Summary of results for STK4011/9011

Distributions

Beta Distribution

The beta family of distributions is a continuous family on $(0,1)$ indexed by two parameters. The *beta*(α, β) *pdf* is

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

where $B(\alpha, \beta)$ denotes the beta function,

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

The beta function is related to the gamma function through the following identity:

(3.3.17) $$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Cauchy Distribution

The *Cauchy distribution* is a symmetric, bell-shaped distribution on $(-\infty, \infty)$ with pdf

(3.3.19) $$f(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

Double Exponential Distribution

The *double exponential distribution* is formed by reflecting the exponential distribution around its mean. The pdf is given by

(3.3.22) $$f(x|\mu,\sigma) = \frac{1}{2\sigma}e^{-|x-\mu|/\sigma}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0.$$

Transformations

Theorem 2.1.5 *Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Let $\mathcal{X}$ and $\mathcal{Y}$ be defined by (2.1.7). Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by*

(2.1.10) $$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| & y \in \mathcal{Y} \\ 0 & \textit{otherwise.} \end{cases}$$

Bivariate transformations

Let (X, Y) be a bivariate random vector with joint pdf $f_{X,Y}(x,y)$ and support $\mathcal{A} = (x,y) : f_{X,Y}(x,y) > 0$. Let (U, V) be given by $U = g_1(X,Y)$ and $V = g_2(X,Y)$. Then the joint pdf $f_{U,V}(u,v)$ of (U, V) has support

$$\mathcal{B} = (u,v) : u = g_1(x,y) \text{ and } v = g_2(x,y) \text{ for some } (x,y) \in \mathcal{A}$$

We assume that $u = g_1(x,y)$ and $v = g_2(x,y)$ is a one-to-one transformation of $\mathcal{A}$ onto $\mathcal{B}$ and let $x = h_1(u,v)$ and $y = h_2(u,v)$ be the inverse transformation.

Let

$$J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}$$

be the Jacobian of the transformation. Then for $(u,v) \in \mathcal{B}$ we have

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))\,|J(u,v)|$$

Theorem 5.2.9 *If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw. \tag{5.2.3}$$

<u>Order statistics</u>

Theorem 5.4.4 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j}. \tag{5.4.4}$$

Theorem 5.4.6 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sampl $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then th joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is*

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} \times [F_X(v) - F_X(u)]^{j-1-i}[1-F_X(v)]^{n-j} \tag{5.4.7}$$

for $-\infty < u < v < \infty$.

Conditional expectation and variance

Theorem 4.4.3 *If X and Y are any two random variables, then*

(4.4.1) $$\mathrm{E}X = \mathrm{E}\left(\mathrm{E}(X|Y)\right),$$

provided that the expectations exist.

Theorem 4.4.7 (Conditional variance identity) *For any two random variables X and Y,*

(4.4.4) $$\mathrm{Var}\,X = \mathrm{E}\left(\mathrm{Var}(X|Y)\right) + \mathrm{Var}\left(\mathrm{E}(X|Y)\right),$$

provided that the expectations exist.

Inequalities

Theorem 3.6.1 (Chebychev's Inequality) *Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,*

$$P\left(g(X) \geq r\right) \leq \frac{\mathrm{E}g(X)}{r}.$$

Theorem 4.7.7 (Jensen's Inequality) *For any random variable X, if $g(x)$ is a convex function, then*

$$\mathrm{E}g(X) \geq g(\mathrm{E}X).$$

Equality holds if and only if, for every line $a + bx$ that is tangent to $g(x)$ at $x = EX$, $P(g(X) = a + bX) = 1$.

Convergence of random variables

Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\ X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.

Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

Theorem 2.3.12 (Convergence of mgfs) *Suppose $\{X_i,\ i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t), \qquad \text{for all } t \text{ in a neighborhood of } 0,$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i \to \infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t| < h$, of mgfs to an mgf implies convergence of cdfs.

Theorem 5.5.13 *The sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a constant μ if and only if the sequence also converges in distribution to μ. That is, the statement*

$$P(|X_n - \mu| > \varepsilon) \to 0 \text{ for every } \varepsilon > 0$$

is equivalent to

$$P(X_n \le x) \to \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu. \end{cases}$$

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\ X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n \to \infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. $Y_n X_n \to aX$ *in distribution.*

b. $X_n + Y_n \to X + a$ *in distribution.*

Theorem 5.5.24 (Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to \mathrm{n}(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

$$\sqrt{n}[g(Y_n) - g(\theta)] \to \mathrm{n}(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.} \tag{5.5.10}$$

Theorem 5.5.26 (Second-order Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to \mathrm{n}(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta) = 0$ and $g''(\theta)$ exists and is not 0. Then*

$$n[g(Y_n) - g(\theta)] \to \sigma^2 \frac{g''(\theta)}{2} \chi_1^2 \text{ in distribution.} \tag{5.5.13}$$

Sufficiency and completeness

Definition 6.2.1 A statistic $T(\mathbf{X})$ is a *sufficient statistic for θ* if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

Theorem 6.2.2 *If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf of $\mathbf{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.*

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \tag{6.2.3}$$

Theorem 6.2.10 *Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ belongs to an exponential family given by*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

Definition 6.2.11 A sufficient statistic $T(\mathbf{X})$ is called a *minimal sufficient statistic* if, for any other sufficient statistic $T'(\mathbf{X}), T(\mathbf{x})$ is a function of $T'(\mathbf{x})$.

Theorem 6.2.13 *Let $f(\mathbf{x}|\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called *complete* if $\mathrm{E}_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic.*

Theorem 6.2.25 (Complete statistics in the exponential family) *Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j)t_j(x)\right), \tag{6.2.7}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in $\Re^k$.

Point estimation

Definition 7.3.7 An estimator W^* is a *best unbiased estimator* of $\tau(\theta)$ if it satisfies $\mathrm{E}_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $\mathrm{E}_\theta W = \tau(\theta)$, we have $\mathrm{Var}_\theta\, W^* \leq \mathrm{Var}_\theta\, W$ for all θ. W^* is also called a *uniform minimum variance unbiased estimator* (UMVUE) of $\tau(\theta)$.

Theorem 7.3.9 (Cramér–Rao Inequality) *Let $X_1, \ldots, X_n$ be a sample with pdf $f(\mathbf{x}|\theta)$, and let $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ be any estimator satisfying*

$$\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\left[W(\mathbf{x})f(\mathbf{x}|\theta)\right]\, d\mathbf{x}$$

(7.3.4) *and*

$$\mathrm{Var}_\theta W(\mathbf{X}) < \infty.$$

Then

$$\mathrm{Var}_\theta\,(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X})\right)^2}{\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right)}. \tag{7.3.5}$$

Corollary 7.3.10 (Cramér–Rao Inequality, iid case) *If the assumptions of Theorem 7.3.9 are satisfied and, additionally, if $X_1, \ldots, X_n$ are iid with pdf $f(x|\theta)$, then*

$$\mathrm{Var}_\theta\, W(\mathbf{X}) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X})\right)^2}{n\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)}.$$

Lemma 7.3.11 *If $f(x|\theta)$ satisfies*

$$\frac{d}{d\theta}\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) f(x|\theta)\right] dx$$

(true for an exponential family), then

$$\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right) = -\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right).$$

Corollary 7.3.15 (Attainment) *Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, where $f(x|\theta)$ satisfies the conditions of the Cramér–Rao Theorem. Let $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ denote the likelihood function. If $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\mathbf{X})$ attains the Cramér–Rao Lower Bound if and only if*

$$a(\theta)[W(\mathbf{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) \tag{7.3.12}$$

for some function $a(\theta)$.

Theorem 7.3.17 (Rao–Blackwell) *Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = \mathrm{E}(W|T)$. Then $\mathrm{E}_\theta\phi(T) = \tau(\theta)$ and $\mathrm{Var}_\theta\,\phi(T) \leq \mathrm{Var}_\theta\, W$ for all θ; that is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.*

Theorem 7.3.19 *If W is a best unbiased estimator of $\tau(\theta)$, then W is unique.*

Theorem 7.3.20 *If $\mathrm{E}_\theta W = \tau(\theta), W$ is the best unbiased estimator of $\tau(\theta)$ if and only if W is uncorrelated with all unbiased estimators of 0.*

Theorem 7.3.23 *Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.*

Point estimation – large sample results

Definition 10.1.1 A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a *consistent sequence of estimators* of the parameter θ if, for every $\epsilon > 0$ and every $\theta \in \Theta$,

(10.1.1) $$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1.$$

Theorem 10.1.3 *If W_n is a sequence of estimators of a parameter θ satisfying*

i. $\lim_{n\to\infty} \mathrm{Var}_\theta\, W_n = 0,$

ii. $\lim_{n\to\infty} \mathrm{Bias}_\theta W_n = 0,$

for every $\theta \in \Theta$, then W_n is a consistent sequence of estimators of θ.

Definition 10.1.9 For an estimator T_n, suppose that $k_n(T_n - \tau(\theta)) \to \mathrm{n}(0, \sigma^2)$ in distribution. The parameter σ^2 is called the *asymptotic variance* or *variance of the limit distribution* of T_n.

Definition 10.1.11 A sequence of estimators W_n is *asymptotically efficient* for a parameter $\tau(\theta)$ if $\sqrt{n}[W_n - \tau(\theta)] \to \mathrm{n}[0, v(\theta)]$ in distribution and

$$v(\theta) = \frac{[\tau'(\theta)]^2}{\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)};$$

that is, the asymptotic variance of W_n achieves the Cramér–Rao Lower Bound.

Theorem 10.1.12 (Asymptotic efficiency of MLEs) *Let $X_1, X_2, \ldots,$ be iid $f(x|\theta)$, let $\hat\theta$ denote the MLE of θ, and let $\tau(\theta)$ be a continuous function of θ. Under the regularity conditions in Miscellanea 10.6.2 on $f(x|\theta)$ and, hence, $L(\theta|\mathbf{x})$,*

$$\sqrt{n}[\tau(\hat\theta) - \tau(\theta)] \to \mathrm{n}[0, v(\theta)],$$

where $v(\theta)$ is the Cramér–Rao Lower Bound. That is, $\tau(\hat\theta)$ is a consistent and asymptotically efficient estimator of $\tau(\theta)$.

Definition 10.1.16 If two estimators W_n and V_n satisfy

$$\sqrt{n}[W_n - \tau(\theta)] \to \mathrm{n}[0, \sigma_W^2]$$
$$\sqrt{n}[V_n - \tau(\theta)] \to \mathrm{n}[0, \sigma_V^2]$$

in distribution, the *asymptotic relative efficiency* (ARE) of V_n with respect to W_n is

$$\mathrm{ARE}(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}.$$

Hypothesis testing

Definition 8.2.1 The *likelihood ratio test statistic* for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}.$$

A *likelihood ratio test* (LRT) is any test that has a rejection region of the form $\{\mathbf{x}: \lambda(\mathbf{x}) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

Theorem 8.2.4 *If $T(\mathbf{X})$ is a sufficient statistic for θ and $\lambda^*(t)$ and $\lambda(\mathbf{x})$ are the LRT statistics based on T and $\mathbf{X}$, respectively, then $\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})$ for every $\mathbf{x}$ in the sample space.*

Definition 8.3.1 The *power function* of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.

Definition 8.3.5 For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a *size α test* if $\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha$.

Definition 8.3.6 For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a *level α test* if $\sup_{\theta\in\Theta_0} \beta(\theta) \leq \alpha$.

Definition 8.3.11 Let $\mathcal{C}$ be a class of tests for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a *uniformly most powerful* (UMP) *class $\mathcal{C}$ test* if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

Theorem 8.3.12 (Neyman–Pearson Lemma) *Consider testing $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$, where the pdf or pmf corresponding to θ_i is $f(\mathbf{x}|\theta_i), i = 0, 1$, using a test with rejection region R that satisfies*

$$\mathbf{x} \in R \quad \textit{if} \quad f(\mathbf{x}|\theta_1) > kf(\mathbf{x}|\theta_0)$$

(8.3.1) *and*

$$\mathbf{x} \in R^c \quad \textit{if} \quad f(\mathbf{x}|\theta_1) < kf(\mathbf{x}|\theta_0),$$

for some $k \geq 0$, and

(8.3.2) $$\alpha = P_{\theta_0}(\mathbf{X} \in R).$$

Then

a. *(Sufficiency) Any test that satisfies (8.3.1) and (8.3.2) is a UMP level α test.*

b. *(Necessity) If there exists a test satisfying (8.3.1) and (8.3.2) with $k > 0$, then every UMP level α test is a size α test (satisfies (8.3.2)) and every UMP level α test satisfies (8.3.1) except perhaps on a set A satisfying $P_{\theta_0}(\mathbf{X} \in A) = P_{\theta_1}(\mathbf{X} \in A) = 0$.*

Corollary 8.3.13 *Consider the hypothesis problem posed in Theorem 8.3.12. Suppose $T(\mathbf{X})$ is a sufficient statistic for θ and $g(t|\theta_i)$ is the pdf or pmf of T corresponding to θ_i, $i = 0, 1$. Then any test based on T with rejection region S (a subset of the sample space of T) is a UMP level α test if it satisfies*

$$t \in S \quad \text{if} \quad g(t|\theta_1) > kg(t|\theta_0)$$

(8.3.4) *and*

$$t \in S^c \quad \text{if} \quad g(t|\theta_1) < kg(t|\theta_0),$$

for some $k \geq 0$, where

(8.3.5)
$$\alpha = P_{\theta_0}(T \in S).$$

Definition 8.3.16 A family of pdfs or pmfs $\{g(t|\theta)\colon \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a *monotone likelihood ratio* (MLR) if, for every $\theta_2 > \theta_1$, $g(t|\theta_2)/g(t|\theta_1)$ is a monotone (nonincreasing or nondecreasing) function of t on $\{t\colon g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$. Note that $c/0$ is defined as ∞ if $0 < c$.

Theorem 8.3.17 (Karlin–Rubin) *Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose that T is a sufficient statistic for θ and the family of pdfs or pmfs $\{g(t|\theta) : \theta \in \Theta\}$ of T has an MLR. Then for any t_0, the test that rejects H_0 if and only if $T > t_0$ is a UMP level α test, where $\alpha = P_{\theta_0}(T > t_0)$.*

(For this result to be valid, the likelihood ratio should be nondecreasing.)

Hypothesis testing – large sample results

Theorem 10.3.1 (Asymptotic distribution of the LRT—simple H_0) *For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, suppose $X_1, \ldots, X_n$ are iid $f(x|\theta)$, $\hat{\theta}$ is the MLE of θ, and $f(x|\theta)$ satisfies the regularity conditions in Miscellanea 10.6.2. Then under H_0, as $n \to \infty$,*

$$-2 \log \lambda(\mathbf{X}) \to \chi_1^2 \text{ in distribution,}$$

where χ_1^2 is a χ^2 random variable with 1 degree of freedom.

Theorem 10.3.3 *Let $X_1, \ldots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. Under the regularity conditions in Miscellanea 10.6.2, if $\theta \in \Theta_0$, then the distribution of the statistic $-2 \log \lambda(\mathbf{X})$ converges to a chi squared distribution as the sample size $n \to \infty$. The degrees of freedom of the limiting distribution is the difference between the number of free parameters specified by $\theta \in \Theta_0$ and the number of free parameters specified by $\theta \in \Theta$.*

Confidence intervals (interval estimators)

Definition 9.1.1 An *interval estimate* of a real-valued parameter θ is any pair of functions, $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$, of a sample that satisfy $L(\mathbf{x}) \leq U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. If $\mathbf{X} = \mathbf{x}$ is observed, the inference $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$ is made. The random interval $[L(\mathbf{X}), U(\mathbf{X})]$ is called an *interval estimator*.

Definition 9.1.4 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *coverage probability* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the probability that the random interval $[L(\mathbf{X}), U(\mathbf{X})]$ covers the true parameter, θ. In symbols, it is denoted by either $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$ or $P(\theta \in [L(\mathbf{X}), U(\mathbf{X})] | \theta)$.

Definition 9.1.5 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *confidence coefficient* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the infimum of the coverage probabilities, $\inf_\theta P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.

Theorem 9.2.2 *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0\colon \theta = \theta_0$. For each $\mathbf{x} \in \mathcal{X}$, define a set $C(\mathbf{x})$ in the parameter space by*

(9.2.1) $$C(\mathbf{x}) = \{\theta_0 \colon \mathbf{x} \in A(\theta_0)\}.$$

Then the random set $C(\mathbf{X})$ is a $1 - \alpha$ confidence set. Conversely, let $C(\mathbf{X})$ be a $1 - \alpha$ confidence set. For any $\theta_0 \in \Theta$, define

$$A(\theta_0) = \{\mathbf{x} \colon \theta_0 \in C(\mathbf{x})\}.$$

Then $A(\theta_0)$ is the acceptance region of a level α test of $H_0\colon \theta = \theta_0$.