556: MATHEMATICAL STATISTICS I

ASYMPTOTIC APPROXIMATIONS AND THE DELTA METHOD

To approximate the distribution of elements in sequence of random variables $\{X_n\}$ for large n, we attempt to find sequences of constants $\{a_n\}$ and $\{b_n\}$ such that $Z_n = a_n X_n + b_n \stackrel{d}{\longrightarrow} Z$, where Z has some distribution characterized by cdf F_Z. Then, for large n, $F_{Z_n}(z) \simeq F_Z(z)$, so

$$F_{X_n}(x) = P[X_n \leq x] = P[a_n X_n + b_n \leq a_n x + b_n] = F_{Z_n}(a_n x + b_n) \simeq F_Z(a_n x + b_n).$$

EXAMPLE Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. such that $X_i \sim Exp(1)$, and let $Y_n = \max\{X_1, X_2, \ldots, X_n\}$. Then by a previous result, $F_{Y_n}(y) = \{F_X(y)\}^n$, so for $y > 0$, $F_{Y_n}(y) = \{1 - e^{-y}\}^n \longrightarrow 0$, and there is no limiting distribution. However, if we take $a_n = 1$ and $b_n = -\log n$, and set $Z_n = a_n Y_n + b_n$, then as $n \longrightarrow \infty$,

$$F_{Z_n}(z) = \Pr[Z_n \leq z] = \Pr[Y_n \leq z + \log n] = \{1 - e^{-z-\log n}\}^n \longrightarrow \exp\{-e^{-z}\} = F_Z(z),$$

$$\therefore \quad F_{Y_n}(y) = P[Y_n \leq y] = P[Z_n \leq y - \log n] \simeq F_Z(y - \log n) = \exp\{-e^{-y+\log n}\} = \exp\{-ne^{-y}\}$$

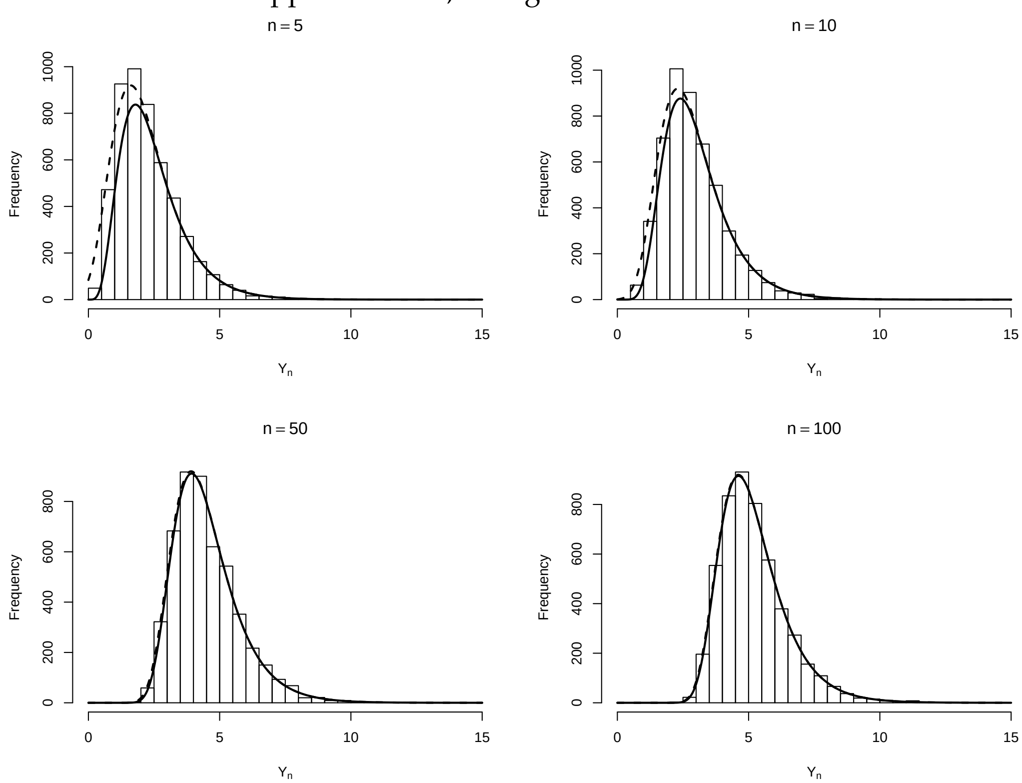
and by differentiating

$$f_{Y_n}(y) \simeq ne^{-y}\exp\{-ne^{-y}\} \quad y > 0.$$

This can be compared with the exact version

$$f_{Y_n}(y) = ne^{-y}(1 - e^{-y})^n \quad y > 0.$$

The figure below compares the approximations for $n = 50, 100, 500, 1000$. Solid lines use the exact formula, dotted lines use the approximation, histograms are 5000 simulated values.



DEFINITION (ASYMPTOTIC NORMALITY)
A sequence of random variables $\{X_n\}$ is **asymptotically normally distributed** as $n \longrightarrow \infty$ if there exist sequences of real constants $\{\mu_n\}$ and $\{\sigma_n\}$ (with $\sigma_n > 0$) such that

$$\frac{X_n - \mu_n}{\sigma_n} \stackrel{d}{\longrightarrow} Z \sim N(0,1).$$

The notation $X_n \stackrel{\cdot}{\sim} N(\mu_n, \sigma_n^2)$ or $X_n \sim AN(\mu_n, \sigma_n^2)$ as $n \longrightarrow \infty$ is commonly used.

THEOREM (THE DELTA METHOD)
Consider sequence of random variables $\{X_n\}$ such that

$$\sqrt{n}(X_n - \mu) \stackrel{d}{\longrightarrow} X.$$

Suppose that $g(.)$ is a function such that first derivative $\dot{g}(.)$ is continuous in a neighbourhood of μ, with $\dot{g}(\mu) \neq 0$. Then

$$\sqrt{n}(g(X_n) - g(\mu)) \stackrel{d}{\longrightarrow} \dot{g}(\mu)X.$$

In particular, if

$$\sqrt{n}(X_n - \mu) \stackrel{d}{\longrightarrow} X \sim N(0, \sigma^2).$$

then

$$\sqrt{n}(g(X_n) - g(\mu)) \stackrel{d}{\longrightarrow} \dot{g}(\mu)X \sim N(0, \{\dot{g}(\mu)\}^2\sigma^2).$$

Proof. Consider a Taylor series expansion of $g(X_n)$ about μ;

$$g(X_n) = g(\mu) + \dot{g}(\mu)(X_n - \mu) + \sum_{r=2}^{\infty} \frac{g^{(r)}(\mu)}{r!}(X_n - \mu)^r \qquad (1)$$

Now as

$$\sqrt{n}(X_n - \mu) \stackrel{d}{\longrightarrow} X \qquad \Longrightarrow \qquad X_n - \mu \stackrel{d}{\longrightarrow} 0 \qquad \Longrightarrow \qquad X_n \stackrel{d}{\longrightarrow} \mu$$

it can be shown that

$$\sum_{r=2}^{\infty} \frac{g^{(r)}(\mu)}{r!}(X_n - \mu)^r \stackrel{d}{\longrightarrow} 0$$

and we can rewrite equation (1) that

$$g(X_n) = g(\mu) + \dot{g}(\mu)(X_n - \mu) + o_p(1)$$

using the **stochastic order notation**, where $o_p(1)$ indicates a term that converges in probability to zero. Thus using **Slutsky's Theorem**, we have that

$$\sqrt{n}(g(X_n) - g(\mu)) = \dot{g}(\mu)\sqrt{n}(X_n - \mu) \stackrel{d}{\longrightarrow} \dot{g}(\mu)X$$

and if $X \sim N(0, \sigma^2)$, it follows from the properties of the Normal distribution that

$$\sqrt{n}(g(X_n) - g(\mu)) \stackrel{d}{\longrightarrow} N(0, \{\dot{g}(\mu)\}^2\sigma^2).$$

∎

Note: This result extends to the multivariate case. Consider a sequence of vector random variables $\{\underset{\sim}{X}_n\}$ such that

$$\sqrt{n}(\underset{\sim}{X}_n - \underset{\sim}{\mu}) \xrightarrow{d} \underset{\sim}{X}.$$

and $\underset{\sim}{g} : \mathbb{R}^k \longrightarrow \mathbb{R}^d$ is a vector-valued function with first derivative matrix $\dot{\underset{\sim}{g}}(.)$ which is continuous in a neighbourhood of $\underset{\sim}{\mu}$, with $\dot{g}(\underset{\sim}{\mu}) \neq \underline{0}$. Note that $\underset{\sim}{g}$ can be considered as a $d \times 1$ vector of scalar functions.

$$\underset{\sim}{g}(\underset{\sim}{x}) = (g_1(\underset{\sim}{x}), \ldots, g_d(\underset{\sim}{x}))^\mathsf{T}.$$

Note that $\dot{\underset{\sim}{g}}(\underset{\sim}{x})$ is a $(d \times k)$ matrix with $(i, j)th$ element

$$\frac{\partial g_i(\underset{\sim}{x})}{\partial x_j}$$

Under these assumptions, in general

$$\sqrt{n}(\underset{\sim}{g}(\underset{\sim}{X}_n) - \underset{\sim}{g}(\underset{\sim}{\mu})) \xrightarrow{d} \dot{\underset{\sim}{g}}(\underset{\sim}{\mu})\underset{\sim}{X}.$$

and in particular, if

$$\sqrt{n}(\underset{\sim}{X}_n - \underset{\sim}{\mu}) \xrightarrow{d} \underset{\sim}{X} \sim N(0, \Sigma).$$

where Σ is a positive definite, symmetric $k \times k$ matrix, then

$$\sqrt{n}(\underset{\sim}{g}(\underset{\sim}{X}_n) - \underset{\sim}{g}(\underset{\sim}{\mu})) \xrightarrow{d} \dot{\underset{\sim}{g}}(\underset{\sim}{\mu})X \sim N\left(0, \left\{\dot{\underset{\sim}{g}}(\mu)\right\} \Sigma \left\{\dot{\underset{\sim}{g}}(\mu)\right\}^\mathsf{T}\right).$$

THEOREM (THE SECOND ORDER DELTA METHOD: Normal case)
Consider sequence of random variables $\{X_n\}$ such that

$$\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Suppose that $g(.)$ is a function such that first derivative $\dot{g}(.)$ is continuous in a neighbourhood of μ, with $\dot{g}(\mu) = 0$, but second derivative exists at μ with $\ddot{g}(\mu) \neq 0$. Then

$$n(g(X_n) - g(\mu)) \xrightarrow{d} \sigma^2 \frac{\ddot{g}(\mu)}{2} X$$

where $X \sim \chi^2_1$.

Proof. Uses a second order Taylor approximation; informally

$$g(X_n) = g(\mu) + \dot{g}(\mu)(X_n - \mu) + \frac{\ddot{g}(\mu)}{2}(X_n - \mu)^2 + o_p(1)$$

thus, as $\dot{g}(\mu) = 0$,

$$g(X_n) - g(\mu) = \frac{\ddot{g}(\mu)}{2}(X_n - \mu)^2 + o_p(1)$$

and thus

$$n(g(X_n) - g(\mu)) = \frac{\ddot{g}(\mu)}{2}\{\sqrt{n}(X_n - \mu)\}^2 \xrightarrow{d} \sigma^2 \frac{\ddot{g}(\mu)}{2} Z^2$$

where $Z^2 \sim \chi^2_1$. ∎

EXAMPLES

1. Under the conditions of the Central Limit Theorem, for random variables $X_1, \ldots, X_n$ and their sample mean random variable $\overline{X}_n$

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

   Consider $g(x) = x^2$, so that $\dot{g}(x) = 2x$, and hence, if $\mu \neq 0$,

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

   and

   $$\overline{X}_n{}^2 \sim AN(\mu^2, 4\mu^2\sigma^2/n)$$

   If $\mu = 0$, we proceed by a different route to compute the approximate distribution of $\overline{X}_n{}^2$; note that, if $\mu = 0$,

   $$\sqrt{n}\overline{X}_n \xrightarrow{d} X \sim N(0, \sigma^2)$$

   so therefore

   $$n\overline{X}_n{}^2 = (\sqrt{n}\overline{X}_n)^2 \xrightarrow{d} X^2 \sim Gamma(1/2, 1/(2\sigma^2))$$

   by elementary transformation results. Hence, for large n,

   $$\overline{X}_n{}^2 \stackrel{\cdot}{\sim} Gamma(1/2, n/(2\sigma^2))$$

2. Again under the conditions of the CLT, consider the distribution of $1/\overline{X}_n$. In this case, we have a function $g(x) = 1/x$, so $\dot{g}(x) = -1/x^2$, and if $\mu \neq 0$, the Delta method gives

   $$\sqrt{n}(1/\overline{X}_n - 1/\mu) \xrightarrow{d} X \sim N(0, \sigma^2/\mu^4)$$

   or,

   $$\frac{1}{\overline{X}_n} \sim AN(1/\mu, n^{-1}\sigma^2/\mu^4).$$