

# DISTRIBUTIONS, VARIANCE, CONDITIONAL EXPECTATION, INTRO TO CONTINUOUS PROBABILITY 9

---

COMPUTER SCIENCE MENTORS 70

October 30 to November 4, 2017

---

## 1 Distributions

---

### 1.1 Introduction

---

**Geometric Distribution:  $\text{Geom}(p)$**  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:

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

The variance of a geometric distribution is:

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

**Binomial Distribution:  $\text{Bin}(n, p)$**  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:

$$E(X) = np$$

The variance of a binomial distribution is:

$$\text{Var}(X) = np(1 - p)$$

**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:

$$E(X) = \lambda$$

The variance of a poisson distribution is:

$$\text{Var}(X) = \lambda$$

## 1.2 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?
- (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.

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?

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

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

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?

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

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

## 2 Variance

### 2.1 Introduction

**Definition:** For a random variable  $X$ , expectation  $E(X) =$

$$\sum_a a * Pr[X = A]$$

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  $X$ . It is usually denoted with the variable  $\sigma$ .

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

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

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

### 2.2 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  $X$ ?

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

(c) What is the variance of  $Y$ ?

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?

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  $E(X)$ .

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

4. a. Prove that for independent random variables  $X$  and  $Y$ ,  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

- b. Is the above result true for non-independent random variables? Prove or give a counterexample.

5. Consider the random variable  $X = X_1 + \dots + X_n$ , where  $X_i$  equals  $i$  with probability  $\frac{1}{i}$  and 0 otherwise.

- (a) What is the variance of  $X$ ? (Assume that  $X_i$  and  $X_j$  are independent for  $i \neq j$ )

- (b) For what value of  $n$  does  $E(X) = \text{Var}(X)$ ?

- (c) For what value of  $n$  does  $E(X) = \text{SD}(X) * \sqrt{2} + 100$ ?

6. An urn contains  $n$  balls numbered  $1, 2, \dots, n$ . We remove  $k$  balls at random (without replacement) and add up their numbers. Find the mean and variance of the total.

## 3 Conditional Expectation

### 3.1 Introduction

The **conditional expectation** of  $Y$  given  $X$  is defined by

$$E[Y|X = x] = \sum_y y \cdot P[Y = y|X = x] = \sum_y y \cdot \frac{P[X = x, Y = y]}{P[X = x]}$$

#### Properties of Conditional Expectation

$$\begin{aligned} E(a|Y) &= a \\ E(aX + bZ|Y) &= a \cdot E(X|Y) + b \cdot E(Z|Y) \\ E(X|Y) &\geq 0 \text{ if } X \geq 0 \\ E(X|Y) &= E(X) \text{ if } X, Y \text{ independent} \\ E(E(X|Y)) &= E(X) \end{aligned}$$

### 3.2 Questions

1. Prove  $E(E(Y|X)) = E(Y)$



2. Prove  $E(h(X) \cdot Y|X) = h(X) \cdot E(Y|X)$

---

## 4 Continuous Probability

---

### 4.1 Questions

---

1. Given the following density functions, identify if they are valid random variables. If yes, find the expectation and variance. If not, what rules does the variable violate?

(a)  $f(x) = \begin{cases} \frac{1}{4} & \text{if } x \in \{\frac{1}{2}, \frac{9}{2}\} \\ 0 & \text{otherwise} \end{cases}$

(b)  $f(x) = \begin{cases} x - \frac{1}{2} & x \in \{0, \infty\} \end{cases}$

2. For a discrete random variable  $X$  we have  $\Pr[X \in [a, b]]$  that we can calculate directly by finding how many points in the probability space fall in the interval and how many total points are in the probability space. How do we find  $\Pr[X \in [a, b]]$  for a continuous random variable?

3. Are there any values of  $a, b$  for the following functions which gives a valid pdf? If not, why? If yes, what values?

(a)  $f(x) = -1, a < x < b$

(b)  $f(x) = 0, a < x < b$

(c)  $f(x) = 10000, a < x < b$

4. For what values of the parameters are the following functions probability density functions? What is the expectation and variance of the random variable that the function represents?

(a)  $f(x) = \begin{cases} ax & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

(b)  $f(x) = \begin{cases} -2x & \text{if } a < x < b \text{ (} a = 0 \vee b = 0 \text{)} \\ 0 & \text{otherwise} \end{cases}$

5. Define a continuous random variable  $R$  as follows: we pick a random point on a disk of radius 1; the value of  $R$  is distance of this point from the center of the disk. We will find the probability density function of this random variable.

- (a) What is (should be) the probability that  $R$  is between 0 and  $\frac{1}{2}$ ? Why?

(b) What is (should be) the probability that  $R$  is between  $a$  and  $b$ , for any  $0 \leq a \leq b \leq 1$ ?

(c) What is a function  $f(x)$ , for which  $\int_a^b f(x)dx$  satisfies these same probabilities?