week 15

5.1 . iid random samples

Def: Suppose $X_1, \dots, X_n \overset{iid}{\sim} f$ (pmf or pdf).
Then we say $(X_1, \dots, X_n)$ is a random sample from the population (with pmf/pdf) $f(x)$.

Remarks: (other variations):
- n-sample of f
- n-iid sample of f
- n-sample of size n

Ex: $X_1, \dots, X_n \overset{iid}{\sim} EXP(\beta)$, $\beta > 0$

Then, we may compute, say
$P(X_1 > a, X_2 > a_2, \dots, X_n > a_n)$, etc.

Def: Let $(X_1, \dots, X_n)$ be a n-sample from a population.
Let $\tau(x_1, \dots, x_n)$ be a real-valued function.
Then $Y = \tau(X_1, \dots, X_n)$ is called a statistic
i.e. a function of a random sample

Remarks: · Statistic is a RV (too)
- telling us something about an (underlying) population w/ pmf/pdf f.
- does so only on random samples (data)

Modern viewpoint: $X_1, \dots, X_n \to \boxed{T} \to Y = T(X_1, \dots, X_n)$

algorithm/application

Examples: 1) $\bar{X} := \frac{1}{n}(X_1 + \cdots + X_n)$; sample mean

2) $S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

3) $X_{(1)} := \min\{X_1, \dots, X_n\}$

$X_{(2)} := \min\{X_1, \dots, X_n\} \setminus \{X_{(1)}\}$

$\vdots$

$X_{(k)} := \min\{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(k-1)}\}$

i.e. the k^{th} smallest number of the sample

order statistics of the n-sample $:= (X_{(1)}, \dots, X_{(n)})$

Thm. Let $(X_1, \dots, X_n)$ be an n-iid sample from a population with mean μ, variance $\sigma^2 < \infty$

Then $\begin{cases} E\bar{X} = \mu \\ \text{var}\,\bar{X} = \sigma^2/n \\ ES^2 = \sigma^2 \end{cases}$

Remark: provides statistical justification for using $\bar{X}$ and s^2 as estimates of μ and σ^2 respectively; (*) and (**) are unbiased estimates

Proof. $E\bar{X} = E\frac{1}{n}(X_1 + \cdots + X_n)$

$= \frac{1}{n}(EX_1 + \cdots + EX_n)$ by LOE

$= \frac{1}{n}(n\mu) = \mu$

$\text{var}\,\bar{X} = \text{var}\frac{1}{n}(X_1 + \cdots + X_n)$

$= \frac{1}{n^2}\text{var}(X_1 + \cdots + X_n)$

$= \frac{1}{n^2}(\text{v}X_1 + \cdots + \text{v}X_n)$, since X_i - independent

$= \frac{1}{n^2} n \cdot \text{var}X_1$

$= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$. ∎

Distribution of $\bar{X}$

two main methods: $\begin{cases} \text{1) method of moments} \\ \text{2) change-of-var formula} \\ \quad \text{(convolution formula)} \end{cases}$

1) $M_{\bar{X}}(t) = Ee^{t\bar{X}}$

$= Ee^{\frac{t}{n}(X_1 + \cdots + X_n)}$

$= Ee^{\frac{t}{n}X_1} \cdots e^{\frac{t}{n}X_n}$

$= (M_X(t/n))^n$

Remark: If $M_{\bar{X}}$ can be recognized as MGF of a known family, then we can find the dist. of $\bar{X}$ easily, e.g.

If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$,
then $\bar{X} \sim N(\mu, \sigma^2/n)$

2) Convolution Formula

Thm. If $X \sim f_X$, $Y \sim f_Y$,
then $Z := X+Y$ has the pdf which is the convolution of f_X and f_Y:

$$\boxed{f_Z(z) = f_X(*) f_Y(z) := \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx}$$

Proof. use the change-of-var formula for the mapping $(X,Y) \to (X+Y, X)$

Remark: useful for deriving pdf/pmf when the transformation $T(X_1, \dots, X_n)$ does not belong to a known (well-recognized) family (e.g. exponential, location-scale, etc.)

5.2: useful classical facts

Thm. If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$,

let $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$

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

then

a) $\bar{X}$ and S^2 are independent RV's

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

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

Recall: χ^2_p w/ p df:

$$f(x) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{p/2-1} e^{-x/2}, \quad x > 0$$

and $\chi^2_p \equiv \text{Gamma}(p/2, 2)$

Other statistics and their distributions

If $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ - unknown

then $\bar{x}$ is a statistic that tells us abt μ

Q: how so? we know $\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

this gives us a way to quantify the uncertainty abt μ, using statistic $\bar{X}$ as an estimate

Q: what if σ is unknown too?

we may want to consider $\frac{\bar{X} - \mu}{S/\sqrt{n}}$

since S is a statistic that can be obtained from the sample

To be useful, we need to know the distribution of

$$T := \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Thm. $T \sim t_{n-1}$: student's t dist. w/ n-1 df

t_p has the pdf

$$\boxed{f_T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+\frac{t^2}{p})^{(p+1)/2}} \Bigg|_{-\infty < t < \infty}}$$

Remarks: • T is the ratio of two independent Random variables as

$$\bar{X} - \mu \sim \text{Normal}, \quad \overbrace{\sqrt{(n-1)S^2} \sim \sqrt{\chi^2_{n-p}}}^{\star}$$

• If $n=2$, then $T \sim$ Ratio of 2 ind. Normal RV's $\Rightarrow T \sim$ Cauchy

5.3: Convergence concepts

Given a sequence of RV's $X_1, \ldots, X_n$, we want to study various notions of convergence to a RV X:

i) $X_n \xrightarrow{P} X$ convergence in probability

ii) $X_n \xrightarrow{a.s.} X$ convergence almost surely / with prob 1

iii) $X_n \overset{d}{\Rightarrow} X$ convergence in dist.

We already encountered (iii):

$X_n \overset{d}{\Rightarrow} X$ if $F_{X_n}(x) \longrightarrow F_X(x)$ at all points where F_X is continuous

Recall (Binomial approximation): a) If $X_n \sim \text{Binomial}(n, p_n)$ and $np_n \to \lambda$ as $n \to \infty$ then $X_n \overset{d}{\Rightarrow} Y$ where $Y \sim \text{Poisson}(\lambda)$

b) If $X_n \sim \text{Binomial}(n, p_n)$, $p_n \to p$ as $n \to \infty$

then $\frac{1}{\sqrt{n}}(X_n - np) \overset{d}{\Rightarrow} Z$ where $Z \sim N(0, p(1-p))$

(a consequence of CLT - see later)

Def: $X_n \xrightarrow{P} X$ if

$$\boxed{\forall \varepsilon > 0,\ P(|X_n - X| \geq \varepsilon) \to 0 \text{ as } n \to \infty}$$

Notice: $\forall \varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) \to 0$ as $n \to \infty$

$= P(s \in |X_n(s) - X(s)| \geq \varepsilon) \to 0$ as $n \to \infty$

where s generally depends on n

Thm (weak law of large numbers - WLLN). Let $X_1, \ldots, X_n \overset{iid}{\sim}$ RV's with $EX_i = \mu$, $\operatorname{Var} X_i = \sigma^2 < \infty$

Define $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$

Then $\bar{X}_n \xrightarrow{P} \mu$

Proof. $\forall \varepsilon > 0$, $P(|\bar{X}_n - \mu| \geq \varepsilon)$

$= P(|\bar{X}_n - \mu|^2 \geq \varepsilon^2)$

$\leq \frac{1}{\varepsilon^2} E|\bar{X}_n - \mu|^2$ by Chebyshev's inequality

$= \frac{1}{\varepsilon^2} \operatorname{Var} \bar{X}_n$

$= \frac{1}{\varepsilon^2} \frac{1}{n} \sigma^2 \to 0$ as $n \to \infty$

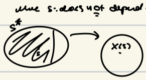
Def: $X_n \to X$ a.s. (w/ prob. 1) if

$$\boxed{P(\lim_{n \to \infty} |X_n - X| = 0) = 1 \equiv P(\lim_{n \to \infty} X_n = X) = 1} \quad (*)$$

Notice: $X_n \xrightarrow{a.s.} X$ if

$P(\{s : X_n(s) \to X(s)\}) = 1$

where s does not depend on n



Here, $X_n(s) \to X(s)\ \forall s \in S^* \Rightarrow P(S^*) = 1$

Remarks: Equivalently,

$\forall \varepsilon > 0\ P(\lim_{n \to \infty} |X_n - X| < \varepsilon) = 1 \iff P(\lim_{n \to \infty} |X_n - X| \geq \varepsilon) = 0$

Then, letting $\{\varepsilon = \frac{1}{2^k}\}_{k \geq 1}$ - a sequence that converges to 0

$P(\lim_{n \to \infty} |X_n - X| < \frac{1}{2^k},\ \forall k \in \mathbb{N})$

$= 1 - P(\exists k \in \mathbb{N},\ \lim_{n \to \infty} |X_n - X| \geq \frac{1}{2^k})$

$= 1 - \sum_{k=0}^{\infty} P(\lim_{n \to \infty} |X_n - X| \geq \frac{1}{2^k})$

$= 1$

$\Rightarrow (*)$ is valid

$\boxed{\textbf{Prop: } \text{If } X_n \xrightarrow{a.s.} X \text{ then } X_n \xrightarrow{P} X}$

Pf. If $X_n \xrightarrow{a.s.} X$ then

$$\forall \varepsilon > 0,\ 1 = P(\lim_{n\to\infty} |X_n - X| < \varepsilon)$$

$$= E\,\mathbb{1}(\lim_{n\to\infty} |X_n - X| < \varepsilon)$$

$$\le \lim_{n\to\infty} P(|X_n - X| < \varepsilon)$$, Fatou's lemma

Hence $\lim_{n\to\infty} P(|X_n - X| \ge \varepsilon) = 0$. ■

Ex: Useful to know the original def. of RV as a function on S: $s \to X(s)$

Let $X_n, X \in \{0,1\}$, $S = [0,1]$, P: uniform dist. on S, i.e. $P \sim \mathrm{Unif}(0,1)$

Define $\begin{cases} X_n(s) = s + s^n \\ X(s) = s \end{cases}$

Then $X_n(s) \longrightarrow X(s)\ \ \forall s \in [0,1)$

So, $P(s : \lim X_n(s) = X(s)) = P([0,1)) = 1$

even though $P(X_n \ne X) = P(s \ne 0) =$

Ex: Let $X(s) = 0\ \ \forall s \in [0,1]$

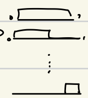
Let $X_1(s) = \mathbb{1}(s \in [0,1])$

$X_2(s) = \mathbb{1}(s \in [0, 1/2])$

$X_3(s) = \mathbb{1}(s \in [1/2, 1])$

$\vdots$

$X_7(s) = \mathbb{1}(s \in [3/4, 1])$



Then $P(s : |X_n(s) - X(s)| > \varepsilon)$

$= P(s : X_n(s) = 1) \le \frac{2}{n}$

$\to 0\ \ \forall \varepsilon > 0$

but $P(s : X_n(s) \to X(s))$

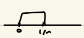
$= P(s : X_n \to 0)$

$= P(s : X_n(s) = 0 \text{ infinitely many times}) = 0$

Hence $X_n \xrightarrow{P} X$ but $X_n \overset{a.s.}{\not\to} X$.

Ex: suppose $X(s) = 0\ \ \forall s$

$X_n(s) = \mathbb{1}(0, \frac{1}{n})$



$= P(s : X_n(s) \to X(s))$

$= P(s : X_n(s) \to 0 \text{ as } n \to \infty)$

$= P(s > 0) = 1$

Ex: Let $X \sim N(0,1)$ and $Y = -X$

then $X \xrightarrow{d} Y$ but $P(X \ne Y) = 1$, i.e. $X \overset{a.s.}{\not\to} Y$.

Remarks:

convergence a.s. $\Rightarrow$ convergence in prob.

convergence prob. $\Rightarrow$ convergence dist.

Thm (strong law of large numbers - SLLN). Let $X_1, \dots, X_n \overset{iid}{\sim}$ RV's with $\begin{array}{|l} EX_i = \mu \\ E|X_i| < \infty \end{array}$

Define $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$

Then $\bar{X}_n \xrightarrow{a.s.} \mu$

Proof. similar in spirit but more technically involved than SLLN

Remark: $E|X_i| < \infty$ is very mild

$\boxed{\begin{array}{l} \underline{\text{Thm (CLT)}}.\ \text{If } X_1, X_2, \ldots \overset{iid}{\sim} \text{RV whe MGF exist in a neighborhood of } 0. \\ \text{Let } \left| \begin{array}{l} \mu = EX_i \\ \sigma^2 = \operatorname{Var} X_i > 0 \end{array} \right. \text{ and } \left| \begin{array}{l} \bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \\ Z \sim N(0,1) \end{array} \right. \\ \text{then } \left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \right) = \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \xRightarrow{d} Z \end{array}}$

<u>Remarks</u>: · perhaps most celebrated thm. in prob.

· widely applicable; only $\operatorname{Var} X_i < \infty$ is required

$\boxed{\begin{array}{l} \underline{\text{Corollary}}.\ \text{Let } \underline{Y_i = (X_i - \mu)/\sigma}, \text{ then } \left| \begin{array}{l} EY_i = 0 \\ EY_i^2 = 1 \end{array} \right. \\ \text{and we may write } \underline{\frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n) \xRightarrow{d} N(0,1)}. \end{array}}$

<u>P.</u> $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma} \right) = \frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \cdots + X_n - n\mu}{n} \right) = \frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n) \xRightarrow{d} N(0,1).$

<u>Proof (sketch)</u>. $M_Z(t) = Ee^{tZ} = e^{\frac{1}{2}t^2}$; it is enough to

show that $M_{\frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n)}(t) \longrightarrow M_Z(t)$ as $n \to \infty$

$\forall\ |t| < \delta$ for some $\delta > 0$

Now

$$M_{\frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n)}(t) = \left(M_Y(t/\sqrt{n}) \right)^n$$

Apply Taylor expansion which is valid for small $|t/(\sqrt{n}\sigma)|$.

$$M_Y(t/\sqrt{n}) = \underbrace{M_Y(0)}_{1} + \underbrace{M_Y^{(1)}(0)}_{EY=0} \frac{t}{\sqrt{n}} + \frac{1}{2} \underbrace{M_Y^{(2)}(0)}_{EY^2=1} \frac{t^2}{n} + o\left(\frac{t^2}{n} \right)$$

where $o\left(\frac{t^2}{n}\right)$ vanishes faster than $\frac{t^2}{n}$ as $n \to \infty$

Thus

$$\left(M_Y(t/\sqrt{n}) \right)^n = \left(1 + \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n} \right) \right)^n$$

$$= \lim_{n \to \infty} \left(1 + \frac{1}{2}\frac{t^2}{n} \right)^{\frac{2n}{t^2} \cdot \frac{t^2}{2}} = e^{t^2/2}. \ \square$$

<u>Lemma (1)</u>: If $\left| \begin{array}{l} X_n \xrightarrow{a.s.} X \\ Y_n \xrightarrow{a.s.} b : \text{constant} \end{array} \right.$

then,

$$aX_n + Y_n \xrightarrow{a.s.} aX + b$$

<u>Lemma (2)</u>: If $\left| \begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} b : \text{constant} \end{array} \right.$

then,

$$aX_n + Y_n \xrightarrow{P} aX + b$$

<u>Lemma (3)</u>: If $\left| \begin{array}{l} X_n \xRightarrow{d} X \\ Y_n \xrightarrow{P} b : \text{const.} \end{array} \right.$

then

$$aX_n + Y_n \xRightarrow{d} aX + b$$

<u>Remarks</u>: · L1 and L2 are almost immediate from definition

· L3 is known as Slutsky's lemma/thm.; <u>P.</u> based on characterization: $X_n \xRightarrow{d} X$

$\Leftrightarrow$

for all continuous and bounded function $f(x)$

$Ef(X_n) \to Ef(X)$ as $n \to \infty$ (higher-dim def. as opposed to cdf - not covered)