Lecture 12

10/19/2

1774 – 1st "Law of Errors". Imagine you're trying to measure something, a quantity v, but your measurements have error, epsilon, so your measurement M is a rv looking like: $M = v + \text{epsilon}$. So what is a good model for the error (epsilon)? It makes sense for $E[\text{epsilon}] = 0$. $Med[\text{epsilon}] = 0$. and symmetric.

How about.... $U(-L, L)$?

It also makes sense for larger errors (in magnitude) to be less probable than smaller errors. $\rightarrow \forall \varepsilon > 0 \; f'(\varepsilon) < 0$.

$\forall \varepsilon > 0 \; f''(\varepsilon) = f'(\varepsilon) \rightarrow$ solve $\rightarrow f(\varepsilon) = ce^{-d\varepsilon} \rightarrow$ Laplace(0,1).

$X \sim Exp(1) = e^{-x} \mathbb{1}_{x \geq 0}$. Let $Y = \frac{1}{\lambda} x^{1/k}$ s.t. $\lambda, k > 0$

$g(x) = \frac{1}{\lambda} x^{\frac{1}{k}}$.

$Y \sim f_Y(y) = ?$

① Find the inverse function.

$\Rightarrow x(y) = x^{1/k} \Rightarrow x = \lambda^k y^k = g^{-1}(y)$

$x = (\lambda y)^k = g^{-1}(y)$

② Derive...

$\left|\frac{d}{dy}[g^{-1}(y)]\right| = \left|\frac{d}{dy}[\lambda^k y^k]\right| = |k\lambda^k y^{k-1}| = k\lambda^k \underbrace{y^{k-1}}_{y \geq 0}$

③ formula....

$f_Y(y) = f_X(g^{-1}(y)) \left|\frac{d}{dy}[g^{-1}(y)]\right| = e^{-(\lambda y)^k} \mathbb{1}_{\lambda^k y^k \geq 0} \cdot k\lambda^k y^{k-1}$

$\mathbb{1}_{\lambda^k y^k \geq 0} \to y^k \geq 0$

① →

$= k\lambda^k y^{k-1} \cdot e^{-(\lambda y)^k} \mathbb{1}_{y\geq 0} = k\lambda(\lambda y)^{k-1} e^{-(\lambda y)^k} \mathbb{1}_{y\geq 0} =$ Weiball(k,λ) 1951.

Note – Weibull $(1,\lambda) = (1)\lambda(1)y^{\cancel{(1-1)}} e^{-(\lambda y)^1} \mathbb{1}_{y\geq 0}$

$= \lambda e^{-\lambda y} \mathbb{1}_{y\geq 0} = Exp(\lambda)$. ✓

K is really cool, this is the main property:

$K=1$.

$P(Y\geq y+c \mid Y\geq c) = P(Y\geq y)$. = memorylessness. = survival equally as likely.

Ex: $y=3, c=14$

$\Rightarrow P(Y\geq 17 \mid Y\geq 14) = P(Y\geq 3)$

If $K>1$, $P(Y\geq y+c \mid Y\geq c) < (P(Y\geq y))$. survival is less likely

If $K<1$, $P(Y\geq y+c \mid Y\geq c) > P(Y\geq y)$. survival is more likely.

Order Statistics (p.160 in the textbook).

NEW TOPIC

Let $X_1, X_2, \ldots, X_n$ be a collection of continuous rv's and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the "order statistic" defined as:

$X_1 = \min\{X_1, \ldots, X_n\}$

$X_{(n)} = \max\{X_1, \ldots, X_n\}$

$\vdots$

$X_{(k)} = k^{th}$ largest $\{X_1, \ldots, X_n\}$

$x_1 = 9, x_2 = 2, x_3 = 12, x_4 = 7$.

$x_{(1)} = 2, x_{(2)} = 7, x_{(3)} = 9$.

$x_{(4)} = 12$

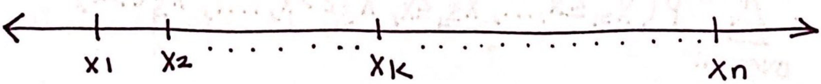
$\rightarrow r = 12-2 = 10$

$R := X_n - X_1$ "range".

We want to find the CDF & PDF of the order statistics.

We'll start by looking at the CDF of the maximum.

$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x \;\&\; X_2 \leq x \;\&\; \ldots \;\&\; X_n \leq x)$ $\overset{\text{if independent}}{=} \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x) \overset{\text{if iid}}{=} F_X(x)^n$



$$f_{X_n}(x) \overset{\text{if iid}}{=} \frac{d}{dx}\left[F_{X_{(n)}}(x)\right] = \frac{d}{dx}\left[F_X(x)^n\right] = n\cdot F_X(x)^{n-1}\cdot f_X(x)$$

Let's now find the CDF / PDF of the minimum.

$$F_{X_{(1)}}(x) = P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x \,\&\, X_2 > x \,\&\, \dots \,\&\, X_n > x)$$

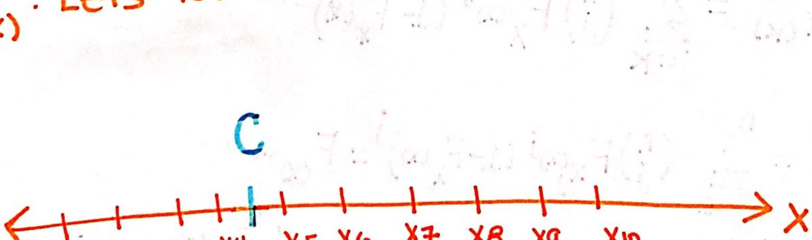
$$\underset{\text{if independent}}{=} 1 - P(X_1>x)(P(X_2>x))\cdots P(X_n>x) = 1 - \prod_{i=1}^{n}\left(1-F_{X_i}(x)\right).$$

if iid

$$\hookrightarrow = 1-\left(1-F_X(x)\right)^n$$

$$f_{X_{(1)}}(x) \overset{\text{if iid}}{=} \frac{d}{dx}\left[1-\left(1-F_X(x)\right)^n\right] = n f_X(x)\left(1-F_X(x)\right)^{n-1}$$

Let's now find the CDF/PDF for the kth order statistics $X_{(k)}$. Let's let $n=10$, $k=4$.



$$P(X_1 \le c \,\&\dots\&\, X_4 \le c \,\&\, X_5 > c \,\&\, \dots \,\&\, X_{10} > c)$$

if independent

$$\hookrightarrow = \prod_{i=1}^{4} P(X_i \le c)\prod_{i=5}^{10} P(X_i > c) = \prod_{i=1}^{4} F_{X_i}(c)\prod_{i=5}^{10}\left(1-F_{X_i}(c)\right) \overset{\text{if iid}}{=} F_X(c)^4\left(1-F_X(c)\right)^6$$

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

$$= \sum_{\substack{\text{over all subsets}\\ S \text{ size } 4,\ S^c \text{ size } 6}} P(X_{S_1} \le x, \ldots, X_{S_4} \le x, X_{S_1^c} > x, \ldots, X_{S_6^c} > x)$$

if independent $\qquad\qquad\qquad\qquad$ if iid

$$\overset{\downarrow}{=} \sum_{\text{'' ''}} \prod_{i=1}^{4} F_{X_{S_i}}(x) \prod_{i=1}^{6} F_{X_{S_i^c}}(x) \overset{\swarrow}{=} \sum_{\text{'' ''}} F_X(x)^4 (1-F_X(x))^6$$

$$= \underbrace{\binom{10}{4}} F_X(x)^4 (1-F_X(x))^6$$

$F_{X_{(4)}}(x) = P(X_{(4)} \le x) = P(4\ X_i\text{'s} \le x,\ 6\ X_i\text{'s} > x) +$
$P(5\ X_i\text{'s} \le x,\ 5\ X_i\text{'s} > x) +$
$\ldots + P(10\ X_i\text{'s} \le x,\ 0\ X_i\text{'s} > x).$

if iid

$$\overset{\downarrow}{=} \sum_{j=4}^{10} \binom{10}{j} F_X(x)^j (1-F_X(x))^{10-j}$$

General case:

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

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

$$F_{X_{(1)}}(x) = \sum_{j=1}^{n} \binom{n}{j} \underbrace{F_X(x)^j}_{a} \underbrace{(1-F_X(x))}_{b}{}^{n-j} = \left(\sum_{j=0}^{n} \binom{n}{j} F_X(x)^j (1-F_X(x))^{n-j}\right)$$

$-\binom{n}{0} F_X(x)^0 (1-F_X(x))^{n-0}$ Binominal Thm:

$= (F_X(x) + 1 - F_X(x))^n - (1-F_X(x))^n = 1 - (1-F_X(x))^n$

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

$= \sum_{j=K}^{n} \binom{n}{j} \frac{d}{dx}\left[\overbrace{F_X(x)^j}^{u}\ \overbrace{(1-F_X(x))^{n-j}}^{v}\right]$

$\frac{d}{dx}[uv] = uv' + u'v$

$u' = j f_X(x) F(x)^{j-1}$

$v' = (n-j) f(x) (1-F_X(x))^{n-j-1}$

To be continued.....