RANDOM VARIABLES, CONDITIONAL PROBABILITY, DISTRIBUTIONS 7

COMPUTER SCIENCE MENTORS 70

March 13 to March 17, 2017

1 Conditional Probability

1.1 Introduction

Bayes' Rule

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$$

Total Probability Rule

$$\mathrm{P}[B] = \mathrm{P}[A \cap B] + \mathrm{P}[\bar{A} \cap B] = \mathrm{P}[B|A] * \mathrm{P}[A] + \mathrm{P}[B|\bar{A}] * (1 - \mathrm{P}[A])$$

Independence

Two events A, B in the same probability space are independent if

$$\mathrm{P}[A \cap B] = \mathrm{P}[A] * \mathrm{P}[B]$$

1.2 Questions

1. A lie detector is known to be 80% reliable when the person is guilty and 95% reliable when the person is innocent. If a suspect is chosen from a group of suspects where only 1% have ever committed a crime, and the test indicates that the person is guilty, what is the probability they are innocent?

Solution: Let I and G be the events that the person is innocent and guilty respectively, and let L_I and L_G be the events that the test says innocent or guilty.

$$P(I|L_G) = \frac{P(L_G|I) * P(I)}{P(L_G|I) * P(I) + P(L_G|G) * P(G)} = \frac{0.05 * 0.99}{0.05 * 0.99 + 0.8 * 0.01} = 0.86$$

2. Jim and George are setting up venture capital portfolios. Suppose that Jim picks $n+1$ startups to fund and George picks n startups to fund. Suppose that the probability of any startup succeeding is $\frac{1}{2}$ and all of the startups succeed or fail independently. What is the probability that Jim picks more successful startups than George?

Solution: Since Jim picks one more startup than George, it is impossible that they pick both the same number of successful startups and the same number of unsuccessful startups. So Jim picks either more successful startups than George or more unsuccessful startups than George (but not both). Since the probability of succeeding is $\frac{1}{2}$, these events are equally likely by symmetry, so both events have probability $\frac{1}{2}$.

3. Oski the bear has lost his dog in either forest A (with a priori probability 0.4) or in forest B (with a priori probability 0.6).
On any given day, if the dog is in A and Oski spends a day searching for it in A, the conditional probability that he will find the dog that day is 0.25. Similarly, if the dog is in B and Oski spends a day looking for it there, the conditional probability that he will find the dog that day is 0.15.
The dog cannot go from one forest to the other. Oski can search only in the daytime, and he can travel from one forest to the other only at night.

   (a) In which forest should Oski look to maximize the probability he finds his dog on the first day of the search?

Solution:

$$P(\text{Finding In A}) = P(\text{Dog In A}) * P(\text{Search Successful}) = \frac{4}{10} * \frac{2}{8} = \frac{1}{10}$$
$$P(\text{Finding In B}) = P(\text{Dog In B}) * P(\text{Search Successful}) = \frac{6}{10} * \frac{3}{20} = \frac{9}{100}$$
$$\frac{1}{10} > \frac{9}{100}$$

so Oski should search in **Forest A**

(b) Given that Oski looked in A on the first day but didnt find his dog, what is the probability that the dog is in A?

Solution:

$$P(\text{Dog In A} \mid \text{Searched A and Failed}) = \frac{P(\text{Dog In A} \cap \text{Searched A and Failed})}{P(\text{Dog In A} \cap \text{Searched A and Failed}) + P(\text{Dog Not In A})} = \frac{\frac{4}{10} * \frac{3}{4}}{\frac{4}{10} * \frac{3}{4} + \frac{6}{10}} = \frac{1}{3}$$

(c) If Oski flips a fair coin to determine where to look on the first day and finds the dog on the first day, what is the probability that he looked in A?

Solution: $\frac{P(\text{Looks In A}) \cap P(\text{Dog In A}) \cap P(\text{Finds Dog})}{P(\text{Finds Dog})} = \frac{\frac{1}{2} * \frac{4}{10} * \frac{1}{4}}{\frac{1}{2} * \frac{4}{10} * \frac{1}{4} + \frac{1}{2} * \frac{6}{10} * \frac{3}{20}}$

(d) If the dog is alive and not found by the Nth day of the search, it will die that evening with probability $\frac{N}{N+2}$. Oski has decided to look in A for the first two days. What is the probability that he will find a live dog for the first time on the second day?

Solution:

$$P(\text{Dog In A}) \cap P(\text{Find Dog First Day})^C \cap P(\text{Dog Dies})^C \cap P(\text{Finds Dog Second Day}) = \frac{4}{10} * \frac{3}{4} * \frac{2}{3} * \frac{1}{4}$$

2 Monty Hall

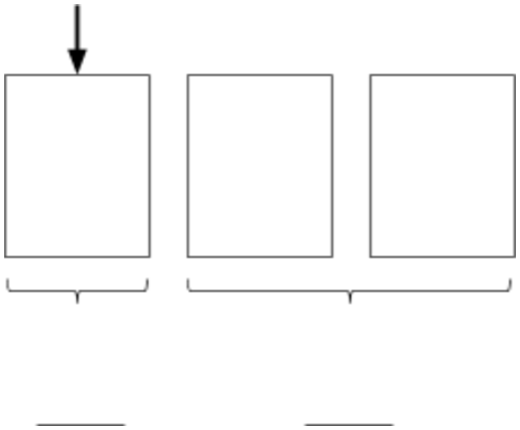
2.1 Introduction

The Problem :
Suppose a contestant is shown 3 doors. There is a car behind one of them and goats behind the rest. Then they do the following:

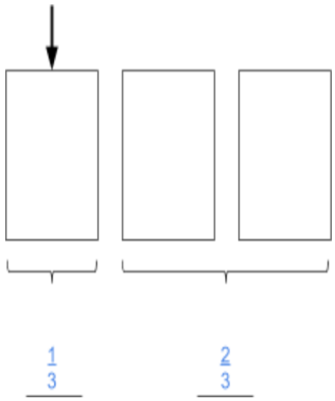
1. Contestant chooses a door.
2. Host opens a door with a goat behind it.
3. Contestant can choose to switch or stick to original choice

Is the contestant more likely to win if they switch?

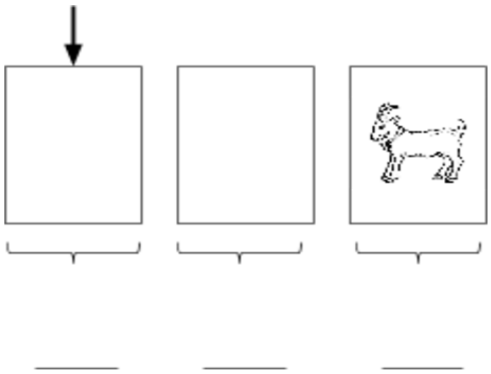
At step 1, what is the probability that the car is behind the door the contestant chose? What is the probability that the car is behind the other two doors?



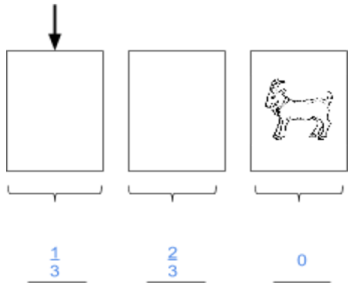
Solution:



After the host opens a door with a goat, what are the probabilities of the car being behind each door?



Solution:



2.2 Questions

1. **Grouping Doors**
   Now we have 6 doors. You pick 1 and the other 5 doors are divided into two groups:

one with 2 doors and the other with 3 doors. He removes doors until each group has 1 door left. Do you switch? What do you switch to?

Solution: Choose the door that was part of the group of three doors. When the doors were split up into three groups, those groups each had a probability of $\frac{1}{6}, \frac{2}{6},$ and $\frac{3}{6}$. These probabilities do not change when doors are removed, so three remaining doors each have probability $\frac{1}{6}, \frac{2}{6},$ and $\frac{3}{6}$.

2. **Macs and Monty**

Suppose instead of the normal Monty Hall scenario in which we have two empty doors and a car residing behind the third we have a car behind one door, a Mac behind another, and nothing behind the third.

Let us assume that the contestant makes an initial pick at his/her discretion (random) and the host proceeds to ALWAYS open the empty door. When the contestant's initial choice corresponds to the empty door, the host will say so and the contestant must switch.

Does the typical Monty Hall paradox of $\frac{2}{3}$ chance of obtaining the car by switching versus a $\frac{1}{3}$ chance of obtaining the car by staying apply in this particular case?

Solution: No. The normal Monty Hall paradox holds because when another door is opened, you learn nothing about your door, so the chance that you picked the right door from the beginning remains $\frac{1}{n}$ (which means the chance of the other door must be $\frac{n-1}{n}$). However, in this case, you may learn something about your door–you may be told that it must be wrong for example–thus shattering the paradox. These are all of the possible scenarios:

Initial Pick	Switch?	Win?
Car	Yes	Lose
Car	No	Win
Mac	Yes	Win
Mac	No	Lose
Empty	Yes (to car)	Win
Empty	Yes (to car)	Lose

Switching results in a win 50% of the time, so there is a 50% chance of winning regardless of strategy.

3 Balls and Bins

3.1 Questions

1. Given n bins and m balls find the largest value of m such that the probability that there is no collision is above $\frac{1}{2}$? (Use the union bound to approximate.)

Solution: Let A be the event that there are no collisions. We know we hve $\binom{m}{2} = k$ pairs of balls. Let A_i be the probability that pair i collides. $\Pr[A_i]$ is simply $\frac{1}{n}$, so by the union bound, we have

$$\Pr[\overline{A}] \leq \sum_{i=1}^{k} \Pr[A_i] = k\left(\frac{1}{n}\right) = \frac{m(m-1)}{2n} \approx \frac{m^2}{2n}$$

So we have that $\frac{m^2}{2n} = \frac{1}{2}$ and so $m \leq \sqrt{2}$.

4 Expectation and Random Variables

4.1 Introduction

Random variable : a function $X : \omega \to R$ that assigns a real number to every outcome ω in the probability space.

Expectation : The expectation of a random variable X is defined as

$$\mathrm{E}(X) = \sum_{\alpha \in A} a * \mathrm{P}[X = a]$$

Where the sum is over all possible values taken by the random variable.

4.2 Questions

1. Does the random variable always take on the value of its expectation?

Solution: No

2. Make a random variable from the probability space: {2, 3, 6, 7} that half the time is 1 and the other half the time is 0. What function can represent this random variable?

Solution: Several answers: $x < 5, x\%2 = 0, (x = 2 or x = 7)$, etc.

3. Given the random variable X defined as taking on the value 1 with probability .25, 2 with probability .5, and 20 with probability .25, what is the expectation of X?

Solution: $E(X) = .25 * 1 + .5 * 2 + .25 * 20 = 6.25$

5 Distributions

5.1 Introduction

Geometric Distribution: Geom(p) Number of trials required to obtain the first success. Each trial has probability of success equal to p. The probability of the first success happening at trial k is:

$$\mathrm{P}[X = k] = (1-p)^{k-1} * p, k > 0$$

The expectation of a geometric distribution is:

$$\mathrm{E}(X) = \frac{1}{p}$$

The variance of a geometric distribution is:

$$Var(X) = \frac{1-p}{p^2}$$

Solution: Derivation of $\mathrm{E}(X)$: The clever way to find the expectation of the geometric distribution uses a method known as the renewal method. $\mathrm{E}(X)$ is the expected number of trials until the first success. Suppose we carry out the first trial, and one of two outcomes occurs. With probability p, we obtain a success and we are done (it only took 1 trial until success). With probability $1 - p$, we obtain a failure, and we are right back where we started. In the latter case, how many trials do we expect until our first success? The answer is $1 + E(X)$: we have already used one trial, and we expect $\mathrm{E}(X)$ more since nothing has changed from our original situation (the geometric distribution is memoryless). Hence $\mathrm{E}(X) = p * 1 + (1 - p) * (1 + E(X))$

Binomial Distribution: Bin(n, p) Number of successes when we do n independent trials. Each trial has a probability p of success. The probability of having k successes:

$$\mathrm{P}[X = k] = \binom{n}{k} * p^k * (1-p)^{n-k}$$

The expectation of a binomial distribution is:

$$\mathrm{E}(X) = np$$

The variance of a binomial distribution is:

$$Var(X) = np(1-p)$$

Solution: Can walk through the derivation of $\mathrm{E}(X)$: We would have to compute this sum:

$$\mathrm{E}(X) = \sum_k k * \mathrm{P}[X = k] = \sum_{k=0}^{n} k * \binom{n}{k} * p^k * (1-p)^{n-k}$$

Instead of doing that just use Bernoulli variables:

$$X = X_1 + \ldots + X_n$$

And now use linearity of expectation:

$$\mathrm{E}(X) = \mathrm{E}(X_1 + \ldots + X_n) = \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_n)$$

Since the probability of a success happening at each step is p, and there are n steps, we are just summing p n times.

Poisson Distribution: Pois(λ) This is an approximation to the binomial distribution. Let the number of trials approach infinity, let the probability of success approach 0, such that $\mathrm{E}(X) = np = \lambda$. This is an accepted model for rare events. The probability of having k successes:

$$\mathrm{P}[X = k] = \frac{e^{-\lambda} * \lambda^k}{k!}$$

The expectation of a poisson distribution is:

$$\mathrm{E}(X) = \lambda$$

The variance of a poisson distribution is:

$$Var(X) = \lambda$$

Solution: Can walk through the derivation of P(X):

$$\begin{aligned}
\mathrm{P}[X = k] = \binom{n}{k} * p^k * (1-p)^{n-k} & \\
&= \frac{n!}{k! * (n-k)!} * p^k * (1-p)^{n-k} \\
&\approx \frac{n^k * p^k}{k!} * (1 - \frac{\lambda}{n}) \\
&\approx \frac{\lambda^k * e^{-\lambda}}{k!}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(X) = \sum_{k=0}^{\infty} k * \frac{e^{-\lambda} * \lambda^k}{k!} & \\
&= \sum_{k=1}^{\infty} k * \frac{e^{-\lambda} * \lambda^k}{k!} \\
&= e^{-\lambda} * \lambda * \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda} * \lambda * \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} * \lambda * e^{\lambda} \\
&= \lambda
\end{aligned}$$

5.2 Questions

1. You are Eve, and as usual, you are trying to break RSA. You are trying to guess the factorization of N, from Bobs public key. You know that N is approximately 1,000,000,000,000. To find the primes p and q, you decide to try random numbers from 2 to $1,000,000 \approx \sqrt{N}$, and see if they divide N.

   To do this, you roll a 999,999-sided die to choose the number, and see if it divides N using your calculator, which takes five seconds. Of course, there will be one number in this range that does divide Nnamely, the smaller of p and q.

   (a) What kind of distribution would you use to model this?

   Solution: Geometric probability of success each time is $p = \frac{1}{999{,}999}$

   (b) What is the expected amount of time until you guess the correct answer, if it takes five seconds per guess (you only have a calculator)? Answer in days.

   Solution:

   $$\mathrm{E}(x) = \frac{1}{p} = 999,999 \text{ tries}$$

   $$(999,999 * 5 \text{ sec}) * \frac{1 \text{ min}}{60 \text{ sec}} * \frac{1 \text{ hr}}{60 \text{ min}} * \frac{1 \text{ day}}{24 \text{ hr}} \approx 57.9 \text{ days}$$

2. Now you are trying to guess the 6-digit factorization digit by digit. Lets assume that when you finish putting these digits together, you can figure out how many digits you got right. Use zeros for blank spaces. For example, to guess 25, you would put 000025

   (a) What kind of distribution would you use to model this?

   Solution: Binomial, since this is multiple independent trials that can either succeed or fail.

   (b) What is the probability that you get exactly 4 digits right?

   Solution: $\binom{6}{4} * \frac{1}{10}^4 * \frac{9}{10}^2$

   (c) What is the probability that you get less than 3 correct?

Solution: $\binom{6}{2} * \frac{1}{10}^2 * \frac{9}{10}^4 + \binom{6}{1} * \frac{1}{10} * \frac{9}{10}^5 + \binom{6}{0} * \frac{1}{10}^0 * \frac{9}{10}^6$

3. You are Alice, and you have a high-quality RSA-based security system. However, Eve is often successful at hacking your system. You know that the number of security breaches averages 3 a day, but varies greatly.

   (a) What kind of distribution would you use to model this?

   Solution: Poisson! That's what we use to model the probably frequencies of rare events.

   (b) What is the probability you experience exactly seven attacks tomorrow? At least seven (no need to simplify your answer)?

   Solution:
   $$\mathrm{P}[X = 7] = \frac{\lambda^7}{7!} * e^{-\lambda} = \frac{3^7}{7!} * e^{-3} \approx 0.0216$$
   $$\mathrm{P}[X \geq 7] = \sum_{i=7}^{\infty} \mathrm{P}[X = i] = \sum_{i=7}^{\infty} \frac{3^i}{i!} * e^{-3}$$

   (c) What is the probability that, on some day in April, you experience exactly six attacks?

   Solution:
   $$\mathrm{P}[X = 6] = \frac{3^6}{6!} * e^{-3} \approx 0.0504$$
   $$1 - (1 - 0.0504)^{30} \approx 0.788 = 78.8\%$$