Probability Primer

1. Introduction

The purpose of this primer is to introduce the reader to the basic theory of probability through commonly played games of chance. By the end of the primer, the reader will be introduced to the concept of a random experiment and its associated sample space, which consists of the set of possible outcomes produced by the experiment. Events represent subsets of the sample space. We assign probabilities to these events to represent how likely they occur. The reader will also be introduced to the concept of a random variable, which maps outcomes in the sample space to real numbers. A random variable has an associated probability distribution. We will show how to build a few popular probability distributions from games of chance. Finally, we conclude by introducing some basic concepts of reliability through an example using the exponential distribution.

2. Sample Space

Consider two games of chance: a coin toss and a roll of a die. These games of chance represent random experiments as they can result in different outcomes, even if repeated in the same manner each time. Each random experiment has an associated sample space. The sample space is the set of all possible outcomes of a random experiment, and is often denoted by the S. The notation for the sample space consists of the possible outcomes listed inside curly brackets.

Example 2.1
In a coin toss, the coin can either land on heads or tails. Thus, the sample space for a coin toss is:

$$\mathrm{S}_{coin\ toss} = \{heads, tails\}$$

For a die roll, the die can either land on one of six faces: 1, 2, 3, 4, 5, or 6. Thus, the state space for a die roll is:

$$\mathrm{S}_{die\ roll} = \{1, 2, 3, 4, 5, 6\}$$

The coin toss and die roll are examples of discrete sample spaces. Discrete sample spaces consist of a finite or countably infinite set of outcomes. Coin tosses and die rolls have a finite set of outcomes, while the number of failures during a time interval is countably infinite. Continuous sample spaces consist of outcomes that lie on continuum. If we were to record the mass of a manufactured component, the sample space for the component mass would be the interval of nonnegative real numbers.

3. Events

In probability theory, we assign values to events, which are subsets of the sample space. For a die roll, an example of an event can be rolling a number greater than 4. Suppose for a die roll, we define A as the event that the die lands on 1 and B as the event that the die lands on 6. Then A and B are mutually exclusive**,** as they do not share any outcomes and cannot occur simultaneously. However, if A denotes the event that the die lands on an even number and B denotes the event that the die lands on a number less than 3, then A and B are not mutually exclusive as they both include the number 2. Let A and B be two events, then the following operations are applicable:

- Union ($A \cup B$): All outcomes contained in either A or B. If A is the event that the die lands on an even number and B is the event that the die lands on a face less than 3, then $A \cup B = \{1, 2, 4, 6\}$
- Intersection ($A \cap B$): All outcomes contained in both A and B. Using the previous example, $A \cap B = \{2\}$
- Complement (A'): All outcomes in the sample space not included in the event A. If A is the event that you roll an even number, $A' = \{1, 3, 5\}$

If two or more events share no outcomes, then their intersection is the empty set $\emptyset$. Thus, the events are mutually exclusive if their intersection is $\emptyset$.

4. Probability

Probability refers to the chance that an outcome occurs. A key assumption in games of chance is that the games are fair and thus, the outcomes are equally likely. If a random experiment has N equally likely outcomes, then the probability of observing each outcome is $\frac{1}{N}$.

> Example 4.1
> For a fair coin toss, there are $N = 2$ possible outcomes: heads or tails. Thus, the probability of the coin landing on heads and the probability of the coin landing on tails are both $\frac{1}{2}$.
> For a roll of a fair six-sided die, there are $N = 6$ possible outcomes. Thus the probability of the die landing on either of the six faces is $\frac{1}{6}$.

For a discrete sample space, we denote the probability of an event as $P(E)$. This probability equals the sum of the probabilities of the outcomes in the event.

> Example 4.2
> Let E be the event that a fair die lands on an even number. The outcomes in E are $\{2, 4, 6\}$. Thus the probability of the die landing on an even number is
>
> $$P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$
>
> Let F be the event that the die lands on a number less than 3. The probability of the union of E and F is
>
> $$P(E \cup F) = P(1) + P(2) + P(4) + P(6) = \frac{4}{6} = \frac{2}{3}$$
>
> The probability of the intersection of E and F is
>
> $$P(E \cap F) = P(2) = \frac{1}{6}$$
>
> The complement of E is the event that the die does not land on an even number.
>
> $$P(E') = P(1) + P(3) + P(5) = \frac{3}{6} = \frac{1}{2}$$

Whether the sample space is continuous or discrete, probability follows three axioms:
Axioms of Probability

1) $P(S) = 1$
2) $0 \leq P(E) \leq 1$
3) The probability of the union of two mutually exclusive events is the sum of the events' respective probabilities. In other words:
$P(E_1 \cup E_2) = P(E_1) + P(E_2)$ when $E_1 \cap E_2 = \emptyset$

Axiom 1 essentially states that some outcome from the sample space must occur. Some consequences of Axiom 1 include:

- $P(\emptyset) = 0$
- $P(E') = 1 - P(E)$

The second axiom states that probabilities range between 0 and 1 where 0 represents impossible and 1 represents certainty. The final axiom provides a way to calculate probabilities of mutually exclusive events. If the events are not mutually exclusive, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Furthermore, if E_1 is contained in E_2: $(E_1 \subseteq E_2)$,

$$P(E_1) \leq P(E_2)$$

5. Conditional Probability and Independence

The conditional probability of event B given A, $P(B|A)$ is the probability of an outcome from event B occurring given that we know an outcome from event A has occurred. When calculating the conditional probability, the sample space reduces to the event A. We then determine the proportion of event A that also contains event B. Mathematically, this is shown using the following formula.

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

Using the above formula, the probability of A given B is:

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

After multiplying both sides by $P(B)$,

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

Thus, we can derive Bayes' Theorem.

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

We now show how to calculate the conditional probability using intuition, the formula for conditional probability, and Bayes' Theorem.

Example 5.1

Suppose you are at a casino playing at the roulette wheel. A roulette wheel has 48 red slots, 48 black slots, and 2 neutral slots. Assume the roulette wheel lands on any slot with equal probability. If you place a bet that the wheel lands on slot Red 1, then you have a $\frac{1}{100}$ chance at winning. Now suppose you are superstitious and you close your eyes while the wheel is coming to a stop. Before checking the result, you are told the wheel stopped on a red slot. Given this information, what is the probability that you won?

Intuition

If it is known that the wheel stopped on a red slot, the number of possible outcomes reduces to 48 and now your chance of winning increases to $\frac{1}{48}$.

Formula for Conditional Probability

Let A denote the event that the roulette wheel lands on a red slot and let B denote the event that the wheel lands on Red 1. The intersection of these two events consists of only one outcome: the Red 1 slot. Thus, $P(A \cap B) = \frac{1}{100}$. The probability of the wheel landing on a red slot is $\frac{48}{100}$. Using the above formula:

$$P(B|A) = \frac{\frac{1}{100}}{\frac{48}{100}} = \frac{1}{48}$$

Bayes' Theorem

$P(B) = \frac{1}{100}$ and $P(A) = \frac{48}{100}$. Given that we know the wheel lands on Red 1, with certainty the wheel will land on a red slot. Thus, $P(A|B) = 1$. From Bayes' Theorem:

$$P(B|A) = 1\left(\frac{\frac{1}{100}}{\frac{48}{100}}\right) = \frac{1}{48}$$

In the case that $P(A|B) = P(A)$, we notice that $P(A \cap B) = P(A)P(B)$ and $P(B|A) = \frac{P(A)P(B)}{P(A)} = P(B)$. If this is the case, then we say events A and B are independent. Independence plays a key role in statistics as a common assumption is that the distributions of multiple observations of a random experiment are independent of each other.

6. Random Variables

The term random variable is a bit of a misnomer. It is not a variable in the traditional algebraic sense (for example, we do not attempt to solve for a missing random variable) and it is not random. A random variable maps outcomes of the sample space of a random experiment to numbers. As an example, consider the random experiment of a coin toss. The outcomes of the toss are either heads or tails. We can design a random number C to output 1 if the coin lands on heads and 0 if the coin lands on tails. As another example, consider the die roll random experiment. Each face of the die has a certain number of dots on it, which we count to determine the value of the roll. Let D denote the number of dots on the face the die lands. Then D is a random number mapping the outcome of the die roll to the number of dots on the die's upper face. Now suppose we roll two dice. Let D_1 denote the number of dots on the face die 1 lands and D_2 denote the number of dots on the face die 2 lands. We can define a random number as a function of other random numbers. For example, let $Y = D_1 + D_2$. Then Y maps the outcome of the dice roll to the sum of the number of dots between the two dice. These are all examples of discrete random variables as their values are countable and finite. Countably infinite random variables such as the number of failures in an interval of time are also discrete. Continuous random variables map outcomes to numbers in a continuum or interval. For example, assuming the manufacture of a part is repeatable; the mass of a part is a continuous random number. Typically, capital letters denote random numbers while lower case letters denote observed realizations of random numbers. Hence the notation $P(X = x)$ means the probability that random number X outputs value x.

7. Discrete Probability Distributions

For a discrete random variable X, the probability distribution is a description of the probability associated with possible values of X: $x_1, x_2, \dots, x_n$. The probability distribution for a discrete random variable is specified by its probability mass function (PMF). The PMF is a function such that:

1) $f(x_i) \geq 0, i = 1, \dots, n$
2) $\sum_{i=1}^{n} f(x_i) = 1$
3) $f(x_i) = P(X = x_i), i = 1, \dots, n$

Using the PMF, we can define the mean $E(X)$ and variance $V(X)$ of random variable X.

$$\mu = E[X] = \sum_{i=1}^{n} x_i f(x_i)$$

$$\sigma^2 = E[(X - \mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i)$$

We can now develop PMFs for the random variables described earlier.

Example 7.1

Coin Toss

Let random variable $X = \begin{cases} 1, & if\ coin\ lands\ heads \\ 0, & if\ coin\ lands\ tails \end{cases}$

Define $p = P(X = 1)$. Since a coin can either land on heads or tails, $P(X = 0) = 1 - p$. Thus, the PMF for a coin toss is:

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

This PMF can be written more succinctly as follows:

$$f(x) = p^x (1 - p)^{1-x}, x = 0,1$$

The mean and variance of random variable X are:

$$\mu = 1p + 0(1 - p) = p$$

$$\sigma^2 = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p)$$

This is the Bernoulli distribution, which is a probability model for random experiments with two possible outcomes. Random experiments such as these are called Bernoulli trials and form the foundation for several other discrete distributions.

Example 7.2

Die Roll

Let X be a random variable equaling the number of dots on the top face of a rolled die. For a six-sided die, X can be any integer ranging from 1 to 6. Assuming a fair die, all six outcomes are equally likely. Thus, we say that X has a discrete uniform distribution. The discrete uniform distribution assigns equal probability to all possible values of random variable X and is defined as follows:

$$f(x_i) = \frac{1}{N}, i = 1, \dots, N$$

Figure 1 provides a graphical representation of the distribution:

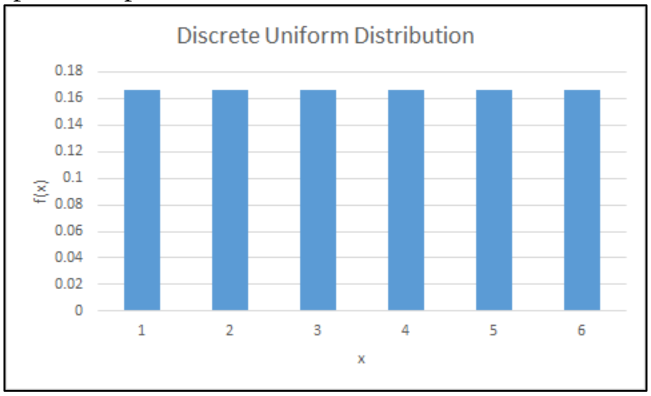

Figure 1: Discrete Uniform Distribution

If X ranges from a to b, the mean and variance for the discrete uniform distribution are:

$$\mu = \frac{(b+a)}{2}$$

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

In this example, $a = 1$ and $b = 6$. Thus, $\mu = \frac{7}{2}$ and $\sigma^2 = \frac{35}{12}$

Example 7.3

Sum of Two Rolled Dice

Let X equal the sum of the values of two dice rolled simultaneously. We would like to build the PMF for X. Let us start by displaying a chart with the column and row headers denoting the outcomes of each die and the cell elements containing their sum. This is shown in Table 1.

Table 1: Outcomes for Rolling Two Dice

	Die 2					
Die 1	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 1 demonstrates that there are 36 outcomes from rolling two dice: $\{(1,1),(1,2),(2,1),\ldots,(6,6)\}$. Assuming all outcomes are equally likely, the probability of each outcome is $\frac{1}{36}$. However, we are interested in PMF for the sum of two dice. This can be done as follows: One outcome results in a sum of 2. Thus, the probability of rolling a 2 is $\frac{1}{36}$. Two outcomes result in a sum of 3, so the probability of rolling a 3 is $\frac{2}{36}=\frac{1}{18}$. Continuing in this manner, we can build the probability distribution as shown in Figure 2.

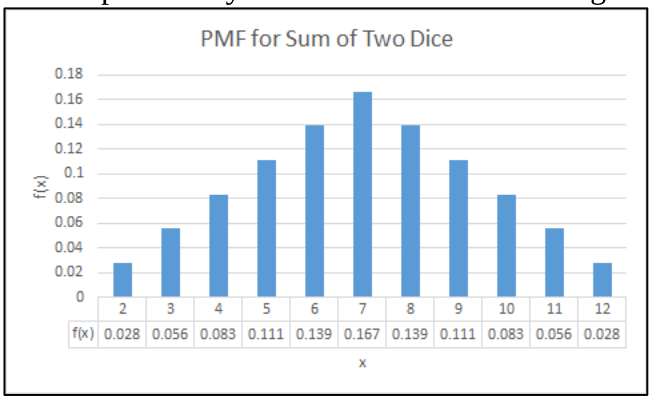

Figure 2: PMF for Sum of Two Dice

The mean and variance are calculated as follows:

$$\mu = \sum_{x=2}^{12} x f(x) = 7$$

$$\sigma^2 = \sum_{x=2}^{12} (x-7)^2 f(x) = 5.833$$

The main takeaway from these examples is that we can build empirical PMFs by counting the number of outcomes that satisfy an event and assigning probability equal to the proportion of the sample space comprised of those outcomes. Furthermore, these PMFs have patterns and characteristics that may match defined probability distributions such as the Bernoulli distribution and the Discrete Uniform distribution.

8. Discrete Cumulative Distribution Function

Suppose we would like to calculate the probability that a discrete random variable X takes on a value within some interval. This can be done by adding the probabilities of all outcomes in the interval. However, we can do this more efficiently using the cumulative distribution function (CDF). The CDF of a discrete random variable X, $F(x)$ is:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x)$$

The CDF has the following properties:

- $0 \leq F(x) \leq 1$
- if $x \leq y$, then $F(x) \leq F(y)$.

Example 8.1
Let X be the random variable denote the sum of two rolled dice. S suppose we want to calculate the probability that the sum of the two dice is between 7 and 10, $P(7 \leq X \leq 10)$. The full CDF for X, the random variable equaling the sum of two dice, is shown in Figure 3.

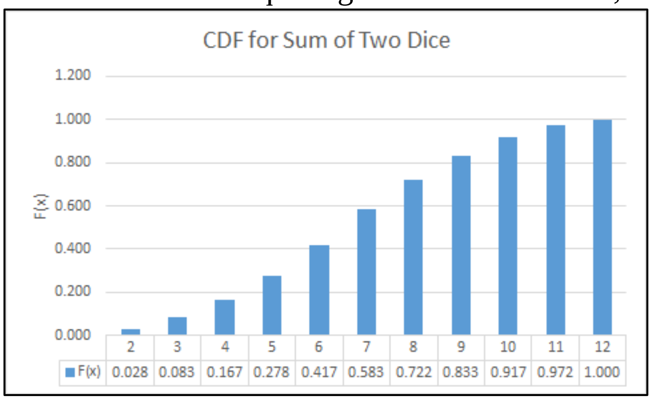

Figure 3: CDF for Sum of Two Dice

Now if we want to calculate $P(7 \leq X \leq 10)$, we only need to subtract $F(6)$ from $F(10)$. Thus, $P(7 \leq X \leq 10) = F(10) - F(6) = 0.917 - 0.417 = 0.5$.

The CDF of a discrete random variable simplifies the calculation of some commonly asked probabilities. For example:

- $P(X \leq x) = F(x)$
- $P(X = x) = F(x) - F(x-1)$
- $P(X > x) = 1 - F(x)$
- $P(X \geq x) = 1 - F(x-1)$
- $P(X < x) = F(x-1)$
- $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1 - 1)$
- $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- $P(x_1 \leq X < x_2) = F(x_2 - 1) - F(x_1 - 1)$
- $P(x_1 < X < x_2) = F(x_2 - 1) - F(x_1)$

9. Continuous Probability Distributions

The analog to the discrete PMF is the probability density function (PDF) for continuous random variables. The PDF is a function such that:

1) $f(x) \geq 0$
2) $\int_{-\infty}^{\infty} f(x)dx = 1$

3) $P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx$ = area under $f(x)$ from x_1 to x_2

The continuous analogs to the mean and variance are as follows:

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

We can also define the CDF for continuous random variables as follows:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du\,, -\infty < x < \infty$$

The largest differences between continuous and discrete probability distributions are that when X is a continuous random variable, $P(X = x) = 0$ and $P(x_1 \le X \le x_2) = P(x_1 < X < x_2)$. In other words, positive probabilities are only defined for intervals of the sample space and there is no distinction between the "less than" and "less than or equal" operators.

10. Exponential Distribution and Reliability

For this primer, we introduce the exponential distribution as it is commonly used in reliability theory and other fields. An exponentially distributed random variable has the following PDF:

$$f(x) = \lambda e^{-\lambda x}, 0 \le x < \infty$$

The CDF of the exponential distribution can be derived by integrating the PDF from 0 to x

$$F(x) = P(X \le x) = \int_{0}^{x} \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}$$

Example 10.1

Suppose we perform an experiment where we run identical parts to failure. The results of the experiment indicate that the time (in hours) until a part failure is exponentially distributed with $\lambda = \frac{1}{8750}$. What is the probability that a part fails in less than 1095 hours? We can solve this using the CDF. If random variable T is the time until failure,

$$F(t = 1095) = P(T \le 1095) = 1 - e^{-\left(\frac{1095}{8750}\right)} = 0.1176$$

Thus, there is a roughly 12 percent chance that a part fails within 1095 hours. In this example, the CDF, $F(t)$, serves as the failure distribution. It calculates the probability of that a part fails within a particular time, t. The part reliability distribution, $R(t)$, is the complement of the failure distribution. It calculates the probability that a part survives up to time t. For the distribution in the example,

$R(t = 1095) = P(T \ge 1095) = \int_{1095}^{\infty} \left(\frac{1}{8750}\right) e^{-\frac{-u}{8750}} du = -e^{-\frac{\infty}{8750}} + e^{-\frac{1095}{8750}} = e^{-\frac{1095}{8750}} =$ 0.8824 Since the reliability distribution is the complement of the failure distribution, $R(t) = 1 - F(t)$. Thus,

$$R(1095) = 1 - F(1095) = e^{-\left(\frac{1095}{8750}\right)} = 1 - 0.1176 = 0.8824$$

confirming the answer we obtained through direct integration.

In addition to the probability of failing before a particular time, we can also evaluate the mean time to failure (MTTF).

$$MTTF = E[T] = \int_0^\infty \left(\frac{1}{8750}\right) te^{-\left(\frac{1}{8750}\right)t} dt = 8750 \text{ hours}$$

To conclude, we demonstrate a special property of the exponential distribution. Suppose a part has been in operation for 2000 operating hours. What is the probability that it fails before 3095 operating hours. We are now calculating the conditional probability

$$P(T \leq 3095|T \geq 2000) = \frac{P(2000 \leq T \leq 3095)}{P(T \geq 2000)} = \frac{F(3095) - F(2000)}{1 - F(2000)}$$
$$= \frac{e^{-\left(\frac{2000}{8750}\right)} - e^{-\left(\frac{3095}{8750}\right)}}{e^{-\left(\frac{2000}{8750}\right)}} = 1 - e^{-\left(\frac{1095}{8750}\right)} = F(1095)$$

Thus, $P(T \leq t_1 + t_2|T > t_1) = P(T \leq t_2)$ where $t_1 = 2000$ and $t_2 = 1095$. This is called the memoryless property. In terms of reliability, it means that the remaining time to failure is independent of how long the part has been operating. Amongst all continuous distributions, this property is unique to the exponential distribution.

References

Montgomery, D. C., & Runger, G. C. (2007). Applied statistics and probability for engineers. John Wiley & Sons, Inc.

Ebeling, C. E. (2010). An introduction to reliability and maintainability engineering. Long Grove, IL: Waveland Press.