STAT 88: Lecture 16

Contents

Probability and Statistics

Probability: A population distribution is known. We draw a sample X (or n many samples $X_1, \ldots, X_n$) 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

$X=x \backslash Y=y$	$y=1$	$y=2$
$x=1$	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 $X=Y$?

$\rightarrow P(X=Y)?$

Statistics: We draw n many samples $X_1, \ldots, X_n$ 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 $\frac{4}{N}$

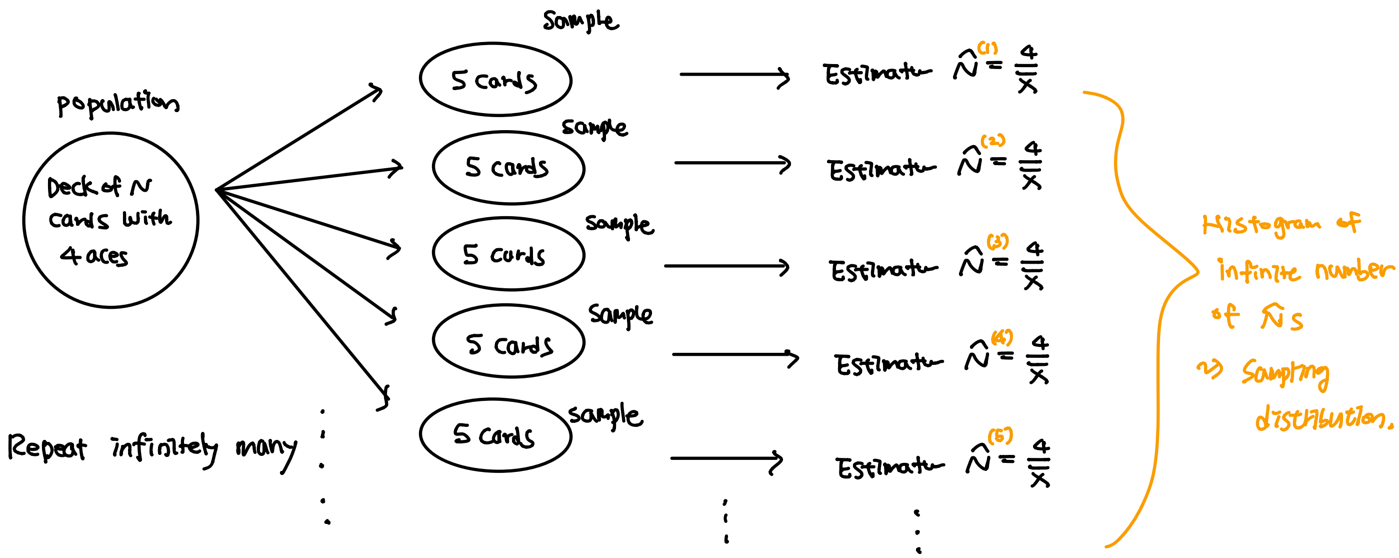
$\rightarrow \hat{N} = \frac{4}{\bar{X}}$

Joint distribution table

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

Draw $(X_1, Y_1), \ldots, (X_n, Y_n)$ 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.



$\hat{N}$ is unbiased if

$$E(\hat{N}) = N$$

$$\approx \frac{1}{m} \sum_{i=1}^{m} \hat{N}^{(i)} \text{ as } m \to \infty$$

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
Conditional distribution of X given $S = 3$	0.5	0.5	0	1.5
Conditional distribution of X given $S = 4$	0.1667	0.6667	0.1667	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	

We define

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

Then it can be shown that

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

$S=s \rightarrow E(X|S)$ 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:

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

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?

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?

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

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

P(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:

$$E(X) = E(X|X = 1)P(X = 1) + E(X|X > 1)P(X > 1).$$

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

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 $E(T_2)$.

Sec 5.5, 5.6 Practice

(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.

Example: A fair coin is tossed 3 times. Let

- X be the number of heads in the first two tosses;
- Y be the number of heads in the last two tosses.

Find $E(Y|X = 2)$.

Find $E(Y|X = 1)$.

Find $E(Y|X = 0)$.

Find $E(Y)$.