Chapter 2
Univariate Limit Theorems

This chapter discusses the central limit theorem, the delta method, asymptotically efficient estimators, convergence in distribution and convergence in probability. This chapter follows Olive (2014, ∮ 8.1-8.5) closely.

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.

2.1 The CLT and Delta Method

The CLT is also known as the Lindeberg-Lévy CLT.

Theorem 2.1: the Central Limit Theorem (CLT). Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Let the sample mean $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Then

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

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 2.4.

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 1.15. The notation $Y_n \xrightarrow{D} X$ 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, and the limiting distribution does not depend on n. For the CLT, notice that

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

is the z–score of $\overline{Y}$ and the z–score of $\sum_{i=1}^n Y_i$. Then $Z_n \xrightarrow{D} N(0,1)$. If $Z_n \xrightarrow{D} N(0,1)$, 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 the cdf of $\overline{Y}_n$ as if $\overline{Y}_n \sim N(\mu, \sigma^2/n)$. Note that the approximate distribution, unlike the limiting distribution, does depend on n. The standard error $S/\sqrt{n}$ approximates the asymptotic standard deviation $\sqrt{\sigma^2/n}$ of $\overline{Y}$.

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 $V(X) = \sigma_X^2$.

Several of the random variables in Theorems 1.24 and 1.25 can be approximated in this way. The CLT says that $\overline{Y}_n \sim AN(\mu, \sigma^2/n)$. The delta method says that if $T_n \sim AN(\theta, \sigma^2/n)$, and if $g'(\theta) \neq 0$, then $g(T_n) \sim AN(g(\theta), \sigma^2[g'(\theta)]^2/n)$. Hence a smooth function $g(T_n)$ of a well behaved statistic T_n tends to be well behaved (asymptotically normal with a $\sqrt{n}$ convergence rate).

Example 2.1. a) Let $Y_1, ..., Y_n$ be iid Ber(ρ). Then $E(Y) = \rho$ and $V(Y) = \rho(1-\rho)$. Hence

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

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

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

since

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

by a).

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

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

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 2.2: 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 2.2. Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. 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 2.3. Let $X \sim$ 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 2.1b 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,

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

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

Example 2.4. Let $X_n \sim$ 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 $E(Y) = \lambda = V(Y)$. Thus by the CLT,

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

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

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

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

Example 2.5. 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 $\sqrt{n}\ \left(\overline{Y} -\ \alpha\beta \right) \stackrel{D}{\to} N(0, \alpha\beta^2)$.

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, $\sqrt{n}\ \left((\overline{Y})^2 -\ c \right) \stackrel{D}{\to} N(0, \sigma^2[g'(\mu)]^2) = N(0, 4\alpha^3\beta^4)$ where $c = \mu^2 = \alpha^2\beta^2$.

Remark 2.1. a) Note that if $\sqrt{n}(T_n - k) \stackrel{D}{\to} N(0, \sigma^2)$, then evaluate the derivative at k. Thus use $g'(k)$ where $k = \alpha\beta$ in the above example. A common error occurs when k is a simple function of θ, for example $k = \theta/2$ with $g(\mu) = \mu^2$. Thus $g'(\mu) = 2\mu$ so $g'(\theta/2) = 2\theta/2 = \theta$. Then the common delta method error is to plug in $g'(\theta) = 2\theta$ instead of $g'(k) = \theta$. See Problems 2.3, 2.33, 2.35, 2.36, and 2.37.

b) For the delta method, also note that the function g can not depend on n since then there would be a sequence of functions g_n rather than one function g. This fact also applies to several other theorems in this chapter.

The following extension of the delta method is sometimes useful.

Theorem 2.3: the Second Order Delta Method. Suppose that $g'(\theta) = 0$, $g''(\theta) \neq 0$ and

$$\sqrt{n}(T_n - \theta) \stackrel{D}{\to} N(0, \tau^2(\theta)).$$

Then

$$n[g(T_n) - g(\theta)] \stackrel{D}{\to} \frac{1}{2}\tau^2(\theta)g''(\theta)\chi^2_1.$$

Example 2.6. 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

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

Barndorff–Nielsen (1982), Casella and Berger (2002, p. 472, 515), Cox and Hinckley (1974, p. 286), Lehmann and Casella (2003, 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

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right) = N[0, FCRLB_1(\tau(\theta))] \tag{2.1}$$

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

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

Hence $\tau(\hat{\theta}) \sim AN[\tau(\theta), FCRLB_n(\tau(\theta))]$ where $FCRLB_n(\tau(\theta)) = FCRLB_1(\tau(\theta))/n$. Notice that if

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

then (2.1) 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 1.43.

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 $\mu_t \equiv \mu_T = E_\theta(\overline{T}_n) = E_\theta[t(Y)]$. Let $\sigma_t^2 = V_\theta[t(Y)]$. These values can be found by using the distribution of $t(Y)$.

Theorem 2.4. 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)

$$\mu_t = E[t(Y)] = \frac{-c'(\theta)}{c(\theta)w'(\theta)} = \frac{-\partial}{\partial \eta}\log(b(\eta)), \tag{2.2}$$

and b)

$$\sigma_t^2 = V[t(Y)] = \frac{\frac{-\partial^2}{\partial \theta^2}\log(c(\theta)) - [w''(\theta)]\mu_t}{[w'(\theta)]^2} = \frac{-\partial^2}{\partial \eta^2}\log(b(\eta)). \tag{2.3}$$

Proof. The proof will be for pdfs. For pmfs replace the integrals by sums. By Theorem 1.31, only the middle equalities need to be shown. By Remark 1.9 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

$$E \left(w'(\theta)t(Y) + \frac{\partial}{\partial \theta} \log(c(\theta)) \right)^2 = -E \left(w''(\theta)t(Y) + \frac{\partial^2}{\partial \theta^2} \log(c(\theta)) \right). \quad (2.4)$$

Using a) shows that the left hand side of (2.4) equals

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

while the right hand side of (2.4) equals

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

and the result follows. □

The simplicity of the following Olive (2014, p. 221) result is rather surprising. When (as is usually the case) $\frac{1}{n} \sum_{i=1}^n t(Y_i)$ is the MLE of μ_t, $\hat{\eta} = g^{-1}(\frac{1}{n} \sum_{i=1}^n t(Y_i))$ is the MLE of η by the invariance principle.

Theorem 2.5. 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 $V(t(Y)) = \sigma_t^2$.

a) Then

$$\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} t(Y_i) - \mu_t] \xrightarrow{D} N(0, I_1(\eta))$$

where

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

b) If $\eta = g^{-1}(\mu_t)$, $\hat{\eta} = g^{-1}(\frac{1}{n}\sum_{i=1}^{n} t(Y_i))$, 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 $V(t(Y)) = \sigma_t^2 = I_1(\eta) = g'(\eta)$. 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 Theorem 2.4 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 Theorem 2.4 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 1.33c, 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. □

Remark 2.2. Following DasGupta (2008, p. 241-242), let $\psi(\eta) = -\log(b(\eta))$. Then $E_\eta[t(Y_1)] = \mu_t = \psi'(\eta) = g(\eta)$ by Theorem 2.4a, and the MLE $\hat{\eta}$ is the solution of $\frac{1}{n}\sum_{i=1}^n t(y_i) \overset{set}{=} E_\eta[t(Y_1)] = g(\eta)$ if the MLE exists. Now $g(\eta) = E_\eta[t(Y_1)]$ is an increasing function of η since $g'(\eta) = \psi''(\eta) = V_\eta(t(Y)) > 0$ (1P–REFs do not contain degenerate distributions). So for large n, with probability tending to one, the MLE $\hat{\eta}$ exists and $\hat{\eta} = g^{-1}(\frac{1}{n}\sum_{i=1}^n t(Y_i))$. Since $g'(\eta)$ exists, $g(\eta)$ and $g^{-1}(\eta)$ are continuous and the delta method can be applied to $\hat{\eta}$ as in Theorem 2.5b. By the proof of Theorem 2.5a), $\psi''(\eta) = I_1(\eta)$. Notice that if $\hat{\eta}$ is the MLE of η, then $\frac{1}{n}\sum_{i=1}^n t(Y_i)$ is the MLE of $\mu_t = E[t(Y_1)]$ by invariance. Hence if n is large enough, Theorem 2.5ab is for the MLE of $E[t(Y_1)]$ and the MLE of η.

2.2 Asymptotically Efficient Estimators

Definition 2.1. 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 $S_A/\sqrt{n}$. If $Y_1, ..., Y_n$ are iid with cdf F, then $\sigma_A^2 \equiv \sigma_A^2(F)$ depends on F.

Remark 2.3. 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 $nV(T_n)$ converges in distribution to σ_A^2, but not always. See Staudte and Sheather (1990, p. 51) and Lehmann (1999, p. 232).

Example 2.7. 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 $V(\overline{Y}_n) = \sigma^2/n = AV(\overline{Y}_n)$ and that the standard error $SE(\overline{Y}_n) = S_n/\sqrt{n}$ where S_n^2 is the sample variance. Note that $\sigma_A^2(F) = \sigma^2$. If F is a $N(\mu, 1)$ cdf then $\sigma_A^2(F) = 1$, but if F is the $G(\nu = 7, \lambda = 1)$ cdf then $\sigma_A^2(F) = 7$.

Definition 2.2. 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

$$P(Y \leq \text{MED}(Y)) \geq 0.5 \text{ and } P(Y \geq \text{MED}(Y)) \geq 0.5. \tag{2.5}$$

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 2.6. 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 2.8. 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 2.6,

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

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 1.43 that $I_1(\theta)$ is the information number for θ based on a sample of size 1. Also recall that $I_1(\tau(\theta)) = I_1(\theta)/[\tau'(\theta)]^2 = 1/FCRLB_1[\tau(\theta)]$. See Definition 1.44.

The following definition says that if T_n is an asymptotically efficient estimator of $\tau(\theta)$, then

$$T_n \sim AN[\tau(\theta), FCRLB_n(\tau(\theta))].$$

Definition 2.3. Assume $\tau'(\theta) \neq 0$. Then an estimator T_n of $\tau(\theta)$ is **asymptotically efficient** if

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right) \sim N(0, FCRLB_1[\tau(\theta)]). \tag{2.6}$$

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

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right) \sim N(0, FCRLB_1[\theta]). \tag{2.7}$$

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 2.5. $\overline{T}_n$ is the UMVUE of $E(t(Y))$ by the LSU theorem.

The following theorem suggests that MLEs and UMVUEs are often asymptotically efficient. The theorem often holds for location families where the support does not depend on θ. The theorem does not hold for the uniform $(0, \theta)$ family. For the MLE $\hat{\theta}$. Geisser (2006, pp. 133-134) shows that if i) the Y_i are iid with pdf $f(y|\theta)$ and likelihood function $L(\theta) = \prod_{i=1}^n f(y_i|\theta)$, ii) $E_\theta\left[\left(\frac{d\ log(L(\theta))}{d\theta}\right)\right] = 0$, and iii) $E_\theta\left[\left(\frac{d\ log(L(\theta))}{d\theta}\right)^2\right] = -E_\theta\left[\frac{d^2\ log(L(\theta))}{d\theta^2}\right]$ exists and is nonzero for all θ in a neighborhood of the true value θ_0, then

$$\sqrt{n}[\hat{\theta}_n - \theta_0] \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta_0)}\right).$$

Conditions ii) and iii) hold for a 1P-REF by Equations (1.44) and (1.47). See Berk (1972) and Wald (1949) for different regularity conditions.

Theorem 2.7: a "Standard Limit Theorem": Let $\hat{\theta}_n$ be the MLE or UMVUE of θ. If $\tau'(\theta) \neq 0$, then under strong regularity conditions,

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

2.3 Modes of Convergence and Consistency

Definition 2.4. 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.

Convergence in distribution is also known as weak convergence or X_n converges weakly to X. An important fact is that **the limiting distribution does not depend on the sample size** n. Notice that the CLT, delta method and Theorem 2.5 give the limiting distributions of $Z_n = \sqrt{n}(\overline{Y}_n - \mu)$, $Z_n = \sqrt{n}(g(T_n) - g(\theta))$ and $Z_n = \sqrt{n}[\frac{1}{n}\sum_{i=1}^n t(Y_i) - E(t(Y))]$, 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 $P(a < X_n \leq b) = F_n(b) - F_n(a) \to F(b) - F(a) = P(a < X \leq b)$ if F is continuous at a and b. Convergence in distribution is useful for constructing large sample confidence intervals and tests of hypotheses. See Chapter 4.

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 $n \to \infty$ 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 $X_n \equiv X_n(\omega)$ converge to the random variable $X \equiv X(\omega)$ for all ω.

Example 2.9. Suppose that $X_n \sim U(-1/n, 1/n)$. Then the cdf $F_n(x)$ of X_n is

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

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 $n \ge 1$. Examining the cases $x < 0$, $x = 0$ and $x > 0$ shows that as $n \to \infty$,

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

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

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

Since $x = 0$ is the only discontinuity point of $F_X(x)$ and since $F_n(x) \to F_X(x)$ for all continuity points of $F_X(x)$ (i.e. for $x \ne 0$),

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

Example 2.10. Suppose $Y_n \sim U(0, n)$. Then $F_n(t) = t/n$ for $0 < t \le n$ and $F_n(t) = 0$ for $t \le 0$. Hence $\lim_{n\to\infty} F_n(t) = 0$ for $t \le 0$. If $t > 0$ and $n > t$, then $F_n(t) = t/n \to 0$ as $n \to \infty$. Thus $\lim_{n\to\infty} F_n(t) = H(t) = 0$ for all t, and Y_n does not converge in distribution to any random variable Y since $H(t) \equiv 0$ is a continuous function but not a cdf.

Definition 2.5. 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* $\tau(\theta)$ or to be a *point mass at* $\tau(\theta)$.

Definition 2.6. a) 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 $\epsilon > 0$,

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

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

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

if for every $\epsilon > 0$,

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

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

Definition 2.7. 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 $\theta \in \Theta$. The probability $P \equiv P_\theta$ is the "true" probability distribution or underlying probability that depends on θ.

Definition 2.8. For a real number $r > 0$, Y_n **converges in** r**th 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

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

if $E[(Y_n - Y)^2] \to 0$ as $n \to \infty$. We say that X_n *converges in* r*th mean* to $\tau(\theta)$, written

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

if $E(|Y_n - \tau(\theta)|^r) \to 0$ as $n \to \infty$.

Convergence in quadratic mean is also known as convergence in mean square and as mean square convergence. From Definition 1.41, the mean square error $MSE_{\tau(\theta)}(X_n) = E_\theta[(X_n - \tau(\theta))^2]$. The notations $Y_n \xrightarrow{r} Y$, $Y_n \xrightarrow{L_r} Y$, and $Y_n \xrightarrow{L^r} Y$ are used in the literature, especially for $r \geq 1$.

Theorem 2.8: Generalized Chebyshev's Inequality or Generalized Markov's Inequality: Let $u : \mathbb{R} \to [0, \infty)$ 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 $\sigma^2 = V(Y)$ exists, then we obtain
Chebyshev's Inequality:

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

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

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

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

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

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

The following theorem 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 $\theta \in \Theta$.

Theorem 2.9. 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

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

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

Proof. a) Using Theorem 2.8 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 1.41,

$$MSE_{\tau(\theta)}(T_n) = V_\theta(T_n) + [\mathrm{Bias}_{\tau(\theta)}(\mathrm{T_n})]^2$$

where $\mathrm{Bias}_{\tau(\theta)}(\mathrm{T_n}) = \mathrm{E}_\theta(\mathrm{T_n}) - \tau(\theta)$. Since $MSE_{\tau(\theta)}(T_n) \to 0$ if both $V_\theta(T_n) \to 0$ and $\mathrm{Bias}_{\tau(\theta)}(\mathrm{T_n}) = \mathrm{E}_\theta(\mathrm{T_n}) - \tau(\theta) \to 0$, the result follows from a). □

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.

Theorem 2.10. a) Let X_θ be a random variable with a distribution depending on θ, and $0 < \delta \leq 1$. If

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

for all $\theta \in \Theta$, 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 2.9. a) A sequence of random variables X_n *converges with probability 1* (or *almost surely*, or *almost everywhere*) to X if

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

This type of convergence will be denoted by

$$X_n \xrightarrow{wp1} X.$$

b)

$$X_n \xrightarrow{wp1} \tau(\theta),$$

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

The convergence in Definition 2.9 is also known as *strong convergence*. Notation such as "X_n converges to X wp1" will also be used. Sometimes "wp1" will be replaced with "as" or "ae." The notations $X_n \xrightarrow{ae} X$, $X_n \xrightarrow{as} X$, and $X_n \xrightarrow{wp1} X$ are often used.

Theorem 2.11. Let Y_n be a sequence of iid random variables with $E(Y_i) = \mu$. Then

a) **Strong Law of Large Numbers (SLLN):** $\overline{Y}_n \stackrel{wp1}{\to} \mu$, and

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

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 $n \to \infty$. $\square$

Remark 2.4. a) For i) $X_n \stackrel{P}{\to} X$, ii) $X_n \stackrel{r}{\to} X$, or iii) $X_n \stackrel{wp1}{\to} X$, the X_n and X need to be defined on the same probability space.

b) For $X_n \stackrel{D}{\to} X$, the probability spaces can differ.

c) For i) $X_n \stackrel{P}{\to} c$, ii) $X_n \stackrel{wp1}{\to} c$, iii) $X_n \stackrel{D}{\to} c$, and iv) $X_n \stackrel{r}{\to} c$, the probability spaces of the X_n can differ.

Theorem 2.12: a) $T_n \stackrel{P}{\to} \tau(\theta)$ iff $\mathrm{T_n} \stackrel{\mathrm{D}}{\to} \tau(\theta)$.

b) If $T_n \stackrel{P}{\to} \theta$ and τ is continuous at θ, then $\tau(T_n) \stackrel{P}{\to} \tau(\theta)$. 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.

Theorem 2.13: Suppose X_n and X are RVs with the same probability space for b) and c). Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function.

a) If $X_n \stackrel{D}{\to} X$, then $g(X_n) \stackrel{D}{\to} g(X)$.

b) If $X_n \stackrel{P}{\to} X$, then $g(X_n) \stackrel{P}{\to} g(X)$.

c) If $X_n \stackrel{ae}{\to} X$, then $g(X_n) \stackrel{wp1}{\to} g(X)$.

Theorem 2.14: Suppose X_n and X are RVs with the same probability space.

a) If $X_n \stackrel{wp1}{\to} X$, then $X_n \stackrel{P}{\to} X$ and $X_n \stackrel{D}{\to} X$.

b) If $X_n \stackrel{P}{\to} X$, then $X_n \stackrel{D}{\to} X$.

c) If $X_n \stackrel{r}{\to} X$, then $X_n \stackrel{P}{\to} X$ and $X_n \stackrel{D}{\to} X$.

d) $X_n \stackrel{P}{\to} \tau(\theta)$ **iff** $X_n \stackrel{D}{\to} \tau(\theta)$ where c is a constant.

Theorem 2.15: a) If $E[(X_n - X)^2] \to 0$ as $n \to \infty$, then $X_n \stackrel{P}{\to} X$.

b) If $E(X_n) \to E(X)$ and $V(X_n - X) \to 0$ as $n \to \infty$, then $X_n \stackrel{P}{\to} X$.

Note: Part a) follows from Theorem 2.14 c) with $r = 2$. See Theorem 2.9 if $P(X = \tau(\theta)) = 1$.

Theorem 2.16: If $X_n \stackrel{r}{\to} X$, then $X_n \stackrel{k}{\to} X$ where $0 < k < r$.

Theorem 2.17: Let X_n have pdf $f_{X_n}(x)$, and let X have pdf $f_X(x)$. If $f_{X_n}(x) \to f_X(x)$ for all x (or for x outside of a set of Lebesgue measure 0), then $X_n \stackrel{D}{\to} X$.

Theorem 2.18: Let $g : \mathbb{R} \to \mathbb{R}$ be continuous at constant c.
a) If $X_n \stackrel{D}{\to} c$, then $g(X_n) \stackrel{D}{\to} c$.
b) If $X_n \stackrel{P}{\to} c$, then $g(X_n) \stackrel{P}{\to} c$.
c) If $X_n \stackrel{wp1}{\to} c$, then $g(X_n) \stackrel{wp1}{\to} c$.

Theorem 2.19: Suppose X_n and X are integer valued RVs with pmfs $f_{X_n}(x)$ and $f_X(x)$. Then $X_n \stackrel{D}{\to} X$ iff $P(X_n = k) \to P(X = k)$ for every integer k iff $f_{X_n}(x) \to f_X(x)$ for every real x.

2.4 Slutsky's Theorem and Related Results

Theorem 2.20: Slutsky's Theorem. Suppose $Y_n \stackrel{D}{\to} Y$ and $W_n \stackrel{P}{\to} w$ for some constant w. Then
a) $Y_n + W_n \stackrel{D}{\to} Y + w$,
b) $Y_n W_n \stackrel{D}{\to} wY$, and
c) $Y_n/W_n \stackrel{D}{\to} Y/w$ if $w \neq 0$.

Remark 2.5. If $Y_n \stackrel{D}{\to} Y$, $a_n \stackrel{P}{\to} a$, and $b_n \stackrel{P}{\to} b$, then $a_n + b_n Y_n \stackrel{D}{\to} a + bY$.

Theorem 2.21. a) If $X_n \stackrel{P}{\to} X$ then $X_n \stackrel{D}{\to} X$.
b) If $X_n \stackrel{wp1}{\to} X$ then $X_n \stackrel{P}{\to} X$ and $X_n \stackrel{D}{\to} X$.
c) If $X_n \stackrel{r}{\to} X$ then $X_n \stackrel{P}{\to} X$ and $X_n \stackrel{D}{\to} X$.
d) $X_n \stackrel{P}{\to} \tau(\theta)$ iff $\mathrm{X_n} \stackrel{\mathrm{D}}{\to} \tau(\theta)$.
e) If $X_n \stackrel{P}{\to} \theta$ and τ is continuous at θ, then $\tau(X_n) \stackrel{P}{\to} \tau(\theta)$.
f) If $X_n \stackrel{D}{\to} \theta$ and τ is continuous at θ, then $\tau(X_n) \stackrel{D}{\to} \tau(\theta)$.

Suppose that for all $\theta \in \Theta$, $T_n \stackrel{D}{\to} \tau(\theta)$, $T_n \stackrel{r}{\to} \tau(\theta)$ or $T_n \stackrel{wp1}{\to} \tau(\theta)$. Then T_n is a consistent estimator of $\tau(\theta)$ by Theorem 2.21. We are assuming that the function τ does not depend on n since we want a single function $\tau(\theta)$ rather than a sequence of functions $\tau_n(\theta)$. See Remark 2.1 b).

Example 2.11. 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 2.11 and 2.21), ii) the WLLN holds and iii) the CLT holds (use Theorem 2.10). Since

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

$\overline{Y}_n$ is also a consistent estimator of μ by Theorem 2.9b. By the delta method and Theorem 2.10b, $T_n = g(\overline{Y}_n)$ is a consistent estimator of $g(\mu)$ if $g'(\mu) \neq 0$

for all $\mu \in \Theta$. By Theorem 2.21e, $g(\overline{Y}_n)$ is a consistent estimator of $g(\mu)$ if g is continuous at μ for all $\mu \in \Theta$.

Theorem 2.22, Helly-Bray-Pormanteau Theorem: $X_n \xrightarrow{D} X$ iff $E[g(X_n)] \to E[g(X)]$ for every bounded, real, continuous function g.

The above theorem is used to prove Theorem 2.23 b).

Theorem 2.23. a) **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 $g(X_n) \xrightarrow{D} g(X)$.

b) **Continuous Mapping Theorem:** If $X_n \xrightarrow{D} X$ and the function g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Remark 2.6. For Theorem 2.21, 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 2.23 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 2.12. (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 2.13. 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 2.21e.

Theorem 2.24: 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)

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

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 $|t| \leq d$ 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 $Y_n \xrightarrow{D} Y$.

Theorem 2.25: If $\lim_{n\to\infty} c_{X_n}(t) = g(t)$ for all t where g is continuous at $t = 0$, then $g(t) = c_X(t)$ is a characteristic function for some RV X, and $X_n \xrightarrow{D} X$.

Remark 2.7. a) Continuity at $t = 0$ implies continuity everywhere since $g(t) = c_X(t)$ is continuous. If $g(t)$ is not continuous at 0, then X_n does not converge in distribution.

b) If $c_{Y_n}(t) \to h(t)$ where $h(t)$ is not continuous, then Y_n does not converge in distribution to any RV Y, by the Continuity Theorem and a).

c) Let $X_1, ..., X_n$ be independent RVs with characteristic functions $c_{X_j}(t)$. Then the characteristic function of $\sum_{j=1}^n X_j$ is $c_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n c_{X_j}(t)$. If the RVs also have mgfs $m_{X_j}(t)$, then the mgf of $\sum_{j=1}^n X_j$ is $m_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n m_{X_j}(t)$.

Application: Proof of a Special Case of the CLT. Following Rohatgi (1984, p. 569-9) and Tardiff (1981), 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 $\psi(x) = \log(m_Z(x))$. Then

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

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

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

Now

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

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

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

Now

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

So

$$\psi''(0) = m''_Z(0) - [m'_Z(0)]^2 = E(Z_i^2) - [E(Z_i)]^2 = 1.$$

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,

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

By Theorem 2.26, $d_n F_{g,d_n,1-\delta} \to \chi^2_{g,1-\delta}$ as $d_n \to \infty$. Here $P(X \le \chi^2_{g,1-\delta}) = 1-\delta$ if $X \sim \chi^2_g$, and $P(X \le F_{g,d_n,1-\delta}) = 1-\delta$ if $X \sim F_{g,d_n}$.

Theorem 2.26. If $W_n \sim F_{r,d_n}$ where the positive integer $d_n \to \infty$ as $n \to \infty$, then $rW_n \xrightarrow{D} \chi^2_r$.

Proof. If $X_1 \sim \chi^2_{d_1} \perp\!\!\!\perp X_2 \sim \chi^2_{d_2}$, then

$$\frac{X_1/d_1}{X_2/d_2} \sim F_{d_1,d_2}.$$

If $U_i \sim \chi^2_1$ are iid then $\sum_{i=1}^k U_i \sim \chi^2_k$. Let $d_1 = r$ and $k = d_2 = d_n$. Hence if $X_2 \sim \chi^2_{d_n}$, then

$$\frac{X_2}{d_n} = \frac{\sum_{i=1}^{d_n} U_i}{d_n} = \overline{U} \xrightarrow{P} E(U_i) = 1$$

by the law of large numbers. Hence if $W \sim F_{r,d_n}$, then $rW_n \xrightarrow{D} \chi^2_r$. □

2.5 Order Relations and Convergence Rates

Definition 2.10. 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

$$P(|W_n| \le D_\epsilon) \ge 1 - \epsilon$$

for all $n \ge N_\epsilon$. 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

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

for all $n \ge N_\epsilon$.

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 2.11. 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.

Theorem 2.27. 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 $\overline{Y}_n \asymp_P n^{-1/2}$.

Theorem 2.28. 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 $W_n \asymp_P X_n$,

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

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

b) Since $W_n \asymp_P X_n$,

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

for all $n \ge N_\epsilon$. 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

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

for all $n \ge N_1$, and

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

for all $n \ge N_2$. Hence

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

and

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

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 $W_n \asymp_P X_n$. $\square$

The following result is used to prove the following Theorem 2.30 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 $\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta})$.

Theorem 2.29: 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

$$W_n = O_P(1). \tag{2.8}$$

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

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

Theorem 2.30. 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

$$\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta}). \tag{2.9}$$

Proof. Let $X_{j,n} = n^\delta \|T_{j,n} - \boldsymbol{\beta}\|$. Then $X_{j,n} = O_P(1)$ so by Theorem 2.29, $n^\delta \|T_n^* - \boldsymbol{\beta}\| = O_P(1)$. Hence $\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta})$. $\square$

2.6 More CLTs

Remark 2.8. For each positive integer n, let $W_{n1}, ..., W_{nr_n}$ be independent. The probability space may change with n, giving a triangular array of random variables. Let $E[W_{nk}] = 0$, $V(W_{nk}) = E[W_{nk}^2] = \sigma_{nk}^2$, and $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 = V[\sum_{k=1}^{r_n} W_{nk}]$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n}$$

is the z-score of $\sum_{k=1}^{r_n} W_{nk}$.

For the above remark, let $r_n = n$. Then the triangular array is shown below.
W_{11}
W_{21}, W_{22}
W_{31}, W_{32}, W_{33}
$\vdots$
$W_{n1}, W_{n2}, W_{n3}, ..., W_{nn}$
$\vdots$

Theorem 2.31, Lyapounov's CLT: Under Remark 2.8), assume the $|W_{nk}|^{2+\delta}$ are integrable for some $\delta > 0$. Assume Lyapounov's condition:

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^{2+\delta}]}{s_n^{2+\delta}} = 0.$$

Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1).$$

Example 2.14. Special cases: i) $r_n = n$ and $W_{nk} = W_k$ has $W_1, ..., W_n, ...$ independent.
ii) $W_{nk} = X_{nk} - E(X_{nk}) = X_{nk} - \mu_{nk}$ has

$$\frac{\sum_{k=1}^{r_n} (X_{nk} - \mu_{nk})}{s_n} \xrightarrow{D} N(0,1).$$

iii) Suppose $X_1, X_2, ...$ are independent with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$. Let

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{(\sum_{i=1}^n \sigma_i^2)^{1/2}}$$

be the z-score of $\sum_{i=1}^n X_i$. Assume $E[|X_i - \mu_i|^3] < \infty$ for all $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E[|X_i - \mu_i|^3]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0. \tag{2.10}$$

Then $Z_n \xrightarrow{D} N(0,1)$.

The (Lindeberg-Lévy) CLT has the X_i iid with $V(X_i) = \sigma^2 < \infty$. The Lyapounov CLT in Example 2.14 iii) has the X_i independent (not necessarily identically distributed), but needs stronger moment conditions to satisfy Equation (2.10).

Theorem 2.32, Lindeberg CLT: Let the W_{nk} satisfy Remark 2.8) and Lindeberg's condition

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E(W_{nk}^2 \; I[|W_{nk}| \ge \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1).$$

Note: The Lindeberg CLT is sometimes called the Lindeberg-Feller CLT. Lindeberg's condition is nearly necessary for $Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1)$.

Example 2.15. a) Special case of the Lindeberg CLT: Let $r_n = n$ and let the $W_{nk} = W_k$ be independent. If

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{E(W_k^2 \; I[|W_k| \ge \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{n} W_k}{s_n} \xrightarrow{D} N(0,1).$$

b) **uniformly bounded sequence**: Let $r_n = n$ and $W_{nk} = W_k$. If there is a constant $c > 0$ such that $P(|W_k| < c) = 1 \; \forall k$, and if $s_n \to \infty$ as $n \to \infty$, then Lindeberg's CLT holds.

c) Let $r_n = n$ and let the $W_{nk} = W_k$ be **iid** with $V(W_k) = \sigma^2 \in (0, \infty)$. Then Lindeberg's CLT holds. (Taking $W_i = X_i - \mu$ proves the usual CLT with the Lindeberg CLT.)

d) If Lyapunov's condition holds, then Lindeberg's condition holds. Hence the Lindeberg CLT proves the Lyapounov CLT.

2.7 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 a) $Z_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) = \left(\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{\sum_{i=1}^n Y_i - n\mu}{\sqrt{n}\sigma}\right)$ is the z–score of $\overline{X}_n$ (and the z-score of $\sum_{i=1}^n Y_i$), and $Z_n \xrightarrow{D} N(0,1)$. b) Two applications of the CLT are to give the limiting distribution of $\sqrt{n}(\overline{Y}_n - \mu)$ and the limiting distribution of $\sqrt{n}(Y_n/n - \mu_Y)$ 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 $V(X) = \sigma_X^2$. See Section 1.4. c) The CLT is the Lindeberg-Lévy CLT.

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

4) **Second Order Delta Method**: Suppose that $g'(\theta) = 0$, $g''(\theta) \neq 0$ and $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \tau^2(\theta))$. Then $n[g(T_n) - g(\theta)] \xrightarrow{D} \frac{1}{2}\tau^2(\theta)g''(\theta)\chi_1^2$.

5) **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 $\sqrt{n}[\frac{1}{n}\sum_{i=1}^n t(Y_i) - \mu_t] \xrightarrow{D} N(0, I_1(\eta))$ where $I_1(\eta) = \sigma_t^2 = g'(\eta)$.

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

7) 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 $\sigma_i^2(F)$.

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

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

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

10) Standard Limit Theorem: Under strong regularity conditions, 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

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

11) $X_n \xrightarrow{D} X$ if
$$\lim_{n\to\infty} F_n(t) = F(t)$$
at each continuity point t of F. **Convergence in distribution** is also known as weak convergence and convergence in law. X is the limiting distribution or asymptotic distribution of X_n. **The limiting distribution does not depend on** the sample size n. $X_n \xrightarrow{D} \tau(\theta)$ if $X_n \xrightarrow{D} X$ where $P(X = \tau(\theta)) = 1$: hence X is *degenerate at* $\tau(\theta)$ or the distribution of X is a *point mass at* $\tau(\theta)$.

12) If $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} Y$, then i) $X \stackrel{D}{=} Y$ and ii) $F_X(x) = F_Y(x)$ for all real x.

13) **Convergence in probability**: a) $X_n \xrightarrow{P} \tau(\theta)$ if for every $\epsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - \tau(\theta)| < \epsilon) = 1 \ \text{ or, equivalently,} \ \lim_{n\to\infty} \mathrm{P}(|\mathrm{X_n} - \tau(\theta)| \geq \epsilon) = 0.$$
b) $X_n \xrightarrow{P} X$ if for every $\epsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1, \ \text{ or, equivalently,} \ \lim_{n\to\infty} \mathrm{P}(|\mathrm{X_n} - \mathrm{X}| \geq \epsilon) = 0.$$

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

15) Theorem: T_n is a **consistent estimator** of $\tau(\theta)$ if any of the following 2 conditions holds:

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

ii) $MSE_{\tau(\theta)}(T_n) = E[(T_n - \tau(\theta))^2] \to 0$ for all $\theta \in \Theta$.

Here
$$MSE_{\tau(\theta)}(T_n) = V_\theta(T_n) + [\mathrm{Bias}_{\tau(\theta)}(\mathrm{T_n})]^2$$
where $\mathrm{Bias}_{\tau(\theta)}(\mathrm{T_n}) = \mathrm{E}_\theta(\mathrm{T_n}) - \tau(\theta)$.

16) Theorem: a) Let X_θ be a random variable with a distribution depending on θ, and $0 < \delta \leq 1$. If
$$n^\delta (T_n - \tau(\theta)) \xrightarrow{D} X_\theta$$
for all $\theta \in \Theta$, 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 $\tau(\theta)$.

Note: If $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$, then $T_n \xrightarrow{P} \theta$. Often $X_\theta \sim N(0, v(\theta))$.

17) **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.

18) Y_n **converges in rth mean** to a random variable Y, $Y_n \xrightarrow{r} Y$, 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 $n \to \infty$. $Y_n \xrightarrow{r} \tau(\theta)$ if $E(|Y_n - \tau(\theta)|^r) \to 0$ as $n \to \infty$. If $r \geq 1$, $Y_n \xrightarrow{r} Y$ is often written as $Y_n \xrightarrow{L^r} Y$ or $Y_n \xrightarrow{L_r} Y$.

19) A sequence of random variables X_n *converges with probability 1* (or *almost surely*, or *almost everywhere*, or *strong convergence*) to X if

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

This type of convergence will be denoted by $X_n \xrightarrow{wp1} X$. Notation such as "X_n converges to X wp1" will also be used. Sometimes "wp1" will be replaced with "as" or "ae."

$$X_n \xrightarrow{wp1} \tau(\theta),$$

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

20) **SLLN**: If $X_1, ..., X_n$ are iid with $E(X_i) = \mu$ finite, then $\overline{X}_n \xrightarrow{wp1} \mu$.

21) a) For i) $X_n \xrightarrow{P} X$, ii) $X_n \xrightarrow{r} X$, or iii) $X_n \xrightarrow{wp1} X$, the X_n and X need to be defined on the same probability space.

b) For $X_n \xrightarrow{D} X$, the probability spaces can differ.

c) For i) $X_n \xrightarrow{P} c$, ii) $X_n \xrightarrow{wp1} c$, iii) $X_n \xrightarrow{D} c$, and iv) $X_n \xrightarrow{r} c$, the probability spaces of the X_n can differ.

22) Theorem: i) $T_n \xrightarrow{P} \tau(\theta)$ iff $\mathrm{T_n} \xrightarrow{\mathrm{D}} \tau(\theta)$.

ii) 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.

23) Theorem: Suppose X_n and X are RVs with the same probability space for b) and c). Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function.
a) If $X_n \xrightarrow{D} X$, then $g(X_n) \xrightarrow{D} g(X)$.
b) If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$.
c) If $X_n \xrightarrow{ae} X$, then $g(X_n) \xrightarrow{wp1} g(X)$.

24) Theorem: Suppose X_n and X are RVs with the same probability space.
a) If $X_n \xrightarrow{wp1} X$, then $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$.
b) If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$.
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** $X_n \xrightarrow{D} \tau(\theta)$ where c is a constant.

25) Theorem: a) If $E[(X_n - X)^2] \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} X$.

b) If $E(X_n) \to E(X)$ and $V(X_n - X) \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} X$.

Note: See 15) if $P(X = \tau(\theta)) = 1$.

26) Theorem: If $X_n \stackrel{r}{\to} X$, then $X_n \stackrel{k}{\to} X$ where $0 < k < r$.

27) Theorem: Let X_n have pdf $f_{X_n}(x)$, and let X have pdf $f_X(x)$. If $f_{X_n}(x) \to f_X(x)$ for all x (or for x outside of a set of Lebesgue measure 0), then $X_n \stackrel{D}{\to} X$.

28) Theorem: Let $g : \mathbb{R} \to \mathbb{R}$ be continuous at constant c.

a) If $X_n \stackrel{D}{\to} c$, then $g(X_n) \stackrel{D}{\to} c$.

b) If $X_n \stackrel{P}{\to} c$, then $g(X_n) \stackrel{P}{\to} c$.

c) If $X_n \stackrel{wp1}{\to} c$, then $g(X_n) \stackrel{wp1}{\to} c$.

29) Theorem: Suppose X_n and X are integer valued RVs with pmfs $f_{X_n}(x)$ and $f_X(x)$. Then $X_n \stackrel{D}{\to} X$ iff $P(X_n = k) \to P(X = k)$ for every integer k iff $f_{X_n}(x) \to f_X(x)$ for every real x.

30) **Slutsky's Theorem**: If $Y_n \stackrel{D}{\to} Y$ and $W_n \stackrel{P}{\to} w$ for some constant w, then i) $Y_n W_n \stackrel{D}{\to} wY$, ii) $Y_n + W_n \stackrel{D}{\to} Y + w$ and iii) $Y_n/W_n \stackrel{D}{\to} Y/w$ for $w \neq 0$.

Note that $Y_n \stackrel{B}{\to} Y$ implies $Y_n \stackrel{D}{\to} Y$ where $B = wp1, r$, or P. Also $W_n \stackrel{P}{\to} c$ iff $W_n \stackrel{D}{\to} c$. If a sequence of constants $c_n \to c$ as $n \to \infty$ (everywhere convergence), then $c_n \stackrel{wp1}{\to} c$ and $c_n \stackrel{P}{\to} c$. (So everywhere convergence is a special case of almost everywhere convergence.)

31) The **cumulative distribution function** (cdf) of any random variable Y is $F(y) = P(Y \leq y)$ for all $y \in \mathbb{R}$. If $F(y)$ is a cumulative distribution function, then i) $F(-\infty) = \lim_{y \to -\infty} F(y) = 0$, ii) $F(\infty) = \lim_{y \to \infty} F(y) = 1$, iii) F is a nondecreasing function: if $y_1 < y_2$, then $F(y_1) \leq F(y_2)$, iv) F is right continuous: $\lim_{h \downarrow 0} F(y+h) = F(y)$ for all real y. v) Since a cdf is a probability for fixed y, $0 \leq F(y) \leq 1$ for all real y. vi) A cdf $F(y)$ can have at most countably many points of discontinuity, vii) $P(a < Y \leq b) = F(b) - F(a)$. viii) If Y is a random variable, then $F_Y(y)$ completely determines the distribution of Y.

32) The **moment generating function** (mgf) of a random variable Y is

$$m(t) = E[e^{tY}] \tag{2.11}$$

if the expectation exists for t in some neighborhood of 0. Otherwise, the mgf does not exist. If Y is discrete, then $m(t) = \sum_y e^{ty} f(y)$, and if Y is continuous, then $m(t) = \int_{-\infty}^{\infty} e^{ty} f(y) dy$. If Y is a random variable and $m_Y(t)$ exists, then $m_Y(t)$ completely determines the distribution of Y.

Notes: a) If X has mgf $m_X(t)$, then $E(X^k)$ exists for all positive integers k.

b) Let j and k be positive integers. If $E(X^k)$ is finite, then $E(X^j)$ is finite for $1 \leq j \leq k$.

33) The **characteristic function** of a random variable Y is $c(t) = E[e^{itY}] = E[\cos(tY)] + iE[\sin(tY)]$ where the complex number $i = \sqrt{-1}$. i) $c(0) = 1$, ii) the modulus $|c(t)| \leq 1$ for all real t, iii) $c(t)$ is a continuous function. iv) If $E(Y) = 0$ and $E(Y^2) = V(Y) = \sigma^2$, then

$$c_Y(t) = 1 + \frac{t^2\sigma^2}{2} + o(t^2) \text{ as } t \to 0.$$

Here $a(t) = o(t^2)$ as $t \to 0$ if $\lim_{t\to 0} \frac{a(t)}{t^2} = 0$. v) If Y is discrete with pmf $f_Y(y)$, then $c_Y(t) = \sum_y e^{ity} f_y(y)$. vi) If Y is a random variable, then $c_Y(t)$ always exists, and completely determines the distribution of Y.

34) **Continuity Theorem:** Let Y_n be sequence of random variables with characteristic functions $c_{Y_n}(t)$. Let Y be a random variable with cf $c_Y(t)$.

a)

$$Y_n \xrightarrow{D} Y \text{ iff } c_{Y_n}(t) \to c_Y(t)\ \forall t \in \mathbb{R}.$$

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

35) Theorem: If $\lim_{n\to\infty} c_{X_n}(t) = g(t)$ for all t where g is continuous at $t = 0$, then $g(t) = c_X(t)$ is a characteristic function for some RV X, and $X_n \xrightarrow{D} X$.

Note: Hence continuity at $t = 0$ implies continuity everywhere since $g(t) = \varphi_X(t)$ is continuous. If $g(t)$ is not continuous at 0, then X_n does not converge in distribution.

36) If $c_{Y_n}(t) \to h(t)$ where $h(t)$ is not continuous, then Y_n does not converge in distribution to any RV Y, by the Continuity Theorem and 35).

37) Let $X_1, ..., X_n$ be independent RVs with characteristic functions $c_{X_j}(t)$. Then the characteristic function of $\sum_{j=1}^n X_j$ is $c_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n c_{X_j}(t)$. If the RVs also have mgfs $m_{X_j}(t)$, then the mgf of $\sum_{j=1}^n X_j$ is $m_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n m_{X_j}(t)$.

38) **Helly-Bray-Pormanteau Theorem**: $X_n \xrightarrow{D} X$ iff $E[g(X_n)] \to E[g(X)]$ for every bounded, real, continuous function g.

Note: 38) is used to prove 39 b).

39) a) **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 $g(X_n) \xrightarrow{D} g(X)$.

Note: $P[X \in C(g)] = 1$ can be replaced by $P[X \in D(g)] = 0$ where $D(g)$ is the set of points where g is not continuous.

b) **Continuous Mapping Theorem:** If $X_n \xrightarrow{D} X$ and the function g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Note: the function g can not depend on n since g_n is a seqeuce of functions rather than a single function.

40) Generalized Chebyshev's Inequality or Generalized Markov's Inequality: Let $u : \mathbb{R} \to [0, \infty)$ 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 $\sigma^2 = V(Y)$ exists, then we obtain
Chebyshev's Inequality:

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

41) a) $\lim_{n\to\infty} \left(1 - \frac{c}{n}\right)^n = e^{-c}$.

b) If $c_n \to c$ as $n \to \infty$, then $\lim_{n\to\infty} \left(1 + \frac{-c_n}{n}\right)^n = e^{-c}$.

c) If c_n is a sequence of complex numbers such that $c_n \to c$ as $n \to \infty$ where c is real, then $\lim_{n\to\infty} \left(1 - \frac{c_n}{n}\right)^n = e^{-c}$.

42) For each positive integer n, let $W_{n1}, ..., W_{nr_n}$ be independent. The probability space may change with n, giving a triangular array of RVs. Let $E[W_{nk}] = 0$, $V(W_{nk}) = E[W_{nk}^2] = \sigma_{nk}^2$, and $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 = V[\sum_{k=1}^{r_n} W_{nk}]$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n}$$

is the z-score of $\sum_{k=1}^{r_n} W_{nk}$.

43) **Lyapounov's CLT**: Under 42), assume the $|W_{nk}|^{2+\delta}$ are integrable for some $\delta > 0$. Assume Lyapounov's condition:

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^{2+\delta}]}{s_n^{2+\delta}} = 0.$$

Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0, 1).$$

44) Special cases: i) $r_n = n$ and $W_{nk} = W_k$ has $W_1, ..., W_n, ...$ independent.
ii) $W_{nk} = X_{nk} - E(X_{nk}) = X_{nk} - \mu_{nk}$ has

$$\frac{\sum_{k=1}^{r_n} (X_{nk} - \mu_{nk})}{s_n} \xrightarrow{D} N(0, 1).$$

iii) Suppose $X_1, X_2, ...$ are independent with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$. Let

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}$$

be the z-score of $\sum_{i=1}^n X_i$. Assume $E[|X_i - \mu_i|^3] < \infty$ for all $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E[|X_i - \mu_i|^3]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0. \qquad (*)$$

Then $Z_n \xrightarrow{D} N(0,1)$.

45) The (Lindeberg-Lévy) CLT has the X_i iid with $V(X_i) = \sigma^2 < \infty$. The Lyapounov CLT in 43 iii) has the X_i independent (not necessarily identically distributed), but needs stronger moment conditions to satisfy $(*)$.

46) **Lindeberg CLT**: Let the W_{nk} satisfy 42) and Lindeberg's condition

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E(W_{nk}^2 \; I[|W_{nk}| \geq \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1).$$

Notes: The Lindeberg CLT is sometimes called the Lindeberg-Feller CLT. Lindeberg's condition is nearly necessary for $Z_n = \dfrac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1)$.

47) Special case of the Lindeberg CLT: Let $r_n = n$ and let the $W_{nk} = W_k$ be independent. If

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{E(W_k^2 \; I[|W_k| \geq \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{n} W_k}{s_n} \xrightarrow{D} N(0,1).$$

48) a) **uniformly bounded sequence**: Let $r_n = n$ and $W_{nk} = W_k$. If there is a constant $c > 0$ such that $P(|W_k| < c) = 1 \; \forall k$, and if $s_n \to \infty$ as $n \to \infty$, then Lindeberg's CLT 46) holds.

b) Let $r_n = n$ and let the $W_{nk} = W_k$ be **iid** with $V(W_k) = \sigma^2 \in (0, \infty)$. Then Lindeberg's CLT 46) holds. (Taking $W_i = X_i - \mu$ proves the usual CLT with the Lindeberg CLT.)

c) If Lyapunov's condition holds, then Lindeberg's condition holds. Hence the Lindeberg CLT proves the Lyapounov CLT.

2.8 Complements

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). Since $\overline{Y} \stackrel{P}{\to} \mu$ iff $\overline{Y} \stackrel{D}{\to} \mu$, the WLLN refers to weak convergence. Almost sure convergence is also called strong convergence. Hence the SLLN refers to strong 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. Theorem 2.5 appears in Olive (2014). 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).

Casella and Berger (2002, p. 112, 133) give results similar to Theorem 2.4. Some of the order relations of Section 2.5 are discussed in Mann and Wald (1943a). See Ver Hoef (2012) for history of the delta method.

2.9 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

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

2.1*. a) Enter the following *R* 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 200. 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 $\overline{Y} \sim N(0, 1/n)$.

2.2*. 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.

2.3^{*Q}. 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.

2.4*. From Theorems 1.24 and 1.25,
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 $V(X) = \sigma^2$ 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) $Y_n \sim$ BIN(n, ρ) 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) $Y_n \sim N(n\mu, n\sigma^2)$.

2.5*. 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?

2.6. 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 if $|t| \le d$ for some $d > 0$. Thus 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?

2.7. 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.

2.8^Q. 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).

Problems from old quizzes and exams. Problems from old qualifying exams are marked with a Q.

2.9. 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.

2.10. 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.

2.11. 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.

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

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

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

2.13. 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 $V(Y_i)$.

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.

2.14. 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.

2.15. 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.

2.16. Suppose that $Y_1, ..., Y_n$ are iid with $E(Y) = (1-\rho)/\rho$ and $V(Y) = (1-\rho)/\rho^2$ 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.

2.17. 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 $\sqrt{n}\ \left[\ \left(\frac{X_n}{n}\right)^2 -\ p^2\ \right]$.

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

2.19. 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

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

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

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

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

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

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

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

2.21. 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?

2.22. 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 (e.g. 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

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

2.23. 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 2.3.

2.24. 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 $n \to \infty$.

2.25. 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 2.3.

2.26^Q. 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 $g(\theta) = \theta^3 - \theta$.)

2.27^Q. 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 $\sqrt{n}\ [\ (\overline{X})^3 - (\lambda)^3\ ]$.

2.28^Q. 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.

2.29. 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?

2.30. 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:

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

2.31. 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.

2.32. Suppose $Y_n \sim$ uniform$(0, n)$. 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.

2.33^Q. Let $Y_1, ..., Y_n$ be independent and identically distributed (iid) from a distribution with probability mass function $f(y) = \rho(1-\rho)^y$ for $y = 0, 1, 2, ...$ and $0 < \rho < 1$. Then $E(Y) = (1-\rho)/\rho$ and $V(Y) = (1-\rho)/\rho^2$.

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

b) Show how to find the limiting distribution of $g(\overline{Y}) = \frac{1}{1+\overline{Y}}$. Deduce it completely. (This bad notation means find the limiting distribution of $\sqrt{n}(g(\overline{Y}) - c)$ for some constant c.)

c) Find the method of moments estimator of ρ.

d) Find the limiting distribution of $\sqrt{n}\ \left((1+\overline{Y}) - \ d \right)$ for appropriate constant d.

e) Note that $1 + E(Y) = 1/\rho$. Find the method of moments estimator of $1/\rho$.

2.34^Q. Let $X_1, ..., X_n$ be independent identically distributed random variables from a normal distribution with mean μ and variance σ^2.

a) Find the approximate distribution of $1/\bar{X}$. Is this valid for all values of μ?

b) Show that $1/\bar{X}$ is asymptotically efficient for $1/\mu$, provided $\mu \neq \mu^*$. Identify μ^*.

2.35^Q. Let $Y_1, ..., Y_n$ be independent and identically distributed (iid) from a distribution with probability density function

$$f(y) = \frac{2y}{\theta^2}$$

for $0 < y \le \theta$ and $f(y) = 0$, otherwise.

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

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

c) Find the method of moments estimator of θ^k.

2.36^Q. Let $Y_1, ..., Y_n$ be independent identically distributed discrete random variables with probability mass function

$$f(y) = P(Y = y) = \binom{r + y - 1}{y} \rho^r (1-\rho)^y$$

for $y = 0, 1, \ldots$ where positive integer r **is known** and $0 < \rho < 1$. Then $E(Y) = r(1-\rho)/\rho$, and $V(Y) = r(1-\rho)/\rho^2$.

a) Find the limiting distribution of $\sqrt{n}\ \left(\overline{Y} - \dfrac{r(1-\rho)}{\rho} \right)$.

b) Let $g(\overline{Y}) = \dfrac{r}{r + \overline{Y}}$. Find the limiting distribution of $\sqrt{n}\ \left(g(\overline{Y}) - \ c \right)$ for appropriate constant c.

c) Find the method of moments estimator of ρ.

2.37^Q. Let $X_1, ..., X_n$ be independent identically distributed uniform $(0, \theta)$ random variables where $\theta > 0$.

a) Find the limiting distribution of $\sqrt{n}(\overline{X} - c_\theta)$ for an appropriate constant c_θ that may depend on θ.

b) Find the limiting distribution of $\sqrt{n}[(\overline{X})^2 - k_\theta]$ for an appropriate constant k_θ that may depend on θ.

2.38^Q. Let $X_1, ..., X_n$ be independent identically distributed (iid) beta(β, β) random variables.

a) Find the limiting distribution of $\sqrt{n}(\ \overline{X}_n - \theta\ )$, for appropriate constant θ.

b) Find the limiting distribution of $\sqrt{n}(\ \log(\overline{X}_n) - d\ )$, for appropriate constant d.

2.39.
2.40.
2.41.
2.42.
2.43.
2.44.
2.45.
2.46.
2.47.

More Problems:

2.48. Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Let $g(\mu) = \mu^2$. For $\mu = 0$, find the limiting distribution of $n[(\overline{Y}_n)^2 - 0^2] = n(\overline{Y}_n)^2$ by using the Second Order Delta Method.