Chapter 20
Functional Limit Theorems

Abstract The chapter begins with Sect. 20.1 presenting the classical Functional Central Limit Theorem in the triangular array scheme. It establishes not only convergence of the distributions of the scaled trajectories of random walks to that of the Wiener process, but also convergence rates for Lipshchitz sets and distribution functions of Lipshchitz functionals in the case of finite third moments when the Lyapunov condition is met. Section 20.2 uses the Law of the Iterated Logarithm for the Wiener process to establish such a low for the trajectory of a random walk with independent non-identically distributed jumps. Section 20.3 is devoted to proving convergence to the Poisson process of the processes of cumulative sums of independent random indicators with low success probabilities and also that of the so-called thinning renewal processes.

20.1 Convergence to the Wiener Process

We have already pointed out in Sect. 19.2 that the Wiener processes are, in a certain sense, limiting to random polygons with vertices at the points $(k/n, S_k/\sqrt{n})$, where $S_k = \xi_1 + \cdots + \xi_k$ are partial sums of independent identically distributed random variables $\xi_1, \xi_2, \ldots$ with zero means and finite variances. Now we will give a more precise and general meaning to this statement.

Let

$$\xi_{1,n}, \ldots, \xi_{n,n} \tag{20.1.1}$$

be independent random variables in the triangular array scheme (see Sects. 8.3, 8.4),

$$\zeta_{k,n} := \sum_{j=1}^{k} \xi_{j,n}, \qquad \mathbf{E}\xi_{k,n} = 0, \qquad \mathbf{E}\xi_{k,n}^2 = \sigma_{k,n}^2,$$

that have finite third moments $\mathbf{E}|\xi_{k,n}|^3 = \mu_{k,n} < \infty$.

We will assume without loss of generality (see Sect. 8.4) that

$$\operatorname{Var}(\zeta_{n,n}) = \sum_{j=1}^{n} \sigma_{j,n}^2 = 1.$$

A.A. Borovkov, *Probability Theory*, Universitext,

DOI 10.1007/978-1-4471-5201-9_20,

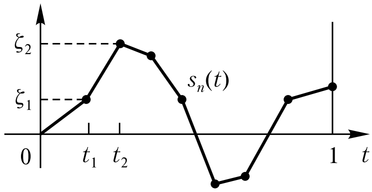

Fig. 20.1 The random polygon $s_n(t)$ constructed from the random walk $\zeta_0, \zeta_1, \zeta_2, \ldots$

Put

$$t_{k,n} := \sum_{j=1}^{k} \sigma_{j,n}^2,$$

so that $t_{0,n} = 0$, $t_{n,n} = 1$, and consider a random polygon with vertices at the points (t_k, ζ_k), where we suppress the second subscript n for brevity's sake: $t_k = t_{k,n}$, $\zeta_k = \zeta_{k,n}$.

We obtain a random process on [0, 1] with continuous trajectories, which will be denoted by $s_n = s_n(t)$ (see Fig. 20.1). The functional limit theorem (or invariance principle; the motivation behind this second name will be commented on below) states that for any functional f given on the space $C(0, 1)$ and continuous in the uniform metric, the distribution of $f(s_n)$ converges weakly to that of $f(w)$ as $n \to \infty$:

$$f(s_n) \Longrightarrow f(w), \tag{20.1.2}$$

where $w = w(t)$ is the standard Wiener process. The conventional central limit theorem is a special case of this statement (one should take $f(x)$ to be $x(1)$).

The above assertion is equivalent to each of the following two statements:

1. For any bounded continuous functional f,

$$\mathbf{E} f(s_n) \to \mathbf{E} f(w), \quad n \to \infty. \tag{20.1.3}$$

2. For any set G from the σ-algebra $\mathfrak{B}_{C(0,1)}$ of Borel sets in the space $C(0, 1)$ ($\mathfrak{B}_{C(0,1)}$ is generated by open balls in the metric space $C(0, 1)$ endowed with the uniform distance ρ; as we already noted, $\mathfrak{B}_{C(0,1)} = \mathfrak{B}_C^{[0,1]}$) such that $\mathbf{P}(w \in \partial G) = 0$, where ∂G is the boundary of the set G, one has

$$\mathbf{P}(s_n \in G) \to \mathbf{P}(w \in G), \quad n \to \infty. \tag{20.1.4}$$

Relations (20.1.3) and (20.1.4) are equivalent definitions of weak convergence of the distributions $\mathbf{P}_n$ of the processes s_n to the distribution $\mathbf{W}$ of the Wiener process w in the space $\langle C(0, 1), \mathfrak{B}_{C(0,1)} \rangle$. More details can be found in Appendix 3 and in [1] and [14].

The main results of the present section are the following theorems.

As before, put $L_3 := \sum_{k=1}^{n} \mu_{k,n}$.

Theorem 20.1.1 *Let* $L_3 \to 0$ *as* $n \to \infty$ (the Lyapunov condition). *Then the convergence relations* (20.1.2)–(20.1.4) *hold true.*

Remark 20.1.1 The condition $L_3 \to 0$ can be relaxed here to the Lindeberg condition. In this version the above convergence theorem is known under the name of the *Donsker–Prokhorov invariance principle*.

Along with Theorem 20.1.1 we will obtain a more precise assertion.

Definition 20.1.1 A set G is said to be *Lipschitz* if $\mathbf{W}(G^{(\varepsilon)}) - \mathbf{W}(G) \le c\varepsilon$ for some $c < \infty$, where $G^{(\varepsilon)}$ is the ε-neighbourhood of G and $\mathbf{W}$ is the measure corresponding to the Wiener process.

In the sequel we will denote by the letter c (with or without subscripts) absolute constants, possibly having different values.

Theorem 20.1.2 *If G is a Lipschitz set, then*

$$\bigl|\mathbf{P}(s_n \in G) - \mathbf{P}(w \in G)\bigr| < cL_3^{1/4}. \tag{20.1.5}$$

In the case when $\xi_{k,n} = \xi_k/\sqrt{n}$, where the ξ_k do not depend on n and are identically distributed with $\mathbf{E}\xi_k = 0$ and $\mathrm{Var}(\xi_k) = 1$, the right-hand side of (20.1.5) becomes $cn^{-1/8}$.

A similar bound can be obtained for functionals. A functional on $C(0, 1)$ is said to be *Lipschitz* if the following two conditions are met:

(1) $|f(x) - f(y)| < c\rho(x, y)$;
(2) the distribution of $f(w)$ has a bounded density.

Corollary 20.1.1 *If f is a Lipschitz functional, then $G_v := \{f(x) < v\}$ is a Lipschitz set (with one and the same constant for all v), so that by Theorem 20.1.2*

$$\sup_v \bigl|\mathbf{P}\bigl(f(w) < v\bigr) - \mathbf{P}\bigl(f(s_n) < v\bigr)\bigr| \le cL_3^{1/4}.$$

The above theorems are consequences of Theorem 20.1.3 to be stated below.

Let

$$\eta_{1,n}, \ldots, \eta_{n,n} \tag{20.1.6}$$

be any other sequence of independent identically distributed random variables in the triangular array scheme with the same two first moments $\mathbf{E}\eta_{k,n} = 0$, $\mathbf{E}\eta_{k,n}^2 = \sigma_{k.n}^4$, and finite third moments. Denote by $F_{k,n}$ and $\Phi_{k,n}$ the distribution functions of $\xi_{k,n}$ and $\eta_{k,n}$, respectively, and put

$$\nu_{k,n} := \mathbf{E}|\eta_{k,n}|^3 < \infty, \qquad N_3 := \sum_{k=1}^n \nu_{k,n},$$

$$\mu_{k,n}^0 := \int |x|^3 \bigl|d\bigl(F_{k,n}(x) - \Phi_{k,n}(x)\bigr)\bigr| \le \mu_{k,n} + \nu_{k,n},$$

$$L_3^0 := \sum_{k=1}^n \mu_{k,n}^0 \le L_3 + N_3.$$

Denote by $s_n'(t)$ the random process constructed in the same way as $s_n(t)$ but using the sequence $\{\eta_{k,n}\}$.

Theorem 20.1.3 *For any $A \in \mathfrak{B}_{C(0,1)}$ and any $\varepsilon > 0$,*

$$\mathbf{P}(s_n \in A) \le \mathbf{P}\bigl(s_n' \in A^{(2\varepsilon)}\bigr) + \frac{cL_3^0}{\varepsilon^3}.$$

In order to prove Theorem 20.1.3, we will first obtain its finite-dimensional analogue. Denote by ζ and η the vectors $\zeta = (\zeta_1, \ldots, \zeta_n)$ and $\eta = (\eta_1, \ldots, \eta_2)$ respectively, where $\zeta_k := \sum_{j=1}^k \zeta_{j,n}$ and $\eta_k := \sum_{j=1}^k \eta_{j,n}$, and by $B^{(\varepsilon)}$ the ε-neighbourhood of a set $B \in \mathbb{R}^n$:

$$B^{(\varepsilon)} := \bigcup_{\substack{x \in B \\ |v| \le \varepsilon}} (x + v),$$

where $x = (x_1, \ldots, x_n)$, $v = (v_1, \ldots, v_n)$, and $|v| = \max_k |v_k|$.

Lemma 20.1.1 *Let B be an arbitrary Borel subset of $\mathbb{R}^n$. Then, for any $\varepsilon > 0$,*

$$\mathbf{P}(\zeta \in B) \le \mathbf{P}\bigl(\eta \in B^{(2\varepsilon)}\bigr) + \frac{cL_3^0}{\varepsilon^3}.$$

Proof[1] Introduce a collection of nested neighbourhoods

$$B^{(\varepsilon)}(k) := \bigcup_{\substack{x \in B \\ |v| \le \varepsilon}} (x_1, \ldots, x_k, x_{k+1} + v_{k+1}, \ldots, x_n + v_n), \quad k = 0, \ldots, n,$$

$$B := B^{(\varepsilon)}(n) \subset B^{(\varepsilon)}(n-1) \subset \cdots \subset B^{(\varepsilon)}(1) \subset B^{(\varepsilon)}(0) = B^{(\varepsilon)}$$

and denote by e_k the vector $(0, \ldots, 0, 1, 0, \ldots, 0)$, where 1 stands in the k-th position. It is obvious that if $x \in B^{(\varepsilon)}(k)$, then

$$x + \mathsf{e}_k v_k \in B^{(\varepsilon)}(k-1) \quad \text{if } |v_k| \le \varepsilon. \tag{20.1.7}$$

Further, together with arrays (20.1.1) and (20.1.6), consider the collection of "transitional" arrays

$$\xi_{1,n}, \ldots, \xi_{k,n}, \eta_{k+1,n}, \ldots, \eta_{n,n}, \quad k = 0, \ldots, n. \tag{20.1.8}$$

Denote by $\zeta(k) = (\zeta_1(k), \ldots, \zeta_n(k))$ the vectors formed by the cumulative sums of random variables from the k-th row (20.1.8), so that

$$\zeta_j(k) = \begin{cases} \zeta_j & \text{for } j \le k, \\ \zeta_k + \eta_{k+1,n} + \cdots + \eta_{j,n} & \text{for } j > k. \end{cases}$$

To continue the proof of Lemma 20.1.1 we need the following.

---

[1] The extension of the approach to proving the central limit theorem used in Sect. 8.5, which is used in this demonstration, was suggested by A.V. Sakhanenko.

Lemma 20.1.2 *For any random variable δ such that $\mathbf{P}(|\delta| \le \varepsilon) = 1$, one has*

$$\mathbf{P}(\zeta \in B) \le \mathbf{P}\big(\eta \in B^{(2\varepsilon)}\big) + \sum_{k=1}^{n} \Delta_k, \tag{20.1.9}$$

where

$$\Delta_k = \mathbf{P}\big(\zeta(k) + \delta \mathsf{e}(k-1) \in B^{(\varepsilon)}(k-1)\big) - \mathbf{P}\big(\zeta(k-1) + \delta \mathsf{e}(k-1) \in B^{(\varepsilon)}(k-1)\big),$$

$$\mathsf{e}(r) = \sum_{j=r+1}^{n} \mathsf{e}_j = (0, \dots, 0, 1, \dots, 1).$$

Proof Indeed, by virtue of (20.1.7),

$$\begin{aligned}\mathbf{P}(\zeta \in B) &= \mathbf{P}\big(\zeta(n) \in B^{(\varepsilon)}(n)\big) \le \mathbf{P}\big(\zeta(n) + \mathsf{e}(n-1)\delta \in B^{(\varepsilon)}(n-1)\big)\\ &\equiv \mathbf{P}\big(\zeta(n-1) + \mathsf{e}(n-1)\delta \in B^{(\varepsilon)}(n-1)\big) + \Delta_n.\end{aligned}$$

Reapplying the same calculations to the right-hand side, we obtain that

$$\begin{aligned}&\mathbf{P}\big(\zeta(n-1) + \mathsf{e}(n-1)\delta \in B^{(\varepsilon)}(n-1)\big)\\ &\quad\le \mathbf{P}\big(\zeta(n-1) + \mathsf{e}(n-1)\delta + \mathsf{e}_{n-1}\delta \in B^{(\varepsilon)}(n-2)\big)\\ &\quad= \mathbf{P}\big(\zeta(n-1) + \mathsf{e}(n-2)\delta \in B^{(\varepsilon)}(n-2)\big)\\ &\quad\equiv \mathbf{P}\big(\zeta(n-2) + \mathsf{e}(n-2)\delta \in B^{(\varepsilon)}(n-2)\big) + \Delta_{n-1}\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ &\mathbf{P}\big(\zeta(1) + \mathsf{e}(1)\delta \in B^{(\varepsilon)}(1)\big) \le \mathbf{P}\big(\zeta(1) + \mathsf{e}(1)\delta + \mathsf{e}_1\delta \in B^{(\varepsilon)}(0)\big)\\ &\quad= \mathbf{P}\big(\zeta(1) + \mathsf{e}(0)\delta \in B^{(\varepsilon)}(0)\big) \equiv \mathbf{P}\big(\zeta(0) + \mathsf{e}(0)\delta \in B^{(\varepsilon)}(0)\big) + \Delta_1.\end{aligned}$$

Since $\zeta(0) = \eta$ and $\mathbf{P}(\eta + \mathsf{e}(0)\delta \in B^{(\varepsilon)}) \le \mathbf{P}(\eta \in B^{(2\varepsilon)})$, inequality (20.1.9) is proved. Lemma 20.1.2 is proved. □

To obtain Lemma 20.1.1, we now have to estimate Δ_k. It will be convenient to consider, along with (20.1.8), the sequences

$$\xi_{1,n}, \dots, \xi_{k-1,n}, y, \eta_{k+1,n}, \dots, \eta_{n,n}$$

and denote by $\zeta(k, y) = (\zeta_1(k, y), \dots, \zeta_n(k, y))$ the respective vectors of cumulative sums, so that

$$\begin{aligned}\zeta(k, \xi_{k,n}) &= \zeta(k) = \zeta(k, 0) + \xi_{k,n}\mathsf{e}_{(k-1)},\\ \zeta(k, \eta_{k,n}) &= \zeta(k-1) = \zeta(k, 0) + \eta_{k,n}\mathsf{e}_{(k-1)}.\end{aligned}$$

Then Δ_k can be written in the form

$$\begin{aligned}\Delta_k = {}& \mathbf{P}\big(\zeta(k, 0) + (\delta + \xi_{k,n})\mathsf{e}(k-1) \in B^{(\varepsilon)}(k-1)\big)\\ &- \mathbf{P}\big(\zeta(k, 0) + (\delta + \eta_{k,n})\mathsf{e}(k-1) \in B^{(\varepsilon)}(k-1)\big).\end{aligned} \tag{20.1.10}$$

Take δ to be a random variable independent of ζ and η. Then it will be convenient to use conditional expectation to estimate the probabilities participating in (20.1.10), because, for instance, in the equality

$$\begin{aligned}&\mathbf{P}\big(\zeta(k,0)+(\delta+\xi_{k,n})\mathsf{e}(k-1)\in B^{(\varepsilon)}(k-1)\big)\\&\quad=\mathbf{E}\mathbf{P}\big((\delta+\xi_{k,n})\mathsf{e}(k-1)\in B^{(\varepsilon)}(k-1)-\zeta(k,0)\bigm|\zeta(k,0)\big)\end{aligned}\tag{20.1.11}$$

the set $C=B^{(\varepsilon)}(k-1)-\zeta(k,0)$ may be assumed fixed (see the properties of conditional expectations; here δ and $\xi_{k,n}$ are independent of $\zeta(k,0)$). Denote by D the set of all ys for which $y\,\mathsf{e}(k-1)\in C$. We have to bound the difference

$$\mathbf{P}(\delta+\xi_{k,n}\in D)-\mathbf{P}(\delta+\eta_{k,n}\in D).\tag{20.1.12}$$

We make use of Lemma 8.5.1. To transform (20.1.12) to a form convenient for applying the lemma, take δ to be a random variable having a thrice continuously differentiable density $g(t)$ and put for brevity $\xi_{k,n}=\xi$ and $\eta_{k,n}=\eta$. Then $\delta+\xi$ will have a density equal to

$$\int dF_\xi(t)g(y-t)=\mathbf{E}g(y-\xi),$$

so that

$$\mathbf{P}(\delta+\xi\in D)=\int_D\mathbf{E}g(y-\xi)\,dy=\mathbf{E}\int_D g(y-\xi)\,dy.$$

Now putting

$$h(x):=\int_D g(y-x)\,dy,$$

we have

$$\mathbf{P}(\delta+\xi\in D)=\mathbf{E}h(\xi),$$

where h is a thrice continuously differentiable function,

$$\big|h'''(x)\big|\le\int\big|g'''(y)\big|\,dy=:h_3.$$

Applying now Lemma 8.5.1 we obtain that

$$\begin{gathered}\big|\mathbf{P}(\delta+\xi\in D)-\mathbf{P}(\delta+\eta\in D)\big|=\big|\mathbf{E}\big(h(\xi)-h(\eta)\big)\big|\le\frac{h_3}{6}\mu^0_{k,n},\\ \mu^0_{k,n}=\int|x^3|\big|d\big(F_{k,n}(x)-\Phi_{k,n}(x)\big)\big|.\end{gathered}$$

Because the right-hand side here does not depend on $\xi(k,0)$ and D in any way, we get, returning to (20.1.10) and (20.1.11), the estimate

$$|\Delta_k|\le\frac{h_3}{6}\mu^0_{n,k}.\tag{20.1.13}$$

Now let $g_1(x)$ be a smooth density concentrated on $[-1,1]$. Then, putting

$$g(x):=g_1\left(\frac{x}{\varepsilon}\right)\frac{1}{\varepsilon},$$

we obtain that

$$h_3 = \int \frac{1}{\varepsilon^4} \left| g_1''' \left(\frac{y}{\varepsilon} \right) \right| dy = \frac{1}{\varepsilon^3} \int \left| g_1'''(y) \right| dy = \frac{c_1}{\varepsilon^3}, \quad c_1 = \text{const.} \qquad (20.1.14)$$

The assertion of Lemma 20.1.1 now follows from (20.1.9), (20.1.13) and (20.1.14). □

Proof of Theorem 20.1.3 This theorem is a consequence of Lemma 20.1.1. Indeed, let $B \in \mathbb{R}^n$ be such that the events $\{s_n \in A\}$ and $\{\zeta \in B\}$ are equivalent (s_n is completely determined by ζ). Then clearly $\{s_n \in A^{(\varepsilon)}\} = \{\zeta \in B^{(\varepsilon)}\}$ and the assertion of Theorem 20.1.3 repeats that of Lemma 20.1.1. Theorem 20.1.3 is proved. □

Proof of Theorem 20.1.1 Let $w(t)$ be the standard Wiener process. Put $\eta_{k,n} := w(t_{k,n}) - w(t_{k-1,n})$. Then the sequence $\eta_{1,n}, \dots, \eta_{n,n}$ satisfies all the required conditions, for

$$\mathbf{E}\eta_{k,n} = 0, \qquad \mathbf{E}\eta_{k,n}^2 = \sigma_{k,n}^2, \qquad \nu_{k,n} = \mathbf{E}|\xi_{k,n}|^3 = c_3\sigma_{k,n}^3 < \infty.$$

Note also that

$$\sigma_{k,n}^3 = \left(\mathbf{E}|\xi_{k,n}|^2\right)^{3/2} \le \mathbf{E}|\xi_{k,n}|^3 = \mu_{k,n},$$

so that

$$N_3 = \sum_{k=1}^{n} \nu_{k,n} = c_3 \sum_{k=1}^{n} \sigma_{k,n}^3 \le c_3 L_3 \to 0.$$

We will need the following

Lemma 20.1.3 $\mathbf{P}(\rho(w, s_n') > \varepsilon) \le cN_3/\varepsilon^3$.

Proof The event $\{\rho(w, s_n') > \varepsilon\}$ is equal to $\bigcup_k A_k$, where

$$A_k := \left\{ \sup_{t \in I_k} \left| w(t) - s'(t) \right| > \varepsilon \right\} \subset \left\{ \sup_{t \in I_k} \left| w(t) \right| > \frac{\varepsilon}{2} \right\}, \qquad I_k := [t_{k-1}, t_k].$$

Therefore, recalling that $t_k - t_{k-1} = \sigma_{k,n}^2$ and $w(t) \stackrel{d}{=} \sigma w(t/\sigma^2)$, we have

$$\mathbf{P}(A_k) \le \mathbf{P}\left(\sup_{t \in [0,1)} \left| w(t) \right| > \frac{\varepsilon}{2\sigma_{k,n}} \right) \le 2\left(1 - \Phi\left(\frac{\varepsilon}{2\sigma_{k,n}} \right) \right).$$

The function $(1 - \Phi(t))$ vanishes as $t \to \infty$ much faster than t^{-3}. Hence

$$2\left(1 - \Phi\left(\frac{\varepsilon}{2\sigma_{k,n}} \right) \right) \le c\frac{\sigma_{k,n}^3}{\varepsilon^3}, \qquad \mathbf{P}\left(\bigcup_k A_k \right) \le \frac{cN_3}{\varepsilon^3}.$$

Lemma 20.1.3 is proved. □

We see from the proof that the bound stated by Lemma 20.1.3 is rather crude.

We return to the proof of Theorem 20.1.1. Because

$$\mathbf{P}\big(s_n' \in G\big) = \mathbf{P}\big(s_n' \in G, \rho\big(w, s_n'\big) \le \varepsilon\big) + \mathbf{P}\big(s_n' \in G, \rho\big(w, s_n'\big) > \varepsilon\big),$$

we have

$$\mathbf{P}\big(s_n' \in G\big) \le \mathbf{P}\big(w \in G^{(\varepsilon)}\big) + \frac{cN_3}{\varepsilon^3} \tag{20.1.15}$$

and, by Theorem 20.1.3,

$$\mathbf{P}(s_n \in G) \le \mathbf{P}\big(w \in G^{(3\varepsilon)}\big) + \frac{c(L_3^0 + N_3)}{\varepsilon^3}.$$

Now we prove the converse inequality. Introduce the set $G^{(-\varepsilon)} := G - (\partial G)^{(\varepsilon)}$. Then $[G^{(-\varepsilon)}]^{(\varepsilon)} =: G^0 \subset G$. Swapping s_n and s_n' in Theorem 20.1.3 and applying the latter to the set $G^{(2\varepsilon)}$, we obtain

$$\mathbf{P}\big(s_n \in G^0\big) \ge \mathbf{P}\big(s_n' \in G^{(-2\varepsilon)}\big) - \frac{c(L_3^0 + N_3)}{\varepsilon^3}. \tag{20.1.16}$$

Swapping w and s_n' in (20.1.15) and applying that relation to $G^{(-3\varepsilon)}$, we find that

$$\mathbf{P}\big(s_n' \in G^{(-2\varepsilon)}\big) \ge \mathbf{P}\big(w \in G^{(-3\varepsilon)}\big) - \frac{cN_3}{\varepsilon^3}.$$

This and (20.1.16) imply that

$$\mathbf{P}(s_n \in G) \ge \mathbf{P}\big(s_n \in G^0\big) \ge \mathbf{P}\big(w \in G^{(-3\varepsilon)}\big) - \frac{c(L_3^0 + N_3)}{\varepsilon^3}.$$

Setting

$$\mathbf{P}\big(w \in G^{(\varepsilon)}\big) - \mathbf{P}(w \in G) = \mathbf{W}\big(G^{(\varepsilon)}\big) - \mathbf{W}(G) =: \mathbf{W}_G(\varepsilon)$$

and taking into account that $N_3 \le cL_3$ and $L_3^0 \le L_3 + N_3$, we will obtain that

$$-\mathbf{W}_G(-3\varepsilon) + \frac{cL_3}{\varepsilon^3} \le \mathbf{P}(s_n \in G) - \mathbf{W}(G) \le \mathbf{W}_G(3\varepsilon) + \frac{cL_3}{\varepsilon^3}. \tag{20.1.17}$$

If $\mathbf{W}(\partial G) = 0$ then clearly

$$\mathbf{W}\big(G^{(3\varepsilon)}\big) - \mathbf{W}\big(G^{(-3\varepsilon)}\big) \to 0$$

as $\varepsilon \to 0$, and $\mathbf{W}_G(\pm 3\varepsilon) \to 0$. From this and (20.1.17) it is easy to derive that

$$\mathbf{P}(s_n \in G) \to \mathbf{P}(w \in G), \quad n \to \infty.$$

Convergence $f(s_n) \Longrightarrow f(w)$ for continuous functionals follows from (20.1.4), since if v is a point of continuity of the distribution of $f(w)$ then the set $G_v = \{x \in C(0,1) : f(x) < v\}$ has the property

$$\mathbf{W}(\partial G_v) = \mathbf{P}\big(f(w) = v\big) = 0$$

and therefore

$$\mathbf{P}\big(f(s_n) < v\big) \to \mathbf{P}\big(f(w) \in v\big).$$

Theorem 20.1.1 is proved. □

Proof of Theorem 20.1.2 If G is a Lipschitz set, then

$$\left|\Delta \mathbf{W}_G(\pm 3\varepsilon)\right| < c\varepsilon,$$

and by (20.1.17)

$$\left|\mathbf{P}(s_n \in G) - \mathbf{W}(G)\right| < c\left(\varepsilon + \frac{L_3}{\varepsilon^3}\right).$$

Putting $\varepsilon := L_3^{1/4}$ we obtain the required assertion. Theorem 20.1.2 is proved. □

The reason for the name "*invariance principle*" used to refer to the main assertions of this section is best illustrated by Theorem 20.1.3. By virtue of the theorem, one can approximate the value of $\mathbf{P}(s_n \in A)$ by $\mathbf{P}(s_n' \in A)$ for any other sequence (20.1.6) having the same first two moments as (20.1.1). In that sense, the asymptotics of $\mathbf{P}(s_n \in A)$ are invariant with respect to particular distributions of the underlying sequences with fixed first two moments. For example, the calculation of $\mathbf{P}(s_n \in G)$ or $\mathbf{P}(w(t) \in G)$ can be replaced with that of $\mathbf{P}(s_n' \in G)$ for a Bernoulli sequence, which is convenient for various numerical methods. On the other hand, the probabilities $\mathbf{P}(w \in G)$ for a whole class of regions G were found in explicit form (see e.g. [32]). We know, for example, that $\mathbf{P}(\sup_{t\in[0,1]} w(t) > y) = 2(1 - \Phi(y))$. (This implies, in particular, that $G = \{x \in C(0,1) \sup_{t\in[0,1]} x(t) > y\}$ is a Lipschitz set.) Hence for the distribution of the maximum $\overline{S}_n = \max_{k\le n} S_k$ of the sums $S_k = \sum_{j=1}^k \xi_j$, when $\mathbf{E}\xi_k = 0$ and $\operatorname{Var}\xi_k = \sigma^2$, we have

$$\mathbf{P}(\overline{S}_n > x\sigma\sqrt{n}\,) \to 2\bigl(1 - \Phi(x)\bigr), \quad n \to \infty,$$

and one can use this relation for the approximate calculation of the distribution of $\overline{S}_n$ which is, as we saw in Chap. 12, of substantial interest in applications.

In the same way we can approximate the joint distribution of S_n, $\overline{S}_n$, and $\underline{S}_n := \min_{k\le n} S_k$ (i.e. the probabilities of the form $\mathbf{P}(\overline{S}_n < x\sqrt{n},\ \underline{S}_n > y\sqrt{n},\ S_n \in B)$) using the respective formulas for the Wiener process given in Skorokhod (1991).

Remark 20.1.2 In conclusion of this section note that all the above assertions will remain true if, instead of $s_n(t)$, we consider in them the step function $s_n^*(t) = \zeta_{k,n}$ for $t \in [t_k, t_{k+1})$. One can verify this by repeating anew all the arguments for s^*. Another way to obtain, say, Theorems 20.1.1 and 20.1.2 for s_n^* is to make use of the already obtained results and bound the distance $\rho(s_n, s_n^*)$. Because

$$\bigl\{\rho\bigl(s_n, s_n^*\bigr) > \varepsilon\bigr\} \subset \bigcup_{k=1}^n \bigl\{|\xi_{k,n}| > \varepsilon\bigr\},$$

one has

$$\mathbf{P}\bigl(\rho\bigl(s_n, s_n^*\bigr) > \varepsilon\bigr) \le \sum_{k=1}^n \mathbf{P}\bigl(|\xi_{k,n}| > \varepsilon\bigr) \le \sum_{k=1}^n \frac{\mu_{k,n}}{\varepsilon^3} = \frac{L_3}{\varepsilon^3}.$$

Recall that a similar bound was obtained for $\rho(s_n', w)$, and this allowed us to replace, where it was needed, the process s_n' with w. Therefore, using the same argument, one can replace s_n with s_n^*. In that case, we can consider convergence of the distributions of functionals $f(s_n^*)$ defined on $D(0,1)$ (and continuous in the uniform metric ρ). Sometimes the use of s_n^* is more convenient than that of s_n. This is the case, for example, when one has to find the limiting distribution of

$$\sum_{k=1}^{n} g(\zeta_{k,n}) = n \int g\big(s_n^*(t)\big)\, dt$$

($\xi_{k,n}$ are identically distributed). It follows from the above representation that

$$\frac{1}{n}\sum_{k=1}^{n} g(\zeta_{k,n}) \Longrightarrow \int g\big(w(t)\big)\, dt, \quad n \to \infty.$$

20.2 The Law of the Iterated Logarithm

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables,

$$\mathbf{E}\xi_k = 0, \qquad \mathbf{E}\xi_k^2 = \sigma_k^2, \qquad \mathbf{E}|\xi_k|^3 = \mu_k,$$
$$S_n = \sum_{k=1}^{n} \xi_k, \qquad B_n^2 = \sum_{k=1}^{n} \sigma_k^2, \qquad M_n = \sum_{k=1}^{n} \mu_k.$$

In this notation, the Lyapunov ratio is equal to

$$L_3 = L_{3,n} = \frac{M_n}{B_n^3}.$$

In the present section, we will show that the law of the iterated logarithm for the Wiener process and Theorem 20.1.2 imply the following.

Theorem 20.2.1 (The law of the iterated logarithm for sums of random variables) *If $B_n \to \infty$ as $n \to \infty$ and $L_{3,n} < c/\ln B_n$ for some $c < \infty$, then*

$$\mathbf{P}\left(\overline{\lim_{n\to\infty}} \frac{S_n}{B_n\sqrt{2\ln\ln B_n}} = 1\right) = 1, \tag{20.2.1}$$

$$\mathbf{P}\left(\underline{\lim_{n\to\infty}} \frac{S_n}{B_n\sqrt{2\ln\ln B_n}} = -1\right) = 1. \tag{20.2.2}$$

Thus all the sequences which lie above

$$(1+\varepsilon)B_n\sqrt{2\ln\ln B_n}$$

will be upper for the sequence of sums S_n, while all the sequences below

$$(1-\varepsilon)B_n\sqrt{2\ln\ln B_n}$$

will be lower.

The conditions of the theorem will clearly be satisfied for identically distributed ξ_k, for in that case $B_n^2 = \sigma_1^2 n$, $L_{3,n} = \mu_1/(\sigma_1^3\sqrt{n})$.

Proof We turn to the proof of the law of the iterated logarithm in Theorem 19.3.2 and apply it to the sequence S_n. We will not need to introduce any essential changes. One just has to consider S_{n_k} instead of $w(a^k)$, where $n_k = \min\{n : B_n^2 \ge a^k\}$, and replace a^k with $B_{n_k}^2$ where it is needed. By the Lyapunov condition, $\max_{k\le n}\sigma_k^2 = o(B_n^2)$, so that $B_{n_k-1}^2 \sim B_{n_k}^2 \sim a^k$ as $k \to \infty$.

The key point in the proof of Theorem 19.3.2 is the proof of convergence (for any $\varepsilon > 0$) of the series

$$\sum_k \mathbf{P}\Bigl(\sup_{u\le a^k} w(u) > (1+\varepsilon)x_{k-1}\Bigr) \tag{20.2.3}$$

and divergence of the series

$$\sum_k \mathbf{P}\biggl(w\bigl(a^k\bigr) - w\bigl(a^{k-1}\bigr) > \biggl(1-\frac{\varepsilon}{2}\biggr)x_k\biggr), \tag{20.2.4}$$

where

$$x_k = \sqrt{2a^k \ln\ln a^k}, \qquad w\bigl(a^k\bigr) - w\bigl(a^{k-1}\bigr) \overset{p}{=} w\bigl(a^k\bigl(1-a^{-1}\bigr)\bigr).$$

In our case, if one follows the same argument, one has to prove the convergence of the series

$$\sum_k \mathbf{P}\bigl(\overline{S}_{n_k} > (1+\varepsilon)\, y_{k-1}\bigr) \tag{20.2.5}$$

and divergence of the series

$$\sum_k \mathbf{P}\biggl(S_{n_k} - S_{n_{k-1}} > \biggl(1-\frac{\varepsilon}{2}\biggr)y_k\biggr), \tag{20.2.6}$$

where $y_k = \sqrt{2B_{n_k}^2 \ln\ln B_{n_k}^2} \sim x_k$. But the asymptotic behaviour of the probabilities of the events in (20.2.3), (20.2.5) and (20.2.4), (20.2.6) under the conditions $L_{3,n} < c/\ln B_n$ will essentially be the same. To establish this, we will make use of the inequality

$$\biggl|\mathbf{P}\biggl(\frac{s_n}{B_n} \in G\biggr) - \mathbf{P}\bigl(w \in G^{(\pm\delta)}\bigr)\biggr| < \frac{cL_{3,n}}{\delta^3}, \tag{20.2.7}$$

which follows from the proof of Theorem 20.1.3. By this inequality,

$$\begin{aligned}\mathbf{P}\biggl(\frac{\overline{S}_n}{B_n} > (1+3\varepsilon)x\biggr) &\le \mathbf{P}\Bigl(\sup_{u\le 1} w(u) > (1+2\varepsilon)\,x\Bigr) + \frac{cL_{3,n}}{(\varepsilon x)^3}\\ &= \mathbf{P}\Bigl(\sup_{u\le B_n^2} w(u) > (1+2\varepsilon)xB_n\Bigr) + \frac{cL_{3,n}}{(\varepsilon x)^3}.\end{aligned}$$

Therefore (see (20.2.5)), putting $n := n_k$ and $x := y_{k-1}/B_n$, we obtain

$$\mathbf{P}\big(\overline{S}_{n_k} > (1+3\varepsilon)y_{k-1}\big) \le \mathbf{P}\Big(\sup_{u\le B^2_{n_k}} w(u) > (1+2\varepsilon)\,y_{k-1}\Big) + \frac{cL_{3,n_k}}{\varepsilon^3(\ln\ln B^2_{n_k})^{3/2}}.$$

Here

$$y_{k-1}\sim x_{k-1}, \qquad B^2_{n_k}\ge a^k, \qquad \ln\ln B^2_{n_k}\sim \ln\ln a^k\sim \ln k,$$
$$L_{3,n_k}\le \frac{c}{\ln B_{n_k}}\sim \frac{c}{\ln a^k}\sim \frac{c_1}{k}.$$

Consequently, for all sufficiently large k (recall that the letter c denotes different constants),

$$\mathbf{P}\big(\overline{S}_{n_k} > (1+3\varepsilon)y_{k-1}\big) \le \mathbf{P}\Big(\sup_{u\le a^k} w(u) > (1+\varepsilon)x_{k-1}\Big) + \frac{c}{\varepsilon^3 k(\ln k)^{3/2}}.$$

Since

$$\sum_{k=1}^{\infty}\frac{1}{k(\ln k)^{3/2}}<\infty, \tag{20.2.8}$$

the above inequality means that the convergence of series (20.2.3) implies that of series (20.2.5). The first part of the theorem is proved.

The second part is proved in a similar way. Consider series (20.2.6). By (20.2.7),

$$\begin{aligned}
&\mathbf{P}\big(S_{n_k}-S_{n_{k-1}} > (1-3\varepsilon)y_k\big)\\
&\quad=\mathbf{P}\Big(s_{n_k}(1)-s_{n_k}(r_k) > (1-3\varepsilon)\frac{y_k}{B_{n_k}}\Big)\\
&\quad\ge \mathbf{P}\Big(w(1)-w(r_k) > (1-2\varepsilon)\frac{y_k}{B_{n_k}}\Big) - \frac{cL_{3,n_k}}{\varepsilon^3}\big(\ln\ln B^2_{n_k}\big)^{-3/2},
\end{aligned}\tag{20.2.9}$$

where $r_k = B^2_{n_{k-1}}B^{-2}_{n_k}\to a^{-1}$ due to the fact that

$$B^2_{n_k}=a^k+\theta_k\sigma^2_{n_k}, \quad 0\le\theta_k\le 1, \quad \sigma^2_{n_k}=o\big(B^2_{n_k}\big).$$

The first term on the right-hand side of (20.2.9) is equal to

$$\begin{aligned}
\mathbf{P}\Big(w(1-r_k) > (1-2\varepsilon)\frac{y_k}{B_{n_k}}\Big) &\ge \mathbf{P}\big(w\big(a^k(1-r_k)\big) > (1-\varepsilon)x_k\big)\\
&= \mathbf{P}\Big(w\big(a^k\big(1-a^{-1}\big)\big) > (1-\varepsilon)x_k\sqrt{\frac{1-a^{-1}}{1-r_k}}\Big)\\
&\ge \mathbf{P}\Big(w\big(a^k\big)-w\big(a^{k-1}\big) > \Big(1-\frac{\varepsilon}{2}\Big)x_k\Big).
\end{aligned}$$

As before, the series consisting of the second terms on the right-hand side of (20.2.9) converges by virtue of (20.2.8). Therefore the established inequalities mean that the divergence of series (20.2.4) implies that of series (20.2.6). The theorem is proved. □

Now we will present an example that we need to complete the proof in Remark 4.4.1.

Example 20.2.1 Let ζ_k be independent and identically distributed, $\mathbf{E}\zeta_k = 0$, $\mathbf{E}\zeta_k^2 = 1$, $\mathbf{E}|\zeta_k|^3 = \mu < \infty$ and $\xi_k = \sqrt{2k}\,\zeta_k$. Here we have $B_n^2 = n^2(1 + 1/n)$. In Remark 4.4.1 we used the assertion that (in a somewhat different notation)

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty}\{S_n < -n\}\right) = 1$$

or, which is the same (as the sign of S_n is inessential),

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty}\{S_n > n\}\right) = \mathbf{P}\left(\bigcup_{n=1}^{\infty}\left\{S_n > B_n\left(1 + O\left(\frac{1}{n}\right)\right)\right\}\right) = 1. \qquad (20.2.10)$$

To verify it, we will show that any sequence of the form $B_n' = B_n(1+O(1/n))$ is lower for $\{S_k\}$. In our case,

$$M_n = \sum_{k=1}^{n}(2k)^{3/2}\mu \sim cn^{5/2}, \qquad L_{3.n} = \frac{M_n}{B_n^3} \sim cn^{-1/2} \ll \frac{1}{\ln B_n}.$$

This means that the conditions of Theorem 20.2.1 are met, and hence any sequence which lies lower than $(1-\varepsilon)n\sqrt{2\ln\ln n}$ (in particular, the sequence $B_n' = n$) is lower for $\{S_k\}$. This proves (20.2.10). □

Let us return to Theorem 20.2.1. As we saw in Sect. 19.3, the proof of the law of the iterated logarithm is based on the asymptotics (the rate of decrease) of the function $1 - \Phi(x)$ as $x \to \infty$. Therefore, the conditions for the law of the iterated logarithm for the sums S_n are related to the width of the range of x values for which the probabilities

$$\mathbf{P}_{n\pm}(x) := \mathbf{P}\left(\pm\frac{S_n}{B_n} > x\right)$$

are approximated by the normal law (i.e. by the function $1 - \Phi(x)$). Here we encounter the problem of large deviations (see Chap. 9). If

$$\frac{\mathbf{P}_{n\pm}(x)}{1 - \Phi(x)} \to 1 \qquad (20.2.11)$$

as $n \to \infty$ for all

$$x \le \sqrt{2\ln\ln B_n}(1 - \varepsilon) \qquad (20.2.12)$$

and some $\varepsilon > 0$ then the proof of the law of the iterated logarithm for the Wiener process given in Sect. 19.3 can easily be extended to the sums S_n/B_n (to estimate $\mathbf{P}(\overline{S}_n/B_n > x)$ one has to use the Kolmogorov inequality; see Corollary 11.2.1).

One way to establish (20.2.11) and (20.2.12) is to use estimates for the rate of convergence in the central limit theorem. This approach was employed in the proof of Theorem 20.2.1, where we used Theorem 20.1.3. However, to ensure that

(20.2.11) and (20.2.12) hold one can use weaker assertions than Theorem 20.1.3. To some extent, this fact is illustrated by the following assertion (see [32]):

Theorem 20.2.2 *If $B_n \to \infty$ and $B_{n+1}/B_n \to 1$ as $n \to \infty$, and*

$$\sup_x \left| \mathbf{P}\left(\frac{S_n}{B_n} < x \right) - \Phi(x) \right| \le c(\ln B_n)^{-1-\delta}$$

for some $\delta > 0$ and $c < \infty$, then the law of the iterated logarithm holds.

If $\xi_k \overset{d}{=} \xi$ are identically distributed then Theorem 20.2.1 implies that the law of the iterated logarithm is valid whenever $\mathbf{E}|\xi|^3$ exists. In fact, however, for identically distributed ξ_k, the law of the iterated logarithm always holds in the case of a finite second moment, without any additional conditions.

Theorem 20.2.3 (Hartman–Wintner, [32]) *If the ξ_k are identically distributed, $\mathbf{E}\xi_k = 0$, and $\mathbf{E}\xi_k^2 = 1$, then (20.2.1) and (20.2.2) hold with B_n^2 replaced with n. Every point from the segment $[-1, 1]$ is a limiting one for the sequence*

$$\frac{S_n}{\sqrt{2n \ln\ln n}}, \quad n \ge 1.$$

The last assertion of the theorem means that, for each $t \in [-1, 1]$ and any $\varepsilon > 0$, the interval $(t - \varepsilon, t + \varepsilon)$ contains, with probability 1, infinitely many elements of the sequence

$$\frac{S_n}{\sqrt{2n \ln\ln n}}.$$

20.3 Convergence to the Poisson Process

20.3.1 Convergence of the Processes of Cumulative Sums

The theorems of Sects. 20.1 and 20.2 show that the Wiener process describes rather well the evolution of the cumulative sums when summing "conventional" random variables $\xi_{k,n}$ satisfying the Lyapunov condition. It turns out that the Poisson process describes in a similar way the evolution of the cumulative sums when the random variables $\xi_{k,n}$ correspond to the occurrence of rare events.

As in Sect. 5.4, first we will not consider the triangular array scheme, but obtain precise inequalities describing the proximity of the processes under study. Consider independent random variables $\xi_1, \dots, \xi_n$ with Bernoulli distributions:

$$\mathbf{P}(\xi_k = 1) = p_k, \qquad \mathbf{P}(\xi_k = 0) = 1 - p_k, \qquad \sum_{k=1}^n p_k = \mu.$$

We will assume that $\overline{p} := \max_{k \le n} p_k$ is small and the number μ is "comparable with 1". Put

$$q_0 := 0, \qquad q_k := \frac{p_k}{\mu}, \qquad Q_k := \sum_{j=0}^{k} q_j, \quad k \ge 0,$$

and form a random function $s_n(t)$ on [0, 1] in the following way. Put $s_n(0) := 0$,

$$s_n(t) := S_k = \sum_{j=1}^{k} \xi_j \quad \text{for } t \in (Q_{k-1}, Q_k],\ k = 1, \dots, n.$$

Here it is more convenient to use a step function rather than a continuous trajectory $s_n(t)$ (cf. Remark 20.1.2). The assertions to be obtained in this section are similar to the invariance principle and state that the process $s_n(t)$ converges in a certain sense to the Poisson process $\xi(t)$ with intensity μ on [0, 1]. This convergence could of course be treated as weak convergence of distributions in the metric space $D(0, 1)$. But in the framework of the present book, it is apparently inexpedient for at least two reasons:

1. To do that, we would have to introduce a metric in $D(0, 1)$ and study its properties, which is somewhat complicated by itself.

2. The trajectories of the processes $s_n(t)$ and $\xi(t)$ are of a simple form, and characterising their closeness can be done in a simpler and more precise way without using more general concepts. Indeed, as we saw, the trajectory of $\xi(t)$ on [0, 1] is completely determined by the collection of random variables $(\pi(1); T_1, \dots, T_{\pi(1)})$, where T_k is the epoch of the k-th jump of the process, $T_{k+1} - T_k \Subset \mathbf{\Gamma}_\mu$. A similar characterisation is valid for the trajectories of $s_n(t)$: they are determined by the vector $(s_n(1), \theta_1, \dots, \theta_{s_n(1)})$, where $\theta_k = Q_{\gamma_k}$, $\gamma_1, \gamma_2, \dots$ are the values j for which $\xi_j = 1$. We will say that the distributions of $s_n(t)$ and $\pi(t)$ are close to each other if the distributions of the above vectors are close. This convention will correspond to convergence of the processes in a rather strong and natural sense.

It is not hard to see from what we said before about the Poisson processes (see Sect. 19.4) that the introduced convergence of the distributions of the jump points of the process $s_n(t)$ is equivalent to convergence of the finite-dimensional distributions of $s_n(t)$ to those of $\pi(t)$ (we know that the trajectories of $s_n(t)$ are step functions).

Theorem 20.3.1 *The processes $s_n(t)$ and $\pi(t)$ can be constructed on a common probability space so that*

$$\mathbf{P}\big(s_n(1) = \pi(1);\ \theta_k - q_{\gamma_k} \le T_k < \theta_k,\ k = 1, \dots, \pi(1)\big) \ge 1 - \sum_{j=1}^{n} p_j^2. \tag{20.3.1}$$

Since $\sum_{j=1}^{n} p_j^2 \le \mu \overline{p}$, the smallness of $\overline{p}$ means that, with probability close to 1, the values of $s_n(1)$ and $\pi(1)$ coincide (cf. Theorem 5.4.2) and the positions of the respective points of jumps of the processes $s_n(t)$ and $\pi(t)$ do not differ much.

Put $\overline{q} = \overline{p}/\mu$ and, for a fixed $k \ge 1$, denote by $B^{(\varepsilon)}$ the ε-neighbourhood of the orthant set $B := \{(x_1, \ldots, x_k) : x_j < v_j,\ j \le k\}$ for some $v_j > 0$. Theorem 20.3.1 implies the following.

Corollary 20.3.1 *For any $k = 1, \ldots, n$,*

$$\mathbf{P}\big(s_n(1) = k, (\theta_1, \ldots, \theta_k) \in B\big) \le \mathbf{P}\big(\pi(1) = k, (T_1, \ldots, T_k) \in B\big) + \sum_{j=1}^{n} p_j^2;$$

$$\mathbf{P}\big(\pi(1) = k, (T_1, \ldots, T_k) \in B\big) \le \mathbf{P}\big(s_n(1) = k, (\theta_1, \ldots, \theta_k) \in B^{(\overline{q})}\big) + \sum_{j=1}^{n} p_j^2.$$

Proof Let A_n denote the event appearing on the left-hand side of (20.3.1),

$$D_n := \big\{s_n(1) = k,\ (\theta_1, \ldots, \theta_k) \in B\big\},$$
$$C_n := \big\{\pi(1) = k,\ (T_1, \ldots, T_k) \in B\big\}.$$

Then, by virtue of (20.3.1),

$$\begin{aligned}\mathbf{P}(D_n) &\le \mathbf{P}(D_n A_n) + \sum_{j=1}^{n} p_j^2 \\ &\le \mathbf{P}\big(D_n, \pi(1) = k,\ (T_1, \ldots, T_k) \in B\big) + \sum_{j=1}^{n} p_j^2 \\ &\le \mathbf{P}\big(\pi(1) = k,\ (T_1, \ldots, T_k) \in B\big) + \sum_{j=1}^{n} p_j^2.\end{aligned}$$

The converse inequality is established similarly. The corollary is proved. □

Proof of Theorem 20.3.1 Let $\eta_k := \pi(Q_k) - \pi(Q_{k-1})$, $k = 1, \ldots, n$. The theorem will be proved if we construct $\{\xi_k\}$ and $\{\eta_k\}$ on a common probability space so that

$$\mathbf{P}\left(\bigcup_{k=1}^{n} \{\xi_k \ne \eta_k\}\right) \le \sum_{j=1}^{n} p_j^2. \tag{20.3.2}$$

A construction leading to (20.3.2) has essentially already been used in Theorem 5.4.2. The required construction will be obtained if we consider independent random variables $\omega_1, \ldots, \omega_n$; $\omega_k \in \mathbf{U}_{0,1}$, and put

$$\xi_k := \begin{cases} 0 & \text{if } \omega_k < 1 - p_k, \\ 1 & \text{if } \omega_k \ge 1 - p_k, \end{cases} \qquad \eta_k := \begin{cases} 0 & \text{if } \omega_k < e^{-p_k} =: \pi_{0,k}, \\ j \ge 1 & \text{if } \omega_k \in [\pi_{j-1,k}, \pi_{j,k}), \end{cases}$$

where $\pi_{j,k} = \mathbf{\Pi}_{p_k}([0, j))$, $j = 0, 1, \ldots$. Then $\eta_k \in \mathbf{\Pi}_{p_k}$, $\sum_{k=1}^{n} \eta_k \in \mathbf{\Pi}_{\mu}$,

$$\{\xi_k \ne \eta_k\} = \big\{\omega_k \in [1 - p_k, e^{-p_k}) \cup \big[e^{-p_k} + p_k e^{-p_k}, 1\big]\big\}.$$

Therefore,

$$\mathbf{P}(\xi_k \neq \eta_k) \leq p_k^2$$

and we get (20.3.2). The theorem is proved. $\square$

If we now consider the triangular array scheme $\xi_{1,n}, \dots, \xi_{n,n}$, for which

$$\mathbf{P}(\xi_{k,n} = 1) = p_{k,n}, \qquad \mathbf{P}(\xi_{k,n} = 0) = 1 - p_{k,n},$$
$$\sum_{k=1}^{n} p_{k,n} =: \mu_n \to \mu, \qquad \overline{p}_n = \max_{k \leq n} p_{k,n} \to 0$$

as $n \to \infty$, then Theorem 20.3.1 easily implies *convergence of the finite-dimensional distributions of the processes $s_n(t)$ to $\pi(t)$*, where $s_n(t)$ is constructed as before and $\pi(t)$ is the Poisson process with parameter μ. Consider, for example, the two-dimensional distributions $\mathbf{P}(s_n(t) \geq j,\ s_n(1) = k)$ for $t \in (0,1)$, $j \leq k$. In the notation of Theorem 20.3.1 (to be precise, we have to add the subscript n where appropriate; e.g., the Poisson processes with parameters μ_n and μ will be denoted by $\pi_n(t)$ and $\pi(t)$, respectively), we obtain

$$\mathbf{P}\big(s_n(t) \geq j,\ s_n(1) = k\big) = \mathbf{P}\big(s_n(1) = k,\ \theta_j < t\big).$$

By Corollary 20.3.1 the right-hand side does not exceed

$$\mathbf{P}\big(\pi_n(1) = k,\ T_j < t\big) + \sum_{l=1}^{n} p_{l,n}^2,$$

where, as is easy to see,

$$\begin{aligned}
\mathbf{P}\big(\pi_n(1) = k,\ T_j < t\big) &= \mathbf{P}\big(\pi_n(1) = k,\ \pi_n(t) \geq j\big) \\
&= \sum_{l=j}^{k} \mathbf{P}\big(\pi_n(t) = l\big)\mathbf{P}\big(\pi_n(1-t) = k - l\big) \\
&\to \mathbf{P}\big(\pi(1) = k,\ \pi(t) \geq j\big)
\end{aligned}$$

as $n \to \infty$, so that

$$\mathbf{P}\big(s_n(t) \geq j,\ s_n(1) = k\big) \leq \mathbf{P}\big(\pi(t) \geq j,\ \pi(1) = k\big) + o(1).$$

The converse inequality is established in a similar way (by using the convergence $\overline{q}_n \to 0$ as $n \to \infty$). The required convergence of the finite-dimensional distributions is proved. $\square$

20.3.2 Convergence of Sums of Thinning Renewal Processes

The Poisson process can appear as a limiting one in a somewhat different set-up—as a limit for the sum of a large number of homogeneous "slow" renewal processes.

We formulate the setting of the problem more precisely. Let $\eta_i(t)$, $i = 1, 2, \dots, n$, be mutually independent *arbitrary homogeneous renewal processes* in the "triangular array scheme" (i.e. they depend on n) generated by sequences $\{\tau_k^{(i)}\}_{k=1}^{\infty}$ for which (see Chap. 10; $\tau_k^{(i)} \overset{d}{=} \tau^{(i)}$ for $k \ge 2$)

$$\mathbf{E}\eta_i(t) = \frac{t}{a_i}, \quad a_i := a_{i,n} = \mathbf{E}\tau^{(i)} \to \infty, \quad \sum_{i=1}^{n} \frac{1}{a_i} \to \mu$$

for a fixed μ, and

$$F_i(t) := \mathbf{P}\big(\tau^{(i)} < t\big) \le r_{t,n} \to 0$$

and for any fixed t as $n \to \infty$, where $r_{t,n}$ does not depend on i.

Theorem 20.3.2 *Under the above conditions, the finite-dimensional distributions of the process*

$$\zeta_n(t) := \sum_{i=1}^{n} \eta_i(t)$$

converge as $n \to \infty$ to those of the Poisson process $\pi(t)$ with the parameter μ: for any $l \ge 1$, $0 \le k_1 \le k_2 \le \dots \le k_l$,

$$\mathbf{P}\big(\zeta_n(t_1) = k_1, \dots, \zeta_n(t_l) = k_l\big) \to \mathbf{P}\big(\pi(t_1) = k_1, \dots, \pi(t_l) = k_l\big).$$

(On convergence to the Poisson process, see the remark preceding Theorem 20.3.1.)

Proof First we will prove convergence of the distributions of the increments

$$\zeta_n(t+u) - \zeta_n(u)$$

to the Poisson distribution with parameter μt. Put $\Delta_i := \eta_i(t+u) - \eta_i(u)$, $p_i := t/a_i$. We have ($\chi_i(u)$ is the excess for the process η_i; see Sects. 10.2, 10.4)

$$\begin{aligned}
\mathbf{E}\Delta_i &= p_i, \\
\mathbf{P}(\Delta_i \ge l) &\le \mathbf{P}\big(\chi_i(u) < t\big)\big[\mathbf{P}\big(\xi_2^{(1)}\big) < t\big]^{l-1} \\
&\le \frac{1}{a_i} \int_0^t \mathbf{P}\big(\xi_2^{(1)} > z\big)\, dz \cdot F_i(t)^{l-1} \le \frac{t}{a_i} (r_{t,n})^{l-1} = p_i r_{t,n}^{l-1}.
\end{aligned}$$

This implies that

$$\begin{gathered}
\mathbf{E}\Delta_i = p_i = \sum_{1}^{\infty} l\mathbf{P}(\Delta_i = l) = \mathbf{P}(\delta_i = 1) + o(p_i), \\
\mathbf{P}(\Delta_i = 1) = p_i + o(p_i), \qquad \mathbf{P}(\Delta_i = 0) = 1 - p_i + o(p_i).
\end{gathered} \tag{20.3.3}$$

Therefore the conditions of Corollary 5.4.2 are met, which implies that

$$\zeta_n(t+u) - \zeta_n(u) = \sum_{i=1}^{n} \Delta_i \Leftrightarrow \mathbf{\Pi}_{\mu t}. \tag{20.3.4}$$

It remains to prove the asymptotic independence of the increments. For simplicity's sake, consider only two increments, on the intervals $(u, 0)$ and $(u, u+t)$, and assume that $\zeta_n(u) = k$. Moreover, suppose that the following event A occurred: the renewals occurred in the processes with numbers $i_1, \dots, i_k$. It suffices to verify that, given this condition, (20.3.4) will still remain true. Let B be the event that there again were renewals on the interval $(u, u+t)$ in the processes with the numbers $i_1, \dots, i_k$. Evidently,

$$\mathbf{P}(B \mid A) \le \sum_{l=1}^{k} \mathbf{P}\big(\tau^{(i_l)} < t + u\big) \le k r_{t+u,n} \to 0.$$

Thus the contribution of the processes η_{i_l}, $l = 1, \dots, k$, to the sum (20.3.4) given condition A is negligibly small. Consider the remaining $n-k$ processes. For them,

$$\begin{aligned}
\mathbf{P}(\Delta_i \ge 1 \mid A) &= \frac{\mathbf{P}(\chi_i(0) \in (u, u+t))}{\mathbf{P}(\chi_i(0) > u)} \\
&= \frac{1}{a_i} \int_u^{u+t} \big(1 - F_i(z)\big)\, da \left[1 - \frac{1}{a_i} \int_0^u \big(1 - F_i(z)\big)\, dz\right]^{-1} \\
&= p_i + o(p_i). \qquad (20.3.5)
\end{aligned}$$

Since relation (20.3.3) remains true for conditional distributions of Δ_i (given A and for $i \ne i_l$, $l = 1, \dots, k$), we obtain, similarly to the above argument (using now instead of the equality $\sum_{i=1}^{\infty} l\mathbf{P}(\Delta_i = l) = p_i$ the relation $\sum_{i=1}^{\infty} l\mathbf{P}(\Delta_i = l \mid A) = p_i + o(p_i)$ which follows from (20.3.5)) that

$$\mathbf{P}(\Delta_i = 1 \mid A) = p_i + o(p_i), \qquad \mathbf{P}(\Delta_i = 0 \mid A) = 1 - p_i + o(p_i).$$

It remains to once again make use of Corollary 5.4.2. $\square$