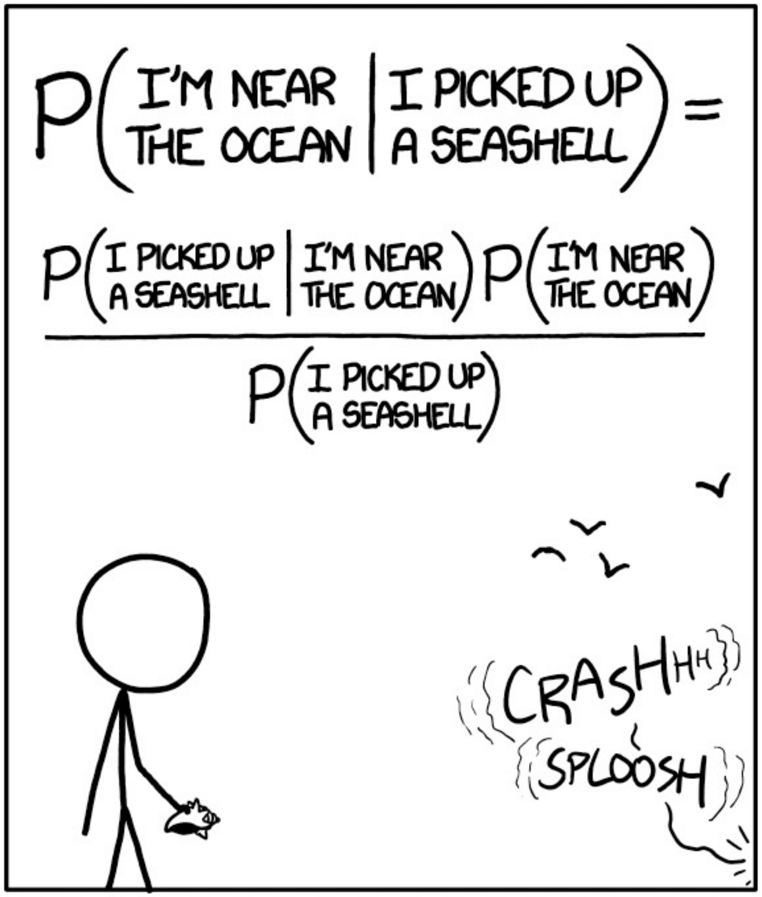

STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

https://xkcd.com/1236/

Stat 88: Probability & Math. Statistics in Data Science

Lecture 22: 3/11/2024

Conditional expectation, Expectation by conditioning, Variance

Sections 5.5, 5.6, 6.1, 6.2

Agenda

- Conditional distributions
- Conditional expectation
- Expectation by conditioning
- Variance definition
- Properties of Variance and SD

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 $\boldsymbol{P(V = v|\, W = w)}$ (using the division rule), then this set of probabilities, which will form a pmf, is called the ***conditional distribution of $\boldsymbol{V}$, given $\boldsymbol{W = w}$***.

- 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 given S.

Conditional distributions: An example

S \ X	1	2	3	Marginal dsn for S
2				
3				
4				
5				
6				
Marginal dsn for X				

Conditional distributions: An example

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

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

Conditional distributions: An example

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

Given: ↓	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$	$E(X\|S = s)$
$S = 2$				
$S = 3$				
$S = 4$				
$S = 5$				
$S = 6$				

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

- $E(X \mid S = s)$ is a function of s. That is, if we change the value of s we get a different value. (Note that it is *not* a function of x, 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 $g(s) = E(X \mid S = s)$?

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:*

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

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

Chapter 6: How to measure variability

- The expected value (μ_X) of a random variable X is a measure of *center.*
- Expectation is a weighted average indicating the center of the distribution of mass.
- How can we describe how the values taken by the random variable ***vary*** about this center of mass? How far does a typical value land from the center?

- The difference between the values X takes and the mean is called the deviation from the mean or the average: ($D = X - \mu_X$)

- We could take each value of X, see how far it is from μ_X, and compute the (weighted) average of this distance.
- Why weighted? Values that are more likely should be counted more.

Measuring variability

- Suppose that X is a rv that takes values -1 and 1 with equal probability.

- We know what a measure of the variability should be, let's see if it works.

- Write out $E(D) = E((X - E(X))$. What problem do you see?
- How can we fix it?

Variance of a random variable

- The *variance* of a random variable is defined by:

$$Var(X) = E(D^2) = E[(X - E(X))^2]$$

- Note that variance is the expectation of a function of X
- We could use the absolute deviation from the mean: $|D| = |X - \mu_X|$ but it isn't as nice a function as the square of the deviation from the mean.

- The only problem with using the variance is the units are off because we squared the deviation. In order to get the proper units back, we have to now take the square root.

Example

- Consider the random variable with distribution shown below:

x	1	2	3
$P(X = x)$	0.2	0.5	0.3
$(x - \mu_X)^2$			

- Find $E(X) = \mu_X$
- Write down the values of $(x - \mu_X)^2$ and find $E(X - \mu_X)^2$

Standard deviation of a random variable

- The *standard deviation* of a random variable is the ***square root of the variance*** of the random variable.

$$SD(X) = \sqrt{Var(X)} = \sqrt{E(D^2)} = \sqrt{E[(X - E(X))^2]}$$

- The variance is more convenient for computations because it doesn't have square roots. However, since the units are squared, it is difficult to interpret. Better to think about SD
- You can think of SD as the RMS deviation from the mean (***root-mean-square***)
- *Deviation* is the amount above or below the expected value. How big is it likely to be?
- The likely or typical size of the deviation is given by the *standard deviation.*
- SD is a *give-or-take* number telling us how far the values of X are from μ_X on average, that is, it gives us a measure of the variability of the random variable.

Shortcuts and alternative formulas

- If X is a random variable, and $E(X)$ is its mean
- Alternative formula for $Var(X)$
- $Var(X) = E\left(\left(X - E(X)\right)^2\right)$

$$= E\left(X^2 - 2 \cdot X \cdot E(X) + \left(E(X)\right)^2\right)$$

$$= \left(E(X^2) - 2E(X) \cdot E(X) + \left(E(X)\right)^2\right)$$

$$= E(X^2) - \left(E(X)\right)^2$$

Therefore, $SD(X) = \sqrt{E(X^2) - \left(E(X)\right)^2}$

Properties of Variance

1. Var(constant) = 0
2. $Var(X + c) = Var(X)$, where c is a constant

3. $Var(cX) = c^2 Var(X) \Rightarrow SD(cX) = |c| SD(X)$, thus $SD(-X) = SD(X)$

4. If X and Y are *independent*, then $Var(X \pm Y) = Var(X) + Var(Y)$

Problems

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.

Exercise 5.7.14 from the text

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 $E(Y)$

Exercises for Chocolate!

1. 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 $E(Y)$.

2. 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.