Stat 134 lec 15

Warmup:

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

(i.e. Find $P(\min(X_1,X_2,X_3) > n)$).

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

$$= P(X_1>n)^3 = (q^n)^3 = q^{3n} = (q^3)^n$$

independent

Fact $X \sim Geom(p)$

iff $P(X>n) = q^n$

It follows that

$\min(X_1,X_2,X_3) \sim Geom(1-q^3)$

More about min of independent geometrics in the appendix.

Announcement! Will start review Friday.
review materials are on stat134.org

last time

sec 3.4 Geometric Distribution

$X \sim Geom(p)$ on $1, 2, \ldots$

X = # trials until the first success.

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

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

More generally:

Negative Binomial Distribution $\left(NegBin(r, p)\right)$

generalization of Geom(p)

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

ex $r = 3$

$\underbrace{qqqp}_{W_1}\underbrace{qqp}_{W_2}\underbrace{p}_{W_3}$ # trials until 3rd success

let $T_r \sim NegBin(r, p)$

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

r-1 p in k-1 slots

$\underbrace{\qquad\qquad\qquad\qquad}$ p

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

$$T_r = W_1 + \dots + W_r \quad \text{where } W_1, \dots, W_r \overset{iid}{\sim} \text{Geom}(p)$$

$$E(T_r) = r E(W_1) = \boxed{\frac{r}{p}}$$

$$\text{Var}(T_r) = r \text{Var}(W_1) = \boxed{\frac{rq}{p^2}}$$

Stat 134

1. The population of a small town is 1000 with 50% democrats. We wish to know what is a more accurate assesment 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

b	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

```r
pop <- 1000
sample_size <- 10
a <- rep(0,each=pop/2)
b <- rep(1,each=pop/2)
box <- c(a,b)
for(boolean in c(TRUE,FALSE)){
  fun <- function(){
    my_sample <- sample(box,size=sample_size,replace=boolean)
    mean(my_sample)*100
  }
 B <- 10000
  vec_percentages <- replicate(B,fun())
  print(sd(vec_percentages))
}
```

replacement ← TRUE

without replacement ← FALSE

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

<u>Today</u>

1. <u>Sec 3.5</u> Poisson distribution
2. Poisson random scatter (PRS) AKA Poisson Process

① Sec 3.5 Poisson distribution ($Pois(\mu)$)

$X \sim Pois(\mu)$

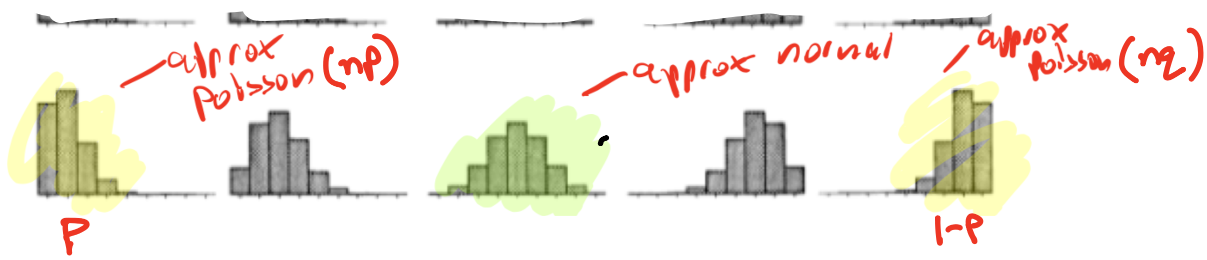
$P(X=k) = \frac{e^{-\mu}\mu^k}{k!}$ $k=0,1,2,\dots$

Intuitively, we know $E(X)=\mu$ and $Var(X)=\mu$

since,

$Bin(n,p) \rightarrow Pois(\mu)$ when $n\to\infty$, $p \to 0$, $np \to \mu$.

n = 8



Also we expect $npq \rightarrow \mu q \approx \mu$ so $var(X)$ should be μ. See appendix for a proof.

ex Let $X \sim Pois(\mu)$

Find $E(X(X+1)) = E(X^2+X) = E(X^2) + E(X)$

$E(X^2) = Var(X) + (E(X))^2$

$= \mu + \mu^2 + \mu$

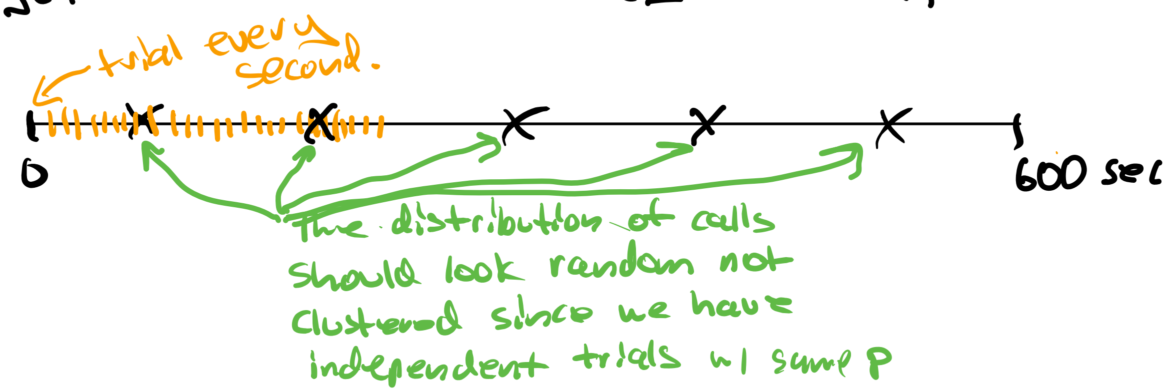
$= \boxed{2\mu + \mu^2}$

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 (ex seconds).



PRS assumptions

1) No time interval gets more than one call

2) Have n iid Bernoulli p trials with $\mu = np$ large n, small p.

(i.e all calls are independent of each other with the same probability) let X = # calls in $\underbrace{t \text{ seconds}}_{\text{time of } n \text{ trials}}$

Then $X \sim Poi(\mu) \longleftarrow$ limit of $Bin(n,p)$ as $n \to \infty$, $p \to 0$, $np \to \mu$.

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

Let $\lambda > 0$ be the rate (or intensity) of calls per second

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

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

λ has units calls/sec

$\mu = \lambda t$ has units calls in t sec

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

tinyurl.com/febr22-2023

Stat 134

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

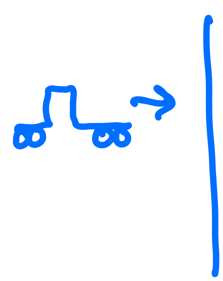
   a The number of blueberries in a 3 cubic inch blueberry muffin

   ~~**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 $X \sim Pois(\mu)$

Then $E(X) = \mu$ and

$Var(X) = \mu$

Pf/

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

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k e^{-\mu} \frac{\mu^k}{k!}$$

$$= \sum_{k=1}^{\infty} \cancel{k} e^{-\mu} \frac{\mu^{k-1}\mu}{(k-1)!\cancel{k}}$$

(note $0 \cdot e^{-\mu} \frac{\mu^0}{0!} = 0$)

$$= \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!}$$

$$= \mu e^{-\mu} \underbrace{\left(1 + \mu + \frac{\mu^2}{2!} + \cdots\right)}_{e^{\mu}} = \boxed{\mu}$$

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

$$Var(X) = E(X^2) - E(X)^2$$

$$= E(X^2) - E(X) + E(X) - E(X)^2$$

$$= E(X(X-1)) + E(X) - E(X)^2$$

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

$$P(X=k) = \frac{e^{-\mu}\mu^{k}}{\cancel{k}(\cancel{k-1})(k-2)!}$$

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

$$= e^{-\mu}\mu^{2} \underbrace{\sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!}}_{= e^{\mu}} = \mu^{2}$$

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

$\square$

Appendix

Minimum of independent geometrics

Adam, Beth and John independently flip a p_1, p_2, p_3 coin respectively, let X = # trials until Adam, Beth or John get a heads.

ex
A TTT $X_1 \sim Geom(p_1)$
B TTT $X_2 \sim Geom(p_2)$
J TTH $X_3 \sim Geom(p_3)$

$\underbrace{TTH}_{X=3}$

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

Two methods

① use inclusion-exclusion (harder)

$$P(A \text{ or } B \text{ or } J \text{ get heads}) = p_1 + p_2 + p_3 - p_1p_2 - p_1p_3 - p_2p_3 + p_1p_2p_3$$

$$= 1 - (1-p_1)(1-p_2)(1-p_3)$$

$$= 1 - q_1q_2q_3$$

② use complement

$$= 1 - P(A, B, J \text{ don't get heads})$$

$$= \boxed{1 - q_1q_2q_3}$$

b) what distribution is X?

$X \sim \text{Geom}(1 - q_1 q_2 q_3)$

Note that $X = \min(X_1, X_2, X_3)$.

Compare this problem with the warmup.