Stat 88: Probability & Statistics in Data Science

https://xkcd.com/612/

Lecture 13: 3/1/2022

Finishing up indicators, Unbiased estimators, Conditional expectation

Sections 5.3, 5.4, 5.5

Agenda

- Finish up examples of computing expectations using indicators
- 5.4
  - Unbiased Estimators
- 5.5
  - Conditional expectation
- 5.6
  - Introduce expectation by conditioning

Example 1

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?

(a) X = # of colors that don't appear in the sample (of n draws from box)

$= I_R + I_B + I_W$

$I_R = \begin{cases} 1 & \text{if } R \text{ does not appear} \\ 0 & \text{o/w} \end{cases}$

I_B, I_W are similarly defined

$\mathbb{E}(I_R) = P(R \text{ does not appear})$
$= (0.65)^n$

$\mathbb{E}(X) = (0.65)^n + (0.7)^n + (0.65)^n$

(b) $\mathbb{E}$(# of colors that do appear)

$= \mathbb{E}(3 - X) = 3 - \mathbb{E}(X)$

exercise

convince yourself that you can compute this directly.

Example 2

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.**

Let X = # of slots that are not picked by any GSI

$$X = I_1 + I_2 + \cdots + I_8$$

$$I_k = \begin{cases} 1, & \text{if } k^{th} \text{ slot is not picked} \\ 0, & \text{o/w} \end{cases}$$

A_k is the event that k^{th} slot is not picked.

$I_k \longleftrightarrow I_{A_k}$

$P(I_k)=1 = \left(\frac{7}{8}\right)^6$

$P(I_k=0) = 1-\left(\frac{7}{8}\right)^6$

$$E(I_k) = 1\cdot P(A_k) + 0\cdot P(A_k^c)$$

$$= 1\cdot\left(\frac{7}{8}\right)^6$$

P(a GSI picks any particular slot) $= \frac{1}{8}$

P(a particular slot is NOT picked by a GSI) $= \frac{7}{8}$

P(none of GSIs pick that slot) $= (7/8)^6$

$$E(X) = \left(\frac{7}{8}\right)^6 + \left(\frac{7}{8}\right)^6 + \cdots + \left(\frac{7}{8}\right)^6$$

$$= 8\cdot\left(7/8\right)^6$$

Example 3

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?

$$X = I_1 + I_2 + \cdots + I_{10}$$

$I_k = I_{A_k}$

A_k is the event that elevator stops at floor k & at least one person gets out

$$P(A_k) = 1 - P(\text{no one gets out at floor } k)$$
$$= 1 - \left(\frac{9}{10}\right)^{12}$$

$$I_k = \begin{cases} 1 & \text{if } A_k \text{ is true} \\ 0 & \text{o/w} \end{cases}$$

$$\mathbb{E}(I_k) = 1 \cdot P(A_k) + 0 \cdot P(A_k^c) = P(A_k) = 1 - \left(\frac{9}{10}\right)^{12}$$

$$\mathbb{E}(X) = \mathbb{E}(I_1 + I_2 + \cdots + I_{10}) = \sum_{k=1}^{10} \mathbb{E}(I_k) = 10 \cdot \left(1 - \left(\frac{9}{10}\right)^{12}\right)$$

5.4 Unbiased Estimators

- We showed the linearity of expectation earlier: $E(aX + b) = aE(X) + b$
- We often want to estimate a *population parameter*: some fixed number associated with the population, possibly unknown

sample average (sample mean)

- 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.

random variable that we use to guess the value 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***.

If θ is the population parameter.

and the statistic is X

$\mathbb{E}(X) = \theta$ then X is an unbiased estimator of θ.

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

- Let $X_1, X_2, \dots, X_n$ 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 $X_1, X_2, \dots, X_n$ which are random draws from the population, all have expectation μ, what is the expectation of $\bar{X}$?

$$\mathbb{E}(X_k) = \mu$$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_1 + X_2 + \dots + X_n)$$

$$= \frac{1}{n}\left[\mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)\right]$$

$$= \frac{1}{n} \cdot n \cdot \mu = \mu.$$

Understanding unbiased parameters

$\mathbb{E}(X_k) = \mu$ for each k.

- Let $X_1, X_2, \ldots, X_n$ be random draws from the population, all have expectation μ.

means "in expectation"

- If an estimator S is unbiased, then ***on average,*** it is equal to the number it is trying to estimate

$\mathbb{E}(S) = \theta$

- Which of the following are unbiased estimators of μ?

✓ (a) X_{15} $\quad \mathbb{E}(X_{15}) = \mu$

✗ (b) $\frac{X_1+X_{15}}{15}$ $\quad \mathbb{E}\left(\frac{X_1+X_{15}}{15}\right) = \frac{1}{15}\left(\mathbb{E}(X_1) + \mathbb{E}(X_{15})\right) = \frac{1}{15}(\mu + \mu) = \frac{2\mu}{15} \neq \mu$

(c) $\frac{X_1+2X_{100}}{3}$ $\quad \mathbb{E}\left(\frac{X_1 + 2X_{100}}{3}\right) = \frac{1}{3}\left(\mathbb{E}(X_1) + 2\mathbb{E}(X_{100})\right) = \frac{1}{3}(\mu + 2\mu) = \mu$

(d) How to make an biased estimator unbiased?

$\mathbb{E}\left(\frac{X_1+X_{15}}{15}\right) = \frac{2}{15}\mu$ so multiply $\frac{X_1+X_{15}}{15}$ by $\frac{15}{2}$ $\rightarrow \mathbb{E}\left(\frac{\cancel{15}}{2} \cdot \left(\frac{X_1+X_{15}}{\cancel{15}}\right)\right) = \mu$

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

$\mathbb{E}(X_1) = \mu$

A special estimator: The sample proportion $\hat{p}$ → $\bar{X}$ when all X_i's are 0-1

- Usual special case of population binary outcomes represented by 0 and 1
- Sum of draws = # of 1s that are in the sample (sample sum) $X_1 + X_2 + \cdots + X_n$ $= \#$ of X_i which are 1
- Sample mean = proportion of 1s in sample

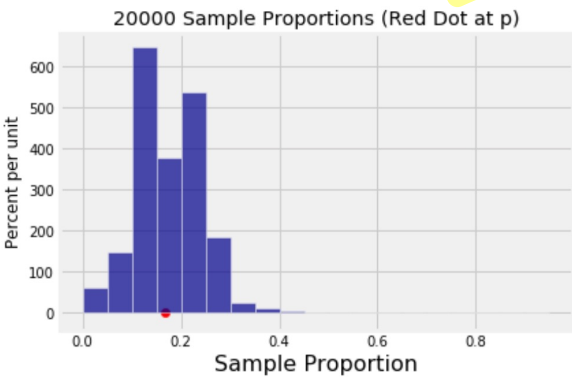
$$= \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\#\text{ of 1's in sample}}{\text{sample size}} = \text{fraction of successes.}$$

- 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.

Estimate p using $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ $\left[\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} \cdot np = p \right]$

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

$\mathbb{E}(\hat{p}) = p$



0.1666667
0.2666667
0.2333333
0.2000000
0.1000000

rolling a fair die looking at # of times get ⚀

← Read example in text

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?

$$\mathbb{E}(\bar{X}) \stackrel{?}{=} N$$

$$\mathbb{E}(\bar{X}) = \mu = \frac{N+1}{2}$$

$$\mu = \mathbb{E}(\bar{X}) = \frac{N+1}{2} \neq N$$

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

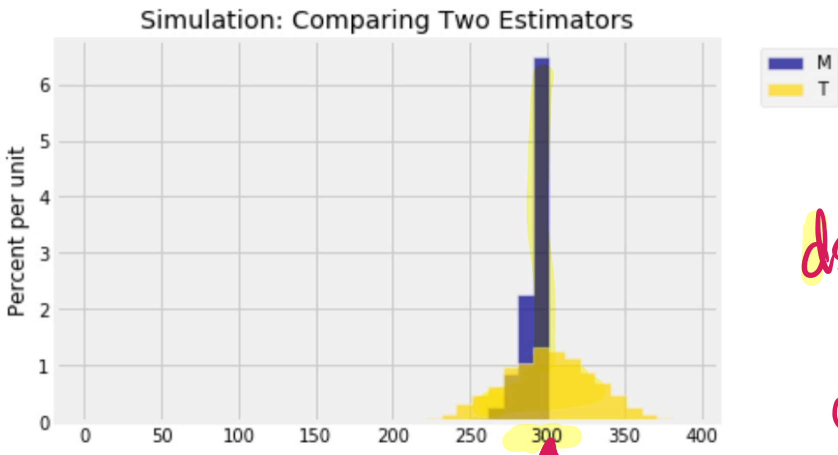
$$T = 2\bar{X} - 1$$

is an unbiased estimator of N.

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

unif. on $\{1, 2, \dots N\}$

- Let $X_1, X_2, \dots, X_n$ be as earlier, and let $M = \max\{X_1, X_2, \dots, X_n\}$. 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).



dsn of M is always to left of 300.

$\mathbb{E}(T) = N$
$\mathbb{E}(M) \leq N$

- pros & cons for M : not unbiased, but small spread.
- pros & cons for T : $\mathbb{E}(T) = N$ (pro)
  spread is large (con)

$\Big\}$ bias-variance tradeoff

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 $E(X)$, where X is the number of cars owned by a randomly selected student (the pmf is above).

(a) Let $X_1, X_2, \dots X_n$ 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.

Conditional Distributions: An example

- Suppose we have two rvs, V and W, and we have the joint dsn for these two rvs. Suppose we fix a value for W – call this value w – and compute, for each value of V, the probability $\boldsymbol{P(V = v|\ W = w)}$, then this set of probabilities, which will form a pmf, is called the ***conditional distribution of V, given $W = w$***.

- 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, and then compute the conditional dsn for X

S \ X	1	2	3
2	$P(X=1, S=2)$ $1/16$		
3			
4			
5			
6			

$S = X + Y$

X, Y are independent

Conditional distributions: An example

Expectation by Conditioning

- In the example we just worked out, once we fix a value s for S, then we have a distribution for X, and can compute its expectation using that distribution that depends on s: $E(\,X \mid S = s\,) = \sum x \cdot P(\,X = x \mid S = s\,)$, with the sum over all values of X.

- Note that $E(\,X \mid S = s\,)$ depends on S , so it is a ***function*** of s. We can think of $E(\,X \mid S\,)$ as a rv as it is a function of s and has a probability distribution on its values.

- This means that if we want to compute $E(X)$, we can just take a weighted average of these conditional expectations $E(\,X \mid S = s\,)$:

$$\boldsymbol{E(X) = \sum_{s} E(\,X \mid S = s\,)\,P(S = s)}$$

- This is called the *law of iterated expectation*

Law of iterated expectation

- Note that $E(X \mid S = s)$ is a function of s. That is, if we change the value of s we get a different value. (It is not a function of x, though since the x is summed out .)

- Therefore, we can define the function $g(s) = E(X \mid S = s)$, and the random variable $g(S) = E(X \mid S)$.

- In general, recall that $E\big(g(S)\big) = \sum_s g(s)f(s) = \sum_s g(s)P(S = s)$.
- How can we use this to find the expected value of the rv $g(S)$?

Examples from the text: Time to reach campus

- 2 routes to campus, student prefers route A (expected time =15 minutes) and uses it 90% of the time. 10% of the time, forced to take route B which has an expected time of 20 minutes. What is the expected duration of her trip on a randomly selected day?

Catching misprints

- The number of misprints is a rv $N \sim Pois(5)$ dsn. Each misprint is caught before printing with chance 0.95 independently of all other misprints. What is the expected number of misprints that are caught before printing?

Expectation of a Geometric waiting time

- $X \sim Geom(p)$: X is the number of trials until the first success
- $P(X = k) = (1-p)^{k-1}\, p, \quad k = 1, 2, 3, \ldots$
- Let $x = E(X)$
- Recall that $P(X > 1) = P(first\ trial\ is\ F) = 1 - p$
- We can split the possible situations into when the first trial is a success and the first trial is a failure, and condition on this and compute the *conditional expectation:*

$$E(X) = \mathbb{E}(\mathbb{E}(X \mid \text{first trial is } S)) = \mathbb{E}(X \mid 1^{st} \text{ trial} = S)\, P(1^{st} \text{ trial} = S) + \mathbb{E}(X \mid 1^{st} \text{ trial} = F)\, P(\underset{\uparrow\ 1^{st}\text{ trial}}{F})$$

$$E(X) = E(X \mid X = 1)P(X = 1) + E(X \mid X > 1)P(X > 1)$$

Expected value of a Geometric distribution

$X \sim Geom(p)$, $\mathbb{E}(X) = \frac{1}{p}$

Expected waiting time until k sixes have been rolled

- Let T_k be the waiting time until the kth six is rolled.
- For example, consider $k = 1$ (geometric) , expected waiting time=6 rolls
- What about if $k = 2$? $k = 3$? $k = 10$?

- General formula for expected waiting time until kth success where P(S) = p

Example

- A fair die is rolled repeatedly. Find the expected waiting time (number of rolls) until two *different* faces are rolled.