Estimation and types of convergence

- When we have IID data $X_1, \ldots, X_n$, $(X_1, \ldots, X_n)$ is called a random sample and n is called the sample size of the random sample. In such case, if the distribution of X_i is D, then we say that $(X_1, \ldots, X_n)$ is a random sample from the distribution D.

- Suppose that $(X_1, \ldots, X_n)$ is a random sample from an unknown distribution D, then we can estimate D based on the random sample. A common approach is to assume that D is determined by a vector $\theta \in R^k$ for some k. Then the problem of estimating D becomes the problem of estimating θ. In such case, the estimation problem is known as a parametric estimation problem and θ is called a parameter vector.

  - Example of a parametric estimation problem. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in R$ and $\sigma > 0$ are unknown. Let $\theta = (\mu, \sigma)$. Then we only need to estimate θ based on the sample to learn about the distribution of X_1.

- Identifiability. Suppose that $\mathcal{C}$ is a collection of distributions given by $\mathcal{C} = \{Q_\theta : \theta \in \Theta\}$, where Θ is a specified subset in R^k. Then $\mathcal{C}$ is also called a family of distributions. The family $\mathcal{C}$ is identifiable means that for $\theta_1, \theta_2 \in \Theta$,
$$Q_{\theta_1} = Q_{\theta_2} \Rightarrow \theta_1 = \theta_2.$$

- Example 1. Let $\mathcal{C} = \{N(\mu, \sigma^2) : (\mu, \sigma) \in R^2\}$. Then the family $\mathcal{C}$ is not identifiable. To see this, let $(\mu_1, \sigma_1) = (0, 1)$ and $(\mu_2, \sigma_2) = (0, -1)$, then (μ_1, σ_1), (μ_2, σ_2) are two points in R^2 such that
$$N(\mu_1, \sigma_1^2) = N(0, 1) = N(\mu_2, \sigma_2^2)$$
yet
$$(\mu_1, \sigma_1) \neq (\mu_2, \sigma_2).$$
Therefore, the family $\mathcal{C}$ is not identifiable.

- Estimator. Suppose that $(X_1, \ldots, X_n)$ is a random sample from a distribution determined by a parameter vector θ. To estimate $g(\theta)$: some quantity determined by θ, we usually use some quantity that can approximate $g(\theta)$ well and be computed based on the sample. Such a quantity is called an estimator of $g(\theta)$ ($g(\theta)$的估計量). A quantity that can be computed based on the sample is called a statistic (統計量).

- To establish good approximation property of an estimator, we often need to apply some results about the convergence of a sequence of random variables (or random vectors), including
  - LLN (law of large numbers)
  - CLT (central limit theorem)

  - continuous mapping theorem
  - Slutsky's theorem
  - Delta method

- We will learn about three types of convergence of sequences of random vectors.

  - Almost surely convergence
  - Convergence in probability
  - Convergence in distribution

- Almost surely convergence. Suppose that $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random vectors in R^k on the probability space $(\Omega, \mathcal{F}, P)$ and Y is also a random vector in R^k on $(\Omega, \mathcal{F}, P)$. If

$$P\left(\{w \in \Omega : \lim_{n\to\infty} Y_n(w) = Y(w)\}\right) = 1, \tag{1}$$

  then we say that Y_n converges to Y almost surely as $n \to \infty$. Here the distance between $Y_n(w)$ and $Y(w)$ is $\|Y_n(w) - Y(w)\|$, where $\|\cdot\|$ denotes the Euclidean norm.

- Convergence in probability. Suppose that $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random vectors in R^k on the probability space $(\Omega, \mathcal{F}, P)$ and Y is also a random vector in R^k on $(\Omega, \mathcal{F}, P)$. If

$$\lim_{n\to\infty} P\left(\{w \in \Omega : \|Y_n(w) - Y(w)\| > \varepsilon\}\right) = 0 \text{ for every } \varepsilon > 0, \tag{2}$$

  then we say that Y_n converges to Y in probability as $n \to \infty$. The convergence is often denoted by $Y_n \stackrel{\mathcal{P}}{\to} Y$ as $n \to \infty$.

- Note. It can be shown that (1) implies that (2).

- Strong law of large numbers (SLLN, 強大數法則). Suppose that $X_1, \ldots, X_n$ are IID random variables and $E(X_1)$ is finite. Let $\bar{X} = \sum_{i=1}^n X_i/n$, then $\bar{X}$ converges to $E(X_1)$ almost surely as $n \to \infty$.

- A version of weak law of large numbers (WLLN, 弱大數法則) is given in Theorem 5.1.1 in the text.

  Fact 1 (Theorem 5.1.1 in the text) Suppose that $X_1, \ldots, X_n$ are IID random variables and $E(X_1)$ and $Var(X_1)$ are finite. Let $\bar{X} = \sum_{i=1}^n X_i/n$, then $\bar{X} \stackrel{\mathcal{P}}{\to} E(X_1)$ as $n \to \infty$.

  Note.

  - The proof of Fact 1 can be based on Chebyshev's inequality or Markov's inequality.

- Since almost surely convergence implies convergence in probability, the assumption that $Var(X_1)$ is finite is not needed in Fact 1.

- Suppose that we use a statistic T_n to estimate some quantity $g(\theta)$, where n is the sample size and the distribution of the sample is determined by θ. If $T_n \xrightarrow{\mathcal{P}} g(\theta)$ as $n \to \infty$, then T_n is called a consistent estimator of $g(\theta)$ (T_n爲$g(\theta)$的一致估計量).

- Example 2. Suppose that $(X_1, \ldots, X_n)$ is a random sample and $\mu = E(X_1)$ is finite. Which of the following statements are true?

  (a) $\sum_{i=1}^{n} X_i/n$ is a consistent estimator of μ.

  (b) $1 + \sum_{i=1}^{n} X_i/n$ is a consistent estimator of $1 + \mu$.

  (c) $\mu + \sum_{i=1}^{n} X_i/n$ is a consistent estimator of 2μ.

  Ans. (a)(b)

- Example 3. Suppose that we have IID data $X_1, \ldots, X_n$, then for $t \in R$, a consistent estimator of $P(X_1 \leq t)$ is

$$\frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(X_i).$$

- In Example 3, for $t \in R$, let

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(X_i).$$

  Then $\hat{F}$ is call the empirical CDF based on $(X_1, \ldots, X_n)$.

- Suppose that $(X_1, \ldots, X_n)$ is a random sample from a distribution with CDF F. Then one can test whether

$$H_0 : F = F_0$$

  based on the sample using the Kolmogorov-Smirnov test, which rejects H_0 when

$$\sup_{x \in R} |\hat{F}(x) - F_0(x)|$$

  is large, where $\hat{F}$ is the empirical CDF based on $(X_1, \ldots, X_n)$.

  - To test whether the CDF F =`F0` based on the sample `x` $= (X_1, \ldots, X_n)$ using the Kolmogorov-Smirnov test, the R command is

```
ks.test(x, F0)
```

- The following result is used to establish convergence in probability under a continuous transformation, which is a version of continuous mapping theorem (for convergence in probability).

  Theorem. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of random vectors in R^k and $X_n \xrightarrow{\mathcal{P}} X$ as $n \to \infty$. Suppose that g is a continuous function on R^k. Then $g(X_n) \xrightarrow{\mathcal{P}} g(X)$ as $n \to \infty$.

- The following is another version of continuous mapping theorem.

  Theorem. Suppose that c is a vector of constants and $X_n \xrightarrow{\mathcal{P}} c$ as $n \to \infty$. Suppose that g is a function that takes values in R^k and is continuous at c. Then $g(X_n) \xrightarrow{\mathcal{P}} g(c)$ as $n \to \infty$. The proof of the above version of continuous mapping theorem is based on the definition of convergence in probability.

- Example 4. 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. 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

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

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

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

  Fact 2 Suppose that $X_n \xrightarrow{\mathcal{D}} c$ as $n \to \infty$ and c is a vector of constants, then $X_n \xrightarrow{\mathcal{P}} c$ as $n \to \infty$.

  Note. The special case where c is a constant is given in Theorem 5.2.2 in the text.

- 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 CLT (central limit theorem).

  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 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 5. 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

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

- Example 6. Suppose that X is a random variable with m possible values $a_1, \ldots, a_m$, and $(X_1, \ldots, X_n)$ is a random sample from the distribution of X. Let $p_j = P(X = j)$ and

  $$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} I_{\{a_j\}}(X_i)$$

  for $j = 1, \ldots, m$. Find the limiting distirbution of

  $$\sqrt{n}(\hat{p}_1 - p_1, \ldots, \hat{p}_m - p_m)^T.$$

  Ans. $N((0, \ldots, 0)^T, \text{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^T)$, where $\boldsymbol{p} = (p_1, \ldots, p_m)^T$, and diag($\boldsymbol{p}$) is the diagonal matrix with the vector $\boldsymbol{p}$ in the diagonal.

- Continuous mapping theorem for convergence in distribution.

  Theorem. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and g is a continuous function, then $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.

- Example 7. In Example 6, find

  (a) the limting distribution of

  $$\sqrt{n}\left(\frac{\hat{p}_1 - p_1}{\sqrt{p_1(1-p_1)}}\right),$$

  (b) the limiting distribution of

  $$n\left(\frac{\hat{p}_1 - p_1}{\sqrt{p_1(1-p_1)}}\right)^2,$$

  and

  (c) the limiting distribution of

  $$n\left(\frac{(\hat{p}_1 - p_1)^2}{p_1} + \cdots + \frac{(\hat{p}_m - p_m)^2}{p_m}\right).$$

  Ans. (a) $N(0,1)$ (b) $\chi^2(1)$ (c) $\chi^2(m-1)$. Note that to find the limiting distribution in (c), we apply the following result:

  Fact 4 Suppose that $\boldsymbol{U} \sim N(\mathbf{0}, \Sigma)$. If $\Sigma^2 = \Sigma$, then $\boldsymbol{U}^T\boldsymbol{U} \sim \chi^2(k)$, where $k = \text{trace}(\Sigma)$.

  Here trace(Σ) denotes the trace of the matrix Σ, which is the sum of the diagonal elements of Σ.

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

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

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

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

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

- A multivarite version of Slutsky's theorem. Suppose that X_n and Y_n are random vectors in R^k, $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{P}} c$ as $n \to \infty$, where c is a constant vector. Then

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

(ii) $X_n * Y_n \xrightarrow{\mathcal{D}} X * c$ as $n \to \infty$.

Here for $w = (w_1, \ldots, w_k)$ and $v = (v_1, \ldots, v_k)$ in R^k, $w * v$ is the vector $(w_1 v_1, \ldots, w_k v_k)$.

- Example 8. Suppose that $X_n \xrightarrow{\mathcal{D}} N(0, \sigma^2)$ 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 $X_n \xrightarrow{\mathcal{D}} Z$ as $n \to \infty$. Take $Y_n = 1/\sigma_n$ for all n, then $Y_n \xrightarrow{\mathcal{P}} 1/\sigma$ as $n \to \infty$. Apply Slutsky's theorem, we have

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

  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 $n \to \infty$.

- Example 9. Suppose that $(X_1, \ldots, X_n)$ is a random sample from a distribution with finite mean μ and finite variance σ^2. Suppose that $\sigma > 0$ and $\hat{\sigma}_n$ is a consistent estimator of σ based on $(X_1, \ldots, X_n)$. Find the limiting distirbution of

  $$\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}_n}.$$

  Sol. By assumption, $\hat{\sigma}_n \xrightarrow{\mathcal{P}} \sigma$. Since the function f defined by $f(x) = \sigma/x$ for $x \in (0, \infty)$ is continuous at σ, by continuous mapping theorem, we have

  $$\frac{\sigma}{\hat{\sigma}_n} = f(\hat{\sigma}_n) \xrightarrow{\mathcal{P}} f(\sigma) = \frac{\sigma}{\sigma} = 1.$$

  From central limit theorem,

  $$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{D}} Z,$$

  where $Z \sim N(0,1)$. Thus by Slutsky's theorem,

  $$\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}_n} = \underbrace{\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}}}_{\xrightarrow{\mathcal{D}} Z} \cdot \underbrace{\frac{\hat{\sigma}}{\hat{\sigma}_n}}_{\xrightarrow{\mathcal{P}} 1} \xrightarrow{\mathcal{D}} Z,$$

  where $Z \sim N(0,1)$. The limiting distribution of

  $$\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}_n}$$

  is $N(0,1)$.

- 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
  $$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} (D_g(\mu))^T Z \tag{3}$$
  as $n \to \infty$, where the column vector $D_g(\mu)$ is the gradient of g evaluated at μ.

  Note.

  - When $k = 1$, (3) 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 2 and the continuous mapping theorem (for convergence in probability).

- Example 10. 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 5 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 5 can be established if

$$\lim_{n\to\infty} P(X_n = c) = 0 = P(X = c) \tag{4}$$

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 (4), 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 11. In Example 10, 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 10, 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 12. Write down R scripts to find estimated coverage probability for the approximate $(1-\alpha)$ confidence interval of θ in Example 11 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.