Chapter 6

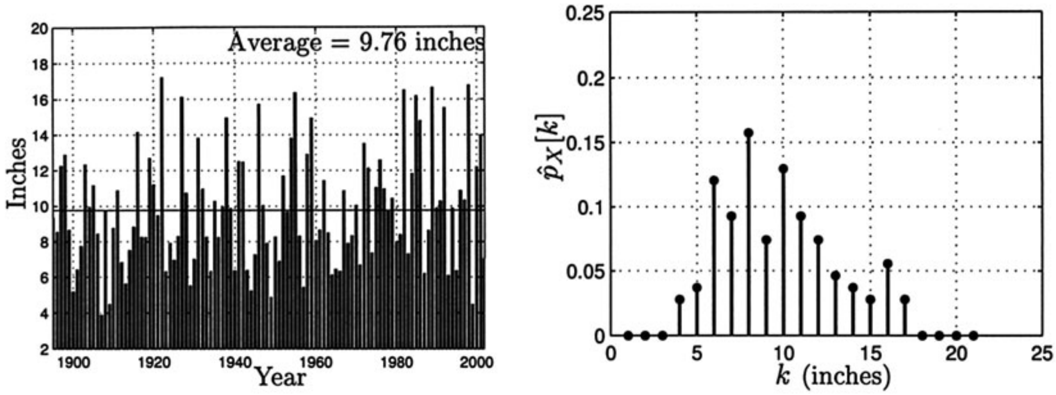
Expected Values for Discrete Random Variables

6.1 Introduction

The probability mass function (PMF) discussed in Chapter 5 is a complete description of a discrete random variable. As we have seen, it allows us to determine probabilities of any event. Once the probability of an event of interest is determined, however, the question of its interpretation arises. Consider, for example, whether there is adequate rainfall in Rhode Island to sustain a farming endeavor. The past history of yearly summer rainfall was shown in Figure 1.1 and is repeated in Figure 6.1a for convenience. Along with it, the estimated PMF of this yearly data is shown in Figure 6.1b (see Section 5.9 for a discussion on how to estimate the PMF). For a particular crop we might need a rainfall of between 8 and 12 inches. This event has probability 0.5278, obtained by $\sum_{k=8}^{12} \hat{p}_X[k]$ for the estimated PMF shown in Figure 6.1b. Is this adequate or should the probability be higher? Answers to such questions are at best problematic. Rather we might be better served by ascertaining the *average* rainfall since this is closer to the requirement of an adequate amount of rainfall. In the case of Figure 6.1a the average is 9.76 inches, and is obtained by summing all the yearly rainfalls and dividing by the number of years. Based on the given data it is a simple matter to estimate the average value of a random variable (the rainfall in this case). Some computer simulation results pertaining to averages have already been presented in Example 2.3. In this chapter we address the topic of the *average* or *expected value* of a discrete random variable and study its properties.

6.2 Summary

The expected value of a random variable is the average value of the outcomes of a large number of experimental trials. It is formally defined by (6.1). For discrete random variables with integer values it is given by (6.2) and some examples of its



(a) Annual summer rainfall (b) Estimated PMF

Figure 6.1: Annual summer rainfall in Rhode Island and its estimated probability mass function.

determination given in Section 6.4. The expected value does not exist for all PMFs as illustrated in Section 6.4. For functions of a random variable the expected value is easily computed via (6.5). It is shown to be a linear operation in Section 6.5. Another interpretation of the expected value is as the best predictor of the outcome of an experiment as shown in Example 6.3. The variability of the values exhibited by a random variable is quantified by the variance. It is defined in (6.6) with examples given in Section 6.6. Some properties of the variance are summarized in Section 6.6 as Properties 1 and 2. An alternative way to determine means and variances of a discrete random variable is by using the characteristic function. It is defined by (6.10) and for integer valued random variables it is evaluated using (6.12), which is a Fourier transform of the PMF. Having determined the characteristic function, one can easily determine the mean and variance by using (6.13). Some examples of this procedure are given in Section 6.7, as are some further important properties of the characteristic function. An important property is that the PMF may be obtained from the characteristic function as an inverse Fourier transform as expressed by (6.19). In Section 6.8 an example is given to illustrate how to estimate the mean and variance of a discrete random variable. Finally, Section 6.9 describes the use of the expected value to reduce the average code length needed to store symbols in a digital format. This is called data compression.

6.3 Determining Averages from the PMF

We now discuss how the average of a discrete random variable can be obtained from the PMF. To motivate the subsequent definition we consider the following game of chance. A barrel is filled with US dollar bills with denominations of \$1, \$5, \$10, and \$20. The proportion of each denomination bill is the same. A person playing the game gets to choose a bill from the barrel, but must do so while blindfolded. He pays \$10 to play the game, which consists of a single draw from the barrel. After he observes the denomination of the bill, the bill is returned to the barrel and he wins that amount of money. Will he make a profit by playing the game many times? A typical sequence of outcomes for the game is shown in Figure 6.2. His average

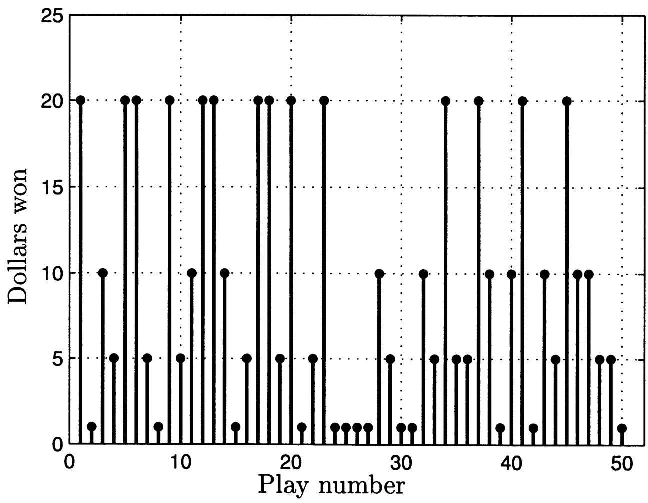


Figure 6.2: Dollar winnings for each play.

winnings per play is found by adding up all his winnings and dividing by the number of plays N. This is computed by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where x_i is his winnings for play i. Alternatively, we can compute $\bar{x}$ using a slightly different approach. From Figure 6.2 the number of times he wins k dollars (where $k = 1, 5, 10, 20$) is given by N_k, where

$$\begin{aligned} N_1 &= 13 \\ N_5 &= 13 \\ N_{10} &= 10 \\ N_{20} &= 14. \end{aligned}$$

As a result, we can determine the average winnings per play by

$$\begin{aligned}
\bar{x} &= \frac{1 \cdot N_1 + 5 \cdot N_5 + 10 \cdot N_{10} + 20 \cdot N_{20}}{N_1 + N_5 + N_{10} + N_{20}} \\
&= 1 \cdot \frac{N_1}{N} + 5 \cdot \frac{N_5}{N} + 10 \cdot \frac{N_{10}}{N} + 20 \cdot \frac{N_{20}}{N} \\
&= 1 \cdot \frac{13}{50} + 5 \cdot \frac{13}{50} + 10 \cdot \frac{10}{50} + 20 \cdot \frac{14}{50} \\
&= 9.16
\end{aligned}$$

since $N = N_1 + N_5 + N_{10} + N_{20} = 50$. If he were to play the game a large number of times, then as $N \to \infty$ we would have $N_k/N \to p_X[k]$, where the latter is just the PMF for choosing a bill with denomination k, and results from the relative frequency interpretation of probability. Then, his average winnings per play would be found as

$$\begin{aligned}
\bar{x} &\to 1 \cdot p_X[1] + 5 \cdot p_X[5] + 10 \cdot p_X[10] + 20 \cdot p_X[20] \\
&= 1 \cdot \frac{1}{4} + 5 \cdot \frac{1}{4} + 10 \cdot \frac{1}{4} + 20 \cdot \frac{1}{4} \\
&= 9
\end{aligned}$$

where $p_X[k] = 1/4$ for $k = 1, 5, 10, 20$ since the proportion of bill denominations in the barrel is the same for each denomination. It is now clear that "on the average" he will lose \$1 per play. The value that the average converges to is called the *expected value* of X, where X is the random variable that describes his winnings for a single play and takes on the values $1, 5, 10, 20$. The expected value is denoted by $E[X]$. For this example, the PMF as well as the expected value is shown in Figure 6.3. The expected value is also called the *expectation* of X, the *average* of X, and the *mean* of X. With this example as motivation we now define the expected value of a discrete random variable X as

$$E[X] = \sum_i x_i p_X[x_i] \tag{6.1}$$

where the sum is over all values of x_i for which $p_X[x_i]$ is nonzero. It is determined from the PMF and as we have seen coincides with our notion of the outcome of an experiment in the "long run" or "on the average." The expected value may also be intepreted as the best prediction of the outcome of a random experiment for a single trial (to be described in Example 6.3). Finally, the expected value is analogous to the center of mass of a system of linearly arranged masses as illustrated in Problem 6.1.

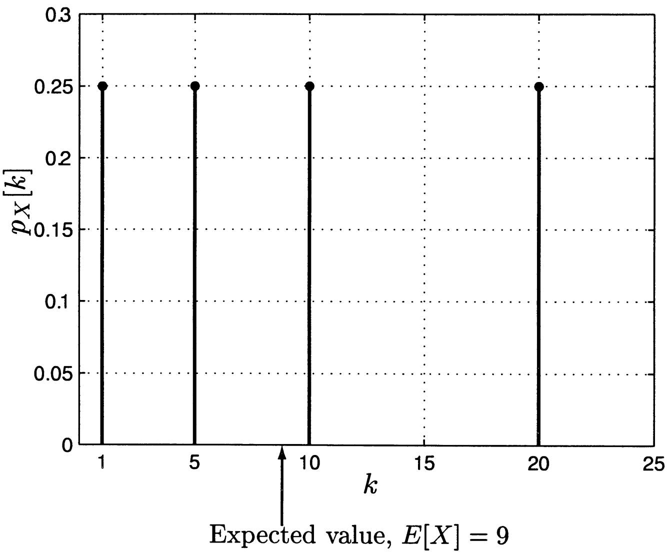


Figure 6.3: PMF and expected value of dollar bill denomination chosen.

6.4 Expected Values of Some Important Random Variables

The definition of the expected value was given by (6.1). When the random variable takes on only integer values, we can rewrite it as

$$E[X] = \sum_{k=-\infty}^{\infty} k p_X[k]. \tag{6.2}$$

We next determine the expected values for some important discrete random variables (see Chapter 5 for a definition of the PMFs).

6.4.1 Bernoulli

If $X \sim \text{Ber}(p)$, then the expected value is

$$\begin{aligned} E[X] &= \sum_{k=0}^{1} k p_X[k] \\ &= 0 \cdot (1-p) + 1 \cdot p \\ &= p. \end{aligned}$$

Note that $E[X]$ need not be a value that the random variable takes on. In this case, it is between $X = 0$ and $X = 1$.

6.4.2 Binomial

If $X \sim \text{bin}(M, p)$, then the expected value is

$$\begin{aligned} E[X] &= \sum_{k=0}^{M} k p_X[k] \\ &= \sum_{k=0}^{M} k \binom{M}{k} p^k (1-p)^{M-k}. \end{aligned}$$

To evaluate this in closed form we will need to find an expression for the sum. Continuing, we have that

$$\begin{aligned} E[X] &= \sum_{k=0}^{M} k \frac{M!}{(M-k)!k!} p^k (1-p)^{M-k} \\ &= Mp \sum_{k=1}^{M} \frac{(M-1)!}{(M-k)!(k-1)!} p^{k-1} (1-p)^{M-1-(k-1)} \end{aligned}$$

and letting $M' = M - 1$, $k' = k - 1$, this becomes

$$\begin{aligned} E[X] &= Mp \sum_{k'=0}^{M'} \frac{M'!}{(M'-k'!)k'!} p^{k'} (1-p)^{M'-k'} \\ &= Mp \sum_{k'=0}^{M'} \binom{M'}{k'} p^{k'} (1-p)^{M'-k'} \\ &= Mp \end{aligned}$$

since the summand is just the PMF of a $\text{bin}(M', p)$ random variable. Therefore, we have that $E[X] = Mp$ for a binomial random variable. This derivation is typical in that we attempt to manipulate the sum into one whose summands are the values of a PMF and so the sum must evaluate to one. Intuitively, we expect that if p is the probability of success for a Bernoulli trial, then the expected number of successes for M independent Bernoulli trials (which is binomially distributed) is Mp.

6.4.3 Geometric

If $X \sim \text{geom}(p)$, the the expected value is

$$E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1} p.$$

To evaluate this in closed form, we need to modify the summand to be a PMF, which in this case will produce a geometric series. To do so we use differentiation by first letting $q = 1 - p$ to produce

$$
\begin{aligned}
E[X] &= p \sum_{k=1}^{\infty} \frac{d}{dq} q^k \\
&= p \frac{d}{dq} \sum_{k=1}^{\infty} q^k.
\end{aligned}
$$

But since $0 < q < 1$ we have upon using the formula for the sum of a geometric series or $\sum_{k=1}^{\infty} q^k = q/(1-q)$ that

$$
\begin{aligned}
E[X] &= p \frac{d}{dq} \left(\frac{q}{1-q} \right) \\
&= p \frac{(1-q) - q(-1)}{(1-q)^2} \\
&= p \frac{1}{(1-q)^2} \\
&= \frac{1}{p}.
\end{aligned}
$$

The expected number of Bernoulli trials until the first success (which is geometrically distributed) is $E[X] = 1/p$. For example, if $p = 1/10$, then on the average it takes 10 trials for a success, an intuitively pleasing result.

6.4.4 Poisson

If $X \sim \text{Pois}(\lambda)$, then it can be shown that $E[X] = \lambda$. The reader is asked to verify this in Problem 6.5. Note that this result is consistent with the Poisson approximation to the binomial PMF since the approximation constrains Mp (the expected value of the binomial random variable) to be λ (the expected value of the Poisson random variable).

Not all PMFs have expected values.

Discrete random variables with a finite number of values always have expected values. In the case of a countably infinite number of values, a discrete random variable *may not have an expected value.* As an example of this, consider the PMF

$$
p_X[k] = \frac{4/\pi^2}{k^2} \qquad k = 1, 2, \ldots. \tag{6.3}
$$

This is a valid PMF since it can be shown to sum to one. Attempting to find the expected value produces

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} k p_X[k] \\ &= \frac{4}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} \to \infty \end{aligned}$$

since $1/k$ is a *harmonic* series which is known not to be summable (meaning that the partial sums do not converge). Hence, the random variable described by the PMF of (6.3) does not have a finite expected value. It is even possible for a sum $\sum_{k=-\infty}^{\infty} k p_X[k]$ that is composed of positive and negative terms to produce different results depending upon the order in which the terms are added together. In this case the value of the sum is said to be ambiguous. These difficulties can be avoided, however, if we require the sum to be absolutely summable or if the sum of the absolute values of the terms is finite [Gaughan 1975]. Hence we will say that the expected value *exists* if

$$E[|X|] = \sum_{k=\infty}^{\infty} |k| p_x[k] < \infty.$$

In Problem 6.6 a further discussion of this point is given.

Lastly, note the following properties of the expected value.

1. It is located at the "center" of the PMF if the PMF is symmetric about some point (see Problem 6.7).

2. It does not generally indicate the most probable value of the random variable (see Problem 6.8).

3. More than one PMF may have the same expected value (see Problem 6.9).

6.5 Expected Value for a Function of a Random Variable

The expected value may easily be found for a *function* of a random variable X if the PMF $p_X[x_i]$ is known. If the function of interest is $Y = g(X)$, then by the definition of expected value

$$E[Y] = \sum_i y_i p_Y[y_i]. \tag{6.4}$$

But as shown in Appendix 6A we can avoid having to find the PMF for Y by using the much more convenient form

$$E[g(X)] = \sum_i g(x_i) p_X[x_i]. \tag{6.5}$$

Otherwise, we would be forced to determine $p_Y[y_i]$ from $p_X[x_i]$ and $g(X)$ using (5.9). This result proves to be very useful, especially when the function is a complicated one such as $g(x) = \sin[(\pi/2)x]$ (see Problem 6.10). Some examples follow.

Example 6.1 – A linear function

If $g(X) = aX + b$, where a and b are constants, then

$$\begin{aligned} E[g(X)] &= E[aX+b] \\ &= \sum_i (ax_i + b)p_X[x_i] \qquad \text{(from (6.5))} \\ &= a\sum_i x_i p_X[x_i] + b\sum_i p_X[x_i] \\ &= aE[X] + b \qquad \text{(definition of } E[X] \text{ and PMF values sum to one.)} \end{aligned}$$

In particular, if we set $a = 1$, then $E[X + b] = E[X] + b$. This allows us to set the expected value of a random variable to any desired value by adding the appropriate constant to X. Finally, a simple extension of this example produces

$$E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[g_2(X)]$$

for any two constants a_1 and a_2 and any two functions g_1 and g_2 (see Problem 6.11). It is said that the *expectation operator E is linear*.

◇

Example 6.2 – A nonlinear function

Assume that X has a PMF given by

$$p_X[k] = \frac{1}{5} \qquad k = 0, 1, 2, 3, 4$$

and determine $E[Y]$ for $Y = g(X) = X^2$. Then, using (6.5) produces

$$\begin{aligned} E[X^2] &= \sum_{k=0}^{4} k^2 p_X[k] \\ &= \sum_{k=0}^{4} k^2 \frac{1}{5} \\ &= 6. \end{aligned}$$

◇

It is not true that $E[g(X)] = g(E[X])$.

From the previous example with $g(X) = X^2$, we had that $E[g(X)] = E[X^2] = 6$ but $g(E[X]) = (E[X])^2 = 2^2 = 4 \neq E[g(X)]$. It is said that the expectation operator *does not commute* (or we cannot just take $E[g(X)]$ and interchange the E and g) for nonlinear functions. This manipulation is valid, however, for *linear* (actually affine) functions as Example 6.1 demonstrates. Henceforth, we will use the notation $E^2[X]$ to replace the more cumbersome $(E[X])^2$.

Example 6.3 – Predicting the outcome of an experiment

It is always of great interest to be able to predict the outcome of an experiment before it has occurred. For example, if the experiment were the summer rainfall in Rhode Island in the coming year, then a farmer would like to have this information before he decides upon which crops to plant. One way to do this is to check the Farmer's almanac, but its accuracy may be in dispute! Another approach would be to *guess* this number based on the PMF (statisticians, however, use the more formal term "predict" or "estimate" which sounds better). Denoting the prediction by the number b, we would like to choose a number so that *on the average* it is close to the true outcome of the random variable X. To measure the error we could use $x - b$, where x is the outcome, and to account for positive and negative errors equally we could use $(x - b)^2$. This *squared error* may at times be small and at other times large, depending on the outcome of X. What we want is the *average value* of the squared error. This is measured by $E[(X - b)^2]$, and is termed the *mean square error* (MSE). We denote it by $\text{mse}(b)$ since it will depend on our choice of b. A reasonable method for choosing b is to choose the value that *minimizes* the MSE. We now proceed to find that value of b.

$$\begin{aligned}
\text{mse}(b) &= E[(X - b)^2] \\
&= E[X^2 - 2bX + b^2] \\
&= E[X^2] - 2bE[X] + E[b^2] && \text{(linearity of } E(\cdot)) \\
&= E[X^2] - 2bE[X] + b^2 && \text{(expected value of constant is the constant).}
\end{aligned}$$

To find the value of b that minimizes the MSE we need only differentiate the MSE, set the derivative equal to zero, and solve for b. This is because the MSE is a quadratic function of b whose minimum is located at the stationary point. Thus, we have

$$\frac{d\text{mse}(b)}{db} = -2E[X] + 2b = 0$$

which produces the minimizing or optimal value of b given by $b_{\text{opt}} = E[X]$. Hence, the best predictor of the outcome of an experiment is the expected value or mean of the random variable. For example, the best predictor of the outcome of a die toss would be 3.5. This result provides another interpretation of the expected value. *The expected value of a random variable is the best predictor of the outcome of the experiment*, where "best" is to be interpreted as the value that minimizes the MSE.

◊

6.6 Variance and Moments of a Random Variable

Another function of a random variable that yields important information about its behavior is that given by $g(X) = (X - E[X])^2$. Whereas $E[X]$ measures the mean of a random variable, $E[(X - E[X])^2]$ measures the *average squared deviation from the mean.* For example, a uniform discrete random variable whose PMF is

$$p_X[k] = \frac{1}{2M+1} \qquad k = -M, -M+1, \ldots, M$$

is easily shown to have a mean of zero for any M. However, as seen in Figure 6.4 the variability of the outcomes of the random variable becomes larger as M increases. This is because the PMF for $M = 10$ can have values exceeding those for $M = 2$. The variability is measured by the *variance* which is defined as

$$\mathrm{var}(X) = E[(X - E[X])^2]. \tag{6.6}$$

Note that the variance is *always greater than or equal to zero.* It is determined from the PMF using (6.5) with $g(X) = (X - E[X])^2$ to yield

$$\mathrm{var}(X) = \sum_i (x_i - E[X])^2 p_X[x_i]. \tag{6.7}$$

For the current example, $E[X] = 0$ due to the symmetry of the PMF about $k = 0$ so that

$$\begin{aligned} \mathrm{var}(X) &= \sum_i x_i^2 p_X[x_i] \\ &= \sum_{k=-M}^{M} k^2 \frac{1}{2M+1} \\ &= \frac{2}{2M+1} \sum_{k=1}^{M} k^2. \end{aligned}$$

But it can be shown that

$$\sum_{k=1}^{M} k^2 = \frac{M(M+1)(2M+1)}{6}$$

which yields

$$\begin{aligned} \mathrm{var}(X) &= \frac{2}{2M+1} \frac{M(M+1)(2M+1)}{6} \\ &= \frac{M(M+1)}{3}. \end{aligned}$$

Clearly, the variance increases with M, or equivalently with the *width* of the PMF, as is also evident from Figure 6.4. We next give another example of the determination of the variance and then summarize the results for several important PMFs.

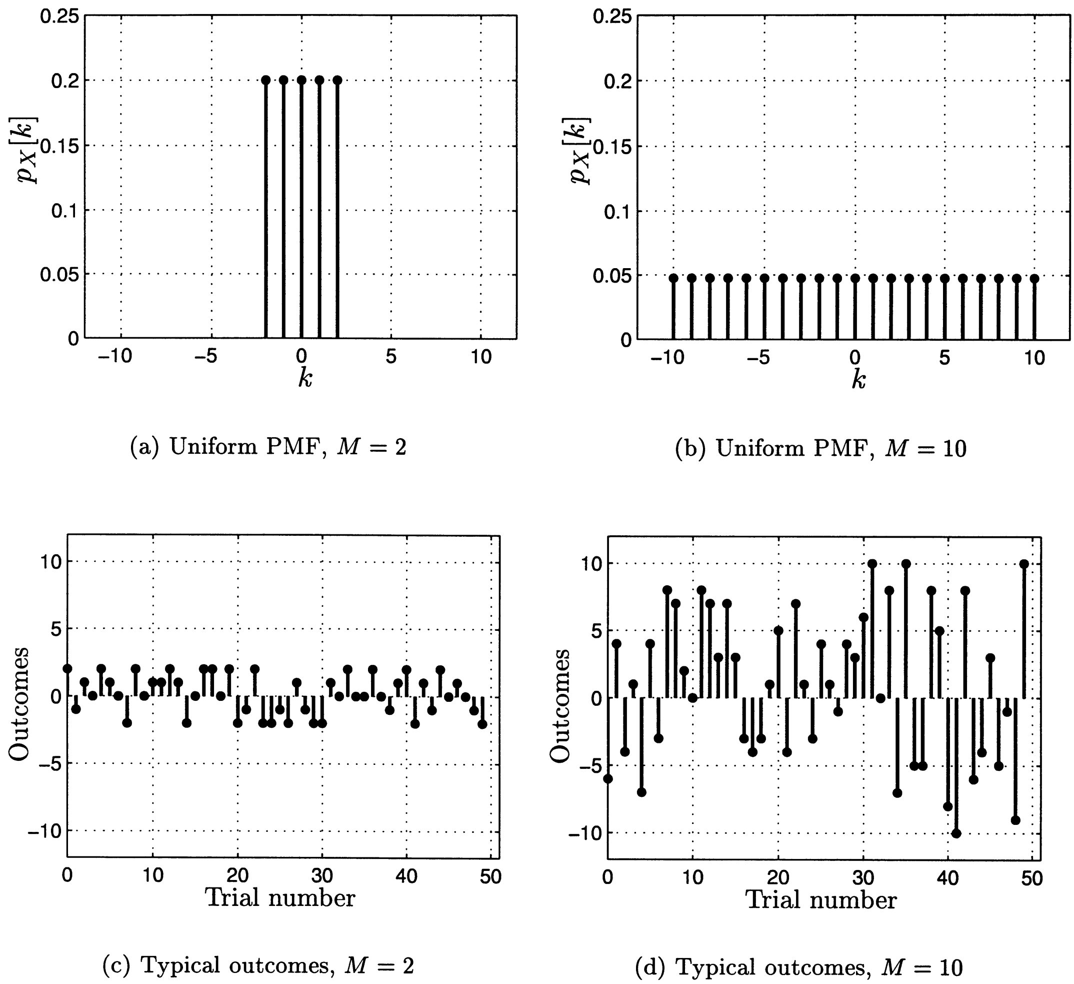


Figure 6.4: Illustration of effect of width of PMF on variability of outcomes.

Example 6.4 – Variance of Bernoulli random variable

If $X \sim \text{Ber}(p)$, then since $E[X] = p$, we have

$$\begin{aligned}
\text{var}(X) &= \sum_i (x_i - E[X])^2 p_X[x_i] \\
&= \sum_{k=0}^{1} (k - p)^2 p_X[k] \\
&= (0 - p)^2(1 - p) + (1 - p)^2 p \\
&= p(1 - p).
\end{aligned}$$

◇

	Values	PMF	$E[X]$	$\text{var}(X)$	$\phi_X(\omega)$
Uniform	$k=-M,\ldots,M$	$\frac{1}{2M+1}$	0	$\frac{M(M+1)}{3}$	$\frac{\sin[(2M+1)\omega/2]}{(2M+1)\sin[\omega/2]}$
Bernoulli	$k=0,1$	$p^k(1-p)^{1-k}$	p	$p(1-p)$	$p\exp(j\omega)+(1-p)$
Binomial	$k=0,1,\ldots,M$	$\binom{M}{k}p^k(1-p)^{M-k}$	Mp	$Mp(1-p)$	$[p\exp(j\omega)+(1-p)]^M$
Geometric	$k=1,2,\ldots$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{\exp(-j\omega)-(1-p)}$
Poisson	$k=0,1,\ldots$	$\exp(-\lambda)\frac{\lambda^k}{k!}$	λ	λ	$\exp[\lambda(\exp(j\omega)-1)]$

Table 6.1: Properties of discrete random variables.

It is interesting to note that the variance is minimized and equals zero if $p = 0$ or $p = 1$. Also, it is maximized for $p = 1/2$. Can you explain this? Important PMFs with their means, variances, and characteristic functions (to be discussed in Section 6.7) are listed in Table 6.1. The reader is asked to derive some of these entries in the Problems.

An alternative useful expression for the variance can be developed based on the properties of the expectation operator. We have that

$$\begin{aligned}\text{var}(X) &= E[(X-E[X])^2] \\ &= E[X^2 - 2XE[X] + E^2[X]] \\ &= E[X^2] - 2E[X]E[X] + E^2[X]\end{aligned}$$

where the last step is due to linearity of the expectation operator and the fact that $E[X]$ is a constant. Hence

$$\text{var}(X) = E[X^2] - E^2[X]$$

and is seen to depend on $E[X]$ and $E[X^2]$. In the case where $E[X] = 0$, we have the simple result that $\text{var}(X) = E[X^2]$. This property of the variance along with some others is now summarized.

Property 6.1 – Alternative expression for variance

$$\text{var}(X) = E[X^2] - E^2[X] \tag{6.8}$$

□

Property 6.2 – Variance for random variable modified by a constant
For c a constant

$$\begin{aligned}\operatorname{var}(c) &= 0\\ \operatorname{var}(X+c) &= \operatorname{var}(X)\\ \operatorname{var}(cX) &= c^2\operatorname{var}(X)\end{aligned}$$

□

The reader is asked to verify Property 6.2 in Problem 6.21.

The expectations $E[X]$ and $E[X^2]$ are called the *first* and *second moments* of X, respectively. The term moment has been borrowed from physics, where $E[X]$ is called the center of mass or moment of mass (see also Problem 6.1). In general, the nth moment is defined as $E[X^n]$ and exists (meaning that the value can be determined unambiguously and is finite) if $E[|X|^n]$ is finite. The latter is called the n absolute moment. It can be shown that if $E[X^s]$ exists, then $E[X^r]$ exists for $r < s$ (see Problem 6.23). As a result, if $E[X^2]$ is finite, then $E[X]$ exists and by (6.8) the variance will also exist. In summary, *the mean and variance of a discrete random variable will exist if the second moment is finite.*

A variant of the notion of moments is that of the *central moments.* They are defined as $E[(X - E[X])^n]$, in which the mean is first subtracted from X before the n moment is computed. They are useful in assessing the average deviations from the mean. In particular, for $n = 2$ we have the usual definition of the variance. See also Problem 6.26 for the relationship between the moments and central moments.

Variance is a nonlinear operator.

The variance of a random variable *does not* have the linearity property of the expectation operator. Hence, in general

$$\operatorname{var}(g_1(X) + g_2(X)) = \operatorname{var}(g_1(X)) + \operatorname{var}(g_2(X)) \qquad \textit{is not true.}$$

Just consider $\operatorname{var}(X + X)$, where $E[X] = 0$ as a simple example.

As explained previously, an alternative interpretation of $E[X]$ is as the best predictor of X. Recall that this predictor is the constant $b_{\text{opt}} = E[X]$ when the mean square error is used as a measure of error. We wish to point out that the *minimum* mse is then

$$\begin{aligned}\text{mse}_{\min} &= E[(X - b_{\text{opt}})^2]\\ &= E[(X - E[X])^2]\\ &= \operatorname{var}(X). \qquad (6.9)\end{aligned}$$

Thus, how well we can predict the outcome of an experiment depends on the variance of the random variable. As an example, consider a coin toss with a probability of heads ($X = 1$) of p and of tails ($X = 0$) of $1 - p$, i.e., a Bernoulli random variable. We would predict the outcome of X to be $b_{\text{opt}} = E[X] = p$ and the minimum mse is the variance which from Example 6.4 is $\text{mse}_{\min} = p(1-p)$. This is plotted in Figure 6.5 versus p. It is seen that the minimum mse is smallest when $p = 0$ or $p = 1$ and largest when $p = 1/2$, or most predictable for $p = 0$ and $p = 1$ and least predictable for $p = 1/2$. Can you explain this?

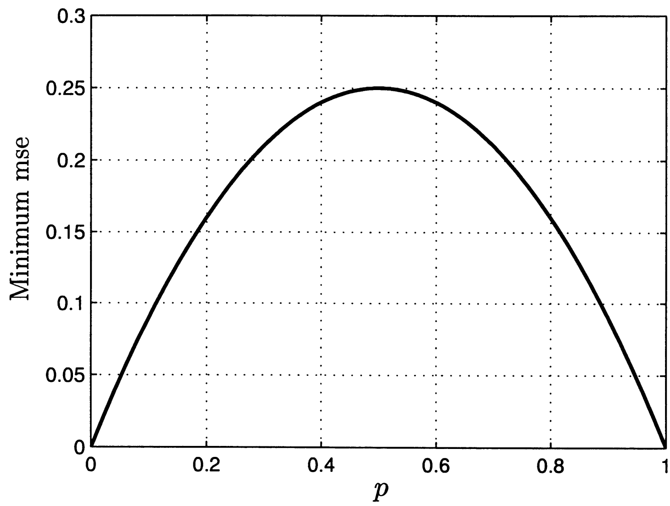


Figure 6.5: Measure of predictability of the outcome of a coin toss.

6.7 Characteristic Functions

Determining the moments $E[X^n]$ of a random variable can be a difficult task for some PMFs. An alternative method that can be considerably easier is based on the *characteristic function*. In addition, the characteristic function can be used to examine convergence of PMFs, as, for example, in the convergence of the binomial PMF to the Poisson PMF, and to determine the PMF for a sum of independent random variables, which will be examined in Chapter 7. In this section we discuss the use of the characteristic function for the calculation of moments and to investigate the convergence of a PMF.

The *characteristic function* of a random variable X is defined as

$$\phi_X(\omega) = E[\exp(j\omega X)] \tag{6.10}$$

where j is the square root of -1 and where ω takes on a suitable range of values. Note that the function $g(X) = \exp(j\omega X)$ is complex but by defining $E[g(X)] =$

$E[\cos(\omega X)+j\sin(\omega X)] = E[\cos(\omega X)]+jE[\sin(\omega X)]$, we can apply (6.5) to the real and imaginary parts of $\phi_X(\omega)$ to yield

$$\begin{aligned}
\phi_X(\omega) &= E[\exp(j\omega X)] \\
&= E[\cos(\omega X) + j\sin(\omega X)] \\
&= E[\cos(\omega X)] + jE[\sin(\omega X)] \\
&= \sum_i \cos(\omega x_i)p_X[x_i] + j\sum_i \sin(\omega x_i)p_X[x_i] \\
&= \sum_i \exp(j\omega x_i)p_X[x_i]. \qquad (6.11)
\end{aligned}$$

To simplify the discussion, yet still be able to apply our results to the important PMFs, we assume that the sample space $\mathcal{S}_X$ is a subset of the integers. Then (6.11) becomes

$$\phi_X(\omega) = \sum_{k=-\infty}^{\infty} \exp(j\omega k)p_X[k]$$

or rearranging

$$\phi_X(\omega) = \sum_{k=-\infty}^{\infty} p_X[k]\exp(j\omega k) \qquad (6.12)$$

where $p_X[k] = 0$ for those integers not included in $\mathcal{S}_X$. For example, in the Poisson PMF the range of summation in (6.12) would be $k \geq 0$. In this form, the characteristic function is immediately recognized as being the Fourier transform of the *sequence* $p_X[k]$ for $-\infty < k < \infty$. Its definition is slightly different than the usual Fourier transform, called the discrete-time Fourier transform, which uses the function $\exp(-j\omega k)$ in its definition [Jackson 1991]. As a Fourier transform, it exhibits all the usual properties. In particular, the *Fourier transform* of a sequence is periodic with period of 2π (see Property 6.4 for a proof). As a result, we need only examine the characteristic function over the interval $-\pi \leq \omega \leq \pi$, which is defined to be the fundamental period. For our purposes the most useful property is that we can differentiate the sum in (6.12) "term by term" or

$$\begin{aligned}
\frac{d\phi_X(\omega)}{d\omega} &= \frac{d}{d\omega}\sum_{k=-\infty}^{\infty} p_X[k]\exp(j\omega k) \\
&= \sum_{k=-\infty}^{\infty} p_X[k]\frac{d}{d\omega}\exp(j\omega k).
\end{aligned}$$

The utility in doing so is to produce a formula for $E[X]$. Carrying out the differentiation

$$\frac{d\phi_X(\omega)}{d\omega} = \sum_{k=-\infty}^{\infty} p_X[k]jk\exp(j\omega k)$$

so that

$$\begin{aligned}\left.\frac{1}{j}\frac{d\phi_X(\omega)}{d\omega}\right|_{\omega=0} &= \sum_{k=-\infty}^{\infty} kp_X[k] \\ &= E[X].\end{aligned}$$

In fact, repeated differentiation produces the formula for the nth moment as

$$E[X^n] = \left.\frac{1}{j^n}\frac{d^n\phi_X(\omega)}{d\omega^n}\right|_{\omega=0}. \tag{6.13}$$

All the moments that exist may be found by repeated differentiation of the characteristic function. An example follows.

Example 6.5 – First two moments of geometric random variable
Since the PMF for a geometric random variable is given by $p_X[k] = (1-p)^{k-1}p$ for $k = 1, 2, \ldots$, we have that

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=1}^{\infty} p_X[k]\exp(j\omega k) \\ &= \sum_{k=1}^{\infty}(1-p)^{k-1}p\exp(j\omega k) \\ &= p\exp(j\omega)\sum_{k=1}^{\infty}[(1-p)\exp(j\omega)]^{k-1}.\end{aligned}$$

But since $|(1-p)\exp(j\omega)| < 1$, we can use the result

$$\sum_{k=1}^{\infty} z^{k-1} = \sum_{k=0}^{\infty} z^k = \frac{1}{1-z}$$

for z a complex number with $|z| < 1$ to yield the characteristic function

$$\begin{aligned}\phi_X(\omega) &= \frac{p\exp(j\omega)}{1-[(1-p)\exp(j\omega)]} \\ &= \frac{p}{\exp(-j\omega)-(1-p)}.\end{aligned} \tag{6.14}$$

Note that as claimed the characteristic function is periodic with period 2π. To find the mean we use (6.13) with $n = 1$ to produce

$$\begin{aligned}E[X] &= \left.\frac{1}{j}\frac{d\phi_X(\omega)}{d\omega}\right|_{\omega=0} \\ &= \frac{1}{j}p(-1)\left.\frac{-j\exp(-j\omega)}{[\exp(-j\omega)-(1-p)]^2}\right|_{\omega=0} && (6.15) \\ &= \frac{1}{j}p\frac{j}{p^2} = \frac{1}{p} && (6.16)\end{aligned}$$

which agrees with our earlier results based on using the definition of expected value. To find the second moment and hence the variance using (6.8)

$$\begin{aligned}
E[X^2] &= \frac{1}{j^2}\left.\frac{d^2\phi_X(\omega)}{d\omega^2}\right|_{\omega=0} \\
&= \frac{p}{j}\frac{d}{d\omega}\left.\frac{\exp(-j\omega)}{[\exp(-j\omega)-(1-p)]^2}\right|_{\omega=0} \qquad \text{(from (6.15))} \\
&= \frac{p}{j}\left.\frac{D^2(-j)\exp(-j\omega)-\exp(-j\omega)2D(-j)\exp(-j\omega)}{D^4}\right|_{\omega=0}
\end{aligned}$$

where $D = \exp(-j\omega) - (1-p)$. Since $D|_{\omega=0} = p$, we have that

$$\begin{aligned}
E[X^2] &= \left(\frac{p}{j}\right)\left(\frac{-jp^2+2jp}{p^4}\right) \\
&= \frac{2p-p^2}{p^3} \\
&= \frac{2}{p^2}-\frac{1}{p}
\end{aligned}$$

so that finally we have

$$\begin{aligned}
\operatorname{var}(X) &= E[X^2]-E^2[X] \\
&= \frac{2}{p^2}-\frac{1}{p}-\frac{1}{p^2} \\
&= \frac{1-p}{p^2}.
\end{aligned}$$

◇

As a second example, we consider the binomial PMF.

Example 6.6 – Expected value of binomial PMF

We first determine the characteristic function as

$$\begin{aligned}
\phi_X(\omega) &= \sum_{k=-\infty}^{\infty} p_X[k]\exp(j\omega k) \\
&= \sum_{k=0}^{M}\binom{M}{k}p^k(1-p)^{M-k}\exp(j\omega k) \\
&= \sum_{k=0}^{M}\binom{M}{k}\Big[\underbrace{p\exp(j\omega)}_{a}\Big]^k\Big[\underbrace{1-p}_{b}\Big]^{M-k} \qquad (6.17) \\
&= (a+b)^M \qquad \text{(binomial theorem)} \\
&= [p\exp(j\omega)+(1-p)]^M. \qquad (6.18)
\end{aligned}$$

The expected value then follows as

$$\begin{aligned}
E[X] &= \frac{1}{j}\left.\frac{d\phi_X(\omega)}{d\omega}\right|_{\omega=0} \\
&= \frac{1}{j}M\left.\left[p\exp(j\omega)+(1-p)\right]^{M-1}pj\exp(j\omega)\right|_{\omega=0} \\
&= Mp
\end{aligned}$$

which is in agreement with our earlier results. The variance can be found by using (6.8) and (6.13) for $n = 2$. It is left as an exercise to the reader to show that (see Problem 6.29)

$$\operatorname{var}(X) = Mp(1-p).$$

◇

The characteristic function for the other important PMFs are given in Table 6.1. Some important properties of the characteristic function are listed next.

Property 6.3 – Characteristic function always exists since $|\phi_X(\omega)| < \infty$

Proof:

$$\begin{aligned}
|\phi_X(\omega)| &= \left|\sum_{k=-\infty}^{\infty} p_X[k]\exp(j\omega k)\right| \\
&\le \sum_{k=-\infty}^{\infty} |p_X[k]\exp(j\omega k)| && \text{(magnitude of sum of complex numbers cannot exceed sum of magnitudes)} \\
&= \sum_{k=-\infty}^{\infty} |p_X[k]| && (|\exp(j\omega k)| = 1) \\
&= \sum_{k=-\infty}^{\infty} p_X[k] && (p_X[k] \ge 0) \\
&= 1.
\end{aligned}$$

□

Property 6.4 – Characteristic function is periodic with period 2π.

Proof: For m an integer

$$\begin{aligned}
\phi_X(\omega + 2\pi m) &= \sum_{k=-\infty}^{\infty} p_X[k]\exp[j(\omega+2\pi m)k] \\
&= \sum_{k=-\infty}^{\infty} p_X[k]\exp[j\omega k]\exp[j2\pi mk] \\
&= \sum_{k=-\infty}^{\infty} p_X[k]\exp[j\omega k] && \text{(since } \exp(j2\pi mk) = 1 \text{ for } mk \text{ an integer)} \\
&= \phi_X(\omega).
\end{aligned}$$

□

Property 6.5 – The PMF may be recovered from the characteristic function.

Given the characteristic function, we may determine the PMF using

$$p_X[k] = \int_{-\pi}^{\pi} \phi_X(\omega) \exp(-j\omega k) \frac{d\omega}{2\pi} \qquad -\infty < k < \infty. \tag{6.19}$$

<u>Proof</u>: Since the characteristic function is the Fourier transform of a sequence (although its definition uses a $+j$ instead of the usual $-j$), it has an inverse Fourier transform. Although any interval of length 2π may be used to perform the integration in the inverse Fourier transform, it is customary to use $[-\pi, \pi]$ which results in (6.19).

□

Property 6.6 – Convergence of characteristic functions guarantees convergence of PMFs.

This property says that if we have a sequence of characteristic functions, say $\phi_X^{(n)}(\omega)$, which converges to a given characteristic function, say $\phi_X(\omega)$, then the corresponding sequence of PMFs, say $p_X^{(n)}[k]$, must converge to a given PMF say $p_X[k]$, where $p_X[k]$ is given by (6.19). The importance of this theorem is that it allows us to approximate PMFs by simpler ones if we can show that the characteristic functions are approximately equal. An illustration is given next. This theorem is known as the *continuity theorem of probability*. Its proof is beyond the scope of this text but can be found in [Pollard 2002].

□

We recall the approximation of the binomial PMF by the Poisson PMF under the conditions that $p \to 0$ and $M \to \infty$ with $Mp = \lambda$ fixed (see Section 5.6). To show this using the characteristic function approach (based on Property 6.6) we let X_b denote a binomial random variable. Its characteristic function is from (6.18)

$$\phi_{X_b}(\omega) = [p \exp(j\omega) + (1-p)]^M$$

and replacing p by λ/M we have

$$\begin{aligned}
\phi_{X_b}(\omega) &= \left[\frac{\lambda}{M}\exp(j\omega) + \left(1 - \frac{\lambda}{M}\right)\right]^M \\
&= \left[1 + \frac{\lambda(\exp(j\omega)-1)}{M}\right]^M \\
&\to \exp[\lambda(\exp(j\omega)-1)] \qquad \text{(see Problem 5.15, results are also valid for a complex variable)}
\end{aligned}$$

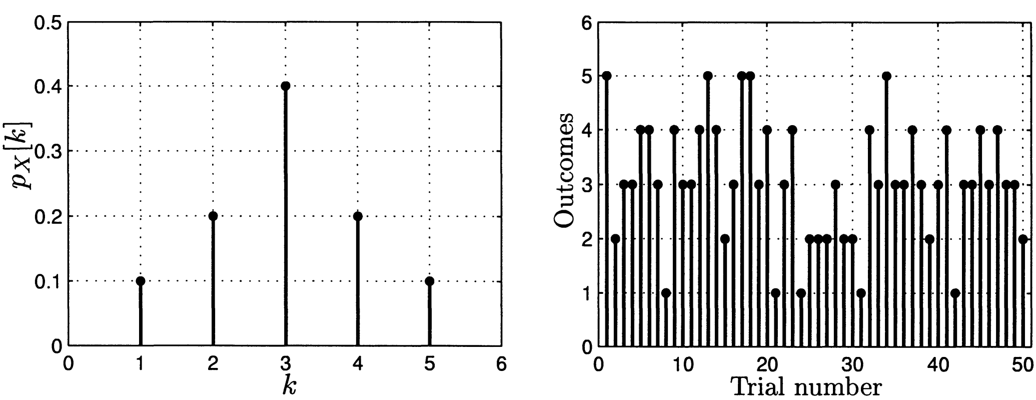
as $M \to \infty$. For a Poisson random variable X_P we have that

$$\begin{aligned}
\phi_{X_P}(\omega) &= \sum_{k=0}^{\infty} \exp(-\lambda)\frac{\lambda^k}{k!}\exp(j\omega k) \\
&= \exp(-\lambda)\sum_{k=0}^{\infty}\frac{[\lambda \exp(j\omega)]^k}{k!} \\
&= \exp(-\lambda)\exp[\lambda\exp(j\omega)] \quad \text{(using results from Problem 5.22 which also hold for a complex variable)} \\
&= \exp[\lambda(\exp(j\omega)-1)].
\end{aligned}$$

Since $\phi_{X_b}(\omega) \to \phi_{X_P}(\omega)$ as $M \to \infty$, by Property 6.6, we must have that $p_{X_b}[k] \to p_{X_P}[k]$ for all k. Hence, under the stated conditions the binomial PMF becomes the Poisson PMF as $M \to \infty$. This was previously proven by other means in Section 5.6. Our derivation here though is considerably simpler.

6.8 Estimating Means and Variances

As alluded to earlier, an important aspect of the mean and variance of a PMF is that they are easily estimated in practice. We have already briefly discussed this in Chapter 2 where it was demonstrated how to do this with computer simulated data (see Example 2.3). We now continue that discussion in more detail. To illustrate the approach we will consider the PMF shown in Figure 6.6a. Since the theoretical



(a) PMF (b) Simulated data

Figure 6.6: PMF and computer generated data used to illustrate estimation of mean and variance.

expected value or mean is given by

$$E[X] = \sum_{k=1}^{5} k p_X[k]$$

then by the relative frequency interpretation of probability we can use the approximation

$$p_X[k] \approx \frac{N_k}{N}$$

where N_k is the number of trials in which a k was the outcome and N is the total number of trials. As a result, we can estimate the mean by

$$\widehat{E[X]} = \sum_{k=1}^{5} \frac{kN_k}{N}.$$

The "hat" will always denote an estimated quantity. But kN_k is just the sum of all the k outcomes that appear in the N trials and therefore $\sum_{k=1}^{5} kN_k$ is the sum of *all* the outcomes in the N trials. Denoting the latter by $\sum_{i=1}^{N} x_i$, we have as our estimate of the mean

$$\widehat{E[X]} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{6.20}$$

where x_i is the outcome of the ith trial. Note that we have just reversed our line of reasoning used in the introduction to motivate the use of $E[X]$ as the definition of the expected value of a random variable. Also, we have previously seen this type of estimate in Example 2.3 where it was referred to as the *sample mean*. It is usually denoted by $\bar{x}$. For the data shown in Figure 6.6b we plot the sample mean in Figure 6.7a versus N. Note that as N becomes larger, we have that $\widehat{E[X]} \to 3 = E[X]$.

The true variance of the PMF shown in Figure 6.6a is computed as

$$\begin{aligned} \text{var}(X) &= E[X^2] - E^2[X] \\ &= \sum_{k=1}^{5} k^2 p_X[k] - E^2[X] \end{aligned}$$

which is easily shown to be $\text{var}(X) = 1.2$. It is estimated as

$$\widehat{\text{var}(X)} = \widehat{E[X^2]} - (\widehat{E[X]})^2$$

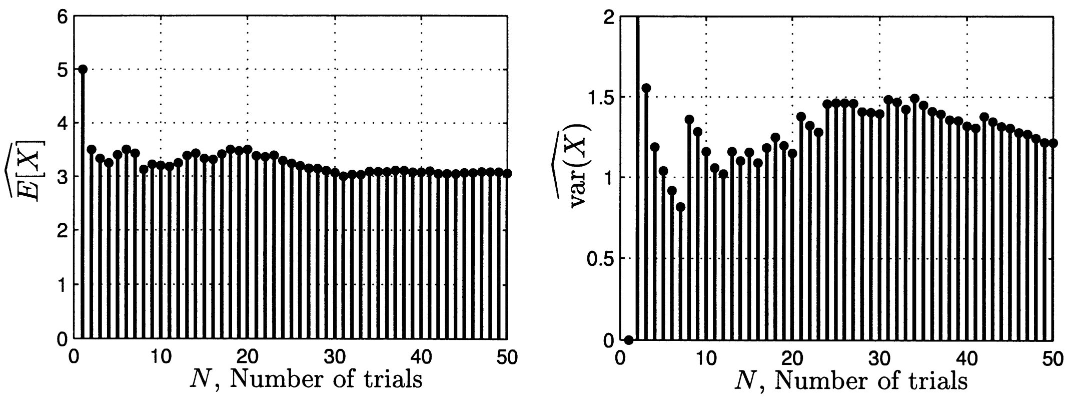
and by the same rationale as before we use

$$\widehat{E[X^2]} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

so that our estimate of the variance becomes

$$\widehat{\text{var}(X)} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2. \tag{6.21}$$

This estimate is shown in Figure 6.7b as a function of N. Note that as the number of



(a) Estimated mean

(b) Estimated variance

Figure 6.7: Estimated mean and variance for computer data shown in Figure 6.6.

trials increases the estimate of variance converges to the true value of $\text{var}(X) = 1.2$. The MATLAB code used to generate the data and estimate the mean and variance is given in Appendix 6B. Also, in that appendix is listed the MATLAB subprogram `PMFdata.m` which allows easier generation of the outcomes of a discrete random variable. In practice, it is customary to use (6.20) and (6.21) to analyze real-world data as a first step in assessing the characteristics of an unknown PMF.

6.9 Real-World Example – Data Compression

The digital revolution of the past 20 years has made it commonplace to record and store information in a digital format. Such information consists of speech data in telephone transmission, music data stored on compact discs, video data stored on digital video discs, and facsimile data, to name but a few. The amount of data can become quite large so that it is important to be able to reduce the amount of storage required. The process of storage reduction is called *data compression*. We now illustrate how this is done. To do so we simplify the discussion by assuming that the data consists of a sequence of the letters A, B, C, D. One could envision these letters as representing the chords of a rudimentary musical instrument, for example. The extension to the entire English alphabet consisting of 26 letters will be apparent. Consider a typical sequence of 50 letters

$$\begin{array}{c}\text{AAAAAAAAAAABAAAAAAAAAAAAA}\\ \text{AAAAAACABADAABAAABAAAAAAD.}\end{array} \quad (6.22)$$

To encode these letters for storage we could use the two-bit code

$$\begin{aligned} \text{A} &\rightarrow 00 \\ \text{B} &\rightarrow 01 \\ \text{C} &\rightarrow 10 \\ \text{D} &\rightarrow 11 \end{aligned} \quad (6.23)$$

which would then require a storage of 2 bits per letter for a total storage of 100 bits. However, as seen above the typical sequence is characterized by a much larger probability of observing an "A" as opposed to the other letters. In fact, there are 43 A's, 4 B's, 1 C, and 2 D's. It makes sense then to attempt a reduction in storage by assigning *shorter code words* to the letters that *occur more often*, in this case, to the "A". As a possible strategy, consider the code assignment

$$\begin{aligned} \text{A} &\rightarrow 0 \\ \text{B} &\rightarrow 10 \\ \text{C} &\rightarrow 110 \\ \text{D} &\rightarrow 111. \end{aligned} \quad (6.24)$$

Using this code assignment for our typical sequence would require only $1 \cdot 43 + 2 \cdot 4 + 3 \cdot 1 + 3 \cdot 2 = 60$ bits or 1.2 bits per letter. The code given by (6.24) is called a *Huffman code*. It can be shown to produce less bits per letter "on the average" [Cover, Thomas 1991].

To determine actual storage savings we need to determine the *average length* of the code word per letter. First we define a discrete random variable that measures the length of the code word. For the sample space $\mathcal{S} = \{\text{A}, \text{B}, \text{C}, \text{D}\}$ we define the random variable

$$X(s_i) = \begin{cases} 1 & s_1 = \text{A} \\ 2 & s_2 = \text{B} \\ 3 & s_3 = \text{C} \\ 3 & s_4 = \text{D} \end{cases}$$

which yields the code length for each letter. The probabilities used to generate the sequence of letters shown in (6.22) are $P[\text{A}] = 7/8$, $P[\text{B}] = 1/16$, $P[\text{C}] = 1/32$, $P[\text{D}] = 1/32$. As a result the PMF for X is

$$p_X[k] = \begin{cases} \frac{7}{8} & k = 1 \\ \frac{1}{16} & k = 2 \\ \frac{1}{16} & k = 3. \end{cases}$$

The average code length is given by

$$\begin{aligned} E[X] &= \sum_{k=1}^{3} k p_X[k] \\ &= 1 \cdot \frac{7}{8} + 2 \cdot \frac{1}{16} + 3 \cdot \frac{1}{16} \\ &= 1.1875 \text{ bits per letter.} \end{aligned}$$

This results in a compression ratio of $2 : 1.1875 = 1.68$ or we require about 40% less storage.

It is also of interest to note that the average code word length per letter can be reduced even further. However, it requires more complexity in coding (and of course in decoding). A fundamental theorem due to Shannon, who in many ways laid the groundwork for the digital revolution, says that the average code word length per letter can be no less than [Shannon 1948]

$$H = \sum_{i=1}^{4} P[s_i] \log_2 \frac{1}{P[s_i]} \qquad \text{bits per letter.} \tag{6.25}$$

This quantity is termed the *entropy* of the source. In addition, he showed that a code exists that can attain, to within any small deviation, this minimum average code length. For our example, the entropy is

$$\begin{aligned} H &= \frac{7}{8}\log_2 \frac{1}{7/8} + \frac{1}{16}\log_2 \frac{1}{1/16} + \frac{1}{32}\log_2 \frac{1}{1/32} + \frac{1}{32}\log_2 \frac{1}{1/32} \\ &= 0.7311 \qquad \text{bits per letter.} \end{aligned}$$

Hence, the potential compression ratio is $2 : 0.7311 = 2.73$ for about a 63% reduction.

Clearly, it is seen from this example that the amount of reduction will depend critically upon the probabilities of the letters occuring. If they are all equally likely to occur, then the minimum average code length is from (6.25) with $P[s_i] = 1/4$

$$H = 4\left(\frac{1}{4}\log_2 \frac{1}{1/4}\right) = 2 \qquad \text{bits per letter.}$$

In this case no compression is possible and the original code given by (6.23) will be optimal. The interested reader should consult [Cover and Thomas 1991] for further details.

References

Cover, T.M., J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

Gaughan, E.D., *Introduction to Analysis*, Brooks/Cole, Monterey, CA, 1975.

Jackson, L.B., *Signals, Systems, and Transforms*, Addison-Wesley, Reading, MA, 1991.

Pollard., D. *A User's Guide to Measure Theoretic Probability*, Cambridge University Press, New York, 2002.

Shannon, C.E., "A Mathematical Theory of Communication," *Bell System Tech. Journal*, Vol. 27, pp. 379–423, 623–656, 1948.

Problems

6.1 (w) The center of mass of a system of masses situated on a line is the point at which the system is balanced. That is to say that at this point the sum of the moments, where the moment is the distance from center of mass times the mass, is zero. If the center of mass is denoted by CM, then

$$\sum_{i=1}^{M}(x_i - \mathrm{CM})m_i = 0$$

where x_i is the position of the ith mass along the x direction and m_i is its corresponding mass. First solve for CM. Then, for the system of weights shown in Figure 6.8 determine the center of mass. How is this analogous to the expected value of a discrete random variable?

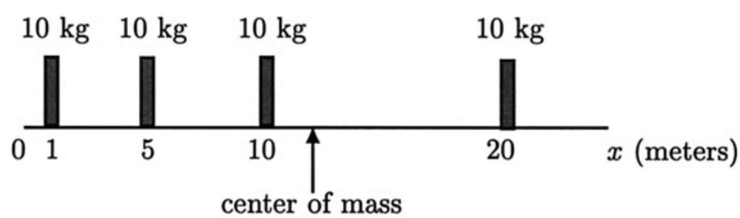


Figure 6.8: Weightless bar supporting four weights.

6.2 (⌣̈) **(f)** For the discrete random variable with PMF

$$p_X[k] = \frac{1}{10} \qquad k = 0, 1, \ldots, 9$$

find the expected value of X.

6.3 (w) A die is tossed. The probability of obtaining a 1, 2, or 3 is the same. Also, the probability of obtaining a 4, 5, or 6 is the same. However, a 5 is twice as likely to be observed as a 1. For a large number of tosses what is the average value observed?

6.4 (⌣) **(f)** A coin is tossed with the probability of heads being 2/3. A head is mapped into $X = 1$ and a tail into $X = 0$. What is the expected outcome of this experiment?

6.5 (f) Determine the expected value of a Poisson random variable. Hint: Differentiate $\sum_{k=0}^{\infty} \lambda^k/k!$ with respect to λ.

6.6 (t) Consider the PMF $p_X[k] = (2/\pi)/k^2$ for $k = \ldots, -1, 0, 1, \ldots$. The expected value is defined as

$$E[X] = \sum_{k=-\infty}^{\infty} k p_X[k]$$

which is actually shorthand for

$$E[X] = \lim_{\substack{N_L \to -\infty \\ N_U \to \infty}} \sum_{k=N_L}^{N_U} k p_X[k]$$

where the L and U represent "lower" and "upper", respectively. This may be written as

$$E[X] = \lim_{N_L \to -\infty} \sum_{k=N_L}^{-1} k p_X[k] + \lim_{N_U \to \infty} \sum_{k=1}^{N_U} k p_X[k]$$

where the limits are taken *independently* of each other. For $E[X]$ to be unambiguous and finite both limits must be finite. As a result, show that the expected value for the given PMF does not exist. If, however, we were to constrain $N_L = N_U$, show that the expected value is zero. Note that if $N_L = N_U$, we are reordering the terms before performing the sum since the partial sums become $\sum_{k=-1}^{1} k p_X[k]$, $\sum_{k=-2}^{2} k p_X[k]$, etc. But for the expected value to be unambiguous, the value should not depend on the ordering. If a sum is *absolutely* summable, any ordering will produce the same result [Gaughan 1975], hence our requirement for the existence of the expected value.

6.7 (t) Assume that a discrete random variable takes on the values $k = \ldots, -1, 0, 1, \ldots$ and that its PMF satisfies $p_X[m + i] = p_X[m - i]$, where m is a fixed integer and $i = 1, 2, \ldots$. This says that the PMF is symmetric about the point $x = m$. Prove that the expected value of the random variable is $E[X] = m$.

6.8 (⌣) **(t)** Give an example where the expected value of a random variable is *not* its most probable value.

6.9 (t) Give an example of two PMFs that have the same expected value.

6.10 (f) A discrete random variable X has the PMF $p_X[k] = 1/5$ for $k = 0, 1, 2, 3, 4$. If $Y = \sin[(\pi/2)X]$, find $E[Y]$ using (6.4) and (6.5). Which way is easier?

6.11 (t) Prove the linearity property of the expectation operator

$$E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[g_2(X)]$$

where a_1 and a_2 are constants.

6.12 (⌣̈) **(f)** Determine $E[X^2]$ for a geom(p) random variable using (6.5). Hint: You will need to differentiate twice.

6.13 (⌣̈) **(t)** Can $E[X^2]$ ever be equal to $E^2[X]$? If so, when?

6.14 (⌣̈) **(w)** A discrete random variable X has the PMF

$$p_X[k] = \begin{cases} \frac{1}{8} & k = 1 \\ \frac{2}{8} & k = 2 \\ \frac{4}{8} & k = 3 \\ \frac{1}{8} & k = 4. \end{cases}$$

If the experiment that produces a value of X is conducted, find the minimum mean square error predictor of the outcome. What is the minimum mean square error of the predictor?

6.15 (⌣̈) **(c)** For Problem 6.14 use a computer to simulate the experiment for many trials. Compare the estimate to the actual outcomes of the computer experiment. Also, compute the minimum mean square error and compare it to the theoretical value obtained in Problem 6.14.

6.16 (w) Of the three PMFs shown in Figure 6.9, which one has the smallest variance? Hint: You do not need to actually calculate the variances.

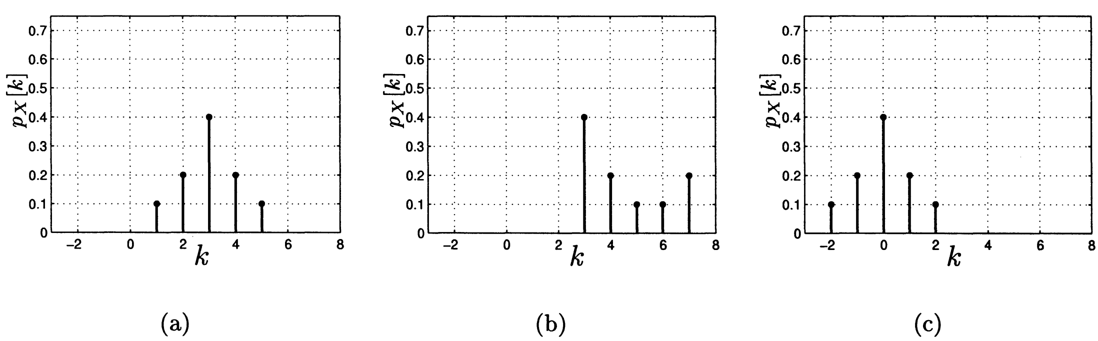


Figure 6.9: PMFs for Problem 6.16.

6.17 (w) If $Y = aX + b$, what is the variance of Y in terms of the variance of X?

6.18 (f) Find the variance of a Poisson random variable. See the hint for Problem 6.12.

6.19 (f) For the PMF given in Problem 6.2 find the variance.

6.20 (⌣̈) (f) Find the second moment for a Poisson random variable by using the characteristic function, which is given in Table 6.1.

6.21 (t) If X is a discrete random variable and c is a constant, prove the following properties of the variance:

$$\begin{aligned} \operatorname{var}(c) &= 0 \\ \operatorname{var}(X+c) &= \operatorname{var}(X) \\ \operatorname{var}(cX) &= c^2\operatorname{var}(X). \end{aligned}$$

6.22 (t) If a discrete random variable X has $\operatorname{var}(X) = 0$, prove that X must be a constant c. This provides a converse to the property that if $X = c$, then $\operatorname{var}(X) = 0$.

6.23 (t) In this problem we prove that if $E[X^s]$ exists, meaning that $E[|X|^s] < \infty$, then $E[X^r]$ also exists for $0 < r < s$. Provide the explanations for the following steps:

a. For $|x| \leq 1$, $|x|^r \leq 1$

b. For $|x| > 1$, $|x|^r \leq |x|^s$

c. For all $|x|$, $|x|^r \leq |x|^s + 1$

d. $E[|X|^r] = \sum_i |x_i|^r p_X[x_i] \leq \sum_i (|x_i|^s + 1) p_X[x_i] = E[|X|^s] + 1 < \infty$.

6.24 (f) If a discrete random variable has the PMF $p_X[k] = 1/4$ for $k = -1$ and $p_X[k] = 3/4$ for $k = 1$, find the mean and variance.

6.25 (t) A symmetric PMF satisfies the relationship $p_X[-k] = p_X[k]$ for $k = \ldots, -1, 0, 1, \ldots$. Prove that all the odd order moments, $E[X^n]$ for n odd, are zero.

6.26 (⌣̈) (t) A central moment of a discrete random variable is defined as $E[(X - E[X])^n]$, for n a positive integer. Derive a formula that relates the central moment to the usual moments. Hint: You will need the binomial formula.

6.27 (⌣̈) (t) If $Y = aX + b$, find the characteristic function of Y in terms of that for X. Next use your result to prove that $E[Y] = aE[X] + b$.

6.28 (⌣̈) (f) Find the characteristic function for the PMF $p_X[k] = 1/5$ for $k = -2, -1, 0, 1, 2$.

6.29 (f) Determine the variance of a binomial random variable by using the properties of the characteristic function. You can assume knowledge of the characteristic function for a binomial random variable.

6.30 (f) Determine the mean and variance of a Poisson random variable by using the properties of the characteristic function. You can assume knowledge of the characteristic function for a Poisson random variable.

6.31 (f) Which PMF $p_X[k]$ for $k = \ldots, -1, 0, 1, \ldots$ has the characteristic function $\phi_X(\omega) = \cos\omega$?

6.32 (⌣̈) (c) For the random variable described in Problem 6.24 perform a computer simulation to estimate its mean and variance. How does it compare to the true mean and variance?

Appendix 6A

Derivation of $E[g(X)]$ Formula

Assume that X is a discrete random variable taking on values in $\mathcal{S}_X = \{x_1, x_2, \ldots\}$ with PMF $p_X[x_i]$. Then, if $Y = g(X)$ we have from the definition of expected value

$$E[Y] = \sum_i y_i p_Y[y_i] \tag{6A.1}$$

where the sum is over all $y_i \in \mathcal{S}_Y$. Note that it is assumed that the y_i are *distinct* (all different). But from (5.9)

$$p_Y[y_i] = \sum_{\{x_j : g(x_j) = y_i\}} p_X[x_j]. \tag{6A.2}$$

To simplify the notation we will define the *indicator function*, which indicates whether a number x is within a given set A, as

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise.} \end{cases}$$

Then (6A.2) can be rewritten as

$$p_Y[y_i] = \sum_{j=1}^{\infty} p_X[x_j] I_{\{0\}}(y_i - g(x_j))$$

since the sum will include the term $p_X[x_j]$ only if $y_i - g(x_j) = 0$. Using this, we have from (6A.1)

$$\begin{aligned} E[Y] &= \sum_i y_i \sum_{j=1}^{\infty} p_X[x_j] I_{\{0\}}(y_i - g(x_j)) \\ &= \sum_{j=1}^{\infty} \left[\sum_i y_i I_{\{0\}}(y_i - g(x_j)) \right] p_X[x_j]. \end{aligned}$$

Now for a given j, $g(x_j)$ is a fixed number and since the y_i's are distinct, there is only one y_i for which $y_i = g(x_j)$. Thus, we have that

$$\sum_i y_i I_{\{0\}}(y_i - g(x_j)) = g(x_j)$$

and finally

$$E[Y] = E[g(X)] = \sum_{j=1}^{\infty} g(x_j) p_X[x_j].$$

Appendix 6B

MATLAB Code Used to Estimate Mean and Variance

Figures 6.6 and 6.7 are based on the following MATLAB code.

```
% PMFdata.m
%
% This program generates the outcomes for N trials
% of an experiment for a discrete random variable.
% Uses the method of Section 5.9.
% It is a function subprogram.
%
% Input parameters:
%
%   N   - number of trials desired
%   xi  - values of x_i's of discrete random variable (M x 1 vector)
%   pX  - PMF of discrete random variable (M x 1 vector)
%
% Output parameters:
%
%   x   - outcomes of N trials (N x 1 vector)
%
function x=PMFdata(N,xi,pX)
M=length(xi);M2=length(pX);
if M~=M2
   message='xi and pX must have the same dimension'
end
for k=1:M ; % see Section 5.9 and Figure 5.14 for approach used here
   if k==1
```

```
    bin(k,1)=pX(k); % set up first interval of CDF as [0,pX(1)]

   else
      bin(k,1)=bin(k-1,1)+pX(k); % set up succeeding intervals
                                 % of CDF
   end
end
u=rand(N,1); % generate N outcomes of uniform random variable
for i=1:N % determine which interval of CDF the outcome lies in
          % and map into value of xi
   if u(i)>0&u(i)<=bin(1)
         x(i,1)=xi(1);
      end
      for k=2:M
         if u(i)>bin(k-1)&u(i)<=bin(k)
            x(i,1)=xi(k);
         end
      end
   end
```