We defer the proof of Theorem 5.3 to the end of this section, choosing to concentrate now on its implications. As an immediate consequence, e.g., in the context of the earlier car accident vignette, we see that the individual cars need not all have the same accident probabilities in order for the Poisson approximation to apply.

5.2.1 The Law of Rare Events and the Poisson Process

Consider events occurring along the positive axis $[0, \infty)$ in the manner shown in Figure 5.2. Concrete examples of such processes are the time points of the X-ray emissions of a substance undergoing radioactive decay, the instances of telephone calls originating in a given locality, the occurrence of accidents at a certain intersection, the location of faults or defects along the length of a fiber or filament, and the successive arrival times of customers for service.

Let $N((a,b])$ denote the number of events that occur during the interval $(a,b]$. That is, if $t_1 < t_2 < t_3 < \cdots$ denote the times (or locations, etc.) of successive events, then $N((a,b])$ is the number of values t_i for which $a < t_i \le b$.

We make the following postulates:

1. The numbers of events happening in disjoint intervals are independent random variables. That is, for every integer $m = 2, 3, \ldots$ and time points $t_0 = 0 < t_1 < t_2 < \cdots < t_m$, the random variables

$$N((t_0,t_1]), N((t_1,t_2]), \ldots, N((t_{m-1},t_m])$$

are independent.

2. For any time t and positive number h, the probability distribution of $N((t,t+h])$, the number of events occurring between time t and $t+h$, depends only on the interval length h and not on the time t.
3. There is a positive constant λ for which the probability of at least one event happening in a time interval of length h is

$$\Pr\{N((t,t+h]) \ge 1\} = \lambda h + o(h) \quad \text{as } h \downarrow 0.$$

(Conforming to a common notation, here $o(h)$ as $h \downarrow 0$ stands for a general and unspecified remainder term for which $o(h)/h \to 0$ as $h \downarrow 0$. That is, a remainder term of smaller order than h as h vanishes.)

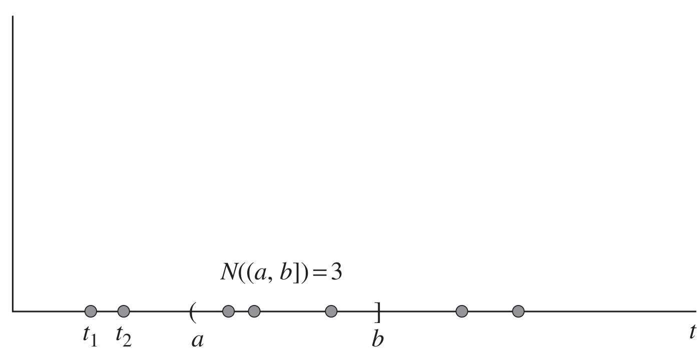

Figure 5.2 A Poisson point process.

4. The probability of two or more events occurring in an interval of length h is $o(h)$, or

$$\Pr\{N((t,t+h]) \geq 2\} = o(h), \quad h \downarrow 0.$$

Postulate 3 is a specific formulation of the notion that events are rare. Postulate 4 is tantamount to excluding the possibility of the simultaneous occurrence of two or more events. In the presence of Postulates 1 and 2, Postulates 3 and 4 are equivalent to the apparently weaker assumption that events occur singly and discretely, with only a finite number in any finite interval. In the concrete illustrations cited earlier, this requirement is usually satisfied.

Disjoint intervals are independent by 1, and 2 asserts that the distribution of $N((s,t])$ is the same as that of $N((0,t-s])$. Therefore, to describe the probability law of the system, it suffices to determine the probability distribution of $N((0,t])$ for an arbitrary value of t. Let

$$P_k(t) = \Pr\{N((0,t]) = k\}.$$

We will show that Postulates 1 through 4 require that $P_k(t)$ be the Poisson distribution

$$P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad \text{for } k = 0, 1, \ldots. \tag{5.8}$$

To establish (5.8), divide the interval $(0,t]$ into n subintervals of equal length $h = t/n$, and let

$$\epsilon_i = \begin{cases} 1 & \text{if there is at least one event in the interval } ((i-1)t/n, it/n], \\ 0 & \text{otherwise.} \end{cases}$$

Then, $S_n = \epsilon_1 + \cdots + \epsilon_n$ counts the total number of subintervals that contain at least one event, and

$$p_i = \Pr\{\epsilon_i = 1\} = \lambda t/n + o(t/n) \tag{5.9}$$

according to Postulate 3. Upon applying (5.7), we see that

$$\begin{aligned} \left|\Pr\{S_n = k\} - \frac{\mu^k e^{-\mu}}{k!}\right| &\leq n[\lambda t/n + o(t/n)]^2 \\ &= \frac{(\lambda t)^2}{n} + 2\lambda t o\left(\frac{t}{n}\right) + no\left(\frac{t}{n}\right)^2, \end{aligned}$$

where

$$\mu = \sum_{i=1}^{n} p_i = \lambda t + no(t/n). \tag{5.10}$$

Because $o(h) = o(t/n)$ is a term of order smaller than $h = t/n$ for large n, it follows that

$$no(t/n) = t\frac{o(t/n)}{t/n} = t\frac{o(h)}{h}$$

vanishes for arbitrarily large n. Passing to the limit as $n \to \infty$, then, we deduce that

$$\lim_{n\to\infty} \Pr\{S_n = k\} = \frac{\mu^k e^{-\mu}}{k!}, \quad \text{with } \mu = \lambda t.$$

To complete the demonstration, we need only show that

$$\lim_{n\to\infty} \Pr\{S_n = k\} = \Pr\{N((0,t]) = k\} = P_k(t).$$

But S_n differs from $N((0,t])$ only if at least one of the subintervals contains two or more events, and Postulate 4 precludes this because

$$\begin{aligned}
|P_k(t) - \Pr\{S_n = k\}| &\le \Pr\{N((0,t]) \ne S_n\} \\
&\le \sum_{i=1}^{n} \Pr\left\{N\left(\left(\frac{(i-1)t}{n}, \frac{it}{n}\right]\right) \ge 2\right\} \\
&\le no(t/n) \quad \text{(by Postulate 4)} \\
&\to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

By making n arbitrarily large, or equivalently, by dividing the interval $(0,t]$ into arbitrarily small subintervals, we see that it must be the case that

$$\Pr\{N((0,t]) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad \text{for } k \ge 0,$$

and Postulates 1 through 4 imply the Poisson distribution.

Postulates 1 through 4 arise as physically plausible assumptions in many circumstances of stochastic modeling. The postulates seem rather weak. Surprisingly, they are sufficiently strong to force the Poisson behavior just derived. This motivates the following definition.

Definition Let $N((s,t])$ be a random variable counting the number of events occurring in an interval $(s,t]$. Then, $N((s,t])$ is a Poisson point process of intensity $\lambda > 0$ if

1. for every $m = 2, 3, \ldots$ and distinct time points $t_0 = 0 < t_1 < t_2 < \cdots < t_m$, the random variables

$$N((t_0, t_1]), N((t_1, t_2]), \ldots, N((t_{m-1}, t_m])$$

are independent; and

2. for any times $s < t$ the random variable $N((s,t])$ has the Poisson distribution

$$\Pr\{N((s,t]) = k\} = \frac{[\lambda(t-s)]^k e^{-\lambda(t-s)}}{k!}, \quad k = 0, 1, \ldots.$$

Poisson point processes often arise in a form where the time parameter is replaced by a suitable spatial parameter. The following formal example illustrates this vein of ideas. Consider an array of points distributed in a space E (E is a Euclidean space of dimension $d \geq 1$). Let $N(A)$ denote the number of points (finite or infinite) contained in the region A of E. We postulate that $N(A)$ is a random variable. The collection $\{N(A)\}$ of random variables, where A varies over all possible subsets of E, is said to be a homogeneous Poisson process if the following assumptions are fulfilled:

1. The numbers of points in nonoverlapping regions are independent random variables.
2. For any region A of finite volume, $N(A)$ is Poisson distributed with mean $\lambda|A|$, where $|A|$ is the volume of A. The parameter λ is fixed and measures in a sense the intensity component of the distribution, which is independent of the size or shape. Spatial Poisson processes arise in considering such phenomena as the distribution of stars or galaxies in space, the spatial distribution of plants and animals, and the spatial distribution of bacteria on a slide. These ideas and concepts will be further studied in Section 5.5.

5.2.2 Proof of Theorem 5.3

First, some notation. Let $\epsilon(p)$ denote a Bernoulli random variable with success probability p, and let $X(\theta)$ be a Poisson distributed random variable with parameter θ. We are given probabilities $p_1, \ldots, p_n$ and let $\mu = p_1 + \cdots + p_n$. With $\epsilon(p_1), \ldots, \epsilon(p_n)$ assumed to be independent, we have $S_n = \epsilon(p_1) + \cdots + \epsilon(p_n)$, and according to Theorem 5.1, we may write $X(\mu)$ as the sum of independent Poisson distributed random variables in the form $X(\mu) = X(p_1) + \cdots + X(p_n)$. We are asked to compare $\Pr\{S_n = k\}$ with $\Pr\{X(\mu) = k\}$, and, as a first step, we observe that if S_n and $X(\mu)$ are unequal, then at least one of the pairs $\epsilon(p_k)$ and $X(p_k)$ must differ, whence

$$|\Pr\{S_n = k\} - \Pr\{X(\mu) = k\}| \leq \sum_{k=1}^{n} \Pr\{\epsilon(p_k) \neq X(p_k)\}. \tag{5.11}$$

As the second step, observe that the quantities that are compared on the left of (5.11) are the marginal distributions of S_n and $X(\mu)$, while the bound on the right is a joint probability. This leaves us free to choose the joint distribution that makes our task the easiest. That is, we are free to specify the joint distribution of each $\epsilon(p_k)$ and $X(p_k)$, as we please, provided only that the marginal distributions are Bernoulli and Poisson, respectively.

To complete the proof, we need to show that $\Pr\{\epsilon(p) \neq X(p)\} \leq p^2$ for some Bernoulli random variable $\epsilon(p)$ and Poisson random variable $X(p)$, since this reduces the right side of (5.11) to that of (5.7). Equivalently, we want to show that $1 - p^2 \leq \Pr\{\epsilon(p) = X(p)\} = \Pr\{\epsilon(p) = X(p) = 0\} + \Pr\{\epsilon(p) = X(p) = 1\}$, and we are free to choose the joint distribution, provided that the marginal distributions are correct.