Hint: See Problem B.2.5.

4. Let $X_1, \ldots, X_n$ be the indicators of n Bernoulli trials with probability of success θ.

(a) Show that $\bar{X}$ is a method of moments estimate of θ.

(b) Exhibit method of moments estimates for $\text{Var}_\theta \bar{X} = \theta(1-\theta)/n$ first using only the first moment and then using only the second moment of the population. Show that these estimates coincide.

(c) Argue that in this case all frequency substitution estimates of $q(\theta)$ must agree with $q(\bar{X})$.

5. Let $X_1, \ldots, X_n$ be a sample from a population with distribution function F and frequency function or density p. The *empirical* distribution function $\widehat{F}$ is defined by $\widehat{F}(x) =$ [No. of $X_i \le x]/n$. If $q(\theta)$ can be written in the form $q(\theta) = s(F)$ for some function s of F we define the *empirical substitution principle estimate* of $q(\theta)$ to be $s(\widehat{F})$.

(a) Show that in the finite discrete case, empirical substitution estimates coincides with frequency substitution estimates.

Hint: Express F in terms of p and $\widehat{F}$ in terms of

$$\widehat{p}(x) = \frac{\text{No. of } X_i = x}{n}.$$

(b) Show that in the continuous case $X \sim \widehat{F}$ means that $X = X_i$ with probability $1/n$.

(c) Show that the empirical substitution estimate of the jth moment μ_j is the jth sample moment $\widehat{\mu}_j$.

Hint: Write $m_j = \int_{-\infty}^{\infty} x^j dF(x)$ or $m_j = E_F(X^j)$ where $X \sim F$.

(d) For $t_1 < \cdots < t_k$, find the joint frequency function of $\widehat{F}(t_1), \ldots, \widehat{F}(t_k)$.

Hint: Consider $(N_1, \ldots, N_{k+1})$ where $N_1 = n\widehat{F}(t_1)$, $N_2 = n(\widehat{F}(t_2) - \widehat{F}(t_1)), \ldots$, $N_{k+1} = n(1 - \widehat{F}(t_k))$.

6. Let $X_{(1)} \le \cdots \le X_{(n)}$ be the order statistics of a sample $X_1, \ldots, X_n$. (See Problem B.2.8.) There is a one-to-one correspondence between the empirical distribution function $\widehat{F}$ and the order statistics in the sense that, given the order statistics we may construct $\widehat{F}$ and given $\widehat{F}$, we know the order statistics. Give the details of this correspondence.

7. The jth cumulant $\widehat{c}_j$ of the empirical distribution function is called the jth *sample cumulant* and is a method of moments estimate of the cumulant c_j. Give the first three sample cumulants. See A.12.

8. Let $(Z_1, Y_1), (Z_2, Y_2), \ldots, (Z_n, Y_n)$ be a set of independent and identically distributed random vectors with common distribution function F. The natural estimate of $F(s,t)$ is the *bivariate empirical* distribution function $\widehat{F}(s,t)$, which we define by

$$\widehat{F}(s,t) = \frac{\text{Number of vectors } (Z_i, Y_i) \text{ such that } Z_i \le s \text{ and } Y_i \le t}{n}.$$

(a) Show that $\widehat{F}(\cdot,\cdot)$ is the distribution function of a probability $\widehat{P}$ on R^2 assigning mass $1/n$ to each point (Z_i, Y_i).

(b) Define the sample product moment of order (i,j), the sample covariance, the sample correlation, and so on, as the corresponding characteristics of the distribution $\widehat{F}$. Show that the sample product moment of order (i,j) is given by

$$\frac{1}{n}\sum_{k=1}^{n} Z_k^i Y_k^j.$$

The sample covariance is given by

$$\frac{1}{n}\sum_{k=1}^{n}(Z_k - \bar{Z})(Y_k - \bar{Y}) = \frac{1}{n}\sum_{k=1}^{n} Z_k Y_k - \bar{Z}\bar{Y},$$

where $\bar{Z}, \bar{Y}$ are the sample means of the $Z_1, \dots, Z_n$ and $Y_1, \dots, Y_n$, respectively. The sample correlation coefficient is given by

$$r = \frac{\sum_{k=1}^{n}(Z_k - \bar{Z})(Y_k - \bar{Y})}{\sqrt{\sum_{k=1}^{n}(Z_k - \bar{Z})^2 \sum_{k=1}^{n}(Y_k - \bar{Y})^2}}.$$

All of these quantities are natural estimates of the corresponding population characteristics and are also called method of moments estimates. (See Problem 2.1.17.) Note that it follows from (A.11.19) that $-1 \le r \le 1$.

9. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ where the X_i are independent $\mathcal{N}(0, \sigma^2)$.

(a) Find an estimate of σ^2 based on the second moment.

(b) Construct an estimate of σ using the estimate of part (a) and the equation $\sigma = \sqrt{\sigma^2}$.

(c) Use the empirical substitution principle to construct an estimate of σ using the relation $E(|X_1|) = \sigma\sqrt{2\pi}$.

10. In Example 2.1.1, suppose that $g(\boldsymbol{\beta}, \mathbf{z})$ is continuous in $\boldsymbol{\beta}$ and that $|g(\boldsymbol{\beta}, \mathbf{z})|$ tends to ∞ as $|\boldsymbol{\beta}|$ tends to ∞. Show that the least squares estimate exists.

Hint: Set $c = \rho(X, \mathbf{0})$. There exists a compact set K such that for $\boldsymbol{\beta}$ in the complement of K, $\rho(X, \boldsymbol{\beta}) > c$. Since $\rho(X, \boldsymbol{\beta})$ is continuous on K, the result follows.

11. In Example 2.1.2 with $X \sim \Gamma(\alpha, \lambda)$, find the method of moments estimate based on $\widehat{\mu}_1$ and $\widehat{\mu}_3$.

Hint: See Problem B.2.4.

12. Let $X_1, \dots, X_n$ be i.i.d. as $X \sim P_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta \subset R^d$, with $\boldsymbol{\theta}$ identifiable. Suppose X has possible values $v_1, \dots, v_k$ and that $q(\boldsymbol{\theta})$ can be written as

$$q(\boldsymbol{\theta}) = h(\mu_1(\boldsymbol{\theta}), \dots, \mu_r(\boldsymbol{\theta}))$$

for some R^k-valued function h. Show that the method of moments estimate $\widehat{q} = h(\widehat{\mu}_1, \dots, \widehat{\mu}_r)$ can be written as a frequency plug-in estimate.

13. *General method of moment estimates*[1]. Suppose $X_1, \dots, X_n$ are i.i.d. as $X \sim P_{\boldsymbol{\theta}}$, with $\boldsymbol{\theta} \in \Theta \subset R^d$ and $\boldsymbol{\theta}$ identifiable. Let $g_1, \dots, g_r$ be given linearly independent functions and write

$$\mu_j(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(g_j(X)),\ \widehat{\mu}_j = n^{-1}\sum_{i=1}^{n} g_j(X_i),\ j = 1, \dots, r.$$

Suppose that X has possible values $v_1, \dots, v_k$ and that

$$q(\boldsymbol{\theta}) = h(\mu_1(\boldsymbol{\theta}), \dots, \mu_r(\boldsymbol{\theta}))$$

for some R^k-valued function h.

(a) Show that the method of moments estimate $\widehat{q} = h(\widehat{\mu}_1, \dots, \widehat{\mu}_r)$ is a frequency plug-in estimate.

(b) Suppose $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is the k-parameter exponential family given by (1.6.10). Let $g_j(X) = T_j(X)$, $1 \le j \le k$. In the following cases, find the method of moments estimates

(i) Beta, $\beta(1, \theta)$

(ii) Beta, $\beta(\theta, 1)$

(iii) Raleigh, $p(x, \theta) = (x/\theta^2)\exp(-x^2/2\theta^2)$, $x > 0$, $\theta > 0$

(iv) Gamma, $\Gamma(p, \theta)$, p fixed

(v) Inverse Gaussian, $IG(\mu, \lambda)$, $\boldsymbol{\theta} = (\mu, \lambda)$. See Problem 1.6.36.

Hint: Use Corollary 1.6.1.

14. When the data are not i.i.d., it may still be possible to express parameters as functions of moments and then use estimates based on replacing population moments with "sample" moments. Consider the Gaussian $AR(1)$ model of Example 1.1.5.

(a) Use $E(X_i)$ to give a method of moments estimate of μ.

(b) Suppose $\mu = \mu_0$ and $\beta = b$ are fixed. Use $E(U_i^2)$, where

$$U_i = (X_i - \mu_0) \Bigg/ \left(\sum_{j=0}^{i-1} b^{2j}\right)^{1/2},$$

to give a method of moments estimate of σ^2.

(c) If μ and σ^2 are fixed, can you give a method of moments estimate of β?

15. *Hardy–Weinberg with six genotypes.* In a large natural population of plants (Mimulus guttatus) there are three possible alleles S, I, and F at one locus resulting in six genotypes labeled SS, II, FF, SI, SF, and IF. Let θ_1, θ_2, and θ_3 denote the probabilities of S, I, and F, respectively, where $\sum_{j=1}^{3} \theta_j = 1$. The Hardy–Weinberg model specifies that the six genotypes have probabilities

Genotype	1	2	3	4	5	6
Genotype	SS	II	FF	SI	SF	IF
Probability	θ_1^2	θ_2^2	θ_3^2	$2\theta_1\theta_2$	$2\theta_1\theta_3$	$2\theta_2\theta_3$

Let N_j be the number of plants of genotype j in a sample of n independent plants, $1 \leq j \leq 6$ and let $\widehat{p}_j = N_j/n$. Show that

$$\begin{aligned}\widehat{\theta}_1 &= \widehat{p}_1 + \tfrac{1}{2}\widehat{p}_4 + \tfrac{1}{2}\widehat{p}_5\\ \widehat{\theta}_2 &= \widehat{p}_2 + \tfrac{1}{2}\widehat{p}_4 + \tfrac{1}{2}\widehat{p}_6\\ \widehat{\theta}_3 &= \widehat{p}_3 + \tfrac{1}{2}\widehat{p}_5 + \tfrac{1}{2}\widehat{p}_6\end{aligned}$$

are frequency plug-in estimates of θ_1, θ_2, and θ_3.

16. Establish (2.1.6).
Hint: $[Y_i - g(\boldsymbol{\beta}, \boldsymbol{z}_i)] = [Y_i - g(\boldsymbol{\beta}_0, \boldsymbol{z}_i)] + [g(\boldsymbol{\beta}_0, \boldsymbol{z}_i) - g(\boldsymbol{\beta}, \boldsymbol{z}_i)]$.

17. *Multivariate method of moments.* For a vector $X = (X_1, \ldots, X_q)$, of observations, let the moments be

$$m_{jkrs} = E(X_r^j X_s^k),\ j \geq 0,\ k \geq 0;\ r, s = 1, \ldots, q.$$

For independent identically distributed $X_i = (X_{i1}, \ldots, X_{iq})$, $i = 1, \ldots, n$, we define the empirical or sample moment to be

$$\widehat{m}_{jkrs} = \frac{1}{n}\sum_{i=1}^{n} X_{ir}^j X_{is}^k,\ j \geq 0,\ k \geq 0;\ r, s = 1, \ldots, q.$$

If $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ can be expressed as a function of the moments, the method of moments estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by replacing m_{jkrs} by $\widehat{m}_{jkrs}$. Let $X = (Z, Y)$ and $\boldsymbol{\theta} = (a_1, b_1)$, where (Z, Y) and (a_1, b_1) are as in Theorem 1.4.3. Show that method of moments estimators of the parameters b_1 and a_1 in the best linear predictor are

$$\widehat{b}_1 = \frac{n^{-1}\sum Z_i Y_i - \bar{Z}\bar{Y}}{n^{-1}\sum Z_i - (\bar{Z})^2},\ \widehat{a}_1 = \bar{Y} - \widehat{b}_1 \bar{Z}.$$

Problems for Section 2.2

1. An object of unit mass is placed in a force field of unknown constant intensity θ. Readings $Y_1, \ldots, Y_n$ are taken at times $t_1, \ldots, t_n$ on the position of the object. The reading Y_i

differs from the true position $(\theta/2)t_i^2$ by a random error ϵ_i. We suppose the ϵ_i to have mean 0 and be uncorrelated with constant variance. Find the LSE of θ.

2. Show that the formulae of Example 2.2.2 may be derived from Theorem 1.4.3, if we consider the distribution assigning mass $1/n$ to each of the points $(z_1, y_1), \ldots, (z_n, y_n)$.

3. Suppose that observations $Y_1, \ldots, Y_n$ have been taken at times $z_1, \ldots, z_n$ and that the linear regression model holds. A new observation Y_{n+1} is to be taken at time z_{n+1}. What is the least squares estimate based on $Y_1, \ldots, Y_n$ of the best (MSPE) predictor of Y_{n+1}?

4. Show that the two sample regression lines coincide (when the axes are interchanged) if and only if the points (z_i, y_i), $i = 1, \ldots, n$, in fact, all lie on a line.

Hint: Write the lines in the form

$$\frac{(z - \bar{z})}{\widehat{\sigma}} = \widehat{\rho}\frac{(y - \bar{y})}{\widehat{\tau}}.$$

5. The regression line minimizes the sum of the squared vertical distances from the points $(z_1, y_1), \ldots, (z_n, y_n)$. Find the line that minimizes the sum of the squared *perpendicular* distance to the same points.

Hint: The quantity to be minimized is

$$\frac{\sum_{i=1}^n (y_i - \theta_1 - \theta_2 z_i)^2}{1 + \theta_2^2}.$$

6. **(a)** Let $Y_1, \ldots, Y_n$ be independent random variables with equal variances such that $E(Y_i) = \alpha z_j$ where the z_j are known constants. Find the least squares estimate of α.

(b) Relate your answer to the formula for the best zero intercept linear predictor of Section 1.4.

7. Show that the least squares estimate is always defined and satisfies the equations (2.1.5) provided that g is differentiable with respect to β_i, $1 \le i \le d$, the range $\{g(\mathbf{z}_1, \boldsymbol{\beta}), \ldots, g(\mathbf{z}_n, \boldsymbol{\beta}), \boldsymbol{\beta} \in R^d\}$ is closed, and $\boldsymbol{\beta}$ ranges over R^d.

8. Find the least squares estimates for the model $Y_i = \theta_1 + \theta_2 z_i + \epsilon_i$ with ϵ_i as given by (2.2.4)–(2.2.6) under the restrictions $\theta_1 \ge 0$, $\theta_2 \le 0$.

9. Suppose $Y_i = \theta_1 + \epsilon_i$, $i = 1, \ldots, n_1$ and $Y_i = \theta_2 + \epsilon_i$, $i = n_1 + 1, \ldots, n_1 + n_2$, where $\epsilon_1, \ldots, \epsilon_{n_1+n_2}$ are independent $\mathcal{N}(0, \sigma^2)$ variables. Find the least squares estimates of θ_1 and θ_2.

10. Let $X_1, \ldots, X_n$ denote a sample from a population with one of the following densities or frequency functions. Find the MLE of θ.

(a) $f(x, \theta) = \theta e^{-\theta x}$, $x \ge 0$; $\theta > 0$. (exponential density)

(b) $f(x, \theta) = \theta c^\theta x^{-(\theta+1)}$, $x \ge c$; c constant > 0; $\theta > 0$. (Pareto density)

(c) $f(x,\theta) = c\theta^c x^{-(c+1)}$, $x \geq \theta$; c constant > 0; $\theta > 0$. (Pareto density)

(d) $f(x,\theta) = \sqrt{\theta}x^{\sqrt{\theta}-1}$, $0 \leq x \leq 1$, $\theta > 0$. (beta, $\beta(\sqrt{\theta}, 1)$, density)

(e) $f(x,\theta) = (x/\theta^2)\exp\{-x^2/2\theta^2\}$, $x > 0$; $\theta > 0$. (Rayleigh density)

(f) $f(x,\theta) = \theta c x^{c-1}\exp\{-\theta x^c\}$, $x \geq 0$; c constant > 0; $\theta > 0$. (Weibull density)

11. Suppose that $X_1, \ldots, X_n$, $n \geq 2$, is a sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution.

(a) Show that if μ and σ^2 are unknown, $\mu \in R$, $\sigma^2 > 0$, then the unique MLEs are $\widehat{\mu} = \bar{X}$ and $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

(b) Suppose μ and σ^2 are both known to be nonnegative but otherwise unspecified. Find maximum likelihood estimates of μ and σ^2.

12. Let $X_1, \ldots, X_n$, $n \geq 2$, be independently and identically distributed with density

$$f(x,\theta) = \frac{1}{\sigma}\exp\{-(x-\mu)/\sigma\},\ x \geq \mu,$$

where $\theta = (\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$.

(a) Find maximum likelihood estimates of μ and σ^2.

(b) Find the maximum likelihood estimate of $P_\theta[X_1 \geq t]$ for $t > \mu$.
Hint: You may use Problem 2.2.16(b).

13. Let $X_1, \ldots, X_n$ be a sample from a $\mathcal{U}[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ distribution. Show that any T such that $X_{(n)} - \frac{1}{2} \leq T \leq X_{(1)} + \frac{1}{2}$ is a maximum likelihood estimate of θ. (We write $\mathcal{U}[a,b]$ to make $p(a) = p(b) = (b-a)^{-1}$ rather than 0.)

14. If $n = 1$ in Example 2.1.5 show that no maximum likelihood estimate of $\theta = (\mu, \sigma^2)$ exists.

15. Suppose that $T(\mathbf{X})$ is sufficient for θ and that $\widehat{\theta}(\mathbf{X})$ is an MLE of θ. Show that $\widehat{\theta}$ depends on $\mathbf{X}$ through $T(\mathbf{X})$ only provided that $\widehat{\theta}$ is unique.
Hint: Use the factorization theorem (Theorem 1.5.1).

16. (a) Let $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$ and let $\widehat{\theta}$ denote the MLE of θ. Suppose that h is a one-to-one function from Θ onto $h(\Theta)$. Define $\eta = h(\theta)$ and let $f(\mathbf{x}, \eta)$ denote the density or frequency function of $\mathbf{X}$ in terms of η (i.e., reparametrize the model using η). Show that the MLE of η is $h(\widehat{\theta})$ (i.e., MLEs are unaffected by reparametrization, they are *equivariant* under one-to-one transformations).

(b) Let $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$, $\Theta \subset R^p$, $p \geq 1$, be a family of models for $\mathbf{X} \in \mathcal{X} \subset R^d$. Let $\mathbf{q}$ be a map from Θ onto Ω, $\Omega \subset R^k$, $1 \leq k \leq p$. Show that if $\widehat{\boldsymbol{\theta}}$ is a MLE of $\boldsymbol{\theta}$, then $\mathbf{q}(\widehat{\boldsymbol{\theta}})$ is an MLE of $\boldsymbol{\omega} = \mathbf{q}(\boldsymbol{\theta})$.
Hint: Let $\Theta(\boldsymbol{\omega}) = \{\boldsymbol{\theta} \in \Theta : \mathbf{q}(\boldsymbol{\theta}) = \boldsymbol{\omega}\}$, then $\{\Theta(\boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$ is a partition of Θ, and $\widehat{\boldsymbol{\theta}}$ belongs to only one member of this partition, say $\Theta(\widehat{\boldsymbol{\omega}})$. Because $\mathbf{q}$ is onto Ω, for each $\boldsymbol{\omega} \in \Omega$ there is $\boldsymbol{\theta} \in \Theta$ such that $\boldsymbol{\omega} = \mathbf{q}(\boldsymbol{\theta})$. Thus, the MLE of $\boldsymbol{\omega}$ is by definition

$$\widehat{\boldsymbol{\omega}}_{MLE} = \arg\sup_{\boldsymbol{\omega} \in \Omega}\sup\{L_{\mathbf{X}}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta(\boldsymbol{\omega})\}.$$

Now show that $\widehat{\omega}_{MLE} = \widehat{\omega} = \mathbf{q}(\widehat{\boldsymbol{\theta}})$.

17. *Censored Geometric Waiting Times.* If time is measured in discrete periods, a model that is often used for the time X to failure of an item is

$$P_\theta[X = k] = \theta^{k-1}(1-\theta),\ k = 1, 2, \ldots$$

where $0 < \theta < 1$. Suppose that we only record the time of failure, if failure occurs on or before time r and otherwise just note that the item has lived at least $(r+1)$ periods. Thus, we observe $Y_1, \ldots, Y_n$ which are independent, identically distributed, and have common frequency function,

$$f(k, \theta) = \theta^{k-1}(1-\theta),\ k = 1, \ldots, r$$
$$f(r+1, \theta) = 1 - P_\theta[X \le r] = 1 - \sum_{k=1}^{r} \theta^{k-1}(1-\theta) = \theta^r.$$

(We denote by "$r+1$" survival for at least $(r+1)$ periods.) Let M = number of indices i such that $Y_i = r+1$. Show that the maximum likelihood estimate of θ based on $Y_1, \ldots, Y_n$ is

$$\widehat{\theta}(\mathbf{Y}) = \frac{\sum_{i=1}^n Y_i - n}{\sum_{i=1}^n Y_i - M}.$$

18. Derive maximum likelihood estimates in the following models.

(a) The observations are indicators of Bernoulli trials with probability of success θ. We want to estimate θ and $\text{Var}_\theta X_1 = \theta(1-\theta)$.

(b) The observations are X_1 = the number of failures before the first success, X_2 = the number of failures between the first and second successes, and so on, in a sequence of binomial trials with probability of success θ. We want to estimate θ.

19. Let $X_1, \ldots, X_n$ be independently distributed with X_i having a $\mathcal{N}(\theta_i, 1)$ distribution, $1 \le i \le n$.

(a) Find maximum likelihood estimates of the θ_i under the assumption that these quantities vary freely.

(b) Solve the problem of part (a) for $n = 2$ when it is known that $\theta_1 \le \theta_2$. A general solution of this and related problems may be found in the book by Barlow, Bartholomew, Bremner, and Brunk (1972).

20. In the "life testing" problem 1.6.16(i), find the MLE of θ.

21. (Kiefer–Wolfowitz) Suppose $(X_1, \ldots, X_n)$ is a sample from a population with density

$$f(x, \theta) = \frac{9}{10\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{10}\varphi(x-\mu)$$

where φ is the standard normal density and $\theta = (\mu, \sigma^2) \in \Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$. Show that maximum likelihood estimates do not exist, but that $\sup_\sigma p(\mathbf{x}, \widehat{\mu}, \sigma^2) = \sup_{\mu,\sigma} p(\mathbf{x}, \mu, \sigma^2)$ if, and only if, $\widehat{\mu}$ equals one of the numbers $x_1, \dots, x_n$. Assume that $x_i \neq x_j$ for $i \neq j$ and that $n \geq 2$.

22. Suppose X has a hypergeometric, $\mathcal{H}(b, N, n)$, distribution. Show that the maximum likelihood estimate of b for N and n fixed is given by

$$\widehat{b}(X) = \left[\frac{X}{n}(N+1)\right]$$

if $\frac{X}{n}(N+1)$ is *not* an integer, and

$$\widehat{b}(X) = \frac{X}{n}(N+1) \text{ or } \frac{X}{n}(N+1) - 1$$

otherwise, where $[t]$ is the largest integer that is $\leq t$.

Hint: Consider the ratio $L(b+1, x)/L(b, x)$ as a function of b.

23. Let $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ be two independent samples from $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ populations, respectively. Show that the MLE of $\theta = (\mu_1, \mu_2, \sigma^2)$ is $\widehat{\theta} = (\bar{X}, \bar{Y}, \widetilde{\sigma}^2)$ where

$$\widetilde{\sigma}^2 = \left[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2\right] / (m+n).$$

24. *Polynomial Regression.* Suppose $Y_i = \mu(\mathbf{z}_i) + \epsilon_i$, where ϵ_i satisfy (2.2.4)–(2.2.6). Set $\mathbf{z}^{\mathbf{j}} = z_1^{j_1} \cdots z_p^{j_p}$ where $\mathbf{j} \in \mathcal{J}$ and $\mathcal{J}$ is a subset of $\{(j_1, \dots, j_p) : 0 \leq j_k \leq J,\ 1 \leq k \leq p\}$, and assume that

$$\mu(\mathbf{z}) = \sum \{\alpha_j \mathbf{z}^{\mathbf{j}} : \mathbf{j} \in \mathcal{J}\}.$$

In an experiment to study tool life (in minutes) of steel-cutting tools as a function of cutting speed (in feet per minute) and feed rate (in thousands of an inch per revolution), the following data were obtained (from S. Weisberg, 1985).

TABLE 2.6.1. Tool life data

Feed	Speed	Life	Feed	Speed	Life
−1	−1	54.5	$-\sqrt{2}$	0	20.1
−1	−1	66.0	$\sqrt{2}$	0	2.9
1	−1	11.8	0	0	3.8
1	−1	14.0	0	0	2.2
−1	1	5.2	0	0	3.2
−1	1	3.0	0	0	4.0
1	1	0.8	0	0	2.8
1	1	0.5	0	0	3.2
0	$-\sqrt{2}$	86.5	0	0	4.0
0	$\sqrt{2}$	0.4	0	0	3.5

The researchers analyzed these data using

$$Y = \log \text{tool life},\ z_1 = (\text{feed rate} - 13)/6,\ z_2 = (\text{cutting speed} - 900)/300.$$

Two models are contemplated

(a) $Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \epsilon$

(b) $Y = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_1^2 + \alpha_4 z_2^2 + \alpha_5 z_1 z_2 + \epsilon$.

Use a least squares computer package to compute estimates of the coefficients (β's and α's) in the two models. Use these estimated coefficients to compute the values of the contrast function (2.1.5) for (a) and (b). Both of these models are approximations to the true mechanism generating the data. Being larger, the second model provides a better approximation. However, this has to be balanced against greater variability in the estimated coefficients. This will be discussed in Volume II.

25. Consider the model (2.2.1), (2.2.4)–(2.2.6) with $g(\boldsymbol{\beta}, \mathbf{z}) = \mathbf{z}^T\boldsymbol{\beta}$. Show that the following are equivalent.

(a) The parameterization $\boldsymbol{\beta} \to \mathbf{Z}_D\boldsymbol{\beta}$ is identifiable.

(b) $\mathbf{Z}_D$ is of rank d.

(c) $\mathbf{Z}_D^T\mathbf{Z}_D$ is of rank d.

26. Let (Z, Y) have joint probability P with joint density $f(z, y)$, let $v(z, y) \geq 0$ be a weight funciton such that $E(v(Z, Y)Z^2)$ and $E(v(Z, Y)Y^2)$ are finite. The best linear weighted mean squared prediction error predictor $\beta_1(P) + \beta_2(P)Z$ of Y is defined as the minimizer of

$$E\{v(Z, Y)[Y - (b_1 + b_2 Z)]^2\}.$$

(a) Let (Z^*, Y^*) have density $v(z, y)f(z, y)/c$ where $c = \int\int v(z, y)f(z, y)dzdy$. Show that $\beta_2(P) = \text{Cov}(Z^*, Y^*)/\text{Var } Z^*$ and $\beta_1(P) = E(Y^*) - \beta_2(P)E(Z^*)$.

(b) Let $\widehat{P}$ be the empirical probability defined in Problem 2.1.8 and let $v(z, y) = 1/\text{Var}(Y \mid Z = z)$. Show that $\beta_1(\widehat{P})$ and $\beta_2(\widehat{P})$ coincide with $\widehat{\beta}_1$ and $\widehat{\beta}_2$ of Example 2.2.3. That is, weighted least squares estimates are plug-in estimates.

27. Derive the weighted least squares normal equations (2.2.19).

28. Let $\boldsymbol{Z}_D = \|z_{ij}\|_{n \times d}$ be a design matrix and let $\boldsymbol{W}_{n \times n}$ be a known symmetric invertible matrix. Consider the model $\boldsymbol{Y} = \boldsymbol{Z}_D\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ has covariance matrix $\sigma^2\boldsymbol{W}$, σ^2 unknown. Let $\boldsymbol{W}^{-\frac{1}{2}}$ be a square root matrix of $\boldsymbol{W}^{-1}$ (see (B.6.6)). Set $\widetilde{\boldsymbol{Y}} = \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{Y}$, $\widetilde{\boldsymbol{Z}}_D = \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{Z}_D$ and $\widetilde{\boldsymbol{\epsilon}} = \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{\epsilon}$.

(a) Show that $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{Z}}_D\boldsymbol{\beta} + \widetilde{\boldsymbol{\epsilon}}$ satisfy the linear regression model (2.2.1), (2.2.4)–(2.2.6) with $g(\boldsymbol{\beta}, \boldsymbol{z}) = \widetilde{\boldsymbol{Z}}_D\boldsymbol{\beta}$.

(b) Show that if $\boldsymbol{Z}_D$ has rank d, then the $\widehat{\boldsymbol{\beta}}$ that minimizes

$$(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{Z}}_D\boldsymbol{\beta})^T(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{Z}}_D\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{Z}_D\boldsymbol{\beta})^T\boldsymbol{W}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}_D\boldsymbol{\beta})$$

is given by (2.2.20).

29. Let $e_i = (\epsilon_i + \epsilon_{i+1})/2$, $i = 1, \dots, n$, where $\epsilon_1, \dots, \epsilon_{n+1}$ are i.i.d. with mean zero and variance σ^2. The e_i are called *moving average errors.*

Consider the model $Y_i = \mu + e_i$, $i = 1, \dots, n$.

(a) Show that $E(Y_{i+1} \mid Y_1, \dots, Y_i) = \frac{1}{2}(\mu + Y_i)$. That is, in this model the optimal MSPE predictor of the future Y_{i+1} given the past $Y_1, \dots, Y_i$ is $\frac{1}{2}(\mu + Y_i)$.

(b) Show that $\bar{Y}$ is a multivariate method of moments estimate of μ. (See Problem 2.1.17.)

(c) Find a matrix $\boldsymbol{A}$ such that $\boldsymbol{e}_{n\times 1} = \boldsymbol{A}_{n\times(n+1)}\boldsymbol{\epsilon}_{(n+1)\times 1}$.

(d) Find the covariance matrix $\boldsymbol{W}$ of $\boldsymbol{e}$.

(e) Find the weighted least squares estimate of μ.

(f) The following data give the elapsed times $Y_1, \dots, Y_n$ spent above a fixed high level for a series of $n = 66$ consecutive wave records at a point on the seashore. Use a weighted least squares computer routine to compute the weighted least squares estimate $\widehat{\mu}$ of μ. Is $\widehat{\mu}$ different from $\bar{Y}$?

TABLE 2.5.1. Elapsed times spent above a certain high level for a series of 66 wave records taken at San Francisco Bay. The data (courtesy S. J. Chou) should be read row by row.

2.968	2.097	1.611	3.038	7.921	5.476	9.858	1.397	0.155	1.301
9.054	1.958	4.058	3.918	2.019	3.689	3.081	4.229	4.669	2.274
1.971	10.379	3.391	2.093	6.053	4.196	2.788	4.511	7.300	5.856
0.860	2.093	0.703	1.182	4.114	2.075	2.834	3.968	6.480	2.360
5.249	5.100	4.131	0.020	1.071	4.455	3.676	2.666	5.457	1.046
1.908	3.064	5.392	8.393	0.916	9.665	5.564	3.599	2.723	2.870
1.582	5.453	4.091	3.716	6.156	2.039				

30. In the multinomial Example 2.2.8, suppose some of the n_j are zero. Show that the MLE of θ_j is $\widehat{\boldsymbol{\theta}}$ with $\widehat{\theta}_j = n_j/n$, $j = 1, \dots, k$.

Hint: Suppose without loss of generality that $n_1 = n_2 = \cdots = n_q = 0, n_{q+1} > 0, \dots, n_k > 0$. Then

$$p(\mathbf{x}, \boldsymbol{\theta}) = \prod_{j=q+1}^{k} \theta_j^{n_j},$$

which vanishes if $\theta_j = 0$ for any $j = q+1, \dots, k$.

31. Suppose $Y_1, \dots, Y_n$ are independent with Y_i uniformly distributed on $[\mu_i - \sigma, \mu_i + \sigma]$, $\sigma > 0$, where $\mu_i = \sum_{j=1}^{p} z_{ij}\beta_j$ for given covariate values $\{z_{ij}\}$. Show that the MLE of

$(\beta_1, \dots, \beta_p, \sigma)^T$ is obtained by finding $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ that minimizes the *maximum absolute value* contrast function $\max_i |y_i - \mu_i|$ and then setting $\widehat{\sigma} = \max_i |y_i - \widehat{\mu}_i|$, where $\widehat{\mu}_i = \sum_{j=1}^{p} z_{ij}\widehat{\beta}_j$.

32. Suppose $Y_1, \dots, Y_n$ are independent with Y_i having the Laplace density

$$\frac{1}{2\sigma}\exp\{-|y_i - \mu_i|/\sigma\},\ \sigma > 0$$

where $\mu_i = \sum_{j=1}^{p} z_{ij}\beta_j$ for given covariate values $\{z_{ij}\}$.

(a) Show that the MLE of $(\beta_1, \dots, \beta_p, \sigma)$ is obtained by finding $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ that minimizes the *least absolute deviation* contrast function $\sum_{i=1}^{n} |y_i - \mu_i|$ and then setting $\widehat{\sigma} = n^{-1}\sum_{i=1}^{n} |y_i - \widehat{\mu}_i|$, where $\widehat{\mu}_i = \sum_{j=1}^{p} z_{ij}\widehat{\beta}_j$. These $\widehat{\beta}_1, \dots, \widehat{\beta}_r$ and $\widehat{\mu}_1, \dots, \widehat{\mu}_n$ are called *least absolute deviation estimates (LADEs)*.

(b) If n is odd, the sample median $\widehat{y}$ is defined as $y_{(k)}$ where $k = \frac{1}{2}(n+1)$ and $y_{(1)}, \dots, y_{(n)}$ denotes $y_1, \dots, y_n$ ordered from smallest to largest. If n is even, the sample median $\widehat{y}$ is defined as $\frac{1}{2}[y_{(r)} + y_{(r+1)}]$ where $r = \frac{1}{2}n$. (See (2.1.17).) Suppose $\mu_i = \mu$ for each i. Show that the sample median $\widehat{y}$ is the minimizer of $\sum_{i=1}^{n} |y_i - \mu|$.

Hint: Use Problem 1.4.7 with Y having the empirical distribution $\widehat{F}$.

33. The *Hodges–Lehmann (location)* estimate $\widehat{x}_{HL}$ is defined to be the median of the $\frac{1}{2}n(n+1)$ pairwise averages $\frac{1}{2}(x_i + x_j)$, $i \le j$. An asymptotically equivalent procedure $\widetilde{x}_{HL}$ is to take the median of the distribution placing mass $\frac{2}{n^2}$ at each point $\frac{x_i + x_j}{2}$, $i < j$ and mass $\frac{1}{n^2}$ at each x_i.

(a) Show that the Hodges–Lehmann estimate is the minimizer of the contrast function

$$\rho(x, \theta) = \sum_{i \le j} |x_i + x_j - 2\theta|.$$

Hint: See Problem 2.2.32(b).

(b) Define θ_{HL} to be the minimizer of

$$\int |x - 2\theta| d(F * F)(x)$$

where $F * F$ denotes convolution. Show that $\widetilde{x}_{HL}$ is a plug-in estimate of θ_{HL}.

34. Let X_i be i.i.d. as $(Z, Y)^T$ where $Y = Z + \sqrt{\lambda}W$, $\lambda > 0$, Z and W are independent $\mathcal{N}(0, 1)$. Find the MLE of λ and give its mean and variance.

Hint: See Example 1.6.3.

35. Let $g(x) = 1/\pi(1 + x^2)$, $x \in R$, be the Cauchy density, let X_1 and X_2 be i.i.d. with density $g(x - \theta)$, $\theta \in R$. Let x_1 and x_2 be the observations and set $\Delta = \frac{1}{2}(x_1 - x_2)$. Let $\widehat{\theta} = \arg\max L_{\mathbf{x}}(\theta)$ be "the" MLE.

(a) Show that if $|\Delta| \le 1$, then the MLE exists and is unique. Give the MLE when $|\Delta| \le 1$.

(b) Show that if $|\Delta| > 1$, then the MLE is not unique. Find the values of θ that maximize the likelihood $L_{\mathbf{x}}(\theta)$ when $|\Delta| > 1$.

Hint: Factor out $(\bar{x} - \theta)$ in the likelihood equation.

36. Problem 35 can be generalized as follows (Dharmadhikari and Joag–Dev, 1985). Let g be a probability density on R satisfying the following three conditions:

1. g is continuous, symmetric about 0, and positive everywhere.
2. g is twice continuously differentiable everywhere except perhaps at 0.
3. If we write $h = \log g$, then $h''(y) > 0$ for some nonzero y.

Let (X_1, X_2) be a random sample from the distribution with density $f(x, \theta) = g(x - \theta)$, where $x \in R$ and $\theta \in R$. Let x_1 and x_2 be the observed values of X_1 and X_2 and write $\bar{x} = (x_1 + x_2)/2$ and $\Delta = (x_1 - x_2)/2$. The likelihood function is given by

$$\begin{aligned} L_{\mathbf{x}}(\theta) &= g(x_1 - \theta) g(x_2 - \theta) \\ &= g(\bar{x} + \Delta - \theta) g(\bar{x} - \Delta - \theta). \end{aligned}$$

Let $\widehat{\theta} = \arg\max L_{\mathbf{x}}(\theta)$ be "the" MLE.

Show that

(a) The likelihood is symmetric about $\bar{x}$.

(b) Either $\widehat{\theta} = \bar{x}$ or $\widehat{\theta}$ is not unique.

(c) There is an interval (a, b), $a < b$, such that for every $y \in (a, b)$ there exists a $\delta > 0$ such that $h(y + \delta) - h(y) > h(y) - h(y - \delta)$.

(d) Use (c) to show that if $\Delta \in (a, b)$, then $\widehat{\theta}$ is not unique.

37. Suppose $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\theta, \sigma^2)$ and let $p(\mathbf{x}, \theta)$ denote their joint density. Show that the entropy of $p(\mathbf{x}, \theta)$ is $\frac{1}{2}n$ and that the Kullback–Liebler divergence between $p(\mathbf{x}, \theta)$ and $p(\mathbf{x}, \theta_0)$ is $\frac{1}{2}n(\theta - \theta_0)^2/\sigma^2$.

38. Let $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$. Suppose h is a 1-1 function from Θ onto $\Omega = h(\Theta)$. Define $\eta = h(\theta)$ and let $p^*(\mathbf{x}, \eta) = p(\mathbf{x}, h^{-1}(\eta))$ denote the density or frequency function of $\mathbf{X}$ for the η parametrization. Let $K(\theta_0, \theta_1)$ $(K^*(\eta_0, \eta_1))$ denote the Kullback–Leibler divergence between $p(\mathbf{x}, \theta_0)$ and $p(x, \theta_1)$ $(p^*(\mathbf{x}, \eta_0)$ and $p^*(\mathbf{x}, \eta_1))$. Show that

$$K^*(\eta_0, \eta_1) = K(h^{-1}(\eta_0), h^{-1}(\eta_1)).$$

39. Let X_i denote the number of hits at a certain Web site on day i, $i = 1, \ldots, n$. Assume that $S = \sum_{i=1}^n X_i$ has a Poisson, $\mathcal{P}(n\lambda)$, distribution. On day $n + 1$ the Web Master decides to keep track of two types of hits (money making and not money making). Let V_j and W_j denote the number of hits of type 1 and 2 on day j, $j = n+1, \ldots, n+m$. Assume that $S_1 = \sum_{j=n+1}^{n+m} V_j$ and $S_2 = \sum_{j=n+1}^{n+m} W_j$ have $\mathcal{P}(m\lambda_1)$ and $\mathcal{P}(m\lambda_2)$ distributions, where $\lambda_1 + \lambda_2 = \lambda$. Also assume that S, S_1, and S_2 are independent. Find the MLEs of λ_1 and λ_2 based on S, S_1, and S_2.

40. Let $X_1, \dots, X_n$ be a sample from the generalized Laplace distribution with density

$$\begin{aligned} f(x, \theta_1, \theta_2) &= \frac{1}{\theta_1 + \theta_2} \exp\{-x/\theta_1\},\ x > 0, \\ &= \frac{1}{\theta_1 + \theta_2} \exp\{x/\theta_2\},\ x < 0 \end{aligned}$$

where $\theta_j > 0$, $j = 1, 2$.

(a) Show that $T_1 = \sum X_i 1[X_i > 0]$ and $T_2 = \sum -X_i 1[X_i < 0]$ are sufficient statistics.

(b) Find the maximum likelihood estimates of θ_1 and θ_2 in terms of T_1 and T_2. Carefully check the "$T_1 = 0$ or $T_2 = 0$" case.

41. The mean relative growth of an organism of size y at time t is sometimes modeled by the equation (Richards, 1959; Seber and Wild, 1989)

$$\frac{1}{y}\frac{dy}{dt} = \beta\left[1 - \left(\frac{y}{\alpha}\right)^{\frac{1}{\delta}}\right],\ y > 0;\ \alpha > 0,\ \beta > 0,\ \delta > 0.$$

(a) Show that a solution to this equation is of the form $y = g(t; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \beta, \mu, \delta)$, $\mu \in R$, and

$$g(t, \boldsymbol{\theta}) = \frac{\alpha}{\{1 + \exp[-\beta(t - \mu)/\delta]\}^{\delta}}.$$

(b) Suppose we have observations $(t_1, y_1), \dots, (t_n, y_n)$, $n \geq 4$, on a population of a large number of organisms. Variation in the population is modeled on the log scale by using the model

$$\log Y_i = \log \alpha - \delta \log\{1 + \exp[-\beta(t_i - \mu)/\delta]\} + \epsilon_i$$

where $\epsilon_1, \dots, \epsilon_n$ are uncorrelated with mean 0 and variance σ^2. Give the least squares estimating equations (2.1.7) for estimating α, β, δ, and μ.

(c) Let Y_i denote the response of the ith organism in a sample and let z_{ij} denote the level of the jth covariate (stimulus) for the ith organism, $i = 1, \dots, n$; $j = 1, \dots, p$. An example of a *neural net model* is

$$Y_i = \sum_{j=1}^{p} h(z_{ij}; \boldsymbol{\lambda}_j) + \epsilon_i,\ i = 1, \dots, n$$

where $\boldsymbol{\lambda} = (\alpha, \beta, \mu)$, $h(z; \lambda) = g(z; \alpha, \beta, \mu, 1)$; and $\epsilon_1, \dots, \epsilon_n$ are uncorrelated with mean zero and variance σ^2. For the case $p = 1$, give the least square estimating equations (2.1.7) for α, β, and μ.

42. Suppose $X_1, \dots, X_n$ satisfy the autoregressive model of Example 1.1.5.

(a) If μ is known, show that the MLE of β is

$$\widehat{\beta} = \frac{-\sum_{i=2}^{n}(x_{i-1}-\mu)(x_i-\mu)}{\sum_{i=1}^{n-1}(x_i-\mu)^2}.$$

(b) If β is known, find the covariance matrix $\mathbf{W}$ of the vector $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ of autoregression errors. (One way to do this is to find a matrix $\mathbf{A}$ such that $\mathbf{e}_{n\times 1} = \mathbf{A}_{n\times n}\epsilon_{n\times 1}$.) Then find the weighted least square estimate of μ. Is this also the MLE of μ?

Problems for Section 2.3

1. Suppose $Y_1, \ldots, Y_n$ are independent

$$P[Y_i = 1] = p(x_i, \alpha, \beta) = 1 - P[Y_i = 0],\ 1 \le i \le n,\ n \ge 2,$$

$$\log \frac{p}{1-p}(x, \alpha, \beta) = \alpha + \beta x,\ x_1 < \cdots < x_n.$$

Show that the MLE of α, β exists iff $(Y_1, \ldots, Y_n)$ is not a sequence of 1's followed by all 0's or the reverse.

Hint:

$$c_1 \sum_{i=1}^{n} y_i + c_2 \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n}(c_1 + c_2 x_i) y_i \le \sum_{i=1}^{n}(c_1 + c_2 x_i) 1(c_2 x_i + c_1 \ge 0).$$

If $c_2 > 0$, the bound is sharp and is attained only if $y_i = 0$ for $x_i \le -\frac{c_1}{c_2}$, $y_i = 1$ for $x_i \ge -\frac{c_1}{c_2}$.

2. Let $X_1, \ldots, X_n$ be i.i.d. gamma, $\Gamma(\lambda, p)$.

(a) Show that the density of $\mathbf{X} = (X_1, \ldots, X_n)^T$ can be written as the rank 2 canonical exponential family generated by $\mathbf{T} = (\Sigma \log X_i, \Sigma X_i)$ and $h(x) = x^{-1}$ with $\eta_1 = p$, $\eta_2 = -\lambda$ and

$$A(\eta_1, \eta_2) = n[\log \Gamma(\eta_1) - \eta_1 \log(-\eta_2)],$$

where Γ denotes the gamma function.

(b) Show that the likelihood equations are equivalent to (2.3.4) and (2.3.5).

3. Consider the Hardy–Weinberg model with the six genotypes given in Problem 2.1.15. Let $\Theta = \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0, \theta_1 + \theta_2 < 1\}$ and let $\theta_3 = 1 - (\theta_1 + \theta_2)$. In a sample of n independent plants, write $x_i = j$ if the ith plant has genotype j, $1 \le j \le 6$. Under what conditions on $(x_1, \ldots, x_n)$ does the MLE exist? What is the MLE? Is it unique?

4. Give details of the proof or Corollary 2.3.1.

5. Prove Lemma 2.3.1.

Hint: Let $c = l(\mathbf{0})$. There exists a compact set $K \subset \Theta$ such that $l(\boldsymbol{\theta}) < c$ for all $\boldsymbol{\theta}$ not in K. This set K will have a point where the max is attained.

6. In the heterogenous regression Example 1.6.10 with $n \geq 3$, $0 < z_1 < \cdots < z_n$, show that the MLE exists and is unique.

7. Let $Y_1, \ldots, Y_n$ denote the duration times of n independent visits to a Web site. Suppose $\mathbf{Y}$ has an exponential, $\mathcal{E}(\lambda_i)$, distribution where

$$\mu_i = E(Y_i) = \lambda_i^{-1} = \exp\{\alpha + \beta z_i\},\ z_1 < \cdots < z_n$$

and z_i is the income of the person whose duration time is Y_i, $0 < z_1 < \cdots < z_n$, $n \geq 2$. Show that the MLE of $(\alpha, \beta)^T$ exists and is unique. See also Problem 1.6.40.

8. Let $X_1, \ldots, X_n \in R^p$ be i.i.d. with density,

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = c(\alpha) \exp\{-|\mathbf{x} - \boldsymbol{\theta}|^\alpha\},\ \boldsymbol{\theta} \in R^p,\ \alpha \geq 1$$

where $c^{-1}(\alpha) = \int_{R_p} \exp\{-|\mathbf{x}|^\alpha\} d\mathbf{x}$ and $|\cdot|$ is the Euclidean norm.

(a) Show that if $\alpha > 1$, the MLE $\widehat{\boldsymbol{\theta}}$ exists and is unique.

(b) Show that if $\alpha = 1$, the MLE $\widehat{\boldsymbol{\theta}}$ exists but is not unique if n is even.

9. Show that the boundary ∂C of a convex C set in R^k has volume 0.

Hint: If ∂C has positive volume, then it must contain a sphere and the center of the sphere is an interior point by (B.9.1).

10. Use Corollary 2.3.1 to show that in the multinomial Example 2.3.3, MLEs of η_j exist iff all $T_j > 0$, $1 \leq j \leq k-1$.

Hint: The k points $(0, \ldots, 0)$, $(0, n, 0, \ldots, 0), \ldots, (0, 0, \ldots, n)$ are the vertices of the convex set $\{(t_1, \ldots, t_{k-1}) : t_j \geq 0, 1 \leq j \leq k-1, \sum_{j=1}^{k-1} t_j \leq n\}$.

11. Prove Theorem 2.3.3.

Hint: If it didn't there would exist $\boldsymbol{\eta}_j = c(\theta_j)$ such that $\boldsymbol{\eta}_j^T t_0 - A(\boldsymbol{\eta}_j) \to \max\{\boldsymbol{\eta}^T \mathbf{t}_0 - A(\boldsymbol{\eta}) : \boldsymbol{\eta} \in c(\Theta)\} > -\infty$. Then $\{\boldsymbol{\eta}_j\}$ has a subsequence that converges to a point $\boldsymbol{\eta}^0 \in \mathcal{E}$. But $c(\Theta)$ is closed so that $\boldsymbol{\eta}^0 = c(\boldsymbol{\theta}^0)$ and $\boldsymbol{\theta}^0$ must satisfy the likelihood equations.

12. Let $X_1, \ldots, X_n$ be i.i.d. $\frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$, $\sigma > 0$, $\mu \in R$, and assume for $w \equiv -\log f_0$ that $w'' > 0$ so that w is strictly convex, $w(\pm\infty) = \infty$.

(a) Show that, if $n \geq 2$, the likelihood equations

$$\sum_{i=1}^n w'\left(\frac{X_i - \mu}{\sigma}\right) = 0$$

$$\sum_{i=1}^n \left\{\frac{(X_i - \mu)}{\sigma} w'\left(\frac{X_i - \mu}{\sigma}\right) - 1\right\} = 0$$

have a unique solution $(\widehat{\mu}, \widehat{\sigma})$.

(b) Give an algorithm such that starting at $\widehat{\mu}^0 = 0$, $\widehat{\sigma}^0 = 1$, $\widehat{\mu}^{(i)} \to \widehat{\mu}$, $\widehat{\sigma}^{(i)} \to \widehat{\sigma}$.