Stat 88: Probability & Math. Statistics in Data Science

Lecture 14: 3/3/2022

Conditional expectation, Expectation by conditioning, Variance

Sections 5.5, 5.6, 6.1, 6.2

https://xkcd.com/2586/

Agenda

- Warm up: Exercise 5.7.11
- Conditional distributions
- Conditional expectation
- Expectation by conditioning
- Variance definition
- Properties of Variance and SD

Warm up: Exercise 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	$1-3\theta$

$f(x) = P(X=x)$

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

$$\mathbb{E}(X) = \sum_x x \cdot f(x) = 0 \cdot 2\theta + 1 \cdot \theta + 2 \cdot (1-3\theta)$$

$$= \underline{2-5\theta} = \mu$$

$X_1, X_2, \ldots X_n$ are i.i.d.

$$\rightarrow \mathbb{E}(\bar{X}) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \cdot n\mu =}_{\text{by additivity \& linearity}} \mu = 2-5\theta$$

$\mathbb{E}(\bar{X}) = 2-5\theta$

& b/c X_i iid

Warm up: Exercise 5.7.11

(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.

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

We want a r.v. Y s.t $\mathbb{E}(Y) = \theta$ (unbiased estimator)

$$\mathbb{E}(\bar{X}) = 2 - 5\theta$$

$$5\theta = 2 - \mathbb{E}(\bar{X})$$

$$\theta = \frac{2 - \mathbb{E}(\bar{X})}{5}$$

$\boxed{Y \text{ is an unbiased estimator of } \theta \text{ that is a function of entire sample}}$

$$Y = \frac{2-\bar{X}}{5} \rightarrow \mathbb{E}(Y) = \mathbb{E}\left(\frac{2-\bar{X}}{5}\right) = \frac{1}{5}\left(2 - \mathbb{E}(\bar{X})\right) = \theta$$

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 $W = w$***.

$$\frac{P(V=v \,\&\, W=w)}{P(W=w)}, \quad P(W=w) \neq 0$$

- Let X and Y be iid (independent, and identically distributed) 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

Conditional distributions: An example : The joint dsn for X, S

$S = X + Y$

$S \downarrow$ \ $X \rightarrow$	1	2	3	Marginal dsn for S
2	$P(X=1, S=2) = P(X=1, Y=1)$ $\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$	0	0	$P(S=2) = 1/16$
3	$P(S=3, X=1)$ $\frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}$	$P(S=3, X=2)$ $\frac{1}{8}$	$P(S=3, X=3)$ 0	$P(S=3) = \frac{4}{16}$
4	$P(S=4, X=1)$ $= \frac{1}{16}$	$P(S=4, X=2) = P(X=2, Y=2) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$	$P(S=4, X=3)$ $P(X=3, Y=1)$ $= 1/16$	$P(S=4) = \frac{6}{16}$
5	0	$\frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}$	$\frac{1}{8}$	$P(S=5) = \frac{4}{16}$
6	0	0	$\frac{1}{16}$	$P(S=6) = \frac{1}{16}$
Marginal dsn for X	$P(X=1) = f_X(1)$ $= 4/16$	$P(X=2)$ $= 8/16$	$P(X=3)$ $= 4/16$	

	$f(x)$
$X = Y = $ 1	1/4
2	1/2
3	1/4

$$P(S=4, X=1) = P(X=1, Y=3) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$$

Conditional distributions: An example

- *Given* $S = 3$, what is $P(X = 1)$?

$$P(X=1 \mid S=3)$$

$$= \frac{P(X=1, S=3)}{P(S=3)} = \frac{1/8}{4/16} = \frac{1}{2}$$

- Write down the conditional distribution for X, given that $S = 3$

$P(X=1 \mid S=3)$	$P(X=2 \mid S=3)$	$P(X=3 \mid S=3)$
$1/2$	$1/2$	0

$$\mathbb{E}(X \mid S=3) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} + 3 \cdot 0 = 1.5$$

Conditional distributions: An example

- Write down the conditional distribution for X, given that $S = s$, for each possible value of S:

conditional probabilities for

Given:	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$	$E(X\|S = s)$
$S = 2$	1	0	0	$E(X\|S=1) = 1$
$S = 3$	$\frac{1}{2}$	$\frac{1}{2}$	0	$E(X\|S=2)$ 1.5
$S = 4$	$1/6$	$4/6$	$1/6$	$E(X\|S=3)$ $1\cdot\frac{1}{6}+2\cdot\frac{4}{6}+3\cdot\frac{1}{6}=2$
$S = 5$	0	$\frac{1}{2}$	$\frac{1}{2}$	$E(X\|S=5)$ $2\cdot\frac{1}{2}+3\cdot\frac{1}{2}=2.5$
$S = 6$	0	0	1	$E(X\|S=6)$ 3

$E(X|S)$ is a function of S.

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.

Suppose $g(S) = S^2$

$\mathbb{E}(g(S)) = \sum_s s^2 P(S=s)$

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

Let $g(S)$ be a function of S

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

$\mathbb{E}(g(S)) = \sum_s g(s) P(S=s)$

- This is called the *law of iterated expectation*

$$\mathbb{E}(\mathbb{E}(X|S)) = \mathbb{E}(X)$$

1 $\mathbb{E}(X) = \sum_x x \cdot P(X=x)$

$\left[\ \mathbb{E}(g(S)) = \sum_s g(s) P(S=s)\ \right]$

2 $\mathbb{E}(\underbrace{\mathbb{E}(X|S)}_{\text{function of } S}) = \sum_s (\mathbb{E}(X|S=s)) \cdot P(S=s)$

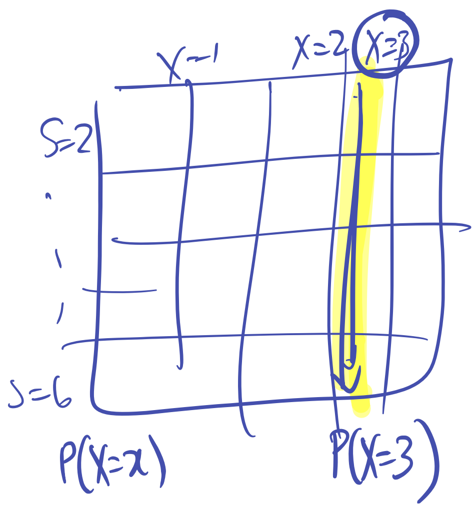
3 $= \sum_s \left[\sum_x x \cdot P(X=x|S=s) \right] P(S=s)$

$$P(X=x|S=s) = \frac{P(X=x, S=s)}{P(S=s)}$$

$$\mathbb{E}(\mathbb{E}(X|S)) = \sum_s \sum_x x \frac{P(X=x, S=s)}{\cancel{P(S=s)}} \cdot \cancel{P(S=s)}$$

$$= \sum_x x \sum_s P(X=x, S=s)$$

$$= \sum_x x P(X=x) = \mathbb{E}(X)$$



$$\mathbb{E}(\underbrace{\mathbb{E}(X|S)}_{\text{r.v.}}) = \mathbb{E}(X)$$

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?

$$S = \begin{cases} \text{route A} & \text{w.p } 0.9 \\ \quad \text{"} \quad B & \text{w.p. } 0.1 \end{cases} \qquad \begin{aligned} E(X \mid S = A) &= 15 \\ E(X \mid S = B) &= 20 \end{aligned}$$

$$E(X) = \mathbb{E}(\mathbb{E}(X \mid S))$$

$$= \mathbb{E}\Big(\sum_s \underbrace{\mathbb{E}(X \mid S = s)} P(S = s) \Big)$$

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

Catching misprints

$\mathbb{E}(N) = 5$

- 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?

$X = \#$ of misprints caught before printing

Suppose # of misprints $= 10$ then $X \sim Bin(10, 0.95)$

$$\mathbb{E}(X \mid n = 10) = (10)(0.95)$$

Suppose document has $N = n$ misprints, $X \sim Bin(n, 0.95)$

$$\mathbb{E}(X \mid N = n) = (0.95)(n)$$

$$\mathbb{E}(X) = \underset{\text{w.r.t } N}{\mathbb{E}}\big(\underset{\text{w.r.t } X}{\mathbb{E}(X \mid N)}\big) = \sum_{n=0}^{\infty} \boxed{\underbrace{\mathbb{E}(X \mid N = n)}_{np}} P(N = n)$$

$$= \sum_{n=0}^{\infty} n (0.95) \cdot P(N = n)$$

$$= (0.95) \underbrace{\sum_{n=0}^{\infty} n P(N = n)}_{\mathbb{E}(N) = 5}$$

$$\boxed{= 5(0.95)}$$

$=\mathbb{E}(N)=3$

Expectation of a Geometric waiting time

1st trial = Succ

- $X \sim Geom(p)$: 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:*

1st trial = Failure

$\mathbb{E}(X)=\mathbb{E}(X|☆)P(☆)+\mathbb{E}(X|☹)P(☹)$

$\mathbb{E}(X)=\mathbb{E}(\mathbb{E}(X|S))$ where S has 2 possible outcomes: first trial Succ. or first trial Failure

$$E(X)=\underline{E(X\mid X=1)}\,\underline{P(X=1)}+\underline{E(X\mid X>1)}\,\underline{P(X>1)}$$

Let $\mathbb{E}(X)=:c$. We need to solve for c.

$c=\quad 1\cdot p+\quad (c+1)(1-p)$

$\cancel{c}=p+(\cancel{c}+1)-p(c+1)$

$0=\cancel{p}+1-pc-\cancel{p}$

$pc=1 \quad \boxed{c=1/p}$

F ← starts over

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 $k = 2$? $k = 3$? $k = 10$?

$E(T_1) = 6$

$\underbrace{- - - - - S}_{E(T_1)=6} \mid - - -$

like starting over

$$E(T_2) = 6 + 6 = 12$$

$$E(T_{20}) = 20 \cdot E(T_1) = 20 \cdot 6$$

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

$$K \cdot \frac{1}{p} \qquad \left(E(T_1) = \frac{1}{p}\right)$$

Problems (5. 7. 14)

1. A fair die is rolled repeatedly.
   a. Find the expected waiting time (number of rolls) till a total of 5 sixes appear
   b. A fair die is rolled repeatedly. Find the expected waiting time (number of rolls) until two *different* faces are rolled.

Problems (similar to 5.7.5)

2. A fair coin is tossed 3 times. Let X be the number of heads in the first 2 tosses and Y the number of heads in the last two tosses.
    a. Find $E(Y \mid X = 2)$.
    b. Find $E(Y \mid X = 1)$
    c. Find $E(Y \mid X = 0)$
    d. Find $E(Y)$

Problems

3. Bella chooses an integer X uniformly at random from 1 to 425. She then chooses an integer Y uniformly at random from $1, \dots, X$. Find $E(Y)$.

4. Bella chooses an integer *N* uniformly at random from *Pois(μ)*. She then picks *N* cards from a deck with replacement. Find the expected number of ace cards.