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, for $y > 0$,

$$F_{Y_n}(y) = \{F_X(y)\}^n = \{1 - e^{-y}\}^n \longrightarrow 0$$

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 $n \longrightarrow \infty$,

$$\begin{aligned} F_{Z_n}(z) &= \mathrm{P}[Z_n \leq z] = \mathrm{P}[Y_n \leq z + \log n] = \{1 - e^{-z-\log n}\}^n \longrightarrow \exp\{-e^{-z}\} = F_Z(z), \\ \therefore \quad F_{Y_n}(y) &= P[Y_n \leq y] = P[Z_n \leq y - \log n] \simeq F_Z(y - \log n) = \exp\{-e^{-y+\log n}\} = \exp\{-ne^{-y}\} \end{aligned}$$

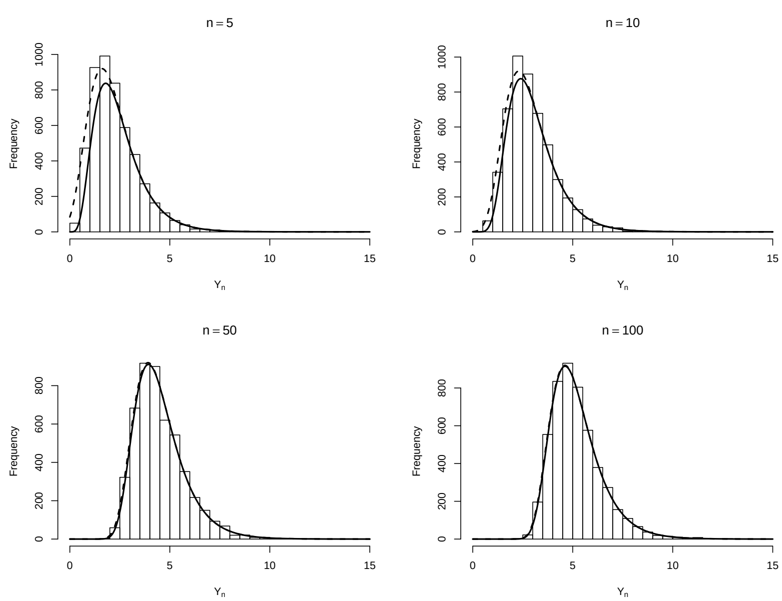
and by differentiating, for $y > 0$

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

This can be compared with the exact version, for $y > 0$

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

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

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

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

DEFINITION (Stochastic Order Notation)
For random variable Z, we write $Z = \mathrm{O}_p(1)$ if for all $\epsilon > 0$, there exists $M < \infty$ such that

$$P[|Z| \geq M] \leq \epsilon.$$

For sequence $\{Z_n\}$, write $Z_n = \mathrm{O}_p(1)$ if for all n

$$P[|Z_n| \geq M] \leq \epsilon.$$

and write $Z_n = \mathrm{O}_p(S_n)$ for sequence of random variables $\{S_n\}$ if

$$\frac{|Z_n|}{|S_n|} = \mathrm{O}_p(1).$$

Note that this includes the case where S_n is a sequence of reals, rather than random variables. Finally, write $Z_n = \mathrm{o}_p(1)$ if $Z_n \xrightarrow{p} 0$, and $Z_n = \mathrm{o}_p(S_n)$ if

$$\frac{|Z_n|}{|S_n|} = \mathrm{o}_p(1).$$

Note that

$$\mathrm{O}_p(1)\mathrm{o}_p(1) = \mathrm{o}_p(1) \qquad \mathrm{O}_p(1) + \mathrm{o}_p(1) = \mathrm{O}_p(1)$$

LEMMA
Suppose $\{X_n\}$ are a sequence of random variables, and that for real sequence $\{a_n\}$ with $a_n \longrightarrow \infty$ as $n \longrightarrow \infty$,

(i) for real constant x_0 and random variable V

$$a_n(X_n - x_0) \xrightarrow{d} V$$

(ii) real function g is differentiable at x_0, with derivative $\dot{g}$.

Then

$$a_n(g(X_n) - g(x_0)) \xrightarrow{d} \dot{g}(x_0)V$$

Proof. Note first that for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$|x - x_0| \leq \delta \quad \Longrightarrow \quad |g(x) - g(x_0) - \dot{g}(x_0)(x - x_0)| \leq \epsilon|x - x_0|$$

Now, from (i) we have

$$a_n(X_n - x_0) = \mathrm{O}_p(1) \qquad \Longrightarrow \qquad X_n - x_0 = \mathrm{O}_p(a_n^{-1}) = \mathrm{o}_p(1)$$

as $a_n \longrightarrow \infty$. Therefore, by definition, for every $\delta > 0$,

$$P[|X_n - x_0| \leq \delta] \longrightarrow 1$$

and therefore from above, for every $\epsilon > 0$,

$$P[|g(X_n) - g(x_0) - \dot{g}(x_0)(X_n - x_0)| \leq \epsilon|X_n - x_0|] \longrightarrow 1.$$

Hence

$$a_n(g(X_n) - g(x_0) - \dot{g}(x_0)(X_n - x_0)) = \mathrm{o}_p(a_n(X_n - x_0)) = \mathrm{o}_p(1)$$

Therefore

$$a_n(g(X_n) - g(x_0)) = \dot{g}(x_0)\{a_n(X_n - x_0)\} + \mathrm{o}_p(1)$$

and hence

$$a_n(g(X_n) - g(x_0)) \xrightarrow{d} \dot{g}(x_0)V.$$

∎

THEOREM (The Delta Method)
Consider sequence of random variables $\{X_n\}$ such that

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

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

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

In particular, if

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

then

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

Proof. Using the Lemma above, with

- $a_n = \sqrt{n}$
- $x_0 = \mu$
- $V = X$.

we have that

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

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

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

Note that this method does not give a useful result if $\dot{g}(\mu) = 0$. ∎

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 $\dot{g}(\underset{\sim}{\mu}) \neq \underset{\sim}{0}$. 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

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

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

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

THEOREM (The Second Order Delta Method: Normal case)
Consider sequence of random variables $\{X_n\}$ such that

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

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

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

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

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

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 $\overline{X}_n$

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

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

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

   and

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

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

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

   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,

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

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

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

   or,

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