CS4830 Probability Recitation

Instructor: Elaine Shi

January 30, 2017

The goal of this lecture is to refresh your memory about probability. You may have seen probability in various different forms and notations. In our course, we will often adopt the "experiment notation". Through this recitation, you'll learn to be comfortable with experiment notation. We will also refresh our minds about fundamental concepts and tools such as the conditional probability, expectation, the Bayes rule, the union bound, and so on.

1 Two Children

A person has two children. Given that at least one of them is a boy, what is the probability that the other child is also a boy?

Sol. $\frac{1}{3}$. Let X be the gender of the first child, $X = 1$ if it is a boy, $X = 0$ otherwise. Let Y be the gender of the second child.

X	Y	event
0	0	
0	1	X or $Y = 1$
1	0	X or $Y = 1$
1	1	X or $Y = 1$, X and $Y = 1$

$$\Pr[X, Y \xleftarrow{\$} \{0, 1\} : X \text{ and } Y = 1 | X \text{ or } Y = 1] = \frac{1}{3}$$

Suppose a person has two children of different birthday. Given that the younger child is a boy, what is the probability that the other child is also a boy?

Sol. $\frac{1}{2}$. Let X be the gender of the older, let Y be the gender of the younger.

$$\Pr[X, Y \xleftarrow{\$} \{0, 1\} : X = 1 | Y = 1] = \frac{1}{2}$$

2 Poker Face

A deck of cards contains 52 cards. There are 13 different kinds of cards, with four suits of cards of each kind. These kinds are twos, threes, fours, fives, sixes, sevens, eights, nines, tens, jacks, queens, kings and aces. There are also four suits: spades, clubs, hearts and diamonds, each containing 13 cards. A pair means two cards in the same kind. A five-card poker hand contains a two pairs if it contains two pairs in different kinds and another card in a third kind. A flush means all cards are of the same suit.

A straight means all cards have consecutive kinds. (Note that an ace can be considered either the lowest card of an A-2-3-4-5 straight or the highest card of a 10-J-Q-K-A straight. However, Q-K-A-2-3 is NOT a straight.)

What is the probability that a five-card poker hand is a flush? Henceforth, we use the short-hand

$$\Pr[\texttt{flush}] := \Pr[\text{Sample } X_1..X_5 \text{ w/o replacement} : X_1..X_5 \text{ is a } \texttt{flush}]$$

$$\Pr[\text{flush}] = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}} = \frac{4 \times 11 \times 9}{4 \times 51 \times 5 \times 49 \times 4} \approx 0.0020$$

What is the probability that a five-card poker hand is two pairs?

$$\Pr[\text{two pairs}] = \frac{\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 11 \cdot \binom{4}{1}}{\binom{52}{5}} \approx 0.0475$$

What is the probability that a five-card poker hand is a flush and two pairs?

$$\Pr[\text{flush and two pairs}] = 0$$

In this case, we say that the two events are *mutually exclusive* (not to be confused with *independence*).

What is the probability that a five-card poker hand does not contains a flush or a two pairs?

$$\Pr[\,\overline{\text{flush or two pairs}}\,] = 1 - \Pr[\text{flush}] - \Pr[\text{two pairs}] + \Pr[\text{flush and two pairs}] \approx 0.713$$

3 Crypton

A space probe near Crypton communicates with Earth using bit strings. Suppose that in its transmissions it sends a '1' 60% of the time and a '0' 40% of the time. When a 0 is sent, the probability that it is received correctly is 0.85, and the probability that it is received incorrectly (as a 1) is 0.15. When a 1 is sent, the probability that it is received correctly is 0.7, and the probability that it is received incorrectly (as a 0) is 0.3.

Bayes Rule: $\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]}$.

Find the probability that a 0 is received. **Sol.** 0.52.

First, here is a rather pendatic way of writing this out as an experiment. Perhaps doesn't give us too much intuition here, but we'll need to be comfortable with such experiment notation.

Expt:
$b := \begin{cases} 0 & \text{w.p. } 40\% \\ 1 & \text{w.p. } 60\% \end{cases}$
if $b = 0$, $b' := \begin{cases} 0 & \text{w.p. } 85\% \\ 1 & \text{w.p. } 15\% \end{cases}$
else $b = 0$, $b' := \begin{cases} 0 & \text{w.p. } 85\% \\ 1 & \text{w.p. } 15\% \end{cases}$
Output b, b'

$$\begin{aligned}
\Pr[b, b' \leftarrow \mathsf{Expt} : b' = 0] &= \Pr[b' = 0 \& b = 0] + \Pr[b' = 0 \& b = 1] \\
&= \Pr[b' = 0 | b = 0] \cdot \Pr[b = 0] + \Pr[b' = 0 | b = 1] \cdot \Pr[b = 1] \\
&= 0.85 \cdot 0.4 + 0.3 \cdot 0.6 \\
&= 0.52
\end{aligned}$$

Find the probability that a 0 was sent, given that a 0 was received. Sol. 0.65

$$\begin{aligned}
\Pr[b, b' \leftarrow \mathsf{Expt} : b = 0 | b' = 0] &= \frac{\Pr[b = 0 \& b' = 0]}{\Pr[b' = 0]} \\
&= \frac{\Pr[b' = 0 | b = 0]}{\Pr[b' = 0]} \\
&= \frac{0.85 \cdot 0.4}{0.85 \cdot 0.4 + 0.3 \cdot 0.6} = \frac{17}{26} \\
&\approx 0.65
\end{aligned}$$

4 Expectation of the Geometric Distribution

Given a biased coin that ends up heads with probability p, how many tosses does it take for the coin to show heads, in expectation? [2]

Sol. $\frac{1}{p}$. Consider the state space $S = \{H, TH, TTH, TTTH, \dots\}$. We define the probability mass function to be

$$f(T^i H) = f(i \text{ tails followed by a head}) = (1-p)^i p.$$

Let X be the random variable that denote the number of coin tosses needed for heads to show up. Then $X(T^i H) = i + 1$. The expectation of X is then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\text{inf}} (i+1) p (1-p)^i \\
&= p \sum_{i=0}^{\text{inf}} (i+1)(1-p)^i = p \frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

5 The Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? [1]

Sol. Yes. Model this game with following experiment:

Expt:
$\mathbf{car} \xleftarrow{\$} \{1,2,3\}$
$X \xleftarrow{\$} \{1,2,3\}$
if $X = \mathbf{car}$, then pick random $Y \leftarrow \{1,2,3\}$ s.t. $Y \neq X$
else let $Y = \{1,2,3\} \backslash \{X, \mathbf{car}\}$
let $\overline{Y} = \{1,2,3\} \backslash \{X, Y\}$

$$\begin{aligned}
\Pr[\overline{Y} = \mathbf{car}] &= \Pr[\overline{Y} = \mathbf{car} | X = \mathbf{car}] \cdot \Pr[X = \mathbf{car}] + \Pr[\overline{Y} = \mathbf{car} | X \neq \mathbf{car}] \cdot \Pr[X \neq \mathbf{car}] \\
&= 0 + 1 \cdot \frac{2}{3} \\
&= \frac{2}{3}
\end{aligned}$$

The probability of winning a car by sticking to initial choice is now $\frac{1}{3}$. So switching is better choice.

6 Union Bound

$G(n,p)$ is the graph G on n nodes, where for every pair of vertices an edge is included with probability p. Show that the probability that there is an isolated vertex is upper bounded by $n(1-p)^{n-1}$.

Proof.

$$\Pr[\text{vertex 1 is isolated}] = (1-p)^{n-1}$$

By the union bound,

$$\Pr[\text{exists isolated vertex }] = n(1-p)^{n-1}$$

□

Note that the event that vertex 1 is isloated and the event that vertex 2 is isolated are *not independent*. The union bound is used everywhere in cryptography, especially because you do not need independence to apply the union bound.

7 The Dating Problem

In our lives, we need to decide who to date and marry. How do we make such a decision? We meet someone, we date him/her for a while, then we need to decide if he/she is a keeper. If you brutally abandon him/her, he/she likely will not come back to you if you change your mind later and decide that there is no one better. Suppose that you aim to get married before you are, say, 35. You estimate that you will be able to meet and date n eligible candidates before that. How can you design a strategy that will allow you to choose the Mr./Ms. Right from the n eligible candidates you will meet and date?

7.1 The Unfaithful World

Say you are allowed to date two people at the same time. Here is a strategy that will guarantee that you find the best date. You date two people you meet, you decide who you like better. You dump the one that you like less. Now if someone comes along, you can date him/her too, since you can date two people simultaneously. This repeats, until eventually you finish dating all n candidates, and the one you have kept is your Mr./Ms. Right. You then make a proposal of marriage to him/her.

7.2 The Faithful World

Suppose we live in a faithful world. You can't date two people simultaneously. When you are in a relationship, no one wants to date you. When you dump someone, the decision sticks, and there is no going back. Now the problem becomes much more difficult. Say, you meet a person, and you now have to make an *online* decision without foresight into whom else you are going to meet in the future. You'd better make the choice carefully to maximize your chance of happiness in life. What is a good strategy in the faithful world?

We are going to make the following assumptions.

- The n people come in random order — note that this assumption may not model reality, but for simplicity this is what we will assume for now. When you apply what you have learned to guide your actions in life, you might want to carefully re-evaluate the assumptions.

- You can assign a preference score to each person you date. This gives a total ordering among all people.

- You can use the relative order between dates to make a decision. You may dump your dates in the process. However, once you decide that someone is a keeper and you marry him/her, you remain faithful for the rest of your life.

- You would like to maximize the probability that you find your Mr./Ms. Right.

Let us consider the following strategy space: say you reject the first $r-1$ eligible people you date. Then you continue, and whenever you date someone who is better than all of the first $r-1$ people, you make a proposal of marriage to him/her. Henceforth we say that r is the *cutoff*. As a special case, if $r = 1$, this means that you marry the first person you date. The probability that this will be the best person is $\frac{1}{n}$. This does not seem like it's the best strategy, you might want to date more people for calibration, before you make a commitment. On the other hand, if we let $r = n$, this means that you always reject the first $n-1$ people and marry the last one. In this case, the probability of finding Mr./Ms. Right is also $\frac{1}{n}$. This also does not seem like the best strategy because if you wait out till the end, then you essentially have no choice left.

It seems like some choice of r in the middle might be the best strategy. Let's now see what choice of r is the best. Let $P(r)$ denote the probability that you select your Mr./Ms. Right for a cutoff choice of r.

$$
\begin{aligned}
P(r) &= \sum_{i=1}^{n} \Pr[i\text{-th person selected and } i\text{-th is best}] \\
&= \sum_{i=1}^{n} \Pr[i\text{-th person selected} \big| i\text{-th is best}] \Pr[i\text{-th is best}] \\
&= \sum_{i=1}^{r-1} 0 + \sum_{i=r-1}^{n} \Pr[\text{best among first } i-1 \text{ appear among the first } r-1 \big| i\text{-th is best}] \cdot \frac{1}{n} \\
&= \sum_{i=r}^{n} \frac{r-1}{i-1} \cdot \frac{1}{n} = \frac{r-1}{n} \sum_{i=r}^{n} \frac{1}{i-1}
\end{aligned}
$$

Claim: as n goes to infinity, the above expression can be approximated by $-x \ln x$ where $x := \frac{r}{n}$. Taking the derivative, you'll see that when $x = \frac{1}{e}$, the expression is maximized.

7.3 Epilogue

Interestingly, it turns out that this strategy is not only optimal among the strategy space we considered, but also the optimal strategy among all strategies that use only the relative order between candidates to make a decision. The dating problem is also referred to as the secretary problem or the hiring problem. There are many variants of the problem based on varying assumptions. There is a branch of theoretical computer science that studies such online optimization problems.

References

[1] Monty hall problem. `https://en.wikipedia.org/wiki/Monty_Hall_problem`.

[2] W. Tseng R. Pass. *A Course in Discrete Structures*.