**Wednesday, September 27** **Fall 2023**

# Section 3: Conditional Probability Examples and Random Variables

Srihari Ganesh

*Based on section note formatting template and a practice problem by Rachel Li and Ginnie Ma '23*

## 1 Summary

### 1.1 Examples from class

Some takeaways:

- Winter girl: define events very specifically and see if you can simplify your problems with logic, not just relying on grinding through math
- Use the Law of Total Probability to condition on anything/everything you wish you knew
- In Monty Hall, condition on the location of the car!
- To mimic Simpson's paradox, set up some "hard" and "easy" tasks and make the better doctor do more of the hard tasks
- Learn how to turn a problem into the gambler's ruin problem:
  - Make losing happen at 0, and winning happen at some fixed $N$
  - Make sure each bet/step is only one dollar in either direction
  - Make sure the probability of winning each individual bet is constant

### 1.2 Random variables

**Definition 1.** **Random variables** are a numerical (real number) summary of the outcome of your experiment.

So for each possible outcome, the random variable takes on a certain value. Multiple outcomes can lead to the same value of the random variable.

**Definition 2.** The **support** of a random variable is the set of possible values it can take on.

**Definition 3.** We define discrete random variables by their **probability mass function**. You should

- define the probability $P(X = x)$ for each $x$ in the random variable's support
- always write probability 0 for any value of $x$ that is not in the support
- The PMF should always give valid probabilities, and should sum to 1 over all possible values

### 1.3 Distributions

**Story 4** (Bernoulli distribution)**.** You conduct a single trial that succeeds with probability $p$. Let $X = 1$ if the trial succeeds and $X = 0$ if it fails. Then $X$ follows a **Bernoulli distribution** with probability $p$, notated $X \sim \text{Bern}(p)$, and has a PMF

$$P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{else} \end{cases}$$

**Story 5** (Binomial distribution)**.** You conducts $n$ independent trials that each succeed with probability $p$. Let $Y$ be the total number of successes among the $n$ trials. Then $Y$ is distributed Binomial with $n$ trials and probability $p$, notated $Y \sim \text{Bin}(n, p)$, and the PMF is

$$P(Y = y) = \begin{cases} \binom{n}{y} p^y (1 - p)^{n-y} & y \in \{0, 1, \ldots, n\} \\ 0 & \text{else} \end{cases}$$

## 2 Practice Problems

1. **Working Bernoulli and Binomial stories and distributions** (NOTE: these are all properties you should know moving forward)

   (a) Suppose $X \sim \text{Bern}(p)$. What is the distribution of $Y = 1 - X$? What about if $X \sim \text{Bin}(n, p)$ and $Y = n - X$?

   > **Solution**
   > $Y$ basically flips $X$: $Y = 1$ when $X = 0$ and $X = 1$ when $Y = 0$. So $P(Y = 1) = 1 - p$ and $P(Y = 0) = p$, which means $Y \sim \text{Bern}(1 - p)$. In less mathy terms, $X$ tracks whether some experiment succeeds; $Y$ tracks whether it fails. The failure probability is $1 - p$, so $Y \sim \text{Bern}(1 - p)$.
   > The Binomial case is very similar. $Y$ ends up counting the number of failed Bernoulli trials among the $n$, so $Y \sim \text{Bin}(n, 1 - p)$.

   (b) Suppose $X_1, X_2, \ldots, X_m \sim \text{Bern}(1/m)$ independently. What is the distribution of $\sum_{i=1}^{m} X_i$?

   > **Solution**
   > We can see that there are $m$ independent Bernoulli trials each with success probability $1/m$, and $\sum_{i=1}^{m} X_i$ ends up counting the number that succeed. This follows the Binomial story; a fixed number of independent Bernoulli trials ($m$) with the same success probability ($1/m$) where our random variable $\sum_{i=1}^{m} X_i$ is the total number of successes. So $\sum_{i=1}^{m} X_i \sim \text{Bin}(m, 1/m)$.

   (c) Suppose $X_1 \sim \text{Bin}(n, p)$ and $X_2 \sim \text{Bin}(m, p)$ with $X_1, X_2$ independent. What is the distribution of $X_1 + X_2$?

   > **Solution**
   > Behind $X_1$ there are $n$ independent Bernoulli trials with probability $p$, and with $X_2$ there are $n$ independent Bernoulli trials with the same probability $p$. We are given that $X_1, X_2$ are independent, so we have a total of $m + n$ independent Bernoulli trials when combining the two. $X_1 + X_2$ is the number of successes across all $m + n$ of those trials, so $X_1 + X_2 \sim \text{Bin}(m + n, p)$.

   (d) Suppose $X \sim \text{Bern}(p)$. What is the distribution of $X^2$?

   > **Solution**
   > Bernoulli variables are often special cases! See that $X^2 = X$ since $1^2 = 1$ and $0^2 = 0$ (and those are the only two values in the support of $X$). So $X^2 \sim \text{Bern}(p)$ since it is the exact same as $X$.

2. **[Working with distributions and probability calculations]**You go to an online store and see 5 one-of-a-kind scarves. For each scarf, you decide to roll a fair die and buy the scarf if you roll a 4. Each scarf costs \$13 (I have now clue how much scarves cost). If you end up ordering anything, you will also have to pay a flat \$8 shipping fee (i.e., it's \$8 whether you get 2 scarves or 5). What's the probability you pay more than \$5, but no more than \$60?

**Solution**
Let $X$ be the number of you buy. Then $X \sim \text{Bin}(10, 1/6)$. If you actually end up buying something, the total price you pay is $13X + 8$ dollars. However, if $X = 0$, we don't pay shipping, so the price is \$0 in that case. So the actual minimum possible price we pay is \$21. We can avoid any numerical problems by using that instead:

$$\begin{aligned} P(21 \leq 13X + 8 \leq 60) &= P(13 \leq 13X \leq 52) \\ &= P(1 \leq X \leq 4). \end{aligned}$$

We can do a bit of complementary counting for out convenience:

$$\begin{aligned} P(1 \leq X \leq 4) &= 1 - P(X = 0) - P(X - 5) \\ &= 1 - (5/6)^5 - (1/6)^5 \approx 0.60 \end{aligned}$$

3. **[Gambler's ruin]** I am pursuing a very unconventional degree program. Each semester, I take a single class that I have a probability $p$ of failing. I am allowed to graduate if, at any point in time, I have passed 20 more classes than I have failed (up to that point). I get kicked out of school if, at any point, I have failed more classes than I have passed (up to that point). I get an unlimited amount of time to try to graduate as long as I haven't been kicked out yet. What's the probability that I eventually end up graduating?

**Solution**
Let $x$ be the probability that I end up graduating. Let's match this up to the gambler's ruin problem. I care about the difference between the number of classes I have passed and failed. I start with a difference of 0. I win (graduate) if I hit 20 and lose (kicked out) if I hit $-1$. So we have to shift to match the gambler's ruin problem (where failure has to occur at 0): I start at $i = 1$, win if I hit $N = 21$, and fail if I hit 0.
Each class only pushes me in a single step either way, so that matches the one-dollar bets. Finally, the probability of winning each bet (passing a class) is $1 - p$. So plugging in the gambler's ruin solution gives

$$x = \begin{cases} \frac{1-(\frac{p}{1-p})}{1-(\frac{p}{1-p})^N} & 1 - p \neq 1/2 \\ \\ \frac{1}{N} & 1 - p = 1/2 \end{cases}$$

4. **[Monty Hall: More Monty]** *(taken from Rachel & Ginnie's 2022 section notes)*

There are three doors, behind one of which there is a car and behind the other two of which there are goats. You want the car (hopefully). Initially, from your perspective, all possibilities are equally likely for where the car is. You choose a door, which we'll say is Door 1. Monty Hall opens a door with a goat in it and then offers you the option of switching. In class, in the case that Monty Hall knows where the car is and would never reveal the car, the probability that the strategy of always switching succeeds is 2/3. Find the probability that the strategy of always switching succeeds if:

(a) Monty does not know what lies behind the doors and opens Door 2 at random (50/50), which happens to not reveal the car.

**Solution**

*This is not the only solution! Check out Karina and Liz's notes if you want to see one of the alternatives!*

Let $W$ be the event we win and $C_i$ be the event that the car is behind door $i$. Let $O_i$ be the event that Monty chooses to reveal door $i$. Then we're calculating $P(W|O_2, C_2^c)$ since Monty revealed the second door and the car was not behind the second door.

$$P(W|O_2, C_2^c) = P(C_3|O_2, C_2^c).$$

I prefer car information in the condition (it tells me everything unknown), so let's use Bayes' rule with extra conditioning to finish this:

$$P(C_3|O_2, C_2^c) = \frac{P(O_2|C_3, C_2^c)P(C_3|C_2^c)}{P(O_2|C_2^c)}.$$

Doing each individually is much more manageable:

- $P(O_2|C_3, C_2^c) = P(O_2|C_3)$ since $C_3$ implies $C_2^c$. From here, since Monty's choice was random we have $O_2$ being independent of $C_3$, so $P(O_2|C_3) = P(O_2) = 1/2$.
- Similarly, $P(O_2|C_2^c) = 1/2$.
- Finally, $P(C_3|C_2^c) = 1/2$ since if the car isn't behind $C_2$, it's equally likely to be behind either of the other two doors (note this is without any extra information about which door Monty opened).

So we end up with

$$P(C_3|O_2, C_2^c) = \frac{(1/2)(1/2)}{1/2} = \frac{1}{2}.$$

(b) Monty always reveals a goat (he knows where the car is) and, if he has a choice, he always reveals the leftmost goat (which may depend on the player's choice). In this case, he has opened Door 2 knowing you chose Door 1.

**Solution**

*This is not the only solution! Check out Karina and Liz's notes if you want to see one of the alternatives!*

Let's use the same events: $W$ is the event we win, $O_i$ is the event Monty opens door $i$, and $C_i$ is the event that the car is behind door $i$. We're again finding $P(W|O_2, C_2^c) = P(C_3|O_2, C_2^c)$. We can even use the same Bayes' rule setup:

$$P(C_3|O_2, C_2^c) = \frac{P(O_2|C_3, C_2^c)P(C_3|C_2^c)}{P(O_2|C_2^c)}$$

And again $P(C_3|C_2^c) = 1/2$ since it's not related to Monty's actions. The actual values are different now, though.

- $P(O_2|C_3, C_2^c) = P(O_2|C_3)$ again. However, now consider that if the car is behind door 3, Monty must pick door 2, so $P(O_2|C_3) = 1$.
- For the denominator $P(O_2|C_2^c)$, consider both cases (this is an informal way to do LOTP). If the car is behind door 1, there are two goat doors, but the leftmost goat is behind door 2. If the car is behind door 3, Monty can only open door 2.

So door 2 will always get opened in this case, which gives $P(O_2|C_2^c) = 1$. Putting this all together yields

$$P(C_3|O_2, C_2^c) = \frac{(1)(1/2)}{1} = 1/2.$$