Recap:

① ~~Q~~. Properties of a random sample

② We have seen the definition of t-distribution, F-distribution and some of their properties

Order Statistic

There is an electronic device which runs on 20 batteries. ~~Each~~ The lifetime of each battery has the same distribution. Let the electronic device dies when 15 batteries die. What is the distribution of the lifetime of the electronic device?

Let $X_1, \dots, X_{20}$ be the lifetimes of 20 batteries. We know $X_1, \dots, X_{20}$ is a random sample.

$$X_{(1)} = \min_{1 \le i \le 20} X_i, \quad X_{(2)} = \min\left[\{X_1, \dots, X_{20}\} \setminus \{X_{(1)}\}\right], \dots$$

$$\dots, \quad X_{(20)} = \max_{1 \le i \le 20} X_i$$

$X_{(1)} < X_{(2)} < \dots < X_{(20)}$ These are known as the order statistics of the random sample.

$X_{(15)} =$ r.v. representing the time when 15 batteries die.

Marginal density of $X_{(i)}$.

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!\,(n-i)!} f_X(x) \left[F_X(x)\right]^{i-1} \left[1-F_X(x)\right]^{n-i}$$

$$-\infty < x < \infty$$

$\binom{n}{i} \longrightarrow$ ith order statistic can be drawn

$n-1$

ith orde.

$\binom{n-1}{i-1} \longrightarrow$ $(i-1)$ order static below the ith order statistic can be done in $\binom{n-1}{i-1}$ ways.

ith order stat. can take the value x with "prob." (density) $f_X(x)$.

And the $(i-1)$ order statistics below the ith one will have to be necessarily smaller than x.

each of them is $\leq x$ with prob. $F_X(x)$

then $(n-i)$ order statistic above the ith one will have to be $> x$. Each of them is $> x$ with prob. $1-F_X(x)$

$$f_{X_{(i)}}(x) = {}^{n}C_{1}\, {}^{n-1}C_{i-1}\, f_X(x) \left[F_X(x)\right]^{i-1} \left[1-F_X(x)\right]^{n-i}$$

$$= \frac{n\,.\,(n-1)!}{(i-1)!\,(n-i)!} f_X(x) \left[F_X(x)\right]^{i-1} \left[1-F_X(x)\right]^{n-i}$$

Joint density of $X_{(i)}, X_{(j)}$

$$f_{X_{(i)}, X_{(j)}}(x_1, x_2) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\left[F_X(x_1)\right]^{i-1}\left[1-F_X(x_2)\right]^{n-j}$$

$$\times \left[F_X(x_2) - F_X(x_1)\right]^{j-i-1} f_X(x_1) f_X(x_2),$$

$$x_1 \le x_2 .$$

ith order statistic can be chosen in ${}^{n}C_{1}$ ways. Once it has been chosen, the j th order statistic can be chosen in ${}^{n-1}C_{1}$ ways.

Once they are both chosen, the $(i-1)$ order statistics below the ith one can be chosen in ${}^{n-2}C_{i-1}$ ways.

Once ith, jth and all order statistic below the ith one are chosen the $(n-j)$ order statistics above the jth one can be chosen in ${}^{(n-2-(i-1))}C_{n-j}$

Now, the density of ith order statistic being equal to x_1 is $f_X(x_1)$ and the density of the jth order statistic being x_2 is $f_X(x_2)$.

the prob. that $(i-1)$ order statistics below the ith one $\le x_1$ is $[F_X(x_1)]^{i-1}$

the prob. that $(n-j)$ order statistics above the jth one $\geq x_2$ is $[1-F_X(x_2)]^{n-j}$

Also, $(j-i-1)$ order statistics which are in between the ith and jth order statistic lie between x_1 and x_2 with prob.

$$[F_X(x_2)-F_X(x_1)]^{j-i-1}$$

$$f_{X_{(i)},X_{(j)}}(x_1,x_2) = {}^{n}C_1\ {}^{(n-1)}C_1\ {}^{(n-2)}C_{i-1}\ {}^{(n-2-i+1)}C_{n-j}\ f_X(x_1)\, f_X(x_2)\ [F_X(x_1)]^{i-1}\,[1-F_X(x_2)]^{n-j}\ [F_X(x_2)-F_X(x_1)]^{j-i-1}, \quad x_1 \leq x_2 .$$

Example: $X_1, \ldots, X_{20}$ are lifetimes of 20 batteries. Let the lifetimes are all independently and identically distributed with density $Exp(\lambda)$.

$X_1, \ldots, X_{20} \overset{iid}{\sim} Exp(\lambda)$.

Goal: To find the distribution of $X_{(15)}$.

$$f_{X_{(15)}}(x) = \frac{20!}{14!\,5!}[\lambda \exp(-\lambda x)][1-\exp(-\lambda x)]^{14}[\exp(-\lambda x)]^{5}, \quad x>0 .$$

~~Example: a policy of five members~~

<u>Example</u>: Five members of a family are in an insurance policy. ~~[illegible]~~ The policy says that they will receive huge amount of money if two people die. If the lifespan distributions of all people in the family are same. What is the distribution of the time when they receive the money?

Convergence of sequence of random variables

Lets say we have a random sample ~~of~~ $X_1, \dots, X_n$ where n is very large. For example $n = 20000$

We are interested to know the distribution of some function of $X_1, \dots, X_{20000}$, let's say the distribution of $\bar{X} = \frac{1}{20000} \sum_{i=1}^{20000} X_i$.

We might encounter a number of situations when the distribution of $\bar{X}$ is difficult to treat analytically, but the distribution can be approximated by a well known distribution for large n.

Convergence concepts

Convergence in probability

A sequence of random variables $X_1, \ldots\ldots$ converges to a random variable X if

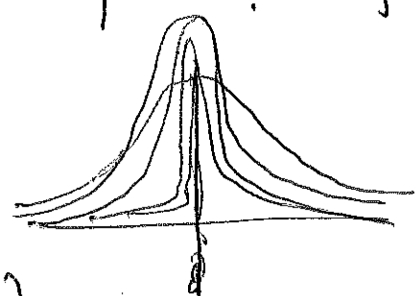
for any $\varepsilon > 0$,

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

$\{\frac{1}{n}\} \to 0$

<u>Example:</u> $X_n \sim N(0, \frac{1}{n})$.

<u>Qn:</u> Does X_n converge in prob? If yes, then to which r.v.?



Guess: $X = 0$ w.p. 1.

$$P(|X_n - X| \geq \varepsilon) = P(|X_n| \geq \varepsilon)$$

$$\leq \frac{E X_n^2}{\varepsilon^2} \quad \text{By Chebychev's inequality}$$

$$= \frac{1}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty,$$

<u>Qn:</u> If $X_1, \ldots\ldots$ converges in prob. to X (formally denoted by $X_n \xrightarrow{P} X$), does $g(X_1), \ldots, \ldots$ converges in prob. to $g(X)$? If yes, under what condition on g?

$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$ when g is a continuous function.

<u>Qn:</u> $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. ~~Where does~~
Does $X_n + Y_n$ converges to some random variable in prob.?

$X_n + Y_n \xrightarrow{P} X + Y$

<u>Convergence in distribution</u>

$X_1, \ldots, \ldots$ is a sequence of random variables. This sequence is said to converge in distribution to a random variable X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, for all continuity points of x of F_X. where F_{X_n} and F_X are the cumulative density functions of X_n and X

<u>Example:</u> $X_1, \ldots, \ldots$ is a random sample from $U(0,1)$. Does $n(1 - X_{(n)})$ converge in distribution?

We know $X_n \sim U(0,1)$. Let $Z_n = n(1 - X_{(n)})$.

~~Comp~~ <u>Step 1</u>: Compute the C.D.F. of Z_n.

$$F_{Z_n}(x) = P(Z_n \le x) = P(n(1 - X_{(n)}) \le x)$$
$$= P\left(X_{(n)} \geq 1 - \frac{x}{n}\right)$$

$$= 1 - P\left(X_{(n)} < 1 - \frac{x}{n}\right)$$

$$= 1 - P\left(X_1 < 1 - \frac{x}{n}, X_2 < 1 - \frac{x}{n}, \ldots, X_n < 1 - \frac{x}{n}\right)$$

$$= 1 - P\left(X_1 \leq 1 - \frac{x}{n}\right) P\left(X_2 < 1 - \frac{x}{n}\right) \cdots P\left(X_n < 1 - \frac{x}{n}\right)$$

as $X_1, \ldots, X_n$ are i.i.d.

$$= 1 - \left(1 - \frac{x}{n}\right)^n$$

$$\xrightarrow{n \to \infty} 1 - e^{-x}$$

$$\lim_{n \to \infty} F_{Z_n}(x) = 1 - e^{-x} = F_Z(x) \text{ when } Z \sim \mathrm{Exp}(1)$$

$n(1 - X_{(n)})$ is a sequence of random variables that converges in distribution to Exp(1) random variable.

When X_n converges in distribution to X (formally written as $X_n \xrightarrow{d} X$), does it mean that $X_n \xrightarrow{P} X$.

This is not true.

<u>Counterexample</u>: X is a random variable s.t. $P(X=0) = P(X=1) = \frac{1}{2}$.

$\Rightarrow$ X and $1-X$ have the same distribution.

Lets take $X_n = X \quad \forall n$.

<u>back ①</u>

clearly each X_n has the same distribution as $1-X$.

By default $X_n \xrightarrow{d} 1-X$

$$P(|X_n-(1-X)|>\tfrac{1}{2}) = P(|X-(1-X)|>\tfrac{1}{2})$$
$$= P(|2X-1|>\tfrac{1}{2})$$

For both $X=0$ and 1 $|2X-1|=1$

which means $P(|X_n-(1-X)|>\tfrac{1}{2}) = 1$

$\Rightarrow X_n \overset{P}{\not\to} 1-X$.

Hence convergence in distribution does not mean convergence in prob.

<u>However</u>: $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.

<u>Example</u>: $X_n = \frac{1}{n}$ w.p. 1

$X = 0$ w.p. 1

$\frac{1}{n} \to 0$ We should expect $X_n \xrightarrow{d} X$ in this case.

However, $F_{X_n}(x) = \begin{cases} 0 & \text{if } x < \frac{1}{n} \\ 1 & \text{o.w.} \end{cases}$

$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{o.w.} \end{cases}$ $\boxed{F_X(x)} = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{o.w.} \end{cases}$

there is a discrepancy at $x=0$ and it occurs as F_X is not cont. at $x=0$.

back ②