Stat 88: Probability & Mathematical Statistics in Data Science

I HATE FEELING DESPERATE ENOUGH TO VISIT THE SECOND PAGE OF GOOGLE RESULTS.

https://xkcd.com/1334/

Lecture 14: 2/22/2021

Method of indicators

Sections 5.2, 5.3

Joint distributions

- Draw two tickets without replacement from a box that has 5 tickets marked 1, 2, 2, 3, 3. Let X_1 and X_2 represent the values of the tickets drawn on the first and second draws respectively.

	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$	Marginal dsn for X_1
$X_1 = 1$	0	2/20	2/20	4/20
$X_1 = 2$	2/20	2/20	4/20	8/20
$X_1 = 3$	2/20	4/20	2/20	8/20
	4/20	8/20	8/20	

Marginal dsn for X_2

$g(X_1, X_2) = S = X_1 + X_2$

$S = X_1 + X_2$

- $S = X_1 + X\_2$, find $E(S)$

$$E(S) = \sum_{(x_1,x_2)} g(x_1,x_2)\, f(x_1,x_2) = g(1,1)\cdot f(1,1) + g(1,2) f(1,2) + \cdots + \cdots + g(3,3) f(3,3)$$

$$= 2\cdot 0 + 3\cdot\frac{2}{20} + 4\cdot\frac{2}{20} + 3\cdot\frac{2}{20} + 4\cdot\frac{2}{20} + 5\cdot\frac{4}{20} + 4\cdot\frac{2}{20} + 5\cdot\frac{4}{20} + 6\cdot\frac{2}{20}$$

$$= \frac{2}{20}(24) + \frac{4}{20}(10) = \frac{48+40}{20} = \frac{88}{20}$$

$f(x_1,x_2) = P(X_1 = x_1, X_2 = x_2)$

$$\mathbb{E}(X_1) = \sum_x x \cdot P(X_1 = x)$$

$$P(X_1 = x) = f_{X_1}(x)$$

$$= 1 \cdot \frac{4}{20} + 2 \cdot \frac{8}{20} + 3 \cdot \frac{8}{20} = \frac{44}{20}$$

$$\mathbb{E}(X_2) = \frac{44}{20}$$

$$\mathbb{E}(X_1) + \mathbb{E}(X_2) = \frac{88}{20} = \mathbb{E}(S) = \mathbb{E}(X_1 + X_2)$$

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Additivity of expectation.

Warm up and review

- A joint distribution for two random variables, M and S, is given below. Find $E(M)$.
- Are M and S independent?

	$M = 2$	$M = 3$	$f_S(x)$
$S = 2$	0	$1/3$	$1/3$
$S = 3$	$1/3$	$1/3$	$2/3$
$f_M(x)$	$1/3$	$2/3$	

$$E(M) = 2 \cdot \frac{1}{3} + 3 \cdot \frac{2}{3} = \frac{2}{3} + \frac{6}{3} = \frac{8}{3}$$

$$f(x,y) \overset{??}{=} f_S(x) \cdot f_M(y), \quad P(M=2, S=2) \overset{??}{=} P(M=2) \cdot P(S=2)$$

$f(x,y)$: joint p.m.f of M & S

$P(M=2, S=2) = 0 \neq P(M=2) \cdot P(S=2) = 1/9$

M & S ARE NOT INDEPENDENT

$E(aX) = aE(X)$

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 holds, making it enormously useful.

$$= E(aX) + E(bY) = aE(X) + bE(Y)$$

- We also have linearity: $E(aX + bY) = aE(X) + bE(Y)$

$$E(aX + bY + c) = aE(X) + bE(Y) + c \neq aE(X+c) + bE(Y+c)$$

- Recall that we talked about "classifying and counting" – so, 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 $p = P(S)$

Each draw defines an indicator rdm variable

$$I_k = \begin{cases} 1 & \text{if } k^{th} \text{ draw is } S \\ 0 & \text{otherwise.} \end{cases}$$

Using indicators and additivity

$S = $ ⚀ $\quad P(S) = 1/6$

- 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 $X = \begin{cases} 0, & w.p. 5/6 \\ 1, & w.p. 1/6 \end{cases}$

Let $A = $ ⚀

- We can also define an event A: let A be the event of rolling a 1 and define a rv I_A that takes the value 1 if A occurs and 0 otherwise.
- This is a Bernoulli rv, what is its expectation?

$I_A = \begin{cases} 1, & \text{if } A \text{ true} \\ 0, & o/w \end{cases}$

$$\mathbb{E}(I_A) = 1 \cdot P(A) + 0 \cdot P(A^c)$$
$$= 1 \cdot P(A) = 1/6$$

$$\boxed{\mathbb{E}(I_A) = P(A)}$$

- 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:

X is the # of ⚀'s in 10 rolls.

$A_1 = $ ⚀ is rolled on 1^{st} roll.

$A_2 = $ ⚀ " " 2^{nd} roll

$A_K = $ ⚀ " " K^{th} roll

$1 \leq K \leq 10$

$I_{A_K} = \begin{cases} 1, & \text{if } A_K \text{ true} \\ 0 & o/w \end{cases}$

$$X = I_{A_1} + I_{A_2} + I_{A_3} + \dots + I_{A_{10}}$$

$\mathbb{E}(X) = \sum_x x \cdot f(x)$ $\quad \mathbb{E}(X) = \mathbb{E}(I_{A_1} + I_{A_2} + \cdots + I_{A_{10}}) = \mathbb{E}(I_{A_1}) + \mathbb{E}(I_{A_2}) + \cdots + \mathbb{E}(I_{A_{10}})$

$= 1/6 + 1/6 + \cdots + 1/6 = 10\left(\frac{1}{6}\right)$

Using indicators

$f(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

- Binomial — In general, if $X \sim Bin(n,p)$, $X = I_{A_1} + I_{A_2} + \cdots + I_{A_n}$

  $\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(I_{A_k}) = \sum_{k=1}^{n} p = n \cdot p.$

- 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: $X \sim HG(N, G, n)$

$P(S \text{ on } 1^{st} \text{ trial}) = \frac{G}{N}$

$P(S \text{ on the } 5^{th} \text{ trial}) = \frac{G}{N}$

b/c of symmetry

$X = I_{A_1} + I_{A_2} + \cdots + I_{A_n} \Rightarrow \mathbb{E}(X) = \mathbb{E}(I_{A_1}) + \cdots + \mathbb{E}(I_{A_n})$

$\mathbb{E}(X) = n \cdot G/N$

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

(a) Let X be the # of times ⁙, $X \sim Bin\left(12, \frac{1}{6}\right)$

$\mathbb{E}(X) = n \cdot p = 12 \cdot \frac{1}{6} = 2$

A_1: 5 rolled on 1^{st} roll, A_2: 5 rolled on 2^{nd}.

$I_{A_1} = \begin{cases} 1 & \text{if } A_1 \text{ true} \\ 0 & \text{if not} \end{cases}$ $\qquad \underline{E(I_{A_1}) = P(A_1)} = 1/6$

$X = I_{A_1} + I_{A_2} + \ldots\ldots + I_{A_{12}}$

$E(X) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{12}) = 12 \cdot \frac{1}{6} = 2$

$P(\text{odd spot on any roll}) = \frac{1}{2}$, $X \sim Bin(12, \frac{1}{2})$

b/c die rolls are indep.

$E(X) = n \cdot p = 12 \cdot \frac{1}{2} = 6$

$E(\#\text{ of faces that don't appear})$

$X = \#$ of faces that don't appear.

Let A_k be the event that k^{th} face appears so $k = 1, 2, \ldots 6$

Let $I_{A_k} = \begin{cases} 1 & \text{if } k^{th} \text{ face appears} \\ 0 & \text{otherwise.} \end{cases}$

$X = I_{A_1} + I_{A_2} + I_{A_3} + I_{A_4} + I_{A_5} + I_{A_6}$

$E(X) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{A_6})$

$E(I_{A_1}) = \left(\frac{5}{6}\right)^{12}$ $\qquad P(\boxdot \text{ doesn't appear on any of 12 rolls})$

$=$

$E(X) = 6 \cdot \left(5/6\right)^{12}$

For any face $P(\boxdot$ never appears$) = \left(\frac{5}{6}\right)^{12}$

$P(\boxdot$ appears at least once$) = 1 - \left(\frac{5}{6}\right)^{12}$.

- X = # of faces that do appear

A_K = event that K^{th} face appears at least once.

$P(A_K) = 1 - \left(\frac{5}{6}\right)^{12}$

$\mathbb{E}(I_{A_K}) = 1 - \left(\frac{5}{6}\right)^{12}$

$$\mathbb{E}(X) = \mathbb{E}(I_{A_1} + I_{A_2} + \cdots + I_{A_6})$$

$$= \mathbb{E}(I_{A_1}) + \cdots + \mathbb{E}(I_{A_6})$$

$$= 6\left[1 - \left(\frac{5}{6}\right)^{12}\right]$$

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

   X = # of spades in hand of 7

2. Define an indicator for the kth trial: I_k

   $k = 1, 2, \dots 7$

   $$I_k = \begin{cases} 1 & \text{if } k^{th} \text{ draw is spade} \\ 0 & \text{if not} \end{cases}$$

3. Find $p = P(I_k = 1)$

   $P(I_k = 1) = \frac{13}{52}$ for each K. (dealing w/repl)

   $= E(I_k)$

4. Write X as a sum of indicators

   $$X = I_1 + I_2 + \dots + I_7$$

   I_1 is indicator for A_1 = spade on first draw

5. Now compute $E(X)$ using additivity

   $$E(X) = E(I_1) + \dots E(I_7) = 7 \cdot \frac{13}{52} = 7/4$$

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

  Same as above b/c of symmetry.

Missing classes (from book)

- 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?

$$I_{NR} = \begin{cases} 1 & \text{if no red balls drawn} \\ 0 & \text{o/w} \end{cases}$$

X = # of colors that don't appear w/ n draws

I_{NW}, I_{NB}

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.

X = # of slots not picked.

2. 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?