[CLAR 1978] J. CLARY, A. JAI, S. WEIKEL, R. SAEKS, and D. SIEWIOREK, "A Preliminary Study of Built-In-Test for the Military Computer Family," Technical Report, Research Triangle Institute, Research Triangle Park, N.C.

[HAMM 1973] R. W. HAMMING, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York.

[KNUT 1969] D. E. KNUTH, *The Art of Computer Programming*, Vol.II, *Seminumerical Algorithms*, Addison-Wesley, Reading, Mass.

[KUCK 1978] D. J. KUCK, *The Structure of Computers and Computations*, Vol.I, John Wiley & Sons, New York.

[MATH 1970] F. P. MATHUR and A. AVIZIENIS, "Reliability Analysis and Architecture of a Hybrid Redundant Digital System: Generalized Triple Modular Redundancy with Self-Repair," *AFI PS Conference Proc., Spring Joint Comp. Conf.*, Vol. 36, pp. 375–83.

[MEYE 1980] J. F. MEYER, D. G. FURCHTGOTT, and L. T. WU, "Performability Evaluation of the SIFT Computer," *IEEE Transactions on Computers*, June 1980, pp. 501–509.

[NG 1976] Y-W. NG, "Reliability Modeling and Analysis for Fault-Tolerant Computers," Ph.D. Dissertation, Computer Science Department, University of California at Los Angeles.

[RUDI 1964] W. RUDIN, *Principles of Mathematical Analysis*, McGraw-Hill, New York.

Chapter 4

Expectation

4.1 INTRODUCTION

The distribution function $F(x)$ or the density $f(x)$ [pmf $p(x_i)$ for a discrete random variable] completely characterizes the behavior of a random variable X. Frequently, however, we need a more concise description such as a single number or a few numbers, rather than an entire function. One such number is the **expectation** or the **mean**, denoted by $E[X]$. Similarly, the **median**, which is defined as any number x such that $P(X < x) \leqslant ½$ and $P(X > x) \leqslant ½$, and the **mode**, defined as the number x, for which $f(x)$ or $p(x_i)$ attains its maximum, are two other quantities sometimes used to describe a random variable X. The mean, median, and mode are often called **measures of central tendency** of a random variable X.

Definition (Expectation). The expectation, $E[X]$, of a random variable X is defined by:

$$E[X] = \begin{cases} \sum_i x_i \, p(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f(x)\, dx, & \text{if } X \text{ is continuous,} \end{cases} \tag{4.1}$$

provided that the relevant sum or integral is absolutely convergent; that is, $\sum_i |x_i| p(x_i) < \infty$ and $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$. If the right hand side in (4.1) is not absolutely convergent, then $E[X]$ does not exist. Most common random variables have finite expectation; however, problem 1 at the end of this section provides an example of a random variable whose expectation does not exist. Definition (4.1) can be extended to the case of mixed random variables through the use of Riemann-Stieltjes integral. Alternatively, the formula given in problem 2 at the end of this section can be used in the general case.

Example 4.1

Consider the problem of searching for a specific name in a table of names. A simple method is to scan the table sequentially, starting from one end, until we either find the name or reach the other end, indicating that the required name is missing from the table. The following is a Pascal program fragment for sequential search:

```
var T: array[0..n] of NAME;
Z: NAME;
I: 0..n;
begin {Z has been initialized elsewhere}
  T[0]:=Z; {T[0] is used as a sentinel or marker}
  I:=n;

  while Z ≠ T[I] do
    I:=I − 1;

  if I > 0 then {found; I points to Z}
   else {not found}.
end
```

In order to analyze the time required for sequential search, let X be the discrete random variable denoting the number of comparisons "$Z \neq T[I]$" made. Clearly, the set of all possible values of X is $\{1, 2, \ldots, n+1\}$, and $X = n+1$ for unsuccessful searches. Since the value of X is fixed for unsuccessful searches, it is more interesting to consider a random variable Y that denotes the number of comparisons on a successful search. The set of all possible values of Y is $\{1, 2, \ldots, n\}$. To compute the average search time for a successful search, we must specify the pmf of Y. In the absence of any specific information, it is natural to assume that Y is uniform over its range; that is:

$$p_Y(i) = \frac{1}{n}, \qquad 1 \leqslant i \leqslant n.$$

Then

$$E[Y] = \sum_{i=1}^{n} i p_Y(i) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Thus, on the average, approximately half the table needs to be searched. #

Example 4.2

The assumption of uniform distribution, used in Example 4.1, rarely holds in practice. It is possible to collect statistics on access patterns and use empirical distributions to reorganize the table so as to reduce the average search time. Unlike Example 4.1, we now assume for convenience that table search starts from the front. If α_i denotes the access probability for name $T[i]$, then the average successful search time $E[Y] = \sum i\alpha_i$. Then $E[Y]$ is minimized when names in the table are in the order of nonincreasing access probabilities; that is, $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n$. As an example, many tables in practice follow Zipf's law [ZIPF 1949]:

$$\alpha_i = \frac{c}{i}, \qquad 1 \leqslant i \leqslant n,$$

where the constant c is determined from the normalization requirement, $\sum_{i=1}^{n} \alpha_i = 1$. Thus:

$$c = \frac{1}{\sum_{i=1}^{n} \frac{1}{i}} = \frac{1}{H_n} \simeq \frac{1}{\ln(n)},$$

where H_n is the partial sum of a harmonic series; that is: $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

Now, if the names in the table are ordered as above, then the average search time is

$$E[Y] = \sum_{i=1}^{n} i\alpha_i = \frac{1}{H_n} \sum_{i=1}^{n} 1 = \frac{n}{H_n} \simeq \frac{n}{\ln(n)},$$

which is considerably less than the previous value $(n+1)/2$, for large n. #

Example 4.3

Recall the example of a computer system with five tape drives (Examples 1.1 and 2.2) and let X be the number of available tape drives. Then:

$$\begin{aligned} E[X] &= \sum_{i=0}^{5} i p_X(i) \\ &= 0 \cdot 1/32 + 1 \cdot 5/32 + 2 \cdot 10/32 + 3 \cdot 10/32 + 4 \cdot 5/32 + 5 \cdot 1/32 \\ &= 2.5. \end{aligned}$$ #

The example above illustrates that $E[X]$ need not correspond to a possible value of the random variable X. The expected value denotes the "center" of a probability distribution in the sense of a weighted average, or better, in the sense of a center of gravity.

Example 4.4

Let X be a continuous random variable with an exponential density given by:

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

Then

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} \lambda x e^{-\lambda x}\, dx.$$

Let $u = \lambda x$, then $du = \lambda dx$, and:

$$E[X] = \frac{1}{\lambda}\int_0^\infty ue^{-u}du = \frac{1}{\lambda}\Gamma 2 = \frac{1}{\lambda}, \quad \text{using formula (3.26).}$$

Thus, if a component obeys an exponential failure law with parameter λ (known as the **failure rate**), then its expected life, or its mean time to failure (MTTF), is $1/\lambda$. Similarly, if the interarrival times of jobs to a computer center are exponentially distributed with parameter λ (known as the **arrival rate**), then the mean (average) interarrival time is $1/\lambda$. Finally, if the service-time requirement of a job is an exponentially distributed random variable with parameter μ (known as the **service rate**), then the mean (average) service time is $1/\mu$. #

Problems

1. Consider a discrete random variable X with the pmf:

$$p_X(x) = \begin{cases} \dfrac{1}{x(x+1)}, & x = 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

Show that the function defined satisfies the properties of a pmf. Show that the formula (4.1) of expectation does not converge in this case and hence $E[X]$ is undefined. [*Hint:* Rewrite

$$\frac{1}{x(x+1)} \quad \text{as} \quad \frac{1}{x} - \frac{1}{x+1}.]$$

*2. Using integration by parts, show (assuming that $E[X]$, $\int_0^\infty [1-F(x)]\,dx$, and $\int_{-\infty}^0 F(x)\,dx$ are all finite) that for a continuous random variable X:

$$E[X] = \int_0^\infty [1 - F(x)]\,dx - \int_{-\infty}^0 F(x)\,dx.$$

This result states that the expectation of a random variable X equals the difference of the areas of the right-hand and left-hand shaded regions in Figure 4.P.1. (This formula applies to the case of discrete and mixed random variables as well.)

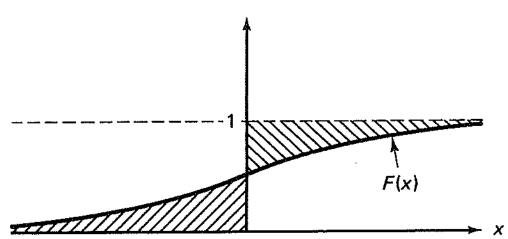

Figure 4.P.1 An alternative method of computing $E[X]$

3. For a given event A show that the expectation of its indicator random variable I_A (refer to Section 2.5.9) is given by:

$$E[I_A] = P(A).$$

4.2 MOMENTS

Let X be a random variable, and define another random variable Y as a function of X so that $Y = \phi(X)$. Suppose we wish to compute $E[Y]$. In order to apply Definition (4.1), we must first compute the pmf (or pdf in the continuous case) of Y by methods of Chapter 2 (or Chapter 3 in the continuous case). An easier method of computing $E[Y]$ is to use the following result:

$$E[Y] = E[\phi(X)] = \begin{cases} \sum_i \phi(x_i)p_X(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^\infty \phi(x) f_X(x)\,dx, & \text{if } X \text{ is continuous,} \end{cases} \tag{4.2}$$

(provided the sum or the integral on the right-hand side is absolutely convergent).

A special case of interest is the power function $\phi(X) = X^k$. For $k = 1, 2, 3, \ldots$, $E[X^k]$ is known as the kth moment of the random variable X. Note that the first moment, $E[X]$, is the ordinary expectation or the mean of X.

It is possible to show that if X and Y are random variables that have matching corresponding moments of *all* orders; that is, $E[X^k] = E[Y^k]$ for $k = 1, 2, \ldots$, then X and Y have the same distribution.

To center the origin of measurement, it is convenient to work with powers of $X - E[X]$. We define the kth central moment, μ_k, of the random variable X by $\mu_k = E[(X - E[X])^k]$. Of special interest is the quantity:

$$\mu_2 = E[(X - E[X])^2], \tag{4.3}$$

known as the variance of X, Var $[X]$, often denoted by σ^2.

Definition (Variance). The variance of a random variable X is

$$\text{Var}[X] = \mu_2 = \sigma^2 = \begin{cases} \int_{-\infty}^\infty (x - E[X])^2 f(x)\,dx & \text{if } X \text{ is continuous,} \\ \sum_i (x_i - E[X])^2 p(x_i) & \text{if } X \text{ is discrete.} \end{cases} \tag{4.4}$$

It is clear that Var $[X]$ is always a nonnegative number. The square root, σ, of the variance is known as the **standard deviation.** The variance and the stan-

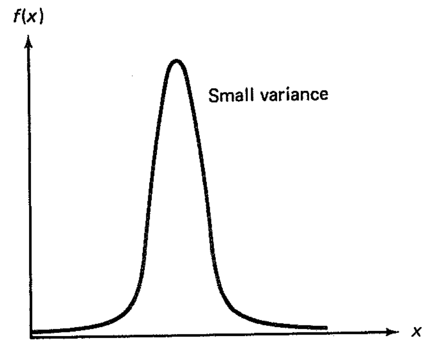

Figure 4.1 The pdf of a "concentrated" distribution

dard deviation are measures of the "spread" or "dispersion" of a distribution. If X has a "concentrated" distribution so that X takes values near to $E[X]$ with a large probability, then the variance is small (see Figure 4.1). Figure 4.2 shows a diffuse distribution—one with a large value of σ^2. Note that variance need not always exist (see problem 3 at the end of Section 4.7).

Example 4.5

Let X be an exponentially distributed random variable with parameter λ. Then, since $E[X] = 1/\lambda$ and $f(x) = \lambda e^{-\lambda x}$:

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

$$= \int_0^\infty \lambda x^2 e^{-\lambda x} dx - 2\int_0^\infty x e^{-\lambda x} dx + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x} dx$$

$$= \frac{1}{\lambda^2}\Gamma 3 - \frac{2}{\lambda^2}\Gamma 2 + \frac{1}{\lambda^2}\Gamma 1 = \frac{1}{\lambda^2}, \qquad \text{using formula (3.26).} \qquad \#$$

The standard deviation is expressed in the same units as the individual value of the random variable. If we divide it by the mean, then we obtain a

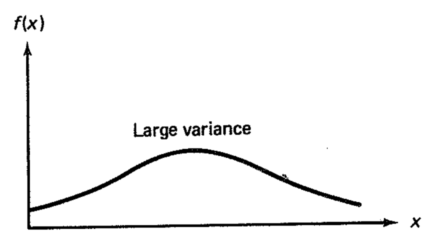

Figure 4.2 The pdf of a diffuse distribution

relative measure of the spread of the distribution of X. The coefficient of variation of a random variable X is denoted by C_X and defined by:

$$C_X = \frac{\sigma_X}{E[X]}. \qquad (4.5)$$

Note that the coefficient of variation of an exponential random variable is 1, so C_X is a measure of deviation from the exponential distribution.

Yet another function of X that is often of interest is $Y = aX + b$, where a and b are constants. It is not difficult to show that:

$$E[Y] = E[aX + b] = aE[X] + b. \qquad (4.6)$$

In particular, if a is zero, then $E[b] = b$; that is, the expectation of a constant random variable is that constant. If we take $a = 1$ and $b = -E[X]$, then we conclude that the first central moment, $\mu_1 = E[X - E[X]] = E[X] - E[X] = 0$.

Problems

***1.** [WOLM 1965]. The problem of dynamic storage allocation in the main memory of a computer system can be simplified by choosing a fixed node size, k, for allocation. Out of the k units (bytes, say) of storage allocated to a node, only $k - b$ bytes are available to the user, since b bytes are required for control information. Let the random variable L denote the length in bytes of a user request. Thus $\lceil L/(k-b) \rceil$ nodes must be allocated to satisfy the user request. Thus the total number of bytes allocated is $X = k \lceil L/(k-b) \rceil$. Find $E[X]$ as a function of k and $E[L]$. Then, by differentiating $E[X]$ with respect to k, show that the optimal value of k is approximately $b + \sqrt{2bE[L]}$.

2. Recall the problem of the mischievous student trying to open a password-protected file, and determine the expected number of trials $E[N_n]$ and the variance Var $[N_n]$ for both techniques described in problem 2, Section 2.5.

3. The number of hardware failures of a computer system in a week of operation has the following pmf:

No. of Failures	0	1	2	3	4	5	6
Probability	.18	.28	.25	.18	.06	.04	.01

(a) Find the expected number of failures in a week.
(b) Find the variance of the number of failures in a week.

4. In a Bell System study made in 1961 regarding the dialing of calls between White Plains, N.Y., and Sacramento, Calif., the pmf of the number of trunks, X, required for a connection was found to be:

i	1	2	3	4	5
$p_X(i)$	.50	.30	.12	.07	.01

Determine the distribution function of X. Compute $E[X]$, Var $[X]$ and mode $[X]$. Let Y denote the number of telephone switching exchanges the above call has to pass through. Then $Y = X + 1$. Determine the pmf, the distribution function, the mean, and the variance of Y.

5. Let X, Y, and Z respectively denote EXP(1), 2-stage hyperexponential with $\alpha_1 = .5 = \alpha_2$, $\lambda_1 = 2$, and $\lambda_2 = 2/3$, and 2-stage Erlang with parameter 2 random variables. Note that $E[X] = E[Y] = E[Z]$. Find the mode, the median, the variance, and the coefficient of variation of each of the random variables. Compare the densities of X, Y, and Z by plotting on the same graph. Similarly compare the three distribution functions.

6. Given a random variable X and two functions $h(x)$ and $g(x)$ satisfying the condition $h(x) \leqslant g(x)$ for all x, show that:

$$E[h(X)] \leqslant E[g(X)]$$

whenever both expectations exist.

4.3 EXPECTATION OF FUNCTIONS OF MORE THAN ONE RANDOM VARIABLE

Let $X_1, X_2, \ldots, X_n$ be n random variables defined on the same probability space and let $Y = \phi(X_1, X_2, \ldots, X_n)$. Then:

$$E[Y] = E[\phi(X_1, X_2, \ldots, X_n)]$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(x_1,x_2,\ldots,x_n) f(x_1,x_2,\ldots,x_n)\, dx_1 dx_2 \cdots dx_n & \text{(continuous case)}, \\ \displaystyle\sum_{x_1}\sum_{x_2} \cdots \sum_{x_n} \phi(x_1,x_2,\ldots,x_n) p(x_1,x_2,\ldots,x_n) & \text{(discrete case)}. \end{cases} \tag{4.7}$$

Example 4.6

Consider a moving head disk with the innermost cylinder of radius a and the outermost cylinder of radius b. We assume that the number of cylinders is very large and the cylinders are very close to each other, so that we may assume a continuum of cylinders. Let the random variables X and Y respectively, denote the current and the desired position of the head. Further assume that X and Y are independent and uniformly distributed over the interval (a, b). Therefore:

$$f_X(x) = f_Y(y) = \frac{1}{b-a}, \qquad a < x,\ y < b,$$

and

$$f(x,y) = \frac{1}{(b-a)^2}, \qquad a < x,\ y < b.$$

Head movement for a seek operation traverses a distance that is a random variable given by $|X - Y|$. The expected seek distance is then given by (see Figure 4.3):

$$\begin{aligned} E[|X-Y|] &= \int_a^b \int_a^b |x-y| f(x,y)\, dx\, dy \\ &= \int_a^b \int_a^b |x-y| \frac{1}{(b-a)^2}\, dx\, dy \\ &= \iint_{a \leqslant y < x \leqslant b} \frac{(x-y)}{(b-a)^2}\, dy\, dx + \iint_{a \leqslant x \leqslant y \leqslant b} \frac{(y-x)}{(b-a)^2}\, dy\, dx \\ &= \frac{2}{(b-a)^2} \int_a^b \int_a^x (x-y)\, dy\, dx, \qquad \text{by symmetry} \\ &= \frac{2}{(b-a)^2} \int_a^b \left(xy - \frac{y^2}{2}\right) \Big|_a^x dx \\ &= \frac{2}{(b-a)^2} \int_a^b \left(x^2 - ax - \frac{x^2}{2} + \frac{a^2}{2}\right) dx \\ &= \frac{2}{(b-a)^2} \left[\frac{b^3-a^3}{6} - \frac{a}{2}(b^2-a^2) + \frac{a^2(b-a)}{2}\right] \\ &= \frac{b-a}{3}. \end{aligned}$$

Thus, the expected seek distance is one third the maximum seek distance. Intuition may have led us to the incorrect conclusion that the expected seek distance is

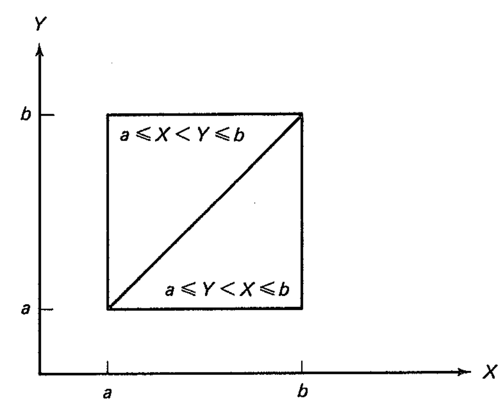

Figure 4.3 Two areas of integration for Example 4.6

half of the maximum. (In practice, the expected seek distance is even smaller due to correlations between successive requests [HUNT 1980].) #

Certain functions of random variables (e.g., sums), are of special interest and are of considerable use.

THEOREM 4.1 (THE LINEARITY PROPERTY OF EXPECTATION). Let X and Y be two random variables. Then the expectation of their sum is the sum of their expectations; that is, if $Z = X + Y$, then $E[Z] = E[X+Y] = E[X] + E[Y]$.

Proof:
We will prove the theorem assuming that X, Y, and hence Z are continuous random variables. Proof for the discrete case is very similar.

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f(x,y)\, dx\, dy$$

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

$$= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

(by definition of the marginal densities)

$$= E[X] + E[Y].$$

Note that the above theorem *does not* require that X and Y be independent. It can be generalized to the case of n variables—that is:

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

and to

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i], \tag{4.8}$$

where $a_1, \ldots, a_n$ are constants. For instance, let $X_1, X_2, \ldots, X_n$ be random variables (not necessarily independent) with a common mean $\mu = E[X_i]$ $(i = 1, 2, \ldots, n)$. Then the expected value of their sample mean (defined in Section 3.9) is equal to μ:

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu. \tag{4.9}$$

Example 4.7
We have noted that the variance:

$$\begin{aligned}\sigma^2 &= E[(X - E[X])^2]\\ &= E[X^2 - 2XE[X] + (E[X])^2]\\ &= E[X^2] - E[2XE[X]] + (E[X])^2, && \text{by (4.8)}\\ &= E[X^2] - 2E[X]E[X] + (E[X])^2\end{aligned}$$

(noting that $E[X]$ is a constant).
Thus,

$$\sigma^2 = E[X^2] - (E[X])^2. \tag{4.10}$$

This formula for Var $[X]$ is usually preferred over the original Definition (4.4). #

Unlike the case of expectation of a sum, the expectation of a product of two random variables does not have a simple form, unless the two random variables are independent.

THEOREM 4.2 $E[XY] = E[X]E[Y]$, if X and Y are independent random variables.

Proof:
We give a proof of the theorem under the assumption that X and Y are discrete random variables. The proof for the continuous case is similar.

$$\begin{aligned}E[XY] &= \sum_i \sum_j x_i y_j p(x_i, y_j)\\ &= \sum_i \sum_j x_i y_j p_X(x_i) p_Y(y_j) && \text{by independence}\\ &= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) = E[X]E[Y].\end{aligned}$$

Note that converse of Theorem 4.2 does not hold; that is, random variables X and Y may satisfy the relation $E[XY] = E[X]\,E[Y]$ without being independent.

The theorem above can be easily generalized to a mutually independent set of n random variables $X_1, X_2, \ldots, X_n$:

$$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i] \tag{4.11}$$

and further to

$$E[\prod_{i=1}^{n} \phi_i(X_i)] = \prod_{i=1}^{n} E[\phi_i(X_i)].$$

Again with the assumption of independence, the variance of a sum takes a simpler form also:

THEOREM 4.3. $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$, if X and Y are independent random variables.

Proof:
From the definition of variance:

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

by the linearity property of expectation.

The quantity $E[(X-E[X])(Y-E[Y])]$ is defined to be the **covariance** of X and Y and is denoted by $\text{Cov}(X,Y)$. It is easy to see that $\text{Cov}(X,Y)$ is zero when X and Y are independent:

$$\begin{aligned}
\text{Cov}(X,Y) &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\
&= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y]
\end{aligned}$$

by the linearity of expectation,

$$\begin{aligned}
&= E[XY] - E[X]E[Y] \\
&= 0,
\end{aligned}$$

by Theorem 4.2, since X and Y are independent.

Therefore, $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ if X and Y are independent random variables.

In case X and Y are not independent we obtain the formula:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}(X,Y). \tag{4.12}$$

The theorem above can be generalized for a set of n mutually independent random variables $X_1, X_2, \ldots, X_n$; and constants $a_1, a_2, \ldots, a_n$:

$$\text{Var}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2\, \text{Var}[X_i]. \tag{4.13}$$

Thus if $X_1, X_2, \ldots, X_n$ are mutually independent random variables with a common variance $\sigma^2 = \text{Var}[X_i]$ $(i = 1, 2, \ldots, n)$ then the variance of their sum is given by:

$$\text{Var}\left[\sum_{i=1}^{n} X_i\right] = n\,\text{Var}[X_i] = n\sigma^2 \tag{4.14}$$

and the variance of their sample mean is:

$$\begin{aligned}
\text{Var}[\bar{X}] = \text{Var}\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] &= \frac{1}{n^2}\,\text{Var}\left[\sum_{i=1}^{n} X_i\right] \qquad (4.15)\\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

We have noted that $\text{Cov}(X,Y) = 0$, if X and Y are independent random variables. However, it is possible for two random variables to satisfy the condition $\text{Cov}(X, Y) = 0$ without being independent.

Definition (Uncorrelated Random Variables). Random variables X and Y are said to be uncorrelated provided $\text{Cov}(X, Y) = 0$.

Since $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$, an equivalent definition of uncorrelated random variables is the condition $E[XY] = E[X]\,E[Y]$. It follows that independent random variables are uncorrelated, but the converse need not hold.

Example 4.8

Let X be uniformly distributed over the interval $(-1, 1)$ and let $Y = X^2$, so Y is completely dependent on X. Noting that for all odd values of $k > 0$, the kth moment $E[X^k] = 0$, we have:

$$E[XY] = E[X^3] = 0 \quad \text{and} \quad E[X]E[Y] = 0 \cdot E[Y] = 0.$$

Therefore X and Y are uncorrelated! #

We have declared that $\text{Cov}(X, Y) = 0$ means X and Y are uncorrelated. On the other hand, if X and Y are linearly related—that is, $X = aY$ for some constant $a \neq 0$—then, since $E[X] = aE[Y]$, we have:

$$\text{Cov}(X, Y) = a\,\text{Var}[Y] = \frac{1}{a}\,\text{Var}[X]$$

or

$$\text{Cov}^2(X, Y) = \text{Var}[X]\,\text{Var}[Y].$$

In the general case, it can be shown that:

$$0 \leqslant \text{Cov}^2(X, Y) \leqslant \text{Var}[X]\,\text{Var}[Y] \tag{4.16}$$

using the following Cauchy-Schwarz inequality:

$$(E[XY])^2 \leqslant E[X^2]E[Y^2]. \tag{4.17}$$

Cov (X, Y) measures the degree of linear dependence (or the degree of correlation) between the two random variables. Recalling Example 4.8, we note that the notion of covariance completely misses the quadratic dependence. It is often useful to define a measure of this dependence in a scale-independent fashion. The correlation coefficient $\rho(X, Y)$ is defined by:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\,\text{Var}[Y]}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{4.18}$$

whenever σ_X and σ_Y are defined.

Using the relation (4.16), we conclude that:

$$-1 \leqslant \rho(X, Y) \leqslant 1. \tag{4.19}$$

Also:

$$\rho(X, Y) = \begin{cases} -1, & \text{if } X = -aY \ (a>0), \\ 0, & \text{if } X \text{ and } Y \text{ are uncorrelated}, \\ +1, & \text{if } X = aY \ (a>0). \end{cases} \tag{4.20}$$

Problems

1. [BLAK 1979] Consider discrete random variables X and Y with the joint pmf as shown below:

X \ Y	-1	0	1
-2	1/16	1/16	1/16
-1	1/8	1/16	1/8
1	1/8	1/16	1/8
2	1/16	1/16	1/16

Are X and Y independent? Are they uncorrelated?

2. Consider the discrete version of Example 4.6 and assume that the records of a file are evenly scattered over n tracks of a moving-head disk. Compute the expected number of gaps X between tracks that the head will pass over between two reads. Also compute the variance of X. [Hint:

$$\sum_{i=1}^{N} i^3 = \left(\sum_{i=1}^{N} i\right)^2$$

is a useful identity.] Next assume that the seek time T is a function $\theta(X)$ of the number of gaps passed over. Compute $E[T]$ and Var $[T]$, assuming $T = 30.0 + 0.5X$.

3. Consider a directed graph G with n nodes. Let X_{ij} be a variable defined so that:

$$X_{ij} = \begin{cases} 0 & \text{if there is no edge between node } i \text{ and node } j, \\ 1 & \text{otherwise.} \end{cases}$$

Assume that the X_{ij}'s are mutually independent Bernoulli random variables with parameter p. The corresponding graph is called a **p-random-graph.** Find the pmf, the expected value, and the variance of the total number of edges X in the graph.

4. Let $X_1, X_2, \ldots, X_n$ be mutually independent and identically distributed random variables with mean μ and variance σ^2. Let $\bar{X} = (\sum_{i=1}^{n} X_i)/n$. Show that:

$$\sum_{k=1}^{n} (X_k - \bar{X})^2 = \sum_{k=1}^{n} (X_k - \mu)^2 - n(\bar{X} - \mu)^2$$

and hence:

$$E\left[\sum_{k=1}^{n} (X_k - \bar{X})^2\right] = (n-1)\sigma^2.$$

5. A certain telephone company charges for calls in the following way: \$0.20 for the first three minutes or less; \$0.08 per minute for any additional time. Thus if X is the duration of a call, the cost Y is given by:

$$Y = \begin{cases} 0.20, & 0 \leqslant X \leqslant 3, \\ 0.20 + 0.08(X-3), & X \geqslant 3. \end{cases}$$

Find the expected value of the cost of a call ($E[Y]$), assuming that the duration of a call is exponentially distributed with a mean of $1/\lambda$ minutes. Use $1/\lambda = 1, 2, 3, 4$, and 5 minutes.

6. Show that $\text{Cov}^2(X, Y) \leqslant \text{Var}[X]\,\text{Var}[Y]$.

7. Random variables X and Y are said to be **orthogonal** if and only if $E[XY] = 0$.
   (a) If X and Y are orthogonal determine the conditions under which they are uncorrelated.
   (b) If X and Y are uncorrelated, determine the conditions under which they are orthogonal.

8. Consider random variables X and Y with the joint pdf (bivariate Gaussian):

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

where $\rho \neq \pm 1$. Show that Cov $(X, Y) = \rho\sigma_X\sigma_Y$. Hence show that if X, Y are jointly Gaussian and uncorrelated (i.e., $\rho = 0$), then they are also independent. Note that this is *not* true in general.

4.4 TRANSFORM METHODS

In many probability problems, the form of the density function (or the pmf in the discrete case) may be so complex so as to make computations difficult, if not impossible. As an example, recall the analysis of the program MAX. A transform can provide a compact description of a distribution, and it is relatively easy to compute the mean, the variance, and other moments directly from a transform rather than resorting to a tedious sum (discrete case) or an equally tedious integral (continuous case). The transform methods are particularly useful in problems involving sums of independent random variables and in solving difference equations (discrete case) and differential equations (continuous case) related to a stochastic process. We will introduce the z-transform (also called the probability generating function), the Laplace transform, and the characteristic function (also called the Fourier transform). We will first define the moment generating function and derive the above three transforms as special cases.

For a random variable X, $e^{X\theta}$ is another random variable. The expectation $E[e^{X\theta}]$ will be a function of θ. Define the moment generating function (MGF) $M_X(\theta)$, abbreviated $M(\theta)$, of the random variable X by:

$$M(\theta) = E[e^{X\theta}] \tag{4.21}$$

so that:

$$M(\theta) = \begin{cases} \sum_j e^{x_j\theta} p(x_j), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} e^{x\theta} f(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{4.22}$$

The expectation defining $M(\theta)$ may not exist for all real numbers θ, but for most problems that we encounter, there will be an interval of θ values within which $M(\theta)$ does exist. Note that the Definition (4.21) allows us to define the moment-generating function for a mixed random variable as well.

The closely related characteristic function of a random variable X is given by

$$N_X(\tau) = N(\tau) = M_X(i\tau). \tag{4.23}$$

Here i denotes $\sqrt{-1}$. $N(\tau)$ is known among electrical engineers as the Fourier transform. The advantage here is that for any X, its characteristic function, $N_X(\tau)$, is always defined for all τ. If X is a nonnegative continuous random variable, then we define the (one-sided) Laplace transform:

$$L_X(s) = L(s) = M_X(-s) = \int_0^{\infty} e^{-sx} f(x)\, dx. \tag{4.24}$$

Finally, if X is a nonnegative integer-valued discrete random variable, then we define its z-transform:

$$G_X(z) = G(z) = E[z^X] = M_X(\ln z) = \sum_{i=0}^{\infty} p_X(i) z^i. \tag{4.25}$$

The reasons for the usefulness of transform methods will be summarized by the following properties of the transforms. We will give the properties of the moment generating function, but by an appropriate substitution for θ, a similar property can be stated for all the other transforms as well.

THEOREM 4.4 (LINEAR TRANSLATION) Let $Y = aX + b$; then:

$$M_Y(\theta) = e^{b\theta} M_X(a\theta).$$

Proof:

$$\begin{aligned} E[e^{Y\theta}] &= E[e^{(aX+b)\theta}] \\ &= E[e^{b\theta} e^{aX\theta}] \\ &= e^{b\theta} E[e^{aX\theta}] \end{aligned}$$

by the linearity property of expectation.

THEOREM 4.5 (THE CONVOLUTION THEOREM) Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables on a given probability space, and let $Y = \sum_{i=1}^{n} X_i$. If $M_{X_i}(\theta)$ exists for all i, then $M_Y(\theta)$ exists, and:

$$M_Y(\theta) = M_{X_1}(\theta) M_{X_2}(\theta) \cdots M_{X_n}(\theta).$$

Thus the moment generating function of a sum of independent random variables is the product of the moment generating functions.

Proof:

$$\begin{aligned} M_Y(\theta) = E[e^{Y\theta}] &= E\left[e^{(X_1+X_2+\cdots+X_n)\theta}\right] \\ &= E[\prod_{i=1}^{n} e^{X_i\theta}] \\ &= \prod_{i=1}^{n} E[e^{X_i\theta}], \quad \text{by independence,} \\ &= \prod_{i=1}^{n} M_{X_i}(\theta). \end{aligned}$$

Thus we may find the transform of a sum of independent random variables without any n-dimensional integration. But the technique will be of little value unless we can recover the distribution function from the transform. The following theorem, which we state without proof, is useful in this regard.

THEOREM 4.6 (CORRESPONDENCE THEOREM OR UNIQUENESS THEOREM) If $M_{X_1}(\theta) = M_{X_2}(\theta)$ for all θ, then $F_{X_1}(x) = F_{X_2}(x)$ for all x.

In other words, if two random variables have the same transform, then they have the same distribution function.

Next we study the **moment generating property** of the MGF. $e^{X\theta}$ can be expanded into a power series:

$$e^{X\theta} = 1 + X\theta + \frac{X^2\theta^2}{2!} + \cdots + \frac{X^k\theta^k}{k!} + \cdots.$$

Taking expectations on both sides (assuming all the expectations exist):

$$\begin{aligned} M(\theta) &= E[e^{X\theta}] \\ &= 1 + E[X]\theta + \cdots + \frac{E[X^k]\theta^k}{k!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{E[X^k]\theta^k}{k!}. \end{aligned}$$

Therefore, the coefficient of $\theta^k/k!$ in the power-series expansion of its MGF yields the kth moment $E[X^k]$ of the random variable X. Alternatively:

$$E[X^k] = \frac{d^kM}{d\theta^k}\Big|_{\theta=0}, \qquad k = 1, 2, \ldots. \tag{4.26}$$

Note that $E[X^0] = M(0) = 1$.

The corresponding property for the Laplace transform is:

$$E[X^k] = (-1)^k \frac{d^kL(s)}{ds^k}\Big|_{s=0}, \qquad k = 1, 2, \ldots. \tag{4.27}$$

For the z-transform:

$$E[X] = \frac{dG}{dz}\Big|_{z=1}, \tag{4.28}$$

$$E[X^2] = \frac{d^2G}{dz^2}\Big|_{z=1} + \frac{dG}{dz}\Big|_{z=1}, \tag{4.29}$$

and for the characteristic function:

$$E[X^k] = (-i)^k \frac{d^kN}{d\tau^k}\Big|_{\tau=0}, \qquad k = 1, 2, \ldots. \tag{4.30}$$

Example 4.9

Let X be exponentially distributed with parameter λ. Then:

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

and

$$\begin{aligned} L_X(s) &= \int_0^{\infty} e^{-sx}\lambda e^{-\lambda x}\,dx \\ &= \frac{\lambda}{\lambda+s}\int_0^{\infty}(\lambda+s)e^{-(s+\lambda)x}\,dx \\ &= \frac{\lambda}{\lambda+s}. \end{aligned} \tag{4.31}$$

Now, using (4.27), we have:

$$E[X] = (-1)\frac{dL_X}{ds}\Big|_{s=0} = (-1)\frac{-\lambda}{(\lambda+s)^2}\Big|_{s=0} = \frac{1}{\lambda},$$

as derived earlier in Example 4.4. Also:

$$E[X^2] = \frac{d^2L_X}{ds^2}\Big|_{s=0} = \frac{2\lambda}{(\lambda+s)^3}\Big|_{s=0} = \frac{2}{\lambda^2}$$

and

$$\mathrm{Var}\,[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

#

Example 4.10

We are now in a position to complete the analysis of program MAX (Section 2.6). Recall that the PGF of the random variable X_n was shown to have the recurrence:

$$G_{X_n}(z) = \frac{(z+n-1)}{n}\,G_{X_{n-1}}(z)$$

with

$$G_{X_1}(z) = 1.$$

Here X_n denotes the number of executions of the **then** clause in program MAX. Then the expected number of executions of the **then** clause is derived using the property (4.28):

$$\begin{aligned} E[X_n] &= \frac{dG_{X_n}}{dz}\Big|_{z=1} \\ &= \frac{1}{n}G_{X_{n-1}}(1) + \frac{z+n-1}{n}\Big|_{z=1}\cdot\frac{dG_{X_{n-1}}(z)}{dz}\Big|_{z=1} \\ &= \frac{1}{n} + E[X_{n-1}] \end{aligned}$$

(since $G_X(1) = 1$ for any PGF). With $E[X_1] = 0$, we have:

$$E[X_n] = \sum_{i=2}^{n} \frac{1}{i} = H_n - 1 \simeq \ln n.$$

To compute the variance of X_n, first observe that if Y_k is a Bernoulli random variable with parameter $p = 1/k$, then

$$G_{Y_k}(z) = (1 - \frac{1}{k}) + \frac{z}{k} = \frac{z + k - 1}{k}.$$

If $Y_2, Y_3, \ldots, Y_n$ are mutually independent, then, using the convolution theorem, $W = \sum_{k=2}^{n} Y_k$ has the PGF:

$$\begin{aligned} G_W(z) &= \prod_{k=2}^{n} G_{Y_k}(z) \\ &= \prod_{k=2}^{n} \frac{z + k - 1}{k} \\ &= G_{X_n}(z). \end{aligned}$$

So we conclude, by the correspondence theorem, that X_n has the same distribution as W; that is:

$$X_n = \sum_{k=2}^{n} Y_k.$$

(Note that although X_n is the sum of $n - 1$ mutually independent Bernoulli random variables, it is not a binomial random variable; why?) Now since the Y_k's are mutually independent, we use formula (4.13) to obtain:

$$\begin{aligned} \text{Var}[X_n] &= \sum_{k=2}^{n} \text{Var}[Y_k] \\ &= \sum_{k=2}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) \\ &= H_n - H_n^{(2)} \end{aligned}$$

where $H_n^{(2)}$ is defined to be $\sum_{k=1}^{n} \frac{1}{k^2}$.

The power of the notion of transforms should now be clear, since we could compute the mean and the variance without the explicit knowledge of pmf, which in this case is quite complex (it involves Stirling numbers!). #

Example 4.11 (Analysis of Straight Selection Sort)

We are given an array:

var a: **array** [1..n] **of** item, where **type** item = **record** key: integer; info: τ **end.**

We are required to sort the array so that keys are in nondecreasing order. We can use the following procedure [WIRT 1976]:


```
for i:= n downto 2 do
begin
1: "assign the index of the item with the largest
    key among the items a[1], a[2], ..., a[i] to k";
2: "exchange a[i] and a[k]"
end.
```

Assume that each element of the array is a large record and, therefore, exchanging (or moving) items is expensive. The total number of moves due to the second statement is easily computed and seen to be a fixed number. But the number of moves in the first statement is variable. Assume that the program MAX is used to perform this operation. Then the number of moves for a fixed value of i will be given by X_i, which was studied in Chapter 2 and in the previous example. Now the total number of moves, W_n, contributed by the first statement is given by:

$$W_n = \sum_{i=2}^{n} X_i.$$

Then:

$$\begin{aligned} E[W_n] &= \sum_{i=2}^{n} E[X_i] = \sum_{i=2}^{n} (H_i - 1) \\ &\simeq n(\ln n) \end{aligned}$$

Example 4.12

Let X_i ($i = 1,2$) be independent exponentially distributed random variables with parameters λ_i. If $\lambda_1 = \lambda_2 = \lambda$, then $X = X_1 + X_2$ will be a two-stage Erlang random variable. Assume $\lambda_1 \neq \lambda_2$, implying that X is a hypoexponentially distributed random variable. Using formula (4.31), we have:

$$L_{X_1}(s) = \frac{\lambda_1}{\lambda_1 + s} \quad \text{and} \quad L_{X_2}(s) = \frac{\lambda_2}{\lambda_2 + s}.$$

By the convolution theorem:

$$L_X(s) = \frac{\lambda_1 \lambda_2}{(\lambda_1 + s)(\lambda_2 + s)}. \tag{4.32}$$

We expand this expression into a partial fraction:

$$L_X(s) = \frac{a_1 \lambda_1}{\lambda_1 + s} + \frac{a_2 \lambda_2}{\lambda_2 + s},$$

where:

$$a_1 = \frac{\lambda_2}{\lambda_2 - \lambda_1} \quad \text{and} \quad a_2 = \frac{\lambda_1}{\lambda_1 - \lambda_2}.$$

Recalling that if Y is EXP (λ), then $L_Y(s) = \lambda/(\lambda + s)$, we conclude (using the uniqueness theorem of Laplace transforms) that:

$$f_X(x) = a_1 \lambda_1 e^{-\lambda_1 x} + a_2 \lambda_2 e^{-\lambda_2 x}$$

$$= \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 x} + \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2} e^{-\lambda_2 x},$$

which is the hypoexponential density. #

More generally, if the X_i's $(i = 1, 2, \ldots, n)$ are mutually independent and exponentially distributed with parameters λ_i $(\lambda_i \neq \lambda_j,\ i \neq j)$, then $X = \sum_{i=1}^{n} X_i$ is an n-stage hypoexponential random variable and

$$L_X(s) = \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i + s}.$$

Using the technique of partial fraction expansion [KOBA 1978], the Laplace transform of X can be rewritten as:

$$L_X(s) = \sum_{i=1}^{n} \frac{a_i\lambda_i}{\lambda_i + s}, \tag{4.33}$$

where:

$$a_i = \prod_{\substack{j=1 \\ j\neq i}}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i}. \tag{4.34}$$

Again, from the uniqueness theorem of Laplace transforms, it follows that

$$f_X(x) = \sum_{i=1}^{n} a_i\lambda_i e^{-\lambda_i x} \tag{4.35}$$

(Although, this form of f_X appears like a hyperexponential density function, it is quite a different hypoexponential density; why?)

Example 4.13

Let X be normally distributed with parameters μ and σ^2. Then:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad -\infty < x < \infty.$$

The characteristic function of X is given by:

$$N_X(\tau) = \int_{-\infty}^{\infty} e^{i\tau x} f_X(x)\, dx.$$

Making the change of variables $y = (x - \mu)/\sigma$, we obtain:

$$N_X(\tau) = \int_{-\infty}^{\infty} e^{i(\sigma y+\mu)\tau} \frac{e^{-(1/2)y^2}dy}{\sqrt{2\pi}}$$

$$= e^{i\tau\mu} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} e^{i\tau\sigma y}\, dy$$

$$= e^{i\tau\mu + \frac{(i\tau\sigma)^2}{2}} \int_{-\infty}^{\infty} e^{-1/2(y-i\tau\sigma)^2} \frac{dy}{\sqrt{2\pi}},$$

(noting that $i^2 = -1$). Thus, the characteristic function of a normal random variable is given by:

$$N_X(\tau) = e^{i\tau\mu - \tau^2\sigma^2/2}, \tag{4.36}$$

since it can be shown that:

$$\int_{-\infty}^{\infty} e^{-1/2(y-i\tau\sigma)^2} \frac{dy}{\sqrt{2\pi}} = 1.$$

[It is the area under the normal density $N(i\tau\sigma, 1)$]. Check that

$$N_X(0) = e^0 = 1.$$

To compute the expected value, we use equation (4.30):

$$E[X] = \frac{1}{i} \frac{dN_X}{d\tau}\Big|_{\tau=0}$$

$$= \frac{1}{i} [(i\mu - \tau\sigma^2)\, e^{i\tau\mu - \frac{\tau^2\sigma^2}{2}}]\Big|_{\tau=0}$$

$$= \frac{1}{i} [i\mu e^0] = \mu.$$

Similarly, it can be shown that:

$$E[X^2] = \frac{1}{i^2} \frac{d^2N_X}{d\tau^2}\Big|_{\tau=0}$$

$$= \sigma^2 + \mu^2$$

(after the computations are worked out). #

Thus the normal distribution $N(\mu, \sigma^2)$ has mean μ and variance σ^2. This distribution is completely specified by the two parameters.

Example 4.14 (Proof of Theorem 3.6)

Let $X_1, X_2, \ldots; X_n$ be mutually independent Gaussian random variables so that X_j is $N(\mu_j, \sigma_j^2)$, $j = 1, 2, \ldots, n$. Then from formula (4.36) we have:

$$N_{X_j}(\tau) = e^{i\tau\mu_j - \tau^2\sigma_j^2/2}, \qquad j = 1, 2, \ldots, n.$$

Let $Y = \sum_{i=1}^{n} X_i$; then, using the convolution theorem, we have:

$$N_Y(\tau) = \prod_{j=1}^{n} N_{X_j}(\tau) = e^{i\tau\mu - \frac{\tau^2\sigma^2}{2}},$$

where

$$\mu = \sum_{j=1}^{n} \mu_j \quad \text{and} \quad \sigma^2 = \sum_{j=1}^{n} \sigma_j^2.$$

Comparing the characteristic function above with that in (4.36), we conclude that Y is $N(\mu, \sigma^2)$. #

Characteristic functions are somewhat more complex than the MGF, but they have two advantages. First, $N_X(\tau)$ is finite for all random variables X and for all real numbers τ. Second, the characteristic function possesses the inversion property, so that the density $f_X(s)$ may be derived from $N_X(\tau)$ by the inversion formula:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix\tau} N_X(\tau)\, d\tau. \tag{4.37}$$

Inversion of a Laplace transform is usually performed using a table lookup. It is helpful first to perform a partial fraction expansion of the transform. See Appendix D for further details.

Problems

1. Show that if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent random variables, then the random variable:

$$Y = X_1 - X_2$$

is also normally distributed with:

$$E[Y] = \mu_1 - \mu_2 \quad \text{and} \quad \text{Var}[Y] = \sigma_1^2 + \sigma_2^2.$$

Generalize to the case of n mutually independent random variables with:

$$X_i \sim N(\mu_i, \sigma_i^2) \quad \text{and} \quad Y = \sum_{i=1}^{n} a_i X_i.$$

2. Take the program to find the maximal element of a given (one dimensional) array B of size n (discussed in Chapter 2 and also on p. 95 of Knuth, Vol. I). Call this subroutine MAX. Write a driver for this subroutine that generates all $n!$ permutations of the set $\{1, 2, \ldots, n\}$ and, for each such permutation, loads it into the array B and calls subroutine MAX. Count the number of exchanges made in subroutine MAX. Add the number of exchanges over all permutations and divide the sum by $(n!)$. Check whether the result equals $H_n - 1$. Similarly compute the variance and check it against the expression $H_n - H_n^{(2)}$. Use $n = 1, 3, 5$, and 10.

To generate $n!$ permutations systematically, you may refer to the following article in the computing surveys: R. Sedgewick, "Permutation Generation Methods," *Computing Surveys*, June 1977, pp. 137-66.

4.5 MOMENTS AND TRANSFORMS OF SOME IMPORTANT DISTRIBUTIONS

4.5.1 Discrete Uniform Distribution

The pmf is given by:

$$p_X(i) = \frac{1}{n}, \qquad 1 \leqslant i \leqslant n.$$

Therefore:

$$E[X^k] = \sum_{i=1}^{n} \frac{i^k}{n}.$$

Then, the mean:

$$E[X] = \frac{n+1}{2},$$

and the variance:

$$\begin{aligned} \text{Var}[X] &= E[X^2] - (E[X])^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{n+1}{12}[2(2n+1) - 3(n+1)] \\ &= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}. \end{aligned}$$

The coefficient of variation:

$$C_X = \sqrt{\frac{n^2-1}{3(n+1)^2}} = \sqrt{\frac{1}{3}\left(1 - \frac{2}{n+1}\right)},$$

so

$$0 \leqslant C_X < \frac{1}{\sqrt{3}}.$$

The generating function in this case is:

$$G_X(z) = \sum_{i=1}^{n} \frac{1}{n} z^i = \frac{1}{n} \sum_{i=1}^{n} z^i.$$

4.5.2 Bernoulli pmf

$$p_X(0) = q, \qquad p_X(1) = p, \qquad p + q = 1.$$
$$E[X^k] = 0^k \cdot q + 1^k \cdot p = p, \qquad k = 1, 2, \ldots .$$

Therefore, the mean:

$$E[X] = p,$$

and the variance:

$$\text{Var }[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) = pq.$$

The coefficient of variation is:

$$C_X = \sqrt{\frac{q}{p}},$$

and the generating function:

$$G_X(z) = (1-p) + pz = q + pz.$$

4.5.3 Binomial Distribution

Note that a binomial random variable X is the sum of n mutually independent Bernoulli random variables $X_1, X_2, \ldots, X_n$. Thus:

$$X = \sum_{i=1}^{n} X_i,$$

and the linearity property of the expectation yields the result:

$$E[X] = \sum_{i=1}^{n} E[X_i] = np.$$

Similarly, using formula (4.13), we get the variance:

$$\text{Var }[X] = \sum_{i=1}^{n} \text{Var }[X_i] = npq.$$

The coefficient of variation:

$$C_X = \sqrt{\frac{npq}{n^2p^2}} = \sqrt{\frac{q}{np}}.$$

Thus, the expected number of successes in a sequence of n Bernoulli trials is np. Also note that the coefficient of variation reduces as n increases, and it approaches zero in the limit as $n \to \infty$. This observation is related to the weak law of large numbers, as will be seen later. We can easily obtain the generating function, using the convolution theorem:

$$G_X(z) = \prod_{i=1}^{n} G_{X_i}(z) = (q + pz)^n.$$

4.5.4 Geometric Distribution

The pmf is given by:

$$p_X(i) = pq^{i-1}, \qquad i = 1, 2, \ldots .$$

The mean is computed by:

$$\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} ipq^{i-1} \\
&= p\sum_{i=1}^{\infty} iq^{i-1} \\
&= p \sum_{i=0}^{\infty} \frac{d}{dq} (q^i) \\
&= p \frac{d}{dq} \left(\sum_{i=0}^{\infty} q^i\right) \\
&= p \frac{d}{dq} \left[\frac{1}{1-q}\right] \\
&= \frac{p}{(1-q)^2} \\
&= \frac{1}{p}.
\end{aligned}$$

Therefore, if we assume that at the end of a CPU burst, a program requests an I/O operation with probability q and it finishes execution with probability p, then the average number of CPU bursts per program is given by $1/p$. Similarly, if a communication channel transmits a message correctly, on each trial, with probability p, then the average number of trials required for a successful transmission is $1/p$.

The generating function of X is given by:

$$G_X(z) = \sum_{i=1}^{\infty} pq^{i-1}z^i$$
$$= pz \sum_{i=1}^{\infty} (qz)^{i-1}$$
$$= pz\sum_{j=0}^{\infty} (qz)^j$$
$$= \frac{pz}{1-qz}.$$

From this, $E[X]$ can be derived in an easier fashion:

$$E[X] = \frac{dG}{dz}\Big|_{z=1}$$
$$= \frac{p(1-qz) - pz(-q)}{(1-qz)^2}\Big|_{z=1}$$
$$= \frac{p(1-q)+pq}{(1-q)^2}$$
$$= \frac{p}{p^2}$$
$$= \frac{1}{p}.$$

The variance is computed in a fashion similar to that used for the mean; we get:

$$\text{Var }[X] = \frac{q}{p^2} \quad \text{and} \quad C_X = \sqrt{\frac{qp^2}{p^2}} = \sqrt{q} = \sqrt{1-p}.$$

For the modified geometric distribution, with the pmf $p_Y(i) = pq^i$, $i = 0, 1, 2, \ldots$:

$$E[Y] = \frac{q}{p}, \qquad \text{Var }[Y] = \frac{q}{p^2}, \qquad C_Y = \sqrt{\frac{qp^2}{p^2q^2}} = \frac{1}{\sqrt{q}},$$

and the generating function is:

$$G_Y(z) = \frac{p}{1-qz}.$$

4.5.5 Poisson pmf

$$p_X(i) = \frac{\alpha^i e^{-\alpha}}{i!}, \qquad 0 \leqslant i < \infty.$$

Then:

$$E[X] = \sum_{i=0}^{\infty} \frac{i\alpha^i}{i!} e^{-\alpha}$$
$$= \alpha e^{-\alpha} \sum_{i=1}^{\infty} \frac{\alpha^{i-1}}{(i-1)!}$$
$$= \alpha e^{-\alpha} e^{\alpha} = \alpha.$$

If the number of job arrivals to a computer center in interval $(0, t]$ is Poisson distributed with parameter $\alpha = \lambda t$, then the average number of arrivals in that interval is λt. Thus, the average arrival rate of jobs is λ.

The Var $[X]$ is easily computed to be α. Therefore:

$$C_X = \frac{1}{\sqrt{\alpha}}.$$

The generating function is given by:

$$G_X(z) = \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} z^k = e^{-\alpha}\sum_{k=0}^{\infty} \frac{(\alpha z)^k}{k!} = e^{-\alpha}e^{\alpha z} = e^{-\alpha(1-z)}.$$

4.5.6 Continuous Uniform Distribution

The density function is given by:

$$f_X(x) = \frac{1}{b-a}, \qquad a < x < b.$$

Then:

$$E[X] = \int_a^b \frac{x}{b-a}\, dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2},$$

the midpoint of the interval (a, b). The kth moment is computed as:

$$E[X^k] = \frac{1}{b-a}\int_a^b x^k\, dx = \frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)}.$$

Therefore:

$$\text{Var }[X] = E[X^2] - (E[X])^2$$
$$= \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4}$$
$$= \frac{(b-a)^2}{12}$$

and

$$C_X = \frac{b-a}{b+a}\sqrt{\frac{1}{3}}.$$

Assuming $0 \leqslant a < b$, the Laplace transform of X is:

$$L_X(s) = \int_a^b e^{-sx} \frac{1}{b-a}\, dx$$
$$= \frac{e^{-as} - e^{-bs}}{s(b-a)}.$$

4.5.7 Exponential Distribution

We have already determined that if the density is given by:

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0,$$

then the mean:

$$E[X] = \frac{1}{\lambda},$$

the variance:

$$\text{Var}\,[X] = \frac{1}{\lambda^2},$$

the coefficient of variation:

$$C_X = 1,$$

and the Laplace transform:

$$L_X(s) = \frac{\lambda}{\lambda + s}.$$

4.5.8 Gamma Distribution

The density function of the random variable X is given by:

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \qquad x > 0.$$

Then, making the substitution $u = \lambda x$, we compute the mean:

$$E[X] = \int_0^\infty \frac{x^\alpha \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)}\, dx = \frac{1}{\lambda \Gamma(\alpha)} \int_0^\infty u^\alpha e^{-u}\, du,$$

and hence, using formula (3.26):

$$E[X] = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

Similarly, the variance is computed to be:

$$\text{Var}\,[X] = \frac{\alpha}{\lambda^2}$$

and thus:

$$C_X = \frac{1}{\sqrt{\alpha}}.$$

Note that if α is an integer, then the above results could be shown by the properties of sums, since X will be the sum of α exponential random variables. Note also that the coefficient of variation of a gamma random variable is less than 1 if $\alpha > 1$; it is equal to 1 if $\alpha = 1$; and otherwise the coefficient of variation is greater than 1. Thus the gamma family is capable of modeling a very powerful class of random variables exhibiting from almost none to a very high degree of variability.

The Laplace transform is given by:

$$L_X(s) = \int_0^\infty e^{-sx} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}\, dx$$
$$= \frac{\lambda^\alpha}{(\lambda+s)^\alpha} \int_0^\infty \frac{(\lambda+s)^\alpha x^{\alpha-1} e^{-(\lambda+s)x}}{\Gamma(\alpha)}\, dx$$
$$= \frac{\lambda^\alpha}{(\lambda+s)^\alpha}$$

since the last integral is the area under a gamma density with parameter $\lambda + s$ and α—that is, 1. If α were an integer, this result could be derived using the convolution property of the Laplace transforms.

4.5.9 Hypoexponential Distribution

We have seen that if $X_1, X_2, \ldots, X_n$ are mutually independent exponentially distributed random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ ($\lambda_i \neq \lambda_j$, $i \neq j$), respectively, then:

$$X = \sum_{i=1}^n X_i$$

is hypoexponentially distributed with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$; that is, X is HYPO $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. The mean of X can then be obtained using the linearity property of expectation, so that:

$$E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \frac{1}{\lambda_i}.$$

Also, because of independence of the X_i's, we get the variance of X as:

$$\text{Var}\,[X] = \sum_{i=1}^n \text{Var}\,[X_i] = \sum_{i=1}^n \frac{1}{\lambda_i^2},$$