Slides for Sessional II

Conditional Probabilities

Suppose that one rolls a pair of dice. The sample space S of this experiment can be taken to be the following set of 36 outcomes.

$$S = \{(i, j), i = 1, 2, 3, 4, 5, 6, j = 1, 2, 3, 4, 5, 6\}$$

Suppose further that we observe that the first die lands on side 3. Then, given this information, what is the probability that the sum of the two dice equals 8?

So let E is the event that the sum of the dice is 8 and F is the event that the first die is a 3, then the probability of E given that F has occurred is called the conditional probability of E given that F has occurred, and is denoted by **P(E | F).**

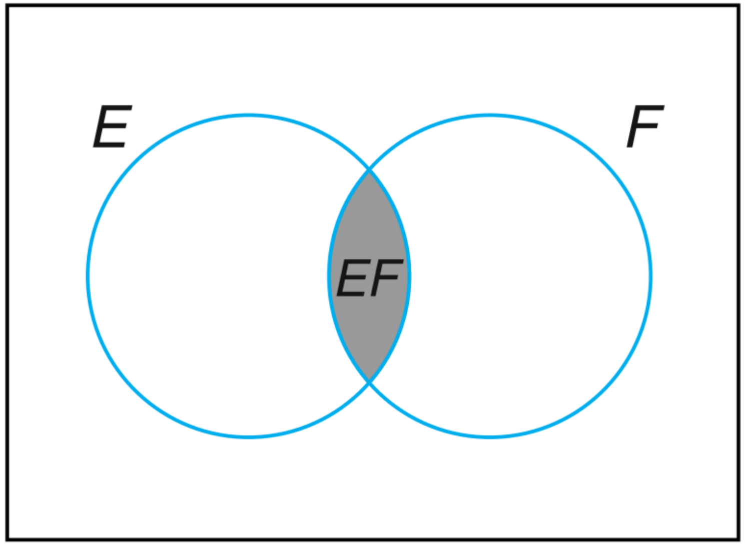
$F = \{$ first die is a 3 $\}$

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$E = \{$ sum of the dice equals 8 $\}$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Note that above Equation is well defined only when $P(F) > 0$ and hence $P(E|F)$ is defined only when $P(F) > 0$.



Problem

A bin contains 5 defective (that immediately fail when put in use), 10 partially defective (that fail after a couple of hours of use), and 25 acceptable transistors. A transistor is chosen at random from the bin and put into use. If it does not immediately fail, what is the probability it is acceptable?

Solution

Since the transistor did not immediately fail, we know that it is not one of the 5 defectives and so the desired probability is:

$$P\{\text{acceptable} \mid \text{not defective}\}$$

$$= \frac{P\{\text{acceptable, not defective}\}}{P\{\text{not defective}\}}$$

$$= \frac{P\{\text{acceptable}\}}{P\{\text{not defective}\}}$$

where the last equality follows since the transistor will be both acceptable and not defective if it is acceptable.

$$P\{\text{acceptable}|\text{not defective}\} = \frac{25/40}{35/40} = 5/7$$

Problem

The organization that Jones works for is running a father–son dinner for those employees having at least one son. Each of these employees is invited to attend along with his youngest son. If Jones is known to have two children, what is the conditional probability that they are both boys given that he is invited to the dinner? Assume that the sample space S is given by S = {(b, b), (b, g), (g, b), (g, g)} and all outcomes are equally likely [(b, g) means, for instance, that the younger child is a boy and the older child is a girl].

SOLUTION The knowledge that Jones has been invited to the dinner is equivalent to knowing that he has at least one son. Hence, letting B denote the event that both children are boys, and A the event that at least one of them is a boy, we have that the desired probability $P(B|A)$ is given by

$$\begin{aligned} P(B|A) &= \frac{P(BA)}{P(A)} \\ &= \frac{P(\{(b,b)\})}{P(\{(b,b),(b,g),(g,b)\})} \\ &= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned}$$

Problem

Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart on time and arrive on time.

(a) Jhon is meeting Ana's flight, which departed on time. What is the probability that Ana will arrive on time?

(b) Jhon has met Ana, and she arrived on time. What is the probability that her flight departed on time?

Solution. Denote the events,

$$\begin{aligned} A &= \{\text{arriving on time}\}, \\ D &= \{\text{departing on time}\}. \end{aligned}$$

We have:

$$\boldsymbol{P}\{A\} = 0.8, \quad \boldsymbol{P}\{D\} = 0.9, \quad \boldsymbol{P}\{A \cap D\} = 0.75.$$

(a) $\boldsymbol{P}\{A \mid D\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{D\}} = \dfrac{0.75}{0.9} = \underline{0.8333}.$

(b) $\boldsymbol{P}\{D \mid A\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{A\}} = \dfrac{0.75}{0.8} = \underline{0.9375}.$

Problem

The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Solution

Consider the events

$$L\text{: length defective,} \qquad T\text{: texture defective.}$$

Given that the strip is length defective, the probability that this strip is texture defective is given by

$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08.$$

$$P(A|B) = \frac{P(AB)}{P(B)} \Rightarrow \boxed{P(AB) = P(A|B)\,P(B)}$$

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

$$\Rightarrow P(B|A) = \frac{P(AB)}{P(A)} \Rightarrow \boxed{P(AB) = P(B|A)\,P(A)}$$

Note:

1) $P(AB) = P(A|B)\,P(B)$

2) $P(AB) = P(B|A)\,P(A)$

Problem

Suppose that two balls are to be selected at random, without replacement, from a box containing r red balls and b blue balls. What is the probability that the first ball will be red and the second ball will be blue?

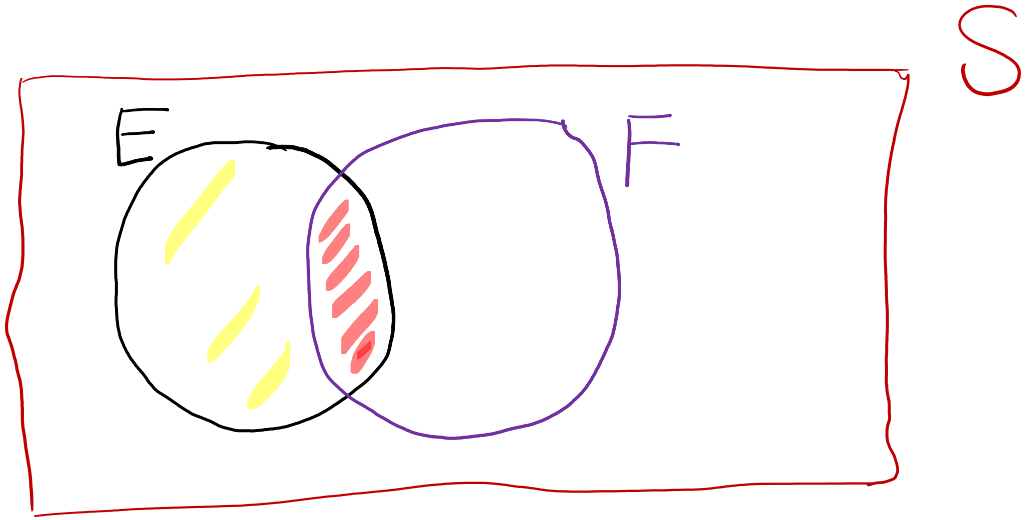
Solution

Let A be the event that the first ball is red, and let B be the event that the second ball is blue. Obviously, $P(A) = r/(r + b)$. Furthermore, if the event A has occurred, then one red ball has been removed from the box on the first draw. Therefore, the probability of obtaining a blue ball on the second draw will be

$$\mathrm{P}(B|A) = \frac{b}{r + b - 1}$$

It follows that

$$P(AB) = \frac{b}{r+b-1} \cdot \frac{r}{r+b}$$

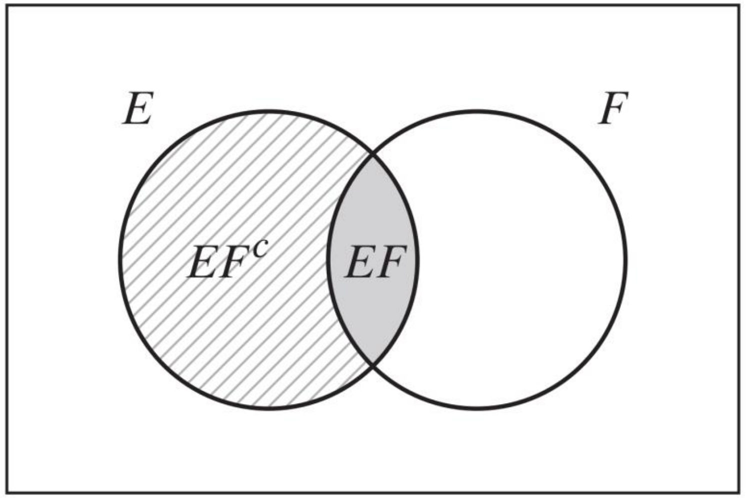


E = Yellow Region $\cup$ Red Region

$\Rightarrow E = EF^c \cup EF$

Let E and F be events. We may express E as

$$E = EF \cup EF^c$$



$E = EF \cup EF^c$. EF = Shaded Area; EF^c = Striped Area.

As EF and EF^c are clearly mutually exclusive, we have by Axiom 3 that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$

$$\Rightarrow \boxed{P(E) = P(E|F)P(F) + P(E|F^c)[1 - P(F)]}$$

Problem

A laboratory blood test is 99 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability .01, the test result will imply he or she has the disease.) If .5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Solution

Let D be the event that the tested person has the disease and E the event that his test result is positive. The desired probability P(D|E) is obtained by

$$P(D|E) = \frac{P(DE)}{P(E)}$$

$$= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)}$$

$$\Rightarrow \quad P(D|E) = \frac{(.99)(.005)}{(.99)(.005) + (.01)(.995)}$$

$$= .3322$$

Thus, only 33 percent of those persons whose test results are positive actually have the disease.

Suppose that $F_1, F_2, F_3, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

In other words, exactly one of the events $F_1, F_2, F_3, \ldots, F_n$ must occur. By writing

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events EF_i, $i = 1, \ldots, n$ are mutually exclusive, we obtain that

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

OR

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n)$$

The Law of Total Probability

Let $F_1, F_2, \ldots, F_n$ be mutually exclusive and exhaustive events. Then for any other event E,

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n)$$
$$= \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Events $F_1, F_2, \ldots, F_n$ are called exhaustive events if $F_1 \cup F_2 \cup \cdots \cup F_n = S$.

Problem

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Solution

$$\begin{aligned} A&: \text{ the product is defective,} \\ B_1&: \text{ the product is made by machine } B_1, \\ B_2&: \text{ the product is made by machine } B_2, \\ B_3&: \text{ the product is made by machine } B_3. \end{aligned}$$

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

and hence

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

Bayes' Formula

Suppose now that E has occurred and we are interested in determining which one of F_j also occurred. We have

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$

$$\Rightarrow \quad \boxed{P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}}$$

In term of two events A and B we can work for Bayes Rule as follows :

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

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

Note:

Most of the time we don't have P(B) so we can use law of total probability to compute P(B).

Problem

Consider a test that can diagnose kidney cancer. The test correctly detects when a patient has cancer 90% of the time. Also, if a person does not have cancer, the test correctly indicates so 99.9% of the time. Finally, suppose it is known that 1 in every 10,000 individuals has kidney cancer. Find the probability that a patient has kidney cancer, given that the test indicates he does.

Solution

Let C denote the event that a patient has kidney cancer then C^c will denote the event that a patient has no kidney cancer.
Po denote the event that the patient tests positive for kidney cancer then $Po^c = N$ denote the event that the patient tests negative for kidney cancer.

Now we want $P(C|Po)$.

Given data is :

Test correctly detects when a patient has cancer 90% of the time : $P(Po|C) = \frac{90}{100} = 0.9$

If a person does not have cancer, the test correctly indicates so 99.9% of the time:

$$P(N|C^C) = \frac{99.9}{100} = 0.999 \Rightarrow P(Po|C^C) = 1 - 0.999 = 0.001$$

It is known that 1 in every 10,000 individuals has kidney cancer:

$P(C) = \frac{1}{10000} = 0.0001 \Rightarrow P(C^C) = 1 - P(C) = 0.9999.$

By Law of Total Probability

$$\begin{aligned} P(Po) &= P(Po|\mathrm{C})P(C) + P(Po|C^c)P(C^C) \\ &= (0.9 * 0.0001) + (0.001 * 0.9999) \\ &= 0.0010899 \end{aligned}$$

By Bayes Rule,

$$P(C|\mathrm{P}o) = \frac{P(Po|C)P(C)}{P(Po)} = \frac{0.9 * 0.0001}{0.0010899} \approx 0.08$$

This implies that only about 8% of patients that test positive under this particular test actually have kidney cancer.

Practice Problems

1. You ask your neighbor to water a sickly plant while you are on vacation. Without water it will die with probability 0.8; with water it will die with probability 0.15. You are 90 percent certain that your neighbor will remember to water the plant.

(a) What is the probability that the plant will be alive when you return?

(b) If it is dead, what is the probability your neighbor forgot to water it?

---

2. There is a 60 percent chance that the event A will occur. If A does not occur, then there is a 10 percent chance that B will occur. What is the probability that at least one of the events A or B occurs?

Independent Events

E is independent of F if knowledge that F has occurred does not change the probability of E occurrence.

Since $P(E|F) = P(EF)/P(F)$, we see that E is independent of F if

$$P(EF) = P(E)P(F)$$

Since this equation is symmetric in E and F, it shows that whenever E is independent of F so is F of E.

Two events E and F are said to be independent if

$$P(EF) = P(E)P(F)$$

Two events E and F that are not independent are said to be dependent.

Problem

Suppose that we roll a pair of fair dice, so each of the 36 possible outcomes is equally likely. Let A denote the event that the first die lands on 3, let B be the event that the sum of the dice is 8, and let C be the event that the sum of the dice is 7.

(a) Are A and B independent?
(b) Are A and C independent?

Solution

(a) Since $A \cap B$ is the event that the first die lands on 3 and the second on 5, we see that

$$P(A \cap B) = P(\{(3, 5)\}) = \frac{1}{36}$$

On the other hand,

$$P(A) = P(\{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}) = \frac{6}{36}$$

and

$$P(B) = P(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}) = \frac{5}{36}$$

Therefore, since $1/36 \neq (6/36) \cdot (5/36)$, we see that

$$P(A \cap B) \neq P(A)P(B)$$

and so events A and B are not independent.

Similar solve part (b).

Problem

Toss two coins and observe the outcome. Define these events:

A: Head on the first coin

B: Tail on the second coin

Are events A and B independent?

Solution

$$P(A)=\frac{1}{2}, P(B)=\frac{1}{2}, \text{and } P(A\cap B)=\frac{1}{4}.$$

Since $P(A)P(B)=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=\frac{1}{4}$ and $P(A\cap B)=\frac{1}{4}$, we have $P(A)P(B)=P(A\cap B)$ and the two events must be independent.

Theorem

If A and B are independent events, then each of the following pair of events are also independent:

(i) A and B^C

(ii) A^C and B

(iii) A^C and B^c

Independency of Three Events

Three events E, F, and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

Note that if E, F, and G are independent, then E will be independent of any event formed from F and G.

Definition of Independence of Several Events

We say that the events $A_1, A_2, \ldots, A_n$ are **independent** if

$$\mathbf{P}\left(\bigcap_{i\in S} A_i\right) = \prod_{i\in S} \mathbf{P}(A_i), \qquad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}.$$

Generalized Chain Rule

$$P(E_1E_2E_3 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \dots P(E_n|E_1E_2 \dots E_{n-1})$$

Practice Problem

1.

Of three cards, one is painted red on both sides; one is painted black on both sides; and one is painted red on one side and black on the other. A card is randomly chosen and placed on a table. If the side facing up is red, what is the probability that the other side is also red?

2.

If $B \subset A$, then show that $P(A\backslash B) = P(A) - P(B)$.

[Hint: Use Venn diagram and third axiom of probability]

Practice Problem

Among employees of a certain firm, 70% know Java, 60% know Python, and 50% know both languages. What portion of programmers

(a) does not know Python?

(b) does not know Python and does not know Java?

(c) knows Java but not Python?

(d) knows Python but not Java?

(e) If someone knows Python, what is the probability that he/she knows Java too?

(f) If someone knows Java, what is the probability that he/she knows Python too?

Practice Problem

(Diagnostics of computer codes): A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

Practice Problem

A computer maker receives parts from three suppliers, S1, S2, and S3. Fifty percent come from S1, twenty percent from S2, and thirty percent from S3. Among all the parts supplied by S1, 5% are defective. For S2 and S3, the portion of defective parts is 3% and 6%, respectively.

(a) What is the probability that a random part is defective?

(b) A customer complains that a certain part in her recently purchased computer is defective. What is the probability that it was supplied by S1?

Practice Problem

1. Prove that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

2. A programming class is composed of 10 juniors, 30 seniors, and 10 graduate students. The final grades show that 5 of the juniors, 10 of the seniors, and 5 of the graduate students received an A for the course. If a student is chosen at random from this class and is found to have earned an A, what is the probability that he or she is a senior?

Practice Problems

1.

All athletes at the Olympic games are tested for performance-enhancing steroid drug use. The imperfect test gives positive results (indicating drug use) for 90% of all steroid users but also (and incorrectly) for 2% of those who do not use steroids. Suppose that 5% of all registered athletes use steroids. If an athlete is tested negative, what is the probability that he/she uses steroids?

2.

At a plant, 20% of all the produced parts are subject to a special electronic inspection. It is known that any produced part which was inspected electronically has no defects with probability 0.95. For a part that was not inspected electronically this probability is only 0.7. A customer receives a part and find defects in it. What is the probability that this part went through an electronic inspection?

Random Variables & Expectation

Random Variables

A **random variable** is a real-valued function whose domain is a sample space.

Random variables are typically denoted by uppercase letters, such as X, Y, and Z. The actual numerical values that a random variable can assume are denoted by lowercase letters, such as x, y, and z.

Mathematically,

$$X : \Omega \longrightarrow \mathbb{R}$$

where Ω represents sample space

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Examples :

Let X denote the random variable that is defined as the sum of two fair dice, then

$P\{X = 2\} = P\{(1, 1)\} = 1/36$

$P\{X = 3\} = P\{(1, 2), (2, 1)\} = 2/36$

$P\{X = 4\} = P\{(1, 3), (2, 2), (3, 1)\} = 3/36$

$P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = 4/36$

$P\{X = 6\} = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = 5/36$

$P\{X = 7\} = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = 6/36$

$P\{X = 8\} = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = 5/36$

$P\{X = 9\} = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = 4/36$

$P\{X = 10\} = P\{(4, 6), (5, 5), (6, 4)\} = 3/36$

$P\{X = 11\} = P\{(5, 6), (6, 5)\} = 2/36$

$P\{X = 12\} = P\{(6, 6)\} = 1/36$

In other words, the random variable X can take on any integral value between 2 and 12.

Note:

$$1 = P(S) = P\left(\bigcup_{i=2}^{12} \{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

$P\{Y = i\} = 1/6, i = 1, 2, 3, 4, 5, 6.$

Next Example :

Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities .09, .21, .21, .49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the indicator random variable for the event A, since I will equal 1 or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I = 1\} = .91$$
$$P\{I = 0\} = .09$$

Distribution of a randam variable X

Collection of all the probabilities related to X is the distribution of X.

Let X is the number of 1's in a random binary string of 3 characters then the distribution of X is

x	$P\{X = x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

Because

$$\Omega = \{000, 001, 010, 100, \\ 111, 110, 101, 011\}$$

Discrete and Continuous Random Variables

In the two foregoing examples, the random variables of interest took on a finite number of possible values. Random variables whose set of possible values can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$ are said to be *discrete*. For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable. However, there also exist random variables that take on a continuum of possible values i.e., a random variable that can take on any real value in an interval (possibly even the entire real line). These are known as *continuous* random variables. In other words a random variable X is said to be continuous if it can take on the infinite number of possible values associated with intervals of real numbers. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The Cumulative Distribution Function (CDF)

Note: CDF is defined for both Discrete & Continuous r.v.s.

The cumulative distribution function, or more simply the distribution function, F of the random variable X is defined for any real number x by

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

That is, F (x) is the probability that the random variable X takes on a value that is less than or equal to x.

***Notation*:** We will use the notation $X \sim F$ to signify that F is the distribution function of X .

Suppose we wanted to compute $P\{a < X \leq b\}$. This can be accomplished by first noting that the event $\{X \leq b\}$ can be expressed as the union of the two mutually exclusive events $\{X \leq a\}$ and $\{a < X \leq b\}$. Therefore, applying Axiom 3, we obtain that

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

$$\longrightarrow P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\}$$

or

$$P\{a < X \leq b\} = F(b) - F(a)$$

Problem

Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

Solution The desired probability is computed as follows:

$$P\{X > 1\} = 1 - P\{X \leq 1\}$$
$$= 1 - F(1)$$
$$= e^{-1}$$

$$\rightarrow \quad P\{X > 1\} = .368$$

Probability Mass Function (PMF)

Note: Only discrete r.v.s have PMF

As was previously mentioned, a random variable whose set of possible values is a sequence is said to be discrete. For a discrete random variable X , we define the probability mass function $p(a)$ of X by

$$p(a) = P\{X = a\}$$

If X assume one of the values $x_1, x_2, \ldots$, then

$$p(x_i) > 0, \qquad i = 1, 2, \ldots$$

and $$p(x) = 0, \qquad \text{all other values of } x$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Problem

Consider a random variable X that is equal to 1, 2, or 3. If we know that

$$p(1) = \tfrac{1}{2} \qquad \text{and} \qquad p(2) = \tfrac{1}{3}$$

Compute $p(3)$.

Solution

since $p(1) + p(2) + p(3) = 1$

$$\longrightarrow p(3) = \tfrac{1}{6}$$

A graph of $p(x)$ is presented in Figure 4.1.

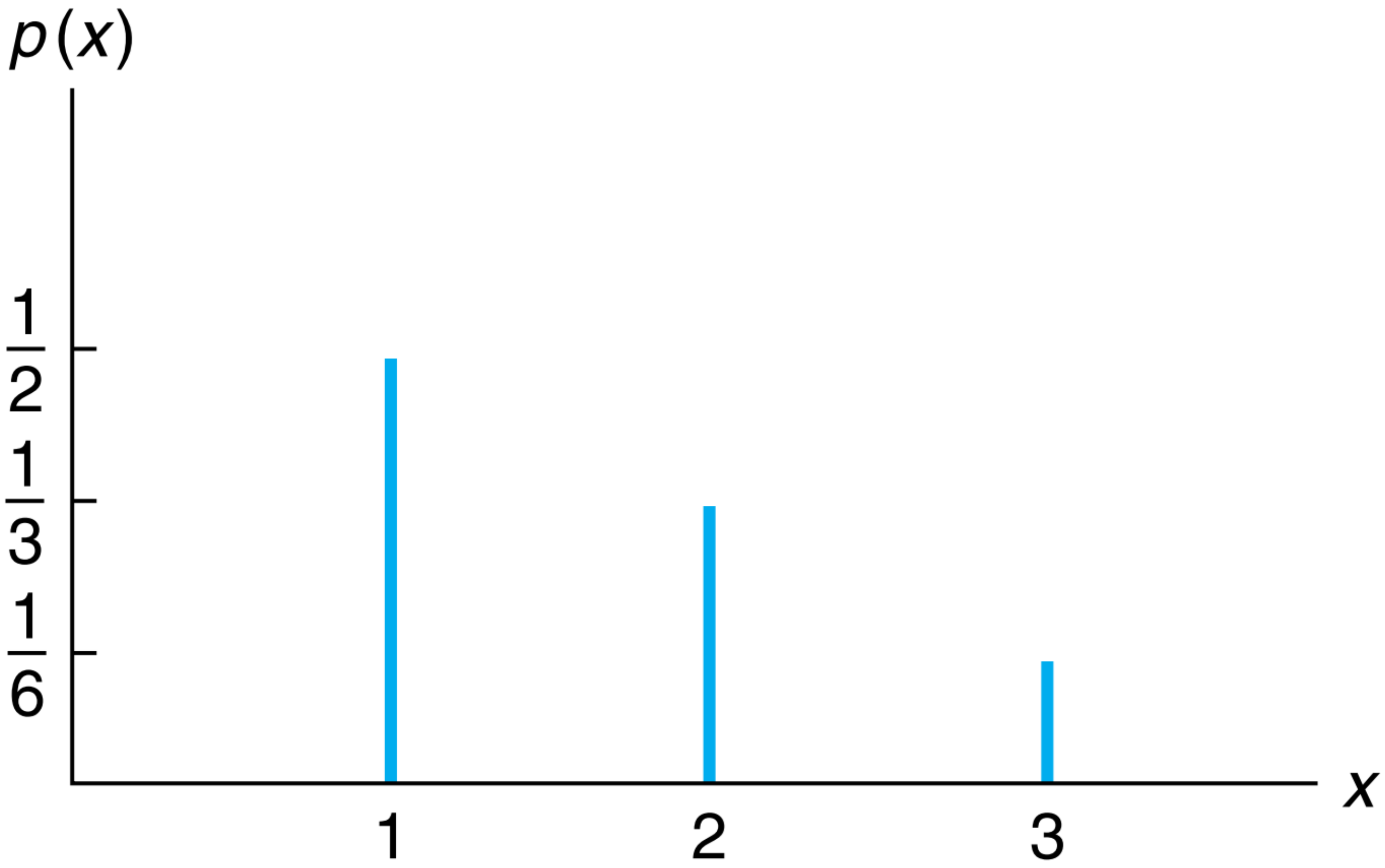


FIGURE 4.1 *Graph of* $p(x)$

Relation of CDF with PMF

The cumulative distribution function F can be expressed in terms of $p(x)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function F is a step function. That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.

For instance, suppose X has a probability mass function given by

$$p(1) = \frac{1}{2}, \qquad p(2) = \frac{1}{3}, \qquad p(3) = \frac{1}{6}$$

Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \leq a < 2 \\ \frac{5}{6} & 2 \leq a < 3 \\ 1 & 3 \leq a \end{cases}$$

This is graphically presented in Figure 4.2.

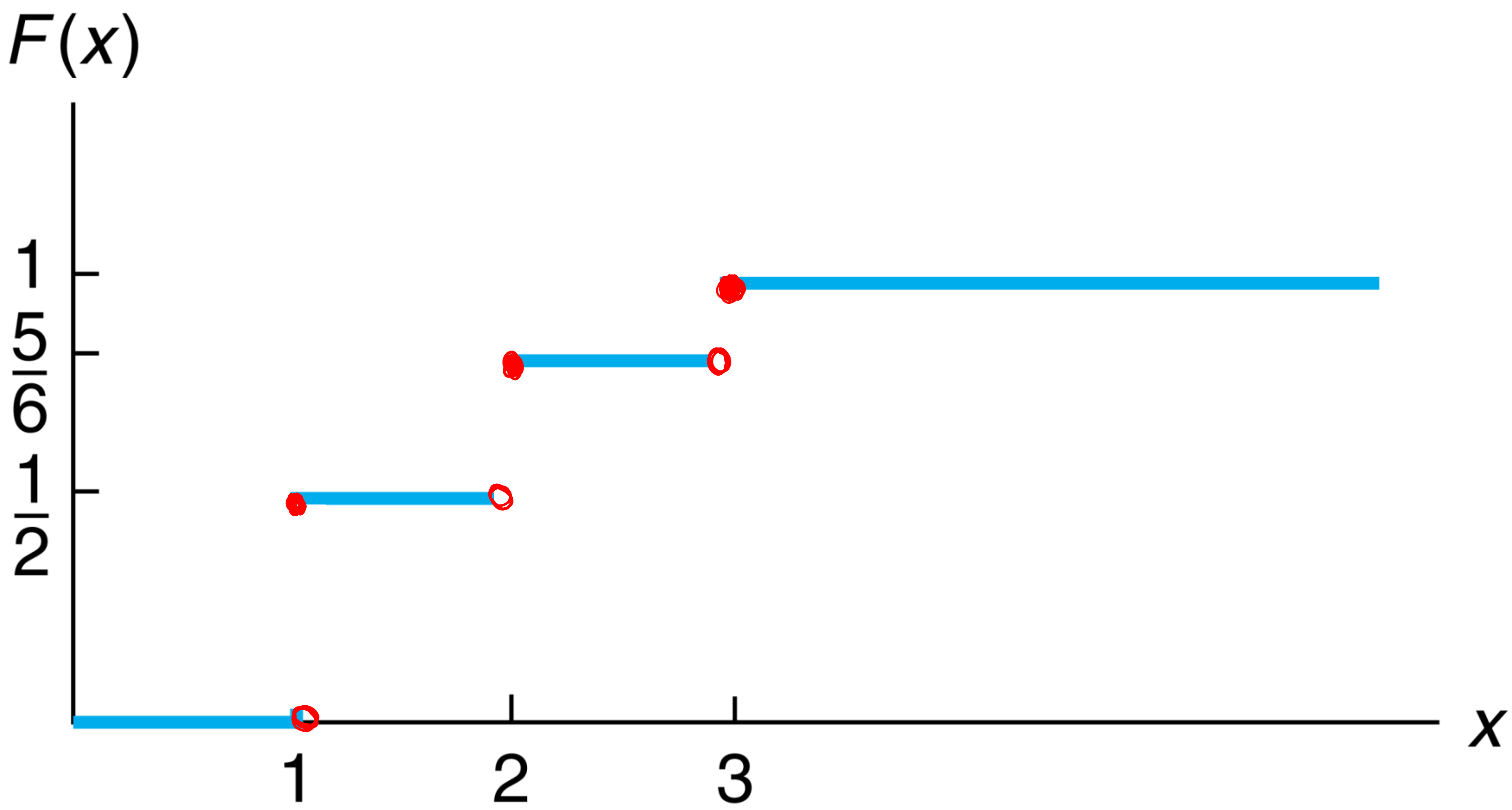


FIGURE 4.2 *Graph of* $F(x)$

Note:

If we are given a CDF and it's a step function, then it is the CDF of a discrete random variable.

(A step function is that one which is made up of disconnected constant functions. This type of function is also referred to as a stair function because its graph is like stairs.)

Method of finding PMF from a CDF of a discrete random variable

We can get an idea about the possible values of the random variable from the CDF by looking to the points where the jumps occurs in the graph of CDF.

If $x_1, x_2, \ldots$ are possible values of the random variable then

$$F(x_1) = P\{X \le x_1\} = P\{X = x_1\} = p(x_1)$$

$$\Rightarrow F(x_1) = p(x_1) \text{ or } p(x_1) = F(x_1)$$

Next

$$F(x_2) = P\{X \le x_2\} = P\{X \le x_1\} + P\{X = x_2\}$$

$$\Rightarrow F(x_2) = F(x_1) + p(x_2) \Rightarrow p(x_2) = F(x_2) - F(x_1)$$

Similarly

$$p(x_3) = F(x_3) - F(x_2) \text{ and so on.}$$

Practice Problem

Find the PMF of the random variable whose CDF is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \leq a < 2 \\ \frac{5}{6} & 2 \leq a < 3 \\ 1 & 3 \leq a \end{cases}$$

Probability Density Function (PDF)

Note: Only continuous r.v.s have PDF

We say that a random variable X has a *continuous distribution* or that X is a *continuous random variable* if there exists a nonnegative function f, defined on the real line, such that for every interval of real numbers (bounded or unbounded), the probability that X takes a value in the interval is the integral of f over the interval.

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

The function *f(x)* is called the *probability density function* of the random variable *X.*

Note that a probability density function $f(x)$ must satisfy the following

$$f(x) \geq 0 \text{ for all } x$$

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

If we let $a = b$ in ★ then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any particular value is zero. The fact that $P(X = a) = 0$ does not imply that $X = a$ is impossible. If it did, all values of X would be impossible and X couldn't assume any value. What happens is that the probability in the distribution of X is spread so thinly that we can only see it on sets like nondegenerate intervals.

As an immediate consequence of the fact that in the continuous case probabilities associated with individual points are always zero, we find that if we speak of the probability associated with the interval from a to b, it does not matter whether either endpoint is included. Symbolically,

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b)$$

Problem

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

Solution

(a) Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that

$$C \int_0^2 (4x - 2x^2)\,dx = 1$$

$$\Rightarrow \quad C \left[2x^2 - \frac{2x^3}{3} \right] \Big|_{x=0}^{x=2} = 1$$

or

$$C = \frac{3}{8}$$

(b)

$$P\{X > 1\} = \int_1^\infty f(x)\,dx = \frac{3}{8}\int_1^2 (4x - 2x^2)\,dx = \frac{1}{2}$$

Problem

The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down?
(b) it will function for fewer than 100 hours?

Solution (a) Since

$$1 = \int_{-\infty}^{\infty} f(x)\, dx = \lambda \int_{0}^{\infty} e^{-x/100}\, dx$$

we obtain

$$1 = -\lambda(100)e^{-x/100}\Big|_0^{\infty} = 100\lambda \quad \text{or} \quad \lambda = \frac{1}{100}$$

Hence, the probability that a computer will function between 50 and 150 hours before breaking down is given by

$$P\{50 < X < 150\} = \int_{50}^{150} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_{50}^{150}$$
$$= e^{-1/2} - e^{-3/2} \approx .383$$

(b) Similarly,

$$P\{X < 100\} = \int_{0}^{100} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_{0}^{100} = 1 - e^{-1} \approx .632$$

In other words, approximately 63.2 percent of the time, a computer will fail before registering 100 hours of use. ■

Relation b/w CDF and PDF

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides yields

$$\frac{d}{da} F(a) = f(a)$$

(By Fundamental Theorem of Calculus)

Problem

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Solution

For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} \frac{t^2}{3} dt = \left.\frac{t^3}{9}\right|_{-1}^{x} = \frac{x^3+1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

Practice Problems

1.

The distribution function of the random variable X is given

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \leq x < 1 \\ \dfrac{2}{3} & 1 \leq x < 2 \\ \dfrac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

(a) Plot this distribution function.

(b) What is $P\{X > \frac{1}{2}\}$?

(c) What is $P\{2 < X \leq 4\}$?

2.

Five men and 5 women are ranked according to their scores on an examination. Assume that no two scores are alike and all 10! possible rankings are equally likely. Let X denote the highest ranking achieved by a woman (for instance, $X = 2$ if the top-ranked person was male and the next-ranked person was female). Find $P\{X = i\}, i = 1, 2, 3, \ldots, 8, 9, 10$.

3. If a random variable has the probability density

$$f(x) = \begin{cases} 2\,e^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

find the probabilities that it will take on a value

(a) between 1 and 3;

(b) greater than 0.5.

4.

Find k so that the following can serve as the probability density of a random variable:

$$f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ kxe^{-4x^2} & \text{for } x > 0 \end{cases}$$

5. The length of satisfactory service (years) provided by a certain model of laptop computer is a random variable having the probability density

$$f(x) = \begin{cases} \dfrac{1}{4.5}\, e^{-x/4.5} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

Find the probabilities that one of these laptops will provide satisfactory service for

(a) at most 2.5 years;

(b) anywhere from 4 to 6 years;

(c) at least 6.75 years.

6.

Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 \le x < 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Evaluate k.

(b) Find $F(x)$ and use it to evaluate $P(3 < X < 4)$.

Jointly Distributed Random Variables

To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

A knowledge of the joint probability distribution function often enable us to compute the probability of statements concerning the values of X and Y .

In the case where X and Y are both discrete random variables whose possible values are, respectively, $x_1, x_2, \ldots,$ and $y_1, y_2, \ldots,$ we define the *joint probability mass function* of X and Y, $p(x_i, y_j)$, by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

The individual probability mass functions of X and Y are easily obtained from the joint probability mass function by the following reasoning. Since Y must take on some value y_j, it follows that the event $\{X = x_i\}$ can be written as the union, over all j, of the mutually exclusive events $\{X = x_i, Y = y_j\}$. That is,

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

and so, using Axiom 3 of the probability function, we see that

$$P\{X = x_i\} = P\left(\bigcup_j \{X = x_i, Y = y_j\}\right)$$

$$= \sum_j P\{X = x_i, Y = y_j\}$$

$$\Rightarrow \quad P\{X = x_i\} = \sum_j p(x_i, y_j)$$

Similarly, we can obtain $P\{Y = y_j\}$ by summing $p(x_i, y_j)$ over all possible values of x_i, that is,

$$P\{Y = y_j\} = \sum_i P\{X = x_i, Y = y_j\}$$

$$\Rightarrow \quad P\{Y = y_j\} = \sum_i p(x_i, y_j)$$

Hence, specifying the joint probability mass function always determines the individual mass functions. However, it should be noted that the reverse is not true. Namely, knowledge of $P\{X = x_i\}$ and $P\{Y = y_j\}$ does not determine the value of $P\{X = x_i, Y = y_j\}$.

Again, $\{(X, Y) = (x, y)\}$ are exhaustive and mutually exclusive events for different pairs (x, y), therefore,

$$\sum_x \sum_y p(x, y) = 1$$

The above equation tell us that if we feed all possible x and y values to the joint PMF and sum up all the resulting values. It should be one. This is the underlying property of a valid joint PMF.

Example

Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability mass function of X and Y , $p(i, j) = P\{X = i, Y = j\}$, is given by

$$p(0, 0) = \binom{5}{3} \Big/ \binom{12}{3} = 10/220$$

$$p(0, 1) = \binom{4}{1} \binom{5}{2} \Big/ \binom{12}{3} = 40/220$$

$$p(0, 2) = \binom{4}{2} \binom{5}{1} \Big/ \binom{12}{3} = 30/220$$

$$p(0,3) = \binom{4}{3} \Big/ \binom{12}{3} = 4/220$$

$$p(1,0) = \binom{3}{1}\binom{5}{2} \Big/ \binom{12}{3} = 30/220$$

$$p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1} \Big/ \binom{12}{3} = 60/220$$

$$p(1,2) = \binom{3}{1}\binom{4}{2} \Big/ \binom{12}{3} = 18/220$$

$$p(2,0) = \binom{3}{2}\binom{5}{1} \Big/ \binom{12}{3} = 15/220$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = 12/220$$

$$p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = 1/220$$

These probabilities can most easily be expressed in tabular form as shown in Table 4.1.

TABLE 4.1 $P\{X = i, Y = j\}$

$i \backslash j$	**0**	**1**	**2**	**3**	**Row Sum $= P\{X = i\}$**
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Because the individual probability mass functions of X and Y thus appear in the margin of such a table, they are often referred to as being the marginal probability mass functions of X and Y , respectively. It should be noted that to check the correctness of such a table we could sum the marginal row (or the marginal column) and verify that its sum is **1**.

Problem

A program consists of two modules. The number of errors, X, in the first module and the number of errors, Y , in the second module have the joint distribution, $p(0,0) = p(0,1) = p(1,0) = 0.2$, $p(1,1) = p(1,2) = p(1,3) = 0.1$, $p(0,2) = p(0,3) = 0.05$. Find (a) the marginal distributions of X and Y , (b) the probability of no errors in the first module, and (c) the distribution of the total number of errors in the program.

[Remember: Collection of all the probabilities related to some random variable is sometime called the distribution of the random variable. i.e. joint distribution mean joint PMF etc.]

Solution

It is convenient to organize the joint pmf of X and Y in a table. Adding row wise and column wise, we get the marginal pmfs,

$p(x,y)$		$y=0$	$y=1$	$y=2$	$y=3$	$p_X(x)$
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$p_Y(y)$		0.40	0.30	0.15	0.15	1.00

This solves (a).

(b) $P_X(0) = 0.50$. (Because the probability of no errors in the first module mean $P\{X = 0\}$.

(c) Let Z = X + Y be the total number of errors. To find the distribution of Z, we first identify its possible values, then find the probability of each value. We see that Z can be as small as 0 and as large as 4. Then,

$$p_Z(0) = P\{X + Y = 0\} = P\{X = 0, Y = 0\} = p(0,0) = 0.20,$$

$$p_Z(1) = P\{X = 0, Y = 1\} + P\{X = 1, Y = 0\}$$

$$= p(0,1) + p(1,0) = 0.20 + 0.20 = 0.40,$$

$$p_Z(2) = p(0,2) + p(1,1) = 0.05 + 0.10 = 0.15,$$

$$p_Z(3) = p(0,3) + p(1,2) = 0.05 + 0.10 = 0.15,$$

$$p_Z(4) = p(1,3) = 0.10.$$

It is a good check to verify that

$$\sum_z p_Z(\mathrm{z}) = 1$$

Practice Problem

An internet service provider charges its customers for the time of the internet use rounding it up to the nearest hour. The joint distribution of the used time (X, hours) and the charge per hour (Y, cents) is given in the table below.

$p(x,y)$		$x=1$	$x=2$	$x=3$	$x=4$
y	1	0	0.06	0.06	0.10
	2	0.10	0.10	0.04	0.04
	3	0.40	0.10	0	0

Each customer is charged $Z = X \cdot Y$ cents, which is the number of hours multiplied by the price of each hour. Find the distribution of Z.

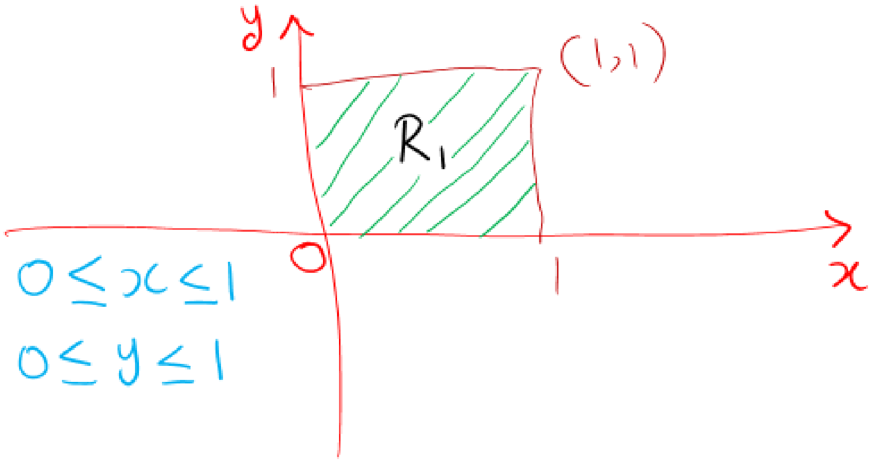
[Hint: First find all the possible values of Z and then find its probabilities.]

$\int_a^b f(x)\,dx$ is an integral defined over an interval $[a,b]$.

We also have integrals that are defined over a region R in xy-plane. e.g the double integral $\iint_R f(x,y)\,dA$ where R is some region in xy-plane.

The question is : How to evaluate a double integral?
Well! it depends upon the region R.

Consider region R_1 shown in the figure. In this case, we can evaluate the double integral quite easily. i.e.



$0 \le x \le 1$
$0 \le y \le 1$

$$\iint_{R_1} (x^2y + x)\, dA = \int_0^1 \int_0^1 (x^2y + x)\, dx\, dy$$

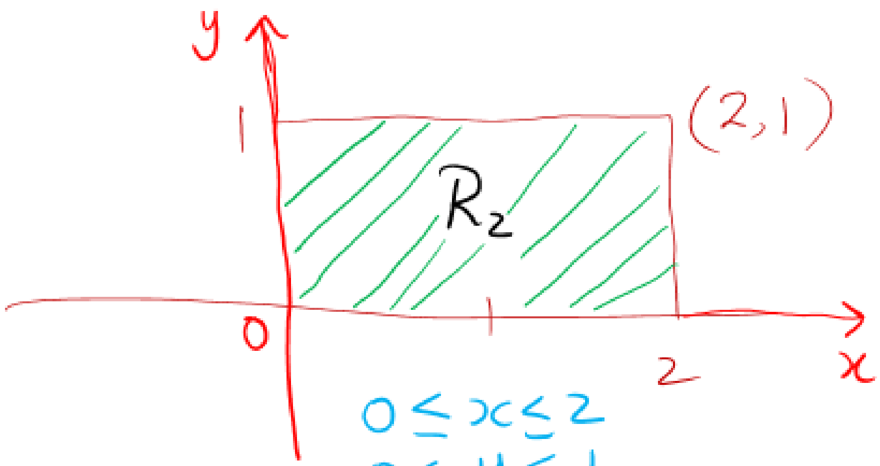
$$= \int_0^1 \left(\frac{x^3y}{3} + \frac{x^2}{2}\right)\Big|_0^1 dy = \int_0^1 \left(\frac{y}{3} + \frac{1}{2}\right) dy$$

$$= \left(\frac{y^2}{6} + \frac{1}{2}x\right)\Big|_0^1 = \frac{1}{6} + \frac{1}{2} = \frac{1+3}{6} = \frac{4}{6}$$

$$\Rightarrow \boxed{\iint_{R_1} (x^2y + x)\, dA = \frac{2}{3}}$$

Note $\int_0^1 \int_0^1 (x^2y + x)\, dy\, dx$ also gives the same result.

Now consider the region shown in the figure.



$0 \le x \le 2$
$0 \le y \le 1$

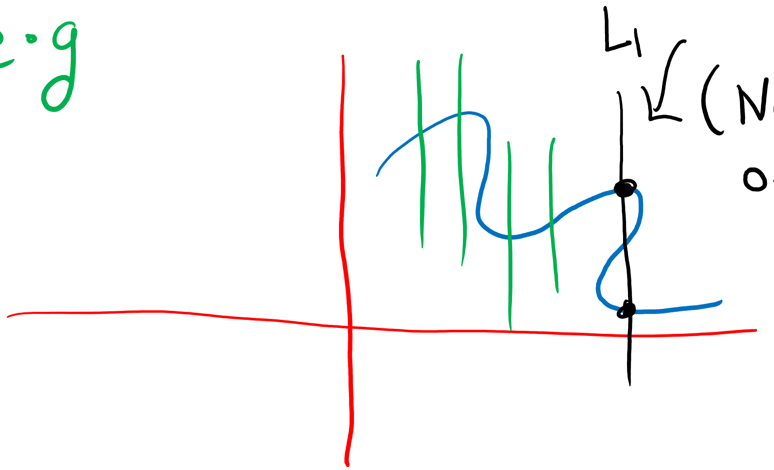
$$\iint_{R_2} (x^2 y + x)\, dA = \int_0^1 \int_0^2 (x^2 y + x)\, dx\, dy = \int_0^2 \int_0^1 (x^2 y + x)\, dy\, dx$$

Vertical Line Test

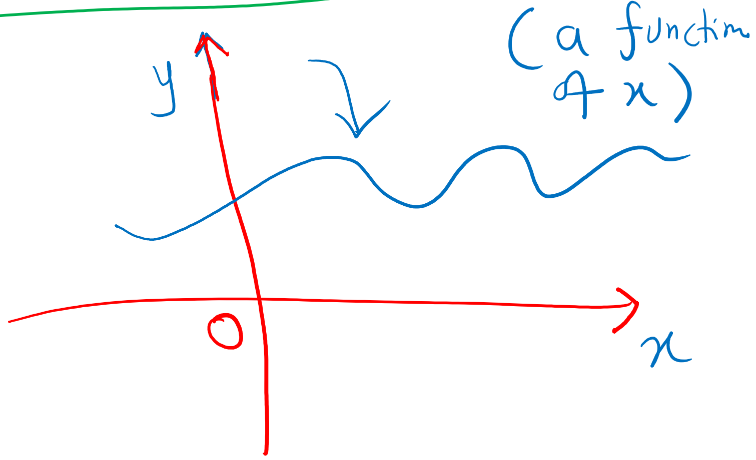
We use V.L.T to check whether a curve is a graph of a function of x.

A curve will be a graph of a function of 'x' if no vertical line intersect the curve more than once.

e.g



(Not a function of x because L_1 intersect the curve at two points)



Some Concepts from Multivariable Calculus

A plane region D is said to be of **type I** if it lies between the graphs of two continuous functions of x, that is,

$$D = \{(x, y) \mid a \leqslant x \leqslant b,\ g_1(x) \leqslant y \leqslant g_2(x)\}$$

where g_1 and g_2 are continuous on $[a, b]$. Some examples of type I regions are shown in Figure 5.

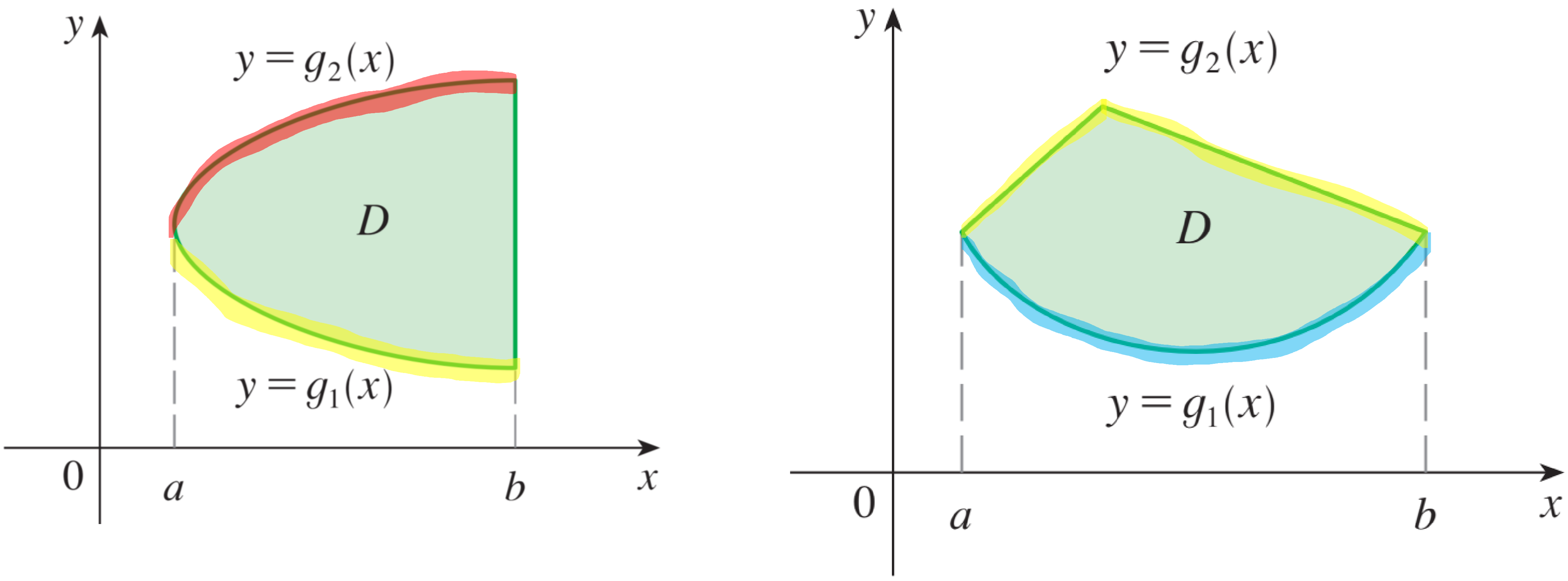


FIGURE 5

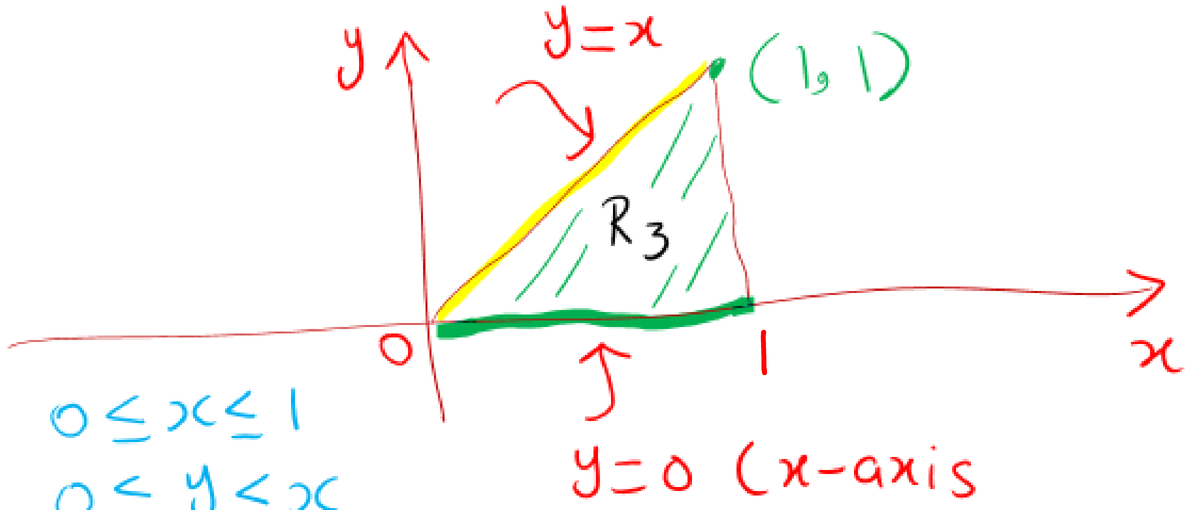
Some type I regions

$\boxed{3}$ If f is continuous on a type I region D described by

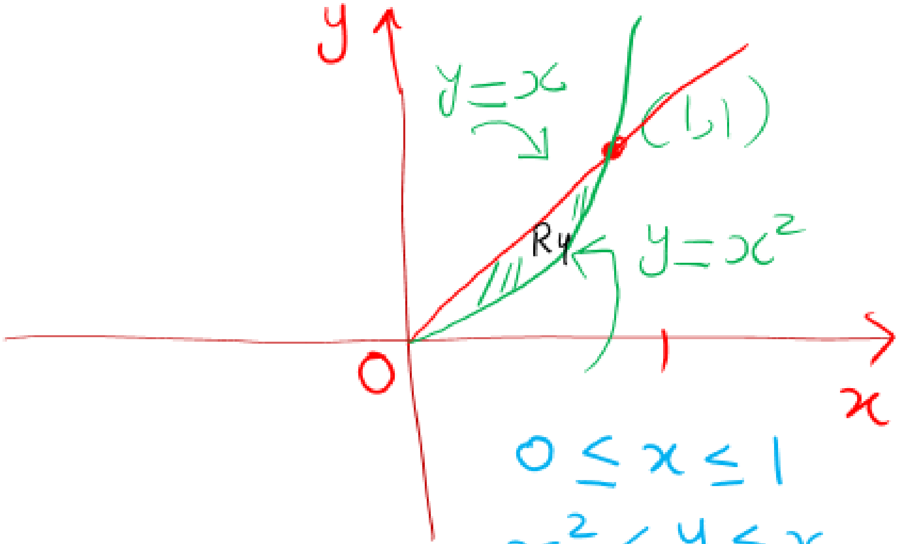
$$D = \{(x, y) \mid a \leqslant x \leqslant b,\ g_1(x) \leqslant y \leqslant g_2(x)\}$$

then
$$\iint_D f(x, y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\, dx$$

Let's evaluate $\iint_{R_3} (x^2y + x)\, dA$ & $\iint_{R_4} (x^2 + x)\, dA$, where R_3 & R_4 are shown in the figure. Both are type I region.



$0 \le x \le 1$
$0 \le y \le x$



$0 \le x \le 1$
$x^2 \le y \le x$

$$\iint_{R_3} (x^2y + x)\,dA = \int_0^1 \int_0^x (x^2y + x)\,dy\,dx = \int_0^1 \left(\frac{x^2y^2}{2} + xy\right)\Bigg|_0^x dx$$

$$= \int_0^1 \left(\frac{x^2 \cdot x^2}{2} + x \cdot x\right) dx = \int_0^1 \left(\frac{x^4}{2} + x^2\right) dx$$

$$= \left(\frac{x^5}{10} + \frac{x^3}{3}\right)\Bigg|_0^1 = \frac{1}{10} + \frac{1}{3} = \frac{3+10}{30} = \frac{13}{30}$$

$$\Rightarrow \boxed{\iint_{R_3} (x^2y + x)\,dA = \frac{13}{30}}$$

Similarly $\iint_{R_4} (x^2y + x)\,dx$

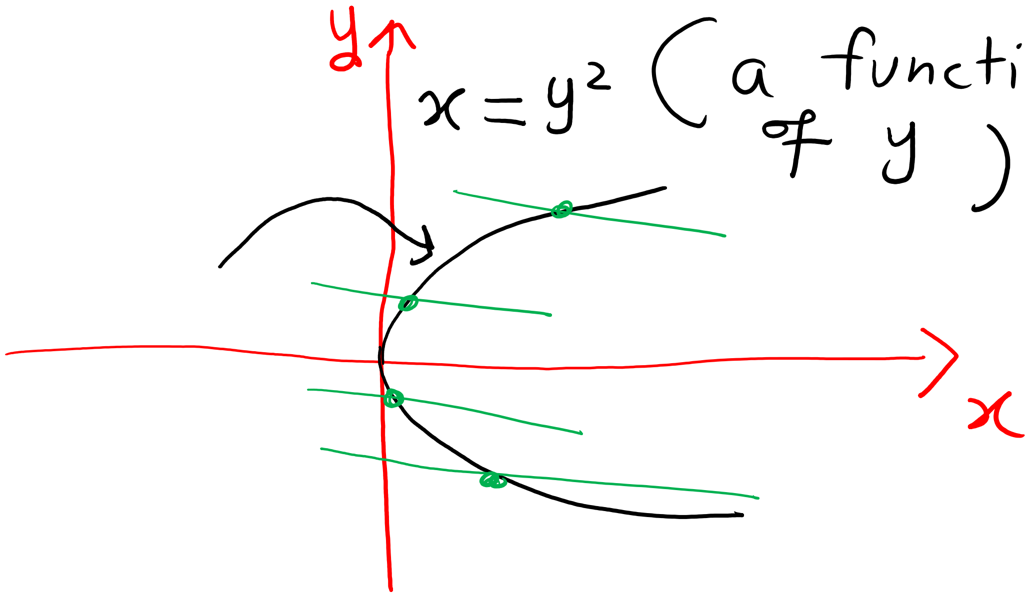
$$= \int_0^1 \int_{x^2}^{x} (x^2y + x)\,dy\,dx$$

Next,

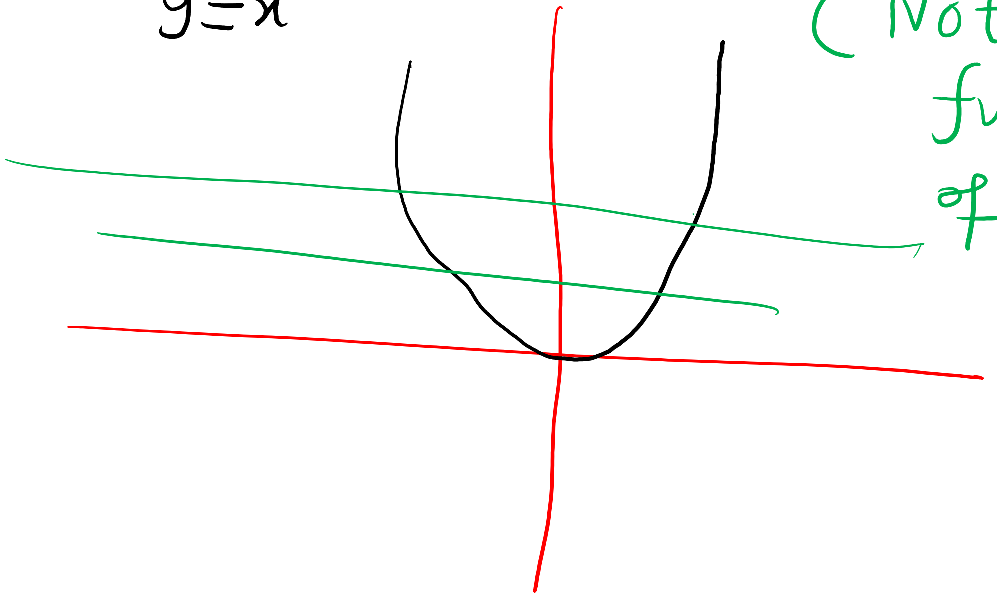
A curve will be a graph of a function of 'y' if no horizontal line intersect the curve more than once.

For example

$x = y^2$ (a function of y)



$y = x^2$ (Not a function of y)



We also consider plane regions of **type II**, which can be expressed as

$$D = \{(x, y) \mid c \leqslant y \leqslant d,\ h_1(y) \leqslant x \leqslant h_2(y)\}$$

where h_1 and h_2 are continuous. Three such regions are illustrated in Figure 7.

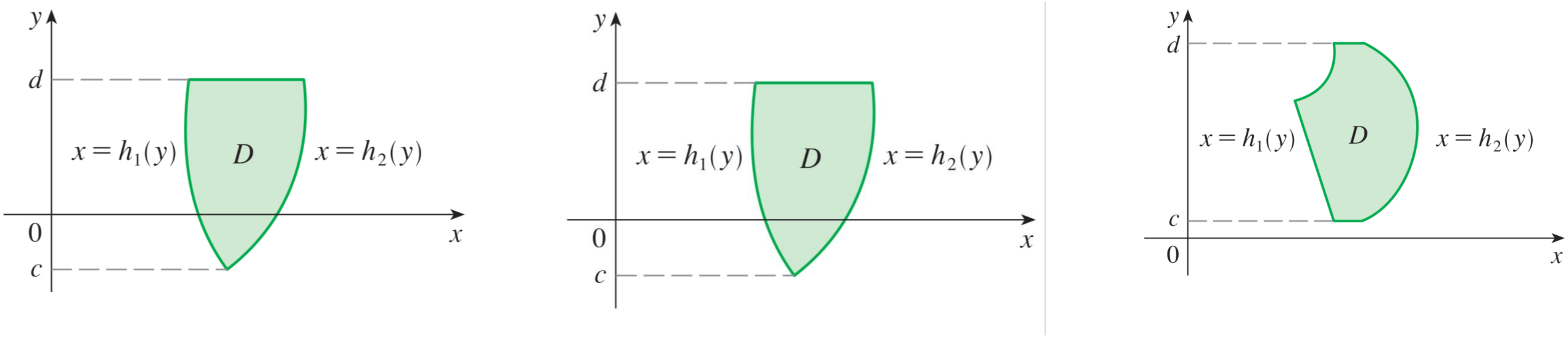


FIGURE 7
Some type II regions

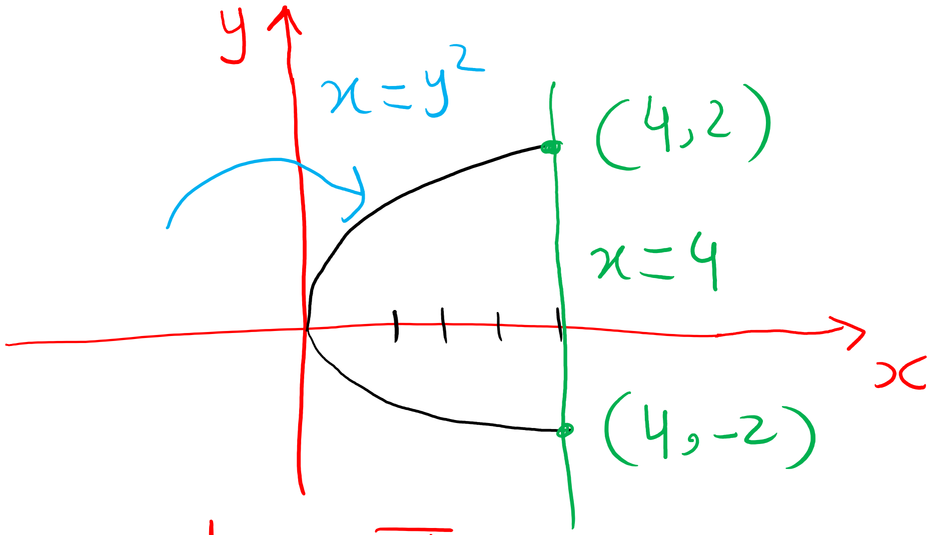
$\boxed{4}$ If f is continuous on a type II region D described by

$$D = \{(x, y) \mid c \leqslant y \leqslant d,\ h_1(y) \leqslant x \leqslant h_2(y)\}$$

then

$$\iint_D f(x, y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, dx\, dy$$

Let's evaluate $\iint\limits_{R_5} (x^2y + x)\, dA$, where R_5 is shown in the figure.



Clearly R_5 is a type II region $-2 \le y \le 2$ and $y^2 \le x \le 4$

So
$$\iint\limits_{R_5} (x^2y + x)\, dA = \int_{-2}^{2} \int_{y^2}^{4} (x^2 y + x)\, dx\, dy$$

Recall

The Fundamental Theorem of Calculus, Part 1 If f is continuous on $[a, b]$, then the function g defined by

$$g(x) = \int_a^x f(t)\,dt \qquad a \leqslant x \leqslant b$$

is continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) = f(x)$.

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane)

$$P\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y)\, dx\, dy \tag{4.3.3}$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from Equation 4.3.3 that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy \tag{4.3.4}$$

Because

$$F(a, b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\}$$

$$= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\, dx\, dy$$

it follows, upon differentiation, that

$$f(a, b) = \frac{\partial^2}{\partial a\, \partial b} F(a, b)$$

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\} \tag{4.3.5}$$

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

$$\Longrightarrow \quad P\{X \in A\} = \int_A f_X(x)\, dx$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$ is thus the probability density function of X.

Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{4.3.6}$$

Note: A joint PDF must satisfy the equation

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

Because probability that x and y belongs to real numbers is 1, just like in case of one continuous random variable we have $\oint_{-\infty}^{\infty} f(x)dx = 1$.

Problem The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute **(a)** $P\{X > 1, Y < 1\}$; **(b)** $P\{X < Y\}$; and **(c)** $P\{X < a\}$.

Solution

(a)

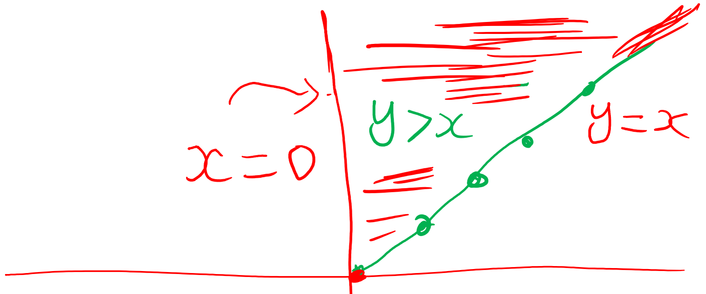
$$P\{X > 1, Y < 1\} = \int_0^1 \int_1^\infty 2e^{-x} e^{-2y}\, dx\, dy$$

$$= \int_0^1 2e^{-2y} (-e^{-x}|_1^\infty)\, dy$$

$$= e^{-1} \int_0^1 2e^{-2y}\, dy$$

$$\Longrightarrow \quad P\{X > 1, Y < 1\} = e^{-1}(1 - e^{-2})$$

$$P\{X < Y\} = \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy$$



$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy = \int_0^\infty 2e^{-2y}(1 - e^{-y})\,dy$$

$$= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy \quad = 1 - \frac{2}{3}$$

$$\Longrightarrow \quad P\{X < Y\} = \frac{1}{3}$$

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-2y} e^{-x}\, dy\, dx \; = \int_0^a e^{-x}\, dx$$

$$\Longrightarrow \quad P\{X < a\} = 1 - e^{-a}$$

Practice Problem 1

Two continuous random variables X and Y have the joint density

$$f(x,y) = C(x^2 + y), \quad -1 \le x \le 1, \quad 0 \le y \le 1.$$

(a) Compute the constant C.

(b) Find the marginal densities of X and Y.

[Hint: In part (a) set $\int_0^1 \int_{-1}^1 f(x,y)dxdy = 1$ and solve the resulting equation.]

Practice Problem 2

The joint probability density function of X and Y is given by

$$f(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 < x < 1, \quad 0 < y < 2$$

(a) Verify that this is indeed a joint density function.
(b) Compute the density function of X.
(c) Find $P\{X > Y\}$.

[Hint: (a) Since $f(x, y)$ is non negative and further show that $\int_0^2 \int_0^1 f(x, y)dxdy = 1.$]