Topics in Stochastic Processes

Michael Zazanis
Department of Statistics
Athens University of Economics and Business

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

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

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

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

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

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

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 $\phi_Z(z) = E[z^{X+Y}] = E[z^X z^Y] = Ez^X Ez^Y$, 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

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

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

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

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, $EX = E[\xi_1 + \cdots + \xi_n] = np$ 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

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

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

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

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

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

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

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

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

with corresponding p.g.f.

$$Ez^Y = \frac{1-q}{1-qz}. \tag{1.7}$$

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

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

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

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

Similarly, $EX(X-1) = \phi''2 \frac{p^\alpha}{(1-q)^{\alpha+2}} = \alpha(\alpha+1) \left(\frac{q}{p}\right)^2$. Thus we have $EX^2 = \alpha(\alpha+1) \left(\frac{q}{p}\right)^2 + \alpha \frac{q}{p}$ and thus $\mathsf{Var}(X) = \alpha(\alpha+1) \left(\frac{q}{p}\right)^2 + \alpha \frac{q}{p} - \left(\alpha \frac{q}{p}\right)^2 = \alpha \frac{q}{p} \left(1 + \frac{q}{p}\right)$ 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 $F(x) := P(X \leq x)$, $x \in \mathbb{R}$. 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 $P(a < X \leq b) = F(b) - F(a)$. 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

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

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

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

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

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

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 $EX = 0$. 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

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

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

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

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 $\theta \ge \beta$ 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

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

with corresponding distribution function

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

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

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

with corresponding density

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

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

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

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,

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

In the above string of equalities we have used the independence of X and Y to write $P(X \le x - y | Y = y) = F(x-y)$ 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 $x \ge 0$ 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

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

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

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

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

$$ES_N = \sum_{n=0}^\infty P(N=n)E[X_1 + \cdots + X_n] = \sum_{n=1}^\infty P(N=n)nEX_1 = ENEX_1. \quad (2.12)$$

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

$$\begin{aligned} ES_N^2 &= \sum_{n=0}^{\infty} P(N=n)E[(X_1+\cdots+X_n)^2] = \sum_{n=1}^{\infty} P(N=n)\left(nEX_1^2 + n(n-1)(EX_1)^2\right) \\ &= E(X_1^2)EN + (EX_1)^2 \sum_{n=1}^{\infty} n(n-1)P(N=n) \\ &= E(X_1^2)EN + (EX_1)^2 EN^2 - (EX_1)^2 EN \\ &= \mathsf{Var}(\mathsf{X}_1)EN + (EX_1)^2 EN^2. \end{aligned} \tag{2.13}$$

From (2.12) and (2.13) we obtain

$$\mathsf{Var}(S_N) = \mathsf{Var}(X_1)EN + \mathsf{Var}(N)(EX_1)^2. \tag{2.14}$$

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

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

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

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

Chapter 3

Stochastic Processes

3.1 Brownian Motion and the Poisson Process

A stochastic process $\{X_t; t \in \mathbb{T}\}$ is a collection of random variables (assumed real here) indexed by a set $\mathbb{T}$ which, in our case is either the natural numbers $\mathbb{N}$ or the real numbers $\mathbb{R}$. In this section we will focus on processes defined in continuous time so $\mathbb{T} = \mathbb{R}$.

To define a stochastic process it suffices to define a system of probability distributions, $\mu_n(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$ which satisfy the Kolmogorov consistency conditions namely that

$$\lim_{x_n \to \infty} \mu_n(x_1, x_2, \ldots, x_{n-1}, x_n; t_1, t_2, \ldots, t_{n-1}, t_n) = \mu_{n-1}(x_1, x_2, \ldots, x_{n-1}; t_1, t_2, \ldots, t_{n-1}).$$

If these conditions are satisfied then Kolmogorov's theorem states that there exists a stochastic process $\{X_t; t \in \mathbb{R}\}$ such that

$$\mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = \mu_n(x_1, \ldots, x_n; t_1, t_2, \ldots, t_n), \quad \text{for all } x_i, t_i \in \mathbb{R}, \;\; n \in \mathbb{N}.$$

3.2 Brownian Motion

This is the most important continuous time stochastic process. It can be thought of as a building block for a variety of other, more complicated stochastic processes. One way of defining it is the following.

Definition 1. *A stochastic process $\{W_t, t \geq 0\}$ is called Standard Brownian Motion if it satisfies the following three postulates*

i) $P(W_0 = 0) = 1$, i.e. the process starts with probability 1 from 0 at time 0.

ii) The increments are independent i.e. if $0 \le t_i < t_2 < \cdots < t_k$ then $P(W_{t_i} - W_{t_{i-1}} \in H_i; i = 1, 2, \ldots, k) = \prod_{i=1}^{k} P(W_{t_i} - W_{t_{i-1}} \in H_i)$ for any (Borel) subsets H_i of $\mathbf{R}$.

iii) For $0 \le s < t$, $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$:

$$P(W_t - W_s \in H) = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{-x^2/2(t-s)} dx$$

From the above postulates it follows that the finite dimensional distributions of the process W_t are given by

$$P\left(W_{t_1} \in (x_1, x_1 + dx_1), \ldots, W_{t_n} \in (x_n, x_n + dx_n)\right) = f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) dx_1 \cdots dx_n$$

with

$$\begin{aligned}
f(x_1, \ldots, x_n; t_1, t_2, \ldots, t_n) & \\
&= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{t_1(t_2 - t_1)\cdots(t_n - t_{n-1})}} e^{-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right\}} \\
&= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2} x^T Q^{-1} x}
\end{aligned}$$

where x^T denotes the transpose of $x = (x_1, \ldots, x_n)$ and

$$\begin{aligned}
Q &= E \begin{pmatrix} W_{t_1} \\ \vdots \\ W_{t_n} \end{pmatrix} (W_{t_1}, \ldots, W_{t_n}) = \begin{bmatrix} EW_{t_1}W_{t_1} & \cdots & EW_{t_1}W_{t_n} \\ \vdots & EW_{t_i}W_{t_j} & \vdots \\ EW_{t_n}W_{t_1} & \cdots & EW_{t_n}W_{t_n} \end{bmatrix} \\
&= [t_i \wedge t_j]_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}
\end{aligned}$$

is the corresponding covariance matrix, i.e. the finite dimensional distributions of brownian motion are normal. This means that brownian motion is a Gaussian process.

3.2.1 Properties of Standard Brownian Motion

1. Markov Property. Brownian motion is a Markov process with stationary transition probabilities

$$\begin{aligned}
P_t(x, A) &= P(W_{t+s} \in A | W_s = x) = P(W_{t+s} - W_s \in A - x | W_s = x) \\
&= P(W_t \in A - x) = \int_{A-x} \phi(u) du
\end{aligned}$$

where $A - x$ is the set $\{y - x : y \in A\}$ and $\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$.

2. <u>Scaling Property.</u> $\forall c > 0$ $\{\sqrt{c}W_{t/c}; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$.

   Indeed, $\sqrt{c}W_{t/c}$ has continuous paths, stationary and independent increments, and the correct distribution.

3. <u>Symmetry.</u> $\{-W_t; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$.

4. <u>Time reversal.</u> $\{tW_{1/t}; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$.

3.3 Maximum of the Standard Brownian Motion

Let $M_t = \max\{W_u; 0 \leq u \leq t\}$, where as usual $\{W_t; t \geq 0\}$ is SBM (Standard Brownian Motion). For fixed t this is a nonnegative random variable, while, if we consider the process $\{M_t; t \geq 0\}$ then we have a process with $M_0 = 0$ and nondecreasing sample paths. Here we will compute the distribution of the random variable M_t using *the reflection principle.* Define $\tau := \inf\{u \geq 0 : W_u = a\}$ the first time when the brownian motion reaches the level a (note that τ is a stopping time). Now define new process, $\widehat{W}_u$, via the relationship

$$\widehat{W}_u = \begin{cases} W_u & u < \tau \\ a - (W_u - a) & u \geq \tau \end{cases}.$$

In the figure below, $\widehat{W}$ is the process that is identical with W up to time τ, and after that time results from the reflection of W around the level a (the grey path in the figure). Note also that $\{M_t \geq a\} = \{\tau \leq t\}$ (i.e. the two events are the same). Now

$$P(M_t \geq a) = P(M_t \geq a, W_t \geq a) + P(M_t \geq a, W_t < a)$$

and $\{M_t \geq a\} \subset \{W_t \geq a\}$, so that $P(M_t \geq a, W_t \geq a) = P(W_t \geq a)$. Also,

$$\begin{aligned} P(M_t \geq a, W_t < a) &= P(\tau \leq t, W_t < a) = P(\tau \leq t)P(W_t < a \mid \tau \leq t) \\ &= P(M_t \geq a)P(\widehat{W}_t < a \mid \tau \leq t) = \frac{1}{2}P(M_t \geq a) \end{aligned}$$

From the above it follows that

$$P(M_t \geq a) = 2P(W_t \geq a) = \int_a^\infty \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx, \qquad a \geq 0. \tag{3.1}$$

Equivalently, we can say that $M_t \stackrel{\mathrm{d}}{=} |W_t|$. This same formula gives the distribution of the time required for SBM to reach a given level $x > 0$: If we denote by $\tau_x := \inf\{t \geq 0 : W_t = x\}$ we have

$$P(\tau_x \leq t) = P(M_t \geq x) = \int_x^\infty \sqrt{\frac{2}{\pi t}} e^{-y^2/2t} dy$$

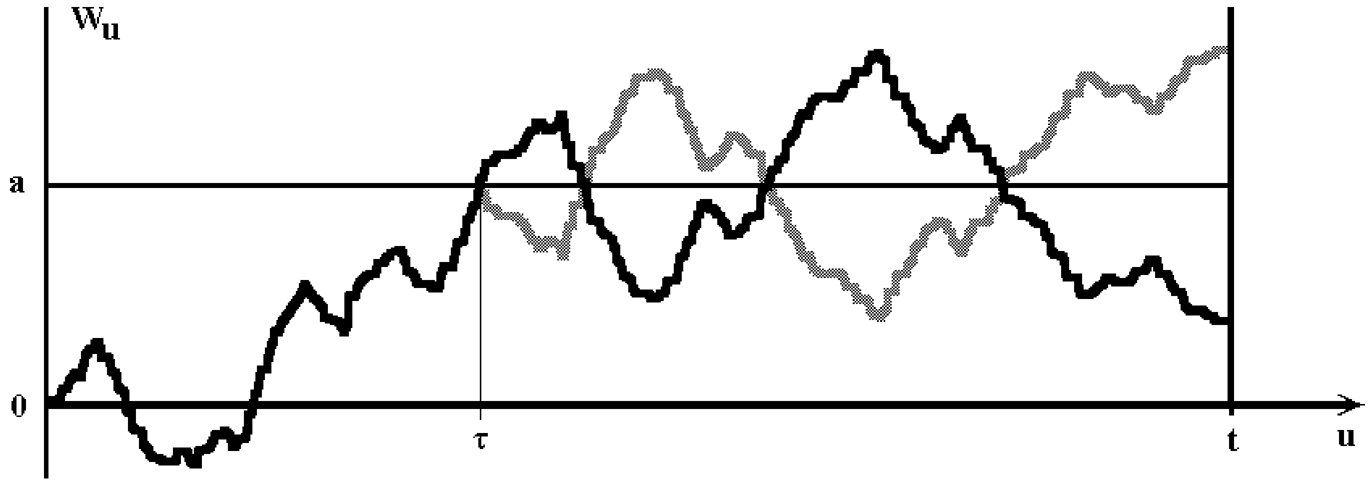

Figure 3.1: The reflection principle

The change of variables $y = z\sqrt{t}$ transforms the above equation into

$$P(\tau_x \le t) = \int_{x/\sqrt{t}}^{\infty} \sqrt{\frac{2}{\pi}} e^{-z^2/2} dz$$

and differentiating with respect to t we obtain the density function of τ_x:

$$f_{\tau_x}(t) = \frac{1}{\sqrt{2\pi t^3}} x e^{-x^2/2t}, \qquad t \ge 0.$$

(This is the inverse Gaussian distribution.)

3.4 Martingales associated with Brownian Motion

It is easy to see that standard brownian motion is a martingale. If we denote by $\mathcal{F}_s := \sigma\{W_u; 0 \le u \le s\}$ the history of the process up to time s then

$$E[W_t|\mathcal{F}_s] = W_s + E[W_t - W_s|\mathcal{F}_s] = W_s$$

the second term in the above equation vanishing as a result of the independent increments property.

This property, together with the *optional stopping theorem* allows us to compute probabilities of reaching boundaries. Suppose that $W_0 = x$ and let $a < x < b$. Set $\tau = \inf\{t \ge 0 : W_t = a \text{ or } b\}$. Then, by the optional stopping theorem we have

$$EW_\tau = EW_0 = x.$$

However $W_\tau = a\mathbf{1}(W_\tau = a) + b\mathbf{1}(W_\tau = b$, and if we denote by $p_a = P(W_\tau = a)$ (and similarly for p_b) we have $ap_a + bp_b = x$ which gives (since $p_a + p_b = 1$)

$$p_a = \frac{b-x}{b-a}.$$

Similarly, one can easily show that the process $S_t = W_t^2 - t$ is also a martingale. Indeed,

$$\begin{aligned} E[W_t^2 - t|\mathcal{F}_s] &= E[(W_t - W_s)^2 + 2W_s(W_t - W_s) + W_s^2 - t|\mathcal{F}_s] \\ &= E[(W_t - W_s)^2|\mathcal{F}_s] + 2W_s E[W_t - W_s|\mathcal{F}_s] + W_s^2 - t \\ &= (t-s) + 0 + W_s^2 - t = W_s^2 - s \end{aligned}$$

With τ defined as before let us use the optional sampling theorem again. This time we obtain

$$EW_\tau^2 - E\tau = x^2$$

which gives

$$p_a a^2 + p_b b^2 - E\tau = x^2$$

or

$$\frac{(b-x)a^2 + (x-a)b^2}{b-a} - x^2 = E\tau$$

from which we obtain

$$E\tau = ab.$$

An important martingale associated with brownian motion is the exponential martingale. Suppose here that W_t is $BM(\mu, \sigma^2)$. Then, if θ is any real number

$$M_t := e^{\theta W_t - q(\theta)t}, \qquad \text{with } q(\theta) = \mu\theta + \frac{1}{2}\theta^2\sigma^2$$

is a martingale. Indeed,

$$E[M_t|\mathcal{F}_s] = E[e^{\theta(W_t - W_s) - q(\theta)(t-s)}|\mathcal{F}_s]M_s = M_s$$

the last equality following from the fact that $Ee^{\theta(W_t - W_s)} = e^{\mu\theta(t-s) + \frac{1}{2}\theta^2\sigma^2(t-s)}$.

We have thus seen that M_t is a martingale for any choice of θ. If we set $\theta = \theta_0 = -\frac{2\mu}{\sigma^2}$ we see that $q(\theta_0) = 0$ and thus the exponential martingale becomes $e^{\theta_0 W_t}$. We can use this to compute p_a and p_b (defined as before) when $\mu \neq 0$. Indeed, in this case, from the optional sampling theorem we have

$$E[e^{\theta_0 W_\tau}] = e^{\theta_0 x}$$

or

$$p_a e^{\theta_0 a} + p_b e^{\theta_0 b} = e^{\theta_0 x}$$

which gives

$$p_a = \frac{e^{\frac{2\mu}{\sigma^2}(b-x)} - 1}{e^{\frac{2\mu}{\sigma^2}(b-a)} - 1}.$$

The optional sampling theorem can also be used to obtain the Laplace transform of the time until we hit the boundary. Here we will assume that $\mu = 0$, $\sigma = 1$ which corresponds to $q(\theta) = \frac{1}{2}\theta^2$, in order to simplify the algebra. We start with

$$E[e^{\theta W_\tau - \tau q(\theta)}] = e^{\theta x}.$$

or

$$\begin{aligned} e^{\theta x} &= p_a E[e^{\theta W_\tau - \tau q(\theta)} | W_\tau = a] + p_b E[e^{\theta W_\tau - \tau q(\theta)} | W_\tau = b] \\ &= p_a e^{\theta a} E[e^{-q(\theta)\tau} | W_\tau = a] + p_b e^{\theta b} E[e^{-q(\theta)\tau} | W_\tau = b]. \end{aligned}$$

We seem to have the problem that this is one equation and we have two unknowns, $E[e^{-q(\theta)\tau}|W_\tau = a]$ and $E[e^{-q(\theta)\tau}|W_\tau = b]$ but in fact we can get around this problem by setting

$$s = q(\theta) = \frac{1}{2}\theta^2.$$

There are two solutions to this equation,

$$\theta_1 = \sqrt{2s}, \qquad \text{and} \qquad \theta_2 = -\sqrt{2s}.$$

Thus, if we set $f_a(s) = E[e^{-s\tau}; W_\tau = a]$ and $f_b(s) = E[e^{-s\tau}; W_\tau = b]$, we have

$$\begin{aligned} e^{x\sqrt{2s}} &= e^{a\sqrt{2s}} f_a(s) + e^{b\sqrt{2s}} f_b(s) \\ e^{-x\sqrt{2s}} &= e^{-a\sqrt{2s}} f_a(s) + e^{-b\sqrt{2s}} f_b(s). \end{aligned}$$

From this system we can compute $f_a(s)$, $f_b(s)$ separately, and hence also $Ee^{-s\tau} = f_a(s) + f_b(s)$. In fact, adding and subtracting the above equations we get

$$\begin{aligned} \cosh(x\sqrt{2s}) &= \cosh(a\sqrt{2s}) f_a(s) + \cosh(b\sqrt{2s}) \\ \sinh(x\sqrt{2s}) &= \sinh(a\sqrt{2s}) f_a(s) + \sinh(b\sqrt{2s}) \end{aligned}$$

or, using the fact that $\sinh(\alpha - \beta) = \sinh\alpha\cosh\beta - \cosh\alpha\sinh\beta$, we obtain

$$f_a(s)\sinh(b-a)\sqrt{2s} = \sinh(b-x)\sqrt{2s}$$

$$f_b(s)\sinh(b-a)\sqrt{2s} = \sinh(x-a)\sqrt{2s}$$

We thus have

$$f(s) = f_a(s) + b_b(s) = \frac{\sinh\left((x-a)\sqrt{2s}\right) + \sinh\left((b-x)\sqrt{2s}\right)}{\sinh\left((b-a)\sqrt{2s}\right)}$$

and using the formulas $\sinh 2\alpha = 2\sinh\alpha\cosh\alpha$, $\sinh\alpha+\sinh\beta = 2\cosh\left(\frac{\beta-\alpha}{2}\right)\sinh\left(\frac{\alpha+\beta}{2}\right)$ we obtain

$$f(s) = \frac{\cosh\left(\left(\frac{b+a}{2} - x\right)\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}$$

Since we can take $x = 0$ without loss of generality, this formula simplifies as follows

$$f(s) = \frac{\cosh\left(\frac{b+a}{2}\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}.$$

In particular, when $b = \ell > 0$, $a = -\ell$, then

$$f(s) = \frac{1}{\cosh\left(\ell\sqrt{2s}\right)}$$

3.5 The Poisson Process

Intuitively speaking the Poisson process on the real line describes the occurrence of random events in time. Let $\{T_n\}$ $n = 1, 2, \ldots$, denote the positions of the random points in the half line $[0, \infty)$. We number them consecutively so that $0 \le T_1 \le T_2 < \cdots$. Denote by N_t the number of points in the interval $(0, t]$. We can then write $N_t = \sum_{n=1}^{\infty} \mathbf{1}(T_n \le t)$. Note that, even though this sum has an infinite number of terms, only a finite number of them is non-zero. The collection of random variables $\{N_t; t \ge 0\}$ is called a counting process.

Definition 2. *Given any positive number λ a process $\{N_t; t \ge 0\}$ is a Poisson process* with rate λ *if*

- *1. $N_0 = 1$ a.s.*
- *2. For any n and $0 < t_1 < t_2 < \cdots < t_n$ the increments $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.*
- *3. For any $s, t > 0$, $\mathbb{P}(N_{t+s} - N_s = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$, for $k = 0, 1, 2, \ldots$.*

Let $\tau_1 = T_1$ and $\tau_n = T_n - T_{n-1}$, $n = 2, 3, \ldots$, denote the interevent times, i.e. the distances that separate the consecutive points. We can prove using the definition that these are independent exponentially distributed random variables with rate λ. For instance $\mathbb{P}(\tau_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$. Similarly,

$$\mathbb{P}(T_n > \tau) \;=\; \mathbb{P}(N_t < 0) = \sum_{k=0}^{n-1} \mathbb{P}(N_t = k) = \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Differentiating the above expression we obtain the density of T_n as $f_{T_n}(t) = \lambda\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t}$, $t \ge 0$. This is in accordance with the fact that $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$ where the τ_i's are independent, exponential with rate λ.

The Poisson process can be used as a building block for more general processes. We will discuss in particular the compound Poisson process, obtained as follows. Let $\{\sigma_n\}$ be a sequence of independent, identically distributed real random variables, independent of the Poisson process $\{N_t; t \geq 0\}$. Define a new process $\{X_t; t \geq 0\}$ via $X_t = \sum_{k=1}^{N_t} \sigma_k$ with the understanding that an empty sum is 0 i.e. if $N_t = 0$ then $X_t = 0$.

The process $\{X_t; t \geq 0\}$ inherits the independent increments property from the Poisson process so that, for any $n \in \mathbb{N}$ and $0 < t_1 < t_2 < \cdots < t_n$ the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent. In order to obtain the distribution of X_t is is best to use the characteristic function or (provided it exists) the moment generating function defined as $\mathbb{E}[e^{\theta X_t}]$ where $\theta \in \mathbb{R}$ and such that the expectation exists. Thus

$$\mathbb{E}[e^{\theta X_t}] = \mathbb{E}[e^{\theta \sum_{k=1}^{N_t} \sigma_k}] = \mathbb{E}\left[\mathbb{E}\left[e^{\theta \sum_{k=1}^{N_t} \sigma_k} | N_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{k=1}^{N_t} e^{\theta \sigma_k} | N_t\right]\right].$$

Since the σ_k's are independent from each other and from N_t we have

$$\mathbb{E}\left[e^{\theta \sum_{k=1}^{N_t} \sigma_k} | N_t\right] = \left(\mathbb{E}[e^{\theta \sigma_1}]\right)^{N_t}.$$

Hence if the denote the moment generating function of σ_1 by $M_\sigma(\theta) := \mathbb{E}[e^{\theta \sigma_1}]$ and the moment generating function of X_t by $M_{X_t}(\theta)$ we have

$$M_{X_t}(\theta) = \mathbb{E}[M_\sigma(\theta)^{N_t}] = \sum_{k=0}^{\infty} M_\sigma(\theta)^k \frac{(\lambda t)^k}{k!} e^{-\lambda t} = e^{-\lambda t(1 - M_\sigma(\theta))}.$$

From the above expression we can obtain the moments of X_t by differentiation. For instance, $\mathbb{E}[X_t] = \lambda t \mathbb{E}[\sigma_1]$ and $Var(X_t) = \lambda t \mathbb{E}[\sigma_1^2]$.

Chapter 4

Martingales in Discrete Time

4.1 Adapted and Predictable processes

Let $\{Y_n\}_{n\geq 0}$ be an sequence of random variables which we will regard as an information sequence. We will often write $\mathcal{F}_n := \{Y_0, Y_1, \ldots, Y_n\}$.

- The process $\{X_n\}_{n\geq 0}$ is **adapted** to $\{Y_n\}_{n\geq 0}$ if there exists a sequence of functions $f_n : \mathbf{R}^{n+1} \to \mathbf{R}$ such that

  $$X_n = f_n(Y_0, Y_1, \ldots, Y_n), \qquad n \geq 0.$$

  We will write

  $$X_n \in \mathcal{F}_n \ .$$

- The process $\{X_n\}_{n\geq 0}$ is **predictable** with respect to $\{Y_n\}_{n\geq 0}$ if there exists a sequence of functions $f_n : \mathbf{R}^{n+1} \to \mathbf{R}$ such that

  $$X_n = f_{n-1}(Y_0, Y_1, \ldots, Y_{n-1}), \qquad n \geq 1, \quad X_0 = \text{constant}.$$

  We will write

  $$X_n \in \mathcal{F}_{n-1} \ .$$

4.2 Stopping Times

Let T be a nonnegative, integer valued, random variable. T is a stopping time w.r.t. the information pattern $\{\mathcal{F}_n\}$ iff the sequence of random variables $\mathbf{1}(T = n)$, $n = 0, 1, 2, \ldots$, is adapted to $\{\mathcal{F}_n\}$. In particular, note that if T is a stopping time then $\{\mathbf{1}(T \leq n)\}$ is also an adapted sequence, while $\{\mathbf{1}(T > n)\}$ is a *predictable* sequence. To see this, write $\mathbf{1}(T \leq n) = \sum_{k=0}^{k=n} \mathbf{1}(T = k)$ and observe that $\mathbf{1}(T = k) \in \mathcal{F}_k \subset \mathcal{F}_n$ for $k \leq n$. This establishes that $\mathbf{1}(T \leq n) \in \mathcal{F}_n$. On the other hand $\mathbf{1}(T > n) = 1 - \mathbf{1}(T \leq n - 1)$ which, in view of the above is a *predictable* sequence.

Proposition 2. *If S, T, are $\mathcal{F}_n$–stopping times then $S+T$, $S \vee T$, $S \wedge T$ are also $\mathcal{F}_n$–stopping times.*

Proof: To prove the first statement, note that $\mathbf{1}(S+T=n) = \sum_{k=0}^{n} \mathbf{1}(S=k)\mathbf{1}(T=n-k) \in \mathcal{F}_n$. The second follows from $\mathbf{1}(S \vee T \le n) = \mathbf{1}(S \le n)\mathbf{1}(T \le n)$, and the fact that both $\mathbf{1}(S \le n)$ and $\mathbf{1}(S \le n)$ are in $\mathcal{F}_n$ since T and S are stopping times. Finally $\mathbf{1}(S \wedge T > n) = \mathbf{1}(S > n)\mathbf{1}(T > n)$.

♠

4.3 Martingales in Discrete Time

Definition 3. *A process $\{X_n\}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$ if*

Theorem 1.

- *X_n is an adapted process, i.e. $X_n \in \mathcal{F}_n$,*
- *$E|X_n| < \infty \ \forall n$,*
- *$E[X_{n+1}|\mathcal{F}_n] = X_n \ \forall n$.*

Example 1: Let $\{Y_i\}$ be independent random variables with $E|Y_i| < \infty$ for all i and denote by $\mathcal{F}_n = \{Y_0, Y_1, \ldots, Y_n\}$. Let $EX_i = \mu_i$. The process $X_n = \sum_{i=0}^{n} Y_i - \mu_i$ is an $\mathcal{F}_n$–martingale.

Example 2: Using the setup of the previous example suppose that, for all i, $\sigma_i^2 = \mathrm{Var}(Y_i) < \infty$. The process $X_n = \left(\sum_{i=0}^{n} Y_i - \mu_i\right)^2 - \sum_{i=0}^{n} \sigma_i^2$ is an $\mathcal{F}_n$–martingale.

Example 3: Using again the same setup we assume that Y_i has distribution F_i and $\tilde{F}_i(s) := \int_{-\infty}^{\infty} e^{-sx} dF_i(x)$ is finite for s in a neighborhood of 0. Then

$$X_n := \frac{e^{-s\sum_{i=0}^{n} Y_i}}{\prod_{i=0}^{n} \tilde{F}_i(s)} ,$$

is an $\mathcal{F}_n$–martingale.

Example 4: Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathcal{S}$ and transition probability matrix $P(i,j)$. Also suppose that $f : \mathcal{S} \to \mathbf{R}$ be a real function. Then

$$X_n := \sum_{k=1}^{n} \left(f(Y_k) - \sum_{j \in \mathcal{S}} P(Y_{k-1}, j) f(j) \right)$$

is an $\mathcal{F}_n$–martingale.

Example 5: [Right Regular Sequences and Induced Martingales for Markov Chains] Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathcal{S}$ and transition probability matrix $P(i,j)$. Let $f : \mathcal{S} \to \mathbf{R}$ be bounded and satisfy

$$f(i) = \sum_{j \in \mathcal{S}} P(i,j) f(j) , \quad \forall i \in \mathcal{S} .$$

Such sequences (right eigenvectors corresponding to eigenvalue 1) are called right regular sequences. Then

$$X_n = f(Y_n)$$

is a martingale.

Example 6: The above example is a special case of the following more general class of martingales. Let f be a right eigenvector corresponding to an eigenvalue λ of P, i.e.

$$\lambda f(i) \sum_{j \in \mathcal{S}} P(i,j) f(j) \ , \quad \forall i \in \mathcal{S} \ .$$

Assuming that $E|f(Y_n)| < \infty$,

$$X_n = \lambda^{-n} f(Y_n)$$

is a martingale.

Example 7: [Likelihood Ratios] Let $\{Y_n\}$ be an i.i.d. sequence with density g. Let f be another density function. Then the process

$$X_n = \frac{f(Y_0)f(Y_1)\cdots f(Y_n)}{g(Y_0)g(Y_1)\cdots g(Y_n)}$$

is a martingale.

4.4 The Optional Sampling Theorem

4.4.1 The Optional Sampling Theorem for Martingales

Let $\{X_n\}$ be a martingale w.r.t. $\{\mathcal{F}_n\}$. We know that $EX_n = EX_0$. If T is a stopping time, under what conditions is $EX_T = EX_0$? We start with

Lemma 1. *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathcal{F}_n\}$. Then, for all $n \geq k$,*

$$E[X_n \mathbf{1}(T = k)] = E[X_k \mathbf{1}(T = k)] \ .$$

Proof: Indeed

$$\begin{aligned} E[X_n \mathbf{1}(T &= k)] = E[E[X_n \mathbf{1}(T = k)|\mathcal{F}_n] = E[\mathbf{1}(T = k) E[X_k|\mathcal{F}_n]] \\ &= E[\mathbf{1}(T = k) X_n] \end{aligned}$$

♠

Lemma 2. *With the assumptions of the previous lemma*

$$E[X_{T \wedge n}] = EX_0 \ .$$

Proof: We can write $X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbf{1}(T = k) + X_n \mathbf{1}(T \geq n)$ and taking expectations,

$$\begin{aligned} E[X_{T\wedge n}] &= \sum_{k=0}^{n-1} E[X_k \mathbf{1}(T = k)] + E[X_n \mathbf{1}(T \geq n)] \\ &= \sum_{k=0}^{n-1} E[X_n \mathbf{1}(T = k)] + E[X_n \mathbf{1}(T \geq n)] \\ &= E\left[X_n \left(\sum_{k=0}^{n-1} \mathbf{1}(T = k) + \mathbf{1}(T \geq n)\right)\right] \\ &= EX_n = EX_0 \ . \end{aligned}$$

Theorem 2. *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathcal{F}_n\}$. Suppose that $P(T < \infty) = 1$ and $E[\sup_k |X_{T\wedge k}|] < \infty$. Then $EX_T = EX_0$.*

Proof: From the previous lemma we have $EX_{T\wedge n} = EX_0 \ \forall n$. Since $P(T < \infty) = 1$, $\lim_{n\to\infty} X_{T\wedge n} = X_T$. Finally, $X_{T\wedge n} \geq \sup_k |X_{T\wedge k}|$. Use the Dominated Convergence Theorem to conclude that

$$\lim_{n\to\infty} E[X_{T\wedge n}] = E[\lim_{n\to\infty} X_{T\wedge n}] = EX_T \ .$$

♠

Chapter 5

Laws of Large Numbers and Central Limit Theorem

Suppose that $\{X_i\}$, $i = 1, 2, \ldots$, are independent, identically distributed random variables with finite mean, i.e. $E|X| < \infty$.

Theorem 3. ***Weak Law of Large Numbers*** *If* $\mu := EX_1$ *then*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0 \qquad \textit{for all } \epsilon > 0.$$

A stronger version also holds under the same assumptions

Theorem 4. ***Strong Law of Large Numbers:***

$$\lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty}\left\{\left|\frac{X_1 + X_2 + \cdots + X_m}{m} - \mu\right| > \epsilon\right\}\right) = 0 \qquad \textit{for all } \epsilon > 0.$$

If, in addition to the above assumptions we also assume that the variance $VarX_1 = \sigma^2$ is finite then

Theorem 5. ***Central Limit Theorem*** *If* $\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}dy$ *is the cumulative distribution function of the standard normal then*

$$\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x).$$

Chapter 6

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

$$\lim_{n\to\infty} P(S_n \geq nx) = 0 \qquad \text{for } x > m \tag{6.1}$$

and similarly that

$$\lim_{n\to\infty} P(S_n \leq nx) = 0 \qquad \text{for } x < m \tag{6.2}$$

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

$$P(S_n \geq nx) \asymp e^{-nI(x)} \qquad \text{for } x > m \tag{6.3}$$

In the above formula note that the exponential rate of decay $I(x)$ is a function of x. The meaning of (6.3) is made precise if we state it as

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq nx) = -I(x) \qquad \text{for } x > m. \tag{6.4}$$

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

$$P\left(S_n - nm \geq n(x-m)\right) = P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \geq \sqrt{n}\left(\frac{x-m}{\sigma}\right)\right) \tag{6.5}$$

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

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

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 (6.5) becomes an exact relationship and we have

$$P(S_n \geq nx) = \int_{\sqrt{n}\left(\frac{x-m}{\sigma}\right)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du. \tag{6.6}$$

Taking into account the bounds in proposition 1 we have

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

or

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

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

$$\lim_{n\to\infty} \frac{1}{n}\log P(S_n \geq nx) = -\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2. \tag{6.7}$$

Hence, setting $I(x) = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2$ we obtain (6.1) 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.

6.1 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

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

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

$$I(x) := \sup_{\theta \in \mathbb{R}} \{x\theta - \log M(\theta)\}. \tag{6.8}$$

With this definition the Chernoff bound becomes

$$P(S_n \geq nx) \leq e^{-nI(x)}. \tag{6.9}$$

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 6. *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

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

Similarly

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

and thus

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

∎

6.2 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 $\log M(\theta) = \log\left(pe^\theta + 1 - p\right)$. 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

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

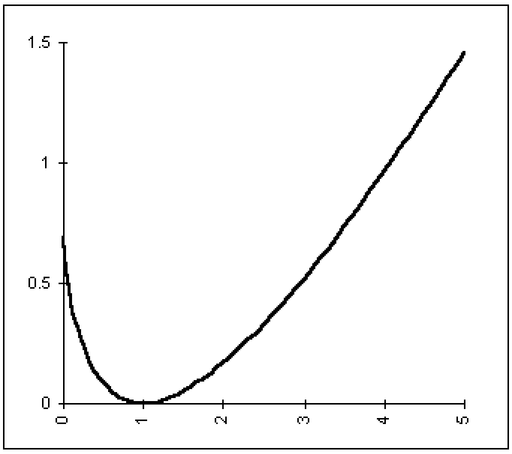

Figure 6.1:

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.

6.3 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,

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

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

$$\frac{M'^{*})}{M(\theta^*)} = x \tag{6.10}$$

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

6.4 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 (6.10),

$$\frac{1}{M(\theta^*)} \int_{-\infty}^{\infty} ye^{\theta^* y} F(dy) = \frac{M'^{*})}{M(\theta^*)} = x. \tag{6.11}$$

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, $i = 1, 2, \ldots$, have distribution F, and the twisted measure $\widetilde{P}$, under which the r.v.'s X_i have distribution $\widetilde{F}$. Expectations with respect to the probability measure $\widetilde{P}$ will be denoted by $\widetilde{E}$.

6.5 Cramér's Theorem

Theorem 7. *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 $S_n := X_1 + \cdots + X_n$,*

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

Bibliography

[1] Asmussen, S. (1987). *Applied Probability and Queues,* John Wiley.

[2] Geoffrey R. Grimmett, David R. Stirzaker, (2001) *Probability and Random Processes*, 3rd edition, Oxford University Press.

[3] Sheldon Ross. (1995). *Stochastic Processes,* 2nd edition, John Wiley.