Chapter 8

Large Sample Theory

8.1 The CLT, Delta Method and an Exponential Family Limit Theorem

Large sample theory, also called asymptotic theory, is used to approximate the distribution of an estimator when the sample size n is large. This theory is extremely useful if the exact sampling distribution of the estimator is complicated or unknown. To use this theory, one must determine what the estimator is estimating, the rate of convergence, the asymptotic distribution, and how large n must be for the approximation to be useful. Moreover, the (asymptotic) standard error (SE), an estimator of the asymptotic standard deviation, must be computable if the estimator is to be useful for inference.

Theorem 8.1: the Central Limit Theorem (CLT). Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $\text{VAR}(Y) = \sigma^2$. Let the sample mean $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Then

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Hence

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n Y_i - n\mu}{n\sigma}\right) \xrightarrow{D} N(0, 1).$$

Note that the sample mean is estimating the *population mean* μ with a $\sqrt{n}$ convergence rate, the asymptotic distribution is normal, and the SE $= S/\sqrt{n}$ where S is the *sample standard deviation.* For many distributions the central limit theorem provides a good approximation if the sample size $n > 30$. A special case of the CLT is proven at the end of Section 4.

Notation. The notation $X \sim Y$ and $X \stackrel{D}{=} Y$ both mean that the random variables X and Y have the same distribution. See Definition 1.24. The notation $Y_n \stackrel{D}{\rightarrow} X$ means that for large n we can approximate the cdf of Y_n by the cdf of X. The distribution of X is the limiting distribution or asymptotic distribution of Y_n. For the CLT, notice that

$$Z_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) = \left(\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}\right)$$

is the z–score of $\overline{Y}$. If $Z_n \stackrel{D}{\rightarrow} N(0,1)$, then the notation $Z_n \approx N(0,1)$, also written as $Z_n \sim AN(0,1)$, means approximate the cdf of Z_n by the standard normal cdf. Similarly, the notation

$$\overline{Y}_n \approx N(\mu, \sigma^2/n),$$

also written as $\overline{Y}_n \sim AN(\mu, \sigma^2/n)$, means approximate cdf of $\overline{Y}_n$ as if $\overline{Y}_n \sim N(\mu, \sigma^2/n)$.

The two main applications of the CLT are to give the limiting distribution of $\sqrt{n}(\overline{Y}_n - \mu)$ and the limiting distribution of $\sqrt{n}(Y_n/n - \mu_X)$ for a random variable Y_n such that $Y_n = \sum_{i=1}^n X_i$ where the X_i are iid with $E(X) = \mu_X$ and $\text{VAR}(X) = \sigma_X^2$. Several of the random variables in Theorems 2.17 and 2.18 can be approximated in this way.

Example 8.1. a) Let $Y_1, ..., Y_n$ be iid Ber(ρ). Then $E(Y) = \rho$ and $\text{VAR}(Y) = \rho(1-\rho)$. Hence

$$\sqrt{n}(\overline{Y}_n - \rho) \stackrel{D}{\rightarrow} N(0, \rho(1-\rho))$$

by the CLT.

b) Now suppose that $Y_n \sim BIN(n, \rho)$. Then $Y_n \stackrel{D}{=} \sum_{i=1}^n X_i$ where $X_1, ..., X_n$ are iid Ber(ρ). Hence

$$\sqrt{n}(\frac{Y_n}{n} - \rho) \stackrel{D}{\rightarrow} N(0, \rho(1-\rho))$$

since

$$\sqrt{n}(\frac{Y_n}{n} - \rho) \stackrel{D}{=} \sqrt{n}(\overline{X}_n - \rho) \stackrel{D}{\rightarrow} N(0, \rho(1-\rho))$$

by a).

c) Now suppose that $Y_n \sim BIN(k_n, \rho)$ where $k_n \to \infty$ as $n \to \infty$. Then

$$\sqrt{k_n}(\frac{Y_n}{k_n} - \rho) \approx N(0, \rho(1-\rho))$$

or

$$\frac{Y_n}{k_n} \approx N\left(\rho, \frac{\rho(1-\rho)}{k_n}\right) \quad \text{or} \quad \mathrm{Y_n \approx N\left(k_n\rho, k_n\rho(1-\rho)\right)}.$$

Theorem 8.2: the Delta Method. If $g'(\theta) \neq 0$ and

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2),$$

then

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2).$$

Example 8.2. Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $\text{VAR}(Y) = \sigma^2$. Then by the CLT,

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Let $g(\mu) = \mu^2$. Then $g'(\mu) = 2\mu \neq 0$ for $\mu \neq 0$. Hence

$$\sqrt{n}((\overline{Y}_n)^2 - \mu^2) \xrightarrow{D} N(0, 4\sigma^2\mu^2)$$

for $\mu \neq 0$ by the delta method.

Example 8.3. Let $X \sim \text{Binomial}(n, p)$ where the positive integer n is large and $0 < p < 1$. Find the limiting distribution of $\sqrt{n} \left[\left(\frac{X}{n}\right)^2 - p^2 \right]$.

Solution. Example 8.1b gives the limiting distribution of $\sqrt{n}(\frac{X}{n} - p)$. Let $g(p) = p^2$. Then $g'(p) = 2p$ and by the delta method,

$$\sqrt{n} \left[\left(\frac{X}{n}\right)^2 - p^2 \right] = \sqrt{n} \left(g(\frac{X}{n}) - g(p) \right) \xrightarrow{D}$$

$$N(0, p(1-p)(g'(p))^2) = N(0, p(1-p)4p^2) = N(0, 4p^3(1-p)).$$

Example 8.4. Let $X_n \sim \text{Poisson}(n\lambda)$ where the positive integer n is large and $0 < \lambda$.

a) Find the limiting distribution of $\sqrt{n} \left(\frac{X_n}{n} - \lambda \right)$.

b) Find the limiting distribution of $\sqrt{n} \left[\sqrt{\frac{X_n}{n}} - \sqrt{\lambda} \right]$.

Solution. a) $X_n \stackrel{D}{=} \sum_{i=1}^n Y_i$ where the Y_i are iid Poisson(λ). Hence $E(Y) = \lambda = Var(Y)$. Thus by the CLT,

$$\sqrt{n} \left(\frac{X_n}{n} - \lambda \right) \stackrel{D}{=} \sqrt{n} \left(\frac{\sum_{i=1}^n Y_i}{n} - \lambda \right) \xrightarrow{D} N(0, \lambda).$$

b) Let $g(\lambda) = \sqrt{\lambda}$. Then $g'(\lambda) = \frac{1}{2\sqrt{\lambda}}$ and by the delta method,

$$\sqrt{n} \left[\sqrt{\frac{X_n}{n}} - \sqrt{\lambda} \right] = \sqrt{n} \left(g(\frac{X_n}{n}) - g(\lambda) \right) \xrightarrow{D}$$

$$N(0, \lambda\ (g'(\lambda))^2) = N(0, \lambda \frac{1}{4\lambda}) = N(0, \frac{1}{4}).$$

Example 8.5. Let $Y_1, ..., Y_n$ be independent and identically distributed (iid) from a Gamma(α, β) distribution.

a) Find the limiting distribution of $\sqrt{n} \left(\overline{Y} - \alpha\beta \right)$.

b) Find the limiting distribution of $\sqrt{n} \left((\overline{Y})^2 - c \right)$ for appropriate constant c.

Solution: a) Since $E(Y) = \alpha\beta$ and $V(Y) = \alpha\beta^2$, by the CLT $\sqrt{n} \left(\overline{Y} - \alpha\beta \right) \xrightarrow{D} N(0, \alpha\beta^2)$.

b) Let $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. Let $g(\mu) = \mu^2$ so $g'(\mu) = 2\mu$ and $[g'(\mu)]^2 = 4\mu^2 = 4\alpha^2\beta^2$. Then by the delta method, $\sqrt{n} \left((\overline{Y})^2 - c \right) \xrightarrow{D} N(0, \sigma^2[g'(\mu)]^2) = N(0, 4\alpha^3\beta^4)$ where $c = \mu^2 = \alpha^2\beta^2$.

Barndorff–Nielsen (1982), Casella and Berger (2002, p. 472, 515), Cox and Hinckley (1974, p. 286), Lehmann and Casella (1998, Section 6.3), Schervish (1995, p. 418), and many others suggest that under regularity conditions if $Y_1, ..., Y_n$ are iid from a one parameter regular exponential family, and if $\hat{\theta}$ is the MLE of θ, then

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right) = N[0, FCRLB_1(\tau(\theta))] \qquad (8.1)$$

where the Fréchet Cramér Rao lower bound for $\tau(\theta)$ is

$$FCRLB_1(\tau(\theta)) = \frac{[\tau'(\theta)]^2}{I_1(\theta)}$$

and the Fisher information based on a sample of size one is

$$I_1(\theta) = -E_\theta[\frac{\partial^2}{\partial\theta^2}\log(f(X|\theta))].$$

Notice that if

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right),$$

then (8.1) follows by the delta method. Also recall that $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$ by the invariance principle and that

$$I_1(\tau(\theta)) = \frac{I_1(\theta)}{[\tau'(\theta)]^2}$$

if $\tau'(\theta) \neq 0$ by Definition 6.3.

For a 1P–REF, $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n t(Y_i)$ is the UMVUE and generally the MLE of its expectation $\mu_t \equiv \mu_T = E_\theta(T_n) = E_\theta[t(Y)]$. Let $\sigma_t^2 = \text{VAR}_\theta[t(Y)]$. These values can be found by using the distribution of $t(Y)$ (see Theorems 3.6 and 3.7) or by the following result.

Proposition 8.3. Suppose Y is a 1P–REF with pdf or pmf

$$f(y|\theta) = h(y)c(\theta)\exp[w(\theta)t(y)]$$

and natural parameterization

$$f(y|\eta) = h(y)b(\eta)\exp[\eta t(y)].$$

Then a)

$$\mu_t = E[t(Y)] = \frac{-c'(\theta)}{c(\theta)w'(\theta)} = \frac{-\partial}{\partial\eta}\log(b(\eta)), \tag{8.2}$$

and b)

$$\sigma_t^2 = V[t(Y)] = \frac{\frac{-\partial^2}{\partial\theta^2}\log(c(\theta)) - [w''(\theta)]\mu_t}{[w'(\theta)]^2} = \frac{-\partial^2}{\partial\eta^2}\log(b(\eta)). \tag{8.3}$$

Proof. The proof will be for pdfs. For pmfs replace the integrals by sums. By Theorem 3.3, only the middle equalities need to be shown. By Remark 3.2 the derivative and integral operators can be interchanged for a 1P–REF. a) Since $1 = \int f(y|\theta)dy$,

$$0 = \frac{\partial}{\partial\theta} 1 = \frac{\partial}{\partial\theta} \int h(y) \exp[w(\theta)t(y) \ + \ \log(c(\theta))]dy$$

$$= \int h(y) \frac{\partial}{\partial\theta} \exp[w(\theta)t(y) \ + \ \log(c(\theta))]dy$$

$$= \int h(y) \exp[w(\theta)t(y) \ + \ \log(c(\theta))] \left(w'(\theta)t(y) + \frac{c'(\theta)}{c(\theta)} \right) dy$$

or

$$E[w'(\theta)t(Y)] = \frac{-c'(\theta)}{c(\theta)}$$

or

$$E[t(Y)] = \frac{-c'(\theta)}{c(\theta)w'(\theta)}.$$

b) Similarly,

$$0 = \int h(y) \frac{\partial^2}{\partial\theta^2} \exp[w(\theta)t(y) \ + \ \log(c(\theta))]dy.$$

From the proof of a) and since $\frac{\partial}{\partial\theta} \log(c(\theta)) = c'(\theta)/c(\theta)$,

$$0 = \int h(y) \frac{\partial}{\partial\theta} \left[\exp[w(\theta)t(y) \ + \ \log(c(\theta))] \left(w'(\theta)t(y) + \frac{\partial}{\partial\theta} \log(c(\theta)) \right) \right] dy$$

$$= \int h(y) \exp[w(\theta)t(y) \ + \ \log(c(\theta))] \ \left(w'(\theta)t(y) + \frac{\partial}{\partial\theta} \log(c(\theta)) \right)^2 dy$$

$$+ \int h(y) \exp[w(\theta)t(y) \ + \ \log(c(\theta))] \ \left(w''(\theta)t(y) + \frac{\partial^2}{\partial\theta^2} \log(c(\theta)) \right) dy.$$

So

$$E \left(w'(\theta)t(Y) + \frac{\partial}{\partial\theta} \log(c(\theta)) \right)^2 = -E \left(w''(\theta)t(Y) + \frac{\partial^2}{\partial\theta^2} \log(c(\theta)) \right). \quad (8.4)$$

Using a) shows that the left hand side of (8.4) equals

$$E\left(w'(\theta)\left(t(Y) + \frac{c'(\theta)}{c(\theta)w'(\theta)}\right)\right)^2 = [w'(\theta)]^2 \ \text{VAR}(t(Y))$$

while the right hand side of (8.4) equals

$$-\left(w''(\theta)\mu_t + \frac{\partial^2}{\partial\theta^2}\log(c(\theta))\right)$$

and the result follows. QED

The simplicity of the following result is rather surprising. When (as is usually the case) $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n t(Y_i)$ is the MLE of μ_t, $\hat{\eta} = g^{-1}(T_n)$ is the MLE of η by the invariance principle.

Theorem 8.4. Let $Y_1, ..., Y_n$ be iid from a 1P–REF with pdf or pmf

$$f(y|\theta) = h(y)c(\theta)\exp[w(\theta)t(y)]$$

and natural parameterization

$$f(y|\eta) = h(y)b(\eta)\exp[\eta t(y)].$$

Let

$$E(t(Y)) = \mu_t \equiv g(\eta)$$

and $\text{VAR}(t(Y)) = \sigma_t^2$.

a) Then

$$\sqrt{n}[\overline{T}_n - \mu_t] \xrightarrow{D} N(0, I_1(\eta))$$

where

$$I_1(\eta) = \sigma_t^2 = g'(\eta) = \frac{[g'(\eta)]^2}{I_1(\eta)}.$$

b) If $\eta = g^{-1}(\mu_t)$, $\hat{\eta} = g^{-1}(\overline{T}_n)$, and $g^{-1'}(\mu_t) \neq 0$ exists, then

$$\sqrt{n}[\hat{\eta} - \eta] \xrightarrow{D} N\left(0, \frac{1}{I_1(\eta)}\right).$$

c) Suppose the conditions in b) hold. If $\theta = w^{-1}(\eta)$, $\hat{\theta} = w^{-1}(\hat{\eta})$, $w^{-1'}$ exists and is continuous, and $w^{-1'}(\eta) \neq 0$, then

$$\sqrt{n}[\hat{\theta} - \theta] \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right).$$

d) If the conditions in c) hold, if τ' is continuous and if $\tau'(\theta) \neq 0$, then

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

Proof: a) The result follows by the central limit theorem if $\sigma_t^2 = I_1(\eta) = g'(\eta)$. Since $\log(f(y|\eta)) = \log(h(y)) + \log(b(\eta)) + \eta t(y)$,

$$\frac{\partial}{\partial \eta} \log(f(y|\eta)) = \frac{\partial}{\partial \eta} \log(b(\eta)) + t(y) = -\mu_t + t(y) = -g(\eta) + t(y)$$

by Proposition 8.3 a). Hence

$$\frac{\partial^2}{\partial \eta^2} \log(f(y|\eta)) = \frac{\partial^2}{\partial \eta^2} \log(b(\eta)) = -g'(\eta),$$

and thus by Proposition 8.3 b)

$$I_1(\eta) = \frac{-\partial^2}{\partial \eta^2} \log(b(\eta)) = \sigma_t^2 = g'(\eta).$$

b) By the delta method,

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{D} N(0, \sigma_t^2 [g^{-1'}(\mu_t)]^2),$$

but

$$g^{-1'}(\mu_t) = \frac{1}{g'(g^{-1}(\mu_t))} = \frac{1}{g'(\eta)}.$$

Since $\sigma_t^2 = I_1(\eta) = g'(\eta)$, it follows that $\sigma_t^2 = [g'(\eta)]^2 / I_1(\eta)$, and

$$\sigma_t^2 [g^{-1'}(\mu_t)]^2 = \frac{[g'(\eta)]^2}{I_1(\eta)} \frac{1}{[g'(\eta)]^2} = \frac{1}{I_1(\eta)}.$$

So

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\eta)}\right).$$

c) By the delta method,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{[w^{-1'}(\eta)]^2}{I_1(\eta)}\right),$$

but

$$\frac{[w^{-1'}(\eta)]^2}{I_1(\eta)} = \frac{1}{I_1(\theta)}.$$

The last equality holds since by Theorem 6.3c, if $\theta = g(\eta)$, if g' exists and is continuous, and if $g'(\theta) \neq 0$, then $I_1(\theta) = I_1(\eta)/[g'(\eta)]^2$. Use $\eta = w(\theta)$ so $\theta = g(\eta) = w^{-1}(\eta)$.

d) The result follows by the delta method. QED

8.2 Asymptotically Efficient Estimators

Definition 8.1. Let $Y_1, ..., Y_n$ be iid random variables. Let $T_n \equiv T_n(Y_1, ..., Y_n)$ be an estimator of a parameter μ_T such that

$$\sqrt{n}(T_n - \mu_T) \xrightarrow{D} N(0, \sigma_A^2).$$

Then the *asymptotic variance* of $\sqrt{n}(T_n - \mu_T)$ is σ_A^2 and the *asymptotic variance* (AV) of T_n is σ_A^2/n. If S_A^2 is a consistent estimator of σ_A^2, then the (asymptotic) *standard error* (SE) of T_n is $S_A/\sqrt{n}$.

Remark 8.1. Consistent estimators are defined in the following section. The parameter σ_A^2 is a function of both the estimator T_n and the underlying distribution F of Y_1. Frequently $n\text{VAR}(T_n)$ converges in distribution to σ_A^2, but not always. See Staudte and Sheather (1990, p. 51) and Lehmann (1999, p. 232).

Example 8.6. If $Y_1, ..., Y_n$ are iid from a distribution with mean μ and variance σ^2, then by the central limit theorem,

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Recall that $\text{VAR}(\overline{Y}_n) = \sigma^2/n = AV(\overline{Y}_n)$ and that the standard error $SE(\overline{Y}_n)$ $= S_n/\sqrt{n}$ where S_n^2 is the sample variance.

Definition 8.2. Let $T_{1,n}$ and $T_{2,n}$ be two estimators of a parameter θ such that

$$n^\delta(T_{1,n} - \theta) \xrightarrow{D} N(0, \sigma_1^2(F))$$

and

$$n^\delta(T_{2,n} - \theta) \xrightarrow{D} N(0, \sigma_2^2(F)),$$

then the **asymptotic relative efficiency** of $T_{1,n}$ with respect to $T_{2,n}$ is

$$ARE(T_{1,n}, T_{2,n}) = \frac{\sigma_2^2(F)}{\sigma_1^2(F)}.$$

This definition brings up several issues. First, both estimators must have the same convergence rate n^δ. Usually $\delta = 0.5$. If $T_{i,n}$ has convergence rate n^{δ_i}, then estimator $T_{1,n}$ is judged to be "better" than $T_{2,n}$ if $\delta_1 > \delta_2$. Secondly, the two estimators need to estimate the same parameter θ. This condition will often not hold unless the distribution is symmetric about μ. Then $\theta = \mu$ is a natural choice. Thirdly, estimators are often judged by their Gaussian efficiency with respect to the sample mean (thus F is the normal distribution). Since the normal distribution is a location–scale family, it is often enough to compute the ARE for the standard normal distribution. If the data come from a distribution F and the ARE can be computed, then $T_{1,n}$ is judged to be a "better" estimator (for the data distribution F) than $T_{2,n}$ if the $ARE > 1$. Similarly, $T_{1,n}$ is judged to be a "worse" estimator than $T_{2,n}$ if the $ARE < 1$. *Notice that the "better" estimator has the smaller asymptotic variance.*

The *population median* is any value MED(Y) such that

$$P(Y \le \text{MED}(Y)) \ge 0.5 \text{ and } P(Y \ge \text{MED}(Y)) \ge 0.5. \tag{8.5}$$

In simulation studies, typically the underlying distribution F belongs to a symmetric location–scale family. There are at least two reasons for using such distributions. First, if the distribution is symmetric, then the population median MED(Y) is the point of symmetry and the natural parameter to estimate. Under the symmetry assumption, there are many estimators of MED(Y) that can be compared via their ARE with respect to the sample mean or the maximum likelihood estimator (MLE). Secondly, once the ARE is obtained for one member of the family, it is typically obtained for *all members of the location–scale family.* That is, suppose that $Y_1, ..., Y_n$ are iid from a location–scale family with parameters μ and σ. Then $Y_i = \mu + \sigma Z_i$ where the Z_i are iid from the same family with $\mu = 0$ and $\sigma = 1$. Typically

$$AV[T_{i,n}(\boldsymbol{Y})] = \sigma^2 AV[T_{i,n}(\boldsymbol{Z})],$$

so

$$ARE[T_{1,n}(\boldsymbol{Y}), T_{2,n}(\boldsymbol{Y})] = ARE[T_{1,n}(\boldsymbol{Z}), T_{2,n}(\boldsymbol{Z})].$$

Theorem 8.5. Let $Y_1, ..., Y_n$ be iid with a pdf f that is positive at the population median: $f(\text{MED}(Y)) > 0$. Then

$$\sqrt{n}(MED(n) - MED(Y)) \xrightarrow{D} N\left(0, \frac{1}{4[f(MED(Y))]^2}\right).$$

Example 8.7. Let $Y_1, ..., Y_n$ be iid $N(\mu, \sigma^2)$, $T_{1,n} = \overline{Y}$ and let $T_{2,n} =$ MED(n) be the sample median. Let $\theta = \mu = E(Y) =$ MED(Y). Find $ARE(T_{1,n}, T_{2,n})$.

Solution: By the CLT, $\sigma_1^2(F) = \sigma^2$ when F is the $N(\mu, \sigma^2)$ distribution. By Theorem 8.5,

$$\sigma_2^2(F) = \frac{1}{4[f(MED(Y))]^2} = \frac{1}{4[\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-0}{2\sigma^2})]^2} = \frac{\pi\sigma^2}{2}.$$

Hence

$$ARE(T_{1,n}, T_{2,n}) = \frac{\pi\sigma^2/2}{\sigma^2} = \frac{\pi}{2} \approx 1.571$$

and the sample mean $\overline{Y}$ is a "better" estimator of μ than the sample median MED(n) for the family of normal distributions.

Recall from Definition 6.3 that $I_1(\theta)$ is the information number for θ based on a sample of size 1. Also recall that $I_1(\tau(\theta)) = I_1(\theta)/[\tau'(\theta)]^2$.

Definition 8.3. Assume $\tau'(\theta) \neq 0$. Then an estimator T_n of $\tau(\theta)$ is **asymptotically efficient** if

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right). \tag{8.6}$$

In particular, the estimator T_n of θ is asymptotically efficient if

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right). \tag{8.7}$$

Following Lehmann (1999, p. 486), if $T_{2,n}$ is an asymptotically efficient estimator of θ, if $I_1(\theta)$ and $v(\theta)$ are continuous functions, and if $T_{1,n}$ is an estimator such that

$$\sqrt{n}(T_{1,n} - \theta) \xrightarrow{D} N(0, v(\theta)),$$

then under regularity conditions, $v(\theta) \geq 1/I_1(\theta)$ and

$$ARE(T_{1,n}, T_{2,n}) = \frac{\frac{1}{I_1(\theta)}}{v(\theta)} = \frac{1}{I_1(\theta)v(\theta)} \leq 1.$$

Hence asymptotically efficient estimators are "better" than estimators of the form $T_{1,n}$. When $T_{2,n}$ is asymptotically efficient,

$$AE(T_{1,n}) = ARE(T_{1,n}, T_{2,n}) = \frac{1}{I_1(\theta)v(\theta)}$$

is sometimes called the asymptotic efficiency of $T_{1,n}$.

Notice that for a 1P–REF, $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n t(Y_i)$ is an asymptotically efficient estimator of $g(\eta) = E(t(Y))$ by Theorem 8.4. $\overline{T}_n$ is the UMVUE of $E(t(Y))$ by the LSU theorem.

The following rule of thumb suggests that MLEs and UMVUEs are often asymptotically efficient. The rule often holds for location families where the support does not depend on θ. The rule does not hold for the uniform $(0, \theta)$ family.

Rule of Thumb 8.1. Let $\hat{\theta}_n$ be the MLE or UMVUE of θ. If $\tau'(\theta) \neq 0$, then

$$\sqrt{n}[\tau(\hat{\theta}_n) - \tau(\theta)] \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

8.3 Modes of Convergence and Consistency

Definition 8.4. Let $\{Z_n, n = 1, 2, ...\}$ be a sequence of random variables with cdfs F_n, and let X be a random variable with cdf F. Then Z_n **converges in distribution to** X, written

$$Z_n \xrightarrow{D} X,$$

or Z_n *converges in law to* X, written $Z_n \xrightarrow{L} X$, if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at each continuity point t of F. The distribution of X is called the **limiting distribution** or the **asymptotic distribution** of Z_n.

Notice that the CLT, delta method and Theorem 8.4 give the limiting distributions of $Z_n = \sqrt{n}(\overline{Y}_n - \mu)$, $Z_n = \sqrt{n}(g(T_n) - g(\theta))$ and $Z_n = \sqrt{n}(\overline{T}_n - E(t(Y)))$, respectively.

Convergence in distribution is useful because if the distribution of X_n is unknown or complicated and the distribution of X is easy to use, then for large n we can approximate the probability that X_n is in an interval by the probability that X is in the interval. To see this, notice that if $X_n \xrightarrow{D} X$, then $P(a < X_n \leq b) = F_n(b) - F_n(a) \rightarrow F(b) - F(a) = P(a < X \leq b)$ if F is continuous at a and b.

Warning: convergence in distribution says that the cdf $F_n(t)$ of X_n gets close to the cdf of F(t) of X as $n \rightarrow \infty$ provided that t is a continuity point of F. Hence for any $\epsilon > 0$ there exists N_t such that if $n > N_t$, then $|F_n(t) - F(t)| < \epsilon$. Notice that N_t depends on the value of t. Convergence in distribution does not imply that the random variables X_n converge to the random variable X.

Example 8.8. Suppose that $X_n \sim U(-1/n, 1/n)$. Then the cdf $F_n(x)$ of X_n is

$$F_n(x) = \begin{cases} 0, & x \leq \frac{-1}{n} \\ \frac{nx}{2} + \frac{1}{2}, & \frac{-1}{n} \leq x \leq \frac{1}{n} \\ 1, & x \geq \frac{1}{n}. \end{cases}$$

Sketching $F_n(x)$ shows that it has a line segment rising from 0 at $x = -1/n$ to 1 at $x = 1/n$ and that $F_n(0) = 0.5$ for all $n \geq 1$. Examining the cases $x < 0$, $x = 0$ and $x > 0$ shows that as $n \rightarrow \infty$,

$$F_n(x) \rightarrow \begin{cases} 0, & x < 0 \\ \frac{1}{2} & x = 0 \\ 1, & x > 0. \end{cases}$$

Notice that if X is a random variable such that $P(X = 0) = 1$, then X has cdf

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases}$$

Since $x = 0$ is the only discontinuity point of $F_X(x)$ and since $F_n(x) \rightarrow F_X(x)$ for all continuity points of $F_X(x)$ (ie for $x \neq 0$),

$$X_n \xrightarrow{D} X.$$

Example 8.9. Suppose $Y_n \sim U(0, n)$. Then $F_n(t) = t/n$ for $0 < t \leq n$ and $F_n(t) = 0$ for $t \leq 0$. Hence $\lim_{n\to\infty} F_n(t) = 0$ for $t \leq 0$. If $t > 0$ and $n > t$, then $F_n(t) = t/n \to 0$ as $n \to \infty$. Thus $\lim_{n\to\infty} F_n(t) = 0$ for all t and Y_n does not converge in distribution to any random variable Y since $H(t) \equiv 0$ is not a cdf.

Definition 8.5. A sequence of random variables X_n converges in distribution to a constant $\tau(\theta)$, written

$$X_n \xrightarrow{D} \tau(\theta), \quad \text{if} \quad \mathrm{X_n} \xrightarrow{\mathrm{D}} \mathrm{X}$$

where $P(X = \tau(\theta)) = 1$. The distribution of the random variable X is said to be degenerate at $\tau(\theta)$.

Definition 8.6. A sequence of random variables X_n converges in probability to a constant $\tau(\theta)$, written

$$X_n \xrightarrow{P} \tau(\theta),$$

if for every $\epsilon > 0$,

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

The sequence X_n **converges in probability to** X, written

$$X_n \xrightarrow{P} X,$$

if $X_n - X \xrightarrow{P} 0$.

Notice that $X_n \xrightarrow{P} X$ if for every $\epsilon > 0$,

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

Definition 8.7. A sequence of estimators T_n of $\tau(\theta)$ is **consistent** for $\tau(\theta)$ if

$$T_n \xrightarrow{P} \tau(\theta)$$

for every $\theta \in \Theta$. If T_n is consistent for $\tau(\theta)$, then T_n is a **consistent estimator** of $\tau(\theta)$.

Consistency is a weak property that is usually satisfied by good estimators. T_n is a consistent estimator for $\tau(\theta)$ if the probability that T_n falls in any neighborhood of $\tau(\theta)$ goes to one, regardless of the value of $\theta \in \Theta$.

Definition 8.8. For a real number $r > 0$, Y_n converges in rth mean to a random variable Y, written

$$Y_n \xrightarrow{r} Y,$$

if

$$E(|Y_n - Y|^r) \to 0$$

as $n \to \infty$. In particular, if $r = 2$, Y_n **converges in quadratic mean** to Y, written

$$Y_n \xrightarrow{2} Y \quad \text{or} \quad \mathrm{Y_n} \xrightarrow{\mathrm{qm}} \mathrm{Y},$$

if

$$E[(Y_n - Y)^2] \to 0$$

as $n \to \infty$.

Lemma 8.6: Generalized Chebyshev's Inequality. Let $u : \Re \to [0, \infty)$ be a nonnegative function. If $E[u(Y)]$ exists then for any $c > 0$,

$$P[u(Y) \geq c] \leq \frac{E[u(Y)]}{c}.$$

If $\mu = E(Y)$ exists, then taking $u(y) = |y - \mu|^r$ and $\tilde{c} = c^r$ gives
Markov's Inequality: for $r > 0$ and any $c > 0$,

$$P(|Y - \mu| \geq c] = P(|Y - \mu|^r \geq c^r] \leq \frac{E[|Y - \mu|^r]}{c^r}.$$

If $r = 2$ and $\sigma^2 = \mathrm{VAR}(Y)$ exists, then we obtain
Chebyshev's Inequality:

$$P(|Y - \mu| \geq c] \leq \frac{\mathrm{VAR}(Y)}{c^2}.$$

Proof. The proof is given for pdfs. For pmfs, replace the integrals by sums. Now

$$E[u(Y)] = \int_{\Re} u(y) f(y) dy = \int_{\{y: u(y) \geq c\}} u(y) f(y) dy + \int_{\{y: u(y) < c\}} u(y) f(y) dy$$

$$\geq \int_{\{y: u(y) \geq c\}} u(y) f(y) dy$$

since the integrand $u(y)f(y) \geq 0$. Hence

$$E[u(Y)] \geq c \int_{\{y: u(y) \geq c\}} f(y) dy = cP[u(Y) \geq c]. \quad QED$$

The following proposition gives sufficient conditions for T_n to be a consistent estimator of $\tau(\theta)$. Notice that $MSE_{\tau(\theta)}(T_n) \to 0$ for all $\theta \in \Theta$ is equivalent to $T_n \stackrel{qm}{\to} \tau(\theta)$ for all $\theta \in \Theta$.

Proposition 8.7. a) If

$$\lim_{n \to \infty} MSE_{\tau(\theta)}(T_n) = 0$$

for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$.

b) If

$$\lim_{n \to \infty} \text{VAR}_\theta(T_n) = 0 \quad \text{and} \quad \lim_{n \to \infty} E_\theta(T_n) = \tau(\theta))$$

for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$.

Proof. a) Using Lemma 8.6 with $Y = T_n$, $u(T_n) = [T_n - \tau(\theta)]^2$ and $c = \epsilon^2$ shows that for any $\epsilon > 0$,

$$P_\theta(|T_n - \tau(\theta)| \geq \epsilon) = P_\theta[(T_n - \tau(\theta))^2 \geq \epsilon^2] \leq \frac{E_\theta[(T_n - \tau(\theta))^2]}{\epsilon^2}.$$

Hence

$$\lim_{n \to \infty} E_\theta[(T_n - \tau(\theta))^2] = \lim_{n \to \infty} MSE_{\tau(\theta)}(T_n) \to 0$$

is a sufficient condition for T_n to be a consistent estimator of $\tau(\theta)$.

b) Referring to Definition 6.1,

$$MSE_{\tau(\theta)}(T_n) = \text{VAR}_\theta(T_n) + [\text{Bias}_{\tau(\theta)}(\text{T}_\text{n})]^2$$

where $\text{Bias}_{\tau(\theta)}(\text{T}_\text{n}) = \text{E}_\theta(\text{T}_\text{n}) - \tau(\theta)$. Since $MSE_{\tau(\theta)}(T_n) \to 0$ if both $\text{VAR}_\theta(T_n) \to 0$ and $\text{Bias}_{\tau(\theta)}(\text{T}_\text{n}) = \text{E}_\theta(\text{T}_\text{n}) - \tau(\theta) \to 0$, the result follows from a). QED

The following result shows estimators that converge at a $\sqrt{n}$ rate are consistent. Use this result and the delta method to show that $g(T_n)$ is a consistent estimator of $g(\theta)$. Note that b) follows from a) with $X_\theta \sim N(0, v(\theta))$.

The WLLN shows that $\overline{Y}$ is a consistent estimator of $E(Y) = \mu$ if $E(Y)$ exists.

Proposition 8.8. a) Let X be a random variable and $0 < \delta \leq 1$. If

$$n^\delta (T_n - \tau(\theta)) \xrightarrow{D} X$$

then $T_n \xrightarrow{P} \tau(\theta)$.

b) If

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$$

for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$.

Definition 8.9. A sequence of random variables X_n *converges almost everywhere* (or *almost surely*, or *with probability 1*) to X if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

This type of convergence will be denoted by

$$X_n \xrightarrow{ae} X.$$

Notation such as "X_n converges to X ae" will also be used. Sometimes "ae" will be replaced with "as" or "wp1." We say that X_n *converges almost everywhere* to $\tau(\theta)$, written

$$X_n \xrightarrow{ae} \tau(\theta),$$

if $P(\lim_{n\to\infty} X_n = \tau(\theta)) = 1$.

Theorem 8.9. Let Y_n be a sequence of iid random variables with $E(Y_i) = \mu$. Then

a) **Strong Law of Large Numbers (SLLN):** $\overline{Y}_n \xrightarrow{ae} \mu$, and

b) **Weak Law of Large Numbers (WLLN):** $\overline{Y}_n \xrightarrow{P} \mu$.

Proof of WLLN when $V(Y_i) = \sigma^2$: By Chebyshev's inequality, for every $\epsilon > 0$,

$$P(|\overline{Y}_n - \mu| \geq \epsilon) \leq \frac{V(\overline{Y}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

as $n \to \infty$. QED

8.4 Slutsky's Theorem and Related Results

Theorem 8.10: Slutsky's Theorem. Suppose $Y_n \xrightarrow{D} Y$ and $W_n \xrightarrow{P} w$ for some constant w. Then

a) $Y_n + W_n \xrightarrow{D} Y + w$,

b) $Y_n W_n \xrightarrow{D} wY$, and

c) $Y_n/W_n \xrightarrow{D} Y/w$ if $w \neq 0$.

Theorem 8.11. a) If $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{D} X$.

b) If $X_n \xrightarrow{ae} X$ then $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$.

c) If $X_n \xrightarrow{r} X$ then $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$.

d) $X_n \xrightarrow{P} \tau(\theta)$ iff $\mathrm{X_n} \xrightarrow{\mathrm{D}} \tau(\theta)$.

e) If $X_n \xrightarrow{P} \theta$ and τ is continuous at θ, then $\tau(X_n) \xrightarrow{P} \tau(\theta)$.

f) If $X_n \xrightarrow{D} \theta$ and τ is continuous at θ, then $\tau(X_n) \xrightarrow{D} \tau(\theta)$.

Suppose that for all $\theta \in \Theta$, $T_n \xrightarrow{D} \tau(\theta)$, $T_n \xrightarrow{r} \tau(\theta)$ or $T_n \xrightarrow{ae} \tau(\theta)$. Then T_n is a consistent estimator of $\tau(\theta)$ by Theorem 8.11.

Example 8.10. Let $Y_1, ..., Y_n$ be iid with mean $E(Y_i) = \mu$ and variance $V(Y_i) = \sigma^2$. Then the sample mean $\overline{Y}_n$ is a consistent estimator of μ since i) the SLLN holds (use Theorem 8.9 and 8.11), ii) the WLLN holds and iii) the CLT holds (use Proposition 8.8). Since

$$\lim_{n\to\infty} \mathrm{VAR}_\mu(\overline{Y}_n) = \lim_{n\to\infty} \sigma^2/n = 0 \quad \text{and} \quad \lim_{n\to\infty} E_\mu(\overline{Y}_n) = \mu,$$

$\overline{Y}_n$ is also a consistent estimator of μ by Proposition 8.7b. By the delta method and Proposition 8.8b, $T_n = g(\overline{Y}_n)$ is a consistent estimator of $g(\mu)$ if $g'(\mu) \neq 0$ for all $\mu \in \Theta$. By Theorem 8.11e, $g(\overline{Y}_n)$ is a consistent estimator of $g(\mu)$ if g is continuous at μ for all $\mu \in \Theta$.

Theorem 8.12: Generalized Continuous Mapping Theorem. If $X_n \xrightarrow{D} X$ and the function g is such that $P[X \in C(g)] = 1$ where $C(g)$ is the set of points where g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Remark 8.2. For Theorem 8.11, a) follows from Slutsky's Theorem by taking $Y_n \equiv X = Y$ and $W_n = X_n - X$. Then $Y_n \xrightarrow{D} Y = X$ and $W_n \xrightarrow{P} 0$. Hence $X_n = Y_n + W_n \xrightarrow{D} Y + 0 = X$. The convergence in distribution parts of b) and c) follow from a). Part f) follows from d) and e). Part e) implies that if T_n is a consistent estimator of θ and τ is a continuous function, then $\tau(T_n)$ is a consistent estimator of $\tau(\theta)$. Theorem 8.12 says that convergence in distribution is preserved by continuous functions, and even some discontinuities are allowed as long as the set of continuity points is assigned probability 1 by the asymptotic distribution. Equivalently, the set of discontinuity points is assigned probability 0.

Example 8.11. (Ferguson 1996, p. 40): If $X_n \xrightarrow{D} X$ then $1/X_n \xrightarrow{D} 1/X$ if X is a continuous random variable since $P(X = 0) = 0$ and $x = 0$ is the only discontinuity point of $g(x) = 1/x$.

Example 8.12. Show that if $Y_n \sim t_n$, a t distribution with n degrees of freedom, then $Y_n \xrightarrow{D} Z$ where $Z \sim N(0, 1)$.

Solution: $Y_n \stackrel{D}{=} Z/\sqrt{V_n/n}$ where $Z \perp\!\!\!\perp V_n \sim \chi^2_n$. If $W_n = \sqrt{V_n/n} \xrightarrow{P} 1$, then the result follows by Slutsky's Theorem. But $V_n \stackrel{D}{=} \sum_{i=1}^n X_i$ where the iid $X_i \sim \chi^2_1$. Hence $V_n/n \xrightarrow{P} 1$ by the WLLN and $\sqrt{V_n/n} \xrightarrow{P} 1$ by Theorem 8.11e.

Theorem 8.13: Continuity Theorem. Let Y_n be sequence of random variables with characteristic functions $\phi_n(t)$. Let Y be a random variable with cf $\phi(t)$.

a)

$$Y_n \xrightarrow{D} Y \quad \text{iff} \quad \phi_{\rm n}({\rm t}) \to \phi({\rm t}) \ \forall {\rm t} \in \Re.$$

b) Also assume that Y_n has mgf m_n and Y has mgf m. Assume that all of the mgfs m_n and m are defined on $|t| \le d$ for some $d > 0$. Then if $m_n(t) \to m(t)$ as $n \to \infty$ for all $|t| < c$ where $0 < c < d$, then $Y_n \xrightarrow{D} Y$.

Application: Proof of a Special Case of the CLT. Following Rohatgi (1984, p. 569-9), let $Y_1, ..., Y_n$ be iid with mean μ, variance σ^2 and mgf $m_Y(t)$ for $|t| < t_o$. Then

$$Z_i = \frac{Y_i - \mu}{\sigma}$$

has mean 0, variance 1 and mgf $m_Z(t) = \exp(-t\mu/\sigma) m_Y(t/\sigma)$ for $|t| < \sigma t_o$. Want to show that

$$W_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) \xrightarrow{D} N(0, 1).$$

Notice that $W_n =$

$$n^{-1/2}\sum_{i=1}^n Z_i = n^{-1/2}\sum_{i=1}^n \left(\frac{Y_i-\mu}{\sigma}\right) = n^{-1/2}\frac{\sum_{i=1}^n Y_i - n\mu}{\sigma} = \frac{n^{-1/2}}{\frac{1}{n}}\ \frac{\overline{Y}_n-\mu}{\sigma}.$$

Thus

$$m_{W_n}(t) = E(e^{tW_n}) = E[\exp(tn^{-1/2}\sum_{i=1}^n Z_i)] = E[\exp(\sum_{i=1}^n tZ_i/\sqrt{n})]$$

$$= \prod_{i=1}^n E[e^{tZ_i/\sqrt{n}}] = \prod_{i=1}^n m_Z(t/\sqrt{n}) = [m_Z(t/\sqrt{n})]^n.$$

Set $\phi(x) = \log(m_Z(x))$. Then

$$\log[m_{W_n}(t)] = n\log[m_Z(t/\sqrt{n})] = n\phi(t/\sqrt{n}) = \frac{\phi(t/\sqrt{n})}{\frac{1}{n}}.$$

Now $\phi(0) = \log[m_Z(0)] = \log(1) = 0$. Thus by L'Hôpital's rule (where the derivative is with respect to n), $\lim_{n\to\infty}\log[m_{W_n}(t)] =$

$$\lim_{n\to\infty}\frac{\phi(t/\sqrt{n}\ )}{\frac{1}{n}} = \lim_{n\to\infty}\frac{\phi'(t/\sqrt{n}\ )[\frac{-t/2}{n^{3/2}}]}{(\frac{-1}{n^2})} = \frac{t}{2}\lim_{n\to\infty}\frac{\phi'(t/\sqrt{n}\ )}{\frac{1}{\sqrt{n}}}.$$

Now

$$\phi'(0) = \frac{m_Z'(0)}{m_Z(0)} = E(Z_i)/1 = 0,$$

so L'Hôpital's rule can be applied again, giving $\lim_{n\to\infty}\log[m_{W_n}(t)] =$

$$\frac{t}{2}\lim_{n\to\infty}\frac{\phi''(t/\sqrt{n}\ )[\frac{-t}{2n^{3/2}}]}{(\frac{-1}{2n^{3/2}})} = \frac{t^2}{2}\lim_{n\to\infty}\phi''(t/\sqrt{n}\ ) = \frac{t^2}{2}\phi''(0).$$

Now

$$\phi''(t) = \frac{d}{dt}\frac{m_Z'(t)}{m_Z(t)} = \frac{m_Z''(t)m_Z(t) - (m_Z'(t))^2}{[m_Z(t)]^2}.$$

So

$$\phi''(o) = m_Z''(0) - [m_Z'(0)]^2 = E(Z_i^2) - [E(Z_i)]^2 = 1.$$

Hence $\lim_{n\to\infty}\log[m_{W_n}(t)] = t^2/2$ and

$$\lim_{n\to\infty} m_{W_n}(t) = \exp(t^2/2)$$

which is the N(0,1) mgf. Thus by the continuity theorem,

$$W_n = \sqrt{n}\left(\frac{\overline{Y}_n-\mu}{\sigma}\right) \xrightarrow{D} N(0,1).$$

8.5 Order Relations and Convergence Rates

Definition 8.10. Lehmann (1999, p. 53-54): a) A sequence of random variables W_n is *tight* or *bounded in probability,* written $W_n = O_P(1)$, if for every $\epsilon > 0$ there exist positive constants D_ϵ and N_ϵ such that

$$P(|W_n| \leq D_\epsilon) \geq 1 - \epsilon$$

for all $n \geq N_\epsilon$. Also $W_n = O_P(X_n)$ if $|W_n/X_n| = O_P(1)$.

b) The sequence $W_n = o_P(n^{-\delta})$ if $n^\delta W_n = o_P(1)$ which means that

$$n^\delta W_n \xrightarrow{P} 0.$$

c) W_n has the same order as X_n in probability, written $W_n \asymp_P X_n$, if for every $\epsilon > 0$ there exist positive constants N_ϵ and $0 < d_\epsilon < D_\epsilon$ such that

$$P\left(d_\epsilon \leq \left|\frac{W_n}{X_n}\right| \leq D_\epsilon\right) \geq 1 - \epsilon$$

for all $n \geq N_\epsilon$.

d) Similar notation is used for a $k \times r$ matrix $\boldsymbol{A}_n = [a_{i,j}(n)]$ if each element $a_{i,j}(n)$ has the desired property. For example, $\boldsymbol{A}_n = O_P(n^{-1/2})$ if each $a_{i,j}(n) = O_P(n^{-1/2})$.

Definition 8.11. Let $\hat{\boldsymbol{\beta}}_n$ be an estimator of a $p \times 1$ vector $\boldsymbol{\beta}$, and let $W_n = \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$.

a) If $W_n \asymp_P n^{-\delta}$ for some $\delta > 0$, then both W_n and $\hat{\boldsymbol{\beta}}_n$ have (tightness) **rate** n^δ.

b) If there exists a constant κ such that

$$n^\delta (W_n - \kappa) \xrightarrow{D} X$$

for some nondegenerate random variable X, then both W_n and $\hat{\boldsymbol{\beta}}_n$ have *convergence rate* n^δ.

Proposition 8.14. Suppose there exists a constant κ such that

$$n^\delta (W_n - \kappa) \xrightarrow{D} X.$$

a) Then $W_n = O_P(n^{-\delta})$.

b) If X is not degenerate, then $W_n \asymp_P n^{-\delta}$.

The above result implies that if W_n has convergence rate n^δ, then W_n has tightness rate n^δ, and the term "tightness" will often be omitted. Part a) is proved, for example, in Lehmann (1999, p. 67).

The following result shows that if $W_n \asymp_P X_n$, then $X_n \asymp_P W_n$, $W_n = O_P(X_n)$ and $X_n = O_P(W_n)$. Notice that if $W_n = O_P(n^{-\delta})$, then n^δ is a lower bound on the rate of W_n. As an example, if the CLT holds then $\overline{Y}_n = O_P(n^{-1/3})$, but $\overline{Y}_n \asymp_P n^{-1/2}$.

Proposition 8.15. a) If $W_n \asymp_P X_n$ then $X_n \asymp_P W_n$.
b) If $W_n \asymp_P X_n$ then $W_n = O_P(X_n)$.
c) If $W_n \asymp_P X_n$ then $X_n = O_P(W_n)$.
d) $W_n \asymp_P X_n$ iff $W_n = O_P(X_n)$ and $X_n = O_P(W_n)$.

Proof. a) Since $W_n \asymp_P X_n$,

$$P(d_\epsilon \leq \left|\frac{W_n}{X_n}\right| \leq D_\epsilon) = P(\frac{1}{D_\epsilon} \leq \left|\frac{X_n}{W_n}\right| \leq \frac{1}{d_\epsilon}) \geq 1 - \epsilon$$

for all $n \geq N_\epsilon$. Hence $X_n \asymp_P W_n$.

b) Since $W_n \asymp_P X_n$,

$$P(|W_n| \leq |X_n D_\epsilon|) \geq P(d_\epsilon \leq \left|\frac{W_n}{X_n}\right| \leq D_\epsilon) \geq 1 - \epsilon$$

for all $n \geq N_\epsilon$. Hence $W_n = O_P(X_n)$.

c) Follows by a) and b).

d) If $W_n \asymp_P X_n$, then $W_n = O_P(X_n)$ and $X_n = O_P(W_n)$ by b) and c). Now suppose $W_n = O_P(X_n)$ and $X_n = O_P(W_n)$. Then

$$P(|W_n| \leq |X_n| D_{\epsilon/2}) \geq 1 - \epsilon/2$$

for all $n \geq N_1$, and

$$P(|X_n| \leq |W_n| 1/d_{\epsilon/2}) \geq 1 - \epsilon/2$$

for all $n \geq N_2$. Hence

$$P(A) \equiv P(\left|\frac{W_n}{X_n}\right| \leq D_{\epsilon/2}) \geq 1 - \epsilon/2$$

and

$$P(B) \equiv P(d_{\epsilon/2} \leq \left|\frac{W_n}{X_n}\right|) \geq 1 - \epsilon/2$$

for all $n \geq N = \max(N_1, N_2)$. Since $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1$,

$$P(A \cap B) = P(d_{\epsilon/2} \leq \left|\frac{W_n}{X_n}\right| \leq D_{\epsilon/2}) \geq 1 - \epsilon/2 + 1 - \epsilon/2 - 1 = 1 - \epsilon$$

for all $n \geq N$. Hence $W_n \asymp_P X_n$. QED

The following result is used to prove the following Theorem 8.17 which says that if there are K estimators $T_{j,n}$ of a parameter $\boldsymbol{\beta}$, such that $\|T_{j,n} - \boldsymbol{\beta}\| = O_P(n^{-\delta})$ where $0 < \delta \leq 1$, and if T_n^* picks one of these estimators, then $\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta})$.

Proposition 8.16: Pratt (1959). Let $X_{1,n}, ..., X_{K,n}$ each be $O_P(1)$ where K is fixed. Suppose $W_n = X_{i_n,n}$ for some $i_n \in \{1, ..., K\}$. Then

$$W_n = O_P(1). \tag{8.8}$$

Proof.

$$P(\max\{X_{1,n}, ..., X_{K,n}\} \leq x) = P(X_{1,n} \leq x, ..., X_{K,n} \leq x) \leq$$

$$F_{W_n}(x) \leq P(\min\{X_{1,n}, ..., X_{K,n}\} \leq x) = 1 - P(X_{1,n} > x, ..., X_{K,n} > x).$$

Since K is finite, there exists $B > 0$ and N such that $P(X_{i,n} \leq B) > 1 - \epsilon/2K$ and $P(X_{i,n} > -B) > 1 - \epsilon/2K$ for all $n > N$ and $i = 1, ..., K$. Bonferroni's inequality states that $P(\cap_{i=1}^K A_i) \geq \sum_{i=1}^K P(A_i) - (K-1)$. Thus

$$F_{W_n}(B) \geq P(X_{1,n} \leq B, ..., X_{K,n} \leq B) \geq$$

$$K(1 - \epsilon/2K) - (K-1) = K - \epsilon/2 - K + 1 = 1 - \epsilon/2$$

and

$$-F_{W_n}(-B) \geq -1 + P(X_{1,n} > -B, ..., X_{K,n} > -B) \geq$$

$$-1 + K(1 - \epsilon/2K) - (K-1) = -1 + K - \epsilon/2 - K + 1 = -\epsilon/2.$$

Hence

$$F_{W_n}(B) - F_{W_n}(-B) \geq 1 - \epsilon \ \text{ for } \ \text{n} > \text{N}. \ \text{ QED}$$

Theorem 8.17. Suppose $\|T_{j,n} - \boldsymbol{\beta}\| = O_P(n^{-\delta})$ for $j = 1, ..., K$ where $0 < \delta \leq 1$. Let $T_n^* = T_{i_n,n}$ for some $i_n \in \{1, ..., K\}$ where, for example, $T_{i_n,n}$ is the $T_{j,n}$ that minimized some criterion function. Then

$$\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta}). \tag{8.9}$$

Proof. Let $X_{j,n} = n^\delta \|T_{j,n} - \boldsymbol{\beta}\|$. Then $X_{j,n} = O_P(1)$ so by Proposition 8.16, $n^\delta \|T_n^* - \boldsymbol{\beta}\| = O_P(1)$. Hence $\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta})$. QED

8.6 Multivariate Limit Theorems

Many of the univariate results of the previous 5 sections can be extended to random vectors. As stated in Section 2.7, the notation for random vectors is rather awkward. For the limit theorems, the vector $\boldsymbol{X}$ is typically a $k \times 1$ column vector and $\boldsymbol{X}^T$ is a row vector. Let $\|\boldsymbol{x}\| = \sqrt{x_1^2 + \cdots + x_k^2}$ be the Euclidean norm of $\boldsymbol{x}$.

Definition 8.12. Let $\boldsymbol{X}_n$ be a sequence of random vectors with joint cdfs $F_n(\boldsymbol{x})$ and let $\boldsymbol{X}$ be a random vector with joint cdf $F(\boldsymbol{x})$.

a) $\boldsymbol{X}_n$ **converges in distribution** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \stackrel{D}{\to} \boldsymbol{X}$, if $F_n(\boldsymbol{x}) \to F(\boldsymbol{x})$ as $n \to \infty$ for all points $\boldsymbol{x}$ at which $F(\boldsymbol{x})$ is continuous. The distribution of $\boldsymbol{X}$ is the **limiting distribution** or **asymptotic distribution** of $\boldsymbol{X}_n$.

b) $\boldsymbol{X}_n$ **converges in probability** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \stackrel{P}{\to} \boldsymbol{X}$, if for every $\epsilon > 0$, $P(\|\boldsymbol{X}_n - \boldsymbol{X}\| > \epsilon) \to 0$ as $n \to \infty$.

c) Let $r > 0$ be a real number. Then $\boldsymbol{X}_n$ **converges in rth mean** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \stackrel{r}{\to} \boldsymbol{X}$, if $E(\|\boldsymbol{X}_n - \boldsymbol{X}\|^r) \to 0$ as $n \to \infty$.

d) $\boldsymbol{X}_n$ **converges almost everywhere** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \stackrel{ae}{\to} \boldsymbol{X}$, if $P(\lim_{n\to\infty} \boldsymbol{X}_n = \boldsymbol{X}) = 1$.

Theorems 8.18, 8.19 and 8.21 below are the multivariate extensions of the limit theorems in Section 8.1. When the limiting distribution of $\boldsymbol{Z}_n = \sqrt{n}(\boldsymbol{g}(\boldsymbol{T}_n) - \boldsymbol{g}(\boldsymbol{\theta}))$ is multivariate normal $N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$, approximate the joint cdf of $\boldsymbol{Z}_n$ with the joint cdf of the $N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$ distribution. Thus to find probabilities, manipulate $\boldsymbol{Z}_n$ as if $\boldsymbol{Z}_n \approx N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. To see that the CLT is a special case of the MCLT below, let $k = 1$, $E(X) = \mu$ and $V(X) = \boldsymbol{\Sigma} = \sigma^2$.

Theorem 8.18: the Multivariate Central Limit Theorem (MCLT). If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid $k \times 1$ random vectors with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then

$$\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where the sample mean

$$\overline{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i.$$

To see that the delta method is a special case of the multivariate delta method, note that if T_n and parameter θ are real valued, then $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = g'(\theta)$.

Theorem 8.19: the Multivariate Delta Method. If

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

then

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{T}_n) - \boldsymbol{g}(\boldsymbol{\theta})) \xrightarrow{D} N_d(\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \boldsymbol{\Sigma} \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})})$$

where the $d \times k$ Jacobian matrix of partial derivatives

$$\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_1(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial \theta_k} g_1(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial \theta_1} g_d(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial \theta_k} g_d(\boldsymbol{\theta}) \end{bmatrix}.$$

Here the mapping $\boldsymbol{g} : \Re^k \to \Re^d$ needs to be differentiable in a neighborhood of $\boldsymbol{\theta} \in \Re^k$.

Example 8.13. If Y has a Weibull distribution, $Y \sim W(\phi, \lambda)$, then the pdf of Y is

$$f(y) = \frac{\phi}{\lambda} y^{\phi-1} e^{-\frac{y^\phi}{\lambda}}$$

where λ, y, and ϕ are all positive. If $\mu = \lambda^{1/\phi}$ so $\mu^\phi = \lambda$, then the Weibull pdf

$$f(y) = \frac{\phi}{\mu} \left(\frac{y}{\mu}\right)^{\phi-1} \exp\left[-\left(\frac{y}{\mu}\right)^\phi\right].$$

Let $(\hat{\mu}, \hat{\phi})$ be the MLE of (μ, ϕ). According to Bain (1978, p. 215),

$$\sqrt{n}\left(\begin{pmatrix} \hat{\mu} \\ \hat{\phi} \end{pmatrix} - \begin{pmatrix} \mu \\ \phi \end{pmatrix} \right) \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.109\frac{\mu^2}{\phi^2} & 0.257\mu \\ 0.257\mu & 0.608\phi^2 \end{pmatrix} \right)$$

$= N_2(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$.

Let column vectors $\boldsymbol{\theta} = (\mu \;\; \phi)^T$ and $\boldsymbol{\eta} = (\lambda \;\; \phi)^T$. Then

$$\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta}) = \begin{pmatrix} \lambda \\ \phi \end{pmatrix} = \begin{pmatrix} \mu^\phi \\ \phi \end{pmatrix} = \begin{pmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \end{pmatrix}.$$

So

$$\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_1(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} g_1(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_1} g_2(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} g_2(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mu} \mu^\phi & \frac{\partial}{\partial \phi} \mu^\phi \\ \frac{\partial}{\partial \mu} \phi & \frac{\partial}{\partial \phi} \phi \end{bmatrix} = \begin{bmatrix} \phi \mu^{\phi-1} & \mu^\phi \log(\mu) \\ 0 & 1 \end{bmatrix}.$$

Thus by the multivariate delta method,

$$\sqrt{n}\left(\begin{pmatrix} \hat{\lambda} \\ \hat{\phi} \end{pmatrix} - \begin{pmatrix} \lambda \\ \phi \end{pmatrix} \right) \xrightarrow{D} N_2(\mathbf{0}, \mathbf{\Sigma})$$

where (see Definition 8.15 below)

$$\mathbf{\Sigma} = \boldsymbol{I}(\boldsymbol{\eta})^{-1} = [\boldsymbol{I}(\boldsymbol{g}(\boldsymbol{\theta}))]^{-1} = \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})} =$$

$$\begin{bmatrix} 1.109\lambda^2(1 + 0.4635\log(\lambda) + 0.5482(\log(\lambda))^2) & 0.257\phi\lambda + 0.608\lambda\phi\log(\lambda) \\ \\ 0.257\phi\lambda + 0.608\lambda\phi\log(\lambda) & 0.608\phi^2 \end{bmatrix}.$$

Definition 8.13. Let X be a random variable with pdf or pmf $f(x|\boldsymbol{\theta})$. Then the **information matrix**

$$\boldsymbol{I}(\boldsymbol{\theta}) = [\boldsymbol{I}_{i,j}]$$

where

$$\boldsymbol{I}_{i,j} = E\left[\frac{\partial}{\partial\theta_i}\log(f(X|\boldsymbol{\theta}))\frac{\partial}{\partial\theta_j}\log(f(X|\boldsymbol{\theta}))\right].$$

Definition 8.14. An estimator $\boldsymbol{T}_n$ of $\boldsymbol{\theta}$ is **asymptotically efficient** if

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})).$$

Following Lehmann (1999, p. 511), if $\boldsymbol{T}_n$ is asymptotically efficient and if the estimator $\boldsymbol{W}_n$ satisfies

$$\sqrt{n}(\boldsymbol{W}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\theta}))$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ and $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$ are continuous functions of $\boldsymbol{\theta}$, then under regularity conditions, $\boldsymbol{J}(\boldsymbol{\theta}) - \boldsymbol{I}^{-1}(\boldsymbol{\theta})$ is a positive semidefinite matrix, and $\boldsymbol{T}_n$ is "better" than $\boldsymbol{W}_n$.

Definition 8.15. Assume that $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta})$. Then

$$\boldsymbol{I}(\boldsymbol{\eta}) = \boldsymbol{I}(\boldsymbol{g}(\boldsymbol{\theta})) = [\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})}]^{-1}.$$

Notice that this definition agrees with the multivariate delta method if

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = \boldsymbol{I}^{-1}(\boldsymbol{\theta})$.

Now suppose that $X_1, ..., X_n$ are iid random variables from a k-parameter REF

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left[\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right] \tag{8.10}$$

with natural parameterization

$$f(x|\boldsymbol{\eta}) = h(x)b(\boldsymbol{\eta}) \exp\left[\sum_{i=1}^{k} \eta_i t_i(x)\right]. \tag{8.11}$$

Then the complete minimal sufficient statistic is

$$\overline{\boldsymbol{T}}_n = \frac{1}{n}(\sum_{i=1}^{n} t_1(X_i), ..., \sum_{i=1}^{n} t_k(X_i))^T.$$

Let $\boldsymbol{\mu}_T = (E(t_1(X), ..., E(t_k(X)))^T$. From Theorem 3.3, for $\boldsymbol{\eta} \in \Omega$,

$$E(t_i(X)) = \frac{-\partial}{\partial \eta_i} \log(b(\boldsymbol{\eta})),$$

and

$$\mathrm{Cov}(\mathrm{t_i(X)}, \mathrm{t_j(X)}) \equiv \sigma_{\mathrm{i,j}} = \frac{-\partial^2}{\partial \eta_{\mathrm{i}} \partial \eta_{\mathrm{j}}} \log(\mathrm{b}(\boldsymbol{\eta})).$$

Proposition 8.20. If the random variable X is a kP–REF with pdf or pdf (8.12), then the information matrix

$$\boldsymbol{I}(\boldsymbol{\eta}) = [\boldsymbol{I}_{i,j}]$$

where

$$\boldsymbol{I}_{i,j} = E\left[\frac{\partial}{\partial \eta_i} \log(f(X|\boldsymbol{\eta})) \frac{\partial}{\partial \eta_j} \log(f(X|\boldsymbol{\eta}))\right] = -E\left[\frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(f(X|\boldsymbol{\eta}))\right].$$

Several authors, including Barndorff–Nielsen (1982), have noted that the multivariate CLT can be used to show that $\sqrt{n}(\overline{\boldsymbol{T}}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. The fact that $\boldsymbol{\Sigma} = \boldsymbol{I}(\boldsymbol{\eta})$ appears in Lehmann (1983, p. 127).

Theorem 8.21. If $X_1, ..., X_n$ are iid from a k-parameter regular exponential family, then

$$\sqrt{n}(\overline{\boldsymbol{T}}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta})).$$

Proof. By the multivariate central limit theorem,

$$\sqrt{n}(\overline{\boldsymbol{T}}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = [\sigma_{i,j}]$. Hence the result follows if $\sigma_{i,j} = \boldsymbol{I}_{i,j}$. Since

$$\log(f(x|\boldsymbol{\eta})) = \log(h(x)) + \log(b(\boldsymbol{\eta})) + \sum_{l=1}^{k} \eta_l t_l(x),$$

$$\frac{\partial}{\partial \eta_i} \log(f(x|\boldsymbol{\eta})) = \frac{\partial}{\partial \eta_i} \log(b(\boldsymbol{\eta})) + t_i(X).$$

Hence

$$-\boldsymbol{I}_{i,j} = E\left[\frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(f(X|\boldsymbol{\eta}))\right] = \frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(b(\boldsymbol{\eta})) = -\sigma_{i,j}. \quad \text{QED}$$

To obtain standard results, use the multivariate delta method, assume that both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are $k \times 1$ vectors, and assume that $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\theta})$ is a one to one mapping so that the inverse mapping is $\boldsymbol{\theta} = \boldsymbol{g}^{-1}(\boldsymbol{\eta})$. If $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}$ is nonsingular, then

$$\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})} \tag{8.12}$$

(see Searle 1982, p. 339), and

$$\boldsymbol{I}(\boldsymbol{\eta}) = [\boldsymbol{D}_{g(\boldsymbol{\theta})} \boldsymbol{I}^{-1}(\boldsymbol{\theta}) \boldsymbol{D}^T_{g(\boldsymbol{\theta})}]^{-1} = [\boldsymbol{D}^{-1}_{g(\boldsymbol{\theta})}]^T \boldsymbol{I}(\boldsymbol{\theta}) \boldsymbol{D}^{-1}_{g(\boldsymbol{\theta})} = \boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})} \boldsymbol{I}(\boldsymbol{\theta}) \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})}. \tag{8.13}$$

Compare Lehmann (1999, p. 500) and Lehmann (1983, p. 127).

For example, suppose that $\boldsymbol{\mu}_T$ and $\boldsymbol{\eta}$ are $k \times 1$ vectors, and

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta}))$$

where $\boldsymbol{\mu}_T = \boldsymbol{g}(\boldsymbol{\eta})$ and $\boldsymbol{\eta} = \boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)$. Also assume that $\overline{\boldsymbol{T}}_n = \boldsymbol{g}(\hat{\boldsymbol{\eta}})$ and $\hat{\boldsymbol{\eta}} = \boldsymbol{g}^{-1}(\overline{\boldsymbol{T}}_n)$. Then by the multivariate delta method and Theorem 8.21,

$$\sqrt{n}(\overline{\boldsymbol{T}}_n - \boldsymbol{\mu}_T) = \sqrt{n}(\boldsymbol{g}(\hat{\boldsymbol{\eta}}) - \boldsymbol{g}(\boldsymbol{\eta})) \xrightarrow{D} N_k[\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\eta})] = N_k[\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})} \boldsymbol{I}^{-1}(\boldsymbol{\eta}) \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})}].$$

Hence
$$\boldsymbol{I}(\boldsymbol{\eta}) = \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})} \boldsymbol{I}^{-1}(\boldsymbol{\eta}) \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})}.$$
Similarly,
$$\sqrt{n}(\boldsymbol{g}^{-1}(\overline{\boldsymbol{T}}_n) - \boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)) = \sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{D} N_k[\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\eta})] =$$
$$N_k[\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)} \boldsymbol{I}(\boldsymbol{\eta}) \boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}].$$
Thus
$$\boldsymbol{I}^{-1}(\boldsymbol{\eta}) = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)} \boldsymbol{I}(\boldsymbol{\eta}) \boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)} \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\eta})} \boldsymbol{I}^{-1}(\boldsymbol{\eta}) \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\eta})} \boldsymbol{D}^T_{\boldsymbol{g}^{-1}(\boldsymbol{\mu}_T)}$$
as expected by Equation (8.13). Typically $\hat{\boldsymbol{\theta}}$ is a function of the sufficient statistic $\boldsymbol{T}_n$ and is the unique MLE of $\boldsymbol{\theta}$. Replacing $\boldsymbol{\eta}$ by $\boldsymbol{\theta}$ in the above discussion shows that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$ is equivalent to $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}_T) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{I}(\boldsymbol{\theta}))$ provided that $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})}$ is nonsingular.

8.7 More Multivariate Results

Definition 8.16. If the estimator $\boldsymbol{g}(\boldsymbol{T}_n) \xrightarrow{P} \boldsymbol{g}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, then $\boldsymbol{g}(\boldsymbol{T}_n)$ is a **consistent estimator** of $\boldsymbol{g}(\boldsymbol{\theta})$.

Proposition 8.22. If $0 < \delta \leq 1$, $\boldsymbol{X}$ is a random vector, and
$$n^{\delta}(\boldsymbol{g}(\boldsymbol{T}_n) - \boldsymbol{g}(\boldsymbol{\theta})) \xrightarrow{D} \boldsymbol{X},$$
then $\boldsymbol{g}(\boldsymbol{T}_n) \xrightarrow{P} \boldsymbol{g}(\boldsymbol{\theta})$.

Theorem 8.23. If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid, $E(\|\boldsymbol{X}\|) < \infty$ and $E(\boldsymbol{X}) = \boldsymbol{\mu}$, then
a) WLLN: $\overline{\boldsymbol{X}}_n \xrightarrow{D} \boldsymbol{\mu}$ and
b) SLLN: $\overline{\boldsymbol{X}}_n \xrightarrow{ae} \boldsymbol{\mu}$.

Theorem 8.24: Continuity Theorem. Let $\boldsymbol{X}_n$ be a sequence of $k \times 1$ random vectors with characteristic function $\phi_n(\boldsymbol{t})$ and let $\boldsymbol{X}$ be a $k \times 1$ random vector with cf $\phi(\boldsymbol{t})$. Then
$$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \text{ iff } \phi_{\rm n}(\boldsymbol{t}) \to \phi(\boldsymbol{t})$$

for all $\boldsymbol{t} \in \Re^k$.

Theorem 8.25: Cramér Wold Device. Let $\boldsymbol{X}_n$ be a sequence of $k \times 1$ random vectors and let $\boldsymbol{X}$ be a $k \times 1$ random vector. Then

$$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \text{ iff } \boldsymbol{t}^{\mathrm{T}} \boldsymbol{X}_{\mathrm{n}} \xrightarrow{\mathrm{D}} \boldsymbol{t}^{\mathrm{T}} \boldsymbol{X}$$

for all $\boldsymbol{t} \in \Re^k$.

Theorem 8.26. a) If $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$, then $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$.
b)

$$\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{g}(\boldsymbol{\theta}) \text{ iff } \boldsymbol{X}_{\mathrm{n}} \xrightarrow{\mathrm{D}} \boldsymbol{g}(\boldsymbol{\theta}).$$

Let $g(n) \geq 1$ be an increasing function of the sample size n: $g(n) \uparrow \infty$, eg $g(n) = \sqrt{n}$. See White (1984, p. 15). If a $k \times 1$ random vector $\boldsymbol{T}_n - \boldsymbol{\mu}$ converges to a nondegenerate multivariate normal distribution with convergence rate $\sqrt{n}$, then $\boldsymbol{T}_n$ has (tightness) rate $\sqrt{n}$.

Definition 8.17. Let $\boldsymbol{A}_n = [a_{i,j}(n)]$ be an $r \times c$ random matrix.
a) $\boldsymbol{A}_n = O_P(X_n)$ if $a_{i,j}(n) = O_P(X_n)$ for $1 \leq i \leq r$ and $1 \leq j \leq c$.
b) $\boldsymbol{A}_n = o_p(X_n)$ if $a_{i,j}(n) = o_p(X_n)$ for $1 \leq i \leq r$ and $1 \leq j \leq c$.
c) $\boldsymbol{A}_n \asymp_P (1/(g(n))$ if $a_{i,j}(n) \asymp_P (1/(g(n))$ for $1 \leq i \leq r$ and $1 \leq j \leq c$.
d) Let $\boldsymbol{A}_{1,n} = \boldsymbol{T}_n - \boldsymbol{\mu}$ and $\boldsymbol{A}_{2,n} = \boldsymbol{C}_n - c\boldsymbol{\Sigma}$ for some constant $c > 0$. If $\boldsymbol{A}_{1,n} \asymp_P (1/(g(n))$ and $\boldsymbol{A}_{2,n} \asymp_P (1/(g(n))$, then $(\boldsymbol{T}_n, \boldsymbol{C}_n)$ has (tightness) rate $g(n)$.

Recall that the smallest integer function $\lceil x \rceil$ rounds up, eg $\lceil 7.7 \rceil = 8$.

Definition 8.18. The *sample* α *quantile* $\hat{\xi}_{n,\alpha} = Y_{(\lceil n\alpha \rceil)}$. The *population quantile* $\xi_\alpha = Q(\alpha) = \inf\{y : F(y) \geq \alpha\}$.

Theorem 8.27: Serfling (1980, p. 80). Let $0 < \rho_1 < \rho_2 < \cdots < \rho_k < 1$. Suppose that F has a density f that is positive and continuous in neighborhoods of $\xi_{\rho_1}, ..., \xi_{\rho_k}$. Then

$$\sqrt{n}[(\hat{\xi}_{n,\rho_1}, ..., \hat{\xi}_{n,\rho_k})^T - (\xi_{\rho_1}, ..., \xi_{\rho_k})^T] \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = (\sigma_{ij})$ and

$$\sigma_{ij} = \frac{\rho_i(1-\rho_j)}{f(\xi_{\rho_i}) f(\xi_{\rho_j})}$$

for $i \leq j$ and $\sigma_{ij} = \sigma_{ji}$ for $i > j$.

Theorem 8.28: Continuous Mapping Theorem. Let $\boldsymbol{X}_n \in \Re^k$. If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ and if the function $\boldsymbol{g} : \Re^k \to \Re^j$ is continuous, then $\boldsymbol{g}(\boldsymbol{X}_n) \xrightarrow{D} \boldsymbol{g}(\boldsymbol{X})$.

8.8 Summary

1) **CLT**: Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Then $\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$.

2) **Delta Method**: If $g'(\theta) \neq 0$ and $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$, then $\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2)$.

3) **1P–REF Limit Theorem**: Let $Y_1, ..., Y_n$ be iid from a 1P–REF with pdf or pmf $f(y|\theta) = h(y)c(\theta)\exp[w(\theta)t(y)]$ and natural parameterization $f(y|\eta) = h(y)b(\eta)\exp[\eta t(y)]$. Let $E(t(Y)) = \mu_t \equiv g(\eta)$ and $V(t(Y)) = \sigma_t^2$. Then $\sqrt{n}[\overline{T}_n - \mu_t] \xrightarrow{D} N(0, I_1(\eta))$ where $I_1(\eta) = \sigma_t^2 = g'(\eta)$ and $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n t(Y_i)$.

4) **Limit theorem for the Sample Median:**
$\sqrt{n}(MED(n) - MED(Y)) \xrightarrow{D} N\left(0, \frac{1}{4f^2(MED(Y))}\right)$.

5) If $n^\delta(T_{1,n} - \theta) \xrightarrow{D} N(0, \sigma_1^2(F))$ and $n^\delta(T_{2,n} - \theta) \xrightarrow{D} N(0, \sigma_2^2(F))$, then the **asymptotic relative efficiency** of $T_{1,n}$ with respect to $T_{2,n}$ is

$$ARE(T_{1,n}, T_{2,n}) = \frac{\sigma_2^2(F)}{\sigma_1^2(F)}.$$

The "better" estimator has the smaller asymptotic variance or $\sigma_i^2(F)$.

6) An estimator T_n of $\tau(\theta)$ is **asymptotically efficient** if

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

7) For a 1P–REF, $\overline{T}_n = \frac{1}{n}\sum_{i=1}^n t(Y_i)$ is an asymptotically efficient estimator of $g(\eta) = E(t(Y))$.

8) Rule of thumb: If $\hat{\theta}_n$ is the MLE or UMVUE of θ, then $T_n = \tau(\hat{\theta}_n)$ is an asymptotically efficient estimator of $\tau(\theta)$. Hence if $\tau'(\theta) \neq 0$, then

$$\sqrt{n}[\tau(\hat{\theta}_n) - \tau(\theta)] \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

9) $X_n \xrightarrow{D} X$ if
$$\lim_{n\to\infty} F_n(t) = F(t)$$
at each continuity point t of F.

10) $X_n \xrightarrow{P} \tau(\theta)$ if for every $\epsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - \tau(\theta)| < \epsilon) = 1 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} \mathrm{P}(|X_n - \tau(\theta)| \geq \epsilon) = 0.$$

11) T_n is a **consistent estimator** of $\tau(\theta)$ if $T_n \xrightarrow{P} \tau(\theta)$ for every $\theta \in \Theta$.

12) T_n is a **consistent estimator** of $\tau(\theta)$ if any of the following 3 conditions holds:

i) $\lim_{n\to\infty} \mathrm{VAR}_\theta(T_n) = 0$ and $\lim_{n\to\infty} E_\theta(T_n) = \tau(\theta)$ for all $\theta \in \Theta$.

ii) $MSE_{\tau(\theta)}(T_n) \to 0$ for all $\theta \in \Theta$.

iii) $E[(T_n - \tau(\theta)^2) \to 0$ for all $\theta \in \Theta$.

13) If
$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$$
for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$.

14) **WLLN:** Let $Y_1, ..., Y_n, ...$ be a sequence of iid random variables with $E(Y_i) = \mu$. Then $\overline{Y}_n \xrightarrow{P} \mu$. Hence $\overline{Y}_n$ is a consistent estimator of μ.

15) i) If $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{D} X$.

ii) $T_n \xrightarrow{P} \tau(\theta)$ iff $\mathrm{T_n} \xrightarrow{D} \tau(\theta)$.

iii) If $T_n \xrightarrow{P} \theta$ and τ is continuous at θ, then $\tau(T_n) \xrightarrow{P} \tau(\theta)$. Hence if T_n is a consistent estimator of θ, then $\tau(T_n)$is a consistent estimator of $\tau(\theta)$ if τ is a continuous function on Θ.

8.9 Complements

The following extension of the delta method is sometimes useful.

Theorem 8.29. Suppose that $g'(\theta) = 0$, $g''(\theta) \neq 0$ and
$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \tau^2(\theta)).$$
Then
$$n[g(T_n) - g(\theta)] \xrightarrow{D} \frac{1}{2}\tau^2(\theta) g''(\theta) \chi^2_1.$$

Example 8.14. Let $X_n \sim$ Binomial(n, p) where the positive integer n is large and $0 < p < 1$. Let $g(\theta) = \theta^3 - \theta$. Find the limiting distribution of $n \left[g\left(\frac{X_n}{n}\right) - c \right]$ for appropriate constant c when $p = \frac{1}{\sqrt{3}}$.

Solution: Since $X_n \stackrel{D}{=} \sum_{i=1}^n Y_i$ where $Y_i \sim BIN(1, p)$,

$$\sqrt{n} \left(\frac{X_n}{n} - p \right) \xrightarrow{D} N(0, p(1-p))$$

by the CLT. Let $\theta = p$. Then $g'(\theta) = 3\theta^2 - 1$ and $g''(\theta) = 6\theta$. Notice that

$$g(1/\sqrt{3}) = (1/\sqrt{3})^3 - 1/\sqrt{3} = (1/\sqrt{3})(\frac{1}{3} - 1) = \frac{-2}{3\sqrt{3}} = c.$$

Also $g'(1/\sqrt{3}) = 0$ and $g''(1/\sqrt{3}) = 6/\sqrt{3}$. Since $\tau^2(p) = p(1-p)$,

$$\tau^2(1/\sqrt{3}) = \frac{1}{\sqrt{3}}(1 - \frac{1}{\sqrt{3}}).$$

Hence

$$n \left[g\left(\frac{X_n}{n}\right) - \left(\frac{-2}{3\sqrt{3}}\right) \right] \xrightarrow{D} \frac{1}{2} \frac{1}{\sqrt{3}}(1 - \frac{1}{\sqrt{3}}) \frac{6}{\sqrt{3}} \; \chi^2_1 = (1 - \frac{1}{\sqrt{3}}) \; \chi^2_1.$$

There are many texts on large sample theory including, in roughly increasing order of difficulty, Lehmann (1999), Ferguson (1996), Sen and Singer (1993), and Serfling (1980). Cramér (1946) is also an important reference, and White (1984) considers asymptotic theory for econometric applications. Lecture notes are available from
(www.stat.psu.edu/∼dhunter/asymp/lectures/). Also see DasGupta (2008), Davidson (1994) and van der Vaart (1998).

In analysis, convergence in probability is a special case of convergence in measure and convergence in distribution is a special case of weak convergence. See Ash (1972, p. 322) and Sen and Singer (1993, p. 39). Almost sure convergence is also known as strong convergence. See Sen and Singer (1993, p. 34). Since $\overline{Y} \xrightarrow{P} \mu$ iff $\overline{Y} \xrightarrow{D} \mu$, the WLLN refers to weak convergence. Technically the X_n and X need to share a common probability space for convergence in probability and almost sure convergence.

Perlman (1972) and Wald (1949) give general results on the consistency of the MLE while Berk (1972), Lehmann (1980) and Schervish (1995, p.

418) discuss the asymptotic normality of the MLE in exponential families. Theorems 8.4 and 8.20 appear in Olive (2007a). Also see Cox (1984) and McCulloch (1988). A similar result to Theorem 8.20 for linear exponential families where $t_i(\boldsymbol{x}) = x_i$, are given by Brown (1986, p. 172). Portnoy (1977) gives large sample theory for unbiased estimators in exponential families. Although $\overline{T}_n$ is the UMVUE of $E(t(Y)) = \mu_t$, asymptotic efficiency of UMVUEs is not simple in general. See Pfanzagl (1993).

The multivariate delta method appears, for example, in Ferguson (1996, p. 45), Lehmann (1999, p. 315), Mardia, Kent and Bibby (1979, p. 52), Sen and Singer (1993, p. 136) or Serfling (1980, p. 122).

In analysis, the fact that

$$\boldsymbol{D}^{-1}_{\boldsymbol{g}(\boldsymbol{\theta})} = \boldsymbol{D}_{\boldsymbol{g}^{-1}(\boldsymbol{\eta})}$$

is a corollary of the inverse mapping theorem (or of the inverse function theorem). See Apostol (1957, p. 146) and Wade (2000, p. 353).

Casella and Berger (2002, p. 112, 133) give results similar to Proposition 8.3.

According to Rohatgi (1984, p. 626), if $Y_1, ..., Y_n$ are iid with pdf $f(y)$, if $Y_{r_n:n}$ is the r_nth order statistic, $r_n/n \to \rho$, $F(\xi_\rho) = \rho$ and if $f(\xi_\rho) > 0$, then

$$\sqrt{n}(Y_{r_n:n} - \xi_\rho) \xrightarrow{D} N\left(0, \frac{\rho(1-\rho)}{[f(\xi_\rho)]^2}\right).$$

So there are many asymptotically equivalent ways of defining the sample ρ quantile.

8.10 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

Refer to Chapter 10 for the pdf or pmf of the distributions in the problems below.

8.1*. a) Enter the following *R/Splus* function that is used to illustrate the central limit theorem when the data $Y_1, ..., Y_n$ are iid from an exponential distribution. The function generates a data set of size n and computes $\overline{Y}_1$ from the data set. This step is repeated $nruns = 100$ times. The output is a vector $(\overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_{100})$. A histogram of these means should resemble a symmetric normal density once n is large enough.

```
cltsim <- function(n=100, nruns=100){
ybar <- 1:nruns
for(i in 1:nruns){
  ybar[i] <- mean(rexp(n))}
list(ybar=ybar)}
```

b) The following commands will plot 4 histograms with $n = 1, 5, 25$ and 100. Save the plot in *Word*.

```
> z1 <- cltsim(n=1)
> z5 <- cltsim(n=5)
> z25 <- cltsim(n=25)
> z200 <- cltsim(n=200)
> par(mfrow=c(2,2))
> hist(z1$ybar)
> hist(z5$ybar)
> hist(z25$ybar)
> hist(z200$ybar)
```

c) Explain how your plot illustrates the central limit theorem.

d) Repeat parts a), b) and c), but in part a), change *rexp(n)* to *rnorm(n)*. Then $Y_1, ..., Y_n$ are iid N(0,1) and $\overline{Y} \sim N(0, 1/n)$.

8.2*. Let $X_1, ..., X_n$ be iid from a normal distribution with unknown mean μ and known variance σ^2. Let

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Find the limiting distribution of $\sqrt{n}(\overline{X}^3 - c)$ for an appropriate constant c.

8.3*. (Aug. 03 QUAL) Let $X_1, ..., X_n$ be a random sample from a population with pdf

$$f(x) = \begin{cases} \frac{\theta x^{\theta-1}}{3^\theta} & 0 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

The method of moments estimator for θ is $T_n = \dfrac{\overline{X}}{3 - \overline{X}}$.

a) Find the limiting distribution of $\sqrt{n}(T_n - \theta)$ as $n \to \infty$.

b) Is T_n asymptotically efficient? Why?

c) Find a consistent estimator for θ and show that it is consistent.

8.4*. From Theorems 2.17 and 2.18, if $Y_n = \sum_{i=1}^n X_i$ where the X_i are iid from a nice distribution, then Y_n also has a nice distribution. If $E(X) = \mu$ and $\text{VAR}(X) = \sigma^2$ then by the CLT

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Hence

$$\sqrt{n}(\frac{Y_n}{n} - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Find μ, σ^2 and the distribution of X_i if

i) $Y_n \sim \text{BIN}(n, \rho)$ where BIN stands for binomial.

ii) $Y_n \sim \chi^2_n$.

iii) $Y_n \sim G(n\nu, \lambda)$ where G stands for gamma.

iv) $Y_n \sim NB(n, \rho)$ where NB stands for negative binomial.

v) $Y_n \sim POIS(n\theta)$ where POIS stands for Poisson.

vi) $Y_n \sim N(n\mu, n\sigma^2)$.

8.5*. Suppose that $X_n \sim U(-1/n, 1/n)$.

a) What is the cdf $F_n(x)$ of X_n?

b) What does $F_n(x)$ converge to?

(Hint: consider $x < 0$, $x = 0$ and $x > 0$.)

c) $X_n \xrightarrow{D} X$. What is X?

8.6. Continuity Theorem problem: Let X_n be sequence of random variables with cdfs F_n and mgfs m_n. Let X be a random variable with cdf F and mgf m. Assume that all of the mgfs m_n and m are defined to $|t| \leq d$ for some $d > 0$. Then if $m_n(t) \to m(t)$ as $n \to \infty$ for all $|t| < c$ where $0 < c < d$, then $X_n \xrightarrow{D} X$.

Let

$$m_n(t) = \frac{1}{[1 - (\lambda + \frac{1}{n})t]}$$

for $t < 1/(\lambda + 1/n)$. Then what is $m(t)$ and what is X?

8.7. Let $Y_1, ..., Y_n$ be iid, $T_{1,n} = \overline{Y}$ and let $T_{2,n} = \text{MED}(n)$ be the sample median. Let $\theta = \mu$.

Then

$$\sqrt{n}(MED(n) - MED(Y)) \xrightarrow{D} N\left(0, \frac{1}{4f^2(MED(Y))}\right)$$

where the population median is MED(Y) (and MED(Y) $= \mu = \theta$ for a) and b) below).

a) Find $ARE(T_{1,n}, T_{2,n})$ if F is the cdf of the normal $N(\mu, \sigma^2)$ distribution.

b) Find $ARE(T_{1,n}, T_{2,n})$ if F is the cdf of the double exponential $DE(\theta, \lambda)$ distribution.

8.8. (Sept. 2005 Qual) Let $X_1, ..., X_n$ be independent identically distributed random variables with probability density function

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

a) Find the MLE of $\dfrac{1}{\theta}$. Is it unbiased? Does it achieve the information inequality lower bound?

b) Find the asymptotic distribution of the MLE of $\dfrac{1}{\theta}$.

c) Show that $\overline{X}_n$ is unbiased for $\dfrac{\theta}{\theta+1}$. Does $\overline{X}_n$ achieve the information inequality lower bound?

d) Find an estimator of $\dfrac{1}{\theta}$ from part (c) above using $\overline{X}_n$ which is different from the MLE in (a). Find the asymptotic distribution of your estimator using the delta method.

e) Find the asymptotic relative efficiency of your estimator in (d) with respect to the MLE in (b).

8.9. Many multiple linear regression estimators $\hat{\boldsymbol{\beta}}$ satisfy

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(0, V(\hat{\boldsymbol{\beta}}, F) \; \boldsymbol{W}) \tag{8.14}$$

when

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \xrightarrow{P} \boldsymbol{W}^{-1}, \tag{8.15}$$

and when the errors e_i are iid with a cdf F and a unimodal pdf f that is symmetric with a unique maximum at 0. When the variance $V(e_i)$ exists,

$$V(OLS, F) = V(e_i) = \sigma^2 \text{ while } \mathrm{V(L_1, F)} = \frac{1}{\mathrm{4[f(0)]^2}}.$$

In the multiple linear regression model,

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i \tag{8.16}$$

for $i = 1, \ldots, n$. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{8.17}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors.

a) What is the ijth element of the matrix

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n}?$$

b) Suppose $x_{k,1} = 1$ and that $x_{k,j} \sim X_j$ are iid with $E(X_j) = 0$ and $V(X_j) = 1$ for $k = 1, ..., n$ and $j = 2, ..., p$. Assume that X_i and X_j are independent for $i \neq j$, $i > 1$ and $j > 1$. (Often $x_{k,j} \sim N(0, 1)$ in simulations.) Then what is $\boldsymbol{W}^{-1}$ in (8.16)?

c) Suppose $p = 2$ and $Y_i = \alpha + \beta X_i + e_i$. Show

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2} & \frac{-\sum X_i}{n\sum(X_i - \overline{X})^2} \\ \\ \frac{-\sum X_i}{n\sum(X_i - \overline{X})^2} & \frac{n}{n\sum(X_i - \overline{X})^2} \end{bmatrix}.$$

d) Under the conditions of c), let $S_x^2 = \sum(X_i - \overline{X})^2/n$. Show that

$$n(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \left(\frac{\boldsymbol{X}^T\boldsymbol{X}}{n}\right)^{-1} = \begin{bmatrix} \frac{\frac{1}{n}\sum X_i^2}{S_x^2} & \frac{-\overline{X}}{S_x^2} \\ \\ \frac{-\overline{X}}{S_x^2} & \frac{1}{S_x^2} \end{bmatrix}.$$

e) If the X_i are iid with variance $V(X)$ then $n(\boldsymbol{X}^T\boldsymbol{X})^{-1} \xrightarrow{P} \boldsymbol{W}$. What is $\boldsymbol{W}$?

f) Now suppose that n is divisible by 5 and the $n/5$ of X_i are at 0.1, $n/5$ at 0.3, $n/5$ at 0.5, $n/5$ at 0.7 and $n/5$ at 0.9. (Hence if $n = 100$, 20 of the X_i are at 0.1, 0.3, 0.5, 0.7 and 0.9.)

Find $\sum X_i^2/n$, $\overline{X}$ and S_x^2. (Your answers should not depend on n.)

g) Under the conditions of f), estimate $V(\hat{\alpha})$ and $V(\hat{\beta})$ if L_1 is used and if the e_i are iid $N(0, 0.01)$.

Hint: Estimate $\boldsymbol{W}$ with $n(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ and $V(\hat{\boldsymbol{\beta}}, F) = V(L_1, F) = \frac{1}{4[f(0)]^2}$. Hence

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \approx N_2 \left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \frac{1}{n} \frac{1}{4[f(0)]^2} \begin{pmatrix} \frac{\frac{1}{n}\sum X_i^2}{S_x^2} & \frac{-\overline{X}}{S_x^2} \\ \frac{-\overline{X}}{S_x^2} & \frac{1}{S_x^2} \end{pmatrix} \right].$$

You should get an answer like $0.0648/n$.

Problems from old quizzes and exams.

8.10. Let $X_1, ..., X_n$ be iid Bernoulli(p) random variables.

a) Find $I_1(p)$.

b) Find the FCRLB for estimating p.

c) Find the limiting distribution of $\sqrt{n}(\ \overline{X}_n - p\ )$.

d) Find the limiting distribution of $\sqrt{n}\ [\ (\overline{X}_n)^2 - c\ ]$ for an appropriate constant c.

8.11. Let $X_1, ..., X_n$ be iid Exponential(β) random variables.

a) Find the FCRLB for estimating β.

b) Find the limiting distribution of $\sqrt{n}(\ \overline{X}_n - \beta\ )$.

c) Find the limiting distribution of $\sqrt{n}\ [\ (\overline{X}_n)^2 - c\ ]$ for an appropriate constant c.

8.12. Let $Y_1, ..., Y_n$ be iid Poisson (λ) random variables.

a) Find the limiting distribution of $\sqrt{n}(\ \overline{Y}_n - \lambda\ )$.

b) Find the limiting distribution of $\sqrt{n}\ [\ (\overline{Y}_n)^2 - c\ ]$ for an appropriate constant c.

8.13. Let $Y_n \sim \chi^2_n$.

a) Find the limiting distribution of $\sqrt{n}\ \left(\frac{Y_n}{n} - 1 \right)$.

b) Find the limiting distribution of $\sqrt{n}\ \left[\left(\frac{Y_n}{n}\right)^3 - 1\right]$.

8.14. Let $X_1, ..., X_n$ be iid with cdf $F(x) = P(X \le x)$. Let $Y_i = I(X_i \le x)$ where the indicator equals 1 if $X_i \le x$ and 0, otherwise.

a) Find $E(Y_i)$.

b) Find VAR(Y_i).

c) Let $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x)$ for some fixed real number x. Find the limiting distribution of $\sqrt{n}\ \left(\hat{F}_n(x) - c_x\right)$ for an appropriate constant c_x.

8.15. Suppose X_n has cdf

$$F_n(x) = 1 - \left(1 - \frac{x}{\theta n}\right)^n$$

for $x \ge 0$ and $F_n(x) = 0$ for $x < 0$. Show that $X_n \xrightarrow{D} X$ by finding the cdf of X.

8.16. Let X_n be a sequence of random variables such that $P(X_n = 1/n) = 1$. Does X_n converge in distribution? If yes, prove it by finding X and the cdf of X. If no, prove it.

8.17. Suppose that $Y_1, ..., Y_n$ are iid with $E(Y) = (1-\rho)/\rho$ and VAR$(Y) = (1-\rho)/\rho^2$ where $0 < \rho < 1$.

a) Find the limiting distribution of

$$\sqrt{n}\ \left(\overline{Y}_n - \frac{1-\rho}{\rho}\right).$$

b) Find the limiting distribution of $\sqrt{n}\ \left[g(\overline{Y}_n) - \rho\right]$ for appropriate function g.

8.18. Let $X_n \sim$ Binomial(n, p) where the positive integer n is large and $0 < p < 1$.

a) Find the limiting distribution of $\sqrt{n}\ \left(\frac{X_n}{n} - p\right)$.

b) Find the limiting distribution of $\sqrt{n}\ \left[\left(\frac{X_n}{n}\right)^2 - p^2\right]$.

8.19. Let $Y_1, ..., Y_n$ be iid exponential (λ) so that $E(Y) = \lambda$ and $\text{MED}(Y) = \log(2)\lambda$.

a) Let $T_{1,n} = \log(2)\overline{Y}$ and find the limiting distribution of $\sqrt{n}(T_{1,n} - \log(2)\lambda)$.

b) Let $T_{2,n} = \text{MED}(n)$ be the sample median and find the limiting distribution of $\sqrt{n}(T_{2,n} - \log(2)\lambda)$.

c) Find $ARE(T_{1,n}, T_{2,n})$.

8.20. Suppose that $\eta = g(\theta)$, $\theta = g^{-1}(\eta)$ and $g'(\theta) > 0$ exists. If X has pdf or pmf $f(x|\theta)$, then in terms of η, the pdf or pmf is $f^*(x|\eta) = f(x|g^{-1}(\eta))$. Now

$$A = \frac{\partial}{\partial \eta} \log[f(x|g^{-1}(\eta))] = \frac{1}{f(x|g^{-1}(\eta))} \frac{\partial}{\partial \eta} f(x|g^{-1}(\eta)) =$$

$$\left[\frac{1}{f(x|g^{-1}(\eta))}\right] \left[\frac{\partial}{\partial \theta} f(x|\theta)\bigg|_{\theta = g^{-1}(\eta)}\right] \left[\frac{\partial}{\partial \eta} g^{-1}(\eta)\right]$$

using the chain rule twice. Since $\theta = g^{-1}(\eta)$,

$$A = \left[\frac{1}{f(x|\theta)}\right] \left[\frac{\partial}{\partial \theta} f(x|\theta)\right] \left[\frac{\partial}{\partial \eta} g^{-1}(\eta)\right].$$

Hence

$$A = \frac{\partial}{\partial \eta} \log[f(x|g^{-1}(\eta))] = \left[\frac{\partial}{\partial \theta} \log[f(x|\theta)]\right] \left[\frac{\partial}{\partial \eta} g^{-1}(\eta)\right].$$

Now show that

$$I_1^*(\eta) = \frac{I_1(\theta)}{[g'(\theta)]^2}.$$

8.21. Let $Y_1, ..., Y_n$ be iid exponential (1) so that $P(Y \leq y) = F(y) = 1 - e^{-y}$ for $y \geq 0$. Let $Y_{(n)} = \max(Y_1, ..., Y_n)$.

a) Show that $F_{Y_{(n)}}(t) = P(Y_{(n)} \leq t) = [1 - e^{-t}]^n$ for $t \geq 0$.

b) Show that $P(Y_{(n)} - \log(n) \leq t) \to \exp(-e^{-t})$ (for all $t \in (-\infty, \infty)$ since $t + \log(n) > 0$ implies $t \in \Re$ as $n \to \infty$).

8.22. Let $Y_1, ..., Y_n$ be iid uniform $(0, 2\theta)$.

a) Let $T_{1,n} = \overline{Y}$ and find the limiting distribution of $\sqrt{n}(T_{1,n} - \theta)$.

b) Let $T_{2,n} = \text{MED}(n)$ be the sample median and find the limiting distribution of $\sqrt{n}(T_{2,n} - \theta)$.

c) Find $ARE(T_{1,n}, T_{2,n})$. Which estimator is better, asymptotically?

8.23. Suppose that $Y_1, ..., Y_n$ are iid from a distribution with pdf $f(y|\theta)$ and that the integral and differentiation operators of all orders can be interchanged (eg the data is from a one parameter exponential family).

a) Show that $0 = E\left[\frac{\partial}{\partial\theta}\log(f(Y|\theta))\right]$ by showing that

$$\frac{\partial}{\partial\theta}1 = 0 = \frac{\partial}{\partial\theta}\int f(y|\theta)dy = \int\left[\frac{\partial}{\partial\theta}\log(f(y|\theta))\right]f(y|\theta)dy. \quad (*)$$

b) Take 2nd derivatives of (*) to show that

$$I_1(\theta) = E_\theta[(\frac{\partial}{\partial\theta}\log f(Y|\theta))^2] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log(f(Y|\theta))\right].$$

8.24. Suppose that $X_1, ..., X_n$ are iid $N(\mu, \sigma^2)$.

a) Find the limiting distribution of $\sqrt{n}\ \left(\overline{X}_n - \ \mu\ \right)$.

b) Let $g(\theta) = [\log(1+\theta)]^2$. Find the limiting distribution of $\sqrt{n}\ \left(g(\overline{X}_n) - \ g(\mu)\ \right)$ for $\mu > 0$.

c) Let $g(\theta) = [\log(1+\theta)]^2$. Find the limiting distribution of $n\ \left(g(\overline{X}_n) - \ g(\mu)\ \right)$ for $\mu = 0$. Hint: Use Theorem 8.29.

8.25. Let $W_n = X_n - X$ and let $r > 0$. Notice that for any $\epsilon > 0$,

$$E|X_n - X|^r \geq E[|X_n - X|^r\ I(|X_n - X| \geq \epsilon)] \geq \epsilon^r P(|X_n - X| \geq \epsilon).$$

Show that $W_n \xrightarrow{P} 0$ if $E|X_n - X|^r \to 0$ as $n \to \infty$.

8.26. Let $X_1, ..., X_n$ be iid with $E(X) = \mu$ and $V(X) = \sigma^2$. What is the limiting distribution of $n[(\overline{X})^2 - \mu^2]$ for the value or values of μ where the delta method does not apply? Hint: use Theorem 8.29.

8.27. (Sept. 05 QUAL) Let $X \sim$ Binomial(n, p) where the positive integer n is large and $0 < p < 1$.

a) Find the limiting distribution of $\sqrt{n} \left(\frac{X}{n} - p \right)$.

b) Find the limiting distribution of $\sqrt{n} \left[\left(\frac{X}{n} \right)^2 - p^2 \right]$.

c) Show how to find the limiting distribution of $\left[\left(\frac{X}{n} \right)^3 - \frac{X}{n} \right]$ when $p = \frac{1}{\sqrt{3}}$.

(Actually want the limiting distribution of

$$n \left(\left[\left(\frac{X}{n} \right)^3 - \frac{X}{n} \right] - g(p) \right)$$

where $g(\theta) = \theta^3 - \theta$.)

8.28. (Aug. 04 QUAL) Let $X_1, ..., X_n$ be independent and identically distributed (iid) from a Poisson(λ) distribution.

a) Find the limiting distribution of $\sqrt{n}$ ($\overline{X} - \lambda$).

b) Find the limiting distribution of $\sqrt{n}$ [$(\overline{X})^3 - (\lambda)^3$].

8.29. (Jan. 04 QUAL) Let $X_1, ..., X_n$ be iid from a normal distribution with unknown mean μ and known variance σ^2. Let $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

a) Show that $\overline{X}$ and S^2 are independent.

b) Find the limiting distribution of $\sqrt{n}(\overline{X}^3 - c)$ for an appropriate constant c.

8.30. Suppose that $Y_1, ..., Y_n$ are iid logistic$(\theta, 1)$ with pdf

$$f(y) = \frac{\exp(-(y-\theta))}{[1+\exp(-(y-\theta))]^2}$$

where and y and θ are real.

a) $I_1(\theta) = 1/3$ and the family is regular so the "standard limit theorem" for the MLE $\hat{\theta}_n$ holds. Using this standard theorem, what is the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$?

b) Find the limiting distribution of $\sqrt{n}(\overline{Y}_n - \theta)$.

c) Find the limiting distribution of $\sqrt{n}(MED(n) - \theta)$.

d) Consider the estimators $\hat{\theta}_n$, $\overline{Y}_n$ and MED(n). Which is the best estimator and which is the worst?

8.31. Let $Y_n \sim$ binomial(n, p). Find the limiting distribution of

$$\sqrt{n}\left(\arcsin\left(\sqrt{\frac{\mathrm{Y_n}}{\mathrm{n}}}\right) - \arcsin(\sqrt{\mathrm{p}})\right).$$

(Hint:

$$\frac{d}{dx}\arcsin(\mathrm{x}) = \frac{1}{\sqrt{1-\mathrm{x}^2}}.)$$

8.32. Suppose $Y_n \sim$ uniform$(-n, n)$. Let $F_n(y)$ be the cdf of Y_n.

a) Find $F(y)$ such that $F_n(y) \to F(y)$ for all y as $n \to \infty$.

b) Does $Y_n \xrightarrow{L} Y$? Explain briefly.