Chapter 5
Sequences of Independent Trials with Two Outcomes

Abstract The weak and strong laws of large numbers are established for the Bernoulli scheme in Sect. 5.1. Then the local limit theorem on approximation of the binomial probabilities is proved in Sect. 5.2 using Stirling's formula (covering both the normal approximation zone and the large deviations zone). The same section also contains a refinement of that result, including a bound for the relative error of the approximation, and an extension of the local limit theorem to polynomial distributions. This is followed by the derivation of the de Moivre–Laplace theorem and its refinements in Sect. 5.3. In Sect. 5.4, the coupling method is used to prove the Poisson theorem for sums of non-identically distributed independent random indicators, together with sharp approximation error bounds for the total variation distance. The chapter ends with derivation of large deviation inequalities for the Bernoulli scheme in Sect. 5.5.

5.1 Laws of Large Numbers

Suppose we have a sequence of trials in each of which a certain event A can occur with probability p independently of the outcomes of other trials. Form a sequence of random variables as follows. Put $\xi_k = 1$ if the event A has occurred in the k-th trial, and $\xi_k = 0$ otherwise. Then $(\xi_k)_{k=1}^{\infty}$ will be a sequence of independent random variables which are identically distributed according to the Bernoulli law: $\mathbf{P}(\xi_k = 1) = p$, $\mathbf{P}(\xi_k = 0) = q = 1 - p$, $\mathbf{E}\xi_k = p$, $\text{Var}(\xi_k) = pq$. The sum $S_n = \xi_1 + \cdots + \xi_n \Subset \mathbf{B}_p^n$ is simply the number of occurrences of the event A in the first n trials. Clearly $\mathbf{E}S_n = np$ and $\text{Var}(S_n) = npq$.

The following assertion is called the *law of large numbers* for the Bernoulli scheme.

Theorem 5.1.1 *For any $\varepsilon > 0$*

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \to 0 \quad as\ n \to \infty.$$

This assertion is a direct consequence of Theorem 4.7.5. One can also obtain the following stronger result:

Theorem 5.1.2 (The Strong Law of Large Numbers for the Bernoulli Scheme)   *For any $\varepsilon > 0$, as $n \to \infty$,*

$$\mathbf{P}\left(\sup_{k \geq n} \left| \frac{S_k}{k} - p \right| > \varepsilon \right) \to 0.$$

The interpretation of this result is that the notion of probability which we introduced in Chaps. 1 and 2 corresponds to the intuitive interpretation of probability as the limiting value of the relative frequency of the occurrence of the event. Indeed, S_n/n could be considered as the relative frequency of the event A for which $\mathbf{P}(A) = p$. It turned out that, in a certain sense, S_n/n converges to p.

Proof of Theorem 5.1.2   One has

$$\mathbf{P}\left(\sup_{k \geq n} \left| \frac{S_k}{k} - p \right| > \varepsilon \right) = \mathbf{P}\left(\bigcup_{k=n}^{\infty} \left\{ \left| \frac{S_k}{k} - p \right| > \varepsilon \right\} \right)$$

$$\leq \sum_{k=n}^{\infty} \mathbf{P}\left(\left| \frac{S_k}{k} - p \right| > \varepsilon \right) \leq \sum_{k=n}^{\infty} \frac{\mathbf{E}(S_k - kp)^4}{k^4 \varepsilon^4}.$$

$$(5.1.1)$$

Here we again made use of Chebyshev's inequality but this time for the fourth moments. Expanding we find that

$$\mathbf{E}(S_k - kp)^4 = \mathbf{E}\left(\sum_{j=1}^{k} (\xi_j - p) \right)^4 = \sum_{j=1}^{k} \mathbf{E}(\xi_j - p)^4 + 6 \sum_{i < j} (\xi_i - p)^2 (\xi_j - p)^2$$

$$= k(pq^4 + qp^4) + 3k(k-1)(pq)^2 \leq k + k(k-1) = k^2.   (5.1.2)$$

Thus the probability we want to estimate does not exceed the sum

$$\varepsilon^{-4} \sum_{k=n}^{\infty} k^{-2} \to 0 \quad \text{as } n \to \infty. \qquad \square$$

It is not hard to see that we would not have found the required bound if we used Chebyshev's inequality with second moments in (5.1.1).

We could also note that one actually has much stronger bounds for $\mathbf{P}(|S_k - kp| > \varepsilon k)$ than those that we made use of above. These will be derived in Sect. 5.5.

Corollary 5.1.1   *If $f(x)$ is a continuous function on $[0, 1]$ then, as $n \to \infty$,*

$$\mathbf{E}f\left(\frac{S_n}{n} \right) \to f(p) \qquad (5.1.3)$$

uniformly in p.

Proof For any $\varepsilon > 0$,

$$
\begin{aligned}
\mathbf{E}\left| f\left(\frac{S_n}{n}\right) - f(p) \right| &\leq \mathbf{E}\left(\left| f\left(\frac{S_n}{n}\right) - f(p) \right|; \left| \frac{S_n}{n} - p \right| \leq \varepsilon \right) \\
&\quad + \mathbf{E}\left(\left| f\left(\frac{S_n}{n}\right) - f(p) \right|; \left| \frac{S_n}{n} - p \right| > \varepsilon \right) \\
&\leq \sup_{|x| \leq \varepsilon} \left| f(p+x) - f(p) \right| + \delta_n(\varepsilon),
\end{aligned}
$$

where the quantity $\delta(\varepsilon)$ is independent of p by virtue of (5.1.1), (5.1.2), and since $\delta_n(\varepsilon) \to 0$ as $n \to \infty$. $\square$

Corollary 5.1.2 *If $f(x)$ is continuous on $[0, 1]$, then, as $n \to \infty$,*

$$
\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} \to f(x)
$$

uniformly in $x \in [0, 1]$.

This relation is just a different form of (5.1.3) since

$$
\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}
$$

(see Chap. 1). This relation implies the well-known *Weierstrass theorem* on approximation of continuous functions by polynomials. Moreover, the required polynomials are given here explicitly—they are *Bernstein polynomials*.

5.2 The Local Limit Theorem and Its Refinements

5.2.1 The Local Limit Theorem

We know that $\mathbf{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}$, $q = 1 - p$. However, this formula becomes very inconvenient for computations with large n and k, which raises the question about the asymptotic behaviour of the probability $\mathbf{P}(S_n = k)$ as $n \to \infty$.

In the sequel, we will write $a_n \sim b_n$ for two number sequences $\{a_n\}$ and $\{b_n\}$ if $a_n/b_n \to 1$ as $n \to \infty$. Such sequences $\{a_n\}$ and $\{b_n\}$ will be said to be *equivalent*.

Set

$$
H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}, \qquad p^* = \frac{k}{n}. \tag{5.2.1}
$$

Theorem 5.2.1 *As $k \to \infty$ and $n - k \to \infty$,*

$$
\mathbf{P}(S_n = k) = \mathbf{P}\left(\frac{S_n}{n} = p^*\right) \sim \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} \exp\{-nH(p^*)\}. \tag{5.2.2}
$$

Proof We will make use of Stirling's formula according to which $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$ as $n \to \infty$. One has

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k q^{n-k} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}} p^k (1-p)^{n-k}$$

$$= \frac{1}{\sqrt{2\pi np^*(1-p^*)}}$$

$$\times \exp\left\{-k\ln\frac{k}{n} - (n-k)\ln\frac{n-k}{n} + k\ln p + (n-k)\ln(1-p)\right\}$$

$$= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\left\{-n\left[p^* \ln p^* + (1-p^*)\ln(1-p^*)\right.\right.$$

$$\left.\left. - p^* \ln p - (1-p^*)\ln(1-p)\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\{nH(p^*)\}. \qquad \qquad \square$$

If $p^* = k/n$ is close to p, then one can find another form for the right-hand side of (5.2.2) which is of significant interest. Note that the function $H(x)$ is analytic on the interval $(0, 1)$. Since

$$H'(x) = \ln\frac{x}{p} - \ln\frac{1-x}{1-p}, \qquad H''(x) = \frac{1}{p} + \frac{1}{1-x}, \qquad (5.2.3)$$

one has $H(p) = H'(p) = 0$ and, as $p^* - p \to 0$,[1]

$$H(p^*) = \frac{1}{2}\left(\frac{1}{p} + \frac{1}{q}\right)(p^* - p)^2 + O(|p^* - p|^3).$$

Therefore if $p^* \sim p$ and $n(p^* - p)^3 \to 0$ then

$$\mathbf{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi pq}} \exp\left\{-\frac{n}{2pq}(p^* - p)^2\right\}.$$

Putting

$$\Delta = \frac{1}{\sqrt{npq}}, \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

one obtains the following assertion.

Corollary 5.2.1 *If $z = n(p^* - p) = k - np = o(n^{2/3})$ then*

$$\mathbf{P}(S_n = k) = \mathbf{P}(S_n - np = z) \sim \varphi(z\Delta)\Delta, \qquad (5.2.4)$$

where $\varphi = \varphi_{0,1}(x)$ is evidently the density of the normal distribution with parameters $(0, 1)$.

---

[1] According to standard conventions, we will write $a(z) = o(b(z))$ as $z \to z_0$ if $b(z) > 0$ and $\lim_{z \to z_0} \frac{a(z)}{b(z)} = 0$, and $a(z) = O(b(z))$ as $z \to z_0$ if $b(z) > 0$ and $\limsup_{z \to z_0} \frac{|a(z)|}{b(z)} < \infty$.

This formula also enables one to estimate the probabilities of the events of the form $\{S_n < k\}$.

If p^* differs substantially from p, then one could estimate the probabilities of such events using the results of Sect. 1.3.

Example 5.2.1 In a jury consisting of an odd number $n = 2m + 1$ of persons, each member makes a correct decision with probability $p = 0.7$ independently of the other members. What is the minimum number of members for which the verdict rendered by the majority of jury members will be correct with a probability of at least 0.99?

Put $\xi_k = 1$ if the k-th jury member made a correct decision and $\xi_k = 0$ otherwise. We are looking for odd numbers n for which $\mathbf{P}(S_n \leq m) \leq 0.01$. It is evident that such a trustworthy decision can be achieved only for large values of n. In that case, as we established in Sect. 1.3, the probability $\mathbf{P}(S_n \leq m)$ is approximately equal to

$$\frac{(n+1-m)p}{(n+1)p-m}\mathbf{P}(S_n = m) \approx \frac{p}{2p-1}\mathbf{P}(S_n = m).$$

Using Theorem 5.2.1 and the fact that in our problem

$$p^* \approx \frac{1}{2}, \qquad H\left(\frac{1}{2}\right) = -\frac{1}{2}\ln 4p(1-p), \qquad H'\left(\frac{1}{2}\right) = \ln\left(\frac{1-p}{p}\right),$$

we get

$$\begin{aligned}\mathbf{P}(S_n \leq m) &\approx \frac{p}{2p-1}\sqrt{\frac{2}{\pi n}}\exp\left\{-nH\left(\frac{1}{2}-\frac{1}{2n}\right)\right\}\\ &\approx \frac{p}{2p-1}\sqrt{\frac{2}{\pi n}}\exp\left\{-nH\left(\frac{1}{2}\right)+\frac{1}{2}H'\left(\frac{1}{2}\right)\right\}\\ &\approx \frac{\sqrt{2\pi(1-p)}}{(2p-1)\sqrt{\pi n}}\left(\sqrt{4p(1-p)}\right)^n \approx 0.915\frac{1}{\sqrt{n}}(0.84)^{n/2}.\end{aligned}$$

On the right-hand side there is a monotonically decreasing function $a(n)$. Solving the equation $a(n) = 0.01$ we get the answer $n = 33$. The same result will be obtained if one makes use of the explicit formulae.

5.2.2   Refinements of the Local Theorem

It is not hard to bound the error of approximation (5.2.2). If, in Stirling's formula $n! = \sqrt{2\pi n}n^n e^{-n+\theta(n)}$, we make use of the well-known inequalities[2]

$$\frac{1}{12n+1} < \theta(n) < \frac{1}{12n},$$

then the same argument will give the following refinement of Theorem 5.2.1.

---

[2] See, e.g., [12], Sect. 2.9.

Theorem 5.2.2

$$\mathbf{P}(S_n = k) = \frac{1}{\sqrt{2\pi np^*(1 - p^*)}} \exp\{nH(p^*) + \theta(k, n)\}, \qquad (5.2.5)$$

where

$$\left|\theta(k, n)\right| = \left|\theta(n) - \theta(k)\theta(n - k)\right| < \frac{1}{12k} + \frac{1}{12(n - k)} = \frac{1}{12np^*(1 - p^*)}.$$
$$(5.2.6)$$

Relation (5.2.4) could also be refined as follows.

Theorem 5.2.3 *For all k such that* $|p^* - p| \leq \frac{1}{2}\min(p, q)$ *one has*

$$\mathbf{P}(S_n = k) = \varphi(z\Delta)\Delta(1 + \varepsilon(k, n)),$$

where

$$1 + \varepsilon(k, n) = \exp\left\{\vartheta\left(\frac{|z|^3 \Delta^4}{3} + \left(|z| + \frac{1}{6}\right)\Delta^2\right)\right\}, \quad |\vartheta| < 1.$$

As one can easily see from the properties of the Taylor expansion of the function e^x, the order of magnitude of the term $\varepsilon(k, n)$ in the above formulae coincides with that of the argument of the exponential. Hence it follows from Theorem 5.2.3 that for $z = k - np = o(\Delta^{-4/3})$ or, which is the same, $z = o(n^{2/3})$, one still has (5.2.4).

Proof We will make use of Theorem 5.2.2. In addition to formulae (5.2.3) one can write:

$$H^{(k)} = \frac{(-1)^k(k - 2)!}{x^{k-1}} + \frac{(k - 2)!}{(1 - x)^{k-1}}, \quad k \geq 2,$$

$$H(p^*) = \frac{1}{2pq}(p^* - p)^2 + R_1,$$

where we can estimate the residual $R_1 = \sum_{k=3}^{\infty} \frac{H^{(k)}(p)}{k!}(p^* - p)$. Taking into account that

$$\left|H^{(k)}(p)\right| \leq (k - 2)!\left(\frac{1}{p^{k-1}} + \frac{1}{q^{k-1}}\right), \quad k \geq 2,$$

and letting for brevity $|p^* - p| = \delta$, we get for $\delta \leq \frac{1}{2}\min(p, q)$ the bounds

$$|R_1| \leq \sum_{k=3}^{\infty} \frac{(k - 2)!}{k!}\left(\frac{1}{p^{k-1}} + \frac{1}{q^{k-1}}\right) \leq \frac{\delta^3}{6}\left(\frac{1}{p^2}\frac{1}{1 - \frac{\delta}{p}} + \frac{1}{q^2}\frac{1}{1 - \frac{\delta}{q}}\right)$$

$$\leq \frac{\delta}{6}\left(\frac{2}{p^2} + \frac{2}{q^2}\right) < \frac{\delta^3}{3(pq)^2}.$$

From this it follows that

$$-nH\left(p^*\right) = -\frac{(k-np)^2}{2npq} + \frac{\vartheta_1|k-np|^3}{3(npq)^2} = -\frac{z^2\Delta^2}{2} + \frac{\vartheta_1|z|^3\Delta^4}{3}, \quad |\vartheta_1| < 1.$$

(5.2.7)

We now turn to the other factors in equality (5.2.5) and consider the product $p^*(1-p^*)$. Since $-p \le 1-p-p^* \le 1-p$, we have

$$\left|p^*(1-p^*) - p(1-p)\right| = \left|(p-p^*)(1-p-p^*)\right| \le |p^*-p|\max(p,q).$$

This implies in particular that, for $|p^*-p| < \frac{1}{2}\min(p,q)$, one has

$$\left|p^*(1-p^*) - pq\right| < \frac{1}{2}pq, \qquad p^*(1-p^*) > \frac{1}{2}pq.$$

Therefore one can write along with (5.2.6) that, for the values of k indicated in Theorem 5.2.3,

$$\left|\theta(k,n)\right| < \frac{1}{6npq} = \frac{\Delta^2}{6}.$$

(5.2.8)

It remains to consider the factor $[p^*(1-p^*)]^{-1/2}$. Since for $|\gamma| < 1/2$

$$\left|\ln(1+\gamma)\right| = \left|\int_1^{1+\gamma} \frac{1}{x}dx\right| < 2|\gamma|,$$

one has for $\delta = |p^*-p| < (1/2)\min(p,q)$ the relations

$$\ln\left(p^*(1-p^*)\right) = \ln pq + \ln\left(1 + \frac{p^*(1-p^*)-pq}{pq}\right)$$

$$= \ln(pq) + \ln\left(1 - \frac{\vartheta^*\delta}{pq}\right), \quad |\vartheta^*| < \max(p,q);$$

(5.2.9)

$$\ln\left(1 - \frac{\vartheta^*\delta}{pq}\right) = -\frac{2\vartheta_2\delta}{pq}, \quad |\vartheta_2| < \max(p,q),$$

$$\left[p^*(1-p^*)\right]^{-1/2} = [pq]^{-1/2}\exp\left\{\frac{\vartheta_2\delta}{pq}\right\}.$$

Using representations (5.2.7)–(5.2.9) and the assertion of Theorem 5.2.2 completes the proof.   □

One can see from the above estimates that the bounds for ϑ in the statement of Theorem 5.2.3 can be narrowed if we consider smaller deviations $|p^*-p|$—if they, say, do not exceed the value $\alpha\min(p,q)$ where $\alpha < 1/2$.

The relations for $\mathbf{P}(S_n = k)$ that we found are the so-called *local limit theorems for the Bernoulli scheme* and their refinements.

5.2.3 The Local Limit Theorem for the Polynomial Distributions

The basic asymptotic formula given in Theorem 5.2.1 admits a natural extension to the polynomial distribution $\mathbf{B}_p^n$, $p = (p_1, \ldots, p_r)$, when, in a sequence of independent trials, in each of the trials one has not two but $r \geq 2$ possible outcomes $A_1, \ldots, A_r$ of which the probabilities are equal to $p_1, \ldots, p_r$, respectively. Let $S_n^{(j)}$ be the number of occurrences of the event A_j in n trials,

$$S_n = \left(S_n^{(1)}, \ldots, S_n^{(r)}\right), \qquad k = (k_1, \ldots, k_r), \qquad p^* = \frac{k}{n},$$

and put $H(x) = \sum x_i \ln (x_i/p_i)$, $x = (x_1, \ldots, x_r)$. Clearly, $S_n \in \mathbf{B}_p^n$. The following assertion is a direct extension of Theorem 5.2.1.

Theorem 5.2.4 *If each of the r variables $k_1, \ldots, k_r$ is either zero or tends to ∞ as $n \to \infty$ then*

$$\mathbf{P}(S_n = k) \sim (2\pi n)^{(1-r_0)/2} \left(\prod_{\substack{j=1 \\ p_j^* \neq 0}}^{r} p_j^* \right)^{-1/2} \exp\{-nH(p^*)\},$$

where r_0 is the number of variables $k_1, \ldots, k_r$ which are not equal to zero.

Proof As in the proof of Theorem 5.2.1, we will use Stirling's formula

$$n! \sim \sqrt{2\pi n}\, e^{-n} n^n$$

as $n \to \infty$. Assuming without loss of generality that all $k_j \to \infty$, $j = 1, \ldots, r$, we get

$$\mathbf{P}(S_n = k) \sim (2\pi)^{(1-r)/2} \left(\frac{n}{\prod_{j=1}^{r} k_j} \right)^{1/2} \prod_{j=1}^{r} \left(\frac{np_j}{k_j} \right)^{k_j}$$

$$= (2\pi n)^{(1-r)/2} \left(\prod_{j=1}^{r} p_j^* \right)^{-1/2} \exp\left\{ n \sum_{j=1}^{r} \frac{k_j}{n} \ln \frac{p_j n}{k_j} \right\}. \qquad \square$$

5.3 The de Moivre–Laplace Theorem and Its Refinements

Let a and b be two fixed numbers and $\zeta_n = (S_n - np)/\sqrt{npq}$. Then

$$\mathbf{P}(a < \zeta_n < b) = \sum_{a\sqrt{npq} < z < b\sqrt{npq}} \mathbf{P}(S_n - np = z).$$

If, instead of $\mathbf{P}(S_n - np = z)$, we substitute here the values $\varphi(z\Delta)\Delta$ (see Corollary 5.2.1), we will get an integral sum $\sum_{a < z\Delta < b} \varphi(z\Delta)\Delta$ corresponding to the integral $\int_a^b \varphi(x)\, dx$.

Thus relations ([5.2.4](#)) make the equality

$$\lim_{n\to\infty} \mathbf{P}(a < \zeta_n < b) = \int_a^b \varphi(x)\,dx = \Phi(b) - \Phi(a) \tag{5.3.1}$$

plausible, where $\Phi(x)$ is the normal distribution function with parameters $(0, 1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}\,dt.$$

This is the *de Moivre–Laplace theorem*, which is one of the so-called *integral limit theorems* that describe probabilities of the form $\mathbf{P}(S_n < x)$. In Chap. 8 we will derive more general integral theorems from which ([5.3.1](#)) will follow as a special case.

Theorem [5.2.3](#) makes it possible to obtain ([5.3.1](#)) together with an error bound or, in other words, *with a bound for the convergence rate*.

Let A and B be integers,

$$a = \frac{A - np}{\sqrt{npq}}, \qquad b = \frac{B - np}{\sqrt{npq}}. \tag{5.3.2}$$

Theorem 5.3.1 *Let $b > a$, $c = \max(|a|, |b|)$, and*

$$\rho = \frac{c^3 + 3c}{3}\Delta + \frac{\Delta^2}{6}.$$

If $\Delta = 1/\sqrt{npq} \le 1/2$ and $\rho \le 1/2$ then

$$\mathbf{P}(A \le S_n < B) = \mathbf{P}(a \le \zeta_n < b) = \int_a^b \varphi(t)\,dt\,(1 + \vartheta_1\Delta c)(1 + 2\vartheta_2\rho), \tag{5.3.3}$$

where $|\vartheta_i| \le 1$, $i = 1, 2$.

This theorem shows that the left-hand side in ([5.3.3](#)) can be equivalent to $\Phi(b) - \Phi(a)$ for growing a and b as well. In that case, $\Phi(b) - \Phi(a)$ can converge to 0, and knowing the *relative* error in ([5.3.1](#)) is more convenient since its smallness enables one to establish that of the absolute error as well, but not vice versa.

Proof First we note that, for all k such that $|z| = |k - np| < c\sqrt{npq}$, the conditions of Theorem [5.2.3](#) will hold. Indeed, to have the inequality $|p^* - p| < (1/2)\min(p, q)$ it suffices that $|k - np| < npq/2 = 1/(2\Delta^2)$. This inequality will hold if $c < 1/(2\Delta)$. But since $\rho \le 1/2$, one has

$$\frac{c(c^2 + 3)\Delta}{3} < 1/2, \qquad c\Delta < 1/2.$$

Thus, for each k such that $a\sqrt{npq} \le z < b\sqrt{npq}$, we can make use of Theorem [5.2.3](#) to conclude that

$$\mathbf{P}(A \le S_n < B)$$

$$= \sum_{a\sqrt{npq}\le z<b\sqrt{npq}} \mathbf{P}(S_n = k)$$

$$= \sum_{a\le z\Delta<b} \varphi(z\Delta)\Delta\left[1 + \left(\exp\left\{\vartheta\left(\frac{|z|^3\Delta^4}{3} + \left(|z| + \frac{1}{6}\right)\Delta^2\right)\right\} - 1\right)\right],$$

$$(5.3.4)$$

where $|\vartheta| < 1$. Since, for $\rho \le 1$,

$$\left|\frac{e^\rho - 1}{\rho}\right| < e - 1 < 2,$$

the absolute value of the correction term in (5.3.4) does not exceed (substituting there $z\Delta = c$)

$$\left|\exp\left\{\vartheta\left(\frac{c^3\Delta}{3} + c\Delta + \frac{\Delta^2}{6}\right)\right\} - 1\right| \le 2\vartheta\left(\frac{c^3\Delta}{3} + c\Delta + \frac{\Delta^2}{6}\right) = 2\vartheta p.$$

Therefore

$$\mathbf{P}(A \le S_n < B) = \sum_{a\le z\Delta<b} \varphi(z\Delta)\Delta[1 + 2\vartheta_1\rho], \qquad (5.3.5)$$

where $|\vartheta_1| < 1$.

Now we transform the sum on the right-hand side of the last equality. To this end, note that, for any smooth function $\varphi(x)$,

$$\left|\Delta\varphi(x) - \int_x^{x+\Delta} \varphi(t)\,dt\right| = \frac{\Delta^2}{2} \max_{x\le t\le x+\Delta} |\varphi'(t)|. \qquad (5.3.6)$$

But for the function $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ one has $\varphi'(x) = -x\varphi(x)$ and the maximum value of $\varphi(t)$ on the segment $[x, x + \Delta]$, $|x| \le c$, differs from the minimum value by not more than the factor $\exp\{c\Delta + \Delta^2/2\}$. Therefore, for $|x| \le c$, one has by virtue of (5.3.6)

$$\left|\Delta\varphi(x) - \int_x^{x+\Delta} \varphi(t)\,dt\right|$$

$$\le \frac{\Delta^2 c}{2}e^{c\Delta+\Delta^2/2} \min_{x\le t\le x+\Delta} \varphi(t) \le \frac{\Delta c}{2}e^{c\Delta+\Delta^2/2}\int_x^{x+\Delta} \varphi(t)\,dt.$$

Since $c\Delta + \Delta^2/2 < 1/2 + 1/8$, $e^{c\Delta+\Delta^2/2} \le 2$, we have the representation

$$\Delta\varphi(x) = \int_x^{x+\Delta} \varphi(t)\,dt\,(1 + \vartheta_1\Delta c), \quad |\vartheta_1| < 1.$$

Substituting this into (5.3.5) we obtain the assertion of the theorem.   □

Thus by Theorem 5.3.1 the difference

$$\left|\mathbf{P}(x \le \zeta_n < y) - \big(\Phi(y) - \Phi(x)\big)\right| \qquad (5.3.7)$$

can be effectively, yet rather roughly, bounded from above by a quantity of the order $1/\sqrt{npq}$ if $x = a$, $y = b$ (assuming that a and b are values which can be represented in the form $(k - np)\Delta$, see (5.3.2)). If x and y do not belong to the mentioned lattice with the span Δ then the error (5.3.7) will still be of the same order since, for instance, when y varies, $\mathbf{P}(x \le \zeta_n < y)$ remains constant on the semi-intervals of the form $(a + k\Delta, a + (k + 1)\Delta]$, while the function $\Phi(y) - \Phi(x)$ increases monotonically with a bounded derivative. A similar argument holds for the left end point x. It is important to note that the error order $1/\sqrt{npq}$ *cannot be improved*, for the jumps of the distribution function of ζ_n are just of this order of magnitude by Theorem 5.2.2.

Theorem 5.3.1 enables one to use the normal approximation for $\mathbf{P}(x \le \zeta_n < y)$ in the so-called *large deviations range* as well, when both x and y grow in absolute value and are of the same sign. In that case, both $\Phi(y) - \Phi(x)$ and the probability to be approximated tend to zero. Therefore the approximation can be considered satisfactory *only if*

$$\frac{\mathbf{P}(x \le \zeta_n < y)}{(\Phi(y) - \Phi(x))} \to 1. \tag{5.3.8}$$

As Theorem 5.3.1 shows, this convergence will take place if

$$c = \max(|x|, |y|) = o(\Delta^{-1/3})$$

or, which is the same, $c = o(n^{1/6})$. For more details about large deviation probabilities, see Chap. 9.

For larger values of c, as one could verify using Theorem 5.2.1, relation (5.3.8) will, generally speaking, not hold.

In conclusion we note that since

$$\mathbf{P}(|\zeta_n| > b) \to 0$$

as $b \to \infty$, it follows immediately from Theorem 5.3.1 that, for any fixed y,

$$\lim_{n \to \infty} \mathbf{P}(\zeta_n < y) = \Phi(y).$$

Later we will show that this assertion remains true under much wider assumptions, when ζ_n is a scaled sum of arbitrary distributed random variables having finite variances.

5.4   The Poisson Theorem and Its Refinements

5.4.1   Quantifying the Closeness of Poisson Distributions to Those of the Sums S_n

As we saw from the bounds in the last section, the de Moivre–Laplace theorem gives a good approximation to the probabilities of interest if the number npq (the variance of S_n) is large. This number will grow together with n if p and q are fixed

positive numbers. But what will happen in a problem where, say, $p = 0.001$ and $n = 1000$ so that $np = 1$? Although n is large here, applying the de Moivre–Laplace theorem in such a problem would be meaningless. It turns out that in this case the distribution $\mathbf{P}(S_n = k)$ can be well approximated by the Poisson distribution $\mathbf{\Pi}_\mu$ with an appropriate parameter value μ (see Sect. 5.4.2). Recall that

$$\mathbf{\Pi}_\mu(B) = \sum_{0 \leq k \in B} e^{-\mu} \frac{\mu^k}{k!}.$$

Put $np = \mu$.

Theorem 5.4.1 *For all sets B,*

$$\left| \mathbf{P}(S_n \in B) - \mathbf{\Pi}_\mu(B) \right| \leq \frac{\mu^2}{n}.$$

We could prove this assertion in the same way as the local theorem, making use of the explicit formula for $\mathbf{P}(S_n = k)$. However, we can prove it in a simpler and nicer way which could be called the *common probability space method*, or *coupling method*. The method is often used in research in probability theory and consists, in our case, of constructing on a common probability space random variables S_n and S_n^*, the latter being as close to S_n as possible and distributed according to the Poisson distribution.

It is also important that the common probability space method admits, without any complications, extension to the case of *non-identically* distributed random variables, when the probability of getting 1 in a particular trial depends on the number of the trial. Thus we will now prove a more general assertion of which Theorem 5.4.1 is a special case.

Assume that we are given a sequence of independent random variables $\xi_1, \ldots, \xi_n$, such that $\xi_j \Subset \mathbf{B}_{p_j}$. Put, as above, $S_n = \sum_{j=1}^n \xi_j$. The theorem we state below is intended for approximating the probability $\mathbf{P}(S_n = k)$ when p_j are small and the number $\mu = \sum_{j=1}^n p_j$ is "comparable" with 1.

Theorem 5.4.2 *For all sets B,*

$$\left| \mathbf{P}(S_n \in B) - \mathbf{\Pi}_\mu(B) \right| \leq \sum_{j=1}^n p_j^2.$$

To prove this theorem we will need an important "stability" property of the Poisson distribution.

Lemma 5.4.1 *If η_1 and η_2 are independent, $\eta_2 \Subset \mathbf{\Pi}_{\mu_1}$ and $\eta_2 \Subset \mathbf{\Pi}_{\mu_2}$, then*[3]

$$\eta_1 + \eta_2 \Subset \mathbf{\Pi}_{\mu_1 + \mu_2}.$$

---

[3]This fact will also easily follow from the properties of characteristic functions dealt with in Chap. 7.

Proof By the total probability formula,

$$\mathbf{P}(\eta_1 + \eta_2 = k) = \sum_{j=0}^{k} \mathbf{P}(\eta_1 = j)\mathbf{P}(\eta_2 = k - j)$$

$$= \sum_{j=0}^{k} \frac{\mu_1^j e^{-\mu_1}}{j!} \cdot \frac{\mu_2^{k-j} e^{-\mu_2}}{(k-j)!} = \frac{1}{k!} e^{-(\mu_1+\mu_2)} \sum_{j=0}^{k} \binom{k}{j} \mu_1^j \mu_2^{k-j}$$

$$= \frac{(\mu_1 + \mu_2)^k e^{-(\mu_1+\mu_2)}}{k!}. \qquad \square$$

Proof of Theorem 5.4.2 Let $\omega_1, \ldots, \omega_n$ be independent random variables, each being the identity function ($\xi(\omega_k) = \omega_k$) on the unit interval with the uniform distribution. We can assume that the vector $\omega = (\omega_1, \ldots, \omega_n)$ is given as the identity function on the unit n-dimensional cube Ω with the uniform distribution.

Now construct the random variables ξ_j and ξ_j^* on Ω as follows:

$$\xi_j(\omega) = \begin{cases} 0 & \text{if } \omega_j < 1 - p_j, \\ 1 & \text{if } \omega_j \geq 1 - p_j, \end{cases} \qquad \xi_j^*(\omega) = \begin{cases} 0 & \text{if } \omega_j < e^{-p_j}, \\ k \geq 1 & \text{if } \omega_j \in [\pi_{k-1}, \pi_k), \end{cases}$$

where $\pi_k = \sum_{m \leq k} e^{-p_j} \frac{(p_j)^m}{m!}$, $k = 0, 1, \ldots$.

It is evident that the $\xi_j(\omega)$ are independent and $\xi_j(\omega) \Subset \mathbf{B}_{p_j}$; $\xi_j^*(\omega)$ are also jointly independent with $\xi_j^*(\omega) \Subset \mathbf{\Pi}_{p_j}$. Now note that since $1 - p_j \leq e^{-p_j}$ one has $\xi_j(\omega) \neq \xi_j^*(\omega)$ only if $\omega_j \in [1 - p_j, e^{-p_j})$ or $\omega_j \in [e^{-p_j} + p_j e^{-p_j}, 1]$. Hence

$$\mathbf{P}(\xi_j \neq \xi_j^j) = (e^{-p_j} - 1 + p_j) + (1 - e^{-p_j} - p_j e^{-p_j}) = p_j(1 - e^{-p_j}) \leq p_j^2$$

and

$$\mathbf{P}(S_n \neq S_n^*) \leq \mathbf{P}\left(\bigcup_j |\xi_j \neq \xi_j^*|\right) \leq \sum p_j^2,$$

where $S_n^* = \sum_{j=1}^{n} \xi_j^* \Subset \mathbf{\Pi}_\mu$.

Now we can write

$$\mathbf{P}(S_n \in B) = \mathbf{P}(S_n \in B, S_n = S_n^*) + \mathbf{P}(S_n \in B, S_n \neq S_n^*)$$
$$= \mathbf{P}(S_n^* \in B) - \mathbf{P}(S_n^* \in B, S_n \neq S_n^*) + \mathbf{P}(S_n \in B, S_n \neq S_n^*),$$

so that

$$\left| \mathbf{P}(S_n \in B) - \mathbf{P}(S_n^* \in B) \right|$$
$$\leq \left| \mathbf{P}(S_n^* \in B, S_n \neq S_n^*) - \mathbf{P}(S_n \in B, S_n \neq S_n^*) \right| \leq \mathbf{P}(S_n \neq S_n^*). \quad (5.4.1)$$

The assertion of the theorem follows from this in an obvious way. $\qquad \square$

Remark 5.4.1 One can give other common probability space constructions as well. One of them will be used now to show that there exists a better Poisson approximation to the distribution of S_n.

Namely, let $\xi_j^*(\omega)$ be independent random variables distributed according to the Poisson laws with parameters $r_j = -\ln(1 - p_j) \geq p_j$, so that $\mathbf{P}(\xi_j^* = 0) = e^{-r_j} = 1 - p_j$. Then $\xi_j(\omega) = \min\{1, \xi_j^*(\omega)\} \in \mathbf{B}_{p_j}$ and, moreover,

$$\mathbf{P}\left(\bigcup_{j=1}^{n}\{\xi_j(\omega) \neq \xi_j^*(\omega)\}\right) \leq \sum_{j=1}^{n}\mathbf{P}\left(\xi_j^*(\omega) \geq 2\right) = \sum_{j=1}^{n}\left(1 - e^{-r_j} - r_j e^{-r_j}\right).$$

But for $r = -\ln(1 - p)$ one has the inequality

$$1 - e^{-r} - re^{-r} = p + (1 - p)\ln(1 - p) \leq p + (1 - p)\left(-p - \frac{p^2}{2}\right)$$

$$= \frac{p^2}{2}(1 + p).$$

Hence for the new Poisson approximation we have

$$\mathbf{P}\left(S_n^* \neq S_n\right) \leq \frac{1}{2}\sum_{j=1}^{n} p_j^2(1 + p_j).$$

Putting $\lambda = -\sum_{j=1}^{n}\ln(1 - p_j) \geq \sum_{j=1}^{n} p_j$, the same argument as above will lead to the bound

$$\sup_{B}\left|\mathbf{P}(S_n \in B) - \mathbf{\Pi}_\lambda(B)\right| \leq \frac{1}{2}\sum_{j=1}^{n} p_j^2(1 + p_j).$$

This bound of the rate of approximation given by the Poisson distribution with a "slightly shifted" parameter is better than that obtained in Theorem 5.4.2. Moreover, one could note that, in the new construction, $\xi_j \leq \xi_j^*$, $S_n \leq S_n^*$, and consequently

$$\mathbf{P}(S_n \geq k) \leq \mathbf{P}\left(S_n^* \geq k\right) = \mathbf{\Pi}_\lambda\left([k, \infty)\right).$$

5.4.2 The Triangular Array Scheme. The Poisson Theorem

Now we will return back to the case of identically distributed ξ_k. To obtain from Theorem 5.4.2 a limit theorem of the type similar to that of the de Moivre–Laplace theorem (see (5.3.1)), one needs a somewhat different setup. In fact, to ensure that np remains bounded as n increases, $p = \mathbf{P}(\xi_k = 1)$ needs to converge to zero which cannot be the case when we consider a fixed sequence of random variables $\xi_1, \xi_2, \ldots$.

We introduce a sequence of rows (of growing length) of random variables:

$$
\begin{array}{ccccc}
\xi_1^{(1)}; & & & & \\
\xi_1^{(2)}, & \xi_2^{(2)}; & & & \\
\xi_1^{(3)}, & \xi_2^{(3)}, & \xi_1^{(1)}; & & \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\xi_1^{(n)}, & \xi_2^{(n)}, & \xi_3^{(n)}, & \ldots, & \xi_n^{(n)}.
\end{array}
$$

This is the so-called *triangular array scheme*. The superscript denotes the row number, while the subscript denotes the number of the variable in the row.

Assume that the variables $\xi_k^{(n)}$ in the n-th row are independent and $\xi_k^{(n)} \Subset \mathbf{B}_{p_n}$, $k = 1, \ldots, n$.

Corollary 5.4.1 (The Poisson theorem)   *If $np_n \to \mu > 0$ as $n \to \infty$ then, for each fixed k,*

$$\mathbf{P}(S_n = k) \to \mathbf{\Pi}_\mu(\{k\}), \tag{5.4.2}$$

where $S_n = \xi_1^{(n)} + \cdots + \xi_n^{(n)}$.

Proof This assertion is an immediate corollary of Theorem 5.4.1. It can also be obtained directly, by noting that it follows from the equality

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

that

$$\mathbf{P}(S_n = 0) = e^{n \ln(1-p)} \sim e^{-\mu}, \qquad \frac{\mathbf{P}(S_n = k+1)}{\mathbf{P}(S_n = k)} = \frac{n-k}{k+1} \frac{p}{1-p} \sim \frac{\mu}{k+1}. \quad \square$$

Theorem 5.4.2 implies an analogue of the Poisson theorem in a more general case as well, when the $\xi_j^{(n)}$ are not necessarily identically distributed[4] and can take values different from 0 and 1.

Corollary 5.4.2 *Assume that $p_{jn} = \mathbf{P}(\xi_j^{(n)} = 1)$ depend on n and j so that*

$$\max_j p_{jn} \to 0, \qquad \sum_{j=1}^n p_{jn} \to \mu > 0, \qquad \mathbf{P}(\xi_j^{(n)} = 0) = 1 - p_{jn} + o(p_{jn}).$$

Then (5.4.2) holds.

Proof To prove the corollary, one has to use Theorem 5.4.2 and the fact that

$$\mathbf{P}\left(\bigcup_{j=1}^n \{\xi_j^{(n)} \neq 0, \xi_j^{(n)} \neq 1\}\right) \leq \sum_{j=1}^n o(p_{jn}) = o(1),$$

which means that, with probability tending to 1, all the variables $\xi_j^{(n)}$ assume the values 0 and 1 only. $\quad \square$

One can clearly obtain from Theorems 5.4.1 and 5.4.2 somewhat stronger assertions than the above. In particular,

$$\sup_B \left|\mathbf{P}(S_n \in B) - \mathbf{\Pi}_\mu(B)\right| \to 0 \quad \text{as } n \to \infty.$$

---

[4]An extension of the de Moivre–Laplace theorem to the case of non-identically distributed random variables is contained in the central limit theorem from Sect. 8.4.

Note that under the assumptions of Theorem 5.4.1 this convergence will also take place in the case where $np \to \infty$ but only if $np^2 \to 0$. At the same time, the refinement of the de Moivre–Laplace theorem from Sect. 5.3 shows that the normal approximation for the distribution of S_n holds if $np \to \infty$ (for simplicity we assume that $p \le q$ so that $npq \ge \frac{1}{2}np \to \infty$).

Thus there exist sequences $p \in \{p : np \to \infty, \ np^2 \to 0\}$ such that *both the normal and the Poisson approximations* are valid. In other words, the domains of applicability of the normal and Poisson approximations *overlap*.

We see further from Theorem 5.4.1 that the convergence rate in Corollary 5.4.1 is determined by a quantity of the order of n^{-1}. Since, as $n \to \infty$,

$$\mathbf{P}(S_n = 0) - \mathbf{\Pi}_\mu(\{0\}) = e^{n \ln(1-p)} - e^{-\mu} \sim \frac{\mu^2}{2\pi} e^{-\mu},$$

this estimate cannot be substantially improved. However, for large k (in the large deviations range, say) such an estimate for the difference

$$\mathbf{P}(S_n = k) - \mathbf{\Pi}_\mu(\{k\})$$

becomes rough. (This is because, in (5.4.1), we neglected not only the different signs of the correction terms but also the rare events $\{S_n = k\}$ and $\{S_n^* = k\}$ that appear in the arguments of the probabilities.) Hence we see, as in Sect. 5.4, the necessity for having approximations of which both *absolute* and *relative* errors are small.

Now we will show that the *asymptotic equivalence relations*

$$\mathbf{P}(S_n = k) \sim \mathbf{\Pi}_\mu(\{k\})$$

remain valid when k and μ grow (along with n) in such a way that

$$k = o\big(n^{2/3}\big), \qquad \mu = o\big(n^{2/3}\big), \qquad |k - \mu| = o(\sqrt{n}).$$

Proof Indeed,

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1 - p)^{n-k}$$

$$= \frac{(nk)^k}{k!} e^{-pn} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)(1 - p)^{n-k} e^{pn}$$

$$= \mathbf{\Pi}_\mu(\{k\}) e^{\varepsilon(k,n)}.$$

Thus we have to prove that, for values of k and μ from the indicated range,

$$\varepsilon(k, n) := \ln\left[\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)(1 - p)^{n-k} e^{pn}\right] = o(1). \qquad (5.4.3)$$

We will obtain this relation together with the form of the correction term. Namely, we will show that

$$\varepsilon(k, n) = \frac{k - (k - \mu)^2}{2n} + O\left(\frac{k^3 + \mu^3}{n^2}\right), \qquad (5.4.4)$$

and hence

$$\mathbf{P}(S_n = k) = \left(1 + \frac{k - (k - \mu)^2}{2n} + O\left(\frac{k^3 + \mu^3}{n^2}\right)\right)\Pi_\mu(\{k\}).$$

We make use of the fact that, as $\alpha \to 0$,

$$\ln(1 - \alpha) = -\alpha - \frac{\alpha^2}{2} + O(\alpha^3).$$

Then relations (5.4.3) and (5.4.4) will follow from the equalities

$$\sum_{j=1}^{k-1}\ln\left(1 - \frac{j}{n}\right) = -\sum_{j=1}^{k-1}\frac{j}{n} + O\left(\frac{k^3}{n^2}\right) = -\frac{k(k-1)}{2n} + O\left(\frac{k^3}{n^2}\right),$$

$$(n - k)\ln(1 - p) + pn = (n - k)\left(-p - \frac{p^2}{2} + O(p^3)\right) + pn$$

$$= -\frac{\mu^2}{2n} + \frac{k\mu}{n} + O\left(\frac{\mu^3}{n^2}\right). \qquad \square$$

In conclusion we note that the approximate Poisson formula

$$\mathbf{P}(S_n = k) \approx \frac{\mu^k}{k!}e^{-\mu}$$

is widely used in various applications and has, as experience and the above estimates show, a rather high accuracy even for moderate values of n.

Now we consider several examples of the use of the de Moivre–Laplace and Poisson theorems for approximate computations.

Example 5.4.1 Suppose we are given 10^4 packets of grain. It is known that there are 5000 tagged grains in the packets. What is the probability that, in a particular fixed packet, there is at least one tagged grain? We can assume that the tagged grains are distributed to packets at random. Then the probability that a particular tagged grain will be in the chosen packet is $p = 10^{-4}$. Since there are 5000 such grains, this will be the number of trials, i.e. $n = 5000$. Define a random variable ξ_k as follows: $\xi_k = 1$ if the k-th grain is in the chosen packet, and $\xi_k = 0$ otherwise. Then

$$S_{5000} = \sum_{k=1}^{5000}\xi_k$$

will be the number of tagged grains in our packet. By Theorem 5.4.1, $\mathbf{P}(S_{5000} = 0) \approx e^{-np} = e^{-0.5}$ so that the desired probability is approximately equal to $1 - e^{-0.5}$. The accuracy of this relation turns out to be rather high (by Theorem 5.4.1, the error does not exceed $2^{-1} \times 10^{-4}$). If we used the Poisson theorem instead of Theorem 5.4.1, we would have to imagine a triangular array of Bernoulli random variables, our ξ_k constituting the 5000-th row of the array. Moreover, we would assume that, for the n-th row, one has $np_n = 0.5$. Thus the conditions of the Poisson theorem would be met and we could make use of the limit theorem to find the approximate equality we have already obtained.

Example 5.4.2 A similar argument can be used in the following problem. There are n dangerous bacteria in a reservoir of capacity V from which we take a sample of volume $v \ll V$. What is the probability that we will find the bacteria in the test sample?

One usually assumes that the probability p that any given bacterium will be in the test sample is equal to the ratio v/V. Moreover, it is also assumed that the presence of a given bacterium in the sample does not depend on whether the remaining $n-1$ bacteria are in the test sample or not. In other words, one usually postulates that the mechanism of bacterial transfer into the test sample is equivalent to a sequence of n independent trials with "success" probability equal to $p = v/V$ in each trial.

Introducing random variables ξ_k as above, we obtain a description of the number of bacteria in the test sample by the sum $S_n = \sum_{k=1}^{n} \xi_k$ in the Bernoulli scheme. If nv is comparable in magnitude with V then by the Poisson theorem the desired probability will be equal to

$$\mathbf{P}(S_n > 0) \approx 1 - e^{-nv/V}.$$

Similar models are also used to describe the number of visible stars in a certain part of the sky far away from the Milky Way. Namely, it is assumed that if there are n visible stars in a region R then the probability that there are k visible stars in a subregion $r \subset R$ is

$$\binom{n}{k} p^k (1 - p)^k,$$

where p is equal to the ratio $S(r)/S(R)$ of the areas of the regions r and R respectively.

Example 5.4.3 Suppose that the probability that a newborn baby is a boy is constant and equals 0.512 (see Sect. 3.4.1).

Consider a group of 10^4 newborn babies and assume that it corresponds to a series of 10^4 independent trials of which the outcomes are the events that either a boy or girl is born. What is the probability that the number of boys among these newborn babies will be greater than the number of girls by at least 200?

Define random variables as follows: $\xi_k = 1$ if the k-th baby is a boy and $\xi_k = 0$ otherwise. Then $S_n = \sum_{k=1}^{10^4} \xi_k$ is the number of boys in the group. The quantity $npq \sim 2.5 \times 10^3$ is rather large here, hence applying the integral limit (de Moivre–Laplace) theorem we obtain for the desired probability the value

$$\mathbf{P}(S_n \geq 5100) = 1 - \mathbf{P}\left(\frac{S_n - np}{\sqrt{npq}} < \frac{5100 - 5120}{\sqrt{2500}} \right)$$
$$\approx 1 - \Phi(-20/50) = 1 - \Phi(-0.4) \approx 0.66.$$

To find the numerical values of $\Phi(x)$ one usually makes use of suitable statistical computer packages or calculators.

In our example, $\Delta = 1/\sqrt{npq} \approx 1/50$, and a satisfactory approximation by the de Moivre–Laplace formula will certainly be ensured (see Theorem 5.3.1) for $c \le 2.5$.

If, however, we have to estimate the probability that the proportion of boys exceeds 0.55, we will be dealing with large deviation probabilities when to estimate $\mathbf{P}(S_n > 5500)$ one would rather use the approximate relation obtained in Sect. 1.3 by virtue of which ($k = 0.45n$, $q = 0.488$) one has

$$\mathbf{P}(S_n > 5500) \approx \frac{(n+1-k)q}{(n+1)q-k}\mathbf{P}(S_n = 5500).$$

Applying Theorem 5.2.1 we find that

$$\mathbf{P}(S_n > 5500) \approx \frac{0.55q}{q-0.45}\frac{1}{\sqrt{2\pi n 0.25}}e^{-nH(0.55)} \le \frac{1}{5}e^{-25} < 10^{-11}.$$

Thus if we assume for a moment that 100 million babies are born on this planet each year and group them into batches of 10 thousand, then, to observe a group in which the proportion of boys exceeds the mean value by just 3.8 % we will have to wait, on average, 10 million years (see Example 4.1.1 in Sect. 4.1).

It is clear that the normal approximation can be used for numerical evaluation of probabilities for the problems from Example 5.4.3 provided that the values of np are large.

5.5 Inequalities for Large Deviation Probabilities in the Bernoulli Scheme

In conclusion of the present chapter we will derive several useful inequalities for the Bernoulli scheme. In Sect. 5.2 we introduced the function

$$H(x) = x \ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p},$$

which plays an important role in Theorems 5.2.1 and 5.2.2 on the asymptotic behaviour of the probability $\mathbf{P}(S_n = k)$. We also considered there the basic properties of this function.

Theorem 5.5.1 *For $z \ge 0$,*

$$\begin{aligned}\mathbf{P}(S_n - np \ge z) &\le \exp\{-nH(p+z/n)\},\\ \mathbf{P}(S_n - np \le -z) &\le \exp\{-nH(p-z/n)\}.\end{aligned} \tag{5.5.1}$$

Moreover, for all p,

$$H(p+x) \ge 2x^2, \tag{5.5.2}$$

so that each of the probabilities in (5.5.1) does not exceed $\exp\{-2z^2/n\}$ for any p.

To compare it with assertion (5.2.2) of Theorem 5.2.1, the first inequality from Theorem 5.5.1 can be re-written in the form

$$\mathbf{P}\left(\frac{S_n}{n} \geq p^*\right) \leq \exp\{-nH(p^*)\}.$$

The inequalities (5.5.1) are close, to some extent, to the de Moivre–Laplace theorem since, for $z = o(n^{2/3})$,

$$-nH\left(p + \frac{z}{n}\right) = -\frac{z^2}{2npq} + o(1).$$

The last assertion, together with (5.5.2), can be interpreted as follows: deviating by z or more from the mean value np has the maximum probability when $p = 1/2$.

If $z/\sqrt{n} \to \infty$, then both probabilities in (5.5.1) converge to zero as $n \to \infty$ for they correspond to large deviations of the sum S_n from the mean np. As we have already said, they are called *large deviation probabilities*.

Proof of Theorem 5.5.1 In Corollary 4.7.2 of the previous chapter we established the inequality

$$\mathbf{P}(\xi \geq x) \leq e^{-\lambda x}\mathbf{E}e^{\lambda \xi}.$$

Applying it to the sum S_n we get

$$\mathbf{P}(S_n \geq np + z) \leq e^{-\lambda(np+z)}\mathbf{E}e^{\lambda S_n}.$$

Since $\mathbf{E}e^{\lambda S_n} = \prod_{k=1}^{n} \mathbf{E}e^{\lambda \xi_k}$ and the random variables $e^{\lambda \xi_k}$ are independent,

$$\mathbf{E}e^{\lambda S_n} = \prod_{k=1}^{n} \mathbf{E}e^{\lambda \xi_k} = \left(pe^{\lambda} + q\right)^n = \left(1 + p\left(e^{\lambda} - 1\right)\right)^n,$$

$$\mathbf{P}(S_n \geq np + z) \leq \left[\left(1 + p\left(e^{\lambda} - 1\right)\right)e^{-\lambda(p+\alpha)}\right]^n, \quad \alpha = z/n.$$

The expression in brackets is equal to

$$\mathbf{E}e^{-\lambda[\xi_k - (p+\alpha)]} = pe^{\lambda(1-p-\alpha)} + (1-p)e^{-\lambda(p+\alpha)}.$$

Therefore, being the sum of two convex functions, it is a convex function of λ. The equation for the minimum point $\lambda(\alpha)$ of the function has the form

$$-(p - \alpha)\left(1 + p\left(e^{\lambda} - 1\right)\right) + pe^{\lambda} = 0,$$

from which we find that

$$e^{\lambda(\alpha)} = \frac{(p+\alpha)q}{p(q-\alpha)},$$

$$\left(1 + p\left(e^{\lambda(\alpha)} - 1\right)\right)e^{-\lambda(\alpha)(p+\alpha)} = \frac{q}{q-\alpha}\left[\frac{p(q-\alpha)}{(p+\alpha)q}\right]^{p+\alpha}$$

$$= \frac{p^{p+\alpha}q^{q-\alpha}}{(p+\alpha)^{p+\alpha}(q-\alpha)^{q-\alpha}}$$

$$= \exp\left\{-(p+\alpha)\ln\frac{p+\alpha}{p} - (q-\alpha)\ln\frac{q-\alpha}{q}\right\}$$

$$= \exp\{-H(p+\alpha)\}.$$

The first of the inequalities (5.5.1) is proved. The second inequality follows from the first if we consider the latter as the inequality for the number of zeros.

It follows further from (5.2.1) that $H(p) = H'(p) = 0$ and $H''(x) = 1/x(1 - x)$. Since the function $x(1 - x)$ attains its maximum value on the interval $[0, 1]$ at the point $x = 1/2$, one has $H''(x) \geq 4$ and hence

$$H(p + \alpha) \geq \frac{\alpha^2}{2} \cdot 4 = 2\alpha^2. \qquad \square$$

For analogues of Theorem 5.5.1 for sums of arbitrary random variables, see Chap. 9 and Appendix 8. Example 9.1.2 shows that the function $H(\alpha)$ is the so-called *deviation function* for the Bernoulli scheme. This function is important in describing large deviation probabilities.