November 1, 2016

Last Week:

- $\mu := E[X] := \sum_{x \in supp[X]} x\, p(x)$
- $\bar{X} \longrightarrow \mu$ (Law of large numbers)
- $X \sim Bern(p)$
  $E[X] := p$
- $X \sim Binom(n,p)$
  $E[X] := np$
- $X \sim Hyper(n,K,N)$
  $E[X] := n\frac{K}{N}$
- $X \sim Geom(p)$
  $E[X] := ?$
- $X \sim NegBin(r,p)$
  $E[X] := ?$

Recall:

$X \sim Geom(p) := (1-p)^{x-1} p$

$X \sim Geom(0.2) = (0.8)^{x-1}(0.2)$

x	$p(x)$	$F(x)$	x	$p(x)$	$F(x)$
1	0.200	0.200	15	0.009	0.965
2	0.160	0.360	16	0.007	0.972
3	0.128	0.488	17	0.006	0.978
4	0.102	0.590	18	0.005	0.983
5	0.082	0.692	19	0.004	0.987
6	0.066	0.738	20	0.003	0.990
7	0.052	0.790	21	0.002	0.992
8	0.042	0.832	22	0.001	0.993
9	0.034	0.866	23	0.001	0.994
10	0.027	0.893	24	0.001	0.995
11	0.021	0.914	25	0.001	0.996
12	0.017	0.931	26	0.001	0.997
13	0.014	0.945	27	0.001	0.998
14	0.011	0.950	28	0.000	0.999

← Never hits 1 in a finite value. It'll only be ≈ 1.

Approximate / Effective Support

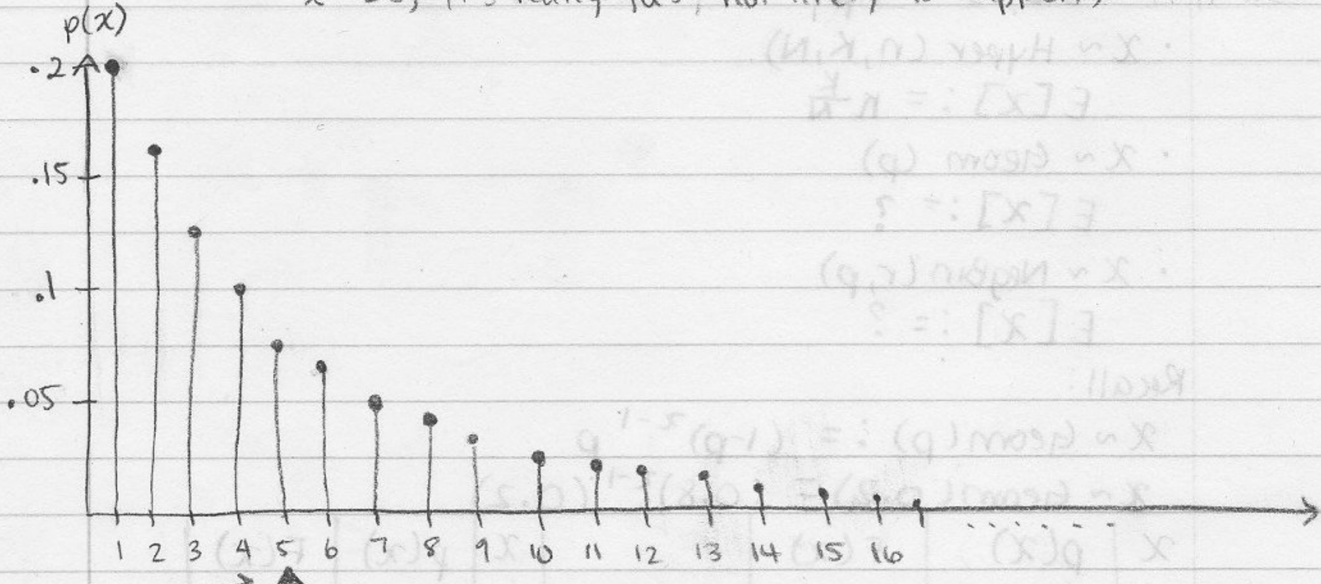
$\{x : p(x) > .001\} \subset Supp[X] := \mathbb{N}$

or

smallest subset of $Supp[X]$, A, s.t.

$\sum_{x \in A} p(x) = 0.999$

(Basically gives you the most likely range of support. The true support goes on to ∞ but if you're at $x = 28$, it's really just not likely to happen)



Where is the balance point / expected value / mean?

$$X \sim Geom(p) \Rightarrow \mu := \sum_{x \in Supp[X]} x\, p(x) = \sum_{x=1}^{\infty} x (1-p)^{x-1} p$$

Geometric Series $= \frac{1}{p}$

$$= \sum_{y=0}^{\infty} (y+1)(1-p)^{y} p = p\left(\sum_{y=0}^{\infty} y (1-p)^{y} + \sum_{y=0}^{\infty} (1-p)^{y}\right) = \sum_{y=0}^{\infty} y (1-p)^{y} p + 1$$

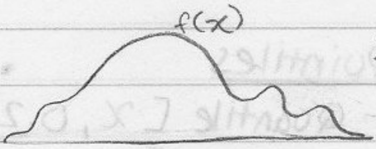
Reindex: Let $y = x - 1 \Rightarrow x = y + 1$

$x = 1 \ldots \infty \Rightarrow y = 0 \ldots \infty$

$$= \underbrace{(1-p)\sum_{y=1}^{\infty} y (1-p)^{y-1} p}_{\mu} + 1 \Rightarrow \mu = \mu - p\mu + 1 \Rightarrow \mu = \frac{1}{p}$$

$X \sim Geom(p) :\Rightarrow \mu = \frac{1}{p}$

$$X \sim Geom(0.2) \Rightarrow \mu = \frac{1}{0.2} = 5$$

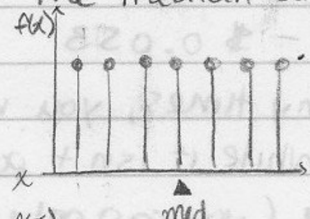


function of a function

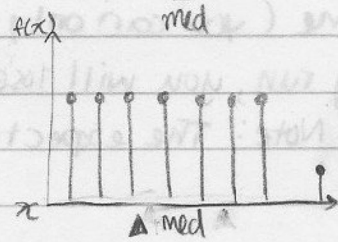
$G[f] = \int_{\mathbb{R}} f(x)\,dx = 17$

↳ you go from infinite info to 17. Likewise, $E[X]$ is also "functional". Sums up r.v. to one number.

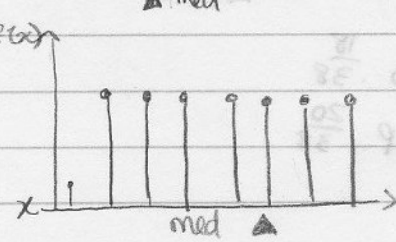
- $\text{mode}[X] := \text{argmax}\{p(x)\}$, $x \in \text{Supp}[X]$
  - "Look at all the probabilities and find the biggest one"
  - $\text{mode}[X] = 1$
  - The most likely value you're going to get.
- $\min[X] := \min\{\text{supp}[X]\} = 1$
- $\max[X] := \max\{\text{supp}[X]\}$ = No max.
- $\text{Range}[X] := \max[X] - \min[X]$: No range
- $\text{Quantile}[x, p] := \text{argmin}\{F(x) \geq p\}$
  - "percentile if p is a percent."
  - "The 95 percentile on an exam was 17." Means that at most, 5% of people scored higher.
  - The 10 percentile of this r.v. is 6.
- $\text{Median}[X] := \text{Quantile}[X, 0.5]$
  - The median of this r.v. is 4.
  - The median and the mean are not the same. Why?



Here, the median value and the mean value are the same.



Here, the median will only move slightly, while the expectation will have to adjust itself to balance the "seesaw".



If...	Type of Distribution
$E[X] = \text{Median}[X]$	Symmetric
$E[X] > \text{Median}[X]$	Skew Right
$E[X] < \text{Median}[X]$	Skew Left
One mode	Unimodal
$E[X] = \text{Median}[X] = \text{Mode}[X]$	Symmetric Unimodal

Tertiles
- Quantile $[x, 0.33]$
- Quantile $[x, 0.66]$

Quartiles
- Quantile $[x, 0.25]$
- Quantile $[x, 0.5]$
- Quantile $[x, 0.75]$

Quintiles
- Quantile $[x, 0.2]$
- Quanite $[x, 0.4]$
- Quantile $[x, 0.6]$
- Quantile $[x, 0.8]$

Deciles
- Quantile $[x, 0.1]$
- Quantile $[x, 0.2]$
- ⋮
- Quantile $[x, 0.9]$

InterQuantile Range (IQR)

$IQR[x] := Q[x, 0.75] - Q[x, 0.25]$

Represents the middle 50 percentile.

> Roulette in America
> Bet on black. Payout is 1 : 1

$$x \sim \begin{cases} \$1 & \text{wp } \frac{18}{38} \\ -\$1 & \text{wp } \frac{20}{38} \end{cases}$$

$$E[x] = \left(1 \cdot \frac{18}{38}\right) - \left(1 \cdot \frac{20}{38}\right) = -\$0.053$$

This means that if you play many times, you will lose 5 cents on average. While it isn't possible to just lose 5¢ in one game (you can only win \$1 or lose \$1), in the long run, you will lose 5¢ per game on average. Note: The expectation is not in the support.

$\bar{X} \longrightarrow \mu$ LLN

$$x_1, \ldots, x_n \overset{iid}{\sim} \begin{cases} \$1 & \text{wp } \frac{18}{38} \\ -\$1 & \text{wp } \frac{20}{38} \end{cases}$$

$$\lim_{n \to \infty} T = -\infty$$

The more you play, the more you lose. In the long run, you will lose everything

> Bet on "lucky" #7. Payout is 35 : 1

$$X \sim \begin{cases} \$35 & wp \ \frac{1}{38} \\ -\$1 & wp \ \frac{37}{38} \end{cases}$$

$$E[X] = (35 \cdot \tfrac{1}{38}) - (1 \cdot \tfrac{37}{38}) = -\$0.053$$

When you consider a whole bunch of plays, you lose about 5 cents per play. It's only an approximation because your $\bar{x}$ will never settle for an exact number.

* Expectation is a long run property. It means very little if you only have one r.v. It needs to be done a bunch of times.

> Bet on dozen. Payout is 2 : 1

$$X \sim \begin{cases} \$2 & wp \ \frac{12}{38} \\ -\$1 & wp \ \frac{26}{38} \end{cases}$$

$$E[X] = (2 \cdot \tfrac{12}{38}) - (1 \cdot \tfrac{26}{38}) = -\$0.053$$

Getting \$0.053 is a coincidence, but it is intellegently designed by casinos. If you play any game, or if a bunch of people play a game, no matter what, the casino will get money the more you play. However, they make sure you can also win so you can get hooked.

> Roulette in Europe

> Bet on black. Payout is 1:1

$$X \sim \begin{cases} \$1 & wp \ \frac{18}{37} \\ -\$1 & wp \ \frac{19}{37} \end{cases} \qquad E[X] = -\$0.027$$

You lose your money slower. It feels more fair.

> If X models a payout of a game, the definition of a "fair game" is $E[X] = 0$.