Chapter 5

Conditional Distribution and Conditional Expectation

5.1 INTRODUCTION

We have seen that if two random variables are independent, then their joint distribution can be determined from their marginal distribution functions. In the case of dependent random variables, however, the joint distribution can not be determined in this simple fashion. This leads us to the notions of conditional pmf, conditional pdf, and conditional distribution.

Recalling the definition of conditional probability, $P(A|B)$, for twc events A and B, we can define the **conditional probability** $P(A|X = x)$ of event A, given that the event $[X = x]$ has occurred, as:

$$P(A|X = x) = \frac{P(A \text{ occurs and } X = x)}{P(X = x)}, \tag{5.1}$$

whenever $P(X = x) \neq 0$. In Chapter 3 we noted that if X is a continuous random variable, then $P(X = x) = 0$ for all x. In this case, Definition (5.1) is not satisfactory. On the other hand, if X is a discrete random variable, then Definition (5.1) is adequate:

Definition (Conditional pmf). Let X and Y be discrete random variables having a joint pmf $p(x, y)$. The conditional pmf of Y given X is defined by:

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= \frac{P(Y=y,\ X=x)}{P(X=x)} \\ &= \frac{p(x, y)}{p_X(x)}, \end{aligned} \tag{5.2}$$

if $p_X(x) \neq 0$.

Note that the conditional pmf, as defined above, satisfies properties (p1)–(p3) of a pmf, discussed in Chapter 2. Rewriting the above definition another way, we have:

$$p(x, y) = p_X(x)p_{Y|X}(y|x) = p_Y(y)p_{X|Y}(x|y). \tag{5.3}$$

This is simply another form of the multiplication rule (of Chapter 1), and it gives us a way to compute the joint pmf whether or not X and Y are independent. *If X and Y are independent,* then from (5.3) and the definition of independence (in Chapter 2) we conclude that:

$$p_{Y|X}(y|x) = p_Y(y). \tag{5.4}$$

From (5.3) we also have the marginal probability:

$$p_Y(y) = \sum_{\text{all } x} p(x, y) = \sum_{\text{all } x} p_{Y|X}(y|x)p_X(x). \tag{5.5}$$

This is another form of the theorem of total probability, discussed in Chapter 1.

We can also define the conditional distribution function $F_{Y|X}(y|x)$ of a random variable Y, given a discrete random variable X by:

$$F_{Y|X}(y|x) = P(Y \leqslant y|X = x) = \frac{P(Y \leqslant y \text{ and } X = x)}{P(X = x)} \tag{5.6}$$

for all values of y and for all values of x such that $P(X = x) > 0$.

Definition (5.6) applies even for the case when Y is not discrete.

Note that the conditional distribution function can be obtained from the conditional pmf (in case Y is discrete):

$$F_{Y|X}(y|x) = \frac{\sum_{t \leqslant y} p(x, t)}{p_X(x)} = \sum_{t \leqslant y} p_{Y|X}(t|x). \tag{5.7}$$

Example 5.1

A computer center has two computer systems labeled A and B. Incoming jobs are independently routed to system A with probability p and to system B with probability $(1 - p)$. The number of jobs, X, arriving per unit time is Poisson distributed with parameter λ. Determine the distribution function of the number of jobs, Y, received by system A, per unit time.

Let us determine the conditional probability of the event $[Y = k]$ given that event $[X = n]$ has occurred. Note that routing of the n jobs can be thought of as a sequence of n independent Bernoulli trials. Hence, the conditional probability that $[Y = k]$ given $[X = n]$ is binomial with parameters n and p:

$$p_{Y|X}(k|n) = \begin{cases} P(Y = k \mid X = n) = \binom{n}{k} p^k (1 - p)^{n-k}, & 0 \leqslant k \leqslant n \\ 0, & \text{otherwise.} \end{cases}$$

Recalling that $P(X = n) = e^{-\lambda}\lambda^n/n!$ and using formula (5.5), we get:

$$p_Y(k) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= \frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!}$$

$$= \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda(1-p)}$$

(since the last sum is the Taylor series expansion of $e^{\lambda(1-p)}$)

$$= \frac{(\lambda p)^k e^{-\lambda p}}{k!}.$$

Thus, Y is Poisson distributed with parameter λp. For this reason we often say that the Poisson distribution is preserved under random selection. #

If X and Y are jointly continuous, then we define the conditional pdf of Y given X in a way analogous to the definition of the conditional pmf.

Definition (Conditional pdf). Let X and Y be continuous random variables with joint pdf $f(x, y)$. The conditional density $f_{Y|X}$ is defined by:

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \qquad \text{if } 0 < f_X(x) < \infty. \tag{5.8}$$

It can be easily verified that the function defined in (5.8) satisfies properties (f1) and (f2) of a pdf.

It follows from the definition of conditional density that:

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y)\, f_{X|Y}(x|y). \tag{5.9}$$

This is the continuous analog of the multiplication rule, MR, of Chapter 1. If X and Y are independent, then:

$$f(x, y) = f_X(x)\, f_Y(y),$$

which implies that:

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y having a joint density to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

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

$$= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x, y) = f_X(x) f_{Y|X}(y|x)$, we obtain:

$$f_{X|Y}(x|y) = \frac{f_X(x)\, f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x)\, f_{Y|X}(y|x)\, dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$P(a \leqslant Y \leqslant b \mid X = x) = \int_a^b f_{Y|X}(y|x)\, dy, \qquad a \leqslant b. \tag{5.13}$$

In particular, the conditional distribution function $F_{Y|X}(y|x)$ is *defined*, analogously to (5.6), as:

$$F_{Y|X}(y|x) = P(Y \leqslant y \mid X = x) = \frac{\int_{-\infty}^{y} f(x, t)\, dt}{f_X(x)}$$

$$= \int_{-\infty}^{y} f_{Y|X}(t|x)\, dt. \tag{5.14}$$

As motivation for Definition (5.14) we observe that:

$$F_{Y|X}(y|x) = \lim_{h \to 0} P(Y \leqslant y \mid x \leqslant X \leqslant x + h)$$

$$= \lim_{h \to 0} \frac{P(x \leqslant X \leqslant x + h \text{ and } Y \leqslant y)}{P(x \leqslant X \leqslant x + h)}$$

$$= \lim_{h \to 0} \frac{\int_x^{x+h} \int_{-\infty}^{y} f(s, t)\, dt\, ds}{\int_x^{x+h} f_X(s)\, ds}$$

$$= \lim_{h \to 0} \frac{h \int_{-\infty}^{y} f(x_1^*, t)\, dt}{h f_X(x_2^*)}$$

for some x_1^*, x_2^* with $x \leqslant x_1^*, x_2^* \leqslant x + h$

(by the mean value theorem of integrals)

$$= \lim_{h \to 0} \frac{\int_{-\infty}^{y} f(x_1^*, t)\,dt}{f_X(x_2^*)}$$

$$= \int_{-\infty}^{y} \frac{f(x, t)}{f_X(x)}\,dt$$

(since both x_1^* and x_2^* approach x as h approaches 0.)

$$= \int_{-\infty}^{y} f_{Y|X}(t|x)\,dt.$$

Example 5.2

Consider a series system of two independent components with the respective lifetime distributions $X \sim \text{EXP}(\lambda_1)$ and $Y \sim \text{EXP}(\lambda_2)$. We wish to determine the probability that component 2 is the cause of system failure. Let A denote the event that component 2 is the cause of system failure; then:

$$P(A) = P(X \geqslant Y).$$

To compute this probability, first consider the conditional distribution function:

$$F_{X|Y}(t|t) = P(X \leqslant t \,|\, Y = t) = F_X(t)$$

(by the independence of X and Y). Now by the continuous version of the theorem of total probability:

$$\begin{aligned} P(A) &= \int_0^\infty P(X \geqslant t \,|\, Y = t) f_Y(t)\,dt \\ &= \int_0^\infty [1 - F_X(t)]\, f_Y(t)\,dt \\ &= \int_0^\infty e^{-\lambda_1 t} \lambda_2 e^{-\lambda_2 t}\,dt \\ &= \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

This result generalizes to a series system of n independent components, each with a respective constant failure rate λ_j ($j = 1, 2, \ldots, n$). The probability that the jth component is the cause of system failure is given by:

$$\frac{\lambda_j}{\sum_{i=1}^{n} \lambda_i}. \tag{5.15}$$

#

Example 5.3 [BARL 1975]

Thus far in our reliability computations, we have considered failure mechanisms of components to be independent. We have derived the exponential lifetime distribution from a Poisson shock model. We now model the behavior of a system of two nonindependent components using a bivariate exponential distribution. Assume three independent Poisson shock sources. A shock from source 1 destroys component 1, and the time to the occurrence U_1 of such a shock is exponentially distributed with parameter λ_1, so that $P(U_1 > t) = e^{-\lambda_1 t}$. A shock from source 2 destroys component 2, and $P(U_2 > t) = e^{-\lambda_2 t}$. Finally, a shock from source 3 destroys both components and it occurs at random time U_{12}, so that $P(U_{12} > t) = e^{-\lambda_{12} t}$. Thus the lifetime X of component 1 satisfies:

$$X = \min\{U_1, U_{12}\}$$

and is exponentially distributed with parameter $\lambda_1 + \lambda_{12}$. The lifetime Y of component 2 is given by:

$$Y = \min\{U_2, U_{12}\}$$

and is exponentially distributed with parameter $\lambda_2 + \lambda_{12}$. Therefore:

$$f_X(x) = (\lambda_1 + \lambda_{12}) e^{-(\lambda_1 + \lambda_{12})x}, \qquad x > 0,$$

and

$$f_Y(y) = (\lambda_2 + \lambda_{12}) e^{-(\lambda_2 + \lambda_{12})y}, \qquad y > 0.$$

To compute the joint distribution function $F(x, y) = P(X \leqslant x,\ Y \leqslant y)$, we first compute:

$$\begin{aligned} R(x, y) &= P(X > x,\ Y > y) \\ &= P(\min\{U_1, U_{12}\} > x,\ \min\{U_2, U_{12}\} > y) \\ &= P(U_1 > x,\ U_{12} > \max\{x, y\},\ U_2 > y) \\ &= P(U_1 > x) P(U_{12} > \max\{x, y\}) P(U_2 > y) \\ &= e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} \max\{x, y\}}, \qquad x \geqslant 0,\ y \geqslant 0. \end{aligned}$$

This is true since U_1, U_2, and U_{12} are mutually independent. It is interesting to note that $R(x, y) \geqslant R_X(x)\, R_Y(y)$. Now $F(x, y)$ can be obtained using the relation (see Figure 5.1):

$$\begin{aligned} F(x, y) &= R(x, y) + F_X(x) + F_Y(y) - 1 \\ &= 1 + e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} \max\{x, y\}} - e^{-(\lambda_1 + \lambda_{12})x} - e^{-(\lambda_2 + \lambda_{12})y}. \end{aligned}$$

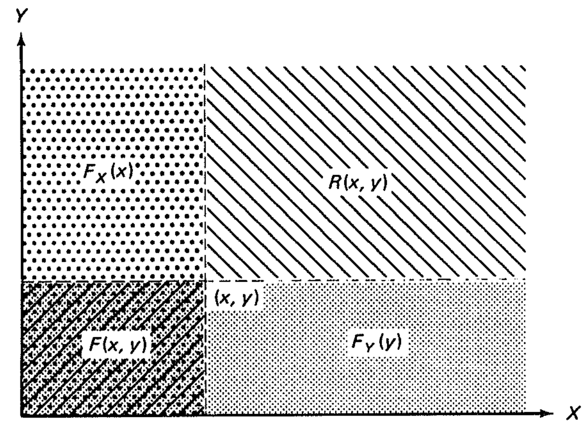

Figure 5.1 Illustration for $R(x, y) + F_x(x) + F_y(y) = 1 + F(x, y)$

In particular:

$$F(x, y) \neq F_X(x) F_Y(y)$$

since:

$$F_X(x) F_Y(y) = 1 - e^{-(\lambda_1+\lambda_{12})x} - e^{-(\lambda_2+\lambda_{12})y} + e^{-(\lambda_1+\lambda_{12})x-(\lambda_2+\lambda_{12})y}$$

Thus X and Y are indeed dependent random variables.

The joint density $f(x, y)$ may be obtained by taking partial derivatives:

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

$$= \begin{cases} \lambda_1(\lambda_2 + \lambda_{12}) e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} y}, & x \leqslant y, \\ \lambda_2(\lambda_1 + \lambda_{12}) e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} x}, & x > y, \end{cases}$$

and the conditional density:

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{\lambda_1 (\lambda_2 + \lambda_{12})}{\lambda_1 + \lambda_{12}} e^{-(\lambda_2+\lambda_{12})y+\lambda_{12}x}, & x \leqslant y, \\ \lambda_2 e^{-\lambda_2 y}, & x > y. \end{cases}$$

Once again, this confirms that X and Y are not independent. #

Problems

1. Consider again the problem of 1K RAM chips supplied by two semiconductor houses (problem 1, Section 3.6). Determine the conditional probability density of the lifetime X, given that the lifetime Y does not exceed 10^6 hours.

2. [MEND 1979] Consider the operation of an on-line file updating system. Let p_i be the probability that a transaction inserts a record into file i ($i = 1, 2, \ldots, n$), so that $\sum_{i=1}^{n} p_i = 1$. The record size (in bytes) of file i is a random variable denoted by Y_i. Determine:
(a) The average number of bytes added to file i per transaction.
(b) The variance of the number of bytes added to file i per transaction.

[*Hint:* You may define the Bernoulli random variable:

$$A_i = \begin{cases} 1, & \text{transaction updates file } i, \\ 0, & \text{otherwise,} \end{cases}$$

and let the random variable $V_i = A_i\, Y_i$ be the number of bytes added to file i in a transaction.]

3. X_1 and X_2 are independent random variables with Poisson distributions, having respective parameters α_1 and α_2. Show that the conditional pmf of X_1, given $X_1 + X_2$, $p_{X_1|X_1+X_2}(X_1 = x_1 | X_1 + X_2 = y)$, is binomial. Determine its parameters.

4. Let the execution times X and Y of two independent parallel processes be uniformly distributed over $(0, t_X)$ and $(0, t_Y)$, respectively, with $t_X \leqslant t_Y$. Find the probability that the former process finishes execution before the latter.

5.2 MIXTURE DISTRIBUTIONS

The definition of conditional density (and conditional pmf) can be naturally extended to the case where X is a discrete random variable and Y is a continuous random variable (or vice versa).

Example 5.4

Consider a computer system whose workload may be divided into r distinct classes. For job class i ($1 \leqslant i \leqslant r$), the CPU service time is exponentially distributed with parameter λ_i. Let Y denote the service time of a job and let X be the job class. Then:

$$f_{Y|X}(y|i) = \lambda_i e^{-\lambda_i y}, \qquad y > 0.$$

Now let α_i ($\geqslant 0$) be the probability that a randomly chosen job belongs to class i; that is:

$$p_X(i) = \alpha_i, \qquad \sum_{i=1}^{r} \alpha_i = 1.$$

Then the joint density is:

$$f(i, y) = f_{Y|X}(y|i) p_X(i)$$

$$= \alpha_i\, \lambda_i\, e^{-\lambda_i y}, \qquad y > 0,$$

and the marginal density is:

$$
\begin{aligned}
f_Y(y) &= \sum_{i=1}^{r} f(i, y) \\
&= \sum_{i=1}^{r} \alpha_i \, f_{Y|X}(y|i) \\
&= \sum_{i=1}^{r} \alpha_i \, \lambda_i \, e^{-\lambda_i y}, \qquad y > 0.
\end{aligned}
$$

Thus Y has an r-stage hyperexponential distribution, denoted by a set of parallel exponential servers as in Figure 5.2. #

Of course, the conditional distribution of Y does not have to be exponential. In general, if we let:

$$f_{Y|X}(y|i) = f_i(y) = f_{Y_i}(y)$$

and

$$F_{Y|X}(y|i) = F_i(y),$$

then we have the unconditional pdf of Y:

$$f_Y(y) = \sum_{i=1}^{r} \alpha_i f_i(y), \tag{5.16}$$

and the unconditional CDF of Y:

$$F_Y(y) = \sum_{i=1}^{r} \alpha_i F_i(y). \tag{5.17}$$

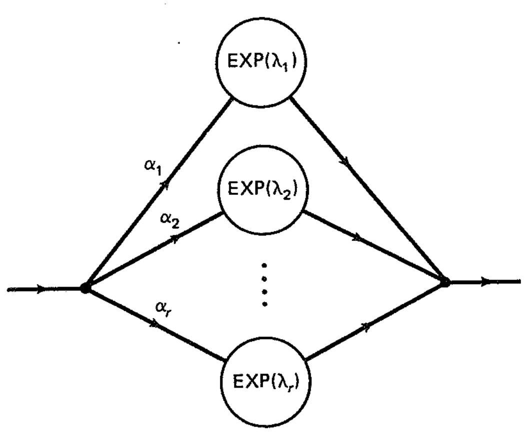

Figure 5.2 The hyperexponential distribution as a set of parallel exponential stages

Taking Laplace transforms on both sides of (5.16), we also have:

$$L_Y(s) = \sum_{i=1}^{r} \alpha_i L_{Y_i}(s). \tag{5.18}$$

Finally applying the definitions of the mean and higher moments to (5.16), we have:

$$E[Y] = \sum_{i=1}^{r} \alpha_i E[Y_i], \tag{5.19}$$

$$E[Y^k] = \sum_{i=1}^{r} \alpha_i E[Y_i^k]. \tag{5.20}$$

Such mixture distributions often arise in a number of reliability situations. For example, suppose a manufacturer produces α_i fraction of a certain product in assembly line i, and the life length of a unit produced in assembly line i has a distribution F_i. Now if the outputs of the assembly lines are merged, then a randomly chosen unit from the merged stream will possess the life-length distribution given by equation (5.17) above.

Example 5.5

Assume that in a mixture of two groups, one group consists of components in the chance-failure period (with constant hazard rate λ_1) and the other of aging items (modeled by an r-stage Erlang lifetime distribution with parameter λ_2). If α is the fraction of group-one components, then the distribution of the lifetime Y of a component from the merged stream is given by:

$$F_Y(y) = \alpha(1 - e^{-\lambda_1 y}) + (1-\alpha)(1 - \sum_{k=0}^{r-1} \frac{(\lambda_2 y)^k}{k!} e^{-\lambda_2 y})$$

and

$$f_Y(y) = \alpha\lambda_1 e^{-\lambda_1 y} + (1-\alpha)\frac{\lambda_2^r y^{r-1}}{(r-1)!} e^{-\lambda_2 y}.$$

This density and the corresponding hazard rate are shown in Figures 5.3 and 5.4. Note that this distribution has a nonmonotonic hazard function. #

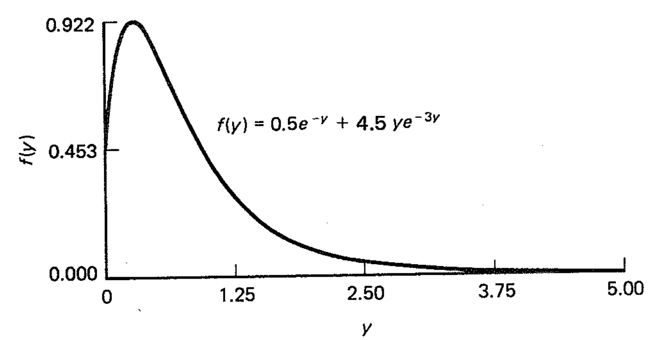

Figure 5.3 The pdf of a mixture of exponential and Erlang

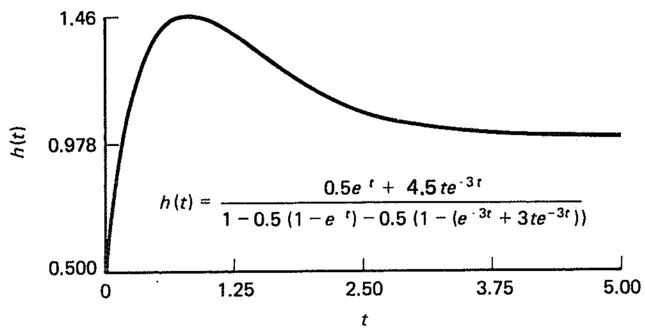

Figure 5.4 Hazard rate of a mixture of exponental and Erlang

More generally, the distributions being mixed may be uncountably infinite in number; that is, X may be a continuous random variable. For instance, the lifetime of a product may depend upon the amount X of impurity present in the raw material. Let the conditional distribution of the lifetime Y be given by:

$$F_{Y|X}(y|x) = G_x(y) = \int_{-\infty}^{y} \frac{f(x,t)\,dt}{f_X(x)},$$

where the impurity X has a density function $f_X(x)$. Then the resultant lifetime distribution F_Y is given by:

$$F_Y(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f(x,t)\;dt\;dx = \int_{-\infty}^{\infty} f_X(x)\;G_x(y)\;dx.$$

In the next example we let Y be discrete and X continuous.

Example 5.6 [CLAR 1970]

Let X be the service time of a customer in a computing center and let it be exponentially distributed with parameter μ, so that:

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

Let the number of customers arriving in the interval $(0, t]$ be Poisson distributed with parameter λt. Finally, let Y be the number of customers arriving while one is being served.

If we fix the value of X to be x, the Poisson arrival assumption can be used to obtain the conditional pmf of Y given $[X = x]$:

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= e^{-\lambda x}\frac{(\lambda x)^y}{y!}, \qquad y = 0, 1, 2, \ldots. \end{aligned}$$

The joint probability density function of X and Y is then given by:

$$\begin{aligned} f(x,y) &= f_X(x)p_{Y|X}(y|x) \\ &= \frac{\mu e^{-(\lambda+\mu)x}(\lambda x)^y}{y!}, \qquad y = 0, 1, 2, \ldots; \qquad x > 0. \end{aligned}$$

The unconditional (or marginal) pmf of Y can now be obtained by integration:

$$\begin{aligned} p_Y(y) &= P(Y = y) \\ &= \int_0^{\infty} f(x,y)\;dx \\ &= \frac{\mu}{y!}\int_0^{\infty} e^{-(\lambda+\mu)x}(\lambda x)^y\,dx. \end{aligned}$$

Substituting $(\lambda + \mu)x = w$, we get:

$$\begin{aligned} p_Y(y) &= \frac{\mu\lambda^y}{y!(\lambda+\mu)^{y+1}}\int_0^{\infty} e^{-w}w^y\,dw \\ &= \frac{\mu\lambda^y y!}{y!(\lambda+\mu)^{y+1}} \end{aligned}$$

[since the last integral is equal to $\Gamma(y + 1) = y!$ by formulas (3.26) and (3.24)]. Thus:

$$\begin{aligned} p_Y(y) &= \frac{\rho^y}{(1+\rho)^{y+1}}, \qquad \text{where } \rho = \frac{\lambda}{\mu}, \\ &= \left(\frac{\rho}{1+\rho}\right)^y \frac{1}{1+\rho}, \qquad y = 0, 1, 2, \ldots. \end{aligned}$$

Thus Y has a modified geometric distribution with parameter $\frac{1}{1+\rho}$; hence the expected value is:

$$E[Y] = \frac{\frac{\rho}{1+\rho}}{\frac{1}{1+\rho}} = \rho = \frac{\lambda}{\mu}.$$

This is an example of the so-called $M/M/1$ queuing system to be discussed in a later chapter. We may argue that an undesirable backlog of customers will not occur provided the average number of customers arriving in the interval representing the service time of a typical customer is less than 1. In other words, the queuing system will remain stable provided:

$$E[Y] = \rho < 1 \quad \text{or} \quad \lambda < \mu.$$

This last condition says that the rate at which customers arrive is less than the rate at which work can be completed. #

Example 5.7 [GAVE 1973]

Consider a series system with n components, each with a lifetime distribution function $G(t)$ and density $g(t)$. Because of options offered, the number of components, Y, in a specific system is a random variable. Let X denote the lifetime of the series system. Then clearly:

$$F_{X|Y}(t|n) = 1 - [1 - G(t)]^n, \quad n = 0, 1, 2, \ldots, \quad t > 0,$$

$$f_{X|Y}(t|n) = n[1 - G(t)]^{n-1} g(t), \quad n = 0, 1, 2, \ldots, \quad t > 0.$$

Assume that the number of components, Y, has a Poisson distribution with parameter α. Then:

$$p_Y(n) = e^{-\alpha} \frac{\alpha^n}{n!}, \qquad \alpha > 0, \quad n = 0, 1, 2, \ldots,$$

and the joint density is:

$$f(t, n) = f_{X|Y}(t \mid n)\, p_Y(n)$$

$$= \begin{cases} e^{-\alpha} \dfrac{\alpha^n}{(n-1)!} [1 - G(t)]^{n-1} g(t), & t > 0, \quad n = 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

We can now determine the marginal density:

$$f_X(t) = \sum_{n=1}^{\infty} [1 - G(t)]^{n-1} g(t)\, e^{-\alpha} \frac{\alpha^n}{(n-1)!}.$$

The system reliability is given by:

$$R_X(t) = P(X > t)$$

$$= \sum_{n=0}^{\infty} [1 - F_{X|Y}(t|n)]\, p_Y(n)$$

(by the theorem of total probability)

$$= \sum_{n=0}^{\infty} [1 - G(t)]^n e^{-\alpha} \frac{\alpha^n}{n!}$$

$$= e^{-\alpha} \sum_{n=0}^{\infty} \frac{\{\alpha[1 - G(t)]\}^n}{n!}$$

$$= e^{-\alpha} e^{\alpha[1 - G(t)]}$$

$$= e^{-\alpha G(t)}.$$

Now suppose that the system has survived until time t. We are interested in computing the conditional pmf of the number of components Y it has:

$$P(Y = n | X > t) = \frac{P(X > t,\ Y = n)}{P(X > t)}$$

$$= \frac{[1 - F_{X|Y}(t|n)]\, p_Y(n)}{R_X(t)}$$

$$= e^{-\alpha[1-G(t)]} \frac{[\alpha(1 - G(t))]^n}{n!}.$$

Thus the conditional pmf of Y, given that no failure has occurred until time t, is Poisson with parameter $\alpha[1 - G(t)]$. Since $G(t)$ is a monotonically increasing function of t, the parameter of the Poisson distribution decreases with t. In other words, the longer the system survives, the greater is the evidence that it has a small number of components. #

Yet another case of a mixture distribution occurs when we mix two distributions, one discrete and the other continuous. The mixture distribution then represents a mixed random variable (see the distribution (3.2) in Chapter 3).

Problems

1. Consider the **if** statement:

   if B **then** S_1 **else** S_2.

   Let the random variables X_1 and X_2 respectively, denote the execution times of the statement groups S_1 and S_2. Assuming the probability that the Boolean expression $B =$ true is p, derive an expression for the distribution of the total execution time X of the **if** statement. Compute $E[X]$ and Var $[X]$ as functions of the means and variances of X_1 and X_2. Generalize your results to a **case** statement with k clauses.

2. Describe a method of generating a random deviate of a two-stage hyperexponential distribution.

3. One of the inputs to a certain program is a random variable whose value is a non-negative real number; call it Λ. The probability density function of Λ is given by:

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

   Conditioned on $\Lambda = \lambda$, the execution time of the program is an exponentially distributed random variable with parameter λ. Compute the distribution function of the program execution time X.

5.3. CONDITIONAL EXPECTATION

If X and Y are continuous random variables, then the conditional density $f_{Y|X}$ is given by formula (5.8). Since $f_{Y|X}$ is a density of a continuous random variable, we can talk about its various moments. Its mean (if it exists) is called the **conditional expectation** of Y given $[X = x]$ and will be denoted by $E[Y|X = x]$ or $E[Y|x]$. Thus:

$$E[Y|x] = \int_{-\infty}^{\infty} y\, f(y|x)\, dy$$
$$= \frac{\int_{-\infty}^{\infty} y\, f(x, y)\, dy}{f_X(x)}, \qquad 0 < f_X(x) < \infty. \tag{5.21}$$

We will define $E[Y|x] = 0$ elsewhere. The quantity $m(x) = E[Y|x]$, considered as a function of x, is known as the **regression function** of Y on X.

In case the random variables X and Y are discrete, the conditional expectation $E[Y|x]$ is defined as:

$$E[Y|X = x] = \sum_y yP(Y = y|X = x)$$
$$= \sum_y y p_{Y|X}(y|x). \tag{5.22}$$

Similar definitions can be given in mixed situations. These definitions can be easily generalized to define the conditional expectation of a function $\phi(Y)$:

$$E[\phi(Y)|X = x] = \begin{cases} \int_{-\infty}^{\infty} \phi(y) f_{Y|X}(y|x)\, dy, & \text{if } Y \text{ is continuous,} \\ \sum_i \phi(y_i) p_{Y|X}(y_i|x), & \text{if } Y \text{ is discrete.} \end{cases} \tag{5.23}$$

As a special case of definition (5.23), we have the conditional kth moment of Y, $E[Y^k|X = x]$, and the conditional moment generating function of Y, $M_{Y|X}(\theta|x) = E[e^{\theta Y}|X = x]$. From the conditional moment generating function we also obtain the definition of the conditional Laplace transform, $L_{Y|X}(s|x) = E[e^{-sY}|X = x]$, and the conditional PGF, $G_{Y|X}(z|x) = E[z^Y|X = x]$.

We may take the expectation of the regression function $m(X)$ to obtain the unconditional expectation of Y:

$$E[m(X)] = E[E[Y|X]] = E[Y]$$

that is to say:

$$E[Y] = \begin{cases} \sum_x E[Y|X = x] p_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} E[Y|X = x] f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{5.24}$$

This last formula, known as the **theorem of total expectation**, is found to be quite useful in practice. A similar result called the theorem of total moments is given by:

$$E[Y^k] = \begin{cases} \sum_x E[Y^k|X = x] p_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} E[Y^k|X = x] f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{5.25}$$

Similarly, we have theorems of total transforms. For example, the theorem of total Laplace transform is (assuming Y is a nonnegative continuous random variable):

$$L_Y(s) = \begin{cases} \sum_x L_{Y|X}(s|x) p_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} L_{Y|X}(s|x) f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{5.26}$$

Example 5.8

Consider the Example 5.4 of r job classes in a computer system. Since:

$$f_{Y|X}(y|i) = \lambda_i e^{-\lambda_i y},$$

then:

$$E[Y|X = i] = \frac{1}{\lambda_i}$$

and

$$E[Y^2|X = i] = \frac{2}{\lambda_i^2}.$$

Then by the theorem of total expectation:

$$E[Y] = \sum_{i=1}^{r} \frac{\alpha_i}{\lambda_i}$$

and

$$E[Y^2] = \sum_{i=1}^{r} \frac{2\alpha_i}{\lambda_i^2}.$$

Then:

$$\text{Var}\,[Y] = \sum_{i=1}^{r} \frac{2\alpha_i}{\lambda_i^2} - \left(\sum_{i=1}^{r} \frac{\alpha_i}{\lambda_i}\right)^2.$$

#

Example 5.9

Refer to Example 5.7 of a series system with a random number of components, where:

$$f_{X|Y}(t|n) = n\,[1 - G(t)]^{n-1}\, g(t), \qquad t > 0.$$

Let:

$$G(t) = 1 - e^{-\lambda t}, \qquad \lambda > 0,\ t \geqslant 0.$$

Then:

$$\begin{aligned} f_{X|Y}(t|n) &= ne^{-\lambda(n-1)t}\lambda e^{-\lambda t} \\ &= n\lambda e^{-n\lambda t}, \end{aligned}$$

which is the exponential pdf with parameter $n\lambda$. It follows that:

$$E[X|Y=n] = \frac{1}{n\lambda}$$

and

$$E[X] = \sum_{n=1}^{\infty} \frac{1}{n\lambda} e^{-\alpha} \frac{\alpha^n}{n!}. \qquad \#$$

Example 5.10

(Analysis of uniform hashing) [KNUT 1973b]. A popular method of storing tables for fast searching is known as **hashing.** The table has M entries indexed from 0 to $M-1$. Given a search key k, an application of the hash function h produces an index, $h(k)$, into the table, where we generally expect to find the required entry. Since there are distinct keys $k_i \neq k_j$ that hash to the same value $h(k_i) = h(k_j)$, a situation known as collision, we have to derive some method for producing secondary indices for search.

Assume that k entries out of M in the table are currently occupied. As a consequence of the assumption that h distributes values uniformly over the table, all $\binom{M}{k}$ possible configurations are equally likely. Let the random variable X denote the number of probes necessary to insert the next item in the table, and let Y denote the number of occupied entries in the table. For a given number of occupied entries $Y = k$, if the number of probes is equal to r, then $(r-1)$ given cells are known to be occupied and the last inspected cell is known to be unoccupied. Out of the remaining $M - r$ cells, $(k - r + 1)$ can be occupied in $\binom{M-r}{k-r+1}$ ways. Therefore:

$$\begin{aligned} P(X = r|Y = k) &= p_{X|Y}(r|k) \\ &= \frac{\binom{M-r}{k-r+1}}{\binom{M}{k}}, \qquad 1 \leqslant r \leqslant M. \end{aligned} \tag{5.27}$$

This implies that:

$$\begin{aligned} E[X|Y=k] &= \sum_{r=1}^{M} r p_{X|Y}(r|k) \\ &= \sum_{r=1}^{M} (M+1)p_{X|Y}(r|k) - \sum_{r=1}^{M} (M+1-r)p_{X|Y}(r|k). \end{aligned}$$

Now since $p_{X|Y}$ is a pmf, the first sum on the right-hand side equals $M+1$. We substitute expression (5.27) in the second sum to obtain:

$$\begin{aligned} E[X|Y=k] &= (M+1) - \sum_{r=1}^{M} (M+1-r) \frac{\binom{M-r}{k-r+1}}{\binom{M}{k}} \\ &= (M+1) - \sum_{r=1}^{M} \frac{(M+1-r)\,(M-r)!}{(k-r+1)!(M-k-1)!\cdot\binom{M}{k}} \\ &= (M+1) - \sum_{r=1}^{M} \frac{(M-r+1)!(M-k)}{(k-r+1)!(M-k)!\cdot\binom{M}{k}} \\ &= (M+1) - \sum_{r=1}^{M} \frac{(M-k)\binom{M-r+1}{M-k}}{\binom{M}{k}} \end{aligned}$$

Now the sum:

$$\sum_{r=1}^{M} \binom{M-r+1}{M-k} = \sum_{i=1}^{M} \binom{i}{M-k} = \sum_{i=0}^{M} \binom{i}{M-k} = \binom{M+1}{M-k+1}$$

(using formula (11) from [KNUT 1973a, p. 54]).

After substitution and simplification, we have:

$$E[X|Y=k] = \frac{M+1}{M-k+1}, \qquad 0 \leqslant k \leqslant M-1.$$

Now, assuming that Y is uniformly distributed over $0 \leqslant k < N \leqslant M$, we get:

$$\begin{aligned} p_Y(k) &= \frac{1}{N}, \\ E[X] &= \sum_{k=0}^{N-1} \frac{1}{N} E[X|Y=k] \\ &= \frac{M+1}{N}\left(\frac{1}{M+1} + \frac{1}{M} + \cdots + \frac{1}{M-N+2}\right) \\ &= \frac{M+1}{N}(H_{M+1} - H_{M-N+1}) \\ &\simeq \frac{1}{\alpha}\ln\frac{1}{1-\alpha}, \end{aligned}$$

where $\alpha = N/(M+1)$, the table occupancy factor. This is the expected number of probes necessary to locate an entry in the table, provided the search is successful.

Note that if the table occupancy factor is low (below 80 percent), the average number of probes is nearly equal to 1. In other words, where applicable, this is an efficient method of search. #

Problems

*1. The notion of a recovery block was introduced by Randell [RAND 1975] to facilitate software fault tolerance in presence of software design errors. This construct provides a "normal" algorithm to perform the required function together with an acceptance test of its results. If the test results are unsatisfactory then an alternative algorithm is executed. Assume that X is the execution time of the normal algorithm and Y is the execution time of the alternative algorithm. Assume p is the probability that the results of the normal execution satisfy the acceptance test. Determine the distribution function of the total execution time T of the recovery block assuming that X and Y are uniformly distributed over (a, b). Repeat, assuming that X and Y are exponentially distributed with parameters λ_1 and λ_2, respectively. In each case determine $E[T]$, Var $[T]$, and in the latter case $L_T(s)$.

2. Consider the flowchart model of fault recovery in a computer system (such as Bell System's Electronic Switching system) as shown in Figure 5.P.1. Assuming that the random variables D, L, R, M_D, and M_L are exponentially distributed with

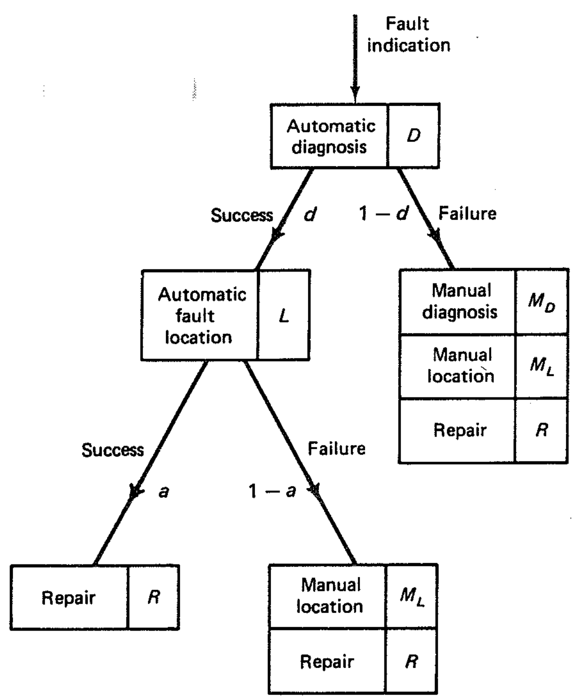

Figure 5.P.1 Flowchart of automatic fault recovery

parameters δ, λ, ρ, μ_1 and μ_2, determine the distribution function of the random variable X, denoting the total recovery time. Also compute $E[X]$ and Var $[X]$.

3. (Linear searching problem) We are given an unordered list with n distinct keys. We are searching linearly for a specific key that has a probability p of being present in the list (and probability q of being absent). Given that the key is in the list, the probability of its being in position i is $1/n$, $i = 1, 2, \ldots, n$. Compute the expected number of comparisons for:
   (a) A successful search.
   (b) An unsuccessful search.
   (c) A search (unconditionally).

4. [MEND 1979] Let V_1 be the random variable denoting the length (in bytes) of a source program. Let p be the probability of successful compilation of the program. Let V_2 be the length of the compiled code (load module). Clearly, V_2 and V_1 will not be independent. Assume $V_2 = BV_1$ where B is a random variable, and B and V_1 are independent. After the compilation, the load module will be entered into a library. Let X be the length of a request for space allocation to the library due to the above source program. Determine $E[X]$ and Var $[X]$ in terms of $E[B]$, $E[V_1]$, Var $[B]$, and Var $[V_1]$.

5.4 IMPERFECT FAULT COVERAGE AND RELIABILITY

Reliability models of systems with dynamic redundancy (e.g., standby redundancy, hybrid NMR) developed earlier are not very realistic. It has been demonstrated that the reliability of such systems depends strongly on the effectiveness of recovery mechanisms. In particular, it may be impossible to switch in an existing spare module and thus recover from a failure. Faults such as these are said to be "uncovered", and the probability that a given fault belongs to this class is denoted by $1 - c$, where c denotes the probability of occurrence of "covered" faults, and is known as the **coverage parameter** [BOUR 1969].

Example 5.11

Let X denote the lifetime of a system with two units, one active and the other a cold standby spare. The failure rate of an active unit is λ, and a cold spare does not fail. Let Y be the indicator random variable of the fault class; that is:

$$Y = 0 \text{ if the fault is uncovered,}$$
$$Y = 1 \text{ if the fault is covered.}$$

Then:

$$p_Y(0) = 1 - c \quad \text{and} \quad p_Y(1) = c.$$

To compute the MTTF of this system, we first obtain the conditional expectation of X given Y by noting that if an uncovered fault occurs, the mean life of the system equals the mean life of the initially active unit. That is: