Probability Theory

Lecture Notes for Math. 232

Feras Awad

Philadelphia University

May 15, 2019

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Current Subject

1. Probability
   - Sample Spaces and Events
   - The Probability of an Event
   - Combinatorial Probability
   - Conditional Probability and Independent Events
   - Bayes' Rule

Definition 1 (SAMPLE SPACE)

The set of all possible outcomes of an experiment is called the sample space and it is usually denoted by the letter S or Ω. Each outcome in a sample space is called an element of the sample space, or simply a sample point.

Definition 2 (EVENT)

An event is a subset of a sample space.

Example 1 (Flipping 2 Coins)

The sample space for the possible outcomes of two flips of a coin may be written

$$S = \left\{HH, HT, TH, TT\right\}$$

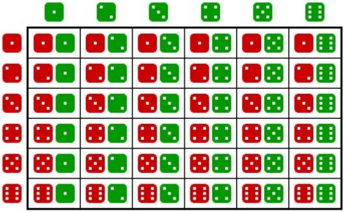
Getting different flips is an event of this experiment, and we write $A = \left\{HT, TH\right\}$.

Note: The sample space S in the previous example contained a finite number of elements; but if a coin is flipped until a head appears for the first time, we obtain the sample space with unending elements.

$$S = \left\{H, TH, TTH, TTTH, TTTTH, \cdots\right\}$$

Example 2 (Rolling 2 Dice)

The sample space that might be appropriate for an experiment in which we roll a pair of dice, one red and one green, where the elements are the totals of the numbers turned up by the two dice is $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$



The event $B = \{3, 6, 9, 12\}$ describes the number in S that is divisible by 3.

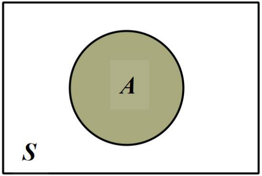
Exercise 1

(1) *Three dice are tossed, one red, one blue, and one green. What outcomes make up the event A that the sum of the three faces showing equals 5?*

(2) *An urn contains six chips numbered 1 through 6. Three are drawn out. What outcomes are in the event "Second smallest chip is a 3"? Assume that the order of the chips is irrelevant.*

Unions, Intersections, and Complements

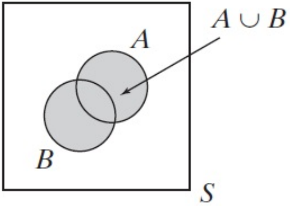
* In many problems of probability we are interested in events that are actually combinations of two or more events, formed by taking unions, intersections, and complements.
* Sample spaces and events, particularly relationships among events, are often depicted by means of Venn diagrams, in which the sample space is represented by a rectangle, while events are represented by regions within the rectangle, usually by circles or parts of circles.



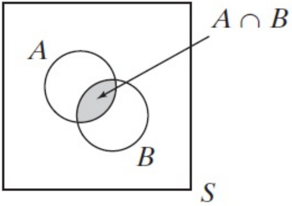
Definition 3

Let A and B be any two events defined over the same sample space S. Then

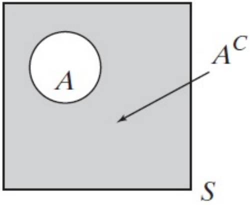
(1) The union of A and B, written $A \cup B$, is the subset of S that contains all the elements that are either in A, in B, or in both.



(2) The intersection of A and B, written $A \cap B$, is the subset of S that contains all the elements that are in both A and B.



(3) The complement of A, written A^c, is the subset of S that contains all the elements of S that are not in A.



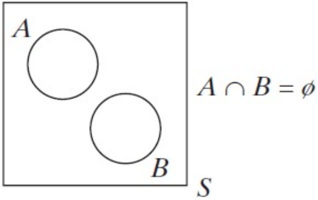
Example 3

Let A be the set of x's for which $x^2 + 2x = 8$; let B be the set for which $x^2 + x = 6$. Find $A \cap B$ and $A \cup B$.

Solution: Since the first equation factors into $(x + 4)(x - 2) = 0$, its solution set is $A = \left\{ -4, 2 \right\}$. Similarly, the second equation can be written $(x + 3)(x - 2) = 0$, making $B = \left\{ -3, 2 \right\}$. Therefore, $A \cap B = \left\{ 2 \right\}$ and $A \cup B = \left\{ -4, -3, 2 \right\}$.

Definition 4 (MUTUALLY EXCLUSIVE EVENTS)

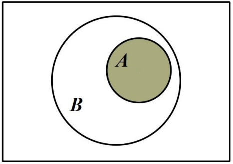
Two events A and B having no elements in common are said to be mutually exclusive – that is $A \cap B = \Phi$ where Φ is the empty set.



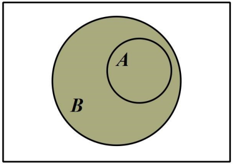
Example 4

In a single throw of two dice, define A to be the event that the two faces showing an odd sum. Let B be the event that the two faces themselves are odd. Then clearly A and B are mutually exclusive since the sum of two odd numbers necessarily being even.

Note: If A is subset of B, in symbols $A \subseteq B$, then $A \cup B = B$ and $A \cap B = A$.



$A \cap B$



$A \cup B$

Example 5

Suppose the events $A_1, A_2, \cdots, A_k$ are telescoping intervals of real numbers such that $A_i = [0, 1/i)$; $i = 1, 2, \cdots, k$. Then

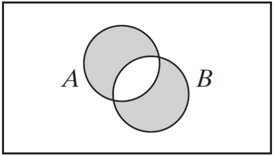
$$A_1 \cap A_2 \cdots \cap A_k = [0, 1) \cap [0, 1/2) \cap \cdots \cap [0, 1/k) = [0, 1/k)$$
$$A_1 \cup A_2 \cdots \cup A_k = [0, 1) \cup [0, 1/2) \cup \cdots \cup [0, 1/k) = [0, 1)$$

The Use of Venn Diagrams

Example 6

When two events A and B are defined on a sample space, we will frequently need to consider

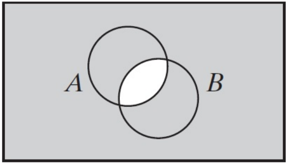
(1) the event that exactly one (of the two) occurs.
The shaded area in the figure represents the event E that either A or B, but not both, occurs (that is, exactly one occurs).



$E = (A \cap B^C) \cup (B \cap A^C)$

Note: The event that only A occurs is $A \cap B^c$, and the event that only B occurs is $B \cap A^c$.

(2) the event that at most one (of the two) occurs.
The shaded area in the figure represents the event F that at most one of the two events occurs.



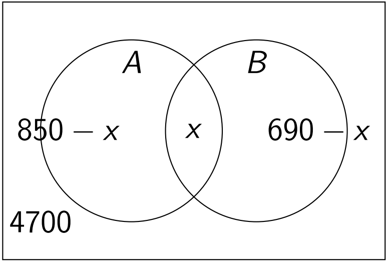
$$F = (A \cap B)^C$$

Note: $F = (A \cap B)^c = A^c \cup B^c$ (De Morgan Laws).

Example 7

During orientation week, the latest Spiderman movie was shown twice at State University. Among the entering class of 6000 freshmen, 850 went to see it the first time, 690 the second time, while 4700 failed to see it either time. How many saw it twice?

Solution: Since $6000 = (850 - x) + x + (690 - x) + 4700$, then $x = 240$.



Exercise 2

(1) ***Use Venn diagrams to verify:***

* ***Associative Laws:***
  (a) $A \cup (B \cup C) = (A \cup B) \cup C$.
  (b) $A \cap (B \cap C) = (A \cap B) \cap C$.
* ***Distributive Laws:***
  (a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
  (b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
* ***De Morgan Laws:***
  (a) $(A \cup B)^c = A^c \cap B^c$.
  (b) $(A \cap B)^c = A^c \cup B^c$.

(2) *For two events A and B defined on a sample space S, $N(A \cap B^c) = 15$, $N(A^c \cap B) = 50$, and $N(A \cap B) = 2$. Given that $N(S) = 120$, how many outcomes belong to neither A nor B?*

Current Subject

1. Probability
   - Sample Spaces and Events
   - The Probability of an Event
   - Combinatorial Probability
   - Conditional Probability and Independent Events
   - Bayes' Rule

If A is any event defined on a sample space S, the symbol $P(A)$ will denote the probability of A, and we will refer to $P : A \to \Big[0, 1\Big]$ as the probability function with the following axioms:

(1) Let A be any event defined over S. Then $P(A) \geq 0$.

(2) $P(S) = 1$.

(3) If $A_1, A_2, A_3, \cdots$ is a finite or infinite sequence of mutually exclusive events of S, $\Big(A_i \cap A_j = \Phi$ for each $i \neq j\Big)$, then

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P\left(A_1\right) + P\left(A_2\right) + P\left(A_3\right) + \cdots$$

Example 8

An experiment has four possible outcomes, A, B, C, and D, that are mutually exclusive. Explain why the following assignments of probabilities are not permissible:

(1) $P(A) = 0.12$, $P(B) = 0.63$, $P(C) = 0.45$, $P(D) = -0.20$;

(2) $P(A) = {}^{9}/_{120}$, $P(B) = {}^{45}/_{120}$, $P(C) = {}^{27}/_{120}$, $P(D) = {}^{46}/_{120}$

Solution:

(1) $P(D) = -0.20$ violates axiom 1;

(2) $P(S) = P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D) = {}^{127}/_{120} \neq 1$, and this violates axiom 2.

Theorem 1

If A is an event in a discrete sample space S, then $P(A)$ equals the sum of the probabilities of the individual outcomes comprising A.

Proof.

Let $O_1, O_2, O_3, \cdots$, be the finite or infinite sequence of outcomes that comprise the event A. Thus,

$$A = O_1 \cup O_2 \cup O_3 \cup \cdots$$

and since the individual outcomes, the O's, are mutually exclusive, the third axiom of probability yields

$$P(A) = P\left(O_1\right) + P\left(O_2\right) + P\left(O_3\right) + \cdots$$

□

Example 9

If we twice flip a balanced coin, what is the probability of getting at least one head?

Solution: Letting A denote the event that we will get at least one head. Since $S = \Big\{HH, HT, TH, TT\Big\}$, we get $A = \Big\{HH, HT, TH\Big\}$, and $P(A) = P(HH) + P(HT) + P(TH) = {}^{1}\!/_{4} + {}^{1}\!/_{4} + {}^{1}\!/_{4} = {}^{3}\!/_{4}$.

Exercise 3

A die is loaded in such a way that each odd number is twice as likely to occur as each even number. Find $P(G)$, where G is the event that a number greater than 3 occurs on a single roll of the die.

Theorem 2

If an experiment can result in any one of N different equally likely outcomes, and if n of these outcomes together constitute event A, then the probability of event A is $P(A) = \frac{n}{N}$.

Proof.

Let $O_1, O_2, \cdots, O_N$ represent the individual outcomes in S, each with probability $1/N$. If A is the union of n of these mutually exclusive outcomes, and it does not matter which ones, then

$$\begin{aligned} P(A) &= P\left(O_1 \cup O_2 \cup \cdots \cup O_n\right) \\ &= P\left(O_1\right) + P\left(O_2\right) + \cdots + P\left(O_n\right) = \underbrace{\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}}_{n-\text{terms}} = \frac{n}{N} \end{aligned}$$

□

Theorem 3

If A and A^c are complementary events in a sample space S, then $P\left(A^c\right) = 1 - P(A)$.

Proof.

Since A and A^c are mutually exclusive, and $S = A \cup A^c$, then

$$\begin{aligned} 1 &= P(S) \\ &= P\left(A \cup A^c\right) \\ &= P\left(A\right) + P\left(A^c\right) \\ P\left(A^c\right) &= 1 - P\left(A\right) \end{aligned}$$

$\square$

Theorem 4

$P(\Phi) = 0$ *for any sample space* S*.*

Proof.

Since S and Φ are mutually exclusive and $S \cup \Phi = S$, it follows that

$$\begin{aligned} P(S) &= P(S \cup \Phi) \\ &= P(S) + P(\Phi) \end{aligned}$$

and, hence, that $P(\Phi) = 0$. $\square$

Theorem 5

If A and B are events in a sample space S and $A \subseteq B$, then $P(A) \leq P(B)$.

Proof.

Since $A \subseteq B$, we can write $B = A \cup (A^c \cap B)$. Then, since A and $A^c \cap B$ are mutually exclusive, we get:
$P(B) = P(A) + P(A^c \cap B) \geq P(A)$. □

Theorem 6

For any event A, $0 \leq P(A) \leq 1$.

Proof.

Since $\Phi \subseteq A \subseteq S$ for any event A in S, then $P(\Phi) \leq P(A) \leq P(S)$ which leads to $0 \leq P(A) \leq 1$. □

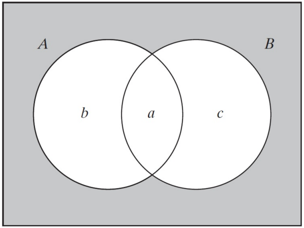
Theorem 7

If A and B are any two events in a sample space S, then
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof:

Assigning the probabilities a, b, and c to the mutually exclusive events $A \cap B$, $A \cap B^c$, and $A^c \cap B$ as in the Venn diagram below, we find that

$$\begin{aligned} P(A \cup B) &= a + b + c \\ &= (a + b) + (a + c) - a \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$



Example 10

Let A and B be two events defined on a sample space S such that $P(A) = 0.3$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$. Find

(1) $P(A \cap B)$ (2) $P(A^c \cup B^c)$ (3) $P(A^c \cap B)$

Solution:

(1) $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.3 + 0.5 - 0.7 = 0.1.$

(2) $P(A^c \cup B^c) = P\left((A \cap B)^c\right) = 1 - P(A \cap B) = 1 - 0.1 = 0.9.$

(3) $P(A^c \cap B) = P(B) - P(A \cap B) = 0.5 - 0.1 = 0.4.$

Exercise 4

(1) *Let A and B be any two events defined on S. Suppose that $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.1$. What is the probability that A or B but not both occur?*

(2) *Let A and B be two events defined on S. If the probability that at least one of them occurs is 0.3 and the probability that A occurs but B does not occur is 0.1, what is $P(B)$?*

(3) *An urn contains 24 chips, numbered 1 through 24. One is drawn at random. Let A be the event that the number is divisible by 2 and let B be the event that the number is divisible by 3. Find $P(A \cup B)$.*

(4) *Three events A, B, and C are defined on a sample space, S. Given that $P(A) = 0.2$, $P(B) = 0.1$, and $P(C) = 0.3$, what is the smallest possible value for $P\left((A \cup B \cup C)^c\right)$?*

(5) *Two dice are tossed. Assume that each possible outcome has a ${}^1\!/_{36}$ probability. Let A be the event that the sum of the faces showing is 6, and let B be the event that the face showing on one die is twice the face showing on the other. Calculate $P\left(A \cap B^c\right)$.*

Current Subject

1. Probability
   - Sample Spaces and Events
   - The Probability of an Event
   - Combinatorial Probability
   - Conditional Probability and Independent Events
   - Bayes' Rule

The mathematical methods of counting are formally known as combinatorial methods.

Theorem 8 (THE MULTIPLICATION RULE)

If an operation consists of k steps, of which the first can be done in n_1 ways, for each of these the second step can be done in n_2 ways, for each of the first two the third step can be done in n_3 ways, and so forth, then the whole operation can be done in $n_1 \times n_2 \times n_3 \times \cdots \times n_k$ ways.

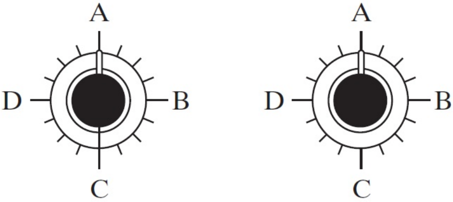
Example 11

In how many different ways can one answer all the questions of a true-false test consisting of 20 questions?

Solution: There are $2 \times 2 \times \cdots \times 2 = 2^{20} = 1048576$ different ways.

Example 12

The combination lock on a briefcase has two dials, each marked off with sixteen notches. To open the case, a person first turns the left dial in a certain direction for two revolutions and then stops on a particular mark. The right dial is set in a similar fashion, after having been turned in a certain direction for two revolutions. How many different settings are possible?



Solution: Opening the briefcase corresponds to the four-step sequence (A_1, A_2, A_3, A_4) detailed in the table. Applying the multiplication rule, we see that one $2 \times 16 \times 2 \times 16 = 1024$ different settings are possible.

Operation	Purpose	Number of Options
A_1	Rotating the left dial in a particular direction	2
A_2	Choosing an endpoint for the left dial	16
A_3	Rotating the right dial in a particular direction	2
A_4	Choosing an endpoint for the right dial	16

Example 13

How many different 7−place license plates are possible if the first 3 places are to be occupied by letters and the final 4 by numbers if repetition among letters or numbers were prohibited?

Solution: There would be $26 \times 25 \times 24 \times 10 \times 9 \times 8 \times 7 = 78624000$ possible license plates.

Exercise 5

How many integers between 100 *and* 999 *have distinct digits? How many of those are odd numbers?*

Definition 5 (PERMUTATIONS)

A permutation is a distinct arrangement of n different elements of a set.

Example 14

How many permutations are there of the letters a, b, and c?

Solution: The possible arrangements are abc, acb, bac, bca, cab, and cba, so the number of distinct permutations is $6 = 3!$.

Theorem 9

The number of permutations of n distinct objects is $n!$.

Example 15

A class in probability theory consists of 6 men and 4 women. An examination is given, and the students are ranked according to their performance. Assume that no two students obtain the same score.

(1) How many different rankings are possible?
Solution: $(6+4)! = 10! = 3628800$.

(2) If the men are ranked just among themselves and the women just among themselves, how many different rankings are possible?
Solution: $6! \times 4! = 720 \times 24 = 17280$.

Example 16

Saleem has 10 books that he is going to put on his bookshelf. Of these, 4 are mathematics books, 3 are chemistry books, 2 are history books, and 1 is a language book. Saleem wants to arrange his books so that all the books dealing with the same subject are together on the shelf. How many different arrangements are possible?

Solution: $\underbrace{4!}_{\text{subjects}} \times \left(\underbrace{4!}_{\text{math.}} \times \underbrace{3!}_{\text{chem.}} \times \underbrace{2!}_{\text{his.}} \times \underbrace{1!}_{\text{lang.}} \right) = 6912.$

Note: We shall now determine the number of permutations of a set of n objects when certain of the objects are indistinguishable from each other.

Theorem 10

The number of ways to arrange n objects, n_1 being of one kind, n_2 of a second kind, $\cdots$, and n_r of an rth kind, is

$$\frac{n!}{n_1!\, n_2! \,\cdots\, n_r!}$$

where $n_1 + n_2 + \cdots + n_r = n$.

Example 17

How many different letter arrangements can be formed from the letters ABB?

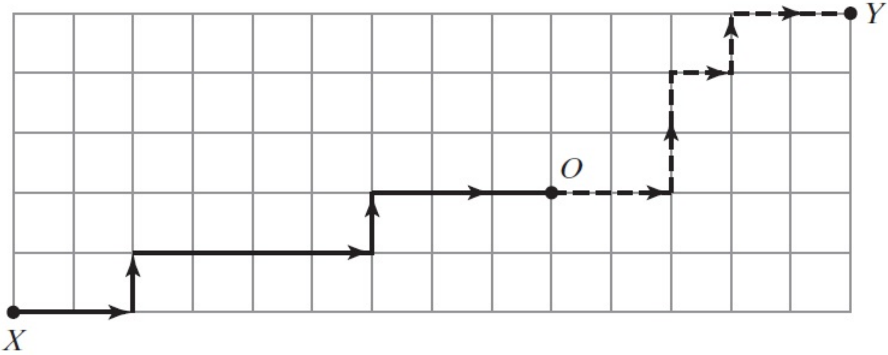
Solution: We first note that there are $3! = 6$ permutations of the letters AB_1B_2.

$$\underbrace{AB_1B_2 \quad AB_2B_1}_{ABB} \quad \underbrace{B_1AB_2 \quad B_2AB_1}_{BAB} \quad \underbrace{B_1B_2A \quad B_2B_1A}_{BBA}$$

But, two of the three are identical, then the number of permutations is $\dfrac{3!}{2!\,1!} = 3$.

Example 18

A deliveryman is currently at Point X and needs to stop at Point O before driving through to Point Y as in the figure. How many different routes can he take without ever going out of his way?



Solution: Notice that any admissible path from, say, X to O is an ordered sequence of 11 "moves" - nine east (E) and two north (N). Pictured in the figure, for example, is the particular X to O route $EENEEEEENEEE$. Similarly, any acceptable path from O to Y will necessarily consist of five moves east and three moves north (the one indicated is $EENNENEE$). By the multiplication rule, the total number of admissible routes from X to Y that pass through 0 is

$$\frac{11!}{9!\,2!} \times \frac{8!}{5!\,3!} = 55 \times 56 = 3080.$$

Theorem 11

The number of permutations of n distinct objects taken r at a time is

$$P_r^n = \frac{n!}{(n-r)!} \quad \text{for } r = 0, 1, 2, \cdots, n.$$

Example 19

Four names are drawn from among the 24 members of a club for the offices of president, vice president, treasurer, and secretary. In how many different ways can this be done?

Solution: $P_4^{24} = \frac{24!}{20!} = 24 \times 23 \times 22 \times 21 = 255024.$

Theorem 12

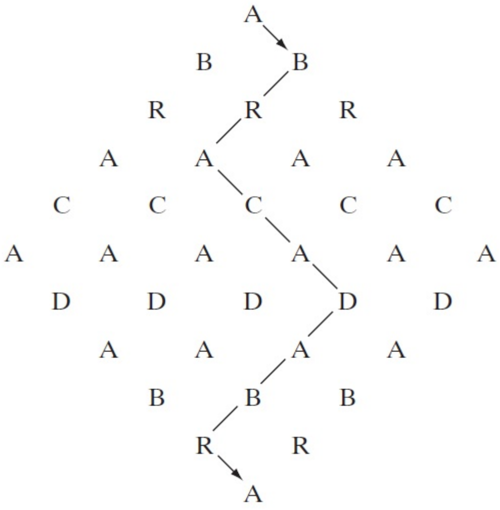
The number of permutations of n distinct objects arranged in a circle is $(n-1)!$.

Exercise 6

(1) (a) *In how many ways can 3 boys and 3 girls sit in a row?*
  (b) *In how many ways can 3 boys and 3 girls sit in a row if the boys and the girls are each to sit together?*
  (c) *In how many ways if only the boys must sit together?*
  (d) *In how many ways if no two people of the same gender are allowed to sit together?*

(2) *How many numbers greater than four million can be formed from the digits $2, 3, 4, 4, 5, 5, 5$?*

(3) *In how many ways can the word ABRACADABRA be formed in the array pictured below? Assume that the word must begin with the top A and progress diagonally downward to the bottom A.*



Definition 6 (COMBINATIONS)

A combination is a selection of r objects taken from n distinct objects without regard to the order of selection.

Note: Actually, "combination" means the same as "subset," and when we ask for the number of combinations of r objects selected from a set of n distinct objects, we are simply asking for the total number of subsets of r objects that can be selected from a set of n distinct objects.

Theorem 13

The number of combinations of n distinct objects taken r at a time is

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{P_r^n}{r!} \quad \text{for } r = 0, 1, 2, \cdots, n$$

Example 20

In how many different ways can six tosses of a coin yield two heads and four tails?

Solution: $C_2^6 = \binom{6}{2} = \frac{6!}{2!\,4!} = 15 = C_4^6.$

Example 21

From a group of 5 women and 7 men, how many different committees consisting of 2 women and 3 men can be formed? What if 2 of the men are feuding and refuse to serve on the committee together?

Solution: There are $C_2^5 \times C_3^7 = 350$ possible committees consisting of 2 women and 3 men. Now, the number of groups that do not contain both of the feuding men is

$$C_2^5 \times \left[C_0^2 C_3^5 + C_1^2 C_2^5 \right] = 300.$$

Exercise 7

(1) *How many straight lines can be drawn between five points (A, B, C, D, and E), no three of which are collinear?*

(2) *Nine students, five men and four women, interview for four summer internships sponsored by a city newspaper.*

  (a) *In how many ways can the newspaper choose a set of four interns?*

  (b) *In how many ways can the newspaper choose a set of four interns if it must include two men and two women in each set?*

  (c) *How many sets of four can be picked such that not everyone in a set is of the same gender?*

(3) *Consider a group of* 20 *people. If everyone shakes hands with everyone else, how many handshakes take place?*

(4) *From a group of* 8 *women and* 6 *men, a committee consisting of* 3 *men and* 3 *women is to be formed. How many different committees are possible if* 1 *man and* 1 *woman refuse to serve together?*

(5) *A person has* 8 *friends, of whom* 5 *will be invited to a party. How many choices if* 2 *of the friends will only attend together?*

Note:

* In the previous slides, our concern focused on counting the number of ways a given operation, or sequence of operations, could be performed.
* Now we want to couple those enumeration results with the notion of probability. Putting the two together makes a lot of sense, since there are many combinatorial problems where an enumeration, by itself, is not particularly relevant.
* In a combinatorial setting, making the transition from an enumeration to a probability is easy. If there are n ways to perform a certain operation and a total of m of those satisfy some stated condition, call it A, then $P(A)$ is defined to be the ratio $^{m}/_{n}$. This assumes, of course, that all possible outcomes are equally likely.

Example 22

An urn contains six chips, numbered 1 through 6. Two are chosen at random and their numbers are added together. What is the probability that the resulting sum is equal to 5?

Solution: $\dfrac{C_1^2}{C_2^6} = \dfrac{2}{15}$.

Example 23

An urn contains eight chips, numbered 1 through 8. A sample of three is drawn without replacement. What is the probability that the largest chip in the sample is a 5?

Solution: $\dfrac{C_1^1 \times C_2^4}{C_3^8} = \dfrac{3}{28}$.

Example 24

Suppose that n fair dice are rolled. What are the chances that all n faces will be the same?

Solution: $\dfrac{6}{6^n} = 6^{1-n}$.

Example 25

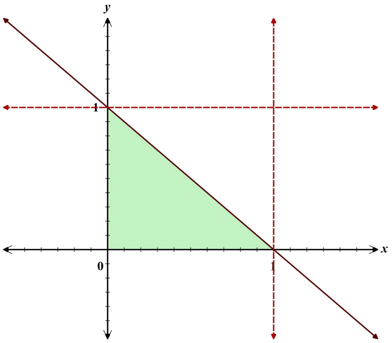
A group of 6 men and 6 women is randomly divided into 2 groups of size 6 each. What is the probability that both groups will have the same number of men?

Solution: $\dfrac{C_3^6 \times C_3^6}{C_6^{12}} = \dfrac{100}{231}$.

Example 26

What is the probability that two drawn random numbers between 0 and 1 have a sum less than or equal to 1?

Solution: The shape described by $0 \leq x, y \leq 1$ and $x + y \leq 1$ is a triangle with three vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$.



The required probability equals the area of the triangle relative to the area of the square, which is $\dfrac{0.5 \times 1 \times 1}{1} = 0.5$.

Exercise 8

(1) *Six dice are rolled one time. What is the probability that each of the six faces appears?*

(2) *Group of children,* 85 *of them are boys. If I pick two children at random, there's a* 50% *chance both are boys. How many girls are there in the group?*

(3) *In a drawer* 5 *red,* 8 *blue, and* 5 *green socks. Two drawn at random. What is the probability of getting a matching pair?*

(4) *Place the numbers* $1, 2, \cdots, 9$ *at random so that they fill a* 3×3 *grid. What is the probability that each of the row sums and each of the column sums is odd?*

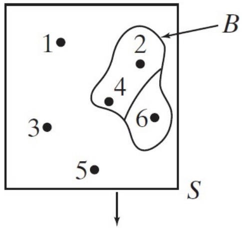
(5) *A box contains* 20 *balls numbered* $1, 2, \cdots, 20$. *If* 3 *balls are randomly taken from the box without replacement, what is the probability that one of them is the average of the other two?*

(6) *If the number is selected at random from the set of all five-digit numbers in which the sum of the digits is equal to* 43, *what is the probability that the number will be divisible by* 11*?*

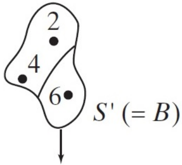
Current Subject

1. Probability
   - Sample Spaces and Events
   - The Probability of an Event
   - Combinatorial Probability
   - Conditional Probability and Independent Events
   - Bayes' Rule

Introduction: Consider a fair die being tossed, with A defined as the event "6 appears." Clearly, $P(A) = 1/6$. But suppose that the die has already been tossed by someone who refuses to tell us whether or not A occurred but does enlighten us to the extent of confirming that B occurred, where B is the event "Even number appears." What are the chances of A now? Here, common sense can help us: There are three equally likely even numbers making up the event B, one of which satisfies the event A, so the "updated" probability is $1/3$.



P (6, relative to S) = 1/6



P (6, relative to S') = 1/3

Definition 7 (CONDITIONAL PROBABILITY)

If A and B are any two events in a sample space S and $P(B) \neq 0$, the conditional probability of A given B is

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

Theorem 14 (MULTIPLICATION RULE)

If A and B are any two events in a sample space S and $P(B) \neq 0$, then $P(A \cap B) = P(A|B)P(B)$.

Example 27

Ten cards numbered 1 through 10 are placed in a hat, mixed and then one card is pulled at random. If the card is an even numbered card, what is the probability that its number is divisible by 3?

Solution: Let A be the event "the card's number is divisible by 3" and B be the event "the card is an even numbered card." Observe that $P(B) = \frac{5}{10}$. Now the event $A \cap B$ is the event that the card's number is both even and divisible by 3, which happens only when the number of the card is 6. Hence $P(A \cap B) = \frac{1}{10}$. Then

$$P(A|B) = \frac{1/10}{5/10} = \frac{1}{5}$$

Example 28

Suppose that two fair dice are tossed. What is the probability that the sum equals 10 given that it exceeds 8?

Solution: Let A be the sum equals 10, then

$$A = \Big\{(4,6),(6,4),(5,5)\Big\}$$

Also, let B be the sum exceeds 8, then

$$B = \Big\{(3,6),(6,3),(4,5),(5,4),(4,6),(6,4),(5,5),(5,6),(6,5),(6,6)\Big\}$$

Hence, $P(A|B) = \dfrac{3}{10}$.

Example 29

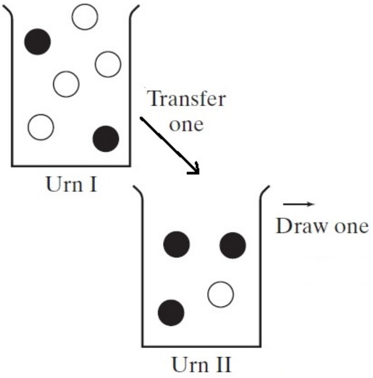
An urn contains one red chip and one white chip. One chip is drawn at random. If the chip selected is red, that chip together with two additional red chips are put back into the urn. If a white chip is drawn, the chip is returned to the urn. Then a second chip is drawn. What is the probability that both selections are red?

Solution: Let R_1 be the first chip is red, and R_2 is the second chip is red, then

$$\begin{aligned} P\left(R_1 \cap R_2\right) &= P\left(R_2 | R_1\right)\, P\left(R_1\right) \\ &= \frac{3}{4} \times \frac{1}{2} \\ &= \frac{3}{8} \end{aligned}$$

Example 30

Urn I contains two black balls and four white balls; urn II, three black and one white. A ball is drawn at random from urn I and transferred to urn II. Then a ball is drawn from urn II. What is the probability that the ball drawn from urn II is black?



Solution:

$$P\left(B_2\right) = P\left(B_2|B_1\right)P\left(B_1\right) + P\left(B_2|W_1\right)P\left(W_1\right)$$
$$= \frac{4}{5}\times\frac{2}{6} + \frac{3}{5}\times\frac{4}{6} = \frac{2}{3}.$$

Exercise 9

(1) *Find* $P(A \cap B)$ *if* $P(A) = 0.2$, $P(B) = 0.4$, *and* $P(A|B) + P(B|A) = 0.75$.

(2) *If* $P(A|B) < P(A)$, *show that* $P(B|A) < P(B)$.

(3) *Let* A *and* B *be two events such that* $P\left((A \cup B)^c\right) = 0.6$ *and* $P(A \cap B) = 0.1$. *Let* E *be the event that either* A *or* B *but not both will occur. Find* $P(E|A \cup B)$.

(4) *A fair coin is tossed three times. What is the probability that at least two heads will occur given that at most two heads have occurred?*

(5) *Two fair dice are rolled. What is the probability that the number on the first die was at least as large as* 4 *given that the sum of the two dice was* 8*?*

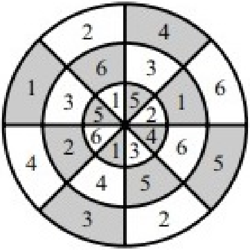
(6) *Urn I contains three red chips and one white chip. Urn II contains two red chips and two white chips. One chip is drawn from each urn and transferred to the other urn. Then a chip is drawn from the first urn. What is the probability that the chip ultimately drawn from urn I is red?*

(7) *Two events A and B are defined such that*

- *the probability that A occurs but B does not occur is* 0.2,
- *the probability that B occurs but A does not occur is* 0.1, *and*
- *the probability that neither occurs is* 0.6.

What is $P(A|B)$?

(8) *Sarah throws a dart that lands within one of the 24 numbered regions on the dartboard shown. What is the probability that the number of the region her dart hits is even? Assume that the probability of hitting on the dartboard is proportional to its area.*



Definition 8 (INDEPENDENCE)

Two events A and B are independent if and only if one of the following holds:

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

Otherwise, the events are said to be dependent.

Definition 9 (INDEPENDENCE OF MORE THAN TWO EVENTS)

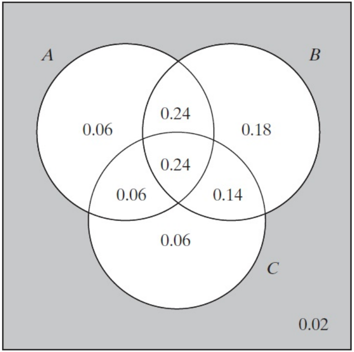
Events $A_1, A_2, \cdots$, and A_k are independent if and only if the probability of the intersections of any 2, 3, $\cdots$, or k of these events equals the product of their respective probabilities.

Example 31

The figure shows a Venn diagram with probabilities assigned to its various regions. Show that

(1) $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$ does not necessarily imply that A, B, and C are all pairwise independent.

(2) if A is independent of B and A is independent of C, then B is not necessarily independent of C.



Solution: As can seen from the diagram, $P(A) = 0.60$, $P(B) = 0.80$, $P(C) = 0.50$, $P(A \cap B) = 0.48$, $P(A \cap C) = 0.30$, $P(B \cap C) = 0.38$, and $P(A \cap B \cap C) = 0.24$. So,

(1) $P(A \cap B \cap C) = 0.24 = P(A) \times P(B) \times P(C)$, and

$$\begin{aligned} 0.48 &= P(A \cap B) = P(A) \times P(B) = 0.48 \quad ✓ \\ 0.30 &= P(A \cap C) = P(A) \times P(C) = 0.30 \quad ✓ \\ 0.38 &= P(B \cap C) \neq P(B) \times P(C) = 0.24 \quad ✗ \end{aligned}$$

(2) It is clear from part (1).

Example 32

Prove that if A and B are independent, then A and B^c are also independent.

Proof:

$$\begin{aligned} P\left(A \cap B^c\right) &= P(A) - P(A \cap B) = P(A) - P(A) \times P(B) \\ &= P(A)\Big[1 - P(B)\Big] = P(A) \times P\left(B^c\right) \end{aligned}$$

Example 33

A coin is tossed and a die is rolled. Find the probability of tossing a tail and then rolling a number greater than 2.

Solution: Let A be tossing a tail, and B be rolling a number greater than 2, then $P(A \cap B) = P(A) \times P(B) = \frac{1}{2} \times \frac{4}{6} = \frac{1}{3}$.

Exercise 10

(1) *Show that if events A and B are independent, then events A^c and B^c are independent.*

(2) *If events A, B, and C are independent, show that A and $B \cap C$ are independent. Then show that A and $B \cup C$ are independent.*

(3) *A sharpshooter hits a target with probability 0.75. Assuming independence, find the probabilities of getting*

  (a) *a hit followed by two misses,*

  (b) *two hits and a miss in any order.*

(4) *If two fair dice are rolled, what is the conditional probability that the first one lands on 6 given that the sum of the dice is 8?*

(5) *An urn contains* 6 *white and* 9 *black balls. If* 4 *balls are to be randomly selected without replacement, what is the probability that the first* 2 *selected are white and the last* 2 *black?*

(6) *Consider* 3 *urns. Urn A contains* 2 *white and* 4 *red balls, urn B contains* 8 *white and* 4 *red balls, and urn C contains* 1 *white and* 3 *red balls. If* 1 *ball is selected from each urn, what is the probability that the ball chosen from urn A was white given that exactly* 2 *white balls were selected?*

(7) Two men, A and B are shooting a target. The probability that A hits the target is $P(A) = \frac{1}{3}$, and the probability that B shoots the target is $P(B) = \frac{1}{5}$, one independently of the other. Find the probability that

  (a) A misses the target,
  (b) both men hit the target,
  (c) at least one of them hits the target,
  (d) none of them hits the target.

Current Subject

1. Probability
   - Sample Spaces and Events
   - The Probability of an Event
   - Combinatorial Probability
   - Conditional Probability and Independent Events
   - Bayes' Rule

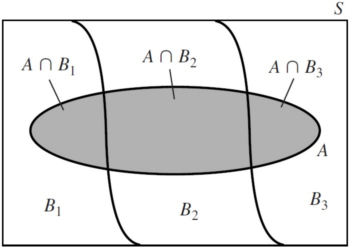
Definition 10

For some positive integer k, let the sets $B_1, B_2, \cdots, B_k$ be such that

* $S = B_1 \cup B_2 \cup \cdots \cup B_k$,
* $B_i \cap B_j = \Phi$ for $i \neq j$.

Then the collection of sets $\{B_1, B_2, \cdots, B_k\}$ is said to be a partition of S.

Note: If A is any subset of S and $\{B_1, B_2, \cdots, B_k\}$ is a partition of S, then A can be decomposed as follows:
$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$.



Theorem 15

Assume that $\{B_1, B_2, \cdots, B_k\}$ *is a partition of* S *such that* $P(B_i) \neq 0$, *for* $i = 1, 2, \cdots, k$. *Then for any event* A*:*

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)\, P(B_i)$$

Theorem 16 (Bayes' Rule)

Assume that $\{B_1, B_2, \cdots, B_k\}$ *is a partition of* S *such that* $P(B_i) \neq 0$, *for* $i = 1, 2, \cdots, k$. *Then*

$$P(B_j|A) = \frac{P(A|B_j)\, P(B_j)}{\sum\limits_{i=1}^{k} P(A|B_i)\, P(B_i)}$$

Example 34

A student answers a multiple-choice examination question that offers four possible answers. Suppose the probability that the student knows the answer to the question is 0.8 and the probability that the student will guess is 0.2. If the student correctly answers a question, what is the probability that the student really knew the correct answer?

Solution: Let A be "the student correctly answers the question", B_g be "the correct answer is by guessing", and B_s be "the correct answer by sure knowledge". Then

$$\begin{aligned} P\left(B_s|A\right) &= \frac{P\left(A|B_s\right)P\left(B_s\right)}{P\left(A|B_s\right)P\left(B_s\right)+P\left(A|B_g\right)P\left(B_g\right)} \\ &= \frac{(1)(0.8)}{(1)(0.8)+(0.25)(0.2)} = \frac{16}{17} \end{aligned}$$

Example 35

A personnel director has two lists of applicants for jobs. List 1 contains the names of five women and two men, whereas list 2 contains the names of two women and six men. A name is randomly selected from list 1 and added to list 2. A name is then randomly selected from the augmented list 2. Given that the name selected is that of a man, what is the probability that a woman's name was originally selected from list 1?

Solution: Let A be "the name selected from list 2 is a man", B_w be "the name selected from list 1 is a woman", and B_m be "the name selected from list 1 is a man". Then

$$P(B_w|A) = \frac{P(A|B_w)P(B_w)}{P(A|B_w)P(B_w) + P(A|B_m)P(B_m)}$$
$$= \frac{\left(\frac{6}{9}\right)\left(\frac{5}{7}\right)}{\left(\frac{6}{9}\right)\left(\frac{5}{7}\right) + \left(\frac{7}{9}\right)\left(\frac{2}{7}\right)} = \frac{15}{22}$$

Exercise 11

(1) *On rainy days, Saleem is late to work with probability* 0.3*; on non-rainy days, he is late with probability* 0.1*. With probability* 0.7*, it will rain tomorrow.*

  (a) *Find the probability that Saleem is early tomorrow.*
  (b) *Given that Saleem was early, what is the conditional probability that it rained?*

(2) *With probability* 0.6*, the present was hidden by mom; with probability* 0.4*, it was hidden by dad. When mom hides the present, she hides it upstairs* 70% *of the time and downstairs* 30% *of the time. Dad is equally likely to hide it upstairs or downstairs.*

  (a) *What is the probability that the present is upstairs?*
  (b) *Given that it is downstairs, what is the probability it was hidden by dad?*

(3) *In a T-maze, a rat is given food if it turns left and an electric shock if it turns right. On the first trial there is a* $50-50$ *chance that a rat will turn either way; then, if it receives food on the first trial, the probability is* 0.68 *that it will turn left on the next trial, and if it receives a shock on the first trial, the probability is* 0.84 *that it will turn left on the next trial. What is the probability that a rat will turn left on the second trial?*

(4) *Saleem takes a twenty-question multiple-choice exam where each question has five possible answers. Some of the answers he knows, while others he gets right just by making lucky guesses. Suppose that the conditional probability of his knowing the answer to a randomly selected question given that he got it right is* 0.92*. How many of the twenty questions was he prepared for?*

(5) *Two sections of a senior probability course are being taught. From what she has heard about the two instructors listed, Sarah estimates that her chances of passing the course are* 0.85 *if she gets Professor X and* 0.60 *if she gets Professor Y. The section into which she is put is determined by the registrar. Suppose that her chances of being assigned to Professor X are four out of ten. Fifteen weeks later we learn that Sarah did, indeed, pass the course. What is the probability she was enrolled in Professor X's section?*

(6) *There are* 3 *coins in a box. One is a two-headed coin, another is a fair coin, and the third is a biased coin that comes up heads* 70% *of the time. When one of the* 3 *coins is selected at random and flipped, it shows heads. What is the probability that it was the two-headed coin?*

Current Subject

2. Probability Distributions and Probability Densities
   - Discrete Random Variables and Probability Distributions
   - Continuous Random Variables and Probability Density Functions
   - Multivariate Distributions
   - Marginal Distributions
   - Conditional Distributions
   - Independent Random Variables

Definition 11 (RANDOM VARIABLE)

If S is a sample space with a probability measure and X is a real-valued function defined over the elements of S, then X is called a random variable.

Example 36

If a balanced coin tossed twice. List the equally likely elements of the sample space, and the corresponding values x of the random variable X, the total number of heads.

Solution:

S	HH	HT	TH	TT
Prob.	$1/4$	$1/4$	$1/4$	$1/4$
X	2	1	1	0

Thus,

$$P\Big(\text{at least one head}\Big) = P(X \geq 1) = P(X = 1) + P(X = 2) = \frac{3}{4}.$$

Definition 12 (DISCRETE RANDOM VARIABLES)

A random variable X is said to be discrete if it can assume only a finite or countably infinite number of distinct values.

Note: Recall that a set of elements is countably infinite if the elements in the set can be put into one-to-one correspondence with the positive integers.

Definition 13 (PROBABILITY DISTRIBUTION)

If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the probability distribution of X.

Note: The probability distribution for a discrete random variable X can be represented by a formula, a table, or a graph.

Theorem 17

A function can serve as the probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions

(1) $f(x) \geq 0$ *for each value within its domain;*

(2) $\sum_{x} f(x) = 1$, *where the summation extends over all the values within its domain.*

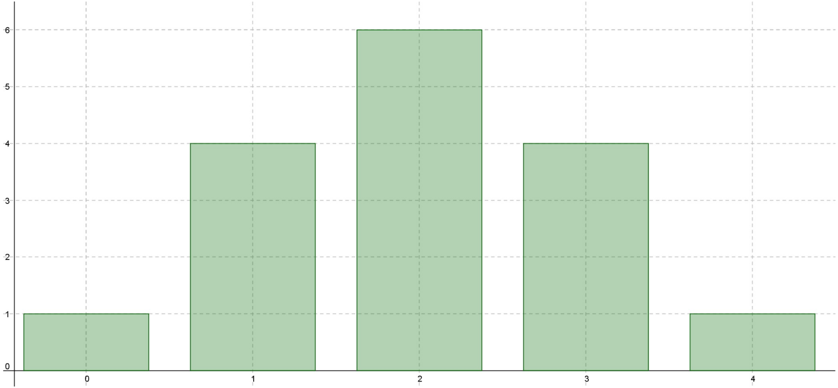
Example 37

Find a formula for the probability distribution of the total number of heads obtained in four tosses of a balanced coin.

Solution: Note that $X = 0, 1, 2, 3, 4$. Thus, the formula for the probability distribution can be written as

$$f(x) = \frac{C_x^4}{2^4}, \quad \text{for } x = 0, 1, 2, 3, 4$$

$X = x$	0	1	2	3	4
$P(X = x)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$



The graph of the figure is called a bar chart, but it is also referred to as a histogram.

There are many problems in which it is of interest to know the probability that the value of a random variable is less than or equal to some real number x.

Definition 14 (DISTRIBUTION FUNCTION)

If X is a discrete random variable, the function given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty < x < \infty$$

where $f(t)$ is the value of the probability distribution of X at t, is called the distribution function, or the cumulative distribution of X.

Note: The distribution function is defined not only for the values taken on by the given random variable, but for all real numbers.

Theorem 18

The values $F(x)$ of the distribution function of a discrete random variable X satisfy the conditions

(1) $F(-\infty) = 0$ *and* $F(\infty) = 1$;

(2) *if* $a < b$, *then* $F(a) \leq F(b)$ *for any real numbers* a *and* b.

Example 38

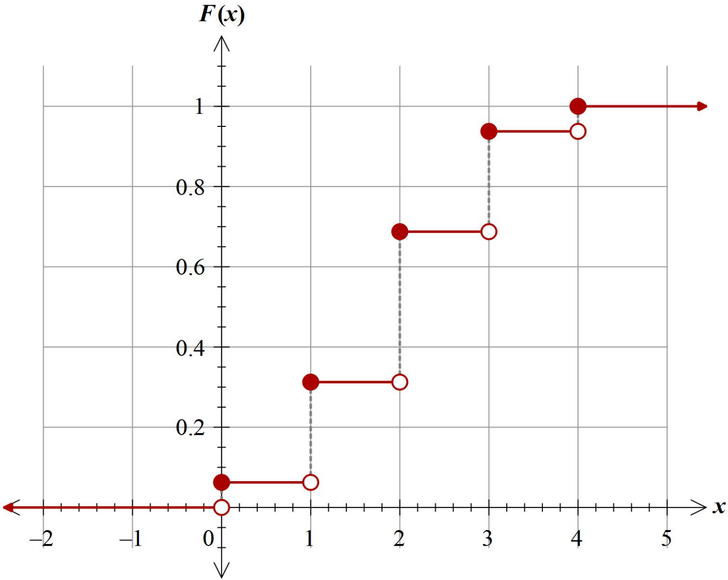
Find the distribution function of the total number of heads obtained in four tosses of a balanced coin.

Solution: From example (37), we have

$$
\begin{aligned}
F(0) &= f(0) = 1/16 \\
F(1) &= f(0) + f(1) = 5/16 \\
F(2) &= f(0) + f(1) + f(2) = 11/16 \\
F(3) &= f(0) + f(1) + f(2) + f(3) = 15/16 \\
F(4) &= f(0) + f(1) + f(2) + f(3) + f(4) = 1
\end{aligned}
$$

Hence, the distribution function is given by

$$
F(x) = \begin{cases} 0 & : \ x < 0 \\ 1/16 & : \ 0 \le x < 1 \\ 5/16 & : \ 1 \le x < 2 \\ 11/16 & : \ 2 \le x < 3 \\ 15/16 & : \ 3 \le x < 4 \\ 1 & : \ x \ge 4 \end{cases}
$$


F(x)
1
0.8
0.6
0.4
0.2
x
−2
−1
0
1
2
3
4
5

Theorem 19

If the range of a random variable X consists of the values $x_1 < x_2 < x_3 < \cdots < x_n$, then $f(x_1) = F(x_1)$, and

$$f(x_i) = F(x_i) - F(x_{i-1}) \text{ for } i = 2, 3, \cdots, n$$

Example 39

Find the probability distribution of the random variable X if its distribution function is given by

$$F(x) = \begin{cases} 0 & : \ x < 0 \\ 0.25 & : \ 0 \leq x < 1 \\ 0.75 & : \ 1 \leq x < 2 \\ 1 & : \ x \geq 2 \end{cases}$$

Solution: Making use of theorem (19), the range of X is $x_1 = 0$, $x_2 = 1$, and $x_3 = 2$, and

$$\begin{aligned} f(0) &= F(0) = 0.25 \\ f(1) &= F(1) - F(0) = 0.75 - 0.25 = 0.5 \\ f(2) &= F(2) - F(1) = 1 - 0.27 = 0.25 \end{aligned}$$

and comparison with the probabilities in the table in example (36) reveals that the random variable with which we are concerned here is the total number of heads appears if a balanced coin tossed twice.

$X = x$	0	1	2
$P(X = x)$	0.25	0.5	0.25

Exercise 12

(1) *For each of the following, determine k so that the function can serve as the probability distribution of a random variable with the given range:*

(a) $f(x) = \dfrac{x+k}{25}$ *for* $x = 1, 2, 3, 4, 5.$

(b) $f(x) = (1-k)k^x$ *for* $x = 0, 1, 2, \cdots.$

(2) *Verify the following:*

(a) $P(X > x_i) = 1 - F(x_i).$

(b) $P(X \geq x_i) = 1 - F(x_{i-1}).$

(c) $P(x_i < X \leq x_j) = F(x_j) - F(x_i).$

(d) $P(x_i \leq X \leq x_j) = F(x_j) - F(x_{i-1}).$

(3) *If X has the distribution function,*

$$F(x) = \begin{cases} 0 & : \ x < -1 \\ 0.25 & : \ -1 \leq x < 1 \\ 0.50 & : \ 1 \leq x < 3 \\ 0.75 & : \ 3 \leq x < 5 \\ 1 & : \ x \geq 5 \end{cases}$$

find:

(a) $P(X \leq 3)$, $P(X = 3)$, $P(X < 3)$,
(b) $P(X \geq 1)$,
(c) $P(-0.4 < X \leq 4)$,
(d) the probability distribution of X.

Current Subject

2. Probability Distributions and Probability Densities
   - Discrete Random Variables and Probability Distributions
   - Continuous Random Variables and Probability Density Functions
   - Multivariate Distributions
   - Marginal Distributions
   - Conditional Distributions
   - Independent Random Variables

A continuous random variable is a random variable where the data can take infinitely many, uncountable values. For example, a random variable measuring the time taken for something to be done is continuous since there are an infinite number of possible times that can be taken.

Definition 15 (PROBABILITY DENSITY FUNCTION)

A function with values $f(x)$, defined over the set of all real numbers, is called a probability density function of the continuous random variable X if and only if

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

for any $a, b \in \mathbb{R}$ with $a \leq b$.

Note:

* Probability density functions are also referred to, more briefly, as probability densities, density functions, densities, or p.d.f.'s.
* Note that $f(c)$, the value of the probability density of X at c, does not give $P(X = c)$ as in the discrete case. In connection with continuous random variables, probabilities are always associated with intervals and $P(X = c) = 0$ for any $c \in \mathbb{R}$.

$$\begin{aligned} P(X = c) &= P(c \leq X \leq c) \\ &= \int_c^c f(x)\, dx = 0 \end{aligned}$$

Theorem 20

If X is a continuous random variable and $a, b \in \mathbb{R}$ with $a \leq b$, then

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) = P(a < X < b) \end{aligned}$$

Theorem 21

A function $f(x)$ can serve as a probability density of a continuous random variable X if its values satisfy the conditions

(1) $f(x) \geq 0$ *for* $-\infty < x < \infty$;

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1.$

Example 40

Given $f(x) = \begin{cases} cx^2 & : \ 0 \leq x \leq 2 \\ 0 & : \ \text{elsewhere} \end{cases}$

(1) Find the value of c for which $f(x)$ is a valid probability density function.

(2) Evaluate $P(1 \leq X < 2)$.

Solution:

(1) We require a value for c such that

$$\int_{-\infty}^{\infty} cx^2 \, dx = 1 \rightarrow \int_{0}^{2} cx^2 \, dx = 1 \rightarrow \left. \frac{cx^3}{3} \right]_0^2 = 1 \rightarrow c = \frac{3}{8}$$

(2) $P(1 \leq X < 2) = \int_{1}^{2} \frac{3}{8} x^2 \, dx = \frac{7}{8}$.

Definition 16 (DISTRIBUTION FUNCTION)

If X is a continuous random variable and the value of its probability density at t is $f(t)$, then the function given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \quad \text{for } -\infty < x < \infty$$

is called the distribution function or the cumulative distribution function of X.

Theorem 22

If $f(x)$ and $F(x)$ are the values of the probability density and the distribution function of X at x, then $P(a \leq X \leq b) = F(b) - F(a)$ for any constants $a, b \in \mathbb{R}$ with $a \leq b$, and

$$f(x) = \frac{dF(x)}{dx}; \quad \textit{where the derivative exists.}$$

Example 41

Find the distribution function of the random variable X whose probability density is given by

$$f(x) = \begin{cases} x & : \ 0 < x < 1 \\ 2 - x & : \ 1 \leq x < 2 \\ 0 & : \ \text{elsewhere} \end{cases}$$

Solution:

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0 \; dt & : \; x \leq 0 \\ 0 + \int_{0}^{x} t \; dt & : \; 0 < x < 1 \\ \frac{1}{2} + \int_{1}^{x} (2-t) \; dt & : \; 1 \leq x < 2 \\ 1 + \int_{2}^{\infty} 0 \; dt & : \; x \geq 2 \end{cases}$$

$$= \begin{cases} 0 & : \; x \leq 0 \\ \frac{x^2}{2} & : \; 0 < x < 1 \\ -\frac{x^2}{2} + 2x - 1 & : \; 1 \leq x < 2 \\ 1 & : \; x \geq 2 \end{cases}$$

Example 42

Find a probability density function for the random variable whose distribution function is given by

$$F(x) = \begin{cases} 0 & : \; x \leq 0 \\ x & : \; 0 < x < 1 \\ 1 & : \; x \geq 1 \end{cases}$$

Solution:

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & : \; x < 0 \\ 1 & : \; 0 < x < 1 \\ 0 & : \; x > 1 \end{cases} = \begin{cases} 1 & : \; 0 < x < 1 \\ 0 & : \; \text{elsewhere} \end{cases}$$

Exercise 13

(1) *The p.d.f. of the random variable X is given by*

$$f(x) = \begin{cases} \dfrac{k}{\sqrt{x}} & : \ 0 < x < 4 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

Find:

(a) *the value of k;*
(b) *the distribution function of the random variable X;*
(c) $P\left(X < \frac{1}{4}\right)$ *and* $P(X > 1)$.

(2) *Find the distribution function of the random variable X whose probability density is given by*

$$f(x) = \begin{cases} \dfrac{x}{2} & : \ 0 < x \leq 1 \\ \dfrac{1}{2} & : \ 1 < x \leq 2 \\ \dfrac{3-x}{2} & : \ 2 < x < 3 \\ 0 & : \ \text{elsewhere} \end{cases}$$

(3) *The distribution function of the random variable Y is given by*

$$F(y) = \begin{cases} 0 & : \; y < -1 \\ \dfrac{y+1}{2} & : \; -1 \leq y \leq 1 \\ 1 & : \; y \geq 1 \end{cases}$$

Find:

(a) $P(Y = -1)$ *and* $P\left(-\frac{1}{2} \leq Y \leq \frac{1}{2}\right)$;

(b) *the probability density of Y.*

Current Subject

2. Probability Distributions and Probability Densities
   - Discrete Random Variables and Probability Distributions
   - Continuous Random Variables and Probability Density Functions
   - Multivariate Distributions
   - Marginal Distributions
   - Conditional Distributions
   - Independent Random Variables

In this section we shall be concerned first with the bivariate case, that is, with situations where we are interested at the same time in a pair of random variables defined over a joint sample space. Later, we shall extend this discussion to the multivariate case, covering any finite number of random variables.

Definition 17 (JOINT PROBABILITY DISTRIBUTION)

If X and Y are discrete random variables, the function given by $f(x, y) = P(X = x, Y = y)$ for each pair of values (x, y) within the range of X and Y is called the joint probability distribution of X and Y.

Theorem 23

A bivariate function can serve as the joint probability distribution of a pair of discrete random variables X and Y if and only if its values, $f(x, y)$, satisfy the conditions

(1) $f(x, y) \geq 0$ *for each pair of values* (x, y) *within its domain;*

(2) $\sum_{x} \sum_{y} f(x, y) = 1$, *where the double summation extends over all possible pairs* (x, y) *within its domain.*

Example 43

Two cards are selected at random from a box containing 3 red, 2 blue, and 4 white cards. If X and Y are, respectively, the numbers of red and blue cards included among the 2 cards drawn from the box, find the probabilities associated with all possible pairs of values of X and Y.

Solution: Since $X = 0, 1, 2$ and $Y = 0, 1, 2$, then the possible pairs are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$. The probability function $f(x, y)$ associated with any pair of values (x, y) within the range of the random variables X and Y is

$$f(x, y) = P(X = x, Y = y) = \frac{C_x^3 \; C_y^2 \; C_{2-x-y}^4}{C_2^9}$$

for $x = 0, 1, 2$, $y = 0, 1, 2$, and $0 \leq x + y \leq 2$. We obtain the values shown in the following table:

			x	
		0	1	2
	0	$1/6$	$1/3$	$1/12$
y	1	$2/9$	$1/6$	
	2	$1/36$		

Example 44

Determine the value of k for which the function given by

$$f(x, y) = kxy \quad \text{for } \; x = 1, 2, 3; \; y = 1, 2, 3$$

can serve as a joint probability distribution.

Solution: Since $\sum_x \sum_y f(x, y) = 1$, then

$$k + 2k + 3k + 2k + 4k + 6k + 3k + 6k + 9k = 1$$

$$36k = 1 \rightarrow k = 1/36$$

Definition 18 (JOINT DISTRIBUTION FUNCTION)

If X and Y are discrete random variables, the function given by

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} f(x,y) \text{ for } \begin{array}{c} -\infty < x < \infty \\ -\infty < y < \infty \end{array}$$

where $f(s,t)$ is the value of the joint probability distribution of X and Y at (s,t), is called the joint distribution function, or the joint cumulative distribution of X and Y.

Example 45

If the values of the joint probability distribution of X and Y are as shown in the table

	x	-1	0	1	3
y	-1	$^{2}/_{89}$	$^{1}/_{89}$	$^{2}/_{89}$	$^{10}/_{89}$
	2	$^{5}/_{89}$	$^{4}/_{89}$	$^{5}/_{89}$	$^{13}/_{89}$
	3	$^{10}/_{89}$	$^{9}/_{89}$	$^{10}/_{89}$	$^{18}/_{89}$

find:

(1) $F(1, 1)$;

(2) $P(X < 1, Y \geq 2)$;

(3) $P(X + Y > 2)$.

Solution:

(1) $F(1,1) = P(X \leq 1, Y \leq 1) = f(-1,-1) + f(0,-1) + f(1,-1)$

$$= \frac{2}{89} + \frac{1}{89} + \frac{2}{89} = \frac{5}{89}.$$

(2) $P(X < 1, Y \geq 2) = f(-1,2) + f(0,2) + f(-1,3) + f(0,3)$

$$= \frac{5}{89} + \frac{4}{89} + \frac{10}{89} + \frac{9}{89} = \frac{28}{89}.$$

(3) $P(X + Y > 2) = f(0,3) + f(1,2) + f(1,3) + f(3,2) + f(3,3)$

$$= \frac{9}{89} + \frac{5}{89} + \frac{10}{89} + \frac{13}{89} + \frac{18}{89} = \frac{55}{89}.$$

Exercise 14

If the values of the joint probability distribution of X and Y are as shown in the table

y \ x	0	1	2
0	$1/12$	$1/6$	$1/24$
1	$1/4$	$1/4$	$1/40$
2	$1/8$	$1/20$	
3	$1/120$		

find:

(1) $P(X = 1, Y = 2)$;
(2) $P(X = 0, 1 \leq Y < 3)$;
(3) $P(X + Y \leq 1)$;
(4) $P(X > Y)$;
(5) $F(2, 0)$;
(6) $F(4, 2.7)$.

Definition 19 (JOINT PROBABILITY DENSITY FUNCTION)

A bivariate function with values $f(x, y)$ defined over the $xy-$plane is called a joint probability density function of the continuous random variables X and Y if and only if

$$P\Big[(X, Y) \in A\Big] = \iint\limits_{A} f(x, y)\, dx\, dy$$

for any region A in the $xy-$plane.

Theorem 24

A bivariate function can serve as a joint probability density function of a pair of continuous random variables X and Y if its values, $f(x, y)$, satisfy the conditions

(1) $f(x, y) \geq 0$ *for* $-\infty < x < \infty \;,\; -\infty < y < \infty$

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$

Example 46

Given the joint probability density function

$$f(x,y) = \begin{cases} \frac{3}{5}x(y+x) & : \ 0 < x < 1\ ,\ 0 < y < 2 \\ 0 & : \ \text{elsewhere} \end{cases}$$

of two random variables X and Y, find $P\Big[(X,Y) \in A\Big]$, where A is the region $\Big\{(x,y) \mid 0 < x < {}^1\!/_2, 1 < y < 2\Big\}$.

Solution: $P\Big[(X,Y) \in A\Big] = P\left(0 < X < {}^1\!/_2, 1 < Y < 2\right)$

$$= \int_1^2 \int_0^{1/2} \frac{3}{5}x(y+x)\, dx\, dy = \frac{11}{80}$$

Definition 20 (JOINT DISTRIBUTION FUNCTION)

If X and Y are continuous random variables, the function given by

$$F(x,y) = P\Big(X \leq x, Y \leq y\Big) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\, dsdt$$

where $x, y \in \mathbb{R}$ and $f(s,t)$ is the joint probability density of X and Y at (s,t), is called the joint distribution function of X and Y.

Note: Partial differentiation in Definition 20 leads to

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$$

Example 47

If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{3}{5}x(y+x) & : \ 0 < x < 1 \ , \ 0 < y < 2 \\ 0 & : \ \text{elsewhere} \end{cases}$$

find the joint distribution function of these two random variables.

Solution:

* For $x \leq 0$ or $y \leq 0$, it follows immediately that $F(x,y) = 0$.
* For $0 < x < 1$ and $0 < y < 2$, we get

$$F(x,y) = \int_0^x \int_0^y \frac{3}{5}s(t+s)\, dtds = \frac{1}{20}x^2y(4x+3y)$$

* For $x \geq 1$ and $0 < y < 2$, we get

$$F(x, y) = \int_0^1 \int_0^y \frac{3}{5} s(t+s)\ dtds = \frac{1}{20} y(3y+4)$$

* For $0 < x < 1$ and $y \geq 2$, we get

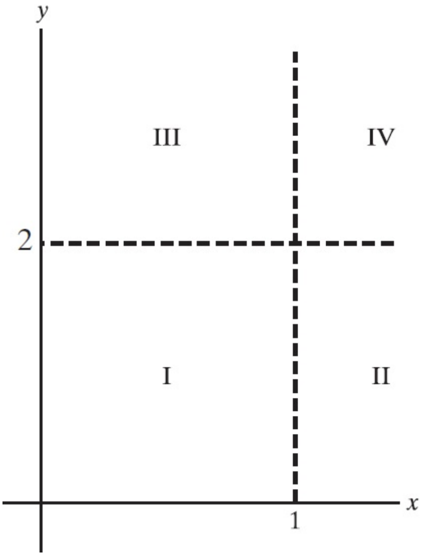
$$F(x, y) = \int_0^x \int_0^2 \frac{3}{5} s(t+s)\ dtds = \frac{1}{5} x^2(2x+3)$$

* For $x > 1$ and $y > 2$, we get

$$F(x, y) = \int_0^1 \int_0^2 \frac{3}{5} s(t+s)\ dtds = 1$$

Since the joint distribution function is everywhere continuous, the boundaries between any two of these regions can be included in either one, and we can write

$$F(x,y) = \begin{cases} 0 & : \text{ for } x \le 0 \text{ or } y \le 0 \\ \frac{1}{20}x^2y(4x+3y) & : \text{ for } 0 < x < 1 \text{ and } 0 < y < 2 \\ \frac{1}{20}y(3y+4) & : \text{ for } x \ge 1 \text{ and } 0 < y < 2 \\ \frac{1}{5}x^2(2x+3) & : \text{ for } 0 < x < 1 \text{ and } y \ge 2 \\ 1 & : \text{ for } x > 1 \text{ and } y > 2 \end{cases}$$


y
III
IV
2
I
II
x
1

Example 48

Find the joint probability density of the two random variables X and Y whose joint distribution function is given by

$$F(x,y)=\begin{cases}(1-e^{-x})(1-e^{-y}) & : \text{ for } x>0 \text{ and } y>0\\ 0 & : \text{ elsewhere}\end{cases}$$

Solution: Since

$$\begin{aligned}\frac{\partial^2}{\partial x\partial y}\left[\left(1-e^{-x}\right)\left(1-e^{-y}\right)\right] &= \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}\left(1-e^{-x}\right)\left(1-e^{-y}\right)\right]\\ &= \frac{\partial}{\partial x}\left[\left(1-e^{-x}\right)e^{-y}\right]\\ &= e^{-x}e^{-y}=e^{-(x+y)}\end{aligned}$$

then $f(x,y)=\begin{cases}e^{-(x+y)} & : \text{ for } x>0 \text{ and } y>0\\ 0 & : \text{ elsewhere}\end{cases}$

Note: All the definitions of this section can be generalized to the multivariate case, where there are n random variables.

Example 49

If the joint probability distribution of three discrete random variables X, Y, and Z is given by

$$f(x,y,z) = \frac{(x+y)z}{63} \quad \text{for } x = 1,2; \quad y = 1,2,3; \quad z = 1,2$$

find $P\Big(X = 2, Y + Z \leq 3\Big)$.

Solution: $P\Big(X = 2, Y + Z \leq 3\Big) = f(2,1,1) + f(2,1,2) + f(2,2,1)$

$$= \frac{3}{63} + \frac{6}{63} + \frac{4}{63} = \frac{13}{63}$$

Example 50

If the **trivariate** probability density of X_1,X_2, and X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)\, e^{-x_3} & : \text{ for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & : \text{ elsewhere} \end{cases}$$

find $P\Big[(X_1, X_2, X_3) \in A\Big]$, where A is the region

$$\left\{(x_1, x_2, x_3) \;\Big|\; 0 < x_1 < \frac{1}{2},\ \frac{1}{2} < x_2 < 1,\ x_3 < 1\right\}$$

Solution:

$$
\begin{aligned}
P\Big[\,(X_1, X_2, X_3) \in A\Big] &= P\left(0 < X_1 < \frac{1}{2}, \frac{1}{2} < X_2 < 1, X_3 < 1\right) \\
&= \int_0^1 \int_{1/2}^1 \int_0^{1/2} (x_1 + x_2)\, e^{-x_3}\ dx_1 dx_2 dx_3 \\
&= \int_0^1 \int_{1/2}^1 \left(\frac{1}{8} + \frac{x_2}{2}\right) e^{-x_3}\ dx_2 dx_3 \\
&= \int_0^1 \frac{1}{4} e^{-x_3}\ dx_3 \\
&= \frac{1}{4}\left(1 - e^{-1}\right)
\end{aligned}
$$

Exercise 15

(1) *Determine k so that*

$$f(x,y) = \begin{cases} kx(x-y) & : \text{for } \ 0 < x < 1, \ -x < y < x \\ 0 & : \ \text{elsewhere} \end{cases}$$

can serve as a joint probability density.

(2) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} 2 & : \text{for } \ x > 0, \ y > 0, \ x + y < 1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

find:

a) $P\left(X \leq 1/2, Y \leq 1/2\right)$

b) $P\left(X + Y \geq 2/3\right)$

c) $P\left(X \geq 2Y\right)$

(3) *Find the joint probability density of the two random variables X and Y whose joint distribution function is given by*

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & : \textit{for } \ x > 0, \ y > 0 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

then use it to find $P(X + Y > 3)$.

(4) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} 3x & : \textit{for } \ 0 \le y \le x \le 1 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

find:

a) $F\left(1/2, 1/3\right)$,

b) $P\left(Y \le \dfrac{X}{2}\right)$.

(5) *If the joint probability density of X and Y is given by*

$$f(x, y) = \begin{cases} 1 & : \textit{for } \; 0 \leq x \leq 1, \; 0 \leq y \leq 1 \\ 0 & : \; \textit{elsewhere} \end{cases}$$

a) *What is* $P\left(X - Y > 1/2\right)$ *?*
b) *What is* $P\left(XY < 1/2\right)$ *?*

Current Subject

2. Probability Distributions and Probability Densities
   - Discrete Random Variables and Probability Distributions
   - Continuous Random Variables and Probability Density Functions
   - Multivariate Distributions
   - Marginal Distributions
   - Conditional Distributions
   - Independent Random Variables

To introduce the concept of a marginal distribution, let us consider the following example.

Example 51

In Example 43 we derived the joint probability distribution of two random variables X and Y, the numbers of red and blue cards included among the 2 cards drawn from the box containing 3 red, 2 blue, and 4 white cards. Find the probability distribution of X alone and that of Y alone.

Solution: The results of Example 43 are shown in the following table, together with the marginal totals, that is, the totals of the respective rows and columns:

			x		
		0	1	2	
	0	$1/6$	$1/3$	$1/12$	$7/12$
y	1	$2/9$	$1/6$		$7/18$
	2	$1/36$			$1/36$
		$5/12$	$1/2$	$1/12$	

The column totals are the probabilities that X will take on the values 0, 1, and 2. By the same token, the row totals are the probabilities that Y will take on the values 0, 1, and 2.

x	0	1	2
$g(x)$	$5/12$	$1/2$	$1/12$

y	0	1	2
$h(y)$	$7/12$	$7/18$	$1/36$

Definition 21 (MARGINAL DISTRIBUTION.)

If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the function given by

$$g(x) = \sum_{y} f(x, y)$$

for each x within the range of X is called the marginal distribution of X. Correspondingly, the function given by

$$h(y) = \sum_{x} f(x, y)$$

for each y within the range of Y is called the marginal distribution of Y.

Definition 22 (MARGINAL DENSITY.)

If X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y), the function given by

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{for} \quad -\infty < x < \infty$$

is called the marginal density of X. Correspondingly, the function given by

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \quad \text{for} \quad -\infty < y < \infty$$

is called the marginal density of Y.

Example 52

Given the joint probability density

$$f(x,y) = \begin{cases} \frac{3}{5}x(y+x) & : \ 0 < x < 1 \ , \ 0 < y < 2 \\ 0 & : \ \text{elsewhere} \end{cases}$$

find the marginal densities of X and Y.

Solution: Performing the necessary integrations, we get

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_0^2 \frac{3}{5}x(y+x)\, dy = \frac{6}{5}x(x+1)$$

for $0 < x < 1$ and $g(x) = 0$ elsewhere. Likewise,

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_0^1 \frac{3}{5}x(y+x)\, dx = \frac{1}{10}(3y+2)$$

for $0 < y < 2$ and $h(y) = 0$ elsewhere.

Current Subject

2. Probability Distributions and Probability Densities
   - Discrete Random Variables and Probability Distributions
   - Continuous Random Variables and Probability Density Functions
   - Multivariate Distributions
   - Marginal Distributions
   - Conditional Distributions
   - Independent Random Variables

Suppose that A and B are the events $X = x$ and $Y = y$ so that we can write the conditional probability of the event $X = x$ given $Y = y$ as

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{h(y)}$$

provided $P(Y = y) = h(y) \neq 0$, where $f(x, y)$ is the value of the joint probability distribution of X and Y at (x, y), and $h(y)$ is the value of the marginal distribution of Y at y.

Definition 23 (CONDITIONAL DISTRIBUTION.)

If $f(x,y)$ is the value of the joint probability distribution of the discrete random variables X and Y at (x,y) and $h(y)$ is the value of the marginal distribution of Y at y, the function given by

$$f(x|y) = \frac{f(x,y)}{h(y)} \; ; \quad h(y) \neq 0$$

for each x within the range of X is called the conditional distribution of X given $Y = y$. Correspondingly, if $g(x)$ is the value of the marginal distribution of X at x, the function given by

$$w(y|x) = \frac{f(x,y)}{g(x)} \; ; \quad g(x) \neq 0$$

for each y within the range of Y is called the conditional distribution of Y given $X = x$.

Example 53

Given the values of the joint probability distribution of X and Y shown in the table

		x		
		0	1	2
	0	$1/9$	$2/9$	$1/9$
y	1	$2/9$	$2/9$	
	2	$1/9$		

find:

a) the marginal distribution of X;

b) the marginal distribution of Y;

c) $P(Y \geq 1|X = 0)$;

d) the conditional distribution of X given $Y = 1$.

Solution: Note that the marginal totals are

		$x=0$	$x=1$	$x=2$	
y	0	$1/9$	$2/9$	$1/9$	$4/9$
	1	$2/9$	$2/9$		$4/9$
	2	$1/9$			$1/9$
		$4/9$	$4/9$	$1/9$	

a)

x	0	1	2
$g(x)$	$4/9$	$4/9$	$1/9$

b)

y	0	1	2
$h(y)$	$4/9$	$4/9$	$1/9$

c) $$P(Y \geq 1|X=0) = \frac{P(Y \geq 1, X=0)}{P(X=0)} = \frac{f(0,1)+f(0,2)}{g(0)}$$
$$= \frac{2/9 + 1/9}{4/9} = \frac{3}{4}$$

d) Since $$f(x=0|y=1) = \frac{f(0,1)}{h(1)} = \frac{2/9}{4/9} = \frac{1}{2}$$
$$f(x=1|y=1) = \frac{f(1,1)}{h(1)} = \frac{2/9}{4/9} = \frac{1}{2}$$
$$f(x=2|y=1) = \frac{f(2,1)}{h(1)} = \frac{0}{4/9} = 0$$

then

x	0	1	2
$f(x\|1)$	$1/2$	$1/2$	0

Definition 24 (CONDITIONAL DENSITY.)

If $f(x, y)$ is the value of the joint density of the continuous random variables X and Y at (x, y) and $h(y)$ is the value of the marginal distribution of Y at y, the function given by

$$f(x|y) = \frac{f(x, y)}{h(y)}\ ; \quad h(y) \neq 0$$

for $x \in \mathbb{R}$, is called the conditional density of X given $Y = y$. Correspondingly, if $g(x)$ is the value of the marginal density of X at x, the function given by

$$w(y|x) = \frac{f(x, y)}{g(x)}\ ; \quad g(x) \neq 0$$

for $y \in \mathbb{R}$, is called the conditional density of Y given $X = x$.

Example 54

With reference to Example 52, find the conditional density of X given $Y = y$, and use it to evaluate $P\left(X \leq {}^1\!/_2 | Y = {}^1\!/_2\right)$.

Solution: $f(x|y) = \dfrac{f(x,y)}{h(y)} = \dfrac{3x(y+x)/5}{(3y+2)/10} = \dfrac{6x(x+y)}{3y+2}$ for $0 < x < 1$ and $f(x|y) = 0$ elsewhere. Now,

$$f\left(x \middle| \frac{1}{2}\right) = \frac{6}{7}x(2x+1)$$

and then

$$P\left(X \leq {}^1\!/_2 | Y = {}^1\!/_2\right) = \int_0^{1/2} \frac{6}{7}x(2x+1)\, dx = \frac{5}{28}$$

Exercise 16

(1) *Given the values of the joint probability distribution of X and Y shown in the table*

		$x = -1$	$x = 1$
y	-1	$1/8$	$1/2$
	0	0	$1/4$
	1	$1/8$	0

find:

a) *the marginal distribution of X;*
b) *the marginal distribution of Y;*
c) *the conditional distribution of X given $Y = -1$.*
d) *the conditional distribution of Y given $X = 1$.*

(2) If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{4}(2x+y) & : \quad \text{for} \;\; 0 < x < 1, \; 0 < y < 2 \\ 0 & : \quad \text{elsewhere} \end{cases}$$

a) Find the marginal density of X.
b) Find the conditional density of Y given $X = \frac{1}{4}$.
c) Find the marginal density of Y.

Current Subject

2 Probability Distributions and Probability Densities
- Discrete Random Variables and Probability Distributions
- Continuous Random Variables and Probability Density Functions
- Multivariate Distributions
- Marginal Distributions
- Conditional Distributions
- Independent Random Variables

Definition 25 (INDEPENDENCE OF RANDOM VARIABLES.)

If $f(x, y)$ is the value of the joint probability distribution (density) of the discrete (continuous) random variables X, Y at (x, y) and $g(x)$ is the value of the marginal distribution (density) of X and $h(y)$ is the value of the marginal distribution (density) of Y, then the random variables X and Y are independent if and only if

$$f(x, y) = g(x) \cdot h(y)$$

for all (x, y) within their range. If X and Y are not independent, they are said to be dependent.

Example 55

Check whether X and Y are independent if their joint probability distribution $f(x, y)$ is given by

$$\begin{array}{ccc|cc|} & & & \multicolumn{2}{c}{x} \\ & & & -1 & 1 \\ \cline{4-5} y & & -1 & 1/4 & 1/4 \\ & & 1 & 1/4 & 1/4 \\ \cline{4-5} \end{array}$$

Solution: Note that the probability distribution of X is

x	-1	1
$g(x)$	$1/2$	$1/2$

and the probability distribution of Y is

y	-1	1
$h(y)$	$1/2$	$1/2$

Since

$$f(-1,-1) = g(-1) \cdot h(-1)$$
$$f(-1,1) = g(-1) \cdot h(1)$$
$$f(1,-1) = g(1) \cdot h(-1)$$
$$f(1,1) = g(1) \cdot h(1)$$

then X, Y are independent.

Example 56

Check whether X and Y are independent if their joint probability distribution $f(x, y)$ is given by

$$\begin{array}{cc|cc|} & & \multicolumn{2}{c}{x} \\ & & 0 & 1 \\ \cline{3-4} \multirow{2}{*}{y} & 0 & 1/3 & 1/3 \\ & 1 & 1/3 & 0 \\ \cline{3-4} \end{array}$$

Solution: Note that the probability distribution of X is

x	0	1
$g(x)$	$2/3$	$1/3$

and the probability distribution of Y is

y	0	1
$h(y)$	$2/3$	$1/3$

Since, for example, $f(0, 0) \neq g(0) \cdot h(0)$ then X, Y are dependent.

Example 57

If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} 6xy^2 & : \text{ if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & : \text{ otherwise} \end{cases}$$

Show that X and Y are independent.

Solution: Since

$$g(x) = \begin{cases} \int_0^1 6xy^2 \, dy & : \text{ if } 0 \leq x \leq 1 \\ 0 & : \text{ otherwise} \end{cases}$$

$$= \begin{cases} 2x & : \text{ if } 0 \leq x \leq 1 \\ 0 & : \text{ otherwise} \end{cases}$$

$$h(y) = \begin{cases} \int_0^1 6xy^2 \, dx & : \text{ if } 0 \le y \le 1 \\ 0 & : \text{ otherwise} \end{cases}$$

$$= \begin{cases} 3y^2 & : \text{ if } 0 \le y \le 1 \\ 0 & : \text{ otherwise} \end{cases}$$

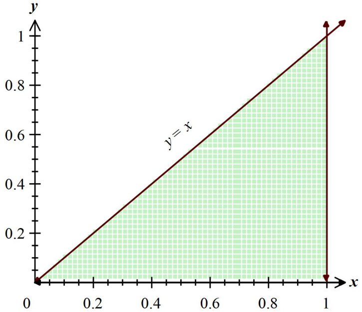
and $f(x, y) = g(x) \cdot h(y)$, then X, Y are independent.

Example 58

If the joint probability density of X and Y is given by

$$f(x,y)=\begin{cases}2 & : \text{ if } 0\le y\le x\le 1\\ 0 & : \text{ otherwise}\end{cases}$$

Show that X and Y are dependent.



Solution: Since

$$g(x) = \begin{cases} \int_0^x 2\ dy & : \quad \text{if } \ 0 \leq x \leq 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$

$$= \begin{cases} 2x & : \quad \text{if } \ 0 \leq x \leq 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$

$$h(y) = \begin{cases} \int_y^1 2\ dx & : \quad \text{if } \ 0 \leq y \leq 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$

$$= \begin{cases} 2(1-y) & : \quad \text{if } \ 0 \leq y \leq 1 \\ 0 & : \quad \text{otherwise} \end{cases}$$

and $f(x, y) \neq g(x) \cdot h(y)$, then X, Y are dependent.

Exercise 17

(1) *Check whether X and Y are independent or not if their joint probability distribution is given by*

$$\begin{array}{ccc|cc|} & & & \multicolumn{2}{c}{x} \\ & & & -1 & 1 \\ \cline{4-5} & & -1 & 1/8 & 1/2 \\ & y & 0 & 0 & 1/4 \\ & & 1 & 1/8 & 0 \\ \cline{4-5} \end{array}$$

(2) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} 24y(1-x-y) & : \quad \textit{for } x>0,\ y>0,\ x+y<1 \\ 0 & : \quad \textit{elsewhere} \end{cases}$$

Determine whether the two random variables are independent or not.

Current Subject

3. Mathematical Expectation
   - The Expected Value of a Random Variable
   - Moments and Moment-Generating Functions
   - Product Moments
   - Moments of Linear Combinations of Random Variables
   - Conditional Expectations

* Originally, the concept of a mathematical expectation arose in connection with games of chance.
* The mean for a sample or population was computed by adding the values and dividing by the total number of values, as shown in these formulas:

$$\overline{X} = \frac{\sum x}{n} \qquad \mu = \frac{\sum x}{N}$$

* Experiment [1]:
  How would you compute the mean of the number of spots that show on top when a die is rolled? You could try rolling the die, say, 10 times, recording the number of spots, and finding the mean; however, this answer would only approximate the true mean. What about 100 rolls or 1000 rolls?

Using *Mathematica*, we easily simulate the experiment:

```
rolls = 10^Range[6];
N@Total[RandomInteger[{1, 6}, #]]/# & /@ rolls
```

rolls	10^1	10^2	10^3	10^4	10^5	10^6
mean	3.8	3.67	3.487	3.5197	3.50123	3.50172

That is, if it were possible to roll the dice many times or an infinite number of times, the *average* of the number of spots would be 3.5.

* Experiment [2]:
Suppose two coins are tossed repeatedly, and the number of heads that occurred is recorded. What will be the mean of the number of heads? Using *Mathematica*, we easily simulate this experiment:

```
flips = 10^Range[6];
X = {0, 1, 2};
prob = {0.25, 0.50, 0.25};
N@Total[RandomChoice[prob -> X, #]]/# & /@ flips
```

flips	10^1	10^2	10^3	10^4	10^5	10^6
mean	1.3	0.96	0.965	1.0001	0.99503	1.00052

That is, if it were possible to toss the coins many times or an infinite number of times, the average of the number of heads would be 1.

The sample space of experiment (2) is HH, HT, TH, TT and each outcome has a probability of $1/4$. Now, in the long run, you would expect two heads (HH) to occur approximately $1/4$ of the time, one head to occur approximately $1/2$ of the time (HT or TH), and no heads (TT) to occur approximately $1/4$ of the time. Hence, on average, you would expect the number of heads to be

$$\frac{1}{4} \times 2 + \frac{1}{2} \times 1 + \frac{1}{4} \times 0 = 1.$$

* Hence, to find the mean (Expected Value) for a probability distribution, you must multiply each possible outcome by its corresponding probability and find the sum of the products.

Definition 26 (EXPECTED VALUE.)

If X is a discrete random variable and $f(x)$ is the value of its probability distribution at x, the expected value of X is

$$E(X) = \sum_{x} x \cdot f(x)$$

Correspondingly, if X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of X is

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Example 59

If X is the random variable corresponding to the number oh heads when we flip two coins, then the probability distribution of X is:

x	0	1	2
$f(x)$	$1/4$	$1/2$	$1/4$

So that, $E(x) = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1$

Example 60

If the probability density of X is given by

$$f(x) = \begin{cases} \dfrac{4}{\pi\left(1+x^2\right)} & : \ 0 < x < 1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

then

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\ &= \frac{4}{\pi}\int_{0}^{1} \frac{x}{1+x^2}\, dx = \frac{2}{\pi}\Big[\ln\left(1+x^2\right)\Big]_0^1 \\ &= \frac{\ln 4}{\pi} \end{aligned}$$

Theorem 25

If X is a discrete random variable and $f(x)$ is the value of its probability distribution at x, the expected value of $g(X)$ is given by

$$E\Big(g(X)\Big) = \sum_{x} g(x) \cdot f(x)$$

Correspondingly, if X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of $g(X)$ is given by

$$E\Big(g(X)\Big) = \int_{-\infty}^{\infty} g(x)f(x)\, dx$$

Example 61

If X is the number of points rolled with a balanced die, find the expected value of $g(X) = 2X^2 + 1$.

Solution: Since each possible outcome has the probability $\frac{1}{6}$, we get

$$E\Big(g(X)\Big) = \sum_{x=1}^{6} \Big(2x^2 + 1\Big) \cdot \frac{1}{6}$$
$$= \Big(2 \cdot 1^2 + 1\Big) \cdot \frac{1}{6} + \cdots + \Big(2 \cdot 6^2 + 1\Big) \cdot \frac{1}{6} = \frac{94}{3}$$

X	1	2	3	4	5	6
$2X^2+1$	3	9	19	33	51	73
$f(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Example 62

If X has the probability density $f(x) = \begin{cases} e^{-x} & : \ x > 0 \\ 0 & : \ \text{elsewhere} \end{cases}$, find the expected value of $g(X) = e^{3X/4}$.

Solution:

$$\begin{aligned} E\left(e^{3X/4}\right) &= \int_0^\infty e^{3x/4} \cdot e^{-x} \, dx \\ &= \int_0^\infty e^{-x/4} \, dx \\ &= -\frac{1}{4}\left[e^{-x/4}\right]_0^\infty = \frac{1}{4}. \end{aligned}$$

Theorem 26

If a and b are constants, then $E\Big(aX + b\Big) = aE\Big(X\Big) + b$.

Corollary 1

If a is a constant, then $E\Big(aX\Big) = aE\Big(X\Big)$.

Corollary 2

If b is a constant, then $E\Big(b\Big) = b$.

Theorem 27

If $c_1, c_2, \cdots, c_n$ are constants, then $E\left[\sum_{i=1}^{n} c_i g_i(x)\right] = \sum_{i=1}^{n} c_i E\left[g_i(x)\right]$.

Example 63

Let Y be a random variable with $f(y)$ given in the accompanying table. $E(Y)$, $E(Y^2)$, $E\left[(2Y-1)^2\right]$.

y	1	2	3	4
$f(y)$	0.4	0.3	0.2	0.1

Solution:

(1) $E(Y) = (1)(0.4) + (2)(0.3) + (3)(0.2) + (4)(0.1) = 2.$

(2) $E(Y^2) = (1)^2(0.4) + (2)^2(0.3) + (3)^2(0.2) + (4)^2(0.1) = 5.$

(3) $$E\left[(2Y-1)^2\right] = E\left(4Y^2 - 4Y + 1\right)$$
$$= 4E\left(Y^2\right) - 4E(Y) + 1 = 13$$

Example 64

If the probability density of X is given by

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

show that $E\left(X^r\right) = \dfrac{2}{(r+1)(r+2)}$, and use this result to evaluate $E\left[(3+X)^2\right]$.

Solution: $E\left(X^r\right) = \int_0^1 2x^r(1-x)\,dx = 2\int_0^1 \left(x^r - x^{r+1}\right)\,dx$

$$= 2\left[\frac{1}{r+1} - \frac{1}{r+2}\right] = \frac{2}{(r+1)(r+2)}$$

So, $E\left[(3+X)^2\right] = E\left(X^2 + 6X + 9\right) = E\left(X^2\right) + 6E(X) + 9 = \dfrac{67}{6}$

Theorem 28

If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the expected value of $g(X, Y)$ is

$$E\Big[g(X, Y)\Big] = \sum_x \sum_y g(x, y) \cdot f(x, y)$$

Correspondingly, if X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y), the expected value of $g(X, Y)$ is

$$E\Big[g(X, Y)\Big] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dxdy$$

Example 65

If the values of the joint probability distribution of X and Y are as shown in the table, find the expected value of $g(X, Y) = X + Y$.

		x = 0	1	2
y	0	$1/6$	$1/3$	$1/12$
	1	$2/9$	$1/6$	
	2	$1/36$		

Solution: $$E(X+Y) = \sum_{x=0}^{2}\sum_{y=0}^{2}(x+y)f(x,y)$$

$$= (0+0)\cdot\frac{1}{6} + (0+1)\cdot\frac{2}{9} + (0+2)\cdot\frac{1}{36}$$

$$+ (1+0)\cdot\frac{1}{3} + (1+1)\cdot\frac{1}{6} + (2+0)\cdot\frac{1}{12}$$

$$= 10/9$$

Example 66

If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{2}{7}(x+2y) & : \text{ for } 0 < x < 1,\ 1 < y < 2 \\ 0 & : \text{ elsewhere} \end{cases}$$

find the expected value of $g(X,Y) = \dfrac{X}{Y^3}$.

Solution: $$E\left(\frac{X}{Y^3}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x}{y^3} \cdot f(x,y)\, dxdy$$
$$= \int_1^2 \int_0^1 \frac{x}{y^3} \cdot \frac{2}{7}(x+2y)\, dxdy$$
$$= \int_1^2 \int_0^1 \left(\frac{2x^2}{7y^3} + \frac{4x}{7y^2}\right)\, dxdy = \frac{15}{84}$$

Exercise 18

(1) *Find the expected value of the discrete random variable X having the probability distribution*

$$f(x) = \frac{\left|x-2\right|}{7} \quad \textit{for} \quad x = -1, 0, 1, 3$$

Also, find $E\left(X^2\right)$, then find $E\left[X \cdot (2-5X)\right]$.

(2) *Find the expected value of the random variable Y whose probability density is given by*

$$f(y) = \begin{cases} \frac{1}{8}(y+1) & : \ 2 < y < 4 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

(3) *Find the expected value of the random variable X whose probability density is given by*

$$f(x) = \begin{cases} x & : \ 0 < x < 1 \\ 2 - x & : \ 1 \leq x < 2 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

(4) *If the values of the joint probability distribution of X and Y are as shown in the table, find: $E(X)$, $E(Y)$, $E(X \cdot Y)$.*

		x		
		0	1	2
	0	$^1/_9$	$^2/_9$	$^1/_9$
y	1	$^2/_9$	$^2/_9$	
	2	$^1/_9$		

(5) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} 6(1-y) & : \ 0 \le x \le y \le 1 \\ 0 & : \ elsewhere \end{cases}$$

find: $E(X)$, $E(Y)$, $E(X-3Y)$.

(6) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} 1 & : \ 0 \le x \le 2, \ 0 \le y \le 1, \ 2y \le x \\ 0 & : \ elsewhere \end{cases}$$

find: $E(X)$, $E(Y)$, $E(X+Y)$.

Current Subject

3. Mathematical Expectation
   - The Expected Value of a Random Variable
   - Moments and Moment-Generating Functions
   - Product Moments
   - Moments of Linear Combinations of Random Variables
   - Conditional Expectations

* An important class of discrete random variables is one in which X represents a count and consequently takes integer values: $X = 0, 1, 2, 3, \cdots$.

x_i	0	1	2	$\cdots$	n	$\cdots$
p_i	p_0	p_1	p_2	$\cdots$	p_n	$\cdots$

* A mathematical device useful in finding the probability distributions and other properties of integer-valued random variables is the moment-generating function.
* In statistics, the mathematical expectations called the moments of the distribution of a random variable or simply the moments of a random variable, are of special importance.

Definition 27 (MOMENTS ABOUT THE ORIGIN)

The rth moment about the origin of a random variable X, denoted by μ'_r, is the expected value of X^r; symbolically

$$\mu'_r = E\left[X^r\right] = \sum_x x^r f(x)$$

for $r = 0, 1, 2, \cdots$ when X is discrete, and

$$\mu'_r = E\left[X^r\right] = \int_{-\infty}^{\infty} x^r f(x)\, dx$$

when X is continuous.

Definition 28 (MEAN OF A DISTRIBUTION)

μ'_1 is called the mean of the distribution of X, or simply the mean of X, and it is denoted simply by μ.

Definition 29 (MOMENTS ABOUT THE MEAN)

The rth moment about the mean of a random variable X, denoted by μ_r, is the expected value of $(X-\mu)^r$, symbolically

$$\mu_r = E\left[(X-\mu)^r\right] = \sum_x (x-\mu)^r f(x)$$

for $r = 0, 1, 2, \cdots$ when X is discrete, and

$$\mu_r = E\left[(X-\mu)^r\right] = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\, dx$$

when X is continuous.

Definition 30 (VARIANCE)

μ_2 is called the variance of the distribution of X, or simply the variance of X, and it is denoted by σ^2, σ_X^2, $var(X)$, or $V(X)$. The standard deviation of X is $\sigma = \sqrt{\sigma^2}$.

Note: The parameters μ and σ are meaningful numerical descriptive measures that locate the center and describe the spread associated with the values of a random variable X. They do not, however, provide a unique characterization of the distribution of X.

(1) $\mu_1' = \mu$ is the mean of the distribution of X,

(2) μ_2 is the variance of the distribution of X,

(3) μ_3 is used to construct a measure of skewness, which describes whether the probability mass is more to the left or the right of the mean, compared to a normal distribution.

(4) μ_4 is used to construct a measure of kurtosis, which measures the "width" of a distribution.

Theorem 29

$$\sigma^2 = \mu_2' - \mu^2$$

Proof.

$$\begin{aligned}\sigma^2 &= \mu_2 = E\left[(X-\mu)^2\right] \\ &= E\left[X^2 - 2\mu X + \mu^2\right] = \underbrace{E\left[X^2\right]}_{\mu_2'} - 2\mu \underbrace{E\left[X\right]}_{\mu} + \mu^2 = \mu_2' - \mu^2\end{aligned}$$

$\square$

Theorem 30

If X has the variance σ^2, then $var\left(aX + b\right) = a^2\sigma^2$.

Example 67

A fair coin is tossed three times. Let X be the random variable defined by $X = \begin{bmatrix} \text{number} \\ \text{of heads} \end{bmatrix} - \begin{bmatrix} \text{number} \\ \text{of tails} \end{bmatrix}$. Find the mean and the variance of the distribution of X.

Solution: The probability distribution of X is given by:

X	-3	-1	1	3
$P(X)$	$1/8$	$3/8$	$3/8$	$1/8$

Hence,

$$\mu = E\left[X\right] = \left(-3 \times \frac{1}{8}\right) + \left(-1 \times \frac{3}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(3 \times \frac{1}{8}\right) = 0$$

$$\mu_2' = E\left[X^2\right] = \left(9 \times \frac{1}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(9 \times \frac{1}{8}\right) = 3$$

$$\therefore\ \sigma^2 = \mu_2' - \mu = 3 - 0 = 3.$$

Example 68

Find $var(X)$ if the probability density of X is given by

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

Solution: We have shown, in Example 64, that $\mu'_r = E\left(X^r\right) = \dfrac{2}{(r+1)(r+2)}$. So,

$$\begin{aligned} var(X) &= \mu'_2 - \mu^2 = E\left[X^2\right] - E\left[X\right]^2 \\ &= \frac{1}{6} - \frac{1}{9} = \frac{1}{18}. \end{aligned}$$

Although the moments of most distributions can be determined directly by evaluating the necessary integrals or sums, an alternative procedure sometimes provides considerable simplifications.

Definition 31 (MOMENT GENERATING FUNCTION)

The moment generating function of a random variable X, where it exists, is given by

$$M_X(t) = E\left[e^{tX}\right] = \sum_x e^{tx} f(x)$$

when X is discrete, and

$$M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

when X is continuous.

To explain why we refer to this function as a "moment-generating" function, let us substitute for e^{tx} its Maclaurin's series expansion, that is,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^rx^r}{r!} + \cdots$$

For the discrete case (*in the continuous case, the argument is the same*), we thus get

$$\begin{aligned} M_X(t) &= \sum_x \left[1 + tx + \frac{t^2x^2}{2!} + \cdots + \frac{t^rx^r}{r!} + \cdots\right] f(x) \\ &= \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \cdots + \frac{t^r}{r!}\sum_x x^r f(x) + \cdots \\ &= 1 + \mu \cdot t + \mu_2' \cdot \frac{t^2}{2!} + \cdots + \mu_r' \cdot \frac{t^r}{r!} + \cdots \end{aligned}$$

Example 69

Find the moment-generating function of the random variable whose probability density is given by $f(x) = \begin{cases} e^{-x} & : \ x > 0 \\ 0 & : \ \text{elsewhere} \end{cases}$ and use it to find an expression for μ_r'.

Solution: $M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$

$$= \int_0^{\infty} e^{x(t-1)}\, dx = \frac{1}{t-1} e^{x(t-1)}\Big]_0^{\infty} = \frac{1}{1-t}$$

for $t < 1$. As is well known, when $|t| < 1$, the Maclaurin's series for this moment-generating function is

$$M_X(t) = 1 + t + t^2 + \cdots + t^r + \cdots$$
$$= 1 + t + 2! \cdot \frac{t^2}{2!} + \cdots + r! \cdot \frac{t^r}{r!} + \cdots$$

and hence $\mu_r' = r!$ for $r = 0, 1, 2, \cdots$.

Theorem 31

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'$$

Example 70

Given that X has the probability distribution $f(x) = \frac{1}{8}\binom{3}{x}$ for $x = 0, 1, 2, 3$. Find the moment-generating function of this random variable and use it to determine μ_1' and μ_2'.

Solution: Since $M_X(t) = E\left[e^{Xt}\right] = \frac{1}{8}\sum_{x=0}^{3} e^{xt}\binom{3}{x}$

$$= \frac{1}{8}\left(1 + 3e^t + 3e^{2t} + e^{3t}\right) = \frac{1}{8}\left(1 + e^t\right)^3$$

then

$$\mu_1' = M_X'(0) = \left[\frac{3}{8}\left(1+e^t\right)^2 e^t\right]_{t=0} = \frac{3}{2}$$

$$\mu_2' = M_X''(0) = \left[\frac{3}{4}\left(1+e^t\right)^2 e^{2t} + \frac{3}{8}\left(1+e^t\right)^2 e^t\right]_{t=0} = 3$$

Theorem 32

If a and b are constants, then

(1) $M_{X+a} = E\left[e^{(X+a)t}\right] = e^{at} M_X(t)$,

(2) $M_{bX} = E\left[e^{bXt}\right] = M_X(bt)$,

(3) $M_{\frac{X+a}{b}} = E\left[e^{\left(\frac{X+a}{b}\right)t}\right] = e^{\frac{a}{b}t} M_X\left(\frac{t}{b}\right)$.

Exercise 19

(1) *Show that* $\mu_0 = 1$ *and that* $\mu_1 = 0$ *for any random variable for which* $E(X)$ *exists.*

(2) *Find* μ, μ_2', *and* σ^2 *for the random variable* X *that has the probability distribution* $f(x) = \frac{1}{2}$ *for* $x = -2$ *and* $x = 2$.

(3) *Find* μ, μ_2', *and* σ^2 *for the random variable* X *that has the probability density*

$$f(x) = \begin{cases} \frac{x}{2} & : \ 0 < x < 2 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

(4) *Find μ_r' and σ^2 for the random variable X that has the probability density*

$$f(x) = \begin{cases} \dfrac{1}{x \ln 3} & : \ 1 < x < 3 \\ 0 & : \ elsewhere \end{cases}$$

(5) *If the probability density of X is given by*

$$f(x) = \begin{cases} 2x^{-3} & : \ x > 1 \\ 0 & : \ elsewhere \end{cases}$$

check whether its mean and its variance exist.

(6) *Find the moment-generating function of the continuous random variable X whose probability density is given by*

$$f(x) = \begin{cases} 1 & : \ 0 < x < 1 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

and use it to find μ_1', μ_2', and σ^2.

(7) *Find the moment-generating function of the discrete random variable X that has the probability distribution*

$$f(x) = 2\left(\frac{1}{3}\right)^x \quad \textit{for} \ \ x = 1, 2, 3, \cdots$$

and use it to find μ_1', μ_2', and σ^2.

(8) *Explain why there can be no random variable for which* $M_X(t) = \dfrac{t}{1-t}$.

(9) *Given the moment-generating function* $M_X(t) = e^{3t+8t^2}$, *find the moment-generating function of the random variable* $Y = \dfrac{1}{4}\Big(X - 3\Big)$, *and use it to determine the mean and the variance of* Y.

Current Subject

3 Mathematical Expectation
- The Expected Value of a Random Variable
- Moments and Moment-Generating Functions
- Product Moments
- Moments of Linear Combinations of Random Variables
- Conditional Expectations

Definition 32 (PRODUCT MOMENTS ABOUT THE ORIGIN.)

The rth and sth product moment about the origin of the random variables X and Y, denoted by $\mu'_{r,s}$, is the expected value of $X^r Y^s$; symbolically,

$$\mu'_{r,s} = E[X^r Y^s] = \sum_x \sum_y x^r y^s f(x, y)$$

for $r = 0, 1, 2, \cdots$ and $s = 0, 1, 2, \cdots$ when X and Y are discrete, and

$$\mu'_{r,s} = E[X^r Y^s] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x, y)\, dxdy$$

when X and Y are continuous.

Definition 33 (Product Moments About the Mean.)

The rth and sth product moment about the means of the random variables X and Y, denoted by $\mu_{r,s}$, is the expected value of $(X - \mu_X)^r (Y - \mu_Y)^s$; symbolically,

$$\begin{aligned}\mu_{r,s} &= E\left[(X - \mu_X)^r (Y - \mu_Y)^s\right] \\ &= \sum_x \sum_y (x - \mu_X)^r (y - \mu_Y)^s f(x, y)\end{aligned}$$

for $r = 0, 1, 2, \cdots$ and $s = 0, 1, 2, \cdots$ when X and Y are discrete, and

$$\begin{aligned}\mu_{r,s} &= E\left[(X - \mu_X)^r (Y - \mu_Y)^s\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y)\, dxdy\end{aligned}$$

when X and Y are continuous.

In statistics, $\mu_{1,1}$ is of special importance because it is indicative of the relationship, if any, between the values of X and Y; thus, it is given a special symbol and a special name.

Definition 34 (COVARIANCE.)

$\mu_{1,1}$ is called the covariance of X and Y, and it is denoted by σ_{XY}, $cov(X, Y)$, or $C(X, Y)$.

Theorem 33

$$\sigma_{XY} = \mu'_{1,1} - \mu_X \mu_Y$$

Proof.

Homework. □

Example 71

The joint and marginal probabilities of X and Y are recorded as follows:

		$x=0$	$x=1$	$x=2$	
y	0	$^1/_6$	$^1/_3$	$^1/_{12}$	$^7/_{12}$
	1	$^2/_9$	$^1/_6$		$^7/_{18}$
	2	$^1/_{36}$			$^1/_{36}$
		$^5/_{12}$	$^1/_2$	$^1/_{12}$	

Find the covariance of X and Y.

Solution: Referring to the joint probabilities given here, we get

$$
\begin{aligned}
\mu'_{1,1} &= E\Big[XY\Big] \\
&= 0 \cdot 0 \cdot \frac{1}{6} + 0 \cdot 1 \cdot \frac{2}{9} + 0 \cdot 2 \cdot \frac{1}{36} + 1 \cdot 0 \cdot \frac{1}{3} + 1 \cdot 1 \cdot \frac{1}{6} + 2 \cdot 0 \cdot \frac{1}{12} \\
&= \frac{1}{6}
\end{aligned}
$$

and using the marginal probabilities, we get

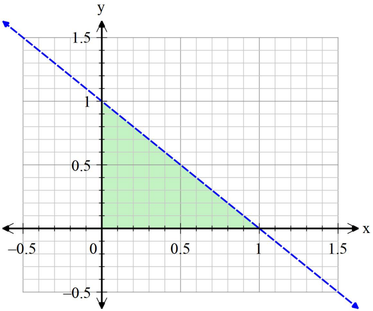
$$
\begin{aligned}
\mu_X &= 0 \cdot \frac{5}{12} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{12} = \frac{2}{3} \\
\mu_Y &= 0 \cdot \frac{7}{12} + 1 \cdot \frac{7}{18} + 2 \cdot \frac{1}{36} = \frac{4}{9}
\end{aligned}
$$

It follows that $\sigma_{XY} = \mu'_{1,1} - \mu_X \mu_Y = \frac{1}{6} - \frac{2}{3} \times \frac{4}{9} = -\frac{7}{54}$.

Example 72

Find the covariance of the random variables whose joint probability density is given by

$$f(x,y) = \begin{cases} 2 & : \ x > 0, \ y > 0, \ x + y < 1 \\ 0 & : \ \text{elsewhere} \end{cases}$$



Solution:

$$\mu_X = E\Big[X\Big] = \int_0^1 \int_0^{1-x} 2x \; dydx = \frac{1}{3}$$
$$\mu_Y = E\Big[Y\Big] = \int_0^1 \int_0^{1-x} 2y \; dydx = \frac{1}{3}$$
$$\mu'_{1,1} = E\Big[XY\Big] = \int_0^1 \int_0^{1-x} 2xy \; dydx = \frac{1}{12}$$
$$\therefore \; \sigma_{XY} = \mu'_{1,1} - \mu_X\mu_Y = \frac{1}{12} - \frac{1}{3} \times \frac{1}{3} = -\frac{1}{36}$$

Theorem 34

If X and Y are independent, then $E\Big[XY\Big] = E\Big[X\Big]E\Big[Y\Big]$ and $\sigma_{XY} = 0$.

Example 73

If the joint probability distribution of X and Y is given by

$$
\begin{array}{cc|ccc|c}
 & & & x & & \\
 & & -1 & 0 & 1 & \\
\hline
 & -1 & 1/6 & 1/3 & 1/6 & 2/3 \\
y & 0 & 0 & 0 & 0 & 0 \\
 & 1 & 1/6 & 0 & 1/6 & 1/3 \\
\hline
 & & 1/3 & 1/3 & 1/3 &
\end{array}
$$

show that their covariance is zero even though the two random variables are not independent.

Solution: Using the probabilities shown in the margins, we get

$$\mu_X = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$\mu_Y = -1 \cdot \frac{2}{3} + 0 \cdot 0 + 1 \cdot \frac{1}{3} = -\frac{1}{3}$$

$$\begin{aligned}\mu'_{1,1} &= (-1) \cdot (-1) \cdot \frac{1}{6} + (-1) \cdot (1) \cdot \frac{1}{6} + (1) \cdot (-1) \cdot \frac{1}{6} + (1) \cdot (1) \cdot \frac{1}{6} \\ &= 0\end{aligned}$$

Thus,$\sigma_{XY} = 0 - 0 \cdot {}^{-1}\!/_{3} = 0$, the covariance is zero, but the two random variables are not independent. For instance, $f(-1,-1) \neq g(-1)h(-1)$.

Product moments can also be defined for the case where there are more than two random variables. Here let us merely state the important result, in the following theorem.

Theorem 35

If $X_1, X_2, \cdots, X_n$ are independent, then

$$E\left[X_1 X_2 \cdots X_n\right] = E\left[X_1\right] E\left[X_2\right] \cdots E\left[X_n\right]$$

Exercise 20

(1) *If X and Y have the joint probability distribution*

		$x = -1$	$x = 0$	$x = 1$
y	0	0	$1/6$	$1/12$
	1	$1/4$	0	$1/2$

show that:

(a) $cov(X, Y) = 0$,

(b) *the two random variables are not independent.*

(2) *If the probability density of* X *is given by*

$$f(x) = \begin{cases} 1+x & : \ -1 < x \leq 0 \\ 1-x & : \ 0 < x < 1 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

and $U = X$ *and* $V = X^2$*, show that*

(a) $cov(U, V) = 0$,

(b) U *and* V *are dependent.*

Current Subject

3 Mathematical Expectation

In this section we shall derive expressions for the mean and the variance of a linear combination of n random variables and the covariance of two linear combinations of n random variables.

Theorem 36

If $X_1, X_2, \cdots, X_n$ are random variables, and $Y = \sum_{i=1}^{n} a_i X_i$ where $a_1, a_2, \cdots, a_n$ are constants, then

$$E\left[Y\right] = \sum_{i=1}^{n} a_i E\left[X_i\right] = \sum_{i=1}^{n} a_i \mu_{X_i}$$

$$\text{and} \quad var\left(Y\right) = \sum_{i=1}^{n} a_i^2 var\left(X_i\right) + 2 \sum_{i<j}\sum a_i a_j cov\left(X_i, X_j\right)$$

where the double summation extends over all values of i and j, from 1 to n, for which $i < j$.

Since $cov\left(X_i, X_j\right) = 0$ when X_i and X_j are independent, we obtain the following corollary.

Corollary 3

If the random variables $X_1, X_2, \cdots, X_n$ are independent and $Y = \sum_{i=1}^{n} a_i X_i$, then

$$var\left(Y\right) = \sum_{i=1}^{n} a_i^2 var\left(X_i\right)$$

Example 74

If the random variables X, Y, and Z have the

$$\begin{aligned} \text{means} &: \mu_X = 2, \mu_Y = -3, \mu_Z = 4, \\ \text{variances} &: \sigma_X^2 = 1, \sigma_Y^2 = 5, \sigma_Z^2 = 2, \\ \text{covariances} &: cov(X, Y) = -2, cov(X, Z) = -1, cov(Y, Z) = 1, \end{aligned}$$

find the mean and the variance of $W = 3X - Y + 2Z$.

Solution:

$$\begin{aligned} E\left[W\right] &= 3\mu_X - \mu_Y + 2\mu_Z = 17, \\ var\left(W\right) &= 3^2 \cdot \sigma_X^2 + (-1)^2 \cdot \sigma_Y^2 + 2^2 \cdot \sigma_Z^2 \\ &\quad + 2(3)(-1)\sigma_{XY} + 2(3)(2)\sigma_{XZ} + 2(-1)(2)\sigma_{YZ} \\ &= 18 \end{aligned}$$

Theorem 37

If $X_1, X_2, \cdots, X_n$ are random variables and $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$ where $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ are constants, then

$$cov\left(Y_1, Y_2\right) = \sum_{i=1}^{n} a_i b_j var\left(X_i\right) + \sum_{i<j}\sum \left(a_i b_j + a_j b_i\right) \cdot cov\left(X_i, X_j\right).$$

Corollary 4

If the random variables $X_1, X_2, \cdots, X_n$ are independent, and $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$, then
$cov\left(Y_1, Y_2\right) = \sum_{i=1}^{n} a_i b_j var\left(X_i\right)$.

Example 75

If the random variables X, Y, and Z have the

$$\begin{aligned} \text{means} &: \mu_X = 3, \mu_Y = 5, \mu_Z = 2, \\ \text{variances} &: \sigma_X^2 = 8, \sigma_Y^2 = 12, \sigma_Z^2 = 18, \\ \text{covariances} &: cov(X, Y) = 1, cov(X, Z) = -3, cov(Y, Z) = 2, \end{aligned}$$

find the covariance of $U = X + 4Y + 2Z$ and $V = 3X - Y - Z$.

Solution:

$$\begin{aligned} cov\left(U, V\right) &= cov\Big(X + 4Y + 2Z, 3X - Y - Z\Big) \\ &= 3\sigma_X^2 - 4\sigma_Y^2 - 2\sigma_Z^2 + 11cov(X, Y) \\ &\quad + 5cov(X, Z) - 6cov(Y, Z) \\ &= -76 \end{aligned}$$

Exercise 21

(1) *If* X_1, X_2, *and* X_3 *are random variables have the means* 4, 9, *and* 3 *and the variances* 3, 7, *and* 5, *and covariances* $cov(X_1, X_2) = 1$, $cov(X_1, X_3) = -3$, $cov(X_2, X_3) = -2$, *find the mean and the variance of* $Y = 2X_1 - 3X_2 + 4X_3$.

(2) *If* $var(X_1) = 5$, $var(X_2) = 4$, $var(X_3) = 7$, $cov(X_1, X_2) = 3$, $cov(X_1, X_3) = -2$, *and* X_2 *and* X_3 *are independent, find the covariance of* $Y_1 = X_1 - 2X_2 + 3X_3$ *and* $Y_2 = -2X_1 + 3X_2 + 4X_3$.

(3) *If the joint probability density of* X *and* Y *is given by*

$$f(x, y) = \begin{cases} \frac{1}{3}\left(x + y\right) & : \ 0 < x < 1, 0 < y < 2 \\ 0 & : \ elsewhere \end{cases}$$

find the variance of $W = 3X + 4Y - 5$.

Current Subject

3. Mathematical Expectation
   - The Expected Value of a Random Variable
   - Moments and Moment-Generating Functions
   - Product Moments
   - Moments of Linear Combinations of Random Variables
   - Conditional Expectations

Definition 35 (CONDITIONAL EXPECTATION)

If X is a discrete random variable, and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at x, the conditional expectation of $u(X)$ given $Y = y$ is

$$E\Big[u(X)|y\Big] = \sum_{x} u(x)f(x|y)$$

Correspondingly, if X is a continuous variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at x, the conditional expectation of $u(X)$ given $Y = y$ is

$$E\Big[u(X)|y\Big] = \int_{-\infty}^{\infty} u(x)f(x|y)\,dx$$

Note:

(1) Similar expressions based on the conditional probability distribution or density of Y given $X = x$ define the conditional expectation of $v(Y)$ given $X = x$.

(2) If we let $u(X) = X$ in Definition 35, we obtain the conditional mean of the random variable X given $Y = y$, which we denote by $\mu_{X|y} = E\Big[X|y\Big]$.

(3) Correspondingly, the conditional variance of X given $Y = y$ is

$$\sigma^2_{X|y} = E\left[(X - \mu_{X|y})^2|y\right] = E\left[X^2|y\right] - \mu^2_{X|y}$$

where $E\left[X^2|y\right]$ is given by Definition 35 with $u(X) = X^2$.

Example 76

If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{2}{3}\Big(x+2y\Big) & : \ 0<x<1, 0<y<1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

find the conditional mean and the conditional variance of X given $Y=\frac{1}{2}$.

Solution: Since the marginal density of Y is

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} \displaystyle\int_0^1 \frac{2}{3}\Big(x+2y\Big)\,dx & : \ 0<y<1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

$$= \begin{cases} \dfrac{1}{3}\Big(1+4y\Big) & : \ 0<y<1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

then
$$f(x|y) = \frac{f(x,y)}{h(y)} = \begin{cases} \dfrac{2x+4y}{1+4y} & : \ 0 < x < 1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

$$\therefore \quad f\left(x \,\middle|\, \frac{1}{2}\right) = \begin{cases} \dfrac{2}{3}(x+1) & : \ 0 < x < 1 \\ 0 & : \ \text{elsewhere} \end{cases}$$

Thus,
$$\mu_{X|\frac{1}{2}} = E\left[X \,\middle|\, \frac{1}{2}\right] = \int_0^1 \frac{2}{3}x(x+1)\,dx = \frac{5}{9}, \qquad .$$

$$E\left[X^2 \,\middle|\, \frac{1}{2}\right] = \int_0^1 \frac{2}{3}x^2(x+1)\,dx = \frac{7}{18},$$

$$\sigma^2_{X|\frac{1}{2}} = E\left[X^2 \,\middle|\, \frac{1}{2}\right] - \mu^2_{X|\frac{1}{2}} = \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162}$$

Exercise 22

(1) *If the joint probability density of X and Y is given by*

$$f(x,y) = \begin{cases} \frac{1}{3}\left(x+y\right) & : \ 0 < x < 1, 0 < y < 2 \\ 0 & : \ \textit{elsewhere} \end{cases}$$

find the conditional mean and the conditional variance of Y given $X = \frac{3}{4}$.

(2) *If X and Y have the joint probability distribution*

		$x = -1$	$x = 0$	$x = 1$
y	0	0	$1/6$	$1/12$
	1	$1/4$	0	$1/2$

find the conditional mean and the conditional variance of X given $Y = 1$.

Current Subject

If a random variable can take on k different values with equal probability, we say that it has a discrete uniform distribution.

Definition 36 (DISCRETE UNIFORM DISTRIBUTION.)

A random variable X has a discrete uniform distribution and it is referred to as a discrete uniform random variable if and only if its probability distribution is given by

$$f(x) = \frac{1}{k} \quad \text{for} \quad x = x_1, x_2, \cdots, x_k$$

where $x_i \neq x_j$ when $i \neq j$.

Exercise 23

If X has the discrete uniform distribution $f(x) = \frac{1}{k}$ for $x = 1, 2, \cdots, k$, show that:

(a) *its mean is $\mu = \frac{k+1}{2}$;*

(b) *its variance is $\sigma^2 = \frac{k^2-1}{12}$;*

(c) *moment-generating function is $M_X(t) = \frac{e^t\left(1-e^{kt}\right)}{k\left(1-e^t\right)}$.*

Current Subject

If an experiment has two possible outcomes, "success" and "failure," and their probabilities are, respectively, θ and $1 - \theta$, then the number of successes, 0 or 1, has a Bernoulli distribution.

Definition 37 (BERNOULLI DISTRIBUTION)

A random variable X has a Bernoulli distribution and it is referred to as a Bernoulli random variable if and only if its probability distribution is given by

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} \quad \text{for } x = 0, 1$$

Definition 38 (BINOMIAL DISTRIBUTION.)

A random variable X has a binomial distribution and it is referred to as a binomial random variable if and only if its probability distribution is given by

$$b\left(x; n, \theta\right) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{for } x = 0, 1, \cdots, n$$

Example 77

Find the probability of getting five heads in 12 flips of a balanced coin.

Solution: $b\left(5; 12, {}^1\!/_2\right) = \binom{12}{5}\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^{12-5} = \frac{99}{512}.$

Theorem 38

$$b\left(x; n, \theta\right) = b\left(n - x; n, 1 - \theta\right)$$

Theorem 39

The mean and the variance of the binomial distribution are

$$\mu = n\theta \quad \textit{and} \quad \sigma^2 = n\theta(1 - \theta)$$

Theorem 40

The moment-generating function of the binomial distribution is given by

$$M_X(t) = \left[1 + \theta\left(e^t - 1\right)\right]^n$$

Current Subject

In connection with repeated Bernoulli trials, we are sometimes interested in the number of the trial on which the kth success occurs.

Definition 39 (NEGATIVE BINOMIAL DISTRIBUTION.)

A random variable X has a negative binomial distribution and it is referred to as a negative binomial random variable if and only if

$$b^*\left(x; k, \theta\right) = \binom{x-1}{k-1}\theta^k(1-\theta)^{x-k} \quad \text{for } x = k, k+1, k+2, \cdots$$

Example 78

If the probability is 0.40 that a child exposed to a certain contagious disease will catch it, what is the probability that the tenth child exposed to the disease will be the third to catch it?

Solution: $b^*\Big(10; 3, 0.4\Big) = \binom{9}{2}(0.4)^3(0.6)^7 = 0.0645.$

Theorem 41

$$b^*\Big(x; k, \theta\Big) = \frac{k}{x} \cdot b\Big(k; x, \theta\Big)$$

Theorem 42

The mean and the variance of the negative binomial distribution are

$$\mu = \frac{k}{\theta} \quad \textit{and} \quad \sigma^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right)$$

Definition 40 (GEOMETRIC DISTRIBUTION.)

A random variable X has a geometric distribution and it is referred to as a geometric random variable if and only if its probability distribution is given by

$$g(x;\theta) = \theta(1-\theta)^{x-1} \quad \text{for} \quad x = 1, 2, 3, \cdots$$

Theorem 43

The moment-generating function of the geometric distribution is given by

$$M_X(t) = \frac{\theta e^t}{1 - e^t(1-\theta)}$$

Example 79

If the probability is 0.75 that an applicant for a driver's license will pass the road test on any given try, what is the probability that an applicant will finally pass the test on the fourth try?

Solution: $g(4; 0.75) = 0.75 \times (0.25)^3 = 0.0117.$

Current Subject

4 Special Probability Distributions and Densities
- The Discrete Uniform Distribution
- The Binomial Distribution
- The Negative Binomial and Geometric Distributions
- The Poisson Distribution
- The Exponential Distribution

When n is large, the calculation of binomial probabilities with the formula of Definition 38 will usually involve a prohibitive amount of work. In this section we shall present a probability distribution that can be used to approximate binomial probabilities of this kind. Specifically, we shall investigate the limiting form of the binomial distribution when $n \to \infty$, $\theta \to 0$, while $n\theta$ remains constant. Letting this constant be λ, that is, $n\theta = \lambda$ and, hence, $\theta = \dfrac{\lambda}{n}$.

Definition 41 (POISSON DISTRIBUTION.)

A random variable has a Poisson distribution and it is referred to as a Poisson random variable if and only if its probability distribution is given by

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for} \quad x = 0, 1, 2, \cdots$$

Note: In general, the Poisson distribution will provide a good approximation to binomial probabilities when $n \geq 20$ and $\theta \leq 0.05$. When $n \geq 100$ and $n\theta < 10$, the approximation will generally be excellent.

Example 80

If 2 percent of the books bound at a certain bindery have defective bindings, use the Poisson approximation to the binomial distribution to determine the probability that 5 of 400 books bound by this bindery will have defective bindings.

Solution: Since $x = 5$ and $\lambda = 400 \times 0.02 = 8$, then

$$p(5;8) = \frac{8^5 \times e^{-8}}{5!} \approx 0.093$$

Theorem 44

The mean and the variance of the Poisson distribution are given by

$$\mu = \lambda \quad \textit{and} \quad \sigma^2 = \lambda$$

Theorem 45

The moment-generating function of the Poisson distribution is given by

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Current Subject

4 Special Probability Distributions and Densities
- The Discrete Uniform Distribution
- The Binomial Distribution
- The Negative Binomial and Geometric Distributions
- The Poisson Distribution
- The Exponential Distribution

Definition 42 (EXPONENTIAL DISTRIBUTION.)

A random variable X has an exponential distribution and it is referred to as an exponential random variable if and only if its probability density is given by

$$g(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & : \ x > 0 \\ 0 & : \ \text{elsewhere} \end{cases}$$

where $\theta > 0$.

Note: The exponential distribution applies not only to the occurrence of the first success in a Poisson process but it applies also to the waiting times between successes.

Theorem 46

The mean and the variance of the exponential distribution are given by

$$\mu = \theta \text{ and } \sigma^2 = \theta^2$$