Stat 134 lec 16

last time

correction factor

sec 3.6 $X \sim Bin(n,p) \Rightarrow Var(X) = np(1-p)$

$X \sim HG(n,N,G) \Rightarrow Var(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$

where $p = \frac{G}{N}$

sec 3.4 Geometric distribution $(Geom(p))$

X = # p-coin tosses until the first heads

takes values 1, 2, 3, ...

$P(X=k) = q^{k-1}p$

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

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

Negative Binomial distribution $(NegBin(r,p))$

$T_r = W_1 + \cdots + W_r$ where $W_1, W_2, \ldots, W_r \overset{iid}{\sim} Geom(p)$

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

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

$Var(T_r) = \frac{rq}{p^2}$

Today

0. Review concept test from last time
1. Finish sec 3.4 Minimum of independent geometrics
2. Poisson distribution
3. Poisson random scatter (PRS) AKA Poisson Process
4. Poisson thinning

6

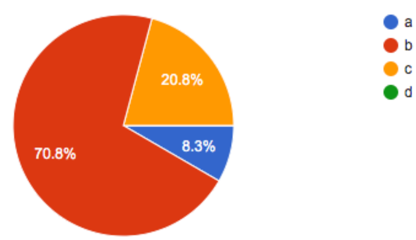
The population of a small town is 1000 with 50% democrats. We wish to know what is a more accurate assesment of the % of democrats in the town, to randomly sample with or without replacement? The sample size is 10.

a with replacement

b without replacement

c same accuracy with or without replacement

d not enough info to answer the question



c	When you calculate the var. It's the same

b	The sample size is 10. The population is 1000 with 50% demos. We are expecting 5 out of 10 are demos. We need a tighter percentage. If we do with replacement, the chance of getting demo people will be higher. So we need to do without replacement.

1) Sec 3.4 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.

eg

A	TTT	$X_1 \sim \text{Geom}(p_1)$
B	TTT	$X_2 \sim \text{Geom}(p_2)$
J	$\underbrace{\text{TTH}}_{X=3}$	$X_3 \sim \text{Geom}(p_3)$

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

$P = \text{Prob}(A \text{ or } B \text{ or } J \text{ get heads})$

$= 1 - \text{Prob}(A, B, J \text{ dont get heads})$

$= \boxed{1 - q_1 q_2 q_3}$

b) what distribution is X?

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

② Sec 3.5 Poisson distribution (Pois(μ))

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

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

Find $E(X)$, $\text{Var}(X)$.

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

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

$$= \sum_{k=1}^{\infty} \cancel{k} e^{-\mu}\frac{\mu^{k-1}\mu}{(k-1)!\,\cancel{k}} \qquad \left(\text{note } 0\cdot e^{-\mu}\frac{\mu^0}{0!} = 0\right)$$

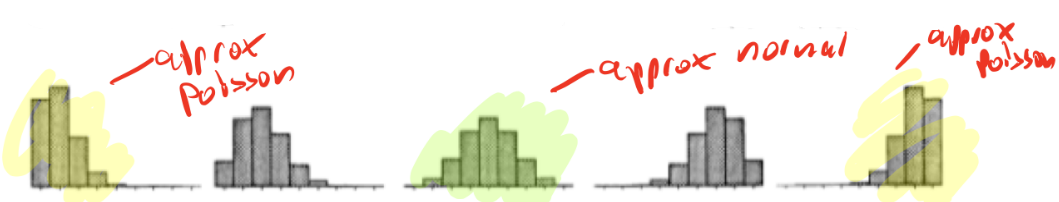
$$= \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!}$$

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

This makes sence since

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

$n = 8$



Also we expect

$$npq \longrightarrow \mu q \approx \mu \quad \text{since } q \to 1$$

So $\text{Var}(X)$ should be μ,

lets check:

$$\text{Var}(X) = E(X^2) - E(X)^2$$
$$= E(X^2) - E(X) + E(X) - E(X)^2$$
$$= E(X(X-1)) + E(X) - E(X)^2$$

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

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

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

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

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

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

Find $E(X(X+1))$

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

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

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

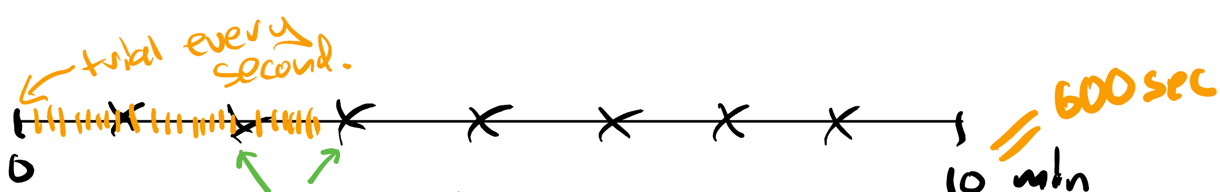
③ Poisson Random Scatter (PRS)

$\text{Pois}(np) \approx \text{Bin}(n,p)$ with large n and small p

models situations with low probability, independent trials.

ex $X =$ number of calls coming into a hotel reservation center in 10 minutes.

Say $\mu = 5$ call in 10 min



The distribution of calls should look random not clustered since we have independent trials w/ same p.

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)

As $n \to \infty$ and $p \to 0$ and $np \to \mu$
let $X =$ # calls in $\underbrace{t \text{ minutes}}_{\text{time of } n \text{ trials}}$.

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

By independence of counts in disjoint time intervals there is a constant rate (intensity) λ of calls per unit time.

ex $\lambda = 5/10 = 1/2$ calls/min in above example.
and $\mu = \lambda t = \frac{1}{2} \cdot 10 = 5$ calls

tinyurl.com/october4-pt1

Stat 134

Friday September 28 2018

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. — no patients outside regular hours

   c The number of times a day a person feels hungry — not indep hits.

   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. — not indep hits since come in pairs.

   e more than one of the above

ex (number of bb in a bb muffin)

A tub of bb muffin batter has

$\lambda = 2 \text{ bb}/\text{in}^3$.

A muffin is 3 in^3.

On average how many bb are there per muffin?

$\mu = \lambda \cdot 3 = 2 \cdot 3 = 6$ bb

Let $X_1 = \#$ bb in 1st muffin

$X_1 \sim \text{Pois}(6)$

Another muffin is 4 in^3 (from the same batter)

let $X_2 = \#$ bb in 2nd muffin.

a) Find P(5 bb in each muffin)

$X_1 \sim \text{Pois}(6)$

$X_2 \sim \text{Pois}(8)$

$$P(X_1 = 5, X_2 = 5) = \frac{e^{-6} 6^5}{5!} \cdot \frac{e^{-8} 8^5}{5!}$$

b) Find P (10 bb total in both muffin together)

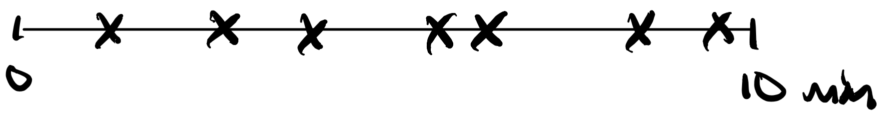
$X_1 + X_2 \sim \text{Pois}(14)$

$$P(X_1 + X_2 = 10) = \frac{e^{-14} 14^{10}}{10!}$$

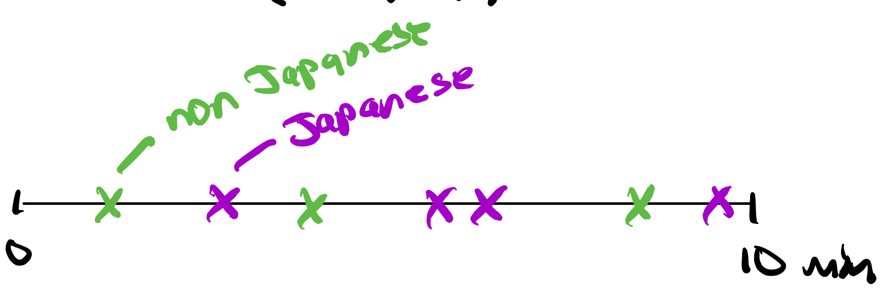
④ Sec 3.5 Poisson thinning

Cars arrive at a toll booth according to a Poisson process at a rate $\lambda = 3$ arrivals/min

X = # cars arriving at a toll booth in 10 min. $X \sim \text{Pois}(\underbrace{\lambda \cdot 10}_{30})$



Of cars arriving, it is known, over the long term, that 60% are Japanese imports.



Call Japanese cars a success and non Japanese a failure.

Each hit is a success with $p = .6$, independent of all other hits.

Then the process of "success" hits in your PRS is a PRS with intensity λp and the process "failure" hits in your PRS is a PRS with intensity λq