* Announcement :

① HW6 due Wed (10/7)

② Midterm practice problems on Piazza

③ Next week : Midterm review / practice
No HW/Quiz

STAT 88: Lecture 16

Contents

Probability and Statistics

Probability: A population distribution is known. We draw a sample X (or n many samples $X_1, \dots, X_n$) from the population and calculate the likelihood of an event, i.e. $P(X \in A)$.

E.x. Pick 5 cards from a deck of size 52 with replacement. What is chance you get 2 ace cards?

$\rightarrow X_i = \begin{cases} 1 & \text{if ith card is ace} \\ 0 & \text{else} \end{cases} \sim Ber(\frac{4}{52})$

$P(\underbrace{X_1 + X_2 + \cdots + X_5}_{= X \sim Binom(5, \frac{4}{52})} = 2) = ?$

Joint distribution table

$X=x$ \ $Y=y$	$y=1$	$y=2$
$x=1$	0.5	0.1
$x=2$	0.2	0.2

Draw a sample (X, Y) from the dist'n table above. What is chance that $X=Y$?

$\rightarrow P(X=Y)? = 0.5 + 0.2 = 0.7$

Statistics: We draw n many samples $X_1, \dots, X_n$ from a population distribution which has unknown parameter θ (e.g. population mean). We then use samples to estimate/draw inference about θ.

E.x. Pick 5 cards from a deck of size N with replacement. If we know the deck contains 4 ace cards, what is N?

$\rightarrow X_i = \begin{cases} 1 & \text{if ith card is ace} \\ 0 & \text{else} \end{cases} \sim Ber(\frac{4}{N})$

$\bar{X} = \frac{X_1 + \cdots + X_5}{5}$ estimates $\frac{4}{N}$

$\rightarrow \boxed{\hat{N} = \frac{4}{\bar{X}}}$

$\bar{X} \rightsquigarrow \frac{4}{N}$

$\Rightarrow \frac{1}{\bar{X}} \rightsquigarrow \frac{N}{4}$

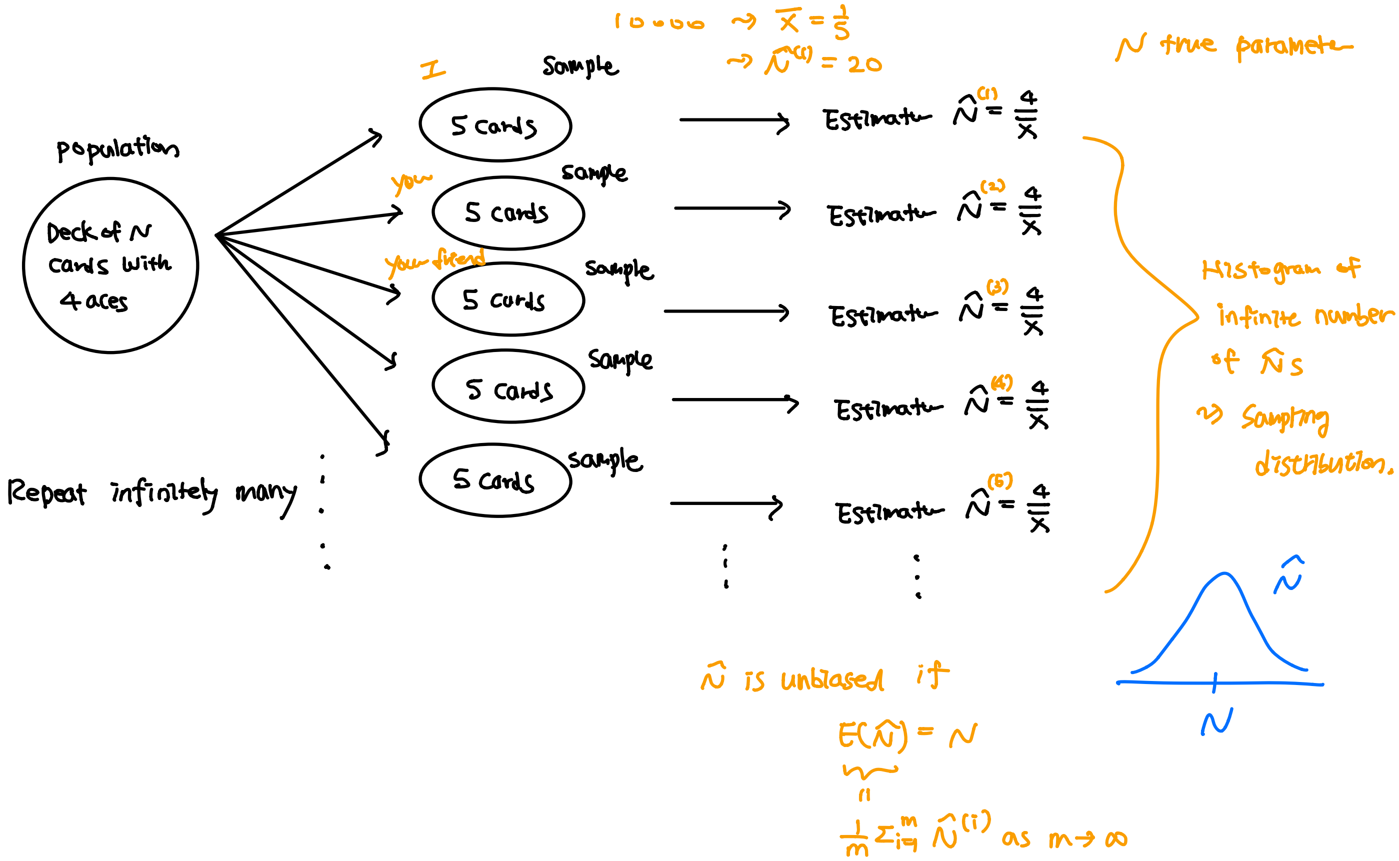
$\Rightarrow \frac{4}{\bar{X}} \rightsquigarrow N$

Joint distribution table

$X=x$ \ $Y=y$	$y=1$	$y=2$
$x=1$	5θ	2θ
$x=2$	θ	$1-8\theta$

Draw $(X_1, Y_1), \dots, (X_n, Y_n)$ from the dist'n table above. How to estimate θ?

Sampling distribution and unbiased estimator: The distribution of statistic or estimator is called sampling distribution. If an estimator is unbiased, its sampling distribution has expectation that equals to the population parameter θ.



Last time

Conditional Expectation

Let X and S be two random variables with joint distribution

	$X = 1$	$X = 2$	$X = 3$
$S = 2$	0.0625	0	0
$S = 3$	0.125	0.125	0
$S = 4$	0.0625	0.25	0.0625
$S = 5$	0	0.125	0.125
$S = 6$	0	0	0.0625

The conditional distribution table:

	$X = 1$	$X = 2$	$X = 3$	$E(X \mid S = s)$
Conditional distribution of X given $S = 2$	* 1 $= \frac{P(X=1, S=2)}{P(S=2)}$	* 0	* 0 $=1$	1
Conditional distribution of X given $S = 3$	0.5	0.5	0 $=1$	1.5
Conditional distribution of X given $S = 4$	0.1667	0.6667	0.1667 $=1$	2
Conditional distribution of X given $S = 5$	0	0.5	0.5	2.5
Conditional distribution of X given $S = 6$	0	0	1	3

We define

$$E(X|S = s) = \sum_{\text{all } x} x \cdot P(X = x|S = s).$$

Then it can be shown that

$$E(X) = \sum_{\text{all } s} E(X|S = s)P(S = s).$$

Y
||

S=s → E(X|S) takes value E(X|S=s)

Note that $E(X|S)$ is a random variable, i.e. a function of S. Recall from Ch5.2 that a function of random variable $f(S)$ is also a random variable. The expectation of $f(S)$ is given by

$$E(f(S)) = \sum_{\text{all } s} f(s)P(S = s).$$

Let $f(S) = E(X|S)$. Then

$$E(E(X|S)) = \sum_{\text{all } s} E(X|S = s)P(S = s) = E(X).$$

This proves the law of iterated expectation:

w.r.t. distribution of S

$$E(X) = E(E(X|S)).$$

taking expectation w.r.t. cond'n dist'n X|S=s

5.6. Expectation by Conditioning

To find expectation of one random variable, it sometimes helps to condition on another random variable.

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.

$R = \begin{cases} \text{route A} & 90\% \\ \text{route B} & 10\% \end{cases}$

$E(D) = E(D \mid R=A)\,P(R=A) + E(D \mid R=B)\,P(R=B)$

$= 15 \cdot 0.9 + 20 \cdot 0.1$

$= 15.5$ mins.

Example: (Exercise 5.7.13) 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?

$M = \#$ of male children

Find $E(M)$?

$N = \#$ children

$M \mid \underline{N=n} \sim \text{Binomial}(\underline{n, 0.51})$

Should depend on n

$E(M \mid N=n) = 0.51 \cdot n$

$E(M) = \sum_n E(M \mid N=n) P(N=n)$

$= \sum_n 0.51 \cdot n \cdot P(N=n)$

$= 0.51(1 \cdot 0.2 + 2 \cdot 0.4 + 3 \cdot 0.2 + 4 \cdot 0.15 + 5 \cdot 0.05)$

$= 1.25$

Example: 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?

$X =$ # heads

$X \mid N=n \sim$ Binomial$(n, 0.5)$

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

$E(X) = \sum_n E(X \mid N=n) P(N=n)$

$= \sum_n \frac{n}{2} \cdot P(N=n)$

$= \frac{1}{2} \underbrace{\sum_n n \cdot P(N=n)}_{=}$, $N \sim$ Poisson$(\underline{\underline{5}})$

$= \frac{1}{2} \cdot 5$

Expectation of a Geometric Waiting Time Let X be # p-coin tosses til first heads. Then $X \sim \text{Geom}(p)$.

$P(\text{Head}) = p$

X takes values on $1, 2, 3, \ldots$. Recall

$$\begin{aligned} P(X > 1) &= P(\text{You need more than 1 trial to get 1st head}) \\ &= P(\text{First trial is failure}) \\ &= 1 - p. \end{aligned}$$

Don't use method of indicators to find expected waiting time since you don't know how many indicators you need.

Use conditional expectation:

$$E(X) = \underbrace{E(X|X=1)}_{=1}\underbrace{P(X=1)}_{=p} + \underbrace{E(X|X>1)}_{=\,?\,=\,1+E(X)}\underbrace{P(X>1)}_{=1-p}.$$

$E(X|X > 1) = ?$

$= E(X \mid \text{First trial is failure})$

$= 1 + E(X)$

Knowing that first trial is failure increases the expected number of trials to get 1st head by 1.

Just thinking of starting over after the first trial.

Future coin tosses are all independent of the 1st coin toss.

$E(X) = 1 \cdot p + (1 + E(X)) \cdot (1-p)$

$\Rightarrow E(X) = p + (1-p) + (1-p) \cdot E(X)$

$= 1 + (1-p) \cdot E(X)$

$\Rightarrow p \cdot E(X) = 1$

$\Rightarrow E(X) = 1/p$

$E(X|X>1)$

$= E(X \mid \text{First failure})$

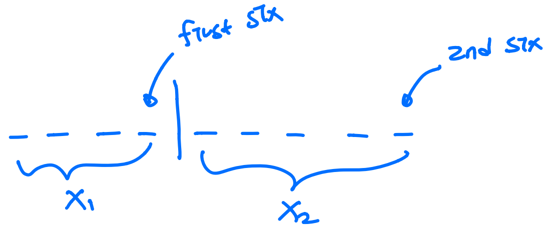
First success

F

$E(X|\boxed{X>1}) = 1 + E(X)$

Expected # trials until first success

Example: (Waiting time til 2 sixes) Let T_2 be the number of rolls of a die til a total of 2 sixes have appeared. Find $E(T_2)$.



$$T_2 = X_1 + X_2$$

$$X_1 \sim \text{Geom}(\tfrac{1}{6})$$

$$X_2 \sim \text{Geom}(\tfrac{1}{6})$$

$$\begin{aligned} E(T_2) &= E(X_1) + E(X_2) \\ &= 6 + 6 \\ &= 12 \end{aligned}$$

(a) A die is rolled repeatedly. Find the expected number of rolls till a total of 5 sixes appear.

(b) A die is rolled repeatedly. Find the expected number of rolls till two different faces appear.

(a) $T_5 = X_1 + X_2 + X_3 + X_4 + X_5$

$E(T_5) = E(X_1) + \cdots + E(X_5)$

$= 5 \cdot 6$

$= 30$

(b) anything

_ _ _ _ _

X = # rolls till any face $= 1$

$E(X) = 1$

Y = # rolls till different face after the first trial

$\sim \text{Geom}(\frac{5}{6})$

$E(X+Y) = E(X) + E(Y)$

$= 11/5$