STA 522, Spring 2022
Introduction to Theoretical Statistics II

Lecture 1

Department of Biostatistics
University at Buffalo

Agenda

- Review random samples
- Order Statistics
- Convergence Concepts

Review: Random Samples

Definition: The random variables $X_1, X_2, \ldots, X_n$ are called a **random sample** of size n from the population $f(x)$ if $X_1, X_2, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$.

Notation: $X_1, X_2, \ldots, X_n \sim \text{iid}\, f$. Joint pdf/pmf:
$f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) := \prod_{i=1}^n f(x_i)$

If f is a member of a parametric family with parameter(s) θ, then we may write $f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta)$

Example: $X_1, X_2, \ldots, X_n \sim \text{iid}\, \mathrm{N}(\mu, \sigma^2)$ with
$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$

Review: Statistics and Sampling Distributions

Definition: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, x_2, \ldots, x_n)$ be a function (real-valued or vector-valued) whose domain includes the sample space of $(X_1, X_2, \ldots, X_n)$. The random variable (or vector) $Y = T(X_1, X_2, \ldots, X_n)$ is called a **statistic**. The probability distribution of a statistic is called its **sampling distribution**.

Note: A statistic cannot contain a parameter.

Examples:

(i) sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$,

(ii) sample variance
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2$

(iii) sample standard deviation $S = \sqrt{S^2}$.

(iv) sample minimum, sample maximum, sample quantiles.

Result (Lemma 5.2.5): Let $X_1, X_2, \ldots, X_n$ be a random sample from a population, and let $g(x)$ be a function such that $\mathrm{E}\left(g(X_1)\right)$ and $\mathrm{Var}\left(g(X_1)\right)$ exist. Then

$$\mathrm{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n\,\mathrm{E}\left(g(X_1)\right)$$

and

$$\mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\,\mathrm{Var}\left(g(X_1)\right).$$

Result (Theorem 5.2.6): Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

(a) $\mathrm{E}\left(\overline{X}\right) = \mu$

(b) $\mathrm{Var}\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$; and

(c) $\mathrm{E}\left(S^2\right) = \sigma^2$.

How to determine the sampling distribution of $\overline{X}$?

(i) **Using transformations.** Let $Y = \sum_{i=1}^{n} X_i$, so that $\overline{X} = \frac{1}{n} Y$. If $f(x)$ is the pdf of Y, then the pdf of $\overline{X}$ is $f_{\overline{X}}(x) = nf(nx)$.

(ii) **Using mgf (Theorem** 5.2.7**).** $M_{\overline{X}}(t) = M_Y\left(\frac{t}{n}\right) = \left[M_X\left(\frac{t}{n}\right)\right]^n$ where $M_X(t)$ is the mgf of the underlying population. Then identify the distribution of $\overline{X}$.

Theorem 5.3.1. Let $X_1, X_2, \ldots, X_n \sim \text{iid}\, \mathrm{N}(\mu, \sigma^2)$. Then

a. $\overline{X}$ and S^2 are independent random variables.

b. $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$.

c. $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

Order Statistics

Definition: The order statistics of a random sample $X_1, X_2, \dots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ and satisfy $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$.

Examples:

(a) **sample minimum:** $X_{(1)}$ and **sample maximum:** $X_{(n)}$ are called the extreme order statistics.

(b) **sample range:** $R = X_{(n)} - X_{(1)}$.

(c) **sample median:** M where

$$M = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd;} \\ \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2} & \text{if } n \text{ is even.} \end{cases}$$

Sampling Distributions of Extreme Order Statistics from a Continuous Population

Suppose $X_1, X_2, \dots, X_n$ is a random sample from a population with continuous cdf F and pdf f. Then

1. $\{X_{(n)} \leq x\} = \{\text{all } X_i \leq x\} = \{X_1 \leq x, \dots, X_n \leq x\}$. So

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) \\ &= P(X_1 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \dots P(X_n \leq x) \\ &= F(x) \dots F(x) = [F(x)]^n \end{aligned}$$

   Differentiating, $f_{X_{(n)}}(x) = n\, f(x)\, [F(x)]^{n-1}$.

2. $\{X_{(1)} \geq x\} = \{\text{all } X_i \geq x\} = \{X_1 \geq x, \dots, X_n \geq x\}$. Implies $F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$ & $f_{X_{(1)}}(x) = n\, f(x)\, [1 - F(x)]^{n-1}$.

Example: $X_1, X_2, \ldots, X_n \sim$ iid Uniform$(0, \theta)$. Find the pdf and the expected value of $X_{(n)}$.

$$\text{Here } f(x \mid \theta) = \frac{1}{\theta}\; I(0 \le x \le \theta) \text{ and } F(x \mid \theta) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \le x \le \theta \\ 1, & x > 1 \end{cases}$$

so that

$$\begin{aligned} f_{X_{(n)}}(x \mid \theta) &= n\ f(x \mid \theta)\,[F(x \mid \theta)]^{n-1} \\ &= n\left(\frac{1}{\theta}\right)\left(\frac{x}{\theta}\right)^{n-1} I(0 \le x \le \theta) \\ &= \frac{n\ x^{n-1}}{\theta^n}\; I(0 \le x \le \theta) \end{aligned}$$

Find expected value $\mathrm{E}\left[X_{(n)}\right] = \mathrm{E}\left[X_{(n)} \mid \theta\right]$ using integration:

$$\mathrm{E}\left[X_{(n)}\right] = \int_{-\infty}^{\infty} x\ f_{X_{(n)}}(x \mid \theta)\ dx = \frac{n}{\theta^n}\int_0^{\theta} x^n\ dx = \frac{n}{n+1}\ \theta$$

Distribution of a general order statistic from a continuous population

Theorem 5.4.4. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics of a random sample $X_1, X_2, \ldots, X_n$ from a continuous population with cdf $F(x)$ and pdf $f(x)$. The pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \, f(x) \, \left[F(x)\right]^{j-1} \left[1 - F(x)\right]^{n-j}.$$

Partial Proof. Call $\{X_i \le x\}$ a "success", $\{X_i > x\}$ a "failure". Define $Z_i = I(X_i \le x)$ and $Y = \sum_{i=1}^{n} Z_i$. Note that $Z_i \sim$ iid Bernoulli$(F(x)) \implies Y \sim$ Binomial$(n, F(x))$. Note that,

$$F_{X_{(j)}}(x) = P(X_{(j)} \le x) = P(Y \ge j) = \sum_{k=j}^{n} \binom{n}{k} \left[F(x)\right]^{k} \left[1 - F(x)\right]^{n-k}$$

The pdf is obtained using differentiation.

Distribution of a general order statistic from a discrete population

Theorem 5.4.3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a discrete distribution with pmf $f(x_i) = p_i$, where $x_1 < x_2 < \ldots$ are the possible values of X in ascending order. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics from the sample. Define

$$\begin{aligned} P_0 &= 0 \\ P_i &= p_1 + p_2 + \cdots + p_i \qquad \text{for } i \geq 1 \end{aligned}$$

Then

$$\begin{aligned} P(X_{(j)} \leq x_i) &= \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \\ P(X_{(j)} = x_i) &= \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right]. \end{aligned}$$

Proof: cdf is similar to the continuous case. The pmf is obtained from the cdf through $P(X_{(j)} = x_i) = P(X_{(j)} \leq x_i) - P(X_{(j)} \leq x_{i-1})$.

Example: $X_1, X_2, \ldots, X_n \sim$ iid Uniform$(0,1)$. Find the distribution of the j^{th} order statistic, along with its mean and variance.

Here $f(x) = I(0 < x < 1)$ and $F(x) = x$ for $0 < x < 1$. Therefore

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(j-1)!(n-j)!}\ f(x)\ \left[F(x)\right]^{j-1}\left[1 - F(x)\right]^{n-j} \\
&= \frac{n!}{(j-1)!(n-j)!}\ x^{j-1}\ (1-x)^{n-j}\ I(0 < x < 1) \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)}\ x^{j-1}\ (1-x)^{n-j} I(0 < x < 1)
\end{aligned}$$

This shows that $X_{(j)} \sim \text{Beta}(j, n-j+1)$. From this, we can deduce that

$$\mathrm{E}\left[X_{(j)}\right] = \frac{j}{n+1}$$

and

$$\mathrm{Var}\left[X_{(j)}\right] = \frac{j(n-j+1)}{(n+1)^2(n+2)}.$$

Joint Distribution of Order Statistics

Theorem 5.4.6. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics of a random sample $X_1, X_2, \ldots, X_n$ from a continuous population with cdf $F(x)$ and pdf $f(x)$. The joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is

$$f_{X_{(i)},X_{(j)}}(u,v) = c\, f(u)\, f(v)\, [F(u)]^{i-1}\, [F(v) - F(u)]^{j-1-i}\, [1 - F(v)]^{n-j}$$

for $-\infty < u < v < \infty$, where $c = \frac{n!}{(i-1)!(j-1-i)!(n-j)!}$.

Joint distribution pdf of all the order statistics from a continuous population:

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n) = \begin{cases} n! f(x_1)\ldots f(x_n), & -\infty < x_1 < \cdots < x_n < \infty \\ 0, & \text{otherwise} \end{cases}$$

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid uniform$(0, a)$, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics. Find the joint pdf of the sample range $R = X_{(n)} - X_{(1)}$ and the mid-range $V = \frac{X_{(1)}+X_{(n)}}{2}$. Hence find the marginal pdf of R.

First obtain the joint pdf of $X_{(1)}$ and $X_{(n)}$:

$$
\begin{aligned}
f_{X_{(1)},X_{(n)}}(x_1, x_n) &= \frac{n(n-1)}{a^2}\left(\frac{x_n}{a} - \frac{x_1}{a}\right)^{n-2} \; I(0 < x_1 < x_n < a) \\
&= \frac{n(n-1)(x_n - x_1)^{n-2}}{a^n} \; I(0 < x_1 < x_n < a)
\end{aligned}
$$

Solve for $X_{(1)}, X_{(n)}$ to obtain $X_{(1)} = V - R/2$ and $X_{(n)} = V + R/2$. Jacobian of this transformation is -1.

Support of (R, V):

$$\begin{aligned} & 0 < x_1 < x_n < a \\ \Longrightarrow\ & 0 < v - r/2 < v + r/2 < a \\ \Longrightarrow\ & 0 < r < a,\ r/2 < v < a - r/2 \end{aligned}$$

The joint pdf of (R, V) is

$$f_{R,V}(r, v) = \frac{n(n-1)\ r^{n-2}}{a^n};\quad 0 < r < a,\ r/2 < v < a - r/2$$

The marginal pdf of R is

$$f_R(r) = \int_{r/2}^{a-r/2} \frac{n(n-1)\ r^{n-2}}{a^n}\ dv = \frac{n(n-1)\ r^{n-2}\ (a-r)}{a^n};\ 0 < r < a$$

It is easy to see that $\frac{R}{a} \sim \text{Beta}(n-1, 2)$ distribution.

HW: find the marginal pdf of V.

Convergence Concepts

What happens to sample statistics, particularly $\overline{X} = \overline{X}_n$, when the sample size $n \to \infty$?

For a real sequence $(a_n)_{n=1}^{\infty}$ defining convergence is straightforward: $(a_n)_{n=1}^{\infty}$ is said to converge to a point a if $\lim_{n\to\infty} |a_n - a| = 0$.

How to define convergence of random variables?

- convergence in probability
- convergence in almost sure sense
- convergence in distribution (or law)
- convergence in mean (or norm) [may be later]

Convergence in Probability (or Weak Convergence)

Definition (5.5.1): A sequence of random variables $X_1, X_2, \ldots$ **converges in probability** to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0,$$

or, equivalently,

$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1.$$

To indicate this, we write $X_n \xrightarrow{P} X$.

Notes:

1. The random variables $X_1, X_2, \ldots$ are NOT assumed to be iid, i.e., we are in a more general setting than what we have discussed so far.
2. We are often interested in the case where the "limiting" random variable X is a constant (degenerate)

Weak Law of Large Numbers (WLLN)

Theorem 5.5.1. Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|\overline{X}_n - \mu| < \varepsilon\right) = 1,$$

so that

$$\overline{X}_n \xrightarrow{P} \mu.$$

Proof: Using Chebyshev. For any $\varepsilon > 0$

$$P\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\mathrm{E}\left(\overline{X}_n - \mu\right)^2}{\varepsilon^2} = \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Hence $0 \leq P\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$.

Remarks:

1. In other words, there is a very high probability that the sample mean can be made as close as we'd like to the population mean by taking n sufficiently large.
2. **Consistency:** when a sample quantity (statistic) converges in probability to a constant (more later).

Example: Suppose we have a sequence $X_1, X_2, \ldots, X_n$ are iid random variables with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Define $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Can we prove a WLLN for S_n^2?

Using Chebyshev

$$P\left(|S_n^2 - \sigma^2| \geq \varepsilon\right) \leq \frac{\mathrm{E}\left(S_n^2 - \sigma^2\right)^2}{\varepsilon^2} = \frac{\mathrm{Var}(S_n^2)}{\varepsilon^2}$$

So a sufficient condition that $S_n^2 \xrightarrow{P} \sigma^2$ is that $\mathrm{Var}(S_n^2) \to 0$ as $n \to \infty$.

This sufficient condition holds if $X_1, X_2, \ldots, X_n \sim$ iid $\mathrm{N}(\mu, \sigma^2)$.
Then $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1} \implies$

$$\mathrm{Var}\left(\frac{(n-1)S_n^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4}\,\mathrm{Var}(S_n^2) = \mathrm{Var}(\chi^2_{n-1}) = 2(n-1)$$

This implies $\mathrm{Var}(S_n^2) = \frac{2\sigma^4}{(n-1)} \to 0$ as $n \to \infty$.

Theorem 5.5.4. Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X, and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

Proof Homework (see exercise 5.39).

Example (contd.): If $S_n^2 \to \sigma^2$ then $S_n = \sqrt{S_n^2} \to \sigma$.

Almost Sure Convergence (or Strong Convergence)

Definition (5.5.6): A sequence of random variables $X_1, X_2, \ldots$ **converges almost surely** to a random variable X if, for every $\varepsilon > 0$,

$$P\left(\lim_{n\to\infty} |X_n - X| < \varepsilon\right) = 1.$$

To indicate this, we write $X_n \xrightarrow{a.s.} X$.

Notes:

- Contrast this with the definition of convergence in probability.
- Recall that a random variable is a function from a sample space $\mathcal{S}$ into the real numbers: $X_n : \mathcal{S} \longrightarrow \mathbb{R}$. For each $s \in S$, $X_n(s) = r \in \mathbb{R}$.
- Definition 5.5.6 states that $X_n \xrightarrow{a.s.} X$ if the functions $X_n(s) \longrightarrow X(s)$ for all $s \in S$, except perhaps for $s \in N$, where $N \subseteq \mathcal{S}$ and $P(N) = 0$ (point-wise convergence on all but a few "null" points).

Homework

- Order Statistics: Read p. $226 - 232$. Exercises $5.22, 5.26$.
- Convergence: Read p. $232 - 236$. Exercises $5.32, 5.38, 5.39a$.