More on Convergence and Large Sample Approximations

A. Stepanov
D. Simpson
S. Culpepper

Central Limit Theorem

$X_1, X_2, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2.

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\left(\sum_{i=1}^n X_i - n\mu\right)}{\sqrt{n}\sigma} \xrightarrow{D} Z \sim N(0,1).$$

Δ-Method $\quad \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$

$g(x)$ is differentiable at θ and $g'(\mu) \neq 0$

$$\Rightarrow \sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{D} N\left[0, \left(g'(\mu)\right)^2 \sigma^2\right]$$

Intuition:

By CLT, $\bar{X}_n - \mu$ is approximately $N\left(0, \frac{\sigma^2}{n}\right)$ for large *n*. If $g(x)$ is differentiable at μ and x is "close" to μ,

$$g(x) \approx g(\mu) + g'(\mu)(x - \mu).$$

Therefore, if $g'(\mu) \neq 0$, $g(\bar{X}_n)$ is approximately $N\left[g(\mu), \left(g'(\mu)\right)^2 \frac{\sigma^2}{n}\right]$ for large *n*.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a Geometric(*p*) distribution (the number of independent trials until the first "success"). That is,

$$P(X_i = k) = (1-p)^{k-1} p, \qquad k = 1, 2, 3, \ldots$$

Show that $\hat{p} = \tilde{p} = 1/\bar{X}_n$ is asymptotically normally distributed (as $n \to \infty$).

$$Var(X) = \frac{1-p}{p^2}$$

By CLT, $\quad \sqrt{n}\left(\bar{X}_n - \frac{1}{p}\right) \xrightarrow{D} N\left(0, \frac{1-p}{p^2}\right).$

$g(x) = \frac{1}{x}$ is differentiable at $\frac{1}{p}$, $g'\left(\frac{1}{p}\right) = -p^2 \neq 0.$

$$\Rightarrow \sqrt{n}\left(g(\bar{X}_n) - g\left(\frac{1}{p}\right)\right) \xrightarrow{D} N\left(0, (-p^2)^2 \frac{1-p}{p^2}\right)$$

$$\Rightarrow \sqrt{n}(\hat{p} - p) \xrightarrow{D} N\big(0, p^2(1-p)\big).$$

$\Rightarrow$ For large n, $\hat{p} \sim N\left(p, \frac{p^2(1-p)}{n}\right)$.

$$\Rightarrow \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}^2(1-\hat{p})}{n}}$$

would have approximate 100(1 – α) % confidence level for large n.

Example 2. Let $X_n \sim \chi^2(n)$. What is the limiting distribution of $W_n = \sqrt{X_n} - \sqrt{n}$?

Hint: We already know that

(a) $Y_n = \frac{X_n}{n} \xrightarrow{P} 1$

(b) $Z_n = \frac{X_n - n}{\sqrt{2n}} = \sqrt{\frac{n}{2}}\left(\frac{X_n}{n} - 1\right) \xrightarrow{D} N(0,1)$.

$$W_n = \sqrt{X_n} - \sqrt{n} = \sqrt{n}\left(\sqrt{\frac{X_n}{n}} - 1\right)$$

$$Z_n = \frac{X_n - n}{\sqrt{2n}} = \sqrt{\frac{n}{2}}\left(\frac{X_n}{n} - 1\right) \xrightarrow{D} N(0,1) \Rightarrow \sqrt{n}\left(\frac{X_n}{n} - 1\right) \xrightarrow{D} N(0,2)$$

Let $g(x) = \sqrt{x}$. Then $g(x)$ is differentiable, $g'(x) = \frac{1}{2} x^{-\frac{1}{2}}$ and $g'(1) = \frac{1}{2}$.

$$\Rightarrow W_n = \sqrt{X_n} - \sqrt{n} = \sqrt{n}\left(\sqrt{\frac{X_n}{n}} - \sqrt{1}\right) \xrightarrow{D} N\left(0, \left(g'(1)\right)^2 \cdot 2\right) = N\left(0, \frac{1}{2}\right).$$

Example 3. Consider the sample mean $\bar{X}_n$ from a random sample of size n from a distribution with pdf $f(x) = e^{-x},\ x > 0$.

a) Obtain the MGF for $Y_n = \sqrt{n}(\bar{X}_n - 1)$

$$M_X(t) = (1-t)^{-1}, \qquad t < 1$$

$$M_{Y_n}(t) = E\left[e^{t\sqrt{n}(\bar{X}_n - 1)}\right] = e^{-t\sqrt{n}} E\left[\exp\left(\frac{t}{\sqrt{n}} \sum_i^n X_i\right)\right]$$

$$= e^{-t\sqrt{n}} \left[M_X\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{-t\sqrt{n}} \left(1 - \frac{t}{\sqrt{n}}\right)^{-n}$$

$$= \left(e^{\frac{t}{\sqrt{n}}} - \frac{t}{\sqrt{n}} e^{\frac{t}{\sqrt{n}}}\right)^{-n}, \frac{t}{\sqrt{n}} < 1.$$

b) Find the limiting distribution of Y_n via the MGF

$$e^{\frac{t}{\sqrt{n}}} = 1 + \frac{t}{\sqrt{n}} + \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

$$\Rightarrow M_{Y_n}(t) = \left(1 + \frac{t}{\sqrt{n}} + \frac{t^2}{2n} - \frac{t}{\sqrt{n}} - \frac{t^2}{n} + o\left(\frac{1}{n}\right)\right)^{-n} = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^{-n}$$

As $n \to \infty$, $M_{Y_n}(t) \to e^{\frac{t^2}{2}} = M_Z(t)$, where $Z \sim N(0,1)$.

$$\Rightarrow Y_n \xrightarrow{D} Z \sim N(0,1).$$

Example 4. Let $X_1, \dots, X_n$ be a random sample from the distribution with probability density function

$$f_X(x;\theta) = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, 0 < x < 1, 0 < \theta < \infty$$

Let $Y_1 < Y_2 < \cdots Y_n$ denote the corresponding order statistics.

a) The *method of moments* estimator of θ is $\tilde{\theta} = \frac{1-\bar{X}}{\bar{X}} = \frac{1}{\bar{X}} - 1$. Determine its large sample behavior.

What is method of moments? Note that

$$\mu = E(X) = (1+\theta)^{-1}.$$

If we don't know θ consider inserting the sample mean for μ:

$$\bar{X}_n = \frac{1}{1+\tilde{\theta}} \Leftrightarrow \tilde{\theta} = \frac{1-\bar{X}_n}{\bar{X}_n}$$

Using the weak law of large numbers we know $\bar{X}_n \xrightarrow{P} \mu$ so we conclude that

$$\tilde{\theta} \xrightarrow{P} \frac{1-\mu}{\mu} = \theta$$

In other words, the MoM estimator is a *consistent estimator* of θ.

Next, show that $\tilde{\theta}$ is asymptotically normally distributed as $n \to \infty$.

For the CLT find the variance,

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \int_0^1 x^2 \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} dx - \frac{1}{(1+\theta)^2}$$

$$= \frac{1}{1+2\theta} - \frac{1}{(1+\theta)^2} = \frac{\theta^2}{(1+2\theta)(1+\theta)^2}.$$

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

Since $g(x) = \frac{1-x}{x}$ is differentiable at $\mu = (1+\theta)^{-1}$, $g'(\mu) = -(1+\theta)^{-2} \neq 0$,

$$\Rightarrow \sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{D} N\left[0, (-(1+\theta)^{-2})^2 \frac{\theta^2}{(1+2\theta)(1+\theta)^2}\right]$$

$$\Rightarrow \sqrt{n}[\tilde{\theta} - \theta] \xrightarrow{D} N\left[0, \frac{\theta^2(1+\theta)^2}{1+2\theta}\right].$$

$\Rightarrow$ For large n,

$$\tilde{\theta} \sim N\left[\theta, \frac{\theta^2(1+\theta)^2}{(1+2\theta)n}\right]$$

b) Suggest a $100(1-\alpha)\%$ confidence interval for θ.

For large n,

$$\tilde{\theta} = \frac{1-\bar{X}_n}{\bar{X}_n} \sim N\left[\theta, \frac{\theta^2(1+\theta)^2}{(1+2\theta)n}\right]$$

$$\tilde{\theta} \pm z_{\alpha/2} \frac{\tilde{\theta}(1+\tilde{\theta})}{\sqrt{(1+2\tilde{\theta})n}}$$

would have an approximate $100(1-\alpha)\%$ confidence level for large n.

c) The maximum likelihood estimator of θ, $\hat{\theta} = -\frac{1}{n}\sum_i^n \ln X_i$, is a consistent estimator of θ. Show that $\hat{\theta}$ is asymptotically normally distributed (as $n \to \infty$).

Find the parameters.

Let $Y_i = -\ln X_i, i = 1, \dots, n$. Then $E(Y) = \theta, Var(Y) = \theta^2$.

By CLT, $\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{D} N(0, \sigma_Y^2)$.

$$\Rightarrow \sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{D} N(0, \theta^2)$$

$\Rightarrow$ For large n, $\hat{\theta} \sim N\left(\theta, \frac{\theta^2}{n}\right)$

d) Suggest a $100\,(1 - \alpha)\,\%$ confidence interval for θ.

For large n, $\hat{\theta} = -\frac{1}{n}\sum_i^n \ln X_i \sim N\left(\theta, \frac{\theta^2}{n}\right)$

$$\Rightarrow P\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\frac{\theta}{\sqrt{n}}} z_{\alpha/2}\right) \approx 1 - \alpha$$

$$\Rightarrow P\left(\frac{\hat{\theta}}{1 + \frac{z_{\alpha/2}}{\sqrt{n}}} < \theta < \frac{\hat{\theta}}{1 - \frac{z_{\alpha/2}}{\sqrt{n}}}\right) \approx 1 - \alpha$$

$\Rightarrow \left(\frac{\hat{\theta}}{1 + \frac{z_{\alpha/2}}{\sqrt{n}}}, \frac{\hat{\theta}}{1 - \frac{z_{\alpha/2}}{\sqrt{n}}}\right)$ would have an approximate $100(1-\alpha)\,\%$ confidence level for large n.

OR

For large n, $\hat{\theta} = -\frac{1}{n}\sum_{i}^{n} \ln X_i \sim N\left(\theta, \frac{\theta^2}{n}\right)$

$\hat{\theta} \pm z_{\alpha/2}\frac{\hat{\theta}}{\sqrt{n}}$ would have an approximate 100(1–α) % confidence level for large n.

OR

For all *n*, let $Y_i = -\ln X_i\,, i = 1, \dots, n$.

$$F_X(x) = x^{\frac{1}{\theta}}, 0 < x < 1$$

$$\Rightarrow F_Y(y) = P(Y \le y) = P(X \ge e^{-y}) = 1 - e^{-\frac{y}{\theta}}, y > 0$$

$$\Rightarrow Y_1, \dots, Y_n \text{ are i.i.d. } Exponential(\theta)$$

$$M_{\hat{\theta}}(t) = M_{\bar{Y}}(t) = \left[M_Y\left(\frac{t}{n}\right)\right]^n = \frac{1}{\left(1 - \frac{\theta}{n}t\right)^n}, t < \frac{n}{\theta}.$$

$$\Rightarrow \hat{\theta} \sim Gamma\left(\alpha = n, \frac{\theta}{n}\right)$$

$$\Rightarrow \frac{2n}{\theta}\hat{\theta} \sim \chi^2(2n)$$

$$\Rightarrow P\left(\chi^2_{1-\alpha/2}(2n) < \frac{2n}{\theta}\hat{\theta} < \chi^2_{\alpha/2}(2n)\right) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{2n\hat{\theta}}{\chi^2_{\alpha/2}(2n)} < \theta < \frac{2n\hat{\theta}}{\chi^2_{1-\alpha/2}(2n)}\right) = 1 - \alpha$$

$\Rightarrow \left(\frac{2n\hat{\theta}}{\chi^2_{\alpha/2}(2n)} < \theta < \frac{2n\hat{\theta}}{\chi^2_{1-\alpha/2}(2n)}\right)$ would have a 100(1–α) % confidence level **for any *n***.

Example 5. Let $\lambda > 0$ and let $X_1, \dots, X_n$ be a random sample from the distribution with the probability density function

$$f(x;\lambda) = 2\lambda^2 x^3 e^{-\lambda x^2}, x > 0$$

Recall:

$$Y_n = \sum_{i=1}^{n} X_i^2 \sim Gamma\left(\alpha = 2n, \text{"}usual\ \theta\text{"} = \frac{1}{\lambda}\right).$$

a) Suggest a confidence interval for λ with $(1-\alpha)100\%$ confidence level.

If $Y_n \sim Gamma\left(\alpha = 2n, \text{"}usual\ \theta\text{"} = \frac{1}{\lambda}\right)$, where α is an integer, then

$$\frac{2}{\theta} Y_n = 2\lambda Y_n = 2\lambda \sum_{i=1}^{n} X_i^2 \sim \chi^2(4n)$$

$$\Rightarrow P\left(\chi^2_{1-\alpha/2}(4n) < 2\lambda \sum_{i=1}^{n} X_i^2 < \chi^2_{\alpha/2}(4n)\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{\chi^2_{1-\alpha/2}(4n)}{2\sum_{i=1}^{n} X_i^2} < \lambda < \frac{\chi^2_{\alpha/2}(4n)}{2\sum_{i=1}^{n} X_i^2}\right) = 1-\alpha$$

A 100(1–α)% confidence interval for λ, $\left(\frac{\chi^2_{1-\alpha/2}(4n)}{2\sum_{i=1}^{n} X_i^2}, \frac{\chi^2_{\alpha/2}(4n)}{2\sum_{i=1}^{n} X_i^2}\right)$.

b) Suppose $n = 5$, and $x_1 = 0.6$, $x_2 = 1.1$, $x_3 = 2.7$, $x_4 = 3.3$, $x_5 = 4.5$

$$\sum_{i=1}^{5} x_i^2 = 40.$$

(i) Use part (a) to construct a 90% confidence interval for λ.

$$\chi^2_{0.95} = 10.85, \chi^2_{0.05} = 31.41$$

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2}}(4n)}{2\sum_{i=1}^n X_i^2}, \frac{\chi^2_{\frac{\alpha}{2}}(4n)}{2\sum_{i=1}^n X_i^2}\right) = \left(\frac{10.85}{2\cdot 40}, \frac{31.41}{2\cdot 40}\right) \approx (\mathbf{0.1356}, \mathbf{0.3926}).$$

(ii) Use part (a) to construct a 95% confidence interval for λ.

$$\chi^2_{0.95} = 9.591, \chi^2_{0.05} = 34.17$$

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2}}(4n)}{2\sum_{i=1}^n X_i^2}, \frac{\chi^2_{\frac{\alpha}{2}}(4n)}{2\sum_{i=1}^n X_i^2}\right) = \left(\frac{9.591}{2\cdot 40}, \frac{34.17}{2\cdot 40}\right) \approx (\mathbf{0.120}, \mathbf{0.427}).$$

Recall:

$$\hat{\lambda} = \frac{2n}{\sum_{i=1}^n X_i^2} = 0.25$$

Example 6. Let $X_1, \dots, X_n$ be a random sample of size n such that $X_i \sim U(0,\theta)$.

$$f(x:\theta) = \begin{cases} \frac{1}{\theta} & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases} \qquad F(x:\theta) = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & 0 \le x < \theta \\ 1 & x \ge \theta \end{cases}$$

$$E(X) = \frac{\theta}{2}, Var(X) = \frac{\theta^2}{12}$$

The method of moments estimator of θ is $\tilde{\theta} = 2\bar{X}_n$.

By CLT,

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

$\Rightarrow$ For large n, $\tilde{\theta} \sim N\left(\theta, \frac{\theta^2}{3n}\right)$.

$$\Rightarrow \left(\frac{\tilde{\theta}}{1 + \frac{1}{\sqrt{3n}} z_{\alpha/2}}, \frac{\tilde{\theta}}{1 - \frac{1}{\sqrt{3n}} z_{\alpha/2}} \right)$$

would have an approximate 100(1–α) % confidence level for large n.

OR

$$\tilde{\theta} \pm z_{\alpha/2} \frac{\tilde{\theta}}{\sqrt{3n}}$$

would have an approximate 100(1–α) % confidence level for large n.

OR

The maximum likelihood estimator of θ is $\hat{\theta} = \max X_i$.

$$F_{\hat{\theta}}(x) = F_{\max X_i}(x) = [F_X(x)]^n = \left(\frac{x}{\theta}\right)^n, 0 < x < \theta.$$

$$f_{\hat{\theta}}(x) = f_{\max X_i}(x) = n\frac{x^{n-1}}{\theta^n}, 0 < x < \theta.$$

$$P\left(c\theta < \hat{\theta} < \theta\right) = F_{\max X_i}(\theta) - F_{\max X_i}(c\theta) = 1 - c^n$$

$$\Rightarrow P\left(\max X_i < \theta < \frac{\max X_i}{c}\right) = 1 - c^n$$

$$\Rightarrow \left(\max X_i\,, \frac{\max X_i}{\alpha^{\frac{1}{n}}}\right)$$ has 100(1 – α)% confidence level for any n

OR

Recall $n(\theta - \max X_i) \xrightarrow{D} Exponential(\theta)$. (Example 5a in convergence1 notes)

Also, recall that the p^{th} quantile, π_p, for the exponential distribution is defined as,

$$F(\pi_{1-\alpha/2}) = 1 - \frac{\alpha}{2} \Rightarrow \pi_{1-\alpha/2} = F^{-1}\left(1 - \frac{\alpha}{2}\right)$$

where the inverse exponential cdf is $\pi_p = F^{-1}(p) = -\theta\ln(1-p)$.

$$\Rightarrow P\left[-\theta\ln\left(1 - \frac{\alpha}{2}\right) < n(\theta - \max X_i) < -\theta\ln\left(\frac{\alpha}{2}\right)\right] \approx 1 - \alpha \text{ for large } n.$$

$$\Rightarrow P\left[\frac{\max X_i}{1 + \frac{1}{n}\ln\left(1 - \frac{\alpha}{2}\right)} < \theta < \frac{\max X_i}{1 + \frac{1}{n}\ln\left(\frac{\alpha}{2}\right)}\right] \approx 1 - \alpha$$

$$\left(\frac{\max X_i}{1 + \frac{1}{n}\ln\left(1 - \frac{\alpha}{2}\right)}, \frac{\max X_i}{1 + \frac{1}{n}\ln\left(\frac{\alpha}{2}\right)}\right)$$ would have approximate 100(1 – α) % confidence level for large n.

Example 7. Let $X \sim Binomial(n, p)$. Recall that for large n,

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1),$$

and

$$P\left(-z_{\alpha/2} < \frac{X - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

Show that an approximate $100(1 - \alpha)\%$ confidence interval for p is,

$$\frac{\hat{p} + \frac{1}{2n} z_{\alpha/2}^2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{1}{4n} z_{\alpha/2}^2}}{1 + \frac{1}{n} z_{\alpha/2}^2}, \qquad \hat{p} = \frac{X}{n}.$$

This interval is called the Wilson interval. Note that for large n, this interval is approximately equal to

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

$$\left|\frac{X - np}{\sqrt{np(1-p)}}\right| < z_{\alpha/2} \Leftrightarrow \frac{(X - np)^2}{np(1-p)} < z_{\alpha/2}^2$$

$$\Leftrightarrow X^2 - 2npX + n^2p^2 < npz_{\alpha/2}^2 - np^2 z_{\alpha/2}^2$$

$$\Leftrightarrow \hat{p}^2 - 2\hat{p}p + p^2 < \frac{z_{\alpha/2}^2}{n} p - \frac{z_{\alpha/2}^2}{n} p^2$$

$$\Leftrightarrow \left(1+\frac{z_{\alpha/2}^2}{n}\right)p^2-\left(2\hat{p}+\frac{z_{\alpha/2}^2}{n}\right)p+\hat{p}^2<0$$

$$ap^2-bp+c=0 \Rightarrow p_1,p_2=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

$$ap^2-bp+c=<0 \Leftrightarrow p_1<p<p_2 \qquad (a>0)$$

$$p_1,p_2=\frac{\left(2\hat{p}+\frac{z_{\alpha/2}^2}{n}\right)\pm\sqrt{\left(2\hat{p}+\frac{z_{\alpha/2}^2}{n}\right)^2-4\left(1+\frac{z_{\alpha/2}^2}{n}\right)\hat{p}^2}}{2\left(1+\frac{z_{\alpha/2}^2}{n}\right)}$$

$$=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}\pm\sqrt{\left(\hat{p}+\frac{z_{\alpha/2}^2}{2n}\right)^2-\left(1+\frac{z_{\alpha/2}^2}{n}\right)\hat{p}^2}}{1+\frac{z_{\alpha/2}^2}{n}}$$

$$=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}\pm\sqrt{\hat{p}^2+2\hat{p}\frac{z_{\alpha/2}^2}{2n}+\frac{z_{\alpha/2}^4}{4n^2}-\left(1+\frac{z_{\alpha/2}^2}{n}\right)\hat{p}^2}}{1+\frac{z_{\alpha/2}^2}{n}}$$

$$=\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}\pm z_{\alpha/2}^2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z_{\alpha/2}^2}{4n^2}}}{1+\frac{z_{\alpha/2}^2}{n}}$$