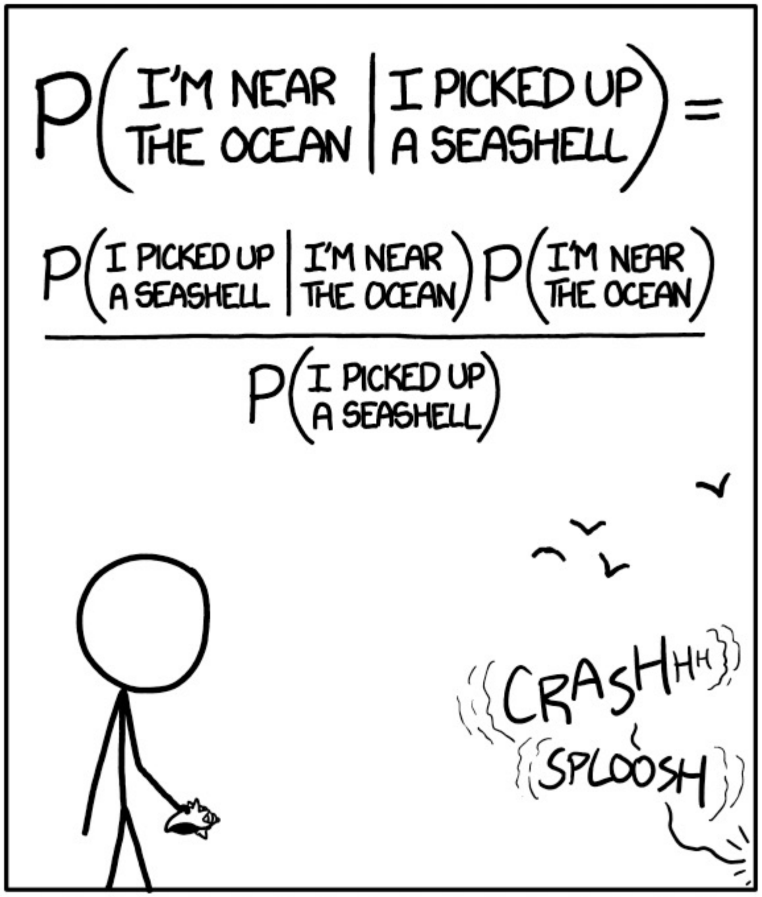

STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

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.

$S = X + Y$

$P(X=1, S=2) = P(X=1, Y=1)$

$X = Y = \begin{cases} 1 & \text{w.p. } 1/4 \\ 2 & \text{w.p. } 1/2 \\ 3 & \text{w.p. } 1/4 \end{cases}$

Conditional distributions: An example

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

Conditional distributions: An example

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

$$P(X=1 \mid S=3) = \frac{P(X=1, S=3)}{P(S=3)}$$

$$= \frac{2/\cancel{16}}{4/\cancel{16}} = \frac{1}{2}$$

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

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

→ NOTICE that this is a function of both X & S. If I change S, then pmf will change

$$\mathbb{E}(X \mid S=3) = 1\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{2}\right) + 0$$

$$= 1.5$$

$$P(X=1 \mid S=2) = \frac{P(X=1, S=2)}{P(S=2)} = \frac{1/\cancel{16}}{1/16} = 1$$

Conditional distributions: An example

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

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

conditional exp. is a function of S, NOT X

$$P(X=1|S=4) = \frac{P(X=1, S=4)}{P(S=4)} = \frac{1/16}{6/16} = \frac{1}{6}$$

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: $E(X \mid S = s) = \sum x \cdot \underbrace{P(X = x \mid S = s)}_{\text{conditional pmf of } X \text{ given } S=s}$, with the sum over all values of X.

- Note that $E(X \mid S = s) = g(s)$ 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.

E(final score in Data 88S | # of hours of study)

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

$$\boxed{\mathbf{E(X) = \sum_s \overset{g(s)}{E(X \mid S = s)}\, P(S = s)}}$$

- This is called the *law of iterated expectation*

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

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 $\boxed{g(s) = E(X \mid S = s)}$, 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 $g(s) = E(X \mid S = s)$?

$$E(X) = \sum_x x \cdot P(X=x)$$

$$E\big(\underbrace{E(X \mid S)}_{g(S)}\big) = E\big(g(S)\big) = \sum_s g(s) P(S=s)$$

$$g(s) = E(X \mid S=s)$$

$$= \sum_s E(X \mid S=s) P(S=s)$$

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

$$= \sum_{s} \left[\sum_{x} x \cdot \frac{P(X=x, S=s)}{\cancel{P(S=s)}} \right] \cancel{P(S=s)}$$

$$= \sum_{x} \sum_{s} x \cdot P(X=x, S=s)$$

$$= \sum_{x} x \underbrace{\sum_{s} P(X=x, S=s)}$$

$$= \sum_{x} x \cdot P(X=x)$$

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

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

Examples from the text: Time to reach campus

$X =$ duration of trip.

- 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?

$$S = \begin{cases} \text{route A} & \text{w.p. } 0.9 \\ \text{route B} & \text{w.p. } 0.1 \end{cases}$$

$$E(X \mid S = A) = 15 \text{ minutes}$$

$$E(X \mid S = B) = 20 \text{ minutes}$$

$$E(X) = (0.9)(15) + (0.1)(20)$$

$$= E_s(E_x(X \mid S)) = \sum_s \left(\sum_x x \cdot P(X = x \mid S = s) \right) P(S = s)$$

Catching misprints

$E(N) = 5$

- 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 we have 10 misprints

$X \sim Bin(n=10, p=0.95)$

Sp we have n misprints $X \sim Bin(n, 0.95)$

$E(X \mid n) = (0.95)(n)$

$E(X \mid N=n) = 0.95n$

$$E(X) = \underset{\substack{\uparrow \\ N}}{E}\Big(\underbrace{\underset{\substack{\uparrow \\ X}}{E}(X \mid N)}_{g(N)} \Big) = \sum_{n=0}^{\infty} P(N=n) E(X \mid N=n)$$

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

$n=0$

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

$$= (0.95)(5)$$

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:*

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