CONDITIONAL PROBABILITY, RANDOM VARIABLES 7

COMPUTER SCIENCE MENTORS 70

March 12 - 16, 2018

1 Monty Hall

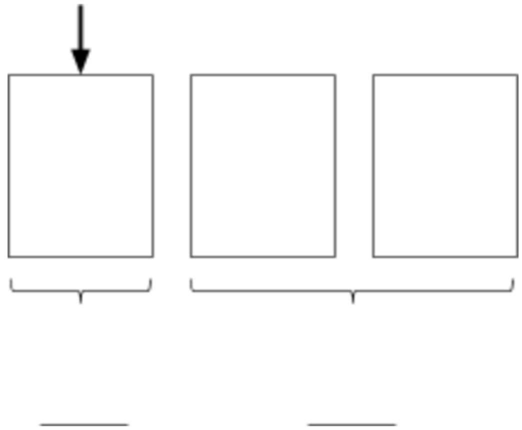
1.1 Introduction

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:

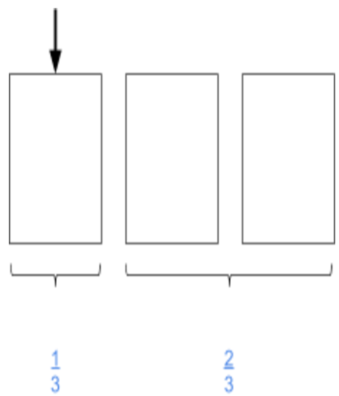
1. Contestant chooses a door.
2. Host opens a door with a goat behind it.
3. Contestant can choose to switch or stick to original choice

Is the contestant more likely to win if they switch?

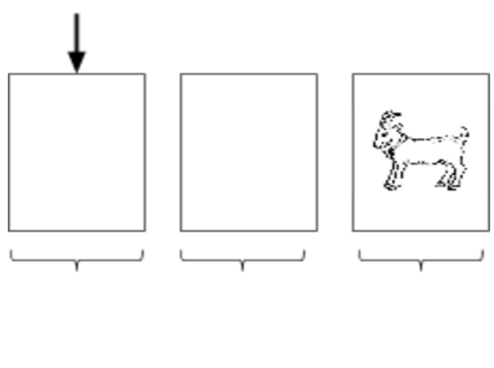
At step 1, what is the probability that the car is behind the door the contestant chose? What is the probability that the car is behind the other two doors?



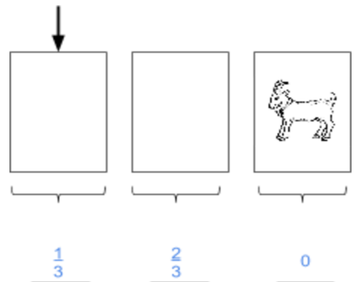
Solution:



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



Solution:

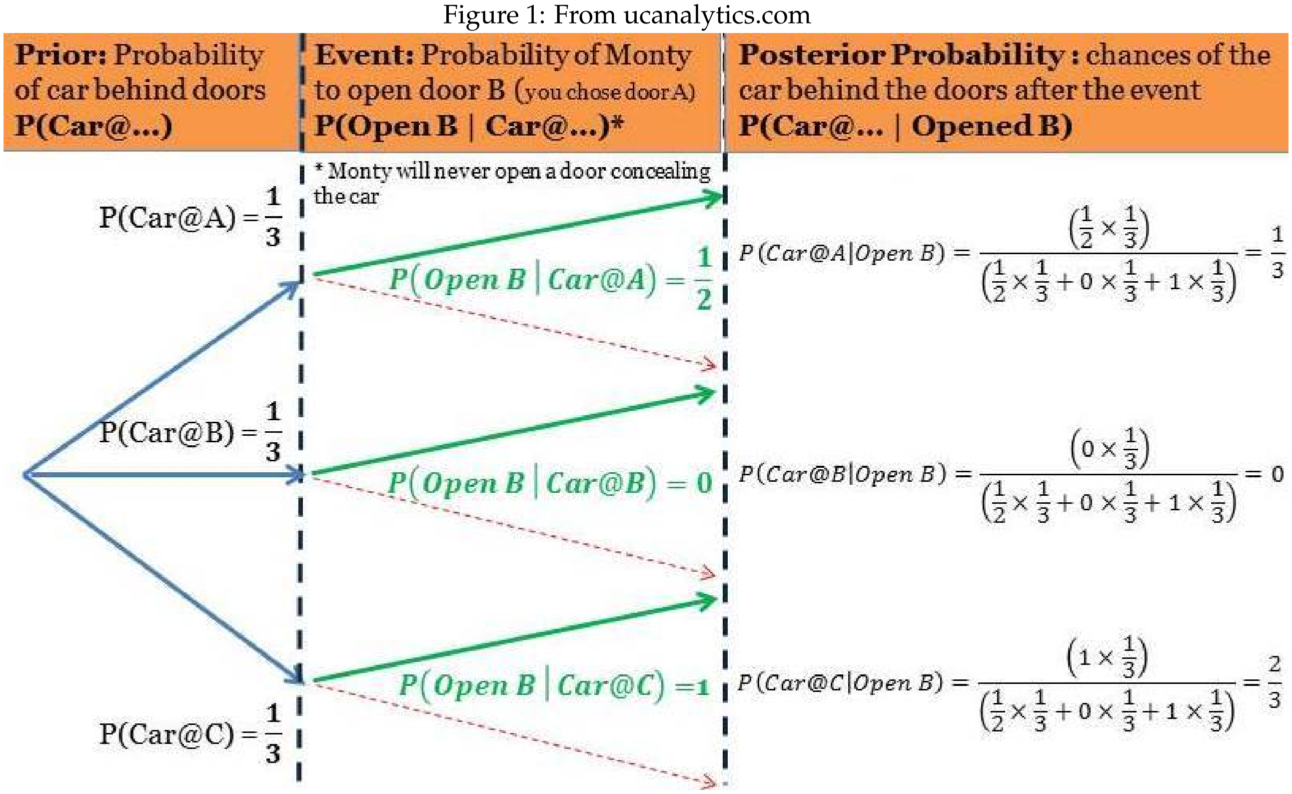


These probabilities might initially seem counterintuitive, but can be explained through an important probability theorem, Bayes' rule, which says that:

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

We can apply this rule to the Monty Hall problem, as shown in the figure below. We assume that the contestant has chosen Door A.

Figure 1: From ucanalytics.com



Beyond Monty Hall, Bayes' rule can be applied to a wide variety of problems, as will be shown in the next section on conditional probability.

2 Conditional Probability

2.1 Introduction

Definition of Conditional Probability

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

Bayes' Rule

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

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

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

2.2 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. 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:

   $$P(\text{Dog In A} \mid \text{No Success in A})$$
   $$= \frac{P(\text{Dog In A} \cap \text{No Success in A})}{P(\text{Dog In A} \cap \text{No Success in A}) + P(\text{Dog Not In A} \cap \text{No Success in A})}$$
   $$= \frac{\frac{4}{10} * \frac{3}{4}}{\frac{4}{10} * \frac{3}{4} + \frac{6}{10} * \frac{1}{1}} = \frac{1}{3}$$

(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:

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

Note that A^C means the complement of event C above.

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 $\frac{1}{2}$? *Hint: Use the union bound to approximate. Recall, the union bound states that* $P(A_1 \cup A_2 \cup ... \cup A_n) \leq P(A_1) + P(A_2) + ... + P(A_n)$.

Solution: Let A be the event that there are no collisions. We know we have $\binom{m}{2} = k$ 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 $m \leq \sqrt{n}$.

4 Expectation and Variance of 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. Expectation is usually denoted with the symbol μ.

Linearity of Expectation: For any random variables $X_1, X_2, ...X_n$, expectation is linear, i.e.:

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

This is true even when these random variables aren't independent.

Variance: The variance of a random variable X is defined as

$$\mathrm{Var}(X) = E((X - E(X))^2) \ = E(X^2) - E(X)^2$$

The latter version of variance is the one we usually use in computations. The square root of Var(X) is called the standard deviation of X. It is usually denoted with the variable σ.

4.2 Questions

1. Show that $E(aX + b) = aE[X] + b$ and that $var(aX + b) = a^2 var(X)$, where X is any random variable.

Solution: First, for the expectation:

$$E[aX + b] = \sum_{x \in X}(aX + b)P(X = x) = a\sum_{x \in X} X \cdot P(X = x) + \sum_{x \in X} bP(X = x)$$

$$= aE[X] + b\sum_{x \in X} P(X = x) = aE[X] + b$$

For the variance:

$$\begin{aligned} var(aX+b) &= E(aX+b-E(aX+b))^2 \\ &= E(aX+b-aE[X]-b)^2 \\ &= E(aX-aE[X])^2 \\ &= a^2E(X-E[X])^2 = a^2var(X) \end{aligned}$$

2. Suppose X is a random variable. Does X always have to take on the value $E(X)$ at some point?

Solution: No. Consider X with probability mass function $P(X=x)=0.1x$ for $x=1,9$ and 0 elsewhere. The expectation of X is $0.1*1+0.9*9=8.2$, however X never takes on the value 8.2.

3. Given the random variable X defined as taking on the value 1 with probability 0.25, 2 with probability 0.5, and 20 with probability 0.25, what is the expectation of X? The variance of X?

Solution: $E(X)=0.25*1+0.5*2+0.25*20=6.25$
To find $var(X)$, we will first find $E(X^2)$. $E(X^2)=.25*1^2+.5*2^2+.25*20^2=102.25$
Then, $var(X)=E(X^2)-E(X)^2=102.25-(6.25)^2=63.19$.

5 Distributions

5.1 Bernoulli Distribution

Bernoulli Distribution: Bernoulli(p)

We say X has the Bernoulli distribution if it takes on value 1 with probability p, and value 0 with probability $1-p$. With the Bernoulli distribution we can model a single countable event, i.e. a single coin flip.
Expectation:

$$E(X)=0*(1-p)+1*p=p$$

Variance:

$$var(X)=E(X^2)-E(X)^2=0^2*(1-p)+1^2*p-p^2=p(1-p)$$

5.2 Binomial Distribution

Binomial Distribution: Bin(n, p)

The binomial distribution counts the number of successes when we conduct n independent trials. Each trial has a probability p of success. For this reason, we can think of the binomial distribution as a sum of n independent Bernoulli trials, each with probability p.

The probability of having k successes:

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

For example, if we flip a fair coin 10 times, the probability of 6 heads is

$$P(H = 6) = \binom{10}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4$$

Expectation:
If we were to compute the sum the traditional way, we would have to compute the sum

$$E(X) = \sum_{x \in X} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

Instead of doing that, we can use the fact that the binomial distribution is the sum of n independent Bernoulli distributions:

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

And now use linearity of expectation:

$$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = p + p + ... + p = np$$

Variance:
We know that variance is only separable when variables are mutually independent, i.e. $var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n)$ only when $X_1, X_2, ...X_n$ are mutually independent. Since our sum of Bernoulli trials is independent, we can do the following:

$$\begin{aligned} var(X) &= var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n) \\ &= p(1-p) + p(1-p) + ... + p(1-p) = np(1-p) \end{aligned}$$

5.3 Poisson Distribution

Poisson Distribution: Pois(λ) The Poisson distribution is an approximation of the binomial distribution under two conditions:

- n is very large
- p is very small

Let $\lambda = np$ represent the "rate" at which some event occurs. We usually use this distribution when these events are rare, such as a lightbulb failing.

The probability of k occurrences is

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

It turns out that the expectation and variance of the Poisson distribution are both equal to λ. This will be clear after we walk through the derivation of the Poisson distribution.

Derivation:
Recall, $\lambda = np$. Also, recall from calculus we have $\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x$, implying that $\lim_{n\to\infty}\left(1 - \frac{\alpha}{n}\right)^n = e^{-\alpha}$. We will also use the fact that for large n, $\frac{n!}{(n-k)!} \approx n^k$. We will use these facts below.

$$\begin{aligned}
\mathrm{P}[X = k] &= \binom{n}{k} * p^k * (1-p)^{n-k} && (1)\\
&= \frac{n!}{k! * (n-k)!} * p^k * (1-p)^{n-k} && (2)\\
&\approx \frac{n^k * p^k}{k!} * (1 - \frac{\lambda}{n})^{n-k} && (3)\\
&\approx \frac{\lambda^k * e^{-\lambda}}{k!} && (4)
\end{aligned}$$

Since we started with a binomial distribution, our expectation and variance should remain the same.

Expectation:
Since the expectation of a binomial is np, and we set $\lambda = np$, our expectation is also $E(X) = np$. We can also show this from scratch:

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

$$= e^{-\lambda} * \lambda * \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$
$$= e^{-\lambda} * \lambda * \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!}$$
$$= e^{-\lambda} * \lambda * e^{\lambda}$$
$$= \lambda$$

Variance:
For variance, it is much easier to start with the binomial case and reason from there. The variance of a binomial is $np(1-p)$, which looks like $\lambda(1-p)$. However, we started with the assumption that p is very small, so we can assume $(1-p)$ is very close to 1 and thus $\lambda(1-p)$ is very close to λ. Therefore, $var(X) = \lambda$.

5.4 Geometric Distribution

Geometric Distribution: Geom(p)

With the geometric distribution, we count the number of failures until the first success. For example, we could count the number of rolls of a dice until we roll a 6. The probability that the first success occurs on trial k is:

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

In what way can we derive the geometric distribution from the binomial distribution?

Expectation:
We know that $E(X)$ is the number of trials until the first success occurs, including that first success. There are two cases:

1. The first success occurs, with probability p
2. We obtain a failure, with probability $1-p$, meaning that we are back where we started but already used one trial

Putting this together, we get:

$$E(X) = p * 1 + (1-p) * (1 + E(X)) \implies E(X) = \frac{1}{p}$$

Variance:

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

5.5 Questions

1. In this problem, we will explore how we can apply multiple distributions to the same problem.

   Suppose you are a professor doing research in *machine learning*. On average, you receive 12 emails a day from students wanting to do research in your lab, but this number varies greatly.

   (a) Which distribution would you use to model the number of emails you receive from students on any one day?

   Solution: Poisson with parameter $\lambda = 12$.

   (b) What is the probability that you receive 7 emails tomorrow? At least 7?

   Solution: The probability we receive exactly 7 emails tomorrow is

   $$P(X = 7) = \frac{e^{-12} 12^7}{7!} \approx 0.0437$$

   The probability we receive at least 7 emails tomorrow is

   $$P(X \geq 7) = P(X = 7) + P(X = 8) + ... = e^{-12} \sum_{k=7}^{\infty} \frac{12^k}{k!}$$

   Equivalently, we can calculate it as:

   $$P(X \geq 7) = 1 - P(X \leq 6)$$

   $$= 1 - P(X = 0) - P(X = 1) - ... - P(X = 5) - P(X = 6)$$

   which gets rid of the infinite sum.

   (c) Now, let's look at the month of April, in which lots of students are emailing you to secure a summer position. What is the probability that the first day in April that you receive exactly 15 emails is April 7th? *Hint: Break this problem down into parts, and assign your result to the first part to the variable p.*

   Solution: It is worth mentioning that "receiving exactly 15 emails in one day" is an event, and either it happens or it does not. We will use the geometric distribution to model this. First, though, we need to find the probability p:

$$p = e^{-15}\frac{12^{15}}{15!} \approx 0.003604$$

Now, for days April 1, April 2, ... April 7, we know that we receive some number of emails that isn't 15, followed by receiving exactly 15 emails on April 8. This corresponds to 7 failures and 1 success in geometric:

$$P(\text{April 8th is first day with exactly 15 emails}) = (1-p)^7 p$$
$$\approx 0.996396^7 * 0.003604 \approx 0.003514$$

(d) Now, calculate the probability that April 8th is the first day that we receive **at least** 15 emails.

Solution: Our geometric model is the same, but we have a different p now.

$$p = e^{-12}\sum_{k=15}^{\infty}\frac{12^k}{k!} \approx 0.22798$$

$$P(\text{April 8th is first day with at least 15 emails}) = (1-p)^7 p$$
$$\approx (0.77202)^7 * 0.22798 \approx 0.03726$$

(e) What is the probability that you receive at least 15 emails on 10 different days in April?

Solution: We can take $p = 0.22798$ from the previous part. Now, we will count with the binomial distribution we have 30 trials (one for each day in April), and each is "successful" with probability p. We want the probability of exactly 10 "successes". Let X be the random variable that counts the number of days that we receive at least 15 emails.

$$P(X = 10) = \binom{30}{10} p^{10}(1-p)^{20}$$
$$\approx 0.06446$$

(f) What is the probability that you receive at least 15 emails on at least 15 days in April?

Solution:

$$P(X \geq 15) = P(X = 15) + P(X = 16) + ... + P(X = 30)$$
$$= \sum_{k=15}^{30} \binom{30}{k} 0.22798^k 0.77202^{30-k} \approx 0.00102$$