1 Conditional Probability and Moments

So far, we have discussed the fundamental properties of probability that are drawn from basic set theory. We will continue to apply these practices when talking about the upcoming sections.

1.1 Conditional Probability and Bayes' Theorem

Let's use a six-sided die to motivate how conditional probability works. We know that the probability of rolling an odd number is 0.5. However, within the odd numbers, what is the chance of rolling a 3 or higher? We can merely compute the probability by counting the number of odd numbers between 3 to 6 and the number of odd numbers on a die.

$$P(\text{roll 3 or higher} \mid \text{odd}) = \frac{\text{cardinality of the set } \{3, 5\}}{\text{cardinality of the set } \{1, 3, 5\}} = \frac{2}{3}.$$

The notation reads: "the probability that a 3 or higher was rolled **given** the number rolled was odd."

Example 1.1. In a group of 635 men who died in 1990, 160 of the men died from causes related to heart disease. Moreover, 275 of the 635 men had at least one parent who suffered from heart disease, and of those 275 men, 95 died from causes related to heart disease. Find the probability that a man randomly selected from this group died of causes not related to heart disease, **given** that neither of his parents suffered from heart disease.

95	180	At least 1 parent with HD	275
65	295	Neither parent with HD	360
HD 160	No HD 475		

By constructing a simple table and some quick math, we can summarize our sample space. We need only look at the bottom row as given by the question. So, out of the 360 men whose parents never had HD, 295 of them died to a cause that was not related to HD. Therefore, the probability is simply

$$P(\text{No HD} \mid \text{Neither parent with HD}) = \frac{295}{360} \approx 81.94\% \text{ or } 0.8194.$$

Definition 1.2. Let A and B be two events. The **conditional probability** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}.$$

Rearranging terms from the definition gives

$$P(B) \cdot P(A|B) = P(AB) = P(A) \cdot P(B|A).$$

Example 1.3. The blood pressure (high, low, or normal) and heartbeats (regular or irregular) of a random sample of patients are measured. Of the patients,

1. 36% have high blood pressure and 16% have low blood pressure.
2. 21% have an irregular heartbeat.
3. Of those with an irregular heartbeat, one-third have high blood pressure.
4. Of those with normal blood pressure, one-eighth have an irregular heartbeat.

What portion have a regular heartbeat and low blood pressure?

By constructing and filling out a table, we can identify the percentage rather quickly. We also need to verify that the sum of entries add up to 1.

0.08	0.42	0.29	0.79 regular
0.08	0.06	0.07	0.21 irregular
0.16 low	0.48 normal	0.36 high	

Definition 1.4. Events A and B are **independent** if $P(A \cap B) = P(A) \cdot P(B)$
Intuitively, independence means $P(A) = P(A|B)$ and $P(B) = P(B|A)$ so knowing if A or B occurred gives no information on whether or not the event occurred.

Say A and B are two independent events. If we want to find the probability of A given B, we already know that B occurring does not influence the outcome of A, so the probability is just $P(A)$.

Example 1.5. Suppose A and B are independent events with $P(A) = 0.6$ and $P(A \cap B) = 0.3$. Find $P(B)$ and $P(A|B)$.
By independence, we have

$$P(A \cap B) = P(A) \cdot P(B)$$

$$0.3 = 0.6P(B) \iff \boxed{P(B) = 0.5}$$

As alluded to earlier, $P(A|B) = P(A) = 0.6$.

Example 1.6. A and B are events such that $P(A) = 0.4, P(B) = 0.1$, and $P(A \cap B) = 0.05$. Are they independent? What is $P(B|A)$?

$$P(A) \cdot P(B) = 0.4 \cdot 0.1 = 0.04 \neq 0.05 = P(A \cap B) \quad \text{Not independent}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.05}{0.4} = \boxed{0.125}$$

Example 1.7. If $P(A) = 0.2$ and $P(B) = 0.3$, find $P(A \cup B)$ if (a) the events are independent and (b) the events are mutually exclusive.
(a)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.3 - (0.2 \cdot 0.3) = 0.44$$

(b)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.3 - 0 = 0.5$$

Let's revisit conditional probability, namely this equation

$$P(A) \cdot P(B|A) = P(A \cap B).$$

This equation can be thought of a sequence of events: first we need A to occur, and then second we need B to occur, taking into account the fact that A occurred.

Example 1.8.

Find the probability of having a flush after being dealt five cards from a standard deck? Recall that a flush contains at least 5 cars of the same suit. Assume we use a standard deck: 52 cards, with 4 suits and 13 ranks.

We present two approaches to reach the same answer.

We can count the probability of 5 independent events, and multiply by 4 for the total number of unique suits. The probability of picking a card from one suit is $\frac{13}{52}$. Since there are now 51 cards and 12 cards of that suit, the chance of pulling another card from that suit is $\frac{12}{51}$. Repeating this process yields

$$4 \cdot \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \cdot \frac{10}{49} \cdot \frac{9}{48} \approx 0.198\%.$$

Similarly, $4 \cdot \frac{13}{52} = 1$ is simply the probability of picking any card, and the resulting probabilities are from picking the same suit.

The following examples look at variations of the previous problem.

Example 1.9.

If exactly three of the first 5 cards dealt are spades, what is the probability of being dealt a flush in the first 7 cards?

Since 5 cards have already been dealt, there are 47 cards left in the deck with 10 spades in there. If we need a flush in the first 7 cards, the next 2 cards dealt must spades. We can compute the probability of cards 6 **and** 7 being spades as

$$P(\text{next 2 cards are spades}) = P(\text{6th card is a spade}) \cdot P(\text{7th is a spade} \mid \text{6th is a spade})$$
$$= \frac{10}{47} \cdot \frac{9}{46} \approx 0.0416 = 4.16\%$$

Example 1.10.

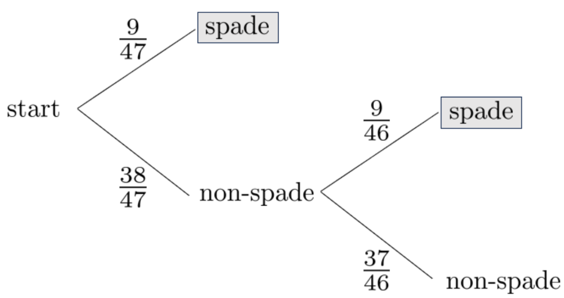
If exactly four of the first 5 cards dealt are spades, what is the probability of being dealt a flush in the first 7 cards?

We consider two methods of solving this problem:

Method 1: We are searching for the probability that at least one of the next two cards. One possibility is to consider the probability of its complement−that neither of the next two cards are spades−and subtract it from 1. We would have

$$P(\text{flush in 7 cards}) = 1 - P(\text{next two cards are not spades}) = 1 - \left(\frac{38}{47} \cdot \frac{37}{46}\right) = 0.35 = 35\%$$

Method 2: Use a tree diagram to write out the possible outcomes.

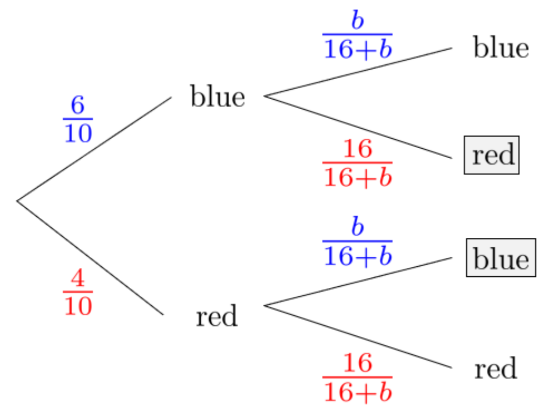


According to the diagram, there are two ways to complete a flush. Once by getting a spade on the 6th card, or getting a non-spade and a spade right after. Their respective probabilities are $\frac{9}{47}$ and $\frac{38}{47} \cdot \frac{9}{46}$. Add them together:

$$\frac{9}{47} + \left(\frac{38}{47} \cdot \frac{9}{46}\right) = \boxed{0.35}$$

Example 1.11. An urn contains 10 balls: 4 red and 6 blue. A second urn contains 16 red balls and an unknown of blue balls. A single ball is drawn from each urn. The probability that both balls are different colors is 0.528. Calculate the number of blue balls in the second urn.

As with the previous example, we will draw a tree diagram to describe all four outcomes.

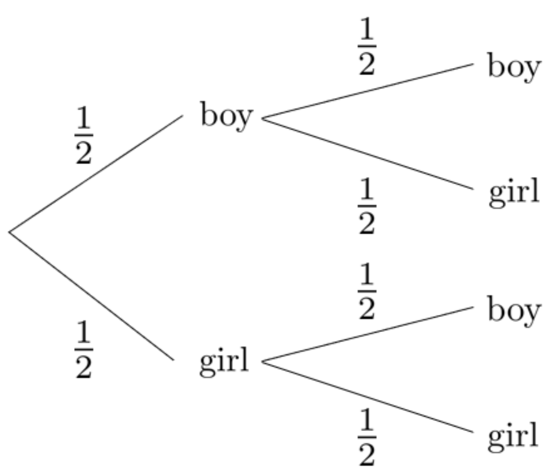


The expressions on the second branch describe the probability given an unknown quantity of blue balls b. There are two possibilities that yield the desired outcome: pulling a blue ball then a red ball, or a red ball then blue ball. Mathematically, we can write this as

$$0.528 = \frac{6}{10}\left(\frac{16}{16+b}\right) + \frac{4}{10}\left(\frac{b}{16+b}\right)$$

Through expansion and rearranging, we find that $\boxed{b = 9}$.

Example 1.12. A family has two children, and they are not twins. Given that at least one of the children is a boy, what is the probability that both children are boys?



Contrary to the previous two examples, the events are independent of each other (i.e. having a boy does not affect the probability of having another boy). By conditional probability, we can write

$$P(2 \text{ boys} \mid \text{at least 1 boy}) = \frac{P(2 \text{ boys})}{P(\text{at least one boy})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Or, we can look at the tree diagram and observe that 3 outcomes have at least one boy (the top 3 on the second branch). Out of these three outcomes, only one results in two boys, thus giving us the probability of $\frac{1}{3}$.

These examples described how probability works in sequences of events, or finding the probability given the multiple outcomes of two or more events.

Once again, let's revisit the formula for conditional probability. Suppose we want to find $P(A|B)$ but we are given $P(B|A)$ instead. By rewriting $P(A \cap B) = P(A)P(B|A)$, we can obtain the simplest form of Bayes' Theorem:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Example 1.13. Below is a table relating one's age range, likelihood of a car accident, and the proportion of drivers in each age range.

Age of Driver	Probability of Accident	Portion of Company's Insured Driver
16-20	0.06	0.08
21-30	0.03	0.15
31-65	0.02	0.49
66-99	0.04	0.28

Given that the driver got into an accident, what is the probability that the driver's age is between 31 and 65?

We will use Bayes' Formula to solve this. By doing so, we must uncover 3 probabilities.

$P(\text{age 31-65}) = 0.49$ and $P(\text{accident} \mid \text{age 31-65}) = 0.02$ by the table. We also need to compute the probability of an accident occurring. This is done by multiplying the accident probability by the respective proportion for each age range, and summing them up:

$$P(\text{accident}) = \sum_{\text{age groups}} P(\text{accident, age group})$$

$$= (0.08)(0.06) + (0.15)(0.03) + (0.49)(0.02) + (0.28)(0.04) = 0.0303$$

Plugging in our known values,

$$P(\text{age 31-65} \mid \text{accident}) = \frac{P(\text{age 31-65, accident})}{P(\text{accident})} = \frac{P(\text{age 31-65}) \cdot P(\text{accident} \mid \text{age 31-65})}{P(\text{accident})}$$

$$= \frac{(0.49)(0.02)}{0.0303} = \boxed{0.3234 \text{ or } 32.34\%}.$$

The computation used to find the probability of an accident is a common result of the upcoming theorem.

Theorem 1.14 (Law of Total Probability). *If $A_1, A_2, \ldots, A_k$ are disjoint and $P(A_1) + P(A_2) + \cdots + P(A_k) = 1$ then*

- $P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k)$
- $P(B) = P(A_1)P(B|A_1) + \cdots + P(A_k)P(B|A_k)$

The sets $A_1, \ldots, A_k$ are called a ***partition*** *of the sample space. We will often refer to them as a list of all possible cases.*

In the previous example, the age groups were the A_i and B was the event of an accident.

The previous theorem is crucial to generalize Bayes' Theorem to multiple events and groups.

Theorem 1.15 (Bayes' Theorem). *Suppose $A_1, \ldots, A_k$ are a partition of the sample space. Then*

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{k} P(B \cap A_i)} = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}$$

The final denominator sums one event over all cases.

Bayes' Theorem tells us the probability of a past event occurring given a present observation, making it a crucial role for inverting mathematical probabilities. In Bayesian terms, $P(A_1|B)$ is the **posterior** probability and $P(A_i)$ is the **prior**. Additionally, the denominator is the result found in Theorem 1.14!

Example 1.16. Life insurance policy holders are categorized as standard, preferred, and ultra-preferred. Of a company's policyholders, 50% are standard, 40% are preferred, and 10% and ultra-preferred. The probability of dying in the next year is 0.01 for each standard policyholder, 0.005 for preferred policyholders, and 0.001 for ultra-preferred. A policyholder dies in the next year. What is the probability that the deceased policyholder was standard?

Let S denote someone who is standard

$$P(S \mid \text{died}) = \frac{P(S \cap \text{died})}{P(\text{died})} = \frac{(0.5)(0.01)}{(0.5)(0.01) + (0.4)(0.005) + (0.1)(0.001)} \approx \boxed{70.4\%}$$

Example 1.17. Taxicabs in Crobuzon are all either green or blue. On Tuesday, a taxicab got into an accident. A witness to the accident thought that the cab involved was blue, and further tests showed that the witness has an 80% chance of correctly identifying the color of a taxicab, independently of its color. If 85% of the taxicabs on the streets on Tuesday were green, what was the probability that the taxicab involved in the accident was blue?

$$P(\text{Blue cab} \mid \text{Witness said blue}) = \frac{P(\text{Blue cab and witness said blue})}{P(\text{Witness said blue})}$$

$$= \frac{(0.15)(0.8)}{(0.15)(0.8) + (0.85)(0.2)} \approx \boxed{41\%}$$

Recall that we use 0.15 as the complement of the 85% of green taxicabs, and 0.2 as the complement of the 80% correct identifications. The probability of a witness saying blue is the linear combination of blue cars · correct identification and green cars · incorrect identification.

1.2 Discrete Moments

In this section, we'll go over basic characteristics of probability distributions of random variables, or the primary measures of central tendency. Generally, the most common measures of a random variable are:

- Mean = average value
- Median = "middle" value
- Mode = average value

Median and mode are fairly straightforward to find, however computing the mean can be more complicated, depending on the distribution.

Definition 1.18. We say X is a **random variable** if it is a number whose value depends on chance. More formally,

$$X : S \to \mathbb{R} \text{ where } S \text{ is the sample space}$$

X is a **discrete random variable** if we can list all of the possible values. For discrete variables,

$$1 = \sum_x P(X = x)$$

Typically we use capital letters for random variables and lower case letters for possible, or non-random, values.

Definition 1.19. For a discrete random variable X, y is the **mode** of X if $P(X = y) \geq P(X = x)$ for all x (i.e. the mode y is the input that maximizes $P(X = y)$).

The mode is NOT unique, a random variable X can have multiple modes, but it will always have at least one.

Example 1.20. Suppose I roll an otherwise fair 7 sided die whose faces are 1, 1, 1, 2, 4, 4, and 6. Find the mode.

Let X be the result of the roll. Then,

$$P(X = 1) = \frac{3}{7} \quad P(X = 2) = \frac{1}{7} \quad P(X = 4) = \frac{2}{7} \quad P(X = 6) = \frac{1}{7}$$

The mode is 1 because when $y = 1$, $P(X = y)$ reaches its max of $\frac{3}{7}$.

The following example uses a common random variable that has not yet been covered.

Example 1.21. Find the mode of a Poisson random variable with mean 2.8, meaning that

$$P(N = n) = \frac{2.8^n}{n!}e^{-2.8} \text{ for } n = 0, 1, 2, \ldots$$

We can plot the above function using a graphic calculator and construct the table below:

n	0	1	2	3	4	5	6
$P(N = n)$	0.0608	0.1703	0.2384	0.2225	0.1557	0.0872	0.0407

The table tells us that the mode is 2, for it maximizes $P(N = y)$.

Before introducing the median, we go over the cumulative distribution function, or the probability that a random variable will take a value less than or equal to.

Definition 1.22. Let X be a random variable. The function $F(x) = P(X \leq x)$ is the **cumulative distribution function (CDF)** of X.

For example, $F(2) = P(X \leq 2)$, or the probability that the value will be less than 2. Say k is the maximum value of X. Then $P(X \leq k) = 1$.

Definition 1.23. The **median** of a random variable X is the smallest m such that $P(X \leq m) = F(m) \geq \frac{1}{2}$.

Remark: This is more of a black-box definition and does not cover all corner cases. For instance, the median may not be uniquely defined for some random variables.

Example 1.24. Suppose I roll an otherwise fair 7 sided die whose faces are 1, 1, 1, 2, 4, 4, and 6. Find the median.

Recall the probabilities found in Example 1.20. We will use these to evaluate the CDF for $X = 1, 2, 4$, and 6

$$P(X = 1) = \frac{3}{7} \qquad P(X \leq 1) = \frac{3}{7}$$

$$P(X = 2) = \frac{1}{7} \qquad P(X \leq 2) = \frac{4}{7}$$

$$P(X = 4) = \frac{2}{7} \qquad P(X \leq 4) = \frac{6}{7}$$

$$P(X = 6) = \frac{1}{7} \qquad P(X \leq 6) = 1$$

Since $P(X \leq 1) \leq \frac{1}{2}$ and $P(X \leq 2) \geq \frac{1}{2}$, by Def 1.23, the median is 2.

Definition 1.25 (Percentile). The $100\% \cdot p^{th}$ percentile π_p is the smallest possible x such that $P(X \leq x) \geq p$.

Let's continue the 7-sided die from Examples 1.20 and 1.24. If X is our die and F is our CDF:

x	0	1	2	4	6
$F(x)$	0	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{6}{7}$	1

We make the following observations

1. 5th percentile = 1 since $P(N \leq 1) \geq 0.05$ but $P(N \leq 0) < 0.05$
2. 10th percentile = 1 since $P(N \leq 1) \geq 0.1$ but $P(N \leq 0) < 0.1$
3. Median = 50th percentile = 2
4. 90th percentile = 6 since $P(N \leq 4) < 0.9$ but $P(N \leq 6) \geq 0.9$

Percentiles are a generalized extrapolation of the median, where $0 \leq m \leq 1$. Recall that the 50th percentile is when $m = \frac{1}{2}$ and also the median.

The following example dives into one of the corner cases alluded to earlier.

Example 1.26. Roll a fair 6-sided die. What is the median outcome?

n	$P(N = n)$	$F(n) = P(N \leq n)$
1	1/6	1/6
2	1/6	2/6
3	1/6	1/2
4	1/6	4/6
5	1/6	5/6
6	1/6	1

By our definition, 3 is the first time $F(n) \geq \frac{1}{2}$ so the median is 3.
But $F(x) = \frac{1}{2}$ for any x such that $3 \leq x < 4$ (e.g. $F(3.5) = 1/2$).
This suggests infinitely many medians, for anything in the interval $3 \leq x < 4$ would qualify as one.

Example 1.27. Suppose that $P(N = n) = \frac{n}{15}$ for $n = 1, 2, 3, 4, 5$. Find the median of N.

Evaluating the CDF:

n	$P(N = n)$	$F(n) = P(N \leq n)$
1	1/15	1/15
2	2/15	1/5
3	1/5	2/5
4	4/15	2/3
5	1/3	1

Since $P(N \leq 3) < \frac{1}{2}$ and $P(N \leq 4) \geq \frac{1}{2}$, we say the median is $\boxed{4}$.

We now motivate the mean with a simple example. Suppose that, in a group of 10 people, I owe \$4 to two of them, \$2 to one of them, \$1 to one of them, and nothing to the others. On average, how much do I owe to these 10 people?

We can straightforwardly compute this as the total amount owed divided by the number of people:

$$\frac{(4)(2) + (2)(1) + (1)(2)}{10} = \frac{12}{10} = \$1.20$$

For simple problems such as these, we can compute the mean using a fraction.

Definition 1.28 (Expected Value). If X is a discrete random variable, then

$$E[X] = \sum_x x \cdot P(X = x)$$

$E[X]$ is the **expected value**, or the mean, of X. More generally, let $g(X)$ be a function of random variable X. Then,

$$E[g(X)] = \sum_x g(x) \cdot P(X = x)$$

Example 1.29. An insurance policy pays 100 per day for up to 3 days of hospitalization and 50 per day of hospitalization thereafter. Find the expected payment for hospitalization if the number of days of hospitalization, X, is a discrete random variable with

$$P(X = k) = \begin{cases} \frac{6-k}{15} & \text{for } k = 1, 2, 3, 4, 5 \\ 0 & \text{otherwise} \end{cases}$$

Let $g(k)$ = payment for k days in the hospital

$$E[g(x)] = \sum g(k) P(X = k)$$

k	1	2	3	4	5
$F(x)$	5/15	4/15	3/15	2/15	1/15

Note that insurance pays \$100 for the first day, \$200 for the second, \$300 for the third, \$350 for the fourth, and \$400 for the fifth. The expected value is the cumulative payment on the k-th day multiplied by the probability on the k-th day. Or, more precisely,

$$E[g(X)] = 100\left(\frac{5}{15}\right) + 200\left(\frac{4}{15}\right) + 300\left(\frac{3}{15}\right) + 350\left(\frac{2}{15}\right) + 400\left(\frac{1}{15}\right) = \boxed{\$220}$$

Example 1.30. Suppose that $P(N = n) = \frac{n}{15}$ for $n = 1, 2, 3, 4, 5$. Find $E[N]$

$$E[N] = \sum_{n} nP(N = n) = \frac{1}{15} + \frac{4}{15} + \frac{9}{15} + \frac{16}{15} + \frac{25}{15} = \boxed{\frac{55}{15} = \frac{11}{3}}$$

Remember the repayment example from earlier:

Suppose that, in a group of 10 people, I owe \$4 to two of them, \$2 to one of them, \$1 to one of them, and nothing to the others. On average, how much do I owe to these 10 people?

We used the most direct definition of averages to solve this. Had we used expected value, the answers should agree. What if we paid the money in steps?

1. Pay \$1 to everyone who is owed money.
2. Pay \$1 more to everyone who is still owed money (i.e. people who were initially owed \$2 or \$4), repeating in \$1 increments until everyone has been paid in full.

The total payment would be \$5 in step 1, \$3 in step 2, \$2 in step 3, and \$2 in step 4, in which everyone will have been paid after. Dividing the total payment by 10 yields the same average as before.

Let's try to generalize this to infinitely many steps. If $N \geq 0$ and N is an integer valued variable,

$$E[N] = P(N > 0) + P(N > 1) + P(N > 2) + \ldots = \sum_{n=0}^{\infty} P(N > n)$$

By letting $k = n + 1$ we have

$$P(N > n) = P(N \geq n + 1) = P(N \geq k) \Longrightarrow E[N] = \sum_{k=1}^{\infty} P(N \geq k)$$

This is mainly done to clean up notation and replace $>$ with $\geq$. The function $P(N > n) = 1 - F(n)$ is called the **survival function** and was appropriately given its name to approximate the probability that something will last longer than expected. The discrete representation of the survival function doesn't hold up as well whereas the continuous form does.

Example 1.31. Following a certain type of surgery, patients are hospitalized for N days, with $P(N \geq k) = \frac{5-k}{5}$ for $k = 0, 1, 2, 3, 4, 5$. Find $E[N]$ using the survival method.

$$E[N] = \sum_{n=0}^{\infty} P(N > n) = P(N > 0) + P(N > 1) + P(N > 2) + P(N > 3) + P(N > 4)$$

$$= P(N \geq 1) + \ldots + P(N \geq 5) = \frac{4}{5} + \frac{3}{5} + \frac{2}{5} + \frac{1}{5} + 0 = \boxed{2}$$

What if we found $E[N]$ through its definition? Recall that

$$P(N = k) = P(N \geq k) - P(N \geq k+1) = \frac{5-k}{5} - \frac{5-(k+1)}{5} = \frac{1}{5}$$

$$E[N] = \left(0 \cdot \frac{1}{5}\right) + \left(1 \cdot \frac{1}{5}\right) + \left(2 \cdot \frac{1}{5}\right) + \left(3 \cdot \frac{1}{5}\right) + \left(4 \cdot \frac{1}{5}\right) = 2$$

Now, we will explore an important metric that is often used in conjunction with the mean. Suppose X, Y, and Z are three random variables such that

$$P(X = 2) = 2$$

$$P(Y = 1) = \frac{1}{3} \quad P(Y = 2) = \frac{1}{3} \quad P(Y = 3) = \frac{1}{3}$$

$$P(Z = 1) = \frac{1}{2} \quad P(Z = 3) = \frac{1}{2}$$

Then $E[X] = E[Y] = E[Z] = 2$. However, Y is more likely to deviate from the mean than X, and Z is even more likely to do so. Here we just compared the variance between the random variables!

Definition 1.32. The **variance** of a variable quantifies how much it differs from its mean. Given a discrete random variable X and its expected value $E[X]$,

$$\text{Var}(X) = E[(X - E[X])^2] \text{ or } \sum_k P(X = k)(X_k - E[X])^2$$

Alternatively, we can write it as

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

Let's use both definitions to compute $\text{Var}(X), \text{Var}(Y), \text{Var}(Z)$:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = X^2 P(X = 2) - 2^2 = 4(1) - 4 = 0$$

$$\text{Var}(Y) = E[(Y-\mu_Y)^2] = \frac{1}{3}(1-2)^2 + \frac{1}{3}(2-2)^2 + \frac{1}{3}(3-2)^2 = \frac{2}{3}$$

$$\text{Var}(Z) = E[(Z-\mu_Z)^2] = \frac{1}{2}(1-2)^2 + \frac{1}{2}(3-2)^2 = 1$$

Definition 1.33 (Moments).

$E[X^k]$ is the k-th moment, or sometimes called the k-th raw moment, of X
$\mu = E[X] =$ mean $=$ average
$E[X^2]$ is the second (raw) moment of X
$\text{Var}(X) = E[(X-\mu)^2] = \sigma^2 =$ 2nd central moment of X.
$\text{Var}(X) = E[(X-\mu)^k] = \sigma^k =$ k-th central moment of X.
$\text{Var}(X) = E[(X-a)^k] =$ k-th moment about a.

Example 1.34. Refer to Example 1.29: An insurance policy pays 100 per day for up to 3 days of hospitalization and 50 per day of hospitalization thereafter. The number of days of hospitalization, X, is a discrete random variable with probability function

$$P(X=k) = \begin{cases} \frac{6-k}{15} & \text{for } k = 1,2,3,4,5 \\ 0 & \text{otherwise} \end{cases}$$

The mean payment is \$220. Find the variance of a payment for the hospitalization.

Let Y denote the payment amount.

k	1	2	3	4	5
Y	100	200	300	350	400
$P(X=k)$	5/15	4/15	3/15	2/15	1/15

$$\text{Var}(X) = \sum_k P(X=k)(Y_k - E[X])^2$$

$$= \frac{5(-120)^2}{15} + \frac{4(-20)^2}{15} + \frac{3(80)^2}{15} + \frac{2(130)^2}{15} + \frac{180^2}{15} = 10{,}600$$

We will now verify the second definition of variance. Recall that if an operator f is linear, it holds that $f(x+y) = f(x) + f(y)$ and $f(ax) = af(x)$. In fact, $E[X]$ is a linear operator, so $E[X+Y] = E[X] + E[Y]$ and $E[aX] = aE[X]$!

Proof. Let $\mu = E[X]$. Then,

$$E[(X-\mu)^2] = E[(X^2 - 2\mu X + \mu^2)] = E[X^2] - 2\mu E[X] + E[\mu^2]$$

$$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 = E[X^2] - (E[X])^2$$

□

As implied by the definition of variance, it follows that $\text{Var}(X) \geq 0$ and $\text{Var}(X) = 0$ if and only if $P(X = E[X]) = 1$. This implies $E[X^2] \geq (E[X])^2$.

We proceed to two properties of variance:

Theorem 1.35 (Transformations on Variance). *Let X be a random variable, $a, b \in \mathbb{R}$ (constants). Then,*

1. $\text{Var}(aX) = a^2 \text{Var}(X)$
2. $\text{Var}(X + b) = \text{Var}(X)$

Combining items (1) and (2), $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Proof.

$$\text{Var}(aX) = E[a^2X^2] - (E[aX])^2 = a^2E[X^2] - (aE[X])^2$$
$$= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2) = a^2\text{Var}(X)$$

This completes the proof for item (1).

$$\text{Var}(X + b) = E[(X + b)^2] - (E[X + b])^2 = E[X^2 + 2bX + b^2] - (E[X] + E[b])^2$$
$$= E[X^2] + 2bE[X] + E[b^2] - (E[X])^2 - 2E[X]E[b] - (E[b])^2$$
$$= E[X^2] + 2bE[X] + b^2 - (E[X])^2 - 2bE[X] - b^2 = E[X^2] - (E[X])^2 = \text{Var}(X)$$

This completes the proof for item (2). □

As expected, item (2) explains how translations on a probability distribution will not impact the variance.

Definition 1.36. Let X be a random variable with variance $\text{Var}(x)$. The **standard deviation** and **coefficient of variation** of X satisfy

$$\text{SD}(X) = \sigma_X = \sqrt{\text{Var}(X)} \qquad \text{CV}(X) = \frac{\sigma}{\mu} = \frac{\text{SD}(X)}{\text{E}[X]}$$

If $c \in \mathbb{R}$ is a constant,

$$\text{SD}(cX) = |c|\text{Var}(X) \qquad \text{CV}(cX) = \text{CV}(X)$$

Remark: If $E[X]$ is held constant and σ increases, then $\text{CV}(X)$ also increases.

Example 1.37. A random variable X satisfies $E[X] = 5$ and $\text{SD}(X) = 3$. Find (a) $E[X^2]$ (b) $\text{Var}(2X + 6)$ (c) $\text{CV}(2X + 6)$

(a) We can directly compute $\text{Var}(X) = 9$ because we are given the standard deviation, allowing us to use the definition of $\text{Var}(X)$ to find $E[X^2]$.

$$9 = E[X^2] - (E[X])^2 \implies 9 = E[X^2] - 5^2 \implies \boxed{E[X^2] = 34}$$

(b) Since $\text{Var}(X) = 9$, it follows from Theorem 1.35 that $\boxed{\text{Var}(2X+6) = 36}$.

(c) We have that $\text{Var}(2X+6) = 36$ implies $\text{SD}(2X+6) = 6$, and $E[2X+6] = 2E[X] + E[6] = 10+6 = 16$ using properties of linearity and the fact that $E[X] = 5$. Therefore,

$$\boxed{\text{CV}(2X+6) = \frac{6}{16} = \frac{3}{8}}.$$

Lastly, we will cover a unique case of discrete random variables, in which the probability is the same for each outcome.

Definition 1.38 (Discrete Uniform). X is **(discrete) uniform** on $\{1, 2, \ldots, n-1, n\}$ if $P(X = i) = \frac{1}{n}$ for those n choices.

Note that the average of 1 and n is $\frac{n+1}{2}$, as is the average of 2 and $n-1$ and 3 and $n-2$ and so forth. Therefore,

$$E[X] = \frac{n+1}{2} \qquad \text{Var}(X) = \frac{n^2-1}{12}$$

Example 1.39. Suppose X is uniform on $\{3, 4, 5, 6, 7, 8\}$. What is $E[X]$? $\text{Var}(X)$?

As before, we find $E[X]$ by pairing extremes, with each pair having an average of $\frac{11}{2} = E[X]$. Alternatively, X is not a standard uniform because it starts at 3, not 1. But $X - 2$ is standard uniform on $\{1, 2, 3, 4, 5, 6\}$.

$$E[X-2] = \frac{1+6}{2} = \frac{7}{2} \implies E[X] = E[X-2] + 2 = \frac{7}{2} + 2 = \frac{11}{2}$$

$$\text{Var}(X-2) = \text{Var}(X) = \frac{n^2-1}{12} = \frac{6^2-1}{12} = \frac{35}{12}$$

In general, for a discrete uniform variable,

$$\text{Var}(X) = \frac{(\text{number of values in range})^2 - 1}{12}$$

Example 1.40. The number of losses N is uniformly distributed on $\{5, 6, \ldots, 20\}$. Each loss results in a payment of \$100. Find the mean and standard deviation of payment amount.

Let $X = 100N$ denote the payment amount. Then

$$E[X] = 100E[N] = 100\left(\frac{20+5}{2}\right) = \$1{,}250$$

$$\mathrm{Var}(X) = 100^2\mathrm{Var}(N) = 10000\left(\frac{(20-4)^2-1}{12}\right) = 212{,}500$$

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} \approx \boxed{460.98}$$

2 Combinatorics

As implied by the name, combinatorics is a branch about counting combinations, or finding the number of ways to count specific outcomes from a set. Permutations are more specific, which we care about the order, or arrangement of these outcomes.

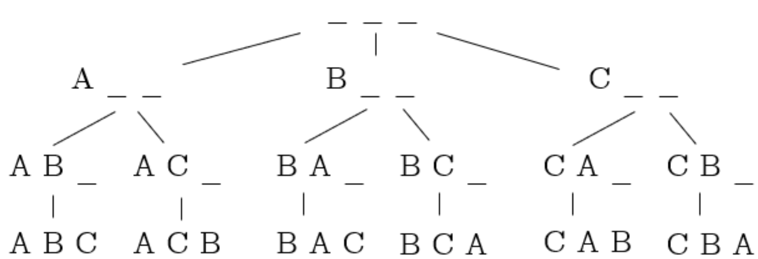
For instance, if we had to choose 2 letters out of A, B, C, and D, combinations would imply that AB = BA, but AB and BA are two different permutations as it accounts for the order of the letters. Or, if we needed to open a combination lock, the order in which the numbers are fixed matter, making it a permutation problem. However, if we only cared about the numbers we choose on the lock, then we have a combination problem.

We can think of combinations as subsets of a set of elements and permutations as arrangements of that subset.

2.1 Combinations and Permutations

Combination and permutation problems can get messy rather quickly, so we will start with a simple scenario: Given people A, B, and C, how many different rankings are possible?

Because we are looking at the number of rankings, we have a permutation problem. There are 3 choices of who comes first, 2 people left who can be second, and 1 person who is last, as displayed in the tree diagram below.



Thus, the total number of possibilities is $3 \cdot 2 \cdot 1 = 6$.

How many possible rankings are there of a group of 12 people?

$$\underline{12}\ \underline{11}\ \underline{10}\ \dots\ \underline{3}\ \underline{2}\ \underline{1}$$

There are 12 choices of who can be first, then 11 left who can be second, 10 left who can be third. Repeat until there is just one who can be last. At each step, multiply the number of choices for the new step with choices so far. This yields $12! = 12 \cdot 11 \cdot 10 \dots 2 \cdot 1$.

The $n!$ notation indicates a *factorial*, which is the product of the positive integer n and all of the ones less than it. We have that

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1, \quad 1! = 1, \quad 0! = 1$$

Each order is called a **permutation**, where $n!$ is the number of permutations of n objects.

Example 2.1. A contest with 12 people gives out 3 distinct prizes. How many ways are there to give out these prizes?

There are 12 choices of who can pick prize 1, 11 who can pick prize 2, and 10 who can pick prize 3. So, there are $12 \cdot 11 \cdot 10$ ways to do this.

This is an example of a partial permutation, or a k-permutation. If there are k items being chosen among n total items, then there are

$$\frac{n!}{(n-k)!} \text{ total permutations.}$$

In simple situations this formula is convenient to have; with more complicated setups, however, it is more useful to think it out from scratch.

Example 2.2. How many 3 digit numbers are there with all even digits?

The first digit cannot be 0 or else our number would be at most 2 digits. So, we can choose between 2, 4, 6, and 8. The second and third digits can be 0, so there are 5 choices for each of them. Therefore, the number of permutations are

$$\underline{4} \cdot \underline{5} \cdot \underline{5} = \boxed{100}$$

Because we are allowed to repeat digits, we are choosing them *without replacement*.

What if we are not allowed to repeat digits?

The first digit follows the same logic from the previous example (choose from 4 digits). Since the second digit can be 0, and one digit is already used, we can choose from 4 digits. The third digit will have 3 digits available to pick from, leaving us with

$$\underline{4} \cdot \underline{4} \cdot \underline{3} = \boxed{48 \text{ permutations}}$$

Because we are *not* allowed to repeat digits, we are choosing digits *without replacement*.

Example 2.3. A conga line forms at a wedding with 20 people (including you) in it. How many different conga lines are possible with you in one of the last 3 spots?

We are in one of the last 3 spots, so we can be in the 3rd to last, 2nd to last, or last in line. In each case, there are 19! ways to arrange everyone else, making 3(19!) total permutations.

In general, if you have n_k copies of the k-th distinct item in a set for $k = 1, 2, \dots, m$ giving a total of $n = n_1 + n_2 + \cdots + n_m$ items, there are

$$\frac{n!}{n_1!n_2!\cdots n_m!} \text{ ways to order the } n \text{ items}$$

That is, there are $n!$ ways to order n items, $n_1!$ ways to arrange item 1, and so forth to $n_m!$ ways to arrange item m.

Example 2.4. How many different 6 six letter words can be made from the word PEPPER? (they do not have to be actual words).

There are 3 distinct letters: 3 P's, 2 E's, and 1 R. If we let $n = 3 + 2 + 1 = 6, n_1 = 3, n_2 = 2$, and $n_3 = 1$, then we have

$$\frac{6!}{3!2!1!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{12} = \boxed{\text{60 six letter permutations}}$$

Example 2.5. How many different four letter words can you make from the word HASHES (once again, they don't need to be actual words)?

First, we want to find out how many 4 letter words we can make out of 6.

$$\frac{6!}{(6-4)!} = \frac{6!}{2!} = 6 \cdot 5 \cdot 4 \cdot 3 = 360 \text{ four letter words}$$

Moreover, we have 2 H's, 2 S's, 1 A, and 1 E. So, $n_1 = 2, n_2 = 2, n_3 = 1, n_4 = 1$ and

$$\frac{360}{2!2!1!1!} = \boxed{\text{90 total permutations}}$$

Sometimes we are interested in the number of ways to select a group, but the order they are selected does not matter.

Example 2.6. A contest with 12 people gives out 3 prizes. How many ways are there to give out the prizes if all 3 are the same?

Recall that when the three prizes were unique, we found that there were 6 permutations in which they could be distributed. To find the total of combinations, we need to divide our previous answer by 6.

$$\frac{12 \cdot 11 \cdot 10}{6} = 220 \text{ combinations}$$

Definition 2.7 (Combinations). If there are n distinct items and want to select a group of k items, the number of **combinations** can be written as

$${}_nC_k \text{ or } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

and we call this quantity "n choose k."

There is a similar notation for permutations, ${}_nP_k$. In addition, it holds that

$$\binom{n}{k} = \binom{n}{n-k} \qquad \text{and} \quad \binom{n}{0} = \binom{n}{n}$$

since $0! = 1$ as there is only 1 way to choose the entire set.

Example 2.8. 18 people are to be divided into 3 groups, one with 8 people, one with 6, and one with 4. How many such divisions are possible?

A standard approach is to initially assume that we have a complete rank, and then divide by the amount of "overcounting." So, we first rank all 18 people, and let group A be the top 8, group B the next 6, and group C the bottom 4.

A	B	C
8 slots	6 slots	4 slots

There are 18! ways to rank everyone, but within each group people are equal, so we over counted by $8! \cdot 6! \cdot 4!$. Therefore, our answer becomes $\frac{18!}{8!6!4!}$.

Alternatively, we can pick the group with 8 people, and we can do so in

$$\binom{18}{8} = \frac{18!}{8!10!} \text{ ways}$$

From the remaining 10 people, pick the group with 6 people, also deciding the group with 4 people. There are

$$\binom{10}{6} = \frac{10!}{6!4!} \text{ ways to do so}$$

That gives a final answer of

$$\boxed{\binom{18}{8} \cdot \binom{10}{6} = \frac{18!}{8!10!}}$$

Example 2.9. 4 distinct numbers are picked from the integers $\{1, 2, \ldots, 30\}$. How many ways are there to draw them such that all of them are divisible by 3?

Within this list, 10 numbers are divisible by 3. Out of these 10, we want to choose 4 of them.

$$\binom{10}{4} = \frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = \boxed{210 \text{ ways}}$$

Example 2.10. 4 distinct numbers are picked from the integers $\{1, 2, \ldots, 30\}$. How many ways are there to draw them such that 3 are divisible by 5 and one is divisible by 7?

The first number that is divisible by both 5 and 7 is 35, so we do not need to worry about duplicates. There are 6 numbers that are multiples of 5 and 4 that are multiples of 7. So, we have

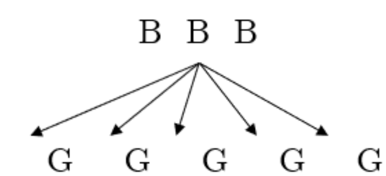
$$\binom{6}{3} \cdot \binom{4}{1} = \frac{6!}{3!3!} \cdot \frac{4!}{3!1!} = 20 \cdot 4 = \boxed{80 \text{ combinations}}$$

Example 2.11. You buy a dozen eggs at the farmers market but 3 of them are rotten. How many different ways can you select two eggs that are not BOTH rotten?

For this problem, we can compute all possible combinations, and subtract the amount of ways to get two rotten eggs:

$$\binom{12}{2} - \binom{3}{2} = \frac{12!}{10!2!} - \frac{6!}{2!1!} = 66 - 3 = \boxed{63 \text{ ways}}$$

Example 2.12. A woman has a set of identical triplet sons and quadruplet daughters. How many ways can she line them up for a picture so that no two sons are standing next to each other?



There are 3 boys that we can put in 5 spots to ensure that no boy is next to each other. Thus, there are

$$\binom{5}{3} = \frac{5!}{3!2!} = \boxed{10 \text{ ways to do so}}$$

2.2 Common Distributions

Now, we proceed to types of random variables and distributions. Before proceeding, we motivate the simplest distribution with an example:

Avery is practicing free throws. If they make each shot with probability 0.7 and each shot is independent, what is the probability that they make the next 4 shots and then miss the 2 after that? What is the probability that they make exactly 4 of the next 6 shots?

Each shot is independent, so each order has probability

$$P([\text{Make}])^4 \cdot P([\text{Miss}])^2 = 0.7^4 \cdot 0.3^2 \approx \boxed{0.0216 = 2.16\%}$$

To make exactly 4 of 6, there are $\binom{6}{4}$ ways to choose 4 shots are successful.

Definition 2.13. A **Bernoulli**(p) random variable, or a Bernoulli 0-1 random variable, is a variable that can only be 0 or 1. Usually 1 is considered a success. If p is the probability of success, then

$$P(X = 1) = p \quad P(X = 0) = 1 - p$$

The mean and variance follow:

$$E[X] = p \quad \text{Var}(X) = p(1 - p)$$

Say that X is a random variable that can take on two values a and b (not necessarily 0 and 1), with

$$P(X = b) = p \quad P(X = a) = 1 - p = q$$

Then the mean and variance are

$$E[X] = aq + bp = a + p(b - a) \qquad E[X^2] = a^2q + b^2p$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = (b - a)^2pq$$

$$\text{Or:} \qquad X = (b - a)Y + a \qquad Y \sim \text{Bernoulli}(p)$$

$$E[X] = (b - a)E[Y] + a = p(b - a) + a$$

$$\text{Var}(X) = (b - a)^2\text{Var}(Y)$$

$$\boxed{\text{Var}(X) = (b - a)^2pq}$$

Definition 2.14. X is a **binomial** (n, p) random variable if X is the number of successes in n independent trials, each of which is a success with the same probability p.

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

The Bernoulli distribution is a special case of the Binomial distribution where $n = 1$, or they can be considered as individual trials. The essential components of a binomial distribution include

- A fixed number of trials
- Independent trials
- Success probability is the same in all trials

Theorem 2.15 (Mean and Variance). *Let $X \sim$ Binomial(n, p). Then,*

$$E[X] = np \qquad \text{Var}(X) = np(1-p)$$

Proof. The results follow immediately from what we know about Bernoulli distributions. If $X_i \sim \text{Bernoulli}(p)$, then

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) \text{ by independence}$$

$$\text{Var}(X) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

□

Example 2.16. A commuter airline sells 32 tickets for a flight on a plane that has 30 seats. The probability that any particular passenger will not show up for a flight is 0.1, independent of other passengers. Find the probability that more passengers show up for the flight than there are seats available.

Let $N =$ number of passengers that show up. Then $N \sim \text{Binomial}(n = 32, p = 0.9)$. To find the probability $P(N > 30)$, we need to sum $P(N = 31) + P(N = 32)$. This can be achieved through the formula provided in Definition 2.14:

$$P(N = 31) + P(N = 32) = \binom{32}{31}(0.9)^{31}(0.1) + \binom{32}{32}(0.9)^{32} \approx \boxed{0.156 = 15.6\%}$$

What are the mean, variance, and standard deviation of number of passengers who show up?

With $n = 32$ and $p = 0.9$,

$$E[N] = 32(0.9) = 28.8 \qquad \text{Var}(N) = 32(0.9)(0.1) = 2.88 \qquad \text{SD}(N) = 1.69$$

Theorem 2.17 (Binomial Expansion). *For any real numbers a, b and positive integer n:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

The proof is rather lengthy so we will not dive into it; however, it does involve knowledge of Pascal's Triangle and induction.

Example 2.18. Use Theorem 2.17 to evaluate the following sums:

1. $\displaystyle\sum_{k=0}^{6} \frac{6!}{k!(6-k)!}$

In this case, $n = 6$. $a = b = 1$ since there are no exponent terms. Therefore, this sum is equal to

$$(1+1)^6 = 2^6 = 64$$

2. $\displaystyle\sum_{k=0}^{n} \frac{n!}{k!(n-k)!}(-2)^k$

In this case, n is unknown, but $a = -2$ and $b = 1$, so this sum is equal to $(-1)^n$.

3. $\displaystyle\sum_{k=1}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

Here, the index starts at $k = 1$. We can rewrite this as an expression where we have a sum starting at $k = 0$ and subtracting off the first term ($k = 0$).

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} - \binom{n}{0} p^0 (1-p)^{n-0}$$

In the first sum, $a = p$ and $b = 1 - p$, so it is equal to $(p + (1-p))^n = 1$. The second sum is equal to $1 \cdot 1 \cdot (1-p)^n$. Therefore, this sum is equal to

$$\boxed{1 - (1-p)^n}$$

In some cases, we are looking at more than just two outcomes. This is where the multinomial distribution enters.

Example 2.19. Accidents are categorized into three groups: minor, moderate, and severe. These occur with probabilities 0.5 for minor, 0.4 for moderate, and 0.1 for severe. Two accidents occur independently in one month. Find the probability that neither accident is severe and at most one is moderate.

This is not binomial because each accident (trial) has more than 2 possible outcomes. We want to sum two probabilities: either we have 1 minor and 1 moderate accident or 2 minor accidents.

$$\binom{2}{1}(0.5)(0.4)+\binom{2}{2}(0.5)^2 = 2(0.4)(0.5)+0.25 = 0.65$$

This was an example of the multinomial distribution.

Suppose there are n trials with 3 possible outcomes. Let the probabilities of those outcomes be p_1, p_2, and p_3 such that $\sum p_i = 1$. Let X_i be the number of trials that have outcome i. Then,

$$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3} p_1^{k_1} p_2^{k_2} p_3^{k_3}$$
$$= \frac{n!}{k_1!(n-k_1)!} \cdot \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\binom{k_3}{k_3} p_1^{k_1} p_2^{k_2} p_3^{k_3} = \frac{n!}{k_1!k_2!k_3!}$$

The final term is called the *multinomial coefficient*.

Definition 2.20 (Multinomial Distribution). Suppose there are n independent trials, each with the same r possible outcomes. Let $p_1, p_2, \ldots p_r$ be the probabilities of the outcomes, and X_i the number of trials resulting in the i-th outcome. Then,

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r) = \frac{n!}{k_1!k_2!\cdots k_n!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

If X_i, X_j are trials whose respective probabilities of success are p_i and p_j, then

$$E[X_i] = np_i \qquad \text{Var}(X_i) = np_i(1-p_i) \qquad \text{Cov}(X_i, X_j) = -np_ip_j$$

Cov is the covariance between two random variable, which measures how two random variables change together, indicating the direction and magnitude of their relationship. This will be covered more when we talk about joint variability.

As with binomial, we need:

- A fixed number of trials
- Different trials are independent
- All trials have the same distribution

Example 2.21. Accidents are categorized into three groups: minor, moderate, and severe. These occur with probabilities 0.5 for minor, 0.4 for moderate, and 0.1 for severe. Four accidents occur independently in one month. Find the probability that there is at least one accident of each type.

There are three cases:

1. $P(2 \text{ minor}, 1 \text{ moderate}, 1 \text{ severe}) = \frac{4!}{2!1!1!}(0.5)^2(0.4)(0.1) = 0.12$
2. $P(1 \text{ minor}, 2 \text{ moderate}, 1 \text{ severe}) = \frac{4!}{1!2!1!}(0.5)(0.4)^2(0.1) = 0.096$
3. $P(1 \text{ minor}, 1 \text{ moderate}, 1 \text{ severe}) = \frac{4!}{1!1!2!}(0.5)(0.4)(0.1)^2 = 0.024$

$$P(\text{Total}) = 0.12 + 0.096 + 0.024 = \boxed{0.24 = 24\%}$$

Example 2.22. San Diego Fire Department has 4 firehouses to store their trucks (North, West, East, South). These occur with probabilities 0.34 for North, 0.18 for South, 0.21 for West, and 0.27 for East. 40 trucks are parked between all four stations. What is the probability that each station has an equal amount of trucks? Assume these events are independent.

The question implies that each station has exactly 10 trucks inside.

$$P(10 \text{ North}, 10 \text{ West}, 10 \text{ East}, 10 \text{ South}) = \frac{40!}{(10!)^4}(0.34)^{10}(0.18)^{10}(0.21)^{10}(0.27)^{10}$$

$$= \boxed{0.0012 \approx 0.12\%}$$

Theorem 2.23 (Multinomial Theorem). *For any real numbers $x_1, x_2, \ldots, x_k \in \mathbb{R}$ we have:*

$$(x_1 + \cdots + x_k)^n = \sum_{n_1+\cdots+n_k=n} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$$

Example 2.24. Evaluate the sum $\displaystyle\sum_{i+j+k=5} \frac{5!}{i!j!k!} 2^j$

In this case, we let $n_1 = i, n_2 = j$, and $n_3 = k$. So, we have $x_1 = 1, x_2 = 2, x_3 = 1$, and $n = 5$.

$$\sum_{i+j+k=5} \binom{5}{i, j, k} = 1^i 2^j 1^k = (1 + 2 + 1)^5 = 1024$$

Example 2.25. What is the coefficient of $a^2b^2c^3$ in $(a + b + 2c)^7$?

$$(a + b + 2c)^7 = \sum_{i+j+k=7} \binom{7}{i, j, k} a^i b^j (2c)^k = \sum_{i+j+k=7} \frac{2^k 7!}{i!j!k!} a^i b^j c^k$$

where $i = 2, j = 2, k = 3$. So, the coefficient is

$$\frac{2^3(7!)}{2!2!3!} = \boxed{1680}$$

The last distribution to be discussed in this section is the hypergeometric distribution.

Example 2.26. A crate of 10 electrical components has 4 defective components. If 3 components are randomly selected, find the probability that at most one of them is defective.

There are two cases: none of the 3 are defective, or 1 of the three is defective. Those are mutually disjoint, so we can sum their probabilities, giving

$$P(\text{none defective}) + P(1 \text{ defective}) = \frac{\binom{6}{3}}{\binom{10}{3}} + \frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}} = \boxed{\frac{20}{120} = \frac{1}{6}}$$

Definition 2.27 (Hypergeometric Distribution). Say we have N trials/objects with m successes. If you randomly select n of them without replacement, then $X \sim \text{Hyp}(n, N, m)$ is **hypergeometric** and has distribution

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

for $k = 0, 1, \ldots, \min(m, n)$. If X follows a hypergeometric distribution,

$$E[X] = \frac{mn}{N} \qquad \text{Var}(X) = \frac{mn(N-n)(N-m)}{N^2(N-1)}$$

In summary, it is the number of ways to choose exactly k good items over the number of ways to choose n total items.

This is not a binomial distribution! Knowing whether or not the first item is good gives information about whether or not the second one will be good *(based on dependent events)*.

If $n > k$, then it isn't possible to have n successes.

Keep in mind that a binomial distribution is sampling *with* replacement, and a hypergeometric distribution is sampling *without* replacement.

Example 2.28. When packing for a trip, I draw 6 socks without replacement from a drawer that contains 16 black socks and 4 white socks. What is the probability I will draw 4 black socks and 2 white socks?

$$P(W = 2) = \frac{\binom{4}{2}\binom{16}{4}}{\binom{20}{6}} \approx \boxed{0.282 = 28.2\%}$$

Example 2.29. For each problem, determine which type of distribution should be used and compute the desired probability.

1. I randomly select 6 socks from a drawer. Each sock has a 50% chance of being black, a 30% chance of being brown, and a 20% chance of being white, independently of the other socks. Find the probability that I will draw 2 socks of each color.

This is a multinomial distribution problem, whose probability is given by

$$P(2 \text{ black, } 2 \text{ brown, } 2 \text{ white}) = \frac{6!}{2!2!2!}(0.5)^2(0.3)^2(0.2)^2 \approx \boxed{0.081 = 8.1\%}$$

2. I draw 6 socks without replacement from a drawer that contains 10 black socks, 6 brown socks, and 4 white socks. Find the probability that I will draw 2 socks of each color.

This is a hypergeometric distribution problem: draws are not independent because we are sampling without replacement. The probability is computed accordingly:

$$P(2 \text{ black, } 2 \text{ brown, } 2 \text{ white}) = \frac{\binom{10}{2}\binom{6}{2}\binom{4}{2}}{\binom{20}{6}} \approx \boxed{0.104 = 10.4\%}$$

3. Suppose that I draw 6 socks with replacement from a drawer that contains 16 black socks and 4 white socks. What is the probability that I will draw 4 black socks and 2 white socks?

This is a binomial distribution problem since the different draws are independent and there are two possible outcomes. The probability is

$$P(4 \text{ black, } 2 \text{ white}) = P(2 \text{ white}) = \binom{6}{2}(0.2)^2(0.8)^4 \approx \boxed{0.246 = 24.6\%}$$

Summary and Comparison between Binomial/Multinomial/Hypergeometric Distributions

	Binomial	**Multinomial**	**Hypergeometric**
Experiment Setup	• Repeated, Bernoulli trials (success/failure)	• Repeated trials with *k* possible outcomes per trial	• Sampling without replacement from a finite population
Independence	• Trials are independent	• Trials are independent	• Trials/draws are dependent (without replacement)
Number of categories	• 2 outcomes	• 2 or more outcomes	• 2 outcomes or extended to multiple categories
Parameters	• n: number of trials • p: probability of success	• n: number of trials • $\mathbf{p} = (p_1, \ldots, p_n)$ category probabilities with $\sum p_i = 1$	• N: population size • K: number of successes in population • n: draws
Random Variable	• Number of successes in n trials	• Counts of occurrences in each of k categories	• Number of successes in the sample

3 Key Discrete Distributions

In this section, we are going to discuss the Geometric, Negative Binomial, and Poisson Distribution.

3.1 Geometric Series and Distributions

Many properties of geometric series will be derived here and will be important to have in our toolkit when we introduce geometric distributions.

Many discrete distributions can equal any non-negative integer. Deriving the mean/variance of these requires using infinite series.

Theorem 3.1 (Geometric Series Convergence). *Let $|r| < 1$ and a be a real number. Then,*

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

Proof. Let $S = \sum_{n=0}^{\infty} ar^n$. Then,

$$S = a + ar + ar^2 + \dots$$

$$Sr = ar + ar^2 + ar^3 + \dots$$

Take $S - Sr$ to get

$$(a + ar + ar^2 + \dots) - (ar + ar^2 + ar^3 + \dots) = a \implies S(1-r) = a \iff S = \frac{a}{1-r}$$

$\square$

More generally, for a partial sum,

$$\sum_{n=0}^{m} ar^n = \sum_{n=0}^{\infty} ar^n - \sum_{n=m+1}^{\infty} ar^n = \frac{a}{1-r} - \frac{ar^{m+1}}{1-r} = \frac{a(1-r^{m+1})}{1-r}$$

Or, more plainly,

$$\frac{\text{first term} - \text{ first missing term}}{1-r}$$

Example 3.2. Evaluate the sum $\sum_{n=3}^{17} 5 \cdot \frac{e^{n+2}}{2^{3n}3^{-n}}$.

Note that $2^{3n} = 8^n$, $e^{n+2} = e^2e^n$, and $\frac{1}{3^{-n}} = 3^n$. Therefore, $r = \frac{3e}{8}$. The sum is therefore equal to

$$\frac{\frac{5(3^3)e^5}{2^9} - \frac{5 \cdot 3^{18} e^{20}}{2^{54}}}{1 - \frac{3e}{8}}.$$

Whew!

What if we want the sums $\sum_{n=0}^{\infty} n \cdot ar^n$ and $\sum_{n=0}^{\infty} n^2 \cdot ar^n$?

This can be achieved with a little bit of calculus! We know that this sum is differentiable term-by-term, so we can differentiate the series with respect to r! Differentiate the default geometric series

$$\frac{d}{dr} \sum_{n=0}^{\infty} ar^n = \sum_{n=0}^{\infty} anr^{n-1} = \frac{1}{r} \sum_{n=0}^{\infty} anr^n$$

We also differentiate $\sum_{n=0}^{\infty} ar^n$

$$\frac{d}{dr} \left(\frac{a}{1-r} \right) = \frac{a}{(1-r)^2}$$

Set these two results equal to each other:

$$\frac{1}{r} \sum_{n=0}^{\infty} n \cdot ar^n = \frac{a}{(1-r)^2} \quad \Longleftrightarrow \quad \boxed{\sum_{n=0}^{\infty} n \cdot ar^n = \frac{ar}{(1-r)^2}}$$

We use this result to compute $\sum_{n=0}^{\infty} n^2 ar^n$. We differentiate $\sum_{n=0}^{\infty} n \cdot ar^n$ with respect to r:

$$\frac{d}{dr} \sum_{n=0}^{\infty} nar^n = \frac{1}{r} \sum_{n=0}^{\infty} n^2 ar^n$$

$$\frac{d}{dr} \left(\frac{ar}{(1-r)^2} \right) = a \left(\frac{(1-r)^2 + 2r(1-r)}{(1-r)^4} \right) = a \cdot \frac{(1-r)(1+r)}{(1-r)^4} = \frac{a(1+r)}{(1-r)^3}$$

Setting these two equal to each other:

$$\frac{1}{r} \sum_{n=0}^{\infty} n^2 ar^n = \frac{a(1+r)}{(1-r)^3} \quad \Longleftrightarrow \quad \boxed{\sum_{n=0}^{\infty} n^2 ar^n = \frac{ar(1+r)}{(1-r)^3}}$$

These results will come in handy as we discuss geometric distributions.

Example 3.3. Suppose I roll a die until I get a 6. Let N be the total number of rolls. What is the distribution of N?

Let's go over the probability as N increases:

$$P(N>0)=1 \qquad P(N=1)=\frac{1}{6}$$
$$P(N>1)=\frac{5}{6} \qquad P(N=2)=\frac{5}{6}\cdot\frac{1}{6}$$
$$P(N>2)=\left(\frac{5}{6}\right)^2 \qquad P(N=3)=\left(\frac{5}{6}\right)^2\cdot\frac{1}{6}$$
$$\vdots \qquad \vdots$$
$$P(N>k)=\left(\frac{5}{6}\right)^k \qquad P(N=n)=\frac{1}{6}\left(\frac{5}{6}\right)^{n-1}$$

What are the expected value and variance in this distribution?

$$E[N]=\sum_{n=1}^{\infty} nP(N=n)=\sum_{k=0}^{\infty} P(N>k)$$
$$=\sum_{n=1}^{\infty} n\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{n-1}=\sum_{k=0}^{\infty}\left(\frac{5}{6}\right)^k$$

We can use Theorem 3.1 to evaluate the sum

$$E[N]=\frac{1}{1-\frac{5}{6}}=6$$

To compute the variance, we use the results from earlier:

$$E[N^2]=\sum_{n=1}^{\infty} n^2P(N=n)=\sum_{n=0}^{\infty}\frac{1}{6}n^2\left(\frac{5}{6}\right)^{n-1}$$
$$=\sum_{n=0}^{\infty} n^2\cdot\frac{1}{6}\cdot\frac{6}{5}\cdot\left(\frac{5}{6}\right)^n$$

Letting $a=\frac{1}{6}$ and $r=\frac{5}{6}$,

$$\frac{ar(1+r)}{(1-r)^3}=\frac{\frac{1}{6}\cdot\frac{6}{5}\cdot\frac{5}{6}\cdot\frac{11}{6}}{\left(1-\frac{5}{6}\right)^3}=216\left(\frac{11}{36}\right)=66$$

$$\boxed{\text{Var}(N)=66-6^2=30}$$

Theorem 3.4 (Geometric Series starting at 1). *Suppose X is a geometric random variable on $\{1, 2, \ldots\}$ with parameter p if X is the number of trials up to, and including, the first success. Then,*

$$E[X] = \frac{1}{p} \qquad \text{Var}(X) = \frac{1-p}{p^2}$$

Proof. We have $P(X = n) = p(1-p)^{n-1}$.

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1}$$

Fix $k = n - 1$, then $n = k + 1$ and we rewrite the sum as

$$\sum_{k=0}(k+1)p(1-p)^k = \frac{p}{(1-(1-p))^2}$$

using the fact $\sum_{n=0}^{\infty} n \cdot ar^n = \frac{ar}{(1-r)^2}$, where $a = p$ and $r = 1 - p$.

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

To compute the variance, we want $E[X^2]$

$$E[X^2] = \sum_{k=0}^{\infty} n^2 p(1-p)^{n-1} = \sum_{k=0}^{\infty}(k+1)^2 p(1-p)^k$$

. . . leaving the same substitution as earlier. We expand $(k+1)^2$ to evaluate three different series.

$$\sum_{k=0}^{\infty} k^2 p(1-p)^k + \sum_{k=0}^{\infty} 2k(1-p)^k + \sum_{k=0}^{\infty}(1-p)^k$$
$$= \frac{p(1-p)(2-p)}{p^3} + \frac{p(1-p)}{p^2} + \frac{1}{p}$$

We establish a common denominator of p^2 by dividing p by p^3 on the first fraction and by multiplying both sides by p on the third fraction:

$$E[X^2] = \frac{(p^2 - 3p + 2) + (p - p^2) + p}{p^2} = \frac{2-p}{p^2} = \frac{2}{p^2} - \frac{1}{p}$$

Thus,

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \left(\frac{2}{p^2} - \frac{1}{p}\right) - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$$

□

Theorem 3.5 (Geometric Series starting at 0). *Suppose Y is a geometric random variable on 0, 1, 2, ... if Y counts the number of failures before the first success. Then*

$$E[Y] = E[X] - 1 = \frac{1}{p} - 1 \qquad Var(Y) = \frac{1-p}{p^2}$$

Proof. The results are hopefully straightforward to understand. By letting $Y = X - 1$ be a translation of X, we revisit the properties of $E[X]$ and $\text{Var}(X)$ in Section 1.2.

$$E[Y] = E[X-1] = E[X] - E[1] = \frac{1}{p} - 1$$

Recall that variances do not change if a translation is applied

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

□

Example 3.6. Let N be the number of visits (possibly 0) that a randomly chosen insured patient makes to the doctor in a year. If N has a geometric distribution with mean 3, what is the probability that a randomly chosen insured makes at least 2 visits to the doctor in a year?

We are told that N can be 0 and that $E[N] = 3$. This implies

$$3 = \frac{1}{p} - 1 \Longleftrightarrow p = \frac{1}{4}$$

To compute $P(N \geq 2)$, we can use the survival method and subtract $P(N = 0)$ and $P(N = 1)$ from 1:

$$P(N \geq 2) = 1 - P(N=0) - P(N=1) = 1 - p - p(1-p) = 1 - 2p + p^2 = (1-p)^2$$
$$= \boxed{\frac{9}{16}}$$

3.2 Memoryless Property and Negative Binomial Distributions

We motivate this concept with an example:

Example 3.7. In each round of the dice game "Nines" I roll two fair six-sided dice. The game ends if either a 7 or 9 is rolled, and continues to the next round on any other outcome. If I play a game of Nines, what is the expected number of rounds I will play?

If I roll two die, the most probable outcome is a 7 with 6 out of 36 possible ways to roll it. In order to roll a nine, you must roll a 4 and 5 or 3 and 6. There are 4 ways to achieve this. Therefore, the game ends on a given round with probability

$$\frac{6}{36} + \frac{4}{36} = \frac{5}{18}$$

The length of the game is therefore a geometric random variable (starting at 1) with $p = \frac{5}{18}$. The expected game length is $\frac{1}{p} = \frac{18}{5}$.

Suppose I watch someone play Nines after the 3rd round. How many more rounds will I watch?

Intuitively, the concept remains the same. Each round I watch will end the game with probability $\frac{5}{18}$, and the number of rounds I watch is a geometric series starting at 1, so the answer remains as $\frac{18}{5}$.

Let's consider an algebraic approach to the previous example. Fix $N =$ game length. We start watching after 3 rounds, so we watch for $N - 3$ rounds. We know the game lasts for more than 3 rounds, implying $N > 3$. Using what we know about conditional probability:

$$P(N - 3 = k \mid N > 3) = \frac{P(N = k + 3, N > 3)}{P(N > 3)} = \frac{p(1-p)^{k+3-1}}{(1-p)^3}$$
$$= p(1-p)^{k-1} = P(N = k)$$

so $(N - 3 \mid N > 3)$ and (the original) N have the same distribution, and $E[N - 3 \mid N > 3] = E[N]$. This is a unique property of discrete geometric distributions, known as the *memoryless property.*

Theorem 3.8 (Memoryless Property). *If N follows a discrete geometric distribution with parameter p, then $(N - k \mid N > k)$ is a geometric distribution starting at 1 with the same p. This holds whether N starts at 0 or 1.*

Example 3.9. A game of Nines lasts for at least 4 rounds. What are the mean and variance of the length of the game?

We are given that $N \geq 4 \implies N > 3$. Apply the Memoryless Property as such:

$$E[N \mid N > 3] = E[N - 3 + 3 \mid N > 3]$$

$$= E[N - 3 \mid N > 3] + 3 = E[N] + 3 = \frac{18}{5} + 3 = \boxed{\frac{33}{5}}$$

Once again, variances are not impacted by translation, so

$$\text{Var}(N \mid N > 3) = \text{Var}(N) = \frac{1-p}{p^2} \approx \boxed{9.36}$$

Example 3.10. Suppose X satisfies $P(X = k) = 0.2(0.8)^k$ for $k = 0, 1, 2, \ldots$. Find $E[X \mid X > 6]$ and $\text{Var}(X \mid X > 6)$.

X can be 0 and $P(X = k)$ decays geometrically, so X is a geometric starting at 0. $P(X = 0) = p = 0.2(0.8)^0 = 0.2$.

$$\boxed{E[X] = \frac{1}{0.2} + 6 = 11 \qquad \text{Var}(X) = \frac{1-p}{p^2} = 20}$$

Let us observe how the memoryless property extrapolates to other distributions:

Example 3.11. Roll a die until the third time a 6 is rolled. Let N denote the number of non-sixes (failures) that we roll. What is the distribution of N?

If $N = n$, then roll number $n + 3$ was a 6. The first $(n + 3) - 1$ rolls had $3 - 1 = 2$ sixes. Since there were exactly 3 sixes in the first $n + 3$ rolls, there were n non-sixes.

$$P(N = n) = \binom{(n+3)-1}{3-1} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^n$$

$$= \binom{n+(3-1)}{3-1} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^n = \binom{n+(3-1)}{n} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^n$$

Recall that the last two factorial expressions are equivalent by symmetry of combinations.

What is the mean and variance of N?

Let N_1, N_2, N_3 be the number of failures before the 1st, 2nd, and 3rd 6, respectively. Then $N_3 = N$ and

$$N = (N_1 - 0) + (N_2 - N_1) + (N_3 - N_2)$$

$N_1 - 0$ is the number of failures before the first six. $N_2 - N_1$ is the number of failures after the first six, but before the second six. And, lastly, $N_3 - N_2$ is the number of failures after the second six, but before the third six.

The number of failures between sixes is a geometric series that starts at 0!

N is the sum of 3 independent geometrics on $\{0, 1, 2, \ldots\}$ and

$$E[N] = E[N_1] + E[N_2 - N_1] + E[N_3 - N_2] = 3E[N_1] = 3\left(\frac{1}{\frac{1}{6}} - 1\right) = \boxed{15}$$

This is because $p = \frac{1}{6}$ for any given roll and rolling the first six does not affect the probability of rolling a second six.

$$\text{Var}(N) = \frac{1-p}{p^2} = \boxed{90}$$

This problem exhibits the key property of a *negative binomial distribution*:

Definition 3.12 (Negative Binomial Distribution). Suppose N is a negative binomial random variable with parameters r and p if it is the sum of r independent geometric random variables starting at 0. It is the number of failures before the r-th success.

$$P(N = n) = \binom{n + (r-1)}{n} p^r (1-p)^n$$

$$E[N] = r\left(\frac{1}{p} - 1\right) \qquad \text{Var}(N) = \frac{r(1-p)}{p^2}$$

Example 3.13. An insurance policy covers accidents at a manufacturing plant. The probability that one or more accidents will occur during any given month is $\frac{3}{5}$. The number of accidents that occur in any given month is independent of the number of accidents that occur in all other months. Find the probability that June with be the fourth month in 2025 in which at least one accident occurs.

Having an accident = "success," $p = \frac{3}{5}$. We want $r = 4$th success in 6th try, $n = 6 - 4 = 2$ "failures." The probability is therefore

$$P(N = 6) = \binom{5}{3}\left(\frac{3}{5}\right)^4\left(\frac{2}{5}\right)^2 \approx \boxed{0.2074 = 20.74\%}$$

Example 3.14. Let N be the sum of r independent geometrics $\{0, 1, 2, \ldots\}$. Suppose that $E[N] = 12$ and $\text{Var}[N] = 60$. Find the probability that N is no more than 2.

We can establish a relationship between $E[N]$ and $\text{Var}(N)$:

$$\text{Var}(N) = \frac{E[N]}{p} \qquad 60 = \frac{12}{p} \qquad p = \frac{1}{5}$$

Use either equation to find $r = 3$.

$$P(N \leq 2) = P(N = 0) + P(N = 1) + P(N = 2)$$

$$= \binom{2}{0}p^3 + \binom{3}{1}(1-p)p^3 + \binom{4}{2}(1-p)^2p^3 \approx \boxed{0.0579 = 5.79\%}$$

3.3 Poisson Distribution and Variables

Before talking about our final distribution in this section, we need to prove one important theorem.

Theorem 3.15 (Taylor Series for e^x). *The Taylor (or Maclaurin) series for e^x centered at $x = 0$ is equal to the infinite sum of terms*

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Proof. A Taylor Series for a function $f(x)$ centered at $x = 0$ is equal to the infinite sum

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

so long as $f(x)$ is infinitely differentiable. e^x is continuously differentiable, so $f(x) = e^x = f^{(n)}(x)$ for all n. Therefore, $f^{(n)}(0) = 1$ and

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

□

We can identify variations of e^x based on their Taylor Series.

Example 3.16. What is the function representation of the Taylor Series $\sum_{n=0}^{\infty} \frac{5^n e^{tn}}{n!} e^{-5}$?

We want to extract the Taylor Series into a form that is familiar to that of e^x.

$$\sum_{n=0}^{\infty} \frac{5^n e^{tn}}{n!} e^{-5} = e^{-5} \sum_{n=0}^{\infty} \frac{(5e^t)^n}{n!}$$

Fix $u = 5e^t$. Then, we have the series

$$e^{-5} \sum_{n=0}^{\infty} \frac{u^n}{n!} \approx e^{-5} u^n = e^{-5}(5e^t) = \boxed{e^{5e^t - 5}}$$

Example 3.17. Evaluate the series $\sum_{n=2}^{\infty} \frac{2^n}{n!}$

We know $\sum_{n=0}^{\infty} \frac{2^n}{n!} = e^2$. We subtract the first two terms from e^2.

$$\sum_{n=2}^{\infty} \frac{2^n}{n!} = \sum_{n=0}^{\infty} \frac{2^n}{n!} - \left(\frac{2^0}{0!} + \frac{2^1}{1!}\right) = \boxed{e^2 - 3}$$

The Taylor Series for e^x is well represented in the Poisson Distribution.

Definition 3.18. X is a Poisson(λ) random variable if

$$P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!} \qquad \text{for } n = 0, 1, 2, \ldots$$

The Taylor Series for $e^{-\lambda}$ is

$$e^{-\lambda} = 1 - \lambda + \frac{\lambda^2}{2} - \frac{\lambda^3}{6} + \ldots$$

The $e^{-\lambda}$ term is the constant needed to make the probabilities to sum to 1.

Poisson variables arise in nature by mimicking the number of *occurrences* of unusual events if the number of occurrences in disjoint time intervals are independent.

Theorem 3.19 (Properties of Poisson Distribution). *Suppose $N \sim Pois(\lambda)$. Then,*

$$E[N] = \lambda \qquad \text{Var}(N) = \lambda$$

Proof.

$$E[N] = \sum_{n=0}^{\infty} nP(N = n)$$

We start the index at 1 because the first term is equal to 0. Moreover, $\frac{n}{n!} = \frac{1}{(n-1)!}$ and $\lambda^n = \lambda \cdot \lambda^{n-1}$. This will help us get the series into something more familiar.

$$\sum_{n=1}^{\infty} ne^{-\lambda}\frac{\lambda^n}{n!} = \lambda \sum_{n=1}^{\infty} e^{-\lambda}\frac{\lambda^{n-1}}{(n-1)!}$$

By letting $m = n - 1$,

$$\lambda \sum_{n=1}^{\infty} e^{-\lambda}\frac{\lambda^{n-1}}{(n-1)!} = \lambda \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda \cdot 1$$

$\ldots$ since the sum is $\sum_{m=0}^{\infty} P(N = m) = 1$. Therefore, $E[N] = \lambda$. To calculate the variance, we once again need to find $E[N^2]$.

$$E[N^2] = \sum_{n=0}^{\infty} n^2 e^{-\lambda} \frac{\lambda^n}{n!} = \sum_{n=1}^{\infty} n e^{-\lambda} \frac{\lambda \cdot \lambda^{n-1}}{(n-1)!}$$

By letting $m = n - 1$, $n = m + 1$ and the sum is

$$\lambda \sum_{n=0}^{\infty} (m+1) e^{-\lambda} \frac{\lambda^m}{m!} = \lambda \left(\sum_{m=0}^{\infty} m \cdot \frac{e^{-\lambda}\lambda^n}{m!} \right) + \lambda \left(\sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{m!} \right)$$

$$= \lambda \left(\sum_{m=0}^{\infty} mP(N=m) \right) + \lambda \left(\sum_{m=0}^{\infty} P(N=m) \right)$$

$$= \lambda \cdot E[N] + \lambda = \lambda^2 + \lambda$$

$$\text{Var}(N) = E[N^2] - (E[N])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

as desired. □

Example 3.20. Policyholders are three times as likely to file two claims as to file four claims. If the number of claims filed has a Poisson distribution, find the variance of the number of claims filed.

Let N be the number of claims. $\text{Var}(N) = \lambda$, so we need to find λ. We currently know that

$$P(N = 2) = 3P(N = 4)$$

$$e^{-\lambda} \frac{\lambda^2}{2} = 3e^{-\lambda} \frac{\lambda^4}{4!}$$

The $e^{-\lambda}$ term can be removed from both sides. We have

$$\frac{\lambda^4}{\lambda^2} = \frac{4!}{2 \cdot 3} \iff \lambda^2 = 4 \iff \boxed{\lambda = \text{Var}(N) = 2}$$

Example 3.21. The number of annual losses has a Poisson distribution with second moment equal to 12. Find the probability that the number of annual losses is at least 2.

Choose N as the number of annual losses.

$$E[N^2] = \text{Var}(N) - (E[N])^2 \quad \iff \quad 12 = \lambda + \lambda^2 \quad \iff \quad \lambda^2 + \lambda - 12 = 0$$

$$(\lambda - 3)(\lambda + 4) = 0$$

We choose $\lambda = 3$ because it must be positive.

$$P(N \geq 2) = 1 - P(N = 0) - P(N = 1) = 1 - e^{-3} - 3e^{-3} = 1 - 4e^{-3} \approx \boxed{0.8006 = 80.06\%}$$

Example 3.22. If $Y = \text{Pois}(2)$, find $P(1 \le Y \le 3)$.

$$P(1 \le Y \le 3) = P(Y=1) + P(Y=2) + P(Y=3) = \frac{2e^{-2}}{1} + \frac{4e^{-2}}{2} + \frac{8e^{-2}}{6} = \boxed{\frac{16}{3e^2}}$$

Example 3.23. If $W \sim \text{Pois}(\lambda)$ and $P(W=0) = \frac{1}{2}$, what is $E[W^2 - W]$?

We wish to rewrite $E[W^2 - W]$ into terms we are familiar with:

$$E[W^2 - W] = E[W^2] - E[W] = E[W^2] - (E[W])^2 + (E[W])^2 - E[W]$$

$$= \text{Var}(W) + (E[W])^2 - E[W] = (E[W])^2 = \lambda^2$$

Use what we are given for $P(W=0)$ to find λ:

$$e^{-\lambda}\frac{\lambda^0}{0!} = \frac{1}{2} \quad\iff\quad e^{\lambda} = 2 \quad\iff\quad \lambda = \ln 2$$

Therefore, $\boxed{E[W^2 - W] = (\ln 2)^2}$

Example 3.24. The number of Supreme Court judges who die each year is 0.1. What is the probability that a president will be able to replace a Supreme Court judge during a 4 year term?

Let N = the number of Supreme Court deaths in 4 years. Then, $\lambda = E[N] = 0.4$. We want to find $P(N \ge 1)$.

$$P(N \ge 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - e^{-0.4} \approx \boxed{32.97\%}$$

Example 3.25. The average number of times Amazon ships the wrong package to a particular customer in a given year is 2. What is the probability that they ship at least 3 wrong people?

We know $\lambda = E[N] = 2$ and therefore $N \sim \text{Pois}(2)$. We want to find $P(X \ge 3)$:

$$P(X \ge 3) = 1 - P(X < 3) = 1 - P(X=0) - P(X=1) - P(X=2)$$

$$= 1 - e^{-2} - \frac{2e^{-2}}{1!} - \frac{2^2e^{-2}}{2!} = 1 - \frac{5}{e^2} \approx \boxed{32.3\%}$$

While is this mostly irrelevant, there is a neat connection between Poisson and Binomial variables.

Theorem 3.26 (Binomial and Poisson Equivalence). *For small values of p and large values of n,*

$$\textit{Binom}(n, p) \approx \textit{Pois}(np)$$

Recall that Binomial distributions operate under a finite number of small-chance trials and Poisson distributions count the number of rare independent events. The theory states in practice, when you have many small-chance trials, these two situations are basically the same!

Suppose n is very large and p is extremely small. Then

- The expected number of successes $\lambda = np$ is moderate
- It's very unlikely two successes happen in the same "small cross-section" of trials
- Each success is like a rare event occurring independently of the others

. . . such is the basis of a Binomial distribution!

Example 3.27. Scientists are testing for a very rare disease that has a 0.1% chance of being found in each person. If they test 1000 people, what is the probability that at least 1 has it?

We will calculate the probability with both Binomial and Poisson distributions and find that the results are equivalent:

$$X = \text{Binom}(1000, 0.001) \qquad Y = \text{Pois}(1000(0.001)) = \text{Pois}(1)$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{1000}{0}(0.001)^0(.999)^{1000} \approx 0.632$$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-1} \approx 0.632$$

Suppose we had two independent Poisson distributions. How are they summed?

Example 3.28. If X is Poisson with mean 1.7 and Y is an independent Poisson with mean 1.3, find (a) $E[X + Y]$, (b) $\text{Var}(X + Y)$, (c) $P(X + Y = 2)$

(a) $E[X + Y] = E[X] + E[Y] = 1.7 + 1.3 = 3$

(b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = E[X] + E[Y] = 3$ by independence.

(c) Find all combinations in which $X + Y = 2$:

$$P(X + Y = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 1) + P(X = 2, Y = 0)$$

$$= e^{-1.7}\frac{1.3^2}{2}e^{-1.3} + 1.7e^{-1.7} \cdot 1.3e^{-1.3} + \frac{1.7^2}{2}e^{-1.7}e^{-1.3}$$

$$= \boxed{4.5e^{-3} = 22.4\%}$$

The upcoming theorem will show that the answer in the previous example is consistent with $X + Y \sim \text{Pois}(3)$.

Theorem 3.29 (Sums of Poisson Variables). *If $N \sim Pois(\lambda), M \sim Pois(\mu)$ and they are independent, then*

$$P(N + M = n) = e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^n}{n!}$$

Proof.

$$P(N + M = n) = \sum_{k=0}^{n} P(N = k)P(M = n - k)$$

$$= \sum_{k=0}^{n} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \cdot e^{-\mu} \cdot \frac{\mu^{n-k}}{(n-k)!}$$

$$= e^{-(\lambda+\mu)} \left(\sum_{k=0}^{n} \frac{\lambda^k \mu^{n-k}}{k!(n-k)!} n! \right) \cdot \frac{1}{n!} = \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \lambda^k \cdot \mu^{n-k} \binom{n}{k}$$

The summation is equal to the binomial expansion $(\lambda + \mu)^n$ by Theorem 2.17.

$$P(N + M = n) = e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^n}{n!} = P(\text{Pois}(\lambda + \mu) = n)$$

□

The results from the theorem and previous example can be generalized to multiple Poissons. If $N_1, \ldots, N_k$ are *independent* Poissons, then their sum is also Poisson. Moreover,

$$E\left[\sum N_i\right] = \sum E[N_i]$$

Example 3.30. The number of accidents per day at a busy intersection has a Poisson distribution with mean 0.5 during a workday and 0.3 during a weekend day. If the number of accidents on different days is independent, what is the probability that there will be exactly three accidents at the intersection during a week?

The sum of independent Poisson variables is also a Poisson variable, so the number of accidents per week is a Poisson with mean $5(0.5) + 2(0.3) = 3.1$.

$$P(N = 3) = e^{-\lambda} \frac{\lambda^3}{3!} = e^{-3.1} \frac{(3.1)^3}{3!} \approx \boxed{22.37\%}$$

Below is a table summarizing key properties of the Geometric, Negative Binomial, and Poisson Distributions.

Disclaimer: We assume the series start at 0 for Geometric and Negative Binomial. This only affects the expected value.

	Geometric	Negative Binomial	Poisson
Experiment Setup	• Number of trials until first success	• Number of trials until *r* successes	• Number of events in a fixed interval
Parameters	• *p*: probability of success	• *r*: number of successes • *p*: probability of success	• λ: average rate of occurrences
Probability Mass Function (PMF)	$P(X=n)=p(1-p)^{n-1}$	$P(X=n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}$	$P(X=n)=e^{-\lambda}\frac{\lambda^n}{n!}$
Mean	$E[X]=\frac{1}{p}-1$	$E[X]=r\left(\frac{1}{p}-1\right)$	$E[X]=\lambda$
Variance	$\text{Var}(X)=\frac{1-p}{p^2}$	$\text{Var}(X)=\frac{r(1-p)}{p^2}$	$\text{Var}(X)=\lambda$
Memoryless Property?	• Yes	• No	• No

4 Deductibles and Limits

Now, we are going to apply some probability into basic insurance methods.

4.1 Deductibles

In general, a deductible is the amount of money one must pay out-of-pocket before the rest is covered by your insurance provider. They are common among most health, auto, and home insurances.

Here's how deductibles operate:

1. If you are under insurance, you are responsible for the initial costs until your total payments reach the deductible amount.
2. Once your payments clear the deductible, insurance will cover any remaining costs.
3. Most deductibles are based annually, meaning you must meet the deductible amount each year.

How can we express this mathematically?

Definition 4.1 (Payment, Uncovered Cost, Total Loss). Suppose X represents the amount of a loss. If there is a deductible of d, then the resulting **(insurance) payment** is

$$\text{Payment} = (X - d)_+ = \begin{cases} 0, & X \leq d \\ X - d, & X > d \end{cases}$$

The **uncovered cost** to the insured, or the expense not protected/paid for by insurance policy is

$$\text{Uncovered Cost} = \min\{X, d\} = X \wedge d = \begin{cases} X, & X \leq d \\ d, & X > d \end{cases}$$

Lastly, the **total loss** is the sum of the insurance payment and uncovered cost:

$$X = (X - d)_+ + (X \wedge d)$$

Note that $\min\{X, d\}$ and $X \wedge d$ are equivalent notation-wise.

For instance, a health-care provider might offer insurance plans with annual deductibles of \$3000. Once these costs are covered, they will pay the rest. If a surgery costs \$7000 for a certain year, then insurance will cover \$4000, assuming no other payments were made.

Example 4.2. Suppose that loss amounts are uniform on $\{1, 2, 3, 4, 5\}$ and that there is a deductible of 2. What is the expected payment? What is the probability that the uncovered loss will be 2?

We will construct a table to summarize the probability, payment, and uncovered loss

x	$P(X = x)$	Payment	Uncovered Loss
1	1/5	0	1
2	1/5	0	2
3	1/5	1	2
4	1/5	2	2
5	1/5	3	2

$$E[\text{Payment}] = \left(0 \cdot \frac{1}{5}\right) + \left(0 \cdot \frac{1}{5}\right) + \left(1 \cdot \frac{1}{5}\right) + \left(2 \cdot \frac{1}{5}\right) + \left(3 \cdot \frac{1}{5}\right) = \boxed{\frac{6}{5} = 1.2}$$

$$\boxed{P(\text{Uncovered Loss} = 2) = \frac{4}{5}}$$

Theorem 4.3 (Expected Payment). *Suppose X represents the amount of a loss. Then,*

$$E[(X - d)_+] = E[X] - E[X \wedge d]$$

This is very straightfoward to prove, recalling that $X = (X - d)_+ + (X \wedge d)$. There are often fewer possible values for the uncovered loss than for the payment, which means it is often easier to find $E[X \wedge d]$ than $E[(X - d)_+]$. This is why we rearrange the terms to solve for the expected payment.

WARNING! This **only** applies to **first moments**. It is not true that

$$X^2 = (X - d)_+^2 + (X \wedge d)^2 \text{ and } E[X^2] = E[(X - d)_+^2] + E[(X \wedge d)^2]$$

Example 4.4. A farm is insured against tornado damage. During tornado season, each week has either 0 or 1 tornadoes, with a probability of 0.3 of having a tornado. The policy pays \$100 per tornado, with an annual deductible of \$50. Tornado season is 8 weeks long and the number of tornadoes in different weeks are independent. Find the expected annual insurance payment.

Let N be the number of storms, and $X = 100N$ be the total loss. The uncovered loss is either 0 (if there are no tornadoes) or 50 (if there is at least 1 tornado).

$$E[X \wedge 50] = 0 \cdot P(N = 0) + 50 \cdot P(N \geq 1) = 50(1 - 0.7^8) \approx 47.12$$

$$E[X] = 100E[N] = 100 \cdot 8 \cdot 0.3 = 240$$

Therefore, the expected insurance payment is

$$E[\text{Payment}] = 240 - 47.12 = \boxed{\$192.88}$$

Example 4.5. The number of annual N is a geometric on $\{0, 1, 2, \ldots\}$ with mean 2. Losses are insured \$100 each, with an annual deductible of \$150. Find the expected annual payment.

Recall that the mean of a geometric series, given probability p, is

$$E[N] = 2 = \frac{1-p}{p} \quad \iff \quad p = \frac{1}{3}$$

If $N = 1$, then we owe \$100. However, if $N \geq 2$, then we clear the deductible of \$150 and the rest is paid for by insurance.

$$E[\text{Uncovered Loss}] = 0 \cdot P(N = 0) + 100P(N = 1) + 150P(N \geq 2)$$

$$= 100p(1-p) + 150(1-p)^2 = \boxed{88.9}$$

$$E[\text{Payment}] = E[\text{Total Loss}] - E[\text{Uncovered Loss}] = 2(100) - 88.9 = \boxed{111.1}$$

$P(N \geq 2) = (1-p)^2$ is just the number of minimum failures.

4.2 Policy Limits

Another way for the payment to be less than the total loss is to have a policy limit.

Definition 4.6 (Policy Limit). Let X be the loss amount, and u the policy limit. With no deductible,

$$\text{Payment} = \begin{cases} X, & X \leq u \\ u, & u < X \end{cases}$$

In this case, Payment $= \min\{X, u\} = X \wedge u$.

With a deductible of d and a limit of u, then there are different types of limits. Generally, u is the maximum payment allowed. In that case,

$$\text{Payment} = \begin{cases} 0 & X \leq d \\ X - d & d < X \leq d + u \\ u & d + u < X \end{cases}$$

The expected payment is also called the *net premium* or the *benefit premium*.

Example 4.7. The number of annual losses is Poisson with mean 2.4. Each loss results in 50 in damages. Total annual claims are insured with a payment limit of 75. Find the expected annual payment.

Let N be the number of losses. The payment is 0 when $N = 0$, 50 when $N = 1$, and 75 when $N \geq 2$.

$$E[\text{Payment}] = 50P(N = 1) + 75P(N \geq 2)$$

$$= 50\left(2.4e^{-2.4}\right) + 75\left(1 - e^{-2.4} - 2.4e^{-2.4}\right) = \boxed{\$62.75}$$

Example 4.8. Loss amounts X have a binomial distribution with $n = 5$ and $p = 0.4$. If there is a deductible of 1 and a payment limit of 3, find the expected payment for a randomly selected loss.

x	0	1	2	3	4	5
$P(X = x)$	$(0.6)^5$	$5(0.4)(0.6)^4$	$10(0.4)^2(0.6)^3$	$10(0.4)^3(0.6)^2$	$5(0.4)^4(0.6)$	$(0.4)^5$
Payment	0	0	1	2	3	3
Uncovered Loss	0	1	1	1	1	2

$$E[\text{Payment}] = 0.3456 + 2(0.2304) + 3(0.0678 + 0.0102) = \boxed{1.06752}$$

Alternatively, we can compute $E[X] - E[\text{Uncovered Loss}]$

$$E[\text{Uncovered Loss}] = 1 \cdot \left(1 - (0.6)^5 - (0.4)^5\right) + 2(0.4)^5 = 0.93248$$

$$E[X] - E[\text{Uncovered Loss}] = (5 \cdot 0.4) - 0.93248 = 1.06752$$

5 Continuous Distributions and Densities

Some random variables are not discrete. Anytime you can have uncountably many possible outcomes (such as those within a given interval), we must shift our focus to using integrals.

This section highlights key differences between discrete and continuous random variables, as well as discovering properties of continuous distribution functions.

Disclaimer: This section assumes we are well acquainted with basic one-dimensional calculus (limits, derivatives, evaluating integrals with substitution and by parts)

5.1 Overview

As a motivation, we will compare discrete and continuous uniform cases.

Example 5.1. Let N be an *integer* uniformly chosen from $\{1, 2, \dots, 100\}$ and let X be a *real number* chosen from (0, 100). Then

$$P(N = n) = \frac{1}{100} \text{ and } P(N \leq n) = \frac{n}{100} \qquad n = 1, 2, 3, \dots, 100$$

As for the continuous case,

$$P(X = x) = 0 \text{ for all } x, P(X \leq x) = \frac{x}{100} \qquad 0 \leq x \leq 100$$

The cumulative distribution function (CDF) $F(x) = P(X \leq x)$ still makes sense for continuous distributions, and will still be useful.

Additionally, for a purely continuous function, $P(X = x) = 0$. There will be a more intuitive reason for this later.

For discrete random variables, we often summed expressions that involved $P(X = x)$ such as

$$E[X] = \sum xP(X = x)$$

As alluded to earlier, we will need integrals for continuous variables and the sums will become the "density" of X. $f(x)$ will replace $P(X = x)$ in most formulas. For example,

$$E[X] = \int xf(x)dx$$

More on this later.

Not all distributions are purely discrete or purely continuous! A mixed distribution blends them together. For instance, we can add deductibles and limits (discrete) to continuous loss amounts.

Example 5.2. Losses X are uniformly distributed on (0, 100). Let Y be the payment amount after a deductible of 30 is applied to the loss.

The deductible of 30 means that $Y = 0$ if the loss X is less than 30, and $Y = X - 30$ if the loss exceeds 30. Therefore, $X = Y + 30$ and

$$P(Y = 0) = P(X \leq 30) = \frac{30}{100}$$

$$P(Y = y) = 0 \text{ for } y > 0$$

$$P(Y \leq y) = P(X \leq y + 30) = \frac{y + 30}{100} \quad \text{for } 0 < y < 70$$

Y has a discrete piece, where the chance of being 0 is $\frac{30}{100}$, and a continuous piece (from 0 to 70), and the CDF makes sense everywhere.

Example 5.3. If N is uniform on $\{1, 2, 3, 4, 5\}$ and X is uniform on (0, 5), find $P(N \leq 2.3)$ and $P(X \leq 2.3)$.

$$P(N \leq 2.3) = P(N = 1) + P(N = 2) = \frac{2}{5}$$

$$P(X \leq 2.3) = \frac{2.3}{5} = 0.46$$

Example 5.4. Loss amounts are uniform on the interval (0, 6) and insured with a deductible of 1.6. Find the probabilities that (a) the payment for a randomly chosen loss is 0 and (b) the payment for a randomly chosen loss is less than 2.

Let X denote the amount of a randomly chosen loss, and Y the corresponding payment.

(a)

$$P(Y = 0) = P(X \leq 1.6) = \frac{1.6}{6} \approx \boxed{26.67\%}$$

(b)

$$P(Y \leq 2) = P(X \leq 2 + 1.6) = P(X \leq 3.6) = \frac{3.6}{6} = \boxed{60\%}$$

5.2 Densities and CDFs

Generally if X can reach any possible value between (a, b), the probability of it being exactly one of those numbers becomes infinitely small, and we say $P(X = t) = 0$. Instead, we focus on the cumulative distribution function.

Definition 5.5. The **cumulative distribution function (CDF)** of X is given by

$$F(x) = F_X(x) = P(X \leq x).$$

This applies to all random variables, whether they have discrete, continuous, or mixed distributions.

If F_x is differentiable, its derivative

$$f_X(t) = F'(x)$$

is referred to as the **density** of X. By the Fundamental Theorem of Calculus, the CDF is then

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

In the discrete case, $F(x) = P(X \leq x) = \sum_{y \leq x} P(X = y)$. In most formulas, $f(y)dy$ will take the place of $P(X = y)$. In some sense, $f(y)dy$ "=" $P(y < X \leq y + dy)$.

Corollary 5.6 (Properties of CDFs). *The cumulative distribution function satisfies*

1. $0 \leq F(x) \leq 1$
2. *If* $x \leq y$ *then* $F(x) \leq F(y)$
3. $\lim_{x \to \infty} F(x) = 1$
4. $\lim_{x \to -\infty} F(x) = 0$

Corollary 5.7 (Properties of Densities). *For continuous* X*, the density* $f_X(t)$ *satisfies*

1. $f(x) \geq 0$
2. *There need not be an upper bound for* $f(x)$
3. $\int_{-\infty}^{\infty} f(x)dx = 1$
4. $\int_a^b f(x)dx = P(a < X \leq b) = F(b) - F(a)$

Each item should feel intuitive. Since $F(x)$ is a probability function, its range must be 0 to 1. Probabilities will only increase as we widen the range of our interval. The total probability, over the entire x-axis, will approach 1. Lastly, if the bounds of the integrals are equal, the integral becomes 0.

$f(x)$ needs to be strictly non-negative. Probabilities are obtained by taking the area under $f(x)$. If $f(x)$ is negative anywhere, then there exists an interval in which the probability is negative.

Example 5.8. Suppose X is uniform on $(0, 0.1)$. What are $F(x)$ and $f(x)$?

One way to approach this is to come up with $f(x)$ such that its area from 0 to 0.1 is equal to 1. We already know X is uniform, so the values on f must be equal on that range. Say c is this constant, then

$$F(x) = \int_0^{0.1} c dx = 1 \implies [cx]_0^{0.1} = 1 \iff 0.1c = 1 \iff c = 10$$

We have effectively found both $F(x)$ and $f(x)$, since F is the definite integral of f.

$$\boxed{F(x) = \begin{cases} 0 & x < 0 \\ 10x & 0 \leq x \leq 0.1 \\ 0 & x > 0.1 \end{cases} \qquad f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ 10 & 0 < x < 0.1 \\ 0 & 0.1 < x \end{cases}}$$

Example 5.9. A modeled random variable X has the density function

$$f(x) = \begin{cases} cx^2 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Compute the probability $P(1 \leq X \leq 2)$.

We adopt a similar approach to the previous example, where we first want to solve for the constant. Item (3) from Corollary 5.7 is the key component into doing so. Because $f(x)$ is only defined on $[0, 3]$, it follows that

$$\int_{-\infty}^{\infty} f(x) dx = 1 \iff \int_0^3 cx^2 dx = 1$$

Evaluate the integral to solve for c:

$$\left[\frac{c}{3}x^3\right]_0^3 = 1 \iff 9c = 1 \iff c = \frac{1}{9}$$

Therefore, $F(x) = \frac{1}{9} \cdot \frac{1}{3}x^3 = \frac{1}{27}x^3$ and

$$\boxed{P(1 \leq X \leq 2) = F(2) - F(1) = \frac{7}{27}}$$

by item (4) of Corollary 5.7.

Example 5.10. The CDF of X satisfies

$$F(x) = \begin{cases} 0 & x < 1 \\ (x-1) - \frac{1}{4}(x-1)^2 & 1 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

Find $P(X \leq 2), P(1.5 < X \leq 2)$, and $f(x)$.

$$P(X \leq 2) = F(2) = 1 - \frac{1}{4} = \frac{3}{4}$$

$$P(1.5 < X \leq 2) = F(2) - F(1.5) = \frac{3}{4} - \left(\frac{1}{2} - \frac{1}{16}\right) = 0.3125$$

$$f(x) = F'(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{2}(x-1) & 1 < x < 3 \\ 0 & 3 < x \end{cases}$$

Example 5.11. A continuous random variable Y has density $f(y) = \frac{2}{y^3}$ for $1 < y < \infty$ and $f(y) = 0$ otherwise. Find a formula for the CDF $F(y)$ and find $P(Y \leq 4 \mid Y > 2)$.

Once again, $f(y)$ is only defined when $y > 1$. So,

$$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{1}^{y} \frac{2}{t^3}dt = \left[-\frac{1}{t^2}\right]_1^y = 1 - \frac{1}{y^2}$$

$P(Y \leq 4 \mid Y > 2)$ can be computed using what we know about conditional probability. If $A = P(Y \leq 4)$ and $B = P(Y > 2)$, then $P(A \cap B) = P(2 < Y \leq 4)$.

$$P(Y \leq 4 \mid Y > 2) = \frac{P(A \cap B)}{P(B)} = \frac{P(2 < Y \leq 4)}{P(Y > 2)} = \frac{P(2 < Y \leq 4)}{1 - P(Y \leq 2)}$$

$$= \frac{F(4) - F(2)}{1 - F(2)} = \frac{\frac{15}{16} - \frac{3}{4}}{\frac{1}{4}} = \boxed{\frac{3}{4}}$$

Definition 5.12 (Percentiles and Medians). x is a k-th **percentile** of X if $F(x) = k\%$. The **median** is the 50th percentile, so $F(x) = 0.5$ at the median.

We will only work in scenarios where there is a unique median, or where there is only one point x such that $F(x) = k\%$.

Example 5.13. Refer to the CDF in Example 5.9. Where is the 25th percentile? Median?

The 25th percentile satisfies $F(x) = 0.25$

$$\frac{1}{27}x^3 = \frac{1}{4} \quad \Longleftrightarrow \quad x^3 = \frac{27}{4} \Longleftrightarrow \boxed{x_{25\%} \approx 1.89}$$

Similarly, the median satisfies $F(x) = 0.5$

$$\frac{1}{27}x^3 = \frac{1}{2} \quad \Longleftrightarrow \quad x^3 = \frac{27}{2} \Longleftrightarrow \boxed{x_{50\%} \approx 2.38}$$

As a sanity check, we know both percentiles are within [0, 3] and their integrals will come out to 0.25 and 0.5.

5.3 Mixed Distributions

As mentioned previously, a lot of mixed distributions will manifest by merging deductibles and benefit limits with continuous loss functions.

Suppose an insurance policy has a *deductible* of d and a *payment limit* of u. A customer / insured has a loss of L.

- Insured is responsible for the first d of loss
- Insurance company / insurer pays for portion of loss that exceeds d up to a total payment u
- Insured is responsible for the rest

$$\begin{aligned} &\text{If } L < d && \text{Insurance payment} = 0 \\ &\text{If } d \le L \le d+u && \text{Insurance payment} = L - d \\ &\text{If } d+u < L && \text{Insurance payment} = u \end{aligned}$$

Remember that (1) the payment **ALWAYS** refers to the payment made by the insurer and (2) the premium is **ALWAYS** paid by the insured/client to the insurer.

Example 5.14. Suppose that loss amounts X have density $f(x) = 0.02x, 0 < x < 10$. If there is a deductible of 2 and a maximum payment of 6, then what is the probability of a payment of 5 or less? What is the probability of a payment of 6?

Since the deductible is 2, the insurer does not begin paying until the loss exceeds 2. Therefore, a payment of 5 occurs when loss is equal to 7.

$$P(\text{Payment} \le 5) = P(X \le 7) = \int_0^7 0.02x dx = 0.01(7^2) = \boxed{0.49}$$

A payment of 6 occurs when the loss is equal to 8. This is also the maximum payment, so the payment is also equal to 6 if the loss exceeds 8. We integrate from 8 to 10 because the loss density function is only defined from (0, 10).

$$P(\text{Payment} = 6) = P(X \geq 8) = \int_8^{10} 0.02dx = 0.01(10^2 - 8^2) = \boxed{0.36}$$

Example 5.15. Losses, if they occur, are uniformly distributed on the interval (100, 500). If there is a 60% probability of no loss and a 40% probability of exactly one loss, what is the CDF of the total loss amount?

Let L denote the loss amount. It follows that $P(L < 100) = P(L = 0) = 0.6$ and $P(L > 500) = 1$. Our CDF would then be

$$F_L(x) = \begin{cases} 0 & x < 0 \\ 0.6 & 0 \leq x < 100 \\ ??? & 100 \leq x < 500 \\ 1 & x \geq 500 \end{cases}$$

In a previous example, we used a unknown constant approach to find the CDF for a uniform density. Let's try it here. We know that on [100, 500], $f(x) = 0.4$.

$$\int_{100}^{500} cdx = \frac{2}{5} \quad \Longleftrightarrow \quad [cx]_{100}^{500} = 0.4 \quad \Longleftrightarrow \quad c = \frac{1}{1000}$$

What we found is the slope. To make F_L continuous, we must shift x by the upper bound, 500. Therefore,

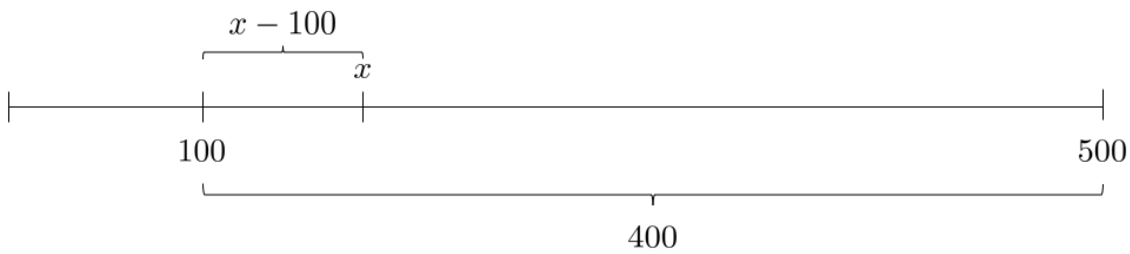
$$F_L(x) = \begin{cases} 0 & x < 0 \\ 0.6 & 0 \leq x < 100 \\ \frac{x+500}{1000} & 100 \leq x < 500 \\ 1 & x \geq 500 \end{cases}$$

Alternatively, we could solve this through conditional probability. If $100 \leq x < 500$,

$$P(L \leq x) = P(\text{no loss}) + P(L \leq x \cap \text{have a loss})$$

$$= P(\text{no loss}) + P(\text{loss}) \cdot P(L \leq x \mid \text{have a loss})$$

The probability of the loss being less than x, given that you have a loss, is equivalent to the difference between x and 100 over the length of [100, 500].

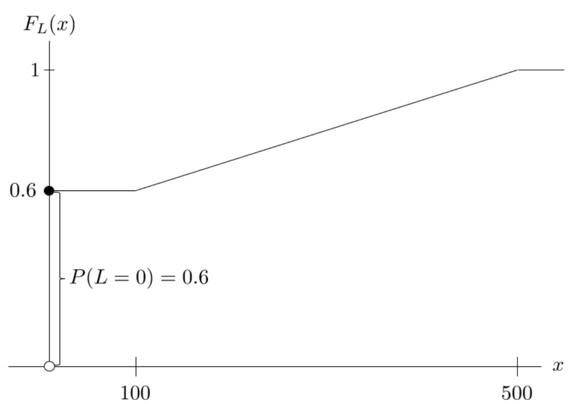


Plainly, the probability is the proportion of the subset length $(x - 100)$ over the original set length (400). This will return a set of probabilities between 0 to 1, and we multiply by 0.4 to extract the range of 0 to 0.4.

$$F_L(x) = 0.6 + 0.4\left(\frac{x - 100}{500 - 100}\right) = 0.6 + 0.4\left(\frac{x - 100}{400}\right)$$

Both results are equivalent.

The main goal of this example was to interpolate two points $(x = 100, x = 500)$ that made our piecewise-defined loss density function continuous while abiding by the uniformity between these two points.



If X is purely continuous, then $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$, and $F(x)$ is continuous.

If $F(x)$ is defined piecewise, it may not be continuous and we may have a mixed distribution. The following example is just one of many scenarios when $F(x)$ is not continuous:

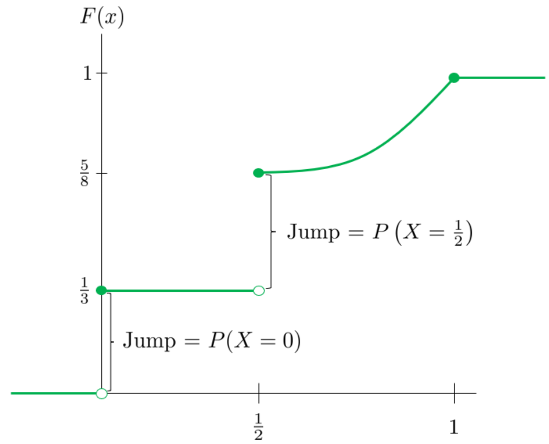
Example 5.16. Suppose X has CDF

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{3} & 0 \le x \le \frac{1}{2} \\ \frac{x^2+1}{2} & \frac{1}{2} \le x < 1 \\ 1 & 1 \le x \end{cases}$$

Then $F(0) = \frac{1}{3}$ and $F\left(\frac{1}{2}\right) = \frac{5}{8}$.

There is no jump discontinuity at 1 because the one-sided limits are equal at that point.

Jump discontinuities are points with non-zero probability.



$$P(X = 0) = \frac{1}{3} = 0 \qquad P\left(X = \frac{1}{2}\right) = \frac{5}{8} - \frac{1}{3}$$

Moreover,

$$f(x) = F'(x) = \frac{d}{dx}\left(\frac{x^2+1}{2}\right) = x$$

Example 5.17. An insurance policy pays for a random loss X subject to a deductible of d. The loss amount is a continuous random variable with density function

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

For a random loss X, the probability that the insurance payment is less than 0.3 is equal to 0.49. Find d.

$$P(\text{Payment} \le 0.3) = P(\text{Loss} \le 0.3 + d)$$

$$0.49 = \int_0^{0.3+d} f(x)dx = \int_0^{0.3+d} 2x dx \quad \Longleftrightarrow \quad 0.49 = (0.3+d)^2$$

$$0.7 = 0.3 + d \quad \Longleftrightarrow \quad \boxed{d = 0.4}$$

Example 5.18. A random variable X has CDF

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{(x-1)^2}{5} & 1 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

Find $P(X = 1), P(X = 3)$, and $f(x)$ for $1 < x < 3$

One can verify that F is continuous except for $x = 3$. Therefore, $\boxed{P(X = 1) = 0}$ because the probability at a single point is zero (if the function is continuous at that point). However, there is a jump discontinuity at $x = 3$.

$$P(X = 3) = F(3) - \lim_{x \to 3} F(x) = 1 - \frac{(3-1)^2}{5} = \boxed{\frac{1}{5}}$$

$$\boxed{f(x) = F'(x) = \frac{2(x-1)}{5} \text{ for } 1 < x < 3}$$

The following example is a sample SOA exam question concerning deductibles and CDFs:

Example 5.19 (SOA Practice Exam Q119). Damages to a car in a crash are modeled by a random variable with density function

$$\begin{cases} c(x^2 - 60x + 800) & 0 < x < 20 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant. A particular car is insured with a deductible of 2. This car was involved in a crash with resulting damages in excess of the deductible. Calculate the probability that the damages exceeded 10.

Solve for c by setting the integral equal to 1.

$$\int_0^{20} c(x^2 - 60x + 800) = 1 \iff c\left[\frac{1}{3}x^3 - 30x^2 + 800x\right]_0^2 0 = c\left(\frac{8000}{3} - 12000 + 16000\right)$$

$$\frac{20000}{3}c = 1 \iff c = \frac{3}{20000} \implies F(x) = \frac{3}{20000}\left(\frac{1}{3}x^3 - 30x^2 + 800x\right)$$

"In excess of the deductible" implies that the loss exceeds 2. So, we want to compute the probability $P(X > 10 \mid X > 2)$. We'll use the survival method to compute the probabilities:

$$P(X > 10 \mid X > 2) = \frac{P(X > 10)}{P(X > 2)} = \frac{1 - F(10)}{1 - F(2)} \approx \boxed{0.2572 = 25.72\%}$$

The probabilities could also be computed through integrals (10 to 20) and (2 to 20).

5.4 Moments of Continuous/Mixed Distributions

Mean and variance translate nicely from discrete to continuous. Recall for discrete variables,

$$E[X] = \sum xP(X = x)$$

Definition 5.20 (Mean/Variance of Continuous Random Variables). If X is random variable whose density function $f(x)$ is purely continuous, then

$$E[X] = \int_x xf(x)dx$$

Once again, if g is a function of X, then

$$E[g(X)] = \int_x g(x)f(x)dx$$

The discrete formula for variance also applies to continuous functions.

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \int_x x^2 f(x)dx - \left(\int_x xf(x)dx\right)^2$$

Example 5.21. A random variable X has density $3x^2$ for $0 < x < 1$. Find its mean and variance.

$$E[X] = \int_0^1 3x^3 dx = \left[\frac{3}{4}x^4\right]_0^1 = \boxed{\frac{3}{4}}$$

$$E[X^2] = \int_0^1 3x^4 dx = \left[\frac{3}{5}x^5\right]_0^1 = \frac{3}{5}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \boxed{\frac{3}{80}}$$

Example 5.22. If Y has density $f(y) = 1 - 0.5y$ for $0 < y < 2$, and $f(y) = 0$ otherwise, find $E[Y]$ and $\text{Var}(Y)$.

$$E[Y] = \int_0^2 \left(y - \frac{1}{2}y^2\right) dy = \left[\frac{1}{2}y^2 - \frac{1}{6}y^3\right]_0^2 = 2 - \frac{4}{3} = \frac{2}{3}$$

$$E[Y^2] = \int_0^2 \left(y^2 - \frac{1}{2}y^3\right) dy = \left[\frac{1}{3}y^3 - \frac{1}{8}y^4\right]_0^2 = \frac{8}{3} - 2 = \frac{2}{3}$$

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{2}{3} - \frac{4}{9} = \boxed{\frac{2}{9}}$$

Example 5.23 (SOA Practice Exam Q129). The proportion X of yearly dental claims that exceed 200 is a random variable with probability density function

$$f(x) = \begin{cases} 60x^3(1-x)^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate Var $\left(\frac{X}{1-X}\right)$.

We want to compute the following items: $E\left[\frac{X}{1-X}\right], E\left[\frac{X^2}{(1-X)^2}\right]$, and $\left(E\left[\frac{X}{1-X}\right]\right)^2$

$$E\left[\frac{X}{1-X}\right] = \int_0^1 60\left(\frac{x}{1-x}\right)x^3(1-x)^2dx = \int_0^1 60x^4(1-x)dx$$

$$= \int_0^1 (60x^4 - 60x^5)dx = \left[12x^5 - 10x^6\right]_0^1 = 2$$

$$E\left[\frac{X^2}{(1-X)^2}\right] = \int_0^1 60\left(\frac{x^2}{(1-x)^2}\right)x^3(1-x)^2dx = \int_0^1 60x^5dx = \left[10x^6\right]_0^1 = 10$$

$$\boxed{\text{Var}\left(\frac{X}{1-X}\right) = E\left[\frac{X^2}{(1-X)^2}\right] - \left(E\left[\frac{X}{1-X}\right]\right)^2 = 10 - 2^2 = 6}$$

What if we have mixed distributions?

1. Use the discrete formula on discrete piece
2. Use the continuous formula on the continuous piece
3. Sum the two parts

Example 5.24. Find the mean and variance of X if

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2-2x+2}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

There is a jump from 0 to $\frac{1}{2}$ as $x \to 1$. So, $P(X = 1) = F(1) - \lim\limits_{x\to 1^-} F(x) = \dfrac{1}{2}$. F is continuous at $x = 2$ because their one-sided limits are equal.
We also need to compute $f(x)$ in order to find $E[X]$.

$$f(x) = F'(x) = x - 1 \quad \text{for } 1 < x < 2$$

$$E[X] = 1 \cdot P(X = 1) + \int_1^2 x(x-1)dx = \frac{1}{2} + \left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^2$$

$$= \frac{1}{2} + \left(\frac{8}{3} - 2\right) - \left(\frac{1}{3} - \frac{1}{2}\right) = \frac{7}{3} - 1 = \boxed{\frac{4}{3}}$$

For the variance,

$$E[X^2] = 1^2 \cdot P(X = 1) + \int_1^2 x^2(x-1)dx = \frac{1}{2} + \int_1^2 (x^3 - x^2)dx$$

$$= \frac{1}{2} + \left[\frac{1}{4}x^4 - \frac{1}{3}x^3\right]_1^2 = \frac{1}{2} + \left(4 - \frac{8}{3}\right) - \left(\frac{1}{4} - \frac{1}{3}\right) = \frac{1}{2} + \frac{4}{3} + \frac{1}{12} = \frac{23}{12}$$

$$\text{Var}(X) = \frac{23}{12} - \left(\frac{4}{3}\right)^2 = \boxed{\frac{5}{36}}$$

Example 5.25. Suppose that X is a mixed random variable such that $P(X = 3) = 0.5$ and X has density $f(x) = x$ for $0 < x < 1$, and 0 otherwise. Find $E[X]$ and $E[X^2]$.

$$E[X] = 3P(X = 3) + \int_0^1 x^2 dx = \frac{3}{2} + \left[\frac{1}{3}x^3\right]_0^1 = \boxed{\frac{11}{6}}$$

$$E[X^2] = 3^2 P(X = 3) + \int_0^1 x^3 dx = \frac{9}{2} + \left[\frac{1}{4}x^4\right]_0^1 = \boxed{\frac{19}{4}}$$

5.5 The Survival Function Approach

In some cases, it may be easier to find the mean of a CDF using the survival function.

Theorem 5.26 (Mean of a CDF). *Suppose that $P(X \geq 0) = 1$ and X is continuous. Then*

$$E[X] = \int_0^\infty P(X > x)dx$$

Proof. We start with the continuous analog of $E[X]$:

$$E[X] = \int_0^\infty x f(x)dx$$

Using integration by parts,

$$u = x \implies du = dx \qquad dv = f(x)dx \implies v = F(x) - 1$$

Subtracting 1 from $F(x)$ will avoid us having problems at infinity and cause us the least trouble moving forward. It also satisfies $dv = f(x)dx$.

$$E[X] = \int_0^\infty xf(x)dx = [uv]_0^\infty - \int_0^\infty (v)du = [x(F(x) - 1)]_0^\infty$$

At $x = 0$, $x(F(x) - 1) = 0$. At $x = \infty$, $F(\infty) - 1 = 1 - 1 = 0$. This leaves us only needing to evaluate

$$-\int_0^\infty (F(x) - 1)dx = \int_0^\infty (1 - F(x))dx = \int_0^\infty P(X > x)dx = E[X]$$

as desired. □

So, for continuous, non-negative X,

$$E[X] = \int_0^\infty P(X > x)dx$$

This actually holds for all non-negative random variables, including discrete and mixed distributions. For discrete distributions,

$$\int_n^{n+1} P(X > x)dx = P(X > n) \qquad \int_0^\infty P(X > x)dx = \sum_{n=0}^{\infty} P(X > n)$$

It holds that for any non-negative variable (continuous, mixed, or discrete), if $g(0) = 0$, then

$$E[g(X)] = \int_0^\infty g'(x)P(X > x)dx$$

Unfortunately, this is rarely useful.

Advantages of Survival Method

- Often saves some steps, especially if $F(x)$ is given but $f(x)$ is not.
- Often faster for mixed distributions.
- Often gives nicer integrals (e.g., if $f(x) = e^{-x}$, integrating $xf(x)$ requires integration by parts, but the survival method does not.

Disadvantages of Survival Method

- Because the integral starts at 0, it can be messier.
- If $f(x)$ is directly given, finding $P(X > x)$ can require an extra step.

There are multiple instances where we have already employed the survival function. See Examples 5.11 and 5.19 as references.

Example 5.27 (Example 5.24 Revisited). Find the mean and variance of X if

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2-2x+2}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

$$E[X] = \int_0^\infty P(X > x)dx = \int_0^\infty (1 - F(x))dx$$

$$= \int_0^1 (1-0)dx + \int_1^2 \frac{2x - x^2}{2}dx + \int_2^\infty (1-1)dx = 1 + \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_1^2 + 0 = \boxed{\frac{4}{3}}$$

$$E[X^2] = \int_0^\infty \frac{d}{dx}(x^2)P(X > x)dx = \int_0^\infty 2x(1 - F(x))dx$$

$$= \int_0^1 2xdx + \int_1^2 2x \cdot \frac{2x - x^2}{2}dx + \int_2^\infty 0dx$$

$$= [x^2]_0^1 + \left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_1^2 = 1 + \left(\frac{16}{3} - 2\right) - \left(\frac{2}{3} - \frac{1}{4}\right) = \frac{23}{12}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{23}{12} - \frac{16}{9} = \boxed{\frac{5}{36}}$$

Example 5.28. Suppose X has density $f(x) = \frac{3(100)^3}{(x+100)^4}$ for $0 < x < \infty$ and 0 otherwise. Find $E[X]$.

One approach is through the standard definition of $E[X]$ and integrating through u-sub. However, we will use the survival method–it is just automating integration by parts!

$$f(x) = \frac{3(100)^3}{(x+100)^4}$$

$$P(X > x) = \int_x^\infty \frac{3(100)^3}{(t+100)^4}dt = \left[-\frac{(100)^3}{(t+100)^3}\right]_x^\infty = \frac{(100)^3}{(x+100)^3}$$

$$E[X] = \int_0^\infty \frac{(100)^3}{(x+100)^3}dx = \left[-\frac{1}{2} \cdot \frac{(100)^3}{(x+100)^2}\right]_0^\infty = \boxed{50}$$

Example 5.29. Use the survival approach to find $E[X]$ if the CDF of X is

$$F(x) = \begin{cases} 1 - \frac{(100)^3}{x^3} & x > 100 \\ 0 & x \le 100 \end{cases}$$

$$E[X] = \int_0^\infty P(X > x)dx = \int_0^\infty (1 - F(x))dx$$

$$= \int_0^{100} (1-0)dx + \int_{100}^\infty \frac{(100)^3}{x^3}dx = 100 - \left[\frac{(100)^3}{2x^2}\right]_{100}^\infty = 100 + 50 = \boxed{150}$$

What if we computed $E[X]$ by finding the density function?

$$f(x) = \frac{3(100)^3}{x^4} \qquad x > 100$$

$$E[X] = \int_{100}^\infty \frac{3(100)^3}{x^3}dx = -\frac{3}{2}\left[\frac{(100)^3}{x^2}\right]_{100}^\infty = 150$$

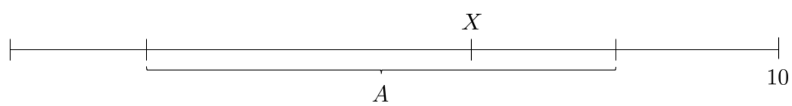
6 Key Continuous Distributions

In this section, we will be going over a myriad of continuous distributions used everywhere!

1. (Continuous) Uniform Random Variables
2. Exponential Random Variables
3. Gamma Random Variables
4. Beta and Pareto Random Variables

6.1 Continuous Uniform Distributions

Pick a point X uniformly between 0 and 10.



Using set notation,

$$P(X \in A) = \frac{\text{length of } A}{\text{total length}}$$

for $0 < x < 10$. We can find the CDF and thus density by

$$P(X \leq x) = F(x) = \frac{x}{10} \qquad f(x) = F'(x) = \frac{1}{10} = \frac{1}{\text{total length}}$$

More generally, if X is uniform on S,

$$P(X \in A) = \frac{\text{length (or area) of } A}{\text{length (or area) of } S} \qquad \text{density } = \frac{1}{\text{length of } S}$$

If X is uniform on (a, b)

$$f(x) = \frac{1}{b-a} \qquad F(x) = \frac{x-a}{b-a}$$

To set us up for moments of this distribution, let us compute the mean and variance of Uniform(0, 1). If $X \sim \text{Uniform}(0,1)$,

$$f(x) = \frac{1}{1-0} = 1$$

$$E[X] = \int_0^1 (x \cdot 1)dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2} \qquad E[X^2] = \int_0^1 (x^2 \cdot 1)dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Theorem 6.1 (Mean and Variance of Uniform Distributions). *Let $X \sim$ Uniform(a,b). Then*

$$E[X] = \frac{a+b}{2} \qquad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Proof. The idea is to shift from Uniform(0, 1) to a general uniform,

$$\text{If } X \sim \text{Uniform}(a,b),$$

$$\text{then } X - a \sim \text{Uniform}(0, b-a) \implies \frac{X-a}{b-a} \sim \text{Uniform}(0,1)$$

$$E\left[\frac{X-a}{b-a}\right] = \frac{1}{2} \implies \frac{1}{b-a}(E[X]-a) = \frac{1}{2}$$

$$E[X] = \frac{b-a}{2} + a \implies E[X] = \frac{b+a}{2} = \text{Average of endpoints}$$

For the variance:

$$\text{Var}\left[\frac{X-a}{b-a}\right] = \text{Var}(\text{Uniform}(0,1)) = \frac{1}{12}$$

We use the property of variance $\text{Var}(aX) = a^2\text{Var}(X)$ on $b-a$:

$$\text{Var}\left[\frac{X-a}{b-a}\right] = \frac{1}{(b-a)^2}\text{Var}(X-a)$$

Translations do not affect variances!

$$\frac{1}{12} = \frac{1}{(b-a)^2}\text{Var}(X) \implies \text{Var}(X) = \frac{(b-a)^2}{12} = \frac{(\text{length of interval})^2}{12}$$

□

With respect to discrete uniform variables, the expected value will be same for both, but the variances are slightly different!

Example 6.2. If N is uniform on $\{7, 8, 9, 10, 11, 12, 13\}$, find the mean and variance of N.

Refer to Def 1.38 and Example 1.39 for the discrete formulas.

$$E[N] = \frac{7+13}{2} = 10 \qquad \text{Var}(N) = \frac{(\text{number of possible vals})^2 - 1}{12} = \frac{7^2-1}{12} = 4$$

Suppose X is continuously uniform on [7, 13]. What is the mean and variance?

The mean is also 10.

$$\text{Var}(X) = \frac{(13-7)^2}{12} = 3$$

We can have mixed distributions with discrete and continuous uniforms. Raw moments (e.g., mean, 2nd moment) can be broken up into pieces.

WARNING: Variance cannot be broken up into cases without an extra correction term.

$$E[X] = \sum_x x \cdot P(X = x)$$

and from the *law of total probability* (see Thm 1.14): if $A_1, A_2, \ldots$ is a list of all possible cases

$$E[X] = \sum_{\text{all } A_i} E[X \mid X \in A_i] \cdot P[X \in A_i]$$

$$E[X^2] = \sum_{\text{all } A_i} E[X^2 \mid X \in A_i] \cdot P[X \in A_i]$$

$$E[g(X)] = \sum_{\text{all } A_i} E[g(X) \mid X \in A_i] \cdot P[X \in A_i]$$

We will mainly be incorporating this idea with deductibles.

Example 6.3. Losses X have a uniform distribution on [0, 100]. Losses are insured with a deductible. At what level must a deductible be set in order for the expected payment to be 40% of what it would be with no deductible?

We are given $X \sim U(0, 100)$, $E[X] = \frac{0+100}{2} = 50$, and d = deductible. Let Y be the payment after the deductible. Then

$$Y = \begin{cases} 0 & X \leq d \\ X - d & X > d \end{cases}$$

We need d such that $E[Y] = 0.4(50) = 20$.

The first approach we can do is to set up the relevant integral for $E[Y]$:

$$P(Y \le y) = P(X \le y + d) = \frac{y+d}{100} \quad \Longrightarrow \quad f_Y(y) = \frac{1}{100} \text{ for } y > 0$$

$$E[Y] = 0 \cdot P(Y = 0) + \int_0^{100-d} \frac{y}{100} dy = \left[\frac{y^2}{200}\right]_0^{100-d} = \frac{(100-d)^2}{200}$$

$$20 = \frac{(100-d)^2}{200} \quad \Longleftrightarrow \quad 4000 = (100-d)^2 \quad \Longleftrightarrow \quad \boxed{d = 36.75}$$

A faster way to find $E[Y]$ is to split into two cases$-X$ is at most the deductible and X exceeds the deductible$-$and use the law of total probability.

$$E[Y] = E[Y \mid X \le d] \cdot P(X \le d) + E[Y \mid X > d] P(X > d)$$

If $X \le d$, $Y = 0$. If $X > d, Y \sim \text{Uniform}(d - d = 0, 100 - d)$.

$$E[Y] = 0 \cdot P(X \ge d) + \frac{100-d}{2} \cdot \frac{100-d}{100}$$

$$20 = \frac{100-d}{2} \cdot \frac{100-d}{100} = \frac{(100-d)^2}{200}$$

This will yield the same answer of $\boxed{d = 36.75}$.

Example 6.4. A homeowner insures their home against storm damage with an insurance policy with a deductible of 50 florins. In the event of storm damage, repair costs are modeled by a uniform random variable on the interval (0, 300). Find the standard deviation of the insurance payment in the event that the home receives storm damage.

Let X = loss, Y = payment. If $X \le 50$, $Y = 0$. If $X > 50, Y \sim U(0, 250)$.

$$E[Y] = 0 \cdot P(X \le 50) + \frac{0 + 250}{2} \cdot P(X > 50) \quad = \quad 125\left(\frac{300-50}{300}\right) = 125 \cdot \frac{5}{6} \approx 104.17$$

The procedure follows identically for $E[Y^2]$:

$$E[Y^2] = E[Y^2 \mid X \le 50] \cdot P(X \le 50) + E[Y^2 \mid X > 50] \cdot P(X > 50)$$

$$0 + \frac{5}{6} E[U^2, U \sim \text{Uniform}(0, 250)]$$

Rearranging the formula for variance gives

$$E[U^2] = \text{Var}(U) + (E[U])^2$$

$$E[U^2] = \left(\frac{(250)^2}{12} + \left(\frac{250}{2}\right)^2\right) \approx 20833.33 \quad \Longrightarrow \quad E[Y^2] = \frac{5}{6} E[U^2] \approx 17361.11$$

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = 17361.11 - (104.17)^2 \approx 6509.72$$

$$\boxed{\text{SD}(Y) = \sqrt{\text{Var}(Y)} \approx 80.68}$$

Try not to get Y and U switched up! U is the uniform distribution that is based on Y, and Y is the payment that is dependent on X.

Example 6.5. Loss amounts are uniform on (0, 20), and insured with a deductible of 3 and a payment limit of 12. Find the expected payment amount and variance of the payment on a randomly selected loss.

Let X denote the loss and Y as the payment.

$$Y = \begin{cases} 0 & X \leq 3 \\ X - 3 & 3 < X \leq 15 \\ 12 & X > 15 \end{cases}$$

$$E[Y] = 0 \cdot P(X < 3) + E[U(3-3, 15-3)] \cdot P(3 < X \leq 15) + 12P(X > 15)$$

$$= \frac{12}{2} \cdot \frac{3}{5} + 12\left(\frac{1}{4}\right) = \boxed{\frac{33}{5} = 6.6}$$

Now, we compute $E[Y^2]$

$$E[Y^2] = E[U^2, U \sim \text{Uniform}(0, 12)] \cdot P(3 < X \leq 15) + 12^2 P(X > 15)$$

$$E[U^2] = \text{Var}(U) + (E[U])^2 = \frac{12^2}{12} + \left(\frac{12}{2}\right)^2 = 48$$

$$E[Y^2] = 48\left(\frac{3}{5}\right) + 144\left(\frac{1}{4}\right) = \frac{144}{5} + 36 = \frac{324}{5} = 64.8$$

$$\boxed{\text{Var}(Y) = 64.8 - (6.6)^2 = 21.24}$$

Example 6.6 (SOA Practice Exam Q336). Losses under an insurance policy are uniformly distributed on the interval [0, 100]. A deductible is set so that the expected claim payment of losses net of the deductible is 32. Calculate the deductible.

This is similar to Example 6.3. Let X denote the losses and Y be the payment, with deductible d unknown. Then

$$Y = \begin{cases} 0 & X \leq d \\ X - d & X > d \end{cases}$$

If $X > d$, then $Y \sim U(0, 100 - d)$. Apply the law of total probability to $E[Y]$:

$$E[Y] = E[Y \mid X \leq d]P(X \leq d) + E[Y \mid X > d]P(X > d)$$

$E[Y \mid X > d]$ is equivalent to finding $E[U] = \frac{100 - d}{2}$.

$$E[Y] = 0 + \frac{100 - d}{2} \cdot \frac{100 - d}{100} \quad \Longleftrightarrow \quad 32 = \frac{(100 - d)^2}{200}$$

Solving this equation gives $\boxed{d = 20}$.

Here's a standard approach to solving continuous uniform distributions with deductibles and policy limits:

1. Construct a piecewise function for Y, which is a function of X. It will always be 0 if X is less than the deductible. Add one element if there is a deductible, and two elements if there are a deductible and payment limit.
2. Make a uniform distribution U on the **length of the interval** in which the insurer is paying. This does include when after the payment limits kicks in!
3. Write an equation using the law of total probability to compute $E[Y]$ (or $E[Y^2]$ for the variance). The intervals and interval lengths of the piecewise function are useful here to compute the relevant items.
4. You should be able to compute everything but $E[U]$ (or $E[U^2]$ for the variance). Use the formula to find $E[U]$.
5. If you need to compute the variance, use the fact $E[U^2] = \text{Var}(U) + (E[U])^2$, which can be found easily. Then plug this in the equation for $E[Y^2]$.

6.2 Exponential Random Variables

Definition 6.7 (Density and CDF of Exponential Distributions). X is an **exponential random variable** with mean θ if

$$F_X(x) = 1 - e^{-\frac{x}{\theta}} \qquad 1 - F(x) = e^{-\frac{x}{\theta}}$$

Sometimes $\lambda = \frac{1}{\theta}$ will be called a rate instead of an exponential.

$$F_x(x) = 1 - e^{-\lambda x}$$

$$f(x) = F'(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} = \lambda e^{-\lambda x} \quad \text{for } x > 0$$

Exponentials often are used to model waiting times (e.g., time between hits of a webpage, time between rain drops, etc.)

Theorem 6.8 (Mean and Variance of Exponential Distributions). *Suppose X follows an exponential distribution. Then,*

$$E[X] = \theta \qquad \text{Var}(X) = \theta^2$$

Proof. We will use the survival function to prove the mean:

$$P(X > x) = 1 - F(x) = e^{-\frac{x}{\theta}}$$

$$E[X] = \int_0^\infty x f(x) dx = \int_0^\infty P(X > x) = \int_0^\infty e^{-\frac{x}{\theta}} dx$$

$$= \left[-\theta e^{-\frac{x}{\theta}}\right]_0^\infty = 0 + \theta = \theta$$

For the variance, use tabular integration to compute $E[X^2]$:

$$E[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

Differentiate	x^2	$2x$	2	0
	$+$	$-$	$+$	
Integrate	$\frac{1}{\theta} e^{-\frac{x}{\theta}}$	$-e^{-\frac{x}{\theta}}$	$\theta e^{-\frac{x}{\theta}}$	$-\theta^2 e^{-\frac{x}{\theta}}$

$$E[X^2] = \left[x^2 \left(-e^{-\frac{x}{\theta}}\right) - (2x)\left(\theta e^{-\frac{x}{\theta}}\right) + 2\left(-\theta^2 e^{-\frac{x}{\theta}}\right)\right]_0^\infty$$

At infinity, all terms will reduce to 0.

$$E[X^2] = 2\theta^2$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \theta^2$$

Therefore, $\text{Var}(X) = (E[X])^2$. □

We can think of exponential distributions as continuous analogs of geometric distributions in two senses.

For a geometric, instead of $\text{Var}(X) = (E[X])^2$,

$$\text{Var}(X) = E[\text{Geo starting at } 0] E[\text{Geo starting at } 1]$$

The second case is the memoryless property!

Suppose X is an exponential random variable with mean θ

$$P(X > x) = e^{-\frac{x}{\theta}}$$

What is $P(X > x + a \mid X > a)$?

$$P(X > x + a \mid X > a) = \frac{P(X > x + a, X > a)}{P(X > a)} = \frac{e^{-(x+a)/\theta}}{e^{-a/\theta}}$$

$$= e^{(-x-a)/\theta} e^{a/\theta} = e^{-\frac{x}{\theta}}$$

i.e., $P(X > x + a \mid X > a) = P(X > x)$. Similarly, $P(X - a > x \mid X > a) = P(X > x)$.

In other words, given that $X > a$, $X - a$ has the same distribution as the original variable X. This means that exponential distributions are no different than if a translation is applied to them!

For example, if the time between buses is exponential with mean 15 minutes, the amount of time I need to wait $(X - a)$ is an exponential with mean 15 minutes no matter how long it has been (a minutes since the last bus).

In an actuarial scenario, the key application is payment amounts $X - d$ with a deductible d conditioned on a payment being made (i.e. given $X > d$) have same distribution as losses X.

$$E[X - a \mid X > a] = E[X] = \theta$$

$$E[X \mid X > a] = E[X - a \mid X > a] + a = \theta + a$$

$$\text{Var}(X \mid X > a) = \text{Var}(X - a \mid X > a) = \text{Var}(X)$$

Example 6.9. Loss amounts are exponential with rate 0.02. If losses are insured with a deductible of 10, find the probability of a loss exceeding 40 given that a positive payment is made.

Let X denote the loss amount and $\lambda = 0.02$ as the rate. X is exponential with mean $\theta = \frac{1}{\lambda} = 50$. We want to find $P(X > 40 \mid X > 10)$ since we are told a positive payment is made by the insurer (implying $X > 10$!).

$$P(X > 40 \mid X > 10) = P(X - 10 > 30 \mid X > 10)$$

$$= P(X > 30) \text{ by the memoryless property}$$

$$= e^{-30/\theta} = \boxed{e^{-3/5} \approx 54.88\%}$$

Alternatively, one could compute $\frac{P(X>40)}{P(X>10)}$ and get the same result.

Example 6.10. Losses are exponential with mean 50, and are insured with a deductible of 10. Find the median loss amount given that a positive payment is made.

We are tasked to compute the loss L that will make the probability

$$P(X > L \mid X > 10) = P(X - 10 > L - 10 \mid X > 10) = \frac{1}{2}$$

$$\implies P(X > L - 10) = \frac{1}{2}$$

$$e^{-(L-10)/50} = \frac{1}{2} \quad \Longleftrightarrow \quad -\frac{L-10}{50} = -\ln 2 \quad \Longleftrightarrow \quad \boxed{L = 10 + 50\ln 2 \approx 44.66}$$

Example 6.11. Losses have density $f(x) = 0.1e^{-0.1x}$ for $x > 0$, and 0 otherwise. If losses are insured with a deductible of 3, find the expected payment for a randomly selected loss.

Let X denote our loss, and Y the payment. $f(x)$ is delivered in the form $\lambda e^{-\lambda x}$, so $E[X] = \theta = \frac{1}{\lambda} = 10$. Use the law of total probability to find $E[Y]$:

$$E[Y] = E[Y \mid X \leq 3]P(X \leq 3) + E[Y \mid X > 3]P(X > 3) = 0 + E[Y \mid X > 3]P(X > 3)$$

Use Definition 6.7 to compute $P(X > 3)$:

$$P(X > 3) = 1 - F(3) = e^{-0.3}$$

Recall that $Y = X - 3$ when $X > 3$ because of the deductible. The mean of $E[X - 3 \mid X > 3]$ follows immediately from the memoryless property:

$$E[Y] = E[X - 3 \mid X > 3]e^{-0.3} = 10e^{-0.3} \approx \boxed{7.41}$$

Example 6.12 (SOA Practice Exam Q28). The number of days that elapse between the beginning of a calendar year and the moment a high-risk driver is involved in an accident is exponentially distributed. An insurance company expects that 30% of high-risk drivers will be involved in an accident during the first 50 days of a calendar year. Calculate the portion of high-risk drivers are expected to be involved in an accident during the first 80 days of a calender year.

The problem tells us that $P(0 < X \leq 50) = 0.3$ given $X \sim \text{Exp}(\lambda)$ and X is the number of days between January 1 and a high-risk driver's first accident. We will use the first known probability to solve for θ:

$$P(0 < X \leq 50) = 1 - e^{-50\lambda} = 0.3 \quad \Longleftrightarrow \quad \ln(0.7) = -50\lambda$$

$$\lambda = -\frac{1}{50}\ln(.7) \approx 0.0071$$

Now, compute $P(0 < X \leq 80)$:

$$\boxed{P(0 < X \leq 80) = 1 - e^{-80\lambda} = 1 - e^{-80(0.00713)} \approx 0.435}$$

Example 6.13 (SOA Practice Exam Q115). An auto insurance policy has a deductible of 1 and a maximum claim payment of 5. Auto loss amounts follow an exponential distribution with mean 2. Calculate the expected claim payment made for an auto loss.

Let Y be the claim payment **made by the insurer** and $X \sim \text{Exp}(2)$. Then,

$$Y = \begin{cases} 0 & X \leq 1 \\ X - 1 & 1 < X \leq 6 \\ 5 & X > 6 \end{cases}$$

$$E[Y] = \int_0^\infty P(Y > y) = \int_0^\infty yf(x)$$

Where $f(x) = \frac{1}{2}e^{-x/2}$

$$E[Y] = \int_1^6 \frac{1}{2}(x-1)e^{-x/2}dx + \int_6^\infty \frac{5}{2}e^{-x/2}dx$$

The first integral requires integration by parts. Let $u = x - 1$, then $du = dx$. Let $dv = e^{-x/2}dx$, then $v = -2e^{-x/2}$:

$$\int_1^6 \frac{1}{2}(x-1)e^{-x/2}dx = \frac{1}{2}\left(\left[-2(x-1)e^{-x/2}\right]_1^6 + \int_1^6 2e^{-x/2}dx\right)$$

$$= \frac{1}{2}\left(-10e^{-3} - \left[4e^{-x/2}\right]_1^6\right) = -7e^{-3} + 2e^{-1/2}$$

Onto the second integral:

$$\int_6^\infty \frac{5}{2}e^{-x/2}dx = \frac{5}{2}\left[-5e^{-x/2}\right]_6^\infty = 5e^{-3}$$

Therefore,

$$\boxed{E[Y] = 2e^{-1/2} - 2e^{-3} \approx 1.113}$$

The integral could have also been done using tabular integration.

Example 6.14 (SOA Practice Exam Q35). The lifetime of a printer costing 200 is exponentially distributed with mean 2 years. The manufacturer agrees to pay a full refund to a buyer if the printer fails during the first year following its purchase, a one-half refund if it fails during the second year, and no refund for failure after the second year. Calculate the expected total amount of refunds from the sale of 100 printers.

Let Y be the payment of the manufacturer and $X \sim \text{Exp}(2)$. Then,

$$\begin{cases} 200 & 0 < X \leq 1 \\ 100 & 1 < X \leq 2 \\ 0 & X > 2 \end{cases}$$

We can compute using integration. Using the given distribution information, $f(x) = \frac{1}{2}e^{-x/2}$.

$$E[Y] = \int_0^\infty yf(x) = \int_0^1 200\left(\frac{1}{2}e^{-x/2}\right)dx + \int_1^2 100\left(\frac{1}{2}e^{-x/2}\right)dx$$

$$= \left[-200e^{-x/2}\right]_0^1 - \left[100e^{-x/2}\right]_1^2 = -200e^{-1/2} + 200 - 100e^{-1} + 100e^{-1/2}$$

$$= 200 - 100e^{-1} - 100e^{-0.5} \approx 102.56$$

Since we want the expect refund from 100 printers, we multiply this quantity by 100:

$$\boxed{E[100Y] = 100E[Y] = 10256}$$

Alternatively, you could compute

$$E[Y] = E[Y \mid 0 < X \leq 1]P(0 < X \leq 1) + E[Y \mid 1 < X \leq 2]P(1 < X \leq 2)$$

$$200F(1) + 100(F(2) - F(1)) \quad \text{where} \quad F(x) = 1 - e^{-x/2}$$

. . . and obtain the same answer when multiplied by 100.

6.3 Gamma, Exponential, and Poisson

Recall the general form for the density of an exponential distribution. Suppose that $X \sim \text{Exp}(\theta)$. Then

$$f(x) = \frac{1}{\theta}e^{-x/\theta} \quad \text{for } x > 0$$

The $\frac{1}{\theta}$ in front is added so $\int_0^\infty f(x)dx = 1$.

To "generalize" this, we can add a factor of $x^{\alpha-1}$ in front of the exponential, so

$$f(x) = cx^{\alpha-1}e^{-x/\theta}$$

where c is designed to make $f(x)$ a density function. The resulting distribution is called a **Gamma**(α, θ) distribution.

When α is an integer, it turns out that a Gamma is the sum of α independent identically distributed (iid) exponentials.

If $\alpha = 1$, then Gamma reduces to an exponential.

$$\alpha = 1: \qquad f(x) = \frac{1}{\theta}e^{-x/\theta}$$

$$\alpha = 2: \qquad f(x) = \frac{x}{\theta^2}e^{-x/\theta}$$

As the powers of x increase, we also need to increase the power of θ by 1 and divide by $\alpha - 1$ to neutralize and keep the total probability to 1

$$\alpha = 3: \qquad f(x) = \frac{x^2}{\theta^3}e^{-x/\theta}$$

Thus, the general form is

$$\boxed{f(x) = \frac{1}{(\alpha-1)!} \cdot \frac{x^{\alpha-1}}{\theta^\alpha}e^{-x/\theta}}$$

What are the CDFs for the corresponding values of α? At $\alpha = 1$ we have the exponential CDF.

$$\alpha = 1: \qquad F(x) = 1 - e^{-x/\theta}$$

$$\alpha = 2: \qquad F(x) = 1 - e^{-x/\theta} - \frac{x}{\theta}e^{-x/\theta}$$

$$\alpha = 3: \qquad F(x) = 1 - e^{-x/\theta} - \frac{x}{\theta}e^{-x/\theta} - \left(\frac{x}{\theta}\right)^2 \cdot \frac{1}{2}e^{-x/\theta}$$

Then, the probability $P(X < x)$ includes something are familiar with:

$$P(X \leq x) = P(X < x) = F(x) = 1 - \sum_{i=0}^{\alpha-1} P\left(\text{Poisson}\left(\frac{x}{\theta}\right) = i\right)$$

$$P(X \geq x) = P(X > x) = \sum_{i=0}^{\alpha-1} P\left(\text{Poisson}\left(\frac{x}{\theta}\right) = i\right)$$

Theorem 6.15 (Mean and Variance of Gamma Distributions). *Let $X \sim Exp(\theta)$ and $Y \sim Gamma(\alpha, \theta)$. If α is an integer, then Y is a sum of α iid $Exp(\theta)$ variables.*

$$E[Y] = \alpha E[X] = \alpha\theta \qquad Var(Y) = \alpha Var(X) = \alpha\theta^2$$

Therefore, if α is an integer, these formulas are basic properties of sums! This even holds if α is not an integer!

Example 6.16. If X is Gamma distributed with mean 10 and variance 50, find $P(X > 10)$.

Use the fact $\text{Var}(X) = \theta E[X] \implies 50 = 10\theta \implies \theta = 5$. Then, we quickly find $\alpha = 2$. Therefore, we sum up $F(x)$ at $\alpha = 1$ and $\alpha = 2$, plugging in $x = 10, \theta = 5$:

$$P(X > 10) = e^{-10/\theta} + \frac{10}{\theta}e^{-10/\theta} = e^{-2} + 2e^{-2} = \boxed{3e^{-2} \approx 0.406}$$

Example 6.17. A company has two electrical generators. The time until failure for each generator follows an exponential distribution with mean 10. The company will begin using the second generator immediately after the first one fails. What is the probability that the total time that the generators produce electricity is less than 30 hours?

Let X_1 and X_2 be generators such that

$$X_1, X_2 \sim \text{Exp}(10)$$

Define $Y = X_1 + X_2 \sim \text{Gamma}(\alpha = 2, \theta = 10)$. We want $P(Y \leq 30)$:

$$P(Y \leq 30) = P\left(\text{Poisson}\left(\frac{30}{10}\right) \geq 2\right) = F_{\alpha=2}(30)$$

$$= 1 - e^{-3} - 3e^{-3} = \boxed{1 - 4e^{-3} \approx 0.8009}$$

Example 6.18. An insured has 3 losses. If loss amounts are independent and exponentially distributed with mean 5, find the probability that the sum of the 3 losses is no more than 11.2.

The sum of independent exponentials is a Gamma:

- $\alpha = 3 =$ number of variables in sum
- $\theta = 5 =$ mean of each exponential
- Total is $\text{Gamma}(\alpha = 3, \theta = 5)$

$$F_{\alpha=3}(11.2) = 1 - e^{-11.2/5} - \frac{11.2}{5}e^{-11.2/5} - \frac{1}{2}\left(\frac{11.2}{5}\right)^2 e^{-11.2/5} \approx \boxed{0.388}$$

X is a gamma (α, θ) random variable if for $x > 0$ the density is

$$\frac{1}{(\alpha-1)!} \cdot \frac{x^{\alpha-1}}{\theta^a} e^{-x/\theta} \quad \text{for } \alpha \text{ an integer}$$

We introduce a new notation,

$$\frac{1}{\Gamma(\alpha)} \cdot \frac{x^{\alpha-1}}{\theta^a} e^{-x/\theta} \text{ for general } \alpha$$

where $\Gamma(\alpha)$ is the number such that $\int_0^\infty f(x)dx = 1$.

Some neat facts about $\Gamma(\alpha)$:

1. $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ if $\alpha - 1 > 0$
2. $\Gamma(\alpha) = (\alpha-1)!$ for a positive integer
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, $\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2} - \frac{1}{2}\right) \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$

6.4 Beta and Pareto Distributions

Historically, there have been ambiguities on Beta and Pareto Distributions−their definitions aren't consistent between readings. However, we will stick to the following definitions for future use:

Definition 6.19 (Beta Distributions). X is Beta(a, b) if $f(x) = cx^{a-1}(1-x)^{b-1}$ for $0 < x < 1$, and 0 otherwise,

$$\text{where } c = \frac{(a+b-1)!}{(a-1)!(b-1)!}$$

Moreover,

$$E[X] = \frac{a}{a+b} \qquad E[X^2] = \frac{a(a+1)}{(a+b)(a+b+1)}$$

Example 6.20. Find $E[X]$ given $f(x) = 6x(1-x)$ for $0 < x < 1$ and 0 otherwise.

$$E[X] = \int_0^1 xf(x)dx = \int_0^1 6x^2(1-x)dx = \int_0^1 (6x^2 - 6x^3)dx$$

$$= \left[2x^3 - \frac{3}{2}\right]_0^1 = \frac{1}{2}$$

Or, we can notice from Definition 6.19 that $a = 2$ and $b = 2$, so $E[X] = \frac{2}{2+2} = \frac{1}{2}$.

Now, let's move on to Pareto Distributions. Suppose $X \geq 0$ but is unbounded. We want $\int_0^1 f(x)dx = 1$, which requires $\lim_{x\to\infty} f(x) = 0$.

An exponential distribution does this with $f(x) = \lambda e^{-\lambda x}$. But x^{-p} also goes to 0. Can it be a basis for a density?

The problem lies in the fact that x^{-p} is asymptotic at 0 and we want to avoid division by 0 at 0. Here lies two solutions:

- Have $f(x) \propto (x + \theta)^{-p}$ for $x > 0$, and 0 otherwise
- Have $f(x) \propto x^{-p}$, but requires $x > \theta > 0$

Though there are other names used for both solutions, we will say the first and second are Pareto and Single parameter Pareto, respectively.

Definition 6.21 (Pareto Distributions). For $\alpha > 0, \theta > 0$, X is Pareto(α, θ) if $f(x) = 0$ for $x < 0$ and for $x > 0$,

$$f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$$

If $\alpha > 1$ then

$$E[X] = \int_0^\infty (1 - F(x))dx = \int_0^\infty \frac{\theta^\alpha}{(x+\theta)^\alpha}dx = \frac{\theta}{\alpha - 1}$$

A single parameter Pareto avoids division by 0 by starting at $\theta > 0$.

Definition 6.22 (Single Parameter Pareto Distributions). X is a single parameter Pareto (α, θ) if for $x > \theta$,

$$f(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}$$

and $f(x) = 0$ for $x < \theta$. Moreover,

$$E[X] = \int_0^\infty x\frac{\alpha\theta^\alpha}{x^{\alpha+1}}dx = \frac{\alpha\theta}{\alpha - 1}$$

Example 6.23. If $f(x) = 12x^2(1 - x)$ for $0 < x < 1$ and 0 otherwise, find Var(X)

$$E[X] = \int_0^1 x \cdot 12x^2(1-x)dx = \int_0^1 (12x^3 - 12x^4)dx = \frac{3}{5}$$

$$E[X^2] = \int_0^1 x^2 \cdot 12x^2(1-x)dx = \int_0^1 \frac{2}{5}$$

$$\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}}$$

Alternatively, we use the fact that $X \sim \mathrm{Beta}(3,2)$ and conclude that $E[X] = \frac{3}{5}$ and $E[X^2] = \frac{2}{5}$, thereby yielding the same variance.

Example 6.24. Y has density $f(y) = \frac{2(100)^2}{(y+100)^3}$ for $0 < y < \infty$ and $f(y) = 0$ otherwise. Find the 75th percentile of Y.

Let t be the 75th percentile. So $F(t) = 0.75$ and $1 - F(t) = 0.25$

$$1 - F(t) = \int_t^\infty f(y)dy \iff 0.25 = \int_t^\infty \frac{2(100)^2}{(y+100)^3}dy$$

$$0.25 = \left[\frac{(-100)^2}{(y+100)^2}\right]_t^\infty = \frac{(100)^2}{(t+100)^2}$$

$$0.5 = \frac{100}{t+100} \iff \boxed{t = 100}$$

7 Normal Approximations

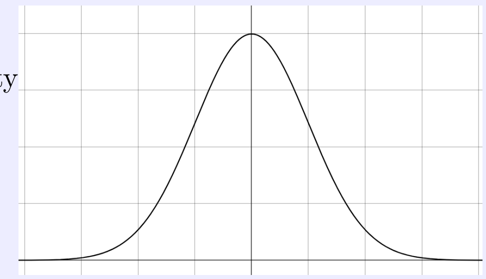
We proceed to the last main family of continuous distributions: the normal distribution!

7.1 Normal Distributions

Definition 7.1 (Normal Distributions).

A standard normal distribution has the density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



The $\frac{1}{\sqrt{2\pi}}$ term makes $\int_{-\infty}^{\infty} f(x)dx = 1$ and the $e^{-x/2}$ term in particular makes $SD(X) = 1$.

A standard normal Z has mean $\mu = 0$ and variance $\sigma = 1$.

What if we want Y to be normal with mean μ and variance σ? We can construct such a Y by rescaling Z:

$$Z \sim \mathcal{N}(0,1) \qquad Y = \sigma Z + \mu$$

$$E[Y] = \sigma E[Z] + \mu = (1 \cdot 0) + \mu = \mu$$

$$\text{Var}(Y) = \text{Var}(Y) = \text{Var}(\sigma Z) = \sigma^2 \text{Var}(Z) = \sigma^2$$

It turns out that Y is still normal, so

$$Y = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$$

Theorem 7.2 (Densities of Normal Variables). *If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then*

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}$$

Proof. Let $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ be the standard normal. Fix $Y = \sigma Z + \mu$. Then, we say $Y = g(z)$ is a function of Z. Then, equivalently,

$$z = g^{-1}(y) = \frac{y - \mu}{\sigma}$$

is the inverse of g. So, if $Y = g(Z)$,

$$f_Y(y) = f_Z(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

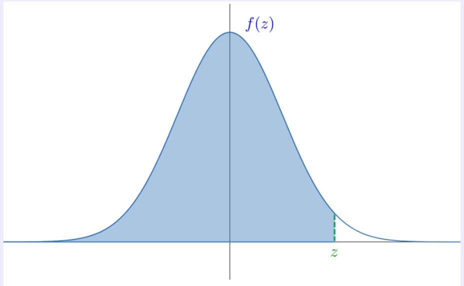
$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right) \right] \cdot \frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}$$

$\square$

Definition 7.3 (Normal CDFs). Suppose that Z is a standard normal ($Z \sim \mathcal{N}(0,1)$). Then

$$\Phi(z) = P(Z \leq z)$$

denotes the CDF of Z (the shaded area shown below)



We will be using tables to access values of $\Phi(z)$. However, the range of values for $\Phi(z)$ only include $z \geq 0$. To find the CDF for negative values, we compare $\Phi(-z)$ with $\Phi(z)$.

$$P(Z > z) = P(Z \leq -z)$$

Entries represent the area under the standardized normal distribution from $-\infty$ to z, i.e. $\Phi(z) = P(Z \leq z)$ is the CDF. The value of z to the first decimal is given in the left column. The second decimal place is given in the top row. The table shown below is what the top few entries look like:

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331

For example, $\Phi(0.12) = 0.5478$, and $\Phi(-0.33) = 1 - \Phi(0.33) = 1 - 0.6293 = 0.3707$

Example 7.4. X is normal with mean $\mu = 2$ and variance $\sigma^2 = 9$. Find $P(X > 3.86), P(X > 1.49)$ and the 95th percentile of X.

Recall that $X = \sigma Z + \mu$ for a standard normal, so using $Z = \dfrac{X-\mu}{\sigma}$ will standardize X. Convert these values to Z and use the tables!

$$P(X > 3.86) = P\left(Z > \frac{3.86-2}{\sqrt{9}}\right) = P(Z > 0.62) = 1 - \Phi(0.62) \approx \boxed{0.2676}$$

$$P(X > 1.49) = P\left(Z > \frac{1.49-2}{\sqrt{9}}\right) = P(Z > -0.17) = \Phi(0.17) \approx \boxed{0.5675}$$

To compute the 95th percentile, use the corresponding Z value and solve for X. We have that

$$\Phi(1.6449) = 0.95 \iff 1.6449 = \Phi^{-1}(0.95)$$

$$\frac{X-2}{3} = 1.6449 \iff \boxed{X \approx 6.935}$$

Theorem 7.5 (Sums of Normal Distributions). *f X and Y are independent normal distributions, then $X+Y$ is also a normal distribution.*

$$E[X+Y] = E[X] + E[Y] \qquad \mathit{Var}(X+Y) = \mathit{Var}(X) + \mathit{Var}(Y)$$

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Using general properties of mean and variance, we can take averages:

$$\frac{X+Y}{2} \sim \mathcal{N}\left(\frac{E[X]+E[Y]}{2}, \frac{1}{4}(\text{Var}(X) + \text{Var}(Y))\right)$$

Or, more generally, for any c, we have

$$cX \sim \mathcal{N}(cE[X], c^2\text{Var}(X))$$

Example 7.6. X is normal with mean -2.47 and variance 1.69. Find $P(X > 0)$.

$$P(X > 0) = P\left(Z > \frac{2.47}{\sqrt{1.69}}\right) = \boxed{1 - \Phi(1.9) \approx 0.0287}$$

As a sanity check, we know X is much larger than its mean, so we expect a small probability as the answer.

Example 7.7. If X and Y are independent normal random variables with $E[X] = 0.6, E[Y] = 3.2$, $\text{Var}(X) = 1.08$ and $\text{Var}(Y) = 1.17$, find the 75th percentile of the average of X and Y.

Let $W = \frac{X+Y}{2}$, then $\mu_W = \frac{0.6+3.2}{2} = 1.9$ and $\text{Var}(W) = \frac{1}{4}(1.08 + 1.17) = 0.5625$.
We have that $\Phi^{-1}(0.75) = 0.6745$, so

$$0.6745 = \frac{W - 1.9}{\sqrt{0.5625}} \iff \boxed{W = \frac{X+Y}{2} \approx 2.406}$$

Example 7.8 (SOA Practice Exam Q110). A wheel is spun with the numbers 1, 2, and 3 appearing with equal probability of $\frac{1}{3}$ each. If the number 1 appears, the player gets a score of 2, the player gets a score of 2. If the number 3 appears, the player gets a score of X, where X is a normal random variable with mean 3 and standard deviation 1. If W represents the player's score on 1 spin of the wheel, what is $P(W \leq 1.5)$?

We already know that rolling a 1 yields a score of less than 1.5, so the probability must be at least $\frac{1}{3}$. Rolling a 2 yields a score higher than 1.5, so the probability must be less than $\frac{2}{3}$. When we roll a 3, we have to standardize the normal distribution:

$$P(X \leq 1.5) = P\left(Z \leq \frac{1.5 - 3}{1}\right) = P(Z \leq -1.5) = 1 - \Phi(1.5) = 0.0668$$

We must multiply this by the original probability of $\frac{1}{3}$. So, the total probability is

$$P(W \leq 1.5) = \frac{1}{3}(1 + 0.0668) \approx \boxed{0.3556}$$

Example 7.9. Suppose Z is a standard normal random variable. What is $P(Z^2 > 2Z - 2)$?

This is equivalent to solving $P(Z^2 - 2Z + 2 > 0)$. The quadratic formula gives $Z < -0.73$ and $Z > 2.73$.

$$P(Z^2 > 2Z + 2) = P(Z < -0.73) + P(Z > 2.73)$$
$$= (1 - \Phi(0.73)) + (1 - \Phi(2.73)) = 2 - 0.7673 - 0.9968 \approx \boxed{0.2359}$$

7.2 Linear Interpolation

Some questions might ask for probabilities rounded to multiple decimal places. This may not be immediately obtainable from a z-score table.

Example 7.10. Let $X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2.2)$. Find $P(X > 3.3)$ to the nearest thousandth.

$$P(X > 3.3) = P\left(Z > \frac{3.3 - 4}{\sqrt{2.2}}\right) = P(Z > -0.472) = \Phi(0.472)$$

According to the table, $\Phi(0.47) = 0.6808$ and $\Phi(0.48) = 0.6844$. What is $\Phi(0.472)$?

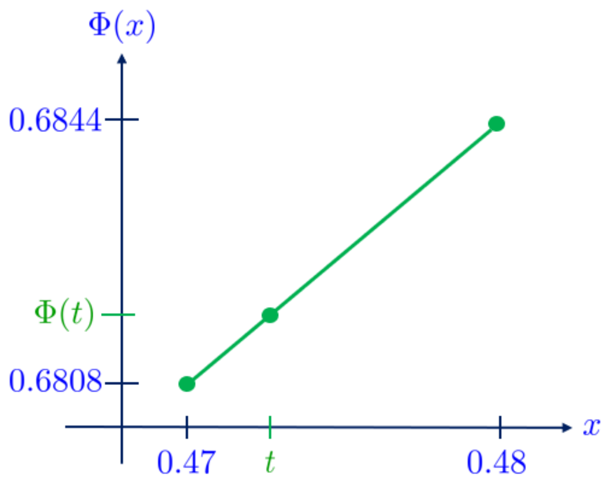
To the nearest hundredth, it is 0.68. To the nearest thousandth, it is between 0.681 and 0.684 but we cannot immediately tell. So,

$$0.6806 = \Phi(0.47) < \Phi(0.472) < \Phi(0.48) = 0.6844.$$

The true value is $\Phi(0.472) = 0.6815 \approx 0.682$.

On an exam, between the range of possible values we just described, the answer will be the true value. While this is generally a non-issue, there are two options:

1. Guesstimate. 0.472 is closer to 0.47 than 0.48. $\Phi(0.472$ will be in the middle, but closer to $\Phi(0.47)$, so probably 0682.
2. Use linear interpolation to get a better approximation.

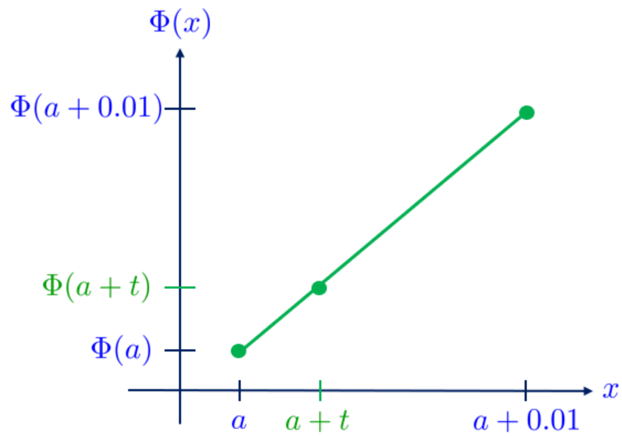


$$\frac{\Phi(t) - 0.6808}{t - 0.47} \approx \frac{0.6844 - 0.6806}{0.48 - 0.47}$$

$$\Phi(0.472) \approx \frac{0.472 - 0.47}{0.48 - 0.47}(0.6844 - 0.6808) + 0.6808$$

$$\Phi(0.472) \approx 0.2(0.6844 - 0.6808) + 0.6808$$

$$= \boxed{0.6815}$$



From the tables, we have $\Phi(a)$ and $\Phi(a+0.01)$. We want $\Phi(a+t)$.

$$\frac{\Phi(a+t)-\Phi(a)}{(t+a)-a}$$

$$= \frac{\Phi(a+0.01)-\Phi(a)}{0.01}$$

$$\Phi(a+t) = \Phi(a) + \frac{t}{0.01}\left[\Phi(a+0.01)-\Phi(a)\right]$$

Example 7.11. Suppose $X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2.2)$. Find the 13th percentile of X.

For a standard normal, the 13th percentile z is:

$$z = \Phi^{-1}(0.13) = -\Phi^{-1}(1-0.13) = -\Phi^{-1}(0.87)$$

From the table, $\Phi(1.12) = 0.8686, \Phi(1.13) = 0.8708$, so $-1.13 < z < -1.12$.

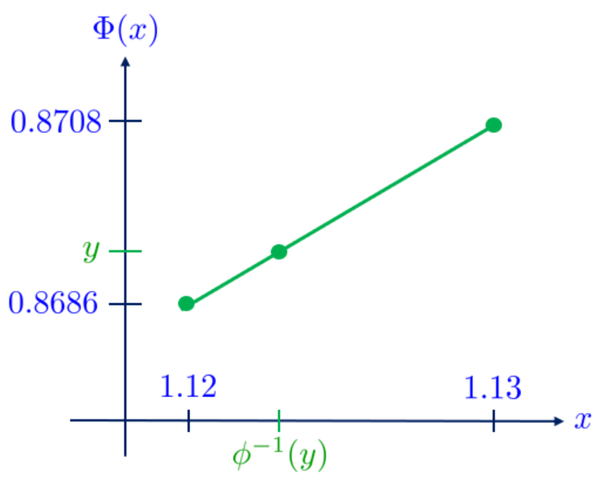
Let t denote the 13th percentile of X:

$$X = \sigma Z + \mu = \sqrt{2.2}Z + 4 \qquad \Longleftrightarrow \qquad t = \sqrt{2.2}z + 4$$

$$-1.13\sqrt{2.2} + 4 < t < -1.12\sqrt{2.2} + 4$$

$$2.324 < t < 2.339$$

$\Phi(1.13) = 0.8708$ is closer to 0.87 than $\Phi(1.12) = 0.8686$, so exact value is closer to 2.324, which corresponds to $z = 1.13$, than 2.339. So, we want $t = 4 - \sqrt{2.2}\phi^{-1}(0.87)$



$$\frac{y - 0.8686}{\Phi^{-1}(y) - 1.12} \approx \frac{0.8708 - 0.8686}{1.13 - 1.12}$$

$$\frac{0.87 - 0.8686}{\Phi^{-1}(0.87) - 1.12} \approx \frac{0.8707 - 0.8686}{1.13 - 1.12}$$

$$\frac{0.87 - 0.8686}{0.8708 - 0.8686} = \frac{\Phi^{-1}(0.87) - 1.12}{1.13 - 1.12}$$

$$\frac{14}{22} = \frac{\Phi^{-1}(0.87) - 1.12}{0.01}$$

$$\Phi^{-1}(0.87) = 1.12 + \frac{14}{22}(0.01) = 1.1264$$

$$t = 4 - \sqrt{2.2}(1.1264) = 2.329$$

Example 7.12. If X is normal with mean 0.6 and variance 1.3, find $P(X \leq 0.8)$

$$P(X \leq 0.8) = P(Z \leq 0.1754)$$

$$P(Z \leq 0.17) \approx 0.5675 \qquad P(Z \leq 0.18) \approx 0.5714$$

$$P(Z \leq 0.1754) = 0.5675 + \frac{0.0054}{0.01}(0.5714 - 0.5675) = \boxed{0.5696}$$

Example 7.13. Find the 60th percentile of X if $E[X] = 81.2$ and $\text{Var}(X) = 33.8$.

Let z denote the 60th percentile of a standard normal, and t the 60th percentile of X. We have

$$\Phi^{-1}(0.5987) < \Phi^{-1}(0.6) < \Phi^{-1}(0.6026) \quad \Leftrightarrow \quad 0.25 < \Phi^{-1}(0.6) < 0.26$$

Use linear interpolation to approximate $\Phi^{-1}(0.6)$:

$$\Phi^{-1}(0.6) = 0.25 + \frac{0.6 - 0.5987}{0.6026 - 0.5987}(0.26 - 0.25) \approx 0.2533$$

$$\boxed{x = \sigma z + \mu = \sqrt{33.8}z + 81.2 \approx 82.67}$$

7.3 The Central Limit Theorem

The main idea of the Central Limit Theorem ties in other random variables with normal distributions!

Theorem 7.14 (Central Limit Theorem (CLT)). *If $X_1, \ldots, X_n$ are identically independently distributed random variables, then*

$$\frac{(X_1 + \cdots + X_n) - nE[X_1]}{\sqrt{n\text{Var}(X_1)}} \sim \mathcal{N}(0, 1)$$

Alternatively, let $S_n = X_1 + \cdots + X_n$. Then

$$E[S_n] = nE[X_1] = n\mu \qquad \text{SD}(S_n) = \sigma\sqrt{n}$$

$$\boxed{\frac{S_n - n\mu}{\sqrt{n}\sigma} \Longrightarrow \mathcal{N}(0, 1)}$$

The theorem states that the **sum of a large number of random variables converges to a normal approximation**!

This lets us approximate the distribution of sums of random variables.

Let $S_n = X_1 + \cdots + X_n$. Finding the exact distribution of S_n is hard, but

$$P(S_n \le x) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

The first term is the normalized standard deviation, and the second term is the z-value. In summary, this is the approximation of

$$\boxed{P(S_n \le x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)}$$

and we can find these values using the z-score table!

Example 7.15. An insurance company pays claims on 625 losses. Losses are independent and exponentially distributed with mean 3. Find the approximate probability that the total payment is between 1800 and 2010.

The words "approximate probability" and the large number of random variables generally implies that the probability should be computed using a normal distribution.

We are given $S_n = X_1 + \cdots + X_{625}$ where each $X_i \sim \text{Exp}(3)$. The mean, variance, and standard deviation follow:

$$E[S_n] = 625 \cdot 3 = 1875 \qquad \text{Var}(S_n) = 625 \cdot 9 = 5625 \qquad \text{SD}(S_n) = \sqrt{\text{Var}(S_n)} = 75$$

$$\begin{aligned} P(1800 < S_n < 2010) &= P\left(\frac{1800 - 1875}{75} < \frac{S_n - E(S_n)}{\text{SD}(S_n)} < \frac{2010 - 1875}{75}\right) \\ &= P(-1 < Z < 1.8) = \Phi(1.8) - \Phi(-1) = \Phi(1.8) - (1 - \Phi(1)) \\ &= 0.9641 + 0.8413 - 1 \approx \boxed{0.8054} \end{aligned}$$

The previous example imposed the constraint that each loss was independent−we were summing them up. This is different than multiplying one loss by the number of losses. Here we distinguish the sum of losses versus a single product.

Suppose $X_1, X_2, \ldots, X_{100}$ are iid random variables with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Think of X_i as the amount that we win in bet number i.

$100X_1$ is the payoff if we bet \$100 on the first bet. It is either -100 or 100.

$\sum_{i=1}^{100} X_i$ is the net payoff after all 100 bets. We will win some and lose some, so $\sum_{i=1}^{100} X_i$ will probably be close to 0. In particular, it is closer to 0 than $100X_i$ and therefore has smaller variance.

$$\text{Var}(100X_1) = 100^2\text{Var}(X_1) = 100^2$$

$$\text{Var}\left(\sum X_i\right) = 100\text{Var}(X) = 100$$

The sum will be approximately normal, but the product will be way off! This is a key distinction and is why we do NOT take products for approximations.

Instead, suppose that $X_1, \ldots, X_n$ are iid random variables, and let $\overline{X}$ be their average. What is the distribution of $\overline{X}$?

$$S = X_1 + X_2 + \cdots + X_n \qquad \Longrightarrow \overline{X} = \frac{S}{n} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$\Longrightarrow E[\overline{X}] = \frac{1}{n}(E[X_1] + E[X_2] + \cdots + E[X_n])$$

As for the variance...

$$\text{Var}(\overline{X}) = \text{Var}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n^2}(\text{Var}(X_1) + \cdots + \text{Var}(X_n))$$

$$= \frac{1}{n^2} \cdot n\text{Var}(X) = \frac{\text{Var}(X)}{n}$$

Since S is roughly normal, so is $\frac{S}{n}$, and so we have that

$$\boxed{\overline{X} \sim \mathcal{N}\left(E[X], \frac{\text{Var}(X)}{n}\right)}$$

Example 7.16. Losses have mean 4 and standard deviation 3. If losses are independent, use a normal approximation to estimate the probability that the sum of 30 losses is at least 100.

Let X_i denote the loss amount, S the sum, and Z the standard normal. X has $\mu = 4, \sigma = 3$. Then, it follows that

$$E[S] = 30E[X] = 120, \qquad \text{Var}(S) = 30(3^2) = 270, \qquad \text{SD}(S) = \sqrt{\text{Var}(S)} \approx 16.43$$

$$P(S \geq 100) = P\left(Z \geq \frac{100 - 120}{16.43}\right) = P(Z \geq -1.22)$$

$$\boxed{P(S \geq 100) = \Phi(1.22) \approx 0.888}$$

We could obtain a more precise answer using linear interpolation:

$$\Phi(1.217) \approx \Phi(1.21) + 0.7(\Phi(1.22) - \Phi(1.21)) \approx 0.8882$$

Example 7.17. Losses are independent, each with density $0.2e^{-0.2x}$ for $x > 0$ and 0 otherwise. Losses are insured with a deductible of 5. The first 60 randomly selected positive payments are averaged. Using a normal approximation, estimate the 88th percentile of the average.

The density function tells us that $X \sim \text{Exp}(5)$. Since we are only counting positive payments, by the memoryless property, if X exceeds the deductible, the payments Y_i have the same distribution as the losses, so they are also exponential with mean 5.

$$E[\overline{Y}] = E[Y] = 5, \quad \text{Var}(\overline{Y}) = \frac{\text{Var}(Y)}{60} = \frac{5^2}{60} = 0.4167, \quad \text{SD}[\overline{Y}] = \sqrt{0.4167} = 0.6455$$

To find Z that corresponds to the 88th percentile, we can once estimate it using linear interpolation. Since $\Phi(1.17) = 0.8790$ and $\Phi(1.18) = 0.8810$, 0.88 is the midpoint between the two, making $Z = 1.175$ the best approximation.

$$1.175 = \frac{\overline{Y} - 5}{0.6455} \iff \overline{Y} \approx \boxed{5.76}$$

Therefore, the 88th percentile of the average is 5.76.

Remember that *given that a loss already exceeds the deductible, the remaining amount beyond said deductible behaves like a brand new exponential random variable with the same mean. The memoryless property applies conditionally on the loss exceeding the deductible.*

Example 7.18 (SOA Practice Exam Q66). Claims filed under auto insurance policies follow a normal distribution with mean 19,400 and standard deviation 5,000. What is the probability that the average of 25 randomly selected claims exceeds 20,000?

Use the formula that was previously derived for averages. Need to compute the variance because the standard deviation is given.

$$\overline{X} \sim \mathcal{N}\left(19400, \frac{(5000)^2}{25}\right) = \mathcal{N}(19400, 1000000)$$

We want to compute the probability $P(X \geq 20000)$:

$$P(X \geq 20000) = P\left(Z \geq \frac{20000 - 19400}{\sqrt{1000000}}\right) = P(Z \geq 0.6)$$

$$\boxed{P(X \geq 20000) = 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743}$$

Example 7.19 (SOA Practice Exam Q65). A charity receives 2025 contributions. Contributions are assumed to be independent and identically distributed with mean 3125 an standard deviation 250. Calculate the approximate 90th percentile for the distribution of the total contributions received.

Let $S = X_1 + \cdots + X_{2025}$ where each X_i has $E[X] = 3125$ and $\text{Var}(X) = 250^2 = 62500$. So,

$$E[S] = 2025(3125) = 6{,}328{,}125 \qquad \text{Var}(S) = 2025(62500)$$

We have that $P(Z < 0.9) \approx 1.2816$ using the values on the bottom. We now have all of the information required to solve for S_{90}:

$$1.2816 = \frac{S_{90} - 6{,}328{,}125}{\sqrt{2025(62500)}} = \frac{S_{90} - 6{,}328{,}125}{11{,}250}$$

$$\boxed{S_{90} = 6{,}328{,}125 + 11{,}250(1.2816) = 6{,}342{,}543}$$

7.4 Continuity Correction

In almost all cases, we will run into weird interactions with approximating discrete probabilities using continuous normal distributions.

Example 7.20. Suppose X is binomial with $n = 25$ and $p = 0.2$. Find (a) $E[X]$ and $\text{Var}(X)$, (b) An expression for the probability that X is at least 8, and (c) An expression for the probability that X is more than 8.

(a) $E[X] = np = 5, \text{Var}(X) = np(1-p) = 25 \cdot 0.2 \cdot 0.8 = 4$

(b)

$$P(X \geq 8) = \sum_{k=8}^{25} P(X = k) = \sum_{k=8}^{25} \binom{25}{k} 0.2^k (0.8)^{25-k} \approx 0.109 \text{ using a computer}$$

(c)

$$P(X > 8) = \sum_{k=9}^{25} P(X = k) = \sum_{k=9}^{25} \binom{25}{k} 0.2^k (0.8)^{25-k} \approx 0.047 \text{ using a computer}$$

Now, suppose W is normal with the same mean and variance ($\mu = 5, \sigma^2 = 4$). Find the same probabilities.

The problem lies in the fact that the probability at a single point of a continuous distribution is 0, so it would follow that $P(W > 8) = P(W \geq 8)$ for a normal distribution.

$$P(W) = P\left(\frac{W - E[W]}{\text{SD}(W)} > \frac{8-5}{\sqrt{4}}\right) = 1 - \Phi(1.5) = 0.0668$$

In those instances, X and W had the same μ and σ^2, but

- $P(X \geq 8) = 0.1091$ and $P(X > 8) = 0.0468$
- $P(W \geq 8) = 0.0668$ and $P(W > 8) = 0.0668$

While we would expect them to be closer by the Central Limit Theorem, the issue is X must be an integer, while W does not. With this in mind, how could we get a better approximation for $P(X \geq 8)$?

The idea is to let W round to something that is ≥ 8. For instance,
$P(X \geq 8) \approx P(W > 7.5)$, or $P(X > 8) = P(W > 8.5)$

$$P(W > 7.5) = 1 - \Phi\left(\frac{7.5 - 5}{2}\right) = 0.1056$$

$$P(W > 8.5) = 1 - \Phi\left(\frac{8.5 - 5}{2}\right) = 0.0401$$

There is definitely room for improvement (one way this could be more precise is by increasing n), but our approximation is much better than what is was before.

Definition 7.21 (Continuity Correction of Normal Approximations).

A **continuity correction** is used with normal approximations of integer valued variables. It corrects for the normal being continuous while the variable we care about is not. It is used when approximating a sum of S_n of *discrete* variables with a normal W.

$$P(S_n \in A) \approx P(W \text{ rounds to something} \in A)$$

There are four possible instances of rounding we can do (for the probabilities $<, \leq, >, \geq$). Rather than listing them out, it is a lot more intuitive to think them through.

Example 7.22. A roulette player betting on black has an $\frac{18}{38}$ probability of winning on each spin of the wheel. Using a normal approximation with a continuity correction, find the approximate probability of betting on black and winning at least 45 out of 100 spins.

This question implies we are dealing with a binomial distribution, with the "success" being when the ball lands on a black section. Therefore, $S_n \sim \text{Bin}\left(100, \frac{18}{38}\right)$

$$E[S_n] = np = 47.37 \qquad \text{Var}(S_n) = np(1-p) = 24.93$$

We want to approximate at least 45 wins, but with continuity correction, we should calculate the probability $P(W > 44.5)$:

$$P(W > 44.5) = P\left(Z > \frac{44.5 - 47.37}{24.93}\right) \approx P(Z > -0.575)$$

$$= \Phi(0.575) \approx \frac{1}{2}(\Phi(0.57) + \Phi(0.58)) \boxed{\approx 0.7174}$$

Example 7.23. Using a normal approximation with a continuity correction, estimate $P(X \leq 4)$ if X is Poisson with mean 2.3.

Let W be normal with the same mean and variance as X.

$$E[X] = 2.3 \qquad \text{SD}(X) = \sqrt{\text{Var}(X)} = \sqrt{2.3}$$

$$P(X \leq 4) \approx P(W < 4.5) = \Phi\left(\frac{4.5 - 2.3}{\sqrt{2.3}}\right) = \Phi(1.45) = 0.9265$$

Now, estimate $P(X = 4)$ using a normal approximation with a continuity correction. How does that compare to the true value?

$$P(X = 4) \approx P(3.5 < W < 4.5) = P(W \leq 4.5) - P(W \leq 3.5)$$

$$= \Phi(1.45) - \Phi\left(\frac{3.5 - 2.3}{\sqrt{2.3}}\right) = \Phi(1.45) - \Phi(0.79)$$

$$P(X = 4) \approx 0.9265 - 0.7852 - 0.1413$$

The true value with a Poisson distribution yields:

$$P(X = 4) = e^{-2.3} \cdot \frac{2.3^4}{4!} = 0.117$$

Generally, normal approximations for Poisson distributions improve as the mean increases.

Example 7.24. Using a normal approximation with a continuity correction, what is the approximate probability that the sum of the rolls of 15 independent six-sided dice will be at least 50? Greater than 50?

This can be modeled through a multinomial distribution, where $n = 15$ and each $p_i = \frac{1}{6}$. We expect each number to get rolled 2.5 times, so

$$E[S] = 2.5 + 5 + 7.5 + 10 + 12.5 + 15 = 52.5, \qquad \text{Var}(S) = \frac{5}{6}(52.5) = 43.75$$

We want the probability of the sum being at least 50, so we will compute $P(W > 49.5)$:

$$P(W > 49.5) = P\left(Z > \frac{49.5 - 52.5}{\sqrt{43.75}}\right) \approx P(Z > -0.454)$$

$$= \Phi(0.454) = 0.6\Phi(0.45) + 0.4\Phi(0.46) = 0.6(0.6736) + 0.4(0.6772)$$

$$\boxed{P(S \geq 45) \approx 0.675}$$

The best continuity correction for $P(S > 50)$ is by computing $P(W > 50.5)$:

$$P(W > 50.5) = P\left(Z > \frac{50.5 - 52.5}{\sqrt{43.75}}\right) \approx P(Z > -0.302)$$

$$= \Phi(0.302) = 0.8\Phi(0.30) + 0.2\Phi(0.31) = 0.8(0.6179) + 0.2(0.6217)$$

$$\boxed{P(S > 50) \approx 0.619}$$

7.5 Lognormal Random Variables

To better understand how lognormal distributions work, we motivate this concept with an example:

Example 7.25. Losses in a year have a lognormal distribution, $Y = e^X$, where X is a normal random variable with mean 3 and variance 0.5. What is the probability that losses exceed 80?

$$P(Y > 80) = P\left(e^X > 80\right) = P(X > \ln 80)$$

$$= P\left(\frac{X - E[X]}{\mathrm{SD}(X)} > \frac{4.38 - 3}{\sqrt{0.5}}\right) = 1 - \Phi\left(\frac{4.38 - 3}{0.25}\right) = 1 - \Phi(1.95) = \boxed{0.03}$$

Definition 7.26 (Lognormal Random Variables). Y is a **lognormal** if $Y = e^X$, X is normal. In words, the log of a lognormal distribution is normal.

Recall that for a normal random variable X, $-\infty < X < \infty$. Although e^X is defined, $\ln(X)$ is not.

Theorem 7.27 (Lognormal Moments). *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = e^X$ is the corresponding lognormal. Then*

$$E[Y] = e^{\mu + \sigma^2/2} \qquad E[Y^2] = e^{2\mu + 2\sigma^2} \qquad E[Y^n] = e^{n\mu + (n\sigma)^2/2}$$

We leave the proof aside due to its complexity. However, it is important to know that *for finding moments, we use these formulas about the lognormal. For finding probabilities, we take logs and work with the underlying normal.*

Example 7.28. Losses Y have a lognormal distribution with mean 50 and variance 400. What is the probability that losses exceed 75?

Use the formulas in Theorem 7.27 to find μ and σ:

$$E[Y] = 50 = e^{\mu+\sigma^2/2} \iff \ln(50) = \mu + \frac{1}{2}\sigma^2$$

$$E[Y^2] = (E[Y])^2 + \text{Var}(Y) = 50^2 + 400 = e^{2\mu+2\sigma^2} \iff \ln(50^2 + 400) = 2\mu + 2\sigma^2$$

This is a system of equations we can solve for. Double the first equation and subtract it from the second:

$$\ln(50^2 + 400) - 2\ln(50) = \sigma^2 \iff \sigma^2 \approx 0.148$$

Now, plug σ^2 back into either equation to solve for μ. We will use the first equation.

$$\mu = \ln(50) - \frac{1}{2}(0.148) \approx 3.84$$

We want to compute $P(Y > 75)$:

$$P(Y > 75) = P(X > \ln(75)) = P\left(Z > \frac{\ln(75) - 3.84}{\sqrt{.148}}\right) = P(Z > 1.24)$$

$$\boxed{P(Y > 75) = 1 - \Phi(1.24) \approx 0.11}$$

Example 7.29. Suppose X is normal with $P(X > 5) = 0.5$ and $P(X > 8) = 0.05$. Find $E[e^{2X}]$.

By symmetry of the normal distribution, we know $\mu = 5$. We also know that 8 is the 95th percentile of X. Use this to find σ:

$$0.95 = \Phi(1.6449) = P(Z < 1.6449) \iff 1.6449 = \frac{8-5}{\sigma}$$

$$\iff \sigma = \frac{3}{1.6449} = 1.824 \iff \sigma^2 = 3.33$$

Therefore, $Y = e^X \sim \ln(\mu = 5, \sigma^2 = 3.33)$.

$$E[e^{2X}] = E[Y^2] = e^{2\mu+2\sigma^2} = e^{10+2(3.33)} \approx \boxed{17{,}192{,}779}$$

Example 7.30. X is lognormal with mean 10 and variance 200. Find $P(X > 10)$.

$$E[X] = 10 = e^{\mu+\sigma^2/2} \iff \ln(10) = \mu + \frac{1}{2}\sigma^2$$

$$E[X^2] = (E[X])^2 + \text{Var}(X) = 300 = e^{2\mu+2\sigma^2} \iff \ln(300) = 2\mu + 2\sigma^2$$

Solving the system of equations yields $\sigma^2 = 1.099, \mu = 1.753$.

$$P(X > 10) = P(\ln(X) > \ln(10) = P\left(Z > \frac{\ln(10) - 1.753}{\sqrt{1.099}}\right) = 1 - \Phi(0.524)$$

$$= 1 - 0.6\Phi(0.52) - 0.4\Phi(0.53) = 1 - 0.6(0.6985) - 0.4(0.7019)$$

$$\boxed{P(X > 10) \approx 0.3}$$

Example 7.31. You are given the following: X is a random variable representing size of loss and $Y = \ln(X)$ is a random variable having a normal distribution mean of 6.503 and standard deviation of 1.5. Determine the probability that X is greater than 1000.

We have that $Y = \ln(X) \sim \mathcal{N}(\mu = 6.503, \sigma = 1.5)$.

$$P(X > 1000) = P(\ln(X) > \ln(1000)) = P\left(Z > \frac{Y - \mu_Y}{\sigma_Y}\right)$$

$$= P\left(Z > \frac{\ln(1000) - 6.503}{1.5}\right) = P(Z > 0.27)$$

$$\boxed{P(X > 1000) = 1 - \Phi(0.27) \approx 0.3936}$$

Example 7.32. Losses Y are such that $\ln(Y)$ is normally distributed. The median loss amount is 206.438, and there is a 6.68% chance that losses will exceed 354.249. Find $\text{Var}(\ln(Y))$.

We work with the underlying normal $\ln(Y)$. The medians correspond, so $\ln(Y)$ has median $\ln(206.438) = 5.33$. Since $\ln(Y)$ is normal, the median and mean are equal by symmetry. So, $\mu = 5.33$.

$$0.0668 = P\left(\frac{\ln(Y) - \mu}{\sigma} > \frac{5.87 - 5.33}{\sigma}\right)$$

$$1.5 = \frac{5.87 - 5.33}{\sigma} \iff \sigma^2 = \left(\frac{0.54}{1.5}\right)^2 = 0.1296$$

Therefore, $\boxed{\text{Var}(\ln(Y)) = 0.1296}$.

Because the ideas of a lognormal distribution can feel a bit overwhelming at first, below is a summary of the key points:

Y is lognormal if $\ln(Y)$ follows a normal distribution. If X is normal, then $Y = e^X$, because $\ln(e^X) = X$, and we already know X is normal. Moreover, $X = \ln Y$, which is important to know when we compute probabilities.

For problems with a lognormal distribution, denoting X and Y as two different distributions is important for organizing what we know about both. For instance, μ_X, σ_X^2 are the mean and variance of our normal distribution and μ_Y, σ_Y^2 is are the mean and variance of our lognormal distribution. Y, the lognormal distribution, uses μ_X and σ_X to compute its moments:

$$\mu_Y = E[Y] = e^{\mu_X + 0.5\sigma_X^2} \qquad E[Y^2] = e^{2\mu_X + 2\sigma_X^2}$$

If we have μ_X and σ_X, we are ready to compute the probability! If we are tasked to find the probability that X exceeds some value k,

$$P(X > k) = P(\ln(Y) > \ln(k)) = P\left(Z > \frac{\ln(k) - \mu_X}{\sigma_X}\right)$$

8 Multivariate Probability

In this section, we will only be talking amount multivariate probability with discrete random variables.

8.1 Joint Distributions

In 1-dimension, for a random variable X we had $\sum_x P(X = x) = 1$. This will translate nicely into 2 dimensions!

Definition 8.1 (Joint PMF). Suppose we have 2 random variables, X and Y. The **joint probability mass function** is $P(X = x, Y = y)$. The total probability is still 1:

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

Example 8.2. The joint distribution of X and Y is given by the table below:

		X		
		0	1	2
Y	1	0.1	0.2	0.3
	2	0.1	0.1	0.2

We have

$$P(X = 0, Y = 1) = 0.1$$
$$P(X = 2, Y = 2) = 0.2$$

We also verify the total probability sums to 1:

$$\sum_{x,y} P(X = x, Y = y) = 0.1+0.2+0.3+0.1+0.1+0.2 = 1$$

Example 8.3. Suppose X and Y are integer valued random variables with $1 \leq X \leq 4$ and $1 \leq Y \leq X$. If every possible outcome is equally likely, find $P(X = Y = 3)$

The possible outcomes are:

$$\{X = 1, Y = 1\}$$
$$\{X = 2, Y = 1\} \quad \{X = 2, Y = 2\}$$
$$\{X = 3, Y = 1\} \quad \{X = 3, Y = 2\} \quad \{X = 3, Y = 3\}$$
$$\{X = 4, Y = 1\} \quad \{X = 4, Y = 2\} \quad \{X = 4, Y = 3\} \quad \{X = 4, Y = 4\}$$

With 10 outcomes of equal probability, the probability is $P(X = Y = 3) = 0.1$.

Recall that in 1-dimension, the CDF of X is $F_X(x) = P(X \leq x)$. In 2-dimensions, the joint CDF of X and Y is

$$\boxed{F_{X,Y}(x, y) = P(X \leq x, Y \leq y)}$$

Theorem 8.4 (Properties of Joint CDFs).

1. $0 \leq F(x, y) \leq 1$
2. $F(x, \infty) = P(X \leq x, Y < \infty) = P(X \leq x) = F_X(x)$
3. $F(\infty, y) = P(X < \infty, Y \leq y) = P(Y \leq y) = F_y(y)$
4. $F(\infty, \infty) = 1$
5. $F(-\infty, y) = 0 = F(x, -\infty)$

Example 8.5. The joint distribution of X and Y is given by the table below:

		X = 0	X = 1	X = 2
Y	1	0.1	0.2	0.3
	2	0.1	0.1	0.2

Some values of the CDF:

$$F_{X,Y}(0, 1) = P(X \leq 0, Y \leq 1) = 0.1$$

$$F_{X,Y}(1, 1) = P(X \leq 1, Y \leq 1) = 0.3$$

$$F_{X,Y}(1, 2) = P(X \leq 1, Y \leq 2) = 0.5$$

Example 8.6. $P(X = x, Y = y) = c(x + y)$ for x and y integers such that $1 \leq x \leq 3$ and $1 \leq y \leq x$. Find c.

The possible combinations are (1, 1), (2, 1), (2, 2), (3, 1), (3, 2), (3, 3). The total probability is:

$$c(2 + 3 + 4 + 4 + 5 + 6) = 1 \quad \iff \quad \boxed{c = \frac{1}{24}}$$

Example 8.7. X and Y are integer valued random variables. If you know $F_{X,Y}(3, 3) = 1$, $F_{X,Y}(3, 2) = 0.7$, $F_{X,Y}(2, 3) = 0.6$, $F_{X,Y}(2, 2) = 0.4$, $F_{X,Y}(1, 3) = 0.3$, and $F_{X,Y}(1, 2) = 0.2$, find $P(X = 2)$.

Since $P(3, 3) = \infty$, the ranges of X and Y are given by $X \leq 3$ and $Y \leq 3$. We know that

$$P(X \leq 1, Y \leq 3) = 0.3, \qquad P(X \leq 2, Y \leq 3) = 0.6$$

Therefore, $\boxed{P(X = 2) = 0.6 - 0.3 = 0.3}$.

Example 8.8 (SOA Practice Exam Q113). Two fair dice are rolled. Let X be the absolute value of the difference between the two numbers on the dice. Calculate the probability that $X < 3$.

Intuitively, the difference is less than 3 if the two dice roll the same number, if their values are 1 apart, or if their values are 2 apart. There are 6 ways for two dice to roll the same number. There are 10 ways for two dice to be one apart (i.e. (1, 2) and (2, 1) are two different outcomes). Lastly, there are 8 ways for two dice to be two apart. There are 36 total possible outcomes. So, if X is the absolute difference between two numbers, then

$$P(X < 3) = \frac{6+10+8}{36} = \boxed{\frac{2}{3}}$$

Example 8.9 (SOA Practice Exam Q238). Skateboarders A and B practice one difficult stunt until becoming injured while attempting the stunt. On each attempt, the probability of becoming injured is p, independent of the outcomes of all previous attempts. Let $F(x, y)$ represent the probability that skateboarders A and B make no more than x and y attempts, respectively, where x and y are positive integers.

It is given that $F(2, 2) = 0.0441$. Calculate $F(1, 5)$.

We have that $X \geq 1$ and $Y \geq 1$. Each probability is independent and is equally likely (uniform), so

$$F(2,2) = F_A(2)F_B(2) = (F(2))^2 \quad \iff \quad F(2) = \sqrt{0.0441} = 0.21$$

Since we are counting the number of attempts they make, and they stop with probability p, the number of attempts each makes is a geometric starting at 1, and

$$F(2) = p + p(1-p) \quad \iff \quad 0.21 = -p^2 + 2p$$

Applying the quadratic formula:

$$p = \frac{2 \pm \sqrt{2^2 - 4(0.21)}}{2} = 0.11118$$

We use p to compute $F(1, 5)$:

$$F(1,5) = P(A = 1, B \leq 5) = P(A = 1)P(B \leq 5)$$

Since $P(B \leq 5)$ is the probability that skateboarder B does not get injured 5 times, the probability is $(1-p)^5$:

$$\boxed{F(1,5) = p(1 - P(B > 5)) = p(1 - (1-p)^5) \approx 0.0495}$$

Definition 8.10 (Marginal Distributions). For discrete variables,

$$P(X = x) = \sum_y P(X = x, Y = y)$$

is the **marginal distribution** of X. In words, it's the distribution of X without knowing Y.

Example 8.11. Let X and Y be discrete random variables with joint probability function

$$p(x, y) = \frac{2x + y}{12}$$

for $(x, y) = (0, 1), (0, 2), (1, 2), (1, 3)$, and 0 otherwise. Determine the marginal probability function of X.

X has two possible values: 0 and 1. Compute $P(X = 0)$ and $P(X = 1)$

$$P(X = 0) = P((0, 1)) + P((0, 2)) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$$

$$P(X = 1) = P((1, 2)) + P((1, 3)) = \frac{1}{3} + \frac{5}{12} = \frac{3}{4}$$

$$P(X = x) = 0 \text{ otherwise}$$

Given the same joint probability function, find $P(X = 0 \mid Y = 2)$.

Sum up the probabilities where $Y = 2$. Divide $P(X = 0 \cap Y = 2)$ by the resulting value:

$$P(X = 0 \mid Y = 2) = \frac{P(X = 0 \cap Y = 2)}{P(Y = 2)} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3}} = \frac{1}{3}$$

Definition 8.12 (Conditional Distributions). For discrete random variables X, Y,

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x, Y = y)}{\sum_x P(X = x, Y = y)}$$

The denominator makes the conditional distribution itself a probability distribution (because the sum of marginal distributions add up to 1!)

Definition 8.13 (Multivariate Independence). X and Y are **independent** if $P(X = x \mid Y = y) = P(X = x)$. For discrete variables,

$$P(X = x, Y = y) = P(Y = y) \cdot P(X = x) \quad \text{if } X \text{ and } Y \text{ are independent}$$

In particular, X and Y are independent if and only if the joint probability mass function factors as a function of x times a function of y and the range of $\{X, Y\}$ is a discrete rectangle.

Example 8.14. The joint distribution of X and Y is given by the table below:

		X = 0	X = 1	X = 2
Y	1	0.1	0.2	0.3
	2	0.1	0.1	0.2

Some conditional probabilities:

$$P(X = 0 \mid Y = 1) = \frac{0.1}{0.6} = \frac{1}{6}$$

$$P(X = 1 \mid Y = 1) = \frac{0.2}{0.6} = \frac{1}{3}$$

$$P(Y = 1 \mid X = 2) = \frac{0.3}{0.5} = \frac{3}{5}$$

$$P(Y = 2 \mid X = 2) = \frac{0.2}{0.5} = \frac{2}{5}$$

Some important points:

- The marginal distribution $P(X = x)$ of X can only involve x. It cannot involve any other variable.
- The conditional distribution $P(X = x \mid Y = y)$ can involve x and y.
- If X and Y are independent, $P(X = x \mid Y = y) = P(X = x)$. More precisely, the conditional distribution of X will only involve x. Even the range cannot depend on y.

Example 8.15. Let X and Y be discrete random variables with joint probability function

$$p(x, y) = \frac{x^2 y}{23}$$

for $(x, y) = (1, 1), (1, 2), (2, 2), (2, 3)$, and 0 otherwise. Determine the marginal probability function of X.

$$P(X = 1) = \frac{1}{23} + \frac{2}{23} = \frac{3}{23}$$

$$P(X = 2) = \frac{8}{23} + \frac{12}{23} = \frac{20}{23}$$

$$P(X = x) = 0 \text{ otherwise}$$

Now, compute $P(X = 1 \mid Y = 2)$:

$$P(X = 1 \mid Y = 2) = \frac{P(X = 1 \cap Y = 2)}{P(1, 2) + P(2, 2)} = \frac{\frac{2}{23}}{\frac{2}{23} + \frac{8}{23}} = \boxed{\frac{1}{5}}$$

Example 8.16. Suppose X and Y are discrete random variables such that

$$P(X = x, Y = y) = \begin{cases} \frac{x+y}{21} & x = 1, 2 \text{ or } 3, y = 1 \text{ or } 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $P(X = x \mid Y = 2)$.

$$P(X = x \mid Y = 2) = \frac{P(X = x \cap Y = 2)}{P(Y = 2)} = \frac{x+2}{21(P(1,2) + P(2,2) + P(3,2))}$$

$$= \frac{x+2}{21\left(\frac{3}{21} + \frac{4}{21} + \frac{5}{21}\right)} = \boxed{\frac{x+2}{12}}$$

Example 8.17 (SOA Practice Exam Q110). The probability of x losses occurring in year 1 is $(0.5)^{x+1}$ for $x = 0, 1, 2, \ldots$. The probability of y losses in year 2 given x losses in year 1 is given by the table:

Number of losses in year 1 (x)	Number of losses in year 2 (y) given x losses in year 1				
	0	1	2	3	4+
0	0.60	0.25	0.05	0.05	0.05
1	0.45	0.30	0.10	0.10	0.05
2	0.25	0.30	0.20	0.20	0.05
3	0.15	0.20	0.20	0.30	0.15
4+	0.05	0.15	0.25	0.35	0.20

Calculate the probability of exactly 2 losses in 2 years.

The possible ways to have exactly 2 losses is (2, 0), (0, 2), and (1, 1). Rearrange the formula for conditional probability to compute each:

$$P(X = 2 \cap Y = 0) = P(X = 2)P(Y = 0 \mid X = 2) = (0.5)^3(0.25) = 0.03125$$

$$P(X = 0 \cap Y = 2) = P(X = 0)P(Y = 2 \mid X = 0) = (0.5)(0.05) = 0.025$$

$$P(X = 1 \cap Y = 1) = P(X = 1)P(Y = 1 \mid Y = 1) = (0.5)^2(0.3) = 0.075$$

Therefore, the probability of exactly 2 losses in 2 years is

$$\boxed{P(2 \text{ losses in } 2 \text{ years}) = 0.03125 + 0.025 + 0.075 = 0.13125}$$

8.2 Joint Moments

Before establishing a concrete definition for the mean, let us get a sense for its computation with an example:

Example 8.18. Let X denote the number of years I will use my current computer, and Y the number of years I will use my tablet. The joint distribution of X and Y is

		X		
		1	2	3
Y	1	0.05	0.16	0.19
	2	0.10	0.13	0.12
	3	0.07	0.10	0.08

What is the average number of years that I will use my tablet?

One approach is to first find the marginal distribution of Y.

$$P(Y = 1) = 0.40 \quad P(Y = 2) = 0.35 \quad P(Y = 3) = 0.25$$

$$E[Y] = 1(0.40) + 2(0.35) + 3(0.25) = 1.85$$

A second approach is to sum over all 9 cases

$$E[Y] = 0.05 + 0.16 + 0.19 + 2(0.1) + 2(0.13) + 2(0.12) + 3(0.07) + 3(0.1) + 3(0.08) = 1.85$$

What is $E[X + Y]$?

The second approach can handle the $X + Y$ much more easily than the first:

$$\begin{aligned} E[X+Y] &= (1+1)(0.05) + (2+1)(0.16) + (3+1)(0.19) + (1+2)(0.1) \\ &\quad + (2+2)(0.13) + (3+2)(0.12) + (1+3)(0.07) + (2+3)(0.1) + (3+3)(0.08) \\ &= 4.02 \end{aligned}$$

Definition 8.19 (Multivariate Mean). For discrete random variables X and Y,

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot P(X = x, Y = y)$$

Some special cases:

$$E[X] = \sum_x \sum_y x \cdot P(X = x, Y = y)$$

$$E[Y^2] = \sum_x \sum_y y^2 P(X = x, Y = y)$$

Example 8.20. Let X and Y be discrete random variables with joint probability function

$$p(x, y) = \frac{2x + y}{12}$$

for $(x, y) = (0, 1), (0, 2), (1, 2), (1, 3)$, and 0 otherwise. Find $E[X]$ and $E[(X + 1)Y]$

$$E[X] = P(1, 2) + P(1, 3) = \frac{1}{3} + \frac{5}{12} = \frac{3}{4}$$

$$E[(X + 1)Y] = (0 + 1)(1)P(0, 1) + (0 + 1)(2)P(0, 2) + (1 + 1)(2)P(1, 2) + (1 + 1)(3)P(1, 3)$$

$$= \frac{1}{12} + \frac{1}{3} + \frac{4}{3} + \frac{5}{2} = \frac{51}{12} = \boxed{\frac{17}{4}}$$

Example 8.21 (SOA Practice Exam Q76). A car dealership sells 0, 1, or 2 luxury cars on any day. When selling a car, the dealer also tries to persuade the customer to buy an extended warranty for the car. Let X denote the number of luxury cars sold in a given day, and let Y denote the number of extended warranties sold.

$$P(X = 0, Y = 0) = \frac{1}{6} \quad P(X = 1, Y = 0) = \frac{1}{12} \quad P(X = 1, Y = 1) = \frac{1}{6}$$

$$P(X = 2, Y = 0) = \frac{1}{12} \quad P(X = 2, Y = 1) = \frac{1}{3} \quad P(X = 2, Y = 2) = \frac{1}{6}$$

What is the variance of X?

$$E[X] = \frac{1}{12} + \frac{1}{6} + 2\left(\frac{1}{12}\right) + 2\left(\frac{1}{3}\right) + 2\left(\frac{1}{6}\right) = \frac{17}{12} \approx 1.417$$

$$E[X^2] = \frac{1}{12} + \frac{1}{6} + 2^2\left(\frac{1}{12}\right) + 2^2\left(\frac{1}{3}\right) + 2^2\left(\frac{1}{6}\right) = \frac{31}{12} \approx 2.583$$

$$\boxed{\text{Var}(X) = E[X^2] - (E[X])^2 \approx 0.575}$$

Example 8.22. Let X and Y be discrete random variables such that

$$P(X = x, Y = y) = \begin{cases} \frac{2^{x+1-y}}{9} & x = 1 \text{ or } 2, y = 1 \text{ or } 2 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $E\left[\frac{X}{Y}\right]$.

$$E\left[\frac{X}{Y}\right] = P(1, 1) + \frac{1}{2}P(1, 2) + 2P(2, 1) + P(2, 2)$$

$$= \frac{2}{9} + \frac{1}{2}\left(\frac{1}{9}\right) + 2\left(\frac{4}{9}\right) + \frac{2}{9} = \boxed{\frac{25}{18}}$$

Now we take a look at metrics that compare the relationship between X and Y:

Definition 8.23 (Covariance). Suppose X, Y are two random variables. Then

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

If k is some constant, then $\text{Cov}(X,k) = 0$.

Covariances are **bilinear**, i.e.,

$$\text{Cov}(aX + bY, cZ) = \text{Cov}(aX, cZ) + \text{Cov}(bY, cZ) = ac\text{Cov}(X,Z) + bc\text{Cov}(Y,Z)$$

If we take the covariance of a random variable with respect to itself, the result reduces to the variance:

$$\boxed{\text{Cov}(X,X) = E[X \cdot X] - E(X)E(X) = \text{Var}(X)}$$

When we compute the variance of sums, the covariance pops out nicely:

$$\text{Var}(X+Y) = \text{Cov}(X+Y, X+Y) = \text{Cov}(X,X) + 2\text{Cov}(X,Y) + \text{Cov}(Y,Y)$$

$$\boxed{\text{Var}(X+Y) = \text{Var}(X) + 2\text{Cov}(X,Y) + \text{Var}(Y)}$$

A useful analog for this formula is the **sum of squares formula**, where $(a+b)^2 = a^2 + 2ab + b^2$. It translates nicely to any argument inside the variance.

If a and b are constants, then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + 2ab\text{Cov}(X,Y) + b^2\text{Var}(Y)$$

If X and Y are independent, then $E[XY] = E[X]E[Y]$ so $\text{Cov}(X,Y) = 0$. That means if X and Y are independent,

$$\boxed{\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \text{ and } \text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)}$$

Definition 8.24 (Correlation). For two random variables X and Y, the **correlation** between them is given by

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)}$$

Once again, if X and Y are independent, then $\text{Cov}(X,Y) = 0 \Longrightarrow \text{Corr}(X,Y) = 0$.

Theorem 8.25 (Properties of Correlations). *Suppose $Y = aX + b$. Then,*

$$Corr(X,Y) = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

In general, $-1 \leq Corr(X,Y) \leq 1$.

Proof. Let a, b be real numbers and $Y = aX + b$. Then,

$$\begin{aligned} \text{Cov}(X,Y) = \text{Cov}(X, aX+b) &= a\text{Cov}(X,X) + \text{Cov}(X,b) \\ &= a\text{Var}(X) \end{aligned}$$

So if $a > 0$, then $\text{Cov}(X,Y) > 0$, and if $a < 0$, $\text{Cov}(X,Y) < 0$.

$$\text{SD}(Y) = |a|\text{SD}(X)$$

$$\text{so Corr}(X,Y) = \frac{a \cdot \text{Var}(X)}{[\text{SD}(X)]^2|a|} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

□

Example 8.26. X and Y are integer valued random variables with $0 < X \leq 3$ and $1 \leq Y \leq 2$. $P(X = x, Y = y) = kxy$ when positive, for some constant k. Find $\text{Cov}(X,Y)$.

Use the Joint PDF to solve for k:

$$1 = k(1+2+2+3+4+6) \quad \iff \quad k = \frac{1}{18}$$

$$E[X] = \frac{1}{18}(P(1,1) + P(1,2) + 2P(2,1) + 2P(2,2) + 3P(3,1) + 3P(3,2)) = \frac{7}{3}$$
$$E[Y] = \frac{1}{18}(P(1,1) + P(2,1) + P(3,1) + 2P(1,2) + 2P(2,2) + 2P(3,2)) = \frac{5}{3}$$
$$E[XY] = \frac{1}{18}(P(1,1) + 2P(1,2) + 2P(2,1) + 4P(2,2) + 3P(3,1) + 6P(3,2)) = \frac{35}{9}$$

Therefore,

$$\boxed{\text{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{35}{9} - \frac{7 \cdot 5}{3 \cdot 3} = 0}$$

Recall from earlier that X and Y are independent if both

1. The support of (X,Y), i.e., the points such that $P(X = x, Y = y) > 0$, is a rectangular lattice

2. $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$

Since $P(X = x, Y = y) = kxy$, we can immediately say that $f(x, y)$ factors as a function of x times a function of y, implying these two variables are independent, and $\text{Cov}(X, Y) = 0$. This is a case where drawing out the entire calculation was unnecessary!

Corollary 8.27. *If X and Y are independent, then*

$$E[XY] = E[X]E[Y]$$

More generally,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

For instance, $E[X^2Y^3] = E[X^2]E[Y^3]$ and it would follow that $\text{Cov}(X^2, Y^3) = 0$.

Example 8.28 (SOA Practice Exam Q75). An insurance policy pays a total medical benefit consisting of two parts for each claim. Let X represent the part of the benefit that is paid to the surgeon, and let Y represent the part that is paid to the hospital. The variance of X is 5,000, the variance of Y is 10,000, and the variance of the total benefit, $X + Y$, is 17,000.

Due to increasing medical costs, the company that issues the policy decides to increase X by a flat amount of 100 per claim and Y by 10% per claim. Calculate the total variance of the total benefit after these revisions have been made.

The total benefit is given by $X + 100 + 1.1Y$.

$$\text{Var}(X + 100 + 1.1Y) = \text{Var}(X + 1.1Y) = \text{Var}(X) + 2(1.1)\text{Cov}(X, Y) + (1.1)^2\text{Var}(Y)$$

We are told that before the revisions,

$$17{,}000 = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) \iff 17{,}000 = 5{,}000 + 2\text{Cov}(X, Y) + 10{,}000$$

$$\iff \text{Cov}(X, Y) = 1{,}000$$

Plug in our known values to evaluate the new total variance:

$$\boxed{\text{Var}(X + 100 + 1.1Y) = 5{,}000 + 2(1.1)(1{,}000) + 1.21(10{,}000) = 19{,}300}$$

Example 8.29. Suppose $P(X = Y = 0) = \frac{1}{2}$ and $P(X = 1, Y = 1) = P(X = -1, Y = 1) = \frac{1}{4}$. Are X and Y independent?

We want to check if $\text{Cov}(X, Y) = 0$.

$$E[X] = \frac{1}{4} - \frac{1}{4} = 0 \qquad E[Y] = \frac{1}{4} - \frac{1}{4} = 0 \qquad E[XY] = 0$$

Therefore, $\text{Cov}(X,Y) = 0$. However, they are dependent because the support isn't rectangular. Or,

$$P(X = Y = 0) = \frac{1}{2} \neq P(X = 0)(Y = 0)$$

Example 8.30. X and Y are random variables with $\text{Corr}(X,Y) = 0.6, \text{Var}(X) = 64$, and $\text{Var}(Y) = 100$. Find $\text{Var}(2X - 3Y)$.

The formula for correlation (see Def. 8.24) requires us to know $\text{SD}(X)$ and $\text{SD}(Y)$, which are easily obtainable by taking square roots of our known variances
$\implies \text{SD}(X) = 8, \text{SD}(Y) = 10$. Now, using the formula for correlation:

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)} \iff 0.6 = \frac{\text{Cov}(X,Y)}{80}$$

$$\text{Cov}(X,Y) = 48$$

We can now calculate $\text{Var}(2X - 3Y)$:

$$\text{Var}(2X - 3Y) = 4\text{Var}(X) - 12\text{Cov}(X,Y) + 9\text{Var}(Y)$$

$$\boxed{\text{Var}(2X - 3Y) = 4(64) - 12(48) + 9(100) = 580}$$

Example 8.31 (SOA Practice Exam Q77). The profit for a new product is given by $Z = 3X - Y - 5$. X and Y are independent random variables with $\text{Var}(X) = 1$ and $\text{Var}(Y) = 2$. What is the variance of Z?

We are told X and Y are independent, so $\text{Cov}(X,Y) = 0$. Therefore,

$$\text{Var}(Z) = \text{Var}(3X - Y - 5) = \text{Var}(3X - Y)$$

$$\boxed{\text{Var}(Z) = 9\text{Var}(X) + \text{Var}(Y) = 11}$$

Example 8.32 (SOA Practice Exam Q80). Let X denote the size of a surgical claim and let Y denote the size of the associated hospital claim. An actuary is using a model in which $E[X] = 5, E[X^2] = 27.4, E[Y] = 7, E[Y^2] = 51.4$, and $\text{Var}(X + Y) = 8$. Let $C_1 = X + Y$ denote the size of the combined claims before the application of a 20% surcharge on the hospital portion of the claim, and let C_2 denote the size of the combined claims after the application of that surcharge. Calculate $\text{Cov}(C_1, C_2)$.

We are told $C_1 = X + Y$ and $C_2 = X + 1.2Y$. We can use the given information to compute $\text{Cov}(X,Y)$:

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X,Y) + \text{Var}(Y)$$

$$8 = E[X^2] - (E[X])^2 + 2\text{Cov}(X,Y) + E[Y^2] - (E[Y])^2 \iff \text{Cov}(X,Y) = 1.6$$

Apply the bilinearity of covariances to compute $\text{Cov}(C_1, C_2)$:

$$\text{Cov}(C_1, C_2) = \text{Cov}(X + Y, X + 1.2Y)$$

$$= \text{Cov}(X, X) + \text{Cov}(X, 1.2Y) + \text{Cov}(Y, X) + \text{Cov}(Y, 1.2Y)$$

By symmetry, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$:

$$= \text{Var}(X) + 1.2\text{Cov}(X, Y) + \text{Cov}(X, Y) + 1.2\text{Var}(Y)$$

$$\boxed{\text{Cov}(C_1, C_2) = (27.4 - 25) + 1.2(1.6) + 1.6 + 1.2(51.4 - 49) = 8.8}$$

8.3 Conditional Moments

Definition 8.33 (Conditional Moments). Let X and Y be two (discrete) random variables. Then,

$$E(X \,|Y = y) = \sum_x x \cdot P(X = x \mid Y = y)$$

Earlier we saw

		X		
		0	1	2
Y	1	0.1	0.2	0.3
	2	0.1	0.1	0.2

$$P(X = 0 \mid Y = 1) = \frac{1}{6} \quad P(X = 1 \mid Y = 1) = \frac{1}{3} \quad P(X = 0 \mid Y = 1) = \frac{1}{2}$$

Using conditional moments,

$$E[X \mid Y = 1] = \left(0 \cdot \frac{1}{6}\right) + \left(1 \cdot \frac{1}{3}\right) + \left(2 \cdot \frac{1}{2}\right) = \frac{4}{3}$$

$$E[X \mid Y = 2] = \left(0 \cdot \frac{0.1}{0.4}\right) + \left(1 \cdot \frac{0.1}{0.4}\right) + \left(2 \cdot \frac{0.2}{0.4}\right) = \frac{5}{4}$$

Note that $E[X \mid Y]$ is a function of Y. As a result, **it is also a random variable**!

$$P\left(E[X \mid Y] = \frac{4}{3}\right) = P(Y = 1) = 0.6$$

$$P\left(E[X \mid Y] = \frac{5}{4}\right) = P(Y = 2) = 0.4$$

$$E[E[X \mid Y] = 0.6\left(\frac{4}{3}\right) + 0.4\left(\frac{5}{4}\right) = 1.3 = E[X]$$

That is not a coincidence! $E[X \mid Y]$ is random, but $E[X]$ is non-random.

When $E[X \mid Y = y]$, the conditional expectation of X, is a nice function of y, the unconditional expectations are also closely related.

Theorem 8.34 (Double Expectation Theorem). *If X and Y are (discrete) random variables,*

$$E[X] = E[E[X \mid Y = y]]$$

$$E[X] = \sum_{\text{all } y} E[X \mid Y = y] \cdot P(Y = y)$$

The second line is the law of total probability with respect to two variables!

This is useful when the distribution of X is complicated (making $E[X]$ hard to find), but the conditional distribution $(X \mid Y)$ and distribution of Y are both nice.

Example 8.35. Let N be the value rolled by a fair six-sided die. Suppose that I then flip N independent fair coins. What is the expected number of heads? What is the variance in the number of heads?

The key is that if we know N then it is easy to find the first and second moment

$$E[\text{Heads} \mid N] = \frac{N}{2}$$

So, by double expectation,

$$E[\text{Heads}] = E[E[\text{Heads} \mid N]] = E\left[\frac{N}{2}\right] = \boxed{\frac{7}{4}}$$

For the second moment, if we know N then the number of heads (H) is binomial with N trials and $p = \frac{1}{2}$. That means

$$E[H \mid N] = \frac{N}{2}, \qquad \text{Var}(H \mid N) = \frac{N}{2^2} = \frac{N}{4}$$

$$E[H^2 \mid N] = \text{Var}(H \mid N) + (E[H \mid N])^2 = \frac{N}{4} + \frac{N^2}{4}$$

$$E[H^2] = E[E[H^2 \mid N]] = E\left[\frac{N}{4} + \frac{N^2}{4}\right]$$

Since N is uniform on $\{1, 2, \ldots, 6\}$,

$$E[N] = \frac{1+6}{2} \qquad \text{Var}(N) = \frac{6^2 - 1}{12} = \frac{35}{12}$$

$$E[N^2] = \frac{35}{12} + \left(\frac{7}{2}\right)^2 = \frac{91}{6}$$

$$E[H^2] = E\left[\frac{N}{4} + \frac{N^2}{4}\right] = \frac{7}{2 \cdot 4} + \frac{91}{6 \cdot 4} = \frac{14}{3}$$

$$\text{Var}(H) = \frac{14}{3} - \left(\frac{7}{4}\right)^2 = \boxed{\frac{77}{48}}$$

Theorem 8.36 (Law of Total Variation). *Let X, Y be random variables. Then,*

$$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E(X \mid Y))$$

Theorem 8.37 (Variance of Random Sums). *Suppose $S = X_1 + \cdots + X_N$, where $X_1, X_2, \cdots,$ are iid and N is an independent integer valued random variable. Then*

$$\text{Var}(S) = E[N]\text{Var}(X) + \text{Var}(N)(E[X])^2$$

Proof. Use the Law of Total Variation (Thm 8.36). First compute $E[S \mid N]$ and $\text{Var}(S \mid N)$:

$$E[S \mid N] = NE[X], \qquad \text{Var}(S \mid N) = N\text{Var}(X)$$

$$\text{Var}(S) = E[\text{Var}(S \mid N)] + \text{Var}(E[S \mid N])$$

$$= E[N\text{Var}(X)] + \text{Var}(NE[X])) = \boxed{E[N]\text{Var}(X) + \text{Var}(N)(E[X])^2}$$

□

Example 8.38. The number of losses N is Poisson with mean 3. Loss amounts are mutually independent, and also independent of the number of losses, and have mean 5 and variance 20. What is the expected value and variance of the sum of all losses?

Let X denote a loss amount.

$$E[S] = E[E[S \mid N]] = E[NE[X]] = E[N]E[X] = 3(5) = 15$$

As for the variance, use the Variance of Random Sums formula from Theorem 8.37. For a Poisson random variable, the mean and variance are equivalent (i.e., $\text{Var}(N) = E[N] = 3$):

$$\text{Var}(S) = E[N]\text{Var}(X) + \text{Var}(N)(E[X])^2 = 3(20) + 3(25) = \boxed{135}$$

Example 8.39 (SOA Practice Exam Q81). Two life insurance policies, each with a death benefit of 10,000 and a one-time premium of 500, are sold to a couple, one for each person. The policies will expire at the end of the tenth year. The probability that only the wife will survive at least ten years is 0.025, the probability that only the husband will survive at least ten years is 0.01, and the probability that both of them will survive at least ten years is 0.96.

What is the expected excess of premiums over claims, given that the husband survives at least ten years?

With no conditioning

	W lives	W dies
H lives	0.96	0.01
H dies	0.025	0.005

The total premium paid is 1,000. Since we are conditioning on the husband surviving, the possible total claims are either 0 (if the wife survives) or 10,000 (if the wife dies).
We have $P(H \text{ lives } \cap W \text{ lives}) = 0.96$, $P(H \text{ lives } \cap W \text{ dies}) = 0.01$, and $P(W \text{ lives } \cap H \text{ dies}) = 0.025$. Use this to compute the remaining probability. We will also need the probabilities:

$$P(W \text{ lives} \mid H \text{ lives}) = \frac{0.96}{0.97}$$

$$P(W \text{ dies} \mid H \text{ lives}) = \frac{0.01}{0.97}$$

We are tasked to compute

$$E[\text{total premiums } - \text{ total claims} \mid H \text{ lives}] = E[1{,}000 \ - \text{ total claims} \mid H \text{ lives}]$$

Use the conditional moment formula (Def. 8.33):

$$= 1000P(W \text{ lives} \mid H \text{ lives}) + (1000 - 10000)P(W \text{ dies} \mid H \text{ lives})$$

$$= 1000\left(\frac{96}{97}\right) - 9000\left(\frac{1}{97}\right) = \boxed{897}$$

Example 8.40 (SOA Practice Exam Q82). A diagnostic test for the presence of a disease has two possible outcomes: 1 for disease present and 0 for disease not present. Let X denote the state of a patient, and let Y denote the outcome of the diagnostic test. The joint probability of X and Y is given by:

$$P(X=0,Y=0)=0.8 \qquad P(X=1,Y=0)=0.050$$

$$P(X=0,Y=1)=0.025 \qquad P(X=1,Y=1)=0.125$$

Calculate $\text{Var}(Y \mid X=1)$.

First, find that $P(X=1)=0.05+0.125=0.175$.

$$P(Y=0 \mid X=1)=\frac{0.05}{0.175}, \quad P(Y=1 \mid X=1)=\frac{0.125}{0.175}$$

Since Y is binary conditional on $X=1$ with two outcomes (namely 0 and 1), we can use what we know about Bernoulli random variables:

$$\text{Var}(Y \mid X=1)=P(Y=0 \mid X=1)P(Y=1 \mid X=1)\approx \boxed{0.204}$$

...because $P(Y=0 \mid X=1)$ can be thought of as a failure (diagnostic did not detect given that the disease is present) whereas $P(Y=1 \mid X=1)$ is a success (diagnostic detected disease given the disease is present).

9 Order Statistics

9.1 Order Statistics

In short, order statistics is a way of describing the maximums and minimums of a group of random variables.

Example 9.1. Claim amounts for flood damage are independent random variables with common density function

$$f(x) = \begin{cases} \frac{4}{x^5} & \text{for } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

where x is the amount of a claim in thousands. Suppose 3 such claims X_1, X_2, X_3 will be made. Find the CDF and density of the largest of 3 claims.

Let M denote the maximum loss amount. The key idea is that $M \leq x$ if and only if each of the 3 claims are at most x.

For one claim, $f(x) = \frac{4}{x^5}$ for $x > 1$

$$P(M \leq x) = P(\text{all 3 losses } \leq x) = (P(X_1 \leq x))^3 = \left(\int_1^x \frac{4}{t^5} dt\right)^3$$

$$\left(\left[-\frac{1}{t^4}\right]_1^x\right)^3 = \left(1 - \frac{1}{x^4}\right)^3 \text{ for } x > 1$$

What about the density? We derived the CDF for M,

$$F_M(x) = P(\text{all 3 losses } \leq x) = \left(1 - \frac{1}{x^4}\right)^3, \quad x > 1$$

The density is obtained by taking the derivative:

$$f_M(x) = 3\left(1 - \frac{1}{x^4}\right)^2 \cdot \frac{4}{x^5}$$

and the density is 0 for $x < 1$ as the loss amounts must all be at least 1 so their maximum must also be at least 1 (and we know $F_M(1) = 0$).

Example 9.2. Refer to the CDF in the previous example, and once again assume 3 such claims will be made. What is the *expected value of the smallest of the three claims*?

Let Y denote the *minimum* loss amount. For minimums, the key idea is $Y > x$ only if all the individual losses exceed x. We can find $E[Y]$ in one of two ways:

$$E[Y] = \int_0^\infty P(Y > y)dy = \int_1^\infty y f_Y(y)dy$$

...of which the first is easier to compute:

For one claim, $f(x) = \frac{4}{x^5}$ for $x > 1$

$$P(Y > y) = P(\text{all 3 losses } > y) = \left(\int_y^\infty \frac{4}{t^5}dt\right)^3$$

$$= \left(\left[-\frac{1}{t^4}\right]_y^\infty\right)^3 = \left(\frac{1}{y^4}\right)^3 = \frac{1}{y^{12}}, \quad y > 1$$

Now we integrate the survival probability. For $y > 1$, we integrate our new density function. From $0 < y < 1$, the minimum must be at least 1 because $Y > x > 1$, so the survival probability is 1.

$$E[Y] = \int_0^1 1dy + \int_1^\infty \frac{1}{y^{12}}dy = 1 - \left[\frac{1}{11y^{11}}\right]_1^\infty = 1 + \frac{1}{11} = \boxed{\frac{12}{11}}$$

Alternatively, we have $F_Y(y) = 1 - \frac{1}{y^{12}}$ for $y > 1$. Then

$$f_Y(y) = \frac{12}{y^{13}} \quad \implies E[Y] = \int_1^\infty y \cdot \frac{12}{y^{13}}dy = \frac{12}{11}$$

Now, let's look at a discrete case:

Example 9.3. If I roll a fair die 5 times, what is the probability that the maximum roll is 4?

Rather than counting cases, using inequalities is a much faster approach. Let M denote the maximum roll. Then

$$P(M \leq 4) = P(\text{All rolls are } \leq 4) = \left(\frac{4}{6}\right)^5$$

Here we counted the probability of 5 rolls being at most 4. However, it is very possible that none of those 5 rolls are actually 4. So, we need to subtract the probability that $M \leq 3$:

$$\boxed{P(M = 4) = P(M \leq 4) - P(M \leq 3) = \left(\frac{4}{6}\right)^5 - \left(\frac{3}{6}\right)^5 \approx 0.1}$$

Definition 9.4 (General Max/Min Formulas). Suppose $X_1, \dots X_n$ are iid random variables. Let $Y_1 = \min\{X_1, \dots, X_n\}$ and let $Y_n = \max\{X_1, \dots X_n\}$. Then

$$P(\max\{X_1, \dots, X_n\} \leq x) = P(Y_n \leq x) = (F_X(x))^n$$

$$P(\min\{X_1, \dots, X_n\} > x) = P(Y_1 > x) = (P(X > x))^n$$

If $X_1, \dots, X_n$ are iid discrete random variables, then

$$P(Y_n = x) = P(Y_n \leq x) - P(Y_n \leq x - 1) = (F_X(x))^n - (F_X(x-1))^n$$

$$P(Y_1 = x) = P(Y_1 \geq x) - P(Y_1 \geq x + 1)$$

Example 9.5. Suppose X_1, X_2, X_3, X_4 are iid exponential random variables, each with mean 3. Find the probability that at least one of them exceeds 5.

If one of them exceeds 5, then the maximum must be greater than 5, so it is easier to work with the CDF. Let M denote the maximum of our X_i. Then,

$$P(M > 5) = 1 - P(M \leq 5) = 1 - (F_X(5))^4$$

$$= 1 - \left(1 - e^{-5/3}\right)^4 \approx \boxed{0.567}$$

Example 9.6. Let $N_1, N_2, \dots, N_5$ be 5 iid Poisson random variables with mean 1.2. Find the probability that the maximum of these 5 variables is 2.

$$P(M = 2) = (P(M \leq 2))^5 - (P(M \leq 1))^5$$

Where

$$P(M \leq 2) = e^{-1.2} \cdot \frac{(1.2)^2}{2} + e^{-1.2}(1.2) + e^{-1.2} \approx 0.8795$$

$$P(M \leq 1) = e^{-1.2}(1.2) + e^{-1.2} = 0.6626$$

Therefore,

$$\boxed{P(M = 2) = (0.8795)^5 - (0.6626)^5 \approx 0.3985}$$

Example 9.7 (SOA Practice Exam Q64). Claim amounts for wind damage to insured homes are independent random variables with common density function

$$f(x) = \begin{cases} \frac{3}{x^4} & x > 1 \\ 0 & \text{otherwise} \end{cases}$$

where x is the amount of a claim in thousands. Suppose 3 such claims will be made.

What is the expected value of the largest of the three claims?

Let M be the maximum claim. We want to compute $(P(X \leq M))^3$ using the given density function:

$$(P(X \leq M))^3 = \left(\int_1^M \frac{3}{x^4} dx\right)^3 = \left(\left[-\frac{1}{x^3}\right]_1^M\right)^3 = \left(1 - \frac{1}{M^3}\right)^3$$

The resulting function is our new CDF. To compute the expected value, we want to compute the density of $F(M)$:

$$f(M) = F'(M) = 3\left(1 - \frac{1}{M^3}\right)^2 \left(\frac{3}{M^4}\right)$$

$$= \frac{9}{M^4}\left(1 - \frac{2}{M^3} + \frac{1}{M^6}\right) = \frac{9}{M^4} - \frac{18}{M^7} + \frac{9}{M^{10}}$$

Now we can compute $E[M]$:

$$E[M] = \int_0^\infty Mf(M)dM = \int_1^\infty \left(\frac{9}{M^3} - \frac{18}{M^6} + \frac{9}{M^9}\right) dM$$

$$\left[-\frac{9}{2M^2} + \frac{18}{5M^5} - \frac{9}{8M^8}\right]_1^\infty = \frac{9}{2} - \frac{18}{5} + \frac{9}{8} = 2.025$$

Since M is to be measured in thousands, the expected maximum value of any of the 3 claims is $\boxed{2,025}$.

9.2 General Order Statistics

Suppose we have some data, and want to know whether or not the median m of the sample is ≤ 4.

- Sample: 1.4, 5.3, 3.8: $m = 3.8 \leq 4$
- Sample: 3.7, 2.3, 1.8: $m = 2.3 \leq 4$
- Sample: 4.4, 5.3, 3.8: $m = 4.4 > 4$
- Sample: 4.1, 6.2, 5.5: $m = 5.5 > 4$
- Sample: 2.6, 1.5, 3.2: $m = 2.6 \leq 4$

Note that the median is less than 4 if 2 or 3 of the data points are ≤ 4 and greater than 4 if 0 or 1 of the data points are ≤ 4. This idea will come in handy as we discuss medians of iid random variables.

Example 9.8. Claim amounts for wind damage to insured homes are independent random variables with common density function

$$f(x) = \begin{cases} \frac{3}{x^4} & x > 1 \\ 0 & \text{otherwise} \end{cases}$$

where x is the amount of a claim in thousands. Suppose 3 such claims will be made. Find the density and CDF of the *median* of the 3 claims.

Let Y denote the median. $Y \leq y$ if either 2 or 3 claims are $\leq y$.

$$P(X \leq y) = \int_1^y \frac{3}{x^4} dx = 1 - \frac{1}{y^3}$$

$$\begin{aligned} P(Y \leq y) = P(\text{at least 2 claims } \leq y) &= P(\text{exactly 2 claims } \leq y) + P(\text{all 3 claims } \leq y) \\ &= \binom{3}{2} \left(1 - \frac{1}{y^3}\right)^2 \left(\frac{1}{y^3}\right) + \left(1 - \frac{1}{y^3}\right)^3 \end{aligned}$$

We can factor out $\left(1 - \frac{1}{y^3}\right)^2$:

$$\left(1 - \frac{1}{y^3}\right)^2 \left(\frac{3}{y^3} + 1 - \frac{1}{y^3}\right) = \left(1 - \frac{1}{y^3}\right)^2 \left(1 + \frac{2}{y^3}\right)$$

Before taking the derivative, we will expand the terms:

$$\left(1 - \frac{1}{y^3}\right)^2 \left(1 + \frac{2}{y^3}\right) = \left(1 - \frac{2}{y^3} + \frac{1}{y^6}\right) \left(1 + \frac{2}{y^3}\right) = 1 - \frac{3}{y^6} + \frac{2}{y^9}$$

$$\boxed{F'(y) = f(y) = \frac{18}{y^7} - \frac{18}{y^{10}}}$$

Definition 9.9 (Distribution of Order Statistics). Suppose that we have n data points, denoted as $X_1, X_2, \dots, X_n$, where the X_i are iid random variables. We can sort the n data points from smallest to largest:

$$X_{(1)} < X_{(2)} < X_{(3)} < \cdots < X_{(n)}$$

$$Y_1 < Y_2 < Y_3 < \ldots < Y_n$$

The i-th smallest data point is called the i-th order statistic, and is denoted either as $X_{(i)}$ or Y_i.

So Y_1 is the smallest data value, aka the minimum.

$$P(Y_1 \leq y) = P(\text{at least one } X_i \text{ is } \leq y)$$

$$= 1 - P(\text{all } X_i > y) = 1 - (1 - P(X_q \leq y))^n = 1 - (1 - F_X(y))^n$$

Y_n is the largest data value (aka the max), and

$$P(Y_n \leq y) = P(\text{all } n \text{ points are } \leq y) = [F_X(y)]^n$$

The other order statistics are messier:

$$P(Y_n \leq y) = (F_X(y))^n$$

$$P(Y_{n-1} \leq y) = P(\text{at least } n-1 \text{ are } \leq y)$$

$$= P(\text{exactly } n{-}1 \text{ are } \leq y) + P(\text{all } n \text{ are } \leq y) = \binom{n}{n-1}(F_X(y))^{n-1}(1-F_X(y)) + (F_X(y))^n$$

Similarly,

$$P(Y_{n-2} \leq y) = P(\text{at least } n-1 \text{ are } \leq y)$$

$$= P(\text{exactly } n{-}2 \text{ are } \leq y) + P(\text{all } n{-}1 \text{ are } \leq y) = \binom{n}{n-2}(F_X(y))^{n-2}(1-F_X(y)) + (F_X(y))^2$$

Definition 9.10 (Densities of Order Stats). Let $f_i(y)$ be the density of Y_i. Then

$$f_i(y) = \frac{n!}{(i-1)!(n-i)!}(1 - F_X(y))^{n-i} F_X(y)^{i-1} f_X(y)$$

Example 9.11. Suppose that X_1, X_2, X_3, X_4, X_5 are iid exponential random variables, each with mean 4. Find the density of the median of those variables.

Let Y_3 denote the median order statistic. From the formula in Definition 9.10,

$$f_3(y) = \frac{5!}{2!3!}\left(e^{-y/4}\right)^2 (1 - e^{-y/4})^2 \left(\frac{1}{4}e^{-y/4}\right)$$

$$= \frac{15}{2}e^{-3y/4}(1 - e^{-y/4})^2$$

Example 9.12. Let $W_1, W_2, \ldots, W_4$ be 4 idd Poisson random variables with mean 3.2. Find the probability that the minimum of these 4 variables is 2.

Let m denote the minimum:

$$P(m = 2) = (P(m \geq 2))^4 - (P(m \geq 3))^3$$

$$P(m = 2) = (1 - P(m = 0) - P(m = 1))^4 - P(1 - P(m \leq 2) - P(m = 2))^4$$

We have

$$P(m \leq 2) = 1 - e^{-3.2} - 3.2e^{-3.2} = 0.829$$

$$P(m = 2) = 0.5(3.2)^2 e^{-3.2} \approx 0.209$$

Therefore $\boxed{P(m = 2) = (0.829)^4 - (0.829 - 0.209)^4 \approx 0.324}$

Example 9.13. Let X_1, X_2 and X_3 be independent continuous random variables with the following density function:

$$f(x) = \begin{cases} \sqrt{2} - x & 0 < x < \sqrt{2} \\ 0 & \text{otherwise} \end{cases}$$

What is the probability that exactly 2 of the 3 random variables exceed 1?

For any one of the variables,

$$P(X \leq 1) = \int_0^1 (\sqrt{2} - x)dx = \sqrt{2} - \frac{1}{2}$$

To have exactly 2 of the 3 variables exceed 1, we also need exactly 1 less than 1, so

$$P(\text{Exactly } 2\ X_i > 1) = \binom{3}{1} P(X \leq 1)(P(X > 1))^2 = 3\left(\sqrt{2} - \frac{1}{2}\right)\left(1 - \left(\sqrt{2} - \frac{1}{2}\right)\right)^2$$

$$= \boxed{3\left(\sqrt{2} - \frac{1}{2}\right)\left(\frac{3}{2} - \sqrt{2}\right)^2}$$

Example 9.14. Let $Y_1 < Y_2 < \cdots < Y_5$ be the order statistics of a random sample of size 5 from a continuous distribution with median m. What is $P(Y_2 < m < Y_4)$?

In order to have $Y_2 < m < Y_4$, we need to have either exactly 2 of the data points be less than m, or exactly 3. The probability $P(x < m) = \frac{1}{2}$, so that gives us

$$P(Y_2 < m < Y_4) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 + \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \boxed{\frac{1}{32}(10 + 10) = \frac{5}{8}}$$

If there were only 4 sample points for instance, exactly 2 of the points need to be less than m, which means

$$P(Y_2 < m < Y_3) = \binom{4}{2}\left(\frac{1}{2}\right)^4 = \frac{3}{8}$$

Example 9.15. I roll a fair die 3 times. Let X be the maximum value of the rolls, and Y the minimum value. What is $P(X = 5, Y = 2)$?

There are 4 cases:

- Two rolls equal 5, one roll equals 2

$$P(\text{Case 1}) = \frac{3!}{1!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right) = 3\left(\frac{1}{6}\right)^3$$

- One roll equals 5, one roll equals 4, one roll equals 2

$$P(\text{Case 2}) = \frac{3!}{1!1!1!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right) = 3\left(\frac{1}{6}\right)^3$$

- One roll equals 5, one roll equals 3, one roll equals 2. This has the same probability as Case 2.

- One roll equals 5, two rolls equal 2

$$P(\text{Case 4}) = \frac{3!}{1!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right) = 3\left(\frac{1}{6}\right)^3$$

The total probability is obtained by summing up each case:

$$\boxed{(X = 5, Y = 2) = (3 + 6 + 6 + 3)\left(\frac{1}{6}\right)^3 = \frac{1}{12}}$$

10 Important Formulas and Theorems

10.1 Discrete Probability

Fundamentals of Probability

Theorem 10.1. *If A is a list of events and S is the sample space:*

$$0 \leq P(A) \leq 1 \qquad P(S) = 1$$

If $A_1 \cap A_2 = \emptyset$ then

$$P(A_1 \cup A_2) = P(A_1 + A_2)$$

Definition 10.2 (Inclusion-Exclusion Principle).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Definition 10.3 (Complements).

$$A' = \text{ everything in sample space } S \text{ but not in } A$$

$$A \cap A' = \emptyset \quad A \cup A' = S$$

Theorem 10.4 (Probabilities with Complements). *Let A, B, C be events, then*

1. $P(A') = 1 - P(A)$
2. $(P(A'))' = P(A)$
3. $P(A \cap B) + P(A' \cap B) = P(B) \iff P(A \cap B) = P(B) - P(A' \cap B)$
4. $P(A \cap B' \cap C') + P(A \cap B \cap C) = P(A)$

Definition 10.5 (Mutually Exclusive). Two events A and B are mutually exclusive if

$$P(A \cap B) = \emptyset$$

Theorem 10.6 (DeMorgans Laws). *Let $A_1, \ldots, A_k$ be events. Then,*

$$\left[\bigcup_{i=1}^{k} A_i\right]' = \bigcap_{i=1}^{k} A_i' \qquad \left[\bigcap_{i=1}^{k} A_i\right]' = \bigcup_{i=1}^{k} A_i'$$

Conditional Probability

Definition 10.7 (Conditional Probability). Let A and B be two events. The **conditional probability** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}.$$

Definition 10.8 (Independence). Events A and B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B)$$

Theorem 10.9 (Law of Total Probability). *If $A_1, A_2, \ldots, A_k$ are disjoint and $P(A_1) + P(A_2) + \cdots + P(A_k) = 1$ then*

- $P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k)$
- $P(B) = P(A_1)P(B|A_1) + \cdots + P(A_k)P(B|A_k)$

Theorem 10.10 (Bayes' Theorem). *Suppose $A_1, \ldots, A_k$ are a partition of the sample space. Then*

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{k} P(B \cap A_i)} = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}$$

The final denominator sums one event over all cases.

Discrete Moments

Definition 10.11. For a discrete random variable X, y is the **mode** of X if $P(X = y) \geq P(X = x)$ for all x (i.e. the mode y is the input that maximizes $P(X = y)$).

Definition 10.12. The **median** of a random variable X is the smallest m such that $P(X \leq m) = F(m) \geq \frac{1}{2}$.

Definition 10.13 (Percentile). The $100\% \cdot p^{th}$ percentile π_p is the smallest possible x such that $P(X \leq x) \geq p$.

Definition 10.14 (Expected Value). If X is a discrete random variable, then

$$E[X] = \sum_x x \cdot P(X = x)$$

$$E[g(X)] = \sum_x g(x) \cdot P(X = x)$$

Theorem 10.15 (Transformations on Variance). *Let X be a random variable, $a, b \in \mathbb{R}$ (constants). Then,*

1. $Var(aX) = a^2 Var(X)$
2. $Var(X + b) = Var(X)$
3. $Var(aX + b) = a^2 Var(X)$

Definition 10.16 (Std. Deviation and Coefficient of Variance). Let X be a random variable with variance $\text{Var}(x)$. Then

$$\text{SD}(X) = \sigma_X = \sqrt{\text{Var}(X)} \qquad \text{CV}(X) = \frac{\sigma}{\mu} = \frac{\text{SD}(X)}{\text{E}[X]}$$

If $c \in \mathbb{R}$ is a constant,

$$\text{SD}(cX) = |c|\text{Var}(X) \qquad \text{CV}(cX) = \text{CV}(X)$$

Combinations and Permutations

Definition 10.17 (Combinations and Permutations). If there are n distinct items and want to select a group of k items, the number of **combinations** can be written as

$${}_nC_k \text{ or } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If an ordering is involved, the number of permutations are

$${}_nP_k = \frac{n!}{(n-k)!}$$

Common Distributions

Definition 10.18 (Bernoulli Random Variables). A **Bernoulli**(p) random variable, or a Bernoulli 0-1 random variable, is a variable that can only be 0 or 1. Usually 1 is considered a success. If p is the probability of success, then

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

The mean and variance follow:

$$E[X] = p \qquad \text{Var}(X) = p(1 - p)$$

Definition 10.19 (Binomial Random Variables). X is a **binomial** (n, p) random variable if X is the number of successes in n independent trials, each of which is a success with the same probability p.

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

$$E[X] = np \qquad \text{Var}(X) = np(1 - p)$$

Theorem 10.20 (Binomial Expansion). *For any real numbers a, b and positive integer n:*

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Definition 10.21 (Multinomial Distribution). Suppose there are n independent trials, each with the same r possible outcomes. Let $p_1, p_2, \ldots p_r$ be the probabilities of the outcomes, and X_i the number of trials resulting in the i-th outcome. Then,

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r) = \frac{n!}{k_1!k_2! \cdots k_n!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

If X_i, X_j are trials whose respective probabilities of success are p_i and p_j, then

$$E[X_i] = np_i \qquad \text{Var}(X_i) = np_i(1 - p_i) \qquad \text{Cov}(X_i, X_j) = -np_i p_j$$

Definition 10.22 (Hypergeometric Distribution). Say we have N trials/objects with m successes. If you randomly select n of them without replacement, then $X \sim \text{Hyp}(n, N, m)$ is **hypergeometric** and has distribution

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

for $k = 0, 1, \ldots, \min(m, n)$. If X follows a hypergeometric distribution,

$$E[X] = \frac{mn}{N} \qquad \text{Var}(X) = \frac{mn(N-n)(N-m)}{N^2(N-1)}$$

Key Discrete Distributions

Theorem 10.23 (Geometric Series Convergence). *Let $|r| < 1$ and a be a real number. Then,*

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

Theorem 10.24 (Geometric Series starting at 1). *Suppose X is a geometric random variable on $\{1, 2, \ldots\}$ with parameter p if X is the number of trials up to, and including, the first success. Then,*

$$E[X] = \frac{1}{p} \qquad \text{Var}(X) = \frac{1-p}{p^2}$$

Theorem 10.25 (Geometric Series starting at 0). *Suppose Y is a geometric random variable on 0, 1, 2, ... if Y counts the number of failures before the first success. Then*

$$E[Y] = E[X] - 1 = \frac{1}{p} - 1 \qquad \text{Var}(Y) = \frac{1-p}{p^2}$$

Theorem 10.26 (Memoryless Property). *If N follows a discrete geometric distribution with parameter p, then $(N - k \mid N > k)$ is a geometric distribution starting at 1 with the same p. This holds whether N starts at 0 or 1.*

Definition 10.27 (Negative Binomial Distribution). Suppose N is a negative binomial random variable with parameters r and p if it is the sum of r independent geometric random variables starting at 0. It is the number of failures before the r-th success.

$$P(N = n) = \binom{n + (r-1)}{n} p^r (1-p)^n$$

$$E[N] = r\left(\frac{1}{p} - 1\right) \qquad \mathrm{Var}(N) = \frac{r(1-p)}{p^2}$$

Definition 10.28. X is a Poisson(λ) random variable if

$$P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!} \quad \text{for } n = 0, 1, 2, \ldots$$

$$E[N] = \mathrm{Var}(N) = \lambda$$

Theorem 10.29 (Sums of Poisson Variables). *If $N \sim Pois(\lambda), M \sim Pois(\mu)$ and they are independent, then*

$$P(N + M = n) = e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^n}{n!}$$

Deductibles and Limits

Definition 10.30 (Payment, Uncovered Cost, Total Loss). Suppose X represents the amount of a loss. If there is a deductible of d, then the resulting **(insurance) payment** is

$$\text{Payment} = (X - d)_+ = \begin{cases} 0, & X \leq d \\ X - d, & X > d \end{cases}$$

The **uncovered cost** to the insured, or the expense not protected/paid for by insurance policy is

$$\text{Uncovered Cost} = \min\{X, d\} = X \wedge d = \begin{cases} X, & X \leq d \\ d, & X > d \end{cases}$$

Lastly, the **total loss** is the sum of the insurance payment and uncovered cost:

$$X = (X - d)_+ + (X \wedge d)$$

Note that $\min\{X, d\}$ and $X \wedge d$ are equivalent notation-wise.

Theorem 10.31 (Expected Payment). *Suppose X represents the amount of a loss. Then,*

$$E[(X-d)_+] = E[X] - E[X \wedge d]$$

Definition 10.32 (Policy Limit). Let X be the loss amount, and u the policy limit. With no deductible,

$$\text{Payment} = \begin{cases} X, & X \leq u \\ u, & u < X \end{cases}$$

In this case, Payment $= \min\{X, u\} = X \wedge u$.

10.2 Continuous Probability

Continuous Distributions I

Definition 10.33. The **cumulative distribution function (CDF)** of X is given by

$$F(x) = F_X(x) = P(X \leq x).$$

If F_x is differentiable, its derivative

$$f_X(t) = F'(x)$$

is referred to as the **density** of X. By the Fundamental Theorem of Calculus, the CDF is then

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

Definition 10.34 (Percentiles and Medians). x is a k-th **percentile** of X if $F(x) = k\%$. The **median** is the 50th percentile, so $F(x) = 0.5$ at the median.

Definition 10.35 (Mean/Variance of Continuous Random Variables). If X is random variable whose density function $f(x)$ is purely continuous, then

$$E[X] = \int_x xf(x)dx$$

Once again, if g is a function of X, then

$$E[g(X)] = \int_x g(x)f(x)dx$$

The discrete formula for variance also applies to continuous functions.

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \int_x x^2 f(x)dx - \left(\int_x xf(x)dx\right)^2$$

Theorem 10.36 (Mean of a CDF using the Survival Method). *Suppose that $P(X \geq 0) = 1$ and X is continuous. Then*

$$E[X] = \int_0^\infty P(X > x)dx$$

Key Continuous Distributions

Theorem 10.37 (Mean and Variance of Uniform Distributions). *Let $X \sim$ Uniform(a, b). Then*

$$E[X] = \frac{a+b}{2} \qquad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Definition 10.38 (Density and CDF of Exponential Distributions). X is an **exponential random variable** with mean θ if

$$F_X(x) = 1 - e^{-\frac{x}{\theta}} \qquad 1 - F(x) = e^{-\frac{x}{\theta}}$$

Sometimes $\lambda = \frac{1}{\theta}$ will be called a rate instead of an exponential.

$$F_x(x) = 1 - e^{-\lambda x}$$

$$f(x) = F'(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} = \lambda e^{-\lambda x} \quad \text{for } x > 0$$

$$E[X] = \theta \qquad \text{Var}(X) = \theta^2$$

Definition 10.39 (Gamma Distribution CDFs). At $\alpha = 1$ we have the exponential CDF.

$$\alpha = 1: \qquad F(x) = 1 - e^{-x/\theta}$$

$$\alpha = 2: \qquad F(x) = 1 - e^{-x/\theta} - \frac{x}{\theta}e^{-x/\theta}$$

$$\alpha = 3: \qquad F(x) = 1 - e^{-x/\theta} - \frac{x}{\theta}e^{-x/\theta} - \left(\frac{x}{\theta}\right)^2 \cdot \frac{1}{2}e^{-x/\theta}$$

Definition 10.40 (Gamma Distribution Densities).

$$f(x) = \frac{1}{(\alpha-1)!} \cdot \frac{x^{\alpha-1}}{\theta^\alpha}e^{-x/\theta}$$

Theorem 10.41 (Mean and Variance of Gamma Distributions). *Let $X \sim Exp(\theta)$ and $Y \sim Gamma(\alpha, \theta)$. If α is an integer, then Y is a sum of α iid $Exp(\theta)$ variables.*

$$E[Y] = \alpha E[X] = \alpha\theta \qquad Var(Y) = \alpha Var(X) = \alpha\theta^2$$

Definition 10.42 (Beta Distributions). X is Beta(a, b) if $f(x) = cx^{a-1}(1-x)^{b-1}$ for $0 < x < 1$, and 0 otherwise,

$$\text{where } c = \frac{(a+b-1)!}{(a-1)!(b-1)!}$$

Moreover,

$$E[X] = \frac{a}{a+b} \qquad E[X^2] = \frac{a(a+1)}{(a+b)(a+b+1)}$$

Normal Approximations

Definition 10.43 (Normal Distributions). A standard normal distribution has the density

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

A standard normal Z has mean $\mu = 0$ and variance $\sigma = 1$.

Theorem 10.44 (Densities of Normal Variables). *If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then*

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}$$

Definition 10.45 (Normal CDFs). Suppose that Z is a standard normal ($Z \sim \mathcal{N}(0,1)$). Then

$$\Phi(z) = P(Z \leq z)$$

denotes the CDF of Z.

Theorem 10.46 (Sums of Normal Distributions). *f X and Y are independent normal distributions, then $X + Y$ is also a normal distribution.*

$$E[X+Y] = E[X] + E[Y] \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Theorem 10.47 (Central Limit Theorem (CLT)). *If $X_1, \ldots, X_n$ are identically independently distributed random variables, then*

$$\frac{(X_1 + \cdots + X_n) - nE[X_1]}{\sqrt{n\mathrm{Var}(X_1)}} \sim \mathcal{N}(0,1)$$

Definition 10.48 (Lognormal Random Variables). Y is a **lognormal** if $Y = e^X$, X is normal. In words, the log of a lognormal distribution is normal.

Theorem 10.49 (Lognormal Moments). *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = e^X$ is the corresponding lognormal. Then*

$$E[Y] = e^{\mu+\sigma^2/2} \qquad E[Y^2] = e^{2\mu+2\sigma^2} \qquad E[Y^n] = e^{n\mu+(n\sigma)^2/2}$$

10.3 Multivariate Probability

Joint Distributions and Moments

Definition 10.50 (Joint PMF). Suppose we have 2 random variables, X and Y. The **joint probability mass function** is $P(X = x, Y = y)$. The total probability is still 1:

$$\sum_x \sum_y P(X = x, Y = y) = 1$$

Theorem 10.51 (Properties of Joint CDFs).

1. $0 \leq F(x, y) \leq 1$
2. $F(x, \infty) = P(X \leq x, Y < \infty) = P(X \leq x) = F_X(x)$
3. $F(\infty, y) = P(X < \infty, Y \leq y) = P(Y \leq y) = F_y(y)$
4. $F(\infty, \infty) = 1$
5. $F(-\infty, y) = 0 = F(x, -\infty)$

Definition 10.52 (Marginal Distributions). For discrete variables,

$$P(X = x) = \sum_y P(X = x, Y = y)$$

is the **marginal distribution** of X. In words, it's the distribution of X without knowing Y.

Definition 10.53 (Conditional Distributions). For discrete random variables X, Y,

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x, Y = y)}{\sum_x P(X = x, Y = y)}$$

The denominator makes the conditional distribution itself a probability distribution (because the sum of marginal distributions add up to 1!)

Definition 10.54 (Multivariate Independence). X and Y are **independent** if $P(X = x \mid Y = y) = P(X = x)$. For discrete variables,

$$P(X = x, Y = y) = P(Y = y) \cdot P(X = x) \quad \text{if } X \text{ and } Y \text{ are independent}$$

Definition 10.55 (Multivariate Mean). For discrete random variables X and Y,

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot P(X = x, Y = y)$$

Definition 10.56 (Covariance). Suppose X, Y are two random variables. Then

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Corollary 10.57 (Bilinearity of Covariances).

$$Cov(aX + bY, cW + dZ) = Cov(aX, cW) + Cov(aX, dZ) + Cov(bY, cW) + Cov(bY, dZ)$$

$$= acCov(X,W) + adCov(X,Z) + bcCov(Y,W) + bdCov(Y,Z)$$

Corollary 10.58 (Variance of a Sum).

$$Var(aX + bY) = a^2 Var(X) + 2abCov(X,Y) + b^2 Var(Y)$$

$$Var(aX - bY) = a^2 Var(X) - 2abCov(X,Y) + b^2 Var(Y)$$

If X and Y are independent,

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) = Var(aX - bY)$$

Definition 10.59 (Correlation). For two random variables X and Y, the **correlation** between them is given by

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)}$$

Corollary 10.60. *If X and Y are independent, then*

$$E[XY] = E[X]E[Y]$$

More generally,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Definition 10.61 (Conditional Moments). Let X and Y be two (discrete) random variables. Then,

$$E(X \mid Y = y) = \sum_x x \cdot P(X = x \mid Y = y)$$

Theorem 10.62 (Double Expectation Theorem). *If X and Y are (discrete) random variables,*

$$E[X] = E[E[X \mid Y = y]]$$

$$E[X] = \sum_{all\ y} E[X \mid Y = y] \cdot P(Y = y)$$

Theorem 10.63 (Law of Total Variation). *Let X, Y be random variables. Then,*

$$Var(X) = E[Var(X \mid Y)] + Var(E(X \mid Y))$$

Theorem 10.64 (Variance of Random Sums). *Suppose $S = X_1 + \cdots + X_N$, where $X_1, X_2, \cdots,$ are iid and N is an independent integer valued random variable. Then*

$$Var(S) = E[N]Var(X) + Var(N)(E[X])^2$$

Order Statistics

Definition 10.65 (General Max/Min Formulas). Suppose $X_1, \ldots X_n$ are iid random variables. Let $Y_1 = \min\{X_1, \ldots, X_n\}$ and let $Y_n = \max\{X_1, \ldots X_n\}$. Then

$$P(\max\{X_1, \ldots, X_n\} \le x) = P(Y_n \le x) = (F_X(x))^n$$

$$P(\min\{X_1, \ldots, X_n\} > x) = P(Y_1 > x) = (P(X > x))^n$$

If $X_1, \ldots, X_n$ are iid discrete random variables, then

$$P(Y_n = x) = P(Y_n \le x) - P(Y_n \le x - 1) = (F_X(x))^n - (F_X(x-1))^n$$

$$P(Y_1 = x) = P(Y_1 \ge x) - P(Y_1 \ge x + 1)$$

Definition 10.66 (Distribution of Order Statistics). Suppose that we have n data points, denoted as $X_1, X_2, \ldots, X_n$, where the X_i are iid random variables. We can sort the n data points from smallest to largest:

$$X_{(1)} < X_{(2)} < X_{(3)} < \cdots < X_{(n)}$$

$$Y_1 < Y_2 < Y_3 < \ldots < Y_n$$

The i-th smallest data point is called the i-th order statistic, and is denoted either as $X_{(i)}$ or Y_i.

Definition 10.67 (Densities of Order Stats). Let $f_i(y)$ be the density of Y_i. Then

$$f_i(y) = \frac{n!}{(i-1)!(n-i)!}(1 - F_X(y))^{n-i} F_X(y)^{i-1} f_X(y)$$

11 References

All diagrams are self-curated. However, a lot of the material and diagrams were heavily inspired from `https://www.theinfiniteactuary.com/`