

Stochastic Models in Risk Theory

Michael Zazanis
Department of Statistics
Athens University of Economics and Business

Summer 2005

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 $a_n = P(X = n)$, $b_n = P(Y = n)$ denotes the distributions of X and Y respectively. The distribution of $Z := X + Y$ is then given by

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

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

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

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

$$\phi(z) := Ez^X = \sum_{n=0}^{\infty} P(X = n)z^n. \quad (1.1)$$

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

$$p_k = \frac{1}{k!} \phi^{(k)}(0), \quad k = 0, 1, 2, \dots, \quad (1.2)$$

and

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

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 $\phi_Z(z) = E[z^{X+Y}] = E[z^X z^Y] = E z^X E z^Y$, 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 $X_i, i = 1, 2, \dots, n$ 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

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

a quantity that, in the general case, is not easy to evaluate. Alternatively, if $p_k = P(X = k)$ then $P(S_n = k) = p_k^{*n} := (p \star \cdots \star p)_k$, 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 $\xi_i, i = 1, 2, \dots, n$ are independent Bernoulli random variables with the same parameter p then their sum $X := \xi_1 + \xi_2 + \dots + \xi_n$ is Binomial with parameters n and p . Its distribution is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n,$$

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, $EX = E[\xi_1 + \dots + \xi_n] = np$ and $\text{Var}(X) = \text{Var}(\xi_1 + \dots + \xi_n) = \text{Var}(\xi_1) + \dots + \text{Var}(\xi_n) = nqp$.

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

1.3.2 The Poisson distribution

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

$$P(X = k) = \frac{1}{k!} \alpha^k e^{-\alpha}, \quad k = 0, 1, 2, \dots$$

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 $EX = \text{Var}(X) = \alpha$.

One of the most important properties of the Poisson distribution is that it arises as the limit of the binomial distribution $\text{Binom}(n, \alpha/n)$ when $n \rightarrow \infty$ (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 $n \rightarrow \infty$. Indeed,

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

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

We also point out that, if X_1, X_2 are independent Poisson random variables with parameters α_1, α_2 respectively, then $X_1 + X_2 \sim \text{Poi}(\alpha_1 + \alpha_2)$. 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

$$P(X = k) = q^{k-1}p, \quad k = 1, 2, 3, \dots, \quad (1.4)$$

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

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

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

$$P(Y = k) = q^k p, \quad k = 0, 1, 2, \dots, \quad (1.6)$$

with corresponding p.g.f.

$$Ez^Y = \frac{1-q}{1-qz}. \quad (1.7)$$

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

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 $\alpha \in \mathbb{R}$:

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

(If α is a positive integer then $\binom{\alpha}{k} = 0$ for all $k = \alpha + 1, \alpha + 2, \dots$ 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

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

Thus we have the identity

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

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

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

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 $EX = \phi'(1) = \alpha q \frac{p^\alpha}{(1-q)^{\alpha+1}}$ or

$$EX = \alpha \frac{q}{p}.$$

Similarly, $EX(X-1) = \phi''(1) = \alpha(\alpha+1)q^2 \frac{p^\alpha}{(1-q)^{\alpha+2}} = \alpha(\alpha+1) \left(\frac{q}{p} \right)^2$. Thus we have

$EX^2 = \alpha(\alpha+1) \left(\frac{q}{p} \right)^2 + \alpha \frac{q}{p}$ and thus $\text{Var}(X) = \alpha(\alpha+1) \left(\frac{q}{p} \right)^2 + \alpha \frac{q}{p} - \left(\alpha \frac{q}{p} \right)^2 = \alpha \frac{q}{p} \left(1 + \frac{q}{p} \right)$ or

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

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

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

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 $\bar{F}(x) := 1 - F(x)$ 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

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

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 $G(x) := P(Y \leq x)$ of Y is given by

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

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 $\text{Var}(Y) = a^2\text{Var}(X)$. Thus if X is “standardized” with mean 0 and standard deviation 1, then Y has mean b and standard deviation a . Also, if $M_X(\theta) = Ee^{\theta X}$ is the moment generating function of X , then the moment generating function of Y is

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

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

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

The distribution function of the standard normal, denoted by

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

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 $EX = 0$. Also, an integration by parts shows that $\text{Var}(X) = 1$. 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

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

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

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

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

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

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

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, $\bar{\Phi}(x) := 1 - \Phi(x)$ are easy to derive. We mention them here for future reference.

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

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

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

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

(remember that $x > 0$).

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

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

■

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

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

2.1.3 The Gamma distribution

The density function is given by

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

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

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

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

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

which can be expressed in terms of the incomplete gamma function defined 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

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

with corresponding distribution function

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

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

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

with corresponding distribution function

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

where $\alpha > 0$. The Pareto distribution is a typical example of a *subexponential distribution*. The n th 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 n th 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

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

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

where $\alpha > 0$.

2.1.5 The Cauchy distribution

The standardized Cauchy density is given by

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

with distribution function

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

It has “fat” polynomial tails: In fact using de l’Hôpital’s rule we see that

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

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 n th 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

$$F \star G(x) = \int_0^x F(x-y) dG(y), \quad x \geq 0. \quad (2.10)$$

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,

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

In the above string of equalities we have used the independence of X and Y to write $P(X \leq x - y | Y = y) = F(x - y)$ 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 $F \star F \star \dots \star F$ (with n factors) with the understanding that $F^{\star 1} = F$ 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

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

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

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

If $X_i, i = 1, 2, \dots, n$ are independent, identically distributed random variables with distribution F and moment generating function $M_X(\theta)$ then $S := X_1 + \dots + X_n$ has distribution function F^{*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 and the uniform distribution. In the case of the exponential distribution, $F(x) = 1 - e^{-\lambda x}, x \geq 0$, 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^{*n} is Gamma($n\alpha, \lambda$).

The uniform distribution is another case in point. Note that the Laplace transform of $x^n, x \geq 0$ is given by $\int_0^\infty x^n e^{-sx} dx = n!s^{n+1}$. Also, the Laplace transform of $(x - \tau)_+^n, x \geq 0, \tau \geq 0$, (where x_+ denotes the positive part of the real number x , i.e. $x_+ = x$ if $x \geq 0$ and $x_+ = 0$ if $x < 0$) is given by

$$\int_0^\infty (x - \tau)_+^n e^{-sx} dx = e^{-\tau s} \int_\tau^\infty (x - \tau)^n e^{-s(x-\tau)} dx = n!e^{-\tau s} s^{n+1}. \quad (2.11)$$

Consider now the sum of n independent random variables, uniformly distributed in $[0, a]$. The Laplace transform of each one of them is given by

$$\frac{1}{a} \int_0^a e^{-sx} dx = \frac{1 - e^{-sa}}{sa}.$$

Thus, the Laplace transform of the n -fold convolution is

$$\frac{(1 - e^{-sa})^n}{(sa)^n} = \sum_{k=0}^n \binom{n}{k} (-1)^k \frac{e^{-ska}}{s^n a^n}.$$

Based on (2.11), we have the identity

$$\frac{e^{-ska}}{s^n a^n} = \int_0^\infty \frac{1}{a(n-1)!} \left(\frac{x}{a} - k\right)_+^{n-1} dx.$$

Thus

$$dF^{*n}(x) = \sum_{k=0}^n \binom{n}{k} (-1)^k \frac{1}{a(n-1)!} \left(\frac{x}{a} - k\right)_+^{n-1} dx$$

and, integrating,

$$F^{*n}(x) = \sum_{k=0}^n (-1)^k \frac{1}{k!(n-k)!} \left(\frac{x}{a} - k\right)_+^n.$$

In, particular, for the uniform distribution in $[0,1]$ the above expression becomes

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

x	$U^{*10}(x)$	$\mathcal{N}(5, 5/6)$
0.0	0.00000000	0.00000002
0.5	0.00000000	0.00000041
1.0	0.00000028	0.00000589
1.5	0.00001589	0.00006305
2.0	0.00027943	0.00050756
2.5	0.00246917	0.00308500
3.0	0.01346285	0.01422982
3.5	0.05045252	0.05017411
4.0	0.13890157	0.13666088
4.5	0.29451868	0.29194118
5.0	0.50000000	0.50000000
5.5	0.70548132	0.70805882
6.0	0.86109843	0.86333912
6.5	0.94954748	0.94982589
7.0	0.98653715	0.98577018
7.5	0.99753083	0.99691500
8.0	0.99972057	0.99949244
8.5	0.99998411	0.99993695
9.0	0.99999972	0.99999411
9.5	1.00000000	0.99999959
10.0	1.00000000	0.99999998

Table 2.1: Comparison of the distribution $U^{*10}(x)$ (sum of 10 independent uniform random variables in $[0,1]$) with the approximating normal which has mean 5 and variance $5/6$. The above table shows that the approximation based on the Central Limit Theorem is quite good in this case.

2.3 Random Sums

Suppose that $X_i, i = 1, 2, \dots$ 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

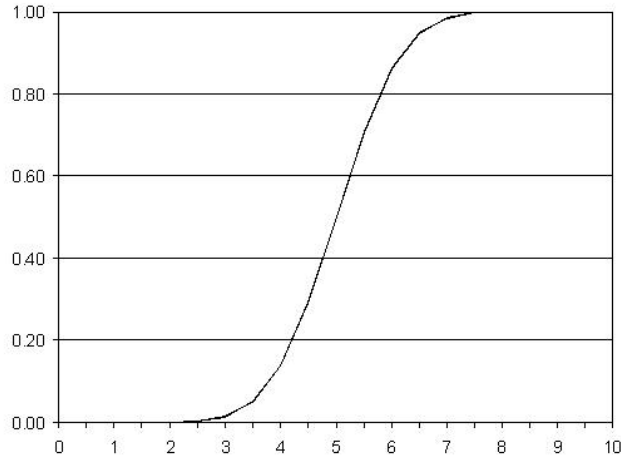


Figure 2.1: Convolution of the uniform distribution, 10 times with itself and the approximating normal. The agreement is too close for the graph to show any difference

also that N is a discrete random variable, independent of the X_i 's, $i = 1, 2, \dots$. 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

$$P(S_N \leq x) = \sum_{n=0}^{\infty} P(N = n)P(X_1 + \dots + X_n \leq x) = \sum_{n=0}^{\infty} P(N = n)F^{*n}(x). \quad (2.12)$$

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

$$ES_N = \sum_{n=0}^{\infty} P(N = n)E[X_1 + \dots + X_n] = \sum_{n=1}^{\infty} P(N = n)nEX_1 = ENEX_1. \quad (2.13)$$

Also

$$E\left(\sum_{i=1}^n X_i\right)^2 = E\left[\sum_{n=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right] = nEX_1^2 + n(n-1)(EX_1)^2$$

and thus

$$\begin{aligned} ES_N^2 &= \sum_{n=0}^{\infty} P(N = n)E[(X_1 + \dots + X_n)^2] = \sum_{n=1}^{\infty} P(N = n) (nEX_1^2 + n(n-1)(EX_1)^2) \\ &= E(X_1^2)EN + (EX_1)^2 \sum_{n=1}^{\infty} n(n-1)P(N = n) \\ &= E(X_1^2)EN + (EX_1)^2 EN^2 - (EX_1)^2 EN \\ &= \text{Var}(X_1)EN + (EX_1)^2 EN^2. \end{aligned} \quad (2.14)$$

From (2.13) and (2.14) we obtain

$$\text{Var}(S_N) = \text{Var}(X_1)EN + \text{Var}(N)(EX_1)^2. \quad (2.15)$$

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

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

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

$$M_{S_N}(\theta) = \phi_N(M_X(\theta)). \quad (2.16)$$

Chapter 3

Simple Models for Premium Calculation

3.1 The rationale for insurance from the point of view of the buyer

The simplest formulation of the problem of insurance from the point of view of the insurance buyer assumes that the buyer, who has a utility function u such that $u' > 0$ and $u'' < 0$ has initial wealth W and faces the possibility of a loss which will be represented by a random variable X with distribution function $F(x) := P(X \leq x)$. (We assume that $F(x) = 0$ for $x < 0$.) He is willing to buy insurance which will cover his losses completely by paying a premium p provided that

$$u(W - p) \geq \int_0^{\infty} u(W - x)dF(x). \quad (3.1)$$

The above equation states that the utility of the buyer if he pays the premium and thus reduces his fortune to $W - p$ is greater than the expected utility if he doesn't. In this latter case, his fortune on the event that $X = x$ is $u(W - x)$ and the right hand side of the above equation represents his expected utility.

It is clear from (3.1) that the insurance buyer would be willing to pay any premium p less than p^* in order to buy full insurance for the risk X , where p^* satisfies the equation

$$u(W - p^*) = \int_0^{\infty} u(W - x)dF(x).$$

It is interesting to note that, in view of our assumptions regarding the utility function u , $p^* > EX$ and thus there exist values of $p > EX$ that would be acceptable to the insurance buyer.

This can be seen by means of Jensen's inequality. If u is a concave function (as is the case here since $u'' < 0$) then for every y ,

$$u(z) < u(y) + (z - y)u'(y).$$

Setting $y = W - p$ and $z = W - x$ we obtain $u(W - x) < u(W - p) + (p - x)u'(W - p)$. Integrating this last equation against $dF(x)$ we obtain

$$\int_0^\infty u(W - x)dF(x) < u(W - p) \int_0^\infty dF(x) + u'(W - p) \int_0^\infty (p - x)dF(x)$$

or

$$\int_0^\infty u(W - x)dF(x) < u(W - p) + (p - EX)u'(W - p).$$

Since $u'(W - p) > 0$ by assumption it is clear that there exist values of p greater than EX and such that

$$\int_0^\infty u(W - x)dF(x) < u(W - p).$$

Thus the insurance buyer would be willing to buy insurance and pay for it a premium higher than the expected size of his loss, EX .

3.2 Safety in numbers

While the decision to buy insurance is based on the buyer's risk aversion, from the point of view of the insurance company a statistical argument based on the law of large numbers applies. Before we explain the situation further we remind the reader of two fundamental results of probability theory.

Theorem 1 (Strong Law of Large Numbers). *If X_i , $i = 1, 2, \dots$, are independent, identically distributed random variables with finite expectation $m = EX_1$, then, as $n \rightarrow \infty$,*

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow m \quad \text{with probability 1.}$$

(The statement above means that $P(\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = m) = 1$.) A less strict law of large numbers is the following

Theorem 2 (Weak Law of Large Numbers). *If X_i , $i = 1, 2, \dots$, are independent, identically distributed random variables with finite expectation $m = EX_1$ then, for any $\epsilon > 0$, as $n \rightarrow \infty$*

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{X_1 + X_2 + \dots + X_n}{n} - m \right| \geq \epsilon \right) = 0.$$

The second fundamental result is the Central Limit theorem which may be formulated as follows.

Theorem 3 (Central Limit Theorem). *Let $\{X_i\}$ be a sequence of independent, identically distributed random variables with common distribution $F(x) := P(X \leq x)$ such that their*

common mean $m := \int x dF(x)$ and variance $\sigma^2 := \int (x - m)^2 dF(x)$ are both finite. Then, as $n \rightarrow \infty$,

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} \leq x\right) \longrightarrow \Phi(x) \quad \forall x \in \mathbb{R} \quad (3.2)$$

where

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

is the standard normal distribution.

Note that the Central Limit Theorem can be seen to describe the statistical fluctuations around the Law of Large Numbers since it can also be stated as follows:

$$P(m - x\sigma\sqrt{n} \leq X_1 + \cdots + X_n \leq m + x\sigma\sqrt{n}) \approx \Phi(x) - \Phi(-x).$$

3.3 The point of view of the insurance company

Suppose then that the insurance company has a portfolio consisting of n policies. Over one period each one of these policies generates a claim, the i th 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 $P(\frac{1}{n} \sum_{i=1}^n X_i > p + u/n)$. If $p > m$ then by virtue of the weak law of large numbers we have that $\lim_{n \rightarrow \infty} P(\frac{1}{n} \sum_{i=1}^n X_i > p + u/n) = 0$.

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.

3.3.1 Net premium computation using 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

$$P(X_1 + X_2 + \cdots + X_n \geq nc) = \alpha.$$

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

$$P\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} \geq \frac{c - m}{\sigma}\sqrt{n}\right) = \alpha. \quad (3.3)$$

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

$$c = m + z_{1-\alpha}\sigma\frac{1}{\sqrt{n}} \quad (3.4)$$

and hence

$$p = \frac{u}{n} + m + z_{1-\alpha}\sigma\frac{1}{\sqrt{n}}. \quad (3.5)$$

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 (3.3). In order to be justified in applying the CLT we should be prepared to let $n \rightarrow \infty$. However, the right hand side of the inequality in (3.3) also goes to infinity with n and therefore the only thing we learn from the CLT is that $P\left(\frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}} > \frac{c - m}{\sigma}\sqrt{n}\right) \rightarrow 0$ as $n \rightarrow \infty$.

Assuming that the CLT approximation can be used,

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

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

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

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

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

We conclude from the above analysis that, if the probability of ruin $P(X_1 + X_2 + \dots + X_n \geq n(c - m))$ 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

$$c = m + \sigma \frac{2 \log(1/\alpha)}{\sqrt{n}}. \quad (3.6)$$

This last equation is to be compared with (3.4).

3.4 Logarithmic Asymptotics

Suppose that $X_i, i = 1, 2, 3, \dots$, 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

$$\lim_{n \rightarrow \infty} P(S_n \geq nx) = 0 \quad \text{for } x > m \quad (3.7)$$

and similarly that

$$\lim_{n \rightarrow \infty} P(S_n \leq nx) = 0 \quad \text{for } x < m \quad (3.8)$$

Correspondingly, if the premium charged per policy, x , is higher than the expected claim size, m , then the probability of ruin goes to zero, whereas if 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

$$P(S_n \geq nx) \asymp e^{-nI(x)} \quad \text{for } x > m \quad (3.9)$$

In the above formula note that the exponential rate of decay $I(x)$ is a function of x . The meaning of (3.9) is made precise if we state it as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq nx) = -I(x) \quad \text{for } x > m. \quad (3.10)$$

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

$$P(S_n - nm \geq n(x - m)) = P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \geq \sqrt{n} \left(\frac{x - m}{\sigma}\right)\right) \quad (3.11)$$

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

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

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 (3.11) becomes an exact relationship and we have

$$P(S_n \geq nx) = \int_{\sqrt{n}\left(\frac{x-m}{\sigma}\right)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du. \quad (3.12)$$

Taking into account the bounds in proposition 1 we have

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

or

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

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq nx) = -\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2. \quad (3.13)$$

Hence, setting $I(x) = \frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2$ we obtain (3.7) 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.

3.5 Chernoff bounds

In the same framework as before $X_i, i = 1, 2, \dots$ 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

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

The above inequality provides an upper bound for $P(S_n \geq nx)$ 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

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

Define now the *rate function*

$$I(x) := \sup_{\theta \in \mathbb{R}} \{x\theta - \log M(\theta)\}. \quad (3.14)$$

With this definition the Chernoff bound becomes

$$P(S_n \geq nx) \leq e^{-nI(x)}. \quad (3.15)$$

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 4. *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

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

and hence

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

Similarly

$$\frac{M'(\theta)}{M(\theta)} = E\left[X \frac{e^{\theta X}}{M(\theta)}\right] = E_{\tilde{P}}[X]$$

and thus

$$\frac{d^2}{d\theta^2} \log M(\theta) = E_{\tilde{P}}[X^2] - (E_{\tilde{P}}[X])^2 = E_{\tilde{P}}(X - E_{\tilde{P}}[X])^2 \geq 0.$$

■

3.6 Examples of rate functions

Bernoulli Random Variables Suppose that $P(X_i = 1) = 1 - P(X_i = 0) = p$ (i.e. the random variables take only the values 0 and 1 with probabilities $1 - p$ and p respectively).

In this case $\log M(\theta) = \log (pe^\theta + 1 - p)$. 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

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

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

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

Geometric (probability of success p) Here

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

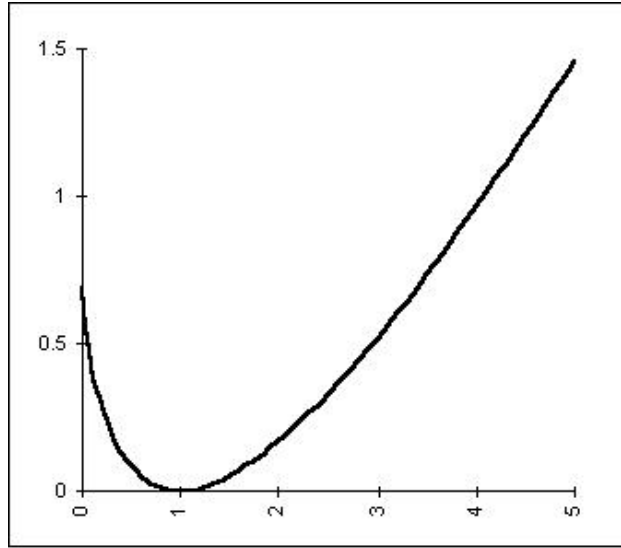


Figure 3.1: Rate function for the geometric distribution

Following the same procedure as before we obtain

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

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

3.7 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,

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

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: $M(\theta) = Ee^{\theta X} \geq e^{\theta EX}$ 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

$$\frac{M'(\theta^*)}{M(\theta^*)} = x \quad (3.16)$$

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 $x - M'(\theta^*)/M(\theta^*) = 0$.

3.8 The twisted distribution

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

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

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

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

is a non-decreasing function of y and as $y \rightarrow \infty$, $\tilde{F}(y) \rightarrow 1$.

The mean of the twisted distribution is given by

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

In particular when $\theta = \theta^*$, the solution of (3.16),

$$\frac{1}{M(\theta^*)} \int_{-\infty}^{\infty} y e^{\theta^* y} F(dy) = \frac{M'(\theta^*)}{M(\theta^*)} = x. \quad (3.17)$$

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

3.9 Cramér's Theorem

Theorem 5. 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 $m = EX_1$ and $S_n := X_1 + \cdots + X_n$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq nx) = -I(x), \quad x \geq m.$$

Proof: Note that, in view of the discussion regarding Chernoff's theorem, $\frac{1}{n} \log P(S_n \geq nx) \leq -I(x)$ whence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq nx) \leq -I(x), \quad x \geq m.$$

Thus, in order to establish the theorem it suffices to show the corresponding inequality for the lim inf, i.e.

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq nx) \geq -I(x) \quad \text{for } x > m. \quad (3.18)$$

Start with the inequality

$$\begin{aligned} P(S_n \geq nx) &\geq P(n(x + \epsilon) \geq S_n \geq nx) = E[\mathbf{1}(n(x + \epsilon) \geq S_n \geq nx)] \\ P(S_n \geq nx) &\geq e^{-n(x+\epsilon)} M(\theta^*)^n E \left[\frac{e^{\theta^* S_n}}{M(\theta^*)^n} \mathbf{1}(n(x + \epsilon) \geq S_n \geq nx) \right] \\ &= e^{-n(x+\epsilon)} M(\theta^*)^n \tilde{P}(n(x + \epsilon) \geq S_n \geq nx) \end{aligned} \quad (3.19)$$

The twisted distribution can be used to establish (3.18) as follows: Set $\theta = \theta^*$ so that, under \tilde{P} , the mean of X_i is x . Also

$$\tilde{P}(n(x + \epsilon) \geq S_n \geq nx) = \tilde{P} \left(\sqrt{n}\epsilon \geq \frac{S_n - nx}{\sqrt{n}} \geq 0 \right) \quad (3.20)$$

and since $\tilde{E}X = x$, we can appeal to the Central Limit Theorem to conclude that

$$\lim_{n \rightarrow \infty} \tilde{P} \left(\sqrt{n}\epsilon \geq \frac{S_n - nx}{\sqrt{n}} \geq 0 \right) = \frac{1}{2}$$

or equivalently

$$\lim_{n \rightarrow \infty} \tilde{P}(n(x + \epsilon) \geq S_n \geq nx) = \frac{1}{2}. \quad (3.21)$$

Taking logarithms in (3.19) we have

$$\log P(S_n \geq nx) \geq -n(x + \epsilon)\theta^* + n \log M(\theta^*) + \log \tilde{P}(n(x + \epsilon) \geq S_n \geq nx)$$

from which we obtain

$$\begin{aligned} \liminf_n \frac{1}{n} \log P(S_n \geq nx) &\geq -(x + \epsilon)\theta^* + \log M(\theta^*) \\ &+ \liminf_n \frac{1}{n} \log \tilde{P}(n(x + \epsilon) \geq S_n \geq nx). \end{aligned} \quad (3.22)$$

In view of (3.21) and the fact that ϵ was arbitrary we obtain

$$\liminf_n \frac{1}{n} \log P(S_n \geq nx) \geq -(x\theta^* - \log M(\theta^*)) = -I(x),$$

the last equality following from the fact that θ^* is the value that maximizes the quantity $x\theta - \log M(\theta)$. ■

Chapter 4

Large claim risk theory

4.1 Subexponential Distributions

Definition 1. (Subexponential Distribution Functions) A distribution function F with support $(0, \infty)$ is subexponential if, for all $n \geq 2$,

$$\lim_{x \uparrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n. \quad (4.1)$$

The class of subexponential distribution functions will be denoted by \mathcal{S} .

The class of subexponential distributions includes all distributions with polynomial (i.e. “fat”) tails.

Let X_i , $i = 1, 2, \dots, n$, be i.i.d. random variables with distribution F . Denote by $S_n := X_1 + \dots + X_n$ their sum and by $M_n := \max(X_1, X_2, \dots, X_n)$. Then

$$\begin{aligned} P(S_n > x) &= \overline{F^{*n}}(x) \\ P(M_n > x) &= \overline{F^n}(x) = 1 - F^n(x) \\ &= (1 - F(x))(1 + F(x) + \dots + F^{n-1}(x)) \\ &= \overline{F}(x) \sum_{k=0}^{n-1} F^k(x). \end{aligned}$$

Hence,

$$\lim_{x \rightarrow \infty} \frac{P(M_n > x)}{\overline{F}(x)} = \lim_{x \rightarrow \infty} \sum_{k=0}^{n-1} F^k(x) = n. \quad (4.2)$$

From (4.1) we see that $P(S_n > x) \sim n\overline{F}(x)$ while from (4.2) it follows that $P(M_n > x) \sim n\overline{F}(x)$. Thus subexponentiality implies that

$$P(S_n > x) \sim P(M_n > x), \quad x \rightarrow \infty.$$

which means that the tail of the sum and the tail of the maximum are the same, i.e. that *very large values of the sum are mainly due to a single large contribution*.

This behavior does not occur with distributions that have exponential tails.

Example: Let $F(x) = 1 - e^{-\lambda x}$. Then

$$P(M_n > x) = \overline{F}(x) \sum_{k=0}^{n-1} F^k(x) \sim ne^{-\lambda x}, \quad \text{as } x \rightarrow \infty.$$

On the other hand

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

Hence in that case

$$\lim_{x \rightarrow \infty} \frac{P(M_n > x)}{P(S_n > x)} = 0.$$

Lemma 1. *If $F \in \mathcal{S}$ then, given $\epsilon > 0$, there exists $K < \infty$ such that, for all $n \geq 2$,*

$$\frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \leq K(1 + \epsilon)^n, \quad x \geq 0. \quad (4.3)$$

Proof: We start with the following simple identity:

$$\frac{\overline{F^{*(n+1)}}(x)}{\overline{F}(x)} = 1 + \frac{F(x) - F^{(n+1)*}(x)}{\overline{F}(x)} = 1 + \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y). \quad (4.4)$$

Let

$$\alpha_n = \sup_{x \geq 0} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)}. \quad (4.5)$$

Then

$$\begin{aligned} \alpha_{n+1} &= \sup_{x \geq 0} \frac{\overline{F^{*(n+1)}}(x)}{\overline{F}(x)} = 1 + \sup_{x \geq 0} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y) \\ &= 1 + \sup_{0 \leq x \leq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y) + \sup_{x \geq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y) \\ &= 1 + \sup_{0 \leq x \leq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y) + \sup_{x \geq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x-y)} \frac{\overline{F}(x-y)}{\overline{F}(x)} dF(y) \\ &= 1 + A_T + \alpha_n \sup_{x \geq T} \frac{F(x) - F^{*2}(x)}{\overline{F}(x)}, \end{aligned}$$

where $A_T = \frac{1}{\overline{F}(T)} < \infty$. Also,

$$\frac{F(x) - F^{*2}(x)}{\overline{F}(x)} = \frac{-\overline{F}(x) + \overline{F^{*2}}(x)}{\overline{F}(x)} = \frac{\overline{F^{*2}}(x)}{\overline{F}(x)} - 1 \rightarrow 2 - 1 = 1 \quad \text{as } x \rightarrow \infty$$

since $F \in \mathcal{S}$. Hence we can choose T large enough so that

$$\sup_{x \geq T} \frac{F(x) - F^{*2}(x)}{\overline{F}(x)} \leq (1 + \epsilon).$$

Thus $\alpha_{n+1} \leq 1 + A_T + \alpha_n(1 + \epsilon)$, hence

$$\alpha_n \leq (1 + A_T)\epsilon^{-1}(1 + \epsilon)^n.$$

■

4.2 An asymptotic expression for the ruin probability in the subexponential case

Consider now the random sum $Y = \sum_{i=1}^N X_i$ where X_i are i.i.d. subexponential random variable with distribution function F and N an independent random variable with values in \mathbb{N} and distribution $P(N = n) = p_n$, $n = 0, 1, 2, \dots$. We will assume that $EN = m < \infty$. Then the following proposition holds

Proposition 2. *If $\Psi(u) = P(Y > u)$ then*

$$\Psi(u) \sim m\overline{F}(u). \quad (4.6)$$

Proof: Note that

$$\frac{\Psi(u)}{\overline{F}(u)} = \sum_{k=0}^{\infty} p_k \frac{\overline{F^{*k}}(u)}{\overline{F}(u)}. \quad (4.7)$$

Letting u go to infinity and using Lemma 3 to allow us to pass the limit inside the infinite sum we obtain

$$\lim_{u \rightarrow \infty} \frac{\Psi(u)}{\overline{F}(u)} = \sum_{k=0}^{\infty} p_k \lim_{u \rightarrow \infty} \frac{\overline{F^{*k}}(u)}{\overline{F}(u)} = \sum_{k=0}^{\infty} p_k k = m.$$

■

Chapter 5

Exact and approximate computation of the loss distribution in a Portfolio

5.1 The compound Poisson distribution for integer-valued claims

Consider $S = X_1 + X_2 + \cdots + X_N$ where N is a Poisson random variable with mean $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_m$ and X_i are independent identically distributed, integer-valued random variables with distribution given by $P(X_i = j) = p_j$, for $j = 1, 2, \dots, n$ (note that the distributions are assumed to have finite support). Then

$$P(S = n) = \sum_{k=0}^{\infty} P^{*k}(n) e^{-\alpha} \frac{\alpha^k}{k!} \quad (5.1)$$

where $P^{*0}(n) = \mathbf{1}(n = 0)$, $P^{*1}(n) = p_n$, and

$$P^{*k}(n) = \sum_{l=0}^{n-k} P^{*(k-1)}(n-l) p_l, \quad k = 2, 3, \dots, \quad n = 0, 1, \dots, nm.$$

The computation of the above convolutions is of course far from simple in general. Fortunately, in the discrete case there exists a rapid recursive way for this.

Let $f(n) := P(S = n)$. Then

$$\begin{aligned} f(0) &= P(S = 0) = P(N = 0) = e^{-\alpha}, \\ f(n) &= \frac{1}{n} \sum_{j=1}^n j \alpha p_j f(n-j), \quad n = 1, 2, 3, \dots \end{aligned} \quad (5.2)$$

To derive this recursive relationship note that

$$g(z) = \sum_{n=0}^{\infty} f(n) z^n \quad (5.3)$$

and

$$\log g(z) = \sum_{j=1}^m \alpha p_j (z^j - 1). \quad (5.4)$$

Since $\frac{d}{dz}g(z) = g(z)\frac{d}{dz}\log g(z)$,

$$\sum_{n=1}^{\infty} n f(n) z^{n-1} = \left(\sum_{i=0}^{\infty} f(i) z^i \right) \left(\sum_{j=1}^m j \alpha p_j z^{j-1} \right). \quad (5.5)$$

Comparing coefficients we have $n f(n) = \sum_{j=1}^m f(n-j) j \alpha p_j$.

5.2 The normal approximation for random sums

Here we consider the random sum $S := \sum_{n=1}^N X_i$ where N is independent of the X_i that are assumed to be i.i.d. with finite second moment. We denote the mean of X_1 by m and its variance by σ^2 and otherwise make no other assumptions regarding the statistics of X or N . The normal approximation then suggests that the random variable $\frac{S-ES}{\sqrt{\text{Var}(S)}}$ is approximately normally distributed. Of course we have seen that $ES = mEN$ and $\text{Var}(S) = m^2\text{Var}(N) + \sigma^2EN$ hence, if $G(x) := P(S \leq x)$,

$$G(x) \approx \Phi \left(\frac{x - mEN}{\sqrt{m^2\text{Var}(N) + \sigma^2EN}} \right). \quad (5.6)$$

The above relationship is of course not exact. To obtain a bound for the error we start with the Berry-Esseen bound for the error in the central limit theorem:

Theorem 6 (Berry-Esseen). *Suppose that $\{X_i\}$ are i.i.d. with zero mean, variance $\sigma^2 = EX_1^2$, and finite third moment, $\gamma^3 = EX_1^3$. If $S_n = X_1 + \dots + X_n$, then*

$$\left| P(S_n/\sigma\sqrt{n} \leq x) - \Phi(x) \right| \leq \frac{c_1}{\sqrt{n}} \left(\frac{\gamma}{\sigma} \right)^3 \quad \text{for all } x \in \mathbb{R}.$$

The value of the constant c_1 is less than 1.

The above theorem can be used to obtain a bound for the normal approximation error. Then, since from the Berry-Esseen bound

$$-\frac{c_1}{\sqrt{n}} \left(\frac{\gamma}{\sigma} \right)^3 \leq P(S_n/\sigma\sqrt{n} \leq x) - \Phi(x) \leq \frac{c_1}{\sqrt{n}} \left(\frac{\gamma}{\sigma} \right)^3 \quad (5.7)$$

The above is valid for $n = 1, 2, \dots$, while for $n = 0$, regardless of how S_n/\sqrt{n} is defined, $|P(S_n/\sigma\sqrt{n} \leq x) - \Phi(x)| \leq 1$. (We could agree for instance to set $S_N/\sqrt{N} = 0$ when

$N = 0$.) Then from (5.7) we have

$$\begin{aligned}
\left| P\left(S_N/\sigma\sqrt{N} \leq x\right) - \Phi(x) \right| &= (\mathbf{1}(x \geq 0) - \Phi(x))P(N = 0) \\
&\quad + \sum_{n=1}^{\infty} P(N = n) \left| P\left(S_n/\sigma\sqrt{n} \leq x\right) - \Phi(x) \right| \\
&\leq P(N = 0) + c_1 \left(\frac{\gamma}{\sigma}\right)^3 \sum_{n=1}^{\infty} P(N = n) \frac{1}{\sqrt{n}} \\
&= P(N = 0) + c_1 P(N > 0) \left(\frac{\gamma}{\sigma}\right)^3 E[N^{-1/2} \mid N > 0].
\end{aligned}$$

We will need the following

Lemma 2. *For any random variable M with values on the positive integers, $1, 2, 3, \dots$ we have*

$$(EM)^{-1/2} \leq EM^{-1/2} \leq (EM^{-1})^{1/2}. \quad (5.8)$$

Proof: We have $(EM^{-1/2})^2 \leq E(M^{-1/2})^2 = EM^{-1}$ (from the Cauchy-Schwarz inequality) or

$$EM^{-1/2} \leq (EM^{-1})^{1/2}. \quad (5.9)$$

On the other hand, from Jensen's inequality we have

$$EM^{-1/2} \geq (EM)^{-1/2}.$$

This last equation, together with (5.9) establish the proof of the lemma. ■

Let us now consider the case when N is Poisson with mean α .

Lemma 3. *Let N_α be a Poisson random variable with mean α . Then $E[N_\alpha^{-1/2} \mid N_\alpha > 0] \asymp \alpha^{-1/2}$ as $\alpha \rightarrow \infty$.*

Proof: We have $P(N_\alpha > 0) = 1 - e^{-\alpha}$ and

$$E[N_\alpha \mid N_\alpha > 0] = (1 - e^{-\alpha})^{-1} \sum_{n=1}^{\infty} n \frac{\alpha^n}{n!} e^{-\alpha} = \alpha(1 - e^{-\alpha}).$$

Then we can use inequality (5.8) on the random variable N_α , conditioned on the event $N_\alpha > 0$ to obtain

$$E[N_\alpha^{-1/2} \mid N_\alpha > 0] \leq (E[N_\alpha^{-1} \mid N_\alpha > 0])^{1/2}.$$

In order to estimate the size of

$$E[N_\alpha^{-1} \mid N_\alpha > 0] = \frac{e^{-\alpha}}{1 - e^{-\alpha}} \sum_{n=1}^{\infty} \frac{1}{n} \frac{\alpha^n}{n!}$$

we set

$$g(\alpha) := \sum_{n=1}^{\infty} \frac{1}{n} \frac{\alpha^n}{n!}$$

and note that $g'(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha^{n-1}}{n!}$ whence we obtain that $g'(\alpha) = \alpha^{-1}(e^\alpha - 1)$. Thus

$$g(\alpha) = \int_0^\alpha (e^x - 1)x^{-1} dx$$

and

$$E[N^{-1}|N > 0] = \frac{1}{e^\alpha - 1} \int_0^\alpha \frac{e^x - 1}{x} dx$$

by virtue of lemma 2 we have

$$\frac{1}{\sqrt{\alpha(1 - e^{-\alpha})}} \leq E[N^{-1/2}|N > 0] \leq \frac{1}{\sqrt{\alpha}} \left(\frac{\alpha}{e^\alpha - 1} \int_0^\alpha \frac{e^x - 1}{x} dx \right)^{1/2}.$$

Since

$$\lim_{\alpha \rightarrow \infty} \frac{\alpha}{e^\alpha - 1} \int_0^\alpha \frac{e^x - 1}{x} dx = 1$$

(e.g. use de L'Hôpital's rule) we conclude that

$$E[N_\alpha^{-1/2}|N_\alpha > 0] \asymp \frac{1}{\sqrt{\alpha}}.$$

■

We can thus state the following

Theorem 7. *For the compound Poisson model we have the asymptotic relationship*

$$\left| P\left(S_{N_\alpha}/\sigma\sqrt{N_\alpha} \leq x\right) - \Phi(x) \right| \asymp \frac{c_1}{\sqrt{\alpha}} \left(\frac{\gamma}{\sigma}\right)^3.$$

5.3 The translated Gamma approximation

The idea here is to approximate $S := \sum_{i=1}^N X_i$ by means of a *translated Gamma* distribution which has density

$$f(x) = \frac{\beta^\alpha (x - k)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta(x-k)}; \quad x \geq k, \quad (5.10)$$

where, as usual, $\beta > 0$, $\alpha > 0$. In particular, if Z has distribution $\text{Gamma}(\alpha, \beta)$ then $X := k + Z$ has the translated Gamma distribution with density given by (5.10). Let us denote by $m := ES$ and $\sigma^2 := \text{Var}(S)$ the mean and variance of the random sum S and by

$$\gamma := \frac{E[S^3]}{(\text{Var}(S))^{3/2}}$$

the *skewness coefficient*. On the other hand, the mean, variance, and skewness coefficient for the translated normal can be obtained easily as follows: Clearly, $EX = k + EZ$ while $\text{Var}(X) = \text{Var}(Z)$ and similarly, the skewness coefficient of X and Z are the same. Since Z follows the Gamma(α, β) distribution we have $EZ = \alpha/\beta$ and $\text{Var}(Z) = \alpha/\beta^2$. Also, $EZ^3 = \alpha(\alpha + 1)(\alpha + 2)/\beta^3$ and thus

$$\begin{aligned} E[(Z - EZ)^3] &= E[Z^3 - 3Z^2EZ + 3Z(EZ)^2 - (EZ)^3] \\ &= \frac{\alpha(\alpha + 1)(\alpha + 2)}{\beta^3} - 3\frac{\alpha(\alpha + 1)}{\beta^2}\frac{\alpha}{\beta} + 3\frac{\alpha^3}{\beta^3} - \frac{\alpha^3}{\beta^3} \\ &= \frac{2\alpha}{\beta^3}. \end{aligned}$$

Thus

$$\text{Skewness coeff. of } X = \text{Skewness coeff. of } Z = \frac{2}{\sqrt{\alpha}}.$$

The unknown parameters of the translated Gamma approximation are then determined from the known m, σ, γ by solving the system

$$\begin{aligned} m &= k + \frac{a}{\beta} \\ \sigma^2 &= \frac{a}{\beta^2} \\ \gamma &= \frac{2}{\sqrt{\alpha}}. \end{aligned}$$

Bibliography

- [1] Asmussen, S. (1987). *Applied Probability and Queues*, John Wiley.
- [2] Beirlant, J., J.L. Teugels, and P. Vynckier. (1996). *Practical Analysis of Extreme Values*, Leuven University Press.
- [3] Bühlman, H. (1971). *Mathematical Methods in Risk Theory*, Springer-Verlag.
- [4] Feller, W. (1971). *Introduction to Probability Theory, Vol. 2*, John Wiley.
- [5] Grandel, J. (1991). *Aspects of Risk Theory*, Springer-Verlag.