CHAPTER 3

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

3.1 Concept of a Random Variable

Random Variable

A *random variable* is a function that associates a real number with each element in the sample space.

In other words, a random variable is a function $X : S \to R$, where S is the sample space of the random experiment under consideration.

NOTE. By convention, we use a capital letter, say X, to denote a random variable, and use the corresponding lower-case letter x to denote the realization (values) of the random variable.

EXAMPLE 3.1 (Coin). Consider the random experiment of tossing a coin three times and observing the result (a Head or a Tail) for each toss. Let X denote the total number of heads obtained in the three tosses of the coin.

(a) Construct a table that shows the values of the random variable X for each possible outcome of the random experiment.

(b) Identify the event $\{X \le 1\}$ in words.

Let Y denote the difference between the number of heads obtained and the number of tails obtained.

(c) Construct a table showing the value of Y for each possible outcome.

(d) Identify the event $\{Y = 0\}$ in words.

EXAMPLE 3.2. Suppose that you play a certain lottery by buying one ticket per week. Let W be the number of weeks until you win a prize.

(a) Is W a random variable? Brief explain.

(b) Identify the following events in words: (i) $\{W > 1\}$, (ii) $\{W \le 10\}$, and (iii) $\{15 \le W < 20\}$

EXAMPLE 3.3. A major metropolitan newspaper asked whether the paper should increase its coverage of local news. Let X denote the proportion of readers who would like more coverage of local news. Is X a random variable? What does it mean by saying $0.6 < x < 0.7$?

EXAMPLE 3.4. Denote T be the survival time for the prostate cancer patients in a local hospital. Explain why T is a random variable.

Discrete Random Variable

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers (countable), it is called a **discrete** sample space. A random variable is called a *discrete random variable* if its set of possible outcomes is countable.

Continuous Random Variable

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous** sample space. When a random variable can take on values on a continuous scale, it is called a *continuous random variable*.

EXAMPLE 3.5. Categorize the random variables in the above examples to be discrete or continuous.

3.2 Discrete Probability Distributions

The probability distribution of a discrete random variable X lists the values and their probabilities.

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

where

$$0 \le p_i \le 1 \quad \text{and} \quad p_1 + p_2 + \cdots + p_k = 1$$

Example 3.6. Determine the value of k so that the function $f(x) = k\left(x^2+1\right)$ for $x = 0, 1, 3, 5$ can be a legitimate probability distribution of a discrete random variable.

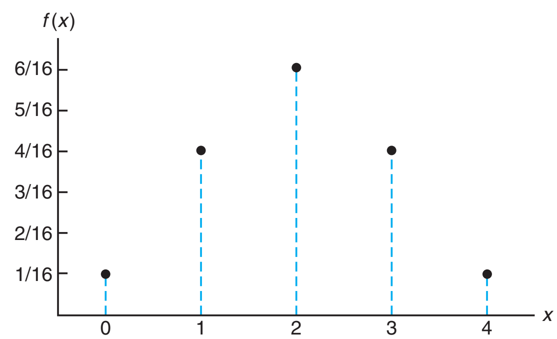
Probability Mass Function (PMF)

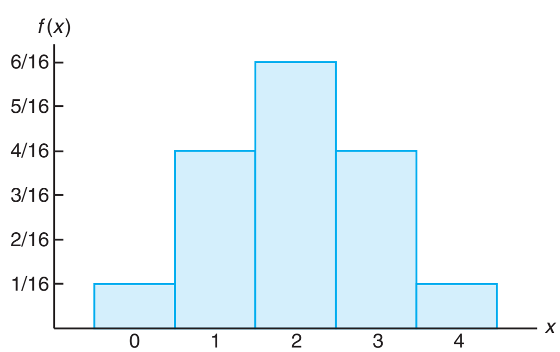
The set of ordered pairs $(x, f(x))$ is a probability function, *probability mass function*, or probability distribution of the discrete random variable X if, for each possible outcome x,

i). $f(x) \geq 0$,

ii). $\sum_{x} f(x) = 1$,

iii). $\mathrm{P}(X = x) = f(x)$.





Example 3.7. Refer to Example 3.1 (Coin). Find a formula for the PMF of X when

(a) the coin is balanced.

(b) the coin has probability 0.2 of a head on any given tosses.

(c) the coin has probability p $(0 \leq p \leq 1)$ of a head on any given toss.

Example 3.8 (Interview). Six men and five women apply for an executive position in a small company. Two of the applicants are selected for interview. Let X denote the number of women in the interview pool.

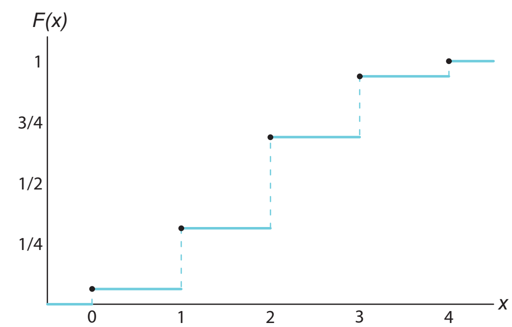
(a) Find the PMF of X, assuming that the selection is done randomly. Plot it.

(b) What is the probability that at least one woman is included in the interview pool?

Cumulative Distribution Function (CDF) of a disc. r.v.

The *cumulative distribution function* $F(x)$ of a discrete random variable X with probability mass function $f(x)$ is

$$\begin{aligned} F(x) &= \mathrm{P}(X \leq x) \\ &= \sum_{t \leq x} f(t), \qquad \text{for } -\infty < x < \infty. \end{aligned}$$



Example 3.9. Given that the CDF

$$F(x) = \begin{cases} 0, & x < 1 \\ 1/3, & 1 \leq x < 2 \\ 1/2, & 2 \leq x < 4 \\ 4/7, & 4 \leq x < 7 \\ 3/4, & 7 \leq x < 10 \\ 1, & x \geq 10 \end{cases}$$

Find

(a) $\mathrm{P}(X \leq 4)$, $\mathrm{P}(X < 4)$ and $\mathrm{P}(X = 4)$

(b) $\mathrm{P}(X \leq 8)$, $\mathrm{P}(X < 8)$ and $\mathrm{P}(X = 8)$

(c) $\mathrm{P}(X > 2)$ and $\mathrm{P}(X > 6)$

(d) $\mathrm{P}(3 < X \leq 5)$

(e) $\mathrm{P}(X \leq 7 | X > 4)$

Example 3.10. Refer to Example 3.8 (Interview).

(a) Find the CDF of X.

(b) Use the CDF to find the probability that exactly two women are included in the interview pool.

(c) Use the CDF to verify 3.8 (b).

3.3 Continuous Probability Distributions

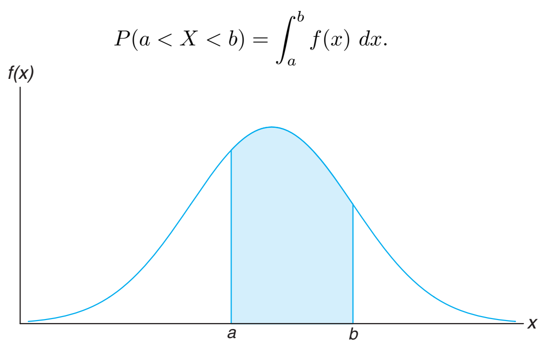
Probability Density Function (PDF)

The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

i). $f(x) \geq 0$ for all $x \in R$.

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

iii). $\mathrm{P}(a < X < b) = \int_a^b f(x)\, dx$.



NOTE. If X is a **continuous** random variable, then

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

This is NOT the case for the *discrete* situation.

EXAMPLE 3.11. Consider

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

This is called a continue uniform distribution Unif$(1,\ 3)$. Calculate

(a) $\mathrm{P}(1.2 \leq X \leq 2.6)$

(b) $\mathrm{P}(X > 2.5)$

(c) $\mathrm{P}(X = 2)$.

EXAMPLE 3.12. Consider

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

(a) Determine the value k that makes f be a legitimate density function.

(b) Evaluate $\mathrm{P}(0 \leq X \leq 1)$.

EXAMPLE 3.13. Consider the density function

$$h(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

Find

(a) $\mathrm{P}(0 < X < 1)$

(b) $\mathrm{P}(X > 5)$

Cumulative Distribution Function (CDF) of a continuous r.v.

The *cumulative distribution function* $F(x)$ of a continuous random variable X with probability density function $f(x)$ is

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

NOTE. A random variable is continuous if and only if its CDF is an everywhere continuous function.

EXAMPLE 3.14. Find the CDF's of the random variables in Examples 3.11, 3.12 and 3.13. Graph them.

3.4 Joint Probability Distributions

We frequently need to examine two more more (discrete or continuous) random variables simultaneously.

3.4.1 Joint, Marginal and Conditional PMFs

Joint Probability Mass Function (Joint PMF)

The function $f(x,y)$ is a *joint* probability function, or *probability mass function* of the discrete random variables X and Y if

i). $f(x,y) \geq 0$ for all (x,y),

ii). $\sum_x \sum_y f(x,y) = 1$,

iii). $\mathrm{P}(X = x, Y = y) = f(x,y)$.

For any region A in the xy plane,

$$\mathrm{P}[(X,Y) \in A] = \sum_A \sum f(x,y).$$

Example 3.15 (Home). Data supplied by a company in Duluth, Minnesota, resulted in the contingency table displayed as below for number of bedroom and number of bathrooms for 50 homes currently for sale. Suppose that one of these 50 homes is selected at random. Let X and Y denote the number of bedrooms and the number of bathrooms, respectively, of the home obtained.

		x		
		2	3	4
y	2	3	14	2
	3	0	12	11
	4	0	2	5
	5	0	0	1

(a) Determine and interpret $f(3,2)$.

(b) Obtain the joint PMF of X and Y.

(c) Find the probability that the home obtained has the same number of bedrooms and bathrooms, i.e., $\mathsf{P}(X=Y)$.

(d) Find the probability that the home obtained has more bedrooms than bathrooms, i.e., $\mathsf{P}(X>Y)$.

(e) Interpret the event $\{X+Y\geq 8\}$ and find its probability.

Marginal Probability Mass Function (Marginal PMF)

The *marginal* distributions of X alone and Y along are, respectively

$$g(x)=\sum_y f(x,y) \quad \text{and} \quad h(y)=\sum_x f(x,y).$$

Example 3.16. Refer to Example 3.15 (Home).

(a) Find the marginal distribution of X alone.

(b) Find the marginal distribution of Y alone.

Example 3.17 (Moive). Two movies are randomly selected from a shelf containing 5 romance movies , 3 action movies, and 2 horror movies. Denote X be the number of romance movies selected and Y be the number of action movies selected.

(a) Find the joint PMF of X and Y.

(b) Construct a table that gives the joint and marginal PMFs of X and Y.

(c) Express that event that no horror movies are selected in terms of X and Y, and then determine its probability.

Conditional Probability Mass Function (Conditional PMF)

Let X and Y be two discrete random variables. The conditional distribution of Y given that $X=x$ is

$$\begin{aligned} f(y|x) &= \mathsf{P}(Y=y|X=x) \\ &= \frac{\mathsf{P}(X=x,Y=y)}{\mathsf{P}(X=x)} \\ &= \frac{f(x,y)}{g(x)}, \end{aligned}$$

provided $\mathsf{P}(X=x)=g(x)>0$.

Similarly, the conditional distribution of X given that $Y=y$ is

$$\begin{aligned} f(x|y) &= \mathsf{P}(X=x|Y=y) \\ &= \frac{\mathsf{P}(X=x,Y=y)}{\mathsf{P}(Y=y)} \\ &= \frac{f(x,y)}{h(y)}, \end{aligned}$$

provided $\mathsf{P}(Y=y)=h(y)>0$.

Example 3.18. Refer to Example 3.15 (Home).

(a) Find the distribution of $X|Y=2$?

(b) Use the result to determine $f(3|2)=\mathsf{P}(X=3|Y=2)$.

Example 3.19. Refer to Example 3.17 (Movie). Find the conditional distribution of Y, given that $X=1$.

3.4.2 Joint, Marginal and Conditional PDFs

Joint Probability Density Function (Joint PDF)

The function $f(x,y)$ is a *joint probability density function* of the continuous random variables X and Y if

i). $f(x,y)\geq 0$ for all (x,y),

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

iii). $\mathsf{P}[(X,Y)\in A]=\int\int_A f(x,y)\,dx\,dy$, for any region A in the xy plane.

Marginal Probability Density Function (Marginal PDF)

The *marginal* distributions of X alone and Y along are, respectively,

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

and

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

Conditional Probability Density Function (Conditional PDF)

Let X and Y be two continuous random variables. The conditional distribution of $Y|X=x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \quad \text{provided } g(x) > 0.$$

Similarly, the conditional distribution of $X|Y=y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad \text{provided } h(y) > 0.$$

EXAMPLE 3.20. Consider the joint probability density function of the random variables X and Y:

$$f(x,y) = \begin{cases} \dfrac{3x-y}{9}, & \text{if } 1<x<3,\ 1<y<2 \\ 0, & \text{elsewhere} \end{cases}.$$

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

(b) Find the marginal distribution of X alone.

(c) Find $\mathrm{P}(1<X<2)$.

(d) Find the conditional distribution of $Y|X=x$.

(e) Find $\mathrm{P}(0.5<Y<1|X=2)$.

EXAMPLE 3.21. Consider the joint density function of the random variables X and Y:

$$f(x,y) = \begin{cases} ke^{-(3x+4y)} & \text{if } x>0,\ y>0 \\ 0 & \text{otherwise} \end{cases}.$$

(a) Determine k.

(b) Find $\mathrm{P}(0<X<1, Y>0)$.

(c) Find the marginal distribution of X alone.

(d) Find $\mathrm{P}(0<X<1)$.

(e) Find the marginal distribution of Y alone.

(f) Find $\mathrm{P}(0<X<1|Y=1)$.

3.4.3 Statistical Independence

For two variables

Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x,y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be *statistically independent* if and only if

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

for all (x,y) within their range.

EXAMPLE 3.22. Determine whether the variables X and Y in Example 3.21 are statistically independent.

NOTE. We can also determine whether X and Y are statistically independent or not by checking if the following holds:

$$f(x|y) = g(x) \quad \text{or} \quad f(y|x) = h(y)$$

EXAMPLE 3.23. Determine whether the variables X and Y in Example 3.20 are statistically independent.

NOTE. If you wish to show X and Y are NOT statistically independent, it suffices to give a pair of (a,b) such that $f(a,b) \neq g(a)h(b)$. However, you must formally verify the definition $f(x,y) = g(x)h(y)$ for ALL the pairs (x,y) to claim the statistical independence.

EXAMPLE 3.24. Revisit Examples 3.15 (Home). Do you think that X and Y are statistically independent?

EXAMPLE 3.25. Consider that the variables X and Y have the following joint probability distribution.

		x		
		0	1	2
y	0	1/18	1/9	1/6
	1	1/9	2/9	1/3

Determine if X and Y are statistically independent.

For three or more variables

Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be *mutually statistically independent* if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

EXAMPLE 3.26. The joint PMF of the random variables X, Y and Z is given by

$$f(x,y,z) = e^{-(\lambda+\mu+\kappa)} \frac{\lambda^x \mu^y \kappa^z}{x!y!z!}, \quad x,y,z = 0,1,2,\ldots$$

and $f(x,y,z) = 0$ otherwise. Are X, Y and Z of statistical independence?

This page is intentionally blank.