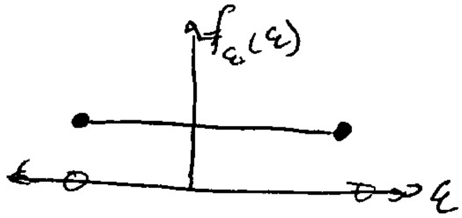
Lecture 12 MATH 621 10/19/2020

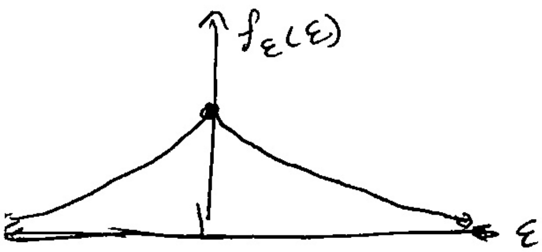
Laplace first published the Laplace dist. In 1774. calling it the first law of errors when you measure a quantity v, you measure it with error, epsilon so that:

$M = v + \text{epsilon}$.

What makes a good distribution for the error, epsilon? The expectation should be zero. and should be symmetric. How about ....



Not very good. It should have the property that the probab. of error. should decrease with its magnitude.



Another good property is that the density should be decreasing in magnitude error.

Laplace assumed for all positive error that

$f''_{\varepsilon}(\varepsilon) = f'_{\varepsilon}(\varepsilon)$

$f(\varepsilon) = c e^{-d\varepsilon} \Rightarrow \varepsilon \sim \text{Laplace}(0,1)$

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

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

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

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

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

famous waiting time / survival r.v model and its used in insurance companies.

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

The k param is cool. Here's a property of Weibull r.v under different values of k:

$c = 14,\ y = 3$

$k=1$, $P(Y\geq y+c \mid Y\geq c) = P(Y\geq y)$ e.g $P(Y\geq 17 \mid Y\geq 14) = P(Y\geq 3)$

This is equality is called "memorylessness"

$k>1$, $P(Y\geq y+c \mid Y\geq c) < P(Y\geq y)$ e.g old lifespan of humen, waiting for bus.

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

e.g startup company.

Order Statistics (p160)

Let $X_1, X_2, \ldots, X_n$ be a collecton of continuous r.v.

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.

$\vdots$

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

e.g $X_1 = 9,\ X_2 = 2,\ X_3 = 12,\ X_4 = 7.$

$X_{(1)} = 2,\ X_{(2)} = 7,\ X_{(3)} = 9,\ X_{(4)} = 17$

$r = 12 - 9 = 10.$

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

We want to find both the CDF and PDF of the k^{th} order stats.

The first thing we'll do is find the CDF and PDF of the maximum.

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, X_3, \ldots, X_n \leq x)$$

$$\overset{X_1,\ldots,X_n \text{ indep}}{=} P(X_1 \leq x) \cdots P(X_n \leq x) = \prod_{i=1}^{n} F_{X_i}(x) \overset{\text{iid}}{=} \prod_{i=1}^{n} F_X(x) = F_X(x)^n$$

$$f_{X_n}(x) = \frac{d}{dx}\left[F(x)^n\right] = nf(x)F(x)^{n-1}$$

The next thing we'll do is to find the CDF and PDF of the min.

$$F_{X_{(1)}}(x) = P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, X_2 > x, X_3 > x, .. X_n > x)$$

$$\overset{\text{indep}}{=} 1 - P(X_1 > x) \cdots P(X_n > x) = 1 - \prod_{i=1}^{n}(1 - F_X(x)) \overset{\text{iid}}{=} 1 - (1 - F(x))^n$$

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

Assume $n = 10$, $k = 4$ and derive the $k = 4^{th}$ order statistic CDF and PDF. Lets find proba that the first ~~four~~ numbers are less than x and the last six nums are greater than x.

$$= P(X_1 \le x, \ldots, X_4 \le x, X_5 > x, \ldots, X_{10} > x)$$

$$\overset{\text{indep}}{=} \prod_{i=1}^{4} F_{X_i}(x) \prod_{i=5}^{10} (1 - F_{X_i}(x)) \overset{\text{iid}}{=} F(x)^4 (1 - F(x))^6$$

Lets find the proba any 4 of the 10 are below x and the remaining are above x. Let S be a subset of size 4 of the index set $\{1, 2, \ldots, 10\}$

$$= \sum_{\text{all } S} P(X_{S_1} \le x, \ldots, X_{S_4} \le x, X_{S_1^c} > x, \ldots, X_{S_6^c} > x)$$

$$\overset{\text{indep}}{=} \sum_{\text{all } S} \prod_{i=1}^{4} F_{X_{S_i}}(x) \prod_{i=1}^{6} (-F_{X_{S_i^c}}(x) + 1) \overset{\text{iid}}{=} \sum_{\text{all } S} F(x)^4 (1 - F(x))^6$$

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

Now lets derive the CDF for the $k = 4^{th}$ order statistic.

$$F_{X_{(4)}}(x) = P(X_{(4)} \le x) = P(\text{a subset of } 4\ X_i\text{'s} \le x \text{ and the remaining 6 are} > x)$$
$$+ P(\text{a subset of } 5\ X_i \le x \text{ and the remaining 5 are} > x)$$
$$+ \ldots + P(\text{all } 10\ X_i\text{'s} \le x)$$

$$\overset{\text{iid}}{=} \binom{10}{4} F(x)^4 (1 - F(x))^6 + \binom{10}{5} F(x)^5 (1 - F(x))^5 + \ldots + \binom{10}{10} F(x)^{10} (1 - F(x))^{10-10}$$

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

For iid cont. r.v's $X_1, \dots X_n$ the CDF and PDF for the k^{th} order statistic is:

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

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

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

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