8

Random Processes with Independent Increments

8.1 Introduction

8.1.1. In Chapter 7, we saw that viewing Heads and Tails as a random process enabled us to present certain problems (laws of large numbers, the central limit theorem) in a more expressive form – as well as giving us insights into their solution. It was, essentially, a question of obtaining a deeper understanding of the problems by looking at them from an appropriate dynamic viewpoint.

This same dynamic viewpoint lends itself, in a natural way, to the study of a number of other problems. Not only does it serve as an aid to one's intuition, but also, and more importantly, it reveals connections between topics and problems that otherwise appear unconnected (a common circumstance, which results in solutions being discovered twice over, and hence not appearing in their true perspective); in so doing, it provides us with a unified overall view.

We have already seen, in the case of Heads and Tails, how the representation as a process enabled us to derive, in an elegant manner, results which could then easily be extended to more general cases. We now proceed by following up this idea in two different, but related, directions:

making precise the kind of process to be considered as a first development of the case of Heads and Tails;

considering (first for the case of Heads and Tails, and then in the wider ambit mentioned above) problems of a more complicated nature than those studied so far.

8.1.2. The random processes that we shall consider first are those *with independent increments*,[1] and we shall pay special attention to the *homogeneous* processes. This will be the case both for processes *in discrete time* (to which we have restricted ourselves so far) and for those in *continuous time*.

In *discrete time* (where t assumes integer values), the processes will be of the same form as those already considered: $Y(t) = X_1 + X_2 + \dots + X_t$, the sum of *independent* random quantities (increments) X_i. In terms of the Y, one can describe the process by

1 Here, and in what follows, *independent* always means *stochastically independent*.

Theory of Probability: A Critical Introductory Treatment, First Edition. Bruno de Finetti.

saying that the increments in $Y(t)$ over disjoint time intervals are independent: that is $Y(t_2) - Y(t_1)$ and $Y(t_4) - Y(t_3)$ are independent for $t_1 < t_2 \leqslant t_3 < t_4$. The increments are, in fact, either the Xs themselves, or sums of distinct Xs, according to whether we are dealing with unit time, or with a longer time interval.

Such a process is called *homogeneous* if all the X_i have precisely the same distribution, F. More generally, all increments $Y(t + \tau) - Y(t)$ relating to intervals with the same length, τ, have exactly the same distribution (given by the convolution $F^{\tau *}$).

In *continuous time* (a case which we have so far only mentioned in passing) the above conditions, expressed in terms of the increments of $Y(t)$, remain unchanged; except that now, of course, they must hold for arbitrary times t (instead of only for integer values, as in the discrete time case). This makes the conditions *much more restrictive*. The consequence of this is that whereas in the discrete case all distributions were possible for the Xs (and all distributions decomposable into a t-fold convolution were admissible for the $Y(t)$), in continuous time we can only have, for the $Y(t)$ and their increments $Y(t + \tau) - Y(t)$, those distributions whose decomposition can be taken a great deal further: in other words, the *infinitely divisible* distributions (which we mentioned in Chapter 6, 6.11.3).[2]

In such a process, the function $Y(t)$ can be thought of as decomposed into two components, $Y(t) = Y_J(t) + Y_C(t)$, the first of which varies *by jumps*, and the second *continuously*. The arguments we shall use, based on this idea, incomplete though they are at the moment, are essentially correct so far as the conclusions are concerned, although critical comments are required at one point (Section 8.3.1) in connection with the interpretation of these conclusions and the initial concepts.

The conclusion will be that the distribution of $Y(t)$ (or of the increment $Y(t + \tau) - Y(t)$, respectively) can be completely and meaningfully expressed by considering the two components Y_J and Y_C separately:

in order to characterize the distribution of $Y_J(t)$ it is necessary and sufficient to give *the prevision*, over the interval $[0, t]$ ($[t, t + \tau]$, respectively), *of the number of jumps of various sizes*;

in order to characterize the distribution of $Y_C(t)$ it is necessary and sufficient to give *the prevision and the variance of the continuous component* (since, as we shall see, its distribution is necessarily *normal*);

taken together, these characterize $Y(t)$.

All the previsions are additive functions of the intervals and, except for that of the increment of the continuous component, Y_C, they are essentially non-negative, and therefore nondecreasing.

In the *homogeneous* case, they depend only on the length of the interval, and are therefore proportional to it.

8.1.3. We shall illustrate straightaway, by presenting some of the most important cases, the structure that derives from what we have described as the most general form

2 These statements are not quite correct as they stand. Firstly, X_i could be a *certain* number; secondly, the jumps might occur at *known* instants. These are trivial special cases, however, which could be considered separately (see Section 8.1.4).

of random process with independent increments. Although this will only be a summary, it should help to make clear the scope of our investigation and also give an idea of the kinds of problems we shall encounter.

Let us first restrict ourselves to the *homogeneous* case (it will be seen subsequently (Section 8.1.4) that the extension to the general case is reasonably straightforward and only involves minor additional considerations).

It is convenient to indicate and collect together at this point the notation that will be used in what follows. This will be given for the homogeneous case; hence a single distribution suffices for the increments $Y(t_0 + t) - Y(t_0)$. This will depend on t but not on t_0 and will also be the distribution of $Y(t)$ if the initial condition $Y(0) = 0$ is assumed (as will usually be convenient). The distribution function, density (if any) and characteristic function of this distribution will be denoted by $F^t(y)$, $f^t(y)$ and $\phi^t(u)$ respectively. The t is, in fact, an exponent of $\phi(u)$ (as is obvious from the homogeneity and the independence of the increments) and is used as a superscript for F and f both for uniformity and to leave room for possible subscripts. Its use is also partially justified by the fact that F^t and f^t are, actually t-fold convolutions, $(F^1)^{t*}$ and $(f^1)^{t*}$, of F^1 and f^1 with themselves, provided that the concept (where it makes sense, as is the case here for all t, because of infinite divisibility) is suitably extended to the case of noninteger exponent t. The distribution and density (if any) of the *jumps* will be denoted by $F(x)$ and $f(x)$, respectively (with no superscripts), and the characteristic function by $\chi(u)$.

Let us examine now the various cases.

The simplest example, and the one which forms the basis for the construction of all the random processes of the type under consideration, is that of the *Poisson process*. This is a jump process, all jumps being of the same size, x; for the time being, we shall take $x = 1$, so that $Y(t) = N(t) =$ the *number of jumps* in $[0, t]$ (see Figure 8.1). We shall denote by μ the *prevision* of the number of jumps occurring per unit time (i.e. μt is the prevision of the number of jumps in a time interval of length t). In an infinitesimal time, $\mathrm{d}t$, the prevision of the number of jumps is $\mu\mathrm{d}t$; up to an infinitesimal quantity of greater order, this is also the probability that *one* jump occurs within the small time interval (the probability of more than one jump occurring is, in comparison, negligible). We call μ the *intensity* of the process.

The distribution of $N(t)$, the number of jumps occurring before time t, is Poisson, with prevision $a = \mu t$ (see Chapter 6, 6.11.2, equation 6.39, for the explicit form of the probabilities and the characteristic function).

We recall that the variance in this case is also equal to μt. It is better, however, to note explicitly that the prevision is $x\mu t$, and the variance is $x^2\mu t$, where x is the magnitude of the jump. In this way, one avoids the ambiguities which arise from ignoring dimensional questions (i.e. taking x = pure number) and, subsequently, from assuming the special value $x = 1$ (for which $x = x^2$).

A superposition of several Poisson processes, having jumps of different sizes, x_k, and different intensities, μ_k ($k = 1, 2, \ldots, m$), is also a jump process, homogeneous with independent increments, $Y(t) = \Sigma x_k N_k(t)$. It also has an alternative interpretation as a process of intensity $\mu = \mu_1 + \mu_2 + \ldots + \mu_m$, where each jump has a random size X (independently of the others), X taking the value x_k with probability μ_k/μ.

Instead of considering the sum of a finite number of terms, we could also consider an infinite series, or even an integral (in general, a Stieltjes integral). Provided the total intensity remains finite, the above interpretations continue to apply (except that X now

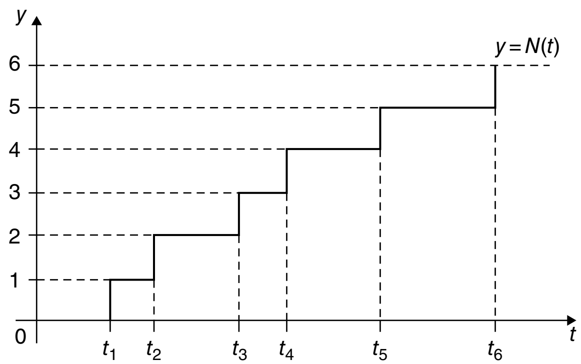


Figure 8.1 The simple Poisson process.

has an arbitrary distribution, rather than the discrete one given above). This case is referred to as a *compound Poisson process* and provides the most general homogeneous process with independent increments and *discrete jumps* (i.e. finite in number in any bounded interval).

Using the same procedure, one can also obtain the *generalized Poisson processes*: that is those *with everywhere dense jumps* (discontinuity points), with $\mu = \infty$. It is necessary, of course, to check that the process does not diverge: for this, we require that the intensity μ_ε of jumps X with $|X| > \varepsilon(\varepsilon > 0)$ remains finite, diverging only as $\varepsilon \to 0$, and then not too rapidly.

The Poisson processes, together with the compound and generalized forms, exhaust all the possibilities for the jump component $Y_J(t)$.

It remains to consider the continuous component $Y_C(t)$. In every case, $Y(t)$ could be considered as the sum of N increments (N arbitrary), corresponding to the N small intervals of length t/N into which the interval $[0, t]$ could be decomposed. If one makes precise the idea of separating off the 'large' increments (large in absolute value), which correspond to the jumps, it can be shown that the others (the 'small' increments) satisfy the conditions of the central limit theorem, and hence the distribution is necessarily normal (as we mentioned above).

As we have already stated, its prevision varies in a linear fashion, $\mathbf{P}[Y_C(t)] = mt$, and the same is true of the variance, $\mathbf{P}[\{Y_C(t) - mt\}^2] = \sigma^2 t$ (where m and σ^2 denote the prevision and variance corresponding to $t = 1$).[3] In some cases, it may also turn out to be convenient to separate the certain linear function, mt, and the fair component (with zero prevision), $Y_C(t) - mt$ (whose variance is still $\sigma^2 t$).

The set-up we have just described is called the *Wiener–Lévy process* (see also Chapter 7, 7.6.5).

8.1.4. So far as the increment over some given interval is concerned, the conclusions arrived at in the homogeneous case carry over without modification to the general case. This is clear, because the conclusions depend on the prevision (of the number of certain jumps etc.) over such an interval as a whole, and not on the way the prevision is distributed

3 These formulae hold, of course, for every $Y(t)$ that is homogeneous with independent increments (with finite m and σ^2). We mention them here, for the particular case of Y_C, only because of their importance in specifying the distribution in this case.

over the subintervals. In the inhomogeneous case, the only new feature is that, within an interval, each prevision can be increasing in an arbitrary way, not necessarily linearly.

The one thing that is required is that one exclude (or, better, consider separately, if they exist) the points of discontinuity for such a prevision. In fact, at such points one would have a nonzero probability of discontinuity (a 'fixed discontinuity'; see the remark in the footnote to Section 8.1.2). This is equivalent to saying that at these points $Y(t)$ receives a random (instantaneous) increment, incompatible with the nature of the 'process in continuous time', because (being instantaneous) it is not decomposable, and therefore does not have to obey the 'infinite divisibility' requirement. In what follows, we shall always tacitly assume that such fixed discontinuities have been excluded.

Observe that, if all the previsions are proportional to one another, the process can be said to be homogeneous with respect to a different time scale (one which is proportional to them). In general, however, the previsions will vary in different ways, and then we have no recourse to anything of this kind.

8.1.5. Let us now consider more specifically some of the problems that one encounters. These are of interest for a number of reasons : by virtue of their probabilistic significance and their range of application; because of the various mathematical aspects involved; and, above all, because of the unified and intuitively meaningful presentation that can be given of a vast collection of seemingly distinct problems. The problems that we shall mention are only a tiny sample from this collection and the treatment that we shall give will only touch upon some of the more essential topics, presenting them in their simplest forms.

First of all, we have to translate into actual mathematical terms (through the distribution, by means of the characteristic function) the characterization of the general process with independent increments, and hence of the most general infinitely divisible distribution (which we have already given above, in terms of the intensities of the jumps and the normal component).

Particular attention must be paid, however, to the limiting arguments that lead to the generalized Poisson process, since the latter gives rise to rather different problems. For example, in our preliminary remarks we did not mention that sometimes convergence can only be achieved by 'compensating' the jumps by means of a certain linear function and that, in this case, the intuitive idea of behaviour similar to that of the discrete case (apart from minor details) must undergo a radical change.

Even the behaviour of the continuous component (the Wiener–Lévy process), despite what one might at first sight be led to expect from the regular and familiar shape of the normal distribution, turns out to be extremely 'pathological'. The study of the behaviour of the function (or, more precisely, the behaviour which it 'almost certainly' enjoys) is, however, a more advanced problem. We shall begin by saying something about the distribution.

How does the distribution of $Y(t)$ behave as t increases? We already know that it tends to normality in the case of finite variances, but there also exist processes (of the generalized Poisson type) with infinite variances. The answer in this case in that there exist other types of stable distribution, all corresponding to generalized Poisson processes (more precisely, as we shall see in Sections 8.4.1–8.4.4, there are a doubly infinite collection of them, reducible, essentially, to a single infinite collection). The processes which are not stable either tend to a stable form, or do not tend to anything at all.

The key to the whole question lies in the behaviour at the other extreme, as $t \to 0$. This is directly connected with the intensity of the jumps of different sizes: one has stability if the intensity of the jumps $>x$ or $<-x$ decreases proportionally to $x^{-\alpha}(x > 0, 0 < \alpha < 2)$; one has a tendency to a stable form if the process, in some sense, approximately satisfies this condition. Referring back to a remark made previously, we note that the 'sum of the jumps' does not require 'compensation' if $\alpha < 1$, but does if $\alpha > 1$ ($\alpha = 1$ constitutes a subcase of its own). It follows that the stable distributions generated by just the positive (negative) jumps extend only over the positive (negative) values if $\alpha < 1$, whereas they extend over all positive and negative values if $\alpha \geqslant 1$.

An important special case is that of 'lattice' processes, in which $Y(t)$ can only assume integer values (or those of an arithmetic progression; there is no essential difference). They must, of course, be compound Poisson processes, but this does not mean that they cannot tend to a stable (continuous distribution as t increases. Indeed, if the variances are finite they necessarily tend to the normal distribution (as we already saw, in the first instance, in the case of Heads and Tails).

8.1.6. The study of the way in which the distribution of $Y(t)$ varies with t does not necessarily entail the study of the behaviour of the actual process $Y(t)$ (i.e. of the function $Y(t)$). The essential thing is to examine the characteristics of the behaviour of the latter, which are of interest to us, behaviour which can only be studied by the simultaneous consideration and comparison of values of $Y(t)$ at different times t (possibly a large or infinite number of them).

So far as the phrase 'infinite number' is concerned, we should make it clear at once that it means 'an arbitrarily large, but finite, number' (unless, in the case under consideration, one makes some additional assumption, such as the validity of countable additivity, or gives some further explanation). However, in order to avoid making heavy weather of the presentation with subtle, critical arguments, we shall often resort to intuitive explanations of this kind (as well as to the corresponding 'practical' justifications).

Problems that have already been encountered in the discrete cases – like that of the asymptotic behaviour of $Y(t)/t$ (as $t \to \infty$), to which the strong law of large numbers gives the solution – are restated and shown to have, generally speaking, similar solutions in the continuous case. One also meets, in the latter case, problems that are, in a certain sense, reciprocal; involving what happens as $t \to 0$ ('local' behaviour – like 'continuity' etc. – at the origin and, hence, in the homogeneous case, at any arbitrary point in time). In the Wiener–Lévy process, the two problems correspond exactly to one another by reciprocity.

If one confines attention to considerations based on the Tchebychev inequality, the conclusions hold for every homogeneous process with independent increments for which the variance is finite. One such process, viewing the jumps and possible horizontal segments 'in the large' (i.e. with respect to large time intervals and intervals of the ordinate), in such a way as to make them imperceptible, is the Wiener–Lévy process. Indeed, Lévy calls it a Brownian motion process, which corresponds to the perception of an observer who is not able to single out the numerous, tiny collisions that, in any imperceptibly small time interval, suddenly change the motion of the particle under observation.

More generally, even for arbitrary random processes (provided they have finite variance), answers can be obtained to a number of problems, even though, in general, they may only be qualitative and based on second-order characteristics (as in the case of a single random quantity, or of several). Here, the random quantities to be considered are the values $Y(t)$ and the characteristics to be used are the previsions, variances and covariances (or, equivalently, $\mathbf{P}[Y(t)]$, $\mathbf{P}[Y^2(t)$, $\mathbf{P}[Y(t_1)Y(t_2)]$). In our case, this is trivial: evaluated at $t = 0$, the prevision and variance of $Y(t)$ are, as we know, mt and $\sigma^2 t$, and the covariance of $Y(t_1)$ and $Y(t_2)$ is $\sigma^2 t_1$ (if $t_1 \leqslant t_2$); the correlation coefficient is, therefore, $r(t_1, t_2) = \sigma^2 t_1/\sigma\sqrt{t_1} \cdot \sigma\sqrt{t_2} = \sqrt{(t_1/t_2)}$ (it is sufficient to observe that $Y(t_2) = Y(t_1) + [Y(t_2) - Y(t_1)]$, and that the two summands are independent).

There are other problems, however, that require one to take all the characteristics into account and to have recourse to new methods of approach. (This also applies to the problems considered previously, if one wishes to obtain conclusions that are more precise, in a quantitative sense, and more specifically related to the particular process under consideration.) A concept that can usefully be applied to a number of problems is that of a *barrier* (a line in the (t, y)-plane on which $y = Y(t)$ is represented). One observes when the barrier is first reached, or when it is subsequently reached, or, sometimes, one assumes that the barrier modifies the process (it may be absorbing – i.e. the process stops – or reflecting, and so on).

The classical problem of this kind, and one that finds immediate application, is that of the 'gambler's ruin' (corresponding to the point when his gain reaches the level $-c$, where c is his initial capital). There are a number of obvious variants: one could consider the capital as being variable (an arbitrary barrier rather than the horizontal line $y = -c$), or one could think of two gamblers, both having a bounded initial capital (or variable capital), and so on.

In addition to this and other interesting and practical applications, problems of this kind also find application in studying various aspects of the behaviour of the function $Y(t)$. In particular, and this question has been studied more than any other, they are useful for specifying the asymptotic behaviour, indicating which functions tend to zero too rapidly, or not rapidly enough, to provide a (practically certain) bound for $Y(t)/t$ from some point $t = T$ on (and similarly as $t \to 0$).

Reflecting barriers, and others which modify the process, take us beyond the scope of this present chapter. In any case, they will enter, in a certain sense, into the considerations we shall make later, and will provide an instructive and useful technique. In particular, they will enable us to make use of the elegant and powerful arguments of Desiré André (and others) based on symmetries.

8.2 The General Case: The Case of Asymptotic Normality

8.2.1. Let us first give a precise, analytic statement of what we have hitherto presented in a descriptive form concerning the structure and properties of the general homogeneous process with independent increments.

When the intensity of the jumps, μ, and the variance, σ^2, are finite, the process is either normal (if $\mu = 0$), or asymptotically normal (as we have already seen from the central limit theorem). In other words, we either have the Wiener–Lévy process, or something which approximates to it asymptotically. We are not saying that the restrictions made are necessary for such behaviour: the restriction on μ, that is $\mu < \infty$, has no direct relevance, and the restriction $\sigma^2 < \infty$ could be weakened somewhat (see Chapter 7, 7.7.3). However, for our immediate purpose it is more appropriate to concentrate attention on the simplest case, avoiding tiresome complications which do not really contribute anything to our understanding.

Our first task, once we have set up the general analytical framework, is to provide insight into the way in which, and in what sense, processes of this kind (i.e. those which are asymptotically normal) can be considered as an approximation, on some suitable scale, to the Wiener–Lévy process, and conversely. In this way, any conclusion established for a special case, for example, that of Heads and Tails, turns out to be necessarily valid (in the appropriate asymptotic version) in the general case. Note that this enables us, among other things, to establish properties of the Wiener–Lévy process by means of elementary combinatorial techniques, which are, in themselves, applicable only to the case of Heads and Tails. Conversely, it enables us to derive properties of the latter, and of similar examples, that cannot be obtained directly (for example, asymptotic properties) by using approaches – often much simpler – which deal with the limit case of the Wiener–Lévy process (itself based on the normal distribution). This is just one example, out of many, where the possibility exists of advantageously switching from a discrete schematization to a continuous one, or vice versa, as the case may be.

8.2.2. The *Wiener–Lévy process* can be derived as the limit case of the *Heads and Tails* process (in discrete time).

Suppose, in fact, that we change the scale, performing tosses at shorter time intervals, and with smaller stakes, so that the variance per unit time remains the same. For this to be so, if the stake is reduced by a factor of N ($a = 1/N$) the number of tosses per unit time must be increased to N^2 (with time intervals $\tau = 1/N^2$). To see this, note that the variance for each small time interval τ is $a^2 = (1/N)^2$, and in order for it to be equal to 1 per unit time, the number of small intervals by which it has to be multiplied must be N^2.

By taking N sufficiently large, one can arrange things in such a way that the increment per unit time has a distribution arbitrarily close to that of the standardized normal. Taking N even larger, one can arrange for the same properties to hold, also, for the small intervals. In other words, one can arrange for the distribution of $Y(t)/\sqrt{t}$ to be arbitrarily close, to any preassigned degree, to the standardized normal, for every t exceeding some arbitrarily chosen value. This could also be expressed by saying that the Heads and Tails process can be made to resemble (with a suitable change of scale) a process in discrete time with normally distributed jumps, and (with a more pronounced change of scale) even to resemble the Wiener–Lévy process (provided, of course, that one regards as meaningless any claim that the scheme is valid, or, anyway, observable, for arbitrarily small time periods).

In terms of the characteristic function, these considerations reduce, in the first case, to the straightforward and obvious observation that if we substitute e^{-u^2} for $\cos u$ then $[\cos(u/\sqrt{n}]^n \to \mathrm{e}^{-u^2}$ becomes the identity $\left[\mathrm{e}^{-(u/\sqrt{n})^2}\right]^n = \mathrm{e}^{-u^2}$; whereas, in the second case, they simply repeat the procedure used in Chapter 7, 7.7.1. In the Heads and Tails process, $Y(t)$ has characteristic function $\phi^t(u) = (\cos u)^t$ (t = integer); under the above-mentioned change of scale, this becomes $\left[\cos(u/N)\right]^{tN^2}$ (t = integer/N^2) and in the limit (as $N \to \infty$) it becomes e^{-u^2} (t arbitrary).

8.2.3. The Wiener–Lévy process can be obtained in an analogous manner from the *Poisson version of the Heads and Tails process* (a compound Poisson process, with the intensity of the jumps given by $\mu = 1$, and with jumps $\pm$ 1, each with probability $\frac{1}{2}$).

The difference is that instead of there certainly being a toss after each unit time interval the tosses occur at random, with a *prevision* of one per unit time (the probability being given by dt in each infinitesimal time interval of length dt). Alternatively (as we mentioned already in Section 8.1.3), we can say that $Y(t) = Y_1(t) - Y_2(t)$, where Y_1 and Y_2 are the number of positive and negative gains, respectively, both occurring at random and independently, each with intensity $\frac{1}{2}$.

The distribution of $Y(t)$ in such a process is the Poisson mixture of the distributions of Heads and Tails. In terms of the characteristic function, $\phi^t(u)$ is the Poisson mixture (with 'weights' given by the probabilities $e^{-t}t^n/n!$ of there being n jumps in $[0, t]$) of the $(\cos u)^n$ (the characteristic functions of the sums of n jumps; i.e. of $Y(t)$, assuming that there are n jumps up until time t):

$$\phi^t(u) = \sum e^{-t}(t\cos u)^n/n! = e^{t(\cos u - 1)}. \tag{8.1}$$

In this case, too, the same change of scale (jumps reduced by $1/N$, and the intensity increased by N^2) leads to the Wiener–Lévy process. In fact, as $N \to \infty$, the characteristic function $\exp\{tN^2[\cos(u/N) - 1]\}$ tends to e^{-tu^2}.

We therefore obtain the conclusion mentioned above, and it is worth pausing to consider what it actually means. It establishes that the distribution of the gain in a game of Heads and Tails is, after a sufficiently long period of time has elapsed, practically the same, no matter whether tosses were performed in a regular fashion (one after each unit time period), or were randomly distributed (with a Poisson distribution, yielding, *in prevision*, one toss per each unit time period).

8.2.4. The three examples given above (the Heads and Tails process, in both the regular, discrete case and its Poisson variant, and the normally distributed jump process in discrete time) provide the simplest approaches to approximate representations of the Wiener–Lévy process and should be borne in mind in this connection. If we wished, we could also include a fourth such example; the Poisson variant of the normal jump process:

$$\phi^t(u) = \sum e^{-t}\left(t\,e^{-u^2}\right)^n/n! = \exp\left[t\left(e^{-u^2} - 1\right)\right]. \tag{8.2}$$

Strictly speaking, however, if one ignores the psychological case for presenting these introductory examples, the above discussion is entirely superfluous. We have merely anticipated, in a few special cases, ideas which can be examined with equal facility in the general case.

Let us now turn, therefore, to a systematic study of the general case. We begin with the Poisson process, and then proceed to a study of the compound Poisson processes.

8.2.5. The (simple) *Poisson process* deals with the number of occurrences, $N(t)$, of some given phenomenon within a time period $[0, t]$. In other words, it counts the *jumps*, each considered as being of unit size (like a meter that clicks once each time it records a phenomenon – such as the beginning of a telephone conversation, a particle hitting a screen, a visitor entering a museum, or a traveller entering an underground station etc.).

The conditions given in Section 8.1.3 simply mean that we must be dealing with a homogeneous process with independent increments, and with jumps all of unit size.

It is instructive to go back to these conditions and, in the context of the present problem of deriving the probabilities of the Poisson distribution, to provide two alternative derivations in addition to that given in Chapter 6, 6.11.2.

First method. Let a ($a = \mu$) be the prevision of the number of jumps occurring in a given interval (of length t, where μ denotes the intensity). If we subdivide the interval into n equal parts (n large, so that a/n is small compared with 1; i.e. $a/n < \varepsilon$, for some preassigned $\varepsilon > 0$), then a/n is the prevision of the number of jumps occurring in each small interval. We also have $(a/n) = q_n m_n$, where q_n is the probability of *at least one* jump occurring in a small time interval of length t/n, and m_n is the prevision of the number of jumps occurring in intervals containing at least one jump. It follows that we must have $m_n \to 1$ (as $n \to \infty$). If this were not so, it would mean that each discontinuity point had a positive probability of having further jumps in any arbitrarily small neighbourhood of itself; in other words, practically speaking, of being a multiple jump (contrary to the hypothesis that all jumps are of unit size).[4]

The probability that h out of the n small intervals contain discontinuities, and $n - h$ do not, is given by $\binom{n}{h} q_n^h (1-q_n)^{n-h}$. As $n \to \infty$, the probability that there are small intervals containing more than one jump becomes negligible (so that h gives the actual number of jumps). On the other hand, we also have $q_n \simeq a/n$ and, therefore

$$p_h(t) = \lim \binom{n}{h} \left(\frac{a}{n}\right)^h \left(1 - \frac{a}{n}\right)^{n-h}.$$

In this way, we reduce to the formulation and procedure that we have already seen (in Chapter 6, 6.11.2) for so-called 'rare events' (which the occurrences of jumps in very small intervals certainly are).

Second method. We can establish immediately that $p_0(t)$, the probability of no jumps in a time interval of length t, must be of the form e^{-kt}. To see this, we note that, because of the independence assumption,

$$p_0(t' + t'') = p_0(t') p_0(t'')$$

and this relation characterizes the exponential function. The probability that the *waiting time*, T_1, until the occurrence of the first jump does not exceed t is given by $F(t) = 1 - p_0(t) = 1 - e^{-kt}$ (which is equivalent to saying that it is not the case that no jump occurred between 0 and t). From the distribution function $F(t)$, we can derive the density function $f(t) = k\, e^{-kt}$, and we then know that the characteristic function is given by

$$\phi(u) = 1/(1 - ku).$$

4 It would certainly be more direct to impose an additional condition requiring that the probability $p^*(t) = 1 - p_0(t) - p_1(t)$ of there being *two or more* jumps in an interval of length t be an infinitesimal of second order or above. This is, in fact, the approach adopted in many treatments, but it carries the risk of being interpreted as an additional restriction, without which there could be different processes, each compatible with the initial assumptions.

We recall (although, in fact, it follows directly from the above) that the gamma distribution is obtained by convolution:

$$\left[\phi(u)\right]^h = (1-ku)^{-h}, \quad f^{h^*} = Kx^{h-1}\mathrm{e}^{-kt}\,(x \geqslant 0),$$
$$F^{h^*} = 1-\mathrm{e}^{-kt}\left[1+\frac{kt}{1!}+\frac{(kt)^2}{2!}+\ldots+\frac{(kt)^h}{h!}\right].$$

This, therefore, gives the distribution of S_h, the waiting time for the occurrence of the hth jump, which is the sum of the first h waiting times (independent, and exponentially distributed):

$$S_h = T_1 + T_2 + \ldots + T_h.$$

This method of approach is, in a sense, the converse of the first one. The connection is provided by noting that $N(t) < h$ is equivalent to $S_h > t$ ('less than h jumps in $[0, t]$' = 'the hth jump takes place after time t'), and that $N(t) = h$ is equivalent to

$$\{N(t) < h+1\} - \{N(t) < h\} = \{S_{h+1} > t\} - \{S_h > t\}.$$

From this we see that

$$p_h(t) = \mathbf{P}\{N(t) = h\} = \mathbf{P}\{S_{h+1} > t\} - \mathbf{P}\{S_h > t\} = \mathrm{e}^{-kt}(kt)^h/h!, \tag{8.3}$$

again yielding the Poisson distribution. Given that its prevision is kt, it turns out that k, introduced as an arbitrary constant, is actually μ: imagine the latter in place of k, therefore, in the preceding formulae.

Third method. This is perhaps the most intuitive approach and the most useful in that it can be applied to any scheme involving passages through different 'states' with intensities μ_{ij}, constant or variable, where $\mu_{ij}\,\mathrm{d}t$ = the probability that from an initial state 'i' at time t one passes to state 'j' within an infinitesimal time $\mathrm{d}t$.

Let us denote by $p_h(t)$ the functions (assumed to be unknown) that express the probabilities of being in state h at time t (in the Poisson process, state h at time t corresponds to $N(t) = h$, $h = 0, 1, 2, \ldots$). In the general case, the probability of a passage from i to j in an infinitesimal time $\mathrm{d}t$ is given by $p_i(t)\mu_{ij}\,\mathrm{d}t$ (the probability of two passages, from i to some h, and then from h to j, within time $\mathrm{d}t$, is negligible, since it is an infinitesimal of the second order). The change in $p_h(t)$ is given by $\mathrm{d}p_h = p_h\,\mathrm{d}t$ and consists of the positive contribution of all the incoming terms (from all the other i to h), and the negative contribution of the outgoing terms (from h to all the other j). One has, therefore, in the general case (which has been mentioned merely to provide a proper setting for the case of special interest to us), a system of differential equations (which requires the addition of suitable initial conditions).

Our case is much simpler, however: we have only one probability, that of passing from an h to the next $h + 1$, the intensity remaining constant throughout. F or $h = 0$, we have only the outgoing term, $-\mu p_0(t)\,\mathrm{d}t$, whereas for $h > 0$, we have, in addition to $-\mu p_h(t)\,\mathrm{d}t$, the incoming term $\mu p_{h-1}(t)\,\mathrm{d}t$. This reduces to the (recursive) system of equations

$$p_0'(t) = -\mu p_0(t), \qquad p_h'(t) = \mu p_{h-1}(t) - \mu p_h(t), \tag{8.4}$$

together with the initial conditions, $p_0(0) = 1$, $p_h(0) = 0$ $(h \neq 0)$.

From the first equation, we obtain immediately that $p_0(t) = e^{-\mu t}$ and, hence, from the second equation we obtain

$$p_1(t) = \mu t \; e^{-\mu t},$$

and so on. If (realizing from the first terms that it is convenient to extract the factor $e^{-\mu t}$) we set $p_h(t) = e^{-\mu t} g_h(t)$ (with $g_0(t) = 1$, and $g_h(0) = 0$ for $h \neq 0$), we can virtually eliminate any need for calculations: the recursive relation for the $g_h(t)$ reduces to the extremely simple form $g_h'(t) = \mu g_{h-1}(t)$, so that $g_h(t) = (\mu t)^h / h!$.

8.2.6. As an alternative to the method used in Chapter 6, 6.11.2, the characteristic function of the Poisson distribution can be obtained by a direct calculation:

$$\phi^t(u) = \sum_h e^{-\mu t} (\mu t)^h e^{iuh} / h! = \exp\left\{\mu t\left(e^{iu} - 1\right)\right\}. \tag{8.5}$$

So far as the random process is concerned, it is very instructive and meaningful to observe that, as $t \to 0$, we have, asymptotically,

$$\phi^t(u) = 1 + \mu t\left(e^{iu} - 1\right) = (1 - \mu t) + \mu t \; e^{iu}$$

(probability $1 - \mu t$ of 0, and μt of 1). This is the 'infinitesimal transformation' from which the process derives. The simplest way of seeing this is, perhaps, to observe that

$$\phi^t(u) = \lim_{n \to \infty} \left[1 + \frac{1}{n} \mu t\left(e^{iu} - 1\right)\right]^n.$$

The Poisson process also tends to a normal form (as is obvious, given that it has finite variance); in other words, asymptotically it approximates the Wiener–Lévy process. However, the prevision is no longer zero, but equals μt (as does the variance). In order to obtain zero prevision, that is in order to have a finite process, it is necessary to subtract off a linear term and to consider a new process consisting of $N(t) - \mu t$: instead of the number of jumps, one considers the difference between this number and its prevision. The behaviour of

$$Y(t) = N(t) - \mu t$$

gives rise to the saw-tooth appearance of Figure 8.4:[5] all the jumps are equal to +1, and the segments in between have slope $-\mu$. With the introduction of the correction term $-\mu t$, the characteristic function is multiplied by $e^{-i\mu t u}$, and we obtain

$$\exp\left\{\mu t\left(e^{iu} - 1 - iu\right)\right\}. \tag{8.6}$$

This was obvious: when we take the logarithm of the characteristic function, the linear term in u must vanish, and we have

$$e^{iu} - 1 - iu = -\tfrac{1}{2}u^2\left[1 + \varepsilon(u)\right] \qquad \left(\varepsilon(u) \to 0 \qquad \text{as } u \to 0\right).$$

5 Figure 8.4 has been placed later in the text (Section 8.4.3), in order to emphasize its connection with Figure 8.5.

Replacing u by $u/\sqrt{(\mu t)}$, in order to obtain the standardized distribution, we obtain

$$\exp\left\{\mu t\left[-\tfrac{1}{2}\left(u/\sqrt{(\mu t)}\right)^2\right]\left[1+\varepsilon\left(u/\sqrt{(\mu t)}\right)\right]\right\}\to \mathrm{e}^{-\frac{1}{2}u^2}.$$

This provides, if required, a fifth approach to approximating the Wiener–Lévy process. Simple though it is, we thought it worth spending some time on this example, because the idea of adjusting (in the mean) the jumps by a certain linear term, that is of considering jumps with respect to an inclined line rather than a horizontal one, turns out, in a number of cases, to be necessary in order to ensure the convergence of certain procedures (and we shall see examples of this in Sections 8.2.7 and 8.2.9).

8.2.7. The compound Poisson process can be developed along the same lines as were followed in the case of the (simple) Poisson process. The analysis of the most general cases can then be attempted, recognizing that these are, in fact, the generalized Poisson processes.

In the case of a compound Poisson process – with intensity μ, and each jump X having distribution function $F(x)$ and characteristic function $\chi(u) = \mathbf{P}(\mathrm{e}^{\mathrm{i}uX})$ – the characteristic function $\phi^t(u)$ is obtained in exactly the same way as for the simple Poisson process: that is by substituting $\chi(u)$ in place of $\mathrm{e}^{\mathrm{i}u}$ (the latter being the $\chi(u)$ of the simple case, where $X = 1$ with certainty; that is $F(x)$ consists of a single mass concentrated at the point $x = 1$).

This is immediate: one can either note that the 'infinitesimal transformation' is now $1 + \mu t[x(u) - 1] = (1 - \mu t) + \mu t\chi(u)$ (probability $1 - \mu t$ of 0, and probability μt distributed according to the distribution of a jump), from which it follows that

$$\phi^t(u)=\lim_{n\to\infty}\left\{1+\frac{1}{n}\mu t\left[\chi(u)-1\right]\right\}^n=\exp\left\{\mu t\left[\chi(u)-1\right]\right\}, \tag{8.7}$$

or one can simply observe that, conditional on the number of jumps $N(t)$ being equal to h, the characteristic function is given by $\chi^h(u)$, and hence that $\phi^t(u)$ is a mixture of the latter, with weights equal to the probabilities of the individual h. In other words,

$$\phi^t(u)=\sum_{h=0}^{\infty}\left[e^{-\mu t}(\mu t)^h/h!\right]\chi^h(u)=\mathrm{e}^{-\mu t}\sum_{h=0}^{\infty}\left[\mu t\chi(u)\right]^h/h!, \tag{8.8}$$

which is the series expansion of the form given in (8.7). Here too, of course, we are merely rewriting (8.5) with $\chi(u)$ substituted in place of $\mathrm{e}^{\mathrm{i}u}$.

If one wishes to give an expression in terms of the distribution of the jumps, $F(x)$, one can write the characteristic function in the form

$$\phi^t(u)=\exp\left\{\mu t\left[\int \mathrm{e}^{\mathrm{i}ux}\,\mathrm{d}F(x)-1\right]\right\}=\exp\left\{t\int\left(\mathrm{e}^{\mathrm{i}ux}-1\right)\mu\,\mathrm{d}F(x)\right\}, \tag{8.9}$$

or, alternatively,

$$\phi^t(u)=\exp\left\{t\int\left(\mathrm{e}^{\mathrm{i}ux}-1\right)\mathrm{d}M(x)\right\}, \tag{8.10}$$

where $M(x) = \mu F(x)$. As another alternative (in order to deal more satisfactorily with the arbitrariness of the additive constant), we could take

$$\begin{aligned} M(x) &= \mu F(x) && \text{for } x < 0, \\ M(x) &= \mu\left[F(x)-1\right] = -\mu\left[1-F(x)\right] && \text{for } x > 0, \end{aligned} \tag{8.11}$$

so that, to make the situation clear in words,

$M(x)$ = The intensity of jumps having the same sign as x, and greater then x in absolute value, taken with the opposite sign to that of x.

With this definition, $M(x'') - M(x')$ is always the intensity of jumps whose magnitude lies between x' and x'', provided they have the same sign and $x'' > x'$. For $x = 0$, $M(x)$ has a jump $M(+0) - M(-0) = -\mu$, because $M(-0)$ is precisely the intensity of negative jumps, and $M(+0)$ is (with a minus sign) the intensity of positive jumps. The intensity of jumps between some $x' < 0$ and $x'' > 0$ is given by $M(x'') - M(x') + \mu$ (but, usually, one needs to consider separately jumps of opposite sign).

Remark. One can always assume (and we shall always do so, unless we state to the contrary) that there does not exist a probability concentrated at $x = 0$ (i.e. that one can speak of $F(0)$ without having to distinguish between '+0' and '−0', as we have tacitly done when stating that $M(+0) - M(-0) = -\mu[1 - F(0)] - \mu F(0) = -\mu$). In actual fact, so far as any effect on the process is concerned, a 'jump of magnitude $x = 0$' and 'no jump' are the same thing. Mathematically speaking, an increment of F (and hence of M) at $x = 0$ gives a contribution of zero to the integral in equation 8.9, since the integrand vanishes at that point. Sometimes, however, one may make the convention of including in $N(t)$ occurrences of a phenomenon 'able to give rise to a jump', even if the jump does not take place, or, so to speak, is zero. An example of this arises in the field of motor car insurance: if the process $Y(t)$ of interest is the total compensation per accident occurring before time t, it is quite natural (as well as more convenient and meaningful) to count up all accidents, or, to be technical, all claims arising from accidents, without picking out, and excluding, the occasional case for which the compensation was zero. The same principle applies if the process of interest concerns the number of dead, or injured, or those suffering damage to property, and so on.

Formally, in this case one would merely replace μ (the intensity of the jumps) by $\mu + \mu_0$ (where μ_0 is the intensity of the phenomenon with 'zero jumps'), including in $M(x)$ a jump μ_0 at $x = 0$, and consequently altering $F(x)$ and the characteristic function $\chi(u)$, which would be replaced by a mixture of $\chi(u)$ and 1, with weights μ and μ_0. This would be irrelevant, as it must be, since the product $\mu[\chi(u) - 1]$ remains unchanged, and this is all that really matters.

Recall (from Chapter 6, 6.11.6, equation 6.69) that, in order to obtain expressions in normal form ($\mu_0 = 0$), it is necessary and sufficient that we have $(1/2a)\int_{-a}^{a}\chi(u)\,\mathrm{d}u \to 0$ (as $a \to \infty$). Were the limit to equal $c \neq 0$ (necessarily >0), it would suffice to remove $\chi(u)$ and replace it by $[\chi(u) - c]/(1 - c)$.

In the case of a compound process, consisting of a finite number of simple processes (like that considered in Section 8.1.3, the notation of which we continue to use), we have the following:

the $dM(x)$ are the masses (intensities) μ_k concentrated at the values x_k;
$M(x)$ is the sum of the μ_k corresponding to the x_k lying between x and $+\infty$ if $x > 0$, and to those lying between $-\infty$ and x if $x < 0$ (in this case, with the sign changed);
$F(x)$ is the same sum, but running always from $-\infty$ to x, and normalized (i.e. divided by $\mu = \mu_1 + \mu_2 + \ldots + \mu_n$);
the characteristic function for the jumps is given by $\chi(u) = \sum_k e^{iux_k} \mu_k / \mu$, that of the process by

$$\phi^t(u) = \exp\left\{t \sum_k \mu_k \left(e^{iux_k} - 1\right)\right\}, \tag{8.12}$$

which, as is obvious, can also be obtained as the product of the characteristic functions of the superimposed simple processes; that is of the $\exp\left\{t\mu_k\left(e^{iux_k} - 1\right)\right\}$.

Equation 8.10 has the same interpretation in the case of an arbitrary compound Poisson process: it reveals it to be a mixture of simple processes, but no longer necessarily a finite mixture.

Finally, we observe that the prevision $\mathbf{P}[Y(t)]$, and the variance $\boldsymbol{\sigma}^2[Y(t)]$, both exist and are determined by the distribution (both in the strict sense, and in terms of $\hat{\mathbf{P}}$; see Chapter 6, 6.5.7), so long as the same holds true for the jumps; that is if $\mathbf{P}(X)$ and $\boldsymbol{\sigma}^2(X)$, respectively, exist (in the same sense). In this case, we have

$$\mathbf{P}\left[Y(t)\right] = \mu t \mathbf{P}(X), \qquad \sigma^2\left[Y(t)\right] = \mu t \sigma^2(X)$$

(where $\mathbf{P}$ can be replaced by $\hat{\mathbf{P}}$, provided we do so on both sides).

If the prevision makes sense, it also makes sense to consider the process minus the prevision; in other words, modified by subtracting the certain linear function $\mu t\mathbf{P}(X)$, so that we obtain a process with zero prevision (i.e. a finite process). In other words, one considers $Y(t) - \mu t\mathbf{P}(X)$, which is the amount by which $Y(t)$ exceeds the prevision, as in the simple case of equation 8.6. The characteristic function is also similar to that of the latter case, and has the form

$$\phi^t(u) = \exp\left\{\mu t \int (e^{iux} - 1 - iux) dF(x)\right\} = \exp\left\{t \int \left(e^{iux} - 1 - iux\right) dM(x)\right\}. \tag{8.13}$$

8.2.8. We are now in a position to characterize the most general form of homogeneous process with independent increments; that is to say, the most general infinitely divisible distribution (see Chapter 6, 6.11.3 and 6.12). In fact, in either formulation it is a question of characterizing the characteristic functions $\phi(u)$ for which $[\phi(u)]^t$ turns out to be a characteristic function for any arbitrary $t > 0$.[6]

We have already encountered an enormous class of infinitely divisible characteristic functions; those of the form $\phi(u) = \exp\{a[\chi(u) - 1]\}$, where $\chi(u)$ is a characteristic function

6 For any $t < 0$, this is impossible (except in the degenerate case $\phi(u) = e^{iua}$, in which $|\phi(u)| = 1$; in other cases, for some u we have $|\phi(u)| < 1$, and then, for $t < 0$, we would have $|\phi(u)|^t > 1$). Moreover, it is sufficient to verify the condition for the sequence $t = 1/n$ (or any other sequence tending to zero), rather than for all t. In fact, it holds for all multiples, and hence for an everywhere dense set of values; by the continuity property of Chapter 6, 6.10.3, it therefore holds for all $t > 0$.

(the $\phi^t(u)$ of the compound Poisson processes). A function that is a limit of characteristic functions of this kind (in the sense of uniform convergence in any bounded interval) is again an infinitely divisible characteristic function. To see this, note that the limit of characteristic functions is again a characteristic function, and if some sequence of $\phi_n(u)$, such that $\phi_n \to \phi$, are infinitely divisible, then ϕ_n^t is a characteristic function, $\phi_n^t \to \phi^t$, and hence ϕ^t is a characteristic function (for every $t > 0$); in other words, ϕ is infinitely divisible. Conversely, it can be shown that an infinitely divisible characteristic function is necessarily either of the 'compound Poisson' form, or is a limit case; that is *the set of infinitely divisible characteristic functions coincides with the closure of the set of characteristic functions of the form* $\phi(u) = \exp\{a[\chi(u) - 1]\}$, *where* $\chi(u)$ *is a characteristic function.*

In order to prove this, it is sufficient to observe that if $[\phi(u)]^t$ is a characteristic function for every t, then $\phi_n(u) = \exp\{n[[\phi(u)]^{1/n} - 1]\}$ is also a characteristic function of the compound Poisson type, and tends to $\phi(u)$ (in fact, we are dealing with the well-known elementary limit $n(x^{1/n} - 1) \to \log x$). The process $\phi^t(u)$ is thus approximated by means of the processes $\phi_n^t(u)$ having, perhaps only apparently (see the *Remark* in Section 8.2.7), intensities $\mu_n = n$, and jump distributions $\chi_n(u) = [\phi(u)]^{1/n}$. More precisely, this 'apparently' holds in the cases we have already considered (the compound Poisson processes) and *actually* holds in the new limit cases which we are trying to characterize. In fact, in the compound Poisson cases, with finite intensity μ, the probability

$$p_t(0) = \mathbf{P}\left[Y(t) = 0\right] = \lim(1/2a)\int_{-a}^{a} \phi^t(u)\mathrm{d}u \qquad \left(\text{as } a \to \infty\right)$$

(the mass concentrated at 0 in the distribution having characteristic function $\phi^t(u)$) is $\geqslant e^{-\mu t}$ (which is the probability of no jump occurring before time t).[7] In this case, all the $\chi_n(u) = \phi^{1/n}(u)$ contain a constant term at least equal to $e^{-\mu/n}$ (corresponding to the mass at 0) and the actual intensity, instead of being $\mu_n = n$, is at most $n(1 - e^{-\mu/n}) \sim \mu$ (and it is easily verified that it actually tends to μ, as we might have guessed).

The new cases arise, therefore, when $Y(t) = 0$ has zero probability for every $t > 0$, no matter how small; that is when there is zero probability of $Y(t)$ remaining unchanged during any finite time interval, however small. We must have either a continuous variation, or a variation whose jumps are everywhere dense; that is with infinite intensity. In the approximation we have considered, the μ_n will all actually be equal to n.

These remarks and the treatment to follow are rather informal. We shall subsequently often have occasion to dwell somewhat more closely on certain critical aspects of the problem, but for the more rigorous mathematical developments we shall refer the reader to other works (for example, Feller, Vol. II, Chapter XVII, Section 2).

---

7 We have $p_t(0) > e^{-\mu t}$ if and only if there are jump values having nonzero probabilities (concentrated masses in the distribution whose characteristic function is $\chi(u)$) and some sum of them is 0. For example, in the case of Heads and Tails, values ± 1, we have $1 + (-1) = 0$ (i.e. we can return to 0 after two jumps). If, in this same example, the gains had been fixed at +2 and −3, then it would be possible to return to 0 after 5 jumps $(2 + 2 + 2 - 3 - 3 = 0)$ etc. In general, $p_t(0) = \sum_{h=2}^{\infty} \mathbf{P}[N(t) = h]\mathbf{P}(X_1 + X_2 + \ldots + X_h = 0)$, where $\mathbf{P}[N(t) = h) = e^{-\mu t}(\mu t)^h/h!$, and $\mathbf{P}(X_1 + X_2 + \ldots + X_h = 0)$ is the mass concentrated at 0 in the distribution whose characteristic function is $[\chi(u)]^h$.

8.2.9. As a first step in getting to grips with the general case, let us begin by extending to this case the considerations concerning the distribution of the intensity of the jumps, $M(x)$, as defined in Section 8.2.7 for the compound Poisson process. A definition which (making the previous considerations more precise) would be equivalent and which is also directly applicable to the general case is the following: $M(x)$ (taken with a plus or minus sign opposite to that of x) is the prevision of the number of increments having the same sign as x, and greater than x in absolute value, that occur in a unit time interval (subdivided into a large number of very small time intervals). More simply, and more concretely, we shall restrict ourselves to considering the subdivision into n small subintervals, each of length $1/n$, subsequently passing to the limit as $n \to \infty$.

The increment of $Y(t)$ in any one of these subintervals,

$$Y\left(t+1/n\right)-Y\left(t\right),$$

has distribution function $F^{1/n}(y)$ (see Section 8.1.3). The probability that it is greater than some positive x is $1 - F^{1/n}(x)$ and the prevision of the number of increments greater than x is $n[1 - F^{1/n}(x)]$, or, if one prefers, $[1 - F^t(x)]/t$. Similarly, for increments 'exceeding' some negative x (i.e. negative, and greater in absolute value), the probability and prevision are given by $F^t(x)$ and $F^t(x)/t$ ($t = 1/n$). We define $M(x)$ as the limit (as $t = 1/n \to 0$) of $-[1 - F^t(x)]/t$ for positive x, and of $F^t(x)/t$ for negative x. Assuming (as is, in fact, the case) that these limits exist, we can say, to a first approximation, that, for $t \sim 0$, we have $F^t(x) = 1 + tM(x)$ (for $x > 0$) and $F^t(x) = tM(x)$ (for $x < 0$). In other words (in a unified form),

$$F^t\left(x\right)=F^0\left(x\right)+tM\left(x\right),$$

where $F^0(x)$ (the limit case for $t = 0$) represents the distribution concentrated at the origin ($F^0(x) = 0$ for $x < 0$, and $= 1$ for $x > 0$).

This agrees intuitively with the idea that $M(x)$ is the intensity of the jumps 'exceeding' x, and, in particular, in the compound Poisson case, with $M(x) = \mu[F(x) - F^0(x)]$. In the general case, the meaning is the same, except that $M(-0)$ and $M(+0)$ can become infinitely large (either $M(-0) = +\infty$, or $M(+0) = -\infty$, or both), as shown in Figure 8.2.

The passage to the limit, which enables one to obtain the generalized Poisson processes, thus reduces to the construction of the $\phi^t(u)$ on the basis of formulae 8.10 and 8.13 of Section 8.2.7, allowing the function $M(x)$ to become infinite as $x \to \pm 0$, along with appropriate restrictions to ensure that the function converges, and that the process it represents makes sense (but we limit ourselves here to simply indicating how this can be done, and that it is, in fact, possible).

8.2.10. A new form, intermediate between the two previous forms in so far as it provides compensation only for the small jumps, proves more suitable as a basis for a unified account. This is defined by equation 8.14 below, and has to be constructed (within largely arbitrary limits) in such a way that it turns out to be equivalent to equation 8.13 in the neighbourhood of $x = \pm 0$, and to equation 8.10 in the neighbourhood of $x = \pm\infty$. We consider

$$\phi^t\left(u\right)=\exp\left\{t\int\left[\mathrm{e}^{\mathrm{i}ux}-1-\mathrm{i}ux.\tau\left(x\right)\right]\mathrm{d}M\left(x\right)\right\}, \tag{8.14}$$

where $\tau(x)$ is an arbitrary bounded function, tending to 1 as $x \to 0$, and to 0 as $x \to \pm\infty$.

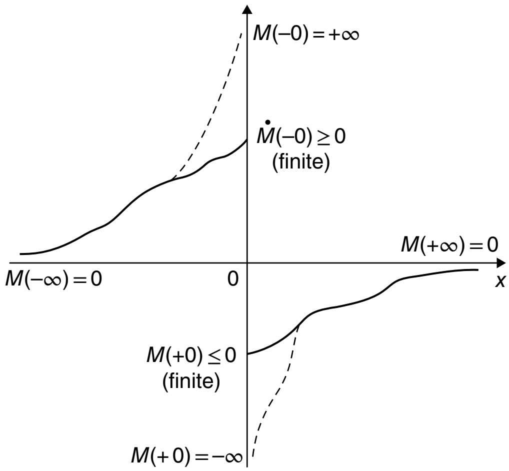


Figure 8.2 Distribution of the intensity of the jumps.

Possible choices are :

$$\tau(x)=(|x|<1)\quad \big(\text{i.e. } 1 \text{ in}[-1,1], \text{ and } 0 \text{ else where; P.Lévy}\big),$$

or

$$\tau(x)=1/\left(1+x^2\right)\quad (\text{Khintchin}),$$

or

$$\tau(x)=\sin x/x\quad (\text{Feller}).^8$$

A necessary and sufficient condition for the expression 8.14 (with any $\tau(x)$ whatsoever) to make sense as the characteristic function of a random process – and, in this way, to provide all the infinitely divisible distributions, apart from the normal, which derives from it as a limit case – is that the contribution to the variance from the 'small jumps' be finite. In other words, we must have $\int x^2 \mathrm{d}M(x)<\infty$ (the integral being taken, for example, over [−1, 1]; the actual interval does not matter, provided it is finite, and contains the origin).[9] We note that, in the more regular case in which the intensity admits a density $M'(x)$, and in which it makes sense to speak of an 'order of infinity' as it tends to 0 (from the left or right), the necessary and sufficient condition is that this order of infinity (from both sides) be <2: if

$$M'(x)\sim 1/x^{\alpha}\ (\alpha<2),$$

things go through; if $M'(x) \sim 1/x^2$, they no longer do.

---

8 Some authors prefer, in place of $\mathrm{d}M(x)$ as the differential element in the integral, to adopt variants like $\mathrm{d}K(x) = [x/(1 + x^2)]\mathrm{d}M(x)$ (Khintchin), or $\mathrm{d}H(x) = x\,\mathrm{d}M(x)$ (Feller). These have distinct formal advantages, but do not seem to me to compensate for the loss of direct meaning (see Feller, Vol. II, p. 536 *et passim*, and P. Lévy (1965), p. 141).

9 See P. Lévy and Feller, *loc. cit.*

Having said this, it is now easy to make precise the circumstances under which, and the reasons why, the expression in equation 8.14 can be replaced by one or other of the two simpler forms given previously. These can be considered as special cases of equation 8.14, in which $\tau(x)$ (instead of satisfying the imposed conditions) is set equal to 0 for equation 8.10 (the term iux is omitted), and equal to 1 for equation 8.13 (the term iux is always present).

The term iux is innocuous (it merely produces an addition to $Y(t)$ of a certain linear function ct) so long as it is applied to jumps that are neither too large nor too small (e.g. for $\varepsilon < |x| < 1/\varepsilon$, with $\varepsilon > 0$ arbitrarily small). When applied in a neighbourhood of $x = 0$, it is either innocuous or *useful;* when applied in a neighbourhood of ∞ (e.g. for $1/|x| < \varepsilon$), it is either innocuous or *harmful*. It can be useful, and indeed *necessary*, when the *small jumps*, if not 'compensated', do not lead to convergence; this happens if $\int |x| \, \mathrm{d}M(x)$ diverges in $[-1, 1]$ (or, equivalently, in any neighbourhood of 0). It can be harmful, because the 'compensation' of the '*big jumps*' destroys convergence if they give too large a contribution, and this happens precisely when the previous integral diverges over $|x| \geqslant 1$. Observe, also, that this may very well happen even in a compound Poisson process (μ finite); it is enough that the distribution of the jumps should fail to have a 'mean value' (as, for example, with the Cauchy distribution). This does not affect the process, but any attempt to pass to the limit for the 'compensated' jumps would destroy the convergence rather than assist it.

In conclusion: the condition $\int x^2 \mathrm{d}M(x) < \infty$ over $[-1, 1]$ is necessary and sufficient, and expression 8.14 always holds if the condition is satisfied. Both of the simple forms equations 8.10 and 8.13 can be applied if $\int |x| \mathrm{d}M(x) < \infty$ over $[-\infty, +\infty]$ (and we observe that this condition implies the general condition and is, therefore, itself sufficient). If, instead, the integral diverges, it is necessary to distinguish whether this is due to contributions in the neighbourhood of the origin, or in the neighbourhood of $\pm\infty$; in the first case equation 8.10 is ruled out, and in the second equation 8.13 is ruled out (and both are if there is trouble both at the origin and at infinity).

8.3 The Wiener–Lévy Process

8.3.1. We now turn to an examination of the continuous component of a homogeneous random process with independent increments, which we described briefly in Section 8.1.3: this is the Wiener–Lévy process. The points already made, together with some further observations, will suffice here to provide a preliminary understanding of the process and will be all that is required for the discussion to follow.

Let us first make clear what is meant by calling the process 'continuous'. It amounts to saying that, for any preassigned $\varepsilon > 0$, if we consider the increments of $Y(t)$ in $0 \leqslant t \leqslant 1$, divided up into N equal intervals, the probability that even one of the increments is, in absolute value, greater than ε, that is the possibility that $|Y(t + 1/N) - Y(t)| > \varepsilon$ for some $t = h/N < 1$, tends to 0 as N increases. In short, we are dealing with 'what escapes the sieve laid down for the selection of the jumps'. There is no harm in thinking (on a superficial level) of this as being equivalent to the continuity of the function $Y(t)$. From a conceptual point of view, however, this would be a distortion of the situation, as can be seen from the few critical comments we have already made, and from the others we shall be presenting later, albeit rather concisely, in a more systematic form (see, for example, Section 8.9.1; especially the final paragraph).

We shall usually deal with the *standardized* Wiener–Lévy process, having zero prevision, $m(t) = 0$, unit variance per unit time period, $\sigma^2(t) = t$, and initial condition $Y(0) = 0$.

In this case, the density $f^t(y)$ and the characteristic function $\phi^t(u)$ (both at time t) are given by

$$f^t(y) = K_t^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}y^2/t} \quad \left(K = 1/\sqrt{(2\pi)}\right), \tag{8.15}$$

$$\phi^t(u) = \mathrm{e}^{-\frac{1}{2}tu^2}. \tag{8.16}$$

The general case ($Y(0) = y_0$, $m(t) = mt$, $\sigma^2(t) = t\sigma^2$) can be reduced to the standardized form by noting that it can be written as $a + mt + \sigma Y(t)$, where $Y(t)$ is in standardized form. If we wish to consider it explicitly in the general form, we have

$$f^t(y)Kt^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left[(y-y_0-mt)/\sigma\right]_{2/t}} \quad \left(K = 1/\sqrt{(2\pi)}\sigma\right),$$

$$\phi^t(u) = \mathrm{e}^{\mathrm{i}uy_0}.\mathrm{e}^{t\left(\mathrm{i}um-\frac{1}{2}u^2\sigma^2\right)}$$

(where, for greater clarity, the terms depending on the initial value, y_0, and those depending on the process, i.e. on t, are written separately).

8.3.2. The same approach holds good when we wish to examine the random process and its behaviour, rather than just isolated values assumed by it. In fact, the joint distribution of the values of $Y(t)$ at an arbitrary number of instants $t_1, t_2 \ldots, t_n$, is also a normal distribution, with density![10] given by

$$f\left(y_1, y_2, \ldots, y_n\right) = K \quad \mathrm{e}^{-\frac{1}{2}Q\left(y_1, y_2, \ldots, y_n\right)},$$

where Q is a positive definite quadratic form determined by the covariances $\mathbf{P}[Y(t_i)Y(t_j)] = \sigma_i\sigma_j r_{ij}$ (if $i = j$, $r_{ii} = 1$, and covariance = variance = σ_i^2). We have already seen (in Section 8.1.6) that (if $t_i \leqslant t_j$) the covariance is t_i^2, and therefore $r_{ij} = \sqrt{(t_i/t_j)}$; this gives all the information required for any application.

It is simpler and more practical, however, to observe that all the $Y(t_i)$ can be expressed in terms of increments, $\Delta_i = Y(t_i) - Y(t_{i-1})$, which follow consecutively and are independent: $Y(t_i) = \Delta_1 + \Delta_2 + \ldots + \Delta_i$. But Δ_i has a centred normal distribution, with variance $(t_i - t_{i-1})$, and Q, as a function of the variables $(y_i - y_{i-1})$, is a sum of squares:

$$Q\left(y_1, y_2, \ldots, y_n\right) = \sum_i \left[\left(y_i - y_{i-1}\right)^2 / \left(t_i - t_{i-1}\right)\right]. \tag{8.17}$$

We shall now make use of this to draw certain conclusions, which we shall require in what follows.

10 We assume that $m = 0$, $\sigma = 1$ (the standardized case); only trivial modifications are required for the general case.

What we have been considering so far, to be absolutely precise, is the Wiener–Lévy process on the half-line $t \geqslant 0$, given that $Y(0) = 0$; the case of $t \geqslant t_0$, given that $Y(t_0) = y_0$, is identical (and the same is true for $t \leqslant 0$, $t \leqslant t_0$, respectively).[11]

In order to consider the case in which several values are given (at $t = t_1, t_2, \ldots$), it is sufficient to consider the problem inside one of the (finite) intervals; on the unbounded intervals things proceed as above. Let us, therefore, consider the process over the interval (t_1, t_2), given the values $Y(t_1) = y_1$ and $Y(t_2) = y_2$ at the end-points. In order to characterize it completely, it will suffice, in this case also, to determine the prevision (no longer necessarily zero!) and the variance of $Y(t)$ for each $t_1 \leqslant t \leqslant t_2$, and the covariance (or correlation coefficient) between $Y(t')$ and $Y(t'')$ for each pair of instants $t_1 \leqslant t' \leqslant t'' \leqslant t_2$.

We now decompose $Y(t)$ into the sum of a certain linear component (a straight line through the two given points) and the deviation from it:

$$Y(t) = y_1 + \left[(t - t_1)/(t_2 - t_1)\right](y_2 - y_1) + Y_0(t),$$

where $Y_0(t)$ corresponds to the same problem with $y_1 = y_2 = 0$. We now consider the two components as if the end-points were not yet fixed, so that the increments $\Delta_1 = Y - Y_1$ and $\Delta_2 = Y_2 - Y$ are independent, with standard deviations $\sigma_1 = \sqrt{(t - t_1)}$ and $\sigma_2 = \sqrt{(t_2 - t)}$. The linear component is then the random quantity

$$\left[1/(t_2 - t_1)\right]\left[(t_2 - t)Y_1 + (t - t_1)Y_2\right],$$

and, by subtraction, $Y_0(t)$ is given by

$$Y_0(t) = \left[1/(t_2 - t_1)\right]\left[(t_2 - t)\Delta_1 - (t - t_1)\Delta_2\right].$$

It is easily seen that $Y_0(t)$ has zero prevision (as was obvious) and standard deviation given by

$$\sigma(Y_0(t)) = \sqrt{\left[(t - t_1)(t_2 - t)/(t_2 - t_1)\right]}; \tag{8.18}$$

moreover, it is uncorrelated with the linear component

$$Y_1 + \left[(t - t_1)/(t_2 - t_1)\right](\Delta_1 + \Delta_2)$$

(and hence, by normality, they are independent). In fact, we have

$$\begin{aligned} \text{covariance} &= K\left[(t - t_1)(t_2 - t)\mathbf{P}(\Delta_1^2) - (t - t_1)^2\,\mathbf{P}(\Delta_2^2)\right] \\ &= K\left[(t_2 - t)\sigma_1^2 - (t - t_1)\sigma_2^2\right] = 0. \end{aligned}$$

It can be seen (as a check, and in order to realize the difference between this and $\sqrt{(t - t_1)}$, which would apply in the absence of the condition on the second end-point) that the

11 Provided that the process is assumed to make sense, even in the past, and that, in fact, no knowledge of the past leads us to adopt different previsions (these assumptions are not, in general, very realistic).

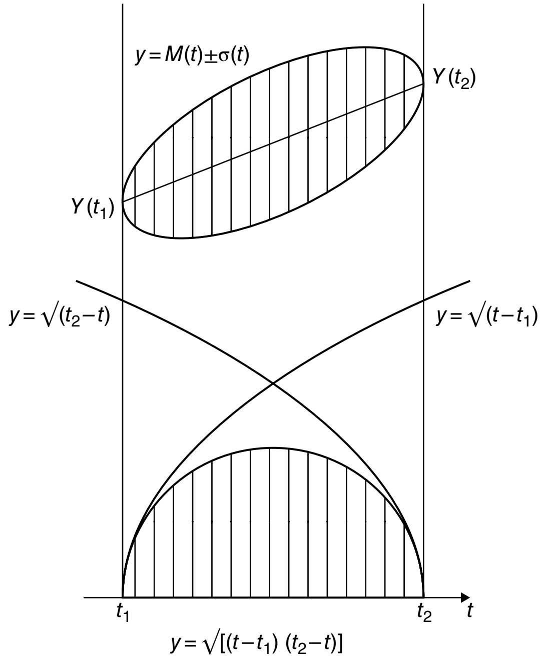


Figure 8.3 Interpolation between two known points in the Wiener–Lévy process (straight line and semi-ellipses: graphs of *prevision* and *prevision* ± *standard deviation*). The lower diagram represents the behaviour of the standard deviation given the point of origin, the final point, or both.

standard deviation of the linear component (considering the value of the first end-point, Y_1, as fixed) is given by

$$\left[\left(t-t_1\right)/\left(t_2-t_1\right)\right]\sqrt{\left[\mathbf{P}\left(\Delta_1+\Delta_2\right)^2\right]}=\sqrt{\left(t-t_1\right)}.\sqrt{\left[\left(t-t_1\right)/\left(t_2-t_1\right)\right]}.$$

Summing the squares of the standard deviations of the two summands, we obtain the square of $\sigma_1 = \sqrt{(t - t_1)}$ (as, indeed, we should). It is useful to note that, as is shown in Figure 8.3, the standard deviation (equation 8.18) of $Y(t)$ given the values at t_1 and t_2, is represented by the (semi) circle resting on the segment (t_1, t_2) (provided the appropriate scale is used; i.e. taking the segment $t_2 - t_1$ on the t-axis equal to the unit of measure on the y-axis: on the other hand, this is really irrelevant, except as an aid to graphical representation and description). If we consider the parabolae that represent, in a similar fashion, $\sigma(t)$ given only Y_1 (i.e. $y = \sqrt{(t - t_1)}$), or given only Y_2 (i.e. $y = \sqrt{(t_2 - t)}$), we see that the product of these two functions is represented by the circle, which, therefore, touches the parabolae (at the end-points), because when one of the two factors vanishes, the other has value 1.

We can determine, in a similar manner, the covariance between

$$Y\left(t'\right)\text{and }Y\left(t''\right),\quad t_1 \le t' \le t'' \le t_2.$$

We denote the successive, independent increments by

$$\Delta_1 = Y\left(t'\right)-Y\left(t_1\right),\ \ \Delta_2 = Y\left(t''\right)-Y\left(t'\right),\ \ \Delta_3 = Y\left(t_2\right)-Y\left(t''\right);$$

$Y_0(t)$ will be the same as before, but writing $y_2 - = y_1 = \Delta_1 + \Delta_2 + \Delta_3$, $T = t_2 - t_1$ and assuming, simply for notational convenience, that $t_1 = 0$ and $y_1 = 0$, we have

$$Y_0(t') = \Delta_1 \qquad - t'(\Delta_1 + \Delta_2 + \Delta_3) = (T - t')\Delta_1 \qquad - t'\Delta_2 - t'\Delta_3,$$
$$Y_0(t'') = \Delta_1 + \Delta_2 - t''(\Delta_1 + \Delta_2 + \Delta_3) = (T - t'')\Delta_1 + (T - t'')\Delta_2 - t''\Delta_3,$$

and hence

$$Y_0(t')Y_0(t'') = (T - t')(T - t'')\Delta_1^2 - t'(T - t'')\Delta_2^2 + t't''\Delta_3^2$$
$$+ \text{ cross-product terms (in } \Delta_i\Delta_j,\ i \neq j, \text{ with zero prevision).}$$

Taking prevision, and bearing in mind that the previsions of the Δ_i^2 are, respectively, $\sigma_1^2 = t'$, $\sigma_2^2 = t'' - t'$, $\sigma_3^2 = T - t''$ we have

$$\begin{aligned}\mathrm{Convar}(t',\ t'') &= (T - t')(T - t'')t' - t'(T - t'')(t'' - t') + t't''(T - t'') \\ &= t'(T - t'')\big[(T - t') - (t'' - t') + t''\big] = t'(T - t'')T.\end{aligned}$$

Dividing by $\sigma' = \sqrt{\left[t'(T - t')/T\right]}$ and $\sigma'' = \sqrt{\left[t''(T - t'')/T\right]}$, we obtain, finally, the correlation coefficient

$$r(t', t'') = \sqrt{\left[\frac{t'(T - t'')}{(T - t')t''}\right]} = \sqrt{\left[\frac{(t' - t_1)(t_2 - t'')}{(t_2 - t')(t'' - t_1)}\right]}, \tag{8.19}$$

and, in this way, we return to our original notation.

8.3.3. It will certainly be no surprise to find that if we put $t_2 = \infty$ we reduce to the results obtained in the case of the single condition at t_1. This is, however, simply a special case of a remarkable fact first brought to light by P. Lévy, and which is often useful for inverting this conclusion by reducing general cases (with two fixed values) to the special limit case (with only one fixed value). The fact referred to is the *projective invariance* of problems concerning processes of this kind,[12] and derives from the expression under the square root in $r(t', t'')$ being the cross-ratio of the four instants involved. It therefore remains invariant under any homographic substitution for the time t, provided the substitution does not make the *finite* interval $[t_1, t_2)$ correspond to the complement of a finite interval (but instead, to either a finite interval, or to a half-line; in other words, the inequalities $t_1 \leqslant t' \leqslant t'' \leqslant t_2$ must either be all preserved, or all inverted). Consequently, the stochastic nature of the function $Y(t)$ remains invariant if we ignore multiplication by a certain arbitrary function; invariance holds for any random function of the form $Z(t) = g(t)Y(t)$, or, in particular, for the standardized process (always having $\sigma = 1$) that can be obtained by taking

$$g(t) = 1/\sigma\big[Y(t)\big] = 1/\sqrt{t}. \tag{8.20}$$

12 The meaning of this is most easily understood by introducing the projective coordinate

$$\tau = \tau(t) = (t - t')(t'' - t_2)/(t'' - t)(t_2 - t');$$

i.e. (as is obvious) taking t', t_2 and t'' to 0, 1 and $\infty : \tau' = \tau(t') = 0, \tau_2 = \tau(t_2) = 1, \tau'' = \tau(t'') = \infty$.
It follows that $r = \sqrt{\tau_1}$, where $\tau_1 = \tau(t_1)$ is the abscissa of the point t_1 after the projective transformation has been performed.

This device will prove useful for, among other things, reducing the study of the asymptotic behaviour of $Y(t)$ in the neighbourhood of the origin (for $0 < t < \varepsilon$, $\varepsilon \to 0$) to that at infinity (for $t > T$, $T \to \infty$); see Section 8.9.7.

The basic properties of the Wiener–Lévy process will be derived later, in contexts where they correspond to actual problems of interest.

8.4 Stable Distributions and Other Important Examples

8.4.1. We have already encountered (in Chapter 6, 6.11.3) two families of stable distributions: the normal (which, in Chapter 7, 7.6.7, we saw to be the only stable distribution having finite variance) and the Cauchy (which has infinite variance).

We are now in a position to determine all stable distributions. It is clear – and we shall see this shortly – that they must be infinitely divisible. Our study can, therefore, be restricted to a consideration of distributions having this latter property and, since we know what the explicit form of their characteristic functions must be, this will not be a difficult task.

Knowledge of these new stable distributions will also prove useful for clarifying the various necessary conditions and other circumstances occurring in the study of the asymptotic behaviour of random processes.

We begin by observing that the convolution of two compound or generalized Poisson distributions, represented by the distributions of the intensities of the jumps, $M_1(x)$ and $M_2(x)$, is obtained by summing them: $M(x) = M_1(x) + M_2(x)$. In fact, $M(x)$ determines the logarithm of the characteristic function in a linear fashion, and the sum in this case corresponds to the product of the characteristic functions; that is to the convolution.

This makes clear, conversely, what the condition for an infinitely divisible distribution to be a factor in the decomposition of some other distribution (also infinitely divisible) must be. The distribution defined by $M_1(x)$ 'divides' the distribution defined by $M(x)$ if and only if the difference

$$M_2(x) = M(x) - M_1(x)$$

is also a distribution function of intensities. This implies that it must never decrease, so that every interval receives positive mass (or, at worst, zero mass), and this implies, simply and intuitively, that in any interval (of the positive or negative semi-axis) $M_1(x)$ must have an increment not exceeding that of $M(x)$. In particular, the concentrated masses and the density (if they exist) must, at every individual point, not exceed those of $M(x)$. If one wishes to include in the statement the case in which a normal component exists (and then we have the most general infinitely divisible distribution) it is sufficient to state that here, too, this component must be the smaller (as a measure, one can take the variance).[13]

In order to prove that stability implies infinite divisibility, it is sufficient to observe that, in the case of stability, the sum of n independent random quantities that are

13 Note that what we have said concerns divisibility within the class of infinitely divisible distributions. However, there may exist indivisible factors of infinitely divisible distributions (and, of course, conversely), as we mentioned already in Chapter 6, 6.12.

identically distributed has again the same distribution (up to a change of scale). It is itself, therefore, a convolution product of an arbitrary number, n, of identical factors, and is therefore infinitely divisible. If, for each factor, the distribution of the intensities of the jumps is $M(x)$, then for the convolution of n factors it is $nM(x)$.

For stability it is necessary and sufficient that the distribution defined by $nM(x)$ belongs to the same family as that defined by $M(x)$. This requires that it differs only by a (positive) scale factor $\lambda(n)$: that is $nM(x) = M(\lambda(n)x)$. It follows immediately that

$$knM(x) = kM(\lambda(n)x) = M(\lambda(k)\lambda(n)x) = M(\lambda(kn)x);$$

in other words, $\lambda(k)\lambda(n) = \lambda(kn)$ for k and n integer. The same relation also holds for rationals if we set $\lambda(1/n) = 1/\lambda(n)$ and hence $\lambda(m/n) = \lambda(m)/\lambda(n)$. By continuity, we then have $\lambda(\nu)$ for all positive reals ν. The functional equation $\lambda(\nu_1)\lambda(\nu_2) = \lambda(\nu_1\,\nu_2)$ characterizes powers, so we have an explicit expression for λ:

$$\lambda(\nu) = \nu^{-1/\alpha}; \qquad \text{in particular}, \lambda(n) = n^{-1/\alpha}. \tag{8.21}$$

We have written the exponent in the form $-1/\alpha$, because it is the reciprocal, $-\alpha$, which appears as the exponent in the expression for $M(x)$, and which will be of more direct use in what follows. It is for this reason that α is known as the 'characteristic exponent' of the distribution for which we have

$$nM(x) = M\left(n^{-1/\alpha}x\right) \quad \left(\text{and, in general, } \nu M(x) = M\left(\nu^{-1/\alpha}x\right), 0 < \nu < +\infty\right). \tag{8.22}$$

In fact, we can immediately obtain an explicit expression for $M(x)$. When $x = 1$, the above expression reduces to $\nu M(1) = M(\nu^{-1/\alpha})$, and when $x = \nu^{-1/\alpha}$, we have $M(x) = -Kx^{-\alpha}$, where $-K = M(1)$, a constant: this holds for all positive x (running from 0 to $+\infty$ as ν varies in the opposite direction from $+\infty$ to 0). Setting $x = -1$ (instead of $+1$), we can obtain the same result for negative x, except that we must now write $|x|^{-\alpha}$ in place of $x^{-\alpha}$. Allowing for the fact that the constant K could assume different values on the positive and negative semi-axes, we have, finally,

$$M(x) = -K^{+}|x|^{-\alpha}(x > 0) + K^{-}|x|^{-\alpha}(x < 0) = K^{\pm}|x|^{-\alpha}, \tag{8.23}$$

where K^{+} and K^{-} are positive, and are therefore written preceded by the appropriate sign (this ensures that $M(x)$ is increasing, in line with what we said in Section 8.2.9). It is obviously unnecessary to write $|x|$ when x is positive, but it serves to stress the identical nature of the expressions over the two semi-axes: $K^{\pm}$ is merely a compact way of writing either $-K^{+}$ or $+K^{-}$ for $x \gtrless 0$ (it could be written in the form $K^{\pm} = K^{-}(x < 0) - K^{+}(x > 0)$).

It remains to examine which values are admissible for the characteristic exponent, We see immediately that these are the α for which $0 < \alpha \leqslant 2$, and it turns out that there are good reasons for considering these as four separate subcases;

$$0 < \alpha < 1, \qquad \alpha = 1, \qquad 1 < \alpha < 2, \qquad \alpha = 2.$$

The value $\alpha = 1$ is a somewhat special case, and corresponds to the Cauchy distribution (we have already met this in Chapter 6, 6.11.3; the correspondence is established by examining the characteristic function given in equation 6.59 of the section mentioned).

8.4.2. For $\alpha = 2$, we cannot proceed in the above manner – the expression for the characteristic function diverges – but we can consider it as a limit case (or we could include it by using the kind of procedures mentioned in the footnote 8). The limit case turns out to be the by now familiar normal distribution.

In fact, this corresponds to the characteristic exponent

$$\alpha = 2 \qquad \left(\text{or}, 1/\alpha = \tfrac{1}{2}\right)$$

because the scale factor (in this case, the standard deviation) for the sum Y_n of n identically distributed summands is multiplied by $\sqrt{n}$, that is $n^{½}$ (and hence, for the mean Y_n/n we multiply by $n^{-½}$).

More generally, even in the case of distributions from the same family but with different scale factors, the well-known formula for the standard deviation (for sums of independent quantities) holds for the normal distribution,

$$\sigma = \left(\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2\right)^{\frac{1}{2}}, \tag{8.24}$$

and also for all stable distributions if adapted to the appropriate characteristic exponent α. Explicitly, if $X_1, X_2, \ldots, X_n$ have distributions all belonging to the same family (stable, with characteristic exponent α), and, $a_1, a_2, \ldots, a_n$ are the respective scale factors, then the random quantity defined by the sum $aX = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ again has a distribution belonging to this family, with scale factor given by

$$a = \left(a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha\right)^{1/\alpha} \qquad \left(i.e. \quad a^\alpha = \sum a_i^\alpha\right). \tag{8.25}$$

In particular, if all the a_i are equal to 1,

$$a = n^{1/\alpha}. \tag{8.26}$$

This is an immediate consequence of the expression for $M(x)$ given in equation 8.23; setting $\nu = \nu_1 + \nu_2$ in $\nu M(x) = M(\nu^{-1/\alpha}x)$, we obtain

$$M\left(\nu^{-1/\alpha}x\right) = M\left(\nu_1^{-1/\alpha}x\right) + M\left(\nu_2^{-1/\alpha}x\right).$$

We began with the case $\alpha = 2$, not only because of its importance and familiarity but also because it enables us to establish immediately that values $\alpha > 2$ are not admissible. This is not only because, *a fortiori*, the integral would diverge; there is another, elementary, or, at least, familiar, reason (which we shall just mention briefly). If the variance is finite, the formula for standard deviations holds when $\alpha = 2$; if it is infinite, we must have $\alpha \leqslant 2$, because $\alpha = 2$ holds for any bounded portion of the distribution (considering, for example, the *truncated* X_h; $-K \vee X_h \wedge K$).

In connection with the idea of *compensation* (e.g. for errors of measurement), there is, by virtue of the (magic?) properties of the arithmetic mean, a point of some conceptual and practical importance which is worth making. (Of course, we are dealing with a mathematical property which we would have had to mention anyway, in order to deal with an important aspect of the behaviour of the means Y_n/n, of n summands, each of which follows a stable distribution with some exponent α.)

The form given in equation 8.26 asserts that, compared with the individual summands, Y_n has scale factor $n^{1/\alpha}$; it follows, therefore, that, for the arithmetic mean Y_n/n, the scale factor is $n^{(1/\alpha)-1}$.

Taking, for example, the factor is	$\alpha =$	2	3/2	4/3	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
	$=$	$n^{-\frac{1}{2}}$	$n^{-\frac{1}{3}}$	$n^{-\frac{1}{4}}$	1	$n^{\frac{1}{4}}$	n	n^3
giving, forexample, for	$n=2$	0·707	0·793	0·841	1	1·189	2	8
	$n=10$	0·316	0·464	0·681	1	1·779	10	1000

The usual amount of 'compensation' (i.e. an increase in precision in the ratio 1: $\sqrt{n}$ for the mean of n values) is only attained for $\alpha = 2$; the increase diminishes as α moves from 2 to 1, and for $\alpha = 1$ (the Cauchy distribution) the precision is unaltered, implying no advantage (or disadvantage) in using a mean based on several values rather than just using a single value; for $\alpha < 1$, the situation is reversed and worsens very rapidly as we approach zero (a limit value which must be excluded since it would give ∞!). The values given above should be sufficient to provide a concrete numerical feeling for the situation.

One should not conclude, however, that in these latter cases there is no advantage in having more information (this would be a narrow, short-sighted mistake, resulting from the assumption that the only way to utilize repeated observations is by forming their mean). There is always an advantage in having more observations (there is more information!) but, to make the most of it, it is necessary to pose the problem properly, in a form corresponding to the actual circumstances. This kind of problem of mathematical statistics is dealt with by the Bayesian formulation given in Chapters 11 and 12.

8.4.3. For $\alpha < 2$, we have, in fact, a generalized Poisson distribution, with $M(x)$ corresponding to the distribution function of the intensities of the jumps. Moreover, since $M(x) = K^{+}.|x|^{-\alpha}$, the density exists and is given by

$$M'(x) = \alpha\left|K^{\pm}\right|\left|x\right|^{-(\alpha+1)}. \tag{8.27}$$

It is simpler and clearer (apart from an exceptional case which arises for $\alpha = 1$) to consider separately the distribution generated by the positive jumps (the other is symmetric). We then have, taking $K^{+} = -1/\alpha$ in order to obtain the simplest form of the density,

$$M(x) = -(1/\alpha)x^{-\alpha}, \quad M'(x) = x^{-(\alpha+1)}, \quad \mathrm{d}M(x) = \mathrm{d}x/x^{\alpha+1}. \tag{8.27$'$}$$

It is here that we meet the circumstance which forces us to distinguish between the two cases of α less than or greater than 1 (the case $\alpha = 1$ will appear later on). The fact is that in the first case we have convergence without requiring the correction term iux in equation 8.13, whereas in the second case this term is required. The reason for this is (roughly speaking) that $\mathrm{e}^{\mathrm{i}ux} - 1 \sim \mathrm{i}ux$ is an infinitesimal of the first order in x for $x \sim 0$; if multiplied by $\mathrm{d}M(x) = \mathrm{d}x/x^{\alpha+1}$, it gives $\mathrm{d}x/x^{\alpha}$, and the integral converges or diverges as $x \to 0$, according to whether $\alpha < 1$ or $\alpha > 1$. This is not merely a question of analysis, however; there is a point of substance involved here. For $\alpha < 1$, the generalized Poisson

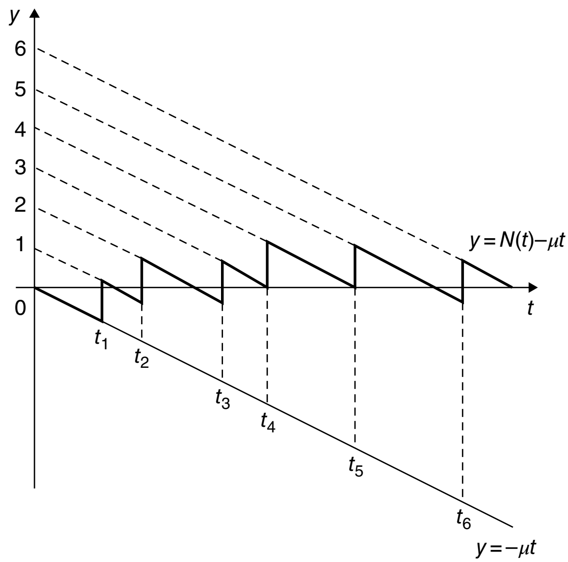


Figure 8.4 The simple, compensated Poisson process (prevision = 0).

random process produced by the positive jumps only, with distribution of intensities $M(x) = Kx^{-\alpha}$ (which is therefore always increasing), makes sense. For $\alpha > 1$, however, only the *compensated sum* of the jumps makes sense. With reference to Figures 8.4 and 8.5,[14] we can give some idea of this behaviour by saying that (as more and more of the numerous very small jumps are added) the sum of the jumps (per unit time) becomes infinite, but the sloping straight line from which we start also becomes infinitely inclined downwards. The process, under these conditions, cannot have monotonic behaviour (in any time interval; no matter how short).

8.4.4. Notwithstanding the diversity of their behaviour, both mathematically and in terms of the actual processes, there are no differences in the form of the characteristic functions. Simple qualitative considerations will suffice to establish that they must have the form $\exp(C|u|^\alpha)$ (where C is replaced by its complex conjugate, C^*, if u is negative). Detailed calculation (see, for example, **P**. Lévy (1965), p. 163) shows that in the case of positive jumps we have

$$\phi(u) = \exp\left\{-e^{\pm\frac{1}{2}i\pi\alpha}|u|^\alpha\right\}, \tag{8.28}$$

14 Figure 8.4 shows a simple Poisson process with its prevision subtracted off (see Section 8.2.6). Figure 8.5 shows what happens in compound processes obtained by superposing, successively, on top of the previous one, simple processes having, in each case, a smaller jump. If one imagines, following on from the three shown, a fourth step, a fifth step, …, and so on, with the slope of the straight line of the prevision increasing indefinitely, one gets some idea of the generalized Poisson processes with only positive jumps.

Note that, in order not to make the diagram too complicated, it has been drawn as if all the additional processes vanish at the instants of the preceding jumps (this is very unlikely, but does not alter the accuracy of the visual impression; we merely wish to warn those who realize that this device has been adopted not to imagine that it reflects some actual property of the processes in question).

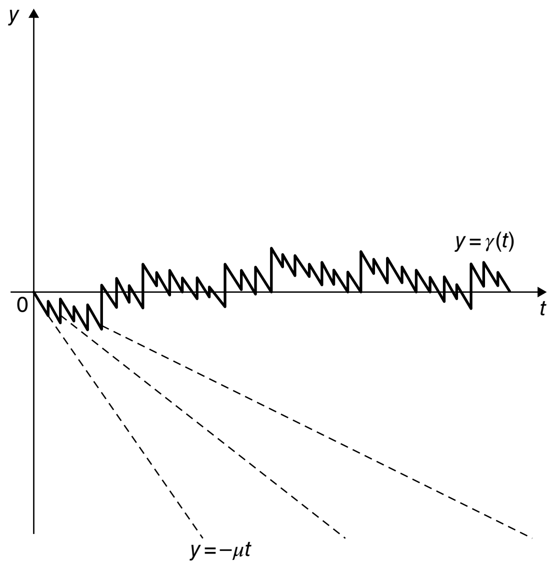


Figure 8.5 The compound Poisson process with successive sums from simple, compensated Poisson processes.

with $\pm$, depending on the sign of u, whereas for negative jumps the $\mp$ signs in the exponent are interchanged. In the general case, it is sufficient to replace α or $-\alpha$ in the exponent by an intermediate value; in particular, in the symmetric case, by 0.

For $\alpha = 1$, the symmetric case gives the Cauchy distribution. The latter can, therefore, be thought of as generated by a generalized Poisson process in which the distribution of the intensities of the jumps is given by

$$M(x) = \pm 1/x,$$

with density $M'(x) = 1/x^2$. In this case, however, we can only ensure convergence by having recourse to the form given in equation 8.14 (this is because the term iux is necessary in the neighbourhood of the origin, but gives trouble at infinity). By doing this, however, we effect a *partial* compensation in the jumps and this reintroduces a certain, arbitrary, additive constant, which prevents the distribution from being stable (following Lévy, we could term it *quasi-stable*; the convolution involves not merely a change of scale, but also a translation). In the symmetric case, we obtain stability by using the same criterion of compensation for the contributions of both negative and positive jumps (or by implicitly compensating, by first integrating between $\pm a$ and then letting $a \to \infty$).

Apart from the two cases $\alpha = 2$ (normal) and $\alpha = 1$ (Cauchy), the densities of stable distributions cannot, in general, be expressed in simple forms (although they exist, and are regular). One exception is the case $\alpha = \frac{1}{2}$, to which corresponds, as an increasing process (positive jumps; $x > 0$),

$$M(x) = -2x^{-\frac{1}{2}}, \quad M'(x) = x^{-\frac{3}{2}}, \quad \mathrm{d}M(x) = \mathrm{d}x / \sqrt{x^3},$$

and the density

$$f(x) = Kx^{-\frac{3}{2}}e^{-1/2x}. \tag{8.29}$$

Finally, we should also mention the case $\alpha = \frac{3}{2}$, which is important on account of an interpretation of it given by Holtsmark in connection with a problem in astronomy (and by virtue of the fact that it precedes knowledge of the problem on the part of mathematicians): the case $\alpha = \frac{4}{3}$ can be given an analogous interpretation (in a four-dimensional world). (See Feller, Vol. II, pp. 170, 215.)

8.4.5. Other important examples of Poisson-type processes are the *gamma* processes (and those derived from them), and those of *Bessel* and *Pascal*.

The gamma distribution (see equations 6.55 and 6.56 in Chapter 6, 6.11.3) has density and characteristic function defined by

$$f^t(x) = Kx^{t-1}e^{-x} \quad (x \geqslant 0), \quad K = \frac{1}{\Gamma(t)}, \tag{8.30}$$

$$\phi^t(u) = 1/(1 - iu)^t. \tag{8.31}$$

As $t \to 0, f^t(x)/t \to e^{-x}/x$, and so the gamma process, in which $Y(t)$ has a gamma distribution with exponent t, derives from jumps whose intensities have the distribution

$$M(x) = \int_x^\infty (e^{-x}/x)dx, \quad M'(x) = e^{-x}/x. \tag{8.32}$$

In interpreting this, we note the connection with the Poisson process: $f^t(x)$, for $t = h$ integer, gives the distribution of the waiting-time, T_h, for the hth occurrence of the phenomenon. We can also obtain this by arguing in terms of the $p_h(t)$ (see Section 8.2.5); this makes it completely obvious, because for $h = 1$ we have the exponential distribution (for T_1, and, independently, for any waiting time $T_h - T_{h-1}$) and for an arbitrary integer h we have the convolution corresponding to the sum T_h of the h individual waiting times.

It is interesting to note that one possible interpretation of the gamma process is as the *inverse* of the (simple) Poisson process. To see this, we interchange the notation, writing this process as $t = T(y)$, the inverse of the other, for which we keep the standard notation, $y = Y(t)$. The inverse function $y = T^{-1}(t)$ (which, of course, does not give a process with independent increments), considered only at those points for which y = integer (or taking $Y(t)$ = the integral part of $y = T^{-1}(t)$), gives precisely the simple Poisson process (with $\mu = 1$).

Remark. We obtain a perfect interpretation if we think, for example, of $y = T^{-1}(t)$ as representing the number of turns (or fractions thereof) made by a point moving in a series of jerks around the circumference of a circle. The standard Poisson situation then corresponds to that of someone who is only able to observe the point when it passes certain given marks.

We are in no way suggesting, however, that this mathematical possibility of considering an 'explanation' in terms of some 'hidden mechanism' provides any automatic justification for metaphysical flights of fancy leading on to assertions about its 'existence' (whereas it may, of course, be useful to explore the possibility if there are some concrete reasons for considering it plausible). There are a number of cases (but not, so far as I know, the present one) in which these kinds of metaphysical interpretation (or so they appear to me, anyway) are accepted, or, at any rate, seriously discussed.

More generally, changing the scale and intensity, we have

$$f^t(x) = Kx^{\mu t-1}e^{-\lambda x}(x>0),\ K = \lambda^{\mu t}/\Gamma(\mu t), \tag{8.33}$$

$$\phi^t(\mu) = (1 - iu/\lambda)^{\mu t}. \tag{8.34}$$

Moreover, we can also reflect the distribution onto the negative axis; $f^t(x)$ is unchanged, apart from writing $|x|$ in place of x and changing the final term to $(x < 0)$, and the characteristic function is given by

$$\phi^t(\mu) = (1 + iu/\lambda)^{\mu t}. \tag{8.35}$$

By convolution, we can construct other processes corresponding to sums of gamma processes. The most important case is that obtained by symmetrization (see equations 6.57 and 6.59 in Chapter 6, 6.11.3), which, for $t = 1$, gives the double-exponential distribution. The general case is obtained from products of the form

$$\phi^t(\mu) = (1 - iu/\lambda)^{\mu_1 t}(1 - iu/\lambda_2)^{\mu_2 t}\ldots(1 - iu/\lambda_n)^{\mu_n t}$$

(where the signs can be either − or +; we could, alternatively, say that the λ_h can be positive or negative); the symmetric case arises when the products are taken in pairs with opposite signs: that is

$$\phi^t(\mu) = \left(1 + u^2/\lambda_1^2\right)^{\mu_1 t}\ldots\left(1 - u^2/\lambda_n^2\right)^{\mu_n t}.$$

8.4.6. The Bessel process acquires its name because the form of the density involves a Bessel function, $I_t(x)$ (for $x > 0$):

$$f^t(x) = \left(e^{-x}/x\right)tI_t(x), \quad \text{where } I_t(x) = \sum_{k=0}^{\infty}\frac{1}{k!\Gamma(k+t+1)}\left(\frac{x}{2}\right)^{2k+1}; \tag{8.36}$$

the characteristic function is given by

$$\phi^t(u) = \left\{1 - iu - \sqrt{\left[(1-iu)^2 - 1\right]}\right\}^t. \tag{8.37}$$

The process deserves mentioning because it has the same interpretation as we put forward for the simple Poisson process, but referred instead to the Poisson variant of the Heads and Tails process (each occurrence of the phenomenon consists of a toss giving a gain of ± 1). Using the same notation, $t = T(y)$ and $y = T^{-1}(t)$ for the inverse function, we could say that the points at which y = an integer correspond to the instants at which the gain $Y(t)$ first reaches level y (or that the 'integer part of $y = T^{-1}(t)$' is the maximum of $Y(t)$ in $[0, t]$, where $Y(t)$ is a Poisson Heads and Tails process).

8.4.7. The Pascal process is that for which $Y(t)$ has, for $t = 1$, a geometric distribution; for t = integer, a Pascal distribution; and, for general t, a negative binomial distribution (see Chapter 7, 7.4.4). Our previous discussion reveals that we are dealing with a compound Poisson process having a *logarithmic* distribution of jumps (see Chapter 6, 6.11.2) and having the positive integers as the set of possible values (with intensity $\mu = 1$).

The interpretation is similar to that of the two previous cases, except that here we do not have a density but, instead, concentrated masses (at the integer values). For t = integer, $Y(t)$ is the number of failures (in a Bernoulli process with probability p, where $\tilde{p}$ is the factor in the geometric distribution corresponding to $t = 1$) occurring before the tth success. In any interval of unit length (from t to $t + 1$), the increment has the geometric distribution; here, it can be thought of as generated by logarithmically distributed jumps, one per interval 'on average'. A noninteger t could be interpreted as 'the number of successes already obtained, plus the *elapsed fraction* of the next one'; $Y(t)$ would then be the number of failures that had actually occurred so far.

Of course, the remark made in the context of the gamma process (Section 8.4.5) applies equally to the other two processes (Sections 8.4.6 and 8.4.7).

8.5 Behaviour and Asymptotic Behaviour

8.5.1. We now turn to a probabilistic study of various aspects of the behaviour of the function $Y(t)$; as we pointed out in Section 8.1.6, these are, in fact, the most important questions. They might be concerned with the behaviour of $Y(t)$ in the neighbourhood of some given instant t (local properties), or in an interval $[t_1, t_2]$, or in the neighbourhood of $t = \infty$ (asymptotic properties). We might ask, for instance, whether $Y(t)$ vanishes somewhere in a given interval (and, if so, how many times), or if it remains bounded above by some given value M_1, or below by M_2 (or, more generally, by functions $M_1(\mathrm{t})$ and $M_2(t)$ rather than constants), and so on. Asymptotically, we might ask whether these or other circumstances will occur from some point on, or locally, or only in the neighbourhood of some particular instant.

Phrases such as we have used here will have to be interpreted with due care, especially if one is not admitting the assumption of countable additivity. Particular attention must then be paid to expressing things always in terms of a finite number of instants (which could be taken arbitrarily large) and not in terms of an infinite number.

We have already come across a typical example of just this kind of question in Chapter 7, 7.5.3. This was the strong law of large numbers, which, in the case of a discrete process, consisted in the study of the asymptotic validity of inequalities of the form $C_1 \leqslant Y(t)/t \leqslant C_2$, or, equivalently,

$$C_1 t \leqslant Y(t) \leqslant C_2 t.$$

We shall see now how this problem, and generalizations of it, together with a number of other problems of a similar kind, can be better formulated, studied and appreciated by setting them within the more general context of the study of random processes. In the most straightforward and intensively studied cases, the processes that will turn out to be useful are precisely those which we have been considering; specifically, the homogeneous processes with independent increments and, from time to time, the Wiener–Lévy process (in situations where it is valid as an asymptotic approximation).

It is both interesting and instructive to observe how the conjunction of the two different modes of presentation and approach serves to highlight conceptual links which are otherwise difficult to uncover, and to encourage the use of the most appropriate methods and techniques for each individual problem. In particular, later in this chapter (in which we shall only deal with the simplest cases) we shall see how the studies of the Heads and Tails process (developed directly using a combinatorial approach) and of the Wiener–Lévy process (which can be considered in a variety of ways) complement each other, enabling us in each case to use the more convenient form, or even to use both together.

8.5.2. We now consider two sets of questions, each set related to the other, and each collecting together, in a unified form, problems of different kinds which admit a variety of interpretations and applications, both theoretical and practical.

The first set brings together problems that reduce to the consideration of whether or not the path $y = Y(t)$ leaves some given region. In other words, whether or not it crosses some given barrier (the region may consist of a strip, $C_1 \leqslant y \leqslant C_2$, or $C_1(t) \leqslant y \leqslant C_2(t)$, or one of the bounds may not be present; i.e. $C_1 = -\infty$ or $C_2 = +\infty$). For this kind of problem, what we are usually interested in knowing is whether or not the process leaves the region and, if so, when this occurs for the first time, and if the process then comes to a halt. In the latter case, we have a so-called absorbing barrier. In general, however, it is useful to argue as if the process were to continue.

An analysis of this kind will serve to make the strong law of large numbers more precise, by examining the rate of convergence that can be expected (we shall, of course, have to make clear what we mean by this !). From the point of view of applications, there will be a number of possible interpretations, among which we note: the gambler's ruin problem (which could be thought of in the context of an insurance company); the termination of a sequential decision-making process because sufficient information has been acquired; the end of a random walk – for example, the motion of a particle – due to arrival at an absorbing barrier, and so on.

The second set of problems are all concerned with recurrence and involve the repetition of some given phenomenon (such as return to the origin, passing a check point etc.). We shall see that in this case the process divides rather naturally into segments and that it may be of interest to study various characteristics of these segments, which, in turn, often shed new light on the original recurrence problems.

In both cases, we shall first consider the Heads and Tails process and the Wiener–Lévy process, afterwards extending the study to the asymptotically normal cases. We shall also have occasion to consider other processes (the Bernoulli process with $p \neq \frac{1}{2}$, stable processes with $\alpha < 2$ and the Poisson process), mainly in order to have the opportunity of presenting further points of possible interest (explanations of unexpected behaviour, drawing attention to unexpected properties, and so on).

8.5.3. The three cases that we shall consider first involve the comparison of $|Y(\mathrm{t})|$ with $y = C$, $y = Ct$, $y = C\sqrt{t}$ $(c > 0)$. Even though we shall dwell, in each case, on possible interpretations in an applied context, developing each topic as required, it will be useful to bear in mind that these are in the way of preliminaries, enabling us subsequently to determine the foreseeable order of increase of $Y(t)$.

The first case, that of comparison with a constant, is the one of greatest practical interest and corresponds to the gambler's ruin problem under the assumption of fixed capital (i.e. with no increases or decreases, other than those caused by the game). We shall now consider the complex of problems which arise in this case, beginning with the simplest.

We first observe that the probability of $Y(t)$ lying between $\pm C$ tends to zero at least as fast as $K/\sqrt{t}$ (if $Y(t)$ has finite variance σ^2 per unit time, it tends to zero as $K/\sqrt{t}$, with $K = 2C/[\sqrt{(2\pi)}\sigma]$; if the process has infinite variance, then, whatever K may be, it tends faster than $K/\sqrt{t}$). The same holds if we are dealing with just a single t. If $\sigma = 1$, we have, in a numerical form, $K = 0{\cdot}8C$ (see Chapter 7, 7.6.3) and, for ease of exposition, we shall, as a rule, refer to this case (the bound is $0{\cdot}8C/\sqrt{t}$, and is approximately attained if $\sigma = 1$; if $\sigma = \infty$, we have strict inequality). The same holds for every interval $[a - C, a + C]$ of length $2C$.

To see this, note that, for $t = n$, in the case of Heads and Tails the maximum probability is given by $u_n \simeq 0{\cdot}8/\sqrt{n}$ and there are $2C$ integers lying between $\pm C$ (give or take $\pm$ 1[15]). In the normal case (Wiener–Lévy), the maximum density is $1/[\sqrt{(2\pi)}\sigma(t)]$, and $\sigma(t) = \sqrt{t}$. In the general case, with $\sigma < \infty$ (and, without loss of generality, we assume $\sigma = 1$), we have, asymptotically, the same process.

If $\sigma = \infty$, the bound corresponding to an arbitrary finite σ is *a fortiori* satisfied from a certain point onwards. Suppose $a > 0$ and arbitrary, but such that

$$\mathbf{P}\left(\left|X_h\right| < a\right) = p > 0,$$

further, let us distinguish the increments $X_h = Y(h) - Y(h - 1)$ according to whether they are $<a$ or $>a$ in absolute value; $Y(n) = \Sigma X_h$ then contains about np summands which are $<a$ (*truncated* distribution, with $\sigma < a$ finite), and the sum is already practically normal with density $<1/[\sqrt{(2\pi)}\sigma/t]$. Adding the sum of the other terms, we have, *a fortiori*, the same circumstance holding (theorem of the increase in dispersion; Chapter 6, 6.9.8).

What matters more than the quantitative result, is the qualitative conclusion: *however large C is chosen to be, the probability that* $|Y(t)|$ (*or* $|Y_n|$) *is greater than C differs from* 1 *by less than any given* $\varepsilon > 0$, *provided that t* (*respectively, n*) *is taken sufficiently large* (more precisely, from

$$t = n = (2/\pi)C^2/\varepsilon^2$$

onwards).

A fortiori, it follows that the probability of $|Y(\tau)| > C$ (or $Y_h > C$) for at least one $\tau \leqslant t$, or $h \leqslant n = t$, tends to 1 (more rapidly). In terms of the gambler's ruin problem, this implies that in a game composed of identical and independent trials between two gamblers each having finite initial capital, provided the game goes on long enough, the probability of it ending with the ruin of one of them tends to 1. Equivalently, the probability that the game does not come to an end is zero.

15 In order for this difference not to matter, it is, of course, necessary that C be much greater than 1. In general, in the case of lattice distributions, or distributions of similar kind, it is necessary that C be large enough for the concentrated masses to be regarded as a 'density'.

8.5.4. A warning against superstitious interpretations of the 'laws of large numbers'. *It is not only true that the absolute deviations, that is the gains and losses* (unlike the relative gains, the average gains per toss), *do not tend to offset one another and, in fact, tend to increase indefinitely in quadratic prevision, but also that it is 'practically certain that, for large, they will be large'* (and it is only in the light of the above considerations that we are now able to see this).

One should be careful, however, not to exaggerate the significance of this statement, turning it, too, into something misleading or superstitious. It holds for each individual instant t (or number of tosses n), but not for a number of them simultaneously. In fact, it does not exclude the possibility (and, indeed, we shall see that this is practically certain) of the process *returning to equilibrium* (and hence of there being segments in which $|Y(t)| < C$) every now and again (and although this happens more and more rarely, it is a never-ending process).

8.6 Ruin Problems; the Probability of Ruin; the Prevision of the Duration of the Game

8.6.1. We shall use p_h and q_n to denote the probabilities of ruin at the hth trial, or before the nth, respectively

$$\left(q_n = p_1 + p_2 + \ldots + p_n,\ p_h = q_h - q_{h-1}\right).$$

In the case of two gamblers, G_1 and G_2, we shall use p'_h and p''_h, q'_n and q''_n for the probabilities of ruin of G_1 and G_2, respectively

$$\left(p_h = p'_h + p''_h, \quad q_n = q'_n + q''_n\right),$$

and c' and c'' for their respective initial fortunes.

By q' and q'' (or q'_∞ and q''_∞), we shall denote the probabilities of ruin within an infinite time, to be interpreted as limits as $n \to \infty$: under the assumptions of Section 8.5.3, we have $q'_n + q''_n \to 1$, and therefore $q' + q'' = 1$.

The probability of ruin, in a fair game, is an immediate consequence of the fairness condition: the previsions of the gains of the two gamblers must balance; that is $q'c' = q''c''$, from which we deduce that $q' = K/c'$, $q'' = K/c''$ $(K = c'c''/(c' + c''))$. In other words, the probabilities of ruin are inversely proportional to the initial fortunes. More explicitly,

$$q' = c''/\left(c' + c''\right),\ q'' = c'/\left(c' + c''\right). \tag{8.38}$$

Comments. In these respects, we might also apply the term 'fair game' to a nonhomogeneous process with nonindependent increments, provided that the prevision conditional on any past behaviour is zero (such processes are called *martingales*). One could think, for example, of the game of Heads and Tails with the stakes depending in some way on the preceding outcomes. With these assumptions, any mode of participation in the game is always fair: it does not matter if one interrupts the play in order to alter the stakes, or even if one decides to stop playing on account of a momentary impulse, or when something happens – such as someone's ruin.

The relation $q'c' = q''c''$ is an exact one if ruin is taken to mean the exact loss of the initial capital with no unpaid residue; in the latter case, it would be necessary to take this into account separately. If, for example, the jumps which are disadvantageous to G_1 and G_2 cannot exceed Δ' and Δ'', respectively, then c' must be replaced by some $c' + \theta'\Delta'$ $(0 \leqslant \theta' \leqslant 1)$, with a similar substitution for c''; the error is negligible if the probable residues are small compared with the initial capital.

The conclusion holds exactly for the Wiener–Lévy process because the continuity of $Y(t)$ ensures that it cannot exceed c' and c'' by jumping past them. The same holds true for the Heads and Tails process – including the Poisson variant – provided c' and c'' are integers (because it is then not possible, with steps of ± 1, to jump over them).

It is clear, particularly if we use the alternative form

$$q' = 1 - 1/\left[1 + \left(c''/c'\right)\right],$$

that the probability of G_1's ruin tends to 1 if the opponent has a fortune that is always greater than his. If he plays against an opponent with infinite capital, the probability of ruin is, therefore, $q' = 1$ (and this is also true if one plays against the general public – who cannot be ruined). This is the *theorem of gambler's ruin* (for fair games).

The case of unfair games reduces to the previous case if one employs a device that goes back to De Moivre: in place of the process $Y(t)$, we consider $Z(t) = \exp[\lambda Y(t)]$. If λ is chosen in such a way as to make the prevision of $Z(t)$ constant (=1, say), then the process $Z(t)$ is fair, and ruin (starting from Z(0) = 1, which corresponds to $Y(0) = 0$) occurs when one goes down by $\bar{c}'' = 1 - \exp(-\lambda c')$, or up by

$$\bar{c}'' = \exp\left(\lambda c''\right) - 1.$$

The probabilities of ruin are, therefore, inversely proportional to $\bar{c}'$ and $\bar{c}''$. It only remains to say how λ is determined. We observe that $\exp[\lambda Y(t)] = \phi^t(-\mathrm{i}\lambda)$ and that, if it exists (see Chapter 6, 6.10.4), ϕ is real and concave on the imaginary axis, taking the value 1 (apart from at the origin) only at the point $u = -\mathrm{i}\lambda$, with λ positive if the game is unfavourable ($\mathbf{P}[Y(t)] < 0$).

Example. We consider the case of Heads and Tails with an unfair coin ($p \neq \frac{1}{2}$) and with gains ± 1. We have $\exp[\lambda Y(1)] = p\mathrm{e}^{\lambda} + \tilde{p}\mathrm{e}^{-\lambda} = 1$, in other words (putting $x = \mathrm{e}^{\lambda}$), $px^2 - x + (1 - p) = 0$ for $x = 1$ and $x = \tilde{p}/p$; $x = \mathrm{e}^{\lambda} = 1$ would yield $\lambda = 0$ (which is meaningless), so we take $\mathrm{e}^{\lambda} = \tilde{p}/p$, $\mathrm{e}^{-\lambda c'} = (\tilde{p}/\mathrm{p})^{-c'}$, $\mathrm{e}^{\lambda c''} = (\tilde{p}/p)^{c''}$ from which we obtain

$$q' = \frac{\left(\tilde{p}/p\right)^{c''} - 1}{\left(\tilde{p}/p\right)^{c''} - \left(\tilde{p}/p\right)^{-c'}}, \quad q'' = \frac{1 - \left(\tilde{p}/p\right)^{-c'}}{\left(\tilde{p}/p\right)^{c''}\left(\tilde{p}/p\right)^{-c'}}. \tag{8.39}$$

If one plays against an infinitely rich opponent, the passage to the limit as $c'' \to \infty$ gives two different results, according to whether the game is favourable, $(\tilde{p}/p) < 1$, or unfavourable $(\tilde{p}/p) > 1$; in the latter case, $q' = 1$ (as was obvious *a fortiori*; ruin is practically certain in the fair case); if, instead, the game is favourable, the probability of ruin is

$q' = (\tilde{p}/p)^{c'}$ and $1 - q' = 1 - (\tilde{p}/p)^{c'}$ is the probability that the game continues indefinitely.

8.6.2. The prevision $\mathbf{P}(T)$ of the duration T of the game until ruin occurs can also be determined in an elementary fashion for the game of Heads and Tails (even for the unfair case) and then carried over to the Wiener–Lévy process.

Instead of merely determining $\mathbf{P}(T)$ (starting from $Y(0) = 0$), it is convenient to determine the prevision of the future duration for any possible initial value y $(-c' \leqslant y \leqslant c'')$ using a recursive argument; we denote this general prevision by $\mathbf{P}_y(T)$. Obviously, we must have $\mathbf{P}_y(T) = 0$ at the end-points $(y = -c', y = c'')$ because there ruin has already occurred. For y in the interval between the end-points, we have, instead, the relation

$$\mathbf{P}_y(T) = 1 + \tfrac{1}{2}\left[\mathbf{P}_{y-1}(T) + \mathbf{P}_{y+1}(T)\right]$$

(because we can always make a first toss, and then the prevision of the remaining duration can be thought of as starting from either $y+1$ or $y-1$, each with probability $\frac{1}{2}$). We then obtain a parabolic form of behaviour (the second difference is constant!) with zeroes at the end-points; explicitly, we obtain

$$\mathbf{P}_y(T) = -(y + c')(y - c''), \qquad \text{and hence } \mathbf{P}(T) = \mathbf{P}_0(T) = c'c''. \tag{8.40}$$

As c'' increases, $\mathbf{P}(T)$ tends to ∞ no matter what $c' > 0$ is; it follows that $\mathbf{P}(T) = \infty$ for a game against an infinitely rich opponent, so that although ruin is practically certain $(q' = 1)$, the expected time before it occurs is infinite.

Even in the case where c' and c'' are finite, the expected duration of the game, although finite, is much longer than one might at first imagine. For example, in the symmetric case, $c' = c'' = c$, the expected duration of the game is c^2 tosses: 100 tosses if each gambler starts with 10 lire; 40 000 tosses if each starts with 200 lire; 25 million tosses if each starts with 5000 lire. In the most extremely asymmetric case, $c' = 1$, $c'' = c$, the expected duration is c tosses; 1000 tosses if initially the fortunes are 1 lira versus 1000 lire; 1 million if initially we have 1 lira versus 1 million. One should note, however, that in this asymmetric case the gambler whose initial fortune is 1 lira always has the same (high) probabilities of coming to grief almost immediately, whatever the initial capital of his opponent (be it finite or infinite), provided that it is sufficient to preclude the opponent's ruin within a few tosses. Specifically, the probability is 75% that the gambler with 1 lira is ruined within 10 tosses, 92% that he is ruined within 100 (in general, $1 - u_n \simeq 1 - 0{\cdot}8/\sqrt{n}$); in these cases, fortunes of 10 or 100 lire, respectively, will ensure that the opponent cannot be ruined within this initial sequence of tosses. On the other hand, there is a chance, albeit very small, that the opponent will be the one who is ruined (this is about $1/c$; one thousandth if $c = 1000$, for example). In order for this to happen, it is necessary that the gambler who begins with 1 lira reaches a situation of parity (500 versus 500) without being ruined; after this, the expected duration of the game will be $500^2 = 250\,000$ tosses, there then being equal probabilities of ruin for the two parties. There is, therefore, a probability of two in a thousand of reaching parity but, in such a case, the subsequent duration of the game is almost certainly very long. As always, one should remember that prevision is not prediction.

For the Wiener–Lévy process, thinking of it as a limit case of Heads and Tails, one sees immediately that exactly the same conclusion holds. It is sufficient to observe that the change of scale ($1/N$ for the stakes, $1/N^2$ for the intervals between tosses) leaves the duration unchanged: the initial capitals are Nc' and Nc'', the duration $N^2c'c''$, with $1/N^2$ as unit. More generally, we can say that, roughly speaking, the conclusion holds for all processes with finite variances ($\sigma = 1$ per unit time; otherwise, $\mathbf{P}(T) = c'c''/\sigma^2$) provided c' and c'' are large enough to make the ruin very unlikely after a few large jumps.

In the case of games which are not fair, one can apply the same argument, but the result is different. In the case of Heads and Tails with an unfair coin ($p \neq \frac{1}{2}$) the relation

$$\mathbf{P}_y(T) = 1 + (1-p)\mathbf{P}_{y-1}(T) + p\mathbf{P}_{y+1}(T)$$

reduces to the characteristic equation $py^2 - y + (1 - p) = 0$, with roots 1 and $(1 - p)/p$, which gives $A + B(\tilde{p}/p)^y$ as the solution of the homogeneous equation. It is easily seen that $y(1 - 2p)$ (or $y/(\tilde{p} - p)$) is a particular solution of the complete equation and, taking into account the fact that $\mathbf{P}_y(T) = 0$ for $y = -c'$ and $y = c''$, we have

$$\mathbf{P}_y(T) = \frac{1}{1-2p}\left[(y+c') - (c'+c'')\frac{1-(\tilde{p}/p)^{y+c'}}{1-(\tilde{p}/p)^{c'+c''}}\right]. \tag{8.41}$$

For the extension to the Wiener–Lévy process (and, more or less as we have said, to cases which approximate to it), it is sufficient to observe that, in the case we have studied, $m = 2p - 1$, $\sigma^2 = 1 - m^2$, from which we obtain $p = \frac{1}{2} + \frac{1}{2}m/\sqrt{(m^2 + \sigma^2)}$. Given the m and σ of a Wiener–Lévy process (or a general process), it suffices to evaluate p in this way.

If one plays against an infinitely rich opponent ($c'' = \infty$), we have $\mathbf{P}(T) = \infty$, provided the game is advantageous ($p > \frac{1}{2}$), and given that, with nonzero probability, it can last indefinitely. If it is disadvantageous ($p < \frac{1}{2}$) only the first term remains:

$$\mathbf{P}(T) = c'/(1-2p). \tag{8.42}$$

8.6.3. *The probabilities of ruin within given time periods* (i.e. within time t, or within $n = t$ tosses) provide the most detailed answer to the problem. Let us consider, for the time being, the case of one barrier ($c' = c$, $c'' = \infty$), and let us begin with Heads and Tails. We shall attempt to determine the probability, q_n, of ruin within n tosses; that is the probability that $Y_h = -c$ for at least one $h \leqslant n$.

The solution is obtained by making use of the celebrated, elegant *argument of Desiré André*. In the case of the game of Heads and Tails, we adopt the following procedure for counting the number of paths (out of the 2^n possible paths between 0 and n) which reach the level $y = -c$ at some stage. First of all, we consider those which terminate beyond that level, $Y(n) < -c$, and we note that there are exactly the same number for which $Y(n) > -c$, since any path in this latter category can be obtained in one and only one way from one of the paths in the former category by reflecting (in the straight line

$y = -c$) the final segment starting from the instant $t = k$ at which the level $y = -c$ is reached for the first time. In other words, we use the reflection

$$Y^*(t) = -c - \left(Y(t) + c\right) \qquad \left(k \leqslant t \leqslant n\right).$$

Finally, we note that (if $n - c$ is even) there are some paths for which $Y(n) = -c$. In terms of the probability (the number of paths divided by 2^n), the first group contribute to $\mathbf{P}(Y_n < -c)$ and, because of the symmetry revealed by André's reflection principle, so do the second group; the final group contribute to $\mathbf{P}(Y_n = -c)$. Expressing the result in terms of c rather than $-c$, we have therefore

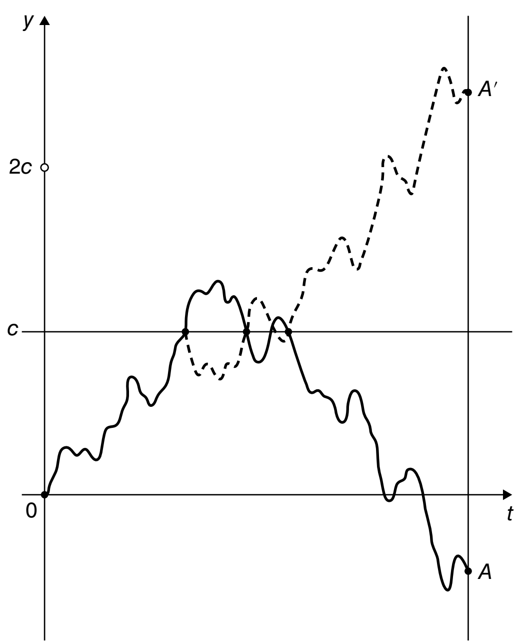
$$q_n = 2\mathbf{P}\left(Y_n > c\right) + \mathbf{P}\left(Y_n = c\right) = \mathbf{P}\left(|Y_n| > c\right) + \tfrac{1}{2}\mathbf{P}\left(|Y_n| = c\right). \tag{8.43}$$

The basic idea is illustrated most clearly in Figure 8.6; to any path which having reached $y = -c$ finds itself at $t = n$ above that level, there corresponds, by symmetry, another path which terminates below it (and, indeed, if the first path terminates at $-c + d$, the second will terminate at $-c - d$; the reader should interpret this fact). Essentially, we could say that reflection corresponds to exchanging the rôles of Heads and Tails (from the instant of ruin onwards), a device that we have already come across (in example (D) of Chapter 7, 7.2.2).

A further principle – similar to that of Desiré André – was introduced by Feller (Vol. I, p. 20) under the heading of the duality principle; we prefer to call it the *reversal principle*, because the other term suggests connections, which do not actually exist, with other, unrelated, notions of 'duality'. The idea is that one reverses the time order; that is

Figure 8.6 Desiré André's argument in the case of a single barrier. The paths which, after having reached level c, are below it at the end of the interval of interest (point A), correspond in a one-to-one manner, by symmetry, to those terminating at A' (symmetric with respect to the barrier $y = c$). It follows (in a symmetric process) that the probability of ending up at A', or at A, having reached level c, is the same.

The point $2c$ on the y-axis has been marked in because it corresponds to the 'cold source' in Lord Kelvin's method (Section 8.6.7).



the ordered events $E_1 \ldots E_n$ become $E_n \ldots E_1$, by setting $E_i = E_{n-i+1}$. The reversed gain is then

$$Y^*(h) = Y(n) - Y(n-h),$$

and the path is reversed (i.e. rotated by 180°) with respect to the central point $(\frac{1}{2}n, \frac{1}{2}\mathrm{Y}(n))$.

The argument and the result given above have a far more general range of application. Indeed, the only facts about the distribution of the increments that we have used are its symmetry ($Y(t_0 + t) - Y(t_0)$ has equal probability of being $>a$ or $<-a$, and, in particular, of being $\lesseqgtr 0$), and the fact that the level $y = -c$ cannot be exceeded by 'skipping' it (i.e. anything exceeding it must actually pass through it). This holds for the Heads and Tails process if c = integer,[16] and for the Wiener–Lévy process (assuming it to be continuous) for arbitrary c. For other cases, it may have approximate, or asymptotic, validity (as in the *Remark* of Section 8.6.1), provided that the jumps in the direction in which the fixed level must be exceeded are small, or, at any rate, the large jumps are relatively rare (this brief comment will suffice; we do not wish to complicate matters by going into all the details).

Some further terminology and notation will be needed for certain more general problems and results that we wish to consider. The *maximum* of $Y(\tau)$ in $0 < \tau \leqslant t$[17] will be denoted by $\vee Y_n$ (an abbreviated form of $Y_1 \vee Y_2 \vee \ldots \vee Y_n$[18]) in the discrete case, and by $\wedge Y(t)$ in the continuous case; similarly, $\wedge Y_n$ and $\wedge Y(t)$ denote the *minimum*;

$$|\wedge Y(t)| = -\wedge Y(t) = \vee(-Y(t))$$

is the absolute value of the minimum, and we shall refer to it as |minimum|.

With this notation, the q_n (or, to be more accurate, the $q_n(c)$) of equation 8.43 determine the probability distribution of $\vee Y_n$ (and of $|\wedge Y_n|$; it is the same, by symmetry);

$$q_n(c) = \mathbf{P}(\wedge Y_n \leqslant -c) = \mathbf{P}(\vee Y_n \geqslant c).$$

By subtraction, we obtain the probabilities

$$\begin{aligned} \mathbf{P}(|\wedge Y_n| = c) &= \mathbf{P}(\vee Y_n = c) = q_n(c) - q_n(c-1) \\ &= \mathbf{P}(Y_n = c) + \mathbf{P}(Y_n = c+1) \end{aligned} \tag{8.43$'$}$$

16 The set of levels which cannot be skipped may take various forms: either all c (positive, negative, or both) if $Y(t)$ varies continuously (nondecreasing, nonincreasing, or completely general); or all multiples of some given k (positive, negative, or both) if all positive jumps are $= k$, and the negative ones are multiples of k (or conversely, or if they are all $\pm k$); in all other cases there are no such levels.

17 We refer to $0 < \tau$ (instead of $0 \leqslant \tau$) in order to facilitate comparison with the discrete case (although the distinction loses its meaning in the continuous case); what is important is to stress that $\tau = t$ is to be included, and, indeed, we should stress that $Y(t)$ is to be understood as $Y(t + 0)$ (i.e. taking into account a possible jump occurring exactly at t).

18 The omission of Y_0 is irrelevant, except when it is useful to distinguish two cases that otherwise would both yield $\vee Y_n = 0$ (all $Y_n \leqslant 0$); with the convention adopted, we have, instead, $\vee Y_n = -1$, if $Y_1 = -1$, and the successive values are all $\leqslant -1$ (or, in general, stepping outside the example of Heads and Tails, $\vee Y_n$ can be an arbitrary negative value).

(only one of the two summands is ever present; the first if n and c have an even sum, the second if the sum is odd). Finally, we obtain

$$\mathbf{P}\left(|\wedge Y_n| = c\right) = \omega_h^{(n)} = \binom{n}{h} 2^{-n},$$

where either $2h - n = c$, or $2h - n = c + 1$.

It is important to pay particular attention to the cases $\wedge Y_n = 0$ and $\wedge Y_n = -1$. They are not contained in the general case (which is based on the assumption $c > 0$) but they can easily be reduced to the appropriate form. For $c = 1$, we have $\mathbf{P}(\vee Y_n < 1) = u_n$ (i.e. $\mathbf{P}(Y_n = 0)$ or $\frac{1}{2}\mathbf{P}(|Y_n| = 1)$, according to whether n is even or odd), and the two cases $\vee Y_n = 0$ and $\vee Y_n = -1$ are equally probable if n is even (and they do not differ by very much if n is odd). More precisely, we have $\mathbf{P}(\vee Y_n = -1) = \frac{1}{2}u_{n-1}$ (the first step is -1, and we do not then go up by $+1$), and, by subtraction

$$\mathbf{P}\left(\vee Y_n = 0\right) = u_n - \frac{1}{2}u_{n-1}$$

($u_n = u_{n-1}$ if n is even; otherwise $u_{n-1} = u_n n/(n - 1)$). In words (for n even): u_n is also the probability that Y_h remains *non-negative* in $0 < t \leqslant n$ (and a similar argument holds for the *nonpositive* case).

Let us also draw attention to the following (interesting) interpretation of equation 8.43′.

The probability of attaining $y = c$ ($c > 0$) as the maximum level for $t \leqslant n$ is precisely the same as that of attaining the same level c (or $c + 1$, according to whether we are dealing with the even or odd case) at $t = n$ (but not, in general, as the maximum level).

To put it another way: the probability $2\omega_h^{(n)}$ that $|Y_n|$ assumes the value $c = 2h - n$ ($h > n/2$) splits into two halves for $\vee Y_n$; one half remains at c and the other at $c - 1$. If this partial shift of one is negligible in a given problem (as it is, in any case, asymptotically) we may say that the distributions of the absolute value, $\vee Y_n$, of the maximum, $\vee Y_n$, and of the |minimum|, $|\wedge Y_n|$, are all identical. Obviously, the maximum and the |minimum| are nondecreasing functions, and hence we can define their inverses. We denote by $T(y)$ ($y \geqslant 0$) the inverse of $\vee Y(t)$ and by $T(-y)$ the inverse of $\vee(-Y(t))$ (they also have the same probability distribution as processes; however, $T(y)$ for $y \lesseqgtr 0$ is not to be understood as a single process for $-\infty < y < +\infty$ but rather as a unified notation for two symmetric but distinct processes):

$T(y)$ = the minimum t for which

$$\vee Y(t) \geqslant y \,(y > 0), \quad \text{or} \quad \vee\left(-Y(t)\right) \geqslant -y\,(y < 0),$$

so that

$$\left(T(y) \leqslant t\right) = \left(\vee Y(t) \geqslant y\right)(y > 0) \vee \left(\vee\left(-Y(t)\right) \geqslant -y\right)(y < 0).$$

For every y, $T(y)$ is the random quantity expressing the instant (or, equivalently, the waiting time) either until ruin occurs, or until the first arrival at level (or point) y occurs, or until a particle is absorbed by a possible absorbing barrier placed at, and so on.

With the present notation, we can express, in a direct form, the probabilities of ruin (or of absorption etc.) as obtained in equation 8.43 on the basis of Desiré André's argument:

$$\begin{aligned}\mathbf{P}\big(\vee Y(t)\geq y\big)=\mathbf{P}\big(\wedge Y(t)\leq -y\big)&=\mathbf{P}\big(T(y)\leq t\big)=\mathbf{P}\big(T(-y)\leq t\big)\\ &=2\mathbf{P}\big(Y(t)>c\big)+\mathbf{P}\big(Y(t)=c\big)\\ &=\mathbf{P}\big(|Y(t)|>c\big)+\frac{1}{2}\mathbf{P}\big(|Y(t)|=c\big).\end{aligned} \tag{8.44}$$

We omit the term corresponding to $Y(t) = c$, which is only required in order to obtain exact expressions for the Heads and Tails case, and which is either zero or negligible in general (for the exact expressions of the Wiener–Lévy process, the asymptotic expressions for Heads and Tails, and for other cases). If we denote by $F^t(y)$ and $f^t(y)$ the distribution function and the density (if it exists) of $Y(t$[19]), then the distribution function and the density of $|Y(t)|$ and therefore (either exactly or approximately) of $\vee Y(t)$ and $\vee(-Y(t))$ are given by

$$2F^t(y)-1 \quad (\text{for}\quad 0\leq y<\infty),\quad 2f^t(y)\quad (\text{for}\quad 0\leq y<\infty). \tag{8.45}$$

8.6.4. Substituting the exact forms for the Heads and Tails case into equations 8.44 or 8.45, we would have

$$q_n=\left(\tfrac{1}{2}\right)^{n-1}\sum_h\binom{n}{h}\left[0\leq h<\tfrac{1}{2}(n-c)\right]+\left(\tfrac{1}{2}\right)^n\binom{n}{(n-c)/2}\quad \big(\text{if } n-c \text{ is even}\big). \tag{8.46}$$

It is more interesting, however, to consider the approximation provided by the normal distribution; this will, of course, be exact for the Wiener–Lévy process and will hold asymptotically in the case of Heads and Tails, and for any other case with finite variance (which we shall always assume to be 1 per unit time). We have (for $y > 0$)

$$\begin{aligned}q_y(t)=p_t(y)&=\mathbf{P}\big(\vee Y(t)\geqslant y\big)=\mathbf{P}\big(T(y)\leqslant t\big)\\ &=2\mathbf{P}\big(Y(t)\geqslant y\big)=\sqrt{(2/\pi)}\int_{y\sqrt{t}}^{\infty}\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x.\end{aligned} \tag{8.47}$$

[20]

Clearly, the form of equation 8.47, interpreted either as a function of y (with t as a parameter) or, alternatively, as a function of t (with y as a parameter), provides the distribution function of $|Y(t)|$ and $\vee Y(t)$ and that of $T(y)$ and $T(-y)$. It is often useful to have this expressed in terms of both the parameters y and t; it can then be interpreted

19 This would also be valid for processes other than those which are homogeneous with independent increments (satisfying the conditions stated for Desiré André's argument) where there is less justification for writing F^t and f^t. In fact, however, we shall not be dealing with the general case.

20 For y large (compared with $\sqrt{t}$), the approximation given by equation 7.20 (in Chapter 7, 7.5.4) can be used, and gives

$$q_y(t)=p_t(y)\simeq K\left(\sqrt{t}/y\right)\mathrm{e}^{-y^2/2t},\qquad K=\sqrt{(2/\pi)}\simeq 0\cdot 8. \tag{8.47′}$$

according to whichever is appropriate. With this notation, we shall denote it in all cases by $\mathbf{P}(\vee Y(t) \geqslant y)$ (even in the Heads and Tails context).

The distribution of the maximum (or |minimum|) $\vee Y(t)$ (or $|\wedge Y(t)|$), and of $|Y(t)|$, is clearly the *semi-normal* (the normal distribution confined to the positive real axis), whose density is given by

$$f(x) = Kt^{-\frac{1}{2}} e^{-\frac{1}{2}x^2/t} \left(x \geqslant 0\right), \qquad K = \sqrt{(2/\pi)} \simeq 0 \cdot 8. \tag{8.48}$$

This is half of the normal distribution with $\bar{m} = 0$ and $\bar{\sigma}^2 = t$; the mean and variance are given by $m = \sqrt{(2/\pi)}\bar{\sigma}$[21] and $\sigma = \sqrt{(1 - 2/\pi)}\bar{\sigma}$; that is, numerically, $m \simeq 0{\cdot}8\bar{\sigma}$ and $\sigma \simeq 0{\cdot}6\bar{\sigma}$ (these should not be confused!).

On the other hand, $T(y)$ (or $T(-y)$) has density

$$f(t) = Kyt^{-\frac{3}{2}} e^{-\frac{1}{2}y^2/t} \left(t \geqslant 0\right), \qquad K = 1/\sqrt{(2\pi)} \simeq 0 \cdot 4. \tag{8.49}$$

This is the stable distribution with characteristic exponent $\alpha = \frac{1}{2}$ (which we mentioned in Section 8.4.4); in other words, it corresponds to jumps x whose density of intensity is $x^{-\frac{3}{2}}$. Since $\alpha < 1$, it has infinite prevision (in line with what we established directly in Section 8.6.2).

The fact that $T(y)$ had to have the stable distribution with $\alpha = \frac{1}{2}$ could have been deduced directly from the fact that

$$T(y_1 + y_2) = T(y_1) + \left[T(y_1 + y_2) - T(y_1)\right] = T(y_1) + T(y_2).$$

The time required in order to reach level $y_1 + y_2$ is, in fact, that required to reach y_1 plus that then required to proceed further to $y_1 + y_2$. However, given that, by the continuity of the Wiener–Lévy process, the level y_1 is reached (and not bypassed) at $T(y_1)$ with a jump, it is a question of going up by another y_2 under the same conditions as at the beginning. By virtue of the homogeneity, however, the distribution can only depend on y^2/t (and the density, in terms of y^2/t, could therefore only be – as, in fact, it actually is – a function of y^2/t divided by t).

We shall return (in Section 8.7.9) to the exact form for the Heads and Tails case, having encountered (for Ballot problems, in Section 8.7.1) an argument which enables us to establish it in a straightforward and meaningful way.

8.6.5. In the case where we consider *two gamblers*, G_1 and G_2 (with initial fortunes c' and c'', where $c' + c'' = c^*$), Desiré André's argument still applies, but now, of course, in a more complicated form. If we denote by A and B passages through levels c' and c'', respectively, a path whose successive passages are $ABABAB$ … signifies the ruin of G_1; if the sequence begins with B, it signifies the ruin of G_2. (It does not matter how many

21 $m = \bar{\sigma}\left[2/\sqrt{(2\pi)}\right]\int_0^\infty x \exp\left(-\frac{1}{2}x^2\right)\,\mathrm{d}x;$

it can easily be shown that the integral is equal to one.

times A is followed by B or B by A, nor does it matter whether the sequence ends in an A or a B; successive passages through the same level are not counted; e.g. $ABBAAABAABB$ = $ABABAB$.) Desiré André's argument (as applied in the one-sided case) does not directly enable one to count the paths {A} which signify G_1's ruin, nor the paths {B} which signify G_2's ruin, nor the paths {0} which indicate that neither is ruined (all belonging to the 2^n paths in the interval $[0, n]$). It does, however, enable one to count those of 'type (A)' 'type (B)' 'type (AB)', 'type (BA)', 'type (ABA)' and so on, where these refer to paths containing, in the sequence, the groups of letters indicated (which might be sandwiched between any number of letters).

Everything then reduces to the previous case of only one gambler; in other words, to

$$p_t(y) = p(y) = \mathbf{P}\left[\vee Y_n \geqslant y\right].$$

The probability of paths of 'type (A)' is, in fact, $p(c')$; that of paths of 'Type (AB)' is $p(c' + c^*)$ (because to first reach $-c'$ and then to reach $-c''$ requires a zigzag path along $c' + (c' + c'')$; this amounts to reflecting the path with respect to $y = -c'$, starting from the instant it reaches this level and continuing up until when it reaches c''); for 'type (ABA)' we have $p(c' + 2c^*)$, and so on. Similarly, for 'types' (B), (BA), (BAB), ..., we have

$$p(c''), p(c'' + c^*), \qquad p(c'' + 2c^*), \ldots,$$

and in this way we arrive at the required conclusion.

The paths {A} are, in fact, those given by

$$(A) - (BA) + (ABA) - (BABA) + (ABABA) - \ldots$$

(i.e. we start with those reaching $-c'$ and we exclude those first reaching c''; in this way, however, we exclude those that reach $-c'$ prior to c''; and so on). The same thing holds for {B}; the {0} are those remaining (i.e. neither {A} nor {B}) (see Figure 8.7).

It follows that the probabilities of ruin within n tosses are, for G_1, given by

$$q_n' = p(c') - p(c'' + c^*) + p(c' + 2c^*) - p(c'' + 3c^*) + \ldots \tag{8.50}$$

or (in terms of c''),

$$p(c^* - c'') - p(c^* + c'') + p(3c^* - c'') - p(3c^* + c'') + \ldots \tag{8.50'}$$

(there are a finite number of terms because $p(y) = 0$ when $y > n$).

The probabilities q_n'' (of ruin for G_2) are obviously expressed by the same formulae, provided we interchange the rôles of c' and c''.

In particular, in the symmetric case, $c' = c'' = c$, we have

$$q_n' = q_n'' = p(c) - p(3c) + p(5c) - p(7c) + \ldots \tag{8.51}$$

8.6.6. In the case of the Wiener–Lévy process (and, asymptotically, for Heads and Tails and for the asymptotically normal processes), we shall restrict ourselves, for simplicity and ease of exposition, to the symmetric case, which provides us with the

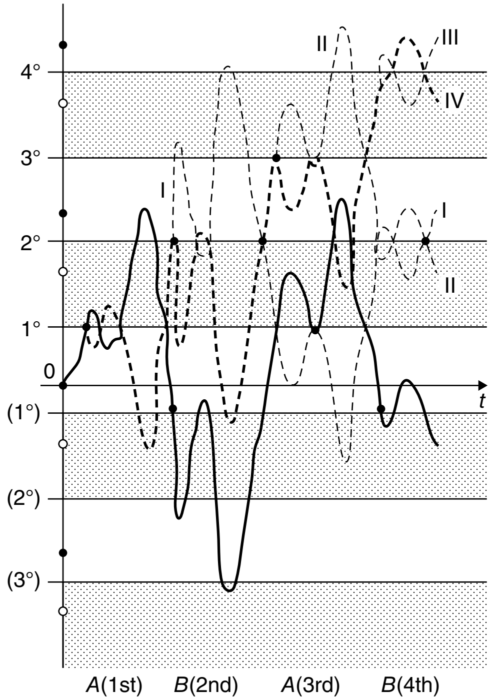


Figure 8.7 Desiré André's argument in the case of two barriers. The barriers are the straight lines bounding the white strip around the origin 0; the other strips are the *proper* images (white strips) and *reversed* images (dark strips) with the respective *hot* (black) and *cold* (white) sources (corresponding to Lord Kelvin's method; see Section 8.6.7). The actual path is indicated by the solid black line; its four successive crossings are denoted by *A*(1st), *B*(2nd), *A*(3rd), B(4th) (consecutive repeat crossings of *A* or *B* are not counted).

The final image of the path (obtained by repeated application of André's reflection principle) is indicated by the heavy broken line; it follows the same path up until *A*(1st) and is then given by the reflection (I) of it with respect to the 1st level. Then, after *B*(2nd), it is given by the reflection (II) of (I) with respect to the 2nd level, and so on. The continuations of the reflected paths (after the section in which they constitute the final image) are indicated by the lighter, dashed line. The image paths reaching the 1st, 2nd, 3rd levels, etc., correspond to paths of types *A*, *AB*, *ABA*, etc. (instant by instant); the same is true in the opposite direction (1st, 2nd, 3rd levels, etc. in the negative halfplane) for paths of types *BA*, *BAB*, etc.

distribution of $\vee|Y(t)|$, the maximum of the absolute values of Y in $[0, t]$. The ruin of one of the two gamblers within time t means, in fact, that, in the interval in question, Y reaches $\pm y$; that is that $|Y|$ reaches y.

In the case of Heads and Tails, we have

$$\mathbf{P}\left(\vee\left|Y(t)\right|\geqslant y\right)=q_n'+q_n''=2\sum_h(-1)^h\,p\left[(2h+1)y\right] \tag{8.52}$$

(there are, in fact, only a finite number of terms as we saw above). In the Wiener–Lévy case, $p(y)$ is given by equation 8.47 in Section 8.6.4, and hence

$$\mathbf{P}\left(\vee\left|Y(t)\right|\geqslant y\right)=2\sqrt{(2/\pi)}\sum_{h=0}^{\infty}(-1)^h\int\limits_{(2h+1)y/\sqrt{t}}^{\infty}\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x. \tag{8.53}$$

Differentiating with respect to y, we obtain the density

$$\begin{aligned} f(y) &= K\sum_{h=0}^{\infty}(-1)^h(2h+1) \quad e^{-\frac{1}{2}\left[(2h+1)y\right]^2/t} \\ &= K\left(e^{-y^2/2t} - e^{-(3y)^2/2t} + e^{-(5y)^2/2t} - e^{-(7y)^2/2t} + \dots\right), \\ K &= 2\sqrt{(2/\pi t)}. \end{aligned} \tag{8.54}$$

8.6.7. It is instructive to compare the present considerations, based on Desiré André's argument, with those, essentially identical, based on Lord Kelvin's *method of images*, which is often applied to diffusion problems. We have already observed (Chapter 7, 7.6.5) that the Heads and Tails process can be seen, in a heuristic fashion, to approach a diffusion process, and that the analogy becomes an identity when we consider the passage to the limit which transforms the Heads and Tails process into the Wiener–Lévy process.

In order to use the method to formulate the problem of the ruin of *one gambler* (occurring when level $y = c$ is reached), it suffices to find the solution of the diffusion equation (given by equation 7.32 of Chapter 7, 7.6.5) in the region $y \leqslant c$ (where we assume $c > 0$), satisfying the initial condition of concentration at the origin, *together with the condition that it vanishes at* $y = c$. By virtue of the obvious symmetry (giving a physical equivalent to Desiré André's argument), it suffices to place initially, at the point $y = 2c$, a mass equal and opposite to that at the origin (the 'cold source'); this gives a process that is identical to the other, but opposite in sign, and symmetric with respect to the line $y = 2c$ instead of to $y = 0$. On the intermediate line, $y = c$, the two functions therefore cancel one another out and their sum provides the desired solution. The probability of ruin at any instant can be interpreted in terms of the flux of heat out past the barrier, together with an incoming cold flux; and so on.

In the case of two gamblers (i.e. barriers at $\pm c$, and the initial position of the mass at $y = 0$, or at $y = a$, $|a| < c$[22]), if we are to use the same trick we have to introduce an *infinite* number of sources, alternatively hot and cold – like alternate images of the face and the back of the head in a barber's shop with two mirrors on opposite walls. We have an infinite number of images of the mirrors (lines $y = (2k + 1)c$, k being an integer between $-\infty$ and $+\infty$) and in between them an infinite number of strips (proper and reversed images of the shop), and inside each of these strips the image ('hot' or 'cold') of the source. If the source is at the centre ($y = 0$), the others are at $y = 2kc$ (hot if k is even, cold if k is odd); otherwise – if it is at $y = a$ – the hot sources are at $2kc + a$ and the cold ones at $2kc - a$ (still with hot corresponding to k even, cold to k odd).

Using the techniques of this theory (Green's functions etc.) one can obtain solutions to even more complicated problems of this nature (e.g. those with curved barriers), where this kind of intuitive interpretation would necessitate one thinking in terms of something like a continuous distribution of hot and cold sources.

22 This is just a more convenient way of saying that one starts from 0 but places barriers at $c - a$ and $c + a$.

8.7 Ballot Problems; Returns to Equilibrium; Strings

8.7.1. We turn now to what we referred to in Section 8.5.2 as the second group of questions concerning random processes; those in which the study of the process reduces to an examination of certain segments into which it may be useful to subdivide it. More precisely, we shall consider the decomposition into *strings*; that is into the segments between successive returns to equilibrium (i.e. between successive *zeroes* of $Y(t)$).

Over and above their intrinsic interest, these questions often point the way to the formulation, understanding and solution of other problems that entail *recurrence* in some shape or form. In other words, problems relating to processes that, after every repetition of some given phenomenon (in this case, the return to equilibrium), start all over again, with the same initial conditions (or with modifications thereof which are easily taken into account).

The simplest scheme involving the notion of recurrence is that of events E_h forming a *recurrent sequence*[23] (such as the $E_h = (Y_h = 0)$ in our example) for which, when the outcomes of the preceding events are known, the probabilities depend on the number of events since the last success. In other words, the index (or 'time') $T^{(k)}$ of the kth success is the sum of the k *independent* waiting times $T_t, T_2, \ldots, T_k$: the T are integers in this case, but this is merely a special feature of this simple scheme.

We now provide a brief account of the most important aspects of the theory (for a fuller account, see Feller, Vol. I, Chapter XIII). We denote by f_h the probability that E_h is the first success (or, equivalently, that, following on from the last success obtained, the first success occurs in the hth place); in other words, $f_h = \mathbf{P}(T = h)$, where T = waiting time. It follows immediately that either $f = \Sigma f_h = 1$, or it is <1 (if the probability of a success does not tend to certainty as the number of trials increases indefinitely). We adopt the convention of denoting the difference $1 - f$ by f_∞ (the probability that the waiting time is infinite).

By convolution, we obtain the probabilities $f_h^{(2)}$ of E_h being the second success, and so on for $f_h^{(3)}$ and the rest. In general, we have

$$f_h^{(r)} = f_1 f_{h-1}^{(r-1)} + f_2 f_{h-2}^{(r-1)} + \ldots + f_{h-1} f_1^{(r-1)}. \tag{8.55}$$

Summing over r, we obtain the probability u_h of E_h being a success (without taking into account whether it was the first, second, …, or whatever; as above, this holds also for a success at the hth place following on from some success already obtained, assuming that nothing is known about successes for subsequent events):[24]

$$u_h = f_h + f_h^{(2)} + f_h^{(3)} + \ldots + f_h^{(h)} \qquad \left(\text{obviously, } f_h^{(r)} = 0 \text{ for } r > h\right). \tag{8.56}$$

23 These are usually called 'recurrent events', but this terminology does not fit in with ours (nor, in a certain sense, with the point of view we have adopted).

24 I would prefer to write $\mathbf{P}(E_h) = p_h$ (instead of u_h) as usual, in order not to make it appear that we are dealing with a special case, and in order to avoid confusion with the standard use of u_n (as the maximum probability for Heads and Tails). The reason we have used u_h is for the convenience of the reader who wishes to pursue this topic (which we are only scratching the surface of) using Feller: however, it only occurs in this section, so that there should be no cause for any confusion.

The sum of the $f_h^{(r)}$ gives f^r (=1 or <1, the same as f), and provides the probability that an rth repetition takes place. The sum

$$f + f^2 + f^3 + \ldots + f^r + \ldots$$

therefore gives the prevision of the total number of successes (finite or infinite, according to whether $f < 1$ or $f = 1$). The same prevision can be expressed in a different way, however, by the sum of the u_h; we therefore obtain

$$u = u_1 + u_2 + \ldots + u_h + \ldots = f/(1-f), \tag{8.57}$$

$$f = u/(u+1). \tag{8.57'}$$

If $f = 1$, $u = \infty$, the events E_h are called *persistent*; in the opposite case, $f < 1$, $u < \infty$, they are called *transient*. In the case of persistent events,[25] if we denote the prevision of the waiting time by τ, that is

$$\tau = f_1 + 2f_2 + 3f_3 + \ldots + hf_h + \ldots \qquad (\text{which could be } \infty), \tag{8.58}$$

then, as h increases, the probability u_h of success tends to the limit

$$\bar{u} = 1/\tau \qquad (\text{in particlar, } u_h \to \bar{u} = 0 \quad \text{if} \quad \tau = \infty). \tag{8.59}$$

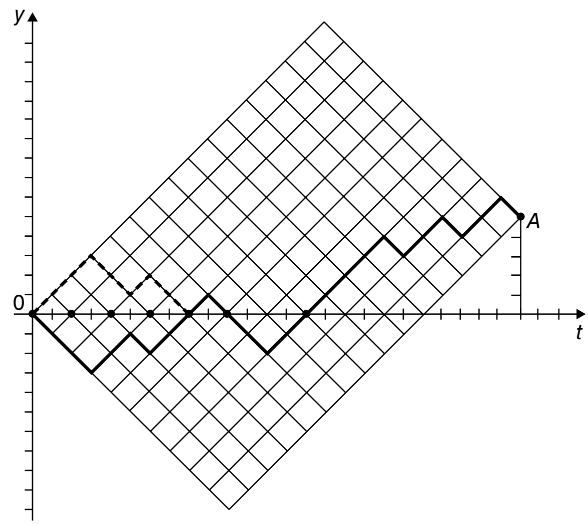
Let us now return to the central topic of this section and explain how the 'Ballot problem' enters into the picture. This is simply a traditional way of referring to, and interpreting, a set of problems similar to those encountered under the heading of 'gambler's ruin', but relating now to drawings from an urn without replacement (which provides a model for the process of counting votes). The results also find application in statistics, where they form the basis of certain criteria (due to Kolmogorov and Smirnov) for considering the deviation of an empirical distribution function from a given hypothetical theoretical distribution. Anyway, although we shall refer to Ballot problems as a convenient aid to intuition, we shall think of this new case always in the context of Heads and Tails.

In fact, we shall study problems of Heads and Tails conditional on the knowledge of the number of successes, H, occurring in the N trials with which we are concerned (we have seen this argument before in Chapter 7, 7.4.3, where we used it to derive the hypergeometric distribution). Proceeding in this way, it is evident that, among other things, the graphical representation of this problem consists simply of the rectangular portion of the Heads and Tails lattice having opposite vertices at the origin (the starting point) and at $[N, H] = (N, 2H - N)$ (the point where the process terminates).

There are $\binom{N}{H}$ possible paths joining these two points (Figure 8.8); in order to fix ideas, we assume that $Y_N = 2H - N \geqslant 0$, that is $H \geqslant N - H$. For $0 < y < Y_n$, all paths either

25 The complication of 'periodic events' (E_h only possible for a multiple of some 'period' λ) can be avoided (and we assume this done) by confining attention to events $E_{h\lambda}$. Of course, this might not always be convenient in practice (e.g., if a sequence defined in terms of another sequence A_n, which is not periodic, turns out to be periodic: see Heads and Tails; returns to zero are only possible for n even).

Figure 8.8 Desiré André's argument: the Ballot problem (i.e. the hypergeometric distribution). Paths from 0 to A which begin with a downward step correspond in a one-to-one fashion to those which start off upwards and then subsequently touch the t-axis (by symmetry in the interval before the t-axis is first reached).



cross or touch each barrier of the form y = constant. For the case $y = 0$ (and $y = Y_N$) we wish to consider how many paths touch it (either crossing it or not) after the initial $Y_0 = 0$ (or before the final $Y_N = N$, respectively). The same question (without the qualifications at the end-points) also arises for levels $y = Y_N + c$ ($c > 0$), or, equivalently,[26] for $y = -c$.

(A) The case $y = 0$, the Ballot problem, is the simplest one (we restrict ourselves to $Y_N > 0$ for now, postponing the case of $Y_N = 0$ to Section 8.7.3). All paths whose first step is downward must cross $y = 0$ again and, by reflecting the initial segment up to the first crossing, one obtains (with a one-to-one correspondence) all those having an upward first step which subsequently touch $y = 0$. But the first step has (like any other step) probability $(N - H)/N$ of being one of the $N - H$ downward steps; the probability of the eventual winner not being in the lead at some stage during the counting is therefore equal to twice this value, $2(N - H)/N$, and the probability that he is *always in the lead* during the counting is given by

$$1-2(N-H)/N=(2H-N)/N=Y_N/N. \tag{8.60}$$

(B) If we turn to the case $y = Y_N + c$ (or $y = -c$), $c > 0$, the principle of reflection (Desiré André) shows immediately that there are $\binom{N}{H+c}$ paths which touch this barrier in $0 \leqslant t \leqslant N$ (either crossing it or not, as the case may be). This is the number which finish up at the point whose ordinate is $(2H - N) + 2c = 2(H + c) - N$, the symmetric image of the given final point with respect to the barrier. In fact, these paths are obtained from the others (in a one-to-one onto fashion) by reflecting, with respect to the barrier, the final portion, starting from when the barrier is first reached (for $t = h$; $Y_h = Y_N + c$).

26 This is an obvious application of the reversal principle; see Section 8.6.3.

The probability s_c of reaching (and possibly going beyond) level $y = Y_N + c$ (or $y = -c$) is therefore given by

$$s_c = \binom{N}{H+c} / \binom{N}{H} = \frac{(N-H)(N-H-1)(N-H-2)\ldots(N-H-c+1)}{(H+1)(H+2)(H+3)\ldots(H+c)}. \tag{8.61}$$

The explicit expression is particularly useful when c is small (note that for $c = 1$ we have $s_1 = (N - H)/(H + 1)$), and is instructive in that it shows how the successive ratios $(N - H - c + 1)/(H + c)$ give the probability of reaching the required level $(Y_N + c)$ given that one knows that the level immediately below it (i.e. $Y_N + c - 1$) has been reached. The complementary probability is

$$(2H - N + 2c - 1)/(H + c),$$

and so the probability r_{c-1} that the *maximum* level reached is $Y_N + c - 1$ is given by the formula for s_c with $(2H - N + 2c - 1)$ replacing

$$(N - H - c + 1)$$

in the final factor; this is also the probability that the *minimum* level is $-(c - 1)$ (alternatively, one can obtain this by noting that $r_{c-1} = s_{c-1} - s_c$). Observe that $s_c = 0$ for $c \geqslant N - H + 1$ (Why is this so?).

(*C*) *The case of two barriers*. In the case of two barriers at levels $y = -c'$ and $y = Y_N + c''$ (c' and c'' positive), by performing successive reflections, as in the previous case, one obtains the paths terminating at the image points of the given final point ($Y_N = 2H - N$) with respect to the two barriers (thought of as parallel mirrors: there are an infinite number of images, but only those with ordinates lying between $\pm N$ can be reached). Setting $c^* = Y_n + c' + c''$, the distance between the barriers, the ordinates of the images are given by

$$(2k + 1)c^* - c' \pm c''.$$

(N.B.: for $k = 0$, we have $c^* - c' - c'' = Y_N = 2H - N$, the given final point, and

$$c^* - c' + c'' = Y_N + 2c'' = 2(H + c'') - N,$$

the unique image in the case of a single barrier $y = Y_N + c''$; in that case, we used c instead of c''.)

It follows that the probability of the lower barrier being reached first is given by

$$q'_N = \left(1/\binom{N}{H}\right)\left[\binom{N}{H+c'} - \binom{N}{H+c^*} + \binom{N}{H+2c^*+c'} - \binom{N}{H+3c^*} + \ldots\right] \tag{8.62}$$

(where we argue as in Section 8.6.5): the result for q''_N is similar (with c'' in place of c'). The sum, $q_N = q'_N + q''_N$, gives the probability of reaching a barrier (not distinguishing which, or which was reached first), and $1 - q_N$ that of not reaching either of them.

8.7.2. When studying a random process, it is often useful to consider it as subdivided into successive *strings*; that is into segments within which it retains the same sign. In our case,[27] this means those segments separated by successive *zeroes*, $Y_t = 0$, and necessarily having even length (since Y_t can only vanish for t even). Strings are either positive or negative (i.e. paths on the positive or negative half-plane, $Y_t > 0$ or $Y_t < 0$; see footnote 12 in Chapter 7, 7.3.2) and between any two strings the path has a zero at which it either *touches* the t-axis or *crosses* it, according to whether the two strings have equal or opposite signs.

If one thinks in terms of gains, that is of the excess of the number of successes over the number of failures, a zero represents a *return to equilibrium* (equal numbers of successes and failures; gain zero), and the string represents a period during which one of the players has a strict lead over the other. Omitting the word 'strict' and including the zeroes, we obtain periods in which one or other player does not lose the lead (i.e. the union of several consecutive strings having the same sign). One might also be interested in knowing the length of time, within some given period up to $t = N$, during which either player has had the lead. If one thinks of a random walk, the zero is a *return to the origin* and a string is a portion of the walk between two returns to the origin.

The above discussion, together with the results obtained so far, leads us directly into this kind of question, either with reference to the special case of Heads and Tails, or to that of the Ballot problem (which reduces to the former, if one thinks of $Y_N = 2H - N$ as being known).

8.7.3. *The Ballot problem in the case of parity:* $Y_N = 0$, *that is there are equal numbers,* $H = N - H = N/2$, *of votes for and against. What is the probability that one of the two candidates has been in the lead throughout the count*? In our new terminology, this means that the process forms a single string; that is there are no zeroes except at the end-points (if we are thinking in terms of a particular one of the candidates, the string must be of a given sign and the probability will be one half of that referred to in the question above).

This can easily be reduced to the form of case (A) considered in Section 8.7.1. In terms of the candidate who is leading before the last vote is counted, we must have $Y_{N-1} = 1$ (since we know that at the final step the lead disappears, and we end up with $Y_N = 0$); it follows that the required probability is given by

$$Y_{N-1}/\left(N-1\right)=1/\left(N-1\right)$$

This is the probability that one of the two candidates (no matter which) remains strictly ahead until the final vote is counted; the probability of this happening for a particular candidate is $1/2(N-1)$.

8.7.4. *What is the probability that in a Heads and Tails process – or, more generally, in an arbitrary Bernoulli process – the first zero* (*return to equilibrium, passage through*

27 That of processes with jumps of ± 1, with paths on a lattice, and (for convenience) starting at the origin, $Y_0 = 0$. In other cases, one could have changes in sign without passages through zero occurring (and one could even have, in continuous time, a discontinuous process $Y(t)$ with an interval in which changes of sign occur within neighbourhoods of each point).

the origin) *occurs at time* $t = n$ (where n, of course, is even)? For this to happen, it is first of all necessary that $Y_n = 0$; the problem is then that considered in Section 8.7.3. The probability that if a zero occurs it is the first one is therefore equal to $1/(n - 1)$. The probability of the first zero at $t = n$ is thus $\mathbf{P}(Y_n = 0)/(n - 1)$. In other words, this is the probability that the first *string* (and hence any string, since the process can be thought of as starting all over again after every zero) has *length n*.

(A) In the case of Heads and Tails, we have $\mathbf{P}(Y_n = 0) = u_n$, so the probability of the first zero occurring at $t = n$ is given by

$$u_n/(n-1) \simeq (0 \cdot 8/\sqrt{n})/n = 0 \cdot 8 n^{-\frac{3}{2}}.$$

The probability that the string is of length n and has a given sign (i.e. is to the advantage of a particular one of two gamblers) is one half of this.

More precisely, we have (setting $n = 2m$)

$$\frac{u_n}{n-1} = \binom{2m}{m} / 2^{2m}(2m-1), \text{ and also } \frac{u_n}{n-1} = \frac{u_{n-2}}{n} = u_{n-2} - u_n.$$

Since

$$u_n \cdot 2^n = \binom{2m}{m} = \frac{2m!}{m!m!} = \frac{2m(2m-1)(2m-2)!}{m \cdot m \cdot (m-1)!(m-1)!}$$
$$= \frac{4(n-1)}{n}\binom{2m-2}{m-1} = \frac{4(n-1)}{n} u_{n-2} \cdot 2^{n-2},$$

we can verify at once that $u_n = u_{n-2}(n - 1)/n$.

This establishes the following important conclusions:

(a) u_n is also the probability that there are no zeros up to and including $t = n$ (this is true for $u_2 = \frac{1}{2}$, and hence is true by induction, since $u_{n-2} - u_n$ is the probability of the first zero occurring at $t = n$);[28]
(a') as in the footnote;
(b) since $u_n \to 0$, the probability that (as the process proceeds indefinitely) there is at least one return to equilibrium tends to 1 (and the same is therefore true for two, three or any arbitrary k returns to equilibrium);
(c) the form u_{n-2}/n tells us that $1/n$ is the probability that the string terminates at $t = n$ (since Y_n becomes 0), assuming that it did not terminate earlier (since u_{n-2} is the probability that $Y_t \neq 0$ for $t = 1, 2, \ldots, n - 2$, and this is necessarily so for $t = n - 1 =$ odd);

28 It follows that the probability of Y_t ($0 < t \leqslant n$) being *always positive* (or *always negative*) is $u_n/2$. If, instead, one requires only that (a') is non-negative (or nonpositive), the probability is double: i.e. it is still u_n (as can be seen from Section 8.6.3; special cases of equation 8.43′ for $c = 0$ and $c = -1$).

(*d*) from this, we can deduce that u_n can be written in the form

$$u_n = \left(1-\frac{1}{2}\right)\left(1-\frac{1}{4}\right)\left(1-\frac{1}{6}\right)\left(1-\frac{1}{8}\right)\ldots(1-1/n);^{29} \tag{8.63}$$

in other words,

$$u_{n+2} = u_n\left(1-\frac{1}{n+2}\right) = \left(u_n\frac{n+1}{n+2}\right)$$

(as a product of complementary probabilities; a demographic analogy goes as follows: the probability of being alive at age n can be expressed as a product of the probabilities of not dying at each previous age);

(*e*) a further meaningful expression for u_n is given by

$$u_n = \sum_k \frac{u_k}{k-1}u_{n-k} \ \left(\text{the sum being over } k, k \leqslant n\right); \tag{8.64}$$

observe that, in fact, each summand expresses the probability that the first zero is at $t = k$, and that there is another zero at $t = n$ (i.e. after time $n - k$); for $k = n$ (the final summand), we must take $u_0 = 1$, and hence $u_n/(n - 1)$, a term which can be taken over to the left-hand side to give the explicit expression $u_n = [(n - 1)/(n - 2)]$ Σ' (where Σ' denotes the same sum, but without the final term).

We shall encounter further properties of u_n and $u_n/(n - 1)$ later.

(*B*) In the general case (the Bernoulli process with $p \neq \frac{1}{2}$) we have instead

$$\mathbf{P}(Y_n = 0) = \binom{2m}{m}(p\tilde{p})^m = u_n(4p\tilde{p})^m = u_n\left[2\sqrt{(p\tilde{p})}\right]^n, \ 2\sqrt{(p\tilde{p})} < 1.$$

The probability of the first zero at $t = n$ is, therefore,

$$\left[u_n/(n-1)\right]\left[2\sqrt{(p\tilde{p})}\right]^n < u_n/(n-1) \ \left(n \text{ even}\right), \tag{8.65}$$

and the sum of such probabilities is <1.

The remaining probability, P, given by $P(x) = 1 - \Sigma_n[u_n/(n - 1)]\ (1 - x)^n$ with $x - 1 - 4p\tilde{p}$, is the probability that a string has infinite length (which, with probability = 1, will be to the advantage of the favourite; i.e. the player with $p > \frac{1}{2}$). At the beginning of each new

29 We note that this enables us to establish Wallis's formula; from

$$u_{2m} = \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\frac{7}{8}\cdots\frac{2m-1}{2m} \simeq \frac{\sqrt{(2/2m)}}{\sqrt{\pi}},$$

we obtain

$$\sqrt{\frac{\pi}{2}} = \lim_{m\to\infty} \frac{2.4.6\ldots 2m}{3.5.7\ldots(2m-1)}\cdot\frac{1}{\sqrt{(2m)}},$$

where π has its usual meaning, given by the double integral of Chapter 7, 7.6.7:

$$\int e^{-\frac{1}{2}\rho^2}\,\mathrm{d}x\mathrm{d}y = \int e^{-\frac{1}{2}\rho^2}\rho\, d\rho\, \mathrm{d}\theta = 2\pi.$$

string, there is probability $(1 - P)$ of a finite string, advantageous to one or other of the players, and probability P that the favourite embarks on an infinite string. The probability that the kth string turns out to be infinite is given by $P(1 - P)^{k-1}$.

Note, in particular, that property (*b*) only holds in the case of Heads and Tails (i.e. *only* if we have $p = \frac{1}{2}$). Otherwise, it is not at all *asymptotically certain* that a return to equilibrium takes place (and even less is it certain that such a return to equilibrium takes place an arbitrarily large number of times). On the contrary, it is asymptotically certain that the *favourite* (the player with $p > \frac{1}{2}$) maintains his lead from some given time onwards.

Remark. We have introduced the phrase '*asymptotically certain*' to mean that some given fact – for example, in the case under consideration, the occurrence of a return to equilibrium, or of k returns to equilibrium – has a probability tending to 1 of occurring in a random process, provided the process goes on indefinitely; that is if p_N is the probability that it occurs before time N, then $p_N \to 1$ as $N \to \infty$.

We note that if some given fact is asymptotically certain, then its occurrence k times (k arbitrary) is also asymptotically certain, provided (as in our case) that each time it occurs we find ourselves with the same initial conditions.[30] Without this latter stipulation, the conclusion of (*b*) – 'and the same is therefore true for two, …' – no longer holds (this is obvious but was not mentioned explicitly in (*b*) for the sake of brevity).

We note also that 'asymptotically certain' in no way (logically) implies 'certain' provided the process continues indefinitely (not even if we were to assume that we could examine the process in its entirety, placing ourselves beyond the end of time). It is even more important to note that the fact that the occurrence of an event k times (with k arbitrarily large) is asymptotically certain *does not imply* that, in a process of infinite duration, its occurrence *an infinite number of times* is certain (necessary), nor even that it has probability 1 (nor even that it is probable, or even possible). We can only say that this number of repetitions N (assuming that it makes sense to speak about it) is a random quantity (either integer or $+\infty$), which has probability 0 of taking on any individual finite value, and hence of belonging to any given finite subset of integers, such as those less than some preassigned integer k. It could, however, be certainly finite, like an 'integer chosen at random' (see Chapter 4, 4.18.3).

8.7.5. *What is the prevision of the length L of a string* (*i.e. of the waiting time t = L until the first zero*)? In the case of Heads and Tails, we see immediately that $\mathbf{P}(L)$ is infinite. In fact, $n(u_{n-2}/n) = u_{n-2}$; that is the contribution to the prevision corresponding to $L = n$ tends to zero like $n^{-\frac{1}{2}}$, and the sum diverges.

As we remarked above (see the discussion following equation 8.65), if $p \neq \frac{1}{2}$ this no longer happens (because of the presence of the factor $2\sqrt{(p\tilde{p})} < 1$ in the geometric progression): the (finite) strings have, in prevision, finite length. However, the prevision becomes infinite if we take into account the fact that each string could be the *final* one,

30 We are referring, therefore, to a recurrent sequence of events (see Section 8.7.1).

of *infinite*[31] length (but, if we distinguish between the two players, this only happens for the favourite: the first player if $p > \frac{1}{2}$).

Remark. The result that we are interested in (for Heads and Tails) is more conveniently expressed in the following form (which also gives us the opportunity to make an observation of a more general character). Each string consists, in prevision, of $\frac{1}{2}$ a string of length 2, of $\frac{1}{8}$ a string of length 4, …, of $u_n/(n-1)$ a string of length n, and so on. In terms of the prevision of length, we have, in contrast, 2/2 = 1 for strings of length 2, $4/8 = \frac{1}{2}$ for strings of length 4, …, $n(u_{n-2}/n)$ for strings of length n and so on. Observe, in particular, that, for an individual string, the prevision of long strings is negligible (i.e. the prevision of strings $>n$ is less than any given $\varepsilon > 0$, provided n is taken sufficiently large), whereas, for the prevision of length (which is infinite), the prevision of the length of short strings (i.e. of those less than some preassigned, arbitrary, finite n) is negligible. It makes no difference if one makes the same statements but multiplies by 1000 (or a million, or whatever): 'out of 1000 strings, in prevision 500 have length 2, and their total length is, in prevision, 1000', and so on. Usually, one says 'on average'. We shall see later (Section 8.8.4) that there are dangers in using this form of expression.

8.7.6. *For the Ballot problem in the case of parity, what is the probability that one of the two candidates has never been behind during the count*? This is almost the same question as we asked in Section 8.7.3, except that we are here asking for something less: we admit the possibility that at some stages during the count the votes for the two candidates may also have been equal. It is sufficient that the lead of the candidate in question never becomes negative; that is that it never reaches the level $y = -1$. As in Section 8.7.3, we assume $Y_{N-1} = 1$; we are then back in case (B) of Section 8.7.1 and we can apply s_c for the case $c = 1$, which we have already given explicitly:

$$s_c = (N - H)/(H + 1),$$

where we have to put $N - 1$ in place of N and $N/2$ in place of H (since $Y_{n-1} = 2H - (N-1) = 1$). We hence obtain

$$(N-2)/(N+2) = 1 - 4/(N+2)$$

for the probability that level $y = -1$ is reached and $4/(N + 2)$ is then the required probability. If we are thinking in terms of one particular candidate, the probability that the lead is never negative is one half of this, that is $2/(N + 2)$.

Since the probability of the lead always being positive is $1/2(N - 1)$ (case (A) in Section 8.7.1), we obtain, by subtraction, the probability that the lead is always non-negative, but sometimes zero: this probability is equal to the previous one multiplied by

$$3(N-2)/(N+2).$$

31 It seems out of place here to complicate such expressions in order to repeat critical comments of the kind mentioned in the previous 'Remark' (following (B)).

In a different form, assuming that the lead is always non-negative, the probability that it is always positive is $(N + 2)/4(N - 1)$, and the probability that it is zero is $3(N - 2)/4(N - 1)$ (i.e., for large N, they are practically equal to $\frac{1}{4}$ and $\frac{3}{4}$, respectively).

Remark. We have seen that $2/(N + 2)$ is the probability of a given candidate being ahead (whether strictly or not) either *always* or *never* (i.e. for N or 0 steps, paths either all in the positive or all in the negative half-plane). The possible values for the number of steps when he is in the lead are

$$0, 2, 4, \ldots, N-2, N,$$

and since there are $(N + 2)/2$ possible values, their average probability must be $2/(N + 2)$. But this is the probability in the two extreme cases we have considered and, moreover, by an obvious symmetry, the probabilities for h and $N - h$ are equal. It follows that either they are all equal, or they have a strange wavy behaviour – with at least three turning points. Actually, they are all equal: in other words, *we have a discrete uniform distribution over the steps spent in the lead*. The proof is not as straightforward as the statement and will be omitted,[32] in order not to interrupt the discussion and overcomplicate matters. A further justification for this is that the previous considerations have already made the conclusion highly plausible.

8.7.7. *What is the probability that in the Heads and Tails process* (*or, more generally, in any Bernoulli process*) *the first crossing of* $y = 0$ (i.e. the first zero where the path does not simply touch the axis) *occurs at time* $t = n$ (*where* n, *of course, is even*)? In other words, we are asking for the probability that the duration of the initial period during which one of the players *never falls behind* is equal to n; that is that this is the sum of the lengths of the initial consecutive strings whose sign is that of the first string. In order for this to happen, it is first of all necessary that $Y_n = 0$, and that no crossings have taken place for $t < n$; in addition, we require that the first toss after $t = n$ (i.e. the $(n + 1)$st) is opposite in sign to the very first toss (and thus to the strings already obtained). The probability we seek is therefore given by

$$\mathbf{P}(Y_n = 0).\left[4/(n+2)\right].2p\tilde{p},$$

where $4/(n + 2)$ is the probability of no crossing occurring (as determined in Section 8.7.6), and $2p\tilde{p}$ is the probability that the first and $(n + 1)$st tosses have opposite signs. In the case of Heads and Tails, $p = \tilde{p} = \frac{1}{2}$, this reduces to $2u_n/(n + 2)$ (or to $u_n/(n + 2)$ if one specifies which of the two players is to be ahead initially). In the general case $\left(p \neq \frac{1}{2}\right)$, the probabilities (for each finite n) are smaller and one has the residual probability of the lead being maintained indefinitely (by the favourite, as for the strings). Apart from a comment of this kind – made for comparative purposes – we shall restrict ourselves to the case of Heads and Tails.

---

32 We merely note that it follows from equation 8.64 of Section 8.7.4(*e*), and that the arguments are similar to those mentioned in Section 8.7.10 (for the arc sine distribution).

If we compare the result obtained for the length L of a string with that for the lead, V say, we have

$$\mathbf{P}(L=n)=u_{n-2}/n,\ \mathbf{P}(V=n)=2u_n/(n+2);$$

from this it is clear that

$$\mathbf{P}(V=n)=2\mathbf{P}(L=n+2),$$

and hence that

$$\mathbf{P}\Big(V\geqslant n\Big)=2\mathbf{P}\Big(L\geqslant n+2\Big)=u_{n+2}.$$

It is instructive to consider the implications of this; on the one hand for the first few values (small values, corresponding to short strings) and, on the other hand, asymptotically (large values, corresponding to long periods of lead).

In the case of the first few possible (even) values, we have:

$$\begin{array}{llllll} n= 2 & 4 & 6 & 8 & 10\ldots \\ u_n = \frac{1}{2} & \frac{3}{8} & \frac{5}{16} & \frac{35}{128} & \frac{63}{256}\ldots \end{array}$$

and hence

$$\mathbf{P}(L=n)=u_n/(n-1)=\frac{1}{2}\,\frac{1}{8}\,\frac{1}{16}\,\frac{5}{128}\,\frac{7}{256}\ldots$$

and

$$\mathbf{P}(V=n)=2u_n/(n+2)=\frac{1}{4}\,\frac{1}{8}\,\frac{5}{64}\,\frac{7}{128}\,\frac{21}{512}\ldots$$

As we know from the last remark, the values in the final line are twice those of the penultimate line each shifted one place to the left (except for the final one, which equals $\frac{1}{2}$; doubling the remaining values, whose sum is $\frac{1}{2}$, we obtain again the total probability 1). This direct comparison shows that for $n = 2$ the probability for L is greater (as is obvious: in order to have $V = 2$, we must have $L = 2$ and, moreover, the subsequent string must be of opposite sign). For $n = 4$, they are equal, and subsequently the probabilities for V become greater $\left(\frac{1}{16}=\frac{4}{64}<\frac{5}{64};\frac{5}{128}<\frac{7}{128};\frac{7}{256}=\frac{14}{512}<\frac{21}{512}\right)$. All this could be seen directly, by simply noting that the ratio $2(n-1)/(n+2)$ is equal to $2-6/(n+2)$.

For large values, this ratio is (asymptotically) equal to 2, and, in any case, $\mathbf{P}(V\geqslant n)\simeq 2\mathbf{P}(L\geqslant n)\simeq 2(0\cdot 8/\sqrt{n})=0\cdot 8/\sqrt{(n/4)}$. This can be expressed by saying that, in a sense, long periods in the lead are four times as long as long strings (more precisely, this is true in the sense that V has the same probability of reaching some (long) length n as L has of reaching length $n/4$).

8.7.8. *For the Ballot problem in the case of parity, what is the probability distribution of the maximum lead attained during the count by a particular candidate? What is the probability distribution of the absolute value of the lead? And why does it become conditional on the fact that a given candidate never lost the lead throughout the count? Or was strictly in the lead throughout the count*? Clearly, if we drop the reference to the Ballot problem, we see that we are dealing with the most general question of the probability distribution of $\vee Y_N$, or of $\mathrm{V}|Y_N|$ (either for Heads and Tails, or for any Bernoulli process), assuming $Y_N = 0$, and possibly, also, $Y_t \geqslant 0$, or even $Y_t > 0$, for $0 < t < N$. This last assumption is the most restrictive and it amounts to seeking the probability distribution of the maximum *in a single string*. Under the next to last assumption, we could be dealing with a segment composed *either of a single string or of several consecutive strings all having the same sign*. In the general case, on the other hand, the segment might consist of a single string or of several, with arbitrary signs: it is only in this latter case that we need to distinguish between $\vee Y_N$ and $\mathrm{V}|Y_N|$.

In fact, these are simply variants of problems (A) and (B) in Section 8.7.1. We shall consider them separately.

(a) The only assumption is that $Y_N = 0$ and we seek the distribution of the maximum of Y_t. From (B) of Section 8.7.1, we see, taking $H = N/2$, that $s_0 = 1$, and, for $c \geqslant 1$,

$$s_c = \mathbf{P}\left(\vee Y_N \geqslant c\right) = \frac{N(N-2)(N-4)\ldots(N-2c+2)}{(N+2)(N+4)(N+6)\ldots(N+2c)} \qquad (8.66)$$
$$\left(s_c = 0 \text{ for } c \geqslant (N+2)/2\right),$$

$$t_c = \mathbf{P}\left(\vee Y_N = c\right) = s_c - s_{c+1} = \frac{4c+2}{N+2c+2} s_c \qquad (8.67)$$

(in particular, $r_0 = 2/(N + 2)$, as we already know). Applying Stirling's formula as given in equation 7.30 of Chapter 7, 7.6.4, we have, approximately (for c large, but $2c/N$ small, i.e. N much larger still), $s_c \simeq e^{-2c^2/N}$, $r_c\left(4c/N\right)\mathrm{e}^{-2c^2/N}$.

(b) Continuing with $Y_N = 0$ as the only assumption, we seek the distribution of the maximum of $|Y_t|$. Arguing as in (C) of Section 8.7.1 but also taking into account the symmetry ($N = 2H$; i.e. $Y_N = 0$, $c' = c'' = c$, $c^* = 2c$), we find that the probability $\overline{s}_c$ of either reaching or crossing $\pm c$ (which was denoted in (C) by $q_N = q'_N + q''_N$, here $q'_N = q''_N$) is given by

$$\overline{s}_c = \left(2/\tbinom{2H}{H}\right)\left[\binom{2H}{H+c} - \binom{2H}{H+2c} + \binom{2H}{H+3c} - \binom{2H}{H+4c} + \ldots\right]. \qquad (8.68)$$

Expressed more simply, using the s_c from (a) above, we have

$$\overline{s}_c = s_c - s_{3c} + s_{5c} - s_{7c} + \ldots, \qquad (8.68')$$

and similarly

$$\overline{r}_c = r_c - r_{3c} + r_{5c} - r_{7c} + \ldots.$$

Asymptotically, we therefore see from the previous expression that

$$\overline{s}_c = 2\sum_{k=1}^{\infty}(-1)^{k+1}\,\mathrm{e}^{-2k^2c^2/N},\ \overline{r}_c = (8c/N)\sum_{k=1}^{\infty}(-1)^{k+1}\,k\mathrm{e}^{-2k^2c^2/N}. \tag{8.68''}$$

(c) Let us now assume, in addition to $Y_N = 0$, that Y_t has not changed sign throughout $0 \leqslant t \leqslant N$ and, in order to fix ideas, let us assume that it is non-negative. It therefore makes no difference whether we talk in terms of the maximum of $|Y_t|$ or of Y_t (or of $-Y_t$, had we made the opposite assumption). We again argue as in (C) of Section 8.7.1, but now with $c' = 1$, $c'' = c$, in order to obtain the probability that Y_t always remains strictly between -1 and c. Dividing this by $2/(N + 2)$ (the probability that $0 \leqslant Y_t$, i.e. that $-1 < Y_t$), we obtain the probability of $\vee Y_t < c$ conditional on the given hypothesis; that is the $1-\overline{s}_c$ of the present case. If we use s_c to denote the same probability as in case (a), we have, therefore,

$$\begin{aligned}\overline{s}_c &= 1-\frac{N+2}{2}\\ &\times\left(s_1+s_c-s_{c+2}-s_{2c+1}+s_{2c+3}+s_{3c+2}-s_{3c+4}-s_{4c+3}+\ldots\right),\ \text{etc.}\end{aligned} \tag{8.69}$$

(d) This proceeds similarly: under the assumption of having been strictly in the lead ($Y_t \geqslant 1$, $0 < t < N$), we can reduce to the previous case by taking the axis $y = 1$ as the base line on the interval from $t = 1$ *to* $t = N - 1$ ($Y_1 = Y_{N-1} = 1$); for large N, the difference is very small.

8.7.9. *Similar problems for an arbitrary segment of the Heads and Tails process* (i.e. for a segment $0 \leqslant t \leqslant n$, where we do not assume, as in the previous examples, that $Y_N = 0$). The segment could consist of one string, or several strings, or none, and, in general, it will end in an incomplete string. We shall give a brief review of certain problems and their solutions, in order to draw attention to various of the points which need considering.

(a) So far as *periods in the lead* are concerned (see Section 8.7.6 and the *Remark*), we know, from Section 8.6.3, that u_n is the probability of Y_t remaining non-negative (in $0 \leqslant t \leqslant n$); that is that the lead is maintained for n steps out of the n (and the same holds true, obviously, for 0 out of n). Assuming n to be even, the number of steps spent in the lead can be any of

$$0, 2, 4\ldots, n-2, n.$$

We therefore have $(n + 2)/2$ possible values and the average probability is $2/(n + 2)$. However, the extreme cases have probabilities $u_n > 2/(n + 2)$,[33] so it is likely (by the same kind of argument that the *Remark* of Section 8.7.6 led us to believe that in that case, a segment consisting of complete strings, the probabilities were equal) that, in the general

33 Since it is the maximum of $n + 1$ probabilities $\omega_h^{(n)}\,(h = 0,\ldots,n)$, u_n is certainly $>1/(n + 1)$. It is, in fact, much greater than this, becoming more and more so as n gets larger: asymptotically, $u_n \simeq (2/(n+2)).0.4\sqrt{n}$.

case, there is a very small probability that a subdivision into periods of lead will consist of nearly equal lengths, and much greater probabilities for the less equal subdivisions. There is another consideration that makes this plausible. We already know that for the segment leading up to the last zero we have equal probabilities for all subdivisions, and that, from the last zero on, the lead does not change hands. In any case, the fact is that this turns out to be true and, in precise terms, we obtain $p_h = u_n u_{n-h}$ (h and n even) for the probability of being in the lead for h out of the n steps. The proof is much more difficult than one might expect from the simplicity of this formula, and we shall omit it.[34] We shall restrict ourselves to a consideration of how this probability behaves.

Recalling that $u_{h+2}/u_h = (h+1)/(h+2)$, we see that the ratio

$$p_{h+2}/p_h = (h+1)(n-h)/(h+2)(n-h-1)$$

is less than, equal to, or greater than 1, according to whether $h + 1 \lesseqqgtr n/2$: taking the asymptotic expression for u_n, we have $p_h = u_h u_{n-h} = (2/\pi)\sqrt{}/[h(n-h)]$. In the limit, we can say that the proportion of time during which a gambler is ahead, in a long period of play, is a random quantity X, whose probability distribution has density $f(x) = 1/(\pi\sqrt{}/[x(1-x)])$. This is the '*arc sine*' distribution, so-called because the distribution function, $F(x) = \int f(x)\mathrm{d}x$, is equal to $(2/\pi)\sin^{-1}(\sqrt{x})$, which might be better written as

$$(1/\pi)\cos^{-1}(2x-1).$$

We shall come back to this again and again.

(b) We know that the *probability distribution of* Y_n is the *Bernoulli* (or binomial) distribution ($p_h = \omega_h^{(n)}, h = 0,1,\ldots,n$), provided we know only that $Y_0 = 0$ (this holds similarly if we are given certain values, of which the last one is $Y_k = y$ with $k < n$; we then have $p_{y+h} = \omega_h^{(n-k)}$). The distribution is the *hypergeometric* if, in addition to Y_0, we are also given a value Y_N, $N > n$ (and similarly if two arbitrary values are known, one before and one after; $Y_{t'}$ and $Y_{t''}$, say, with $t' < n < t''$).

In general, we can say that any change in the state of information produces a change in the probabilities. In particular, if, in addition to knowing the initial value $Y_0 = 0$ (and, possibly, a subsequent value, or two arbitrary values, one before and one after), one also knows that Y_t has remained non-negative throughout $0 \leqslant t < n$, we have a range of possibilities as discussed above. The same holds if we have non-negativity throughout the entire process, $0 \leqslant t \leqslant N$, or throughout $t' \leqslant t \leqslant t''$ in the case where we know the values at the two points on either side, or even only for $t' \leqslant t < n$, or $n < t \leqslant t''$. Different probabilities also result if we assume the process to be strictly positive or strictly negative, or above or below some given level, or in between two levels, and so on. All this is obvious, but it needs emphasizing, and should be borne in mind.

(c) *The probability that level* $y = c$ *is reached for the first time at* $t = n$ ($Y_n = c$, $\wedge Y_{n-1} < c$, $c > 0$; symmetrically if $c < 0$) *is equal* (by the principle of reversal the problem is unchanged) *to the probability that* $Y_n = c$ *without there being any zeroes beforehand* (that is all Y_t, $0 < t < n$, have the same sign as $c = Y_n$). This probability is given by $\mathbf{P}(Y_n = c)$ multiplied by c/n, where c/n is *the probability that*, the value c of Y_t at $t = n$ being known,

34 See the notes in Section 8.7.10.

this level c has been reached for the first time at that point, and also the probability that the starting level ($y = 0$) has not been reached again; that is that Y_t *has always had the same sign as c*.

If we denote by H one or other of the two hypotheses mentioned, we have that $\mathbf{P}(H) = u_{n-1}/2$ (Section 8.6.3) and that $\mathbf{P}(H|Y_n = c) = c/n$ (the Ballot problem: case (A) of Section 8.7.1). The probability that we seek (which can also be obtained as the probability of ruin occurring precisely at the nth toss, $p_n(c) = q_n(c) - q_{n-1}(c)$; see Sections 8.6.1 and 8.6.3) therefore has the value stated:

$$\mathbf{P}\left[(Y_n = c).H\right] = (c/n)\mathbf{P}(Y_n = c) = (c/n)\omega_h^{(n)} \quad (c = 2h - n, c > 0). \tag{8.70}$$

The probability distribution of Y_n conditional on H (one of the two hypotheses) is, therefore, proportional to $c\omega_h^{(n)}$; in fact, we have

$$\begin{gathered}\mathbf{P}(Y_n = c|H) = \mathbf{P}\left[(Y_n = c).H\right] / \mathbf{P}(H) = (2/nu_{n-1})c\omega_h^{(n)} \\ (c = 2h - n > 0).\end{gathered} \tag{8.71}$$

Expressed in words: the probabilities of the possible values c for Y_n are altered (by the condition H) in a manner proportional to c (those for $c \leqslant 0$ are clearly 0), and the normalization factor K has been found explicitly in passing. The same result (and proof) goes through for the opposite hypothesis and provides the probability that $Y_n = c$ given that Y_n is greater than any value obtained previously (for $0 \leqslant t < n$, all Y_t are $<Y_n$; in this case, of course, we do not exclude the possibility of negative values). Asymptotically, we have a distribution of the form $f(x) = Kxe^{-x^2/2} (x \geqslant 0)$.

(d) We now restrict ourselves to the special case of knowledge of one value before and one after, $Y_0 = 0$ and $Y_N = 0$, together with the condition that $Y_t > 0$ throughout the given interval. The distribution of Y_n (for any integer n, $0 < n < N$) is obtained in a similar way, by observing that the probability that $Y_n = c$ ($c = 2h - n > 0$), and that the given conditions hold, is, setting $N = 2H$, given by

$$\left[\binom{H}{h}\binom{H}{n-h}\Big/\binom{2H}{n}\right].(c/n).\left(c/(N-n)\right)\Big] \tag{8.72}$$

(the product formed by taking $\mathbf{P}(Y_n = c)$ from the hypergeometric case and multiplying it by the probability that the process does not vanish in the passage from 0 to c during the first n and subsequent $N - n$ steps). We note, however, that this probability can also be thought of as the product of the p_h we are after, multiplied by the probability of the hypothesis that at time $t = N$ we obtain the first zero, the sign having previously been positive. This latter probability is equal to $u_n/2$, so we obtain

$$p_h = \left[\frac{2}{u_n n(N-n)\binom{N}{n}}\right] c^2 \binom{H}{h}\binom{H}{n-h} = K.c^2.\bar{\omega}_h^{(n)} \tag{8.73}$$

(we have placed a bar over the ω in order to stress the fact that it is the ω of the hypergeometric rather than the Bernoulli process, as above). One should note, however, the meaningful analogy between the two (every condition of positivity, on the left and on the right, entails a modification proportional to c). Asymptotically, we have a distribution of the form $f(x) = Kx^2 e^{-x^2/2} (x > 0)$.

8.7.10. *Remark*. It is instructive to consider in more detail some of the problems which lead (asymptotically) to the arc sine distribution (see Section 8.7.6 and the *Remark* in Section 8.7.9 (a)). We have omitted the proofs and we suggest that the reader refers to the third edition of Feller, Vol. I (1968). Comparison with earlier editions will reveal the simplifications in formulation that took place from one edition to the other, partly as a result of greater insight into the problems and their connections with one another, and partly for purely fortuitous reasons (see Chapter III, Section 4 of Feller, 'Last visit and long leads', and, in particular, the historical notes on pages 78 and 82).

The arc sine distribution was considered by P. Lévy (1939) in connection with the Wiener–Lévy process (see Section 8.9.8).[35] The application to Heads and Tails and other cases (obvious in the asymptotic case) seemed to require 'mysterious' forms of explanation, until their combinatorial character was revealed (by Sparre Andersen, in 1953). The methods used were quite complicated, and still were in the first edition of Feller (where they were due to Chung and Feller); the new feature, which arises in the passage from the second to the third edition, lies in the preliminary statement of a simpler theorem, which, from a qualitative viewpoint, begins to explain the weighting towards very unequal subdivisions of periods in which the lead does not change hands.

We can prove this in just a few lines. *The probability that in* $2m$ *tosses at Heads and Tails the final return to zero,* $Y_t = 0$, *occurs at* $t = 2k$ *is given by* $u_{2k}u_{2m-2k}$ (which is the discrete version of the arc sine distribution). In fact, the probability that $Y_{2k} = 0$ is u_{2k}, and the probability of no zeroes in the $2m - 2k$ subsequent tosses is u_{2m-2k} (see Section 8.7.4(a)). This is trivial, and yet the theorem is new (according to Feller); moreover, it was discovered by chance, experimentally, on the basis of observed statistical properties of random sequences produced by a computer. These were detected by capable mathematical statisticians, who then simply pointed out, and subsequently proved, that the distribution was symmetric (without realizing that it was the arc sine distribution).

This is by no means intended as in any way disrespectful to the number of authors who have made valuable contributions to this topic. It merely goes to show that tucked away in the vast rock-pile of problems there is the odd nugget lying unobserved; once noticed, of course, it appears obvious.

The following little calculation, which I made simply out of curiosity, may be new and possibly of some interest. I observed that it was not appropriate to call one gambler 'luckier' simply because he has led for most of the game so far (see the footnote to

35 I (vaguely) remember that an obvious property of the arc sine distribution – the density taking its maxima at the end-points – was considered paradoxical, even in cases where it was natural, as for observations of periodic phenomena (for example, in the case of a river flooding, the level remains around the maximum longer than it does around intermediate values, which are passed through more rapidly, both when the level is increasing and decreasing): see Figure 8.9. Of course, when periodicity is crude (for example, seasonal temperature changes, where maxima and minima vary from year to year) there are smoothed peaks, or sequences of peaks.

Section 8.8.1): it might well turn out that his luck runs out at the end of the game – 'he who laughs last laughs longest'. The probability of this happening is given by

$$1/\sqrt{3\pi} = 0\cdot 184 = \int_{\frac{1}{2}}^{1} \left(2\pi t\sqrt{\left[t\left(1-t\right)\right]}\right)^{-1} \mathrm{d}t. \tag{8.74}$$

To see this, let t denote the time when the final zero occurs (taking [0, 1] as the whole interval). If $t < \frac{1}{2}$, the one who is leading at the end is also the one who has spent most time in the lead. If $t > \frac{1}{2}$ in order for the one who is leading at the end to have spent most time in the lead, he must previously (i.e. before time t) have spent at least an additional time $t - \frac{1}{2}$ in the lead. Because of the uniformity of the distribution of the lengths in an interval between two zeroes, this, for given t, has probability $\left(t-\frac{1}{2}\right)/t = 1-\frac{1}{2}t$. This leaves a probability $\frac{1}{2}t$ for the opponent, conditional on t (having the arc sine distribution) being greater than $\frac{1}{2}$; we thus obtain equation 8.74)

8.8 The Clarification of Some So-Called Paradoxes

8.8.1. We have already found (and will do so again) that certain of the conclusions we have arrived at have had a paradoxical air, or, at any rate, have been easy to misinterpret. We have discussed various examples where such misinterpretation arises and, in so doing, have attempted to clarify the issues involved. In particular, we recall the laws of large numbers and, in the gambling context, the long expected time to ruin. The topics we have just been considering also lend themselves to discussions of this kind. Indeed, it is hard to decide whether their main value lies in the knowledge they provide, and the light they throw on a number of important theoretical and practical questions, or in the opportunity they give one to clear up a number of misconceptions and confusions, which otherwise could make one rather wary of entering into the probabilistic domain at all.

In those aspects of the Heads and Tails process that we have just been studying, it surely seems rather strange and mystifying that some kind of 'stationarity' or regularity does not hold. In particular, why is there not a tendency for the periods of unbroken lead to be equally distributed in the two opposite directions (all the more so after having seen that the process can be considered as an indefinite sequence of *strings*, at the end of each of which the process begins all over again under identical conditions)?

In particular, since the alternation of strings in the two directions (i.e. in the positive and negative half-planes) is itself a Heads and Tails process when the strings are considered as 'tosses', it seems obvious that the balancing of periods in the lead should hold by analogy with the balancing up of the frequencies of Heads and Tails. In actual fact, this conclusion is true if one considers the *number* of strings giving the lead in one or other of the two directions, but it is *not* true if one wishes to consider the respective total *durations* of periods in the lead. In fact, we have seen (see the *Remark* in Section 8.7.6) that in an interval formed by complete strings (i.e. those ending in a zero) all durations are equally probable, instead of, as one might have expected, those of intermediate length being more probable (i.e. we have a distribution into almost equal periods). In the general case (where the final string may not be complete, i.e. the interval does not end in a zero; see the *Remark* in Section 8.7.9), the situation is, in fact, the very opposite; it is the most unbalanced distributions which are most likely.

Both the form of the density, $f(x) = K/\sqrt{[x(1-x)]}$, and that of the distribution function, $F(x) = K \cos^{-1}(2x - 1)$, show clearly that the extremely asymmetric values are favoured (although in a symmetric fashion so far as the two opposite directions are concerned). The best way to visualize the result is to note that the splitting of the total duration of a long game into the fractions x and $1 - x$ of the duration in which one or other of the two gamblers is ahead can be thought of as brought about by choosing 'at random' (i.e. with uniform density) a point on the circumference of the semicircle having the segment [0, 1] as diameter, and then obtaining x by projecting that point onto this diameter. In other words, if the circumference is divided into an arbitrary number of equal arcs, their projections onto the diameter (which will clearly be smaller the nearer they are to the end-points) are equally probable (because they contain the point dividing the two parts x and $1 - x$).

The reader should now examine Figures 8.9c, 8.9b and8.9a (in reverse order, from bottom to top), together with the notes which accompany it.

Feller (Vol. I, Chapter III) provides numerical data that illustrate this phenomenon and make clear why it is not, in fact, surprising. Imagine that two gamblers play continuously for a year (making a toss every hour, minute or second; it does not matter which). It turns out that there is only a 30% probability of both being ahead for more than 100 days (about 28% of the total time), whereas there is a 50% probability that one of them remains ahead for less than 54 days (15% of the time), 20% that he remains ahead for less than nine days (2·4% of the time), 10% for less than 2.25 days (i.e. less than 0·6% of the time – more than 99·4% for his opponent!).[36] Feller also provides the details of the behaviour resulting from a computer experiment.

8.8.2. It is not really surprising that these numbers are not what we would imagine intuitively. Intuition cannot guide us – not even roughly sometimes – in foreseeing the results from analyses of complicated situations. This is precisely why mathematics is so useful, particularly in probability theory.

We should ask ourselves, however, whether, even from a qualitative point of view, the above conclusions are paradoxical (and, if so, for what reasons), and how one might set about correcting and altering this impression by showing that it is, in fact, perfectly natural for things to be thus. Despite the fact that the example which has given rise to this discussion is an especially striking one, it is by no means a unique and isolated case and it provides us with an excellent basis for discussion and considerations relating more or less directly to more general problems. On the other hand, it is not so much the individual result itself that merits and requires illustration but rather the nature of random processes which – like the very simple case of Heads and Tails – are based on the simple idea of stochastic independence (or lack of memory, if one prefers to think of it in this way). Although this is a simple notion, it is difficult to understand it sufficiently well to avoid finding certain of its consequences paradoxical. We have already commented upon this on a number of occasions, some of which we recalled above (the laws

36 Sometimes people refer to 'the less fortunate' player. This is not quite right, however, since it is possible (although not very probable) that the one who has been in the lead for most of the time finds himself behind at the end (see Section 8.7.10).

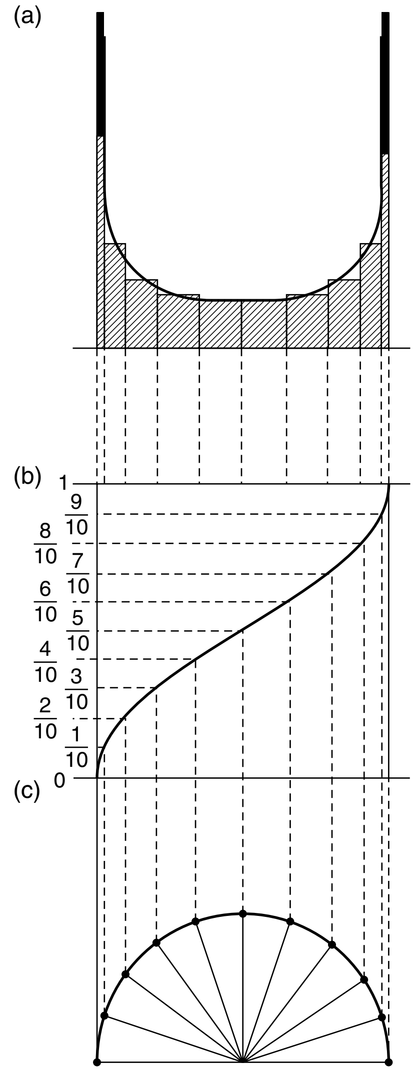


Figure 8.9 These should be read in reverse order (i.e. (c), then (b) and finally (a)).

a) The density of the arc sine distribution. The histogram shows the average density in each of the intervals between the deciles. The graph shows the density, whose equation, taking the interval to be [0, 1], is

$$f(x) = K/\sqrt{[x(1-x)]},$$

and is infinite at the end-points.

b) The distribution function of the arc sine distribution (obtainable using the device shown in (c)).

The abscissae shown correspond to the 'deciles' (see Chapter 6, 6.6.6) since they are obtained from the corresponding ordinates.

The ten intervals between the deciles are equally probable (with probability $\frac{1}{10}$). Note how much more dense the probability is near the end-points.

c) The probability distribution of the projection (onto the diameter) of a point 'chosen at random' (with uniform density) on the circumference.

This distribution occurs, for example, if one measures, at a 'random' instant, the position (or velocity) of a point performing harmonic oscillations.

The division of the circumference into 10 equal parts (18°) gives the deciles.

of large numbers and the long-expected time to ruin). These are – like the present example, and others which we shall soon come across – topics that are interrelated and deal with the same kinds of questions.

The reasons why these results appear paradoxical are all related to various kinds of distortion of the relations between probability and frequency:

by assuming connections without taking into account that they only exist under certain restrictive conditions;
by thinking that they virtually entitle one to make a prediction rather than a prevision;
by assuming that they systematically fall into familiar patterns of statistical 'regularity';
by having such a strong belief in such regularity as to make of it an autonomous principle; this leads one, inadvertently, to expect 'compensations' to take place in a

more extreme form and providential manner and sense than can be derived legitimately from probabilistic assumptions.

The danger of falling into these traps is even greater when one has been taught statistical concepts in a grossly oversimplified form, easily misunderstood and without the necessary warnings being given. The use of certain forms of terminology – for example, saying that something occurs *on average* a given number of times per unit time (instead of saying *in prevision*) – can lead one to regard such 'regularities' as certain, instead of merely being probable; that is, as predictions instead of previsions.

If 'regularity' is assumed as an 'article of faith' (and there is a book on statistics, inspired by this outlook, which is entitled 'Gleichförmigkeit der Welt'), how can it be, one might ask, that phenomena like returns to equilibrium and the distribution of leads could violate this regularity, thus challenging the supreme dictate of the ordered universe? If one thinks of returns to equilibrium (which are practically certain) as revealing a tendency towards, or desire for, such 'regularity', one would expect that a particle, having gone a long way below the origin while performing a random walk, would make an about turn in order to return to the fold. On the contrary, it has no memory and there is no fold to return to. It might carry on until it is twice as far from the origin before it turns back towards it, or it might have only gone half as far. If it does end up by going back to the origin with certainty, this is simply because, being on a random walk, it will sooner or later pass through all the points (but without any possibility of recognizing the point we have labelled the 'origin', nor any desire to do so).[37] One would have a stronger case (and, indeed, a valid one were it not for omitting to point out the fact that the expected duration is infinite, or, at least, for not taking it into account) if one were to argue that the phenomenon should reproduce itself 'regularly' because after each return to the origin a new string begins under precisely the same conditions.

If we attempt to identify and explain those reasons that we assume to underlie the tendency to talk in terms of 'paradoxes', we find the answers staring us in the face. The probability–frequency relation as it occurs in the law of large numbers should not be assumed to hold, since the successive Y_n are not independent. They depend upon a 'cumulative effect', which tends to be dominant; deviations take place only slowly and returns to equilibrium and changes of sign, that is of 'lead', only seldom. We have already mentioned above the idea of some kind of restoring force causing returns to equilibrium. However, only the last point is really important, because it pin-points a subtle and basic difference (whereas the other points simply caution one against the possibility of trivial and rather absurd misunderstandings).

The very fact that the probability of a return to the origin at time $t = n$ *tends to zero* (for n even, $u_n = 0{\cdot}8/\sqrt{n}$) should be sufficient to rule out the 'regularity' or 'stationarity' of behaviour that is the implicit and unconscious assumption occasioning all the 'astonishment' at these 'paradoxes'. The latter appear as such simply because they do not fit

37 I do not mean to imply that fallacious ideas of this kind are accepted statistical doctrine in some other approach differing from the one we follow. However, the environment created by a few introductory sentences followed by empirical clarifications etc. does not seem to be sufficiently antiseptic to prevent the germs of these dangerous distortions from multiplying in the subconscious.

into that particular framework; a framework that has created, in its own image and likeness, psychological attitudes whose tendency to rise to the surface becomes, in the absence of any process of re-education, general and indiscriminate. The following points may serve to provide a better appreciation of just how 'sensational' are the consequences of this probability tending to zero (a fact which is so simple when it is accepted as it is, without further thought). In a consecutive sequence of k tosses (e.g. 100 000), the probability that Heads always occurs from some n onwards is very small, but nonetheless finite; and this is also true for the probability that, starting from n, the sequence 11001… (where 1 = Heads and 0 = Tails) represents, in the binary code, either the first 100 000 decimal places of π, or the text of the Divine Comedy, or that it reproduces the initial segment obtained with the first k tosses, or any other preassigned segment of fixed length. Something of this kind occurs, in prevision, once every 2^k tosses, and it is practically certain that it occurs at least once within every segment of length N, if N is considerably larger than 2^k (and also that it occurs at least 10 times, or at least 1000 etc., provided we take a long enough sequence; the details are straightforward and we shall not bother with them here). On the other hand, the expected number of returns to equilibrium with a given number, N, of consecutive tosses starting from $t = n$ is, approximately,

$$\left(0\cdot 8/\surd n\right).N/2 = 0\cdot 4N/\surd n \to 0,$$

and the probability of at least one is even smaller. This means that, if one proceeds far enough to have an interval sufficiently long to give a non-negligible probability of containing a return to equilibrium (e.g. 1% or 10%), then one has an interval which almost certainly (e.g. with probability 90% or 99 %) contains the Divine Comedy at least once, and, if one carries on, at least 10 times, 1 million times, and so on, indefinitely.

8.8.3. Turning to a consideration of the values of Y_t (and not only at those instants for which $Y_t = 0$), a topic we shall be dealing with shortly, we can deduce an immediate and straightforward result about the extreme length of the strings at times far away from the starting time $t = 0$. We know that $|Y_n|$ has probability $\simeq 0{\cdot}8\ M/\surd n$ of being less than a preassigned M. For sufficiently large n, it is therefore almost certain that $|Y_n| > M$, in which case the string containing the instant $t = n$ *necessarily* has length $>2M$. In fact, the length would equal $2M$ under the assumption that the increase from the previous zero to Y_n and then the decrease to the following zero take place in an unbroken sequence of M successes and M failures, respectively. From considerations made in the context of the ruin problem, however, we know that it is probable that the increase and the decrease take much longer.

But the above remarks only have an illustrative and introductory value: they help us to see what is happening but they do not yet provide us with an explanation; neither do they resolve the confusion by going back to the source. At most, there is a restatement of the problem: instead of asking ourselves *why do the lengths of the strings become longer and longer* (despite the fact that they begin again from zero under the same conditions), we can ask why, given the same assumptions, do the ordinates Y_n become larger and larger (in absolute value); that is *why do the strings get further and further away from the axis* $y = 0$ (and it is clear that the two questions, even if not identical, are closely related).

Let us study then the history of each string. We might as well consider the first one, starting at $t = 0$. The probability that its length is $n = 2m$, that is that it terminates at the nth toss (with a return to equilibrium), is

$$u_n / (n-1) = u_{n-2} / n.$$

But u_{n-2} is the probability that no zero has previously occurred, so $1/n$ is the probability that the string terminates at time *n assuming that it does not terminate earlier* (for n even; otherwise, the probability is zero). It follows that a string has probability $\frac{1}{2}$ of terminating at the second toss, $\frac{1}{4}$ at the fourth (provided it did not terminate at the second), $\frac{1}{6}$ at the sixth (provided it did not terminate at the fourth) and so on. In demographic terms, a string can be thought of as an individual whose probability of dying decreases with age (this happens for newborn babies, the probability of survival increases each day they survive; the difference is that they grow old and the conditions become worse, whereas for the strings they continue to improve). We therefore see why the probability of 'the string enclosing the instant t' is smaller if t is smaller than if t is large. In the first case, it is necessarily 'young' (age at most t) and this bounds the past duration (the age that has been attained) and provides less favourable possibilities for the future (since an individual gets stronger with age): it has probability at least $1/(t + 2)$ of terminating at the next even toss, and so on.

And what are the probabilities of the various possible values for Y_n? They are no longer those of the Bernoulli distribution that we had before. We are now in a different *state of information*, because we are discussing a string; in other words, we know (or assume) that Y_t has not vanished in the meantime, and that the probability we seek is that conditional on this hypothesis, H, as we have already seen (Section 8.7.9(c)). This means that knowing that there are no zeroes modifies the distribution in favour of the larger values (as is natural); more precisely, it alters the probabilities in proportion to their sizes.

Every change in the state of information brings about a modification. For instance, if I knew the values of Y_n at every instant, then the probability of the string ending at the next toss – let us take n to be odd – is no longer $1/(n + 1)$, but is $\frac{1}{2}$ if $Y_n = \pm 1$, and zero otherwise. The situation would be different again if I knew just a *few* of the past values. If I knew Y_t at certain instants $t = n_1, \ldots, n_k$, the probabilities would be those conditional on the last value; that is on the hypothesis $H = (Y_{n_k} = c)$. But beware! This is only true if I have no knowledge, no clue whatsoever, relating to the results following the last known result. For example, if I obtain information every time $Y_t = 0$, I not only know the position of this last zero but I also know that no other zeroes have occurred since (and I then have the distribution given above, for the present Y_n, whereas, otherwise, I would have the Bernoulli distribution). Yet another situation would arise if I knew it to be more probable for information to be available in the case of returns to the origin, or in the case of large values being attained, and so on, than in other cases: *absence of information can itself be informative.* In the cases just mentioned, it increases the probability of the nonoccurrence of those things which, had they occurred, would probably have been reported.

Many of the mistakes that are made in the probabilistic treatment of problems and phenomena derive precisely from either ignoring, or forgetting, or giving insufficient weight to, the following fact: that everything depends upon the current, actual state of information (with all the attendant flexibility that attaches to this notion in practice). In interpreting deterministic laws, also, we need to keep circumstances of this kind in mind. An example is provided by the treatment of 'hereditary' phenomena, such as hysteresis, using integral or integral–differential equations. If one assumes that knowledge of the past enters the picture indirectly through the modification it produces in the structure determining the present state (which is itself not directly observable), then we see that certain information (in our case, 'the past') may or may not be 'informative' in so far as the effects we are interested in are concerned (here, *deterministic* prevision of the future – i.e. 'prediction' – rather than prevision), depending on whether certain other information is, or is not, available (here, complete information about the structure of the 'present state'). If this latter information is not available, the information concerning the past serves as a substitute. It may be a completely adequate substitute, or only partially so, according to whether the knowledge of the outside influences in the past are, or are not, considered sufficient to determine, completely and with certainty, the present unobservable situation. In the latter case, we are essentially back in the realm of probability, even if this remains obscured by the fact that the treatment deals only with macroscopic behaviour, neglecting the random aspects which are, in that context, negligible.

In the probabilistic field, however, information is always incomplete and derives from the distinction – which, in any case, is never very clear-cut – between what one knows, or believes one knows, or definitely remembers, and what one does not know. A fundamental rôle is played by that certain something which is, in a sense, complementary to information, and which comes about by interpreting the reasons for the absence of information. We have illustrated this with the Heads and Tails process, and we shall return to it again and again, sometimes with illustrations which are particularly instructive because they are, at first sight, rather disconcerting.[38] For examples from more familiar fields, note that the knowledge of a person's age is, to some extent, a substitute for a medical examination in so far as the evaluation of the probability of death is concerned and that, for a person who is insured, present age, together with the medical report dating back to when the policy was originally taken out, are taken as substitutes for a medical examination at the present time. In like manner, the fact of whether or not one receives direct news, or whether or not the newspapers carry reports of a certain situation, individual, firm, institution and so on, itself constitutes information (satisfactory or not, as the case may be). Any attempt – and these are still frequent – to base the theory of probability on some distinction between those things of which one is *perfectly certain* and others of which one is *perfectly ignorant* precludes, for the reasons we have given (and by not taking into account objections of principle), an understanding of the most meaningful aspects of problems requiring the use of probability theory.

Going back to the discussion of the questions we considered for the Heads and Tails process and to the doubts expressed in this context ('*How can it be…* ?'), we see, therefore, that, in line with what has been said, the answer lies in making it clear that the situation – for example, the probability of a return to equilibrium at a given instant – does not alter merely because of the passage of time, or because time modifies something or other, but rather because our state of information changes. Initially, that is not

38 For example, the equivalence of the Bayes–Laplace scheme to that of Pólya's 'contagion probabilities'; see Chapter 11, 11.4.4.

conditional on any subsequent information, our state of information consists solely of the knowledge that $Y_0 = 0$. When we later place ourselves at times $t = n$, however, we have a *changed* state of information; one that consists in knowing that there was a passage through the origin *n steps ago* (without knowing whether or not this was the last zero, nor anything else which would lead us to depart from an identical state of information regarding the 2^n possible paths that could have been followed meanwhile).

8.8.4. Despite all this, the doubt might linger on, transplanted to the new ground opened up by the new information. How can it be that the variation in the state of information to which we referred continues to exert an influence even when we move indefinitely far away? There is, in fact, a case in which knowledge of the initial state, provided it is sufficiently remote, ceases to have any influence (this is the so-called *ergodic* case, which we shall be concerned with briefly in Chapter 9, 9.1, when we deal with 'Markov Chains'). This is, however, something that only occurs under certain specified conditions and the fact that it often crops up is merely an indication that these conditions are satisfied in many of the problems one considers, rather than because of some principle which permits one to use the idea indiscriminately.

Here, too, as in the case of belief in a tendency to equilibrium, it happens that a special circumstance is assumed as some kind of autonomous 'principle', rather than as a simple and direct consequence of conditions that may hold in some cases and not in others. In this way – partly by accident and partly as a result of the usual obsession with replacing probability theory by something which is apparently similar, but which can, in fact, be reduced to the ordinary logic of certainty – one ends up by seizing upon the most fascinating results (like the laws of large numbers and the ergodic theorem) and raising them to the status of principles. When the applications of these principles to situations in which the theorems they misrepresent do not hold turn out to be contradictory, the results are then held to be paradoxical. As an analogy, it is as if, instead of the principle of conservation of energy, one took as a 'principle' the statement that a field of forces must be conservative, and then were faced with justifying the 'paradoxical' cases (like magnetic fields) where the 'principle' no longer holds.

As an example, it is often asserted – especially by philosophers – that the calculus of probability *proves* that the 'ergodic death' of the universe is inevitable. On the contrary, the calculus of probability (the logic of uncertainty) is completely neutral with respect to facts and behaviour relating to natural phenomena, and with respect to any other kind of 'reality' (in just the same way as the logic of certainty is). It is absurd to believe that the calculus of probability can itself rule out any particular belief or that it can force one to adopt it; whether it be a belief in 'ergodic death', or whatever. All that it does do is to rule out 'incoherent previsions', on the grounds that these are not previsions; in the same way as the logic of certainty precludes one making the assertion that a horse has three fore legs and four hind legs, making a total of five (whereas it would be admissible to say 3 + 4 = 7, or 3 + 2 = 5, or 1 + 4 = 5). The point is that it must not be inconsistent; the question of whether or not the statements conform to what zoologists regard as admissible is not relevant. Ergodic death is very probable if one accepts, or at least assumes as the most plausible, that model of physical phenomena which regards things as deriving from the destruction of an initial state of order (as in the mixing of gases; kinetic theory). But the calculus of probability in no way precludes phenomena in which a new order is created (as in biology; in particular, the mechanisms of reproduction for

DNA, and hence of cells, human beings, and – who knows? – new stars or galaxies);[39] on the contrary, its techniques provide the means of analysing them.

Returning to our case, one could say that the 'ergodic principle' *no longer applies* (which is another way of saying that the 'ergodic theorem does not apply unless the necessary assumptions are satisfied'), because if, at time $t = n$, we know that $Y_0 = 0$, then, amongst other things, we already know that $|Y_n| \leqslant n$ with certainty (not to mention our knowledge of the distribution); this information is significant, although its significance varies a great deal with n. The opposite situation – the ergodic case – occurs if we think of the same random walk on an m-sided polygon (m odd; a step clockwise or anticlockwise, according to whether we get a Head or a Tail). It is clear that knowledge of the starting point is practically irrelevant for evaluating the probabilities of the m positions after n steps (for n large); these probabilities will all be practically equal (to $1/m$).

There is one thing that we have seen, however, which seems to contradict the (obvious) fact that the process begins again under identical initial conditions after every return to zero. If this is so, how is it that the first zeroes can be expected to be very close and then subsequently get further and further apart in the startling way brought home by our discussion of the interposed repetitions of the Divine Comedy? There is, in fact, no contradiction. Each time a zero occurs, it is to be expected that there will be several close to each other; the initial period is a special case of this. As a result, obviously, the groups of zeroes are even further from one another than the individual zeroes would be if we took the same number of them but assumed them approximately equidistant. For an arbitrary zero, for example, the kth, conveying *no information* about the length of adjacent strings (as would be the case if one said, for instance, 'the first zero after the nth toss', because it is likely that the n*th* toss will fall within a long string), the probability is always $\frac{1}{4}$ that the two adjacent zeroes are the minimal distance away (=2; in other words, that the two adjacent strings have the minimal lengths possible, i.e. 2); $\left(\frac{5}{8}\right)^2 \simeq 0 \cdot 39$ that they are both not more than a distance of 4 away; $\left(\frac{193}{256}\right)^2 \simeq 0 \cdot 75$ for not more than 10 away; and so on (in general, the probability is $\left(1 - u_n\right)^2 \simeq 1 - 2u_n \simeq 1 - 1 \cdot 60/\sqrt{n}$ that both adjacent strings have lengths not exceeding n). Briefly, there are a great number of short strings, but here and there we find long strings, some extremely long; when we count not the numbers of strings but the number of steps they contain, however, the proportionate contribution from short and long strings is inverted. This is what we saw (see the *Remark* in Section 8.7.6) when we compared the probabilities u_{n-2}/n with the previsions of the lengths $n(u_{n-2}/n) = u_{n-2}$. In the context of this conceptual discussion, it is convenient just to take up the final point (mentioned above) concerning the trouble caused by the expression 'on average', a notion inspired by the statistical formulation.

Once again, we are dealing with the attempt to replace a genuine and valid probabilistic concept, which applies under all circumstances, by a counterfeit notion, only partially valid and not always applicable (it does not apply here, for example).

39 For a brief summary of how 'chance' comes to intervene continuously in thousands of complicated ways to bring about evolution (albeit, of course, according to our present conceptions), it is sufficient to read the two sections entitled 'The Development of Life' and 'A Chance Happening' in V.F. Weisskopf, *Knowledge and Wonder*, Heinemann, London (1962). As for the 'continuous creation of matter' and the formation of the galaxies, see the section entitled 'What happened at the beginning?', pp. 165–167; also see D.W. Sciama, *The Unity of the Universe*, Faber and Faber, London (1959); in particular, the section on the 'Steady State Model', pp. 155–157.

The probabilistic meaning is expressed perfectly – even though the expression may be rather unpalatable – by saying that each *individual* string has an infinite *expected length*, which is a result of the possible lengths 2, 4, 6, 8,... having probabilities $\frac{1}{2}, \frac{3}{8}, \frac{5}{16}, \frac{35}{128}, \ldots$, respectively. It makes no difference (just as it makes no difference whether we quote an interest rate as 45 lire per 1000 lire, or as 0–045) if we say that in 1000 strings we have, in prevision, a total length of 1000 deriving from strings of length *two*, 375 from strings of length *four*, 312·50 from length *six*, 273·44 from length *eight* and so on. There is no harm in this, and, indeed, it could be more expressive to consider that an expected length of 312·50 steps from 1000 strings derives from an expected number of strings of length six equal to 52·087. The trouble arises when one tries to interpret the phrase in a nonprobabilistic sense, as if it were possible to state that in any 1000 strings *things will turn out in this way* (in some vague sense: no-one goes so far as to claim this to be logically true – i.e. in a definite necessary sense – but people omit to state that it is at most 'very likely'; it is as though the possibility of something else happening could be avoided by recourse to some hybrid notion of 'practically certain'). Anyway, this very fundamental objection, one which always applies, precludes one from making statements of this kind without qualifying them as being *almost certain*, where by *almost-certainty* – by the mere fact of it not being *certainty* – we mean simply a rather high degree of probability (the latter being always subjective).

But it is not enough merely to correct a conceptually and formally inadequate form of expression. We must also make clear that statements which assert that in a large number of trials (in our case, strings) the actual outcomes will very likely be close to the 'previsions' do not hold except under appropriate conditions. First of all, we need conditions like independence, and this holds in our case; so far as the lengths and the signs are concerned, the strings are stochastically independent. For this reason, we can conclude that we may be almost certain that the proportion of positive and negative strings is fifty-fifty; the sequence of strings thought of in terms of their signs is a Heads and Tails process. However, we cannot claim that the same is true for the number of steps made on the two half-planes: despite the independence, the conclusion fails to hold because the prevision of the length of a string is infinite. *A fortiori*, for precisely the same reason, we cannot make statements of almost certainty about the frequency distribution of the lengths of the strings (or of the number of steps, considered in terms of the length of the string to which they belong). On the contrary, it would be very difficult to formulate this problem (even if the expected length were finite, and the statement therefore essentially true), not least because there are an infinite number of cases (lengths) to be distinguished.

We have shown how even a fairly superficial examination of very simple cases, like that of the Heads and Tails process, can reveal a number of features which are both unsuspected and fascinating in their own right. The intrinsic interest of these results has interesting conceptual implications when one considers more deeply the reasons for the surprise and the air of paradox which they generate.

8.9 Properties of the Wiener–Lévy Process

8.9.1. In Section 8.3, we had a brief look at those properties of the Wiener–Lévy process that could be established immediately and which served to enable us to make reference to the process. We now return to this topic, both in order to consider it in more depth and to show how certain asymptotic properties, which hold in many cases of asymptotically normal processes, gain in simplicity and clarity when we observe that they are exact properties in the Wiener–Lévy case.

This is, paradoxically, both the simplest and the most pathological case. The disarmingly simple features we have seen already: all the quantities we consider, whether individually, in pairs, or n at a time, and under all the circumstances we have examined, are normally distributed (in either 1, 2 or n dimensions). What could be better?

There was one feature, however, which might perhaps have given us grounds for suspecting that troubles might lie ahead. We are referring to the property of projective invariance, which, as we mentioned, enables us to reduce the study of asymptotic behaviour to that of the behaviour in the neighbourhood of the origin. At the time, we did not wish to frighten the reader by drawing attention to certain things that happen at infinity and cause awful trouble when one considers them concentrated near the origin. What is even worse is that everything that happens in the neighbourhood of the origin also happens in the neighbourhood of every other point of the curve $y = Y(t)$ (since the process is homogeneous with independent increments).

8.9.2. First of all, we recall how the Heads and Tails process (along with many others) provides – with an appropriate change of scale – a good approximation to the Wiener–Lévy process (to any desired degree of accuracy).

Let us consider the standardized Wiener–Lévy process ($m = 0$, $\sigma = 1$). A Heads and Tails process, in order to preserve these characteristics and to approximate the continuous process, must consist of a large number of small jumps (e.g. very frequent bets with very small stakes) and, instead of a single jump +1 per unit time, requires N^2 jumps of size $\pm 1/N$ per unit time. In this way, the standard deviation per unit time is, in fact, given by

$$\sigma\sqrt{n} = (1/N).\sqrt{N^2} = 1$$

as required.

If N is large – in the sense that the time intervals $\tau = 1/N^2$ and stakes $s = 1/N$ are small in comparison with the precision with which we wish, or are able, to measure intervals of time and amounts of money – this process is practically indistinguishable from the Wiener–Lévy process. In fact, if all the time intervals we wish, or are able, to consider contain a large number of small time intervals, τ, then the increments of $Y(t)$ are made up of the sums of a large number of independent increments and are, therefore, approximately normally distributed.

If we think in graphical terms, we can say that if the graph of the Heads and Tails process (Figure 7.1 in Chapter 7, 7.3.2) is collapsed by dividing the ordinates by $1/N$ and the abscissae by $1/N^2$, with N large enough to render imperceptible the segments of the broken line corresponding to the individual tosses, then we have the most exact obtainable representation of a Wiener–Lévy process. In a certain sense, this is precisely the same old process of approximation and idealization as is used when we consider changes in population (the number of inhabitants of some region etc.) as a continuous graph: even though one wishes to consider them as drawn in, one chooses the scale in such a way as to render imperceptible the jumps that represent the individual births and deaths on which the behaviour of the curve actually depends.

8.9.3. Of course, as we remarked at the time, instead of starting from the Heads and Tails process in discrete time, we could start from that in continuous time (the Poisson variant, with jumps $\pm 1/N$, N^2 of them, *in prevision*, per unit time), or with any other distribution of jumps (e.g. normal, always taking the standard deviation to be $1/N$ etc.).

Conversely, the Wiener–Lévy process can be a useful representation (giving an excellent approximation on some given scale) of phenomena whose 'microscopic' behaviour may well be very different. Among other things, it provides a useful model of the *Brownian motion* of a particle (or, better, if we restrict ourselves to one dimension, the projection of its motion onto one of the axes). Of course, the scale must be chosen so that it no longer makes sense to attempt to follow the actual mechanism of the phenomena, with its free paths, collisions and so on. We also note that P. Lévy often refers to the Wiener–Lévy process as the 'Brownian motion process' (the name Wiener–Lévy has come about in recognition of the two authors who made the greatest contributions to the study of this process; Bachelier also deserves a mention, however – he had previously discovered many of the properties and results, although in not such a rigorous manner).

8.9.4. We shall restrict ourselves in what follows to collecting together, as a survey, some of the more interesting facts about the process, but without providing proofs. In general, however, we shall be dealing with results that have already been proved implicitly – or at least made plausible – by virtue of results established for the Heads and Tails process.

Problems relating to the Wiener–Lévy process can be tackled in many different ways; in a certain sense, this reflects the various ways of looking at the normal distribution, which we noted when the latter was first introduced (the Wiener–Lévy process can be considered, roughly speaking, as a particular form of extension of the normal distribution to the infinite-dimensional case). On the other hand, we get a better overall view if we say something briefly about each of the most important procedures.

Those procedures, which derive basically from the Heads and Tails process, or something similar, are essentially rooted in combinatorial theory (together with whatever else may be required). The greater part of Chapters 7 and 8 are, in fact, devoted to this kind of procedure and we have often pointed out how it might be used in the context of the Wiener–Lévy process. We shall shortly give the details of this.

Rather more direct procedures are derived from the properties of the normal distribution itself, together with the various techniques for dealing with distributions. This means that knowledge of the second-order characteristics (variances and covariances) are sufficient to determine the process. We have already given examples of this when introducing the preliminary properties of the Wiener–Lévy process.

The third kind of procedure will require a more thorough discussion. It involves the study of diffusion problems (the heat equation and so on) and was briefly mentioned in Chapter 7, 7.6.5, and again in Chapter 8, 8.6.7. Despite the elegance and power of these methods, we shall not be able to say a great deal about them here. This is unfortunate, because, besides their power, they can be made very expressive in terms of the image of the spread of probability, considered as mass. However, we clearly cannot include everything and this seemed a reasonable candidate for exclusion, as – from a conceptual viewpoint – it is more in the nature of an analogy than a genuine representation of the problems. This is in contrast to the other approaches, which, in their various ways, stick closely to the probabilistic meaning and enable one to shed light on every aspect of it.

Anyway, we shall restrict ourselves here to illustrating, using the perspective provided by diffusion theory, some of the problems with which we are already familiar through other approaches.

The Wiener–Lévy process corresponds precisely to the basic case of diffusion starting from a single source. The so-called 'dynamic' considerations and conclusions, in which t is considered as the time variable (instead of just a constant), involve precisely this process (and not just the individual distributions at individual instants).

The gambler's ruin problem (in the version provided by the Wiener–Lévy process) requires the introduction of an absorbing barrier; the straight line $y = -c$, where c = the initial capital. This problem can be solved, in the theory of heat transfer, by the *method of images* (due to Lord Kelvin). This involves placing an opposite (cold) source at the point $t = 0$, $y = -2c$ (a mirror-like image of the origin, a hot source, the image being taken with respect to the barrier). The resulting process, which, for reasons of symmetry, clearly has zero density on the barrier, gives, at each instant t, the density of the distribution of the gain. The missing part (the integral of the density is less than 1) is the mass absorbed by the barrier; that is the probability of ruin before the instant under consideration. It can be seen, without the need for any calculations, that this is twice the 'tail' that would go beyond the barrier (this tail is itself missing and there is also the negative tail that has come in from the cold source). We note that this corresponds precisely to Desiré André's argument.

Similarly, in the case of the two-sided problem, the method of images leads to the introduction of an infinite number of hot and cold sources (images of the actual source, with an even or odd number of reflections in the absorbing barriers). This is the 'physical' interpretation of the formulae in Section 8.6.6.

8.9.5. Our survey of the results relating to the Wiener–Lévy process should begin, naturally enough, with those we gave in Section 8.3, and with those we came across subsequently. We shall only repeat those things which are required to make the survey sufficiently complete.

We begin with the results relating to the ruin problem (i.e. to the case with an absorbing barrier).

In the case of a single barrier (at $y = c$, say), the probability of ruin at or before time t, $F(c, t)$,[40] that is the distribution function for the time T spent by the process before ruin occurs, $F(c, t) = P(T \leqslant t)$, is given by

$$\begin{aligned} F(c,t) &= \mathbf{P}\big(|Y(t)| > |c|\big) = 2\mathbf{P}\big(Y(t) > |c|\big) \\ &= \frac{2}{\sqrt{(2\pi t)}} \int_{|c|}^{\infty} \mathrm{e}^{-y^2/2t}\,\mathrm{d}y = \frac{2}{\sqrt{(2\pi)}} \int_{|c|/\sqrt{t}}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x \end{aligned} \tag{8.47}$$

(the above holds in a real sense, because the probability of $T < \infty$ is 1). The density has the form

$$\frac{\partial F}{\partial t} = f_c(t) = \frac{|c|}{\sqrt{(2\pi)}} t^{-\frac{3}{2}}\, \mathrm{e}^{-c^2/2t} = \frac{|c|}{Nt} \cdot \frac{N}{\sqrt{(2\pi t)}} \mathrm{e}^{-c^2/2t}. \tag{8.49}$$

40 In Sections 8.6.4–8.6.5, this was denoted by $q(t)$ and $p(c)$ (or $q_c(t)$ and $p_t(c)$) because it was then convenient to think of one of the variables as fixed (i.e. included as a parameter).

We recall that we are dealing with the stable distribution with index $\alpha = \frac{1}{2}$. The second form emphasizes the relationship with the Heads and Tails process, involving N^2 tosses per unit time, each involving a gain of $\pm 1/N$. The second factor expresses (asymptotically) the probability of a gain of $|c| = (N|c|)(1/N)$ in N^2t tosses, the first factor the probability that we are dealing with the first passage through level $y = |c|$ (see Section 8.7.9(c)).

When considered as a function of $|c|$ (and we shall write y rather than $|c|$), the distribution becomes the half-normal, with density

$$f_t(y) = \sqrt{(2/\pi)}\,\mathrm{e}^{-y^2/2t}\left(y \geqslant 0\right)\left(\text{i.e. zero for } y < 0\right). \tag{8.48}$$

We recall that, in addition, this holds for the following cases:

the absolute value of $Y(t)$
the absolute value of $\vee Y(t)$ (the maximum of $Y(\tau)$ in $0 \leqslant \tau \leqslant t$)
the absolute value of $\wedge Y(t)$ (the minimum of $Y(\tau)$ in $0 \leqslant \tau \leqslant t$)
the absolute value of $\vee Y(t) - Y(t)$
(the deviation from the maximum)
the absolute value of $\wedge Y(t) - Y(t)$
(the deviation from the minimum).

We now give the probability distributions of $Y(t)$ conditional on three different assumptions concerning the maximum of $Y(\tau)$ in $[0, t]$. The three assumptions are as follows:

that, with respect to some given $c > 0$, we have $\vee Y(t) \geqslant c$ (in equation 8.75);
that $\vee Y(t) \geqslant c$ (in equation 8.76);
that $\vee Y(t) = Y(t)$(in equation 8.77).

In the first two cases we have:

$$f(y) = K\exp\left\{-\left(c + |y - c|\right)^2/2t\right\}, \quad 1/K = F(c,t), \tag{8.75}$$

$$f(y) = K\left[\exp\left\{-y^2/2t\right\} - \exp\left\{-\left(2c - y\right)^2/2t\right\}\right]\left(y \leqslant c\right), \quad 1/K = 1 - F(c,t). \tag{8.76}$$

The first follows immediately from the reflection principle. Note that $c + |y - c|$ is equal to y (for $y \geqslant c$) or $2c - y$ (for $y \leqslant c$), and that the distribution is therefore the normal with the central portion (between $\pm c$) removed, and the two remaining tails attached to one another. For the second, it is sufficient to observe that, multiplying it and the first one by $1 - F(c, t)$ and $F(c, t)$, respectively (i.e. by suppressing the K), and summing them, we must again obtain $K\exp(-y^2/2t)$.

Finally, suppose we assume that we know that either the value $Y(t)$ is greater than all those previously obtained, that is that $Y(t) = \vee Y(t)$ (without knowing anything more about the actual value), or that we know that $\wedge Y(t) = 0$, that is that the minimum is the initial zero, $Y(0) = 0$ (by the reversal principle, the two cases are equivalent).

Conditional on either of these, the density of the distribution is like $\partial F/\partial t$ in equation 8.49, except that we now have to take $y = |c|$ as the variable rather than t. This changes K and we obtain

$$f(y) = \frac{y}{t} e^{-y^2/2t} \left(y \geqslant 0\right), \tag{8.77}$$

giving the distribution function

$$F(y) = \left[1 - e^{-y^2/2t}\right]\left(y \geqslant 0\right). \tag{8.78}$$

The same thing holds, with the range of values reversed, when we take $Y(t)$ to be equal to the minimum rather than the maximum (or if $Y(0) = 0$ is the maximum).

This can be justified by considering (in a somewhat roundabout manner) the meaning of $\partial F/\partial t$, or, alternatively, by considering equation 8.71 of Section 8.7.9, which refers to the Heads and Tails process.

In the case of two barriers (the ruin problem for two gamblers), the distribution of $Y(t)$ conditional on the fact that neither has been ruined in $[0, t]$, that is conditional on

$$-c' \leqslant \wedge Y(t) \leqslant \vee Y(t) \leqslant c'' \qquad \left(\text{with } c' > 0 \text{ and } c'' > 0\right),$$

is given by

$$f(y) = K \sum_{h=-\infty}^{+\infty} \left[\exp\left\{-\left(y - 2hc^*\right)^2/2t\right\} - \exp\left\{-\left(y + 2c' + 2hc^*\right)^2/2t\right\}\right], \tag{8.79}$$

where $c^* = c' + c''$, and $-c' \leqslant y \leqslant c''$.

In the symmetric case, $c' = c'' = c$, $c^* = 2c$, this becomes

$$f(y) = K \sum_{h=-\infty}^{+\infty} (-1)^h \exp\left\{-\left(y - 2hc\right)^2/2t\right\} \qquad \left(-c \leqslant y \leqslant c\right). \tag{8.80}$$

Clearly, the probability, $1 - q(t)$, of neither barrier being crossed until time t is equal to $1/K$ (which is given by the integral of the Σ between $\pm c$, or, in the general case, between $-c'$ and $+c''$). This $q(t)$ also appeared, in a slightly different form, in equation 8.53 (see Section 8.6.5). Note that, if we ignored the K, equation 8.79 would give $f(y)\,dy$ = the probability that $Y(t)$ lies in $[y, y + dy]$ *and* has never previously gone outside the interval $[-c', c'']$: the point is that we would be saying 'and', rather than 'assuming that'. Similar comments apply in all other cases of this kind.

The few remarks we have made about Lord Kelvin's 'method of images' (Section 8.6.7) suffice to explain the result. If one so wished, one could verify it by checking that both the diffusion equation (equation 8.32 of Chapter 7, 7.6.5) and the boundary conditions, $f(y) = 0$ on the half-lines $y = -c'$ and $y = c''$ for $0 < t < \infty$, are satisfied.

8.9.6. In the case of the Wiener–Lévy process, we can provide complete answers to questions concerning the asymptotic behaviour of $Y(t)$ as $t \to \infty$. In principle, these answers are provided by a celebrated result of Petrowsky and Kolmogorov; 'in practice', that is in a less complete but more expressive way, they are given by a famous theorem of Khintchin, the so-called 'law of the iterated logarithm' (see the brief comment in Chapter 7, 7.5.4).

What we do is to compare $Y(t)$ with some function $\omega(t)$ (which we assume to be continuous, increasing and tending to $+\infty$), and then calculate the probability that the inequality $Y(t) < \omega(t)$ holds from some arbitrary time t onwards. More precisely, we examine the limit, as $t' \to \infty$, of the probability that the inequality holds from t' onwards. To be even more precise,[41] this latter probability is itself to be understood as the limit, as $t' \to \infty$, of the probability that the inequality holds in (t', t''). The p of interest is thus given by

$$p = \lim_{t' \to \infty} \left[\lim_{t'' \to \infty} p\left(t', t''\right) \right];$$

the limit certainly exists, since $p(t', t'')$ increases as t'' increases and decreases as t' increases.

We can say a great deal more, however. By the Zero–One law, only the two values $p = 0$ or $p = 1$ are possible: either it is practically certain that $Y(t)$ remains below $\omega(t)$ from a certain time onwards, or it is practically certain that this does not happen; that is there will always be segments in which $Y(t)$ is greater than $\omega(t)$. The class of functions $\omega(t)$ can therefore be divided into two subclasses, which could be said to contain 'those which increase more (less) rapidly that the "large values" of $Y(t)$' The general distinction (given by Petrowsky and Kolmogorov) is that $\omega(t)$ belongs to the upper or lower class according to whether the improper integral (from an arbitrary positive t_0 to $+\infty$) of

$$\psi(t).t^{-1} \exp\left\{-\frac{1}{2}\psi^2(t)\right\} \mathrm{d}t, \quad \text{where } \psi(t) = \omega(t)/\sqrt{t}, \tag{8.81}$$

converges or diverges.

In terms of $\psi(t)$, the condition $Y(t) < \omega(t)$ can be written as $Y(t)/\sqrt{}/t < \psi(t)$; that is in terms of a *standardized* function (for which we have made σ = constant = 1.

The more expressive distinction (that of Khintchin) simply considers the class of functions

$$\omega(t) = k\sqrt{(2t \ \log \ \log t)} \qquad \text{(i.e.} \quad \psi(t) = k\sqrt{(2\log\log t)}) \tag{8.82}$$

41 This further qualification is unnecessary if countable additivity is assumed. We recall similar caveats in the case of the strong law of large numbers (Chapter 7, 7.7.3), etc. For simplicity, we shall give an informal discussion here.

and asserts that these belong to the lower class when $k \leqslant 1$, and to the upper class when $k > 1$. The result can be strengthened by considering the functions

$$\omega(t) = \sqrt{\left[2t\left(\log\log t + k\log\log\log t\right)\right]}: \tag{8.83}$$

these belong to the lower class if $k \leqslant \frac{3}{2}$, and to the upper if $k > \frac{3}{2}$ (a generalization due to P. Lévy, which is proved by first using a direct approach, which leaves the cases $\frac{1}{2} \leqslant k \leqslant \frac{3}{2}$ undecided, and then removing the gap by using the Petrowsky and Kolmogorov result).

The proof of the general criterion is based on diffusion theory ideas (which relate to the theory of heat flow). For the law of the iterated logarithm, we recall the previous comments given for the case of Heads and Tails.

8.9.7. *Small-scale behaviour* is extraordinarily complicated and irregular. Not only do all the large-scale peculiarities reappear (shrunk by a factor of N^2 in the abscissa, corresponding to a factor of N in the ordinate) but, also, if we study the behaviour in the neighbourhood of a point – the origin, for instance – we find all the asymptotic properties corresponding to $t \to \infty$ reappearing in an inverted way. This can be seen most simply by observing that, if $Y(t)$ is given by a Wiener–Lévy process, then the same is true for the function

$$Z(t) = tY(1/t);$$

this has $m_t = 0$, $\sigma_t = t\sqrt{(1/t)} = \sqrt{t}$, the distribution is normal, and the correlation coefficient between $Z(t_1)$ and $Z(t_2)$ is the same as that between $Y(1/t_1)$ and $Y(1/t_2)$ (if $t_2 > t_1$, and hence $1/t_1 > 1/t_2$, it is equal to $\sqrt{[(1/t_1)/(1/t_2)]} = \sqrt{(t_2/t_1)}$); this is all we need.

It is practically certain, therefore, that in every neighbourhood of zero ($Y(0) = 0$) $Y(t)$ vanishes *an infinite number of times* (as in the case when $t \to \infty$) and that it touches infinitely often every curve

$$y = \omega(t) = k\sqrt{\left[2t\log\log(1/t)\right]}$$

with $k \leqslant 1$, but not those with $k > 1$ (which gives, *locally*, an almost certain 'modulus of continuity'; $|Y(t_0 + t) - Y(t_0)| < \omega(t)$ in a neighbourhood of t_0, with $0 < t < \varepsilon$). If, however, we want this to hold almost certainly for all the t_0 of some given interval simultaneously (still for all t between 0 and ε), we have to take

$$\omega(t) = k\sqrt{\left[2t\log(1/t)\right]}, \qquad k > 1 \tag{8.84}$$

(the simple rather than the iterated logarithm).

In order to be brief, the presentation of the results in these cases has been rather informal. We should point out, however – for reasons we shall see shortl y– that there are grave dangers in treating these topics without sufficient care and attention. For every point t_0 at which $Y(t_0) = 0$, it is practically certain (probability = 1) that there are other roots (an infinite number of them) in every interval of the point, either to the left or to the right (and the same holds true at every other point if we consider crossings of the horizontal line $y = Y(t_0)$). On the other hand, between two roots there are always several intervals (and almost certainly a countable infinity) in which $Y(t)$ is either positive or negative; hence there are isolated roots to the right or to the left – the end-points of such intervals. Since this can be repeated for all horizontal lines y = constant (an *uncountably infinite* set), the points $y = Y(t)$, which are isolated (on at least one side) from points of the curve at precisely the same level y, form, in every interval, an uncountably infinite set and among them there are always an infinite number of points isolated on either side (at least the maxima and minima).

This having been said, the length of the segment starting from the origin, where we assume $Y(0) = 0$ (or starting from some arbitrary t' at which we know that there is a root, $Y(t') = 0$), and containing no roots, is a random quantity X, which has probability 1 of being precisely zero (if 0, or, in general, t' is a root which is adherent on the left to the set of roots). Indeed, such a random quantity, $X(t')$, can be considered, without changing the problem, for any arbitrary t' – even if $y' = Y(t')$ is not zero – as the length of the interval on the left of t' not containing points t at which $Y(t)$ again takes the value y'. In any case, we know that we necessarily have $X(t') > 0$ for an uncountable infinity of points in any arbitrarily small interval, and it can be shown that, *assuming* the length X to be greater than some given $x_0 > 0$, the probability of it being greater than some $x \geqslant x_0$ is $\sqrt{(x/x_0)}$. In other words, conditional on the hypothesis $X \geqslant x_0$ ($x_0 > 0$), we can say that X has distribution function and density given by

$$F(x) = 1 - K\sqrt{x}, \tag{8.85}$$

$$f(x) = \tfrac{1}{2}K/\sqrt{x}, \tag{8.86}$$

where $K = 1/\sqrt{x_0}$ (so that $F(x_0) = 0$); as $x_0 \to 0$, we also have $K \to 0$.[42]

The result for Heads and Tails (where $X \geqslant n$ has probability $u_n \simeq 0{\cdot}8/\sqrt{n}$[43]) corresponds to the case $K \neq 0$ (because, clearly, in discrete time there is no way for 'peculiarities on the small scale' to occur).

---

42 This means not only that, in the absence of any contrary hypothesis (the case of an infinite number of roots adherent on the left), there is a probability = 1 that $X = 0$ (i.e. X is concentrated at the point $x = 0$), but also that with the single assumption that $x > 0$, all the probability is *adherent* to zero (i.e. however $x_0 > 0$ is chosen, the probability that $X \geqslant x_0$ is zero; this is obvious, because, in any finite interval, however large, there can only be a finite number of intervals containing no roots and of length greater than x_0, whereas there are an infinite number of 'small' intervals containing no roots in every interval of almost all the roots; i.e. excluding the isolated ones).

Note, of course, that the problem would be different if we were talking about an interval containing no roots, and chosen by picking out some point in it. As usual – recall 'sums' at Heads and Tails, 'number' and 'length' of strings, etc. – this procedure would favour the choice of the longest intervals (see the comments to follow in the text). The choice must be made by saying, for example, 't' = the starting point of the third interval of length $\geqslant x_0$ (possibly with some additional complications, in order that the restriction to a simple example is seen not to be necessary) after the level $Y(t) = c$ has been reached'. For the case $x_0 \to 0$ (under the assumption $X > 0$), this explicit method of choice does not exist. We can, however, reduce to the previous case by thinking of t' as having been determined in this way by some other person, with x_0 unknown to us, but given by $x_0 = 1/N$, where 'N is an integer chosen at random' (in the sense we discussed in Chapter 3, Section 3.2, and Chapter 4, Section 4.18).

43 Or $u_n/2 \simeq 0{\cdot}4/\sqrt{n}$: it does not make any difference whether we consider the X of the continuous case as a generalization of the length of a *string*, or of the period spent in the *lead* (i. as L or V of Chapter 8, Section 8.7).

8.9.8. If, instead, we begin by fixing a time t_0, knowing only that $Y(0) = 0$, and we consider $X = T''' - T'$, the length of the interval containing t_0 and no roots (i.e. T' = the last root of $Y(t) = 0$ with $t \leqslant t_0$, and T''' = the first root of $Y(t) = 0$ with $t \geqslant t_0$[44]), then we have the following probability distributions :

$$\text{for}\, T': \;\; f(t) = K/\sqrt{[t(t_0 - t)]},\; F(t) = K \sin^{-1}\sqrt{(t/t_0)},\; (0 \le t \le t_0), \tag{8.87}$$

$$\text{for}\, T'': \;\; f(t) = K/\sqrt{[t(t - t_0)]},\; F(t) = K \cos^{-1}\sqrt{(t_0/t)},\; (t \ge t_0), \tag{8.88}$$

$$\text{for} \quad X: \quad f(x) = K\int_{\alpha}^{t_0} \mathrm{d}t/\sqrt{[t(t_0 - t)(t + x)(t + x - t_0)]}, \qquad (x \geqslant 0), \tag{8.89}$$

where $\alpha = 0 \wedge (t_0 - x)$. Similarly, we have the result that, given t' and t'' ($t' < t''$) the probability of at least one root of $X(t)$ in the interval (t', t'') is equal to $K \cos^{-1}\sqrt{(t'/t'')}$.

The same results hold (by virtue of the usual transformations) if we think, for example, of T' as denoting the abscissa of the maximum (or of the minimum) of $X(t)$ between 0 and t_0 (rather than the last root) and, correspondingly, of (T', T''') as the interval in which the maximum (or minimum) remains constant (i.e. T''' is the last instant up to which $X(t)$ does not exceed the maximum value attained in (0, t_0); and similarly for the minimum) and so on. It is interesting to note – and this ties in with what we drew attention to in Section 9.8 as *seemingly 'paradoxical'* – that these points (maximum, minimum, last root) are more likely to be near the end-points of the interval (0, t_0) than near the centre. More precisely, as a more expressive interpretation, recall that T' is the abscissa of a point 'chosen at random' (i.e. with uniform probability density) on the circumference of a semi-circle having (0, t_0) as diameter (see Figure 8.9c).

In the case of Heads and Tails, we saw that, asymptotically, in any interval (0, t_0) with $Y(0) = 0$ (or in any interval (t', t'') with $Y(t') = 0$), the proportion of time during which $Y(t)$ is positive had the arc sine distribution. This property continues to hold, exactly, for the Wiener–Lévy process.

8.9.9. The 'pathological' character of the 'small-scale' behaviour might leave one somewhat puzzled as to the possibility of interpreting the process in a constructive way. For this purpose, Lévy suggests a procedure of definition by successive approximations. It consists of subdividing the interval under consideration ($0 \leqslant t \leqslant 1$, say) into 2, 4, 8,…, 2^k,… equal parts, in determining $Y(t)$ at the division points and in taking as the kth approximation a function $Y_k(t)$, coinciding with $Y(t)$ for those t which are multiples of $1/2^k$ and linear in between them. Given $Y(0) = 0$, and $Y(1)$ determined as a random quantity with a standard normal distribution ($m = 0$, $\sigma = 1$), the intermediate points are successively determined by means of the considerations of Section 8.3.2. If t' and t'' are two consecutive multiples of $1/2^k$, and $t = (t' + t'')/2$ is the point at which $Y_{k+1}(t) = Y(t)$ is to be determined, we know that it is sufficient to add to the prevision given by $Y_k(t)$

44 If it so happened that $Y(t_0) = 0$ (and this has probability 0), we would have $T' = T''' = t_0$ and $X = 0$.

(= $[Y_k(t') + Y_k(t'')]/2 = [Y(t') + Y(t'')]/2$) a random quantity having a centred normal distribution ($m = 0$) and standard deviation given by

$$\sigma = \sqrt{\left[(t-t')(t''-t)/(t''-t')\right]} = 1/2^{(k+2)/2} \qquad \left(\text{see Figure 8.3}\right).$$

By bounding the probabilities of large values for these successive correction terms, we can conclude (following Lévy) that the $Y_k(t)$ converge almost certainly to a continuous $Y(t)$.

Of course, this means that we are using countable additivity. If one wishes to avoid this, all the difficulties relating to 'small-scale' behaviour could be avoided by imagining, for instance, that the process only appears to take place in continuous time but, in fact, takes place in discrete time, with time intervals $1/N$ (with N unknown and having probability 0 of being smaller than any arbitrary preassigned integer).[45]

45 Note that the same idea can be used in reverse, making any discontinuous process, for example, the Poisson process, *continuous*. It is sufficient to think of the 'jump' +1, at any instant t, as actually a continuous increase taking place in a very short time interval from t to $t + 1/N$ (with N as above; for example, one could take an increment $N\tau$ in $0 \leqslant \tau \leqslant 1/N$, or take $\sin^2\left(\frac{1}{2}\pi N\tau\right)$; one could even assume behaviour of the form $1 - e^{-N\tau}$, or $1 - e^{-N\tau} \cos N\tau$, and so on, in $0 \leqslant \tau \leqslant \infty$).

Without countable additivity, there is no unique answer to certain of the more subtle questions. Countable additivity certainly provides unique answers, but this, of course, is no reason to consider the latter as 'well founded' in any special sense.