Math 621 Lec 9 9/28/17

$X \sim U(0,1)$, $Y = -\ln\left(\frac{1}{X} - 1\right) = g(X)$

Supp(Y) ? $= \mathbb{R}$

$X \in (0,1) \Rightarrow \frac{1}{X} \in (1,\infty) \Rightarrow \frac{1}{X} - 1 \in (0,\infty)$
$\Rightarrow \ln\left(\frac{1}{X}-1\right) \in \mathbb{R} \Rightarrow -\ln\left(\frac{1}{X}-1\right) \in \mathbb{R}$

$g^{-1}(y) = \frac{1}{1+e^{-x}}$

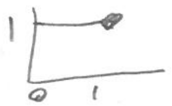
$-(1+e^{-x})^{-2}(-e^{-x})$

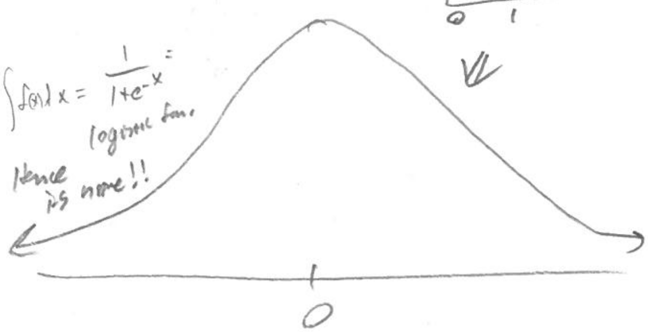
$= \frac{e^{-x}}{(1+e^{-x})^2}$

$$f_Y(y) = \underbrace{f_X(g^{-1}(y))}_{1} \left| \frac{d}{dy}\left[g^{-1}(y)\right] \right|$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \text{Logistic}(0,1)$$

stretch $(0,1) \to \mathbb{R}$





$F(x) = \int f(x)\,dx = \frac{1}{1+e^{-x}}$ = logistic func.

Hence its name!!

Due to heavier tails it's used for many systems e.g.

Elo system, US Chess Federation.

$f(x) = \frac{L}{1+e^{-K(x-x_0)}}$ "Logistic function"

L = max val.

K = steepness

x_0 = midpoint

$f(x) = \frac{1}{1+e^{-x}}$ Standard logistic function $L=1, x_0=0, K=1$

$= \frac{e^x}{1+e^x}$

exponential increase $\Rightarrow$ exponential decrease

$f : \mathbb{R} \to (0,1)$

$y = \frac{1}{1+e^{-x}}$

$\Rightarrow \frac{1}{y} = 1+e^{-x}$

$\Rightarrow \frac{1}{y} - 1 = e^{-x}$

$\Rightarrow \underbrace{-\ln\left(\frac{1}{y}-1\right)}_{f^{-1}(y)} = x$

$f^{-1} : [0,1] \to \mathbb{R}$

Just like the normal distr. but has heavier tails.

It is the basis for logistic regression and is useful in deep learning (neural networks).

if $X \sim \text{Exp}(1) \Rightarrow Y = -\ln\left(\frac{e^{-X}}{1-e^{-X}}\right) \sim \text{Logistic}(0,1)$ (HW)

Assume $k \in (0, \infty)$

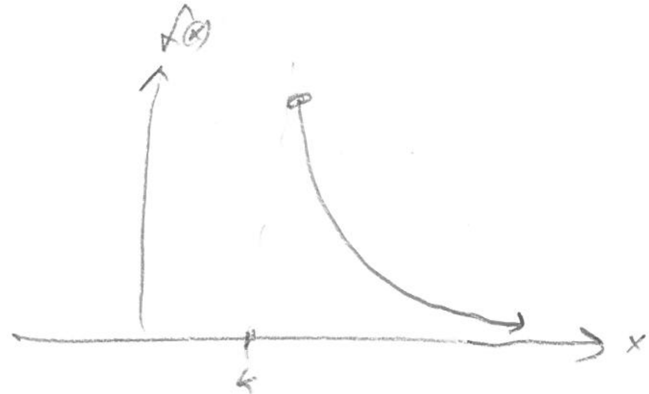
$X \sim Exp(\lambda)$ $\quad Y = ke^X$ $\quad Supp(Y) = (1, \infty)$ if $k=1$ $\quad Supp(Y) = (k, \infty)$

$\Rightarrow \frac{y}{k} = e^x \Rightarrow x = \ln\left(\frac{y}{k}\right) = g^{-1}(y) \quad \left|\frac{d}{dy}\left[g^{-1}(y)\right]\right| = y^{-1}$

$f_Y(y) = f_X\left(\ln\left(\frac{y}{k}\right)\right) y^{-1} = \lambda e^{-\lambda \ln\left(\frac{y}{k}\right)} y^{-1} = \frac{\lambda}{y} e^{\ln\left(\frac{k}{y}\right)^\lambda} = \frac{\lambda}{y}\left(\frac{k}{y}\right)^\lambda$

$= \frac{\lambda k^\lambda}{y^{\lambda+1}} = ParetoI(k, \lambda)$

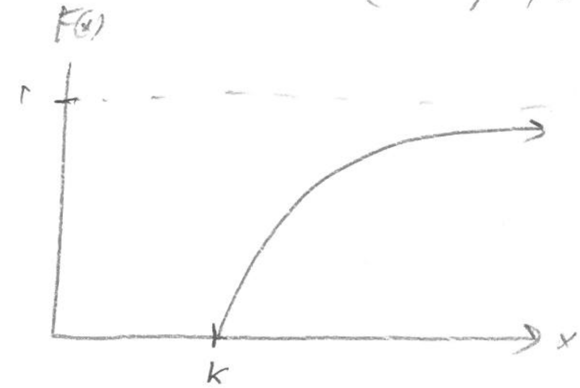
Type I



$F_Y(y) = \int_k^y \frac{\lambda k^\lambda}{x^{\lambda+1}} dx = \lambda k^\lambda \left.\frac{x^{-\lambda-1+1}}{-\lambda-1+1}\right]_k^y = -k^\lambda\left[x^{-\lambda}\right]_k^y = -k^\lambda\left(y^{-\lambda} - k^{-\lambda}\right)$

$= k^\lambda\left(k^{-\lambda} - y^{-\lambda}\right) = 1 - \left(\frac{k}{y}\right)^\lambda$

Used to model
- population spread towns/cities
- HD disk failure
- sizes of sand powders
- file size distr. in Internet routers
- AND... where it gets its name...



Named Pareto due to the "Pareto Principle". 1896 noticed that 80% of the land in Italy was owned by 20% of the pop.

We will now demonstrate this...

$F_Y^{-1}(p)$

What is this?

Quantile Function!

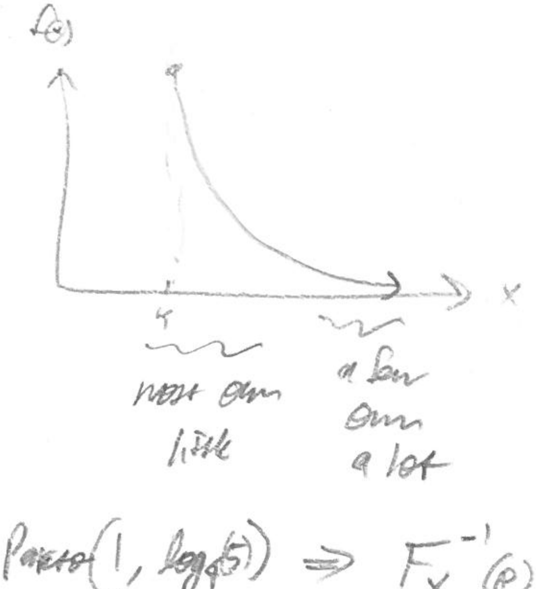
$\text{Quantile}[X, p] = \inf_x \{F(x) \geq p\}$ at which value of x is $p = P(X \leq x) = F(x)$

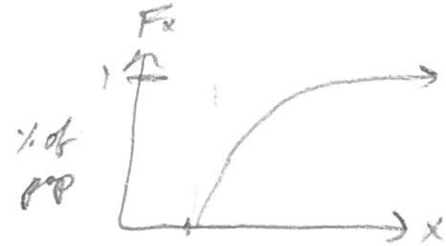
What is the 99%ile or .99 quantile of the SAT? The score that is better than 99% of the pop.

For continuous functions $F_X^{-1}(p) = \text{Quantile}[X, p]$ For discrete functions $\Rightarrow$ not so easy!

For Pareto,

$$p = F_Y(y) = 1 - \left(\frac{k}{y}\right)^{\lambda} \Rightarrow 1-p = \left(\frac{k}{y}\right)^{\lambda} \Rightarrow (1-p)^{\frac{1}{\lambda}} = \frac{k}{y} \Rightarrow y = k(1-p)^{-\frac{1}{\lambda}} = F_Y^{-1}(p)$$





How much land owned / how much wealth

most own little

a few own a lot

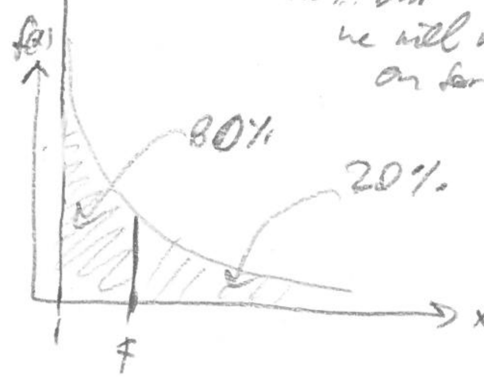
$X \sim \text{Pareto}(1, \log_4 5) \Rightarrow F_X^{-1}(p) = (1-p)^{-0.861}$

How much do the bottom 80% own?

$F_X^{-1}(0.8) = (1-0.8)^{-.861} = 4$

$1 - F_X(4) = 1 - \left(\frac{1}{4}\right)^{1.161} = 0.8$

There is a lot more to do... but we will move on for now...



$X, Y \overset{iid}{\sim} \text{Exp}(1) = e^{-x}$ difference in survival times. Imagine two lightbulbs

$D := X - Y$ $\quad\quad$ $\text{Supp}[D] = \mathbb{R}$

let $Z = -Y$ $\quad f_Z(z) = f_Y(-y) = e^{y}$

$D = X + Z$

$$f_D(d) = \int_{\text{Supp}(X)} f_X(x)\, f_Z(d-x)\, dx = \int_0^\infty \overbrace{\underbrace{e^{-x}\, \overbrace{e^{d-x}}^{e^d e^{-x}}}_{}}^{e^d e^{-2x}}\, \underbrace{\mathbb{1}_{d-x \in (-\infty, 0)}}\, dx$$

$x - d \in (0, \infty)$

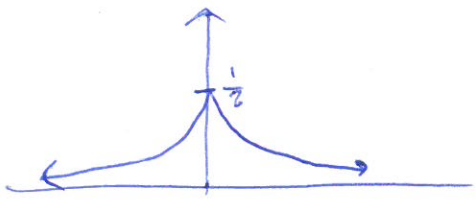
$x \in (d, \infty)$

$x \geq d$

$$= e^d \int_{\max\{0,d\}}^{\infty} e^{-2x}\, dx = e^d \left[-\frac{1}{2} e^{-2x}\right]_{\max\{0,d\}}^{\infty} = \frac{1}{2} e^d e^{-2\max\{0,d\}}$$

$$= \frac{1}{2}\begin{cases} e^d & \text{if } d < 0 \\ e^{-d} & \text{if } d \geq 0 \end{cases} = \frac{1}{2} e^{-|d|} = \text{Laplace}(0,1)$$

AKA double-exponential - why?



$P(X \geq Y) = P(X - Y \geq 0) = \frac{1}{2}$ due to symmetry around $d = 0$.

1774 Laplace published this and called it "the first law of errors".

Laplace(0,1) is an "error distr." What is that?

Imagine you are trying to measure V, a fixed value. But your measurement system and/or your model for V is not perfect. So you measure $Y \neq V$ but $Y \approx V$. What is the distr of Y?

You can say $$Y = V + \underbrace{\varepsilon}_{\text{"error"}}$$

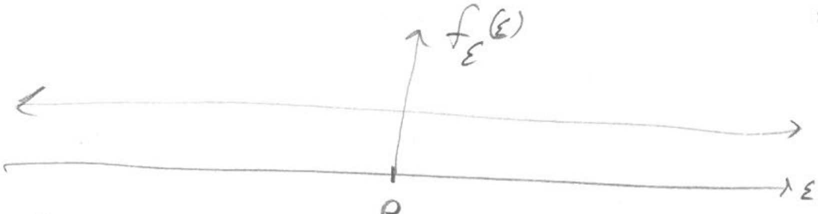
In other disciplines e.g. signal detection it is called noise. What is distr of ε?

It seems reasonable that $E[\varepsilon] = 0$ so that $E(Y) = V$ also,

$$\text{Med}(\varepsilon) = 0 \quad \text{so that} \quad \text{Med}(Y) = V.$$

$\Rightarrow f_\varepsilon(\varepsilon) = f_\varepsilon(-\varepsilon)$

Underestimates and overestimates of same magnitude are equiprobable



Could this be a good idea?

What is a reasonable assumption? Small errors are more probable than large errors

$$\Rightarrow f'_\varepsilon(\varepsilon) < 0 \quad \text{if } \varepsilon > 0$$

Now he then assumes the function and its derivative change at same rate

$$f''(\varepsilon) = f'(\varepsilon) \Rightarrow f(\varepsilon) = c e^{-mx}$$

duplicating this when $\varepsilon < 0$ and normalizing $\Rightarrow$ Laplace distr!

log error

Laplace gives linear relationship of log error

In 1778 he published the "second law of errors"

where

lin rel. of log error squared

$\Rightarrow f(\varepsilon) \propto e^{-\varepsilon^2}$

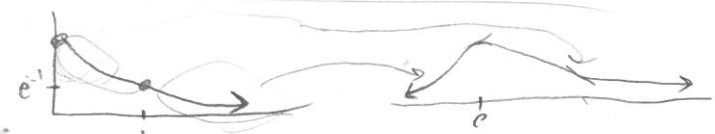
which is the normal distr.

Laplace invented it before Gauss!

Gauss born in 1777!

Is logistic an error distr? YES ... sometimes used too

$X \sim Exp(1) = e^{-x}$



$Y = -\ln(X) = \ln\left(\frac{1}{X}\right)$ $Supp[Y] = \mathbb{R}$ the log inverse waiting time or log inverse survival time

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

$f_Y(y) = f_X(g^{-1}y) \left|\frac{d}{dy}(g^{-1}(y))\right|$

low waiting times $\Rightarrow$ large Gumbel vals

high waiting times $\Rightarrow$ low Gumbel vals

$= e^{-e^{-y}} e^{-y} = e^{-(y+e^{-y})} = Gumbel(0,1)$ "Standard Gumbel"

$X \sim Gumbel(0,1)$, $Y = \mu + \beta X$ "Gumbel"

$f_Y(y) = \frac{1}{|\beta|} f_X\left(\frac{y-\mu}{\beta}\right) = \frac{1}{|\beta|} e^{-\left(\frac{y-\mu}{\beta} + e^{-\left(\frac{y-\mu}{\beta}\right)}\right)} \sim Gumbel(\mu, \beta)$

$\mu \in \mathbb{R}, \beta > 0$ if $\beta < 0$ becomes $+$ $\Rightarrow$ diverges

so delete abs value