Chapter 4

Repeated Trials

4.1 Introduction

Repeated indepentent trials in which there can be only two outcomes are called Bernoulli trials in honor of James Bernoulli (1654-1705). As we shall see, Bernoulli trials lead to the binomial distribution. If a the number of trials is large, then the probability of k successes in n trials can be approximated by the Poisson distribution. The binomial distribution and the Poisson distribution are closely approximated by the normal (Gaussian) distribution. These three distributions are the foundation of much of the analysis of physical systems for detection, communication and storage of information.

4.2 Bernoulli Trials

Consider an experiment $\mathbf{E}$ that has two outcomes, say a and b, with probability p and $q = 1 - p$, respectively. Let $\mathbf{E}_n$ be the experiment that consists of n independent repetitions of $\mathbf{E}$. The outcomes are sequences with the components a and b. The outcomes of $\mathbf{E}_2$ are $\{aa\}, \{ab\}, \{ba\}, \{bb\}$, with probabilities p^2, pq, pq, and q^2, respectively.

Theorem 4.2.1 *The outcomes of* $\mathbf{E}_n$ *are the* 2^n *sequences of length* n. *The number of outcomes of* $\mathbf{E}_n$ *that contain* a *exactly* k *times is given by the binomial coefficient.* $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof: Assume that each of the terms in the expansion of $(a + b)^n$ represents one of the possible outcomes of the experiment

> $\mathbf{E}_n$. Then multiplying by $(a + b)$ to form $(a + b)^{n+1}$ produces an expression in which each term in $(a+b)^n$ appears twice–once with a appended and once with b appended. If the assumption is true, then this constructs all possible distinct arrangements of $n + 1$ terms.
>
> The above is certainly true for $n = 1$, since $(a + b)^1 = a + b$, so that all sequences of length $n = 1$ are represented. Multiplying by $(a + b)$ causes a and b to be suffixed to the terms in the previous expression, yielding $(a + b)^2 = (a + b)(a + b) = aa + ab + ba + bb$. These are all sequences of length $n = 2$. At this point, we have proved the theorem by induction. Since it is true for $n = 1$ it must be true for $n = 3$, and so on.♦

It is known that $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$, so that the number of terms that contain a k-times and b $(n-k)$-times is given by the binomial coefficient $\binom{n}{k}$.

To illustrate, multiplication of $(a + b)^2$ by $(a + b)$ yields $(a + b)^3 = (a + b)^2(a+b) = aaa+aba+baa+bba+aab+abb+bab+bbb. = a^3+3a^2b+3ab^2+b^3$, in which $\binom{3}{0} = 1$, $\binom{3}{1} = 3$, $\binom{3}{2} = 3$, $\binom{3}{3} = 1$. We would not necessarily use this technique to generate all the sequences, but we could. An important insight of this techique is the calculation of the probability the outcome is a term with k appearances of a and $(n - k)$ appearances of b. In the following we will use the common language of "success" and "failure", which correspond to a and b, for the result of each trial in a Bernoulli experiment.

Theorem 4.2.2 *The probability that the outcome of an experiment that consists of n Bernoulli trials has k successes and n − k failures is given by the* binomial distribution

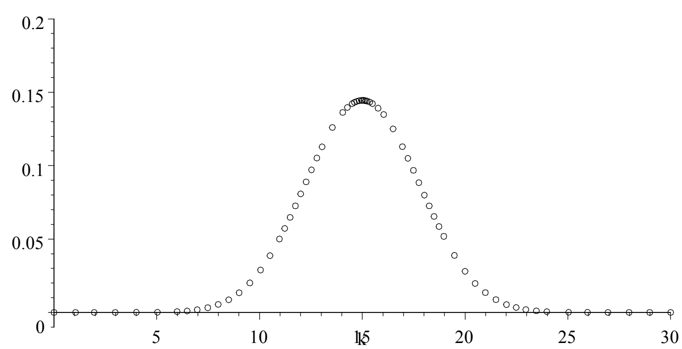
$$b(n, k, p) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{4.1}$$

where the probability of success on an individual trial is given by p.

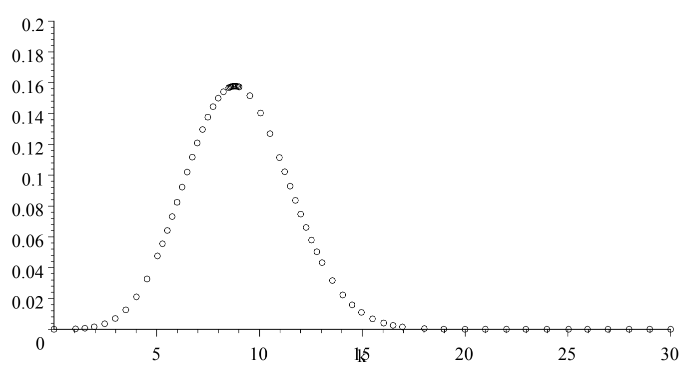
> **Proof:** Since the trials are independent, the probability of any particular sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$. All sequences with the same mixture are equally probable. Hence, the probability of *some* sequence with k successes and $(n-k)$ failures is the number of such sequences times the probability. By

the theorem 4.2.1, the number of such sequences is $\binom{n}{k}$. Multiplying by the sequence probability $p^k(1-p)^{n-k}$ gives the required result.♦

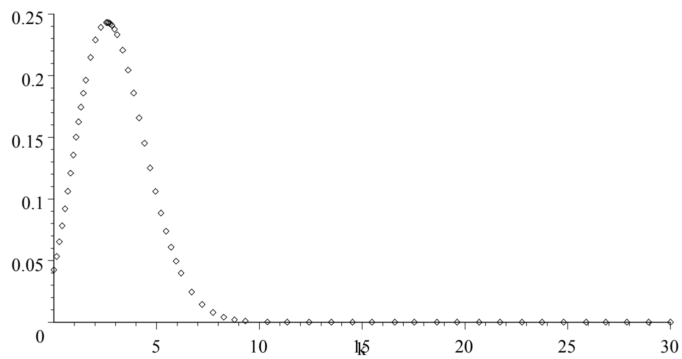
Graphs of the binomial distribution for $n = 30$ and $p = 0.5$, 0.3 and 0.1 are shown in the following figures. We note that the distribution has a peak value that falls near np. This is an important quality of the binomial distribution. It says that if the probability of success is p then on n trials you are most likely to observe about np successes and $n(1-p)$ failures. The fact that the largest value is near $k = np$ was established in Problem 6 of Section 2.6.5.



A plot of the binomial distribution $b(30, k, 0.5)$.



A plot of the binomial distribution $b(30, k, 0.3)$.



A plot of the binomial distribution $b(30, k, 0.1)$.

The variance of a binomial distribution is shown in Exercise 3 to be $np(1-p)$, and hence the standard deviation is $\sqrt{np(1-p)}$ The standard deviation is a measure of the spread of a distribution about its mean value. Both the mean value and the standard deviation increase with the number of trials, but the mean value increases faster. Consider the ratio σ/μ as a measure of the spread relative to the mean value. We see that $\sigma/\mu = \frac{1}{\sqrt{n}}\sqrt{\frac{1-p}{p}}$ which is a function that decreases in proportion to the square root of the number of trials. A graph of the binomial distribution as a function of the fraction k/n, which places it on a normalized scale, is shown in Figure 4.1 for $n = 30$, 100 and 300, which shows the concentration near the mean value as n increases.

4.2.1 Law of Large Numbers

With Bernoulli trials it is natural to ask a question such as "How wide is the central peak?" where by "wide" we may mean the interval that contains, say, 95% of the probability. It is clear from Figure 4.1 that the size of such an interval must shrink with increasing n. This question and many that are related to it can be answered by defining the random variable S_n to be the number of successes on n trials. This corresponds to the index k that we have been using in the binomial distribution. Then

$$P\left[S_n = k\right] = b(n, k, p) \tag{4.2}$$

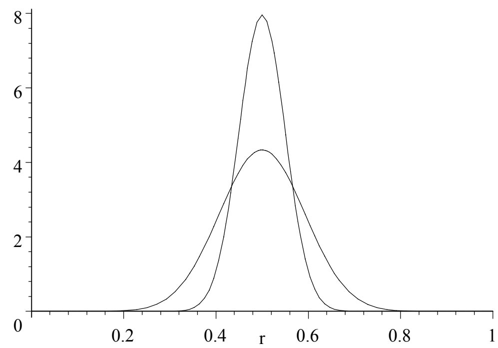


Figure 4.1: Three plots of the binomial distribution, for $n = 30, 100, 300$. The curves become sharper as n increases. The graphs have been normalized so they cover the same area, which visually adjusts for the different number of points in each plot.

We may now ask about the distribution function for S_n. By definition,

$$F_{S_n}(r) = P[S_n \leq r] = \begin{cases} 0 & r < 0 \\ \sum_{k=0}^{r} b(n,k,p) & 0 \leq r \leq n \\ 1 & r > n \end{cases} \tag{4.3}$$

Let us look at the right tail of the distribution. Analysis of the left tail will be symmetric. This will lead us to a method of analyzing the central region. If $r > np$ then $b(n, r, p)$ is on an decreasing part of the curve. The ratio[1] of successive terms is

$$\frac{b(n,r,p)}{b(n,r-1,p)} = \frac{(n-r+1)p}{rq} = 1 + \frac{(n+1)p-r}{rq} < 1 - \frac{r-np}{rq} \tag{4.4}$$

The ratio of successive terms is a number that is decreasing. Therefore, the sum is smaller than the sum over a geometric series, in which the ratio of terms is a constant. A bound on the probability $P[S_n \geq r]$ is therefore given by the geometric sum with ratio ρ if ρ is the ratio for the first pair of terms.

$$P[S_n \geq r] \leq \sum_{k=0}^{\infty} b(n,r,p)\rho^k = b(n,r,p)\frac{1}{1-\rho} \tag{4.5}$$

[1] Here as elsewhere we will use $q = 1 - p$ as a simplification.

Substitution of $\rho = 1 - \frac{r-np}{rq}$ now leads to the upper bound

$$P[S_n \geq r] \leq b(n, r, p)\frac{rq}{r - np} \tag{4.6}$$

We now need to replace $b(n, r, p)$ with an upper bound that is easy to work with. We do this by noting that all of the terms between the center, m, and r are greater than $b(n, r, p)$ and that the total of those terms must be less than 1. The number of such terms is no more than $r - np$, so $(r - np)b(n, r, p) < 1$ so that $b(n, r, p) < 1/(r - np)$. Puttint this into the above equation yields the simple upper bound

$$P[S_n \geq r] \leq \frac{rq}{(r - np)^2} \quad \text{if } r > np \tag{4.7}$$

A similar analysis could be performed on the left tail. However, this can be avoided by observing that saying that there are at most r successes is the same as saying there are at least $(n - r)$ failures. Exchanging $n - r$ for r and p for q on the right side above then yields, after simplification,

$$P[S_n \leq r] \leq \frac{(n - r)p}{(np - r)^2} \quad \text{if } r < np \tag{4.8}$$

Let us now look at the probability that the number of successes is much different from np. We expect that as n increases and the width of the binomial distribution decreases relative to its mean, then almost all of the results will fall near np. We can address this by using the above results. Let $r = n(p+\varepsilon)$. Then

$$P[S_n \geq n(p + \varepsilon)] \leq \frac{n(p + \varepsilon)q}{(n(p + \varepsilon) - np)^2} = \frac{n(p + \varepsilon)q}{(n\varepsilon)^2} \to 0$$

because the denominator grows as n^2 while the numberator grows as n. In the same way, the probability on the left tail also decreases with n, so that $P[S_n \leq n(p - \varepsilon)] \to 0$. Therefore, almost all the probability is in the central region, which is of width $n\varepsilon$. Since the location of the center is $m = np$, the ratio of the width to the center point is ε/p, which can be as small as one wishes. We have therefore established

Theorem 4.2.3 ***Law of Large Numbers****: The probability that the ratio S_n/n differs from p by less than ε in a set of n Bernoulli trials approaches unity as n increases.*

$$P\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] \to 1$$

As n increases, the probability that the average number of successes differs from p by more than ε tends to zero[2]. We find application of the law of large numbers in many areas of science and engineering. One prominent example is in Shannon's development of the noisy channel coding theorem.

4.3 Poisson Distribution

The Poisson[3] distribution can be derived as a limiting form of the binomial distribution in which n is increased without limit as the product $\lambda = np$ is kept constant. This corresponds to conducting a very large number of Bernoulli trials with the probability p of success on any one trial being very small. The Poisson distribution can also be derived directly in a manner that shows how it can be used as a model of real situations. In this sense, it stands alone and is independent of the binomial distribution. The latter insight is worthwhile, and we shall therefore invest the effort. The derivation as a limiting form of the binomial distribution is addressed in Exercise 5.

In order to derive the Poisson process directly and also motivate a model of a physical situation, we will describe a realistic experiment. Imagine that you are able to observe the arrival of photons at a detector. Your detector is designed to count the number of photons that arrive in an interval $\Delta\tau$. We will make the assumption that the probability of one photon arriving in $\Delta\tau$ is proportional to $\Delta\tau$ when $\Delta\tau$ is very small.

$$P(1; \Delta\tau) = a\Delta\tau \quad \text{for small } \Delta\tau \tag{4.9}$$

where a is a constant whose value is not yet determined. We make the second assumption that the probability that more than one photon arrives in $\Delta\tau$ is negligible when $\Delta\tau$ is very small.

$$P(0; \Delta\tau) + P(1; \Delta\tau) = 1 \quad \text{for small } \Delta\tau \tag{4.10}$$

We also assume that the number of photons that arrive in one interval is independent of the number of photons that arrive in some other non-overlapping interval. These three assumptions are all that we need to derive the Poisson distribution. Any process that fits these assumptions will therefore be modeled by the Poisson distribution.

---

[2] For further discussion of the law of large numbers see William Feller, *An Introduction to Probability Theory and its Applications*, Vol I, page 152.

[3] Siméon D. Poisson, (1781-1840).

To derive the distribution we begin by calculating the probability that zero photons arrive in a finite interval of length τ. The probability that a zero photons arrive in τ is equal to the probability that zero photons arrive in $\tau - \Delta\tau$ and no photons arrive in $\Delta\tau$. Since the intervals do not overlap, the events are independent and

$$P(0;\tau) = P(0;\tau - \Delta\tau)P(0;\Delta\tau) \tag{4.11}$$

If we substitute (4.10) and (4.9) and rearrange we find

$$\frac{P(0;\tau) - P(0;\tau - \Delta\tau)}{\Delta\tau} = -aP(0;\tau - \Delta\tau)$$

If we now let $\Delta\tau \to 0$ we have the definition of the derivative on the left. This leads to the differential equation

$$\frac{dP(0;\tau)}{d\tau} = -aP(0;\tau) \tag{4.12}$$

The solution is $P(0;\tau) = Ce^{-a\tau}$. When we apply the boundary condition $\lim_{\tau\to 0} P(0;\tau) = 1$ we find $C = 1$, so that

$$P(0;\tau) = e^{-a\tau} \tag{4.13}$$

Consider next the probability that k photons arrive in interval $\tau + \Delta\tau$. There are only two possiblities. Either k arrive in τ and 0 arrive in $\Delta\tau$ or $k-1$ arrive in τ and 1 arrives in $\Delta\tau$. Since these events are mutually exclusive,

$$P(k;\tau + \Delta\tau) = P(k;\tau)P(0;\Delta\tau) + P(k-1;\tau)P(1;\Delta\tau)$$

Now substitute for $P(0;\Delta\tau)$ and $P(1;\Delta\tau)$ and rearrange.

$$\frac{P(k;\tau + \Delta\tau) - P(k;\tau)}{\Delta\tau} + aP(k;\tau) = aP(k-1;\tau)$$

In the limit we have the differential equation

$$\frac{dP(k;\tau)}{d\tau} + aP(k;\tau) = aP(k-1;\tau) \tag{4.14}$$

This is a recursive equation that ties $P(k;\tau)$ to $P(k-1;\tau)$. To solve it we need to convert it into something we can integrate. Multiply through by $e^{a\tau}$:

$$e^{a\tau}\frac{dP(k;\tau)}{d\tau} + ae^{a\tau}P(k;\tau) = ae^{a\tau}P(k-1;\tau)$$

The term on the left can be expressed as a total derivative

$$\frac{d}{d\tau}\left(e^{a\tau}P(k;\tau)\right) = ae^{a\tau}P(k-1;\tau)$$

Now, upon integrating with respect to τ we have

$$e^{a\tau}P(k;\tau) = \int_0^{\tau} ae^{at}P(k-1;t)dt + C$$

We note that $P(k;0) = 0$ so that the constant of integration is $C = 0$. Rearranging,

$$P(k;\tau) = ae^{-a\tau}\int_0^{\tau} e^{at}P(k-1;t)dt \tag{4.15}$$

We can now apply the recursion relationship starting with $k = 1$ and making use of (4.13) to obtain $P(1;\tau)$. Then we can do the recursion again to obtain $P(2;\tau)$, and so on. We ultimately conclude that the Poisson distribution can be expressed as

$$P(k;\tau) = \frac{(a\tau)^k e^{-a\tau}}{k!} \tag{4.16}$$

The expected number of photons in τ can be obtained by finding the first moment.

$$E[k] = \sum_{k=0}^{\infty} \frac{k(a\tau)^k e^{-a\tau}}{k!} = a\tau \tag{4.17}$$

If τ corresponds to time in seconds then a corresponds to the average rate of photon arrival in photons per second. The quantity $a\tau$ corresponds to the parameter $\lambda = np$ that was discussed in connection with the binomial distribution. The connection is made by imagining that τ is divided into a very large number n of intervals and that the probability of a photon landing in any given interval is p. This corresponds to "success" in Bernoulli trials. Then the the expected number of photons in the n intervals is $np = \lambda$ and is also equal to $a\tau$. Hence, $\lambda = a\tau$. When the rate a is given and the interval τ is fixed, it is common to write the Poisson distribution as

$$P(k;\lambda) = \frac{(\lambda)^k e^{-\lambda}}{k!} \tag{4.18}$$

This is the form you will get in Exercise 5.

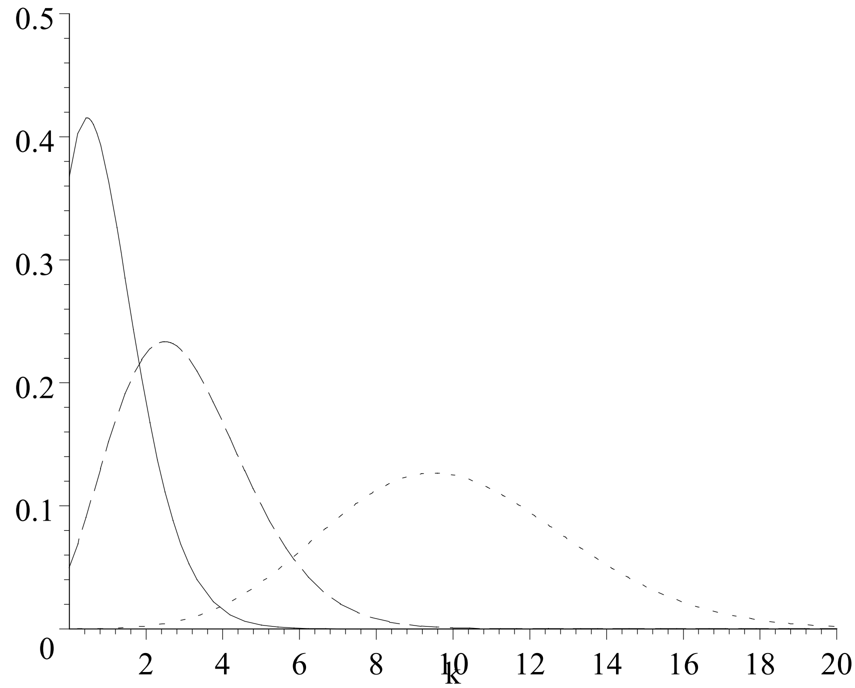


Figure 4.2: Graphs of the Poisson distribution for $\lambda = 1, 3, 10$.

The variance of the Poisson distribution is calculated by

$$\mathrm{var}(k) = E[k^2] - E^2[k] = a\tau \tag{4.19}$$

The calculation is left as an exercise. Note that the mean and the variance of a Poisson distribution are equal to each other. This is illustrated in Figure 4.2 in which the Poisson distribution is plotted for $\lambda = 1, 3, 10$. Note how the location of the peak, which is approximately equal to λ, and the width increase together.

4.3.1 Compared to Binomial Distribution

We expect the Poisson distribution to be a very close approximation to the binomial distribution for small values of p. A comparison can be made in the graphs of Figure 4.3, where we see that improvement in the match as p is decreased from 0.5 to 0.1. This would indicate that we should expect good agreement for conditions in which n is large and p small.

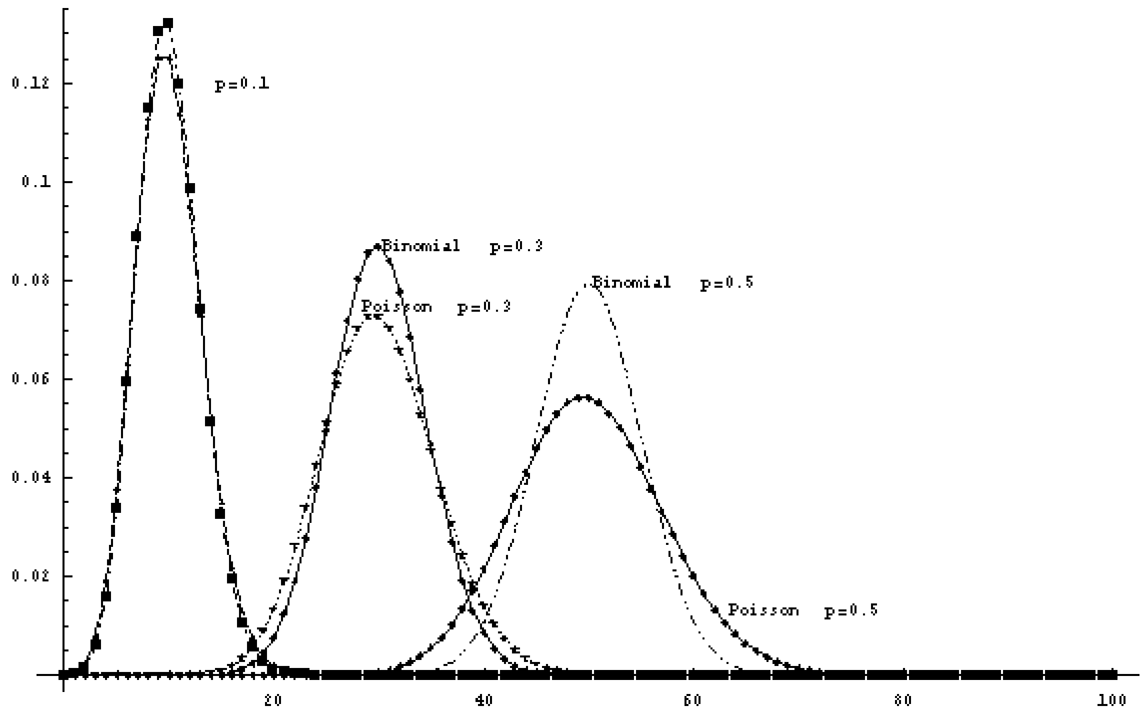


Figure 4.3: Comparison of the binomial and Poisson distributions for $n = 100$ and $p = 0.1, 0.3, 0.5$. In each case the parameter $\lambda = np$ is used in the Poisson distribution, so the graphs are plots of $b(100, k, p)$ and $P(k; \lambda = 100p)$ versus k.

We will illustrate the use of the Poisson distribution and compare it to the binomial distribution with an experiment that uses a random number generator.This example simulates the use of 10×10 array of detector elements in beam of light.

Example 4.3.1 *Suppose that a photon stream that contains n =1000 photons falls on a* 10×10 *array of detector elements. The probability of any given photon hitting any given element is* $p = 0.01$, *so that we expect about* $\lambda = np = 10$ *photons to fall on each element. We will do* $N = 1000$ *trials*

k	$b[1000, k, 0.01]$	$P(k, 10)$	m
0	0.00004	0.00005	0
1	0.00044	0.00045	0
2	0.00220	0.00227	3
3	0.00739	0.00757	8
4	0.01861	0.01892	19
5	0.03745	0.03783	31
6	0.06274	0.06306	55
7	0.08999	0.09008	92
8	0.11282	0.11260	115
9	0.12561	0.12511	126
10	0.12574	0.12511	136
11	0.11431	0.11374	121
12	0.09516	0.09478	98
13	0.07305	0.07291	64
14	0.05202	0.05208	46
15	0.03454	0.03472	42
16	0.02148	0.02170	21
17	0.01256	0.01276	10
18	0.00693	0.00709	8
19	0.00362	0.00373	3
20	0.00179	0.00187	1

Table 4.1: The number of trials (total of 1000) in which k photons strike a particular cell is in the right-hand column. The binomial and Poisson distributions are shown for comparison.

and count the number of trials in which some particular element, such as (7,7), is hit k times. The results are shown in Table 4.1. If one multiplies the binomial and Poisson probabilities by N, then the experimental results compare favorably with the distributions. Every time the experiment is done the results will be a little different. This illustrates the randomness of Poisson arrivals.

On any particular trial there will be a distribution of the photons across the array, with about $\lambda = np = 10$ photons hitting each cell. However, the distribution is quite uneven, with a standard deviation of $\sigma = \sqrt{\lambda} = \sqrt{10} = 3.16$. This is illustrated in Figure 4.4, which is the result of one such experiment. The photon count for each cell is shown by the height of the bar. Note the variation across the array. As the number of photons per cell is increased, say by a longer exposure, both σ and λ will increase, but their ratio will decrease as $1/\sqrt{n}$. That is why longer exposures appear less noisy.

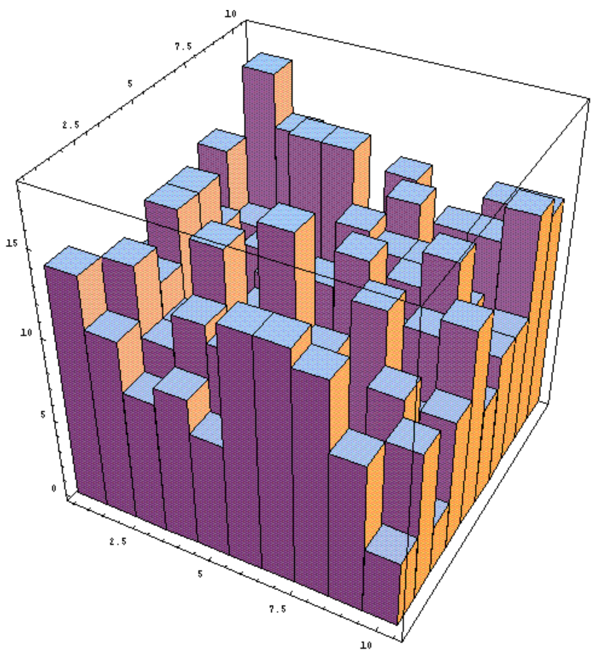


Figure 4.4: An illustration of the distribution of 1000 photons across a 10x10 array.

4.4 The Normal Approximation to the Binomial Distribution

The normal distribution, also known as the Gaussian distribution, is one of the most important in probability analysis. It arises naturally as a limit of both the binomial distribution and the Poisson distribution. It is possible to derive the relationships through limiting operations. Here we will be content with a statement of the results and some examples.

The Normal Distribution

The normal distribution is the "bell-shaped curve" that is familiar in many applications of probability and statistics. We will define it here and illustrate some of its properties before looking at its relationship to the other distributions.

Definition 4.4.1 ***Normal Distribution*** *The function defined by*

$$(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{4.20}$$

is called the normal density function. The normal distribution function is

$$\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt \tag{4.21}$$

The distribution function is often presented in a slightly different form This form, which is called the "error function" is

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-u^2} du \tag{4.22}$$

This form of the error function is built into the IDL language as the function ERRORF. In the exercises you are asked to show that the distribuition function and the error function are related by

$$\mathcal{Q}(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}}\right) \tag{4.23}$$

The normal distribution has a mean value of $\mu = 0$ and a standard deviation of $\sigma = 1$. If a random variable X with a normal distribuition is replaced by $S = \sigma X + m$ then, by a simple change of variable, the probability density function for S will be

$$f_S(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-m)^2}{2\sigma^2}} \tag{4.24}$$

We will refer to this as a Gaussian distribution with mean m and standard deviation σ. The probability distribution function is

$$F_S(s) = \mathcal{Q}\left(\frac{s-m}{\sigma}\right) \tag{4.25}$$

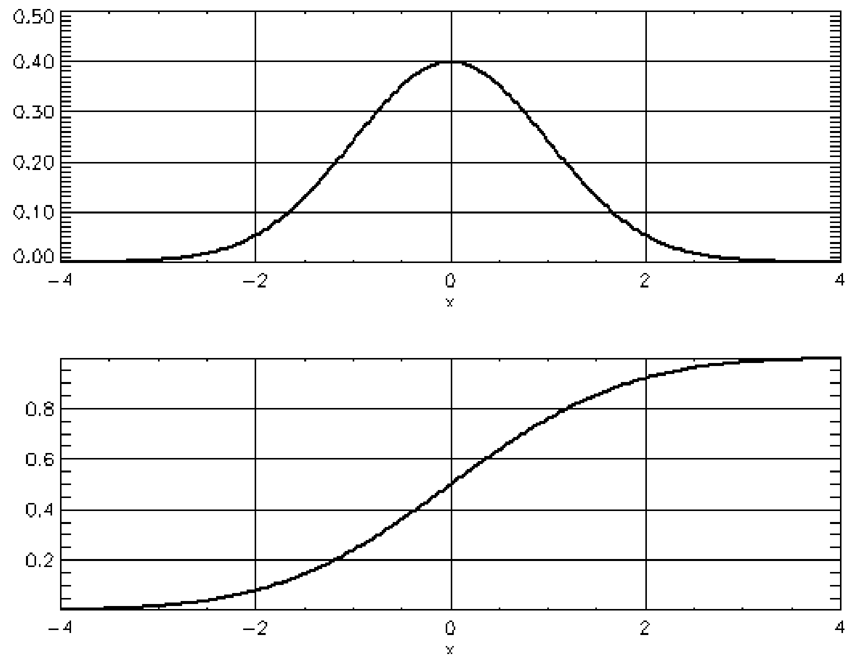


Figure 4.5: The normal density function and the normal distribution function.

4.4.1 Relationship to the Binomial Distribution

Let S_n be a random variable that is equal to the number of successes in n Bernoulli trials. Then S_n has a binomial distribution. The probability that the number of successes is between two values, a and b, is $P[a \leq S_n \leq b]$,

$$P[a \leq S_n \leq b] = \sum_{r=a}^{b} b[n, r, p] \tag{4.26}$$

The following theorem states that this probability can be computed by use of the normal distribution.

Theorem 4.4.1 ***(DeMoivre-Laplace limit theorem)*** *Let $m = np$ and $\sigma = \sqrt{np(1-p)}$. For fixed values of parameters z_1 and z_2, as $n \rightarrow \infty$,*

$$P[m + z_1\sigma \leq S_n \leq m + z_2\sigma] \rightarrow \mathcal{Q}(z_2) - \mathcal{Q}(z_1)$$

The parameters z_1 and z_2 are distances from the mean measured in units of σ. If we define a normalized random variable

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} \tag{4.27}$$

we have the equivalent probability relationship

$$P[z_1 \leq Z_n \leq z_2] \rightarrow \mathcal{Q}(z_2) - \mathcal{Q}(z_1) \tag{4.28}$$

This is an expression that is in standard units, often referred to as a z-score. To find the probability that a normal random variable falls between two limits can be found by using a single normalized table of numbers. Standardized tables of the normal probability distribution function are available, and shortened ones are included in nearly every book on probability and statistics. An equivalent table could be provided in terms of the error function. Rather than using tables we now tend to use a computer, but the principle is the same. An expression of the probability relationship above in terms of the error function is

$$P[z_1 \leq Z_n \leq z_2] \rightarrow \frac{1}{2}\text{erf}\left(\frac{z_2}{\sqrt{2}}\right) - \frac{1}{2}\text{erf}\left(\frac{z_2}{\sqrt{2}}\right) \tag{4.29}$$

The binomial, Poisson and normal probability functions give comparable results for large values of n. Graphs of the distributions are shown in Figures 4.6 and 4.7. From these figures we can see that the agreement improves both as n increases and p decreases. Because the Poisson and normal distributions are analytically convenient, we tend to prefer them to the binomial distribution for calculations.

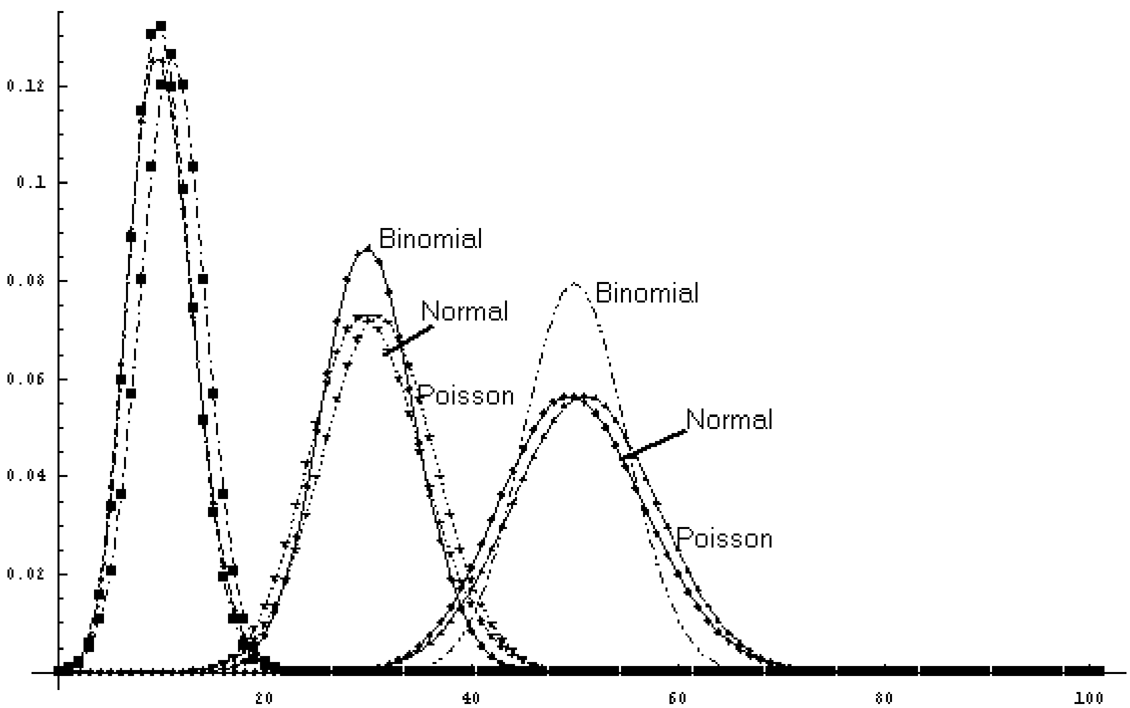


Figure 4.6: A comparison of three probability functions with $n = 100$ and $p = 0.1$, 0.3, and 0.5. The Poisson distribution is plotted with $\lambda = np$, and the normal is plotted with $m = np$, $\sigma = \sqrt{np(1-p)}$. The three distributions are very close for small values of p, and the normal and Poisson distributions are quite close even for larger values of p.

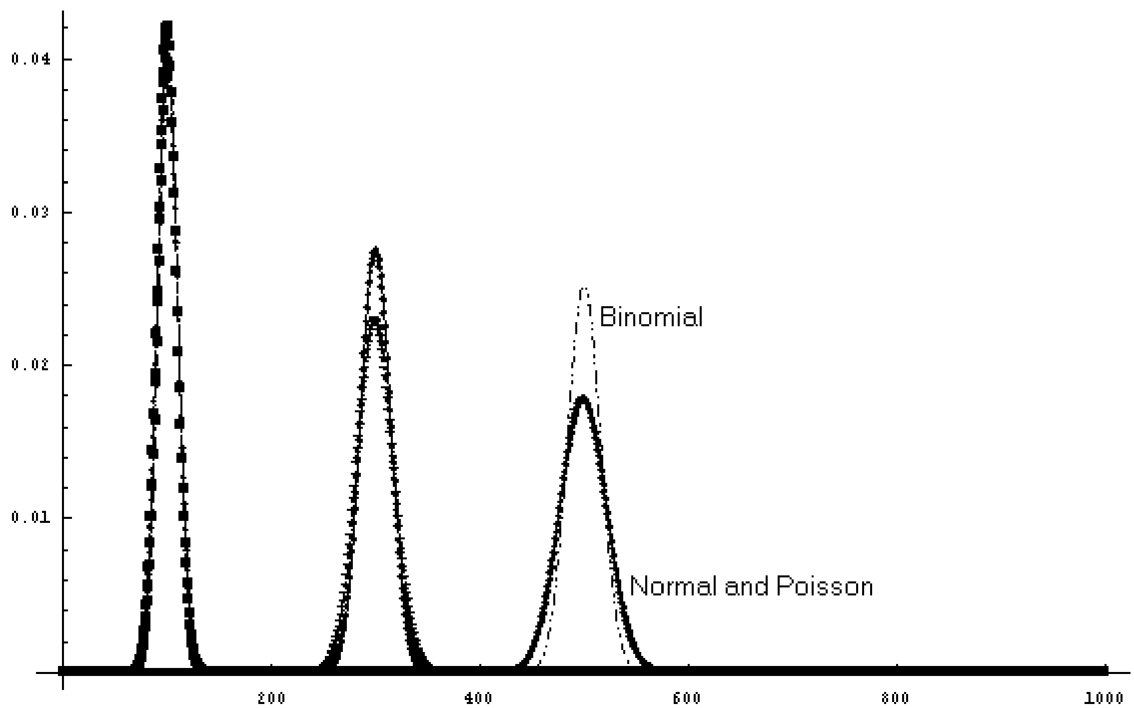


Figure 4.7: A comparison of the binomial, Poisson and normal probability functions for $n = 1000$ and $p = 0.1, 0.3, 0.5$. The normal and Poisson functions agree well for all of the values of p, and agree with the binomial function for $p = 0.1$.

4.5 Exercises

1. Show that the ratio

$$\frac{b(n,k,p)}{b(n,k-1,p)} = 1 + \frac{(n+1)p - k}{k(1-p)} \tag{4.30}$$

and thereby show that each term is greater than its predecessor if $k < (n+1)p$ and smaller than its predecessor if $k > (n+1)p$. The distribution is therefore monotonically increasing until it reaches a peak value and then is monotonically decreasing. If $(n+1)p = m$ is an integer then $b(n,m,p) = b(n,m-1,p)$ and there are two maximum points. Otherwise, there exists exactly point with index m that satisfies

$$(n+1)p - 1 < m \leq (n+1)p \tag{4.31}$$

where the distribution has its maximum value. You are asked to carry out the analysis in Exercise 2 to show that the expected number of successes in np.

2. Let S_n be the number of successes in n Bernoulli trials with p equal to the probability of success on any one trial. Show that $E[S_n] = np$. [Hint: Show that $\sum_{k=0}^{n} kb(n,k,p) = np \sum_{k=0}^{n-1} b(n-1,k,p)$].

3. Let S_n be the number of successes in n Bernoulli trials with p equal to the probability of success on any one trial. Show that $\text{var}[S_n] = np(1-p)$. [Hint: Use a technique similar to the previous problem with $\sum_{k=0}^{n} k^2 b(n,k,p) - (np)^2 = np(1-p)$ after simplification].

4. Consider a sequence of Bernoulli trials with $p = 0.5$. Determine the number of trials that you must conduct such the probability $P[S_n/n > 0.51] \leq 0.01$.

5. Show that the Poisson distribution can be obtained as a limiting form of the binomial distribution in which n is increased without limit and $np = \lambda$ so that p is reduced as n is increased so that the product is constant. [Hint: Show that the ratio $b(n,k,p)/b(n,k-1,p) \approx \lambda/k$. Show that $b(n,0,p) \approx e^{-\lambda}$ by taking $b(n,0,p) = (1-p)^n$ to the limit with $p = \lambda/n$. (Recall the definition of e.) Then, knowing $b(n,0,p)$ you can find $b(n,1,p)$, etc.]

6. Show that the probabilities in the Poisson distribution sum to unity.

7. Carry out the calculation of the sum in (4.17) to obtain $E[k] = a\tau$.

8. Carry out the calculation of the variance of the Poisson distribution in (4.19).

9. Show that the area under the normal probability density function is unity.

10. Verify the relationship between the normal probability density function and the error function, as given in (4.23).

11. Show that the mean and standard deviation of the expression in (4.24) is actually m and σ, respectively. Show by appropriate sketches how the curve changes with variations in these parameters.

12. Show that the probability distribution function for a Gaussian random variable with mean m and standard deviation σ is given by (4.25). Show by appropriate sketches how the curve changes with variations in these parameters. Find the value of s in terms of m and σ for which $F_S(s) = 0.99$.