Stat 88 lec 15

Warmup 2:00–2:00 (exercise 5.7.11)

Let X = the number of cars owned by a Cal student.

Here is the distribution of X

number of cars	0	1	2
probability	2θ	θ	$1-3\theta$

a) Find $E(X)$ (as a function of θ)

b) Let $X_1, \ldots, X_n$ be the number of cars owned by n randomly picked students.

Use $\bar{X}$ to find an unbiased estimator of θ.

a) $E(X) = 0 \cdot 2\theta + 1 \cdot \theta + 2(1-3\theta) = \boxed{2-5\theta}$

b) know $E(\bar{X}) = E(X)$ since $\bar{X}$ is an unbiased estimator of $\mu = E(X)$.

$$E(\bar{X}) = 2-5\theta$$
$$5\theta = 2 - E(\bar{X})$$
$$\theta = \frac{2-E(\bar{X})}{5} = E\left(\boxed{\frac{2-\bar{X}}{5}}\right)$$

unbiased estimator

Last time

sec 5.4 unbiased estimators

Probability distributions often have parameters that we wish to estimate. An estimator is a RV. With an unbiased estimator, on average the estimator will be the parameter.

ex Lets estimate the largest possible value of the distribution Uniform$\{1, 2, 3, \dots, N\}$ where N is an unknown parameter.

let $X_1, \dots, X_n$ be a SRS of $\{1, 2, \dots, N\}$.

Two possible estimators:

① $M = \max\{X_1, \dots X_n\}$ ← biased

② $T = 2\bar{X} - 1$ ← unbiased

T is unbiased since:

$$E(T) = E(2\bar{X} - 1) = 2E(\bar{X}) - 1$$
$$= 2E(X) - 1$$
$$= 2\left(\frac{N+1}{2}\right) - 1 = N$$

There are pros and cons to both estimators

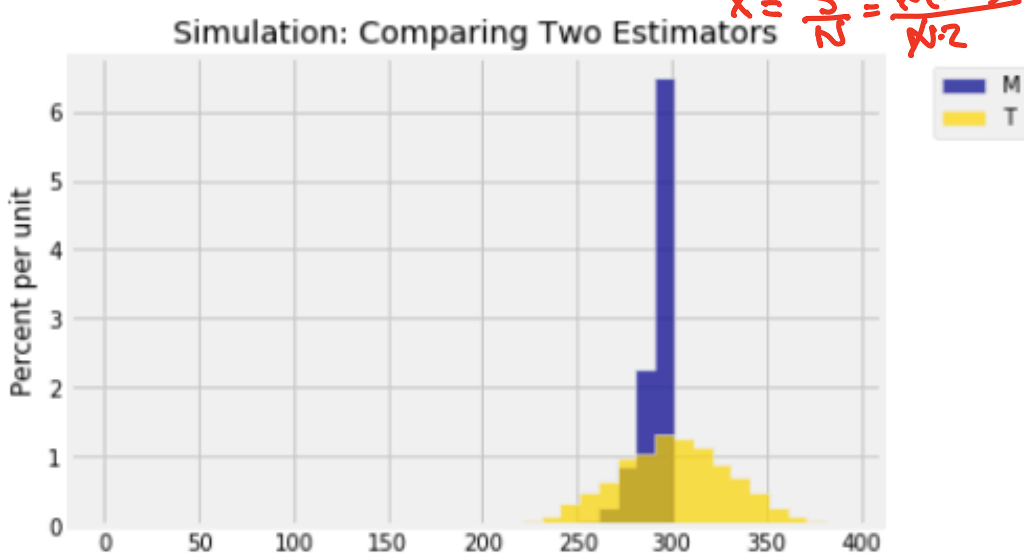
why is $\bar{X} = \frac{N+1}{2}$?

$$\begin{aligned} & 1+2+3+\dots+N = S \\ + \; & N + N-1 + N-2 + \dots + 1 = S \\ \hline & \underbrace{(N+1)+(N+1)+\dots+(N+1)}_{N} = 2S \end{aligned}$$

$$N(N+1) = 2S \Rightarrow S = \frac{N(N+1)}{2}$$

$$\bar{X} = \frac{S}{N} = \frac{\cancel{N}(N+1)}{\cancel{N} \cdot 2} = \boxed{\frac{N+1}{2}}$$

```
N = 300
n = 30
5000 repetitions
```



The histograms show that both estimators have pros and cons.

M pros: small spread of values
cons: biased

T pros: unbiased
cons: big spread of values

Today <u>Sec 5.5</u> Conditional expectation
<u>Sec 5.6</u> applications

sec 5.5 conditional expectation

Expectation

Lets review how to find expectation of a joint distribution.

. A joint distribution for two random variables M and S is given below.

	$M = 2$	$M = 3$
$S = 2$	0	$\frac{1}{3}$
$S = 1$	$\frac{1}{3}$	$\frac{1}{3}$

E(M) equals:

$$E(M) = \sum_{\text{all } s} \sum_{\text{all } m} m P(M=m, S=s)$$

$$= 2 \cdot 0 + 3 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = \boxed{2.67}$$

Conditional expectation

What is $E(M \mid S=2)$? — called conditional expectation

		$M = 2$	$M = 3$
$\frac{1}{3}$	$S = 2$	0	$\frac{1}{3}$
$\frac{2}{3}$	$S = 1$	$\frac{1}{3}$	$\frac{1}{3}$

$$2 \cdot P(M=2 \mid S=2) + 3 \cdot P(M=3 \mid S=2)$$

$$P(M=2 \mid S=2) = \frac{P(M=2, S=2)}{P(S=2)} \qquad P(M=3 \mid S=2) = \frac{P(M=3, S=2)}{P(S=2)}$$

$$2 \cdot 0 + 3 \cdot 1 = \boxed{3}$$

	$M=2$	$M=3$
$S=2$	0	$\frac{1}{3}$
$S=1$	$\frac{1}{3}$	$\frac{1}{3}$

Find $E(M \mid S=1)$

$$2 \cdot \frac{\frac{1}{3}}{\frac{2}{3}} + 3 \frac{\frac{1}{3}}{\frac{2}{3}} = \boxed{2.5}$$

<u>What is relationship between expectation and conditional expectation?</u>

$$E(M) = \sum_{\text{all } s} \sum_{\text{all } m} m P(M=m, S=s)$$

By multiplication rule,

$$P(M=m, S=s) = P(M=m \mid S=s) \cdot P(S=s)$$

So $$E(M) = \sum_{\text{all } s} \sum_{\text{all } m} m P(M=m, S=s)$$

$$= \sum_{\text{all } s} \underbrace{\sum_{\text{all } m} m P(M=m \mid S=s)}_{E(M \mid S=s)} P(S=s)$$

$$\boxed{E(M) = \sum_{\text{all } s} E(M \mid S=s) P(S=s)}$$

$$E(M) = \underset{3}{E(M \mid S=2)} \, \underset{\frac{1}{3}}{P(S=2)} + \underset{2.5}{E(M \mid S=1)} \, \underset{\frac{2}{3}}{P(S=1)} = 1 + \frac{5}{3} = \frac{8}{3} = \boxed{2.67}$$

__Note__

$E(M \mid S=s)$ is a __function__ of S

ex $E(M \mid S=2) = 3$

$E(M \mid S=1) = 2.5$

Sec 5.6 application

To find expectation of one RV it sometimes helps to condition on another RV.

ex

Time to Reach Campus

A student has two routes to campus. Each route has a random duration. The student prefers Route A because its expected duration is 15 minutes compared to 20 minutes for Route B. However, on 10% of her trips she is forced to take Route B because of road work on Route A. On the remaining 90% of the days she takes Route A.

What is the expected duration of her trip on a random day?

D = duration of trip.

Find $E(D)$

R = route — A or B

$$E(D) = \underbrace{E(D \mid R=A)}_{=15}\underbrace{P(R=A)}_{=.9} + \underbrace{E(D \mid R=B)}_{=20}\underbrace{P(R=B)}_{=.1}$$

$$= \boxed{15.5}$$

ex (exercise 5.7.13)

13. A survey organization in a town is studying families with children. Among the families that have children, the distribution of the number of children is as follows.

Number of Children n	1	2	3	4	5
Proportion with n Children	0.2	0.4	0.2	0.15	0.05

Suppose each child has chance 0.51 of being male, independently of all other children. What is the expected number of male children in a family picked at random from those that have children?

M = # male children

find $E(M)$

N = # children

$M|N=n \sim \text{Binomial}(n, .51)$

$E(M|N=n) = n(.51)$

$$E(M) = \sum_{\text{all } n} \underbrace{E(M|N=n)}_{= n(.51)} \underbrace{P(N=n)}_{\text{table}}$$

$$= .51\left(1(.2) + 2(.4) + 3(.2) + 4(.15) + 5(.05)\right)$$

$$= \boxed{1.25}$$

Tinyurl.com/feb26-pt1

Stats 88

Wednesday February 26 2020

1. You flip a fair coin N times where N is a random variable $N \sim$Poisson(5). What is the expected number of heads you will get?

   a 5

   b $\frac{5}{2}$

   c $\frac{N}{2}$

   d none of the above

X = # heads

$X \mid N=n \sim \text{Bin}(n, \frac{1}{2})$

$E(X \mid N=n) = \frac{n}{2}$

$$E(X) = \sum_{\text{all } n} \frac{n}{2} P(N=n) = \frac{1}{2} \boxed{\sum_{\text{all } n} nP(N=n)}$$

$\sum_{\text{all } n} nP(N=n) = E(N) = 5$

$= \boxed{\frac{5}{2}}$