Dis 5C: Random Variables

Thursday, 23 July 2020   10:43 AM

Was going to introduce RVs by establishing formal definitions first ... changed my mind

> Ah yes, this one is tricky too. At least in my experience, most students seem to have an extremely hard time parsing the formal definition, since it's so stripped of any probability.

> I usually don't provide the definition for a good while, and try to have them discover it for themselves through a few of the examples: they've become friends with over time.

> E.g. on the campus map, I'd take some quantity of interest (say, number of water fountains within a 20-meter radius, or number of empty tables in the [illegible]) and have them try and say as much as they can about this quantity.

> Usually, they arrive quite quickly at the fact that this is a random number, that it's a function of the outcome (so we've got the definition covered), and that really the right way to think about random variables is as a convenient tool to partition your sample space.

> (Sometimes, when students are still scared of random variables, I take my class as a whole as the sample space, and have them labelled by random variables like gender, age, year of study etc. That seemed to get them somewhere...)

Example: Three people ask Jonas the question "*will you marry me?*" Assume Jonas always says "yes" with 50% probability to the first person, 25% to the second person and 12.5% to the third person. Also assume the people who ask are indistinguishable. You and I are interested in the number of people Jonas says "yes" to. X

Random Experiment:
Sample Space: $\Omega = \{YYY, YYN, YNY, YNN, NYY, NYN, NNY, NNN\}$

$X = 1$

$\Pr(X=1) =$

$X=2$

$X=3$

$X=0$

Probability Space:
1. Sample Space Ω
2. Probability Function (assigns probabilities to events)

$P(YYY) = \frac{1}{2}\times\frac{1}{4}\times\frac{1}{8}$

Random Variable, let's call it X:

Can you write X in terms of indicator random variables? Let X_i ind. RV for whether Jonas says 'yes' to person i

$X = X_1 + X_2 + X_3$

E[X] using formula (without indicator RVs)?

$\boxed{E(X)} = \sum_{a \in A} a\, P(X=a)$ $\quad X(\omega) = a$ $\quad A = \{0,1,2,3\}$

$= 0 \times P(X=0) + 1\times P(X=1) + 2\times P(X=2) + 3\times P(X=3)$

E[Y] using indicator random variables?

$E(X) = E(X_0 + \dots + X_3) = \boxed{E(X_0)} + E(X_1) + E(X_2) + E(X_3) =$

You might be wondering why you'd split Y into indicators. For this questions, there's no need. If / when you get stuck calculating E[Y] using the expectation formula, you'll see the power of indicators. HINT for questions later on: try indicator random variables.

$$X_0 = \begin{cases} 1 & \text{wp } \frac{1}{2}\times\frac{3}{4}\times\frac{7}{8} \\ 0 & \text{wp } 1-\frac{1}{2}\times\frac{3}{4}\times\frac{7}{8} \end{cases}$$

$\boxed{E(X_0) = \Pr(X_0 = 1)}$

$= \cancel{0 \times \Pr(X_0=0)} + 1\times\Pr(X_0=1)$

Formal definition
A random variable:

- Is a **function**
- Is defined on a **sample space**
- Has a **set of values it can take on** that's usually a subset of R, rather than R itself
- Has an associated **probability distribution**
- Think of it as "... a convenient tool to partition your sample space" - Dan

Definition 14.5 (Expectation). *The expectation of a discrete random variable X is defined as*

$$E[X] = \sum_{a \in A} a \times \Pr[X=a] \qquad (4)$$

where the sum is over all possible values taken by the r.v.

Theorem 14.1. *For any two random variables X and Y on the same probability space, we have*

$$E[X+Y] = E[X] + E[Y].$$

Also, for any constant c, we have

$$E[cX] = cE[X].$$

Indicator Random Variables

A convenient tool to calculate expectations of random variables. Use them when you can separate the 'original' random variable into disjoint 'units' of random variables with 'yes' no 'answers'.

3 How Many Queens? $\quad (!Q\,!Q\,Q),\ (!Q, Q, !Q),\ (Q, !Q, Q)$

You shuffle a standard 52-card deck, before drawing the first three cards from the top of the pile. Let X denote the number of queens you draw.

(a) What is $\Pr[X=0]$?

$= \frac{|X=0|}{|\Omega|} = \frac{\binom{48}{3}}{\binom{52}{3}}$ $\qquad \frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50}$

(b) What is the probability that you draw queens in positions 1 and 2? $\quad \Pr(X=0) = \Pr(1^{st}\ 3 \text{ aren't } Q)$

$= \Pr(1^{st} \text{ isn't } Q \cap 2^{nd} \text{ isn't } Q \cap 3^{rd} \text{ isn't } Q)$

(c) Compute $E[X]$ from the definition of expectation.

$= \sum_{a \in \{0,1,2,3\}} a \times \Pr(X=a) = 0\times P(X=0) + 1\times P(X=1) + 2\times P(X=2) + 3\times P(X=3) = \Pr(3^{rd}\ !Q \mid 1^{st}\ !Q \cap 2^{nd}\ !Q)\ \Pr(1^{st}\ !Q \cap 2^{nd}\ !Q)$

(d) Suppose we define indicators X_i, $1 \le i \le 3$, where X_i is the indicator variable that equals 1 if the ith card is a queen and 0 otherwise. Compute $E[X]$ using linearity of expectation.

$E(X) = E(X_1 + X_2 + X_3)$

$= E(X_1) + E(X_2) + E(X_3) = \Pr(X_1=1) + \Pr(X_2=1) + \Pr(X_3=1)$

(e) Are the X_i indicators independent? Does this affect your solution to part (d)?

X_2 is true $=1$

$\Pr(X_2 = 1 \mid X_1 = 1) \neq \Pr(X_2 = 1)$

$\boxed{E(X_1)} = \sum_{a \in A = \{0,1\}} a\, P(X_1 = a)$

$= \cancel{0 \cdot P(X_1=0)} + 1\times P(X_1=1)$

$= \boxed{P(X_1=1)}$

4 Linearity

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

CS 70, Summer 2020, Dis 5C

$\Pr(X_2 = 1 \mid X_1 = 1) \neq \Pr(X_2 = 1)$

events A, B are indep. if

$\Pr(A \cap B) = \Pr(A)\Pr(B)$

$\Leftrightarrow$

$\Pr(A \mid B) = \Pr(A)$

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability 1/3 (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability 1/5, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

$\boxed{T = 10A + 20B}$

$$A = \begin{cases} 3 & \text{wp } \frac{1}{3} \\ 0 & \text{wp } \frac{2}{3} \end{cases} \qquad B = \begin{cases} 4 & \text{wp } \frac{1}{5} \\ 0 & \text{wp } \frac{4}{5} \end{cases}$$

$\boxed{E(T) = E(10A + 20B)}$

$= 10\,E(A) + 20\,E(B)$

$= 10 \cdot 1 + 20 \cdot \frac{4}{5}$

$\Pr(X_2 = 1) = \frac{4}{52}$

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possible letters. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears?

recall the way to find E(X) if using formula straight up doesn't work...

Let B_i be the ind. RV for whether, say, 'book' appears on ith letter

$\boxed{T = B_1 + B_2 + B_3 + \dots + B_{1000000-3}}$

B_i = indicator RV for whether 'book' appears on letters $(i, i+1, i+2, i+3)$

$E(B_i) = E(B_j) \quad i \neq j$

$E(T) = \sum_{i=1}^{1mil-3} E(B_i) = (1mil - 3)\, E(B_i)$ $\quad \underline{b}\ \underline{o}\ \underline{o}\ \underline{k}\ \underline{z}$

$= (1mil-3) \times \left(\frac{1}{26}\right)^4 = \boxed{}$

in others $(i, i+1, i+2, i+3)$

$E[B_i] = E[B_j]$ $(\forall j)$

$$E[T] = \sum_{i=1}^{|m|-3} E[B_i] = (|m|-3)\, E[B_1] \quad \underline{b}\ \underline{o}\ \underline{o}\ \underline{k}\ \underline{z}$$

$$= \text{"} \times \left(\tfrac{1}{26}\right)^4 = \boxed{}$$

[In case we don't get here] This Q's about probability distributions of RVs: super fundamental to the language of statistics. Left later because it's easier to learn by yourself.

1 Head Count

Consider a coin with $\mathbb{P}(\text{Heads}) = 2/5$. Suppose you flip the coin 20 times, and define X to be the number of heads.

(a) Name the distribution of X and what its parameters are.

(b) What is $\mathbb{P}(X = 7)$?

(c) What is $\mathbb{P}(X \geq 1)$? Hint: You should be able to do this without a summation.

(d) What is $\mathbb{P}(12 \leq X \leq 14)$?

2 Family Planning

Mr. and Mrs. Brown decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Browns have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}(G=0)$	
$\mathbb{P}(G=1)$	

$\mathbb{P}(C=1)$	$\mathbb{P}(C=2)$	$\mathbb{P}(C=3)$

(d) Are G and C independent?

(e) What is the expected number of girls the Browns will have? What is the expected number of children that the Browns will have?

① Defining a 'good' rand. exp.

n blue balls
m red balls   randomly draw balls out of a bag

Q: Whats $\Pr(\underbrace{\text{6th ball is blue}}_{B_6}) = \dfrac{|\text{6th ball is blue}|}{|\Omega|}$

$$\Pr(A) = \frac{|A|}{|\Omega|}$$

$$\omega = \underbrace{bbbrbrbr \ldots brr}_{n+m}$$

$$\Pr(\underbrace{\text{5th ball is B}}_{B_5}) = \frac{|\text{5th ball is B}|}{|\Omega|}$$

$\Pr(B) =$ if $|A| = |B|$

$\to \Pr(B_i) = \Pr(B_j)$ $\forall i,j$

$\Leftrightarrow$ $\boxed{|i^{th} \text{ ball is B}| = |j^{th} \text{ ball is B}|}$

i 4 j

1 4

$$\Pr(B_4) = \frac{\binom{n+m-1}{n-1}}{\binom{n+m}{n}}$$

B _ _ _ _ _ _ _ _ _

4

$n-1$ $\underbrace{n+m-1}$ $\binom{n+m-1}{n-1}$

$|B_4| = |B_1| = |B_2| = |B_3| = |B_5| \ldots$

$- - - \; B \; \underbrace{- \ldots -}_{\binom{n+m-1}{n-1}}$

$$\boxed{\Pr(B_1|D)} = \frac{\Pr(B_1 \cap D)}{\Pr(D)} = \frac{\Pr(D|B_1)\Pr(B_1)}{\Pr(D)}$$

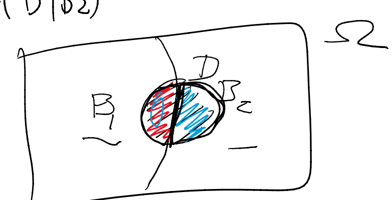
obvious :)

← non obvious :c

$\Pr(B_1)$, $\Pr(B_2)$ Pr(

$\Pr(D|B_1)$, $\Pr(D|B_2)$

Law of ttl prob :



$$\Pr(D) = \underline{\Pr(D \cap B_1)} + \Pr(D \cap B_2)$$

a)

b) $$Pr(B_1 \mid DD) = \frac{Pr(DD \mid B_1)\, Pr(B_1)}{P(DD) = \boxed{Pr(DD \mid B_1)\, Pr(B_1) + Pr(DD \mid B_2)\, Pr(B_2)}}$$

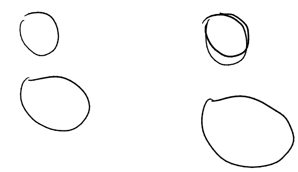
$$0.1 \times \frac{99}{999}$$

$$Pr_1(a_1 \mid a_1 a_2) + Pr_1(a_2 \mid a_1 a_2) = 1 \quad \text{vs} \quad Pr_1(a_1 \mid a_1 a_2) + Pr_2(a_1 \mid a_1 a_2) = 1$$

$$\underbrace{Pr_1(a_1)}_{=0.2} \neq \underbrace{Pr_2(a_1)}_{=0.3}$$

$$\underbrace{Pr_1(a_2)}_{0.4} \neq \underbrace{Pr_2(a_2)}_{0.5}$$

P1 P2



$Pr_1($

$$P_1(w \mid A) = \frac{Pr_1(A \mid w)\, Pr_1(w)}{Pr_1(A)} \overset{?}{=} P_2(w \mid A) = \frac{Pr_2(A \mid w)\, Pr_2(w)}{Pr_2(A)}$$

$Pr(a_1$

$$Pr_1(\overset{a_1 \cap a_2}{A}) \quad \text{vs} \quad Pr_2(\overset{a_1 \cap a_2}{A})$$

$$Pr_1(a_1 \mid a_1 \cup a_2) = Pr_2(a_1 \mid a_1 \cup a_2)$$

$Pr(a_1 \mid$

$Pr($

$$Pr_1(a_1) \text{ vs } Pr(a_2) \qquad Pr_1(a_1 \cup a_2)$$

$$Pr_1(a_1 \mid a_1 \cup a_2) = Pr_2(a_2 \mid a_1 \cup a_2)$$

$$= \frac{Pr_1(a_1 \cap (a_1 \cup a_2))}{Pr_1(a_1 \cup a_2)} = \frac{Pr_2(a_2 \cap (a_1 \cup a_2))}{Pr_2(a_1 \cup a_2)}$$

$Pr($

$$a_1 \neq a_2 \qquad = \frac{Pr_1(a_1)}{Pr_1(a_1 \cup a_2)} = \frac{Pr_2(a_2)}{P_2(a_1 \cup a_2)} = P_2($$

$$\frac{Pr_1(a_1)}{Pr_2(a_2)} = \frac{Pr_1(A)}{Pr_2(A)} \qquad \frac{Pr_1(a_1) + P_1(a_2)}{Pr_1(a_1 \cup a_2)} = \frac{P_2(a_1) + P_2(a_2)}{Pr_2(a_1 \cup a_2)}$$

$$\sum_{i=1}^{n} Pr_1(a_i) = 1 \qquad Pr(a_1) =$$

$$\sum_{i=1}^{n} Pr_2(a_i) = 2 \qquad 1 = Pr_1(w_j) \times \sum_{i=1}^{N} \frac{Pr_1(w_i)}{Pr_1(w_j)}$$