(b) $p(x,\theta) = \{\exp[-2\log\theta + \log(2x)]\}1[x \in (0,\theta)]$

(c) $p(x,\theta) = \frac{1}{9}$, $x \in \{0.1+\theta,\dots,0.9+\theta\}$

(d) The $\mathcal{N}(\theta,\theta^2)$ family, $\theta > 0$

(e) $p(x,\theta) = 2(x+\theta)/(1+2\theta)$, $0 < x < 1$, $\theta > 0$

(f) $p(x,\theta)$ is the conditional frequency function of a binomial, $\mathcal{B}(n,\theta)$, variable X, given that $X > 0$.

5. Show that the following families of distributions are two-parameter exponential families and identify the functions η, B, T, and h.

(a) The beta family.

(b) The gamma family.

6. Let X have the Dirichlet distribution, $\mathcal{D}(\alpha)$, of Problem 1.2.15.

Show the distribution of X form an r-parameter exponential family and identify η, B, T, and h.

7. Let $\mathbf{X} = ((X_1,Y_1),\dots,(X_n,Y_n))$ be a sample from a bivariate normal population. Show that the distributions of $\mathbf{X}$ form a five-parameter exponential family and identify η, B, T, and h.

8. Show that the family of distributions of Example 1.5.3 is not a one parameter exponential family.

Hint: If it were, there would be a set A such that $p(x,\theta) > 0$ on A for all θ.

9. Prove the analogue of Theorem 1.6.1 for discrete k-parameter exponential families.

10. Suppose that $f(x,\theta)$ is a positive density on the real line, which is continuous in x for each θ and such that if (X_1,X_2) is a sample of size 2 from $f(\cdot,\theta)$, then X_1+X_2 is sufficient for θ. Show that $f(\cdot,\theta)$ corresponds to a one-arameter exponential family of distributions with $T(x) = x$.

Hint: There exist functions $g(t,\theta)$, $h(x_1,x_2)$ such that $\log f(x_1,\theta) + \log f(x_2,\theta) = g(x_1+x_2,\theta) + h(x_1,x_2)$. Fix θ_0 and let $r(x,\theta) = \log f(x,\theta) - \log f(x,\theta_0)$, $q(x,\theta) = g(x,\theta) - g(x,\theta_0)$. Then, $q(x_1+x_2,\theta) = r(x_1,\theta) + r(x_2,\theta)$, and hence, $[r(x_1,\theta) - r(0,\theta)] + [r(x_2,\theta) - r(0,\theta)] = r(x_1+x_2,\theta) - r(0,\theta)$.

11. Use Theorems 1.6.2 and 1.6.3 to obtain moment-generating functions for the sufficient statistics when sampling from the following distributions.

(a) normal, $\boldsymbol{\theta} = (\mu,\sigma^2)$

(b) gamma, $\Gamma(p,\lambda)$, $\theta = \lambda$, p fixed

(c) binomial

(d) Poisson

(e) negative binomial (see Problem 1.6.3)

(f) gamma, $\Gamma(p,\lambda)$, $\boldsymbol{\theta} = (p,\lambda)$.

12. Show directly using the definition of the rank of an exponential family that the multinomial distribution, $\mathcal{M}(n; \theta_1, \dots, \theta_k)$, $0 < \theta_j < 1$, $1 \le j \le k$, $\sum_{j=1}^{k} \theta_j = 1$, is of rank $k-1$.

13. Show that in Theorem 1.6.3, the condition that $\mathcal{E}$ has nonempty interior is equivalent to the condition that $\mathcal{E}$ is not contained in any $(k-1)$-dimensional hyperplane.

14. Construct an exponential family of rank k for which $\mathcal{E}$ is not open and $\dot{A}$ is not defined on all of $\mathcal{E}$. Show that if $k = 1$ and $\mathcal{E}^0 \neq \emptyset$ and $\dot{A}, \ddot{A}$ are defined on all of $\mathcal{E}$, then Theorem 1.6.3 continues to hold.

15. Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ where P_θ is discrete and concentrated on $\mathcal{X} = \{x_1, x_2, \dots\}$, and let $p(x, \theta) = P_\theta[X = x]$. Show that if $\mathcal{P}$ is a (discrete) canonical exponential family generated by $(\mathbf{T}, h)$ and $\mathcal{E}^0 \neq \emptyset$, then $\mathbf{T}$ is minimal sufficient.

Hint: $\frac{\partial \log}{\partial \eta_j} L_X(\eta) = T_j(X) - E_\eta T_j(X)$. Use Problem 1.5.12.

16. *Life testing.* Let $X_1, \dots, X_n$ be independently distributed with exponential density $(2\theta)^{-1} e^{-x/2\theta}$ for $x \ge 0$, and let the ordered X's be denoted by $Y_1 \le Y_2 \le \dots \le Y_n$. It is assumed that Y_1 becomes available first, then Y_2, and so on, and that observation is continued until Y_r has been observed. This might arise, for example, in life testing where each X measures the length of life of, say, an electron tube, and n tubes are being tested simultaneously. Another application is to the disintegration of radioactive material, where n is the number of atoms, and observation is continued until r α-particles have been emitted. Show that

(i) The joint distribution of $Y_1, \dots, Y_r$ is an exponential family with density

$$\frac{1}{(2\theta)^r} \frac{n!}{(n-r)!} \exp\left[-\frac{\sum_{i=1}^{r} y_i + (n-r) y_r}{2\theta}\right], \quad 0 \le y_1 \le \dots \le y_r.$$

(ii) The distribution of $[\sum_{i=1}^{r} Y_i + (n-r) Y_r]/\theta$ is χ^2 with $2r$ degrees of freedom.

(iii) Let $Y_1, Y_2, \dots$ denote the time required until the first, second,... event occurs in a Poisson process with parameter $1/2\theta'$ (see A.16). Then $Z_1 = Y_1/\theta'$, $Z_2 = (Y_2 - Y_1)/\theta'$, $Z_3 = (Y_3 - Y_2)/\theta', \dots$ are independently distributed as χ^2 with 2 degrees of freedom, and the joint density of $Y_1, \dots, Y_r$ is an exponential family with density

$$\frac{1}{(2\theta')^r} \exp\left(-\frac{y_r}{2\theta'}\right), \quad 0 \le y_1 \le \dots \le y_r.$$

The distribution of Y_r/θ' is again χ^2 with $2r$ degrees of freedom.

(iv) The same model arises in the application to life testing if the number n of tubes is held constant by replacing each burned-out tube with a new one, and if Y_1 denotes the time at which the first tube burns out, Y_2 the time at which the second tube burns out, and so on, measured from some fixed time.

[(ii): The random variables $Z_i = (n-i+1)(Y_i - Y_{i-1})/\theta$ $(i = 1, \dots, r)$ are independently distributed as χ^2 with 2 degrees of freedom, and $[\sum_{i=1}^{r} Y_i + (n-r)Y_r]/\theta = \sum_{i=1}^{t} Z_i$.]

17. Suppose that $(\mathbf{T}_{k\times 1}, h)$ generate a canonical exponential family $\mathcal{P}$ with parameter $\eta_{k\times 1}$ and $\mathcal{E} = R^k$. Let

$$\mathcal{Q} = \{Q_{\boldsymbol{\theta}} : Q_{\boldsymbol{\theta}} = P_{\boldsymbol{\eta}} \text{ with } \boldsymbol{\eta} = B_{k\times l}\boldsymbol{\theta}_{l\times 1} + \mathbf{c}_{l\times 1}\},\quad l \le k.$$

(a) Show that $\mathcal{Q}$ is the exponential family generated by $\Pi_L \mathbf{T}$ and $h\exp\{c^T\mathbf{T}\}$, where Π_L is the projection matrix of $\mathbf{T}$ onto $\mathcal{L} = \{\boldsymbol{\eta} : \boldsymbol{\eta} = B\boldsymbol{\theta} + \mathbf{c}\}$.

(b) Show that if $\mathcal{P}$ has full rank k and B is of rank l, then $\mathcal{Q}$ has full rank l.
Hint: If B is of rank l, you may assume

$$\Pi_L = B[B^TB]^{-1}B^T.$$

18. Suppose $Y_1, \dots, Y_n$ are independent with $Y_i \sim \mathcal{N}(\beta_1 + \beta_2 z_i, \sigma^2)$, where $z_1, \dots, z_n$ are covariate values not all equal. (See Example 1.6.6.) Show that the family has rank 3. Give the mean vector and the variance matrix of $\mathbf{T}$.

19. *Logistic Regression.* We observe $(\mathbf{z}_1, Y_1), \dots, (\mathbf{z}_n, Y_n)$ where the $Y_1, \dots, Y_n$ are independent, $Y_i \sim \mathcal{B}(n_i, \lambda_i)$. The success probability λ_i depends on the characteristics $\mathbf{z}_i$ of the ith subject, for example, on the covariate vector $\mathbf{z}_i$ = (age, height, blood pressure)^T. The function $l(u) = \log[u/(1-u)]$ is called the *logit* function. In the logistic linear regression model it is assumed that $l(\lambda_i) = \mathbf{z}_i^T\boldsymbol{\beta}$ where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ and $\mathbf{z}_i$ is $d \times 1$. Show that $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ follow an exponential model with rank d iff $\mathbf{z}_1, \dots, \mathbf{z}_d$ are not collinear (linearly independent) (cf. Examples 1.1.4, 1.6.8 and Problem 1.1.9).

20. **(a)** In part II of the proof of Theorem 1.6.4, fill in the details of the arguments that $\mathcal{Q}$ is generated by $(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0)^T T$ and that $\sim$(ii) $\equiv\sim$(i).

(b) Fill in the details of part III of the proof of Theorem 1.6.4.

21. Find $\mu(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}\mathbf{T}(X)$ for the gamma, $\Gamma(\alpha, \lambda)$, distribution, where $\theta = (\alpha, \lambda)$.

22. Let $X_1, \dots, X_n$ be a sample from the k-parameter exponential family distribution (1.6.10). Let $\mathbf{T} = (\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ and let

$$\mathcal{S} = \{(\eta_1(\boldsymbol{\theta}), \dots, \eta_k(\boldsymbol{\theta})) : \boldsymbol{\theta} \in \Theta\}.$$

Show that if $\mathcal{S}$ contains a subset of $k+1$ vectors $\mathbf{v}_0, \dots, \mathbf{v}_{k+1}$ so that $\mathbf{v}_i - \mathbf{v}_0$, $1 \le i \le k$, are not collinear (linearly independent), then $\mathbf{T}$ is minimally sufficient for θ.

23. Using (1.6.20), find a conjugate family of distributions for the gamma and beta families.

(a) With one parameter fixed.

(b) With both parameters free.

24. Using (1.6.20), find a conjugate family of distributions for the normal family using as parameter $\theta = (\theta_1, \theta_2)$ where $\theta_1 = E_\theta(X)$, $\theta_2 = 1/(\text{Var}_\theta X)$ (cf. Problem 1.2.12).

25. Consider the linear Gaussian regression model of Examples 1.5.5 and 1.6.6 except with σ^2 known. Find a conjugate family of prior distributions for $(\beta_1, \beta_2)^T$.

26. Using (1.6.20), find a conjugate family of distributions for the multinomial distribution. See Problem 1.2.15.

27. Let $\mathcal{P}$ denote the canonical exponential family genrated by $\mathbf{T}$ and h. For any $\eta_0 \in \mathcal{E}$, set $h_0(x) = q(x, \eta_0)$ where q is given by (1.6.9). Show that $\mathcal{P}$ is also the canonical exponential family generated by $\mathbf{T}$ and h_0.

28. *Exponential families are maximum entropy distributions.* The *entropy* $h(f)$ of a random variable X with density f is defined by

$$h(f) = E(-\log f(X)) = -\int_{-\infty}^{\infty} [\log f(x)] f(x) dx.$$

This quantity arises naturally in information in theory; see Section 2.2.2 and Cover and Thomas (1991). Let $S = \{x : f(x) > 0\}$.

(a) Show that the canonical k-parameter exponential family density

$$f(x, \eta) = \exp\left\{\eta_0 + \sum_{j=1}^{k} \eta_j r_j(x) - A(\eta)\right\}, \ x \in S$$

maximizes $h(f)$ subject to the constraints

$$f(x) \geq 0, \ \int_S f(x) dx = 1, \ \int_S f(x) r_j(x) = \alpha_j, \ 1 \leq j \leq k,$$

where $\eta_0, \dots, \eta_k$ are chosen so that f satisfies the constraints.

Hint: You may use Lagrange multipliers. Maximize the integrand.

(b) Find the maximum entropy densities when $r_j(x) = x^j$ and (i) $S = (0, \infty)$, $k = 1$, $\alpha_1 > 0$; (ii) $S = R$, $k = 2$, $\alpha_1 \in R$, $\alpha_2 > 0$; (iii) $S = R$, $k = 3$, $\alpha_1 \in R$, $\alpha_2 > 0$, $\alpha_3 \in R$.

29. As in Example 1.6.11, suppose that $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are i.i.d. $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu}$ varies freely in R^p and Σ ranges freely over the class of all $p \times p$ symmetric positive definite matrices. Show that the distribution of $\mathbf{Y} = (Y_1, \dots, Y_n)$ is the $p(p+3)/2$ canonical exponential family generated by $h = 1$ and the $p(p+3)/2$ statistics

$$T_j = \sum_{i=1}^{n} \mathbf{Y}_{ij}, \ 1 \leq j \leq p; \ T_{jl} = \sum_{i=1}^{n} Y_{ij} Y_{il}, \ 1 \leq j \leq l \leq p$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})$. Show that $\mathcal{E}$ is open and that this family is of rank $p(p+3)/2$.

Hint: Without loss of generality, take $n = 1$. We want to show that $h = 1$ and the $m = p(p+3)/2$ statistics $T_j(\mathbf{Y}) = Y_j$, $1 \leq j \leq p$, and $T_{jl}(\mathbf{Y}) = Y_j Y_l$, $1 \leq j \leq l \leq p$, generate $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. As Σ ranges over all $p \times p$ symmetric positive definite matrices, so does Σ^{-1}. Next establish that for symmetric matrices M,

$$\int \exp\{-\mathbf{u}^T M \mathbf{u}\} d\mathbf{u} < \infty \text{ iff } M \text{ is positive definite}$$

by using the spectral decomposition (see B.10.1.2)

$$M = \sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j^T \text{ for } \mathbf{e}_1, \dots, \mathbf{e}_p \text{ orthogonal, } \lambda_j \in R.$$

To show that the family has full rank m, use induction on p to show that if $Z_1, \dots, Z_p$ are i.i.d. $\mathcal{N}(0, 1)$ and if $B_{p \times p} = (b_{jl})$ is symmetric, then

$$P\left(\sum_{j=1}^{p} a_j Z_j + \sum_{j,l} b_{jl} Z_j Z_l = c\right) = P(\mathbf{a}^T \mathbf{Z} + \mathbf{Z}^T B \mathbf{Z} = c) = 0$$

unless $\mathbf{a} = \mathbf{0}$, $B = 0$, $c = 0$. Next recall (Appendix B.6) that since $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, then $\mathbf{Y} = S\mathbf{Z}$ for some nonsingular $p \times p$ matrix S.

30. Show that if $\mathbf{X}_1, \dots, \mathbf{X}_n$ are i.i.d. $\mathcal{N}_p(\boldsymbol{\theta}, \Sigma_0)$ given $\boldsymbol{\theta}$ where Σ_0 is known, then the $\mathcal{N}_p(\boldsymbol{\lambda}, \Gamma)$ family is conjugate to $\mathcal{N}_p(\boldsymbol{\theta}, \Sigma_0)$, where $\boldsymbol{\lambda}$ varies freely in R^p and Γ ranges over all $p \times p$ symmetric positive definite matrices.

31. *Conjugate Normal Mixture Distributions. A Hierarchical Bayesian Normal Model.* Let $\{(\mu_j, \tau_j) : 1 \le j \le k\}$ be a given collection of pairs with $\mu_j \in R$, $\tau_j > 0$. Let $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ be a random pair with $\lambda_j = P((\boldsymbol{\mu}, \boldsymbol{\sigma}) = (\mu_j, \tau_j))$, $0 < \lambda_j < 1$, $\sum_{j=1}^{k} \lambda_j = 1$. Let $\boldsymbol{\theta}$ be a random variable whose conditional distribution given $(\boldsymbol{\mu}, \boldsymbol{\sigma}) = (\mu_j, \tau_j)$ is normal, $\mathcal{N}(\mu_j, \tau_j^2)$. Consider the model $X = \boldsymbol{\theta} + \epsilon$, where $\boldsymbol{\theta}$ and ϵ are independent and $\epsilon \sim \mathcal{N}(0, \sigma_0^2)$, σ_0^2 known. Note that $\boldsymbol{\theta}$ has the prior density

$$\pi(\theta) = \sum_{j=1}^{k} \lambda_j \varphi_{\tau_j}(\theta - \mu_j) \tag{1.7.4}$$

where φ_τ denotes the $\mathcal{N}(0, \tau^2)$ density. Also note that $(X \mid \theta)$ has the $\mathcal{N}(\theta, \sigma_0^2)$ distribution.

(a) Find the posterior

$$\pi(\theta \mid x) = \sum_{j=1}^{k} P((\boldsymbol{\mu}, \boldsymbol{\sigma}) = (\mu_j, \tau_j) \mid x) \pi(\theta \mid (\mu_j, \tau_j), x)$$

and write it in the form

$$\sum_{j=1}^{k} \lambda_j(x) \varphi_{\tau_j(x)}(\theta - \mu_j(x))$$

for appropriate $\lambda_j(x)$, $\tau_j(x)$ and $\mu_j(x)$. This shows that (1.7.4) defines a conjugate prior for the $\mathcal{N}(\theta, \sigma_0^2)$, distribution.

(b) Let $X_i = \theta + \epsilon_i$, $1 \leq i \leq n$, where θ is as previously and $\epsilon_1, \dots, \epsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma_0^2)$. Find the posterior $\pi(\theta \mid x_1, \dots, x_n)$, and show that it belongs to class (1.7.4).

Hint: Consider the sufficient statistic for $p(\mathbf{x} \mid \theta)$.

32. *A Hierarchical Binomial–Beta Model.* Let $\{(r_j, s_j) : 1 \leq j \leq k\}$ be a given collection of pairs with $r_j > 0$, $s_j > 0$, let (R, S) be a random pair with $P(R = r_j, S = s_j) = \lambda_j$, $0 < \lambda_j < 1$, $\sum_{j=1}^k \lambda_j = 1$, and let θ be a random variable whose conditional density $\pi(\theta, r, s)$ given $R = r$, $S = s$ is beta, $\beta(r, s)$. Consider the model in which $(X \mid \theta)$ has the binomial, $\mathcal{B}(n, \theta)$, distribution. Note that θ has the prior density

$$\pi(\theta) = \sum_{j=1}^{k} \lambda_j \pi(\theta, r_j, s_j). \tag{1.7.5}$$

Find the posterior

$$\pi(\theta \mid x) = \sum_{j=1}^{k} P(R = r_j, S = s_j \mid x) \pi(\theta \mid (r_j, s_j), x)$$

and show that it can be written in the form $\sum \lambda_j(x) \pi(\theta, r_j(\mathbf{x}), s_j(\mathbf{x}))$ for appropriate $\lambda_j(x)$, $r_j(x)$ and $s_j(x)$. This shows that (1.7.5) defines a class of conjugate priors for the $\mathcal{B}(n, \theta)$ distribution.

33. Let $p(x, \eta)$ be a one parameter canonical exponential family generated by $T(x) = x$ and $h(x)$, $x \in \mathcal{X} \subset R$, and let $\psi(x)$ be a nonconstant, nondecreasing function. Show that $E_\eta \psi(X)$ is strictly increasing in η.

Hint:

$$\begin{aligned} \frac{\partial}{\partial \eta} E_\eta \psi(X) &= \operatorname{Cov}_\eta(\psi(X), X) \\ &= \frac{1}{2} E\{(X - X')[\psi(X) - \psi(X')]\} \end{aligned}$$

where X and X' are independent identically distributed as X (see A.11.12).

34. Let $(X_1, \dots, X_n)$ be a stationary Markov chain with two states 0 and 1. That is,

$$P[X_i = \epsilon_i \mid X_1 = \epsilon_1, \dots, X_{i-1} = \epsilon_{i-1}] = P[X_i = \epsilon_i \mid X_{i-1} = \epsilon_{i-1}] = p_{\epsilon_{i-1}\epsilon_i}$$

where $\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ is the matrix of transition probabilities. Suppose further that

(i) $p_{00} = p_{11} = p$, so that, $p_{10} = p_{01} = 1 - p$.

(ii) $P[X_1 = 0] = P[X_1 = 1] = \frac{1}{2}$.

(a) Show that if $0 < p < 1$ is unknown this is a full rank, one-parameter exponential family with $T = N_{00} + N_{11}$ where $N_{ij} \equiv$ the number of transitions from i to j. For example, 01011 has $N_{01} = 2, N_{11} = 1, N_{00} = 0, N_{10} = 1$.

(b) Show that $E(T) = (n-1)p$ (by the method of indicators or otherwise).

35. *A Conjugate Prior for the Two-Sample Problem.* Suppose that $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ are independent $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ samples, respectively. Consider the prior π for which for some $r > 0$, $k > 0$, $r\sigma^{-2}$ has a χ^2_k distribution and given σ^2, μ_1 and μ_2 are independent with $\mathcal{N}(\xi_1, \sigma^2/k_1)$ and $\mathcal{N}(\xi_2, \sigma^2/k_2)$ distributions, respectively, where $\xi_j \in R$, $k_j > 0$, $j = 1, 2$. Show that π is a conjugate prior.

36. The *inverse Gaussian* density, $IG(\mu, \lambda)$, is

$$f(x, \mu, \lambda) = [\lambda/2\pi]^{1/2} x^{-3/2} \exp\{-\lambda(x-\mu)^2/2\mu^2 x\}, \quad x > 0, \quad \mu > 0, \quad \lambda > 0.$$

(a) Show that this is an exponential family generated by $\mathbf{T}(X) = -\frac{1}{2}(X, X^{-1})^T$ and $h(x) = (2\pi)^{-1/2} x^{-3/2}$.

(b) Show that the canonical parameters η_1, η_2 are given by $\eta_1 = \mu^{-2}\lambda$, $\eta_2 = \lambda$, and that $A(\eta_1, \eta_2) = -\left[\frac{1}{2}\log(\eta_2) + \sqrt{\eta_1 \eta_2}\right]$, $\mathcal{E} = [0, \infty) \times (0, \infty)$.

(c) Find the moment-generating function of $\mathbf{T}$ and show that $E(X) = \mu$, $\text{Var}(X) = \mu^{-3}\lambda$, $E(X^{-1}) = \mu^{-1} + \lambda^{-1}$, $\text{Var}(X^{-1}) = (\lambda\mu)^{-1} + 2\lambda^{-2}$.

(d) Suppose $\mu = \mu_0$ is known. Show that the gamma family, $\Gamma(\alpha, \beta)$, is a conjugate prior.

(e) Suppose that $\lambda = \lambda_0$ is known. Show that the conjugate prior formula (1.6.20) produces a function that is not integrable with respect to μ. That is, Ω defined in (1.6.19) is empty.

(f) Suppose that μ and λ are both unknown. Show that (1.6.20) produces a function that is not integrable; that is, Ω defined in (1.6.19) is empty.

37. Let $X_1, \dots, X_n$ be i.i.d. as $X \sim \mathcal{N}_p(\boldsymbol{\theta}, \Sigma_0)$ where Σ_0 is known. Show that the conjugate prior generated by (1.6.20) is the $\mathcal{N}_p(\boldsymbol{\eta}_0, \tau_0^2 \mathbf{I})$ family, where $\boldsymbol{\eta}_0$ varies freely in R^p, $\tau_0^2 > 0$ and $\mathbf{I}$ is the $p \times p$ identity matrix.

38. Let $X_i = (Z_i, Y_i)^T$ be i.i.d. as $X = (Z, Y)^T$, $1 \le i \le n$, where X has the density of Example 1.6.3. Write the density of $X_1, \dots, X_n$ as a canonical exponential family and identify T, h, A, and $\mathcal{E}$. Find the expected value and variance of the sufficient statistic.

39. Suppose that $Y_1, \dots, Y_n$ are independent, $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $n \ge 4$.

(a) Write the distribution of $Y_1, \dots, Y_n$ in canonical exponential family form. Identify $\mathbf{T}$, h, $\boldsymbol{\eta}$, A, and $\mathcal{E}$.

(b) Next suppose that μ_i depends on the value z_i of some covariate and consider the submodel defined by the map $\boldsymbol{\eta} : (\theta_1, \theta_2, \theta_3)^T \to (\boldsymbol{\mu}^T, \sigma^2)^T$ where $\boldsymbol{\eta}$ is determined by

$$\mu_i = \exp\{\theta_1 + \theta_2 z_i\}, \; z_1 < z_2 < \dots < z_n; \; \sigma^2 = \theta_3$$

where $\theta_1 \in R$, $\theta_2 \in R$, $\theta_3 > 0$. This model is sometimes used when μ_i is restricted to be positive. Show that $p(\mathbf{y}, \boldsymbol{\theta})$ as given by (1.6.12) is a curved exponential family model with $l = 3$.

40. Suppose $Y_1, \dots, Y_n$ are independent exponentially, $\mathcal{E}(\lambda_i)$, distributed survival times, $n \geq 3$.

(a) Write the distribution of $Y_1, \dots, Y_n$ in canonical exponential family form. Identify $\mathbf{T}$, h, $\boldsymbol{\eta}$, A, and $\mathcal{E}$.

(b) Recall that $\mu_i = E(Y_i) = \lambda_i^{-1}$. Suppose μ_i depends on the value z_i of a covariate. Because $\mu_i > 0$, μ_i is sometimes modeled as

$$\mu_i = \exp\{\theta_1 + \theta_2 z_i\},\ i = 1, \dots, n$$

where not all the z's are equal. Show that $p(\mathbf{y}, \boldsymbol{\theta})$ as given by (1.6.12) is a curved exponential family model with $l = 2$.

1.8 NOTES

Note for Section 1.1

(1) For the measure theoretically minded we can assume more generally that the $P_{\boldsymbol{\theta}}$ are all dominated by a σ finite measure μ and that $p(x, \theta)$ denotes $\frac{dP_\theta}{d\mu}$, the Radon Nikodym derivative.

Notes for Section 1.3

(1) More natural in the sense of measuring the Euclidean distance between the estimate $\widehat{\theta}$ and the "truth" θ. Squared error gives much more weight to those $\widehat{\theta}$ that are far away from θ than those close to θ.

(2) We define the lower boundary of a convex set simply to be the set of all boundary points r such that the set lies completely on or above any tangent to the set at r.

Note for Section 1.4

(1) Source: Hodges, Jr., J. L., D. Kretch, and R. S. Crutchfield. *Statlab: An Empirical Introduction to Statistics*. New York: McGraw-Hill, 1975.

Notes for Section 1.6

(1) Exponential families arose much earlier in the work of Boltzmann in statistical mechanics as laws for the distribution of the states of systems of particles—see Feynman (1963), for instance. The connection is through the concept of entropy, which also plays a key role in information theory—see Cover and Thomas (1991).

(2) The restriction that's $x \in R^q$ and that these families be discrete or continuous is artificial. In general if μ is a σ finite measure on the sample space $\mathcal{X}$, $p(x, \theta)$ as given by (1.6.1)

can be taken to be the density of X with respect to μ—see Lehmann (1997), for instance. This permits consideration of data such as images, positions, and spheres (e.g., the Earth), and so on.

Note for Section 1.7

(1) $\mathbf{u}^T M \mathbf{u} > 0$ for all $p \times 1$ vectors $\mathbf{u} \neq 0$.

1.9 REFERENCES

BERGER, J. O., *Statistical Decision Theory and Bayesian Analysis* New York: Springer, 1985.

BERMAN, S. M., "A Stochastic Model for the Distribution of HIV Latency Time Based on T4 Counts," *Biometika, 77*, 733–741 (1990).

BICKEL, P. J., "Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity," *Ann. Statist. 6*, 266–291 (1978).

BLACKWELL, D. AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions* New York: Wiley, 1954.

BOX, G. E. P., "Sampling and Bayes Inference in Scientific Modelling and Robustness (with Discussion)," *J. Royal Statist. Soc. A 143*, 383–430 (1979).

BROWN, L., *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, IMS Lecture Notes—Monograph Series, Hayward, 1986.

CARROLL, R. J. AND D. RUPPERT, *Transformation and Weighting in Regression* New York: Chapman and Hall, 1988.

COVER, T. M. AND J. A. THOMAS, *Elements of Information Theory* New York: Wiley, 1991.

DE GROOT, M. H., *Optimal Statistical Decisions* New York: McGraw–Hill, 1969.

DOKSUM, K. A. AND A. SAMAROV, "Nonparametric Estimation of Global Functionals and a Measure of the Explanatory Power of Covariates in Regression," *Ann. Statist. 23*, 1443–1473 (1995).

FERGUSON, T. S., *Mathematical Statistics* New York: Academic Press, 1967.

FEYNMAN, R. P., *The Feynman Lectures on Physics*, v. 1, R. P. Feynman, R. B. Leighton, and M. Sands, Eds., Ch. 40 *Statistical Mechanics of Physics* Reading, MA: Addison-Wesley, 1963.

GRENANDER, U. AND M. ROSENBLATT, *Statistical Analysis of Stationary Time Series* New York: Wiley, 1957.

HODGES, JR., J. L., D. KRETCH AND R. S. CRUTCHFIELD, *Statlab: An Empirical Introduction to Statistics* New York: McGraw–Hill, 1975.

KENDALL, M. G. AND A. STUART, *The Advanced Theory of Statistics*, Vols. II, III New York: Hafner Publishing Co., 1961, 1966.

LEHMANN, E. L., "A Theory of Some Multiple Decision Problems, I and II," *Ann. Math. Statist. 22*, 1–25, 547–572 (1957).

LEHMANN, E. L., "Model Specification: The Views of Fisher and Neyman, and Later Developments," *Statist. Science 5*, 160–168 (1990).

LEHMANN, E. L., *Testing Statistical Hypotheses*, 2nd ed. New York: Springer, 1997.