STAT 330: Lectures 14&15

2024 Spring

July 2, 2024

Last Lecture

- Continue on one-to-one bivariate transformation method
  1. Find the inverse transformation
  2. Find the support
  3. Using the formula
- MGF method
- Some important distributions. $\chi^2_n \stackrel{d}{=} \sum_{i=1}^{n} Z_i^2$ $\quad F_{n,m} \stackrel{d}{=} \frac{\chi^2_n / n}{\chi^2_m / m}$

$t_n \stackrel{d}{=} \frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}$

This Lecture

- Some results of a normal sample (end of Chapter 3).
- Start the last chapter on asymptotic statistics (Chapter 4).

Some Distributional Results of Normal Samples $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Norm}(\mu, \sigma^2)$

Sample Mean and Sample Variance:

1. Definitions

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$S_n^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

random variables.

Mean: $E(X) = \mu$
Variance: $Var(X) = \sigma^2$
non-random.

2. Random or Non-random

Mean, Variance: non-random

Sample mean, Sample variance: random

3. Estimator/Estimate/Parameter

$\bar{X}_n$: estimator for parameter μ

S_n^2 : estimator for parameter σ^2

suppose. $x_1 = 1$. $x_2 = 2$. $x_3 = 3$.

$\bar{x}_n = \frac{1+2+3}{3} = 2$. estimate. num.

estimate is the realized version of the estimator. and it is non-random.

4. Distributional Results (proofs skipped)

① $\bar{X}_n \perp S_n^2$

② $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \text{Norm}(0,1)$

③ $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$

NOT a proof!

$$\frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2}$$

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} + 2\cdot\frac{\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \mu)}{\sigma^2}$$

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu)}{\sigma^2} = \frac{(\bar{X}-\mu)}{\sigma^2}\cdot\sum_{i=1}^{n}(X_i-\bar{X}) = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

$$\hookrightarrow \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2_1$$

If $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$, the equation holds.

<u>*More Distributional Results*</u>:

① $\frac{\bar{X}_n-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$\frac{\bar{X}_n-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}}{S/\sigma}$$

1). $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \sim Norm(0,1)$

2) $\frac{S}{\sigma} = \sqrt{\frac{S^2}{\sigma^2}} = \sqrt{\frac{1}{n-1}\cdot\frac{(n-1)S^2}{\sigma^2}} \sim \sqrt{\frac{\chi^2_{n-1}}{n-1}}$

3) $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \perp \frac{S}{\sigma}$

$\hookrightarrow t_{n-1}$

②. $X_1 \cdots X_n \overset{iid}{\sim} Norm(\mu_1, \sigma_1^2)$
$Y_1 \cdots Y_m \overset{iid}{\sim} Norm(\mu_2, \sigma_2^2)$ independent.

Let $S_1^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{n-1}$, $S_2^2 = \frac{\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2}{m-1}$.

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim ?$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{\sigma_1^2}\Big/(n-1)}{\frac{\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2}{\sigma_2^2}\Big/(m-1)} \sim \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)} \overset{d}{=} F_{n-1,m-1}$$

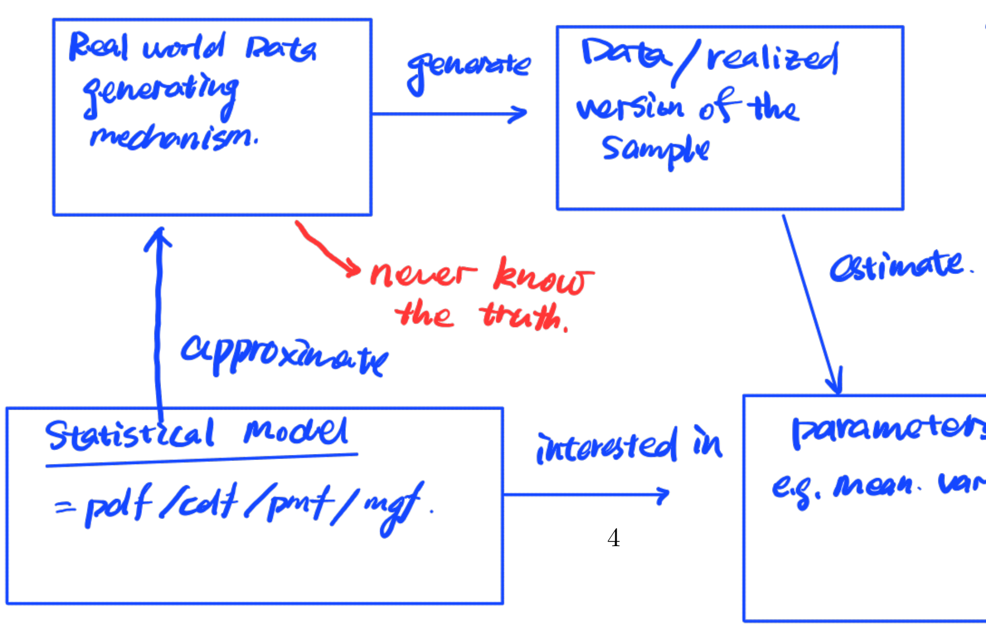
Introduction to Asymptotic Statistics

Where are we?

1. So far, our focus is on how to describe the behaviors of random variables.
   (a) Pdf, pmf, and cdf provide a whole picture (i.e., distribution) of the random variables behaviors.
   (b) Parameters, including mean, variance, and correlation coefficient, summarize the distribution.

What is next? We almost always assume that we have access to true distributions of the random variables. But this is not the case in applications.

1. In real-world applications, we are given a dataset. But we do not know the true distribution.
2. We treat the dataset as realizations $\boldsymbol{X} = \mathbf{x}$ of some random variables $\boldsymbol{X}$.
3. We use some functions $h(\mathbf{x})$ of the data to estimate parameters.
4. If we replace the $\mathbf{x}$ in estimate $h(\mathbf{x})$ with random variable $\boldsymbol{X}$, we call $h(\boldsymbol{X})$ an estimator.
5. $h(\boldsymbol{X})$ is a random variable. So, $h(\boldsymbol{X})$ itself has a distribution, and we hope to describe it.
6. In some cases, the behavior of $h(\boldsymbol{X})$ is easy to describe.
7. In some cases, the finite sample behavior of the estimator is unclear. But it becomes clear when the sample size $n \to \infty$.
8. Asymptotic statistics aims to answer the question of what is the behavior of estimators when $n \to \infty$.



E.g. $x_1 = 1.\ x_2 = 2.\ x_3 = 3.$
Estimate $E(X) = \mu$.
$\hat{\mu} = \frac{x_1 + x_2 + x_3}{3} = 2$
estimate.
replace $x_1 . x_2 . x_3$ with rvs. $X_1 . X_2 . X_3$.
$\hat{\mu} = \frac{X_1 + X_2 + X_3}{3}$
estimator
need to find the dist'n of estimator to do inference.

①. $X_1 \cdots X_n \sim Norm(\mu, 1)$

$\hat{\mu} = \frac{X_1 + \cdots + X_n}{n} \sim Norm(\mu, \frac{1}{n})$

Easy case.

②. $X_1 \cdots X_n \sim Unif(0, \theta)$

$\hat{\mu} = \frac{X_1 + \cdots + X_n}{n} \sim$? (Not easy to find)

If $n \to \infty$, what will happen to the dist'n of $\hat{\mu}$

Convergence in Distribution

Definition: Suppose $X_1, \ldots, X_n$ is a sequence of random variables and the cdf of X_i is denoted as $F_i(x)$. Let X be a random variable with cdf $F(x)$, then if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all x at which $F(x)$ is continuous, we say the sequence $\{X_i\}$ converge to X in distribution.

Notation: $X_n \xrightarrow{d} X, n \to \infty$.

Comments: $Y_1 \cdots Y_n \sim Norm(\mu, \sigma^2)$

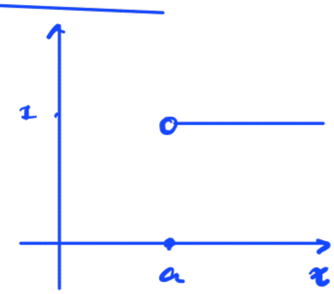
1. E.g. $X_1 = Y_1$, $X_2 = \frac{Y_1+Y_2}{2}$, $X_3 = \frac{Y_1+Y_2+Y_3}{3}$, ...

2. $F(\cdot)$ is called the limiting distribution/asymptotic distribution of $\{X_i\}$.

3. E.g. $F(x) = \begin{cases} 1, & x \geq a \\ 0, & x < a. \end{cases}$ $\quad \lim_{n\to\infty} F_n(x) = \begin{cases} 1 & x > a \\ 0 & x \leq a. \end{cases}$

   $\{X_i\}$ has a limit distribution $F(x)$ even if $\lim_{n\to\infty} F_n(x)$ is not a cdf.

4. It is the cdf that converges, not random variable.



5. The definition holds for both discrete/continuous r.v.s.

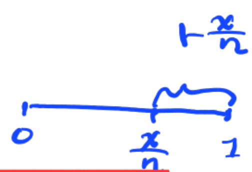
Example: Suppose $X_1, \ldots, X_n$ are i.i.d. Uniform$(0,1)$. Let $X_{(1)}$ and $X_{(n)}$ be the minimal and maximum of $X_1, \ldots, X_n$. Find the limiting distributions of

1. $nX_{(1)}$ and $n(1-X_{(n)})$
2. $X_{(1)}$ and $X_{(n)}$

①. $nX_{(1)} \in (0, n)$

$$\Pr(nX_{(1)} \le x) = \begin{cases} 0 & x \le 0 \\ \Pr(X_{(1)} \le \frac{x}{n}), & x \in (0,n) \\ 1 & x \ge n \end{cases}$$

$$\Pr(X_{(1)} \le \tfrac{x}{n}) = 1 - \Pr(X_{(1)} > \tfrac{x}{n}) = 1 - \{\Pr(X_1 > \tfrac{x}{n})\}^n$$
$$= 1 - (1 - \tfrac{x}{n})^n$$



$$\lim_{n\to\infty} \Pr(X_{(1)} \le \tfrac{x}{n}) = \begin{cases} 0 & , x \le 0 \\ 1 - e^{-x} & , x \in (0, +\infty) \end{cases}$$

$$\boxed{\lim_{n\to\infty} (1+\tfrac{x}{n})^n = e^x}$$

$\Rightarrow$ $nX_{(1)} \xrightarrow{d} \text{Exp}(1)$

$n(1-X_{(n)}) \in (0, n)$

$$\Pr\{n(1-X_{(n)}) \le x\} = \begin{cases} 0, & x \le 0 \\ \Pr(X_{(n)} \ge 1-\frac{x}{n}) & x \in (0,n) \\ 1, & x \ge n \end{cases}$$

$$\Pr(X_{(n)} \ge 1-\frac{x}{n}) = 1-\Pr(X_{(n)} < 1-\frac{x}{n})$$
$$= 1-\{\Pr(X_1 < 1-\frac{x}{n})\}^n$$
$$= 1-(1-\frac{x}{n})^n$$

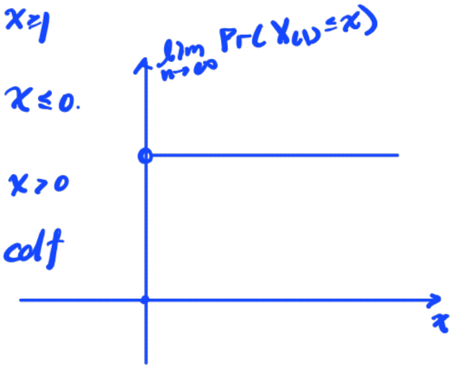
$$\lim_{n\to\infty} \Pr(n(1-X_{(n)}) \le x) = \begin{cases} 0, & x \le 0 \\ 1-e^{-x}, & x \in (0, \infty) \end{cases}$$

$\Rightarrow \quad n(1-X_{(n)}) \xrightarrow{d} Exp(1)$

(2) $X_{(1)} \in (0,1)$

$$\Pr(X_{(1)} \le x) = \begin{cases} 0 & , x \le 0 \\ 1-(1-x)^n & , x \in (0,1) \\ 1 & , x \ge 1 \end{cases}$$

$$\lim_{n\to\infty} \Pr(X_{(1)} \le x) = \begin{cases} 0 & , x \le 0. \\ 1 & , x > 0 \end{cases}$$



not right-continuous. / not a cdf

$$F(x) = \begin{cases} 0. & x < 0 \\ 1. & x \ge 0 \end{cases}$$

$F(x) \neq \lim_{n\to\infty} F_n(x)$. at. $x=0$

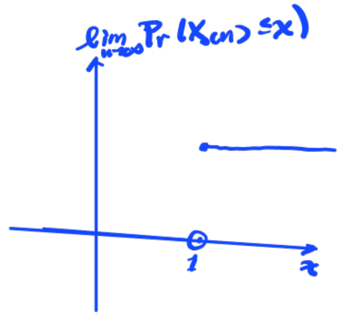
$F(x)$ is not continuous at 0.

$F(x)$ is the limiting distribution of $X_{(1)}$

$X_{(n)} \in (0,1)$

$$\Pr(X_{(n)} \le x) = \begin{cases} 0 & , x \le 0 \\ x^n & , x \in (0,1) \\ 1 & , x \ge 1 \end{cases}$$

$$\lim_{n\to\infty} \Pr(X_{(n)} \le x) = \begin{cases} 0 & , x < 1 \\ 1 & , x \ge 1 \end{cases}$$



its limiting distribution is

$$F(x) = \begin{cases} 0 & , x < 1 \\ 1 & , x \ge 1. \end{cases}$$