1

Probability Theory

> Doubt is disagreeable, but certainty is ridiculous.
>
> Voltaire

1.1 Basic Set Theory and Mathematical Notation

A *set* is a collection of objects. We can represent a set by enumerating its objects. Thus,

$$A = \{1, 3, 5, 7, 9, 34\}$$

is the set of single digit odd numbers plus the number 34. We can also represent the same set by a formula. For instance,

$$A = \{x | x \in \mathbf{N} \wedge (x < 10 \wedge x \text{ is odd}) \vee (x = 34)\}.$$

In interpreting this formula, **N** is the set of natural numbers (positive integers), "|" means "such that," "∈" means "is a element of," $\wedge$ is the logical symbol for "and," and $\vee$ is the logical symbol for "or." See the table of symbols in chapter 14 if you forget the meaning of a mathematical symbol.

The subset of objects in set X that satisfy property p can be written as

$$\{x \in X | p(x)\}.$$

The *union* of two sets $A, B \subset X$ is the subset of X consisting of elements of X that are in *either* A or B:

$$A \cup B = \{x | x \in A \vee x \in B\}.$$

The *intersection* of two sets $A, B \subset X$ is the subset of X consisting of elements of X that are in *both* A or B:

$$A \cap B = \{x | x \in A \wedge x \in B\}.$$

If $a \in A$ and $b \in B$, the *ordered pair* (a, b) is an entity such that if $(a, b) = (c, d)$, then $a = c$ and $b = d$. The set $\{(a, b) | a \in A \wedge b \in B\}$

is called the *product* of A and B and is written $A \times B$. For instance, if $A = B = \mathbf{R}$, where **R** is the set of real numbers, then $A \times B$ is the real plane, or the real two-dimensional vector space. We also write

$$\prod_{i=1}^{n} A_i = A_1 \times \ldots \times A_n.$$

A *function* f can be thought of as a set of ordered pairs $(x, f(x))$. For instance, the function $f(x) = x^2$ is the set

$$\{(x, y) | (x, y \in \mathbf{R}) \wedge (y = x^2)\}$$

The set of arguments for which f is defined is called the *domain* of f and is written $\mathrm{dom}(f)$. The set of values that f takes is called the *range* of f and is written $\mathrm{range}(f)$. The function f is thus a subset of $\mathrm{dom}(f) \times \mathrm{range}(f)$. If f is a function defined on set A with values in set B, we write $f{:}A \rightarrow B$.

1.2 Probability Spaces

We assume a finite *universe* or *sample space* Ω and a set $\mathcal{X}$ of subsets $A, B, C, \ldots$ of Ω, called *events*. We assume $\mathcal{X}$ is closed under finite unions (if $A_1, A_2, \ldots A_n$ are events, so is $\cup_{i=1}^{n} A_i$), finite intersections (if $A_1, \ldots, A_n$ are events, so is $\cap_{i=1}^{n} A_i$), and complementation (if A is an event so is the set of elements of Ω that are not in A, which we write A^c). If A and B are events, we interpret $A \cap B = AB$ as the event "A and B both occur," $A \cup B$ as the event "A or B occurs," and A^c as the event "A does not occur."

For instance, suppose we flip a coin twice, the outcome being HH (heads on both), HT (heads on first and tails on second), TH (tails on first and heads on second), and TT (tails on both). The sample space is then $\Omega = \{HH, TH, HT, TT\}$. Some events are $\{HH, HT\}$ (the coin comes up heads on the first toss), $\{TT\}$ (the coin comes up tails twice), and $\{HH, HT, TH\}$ (the coin comes up heads at least once).

The *probability* of an event $A \in \mathcal{X}$ is a real number $\mathrm{P}[A]$ such that $0 \leq \mathrm{P}[A] \leq 1$. We assume that $\mathrm{P}[\Omega] = 1$, which says that with probability 1 *some* outcome occurs, and we also assume that if $A = \cup_{i=1}^{n} A_i$, where $A_i \in \mathcal{X}$ and the $\{A_i\}$ are disjoint (that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$), then $\mathrm{P}[A] = \sum_{i=1}^{n} P[A_i]$, which says that probabilities are additive over finite disjoint unions.

1.3 De Morgan's Laws

Show that for any two events A and B, we have

$$(A \cup B)^c = A^c \cap B^c$$

and

$$(A \cap B)^c = A^c \cup B^c.$$

These are called *De Morgan's laws*. Express the meaning of these formulas in words.

Show that if we write p for proposition "event A occurs" and q for "event B occurs," then

$$\text{not } (p \text{ or } q) \Leftrightarrow (\text{ not } p \text{ and not } q),$$

$$\text{not } (p \text{ and } q) \Leftrightarrow (\text{ not } p \text{ or not } q).$$

The formulas are also De Morgan's laws. Give examples of both rules.

1.4 Interocitors

An interocitor consists of two kramels and three trums. Let A_k be the event "the kth kramel is in working condition," and B_j is the event "the jth trum is in working condition." An interocitor is in working condition if at least one of its kramels and two of its trums are in working condition. Let C be the event "the interocitor is in working condition." Write C in terms of the A_k and the B_j.

1.5 The Direct Evaluation of Probabilities

THEOREM 1.1 *Given* $a_1, \ldots, a_n$ *and* $b_1, \ldots, b_m$, *all distinct, there are* $n \times m$ *distinct ways of choosing one of the* a_i *and one of the* b_j. *If we also have* $c_1, \ldots, c_r$, *distinct from each other, the* a_i *and the* b_j, *then there are* $n \times m \times r$ *distinct ways of choosing one of the* a_i, *one of the* b_j, *and one of the* c_k.

Apply this theorem to determine how many different elements there are in the sample space of

a. the double coin flip

b. the triple coin flip
c. rolling a pair of dice.

Generalize the theorem.

1.6 Probability as Frequency

Suppose the sample space Ω consists of a finite number n of equally probable elements. Suppose the event A contains m of these elements. Then the *probability of the event* A is m/n.

A second definition: Suppose an experiment has n distinct outcomes, all of which are equally likely. Let A be a subset of the outcomes, and $n(A)$ the number of elements of A. We define the *probability* of A as $\mathrm{P}[A] = n(A)/n$.

For example, in throwing a pair of dice, there are $6 \times 6 = 36$ mutually exclusive, equally likely events, each represented by an ordered pair (a, b), where a is the number of spots showing on the first die and b the number on the second. Let A be the event that both dice show the same number of spots. Then $n(A) = 6$ and $\mathrm{P}[A] = 6/36 = 1/6$.

A third definition: Suppose an experiment can be repeated any number of times, each outcome being independent of the ones before and after it. Let A be an event that either does or does not occur for each outcome. Let $n_t(A)$ be the number of times A occurred on all the tries up to and including the t^{th} try. We define the *relative frequency* of A as $n_t(A)/t$, and we define the *probability of* A as

$$\mathrm{P}[A] = \lim_{t \to \infty} \frac{n_t(A)}{t}.$$

We say two events A and B are *independent* if $\mathrm{P}[A]$ does not depend on whether B occurs or not and, conversely, $\mathrm{P}[B]$ does not depend on whether A occurs or not. If events A and B are independent, the probability that both occur is the product of the probabilities that either occurs: that is,

$$\mathrm{P}[A \text{ and } B] = \mathrm{P}[A] \times \mathrm{P}[B].$$

For example, in flipping coins, let A be the event "the first ten flips are heads." Let B be the event "the eleventh flip is heads." Then the two events are independent.

For another example, suppose there are two urns, one containing 100 white balls and 1 red ball, and the other containing 100 red balls and 1 white ball. You do not know which is which. You choose 2 balls from the first urn. Let A be the event "The first ball is white," and let B be the event "The second ball is white." These events are not independent, because if you draw a white ball the first time, you are more likely to be drawing from the urn with 100 white balls than the urn with 1 white ball.

Determine the following probabilities. Assume all coins and dice are "fair" in the sense that H and T are equiprobable for a coin, and $1,\ldots,6$ are equiprobable for a die.

a. At least one head occurs in a double coin toss.
b. Exactly two tails occur in a triple coin toss.
c. The sum of the two dice equals 7 or 11 in rolling a pair of dice.
d. All six dice show the same number when six dice are thrown.
e. A coin is tossed seven times. The string of outcomes is HHHHHHH.
f. A coin is tossed seven times. The string of outcomes is HTHHTTH.

1.7 Craps

A roller plays against the casino. The roller throws the dice and wins if the sum is 7 or 11, but loses if the sum is 2, 3, or 12. If the sum is any other number (4, 5, 6, 8, 9, or 10), the roller throws the dice repeatedly until either winning by matching the first number rolled or losing if the sum is 2, 7, or 12 ("crapping out"). What is the probability of winning?

1.8 A Marksman Contest

In a head-to-head contest Alice can beat Bonnie with probability p and can beat Carole with probability q. Carole is a better marksman than Bonnie, so $p > q$. To win the contest Alice must win at least two in a row out of three head-to-heads with Bonnie and Carole and cannot play the same person twice in a row (that is, she can play Bonnie-Carole-Bonnie or Carole-Bonnie-Carole). Show that Alice maximizes her probability of winning the contest playing the better marksman, Carole, twice.

1.9 Sampling

The mutually exclusive outcomes of a random action are called *sample points*. The set of sample points is called the *sample space*. An *event* A is a subset of a sample space Ω. The event A is *certain* if $A = \Omega$ and *impossible* if $A = \emptyset$ (that is, A has no elements). The *probability* of an event A is $\mathrm{P}[A] = n(A)/n(\Omega)$, if we assume Ω is finite and all $\omega \in \Omega$ are equally likely.

a. Suppose six dice are thrown. What is the probability all six die show the same number?
b. Suppose we choose r object in succession from a set of n distinct objects $a_1, \ldots, a_n$, each time recording the choice and returning the object to the set before making the next choice. This gives an ordered sample of the form $(b_1, \ldots, b_r)$, where each b_j is some a_i. We call this *sampling with replacement*. Show that, in sampling r times with replacement from a set of n objects, there are n^r distinct ordered samples.
c. Suppose we choose r objects in succession from a set of n distinct objects $a_1, \ldots, a_n$, without returning the object to the set. This gives an ordered sample of the form $(b_1, \ldots, b_r)$, where each b_j is some unique a_i. We call this *sampling without replacement* . Show that in sampling r times without replacement from a set of n objects, there are

$$n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

distinct ordered samples, where $n! = n \times (n-1) \times \ldots \times 2 \times 1$.

1.10 Aces Up

A deck of 52 cards has 4 aces. A player draws 2 cards randomly from the deck. What is the probability that both are aces?

1.11 Permutations

A linear ordering of a set of n distinct objects is called a *permutation* of the objects. It is easy to see that the number of distinct permutations of $n > 0$ distinct objects is $n! = n \times (n-1) \times \ldots \times 2 \times 1$. Suppose we have a deck of cards numbered from 1 to $n > 1$. Shuffle the cards so their new order is a random permutation of the cards. What is the average number of cards that appear in the "correct" order (that is, the k^{th} card is in the k^{th} position) in the shuffled deck?

1.12 Combinations and Sampling

The number of *combinations* of n distinct objects taken r at a time is the number of subsets of size r, taken from the n things without replacement. We write this as $\binom{n}{r}$. In this case, we do not care about the order of the choices. For instance, consider the set of numbers $\{1,2,3,4\}$. The number of samples of size two without replacement = 4!/2! = 12. These are precisely $\{12,13,14,21,23,24,31,32,34,41,42,43\}$. The combinations of the four numbers of size two (that is, taken two at a time) are $\{12,13,14,23,24,34\}$, or six in number. Note that $6 = \binom{4}{2} = 4!/2!2!$. A set of n elements has $n!/r!(n-r)!$ distinct subsets of size r. Thus, we have

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

1.13 Mechanical Defects

A shipment of seven machines has two defective machines. An inspector checks two machines randomly drawn from the shipment, and accepts the shipment if neither is defective. What is the probability the shipment is accepted?

1.14 Mass Defection

A batch of 100 manufactured items is checked by an inspector, who examines 10 items at random. If none is defective, she accepts the whole batch. What is the probability that a batch containing 10 defective items will be accepted?

1.15 House Rules

Suppose you are playing the following game against the house in Las Vegas. You pick a number between one and six. The house rolls three dice, and pays you \$1,000 if your number comes up on one die, \$2,000 if your number comes up on two dice, and \$3,000 if your number comes up on all three dice. If your number does not show up at all, you pay the house \$1,000. At first glance, this looks like a *fair game* (that is, a game in which the expected payoff is zero), but in fact it is not. How much can you expect to win (or lose)?

1.16 The Addition Rule for Probabilities

Let A and B be two events. Then $0 \leq \mathrm{P}[A] \leq 1$ and

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[AB].$$

If A and B are disjoint (that is, the events are mutually exclusive), then

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B].$$

Moreover, if $A_1, \ldots, A_n$ are mutually disjoint, then

$$\mathrm{P}[\cup_i A_i] = \sum_{i=1}^{n} \mathrm{P}[A_i].$$

We call events $A_1, \ldots, A_n$ a *partition* of the sample space Ω if they are mutually disjoint and exhaustive (that is, their union is Ω). In this case for any event B, we have

$$\mathrm{P}[B] = \sum_i \mathrm{P}[BA_i].$$

1.17 A Guessing Game

Each day the call-in program on a local radio station conducts the following game. A number is drawn at random from $\{1, 2, \ldots, n\}$. Callers choose a number randomly and win a prize if correct. Otherwise, the station announces whether the guess was high or low and moves on to the next caller, who chooses randomly from the numbers that can logically be correct, given the previous announcements. What is the expected number $f(n)$ of callers before one guesses the number?

1.18 North Island, South Island

Bob is trying to find a secret treasure buried in the ground somewhere in North Island. According to local custom, if Bob digs and finds the treasure, he can keep it. If the treasure is not at the digging point, though, and Bob happens to hit rock, Bob must go to South Island. On the other hand, if Bob hits clay on North Island, he can stay there and try again. Once on South Island, to get back to North Island, Bob must dig and hit clay. If Bob hits rock on South Island, he forfeits the possibility of obtaining the treasure.

On the other hand, if Bob hits earth on South Island, he can stay on South Island and try again. Suppose q_n is the probability of finding the treasure when digging at a random spot on North Island, r_n is the probability of hitting rock on North Island, r_s is the probability of hitting rock on South Island, and e_s is the probability of hitting earth on South Island. What is the probability, P_n, that Bob will eventually find the treasure before he forfeits, if we assume that he starts on North Island?

1.19 Conditional Probability

If A and B are events, and if the probability $\mathrm{P}[B]$ that B occurs is strictly positive, we define the *conditional probability* of A given B, denoted $\mathrm{P}[A|B]$, by

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

We say $B_1, \ldots, B_n$ are a *partition* of event B if $\cup_i B_i = B$ and $B_i B_j = \emptyset$ for $i \neq j$. We have:

a. If A and B are events, $\mathrm{P}[B] > 0$, and B implies A (that is, $B \subseteq A$), then $\mathrm{P}[A|B] = 1$.
b. If A and B are contradictory (that is, $AB = \emptyset$), then $\mathrm{P}[A|B] = 0$.
c. If $A_1, \ldots, A_n$ are a partition of event A, then

$$\mathrm{P}[A|B] = \sum_{i=1}^{n} \mathrm{P}[A_i|B].$$

d. If $B_1, \ldots, B_n$ are a partition of the sample space Ω, then

$$\mathrm{P}[A] = \sum_{i=1}^{n} \mathrm{P}[A|B_i]\,\mathrm{P}[B_i].$$

1.20 Bayes' Rule

Suppose A and B are events with $\mathrm{P}[A], \mathrm{P}[B], \mathrm{P}[B^c] > 0$. Then we have

$$\mathrm{P}[B|A] = \frac{\mathrm{P}[A|B]\,\mathrm{P}[B]}{\mathrm{P}[A|B]\,\mathrm{P}[B] + \mathrm{P}[A|B^c]\,\mathrm{P}[B^c]}.$$

This follows from the fact that the denominator is just $\mathrm{P}[A]$, and is called *Bayes' rule*.

More generally, if $B_1, \ldots, B_n$ is a partition of the sample space and if $\mathrm{P}[A], \mathrm{P}[B_k] > 0$, then

$$\mathrm{P}[B_k|A] = \frac{\mathrm{P}[A|B_k]\,\mathrm{P}[B_k]}{\sum_{i=1}^{n} \mathrm{P}[A|B_i]\,\mathrm{P}[B_i]}.$$

To see this, note that the denominator on the right-hand side is just $\mathrm{P}[A]$, and the numerator is just $\mathrm{P}[AB_k]$ by definition.

1.21 Extrasensory Perception

Alice claims to have ESP. She says to Bob, “Match me against a series of opponents in picking the high card from a deck with cards numbered 1 to 100. I will do better than chance in either choosing a higher card than my opponent or choosing a higher card on my second try than on my first.” Bob reasons that Alice will win on her first try with probability 1/2, and beat her own card with probability 1/2 if she loses on the first round. Thus, Alice should win with probability $(1/2) + (1/2)(1/2) = 3/4$. He finds, to his surprise, that Alice wins about 5/6 of the time. Does Alice have ESP?

1.22 Les Cinq Tiroirs

You are looking for an object in one of five drawers. There is a 20% chance that it is not in any of the drawers, but if it is in a drawer, it is equally likely to be in each one. Show that as you look in the drawers one by one, the probability of finding the object in the next drawer rises if not found so far, but the probability of not finding it at all also rises.

1.23 Drug Testing

Bayes’ rule is useful because often we know $\mathrm{P}[A|B]$, $\mathrm{P}[A|B^c]$ and $\mathrm{P}[B]$, and we want to find $\mathrm{P}[B|A]$. For example, suppose 5% of the population uses drugs, and there is a drug test that is 95% accurate: it tests positive on a drug user 95% of the time, and it tests negative on a drug nonuser 95% of the time. Show that if an individual tests positive, the probability of his being a drug user is 50%. Hint: Let A be the event “is a drug user,” let “Pos” be the event “tests positive,” let “Neg” be the event “tests negative,” and apply Bayes’ rule.