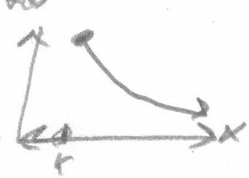
MATH 340/640 Lec 23



Recall $X \sim \text{ParetoI}(k, \lambda) := \frac{\lambda k^\lambda}{x^{\lambda+1}} \mathbb{1}_{x \in (k, \infty)}$, $k, \lambda > 0$

$F(x) = 1 - \left(\frac{k}{x}\right)^\lambda \Rightarrow \underbrace{1 - F(x)}_{q} = \left(\frac{k}{x}\right)^\lambda \Rightarrow (1-q)^{\frac{1}{\lambda}} = \frac{k}{x} \Rightarrow \frac{x}{k} = (1-q)^{-\frac{1}{\lambda}}$

$F^{-1}(q) = k(1-q)^{-\frac{1}{\lambda}}$

Note $Y = X - k \sim \text{Lomax}(\lambda, k) = \frac{\lambda}{k}\left(1 + \frac{x}{k}\right)^{-(\lambda+1)} \mathbb{1}_{x \in (0,\infty)}$ (1954)

let $k = 1 \Rightarrow f(x) = \frac{\lambda}{x^{\lambda+1}} \mathbb{1}_{x \in (1,\infty)}$, $F^{-1}(q) = (1-q)^{-\frac{1}{\lambda}}$

$E[X] = \frac{\lambda}{\lambda - 1}$

In 1896, Vilfredo Pareto looked at land ownership by person in Italy.

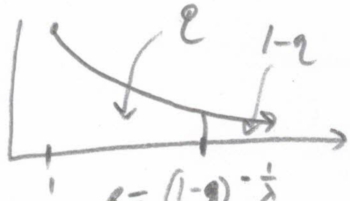
His plot looked like this:

let $x_{min} = 1$.

He found that the bottom 80% of people own 20% of the land

He fit his curve which became the ParetoI distr. Which value of λ fit?

Let $L(a)$ be the amount of land owned by people who own less than land amount a



$a = (1-q)^{-\frac{1}{\lambda}}$

$L(a) = \int_1^a x f(x)\,dx = \int_1^a x \frac{\lambda}{x^{\lambda+1}}\,dx = \lambda \int_1^a x^{-\lambda}\,dx =$

$= \lambda \frac{1}{-\lambda+1}\left[x^{-\lambda+1}\right]_1^a = \frac{\lambda}{\lambda-1}\left(1 - a^{-\lambda+1}\right)$

$\text{Tot} := L(\infty) = \int_1^\infty x f(x)\,dx = \frac{\lambda}{\lambda - 1}$

$\pi(a) = \frac{L(a)}{L(\infty)} = \frac{\cancel{\frac{\lambda}{\lambda-1}}\left(1 - a^{-\lambda+1}\right)}{\cancel{\frac{\lambda}{\lambda-1}}}$

prop of land owned by bottom a prop. of people

$a = (1 - .8)^{-\frac{1}{\lambda}} = .2^{-\frac{1}{\lambda}}$

$0.2 = \pi\left(.2^{-\frac{1}{\lambda}}\right) = 1 - .2^{\frac{\lambda-1}{\lambda}} = 1 - .2^{1 + \frac{1}{\lambda}}$

$\Rightarrow .8 = .2^{1+\frac{1}{\lambda}} \Rightarrow \ln(.8) = \left(1 - \frac{1}{\lambda}\right)\ln(.2)$

$1 - \frac{\ln(.8)}{\ln(.2)} = \frac{1}{\lambda} \Rightarrow \lambda = \frac{\ln(.2)}{\ln(.2) - \ln(.8)} = 1.161 = \frac{\ln\left(\frac{1}{5}\right)}{\ln\left(\frac{1}{5}\right) - \ln\left(\frac{4}{5}\right)} = \frac{\ln\left(\frac{1}{5}\right)}{\ln\left(\frac{1}{4}\right)} = \frac{\ln(5)}{\ln(4)}$

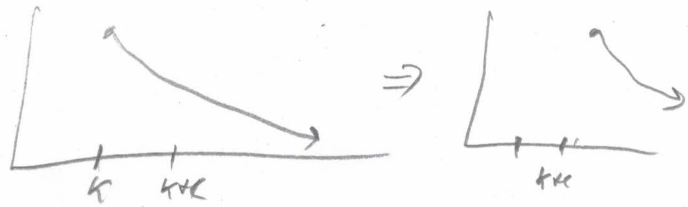
$\lambda = 1.161 = \ldots$

$\Rightarrow$ the 80-20 rule is the Pareto I distr where k doesn't matter.

$\Rightarrow$ the 90-20 rule is the " " " $\lambda = \frac{\ln(10)}{\ln(9)} = 1.048$ Pareto I distr.

---

let $X \sim \text{Pareto}(k, \lambda)$



what is distr of $Y \sim X \mathbb{1}_{X > k+c}$

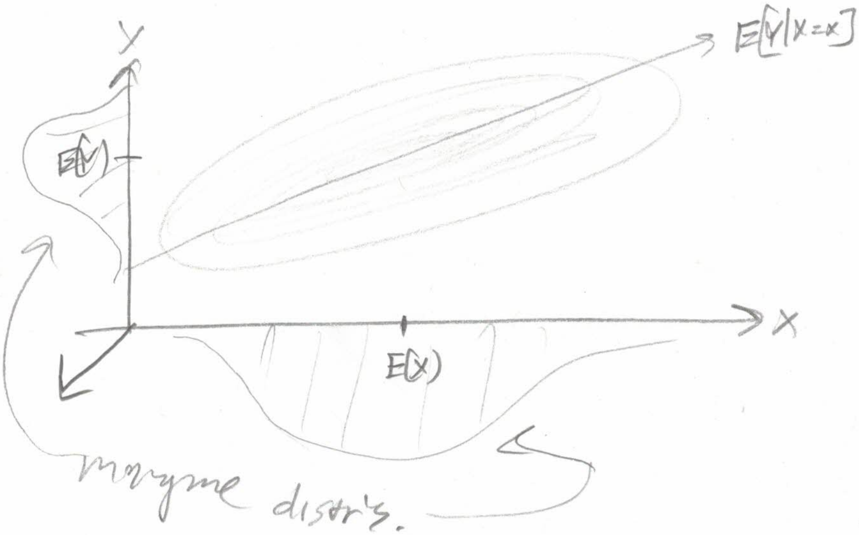
$$f_Y(y) = \frac{f_X(y) \mathbb{1}_{y > k+c}}{\int_{k+c}^{\infty} f_X(u)\,du} = \frac{\frac{\cancel{\lambda k^\lambda}}{y^{\lambda+1}} \mathbb{1}_{y>k} \mathbb{1}_{y>k+c}}{\int_{k+c}^{\infty} \frac{\cancel{\lambda k^\lambda}}{u^{\lambda+1}}\,du} = \frac{\frac{1}{y^{\lambda+1}} \mathbb{1}_{y>k+c}}{\frac{1}{-\lambda}\left[u^{-\lambda}\right]_{k+c}^{\infty}}$$

$$= \frac{\frac{1}{y^{\lambda+1}} \mathbb{1}_{y>k+c}}{\frac{1}{\lambda}(k+c)^{-\lambda}} = \frac{\lambda (k+c)^\lambda}{y^{\lambda+1}} \mathbb{1}_{y>k+c} = \text{Pareto}(k+c, \lambda)$$

i.e. the Pattern repeats itself! Like a fractal!

If the richest 10% own 90%, then the richest 90% of the 10% own 90% of the richest 10%!!!

Imagine two r.v.'s creating a joint density $f_{XY}(x,y)$



We can derive a nice identity:

$$E(Y) = \int_{Supp(Y)} y\, f_Y(y)\, dy = \int_{Supp(Y)} y \int_{Supp(X)} f_{XY}(x,y)\, dx\, dy$$

$$= \int_{Supp(Y)} \int_{Supp(X)} y\, f_{Y|X}(x,y)\, f_X(x)\, dx\, dy = \int_{Supp(X)} \int_{Supp(Y)} y\, f_{Y|X}(x,y)\, f_X(x)\, dy\, dx$$

$$= \int_{Supp(X)} f(x) \int_{Supp(Y)} y\, f_{Y|X}(x,y)\, dy\, dx = \int_{Supp(X)} f(x)\, \underbrace{E[Y|X=x]}_{g(x)}\, dx = E_X[\overbrace{E_Y[Y|X=x]}^{\text{Shorthand } E(Y|X)}]$$

Law of Iterated Expectation

Ex: $X \sim U(0,1)$, $Y|X=x \sim U(0,x)$

$E(Y) = E_X[E(Y|X)] = E_X\left[\frac{X}{2}\right] = \frac{1}{2} E[X] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$X \sim Beta(\alpha,\beta)$, $Y|X=x \sim Binom(n,x)$

$E(Y) = E_X[E(Y|X)] = E_X[nX] = n\,E(X) = n\frac{\alpha}{\alpha+\beta}$

Using the Law of Iterated Expectations

$$\mathrm{Var}(Y) = E[Y^2] - E(Y)^2$$

$$= E_X\left[E_Y[Y^2|X]\right] - E_X\left(E_Y(Y|X)\right)^2$$

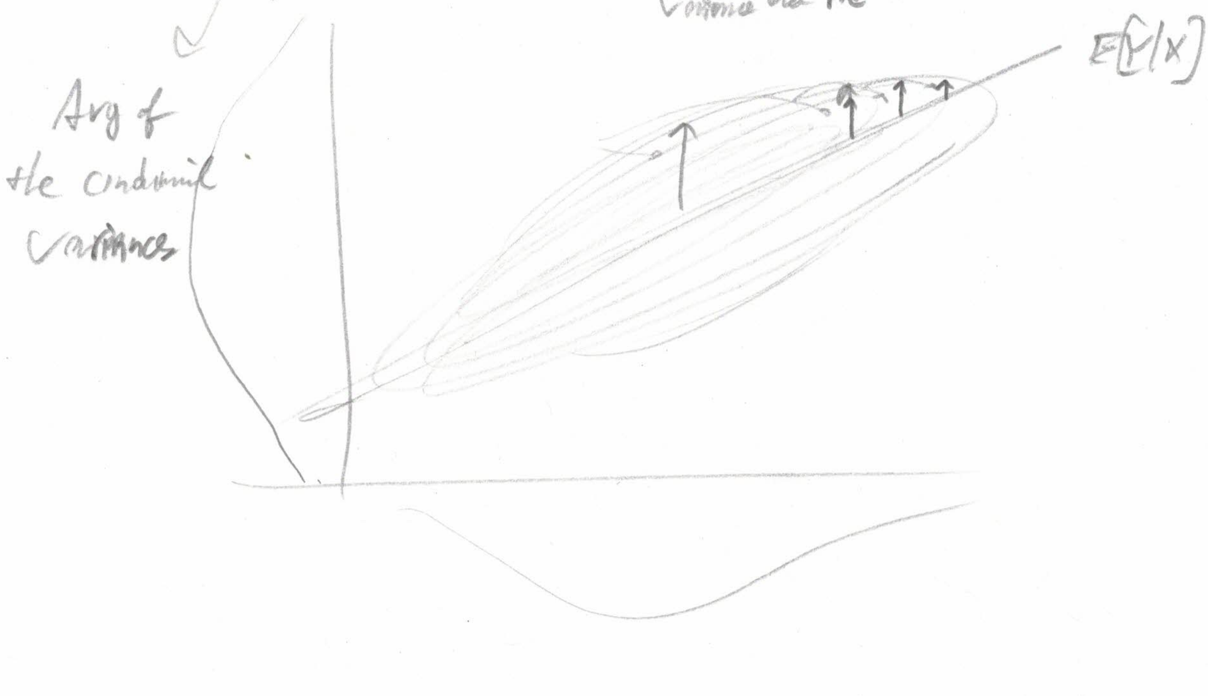
$$= E_X\left[\mathrm{Var}_Y[Y|X] + E_Y(Y|X)^2\right] - E_X\left(E_Y(Y|X)\right)^2$$

$$= E_X\left[\mathrm{Var}_Y(Y|X)\right] + \underbrace{E_X\left[\underbrace{E_Y(Y|X)^2}_{Q^2}\right] - E_X\left[\underbrace{E_Y(Y|X)}_{Q}\right]^2}_{\mathrm{Var}(Q)}$$

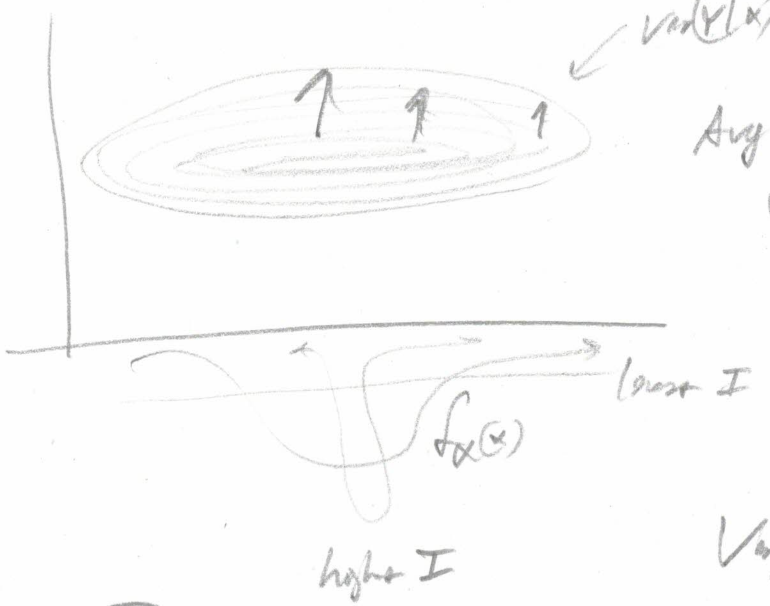
$$\Rightarrow \mathrm{Var}(Y) = \overbrace{E_X\left(\mathrm{Var}_Y(Y|X)\right)}^{I} + \overbrace{\mathrm{Var}_X\left[E_Y(Y|X)\right]}^{II}$$

Law of Total Variance

variance over the cond. mean



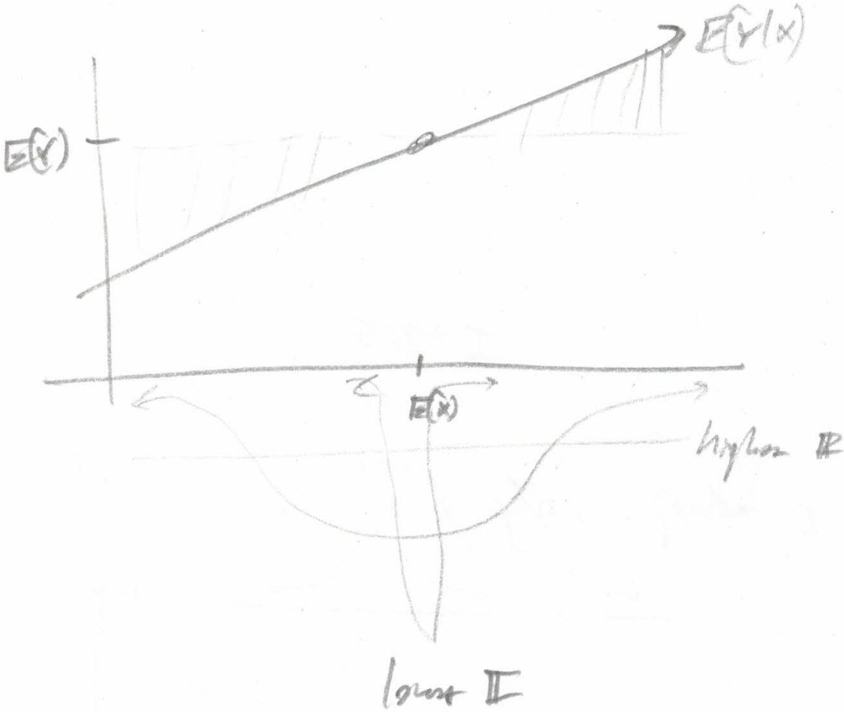
I



Avg of all these variances weighted by $f_X(x)$

$$Var_X(x) = E_X\left[\left(E_Y(Y|x) - E[Y]\right)^2\right]$$

II



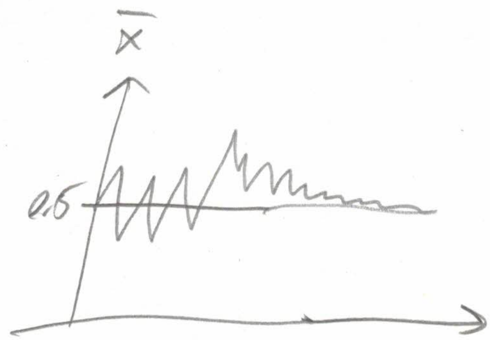
Topics Skipped: ① Poisson process: link between waiting times in Erlang(k,λ) in x∈(0,1) and # events of Poisson(λ).

② Skellam Distr. $X_1 \sim Poisson(\lambda_1)$ indep of $X_2 \sim Poisson(\lambda_2) \implies X_1 - X_2 \sim Skellam(\lambda_1, \lambda_2)$ uses Bessel Functions. This distr. models point spreads in sports

Frequentist

Definitions of prob.

Let $\theta := P(\text{Heads})$



$$\theta = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n}$$

$\Updownarrow$

$\bar{X} \xrightarrow{P} \theta$ if $X_1, X_2, \ldots \overset{iid}{\sim}$

This is "objective" in the sense that only the data speaks!

However, it is absurd because the true value of θ is never actually known, we only believe it exists and that it's fixed and immutable.

Propensity definition:

$\theta = P(\text{Heads}) = f(\text{of the physical nature of the coin})$

For instance the half life of uranium 235 is 700m years

This one atom of uranium has Bern$(\frac{1}{2})$ that it decays in the first 700m years. This is derivable from the quantum physics laws.

Also objective!

What about $P(\text{Dolphins win the superbowl})$. No long run freq can apply. No magic physical function apply

$\Rightarrow$ Must be subjective (based on gambling odds usually).

Back to sampling...

$$\begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix} = \vec{X} \sim \text{Multinom}(n, \vec{p}). \quad \text{How to sample?}$$

$\vec{X} = \vec{X}_1 + \dots + \vec{X}_n$ where $\vec{X}_1, \dots, \vec{X}_n \overset{iid}{\sim} \text{Multinom}(1, \vec{p})$

How to sample $\vec{X}_i$? Same question!!

Consider:

$$P(X_1 \dots X_K) = P(X_2, \dots X_K | X_1)\, P(X_1)$$

We know that $P(X_1) = \text{Bin}(n, p_1)$, $P(X_2 \dots X_K | X_1) = \text{Multinom}\left(n - X_1, \frac{\begin{bmatrix} p_2 \\ \vdots \\ p_K \end{bmatrix}}{1-p_1}\right)$

then $P(X_2, \dots X_K | X_1) = P(X_3, \dots X_K | X_1, X_2)\, P(X_2 | X_1)$

$P(X_2 | X_1) = \text{Binom}\left(n - X_1, \frac{p_2}{1-p_1}\right)$, $P(X_3 \dots X_K | X_1, X_2) = \text{Multinom}\left(n - X_1 - X_2, \frac{\begin{bmatrix} p_3 \\ \vdots \\ p_K \end{bmatrix}}{1-p_1-p_2}\right)$

then $P(X_3, \dots, X_K | X_1, X_2) = P(X_4, \dots, X_K | X_1, X_2, X_3)\, P(X_3 | X_1, X_2)$

Step through dimensions $P(X_3 | X_1, X_2) = \text{Bin}\left(n - X_1 - X_2, \frac{p_3}{1-p_1-p_2}\right)$ etc...

Algorithm:

Step 1: Sample x_1 from $\text{Bin}(n, p_1)$. If $x_1 = n$, stop, all remaining x_j's $= 0$

Step 2 Sample x_2 from $\text{Bin}\left(n - x_1, \frac{p_2}{1-p_1}\right)$ If $x_1 + x_2 = n$ stop, ...

Step 3 " " x_3 " " $\text{Bin}\left(n - x_1 - x_2, \frac{p_3}{1-p_1-p_2}\right)$ " " " "

$\vdots$

Step K $x_K = n - x_1 - x_2 - \dots - x_{K-1}$ degenerate as $\text{Bin}\left(n - x_1 - \dots - x_{K-1}, \underbrace{\frac{p_K}{1-p_1-\dots-p_{K-1}}}_{1}\right)$

This step through is valid for sampling any $\vec{X}$!

Requirement: need all conditional distributions $P(X_j | X_1, \dots, X_{j-1})$ $\forall j \in \{1, \dots, K\}$

If you don't $\Rightarrow$ Gibbs Sampling, Metropolis-Hastings, Hamiltonian MCMC (3A3/6A3)

Markov Chain Monte Carlo (MCMC)