Lecture #7
6/8/11

Admin
- HW coming out ~~today~~, due Thurs
  long!!! would be here, Friday fully due
- Trying to midterm review session Sunday night

Plan - One thing from the review could probably
- Bernoulli ⇒ Binomial ⇒ Poisson (ch 11)
- Ch 1, ch 2, ch 3

Marble Q: What's the probability we pick 4 marbles and get 4 B's?
Remember 4B, 4G, 2R, 2O

~~Use~~ In choose notation: $\dfrac{\binom{4}{4}}{\binom{12}{4}} = \dfrac{1}{495}$

How else can you do this?

$$\frac{4}{12} \cdot \frac{3}{11} \cdot \frac{2}{10} \cdot \frac{1}{9} = \frac{1}{495}$$

What is this?

Imagine the event

P(pick 1st blue AND pick 2nd B AND pick 3rd B AND pick 4th B)

$= P(A, B, C, D) = P(A) \cdot P(B|A) \cdot P(C|A,B) \cdot P(D|A,B,C)$

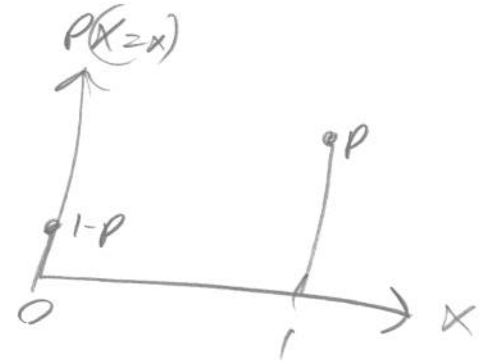
Without knowing it, this is 3 Bayes rules... think about it...

Yesterday we did Bernoulli r.v.'s

$X \sim \text{Bernoulli}(p)$

$E[X] = p$

$\text{Var}[X] = p(1-p)$



Yesterday, we did $S_2 = X_1 + X_2$ of spinners. Let's do $X_1, \ldots, X_{10} \overset{iid}{\sim} \text{Bernoulli}(p)$

and $S_{10} = X_1 + \ldots + X_{10}$

It is clear S_{10} is complicated...

Let $p = \frac{1}{2}$ and let's reason through this...

If $p = \frac{1}{2}$ what's our favorite Ω?

~~take~~ What's the prob I get 4 heads i.e. $S_{10} = 4$

Review: $P(H_1 H_2 H_3 H_4 T_1 T_2 T_3 T_4 T_5 T_6)$

$$= P(H_1) P(H_2) P(H_3) P(H_4) P(T_1) \cdot \ldots \cdot P(T_6)$$
$$= \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6$$

But we can do this in any order...

$$\frac{10!}{4!\ 6!}$$

all sequences of 10

don't care about order of H's

don't care about order of tails

$$\Rightarrow P(S_{10} = 4) = \underbrace{\frac{10!}{4!\,6!}}_{\binom{10}{4}} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6$$

In general $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$

n trials

prob of success

$$\Rightarrow S_n = X_1 + \ldots + X_n \sim \text{Binomial}(n, p) \triangleq \binom{n}{k} p^k (1-p)^{n-k}$$

P(k successes in n trials)

binomial coefficient

PMF of a binomial r.v.

$$\text{Support}(S_n) = \{0, 1, \ldots, n\}$$

Dentist yesterday. 32 teeth. Let's say prob of getting a cavity is $\frac{1}{100}$. Let's build a PMF:

Can we model this as a binom? $S_{32} = X_1 + \dots + X_{32}$ s.t. $X_1, \dots, X_{32} \stackrel{iid}{\sim} \text{Bernoulli}\left(\frac{1}{100}\right)$ What is BAD??

$$P(S_{32} = 0) = \binom{32}{0}\left(\frac{1}{100}\right)^0\left(\frac{99}{100}\right)^{32} = 0.7247$$

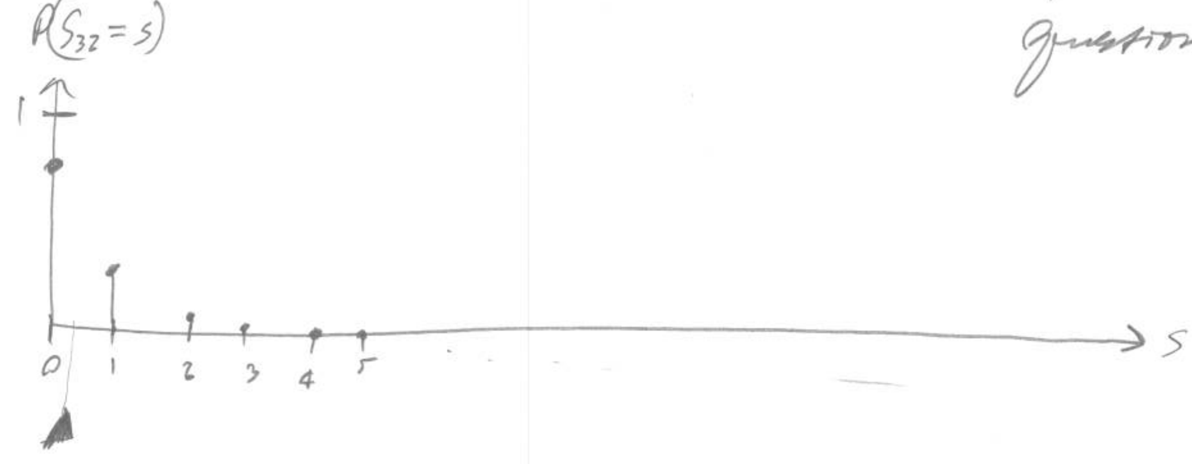
$$P(S_{32} = 1) = \binom{32}{1}\left(\frac{1}{100}\right)^1\left(\frac{99}{100}\right)^{31} = 0.2343$$

$$P(S_{32} = 2) = \binom{32}{2}\left(\frac{1}{100}\right)^2\left(\frac{99}{100}\right)^{30} = 0.0367$$

$$P(S_{32} = 3) = \binom{32}{3}\left(\frac{1}{100}\right)^3\left(\frac{99}{100}\right)^{29} = 0.0003$$

$$P(S_{33} > 3) = 1 - P(0 \le S_{32} \le 3) = 1 - .9962 = .004$$

✓ Perfect test question...



Let's think about expectation, var, SD

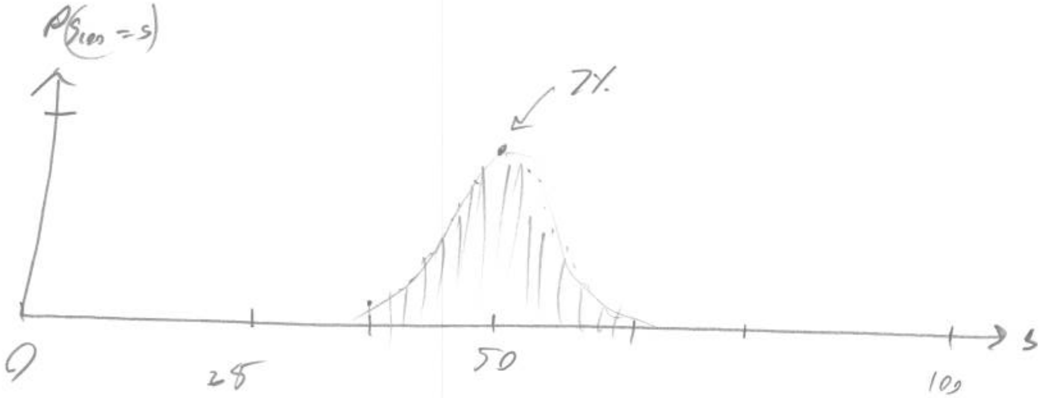
$$E[S_{32}] = E[X_1 + \dots + X_{32}] = 32\, E[X] = 32 \cdot \frac{1}{100} = .32$$

$$\text{Var}[S_{32}] = \text{Var}[X_1 + \dots + X_{32}] = 32\, \text{Var}[X] = 32 \cdot \frac{1}{100} \cdot \frac{99}{100} = .3168$$

$$SD[S_{32}] = \sqrt{.3168} = .56$$

$.32 \pm .56$ teeth to have cavities...

More genl case $n=100$, $p=\frac{1}{2}$



In the genl case where n is large and p is not close to 0 or 1... we get a symmetric, bell-shaped curve... More on this later... This is actually the most special r.v. and one of the most beautiful theories in all of mathematics...

Genl expectation

$S_n \sim \text{Binomial}(n,p)$

$$E[S_n] = \sum_{k=1}^{n} (k) \binom{n}{k} p^k (1-p)^{n-k} \quad \text{tough...}$$

$$= E[X_1 + \dots + X_n] = np$$

$$\text{Var}[S_n] = \sum_{k=1}^{n} (k-np)^2 \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \text{Var}[X_1 + \dots + X_n] = np(1-p) \Rightarrow SD[S_n] = \sqrt{np(1-p)}$$

Now, what if my case was n very large, p very small...

$n \to \infty$, $p \to 0$, but $np = \lambda$

Imagine 911 call center for Philadelphia $n = 1.5$ million

$p \approx \frac{1}{500{,}000}$? For 9AM-10AM? $\lambda = 3$?

Of course we can do this with binomial...

$$P(3 \text{ calls}) = \binom{1{,}500{,}000}{3} \left(\frac{1}{500{,}000}\right)^3 \left(\frac{499{,}999}{500{,}000}\right)^{1{,}499{,}997}$$

very messy... ~~plus~~ what if we don't really know n, p, just some idea of λ... consider

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$p = \frac{\lambda}{n}$ by def of λ

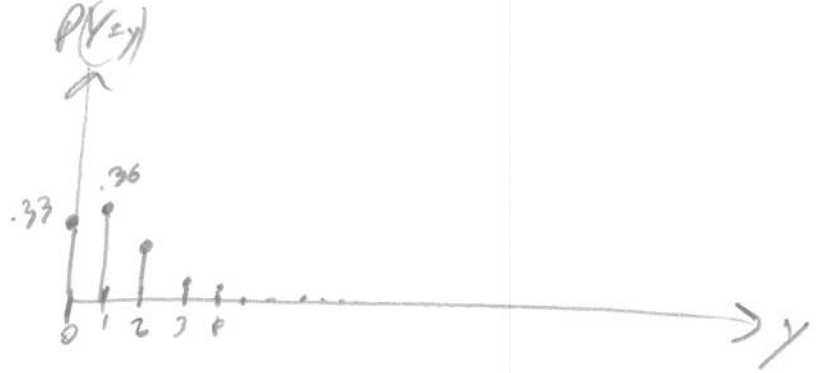
$$= \frac{n!}{(n-k)!\,n^k} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}$$ algebra

$P(k \text{ successes infinite trials})$

$$= \left(\underbrace{\lim_{n\to\infty} \frac{n!}{(n-k)!\,n^k}}\right) \left(\frac{\lambda^k}{k!}\right) \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\lim_{n\to\infty} \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n\cdot\ldots\cdot n} \quad \frac{\lambda^k}{k!} \quad e^{-\lambda} \quad 1 \quad = \frac{e^{-\lambda}\lambda^k}{k!} \triangleq \text{Poisson}(\lambda)$$

$Y \sim \text{Poisson}(\lambda) \quad \lambda = 1.1$



$\text{Support}(Y) = \{0, 1, 2, \dots\}$

goes on forever...

$S_{10} = X_1 + \dots + X_{10}$ s.t. $X_1, \dots, X_{10} \overset{iid}{\sim} \text{Bernoulli}(\frac{1}{2})$

What does $X_1 = ?$ either 0 or 1 don't know...

Best we can do $P(X_1 = 1) = \frac{1}{2}$, $P(X_1 = 0) = \frac{1}{2}$

$x_1 = 1, x_2 = 0$

fill in 10

Data: 1, 0, ___

Data is realizations of r.v.'s!

X_1 is a r.v. $\quad$ x_1 is its realization, after the fact, what happened?? After we draw from Ω

r Ch1, ch2, ch3,
Ch ~~5~~ → stats

Let's define ~~stuff~~ from high school:

$\overline{x} = \frac{x_1 + \dots + x_{10}}{10} = ?$ (small letters over) $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

sample average

Stat 434

$\overline{x}$ is realized from $\overline{X}$ the r.v. for the avg (estimator)
(estimate)

Do not call sample average the mean (even though that's what JMP does)

The mean is reserved for $\mu = E[X]$

The $\overline{x}$ is good for making a guess at μ!

Also... another reason but will do tomorrow

$s^2 = \frac{(x_1 - \overline{x})^2 + \dots + (x_{10} - \overline{x})^2}{9} = ?$ $s^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2$, $s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2}$

$s = \sqrt{s^2}$

What the hell?

s^2 estimates σ^2, s estimates σ

extra credit problem on HW

Couple reasons

① If $n=1$? ② $\sum(x_i - \overline{x})^2 \leq \sum(x_i - \mu)^2$ since $\overline{x} \neq \mu$
⇒ divisor must be smaller to correct for this...

In the case of binomials ... one more bit of notation

$$\hat{p} \triangleq \bar{X} = \frac{X_1 + \dots + X_n}{n}$$ ← Yes/No's 0/1's

is a percentage of Yes/No's 0/1's

we denote this $\hat{p}$ for "sample proportion"

Because we can!

It's better this way... you know for sure the answer is between 0 and 1!

What are statistics? $\bar{X}$, S^2 examples

they are functions of data which was sampled from a population. More on this tomorrow

If time ... do:

build a data table...

binary, nominal, ordinal } categorical

interval, ratio } numerical

DATA Table ch2

observations, cases,

variables, attributes, features, responses, objects

$n \times p$

Not covered: dotplots 2.4, scatterplots ⇒ next week