# MATH 340/640 Lec 16

Order statistics (p160 in Hoerl, Port, Stone)

Let $X_1, X_2, \ldots, X_n$ be a collection of cont. rv's. Let

$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be called the order statistics; which are rv's defined as

$$X_{(1)} = \min\{X_1, \ldots, X_n\} = \text{smallest}\{X_1, \ldots, X_n\}$$
$$X_{(2)} = 2^{nd} \text{ smallest } \{X_1, \ldots, X_n\}$$
$$\vdots$$
$$X_{(n)} = \max\{X_1, \ldots, X_n\} = \text{largest}\{X_1, \ldots, X_n\}$$

Let's let $n = 17$. Let $X_1, \ldots, X_{17} \overset{iid}{\sim} U(0,1)$.

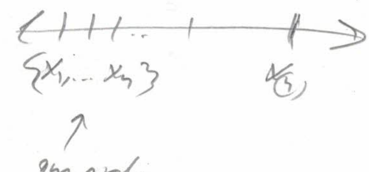
What does min look like? Max? $X_{(9)}$? the middle $X_{(3)}$? $X_{(14)}$?

$\boxed{\text{DEMO}}$

Let's find CDF & PDF of the maximum in general

best you can do!

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \ldots, X_n \le x)$$




if $X_1, \ldots, X_n$ ind

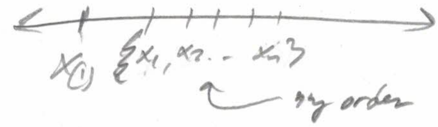
$$= \prod_{i=1}^{n} P(X_i \le x) = \prod_{i=1}^{n} F_{X_i}(x)$$

if $X_1, \ldots, X_n$ iid

$$= \prod_{i=1}^{n} F(x) = F(x)^n$$

$$\Rightarrow f_{X_{(n)}}(x) = n\, f_X(x)\, F_X(x)^{n-1}$$

Let's find the CDF & PDF of the minimum




$$F_{X_{(1)}}(x) = P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, X_2 > x, \ldots, X_n > x)$$ best you can do!

if $X_1 \ldots X_n \overset{ind}{\sim}$ $$= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \prod_{i=1}^{n} (1 - F_{X_i}(x))$$

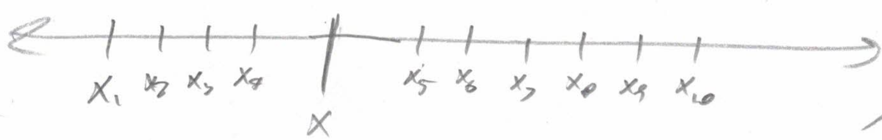
if $X_1 \ldots X_n \overset{iid}{\sim}$ $$= 1 - \prod_{i=1}^{n} (1 - F(x)) = 1 - (1 - F(x))^n$$

$$\Rightarrow f_{X_{(1)}}(x) = -n(-f(x))(1 - F(x))^{n-1} = n f(x)(1 - F(x))^{n-1}$$

These two exercises give us intuition on how to solve the general problem of finding $F_{X_{(k)}}(x)$, $f_{X_{(k)}}(x)$

Let's let $n = 10$, Consider $X_{(4)}$. This means the dens. of the $4^{th}$ smallest of the 10 realizations. What is the probability the following happens? In this precise order...

P(




)

$$= P(X_1 \le x, X_2 \le x, X_3 \le x, X_4 \le x, X_5 > x, X_6 > x, X_7 > x, X_8 > x, X_9 > x, X_{10} > x)$$ best you can do

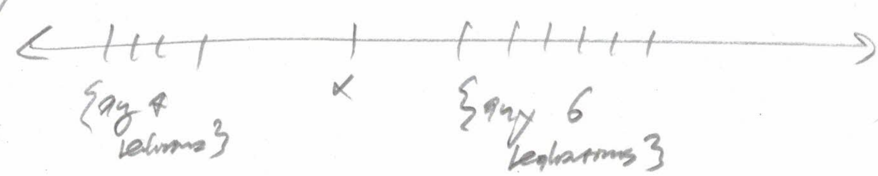
if $X_1, \ldots, X_{10} \overset{ind}{\sim}$

$$= \prod_{i=1}^{4} P(X_i \le x) \prod_{i=5}^{10} P(X_i > x) = \prod_{i=1}^{4} F_{X_i}(x) \prod_{i=5}^{10} (1 - F_{X_i}(x))$$

if $X_1, \ldots X_{10} \overset{iid}{\sim}$

$$= F(x)^4 (1 - F(x))^6$$

How do we compute the following prob?

$P\Big($  $\Big)$


$$= P(\text{any 4 } X_i\text{'s} \le x \text{ and the other 6 } X_i\text{'s} > x)$$

$$= \sum_{\substack{\text{all subsets} \\ S \subset \{1,2,\dots,10\} \\ \text{s.t. } |S|=4}} P\left(X_{S_1} \le x, \dots X_{S_4} \le x,\ X_{S_1^c} > x, \dots X_{S_6^c} > x\right)$$ best you can do!

if $X_1, \dots X_n$ ind

$$= \sum_{\text{all subsets}} \prod_{i=1}^{4} F_{X_{S_i}}(x) \prod_{i=1}^{6}\left(1 - F_{X_{S_i^c}}(x)\right)$$

if $X_1, \dots X_n \overset{iid}{\sim}$

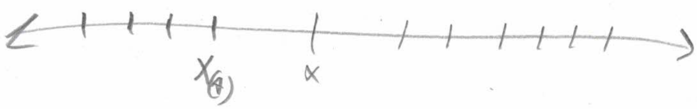
$$= \sum_{\text{all subsets}} \prod_{i=1}^{4} F(x) \prod_{i=1}^{6}(1 - F(x)) = \binom{10}{4} F(x)^4 (1 - F(x))^6$$

Now we solve the problem for real.

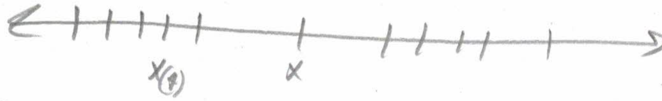
if $X_1 \ldots X_{10} \overset{iid}{\sim}$

$$F_{X_{(4)}}(x) = P(X_{(4)} \le x) = \binom{10}{4} F(x)^4 (1-F(x))^6 + \binom{10}{5} F(x)^5 (1-F(x))^5 + \ldots + \binom{10}{10} F(x)^{10} (1-F(x))^0$$

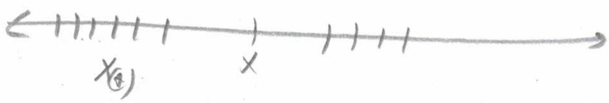
$$= \sum_{j=4}^{10} \binom{10}{j} F(x)^j (1-F(x))^{n-j}$$

$P\Big($  $\Big)$



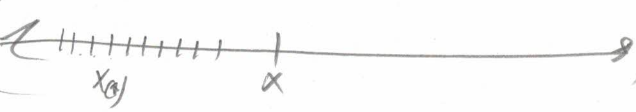
$+P\Big($  $\Big)$



$+P\Big($ 



⋮

$+P\Big($  $\Big)$



If you look at this proof, there was nothing special about $n=10$, $k=4$, so...

if $X_1, \ldots, X_n \overset{iid}{\sim}$

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j}$$

Make sure this works for min/max

$$F_{X_{(n)}} = \sum_{j=n}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} = \binom{n}{n} F(x)^n (1-F(x))^{n-n} = F(x)^n \checkmark$$

$$F_{X_{(1)}}(x) = \sum_{j=1}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} = \left( \sum_{j=0}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} \right) - \binom{n}{0} F(x)^0 (1-F(x))^n$$

$$= \big(F(x) + (1-F(x))\big)^n - (1-F(x))^n = 1-(1-F(x))^n. \checkmark$$

Now, let's get the PDF of $X_{(k)}$

$$f_{X_{(k)}}(x) = \frac{d}{dx}\left[\sum_{j=k}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j}\right]$$

$$= \sum_{j=k}^{n} \binom{n}{j} \frac{d}{dx}\left[\underbrace{F(x)^j}_{u}\underbrace{(1-F(x))^{n-j}}_{v}\right]$$

$\frac{d}{dx}[uv] = u\,dv + v\,du$

$$= \sum_{j=k}^{n} \frac{n!}{j!(n-j)!}\left(F(x)^j (n-j)(-f(x))(1-F(x))^{n-j-1} + (1-F(x))^{n-j}\, j\, f(x)\, F(x)^{j-1}\right)$$

$$= \sum_{j=k}^{n} \frac{n!}{\underset{(j-1)!}{\cancel{j!}}\,(n-j)!} \cancel{j} f(x)\, F(x)^{j-1}(1-F(x))^{n-j} - \sum_{j=k}^{n-1} \frac{n!}{j!\,\underset{(n-j-1)!}{\cancel{(n-j)!}}} \cancel{(n-j)} f(x)\, F(x)^{j}(1-F(x))^{n-j-1}$$

if $j=n$ this is zero

let $l = j+1 \Rightarrow j = l-1 \Rightarrow n-j-1 = n-(l-1)-1 = n-l$

$$= \underbrace{\sum_{j=k}^{n} \frac{n!}{(j-1)!(n-j)!} f(x)\, F(x)^{j-1}(1-F(x))^{n-j}} - \underbrace{\sum_{l=k+1}^{n} \frac{n!}{(l-1)!(n-l)!} f(x)\, F(x)^{l-1}(1-F(x))^{n-l}}$$

¡identical expressions!!!

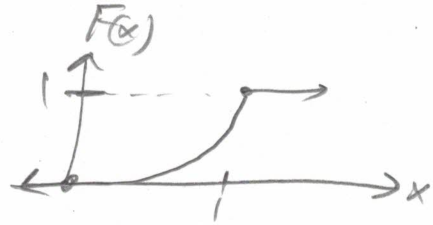
$$\boxed{f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x)\, F(x)^{k-1}(1-F(x))^{n-k}}$$

if $X_1, \dots, X_n \overset{iid}{\sim}$

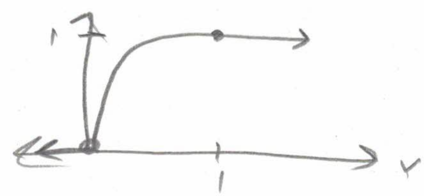
let's do an example. let $X_1, \ldots, X_n \overset{iid}{\sim} U(0,1)$

$$f(x) = \mathbb{1}_{x \in [0,1]}, \qquad F(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } x \in [0,1) \\ 1 \text{ if } x > 1 \end{cases}$$

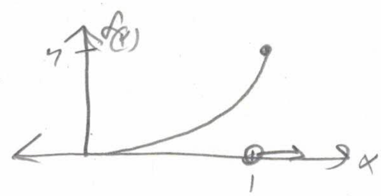
$$F_{X_{(n)}}(x) = F(x)^n = \begin{cases} 0 \text{ if } x < 0 \\ x^n \text{ if } x \in [0,1] \\ 1 \text{ if } x > 1 \end{cases}$$



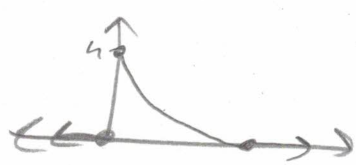

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n = \begin{cases} 0 \text{ if } x < 0 \\ 1-(1-x)^n \text{ if } x \in [0,1) \\ 1 \text{ if } x > 1 \end{cases}$$



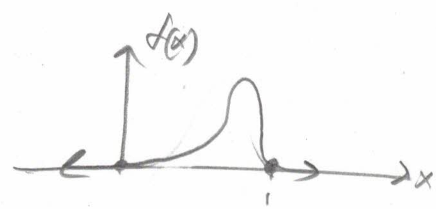

$$f_{X_{(n)}}(x) = n x^{n-1} \mathbb{1}_{x \in [0,1]}$$




$$f_{X_{(1)}}(x) = n(1-x)^{n-1} \mathbb{1}_{x \in [0,1]}$$




$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k} \mathbb{1}_{x \in [0,1]}$$




$$= \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)} x^{k-1}(1-x)^{n-k} \mathbb{1}_{x \in [0,1]}$$

$$f_{X_{(n)}} = \frac{\Gamma(n+1)}{\Gamma(n)\,\Gamma(1)} x^{n-1}(1-x)^0 \mathbb{1}_{x \in [0,1]}$$

$$f_{X_{(1)}} = \frac{\Gamma(n+1)}{\Gamma(1)\,\Gamma(n)} x^0(1-x)^{n-1} \mathbb{1}_{x \in [0,1]}$$

New topic: kernels. We define

$p_X(x) \propto k_X(x)$ for discrete $X$ to mean $\exists\, c \in \mathbb{R}$ $p_X(x) = c\,k_X(x)$

$f_X(x) \propto k_X(x)$ for cont. $X$ to mean $\exists\, c \in \mathbb{R}$ $f_X(x) = c\,k_X(x)$

and $k(x)$ is called the "kernel" of the PMF / PDF and $c$, which is not a function of $x$, is called the normalization constant.

The kernel, like the ch.f., determines the r.v. since you can always use $k(x)$ to rehydrate $p(x)$ or $f(x)$. How?

$$1 = \sum_{x \in \mathbb{R}} p(x) = \sum_{x \in \mathbb{R}} c\,k(x) \Rightarrow \frac{1}{c} = \sum k(x) \Rightarrow c = \frac{1}{\sum k(x)}$$

$$1 = \int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} c\,k(x)\,dx \Rightarrow \frac{1}{c} = \int_{\mathbb{R}} k(x)\,dx \Rightarrow c = \frac{1}{\int k(x)\,dx}$$

---

Time to review all of our r.v.s!

$$X \sim \text{Bin}(n,p) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^n (1-p)^{-x} \mathbb{1}_{x \in \{0,\ldots,n\}} \underbrace{n!(1-p)^n}_{c} \underbrace{\frac{1}{x!(n-x)!}\left(\frac{p}{1-p}\right)^x \mathbb{1}_{x \in \{0,\ldots,n\}}}_{k(x)}$$

$$\propto \frac{1}{x!(n-x)!}\left(\frac{p}{1-p}\right)^x \mathbb{1}_{x \in \{0,\ldots,n\}}$$

this means you have a binomial. Strange...

$$c = \frac{1}{\sum_{x \in \mathbb{R}} x!(n-x)!\left(\frac{p}{1-p}\right)^x \mathbb{1}_{x \in \{0,\ldots,n\}}}$$

$$X \sim N(\mu, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$= e^{-\frac{1}{2\sigma^2}x^2}\, e^{-\frac{1}{2\sigma^2}(-2x\mu)}\, \underbrace{e^{-\frac{1}{2\sigma^2}\mu^2}}$$

$$\propto \underbrace{e^{-\frac{1}{2\sigma^2}x^2}\, e^{\frac{x\mu}{\sigma^2}}}_{k(x)}$$

$$\Rightarrow c = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}\mu^2}$$