Central Limit Theorem and approximate confidence intervals

- Convergence in distribution. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of random vectors in R^k and X is a random vector in R^k with CDF F_X. If

$$\lim_{n\to\infty} P(X_n \le t) = F_X(t)$$

for every t that is a continuity point of F_X, then we say that X_n converges in distribution to X as $n \to \infty$, denoted by

$$X_n \stackrel{\mathcal{D}}{\to} X \text{ or } X_n \stackrel{\mathcal{D}}{\to} D_X,$$

where D_X is the distribution of X. We will say that D_X is the limiting distribution of X_n.

- Note. $X_n \stackrel{\mathcal{P}}{\to} X$ implies that $X_n \stackrel{\mathcal{D}}{\to} X$ but not vice versa. However, we have the following result:

Fact 1 *Suppose that* $X_n \stackrel{\mathcal{D}}{\to} 0$ *as* $n \to \infty$*, then* $X_n \stackrel{\mathcal{P}}{\to} 0$ *as* $n \to \infty$.

Fact 1 is a special case of Theorem 5.2.2 in the text and the proof can be found thereof.

- The following result is known as Slutsky's theorem.

Theorem. *Suppose that* $X_n \stackrel{\mathcal{D}}{\to} X$ *and* $Y_n \stackrel{\mathcal{P}}{\to} c$ *as* $n \to \infty$*, where* c *is a constant. Then*

(i) $X_n + Y_n \stackrel{\mathcal{D}}{\to} X + c$ *as* $n \to \infty$*, and*

(ii) $X_n Y_n \stackrel{\mathcal{D}}{\to} cX$ *as* $n \to \infty$.

The proof of Slutsky's theorem (Theorem 5.2.5 in the text) is omitted.

- Example 1. Suppose that $X_n \stackrel{\mathcal{D}}{\to} N(0, \sigma^2)$ as $n \to \infty$, where $\sigma > 0$. Suppose that $\{\sigma_n\}_{n=1}^{\infty}$ is a sequence of positive numbers so that $\lim_{n\to\infty} \sigma_n = \sigma$. Find the limiting distribution of X_n/σ_n as $n \to \infty$.

Sol. Let Z be a random variable such that $Z \sim N(0, \sigma^2)$, then $X_n \stackrel{\mathcal{D}}{\to} Z$ as $n \to \infty$. Take $Y_n = 1/\sigma_n$ for all n, then $Y_n \stackrel{\mathcal{P}}{\to} 1/\sigma$ as $n \to \infty$. Apply Slutsky's theorem, we have

$$X_n Y_n \stackrel{\mathcal{D}}{\to} \sigma^{-1} Z$$

as $n \to \infty$. Since $\sigma^{-1} Z \sim N(0, 1)$, $N(0, 1)$ is the limiting distribution of $X_n Y_n = X_n/\sigma_n$ as $n \to \infty$.

- The following result is used to establish convergence in probability under a continuous transformation.

Fact 2 *Suppose that* $X_n \xrightarrow{\mathcal{P}} c$ *as* $n \to \infty$ *and* g *is continuous at* c*. Then* $g(X_n) \xrightarrow{\mathcal{P}} g(c)$ *as* $n \to \infty$*. Here each one of* c *and* $g(c)$ *can be a constant or a vector of constants.*

The proof of Fact 2 is based on the definition of convergence in probability.

- Example 2. Suppose that $(X_1, \ldots, X_n)$ is a random sample, both $E(X_1)$ and $Var(X_1)$ are finite, and $Var(X_1) > 0$. Let $\mu = E(X_1)$ and $\sigma^2 = Var(X_1)$. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$ and $\bar{Y}_n = \sum_{i=1}^n X_i^2/n$, then from the WLLN, $\bar{X}_n$ and $\bar{Y}_n$ are consistent estimators of μ and $\sigma^2 + \mu^2$ respectively. Find a consistent estimator of σ.

  Ans. $(\bar{Y}_n - \bar{X}_n^2)^{1/2}$.

- Convergence in distribution for the univariate case can be established using convergence of MGFs according to Theorem 3 in [1]. Below is a modified version.

  Fact 3 *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of random variables and* X *is a random variable. Let* M_{X_n} *and* M_X *denote the MGFs of* X_n *and* X *respectively. If there exists* n_0 *and* $\delta > 0$ *such that* M_X *and* M_{X_n} *are finite on* $(-\delta, \delta)$ *for* $n \geq n_0$*, and*

  $$\lim_{n \to \infty} M_{X_n}(t) = M_X(t)$$

  for $t \in (-\delta, \delta)$*, then* $X_n \xrightarrow{\mathcal{D}} X$ *as* $n \to \infty$*.*

- The following is a multivariate version of Central Limit Theorem (CLT).

  Theorem. *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of IID random vectors. Let* μ *be the mean vector of* X_1 *and* Σ *be the covariance matrix of* X_1*. Suppose that all elements in* μ *and* Σ *are finite. Let* $\bar{X} = \sum_{i=1}^n X_i/n$*, then*

  $$\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{D}} N(0, \Sigma) \text{ as } n \to \infty.$$

- The following is a univariate version of Central Limit Theorem (CLT).

  Theorem. *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of IID random variables. Let* $\mu = E(X_1)$ *and* $\sigma^2 = Var(X_1)$*. Suppose that both* μ *and* Σ *are finite. Let* $\bar{X} = \sum_{i=1}^n X_i/n$*, then*

  $$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1) \text{ as } n \to \infty.$$

  We can prove the univariate version of CLT using Fact 3 assuming that there exists $\delta > 0$ such that the MGF of X_1 is finite on $(-\delta, \delta)$.

- Example 3. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Let $\mu_2 = E(X_1^2) = \mu + \sigma^2$ and $Y_i = X_i^2$ for $i = 1, \ldots, n$. Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n X_i^2/n$. Find the limiting distribution of $(\sqrt{n}(\bar{X} - \mu), \sqrt{n}(\bar{Y} - \mu_2))^T$ as $n \to \infty$.

  Ans. $N((0,0)^T, \Sigma)$, where

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, Y_1) \\ Cov(X_1, Y_1) & Var(Y_1) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 4\mu^2\sigma^2 + 2\sigma^4 \end{pmatrix}.$$

- The following result is known as Delta method.

  Theorem. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of $k \times 1$ random vectors such that*

$$\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{D}} Z$$

  as $n \to \infty$, where Z is a $k \times 1$ random vector. Suppose that g is a differentiable function such that $g(X_n)$ is defined, then

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} (D_g(\mu))^T Z \qquad (1)$$

  as $n \to \infty$, where the column vector $D_g(\mu)$ is the gradient of g evaluated at μ.

  Note.

  - When $k = 1$, (1) becomes

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} g'(\mu) Z$$

    as $n \to \infty$. In such case, Delta method can be proved using Slutsky's theorem, Fact 1 and Fact 2.

- Example 4. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $\Gamma(\theta, 1)$, where $\theta > 0$.

  (a) Find T_n: an estimator of θ based on $(X_1, \ldots, X_n)$ so that

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

  for some $\sigma > 0$. Express σ as a function of θ.

  (b) Based on the estimator T_n in Part (a), find W_n: an estimator of θ based on $(X_1, \ldots, X_n)$ so that

$$\sqrt{n}(W_n - \theta^2) \xrightarrow{\mathcal{D}} N(0, \tau^2)$$

  for some $\tau > 0$. Express τ as a function of θ.

  Ans. (a) We can take $T_n = \sum_{i=1}^n X_i/n$, then $\sigma = \sqrt{\theta}$. (b) $W_n = T_n^2$, $\tau = 2\theta\sigma = 2\theta^{3/2}$.

- Suppose that we would like to construct a confidence interval of $g(\theta)$ based on a sample $X = (X_1, \ldots, X_n)$. Suppose that we can find a function h so that
$$h(n, X, g(\theta)) \xrightarrow{\mathcal{D}} D_0$$
as $n \to \infty$, where D_0 is some distribution that does not depend on θ. Then an approximate confidence interval of $g(\theta)$ can be constructed by treating $h(n, X, g(\theta))$ as a pivot with distribution D_0. This approach can be justified by the following result when D_0 has a continuous CDF:

Fact 4 *Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables and X is a random variable such that $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$. Suppose that the CDF of X is continuous on $(-\infty, \infty)$. Then for every interval $I \subset (-\infty, \infty)$,*
$$\lim_{n\to\infty} P(X_n \in I) = P(X \in I).$$

The proof of Fact 4 can be established if
$$\lim_{n\to\infty} P(X_n = c) = 0 = P(X = c) \tag{2}$$
for every constant c. The reason is that $P(X \in I)$ can be computed using $P(X \le x)$ and $P(X = c)$ for some c, x, and we have
$$\lim_{n\to\infty} P(X_n \le x) = P(X \le x)$$
for every x since $X_n \xrightarrow{\mathcal{D}} X$ and the CDF of X is continuous everywhere. To prove (2), note that
$$\begin{aligned}\limsup_{n\to\infty} P(X_n = c) &\le \limsup_{n\to\infty} [P(X_n \le c) - P(X_n \le c - h)] \\ &= P(X \le c) - P(X \le c - h)\end{aligned}$$
for every $h > 0$. Since the CDF of X is continuous at c, $\lim_{h\to 0^+} P(X \le c) - P(X \le c - h) = 0$, so $\limsup_{n\to\infty} P(X_n = c) \le 0$. It is clear that $\liminf_{n\to\infty} P(X_n = c) \ge 0$, so we must have
$$\liminf_{n\to\infty} P(X_n = c) = \limsup_{n\to\infty} P(X_n = c) = 0.$$

- Example 5. In Example 4, construct an approximate $(1-\alpha)$ confidence interval of θ for $\alpha \in (0, 1)$.

Sol. Let $\bar{X} = \sum_{i=1}^{n} X_i/n$. In Example 4, we have for $Z \sim N(0, \theta)$,
$$\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{D}} Z$$
as $n \to \infty$. By WLLN, $\bar{X} \xrightarrow{\mathcal{P}} E(X_1) = \theta$, so
$$\frac{1}{\sqrt{\bar{X}}} \xrightarrow{\mathcal{P}} \frac{1}{\sqrt{\theta}},$$

which implies that

$$\sqrt{n}(\bar{X}-\theta)\left(\frac{1}{\sqrt{\bar{X}}}\right) \xrightarrow{\mathcal{D}} \left(\frac{1}{\sqrt{\theta}}\right) Z \sim N(0,1) \text{ as } n\to\infty$$

by Slutsky's theorem. That is,

$$\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}}} \xrightarrow{\mathcal{D}} N(0,1) \text{ as } n\to\infty.$$

Let $z_{\alpha/2}$ be the $(1-\alpha/2)$ quantile of $N(0,1)$, then $P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 1-\alpha$, so

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

for large n. Since

$$\begin{aligned} & -z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}}} < z_{\alpha/2} \\ \Leftrightarrow\ & \theta \in \left(\bar{X} - \frac{z_{\alpha/2}\sqrt{\bar{X}}}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2}\sqrt{\bar{X}}}{\sqrt{n}}\right), \end{aligned}$$

an approximate $(1-\alpha)$ confidence interval of θ is

$$\left(\bar{X} - \frac{z_{\alpha/2}\sqrt{\bar{X}}}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2}\sqrt{\bar{X}}}{\sqrt{n}}\right).$$

- Example 6. Write down R scripts to find estimated coverage probability for the approximate $(1-\alpha)$ confidence interval of θ in Example 5 for $\alpha = 0.05$, $n = 100$ and $\theta \in \{1, 10\}$. The coverage probability is estimated by $\hat{p}$, where $\hat{p}$ is obtained by carrying out Steps (a)–(d):

  (a) Generate 500 random samples of size n from $\Gamma(\theta, 1)$.

  (b) Compute the 500 observed confidence intervals.

  (c) Compute N: the number of observed confidence intervals that contains θ.

  (d) Take $\hat{p} = N/500$.

  Sol. We first write an R function `test` with two input variables `theta` and `x`, where `x` is a sample. The function output is 1 if the observed confidence interval of θ contains the input `theta` and is 0 otherwise.

```
test <- function(theta, x){
 alpha <- 0.05
```

```
 z <- qnorm(1-alpha/2)
 n <- length(x)
 x.bar <- mean(x)
 d <- z*sqrt(x.bar)/sqrt(n)
 ci.lb <- x.bar - d
 ci.ub <- x.bar + d
 if ( (ci.lb < theta)&(theta < ci.ub) ) { ans <- 1 } else { ans <- 0 }
 return(ans)
}
```

The R scripts for finding estimated estimated coverage probability when $n = 100$ and $\theta = 1$ is given below:

```
theta <- 1
n <- 100
set.seed(1)
res <- rep(0, 500)   #vector for storing coverage results for the 500 samples
for ( i in 1:500){
 x <- rgamma(n, shape = theta, scale =1)
 res[i] <- test(theta, x)
}
mean(res) #0.95
```

The estimated coverage probability when $n = 100$ and $\theta = 10$ can be obtained by replacing

```
theta <- 1
```

with

```
theta <- 10
```

Reference

[1] J. H. Curtiss. A note on the theory of moment generating functions. *The Annals of Mathematical Statistics*, 13(4):430–433, 1942. https://www.jstor.org/stable/2235846.