - \left\{\frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}}\right\}\right]^2.$$

Explain briefly how you could construct a more efficient unbiased estimator. Does there exist an unbiased estimator with efficiency 1?

Exercise 2.24: Observations are made of the value of a r.v. Y under two experimental conditions and in association with various values of a variable x. For each condition the model proposed for the dependence of Y on x is one of linear regression; observations (y_{ij}, x_{ij}), $j = 1, 2, \ldots, n_i$ are made under condition $i (i = 1, 2)$, and the model may be written as

$$Y_{ij} = \alpha_i + \beta_i(x_{ij} - \bar{x}_i) + e_{ij}$$

$(i = 1, 2;\ j = 1, 2, \ldots, n_i)$ where $\bar{x} = (1/n_i)\sum_{j=1}^{n_i} x_{ij}$, and the e_{ij} are independent $N(0, \sigma^2)$ r.vs. Find expressions for the minimum-variance unbiased estimators of α_1, α_2, β_1, β_2, and σ^2, and identify their joint distribution.

Exercise 2.25: A random sample of size $n (n \geq 3)$ is available from the exponential distribution with density

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Find the minimum-variance unbiased estimator of λ, and show that the ratio of its variance to that given by the Cramér–Rao lower bound is $n/(n-2)$. Why would you expect this ratio to be greater than 1?

Exercise 2.26: The r.v. X has a discrete distribution such that $\Pr[X = r] = \theta^{-1}$ for $r = 1, 2, \ldots, \theta$, where θ is an unknown positive integer. Show that Y, the maximum of a sample of n independent observations of X, is a complete sufficient

statistic for θ and hence verify that $[Y^{n+1} - (Y-1)^{n+1}]/[Y^n - (Y-1)^n]$ is a minimum-variance unbiased estimator for θ.

Exercise 2.27: Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\mu > 0$. Find the minimum-variance unbiased estimator for

$$\Pr[X \leq 1] = (1 + \mu)e^{-\mu}.$$

Does this estimator attain the Cramér–Rao lower bound?

widespread use in statistical inference. We have seen that this popularity is justified by the general good properties of MLEs. However, like other inference procedures, there are circumstances where MLEs are less than ideal.

The final section of this chapter briefly describes four alternative methods of estimation which may be of use when MLEs are not satisfactory, or for finding initial values in an iterative ML scheme. All these alternative methods remain within the classical framework of inference, based on sampling distributions, and there are others within this framework, such as methods based on empirical transforms—see, for example, Morgan (2000, Section 5.7.4). Other methods outside this framework will be discussed in subsequent chapters. In particular, estimators based on decision theory and on the Bayesian approach are introduced in Chapters 6 and 7. Robust estimation and estimation based on computationally intensive methods are described in Chapters 8 and 9 respectively.

3.6 Exercises

Exercise 3.1: Find the MLE for θ in the following cases, where in each case we have available a random sample $X_1, X_2, \ldots, X_n$ from the relevant distribution.

(a) The geometric distribution with probability function $\theta(1-\theta)^{x-1}$, $x = 1, 2, \ldots$.

(b) The uniform distribution on the interval $(-\theta/2, \theta/2)$.

(c) The gamma distribution with parameters 2 and θ, that is, with p.d.f.

$$f(x \mid \theta) = \frac{1}{\theta^2} x\, e^{-x/\theta}, \quad x > 0.$$

(d) The Poisson distribution with mean θ.

Exercise 3.2: Consider the situation in Exercise 3.1(c) when there are n_1 observations, $X_1, X_2, \ldots, X_{n_1}$, whose values are given, and are all $\leq a$, for some constant $a > 0$. In addition it is known that $n_2 = (n - n_1)$ observations have values $\geq a$, but their exact values are unknown. Write down the likelihood function in this case, and find an equation whose solution will give the MLE for θ.

Exercise 3.3: Three independent binomial experiments are conducted with n_1, n_2, n_3 trials and x_1, x_2, x_3 are the respective numbers of successes observed.

(a) Suppose that the probability of success, p, is the same in each trial. Find the MLE of p, based on all three trials.

(b) Now suppose that the probability of success varies between trials, and is α, $\alpha + \beta$, α respectively. Find MLEs of α and β, based on all three trials.

Exercise 3.4: The r.v. X has an exponential distribution with p.d.f. $f(x;\theta) = \theta e^{-\theta x}$, $x > 0$. For a fixed positive constant τ, show that $E[X | X \geq \tau] = \tau + 1/\theta$.

Exercise 3.5: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a distribution belonging to the k-parameter exponential family. Show that the E-step in the EM algorithm reduces to finding the conditional expectations of the k minimal sufficient statistics, and then substituting these expected values into the complete log-likelihood function.

Exercise 3.6: Consider the complete log-likelihood in Example 3.2 in Section 3.2. Using Exercise 3.5, or otherwise, show that it is a linear function of the unknown values $x_{m+1}, x_{m+2}, \ldots, x_n$.

Exercise 3.7: For $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ with both μ and σ^2 unknown, differentiate the log-likelihood with respect to μ and σ, and hence show that the MLE of σ is the square root of the MLE for $\hat{\sigma}^2$ found in Example 3.5.

Exercise 3.8: Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution having p.d.f.

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2}, & x \geq \theta_1, \theta_2 > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the MLEs of θ_1 and θ_2.

Exercise 3.9: Observations are made on independent r.vs X_{ij} where $X_{ij} \sim N(\mu_i, \theta)$, $j = 1, 2$; $i = 1, \ldots, n$. Write down the likelihood function for θ, $\mu_1, \mu_2, \ldots, \mu_n$. Obtain the MLEs of the parameters $\mu_1, \ldots, \mu_n$ and show that the MLE $\hat{\theta}$ of θ is $(1/4n) \cdot \sum Z_i^2$, where $Z_i = X_{i1} - X_{i2}$. Obtain the expectation and variance of $\hat{\theta}$; hence or otherwise prove that, as $n \to \infty, \hat{\theta}$ is not a consistent estimator for θ.

Exercise 3.10: The r.vs $X_1, X_2, \ldots, X_m, X_{m+1}, \ldots, X_n$ are independently, normally distributed with unknown mean μ and variance 1. After $X_1, X_2, \ldots, X_m$ have been observed it is decided to record only the signs of $X_{m+1}, \ldots, X_n$. Obtain the equation satisfied by the MLE $\hat{\mu}$ of μ.

Exercise 3.11: A random sample of n observations is taken on a r.v. X which has a Poisson distribution with mean θ. Suppose that $\phi = \theta^2$. Find the MLE, $\hat{\phi}$, for ϕ, and show that $\hat{\phi}$ is a biased, but consistent, estimator. (Note that $E[X^4] = \theta^4 + 6\theta^3 + 7\theta^2 + \theta$.)

Exercise 3.12: Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution having p.d.f.

$$f(x \mid \theta) = \begin{cases} \dfrac{2x}{\theta^2}, & 0 < x < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the MLE of the median of this distribution and show that this estimator is sufficient. Is it also minimal sufficient?

Exercise 3.13: Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution $U[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]$. Find the MLEs of μ and σ.

Exercise 3.14: A random sample of size n is obtained from a bivariate normal distribution with density function

$$f(x,y) = (2\pi)^{-1}(1-\rho^2)^{-1/2}\sigma^{-2} \times \exp\left[-\frac{1}{2(1-\rho^2)\sigma^2}\{(x-\mu_1)^2 - 2\rho(x-\mu_1)(y-\mu_2) + (y-\mu_2)^2\}\right]$$
$$-\infty < x, y < \infty.$$

Write down the likelihood function and show that its logarithm may be written as

$$\ell = \text{constant} - n\ln(\sigma^2) - \tfrac{1}{2}n\ln(1-\rho^2) - \tfrac{1}{2}n\sigma^{-2}(1-\rho^2)^{-1}\Big[(s_1^2 + s_2^2 - 2\rho s_{12}) + (\bar{x}-\mu_1)^2 - 2\rho(\bar{x}-\mu_1)(\bar{y}-\mu_2) + (\bar{y}-\mu_2)^2\Big],$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$s_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

Hence obtain the MLEs of μ_1, μ_2, σ^2 and ρ.

Exercise 3.15: Consider the bivariate normal distribution in Exercise 3.14, modified so that μ_1, μ_2 take the common value μ, but X, Y now have different (known) variances σ_1^2, σ_2^2. The part of the log-likelihood function which involves μ may be written as

$$\ell = -\frac{1}{2}n(1-\rho^2)^{-1}\sum_{i=1}^{n}\left[\frac{(x_i-\mu)^2}{\sigma_1^2} - \frac{2\rho(x_i-\mu)(y_i-\mu)}{\sigma_1\sigma_2} + \frac{(y_i-\mu)^2}{\sigma_2^2}\right].$$

Show that the MLE of μ is

$$\hat{\mu} = \frac{\bar{x}(\sigma_2^2 - \sigma_{12}) + \bar{y}(\sigma_1^2 - \sigma_{12})}{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})},$$

where $\sigma_{12} = \rho\sigma_1\sigma_2$, and ρ is known.
Suppose that $\bar{x} = 2$, $\bar{y} = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 3$, $\sigma_{12} = 1.5$. Find $\hat{\mu}$ and comment on its value.

Exercise 3.16: A sample of n independent observations is taken on a r.v. X having a logarithmic series distribution,

$$P(X = x) = \frac{-\theta^x}{x\ln(1-\theta)}, \quad x = 1, 2, \ldots,$$

where θ is an unknown parameter in the range $(0,1)$. Show that the MLE $\hat{\theta}$ of θ satisfies the equation

$$\hat{\theta} + \bar{x}(1-\hat{\theta})\ln(1-\hat{\theta}) = 0,$$

where $\bar{x}$ is the sample mean. Find the asymptotic distribution of $\hat{\theta}$.

Exercise 3.17: (a) The gamma distribution

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0,$$

has mean α/β and variance α/β^2. If α is known, find the MLE of β and its asymptotic variance.

Find the MLE of $1/\beta$. What is its mean and variance? Does this estimator attain the Cramér–Rao lower bound? Does the MLE for β attain the Cramér–Rao lower bound?

(b) Verify that the MLE of $1/\beta$ obtained in part (a) is sufficient. Is the estimator of β also sufficient?

Exercise 3.18: Independent observations $x_1, x_2, \ldots, x_n$ are available on a r.v. which is normally distributed with mean and standard deviation both equal to μ.

Find the MLE $\hat{\mu}$ of μ. Show that the asymptotic variance of $\hat{\mu}$ is $\mu^2/3n$.

Compare this with the variance of the MLE which would have been obtained for the mean if the functional relationship between the mean and variance had been ignored. Comment on your results.

Exercise 3.19: A closed population of animals contains θ males and θ females. A random sample is taken in which each animal, independently of all others, has a known probability p of being caught. Altogether m males and f females are caught. Write down the likelihood function $L(\theta)$ for θ.

Let $t = \text{Max}(m, f), s = \text{Min}(m, f)$. By considering the ratio of $L(\theta+1)$ to $L(\theta)$, or otherwise, show that if

$$\{(t+1)(1-p)\}^2 < (t-s+1),$$

then the MLE $\hat{\theta}$ of θ is t, and that if the condition above is not satisfied then $\hat{\theta}$ is the largest integer smaller than the larger root of the equation

$$(2p - p^2)\theta^2 - (m+f)\theta + mf = 0.$$

Exercise 3.20: The number of particles emitted by a radioactive source in unit time has a Poisson distribution with mean ρ. The strength of the source decreases as time goes by, and on days 0, 1, 2, ..., n it is assumed that ρ is α, $\alpha\beta$, $\alpha\beta^2$, ..., $\alpha\beta^n$ respectively, where α, β are unknown parameters. Independent counts of particles $x_0, x_1, x_2, \ldots, x_n$ are obtained over unit time on days 0, 1, 2, ..., n. Show that

$$\left(\sum_{i=0}^{n} x_i, \sum_{i=0}^{n} ix_i\right)$$

is minimal sufficient for (α, β), and find equations whose solution will give $\hat{\alpha}, \hat{\beta}$, the MLEs for α and β.

Write down approximate expressions for the variances of $\hat{\alpha}, \hat{\beta}$.

Exercise 3.21: In families where one parent has a rare hereditary disease the probability that a particular child inherits the disease is p, where $0 < p < 1$. Show that, in a family of fixed size k, the probability of at least one abnormal (diseased) child is $1-(1-p)^k$. In a survey, only families of size k with at least one abnormal child are sampled. In all, n such families are observed independently and there are r_i abnormal children in the ith family ($i = 1, 2, \ldots, n$). Show that $\hat{p}$, the MLE for p, satisfies the equation

$$nk\hat{p} = [1-(1-\hat{p})^k]\sum_{i=1}^{n} r_i.$$

Exercise 3.22: Individuals are given a measurable stimulus, to which they may or may not respond. When the stimulus is x, the probability that an individual responds is $p(x)$ where $p(x)$ and x are related by

$$p(x) = \frac{e^{\alpha+\beta x}}{\{1 + e^{\alpha+\beta x}\}}.$$

Here α and β ($\beta > 0$) are fixed and unknown constants. Thus as x increases from $-\infty$ to ∞, $p(x)$ increases from 0 to 1. Experimental results are available for k groups of individuals as follows. Each of the n_i individuals in the group was given stimulus x_i, and r_i individuals responded out of these $n_i (i = 1, 2, \ldots, k)$.

Find two equations whose solution gives the MLEs $\hat{\alpha}$ and $\hat{\beta}$ for the parameters α and β, and describe briefly how you would set about solving for $\hat{\alpha}$ and $\hat{\beta}$ when numerical values for $n_1, n_2, \ldots, n_k; r_1, r_2, \ldots, r_k$ are given.

Exercise 3.23: A cosmetic company is considering the marketing of a new product for men and wishes to estimate the proportion, θ, of males in a certain age group that would buy the product. Because a direct question may cause embarrassment, a so-called randomized response procedure is used to disguise the interviewee's actual willingness to buy the product.

Each person interviewed throws a fair die, and instead of giving the interviewer his true response 'Yes (will buy the product)' or 'No (will not buy it)', he gives a coded response A, B or C, as indicated by the table below. The interviewer does not see the score on the die.

		Die Score					
		1	2	3	4	5	6
True	Yes	C	C	C	A	B	A
Response	No	C	A	A	B	A	B

In a random sample of 1000 men, the numbers of A, B, C responses were 440, 310, and 250, respectively. If each man in the sample has the same probability, θ, of having the response 'Yes', show that the log-likelihood for θ is

$$440 \ln(3-\theta) + 310 \ln(2-\theta) + 250 \ln(1+2\theta) + \text{constant},$$

and obtain the maximum likelihood estimate of θ.

Exercise 3.24: Suppose that the number of eggs laid by a particular parasite is a r.v. N, where N has a Poisson distribution with mean μ. Each egg, independently of all other eggs, has a probability p of hatching. Given that M denotes the number of hatched eggs, determine the joint distribution of (N, M). Given that $(n_1, m_1), (n_2, m_2), \ldots, (n_s, m_s)$ is a random sample of size s, determine $\hat{\theta}$, the MLE of $\theta' = (\mu, p)$ and find the asymptotic distribution of $\hat{\theta}$.

Exercise 3.25: Consider a time series $X_1, X_2, \ldots, X_n$ which follows a first order autoregressive model so that

$$(X_t - \mu) = \phi(X_{t-1} - \mu) + \varepsilon_t,$$

where $\mu = E[X_t], t = 1, 2, \ldots, n$, ϕ is an unknown parameter which is to be estimated, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent r.vs, each with mean zero and variance σ^2. Show that the conditional distribution of X_t, given $X_{t-1}, \ldots, X_1$ for $t = 2, \ldots, n$, is $N[\phi(X_{t-1} - \mu), \sigma^2]$, and that the marginal distribution of X_1 is $N[\mu, \sigma^2/(1-\phi^2)]$.

Hence, show that the log-likelihood function can be written as

$$\ell(\phi, \sigma^2; x) = \text{constant} - \frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\ln(1-\phi^2) - \frac{(1-\phi^2)}{2\sigma^2}(x_1 - \mu)^2$$
$$- \frac{1}{2\sigma^2}\sum_{t=2}^{n}[(x_t - \mu) - \phi(x_{t-1} - \mu)]^2.$$

Exercise 3.26: Find the method-of-moments estimator of θ in (a), (b), (c), (d) of Exercise 3.1.

Exercise 3.27: Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$. Show that the MLE of θ is any value $\hat{\theta}$ in the interval $[\max(X_i) - \frac{1}{2}, \min(X_i) + \frac{1}{2}]$. What is the method-of-moments estimator?

Exercise 3.28: Let $X_1, X_2, \ldots, X_n$ be a random sample from the density

$$f(x) = e^{-(x-\theta)}, \quad x > \theta.$$

(a) Show that the MLE $\hat{\theta}$ of θ is the minimum of $X_1, \ldots, X_n$.

(b) By finding the density function of $\hat{\theta}$ show that $\hat{\theta}$ is a consistent, but biased, estimator of θ with $E(\hat{\theta}) = \theta + 1/n$. Suggest an unbiased and consistent estimator and find its variance.

(c) Compare the sampling properties of the MLE with those of the method of moments estimator. Is it appropriate to compare the variances of these estimators with that suggested by the Cramér–Rao inequality?

Exercise 3.29: Consider a random sample of size n from the distribution with p.d.f.

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, \ \theta > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find

(a) the MLE for θ;

(b) an estimator based on the method of moments.

Exercise 3.30: For the general linear model described in Section 3.4.2 show that the MLE for β is identical to the least squares estimator.

Exercise 3.31: The minimum χ^2 estimate of a scalar parameter θ minimizes

$$\sum_{i=1}^{k} \frac{(\hat{p}_i - p_i(\theta))^2}{p_i(\theta)},$$

and the modified minimum χ^2 estimate minimizes

$$\sum_{i=1}^{k} \frac{(\hat{p}_i - p_i(\theta))^2}{\hat{p}_i},$$

(see Section 3.4.3). Show that the two estimates can be obtained by solving the equations

$$\sum_{i=1}^{k} \left(\frac{\hat{p}_i}{p_i(\theta)}\right)^2 \frac{\partial p_i(\theta)}{\partial \theta} = 0,$$

and

$$\sum_{i=1}^{k} \left(\frac{p_i(\theta)}{\hat{p}_i} \right) \frac{\partial p_i(\theta)}{\partial \theta} = 0,$$

respectively. (Recall that $\sum_{i=1}^{k} \hat{p}_i = \sum_{i=1}^{k} p_i(\theta) = 1$.)

Exercise 3.32: Using Exercise 3.31, or otherwise, find equations whose solutions give the minimum χ^2 estimate and the modified minimum χ^2 estimate, respectively for the mean θ of a Poisson distribution. Assume that a random sample of n observations is available, and that the classes used are all the distinct values of the r.v. which have been observed.

4 Hypothesis testing

4.1 Introduction and some basic definitions

The most usual formulation of a hypothesis-testing problem is that we have to decide between two hypotheses regarding one or more parameters θ. The two hypotheses are the **null hypothesis**, which we denote by H_0: $\theta \in \omega$, and the **alternative hypothesis**, which we denote by H_1: $\theta \in \Omega - \omega$. Thus Ω is the set of all possible values for θ, called the **parameter space** and ω is some subset of Ω.

To make our choice between H_0 and H_1 we have a random sample $X_1, X_2, \ldots, X_n$ from a distribution with p.d.f. $f(x; \theta)$. We choose some subset C of possible values of $X_1, X_2, \ldots, X_n$ and reject H_0 if and only if $\boldsymbol{X} \in C$. Usually C is defined in terms of (extreme) values of some statistic $T(\boldsymbol{X})$. T is the **test statistic** and C is the **critical region** or **rejection region**; $\bar{C}$, the complement of C, is the **acceptance region**.

A hypothesis is **simple** if it specifies a single value for θ (i.e. ω or $\Omega - \omega$ contain only one point); otherwise it is **composite**.

It seems likely that it will be easier to deal with simple hypotheses than with composite ones, so we start in the next section by looking at the simplest case of all when H_0, H_1 are both simple.

4.2 Simple null and alternative hypotheses—the Neyman–Pearson approach

If both H_0 and H_1 are simple, they can be written as

$$H_0\colon \theta = \theta_0, \quad H_1\colon \theta = \theta_1$$

for some values θ_0, θ_1 of θ. When we decide which of H_0, H_1 to choose there are two ways in which a mistake can be made.

Definition 4.1 *Rejection of H_0 when it is true is called a **Type I error**; acceptance of H_0 when it is false is a **Type II error**. The probabilities of making a Type I error or Type II error are denoted by α and β, respectively.*

Ghosh and Mukerjee (1994) discuss some problems related to tests based on adjusted conditional likelihood and adjusted profile likelihood, from both frequentist and Bayesian points of view—see also DiCiccio and Stern (1994). Adjustments analogous to Bartlett corrections may also be made to score tests and Wald tests—see, for example, Cordeiro *et al.* (1993), Rao and Mukerjee (1995). DiCiccio *et al.* (1996) and Stern (1997) describe adjustments to profile likelihood-based score statistics. The results are relevant to interval estimation as well as to hypothesis testing.

Many of the distributional approximations outlined in this section, as well as the original χ^2 approximation to $-2\ln(\lambda)$, rely, for their validity, on an assumption of some version of the usual regularity conditions. When these conditions do not hold, the distributional approximations will often break down, although it is sometimes possible to derive alternative distributional results. For example, Stuart and Ord (1991, pp. 876–80) discuss an exact distributional result, $-2\ln(\lambda) \sim \chi^2_{2d}$, which holds in some situations where the range of X depends on $\boldsymbol{\theta}$. Cox and Hinkley (1974, Section 9.3) describe some implications of a null value $\boldsymbol{\theta}_0$ falling on the boundary of $\boldsymbol{\Omega}$. Feng and McCulloch (1994) discuss the MLRT in the case of a simple mixture model where the likelihood is unbounded so that the global maximum does not exist.

4.8 Discussion

In this chapter we have discussed the standard frequentist approach to hypothesis testing. Discussion of alternative approaches based on decision theory and Bayesian ideas is deferred until Chapters 6 and 7, respectively. A number of topics, both elementary and advanced, have been omitted entirely.

At the elementary level, some of the more practical aspects of hypothesis testing, while indisputably important, do not fall naturally into the framework of this text. These include the choice of sample size to achieve a desired power, and the distinction between practical significance and statistical significance.

At an advanced level, details have been omitted for many of the more complicated techniques, for example in Sections 4.4.2 and 4.7.3. A further topic for which no details have been given is the comparison of **non-nested** composite hypotheses. Work in this area dates back to Cox (1961). One type of approach to problems of this type is to optimize a criterion such as Akaike's information criterion (AIC)—Akaike (1973). Such criteria are based on the likelihood function, but also include a penalty function, which increases as the number of parameters in the model increases. Optimization of AIC, or similar criteria, therefore involves a trade-off between those models which give a good fit to the data in terms of likelihood, and those models which are parsimonious (have few parameters). Morgan (2000, Section 4.5) gives an example of the use of AIC, while noting that simulation studies have demonstrated a superior performance for the Bayesian Information Criteria (BIC) which has a different penalty function to that of AIC.

4.9 Exercises

Exercise 4.1: Use the Neyman–Pearson lemma to find the form of the critical region for the best test of H_0 against H_1 when

(a) $X_1, X_2, \ldots, X_n$ are a random sample from a Poisson distribution with mean θ, and $H_0\colon \theta = \theta_0$, $H_1\colon \theta = \theta_1, \theta_1 > \theta_0$.

T (b) $X_1, X_2, \ldots, X_n$ are a random sample from the exponential distribution with p.d.f. $f(x;\theta) = (1/\theta)e^{-x/\theta}, x > 0$, and $H_0\colon \theta = \theta_0$, $H_1\colon \theta = \theta_1 > \theta_0$.

T (c) $X_1, X_2, \ldots, X_n$ are a random sample from the exponential distribution with p.d.f. $f(x;\theta) = \theta e^{-\theta x}, x > 0$, and $H_0\colon \theta = \theta_0, H_1\colon \theta = \theta_1, \theta_1 > \theta_0$.

(d) $X_{11}, X_{12}, \ldots, X_{1n_1} \sim N(\mu_1, \sigma_1^2), X_{21}, X_{22}, \ldots, X_{2n_2} \sim N(\mu_2, \sigma_2^2)$, all X_{ij} are independent of each other, σ_1^2, σ_2^2 are known, and $H_0\colon \mu_2 = \mu_1, H_1\colon \mu_2 = \mu_1 + \delta$ with $\delta > 0$ (μ_1 and δ are both known constants).

Exercise 4.2: In Exercise 1(d), suppose that $\sigma_1^2 = \sigma_2^2 = \delta = 1, n_1 = n_2 = n$, and that we wish to perform a best test with $\alpha = 0.01$. Find

(a) the power of the test when $n = 10$;

(b) the smallest value for n, for which we can achieve a power ≥ 0.95.

Exercise 4.3: A random sample $X_1, X_2, \ldots, X_n$ is taken from the gamma distribution with p.d.f.

$$f(x;\theta) = \frac{\theta^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\theta x}, \quad x > 0,$$

where $\lambda > 0$ is known, but $\theta > 0$ is an unknown parameter.

Show that the critical region for the most powerful α-level test of $H_0\colon \theta = \theta_0$ against $H_1\colon \theta = \theta_1$, where $\theta_1 > \theta_0$, is of the form $\sum_{i=1}^n X_i \leq C$, for some constant C. Show also that when $n = 1$ and $\lambda = 1$, the power of this test is $1 - (1-\alpha)^{\theta_1/\theta_0}$.

Ex **Exercise 4.4:** The r.v. X has p.d.f. $f(x)$, the functional form of which is unknown. A random sample of size n is drawn to test the null hypothesis

$$H_0\colon f(x) = f_0(x),$$

against the alternative

$$H_1\colon f(x) = f_1(x).$$

The functional forms of f_0 and f_1 are known. They have no unknown parameters and they have the same domain.

By considering the p.d.f.

$$\lambda f_0(x) + (1-\lambda) f_1(x),$$

show that H_0 and H_1 may be expressed parametrically. Hence show that if

$$f_0(x) = (2\pi)^{-1/2}\left(-\tfrac{1}{2}x^2\right), \quad -\infty < x < \infty,$$

$$f_1(x) = \tfrac{1}{2}\exp(-|x|), \quad -\infty < x < \infty,$$

then the best critical region for the test of H_0 against H_1 is given by

$$\sum_{i=1}^n (|x_i| - 1)^2 \geq k, \quad \text{for some constant } k.$$

Find the probabilities of type I and type II errors for the best critical region for $n = 1$ with (a) $k = 1$ and (b) $k = \frac{1}{4}$. In the case (a) is the test unbiased?

Exercise 4.5: $X_1, X_2, \ldots, X_n$ are a random sample from the gamma distribution with parameters 3 and θ, which has p.d.f.

$$f(x;\theta) = \frac{1}{2\theta^3}x^2 e^{-x/\theta}, \quad x > 0.$$

(a) Find the best test of H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ where $\theta_1 > \theta_0$.

(b) Use the result $2\sum_{i=1}^n X_i/\theta \sim \chi^2_{(6n)}$ to construct the best test of H_0 against H_1 when $n = 4$, $\theta_0 = 1$, $\theta_1 = 3$ and $\alpha = 0{\cdot}05$. Calculate (approximately) the power of this test, that is, find the probability of falling in the critical region when H_1 is true.

(c) Use (a) to find the UMP test of H_0 against H_1': $\theta > \theta_0$.

Exercise 4.6: $X_1, X_2, \ldots, X_n$ form a random sample from a uniform distribution on $[0, \theta]$. Find the form of a best test of size α for H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$, where $\theta_1 > \theta_0$. Suppose now that the alternative hypothesis is H_1: $\theta > \theta_0$; show that there exists a UMP test, and plot the power function of such a test.

Exercise 4.7: A random sample $X_1, X_2, \ldots, X_n$ is taken from the distribution with p.d.f.

$$f(x;\theta) = \frac{1}{(2\pi)^{1/2}\theta x}\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)}{\theta}\right)^2\right\}, \quad x > 0,$$

where $\theta > 0$. Show that there is a UMP test of the null hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta > \theta_0$, and find the form of this test.

Exercise 4.8: Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of observations on a r.v. X, which takes values only in the range (0, 1). Under the null hypothesis H_0, the distribution of X is uniform on (0, 1), whereas under an alternative hypothesis, H_1, the distribution is the truncated exponential with p.d.f.

$$f(x;\theta) = \frac{\theta e^{\theta x}}{e^\theta - 1}, \quad 0 \le x \le 1, \ \theta > 0,$$

where θ is unknown. Show that there is a UMP test of H_0 vs H_1 and find, approximately, the critical region for such a test when n is large.

Exercise 4.9: For a random sample of observations from a Poisson distribution, as in Exercise 1(a), consider testing H_0 against the composite alternative H_1': $\theta \ne \theta_0$. Show that no UMP test exists for H_0 against H_1'.

An intuitively plausible acceptance region for H_0 against H_1' above is of the form $C_1 < \sum_{i=1}^n X_i < C_2$, where C_1, C_2 are suitably chosen integers such that $C_2 > C_1 + 1$. Using the result that $Y = \sum_{i=1}^n X_i$ has a Poisson distribution with mean $n\theta$, show that the power function of the test with this acceptance region has the form

$$1 - e^{-n\theta}\sum_{y=C_1+1}^{C_2-1}(n\theta)^y/y!.$$

Exercise 4.10: Suppose that $X_1, X_2, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution with both parameters unknown. We wish to test the null hypothesis H_0: $\mu = 0$, $0 < \sigma^2 < \infty$ against the alternative H_1: $\mu \ne 0$, $0 < \sigma^2 < \infty$. Find the form of

(a) a UMP similar test,

(b) a UMP invariant test.

Exercise 4.11: Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and variance σ^2. Show that the MLRT of the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ against the alternative H_1: $\sigma^2 \ne \sigma_0^2$ has a test statistic

$$\left(\frac{Q}{n}\right)^{n/2} e^{(n-Q)/2} \quad \text{where } Q = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma_0^2}.$$

Show that this test has a critical region consisting of both large and small values of Q, that is, H_0 is rejected for $Q \le C_1$ or $Q \ge C_2$ where $C_1 < C_2$ are appropriately chosen constants. By considering the minimum value of the power function of such a test, derive conditions that C_1 and C_2 must satisfy if the test is to be unbiased.

Exercise 4.12: A survey of the use of a particular product was conducted in four areas, with a random sample of 200 potential users interviewed in each area. The results were that in the four areas, respectively, x_1, x_2, x_3, and x_4 of the 200 interviewees said that they used the product. Construct an MLRT to test whether the proportion of the population using the product is the same in each area.

Carry out the test, with $\alpha = 0.05$, when $x_1 = 76$, $x_2 = 53$, $x_3 = 59$, and $x_4 = 48$, using the large sample approximation for the distribution of the test statistic.

Exercise 4.13: Random samples are available from k exponential distributions; the ith distribution has parameter θ_i, and the sample from this distribution is of size n_i. Construct a MLRT of the null hypothesis $\theta_1 = \theta_2 = \cdots = \theta_k$ against a general alternative, and show how an approximate test may be performed by comparing the value of a suitable statistic with tabulated values of a χ^2 distribution. Show also that, if $k = 2$ and $n_1 = n_2$, an exact test based on an F-distribution is possible.

Exercise 4.14: Random samples, of size n_1 and n_2 respectively, are drawn from the distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, all four parameters being unknown. Suppose that $\sigma_2^2/\sigma_1^2 = c$, and we wish to test the null hypothesis H_0: $c = c_0$, for some specified c_0. Construct a MLRT for H_0 against the two-sided alternative H_1: $c \ne c_0$ and, in the particular case when $n_1 = n_2$, show that the test may be performed by comparing the observed value of a suitable statistic with a standard tabulation of critical values of the F-distribution.

Exercise 4.15: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a distribution with p.d.f.

$$f(x;\theta) = \begin{cases} c(\theta)d(x), & a \le x \le b(\theta), \\ 0, & \text{elsewhere,} \end{cases}$$

where $b(\theta)$ is a monotonically increasing function of the single parameter θ. Show that the MLRT statistic for testing H_0: $\theta = \theta_0$ against the two-sided alternative H_1: $\theta \ne \theta_0$ is given by

$$W = -2\ln(\lambda) = -2\ln\left[\int_a^{x_{(n)}} c(\theta_0)d(x)\,dx\right],$$

where $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$.

Given that

$$f(x;\theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 \le x \le \theta, \\ 0, & \text{elsewhere,} \end{cases}$$

obtain the MLRT statistic W, as above, and show that when H_0 is true W follows, exactly, a χ^2 distribution on two degrees of freedom.

Exercise 4.16: Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from two exponential distributions with unknown parameters λ and μ, that is, the Xs and Ys have, respectively, p.d.fs

$$f(x;\,\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0,$$
$$g(y;\,\mu) = \mu e^{-\mu y}, \quad y \geq 0.$$

Show that the critical region for the MLRT of the null hypothesis $H_0\colon \lambda = \mu$ against the alternative hypothesis $H_1\colon \lambda \neq \mu$ depends on the data only through the ratio $\bar{y}/\bar{x}$ of the sample means.

Given that $2\lambda \sum_{i=1}^{m} X_i$ and $2\mu \sum_{j=1}^{n} Y_j$ each has a χ^2 distribution, with $2m$ and $2n$ degrees of freedom respectively, derive the distribution of $\bar{Y}/\bar{X}$. Use this distribution to find an unbiased test of H_0, above, against the one-sided alternative $H_1'\colon \lambda > \mu$.

Exercise 4.17: A random sample of n observations $X_1, X_2, \ldots, X_n$ is taken from a Poisson distribution with standard deviation θ (i.e. the usual parameter of the distribution is θ^2). Construct a score test for the null hypothesis $H_0\colon \theta = \theta_0$ against the alternative $H_1\colon \theta > \theta_0$, and show that this test is UMP.

If $\theta_0 = 1$ and $n = 5$, find a critical region for the above test with size $\alpha = 0.0318$. What is the power of this test when $\theta = 2$?

Exercise 4.18: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from $N(\mu, \sigma^2)$ where μ and σ^2 are both unknown. Use the natural parameterization for $N(\mu, \sigma^2)$ to find an expression for $-2\ln(\lambda)$ in a MLRT of $H_0\colon \mu = \mu_0$ vs $H_1\colon \mu \neq \mu_0$—see Section 4.6.2.

Exercise 4.19: Using a random sample of size n from a Poisson distribution with mean θ, it is required to test $H_0\colon \theta = \theta_0$ against $H_1\colon \theta \neq \theta_0$. Find the test statistic for

(a) a score test of H_0 vs H_1;
(b) a Wald test of H_0 vs H_1.

Compare these statistics with that of the MLRT for the same pair of hypotheses.

Exercise 4.20: For an exponential distribution with p.d.f. $f(x;\,\theta) = \theta\, e^{-\theta x}$, $x > 0$, construct

(a) the score test;
(b) the Wald test;

for the hypotheses $H_0\colon \theta = \theta_0$ vs $H_1\colon \theta \neq \theta_0$.

Exercise 4.21: Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a distribution with p.d.f. $f(x;\,\theta)$, where $\theta = (\theta_1, \theta_2)$. Here θ_1 is a parameter of interest and θ_2 is a nuisance parameter, and it is required to test $H_0\colon \theta_1 = \theta_{10}$ vs $H_1\colon \theta_1 \neq \theta_{10}$. The score test statistic in this case is $u^2(\theta_{10}) I_{\tilde{\theta}}^{-1}$ and the Wald test statistic is $(\hat{\theta}_1 - \theta_{10})^2 I_{\hat{\theta}_1|\hat{\theta}_2}$, where $\tilde{\theta}$, $\hat{\theta}$ are the MLEs of θ under H_0 and $H_0 \cup H_1$ respectively,

$$u(\theta_{10}) = \left.\frac{\partial \ell}{\partial \theta_1}\right|_{\theta_1 = \theta_{10}}, \qquad I_{\tilde{\theta}} = E\left[-\frac{\partial^2 \ell}{\partial \theta_1^2}\right]_{\theta = \tilde{\theta}}$$

$$I_{\hat{\theta}_1|\hat{\theta}_2} = I_{11} - I_{12}^2/I_{22} \quad \text{and} \quad I_{ij} = E\left[-\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\right]_{\theta = \hat{\theta}}.$$

Find the Wald and score test statistics for the hypotheses in Exercise 4.11.

Exercise 4.22: Show that the MLRT in Example 4.10 (Section 4.7.2) is asymptotically equivalent to the Wald and score tests.

5 Interval estimation

5.1 Introduction

In Chapters 2 and 3 we discussed point estimation at some length. However, as was noted in Chapter 1, a point estimate on its own is of little use—some measure of its precision is also necessary. This leads naturally to the idea of interval estimation, which is of great practical importance. Nevertheless, we devote less space in this book to interval estimation than to either point estimation or hypothesis testing. This is because the theory underlying interval estimation is closely related to that already covered for point estimation and hypothesis testing, and the theory is most conveniently developed in these latter contexts. In fact, there is a close relationship between interval estimation and hypothesis testing, so that many of the ideas of hypothesis testing carry over directly to interval estimation; we show this later for some of the more useful ideas. The approach used in most of this chapter is known as the frequentist approach.

After defining confidence sets in this introductory section, Section 2 discusses various ways of constructing such sets, mainly based on ideas from point estimation and hypothesis testing. Section 3 defines a number of desirable (optimal) properties of confidence sets, and Section 4 describes some problems associated with interval estimation. Some of these problems can be overcome by fiducial intervals which are discussed briefly in Section 4 or by using a Bayesian approach, which is described in detail in Chapter 7.

Suppose that we are interested in a vector of parameters $\boldsymbol{\theta}$. Then we divide the possible values of $\boldsymbol{\theta}$ into a 'plausible' region and a 'less plausible' region with the plausible region usually constructed to have a predetermined probability of including the true value of $\boldsymbol{\theta}$. Note that the region depends on data and is random, whereas $\boldsymbol{\theta}$ is fixed. More formally we have the following definition.

Definition 5.1 *Suppose that $\{f(x;\,\boldsymbol{\theta});\ \boldsymbol{\theta} \in \Omega\}$ defines a family of distributions, indexed by a vector parameter, $\boldsymbol{\theta}$, and that a random sample of observations, denoted by $\mathbf{X}$, is taken from $f(x;\,\boldsymbol{\theta})$ with $\boldsymbol{\theta}$ fixed, but unknown. If S_X is a subset of Ω, depending on $\mathbf{X}$, such that*

$$\Pr[\mathbf{X}\colon S_X \supset \boldsymbol{\theta}] = 1 - \alpha,$$

*then S_X is a **confidence set** for $\boldsymbol{\theta}$ with **confidence coefficient** $1 - \alpha$.*

latter property which causes some of the problems associated with frequentist intervals which were described earlier in the section. The post-data nature of fiducial intervals means that it is possible to evaluate probabilities concerning values of θ, given x, which is impossible in the frequential framework. The way in which Wang (2000) constructs fiducial intervals ensures that the value of θ for which the observed x is most likely (the MLE) is included in the interval, something which does not always happen with frequentist intervals. For a location or scale parameter θ, for which an associated statistic T has a probability distribution which is stochastically increasing in θ, Wang (2000) shows that the lower and upper end-points θ_L, θ_U of a fiducial interval with confidence coefficient $1-\alpha$, given an observed value t for T, are obtained as the solutions to the equations

$$P(T \geq t \mid \theta_L(t)) = \alpha_2, \qquad P(T \leq t \mid \theta_U(t)) = \alpha_1,$$

where $\alpha_1 + \alpha_2 = \alpha$. The equations giving the 'exact' interval for a binomial parameter in Section 5.2.3 are of this form. More generally, in the binomial case, Wang (2000) shows how fiducial confidence coefficients can help in understanding the complexities of coverage probabilities for the various binomial confidence intervals discussed in Example 5.5.

Another complication arises with all types of confidence sets or regions if we wish to construct several regions *simultaneously*. If the confidence level is $1-\alpha$, for each region, then the *overall confidence level* relating to the simultaneous coverage of all parameters of interest will be less than $1-\alpha$. In some circumstances this overall level may be of more interest than the confidence level for individual parameters. For example in simple linear regression, with the model $E[Y] = \beta_0 + \beta_1 x$, the usual confidence interval for $E[Y]$, based on n pairs of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, has end points

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{n-2;\alpha/2}\, s \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{1/2},$$

where $\hat{\beta}_0, \hat{\beta}_1$ are least squares estimates of β_0, β_1 (Section 3.4.2), s^2 is the residual variance, and $t_{n-2;\alpha/2}$ is a critical value of the t-distribution with $(n-2)$ degrees of freedom (Hogg and Tanis, 1993, Section 8.5). This interval has probability $(1-\alpha)$ of including $E[Y]$ when we consider *one single value of* x. However, if we require a probability $1-\alpha$ of *simultaneously* including $E[Y]$ in intervals constructed *for all possible* x, we must replace $t_{n-2;\alpha/2}$ by $[2F_{2,n-2;\alpha}]^{1/2}$—see Miller (1966, p. 111).

Yet another problem with confidence sets or regions is the lack of exact confidence regions in examples including nuisance parameters. Weerahandi (1995) addresses this problem by introducing the notion of **generalized confidence intervals**, in which the requirements placed on the behaviour of the interval in repetitions of the *same* experiment are relaxed. Weerahandi's (1995) approach also provides an alternative resolution of the problem in Example 5.7.

Finally, we note that other types of intervals, such as **tolerance intervals** and **prediction intervals**, each with their own interpretation, may be useful in some circumstances—see Vardeman (1992).

5.5 Exercises

Exercise 5.1: $X_1, X_2, \ldots, X_n$ are a random sample from the exponential distribution with p.d.f. $f(x;\theta) = (1/\theta)e^{-x/\theta}, x > 0$. Using the result that $Y = 2\Sigma_{i=1}^{n} X_i/\theta$ has a $\chi^2_{(2n)}$ distribution, construct a confidence interval for θ based on the pivotal quantity Y.

Exercise 5.2: Find a confidence interval for the variance of the exponential distribution given in Exercise 5.1.

Exercise 5.3: Independent normally distributed r.vs $X_1, X_2, \ldots, X_n$ are such that X_k has expectation $\theta f(k)$ and variance $g(k)$, where $f(k)$ and $g(k)$ are known functions of k, $k = 1, \ldots, n$. Find a sufficient statistic for θ, and from it construct a two-sided 95% confidence interval for θ.

If $f(k) = g(k) = k$, $k = 1, 2, \ldots, n$, what is the smallest value of n for which the length of this confidence interval is less than 0.5?

Exercise 5.4: Use the MLE for θ (see Exercise 3.29) to find an approximate $100(1-\alpha)\%$ confidence interval for θ, given a random sample of size n from the distribution with p.d.f.

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \ \theta > 0.$$

Exercise 5.5: $X_1, X_2, \ldots, X_n$ are a random sample from a normal distribution with known mean μ and unknown variance σ^2. Three possible confidence intervals for σ^2 are:

(a) $\left(\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{a_1}, \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{a_2}\right)$

(b) $\left(\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{b_1}, \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{b_2}\right)$

(c) $\left(\frac{n(\bar{X}-\mu)^2}{c_1}, \frac{n(\bar{X}-\mu)^2}{c_2}\right)$,

where $a_1, a_2, b_1, b_2, c_1, c_2$ are constants.

Find values of these six constants which give confidence coefficient 0.90 for each of the three intervals when $n = 10$ and compare the expected widths of the three intervals in this case.

With $\sigma^2 = 1$, what is the value of n required to achieve a 90% confidence interval of width less than 2 in cases (b) and (c) above?

Exercise 5.6: For a random sample of size n from an exponential distribution with mean θ, find a pivotal quantity based on the approximate large sample distribution of the MLE for θ. Hence show that

$$\left(\frac{\sqrt{n}\bar{X}}{(\sqrt{n}+z_{\alpha/2})}, \frac{\sqrt{n}\bar{X}}{(\sqrt{n}-z_{\alpha/2})}\right)$$

provides an approximate confidence interval for θ, with confidence coefficient $1-\alpha$.

Exercise 5.7: Use the results of Exercise 3.18 to find an approximate confidence interval for μ, when μ is both the mean and standard deviation of a normal distribution.

Exercise 5.8: For the data given in Exercise 3.23 find an approximate 95% confidence interval for θ. Suppose that you are able to ascertain the true response for a random sample of n men. Estimate how large n must be in order to achieve a 95% confidence interval for θ with the same width as that obtained above.

Exercise 5.9: A random sample $X_1, X_2, \ldots, X_n$ is obtained from a distribution whose p.d.f. depends upon a scalar parameter θ. Given that $\ell(\theta; x)$ is the log likelihood function, write down the asymptotic distribution of $\partial\ell/\partial\theta$, and show how it can be used to provide a pivotal quantity.

Hence, or otherwise, derive an asymptotically valid confidence interval for σ using a sample of size n drawn from the normal distribution with mean zero and standard deviation σ.

Exercise 5.10: One success occurs in 10 trials of a binomial experiment. By considering the simple null hypothesis H_0: $p = p_0$, and the fact that a confidence interval can be constructed by finding values of p_0 for which H_0 is accepted, find an 'exact' two-sided confidence interval for p, the probability of success, with confidence coefficient 0.95—see the continuation of Example 5.5 in Section 5.2.3.

Exercise 5.11: Given that X, a Poisson r.v. with mean θ, takes the value x, show that a confidence interval (θ_L, θ_U) has confidence coefficient $1-\alpha$ when θ_L, θ_U satisfy the equations

$$\sum_{i=x}^{\infty} \frac{e^{-\theta_L}\theta_L^i}{i!} = \frac{\alpha}{2}, \qquad \sum_{i=0}^{x} \frac{e^{-\theta_U}\theta_U^i}{i!} = \frac{\alpha}{2}.$$

Exercise 5.12: Suppose that Y_{2k} has a χ^2-distribution with $2k$ degrees of freedom, and hence p.d.f.

$$f(x) = \frac{x^{k-1}e^{-x/2}}{2^k(k-1)!} \qquad x > 0.$$

By integrating $f(x)$ by parts, show that, for any $c > 0$,

$$\Pr[Y_{2k} \geq c] = \frac{e^{-c/2}(c/2)^{k-1}}{(k-1)!} + \Pr[Y_{2(k-1)} \geq c],$$

where $Y_{2(k-1)} \sim \chi^2_{2(k-1)}$.

Use this result together with an expression for $\Pr[Y_2 \geq c]$, where $Y_2 \sim \chi^2_2$, to derive a relationship between the c.d.fs of the χ^2 and Poisson distributions.

Deduce from this relationship and from Exercise 5.11 that the interval

$$\left(\tfrac{1}{2}\chi^2_{2x;(1-\alpha/2)}, \tfrac{1}{2}\chi^2_{2x+2;\alpha/2}\right)$$

based on a single observation, x, from a Poisson distribution, gives a confidence interval, with confidence coefficient $1-\alpha$, for the mean of that distribution.

Exercise 5.13: The r.v. X_1 has a Poisson distribution with parameter θ_1 and, independently, the r.v. X_2 has a Poisson distribution with parameter θ_2. By considering the conditional distribution of X_1, given the value of $X_1 + X_2$, show how a test of the null hypothesis that $\theta_1 = k\theta_2$, for fixed k, reduces to testing that the 'success probability' parameter of a binomial distribution is $k/(1+k)$.

Modifications are made to a machine, which runs continuously (except for breakdowns, which result in a negligible loss of running time), in an attempt to reduce the number of breakdowns which occur. In a one-month period prior to the modifications 15 breakdowns occurred, while in a three-month period after the modifications, 20 breakdowns occurred. Find an approximate 95% confidence interval for the ratio, θ_1/θ_2, of the breakdown rates before and after the modifications were made.

Exercise 5.14: A r.v., X, having a Poisson distribution, is observed to take the value x. Using the result that, for large θ, $X \dot{\sim} N(\theta, \theta)$, obtain a quadratic equation in θ whose roots give the end points of a confidence interval for the mean, θ, of the distribution.

The following table gives the frequency distribution of the number of breakdowns in a year for 550 army vehicles. Assuming the data arise as a random sample from a Poisson distribution, find a confidence interval with approximate confidence coefficient 95% for the expected number of breakdowns per vehicle per year.

Number of breakdowns	0	1	2	3	4	5
Number of vehicles	295	190	53	5	5	2

Exercise 5.15: Random samples, each of size n, are drawn from the distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, all four parameters being unknown. Use the results of Exercise 4.14 to find a confidence interval for σ_1^2/σ_2^2.

Exercise 5.16: Given that λ, μ are parameters of two exponential distributions as defined in Exercise 4.16, use the results of that Exercise to find a lower confidence limit for λ/μ.

Exercise 5.17: Show that the two confidence intervals for p found in Example 5.5 in Section 5.2.2 can be derived by inverting the score test and the Wald test for H_0: $p = p_0$ vs H_1: $p \neq p_0$ (see Example 4.10).

Exercise 5.18: Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution on the range $[0, \theta]$. Use a pivotal quantity based on the sufficient statistic $Y = X_{(n)}$, to derive a family of $100(1-\alpha)\%$ confidence intervals for θ.

Show that the shortest $100(1-\alpha)\%$ confidence interval for θ in this family is of length $Y(\alpha^{-1/n} - 1)$.

Exercise 5.19: Show that the critical region C provides an unbiased test of H_0: $\theta = \theta_0$ if and only if the confidence set for θ based on C is unbiased.

When the cost of each observation (c) is small, a large number of observations are usually taken. The consequence is that K_1 will be small and K_2 large. It can then be shown (Exercise 6.33) that

$$K_1 \approx \frac{-\Pr[H_0]c}{\Pr[H_1]b\mu_0} \quad \text{and} \quad K_2 \approx \frac{\Pr[H_0]a\mu_1}{\Pr[H_1]c}.$$

The SPRT tends to overshoot the thresholds, K_1 and K_2, when it stops. That is, the test stops with $\mathrm{LR}(n) < K_1$ or $\mathrm{LR}(n) > K_2$, and not $\mathrm{LR}(n) = K_1$ or $\mathrm{LR}(n) = K_2$. It is because of the 'overshoot' that results given in Properties 6.2–6.4 are approximations, rather than exact equalities. The approximations will usually be quite accurate as the overshoot is generally small.

6.8 Exercises

Exercise 6.1: The risks for five decision procedures $\delta_1, \delta_2, \ldots, \delta_5$ depend on the value of a positive-valued parameter θ. The risks are given in the table below.

	δ_1	δ_2	δ_3	δ_4	δ_5
$0 \le \theta < 1$	10	10	7	6	8
$1 \le \theta < 2$	8	11	8	5	10
$2 \le \theta$	15	11	12	14	14

(a) Which decision procedures are at least as good as δ_1 for all θ?
(b) Which decision procedures are admissible?
(c) Which is the minimax procedure?
(d) Suppose $\theta \sim U[0, 5]$. Which is the Bayes procedure and what is the Bayes risk for that procedure?

Exercise 6.2: Suppose $\theta_1, \theta_2, \ldots, \theta_k$ are the only possible values of θ and that $p(\theta_i) > 0$ for $i = 1, 2, \ldots, k$ (k finite), where $p(\theta)$ is the prior distribution. Show that the Bayes estimator is admissible.

Exercise 6.3: Suppose θ^* is the Bayes estimator of θ both if $p_1(\theta)$ is the prior distribution and if $p_2(\theta)$ is the prior distribution. Show that θ^* is then also the Bayes estimator for the prior distribution $p(\theta) = \alpha p_1(\theta) + (1-\alpha)p_2(\theta)$, $0 \le \alpha \le 1$.

Exercise 6.4: Suppose the posterior distribution of θ is

$$q(\theta) = 4\theta - 3\theta^2, \quad 0 \le \theta \le 1.$$

(a) Show that the point estimator of θ for an absolute error loss function is $\hat{\theta} \approx 0.597$.
(b) Show that $\hat{\theta} \approx 0.567$ is the point estimator of θ under the loss function

$$L_S(\theta, \hat{\theta}) = \begin{cases} 0, & |\hat{\theta} - \theta| < 0.1 \\ a, & |\hat{\theta} - \theta| \ge 0.1, \end{cases}$$

where $a > 0$.
(c) Find the point estimator of θ for a quadratic loss function.

Exercise 6.5: A baker has to decide how many loaves to bake tomorrow. If he bakes too many, he loses k_1 pence for each loaf he has left. If he bakes too few, he loses k_2 pence in profit on each loaf that he could have sold. Suppose his opinion about θ, the number of loaves he could sell tomorrow (if he does not run out) is represented by the p.d.f. $f(\theta)$ whose distribution function is $F(\theta)$. Show that $\hat{\theta}$, the number of loaves he should bake tomorrow, satisfies $F(\hat{\theta}) = k_2/(k_1 + k_2)$ if $k_2/(k_1 + k_2)$ is an integer.

Exercise 6.6: Let θ denote the unknown variance of the variable X. θ is to be estimated from a sample of values of X using an estimator of the form $\hat{\theta} = bs^2$, where s^2 is the sample variance. It has been decided to use the loss function

$$L_S(\theta, \hat{\theta}) = \frac{\hat{\theta}}{\theta} - 1 - \ln\left(\frac{\hat{\theta}}{\theta}\right), \quad \hat{\theta} > 0, \ \theta > 0.$$

(a) Show that the risk function has the form

$$R(\theta, \hat{\theta}) = b - \ln(b) - 1 - c,$$

where c does not depend on b.
(b) Show that $R(\theta, \hat{\theta})$ is minimized (for estimators of the form $\hat{\theta} = bs^2$) when $b = 1$.

Exercise 6.7: Evaluate the following integrals by comparing them to standard distributions

(a) $\int_0^\infty \frac{\Gamma(\alpha)}{\Gamma(\alpha+\beta)} x^\alpha e^{-\beta x+3}\, dx$,
(b) $\int_0^1 15x^{10}(1-x)^5\, dx$.

Exercise 6.8: Suppose X is an observation from the distribution

$$f(x;\, \theta) = (x-1)\theta^2(1-\theta)^{x-2}, \quad x = 2, 3, \ldots; \ 0 < \theta < 1.$$

The prior distribution for θ is a beta distribution:

$$p(\theta) = \frac{\Gamma(7)}{\Gamma(4)\,\Gamma(3)}\theta^3(1-\theta)^2, \quad 0 \le \theta \le 1,$$

and the loss from estimating θ by $\hat{\theta}$ is $(\theta - \hat{\theta})^2$.

(a) Find the posterior distribution, $q(\theta;\, x)$.
(b) Show that the Bayes estimator of θ is $6/(7+x)$.
(c) Find the Bayes risk associated with this estimator.

Exercise 6.9: Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, σ^2 known, and let the prior distribution be $\theta \sim N(\phi, \tau^2)$. Suppose θ is to be estimated under a quadratic loss function.

(a) State the Bayes estimator of θ and show that this estimator has constant risk as $\tau \to \infty$.
(b) Deduce the minimax estimator of θ and state its risk.

Exercise 6.10: Suppose the prior distribution for θ is to be modelled by a normal distribution. An expert assessed 20 as his median estimate of θ and 25 for its upper quartile. Determine the normal distribution that is consistent with his assessments.

Exercise 6.11: Suppose $\theta > 0$ and the prior distribution for θ is to be modelled by a gamma distribution. Give the distribution if the prior mean is 20 and the prior standard deviation is 10.

Exercise 6.12: Suppose θ is the probability of 'success' in a trial and the prior distribution for θ is to be modelled by a beta distribution. Before observing any trials, the prior mean of θ is 0.4 while, after observing 10 independent trials in which there are 7 successes, the posterior mean of θ is 0.5. Determine the prior distribution (before observing the 10 trials).

Exercise 6.13: Let $x_1, x_2, \ldots, x_n$ denote a random sample from the Weibull distribution with index 2:

$$f(x;\,\theta) = 2\theta^2 x \exp(-\theta^2 x^2), \quad (x > 0),$$

where θ is an unknown positive parameter. Assume that the prior distribution for θ is the improper distribution

$$p(\theta) = \text{constant}.$$

Find the mode of the posterior distribution of θ in terms of $x_1, x_2, \ldots, x_n$.

Exercise 6.14: Suppose $f(x;\,\theta) = \theta a x^{a-1}\exp(-\theta x^a), (x > 0, a > 0)$, where θ is an unknown positive parameter and a is a known constant.

(a) Show that this distribution is a member of the one-parameter exponential family of distributions.

(b) Hence suggest a conjugate prior distribution for θ.

Exercise 6.15: The time to failure of a particular type of component follows an exponential distribution with mean $1/\theta$.

(a) Determine the probability that a component will be working after 10 h.

In a trial, the times to failure (in hours) of five components were 4, 1, 3, 3, and 6, and a sixth component was still working after 10 h when the trial was terminated. Prior to the trial, opinion about the value of θ corresponded to the gamma distribution:

$$p(\theta) = \tfrac{1}{2}\theta^2 e^{-\theta}, \quad \theta > 0.$$

(b) Show that the posterior distribution of θ is

$$q(\theta \,|\, \text{data}) = \frac{28^8}{7!}\theta^7 e^{-28\theta}, \quad \theta > 0.$$

Exercise 6.16: Suppose $X_1, X_2, \ldots, X_n$ are random independent observations from a Poisson distribution with unknown mean θ.

(a) Give the likelihood for θ.

(b) Form the conjugate prior distribution and identify it.

(c) Determine the posterior distribution.

(d) State the mean and variance of the prior distribution and the prior expectation of the variance of $\bar{X} = \sum X_i/n$. Show that the posterior mean is a weighted average of the prior mean and the sample mean, with weights proportional to the reciprocals of the prior variance and $E[\text{Var}(\bar{X})]$.

(e) Suggest an uninformative prior distribution for θ.

Exercise 6.17: A large population must be screened for the presence of a certain antibody in the blood. An infallible test on a blood sample from a single person costs c_1 and gives a positive result if the antibody is present and a negative result otherwise. If a batch of blood samples from n people are mixed together and then tested, the test costs $c_1 + c_2 n$ and gives a negative result only if none of the n samples contained the antibody. If a positive result is obtained then all n samples have to be tested individually. The proportion θ of people who have the antibody has prior distribution

$$p(\theta) = \beta(1-\theta)^{\beta-1} \quad \text{with } \beta \text{ large.}$$

(a) Show that individual testing gives a lower expected cost than testing in batches of size n if

$$c_2 > c_1\left(\frac{\beta}{n+\beta} - \frac{1}{n}\right).$$

(b) Show that if batch testing is adopted, the optimal value of n is approximately $\sqrt{\beta}$.

(c) As screening goes along, suggest how increasing knowledge about θ can be used to modify the value of n used in successive screening tests.

Exercise 6.18: When $X_1, X_2, \ldots, X_n$ is a random sample from each of the following distributions, determine the Fisher information, I_θ, and hence find Jeffreys prior distribution for θ.

(a) The geometric distribution, $f(x;\,\theta) = \theta(1-\theta)^{x-1}$, $x = 1, 2, \ldots$.

(b) The Poisson distribution, $f(x;\,\theta) = e^{-\theta}\theta^x/x!$, $x = 1, 2, \ldots$.

(c) The gamma distribution, $f(x;\,\theta) = \frac{1}{2}\theta^3 x^2 e^{-x\theta}$, $x > 0$.

Exercise 6.19: Suppose a random sample of size 12 is taken from a normal distribution with a variance of 15 and an unknown mean, μ. H_0: $\mu = 4$ is to be tested against H_1: $\mu = 1$. If $p_0 = 0.1, p_1 = 0.9$, and $a = b$, should H_0 be rejected when the sample mean is 2.0? Would the result be the same if a sample of size 24 had given the sample mean of 2.0?

Exercise 6.20: Show that δ_7 is the Bayes test for $0.177 \leq p_0 \leq 0.462$ in Example 6.7. For what value of p_0 is δ_5 the Bayes test?

Exercise 6.21: Suppose we have a single observation, x_1, which comes from a distribution with p.d.f. $f(x)$, and we want to test

$$H_0\colon f(x) = 2(1-x), \quad 0 \leq x \leq 1,$$

against

$$H_1\colon f(x) = 2x, \quad 0 \leq x \leq 1.$$

Show that the best critical region for the likelihood ratio test of H_0 vs H_1 is given by $x_1 \geq B$ for some constant B. Given that the losses incurred when Type I and Type II errors occur are equal, find the values of B which give

(a) the minimax procedure;

(b) the Bayes procedure, when the prior probabilities are $\frac{1}{4}, \frac{3}{4}$ for H_0 and H_1, respectively.

Find the values of the Type I and Type II errors in (a) and (b), and find also the prior probabilities of H_0 and H_1 for which the Bayes procedure is equivalent to the minimax procedure.

Exercise 6.22: A random sample $X_1, X_2, \ldots, X_n (n \geq 1)$ is to be taken from a uniform distribution whose range is $(0, \theta)$. One of the hypotheses, H_0: $\theta = 1$ or H_1: $\theta = 3$, holds and the prior probabilities of these hypotheses are $\frac{1}{5}$ and $\frac{4}{5}$, respectively. On the basis of a sample of size n, H_0 must be accepted or rejected. There is zero loss for a correct decision, a loss of one if H_0 is incorrectly rejected and a loss of ten if H_0 is incorrectly accepted.

(a) Show that for $n \leq 3$, the Bayes test always rejects H_0 while, for $n \geq 4$, H_0 is accepted if no observation in the range $(1, 3)$ is observed.

(b) For each n, determine the expected risk for the Bayes test.

Exercise 6.23: Two outwardly identical measuring instruments M_1 and M_2 have normally distributed errors with variance σ_1^2 and σ_2^2, respectively, where $\sigma_1^2 < \sigma_2^2$. A technician brings you one of the instruments, but can shed no light on which instrument he has brought. You therefore make n observations $x_1, x_2, \ldots, x_n$ of a fixed quantity θ and calculate $U = \sum(x_i - \bar{x})^2$. Suppose the loss from incorrectly deciding the instrument is M_1 equals the loss from incorrectly deciding it is M_2. Show that you will decide it is M_2, if and only if

$$U > 2(n-1)(\ln(\sigma_2) - \ln(\sigma_1))/(\sigma_1^{-2} - \sigma_2^{-2}).$$

An alternative approach to the problem is to argue on intuitive grounds that the decision rule should take the form: decide it is M_2 if $U > k$, and choose k to minimize the probability of making a mistake. Show this leads to the same rule as above. (Note $U/\sigma_i^2 \sim \chi^2_{(n-1)}$ if M_i is the instrument being used.)

Exercise 6.24: Consider a binomial experiment with two trials, where H_0: $p = 0.2$ is to be tested against H_1: $p = 0.7$. Suppose the loss is zero for a correct decision and five for an incorrect decision (wrongly rejecting H_0 or wrongly accepting H_0). Let p_0 be the prior probability that H_0 is true. State the possible critical regions and, for each region, determine the range of values for p_0 for which that region is the critical region. If $p_0 = 0.4$, what is the expected risk for the Bayes test?

Exercise 6.25: Suppose $f(x; \theta) = \theta e^{-\theta x}, x < 0$. Two alternative theories give values for θ. The first theory, H_0, states $\theta = 1$ and the second theory H_1, states $\theta = 2$. The prior probabilities are $\Pr(H_0) = \frac{2}{5}$ and $\Pr(H_1) = \frac{3}{5}$.

(a) Given a single datum, x, show that $e^x/(3+e^x)$ is the posterior probability that H_0 is the correct theory.

A decision as to which hypothesis holds must be made, the loss being 10 for an incorrect decision with no loss for a correct decision.

(b) If a decision is to be made without taking any data, which theory should be chosen and what is the expected loss?

(c) If one datum x is taken and then a decision made, show that H_0 should be chosen if $x > 1.0986$.

(d) Suppose the cost of taking a datum is c and a decision can be made either immediately (without taking any data) or after taking just one datum. Show that the datum should be taken if $c < \frac{2}{3}$.

Exercise 6.26: A stream of observations $X_1, X_2, \ldots$ may be taken one at a time at a cost of c for each observation. The X_i are i.i.d. with $X_i \sim N(\theta, \sigma^2), \sigma^2$ known, and the prior distribution for θ is $\theta \sim N(\phi, \tau^2)$. The value of θ must be estimated under the quadratic loss function, $L_S(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$, where $\hat{\theta}$ is the estimate of θ.

Suppose n observations, $x_1, x_2, \ldots, x_n$, are taken.

(a) Give the posterior distribution of θ and the Bayes risk (including sampling costs) from stopping at this stage.

(b) Hence show that sampling should stop when

$$n \approx \frac{\sigma}{\sqrt{c}} - \frac{\sigma^2}{\tau}.$$

(Notice that we know how many observations to take before sampling starts, because the Bayes risk does not depend on their values.)

Exercise 6.27: Continue Example 6.8 and determine

(a) when sampling should be continued after one observation has been taken;

(b) whether the first observation should be taken.

Exercise 6.28: Consider the situation examined in Example 6.9 and suppose each observation costs the same amount. For this case, explain whether the fixed sample look ahead procedure is the same as the m-step look ahead procedure. Does your answer depend on the value of m?

Exercise 6.29: Consider the problem presented in Exercise 6.22. If $x_i \in (0, 1)$ for $i = 1, 2, \ldots, n$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, show that

$$\Pr[H_1 \mid \boldsymbol{x}] = 4\left(\tfrac{1}{3}\right)^n \Big/ \left\{1 + 4\left(\tfrac{1}{3}\right)^n\right\}.$$

Suppose a fixed-size sample is not taken but, instead, observations are taken sequentially at a cost of c for each observation, where c is small and positive. Show that the fixed sample size look ahead procedure stops when either

(a) an observation in the range $(1, 3)$ is obtained (in which case H_1 is accepted), or

(b) n^* observations in the range $(0, 1)$ are obtained, where n^* is the smallest integer for which

$$c > 80/(12 + 3^{n^*+1}),$$

(in which case H_0 is accepted).

Exercise 6.30: Suppose $X_1, X_2, \ldots$ is a sequential sample of i.i.d. observations with

$$f(x; \theta) = \frac{1}{x \ln(\theta)} \left(\frac{\theta - 1}{\theta}\right)^x, \quad x = 1, 2, \ldots,$$

and we wish to test H_0: $\theta = 2$ against H_1: $\theta = 4$.

(a) From Property 6.3, find an SPRT for which the probabilities of incorrectly rejecting H_0 and incorrectly rejecting H_1 are both 0.1.

(b) For this SPRT, determine the expected number of observations that will be taken if H_0 is true and if H_1 is true.

Exercise 6.31: A sequential sample of independent observations $X_1, X_2, \ldots$ is taken from the binomial distribution $B(5, \theta)$. H_0: $p = 0.2$ is to be tested against

H_1: $p = 0.3$ and the prior probabilities are $\Pr[H_0] = 0.3$ and $\Pr[H_1] = 0.7$. If the probability of incorrectly rejecting H_0 and incorrectly rejecting H_1 are both to be 0.1,

(a) determine an appropriate SPRT.

(b) For this SPRT, determine the expected sample sizes $E(N \mid H_0), E(N \mid H_1)$, and $E(N)$.

Exercise 6.32: Suggest conditions under which Property 6.2 yields the further approximations $E(N \mid H_0) = \ln(K_1)/\mu_0$ and $E(N \mid H_1) = \ln(K_2)/\mu_1$.

Exercise 6.33: Suppose $1/K_1$ and K_2 are both large and of similar orders of magnitude. Show that the Bayes risk given in Property 6.4 is then minimized when $K_1 \approx -\Pr[H_0]c/\{\Pr[H_1]b\mu_0\}$ and $K_2 \approx \Pr[H_0]a\mu_1/\{\Pr[H_1]c\}$.

Exercise 6.34: Suppose $X_1, X_2, \ldots$ is a sequential sample from a Poisson distribution with mean θ and H_0: $\theta = 2$ is to be tested against H_1: $\theta = 3$. The prior probabilities are $\frac{3}{4}$ that $\theta = 2$ and $\frac{1}{4}$ that $\theta = 3$. The loss in incorrectly accepting H_0 is 50, from incorrectly accepting H_1 is 25, and there is no loss from a correct decision. The cost of each observation is $c = 0.001$. Using the approximations in Exercise 6.33, determine

(a) K_1 and K_2 for the Bayes SPRT,

(b) the Bayes risk for this SPRT.

Exercise 6.35: Consider the situation in Exercise 6.31. Suppose each observation costs 0.01 and the loss for incorrectly rejecting H_0 or H_1 is 10.

(a) Determine the Bayes risk for the SPRT found in Exercise 6.31.

(b) Suppose the SPRT is changed so that K_1 and K_2 are given by the approximations found in Exercise 6.33. Determine the expected sample sizes $E(N \mid H_0)$ and $E(N \mid H_1)$ for this SPRT and also determine its Bayes risk.

7 Bayesian inference

7.1 Introduction

The previous chapter introduced some elements of Bayesian inference. These will be discussed further now, together with additional topics. In some cases the Bayesian approach and the frequentist (or classical) approach have clear similarities. For example, in the next section we give the Bayesian definition of sufficiency and show its equivalence to the frequentist definition. We also find that Bayesian interval estimates are often numerically equal to frequentist confidence intervals when an uninformative prior distribution is used. With other topics, such as hypothesis testing, we find marked differences between the two approaches.

The form of the prior distribution plays an important role in much of this chapter. Improper prior distributions can cause problems in hypothesis testing. Hierarchical prior distributions (introduced in Section 7.7) can be used to model structural relationships between parameters, illustrating the flexibility that results from treating unknown parameters as r.vs. Empirical Bayes methods are also discussed, in which past data are used to construct a prior distribution.

7.2 Sufficiency

The Bayesian definition of sufficiency is based on the distribution of θ, but, as we will show, it is equivalent to the frequentist definition given earlier. The notation used is the same as in Chapter 6.

Definition 7.1 *A statistic $T(X_1, X_2, \ldots, X_n)$ is sufficient for θ if and only if the posterior distribution of θ given $X_1, X_2, \ldots, X_n$ is the same as the posterior distribution of θ given T.*

Theorem 7.1 *Definitions 2.5 and 7.1 are equivalent.*

Proof Suppose that T satisfies Definition 2.5. Then, leaving out terms which do not depend on θ, we have

$$f(x;\theta) = g(x \mid t, \theta)h(t \mid \theta) \propto h(t \mid \theta),$$

because g does not depend on θ.

and H_1, respectively, and there is no loss for a correct decision. Then, from Lemma 6.1, the EB test is to reject H_0 if

$$\frac{L(\theta_1^*; x_k)}{L(\theta_0^*; x_k)} \geq \frac{\hat{p}a}{(1-\hat{p})b}. \qquad (7.28)$$

Example 7.10 For $i = 1, 2, \ldots, k$, suppose $X_i \sim N(\theta_i, \sigma^2)$ with σ^2 known and the hypotheses are H_0: $\theta_k = \mu_0$ vs H_1: $\theta_k = \mu_1$, with $\mu_1 > \mu_0$. From Example 6.6, H_0 is rejected if

$$x_k > \frac{1}{2}(\mu_0 + \mu_1) + \frac{\sigma^2}{(\mu_1 - \mu_0)} \ln\left(\frac{\hat{p}a}{(1-\hat{p})b}\right),$$

provided $\hat{p} \in (0, 1)$. If $\hat{p} = 0$, then H_0 is rejected regardless of the value of x_k, while if $\hat{p} = 1$, then H_0 is always accepted. □

A disadvantage of many EB methods is that the estimate $\hat{p}(\theta)$ is treated as if it were actually $p(\theta)$; no allowance is made for the uncertainty of the estimate. In the last example, for instance, if $\bar{x} > \mu_1$, then $\hat{p} = 0$ and H_0 is rejected for any value of x_k. This is unreasonable if $\bar{x}$ is based on a small set of data.

Forming confidence intervals is one task where allowance is sometimes made for the uncertainty in $\hat{p}(\theta)$. Let $I(\hat{p}(\theta))$ be the α% Bayesian credible interval for θ_k when $\hat{p}(\theta)$ is taken as the prior distribution for θ. The **coverage** of the interval, $C[I(\hat{p}(\theta))] = \Pr\{\theta \in I(\hat{p}(\theta))\}$, is the posterior probability, given $p(\theta)$ and the data, that the interval actually contains θ. One useful measure of the level of confidence to associate with $I(\hat{p}(\theta))$ is its expected coverage,

$$E[C[I(\hat{p}(\theta))]],$$

where the expectation is with respect to variation of $\hat{p}(\theta)$. Determining the expected coverage is usually difficult, but approximations using Taylor expansions are possible when $p(\theta)$ has a known parametric form. Some examples are given in Maritz and Lwin (1989, Chapter 6) and a practical application is given by Martz and Zimmer (1992).

In discussing EB we have assumed throughout that a single observation, x_i, is obtained for each θ_i. Clearly the methods can be extended straightforwardly to cases where there are several observations for each θ_i.

In situations where EB can be used, an obvious alternative is to use hierarchical Bayesian models, as the θ_i are exchangeable. When the two methods use the same prior distribution, $p(\theta)$, they typically give similar estimators.

7.9 Exercises

Exercise 7.1:

(a) Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. observations from a Poisson distribution with mean θ and let $p(\theta)$ be the prior distribution for θ. Give an expression for the posterior distribution of θ, $q_1(\theta; \boldsymbol{x})$, ignoring constants of proportionality.

(b) Suppose Y is an observation from a Poisson distribution with mean $n\theta$ and $p(\theta)$ is the prior distribution of θ. Show that the posterior distribution of θ, $q_2(\theta; y)$, is the same as $q_1(\theta; \boldsymbol{x})$.

(c) Use Definition 7.1 to conclude that $Y = \sum_{i=1}^n X_i$ is sufficient for θ.

Exercise 7.2: Suppose $X_1, X_2, \ldots, X_n$ form a random sample from a distribution parameterized by θ.

(a) Suppose also that, in the likelihood, the Xs can be replaced by functions $T_1(X_1, \ldots, X_n)$ and $T_2(X_1, \ldots, X_n)$. Show that T_1 and T_2 are sufficient for θ.

(b) If observations are from the gamma distribution, $X_i \sim \Gamma(\alpha, \beta)$, determine sufficient statistics for $\theta = (\alpha, \beta)$ and give a conjugate prior distribution for θ up to a constant of proportionality.

Exercise 7.3: Show that

$$f(x; \theta) = \frac{\theta^2}{\theta + 1}(x+1)e^{-x\theta}, \quad (x > 0)$$

is, for any $\theta > 0$, a p.d.f. A random sample of size n is taken from this p.d.f. giving values $x_1, x_2, \ldots, x_n$.

(a) What is the likelihood function (ignore constants of proportionality)?

(b) Give function(s) of $x_1, \ldots, x_n$ which are sufficient for θ.

(c) Give the conjugate prior distribution for θ.

(d) Obtain the posterior distribution for θ using this prior and compare its form with that of the prior distribution.

Exercise 7.4: Suppose $X_1, \ldots, X_n$ are i.i.d. with probability densities

$$f(x_i; \theta) = \theta b x_i^{b-1} \exp(-\theta x_i^b), \quad x_i > 0, \quad \theta > 0,$$

where b is a known positive constant. Obtain the natural conjugate prior distribution for θ (ignore constants of proportionality) and show that it is a conjugate prior distribution.

Exercise 7.5: Suppose $X_1, X_2, \ldots, X_n$ are a random sample from the uniform distribution $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, where $\theta \in (-\infty, \infty)$ is unknown.

(a) Show that the smallest and largest of $X_1, \ldots, X_n$ are jointly sufficient for θ.

(b) If $p(\theta) = \text{constant}$, $\theta \in (-\infty, \infty)$, is the prior distribution of θ, find its posterior distribution.

Exercise 7.6: Suppose the posterior distribution of θ is a t-distribution with 10 degrees of freedom. Give a 95% equal-tailed credible interval for θ. Is this interval also an HPD interval?

Exercise 7.7: Suppose the posterior distribution of θ is

$$q(\theta; x) = 4\theta - 3\theta^2, \quad 0 \leq \theta \leq 1.$$

Determine a 90% equal-tailed credible interval for θ.

Exercise 7.8: Suppose (a, b) is the $1-\alpha$ HPD credible interval for θ and $\phi = c + k\theta$ $(c \geq 0, k > 0)$. Show that $(c + ka, c + kb)$ is the $1-\alpha$ HPD credible interval for ϕ.

Exercise 7.9: A random sample of size 25 from $N(\theta, 1)$ has mean 0.30.

(a) Given that $p(\theta) \propto$ constant, $\theta \in (-\infty, \infty)$, determine a 95% HPD credible interval for θ.

(b) Given that θ is known to be positive, so that

$$p(\theta) \propto \text{constant}, \quad \theta > 0$$
$$p(\theta) = 0, \quad \theta \leq 0,$$

determine a 95% HPD credible interval for θ.

Exercise 7.10: Prior information on the mean θ of a Poisson distribution is such that the prior mean and prior variance of θ are both unity.

(a) Determine a conjugate prior density having these properties.

(b) Find the corresponding posterior density of θ given a random sample of n observations from the Poisson distribution.

(c) Derive expressions, in terms of n, the sample observations, and probability points of a χ^2 distribution, for the limits of the 1–α equal-tailed credible interval for θ. [*Hint*: If $f(x) \propto x^{\alpha-1}e^{\beta x}$ and $Y = 2\beta X$, then $Y \sim \chi^2(2\alpha)$.]

Exercise 7.11: Suppose the posterior distribution of θ is

$$q(\theta) = \begin{cases} \theta, & 0 \leq \theta \leq 1 \\ 2-\theta, & 1 < \theta \leq 2. \end{cases}$$

(a) Show that $(\sqrt{0.05}, 2 - \sqrt{0.05})$ is a 95% HPD credible interval for θ.

(b) Let $\phi = \theta^2$.

(i) Show that $0.05, \{2 - \sqrt{0.05}\}^2)$ is a 95% credible interval for ϕ and verify that it is not a 95% HPD interval.

(ii) Find a 95% HPD interval for ϕ.

(iii) Explain why there are an infinite number of 40% HPD credible intervals for ϕ.

Exercise 7.12: Suppose 1.03, 0.20, 1.21, 3.31, and 1.49 are a random sample of five observations from the Weibull distribution

$$f(x) = 2\theta^{-2}x e^{-(x/\theta)^2}, \quad x > 0.$$

H_0: $\theta = 2$ is to be tested against H_1: $\theta = 3$. Determine the posterior odds, Q^*.

Exercise 7.13: Suppose that X is the number of successes in a binomial experiment with n trials and probability of success θ. Either H_0: $\theta = \frac{1}{2}$ or H_1: $\theta = \frac{3}{4}$ is true. Show that the posterior probability that H_0 is true is greater than the prior probability for H_0 if and only if

$$x \ln(3) < n \ln(2).$$

Exercise 7.14: Let θ be a person's IQ and suppose that in the population $\theta \sim N(100, 225)$. A person's score on an IQ test, X, is an unbiased but imprecise measure of $\theta, X \sim N(\theta, 50)$. Jane takes the test. Let θ_J denote her true IQ and suppose she scored 120 on the test.

(a) Calculate the posterior distribution of θ_J.

(b) Define the hypotheses H_0: $\theta_J \geq 110$ and H_1: $\theta_J < 110$. Using tables for the standard normal distribution, determine the posterior odds ratio that Jane's IQ is above 110.

Exercise 7.15: Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution, $f(x; \theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Suppose we wish to test H_0: $\theta = 1$ against H_1: $\theta \neq 1$, where $\Pr[H_0] = p$ and $\Pr[H_1] = 1 - p$. If the prior distribution for θ, given H_1, is

$$p_1(\theta \mid H_1) = \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}/\Gamma(\alpha), \quad \theta \neq 1,$$

determine the posterior odds $Q^* = \Pr[H_0 \mid x]/\Pr[H_1 \mid x]$ using (a) eqn (7.10) and (b) eqn (7.13).

Exercise 7.16: Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, 1)$. There are two hypotheses, H_0: $\theta = 1$ and H_1, where $\Pr[H_0] = p$ and $\Pr[H_1] = 1 - p$.

(a) Given that H_1 specifies $\theta = -1$, show that

$$\Pr[H_0 \mid x] = \frac{p\, e^{\sum x_i}}{p\, e^{\sum x_i} + (1-p)e^{-\sum x_i}}.$$

(b) Given that H_1 specifies $\theta \neq 1$ and gives θ the prior distribution

$$p(\theta \mid H_1) = \frac{1}{\sqrt{2\pi}} \exp(-\theta^2/2), \quad \theta \neq 1,$$

determine $\Pr[H_0 \mid x]$ when $\sum x_i = n$.

Exercise 7.17: Determine the marginal posterior distributions of θ and λ when their joint distribution is

(a) $q(\theta, \lambda) = \dfrac{2}{(1+\theta+\lambda)^3}, \quad \theta > 0, \quad \lambda > 0,$

(b) $q(\theta, \lambda) = \dfrac{1}{81}\theta^2\lambda^2, \quad 0 < \theta < 3, \quad 0 < \lambda < 3,$

(c) $q(\theta, \lambda) = 21\lambda\theta^2/4, \quad \theta^2 \leq \lambda \leq 1.$

Exercise 7.18: In order to measure the intensity, θ, of a source of radiation in a noisy environment a measurement X_1 is taken without the source present and a second, independent measurement X_2 is taken with it present. It is known that X_1 is $N(\mu, 1)$ and X_2 is $N(\mu + \theta, 1)$, where μ is the mean noise level. The prior distribution for μ is $N(\mu_0, 1)$ while the prior distribution for θ is constant. (Thus μ is known with some accuracy while little is known about θ.)

(a) Write down (apart from a constant of proportionality) the joint posterior distribution of μ and θ.

(b) Hence obtain the posterior marginal distribution of θ.

(c) The usual estimate of θ is $x_2 - x_1$. Explain why $\frac{1}{2}(2x_2 - x_1 - \mu_0)$ might be better.

Exercise 7.19: The number of phone calls a man receives in a week follows a Poisson distribution with mean θ. At one point in time, the man's opinion about the value of θ corresponds to the gamma distribution:

$$p(\theta) = \tfrac{1}{54}\theta^2 e^{-\theta/3}, \quad \theta > 0.$$

In the next 4 weeks the man received 3, 7, 6, and 10 phone calls, respectively. Determine the distribution that should now represent his opinion about θ and find the predictive distribution for the number of calls he will receive in week 5.

Exercise 7.20: Independent observations $X_1, X_2, \ldots, X_{10}$ from a power distribution with p.d.f.

$$f(x; \theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \ \theta > 0$$

take the values 0.91, 0.41, 0.99, 0.05, 0.90, 0.76, 0.59, 0.98, 0.78, 0.51. If the prior distribution for θ is such that $\theta/0.075$ has a χ^2_{20} distribution (i.e. $p(\theta) \propto \theta^9 e^{-6.67\theta}$), find the posterior distribution of θ.

Find also the predictive distribution for a further independent observation, X_{11}. Given that m is the best prediction of X_{11} under a linear loss function, show that m satisfies

$$\left(\frac{12.51}{12.51 - \ln(m)}\right)^{20} = 0.5.$$

Exercise 7.21: Suppose a random sample of 16 independent observations from $N(\theta_1, 400)$ has a mean of 70 and an independent random sample of 25 observations from $N(\theta_2, 600)$ has a mean of 63. A priori nothing is known about θ_1 or θ_2, so non-informative prior distributions $p_1(\theta_1)$ = constant and $p_2(\theta_2)$ = constant are deemed appropriate.

(a) What is the posterior distribution of (i) θ_1 and (ii) θ_2?

(b) Put $\eta = \theta_1 - \theta_2$ and determine the joint posterior distribution of η and θ_2 through a change of variables.

(c) Hence find the marginal posterior distribution of η.

Exercise 7.22: The unknown means of two Poisson distributions are θ_1 and θ_2, and these have independent and identical prior distributions

$$p(\theta) \propto \theta^{\alpha_1} e^{-\alpha_2\theta}, \quad \alpha_1 > -1, \ \alpha_2 > 0.$$

A single observation from each Poisson distribution is taken, giving values x_1 and x_2, respectively. If $\eta = \theta_1/(\theta_1 + \theta_2)$, show that the posterior distribution of η is beta$(x_1 + \alpha_1 + 1, x_2 + \alpha_1 + 1)$.

Exercise 7.23: Suppose independent samples of sizes n_1 and n_2 are taken from two normal distributions with known means, μ_1 and μ_2, and unknown precisions, θ_1 and θ_2. For $i = 1, 2$, let $r_i = \sum_j (x_{ij} - \mu_i)^2$, where $\{x_{i1}, x_{i2}, \ldots, x_{in_i}\} = x_i$ are the observations from population i. Show that

$$(r_i + \alpha_i)\theta_i \sim \chi^2_{(n_i+\nu_i+2)} \quad \text{for } i = 1, 2.$$

Hence derive eqn (7.17).

Exercise 7.24: Suppose $X_1, X_2, \ldots$ is a sequential sample from a $B(1, \theta)$ distribution and we have an uninformative prior distribution for $\theta, p(\theta) = \theta^{-1}(1-\theta)^{-1}$. Suppose also that θ is to be estimated under a quadratic loss function.

(a) After observing $x_1, x_2, \ldots, x_n$, show that $\bar{x}_n = (1/n)\sum x_i$ is the Bayes's estimate of θ.

(b) Suppose the following stopping rule is used. Take X_1 and stop if $X_1 = 1$. If $X_1 = 0$, take X_2 and then stop, regardless of the value of X_2. Show that, using frequentist methods, $E[\bar{X}] = \theta + \frac{1}{2}\theta(1-\theta)$. (Thus, from a frequentist perspective, the Bayes's estimator is biased under this stopping rule. A Bayesian might argue that it is not important for an estimator to satisfy the frequentist definition of unbiasedness.)

Exercise 7.25: Suppose k different people repeatedly perform a test that results in success or failure. Let X_i be the number of successes the ith person obtains in n_i trials. Assume $X_i \sim B(n_i, p_i)$, where the p_i vary between people but assume that, a priori, the p_i are exchangeable. Construct a hierarchical model in which the p_i are independent values from a beta distribution whose parameters have an uninformative prior distribution. Form an equation equivalent to eqn (7.19) and explain how estimates of the p_i could be obtained.

Exercise 7.26: Random samples of sizes n_1, n_2, and n_3 are taken from three Poisson distributions whose means are θ_1, θ_2, and θ_3. Specifically,

$$X_{ij} \sim \text{Poisson } (\theta_i) \quad \text{for } i = 1, 2, 3; \quad j = 1, \ldots, n_i.$$

A priori, the θ_i are exchangeable. Explain what is meant by *exchangeable* in this context and suggest an appropriate hierarchical model for the data.

Exercise 7.27: Derive eqn (7.21) from eqn (7.20). [Note that $\delta(\theta_i - \bar{\theta}/\delta\theta_i = (k-1)(\theta_i - \bar{\theta})/k$.]

Exercise 7.28: For Example 7.8:

(a) show that $\bar{X}_1, \ldots, \bar{X}_k$ are jointly sufficient for $(\theta_i, \ldots, \theta_k, \phi, \tau^2)$;

(b) put $f(\bar{x}_i; \phi, \tau^2) = \int f(\bar{x}_i; \theta_i, \phi, \tau^2) p(\theta_i; \phi, \tau^2)\, d\theta_i$ and show that

$$\bar{x}_i \,|\, \phi, \tau^2 \sim N(\phi, \tau^2 + \sigma^2/n_i);$$

(c) suppose the same number of children are tested in each school, so that $n_i = n$ for all i. Using (b) and the result found in Example 6.3, show that $\phi \,|\, \tau, \bar{x}_1, \ldots, \bar{x}_k \sim N(\sum \bar{x}_i/k, \{\tau + \sigma^2/n\}/k)$.

Exercise 7.29: Suppose $(x_i, \theta_i), i = 1, 2, \ldots, k$ is a sequence of i.i.d. bivariate r.vs for which only the x_i are observable. Suppose also that each x_i has a binomial distribution, $B(n, \theta_i)$, and $\theta_i \sim$ beta(α, β). Show that, using the method of moments, the EB estimates of α and β are

$$\alpha = \frac{n\bar{x}^2 - \bar{x}^3 - \bar{x}s^2}{ns^2 - n\bar{x} + \bar{x}^2}, \quad \beta = \frac{(n-\bar{x})(n\bar{x} - \bar{x}^2 - s^2)}{ns^2 - n\bar{x} + \bar{x}^2},$$

where $\bar{x}$ and s^2 are the sample mean and variance of $x_1, \ldots, x_k$.

Exercise 7.30: Suppose $X_i \,|\, \theta_i \sim N(\theta_i, 1)$, where $X_1, \ldots, X_k$ are independent and $\theta_i \sim N(\mu, \sigma^2)$, where $\theta_1, \ldots, \theta_k$ are independent. Determine EB estimates of μ and σ^2

(a) using the method of moments;

(b) using maximum likelihood.

Exercise 7.31: Suppose $f(x_i;\, \theta_i)$ is the geometric distribution $f(x_i;\, \theta_i) = \theta_i(1-\theta_i)^{x_i-1}, i = 1, 2, \ldots, k$, and θ_k must be estimated under a squared-error loss function. Show that $\{f(x_k) - f(x_k+1)\}/f(x_k)$ is a non-parametric EB estimator of θ_k. Given that the θ_i are independent and identically distributed, explain how this expression can be used to estimate θ_k from $x_1, x_2, \ldots, x_k$.

Exercise 7.32: Suppose $(X_i,\, \theta_i)$ is a sequence of binary r.vs in which the θ_i are identically distributed, $i = 1, 2, \ldots, k$. Suppose $X_i \mid \theta_i \sim B(n_i,\, \theta_i)$, where $n_1, \ldots, n_k$ are predetermined constants and let $f_n(x;\, \theta)$ denote the p.d.f. of the binomial distribution $B(n,\, \theta)$.

(a) Show that

$$\frac{(x+1)f_{n+1}(x+1)}{(n+1)f_n(x)} = E(\theta \mid x).$$

(b) Suggest how this result might be used to obtain a non-parametric EB estimator of θ_k under a squared-error loss function.

Exercise 7.33: Suppose an observation comes from the uniform distribution $U(0,\, \theta_0)$ with probability p and from $U(0,\, \theta_1)$ with probability $1-p$, where $\theta_0 < \theta_1$. Out of k independent observations, n were in the range $(0,\, \theta_0)$ while the remaining $k-n$ were in the range $(\theta_0,\, \theta_1)$. Show that the MLE of p is

$$\hat{p} = \begin{cases} \frac{n\theta_1 - k\theta_0}{k(\theta_1-\theta_0)}, & \text{if } n > k\theta_0/\theta_1 \\ 0, & \text{if } n \le k\theta_0/\theta_1. \end{cases}$$

Given that $n > k\theta_0/\theta_1$, predict the value of the next observation when error in prediction is penalized by a linear loss function.

Exercise 7.34: For $i = 1, 2, \ldots$, suppose θ_i can take one of two values, 0.2 or 0.5 (some θ_i equal 0.2, while others equal 0.5). Let $p = \Pr[\theta_i = 0.2]$ and $1-p = \Pr[\theta_i = 0.5]$. Suppose we have 20 independent observations $x_1, x_2, \ldots, x_{20}$, where $f(x_i;\, \theta_i) = \theta_i e^{-\theta_i x_i}$ for $x_i > 0$, and that $\Sigma x_i = 60$.

(a) Use the method of moments to estimate p.

(b) Calculate the probability that the next observation, X_{20} will exceed 4.

(c) Suppose one must choose between H_0: $\theta_{21} = 0.2$ and H_1: $\theta_{21} = 0.5$, and the loss for incorrectly rejecting H_0 is 10 and for incorrectly rejecting H_1 is 5. For what values of x_{21} should H_0 be accepted and for what values should it be rejected?

Exercise 7.35: Suppose 1.5, 1.7, 2.1, 1.8, and 1.3 are a random sample of five observations from $N(\theta,\, \sigma^2)$, where both θ and σ^2 are unknown.

(a) Stating any formulae you use, give the MLEs of θ and σ^2.

(b) Using these estimates, form a naïve 95% EB confidence interval for θ. Criticize this interval.

8 Non-parametric and robust inference

8.1 Introduction

In much of inference, for both the frequentist and Bayesian approaches, there is an assumption that we know the form of the p.d.f. $f(x;\theta)$, apart from the values of one or more parameters θ. In practice, this is often an unrealistic assumption. For example, a distribution may be approximately normal, but it is rarely exactly so. In more extreme cases, we have little confidence that we can correctly specify a parametric form for a p.d.f. This occurs particularly when only a small sample of observations is available, so that the shape of the distribution is not easily estimated.

It is therefore desirable to construct methods of inference which do not depend on distributional assumptions for their validity, or which are relatively insensitive to any distributional assumptions made. These two requirements lead us to the topics of non-parametric inference and robust inference, which are discussed in this chapter. Non-parametric methods require relatively weak distributional assumptions for their validity, while robust methods make inferences that are little affected by a small number of outliers in the data or slight departures from the distributional assumptions. We also consider the important problem of testing to see if a specified parametric model is appropriate. In addition, there is a brief description of semi-parametric methods which occupy an intermediate position between non-parametric inference and approaches described in earlier chapters. (Approaches that make full distributional assumptions are termed parametric methods.) Some types of computationally intensive methods are non-parametric in nature. Among these, permutation tests are introduced in this chapter, but discussion of other computationally intensive approaches is deferred until Chapter 9.

Suppose, as usual, that we have a random sample $(x_1, x_2, \ldots, x_n)$ from a probability distribution with p.d.f. $f(x;\theta)$, but that we do not know $f(x;\theta)$. In previous chapters the form of $f(x;\theta)$ was assumed known, except for the value of the parameter(s), θ, and our objective was to make inferences about θ. If we do not know the form of $f(x;\theta)$ then it is not immediately obvious what we mean by 'inference about θ'. However, there are certain parameters of