

RANDOM VARIABLES, CONDITIONAL PROBABILITY, DISTRIBUTIONS

COMPUTER SCIENCE MENTORS 70

October 24 to October 28, 2016

1 Conditional Probability

1.1 Introduction

Bayes' Rule

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

Total Probability Rule

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

Independence

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

$$P[A \cap B] = P[A] * P[B]$$

1.2 Questions

1. 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:

$$P(\text{Dog In } A \mid \text{Searched } A \text{ and Failed})$$

$$= \frac{P(\text{Dog In } A \cap \text{Searched } A \text{ and Failed})}{P(\text{Dog In } A \cap \text{Searched } A \text{ and Failed}) + P(\text{Dog Not In } A)}$$

$$= \frac{\frac{4}{10} * \frac{3}{4}}{\frac{4}{10} * \frac{3}{4} + \frac{6}{10}} = \frac{1}{3}$$

- (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 N th 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 Expectation and Random Variables

2.1 Introduction

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

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

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

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

2.2 Questions

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

Solution: No

- 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: $x < 5$, $x \% 2 = 0$, $(x = 2 \text{ or } x = 7)$, 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: $E(X) = .25 * 1 + .5 * 2 + .25 * 20 = 6.25$

3 Variance

3.1 Introduction

For a random variable X with expectation $E(X) = \mu$, the variance of X is:

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

The square root of $\text{Var}(X)$ is called the standard deviation of X

Theorem: For a random variable X with expectation $E(X) = \mu$ and a constant c ,

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

$$\text{Var}(cX) = c^2 * \text{Var}(X)$$

3.2 Questions

1. Let's consider the classic problems of flipping coins and rolling dice. Let X be a random variable for the number of coins that land on heads and Y be the value of the die roll.

- (a) What is the expected value of X after flipping 3 coins? What is the variance of X ?

Solution:

$$E(X) = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = \frac{3}{2}$$
$$E(X^2) = 0^2 * \frac{1}{8} + 1^2 * \frac{3}{8} + 2^2 * \frac{3}{8} + 3^2 * \frac{1}{8} = \frac{24}{8} = 3$$

$$E(X)^2 = \frac{9}{4}\text{Var}(X) = 3 - \frac{9}{4} = \frac{3}{4}$$

(b) Let Y be the sum of rolling a dice 1 time. What is the expected value of Y ?

Solution: $E(Y) = \frac{1}{6} * (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$

(c) What is the variance of Y ?

Solution: $E(Y^2) = [\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)] = \frac{91}{6}$ $\text{Var}(Y) = E(Y^2) - (E(Y))^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$

2. Say you're playing a game with a coin and die, where you flip the coin 3 times and roll the die once. In this game, your score is given by the number of heads that show multiplied with the die result. What is the expected value of your score? What's the variance?

Solution: $E(XY) = E(X)E(Y) = \frac{21}{4}$ since X and Y are independent. $\text{Var}(XY) = E(X^2Y^2) - E(XY)^2 = E(X^2)E(Y^2) - E(X)^2E(Y)^2 = 3 * \frac{91}{6} - \frac{3^2}{2} * \frac{7^2}{2} = \frac{91}{2} - \frac{9}{4} * \frac{49}{4} = 17.9375 = \frac{287}{16}$

3. You are at a party with n people where you have prepared a red solo cup labeled with their name. Before handing red cups to your friends, you pick up each cup and put a sticker on it with probability $\frac{1}{2}$ (independently of the other cups). Then you hand back the cups according to a uniformly random permutation. Let X be the number of people who get their own cup back AND it has a sticker on it.

(a) Compute the expectation $E(X)$.

Solution: Define $X_i = 1$ if the i -th person gets their own cup back and it has a sticker on it 0 otherwise Hence $E(X) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$ $E(X_i) = P[X_i = 1] = \frac{1}{2n}$ since the i -th student will get his/her cup with probability $\frac{1}{n}$ and has a sticker on it with probability $\frac{1}{2}$ and stickers are put independently. Hence $E(X) = n \cdot \frac{1}{2n} = \frac{1}{2}$.

(b) Compute the variance $\text{Var}(X)$

Solution: To calculate $\text{Var}(X)$, we need to know $E(X^2)$

$$E(X^2) = E(X_1 + X_2 + \dots + X_n)^2 = E\left(\sum_{i,j} (X_i * X_j)\right) = \sum_{i,j} (E(X_i * X_j))$$

(by linearity of expectation)

Then we consider two cases, either $i = j$ or $i \neq j$. Hence

$$\sum_{i,j} E(X_i * X_j) = \sum_i E(X_i^2) + \sum_{i \neq j} E(X_i * X_j)$$

$E(X_i^2) = \frac{1}{2n}$ for all i . To find $E(X_i * X_j)$, we need to calculate $P[X_i X_j = 1]$. $P[X_i * X_j = 1] = P[X_i = 1]P[X_j = 1 | X_i = 1] = \frac{1}{2n} * \frac{1}{2*(n-1)}$ since if student i has received his/her own cup, student j has $n - 1$ choices left. Hence

$$E(X^2) = n * \frac{1}{2n} + n * (n - 1) * \frac{1}{2n} * \frac{1}{2 * (n - 1)} = \frac{3}{4}$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}.$$

4. True or False?

Assume that X is a discrete random variable. If $\text{Var}(X) = 0$, then X is a constant.

Solution: TRUE. Let $\mu = E[X]$. By definition,

$$0 = \text{Var}(X) = E[(X - \mu)^2] = \sum \omega P[\omega] (X(\omega) - \mu)^2$$

The RHS is the sum of non-negative numbers, so if the sum is 0, each term must be 0. So

$$P[\omega] > 0 \rightarrow (X(\omega) - \mu)^2 = 0 = X(\omega) = \mu.$$

Therefore X is constant (equal to $\mu = E[X]$).

4 Distributions

4.1 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:

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

The expectation of a geometric distribution is:

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

Solution: Derivation of $E(X)$: The clever way to find the expectation of the geometric distribution uses a method known as the renewal method. $E(X)$ 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 $E(X)$ more since nothing has changed from our original situation (the geometric distribution is memoryless). Hence $E(X) = p * 1 + (1 - p) * (1 + E(X))$

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:

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

The expectation of a binomial distribution is:

$$E(X) = np$$

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

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

Instead of doing that just use Bernoulli variables:

$$X = X_1 + \dots + X_n$$

And now use linearity of expectation:

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

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

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

The expectation of a poisson distribution is:

$$E(X) = \lambda$$

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

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

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

4.2 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 N namely, 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:

$$\begin{aligned} E(x) &= \frac{1}{p} = 999,999 \text{ tries} \\ (999,999 * 5 \text{ sec}) * \frac{1 \text{ min}}{60 \text{ sec}} * \frac{1 \text{ hr}}{60 \text{ min}} * \frac{1 \text{ day}}{24 \text{ hr}} &\approx 57.9 \text{ days} \end{aligned}$$

(c) What is the variance in the amount of time? (Answer in seconds, approximately.)

Solution:

$$\begin{aligned}\text{Var}(x) &= \frac{1-p}{p^2} = \frac{1 - \frac{1}{999,999}}{999,999^2} \\ &= 999,999 * 999,998 \\ &\approx 1 \text{ trillion tries}\end{aligned}$$

When we scale a variable, we scale the variance by the square of that factor, so

$$\text{Var}(x) = 25 \text{ trillion sec}^2$$

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

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:

$$P[X = 7] = \frac{\lambda^7}{7!} * e^{-\lambda} = \frac{3^7}{7!} * e^{-3} \approx 0.0216$$

$$P[X \geq 7] = \sum_{i=7}^{\infty} P[X = i] = \sum_{i=7}^{\infty} \frac{3^i}{i!} * e^{-3}$$

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

Solution:

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