Statistics II: Introduction to Inference

Week 9: Large Sample Theory

1 Sampling distributions

Recall the sampling distributions of some important statistics.

Example 1. *Let* $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. *Define* $T_{1n} = \bar{X}_n$ *(Sample mean) and* $T_{2n} = S_n^2$ *(Sample variance). Then,* $T_{1n} \sim \texttt{Normal}(\mu, n^{-1}\sigma^2)$ *and* $nT_{2n} \sim \sigma^2\chi^2_{(n-1)}$.

Example 2. *Let* $X_1, \ldots, X_n \overset{iid}{\sim} \texttt{Gamma}(\alpha, \beta)$, *then* $T_{3n} = \bar{X}_n \sim n^{-1}\texttt{Gamma}(n\alpha, \beta)$.

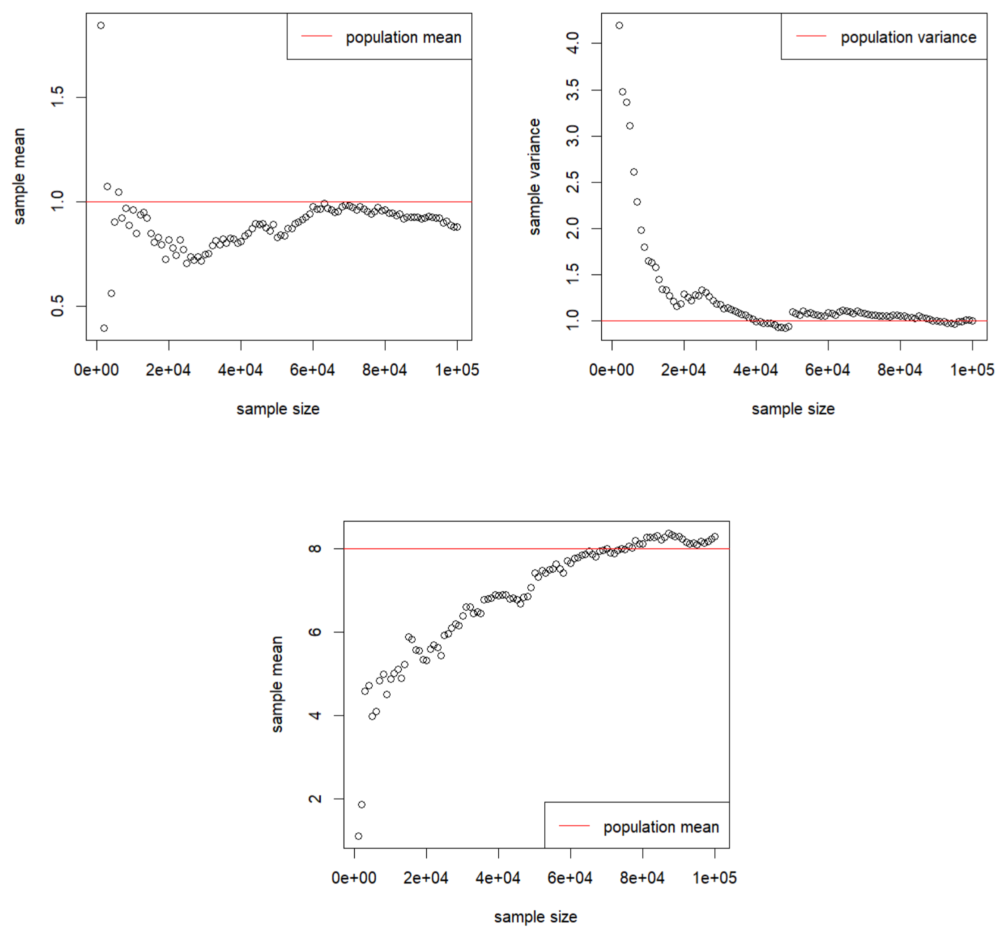

Fig. 1: Plots of three statistics T_{1n}, T_{2n} generated from $\texttt{Normal}(1, 1)$ and T_{3n} generated from $\texttt{Gamma}(2, 4)$ for varying sample sizes (n).

- As we see in the above examples, usually distribution of good statistics or estimators depend on n.
- Usually, a good statistic or estimator, say, T_n is a function of $X_1, \ldots, X_n$. For example,
  - $T_{11} = X_1$,
  - $T_{12} = (X_1 + X_2)/2$
  - $T_{13} = (X_1 + X_2 + X_3)/3$ and so on
  - $T_n = \bar{X}_n$.
- Therefore, it is natural to ask, **how the statistic behaves when we have a large number of representative samples**?
- In this module, we will learn the **large sample behaviour of statistic**.

2 Convergence of a sequence of random variables

Towards understanding the large sample behavior of a statistic, say T_n, it is important to understand what do we mean by convergence of a sequence of random variable.

Example 3. *Let $X_1, \cdots, X_n \overset{iid}{\sim} \texttt{Ber}(p)$ and consider the statistic $T_n = \bar{X}_n$. The distribution of T_n changes over n as follows:*

$$T_1 = \begin{cases} 0 & w.p.\ (1-p) \\ 1 & w.p.\ p \end{cases}, \ T_2 = \begin{cases} 0 & w.p.\ (1-p)^2 \\ \frac{1}{2} & w.p.\ 2p(1-p) \\ 1 & w.p.\ p^2 \end{cases}, \ T_3 = \begin{cases} 0 & w.p.\ (1-p)^3 \\ 1/3 & w.p.\ 3p(1-p)^2 \\ 2/3 & w.p.\ 3p^2(1-p)^2 \\ 1 & w.p.\ p^3 \end{cases} \ \textit{and so on.}$$

So, one can not directly generalize the concept of convergence of real sequence for convergence of random variables (RV). [Recall, a real sequence $\{a_n\}$ converges to a point a if for any $\epsilon > 0, \exists\ n_\epsilon$ s.t. $|a_n - a| < \epsilon$ for $n \geq n_\epsilon$.] Therefore, to define convergence of a sequence of RVs to a RV, i.e., to define $T_n \to T$, where $\{T_n\}$ and T are RVs, one first,

- reduces the difference of T_n and T to a real number,
- then apply the concept of real convergence.

Definition 1 (Convergence in probability)**.** *Let $\{T_n\}$ be a sequence of RVs and T be another RV such that all of them are defined on the same probability space (so that the RV $T_n - T$ is well-defined), then we say that T_n converges in probability to T, notationally, $T_n \xrightarrow{p} T$ if for any $\epsilon > 0$,*

$$p_{n,\epsilon} = P(|T_n - T| > \epsilon) \to 0 \qquad \textit{as } n \to \infty.$$

Remark 1. *Observe that, here $\{p_{n,\epsilon}\}$ is a real sequence measuring the difference between T_n and T (at ϵ level).*

Remark 2. *Sometimes the limiting RV, T, is of degenerate type, i.e., $P(T = c) = 1$. In such cases, $T_n \xrightarrow{p} T$ can equivalently be expressed as $T_n \xrightarrow{p} c$.*

Example 4. *Let $X_1, \cdots, X_n \overset{iid}{\sim} U(0, \theta)$, and $T_n = X_{(n)}$. Show that $T_n \xrightarrow{p} \theta$.*

Fix any $\epsilon > 0$,

$$\begin{aligned} P(|T_n - \theta| < \epsilon) &= P(\theta - \epsilon < T_n < \theta + \epsilon) \\ &= \begin{cases} P(\theta - \epsilon < T_n \le \theta) & \text{if } \epsilon \le \theta \\ P(0 \le T_n \le \theta) & \text{if } \epsilon > \theta \end{cases} \\ &= \begin{cases} 1 - (\epsilon/\theta)^n & \text{if } \epsilon \le \theta \\ 1 & \text{if } \epsilon > \theta \end{cases} \\ &\to 1 \text{ as } n \to \infty. \end{aligned}$$

Thus, $P(|T_n - \theta| > \epsilon) = p_{n,\epsilon} \to 0$ as $n \to \infty$.

Definition 2 (Consistency)**.** *Let $\{T_n\}$ be a sequence of estimators for the parameter $\psi(\theta)$. Then $\{T_n\}$ is called a* consistent *sequence of estimators for $\psi(\theta)$ if $T_n \xrightarrow{p} \psi(\theta)$ as $n \to \infty$.*

Example 5. *In the above **Example 4**, $X_{(n)}$ is consistent for θ.*

Checking Consistency: The following inequalities are quite helpful in checking consistency of estimators.

(I) Markov's inequality: Let X be a non-negative random variable. Then,

$$P(X > t) \le \frac{\mathbb{E}(X)}{t}, \quad \text{for any } t > 0.$$

Applications of Markov's inequality:

(1) **Chebyshev's inequality**: Let X be any random variable with expectation $\mathbb{E}(X) = \mu$ and finite variance $\mathrm{Var}(X) = \sigma^2$. Then,

$$\begin{aligned} P(|X - \mu| > \epsilon) &= P\left(|X - \mu|^2 > \epsilon^2\right) \\ &\le \frac{\mathbb{E}\left(|X - \mu|^2\right)}{\epsilon^2} \\ &= \frac{\sigma^2}{\epsilon^2} \end{aligned}$$

by Markov's inequality. The above inequality is called the Chebyshev's inequality.

(2) One can further sharpen the upper bound of $P(|X - \mu| > \epsilon)$ (the tail probability) when higher order moments are finite.

Let X be a RV with expectation μ, and finite r-th order moment $\mu_r = \mathbb{E}|X - \mu|^r$. Then,

$$P(|X - \mu| > \epsilon) \le \frac{\mu_r}{\epsilon^r}.$$

Note: Usually one obtains a sharper bound by taking a larger value of r. However, for a larger value of r, μ_r may not be finite.

(II) Application of the inequalities:

Sufficient conditions for consistency: A sequence of estimators $\{T_n\}$ is consistent for $\psi(\theta)$, i.e., $T_n \xrightarrow{p} \psi(\theta)$ if

$$\mathbb{E}(T_n) \to \psi(\theta) \text{ as } n \to \infty \quad \text{and} \tag{1}$$

$$\mathrm{Var}(T_n) \to 0 \text{ as } n \to \infty \tag{2}$$

Note: If (1) and (2) holds then $T_n \xrightarrow{p} \psi(\theta)$. But the converse is **NOT** true.

Proof. Observe that, for any $\epsilon > 0$,

$$\begin{aligned} P\left(|T_n - \psi(\theta)| > \epsilon\right) &= P\left(\{T_n - \psi(\theta)\}^2 > \epsilon^2\right) \\ &\le \frac{\mathbb{E}\left[\{T_n - \psi(\theta)\}^2\right]}{\epsilon^2} \end{aligned}$$

Now,

$$\mathbb{E}\left[\{T_n - \psi(\theta)\}^2\right] \le 3\,\mathrm{Var}\left(T_n\right) + 3\left[\mathbb{E}\left(T_n\right) - \psi(\theta)\right]^2 \to 0 \quad \text{by (1) and (2).}$$

$\square$

Example 6 (WLLN: IID case)**.** *Let $X_1, \cdots, X_n$ be IID samples from some distribution with expectation μ and finite variance σ^2. Then $\mathbb{E}\left(\bar{X}_n\right) = \mu$ and $\mathrm{Var}\left(\bar{X}_n\right) = \sigma^2/n \to 0$ as $n \to \infty$. So, $\bar{X}_n \xrightarrow{p} \mu$ as $n \to \infty$.*

2.1 Weak Law of Large Numbers (WLLN):

As shown in the last example, sample mean, $\bar{X}_n$, is a consistent estimator of population mean (expectation). This result is called law of large numbers (LLN). WLLN can be extended to the case with independent (not identically distributed) samples.

Result 1 (WLLN : Independent case)**.** *Let $X_1, \cdots, X_n$ be independent samples such that $\mathbb{E}\left(X_i\right) = \mu_i$ and $\mathrm{Var}\left(X_i\right) = \sigma_i^2; i = 1, \ldots, n$. Further, let $\sigma_i^2 \le M$ for some $M > 0$ for all $i = 1, \ldots, n$. Then*

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0 \text{ as } n \to \infty, \quad \text{where} \quad \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i.$$

Proof. Fix $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\bar{X}_n - \bar{\mu}_n\right| > \epsilon\right) &\le \frac{\mathbb{E}\left[\left(\bar{X}_n - \bar{\mu}_n\right)^2\right]}{\epsilon^2} \\ &= \frac{\mathbb{E}\left[\frac{1}{n^2}\left\{\sum_{i=1}^{n}\left(X_i - \mu_i\right)^2\right\}\right]}{\epsilon^2} \\ &= \frac{\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) + \sum_{i \neq j}\mathrm{Cov}\left(X_i, X_j\right)}{\epsilon^2 n^2} \\ &= \frac{\sum_{i=1}^{n}\sigma_i^2}{n^2\epsilon^2} \\ &\le \frac{nM}{n^2\epsilon^2} \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

$\square$

Remark 3. *Observe that, in the above proof it is enough to have*

$$\frac{\sum_{i=1}^{n}\sigma_i^2}{n^2} \to 0 \text{ as } n \to \infty.$$

Example 7. *Let $X_i \sim \texttt{Bern}\left(p_i\right)$ for $i = 1, \ldots, n$ and $X_1, \cdots, X_n$ are mutually independent. Observe that,*

$$\mathbb{E}\left(X_i\right) = p_i, \text{ and } \mathrm{Var}\left(X_i\right) = p_i\left(1 - p_i\right) \le \frac{1}{4} \quad \forall i.$$

Thus,

$$\left(\bar{X}_n - \bar{p}_n\right) \xrightarrow{p} 0 \text{ as } n \to \infty.$$

3 Convergence in distribution

Another concept of convergence of RVs is convergence in distribution.

Definition 3 (Convergence in distribution)**.** *Let $\{T_n\}$ be a sequence of RVs with CDFs $\{F_n\}$ and T be another random variable with CDF F. Then T_n is said to converge in distribution to T, notationally, $T_n \xrightarrow{d} T$ if*

$$F_n(x) \longrightarrow F(x) \text{ as } n \to \infty,$$

for any continuity point x of $F(\cdot)$.

Remark 4. *A point x is a continuity point of a CDF F if*

$$F\left(x^-\right) = F(x) = F\left(x^+\right) \tag{3}$$

We know that any CDF F is right continuous (i.e., $F\left(x^+\right) = F(x)$ for all $x \in \mathbb{R}$), and non-decreasing. However, for some points $x \in \mathbb{R}$, it might happen that

$$F\left(x^-\right) < F(x) = F\left(x^+\right). \tag{4}$$

Such points are called discontinuity points of F.

Note: There can be at most countable number of discontinuity points in any CDF F.

Remark 5. *If the limiting RV T is degenerate with $P(T = c) = 1$, then $T_n \xrightarrow{d} T$ can equivalently be written as $T_n \xrightarrow{d} c$.*

Example 8. *Let $X_1, \cdots, X_n \overset{iid}{\sim} \texttt{uniform}(0, \theta)$ and $T_n = X_{(n)}$. Show that $T_n \xrightarrow{d} \theta$.*

Proof. Let T be such that $P(T = \theta) = 1$. So, we'll show $T_n \xrightarrow{d} T$. The CDF of T, denoted by F_T, is as follows:

$$F_T(t) = \begin{cases} 0 & \text{if} \quad -\infty < t < \theta \\ 1 & \text{if} \quad t \geqslant \theta \end{cases}$$

Thus, F_T is discontinuous at the point θ only. So, the set of continuity points of F_T,

$$e = \{x : x \in \mathbb{R} \backslash \{\theta\}\}$$

Now, The CDF of T_n, F_n is given by

$$F_n(t) = \begin{cases} 0 & \text{if} \quad -\infty < t < 0 \\ (t/\theta)^n & \text{if} \quad 0 \leq t \leq \theta \\ 1 & \text{if} \quad t > \theta \end{cases}$$

Consider any point $x \in e$.

$$\begin{array}{lll} F_n(x) = 0 = F_T(x) & \text{if} & -\infty < x < 0, \\ F_n(x) = (x/\theta)^n \to 0 = F_T(x) & \text{if} & 0 \leq x < \theta, \\ F_n(x) = 1 = F(x) & \text{if} & x > \theta \end{array}$$

So, $T_n \xrightarrow{d} T$. $\square$

3.1 Some important remarks

Remark 6. *If* $T_n \xrightarrow{p} T$ *then* $T_n \xrightarrow{d} T$*, but the converse is not true in general. However, if* $T_n \xrightarrow{d} c$ *then* $T_n \xrightarrow{P} c$.

Remark 7 (Continuous mapping theorem)**.** *If* $T_n \xrightarrow{d/p} T$ *and* $g : \mathbb{R} \to \mathbb{R}$ *is a continuous function then* $g(T_n) \xrightarrow{d/p} g(T)$.

Remark 8 (Slutsky's Lemma)**.** *Let* $T_n \xrightarrow{d} T$ *and* $W_n \xrightarrow{p} c$ *then*

(i) $T_n \pm W_n \xrightarrow{d} T \pm c$,

(ii) $T_n W_n \xrightarrow{d} cT$, *and*

(iii) if $c \neq 0$ *then* $T_n/W_n \xrightarrow{d} T/c$.

4 Central Limit Theorem (CLT)

The CLT concerns the large sample behavior of sample mean, after proper scaling (magnification). Like WLLN, there are various versions of CLT. Here we provide only the result for **IID** samples.

Theorem 1 (CLT, IID case)**.** *Let* $X_1, \ldots, X_n$ *be IID samples from some distribution with expectation* μ*, and finite variance* σ^2*. Then,*

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z,$$

where $Z \sim N(0,1)$ *and* $S_n = (X_1 + \cdots + X_n)$.

Example 9. *Let* $X_1, \ldots, X_n \overset{iid}{\sim} N\left(0, \sigma^2\right)$ *distribution. Then,* $n^{-1}\sum_{i=1}^{n} X_i^2$ *is an estimator of* σ^2*. By CLT,*

$$\frac{\sqrt{n}\left[\sum_{i=1}^{n} X_i^2 - \sigma^2\right]}{\sqrt{3\sigma^4}} \xrightarrow{d} Z \sim N(0,1)$$

We say that $n^{-1}\sum_{i=1}^{n} X_i^2$ *is asymptotically normally distributed with asymptotic mean* σ^2 *and asymptotic variance* $3\sigma^4/n$.

5 Reference

Chapter 6 of *Probability and Statistics* by M. H. DeGroot and M. J. Schervish.