

Warmup: 9:00 - 9:10

Lucy and two friends each have a p-coin and toss it independently at the same time.

What is the probability that the first person to get a head has to toss more than n times

$$\left(\text{i.e. Find } P\left(\min(X_1, X_2, X_3) > n\right) \right), \quad \begin{array}{c} + \\ + \\ \hline \min(X_1, X_2, X_3) \end{array}$$

$$= P(X_1 > n, X_2 > n, X_3 > n)$$

$$= P(X_1 > n)^3 \text{ by independence}$$

$$= (q^n)^3 \text{ since } X_1 \sim \text{Geom}(p)$$

$$= \boxed{q^{3n}} \quad \begin{array}{l} \text{Fact } X \sim \text{Geom}(p) \\ \text{iff } P(X > n) = q^n \end{array}$$

$$\Leftrightarrow \min(X_1, X_2, X_3) \sim \text{Geom}[1 - q^3]$$

More about min of independent geometrics in the appendix.

Announcement! Will start review Friday.

Last time

Sec 3.4 Geometric Distribution

$X \sim \text{Geom}(p)$ on $1, 2, \dots$

$X = \# \text{ trials until } \text{ first success.}$

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{q}{p^2}$$

More generally:

Negative Binomial Distribution ($\text{NegBin}(r, p)$)

generalization of $\text{Geom}(p)$

Sum of
 r indep $\text{Geom}(p)$
on $\{r, r+1, r+2, \dots\}$

ex $r=3$

$\underbrace{w_1 w_2 w_3}_{\# \text{ trials until 3rd success}}$

let $T_r \sim \text{NegBin}(r, p)$

$T_r = \# \text{ indep } p\text{-trials until } r^{\text{th}} \text{ success}$

$\underbrace{\quad}_{\text{P}}$

$$P(T_r = k) = \binom{k-1}{r-1} p^{r-1} q^{k-r}$$

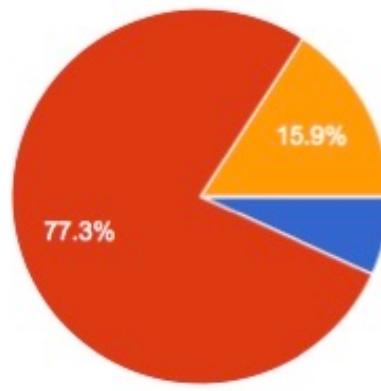
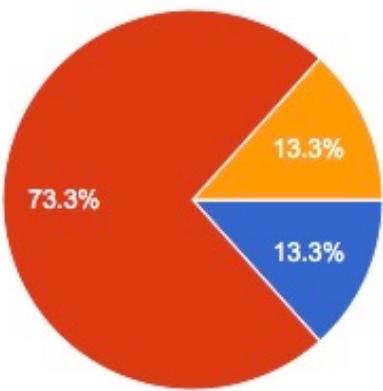
$\binom{k-1}{r-1}$ is highlighted with a green oval and labeled $k-r$

$$= \binom{k-1}{r-1} p^r q^{k-r}$$

$T_r = w_1 + \dots + w_r$ where $w_1, \dots, w_r \stackrel{iid}{\sim} \text{Geom}(p)$

$$E(T_r) = r E(w_i) = \boxed{\frac{r}{p}}$$

$$\text{Var}(T_r) = r \text{Var}(w_i) = \boxed{\frac{rq}{p^2}}$$



1. The population of a small town is 1000 with 50% democrats. We wish to know what is a more accurate assessment of the % of democrats in the town, to randomly sample with or without replacement? The sample size is 10.

a with replacement

b without replacement

c same accuracy with or without replacement

d not enough info to answer the question

c

N large enough that sampling with or without replacement has minimal effect. Also, p close to 1/2

Correction factor is $(1000-10)/(999) = 0.99$

The variance of hypergeometric is smaller than the variance of binomial. Therefore, sampling without replacement will give more accurate results.

R simulation for SD of # democrats in sample

Code Start Over Solution

```
1 pop <- 1000
2 sample_size <- 10
3 a <- rep(0,each=pop/2)
4 b <- rep(1,each=pop/2)
5 box <- c(a,b)
6 for(boolean in c(TRUE,FALSE)){
7   fun <- function(){
8     my_sample <- sample(box,size=sample_size,replace=boolean)
9     mean(my_sample)*100
10  }
11  B <- 10000
12  vec_percentages <- replicate(B,fun())
13  print(sd(vec_percentages))
14 }
```

replacement without replacement

SD =
[1] 15.77621 ← replacement
SD =
[1] 15.24316 ← without replacement. (Smaller)

Today

- ① Sec 3.5 Poisson distribution
- ② Poisson random scatter (PRS) AKA
Poisson process

① Sec 3.5 Poisson distribution ($\text{Pois}(n)$)

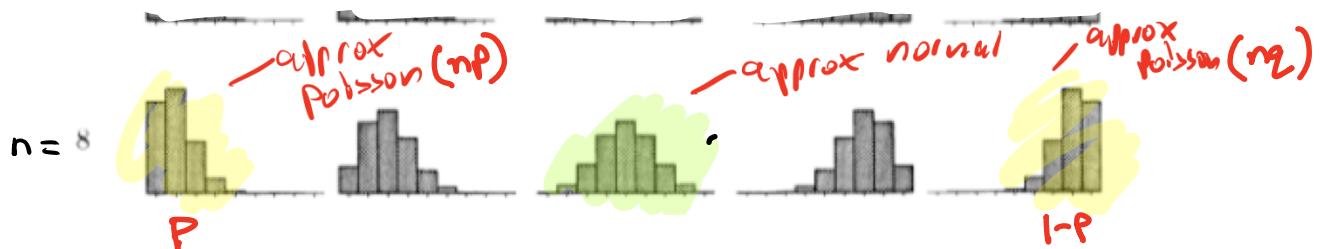
$$X \sim \text{Pois}(n)$$

$$P(X=k) = \frac{e^{-n} n^k}{k!} \quad k=0, 1, 2, \dots$$

Intuitively, we know $E(X)=n$ and $\text{Var}(X)=n$

since,

$$\text{Bin}(n, p) \rightarrow \text{Pois}(n) \quad \text{when } \begin{cases} n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow n \end{cases}$$



Also we expect $npq \rightarrow nq \times 1$ so $\text{var}(X)$ should be n . See appendix for a proof.

Ex Let $X \sim \text{Pois}(n)$

$$\text{Find } E(\underbrace{X(X+1)}_{X^2+X}) = E(X^2) + E(X)$$

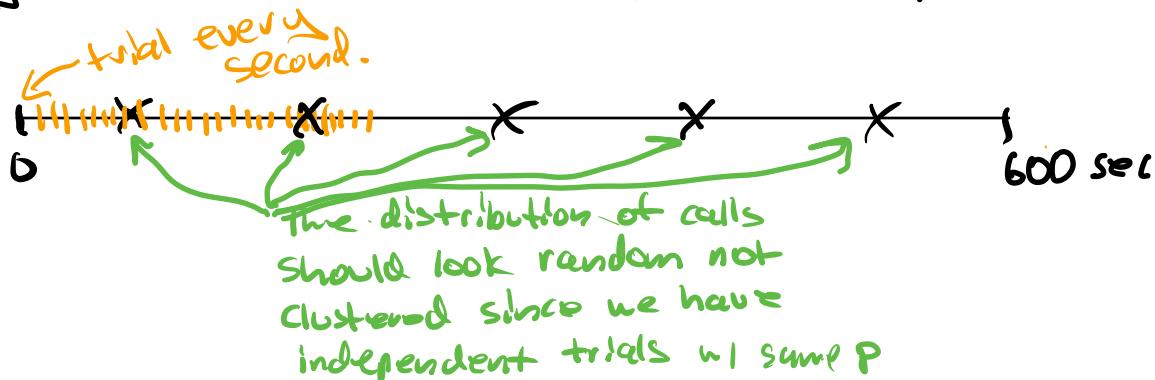
$$\text{II} \quad \text{Var}(X) + (E(X))^2$$

$$= n + n + n^2 \underbrace{\{2n + n^2\}}$$

(2) Poisson Random Scatter (PRS)

A random scatter of points in a time line is an example of a Poisson random scatter,

ex X = number of calls coming into a hotel reservation center in 600 seconds
Choose an interval of time so no time interval gets more than one call (~~or~~ seconds),



PRS assumption

1) No time interval gets more than one call

2) Have n iid Bernoulli P trials with $M = np$ large n , small P .

(i.e. all calls are independent of each other with the same probability)

Let $X = \# \text{ calls in } \underline{t \text{ seconds}}$,
time of n trials

Then $X \sim \text{Pois}(n)$ ← limit of $\text{Bin}(n, p)$ as $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \mu$.

Say on average there are $\mu = 5$ calls in 600 seconds

Let λ be the rate (or intensity)
of calls per second

ex $\lambda = \frac{5}{600}$ calls/sec in above example.

Since λ is the same every time interval
(Pois assumptions) $n = \lambda t$.

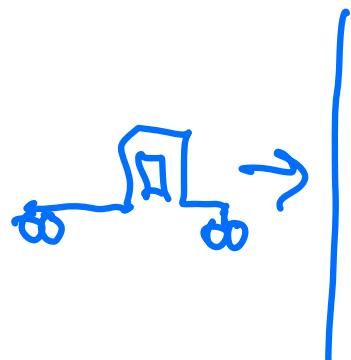
λ has units calls/sec so $n = \lambda t$ has units calls
in t sec

ex $n = \lambda t = \frac{5}{600} \cdot 600 = 5$ calls in 600 sec.

Stat 134

1. Which of the following can be modeled as a Poisson Random Scatter with intensity $\lambda > 0$?

- a The number of blueberries in a 3 cubic inch blueberry muffin
- b The number of patients entering a doctor's office in a 24 hour period.
- c The number of times a day a person feels hungry
- d The number of air pulses counted every second from cars driving over an empty rubber hose lying across a highway between noon and 1pm.
- e more than one of the above



Appendix

Let $X \sim \text{Pois}(\mu)$

Then $E(X) = \mu$ and

$$\text{Var}(X) = \mu$$

Pf/

Recall $\bar{e}^{\bar{\mu}} = 1 + \mu + \frac{\mu^2}{2!} + \dots = \sum_{k=0}^{\infty} \frac{\mu^k}{k!}$ Taylor Series.

$$\begin{aligned}
 E(X) &= \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k \bar{e}^{-\mu} \frac{\mu^k}{k!} \\
 &= \sum_{k=1}^{\infty} k \bar{e}^{-\mu} \frac{\mu^{k-1} \mu}{(k-1)! k} \quad (\text{note } 0 \cdot \bar{e}^{-\mu} \frac{\mu^0}{0!} = 0) \\
 &= \mu \bar{e}^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \\
 &= \mu \bar{e}^{-\mu} \left(\underbrace{1 + \mu + \frac{\mu^2}{2!} + \dots}_{\bar{e}^{\bar{\mu}}} \right) = \boxed{\mu}
 \end{aligned}$$

Next we show $\text{Var}(X) = \mu$:

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - E(X)^2 \\
 &= E(X^2) - E(X) + E(X) - E(X)^2 \\
 &= \boxed{E(X(X-1))} + E(X) - E(X)^2
 \end{aligned}$$

$$E(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) P(X=k)$$

!!
 $e^{-\mu} \frac{\mu^k}{k(k-1)(k-2)!}$

$$\begin{aligned}
 &= e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} \\
 &= e^{-\mu} \mu^2 \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} = \mu^2
 \end{aligned}$$

$$\Rightarrow \text{Var}(X) = \mu^2 + \mu - \mu^2 = \boxed{\mu}$$

□

Answers

Minimum of independent geometrics

Adam, Beth and John independently flip a P_1, P_2, P_3 coin respectively.
let $X = \# \text{ trials until Adam, Beth or John get a heads.}$

A	TTT	$X_1 \sim \text{Geom}(P_1)$
B	TTT	$X_2 \sim \text{Geom}(P_2)$
J	$\underbrace{\text{TTH}}$	$X_3 \sim \text{Geom}(P_3)$
	$X = 3$	

a) What is probability Adam, Beth or John get a head?

$$\begin{aligned} P &= \text{Prob}[A \text{ or } B \text{ or } J \text{ get heads}] \\ &= 1 - \text{Prob}(A, B, J \text{ don't get heads}) \\ &= 1 - q_1 q_2 q_3 \end{aligned}$$

b) What distribution is X ?

$$X = \min(X_1, X_2, X_3) \sim \text{Geom}(1 - q_1 q_2 q_3)$$

