2

Fluctuations of Sums

In this chapter we consider some basic theory for sums of independent rvs. This includes classical results such as the strong law of large numbers (SLLN) in Section 2.1 and the central limit theorem (CLT) in Section 2.2, but also refinements on these theorems. In Section 2.3 refinements on the CLT are given (asymptotic expansions, large deviations, rates of convergence). Brownian and α–stable motion are introduced in Section 2.4 as weak limits of partial sum processes. They are fundamental stochastic processes which are used throughout this book. This is also the case for the homogeneous Poisson process which occurs as a special renewal counting process in Section 2.5.2. In Sections 2.5.2 and 2.5.3 we study the fluctuations of renewal counting processes and of random sums indexed by a renewal counting process. As we saw in Chapter 1, random sums are of particular interest in insurance; for example, the compound Poisson process is one of the fundamental notions in the field.

The present chapter is the basis for many other results provided in this book. Poisson random measures occur as generalisations of the homogeneous Poisson process in Chapter 5. Since most of the theory given below is classical we only sketch the main ideas of the proofs and refer to some of the relevant literature for details. We also consider extensions and generalisations of the theory for sums in Sections 8.5 and 8.6. There we look at the fine structure of a random walk, in particular at the longest success–run and large deviation

P. Embrechts et al., *Modelling Extremal Events*, Stochastic Modelling and Applied Probability 33, DOI 10.1007/978-3-642-33483-2_3,

results. The latter will find some natural applications in reinsurance (Section 8.7). An introduction to general stable processes is given in Section 8.8.

2.1 The Laws of Large Numbers

Throughout this chapter $X, X_1, X_2, \ldots$ is a sequence of iid non–degenerate rvs defined on a probability space $[\Omega, \mathcal{F}, P]$ with common df F. If we want to get a rough idea about the fluctuations of the X_n we might ask for convergence of the sequence (X_n). Unfortunately, for almost all $\omega \in \Omega$ this sequence does not converge. However, we can obtain some information about how the X_n "behave in the mean". This leads us to the consideration of the cumulative sums

$$S_0 = 0\,, \quad S_n = X_1 + \cdots + X_n, \quad n \geq 1\,,$$

and of the arithmetic (or sample) means

$$\overline{X}_n = n^{-1} S_n\,, \quad n \geq 1\,.$$

Mean values accompany our daily life. For instance, in the newspapers we are often confronted with average values in articles on statistical, actuarial or financial topics. Sometimes they occur in hidden form such as the NIKKEI, DAX, Dow Jones or other indices.

Intuitively, it is clear that an arithmetic mean should possess some sort of "stability" in n. So we expect that for large n the individual values X_i will have less influence on the order of $\overline{X}_n$, i.e. the sequence $(\overline{X}_n)$ stabilises around a fixed value (converges) as $n \to \infty$. This well–known effect is called a *law of large numbers.*

Suppose for the moment that $\sigma^2 = \operatorname{var}(X)$ is finite. Write $\mu = EX$. From Chebyshev's inequality we conclude that for $\epsilon > 0$,

$$P\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \quad \leq \quad \epsilon^{-2} \operatorname{var}\left(\overline{X}_n\right) = (n\epsilon^2)^{-1}\sigma^2 \to 0\,, \quad n \to \infty\,.$$

Hence

$$\overline{X}_n \xrightarrow{P} \mu\,, \quad n \to \infty\,.$$

This relation is called the *weak law of large numbers* (*WLLN*) or simply the *law of large numbers* (*LLN*) for the sequence (X_n). If we interpret the index of X_n as time n then $\overline{X}_n$ is an average over time. On the other hand, the expectation

$$EX = \int_\Omega X(\omega)\, dP(\omega)$$

is a weighted average over the probability space Ω. Hence the LLN tells us that, over long periods of time, the time average $\overline{X}_n$ converges to the space average EX. This is the physical interpretation of the LLN which gained it the special name *ergodic theorem*.

We saw that (X_n) obeys the WLLN if the variance σ^2 is finite. This condition can be weakened substantially:

Theorem 2.1.1 (Criterion for the WLLN)
The WLLN

$$\overline{X}_n \stackrel{P}{\to} 0$$

holds if and only if the following two conditions are satisfied:

$$\begin{aligned} nP(|X| > n) &\to 0\,, \\ EXI_{\{|X|\le n\}} &\to 0\,. \end{aligned}$$

□

Here and throughout we use the notation

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

for the *indicator function of the event* (of the set) A.

The assumptions of Theorem 2.1.1 are easily checked for the centred sequence $(X_n - \mu)$. Thus we conclude that the WLLN $\overline{X}_n \stackrel{P}{\to} \mu$ holds provided the expectation of X is finite. But we also see that the existence of a first moment is not necessary for the WLLN in the form $\overline{X}_n \stackrel{P}{\to} 0$:

Example 2.1.2 Let X be symmetric with tail

$$P(|X| > x) \sim \frac{c}{x \ln x}, \quad x \to \infty\,,$$

for some constant $c > 0$. The conditions of Theorem 2.1.1 are easily checked. Hence $\overline{X}_n \stackrel{P}{\to} 0$. However,

$$E|X| = \int_0^\infty P(|X| > x)\, dx = \infty\,.$$

□

Next we ask what conditions are needed to ensure that $\overline{X}_n$ does not only converge in probability but also *with probability 1* or *almost surely* (*a.s.*). Such a result is then called a *strong law of large numbers* (*SLLN*) for the sequence (X_n). The existence of the first moment is a necessary condition for the SLLN: given that $\overline{X}_n \stackrel{\text{a.s.}}{\to} a$ for some finite constant a we have that

$$n^{-1} X_n = n^{-1}\left(S_n - S_{n-1}\right) \stackrel{\text{a.s.}}{\to} a - a = 0\,.$$

Hence, for $\epsilon > 0$,

$$P\left(n^{-1}|X_n| > \epsilon \quad \text{i.o.}\right) = 0\,.$$

This and the Borel–Cantelli lemma (see Section 3.5) imply that for $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P\left(\left|n^{-1}X_n\right| > \epsilon\right) = \sum_{n=1}^{\infty} P\left(|X| > \epsilon n\right) < \infty\,,$$

which means that $E|X| < \infty$. This condition is also sufficient for the SLLN:

Theorem 2.1.3 (Kolmogorov's SLLN)
The SLLN

$$\overline{X}_n \stackrel{\text{a.s.}}{\to} a$$

holds for the sequence (X_n) *and some real constant* a *if and only if* $E|X| < \infty$. *Moreover, if* (X_n) *obeys the SLLN then* $a = \mu$. □

Formally, Kolmogorov's SLLN remains valid for positive (negative) rvs with infinite mean, i.e. in that case we have

$$\overline{X}_n \stackrel{\text{a.s.}}{\to} EX = \infty \quad (= -\infty)\,.$$

Example 2.1.4 (Glivenko–Cantelli theorem)
Denote by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}\,, \quad x \in \mathbb{R}\,,$$

the *empirical df* of the iid sample $X_1, \ldots, X_n$. An application of the SLLN yields that

$$F_n(x) \stackrel{\text{a.s.}}{\to} EI_{\{X \le x\}} = F(x)$$

for every $x \in \mathbb{R}$. The latter can be strengthened (and is indeed equivalent) to

$$\varDelta_n = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \stackrel{\text{a.s.}}{\to} 0\,. \tag{2.1}$$

The latter is known as the *Glivenko–Cantelli theorem*. It is one of the fundamental results in non–parametric statistics. In what follows we will frequently make use of it.
We give a proof of (2.1) for a continuous df F. For general F see Theorem 20.6 in Billingsley [70]. Let

$$-\infty = x_0 < x_1 < \cdots < x_k < x_{k+1} = \infty$$

be points such that $F(x_{i+1}) - F(x_i) < \varepsilon$ for a given $\varepsilon > 0$, $i = 0, \ldots, k$. $F(\pm\infty)$ are interpreted in the natural way as limits. By the monotonicity of F and F_n we obtain

$$\begin{aligned}\varDelta_n &= \max_{i=0,\ldots,k} \sup_{x\in(x_i,x_{i+1}]} |F_n(x) - F(x)| \\ &\le \max_{i=0,\ldots,k} \left(F_n(x_{i+1}) - F(x_i), F(x_{i+1}) - F_n(x_i)\right)\,.\end{aligned}$$

An application of the SLLN to the rhs yields

$$\limsup_{n\to\infty} \Delta_n \le \max_{i=0,\ldots,k} (F(x_{i+1}) - F(x_i)) < \varepsilon \quad \text{a.s.}$$

This concludes the proof of (2.1). The latter remains valid for stationary ergodic sequences (X_n). This is a consequence of Birkhoff's ergodic theorem (for instance Billingsley [68]) which implies that $F_n(x) \stackrel{\text{a.s.}}{\to} F(x)$ for every fixed x. $\square$

The SLLN yields an a.s. first–order approximation of the rv $\overline{X}_n$ by the deterministic quantity μ:

$$\overline{X}_n = \mu + o(1) \quad \text{a.s.}$$

The natural question that arises is:

What is the quality of this approximation?

Refinements of the SLLN are the aim of some of our future considerations. We pose a further question:

What can we conclude about the a.s. fluctuations of the sums S_n if we choose another normalising sequence?

A natural choice of normalising constants is given by the powers of n.

Theorem 2.1.5 (Marcinkiewicz–Zygmund SLLNs)
Suppose that $p \in (0,2)$. The SLLN

$$n^{-1/p}\,(S_n - a\,n) \stackrel{\text{a.s.}}{\to} 0 \tag{2.2}$$

holds for some real constant a if and only if $E|X|^p < \infty$. If (X_n) obeys the SLLN (2.2) *then we can choose*

$$a = \begin{cases} 0 & \text{if} \quad p < 1\,, \\ \mu & \text{if} \quad p \in [1,2)\,. \end{cases}$$

Moreover, if $E|X|^p = \infty$ for some $p \in (0,2)$ then for every real a,

$$\limsup_{n\to\infty} n^{-1/p}\,|S_n - a\,n| = \infty \quad \text{a.s.}$$ $\square$

This theorem gives a complete characterisation of the SLLN with normalising power functions of n. Under the conditions of Theorem 2.1.5 we obtain the following refined a.s. first–order approximation of $\overline{X}_n$:

$$\overline{X}_n = \mu + o(n^{1/p-1}) \quad \text{a.s.}\,, \tag{2.3}$$

which is valid if $E|X|^p < \infty$ for some $p \in [1,2)$.

Theorem 2.1.5 allows us to derive an elementary relationship between the large fluctuations of the sums S_n, the summands X_n and the maxima

$$M_1 = |X_1|\,, \quad M_n = \max\left(|X_1|, \ldots, |X_n|\right)\,, \quad n \ge 2\,.$$

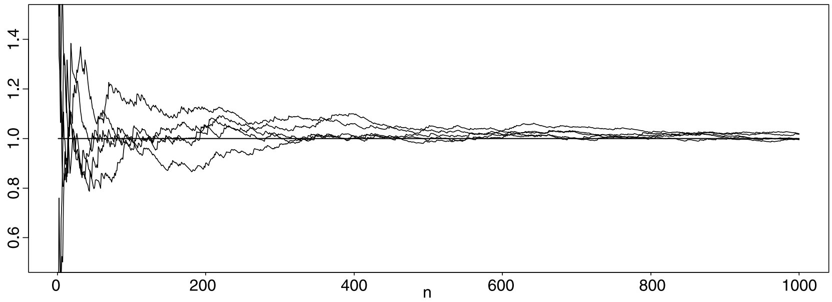

Figure 2.1.6 *Visualisation of the convergence in the SLLN: five sample paths of the process* (S_n/n) *for iid standard exponential* X_n.

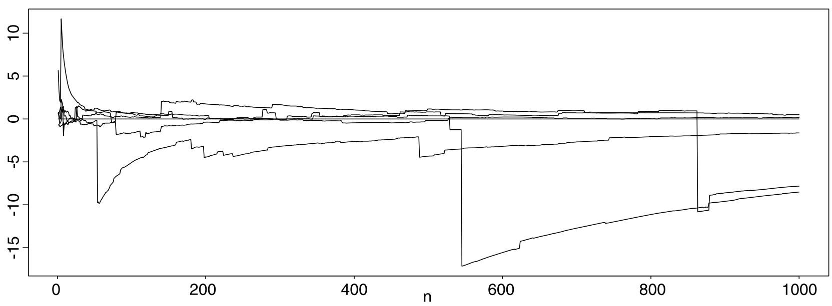

Figure 2.1.7 *Failure of the SLLN: five sample paths of the process* (S_n/n) *for iid standard symmetric Cauchy rvs* X_n *with* $E|X| = \infty$, *hence the wild oscillations of the sample paths.*

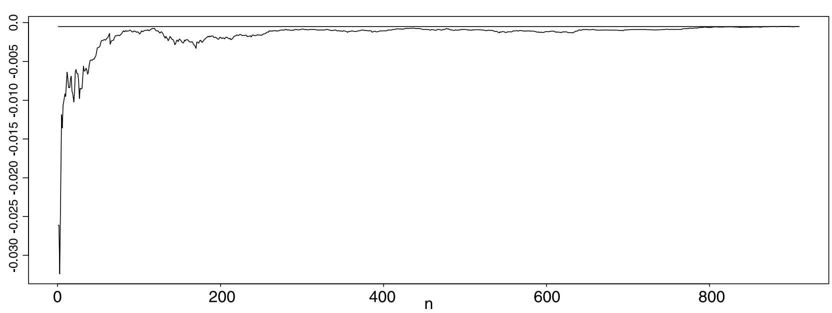

Figure 2.1.8 *The SLLN for daily log–returns of the* NIKKEI *index February 22, 1990 – October 8, 1993. The solid straight line shows the mean* -0.000501 *of these* 910 *values.*

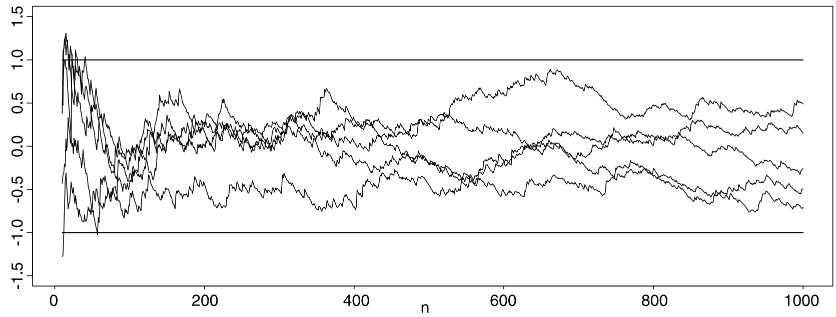

Figure 2.1.9 *Visualisation of the LIL: five sample paths of the process* $((2n \ln \ln n)^{-1/2}(S_n - n))$ *for iid standard exponential rvs* X_n.

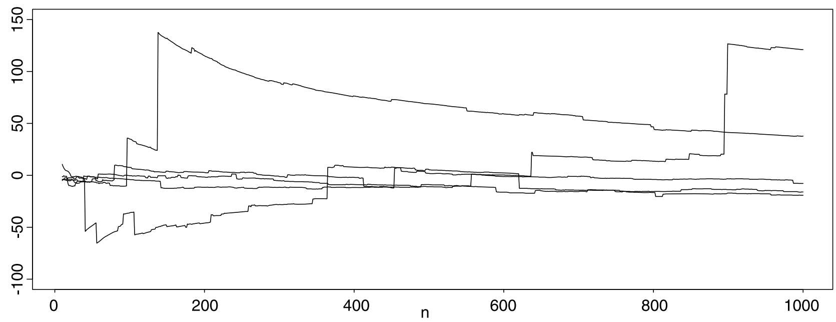

Figure 2.1.10 *Failure of the LIL: five sample paths of the process* $((2n \ln \ln n)^{-1/2} S_n)$ *for iid standard symmetric Cauchy rvs* X_n. *Notice the difference in the vertical scale!*

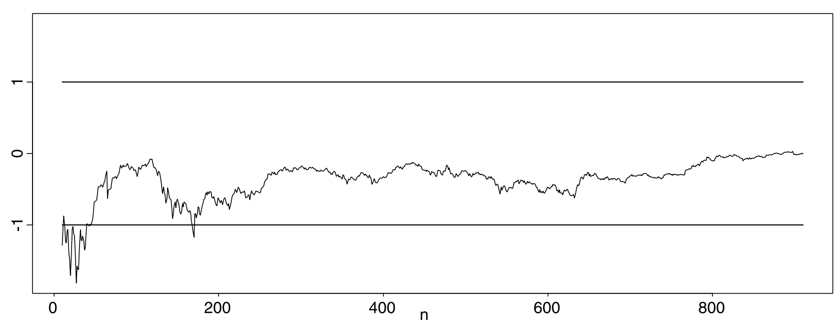

Figure 2.1.11 *The LIL for daily log–returns of the* NIKKEI *index February 22, 1990 – October 8, 1993.*

Corollary 2.1.12 *Suppose that $p \in (0,2)$. Then*

$$E|X|^p < \infty \qquad (= \infty) \tag{2.4}$$

according as

$$\limsup_{n\to\infty} n^{-1/p}\,|X_n| = 0 \qquad (= \infty) \quad \text{a.s.} \tag{2.5}$$

according as

$$\limsup_{n\to\infty} n^{-1/p} M_n = 0 \qquad (= \infty) \quad \text{a.s.} \tag{2.6}$$

according as

$$\limsup_{n\to\infty} n^{-1/p}\,|S_n - a\,n| = 0 \qquad (= \infty) \quad \text{a.s.} \tag{2.7}$$

Here a has to be chosen as in Theorem 2.1.5.

Proof. It is not difficult to see that (2.4) holds if and only if

$$\sum_{n=1}^{\infty} P\left(|X| > \epsilon n^{1/p}\right) < \infty \qquad (= \infty) \quad \forall \epsilon > 0\,.$$

A Borel–Cantelli argument yields that this is equivalent to

$$P\left(|X_n| > \epsilon n^{1/p} \quad \text{i.o.}\right) = 0 \qquad (= 1) \quad \forall \epsilon > 0\,.$$

Combining this and Theorem 2.1.5 we see that (2.4), (2.5) and (2.7) are equivalent.
The equivalence of (2.5) and (2.6) is a consequence of the elementary relation

$$\begin{aligned} n^{-1/p}\,|X_n| \quad &\le \quad n^{-1/p} M_n \\ &\le \quad \max\left(\frac{|X_1|}{n^{1/p}}, \ldots, \frac{|X_{n_0}|}{n^{1/p}}, \frac{|X_{n_0+1}|}{(n_0+1)^{1/p}}, \frac{|X_{n_0+2}|}{(n_0+2)^{1/p}}, \ldots, \frac{|X_n|}{n^{1/p}}\right) \end{aligned}$$

for every fixed $n_0 \le n$. □

This means that the asymptotic order of magnitude of the sums S_n, of the summands X_n and of the maxima M_n is roughly the same.

Another question arises from Theorem 2.1.5:

Can we choose $p = 2$ or even $p > 2$ in (2.2) ?

The answer is unfortunately *no*. More precisely, for all non–degenerate rvs X and deterministic sequences (a_n),

$$\limsup_{n\to\infty} n^{-1/2}\,|S_n - a_n| = \infty \quad \text{a.s.}$$

This is somewhat surprising because we might have expected that the more moments of X exist the smaller the fluctuations of the sums S_n. This is not the case by virtue of the central limit theorem (CLT) which we will consider in more detail in Sections 2.2 and 2.3. Indeed, the CLT requires the normalisation $n^{1/2}$ which makes a result like (2.3) for $p = 2$ impossible. However, a last a.s. refinement can still be done if the second moment of X exists:

Theorem 2.1.13 (Hartman–Wintner law of the iterated logarithm)
If $\sigma^2 = \operatorname{var}(X) < \infty$ then

$$\begin{aligned}\limsup_{n\to\infty}\,(2n\ln\ln n)^{-1/2}\,(S_n - \mu\, n) &= -\liminf_{n\to\infty}\,(2n\ln\ln n)^{-1/2}\,(S_n - \mu\, n)\\ &= \sigma \quad \text{a.s.}\end{aligned}$$

If $\sigma^2 = \infty$ then for every real sequence (a_n)

$$\limsup_{n\to\infty}\,(2n\ln\ln n)^{-1/2}\,|S_n - a_n| = \infty \quad \text{a.s.} \qquad \square$$

Hence the *law of the iterated logarithm* (*LIL*) as stated by Theorem 2.1.13 gives us the a.s. first–order approximation

$$\overline{X}_n = \mu + O\left((\ln\ln n/n)^{-1/2}\right) \quad \text{a.s.}\,,$$

which is the best possible a.s. approximation of $\overline{X}_n$ by its expectation μ. We will see in the next section that we have to change the mode of convergence if we want to derive more information about the fluctuations of the sums S_n. There we will commence with their distributional behaviour.

Notes and Comments

The WLLN for iid sequences (Theorem 2.1.1) can be found in any standard textbook on probability theory; see for instance Breiman [90], Chow and Teicher [118], Feller [235], Loève [427]. The WLLN with other normalisations and for the non–iid case has been treated for instance in Feller [235], Petrov [495, 496], or in the martingale case in Hall and Heyde [312].

More insight into the weak limit behaviour of sums is given by so–called *rates of convergence* in the LLN, i.e. by statements about the order of the probabilities

$$P\left(|S_n - a_n| > b_n\right), \quad n \to \infty,$$

for appropriate normalising and centring sequences (b_n), (a_n). We refer to Petrov [495, 496] and the literature cited therein.

The classical Kolmogorov SLLN (Theorem 2.1.3) is part of every standard textbook on probability theory, and the Marcinkiewicz–Zygmund SLLNs can be found for instance in Stout [608]. Necessary and sufficient conditions under non–standard normalisations and for the non–iid case are given for instance in Petrov [495, 496] or in Stout [608]. In Révész [538] and Stout [608] various SLLNs are proved for sequences of dependent rvs. Some remarks on the *convergence rate in the SLLN*, i.e. on the order of the probabilities

$$P\left(\sup_{k \geq n} |S_k - a_k| > b_n\right), \quad n \to \infty,$$

can be found in Petrov [495, 496].

The *ergodic theorem* as mentioned above is a classical result which holds for stationary ergodic (X_n); see for instance Billingsley [68] or Stout [608].

The limit in the SLLN for a sequence (X_n) of independent rvs is necessarily a constant. This is due to the so–called 0–1 *law*; for different versions see for instance Stout [608]. The limit in the SLLN for a sequence of dependent rvs can be a genuine rv.

The Marcinkiewicz–Zygmund SLLNs for an iid sequence exhibit another kind of 0–1 *behaviour:* either the SLLN holds with a constant limit for the normalisation $n^{1/p}$ or, with the same normalisation, the sums fluctuate wildly with upper or lower limit equal to $\pm\infty$. This behaviour is typical for a large class of normalisations (cf. *Feller's SLLN*; see Feller [233], Petrov [495, 496], Stout [608]). Similar behaviour can be observed for a large class of rvs with infinite variance which includes the class of α–stable rvs, $\alpha < 2$, and their domains of attraction; see Section 2.2. To be precise, suppose that for some constant $c > 0$,

$$x^{-2} E X^2 I_{\{|X| \leq x\}} \leq c\, P(|X| > x), \quad x > 0. \tag{2.8}$$

Let (b_n) be any real sequence such that $b_n \uparrow \infty$ and if $E|X| < \infty$ suppose that $\mu = 0$. Then

$$\limsup_{n \to \infty} b_n^{-1} |S_n| = 0 \qquad (= \infty) \quad \text{a.s.}$$

according as

$$\sum_{n=1}^{\infty} P\left(|X| > b_n\right) < \infty \qquad (= \infty).$$

The latter relation is a moment condition. Moreover, the relations between S_n, X_n and M_n (with normalisation b_n) corresponding to Corollary 2.1.12 hold. This SLLN is basically due to Heyde [323]; see Stout [608].

The SLLN can also be extended to sums S_n of *independent but not identically distributed rvs*. There exist results under the condition of finiteness of the second moments of the X_n. The results are typically of the form

$$b_n^{-1}(S_n - ES_n) \stackrel{\text{a.s.}}{\to} 0\,,$$

where b_n is the variance of S_n; see for instance Petrov [495, 496]. However, it seems difficult to use such a model for statistical inference as long as the class of distributions of the X_n is not specified. A more sensitive study is possible for sequences of iid rvs with given deterministic weights. *Weighted sums*

$$T_n = \sum_k w_n(k)\, X_k$$

are standard models in the statistical literature. For example, in time series analysis the linear processes, including the important class of ARMA processes, are weighted sums of iid rvs; see Chapter 7. The rvs T_n can be considered as a mean which, in contrast to $\overline{X}_n$, gives different weight to the observations X_k. Examples are the discounted sums $\sum_{k\geq 0} z^k\, X_k$ whose asymptotic behaviour (as $z \uparrow 1$) is well studied (so–called Abel summation). There is quite a mass of literature on the a.s. behaviour of the weighted sums T_n. Results of SLLN–type can be found for instance in Mikosch and Norvaiša [461] or in Stout [608]. Overviews of *summability methods* have been given in Bingham and Maejima [71], Maejima [433], Mikosch and Norvaiša [459, 460].

The *Hartman–Wintner LIL* (Theorem 2.1.13) is included in standard textbooks; see for instance Feller [235]. Different proofs and ramifications for non–identically distributed rvs and dependent observations can be found in Csörgő and Révész [145], Hall and Heyde [312], Petrov [495, 496], Stout [608].

There exists a well developed theory about fluctuations of sums of iid rvs with or without normalisation. The latter is also called a *random walk*. For example, necessary and sufficient conditions have been derived for relations of type

$$\begin{aligned} \limsup_{n\to\infty} \quad b_n^{-1}\,|S_n - a_n| &= c_1 \in (0,\infty) && \text{a.s.}\,,\\ \limsup_{n\to\infty} \quad b_n^{-1}\,(S_n - a_n) &= c_2 \in (-\infty,\infty) && \text{a.s.} \end{aligned}$$

(*generalised LIL, one–sided LIL*) for given (a_n), (b_n) and constants c_1, c_2, and also results about the existence of such normalising or centring constants; see for instance Kesten [378], Klass [381, 382], Martikainen [441, 442], Pruitt

[515, 517]. These results give some insight into the complicated nature of the fluctuations of sums. However, they are very difficult to apply: the sequence (b_n) is usually constructed in such a way that one has to know the whole distribution tail of X. Thus these results are very sensitive to changes in the distribution.

A further topic of research has been concentrated around *cluster phenomena* of the sums S_n (normalised or non–normalised) and the general properties of random walks. We refer to Cohen [130], Erickson [222, 223], Kesten [378], Révész [540], Spitzer [604], Stout [608]. The set of *a.s. limit points* of the sequence of normalised sums can be very complicated. However, in many situations the set of a.s. limit points coincides with a closed interval (finite or infinite). The following basic idea from elementary calculus is helpful: let (a_n) be a sequence of real numbers such that $a_n - a_{n-1} \to 0$. Then every point in the interval $[\liminf_{n\to\infty} a_n, \limsup_{n\to\infty} a_n]$ is a limit point of (a_n). Applying this to the Hartman–Wintner LIL with $EX^2 < \infty$ and $A_n = (2n \ln\ln n)^{-1/2}(S_n - \mu\, n)$ we see that $A_n - A_{n-1} \stackrel{\text{a.s.}}{\to} 0$ and hence every point in $[-\sigma, \sigma]$ is a limit point of (A_n) for almost every sample path. This remarkable property means that the points $A_1, A_2, \ldots$ fill the interval $[-\sigma, \sigma]$ densely for almost every sample path. This is somehow counter–intuitive since at the same time $A_n \stackrel{P}{\to} 0$.

2.2 The Central Limit Problem

In the preceding section we saw that the sums S_n of the iid sequence (X_n) diverge a.s. when normalised with $n^{1/2}$. However, we can still get information about the growth of $n^{-1/2} S_n$ if we change to *convergence in distribution* (*weak convergence*).

We will approach the problem from a more general point of view. We ask:

What are the possible (non–degenerate) limit laws for the sums S_n when properly normalised and centred?

This is a classical question in probability theory. Many famous probabilists of this century have contributed to its complete solution: Khinchin, Lévy, Kolmogorov, Gnedenko, Feller,.... It turns out that this question is closely related to another one:

Which distributions satisfy the identity in law

$$c_1 X_1 + c_2 X_2 \stackrel{d}{=} b(c_1, c_2) X + a(c_1, c_2) \tag{2.9}$$

for all non–negative numbers c_1, c_2 and appropriate real numbers $b(c_1, c_2) > 0$ and $a(c_1, c_2)$?

In other words, which classes of distributions are closed (up to changes of location and scale) under convolution and multiplication with real numbers? The possible limit laws for sums of iid rvs are just the distributions which satisfy (2.9) for all non–negative c_1, c_2. Many classes of distributions are closed with respect to convolution but the requirement (2.9) is more stringent. For example, the convolution of two Poisson distributions is a Poisson distribution. However, the Poisson distributions do not satisfy (2.9).

Definition 2.2.1 (Stable distribution and rv)
A rv (a distribution, a df) is called stable *if it satisfies* (2.9) *for iid X, X_1, X_2, for all non–negative numbers c_1, c_2 and appropriate real numbers $b(c_1, c_2) > 0$ and $a(c_1, c_2)$.* □

Now consider the sum S_n of iid stable rvs. By (2.9) we have for some real constants a_n and $b_n > 0$ and $X = X_1$,

$$S_n = X_1 + \cdots + X_n \stackrel{d}{=} b_n X + a_n\,, \quad n \geq 1\,,$$

which we can rewrite as

$$b_n^{-1}(S_n - a_n) \stackrel{d}{=} X\,.$$

We conclude that, if a distribution is stable, then it is a limit distribution for sums of iid rvs. Are there any other possible limit distributions? The answer is NO:

Theorem 2.2.2 (Limit property of a stable law)
The class of the stable (non–degenerate) distributions coincides with the class of all possible (non–degenerate) limit laws for (properly normalised and centred) sums of iid rvs. □

Because of the importance of the class of stable distributions it is necessary to describe them analytically. The most common way is to determine their characteristic functions (chfs):

Theorem 2.2.3 (Spectral representation of a stable law)
A stable distribution has chf

$$\phi_X(t) = E\exp\{iXt\} = \exp\{i\gamma t - c|t|^\alpha(1 - i\beta\,\mathrm{sign}\,(t)\,z(t,\alpha))\}\,, \quad t \in \mathbb{R}\,, \tag{2.10}$$

where γ is a real constant, $c > 0$, $\alpha \in (0,2]$, $\beta \in [-1,1]$, and

$$z(t,\alpha) = \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right) & \text{if} \quad \alpha \neq 1\,, \\ -\dfrac{2}{\pi}\ln|t| & \text{if} \quad \alpha = 1\,. \end{cases}$$

□

Remarks. 1) We note that we can formally include the case $c = 0$ which corresponds to a degenerate distribution. Every sequence (S_n) can be normalised and centred in such a way that it converges to a constant (for instance zero) in probability. Thus this trivial limit belongs to the class of the possible limit rvs. However, it is not of interest in the context of weak convergence and therefore excluded from our considerations.

2) The quantity γ is just a location parameter. For the rest of this section we assume $\gamma = 0$.

3) The most important parameter in this representation is α. It determines the basic properties of this class of distributions (moments, tails, asymptotic behaviour of sums, normalisation etc.). □

Definition 2.2.4 *The number* α *in the chf* (2.10) *is called the* characteristic exponent, *the corresponding distribution* α–stable. □

Remarks. 4) For $\alpha = 2$ we obtain the *normal* or *Gaussian* distributions. In this case, we derive from (2.10) the well known chf

$$\phi_X(t) = \exp\left\{-ct^2\right\}$$

of a Gaussian rv with mean zero and variance $2c$. Thus one of the most important distributions in probability theory and mathematical statistics is a stable law. We also see that the normal law is determined just by two parameters (mean and variance) whereas the other α–stable distributions depend on four parameters. This is due to the fact that a normal distribution is always symmetric (around its expectation) whereas a stable law for $\alpha < 2$ can be asymmetric and even be concentrated on a half–axis (for $\alpha < 1$).

5) Another well–known class of stable distributions corresponds to $\alpha = 1$: the *Cauchy laws* with chf

$$\phi_X(t) = \exp\left\{-c|t|\left(1 + i\beta\frac{2}{\pi}\,\mathrm{sign}\,(t)\ln|t|\right)\right\}.$$

6) For fixed α, the parameters c and β determine the nature of the distribution. The parameter c is a scaling constant which corresponds to $c = \sigma^2/2$ in the Gaussian case and has a similar function as the variance in the non–Gaussian case (where the variance is infinite). The parameter β describes the skewness of the distribution. We see that the chf $\phi_X(t)$ is real–valued if and only if $\beta = 0$. The chf

$$\phi_X(t) = \exp\{-c|t|^\alpha\} \tag{2.11}$$

corresponds to a symmetric rv X. We will sometimes use the notation *sαs* for *symmetric α–stable.* If $\beta = 1$ and $\alpha < 1$ the rv X is positive, and for

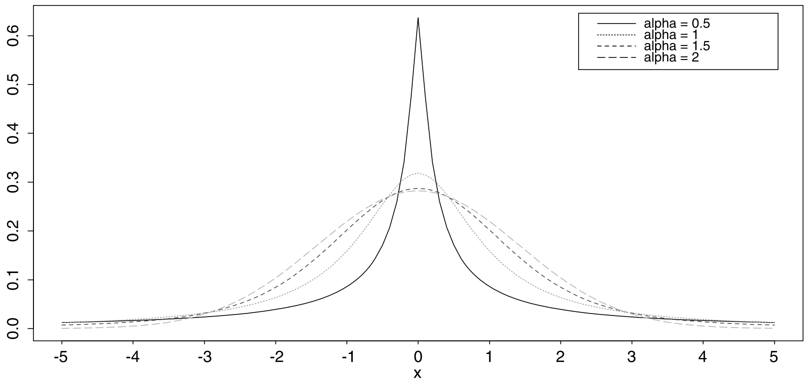

Figure 2.2.5 *Densities of $s\alpha s$ rvs with $c = 1$.*

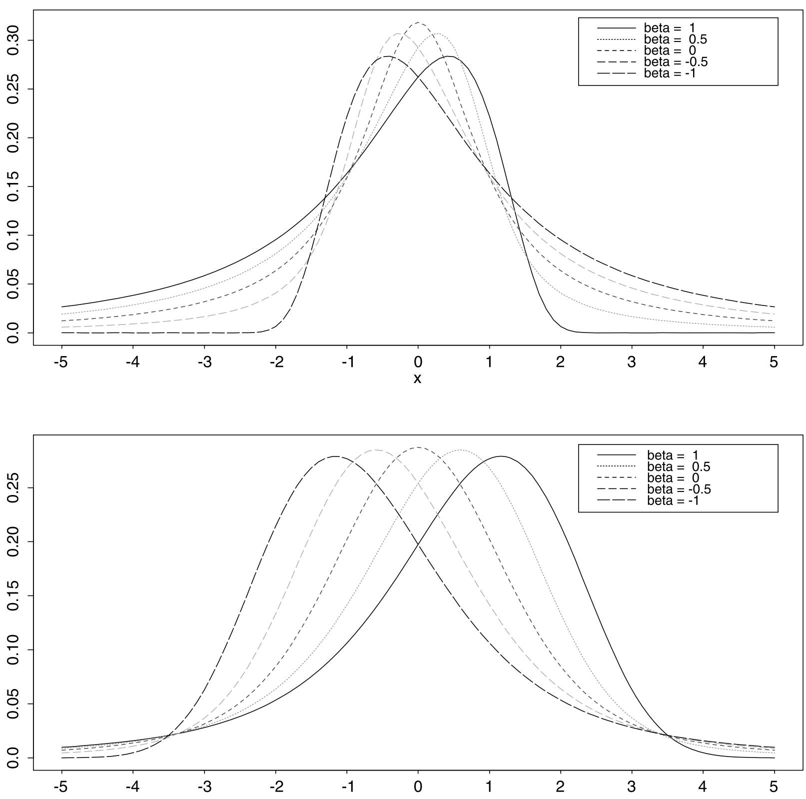

Figure 2.2.6 *Densities of $1-$ and of $1.5-$stable rvs (top, bottom) with $c = 1$.*

$\beta = -1$ and $\alpha < 1$ it is negative. In the cases $|\beta| < 1$, $\alpha < 1$, or $\alpha \in [1, 2]$, the rv X has the whole real axis as support. However, if $\beta = 1$, $\alpha \in [1, 2)$ then $P(X \le -x) = o(P(X > x))$ as $x \to \infty$. From the chf (2.10) we also deduce that $-X$ is α–stable with parameters c and $-\beta$. It can be shown that every α–stable rv X with $|\beta| < 1$ is equal in law to $X' - X'' + c_0$ for independent α–stable rvs X', X'' both with parameter $\beta = 1$ and a certain constant c_0.

7) We might ask why we used the inconvenient (from a practical point of view) representation of α–stable rvs via their chf. The answer is simple: it is the best analytic way of characterising all members of this class. Although the α–stable laws are absolutely continuous, their densities can be expressed only by complicated special functions; see Hoffmann–Jørgensen [329] and Zolotarev [646]. Only in a few cases which include the Gaussian ($\alpha = 2$), the symmetric Cauchy ($\alpha = 1, \beta = 0$) and the stable inverse Gaussian ($\alpha = 1/2, \beta = 1$), are these densities expressible explicitly via elementary functions. But there exist asymptotic expansions of the α–stable densities in a neighbourhood of the origin or of infinity; see Ibragimov and Linnik [350] and Zolotarev [645]. Therefore the α–stable distributions (with a few exceptions) are not easy to handle. In particular, they are difficult to simulate; see for instance Janicki and Weron [354]. □

Next we ask:

Given an α–stable distribution G_α, what conditions imply that the normalised and centred sums S_n converge weakly to G_α?

This question induces some further problems:

How must we choose constants $a_n \in \mathbb{R}$ and $b_n > 0$ such that

$$b_n^{-1}(S_n - a_n) \xrightarrow{d} G_\alpha\,? \tag{2.12}$$

Can it happen that different normalising or centring constants imply convergence to different limit laws?

The last question can be answered immediately: the convergence to types theorem (Theorem A1.5) ensures that the limit law is uniquely determined up to positive affine transformations.

Before we answer the other questions we introduce some further notions:

Definition 2.2.7 (Domain of attraction)
We say that the rv X (the df F of X, the distribution of X) belongs to the domain of attraction of the α–stable distribution G_α *if there exist constants $a_n \in \mathbb{R}$, $b_n > 0$ such that* (2.12) *holds. We write $X \in \mathrm{DA}(G_\alpha)$ ($F \in \mathrm{DA}(G_\alpha)$) and say that (X_n)* satisfies the central limit theorem (CLT) with limit G_α. □

If we are interested only in the fact that X (or F) is attracted by some α–stable law whose concrete form is not of interest we will simply write $X \in \mathrm{DA}(\alpha)$ (or $F \in \mathrm{DA}(\alpha)$).

The following result characterises the domain of attraction of a stable law completely. Here and in the remainder of this section we will need some facts about *slowly and regularly varying* functions which are given in Appendix A3.1. We recall that a (measurable) function L is slowly varying if $\lim_{x\to\infty} L(tx)/L(x) = 1$ for all $t > 0$.

Theorem 2.2.8 (Characterisation of domain of attraction)

(a) *The df F belongs to the domain of attraction of a normal law if and only if*

$$\int_{|y|\le x} y^2 \, dF(y)$$

is slowly varying.

(b) *The df F belongs to the domain of attraction of an α–stable law for some $\alpha < 2$ if and only if*

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha} \, L(x) \,, \quad 1 - F(x) = \frac{c_2 + o(1)}{x^\alpha} \, L(x) \,, \quad x \to \infty \,,$$

where L is slowly varying and c_1, c_2 are non–negative constants such that $c_1 + c_2 > 0$. □

First we study the case $\alpha = 2$ more in detail. If $EX^2 < \infty$ then

$$\int_{|y|\le x} y^2 \, dF(y) \to EX^2 \,, \quad x \to \infty \,,$$

hence $X \in \mathrm{DA}(2)$. Moreover, by Proposition A3.8(f) we conclude that slow variation of $\int_{|y|\le x} y^2 \, dF(y)$ is equivalent to the tail condition

$$G(x) = P(|X| > x) = o\left(x^{-2} \int_{|y|\le x} y^2 \, dF(y) \right) \,, \quad x \to \infty \,. \tag{2.13}$$

Thus we derived

Corollary 2.2.9 (Domain of attraction of a normal distribution)
A rv X is in the domain of attraction of a normal law if and only if one of the following conditions holds:

(a) $EX^2 < \infty$,
(b) $EX^2 = \infty$ *and* (2.13). □

The situation is completely different for $\alpha < 2$: $X \in \mathrm{DA}(\alpha)$ implies that

$$G(x) = x^{-\alpha} L(x)\,, \quad x > 0\,, \tag{2.14}$$

for a slowly varying function L and

$$x^2 G(x) \Big/ \int_{|y| \le x} y^2 \, dF(y) \to \frac{2-\alpha}{\alpha}, \quad x \to \infty\,. \tag{2.15}$$

The latter follows from Proposition A3.8(e). Hence the second moment of X is infinite. Relation (2.14) and Corollary 2.2.9 show that the domain of attraction of the normal distribution is much more general than the domain of attraction of an α–stable law with exponent $\alpha < 2$. We see that DA(2) contains at least all distributions that have a second finite moment.

From Corollary 2.2.9 and from (2.14) we conclude the following about the moments of distributions in $\mathrm{DA}(\alpha)$:

Corollary 2.2.10 (Moments of distributions in $\mathrm{DA}(\alpha)$)
If $X \in \mathrm{DA}(\alpha)$ then

$$\begin{aligned} E|X|^\delta &< \infty \quad \textit{for } \delta < \alpha\,, \\ E|X|^\delta &= \infty \quad \textit{for } \delta > \alpha \textit{ and } \alpha < 2\,. \end{aligned}$$

In particular,

$$\begin{aligned} \mathrm{var}(X) &= \infty \quad \textit{for } \alpha < 2\,, \\ E|X| &< \infty \quad \textit{for } \alpha > 1\,, \\ E|X| &= \infty \quad \textit{for } \alpha < 1\,. \end{aligned}$$

□

Note that $E|X|^\alpha = \int_0^\infty P(|X|^\alpha > x) dx < \infty$ is possible for certain $X \in \mathrm{DA}(\alpha)$, but $E|X|^\alpha = \infty$ for an α–stable X for $\alpha < 2$. Recalling the results of the preceding section we can apply the Marcinkiewicz–Zygmund SLLNs (Theorem 2.1.5) as well as Heyde's SLLN ((2.8) is satisfied in view of (2.15)) to α–stable rvs with $\alpha < 2$, and these theorems show that the sample paths of (S_n) fluctuate wildly because of the non–existence of the second moment.

Next we want to find appropriate normalising and centring constants for the CLT (2.12). Suppose for the moment that X is $s\alpha s$ with chf $\phi_X(t) = \exp\{-c|t|^\alpha\}$; cf. (2.11). We see that

$$\begin{aligned} E \exp\left\{ itn^{-1/\alpha} S_n \right\} &= \left(\exp\left\{ -c|n^{-1/\alpha} t|^\alpha \right\} \right)^n \\ &= \exp\left\{ -c|t|^\alpha \right\} \\ &= E \exp\{itX\}\,. \end{aligned} \tag{2.16}$$

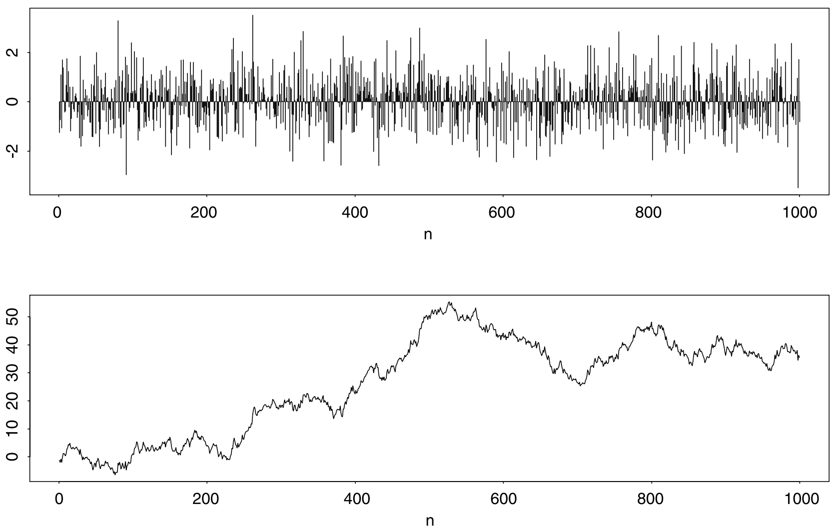

Figure 2.2.11 *One sample path of the process* (X_n) (*top*) *and of the corresponding path* (S_n) (*bottom*) *for* X *with a standard normal distribution.*

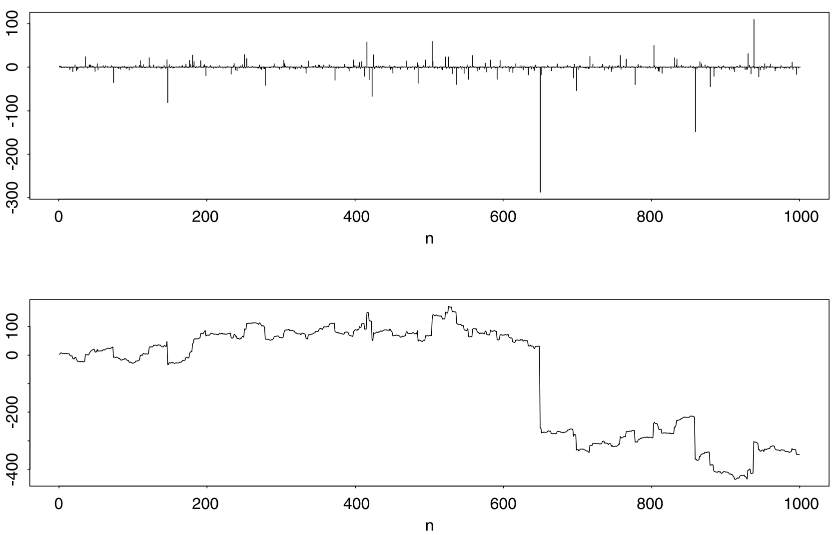

Figure 2.2.12 *One sample path of the process* (X_n) (*top*) *and of the corresponding path* (S_n) (*bottom*) *for* X *with a standard symmetric Cauchy distribution.*

Thus

$$n^{-1/\alpha} S_n \stackrel{d}{=} X$$

which gives us a rough impression of the order of the normalising constants.

For symmetric $X \in \mathrm{DA}(\alpha)$ one can show that the relation

$$E \exp\{itX\} = \exp\{-|t|^\alpha L_1(1/t)\}$$

holds in a neighbourhood of the origin, with a slowly varying function L_1 which is closely related to the slowly varying functions which occur in Theorem 2.2.8; see Theorem 2.6.5 in Ibragimov and Linnik [350]. Now we can apply the same idea as in (2.16) although this time we will have to compensate for the slowly varying function L_1. Thus it is not surprising that the normalising constants in the CLT (2.12) are of the form

$$b_n = n^{1/\alpha} L_2(n)$$

for a slowly varying function L_2. To be more precise, introduce the quantities

$$\begin{aligned} K(x) &= x^{-2} \int_{|y| \le x} y^2 \, dF(y) \,, \quad x > 0 \,, \\ Q(x) &= G(x) + K(x) = P(|X| > x) + K(x) \,, \quad x > 0 \,. \end{aligned}$$

Note that $Q(x)$ is continuous and decreasing on $[x_0, \infty)$ where x_0 denotes the infimum of the support of $|X|$.

Proposition 2.2.13 (Normalising constants in the CLT)
The normalising constants in the CLT for $F \in \mathrm{DA}(\alpha)$ can be chosen as the unique solution of the equation

$$Q(b_n) = n^{-1} \,, \quad n \ge 1 \,. \tag{2.17}$$

In particular, if $\sigma^2 = \mathrm{var}(X) < \infty$ and $EX = 0$ then

$$b_n \sim n^{1/2} \sigma, \quad n \to \infty \,.$$

If $\alpha < 2$ we can alternatively choose (b_n) such that

$$b_n = \inf\{y : G(y) < n^{-1}\} \,, \quad n \ge 1 \,. \tag{2.18}$$

We note that (2.18) implies that

$$G(b_n) \sim n^{-1} \,, \quad n \to \infty \,,$$

and that, in view of (2.14),

$$b_n = n^{1/\alpha} L_3(n) \,, \quad n \ge 1 \,,$$

for an appropriate slowly varying function L_3.

Sketch of the proof. We omit the calculations leading to (2.17) and restrict ourselves to the particular cases. For a proof we refer to Ibragimov and Linnik [350], Section II.6.
If $EX^2 < \infty$ then

$$G(x) = o\left(x^{-2}\right), \quad K(x) = x^{-2}\, EX^2(1+o(1)), \quad x \to \infty.$$

Hence, if $EX = 0$,

$$n^{-1} = Q(b_n) \sim b_n^{-2}\sigma^2 .$$

If $\alpha < 2$ then, using (2.15), we see immediately that (2.17) and (2.18) yield asymptotically equivalent sequences (b_n) and (b_n'), say, which means that $b_n \sim c\, b_n'$ for a positive constant c. □

Proposition 2.2.14 (Centring constants in the CLT)
The centring constants a_n in the CLT (2.12) *can be chosen as*

$$a_n = n \int_{|y| \le b_n} y\, dF(y)\,, \tag{2.19}$$

where b_n is given in Proposition 2.2.13. In particular, we can take $a_n = \widetilde{a}\, n$, where

$$\widetilde{a} = \begin{cases} \mu & \text{if} \quad \alpha \in (1,2]\,, \\ 0 & \text{if} \quad \alpha \in (0,1)\,, \\ 0 & \text{if} \quad \alpha = 1 \text{ and } F \text{ is symmetric.} \end{cases} \tag{2.20}$$

□

For a proof we refer to Ibragimov and Linnik [350], Section II.6. Now we formulate a general version of the CLT.

Theorem 2.2.15 (General CLT)
Suppose that $F \in \mathrm{DA}(\alpha)$ for some $\alpha \in (0,2]$.

(a) *If $EX^2 < \infty$ then*

$$\left(\sigma n^{1/2}\right)^{-1} (S_n - \mu\, n) \xrightarrow{d} \Phi$$

for the standard normal distribution Φ with mean zero and variance 1.

(b) *If $EX^2 = \infty$ and $\alpha = 2$ or if $\alpha < 2$ then*

$$\left(n^{1/\alpha} L_4(n)\right)^{-1} (S_n - a_n) \xrightarrow{d} G_\alpha$$

for an α–stable distribution G_α, an appropriate slowly varying function L_4 and centring constants as in (2.19).
In particular,

$$\left(n^{1/\alpha} L_4(n)\right)^{-1} \, (S_n - \widetilde{a}\, n) \xrightarrow{d} G_\alpha \,,$$

where $\widetilde{a}$ *is defined in* (2.20). □

We notice that it is possible for the normalising constants in the CLT to be of the special form $b_n = c\,n^{1/\alpha}$ for some constant c. This happens for instance if $EX^2 < \infty$ or if X is α–stable. There is a special name for this situation:

Definition 2.2.16 (Domain of normal attraction)
We say that X (*or* F) *belongs to the* domain of normal attraction of an α–stable distribution G_α ($X \in \mathrm{DNA}(G_\alpha)$ or $F \in \mathrm{DNA}(G_\alpha)$) *if* $X \in \mathrm{DA}(G_\alpha)$ *and if in the CLT we can choose the normalisation* $b_n = c\,n^{1/\alpha}$ *for some positive constant* c. □

If we are interested only in the fact that X (or F) belongs to the DNA of some α–stable distribution we write $X \in \mathrm{DNA}(\alpha)$ (or $F \in \mathrm{DNA}(\alpha)$). We recall the characterisation of the domains of attraction via tails; see Theorem 2.2.8. Then (2.17) implies the following:

Corollary 2.2.17 (Characterisation of DNA)

(*a*) *The relation* $F \in \mathrm{DNA}(2)$ *holds if and only if* $EX^2 < \infty$.
(*b*) *For* $\alpha < 2$, $F \in \mathrm{DNA}(\alpha)$ *if and only if*

$$F(-x) \sim c_1 x^{-\alpha} \quad \text{and} \quad 1 - F(x) \sim c_2 x^{-\alpha} \,, \quad x \to \infty \,,$$

for non–negative constants c_1, c_2 *such that* $c_1 + c_2 > 0$.

In particular, every α*–stable distribution is in its own DNA.* □

So we see that $F \in \mathrm{DNA}(\alpha)$, $\alpha < 2$, actually means that the corresponding tail $\overline{G}(x)$ has *power law* or *Pareto–like* behaviour. Note that a df F with Pareto–like tail $\overline{G}(x) \sim cx^{-\alpha}$ for some $\alpha \geq 2$ is in DA(2), and if $\alpha > 2$, then $F \in \mathrm{DNA}(2)$.

Notes and Comments

The theory above is classical and can be found in detail in Araujo and Giné [19], Bingham, Goldie and Teugels [72], Feller [235], Gnedenko and Kolmogorov [267], Ibragimov and Linnik [350], Loève [427] and many other textbooks. For applications of the CLT and related weak convergence results to asymptotic inference in statistics we refer to Ferguson [236] or Serfling [576].

There exists some more specialised literature on stable distributions and stable processes. Mijnheer [456] is one of the first monographs on the topic. Zolotarev [645] covers a wide range of interesting properties of stable distributions, including asymptotic expansions of the stable densities and many useful representations and transformation formulae. Some limit theory for distributions in the domain of attraction of a stable law is given in Christoph and Wolf [119]. An encyclopaedic treatment of stable laws, multivariate stable distributions and stable processes can be found in Samorodnitsky and Taqqu [565]; see also Kwapień and Woyczyński [411] and Janicki and Weron [354]. The latter book also deals with numerical aspects, in particular the simulation of stable rvs and processes. An introduction to stable random vectors and processes will be provided in Section 8.8.

Recently there have been some efforts to obtain efficient methods for the numerical calculation of stable densities. This has been a problem for many years and was one of the reasons that practitioners expressed doubts about the applicability of stable distributions for modelling purposes. McCulloch and Panton [451] and Nolan [480, 481] provided tables and software for calculating stable densities for a large variety of parameters α and β. Their methods allow one to determine those densities for small and moderate arguments with high accuracy; the determination of the densities in their tails needs further investigation. Figures 2.2.5 and 2.2.6 were obtained using software kindly provided to us by John Nolan.

The central limit problem has also been solved for independent non–iid rvs; see for instance Feller [235], Gnedenko and Kolmogorov [267], Ibragimov and Linnik [350], Petrov [495, 496]. To be precise, let $(X_{nk})_{k=1,\dots,n}$, $n = 1, 2, \dots$ be a triangular scheme of row–wise independent rvs satisfying the condition of *infinitesimality*:

$$\max_{k=1,\dots,n} P(|X_{nk}| > \epsilon) \to 0\,, \quad n \to \infty\,, \quad \epsilon > 0\,.$$

The class of possible limit laws for the sums $\sum_{k=1}^{n} X_{nk}$ consists of the *infinitely divisible distributions* including most distributions of interest in statistics. For example, the stable distributions and the Poisson distribution belong to this class. A rv Y (and its distribution) is infinitely divisible if and only if we can decompose it in law:

$$Y \stackrel{d}{=} Y_{n1} + \cdots + Y_{nn}$$

for every n, where $(Y_{nk})_{k=1,\dots,n}$ are iid rvs with possibly different common distribution for different n. There exist representations of the chf of an infinitely divisible law. Theorem 2.2.3 is a particular case for stable laws.

As in the case of a.s. convergence, see Section 2.1, *weighted sums* are particularly important for applications in statistics. The general limit theory for non–iid rvs can sometimes be applied to weighted sums. However, there exist quite a few results for special *summability methods*; for references see the Notes and Comments of Section 2.1.

2.3 Refinements of the CLT

In this section we consider some refinements of the results of the previous section. We will basically restrict ourselves to the case when $EX^2 < \infty$ and briefly comment on the other ones. It is natural to ask:

How can we determine and improve the quality of the approximation in the CLT?

Berry–Esséen Theorem

Let Φ denote the df of the standard normal distribution and write

$$G_n(x) = P\left(\frac{S_n - n\,\mu}{\sigma\sqrt{n}} \le x\right), \quad x \in \mathbb{R}.$$

From the previous section we know that

$$\Delta_n = \sup_{x\in\mathbb{R}} |G_n(x) - \Phi(x)| \to 0\,. \tag{2.21}$$

There we formulated only a weak convergence result, i.e. convergence of G_n at every continuity point of Φ. However, Φ is continuous and therefore (2.21) holds; see Appendix A1.1.

One can show that the rate at which Δ_n converges to zero can be arbitrarily slow if we require no more than a finite second moment of X. The typical rate of convergence is $1/\sqrt{n}$ provided the third moment of X exists. We give here a *non–uniform* version of the well–known *Berry–Esséen theorem*:

Theorem 2.3.1 (Berry–Esséen theorem)
Suppose that $E|X|^3 < \infty$. Then

$$|G_n(x) - \Phi(x)| \le \frac{c}{\sqrt{n}(1+|x|)^3}\,\frac{E|X-\mu|^3}{\sigma^3} \tag{2.22}$$

for all x, where c is a universal constant. In particular,

$$\Delta_n \le \frac{c}{\sqrt{n}}\,\frac{E|X-\mu|^3}{\sigma^3}\,. \tag{2.23}$$

□

From (2.22) we have learnt that the quality of the approximation can be improved substantially for large x. Moreover, the rate in (2.22) and (2.23) is influenced by the order of magnitude of the ratio $E|X-\mu|^3/\sigma^3$ and of the constant c. This is of crucial importance if n is small.

The rates in (2.22) and (2.23) are optimal in the sense that there exist sequences (X_n) such that $\Delta_n \asymp (1/\sqrt{n})$. For example, this is true for symmetric

Bernoulli rvs assuming the values $+1$ and -1 with equal probability. On the other hand, the Berry–Esséen estimate is rather pessimistic and can be improved when special conditions on X are satisfied, for instance the existence of a smooth density, moment generating function etc.

Results of Berry–Esséen type have been studied for $X \in \mathrm{DNA}(\alpha)$ and $X \in \mathrm{DA}(\alpha)$ with $\alpha < 2$ as well. A unifying result such as Theorem 2.3.1 does not exist and cannot be expected. The results require very special knowledge about the structure of the df in DNA and DA and are difficult to apply.

Notes and Comments

References for the speed of convergence in the CLT are Petrov [495, 496] and Rachev [520].

A proof of the classical Berry–Esséen theorem and its non–uniform version using Fourier methods can be found in Petrov [495, 496]. Also given there are results of Berry–Esséen type in the non–iid situation and for iid rvs under the existence of the $(2+\delta)$th moment for some $\delta \in (0,1]$. The rate of convergence is the slower, the less δ is. In the iid situation the speed is just $n^{-\delta/2}$. The rate can be improved under special conditions on the rv X, although, as mentioned before, an increase of the power of the moments is not sufficient for this.

Attempts have been made to calculate the best constants in the Berry–Esséen theorem; see Petrov [495, 496]. One can take 0.7655 in (2.23) and $0.7655 + 8(1+e)$ in (2.22).

Studies of the rate of convergence in $\mathrm{DA}(\alpha)$ for $\alpha < 2$ can be found in Christoph and Wolf [119] or in Rachev [520]. The former concentrates more on classical methods whereas the latter proposes other techniques for estimating the rate of convergence. For example, appropriate metrics (L^p, Lévy and Lévy–Prokhorov metrics) for weak convergence are introduced and then applied to sums of iid rvs. We also refer to results by de Haan and Peng [298] and Hall [308] who study rates of convergence under second–order regular variation conditions on the tail $\overline{F}$.

The approximation of the df of the cumulative sum by a stable limit distribution and its refinements is not always optimal. There exist powerful direct estimates for these probabilities assuming conditions on the tails, the moments or the bounds of the support of these rvs; see for instance Petrov [495, 496], Shorack and Wellner [579].

Asymptotic Expansions

As mentioned above, the Berry–Esséen estimate (2.23) is optimal for certain dfs F. However, in some cases one can approximate the df G_n by the standard normal df Φ and some additional terms. The approximating function is then *not* a df. A common approximation method is called *Edgeworth* or *asymptotic expansion*: formally we write

$$G_n(x) = \Phi(x) + \sum_{k=1}^{\infty} n^{-k/2} Q_k(x), \quad x \in \mathbb{R}, \tag{2.24}$$

where the Q_k are expressions involving the Hermite polynomials, the precise form of the expression depending on the moments of X. The expansion (2.24) is derived from a formal Taylor expansion of the logarithm of the corresponding chf. To the latter is then applied a Fourier inversion. This approach does not depend on the specific form of the df G_n and is applicable to much wider classes of distributions, but here we restrict ourselves to G_n for illustrational purposes.

In practice one can take only a finite number of terms Q_k into account. To get an impression we consider the first two terms: let

$$\varphi(x) = (2\pi)^{-1/2} \exp\left\{-x^2/2\right\}, \quad x \in \mathbb{R},$$

denote the density function of the standard normal df Φ. Then, for $x \in \mathbb{R}$,

$$\begin{aligned} Q_1(x) &= -\varphi(x) \frac{H_2(x)}{6} \frac{E(X-\mu)^3}{\sigma^3}, \\ Q_2(x) &= -\varphi(x) \left\{ \frac{H_5(x)}{72} \left(\frac{E(X-\mu)^3}{\sigma^3} \right)^2 + \frac{H_3(x)}{24} \left(\frac{E(X-\mu)^4}{\sigma^4} - 3 \right) \right\}, \end{aligned} \tag{2.25}$$

where H_i denotes the Hermite polynomial of degree i:

$$\begin{aligned} H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_5(x) &= x^5 - 10x^3 + 15x. \end{aligned}$$

Notice that the Q_k in (2.25) vanish if X is Gaussian, and the quantities

$$E(X-\mu)^3/\sigma^3, \qquad E(X-\mu)^4/\sigma^4$$

are the *skewness* and *kurtosis* of X, respectively. They measure the "closeness" of the df F to Φ.

If we want to expand G_n with an asymptotically negligible remainder term special conditions on the df F must be satisfied. For example, F must be absolutely continuous or distributed on a lattice. We provide here just one example to illustrate the power of the method.

Theorem 2.3.2 (Asymptotic expansion in the absolutely continuous case)
Suppose that $E|X|^k < \infty$ for some integer $k \geq 3$. If F is absolutely continuous then

$$(1+|x|)^k \left| G_n(x) - \Phi(x) - \sum_{i=1}^{k-2} \frac{Q_i(x)}{n^{i/2}} \right| = o\left(\frac{1}{n^{(k-2)/2}}\right),$$

uniformly in x. In particular,

$$G_n(x) = \Phi(x) + \sum_{i=1}^{k-2} \frac{Q_i(x)}{n^{i/2}} + o\left(\frac{1}{n^{(k-2)/2}}\right),$$

uniformly in x. □

Asymptotic expansions can also be applied to the derivatives of G_n. In particular, if F is absolutely continuous then one can obtain asymptotic expansions for the density of G_n.

Notes and Comments

Results on asymptotic expansions for the iid and non–iid case can be found in Hall [311] or in Petrov [495, 496]. Asymptotic expansions for an arbitrary df have been treated in Field and Ronchetti [239] and Jensen [356]. In Christoph and Wolf [119], Ibragimov and Linnik [350] and Zolotarev [645] one can find some ideas about the construction of asymptotic expansions in the α–stable case.

Large Deviations

The CLT can be further refined if one starts looking at G_n for x taken from certain regions (depending on n) or if $x = x_n \to \infty$ at a given rate. This is the objective of the so–called *large deviation* techniques. Nowadays the theory of large deviations has been developed quite rapidly in different directions with applications in mathematics, statistics, engineering and physics. We will restrict ourselves to large deviations in the classical framework of Cramér [139].

Theorem 2.3.3 (Cramér's theorem on large deviations)
Suppose that the moment generating function $M(h) = E\exp\{hX\}$ exists in a neighbourhood of the origin. Then

$$\begin{aligned}
\frac{1-G_n(x)}{1-\Phi(x)} &= \exp\left\{\frac{x^3}{\sqrt{n}}\lambda\left(\frac{x}{n}\right)\right\}\left[1+O\left(\frac{x+1}{\sqrt{n}}\varphi(x)\right)\right], \\
\frac{G_n(-x)}{\Phi(-x)} &= \exp\left\{\frac{-x^3}{\sqrt{n}}\lambda\left(\frac{-x}{n}\right)\right\}\left[1+O\left(\frac{x+1}{\sqrt{n}}\varphi(x)\right)\right],
\end{aligned}$$

uniformly for positive $x = o(\sqrt{n})$. *Here* $\lambda(z)$ *is a power series which converges in a certain neighbourhood of the origin and whose coefficients depend only on the moments of* X. □

The power series $\lambda(z)$ is called *Cramér's series.* Instead of determining the general coefficients of this series we consider a particular case of Theorem 2.3.3.

Corollary 2.3.4 *Suppose that the conditions of Theorem 2.3.3 are satisfied. Then*

$$\begin{aligned} 1 - G_n(x) &= (1 - \Phi(x)) \exp\left\{ \frac{x^3}{6\sqrt{n}} \frac{E(X-\mu)^3}{\sigma^3} \right\} + O\left(\frac{1}{\sqrt{n}} \varphi(x) \right), \\ G_n(-x) &= \Phi(-x) \exp\left\{ \frac{-x^3}{6\sqrt{n}} \frac{E(X-\mu)^3}{\sigma^3} \right\} + O\left(\frac{1}{\sqrt{n}} \varphi(x) \right), \end{aligned}$$

for $x \geq 0$, $x = O(n^{1/6})$. *In particular, if* $E(X-\mu)^3 = 0$ *then*

$$G_n(x) - \Phi(x) = O\left(\frac{1}{\sqrt{n}} \varphi(x) \right), \quad x \in \mathbb{R}.$$ □

Large deviation results can be interpreted as refinements of the convergence rates in the CLT. Indeed, let $x = x_n \to \infty$ in such a way that $x_n = o(n^{1/6})$. Then we conclude from Corollary 2.3.4 that

$$P\left(\left| \frac{S_n - \mu\, n}{\sigma\sqrt{n}} \right| > x_n \right) = 2\left((1 - \Phi(x_n)) + O\left(\frac{1}{\sqrt{n}} \exp\left\{ -\frac{x_n^2}{2} \right\} \right) \right).$$

Note that the x_n are chosen such that

$$\frac{S_n - \mu\, n}{\sigma\sqrt{n}\, x_n} \xrightarrow{P} 0. \tag{2.26}$$

In an analogous way we can also consider large deviation results for $X \in \mathrm{DA}(\alpha)$, $\alpha < 2$. These must be of a completely different nature since the moment generating function $M(h)$ does not exist in any neighbourhood of the origin. However, one can get an impression of the order of decrease for the tail probabilities of S_n. For simplicity we restrict ourselves to symmetric rvs.

Theorem 2.3.5 (Heyde's theorem on large deviations)
Let $X \in \mathrm{DA}(\alpha)$ *be symmetric and* $\alpha \in (0,2)$. *Let* (b_n) *be any sequence such that* $b_n \uparrow \infty$ *and* $P(X > b_n) \sim 1/n$, *and denote by*

$$M_1 = X_1, \quad M_n = \max(X_1, \ldots, X_n), \quad n \geq 2,$$

the sample maxima. Then

$$\lim_{n\to\infty} \frac{P(S_n > b_n x_n)}{nP(X > b_n x_n)} = \lim_{n\to\infty} \frac{P(S_n > b_n x_n)}{P(M_n > b_n x_n)} = 1$$

for every sequence $x_n \to \infty$. □

In view of Theorem 2.2.15, the conditions of Theorem 2.3.5 ensure that $b_n^{-1} S_n \xrightarrow{d} G_\alpha$ for an α–stable law G_α. Thus the relation

$$\frac{S_n}{b_n x_n} \xrightarrow{P} 0$$

is directly comparable with (2.26). Similar results can be established for rvs with regularly varying tails $1 - F(x) \sim x^{-\alpha} L(x)$, where $\alpha \geq 2$, as $x \to \infty$; see Section 8.6. This kind of result is another example of the interplay between sums and maxima of a sample $X_1, \ldots, X_n$ as $n \to \infty$. Notice that Theorem 2.3.5 can be understood as a supplementary result to the limit relation

$$\lim_{x\to\infty} \frac{P(S_n > x)}{P(M_n > x)} = 1\,, \quad n = 1, 2, \ldots\,,$$

which is a defining property of subexponentiality and is studied in detail in Section 1.3.2 and Appendix A3.

Notes and Comments

Cramér's theorem and other versions of large deviation results (including the non–iid case) can be found in Petrov [495, 496]. Theorem 2.3.5 is due to Heyde [321, 322].

The general theory of large deviations has become an important part of probability theory with applications in different fields, including insurance and finance. By now it has become a theory which can be applied to sequences of arbitrary rvs which do not necessarily have sum structure and which can satisfy very general dependence conditions; see for instance the monographs by Bucklew [96], Dembo and Zeitouni [177], Deuschel and Strook [178], or Ellis [200]. We also mention that large deviation results are closely related to *saddlepoint approximations* in statistics, for instance Barndorff–Nielsen and Cox [48], Field and Ronchetti [239], Jensen [356]. The latter contains various applications to insurance risk theory. It should also be mentioned that, whereas Edgeworth expansions yield good approximations around the mean, they become unreliable in the tails. Saddlepoint approximations remedy this problem.

We give some more specific results on large deviations in Section 8.6. They find immediate applications for the valuation of certain quantities which are closely related to reinsurance problems; see Section 8.7.

2.4 The Functional CLT: Brownian Motion Appears

Let (X_n) be an iid sequence with $0 < \sigma^2 < \infty$. In this section we embed the sequence of partial sums (S_n) in a process on $[0,1]$ and consider the limit process which turns out to be Brownian motion. First consider the process $S_n(\cdot)$ on $[0,1]$ such that

$$S_n\left(n^{-1}k\right) = \frac{1}{\sigma\sqrt{n}}\,(S_k - \mu\,k)\,, \quad k = 0,\ldots,n\,,$$

and define the graph of the process $S_n(\cdot)$ at every point of $[0,1]$ by linear interpolation between the points $(k/n\,, S_n(k/n))$. This graph is just a "broken line" and the sample paths are continuous (polygonal) functions. Suppose for the moment that (X_n) is a sequence of iid standard normal rvs. Then the increments $S_n(k/n) - S_n(l/n)$ for $l < k$ are Gaussian with mean zero and variance $(k-l)/n$. Moreover, the process has independent increments when restricted to the points $(k/n)_{k=0,\ldots,n}$. These properties remind us of one of the most important processes which is used in probability theory, Brownian motion, the definition of which follows:

Definition 2.4.1 (Brownian motion)
Let $(B_t)_{t\in[0,1]}$ be a stochastic process which satisfies the following conditions:

(a) *It starts at zero:* $B_0 = 0$ a.s.
(b) *It has independent increments: for any partition* $0 \le t_0 < t_1 < \cdots < t_m \le 1$ *and any* m *the rvs* $B_{t_1} - B_{t_0}, \ldots, B_{t_m} - B_{t_{m-1}}$ *are independent.*
(c) *For every* $t \in [0,1]$, B_t *has a Gaussian distribution with mean zero and variance* t.
(d) *The sample paths are continuous with probability* 1.

This process is called (standard) Brownian motion *or* Wiener process on $[0,1]$. □

A consequence of this definition is that the increments $B_t - B_s$, $t > s$, have a $N(0, t-s)$ distribution. Brownian motion on $[0,T]$ and on $[0,\infty)$ is defined in a straightforward way by suitably modifying Definition 2.4.1. We mention that one can give a "minimal" definition of Brownian motion as a process with stationary, independent increments and a.s. continuous sample paths. It can be shown that from these properties alone it follows that the increments *must be* normally distributed.

We write $\mathbb{C}\,[0,1]$ for the vector space of continuous functions which is equipped with the supremum norm; see Appendix A2.2: for $x \in \mathbb{C}\,[0,1]$

$$\|x\| = \sup_{0\le t\le 1} |x(t)|\,.$$

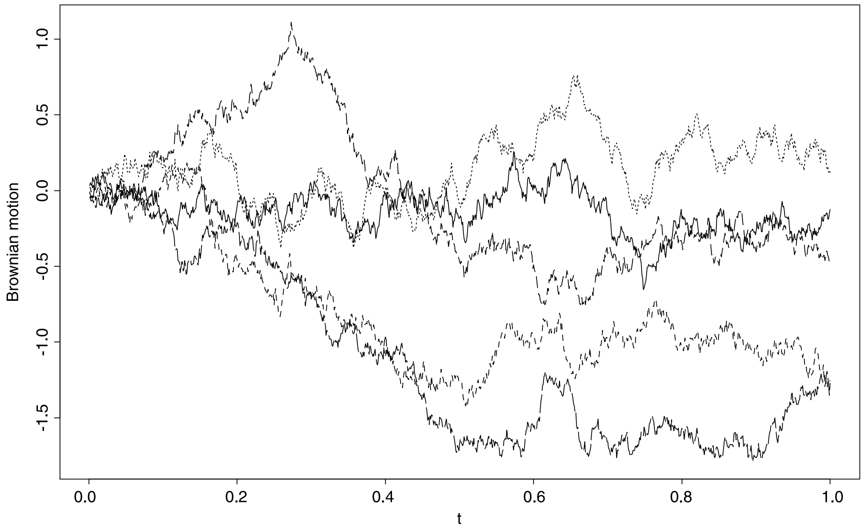

Figure 2.4.2 *Visualisation of Brownian motion: five sample paths of standard Brownian motion on* $[0,1]$.

Notice that the processes $S_n(\cdot)$ and $B.$ assume values in $\mathbb{C}\,[0,1]$.

We introduce still another process on $[0,1]$ which coincides with $S_n(\cdot)$ at the points k/n, $k = 0, 1, \ldots, n$. It is easier to construct but more difficult to deal with theoretically:

$$\widetilde{S}_n(t) = \frac{1}{\sigma\sqrt{n}} \left(S_{[nt]} - \mu\,[nt]\right), \quad 0 \le t \le 1,$$

where $[y]$ denotes the integer part of the real number y. This process has independent increments which are Gaussian (possibly degenerate) if X is Gaussian. Its sample paths are not continuous but have possible jumps at the points k/n. At each point of $[0,1)$ they are continuous from the right and at each point of $(0,1]$ the limit from the left exists. Thus the process $\widetilde{S}_n(\cdot)$ has *cadlag* sample paths, see Appendix A2.2, i.e. they belong to the space $\mathbb{D}\,[0,1]$. The space $\mathbb{D}\,[0,1]$ of cadlag functions can be equipped with different metrics in order to define weak convergence on it. However, our limit process will be Brownian motion which assumes values in $\mathbb{C}\,[0,1]$ so that we are allowed to take the sup–norm as an appropriate metric in $\mathbb{D}\,[0,1]$; see Theorem A2.5.

The following result is known as the *Donsker invariance principle* or *functional CLT* (*FCLT*). In this context it is worthwhile to recall the continuous mapping theorem (Theorem A2.6) and the notion of weak convergence in the

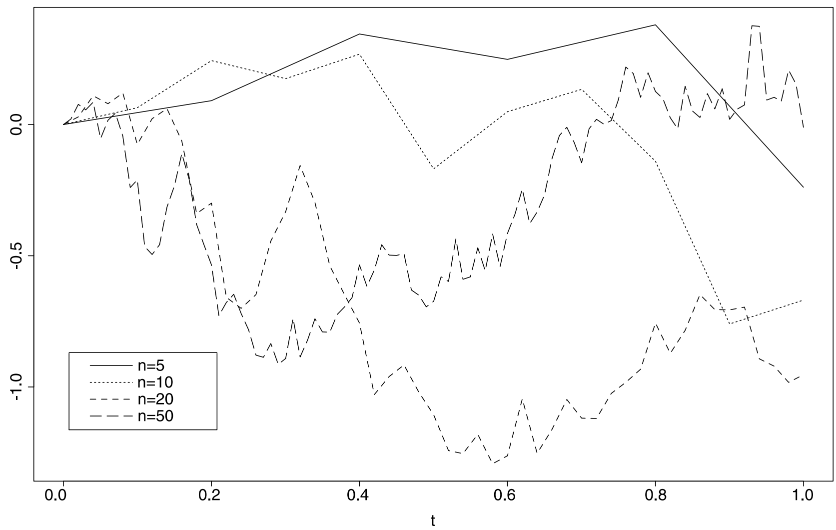

Figure 2.4.3 *Visualisation of the Donsker invariance principle: sample paths of the process $S_n(\cdot)$ for the same realisation of (X_n) for $n = 5, 10, 20, 50$.*

function spaces $\mathbb{C}\,[0,1]$ and in $\mathbb{D}\,[0,1]$ from Appendix A2.4. To those of you not familiar with this abstract terminology we simply recommend a glance at Figure 2.4.3 which explains how a sample path of Brownian motion is built up from sums of iid rvs.

Theorem 2.4.4 (FCLT, Donsker invariance principle)
Suppose that $EX^2 < \infty$. Then

(a) $S_n(\cdot) \xrightarrow{d} B.$ *in* $\mathbb{C}\,[0,1]$ (*equipped with the sup–norm and the σ–algebra generated by the open subsets*),

(b) $\widetilde{S}_n(\cdot) \xrightarrow{d} B.$ *in* $\mathbb{D}\,[0,1]$ (*equipped with the sup–norm and the σ–algebra generated by the open balls*).

In particular, if f_1 (f_2) is continuous except possibly on a subset $A \subset \mathbb{C}\,[0,1]$ ($A \subset \mathbb{D}\,[0,1]$) for which $P(B. \in A) = 0$, then $f_1(S_n(\cdot)) \xrightarrow{d} f_1(B.)$ and $f_2(\widetilde{S}_n(\cdot)) \xrightarrow{d} f_2(B.)$. □

Remarks. 1) The Donsker invariance principle is a very powerful result. It explains why Brownian motion can be taken as a reasonable approximation to many real processes which are in some way related to sums of independent rvs. In finance, the celebrated Black–Scholes model is based on geometric

Brownian motion $X_t = \exp\{ct + \sigma B_t\}$ for constants c and $\sigma > 0$. The rationale of such an approach is that the logarithmic price $\ln X_t$ of a risky asset can be understood as the result of actions and interactions caused by a large number of different independent activities in economy and politics, or indeed individual traders, i.e. it can be understood as a sum process. And indeed, geometric Brownian motion can be viewed as a weak limit of the binomial pricing model of Cox–Ross–Rubinstein; see for instance Lamberton and Lapeyre [412]. In physics a sample path of Brownian motion is often interpreted as the movement of a small particle which is pushed by small independent forces from different directions. Here again the interpretation as a sum process is applicable. As a limit process of normalised and centred random walks, we can consider Brownian motion as a random walk in continuous time.

2) The Donsker invariance principle suggests an easy way of simulating Brownian sample paths by the approximating processes $S_n(\cdot)$ or $\widetilde{S}_n(\cdot)$. They can easily be simulated, for example, if (X_n) is iid Gaussian noise or if (X_n) is a sequence of iid Bernoulli rvs assuming the two values $+1$ and -1 with equal probability. Again, back to the finance world: Donsker explains how to generate from one fair coin the basic process underlying modern mathematical finance. $\square$

The power of a functional limit theorem is considerably increased by the continuous mapping theorem (Theorem A2.6):

Example 2.4.5 (Donsker and continuous mapping theorem)
We may conclude from Theorem 2.4.4 that the finite–dimensional distributions of the processes $S_n(\cdot)$ and $\widetilde{S}_n(\cdot)$ converge. Indeed, consider the mapping $f : \mathbb{D}[0,1] \to \mathbb{R}^m$ defined by

$$f(x) = (x_{t_1}, \ldots, x_{t_m})$$

for any $0 \le t_1 < \cdots < t_m \le 1$. It is continuous at elements $x \in \mathbb{C}[0,1]$. Then

$$\begin{aligned} f(S_n(\cdot)) &= (S_n(t_1), \ldots, S_n(t_m)) \xrightarrow{d} f(B_\cdot) = (B_{t_1}, \ldots, B_{t_m}), \\ f(\widetilde{S}_n(\cdot)) &= (\widetilde{S}_n(t_1), \ldots, \widetilde{S}_n(t_m)) \xrightarrow{d} f(B_\cdot) = (B_{t_1}, \ldots, B_{t_m}). \end{aligned}$$

Hence weak convergence of the processes $S_n(\cdot)$ and $\widetilde{S}_n(\cdot)$ implies convergence of the finite–dimensional distributions.

Moreover, the following functionals are continuous on both spaces $\mathbb{C}[0,1]$ and $\mathbb{D}[0,1]$ when endowed with the sup–norm:

$$f_1(x) = x(1), \quad f_2(x) = \sup_{0 \le t \le 1} x(t), \quad f_3(x) = \inf_{0 \le t \le 1} x(t).$$

In particular,

$$\begin{aligned} f_1(S_n(\cdot)) &= f_1(\widetilde{S}_n(\cdot)) &= \frac{1}{\sigma\sqrt{n}}\,(S_n - n\,\mu)\,, \\ f_2(S_n(\cdot)) &= f_2(\widetilde{S}_n(\cdot)) &= \frac{1}{\sigma\sqrt{n}}\,\max_{0\le k\le n}\,(S_k - k\,\mu)\,, \\ f_3(S_n(\cdot)) &= f_3(\widetilde{S}_n(\cdot)) &= \frac{1}{\sigma\sqrt{n}}\,\min_{0\le k\le n}\,(S_k - k\,\mu)\,. \end{aligned}$$

Moreover, the multivariate function (f_1, f_2, f_3) is continuous on both spaces $\mathbb{C}\,[0,1]$ and $\mathbb{D}\,[0,1]$. From Theorem 2.4.4 and the continuous mapping theorem we immediately obtain

$$\frac{1}{\sigma\sqrt{n}}\left(S_n - n\mu\,,\,\max_{0\le k\le n}\,(S_k - k\,\mu)\,,\,\min_{0\le k\le n}\,(S_k - k\,\mu)\right)$$
$$\stackrel{d}{\to}\left(B_1\,,\,\max_{0\le t\le 1}\,B_t\,,\,\min_{0\le t\le 1}\,B_t\right)\,.$$

The joint distribution of B_1, the minimum and maximum of Brownian motion on $[0,1]$ is well known. A derivation is given in Billingsley [69], Chapter 2.11.

At this point it is still worth stressing that, whereas Donsker in conjunction with the continuous mapping theorem offers indeed a very powerful tool, in many applications actually proving that certain functionals on either $\mathbb{C}$ or $\mathbb{D}$ are continuous may be the hard part. Also, once we have a weak convergence result, we may want to use it in two ways. First, in some cases we may derive distributional properties of the limit process through known properties of the approximating process; the latter can for instance be taken to be based on iid Bernoulli rvs. For several examples see Billingsley [69]. However, we may also use the limit process as a useful approximation of a less tangible underlying process; a typical example will be discussed in the diffusion approximation for risk processes, see Example 2.5.18. □

As already stated, Brownian motion is a particular *process with independent, stationary increments*:

Definition 2.4.6 (Process with independent, stationary increments)
Let $\xi = (\xi_t)_{0\le t\le 1}$ *be a stochastic process. Then* ξ *has* independent increments *if for any* $0 \le t_0 < \cdots < t_m \le 1$ *and any* $m \ge 1$ *the rvs*

$$\xi_{t_1} - \xi_{t_0}, \ldots, \xi_{t_m} - \xi_{t_{m-1}}\,,$$

are independent. This process is said to have stationary increments *if for any* $0 \le s < t \le 1$ *the rvs* $\xi_t - \xi_s$ *and* ξ_{t-s} *have the same distribution.*

A process with independent, stationary increments and sample paths in $\mathbb{D}\,[0,1]$ *is also called a* Lévy process. □

By a straightforward modification of this definition we can also define processes with independent, stationary increments on $[0,T]$ or on $[0,\infty)$.

We introduce another class of stochastic processes which contains Brownian motion as a special case. Recall from Section 2.2 the definition of an α–stable rv.

Definition 2.4.7 (α–stable motion)
A stochastic process $(\xi_t)_{0\le t\le 1}$ *with sample paths in* $\mathbb{D}\,[0,1]$ *is said to be* α–stable motion *if the following properties hold:*

(a) *It starts at zero:* $\xi_0 = 0$ *a.s.*
(b) *It has independent, stationary increments.*
(c) *For every* $t \in [0,1]$, ξ_t *has an* α*–stable distribution with fixed parameters* $\beta \in [-1,1]$ *and* $\gamma = 0$ *in the spectral representation* (2.10). □

It is straightforward that we can extend this definition to processes on $[0,T]$ or on $[0,\infty)$. We see that Brownian motion (cf. Definition 2.4.1) is just a 2–stable motion. For simplicity, α–stable motions are often called *stable processes* although this might be confusing since in the specialised literature more general stable processes (with dependent or non–stationary stable increments) occur. In Section 8.8 we give an introduction to the world of multivariate stable random vectors and of stable processes.

We need the following elementary relation:

Lemma 2.4.8 *For an* α*–stable motion* $(\xi_t)_{0\le t\le 1}$ *we have*

$$\xi_t - \xi_s \stackrel{d}{=} (t-s)^{1/\alpha}\xi_1\,, \quad 0 \le s < t \le 1\,.$$

Proof. Using the spectral representation (2.10) and the stationary, independent α–stable increments we conclude that

$$\begin{aligned} E\exp\{i\lambda\xi_t\} &= \exp\{-c_t|\lambda|^\alpha\,(1 - i\beta\,\mathrm{sign}(\lambda)z(\lambda,\alpha))\} \\ &= E\exp\{i\lambda\xi_s\}\,E\exp\{i\lambda\,(\xi_t - \xi_s)\} \\ &= E\exp\{i\lambda\xi_s\}\,E\exp\{i\lambda\xi_{t-s}\} \\ &= \exp\{-\,(c_s + c_{t-s})\,|\lambda|^\alpha\,(1 - i\beta\,\mathrm{sign}(\lambda)z(\lambda,\alpha))\}\;, \end{aligned}$$

for $\lambda \in \mathbb{R}$ and positive constants c_s, c_t and c_{t-s} which satisfy the relation

$$c_s + c_{t-s} = c_t\,, \quad c_0 = 0\,, \quad 0 \le s < t \le 1\,.$$

The well known measurable solution to this Cauchy functional equation is $c_s = cs$ for a constant c (see Bingham, Goldie and Teugels [72], Theorem 1.1.7), and c must be positive because of the properties of chfs. This proves the lemma. □

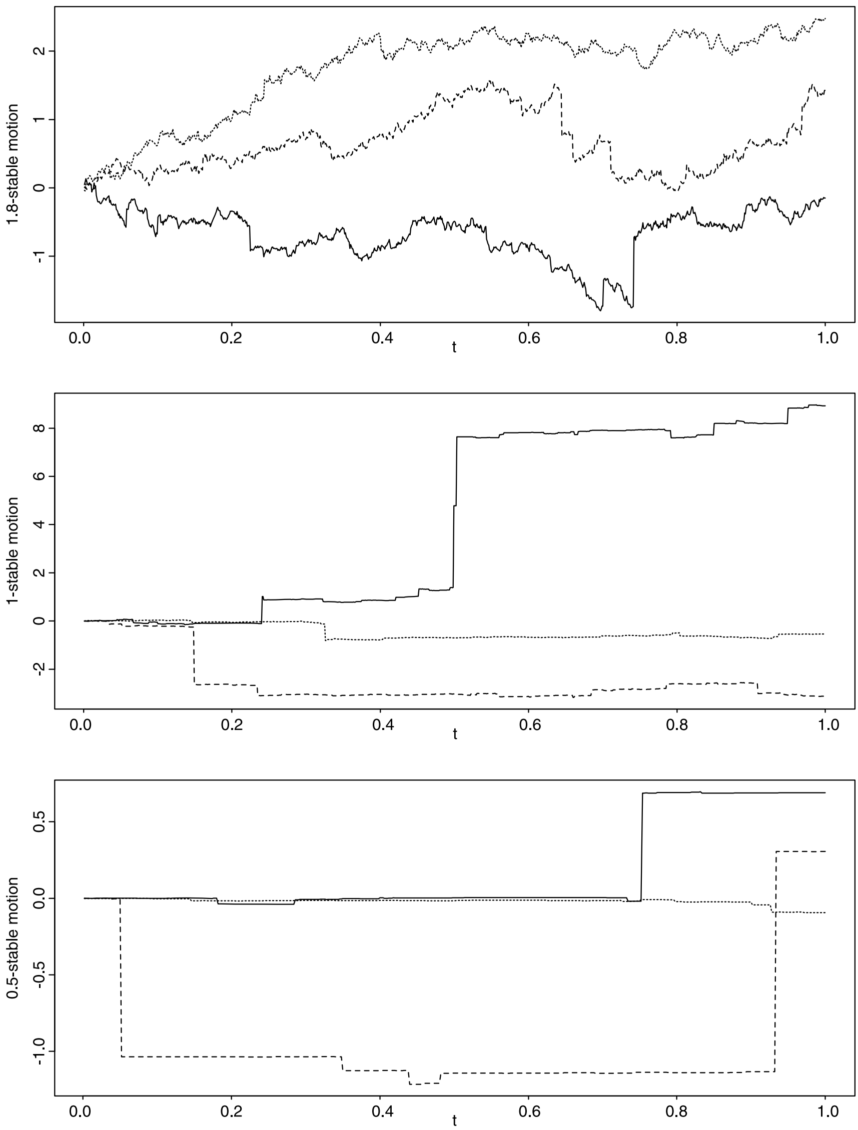

Figure 2.4.9 *Visualisation of symmetric* $1.8-$, $1-$ *and* $0.5-$*stable motion* (*top, middle and bottom*)*: three sample paths of* (ξ_t) *on* $[0,1]$. *The lower two graphs suggest that the sample paths are piecewise constant. This is by no means the case; the set of jumps of almost every sample path is a dense set in* $[0,1]$. *However, the jump heights are in general so tiny that we cannot see them, we only see the large ones.*

From Lemma 2.4.8 we can easily derive the finite–dimensional distributions of an α–stable motion:

$$\begin{aligned}
&(\xi_{t_1}, \xi_{t_2}, \ldots, \xi_{t_m}) \\
&= \left(\xi_{t_1}, \xi_{t_1} + (\xi_{t_2} - \xi_{t_1}), \ldots, \xi_{t_1} + (\xi_{t_2} - \xi_{t_1}) + \cdots + \left(\xi_{t_m} - \xi_{t_{m-1}}\right)\right) \\
&\stackrel{d}{=} \Big(t_1^{1/\alpha} Y_1, t_1^{1/\alpha} Y_1 + (t_2 - t_1)^{1/\alpha} Y_2, \ldots, \\
&\qquad t_1^{1/\alpha} Y_1 + (t_2 - t_1)^{1/\alpha} Y_2 + \cdots + (t_m - t_{m-1})^{1/\alpha} Y_m\Big)
\end{aligned}$$

for any real numbers $0 \le t_1 < \cdots < t_m \le 1$ and iid α–stable rvs $Y_1, \ldots, Y_m$ such that $Y_1 \stackrel{d}{=} \xi_1$.

Analogously to the Donsker invariance principle we might ask:

Can every α–stable motion be derived as the weak limit of an appropriate sum process?

The answer is YES as the following theorem shows. We refer to Section 2.2 for the definition of domains of attraction and to Appendix A2.4 for the notion of weak convergence of processes.

Theorem 2.4.10 (Stable FCLT)
Let (X_n) be iid rvs in the domain of attraction of an α–stable rv Z_α with parameter $\gamma = 0$ in (2.10). *Suppose that*

$$\left(n^{1/\alpha} L(n)\right)^{-1} (S_n - a_n) \stackrel{d}{\to} Z_\alpha, \quad n \to \infty,$$

for an appropriate slowly varying function L. Then the process

$$\left(n^{1/\alpha} L(n)\right)^{-1} \left(S_{[nt]} - a_{[nt]}\right), \quad 0 \le t \le 1,$$

converges weakly to an α–stable motion $(\xi_t)_{0 \le t \le 1}$, and $\xi_1 \stackrel{d}{=} Z_\alpha$. Here convergence is understood as weak convergence in $\mathbb{D}[0,1]$ equipped with the J_1–metric and the σ–algebra generated by the open sets. □

We know that Brownian motion has a.s. continuous sample paths. This is not the case for α–stable motions with $\alpha < 2$. Apart from a drift, their sample paths are pure jump processes, and all jumps occur at random instants of time. If we restrict the sample paths of ξ to the interval $[0,1]$ then ξ is a stochastic process which assumes values in $\mathbb{D}[0,1]$, i.e. these sample paths are cadlag. Again we can apply the continuous mapping theorem. For example, the results of Example 2.4.5 remain valid with Brownian motion replaced by a general α–stable motion as limit process.

Notes and Comments

Proofs of the Donsker invariance principle (Theorem 2.4.4) can be found in Billingsley [69] and Pollard [504]. Generalisations to martingales are given in Hall and Heyde [312] and to more general processes in Jacod and Shiryaev [352].

Monographs on Brownian motion and its properties are Hida [325], Karatzas and Shreve [368], Revuz and Yor [541]. An encyclopaedic compendium of facts and formulae for Brownian motion is Borodin and Salminen [82].

FCLTs are applied in insurance mathematics for determining the *probability of ruin* via the so–called *diffusion approximation*; see Grandell [282]. The idea is due to Iglehart [351]. We explain this method in Example 2.5.18.

Methods for simulating Brownian motion are given for instance in Janicki and Weron [354], Kloeden and Platen [385] and the companion book by Kloeden, Platen and Schurz [386].

Definitions of α–stable motion and more general α–stable processes can be found in the literature cited below. A proof of the FCLT in the form of Theorem 2.4.10 follows from the general theory of processes with independent increments; see for instance Gikhman and Skorokhod [262], Jacod and Shiryaev [352], Chapter VII, see also Resnick [529]. Stable motions and processes are treated in various books: Mijnheer [456] concentrates on a.s. properties of the sample paths of α–stable motions. Janicki and Weron [354] discuss various methods for simulating α–stable processes and consider applications. Samorodnitsky and Taqqu [565] give a general theory for α–stable processes including several representations of stable rvs, stable processes and stable integrals. They also develop a theory of stochastic integration with respect to α–stable processes. Kwapień and Woyczyński [411] consider the case of single and multiple stochastic integrals with respect to α–stable processes. Lévy processes are considered in Bertoin [63], Jacod and Shiryaev [352] and Sato [566].

In Section 8.8 we give an introduction to stable processes more general than stable motion.

2.5 Random Sums

2.5.1 General Randomly Indexed Sequences

Random (i.e. randomly indexed) sums are the bread and butter of insurance mathematics. The total claim amount of an insurance portfolio is classically modelled by random sums

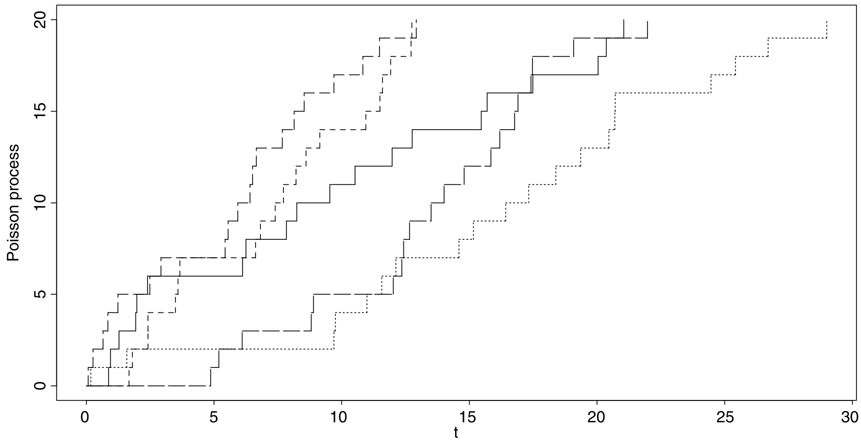

Figure 2.5.1 *Visualisation of a Poisson process with intensity* 1*: five sample paths of* $(N(t))$.

$$S(t) = S_{N(t)} = \begin{cases} 0 & \text{if} \quad N(t) = 0\,, \\ X_1 + \cdots + X_{N(t)} & \text{if} \quad N(t) \geq 1\,, \end{cases} \qquad t \geq 0\,,$$

where $(N(t))_{t \geq 0}$ is a stochastic process on $[0, \infty)$ such that the rvs $N(t)$ are non–negative integer–valued. Usually, $(N(t))$ is assumed to be generated by a sequence $(T_n)_{n \geq 1}$ of non–negative rvs such that

$$0 \leq T_1 \leq T_2 \leq \cdots \quad \text{a.s.}$$

and

$$N(t) = \sup \{n \geq 1 : T_n \leq t\}\ , \quad t \geq 0\,. \tag{2.27}$$

As usual, $\sup A = 0$ if $A = \emptyset$. This is then called a *counting process*. The rv X_n can be interpreted as an individual claim which arrives at the random time T_n, $N(t)$ counts the total number of individual claims and $S(t)$ is the total claim amount in the portfolio up to time t. In the context of finance, $N(t)$ could for instance represent the (random) number of position changes in a foreign exchange portfolio based on tick–by–tick (high frequency) observations. The quantity $S(t)$ then represents the total return over $[0, t]$.

Example 2.5.2 (Homogeneous Poisson process and compound Poisson process)
In the Cramér–Lundberg model (Definition 1.1.1) it is assumed that (X_n) and $(N(t))$ are independent and that $(N(t))$ is a homogeneous *Poisson process with intensity parameter* $\lambda > 0$, i.e. it is a counting process (2.27) with

$$T_n = Y_1 + \cdots + Y_n \,, \quad n \geq 1 \,,$$

and (Y_n) (the inter–arrival times of the claims) are iid exponential rvs with expectation $1/\lambda$. Any counting process which is generated by an iid sum process (T_n) is also called a *renewal counting process.*
Alternatively, a (homogeneous) *Poisson process* is defined by the following three properties:

(a) It starts at zero: $N(0) = 0$.
(b) It has independent, stationary increments.
(c) For every $t > 0$, $N(t)$ is a Poisson rv with parameter $\lambda\, t$:

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!}\, e^{-\lambda\, t} \,, \quad n = 0, 1, 2, \ldots \,.$$

The Poisson process $(N(t))$ is a pure jump process with sample paths in $\mathbb{D}\,[0,\infty)$ which increase to ∞ as $t \to \infty$ and have jumps of height 1 at the random times T_n. It is also a Lévy process; see Definition 2.4.6. A homogeneous Poisson process can be interpreted as a special point process; see Section 5.1.3.
If $(N(t))$ and (X_n) are independent then the process $(S(t))_{t\geq 0}$ is called a *compound Poisson process.*
The Poisson process and Brownian motion and their modifications and generalisations are the most important stochastic processes in probability theory and mathematical statistics. □

The fluctuations of the random sums $S(t)$ for large t can again be described via limit theorems. In what follows we provide some basic tools which show that the asymptotic behaviour of (S_n) and $(S(t))$ is closely linked.

In this section, $(Z_n)_{n\geq 0}$ is a general sequence of rvs and $(N(t))_{t\geq 0}$ is a process of non–negative integer–valued rvs $N(t)$.

Lemma 2.5.3 *Suppose that* $Z_n \stackrel{\text{a.s.}}{\to} Z$ *as* $n \to \infty$ *and* $N(t) \stackrel{\text{a.s.}}{\to} \infty$ ($N(t) \stackrel{P}{\to} \infty$) *as* $t \to \infty$. *Then*

$$Z_{N(t)} \stackrel{\text{a.s.}}{\to} Z \,, \qquad \left(Z_{N(t)} \stackrel{P}{\to} Z \right) \,, \quad t \to \infty \,.$$

Proof. Suppose $N(t) \stackrel{\text{a.s.}}{\to} \infty$. Set

$$A_1 = \{\omega : N(t)(\omega) \to \infty \,, \quad t \to \infty\} \,, \quad A_2 = \{\omega : Z_n(\omega) \to Z(\omega) \,, \; n \to \infty\} \,,$$

and note that $P(A_1) = P(A_2) = 1$. Then

$$P\left(\left\{\omega : Z_{N(t)(\omega)}(\omega) \to Z(\omega) \,, \quad t \to \infty\right\}\right) \geq P\left(A_1 \cap A_2\right) = 1 \,,$$

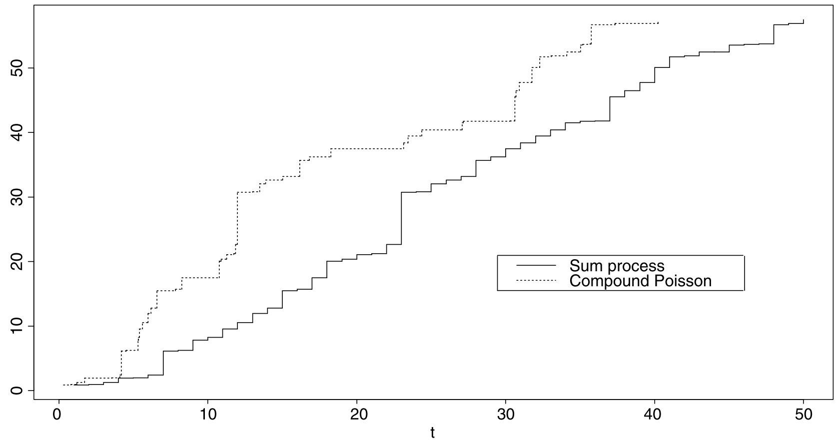

Figure 2.5.4 *One sample path of* (S_n) *and one sample path of the compound Poisson process* $(S(t))$ $(\lambda = 1)$ *for the same realisation of iid standard exponential* X_n.

i.e. $Z_{N(t)} \stackrel{\text{a.s.}}{\to} Z$.

Now suppose that $N(t) \stackrel{P}{\to} \infty$ as $t \to \infty$. For every sequence $t_k \to \infty$, $N(t_k) \stackrel{P}{\to} \infty$ as $k \to \infty$, and there exists a subsequence $t_{k_j} \uparrow \infty$ such that $N(t_{k_j}) \stackrel{\text{a.s.}}{\to} \infty$ as $j \to \infty$; see Appendix A1.3. From the first part of the proof, $Z_{N(t_{k_j})} \stackrel{\text{a.s.}}{\to} Z$, hence $Z_{N(t_{k_j})} \stackrel{P}{\to} Z$. Thus every sequence $(Z_{N(t_k)})$ contains a subsequence which converges in probability to Z. Since convergence in probability is metrizable (see Appendix A1.2) this means that $Z_{N(t)} \stackrel{P}{\to} Z$.

□

Combining Theorem 2.1.5 and Lemma 2.5.3 we immediately obtain

Theorem 2.5.5 (Marcinkiewicz–Zygmund SLLNs for random sums)
Suppose that $E|X|^p < \infty$ *for some* $p \in (0,2)$ *and* $N(t) \stackrel{\text{a.s.}}{\to} \infty$. *Then*

$$(N(t))^{-1/p} \ (S(t) - aN(t)) \stackrel{\text{a.s.}}{\to} 0\,, \tag{2.28}$$

where

$$a = \begin{cases} 0 & \textit{if} \quad p < 1\,, \\ \mu = EX & \textit{if} \quad p \in [1,2)\,. \end{cases}$$

□

We will see in Section 2.5.3 that, if we restrict ourselves to renewal counting processes $(N(t))$, we can replace the random normalising and centring processes in (2.28) by deterministic functions. Moreover, (2.28) can be extended to a LIL.

Now we turn to the case of weak convergence. In particular, we wish to derive the CLT for random sums. The following lemma covers many cases of practical interest, for example, the compound Poisson case as considered in Example 2.5.2.

Lemma 2.5.6 *Suppose that (Z_n) and $(N(t))$ are independent and $N(t) \stackrel{P}{\to} \infty$ as $t \to \infty$. If $Z_n \stackrel{d}{\to} Z$ as $n \to \infty$ then $Z_{N(t)} \stackrel{d}{\to} Z$ as $t \to \infty$.*

Proof. Write $\phi_A(s) = E \exp\{isA\}$ for the chf of any rv A and $f_n(s) = \phi_{Z_n}(s)$ for real s. By independence,

$$E\left(\exp\left\{isZ_{N(t)}\right\}\middle|\, N(t)\right) = f_{N(t)}(s) \quad \text{a.s.}$$

Note that $f_n(s) \to \phi_Z(s)$ as $n \to \infty$ and $N(t) \stackrel{P}{\to} \infty$. By Lemma 2.5.3,

$$f_{N(t)}(s) \stackrel{P}{\to} \phi_Z(s)\,, \quad t \to \infty\,,$$

and since $(f_{N(t)})$ is uniformly integrable,

$$E\exp\left\{isZ_{N(t)}\right\} = E_{N(t)}\left(f_{N(t)}(s)\right) \to E\left(\phi_Z(s)\right) = \phi_Z(s)\,, \quad s \in \mathbb{R}\,.$$

This proves that $Z_{N(t)} \stackrel{d}{\to} Z$. □

As an immediate consequence we derive an analogue of Theorem 2.2.15 for random sums:

Theorem 2.5.7 (CLT for random sums)
Suppose that (X_n) and $(N(t))$ are independent and that $N(t) \stackrel{P}{\to} \infty$. Assume that $F \in \mathrm{DA}(\alpha)$ for some $\alpha \in (0,2]$. Then Theorem 2.2.15 remains valid if n is everywhere replaced by $N(t)$, i.e. there exist appropriate centring constants a_n and a slowly varying function L such that

$$\left((N(t))^{1/\alpha} L(N(t))\right)^{-1} \left(S(t) - a_{N(t)}\right) \stackrel{d}{\to} G_\alpha\,, \quad t \to \infty\,, \tag{2.29}$$

for an α–stable distribution G_α. □

In Section 2.5.3 we will specify conditions which ensure that the random normalising and centring processes in (2.29) can be replaced by deterministic functions.

The condition that the processes (Z_n) and $(N(t))$ are independent can be relaxed substantially:

Lemma 2.5.8 (Anscombe's theorem)
Suppose that there exists an integer-valued function $b(t) \uparrow \infty$ *such that*

$$\frac{N(t)}{b(t)} \xrightarrow{P} 1\,, \quad t \to \infty\,, \tag{2.30}$$

and that the following, so–called Anscombe condition *holds:*

$$\forall \epsilon > 0\, \forall \eta > 0\, \exists \delta > 0\, \exists n_0 \quad \textit{such that}$$

$$P\left(\max_{m:|m-n|<n\delta} |Z_m - Z_n| > \epsilon\right) < \eta\,, \quad n > n_0\,. \tag{2.31}$$

If $Z_n \xrightarrow{d} Z$ *as* $n \to \infty$ *then* $Z_{N(t)} \xrightarrow{d} Z$ *as* $t \to \infty$. □

Roughly speaking, condition (2.30) ensures that the random index $N(t)$ can be replaced by the deterministic function $b(t)$, and if we do so with $Z_{N(t)}$, i.e. if we replace $Z_{N(t)}$ by $Z_{b(t)}$, then (2.31) guarantees that the error $|Z_{N(t)} - Z_{b(t)}|$ is negligible. In other words, Anscombe's condition is a specific stochastic continuity property of the sequence (Z_n).

We note that (2.30) is satisfied for wide classes of renewal counting processes (see Section 2.5.3), including the homogeneous Poisson process. Moreover, (2.31) holds for the (properly normalised and centred) sums S_n. This is the content of the following result which is analogous to Theorems 2.2.15 and 2.5.7. The use of the Anscombe condition in the proof below is not obvious; it is hidden by Kolmogorov's inequality.

Theorem 2.5.9 (Anscombe–type CLT for random sums)
Suppose that

$$\frac{N(t)}{t} \xrightarrow{P} \lambda\,, \quad t \to \infty\,, \tag{2.32}$$

for some positive λ *and that* $F \in \mathrm{DA}(\alpha)$ *for some* $\alpha \in (0,2]$ *with*

$$\left(n^{1/\alpha} L(n)\right)^{-1} (S_n - \widetilde{a}\, n) \xrightarrow{d} G_\alpha\,, \tag{2.33}$$

for an α*–stable distribution* G_α *and a slowly varying function* L. *Here*

$$\widetilde{a} = \begin{cases} 0 & \textit{if} \quad \alpha \le 1\,, \\ \mu & \textit{if} \quad \alpha \in (1,2]\,. \end{cases}$$

Then

$$\left((N(t))^{1/\alpha} L(N(t))\right)^{-1} (S(t) - \widetilde{a} N(t)) \quad \xrightarrow{d} \quad G_\alpha\,, \tag{2.34}$$

$$\left((\lambda t)^{1/\alpha} L(t)\right)^{-1} (S(t) - \widetilde{a} N(t)) \quad \xrightarrow{d} \quad G_\alpha\,. \tag{2.35}$$

In particular, if $\sigma^2 < \infty$ *then*

$$\left(\lambda\sigma^2 t\right)^{-1/2}\ (S(t) - \mu N(t)) \xrightarrow{d} \Phi\,,$$

where Φ *is the standard normal distribution.*

In view of Theorem 2.2.15, (2.33) is only a restriction on the distribution of X in the case $\alpha = 1$. It is satisfied for instance for symmetric F. In Section 2.5.3 we will find conditions which ensure that the random centring process in Theorem 2.5.9 can be replaced by a deterministic function.

Sketch of the proof. We restrict ourselves to the case $\alpha = 2$ and $\sigma^2 = \mathrm{var}(X) < \infty$. Without loss of generality we may and do assume that $\sigma^2 = 1$ and $\mu = 0$. We write

$$\frac{S(t)}{(N(t))^{1/2}} = \left(\frac{S_{[\lambda t]}}{(\lambda t)^{1/2}} + \frac{S_{N(t)} - S_{[\lambda t]}}{(\lambda t)^{1/2}}\right)\left(\frac{\lambda t}{N(t)}\right)^{1/2}\,.$$

By (2.32), the term $(\lambda t/N(t))^{1/2}$ converges to 1 in probability and the classical CLT yields that

$$\frac{S_{[\lambda t]}}{(\lambda t)^{1/2}} \xrightarrow{d} \Phi\,.$$

By virtue of the continuous mapping theorem (Theorem A2.6) it suffices to show that

$$\frac{S_{N(t)} - S_{[\lambda t]}}{(\lambda t)^{1/2}} \xrightarrow{P} 0\,.$$

For every $\epsilon > 0$ and $\delta > 0$ we have that

$$\begin{aligned}
C_\epsilon &= \left\{\frac{\left|S_{N(t)} - S_{[\lambda t]}\right|}{(\lambda t)^{1/2}} > \epsilon\right\}\\
&\subset \left\{\left|\frac{N(t)}{t} - \lambda\right| > \delta\right\} \bigcup \left\{\left|\frac{N(t)}{t} - \lambda\right| \le \delta\,,\ \frac{\left|S_{N(t)} - S_{[\lambda t]}\right|}{(\lambda t)^{1/2}} > \epsilon\right\}\\
&= A_1 \cup A_2\,.
\end{aligned}$$

By (2.32), $P(A_1) \to 0$ as $t \to \infty$. By Kolmogorov's inequality,

$$\begin{aligned}
P\left(A_2\right) &\le P\left(\max_{|n/t-\lambda|\le\delta}\left|S_n - S_{[\lambda t]}\right| > \epsilon(\lambda t)^{1/2}\right)\\
&\le \frac{2}{\epsilon^2\lambda t}\,\mathrm{var}\left(S_{[(\lambda+\delta)t]} - S_{[\lambda t]}\right)\\
&\le c\,\frac{\delta}{\epsilon^2}
\end{aligned}$$

for a positive constant c. Thus letting first t tend to ∞ and then δ to 0 we see that $P(C_\epsilon) \to 0$ for every fixed $\epsilon > 0$. This proves (2.34) for $\alpha = 2$.
For $\alpha < 2$ the proof is analogous. Instead of Kolmogorov's inequality one can apply Skorokhod–Ottaviani–type inequalities; see for instance Petrov [496], Theorem 2.3. The details are technical in nature and therefore omitted.
The equivalence of (2.34) and (2.35) is a consequence of (2.32) and of the slow variation of L. □

2.5.2 Renewal Counting Processes

We consider a *renewal counting process* $(N(t))_{t\geq 0}$, i.e.

$$N(t) = \sup\{n \geq 1 : T_n \leq t\}\ , \quad t \geq 0\,, \tag{2.36}$$

and

$$T_n = Y_1 + \cdots + Y_n\,, \quad n \geq 1\,,$$

for iid non–negative (non–zero) rvs $Y, Y_1, Y_2, \ldots$. For applications of this kind of processes to risk theory see Chapter 1. The homogeneous Poisson process (see Example 2.5.2) is such a renewal counting process where Y is exponential with expectation $1/\lambda$.

In this section we answer the question:

What is the order of magnitude of $N(t)$ as $t \to \infty$?

Observe that

$$\{T_n \leq t\} = \{N(t) \geq n\}\ .$$

Kolmogorov's SLLN implies that $T_n \stackrel{\text{a.s.}}{\to} \infty$ and therefore $N(t) \stackrel{\text{a.s.}}{\to} \infty$. However, we can derive much more precise information:

Theorem 2.5.10 (Marcinkiewicz–Zygmund SLLNs/LIL for renewal counting processes)
Suppose that $EY = 1/\lambda \leq \infty$ (if $EY = \infty$ set $\lambda = 0$). Then

$$t^{-1} N(t) \stackrel{\text{a.s.}}{\to} \lambda\,. \tag{2.37}$$

If $EY^p < \infty$ for some $p \in (1,2)$ then

$$t^{-1/p}\ (N(t) - \lambda\, t) \stackrel{\text{a.s.}}{\to} 0\,. \tag{2.38}$$

If $\sigma_Y^2 = \mathrm{var}(Y) < \infty$ then

$$\begin{aligned}
&\limsup_{t\to\infty}\,(2t\ln\ln t)^{-1/2}\,(N(t) - \lambda\, t)\\
&= \ -\liminf_{t\to\infty}\,(2t\ln\ln t)^{-1/2}\,(N(t) - \lambda\, t)\\
&= \ \sigma_Y \lambda^{3/2} \quad \text{a.s.}
\end{aligned}$$

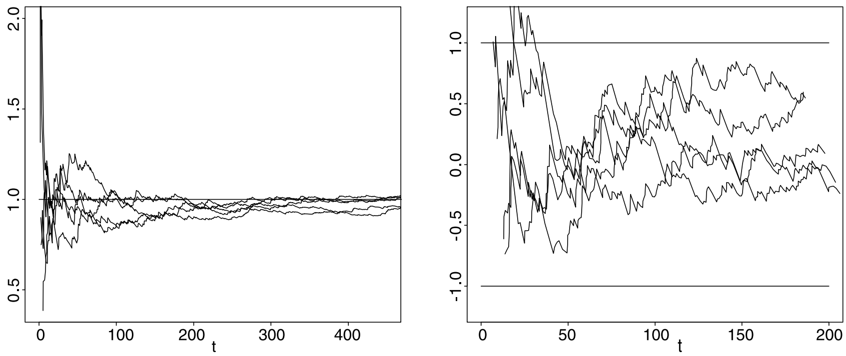

Figure 2.5.11 *Visualisation of the SLLN (left) and of the LIL (right) for the homogeneous Poisson process with intensity 1: five sample paths.*

Sketch of the proof. We restrict ourselves to show the SLLNs (2.37) and (2.38). Kolmogorov's SLLN for random sums yields that

$$\frac{T_{N(t)}}{N(t)} \stackrel{\text{a.s.}}{\to} \frac{1}{\lambda}\,.$$

By this and a sandwich argument applied to

$$\frac{T_{N(t)}}{N(t)} \le \frac{t}{N(t)} \le \frac{T_{N(t)+1}}{N(t)+1}\;\frac{N(t)+1}{N(t)}\;,$$

we prove (2.37).

Now suppose that $EY^p < \infty$ for some $p \in (1,2)$. Notice that

$$n^{-1/p}(T_{n+1} - T_n) = n^{-1/p} Y_{n+1} \stackrel{\text{a.s.}}{\to} 0\;.$$

This, the Marcinkiewicz–Zygmund SLLNs of Theorem 2.5.5, and (2.37) imply that

$$t^{-1/p}\;\left(T_{N(t)+1} - T_{N(t)}\right) = t^{-1/p} Y_{N(t)+1} \stackrel{\text{a.s.}}{\to} 0\,.$$

This, Theorem 2.5.5 and a sandwich argument applied to

$$\frac{\lambda t - N(t)}{t^{1/p}} \le \frac{\lambda T_{N(t)+1} - N(t)}{t^{1/p}} \le \frac{\lambda t - N(t)}{t^{1/p}} + \frac{\lambda Y_{N(t)+1}}{t^{1/p}} \qquad (2.39)$$

gives us (2.38). □

Theorem 2.5.10 suggests that $EN(t) \sim \lambda t$ and $\mathrm{var}(N(t)) \sim \sigma_Y^2 \lambda^3 t$. In the case of the Poisson process we even have that $EN(t) = \lambda t$ and $\mathrm{var}(N(t)) = \sigma_Y^2 \lambda^3 t$ since $EY = 1/\lambda$ and $\sigma_Y^2 = 1/\lambda^2$. For the following results see Gut [291], Theorems 5.1 and 5.2.

Proposition 2.5.12 (Moments of renewal counting process)
The following relations hold:

(a) $EN(t) = (\lambda + o(1))t$ *as* $t \to \infty$.
(b) *Suppose that* $\sigma_Y^2 = \mathrm{var}(Y) < \infty$. *Then*

$$\begin{aligned} EN(t) &= \lambda t + O(1)\,, \qquad & t \to \infty\,, \\ \mathrm{var}(N(t)) &= \sigma_Y^2 \lambda^3 t + o(t)\,, \qquad & t \to \infty\,. \end{aligned}$$

□

From Theorem 2.5.10 and Proposition 2.5.12 we have gained a first impression on the growth of a renewal counting process. Next we study the weak convergence of $(N(t))$.

Theorem 2.5.13 (CLT for renewal counting process)
Suppose that $\sigma_Y^2 < \infty$. *Then*

$$\left(\sigma_Y^2 \lambda^3 t\right)^{-1/2} \; (N(t) - \lambda\, t) \xrightarrow{d} \Phi\,, \tag{2.40}$$

where Φ *is the standard normal distribution.*

Recall from Proposition 2.5.12 that $EN(t) \sim \lambda\, t$ and $\mathrm{var}(N(t)) \sim \sigma_Y^2 \lambda^3 t$. Thus Theorem 2.5.13 is similar to the classical CLT for iid sums. We note that one can prove an analogous result for $(N(t))$ with an α–stable limit.

Proof. We proceed as in (2.39):

$$\frac{\lambda t - N(t)}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}} \le \frac{\lambda T_{N(t)+1} - N(t)}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}} \le \frac{\lambda t - N(t)}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}} + \frac{\lambda Y_{N(t)+1}}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}}\,.$$

We have, by independence,

$$\begin{aligned} P\left(Y_{N(t)+1} > \epsilon t^{1/2}\right) &= E\left(P\left(Y_{N(t)+1} > \epsilon t^{1/2} \,\middle|\, N(t)\right)\right) \\ &= P\left(Y > \epsilon t^{1/2}\right)\,, \quad \forall \epsilon > 0\,. \end{aligned}$$

Hence

$$\frac{\lambda t - N(t)}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}} = \frac{\lambda T_{N(t)} - N(t)}{\left(\sigma_Y^2 \lambda^3 t\right)^{1/2}} + o_P(1)\,. \tag{2.41}$$

In view of Theorem 2.5.9 and by the continuous mapping theorem the rhs converges weakly to Φ. This proves the theorem. □

2.5.3 Random Sums Driven by Renewal Counting Processes

In this section we consider one of the most important models in insurance mathematics. Throughout, we assume that the random sums $S(t) = S_{N(t)}$ are driven by a *renewal counting process* as defined in (2.36). The process $(S(t))$ is a model for the total claim amount of an insurance portfolio. The renewal and the Cramér–Lundberg models (Definition 1.1.1) are included in this setting as particular cases when $(N(t))$ and (X_n) are independent. In general, we do not require this assumption. In what follows we are interested in the asymptotic properties of the process $(S(t))$.

Recall from Section 2.5.2 that $N(t) \stackrel{\text{a.s.}}{\rightarrow} \infty$. Hence we may apply the Marcinkiewicz–Zygmund SLLNs for random sums (Theorem 2.5.5),

$$(N(t))^{-1/p} \; (S(t) - a\,N(t)) \stackrel{\text{a.s.}}{\rightarrow} 0\,, \tag{2.42}$$

provided $E|X|^p < \infty$ for some $p \in (0,2)$ and

$$a = \begin{cases} 0 & \text{if} \quad p < 1\,, \\ \mu = EX & \text{if} \quad p \in [1,2)\,. \end{cases} \tag{2.43}$$

The following question arises naturally:

May we replace $N(t)$ in (2.42) by a deterministic function, for instance λt?

The answer is

In general: NO.

However, by Theorem 2.5.10, $N(t)/t \stackrel{\text{a.s.}}{\rightarrow} \lambda$ provided $EY < \infty$. Hence we may replace the normalising process $(N(t))^{1/p}$ by $(\lambda t)^{1/p}$. The centring process causes some problems. To proceed, suppose $E|X|^p < \infty$ for some $p \in [1,2)$. We write

$$t^{-1/p} \; (S(t) - \lambda\mu t) = t^{-1/p} \; (S(t) - \mu\,N(t)) + \mu t^{-1/p} \; (N(t) - \lambda\,t)\;.$$

In view of (2.42), the first term on the rhs converges to zero a.s. provided the first moment of Y is finite. On the other hand,

$$\mu t^{-1/p} \; (N(t) - \lambda\,t) \stackrel{\text{a.s.}}{\rightarrow} 0 \tag{2.44}$$

does not hold in general. But if $EY^p < \infty$ we conclude from Theorem 2.5.10 that (2.44) is satisfied. In summary we obtain:

Theorem 2.5.14 (Marcinkiewicz–Zygmund SLLNs for random sums)
Suppose that $E|X|^p < \infty$ for some $p \in (0,2)$.

(*a*) *If* $EY < \infty$ *then*

$$t^{-1/p}\ \left(S(t) - a\,N(t)\right) \stackrel{\text{a.s.}}{\rightarrow} 0\,,$$

where a *is defined by* (2.43)*.*

(*b*) *If* $p \geq 1$ *and* $EY^p < \infty$ *then*

$$t^{-1/p}\ \left(S(t) - \mu\lambda\, t\right) \stackrel{\text{a.s.}}{\rightarrow} 0\,. \qquad \square$$

In the weak convergence case we can proceed analogously. Since $N(t) \to \infty$ a.s. for a renewal counting process, the CLT for random sums applies under mild conditions:

Theorem 2.5.15 (Anscombe–type CLT for random sums)
Assume that $F \in \mathrm{DA}(\alpha)$ *for some* $\alpha \in (0,2]$ *and that*

$$\left(n^{1/\alpha}L(n)\right)^{-1}\ \left(S_n - \widetilde{a}n\right) \stackrel{d}{\rightarrow} G_\alpha\ ,$$

for some α*–stable distribution* G_α *and a slowly varying function* L*. Here*

$$\widetilde{a} = \begin{cases} 0 & \textit{if} \quad \alpha \leq 1\,, \\ \mu & \textit{if} \quad \alpha \in (1,2]\,. \end{cases}$$

(*a*) *If* $EY < \infty$ *then*

$$\left((\lambda t)^{1/\alpha}L(t)\right)^{-1}\ \left(S(t) - \widetilde{a}N(t)\right) \stackrel{d}{\rightarrow} G_\alpha\,. \tag{2.45}$$

In particular, if $\sigma^2 = \mathrm{var}(X) < \infty$ *then*

$$\left(\lambda\sigma^2 t\right)^{-1/2}\ \left(S(t) - \mu N(t)\right) \stackrel{d}{\rightarrow} \Phi\ , \tag{2.46}$$

where Φ *is the standard normal distribution.*

(*b*) *If* $\alpha \in (1,2)$ *and* $EY^p < \infty$ *for some* $p > \alpha$ *then*

$$\left((\lambda t)^{1/\alpha}L(t)\right)^{-1}\ \left(S(t) - \lambda\mu\, t\right) \stackrel{d}{\rightarrow} G_\alpha\,. \tag{2.47}$$

Proof. If $EY < \infty$ then, by Theorem 2.5.10, $N(t)/t \stackrel{\text{a.s.}}{\rightarrow} \lambda$, and Theorem 2.5.9 applies immediately. This yields (2.45) and (2.46).
If $EY^p < \infty$ and $p \in [1,2)$ then, by Theorem 2.5.10,

$$t^{-1/p}\ \left(N(t) - \lambda t\right) \stackrel{\text{a.s.}}{\rightarrow} 0\,. \tag{2.48}$$

Hence

$$
\begin{aligned}
& \left((\lambda t)^{1/\alpha} L(t)\right)^{-1} (S(t) - \lambda\mu\, t) \\
= \; & \left((\lambda t)^{1/\alpha} L(t)\right)^{-1} ((S(t) - \mu\, N(t)) + \mu\, (N(t) - \lambda\, t)) \\
= \; & \left((\lambda t)^{1/\alpha} L(t)\right)^{-1} \; (S(t) - \mu\, N(t)) + o(1) \quad \text{a.s.}
\end{aligned} \tag{2.49}
$$

Here we used (2.48) and the fact that, for $p > \alpha$,

$$\lim_{t\to\infty} \frac{t^{1/\alpha} L(t)}{t^{1/p}} = \infty$$

which is a consequence of the slow variation of L. Relation (2.47) is now immediate from (2.45) and (2.49) by the continuous mapping theorem (Theorem A2.6). □

Note that we excluded the case $\alpha = 2$ from (2.47). In that case the method of proof fails. Indeed, (2.49) is no longer applicable if $\sigma^2 < \infty$. This follows from the CLT in Theorem 2.5.13.

Now we try to combine the CLT for $(S(t))$ and for $(N(t))$. Assume that $\sigma^2 < \infty$ and $\sigma_Y^2 < \infty$. Using (2.41), we obtain

$$
\begin{aligned}
t^{-1/2}\, (S(t) - \lambda\mu\, t) \;&=\; t^{-1/2}\, ((S(t) - \mu\, N(t)) + \mu(N(t) - \lambda\, t)) \\
&=\; t^{-1/2} \left((S(t) - \mu\, N(t)) + \mu \left(N(t) - \lambda T_{N(t)}\right)\right) + o_P(1) \\
&=\; t^{-1/2} \sum_{i=1}^{N(t)} (X_i - \mu\lambda Y_i) + o_P(1)\,.
\end{aligned} \tag{2.50}
$$

Notice that the rvs $X_i' = X_i - \mu\lambda Y_i$ have mean zero. Moreover, the sequence (X_i') is iid if $((X_n, Y_n))$ is iid. Under the latter condition, Theorem 2.5.9 applies immediately to (2.50) and yields the following result:

Theorem 2.5.16 *Suppose that $((X_n, Y_n))$ is a sequence of iid random vectors and that $\sigma^2 < \infty$ and $\sigma_Y^2 < \infty$. Then*

$$(\operatorname{var}(X - \mu\lambda Y)\lambda t)^{-1/2}\, (S(t) - \mu\lambda t) \xrightarrow{d} \Phi\,,$$

where Φ denotes the standard normal distribution.
In particular, if (X_n) and (Y_n) are independent then

$$\left(\left(\sigma^2 + (\mu\lambda\sigma_Y)^2\right)\lambda t\right)^{-1/2} (S(t) - \mu\lambda\, t) \xrightarrow{d} \Phi\,. \qquad \square$$

Following the idea of proof of Theorem 2.5.16, one can derive results if $\sigma^2 = \infty$ or $\sigma_Y^2 = \infty$ with appropriate stable limit distributions; see for instance Kotulski [405].

It is also possible to derive different versions of FCLTs with Gaussian or α–stable limit processes for $(S(t))$. We state here one standard result, versions of which can be found in Billingsley [69], Section 17, and in Gut [291], Theorem 2.1 in Chapter V. In our presentation we follow Grandell [281], p. 47. Recall the notion of weak convergence from Appendix A2.4 and compare the following result with the Donsker invariance principle (Theorem 2.4.4).

Theorem 2.5.17 (FCLT for random sum process)
Let (X_n) be a sequence of iid rvs such that $\sigma^2 < \infty$. Assume that the renewal counting process $(N(t))$ and (X_n) are independent and that $EY = 1/\lambda$ and $\sigma_Y^2 < \infty$. Let $B.$ be standard Brownian motion on $[0,\infty)$. Then

$$\left(\left(\sigma^2 + (\mu\lambda\sigma_Y)^2\right)\lambda n\right)^{-1/2}\left(S_{N(n\cdot)} - \lambda\mu\, n\cdot\right) \xrightarrow{d} B.$$

in $\mathbb{D}\,[0,\infty)$ equipped with the J_1–metric and the corresponding σ–algebra of the open sets. □

This theorem has quite an interesting application in insurance mathematics:

Example 2.5.18 (Diffusion approximation of the risk process)
Consider the Cramér–Lundberg model (Definition 1.1.1), i.e. $(S(t))$ is compound Poisson with positive iid claims (X_n) independent of the homogeneous Poisson process $(N(t))$ with intensity $\lambda > 0$. The corresponding risk process with initial capital u and premium income rate $c = (1+\rho)\lambda\mu > 0$ (with safety loading $\rho > 0$) is given in (1.4) as

$$U(t) = u + ct - S(t)\,, \quad t \geq 0\,.$$

In Chapter 1 we mainly studied the ruin probability in infinite time. One method to obtain approximations to the ruin probability $\psi(u,T)$ in finite time T, i.e.

$$\begin{aligned}\psi(u,T) &= P\left(U(t) < 0 \text{ for some } t \leq T\right)\\ &= P\left(\inf_{0\leq t\leq T}(ct - S(t)) < -u\right)\,,\end{aligned}$$

is the so–called *diffusion approximation* which was introduced in insurance mathematics by Iglehart [351]; see also Grandell [282], Appendix A.4, for an extensive discussion of the method. Define

$$\widetilde{\sigma}^2 = \left(\sigma^2 + (\mu\lambda\sigma_Y)^2\right)\lambda = \left(\sigma^2 + \mu^2\right)\lambda\,.$$

Then

$$
\begin{aligned}
&\psi(u,T)\\
&= P\left(\inf_{0\le t\le T}((1+\rho)\lambda\mu t - S(t)) < -u\right)\\
&= P\left(\inf_{0\le t\le T/n}((1+\rho)\lambda\mu tn - S(tn)) < -u\right)\\
&= P\left(\inf_{0\le t\le T/n}\left(\widetilde{\sigma}^2 n\right)^{-1/2}((1+\rho)\lambda\mu tn - S(tn)) < -u\left(\widetilde{\sigma}^2 n\right)^{-1/2}\right).
\end{aligned}
$$

Now assume that $T_0 = T/n$, $\rho_0 = \rho\lambda\mu\widetilde{\sigma}^{-1}\sqrt{n}$ and $u_0 = u\,(\widetilde{\sigma}^2 n)^{-1/2}$ are constants, i.e. we increase T and u with n, and decrease at the same time the safety loading ρ with n. This means that a small safety loading is compensated by a large initial capital. Then we obtain

$$
\psi(u,T) = P\left(\inf_{0\le t\le T_0}\left(\left(\widetilde{\sigma}^2 n\right)^{-1/2}(\lambda\mu tn - S(tn)) + \rho_0 t\right) < -u_0\right).
$$

The functional $x(f) = \inf_{0\le t\le T_0} f(t)$ is continuous on $\mathbb{D}\,[0,T_0]$. Thus we may conclude from Theorem 2.5.17 by the continuous mapping theorem (note that u, ρ and T depend on n) that

$$
\begin{aligned}
\psi(u,T) &\to P\left(\inf_{0\le t\le T_0}(\rho_0 t - B_t) < -u_0\right)\\
&= P\left(\sup_{0\le t\le T_0}(B_t - \rho_0 t) > u_0\right). \qquad (2.51)
\end{aligned}
$$

The latter approach is called a *diffusion approximation* since Brownian motion is a special diffusion process. The distribution of the supremum functional of Brownian motion with linear drift is well known; see for instance Lerche [421], Example 1 on p. 27:

$$
P\left(\sup_{0\le t\le T_0}(B_t - \rho_0 t) > u_0\right) = \overline{\Phi}\left(\frac{\rho_0 T_0 + u_0}{\sqrt{T_0}}\right) + e^{-2u_0\rho_0}\Phi\left(\frac{\rho_0 T_0 - u_0}{\sqrt{T_0}}\right).
$$

The diffusion approximation has many disadvantages, but also some good aspects. We refer to Grandell [282], Appendix A.4, and Asmussen [28] for a discussion and some recent literature; see also Schmidli [567] and Furrer, Michna and Weron [247]. The latter look at weak approximations of the risk process by α–stable processes. Among the positive aspects of the diffusion approximation is that it is applicable to a wide range of risk processes which deviate from the Cramér–Lundberg model. In that case, the classical methods from renewal theory as developed in Chapter 1 will usually break down, and the diffusion approach is then one of the few tools which work. In such more general models it is usually not possible to choose the premiums as a linear function in time; see for example Klüppelberg and Mikosch [392, 393] for a shot noise risk model. As a contra one might mention that the choice of T_0, ρ_0 and u_0 is perhaps not the most natural one. On the other hand, "large" values of T and u and small values of ρ are relative and down to individual judgement. Notice that in Chapter 1 the probability of ruin in infinite time was approximated for "large" initial capital u. Nevertheless, if one wants to use the diffusion approximation for practical purposes a study of the values of T, ρ and u for which the method yields reasonable results is unavoidable. For example, Grandell [282], Appendix A.4, gives a simulation study. □

Notes and Comments

There are several texts on random sums, renewal counting processes and related questions. They are mainly motivated by renewal theory. A more advanced limit theory, but also the proofs of the standard results above can be found in Gut [291]. The classical theory of random sums relevant for risk theory was reviewed in Panjer and Willmot [489]. Other relevant literature is Asmussen [27] and Grandell [282]. The latter deals with the total claim amount process and related questions of risk and ruin for very general processes. Grandell [284] gives an overview of the corresponding theory for mixed Poisson processes and related risk models. A recent textbook treatment of random sums is Gnedenko and Korolev [268].