Stochastic Processes

June 25, 2012

Chapter 1

Introduction

1.1 Discrete Time

Let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A discrete time stochastic process is a family of real random variables $\{X_n; n \in \mathbb{N}_0\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Its *law* is determined by the family of finite dimensional distributions

$$\begin{aligned} F_m(x_1, \ldots, x_m; n_1, \ldots, n_m) &= \mathbb{P}(X_{n_1} \leq x_1, \ldots, X_{n_m} \leq x_m), \\ & m \in \mathbb{N},\ x_k \in \mathbb{R},\ k = 1, \ldots, m,\ \ 0 \leq n_1 < n_2 < \cdots < n_m \ \in \mathbb{N}_0. \end{aligned} \tag{1.1}$$

The family of distributions must satisfy the consistency conditions

$$\begin{aligned} &F_{m+1}(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_{m+1}; n_1, \ldots, n_{k-1}, n_k, n_{k+1}, \ldots, n_{m+1}) \\ &\quad = F_m(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{m+1}; n_1, \ldots, n_{k-1}, n_{k+1}, \ldots, n_{m+1}) \end{aligned} \tag{1.2}$$

for all $m \in \mathbb{N}$ and all $x_k \in \mathbb{R}$.

Conversely, for any given family of distribution functions $\{F_m(x_1, \ldots, x_m; n_1, \ldots, n_m);\ m \in \mathbb{N}, x \in \mathbb{R}^m, 0 \leq n_1 < \cdots < n_m\}$ that satisfy the consistency conditions (1.2) there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of measurable functions $\{X_n\}$, $X_n : \Omega \to \mathbb{R}$ such that (1.1) holds, i.e. $\{X_n\}$ is a stochastic process with the given family of finite dimensional distributions. (This statement is Kolmogorov's theorem.)

1.2 Continuous Time

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consider a family of real random variables $\{X_t; t \in [0, T]\}$ (where $T > 0$). For each $t \in [0, T]$ X_t is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ (where $\mathcal{B}$

is the Borel σ–field on the real line) while, for each $\omega \in \Omega$ $X.(\omega) : [0,T] \to \mathbb{R}$ is a measurable function. The family of distribution functions

$$F_m(x_1,\ldots,x_m;t_1,\ldots,t_m) = \mathbb{P}(X_{t_1} \le x_1,\ldots,X_{t_m} \le x_m), \quad (1.3)$$
$$m \in \mathbb{N},\ x_k \in \mathbb{R},\ k=1,\ldots,m,\ \ 0 \le t_1 < t_2 < \cdots < t_m \le T$$

are the *finite dimensional distributions* of the process $\{X_t\}$ and must satisfy consistency conditions analogous to (1.2).

1.2.1 Brownian Motion

Consider the family of finite-dimensional densities

$$f(x_1,\ldots,x_n;t_1,\ldots,t_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} e^{-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}} \quad (t_0 := 0). \quad (1.4)$$

This family defines a corresponding family of distribution functions which satisfies the consistency conditions. Therefore, by virtue of the Kolmogorov theorem, there exists a process $\{W_t; t \ge 0\}$ with these finite dimensional distributions. It can be shown that there exists a version of the process that has continuous sample paths with probability 1.

1.2.2 Markov Processes

A process $\{X_t; t \in \mathbb{R}\}$ has the Markov property if for any $n \in \mathbb{N}$, $t_1 < t_2 < \ldots < t_n$ and any $x_1,\ldots,x_n \in \mathbb{R}$

$$\mathbb{P}(X_{t_n} \le x_n | X_{t_{n-1}} = x_{t_{n-1}}, X_{t_{n-2}} = x_{t_{n-2}} \ldots, X_{t_1} = x_{t_1}) = \mathbb{P}(X_{t_n} \le x_n | X_{t_{n-1}} = x_{t_{n-1}}). \quad (1.5)$$

The meaning of the above is that in order to "predict" the value of the process at a future time t_n, only the most recent known value $X_{t_{n-1}}$ is relevant. Past values, $X_{t_{n-2}}, X_{t_{n-3}}, \ldots, X_{t_1}$ provide no further information if the process is markovian.

Clearly, a markovian process can be described completely by the so-called transition kernel, $P(y,x;t,s) := \mathbb{P}(X_t \le y | X_s = x)$ for all $s < t$ and $x, y \in \mathbb{R}$. The process is called *time-homogeneous* if $P(y,x;t,s) = P(y,x;t-s,0)$ which means that $\mathbb{P}(X_t \le y | X_s = x) = \mathbb{P}(X_{t-s} \le y | X_0 = x) = Q(y,x;t-s)$ for some transition kernel Q. Note that the transition kernel Q for each fixed t and x is a distribution function on the real line, the conditional distribution of X_t given that $X_0 = x$.

We will in particular consider the case where the kernel possesses a density $p(y,x;t)dy = \mathbb{P}(X_t \in dy | X_0 = x)$. Let us denote by $f(x_n, x_{n-1}, \ldots, x_1; t_n, t_{n-1}, \ldots, t_1) dx_n dx_{n-1} \ldots dx_1$ the

joint density $\mathbb{P}(X_{t_n} \in dx_n, X_{t_{n-1}} \in dx_{n-1}, \ldots, X_{t_1} \in dx_1)$. Then

$$\begin{aligned} f(x_n, x_{n-1}, \ldots, x_1, x_0; t_n, t_{n-1}, \ldots, t_1, t_0) &= f(x_0; t_0) p(x_1, x_0; t_1 - t_0) p(x_2, x_1; t_2 - t_1) \\ &\quad \cdots p(x_n, x_{n-1}; t_n - t_{n-1}). \end{aligned}$$

Thus if the initial density $f(x_0; t_0)$ and the transition function $p(y, x; t)$ are given the finite dimensional distributions and thus the whole process are determined in the Markovian case. In particular the standard Brownian motion is a Markov process with transition density

$$p(y, x; t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}.$$

1.3 The Multivariate Normal Distribution

1.3.1 Symmetric Nonnegative Definite Matrix

Definition 1. *A symmetric $n \times n$ matrix R is positive definite if, for all $x \neq 0$ in $\mathbb{R}^n$, $x \top R x > 0$. It is non-negative definite if $x^\top R x \geq 0$.*

Recall that any symmetric matrix $n \times n$ has n real eigenvalues (not necessarily distinct) and n corresponding eigenvectors that can be taken to be orthogonal:

$$R\phi_i = \lambda_i \phi_i, \quad i = 1, \ldots, n, \quad \text{with } \phi_i^\top \phi_j = \delta_{ij}, \tag{1.6}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. If $\Phi := [\phi_1, \ldots, \phi_n]$, the matrix whose columns are the eigenvectors of R, and Λ the diagonal matrix of the eigenvalues, then the above relationship can also be written as

$$R = \Phi \Lambda \Phi^\top.$$

An equivalent restatement of the above is the so-called *spectral representation*

$$R = \sum_{i=1}^{n} \lambda_i \phi_i \phi_i^\top.$$

It is easy to see that R is non-negative definite provided that $\lambda_i \geq 0$ for all i and positive definite if $\lambda_i > 0$ for all i.

If R is non-negative definite then there exists a real square matrix V such that

$$R = VV^\top. \tag{1.7}$$

Indeed, since $\lambda_i \geq 0$ we can define $\Lambda^{1/2}$ as the diagonal matrix with elements $\lambda_i^{1/2}$ and then we can take $V := \Phi \Lambda^{1/2}$. The non-uniqueness of V is obvious since we could make the choice $-\lambda_i^{1/2}$ for some of the diagonal elements. However there are other, more interesting possibilities.

Recalling the LDU decomposition, where L and U are lower and upper triangular matrices with unit elements on the diagonal and taking into account that R is symmetric we obtain the decomposition

$$R = LDL^\top \tag{1.8}$$

where D is the diagonal matrix with elements d_i. R is positive definite if and only if $d_i > 0$ for all i. Thus, an essentially different choice for V in the decomposition (1.7) is $V = LD^{1//2}$.

1.3.2 Moment Generating Function and Joint Density

Definition 2. *A random vector $(X_1, \ldots, X_n)$ is normal with mean $\mu \in \mathbb{R}^n$ and covariance matrix R if its moment generating function is*

$$M(u_1, \ldots, u_n) := \mathbb{E}e^{\sum_{i=1}^n u_i X_i} = e^{\sum_{i=1}^n u_i \mu_i + \frac{1}{2}\sum_{i,j=1}^n u_i R_{ij} u_j}. \tag{1.9}$$

An immediate consequence is the following proposition

Proposition 3. *$(X_1, \ldots, X_n)$ is multivariate normal if and only if $\sum_{i=1}^n a_i X_i$ is normal for any vector $(a_1, \ldots, a_n)$.*

Problem 4. *Suppose that R is an $n \times n$ covariance matrix with rank $k \leq n$. Show that there exists an $n \times k$ matrix G such that $R = GG^\top$. If U_i, $i = 1, \ldots, k$ are independent, standard normal random variables, then*

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = G \begin{bmatrix} U_1 \\ \vdots \\ U_k \end{bmatrix}$$

If the rank of R is n (equivalently if all its eigenvalues are strictly positive) then the law of $(X_1, \ldots, X_n)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ with density

$$f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}(\det R)^1/2} e^{-\frac{1}{2}(x-\mu)^\top R^{-1}(x-\mu)}. \tag{1.10}$$

Suppose for simplicity that them mean μ is zero. It is well known that the joint moments are given in terms of the moment generating function via the expression

$$\frac{1}{\prod_{j=1}^n i_j!} \frac{\partial^{i_1+\cdots+i_n}}{\partial u_1^{i_1} \cdots \partial u_n^{i_n}} M(u_1, \ldots, u_n)\Big|_{u_1=\cdots=u_n=0} = \mathbb{E}[X_1^{i_1} \cdots X_n^{i_n}]. \tag{1.11}$$

Equivalently, $\frac{1}{(\sum_{j=1}^n i_j)!}\mathbb{E}[X_1^{i_1} \cdots X_n^{i_n}]$ is the coefficient of $u_1^{i_1} u_2^{i_2} \cdots u_n^{i_n}$ in the Taylor expansion of $M(u_1, \ldots, u_n)$ around 0. Since $M(u_1, \ldots, u_n) = e^{\frac{1}{2}\sum_{kl} u_k u_l r_{kl}}$ this series expansion has the form

$$M(u_1, \ldots, u_n) = 1 + \sum_{m=1}^{\infty} \frac{1}{2^m} \left(\sum_{kl} u_k u_l r_{rl} \right)^m$$

As a simple illustration consider that

$$\mathbb{E}[X_1X_2X_3X_4] = r_{12}r_{34} + r_{13}r_{24} + r_{14}r_{23}.$$

Similarly

$$\mathbb{E}[X_1^2X_2^2] = r_{11}r_{22} + 2r_{12}^2,$$
$$\mathbb{E}[X_1^4] = 3r_{11}^4,$$

and generally

$$\mathbb{E}[X_1^{2n}] = 1 \cdot 3 \cdot 5 \cdots (2n-1).$$

Also,

$$\mathbb{E}[X_1^2X_2^2X_3^2] = r_{11}r_{22}r_{33} + 2r_{11}r_{23}^2 + 2r_{22}r_{13}^2 + 2r_{33}r_{12}^2 + 8r_{12}r_{23}r_{31}$$

Can you discover the general pattern? A relevant combinatorial problem is the following:

The number of pairs that that can be formed by $2n$ different objects is

$$\binom{2n}{n} n! \frac{1}{2^n} = \frac{(2n)!}{2^n n!} = 1 \cdot 3 \cdot 5 \cdots (2n-1).$$

1.4 Expansion in Orthogonal Functions – Karhunen-Loève

Here we give a concrete representation for Brownian motion with variance constant σ^2 in the interval $[0, T]$. The covariance function is given by $R(s,t) = \sigma^2(s \wedge t)$, $s, t \in [0, T]$.. By analogy with the discrete case (1.6) we can attempt to solve the eigenvalue problem

$$\int_0^T R(s,t)\phi_i(s)ds = \lambda_i\phi_i(t). \tag{1.12}$$

This will determine a sequence of eigenvalues $\{\lambda_i\}$, $i = 1, 2, \ldots$ and corresponding eigenfunctions $\{\phi_i\}$ that may be taken to be of unit norm, i.e. to satisfy $\int_0^T \phi_i^2(s)ds = 1$. The integral equation (1.12) then becomes

$$\sigma^2 \int_0^t s\phi_i(s)ds + \sigma^2 \int_t^T \phi_i(s)ds = \lambda_i\phi_i(t)\ . \tag{1.13}$$

Differentiating with respect to t we obtain

$$\sigma^2 \int_t^T \phi_i(s)ds = \lambda_i\phi_i'(t) \tag{1.14}$$

and differentiating once again

$$-\sigma^2\phi_i(t) = \lambda_i\phi_i''(t). \tag{1.15}$$

Setting $t = 0$ in (1.13) and $t = T$ in (1.14) we obtain the boundary conditions

$$\phi_i(0) = 0, \qquad \phi_i'(T) = 0. \tag{1.16}$$

whence

$$\lambda_i = \frac{\sigma^2 T^2}{\pi^2 \left(i - \frac{1}{2}\right)^2}, \qquad \phi_i(t) = \sqrt{\frac{2}{T}} \sin \pi \left(i - \frac{1}{2}\right) \frac{t}{T}, \qquad i = 1, 2, \ldots. \tag{1.17}$$

Thus we obtain the remarkable expression

$$R(s,t) = \sigma^2(s \wedge t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(t)$$

for the covariance function and the following representation for the Brownian motion: If $\{Z_i\}$ are i.i.d. standard normal random variables then

$$X_t := \sum_{i=1}^{\infty} Z_i \lambda_i^{1/2} \phi_i(t), \quad t \in [0, T] \tag{1.18}$$

is Brownian motion with variance constant σ^2.

1.5 Processes with Stationary Independent Increments

Definition 5. *A process $\{X_t; t \geq 0\}$ with stationary independent increments, also called a Lévy process, is one satisfying the following three conditions.*

i) $\mathbb{P}(X_0 = 0) = 1$ *and the process has with probability 1 paths which are right-continuous with left hand limits.*

ii) For all $n \in \mathbb{N}$ *and* $0 \leq t_0 < t_1 < \ldots < t_n$ *the random variables* $X_{t_i} - X_{t_{i-1}}$, $i = 1, \ldots, n$ *are independent.*

iii) For all $s, t > 0$ *the random variables* X_t *and* $X_{t+s} - X_t$ *are identically distributed.*

This family includes some of the best known and simplest processes, namely the Poisson process, the compound Poisson process, and the Brownian motion. The most striking property of a Lévy process is apparent from the definition. Let $\Phi_t(u) := \mathbb{E}e^{iuX_t}$ denote the characteristic function of X_t and $n \in \mathbb{N}$. Then the random variables $X_{kt/n} - X_{(k-1)t/n}$, $k = 1, 2, \ldots, n$ are i.i.d. random variables and hence

$$\Phi_t(u) = \mathbb{E}[e^{iX_t}] = \mathbb{E}[\prod_{k=1}^{n} e^{it(X_{kt/n} - X_{(k-1)t/n})}] = \prod_{k=1}^{n} \mathbb{E}[e^{it(X_{kt/n} - X_{(k-1)t/n})}] = \Phi_{t/n}(u)^n.$$

As we will argue later a consequence of the above identity is that $\Phi_t(u) = e^{t\phi(u)}$ where

$$\phi(u) = iu\beta - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} (e^{iux} - 1 - iux\mathbf{1}(x \leq 1))\nu(dx) \tag{1.19}$$

where $\beta \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a σ–finite measure on $\mathbb{R} \setminus \{0\}$ such that

$$\int_{-1}^{1} x^2 \nu(dx) < \infty \quad \text{and} \int_{|x|>1} \nu(dx) < \infty.$$

The triplet (β, σ^2, ν) is called the Lévy triplet and it characterizes the law of the process. We will see specific examples later on. Suffice it to say at this point that if $\nu \equiv 0$ then we obtain Brownian motion with drift β and variance constant σ^2 since $\Phi_t(u) = e^{it\beta - \frac{1}{2}u^2\sigma^2}$ whereas if $\nu(dx)$ is a finite measure on $\mathbb{R}$ we obtain a compound Poisson process.

Chapter 2

Characteristic Functions

Let X a real random variable with distribution function F. We denote by μ the corresponding *measure* induced on the real line by F via the relationship $\mu(a,b] = F(b) - F(a)$. The characteristic function corresponding to X (or equivalently to F or μ) is

$$\phi(t) = \int_{\mathbb{R}} e^{itx} dF(x) = \int_{\mathbb{R}} e^{itx} \mu(dx) = Ee^{itX} \tag{2.1}$$

where $i = \sqrt{-1}$ is the imaginary unit and $t \in \mathbb{R}$. Thus f is a function from $\mathbb{R}$ to $\mathbb{C}$. Recalling de Moivre's formula for the complex exponential, $e^{ix} = \cos x + i \sin x$ we can also write

$$\phi(t) = \int_{\mathbb{R}} \cos(xt) dF(x) + i \int_{\mathbb{R}} \sin(xt) dF(x)$$

Suppose that the distribution function F is *symmetric*, i.e. $P(X > x) = P(X < -x)$ for every x, or equivalently $1 - F(x) = F(-x-)$. Then, taking into account the fact that $\sin x$ is an *odd* function we can see that the imaginary part of the characteristic function vanishes and we are left with

$$\phi(t) = \int_{\mathbb{R}} \cos(xt) dF(x)$$

From the above definition it is obvious that the probability distribution specifies the characteristic function. Later in this discussion we will also prove the *uniqueness theorem* which states that the characteristic function uniquely specifies the probability measure. Hence, knowledge of the characteristic function of a random variable is enough to determine its distribution. We will begin with some useful elementary results.

If $f(t)$ is a characteristic function then $f(0) = 1$. This follows by direct substitution into (2.1).

A characteristic function is *uniformly continuous*, i.e. $\forall \epsilon > 0\ \exists \delta > 0$ such that $|f(t+h) - f(t)| < \epsilon$ whenever $|h| < \delta$ for all $t \in \mathbb{R}$. Indeed,

$$\begin{aligned} |f(t+h) - f(t)| &= \left| \int_{\mathbb{R}} e^{i(t+h)x} dF(x) - \int_{\mathbb{R}} e^{itx} dF(x) \right| \\ &\leq \int_{\mathbb{R}} \left| e^{itx}(e^{ixh} - 1) \right| dF(x) = \int_{\mathbb{R}} \left| (e^{ixh} - 1) \right| dF(x) \end{aligned}$$

However, $|e^{ixh} - 1| \leq 2$ and $\int_{\mathbb{R}} 2dF(x) < \infty$, hence we can appeal to the Dominated Convergence theorem to argue that $\lim_{h \to 0} \int_{\mathbb{R}} \left| (e^{ixh} - 1) \right| dF(x) = 0$. Thus the result is established.

If the characteristic function (ch. f.) of the random variable X is $f(t)$, then the ch. f. of $aX + b$ is $e^{itb} f(at)$. This follows immediately from $E[e^{it(aX+b)}] = e^{itb} E e^{i(at)X}$.

Let $f_i(t) = E[e^{itX_i}]$, $i = 1, 2$, where X_1, X_2 are *independent* random variables. Then the characteristic function of their sum is the product of the characteristic functions: $E e^{it(X_1+X_2)} = E e^{itX_1} E e^{itX_2} = f_1(t) f_2(t)$. This of course generalizes to sums of independent random variables with arbitrarily many terms.

Let X, X' independent random variables *with the same distribution* and characteristic function $f(t)$. Show that $E e^{it(X+X')} = f(t)^2$ and $E e^{it(X-X')} = f(t)\overline{f(t)} = |f(t)|^2$. This shows that whenever $f(t)$ is a characteristic function, $|f(t)|^2$, which is always real–valued, is also a characteristic function.

Let Z be a standard normal random variable. Then its characteristic function is $f(t) = e^{-t^2/2}$. Indeed, $f(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx e^{itx} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(tx) dx$ since the density of the standard normal is an even function. Thus differentiating with respect to t inside the integral and integrating by parts gives

$$\begin{aligned} f'(t) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x \sin(tx) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(tx) d(e^{-x^2/2}) \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(tx) dx = -f(t) \end{aligned}$$

Thus, $f(t)$ satisfies the differential equation

$$f'(t) = -tf(t), \qquad f(0) = 1$$

which has the solution $f(t) = e^{-t^2/2}$. From the above it follows that the characteristic function of a normal r.v. with mean μ and variance σ^2 is $e^{it\mu - t^2\sigma^2/2}$.

The exponential distribution with density e^{-x}, $x \geq 0$ has characteristic function

$$\int_0^{\infty} e^{-x} e^{-itx} dx = \int_0^{\infty} e^{-x(1+it)} dx = \frac{1}{1+it}.$$

In the same way we can compute the characteristic function of the Laplace distribution with density $\frac{1}{2} e^{-|x|}$, $x \in \mathbb{R}$ as

$$\frac{1}{2} \left(\frac{1}{1+it} + \frac{1}{1-it} \right) = \frac{1}{1+t^2}$$

If $f_i(t)$ are characteristic functions (corresponding to distribution functions $F_i(x)$), $i = 1, 2, 3, \ldots$, and $p_i \geq 0$, $\sum_i p_i = 1$, then $\sum_i p_i f_i(t)$ is the characteristic function that corresponds to the distribution $\sum_i p_i F_i(x)$. This idea of course extends from sums to integrals: If $F(x, a)$ is a distribution depending with a parameter a with characteristic function $f(t, a)$ and G is another distribution function, then $\int f(t, a) dG(a)$ is the characteristic function of the "mixed" distribution $\int F(x, a) dG(a)$.

The following table gives examples of distributions and the characteristic functions that correspond to them.[1]

	Distribution/Density Function	Characteristic Function
1.	Deterministic: $F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$	e^{ita}
2.	Bernoulli: $F(x) = \begin{cases} 0 & x < 0 \\ q & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$	$q + pe^{it}$
3.	Uniform with density $F'(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$	$e^{it/2} \frac{\sin(t/2)}{t/2}$
4.	Standard Normal with density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$e^{-t^2/2}$
5.	Gamma with density $\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$	$\left(\frac{1}{1-it}\right)^{\alpha}$
6.	Triangular density $F'(x) = \begin{cases} (1 - \lvert x \rvert)^+ & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$\left(\frac{\sin(t/2)}{t/2}\right)^2$
7.	Cauchy density $\frac{1}{\pi} \frac{1}{1+x^2}$	$e^{-\lvert t \rvert}$
8.	Geometric $P(X = k) = q^{k-1}p,\ k = 1, 2, 3, \ldots$	$\frac{pe^{it}}{1-qe^{it}}$
9.	Binomial $P(X = k) = \binom{n}{k} p^k q^{n-k}$	$\left(q + pe^{it}\right)^n$
10.	Poisson $P(X = k) = \frac{\alpha^k}{k!} e^{-\alpha}\ k = 0, 1, 2 \ldots$	$e^{\alpha(e^{it}-1)}$

[1] We denote by x^+ the positive part of a real number x, i.e. $x^+ = \max(0, x)$ and by x^- the negative part, $x^- = -\min(0, x)$. Thus $x = x^+ - x^-$

2.1 The Uniquness Theorem

Before we turn to the central result regarding characteristic functions we need the following fact from analysis

$$\int_0^\infty \frac{\sin ax}{x} dx = \frac{\pi}{2}\text{sgn}(a) := \begin{cases} \frac{\pi}{2} & a > 0 \\ 0 & a = 0 \\ -\frac{\pi}{2} & a < 0 \end{cases} \tag{2.2}$$

It is enough to establish the result for $a > 0$ as the others follow from a simple change of variables. To this end, write

$$\begin{aligned} \int_0^\infty \frac{\sin ax}{x} dx &= \int_0^\infty \frac{\sin ax}{ax} d(ax) = \int_0^\infty \frac{\sin x}{x} dx \\ &= \int_0^\infty \sin x \left[\int_0^\infty e^{-ux} du \right] dx = \int_0^\infty \left[\int_0^\infty e^{-xu} \sin x dx \right] du \\ &= \int_0^\infty \frac{1}{1+u^2} du = \frac{\pi}{2} \end{aligned}$$

Theorem 6. *[Uniqueness Theorem] The characteristic function uniquely specifies the probability measure via the relationship*

$$\mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{it} f(t) dt. \tag{2.3}$$

Proof: In view of the definition of the characteristic function the integral on the right hand side of (2.3) is

$$\int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{it} \int_{\mathbb{R}} e^{itx} dF(x) dt = \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{it} e^{itx} dF(x) dt.$$

Since for any real y

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{ity}}{it} dt = \frac{1}{\pi} \int_0^T \frac{e^{ity} - e^{-ity}}{2it} dt = \frac{1}{\pi} \int_0^T \frac{\sin yt}{t} dt =: S_T(y),$$

taking into account the above equation, the right hand side of (2.3) becomes

$$\lim_{T\to\infty} \int_{\mathbb{R}} \left(S_T(x-b) - S_T(x-a) \right) dF(x). \tag{2.4}$$

[1] The interchange of the order of the two integrals here is justified by Fubini's theorem since

$$e^{-itb} - e^{-ita} = -\int_a^b ite^{-itu} du$$

and hence

$$\left| \frac{e^{-itb} - e^{-ita}}{it} e^{itx} \right| \le |e^{itx}| \int_a^b |e^{-itu}| du = |b-a|.$$

.

As we have seen from (2.2), $\lim_{T\to\infty} S_T(y) = \frac{1}{2}\text{sgn}(y)$. Also, $|S_T(y)| \le c$, hence,

$$\left|\int_{\mathbb{R}} (S_T(x-b) - S_T(x-a))\, dF(x)\right| \le \int_{\mathbb{R}} |S_T(x-b) - S_T(x-a)|\, dF(x) \le 2c\int_{\mathbb{R}} dF(x) = 2c$$

and we can appeal to the Dominated Convergence Theorem in order to interchange the order of the limit and the integral in (2.4). Thus

$$\lim_{T\to\infty} S_T(x-b) - S_T(x-a) = (a < x < b) + \frac{1}{2}(x = a) + \frac{1}{2}(x = b)$$

and hence

$$\begin{aligned}\lim_{T\to\infty} &\int_{\mathbb{R}} (S_T(x-b) - S_T(x-a))\, dF(x) \\ &= \int_{\mathbb{R}} \left((a < x < b) + \tfrac{1}{2}(x = a) + \tfrac{1}{2}(x = b)\right) \mu(dx) \\ &= \mu(a,b) + \tfrac{1}{2}\mu\{a\} + \tfrac{1}{2}\mu\{b\}.\end{aligned}$$

♠

Theorem 7. *Suppose that the characteristic function $f(t)$ is integrable, i.e. $\int_{\mathbb{R}} |f(t)|dt < \infty$. Then the corresponding distribution function is absolutely continuous with corresponding density $p(x)$ given by*

$$p(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx} f(t)dt. \tag{2.5}$$

Proof: We begin with the remark that if f is integrable, then the corresponding distribution function has no atoms, i.e. $F(x) = F(x-)$. Indeed, from the uniquness theorem we have

$$\frac{F(x) - F(x-h)}{h} = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{1 - e^{iht}}{ith} e^{-itx} f(t)dt.$$

Since $\left|\frac{1-e^{iht}}{ith}\right| \le 1$, the integrand is bounded by $\left|\frac{1-e^{iht}}{ith} e^{-itx} f(x)\right| \le |f(t)|$, and since by assumption $f(t)$ is integrable, we can appeal to the Dominated Convergence Theorem to obtain

$$\begin{aligned}\lim_{h\to 0} \frac{F(x) - F(x-h)}{h} &= \lim_{h\to 0} \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1 - e^{iht}}{ith} e^{-itx} f(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f(t)dt.\end{aligned}$$

The above argument establishes that the left derivative of $F(x)$ exists and is given by the above expression. An identical argument shows that the right derivative also exists and equals the same quantity. This completes the proof. ♠

The behavior of the characteristic function near the origin determines the "heaviness" of the tails of the deistribution. This idea is formalized in the following inequalities

Theorem 8. ***[Modulus Inequalities]*** *If we denote by $\mu[-A, A]^c$ the probability $P(|X| > A)$ for any $A > 0$, then*

$$\mu[-A, A]^c \le \frac{2}{A}\int_{-1/A}^{1/A} [1 - f(t)]dt. \tag{2.6}$$

Proof:

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T} f(t)dt &= \frac{1}{2T}\int_{-T}^{T}\int_{\mathbb{R}} e^{itx}\mu(dx)dt = \frac{1}{2T}\int_{\mathbb{R}}\mu(dx)\left(\int_{-T}^{T} e^{itx}dt\right) \\ &= \int_{\mathbb{R}}\mu(dx)\int_{0}^{T}\frac{\cos tx}{T}dt = \int_{\mathbb{R}}\frac{\sin Tx}{Tx}\mu(dx). \qquad (2.7)\end{aligned}$$

Note however that

$$\left|\frac{\sin Tx}{Tx}\right| \leq 1\begin{cases} 1 & |x| \leq 2A \\ \frac{1}{2TA} & |x| > 2A \end{cases}$$

and hence

$$\int_{\mathbb{R}}\frac{\sin Tx}{Tx}\mu(dx) \leq \mu[-2A,2A] + \frac{1}{2TA}(1-\mu[-2A,2A]) = \left(1-\frac{1}{2TA}\right)\mu[-2A,2A] + \frac{1}{2TA}.$$

If we set $T = A^{-1}$ in the above we obtain

$$\left|\frac{A}{2}\int_{-A^{-1}}^{A^{-1}} f(t)dt\right| \leq \frac{1}{2}\mu[-2A,2A] + \frac{1}{2}.$$

From this last inequality, (2.6) follows readily. ♠

2.2 Weak Convergence

In this section we sketch briefly (and mostly without proof) some of the most important results regarding weak convergence of distribution functions. The set up is the following: Suppose that a family of random variables $\{X_n\}$ is given with corresponding distribution functions F_n. (It is important to note that we are not concerned at all here with the *joint statistics* of the family X_n, only with their marginal distributions $F_n(x) = P(X_n \leq x)$, so the random variables do not even have to be defined on the same probability space.)

Definition 9. *$\{F_n\}$ converges weakly to a distribution function F if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for each point of continuity of $F(x)$.

Weak convergence is often refered to as *convergence in distribution* and we write $F_n \xrightarrow{\mathrm{d}} F$.

Theorem 10 (Helly). *Let $\{F_n\}$ be an arbitrary collection of distribution functions. Then there exists a subsequence $\{F_{n_k}\}$ such that*

$$F_{n_k} \xrightarrow{\mathrm{d}} F$$

for some distribution F.

Theorem 11. *$\{F_n\}$ converges weakly to F if and only if*

$$\lim_n \int_{\mathbb{R}} f(x)dF_n(x) = \int_{\mathbb{R}} f(x)dF(x)$$

for every bounded, continuous f.

(This is sometimes refered to as Helly's second theorem.)

As we shall see when we discuss the Central Limit Theorem later on, one of the problems that arises very often, both in practice and in theory is the following. If we have a family of distributions $\{F_n\}$ with corresponding characteristic functions f_n then,

a) If F_n converges weakly to some distribution function F can we conclude that f_n will converge to the characteristic function f of F?

b) If $f_n(t)$ converges for all t to $f(t)$, then is $f(t)$ also a characteristic function, and if it is and it corresponds to (say) the distribution F, can we conclude from this that $F_n \xrightarrow{\mathrm{d}} F$?

The first question has an affirmative answer as one can show without much effort (essentially this follows from Helly's second theorem). The answer to the second question however is more complicated as we can see from the following example.

Let

$$F_n(x) = \begin{cases} 0 & x < -n \\ \frac{x+n}{2n} & -n \leq x < n \\ 1 & n \leq x \end{cases}$$

i.e. we have a family of uniform distributions on $[-n, n]$. Their ch.f.'s are

$$f_n(t) = \frac{\sin(nt)}{nt}$$

We thus see that

$$f_n(t) \longrightarrow f(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ 1 & t = 0 \end{cases}$$

It is easy to see that the above limit is not a characteristic function (it is not continuous!). Also, in this case, $F_n(x) \to 0$ for all x so $\{F_n\}$ does not converge to a distribution function. Thus clearly it is not enough for f_n to converge.

Theorem 12. ***[Convergence Theorem]*** *Let $\{F_n\}$ be probability distributions with characteristic functions $\{f_n\}$. If*

a) $f_n(t)$ converges for every t and defines a limit function $f(t)$

b) This limit function $f(t)$ is continuous at $t = 0$

then

$\{F_n\}$ converges weakly to some distribution F with characteristic function F.

2.3 Positive definite functions

Definition 13. *A function $f : \mathbb{R} \to \mathbb{C}$ is* positive definite *if for every $n \in \mathbb{N}$, $t_1, t_2, \ldots, t_n \in \mathbb{R}$ and $c_1, c_2, \ldots, c_n \in \mathbb{C}$,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i \bar{c}_j f(t_i - t_j) \geq 0, \tag{2.8}$$

where $\bar{c}$ denotes the complex conjugate of c.

(The meaning of the above inequality is that the left hand side should be real and nonnegative.) Note that the positive definiteness of f is equivalent to the positive definiteness of the matrix

$$\begin{bmatrix} f(0) & f(t_1 - t_2) & f(t_1 - t_3) & \cdots & f(t_1 - t_n) \\ f(t_2 - t_1) & f(0) & f(t_2 - t_3) & \cdots & f(t_2 - t_n) \\ f(t_3 - t_1) & f(t_3 - t_2) & f(0) & \cdots & f(t_3 - t_n) \\ & & & & \\ f(t_n - t_1) & f(t_n - t_2) & f(t_n - t_3) & \cdots & f(0) \\ & & & & \end{bmatrix} \tag{2.9}$$

for every n and $t_i \in \mathbb{R}$. Also note that we use the term positive in the weak sense following common usage (in other words, here positive means "nonnegative").

Theorem 14. *All characteristic functions are positive definite*

Proof: We start with the remark that e^{itx} is positive definite. Indeed,

$$\sum_{i,j} c_i \bar{c}_j e^{i(t_i - t_j)} = \sum_{i,j} c_i e^{it_i x} \overline{(c_j e^{it_j x})} = \left(\sum_i c_i e^{it_i x} \right) \overline{\left(\sum_j c_j e^{it_j x} \right)} = \left| \sum_i c_i e^{it_i x} \right|^2 \geq 0.$$

To show that a characteristic function is positive definite, it is enough to mimic the above

argument, interchanging summations and expectation:

$$\sum_{i,j} f(t_i - t_j)c_i\bar{c}_j = \sum_{i,j} c_i\bar{c}_j E[e^{i(t_i-t_j)X}] = E\sum_{i,j} c_i\bar{c}_j e^{i(t_i-t_j)X} = E\left|\sum_i c_i e^{it_i x}\right|^2 \geq 0$$

More interesting and far–reaching however is the fact that the converse is also true, namely that All positive definite functions $f : \mathbb{R} \longrightarrow \mathbb{C}$ are characteristic functions of some measure on the real line. This result will be established latter on. We first establish some of the properties of positive definite functions.

1. If f is positive definite, then $f(0) \geq 0$ (as before part of the assertion is that $f(0)$ is real). Indeed, (2.8) with $n = 1$ gives $c\bar{c}f(0) = |c|^2 f(0) \geq 0$.

2. $f(t) = \overline{f(-t)}$. In particular this means that a real positive definite function must be even, i.e. it must satisfy $f(t) = f(-t)$. To prove this assertion apply (2.8) with $n = 2$, $t_1 = 0$, $t_2 = t$, $c_1 = c_2 = 1$ to obtain

$$2f(0) + f(t) + f(-t) \geq 0$$

which implies[2] that $f(t) + f(-t)$ is real, hence $\Im f(t) + \Im f(-t) = 0$, or

$$\Im f(t) = -\Im f(-t) \tag{2.10}$$

If we choose $c_1 = 1$, $c_2 = i$ we obtain

$$f(0) + if(t) - if(-t) - f(0) \geq 0$$

which implies that $f(t) - f(-t)$ is pure imaginary, hence $\Re f(t) - \Re f(-t) = 0$, or

$$\Re f(t) = \Re f(-t). \tag{2.11}$$

Equations (2.10) and (2.11) together establish that

$$f(t) = \overline{f(-t)}. \tag{2.12}$$

3. $|f(t)| \leq f(0)$ for every $t \in \mathbb{R}$. To show this, take $c_1 = f(t)$, $c_2 = -|f(t)|$ to obtain $2f(0)|f(t)|^2 - 2|f(t)|^3 \geq 0$, whence the inequality follows.

4. Any positive definite function for which $f(0) = 1$, satisfies the following inequality:

$$|f(t+h) - f(t)| \leq 2|1 - f(h)|^2. \tag{2.13}$$

(The normalizing assumption $f(0) = 1$ simplifies the algebra without harming the generality of the statement.) The importance of this inequality lies in the fact that it implies that if a positive definite function is continuous at 0 then it must be continuous (and in fact uniformly

[2] If $c = a + ib$ is a complex number (a, $b \in \mathbb{R}$) we denote its real part by $\Re c = a$ and its imaginary part by $\Im c = b$)

continuous) on $\mathbb{R}$. We have already seen this for characteristic functions. To prove this assertion, we will use the positive definiteness of the matrix

$$\begin{bmatrix} 1 & f(-t) & f(-t-h) \\ f(t) & 1 & f(-h) \\ f(t+h) & f(h) & 1 \end{bmatrix} \tag{2.14}$$

which is obtained from (2.9) with $n = 3$, $t_1 = 0$, $t_2 = t$, $t_3 = t + h$ and $f(0) = 1$. (2.14) is positive definite if

$$1 - |f(t)|^2 \geq 0$$

$$1 + f(-t)f(-h)f(t+h) + f(t)f(h)f(-t-h) - |f(h)|^2 - |f(t)|^2 - |f(t+h)|^2 \geq 0.$$

Making use of (2.12), this last inequality can be rewritten as

$$1 + f(t)f(h)\overline{f(t+h)} + \overline{f(t)f(h)}f(t+h) - |f(t)|^2 - |f(h)|^2 - |f(t+h)|^2 \geq 0$$

or

$$1 + 2\Re\{f(t)f(h)\overline{f(t+h)}\} - |f(t)|^2 - |f(h)|^2 - |f(t+h)|^2 \geq 0,$$

which gives

$$|f(t+h)|^2 + |f(t)|^2 \leq 1 - |f(h)|^2 + 2\Re\{f(t)f(h)\overline{f(t+h)}\}.$$

We are now ready to show (2.13)

$$\begin{aligned} |f(t) - f(t+h)|^2 &= |f(t)|^2 + |f(t+h)|^2 - f(t)\overline{f(t+h)} - \overline{f(t)}f(t+h) \\ &= |f(t)|^2 + |f(t+h)|^2 - 2\Re\{f(t)\overline{f(t+h)}\} \\ &\leq 1 - |f(h)|^2 + 2\Re\{f(t)f(h)\overline{f(t+h)}\} - 2\Re\{f(t)\overline{f(t+h)}\} \\ &= 1 - |f(h)|^2 + 2\Re\{f(t)\overline{f(t+h)}[f(h) - 1]\} \\ &\leq 1 - |f(h)|^2 + 2|1 - f(h)| \end{aligned} \tag{2.15}$$

where in this last inequality we have used the fact that

$$\begin{aligned} \Re\{f(t)\overline{f(t+h)}[f(h) - 1]\} &\leq |f(t)\overline{f(t+h)}[f(t) - 1]| \leq |f(t)|\,|f(t+h)|\,|1 - f(t)| \\ &\leq |1 - f(t)| \end{aligned}$$

(since $|f(t)| \leq f(0) = 1$). Finally, note that $1 - |f(h)| = |1 - |f(h)|| \leq |1 - f(h)|$ and hence (2.15) gives

$$\begin{aligned} |f(t) - f(t+h)|^2 &\leq (1 - |f(h)|)(1 + |f(h)|) + 2|1 - f(h)| \leq |1 - f(h)|\,(1 + |f(h)| + 2) \\ &\leq 4|1 - f(h)| \end{aligned}$$

As we have seen, characteristic functions of probability measures are positive definite and positive definite functions that are continuous at zero have the same properties as characteristic functions. This is far from accidental. In fact as the next theorem shows these two classes of functions coincide.

Theorem 15 (Bochner). *Suppose that a function $f : \mathbb{R} \to \mathbb{C}$ is positive definite with $f(0) = 1$ and continuous at 0. Then there exists a probability distribution F on $\mathbb{R}$ such that $f(t) = \int_{\mathbb{R}} e^{itx} dF(x)$.*

Proof: Fix $T > 0$ and consider the function

$$p_T(x) = \frac{1}{T^2} \int_0^T \int_0^T f(t-s) e^{itx} e^{-isx} ds dt \geq 0. \tag{2.16}$$

It is clear that $p_T(x)$ is real and nonnegative since the double integral is the limit of Riemann sums $\sum_j \sum_k f(t_j - s_k) e^{it_j x} e^{-is_k x} \Delta t_j \Delta s_k$ which are nonnegative by positive definiteness. Changing variables in (2.16) gives

$$p_T(x) = \int_{-T}^T \left(1 - \frac{|t|}{T}\right) f(t) e^{-itx} dt. \tag{2.17}$$

Remark: For any $L > 0$,

$$\begin{aligned} \frac{1}{L} \int_0^L dy \int_{-y}^y e^{itx} dx &= \frac{1}{L} \int_0^L dy \int_0^y 2 \cos tx \; dx = \frac{2}{Lt} \int_0^L \sin ty \; dy \\ &= 2 \frac{1 - \cos Lt}{Lt^2} \end{aligned}$$

Hence,

$$\begin{aligned} \frac{1}{L} \int_0^L dy \int_{-y}^y p_T(x) dx &= 2 \frac{1}{2\pi} \int_{-T}^T \left(1 - \frac{|t|}{T}\right) \frac{1 - \cos Lt}{Lt^2} f(t) dt \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left(1 - \frac{|t|}{T}\right)^+ \frac{1 - \cos Lt}{Lt^2} f(t) dt \end{aligned}$$

Define

$$f_T(t) := \left(1 - \frac{|t|}{T}\right)^+ f(t) \tag{2.18}$$

and note that $|f_T(t)| \leq 1$. Also, note that $\left| \int_{-\infty}^{\infty} \frac{1 - \cos t}{t^2} dt \right| < \infty$. Hence, by dominated convergence,

$$\lim_{L \to \infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f_T(t/L) \frac{1 - \cos t}{t^2} dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos t}{t^2} dt = 1.$$

We thus have that $\int_{-y}^y p_T(x) dx \uparrow 1$ as $y \uparrow \infty$. (Here we are using the following result: If g is an increasing function and $\frac{1}{X} \int_0^X g(x) dx \to a$ as $X \to \infty$, then $g(x) \to a$ as $x \to \infty$. $\int_{-y}^y p_T(x) dx$ must be an increasing function of y, since $p_T(x) \geq 0$ for all x. We have thus shown that $p_T(x)$ is integrable with $\int_{-\infty}^{\infty} p_T(x) dx = 1$, hence $p_T(x)$ is the probability density of some distribution. We also have

$$p_T(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(1 - \frac{|t|}{T}\right)^+ f(t) e^{-itx} dx$$

Hence, for each T, $f_T(t)$ is the characteristic function of some distribution function. As $T \to \infty$,

$$f_T(t) = \left(1 - \frac{|t|}{T}\right)^+ f(t) \to f(t) \qquad \text{for all } t \in \mathbb{R}$$

and $f(t)$ is by assumption continuous at 0. Therefore the convergence theorem of the previous section guarrantees that $f(t)$ must also be a characteristic function ♠

2.4 Second Order Stationary Processes

Let $\{X_t; t \in \mathbb{R}\}$ a stochastic process with $EX_t = \mu(t)$. Define $\xi_t := X_t - \mu(t)$. The function $R(s,t) := E\xi_s\xi_t$ is called the *covariance function* of the process X. A process X is called *Gaussian* if, for any $n \in \mathbb{N}$ and any $t_1, t_2, \ldots, t_n$, $(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \mathrm{Var}(X_{t_1}) & \mathrm{Cov}(X_{t_1}, X_{t_2}) & \cdots & \mathrm{Cov}(X_{t_1}, X_{t_n}) \\ \mathrm{Cov}(X_{t_2}, X_{t_1}) & \mathrm{Var}(X_{t_2}) & \cdots & \mathrm{Cov}(X_{t_1}, X_{t_n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_{t_n}, X_{t_1}) & \mathrm{Cov}(X_{t_n}, X_{t_2}) & \cdots & \mathrm{Var}(X_{t_n}, X_{t_1}) \end{bmatrix}.$$

A process $\{X_t; t \in \mathbb{R}\}$ is a second order stationary process if

i) $EX_t = \mu$, $E(X_t - \mu)^2 = \sigma^2 < \infty$, for all $t \in \mathbb{R}$,

ii) There exists a real function $r : \mathbb{R} \longrightarrow \mathbb{R}$ such that $E(X_s - \mu)(X_t - \mu) = r(t-s)$ for all $s, t \in \mathbb{R}$,

iii) The process is mean-square continuous, i.e.

$$\lim_{h \to 0} E(X_{t+h} - X_t)^2 = 0$$

Let us now examine some of the consequences of the above properties. First, since $\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t)$ if follows that $r(0) = \sigma^2$. Also, $E(X_{t+h} - X_t)^2 = \sigma^2 + \sigma^2 - 2r(h) = 2(r(h) - r(0))$ and hence, property *iii)* is equivalent to the requirement that r be continuous at zero, i.e. $r(h) \to r(0)$ as $h \to 0$. Cauchy–Schwartz implies $|r(t)| \leq \sigma^2 \ \forall t$. Without loss of generality suppose $\sigma = 1$.

Theorem 16. *$r(t)$ is a positive definite function.*

Proof:

$$\begin{aligned} 0 &\leq \left|\sum_{j=1}^n X_{t_j} z_j\right|^2 = \overline{\left(\sum_{j=1}^n X_{t_j} z_j\right)} \left(\sum_{i=1}^n X_{t_i} z_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n z_i \overline{z}_j X_{t_i} X_{t_j}. \end{aligned}$$

Taking expectations we have

$$0 \leq \sum_{i,j} z_i \overline{z}_j E X_{t_i} X_{t_j} = \sum_{i,j} r(t_j - t_i) z_i \overline{z}_j$$

This establishes that the covariance function of any stationary second order process is positive definite. Hence Bochner's theorem guarantees that there exists a uniquely determined probability measure on $\mathbb{R}$ such that

$$r(t) = \int_{-\infty}^{\infty} e^{itx} R(dx)$$

In particular, since r is an even function,

$$r(t) = 2\int_{0}^{\infty} \cos(tx) R(dx) \tag{2.19}$$

2.4.1 An example of a stationary second order process

Let $N(t)$ be a Poisson process with rate λ and $X(0)$ a random variable with $P(X(0) = 1) = P(X(0) = -1) = 1/2$, independent of the Poisson process. Consider the process

$$X(t) = X(0)(-1)^{N(t)} \tag{2.20}$$

Clearly, X alternates between the values 1, and -1, changing value at each Poisson point. It is easy to see that $EX(t) = EX(0)E(-1)^{N(t)} = 0$ (since $EX(0) = 0$). The covariance function is easily computed as follows:

$$EX(t)X(t+s) = E\left[X(0)^2(-1)^{N(t)N(t+s)}\right] = E\left[(-1)^{N(t+s)-N(t)}\right]$$

Using the stationary increments property the above expectation is

$$\begin{aligned} E(-1)^{N(s)} &= \sum_{n=0}^{\infty} \frac{(\lambda s)^{2n}}{(2n)!} e^{-\lambda s} - \sum_{n=0}^{\infty} \frac{(\lambda s)^{2n+1}}{(2n+1)!} e^{-\lambda s} \\ &= \cosh(\lambda s) e^{-\lambda s} - \sinh(\lambda s) e^{-\lambda s} = e^{-2\lambda s}. \end{aligned}$$

Hence the covariance function is given by

$$EX(t)X(s) = r(t-s) = e^{-2\lambda|t-s|}.$$

The spectral measure can be easily computed in this case: We must have $r(t) = e^{-2\lambda|t|} = \int_{-\infty}^{\infty} e^{itx} R(dx)$ and hence

$$R(dx) = \frac{1}{2\pi\lambda} \frac{1}{1 + (x/2\lambda)^2} dx.$$

Chapter 3

Infinitely Divisible Laws

Let X be a real random variable with characteristic function $\phi(t) := Ee^{itX}$ and distribution function $F(x) = P(X \leq x)$. We say that X is infinitely divisible (or, equivalently, that F, or ϕ are infinitely divisible) if, for every integer n there exist independent, identically distrtibuted random variables X_i, $i = 1, 2, \ldots, n$, such that

$$X \stackrel{\mathrm{d}}{=} X_1 + X_2 + \cdots + X_n$$

Equivalently we say that the characteristic function $\phi(t)$ is infinitely divisible if, for every n, there exists a *characteristic function* $\phi_n(t)$ such that

$$\phi(t) = (\phi_n(t))^n \tag{3.1}$$

A few examples will convince us that this definition is not vacuous: If X is $\mathcal{N}(0, 1)$ then for each n it can be expressed as a sum of n independent normal $\mathcal{N}(0, \frac{1}{n})$ r.v's. For another example consider a Gamma distributed r.v. with shape parameter α and scale parameter β and corresponding characteristic function $\phi(t) = \left(\frac{\beta}{\beta - it}\right)^{\alpha}$. Since $\left(\frac{\beta}{\beta - it}\right)^{\alpha/n}$ is also a characteristic function (of a Gamma distribution with shape parameter α/n and scale parameter again β) we see that (3.1) is satisfied, hence the Gamma distribution is infinitely divisible.

Theorem 17. *The characteristic function of an i.d. r.v. does not vanish for any real t.*

Theorem 18. *The distribution function of a sum of independent r.v.'s having infinitely divisible distribution function is also infinitely divisible*

Proof: Let X_i, $i = 1, 2, \ldots, k$ be independent r.v.'s with infinitely divisible characteristic functions $\phi_i(t)$. Set $X = X_1 + \cdots + X_k$ and denote by $\phi(t) = Ee^{itx}$ its characteristic function. Clearly $\phi(t) = \phi_1(t)\phi_2(t)\cdots\phi_k(t)$. Since X_i is infinitely divisible, for every integer n $\phi_i^{1/n}(t)$ is also a characteristic function. Hence, since the product of characteristic functions is also a characteristic function, $\phi^{1/n}(t) = \phi_1^{1/n}(t)\phi_2^{1/n}(t)\cdots\phi_k^{1/n}(t)$ is a characteristic function and $\phi(t)$ is infinitely divisible.

Theorem 19. *The limit distribution function of a sequence of infinitely divisible distribution functions is itself infinitely divisible, i.e. if $F_n(x)$ is a sequence of inifinitely divisible distribution functions such that $F_n(x) \to F(x)$ for all continuity points of the distribution function F, then F is infinitely divisible.*

Proof: Let $\phi_n(t)$ denote the characteristic function of $F_n(x)$ and $\phi(t)$ the ch.f. of $F(x)$. By the convergence theorem we know that $\phi_n(t) \to \phi(t)$ for all t, and in fact this convergence is uniform in t. Since F_n is infinitely divisible, $\phi_n^{1/k}$ is also a characteristic function, and by the continuity of the square root

Theorem 20. ***[Lévy–Khinchine Representation]*** *A distribution function F with finite variance (and corresponding characteristic function $\phi(t)$) is infinitely divisible if and only if it has the representation*

$$\log \phi(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - itx\right) \frac{1}{x^2} dG(x) \tag{3.2}$$

where γ is a real constant and G a nondecreasing function of bounded variation.

Proof: Suppose that $\phi(t)$ is i.d. Then from (3.1) for any n we have

$$\log \phi(t) = n \log \phi_n(t) = n \log \left(1 + \phi_n(t) - 1\right)$$

However, for any $T > 0$, as $n \to \infty$, $\phi_n(t) \to 1$ uniformly in $|t| < T$. Hence we can write $\log\left(1 + \phi_n(t) - 1\right) = \left(\phi_n(t) - 1\right)\left(1 + \epsilon_n\right)$ where $\epsilon_n \to 0$ as $n \to \infty$ and

$$\log \phi(t) = n \left(\phi_n(t) - 1\right)\left(1 + \epsilon_n\right)$$

Denote by F_n the distribution function corresponding to the characteristic function ϕ_n. We then have

$$\phi_n(t) - 1 = \int_{\mathbb{R}} \left(e^{itx} - 1\right) dF_n(x)$$

Also,

$$n \int_{\mathbb{R}} x dF_n(x) = \int_{\mathbb{R}} x dF(x) = \gamma.$$

Hence

$$\log \phi(t) = i\gamma t + \lim_{n\to\infty} n \int_{\mathbb{R}} \left(e^{itx} - 1 - itx\right) dF_n(x) \tag{3.3}$$

Set

$$G_n(x) := n \int_{-\infty}^{x} u^2 dF_n(u)$$

Then $\{G_n(x)\}$ is a sequence of increasing functions. Also $\{G_n(\infty)\}$ is bounded. Indeed,

$$n \int_{-\infty}^{\infty} u^2 dF_n(u) = \sigma^2 + \frac{1}{n}\gamma^2$$

With these definitions,

$$\log \phi(t) = i\gamma t + \lim_{n\to\infty} \int_{\mathbb{R}} \frac{e^{itx} - 1 - itx}{x^2} dG_n(x).$$

Helly's first theorem asserts that there exists a subsequence n_k and an increasing function $G(x)$ such that $G_{n_k}(x) \to G(x)$ for all continuity points of $G(x)$. On the other hand $\left|\frac{e^{itx}-1-itx}{x^2}\right| \leq \frac{t^2}{2}$

$$\log \phi(t) = i\gamma t + \int_{\mathbb{R}} \frac{e^{itx} - 1 - itx}{x^2} dG(x).$$

Uniquness of specification: By differentiating twice the above relationship with respect to t we see that

$$-\frac{d^2}{dt^2} \log \phi(t) = \int_{\mathbb{R}} e^{itx} dG(x).$$

Hence, to a given $\phi(t)$ there corresponds a unique function $G(x)$ by the uniqueness theorem for characteristic functions.

3.1 Examples of infinitely divisible distributions

The following infinitely divisible distributions are described by means of their characteristic functions

- The deterministic distribution e^{ita}
- The Normal distribution $e^{-\sigma^2 t^2}$
- The gamma distribution $\frac{1}{(1-it)^\alpha}$, $\alpha > 0$
- The Poisson distribution $e^{-\lambda(1-e^{it})}$
- The compound Poisson distribution $e^{-\lambda(1-\psi(t))}$ where $\psi(t)$ is the characteristic function of some random variable
- The symmetric stable distribution of exponent α: $e^{-c|t|^\alpha}$ where $0 < \alpha < 2$ and $c > 0$.

In the next example we compute the characteristic measure for the gamma distribution: Starting with $\phi(t) = (1-it)^{-\alpha}$ we see that $\frac{d}{dt}\log\phi(t) = \frac{i\alpha}{1-it}$. But $\frac{1}{1-it} = \int_0^\infty e^{itx}e^{-x}dx$ and hence, integrating with respect to t, $\log\phi(t) - \log\phi(0) = i\alpha\int_0^\infty \left[\int_0^t e^{iux}du\right] e^{-x}dx = \alpha\int_0^\infty \left(e^{itx}-1\right)\frac{e^x}{x}dx$. Hence, taking into account that $\int_0^\infty e^{-x}dx = 1$, we have the representation

$$\log\phi(t) = it\alpha + \int_{\mathbb{R}} \left(e^{itx} - 1 - itx\right)\frac{1}{x^2}dG(x) \quad \text{with} \quad dG(x) = \begin{cases} 0 & x < 0 \\ \alpha x e^{-x} & x \geq 0 \end{cases}$$

This corresponds to $G(x) = \alpha\left(1 - (1+x)e^{-x}\right)$ for $x \geq 0$ and $G(x) = 0$ for $x < 0$.

Poisson random variables with non–integer values. Let N be a Poisson random variable with parameter λ: $P(N = k) = \frac{1}{k!}\lambda^k e^{-\lambda}$, $k = 0, 1, 2, \ldots$, and $a > 0$. Then the random

variable $X = aN$ has distribution $\mathcal{P}(\lambda, a)$ given by $P(X = ka) = \frac{1}{k!}\lambda^k e^{-\lambda}$, $k = 0, 1, 2, \ldots$, i.e. it takes values on the integer multiples of a. Its characteristic function is

$$e^{\lambda\left(e^{ita}-1\right)}.$$

The geometric distribution. Let X be a random variable with distribution $P(X = k) = q^k p$, $k = 0, 1, 2, \ldots,$. We will show that X is infinitely divisible. Its characteristic function is given by

$$\phi(t) = E[e^{itX}] = \sum_{k=0}^{\infty} q^k p e^{ikt} = \frac{1-q}{1-qe^{it}}.$$

Since $0 < q < 1$, $\log(1-q)$, and $\log(1-qe^{it})$ are well defined, so

$$\log \phi(t) = \log(1-q) - \log(1-qe^{it}) = \sum_{k=1}^{\infty} \frac{1}{k} q^k \left(e^{ikt} - 1\right).$$

Hence

$$\phi(t) = \prod_{k=1}^{\infty} e^{\frac{1}{k}q^k\left(e^{ikt}-1\right)} =: \prod_{k=1}^{\infty} \varphi_k(t),$$

where, each of the characteristic functions $\varphi_k(t)$ corresponds to a Poisson random variable.

Chapter 4

Homogeneous and Non-Homogeneous Poisson Processes.

In this chapter we discuss the definition and the basic properties of homogeneous and non-homogeneous Poisson process on the real line.

4.1 Counting Processes

A real-valued stochastic process $\{N(t); t \geq 0\}$ is said to be a counting process if it satisfies the following

(i) $N(0) = 0$,

(ii) $N(t)$ is integer valued,

(iii) $N(t)$ is an increasing function of t, i.e. if $s < t$, then $N(s) \leq N(t)$.

We may think of $N(t)$ as the total number of 'events' that have occurred up to time t. In particular, for $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval $(s, t]$. The counting process is said to be simple if it only has jumps of unit size. A counting process has *independent increments* if the number of events that occur in disjoint time intervals are independent. This means in particular that the number of events that have occurred by time t, must be independent of the number of events occurring between times t and $t + s$ (that is, $N(t)$ must be independent of $N(t + s) - N(t)$).

A counting process possesses stationary increments if the distribution of the number of events that occur in any time interval depends only on the length of the interval. In other words,

the process has stationary increments if the number of events in the interval $(t_1 + s, t_2 + s]$ has the same distribution as the number of events in the interval $(t_1, t_2]$ for all $0 < t_1 < t_2$, and $s > 0$.

4.2 Poisson Process

The Poisson process is the simplest and most widely used counting process for modelling purposes in numerous practical applications, e.g. to model arrival processes for queuing systems or demand processes for inventory systems. It is empirically found that in a wide variety of circumstances a Poisson process can represent quite adequately the occurrence of events in time or points in space. An explanation of this fact can be based on limit theorems that establish that in the situation of many individual events, each having a small probability of occurrence, the actual number of events occurring approximately follows a Poisson distribution. Of course the simplest and best known among them is the theorem that establishes the convergence of the binomial distribution to the Poisson distribution when the number of binomial trials goes to infinity while the probability of success goes to zero in such a fashion that their product converge to a strictly positive number.

There are several equivalent definitions of the Poisson process (e.g. see Tijms, 1986). We give the following

Definition 21. *The counting process* $\{N(t); t \geq 0\}$ *is a Poisson process with rate* $\lambda > 0$, *if*

(i) $N(0) = 0$.

(ii) $N(t)$ is piece-wise constant with unit jumps.

(iii) The process has independent increments.

(iv) The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all s, $t \geq 0$,

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, \ldots \tag{4.1}$$

Note that condition (iv) implies that the Poisson process has stationary increments and also that $E[N(t)] = \lambda t$, which explains why λ is called the rate of the process.

Recall the Landau $o(\cdot)$ notation: The function f is said to be $o(h)$ in the vicinity of 0 if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0 \ . \tag{4.2}$$

Theorem 22. *The counting process* $\{N(t); t \geq 0\}$ *is a Poisson process with rate* $\lambda \geq 0$, *if*

(i) $N(0) = 0$.

(ii) *The process has stationary and independent increments.*

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(iv) $P\{N(h) \geq 2\} = o(h)$.

For the proof of the above we refer the reader to any text on stochastic processes, e.g. (Tijms, 1986).

4.3 The Memoryless Property of the Poisson Process

Next we discuss a memoryless property that characterizes the Poisson process (see Tijms, 1986). For each $t > 0$, define the *residual life* variable γ_t, which is the amount of time that elapses from t until the next arrival. For any t we have

$$P\{\gamma_t \leq u\} = 1 - e^{-\lambda u}, \quad u \geq 0, \tag{4.3}$$

that is, at each point in time the waiting time until the next arrival has the same exponential distribution as the original interarrival time, regardless of how long it has been since the last arrival occurred. The Poisson process is the only renewal process with this memoryless property. The lack of memory of the Poisson process explains the mathematical tractability of this process, since in specific problems the analysis does not require a state variable indicating the time elapsed since the last arrival.

4.4 Interarrival and Waiting Time Distribution

Consider a Poisson process, and let X_1 denote the time of the first event. Further, for $n \geq 1$, let X_n denote the interarrival time between the $(n-1)^{th}$ and the n^{th} event. $\{X_n; n \geq 1\}$ is called the *sequence of interarrival times.*

We shall now determine the distribution of X_n. To do so we first note that the event $\{X_1 > t\}$ takes place if and only if no events of the Poisson process occur in the interval $[0, t]$ and thus

$$P\{X_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}.$$

Hence, X_1 has an exponential distribution with mean $\frac{1}{\lambda}$. To obtain the distribution of X_2 condition on X_1. This gives

$$\begin{aligned} P\{X_2 > t \mid X_1 = s\} &= P\{0 \text{ events in } (s, s+t] \mid X_1 = s\} \\ &= P\{0 \text{ events in } (s, s+t]\} = e^{-\lambda t} \end{aligned}$$

Therefore, from the above we conclude that X_2 is also an exponential random variable with mean $\frac{1}{\lambda}$, and furthermore, that X_2 is independent of X_1. Repeating the same argument yields the following

Proposition 23. *The interarrival times of a Poisson process X_n, $n = 1, 2$, are independent identically distributed exponential random variables having mean $\frac{1}{\lambda}$.*

The proposition should not surprise us. The assumption of stationary and independent increments is equivalent to asserting that, at any point in time, the process probabilistically restarts itself. That is, the process from any point on is independent of all events that have occurred previously (by the independent increments property), and also has the same distribution as the original process (by the stationarity of increments). In other words, the process has no memory, and hence exponential interarrival times are to be expected.

Another quantity of interest is S_n, the occurrence time of the nth event, also called the waiting time until the n^{th} event. Since

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \geq 1$$

it is easy to show, using moment generating functions, that Proposition 1 implies that S_n has a gamma distribution with parameters n and λ. That is, its probability density is

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \qquad t \geq 0. \tag{4.4}$$

The above could also have been derived by noting that the n^{th} event occurs prior to, or at, time t if and only if the number of events occurring by time t is at least n, that is

$$N(t) \geq n \Leftrightarrow S_n \leq t.$$

Hence

$$P\{S_n \leq t\} = P\{N(t) \geq n\} = \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} = 1 - \sum_{j=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!}$$

which upon differentiation shows that the density function of S_n is given by (4.4).

4.5 Simulation of Homogeneous Poisson Process

Suppose that we wanted to generate the first n event times of a Poisson process with rate λ. To do so we make use of the result that the times between successive events for such a process are independent, exponential random variables each with rate λ. Thus, one way to generate the process is to generate these interarrival times. So if we generate n random numbers, uniform

in $[0,1]$ and independent, say $U_1, U_2, \ldots, U_n$, and set $X_i = -\frac{1}{\lambda}\log U_i$, then X_i can be regarded as the time between the $(i-1)$th and the ith event of the Poisson process. Since the actual time of the jth event will equal the sum of the first j interarrival times, it thus follows that the generated values of the first n event times are

$$S_j = \sum_{i=1}^{j} X_i, \qquad j = 1, \ldots, n.$$

If we want to generate the first T time units of the Poisson process, we can follow the above procedure of successively generating the interarrival times, stopping when their sum exceeds T. The following algorithm can thus be used to generate all the event times occurring in $(0,T)$ of a Poisson process having rate λ. In the algorithm t refers to time, N is the number of events that have occurred by time t, and $S(N)$ is the most recent event time.

Generating the first T times units of a Poisson process with rate λ

STEP 1: $t = 0$, $N = 0$.

STEP 2: Generate a random number U.

STEP 3: $t = t - \frac{1}{\lambda}\log U$. If $t < T$, stop.

STEP 4: $N = N + 1$, $S(N) = t$.

STEP 5: Go to step 2.

The final value of N in the above algorithm will represent the number of events that occur by time T, and the values $S(1), S(2), \ldots, S(N)$ will be the N event times in increasing order. This algorithm can be easily modified to generate any arrival process where the interarrival times are i.i.d. random variables, whether or not they are exponential.

There is another approach for simulating the first T time units of a Poisson process that is also quite efficient (see Ross, 1997) and has the added advantage that it generalizes to Poisson processes in higher dimensions. It starts by generating $N(T)$, the total number of events that occur by time T, and then makes use of a property of the Poisson process according to which, given $N(T)$, the times at which these events occur are distributed independently and uniformly over $(0,T)$ (see Ross, 1996). Hence we can start by generating the value of $N(T)$, a Poisson random variable with mean λT. If the generated value of $N(T)$ is n, we then generate n uniform random numbers in $[0,1]$-call them $U_1, U_2, \ldots, U_n$ and, as TU_i will be uniformly distributed over $(0,T)$, the set of event times will thus be $\{TU_1, \ldots, TU_n\}$. If we were to stop here, this approach would certainly be more efficient than simulating the exponentially distributed interarrival times. However, we usually desire the event times in increasing order (e.g. so as to be able to know $N(s)$ for all $s < T$); thus, we would also need to sort the values TU_i, $i = 1, \ldots, n$ in ascending order.

4.6 Stationary Processes

A stochastic process $\{X(t); t \geq 0\}$ is said to be stationary (see Karlin and Taylor, 1975) if for all $n, s, t_1, \ldots, t_n$, the random variables $X(t_1), \ldots, X(t_n)$ and $X(t_1 + s), \ldots, X(t_n + s)$ have the same joint distribution. In other words, a process is stationary if choosing any fixed point as the origin, the ensuing process has the same probability law. Some examples of stationary processes are

1. A continuous-time Markov chain $\{X(t); t \geq 0\}$ with countable state space, $\mathcal{S}$, where the initial distribution is chosen to be equal to the equilibrium distribution of the chain, $P\{X(0) = j\} = P_j$, $j \in \mathcal{S}$.
2. $\{X(t); t \geq 0\}$ where $X(t)$ is the age at time t of an equilibrium renewal process,
3. $\{X(t); t \geq 0\}$ where $X(t) = N(t + L) - N(t)$, $t \geq 0$, with $L > 0$ a fixed constant, and $\{N(t), t \geq 0\}$ a Poisson process with rate λ.

The first two of the above processes are stationary for the same reason: they are Markov processes whose initial state is chosen according to the limiting state distribution. That the third example, where $X(t)$ represents the number of events of a Poisson process that occur between t and $t + L$, is stationary follows from the stationary and independent increment property of the Poisson process.

In a number of situations, when the condition for a process to be stationary appears rather stringent, one may define the process $\{X(t), t \geq 0\}$ to be *second-order stationary*, or *stationary in the wide sense* if $E[X(t)] = c$ and $\text{Cov}(X(t), X(t + s))$ does not depend on t. That is, a process is second-order stationary if the first two moments of $X(t)$ do not depend on t and the covariance between $X(s)$ and $X(t)$ depends only on $|t - s|$ i.e. $\text{Cov}(X(t), X(s + t)) = R(s)$. Clearly, every stationary process with finite second moments is also second-order stationary. The converse is of course not true in general. However, it is true for Gaussian processes since, in that case, the finite-dimensional distributions are determined by their means and covariances. Thus, it follows that a second-order stationary Gaussian process is stationary.

4.7 Nonhomogeneous Poisson Processes

From a modelling point of view a major restriction in using the ordinary Poisson process is the assumption that in intervals of the same length we have the same distribution for the number of events that occur, regardless of the position of the interval in the real line. A generalization which relaxes this assumption leads to the so-called nonhomogeneous or non-stationary process Poisson process (see Ross, 1996).

Definition 24. *The counting process* $\{N(t); t \geq 0\}$ *is said to be a nonhomogeneous Poisson process with intensity function* $\lambda(t)$, $t \geq 0$, *if*

(i) $N(0) = 0$.

(ii) $\{N(t); t \geq 0\}$ *has independent increments.*

(iii) $P\{N(t+h) - N(t) \geq 2\} = o(h)$.

(iv) $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$.

The quantity $\lambda(t)$, called the intensity at time t, indicates how likely it is that an event will occur (approximately) at time t. (see *iv* in the definition above). The function

$$\Lambda(t) := \int_0^t \lambda(s)ds \tag{4.5}$$

is called the mean-value function. Then the following holds.

Theorem 25. *If* $\{N(t); t \geq 0\}$ *is a non-stationary Poisson process with mean-value function* $\Lambda(t)$, *then* $N(t+s) - N(t)$ *is a Poisson random variable with mean* $\Lambda(t+s) - \Lambda(t)$*:*

$$P\{N(t+s) - N(t) = n\} = \frac{e^{-\{\Lambda(t+s)-\Lambda(t)\}}\,[\Lambda(t+s) - \Lambda(t)]^n}{n!}, \quad n \geq 0. \tag{4.6}$$

Proof The proof of (4.6) is along the lines of the proof of Theorem 1 with a slight modification: Fix t and define

$$P_n(s) = P\{N(t+s) - N(t) = n\} \tag{4.7}$$

Then,

$$\begin{aligned} P_0(s+h) &= P\{N(t+s+h) - N(t) = 0\} \\ &= P\{0 \text{ events in } (t, t+h), 0 \text{ events in } (t+s, t+s+h)\} \\ &= P\{0 \text{ events in } (t, t+s)\}P\{0 \text{ events in } (t+s, t+s+h)\} \\ &= P_0(s)\,[1 - \lambda(t+s)\,h + o(h)] \end{aligned}$$

where the next-to-last equality follows from (ii) and the last from (iii) and (iv). Hence,

$$\frac{P_0(s+h) - P_0(s)}{h} = -\lambda(t+s)\,P_0(s) + \frac{o(h)}{h}$$

Letting $h \to 0$ yields

$$\frac{d}{ds}P_0(s) = -\lambda(t+s)\,P_0(s)$$

from which, taking into account that $P_0(0) = 1$, we obtain

$$\log P_0(s) = -\int_0^s \lambda(t+u)\,du$$

or

$$P_0(s) = e^{-(\Lambda(t+s)-\Lambda(t))}.$$

The remainder of the verification of (4.6) follows similarly. ∎

An alternative argument, which will be useful in the sequel, is based on the generating functions. The probabilities $P_n(s)$ defined in (4.7) satisfy the system of differential equations

$$\begin{aligned}
\frac{d}{ds}P_0(s) &= -\lambda(t+s)\,P_0(s) \\
\frac{d}{ds}P_n(s) &= -\lambda(t+s)\,P_n(s) + \lambda(t+s)\,P_{n-1}(s), \qquad n = 1, 2, \ldots
\end{aligned}$$

Define the probability generating function

$$G(z,s) := \sum_{n=0}^{\infty} z^n P_n(s).$$

Multiplying the nth equation in the above system by z^n and adding from 0 to infinity we obtain

$$\sum_{n=0}^{\infty} z^n \frac{d}{ds}P_n(s) = -\lambda(t+s)\sum_{n=0}^{\infty} z^n P_n(s) + \lambda(t+s)\,z\sum_{n=1}^{\infty} z^{n-1}P_{n-1}(s)$$

or

$$\frac{\partial}{\partial s}G(z,s) = (z-1)\lambda(t+s)\,G(z,s).$$

The solution of this differential equation gives

$$G(z,s) = C(z)e^{-(1-z)\int_0^s \lambda(t+u)du}.$$

The unknown function, $C(z)$, is determined by setting $s = 0$ in the above equation to obtain $G(z,0) = 1 = C(z)$. Thus, taking into account (4.5) we obtain the following expression for the probability generating function,

$$G(z,s) = e^{-(1-z)[\Lambda(t+s)-\Lambda(t)]}$$

which corresponds to a Poisson distribution with parameter $\Lambda(t+s) - \Lambda(t)$.

The importance of the nonhomogeneous Poisson process resides in the fact that we no longer require stationary increments, and so we allow for the possibility that events may be more likely to occur at certain times.

When the intensity function $\lambda(t)$ is bounded, we can think of the nonhomogeneous process as being a random sample from a homogeneous Poisson process. Specifically, let λ be such that $\lambda(t) \leq \lambda$, for all $t \geq 0$ and consider a Poisson process with rate λ. Now if we suppose that an event of the Poisson process that occurs at time t is counted with probability $\frac{\lambda(t)}{\lambda}$, then the process of counted events is a nonhomogeneous Poisson process with intensity function $\lambda(t)$.

This last statement easily follows from Definition 1. For instance (i), (ii), and (iii) follow since they are also true for the homogeneous Poisson process. (iv) follows since

$$\begin{aligned} P\{\text{one event in, } (t,t+h)\} &= P\{\text{one event in, } (t,t+h)\}\frac{\lambda(t)}{\lambda}+o(h) \\ &= \lambda h\frac{\lambda(t)}{\lambda}+o(h) \\ &= \lambda(t)h+o(h). \end{aligned}$$

The simulation of the NHPP is discussed in the following section.

4.8 Simulation of Non Homogeneous Poisson Process

We begin with a discussion of the procedure for simulating a Non Homogeneous Poisson Process NHPP. It is tempting to modify the algorithm used to simulate a homogeneous Poisson process which we have already discussed in chapter 1, to generate t_i given t_{i-1} by substituting $\lambda(t_{i-1})$ in step 2 for λ. However, this would be incorrect, as can be seen from figure 3.1. (This figure might represent traffic arrival rates at an intersection over a 24-hour day.) If $t_{i-1} = 5$, for example, this erroneous "algorithm" would tend to generate a large interarrival time before t_i, since $\lambda(5)$ is low compared with $\lambda(s)$ for s between 6 and 9. Thus, we would miss this upcoming rise in the arrival rate and would not generate the high traffic density associated with the morning rush; indeed if t_i turned out to be 11, we would miss the morning rush altogether.

Care must be taken, then, to generate a Non Homogeneous Poisson Process in a valid way. There are two methods that can be used, one based on a rejection scheme while the other on the inverse transform method

4.8.1 Simulation based on the acceptance-rejection method

A general and simple method proposed by Lewis and Shedler (see Ross, 1997) known as thinning can be used. We present a special case of the thinning algorithm, which works when $\lambda^* = \sup_{0\le s\le T}\lambda(s)$ is finite. We generate a stationary Poisson process with constant rate λ^* and arrival times $\{t_i^*\}$ (by generating exponential interarrival times with rate λ^*, as described in the algorithm for the simulation of a Poisson process with constant rate), then "thin out" the arrival epochs $\{t_i^*\}$ by throwing away (rejecting) each t_i^* as an arrival, with probability $1-\frac{\lambda(t_i^*)}{\lambda^*}$, independently of all others. Thus, we are more likely to accept t_i^* as an arrival if $\lambda(t_i^*)$ is high, yielding the desired property that arrivals will occur more frequently in intervals for which $\lambda(s)$ is high. An equivalent algorithm, in a more convenient recursive form, is as follows (we assumed that t_{i-1} has been validly generated and we want to generate the net arrival time t_i):

1. `Initialize:` $t_0 = 0$, $i = 0$.

2. Set $i = i + 1$, $t = t_{i-1}$.

3. Generate U_1 and U_2, independent, $U(0,1)$, and also independent of all previously generated random variables.

4. Replace t by $t - \frac{1}{\lambda^*}\log(U_1)$.

5. If $U_2 \leq \frac{\lambda(t)}{\lambda^*}$ set $t_i = t$ and return to step 2. Else go to step 3.

If the evaluation of $\lambda(t)$ is slow (which might be the case if, for example, $\lambda(t)$ is a complicated function involving exponential and trigonometric calculations), computation time might be saved in step 5 by adding an acceptance pretest; i.e., the current value for t is accepted as the next arrival time if $U_2 \leq \frac{\lambda_*}{\lambda^*}$, where $\lambda_* = \inf_{0 \leq s \leq T} \lambda(s)$. This would be useful, especially when $\lambda(s)$ is fairly flat.

4.8.2 Simulation based on the inverse transform method

An alternative method for simulating the non-homogeneous Poisson process is based on the inverse transform method. Set

$$\Lambda(t) = \int_0^t \lambda(s)ds, \qquad t \geq 0,$$

and define the inverse function via

$$\Lambda^{-1}(u) = \inf\{t : \Lambda(t) > u\}, \qquad u \geq 0.$$

If $\{S_n; n \in \mathbb{N}\}$ is a unit rate homogeneous Poisson process then

$$T_n = \Lambda^{-1}(S_n), \qquad n \in \mathbb{N},$$

defines a non-homogeneous Poisson process with rate $\lambda(s)$. The usefulness of the above method in practice depends of course on the ease with which one can obtain the inverse function, $\Lambda^{-1}(u)$, in closed form, suitable for computation.

The above suggests the following recursive generation of the consecutive points of the time varying process. If we denote by $X_n := T_n - T_{n-1}$ the inter–event time for the time–varying process and by $Y_n := S_n - S_{n-1}$ the corresponding inter–event time for the unit rate homogeneous process (using the convention $S_0 = T_0 = 0$) we then have $X_1 = \Lambda^{-1}(Y_1)$ and, generally, $T_n = \Lambda^{-1}(S_n)$. Hence

$$X_{n+1} = T_{n+1} - T_n = \Lambda^{-1}(Y_{n+1} + S_n) - \Lambda^{-1}(S_n).$$

In the special case where Λ is absolutely continuous and $\Lambda(t) = \int_0^t \lambda(s)ds$ with $\lambda(s) > 0$ for all $s > 0$ then Λ^{-1} is also absolutely continuous with $\Lambda^{-1}(t) = \int_0^t \frac{1}{\lambda(s)}ds$. In this case

$$X_{n+1} = \int_0^{Y_{n+1}} \frac{ds}{\lambda(s + S_n)}.$$

4.9 The Poisson Process in Euclidean Spaces

In this section we describe how one can define a Poisson process on the Euclidean space, $\mathbb{R}^n$. Let $\mathcal{B}(\mathbb{R}^n)$ denote the Borel subsets of $\mathbb{R}^n$ and let $\mathbf{M}$ denote the set of all $\sigma-$finite point measures on $\mathbb{R}^n$. Let also $\mathcal{M}$ denote the sigma field of subsets of $\mathbf{M}$ generated by the sets $\{m \in \mathbf{M} : m(B) = k\}$ where $k \in \mathbb{N}$ and $B \in \mathcal{B}(\mathbb{R}^n)$. A random point process N of $\mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$ is defined as a measurable mapping $N : (\Omega, \mathcal{F}) \to (\mathbf{M}, \mathcal{M})$. The distribution of N is given by specifying $P(N(B) = k)$ in a consistent fashion. From the above definition it follows that $\{\omega : N(B, \omega) = k\} \in \mathcal{F}$ for all $k \in \mathbb{N}$ and $B \in \mathcal{B}(\mathbb{R}^n)$.

Definition 26. *The random point measure N is a homogeneous Poisson process on $\mathbb{R}^n$ if the following two conditions hold*

1. *For any Borel set $B \in \mathcal{B}(\mathbb{R}^n)$,*

$$P(N(B) = k) = \frac{(\lambda\,|B|)^k}{k!} e^{-\lambda|B|}, \qquad k = 0, 1, 2, \ldots,$$

   where $|B|$ denotes the Lebesgue measure of and $\lambda > 0$.

2. *For any $n \in \mathbb{N}$, if B_i, $i = 1, 2, \ldots, n$ are disjoint Borel sets in $\mathbb{R}^n$ then the random variables $N(B_i)$, $i = 1, 2, \ldots, n$ are independent.*

More generally, let Λ denote a $\sigma-$finite measure on $\mathbb{R}^n$. If

$$P(N(B) = k) = \frac{\Lambda(B)^k}{k!} e^{-\Lambda(B)}, \qquad k = 0, 1, 2, \ldots,$$

for all $B \in \mathcal{B}(\mathbb{R}^n)$, then N is a Poisson random measure on $\mathbb{R}^n$ with mean measure Λ. In particular, if Λ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ then $\Lambda(B) = \int_B \lambda(x)dx$ for some non-negative Borel function λ which will be called stochastic intensity of the non-homogeneous Poisson random measure N.

Chapter 5

Compound Poisson Processes

5.1 The Homogeneous Compound Poisson Process

Compound Poisson processes (both stationary and non-stationary) are extremely useful modelling tools in many areas of application of probability theory, including queuing models, risk theory and insurance mathematics, reliability, finance, and many others. The homogeneous compound Poisson process is a particularly simple stochastic process with independent, stationary increments. Suppose that $\{N(t); t \geq 0\}$ is a stationary Poisson process with rate λ and $\{\xi_i; i = 1, 2, \ldots\}$ an i.i.d. sequence of real-valued random variables, independent from the Poisson process, with common distribution F and corresponding Laplace transform

$$\phi(s) := Ee^{-s\xi} = \int_{-\infty}^{\infty} e^{-sx} dF(x).$$

Set

$$X_t = \sum_{i=1}^{N(t)} \xi_i.$$

The process $\{X_t; t \geq 0\}$ is then called a compound poisson process with poisson rate λ and increment distribution F. It is easy to see that this process has independent, stationary increments with distribution

$$P(X_{t+u} - X_t \leq x) = P(X_u \leq x) = \sum_{n=0}^{\infty} P(N(u) = n) P(\xi_1 + \cdots + \xi_n \leq x).$$

If we denote by $F^{k*} := F * F * \cdots * F$ the $k-$fold convolution of F with itself, then the above can also be written as

$$P(X_{t+u} - X_t \leq x) = \sum_{n=0}^{\infty} \frac{(\lambda u)^n}{n!} e^{-\lambda u} F^{n*}(x).$$

In general, evaluation of the above convolution is not simple. A simpler expression can be obtained for the Laplace transform of the increments as follows

$$\begin{aligned} Ee^{-sX_t} &= Ee^{-s\sum_{i=1}^{N(t)}\xi_i} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}Ee^{-s(\xi_1+\cdots+\xi_n)} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t\phi(s))^n}{n!} \\ &= e^{-\lambda t(1-\phi(s))}. \end{aligned}$$

5.2 The Gamma Process

A useful generalization of the compound Poisson process is the so-called Gamma process described in the sequel. We consider again a Poisson process on the quarter plane $\mathbb{R}^+ \times \mathbb{R}^+$, this time with mean measure given by $\lambda dt \times x^{-1}e^{-vx}dx$, $t \geq 0, x > 0$. Note however that this case cannot be put in the framework of the preceding sections since $\lim_{a\downarrow 0}\int_a^{\infty}\frac{e^{-vx}}{x}dx = \infty$ and thus we cannot define a proper distribution function F for the size of the jumps. Nontheless, the Poisson process in the quarter plane is well defined. One detail that needs to be taken care of, before we proceed is the problem of numbering the points since now, in the strip $D_t := \{(t,x); 0 \leq t \leq T, 0 \leq x\}$ we have a countably infinite number of points with probability 1. Thus we can use the following numbering scheme: divide the strip into parallel stripes according to some sequence that converges to zero such as $\frac{1}{n}$ and then number the points from right to left starting with those in the first stripe and continuing with those in subsequent stripes. This numbering scheme is illustrated in the figure ?? below. We proceed by defining a sequence of processes indexed by $n \in \mathbb{N}$ as follows: The process $\{Z_t^n; t \in [0,T]\}$ is defined by means of

$$Z_t^n = \sum_{i=1}^{\infty}\sigma_i \mathbf{1}(T_i \leq t; \sigma_i > 1/n).$$

Clearly, for each n, Z_t^n is a compound Poisson process with parameters

$$\lambda_n = \lambda\int_{1/n}^{\infty}\frac{e^{-vy}}{y}dy$$

and

$$F_n(x) = \frac{\int_{1/n}^{x}\frac{e^{-vy}}{y}dy}{\int_{1/n}^{\infty}\frac{e^{-vy}}{y}dy}, \qquad x \geq 1/n.$$

The process $\{Z_t^n; t \in [0,T]\}$ is then a process with independent increments and Laplace transform given by

$$\begin{aligned} \log Ee^{-\theta Z_t^n} &= -\lambda_n t\left(1 - \int_0^{\infty}e^{-\theta x}F_n(dx)\right) \\ &= -\lambda t\left(\int_{1/n}^{\infty}\frac{e^{-vx}}{x}dx - \int_{1/n}^{\infty}e^{-\theta x}\frac{e^{-vx}}{x}dx\right) \\ &= -\lambda t\int_{1/n}^{\infty}\left(1 - e^{-\theta x}\right)\frac{e^{-vx}}{x}dx. \end{aligned}$$

Expressed in this fashion, it becomes clear that the limit $\lim_{n\to\infty} \log Ee^{-\theta Z_t^n} := -\Phi(\theta)$ exists and

$$\Phi(\theta) := \lambda t \int_0^\infty \left(1 - e^{-\theta x}\right) \frac{e^{-vx}}{x} dx.$$

This last integral can be computed explicitly by setting $\frac{1}{x} = \int_0^\infty e^{-ux} du$ and using Fubini's theorem to interchange the order of integration as follows

$$\begin{aligned} \Phi(\theta) &= \lambda t \int_0^\infty \left(e^{-vx} - e^{-(\theta+v)x}\right) \frac{1}{x} dx = \lambda t \int_0^\infty \int_0^\infty (e^{-(u+v)x} - e^{-(u+v+\theta)x}) dx du \\ &= \lambda t \int_0^\infty \left(\frac{1}{u+v} - \frac{1}{u+v+\theta}\right) du. \end{aligned}$$

This last integral can be easily computed if we express it as

$$\begin{aligned} \lim_{M\to\infty} \int_0^M \left(\frac{1}{u+v} - \frac{1}{u+v+\theta}\right) du &= \lim_{M\to\infty} \left(\int_0^M \frac{1}{u+v} du - \int_0^M \frac{1}{u+v+\theta} du\right) \\ &= \lim_{M\to\infty} \left(\log \frac{M}{v} - \log \frac{M}{v+\theta}\right) \\ &= \log \frac{v+\theta}{v}. \end{aligned}$$

It is thus clear that for all $t \in [0, T]$, $Z_t^n \uparrow Z_t$ with

$$Ee^{-\theta Z_t} = e^{-\lambda t \log \frac{v+\theta}{v}} = \left(\frac{v}{v+\theta}\right)^{\lambda t}.$$

This justifies the name *Gamma process* since the marginal distributions of this process are Gamma with shape parameter that depends on the time.

5.3 Non-Homogeneous Compound Poisson Process

We can envision a non-homogeneous Poisson point process by assuming that $\{N(t); t \geq 0\}$ is a Poisson process with time varying intensity $\lambda(t)$, $t \geq 0$, in the model of section 1. More generally, however, we would like the distribution of each jump to depend on the time of the occurence of the jump. Suppose that $\{F_t; t \geq 0\}$ is a family of distribution functions that describe the size of the jumps, given their time of occurrence, $\{T_i; i = 1, 2, \ldots\}$. We can define a non-homogeneous compound Poisson process by setting

$$X_t = \sum_{i=1}^{N(t)} \xi_i$$

where

$$P(\xi_1 \leq x_1, \ldots, \xi_n \leq x_n | N(t) = n; T_1 = t_1, \ldots, T_n = t_n) = \prod_{i=1}^n F_{t_i}(x_i). \tag{5.1}$$

We will assume without much loss of generality that the distributions F_t are absolutely continuous and thus we have a family of densities $\{f_t; t \geq 0\}$ indexed by the time of occurence of the

corresponding event. A better way to describe this process is to consider a Poisson point process on the *half–plane* denoted by $N_p = \{N(x,t); (x,t) \in \mathbb{R}_+ \times \mathbb{R}\}$ and assumed to have intensity $\lambda(x,t) \geq 0$ with respect to the Lebesgue measure on $\mathbb{R}^2_+$ for every $(x,t) \in \mathbb{R}^2_+$. The time component of N_p is a Poisson point process $\{N(t); t \geq 0\}$, with intensity $\lambda(t) = \int_0^\infty \lambda(x,t)dx$ which we will assume to be finite for all $t \in \mathbb{R}_+$. Under the above assumptions, for any t there are, with probability 1, a finite number of points of the process and thus counting them from right to left we can label them $\{(T_k, \xi_k); k \in \mathbb{N}\}$. T_k will be interpreted as the kth time of occurrence while ξ_k as the kth jump size. Also, $N(t)$ is the number of jumps that have occurred in the time interval $(0,t]$. In the above framework, the conditional density of a jump, given that it occurs at time t, is given by

$$f_t(y) = \frac{\lambda(y,t)}{\lambda(t)}. \tag{5.2}$$

The conditional distribution functions in (5.1) are then obtained by integration of the above conditional densities. Also, from standard results on non–homogeneous Poisson processes on the real line, the times of occurrence of the jumps in an interval $(0,t]$, conditional on the event $\{N(t) = k\}$, $k > 0$, have the order statistics of k independent random variables with density

$$h_t(s) = \frac{\lambda(s)}{\Lambda(t)}, \quad 0 < s \leq t,$$

where

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

Chapter 6

Collective Risk Theory

6.1 The classical risk process

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

Figure 1: Sample realization of a risk process

We denote by u the initial capital and by $X(t)$ the process $ct - \sum_{k=1}^{N(t)} Z_k$. Let $\Psi(u) = P\{\inf_{0\leq t} u + X_t < 0\}$ denote the probability of ruin at some point in the future and $\Phi(u) = 1 - \Psi(u)$ denote the non-ruin probability.

Lemma 27. *$\Phi(u)$ is non-decreasing in u and $\lim_{u\to\infty} \Phi(u) = 1$.*

Proof: Indeed, $\Phi(u) := P\{\inf_{0\leq t} X(t) \geq -u\}$ which is non-decreasing in u.

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

$$\boxed{\rho := \frac{c}{\alpha\mu} - 1} \qquad \text{or} \qquad \boxed{1 + \rho = \frac{c}{\lambda\mu}} \tag{6.1}$$

6.2 Integrodifferential Equation for the Non–ruin Probability

A first step analysis gives

$$\Phi(u) = E\Phi(u + cS_1 - Z_1) = \int_0^\infty \lambda e^{-\lambda s} \int_0^{u+cs} \Phi(u + cs - z) dF(z) ds \tag{6.2}$$

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

$$\Phi(u) = \frac{\lambda}{c} e^{\lambda u/c} \int_u^\infty e^{-\lambda x/c} \int_0^x \Phi(x - z) dF(z) dx. \tag{6.3}$$

Differentiation of the above with respect to u gives

$$\Phi'(u) = \frac{a}{c}\Phi(u) - \frac{\lambda}{c} \int_0^u \Phi(u - z) dF(z). \tag{6.4}$$

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

$$\Phi(t) - \Phi(0) = \frac{\lambda}{c} \int_0^t \Phi(u) du + \frac{\lambda}{c} \int_0^t \int_0^u \Phi(u - z) \left[1 - F(z)\right] dz du \tag{6.5}$$

which can be rewritten (after integration by parts) as

$$\Phi(u) = \Phi(0) + \frac{\lambda}{c} \int_0^u \Phi(u - z) \left[1 - F(z)\right] dz \tag{6.6}$$

6.3 Exponentially Distributed Claims

Suppose that $F(z) = 1 - e^{-z/\mu}$, i.e. the claim distribution is exponential with mean μ. Then (6.4) becomes

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

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

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

and integrating twice yields

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

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

$$\boxed{\Psi(u) = \frac{1}{1+\rho} e^{-\frac{\rho}{\mu(1+\rho)} u}}. \tag{6.7}$$

Chapter 7

Finite horizon ruin probabilities–A Monte Carlo approach

Here we assume that the times between claims, $\{L_i, i = 1, 2, \ldots\}$ are i.i.d. with distribution G while the claim sizes $\{Z_i, i = 1, 2, \ldots\}$ are also i.i.d. with distribution F. The initial capital is, as usual, denoted by u and the premium rate by c. To simulate this process let $\{U_i, V_i, i = 1, 2, \ldots\}$ be independent random variables, uniformly distributed in $[0, 1]$.

Theorem 28. ***Inverse Transform Method:*** *If U is a random variable, uniformly distributed on $[0, 1]$ and F^{-1} is the inverse function of F then $X := F^{-1}(U)$ is a random variable with distribution F.*

Proof. We will show that X has distribution F i.e. that $P(X \leq x) = F(x)$ for all x. Indeed, $P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$. (Note that $F(x)$ is always a number between 0 and 1.) □

As before we will denote by $Y(t) = ct - \sum_{k=1}^{N(t)} Z_k$ the risk process which gives the capital (risk reserve) of the company at time t and by $\{S_i, i = 1, 2 \ldots\}$ the times at which claims occur. Thus $S_i = L_1 + L_2 + \cdots + L_i$. Let $W^-{}_i = Y(S_i-)$ denote the risk reserve of the company *at the moment right before the i'th claim occurs* and $W^+{}_i$ the risk reserve *right after the i'th claim occurs.* Thus

$$W_i^- = W_{i-1}^+ + cG^{-1}(U_i), \tag{7.1}$$

$$W_i^+ = W_i^- - F^(V_i), \tag{7.2}$$

$$S_i = S_{i-1} + G^{-1}(U_i). \tag{7.3}$$

The above equations, together with the initial conditions $W_0^+ = u$, $S_0 = 0$, allow us to simulate the risk process.

For instance suppose that claims arrive according to a Poisson process with rate λ and that the claim sizes follow a Pareto distribution with density $f(x) = \lambda \frac{c^\alpha}{x^{a+1}}$ for $x > c$. In this case $G(x) = 1 - e^{-\lambda x}$ and thus to obtain L_i we have to solve the equation $G(L_i) = U_i$ or $e^{-\lambda L_i} = 1 - U_i$ which gives[1]

$$L_i = -\frac{1}{\lambda} \log(1 - U_i) \tag{7.4}$$

Similarly for the claim sizes, in order to determine F^{-1} we have to solve the equation $F(V_i) = Z_i$ which, in the Pareto case becomes $\left(\frac{c}{Z_i}\right)^\alpha = 1 - V_i$ or

$$Z_i = ce^{-\frac{1}{\alpha}\log(1-U_i)}. \tag{7.5}$$

[1] In these notes the function log always designates the *natural logarithm*. Many software packages however still reserve this notation for the decimal logarithm (rarely used anymore) while the natural logarithm is denoted by ln.

Chapter 8

Large Deviations

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

$$\lim_{n\to\infty} P(S_n \geq nx) = 0 \qquad \text{for } x > m \tag{8.1}$$

and similarly that

$$\lim_{n\to\infty} P(S_n \leq nx) = 0 \qquad \text{for } x < m \tag{8.2}$$

One important question is *how fast do the above probabilities go to zero.* We will see that they go to zero exponentially fast, i.e. that

$$P(S_n \geq nx) \asymp e^{-nI(x)} \qquad \text{for } x > m \tag{8.3}$$

In the above formula note that the exponential rate of decay $I(x)$ is a function of x. The meaning of (8.3) is made precise if we state it as

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq nx) = -I(x) \qquad \text{for } x > m. \tag{8.4}$$

Where does the exponential behavior come from? Write $P(S_n \geq nx)$ as $P\left(S_n - nm \geq n(x-m)\right) = P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \geq \sqrt{n}(x-m)\right)$ 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

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

Before we continue we need some results concerning the tail of the standard normal distribution. Define

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

and correspondingly the tail $1 - \Phi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$. Then

$$1 - \Phi(x) \leq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad x > 0, \tag{8.5}$$

and

$$1-\Phi(x) \geq \left(\frac{1}{x}-\frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \qquad x>0. \tag{8.6}$$

The upper bound for the tail (8.5) 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 $x>0$).

To obtain the lower bound (8.6) we start with the following integration by parts formulas

$$\int_x^\infty \frac{3}{u^4}e^{-\frac{1}{2}u^2}du = -\frac{1}{u^3}e^{-\frac{1}{2}u^2}|_x^\infty + \int_x^\infty \frac{3}{u^3}e^{-\frac{1}{2}u^2}du = \frac{1}{x^3}e^{-\frac{1}{2}x^2} + \int_x^\infty \frac{3}{u^3}e^{-\frac{1}{2}u^2}du.$$

and

From (8.5) and (8.6) we have that $\left(\frac{1}{x}-\frac{1}{x^3}\right) \leq \frac{1-\Phi(x)}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}} \leq \frac{1}{x}$. Thus

$$\int_{\sqrt{n}\left(\frac{x-m}{\sigma}\right)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \approx \frac{1}{\frac{\sqrt{n}(x-m)}{\sigma}}\sqrt{2\pi}$$

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 (8.4) becomes an exact relationship

$$P(S_n \geq nx) = \int_{\sqrt{n}(\frac{x-m}{\sigma})}^\infty e^{-\frac{1}{2}u^2}du \tag{8.7}$$

Taking into account the bounds (8.5), (8.6) we have

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

or

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

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

$$\lim_{n\to\infty}\frac{1}{n}\log P(S_n\geq nx) = -\left(\frac{x-m}{\sigma}\right)^2 \tag{8.8}$$

Hence, setting $I(x) = \left(\frac{x-m}{\sigma}\right)^2$ we obtain (8.1) 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.

8.0.1 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 real θ since the exponential is non–negative. Taking expectations in the above inequality we obtain

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

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

$$I(x) := \sup_{\theta \in \mathbb{R}} \{x\theta - \log M(\theta)\}. \tag{8.9}$$

With this definition the Chernoff bound becomes

$$P(S_n \geq nx) \leq e^{-nI(x)} \tag{8.10}$$

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.

Lemma 29. *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

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

Let θ^* be the solution of

$$x - \frac{M'(\theta^*)}{M(\theta^*)} = 0. \tag{8.11}$$

8.1 Examples of rate functions

Bernoulli Random Variables (1) 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 logarigthms $\theta = \log\frac{x}{1-x} + \log\frac{1-p}{p}$. 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) $M(\theta) = \frac{\lambda}{\lambda-\theta}$ 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

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

Binomial (number of trials k, probability of success p) $M(\theta) = (1 - p + pe^{\theta})^k$

Geometric (probability of success p) Here $M(\theta) = \frac{1-p}{1-pe^{\theta}}$. Following the same procedure as before we obtain

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

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

We will next establish Cramér's theorem by showing that

$$\liminf_{n\to\infty} P(S_n \ge nx) \ge -I(x) \qquad \text{for } x > m. \tag{8.12}$$

8.1.1 The twisted distribution

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

$$d\tilde{F}(dx) = \frac{e^{\theta x}}{M(\theta)} F(dx)$$

is called the *twisted distribution* that corresponds to F. (It is easy to see that $\tilde{F}(x) = \int_{-\infty}^{x} \frac{e^{\theta u}}{M(\theta)} F(du)$ is a non–decreasing function of x and as $x \to \infty$, $\tilde{F}(x) \to 1$.

The mean of the twisted distribution is given by $\int_{-\infty}^{\infty} x\tilde{F}(dx) = \frac{1}{M(\theta)} \int_{-\infty}^{\infty} xe^{\theta x} F(dx) = \frac{1}{M(\theta)} \frac{d}{d\theta} \int_{-\infty}^{\infty} e^{\theta x} F(dx) = \frac{M'(\theta)}{M(\theta)}$.

In particular when $\theta = \theta^*$, the solution of (8.11),

$$\frac{1}{M(\theta^*)} \int_{-\infty}^{\infty} xe^{\theta^* x} F(dx) = \frac{M'(\theta^*)}{M(\theta^*)} = x \tag{8.13}$$

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 $\tilde{P}$, 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 $\tilde{E}$.

Start with the inequality

$$P\left(S_n \geq nx\right) \geq P\left(n(x+\epsilon) \geq S_n \geq nx\right) = E\left[\mathbf{1}(n(x+\epsilon) \geq S_n \geq nx)\right] \tag{8.14}$$

$$\begin{aligned} P\left(S_n \geq nx\right) &\geq e^{-n(x+\epsilon)} M(\theta^*)^n E\left[\frac{e^{\theta^* S_n}}{M(\theta^*)^n} \mathbf{1}(n(x+\epsilon) \geq S_n \geq nx)\right] && (8.15) \\ &= e^{-n(x+\epsilon)} M(\theta^*)^n \tilde{P}\left(\sqrt{n}\epsilon \geq \frac{S_n - nx}{\sqrt{n}} \geq 0\right) && (8.16) \end{aligned}$$

The twisted distribution can be used to establish (8.12) as follows: Set $\theta = \theta^*$ so that, under $\tilde{P}$, the mean of X_i is x.

$$\tilde{P}\left(n(x+\epsilon) \geq S_n \geq nx\right) = \tilde{P}\left(\sqrt{n}\epsilon \geq \frac{S_n - nx}{\sqrt{n}} \geq 0\right) \tag{8.17}$$

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

$$\lim_{n\to\infty} \tilde{P}\left(\sqrt{n}\epsilon \geq \frac{S_n - nx}{\sqrt{n}} \geq 0\right) = \frac{1}{2}. \tag{8.18}$$

Hence,

$$\liminf_n \tilde{P}\left(S_n \geq nx\right) \geq \liminf_n \tilde{P}\left(n(x+\epsilon) \geq S_n \geq nx\right) \geq \frac{1}{2} \tag{8.19}$$

Write $\tilde{P}\left(n(x+\epsilon) \geq S_n \geq nx\right) = \tilde{E}[\mathbf{1}(n(x+\epsilon) \geq S_n \geq nx)] = E\left[\frac{e^{\theta^* S_n}}{M(\theta^*)^n}\mathbf{1}(n(x+\epsilon) \geq S_n \geq nx)\right]$

We thus have

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

from which we obtain

$$\begin{aligned} \liminf_n \frac{1}{n}\log P(S_n \geq nx) \quad &\geq \\ &-(x+\epsilon)\theta^* + \log M(\theta^*) + \liminf_n \frac{1}{n}\log \tilde{P}\left(n(x+\epsilon) \geq S_n \geq nx\right) \end{aligned} \tag{8.20}$$

In view of (8.18) 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 $x\theta - \log M(\theta)$.

8.1.2 The principle of the largest term

Suppose that α_i, $i = 1, 2, \ldots, n$ are nonnegative real numbers. Then

$$e^{-\alpha_1 x} + e^{-\alpha_2 x} + \cdots + e^{-\alpha_n x} \asymp e^{-x\min_i \alpha_i}$$

8.1.3 Legendre Transforms

The rate function has been defined as

$$I(x) = \sup_{\theta\in\mathbb{R}}\left[\theta x - \log M(\theta)\right].$$

Define now

$$\lambda(\theta) = \sup_{x\in\mathbb{R}}\left[\theta x - I(x)\right].$$

We will show that $\lambda(\theta) = \log M(\theta)$. Indeed, we have

$$\begin{aligned} \lambda(\theta) &= \sup_{x\in\mathbb{R}}\left[\theta x - \sup_{\eta\in\mathbb{R}}\left[\eta x - \log M(\eta)\right]\right] \\ &= \sup_{x\in\mathbb{R}}\inf_{\eta\in\mathbb{R}}\left[x(\theta-\eta) + \log M(\eta)\right] \\ &= \inf_{\eta\in\mathbb{R}}\sup_{x\in\mathbb{R}}\left[x(\theta-\eta) + \log M(\eta)\right] = \log M(\theta) \end{aligned}$$

The last equality is due to the fact that if $\theta \neq \eta$ then $\sup_{x\in\mathbb{R}}[x(\theta-\eta)+\log M(\eta)] = +\infty$, while on the other hand, if $\theta = \eta$ $\sup_{x\in\mathbb{R}}[x(\theta-\eta)+\log M(\eta)] = \log M(\theta)$.

8.2 Large Deviation Heuristics for the Analysis of Risk Processes

As we have seen, the infinite horizon ruin probability for a risk process is defined as

$$\Psi(u) = \mathbb{P}\left(\sup_{t\geq 0}\left(-ct + \sum_{k=1}^{N_t} Z_k\right) > u\right).$$

Consider a discretization $\{t_i\}$, $i = 0, 1, 2, \ldots$, of $[0, \infty)$, $t_0 = 0$, $t_i = i\delta$, where $\delta > 0$ and set $\xi_i := \sum_{k=N_{t_{i-1}}+1}^{N_{t_i}} Z_k - c\delta$. Note that ξ_i is the sum of all the claims that occur in the interval $(t_{i-1}, t_i]$ from which we subtract $c\delta$, the premium income during this period. (If $N_{t_{i-1}} = N_{t_i}$ then the sum in the definition of ξ_i is empty and its value is zero. Note that

$$\Lambda(\theta) := \mathbb{E}e^{\theta\xi_i} = e^{-c\delta\theta + \lambda\delta(M(\theta)-1)}$$

where $M(\theta) = \mathbb{E}e^{\theta Z_1}$, the moment generating function of the claim distribution. In particular $\mathbb{E}\xi_i = \delta(-c + \lambda\mu)$ where $\mu = \mathbb{E}Z_1$, the mean claim size. We assume that $c > \lambda\mu$ and hence $\mathbb{E}\xi_i < 0$. Then, the ruin probability is approximately (and exactly in the limit, as $\delta \to 0$,)

$$\begin{aligned}\Psi(u,t) &= \mathbb{P}\left(\sup_{j=1,2,\ldots}\sum_{i=1}^{j}\xi_i > u\right) = \mathbb{P}\left(\bigcup_{j=1}^{\infty}\{\sum_{i=1}^{j}\xi_i > u\}\right)\\ &\leq \sum_{j=1}^{\infty}\mathbb{P}\left(\sum_{i=1}^{j}\xi_i > u\right). \qquad (8.21)\end{aligned}$$

From Cramér's theorem we have the asymptotic relationship (for large j)

$$\mathbb{P}\left(\sum_{i=1}^{j}\xi_i > u\right) \asymp e^{-jI(u/j)}$$

where

$$I(x) := \sup_{\theta\in\mathbb{R}}\{\theta x - \log\Lambda(\theta)\} = \sup_{\theta\in\mathbb{R}}\{\theta x + \theta\delta c - \lambda\delta(M(\theta)-1)\}.$$

It can be shown that

$$\Psi(u) \asymp \sum_{j=1}^{\infty} e^{-jI(u/j)} \asymp \max_{1\leq j\leq n} e^{-u\frac{I(u/j)}{u/j}}.$$

From the principle of the largest term it seems plausible that

$$\Psi(u) \asymp e^{-x\inf_{y>0}\frac{I(y)}{y}} = e^{-xr} \qquad (8.22)$$

where

$$r := \inf_{y>0}\frac{I(y)}{y} \qquad (8.23)$$

We can easily express r in terms of $M(\theta)$ as follows:

$$\begin{aligned} & r = \sup\left\{\theta : \theta \leq \frac{I(x)}{x} \quad \forall x\right\} \\ \iff \quad & r = \sup\left\{\theta : \theta x - I(x) \leq 0 \quad \forall x\right\} \\ \iff \quad & r = \sup\left\{\theta : \sup_x \left[\theta x - I(x)\right] \leq 0\right\} \\ \iff \quad & r = \sup\left\{\theta : \log \Lambda(\theta) \leq 0\right\} \end{aligned}$$

Therefore r is the largest root of the equation $\log \Lambda(\theta) = 0$ or equivalently

$$-c\delta\theta + \lambda\delta(M(\theta) - 1) = 0.$$

or

$$M(\theta) - 1 = \frac{c\theta}{\lambda}$$

It can be seen that the above equation has precisely two roots: $\theta = 0$ and $\theta = r > 0$ provided that $\mu = M'(0) < c\lambda^{-1}$.