# Chapter 3

# Statistical Averages

## 3.1 Introduction

Statistical averages are important in the measurement of quantities that are obscured by random variations. As an example to motivate the discussion, consider the problem of measuring a voltage level with a noisy instrument. Suppose that the unknown voltage has value $a$ and that the instrument has an uncertainty $x$. The observed value may be $y = a + x$. Suppose that $n$ independent measurements are made under identical conditions, meaning that neither the unknown value of the voltage nor the statistics of the instrument noise change during the process. Let us call the $n$ measurements $y_i$, $1 \leq i \leq n$. Under our model of the process, it must be the case that $y_i = a + x_i$. Now form the quantity

$$\overline{y}(n) = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.1}$$

This is the *emperical average* of the observed values. It is important to note that $\overline{y}(n)$ is a *random variable* because it is a numerical value that is the outcome of a random experiment. That means that it will not have a single certain value. We expect to obtain a different value if we repeat the experiment and obtain $n$ new measurements. We also expect that the result depends upon the value of $n$, and have the sense that larger values of $n$ should give better results.

This intuition is basically correct. The purpose of this analysis is to refine what we understand intuitively and to identify the ways in which our intuition can be fooled.

If we substitute $y_i = a + x_i$ into the above expression we find

$$\overline{y}(n) = a + \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.2}$$

The summation represents the emperical average $\overline{x}(n)$. Hence, we have arrived at a result $\overline{y}(n) = a + \overline{x}(n)$. We cannot actually observe $\overline{x}(n)$ because it is the result of the unknown measurment noise. So we are led to wonder how $\overline{y}(n) = a + \overline{x}(n)$ tells us more about the value of $a$ than what $y = a + x$ tells us. The answer is that the variance of $\overline{y}(n)$ can be much smaller than the variance of $y$. We have not yet defined the term *variance* but we will shortly. It is a measure of our uncertainty. If the noise on each measurement is independent, then the variance decreases as $1/n$. By making $n$ sufficiently large it can be made as small as we wish.

Let us now turn to setting up the machinery. The approach will be to bring in the probability distributions for the unknown quantities.

## 3.2 Discrete Random Variables

Suppose that a random variable $U$ can take on any one of $L$ random values, say $u_{1,}u_2, \ldots u_L$. Imagine that we make $n$ independent observations of $U$ and that the value $u_k$ is observed $n_k$ times, $k = 1, 2, \ldots, L$. Of course, $n_1 + n_2 + \cdots + n_L = n$. The emperical average can be computed by

$$\overline{u} = \frac{1}{n}\sum_{k=1}^{L} n_k u_k = \sum_{k=1}^{L} \frac{n_k}{n} u_k \tag{3.3}$$

We recognize that the ratio $n_k/n$ is the relative frequency of level $u_k$. Recall that the relative frequency is a stand-in for the probability of observing level $u_k$. This motivates the definition of the expected value.

**Definition 3.2.1** *The* expected value *of a discrete random variable $U$ is*

$$E[U] = \sum_{k=1}^{L} u_k \, P[U = u_k] \tag{3.4}$$

The definition may be motivated by our ideas of calculating an average. However, it stands on its own–it does not depend upon the emperical notions. The expected value is a specific value that can be calculated once the probability distribution is given. The expected value is identical to the mean value of a quantity. The emperical average, on the other hand, is a random variable. So we will have to inquire into the relationship between $\overline{u}$ and $E[U]$.

If we want to determine the value of the voltage $a$ in the example, we need to make use of observed data. Although we may know the statistics of the observation noise, we do not know which values it presented on a particular set of measurements. Although we may know $E[X]$, that fact does us little good when trying to find the value of $a$ from the observations. We are forced to do calculations that are based on the observations we actually have. We would therefore do something like (1) compute $\overline{y}(n)$ and (2) use $\overline{y}(n)$ as an estimate of the value of $a$. In fact, this is the thing we should do, for, as we will show, the expected value of $\overline{y}(n)$ is equal to $a$.

Let us return to a consideration of the relationship between $\overline{u}$ and $E[U]$. An examination of (3.3) and (3.4) shows that all of the quantities in the second formula are fixed whereas the quantities $n_k$, $k = 1, 2, \ldots, L$, in the first formula depend upon the outcome of the measurement experiment. We expect to observe a value of $n_k$ that is close to $nP[U = u_k]$ but that is not certain. Hence, $\overline{u}$ will vary from one experiment to the next. It is a random variable. Its characteristics depend upon the size $n$ of the sample set.

**Example 3.2.1** ***Fair Die*** *Let an experiment consist of $n$ tosses of a fair die, and let $\overline{u}$ be the sum of the faces observed on the $n$ trials divided by $n$. Let $U$ be a random variable that is produced when a die is tossed once and the face count is observed. Then*

$$E[U] = \frac{1}{6} \times (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3\frac{1}{2} \tag{3.5}$$

By repeating the experiment of computing $\overline{u}$ many times one can obtain a distribution of the results that can be observed. The upper plot in Figure 3.1 shows the distribution of values of $\overline{u}$ with $n = 12$ and the lower plot shows the distribution of values of $\overline{u}$ with $n = 120$. It is clear that it is more likely that an emperical average close to $E[U]$ will be obtained when $n$ is large.

One of the most important tools in the analysis of averages of a random variable is the idea of the average of a function of a random variable. We will look at that topic next.

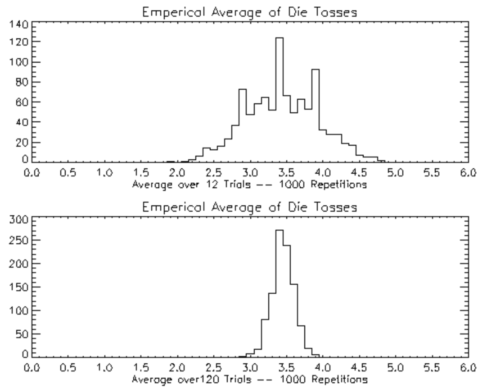


Figure 3.1: The empirical average of the results of tossing a fair die. The averages in the upper figure are over 12 tosses and in the lower figure they are over 120 tosses. The distribution of results from 1000 trials for each experiment produce the histograms.

### 3.2.1 Functions of a Discrete Random Variable

Let $U$ be a random variable and let $V = g(U)$, where $g$ is a bounded function with a finite number of discontinuities. Suppose that we want to compute the average value of $V$ rather than of $U$. Given that $U$ can take on the values $u_k$ with probabilities $P[U = u_k]$, $k = 1, 2, \ldots, L$, we can compute both the values and probabilities that can be taken on by $V$. Let $v_1, v_2, \ldots, v_r$ be the set of values that can be assumed by $V$ and let $P[V = v_j]$, $j = 1, 2, \ldots r$ be the corresponding probabilities. Then

$$E[V] = \sum_{j=1}^{r} v_j P[V = v_j] \tag{3.6}$$

The computation of the values of $V$ is illustrated in Figure 3.2. In the example shown the function $v = g(u)$ has the property that more than one value of $u$ can map to the same value of $v$. This adds a mild degree of complexity to the calculation, but no conceptual difficulty. We see that $v_1 = g(u_1)$, $v_2 = g(u_5) = g(u_6)$, $v_3 = g(u_2) = g(u_4) = g(u_7)$, and $v_4 = g(u_3)$. Therefore the probability can be calculated for each of the $v_j$. $P[V = v_1] = P[U = u_1]$, $P[V = v_2] = P[U = u_5] + P[U = u_6]$, $P[V = v_3] = P[U = u_2] + P[U = u_4] + P[U = u_7]$, and $P[V = v_4] = P[U = u_3]$. We can now

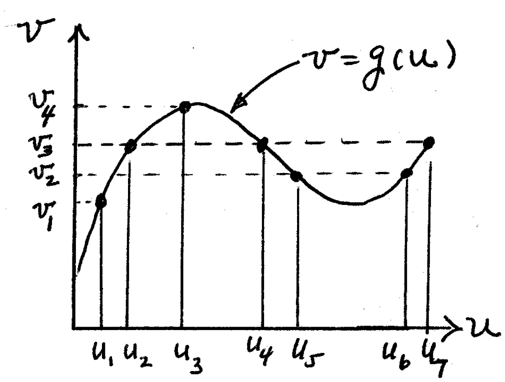


Figure 3.2: Mapping of a discrete random variable by the function $g(u)$. The inverse function is multiple-valued because, for example, $v_3 = g(u_2) = g(u_4) = g(u_6)$.

substitute into (3.6) and obtain

$$\begin{aligned} E[V] = & \quad v_1 P[U = u_1] + v_2(P[U = u_5] + P[U = u_6]) \\ + & \quad v_3(P[U = u_2] + P[U = u_4] + P[U = u_7]) + v_4 P[U = u_3]. \end{aligned} \tag{3.7}$$

By making use of the relationship $v = g(u)$ we can now rewrite this in a form that associates $g(u_k)$ with $P[U = u_k]$. The result is the nice compact equation

$$E[V] = \sum_{k=1}^{L} g(u_k) P[U = u_k] \tag{3.8}$$

You should work through the details that link (3.6) to (3.8). Note that the final result is very compact and that it can be computed directly from knowledge of quantities in the $U$ domain. The transformation is taken care of automatically, even if the inverse function is multiple-valued. This is a powerful and useful result–one which we will use repeatedly.

### 3.2.2 Multivariate Functions

Let $\mathbf{U} = [U_1, U_2, \ldots, U_n]$ be a vector of random variables. Each component of $\mathbf{U}$ is a random variable–which may have been obtained in a number of ways, such as repeating an experiment $n$ times, or by forming $n$ different functions on the outcome of one experiment–and thus **U** may be used to model many situations in which several random variables must be considered. Let $g(\mathbf{U})$ be a function of the random vector. We often need tools to analyze $E[g(\mathbf{U})]$ and interpret its meaning. There are two different cases to be considered: (a) $V = g(\mathbf{U})$ is a scalar function. That is, the result is a scalar value; (b) $\mathbf{V} = g(\mathbf{U})$ is a vector value.The second case can be handled by combining results of the first case, so we start with the scalar case first.

By analogy to (3.8) we can write[1]

$$E[V] = E[g(\mathbf{U})] = \sum_{u_1}\sum_{u_2}\cdots\sum_{u_n} g(u_1, u_2, \ldots u_n) P\left[U_1 = u_1, U_2 = u_2, \ldots U_n = u_n\right] \tag{3.9}$$

where the summation is over all possible values for the components. This computation requires knowledge of the multivariate probability function, but is technically straightforward to do.

The case of a vector function is handled by noting that one can write $\mathbf{V} = g(\mathbf{U})$ in terms of components. Suppose that **V** has $m$ components so that $\mathbf{V} = [V_1, V_2, \ldots, V_m]$ where each component is a scalar. Each component can depend upon $U$ in a different way, so we can write

$$\mathbf{V} = [g_1(\mathbf{U}), g_2(\mathbf{U}), \ldots, g_m(\mathbf{U})] \tag{3.10}$$

Then we define the expected value of **V** as

$$E[\mathbf{V}] = [E[g_1(\mathbf{U})], \mathbf{E}[g_2(\mathbf{U})], \ldots, \mathbf{E}[g_m(\mathbf{U})]] \tag{3.11}$$

where each term is then computed by use of (3.9). Again, all that is required is knowledge of the multivariate probability function and the $m$ scalar functions, $g_k(U)$, $k = 1, 2, \ldots, m$.

[1] The development of this result is reasonably straightforward, but somewhat detailed and tedious. We will simply state it here.

**Example 3.2.2** *Let $V = U_1 + U_2 + \cdots + U_n$. Then, using (3.9) we can write*

$$\begin{aligned} E[V] &= \sum_{u_1}\sum_{u_2}\cdots\sum_{u_n}(u_1+u_2+\cdots+u_n)P\left[U_1=u_1,U_2=u_2,\ldots U_n=u_n\right] \\ &= \sum_{u_1}u_1\sum_{u_2}\cdots\sum_{u_n}(P\left[U_1=u_1,U_2=u_2,\ldots U_n=u_n\right] \\ &\quad+\sum_{u_2}u_2\sum_{u_1}\cdots\sum_{u_n}(P\left[U_1=u_1,U_2=u_2,\ldots U_n=u_n\right]+\cdots \\ &\quad+\sum_{u_n}u_n\sum_{u_1}\cdots\sum_{u_{n-1}}(P\left[U_1=u_1,U_2=u_2,\ldots U_n=u_n\right] \end{aligned}$$

*The inner sum in each case is over the joint probability function with all of the $u_k$ parameters except the one that is the factor in front. The inner sum then reduces to just the marginal probability for the component that is omitted. The result is a very simple expression*

$$E[V] = \sum_{k=1}^{n} u_k P[U_k = u_k] = E[U_1] + E[U_2] + \cdots + E[U_n] \tag{3.12}$$

*The expected value of the sum is the sum of the expected values.*

## 3.3 Continuous Random Variables

We have seen that continuous random variables are described by probability density functions. The basic procedure for dealing with expectations of continuous random variables is similar to that we used with discrete random variables. The approach is to approximate the continuous random variable with a discrete random variable and then use a limiting process. A careful analysis that is built on this approach is technically detailed and will not be done here. You are referred to the probability literature for details.

We will define the expected value of a continuous random variable $U$ by

$$E[U] = \int_{-\infty}^{\infty} u f_U(u)du \tag{3.13}$$

**Example 3.3.1** *Let $U$ be a random variable with an exponential distribution with parameter $a$.*

$$f_U(u) = \begin{cases} ae^{-au} & \text{for } u \geq 0 \\ 0 & \text{for } u < 0 \end{cases} \tag{3.14}$$

*To find the expected value of $U$ we calculate the integral*

$$E[U] = \int_0^\infty uae^{-au}du = \frac{1}{a} \tag{3.15}$$

*The parameter $a$ is therefore the reciprocal of the expected value. You should sketch the exponential distribution for a few values of $a$ and convince yourself that this is a reasonable result.*

By analogy with the discrete case, the expected value of a function of a random variable $V = g(U)$ is

$$E[V] = \int_{-\infty}^\infty g(u)f_U(u)du \tag{3.16}$$

**Example 3.3.2** *Let $U$ have an exponential distribution with parameter $a$, and let $V = U^2$. Find $E[V] = E[U^2]$. By making use of the definition we find*

$$E[U^2] = \int_0^\infty u^2ae^{-au}du = \frac{2}{a^2} \tag{3.17}$$

### Multivariate Functions

Let $\mathbf{U} = [U_1, U_2, \ldots, U_n]$ be a vector of random variables. Just as in the discrete case, we can define a function $V = g(\mathbf{U})$ and seek to compute the expected value of $V$. This requires the multivariate probability density function. The calculation of moments can be done directly using the multivariate pdf

$$E[V] = \int\int\cdots\int g(u_1, u_2, \ldots, u_n)f_{\mathbf{U}}(u_1, u_2, \ldots u_n)du_1du_2 ... du_n \tag{3.18}$$

where the integration is over the infinite range of values for each variable. A function $\mathbf{V} = \mathbf{g}(\mathbf{U})$ is then handled in a manner similar to (3.10).

An interesting example of a multivariate transformation is provided by the following example. This transformation arises in many physical applications, but we will phrase it in terms of a game.

**Example 3.3.3** *Consider the problem of throwing darts at a target. Let $X$ and $Y$ be random variables that represent the horizontal and vertical offset of the dart from the center of the target. Note that the center is at (0,0) so* *that the variables can have both negative and positive values. Suppose that* $X$ *and* $Y$ *are statistically independent normal random variables, with*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{x^2}{2\sigma^2}\right]$$

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{y^2}{2\sigma^2}\right]$$

*We will see below that the parameter* $\sigma$ *is the standard deviation, and controls the spread of the distribution of the dart throws. A small* $\sigma$ *means that the thrower is very accurate. Now, because the horizontal and vertical offsets are statistically independent, the joint distribution is*

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{\sigma^2 2\pi} exp\left[-\frac{x^2+y^2}{2\sigma^2}\right] \tag{3.19}$$

*Suppose that we have to calculate the probability that the dart falls within a radius* $r$ *of the center. Let* $R = \sqrt{X^2+Y^2}$ *be the distance of the dart from the center. We need to find the pdf of* $R$ *given the above information. By substitution of* $r^2 = x^2 + y^2$ *we note that the normal distribution has the same value for all points that are at the same distance from the center. The volume above a ring of thickness* $dr$ *and mean radius* $r$ *is* $2\pi r f_{X,Y}(r)dr$*. This must be the same value as achieved under the distribution of* $f_R(r)$ *over the interval* $dr$*. Therefore,* $f_R(r)dr = 2\pi r f_{X,Y}(r)dr$ *where* $r^2 = x^2 + y^2$ *in the latter expression. Therefore,*

$$f_R(r) = \frac{r}{\sigma^2} exp\left[-\frac{r^2}{2\sigma^2}\right] \tag{3.20}$$

*Let us replace* $\sigma^2$ *with a parameter* $b$*. The result is*

$$f_R(r) = \frac{r}{b} exp\left[-\frac{r^2}{2b}\right]$$

*This is known as the Rayleigh probability density or Rayleigh distribution with parameter* $b$*. A curve for the value* $b = 1$ *is shown in Figure 3.3. It seems to say that one likely to hit the target a ways from the center, since the maximum is some distance away. The way the curve changes with* $b$ *is investigated in Exercise 8.*

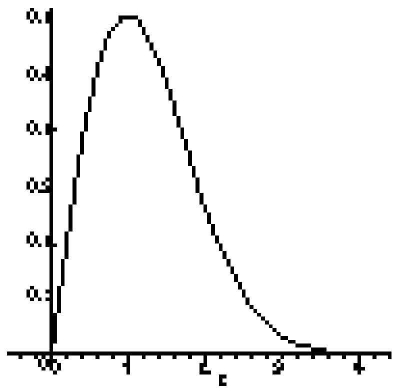

Figure 3.3: A plot of the Rayleigh distribution for the case $b = 1$.

## 3.4 Moments

Moments are parameters that are important in the characterization of probability distributions and probability density functions. The $k^{\text{th}}$ moment of a random variable $U$ is defined as $E[U^k]$, and can be computed by the following formulas for the discrete and continuous cases, respectively.

$$E[U^k] = \sum_{i=1}^{L} u_i^k P[U = u_i] \tag{3.21}$$

$$E[U^k] = \int_{-\infty}^{\infty} u^k f_U(u) du \tag{3.22}$$

The cases $k = 1$ and $k = 2$ are analagous to moments that are used in mechanics. If a body has a linear distribution of mass, $m(u)$ then the total mass is

$$M = \int_{-\infty}^{\infty} m(u) du$$

The center of mass, or centroid, is

$$\mu = \frac{1}{M} \int_{-\infty}^{\infty} u m(u) du$$

If the mass is normalized so that $M = 1$ then the centroid of mass is analagous to the first moment. The distribution of probability is analagous to the distribution of mass, and the expected value is analagous to the centroid.

The probability distribution will "balance" at the point $E[U]$. It is therefore common to use the symbol $\mu$ to represent the expected value. We will follow that practice.

The mechanical moment of inertia about the origin is

$$I = \int_{-\infty}^{\infty} u^2 m(u) du$$

If we identify $m(u)/M \sim f_U(u)$ then $I \sim ME[U^2]$. If the mass is normalized to unity then the moment of inertia about the origin and the second moment are equivalent. We know that a distribution with a high moment of inertia is spread out from the origin more than one with a low moment of inertia.

The location of the origin relative to the centroid influences the moment of inertia. To remove the effect of the location of the origin, it is common to compute the second moment about the centroid. This is

$$I_0 = \int_{-\infty}^{\infty} (u - \mu)^2 m(u) du$$

The analagous operation with probability distributions is the *variance*[2]:

$$\sigma_U^2 = E[(U - E[U])^2] = E[(U - \mu)^2] \tag{3.23}$$

$$\sigma_U^2 = \int_{-\infty}^{\infty} (u - \mu)^2 f_U(u) du \tag{3.24}$$

The relationship $\sigma_U^2 \sim I_0$ is obvious. The *variance* is a measure of the spread of the function about the mean. The equivalent expression for a discrete distribution is

$$\text{var}(U) = \sigma_U^2 = \sum_{k=1}^{L} (u_k - \mu)^2 P[U = u_k] \tag{3.25}$$

By expanding (3.23) one can readily show a useful relationship

$$\text{var}(U) = \sigma_U^2 = E[U^2] - E^2[U] \tag{3.26}$$

If one can compute any two of these quantities then the third is available from this expression. The quantity $\sigma_U$ is called the *standard deviation* of the random variable $U$.

[2] Here we use $\mu$ for $E[U]$.

**Example 3.4.1** *Find the mean and standard deviation of a Rayleigh distribution with parameter b. The Rayleigh distribution is given by*

$$f_R(r) = \frac{r}{b} e^{-r^2/2b} \quad (r \geq 0) \tag{3.27}$$

*The mean value is given by*

$$\mu_R = E[R] = \int_0^\infty \frac{r^2}{b} e^{-r^2/2b} dr = \sqrt{\frac{\pi b}{2}} \tag{3.28}$$

*The integral can be computed using integration by parts. The mean-squared value is, again using integration by parts,*

$$E[R^2] = \int_0^\infty \frac{r^3}{b} e^{-r^2/2b} dr = 2b \tag{3.29}$$

*The variance is, from (3.26)*

$$var(R) = \sigma_R^2 = 2b - \frac{\pi b}{2} \approx 0.43b \tag{3.30}$$

*The standard deviation is* $\sigma_R \approx 0.655\sqrt{b}$*. By a little calculus you can show that the peak of the distribution occurs at* $r = \sqrt{b}$*, so that the standard deviation is about 65% of the location of the peak. As b is increased the curve shifts to the right and broadens in proportion to* $\sqrt{b}$*.*

## 3.5 Exercises

1. Let $U$ be a discrete random variable and let $V = aU + b$ where $a$ and $b$ are constants. Show that $E[V] = aE[U] + b$.

2. Let $U$ be a discrete random variable. Show that $|E[U]| \leq E[|U|]$ and specify the conditions that must be true for equality to hold. [Hint: Use the triangle inequality $|\sum_k r_k| \leq \sum_k |r_k|$].

3. Let $V = a_1U_1 + a_2U_2 + \cdots + a_nU_n$. Carry out an analysis similar to that of Example 3.2.2 to find $E[V]$ in terms of the expectations of the $U_k$.

4. Suppose that $U_k$ is a binomial random variable that takes on the value 1 with probability $p$ and the value 0 with probability $(1 - p)$. Let $V = U_1 + U_2 + \cdots + U_n$ be the sum of $n$ such binomial random variables. Show that $E[V] = np$.

5. Let $U_1$ and $U_2$ be statistically independent random variables, and let $V = U_1U_2$. Show that $E[V] = E[U_1]E[U_2]$. Make specific use of the assumption of statistical independence.

6. Let $X$ be a discrete random variable with the Poisson probability distribution
$$P[X = k] = \frac{\mu^k e^{-\mu}}{k!} \quad k = 0, 1, 2, \ldots \tag{3.31}$$
   (a) Show that $\sum_{k=0}^{\infty} P[X = k] = 1$
   (b) Show that $E[X] = \mu$
   (c) Show that $\text{var}(U) = \mu$. That is, the expected value and the variance of a Poisson distribution are equal.

7. Let $X$ be a normal random variable with the probability density function
$$f_X(x) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2s^2}\right] \tag{3.32}$$
   (a) Show that $\int_{-\infty}^{\infty} f_X(x)dx = 1$
   (b) Show that $E[X] = a$, so that writing $\mu$ in the position occupied by $a$ is a sensible thing to do.

(c) Show that $\text{var}(X) = s^2$ so that writing $\sigma$ in the position occupied by $s$ is a sensible thing to do.

8. Verify the results for the mean and standard deviation of the Rayleigh distribution that are given in Example 3.4.1. Plot the Rayleigh distribution for $b = 1, 4, 9, 25$ and observe the changes in the graph.

## 3.6 Characteristic Functions

The characteristic function of a random variable $X$ is the expected value of the function $e^{juX}$ where $j = \sqrt{-1}$. If $X$ has a continuous probability density function $f_X(x)$ then the characteristic function is

$$M_X(ju) = E\left[e^{juX}\right] = \int_{-\infty}^{\infty} f_X(x)e^{jux}dx \tag{3.33}$$

If $X$ has a discrete distribution with nonzero probability weights $f_X(x_i)$ at the points $x_i$, $i = 0, 1, 2, \ldots$ then the characteristic function is defined by the sum

$$M_X(ju) = E\left[e^{juX}\right] = \sum_i f_X(x_i)e^{jux_i} \tag{3.34}$$

Note that the characteristic function is closely related to the Fourier transform of the probability density function. Since $f_X(x)$ is nonnegative and $e^{jux}$ has unit magnitude for all values of $x$ and $u$,

$$\left|\int_{-\infty}^{\infty} f_X(x)e^{jux}dx\right| \leq \int_{-\infty}^{\infty} \left|f_X(x)e^{jux}\right| dx = \int_{-\infty}^{\infty} f_X(x)dx = 1$$

Hence, the characteristic function always exists and

$$|M_X(ju)| \leq M_X(0) = 1 \tag{3.35}$$

It is sometimes easier to find the characteristic function than it is to find the pdf. Then, under suitable conditions, one can invert the characteristic function to find the pdf.

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(ju)e^{-jux}du \tag{3.36}$$

### 3.6.1 Moment Generation

The function $e^{jux}$ can be expanded in a Taylor series:

$$e^{juX} = \sum_{n=0}^{\infty} \frac{(juX)^n}{n!} = \sum_{n=0}^{\infty} \frac{(ju)^n}{n!} X^n \tag{3.37}$$

Hence,

$$M_X(ju) = \sum_{n=0}^{\infty} \frac{(ju)^n}{n!} E[X^n] = \sum_{n=0}^{\infty} \frac{(ju)^n}{n!} m_n \tag{3.38}$$

where $m_n$ is the $n^{th}$ moment of $X$ about the origin. If we expand the characteristic function in a Taylor series, then $m_n$ can be found from the coefficient of $u^n$ in the expansion. Specifically, if

$$M_X(ju) = \sum_{n=0}^{\infty} k_n u^n \tag{3.39}$$

then $m_n = n!k_n/j^n$. The moment can also be found by taking derivatives in (3.37).

$$E[X^m] = \frac{1}{j^m} \left. \frac{d^m M_X}{du^m} \right|_{u=0} \tag{3.40}$$

### 3.6.2 Example: Exponential Probability Density Function

Consider the exponential pdf given by

$$f_X(x) = be^{-bx}\text{step}(x) \tag{3.41}$$

The characteristic function is

$$M_X(u) = \int_0^{\infty} be^{-(b-ju)x} dx = \frac{b}{b - ju} \tag{3.42}$$

Within the circle $|u| < b$ we can expand the term on the right in a power series by using (for $|t| < 1$ )

$$\frac{1}{1-t} = 1 + t + t^2 + t^3 + \cdots \tag{3.43}$$

Then the expansion, with $t = ju/b$, is

$$M_X(u) = \sum_{n=0}^{\infty} \left(\frac{ju}{b}\right)^n \tag{3.44}$$

In this case $k_n = (j/b)^n$ so that

$$m_n = E\left[X^n\right] = \frac{n!}{b^n} \tag{3.45}$$

### 3.6.3 Example: Binomial Distribution

The binomial distribution describes the results of Bernoulli trials, discussed in the next chapter. The outcomes takes on discrete values $X_k = k$, $\{k = 0, 1, 2, \ldots, n\}$, where the random variable $X_k$ is the event that there are $k$ successes in $n$ trials. The c.f. is then given by the discrete equation (3.34). The probability of $k$ successes in $n$ trials is governed by the binomial probability distribution $b(n,k,p) = \binom{n}{k} p^k (1-p)^{n-k}$. The c.f. is

$$\begin{aligned} M_X(ju) &= \sum_{k=0}^{n} b(n,k,p) e^{juk} = \sum_{k=0}^{n} \tbinom{n}{k} p^k (1-p)^{n-k} e^{juk} \\ &= \sum_{k=0}^{n} \tbinom{n}{k} (pe^{ju})^k (1-p)^{n-k} = (pe^{ju} + 1 - p)^n \end{aligned} \tag{3.46}$$

Here we have made use of the binomial expansion

$$(a+b)^n = \sum_{k=0}^{n} \tbinom{n}{k} a^k b^{n-k} \tag{3.47}$$

The mean value is found by differentiation.

$$m_1 = E[X] = \frac{1}{j} \left.\frac{dM}{du}\right|_{u=0} = npe^{ju}(pe^{ju} + 1 - p)^{n-1}\Big|_{u=0} = np \tag{3.48}$$

Similarly,

$$\begin{aligned} m_2 &= E[X^2] = \frac{1}{j^2} \left.\frac{d^2M}{du^2}\right|_{u=0} \\ &= n(n-1)p^2 e^{ju}(pe^{ju} + 1 - p)^{n-2} + npe^{ju}(pe^{ju} + 1 - p)^{n-1}\Big|_{u=0} \\ &= n(n-1)p^2 + np \end{aligned} \tag{3.49}$$

The variance is easily found by

$$\sigma^2 = \text{var}[X] = E[X^2] - E^2[X] = n(n-1)p^2 + np - n^2p^2 = np(1-p) \quad (3.50)$$

These values are calculated by a direct means in the problems of the next chapter. This example shows the utility of the characteristic function approach.

### 3.6.4 Sum of Independent Random Variables

Let $W = X + Y$ where $X$ and $Y$ are statistically independent random variables. We have seen that the probability density functions are related by the convolution

$$f_W(x) = \int_t f_X(t) f_Y(x-t)dt = f_X \star f_Y \quad (3.51)$$

The c.f. can be found by doing the transform

$$M_W(ju) = \int_{-\infty}^{\infty} f_W(x) e^{jux} dx \quad (3.52)$$

After some manipulation it is found that the c.f. is given by the product

$$M_W(ju) = M_X(ju) M_Y(ju) \quad (3.53)$$

This result can be extended to a sum of any number of random variables.

### 3.6.5 Homework

1. Show by direct integration that $E[X^n] = \frac{n!}{b^n}$ for the exponential distribution $f_X(x) = be^{-bx}\text{step}(x)$. You can do this by mathematical induction and integration by parts.

2. Show that $M(ju) = e^{-\sigma^2 u^2/2}$ is the characteristic function of the normal distribution $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$. [HINT: Set up the integral and then complete the square in the exponent. Recognize the integral that results.]

3. Suppose that $M_X(ju)$ is the c.f. for the random variable $X$ and that $Y = X + a$. What is the c.f. for $Y$?

4. Consider the Poisson distribution, studied in the next chapter, in which the discrete random variable $X_k = k$ has the probability distribution $P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$. Show that the c.f. is

$$M(ju) = e^{\lambda(e^{ju}-1)}$$

   Find the mean and variance of the distribution.

5. Let $W$ be the sum of $n$ statistically independent normal random variables each of which has mean $\mu$ and standard deviation $\sigma$. Find the c.f. of $W$ and from that deduce $\mu_W$, $\sigma_W$, and the ratio of the standard deviation to the mean value in terms of $\mu$, $\sigma$,and $n$.

6. In this problem we remove the stipulation that $X$ has a normal distribution. Let $W$ be the sum of $n$ statistically independent random variables each of which has mean $\mu$ and standard deviation $\sigma$. Express the c.f. of $W$ in terms of the c.f. of $X$. From that general form and the rule for computing moments deduce $\mu_W$, $\sigma_W$, and the ratio of the standard deviation to the mean value in terms of $\mu$, $\sigma$,and $n$. You do not need to know the explicit distribution of $X$ to do this, only that it exists.