The binomial and geometric distributions describe two random variables associated with the same process. The binomial distribution describes the number of "successes" in a given number of trials, and the geometric distribution describes the time until the first success. The exponential probability density function describes the **time** of the first event when events occur at a constant rate. We now study the **Poisson distribution**, the probability distribution for the number of events that occur during a fixed time interval when events occur at a constant rate, as in the exponential distribution. The underlying process generating events is called the **Poisson process.** The following chart relates these four fundamental probability distributions.

count	binomial	Poisson
waiting time	geometric	exponential

Consider a cell floating around in a medium with many molecules that bind permanently to receptors on the cell. Suppose that binding events occur at rate λ. We wish to find the probability distribution for the random variable N that gives the number of molecules binding to the cell during a given interval of time.

The **Poisson process** describes events that obey the following assumptions.

Events occur at a constant probabilistic rate λ.

Events are independent.

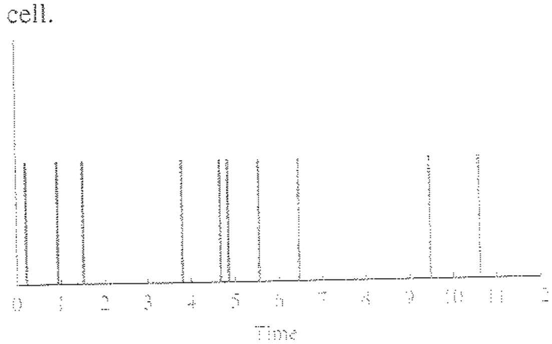
These conditions are similar to those for the process underlying the binomial distribution, where a success occurs independently and with equal probability in each trial, but here events can occur at any time.

A Simulation of the Poisson Process with $\lambda = 1.0$

A simulation of this process with $\lambda = 1.0$ produced the following results after running for 12 s.

1	0.219032	6	4.84409
2	0.929085	7	5.54173
3	1.50534	8	6.45588
4	3.79879	9	9.47364
5	4.66062	10	10.6027

Each black spike in Figure 7.7.47 indicates black when a molecule bound to the cell.



A simulation of the Poisson process

The waiting times between events in the Poisson distribution follow the exponential distribution with parameter λ. The Poisson process can be simulated efficiently by choosing a series of random numbers that obey the appropriate exponential distribution as the waiting times and then adding them to find the times of the events.

Example 7.7.2 Waiting Times Between Events in the Poisson Process

The data in Example 7.7.1 obey the Poisson process with $\lambda = 1.0$. The waiting times between events (after the clock is started at $t = 0$) are found as the differences between the times when the events occurred.

Event	Time	Event	Time
1	0.219032	6	0.18347
2	0.71005	7	0.69764
3	0.57625	8	0.91415
4	2.29345	9	3.01776
5	0.86183	10	1.12906

These waiting times come from an exponential distribution with rate parameter $\lambda = 1.0$.

Example 7.7.3 Summarizing a Simulation of the Poisson Process

We can summarize the results in Example 7.7.1 by asking how many molecules bound during each of the twelve 1 s intervals: Two bound during the first second (between times $t = 0.0$ and $t = 1.0$), one during the next second (between $t = 1.0$ and $t = 2.0$),

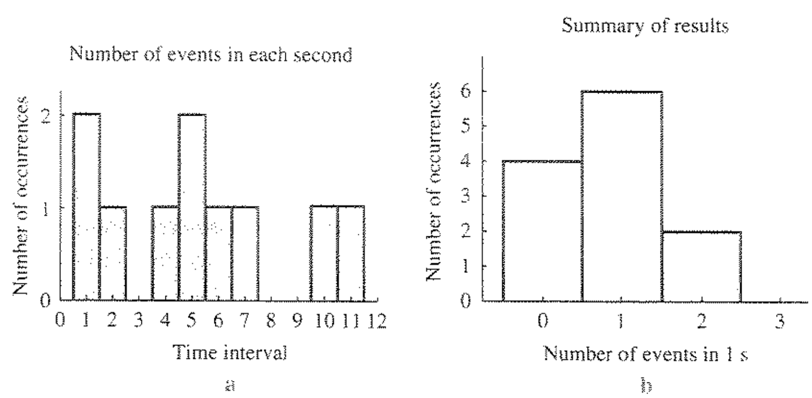


FIGURE 7.7.48
Summaries of a simulation of the Poisson process

none during the third second (between times $t = 2.0$ and $t = 3.0$), and so on. A total of four 1 s intervals had zero events, six had one event, and two had two events (Figure 7.7.48).

Example 7.7.4 A Simulation of the Poisson Process with $\lambda = 4.5$

A simulation with the higher rate $\lambda = 4.5$ produced many more events in 12 s—and much larger numbers in the histogram of results (Figure 7.7.49).

These two simulations provide a sample of the number of events that occur in 1 s. Our goal is to compute the mathematically exact probability that exactly k events occurred during an interval of length t when events occur according to a Poisson process at rate λ.

Let N be the random variable counting the number of events in an interval of length t. The Poisson distribution gives formulas for $\Pr(N = 0)$, $\Pr(N = 1)$, $\Pr(N = 2)$,

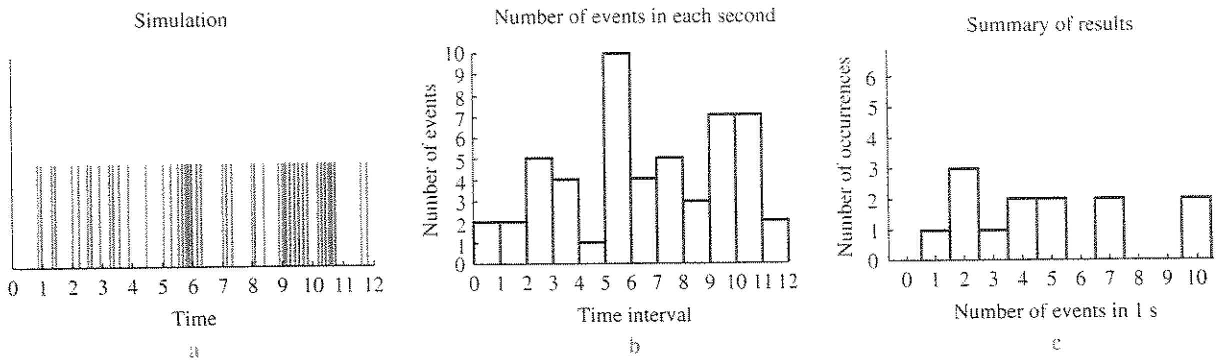


FIGURE 7.7.49
The Poisson process with $\lambda = 4.5$

and so forth. Although the underlying process occurs in continuous time, the Poisson distribution itself describes a **discrete random variable,** because it describes a count. The Poisson distribution, like the binomial distribution, describes a count of independent events. Because events occur continuously in the Poisson process, there is no upper limit on the number of events possible, in contrast to the upper limit of n successes in n trials with the binomial distribution.

We have already computed the first term in the Poisson distribution, $\Pr(N = 0)$. This term describes the probability that no events have occurred, which is given by the **exponential survivorship function**

$$\Pr(N = 0) = e^{-\lambda t}$$

Example 7.7.5 Comparing $\Pr(N = 0)$ with a Simulation Using $\lambda = 1.0$

With $\lambda = 1.0$ and $t = 1.0$, the probability of 0 events is

$$\Pr(N = 0) = e^{-1.0 \cdot 1.0} \approx 0.368$$

The simulation in Example 7.7.1 has rate $\lambda = 1.0$, and we counted the number of events in 1.0 s in Example 7.7.3. We found that four out of twelve 1-s intervals, or a fraction 0.333, experienced 0 events; this is quite close to the probability of 0.368.

Example 7.7.6 Comparing $\Pr(N = 0)$ with a Simulation Using $\lambda = 4.5$

With $\lambda = 4.5$ and $t = 1.0$, the probability of 0 events is

$$\Pr(N = 0) = e^{-4.5 \cdot 1.0} \approx 0.011$$

It is quite unlikely that nothing happens in 1 s, as is consistent with the results in Example 7.7.4, where none of the twelve 1-s intervals observed had 0 events.

The Poisson distribution fills in the rest of the probabilities, giving $\Pr(N = k)$ for every value of k. We denote $\Pr(N = k)$ by $p(k; \lambda t)$.

Theorem 7.19 Suppose events obey a Poisson process and occur at constant rate λ. The probability that exactly k events occur in time t is

$$\Pr(N = k) = p(k; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

The range of the random variable N is all non-negative integers, 0, 1, 2, ...

A hint about one way to derive these probabilities is given in Exercises 35–38.

Example 7.7.7 Checking the Formula for $\Pr(N = 0)$

In Examples 7.7.5 and 7.7.6, we used the exponential survivorship function to deduce $\Pr(N = 0) = e^{-\lambda t}$. The formula for the Poisson distribution with $k = 0$ gives

$$\begin{aligned}\Pr(N = 0) &= p(0; \lambda t)\\ &= \frac{e^{-\lambda t}(\lambda t)^0}{0!}\\ &= e^{-\lambda t}\end{aligned}$$

(recall from Table 7.2 that $0! = 1$), matching the earlier formula.

Because the Poisson distribution depends on λ and t only through their product λt, we often write $\Lambda = \lambda t$ and rewrite the formula for the Poisson distribution.

Simplified Formula for the Poisson Distribution:

$$\Pr(N = k) = p(k; \Lambda) = \frac{e^{-\Lambda}(\Lambda)^k}{k!}$$

Example 7.7.8 The Poisson Distribution with $\lambda = 1.0$ and $t = 1.0$

If we wish to find the probability distribution for the number of events that occur in a time interval $t = 1.0$ at rate $\lambda = 1.0$ each minute (as in Example 7.7.3), we use the formula for the Poisson distribution with parameter $\Lambda = \lambda t = 1.0 \cdot 1.0 = 1.0$.

$$\Pr(N = 0) = p(0; 1) = \frac{e^{-1.0}(1.0^0)}{0!} \approx 0.3679$$

$$\Pr(N = 1) = p(1; 1) = \frac{e^{-1.0}(1.0^1)}{1!} \approx 0.3679$$

$$\Pr(N = 2) = p(2; 1) = \frac{e^{-1.0}(1.0^2)}{2!} \approx 0.1839$$

$$\Pr(N = 3) = p(3; 1) = \frac{e^{-1.0}(1.0^3)}{3!} \approx 0.0613$$

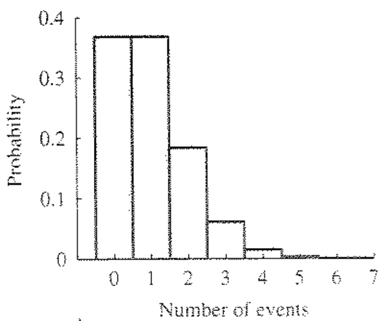


FIGURE 7.7.50
The Poisson distribution with $\Lambda = 1.0$

and so on. The probability distribution, plotted in Figure 7.7.50, matches the simulated results in Example 7.7.3 reasonably well.

Example 7.7.9 The Poisson Distribution with $\lambda = 4.5$ and $t = 1.0$

With $\lambda = 4.5$, many more events occur each minute (Example 7.7.4). In this case, $\Lambda = \lambda t = 4.5 \cdot 1.0 = 4.5$ and

$$\Pr(N = 0) = p(0; 4.5) = \frac{e^{-4.5}(4.5^0)}{0!} \approx 0.01111$$

$$\Pr(N = 1) = p(1; 4.5) = \frac{e^{-4.5}(4.5^1)}{1!} \approx 0.04999$$

$$\Pr(N = 2) = p(2; 4.5) = \frac{e^{-4.5}(4.5^2)}{2!} \approx 0.11248$$

$$\Pr(N = 3) = p(3; 4.5) = \frac{e^{-4.5}(4.5^3)}{3!} \approx 0.16872$$

$$\Pr(N = 4) = p(4; 4.5) = \frac{e^{-4.5}(4.5^4)}{4!} \approx 0.18981$$

$$\Pr(N=5) = p(5; 4.5) = \frac{e^{-4.5}(4.5^5)}{5!} \approx 0.17083$$

$$\Pr(N=6) = p(6; 4.5) = \frac{e^{-4.5}(4.5^6)}{6!} \approx 0.12812$$

$$\Pr(N=7) = p(7; 4.5) = \frac{e^{-4.5}(4.5^7)}{7!} \approx 0.08236$$

$$\Pr(N=8) = p(8; 4.5) = \frac{e^{-4.5}(4.5^8)}{8!} \approx 0.04633$$

$$\Pr(N=9) = p(9; 4.5) = \frac{e^{-4.5}(4.5^9)}{9!} \approx 0.02316$$

$$\Pr(N=10) = p(10; 4.5) = \frac{e^{-4.5}(4.5^{10})}{10!} \approx 0.01042$$

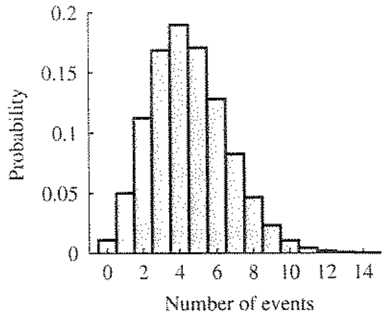


FIGURE 7.7.51
The Poisson distribution with $\Lambda = 4.5$

See Figure 7.7.51.

Example 7.7.10 The Poisson Distribution with $\lambda = 1.0$ and $t = 4.5$

Suppose we count the number of events that occur in intervals of length $t = 4.5$ min when the underlying rate is $\lambda = 1.0$. The parameter Λ is $\Lambda = \lambda t = 1.0 \cdot 4.5 = 4.5$. Because the probability distribution depends only on Λ, the probabilities exactly match those in Example 7.7.9. The number of events that occur with a fast process over a short time interval exactly match the number that occur with a slow process over a long time interval.

Although they are a bit tricky to derive (Exercises 39 and 40), the formulas for the mean and variance of the Poisson distribution are simple.

Theorem 7.20 Suppose the random variable N has a Poisson distribution with parameter Λ. Then

$$E(N) = \Lambda$$
$$\text{Var}(N) = \Lambda$$

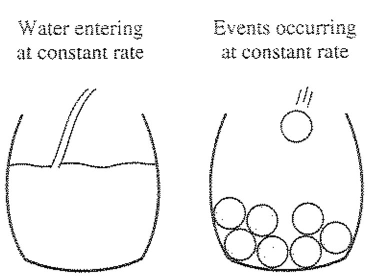


FIGURE 7.7.52
The difference between the Poisson process and an ordinary rate

Why are these formulas so simple? Because λ is a rate, the expectation Λ is the rate times the time. If water is entering a vessel at rate λ for time t, the total amount is exactly λt. The difference between the Poisson process and constant flux (see Figure 7.7.52) is the *variance*. When the flow rate is 1.0 L/min, exactly 4.5 liters enter in 4.5 min. When molecules enter at an average rate of 1.0 each minute, the *expected number* of molecules entering in 4.5 min is 4.5, but with a variance of 4.5 molecules^2 and a standard deviation of $\sqrt{4.5} \approx 2.12$ molecules.

Be careful to distinguish these results from the exponential distribution, which has *standard deviation* equal to the mean. With the Poisson distribution, the coefficient of variation gets *smaller* as the mean gets larger. By sampling more events, we get a more consistent count. The exponential distribution is always highly variable because it concerns only counting a *single* event.

Example 7.7.11 The Behavior of the Poisson Distribution as a Function of Time

Suppose events occur according to a Poisson process at rate $\lambda = 4.5$. Let $N(t)$ be a random variable representing the number of events that have occurred by time t. Then

$$E[N(t)] = \Lambda = \lambda t = 4.5t$$
$$\text{Var}[N(t)] = \Lambda = \lambda t = 4.5t$$
$$\sigma_{N(t)} = \sqrt{4.5t}$$
$$\text{CV}_{N(t)} = \frac{\sigma_{N(t)}}{E[N(t)]} = \frac{\sqrt{4.5t}}{4.5t} = \frac{1}{\sqrt{4.5t}}$$

At $t = 1$,

$$\begin{aligned} E[N(t)] &= 4.5 \\ \text{Var}[N(t)] &= 4.5 \\ \sigma_{N(t)} &= \sqrt{4.5} \approx 2.121 \\ CV_{N(t)} &= \frac{1}{\sqrt{4.5}} \approx 0.471 \end{aligned}$$

At the later time $t = 20$,

$$\begin{aligned} E[N(t)] &= 4.5 \cdot 20 = 90.0 \\ \text{Var}[N(t)] &= 4.5 \cdot 20 = 90.0 \\ \sigma_{N(t)} &= \sqrt{4.5 \cdot 20} \approx 9.487 \\ CV_{N(t)} &= \frac{1}{\sqrt{4.5 \cdot 20}} \approx 0.105 \end{aligned}$$

(Figure 7.7.53). The coefficient of variation decreases over time because there has been time to average out any streaks of unusually dense or unusually sparse events.

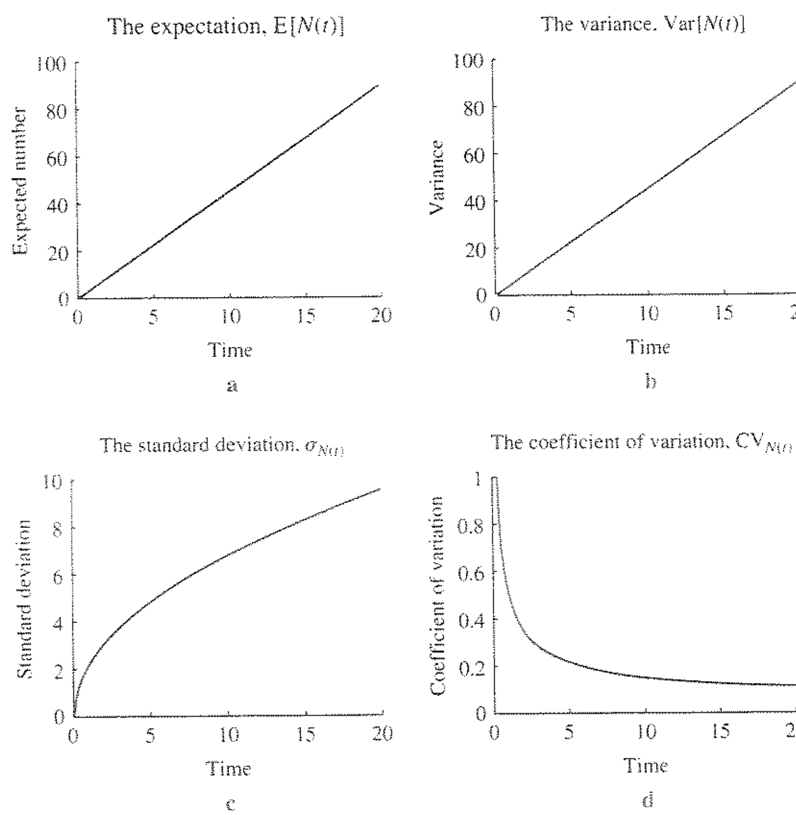


FIGURE 7.7.53
The statistics describing the Poisson distribution as functions of time

The mode of the Poisson distribution, like that for the binomial, is always close to the mean.

Theorem 7.21 Suppose the random variable N has a Poisson distribution with parameter Λ. The mode of N is the largest integer less than Λ.

Proof: The mode can be found with an iterative formula like that for the binomial distribution (Theorem 7.15). We find

$$\begin{aligned} p(k+1; \Lambda) &= \frac{e^{-\Lambda}(\Lambda)^{k+1}}{(k+1)!} \\ &= p(k; \Lambda)\frac{\Lambda}{k+1} \end{aligned}$$

The mode is the smallest value of k for which $p(k+1;\Lambda) < p(k;\Lambda)$. This occurs when $k+1 > \Lambda$. Therefore, the mode is the largest integer less than Λ. If Λ is an integer, there is a double mode of Λ and $\Lambda - 1$.

Example 7.7.12 The Mode with $\Lambda = 4.5$

With $\Lambda = 4.5$, the largest integer less than 4.5 is 4, which matches the mode in Figure 7.7.51.

The Poisson Distribution in Space

Like the other important probability distributions in biology, the Poisson distribution has a remarkable array of applications. One of the most useful is the distribution of objects or events in space rather than in time. The parallel with time is strongest in one dimension. Suppose two bacterial clones have been accumulating mutations since they last had a common ancestor and that an average of 1.3 mutations have become fixed per million nucleotides. We have a piece of DNA 4.7 million nucleotides long from each organism (Figure 7.7.54). In how many sites will the two differ?

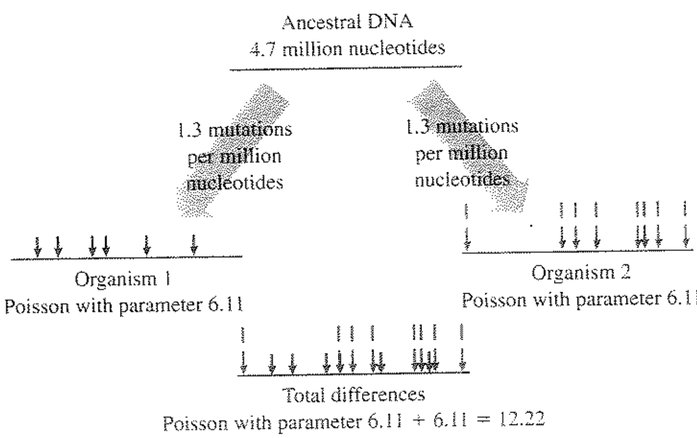


FIGURE 7.7.54
The evolution of two organisms

The picture of the mutations, which occur over the *length* of the DNA, looks exactly like the simulation of events occurring in *time* (Figure 7.7.47). What assumptions must we make to use the Poisson distribution? We need that the rate at which mutations occur along the piece of DNA is constant and that mutations occur independently. Both of these assumptions have been the source of important debates among geneticists. Assuming that they are true, the number of mutations M_1 in the first organism has a Poisson distribution with rate $\lambda = 1.3$ per million and "t" equal to 4.7 million. Therefore, $\Lambda = 1.3 \cdot 4.7 = 6.11$ mutations. The same argument holds for organism 2, so M_2 has a Poisson distribution with $\Lambda = 6.11$.

However, both M_1 and M_2 measure the number of differences from the common ancestor, which is extinct and inaccessible. What is the distribution of the number of **differences** between the two organisms? If we superimpose the two pieces of DNA, the number of differences is the **sum** of the number of mutations in each (except for the unlikely event that two mutations occurred exactly in the same spot). In mathematical notation, the total number of differences D is

$$D = M_1 + M_2$$

The following theorem tells us everything we need to know about D.

Theorem 7.22 Suppose X and Y are independent Poisson-distributed random variables with parameters Λ_X and Λ_Y. Then the sum $Z = X + Y$ has a Poisson distribution with parameter $\Lambda_X + \Lambda_Y$.

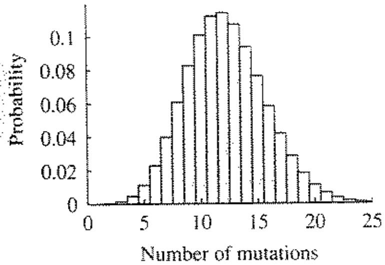


FIGURE 7.7.55
The distribution of the number of mutations separating two organisms

Therefore, D has a Poisson distribution with mean $6.11 + 6.11 = 12.22$ and variance 12.22. Our theorems about expectation and variance of the sum of independent random variables (Theorem 7.4 and Theorem 7.9) guarantee that these must be the mean and variance of D:

$$\mathrm{E}(D) = \mathrm{E}(M_1) + \mathrm{E}(M_2) = 6.11 + 6.11 = 12.22$$
$$\mathrm{Var}(D) = \mathrm{Var}(M_1) + \mathrm{Var}(M_2) = 6.11 + 6.11 = 12.22$$

Theorem 7.22 tells us more—the entire probability distribution of D. By computing the probability distribution for D, we can compare our result (14 mutations) with the expectation (about 12 mutations). Figure 7.7.55 plots the distribution, from which we can see that the probability of 14 mutations (0.094) is very close to the probability of the mode (12 mutations with probability 0.114).

The Poisson distribution also applies to processes in two or more dimensions. Suppose seeds fall independently in a region at a rate of 0.0023 seeds per square centimeter. A simulation of this process in a 9-m^2 region is shown in Figure 7.7.56, with the number of seeds in each block indicated. The number of seeds S per square meter follows a Poisson distribution. In this case, $\lambda = 0.0023$ seeds per square centimeter and "t" is 10,000 cm^2. Therefore,

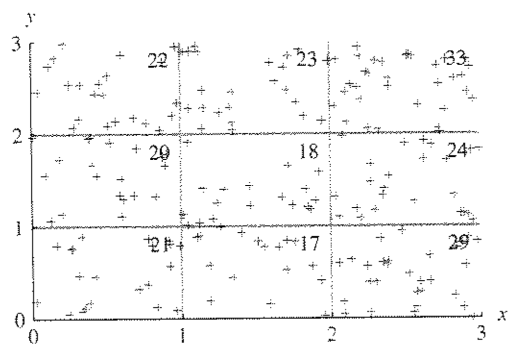


FIGURE 7.7.56
Scatter of seeds in two dimensions

$$\Lambda = 0.0023 \cdot 10000 = 23.0$$

the mean number of seeds per square meter. The probability of k seeds in a given square meter is $p(k; 23.0)$ (Figure 7.7.57). The distribution is less spread out and has more of a "bell" shape than with small Λ. The probability that S is within 5 of the expectation is

$$\Pr(18 \leq S \leq 28) = \sum_{k=18}^{28} p(k; 23.0) \approx 0.750$$

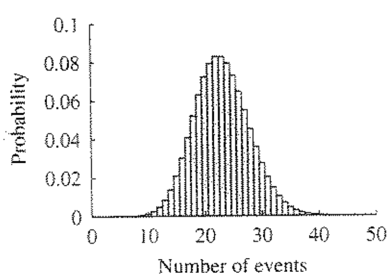


FIGURE 7.7.57
The Poisson distribution with $\Lambda = 23$ and the binomial distribution with $n = 10000$ and $p = 0.0023$

where we evaluated the various values $p(k; 23.0)$ on the computer (as with the binomial distribution, there is no convenient formula for the cumulative distribution). About 75% of square meters surveyed should have within five seeds of the expectation, consistent with the six out of nine in the simulation.

The Poisson and the Binomial

The seed example highlights the link between the Poisson distribution and the binomial distribution. The number of seeds in 1 m^2 is a count of the number in each of 10,000 cm^2, each with the tiny probability 0.0023 of containing a seed. If we ignore the probability that some square centimeter has two seeds, the count follows a binomial distribution with $n = 10000$ and $p = 0.0023$.

.13 Comparison of Probabilities with the Poisson and Binomial Distributions

The probability distribution for the Poisson distribution with $\Lambda = 23$ is nearly identical to the binomial with $n = 10000$ and $p = 0.0023$. For example,

$$p(23; 23.0) \approx 0.08288438$$
$$b(23; 10000, 0.0023) \approx 0.08297986$$

On a histogram, the two distributions appear identical.

In general, we have the following rule.

Rule for Approximating the Binomial Distribution with the Poisson Distribution
If p is small (less than about 0.01) and n is large (greater than about 100), then

$$b(k; n, p) \approx p(k; np) \tag{7.7.1}$$

The Poisson distribution can be thought of as the limit of infinitely many trials with infinitesimally unlikely Bernoulli random variables.

Let B be a random variable following a binomial distribution with parameters n and p, and let P be a random variable following a Poisson distribution with parameter $\Lambda = np$. Then

$$\mathrm{E}(B) = \mathrm{E}(P) = np$$
$$\mathrm{Var}(B) = np(1 - p)$$
$$\mathrm{Var}(P) = np$$

The expectations match, and the variances are very close when p is small. How do B and P differ? The random variable P can take on any non-negative integer value, and B can take on only values from 0 to n. However, the probability that $P > n$ (10,000 in the seed example) is astronomically small.

Example 7.7.14 Comparison of the Mean and Variance for Seed Number

If we model seeds as a random variable S that follows a Poisson distribution with parameter $\Lambda = 23.0$, then $\mathrm{E}(S) = 23.0$ and $\mathrm{Var}(S) = 23.0$. If we think of seed number instead as a random variable S_b that follows a binomial distribution with $n = 10{,}000$ and $p = 0.0023$, then $\mathrm{E}(S_b) = 10{,}000 \cdot 0.0023 = 23.0$ and $\mathrm{Var}(S_b) = 10{,}000 \cdot 0.0023 \cdot (1 - 0.0023) = 22.9471$. The binomial has the same mean and a slightly smaller variance.

Summary Events that occur independently at a constant rate λ follow a **Poisson process.** A random variable that counts the number of events that occur in time t follows a **Poisson distribution** with parameter $\Lambda = \lambda t$, the product of the rate and the time. The mean and variance of the Poisson distribution both equal Λ. The Poisson distribution also describes counts of events that occur independently in space. The sum of independent random variables that all follow Poisson distributions also follows a Poisson distribution. A binomial distribution with small p and large n can be approximated by a Poisson distribution with $\Lambda = np$.

7.7 Exercises

Mathematical Techniques

1–4 ▪ The following figures show the results of simulations of the Poisson process with the given value of the rate λ. Find the number of hits per second in each of the simulated seconds. What is the average number of hits per second? How closely does this match what you would expect from the Poisson distribution?