

## Section #2: Random Variables

---

### Overview of Section Materials

The warmup questions provided will help students practice concepts introduced in lectures. The section problems are meant to apply these concepts in more complex scenarios similar to what you will see in problem sets and quizzes.

## 1 Warmups

### 1.1 Independence

1. Definitions: Cite Bayes' Theorem.
2. True or False. Note that true means true for ALL cases.
  - (a) In general,  $P(AB|C) = P(B|C)P(A|BC)$
  - (b) If  $A$  and  $B$  are independent, so are  $A$  and  $B^C$ .

1. Bayes' Theorem:  $P(E|F) = \frac{P(F|E)P(E)}{P(F)}$

2. (a) True

Using Chain Rule or Def'n of Conditional Probability seems hard with three events, but remember that the intersection of two events is also an event. Therefore, expressions with three events are analogous to two events if we are clever about how we group things.

$$\begin{aligned}
 P(AB|C) &= P((AB)|C) && \text{Group} \\
 &= \frac{P((AB)C)}{P(C)} && \text{Def'n Conditional Prob} \\
 &= \frac{P(A(BC))}{P(C)} && \text{Group} \\
 &= \frac{P(A|BC)P(BC)}{P(C)} && \text{Chain Rule} \\
 &= \frac{P(A|BC)P(B|C)P(C)}{P(C)} && \text{Chain Rule} \\
 &= P(A|BC)P(B|C) && \text{Simplify} \\
 &= P(B|C)P(A|BC) && \text{Rearrange}
 \end{aligned}$$

(b) True

Start with some known equation which relates  $A$  and  $B^c$ . Is there a way we can exploit the assumed independence between  $A$  and  $B$  and then massage the equation into a form which proves  $A$  and  $B^c$  are independent?

$P(A) = P(AB) + P(AB^c)$	Law of Total Probability
$P(A) = P(A)P(B) + P(AB^c)$	Independence
$P(A) - P(A)P(B) = P(AB^c)$	Subtract
$P(A)(1 - P(B)) = P(AB^c)$	Factor
$P(A)P(B^c) = P(AB^c)$	Complement
$P(AB^c) = P(A)P(B^c)$	Rearrange

## 1.2 Random Variables and Expectation

1. Definitions:

(a) If we let  $X$  be a random variable, then what is  $E[X]$ ? What is  $E[g(X)]$ ?

(b) For random variables  $X_1, \dots, X_n$ , what is  $E[\sum_{i=1}^n X_i]$ ?

2. True or False: For any random variable  $X$ ,  $E[X^2] = E[X]^2$ .

1. Definitions:

(a)  $E[X] = \sum_x x p_X(x)$   
 $E[g(X)] = \sum_x g(x) p_X(x).$

(b)  $E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i]$

2. False

## 2 Problems

### 2.1 Taking Expectation: Breaking Vegas

**Preamble:** When a random variable fits neatly into a family we've seen before (e.g. Binomial), we get its expectation for free. When it does not, we have to use the definition of expectation.

**Problem:** If you bet on "Red" in Roulette, there is  $p = 18/38$  that you will win  $\$Y$  and a  $(1 - p)$  probability that you lose  $\$Y$ . Consider this algorithm for a series of bets:

1. Let  $Y = \$1$ .
2. Bet  $Y$ .

3. If you win, then stop.
4. If you lose, then set  $Y$  to be  $2Y$  and goto step (2).

What are your expected winnings when you stop? It will help to recall that the sum of a geometric series  $a^0 + a^1 + a^2 + \dots = \frac{1}{1-a}$  if  $0 < a < 1$ . Vegas breaks you: Why doesn't everyone do this?

Let  $X$  be the number of dollars that you earn.

The possible values of  $x$  are from the outcomes of: winning on your first bet, winning on your second bet, and so on.

$$\begin{aligned} E[X] &= \frac{18}{38} + \frac{20}{38} \frac{18}{38} (2 - 1) + \left(\frac{20}{38}\right)^2 \frac{18}{38} (4 - 2 - 1) + \dots \\ &= \sum_{i=0}^{\infty} \left(\frac{20}{38}\right)^i \left(\frac{18}{38}\right) \left(2^i - \sum_{j=0}^{i-1} 2^j\right) \\ &= \left(\frac{18}{38}\right) \sum_{i=0}^{\infty} \left(\frac{20}{38}\right)^i \\ &= \left(\frac{18}{38}\right) \frac{1}{1 - \frac{20}{38}} = 1 \end{aligned}$$

Real games have maximum bet amounts. You have finite money and casinos can kick you out. But, if you had no betting limits and infinite money, then go for it! (and tell me which planet you are living on).

## 2.2 Linearity of Expectation: Hat-Check

**Preamble:** Typically, it is easier to use linearity of expectation for sums of random variables than it is to manually compute a PMF and apply the definition.

**Problem:**  $n$  people go to a party and drop off their hats to a hat-check person. When the party is over, a different hat-check person is on duty, and returns the  $n$  hats randomly back to each person. Let  $X$  be the random variable representing the number of people who get their own hat back.

- a. For  $n = 3$ , find  $E[X]$  by first computing the probability mass function  $p_X$ , and then applying the definition of expectation.
- b. Find a general formula for  $E[X]$ , for any positive integer  $n$ .

- a. The number of people  $X$  who could get their hat back is in  $\{0, 1, 3\}$  (why not 2?). One can enumerate the possibilities:

$$123 \rightarrow 3$$

$$132 \rightarrow 1$$

$$213 \rightarrow 1$$

231 → 0

312 → 0

321 → 1

Hence,  $p_X(0) = P(X = 0) = 1/3$ ,  $p_X(1) = P(X = 1) = 1/2$ , and  $p_X(3) = P(X = 3) = 1/6$ . We then have that

$$E[X] = \sum_x x p_X(x) = 0 \cdot (1/3) + 1 \cdot (1/2) + 3 \cdot (1/6) = 1.$$

- b. For  $i = 1, \dots, n$ , let  $X_i$  be the indicator variable of whether person  $i$  gets their hat back. That is,  $X_i = 1$  if person  $i$  gets their hat back, and  $X_i = 0$  otherwise. Then,  $X = \sum_{i=1}^n X_i$ . For a particular person  $i$ , the probability they get their hat back is exactly  $1/n$  (why?), and so  $E[X_i] = 1 \cdot (1/n) + 0 \cdot (1 - 1/n) = 1/n$ .

By linearity of expectation,

$$E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \frac{1}{n} = n \cdot (1/n) = 1.$$

Now imagine finding the PMF for this random variable with  $n$  people/hats. There was no nice catch-all formula in part a) for  $n = 3$ , and so it would be extremely difficult/impossible to come up with one for general  $P(X = k)$ . Even if you could, evaluating the sum might be difficult. This is the power of linearity of expectation - though we don't know the PMF, we can still compute it easily by breaking it down into smaller pieces. Notice that people getting their hat backs are not independent events either!

### 2.3 Sending Bits to Space

**Preamble:** When sending binary data to satellites (or really over any noisy channel), the bits can be flipped with high probability. In 1947, Richard Hamming developed a system to more reliably send data. By using Error Correcting Hamming Codes, you can send a stream of 4 bits along with 3 redundant bits. If zero or one of the seven bits are corrupted, using error correcting codes, a receiver can identify the original 4 bits.

**Problem:** Lets consider the case of sending a signal to a satellite where each bit is independently flipped with probability  $p = 0.1$ .

- If you send 4 bits, what is the probability that the correct message was received (i.e. none of the bits are flipped).
- If you send 4 bits, with 3 Hamming error correcting bits, what is the probability that a correctable message was received?

- c. Instead of using Hamming codes, you decide to send 100 copies of each of the four bits. If for every single bit, more than 50 of the copies are not flipped, the signal will be correctable. What is the probability that a correctable message was received?

Hamming codes are super interesting. It's worth looking up if you haven't seen them before!

- a. Let  $Y$  be the number of 4 bits corrupted. Then  $P(Y = k)$  is given as:

$$P(Y = 0) = \binom{4}{0} (0.1)^0 (0.9)^4 = 0.656$$

- b. Let  $Z$  be the number of 7 bits corrupted. A correctable message is received if  $Z$  equals 0 or 1:

$$\begin{aligned} P(\text{correctable}) &= P(Z = 0) + P(Z = 1) \\ &= \binom{7}{0} (0.1)^0 (0.9)^7 + \binom{7}{1} (0.1)^1 (0.9)^6 = 0.850 \end{aligned}$$

That is a 30% improvement!

- c. Let  $X_i$  be the number of copies of bit  $i$  which are not corrupted. We can represent each as a random variable as we did in parts a and b.

$$\begin{aligned} P(\text{correctable}) &= \prod_{i=1}^4 P(X_i > 50) \\ &= \prod_{i=1}^4 \sum_{j=51}^{100} P(X_i = j) \\ &= \prod_{i=1}^4 \sum_{j=51}^{100} \binom{100}{j} (0.9)^j (0.1)^{100-j} \\ &= \left( \sum_{j=51}^{100} \binom{100}{j} (0.9)^j (0.1)^{100-j} \right)^4 > 0.999 \end{aligned}$$

But now you need to send 400 bits, instead of the 7 required by hamming codes :-).

### 3 Previous Exam Questions

#### 3.1 Spring 2021: Quiz 1

Jerry and Doris have abandoned their respective careers in education and puppy paw modeling and have taken up new work as art curators. As luck would have it, the San Francisco Museum of Modern Art has hired them both to independently appraise modern paintings. Doris has a keen eye for art and can spot a forgery - that is, identifies fake painting as fake - with probability 0.91,

whereas Jerry is more easily fooled and only spots a fake painting as fake with probability 0.82. Doris and Jerry also do an excellent good job at certifying authentic art as authentic. When Doris sees an authentic modern art painting, she certifies it as authentic with probability 0.99. Jerry does the same thing, but with probability 0.84.

Those working at SFMoMA are adamant that, based on past experience, the probability that each painting they consider is authentic with probability 0.6. Assuming a single painting is fake, Jerry and Doris independently identify that painting as fake. Similarly, Jerry and Doris independently identify authentic paintings as authentic.

For all of the questions below, in addition to providing an expression, please compute a numeric answer.

- a. Assuming a painting is authentic, what is the probability that neither Jerry nor Doris believe it to be authentic?
- b. Assuming a painting is a forgery, what is the probability that exactly one of Doris and Jerry identify it as a forgery?
- c. Given that both Doris and Jerry certify a painting as authentic, what is the probability the painting is really a forgery?

- a. Lets define some events for the problem. Let  $A$  be the event that a painting is authentic, let  $D$  be the event that Doris says a painting is authentic, and let  $J$  be the event that Jerry says a painting is authentic. We have:

$$\begin{aligned}
 P(A) &= 0.6 \text{ so that } P(A^C) = 0.4 \\
 P(D|A) &= 0.99 \text{ and } P(D^C|A) = 0.01 \\
 P(D^C|A^C) &= 0.91 \text{ and } P(D|A^C) = 0.09 \\
 P(J|A) &= 0.84 \text{ and } P(J^C|A) = 0.16 \\
 P(J^C|A^C) &= 0.82 \text{ so } P(J|A^C) = 0.18
 \end{aligned}$$

The problem is asking for  $P(D^C J^C | A)$ , which is  $P(D^C | A)P(J^C | A)$ , or  $0.01 \cdot 0.16 = 0.0016$ .

- b. The approach is similar to that for part a, except that there are more mutually exclusive conditional events that need to be union-ed. In particular, we need to compute  $P(D J^C | A^C) + P(D^C J | A^C) = P(D | A^C)P(J^C | A^C) + P(D^C | A^C)P(J | A^C)$ . Numerically, this is  $0.09 \cdot 0.82 + 0.91 \cdot 0.18 = 0.2376$ .
- c. Here, were asking for  $P(A^C | DJ)$ . The best approach is the employ Bayes Theorem of, in my presentation below, to effectively derive it from scratch.

$$\begin{aligned} P(A^C|DJ) &= \frac{P(A^C DJ)}{P(DJ)} \\ &= \frac{P(DJ|A^C)P(A^C)}{P(DJ)} \\ &= \frac{P(D|A^C)P(J|A^C)P(A^C)}{P(DJ|A)P(A) + P(DJ|A^C)P(A^C)} \\ &= \frac{P(D|A^C)P(J|A^C)P(A^C)}{P(D|A)P(J|A)P(A) + P(D|A^C)P(J|A^C)P(A^C)} \\ &= \frac{0.09 \cdot 0.18 \cdot 0.4}{0.99 \cdot 0.84 \cdot 0.6 + 0.09 \cdot 0.18 \cdot 0.4} = 0.01282 \end{aligned}$$