Theorem 4.1 (Bayes formula). Assume that $\mathcal{P} = \{P_{x|\theta} : \theta \in \Theta\}$ is dominated by a σ-finite measure ν and $f_\theta(x) = \frac{dP_{x|\theta}}{d\nu}(x)$ is a Borel function on $(\mathfrak{X} \times \Theta, \sigma(\mathcal{B}_{\mathfrak{X}} \times \mathcal{B}_\Theta))$. Let Π be a prior distribution on Θ. Suppose that $m(x) = \int_\Theta f_\theta(x) d\Pi > 0$.
(i) The posterior distribution $P_{\theta|x} \ll \Pi$ and

$$\frac{dP_{\theta|x}}{d\Pi} = \frac{f_\theta(x)}{m(x)}.$$

(ii) If $\Pi \ll \lambda$ and $\frac{d\Pi}{d\lambda} = \pi(\theta)$ for a σ-finite measure λ, then

$$\frac{dP_{\theta|x}}{d\lambda} = \frac{f_\theta(x)\pi(\theta)}{m(x)}. \tag{4.1}$$

Definition 4.1. Let $\mathbb{A}$ be an action space in a decision problem and $L(\theta, a) \geq 0$ be a loss function. For any $x \in \mathfrak{X}$, a *Bayes action* w.r.t. Π is any $\delta(x) \in \mathbb{A}$ such that

$$E[L(\boldsymbol{\theta}, \delta(x))|X = x] = \min_{a \in \mathbb{A}} E[L(\boldsymbol{\theta}, a)|X = x], \tag{4.3}$$

where the expectation is w.r.t. the posterior distribution $P_{\theta|x}$. ∎

Proposition 4.1. Assume that the conditions in Theorem 4.1 hold; $L(\theta, a)$ is convex in a for each fixed θ; and for each $x \in \mathfrak{X}$, $E[L(\boldsymbol{\theta}, a)|X = x] < \infty$ for some a.
(i) If $\mathbb{A}$ is a compact subset of $\mathcal{R}^p$ for some integer $p \geq 1$, then a Bayes action $\delta(x)$ exists for each $x \in \mathfrak{X}$.
(ii) If $\mathbb{A} = \mathcal{R}^p$ and $L(\theta, a)$ tends to ∞ as $\|a\| \to \infty$ uniformly in $\theta \in \Theta_0 \subset \Theta$ with $\Pi(\Theta_0) > 0$, then a Bayes action $\delta(x)$ exists for each $x \in \mathfrak{X}$.
(iii) In (i) or (ii), if $L(\theta, a)$ is strictly convex in a for each fixed θ, then the Bayes action is unique.

Theorem 4.2. In a decision problem, let $\delta(X)$ be a Bayes rule w.r.t. a prior Π.
(i) If $\delta(X)$ is a unique Bayes rule, then $\delta(X)$ is admissible.
(ii) If Θ is a countable set, the Bayes risk $r_\delta(\Pi) < \infty$, and Π gives positive probability to each $\theta \in \Theta$, then $\delta(X)$ is admissible.
(iii) Let $\Im$ be the class of decision rules having continuous risk functions. If $\delta(X) \in \Im$, $r_\delta(\Pi) < \infty$, and Π gives positive probability to any open subset of Θ, then $\delta(X)$ is $\Im$-admissible. ∎

Theorem 4.3. Suppose that Θ is an open set of $\mathcal{R}^k$. In a decision problem, let $\Im$ be the class of decision rules having continuous risk functions. A decision rule $T \in \Im$ is $\Im$-admissible if there exists a sequence $\{\Pi_j\}$ of (possibly improper) priors such that (a) the generalized Bayes risks $r_T(\Pi_j)$ are finite for all j; (b) for any $\theta_0 \in \Theta$ and $\eta > 0$,

$$\lim_{j \to \infty} \frac{r_T(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\theta_0,\eta})} = 0,$$

where $r_j^*(\Pi_j) = \inf_{T \in \Im} r_T(\Pi_j)$ and $O_{\theta_0,\eta} = \{\theta \in \Theta : \|\theta - \theta_0\| < \eta\}$ with $\Pi_j(O_{\theta_0,\eta}) < \infty$ for all j.

Proposition 4.2. Let $\delta(X)$ be a Bayes estimator of $\vartheta = g(\theta)$ under the squared error loss. Then $\delta(X)$ is not unbiased unless the Bayes risk $r_\delta(\Pi) = 0$.

Lemma 4.1. Suppose that X has a p.d.f. $f_\theta(x)$ w.r.t. a σ-finite measure ν. Suppose that $\theta = (\theta_1, \theta_2)$, $\theta_j \in \Theta_j$, and that the prior has a p.d.f.

$$\pi(\theta) = \pi_{\theta_1|\theta_2}(\theta_1)\pi_{\theta_2}(\theta_2),$$

where $\pi_{\theta_2}(\theta_2)$ is a p.d.f. w.r.t. a σ-finite measure ν_2 on Θ_2 and for any given θ_2, $\pi_{\theta_1|\theta_2}(\theta_1)$ is a p.d.f. w.r.t. a σ-finite measure ν_1 on Θ_1. Suppose further that if θ_2 is given, the Bayes estimator of $h(\theta_1) = g(\theta_1, \theta_2)$ under the squared error loss is $\delta(X, \theta_2)$. Then the Bayes estimator of $g(\theta_1, \theta_2)$ under the squared error loss is $\delta(X)$ with

$$\delta(x) = \int_{\Theta_2} \delta(x, \theta_2) p_{\theta_2|x}(\theta_2) d\nu_2,$$

where $p_{\theta_2|x}(\theta_2)$ is the posterior p.d.f. of $\boldsymbol{\theta}_2$ given $X = x$. ∎

Theorem 4.11. Let Π be a proper prior on Θ and δ be a Bayes estimator of ϑ w.r.t. Π. Let $\Theta_\Pi = \{\theta : R_\delta(\theta) = \sup_{\theta \in \Theta} R_\delta(\theta)\}$. If $\Pi(\Theta_\Pi) = 1$, then δ is minimax. If, in addition, δ is the unique Bayes estimator w.r.t. Π, then it is the unique minimax estimator.

Theorem 4.12. Let Π_j, $j = 1, 2, ...$, be a sequence of priors and r_j be the Bayes risk of a Bayes estimator of ϑ w.r.t. Π_j. Let T be a constant risk estimator of ϑ. If $\liminf_j r_j \geq R_T$, then T is minimax. ∎

Lemma 4.3. Let Θ_0 be a subset of Θ and T be a minimax estimator of ϑ when Θ_0 is the parameter space. Then T is a minimax estimator if

$$\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta).$$

Theorem 4.13. Suppose that T as an estimator of ϑ has constant risk and is admissible. Then T is minimax. If the loss function is strictly convex, then T is the unique minimax estimator.

Theorem 4.14. Suppose that X has the p.d.f. $c(\theta)e^{\theta T(x)}$ w.r.t. a σ-finite measure ν, where $T(x)$ is real-valued and $\theta \in (\theta_-, \theta_+) \subset \mathcal{R}$. Consider the estimation of $\vartheta = E[T(X)]$ under the squared error loss. Let $\lambda \geq 0$ and γ be known constants and let $T_{\lambda,\gamma}(X) = (T + \gamma\lambda)/(1 + \lambda)$. Then a sufficient condition for the admissibility of $T_{\lambda,\gamma}$ is that

$$\int_{\theta_0}^{\theta_+} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} d\theta = \int_{\theta_-}^{\theta_0} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} d\theta = \infty, \tag{4.34}$$

where $\theta_0 \in (\theta_-, \theta_+)$.

Corollary 4.3. Assume that X has the p.d.f. as described in Theorem 4.14 with $\theta_- = -\infty$ and $\theta_+ = \infty$.
(i) As an estimator of $\vartheta = E(T)$, $T(X)$ is admissible under the squared error loss and the loss $(a - \vartheta)^2/\mathrm{Var}(T)$.
(ii) T is the unique minimax estimator of ϑ under the loss $(a - \vartheta)^2/\mathrm{Var}(T)$.

Theorem 4.15. Suppose that X is from $N_p(\theta, I_p)$ with $p \geq 3$. Then, under the loss function (4.37), the risks of the following estimators of θ,

$$\delta_{c,r} = X - \frac{r(p-2)}{\|X - c\|^2}(X - c), \tag{4.41}$$

are given by

$$R_{\delta_{c,r}}(\theta) = p - (2r - r^2)(p-2)^2 E(\|X - c\|^{-2}), \tag{4.42}$$

where $c \in \mathcal{R}^p$ and $r \in \mathcal{R}$ are known.

Theorem 4.16. Let $X_1, ..., X_n$ be i.i.d. from a p.d.f. f_θ w.r.t. a σ-finite measure ν on $(\mathcal{R}, \mathcal{B})$, where $\theta \in \Theta$ and Θ is an open set in $\mathcal{R}^k$. Suppose that for every x in the range of X_1, $f_\theta(x)$ is twice continuously differentiable in θ and satisfies

$$\frac{\partial}{\partial\theta} \int \psi_\theta(x) d\nu = \int \frac{\partial}{\partial\theta} \psi_\theta(x) d\nu$$

for $\psi_\theta(x) = f_\theta(x)$ and $= \partial f_\theta(x)/\partial\theta$; the Fisher information matrix

$$I_1(\theta) = E\left\{\frac{\partial}{\partial\theta} \log f_\theta(X_1) \left[\frac{\partial}{\partial\theta} \log f_\theta(X_1)\right]^\tau\right\}$$

is positive definite; and for any given $\theta \in \Theta$, there exists a positive number c_θ and a positive function h_θ such that $E[h_\theta(X_1)] < \infty$ and

$$\sup_{\gamma: \|\gamma - \theta\| < c_\theta} \left\|\frac{\partial^2 \log f_\gamma(x)}{\partial\gamma\partial\gamma^\tau}\right\| \leq h_\theta(x) \tag{4.69}$$

for all x in the range of X_1, where $\|A\| = \sqrt{\mathrm{tr}(A^\tau A)}$ for any matrix A. If $\hat{\theta}_n$ is an estimator of θ (based on $X_1, ..., X_n$) and satisfies (4.67) with $V_n(\theta) = V(\theta)/n$, then there is a $\Theta_0 \subset \Theta$ with Lebesgue measure 0 such that (4.68) holds if $\theta \notin \Theta_0$.

Definition 4.4. Assume that the Fisher information matrix $I_n(\theta)$ is well defined and positive definite for every n. A sequence of estimators $\{\hat{\theta}_n\}$ satisfying (4.67) is said to be *asymptotically efficient* or *asymptotically optimal* if and only if $V_n(\theta) = [I_n(\theta)]^{-1}$. ∎

Theorem 4.17. Assume the conditions of Theorem 4.16.
(i) There is a sequence of estimators $\{\hat{\theta}_n\}$ such that

$$P(s_n(\hat{\theta}_n) = 0) \to 1 \quad \text{and} \quad \hat{\theta}_n \to_p \theta, \tag{4.74}$$

where $s_n(\gamma) = \partial \log \ell(\gamma)/\partial\gamma$.
(ii) Any consistent sequence $\tilde{\theta}_n$ of RLE's is asymptotically efficient.

Theorem 4.18. Consider the GLM (4.55)-(4.58) with t_i's in a fixed interval (t_0, t_∞), $0 < t_0 \leq t_\infty < \infty$. Assume that the range of the unknown parameter β in (4.57) is an open subset of $\mathcal{R}^p$; at the true parameter value β, $0 < \inf_i \varphi(\beta^\tau Z_i) \leq \sup_i \varphi(\beta^\tau Z_i) < \infty$, where $\varphi(t) = [\psi'(t)]^2\zeta''(\psi(t))$; as $n \to \infty$, $\max_{i \leq n} Z_i^\tau(Z^\tau Z)^{-1}Z_i \to 0$ and $\lambda_-[Z^\tau Z] \to \infty$, where Z is the $n \times p$ matrix whose ith row is the vector Z_i and $\lambda_-[A]$ is the smallest eigenvalue of the matrix A.
(i) There is a unique sequence of estimators $\{\hat{\beta}_n\}$ such that

$$P(s_n(\hat{\beta}_n) = 0) \to 1 \quad \text{and} \quad \hat{\beta}_n \to_p \beta, \tag{4.80}$$

where $s_n(\gamma)$ is the score function defined to be the left-hand side of (4.59) with $\gamma = \beta$.
(ii) Let $I_n(\beta) = \mathrm{Var}(s_n(\beta))$. Then

$$[I_n(\beta)]^{1/2}(\hat{\beta}_n - \beta) \to_d N_p(0, I_p). \tag{4.81}$$

(iii) If ϕ in (4.58) is known or the p.d.f. in (4.55) indexed by $\theta = (\beta, \phi)$ satisfies the conditions for f_θ in Theorem 4.16, then $\hat{\beta}_n$ is asymptotically efficient.

Theorem 5.2. Let F_n be the empirical c.d.f. based on i.i.d. random variables $X_1, ..., X_n$ from a c.d.f. $F \in \mathcal{F}_1$. Then
(i) $\varrho_{L_p}(F_n, F) \to_{a.s.} 0$;
(ii) $E[\sqrt{n}\varrho_{L_p}(F_n, F)] = O(1)$ if $1 \leq p < 2$ and $\int \{F(t)[1 - F(t)]\}^{p/2} dt < \infty$, or $p \geq 2$.

Theorem 5.3. Let $X_1, ..., X_n$ be i.i.d. with $F \in \mathcal{F}$ and $\ell(G)$ be defined by (5.8). Then F_n maximizes $\ell(G)$ over $G \in \mathcal{F}$.
Proof. We only need to consider $G \in \mathcal{F}$ such that $\ell(G) > 0$. Let $c \in (0, 1]$ and $\mathcal{F}(c)$ be the subset of $\mathcal{F}$ containing G's satisfying $p_i = P_G(\{x_i\}) > 0$, $i = 1, ..., n$, and $\sum_{i=1}^n p_i = c$. We now apply the Lagrange multiplier method to solve the problem of maximizing $\ell(G)$ over $G \in \mathcal{F}(c)$. Define

$$H(p_1, ..., p_n, \lambda) = \prod_{i=1}^n p_i + \lambda\left(\sum_{i=1}^n p_i - c\right),$$

where λ is the Lagrange multiplier. Set

$$\frac{\partial H}{\partial\lambda} = \sum_{i=1}^n p_i - c = 0, \qquad \frac{\partial H}{\partial p_j} = p_j^{-1}\prod_{i=1}^n p_i + \lambda = 0, \qquad j = 1, ..., n.$$

The solution is $p_i = c/n$, $i = 1, ..., n$, $\lambda = -(c/n)^{n-1}$. It can be shown (exercise) that this solution is a maximum of $H(p_1, ..., p_n, \lambda)$ over $p_i > 0$, $i = 1, ..., n$, $\sum_{i=1}^n p_i = c$. This shows that

$$\max_{G \in \mathcal{F}(c)} \ell(G) = (c/n)^n,$$

which is maximized at $c = 1$ for any fixed n. The result follows from $P_{F_n}(\{x_i\}) = n^{-1}$ for given $X_i = x_i$, $i = 1, ..., n$. ∎

$$p_i > 0, \quad i = 1, ..., n, \quad \sum_{i=1}^n p_i = 1, \quad \text{and} \quad \sum_{i=1}^n p_i u(x_i) = 0, \tag{5.10}$$

where $p_i = P_G(\{x_i\})$. Using the Lagrange multiplier method and an argument similar to the proof of Theorem 5.3, it can be shown (exercise) that an MELE of F is

$$\hat{F}(t) = \sum_{i=1}^n \hat{p}_i I_{(-\infty,t]}(X_i), \tag{5.11}$$

where the notation $(-\infty, t]$ is the same as that in (5.1),

$$\hat{p}_i = n^{-1}[1 + \lambda_n^\tau u(X_i)]^{-1}, \qquad i = 1, ..., n, \tag{5.12}$$

and $\lambda_n \in \mathcal{R}^s$ is the Lagrange multiplier satisfying

$$\sum_{i=1}^n \hat{p}_i u(X_i) = \frac{1}{n}\sum_{i=1}^n \frac{u(X_i)}{1 + \lambda_n^\tau u(X_i)} = 0. \tag{5.13}$$

Theorem 5.4. Let $X_1, ..., X_n$ be i.i.d. with $F \in \mathcal{F}$, u be a Borel function on $\mathcal{R}^d$ satisfying (5.9), and $\hat{F}$ be given by (5.11)-(5.13). Suppose that $U = \mathrm{Var}(u(X_1))$ is positive definite. Then, for any m fixed distinct $t_1, ..., t_m$ in $\mathcal{R}^d$,

$$\sqrt{n}[(\hat{F}(t_1), ..., \hat{F}(t_m)) - (F(t_1), ..., F(t_m))] \to_d N_m(0, \Sigma_u), \tag{5.15}$$

where

$$\Sigma_u = \Sigma - W^\tau U^{-1} W,$$

Σ is given in (5.2), $W = (W(t_1), ..., W(t_m))$, $W(t_j) = E[u(X_1)I_{(-\infty,t_j]}(X_1)]$, and the notation $(-\infty, t]$ is the same as that in (5.1).

$$f_n(t) = \frac{F_n(t + \lambda_n) - F_n(t - \lambda_n)}{2\lambda_n}, \qquad t \in \mathcal{R}, \tag{5.26}$$

$$E[f_n(t)] \to f(t) \qquad \text{if } \lambda_n \to 0 \text{ as } n \to \infty$$

and

$$\mathrm{Var}(f_n(t)) \to 0 \qquad \text{if } \lambda_n \to 0 \text{ and } n\lambda_n \to \infty.$$

Thus, we should choose λ_n converging to 0 slower than n^{-1}. If we assume that $\lambda_n \to 0$, $n\lambda_n \to \infty$, and f is continuously differentiable at t, then it can be shown (exercise) that

$$\mathrm{mse}_{f_n(t)}(F) = \frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right) + O(\lambda_n^2) \tag{5.27}$$

and, under the additional condition that $n\lambda_n^3 \to 0$,

$$\sqrt{n\lambda_n}[f_n(t) - f(t)] \to_d N(0, \tfrac{1}{2}f(t)). \tag{5.28}$$

$u(x)$ replaced by $\psi(x, \theta)$, where ψ is a known function from $\mathcal{R}^d \times \mathcal{R}^k$ to $\mathcal{R}^s$. Maximizing this empirical likelihood is equivalent to maximizing

$$\ell(p_1, ..., p_n, \omega, \lambda, \theta) = \prod_{i=1}^n p_i + \omega\left(1 - \sum_{i=1}^n p_i\right) + \sum_{i=1}^n p_i \lambda^\tau \psi(x_i, \theta),$$

where ω and λ are Lagrange multipliers. It follows from (5.12) and (5.13) that $\omega = n$, $\hat{p}_i(\theta) = n^{-1}\{1 + [\lambda_n(\theta)]^\tau \psi(x_i, \theta)\}^{-1}$ with a $\lambda_n(\theta)$ satisfying

$$\frac{1}{n}\sum_{i=1}^n \frac{\psi(x_i, \theta)}{1 + [\lambda_n(\theta)]^\tau \psi(x_i, \theta)} = 0$$

maximize $\ell(p_1, ...p_n, \omega, \lambda, \theta)$ for any fixed θ. Substituting $\hat{p}_i$ with $\sum_{i=1}^n \hat{p}_i = 1$ into $\ell(p_1, ...p_n, \omega, \lambda, \theta)$ leads to the following profile empirical likelihood for θ:

$$\ell_P(\theta) = \prod_{i=1}^n \frac{1}{n\{1 + [\lambda_n(\theta)]^\tau \psi(x_i, \theta)\}}. \tag{5.36}$$

$$P(\hat{\theta}_p \leq t) = P(F_n(t) \geq p) = \sum_{i=l_p}^n \binom{n}{i}[F(t)]^i[1 - F(t)]^{n-i}, \tag{5.68}$$

where $l_p = np$ if np is an integer and $l_p = 1+$ the integer part of np if np is not an integer. If F has a Lebesgue p.d.f. f, then $\hat{\theta}_p$ has the Lebesgue p.d.f.

$$\varphi_n(t) = n\binom{n-1}{l_p - 1}[F(t)]^{l_p - 1}[1 - F(t)]^{n - l_p} f(t). \tag{5.69}$$

Theorem 5.10. Let $X_1, ..., X_n$ be i.i.d. random variables from F.
(i) If $F(\theta_p) = p$, then $P(\sqrt{n}(\hat{\theta}_p - \theta_p) \leq 0) \to \Phi(0) = \frac{1}{2}$, where Φ is the c.d.f. of the standard normal.
(ii) If F is continuous at θ_p and there exists $F'(\theta_p-) > 0$, then

$$P(\sqrt{n}(\hat{\theta}_p - \theta_p) \leq t) \to \Phi(t/\sigma_F^-), \qquad t < 0,$$

where $\sigma_F^- = \sqrt{p(1-p)}/F'(\theta_p-)$.
(iii) If F is continuous at θ_p and there exists $F'(\theta_p+) > 0$, then

$$P(\sqrt{n}(\hat{\theta}_p - \theta_p) \leq t) \to \Phi(t/\sigma_F^+), \qquad t > 0,$$

where $\sigma_F^+ = \sqrt{p(1-p)}/F'(\theta_p+)$.
(iv) If $F'(\theta_p)$ exists and is positive, then

$$\sqrt{n}(\hat{\theta}_p - \theta_p) \to_d N(0, \sigma_F^2), \tag{5.70}$$

where $\sigma_F = \sqrt{p(1-p)}/F'(\theta_p)$.

Theorem 5.11 (Bahadur's representation). Let $X_1, ..., X_n$ be i.i.d. random variables from F. Suppose that $F'(\theta_p)$ exists and is positive. Then

$$\hat{\theta}_p = \theta_p + \frac{F(\theta_p) - F_n(\theta_p)}{F'(\theta_p)} + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{5.71}$$

Corollary 5.1. Let $X_1, ..., X_n$ be i.i.d. random variables from F having positive derivatives at θ_{p_j}, where $0 < p_1 < \cdots < p_m < 1$ are fixed constants. Then

$$\sqrt{n}[(\hat{\theta}_{p_1}, ..., \hat{\theta}_{p_m}) - (\theta_{p_1}, ..., \theta_{p_m})] \to_d N_m(0, D),$$

where D is the $m \times m$ symmetric matrix whose (i, j)th element is

$$p_i(1 - p_j)/[F'(\theta_{p_i})F'(\theta_{p_j})], \qquad i \leq j. \ ∎$$

$$\sqrt{n}(\bar{X} - \theta) \to_d N(0, \mathrm{Var}(X_1)).$$

By Theorem 5.10(iv),

$$\sqrt{n}(\hat{\theta}_{0.5} - \theta) \to_d N(0, [2F'(\theta)]^{-2}).$$

Hence, the asymptotic relative efficiency of $\hat{\theta}_{0.5}$ w.r.t. $\bar{X}$ is

$$e(F) = 4[F'(\theta)]^2 \mathrm{Var}(X_1).$$

$$\sqrt{n}(\bar{X}_\alpha - \theta) \to_d N(0, \sigma_\alpha^2), \tag{5.78}$$

where

$$\sigma_\alpha^2 = \frac{2}{(1 - 2\alpha)^2}\left\{\int_0^{F_0^{-1}(1-\alpha)} x^2 dF_0(x) + \alpha[F_0^{-1}(1 - \alpha)]^2\right\}.$$

This leads to the following general estimation method. Let $\Theta \subset \mathcal{R}^k$ be the range of θ, ψ_i be a Borel function from $\mathcal{R}^{d_i} \times \Theta$ to $\mathcal{R}^k$, $i = 1, ..., n$, and

$$s_n(\gamma) = \sum_{i=1}^{n} \psi_i(X_i, \gamma), \qquad \gamma \in \Theta. \tag{5.84}$$

If θ is estimated by $\hat{\theta} \in \Theta$ satisfying $s_n(\hat{\theta}) = 0$, then $\hat{\theta}$ is called a GEE estimator. The equation $s_n(\gamma) = 0$ is called a GEE. Apparently, the LSE's, RLE's, MQLE's, and M-estimators are special cases of GEE estimators.

the empirical likelihood

$$\ell(G) = \prod_{i=1}^{n} P_G(\{x_i\}), \qquad G \in \mathfrak{F}$$

subject to

$$p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \text{and} \quad \sum_{i=1}^{n} p_i \psi(x_i, \theta) = 0, \tag{5.91}$$

where $p_i = P_G(\{x_i\})$. However, in this case the dimension of the function ψ is the same as the dimension of the parameter θ and, hence, the last equation in (5.91) does not impose any restriction on p_i's. Then, it follows

Proposition 5.2. Suppose that $X_1, ..., X_n$ are i.i.d. from F and $\psi_i \equiv \psi$, a bounded and continuous function from $\mathcal{R}^d \times \Theta$ to $\mathcal{R}^k$. Let $\Psi(t) = \int \psi(x, t) dF(x)$. Suppose that $\Psi(\theta) = 0$ and $\partial \Psi(t)/\partial t$ exists and is of full rank at $t = \theta$. Then $\hat{\theta}_n \to_p \theta$. ∎

Theorem 5.13. Let $X_1, ..., X_n$ be i.i.d. from F, $\psi_i \equiv \psi$, and $\theta \in \mathcal{R}$. Suppose that $\Psi(\gamma) = \int \psi(x, \gamma) dF(x) = 0$ if and only if $\gamma = \theta$, $\Psi'(\theta)$ exists and $\Psi'(\theta) \neq 0$.
(i) Assume that $\psi(x, \gamma)$ is nonincreasing in γ and that $\int [\psi(x, \gamma)]^2 dF(x)$ is finite for γ in a neighborhood of θ and is continuous at θ. Then, any sequence of GEE estimators (M-estimators) $\{\hat{\theta}_n\}$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma_F^2), \tag{5.98}$$

where

$$\sigma_F^2 = \int [\psi(x, \theta)]^2 dF(x) / [\Psi'(\theta)]^2.$$

(ii) Assume that $\int [\psi(x, \theta)]^2 dF(x) < \infty$, $\psi(x, \gamma)$ is continuous in x, and $\lim_{\gamma \to \theta} \|\psi(\cdot, \gamma) - \psi(\cdot, \theta)\|_V = 0$, where $\|\cdot\|_V$ is the variation norm defined in Lemma 5.2. Then, any consistent sequence of GEE estimators $\{\hat{\theta}_n\}$ satisfies (5.98).

Theorem 5.14. Suppose that $\varphi_i(x, \gamma) = \partial \psi_i(x, \gamma)/\partial \gamma$ exists and the sequence of functions $\{\varphi_{ij}, i = 1, 2, ...\}$ satisfies the conditions in Lemma 5.3 with Θ replaced by a compact neighborhood of θ, where φ_{ij} is the jth row of φ_i; $\sup_i E\|\psi_i(X_i, \theta)\|^{2+\delta} < \infty$ for some $\delta > 0$ (this condition can be replaced by $E\|\psi(X_1, \theta)\|^2 < \infty$ if X_i's are i.i.d. and $\psi_i \equiv \psi$); $E[\psi_i(X_i, \theta)] = 0$; $\liminf_n \lambda_-[n^{-1}\mathrm{Var}(s_n(\theta))] > 0$ and $\liminf_n \lambda_-[n^{-1} M_n(\theta)] > 0$, where $M_n(\theta) = -E[\nabla s_n(\theta)]$ and $\lambda_-[A]$ is the smallest eigenvalue of the matrix A. If $\{\hat{\theta}_n\}$ is a consistent sequence of GEE estimators, then

$$V_n^{-1/2}(\hat{\theta}_n - \theta) \to_d N_k(0, I_k), \tag{5.99}$$

where

$$V_n = [M_n(\theta)]^{-1} \mathrm{Var}(s_n(\theta)) [M_n(\theta)]^{-1}. \tag{5.100}$$

Theorem 6.1 (Neyman-Pearson lemma). Suppose that $\mathcal{P}_0 = \{P_0\}$ and $\mathcal{P}_1 = \{P_1\}$. Let f_j be the p.d.f. of P_j w.r.t. a σ-finite measure ν (e.g., $\nu = P_0 + P_1$), $j = 0, 1$.
(i) (Existence of a UMP test). For every α, there exists a UMP test of size α, which is equal to

$$T_*(X) = \begin{cases} 1 & f_1(X) > c f_0(X) \\ \gamma & f_1(X) = c f_0(X) \\ 0 & f_1(X) < c f_0(X), \end{cases} \tag{6.3}$$

where $\gamma \in [0, 1]$ and $c \geq 0$ are some constants chosen so that $E[T_*(X)] = \alpha$ when $P = P_0$ ($c = \infty$ is allowed).
(ii) (Uniqueness). If T_{**} is a UMP test of size α, then

$$T_{**}(X) = \begin{cases} 1 & f_1(X) > c f_0(X) \\ 0 & f_1(X) < c f_0(X) \end{cases} \quad \text{a.s. } \mathcal{P}. \tag{6.4}$$

Lemma 6.1. Suppose that there is a test T_* of size α such that for every $P_1 \in \mathcal{P}_1$, T_* is UMP for testing H_0 versus the hypothesis $P = P_1$. Then T_* is UMP for testing H_0 versus H_1.

Lemma 6.3. Suppose that the distribution of X is in a parametric family $\mathcal{P}$ indexed by a real-valued θ and that $\mathcal{P}$ has monotone likelihood ratio in $Y(X)$. If ψ is a nondecreasing function of Y, then $g(\theta) = E[\psi(Y)]$ is a nondecreasing function of θ.

Example 6.5. The following families have monotone likelihood ratio:
(a) the double exponential distribution family $\{DE(\theta, c)\}$ with a known c;
(b) the exponential distribution family $\{E(\theta, c)\}$ with a known c;
(c) the logistic distribution family $\{LG(\theta, c)\}$ with a known c;
(d) the uniform distribution family $\{U(\theta, \theta + 1)\}$;
(e) the hypergeometric distribution family $\{HG(r, \theta, N - \theta)\}$ with known r and N (Table 1.1, page 18).

An example of a family that does not have monotone likelihood ratio is the Cauchy distribution family $\{C(\theta, c)\}$ with a known c. ∎

Theorem 6.2. Suppose that X has a distribution in $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ ($\Theta \subset \mathcal{R}$) that has monotone likelihood ratio in $Y(X)$. Consider the problem of testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a given constant.
(i) There exists a UMP test of size α, which is given by

$$T_*(X) = \begin{cases} 1 & Y(X) > c \\ \gamma & Y(X) = c \\ 0 & Y(X) < c, \end{cases} \tag{6.11}$$

where c and γ are determined by $\beta_{T_*}(\theta_0) = \alpha$, and $\beta_T(\theta) = E[T(X)]$ is the power function of a test T.
(ii) $\beta_{T_*}(\theta)$ is strictly increasing for all θ's for which $0 < \beta_{T_*}(\theta) < 1$.
(iii) For any $\theta < \theta_0$, T_* minimizes $\beta_T(\theta)$ (the type I error probability of T) among all tests T satisfying $\beta_T(\theta_0) = \alpha$.
(iv) Assume that $P_\theta(f_\theta(X) = c f_{\theta_0}(X)) = 0$ for any $\theta > \theta_0$ and $c \geq 0$, where f_θ is the p.d.f. of P_θ. If T is a test with $\beta_T(\theta_0) = \beta_{T_*}(\theta_0)$, then for any $\theta > \theta_0$, either $\beta_T(\theta) < \beta_{T_*}(\theta)$ or $T = T_*$ a.s. P_θ.
(v) For any fixed θ_1, T_* is UMP for testing $H_0 : \theta \leq \theta_1$ versus $H_1 : \theta > \theta_1$, with size $\beta_{T_*}(\theta_1)$.

Corollary 6.1. Suppose that X has the p.d.f. given by (6.10) w.r.t. a σ-finite measure, where η is a strictly monotone function of θ. If η is increasing, then T_* given by (6.11) is UMP for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where γ and c are determined by $\beta_{T_*}(\theta_0) = \alpha$. If η is decreasing or $H_0 : \theta \geq \theta_0$ ($H_1 : \theta < \theta_0$), the result is still valid by reversing inequalities in (6.11). ∎

The following hypotheses are called two-sided hypotheses:

$$H_0: \theta \leq \theta_1 \text{ or } \theta \geq \theta_2 \quad \text{versus} \quad H_1: \theta_1 < \theta < \theta_2, \tag{6.12}$$

$$H_0: \theta_1 \leq \theta \leq \theta_2 \quad \text{versus} \quad H_1: \theta < \theta_1 \text{ or } \theta > \theta_2, \tag{6.13}$$

$$H_0: \theta = \theta_0 \quad \text{versus} \quad H_1: \theta \neq \theta_0, \tag{6.14}$$

where θ_0, θ_1, and θ_2 are given constants and $\theta_1 < \theta_2$.

Theorem 6.3. Suppose that X has the p.d.f. given by (6.10) w.r.t. a σ-finite measure, where η is a strictly increasing function of θ.
(i) For testing hypotheses (6.12), a UMP test of size α is

$$T_*(X) = \begin{cases} 1 & c_1 < Y(X) < c_2 \\ \gamma_i & Y(X) = c_i, \ i = 1, 2 \\ 0 & Y(X) < c_1 \text{ or } Y(X) > c_2, \end{cases} \tag{6.15}$$

where c_i's and γ_i's are determined by

$$\beta_{T_*}(\theta_1) = \beta_{T_*}(\theta_2) = \alpha. \tag{6.16}$$

(ii) The test defined by (6.15) minimizes $\beta_T(\theta)$ over all $\theta < \theta_1$, $\theta > \theta_2$, and T satisfying $\beta_T(\theta_1) = \beta_T(\theta_2) = \alpha$.
(iii) If T_* and T_{**} are two tests satisfying (6.15) and $\beta_{T_*}(\theta_1) = \beta_{T_{**}}(\theta_1)$ and if the region $\{T_{**} = 1\}$ is to the right of $\{T_* = 1\}$, then $\beta_{T_*}(\theta) < \beta_{T_{**}}(\theta)$ for $\theta > \theta_1$ and $\beta_{T_*}(\theta) > \beta_{T_{**}}(\theta)$ for $\theta < \theta_1$. If both T_* and T_{**} satisfy (6.15) and (6.16), then $T_* = T_{**}$ a.s. $\mathcal{P}$.

Lemma 6.4. Suppose that X has a p.d.f. in $\{f_\theta(x) : \theta \in \Theta\}$, a parametric family of p.d.f.'s w.r.t. a single σ-finite measure ν on $\mathcal{R}$, where $\Theta \subset \mathcal{R}$. Suppose that this family has monotone likelihood ratio in X. Let ψ be a function with a single change of sign.
(i) There exists $\theta_0 \in \Theta$ such that $E_\theta[\psi(X)] \leq 0$ for $\theta < \theta_0$ and $E_\theta[\psi(X)] \geq 0$

for $\theta > \theta_0$, where E_θ is the expectation w.r.t. f_θ.
(ii) Suppose that $f_\theta(x) > 0$ for all x and θ, that $f_{\theta_1}(x)/f_\theta(x)$ is strictly increasing in x for $\theta < \theta_1$, and that $\nu(\{x : \psi(x) \neq 0\}) > 0$. If $E_{\theta_0}[\psi(X)] = 0$, then $E_\theta[\psi(X)] < 0$ for $\theta < \theta_0$ and $E_\theta[\psi(X)] > 0$ for $\theta > \theta_0$.

A UMP test T of size α has the property that

$$\beta_T(P) \leq \alpha, \quad P \in \mathcal{P}_0 \quad \text{and} \quad \beta_T(P) \geq \alpha, \quad P \in \mathcal{P}_1. \tag{6.19}$$

Definition 6.3. Let α be a given level of significance. A test T for $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ is said to be unbiased of level α if and only if (6.19) holds. A test of size α is called a *uniformly most powerful unbiased* (UMPU) test if and only if it is UMP within the class of unbiased tests of level α. ∎

Lemma 6.5. Consider hypotheses (6.20). Suppose that, for every T, $\beta_T(P)$ is continuous in θ. If T_* is uniformly most powerful among all tests satisfying (6.21) and has size α, then T_* is a UMPU test.

Lemma 6.6. Let $U(X)$ be a sufficient statistic for $P \in \bar{\mathcal{P}}$. Then a necessary and sufficient condition for all tests similar on $\bar{\Theta}_{01}$ to have Neyman structure w.r.t. U is that U is boundedly complete for $P \in \bar{\mathcal{P}}$.

Throughout this section, we consider the following hypotheses:

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1, \tag{6.20}$$

Definition 6.4. Consider the hypotheses specified by (6.20). Let α be a given level of significance and let $\bar{\Theta}_{01}$ be the common boundary of Θ_0 and Θ_1, i.e., the set of points θ that are points or limit points of both Θ_0 and Θ_1. A test T is *similar* on $\bar{\Theta}_{01}$ if and only if

$$\beta_T(P) = \alpha, \qquad \theta \in \bar{\Theta}_{01}. \ ∎ \tag{6.21}$$

$$f_{\theta,\varphi}(x) = \exp\left\{\theta Y(x) + \varphi^\tau U(x) - \zeta(\theta, \varphi)\right\}, \tag{6.23}$$

Theorem 6.4. Suppose that the distribution of X is in a multiparameter natural exponential family given by (6.23).
(i) For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, a UMPU test of size α is

$$T_*(Y, U) = \begin{cases} 1 & Y > c(U) \\ \gamma(U) & Y = c(U) \\ 0 & Y < c(U), \end{cases} \tag{6.24}$$

where $c(u)$ and $\gamma(u)$ are Borel functions determined by

$$E_{\theta_0}[T_*(Y, U) | U = u] = \alpha \tag{6.25}$$

for every u, and E_{θ_0} is the expectation w.r.t. $f_{\theta_0,\varphi}$.
(ii) For testing hypotheses (6.12), a UMPU test of size α is

$$T_*(Y, U) = \begin{cases} 1 & c_1(U) < Y < c_2(U) \\ \gamma_i(U) & Y = c_i(U), \ i = 1, 2, \\ 0 & Y < c_1(U) \text{ or } Y > c_2(U), \end{cases} \tag{6.26}$$

where $c_i(u)$'s and $\gamma_i(u)$'s are Borel functions determined by

$$E_{\theta_1}[T_*(Y, U) | U = u] = E_{\theta_2}[T_*(Y, U) | U = u] = \alpha \tag{6.27}$$

for every u.
(iii) For testing hypotheses (6.13), a UMPU test of size α is

$$T_*(Y, U) = \begin{cases} 1 & Y < c_1(U) \text{ or } Y > c_2(U) \\ \gamma_i(U) & Y = c_i(U), \ i = 1, 2, \\ 0 & c_1(U) < Y < c_2(U), \end{cases} \tag{6.28}$$

where $c_i(u)$'s and $\gamma_i(u)$'s are Borel functions determined by (6.27) for every u.
(iv) For testing hypotheses (6.14), a UMPU test of size α is given by (6.28), where $c_i(u)$'s and $\gamma_i(u)$'s are Borel functions determined by (6.25) and

$$E_{\theta_0}[T_*(Y, U) Y | U = u] = \alpha E_{\theta_0}(Y | U = u) \tag{6.29}$$

for every u.

Lemma 6.7. Suppose that X has the p.d.f. (6.23) and that $V(Y, U)$ is a statistic independent of U when $\theta = \theta_j$, where θ_j's are known values given in the hypotheses in (i)-(iv) of Theorem 6.4.
(i) If $V(y, u)$ is increasing in y for each u, then the UMPU tests in (i)-(iii) of Theorem 6.4 are equivalent to those given by (6.24)-(6.28) with Y and (Y, U) replaced by V and with $c_i(U)$ and $\gamma_i(U)$ replaced by constants c_i and γ_i, respectively.
(ii) If there are Borel functions $a(u) > 0$ and $b(u)$ such that $V(y, u) = a(u) y + b(u)$, then the UMPU test in Theorem 6.4(iv) is equivalent to that given by (6.25), (6.28), and (6.29) with Y and (Y, U) replaced by V and with $c_i(U)$ and $\gamma_i(U)$ replaced by constants c_i and γ_i, respectively.

Definition 6.6. Let $\ell(\theta) = f_\theta(X)$ be the likelihood function. For testing (6.59), a *likelihood ratio* (LR) test is any test that rejects H_0 if and only if $\lambda(X) < c$, where $c \in [0, 1]$ and $\lambda(X)$ is the likelihood ratio defined by

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} \ell(\theta)}{\sup_{\theta \in \Theta} \ell(\theta)}. \ ∎$$

Proposition 6.5. Suppose that X has the p.d.f. given by (6.10) w.r.t. a σ-finite measure ν, where η is a strictly increasing and differentaible function of θ.
(i) For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, there is an LR test whose rejection region is the same as that of the UMP test T_* given by (6.11).
(ii) For testing the hypotheses in (6.12), there is an LR test whose rejection region is the same as that of the UMP test T_* given by (6.15).
(iii) For testing the hypotheses in (6.13) or (6.14), there is an LR test whose rejection region is equivalent to $Y(X) < c_1$ or $Y(X) > c_2$ for some constants c_1 and c_2.

Theorem 6.5. Assume the conditions in Theorem 4.16. Suppose that H_0 is determined by (6.63). Under H_0, $-2 \log \lambda_n \to_d \chi_r^2$, where $\lambda_n = \lambda(X)$ and χ_r^2 is a random variable having the chi-square distribution χ_r^2. Consequently, the LR test with rejection region $\lambda_n < e^{-\chi_{r,\alpha}^2/2}$ has asymptotic significance level α, where $\chi_{r,\alpha}^2$ is the $(1 - \alpha)$th quantile of the chi-square distribution χ_r^2.

Theorem 6.6. Assume the conditions in Theorem 4.16.
(i) Under H_0 given by (6.64), $W_n \to_d \chi_r^2$ and, therefore, the test rejects H_0 if and only if $W_n > \chi_{r,\alpha}^2$ has asymptotic significance level α, where $\chi_{r,\alpha}^2$ is the $(1 - \alpha)$th quantile of the chi-square distribution χ_r^2.
(ii) The result in (i) still holds if W_n is replaced by R_n.

There are two popular asymptotic tests based on likelihoods that are asymptotically equivalent to LR tests. Note that the hypothesis in (6.63) is equivalent to a set of $r \leq k$ equations:

$$H_0 : R(\theta) = 0, \tag{6.64}$$

where $R(\theta)$ is a continuously differentiable function from $\mathcal{R}^k$ to $\mathcal{R}^r$. Wald (1943) introduced a test that rejects H_0 when the value of

$$W_n = [R(\hat{\theta})]^\tau \{[C(\hat{\theta})]^\tau [I_n(\hat{\theta})]^{-1} C(\hat{\theta})\}^{-1} R(\hat{\theta})$$

is large, where $C(\theta) = \partial R(\theta)/\partial \theta$, $I_n(\theta)$ is the Fisher information matrix based on $X_1, ..., X_n$, and $\hat{\theta}$ is an MLE or RLE of θ. For testing $H_0 : \theta = \theta_0$

with a known θ_0, $R(\theta) = \theta - \theta_0$ and W_n simplifies to

$$W_n = (\hat{\theta} - \theta_0)^\tau I_n(\hat{\theta})(\hat{\theta} - \theta_0).$$

Rao (1947) introduced a *score* test that rejects H_0 when the value of

$$R_n = [s_n(\tilde{\theta})]^\tau [I_n(\tilde{\theta})]^{-1} s_n(\tilde{\theta})$$

is large, where $s_n(\theta) = \partial \log \ell(\theta)/\partial \theta$ is the score function and $\tilde{\theta}$ is an MLE or RLE of θ under H_0 in (6.64).

Theorem 6.7. Assume the conditions in Theorem 4.18. Consider the problem of testing H_0 in (6.64) (or equivalently, (6.63)) with $\theta = (\beta, \phi)$. Then the results in Theorems 6.5 and 6.6 still hold. ∎

Let X be a sample of size n from a population P and $\hat{\theta}_n$ be an estimator of θ, a k-vector of parameters related to P. Suppose that under H_0,

$$V_n^{-1/2}(\hat{\theta}_n - \theta) \to_d N_k(0, I_k), \tag{6.92}$$

where V_n is the asymptotic covariance matrix of $\hat{\theta}_n$. If V_n is known when $\theta = \theta_0$, then a test with rejection region

$$(\hat{\theta}_n - \theta_0)^\tau V_n^{-1}(\hat{\theta}_n - \theta_0) > \chi^2_{k,\alpha} \tag{6.93}$$

has asymptotic significance level α, where $\chi^2_{k,\alpha}$ is the $(1-\alpha)$th quantile of the chi-squared distribution χ^2_k. If the distribution of $\hat{\theta}_n$ does not depend

Theorem 6.12. Assume that (6.92) holds for any P and that $\lambda_+[V_n] \to 0$, where $\lambda_+[V_n]$ is the largest eigenvalue of V_n.
(i) The test having rejection region (6.93) (with a known V_n or V_n replaced by an estimator $\hat{V}_n$ that is consistent for any P) is consistent.
(ii) If we choose $\alpha = \alpha_n \to 0$ as $n \to \infty$ and $\chi^2_{k,1-\alpha_n}\lambda_+[V_n] = o(1)$, then the test in (i) is Chernoff-consistent.

Theorem 7.1. Suppose that P is in a parametric family indexed by a real-valued θ. Let $T(X)$ be a real-valued statistic with c.d.f. $F_{T,\theta}(t)$ and let α_1 and α_2 be fixed positive constants such that $\alpha_1 + \alpha_2 = \alpha < \frac{1}{2}$.
(i) Suppose that $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nonincreasing in θ for each fixed t. Define

$$\overline{\theta} = \sup\{\theta : F_{T,\theta}(T) \geq \alpha_1\} \quad \text{and} \quad \underline{\theta} = \inf\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}.$$

Then $[\underline{\theta}(T), \overline{\theta}(T)]$ is a level $1-\alpha$ confidence interval for θ.
(ii) If $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nondecreasing in θ for each t, then the same result holds with

$$\underline{\theta} = \inf\{\theta : F_{T,\theta}(T) \geq \alpha_1\} \quad \text{and} \quad \overline{\theta} = \sup\{\theta : F_{T,\theta}(T-) \leq 1 - \alpha_2\}.$$

(iii) If $F_{T,\theta}$ is a continuous c.d.f. for any θ, then $F_{T,\theta}(T)$ is a pivotal quantity and the confidence interval in (i) or (ii) has confidence coefficient $1-\alpha$.

Theorem 7.2. For each $\theta_0 \in \Theta$, let T_{θ_0} be a test for $H_0 : \theta = \theta_0$ (versus some H_1) with significance level α and acceptance region $A(\theta_0)$. For each x in the range of X, define

$$C(x) = \{\theta : x \in A(\theta)\}.$$

Then $C(X)$ is a level $1-\alpha$ confidence set for θ. If T_{θ_0} is nonrandomized and has size α for every θ_0, then $C(X)$ has confidence coefficient $1-\alpha$.

Proposition 7.2. Let $C(X)$ be a confidence set for θ with significance level (or confidence coefficient) $1-\alpha$. For any $\theta_0 \in \Theta$, define a region $A(\theta_0) = \{x : \theta_0 \in C(x)\}$. Then the test $T(X) = 1 - I_{A(\theta_0)}(X)$ has significance level α for testing $H_0 : \theta = \theta_0$ versus some H_1. ∎

Theorem 7.3. Let θ be a real-valued parameter and $T(X)$ be a real-valued statistic.
(i) Let $U(X)$ be a positive statistic. Suppose that $(T-\theta)/U$ is a pivotal quantity having a Lebesgue p.d.f. f that is *unimodal* at $x_0 \in \mathcal{R}$ in the sense that $f(x)$ is nondecreasing for $x \leq x_0$ and $f(x)$ is nonincreasing for $x \geq x_0$. Consider the following class of confidence intervals for θ:

$$\mathcal{C} = \left\{[T - bU, T - aU] : \ a \in \mathcal{R}, b \in \mathcal{R}, \int_a^b f(x)dx = 1 - \alpha\right\}. \tag{7.10}$$

If $[T - b_* U, T - a_* U] \in \mathcal{C}$, $f(a_*) = f(b_*) > 0$, and $a_* \leq x_0 \leq b_*$, then the interval $[T - b_* U, T - a_* U]$ has the shortest length within $\mathcal{C}$.

(ii) Suppose that $T > 0$, $\theta > 0$, T/θ is a pivotal quantity having a Lebesgue p.d.f. f, and that $x^2 f(x)$ is unimodal at x_0. Consider the following class of confidence intervals for θ:

$$\mathcal{C} = \left\{[b^{-1}T, a^{-1}T] : \ a > 0, b > 0, \int_a^b f(x)dx = 1 - \alpha\right\}. \tag{7.11}$$

If $[b_*^{-1}T, a_*^{-1}T] \in \mathcal{C}$, $a_*^2 f(a_*) = b_*^2 f(b_*) > 0$, and $a_* \leq x_0 \leq b_*$, then the interval $[b_*^{-1}T, a_*^{-1}T]$ has the shortest length within $\mathcal{C}$.

Definition 7.2. Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true parameter value θ. A confidence set $C(X)$ for θ with confidence coefficient $1-\alpha$ is said to be Θ'-*uniformly most accurate* (UMA) if and only if for any other confidence set $C_1(X)$ with significance level $1-\alpha$,

$$P\big(\theta' \in C(X)\big) \leq P\big(\theta' \in C_1(X)\big) \qquad \text{for all } \theta' \in \Theta'. \tag{7.15}$$

$C(X)$ is UMA if and only if it is Θ'-UMA with $\Theta' = \{\theta\}^c$. ∎

Theorem 7.4. Let $C(X)$ be a confidence set for θ obtained by inverting the acceptance regions of nonrandomized tests T_{θ_0} for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \Theta_{\theta_0}$. Suppose that for each θ_0, T_{θ_0} is UMP of size α. Then $C(X)$ is Θ'-UMA with confidence coefficient $1-\alpha$, where $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$.

Definition 7.3. Let $\theta \in \Theta$ be an unknown parameter, Θ' be a subset of Θ that does not contain the true parameter value θ, and $1-\alpha$ be a given significance level.

(i) A level $1-\alpha$ confidence set $C(X)$ is said to be Θ'-*unbiased* (unbiased when $\Theta' = \{\theta\}^c$) if and only if $P(\theta' \in C(X)) \leq 1-\alpha$ for all $\theta' \in \Theta'$.
(ii) Let $C(X)$ be a Θ'-unbiased confidence set with confidence coefficient $1-\alpha$. If (7.15) holds for any other Θ'-unbiased confidence set $C_1(X)$ with significance level $1-\alpha$, then $C(X)$ is Θ'-*uniformly most accurate unbiased* (UMAU). $C(X)$ is UMAU if and only if it is Θ'-UMAU with $\Theta' = \{\theta\}^c$. ∎

Theorem 7.5. Let $C(X)$ be a confidence set for θ obtained by inverting the acceptance regions of nonrandomized tests T_{θ_0} for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \Theta_{\theta_0}$. If T_{θ_0} is unbiased of size α for each θ_0, then $C(X)$ is Θ'-unbiased with confidence coefficient $1-\alpha$, where $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$; if T_{θ_0} is also UMPU for each θ_0, then $C(X)$ is Θ'-UMAU. ∎

Theorem 7.6 (Pratt's theorem). Let X be a sample from P and $C(X)$ be a confidence set for $\theta \in \mathcal{R}^k$. Suppose that $\mathrm{vol}(C(x)) = \int_{C(x)} d\theta'$ is finite a.s. P. Then the expected volume of $C(X)$ is

$$E[\mathrm{vol}(C(X))] = \int_{\theta \neq \theta'} P(\theta' \in C(X)) d\theta'. \tag{7.16}$$

Proposition 7.4. Let $C_j(X)$, $j = 1, 2$, be the confidence sets given in (7.19) with $\hat{\theta}_n = \hat{\theta}_{jn}$ and $\hat{V}_n = \hat{V}_{jn}$, $j = 1, 2$, respectively. Suppose that for each j, (7.18) holds for $\hat{\theta}_{jn}$ and $\hat{V}_{jn}$ is consistent for V_{jn}, the asymptotic covariance matrix of $\hat{\theta}_{jn}$. If $\mathrm{Det}(V_{1n}) < \mathrm{Det}(V_{2n})$ for sufficiently large n, where $\mathrm{Det}(A)$ is the determinant of A, then

$$P\big(\mathrm{vol}(C_1(X)) < \mathrm{vol}(C_2(X))\big) \to 1.$$

the likelihood function based on the observation $X = x$. The acceptance region of the LR test defined in §6.4.1 with $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$ is

$$A(\vartheta_0) = \{x : \ell(\vartheta_0, \hat{\varphi}_{\vartheta_0}) \geq e^{-c_\alpha/2}\ell(\hat{\theta})\},$$

where $\ell(\hat{\theta}) = \sup_{\theta \in \Theta} \ell(\theta)$, $\ell(\vartheta, \hat{\varphi}_\vartheta) = \sup_\varphi \ell(\vartheta, \varphi)$, and c_α is a constant related to the significance level α. Under the conditions of Theorem 6.5, if c_α is chosen to be $\chi^2_{r,\alpha}$, the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r, then

$$C(X) = \{\vartheta : \ell(\vartheta, \hat{\varphi}_\vartheta) \geq e^{-c_\alpha/2}\ell(\hat{\theta})\} \tag{7.20}$$

In §6.4.2 we discussed two asymptotic tests closely related to the LR test: Wald's test and Rao's score test. When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Wald's

test has acceptance region

$$A(\vartheta_0) = \{x : (\hat{\vartheta} - \vartheta_0)^\tau \{C^\tau [I_n(\hat{\theta})]^{-1} C\}^{-1} (\hat{\vartheta} - \vartheta_0) \leq \chi^2_{r,\alpha}\}, \tag{7.21}$$

where $\hat{\theta} = (\hat{\vartheta}, \hat{\varphi})$ is an MLE or RLE of $\theta = (\vartheta, \varphi)$, $I_n(\theta)$ is the Fisher information matrix based on X, $C^\tau = (\ I_r \ \ 0\ )$, and 0 is an $r \times (k-r)$ matrix of 0's. By Theorem 4.17 or 4.18, the confidence set obtained by inverting $A(\vartheta)$ in (7.21) is the same as that in (7.19) with $\theta = \vartheta$ and $\hat{V}_n = C^\tau [I_n(\hat{\theta})]^{-1} C$.

When $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$, Rao's score test has acceptance region

$$A(\vartheta_0) = \{x : [s_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^\tau [I_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^{-1} s_n(\vartheta_0, \hat{\varphi}_{\vartheta_0}) \leq \chi^2_{r,\alpha}\}, \tag{7.22}$$

where $s_n(\theta) = \partial \log \ell(\theta)/\partial\theta$ and $\hat{\varphi}_\vartheta$ is defined in (7.20). The confidence set obtained by inverting $A(\vartheta)$ in (7.22) is also $1-\alpha$ asymptotically correct.

Definition 7.6. Let X be a sample from $P \in \mathcal{P}$, let θ_t, $t \in \mathcal{T}$, be real-valued parameters related to P, and let $C_t(X)$, $t \in \mathcal{T}$, be a class of (one-sided or two-sided) confidence intervals.
(i) Intervals $C_t(X)$, $t \in \mathcal{T}$, are level $1-\alpha$ *simultaneous confidence intervals* for θ_t, $t \in \mathcal{T}$, if and only if

$$\inf_{P \in \mathcal{P}} P\big(\theta_t \in C_t(X) \text{ for all } t \in \mathcal{T}\big) \geq 1 - \alpha. \tag{7.57}$$

The left-hand side of (7.57) is the confidence coefficient of $C_t(X)$, $t \in \mathcal{T}$.
(ii) Intervals $C_t(X)$, $t \in \mathcal{T}$, are simultaneous confidence intervals for θ_t, $t \in \mathcal{T}$, with asymptotic significance level $1-\alpha$ if and only if

$$\lim_{n\to\infty} P\big(\theta_t \in C_t(X) \text{ for all } t \in \mathcal{T}\big) \geq 1 - \alpha. \tag{7.58}$$

Intervals $C_t(X)$, $t \in \mathcal{T}$, are $1-\alpha$ asymptotically correct if and only if the equality in (7.58) holds. ∎

Uniform	p.d.f.	$1/m,\ x = a_1, ..., a_m$
	m.g.f.	$\sum_{j=1}^m e^{a_j t}/m,\ \ t \in \mathcal{R}$
$DU(a_1, ..., a_m)$	Expectation	$\sum_{j=1}^m a_j/m$
	Variance	$\sum_{j=1}^m (a_j - \bar{a})^2/m,\ \bar{a} = \sum_{j=1}^m a_j/m$
	Parameter	$a_i \in \mathcal{R},\ m = 1, 2, ...$
Binomial	p.d.f.	$\binom{n}{x} p^x (1-p)^{n-x},\ \ x = 0, 1, ..., n$
	m.g.f.	$(pe^t + 1 - p)^n,\ \ t \in \mathcal{R}$
$Bi(p, n)$	Expectation	np
	Variance	$np(1-p)$
	Parameter	$p \in [0, 1],\ \ n = 1, 2, ...$
Poisson	p.d.f.	$\theta^x e^{-\theta}/x!,\ \ x = 0, 1, 2, ...$
	m.g.f.	$e^{\theta(e^t - 1)},\ \ t \in \mathcal{R}$
$P(\theta)$	Expectation	θ
	Variance	θ
	Parameter	$\theta > 0$
Geometric	p.d.f.	$(1-p)^{x-1}p,\ \ x = 1, 2, ...$
	m.g.f.	$pe^t/[1 - (1-p)e^t],\ \ t < -\log(1-p)$
$G(p)$	Expectation	$1/p$
	Variance	$(1-p)/p^2$
	Parameter	$p \in [0, 1]$
Hyper-geometric	p.d.f.	$\binom{n}{x}\binom{m}{r-x} / \binom{N}{r}$, $x = 0, 1, ..., \min\{r, n\},\ \ r - x \leq m$
	m.g.f.	No explicit form
$HG(r, n, m)$	Expectation	rn/N
	Variance	$rnm(N-r)/[N^2(N-1)]$
	Parameter	$r, n, m = 1, 2, ...,\ N = n + m$
Negative binomial	p.d.f.	$\binom{x-1}{r-1} p^r (1-p)^{x-r},\ \ x = r, r+1, ...$
	m.g.f.	$p^r e^{rt}/[1 - (1-p)e^t]^r,\ \ t < -\log(1-p)$
	Expectation	r/p
$NB(p, r)$	Variance	$r(1-p)/p^2$
	Parameter	$p \in [0, 1],\ \ r = 1, 2, ...$
Log-distribution	p.d.f.	$-(\log p)^{-1} x^{-1} (1-p)^x,\ \ x = 1, 2, ...$
	m.g.f.	$\log[1 - (1-p)e^t]/\log p,\ \ t \in \mathcal{R}$
	Expectation	$-(1-p)/(p \log p)$
$L(p)$	Variance	$-(1-p)[1 + (1-p)/\log p]/(p^2 \log p)$
	Parameter	$p \in (0, 1)$

Uniform	p.d.f.	$(b-a)^{-1} I_{(a,b)}(x)$
	m.g.f.	$(e^{bt} - e^{at})/[(b-a)t],\ \ t \in \mathcal{R}$
$U(a, b)$	Expectation	$(a+b)/2$
	Variance	$(b-a)^2/12$
	Parameter	$a,\ b \in \mathcal{R},\ \ a < b$
Normal	p.d.f.	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
	m.g.f.	$e^{\mu t + \sigma^2 t^2/2},\ \ t \in \mathcal{R}$
$N(\mu, \sigma^2)$	Expectation	μ
	Variance	σ^2
	Parameter	$\mu \in \mathcal{R},\ \ \sigma > 0$
Exponential	p.d.f.	$\theta^{-1} e^{-(x-a)/\theta} I_{(a,\infty)}(x)$
	m.g.f.	$e^{at}(1 - \theta t)^{-1},\ \ t < \theta^{-1}$
$E(a, \theta)$	Expectation	$\theta + a$
	Variance	θ^2
	Parameter	$\theta > 0,\ \ a \in \mathcal{R}$
Chi-square	p.d.f.	$\frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2} I_{(0,\infty)}(x)$
	m.g.f.	$(1-2t)^{-k/2},\ \ t < 1/2$
χ^2_k	Expectation	k
	Variance	$2k$
	Parameter	$k = 1, 2, ...$
Gamma	p.d.f.	$\frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x)$
	m.g.f.	$(1 - \gamma t)^{-\alpha},\ \ t < \gamma^{-1}$
$\Gamma(\alpha, \gamma)$	Expectation	$\alpha\gamma$
	Variance	$\alpha\gamma^2$
	Parameter	$\gamma > 0,\ \alpha > 0$
Beta	p.d.f.	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$
	m.g.f.	No explicit form
$B(\alpha, \beta)$	Expectation	$\alpha/(\alpha + \beta)$
	Variance	$\alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$
	Parameter	$\alpha > 0,\ \ \beta > 0$
Cauchy	p.d.f.	$\frac{1}{\pi\sigma}\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1}$
	ch.f.	$e^{\sqrt{-1}\mu t - \sigma\lvert t\rvert}$
$C(\mu, \sigma)$	Expectation	Does not exist
	Variance	Does not exist
	Parameter	$\mu \in \mathcal{R},\ \ \sigma > 0$

t-distribution	p.d.f.	$\frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$
	ch.f.	No explicit form
t_n	Expectation	$0,\ \ (n > 1)$
	Variance	$n/(n-2),\ \ (n > 2)$
	Parameter	$n = 1, 2, ...$
F-distribution	p.d.f.	$\frac{n^{n/2} m^{m/2} \Gamma[(n+m)/2] x^{n/2-1}}{\Gamma(n/2)\Gamma(m/2)(m+nx)^{(n+m)/2}} I_{(0,\infty)}(x)$
	ch.f.	No explicit form
$F_{n,m}$	Expectation	$m/(m-2),\ \ (m > 2)$
	Variance	$2m^2(n+m-2)/[n(m-2)^2(m-4)]$, $(m > 4)$
	Parameter	$n = 1, 2, ...,\ \ m = 1, 2, ...$
Log-normal	p.d.f.	$\frac{1}{\sqrt{2\pi}\sigma} x^{-1} e^{-(\log x - \mu)^2/2\sigma^2} I_{(0,\infty)}(x)$
	ch.f.	No explicit form
$LN(\mu, \sigma^2)$	Expectation	$e^{\mu + \sigma^2/2}$
	Variance	$e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$
	Parameter	$\mu \in \mathcal{R},\ \ \sigma > 0$
Weibull	p.d.f.	$\frac{\alpha}{\theta} x^{\alpha-1} e^{-x^\alpha/\theta} I_{(0,\infty)}(x)$
	ch.f.	No explicit form
$W(\alpha, \theta)$	Expectation	$\theta^{1/\alpha}\Gamma(\alpha^{-1} + 1)$
	Variance	$\theta^{2/\alpha}\left\{\Gamma(2\alpha^{-1} + 1) - \left[\Gamma(\alpha^{-1} + 1)\right]^2\right\}$
	Parameter	$\theta > 0,\ \ \alpha > 0$
Double Exponential	p.d.f.	$\frac{1}{2\theta} e^{-\lvert x-\mu\rvert/\theta}$
	m.g.f.	$e^{\mu t}/(1 - \theta^2 t^2),\ \ \lvert t\rvert < \theta^{-1}$
	Expectation	μ
$DE(\mu, \theta)$	Variance	$2\theta^2$
	Parameter	$\mu \in \mathcal{R},\ \ \theta > 0$
Pareto	p.d.f.	$\theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x)$
	ch.f.	No explicit form
$Pa(a, \theta)$	Expectation	$\theta a/(\theta - 1),\ \ (\theta > 1)$
	Variance	$\theta a^2/[(\theta-1)^2(\theta-2)],\ \ (\theta > 2)$
	Parameter	$\theta > 0,\ \ a > 0$
Logistic	p.d.f.	$\sigma^{-1} e^{-(x-\mu)/\sigma}/[1 + e^{-(x-\mu)/\sigma}]^2$
	m.g.f.	$e^{\mu t}\Gamma(1 + \sigma t)\Gamma(1 - \sigma t),\ \lvert t\rvert < \sigma^{-1}$
$LG(\mu, \sigma)$	Expectation	μ
	Variance	$\sigma^2\pi^2/3$
	Parameter	$\mu \in \mathcal{R},\ \sigma > 0$

Inverse gamma	p.d.f.	$\frac{\gamma^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\gamma/y} I_{(0,\infty)}(x)$
$\Gamma^{-1}(\alpha, \gamma)$	Expectation	$\frac{\gamma}{\alpha-1}$
	Variance	$\frac{\gamma^2}{(\alpha-1)(\alpha-2)}$
	Parameter	$\gamma > 0,\ \alpha > 0$

Conjugate Priors:

Data $(X_1, ..., X_n) \sim ?$	Prior $\theta \sim ?$	Posterior $\theta \mid X \sim ?$
$N_k(\theta, I_k)$	$N_k(\mu_0, \Sigma_0)$	$N_k(A^{-1}(\Sigma_0^{-1}\mu_0 + T), A^{-1})$ where $T = \sum_{i=1}^n X_i,\ A = nI_k + \Sigma_0^{-1}$
$\text{Binom}(k, \theta)$	$\text{Beta}(\alpha, \beta)$	$\text{Beta}(T + \alpha, nk - T + \beta)$ where $T = \sum_{i=1}^n X_i$
$U(0, \theta)$	$\text{Pa}(a, b)$	$\text{Pa}(\max(a, X_{(n)}), b + n)$
$E(0, \theta)$	$\Gamma^{-1}(\alpha, \gamma^{-1})$	$\Gamma^{-1}(\alpha + n, (T + \gamma^{-1})^{-1})$

Proposition 1.8. Let X be a random k-vector with a Lebesgue p.d.f. f_X and let $Y = g(X)$, where g is a Borel function from $(\mathcal{R}^k, \mathcal{B}^k)$ to $(\mathcal{R}^k, \mathcal{B}^k)$. Let $A_1, ..., A_m$ be disjoint sets in $\mathcal{B}^k$ such that $\mathcal{R}^k - (A_1 \cup \cdots \cup A_m)$ has Lebesgue measure 0 and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $\text{Det}(\partial g(x)/\partial x) \neq 0$ on A_j, $j = 1, ..., m$. Then Y has the following Lebesgue p.d.f.:

$$f_Y(x) = \sum_{j=1}^m \left|\text{Det}\left(\partial h_j(x)/\partial x\right)\right| f_X\left(h_j(x)\right),$$

where h_j is the inverse function of g on A_j, $j = 1, ..., m$. ∎