Stochastic Models in Risk Theory

Michael Zazanis
Department of Statistics
Athens University of Economics and Business

Winter 2015-16

Chapter 1

Discrete Distributions

1.1 Sums of discrete independent random variables

Let X, Y, be independent random variables with values in $\mathbb{Z}$. Suppose that $a_n = \mathbb{P}(X = n)$, $b_n = \mathbb{P}(Y = n)$ denotes the distributions of X and Y respectively. The distribution of $Z := X + Y$ is then given by

$$\begin{aligned}\mathbb{P}(Z = n) &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X + Y = n, Y = k) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = n - k, Y = k) \\ &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = n - k) P(Y = k) = \sum_{k=-\infty}^{\infty} a_{n-k} b_k.\end{aligned}$$

For the most part we will restrict ourselves to distributions on the non-negative integers. In this case, if X, Y, take values on $\mathbb{N}$, then

$$\mathbb{P}(Z = n) = \sum_{k=0}^{n} a_{n-k} b_k \qquad \text{for } n \in \mathbb{N}.$$

If $\{a_n\}$, $\{b_n\}$, $n \in \mathbb{N}$ are real sequences then the sequence $\{c_n\}$ where $c_n = \sum_{k=0}^{n} a_{n-k} b_k$ is called the *convolution* of the two sequences. We write $c_n = (a \star b)_n$.

1.2 The Probability Generating Function

The *probability generating function* (p.g.f.) of a discrete random variable X (with values in $\mathbb{N}$) is defined as

$$\phi(z) := \mathbb{E} z^X = \sum_{n=0}^{\infty} \mathbb{P}(X = n) z^n. \tag{1.1}$$

The series above converges at least for all $z \in [-1, 1]$. We note that if $p_k = \mathbb{P}(X = k)$, $\phi(z) = \sum_{k=0}^{\infty} p_k z^k$, and by $\phi^{(k)}(z)$ we denote the derivative of order k at z, then

$$p_k = \frac{1}{k!} \phi^{(k)}(0), \qquad k = 0, 1, 2, \ldots, \tag{1.2}$$

and

$$\mathbb{E}[X(X-1)\cdots(X-k+1)] = \phi^{(k)}(1). \tag{1.3}$$

The latter is called the descending factorial moment or order k. Ordinary moments can be easily obtained from these. Finally we note that the probability distribution $\{p_n\}$ obviously determines uniquely the p.g.f. $\phi(z)$ and, reversely, the p.g.f. uniquely determines the probability distribution via (1.2).

In particular we point out that, if X, Y, are *independent* random variables with p.g.f.'s $\phi_X(z)$, $\phi_Y(z)$ respectively, then the p.g.f. of their sum $Z = X + Y$ is given by $\phi_Z(z) = \phi_X(z)\phi_Y(z)$. To see this it suffices to note that $\phi_Z(z) = \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] = \mathbb{E}z^X \mathbb{E}z^Y$, the last equality holding because of the independence of X, Y. The above relation extends readily to the case of any finite number of independent random variables. In particular if X_i, $i = 1, 2, \ldots, n$ are i.i.d. (independent, identically distributed) random variables with (common) probability generating function $\phi_X(z)$ then their sum $S_n := X_1 + \cdots + X_n$ has p.g.f. given by $\phi_{S_n}(z) = (\phi_X(z))^n$.

While the p.g.f. of the sum S_n is readily obtained in terms of the p.g.f. of each of the terms X_i, the corresponding probability distributions are in general hard to compute. Based on the above discussion it should be clear that

$$\mathbb{P}(S_n = k) = \frac{1}{k!}\,\frac{d^k}{dz^k}(\phi_X(z))^n\bigg|_{z=0},$$

a quantity that, in the general case, is not easy to evaluate. Alternatively, if $p_k = \mathbb{P}(X = k)$ then $\mathbb{P}(S_n = k) = p_k^{\star n} := (p \star \cdots \star p)_k$, the n–fold convolution of the sequence $\{p_n\}$ with itself.

We give some examples of discrete probability distributions.

1.3 Discrete distributions

1.3.1 The Bernoulli and the Binomial distribution

The random variable

$$\xi = \begin{cases} 0 & \text{w.p. } q := 1-p, \\ 1 & \text{w.p.} p \end{cases}$$

where $p \in [0, 1]$ is called a Bernoulli random variable. It is the most elementary random variable imaginable and a useful building block for more complicated r.v.'s. Its p.g.f. is given by $\phi(z) = 1 - p + zp$, its mean is p and its variance is pq.

If ξ_i, $i = 1, 2, \ldots, n$ are independent Bernoulli random variables with the same parameter p then their sum $X := \xi_1 + \xi_2 + \cdots + \xi_n$ is Binomial with parameters n and p. Its distribution is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \ldots, n,$$

and its p.g.f. by

$$\phi(z) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = (1 - p + pz)^n.$$

The mean and variance of the binomial can be readily obtained from its representation as a sum of independent Bernoulli random variables. Indeed, $\mathbb{E}X = \mathbb{E}[\xi_1 + \cdots + \xi_n] = np$ and $\mathsf{Var}(X) = \mathsf{Var}(\xi_1 + \cdots + \xi_n) = \mathsf{Var}(\xi_1) + \cdots + \mathsf{Var}(\xi_n) = nqp$.

Note that, if $X \sim \mathsf{Binom}(p, n)$, $Y \sim \mathsf{Binom}(p, m)$, and X, Y, are independent, then $X + Y \sim \mathsf{Binom}(p, n + m)$.

1.3.2 The Poisson distribution

X is Poisson with parameter $\alpha > 0$ if its distribution is given by

$$\mathbb{P}(X = k) = \frac{1}{k!}\alpha^k e^{-\alpha}, \qquad k = 0, 1, 2, \ldots.$$

Its p.g.f. is given by

$$\phi(z) = \sum_{k=0}^{\infty} z^k \frac{1}{k!}\alpha^k e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{1}{k!}(\alpha z)^k = e^{-\alpha} e^{z\alpha} = e^{-\alpha(1-z)}.$$

The mean and variance of the Poisson can be easily computed and are given by $\mathbb{E}X = \mathsf{Var}(X) = \alpha$.

One of the most important properties of the Poisson distribution is that it arises as the limit of the binomial distribution $\mathsf{Binom}(n, \alpha/n)$ when $n \to \infty$ (i.e. in the case of a large number of independent trials, say n, each with a very small probability of success, α/n). This is easy to see by examining the probability generating function of the binomial $(n, \alpha/n)$ and letting $n \to \infty$. Indeed,

$$\lim_{n\to\infty} \left(1 - \frac{\alpha}{n} + z\frac{\alpha}{n}\right)^n = \lim_{n\to\infty} \left(1 - \frac{\alpha(1-z)}{n}\right)^n = e^{-\alpha(1-z)}$$

which establishes that $\mathsf{Binom}(\alpha/n, n) \to \mathsf{Poi}(\alpha)$ as $n \to \infty$.

We also point out that, if X_1, X_2 are independent Poisson random variables with parameters α_1, α_2 respectively, then $X_1 + X_2 \sim \mathsf{Poi}(\alpha_1 + \alpha_2)$. The easiest way to see this is to consider the p.g.f. $Ez^{X_1+X_2} = Ez^{X_1} Ez^{X_2} = e^{-\alpha_1(1-z)} e^{-\alpha_2(1-z)} = e^{-(\alpha_1+\alpha_2)(1-z)}$.

1.3.3 The geometric distribution

If X is geometric with parameter p its distribution function is given by

$$\mathbb{P}(X = k) = q^{k-1}p, \qquad k = 1, 2, 3, \ldots, \tag{1.4}$$

where $p \in (0, 1)$ and $q = 1 - p$, and its p.g.f. by

$$\phi(z) = \sum_{k=1}^{\infty} q^{k-1} p z^k = \frac{(1-q)z}{1-qz}. \tag{1.5}$$

The parameter p is usually referred to as the "probability of success" and X is then the number of independent trials necessary until we obtain the first success. An alternative definition counts not the trials but the failures Y until the first success. Clearly $Y = X - 1$ and

$$\mathbb{P}(Y = k) = q^k p, \qquad k = 0, 1, 2, \ldots, \tag{1.6}$$

with corresponding p.g.f.

$$\mathbb{E}z^Y = \frac{1-q}{1-qz}. \tag{1.7}$$

It is easy to check that $\mathbb{E}Y = q/p$ and $\mathsf{Var}(Y) = q/p^2$. Also, $\mathbb{E}X = 1 + \mathbb{E}Y = 1/p$ and $\mathsf{Var}(X) = \mathsf{Var}(Y) = q/p^2$.

1.3.4 The negative binomial distribution

The last example we will mention here is the *negative binomial (or Pascal) distribution*. Recall that the binomial coefficient is defined for all $a \in \mathbb{R}$ and $n \in \mathbb{N}$ as

$$\binom{a}{n} = \frac{a(a-1)\dots(a-n+1)}{n!}.$$

If a is a positive integer then $\binom{a}{n} = 0$ for all $n > a$. If however a is a negative integer or a (non-integer) real then $\binom{a}{n} \neq 0$ for all $n \in \mathbb{N}$. Also recall the binomial theorem, valid for $|x| < 1$ and all $\alpha \in \mathbb{R}$:

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k. \tag{1.8}$$

(If α is a positive integer then $\binom{\alpha}{k} = 0$ for all $k = \alpha+1, \alpha+2, \cdots$ and thus the infinite series (1.8) turns into a finite sum: $(1+x)^\alpha = \sum_{k=0}^{\alpha} \binom{\alpha}{k} x^k$.)

Note in particular that binomial coefficient $\binom{-\alpha}{n}$ can be written as

$$\begin{aligned}
\binom{-\alpha}{n} &= \frac{(-\alpha)(-\alpha-1)\cdots(-\alpha-n+2)(-\alpha-n+1)}{n!} \\
&= (-1)^n \frac{(\alpha+n-1)(\alpha+n-2)\cdots(\alpha+1)\alpha}{n!} = (-1)^n \binom{\alpha+n-1}{n}.
\end{aligned}$$

Thus we have the identity

$$(1-x)^{-\alpha} = \sum_{k=0}^{\infty} \binom{-\alpha}{k} (-x)^k = \sum_{k=0}^{\infty} \binom{\alpha+k-1}{k} x^k. \tag{1.9}$$

If $p \in (0,1)$ and $q = 1-p$ then the *negative binomial distribution* with parameters p and $\alpha > 0$ is defined as

$$\mathbb{P}(X = k) = \binom{\alpha+k-1}{k} p^\alpha q^k, \qquad k = 0, 1, 2, \dots. \tag{1.10}$$

In order to check that the above is indeed a probability distribution it suffices to note that $\binom{\alpha+k-1}{k} > 0$ when $\alpha > 0$ for all $k \in \mathbb{N}$ and that $\sum_{k=0}^{\infty} \binom{\alpha+k-1}{k} p^\alpha q^k = p^\alpha (1-q)^{-\alpha} = 1$, on account of (1.9).

The probability generating function of the negative binomial distribution is given by

$$\phi(z) = \sum_{k=0}^{\infty} \binom{\alpha+k-1}{k} p^\alpha q^k z^k = \left(\frac{p}{1-qz}\right)^\alpha.$$

If X is a random variable with this distribution then $\mathbb{E}X = \phi'(1) = \alpha q \frac{p^\alpha}{(1-q)^{\alpha+1}}$ or

$$\mathbb{E}X = \alpha \frac{q}{p}.$$

Similarly, $\mathbb{E}X(X-1) = \phi''(1) = \alpha(\alpha+1) q^2 \frac{p^\alpha}{(1-q)^{\alpha+2}} = \alpha(\alpha+1)\left(\frac{q}{p}\right)^2$. Thus we have $\mathbb{E}X^2 = \alpha(\alpha+1)\left(\frac{q}{p}\right)^2 + \alpha\frac{q}{p}$ and thus $\mathsf{Var}(X) = \alpha(\alpha+1)\left(\frac{q}{p}\right)^2 + \alpha\frac{q}{p} - \left(\alpha\frac{q}{p}\right)^2 = \alpha\frac{q}{p}\left(1+\frac{q}{p}\right)$ or

$$\mathsf{Var}(X) = \alpha \frac{q}{p^2}.$$

When $\alpha = m \in \mathbb{N}$ then the negative binomial random variable can be thought of as a sum of m independent geometric random variables with distribution (1.6). This follows readily by comparing the corresponding generating functions.

Chapter 2

Distributions on $\mathbb{R}$

The statistics of a real random variable X are determined by its distribution function $F(x) := \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$. It is clear that F is nondecreasing and that $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. F is defined to be right-continuous. Note that $\mathbb{P}(a < X \leq b) = F(b) - F(a)$. If x is a point of discontinuity of F then x is called an *atom* of the distribution and $\mathbb{P}(X = x) = F(x) - F(x-) > 0$. If on the other hand x is a point of continuity of F then $\mathbb{P}(X = x) = 0$. F can have at most countably many discontinuity points. If there exists a nonnegative f such that

$$F(x) = \int_{-\infty}^{x} f(y)dy, \qquad x \in \mathbb{R}$$

then F is called an absolutely continuous distribution and f is (a version of) the density of F. Most of the distributions we will consider here will have densities though occasionally we will find it useful to think in terms of more general distribution functions. Most of the time we will also be thinking in terms of distributions on $\mathbb{R}^+$, i.e. distributions for which $F(0-) = 0$. The function $\overline{F}(x) := 1 - F(x)$ is called the *tail* of the distribution function. The moment of order k of a distribution is defined as

$$m_k := \int_{-\infty}^{\infty} x^k dF(x),$$

provided that the integral exists.

The *moment generating function* that corresponds to a distribution F is defined as

$$M(\theta) := \mathbb{E}e^{\theta X} = \int_{-\infty}^{\infty} e^{\theta x} dF(x)$$

for all values of θ for which the integral converges. If there exists $\epsilon > 0$ such that $M(\theta)$ is defined in $(-\epsilon, +\epsilon)$ then the corresponding distribution is called *light–tailed*. In that case one can show that repeated differentiation inside the integral is permitted and thus $M^{(k)}(\theta) = \int_{-\infty}^{\infty} x^k e^{\theta x} dF(x)$ for $\theta \in (-\epsilon, +\epsilon)$. Thus we see that F has moments of all orders and

$$M^{(k)}(0) = m_k,$$

$$M(\theta) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} m_k.$$

This justifies the name “moment generating function”. There exist however many distributions for which the moment generating function does not exist for all values of $\theta \in \mathbb{R}$. We shall see such examples in the sequel. In fact it is possible that the integral defining the moment generating function exists only for $\theta = 0$. This is the case for instance in the ”double-sided Pareto” distribution with density $f(x) = \frac{\alpha}{2|x|^{\alpha+1}}$, $|x| \geq 1$, $\alpha > 0$.

Convergence problems, such as the ones just mentioned, are usually sidestepped by examining the characteristic function $\int_{\mathbb{R}} e^{itx} dF(x)$. In this case the defining integral converges for all $t \in \mathbb{R}$. Also, particularly when dealing with nonnegative random variables, it is often customary to examine the so-called Laplace transform which is defined as $\int e^{-sx} dF(x)$. For nonnegative random variables the Laplace transform always exists for $s \geq 0$. The only difference between Laplace transforms and moment generating functions is of course the sign in the exponent and thus all statements regarding moment generating functions carry over to Laplace transforms *mutatis mutandis*.

Scale and location parameters. Let X a random variable with distribution F (and density f). If $Y = aX + b$ where ($a > 0$ and $b \in \mathbb{R}$) then the distribution $G(x) := \mathbb{P}(Y \leq x)$ of Y is given by

$$G(x) = \mathbb{P}(X \leq (x-b)/a) = F\left(\frac{x-b}{a}\right).$$

a is called a *scale* parameter while b a *location* parameter. The density of G, g, is given by

$$g(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right).$$

Note in particular that $EY = aEX + b$ and $\mathsf{Var}(Y) = a^2 \mathsf{Var}(X)$. Thus if X is “standardized” with mean 0 and standard deviation 1, then Y has mean b and standard deviation a. Also, if $M_X(\theta) = \mathbb{E}e^{\theta X}$ is the moment generating function of X, then the moment generating function of Y is

$$M_Y(\theta) = Ee^{\theta(aX+b)} = e^{\theta b} M_X(a\theta). \tag{2.1}$$

2.1 Some distributions and their moment generating functions

In this section we give the definition of several continuous distributions that will play an important role in the sequel. Many of their properties will be explored in later sections.

2.1.1 The normal distribution

This is the most important distribution in probability theory. The standard normal distribution has density given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad x \in \mathbb{R}. \tag{2.2}$$

The distribution function of the standard normal, denoted by

$$\Phi(x) := \int_{-\infty}^{x} \varphi(y) dy, \qquad x \in \mathbb{R} \tag{2.3}$$

cannot be expressed in terms of elementary functions. Its values are available in tables. If X has the standard normal density then one can readily check (by a symmetry argument) that $\mathbb{E}X = 0$.

Also, an integration by parts shows that $\mathsf{Var}(X) = 1$. We denote the standard normal distribution as $\mathcal{N}(0,1)$. The general normal random variable can be obtained via a location-scale transformation: If X is $\mathcal{N}(0,1)$ then $Y = \sigma X + \mu$ (with $\sigma > 0$) has mean μ and variance σ^2. Its density is given by

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.4}$$

and of course its distribution function by $F(x) = \Phi(\frac{x-\mu}{\sigma})$. It is denoted by $\mathcal{N}(\mu, \sigma^2)$.

The moment generating function of the standard normal distribution is given by

$$\begin{aligned} M(\theta) &= \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{\infty} e^{\frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(x^2 - 2\theta x + \theta^2\right)} dx \\ &= e^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-\theta)^2} dx \\ &= e^{\frac{1}{2}\theta^2}, \end{aligned} \tag{2.5}$$

where in the last equality we have used the fact that $\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-\theta)^2}$ is a probability density function. Thus, using (2.1), for a $\mathcal{N}(\mu, \sigma^2)$ normal distribution the corresponding moment generating function is given by

$$M(\theta) = e^{\mu\theta + \frac{1}{2}\theta^2\sigma^2}, \qquad \theta \in \mathbb{R}. \tag{2.6}$$

Note that the moment generating function is defined for all $\theta \in \mathbb{R}$.

While $\Phi(x)$ cannot be expressed in closed form in terms of elementary functions, some particularly useful bounds for the tail of the distribution, $\overline{\Phi}(x) := 1 - \Phi(x)$ are easy to derive. We mention them here for future reference.

Proposition 1. *For all $x > 0$ we have*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \leq 1 - \Phi(x) \leq \frac{1}{x}\, \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \tag{2.7}$$

Proof: The tail is given by $\overline{\Phi}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$. The upper bound for the tail follows immediately from the inequality

$$\int_x^\infty e^{-\frac{1}{2}u^2} du \leq \int_x^\infty \frac{u}{x} e^{-\frac{1}{2}u^2} du = \frac{1}{x}\int_x^\infty e^{-\frac{1}{2}u^2} d(\frac{1}{2}u^2) = \frac{1}{x} e^{-\frac{1}{2}x^2}$$

(remember that $x > 0$).

The lower bound can be obtained by the following integration by parts formula

$$\begin{aligned} 0 \leq \int_x^\infty \frac{3}{u^4} e^{-\frac{1}{2}u^2} du &= -\frac{1}{u^3} e^{-\frac{1}{2}u^2}\Big|_x^\infty - \int_x^\infty \frac{1}{u^2} e^{-\frac{1}{2}u^2} du \\ &= \frac{1}{x^3} e^{-\frac{1}{2}x^2} - \int_x^\infty \frac{1}{u^2} e^{-\frac{1}{2}u^2} du \\ &= \frac{1}{x^3} e^{-\frac{1}{2}x^2} - \frac{1}{x} e^{-\frac{1}{2}x^2} + \int_x^\infty e^{-\frac{1}{2}u^2} du. \end{aligned} \tag{2.8}$$

∎

2.1.2 The exponential distribution

The distribution function is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases},$$

(where $\lambda > 0$ is called the *rate*) with corresponding density

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases}.$$

The mean of the exponential distribution is $\frac{1}{\lambda}$ and the variance $\frac{1}{\lambda^2}$. Its moment generating function is given by

$$\int_0^\infty e^{\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - \theta}, \qquad \text{for } \theta < \lambda.$$

2.1.3 The Gamma distribution

The density function is given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \beta \dfrac{(\beta x)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} & \text{if } x > 0 \end{cases}.$$

β is often called the scale parameter, while α the shape parameter. The Gamma function, which appears in the above expressions is defined via the integral

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \qquad x > 0, \tag{2.9}$$

and satisfies the functional equation

$$\Gamma(x+1) = x\Gamma(x).$$

In particular, when x is an integer, say n,

$$\Gamma(n) = (n-1)!\ .$$

(This can be verified by evaluating the integral in (2.9).) We also mention that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

The corresponding distribution function is

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \displaystyle\int_0^x \beta \frac{(\beta u)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta u} du & \text{if } x > 0 \end{cases} \qquad \alpha > 0,$$

which can be expressed in terms of the incomplete gamma function defied as $\Gamma(z, \alpha) := \int_0^z t^{\alpha-1} e^{-t} dt$.

The moment generating function of the Gamma distribution is

$$M(\theta) = \int_0^\infty e^{x\theta} \beta \frac{(\beta x)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} dx = \left(\frac{\beta}{\beta - \theta}\right)^\alpha.$$

Note that $M(\theta)$ above is defined only in the interval $-\infty < \theta < \beta$ because when $\theta \geq \beta$ the defining interval does not converge. It is easy to see that, for $\alpha = 1$ the Gamma distribution reduces to the exponential.

A special case of the Gamma distribution is the so-called Erlang distribution obtained for integer $\alpha = k - 1$ (We have also renamed β into λ). Its density is given by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \\ \lambda \dfrac{(\lambda x)^k}{k!} e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

with corresponding distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \\ 1 - \displaystyle\sum_{i=0}^{k-1} \lambda \frac{(\lambda x)^i}{i!} e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

Its moment generating function is of course $\left(\frac{\lambda}{\lambda - \theta}\right)^k$. One of the reasons for the importance of the Erlang distribution stems from the fact that it describes the sum of k independent exponential random variables with rate λ.

2.1.4 The Pareto distribution

The Pareto density has the form

$$f(x) = \begin{cases} 0 & \text{if } x \leq c \\ \\ \dfrac{\alpha c^\alpha}{x^{\alpha+1}} & \text{if } x > c \end{cases}$$

with corresponding distribution function

$$F(x) = \begin{cases} 0 & \text{if } x \leq c \\ \\ 1 - \left(\dfrac{c}{x}\right)^\alpha & \text{if } x > c \end{cases}$$

where $\alpha > 0$. The Pareto distribution is a typical example of a *subexponential distribution*. The nth moment of the Pareto distribution is given by the integral $\int_c^\infty x^n \alpha c^\alpha x^{-\alpha-1} dx$ provided that it is finite. Hence the nth moment exists if $\alpha > n$ and in that case it is equal to $\frac{\alpha c^n}{\alpha - n}$. In particular the mean exists only if $\alpha > 1$ and in that case it is equal to $\frac{c\alpha}{\alpha - 1}$

An alternative form of the Pareto which is non-zero for all $x \geq 0$ is given by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \\ \dfrac{\alpha}{c(1 + x/c)^{\alpha+1}} & \text{if } x \geq 0 \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \\ 1 - \dfrac{1}{(1 + x/c)^{\alpha}} & \text{if } x \geq 0 \end{cases}$$

where $\alpha > 0$.

2.1.5 The Cauchy distribution

The standardized Cauchy density is given by

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \qquad x \in \mathbb{R},$$

with distribution function

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x), \qquad x \in \mathbb{R}.$$

It has "fat" polynomial tails: In fact using de l'Hôpital's rule we see that

$$\lim_{x \to \infty} x\overline{F}(x) = \lim_{x \to \infty} \frac{\overline{F}(x)}{x^{-1}} = \lim_{x \to \infty} \frac{f(x)}{x^{-2}} = \lim_{x \to \infty} \frac{x^2}{\pi(1 + x^2)} = \frac{1}{\pi}.$$

This it does not have a mean or a variance because the integrals that define them do not converge. It is useful in modelling phenomena that can produce large claims.

2.1.6 The Weibull distribution

The distribution function is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-x^{\beta}} & \text{if } x > 0 \end{cases}$$

with corresponding density

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \beta x^{\beta - 1} e^{-x^{\beta}} & \text{if } x > 0 \end{cases}$$

The nth moment of this distribution is given by

$$\int_0^{\infty} \beta x^{n + \beta - 1} e^{-x^{\beta}} dx = \int_0^{\infty} y^{n/\beta} e^{-y} dy = \Gamma\left(\frac{n}{\beta} + 1\right).$$

2.2 Sums of independent random variables in $\mathbb{R}^+$

Suppose that F, G, are two distributions on $\mathbb{R}^+$. Their convolution is defined as the function

$$F \star G(x) = \int_0^x F(x - y) dG(y), \qquad x \geq 0. \tag{2.10}$$

If X, Y, are independent random variables with distributions F and G respectively, then $F \star G$ is the distribution of their sum $X+Y$. Indeed,

$$\begin{aligned} \mathbb{P}(X+Y \leq x) &= \int_0^\infty \mathbb{P}(X+Y \leq x | Y=y) dG(y) = \int_0^\infty \mathbb{P}(X \leq x-y | Y=y) dG(y) \\ &= \int_0^\infty F(x-y) dG(y) = \int_0^x F(x-y) dG(y). \end{aligned}$$

In the above string of equalities we have used the independence of X and Y to write $\mathbb{P}(X \leq x-y|Y=y) = F(x-y)$ and the fact that $F(x-y)=0$ for $y>x$ to restrict the range of integration. In view of this last remark it is clear that $F \star G = G \star F$. We will also write $F^{\star n}$ to denote the n–fold convolution $F \star F \star \cdots \star F$ (with n factors) with the understanding that $F^{\star 1} =$ and $F^{\star 0} = I$ where $I(x) = 1$ if $x \geq 0$ and $I(x) = 0$ when $x < 0$. When both F and G are absolutely continuous with densities f and g respectively then $H = F \star G$ is again absolutely continuous with density

$$h(x) = \int_0^x f(x-y) g(y) dy.$$

We will denote the convolution of the two densities by $h = f * g$. For instance, if $f(x) = \lambda e^{-\lambda x}$, $g(x) = \mu e^{-\mu x}$, then

$$f * g(x) = \int_0^x \lambda\mu e^{-\lambda(x-y)} e^{-\mu y} dy = \lambda\mu e^{-\lambda x} \frac{\left(1-e^{-(\mu-\lambda)x}\right)}{\mu-\lambda} = \frac{\lambda\mu}{\mu-\lambda}\left(e^{-\lambda x} - e^{-\mu x}\right).$$

Note that, if X, Y, are independent then the moment generating function of the sum $X+Y$ is given by

$$M_{X+Y}(\theta) = \mathbb{E} e^{\theta(X+Y)} = \mathbb{E} e^{\theta X} e^{\theta Y} = M_X(\theta) M_Y(\theta).$$

If X_i, $i = 1, 2, \ldots, n$ are independent, identically distributed random variables with distribution F and moment generating function $M_X(\theta)$ then $S := X_1 + \cdots + X_n$ has distribution function $F^{\star n}$ and moment generating function $M_S(\theta) = (M_X(\theta))^n$.

Convolutions are in general hard to evaluate explicitly. As an exception to this statement we mention the exponential distribution, $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$. In that case we have

$$F^{*n}(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}.$$

(This is the well known Erlang distribution). More generally, if $F(x) = 1 - \sum_{k=0}^{m-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}$ then $F^{*n}(x) = 1 - \sum_{k=0}^{nm-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}$ and, more generally yet, if F is Gamma(α, λ) then F^* is Gamma$(n\alpha, \lambda)$.

2.3 Random Sums

Suppose that X_i, $i = 1, 2, \ldots$ is a sequence of non-negative random variables with distribution function F and moment generating function $M_X(\theta) := \int_0^\infty e^{\theta x} dF(x)$. Suppose also that N is a discrete random variable, independent of the X_i's, $i = 1, 2, \ldots$. Let $S_N = \sum_{i=1}^N X_i$. The distribution and the moments of S_N can be obtained by conditioning on N. For instance

$$\mathbb{P}(S_N \leq x) = \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{P}(X_1 + \cdots + X_n \leq x) = \sum_{n=0}^{\infty} \mathbb{P}(N = n)F^{\star n}(x). \tag{2.11}$$

The mean and the variance of S_N can be computed in the same fashion.

$$\mathbb{E}S_N = \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}[X_1 + \cdots + X_n] = \sum_{n=1}^{\infty} \mathbb{P}(N = n)n\mathbb{E}X_1 = \mathbb{E}N\mathbb{E}X_1. \tag{2.12}$$

Also

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2 = \mathbb{E}\left[\sum_{n=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right] = n\mathbb{E}X_1^2 + n(n-1)(\mathbb{E}X_1)^2$$

and thus

$$\begin{aligned}
\mathbb{E}S_N^2 &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}[(X_1 + \cdots + X_n)^2] = \sum_{n=1}^{\infty} \mathbb{P}(N = n)\left(n\mathbb{E}X_1^2 + n(n-1)(\mathbb{E}X_1)^2\right) \\
&= \mathbb{E}(X_1^2)\mathbb{E}N + (\mathbb{E}X_1)^2 \sum_{n=1}^{\infty} n(n-1)\mathbb{P}(N = n) \\
&= \mathbb{E}(X_1^2)\mathbb{E}N + (\mathbb{E}X_1)^2\mathbb{E}N^2 - (\mathbb{E}X_1)^2\mathbb{E}N \\
&= \mathsf{Var(X_1)}\mathbb{E}N + (\mathbb{E}X_1)^2\mathbb{E}N^2. \qquad (2.13)
\end{aligned}$$

From (2.12) and (2.13) we obtain

$$\mathsf{Var}(S_N) = \mathsf{Var}(X_1)\mathbb{E}N + \mathsf{Var}(N)(\mathbb{E}X_1)^2. \tag{2.14}$$

Finally we can also compute the moment generating function of S_N by conditioning:

$$\begin{aligned}
M_{S_N}(\theta) &= \mathbb{E}e^{\theta S_N} = \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}e^{\theta \sum_{i=1}^{n} X_i} = \sum_{n=0}^{\infty} \mathbb{P}(N = n)\left(\mathbb{E}e^{\theta X_1}\right)^n \\
&= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\left(M_X(\theta)\right)^n.
\end{aligned}$$

If we denote by $\phi_N(z) = \sum_{n=0}^{\infty} \mathbb{P}(N = n)z^n$ the p.g.f. of N we see from the above that

$$M_{S_N}(\theta) = \phi_N(M_X(\theta)). \tag{2.15}$$

Chapter 3

The Central Limit Theorem and Logarithmic Asymptotics

3.1 Premium Computation Using the Central Limit Theorem

Suppose that the insurance company has a portfolio consisting of n policies. Over one period each one of these policies generates a claim, the ith claim being a random variable X_i, $i = 1, 2, \dots, n$. These are assumed to be independent, identically distributed random variables with common distribution F and mean m. The premium the insurance company receives for each company is p. If at the beginning of the period the free reserves of the insurance company are u, then at the end of the period they are $u + np - \sum_{i=1}^{n} X_i$. If the premium p is greater than the expected size of the claim (a situation which may be acceptable to the insurance buyer as we saw in the previous section) then by virtue of the law of large numbers the insurer is likely to profit and extremely unlikely to be unable to cover the total losses. To see this note that the event that the income from the premiums np plus the initial free reserves u will not suffice to cover the losses, happens with probability $\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > p + u/n)$. If $p > m$ then by virtue of the weak law of large numbers we have that $\lim_{n\to\infty} \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > p + u/n) = 0$.

The above limiting argument makes it very implausible that for large but finite n the losses will exceed the premium income and safety reserves. In order to quantify this a study of the fluctuations of the sums of random variables is necessary. This study starts with the central limit theorem.

Let us start by setting $u/n =: v$, the total free reserves per contract and $c := p + v$. Then, in order to set the probability that the losses will exceed the available reserves equal to α, (a small number, typically 0.1% – 1%) we should choose c such that

$$\mathbb{P}\left(X_1 + X_2 + \cdots + X_n \geq nc\right) = \alpha.$$

In order to take advantage of the Central Limit theorem (CLT) let us rewrite the above equation as

$$\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} \geq \frac{c-m}{\sigma}\sqrt{n}\right) = \alpha. \tag{3.1}$$

Assuming n to be large enough to justify our appeal to the CLT we then have

$$\frac{c-m}{\sigma}\sqrt{n} = z_{1-\alpha}$$

where $\Phi(z_{1-\alpha}) = 1 - \alpha$. Thus

$$c = m + z_{1-\alpha}\sigma\frac{1}{\sqrt{n}} \tag{3.2}$$

and hence

$$p = \frac{u}{n} + m + z_{1-\alpha}\sigma\frac{1}{\sqrt{n}}. \tag{3.3}$$

The above argument often provides the basis for rate setting but it should be used with care. There are three reasons that could lead to erroneous estimation of the probability of ruin:

a) The risks in the portfolio are inhomogeneous to such a degree that the equidistribution approach is not justified

b) The claim distribution is *heavy tailed* ($\sigma = \infty$) and therefore the CLT cannot be applied

c) We are interested in "rare" events, out in the tail of the distribution of $X_1 + X_2 + \cdots + X_n$ where the CLT does not hold

We will have the opportunity to look at the issues raised in a) and b). Regarding c) let us examine more closely the approximation involved in (3.1). In order to be justified in applying the CLT we should be prepared to let $n \to \infty$. However, the right hand side of the inequality in (3.1) also goes to infinity with n and therefore the only thing we learn from the CLT is that $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n-nm}{\sigma\sqrt{n}} > \frac{c-m}{\sigma}\sqrt{n}\right) \to 0$ as $n \to \infty$.

Assuming that the CLT approximation can be used,

$$\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} > \frac{c-m}{\sigma}\sqrt{n}\right) \approx \overline{\Phi}\left(\frac{c-m}{\sigma}\sqrt{n}\right)$$

where $\overline{\Phi}(x) = 1 - \Phi(x)$. Then it is easy to see that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} > \frac{c-m}{\sigma}\sqrt{n}\right) = \lim_{n\to\infty}\frac{1}{n}\log\overline{\Phi}\left(\frac{c-m}{\sigma}\sqrt{n}\right).$$

This last limit will be computed in the next section using the inequalities of proposition 1 and we will see that

$$\lim_{n\to\infty}\frac{1}{n}\log\overline{\Phi}\left(\frac{c-m}{\sigma}\sqrt{n}\right) = -\frac{1}{2}\left(\frac{c-m}{\sigma}\right)^2.$$

We conclude from the above analysis that, if the probability of ruin $\mathbb{P}(X_1 + X_2 + \cdots + X_n \geq n(c-m))$ is set equal to α (a small number) then for large n

$$\frac{1}{n}\log\alpha \approx -\frac{1}{2}\left(\frac{c-m}{\sigma}\right)^2$$

or equivalently

$$c = m + \sigma\frac{2\log(1/\alpha)}{\sqrt{n}}. \tag{3.4}$$

This last equation is to be compared with (3.2).

3.2 Logarithmic Asymptotics

Suppose that X_i, $i = 1, 2, 3, \ldots$, are i.i.d. with distribution function F, corresponding mean $m = \int_{\mathbb{R}} xF(dx)$, variance σ^2, and moment generating function $M(\theta) := \int_{\mathbb{R}} e^{\theta x} F(dx)$. The weak law of large numbers guarantees that

$$\lim_{n\to\infty} \mathbb{P}(S_n \geq nx) = 0 \qquad \text{for } x > m \tag{3.5}$$

and similarly that

$$\lim_{n\to\infty} \mathbb{P}(S_n \leq nx) = 0 \qquad \text{for } x < m \tag{3.6}$$

Correspondingly, if the premium charged per policy, x, is higher than the expected claim size, m, then the probability or ruin goes to zero, whereas if the it is less than m then ruin is certain.

One important question however not answered by the weak law of large numbers is *how fast do the above probabilities go to zero*. We will see that they go to zero exponentially fast, i.e. that

$$\mathbb{P}(S_n \geq nx) \asymp e^{-nI(x)} \qquad \text{for } x > m \tag{3.7}$$

In the above formula note that the exponential rate of decay $I(x)$ is a function of x. The meaning of (3.7) is made precise if we state it as

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nx) = -I(x) \qquad \text{for } x > m. \tag{3.8}$$

Where does the exponential behavior come from? Write $\mathbb{P}(S_n \geq nx)$ as

$$\mathbb{P}\left(S_n - nm \geq n(x-m)\right) = \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} \geq \sqrt{n}\left(\frac{x-m}{\sigma}\right)\right) \tag{3.9}$$

and appeal to the central limit theorem: For n large $\frac{S_n - nm}{\sigma\sqrt{n}}$ is approximately normally distributed with mean 0 and standard deviation 1 and hence

$$\begin{aligned} \mathbb{P}(S_n \geq nx) &= \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} \geq \sqrt{n}\left(\frac{x-m}{\sigma}\right)\right) \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}\left(\frac{x-m}{\sigma}\right)}^{\infty} e^{-\frac{1}{2}u^2} du \\ &\approx \frac{\sigma}{(x-m)\sqrt{2\pi n}} e^{-n\frac{(x-m)^2}{2\sigma^2}}. \end{aligned}$$

Are the above asymptotics justified? In one case at least yes. Suppose that the r.v.'s X_i, are i.i.d. are normal with mean m and variance σ^2 ($N(m, \sigma^2)$). Then S_n/n has distribution $N\left(m, \frac{\sigma^2}{n}\right)$. Hence in this case (3.9) becomes an exact relationship and we have

$$\mathbb{P}(S_n \geq nx) = \int_{\sqrt{n}\left(\frac{x-m}{\sigma}\right)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du. \tag{3.10}$$

Taking into account the bounds in proposition 1 we have

$$\begin{aligned} \log\left(\left(\frac{1}{n^{1/2}}\frac{\sigma}{x-m} - \frac{1}{n^{3/2}}\frac{\sigma^3}{(x-m)^3}\right)\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}n\left(\frac{x-m}{\sigma}\right)^2}\right) &\leq \log \mathbb{P}(S_n \geq nx) \\ &\leq \log\left(\frac{1}{n^{1/2}}\frac{\sigma}{x-m}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}n\left(\frac{x-m}{\sigma}\right)^2}\right) \end{aligned}$$

or

$$-\frac{1}{2}\log n + \log\left(\tfrac{\sigma}{x-m} - \tfrac{1}{n}\tfrac{\sigma^3}{(x-m)^3}\right) - \frac{1}{2}\log 2\pi - \frac{1}{2}n\left(\frac{x-m}{\sigma}\right)^2 \quad \leq \quad \log \mathbb{P}(S_n \geq nx)$$
$$\leq \quad -\frac{1}{2}\log n + \log\tfrac{\sigma}{x-m} - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2}n\left(\tfrac{x-m}{\sigma}\right)^2.$$

Dividing the above inequality with n and letting $n \to \infty$ (taking into account that $\frac{1}{n}\log n \to 0$) we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}(S_n \geq nx) = -\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2. \tag{3.11}$$

Hence, setting $I(x) = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2$ we obtain (3.5) for normal random variables. Can we generalize this to non–normal random variables? Can we generalize it for sequences that are not independent and identically distributed?

As we will see the answer is in the affirmative on both counts. We start with a relatively simple bound known as the Chernoff bound.

3.3 Chernoff bounds

In the same framework as before X_i, $i = 1, 2, \ldots$ are assumed to be i.i.d. r.v.'s with moment generating function $M(\theta)$. We start with the obvious inequality

$$\mathbf{1}(S_n \geq nx)e^{nx\theta} \leq e^{\theta S_n}$$

which holds for all $\theta \geq 0$ since the exponential is non–negative. Taking expectations in the above inequality we obtain

$$\mathbb{P}(S_n \geq nx) \leq e^{-nx\theta}\mathbb{E}[e^{\theta X_1 + X_2 + \cdots + X_n}] = e^{-nx\theta}M(\theta)^n, \qquad \theta \geq 0.$$

The above inequality provides an upper bound for $\mathbb{P}(S_n \geq nx)$ for each $\theta \in \mathbb{R}^+$. Since the left hand side in the above inequality does not depend on θ we can obtain the best possible bound by setting

$$\mathbb{P}(S_n \geq nx) \leq \inf_{\theta \geq 0} e^{-n\{x\theta - \log M(\theta)\}} = e^{-n\sup_{\theta\geq 0}\{x\theta - \log M(\theta)\}}.$$

Define now the *rate* function

$$I(x) := \sup_{\theta\in\mathbb{R}}\{x\theta - \log M(\theta)\}. \tag{3.12}$$

With this definition the Chernoff bound becomes

$$\mathbb{P}(S_n \geq nx) \leq e^{-nI(x)}. \tag{3.13}$$

As we will see in many cases this upper bound can be turned into an asymptotic inequality. This is the content of Cramér's theorem.

Theorem 1. *The cumulant* $\log M(\theta)$ *is a convex function of* θ.

Proof: To establish this we will show that the second derivative $\frac{d^2}{d\theta^2}\log M(\theta)$ is non–negative. Indeed

$$\frac{d^2}{d\theta^2}\log M(\theta) = \frac{M''(\theta)}{M(\theta)} - \left(\frac{M'(\theta)}{M(\theta)}\right)^2$$

However note that

$$M''(\theta) = \frac{d^2}{d\theta^2}\mathbb{E}[e^{\theta X}] = E[X^2 e^{\theta X}]$$

and hence

$$\frac{M''(\theta)}{M(\theta)} = \mathbb{E}[X^2 \frac{e^{\theta X}}{M(\theta)}] = E_{\widetilde{\mathbb{P}}}[X^2].$$

Similarly

$$\frac{M'(\theta)}{M(\theta)} = \mathbb{E}[X \frac{e^{\theta X}}{M(\theta)}] = \mathbb{E}_{\widetilde{\mathbb{P}}}[X]$$

and thus

$$\frac{d^2}{d\theta^2}\log M(\theta) = \mathbb{E}_{\widetilde{\mathbb{P}}}[X^2] - \left(\mathbb{E}_{\widetilde{\mathbb{P}}}[X]\right)^2 = \mathbb{E}_{\widetilde{\mathbb{P}}}\left(X - \mathbb{E}_{\widetilde{\mathbb{P}}}[X]\right)^2 \geq 0.$$

∎

3.4 Examples of rate functions

Bernoulli Random Variables Suppose that $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p$ (i.e. the random variables take only the values 0 and 1 with probabilities $1-p$ and p respectively). In this case $\log M(\theta) = \log\left(pe^\theta + 1 - p\right)$. To maximize $x\theta - \log M(\theta)$ we set its derivative equal to zero: $x = \frac{pe^\theta}{1-p+pe^\theta}$ or $e^\theta = \frac{x}{1-x}\frac{1-p}{p}$ and, taking logarithms,

$$\theta = \log\frac{x}{1-x} + \log\frac{1-p}{p}.$$

Therefore

$$I(x) = \begin{cases} x\log\dfrac{x}{p} + (1-x)\log\dfrac{1-x}{1-p}, & 0 < x < 1 \\ \\ \infty, & \text{otherwise} \end{cases}$$

Normal $N(\mu, \sigma^2)$ Here $M(\theta) = e^{\theta\mu + \frac{1}{2}\theta^2\sigma^2}$. The rate function is given by

$$I(x) = \sup_\theta \left[\theta x - \theta\mu - \frac{1}{2}\theta^2\sigma^2\right].$$

Differentiating we obtain $(x-\mu) - \theta\sigma^2 = 0$ or $\theta = \frac{x-\mu}{\sigma^2}$. Substituting back we get

$$I(x) = \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2.$$

Exponential (rate λ) In this case $M(\theta) = \frac{\lambda}{\lambda-\theta}$ and thus the rate function is obtained by maximizing the expression $\theta x - \log\frac{\lambda}{\lambda-\theta}$. The optimal value of θ is obtained by the solution of the equation $x - \frac{1}{\lambda-\theta} = 0$ or $\theta = \lambda - 1/x$ which gives

$$I(x) = \begin{cases} \lambda x - \log \lambda x - 1, & x > 0 \\ +\infty, & x \leq 0 \end{cases}$$

Binomial (number of trials n, probability of success p) Here $M(\theta) = (1-p+pe^{\theta})^n$ (note the close connection with the Bernoulli distribution) and $\log M(\theta) = n\log(1-p+pe^{\theta})$. Thus, arguing as in the Bernoulli case, we see that $x\theta - \log M(\theta)$ is maximized for $\theta^* = \log\left(\frac{x(1-p)}{(k-x)p}\right)$ and hence

$$I(x) = \begin{cases} x\log\dfrac{x}{p} + (n-x)\log\dfrac{n-x}{1-p} - n\log n, & 0 < x < n \\ \\ \infty, & \text{otherwise} \end{cases}$$

Geometric (probability of success p) Here

$$M(\theta) = \frac{1-p}{1-pe^{\theta}}.$$

Following the same procedure as before we obtain

$$I(x) = \begin{cases} x\log x - (x+1)\log(x+1) + x\log\dfrac{1}{p} - \log(1-p), & x > 0 \\ \\ +\infty, & x \leq 0 \end{cases}$$

In the following graph the rate function of the geometric distribution (with $p = 1/2$) is shown.

3.5 Properties of the rate function

Let $D = \{x : I(x) < \infty\}$ be the domain of definition of I. It is easy to see that D is either the whole of $\mathbb{R}$ or an interval that may extend infinitely to the right or the left. If the upper or lower end of the interval is finite it may or may not belong to D depending on the case. Thus in any case D is a convex set in $\mathbb{R}$.

1. $I(x)$ is a convex function (on its domain of definition). It suffices to show that, for each $\lambda \in [0,1]$, $x, y \in D$, $I(x\lambda + y(1-\lambda)) \leq \lambda I(x) + (1-\lambda)I(y)$. Indeed,

$$\begin{aligned} I(x\lambda + y(1-\lambda)) &= \sup_\theta \{\theta(x\lambda + y(1-\lambda)) - \log M(\theta)\} \\ &= \sup_\theta \{\lambda(\theta x - \log M(\theta)) + (1-\lambda)(\theta x - \log M(\theta))\} \\ &\leq \lambda \sup_\theta \{\theta x - \log M(\theta)\} + (1-\lambda)\sup_\theta \{\theta y - \log M(\theta)\} \\ &= \lambda I(x) + (1-\lambda)I(y) \end{aligned}$$

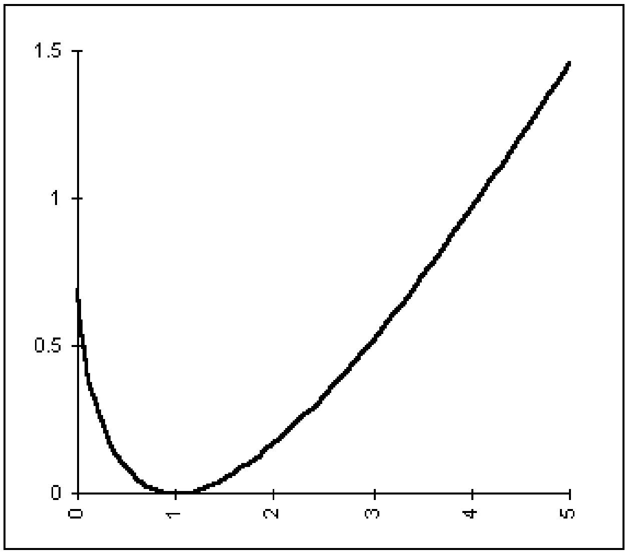

Figure 3.1: Rate function for the geometric distribution

2. $I(x) \geq 0$ *for all* $x \in D$ *and* $I(m) = 0$*.* (In particular this implies that I is minimized at $x = m$.) We begin with the remark that for $\theta = 0$, $\theta x - \log M(\theta) = 0$. Thus $I(x) \geq 0$. Next, use Jensen's inequality: $M(\theta) = \mathbb{E}e^{\theta X} \geq e^{\theta EX}$ for all θ for which $M(\theta) < \infty$. Thus $\log M(\theta) \geq \theta m$ or $\theta m - \log M(\theta) \leq 0$. Since $I(x) \geq 0$, we conclude that $I(m) = 0$.

3. For each $x \in D$ *there exists* θ^* *such that*

$$\frac{M'(\theta^*)}{M(\theta^*)} = x \tag{3.14}$$

We will not present a complete proof of this. A justification might be given along the following lines: since for fixed x the function $f(\theta) = \theta x - \log M(\theta)$ is convex in θ and smooth ($M(\theta)$ has derivatives of all orders) it suffices to find θ^* so that $f(\theta^*) = 0$ or equivalently $x - M'(\theta^*)/M(\theta^*) = 0$.

3.6 The twisted distribution

Let $F(y)$ be a distribution function on $\mathbb{R}$ with moment generating function $M(\theta)$. The distribution function $\widetilde{F}(y)$ defined via

$$d\widetilde{F}(dy) = \frac{e^{\theta y}}{M(\theta)} F(dy)$$

is called the *twisted distribution* that corresponds to F. It is easy to see that

$$\widetilde{F}(y) = \int_{-\infty}^{y} \frac{e^{\theta u}}{M(\theta)} F(du)$$

is a non–decreasing function of y and as $y \to \infty$, $\widetilde{F}(y) \to 1$.

The mean of the twisted distribution is given by

$$\int_{-\infty}^{\infty} y\widetilde{F}(dy) = \frac{1}{M(\theta)}\int_{-\infty}^{\infty} ye^{\theta y}F(dy) = \frac{1}{M(\theta)}\frac{d}{d\theta}\int_{-\infty}^{\infty} e^{\theta y}F(dy) = \frac{M'(\theta)}{M(\theta)}.$$

In particular when $\theta = \theta^*$, the solution of (3.14),

$$\frac{1}{M(\theta^*)}\int_{-\infty}^{\infty} ye^{\theta^* y}F(dy) = \frac{M'(\theta^*)}{M(\theta^*)} = x. \tag{3.15}$$

Regarding our notation, it will be convenient to think of two different probability measures, the probability measure $\mathbb{P}$, under which the random variables X_i, $i = 1, 2, \ldots$, have distribution F, and the twisted measure $\widetilde{\mathbb{P}}$, under which the r.v.'s X_i have distribution $\widetilde{F}$. Expectations with respect to the probability measure $\widetilde{\mathbb{P}}$ will be denoted by $\widetilde{\mathbb{E}}$.

3.7 Cramér's Theorem

Theorem 2. *Suppose that $\{X_n\}$ is an i.i.d. sequence of real random variables with moment generating function $M(\theta)$ which exists in an open neighborhood of zero. Then, if $m = \mathbb{E}X_1$ and $S_n := X_1 + \cdots + X_n$,*

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq nx) = -I(x), \qquad x \geq m,$$
$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \leq nx) = -I(x), \qquad x \leq m.$$

The intuitive content of the above theorem is that $\mathbb{P}(S_n \geq nx) \approx e^{-nI(x)}$ when $x > m$ for large values of n (with the corresponding approximation, $\mathbb{P}(S_n \leq nx) \approx e^{-nI(x)}$ holding when $x < m$).

To see how this works in practice suppose that X_i, $i = 1, 2, \ldots, n$ be independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = q := 1-p$. Then $m = 1\cdot p + 0\cdot q = p$. Let $S_n = \sum_{i=1}^n X_i$. Then S_n is Binomial and, as we saw in section 3.4, $I(x) = x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}$ for $0 < x < 1$. The following program in `R` computes the probabilities $\mathbb{P}(S_n \geq mx)$ using the function `pbinom(x,m,0.5)`. for $m = 200$ and various values of $x > 0.5$. It also computes the Large deviation estimate $e^{-mI(x)}$ and plots the two series of values.

```
# This is a comparison between the tail of the binomial distribution and the
#    approximation using Large Deviations theory

a<-0.02

I<-function(x) (x*log(2*x)+(1-x)*log(2*(1-x))) # This is the rate function of
ldp=matrix(0,1,20)                              #  a Bernoulli random variable
m<-200

for (j in 1:20)
{
```

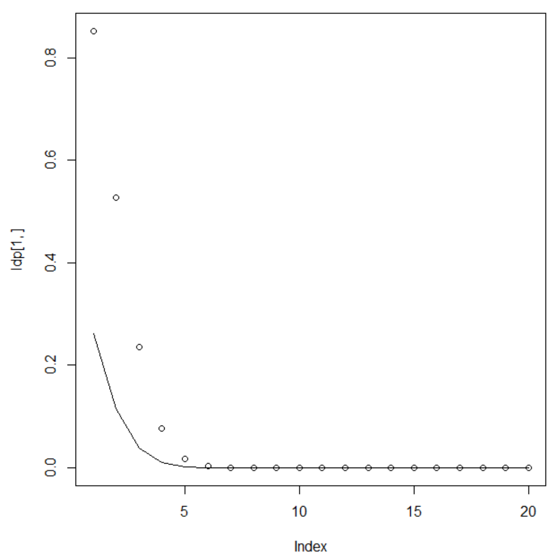

Figure 3.2: Tail of the Binomial Distribution

```
  x<- 0.5+ j*a
  ldp[j]<- exp(-m*I(x))
}

bb=matrix(0,1,20)
for (j in 1:20) {
  bb[j]<- 1-pbinom(m*(0.5+j*a),m,0.5) # Here the tail of the binomial distribution
}                                     # distribution is computed.
plot(ldp[1,])                         # The small circles show the Large Deviation
lines(bb[1,])                         # estimates. The line shows the exact value
                                      # of the tail probability.
```

Chapter 4

Random Walks and the Cramér-Lundberg Model

4.1 The Simple Random Walk

Let $\{\xi_n\}$, $n = 1, 2, \ldots$ be independent, identically distributed random variables with $\mathbb{P}(\xi_n = 1) = p$, $\mathbb{P}(\xi_n = -1) = q := 1 - p$. Consider the process $\{X_n; n = 0, 1, 2, \ldots\}$ defined by $X_0 := a$ (where $a \in \mathbb{Z}$ the given initial value of the process) and $X_n := X_{n-1} + \xi_n$, $n = 1, 2, \ldots$. Consider the following problem known as the *gambler's ruin*.

A gambler with initial fortune a each time places a unit bet which he wins with probability p or loses with probability $q = 1 - p$. Consecutive outcomes are independent and the process continues until either the gambler loses all his fortune or when his fortune reaches a predetermined level $b > a$. Thus after the n'th bet the gambler's fortune is X_n and the process stops at time $T = \min\{n > 0 : X_n = 0, \text{ or } X_n = b\}$. Let $\pi_a := \mathbb{P}(X_T = 0)$ denote the *probability of ruin* when the initial fortune of the gambler is a. Then $1 - \pi_a = \mathbb{P}(X_T = b)$ is the complementary probability that the gambler eventually reaches his goal and wins.

The ruin probability can be determined if we consider the initial fortune as a variable parameter. Then a "first step analysis" shows that

$$\pi_k \;=\; p\pi_{k+1} + q\pi_{k-1}, \qquad k = 1, 2, \ldots, b-1. \tag{4.1}$$

To see this suppose that the initial fortune is k and consider what happens after the first bet. (Mathematically this is justified by the independence of the outcomes of the bets, $\xi_1, \xi_2, \ldots$.) Equation (4.1) is complemented by the *boundary conditions*

$$\pi_0 \;=\; 1, \quad \pi_b \;=\; 1. \tag{4.2}$$

Consider solutions to equation (4.1) of the form $\pi_k = \lambda^k$. This is possible only if $\lambda^k = p\lambda^{k+1} + q\lambda^{k-1}$ or, equivalently, if λ is a root of the *characteristic equation* $\lambda = p\lambda^2 + q$. The general solution of (4.1) can be written as $\pi_k = C_1\lambda_1^k + C_2\lambda_2^k$ where λ_1, λ_2 are the two roots of the characteristic equation, $\lambda_1 = 1$, $\lambda_2 = q/p$. Thus $\pi_k = C_1 + C_2\left(\frac{q}{p}\right)^k$, and the unknown constants C_1, C_2, are

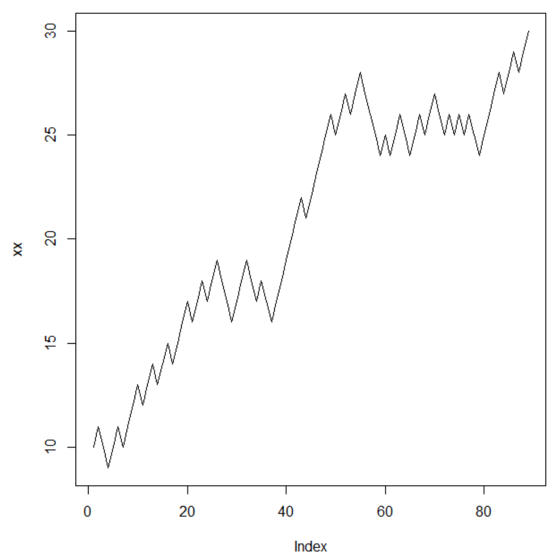

Figure 4.1: Sample path of a random walk

determined from the two boundary conditions (4.2). Thus the ruin probability is given by

$$\pi_a = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}. \tag{4.3}$$

A simulation of the simple random walk using R

```
a<-10
b<-30
p<-0.55
x<-a
xx<-a

while ( (x>0) & (x<b)){
  x<-x+2*rbinom(1,1,p)-1
  xx<-c(xx,x)
}
plot(xx,type="l")
```

The above program gives a single sample path of the process. A realization is shown in figure 4.1. If we need to estimate the probability of ruin many independent replications need to be carried out. The following programm does this.

```
a<-10
b<-30
```

```
p<-0.55
replications<-10000
pr<-{}

for (i in 1:replications){
ruin<-0
x<-a
while ( (x>0) & (x<b)){
  x<-x+2*rbinom(1,1,p)-1
  }
if (x==0) {ruin<-1}
pr<-c(pr,ruin)
}
mean(pr)
sd(pr)

thrp<-function(p,a,b){ (((1-p)/p)^a - ((1-p)/p)^b)/(1 - ((1-p)/p)^b)}
thrp(p,a,b)
```

In the above program `pr` is a string with 10000 elements zeros and ones depending on whether each run ends in non-ruin or ruin respectively. The last two lines provide the function that computes the theoretically expected value from (4.3). The output given by `R` is

```
> mean(pr)
[1] 0.1296
> sd(pr)
[1] 0.3358796
>
> thrp<-function(p,a,b){ (((1-p)/p)^a - ((1-p)/p)^b)/(1 - ((1-p)/p)^b)}
> thrp(p,a,b)
[1] 0.1323227
```

Note that the mean of `pr` is 0.1296, which is relatively close to the theoretically expected value of 0.1323. The standard deviation of the *mean* is $\texttt{sd(pr)}/\sqrt{\texttt{replications}} = 0.003359$. Hence, a 95% confidence interval for the true ruin probability is

$$\texttt{mean(pr)} \pm 1.96 \cdot \texttt{sd(pr)}/\sqrt{\texttt{replications}} \;=\; 0.3358796 \pm 1.96 \cdot 0.0033588$$

Thus the confidence interval is $[0.1230168\,,\,0.1361832]$ which of course contains the theoretical value.

4.2 The classical risk model of Cramér and Lundberg

The simplest model that describes the operation of an insurance company consists of the following elements. Initially the insurance firm starts with initial capital u which increases linearly with time with rate c because of incoming premiums. At times $\{S_n\}$, $n = 1, 2, \ldots$, claims arrive with respective sizes $\{Z_n\}$. Hence, a typical realization of this process has the form shown in figure 4.2.

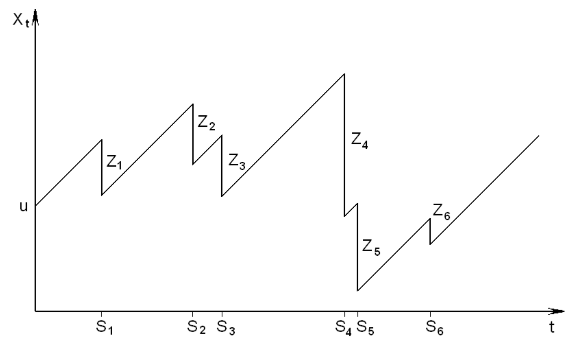

Figure 4.2: Sample path of the Cramér–Lundberg process

We denote by $\{X_t; t \in \mathbb{R}\}$ the process

$$X_t \; = \; u + ct - \sum_{k=1}^{N(t)} Z_k.$$

For any $T \geq 0$ let us denote by

$$\Psi(u;T) = \mathbb{P}\{\inf_{0 \leq t \leq T} X_t < 0\}$$

the probability that "ruin" occurs within the finite horizon $[0,T]$. Similarly, we will denote by

$$\Psi(u) = \mathbb{P}\{\inf_{t \geq 0} X_t < 0\}$$

the *infinite horizon ruin probability*. We will also denote by

$$\Phi(u) := 1 - \Psi(u)$$

the *infinite horizon non-ruin probability*. Suppose that the claim arrival process $\{S_n\}$ is a Poisson process with rate λ and the claims $\{Z_i\}$ are i.i.d. random variables with distribution F and mean μ. Clearly, from the Strong Law of Large Numbers, when $c < \lambda\mu$ then $X(t) \to -\infty$ w.p. 1 and hence ruin is certain eventually. Therefore the premium rate c must exceed the rate with which the company loses money because of the claims which on the average is $\lambda\mu$. (This is called the *net premium rate*.) The factor ρ by which the premium rate charged by the company exceeds the net premium rate is called *safety loading* i.e.

$$\boxed{\rho := \frac{c}{\lambda\mu} - 1} \qquad \text{or} \qquad \boxed{1 + \rho = \frac{c}{\lambda\mu}} \tag{4.4}$$

4.3 Integrodifferential Equation for the Non–ruin Probability

We start with

Lemma 1. $\Phi(u)$ *is non-decreasing in* u *and* $\lim_{u\to\infty} \Phi(u) = 1$.

Proof: Indeed, $\Phi(u) = \mathbb{P}\{\inf_{t\geq 0} X_t \geq 0\}$ and $B_u := \{\inf_{t\geq 0} X_t \geq 0\}$ is a non-decreasing family of sets in u in the sense that $u_1 < u_2$ implies that $B_{u_1} \subset B_{u_2}$. Hence, by the monotone continuity of the probability measure, if $\{u_n\}$ is a sequence such that $u_n \to \infty$ as $n \to \infty$, $\lim_{n\to\infty} \mathbb{P}\{\inf_{t\geq 0} X_t \geq 0\} = \mathbb{P}\left(\lim_{n\to\infty}\{\inf_{t\geq 0} X_t \geq 0\}\right)$. An argument based on the Strong Law of Large numbers shows that, since $\lambda\mu < c$, $\inf_{t\geq 0}\{ct - \sum_{k=1}^{N(t)} Z_k\} > -\infty$ with probability 1 and hence $\lim_{n\to\infty} \inf_{t\geq 0}\{u_n + ct - \sum_{k=1}^{N(t)} Z_k\} = \infty$, hence $\mathbb{P}\left(\lim_{n\to\infty}\{\inf_{t\geq 0} X_t \geq 0\}\right) = 1$. ■

A first step analysis gives

$$\begin{aligned}\Phi(u) &= \mathbb{E}\Phi(u + cS_1 - Z_1) \\ &= \int_0^\infty \lambda e^{-\lambda s} \int_0^{u+cs} \Phi(u + cs - z)dF(z)ds \end{aligned} \tag{4.5}$$

The change of variables $x := u + cs$ transforms the above equation into

$$\Phi(u) = \frac{\lambda}{c} e^{\lambda u/c} \int_u^\infty e^{-\lambda x/c} \int_0^x \Phi(x - z)dF(z)dx. \tag{4.6}$$

Differentiation of the above with respect to u gives

$$\Phi'(u) = \frac{\lambda}{c}\Phi(u) - \frac{\lambda}{c}\int_0^u \Phi(u - z)dF(z). \tag{4.7}$$

Integrating again w.r.t. u from 0 to t we obtain

$$\Phi(t) - \Phi(0) = \frac{\lambda}{c}\int_0^t \Phi(u)du + \frac{\lambda}{c}\int_0^t \int_0^u \Phi(u - z)\left[1 - F(z)\right] dzdu$$

which can be rewritten (after integration by parts) as

$$\Phi(u) = \Phi(0) + \frac{\lambda}{c}\int_0^u \Phi(u - z)\left[1 - F(z)\right] dz. \tag{4.8}$$

4.4 Exponentially Distributed Claims

Suppose that $F(z) = 1 - e^{-z/\mu}$, i.e. the claim distribution is exponential with mean μ. Then (4.7) becomes

$$\begin{aligned}\Phi'(u) &= \frac{\lambda}{c}\Phi(u) - \frac{\lambda}{c\mu}\int_0^u \Phi(u - z)e^{-z/\mu}dz \\ &= \frac{\lambda}{c}\Phi(u) - \frac{\lambda}{c\mu}\int_0^u \Phi(z)e^{-(u-z)/\mu}dz \\ &= \frac{\lambda}{c}\Phi(u) - \frac{\lambda}{c\mu}e^{-u/\mu}\int_0^u \Phi(z)e^{z/\mu}dz.\end{aligned}$$

Differentiating w.r.t. u once more we obtain

$$\Phi''(u) = \left(\frac{\lambda}{c} - \frac{1}{\mu}\right)\Phi'(u) = -\frac{\rho}{\mu(1+\rho)}\Phi'(u),$$

and integrating twice yields

$$\Phi(u) = C_1 + C_2 e^{-\frac{\rho}{\mu(1+\rho)}u}.$$

From the requirement $\lim_{u\to\infty}\Phi(u) = 1$ we see that $C_1 = 1$ whereas from the requirement $\Phi(0) = \frac{\rho}{1+\rho}$ it follows that $C_2 = \frac{1}{1+\rho}$. Thus in the exponential case we have the simple formula $\Phi(u) = 1 - \frac{1}{1+\rho}e^{-\frac{\rho}{\mu(1+\rho)}u}$ and for the ruin probability

$$\boxed{\Psi(u) = \frac{1}{1+\rho}e^{-\frac{\rho}{\mu(1+\rho)}u}.} \tag{4.9}$$

4.5 Simulation of the classic Cramér – Lundberg risk model using R

The following program simulates a risk model with exponentially distributed claims.

```
Thorizon<-100
x<-{}
h<-0.01
rho<-0.02
cp<-1+rho
u<-15
n<-Thorizon/h
y<-u

for (i in 1:n)
{
x<-c(x,y)
if (rbinom(1,1,h)==0) y<-y+h*cp else y<-y+h*cp-rexp(1,1)
}
plot(x,type="l")
```

The simulation horizon is 100 time units. We set the time quantum to $h = 0.01$. In such a small time interval either there is no claim, or there is a single claim, with probablility $1 - \lambda h$ and λh respectively. (λ in the above program is equal to 1 and claims are assumed to be exponential with mean 1 as well.) In view of the properties of the Poisson process the above is a good approximation which becomes exact in the limit as $h \to 0$. (Of course it is not practical to do this in a program, and this is the reason for choosing an appropriately small h compared to λ^{-1}.)

In the next program we simulate this model for 1000 replications and count the number of replications for which ruin occurs. Also, the philosophy of the program is difference. Time is not incremented by a fixed quantum h as before. Instead it is updated each time a new claim arrives. Both the effect of the claim and the income from premiums is taken into account in the statement `y<-y+cp*tinc-claim`. The `while` statement insures that the simulation stops when the time horizon is reached. This provides an exact picture of the process. (Note that ruin can occur only when a claim arrives and therefore nothing is lost by observing the process only at claim arrival epochs.) Also note that a simple counter, `ruincount`, is used to count the number of replications for which ruin occurs. Our estimate for the ruin probability is then given by $\hat{\pi} = \frac{\texttt{ruincount}}{\texttt{replications}}$. The function `Thpr` computes the theoretical *infinite horizon* ruin probability given by (4.9).

```
replications<-1000
```

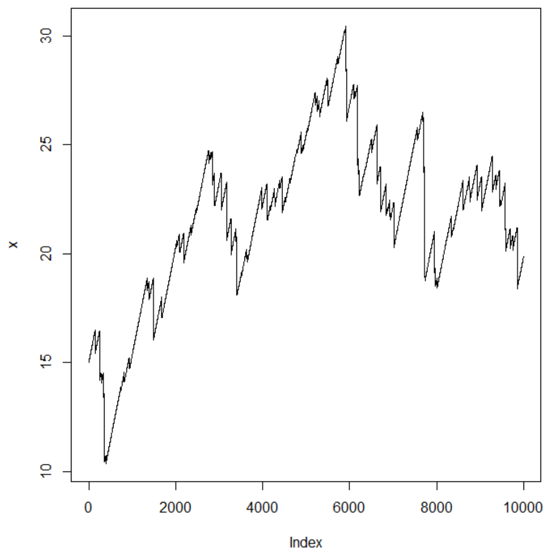

Figure 4.3: Sample path of the Cramér–Lundberg process – R simulation

```
Thorizon<-1000
u<-15
rho<-0.02
cp<-1+rho
ruincount<-0

for (i in 1:replications)
{
y<-u
t<-0
ruin<-0
  while (t < Thorizon)
   {
    tinc<-rexp(1,1)
    claim<-rexp(1,1)
    y<-y+cp*tinc-claim
    t<-t+tinc
    if (y<0) ruin<-1
   }
ruincount<-ruincount+ruin
}
probruin<-ruincount/replications
print(probruin)

Thpr<-function(u) {exp(-u*rho/(1+rho))/(1+rho)}
Thpr(u)
```

The output obtained is

```
> probruin<-ruincount/replications
> print(probruin)
[1] 0.735
>
> Thpr<-function(u) {exp(-u*rho/(1+rho))/(1+rho)}
> Thpr(u)
[1] 0.7305773
```

Thus $\hat{\pi} = 0.735$. We can also obtain an estimate for the standard deviation of the mean (since each trial results in a success or a failure and therefore we deal with Bernoulli random variables) as $\hat{\sigma} = \sqrt{\hat{\pi}(1-\hat{\pi})}/\sqrt{999} = 0.006162408$. Thus the 95% confidence interval for the ruin probability is $0.735 \pm 1.96 \cdot 0.006162408$ or $[0.7229217, 0.7470783]$ which contains the theoretical value 0.7305773. As we mentioned, however, the theoretical value refers to the infinite horizon ruin probability and therefore it is higher than the true (and unknown) theoretical value $\Psi(u, 1000)$.