Chapter 2
Statistical Background

2.1 Probability Set Function

For many interesting physical problems, we need to describe the "long-term" behavior of systems governed by macroscopic laws and microscopic randomness. A random event has an outcome that is uncertain and unpredictable. Sometimes small changes in initial conditions can result in a substantially different outcome – this is the essence of chaos. Quantities that change randomly in time and space are called stochastic processes. Physical systems that are subject to stochastic driving will have a random component and the variables that describe the system are also stochastic processes. Examples of physical problems include the behavior of gases in the presence of microscopic collisions of the constituent particles, the collective propagation of energetic charged particles in a magnetically turbulent medium, the collective behavior of dust particles in an accretion disk subject to coagulation and destruction, the evolution of a gas of charged protons and electrons (a plasma), etc.

Before considering specific physical problems, some basic statistical concepts need to be reviewed. There are numerous excellent texts that introduce the basic elements of probability theory and stochastic theory and this introductory chapter draws heavily from these. The classic treatise is that of Feller (1968), from which almost all introductory texts draw. Much of this chapter is based on the books by Hogg and Craig (1978) and Gibra (1973).

Suppose we perform n independent experiments under identical conditions. If an outcome A results n_A times, then the probability that A occurs is

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}.$$

More formally, let $\mathcal{C}$ be the set of all possible outcomes of a random experiment. $\mathcal{C}$ is the *sample space*. An outcome is a point or an element in the sample space. Thus,

G.P. Zank, *Transport Processes in Space Physics and Astrophysics*, Lecture Notes in Physics 877, DOI 10.1007/978-1-4614-8480-6_2,

a sample space $\mathcal{C}$ is a set of elements or points, each of which corresponds to an outcome of an experiment or observation. A sample space can be finite or infinite.

Example. Toss a coin twice and denote the outcomes as H (head) and T (tail). The possible outcomes are

$$\{(H, H), (H, T), (T, H), (T, T)\},$$

and so the sample space is given by

$$\mathcal{C} = \{(H, H), (H, T), (T, H), (T, T)\}.$$

The subset E corresponding to the *event* of heads occurring on the first toss contains two elements

$$E = \{(H, H), (H, T)\}.$$

Another subset is the event F where a head occurs on the first and second toss, given by

$$F = \{(H, H)\}.$$

An event E is defined as a set of outcomes, and an event has occurred if the outcome of the experiment corresponds to an element of subset E. A *null event* corresponds to the empty set $\emptyset$, i.e., the set of no outcomes. If the subset E consists of all possible outcomes of the experiment, then E is the sample space (and obviously an event).

Define a *probability set function* $P(C)$ such that if $C \subset \mathcal{C}$, then $P(C)$ is the probability that the outcome of the random experiment is an element of C. We take $P(C)$ to be the number about which the relative frequency n_A/n converges after many experiments. The properties that we want of the probability set function may be defined as follows.

Definition. If $P(C)$ is defined for a subset C of the space $\mathcal{C}$, and $C_1, C_2, C_3, \ldots$ are disjoint subsets of $\mathcal{C}$, then $P(C)$ is called the *probability set function* of the outcome of the random experiment if

(i) $P(C) \geq 0$,
(ii) $P(C_1 \cup C_2 \cup C_3 \cdots) = P(C_1) + P(C_2) + P(C_3) + \cdots$,
(iii) $P(\mathcal{C}) = 1$.

Theorem 1. *For each $C \subset \mathcal{C}$, $P(C) = 1 - P(C^*)$, where C^* denotes the complement of C.*

Proof. Since $\mathcal{C} = C \cup C^*$ and $C \cap C^* = \emptyset$, $1 = P(C) + P(C^*)$.

Theorem 2. $P(\emptyset) = 0$.

Proof. In Theorem 1, take $C = \emptyset$ so that $C^* = \mathcal{C}$. Hence

$$P(\emptyset) = 1 - P(\mathcal{C}) = 1 - 1 = 0.$$

Theorem 3. *If C_1 and C_2 are subsets of $\mathcal{C}$ such that $C_1 \subset C_2$, then $P(C_1) \leq P(C_2)$.*

Proof. $C_2 = C_1 \cup (C_1^* \cap C_2)$ and $C_1 \cap (C_1^* \cap C_2) = \emptyset$. Hence

$$P(C_2) = P(C_1) + P(C_1^* \cap C_2).$$

But $P(C_1^* \cap C_2) \geq 0$, hence $P(C_2) \geq P(C_1)$.

Theorem 4. *For each $C \subset \mathcal{C}$, $0 \leq P(C) \leq 1$.*

Proof. Since $\emptyset \subset C \subset \mathcal{C}$, $P(\emptyset) \leq P(C) \leq P(\mathcal{C})$ or $0 \leq P(C) \leq 1$.

Theorem 5. *If $C_1 \subset \mathcal{C}$ and $C_2 \subset \mathcal{C}$, then*

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2).$$

Proof. Since

$$C_1 \cup C_2 = C_1 \cup (C_1^* \cap C_2) \qquad \text{and} \qquad C_2 = (C_1 \cap C_2) \cup (C_1^* \cap C_2),$$

we have

$$P(C_1 \cup C_2) = P(C_1) + P(C_1^* \cap C_2)$$

and

$$P(C_2) = P(C_1 \cap C_2) + P(C_1^* \cap C_2).$$

Hence

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2).$$

Example. Two coins are tossed and the ordered pairs form the sample space

$$\mathcal{C} = \{c : c = (H, H), (H, T), (T, H), (T, T)\},$$

and let $P(c \in \mathcal{C}) = \frac{1}{4}$. Suppose C_1 is the event that a head is tossed with the first coin and C_2 the event that a head is tossed with the second coin. The events C_1 and C_2 therefore correspond to the subsets

$$C_1 = \{c : c = (H, H), (H, T)\} \quad \text{and} \quad C_2 = \{c : c = (H, H), (T, H)\}.$$

To find the probability that the first coin toss corresponds to a head or the second to head, we compute

$$P(C_1) = P(C_2) = \frac{1}{2},$$

and the probability that tossing the two coins each results in a head is given by

$$P(C_1 \cap C_2) = \frac{1}{4}.$$

The probability that a head is tossed by one or the other coin is given by

$$P(C_1 \cup C_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Exercises

1. A positive integer from 1 to 6 is chosen by casting a die. Thus $\mathcal{C} = \{c : c = 1, 2, 3, 4, 5, 6\}$. Let $C_1 = \{c : c = 1, 2, 3, 4\}$, $C_2 = \{c : c = 3, 4, 5, 6\}$. If $P(c \in \mathcal{C}) = \frac{1}{6}$, find $P(C_1)$, $P(C_2)$, $P(C_1 \cap C_2)$, and $P(C_1 \cup C_2)$.
2. Draw a number without replacement from the set $\{1, 2, 3, 4, 5\}$, i.e., choose a number, and then a second from the remaining numbers, etc. Assume that all 20 possible results have the same probability. Find the probability that an odd digit will be selected (a) the first time, (b) the second time, and (c) both times.
3. Draw cards from an ordinary deck of 52 cards and suppose that the probability set function assigns a probability of $\frac{1}{52}$ to each of the possible outcomes. Let C_1 denote the collection of 13 hearts and C_2 the collection of 4 kings. Compute $P(C_1)$, $P(C_2)$, $P(C_1 \cap C_2)$ and $P(C \cup C_2)$.
4. A coin is tossed until a head results. The elements of the sample space $\mathcal{C}$ are therefore H, TH, TTH, TTTH, TTTTH, etc. The probability set function assigns probabilities $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, etc. Show that $P(\mathcal{C}) = 1$. Suppose $C_1 = \{c : c \text{ is } H, TH, TTH, TTTH, \text{ or } TTTTH\}$ and $C_2 = \{c : c \text{ is } TTTTH \text{ or } TTTTTH\}$. Find $P(C_1)$, $P(C_2)$, $P(C_1 \cap C_2)$, and $P(C_1 \cup C_2)$.
5. A coin is tossed until for the first time the same result appears twice in succession. Let the probability for each outcome requiring n tosses be $1/2^{n-1}$. Describe the sample space, and find the probability of the events (a) the tosses end before the sixth toss, (b) an even number of tosses is required.
6. Find $P(C_1 \cap C_2)$ if the sample space is $\mathcal{C} = C_1 \cup C_2$, $P(C_1) = 0.8$ and $P(C_2) = 0.5$.
7. Suppose $C \subset \mathcal{C} = \{c : 0 < c < \infty\}$ with $C = \{c : 4 < c < \infty\}$ and $P(C) = \int_C e^{-x} dx$. Determine $P(C)$, $P(C^*)$, and $P(C \cup C^*)$.
8. If $C \subset \mathcal{C}$ is a set for which $\int_C e^{-|x|} dx$ exists, $\mathcal{C} = \{c : -\infty < x < \infty\}$, then show that this set function is *not* a probability set function. What constant should the integral be multiplied by to make it a probability set function?
9. If $C_1 \subseteq \mathcal{C}$ and $C_2 \subseteq \mathcal{C}$ of the sample space $\mathcal{C}$, show that

$$P(C_1 \cap C_2) \leq P(C_1) \leq P(C_1 \cup C_2) \leq P(C_1) + P(C_2).$$

2.2 Random or Stochastic Variables

As discussed above, elements of a sample space $\mathcal{C}$ may not be numbers, being outcomes such as "heads" or "tails". Since we are typically interested in quantifying the outcome of an experiment, we formulate a rule by which elements $c \in \mathcal{C}$ may be represented by numbers x, pairs or n-tuples of numbers $(x_1, x_2, \ldots, x_n)$.

Definition. Consider a random experiment with a sample space $\mathcal{C}$. A function X that assigns to each $c \in \mathcal{C}$ one and only one real number $x = X(c)$ is a *random variable*, and the *space* of X is the set of real numbers $\mathcal{A} = \{x : x = X(c), c \in \mathcal{C}\}$.

Example. Coin toss: $\mathcal{C} = \{c$: where c is T or $H\}$, and $T \equiv$ Tail, $H \equiv$ Head. Define a function X such that

$$X = \begin{cases} 0 \text{ if } & c = T \\ 1 \text{ if } & c = H \end{cases}$$

Therefore X is a real-valued function defined on $\mathcal{C}$ which maps $c \in \mathcal{C}$ to a set of real numbers $\mathcal{A} = \{x : x = 0, 1\}$. X is a random variable and the space associated with X is $\mathcal{A}$.

Sometimes the set $\mathcal{C}$ has elements that are real numbers, so that if we write $X(c) = c$, then $\mathcal{A} = \mathcal{C}$.

Two forms of random variable can be defined. (1) *Discrete random variables* are those that take on a finite or denumerably infinite number of distinct values. (2) *Continuous random variables* are those that take on a continuum of values within the given range. Random variables are generally denoted by capital Latin letters such as X, Y, Z. Some examples of discrete and continuous random variables are the daily demand for coffee at a Starbucks (discrete), the number of customers at a checkout per hour (discrete), the daily number of absences from a company (discrete), the waiting time of a passenger for a train at a particular train station (continuous), the daily consumption of gas by your car (continuous), and the annual snowfall in Alabama (continuous).

Example. A vendor at a rugby game buys "koeksisters" (a form of unfilled donut from South Africa) for \$1.00 each and sells them for \$2.50 each. Unsold koeksisters cannot be returned. Suppose the demand during a game is a random variable Y. Suppose the vendor orders a quantity X of koeksisters. Let F denote the profit after the game, which can then be computed as

$$F(Y) = \begin{cases} \$1.50X & Y \geq X \\ \$1.50Y - \$1.00(X - Y) & Y < X \end{cases}$$

where $Y = 1, 2, \ldots$.

Example. If a die is rolled twice, the random variable X that describes the sum of the values is $X(c) = c$ where $c = 2, 3, \ldots, 12$.

Example. A sample of five items is drawn randomly from a lot. The random variable X that describes the number of defective items in the sample is $X(c) = c$, where $c = 0, 1, 2, 3, 4, 5$. If the random variable Y is the number of non-defective items in the sample, define the random variable $Z = |X - Y|$. Thus, the random variable $Z(c) = c$ where $c = 1, 3, 5$.

Just as we refer to an "event C" with $C \subset \mathcal{C}$, we can introduce an event A. Like the definition for the probability $P(C)$, we define the probability of the event A, $P(X \in A)$. With $A \subset \mathcal{A}$, let $C \subset \mathcal{C}$ such that $C = \{c : c \in \mathcal{C}$ and $X(c) \in \mathcal{A}\}$. Thus C has as its elements all outcomes in $\mathcal{C}$ for which the random variable X has a value that is in A. This means that we can define $P(A)$ to be equal to be $P(C)$ where $C = \{c : c \in \mathcal{C}$ and $X(c) \in \mathcal{A}\}$. This allows us to use the same notation without confusion.

That $P(A)$ is a probability set function can be seen as follows. For condition (i) above, $P(A) = P(C) \geq 0$.

Consider two mutually exclusive events A_1 and A_2. Here

$$P(A_1 \cup A_2) = P(C),$$

where $C = \{c : c \in \mathcal{C}$ and $X(c) \in A_1 \cup A_2\}$. However,

$$C = \{c : c \in \mathcal{C} \text{ and } X(c) \in A_1\} \cup \{c : c \in \mathcal{C} \text{ and } X(c) \in A_2\} = C_1 \cup C_2,$$

say. Since C_1 and C_2 disjoint, we have

$$P(C) = P(C_1) + P(C_2) = P(A_1) + P(A_2),$$

which is condition (ii).

Finally, since $\mathcal{C} = \{c : c \in \mathcal{C}$ and $X(c) \in \mathcal{A}\}$, it implies that $P(\mathcal{A}) = P(\mathcal{C}) = 1$.

Example. Let a coin be tossed twice, and consider the number of heads observed. The sample space is

$$\mathcal{C} = \{c : \text{where } c = TT, TH, HT, HH\}$$

$$X(c) = \begin{cases} 0 \text{ if c = TT} \\ 1 \text{ if } c = TH \text{ or } HT \\ 2 \text{ if } c = HH \end{cases}$$

Hence, $\mathcal{A} = \{x : X = 0, 1, 2\}$.

Let $A \subset \mathcal{A}$ such that $A = \{x : x = 1\}$. What is $P(A)$?

Since $X(c) = 1$ if c is TH or HT,

$$\Rightarrow C \subset \mathcal{C} \quad \text{such that} \quad C = \{c : c = TH \text{ or } HT\}$$

$$\Rightarrow P(A) = P(C) \qquad \text{i.e.,} \qquad P(X = 1) = P(C).$$

Define

$$C_1 = \{c : c = TT\}$$
$$C_2 = \{c : c = TH\}$$
$$C_3 = \{c : c = HT\}$$
$$C_4 = \{c : c = HH\}$$

Suppose $P(C)$ assigns a probability of $\frac{1}{4}$ to each C_i. Then $P(C_1) = \frac{1}{4}$, $P(C_2 \cup C_3) = \frac{1}{2}$, and $P(C_4) = \frac{1}{4}$, so that $P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$, and $P(X = 2) = \frac{1}{4}$.

Example. An experiment yields a random value in the interval (0, 1), so the sample space is $\mathcal{C} = \{c : 0 < c < 1\}$. Let the probability set function be given by the "length"

$$P(C) = \int_C dc,$$

so, for example, if $C = \{c : \frac{1}{3} < c < \frac{2}{3}\}$ then

$$P(C) = \int_{1/3}^{2/3} dc = \frac{1}{3}.$$

Define a random variable $X = X(c) = 2c + 1$, so that the space $\mathcal{A} = \{x : 1 < x < 3\}$. For $A \subset \mathcal{A}$, such that e.g., $A = \{x : a < x < b, a > 1, b < 3\}$, we have $C = \{c : (a-1)/2 < c < (b-1)/2, a > 1, b < 3\}$. Hence

$$P(A) = P(C) = \int_{(a-1)/2}^{(b-1)/2} dc = \int_a^b \frac{1}{2} dx.$$

Typically, however, we assume a probability distribution for the random variable X rather than introducing the sample space $\mathcal{C}$ and the probability set function $P(C)$.

Example. Suppose the probability set function $P(A)$ of a random variable X is

$$P(A) = \int_A f(x)dx \qquad \text{where} \qquad f(x) = 2x, \quad x \in \mathcal{A} = \{x : 0 < x < 1\}.$$

$A_1 = \{x : 0 < x < \frac{1}{4}\}$ and $A_2 = \{x : \frac{1}{2} < x < \frac{3}{4}\}$ are subsets of $\mathcal{A}$. Then

$$P(A_1) = P(X \in A_1) = \int_0^{1/4} 2x dx = \frac{1}{16}$$

and

$$P(A_2) = P(X \in A_2) = \int_{1/2}^{3/4} 2x dx = \frac{5}{16}.$$

Hence, it follows that since $A_1 \cap A_2 = \emptyset$, $P(A_1 \cup A_2) = P(A_1) + P(A_2) = \frac{3}{8}$.

Example. Consider two random variables X and Y, and let $\mathcal{A} = \{(x, y) : 0 < x < y < 1\}$ be the 2-space. Suppose the probability set function is

$$P(A) = \int \int_A 2dxdy.$$

If $A_1 = \{(x, y) : \frac{1}{2} < x < y < 1\}$, then

$$P(A_1) = P[(X, Y) \in A_1] = \int_{1/2}^{1} \int_{1/2}^{y} 2dxdy = \int_{1/2}^{1} (2y-1)dy = y^2 - y|_{1/2}^{1} = \frac{1}{4}.$$

Suppose $A_2 = \{(x, y) : x < y < 1, 0 < x \leq \frac{1}{2}\}$, then $A_2 = A_1^*$, and

$$P(A_2) = P[(X, Y) \in A_2] = P(A_1^*) = 1 - P(A_1) = \frac{3}{4}.$$

Exercises

1. Select a card from a standard deck of 52 playing cards with outcome c. Let $X(c) = 4$ if c is an ace, $X(c) = 3$ for a king, $X(c) = 2$ for a queen, $X(c) = 1$ for a jack, and $X(c) = 0$ otherwise. Suppose $P(C)$ assigns a probability $\frac{1}{52}$ to each outcome c. Calculate the probability $P(A)$ on the space $\mathcal{A} = \{x : x = 0, 1, 2, 3, 4\}$ of the random variable X.
2. Suppose the probability set function $P(A)$ of the random variable X is $P(A) = \int_A f(x)dx$ where $f(x) = 2x/9$, $x \in \mathcal{A} = \{x : 0 < x < 3\}$. For $A_1 = \{x : 0 < x < 1\}$ and $A_2 = \{x : 2 < x < 3\}$, compute $P(A_1)$, $P(A_2)$, and $P(A_1 \cup A_2)$.
3. Suppose that the random variable X has space $\mathcal{A} = \{x : 0 < x < 1\}$. If $A_1 = \{x : 0 < x < \frac{1}{2}\}$ and $A_2 = \{x : \frac{1}{2} \leq x < 1\}$, find $P(A_2)$ if $P(A_1) = \frac{1}{4}$.

2.3 The Probability Density Function

The distribution of the random variable X refers to the distribution of probability, and this applies even when more than one random variable is involved. We discuss some random variables whose distributions can be described by a *probability density function* of both the *discrete* and *continuous* type. Consider first probability distribution functions (pdfs) of one random variable. Suppose X denotes a random variable with one-dimensional space $\mathcal{A}$ such that $\mathcal{A}$ is a set of discrete points. Let $f(x)$ be a one-to-one function $f(x) > 0$, $x \in \mathcal{A}$ with

$$\sum_{\mathcal{A}} f(x) = 1.$$

Whenever a probability set function $P(A)$, $A \subset \mathcal{A}$, can be expressed as

$$P(A) = \sum_{A} f(x),$$

then X is a random variable of the *discrete type*, and X has a *discrete distribution*.

Example. Consider a discrete random variable X with space $\mathcal{A} = \{x : x = 0, 1, 2, 3\}$, and let

$$P(A) = \sum_{A} f(x),$$

where

$$f(x) = \frac{3!}{x!(3-x)!}\left(\frac{1}{2}\right)^3, \qquad x \in \mathcal{A},$$

(recall $0! = 1$). If $A = \{x : x = 0, 1, 2\}$, then

$$\Rightarrow P(X \in A) = \frac{3!}{0!3!}\left(\frac{1}{2}\right)^3 + \frac{3!}{1!2!}\left(\frac{1}{2}\right)^3 + \frac{3!}{2!1!}\left(\frac{1}{2}\right)^3 = \frac{7}{8}.$$

Note that $P(\mathcal{A}) = 1$.

Example. Consider a discrete random variable X with space $\mathcal{A} = \{x : x = 0, 1, 2, 3, \ldots\}$, and let

$$f(x) = \left(\frac{1}{2}\right)^x, \qquad x \in \mathcal{A}.$$

Thus, $P(X \in A) = \sum_A f(x)$. For $A = \{x : x = 1, 3, 5, 7, \ldots\}$,

$$P(X \in A) = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots = \frac{2}{3}.$$

Suppose that the one-dimensional Riemann integral over the space $\mathcal{A}$ satisfies

$$\int_{\mathcal{A}} f(x)dx = 1,$$

where $f(x)$ is a one-to-one function $f(x) > 0$, $x \in \mathcal{A}$ with at most a finite number of discontinuities in every finite subset (interval) of $\mathcal{A}$. Whenever a probability set function $P(A)$, $A \subset \mathcal{A}$, can be expressed as

$$P(A) = P(X \in A) = \int_{A} f(x)dx,$$

then X is a random variable of the *continuous type*, and X has a *continuous distribution*.

Example. Let $\mathcal{A} = \{x : 0 < x < \infty\}$ and $f(x) = ae^{-3x}$, $x \in \mathcal{A}$. The probability set function is

$$P(X \in A) = \int_{A} f(x)dx = \int_{A} ae^{-3x}dx.$$

Since $P(\mathcal{A}) = 1$,

$$P(\mathcal{A}) = \int_{0}^{\infty} ae^{-3x}dx = 1 \Longrightarrow a = 3.$$

If $A = \{x : 0 < x < 1\}$, then

$$P(X \in A) = \int_{0}^{1} 3e^{-3x}dx = 1 - e^{-3}.$$

The probability $P(A)$ is determined completely by the *probability density function (pdf)* $f(x)$, whether or not X is a discrete or continuous random variable.

The concept of the pdf of one random variable is readily extended to the pdf of multiple random variables. For example, suppose the two random variables X and Y are discrete or continuous and have a distribution such that the probability set function $P(A)$, $A \subset \mathcal{A}$ can be expressed as

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

or

$$P(A) = P[(X, Y) \in A] = \int\int_{A} f(x, y)dxdy.$$

In either case, $f(x, y)$ is the pdf of the two random variables X and Y. Of course, $P(\mathcal{A}) = 1$.

Suppose that the space of a continuous random variable X is $\mathcal{A} = \{x : 0 < x < \infty\}$ and that the pdf is $3e^{-3x}$, $x \in \mathcal{A}$. We can write

$$f(x) = \begin{cases} 3e^{-3x} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and $f(x)$ is the pdf of X. Hence

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{\infty} 3e^{-3x}dx = 1.$$

If $A \subset \mathcal{A}$ such that $A = \{x : a < x < b\}$, then

$$P(A) = P(a < x < b) = \int_{a}^{b} f(x)dx.$$

If $A = \{x : x = a\}$, then $P(A) = 0$, which implies that $P(a < x < b) = P(a \le x \le b)$.

Example. Suppose the random variable X has pdf

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

To find $P(\frac{1}{4} < X < \frac{1}{2})$, we evaluate

$$P(\frac{1}{4} < X < \frac{1}{2}) = \int_{1/4}^{1/2} f(x)dx = \int_{1/4}^{1/2} 3x^2dx = \frac{3}{32}.$$

Similarly,

$$P(-\frac{1}{2} < X < \frac{1}{2}) = \int_{-1/2}^{1/2} f(x)dx = \int_{-1/2}^{0} 0dx + \int_{0}^{1/2} 3x^2dx = \frac{1}{8}.$$

Example. Let

$$f(x, y) = \begin{cases} 6x^2y & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

be the pdf of two random variables X, Y. For example,

$$P(0 < X < \frac{3}{4}, \frac{1}{3} < Y < 2) = \int_{1/3}^{2} \int_{0}^{3/4} f(x, y) dx dy$$
$$= \int_{1/3}^{1} \int_{0}^{3/4} 6x^2 y dx dy + \int_{1}^{2} \int_{0}^{3/4} 0 dx dy$$
$$= \frac{3}{8} + 0 = \frac{3}{8}.$$

Exercises

1. Find the constant a that ensures that $f(x)$ is a pdf of the random variable X: (a) $f(x) = a\left(\frac{2}{3}\right)^x$, $x = 1, 2, 3, \ldots$, 0 elsewhere. (b) $f(x) = axe^{-x}$, $0 < x < \infty$, 0 elsewhere.
2. Consider a function of the random variable X such that

$$f(x) = \begin{cases} ax & 0 \leq x < 10 \\ a(20 - x) & 10 \leq x < 20 \\ 0 & \text{elsewhere} \end{cases}$$

   Find a so that $f(x)$ is a pdf and sketch the graph of the pdf. Compute $P(X \geq 10)$ and $P(15 \leq X \leq 20)$.
3. Let $f(x) = x/15$, $x = 1, 2, 3, 4, 5$, 0 elsewhere, be the pdf of X. Find $P(X = 1 \text{ or } 2)$, $P(\frac{1}{2} < X < \frac{5}{2})$, and $P(1 \leq X \leq 2)$.
4. Compute $P(|X| < 1)$ and $P(X^2 < 9)$ for the following pdfs of X, (a) $f(x) = x^2/18$, $-3 < x < 3$, 0 elsewhere. (b) $f(x) = (x + 2)/18$, $-2 < x < 4$, 0 elsewhere.
5. Given $P(X > a) = e^{-\lambda a}(\lambda a + 1)$, $\lambda > 0$, $a \geq 0$, find the pdf of X and $P(X > \lambda^{-1})$.
6. Let $f(x) = x^{-2}$ $1 < x < \infty$, 0 elsewhere, be the pdf of X. If $A_1 = \{x : 1 < x < 2\}$ and $A_2 = \{x : 4 < x < 5\}$, find $P(A_1 \cup A_2)$ and $P(A_1 \cap A_2)$.
7. Let $f(x, y) = 4xy$, $0 < x < 1$, $0 < y < 1$, 0 elsewhere, be the pdf of X and Y. Find $P(0 < X < \frac{1}{2}, \frac{1}{4} < Y < 1)$, $P(X = Y)$, $P(X < Y)$, and $P(X \leq Y)$.
8. Given that the random variable X has the pdf

$$f(x) = \begin{cases} \frac{5}{a} & -0.1a < x < 0.1a \\ 0 & \text{elsewhere} \end{cases}$$

   and $P(|X| < 2) = 2P(|X| > 2)$, find the value of a.

2.4 The Distribution Function

Suppose a random variable X has the probability set function $P(A)$, and A is a 1D set. For a real number x, let $A = \{y : -\infty < y \le x\}$, so that $P(A) = P(X \in A) = P(X \le x)$. The probability is thus a function of x, say $F(x) = P(X \le x)$. The function $F(x)$ is the *distribution function* of the random variable X. Hence, if $f(x)$ is the pdf of X, we have for a discrete random variable X

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

and for a continuous random variable X

$$F(x) = \int_{y \le x} f(y)dy.$$

The distribution function is therefore of a discrete or continuous type.

Example. A die is rolled once. The sample space is $\mathcal{C} = \{1, 2, 3, 4, 5, 6\}$. To find the probability that the value on the upturned face is less than or equal to three, let X be the random variable whose value is less than or equal to three. Thus, the event $X \le 3$ is the subset $C = \{1, 2, 3\} \subset \mathcal{C}$. The distribution function

$$F(3) = P(X \le 3) = P(C) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Sketch the graph of the distribution function.

A distribution function $F(x)$ possesses the following properties.

(i) $0 \le F(x) \le 1$ because $0 \le P(X \le x) \le 1$.

(ii) $F(x)$ is a non-decreasing function of x. This can be seen as follows. Suppose $x_1 < x_2$. Then

$$\{x : x \le x_2\} = \{x : x \le x_1\} \cup \{x : x_1 < x \le x_2\}$$

and

$$P(X \le x_2) = P(X \le x_1) + P(x_1 < X \le x_2),$$

from which it follows that

$$F(x_2) - F(x_1) = P(x_1 < X \le x_2) \ge 0.$$

(iii) $F(\infty) = 1$ and $F(-\infty) = 0$ ($\{x : x \le -\infty\} = \emptyset$).

(iv) If $a < b$, then, from (ii),

$$P(a < X < b) = F(b) - F(a).$$

Without proof,

$$P(X = b) = F(b) - F(b-),$$

where $F(b-)$ denotes the left-hand limit of $F(x)$ at $x = b$. If the distribution is continuous at $x = b$, then $P(X = b) = 0$.

(v) Also without proof, $F(x)$ is continuous to the right at each point x.

Example. The distribution function of a random variable X is given by

$$F(x) = \begin{cases} 0 \ x < 0 \\ \frac{1}{4} \ 0 \le x < 1 \\ \frac{2}{3} \ 1 \le x < 2 \\ 1 \ x \ge 2 \end{cases}$$

Thus,

$$\begin{aligned} P(X \le 1) &= F(1) = \frac{2}{3}; \\ P(0 < X \le 2) &= F(2) - F(0) = 1 - \frac{1}{4} = \frac{3}{4}; \\ P(1 < X \le 2) &= F(2) - F(1) = 1 - \frac{2}{3} = \frac{1}{3}; \\ P(1 \le X \le 2) &= F(2) - F(0) = 1 - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Example. Suppose the random variable X is continuous with pdf $f(x) = e^{-x}$, $0 < x < \infty$, 0 elsewhere. The distribution function of X is therefore

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} 0 dy, \qquad x < 0 \\ &= \int_{0}^{x} e^{-y} dy = 1 - e^{-x}, \qquad 0 \le x \end{aligned}$$

Sketch this function. $F(x)$ is a continuous function for all real x, and the derivative exists at all points except $x = 0$.

Example. Let a distribution function be given by (sketch this function)

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x+1}{2}, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$$

so that

$$P(-3 < X \le \frac{1}{2}) = F\left(\frac{1}{2}\right) - F(-3) = \frac{3}{4} - 0 = \frac{3}{4},$$

for example, and

$$P(X = 0) = F(0) - F(0-) = \frac{1}{2} - 0 = \frac{1}{2}.$$

Note that $F(x)$ is a mixture of both discrete and continuous type distributions.

Suppose X is a random variable with space $\mathcal{A}$, and $Y = g(X)$ is a function of X. $Y = g(X)$ is a random variable with space $\mathcal{B} = \{y : y = g(x), x \in \mathcal{A}\}$ and a probability set function. If $y \in \mathcal{B}$, the event $Y = g(X) \leq y$ occurs when and only when the event $X \in A \subset \mathcal{A}$ occurs, where $A = \{x : g(x) \leq y\}$. The distribution function of Y is therefore

$$G(y) = P(Y \leq y) = P(g(X) \leq y) = P(A).$$

Example. Suppose a random variable X has the pdf $f(x) = x + \frac{1}{2}$, $0 < x < 1$, 0 elsewhere. The distribution function of X is given by

$$F_X(x) = \int_0^x \left(y + \frac{1}{2}\right) dy = \frac{1}{2}(x^2 + x),$$

so that

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}(1 + x), & 0 \leq x \leq 1 \\ 1, & x \geq 1 \end{cases}$$

Consider now the composite random variable $Y = X^2$. The distribution function of Y is

$$F_Y(a) = P(Y \leq a) = P(X^2 \leq a) = P(-\sqrt{a} \leq X \leq \sqrt{a}).$$

Y is always non-negative,

$$F_Y(a) = P(X \leq \sqrt{a}) = F_X(\sqrt{a}),$$

therefore

$$F_Y(a) = \begin{cases} 0, & a < 0 \\ \frac{\sqrt{a}}{2}(1 + \sqrt{a}), & 0 \leq a \leq 1 \\ 1, & a \geq 1 \end{cases}$$

Differentiation of F_Y yields the pdf as

$$f_Y(y) = \begin{cases} \frac{1}{2}\left(1 + \frac{1}{2\sqrt{y}},\right) & 0 \leq y \leq 1 \\ 0. & \text{elsewhere} \end{cases}$$

Finally, for example,

$$P(Y > 0.49) = 1-P(Y \leq 0.49) = 1-F_Y(0.49) = 1-\frac{1}{2}(0.49-\sqrt{0.49}) = 0.405.$$

Extending these ideas to two or more random variables e.g., X and Y, is readily accomplished. Let $P(A)$ be the probability set function of X and Y, where A is a 2D set. If $A = \{(u, v) : u \leq x, v \leq y\}$, x, y real numbers, then

$$P(A) = P[(X, Y) \in A] = P(X \leq x, Y \leq y).$$

The *distribution function* of X and Y is

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

If X and Y are continuous type random variables with pdf $f(x, y)$, then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v) du dv,$$

so that

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

at continuous points.

Example. The random variables X, Y, and Z have the pdf $f(x, y, z) = e^{-(x+y+z)}$, $0 < x, y, z < \infty$, 0 elsewhere. The distribution function of X, Y, and Z, is

$$\begin{aligned} F(x, y, z) &= P(X \leq x, Y \leq y, Z \leq z) \\ &= \int_0^z \int_0^y \int_0^x e^{-(u+v+w)} du dv dw \\ &= (1 - e^{-x})(1 - e^{-y})(1 - e^{-z}), \qquad 0 \leq x, y, z < \infty, \end{aligned}$$

and 0 elsewhere.

Exercises

1. Let $f(x)$ be the pdf of a random variable X. Find the distribution function $F(x)$ of X and sketch the graph

   $f(x) = 1$, $x = 0$, 0 elsewhere.
   $f(x) = 1/3$, $x = -1, 0, 1$, 0 elsewhere.
   $f(x) = x/15$, $x = 1, 2, 3, 4, 5$, 0 elsewhere.
   $f(x) = 3(1 - x)^2$, $0 < x < 1$, 0 elsewhere.
   $f(x) = x^{-2}$, $1 < x < \infty$, 0 elsewhere.
   $f(x) = 1/3$, $0 < x < 1$ and $2 < x < 4$, 0 elsewhere.

2. Given the distribution function

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{x+2}{4}, & -1 \le x < 1 \\ 1, & 1 \le x \end{cases}$$

   Sketch $F(x)$ and compute $P(-\frac{1}{2} < X \le \frac{1}{2})$, $P(X = 0)$, $P(X = 1)$, and $P(2 < X \le 3)$.
3. Suppose the random variable X has a distribution function $F(x) = 1 - e^{-0.1x}$, $x \ge 0$. Find (a) the pdf of X, (b) $P(X > 100)$, and (c) let $Y = 2X + 5$ and find the corresponding distribution functionF_Y.
4. Let $f(x) = 1$, $0 < x < 1$, 0 elsewhere be the pdf of X. Find the distribution function and pdf of $Y = \sqrt{X}$.
5. Let $f(x) = x/6$, $x = 1, 2, 3$, 0 elsewhere, be the pdf of X. Find the distribution function and pdf of $Y = X^2$.
6. Suppose that a random variable X has the pdf

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

   Determine (a) the distribution function $F(x)$; (b) the probability density function of the random variable $Y = X^2$, and (c) compute $P(Y > 0.36)$.

2.5 Expectations and Moments

One of the most important concepts needed for the transport theory of almost any physical system is that of moments. This section provides the foundation for the remaining chapters. Suppose X is a continuous or discrete random variable with pdf $f(x)$ and let $u(X)$ be a function of X such that

$$E[u(x)] = \int_{-\infty}^{\infty} u(x) f(x) dx, \qquad \text{or} \qquad E[u(x)] = \sum_x u(x) f(x)$$

exists, then $E[u(x)]$ is called the *mathematical expectation* or *expected value* of $u(x)$.

More generally, if $X_1, X_2, \ldots, X_n$ are continuous random variables with *joint pdf* $f(x_1, x_2, \ldots x_n)$, then for the function $u(X_1, X_2, \ldots X_n)$, the expectation is defined by

$$E[u(x_1, x_2, \ldots x_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, x_2, \ldots x_n) f(x_1, x_2, \ldots x_n) dx_1 dx_2 \ldots dx_n.$$

The case for discrete random variables is similarly defined.

The expectation possesses the following properties:

1. For a constant k, $E(k) = \int_{-\infty}^{\infty} kf(x)dx = k\int_{-\infty}^{\infty} f(x)dx = k$.
2. For a constant k, $E(ku) = kE(u)$.
3. For constants k_1, k_2, $E(k_1u_1 + k_2u_2) = k_1E(u_1) + k_2E(u_2)$.
4. For constants c and k, $E[(ku)^c] = k^cE(u^c)$ or $E(kX)^c = k^cE(X^c)$.

Properties (1)–(3) demonstrate that E is a linear operator.

Example. Suppose X has the pdf

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

Corresponding expectations are

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x2xdx = \frac{2}{3} \\ E(X^2) &= \int_{-\infty}^{\infty} x^2f(x)dx = \int_0^1 x^2 2xdx = \frac{1}{2} \\ E(X + 2X^2) &= \frac{2}{3} + 2\frac{1}{2} = \frac{5}{3}. \end{aligned}$$

Example. Suppose X and Y have the joint pdf

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

Then

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^1 xy(x + y)dxdy \\ &= \int_0^1 \left(\frac{1}{3}y + \frac{1}{2}y^2\right) dy = \frac{1}{3}. \end{aligned}$$

Similarly

$$\begin{aligned} E(XY^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy^2 f(x, y)dxdy \\ &= \int_0^1 \int_0^1 xy^2(x + y)dxdy = \frac{17}{12}. \end{aligned}$$

Example. Suppose a rod of length three units is randomly divided into two parts. If X is the length of the left-hand piece, we might assume that X has the pdf

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

Note that

$$\int_0^3 f(x)dx = 1.$$

Hence the expected value of the length X is

$$E(X) = \int_0^3 x\frac{1}{3}dx = \frac{3}{2} \qquad \text{and} \qquad E(3-X) = \frac{3}{2}.$$

However, note that the expected product of the two lengths is

$$E[X(3-X)] = \int_0^3 x(3-x)\frac{1}{3}dx = \frac{9}{6} = \frac{3}{2} \neq \left(\frac{3}{2}\right)^2,$$

illustrating that the expected value of a product is not the product of expected values.

For a random variable X that is either continuous or discrete with pdf $f(x)$, let $u(X) = X$. This defines the *mean value* μ of X with

$$\mu = E(X) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx & \text{continuous} \\ \sum_x xf(x) & \text{discrete} \end{cases}$$

The *variance* σ^2 of X is obtained by taking $u(X) = (X-\mu)^2$ i.e.,

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

for a continuous random variable (an analogous definition for a discrete random variable holds). The *standard deviation* of X is simply σ. Observe that since E is a linear operator,

$$\begin{aligned}\sigma^2 = E[(X-\mu)^2] &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2.\end{aligned}$$

The variance σ^2 of a random variable represents a measure of the variability of observations or fluctuations about the mean μ. If a random variable has a small variance or standard deviation, then most of the values are grouped around the mean. We may therefore expect the probability that a random variable assumes a value within an interval about the mean is greater than for a similar random variable with a larger variance. A useful but rather weak result for estimating the probability of a random variable falling within n standard deviations of its mean is the following inequality.

Theorem (Chebyshev's inequality). *The probability that any random variable X falls within n standard deviations of the mean is at least* $(1 - n^{-2})$*, or equivalently*

$$P(\mu - n\sigma < X < \mu + n\sigma) \geq 1 - n^{-2}.$$

Proof. Since

$$\begin{aligned}
\sigma^2 = E[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\mu-n\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-n\sigma}^{\mu+n\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+n\sigma}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-n\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+n\sigma}^{\infty} (x-\mu)^2 f(x)dx
\end{aligned}$$

since the middle integral is non-negative. Because $|x - \mu| \geq n\sigma$ whenever $x \geq \mu + n\sigma$ or $x \leq \mu - n\sigma$, we have $(x-\mu)^2 \geq n^2\sigma^2$ in both remaining integrals. Thus,

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

from which we obtain

$$\int_{-\infty}^{\mu-n\sigma} f(x)dx + \int_{\mu+n\sigma}^{\infty} f(x)dx \leq n^{-2}.$$

Hence we have established

$$P(\mu - n\sigma < X < \mu + n\sigma) = \int_{\mu-n\sigma}^{\mu+n\sigma} f(x)dx \geq 1 - n^{-2}.$$

By way of example, for $n = 2$, the random variable X has a probability of at least $1 - (2)^{-2} = 3/4$ of being within two standard deviations of the mean, or equivalently, that 3/4 or more of the observations of any distribution fall in the interval $\mu \pm 2\sigma$.

Example. Suppose a random variable X with an unknown probability distribution has a mean $\mu = 8$, a variance $\sigma^2 = 9$. Then, for example,

$$P(-4 < X < 20) = P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}.$$

Consider instead

$$P(|X-8| \geq 6) = 1 - P(|X-8| < 6) = 1 - P(-6 < X - 8 < 6)$$
$$= 1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}.$$

These values are lower bounds only.

Another important concept is the covariance and the related correlation function since it enables the introduction of the idea of statistical independence and stationarity of a random variable. These concepts are particularly important for the statistical description of turbulence, for example. Let X, Y, and Z be random variables with joint pdf $f(x, y, z)$. The mathematical expectation

$$\begin{aligned} E[(X-\mu_1)(Y-\mu_2)] &= E(XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2) \\ &= E(XY) - \mu_2 E(X) - \mu_1 E(Y) + \mu_1\mu_2 \\ &= E(XY) - \mu_1\mu_2, \end{aligned}$$

is the *covariance* of X and Y. Here μ_1, μ_2, μ_3, σ_1^2, σ_2^2, and σ_3^2 denote the means and variances of X, Y, and Z respectively. Similarly, the covariance of X and Z is $E[(X-\mu_1)(Z-\mu_3)]$ and the covariance of Y and Z is $E[(Y-\mu_2)(Z-\mu_3)]$. If the standard deviations $\sigma_1 > 0$ and $\sigma_2 > 0$, we define the *correlation coefficient* of X and Y as

$$\rho_{12} = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2}.$$

In general, if the standard deviations are positive, the correlation coefficient of any two random variables is the covariance of the two random variables divided by the product of the their standard deviations.

Example. The random variables X and Y have the joint pdf

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

To compute the correlation coefficient, we need

$$\mu_1 = E(X) = \int_0^1 \int_0^1 x(x+y)dxdy = \frac{7}{12},$$
$$\mu_2 = E(Y) = \int_0^1 \int_0^1 y(x+y)dxdy = \frac{7}{12},$$

and

$$\sigma_1^2 = E(X^2) - \mu_1^2 = \int_0^1 \int_0^1 x^2(x+y)dxdy - \left(\frac{7}{12}\right)^2 = \frac{11}{144},$$
$$\sigma_2^2 = E(Y^2) - \mu_2^2 = \int_0^1 \int_0^1 y^2(x+y)dxdy - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

The covariance of X and Y is

$$E(XY) - \mu_1\mu_2 = \int_0^1 \int_0^1 xy(x+y)dxdy - \left(\frac{7}{12}\right)^2 = -\frac{1}{144},$$

yielding the correlation coefficient as

$$\rho = \frac{-\frac{1}{144}}{\sqrt{\left(\frac{11}{144}\right)\left(\frac{11}{144}\right)}} = -\frac{1}{11}.$$

An important expectation is the *moment generating function* of a random variable X, either continuous or discrete. Suppose there exists a finite real number t for which the expectation

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \qquad \text{or} \qquad E(e^{tX}) = \sum_x e^{tx} f(x)$$

(continuous or discrete respectively) exists. Then, $M(t) = E(e^{tX})$ is the moment generating function. Setting $t = 0 \Rightarrow M(0) = 1$. Not every distribution has a (real-valued) moment generating function, but if it does, then the moment-generating function is unique and completely determines the distribution of the random variable.

For example, let X be a continuous random variable with

$$M(t) = (1-t)^{-2} = \int_{-\infty}^{\infty} e^{tx} f(x)dx, \qquad t < 1.$$

It is not obvious how to determine $f(x)$. Consider a distribution with pdf

$$f(x) = \begin{cases} xe^{-x}, \ 0 < x < \infty \\ 0 \qquad \text{elsewhere} \end{cases}$$

Then

$$M(t) = \int_0^{\infty} e^{tx} xe^{-x} dx = \int_0^{\infty} xe^{-(1-t)x} dx$$

$$= \int_0^\infty \frac{e^{-(1-t)x}}{1-t} dx = (1-t)^{-2}.$$

Thus, the pdf has the moment generating function $M(t) = (1-t)^{-2}$, $t < 1$.

An important property of moment generating functions is derivatives of all orders exist at $t = 0$,

$$\frac{dM(t)}{dt} = M'(t) = \int_{-\infty}^{\infty} x e^{tx} f(x) dx$$

and analogously for a discrete random variable X. Hence, for both cases, $t = 0$ implies

$$M'(0) = E(X) = \mu.$$

Similarly, the second derivative

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x) dx,$$

yields $M''(0) = E(X^2)$. Hence,

$$\sigma^2 = E(X^2) - \mu^2 = M''(0) - \left[M'(0)\right]^2.$$

Thus, using the previous example $M(t) = (1-t)^{-2}$, $t < 1$, yields

$$M'(t) = 2(1-t)^{-3}, \qquad \text{and} \qquad M''(t) = 6(1-t)^{-4},$$

so that $\mu = M'(0) = 2$ and $\sigma^2 = M''(0) - \mu^2 = 2$.

In general, for $m > 0$ an integer, the mth derivative of the moment generating function generates the mth moment of the distribution,

$$M^m(0) = E(X^m) = \int_{-\infty}^{\infty} x^m f(x) dx \qquad \text{or} \qquad \sum_x x^m f(x),$$

and this is used to define macroscopic quantities for e.g., gases, plasmas, collections of stars, etc.

Note that a mean, or any other higher-order moments or expectations, need not exist even for a well-defined pdf, as the following example illustrates.

Example. Suppose the random variable X has pdf

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

Then

$$\int_1^\infty x x^{-2} dx = \lim_{c\to\infty} \int_1^c \frac{dx}{x} = \lim_{c\to\infty} (\ln c - ln 1),$$

does not exist, and hence neither does the mean value of X, nor other higher-order expectations.

Example. Consider the moment generating function $M(t) = e^{t^2/2}$, $-\infty < x < \infty$. We can use an alternative approach to computing moments rather than simply differentiating. We may write

$$\begin{aligned} e^{t^2/2} &= 1 + \frac{1}{1!}\left(\frac{t^2}{2}\right) + \frac{1}{2!}\left(\frac{t^2}{2}\right)^2 + \cdots + \frac{1}{k!}\left(\frac{t^2}{2}\right)^k + \cdots \\ &= 1 + \frac{1}{2!}t^2 + \frac{(3)(1)}{4!}t^4 + \cdots + \frac{(2k-1)(2k-3)\cdots(3)(1)}{(2k)!}t^{2k} + \cdots \end{aligned}$$

The MacLaurin's series for $M(t)$ is

$$\begin{aligned} M(t) &= M(0) + \frac{M'(0)}{1!}t + \frac{M''(0)}{2!}t^2 + \cdots + \frac{M^m(0)}{m!}t^m + \cdots \\ &= 1 + \frac{E(X)}{1!}t + \frac{E(X^2)}{2!}t^2 + \cdots + \frac{E(X^m)}{m!}t^m + \cdots \end{aligned}$$

Thus the coefficient of $(t^m/m!)$ in the MacLaurin expansion of $M(t)$ is $E(X^m)$. For the example above, we therefore have

$$E(X^{2k}) = (2k-1)(2k-3)\cdots(3)(1) = \frac{(2k)!}{2^k k!},$$

$k = 1, 2, 3, \ldots$, and $E(X^{2k-1}) = 0$, $k = 1, 2, 3, \ldots$.

Before completing this section, we should point out that many functions do not have real-valued moment generating functions. However, if we define $\phi(t) = E(e^{itX})$, t an arbitrary real number, then this expectation exists for every distribution and is called the *characteristic function* of the distribution. In fact, the characteristic function may be defined from the pdf $f(x)$ using the Fourier integral

$$\phi(t) = \langle e^{itx} \rangle = \int_{-\infty}^\infty e^{itx} f(x) dx,$$

and of course the inverse Fourier transform yields the pdf

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt.$$

By expanding the exponential in the Fourier integral, we can compute moments as before,

$$\begin{aligned}
\phi(t) &= \int_{-\infty}^{\infty} \left(1 + itx - \frac{1}{2}t^2x^2 + \cdots\right) f(x)dx \\
&= 1 + iE(X)t - \frac{1}{2}E(X^2)t^2 + \cdots
\end{aligned}$$

Exercises

1. Suppose X has the pdf

$$f(x) = \begin{cases} \frac{x+2}{18}, & -2 < x < 4 \\ 0, & \text{elsewhere} \end{cases}$$

   Find $E(X)$, $E[(X+2)^3]$, and $E[6X - 2(X+2)^3]$.
2. The *median* of a random variable X is the value x such that the distribution function $F(x) = \frac{1}{2}$. Compute the median of the random variable X for the pdf

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

3. The *mode* of a random variable X is the value that occurs most frequently - sometimes called the *most probable value*. The value a is the mode of the random variable X if

$$f(a) = max f(x),$$

   (for a continuous pdf). The mode is not necessarily unique. Compute the mode and median of a random variable X with pdf

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

4. Suppose X and Y have the joint pdf

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

   and that $u(X, Y) = X$, $v(X, Y) = Y$, and $w(X, Y) = XY$. Show that $E[u(X, Y)] \cdot E[v(X, Y)] = E[w(X, Y)]$.

5. If X and Y are two exponentially distributed random variable with pdfs
$$f_X(x) = 2e^{-2x}, \qquad x \geq 0; \qquad f_Y(y) = 4e^{-4y}, \qquad y \geq 0,$$
calculate $E(X + Y)$.
6. Suppose X and Y have the joint pdf
$$f(x, y) = \begin{cases} 2, 0 < x < y, & 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$
and that $u(X, Y) = X$, $v(X, Y) = Y$, and $w(X, Y) = XY$. Show that $E[u(X, Y)] \cdot E[v(X, Y)] \neq E[w(X, Y)]$.
7. Let X have a pdf $f(x)$ that is positive at $x = -1, 0, 1$ and zero elsewhere. (a) If $f(0) = \frac{1}{2}$, find $E(X^2)$. (b) If $f(0) = \frac{1}{2}$, and if $E(X) = \frac{1}{6}$, determine $f(-1)$ and $f(1)$.
8. A random variable X with an unknown probability distribution has a mean $\mu = 12$ and a variance $\sigma^2 = 9$. Use Chebyshev's inequality to bound $P(6 < X < 18)$ and $P(3 < X < 21)$.
9. Two distinct integers are chosen randomly without replacement from the first six positive integers. What is the expected value of the absolute value of the difference of these two numbers?
10. Assume that the random variable X has mean μ, standard deviation σ, and moment generating function $M(t)$. Show that
$$E\left(\frac{X - \mu}{\sigma}\right) = 0; \qquad E\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = 1,$$
and
$$E\left\{\exp\left[t\left(\frac{X - \mu}{\sigma}\right)\right]\right\} = e^{-\mu t/\sigma} M\left(\frac{t}{\sigma}\right).$$
11. Suppose that $E[(x - b)^2]$ exists for a random variable X for all real b. Show that $E[(x - b)^2]$ is a minimum when $b = E(X)$.
12. Suppose that $R(t) = E(e^{t(X-b)})$ exists for a random variable X. Show that $R^m(0)$ is the mth moment of the distribution about the point b, where m is a positive integer.
13. Let $\Psi(t) = \ln M(t)$, where $M(t)$ is the moment generating function of a distribution. Show that $\Psi'(0) = \mu$ and $\Psi''(0) = \sigma^2$.
14. Suppose X is a random variable with mean μ and variance σ^2, and assume that the third moment $E[(X - \mu)^3]$ exists. The ratio $E[(X - \mu)^3]/\sigma^3$ is a measure of the *skewness* of the distribution. Graph the following pdfs and show that the skewness is negative, zero, and positive respectively:

    (a) $f(x) = (x + 1)/2, -1 < x < 1, 0$ elsewhere.
    (b) $f(x) = 1/2, -1 < x < 1, 0$ elsewhere.
    (c) $f(x) = (1 - x)/2, -1 < x < 1, 0$ elsewhere.

15. Suppose X is a random variable with mean μ and variance σ^2, and assume that the fourth moment $E[(X-\mu)^4]$ exists. The ratio $E[(X-\mu)^4]/\sigma^4$ is a measure of the *kurtosis* of the distribution. Graph the following pdfs and show that the kurtosis is smaller for the first distribution

    (a) $f(x) = 1/2, -1 < x < 1, 0$ elsewhere.
    (b) $f(x) = 3(1-x^2)/4, -1 < x < 1, 0$ elsewhere.

2.6 Conditional Probability and Marginal and Conditional Distributions

One is sometimes interested only in outcomes that are elements of a subset C_1 of the sample space $\mathcal{C}$. Thus, the subset becomes effectively the new sample space. Let $P(C)$ be the probability set function defined on $\mathcal{C}$ and let $C_1 \subset \mathcal{C}$ such that $P(C_1) > 0$. Suppose $C_2 \subset \mathcal{C}$. We want to define the probability of the event C_2 relative to the hypothesis of the event C_1. This is called the *conditional probability* of the event C_2 relative to the event C_1, or simply the conditional probability of C_2 given C_1, denoted by $P(C_2|C_1)$. Specifically, we define $P(C_2|C_1)$ such that

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}$$

and $P(C_1) > 0$. We then have

1. $P(C_2|C_1) \geq 0$.
2. $P(C_2 \cup C_3 \cup \cdots |C_1) = P(C_2|C_1) + P(C_3|C_1) + \cdots$, provided $C_2, C_3, \ldots$ are mutually disjoint sets.
3. $P(C_1|C_1) = 1$.

Properties (1) and (3) are obvious and (2) is an exercise. These properties of course ensure that $P(C_2|C_1)$ is a probability set function defined for subsets of C_1 – called the conditional probability set function given C_1.

Consider now a subset A of the event space $\mathcal{A}$ of one or more random variables defined on the sample space $\mathcal{C}$. If P is the probability set function of the induced probability on $\mathcal{A}$, and $A_1 \subset \mathcal{A}$ and $A_2 \subset \mathcal{A}$, then the conditional probability of the event A_2 given the event A_1 is

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)}$$

provided $P(A_1) > 0$.

Note that the above definition yields the multiplication rule for probabilities

$$P(C_1 \cap C_2) = P(C_1)P(C_2|C_1).$$

Example. Suppose we draw cards successively and randomly without replacement from an ordinary deck of cards. Given that three spades were drawn in the first six drawings, what is the probability that the seventh draw will yield a spade? Let C_1 denote the event of three spades in the first six draws and C_2 the event of a spade on the seventh drawing. We want to compute $P(C_1 \cap C_2)$. We therefore have

$$P(C_1) = \frac{\binom{13}{3}\binom{39}{3}}{\binom{52}{6}},$$

and

$$P(C_2|C_1) = \frac{10}{46}.$$

Hence, using

$$P(C_1 \cap C_2) = P(C_1)P(C_2|C_1) \simeq 0.028.$$

Note that the multiplication rule can be extended to multiple events quite straightforwardly. For three events, we have

$$\begin{aligned} P(C_1 \cap C_2 \cap C_3) &= P\left[(C_1 \cap C_2) \cap C_3\right] \\ &= P(C_1 \cap C_2)P(C_3|C_1 \cap C_2) \\ &= P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2). \end{aligned}$$

Let $f(x_1, x_2)$ be the joint pdf of two random variables X_1 and X_2. Consider the event $a < X_1 < b$, $a < b$. This event can occur when and only when $a < X_1 < b, -\infty < X_2 < \infty$.

$$P(a < X_1 < b, -\infty < X_2 < \infty) = \int_a^b \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 dx_1$$

for the continuous case (the extension to the discrete case is obvious). Now $\int_{-\infty}^{\infty} f(x_1, x_2) dx_2$ is a function of x_1 only, say $f_1(x_1)$. Hence, for every $a < b$,

$$P(a < X_1 < b) = \int_a^b f_1(x_1) dx_1$$

so that $f_1(x_1)$ is a function of X_1 only. $f_1(x_1)$ results from integrating (or summing) the joint pdf $f(x_1, x_2)$ over all x_2 for a fixed x_1. The function $f_1(x_1)$ is called the *marginal pdf* for X_1. A marginal pdf for X_2 is defined by

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1.$$

Example. Suppose the joint pdf of X_1 and X_2 is

$$f(x_1, x_2) = \begin{cases} \frac{x_1+x_2}{8}, & 0 \leq x_1 \leq 2, \quad 0 \leq x_2 \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

The marginal pdf of X_1 is

$$f_1(x_1) = \int f(x_1, x_2)dx_2 = \frac{1}{8}\int_0^2 (x_1 + x_2)dx_2 = \frac{x_1 + 1}{4},$$

and the marginal pdf of X_2 is

$$f_2(x_2) = \int f(x_1, x_2)dx_1 = \frac{1}{8}\int_0^2 (x_1 + x_2)dx_1 = \frac{x_2 + 1}{4}.$$

Note that

$$\int_0^2 f_1(x_1)dx_1 = 1 = \int_0^2 f_2(x_2)dx_2.$$

Example. Suppose the joint pdf of X_1 and X_2 is given by

$$f(x_1, x_2) = \begin{cases} \frac{x_1+x_2}{21}, & x_1 = 1, 2, 3, x_2 = 1, 2 \\ 0, & \text{elsewhere} \end{cases}$$

We have, for example,

$$P(X_1 = 3) = f(3, 1) + f(3, 2) = \frac{3}{7} \text{ and } P(X_2 = 2)$$
$$= f(1, 2) + f(2, 2) + f(3, 2) = \frac{4}{7}.$$

The marginal pdf of X_1 is

$$f_1(x_1) = \sum_{x_2=1}^{2} \frac{x_1 + x_2}{21} = \frac{2x_1 + 3}{21}, \qquad x_1 = 1, 2, 3$$

and 0 elsewhere. The marginal pdf of X_2 is

$$f_2(x_2) = \sum_{x_1=1}^{3} \frac{x_1 + x_2}{21} = \frac{6 + 3x_2}{21}, \qquad x_2 = 1, 2$$

and 0 elsewhere. The preceding probabilities can be computed directly from the marginals as $P(X_1 = 3) = f_1(3) = \frac{3}{7}$ and $P(X_2 = 2) = f_2(2) = \frac{4}{7}$.

Consider now the moment generating function (if it exists) $M(t_1, t_2) = E(e^{t_1 X + t_2 Y})$, t_1, t_2 finite, of the pdf $f(x, y)$ of the random variables X and Y. The moment-generating function, like the single random variable case, completely determines the joint distribution of X and Y, and hence the marginal distributions of X and Y. This follows from

$$M(t_1, 0) = E(e^{t_1 X}) = M(t_1), \quad \text{and} \quad M(0, t_2) = E(e^{t_2 Y}) = M(t_2).$$

For continuous random variables,

$$\frac{\partial^{k+m} M(t_1, t_2)}{\partial t_1^k t_2^m} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^m e^{t_1 x + t_2 y} f(x, y) dx dy,$$

implying that

$$\left. \frac{\partial^{k+m} M(t_1, t_2)}{\partial t_1^k t_2^m} \right|_{t_1 = t_2 = 0} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^m f(x, y) dx dy = E(X^k Y^m).$$

This yields the following set of useful relations,

$$\mu_1 = E(X) = \frac{\partial M(0,0)}{\partial t_1}, \qquad \mu_2 = E(Y) = \frac{\partial M(0,0)}{\partial t_2};$$

$$\sigma_1^2 = E(X^2) - \mu_1^2 = \frac{\partial^2 M(0,0)}{\partial t_1^2} - \mu_1^2;$$

$$\sigma_2^2 = E(Y^2) - \mu_2^2 = \frac{\partial^2 M(0,0)}{\partial t_2^2} - \mu_2^2;$$

$$E[(X - \mu_1)(Y - \mu_2)] = \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2} - \mu_1 \mu_2.$$

Thus, the covariance and correlation functions can be computed using the moment generating function of the joint pdf.

Example. Continuous random variables X and Y have the joint pdf

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

The moment generating function is given by

$$\begin{aligned} M(t_1, t_2) &= \int_0^\infty \int_x^\infty \exp(t_1 x + t_2 y - y) dy dx \\ &= \frac{1}{(1 - t_1 - t_2)(1 - t_2)}, \end{aligned}$$

provided $t_1 + t_2 < 1$ and $t_2 < 1$. From the moment-generating formulae above, we can derive (Exercise: Check)

$$\begin{aligned} &\mu_1 = 1, \qquad \mu_2 = 2, \\ &\sigma_1^2 = 1, \qquad \sigma_2^2 = 2, \\ &E[(X - \mu_1)(Y - \mu_2)] = 1. \end{aligned}$$

Hence, the correlation coefficient of X and Y is $\rho = 1/\sqrt{2}$. The moment generating functions of the marginal distributions of X and Y are

$$M(t_1, 0) = \frac{1}{1 - t_1}, \quad t_1 < 1; \quad M(0, t_2) = \frac{1}{(1 - t_2)^2}, \quad t_2 < 1.$$

The corresponding marginal pdfs are

$$f_1(x) = \begin{cases} \int_x^\infty e^{-y} dy = e^{-x}, 0 < x < \infty \\ \qquad 0 \qquad\qquad \text{elsewhere} \end{cases},$$

and

$$f_2(y) = \begin{cases} \int_0^y e^{-y} dx = y e^{-y}, 0 < y < \infty \\ \qquad 0 \qquad\qquad \text{elsewhere} \end{cases}.$$

Let X_1 and X_2 denote continuous random variables with joint pdf $f(x_1, x_2)$ and marginal pdfs $f_1(x_1)$ and $f_2(x_2)$. Provided $f_1(x_1) > 0$, we define the *conditional pdf* of the continuous random variable X_2 as

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}.$$

It is easily seen that $f(x_2|x_1)$ has the properties of a pdf (Exercise: Check!), which means that it can be used to compute probabilities and expectations. Thus, the conditional probability that $a < X_2 < b$, given that $X_1 = x_1$ is given by

$$P(a < X_2 < b|X_1 = x_1) = \int_a^b f(x_2|x_1) dx_2.$$

Similarly, $P(c < X_1 < d|X_2 = x_2) = P(c < X_1 < d|x_2) = \int_c^d f(x_1|x_2)dx_1$.

The expectation of the function $u(X_2)$ of X_2

$$E[u(X_2)|x_1] = \int_{-\infty}^{\infty} u(x_2) f(x_2|x_1)dx_2$$

is the *conditional expectation* of $u(X_2)$ given $X_1 = x_1$. Specifically, $E(X_2|x_1)$ is the mean and $E([X_2 - E(X_2|x_1)]^2|x_1)$ the variance of the conditional distribution of X_2 given $X_1 = x_1$. These are sometimes referred to as the conditional mean and variance. Obviously,

$$E[(X_2 - E(X_2|x_1))^2|x_1] = E(X_2^2|x_1) - [E(X_2|x_1)]^2.$$

Example. Suppose the random variables X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 6x_1, & 0 < x_1 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

The marginal pdfs are

$$f_1(x_1) = \int_{x_1}^{1} 6x_1 dx_2 = 6x_1(1 - x_1), \qquad 0 \le x_1 \le 1$$

$$f_2(x_2) = \int_0^{x_2} 6x_1 dx_1 = 3x_2^2, \qquad 0 \le x_2 \le 1$$

Note that

$$\int_0^1 f_1(x_1)dx_1 = \int_0^1 6x_1(1 - x_1)dx_1 = 1$$

$$\int_0^1 f_2(x_2)dx_2 = \int_0^1 3x_2^2 dx_2 = 1.$$

The conditional pdf of X_1 given $X_2 = x_2$ is

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{6x_1}{3x_2^2} = \frac{2x_1}{x_2^2},$$

and the conditional mean is

$$E(X_1|x_2) = \int_0^{x_2} x_1 f(x_1|x_2)dx_1 = \int_0^{x_2} \frac{2x_1^2}{x_2^2} dx_1 = \frac{2}{3}x_2.$$

Example. The random variables X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 2, & 0 < x_1 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

The marginal pdfs are thus

$$f_1(x_1) = \begin{cases} \int_{x_1}^{1} 2dx_2 = 2(1 - x_1), & 0 < x_1 < 1 \\ 0, & \text{elsewhere} \end{cases},$$

and

$$f_2(x_2) = \begin{cases} \int_{0}^{x_2} 2dx_1 = 2x_2, & 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}.$$

The conditional pdf of X_1, given $X_2 = x_2$, is

$$f(x_1|x_2) = \begin{cases} \frac{2}{2x_2} = x_2^{-1}, & 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

The conditional mean and variance of X_1 given $X_2 = x_2$ are therefore

$$\begin{aligned} E(X_1|x_2) &= \int_{-\infty}^{\infty} x_1 f(x_1|x_2) dx_1 \\ &= \int_0^{x_2} \frac{x_1}{x_2} dx_1 = \frac{1}{2} x_2, \qquad 0 < x_2 < 1 \end{aligned}$$

and

$$\begin{aligned} E([X_1 - E(X_1|x_2)]^2|x_2) &= \int_0^{x_2} \left(x_1 - \frac{1}{2}x_2\right)^2 x_2^{-1} dx_1 \\ &= \frac{x_2^2}{12}, \qquad 0 < x_2 < 1. \end{aligned}$$

Note that

$$P(0 < X_1 < 1/2|X_2 = 3/4) = \int_0^{1/2} f(x_1|3/4)dx_1 = \int_0^{1/2} \frac{4}{3} dx_1 = \frac{2}{3},$$

and

$$P(0 < X_1 < 1/2) = \int_0^{1/2} f_1(x_1)dx_1 = \int_0^{1/2} 2(1 - x_1)dx_1 = \frac{3}{4}.$$

The definitions introduced above are all extended in an obvious way to multivariables. Thus, for continuous random variables $X_1, X_2, \ldots, X_n$ with the joint pdf $f(x_1, x_2, \ldots, x_n)$, we have the following definitions:

1. The marginal pdfs $f_1(x_1)$, $f_2(x_2) \ldots, f_n(x_n)$ are defined by $(n-1)$-fold integrals

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n, \quad 1 \leq i \leq n.$$

2. We can also define a marginal pdf of a set of k variables, $k < n$. For example, suppose $n = 4$, i.e., X_1, X_2, X_3, X_4, and consider the subset <= X_2 and X_4, i.e., $k = 2$. The marginal pdf of X_2 and X_4 is the joint pdf of the two variables

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_1 dx_3.$$

   This is extended in an obvious way to any subset of n random variables.
3. Provided $f_i(x_i) > 0$, the joint conditional pdf of $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n$ given $X_i = x_i$ is

$$f(x_1, \ldots x_{i-1}, x_{i+1}, \ldots x_n | x_i) = \frac{f(x_1, \ldots, x_n)}{f_i(x_i)}.$$

4. As above, more generally, the joint conditional pdf of $n - k$ of the variables for given values of the remaining k variables is defined as the joint pdf of the n variables divided by the marginal pdf of the group of k variables provided it is positive.
5. Provided $f_i(x_i) > 0$, the conditional expectation of $u(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n)$ given $X_i = x_i$ is defined by

$$E[u(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n)|x_i] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n)$$
$$f(x_1, \ldots, x_n | x_i) dx_1, \ldots dx_{i-1} dx_{i+1} \ldots dx_n.$$

Corresponding definitions for discrete random variables X_1, $X_2, \ldots, X_n$ with the joint pdf $f(x_1, x_2, \ldots, x_n)$ hold, now using sums instead of integrals.

Exercises

1. Consider the joint pdf

$$f(x_1, x_2) = \begin{cases} \frac{1}{4} x_1 (1 + 3x_2^2), & 0 < x_1 < 2, \quad 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Show that $\int f(x_1, x_2)dx_1dx_2 = 1$. Find $P[(X_1, X_2) \in A]$ where $A = \{f(x_1, x_2)|0 < x_1 < 1, \frac{1}{4} < x_2 < \frac{1}{2}\}$. Determine $f_1(x_1)$, $f_2(x_2)$, $f(x_1|x_2)$, and $P(1/4 < X_1 < 1/2|X_2 = 1/3)$.

2. The random variables X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, 0 < x_1 < 1, & 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find the conditional mean and variance of X_2 given $X_1 = x_1, 0 < x_1 < 1$.

3. Suppose the conditional pdf of X_1 given $X_2 = x_2$ is

$$f(x_1|x_2) = \begin{cases} c_1 \frac{x_1}{x_2^2}, 0 < x_1 < x_2, & 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

and the marginal pdf of X_2 is

$$f_2(x_2) = \begin{cases} c_2 x_2^4, 0 < x_2 < 1, \\ 0, \quad \text{elsewhere} \end{cases}$$

Find (i) the constants c_1 and c_2; (ii) the joint pdf of X_1 and X_2; (iii) $P(1/4 < X_1 < 1/2|X_2 = 5/8)$; and (iv) $P(1/4 < X_1 < 1/2)$.

4. Suppose that the joint pdf of X_1 and X_2 is

$$f(x_1, x_2) = \begin{cases} c x_1^2 x_2, x_1^2 \le x_2 < 1 \\ 0, \quad \text{elsewhere} \end{cases}$$

Determine the value of the constant c and then $P(X_1 \ge X_2)$. Evaluate $f_1(x_1)$ and $f_2(x_2)$. (Hint: sketch the region where $f(x_1, x_2) \ge 0$.)

5. Let $\Psi(t_1, t_2) \equiv \ln M(t_1, t_2)$, where $M(t_1, t_2)$ is the moment generating function of X and Y. Show that

$$\frac{\partial \Psi(0, 0)}{\partial t_k}, \quad \frac{\partial^2 \Psi(0, 0)}{\partial t_k^2} \quad (k = 1, 2), \quad \frac{\partial^2 \Psi(0, 0)}{\partial t_1 \partial t_2},$$

yields the means, the variances, and the covariance of the two random variables.

6. Given the joint pdf of X_1 and X_2,

$$f(x_1, x_2) = \begin{cases} 21 x_1^2 x_2^3, 0 < x_1 < x_2 < 1, \\ 0 \quad \text{elsewhere} \end{cases}$$

find the conditional mean and variance of X_1 given $X_2 = x_2, 0 < x_2 < 1$.

7. Five cards are drawn at random without replacement from a deck of cards. The random variables X_1, X_2, and X_3 denote the number of spades, the

number of hearts, and the number of diamonds that appear among the 5 cards respectively. Determine the joint pdf of X_1, X_2, and X_3. Find the marginal pdfs of X_1, X_2, and X_3. What is the joint conditional pdf of X_2 and X_3 given that $X_1 = 3$?

8. Suppose that the joint pdf of X and Y is given by

$$f(x, y) = \begin{cases} 2, \; 0 < x < y, 0 < y < 1, \\ 0, \; \text{elsewhere.} \end{cases}.$$

Show that the conditional means are $(1 + x)/2$, $0 < x < 1$ and $y/2$, $0 < y < 1$, and the correlation function of X and Y is $\rho = 1/2$. Show also that the variance of the conditional distribution of Y given $X = x$ is $(1 - x)^2/12$, $0 < x < 1$, and that the variance of the conditional distribution of X given $Y = y$ is $y^2/12$, $0 < y < 1$.

9. Let $f(t)$ and $F(t)$ be the pdf and distribution function of the random variable T. The conditional pdf of T given $T > t_0$, t_0 a fixed time, is defined by $f(t|T > t_0) = f(t)/[1 - F(t_0)]$, $t > t_0$, 0 elsewhere. This kind of pdf is used in survival analysis i.e., problems of time until death, given survival until time t_0. Show that $f(t|T > t_0)$ is a pdf. Let $f(t) = e^{-t}$. $0 < t < \infty$, 0 elsewhere, and compute $P(T > 2|T > 1)$.

2.7 Stochastic Independence

Consider two random variables X_1 and X_2 with joint pdf $f(x_1, x_2)$. The joint pdf may be expressed as

$$f(x_1, x_2) = f(x_2|x_1) f_1(x_1).$$

Suppose that $f(x_2|x_1)$ does not depend on x_1. Then the marginal pdf of X_2 (assuming X_2 is continuous) is

$$\begin{aligned} f_2(x_2) &= \int_{-\infty}^{\infty} f(x_2|x_1) f_1(x_1) dx_1 \\ &= f(x_2|x_1) \int_{-\infty}^{\infty} f_1(x_1) dx_1 \\ &= f(x_2|x_1). \end{aligned}$$

Hence,

$$f_2(x_2) = f(x_2|x_1) \quad \text{and} \quad f(x_1, x_2) = f_1(x_1) f_2(x_2),$$

when $f(x_2|x_1)$ is independent of x_1. Thus, if the conditional distribution of X_2 given $X_1 = x_1$ is independent of x_1, then $f(x_1, x_2) = f_1(x_1)f_2(x_2)$.

Definition. Let the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2)$ and marginal pdfs $f_1(x_1)$ and $f_2(x_2)$. Then the random variables X_1 and X_2 are *stochastically independent* if and only if $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. Otherwise, they are stochastically dependent.

Example. Let the joint pdf of the random variables X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} \frac{1}{2}, 0 \leq x_1 \leq 2, 0 <\leq x_2 \leq 1, \\ 0, \text{ elsewhere.} \end{cases}.$$

The marginal pdfs are

$$f_1(x_1) = \int_0^1 \frac{1}{2} dx_2 = \frac{1}{2}, \qquad 0 \leq x_1 \leq 2$$

$$f_2(x_2) = \int_0^2 \frac{1}{2} dx_1 = 1, \qquad 0 \leq x_2 \leq 1$$

and 0 elsewhere. Thus,

$$f(x_1, x_2) = \frac{1}{2} \times 1 = \frac{1}{2}$$

implies that the random variables X_1 and X_2 are stochastically independent.

Example. Consider the random variables X_1 and X_2 with joint pdf

$$f(x_1, x_2) = \begin{cases} 12x_1x_2(1 - x_2), & 0 < x_1 < 1, 0 < x_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}.$$

Since the marginal pdfs are given by

$$f_1(x_1) = \int_0^1 12x_1x_2(1 - x_2)dx_2 = 2x_1;$$

$$f_2(x_2) = \int_0^1 12x_1x_2(1 - x_2)dx_1 = 6x_2(1 - x_2),$$

we have $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ and thus that X_1 and X_2 are stochastically independent.

Some useful theorems follow. These (i) provide a means of determining whether random variables are stochastically independent without computing the marginal pdfs; and (ii) show that the product property of independence carries over to probabilities, expectations, and moment generating functions.

Theorem. *Let the random variables X_1 and X_2 have joint pdf $f(x_1, x_2)$. Then X_1 and X_2 are stochastically independent if and only if*

$$f(x_1, x_2) \equiv g(x_1)h(x_2),$$

where $g(x_1) > 0$, $\forall x_1 \in \mathcal{A}_1, 0$ elsewhere, and $h(x_2) > 0$, $\forall x_2 \in \mathcal{A}_2, 0$ elsewhere.

Proof. If X_1 and X_2 are stochastically independent, then $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, where $f_1(x_1)$ and $f_2(x_2)$ are marginal pdfs. Thus the condition $f(x_1, x_2) = g(x_1)h(x_2)$ holds.

Conversely, if $f(x_1, x_2) = g(x_1)h(x_2)$ holds, then

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = g(x_1)\int_{-\infty}^{\infty} h(x_2)dx_2 = c_1 g(x_1),$$

and

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = h(x_2)\int_{-\infty}^{\infty} g(x_1)dx_1 = c_2 h(x_2).$$

Here c_1 and c_2 are constants. However,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)dx_1 dx_2 = 1 = \int_{-\infty}^{\infty} g(x_1)dx_1 \int_{-\infty}^{\infty} h(x_2)dx_2 = c_1 c_2.$$

Hence

$$f(x_1, x_2) \equiv g(x_1)h(x_2) = f_1(x_1) f_2(x_2),$$

and X_1 and X_2 are stochastically independent. The related proof for discrete random variables is similar.

Theorem. *If X_1 and X_2 are stochastically independent random variables with marginal pdfs $f_1(x_1)$ and $f_2(x_2)$, then*

$$P(a < X_1 < b, c < X_2 < d) = P(a < X_1 < b)P(c < X_2 < d),$$

for all constants a, b, c, d satisfying $a < b$ and $c < d$.

Proof.

$$\begin{aligned} P(a < X_1 < b, c < X_2 < d) &= \int_a^b \int_c^d f_1(x_1) f_2(x_2) dx_1 dx_2 \\ &= \int_a^b f_1(x_1)dx_1 \int_c^d f_2(x_2)dx_2 \\ &= P(a < X_1 < b)P(c < X_2 < d). \end{aligned}$$

Theorem. *Suppose X_1 and X_2 are stochastically independent random variables with marginal pdfs $f_1(x_1)$ and $f_2(x_2)$, and $u(X_1)$ and $v(X_2)$ are functions of X_1 and X_2 respectively. Then, the expectation*

$$E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)].$$

Proof. This follows immediately from the definition of stochastic independence.

$$\begin{aligned} E[u(X_1)v(X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)f_1(x_1)f_2(x_2)dx_1dx_2 \\ &= \int_{-\infty}^{\infty} u(x_1)f_1(x_1)dx_1 \int_{-\infty}^{\infty} v(x_2)f_2(x_2)dx_2 \\ &= E[u(X_1)]E[v(X_2)]. \end{aligned}$$

Example. Suppose X and Y are stochastically independent random variables with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. Then, we have the important result that the correlation coefficient of X and Y is zero because the covariance

$$\sigma_{12} = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} = \frac{E[(X-\mu_1)]E[(Y-\mu_2)]}{\sigma_1\sigma_2} = 0.$$

Theorem. *Suppose X_1 and X_2 are stochastically independent random variables with joint pdf $f(x_1, x_2)$ and marginal pdfs $f_1(x_1)$ and $f_2(x_2)$. If the moment generating function $M(t_1, t_2)$ of the distribution exists, then X_1 and X_2 are stochastically independent if and only if $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$.*

Proof. If X_1 and X_2 are stochastically independent, then

$$\begin{aligned} M(t_1, t_2) &= E(e^{t_1X_1 + t_2X_2}) \\ &= E(e^{t_1X_1}e^{t_2X_2}) \\ &= E(e^{t_1X_1})E(e^{t_2X_2}) = M(t_1, 0)M(0, t_2). \end{aligned}$$

Hence, stochastic independence of X_1 and X_2 implies that the moment generating function factors into a product of the two marginal moment generating functions of the marginal distributions.

Suppose now instead that $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$. Since

$$M(t_1, 0) = \int_{-\infty}^{\infty} e^{t_1x_1}f_1(x_1)dx_1 \quad \text{and} \quad M(0, t_2) = \int_{-\infty}^{\infty} e^{t_2x_2}f_2(x_2)dx_2,$$

we have

$$
\begin{aligned}
M(t_1,0)M(0,t_2) &= \int_{-\infty}^{\infty} e^{t_1x_1} f_1(x_1)dx_1 \int_{-\infty}^{\infty} e^{t_2x_2} f_2(x_2)dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f_1(x_1)f_2(x_2)dx_1dx_2 \\
&= M(t_1,t_2).
\end{aligned}
$$

But

$$
M(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f(x_1,x_2)dx_1dx_2,
$$

which implies that

$$
f(x_1,x_2) = f_1(x_1)f_2(x_2),
$$

and hence that X_1 and X_2 are stochastically independent random variables. The case for discrete random variables is similar.

The n random variables $X_1, X_2, \ldots, X_n$, with joint pdf $f(x_1, x_2, \ldots, x_n)$ and marginal pdfs $f_1(x_1)$, $f_2(x_2), \ldots, f_n(x_n)$, are *mutually stochastically independent* if and only if $f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n)$. The theorems above can be suitably generalized.

Exercises

1. Let the joint pdf of X_1 and X_2 be

$$
f(x_1,x_2) = \begin{cases} x_1 + x_2, & 0 < x_1 < 1, 0 < x_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}.
$$

   Show that the random variables X_1 and X_2 are stochastically dependent.
2. Show that the random variables X and Y with joint pdf

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

   are stochastically dependent.
3. Consider the joint pdf of random variables X and Y,

$$
f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}.
$$

   Are the random variables X and Y stochastically independent? Compute $f(x|y)$ and hence $P(1/4 < X < 1/2|Y = 3)$.

4. The random variables X and Y have joint pdf

$$f(x, y) = \begin{cases} 4x(1-y), & 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}.$$

Find $P(0 < X < 1/3, 0 < Y < 1/3)$.

5. Let X_1, X_2, and X_3 be three stochastically independent random variables, each with pdf

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

Find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.

6. Show that the random variables X and Y with joint pdf

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

are stochastically independent, and that

$$E(e^{t(X+Y)}) = (1-t)^{-2}, \quad t < 1.$$

2.8 Particular Distributions

We consider three of the most important probability distribution functions, the *binomial distribution*, the *Poisson distribution*, and the *normal* or *Gaussian distribution*, and the latter's connection to the *Maxwell-Boltzmann* distribution.

2.8.1 *The Binomial Distribution*

The binomial theorem is expressed as

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x},$$

for $n > 0$ an integer. Recall, $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ By analogy, introduce the function

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, 3, \ldots, n$$
$$= 0 \quad \text{elsewhere}$$

for $n > 0$ an integer and $0 < p < 1$. Clearly, $f(x) \geq 0$ and

$$\sum_x f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [(1-p) + p]^n = 1.$$

Hence, the function $f(x)$ is the pdf of a discrete random variable X. A random variable with this pdf has a *binomial distribution*, and $f(x)$ is a *binomial pdf*. n and p are the parameters of the binomial distribution, often denoted by $B(n, p)$. For example, $B(4, 1/3)$ has the binomial pdf

$$f(x) = \binom{4}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{4-x}, \qquad x = 0, 1, 2, 3, 4$$
$$= 0 \quad \text{elsewhere}$$

The binomial distribution is a very useful model for any experiment or system that admits an outcome drawn from two possibilities only, such as heads or tails in a coin toss, life or death, red or green, etc. If the experiment is repeated n independent times or a system produces n independent outcomes and the probability of "success" is p on each occasion, then the probability for "failure" is $1 - p$. Define the random variable X_i, $i = 1, 2 \ldots, n$ to be 0 if the outcome of the ith performance is a failure and 1 if the outcome is a success. Thus $P(X_i = 0) = 1 - p$ and $P(X_i = 1) = p$, $i = 1, 2 \ldots, n$. The random variables X_i are mutually stochastically independent since the experiment is repeated n independent times. $Y = X_1 + X_2 + \cdots + X_n$ is the number of successes through the n repetitions of the experiment. Let y be an element of the set $\{y : y = 0, 1, 2, \ldots, n\}$. Then $Y = y$ if and only if y of the random variables X_i have the value 1 and $n - y$ have value 0. The number of combinations of y 1's that can be assigned to the X_i is just $\binom{n}{y}$. The probability for each of these possible combinations is simply $p^y(1-p)^{n-y}$ because the X_i are mutually stochastically independent. The $P(Y = y)$ is the sum of the $\binom{n}{y}$ mutually exclusive events, i.e.,

$$\binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0, 1, 2, 3, \ldots, n$$

and 0 elsewhere. This is the pdf of a binomial distribution.

The moment generating function can be evaluated from

$$\begin{aligned} M(t) = \sum_x e^{tx} f(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ = \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x (1-p)^{n-x} \\ = \left[(1-p) + pe^t\right]^n, \qquad \forall t \in \Re. \end{aligned}$$

The mean and variance are therefore

$$M'(t) = n\left[(1-p) + pe^t\right]^{n-1} pe^t \Longrightarrow \mu = M'(0) = np,$$

and

$$\begin{aligned} M''(t) \;&=\; n\left[(1-p) + pe^t\right]^{n-1} pe^t + n(n-1)\left[(1-p) + pe^t\right]^{n-2} \left(pe^t\right)^2 \\ &\Longrightarrow \sigma^2 = M''(0) - \mu^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p). \end{aligned}$$

Example. The binomial distribution with pdf

$$\begin{aligned} f(x) &= \binom{4}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}, \qquad x = 0, 1, 2, 3, 4 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

and random variable X has moment generating function

$$M(t) = \left(\frac{1}{2} + \frac{1}{2}e^t\right)^4,$$

and mean $\mu = np = 2$ and variance $\sigma^2 = np(1-p) = 1$. We can compute, for example,

$$P(0 \le X \le 1) = \sum_{x=0}^{1} f(x) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16},$$

and

$$P(X = 3) = f(3) = \frac{4!}{3!1!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 = \frac{1}{4}.$$

Exercises

1. If the moment generating function of a random variable X is

$$\left(\frac{1}{3}+\frac{2}{3}e^t\right)^5$$

find $P(X = 2 \text{ or } 3)$.
2. The moment generating function of a random variable X is

$$\left(\frac{2}{3}+\frac{1}{3}e^t\right)^9.$$

Show that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \sum_{x=1}^{5}\binom{9}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{9-x}.$$

3. The probability that a patient recovers from heart surgery is 0.4. If 15 people have had heart surgery, what is the probability that (i) at least 10 survive, (ii) from 3 to 8 survive, (iii) exactly 5 survive? Using Chebyshev's inequality, find and interpret the interval $\mu \pm 2\sigma$.
4. If the random variable X has a binomial distribution with parameters n and p, show that

$$E\left(\frac{X}{n}\right) = p \quad \text{and} \quad E\left[\left(\frac{X}{n}-p\right)^2\right] = \frac{p(1-p)}{n}.$$

2.8.2 *The Poisson Distribution*

For all values of p, the series

$$1 + p + \frac{p^2}{2!} + \frac{p^3}{3!} + \cdots = \sum_{x=0}^{\infty}\frac{p^x}{x!}, \quad x = 0, 1, 2, \ldots$$

converges to e^p. This motivates the introduction of the function $f(x)$

$$\begin{aligned} f(x) &= \frac{p^x e^{-p}}{x!}, \quad x = 0, 1, 2, \ldots \\ &= 0, \quad \text{elsewhere} \qquad \forall p > 0. \end{aligned}$$

Since $p > 0$, $f(x) \geq 0$ and

$$\sum_x f(x) = \sum_{x=0}^{\infty} \frac{p^x e^{-p}}{x!} = e^{-p} \sum_{x=0}^{\infty} \frac{p^x}{x!} = e^{-p} e^p = 1.$$

Hence, $f(x)$ is a pdf of a discrete random variable. A random variable X that has the pdf $f(x)$ is a *Poisson distribution* and $f(x)$ is a *Poisson pdf*.

Examples of Poisson distributions include the random variable X that denotes the number of alpha particles emitted by a radioactive substance that enter some region in a prescribed time interval, or the number of defects in a manufactured article. Even the number of automobile accidents during some unit time is often assumed to be random variable with a Poisson distribution. A process that leads to a Poisson distribution is called a *Poisson process*. The assumptions that underly a Poisson process are essentially that the probability of a change during a sufficiently short interval is independent of changes in other non-overlapping intervals, and is approximately proportional to the length of the interval, and the probability of more than one change during a short interval is essentially zero. One can formalize these assumptions and derive a simple ordinary differential equation that shows that the probability of changes X in an interval of some length has a Poisson distribution.

The moment generating function of a Poisson distribution is

$$\begin{aligned} M(t) = \sum_x e^{tx} f(x) &= \sum_{x=0}^{\infty} e^{tx} \frac{p^x e^{-p}}{x!} \\ &= e^{-p} \sum_{x=0}^{\infty} \frac{(pe^t)^x}{x!} \\ &= e^{-p} e^{pe^t} = e^{p(e^t - 1)}, \quad \forall t \in \Re \end{aligned}$$

The mean and variance are found to be

$$\begin{aligned} M'(t) &= \exp\left[p\left(e^t - 1\right)\right] pe^t \Longrightarrow \mu = M'(0) = p \\ M''(t) &= \exp\left[p\left(e^t - 1\right)\right] pe^t + \exp\left[p\left(e^t - 1\right)\right] \left(pe^t\right)^2 \\ &\Longrightarrow \sigma^2 = M''(0) - \mu^2 = p + p^2 - p^2 = p, \end{aligned}$$

i.e., a Poisson distribution has $\mu = \sigma^2 = p > 0$. This allows us to express the Poisson pdf as

$$\begin{aligned} f(x) &= \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0, 1, 2, \ldots \\ &= 0 \quad \text{elsewhere} \qquad \forall \mu > 0. \end{aligned}$$

Example. Consider a random variable X with a Poisson distribution with $\mu = 3$ and $\sigma^2 = 3$, i.e.,

$$f(x) = \frac{3^x e^{-3}}{x!}, \quad x = 0, 1, 2, \ldots$$
$$= 0 \quad \text{elsewhere.}$$

For example,

$$P(1 \leq X) = 1 - P(X = 0) = 1 - f(0) = 1 - e^{-3} = 0.95.$$

Example. If the moment generating function of a random variable X is

$$M(t) = \exp\left[2\left(e^t - 1\right)\right],$$

then X has a Poisson distribution with $\mu = 2$. For example,

$$P(X = 3) = f(3) = \frac{2^3 e^{-2}}{3!} = \frac{4}{3} e^{-2} = 0.18.$$

Example. The average number of radioactive particles passing through a counter during 1 millisecond in an experiment is 4. What is the probability that six particles enter the counter in a given millisecond?

We may assume a Poisson distribution with $x = 6$ and $\mu = 4$, so that

$$f(6) = \frac{4^6 e^{-4}}{6!} = 0.104.$$

Exercises

1. If the random variable X has a Poisson distribution such that $P(X = 1) = P(X = 2)$, find $P(X = 4)$.
2. Given that $M(t) = \exp\left[4\left(e^t - 1\right)\right]$ is the moment generating function of a random variable X, show that $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.931$.
3. Suppose that during a given rush hour Wednesday, the number of accidents on a certain stretch of highway has a Poisson distribution with mean 0.7. What is the probability that there will be at least three accidents on that stretch of highway at rush hour on Wednesday?
4. Compute the measures of skewness and kurtosis of the Poisson distribution with mean μ.
5. Suppose the random variables X and Y have the joint pdf

$$f(x, y) = \frac{e^{-2}}{x!(y - x)!}, \quad y = 0, 1, 2, \ldots; x = 0, 1, \ldots, y,$$
$$= 0, \quad \text{elsewhere.}$$

(i) Find the moment generating function $M(t_1, t_2)$ of the joint pdf. (ii) Compute the means, variances, and correlation coefficient of X and Y. (iii) Determine the conditional mean $E(X|y)$.

2.8.3 The Normal or Gaussian Distribution

Known by both terms, depending on the context (mathematical statistics, plasma physics, statistical physics), this is the most familiar and important of the many distribution functions that exist. We can evaluate the integral

$$I = \int_{-\infty}^{\infty} \exp\left(-x^2/2\right) dx,$$

by noting that $I > 0$ and that I^2 may be written as

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 + y^2}{2}\right) dxdy.$$

Introducing polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$ yields

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta = \int_0^{2\pi} d\theta = 2\pi,$$

which shows that $I = \sqrt{2\pi}$ and so

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-x^2/2\right) dx = 1.$$

If we replace x by

$$\frac{x-a}{b}, \quad b > 0,$$

we have

$$\frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-a)^2}{2b^2}\right) dx = 1.$$

Consequently, since $b > 0$, the function

$$f(x) = \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2b^2}\right), \quad -\infty < x < \infty,$$

is the pdf of a continuous random variable, and the random variable has a *normal* or *Gaussian* distribution and $f(x)$ is a normal pdf.

The moment generating function for a normal distribution is

$$\begin{aligned} M(t) &= \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\left(-\frac{(x-a)^2}{2b^2}\right) dx \\ &= \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{-2b^2tx + x^2 - 2ax + a^2}{2b^2}\right) dx \\ &= \exp\left(-\frac{a^2 - (a+b^2t)^2}{2b^2}\right) \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-a-b^2t)^2}{2b^2}\right) dx \\ &= \exp\left(at + \frac{b^2t^2}{2}\right), \end{aligned}$$

after completing the square and since the last integrand is a normal distribution with $a + b^2t$. The mean and variance can then be computed as

$$M'(t) = M(t)(a + b^2t) \Longrightarrow \mu = M'(0) = a,$$

and

$$\begin{aligned} M''(t) &= M(t)b^2 + M(t)(a + b^2t)^2, \\ &\Longrightarrow \sigma^2 = M''(0) - \mu^2 = b^2 + a^2 - a^2 = b^2. \end{aligned}$$

The normal or Gaussian pdf can therefore be written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty,$$

and the moment generating function as

$$M(t) = \exp\left(\mu t + \frac{\sigma^2t^2}{2}\right).$$

Example. If X has the moment generating function

$$M(t) = \exp\left(3t + 16t^2\right),$$

then X is normally distributed with mean $\mu = 3$ and variance $\sigma^2 = 32$.

The normal distribution is often expressed simply as $n(\mu, \sigma^2)$, thus, for example, $n(0, 1)$ implies the pdf of X has mean 0 and variance 1 and is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty, \quad M(t) = e^{t^2/2}.$$

The graph of the normal distribution is the familiar "bell shape," symmetric about $x = \mu$ with a maximum there of $1/\sigma\sqrt{2\pi}$. There are points of inflection at $x = \mu \pm \sigma$ (Exercise: Check!).

A useful "renormalization" of the Gaussian distribution is the following

Theorem. *If the random variable X is $n(\mu, \sigma^2)$, $\sigma^2 > 0$, then the random variable $W = (X - \mu)/\sigma$ is $n(0, 1)$.*

Proof. Since $\sigma^2 > 0$, the distribution function $G(w)$ of W is

$$G(w) = P\left(\frac{X - \mu}{\sigma} \leq w\right) = P(X \leq w\sigma + \mu),$$

which corresponds to

$$G(w) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{w\sigma+\mu} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

On setting $y = (x - \mu)/\sigma$, we have

$$G(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-y^2/2} dy,$$

implying that the pdf $g(w) = G'(w)$ of the continuous random variable W is

$$g(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2},$$

meaning that W is $n(0, 1)$,

This theorem is very useful for calculating probabilities of normally distributed variables. Suppose X is $n(\mu, \sigma^2)$. Then if $a < b$, we have

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X < a) \\ &= P\left(\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(b-\mu)/\sigma} e^{-w^2/2} dw - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} e^{-w^2/2} dw \end{aligned}$$

because $W = (X - \mu)/\sigma$ is $n(0, 1)$. Integrals of this form cannot be evaluated so tables are used, based on the $n(0, 1)$ distribution i.e., if

$$N(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2} dw,$$

and X is $n(\mu, \sigma^2)$, then

$$P(a < X < b) = P\left(\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right)$$
$$= N\left(\frac{b-\mu}{\sigma}\right) - N\left(\frac{a-\mu}{\sigma}\right).$$

Note that it can be shown that $N(-x) = 1 - N(x)$.

Example. For X $n(2, 25)$,

$$P(0 < X < 10) = N\left(\frac{10-2}{5}\right) - N\left(\frac{0-2}{5}\right)$$
$$= N(1.6) - N(-0.4) = 0.945 - (1 - 0.655) = 0.600$$

where the last steps involved looking up a table of normal values. In similar fashion,

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = N\left(\frac{\mu + 2\sigma - \mu}{\sigma}\right) - N\left(\frac{\mu - 2\sigma - \mu}{\sigma}\right)$$
$$= N(2) - N(-2) = 0.977 - (1 - 0.977) = 0.954.$$

Exercises

1. If
$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy,$$
show that $N(-x) = 1 - N(x)$.
2. If X is $n(75, 100)$, find $P(X < 60)$ and $P(70 < X < 100)$.
3. If X is $n(\mu, \sigma^2)$, find a so that $P(-a < (X - \mu)/\sigma < a) = 0.90$.
4. If X is $n(\mu, \sigma^2)$, show that $E(|X - \mu|) = \sigma\sqrt{2/\pi}$.
5. Show that the pdf $n(\mu, \sigma^2)$ has points of inflection at $x = \mu \pm \sigma$.
6. Suppose a random variable X has pdf
$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \qquad 0 < x < \infty$$
$$= 0, \qquad \text{elsewhere.}$$
Find the mean and variance of X.
7. Let X_1 and X_2 be two stochastically independent normally distributed random variables with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. Show that $X_1 + X_2$ is

normally distributed with mean $(\mu_1 + \mu_2)$ and variance $(\sigma_1^2 + \sigma_2^2)$. (Hint: use the uniqueness of the moment generating function.)

8. Compute $P(1 < X^2 < 9)$ if X is $n(1, 4)$.
9. Suppose the random variable X is normally distributed with $n(\mu, \sigma^2)$. What will the distribution be if $\sigma^2 = 0$?

2.9 The Central Limit Theorem

The central limit theorem shows that under certain conditions, probability distributions will converge to the normal or Gaussian distribution. We will show briefly the relationship to the Maxwell-Boltzmann distribution. Many different versions exist with different conditions and convergence properties. We will describe the simplest (and most restrictive) version.

Before establishing the central limit theorem, we will need the following useful result. Consider a limit of the form

$$\lim_{n\to\infty}\left[1 + \frac{a}{n} + \frac{\phi(n)}{n}\right]^{bn},$$

for a and b independent of n and where $\lim_{n\to\infty}\phi(n) = 0$. Then

$$\lim_{n\to\infty}\left[1 + \frac{a}{n} + \frac{\phi(n)}{n}\right]^{bn} = \lim_{n\to\infty}\left[1 + \frac{a}{n}\right]^{bn} = e^{ab}.$$

For example,

$$\lim_{n\to\infty}\left[1 - \frac{w^2}{n} + \frac{w}{n^4}\right]^{2n} = \lim_{n\to\infty}\left[1 + \frac{-w^2}{n} + \frac{w/n^3}{n}\right]^{2n} = e^{-2w^2},$$

for all fixed values of w.

Theorem. *Suppose that X_i, $i = 1, 2, \ldots, n$ is a random sample from a distribution that has mean μ and variance σ^2. Then the random variable $Y_n = \left(\sum_{i=1}^{n} X_i - n\mu\right)/\sqrt{n}\sigma = \sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma$ has a limiting distribution that is normal with mean 0 and variance 1.*

Comment By establishing this theorem, the central limit theorem, it implies that whenever the conditions of the theorem are satisfied, the random variable $\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma$ has, for a fixed n, an approximate normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

Proof. Let us assume the existence of a moment generating function $M(t) = E(e^{tX})$ for finite values of t for the distribution. (An alternative more general proof would be based on the characteristic function $\phi(t) = E(e^{itX})$ instead.) Introduce the moment generating function for $X - \mu$,

$$m(t) \equiv E\left(e^{t(X-\mu)}\right) = e^{-\mu t} M(t).$$

Hence, $m(0) = 1$, $m'(0) = E(X - \mu) = 0$, $m''(0) = E[(X - \mu)^2] = \sigma^2$. The function $m(t)$ can be expanded using Taylor's formula, for $0 < \xi < t$, such that

$$\begin{aligned} m(t) &= m(0) + m'(0)t + \frac{1}{2}m''(\xi)t^2 \\ &= 1 + \frac{1}{2}m''(\xi)t^2 \\ &= 1 + \frac{1}{2}\sigma^2 t^2 + \frac{1}{2}\left(m''(\xi) - \sigma^2\right)t^2. \end{aligned}$$

Now consider $M(t;n)$, where

$$\begin{aligned} M(t;n) &= E\left[\exp\left(t\frac{\sum X_i - n\mu}{\sigma\sqrt{n}}\right)\right] \\ &= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\exp\left(t\frac{X_2 - \mu}{\sigma\sqrt{n}}\right)\cdots\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\ &= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\right]E\left[\exp\left(t\frac{X_2 - \mu}{\sigma\sqrt{n}}\right)\right]\cdots E\left[\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\ &= \left\{E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]\right\}^n \\ &= \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n. \end{aligned}$$

Hence,

$$m\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{\left(m''(\xi) - \sigma^2\right)t^2}{2n\sigma^2}, \quad 0 < \xi < \frac{t}{\sigma\sqrt{n}}.$$

Thus,

$$M(t;n) = \left[1 + \frac{t^2}{2n} + \frac{\left(m''(\xi) - \sigma^2\right)t^2}{2n\sigma^2}\right]^n.$$

Since $m''(t)$ is continuous at $t = 0$, and since $\xi \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty}\left(m''(\xi)-\sigma^2\right)=0.$$

Hence, using the result above, $a = t^2/2$ and $b = 1$, and so

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2} \quad \forall t \in \Re.$$

It therefore follows that the random variable $Y_n = \sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma$ has a limiting normal distribution with $\mu = 0$ and variance $\sigma^2 = 1$.

Example. Suppose $\bar{X}$ denotes the mean of a random sample of size 75 from a distribution that has the pdf

$$\begin{aligned} f(x) &= 1, \quad 0 < x < 1 \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

Hence, $\mu = 1/2$ and $\sigma^2 = 1/12$. The limiting distribution of $Y_n = \sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma$ is normally distributed, allowing us to compute for example the $P(0.45 < \bar{X} < 0.55)$ by means of

$$P(0.45 < \bar{X} < 0.55) = P\left[\frac{\sqrt{n}(0.45-\mu)}{\sigma} < \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} < \frac{\sqrt{n}(0.55-\mu)}{\sigma}\right]$$

$$= P\left[\frac{\sqrt{75}(0.45-1/2)}{\sqrt{1/12}} < \frac{\sqrt{75}(\bar{X}_n-1/2)}{\sqrt{1/12}} < \frac{\sqrt{75}(0.55-1/2)}{\sqrt{1/12}}\right]$$

$$= P[-1.5 < 30(\bar{X}-1/2) < 1.5] = 0.866.$$

Example. Suppose X_i, $i = 1, 2, \ldots, n$ is a random sample from a binomial distribution $B(n, p) = B(1, p)$ i.e., $\mu = np = p$ and $\sigma^2 = p(1-p)$, and $M(t)$ exists $\forall t \in \Re$. If $Y_n = X_1 + X_2 + \cdots + X_n$, we know that Y_n is $B(n, p)$. We can use

$$\frac{Y_N - np}{\sqrt{np(1-p)}} = \frac{n(\bar{X}_n - p)}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}$$

as a limiting distribution with mean 0 and variance 1. Suppose $n = 100$ and $p = 1/2$, and that we want to compute $P(Y = 48, 49, 50, 51, 52)$. Since Y is a discrete random variable, the events $Y = 48, 49, 50, 51, 52$ and $47.5 < Y < 52.5$ are equivalent (using the convention of taking 0.5 above and below the limiting discrete values). So instead we compute $P(47.5 < Y < 52.5)$. Thus, with $\mu = np = 50$ and $\sigma^2 = np(1-p) = 25$, we have

$$P(47.5 < Y < 52.5) = P\left(\frac{47.5-50}{5} < \frac{Y-50}{5} < \frac{52.5-50}{5}\right)$$
$$= P\left(-0.5 < \frac{Y-50}{5} < 0.5\right) = 0.382$$

since $(Y-50)/5$ has an approximately normal distribution.

There are many examples of stochastic variables whose values are determined by independent additive increments. The best known example of such a variable may be the momentum of a molecule in a dilute gas. At a given time, the x momentum mv_x is the vector sum of all momentum increments caused by past collisions with other molecules. If we suppose that the increments in mv_x are independent with zero mean, we may conclude from the central limit theorem that v_x is normally distributed, i.e.,

$$f(v_x) = \frac{1}{\sqrt{2\pi kT}} \exp\left(-\frac{v_x^2}{2kT}\right),$$

where k is Boltzmann's constant and T is the temperature. Similarly, we may argue that the y and z momenta are independent of the x momentum, and that the increments in the three directions are independent. Hence the y and z velocities also have normal or Gaussian distributions. The joint pdf $f(v_x, v_y, v_z)$ of the independent velocities is therefore the product

$$f(v_x, v_y, v_z) = f(v_x)f(v_y)f(v_z) = \left(\frac{1}{2\pi kT}\right)^{3/2} \exp\left(-\frac{v_x^2+v_y^2+v_z^2}{2kT}\right).$$

By introducing spherical coordinates in velocity space with $c^2 = v_x^2 + v_y^2 + v_z^2$, we obtain the Maxwell-Boltzmann distribution

$$f(c) = 4\pi\left(\frac{1}{2\pi kT}\right)^{3/2} c^2 \exp\left(-\frac{c^2}{2kT}\right).$$

The Maxwell-Boltzmann distribution for the gas is a consequence of the independence of the successive collisions experienced by a molecule. The three components are identically distributed because there is no preferred direction. This can be different for a magnetized flow in the presence of a large-scale or mean magnetic field. The distributions have a zero mean because the gas is at rest with respect to the chosen coordinate system. A mean flow will introduce an offset in the normal distribution.

Velocities in a turbulent fluid flow do not have a normal distribution because the momentum increments that a fluid parcel experiences at successive times are not necessarily independent. For example, eddies tend to be coherent and interact in sometimes complicated ways with other fluid particles. However, at some scales, the local motions may be nearly independent. Consequently, turbulent velocity fields are not Gaussian, although often not very different from having a Gaussian distribution. This difference is a fundamental property of the dynamics of turbulence. To analyze the dynamics of turbulence, non-Gaussian properties do need to be included typically, but if one is concerned primarily with the effects of turbulence, assuming a Gaussian distribution may be an adequate approximation.

Exercises

1. Compute an approximate probability that the mean of a random sample of size 15 from a distribution having pdf

$$f(x) = 3x^2, \quad 0 < x < 1,$$
$$= 0, \quad \text{elsewhere}$$

is between 3/5 and 4/5.

2. Let Y be $B(72, 1/3)$. Approximate $P(22 \leq Y \leq 28)$.

2.10 Relation Between Microscopic and Macroscopic Descriptions: Particles, the Gibbs Ensemble, and Liouville's Theorem

A gas of particles or plasma of charged particles are both characterized by a very large number of degrees of freedom. One can in principle describe the system in terms the spatial and momentum coordinates of each of the particles in the system. By contrast, a macroscopic description, such as a fluid mechanical model, may have as few as three variables (the density, the velocity, the pressure or the temperature), depending on the closure assumptions imposed. The statistical treatment of the same system may require as many as $6N$ variables, where N is the number of particles in the system. These variables are the $3N$ spatial coordinates $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_N \equiv (\mathbf{x})$ and the $3N$ conjugate momenta $\mathbf{p}_1, \mathbf{p}_2, \cdots \mathbf{p}_N \equiv (\mathbf{p})$ of the constituent particles. This of course neglects effects specific to the particles themselves and treats the particles as point masses. A typical system can be described in terms of a Hamiltonian function $H\,[(\mathbf{x}), (\mathbf{p})]$, where

$$H\,[(\mathbf{x}), (\mathbf{p})] = E\,[(\mathbf{x}), (\mathbf{p})]\,.$$

In the absence of external fields, $E\,[(\mathbf{x}), (\mathbf{p})]$ denotes the total energy, kinetic energy, and potential energy of the system. The equations of motion for the system are given by Hamilton's equations

$$\frac{d\mathbf{x}_i}{dt} = \dot{\mathbf{x}}_i = \frac{\partial H}{\partial \mathbf{p}_i}$$

$$\frac{d\mathbf{p}_i}{dt} = \dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{x}_i}, \qquad i = 1, 2, \cdots, N.$$

The state of the system at any time is given by a representative point in the $6N$ dimensional phase space (also called the Γ space) defined by the mutually orthogonal vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots \mathbf{x}_N$, $\mathbf{p}_1$, $\mathbf{p}_2$, $\cdots \mathbf{p}_N$. Thus, for a given set of initial conditions, the trajectories of a particular system can be computed. Note that the Hamiltonian does not depend on time and so the equations of motion above are invariant under time reversal.

It is evident that a very large number of states of a gas corresponds to a particular macroscopic state of a gas e.g., a gas of a particular density contained in a box of fixed volume can be formed in an infinite number of ways according to the distribution of the particles in space.[1] However, macroscopically we cannot distinguish between one representative point or another i.e., between gases existing in different states. A gas that can be described by certain macroscopic conditions refers therefore to an infinite number of states and not to a single state. Thus, instead of considering a single system, we may consider a collection of systems that are identical in composition and macroscopic conditions but existing in different states. Such a collection of systems is called a Gibbs ensemble, and is the collection of systems that is microscopically equivalent to the system we are considering macroscopically. Each system in the ensemble can be represented by a point in phase space. As the number of systems becomes very large, the representative points become increasingly dense in phase space and we can describe their distribution in phase space by a density function. The density function is a continuous function of $(\mathbf{x})$ and $(\mathbf{p})$, which if normalized can be described by a probability density function $f_N(\mathbf{x}, \mathbf{p}, t)$ i.e., $f_N(\mathbf{x}, \mathbf{p}, t)d^{3N}\mathbf{p}d^{3N}\mathbf{x}$ is the number of representative points that at time t are in the infinitesimal volume $d^{3N}\mathbf{p}d^{3N}\mathbf{x}$ about the point $(\mathbf{x}, \mathbf{p})$ in phase space. Although f_N is a probability distribution function, it evolves in time in a completely deterministic manner, in principle though solving Hamilton's equations.

An ensemble average of the macroscopic property $M(\mathbf{x}, \mathbf{p})$ can be defined by

$$\langle M(t)\rangle = \int M(\mathbf{x}, \mathbf{p}) f_N(\mathbf{x}, \mathbf{p}, t) d^{3N}\mathbf{p} d^{3N}\mathbf{x};$$

i.e., the expectation of the property $M(\mathbf{x}, \mathbf{p})$. Another important and basic postulate of statistical mechanics is the so-called *ergodic statement*, which is that the time average $\bar{M}(\mathbf{x}, \mathbf{p})$

[1]Although it appears that this statement is self-evident, it is essentially a postulate. A basic postulate of both equilibrium and non-equilibrium statistical mechanics is that all macroscopic properties of a given system can be described in terms of the microscopic state of that system.

$$\bar{M}(\mathbf{x}, \mathbf{p}) = \langle M(t) \rangle .$$

The ergodic statement asserts that we can consider ensemble averages rather than time averages as a basis for determining macroscopic properties from the microscopic description. Thus, we need to study the properties and behavior of the probability density function f_N.

The evolution of the pdf f_N is described by Liouville's theorem. The Hamilton equations determine how each ensemble member evolves in phase space. Consider the change df_N in the value of f_N at the point $(\mathbf{x}, \mathbf{p})$ at time t in phase space which results from an arbitrary, infinitesimal change in these variables. This yields

$$df_N = \frac{\partial f_n}{\partial t} dt + \sum_{i=1}^{N} \frac{\partial f_N}{\partial \mathbf{x}_i} \cdot d\mathbf{x}_i + \sum_{i=1}^{N} \frac{\partial f_N}{\partial \mathbf{p}_i} \cdot d\mathbf{p}_i$$

$$\Rightarrow \frac{df_N}{dt} = \frac{\partial f_N}{\partial t} + \sum_{i=1}^{N} \left[\frac{\partial f_N}{\partial \mathbf{x}_i} \cdot \dot{\mathbf{x}}_i + \frac{\partial f_N}{\partial \mathbf{p}_i} \cdot \dot{\mathbf{p}}_i \right]. \tag{2.1}$$

df_N/dt is the total change of f_N along the trajectory in the neighborhood of $(\mathbf{x})$ and $\partial f_N/\partial t$ is the local change in f_N, i.e., at the point $(\mathbf{x})$. Liouville's Theorem is the statement

$$\frac{df_N}{dt} = 0. \tag{2.2}$$

Liouville's theorem states that along the trajectory of any phase point, the probability density in the neighborhood of the point remains constant in time. Since Hamilton's equations have unique solutions, there can be no intersection of trajectories of separate ensemble members in phase space. Thus, an incremental volume about the point $(\mathbf{x})$ in phase space, defined by a specified surface of points in phase space, is also invariant in time, even though it may change its shape (points from inside the volume can never cross the surface since then they would intersect with the points defining the boundary). Since both f_N and the number of points inside the volume $d\mathbf{x}$ remain constant in time, the volume of $d\mathbf{x}$ is unchanged.

Note that since f_N is constant along a trajectory in phase space, so too is any function of f_N. Finally, Liouville's equation is reversible in the sense that the transformation $t \rightarrow -t$ leaves the form of the equation unaltered. Hence, if $f_N((\mathbf{x(t)}), (\mathbf{p(t)}), t)$ is a solution to Liouville's equation, then so is $f_N((\mathbf{x(-t)}), (\mathbf{p(-t)}), -t)$.

2.11 The Language of Fluid Turbulence

The mathematical description in the previous sections provides the statistical tools to understand fluid turbulence, for example, but like many areas of physics, one needs to translate the language of mathematics to that of physics. There will be a slight departure from some of the concepts introduced already in that we will use some of the tools of Fourier theory. As we have seen, an important physical concept is the notion of *ensemble average*, as it allows one to form averages for time-dependent processes.

An example of a random function in space and time in fluid dynamics is the velocity field of a turbulent jet or flow. The macroscopic boundary conditions for the flow field may be independent of time, but the velocity at a point varies in an unpredictable manner in time. The local time-average velocity is different in different locations, as are other averages, such as the square of the velocity departures from the mean $(\mathbf{v} - \mathbf{U})^2$ – the variance. For flows with gross boundary conditions that are constant, we can define time averages. For flows where the boundary conditions are temporal, time averages are not useful, and we need to use ensemble averages.

Consider an ensemble of macroscopically identical experiments, each of which produces as output a variable $u(t)$, where $t > 0$ is the time. The output from the jth experiment is the jth realization of $u(t)$, denoted by ${}^j u(t)$ say. The ${}^j u(t)$ may look like an oscillation with "noise" superimposed, for example, and each realization may be rather different from the others. The ensemble average of the values of $u(t)$ is defined as the limit

$$\langle u(t)\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} {}^j u(t),$$

or one can define the ensemble average of a function $g[u(t)]$ of $u(t)$ in the same way,

$$\langle g[u(t)]\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} g\left[{}^j u(t)\right].$$

Ensemble averages of powers of u are the moments, so the ensemble average of u^k is the kth moment of u at t, i.e.,

$$\langle u(t)^k\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \left[{}^j u(t)\right]^k.$$

If we consider two distinct times t and t' and form the ensemble average of the product $u(t)u(t')$ for each realization, we then define the *covariance* R_{uu} by

$$R_{uu}(t,t') = \langle u(t)u(t')\rangle.$$

A random function is *stationary* if all its moments and joint moments are independent of the choice of time origin. For example, a flow becomes turbulent after passing through a grid at a starting time $t = 0$. After some time $T \gg 0$, the flow will have settled down and initial transients will have damped away. Then, for $t \gg T$, the values of velocities and other variables can be expected to be stationary random functions. Instead of using the above definitions, the stationary ensemble average is equivalent to a time average, e.g.,

$$\langle u\rangle = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} u(t)dt.$$

A stationary random variable is a significant simplification since then averages such as $\langle u(t)\rangle$ are independent if time, as are all expectations and moments of u.

Time covariances of stationary random functions should be independent of the choice of time origin but will depend on the time difference $\tau = t' - t$,

$$R_{uu}(\tau) = \langle u(t)u(t+\tau)\rangle.$$

The double subscript indicates that the covariance is the covariance of u with u, and the argument τ indicates that the velocities are measured at a time interval τ apart. We note too that the velocity is generally three dimensional and so one frequently expresses the tensor $\mathbf{R}$ as

$$R_{ij} = \langle u_i(t)u_j(t+\tau)\rangle.$$

Some properties of the covariance are easily derived.

1. $R_{uu}(\tau)$ is an even function. This can be seen from

$$R_{uu}(\tau) = \langle u(t)u(t+\tau)\rangle = \langle u(t'-\tau)u(t')\rangle = R_{uu}(-\tau).$$

2. The joint covariance of u and its time derivative $\dot{u}$ is the time derivative of R_{uu}. This follows from

$$R_{u\dot{u}}(\tau) = \langle u(t)\dot{u}(t+\tau)\rangle = \frac{\partial}{\partial\tau}\langle u(t)u(t+\tau)\rangle = \frac{\partial}{\partial\tau}R_{uu}(\tau).$$

Related results are derived easily.

Consider two joint random variables – that is, in each realization there are two results (such as the x and y components of velocity, or the velocity and the density at a point in the flow) – say, $u(t)$ and $v(t)$. The joint covariance is then given by

$$R_{uv}(\tau) = \langle u(t)v(t+\tau)\rangle,$$

assuming the process is stationary (since we express the joint covariance in terms of the relative time delay only). Some elementary properties are given in the Exercises.

The covariance and joint covariance functions are assumed to decay for large values of the lag or delay time τ, so that the functions are square integrable and possess Fourier transforms. The Fourier transform of the time *autocovariance* function $R_{uu}(\tau)$ is called the *power spectral density*, defined by

$$S_{uu}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega\tau} R_{uu}(\tau)d\tau.$$

For a function $f(t)$, Fourier's integral theorem yields the expression

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\int_{-\infty}^{\infty} e^{i\omega t'} f(t')dt'd\omega.$$

Using this result in the power spectral density expression then yields

$$R_{uu}(\tau) = \int_{-\infty}^{\infty} e^{-i\omega\tau} S_{uu}d\omega.$$

The joint or cross-spectral density of the joint pair of random functions u and v is given by

$$S_{uv}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega\tau} R_{uv}d\tau = Co_{uv}(\omega) + iQu_{uv}(\omega),$$

where the real part Co_{uv} is called the co-spectrum and the imaginary part Qu_{uv} is called the quadrature spectrum.

Let $u(\mathbf{x}, t)$ be a random function of position $\mathbf{x}$ and time t. The space-time covariance is expressed as

$$R_{uu}(\mathbf{x}, \mathbf{x}', t, t') = \langle u(\mathbf{x}, t)u(\mathbf{x}', t')\rangle.$$

If u is a stationary random function, then R_{uu} is independent of the choice of time origin. If R_{uu} is independent of the choice of spatial origin, and the same is true for other statistical measures, then $u(\mathbf{x}, t)$ is a *homogeneous* function of $\mathbf{x}$. Hence, for a stationary and homogeneous random function, the space-time covariance is

$$R_{uu}(\xi, \tau) = \langle u(\mathbf{x}, t)u(\mathbf{x} + \xi, t + \tau)\rangle.$$

We can Fourier transform the space-time covariance with respect to space and time. The power spectral density of u is the Fourier transform of the time autocovariance $R_{uu}(0, \tau)$, and is given by

$$S_{uu}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega\tau} R_{uu}(0, \tau)d\tau.$$

The *wave number spectrum* $\Phi(k)$ is defined by

$$\Phi(\mathbf{k}) = \left(\frac{1}{2\pi}\right)^3 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-i(\mathbf{k}\cdot\xi)]R_{uu}(\xi,0)d\xi_1 d\xi_2 d\xi_3.$$

Finally, we may define the combined wave number-frequency spectrum as

$$\Phi(\mathbf{k},\omega) = \left(\frac{1}{2\pi}\right)^4 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-i(\mathbf{k}\cdot\xi-\omega\tau)]R_{uu}(\xi,\tau)d\xi_1 d\xi_2 d\xi_3 d\tau.$$

Example. Suppose that we have placed two surface gauges in the middle of the Pacific ocean, separated by a small distance ℓ, so that we can measure the cross-spectral density at each. With these gauges, we want to determine the phase velocity of the random surface waves – these are 2D with velocity (u, v) say. Assuming that the power spectrum is the same at both gauges, we have

$$S_{uv}(\omega,\ell) = [|S_{uu}||S_{vv}|]^{1/2}(\omega)Coh_{uv}(\omega)\exp[i\theta_{uv}(\omega)],$$

where the *coherence* $Coh_{uv}(\omega) = \left[(Co_{uv}^2 + Qu_{uv}^2)/|S_{vv}||S_{uu}|\right]^{1/2}$. The phase velocity of frequency Fourier component can be determined by a straightforward argument. The cross-spectral density is a complex function, so the phase is given by the argument of S_{uv}, i.e.,

$$\theta_{uv}(\omega) = \arg[S_{uv}(\omega,\ell)],$$

and this will give the phase velocity. The phase gives the time interval δt between arrivals of a wave, which is proportional to the phase difference times the wave period divided by 2π,

$$\delta t = \frac{(2\pi/\omega)\theta_{uv}}{2\pi} = \frac{\theta_{uv}}{\omega}.$$

The phase speed is simply the gauge separation distance divided by the time delay, so the phase velocity as a function of frequency is therefore

$$c(\omega) = \frac{\ell}{\delta t} = \frac{\ell\omega}{\theta_{uv}(\omega)}.$$

This expression also obviously gives the average wave number for the frequency as

$$k = \frac{\omega}{c(\omega)} = \frac{\theta_{uv}(\omega)}{\ell}.$$

For surface waves, it is also possible to derive the phase velocity from $S_{uu}(\omega)$. Since, if $u' = \partial u/\partial x$, then $S_{u'u'}(\omega) = k^2 S_{uu}(\omega)$ (Exercise), the mean square wave number is

$$\langle k^2 \rangle = \frac{S_{u'u'}(\omega)}{S_{uu}(\omega)}.$$

The phase speed is then obtained from

$$c_{slope}^2 = \frac{\omega^2}{\langle k^2 \rangle}.$$

Note that the phase speeds calculated from the two methods might well be different. For example, for a standing wave between the two gauges, the phase speed derived from the cross-spectral density would be zero while the slope method may yield a non-zero phase speed. For waves propagating in a single direction, the two methods should give similar results.

Example. Noise is often defined as having all its Fourier components as stochastic variables with zero mean i.e., $\langle u(\omega) \rangle = 0$. A signal therefore corresponds to a definite additive component. In the absence of periodic components, the correlation function or covariance tends to zero as $t \to \infty$. Suppose that the covariance decays exponentially, so that

$$\langle u(\tau)u(0) \rangle = Ce^{-\gamma|\tau|}.$$

The power spectrum is given by (Exercise)

$$S_{uu}(\omega) = \frac{1}{\pi} \frac{C\gamma}{\omega^2 + \gamma^2},$$

and is called the *Lorentz* distribution. Note that *white noise* corresponds to the limit $\gamma \longrightarrow \infty$.

Consider now a periodic component to $u(t)$, say $u(t) = v(t) + Ae^{-i\omega_0 t}$, with $\langle v(t) \rangle = 0$ and $\langle v(\tau)v(0) \rangle = Ce^{-\gamma|\tau|}$. Hence,

$$\langle u(\tau)u(0) \rangle = Ce^{-\gamma|\tau|} + A^2 e^{-i\omega_0 \tau},$$

$$S_{uu}(\omega) = \frac{1}{\pi} \frac{\gamma C}{\omega^2 + 16} + A^2 \delta(\omega - \omega_0),$$

indicating that the periodic signal introduces a spike in the power density spectrum at $\omega = \omega_0$.

Exercises

1. Show that the joint covariance is not symmetric in the time lag τ, i.e., that

$$R_{uv}(\tau) = R_{vu}(-\tau).$$

2. Show that the joint covariance for u and a time derivative of v satisfies

$$R_{u\dot{v}} = \frac{\partial}{\partial\tau} R_{vu}(-\tau).$$

3. Show that the co-spectrum and quadrature spectrum may be expressed as integrals

$$Co_{uv}(\omega) = \frac{1}{2\pi}\int_0^\infty [R_{uv}(\tau) + R_{uv}(-\tau)]\cos(\omega\tau)d\tau;$$
$$Qu_{uv}(\omega) = \frac{1}{2\pi}\int_0^\infty [R_{uv}(\tau) - R_{uv}(-\tau)]\sin(\omega\tau)d\tau.$$

4. By introducing the *coherence* $Coh_{uv}(\omega) = \left[(Co_{uv}^2 + Qu_{uv}^2)/|S_{vv}||S_{uu}|\right]^{1/2}$ and the phase $\theta_{uv}(\omega) = \arg(S_{uv})$ (the argument of S_{uv}), show that the joint- or cross-spectral density can be expressed in terms its magnitude and argument,

$$S_{uv}(\omega) = [|S_{uu}(\omega)||S_{vv}(\omega)|]^{1/2}\, Coh_{uv}(\omega)\exp[i\theta_{uv}(\omega)].$$

5. Show that if $u' = \partial u/\partial x$, then $S_{u'u'} = k^2 S_{uu}(\omega)$.
6. Show that an exponentially decaying covariance $\langle u(\tau)u(0)\rangle = Ce^{-\gamma|\tau|}$ yields a Lorentz distribution for the power spectral density,

$$S_{uu}(\omega) = \frac{1}{\pi}\frac{\gamma C}{\omega^2 + \gamma^2}.$$

Sketch the covariance and the power spectral density.

7. Show that the autocovariance in the last example of the chapter is given by

$$\langle u(\tau)u(0)\rangle = Ce^{-\gamma|\tau|} + A^2 e^{-i\omega_0\tau},$$

and that

$$S_{uu}(\omega) = \frac{1}{\pi}\frac{\gamma C}{\omega^2 + 16} + A^2\delta(\omega - \omega_0).$$

Sketch the covariance and the power spectral density.

References

W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, 3rd edn. (John Wiley, New York, 1968)

R.V. Hogg, A.T. Craig, *Introduction to Mathematical Statistics*, 4th edn. (McMillan, New York, 1978)

I.N. Gibra, *Probability and Statistical Inference for Scientists and Engineers*. (Prentice Hall, Englewood Cliffs, 1973)

http://www.springer.com/978-1-4614-8479-0

Transport Processes in Space Physics and Astrophysics
Zank, G.P.
2014, XI, 286 p. 21 illus., 1 illus. in color., Softcover
ISBN: 978-1-4614-8479-0