Chapter 3
Three Fundamental Distributions: Binomial, Gaussian, and Poisson

Abstract There are three distributions that play a fundamental role in statistics. The binomial distribution describes the number of positive outcomes in binary experiments, and it is the "mother" distribution from which the other two distributions can be obtained. The Gaussian distribution can be considered as a special case of the binomial, when the number of tries is sufficiently large. For this reason, the Gaussian distribution applies to a large number of variables, and it is referred to as the *normal* distribution. The Poisson distribution applies to counting experiments, and it can be obtained as the limit of the binomial distribution when the probability of success is small.

3.1 The Binomial Distribution

Many experiments can be considered as *binary*, meaning that they can only have two possible outcomes which we can interpret as *success* or *failure*. Even complex experiments with a larger number of possible outcomes can be described as binary, when one is simply interested about the occurrence of a specific event A, or its non-occurrence, $\overline{A}$. It is therefore of fundamental importance in statistics to determine the properties of binary experiments, and the distribution of the number of successes when the experiment is repeated for a number of times under the same experimental conditions.

3.1.1 Derivation of the Binomial Distribution

Consider a binary experiment characterized by a probability of success p a therefore a probability of failure $q = 1-p$. The probabilities p and q are determined according to the theory of probability and are assumed to be known for the experiment being considered. We seek the probability of having n successes in N tries, regardless of the order in which the successes take place. For example, consider tossing four coins, and being interested in any two of these coins showing heads up, as an indication of success of the toss. To obtain this probability, we start by counting

M. Bonamente, *Statistics and Analysis of Scientific Data*, Graduate Texts in Physics, DOI 10.1007/978-1-4939-6572-4_3

how many possible outcomes for the experiments are possible, and break down the derivation into three parts:

- *Probability of a specific sequence of n successes out of N tries.* Assume that successive experiments are independent, e.g., one tosses the same coin many times, each time independently of each other. The probability of having n successes and therefore $N-n$ failures occurring *in a specific sequence*, is given by

$$P(\textit{specific sequence} \text{ of } n \text{ successes}) = p^n \times q^{N-n}. \tag{3.1}$$

This result can be seen by using the property of independence among the N events, of which the n successes carry a probability p, and the $(N-n)$ failures a probability q.

Example 3.1 Considering the case of four coin tosses, the probability of a given sequence, for example "heads-tails-tails-heads," is $(1/2) \times (1/2) \times (1/2) \times (1/2) = (1/2)^4$, since $p = q = 1/2$. Successive tosses are assumed to be independent. ◇

- *Number of ordered sequences.* We start by counting how many ordered sequences exist that have n successes out of N tries. If there are no successes ($n = 0$), then there is only one possible sequence with N failures. If $n > 0$, each of the N tries can yield the "first" success, and therefore there are N possibilities for what try is the first success. Continuing on to the "second" success, there are only $N - 1$ possibilities left for what trial will be the second success, and so on. This leads to the following number of sequences containing n time-ordered successes, that is, sequences for which we keep track of the order in which the successes occurred:

$$\mathrm{Perm}(n, N) = N \cdot (N-1) \cdot (N-n+1) = \frac{N!}{(N-n)!}. \tag{3.2}$$

This is called the number of *permutations* of n successes out of N tries. This method of counting sequences can also be imagined as the placement of each success in a "success box": the first place in this box can be filled in N different ways, the second in $(N-1)$ ways corresponding to the remaining tries, and so on.

Example 3.2 Consider the case of $n = 2$ successes out of $N = 4$ trials. According to (3.2), the number of permutations is $4!/2! = 12$. We list explicitly all 12 ordered sequences that give rise to 2 successes out of 4 tries in Table 3.1. Symbol H_1 denotes the "first success," and H_2 the "second success." Consider, for example, lines 5 and 8: both represent the same situation in which the coin 2 and 3 showed heads up, or success, and they are not really different sequences, but the separate entries in this table are the result of our method of counting time-ordered sequences. ◇

Table 3.1 Permutations (ordered sequences) of 2 successes out of 4 tries

Sequence	Number of try 1	2	3	4	Sequence	Number of try 1	2	3	4
1	H_1	H_2	–	–	7	H_2	–	H_1	–
2	H_1	–	H_2	–	8	–	H_2	H_1	–
3	H_1	–	–	H_2	9	–	–	H_1	H_2
4	H_2	H_1	–	–	10	H_2	–	–	H_1
5	–	H_1	H_2	–	11	–	H_2	–	H_1
6	–	H_1	–	H_2	12	–	–	H_2	H_1

In reality, we are not interested in the *time order* in which the n successes occur, since it is of no consequence whether the first or the Nth, or any other, try is the "first" success. We must therefore correct for this artifact in the following.

- *Number of sequences of n successes out of N tries (regardless of order)*. As it is clear from the previous example, the number of permutations is not quite the number we seek, since it is of no consequence which success happened first. According to (3.2), there are $n!$ ways of ordering n successes among themselves, or $\text{Perm}(n,n) = n!$. Since all $n!$ permutations give rise to the same practical situation of n successes, we need to divide the number of (time-ordered) permutations by $n!$ in order to avoid double-counting of permutations with successes in the same trial number. It is therefore clear that, regardless of time order, the number of *combinations* of n successes out of N trials is

$$C(n,N) = \frac{\text{Perm}(n,N)}{n!} = \frac{N!}{(N-n)!n!} \equiv \binom{N}{n}. \tag{3.3}$$

The number of combinations is the number we seek, i.e., the number of possible sequences of n successes in N tries.

Example 3.3 Continue to consider the case of 2 successes out of 4 trials. There are $2! = 2$ ways to order the 2 successes among themselves (either one or the other is the first success). Therefore the number of combinations of 2 successes out of 4 trials is 6, and not 12. As indicated above, in fact, each sequence had its "twin" sequence listed separately, and (3.3) correctly counts only different sequences. ◇

According to the results obtained above, what remains to be done is to use the probability of each sequence (3.1) and multiply it by the number of combinations in (3.3) to obtain the overall probability of having n successes in N trials:

$$P(n) = \binom{N}{n} p^n q^{N-n} \qquad n = 0, \ldots, N. \tag{3.4}$$

This distribution is known as the *binomial distribution* and it describes the probability of n successes in N tries of a binary experiment. It is a discrete distribution that is defined for non-negative values $n \leq N$. The factor in (3.3) is in fact the binomial coefficient and it derives its name from its use in the binomial expansion

$$(p+q)^N = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}. \tag{3.5}$$

3.1.2 Moments of the Binomial Distribution

The moment of mth order for a discrete random variable X of distribution $P(n)$ is given by

$$E[X^m] = \overline{n^m} = \sum_{n=0}^{N} n^m P(n). \tag{3.6}$$

We can show that the mean and the second moment of the binomial distribution are given by

$$\begin{cases} \overline{n} = pN \\ \overline{n^2} = \overline{n}^2 + pqN. \end{cases} \tag{3.7}$$

Proof Start with the mean,

$$\overline{n} = \sum_{n=0}^{N} P(n)n = \sum_{n=0}^{N} \binom{N}{n} np^n q^{N-n} = \sum_{n=0}^{N} \binom{N}{n} \left[p\frac{\partial}{\partial p}\right] p^n q^{N-n};$$

in which we have introduced a linear operator $p\dfrac{\partial}{\partial p}$ that can be conveniently applied to the entire sum,

$$\overline{n} = p\frac{\partial}{\partial p}\left[\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}\right] = p\frac{\partial}{\partial p}(p+q)^N = pN(p+q)^{N-1} = pN.$$

The derivation for the moment $\overline{n^2}$ is similar:

$$\overline{n^2} = \sum_{n=0}^{N} P(n)n^2 = \sum_{n=0}^{N} \binom{N}{n} n^2 p^n q^{N-n} = \sum_{n=0}^{N} \binom{N}{n} \left[p\frac{\partial}{\partial p}\right]^2 q^{N-n}$$
$$= \left[p\frac{\partial}{\partial p}\right]^2 (p+q)^N = p\frac{\partial}{\partial p}\left[pN(p+q)^{N-1}\right] =$$
$$p\left[N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2}\right] =$$
$$pN + p^2N(N-1) = pN + (pN)^2 - p^2N =$$
$$\overline{n}^2 + p(1-p)N = \overline{n}^2 + pqN.$$

□

It follows that the variance of the binomial distribution is given by

$$\sigma^2 = E[(X-\overline{n})^2] = pqN. \tag{3.8}$$

Equations (3.7) and (3.8) describe the most important features of the binomial distribution, shown in Fig. 3.1 for the case of $N = 10$. The mean is naturally given by the product of the number of tries N and the probability of success p in each of the tries. The standard deviation σ measures the root mean square of the deviation and it is the measurement of the width of the distribution.

Example 3.4 (Probability of Overbooking) An airline knows that 5 % of the persons making reservations will not show up at the gate. On a given flight that

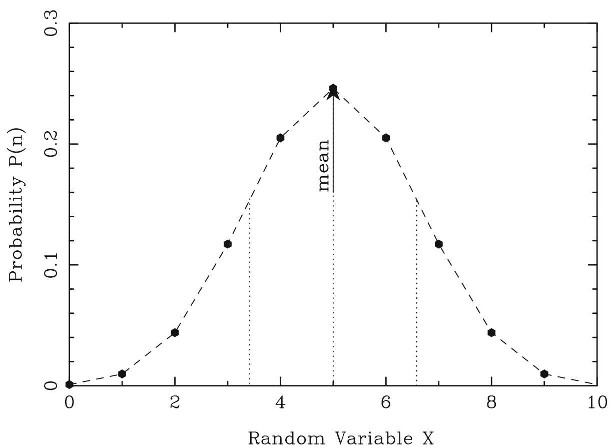

Fig. 3.1 Binomial distribution with $p = q = 0.5$ and $N = 10$. The *dotted lines* around the mean mark the $\pm\sigma$ range

can seat 50 people, 52 tickets have been sold. Calculate the probability that there will be a seat available for every passenger that will arrive at the gate.

This is a binary experiment in which $p = 0.95$ is the probability that a passenger will show. For that specific flight, $N = 52$ passenger have the choice of showing (or not). The probability that there is a seat available for each passenger is therefore given by $P = 1 - P_N(52) + P_N(51)$, which is calculated as

$$P = 1 - \binom{52}{52} p^{52} \cdot 1 - \binom{52}{51} p^{51} \cdot q = 1 - (0.95)^{52} - 52 \cdot (0.95)^{51} \cdot 0.05 = 0.741.$$

Therefore the airline is willing to take a 25.9 % chance of having an overbooked flight. ◇

3.2 The Gaussian Distribution

The Gaussian distribution, often referred to as the *normal* distribution, can be considered as a special case of the binomial distribution in the case of a large number N of experiments performed. In this section we derive the Gaussian distribution from the binomial distribution and describe the salient features of the distribution.

3.2.1 Derivation of the Gaussian Distribution from the Binomial Distribution

The binomial distribution of (3.4) acquires a simpler form when N is large. An alternative analytic expression to the binomial distribution is a great advantage, given the numerical difficulties associated with the evaluation of the factorial of large numbers. As was evident from Fig. 3.1, the binomial distribution has a maximum at value $n = Np$. In the following we prove that the binomial distribution can be approximated as

$$P(n) \simeq \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}} \tag{3.9}$$

in the case in which $N \gg 1$, and for values of the variable that are close to the peak of the distribution.

Proof Expand the logarithm of the binomial probability as a Taylor series in the neighborhood of the peak value $\tilde{n}$,

$$\ln P(n) = \ln P(\tilde{n}) + \sum_{k=1}^{\infty} \frac{B_k}{k!} \Delta n^k$$

where $\Delta n = n - \tilde{n}$ is the deviation from the peak value and

$$B_k = \left. \frac{\partial \ln P(n)^k}{\partial^k n} \right|_{n=\tilde{n}}$$

is the coefficient of the Taylor series expansion. Since, by assumption, $\tilde{n}$ is a point of maximum, the first coefficient is null, $\partial \ln P(n)/\partial n|_{n=\tilde{n}} = 0$. We neglect terms of order $O(\Delta n^3)$ in the expansion, and the approximation results in

$$\ln P(n) \simeq \ln P(\tilde{n}) + \frac{1}{2} B_2 \Delta n^2,$$

where B_2 is negative, since $n = \tilde{n}$ is a point of maximum. It follows that

$$P(n) \simeq P(\tilde{n}) e^{-\frac{|B_2| \Delta n^2}{2}}.$$

Neglecting higher-order terms in Δn means that the approximation will be particularly accurate in regions where Δn is small, i.e., near the peak of the distribution. Away from the peak, the approximation will not hold with the same precision.

In the following we show that the unknown terms can be calculated as

$$\begin{cases} B_2 = -\dfrac{1}{Npq} \\ P(\tilde{n}) = \dfrac{1}{\sqrt{2\pi Npq}}. \end{cases}$$

First, we calculate the value of $|B_2|$. Start with

$$\ln P(n) = \ln \left(\frac{N!}{n!(N-n)!} p^n q^{N-n} \right) =$$
$$\ln N! - \ln n! - \ln(N-n)! + n \ln p + (N-n) \ln q.$$

At this point we need to start treating n as a continuous variable. This approximation is acceptable when $Np \gg 1$, so that values n of the random variable near the peak of the distribution are large numbers. In this case, we can approximate the derivative of the logarithm with a difference,

$$\frac{\partial \ln n!}{\partial n} = (\ln(n+1)! - \ln n!)/1 = \ln(n+1) \simeq \ln n.$$

From this it follows that the first derivative of the probability function, as expected, is zero at the peak value,

$$\begin{aligned}\left.\frac{\partial \ln P(n)}{\partial n}\right|_{n=\tilde{n}} &= -\ln n + \ln(N-n) + \ln p - \ln q|_{n=\tilde{n}} \\ &= \ln\left(\frac{N-n}{n}\frac{p}{q}\right) = 0\end{aligned}$$

so that the familiar result of $\tilde{n} = p \cdot N$ is obtained. This leads to the calculation of the second derivative,

$$\begin{aligned}B_2 = \left.\frac{\partial^2 \ln P(n)}{\partial n^2}\right|_{n=\tilde{n}} &= \left.\frac{\partial}{\partial n}\ln\left(\frac{N-n}{n}\frac{p}{q}\right)\right|_{n=\tilde{n}} \\ = \left.\frac{\partial}{\partial n}(\ln(N-n) - \ln n)\right|_{n=\tilde{n}} &= \left.-\frac{1}{N-n} - \frac{1}{n}\right|_{n=\tilde{n}} \\ = -\frac{1}{\tilde{n}} - \frac{1}{N-\tilde{n}} = -\frac{1}{Np} - \frac{1}{N(1-p)} &= -\frac{p+q}{Npq} = -\frac{1}{Npq}.\end{aligned}$$

Finally, the normalization constant $P(\tilde{n})$ can be calculated making use of the integral

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a}.$$

Enforcing the normalization condition of the probability distribution function,

$$\int_{-\infty}^{\infty} P(\tilde{n}) e^{-\frac{|B_2|\Delta n^2}{2}} d\Delta n = P(\tilde{n})\sqrt{\frac{2\pi}{|B_2|}} = 1$$

we find that $P(\tilde{n}) = 1/\sqrt{2\pi Npq}$. We are therefore now in a position to obtain an approximation to the binomial distribution, valid when $n \gg 1$:

$$P(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2NPq}}.$$

□

Using the fact that the mean of the distribution is $\mu = Np$, and that the variance is $\sigma^2 = Npq$, the approximation takes the form

$$P(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \tag{3.10}$$

which is the standard form of the Gaussian distribution, in which n is a continuous variable. Equation (3.10) read as $P(n)$ being the probability of occurrence of the value n for a given random variable of mean μ and variance σ^2. The Gaussian distribution has the familiar "bell" shape, as shown in Fig. 3.2. When n becomes a continuous variable, which we will call x, we talk about the probability of occurrence of the variable in a given range $x, x + dx$. The Gaussian probability distribution function is thus written as

$$f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \tag{3.11}$$

A Gaussian of mean μ and variance σ^2 is often referred to as $N(\mu, \sigma)$. The standard Gaussian is one with zero mean and unit variance, indicated by $N(0, 1)$.

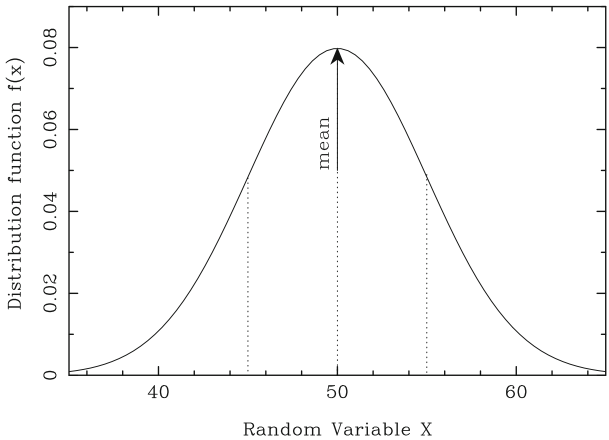

Fig. 3.2 Gaussian distribution with $\mu = 50$ and $\sigma^2 = 12.5$, corresponding to a binomial distribution of $p = q = 0.5$, and $N = 100$. The Gaussian distribution is symmetrical around the mean and therefore the mean, mode, and median coincide. The *dotted lines* around the mean mark the $\pm\sigma$ range

3.2.2 Moments and Properties of the Gaussian Distribution

The parameters μ and σ^2 are, respectively, the mean and variance of the Gaussian distribution. These results follow from the derivation of the Gaussian distribution from the binomial, and can be confirmed by direct calculation of expectations from (3.11). Central moments of odd order are zero, since the Gaussian is symmetric with respect to the mean.

Given its wide use in statistics, it is important to quantify the "effective width" of the Gaussian distribution around its mean. The *half width at half maximum*, or HWHM, is defined as the range of x between the peak and the point where $P(x) = 0.5P(\mu)$. It can be easily shown that the HWHM has a size of approximately 1.18σ, meaning that the half-maximum point is just past one standard deviation of the mean. By the same token, the *full-width at half maximum*, or FWHM, is defined as the range between the two points where $P(x) = 0.5P(\mu)$. It is twice the HWHM, or 2.35σ in size. Tables of the Gaussian distribution are provided in Appendix A.1.

The range between the points $x = \mu \pm \sigma$ is a common measure of the effective range of the random variable. The probability of a Gaussian variable to be in the range from $\mu - \sigma$ to $\mu + \sigma$ is calculated as the integral of the probability distribution function between those limits. In general, we define the integral

$$A(z) = \int_{\mu - z\sigma}^{\mu + z\sigma} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-\frac{x^2}{2}} dx \tag{3.12}$$

where $f(x)$ is the Gaussian distribution; this integral is related to the *error function*,

$$\operatorname{erf} z = \frac{1}{\sqrt{\pi}} \int_{-z}^{z} e^{-x^2} dx. \tag{3.13}$$

The function $A(z)$ is tabulated in Appendix A.1 The probability of the variable to be within one σ of the mean is $A(1) = 0.683$, or 68.3 %. The range of x between $\mu - \sigma$ and $\mu + \sigma$ therefore corresponds to a 68.3 % interval of probability, and it is referred to as the 1σ interval. The correspondence between the 1σ interval and the 68.3 % confidence interval applies strictly only to the Gaussian distribution, for which the value of σ is defined via the distribution function. It is common practice, however, to calculate to the 68.3 % interval (sometimes shortened to 68 %) even for those random variables that do not strictly follow a Gaussian distribution, and refer to it as the 1σ interval. The probability associated with characteristic intervals of a Gaussian variable is also reported in Table 3.2.

The cumulative distribution of a Gaussian random variable $N(0,1)$ is defined by the following integral:

$$B(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx; \tag{3.14}$$

Table 3.2 Probability associated with characteristic intervals of a Gaussian distribution

Interval	Integrated probability
$\mu-\sigma, \mu+\sigma$ (1σ interval)	0.6827, or 68.27 %
$\mu-2\sigma, \mu+2\sigma$ (2σ interval)	0.9545, or 95.45 %
$\mu-3\sigma, \mu+3\sigma$ (3σ interval)	0.9973, or 99.73 %
$\mu-4\sigma, \mu+4\sigma$ (4σ interval)	0.9999, or 99.99 %

the integral can be calculated as $B(z) = 1/2 + A(z)/2$ for $z > 0$ and it is tabulated in Table A.3. For $z < 0$, the table can be used to calculate the cumulative distribution as $B(z) = 1 - B(-z)$.

3.2.3 How to Generate a Gaussian Distribution from a Standard Normal

All Gaussian distributions can be obtained from the standard $N(0, 1)$ via a simple change of variable. If X is a random variable distributed like $N(\mu, \sigma)$, and Z a standard Gaussian $N(0, 1)$, then the relationship between Z and X is given by

$$Z = \frac{X-\mu}{\sigma}. \tag{3.15}$$

The variable Z is also referred to as the *z-score* associated with the variable X. This equation means that if we can generate samples from a standard normal, we can also have samples from any other Gaussian distribution. If we call z a sample from Z, then

$$x = \sigma \cdot z + \mu \tag{3.16}$$

will be a sample drawn from X, according to the equation above. Many programming languages have a built-in function to generate samples from a standard normal, and this simple process can be used to generate samples from any other Gaussian. A more general procedure to generate a given distribution from a uniform distribution will be presented in Sect. 4.8.

3.3 The Poisson Distribution

The Poisson distribution describes the probability of occurrence of events in counting experiments, i.e., when the possible outcome is an integer number describing how many counts have been recorded. The distribution is therefore discrete and can be derived as a limiting case of the binomial distribution.

3.3.1 Derivation of the Poisson Distribution

The binomial distribution has another useful approximation in the case in which $p \ll 1$, or when the probability of success is small. In this case, the number of positive outcomes is much smaller than the number of tries, $n \ll N$, and the factorial function can be approximated as

$$N! = N(N-1)\cdots(N-n+1)\cdot(N-n)! \simeq N^n(N-n)!.$$

We are also interested in finding an approximation for the q^{N-n} term that appears in the binomial. For this we set

$$\ln q^{N-n} = \ln(1-p)^{N-n} = (N-n)\ln(1-p) \simeq -p(N-n) \simeq -pN,$$

and therefore we obtain the approximation

$$q^{N-n} \simeq e^{-pN}.$$

These two approximations can be used into (3.4) to give

$$P(n) \simeq \frac{N^n(N-n)!}{n!(N-n)!}p^n e^{-pN} = \frac{(pN)^n}{n!}e^{-pN}. \tag{3.17}$$

Since pN is the mean of the distribution, we can rewrite our approximation as

$$P(n) = \frac{\mu^n}{n!}e^{-\mu}, \tag{3.18}$$

known as the Poisson distribution. This function describes the probability of obtaining n positive outcomes, or counts, when the expected number of outcomes is μ. It can be immediately seen that the distribution is properly normalized, since

$$\sum_{n=0}^{\infty}\frac{\mu^n}{n!} = e^{\mu}.$$

A fundamental feature of this distribution is that it is described by only one parameter, the mean μ, as opposed to the Gaussian distribution that had two parameters. This clearly does not mean that the Poisson distribution has no variance—in that case, it would not be a random variable!—but that the variance can be written as function of the mean, as will be shown in the following.

3.3.2 Properties and Interpretation of the Poisson Distribution

The approximations used in the derivation of (3.18) caused the loss of any reference to the initial binomial experiment, and only the mean $\mu = Np$ is present. Using the definition of mean and variance, it is easy to prove that the mean is indeed μ, and that the variance is also equal to the mean, $Var(n) = \sigma^2 = \mu$. The fact that the mean equals the variance can be seen using the values for the binomial, $\mu = Np$ and $\sigma^2 = Npq$; since $p \ll 1$, $q \simeq 1$, and $\mu \simeq \sigma^2$. As a result, the Poisson distribution has only one parameter.

The Poisson distribution can be interpreted as the probability of occurrence of n events in the case of an experiment that detects individual counts, when the mean of the counts is μ. This makes the Poisson distribution the primary statistical tool for all experiments that can be expressed in terms of the *counting* of a specific variable associated with the experiment. Typical examples are the counting of photons or the counting of plants with a given characteristic, etc. When an experiment can be cast in terms of a counting experiment, even without a specific reference to an underlying binary experiment, then the Poisson distribution will apply. All reference to the total number of possible events (N) and the probability of occurrence of each event (p) was lost because of the approximation used throughout, i.e., $p << 1$, and only the mean μ remains to describe the primary property of the counting experiment, which is the mean or expectation for the number of counts.

As can be seen in Fig. 3.3, the Poisson distribution is not symmetric with respect of the mean, and the distribution becomes more symmetric for larger values of the mean. As for all discrete distributions, it is only meaningful to calculate the probability at a specific point or for a set of points, and not for an interval of points as in the case of continuous distributions. Moreover, the mean of the distribution itself can be a non-integer number, and still the outcome of the experiment described by the Poisson distribution can only take integer values.

Example 3.5 Consider an astronomical source known to produce photons, which are usually detected by a given detector in the amount of $\mu = 2.5$ in a given time interval. The probability of detecting $n = 4$ photons in a given time interval is therefore

$$P(4) = \frac{2.5^4}{4!} e^{-2.5} = 0.134$$

The reason for such apparently large probability of obtaining a measurement that differs from the expected mean is simply due to the statistical nature of the detection process. ◇

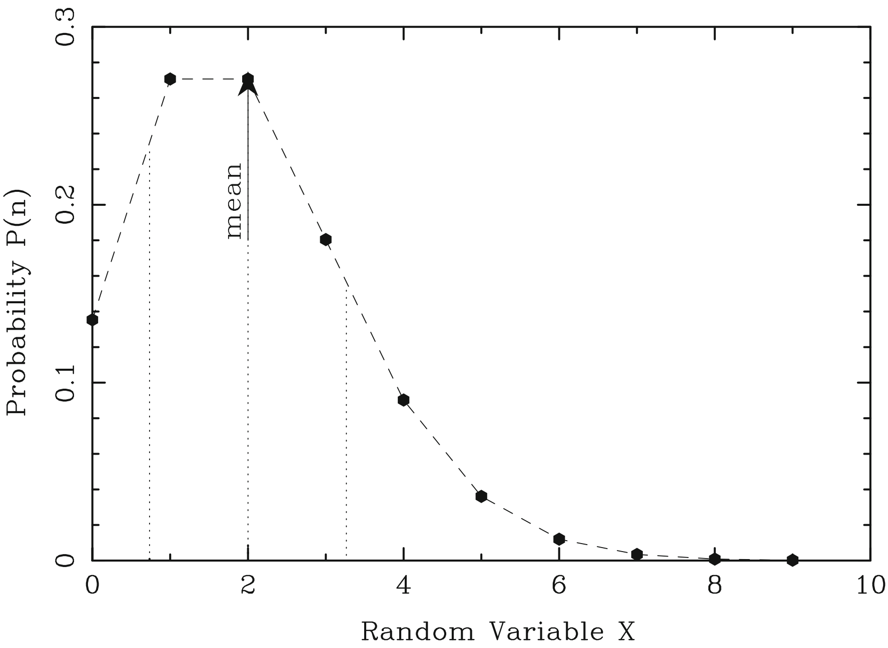

Fig. 3.3 Poisson distribution with $\mu = 2$, corresponding to a binomial distribution with $p = 0.2$ and $N = 10$. The *dotted lines* represent the mean, the $\mu - \sigma$ and $\mu + \sigma$ points

3.3.3 The Poisson Distribution and the Poisson Process

A more formal justification for the interpretation of the Poisson distribution as the distribution of counting experiments comes from the Poisson process. Although a complete treatment of this subject is beyond the scope of this book, a short description of stochastic processes will serve to strengthen the interpretation of (3.18), which is one of the foundations of statistics. More details on stochastic processes can be found, for example, in the textbook by Ross [38].

A *stochastic counting process* $\{N(t), t > 0\}$ is a sequence of random variables $N(t)$, in which t indicates time, and $N(t)$ is a random variable that indicates the number of events occurred up to time t. The stochastic process can be thought of as repeating the experiment of "counting the occurrence of a given event" at various times t; $N(t)$ is the result of the experiment. The *Poisson process with rate* λ is a particular type of stochastic process, with the following properties:

1. $N(0) = 0$, meaning that at time 0 there are no counts detected.
2. The process has *independent increments*, meaning that $N(t + s) - N(s)$ is independent of $N(t)$; this is understood with the events occurring after time t not being influenced by those occurring prior to it.

3. The process has *stationary increments*, i.e., the distribution of the number of events in an interval of time s depends only on the length of the time interval itself.
4. $P(N(h) = 1) = \lambda h + o(h)$ in which $o(h)$ is a function with the property that

$$\lim_{h \to 0} \frac{o(h)}{h} = 0.$$

5. $P(N(h) \geq 2) = o(h)$. The latter two properties mean that the probability of obtaining one count depends on the finite value λ, while it is unlikely that two or more events occur in a short time interval.

It can be shown that under these hypotheses, the number of events $N(t)$ recorded in any interval of length t is Poisson distributed,

$$P\{N(t+s) - N(s) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{3.19}$$

This shows that the Poisson distribution is to be interpreted as the distribution of occurrence of n events during a time interval t, under the hypothesis that the rate of occurrence of events is λ. This interpretation is identical to the one provided above, given that $\mu = \lambda t$ is the mean of the counts in that time interval.

3.3.4 An Example on Likelihood and Posterior Probability of a Poisson Variable

The estimation of parameters of a random variable, such as the mean of the Poisson distribution, will be treated in full detail in Chap. 5. Here we present a simple application that consists of using available measurements to calculate the likelihood and to make inferences on the unknown value of the parent mean μ of a Poisson variable. The following examples illustrate how a single measurement n of a Poisson variable can be used to constrain the true mean μ, and that care must be exercised in not confusing the likelihood of a measurement with the *posterior probability*. We assume for simplicity that the mean is an integer, although in general it may be any real number.

Within the Bayesian framework, a counting experiment can be written in terms of a dataset B, consisting of the measurement n of the variable, and events A_i, representing the fact that the parent mean is $\mu = i$. It follows that the likelihood can be written as

$$P(B/A_i) = \frac{i^n}{n!} e^{-i}.$$

Example 3.6 (Calculation of Data Likelihood) A counting experiment results in a detection of $n = 4$ units, and one wants to make a statement as to what is the probability of such measurement. Using the Poisson distribution, the probability of detecting 4 counts if, for example, $\mu = 0$, 1, or 2, is given by the likelihood

$$P(B/A_{012}) = \sum_{\mu=0}^{2} \frac{\mu^4}{4!} e^{-\mu} = 0 + \frac{1}{4!}\frac{1}{e} + \frac{2^4}{4!}\frac{1}{e^2} = 0.015 + 0.091 = 0.106,$$

or 10.6 %; this is a likelihood of the data with models that assume a specific value for the mean. Notice that if the true value of the mean is zero, there is absolutely no probability of detecting any counts. One can thus conclude that there is slightly more than a 10 % chance of detecting 4 counts, given that the source truly emits 2 or fewer counts. This is not, however, a statement of possible values of the parent mean μ. ◇

According to Bayes' theorem, the posterior distributions are

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{P(B)}$$

where $P(B/A_i)$ is the likelihood, corresponding to each of the three terms in the sum of the example above. In the following example, we determine posterior probabilities.

Example 3.7 (Posterior Probability of the Poisson Mean) We want to calculate the probability of the true mean being less or equal than 2, $P(A_{012}/B)$, and start by calculating the likelihoods required to evaluate $P(B)$. We make an initial and somewhat arbitrary assumption that the mean should be $\mu \leq 10$, so that only 11 likelihoods must be evaluated. This assumption is dictated simply by practical considerations, and can also be stated in terms of assuming a subjective prior knowledge that the mean is somehow known not to exceed 10. We calculate

$$P(B) \simeq \sum_{i=0}^{10} \frac{i^4}{4!} e^{-i} \times P(A_i) = 0.979 \times P(A_i)$$

Also, assuming uniform priors, we have $P(A_i) = 1/11$ and that

$$P(A_{012}/B) = \frac{P(A_i) \times \sum_{i=0}^{2} \frac{i^4}{4!} e^{-i}}{P(A_i) \times 0.979} = \frac{1}{0.979} \sum_{i=0}^{2} \frac{i^4}{4!} e^{-i} = 0.108.$$

◇

The examples presented in this section illustrate the conceptual difference between the likelihood calculation and the estimate of the posterior, though the two calculations yielded similar numerical values.

3.4 Comparison of Binomial, Gaussian, and Poisson Distributions

In this section we provide numerical calculations that compare the binomial and Gaussian functions, and also discuss under what circumstances the Poisson distribution can be approximated by a Gaussian of same mean and variance. In fact practical computations with the Poisson distribution are often hampered by the need to calculate the factorial of large numbers. In Sect. 3.2 we derived the Gaussian distribution from the binomial function, using the approximation that $Np \gg 1$. In fact we assumed that the function has values $n \gg 1$ and, since the mean of the binomial is $\mu = Np$, the value Np sets the order of magnitude for the values of the random variable that have non-negligible probability. In the left panel of Fig. 3.4 we show the binomial distribution with parameters $p = q = 0.5$, showing that for $Np = 5$ the approximation is already at the level of 1 % near the peak of the distribution.

The main limitation of the Poisson distribution (3.18) is the presence of the factorial function, which becomes very rapidly a large number as function of the integer n (for example, $20! = 2.423 \times 10^{18}$), and it may lead to overflow problems in numerical codes. For large values of n, one can use the Stirling approximation to the factorial function, which retains only the first term of the following expansion:

$$n! = \sqrt{2\pi n} \times n^n e^{-n} \left(1 + \frac{1}{12n} + \ldots\right). \tag{3.20}$$

Using this approximation for values of $n \geq 10$, the right panel of Fig. 3.4 shows two Poisson distributions with mean of, respectively, 3 and 20, and the corresponding Gaussian distributions with the same mean and of variance equal to the mean, as is the case for the Poisson distribution. The difference between the Gaussian and the

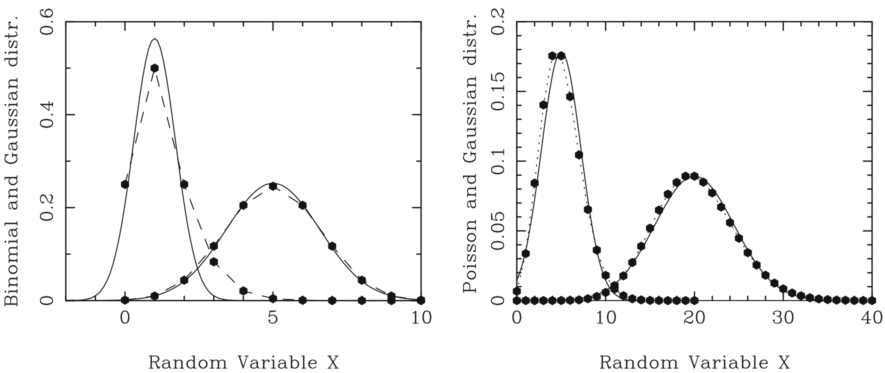

Fig. 3.4 (*Left*) Binomial distributions with $p = q = 0.5$ and, respectively, $N = 2$ and $N = 10$ as points connected by *dashed line*. Matching Gaussian distributions with same mean $\mu = Np$ and variance $\sigma^2 = Npq$ (*solid lines*). (*Right*) Gaussian distribution with $\mu = \sigma^2 = 3$ and $\mu = \sigma^2 = 20$ (*solid lines*) and Poisson distributions with same mean as points connected by a *dotted line*

Poisson distributions for a mean of $\mu = 20$ is at the percent level near the peak of the distribution. The Poisson distribution retains its characteristic asymmetry and a heavier tail at large values, and therefore deviations between the two function are larger away from the mean where, however, the absolute value of the probability becomes negligible. It can also be shown that for the value of $x = \mu$, the two distributions have the same value, when the Stirling approximation is used for the factorial function. A rule of thumb used by many is that for $x \geq 20$ the Gaussian approximation to the Poisson distribution is acceptable.

The approximation of a Poisson distribution with a Gaussian distribution is of great practical importance. Consider a counting experiment in which N counts are measured. The parent distribution of the random variable of interest is Poisson distributed and it is reasonable to assume that the best estimate of its mean is $\mu = N$ (but see Sect. 5.5.1 for a Bayesian approach that gives a slightly different answer). For values of $N > 20$ or so, the standard deviation of the parent Poisson distribution is therefore $\sigma = \sqrt{N}$. The measurement can be reported at $N \pm \sqrt{N}$, where the range of $N \pm \sqrt{N}$ corresponds to the $\mu \pm 1\sigma$ interval for a Gaussian variable.

Summary of Key Concepts for this Chapter

- ☐ *Binomial distribution:* It describes the probability of occurrence of n successes in N tries of a binary event,

$$P(n) = \binom{N}{n} p^n q^{N-n}$$

(mean pN and variance pqN).

- ☐ *Gaussian distribution*: It is an approximation of the binomial distribution when N is large,

$$f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

(mean μ and variance σ^2).

- ☐ *Poisson distribution*: It is an approximation of the binomial distribution when $p \ll 1$ that describes the probability of counting experiments,

$$P(n) = \frac{\mu^n}{n!} e^{-\mu}$$

(mean and variance have a value of μ).

Problems

3.1 Consider the Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Calculate the mean and variance and show that all odd moments $E[(X - \mu)^n]$ of order $n \geq 3$ are zero.

3.2 Assume that scores from an I.Q. test follow a Gaussian distribution, and that the scores are standardized in such a way that the mean is $\mu = 100$, and the standard deviation is $\sigma = 15$.

(a) Calculate the probability that an I.Q. score is greater or equal than 145.
(b) Calculate the probability that the *mean I.Q. score* of a sample of 100 persons, chosen at random, is equal or larger than 105.

3.3 A coin is tossed ten times. Find

(a) The probability of obtaining 5 heads up and 5 tails up;
(b) The probability of having the first 5 tosses show heads up, and the final 5 tosses show tails up;
(c) The probability to have at least 7 heads up.

3.4 In a given course, it is known that 7.3 % of students fail.

(a) What is the expected number of failures in a class of 32 students?
(b) What is the probability that 5 or more students fail?

3.5 The frequency of twins in European population is about 12 in every 1000 maternities. Calculate the probability that there are no twins in 200 births, using (a) the binomial distribution, and (b) the Poisson distribution.

3.6 Given the distribution of a Poisson variable N,

$$P(n) = \frac{\mu^n}{n!} e^{-\mu}$$

show that the mean is given by μ and that the variance is also given by μ.

3.7 Consider Mendel's experiment of Table 1.1 at page 9 and refer to the "Long vs. short stem" data.

(a) Determine the parent distribution for the number of dominants.
(b) Calculate the uncertainty in the measurement of the number of plants that display the dominant character.
(c) Determine the difference between the number of measured plants with the dominant character and the expected number, in units of the standard deviation, to show that this number has an absolute value of less than one.

3.8 For Mendel's experimental data in Table 1.1 at page 9, consider the overall fraction of plants that display the dominant character, for all seven experiments combined.

(a) Determine the parent distribution of the overall fraction X of plants with dominant character and its expected value.
(b) Determine the sample mean of the fraction X;
(c) Using the parent variance of X, determine the value

$$z = \frac{x - E[X]}{\sigma}$$

which is the standardized difference between the measurement and the mean. Assuming that the binomial distribution can be approximated by a Gaussian of same mean and variance, calculate the probability of having a value of z equal or smaller (in absolute value) to the measured value.