Stat 88: Probability & Statistics in Data Science

THE AUTHOR OF THE WINDOWS FILE COPY DIALOG VISITS SOME FRIENDS.

https://xkcd.com/612/

Lecture 12: 2/24/2022

Indicators, Unbiased estimators, Conditional expectation

Sections 5.3, 5.4, 5.5

Agenda

- 5.3
  - Method of indicators
- 5.4
  - Unbiased Estimators
- 5.5
  - Conditional expectation

Method of indicators

- Additivity of Expectation: This is a very useful property – no matter what the joint distribution of X and Y may be, we have:
$$E(X+Y) = E(X) + E(Y)$$
- Whether X and Y are dependent or independent, this additivity property holds, making it enormously useful.
- We also have linearity: $E(aX + bY) = aE(X) + bE(Y)$

- Recall that we talked about "classifying and counting" – that is, we divide up the outcomes into those that we are interested in (successes), and everything else (failures), and then count the number of successes.
- We can represent these outcomes as 0 and 1, where 1 marks a success and 0 and failure, so if we model the trials as **draws from a box**, we can count the number of success by counting up the number of times we drew a 1.

- We can represent each draw as a Bernoulli trial, where $p = P(S)$

Using indicators and additivity

- For example, say we roll a die 10 times, and success is rolling a 1.
- Then $p = 1/6$, and we can define a Bernoulli rv as $X = \begin{cases} 0, & w.p. 5/6 \\ 1, & w.p. 1/6 \end{cases}$
- We can also define an event A: let A be the event of rolling a 1 and define a rv $\boldsymbol{I_A}$ aka $\mathbf{1_A}$ that takes the value 1 if A occurs and 0 otherwise.
- This is a Bernoulli rv, what is its expectation (in terms of A)?
- Now let $X \sim Bin(10, \frac{1}{6})$, so X counts the number of successes in 10 rolls. Let's find $E(X)$ using additivity and indicators:

Using indicators to compute expected value

- Binomial

- Hypergeometric: Did we use the independence of the trials for the binomial? If not, we can use the same method to compute the expected value of a hypergeometric rv:

Using indicators

Exercise 5.7.6: A die is rolled 12 times. Find the expectation of:

a) the number of times the face with five spots appears

b) the number of times an odd number of spots appears

c) the number of faces that don't appear

d) the number of faces that do appear

Example

- Let X be the number of spades in 7 cards dealt **with replacement** from a well shuffled deck of 52 cards containing 13 spades. Find E(X).

1. Write down what X is
2. Define an indicator for the kth trial: I_k
3. Find $p = P(I_k = 1)$
4. Write X as a sum of indicators
5. Now compute $E(X)$ using additivity

- Do the same thing if we deal 7 cards **without replacement.**

Missing classes

- We can use indicators to compute the chance that something *doesn't* occur.

- For example, say we have a box with balls that are red, white, or blue, with 35% being red, 30% being white, and 35% blue. If we draw n times with replacement from this box, what is the expected number of colors that *don't* appear in the sample?

Examples

1. An instructor is trying to set up office hours during RRR week. On one day there are 8 available slots: 10-11, 11-noon, noon-1, 1-2, 2-3, 3-4, 4-5, and 5-6. There are 6 GSIs, each of whom picks one slot. Suppose the GSIs pick the slots at random, independently of each other. Find the expected number of slots that no GSI picks.

Examples

- A building has 10 floors above the basement. If 12 people get into an elevator at the basement, and each chooses a floor at random to get out, independently of the others, at how many floors do you expect the elevator to make a stop to let out one or more of these 12 people?

5.4 Unbiased Estimators

- We showed the linearity of expectation earlier: $\boldsymbol{E(aX+b)=aE(X)+b}$
- We often want to estimate a ***population parameter***: some fixed number associated with the population, possibly unknown
- A statistic is any number that is computed from the data sample. Usually we use a *random sample.*
- Note that the parameter is *constant* and the statistic is a *random variable*.
- We will use a ***statistic*** to *estimate* (guess at the value of; approximate) the parameter. It is called an ***estimator*** of the parameter.
- If the expectation of the statistic is the parameter that it is estimating, we call the statistic an ***unbiased estimator of the parameter***.

An example of an unbiased estimator: $\boldsymbol{E(\bar{X}) = \mu}$

- Let $X_1, X_2, \ldots, X_n$ be our random sample, and the sample mean is $\bar{X}$
- $\bar{X}$ is computed from the sample and will change depending on the sample values, so is a *random variable*.
- If $X_1, X_2, \ldots, X_n$ which are random draws from the population, all have expectation μ, what is the expectation of $\bar{X}$?

Understanding unbiased parameters

- Let $X_1, X_2, \ldots, X_n$ be random draws from the population, all have expectation μ.
- If an estimator S is unbiased, then ***on average,*** it is equal to the number it is trying to estimate
- Which of the following are unbiased estimators of μ?

(a) X_{15}

(b) $\frac{X_1+X_{15}}{15}$

(c) $\frac{X_1+2X_{100}}{3}$

(d) How to make an biased estimator unbiased?

(e) If $X_1$is unbiased, why bother taking the mean? Why not just use X_1?

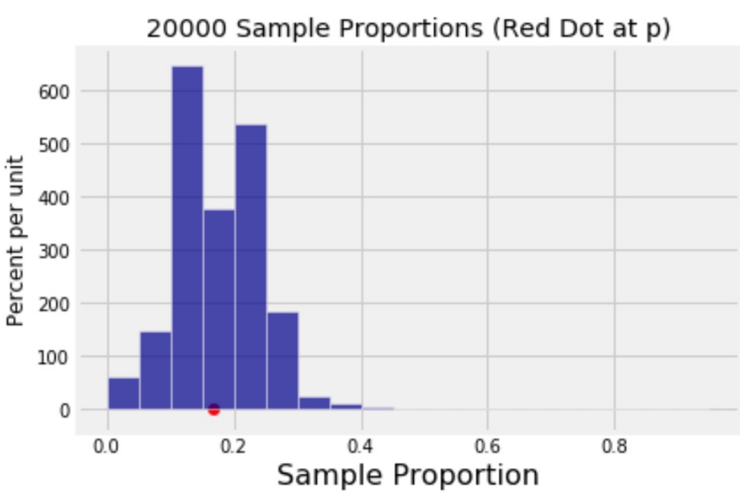
Understanding unbiased parameters

A special estimator: The sample proportion $\hat{p}$

- Usual special case of population binary outcomes represented by 0 and 1
- Sum of draws = # of 1s that are in the sample (sample sum)
- Sample mean = proportion of 1s in sample

- Consider a population of 0s and 1s, and draw n times from this population, **with** replacement: $X_1, X_2, \dots, X_n$ are the draws, note that each of the X_k are Bernoulli or indicator random variables, with parameter p where p = proportion of 1s in the population.

- Note that the population mean $\mu = p$ and the sample mean $\bar{X} = \hat{p}$, and $\bar{X}$ is an unbiased estimator of p



0.1666667
0.2666667
0.2333333
0.2000000
0.1000000

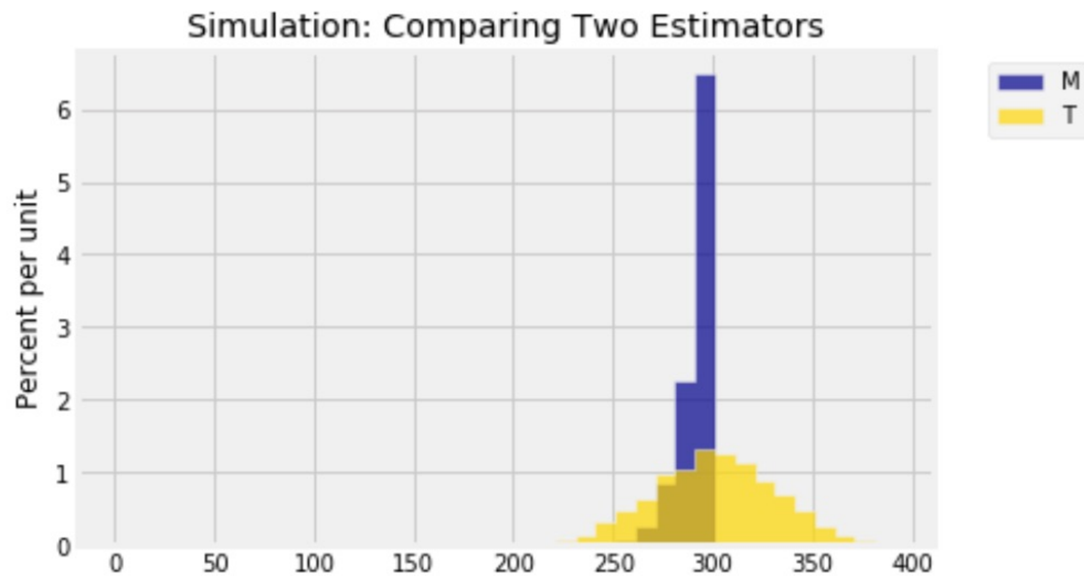
Estimating the largest possible value

- $X_1, X_2, \ldots, X_n$ are drawn at random with replacement from $\{1, 2, \ldots, N\}$. That is, they are independent and identically distributed random variables with the discrete uniform distribution on 1, 2, … N.

- We want to estimate N using an unbiased estimator. Does the sample mean work?

Comparing two estimators: T and M (max sample value)

- Let $X_1, X_2, \ldots, X_n$ be as earlier, and let $M = \max\{X_1, X_2, \ldots, X_n\}$. Below are histograms for M and $T = 2\bar{X} - 1$, from simulations assuming that N=300 and that the sample size is 30 (5,000 repetitions, computing T, M each time).



- pros & cons for M
- pros & cons for T

Example: (5.7.11)

A data scientist believes that a randomly picked student at his school is twice as likely not to own a car as to own one car. He knows that no student has three cars, though some students do have two cars. He therefore models the probability distribution for the number of cars owned by a random student as follows. The model involves an unknown positive parameter θ.

# of cars	0	1	2
Probability	2θ	θ	$1-3\theta$

(a) Find $E(X_k)$

(b) Let $X_1, X_2, \ldots X_n$ be the numbers of cars owned by n random students picked independently of each other. Assuming that the data scientist's model is good, use the entire sample to construct an unbiased estimator of θ.

Example: (5.7.11)

Conditional Expectation: An example

- Let X and Y be iid rvs with the distribution described below, and let $S = X + Y$:

x	1	2	3
$P(X = x)$	$1/4$	$1/2$	$1/4$

- Let's write down the joint distribution of X and S:

Conditional Expectation: An example