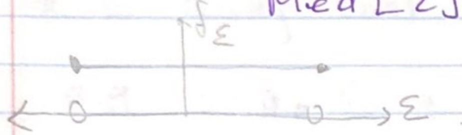
Lecture 12 10/19/2020

1774 - First "law of errors". Imagine you're trying to measure something, a constant quantity ~~variable~~, but your measurements have random error, ε, so your measurement M is a rv looking like: $M = v + \varepsilon$. So what is a good model for the error (ε)?
It makes sense for $E[\varepsilon] = 0$.
$Med[\varepsilon] = 0$ & symmetric.



How about $U(-L, L)$

It also makes sense for larger errors (in magnitude) to be less probable than smaller errors. $\Rightarrow \forall_{\varepsilon} > 0 \quad f'(\varepsilon) < 0$

It . . .

$\forall_{\varepsilon} > 0 \quad f''(\varepsilon) = f'(\varepsilon) \underset{solve.}{\Longrightarrow} f(\varepsilon) = ce^{-d\varepsilon} \Rightarrow \text{Laplace}(0,1)$

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

$$\Downarrow$$

$Y \sim f_Y(y) = ?$ Inverse function $\lambda y = x^{1/k}$

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

$$\left|\tfrac{d}{dy}\left[g^{-1}(y)\right]\right| = \left|\tfrac{d}{dy}\left[\lambda^k y^k\right]\right| = \left|k\lambda^k y^{k-1}\right| = k\lambda^k y^{k-1}$$

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

$y \geq 0$
$y^k \geq 0$

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

$$= k\lambda(\lambda y)^{k-1} e^{-(\lambda y)^k} \mathbb{1}_{y \geq 0} = \text{Waiball}(k, \lambda)$$ 1951

Note

$$\text{Waiball}(1,\lambda) = (1)\lambda(1)\cancel{y^{(1)-1}} e^{-(\lambda y)^1} \mathbb{1}_{y\geq 0}.$$

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

k is really cool... this is the main property:

e.g. $Y = 3$, $C = 14$ $\Big\{ P(Y \geq 17 \mid Y \geq 14) = P(Y \geq 3)$
memorylessness

$k = 1$ $P(Y \geq y + c \mid Y \geq c) = P(Y \geq y)$

$k > 1$ $P(Y \geq y + c \mid Y \geq c) < P(Y \geq y)$ - survival less likely as time goes on.

$k < 1$ $P(Y \geq y + c \mid Y \geq c) > P(Y = y)$ - survival more likely as time goes on.

You will prove these facts on the HW.

Order statistics (p160 in the text book)

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

$X_{(1)} := \min\{X_1, X_2, \ldots, X_n\}$,

$\vdots$

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

$\vdots$

$X_{(n)} := \max\{X_1, X_2, \ldots, X_n\}$

$R := X_{(n)} - X_{(1)}$ Range.

$X_{(1)} = \min\{X_1, \ldots X_n\}$ $\quad X_1 = 9,\ X_2 = 2,\ X_3 = 12,\ X_4 = 7$

$X_{(n)} = \max\{X_1 \ldots X_n\}$ $\quad X_{(1)} = 2,\ X_{(2)} = 7,\ X_{(3)} = 9,\ X_{(4)} = 12$

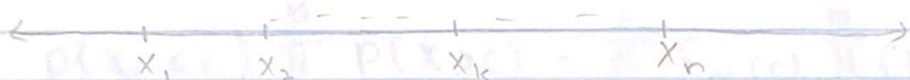
$r = 12 - 2 = 10.$

We want to find CDF and PDF of the k^{th} order statistics. We'll start by looking at the CDF of the maximum

$$F_{X_{(n)}}(x) = P(\underbrace{X_{(n)} \leq x}_{\text{event}}) = P(X_1 \leq x \ \&\ X_2 \leq x \ \& \ldots \&\ X_n \leq x)$$

$$\overset{\text{if indep}}{=} \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x)$$

$$\overset{\text{if iid}}{=} F_X(x)^n.$$



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

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

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

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x)$$

$$= 1 - P(X_1 > x \ \&\ X_2 > x \ \& \ldots \&\ X_n > x)$$

$$\overset{\text{if indep}}{=} 1 - P(X_1 > x)\, P(X_2 > x) \cdot \ldots \cdot (P(X_n > x)$$

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

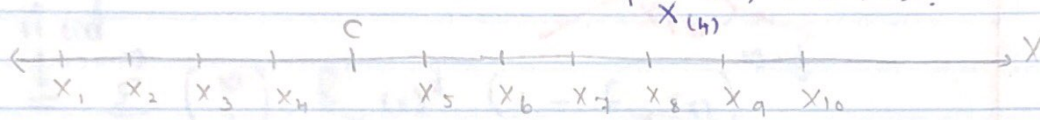
$$\overset{\text{if iid}}{=} 1 - \left(1 - F_X(x)\right)^n.$$

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

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

Let's now find the CDF / PDF for the k^{th} order statistic, $X_{(k)}$.

Let's let $n = 10$, $k = 4$. $\quad F_{X_{(4)}}(x) = \ldots$



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

$$\overset{\text{if indep}}{=} \prod_{i=1}^{4} P(X_i \le c) \prod_{i=5}^{10} P(X > c) = \prod_{i=1}^{4} F_{X_i}(c) \prod_{i=5}^{10} (1 - F_{X_i}(c))$$

$$\overset{\text{if iid}}{=} F_X(c)^4 \left(1 - F_X(c)\right)^6.$$

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

$$= \sum_{\substack{\text{over all subsets} \\ S \text{ size } 4,\ S^c \\ \text{size } 6}} P(X_{s_1} \le x, \ldots, X_{s_4} \le x, X_{s_1^c} > x, \ldots, X_{s_6^c} > x)$$

$$\overset{\text{if indep}}{=} \sum_{\text{same}} \prod_{i=1}^{4} F_{X_{s_i}}(x) \prod_{i=1}^{6} F_{X_{s_i^c}}(x) \overset{\text{if iid}}{=} \sum_{\text{same}} F_X(x)^4 \left(1 - F_X(x)\right)^6.$$

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

$$F_{X_{(4)}}(x) = P\left(X_{(4)} \leq x\right) = P\left(4\ X_i\text{'s} \leq x,\ 6\ X_i\text{'s} > x\right) +$$

$$P\left(5\ X_i\text{'s} \leq x,\ 5\ X_i\text{'s} > x\right) +$$

$$\vdots$$

$$+ P\left(10\ X_i\text{'s} \leq x,\ 0\ X_i\text{'s} > x\right)$$

$$\cancel{+ P\left(3\ X_i\text{'s} \leq x,\ 7\ X_i\text{'s} > x\right)}$$

$$\cancel{X_{(4)} > x}$$

$$\overset{\text{if iid}}{=} \sum_{j=4}^{10} \binom{10}{j} F_X(x)^j \left(1 - F_X(x)\right)^{10-j}$$

General case: k, n

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

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

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

Binomial thm

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}.$$

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

$$\overset{\downarrow}{=} \left(F_X(x) + 1 - F_X(x)\right)^n - \left(1 - F_X(x)\right)^n$$

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

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

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

$$\frac{d}{dx}[uv] = uv' + u'v \qquad u' = j\, f_X(x)\, F_X(x)^{j-1}$$

$$v' = -(n-j)\, f(x)\, (1 - F(x))^{n-j-1}$$