5.5.3 Convergence in Distribution

Definition 5.5.10

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

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

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

Example (Maximum of uniforms)

If $X_1, X_2, \ldots$ are iid uniform(0,1) and $X_{(n)} = \max_{1\le i\le n} X_i$, let us examine if $X_{(n)}$ converges in distribution.

As $n \to \infty$, we have for any $\epsilon > 0$,

$$\begin{aligned} P(|X_n - 1| \ge \epsilon) &= P(X_{(n)} \le 1-\epsilon) \\ &= P(X_i \le 1-\epsilon, i = 1, \ldots, n) = (1-\epsilon)^n, \end{aligned}$$

which goes to 0. However, if we take $\epsilon = t/n$, we then have

$$P(X_{(n)} \le 1 - t/n) = (1 - t/n)^n \to e^{-t},$$

which, upon rearranging, yields

$$P(n(1 - X_{(n)}) \le t) \to 1 - e^{-t};$$

that is, the random variable $n(1-X_{(n)})$ converges in distribution to an exponential(1) random variable.

Note that although we talk of a sequence of random variables converging in distribution, it is really the cdfs that converge, not the random variables. In this very fundamental way convergence in distribution is quite different from convergence in probability or convergence almost surely.

Theorem 5.5.12

If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X.

Theorem 5.5.13

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

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

is equivalent to

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

Theorem 5.5.14 (Central limit theorem)

Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $EX_i = \mu$ and $\text{Var}X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (\frac{1}{n})\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,

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

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

Theorem 5.5.15 (Stronger form of the central limit theorem)

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

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

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

The proof is almost identical to that of Theorem 5.5.14, except that characteristic functions are used instead of mgfs.

Example (Normal approximation to the negative binomial)

Suppose $X_1, \ldots, X_n$ are a random sample from a negative binomial(r, p) distribution. Recall that

$$EX = \frac{r(1-p)}{p}, \quad \text{Var}X = \frac{r(1-p)}{p^2}$$

and the central limit theorem tells us that

$$\frac{\sqrt{n}(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}}$$

is approximately $N(0,1)$. The approximate probability calculation are much easier than the exact calculations. For example, if $r = 10$, $p = \frac{1}{2}$, and $n = 30$, an exact calculation would be

$$\begin{aligned} P(\bar{X} \le 11) &= P(\sum_{i=1}^{30} X_i \le 330) \\ &= \sum_{x=0}^{330} \binom{300 + x - 1}{x} (\frac{1}{2})^{300+x} = 0.8916 \end{aligned}$$

Note $\sum X$ is negative binomial(nr, p). The CLT gives us the approximation

$$P(\bar{X} \le 11) = P(\frac{\sqrt{30}(\bar{X} - 10)}{\sqrt{20}} \le \frac{\sqrt{30}(11 - 10)}{\sqrt{20}}) \approx P(Z \le 1.2247) = .8888.$$

Theorem 5.5.17 (Slutsky's theorem)

If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then

(a) $Y_n X_n \to aX$ in distribution.

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

Example (Normal approximation with estimated variance)

Suppose that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0,1),$$

but the value σ is unknown. We know $S_n \to \sigma$ in probability. By Exercise 5.32, $\sigma/S_n \to 1$ in probability. Hence, Slutsky's theorem tells us

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n}\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0,1).$$

5.5.4 The Delta Method

First, we look at one motivation example. Example 5.5.19 (Estimating the odds)

Suppose we observe $X_1, X_2, \ldots, X_n$ independent Bernoulli(p) random variables. The typical parameter of interest is p, but another population is $\frac{p}{1-p}$. As we would estimate p by $\hat{p} = \sum_i X_i/n$, we might consider using $\frac{\hat{p}}{1-\hat{p}}$ as an estimate of $\frac{p}{1-p}$. But what are the properties of this estimator? How might we estimate the variance of $\frac{\hat{p}}{1-\hat{p}}$?

Definition

If a function $g(x)$ has derivatives of order r, that is, $g^{(r)}(x) = \frac{d^r}{dx^r}g(x)$ exists, then for any constant a, the Taylor polynomial of order r about a is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i.$$

Theorem (Taylor)

If $g^{(r)}(a) = \frac{d^r}{dx^r}g(x)|_{x=a}$ exists, then

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0.$$

Since we are interested in approximations, we are just going to ignore the remainder. There are, however, many explicit forms, one useful one being

$$g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r dt.$$

Now we consider the multivariate case of Taylor series. Let $T_1, \dots, T_k$ be random variables with means $\theta_1, \dots, \theta_k$, and define $\mathbf{T} = (T_1, \dots, T_k)$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Suppose there is a differentiable function $g(T)$ (an estimator of some parameter) for which we want an approximate estimate of variance. Define

$$g_i'(\boldsymbol{\theta}) = \frac{\partial}{\partial t_i} g(\boldsymbol{t})|_{t_1=\theta_1,\dots,t_k=\theta_k}.$$

The first-order Taylor series expansion of g about $\boldsymbol{\theta}$ is

$$g(\boldsymbol{t}) = g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i) + \text{Remainder}.$$

From our statistical approximation we forget about the remainder and write

$$g(\boldsymbol{t}) \approx g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i).$$

Now, take expectation on both sides to get

$$E_\theta g(\mathbf{T}) \approx g(\mathbf{theta}) + \sum_{i=1}^{k} g_i'(\mathbf{theta}) E_\theta (T_i - \theta_i) = g(\mathbf{theta}).$$

We can now approximate the variance of $g(\mathbf{T})$ by

$$\begin{aligned}\mathrm{Var}_\theta g(\mathbf{T}) &\approx E_\theta([g(\mathbf{T}) - g(\mathbf{theta})]^2) \approx E_\theta\big((\sum_{i=1}^{k} g_i'(\mathbf{theta})(T_i - \theta_i)^2\big) \\ &= \sum_{i=1}^{k} [g_i'(\mathbf{theta})]^2 \mathrm{Var}_\theta T_i + 2\sum_{i>j} g_i'(\boldsymbol{\theta}) g_j'(\mathbf{theta}) \mathrm{Cov}_\theta(T_i, T_j).\end{aligned}$$

This approximation is very useful because it gives us a variance formula for a general function, using only simple variance and covariance.

Example (Continuation of Example 5.5.19)

In our above notation, take $g(p) = \frac{p}{1-p}$, so $g'(p) = \frac{1}{(1-p)^2}$ and

$$\begin{aligned}\mathrm{Var}(\frac{\hat{p}}{1-\hat{p}}) &\approx [g'(p)]^2 \mathrm{Var}(\hat{p}) \\ &[\frac{1}{(1-p)^2}]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3},\end{aligned}$$

giving us an approximation for the variance of our estimator.

Example (Approximate mean and variance)

Suppose X is a random variable with $E_\mu X = \mu \neq 0$. If we want to estimate a function $g(\mu)$, a first-order approximation would give us

$$g(X) = g(\mu) + g'(\mu)(X - \mu).$$

If we use $g(X)$ as an estimator of $g(\mu)$, we can say that approximately

$$E_\mu g(X) \approx g(\mu),$$

and

$$\mathrm{Var}_\mu g(X) \approx [g'(\mu)]^2 \mathrm{Var}_\mu X.$$

Theorem 5.5.24 (Delta method)

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ in distribution.

For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}[g(Y_n) - g(\theta)] \rightarrow N(0, \sigma^2[g'(\theta)^2])$$

in distribution.

Proof: The Taylor expansion of $g(Y_n)$ around $Y_n = \theta$ is

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \text{remainder},$$

where the remainder$\rightarrow$ 0 as $Y_n \rightarrow \theta$. Since $Y_n \rightarrow \theta$ in probability it follows that the remainder$\rightarrow$ 0 in probability. By applying Slutsky's theorem (a),

$$g'(\theta)\sqrt{n}(Y_n - \theta) \rightarrow g'(\theta)X,$$

where $X \sim N(0, \sigma^2)$. Therefore

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

$\square$

<u>Example</u>

Suppose now that we have the mean of a random sample $\bar{X}$. For $\mu \neq 0$, we have

$$\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\mu}) \rightarrow N(0, (\frac{1}{\mu})^4 \text{Var}_\mu X_1).$$

in distribution.

There are two extensions of the basic Delta method that we need to deal with to complete our treatment. The first concerns the possibility that $g'(\mu) = 0$.

<u>(Second-order Delta Method)</u>

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

$$n[g(Y_n) - g(\theta)] \rightarrow \sigma^2 \frac{g''(\theta)}{2} \chi_1^2$$

in distribution.

Next we consider the extension of the basic Delta method to the multivariate case.

Theorem 5.5.28

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample with $E(X_{ij}) = \mu_i$ and $\text{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}$. For a given function g with continuous first partial derivatives and a specific value of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ for which $\tau^2 = \sum \sum \sigma_{ij} \frac{\partial g(\mu)}{\partial \mu_i} \frac{\partial g(\mu)}{\partial \mu_j} > 0$,

$$\sqrt{n}[g(\bar{X}_1, \ldots, \bar{X}_p) - g(\mu_1, \ldots, \mu_p)] \rightarrow N(0, \tau^2)$$

in distribution.