Twisting in Applied Probability

Tomasz Rolski[1]

Department of Actuarial Mathematics and Statistics, Heriot–Watt University

Edinburgh, May–July 2004

[1]Visit to Heriot–Watt University was supported by EPSRC grant GR/S66732/01

Contents

Chapter I

Perron-Frobenius theorem in probability

Sequences $(x_1, \ldots, x_l)$ are denoted by $(x_j)_{j=1,\ldots,l}$ or simply (x_j). We use the following matrix and vector notations. Matrices, actually we deal with square matrices only, we denote by bold face capital letters, e.g. $\boldsymbol{A} = (a_{ij})_{i,j=1,\ldots,l}$ is an $l \times l$ matrix with entries a_{ij}. The *transposition* of any matrix $\boldsymbol{A} = (a_{ij})$ is denoted by $\boldsymbol{A}^\top$, i.e. $\boldsymbol{A}^\top = (a_{ji})$, but we rarely use it. For this we denote row vectors by Greek small case bold face letters, e.g. $\boldsymbol{\xi} = (\xi_j)_{j=1,\ldots,l}$, and column vectors by small case, bold face Roman letters, like $\boldsymbol{x}^\top = (x_j)_{j=1,\ldots,l}$, that is

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_l \end{pmatrix}.$$

The row vector of ones is $\boldsymbol{\varepsilon} = (1, \ldots, 1)$, by $\boldsymbol{\varepsilon}_i$ we denote the (row) vector having zeros at all components with the exception of the i-th component, which is equal to 1, i.e.

$$\boldsymbol{\varepsilon}_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots).$$

Similarly the column vector of ones is $\boldsymbol{e}$, and $\boldsymbol{e}_i = \boldsymbol{\varepsilon}^\top$. We always denote by $\mathbf{0}$ a vector or matrix having all entries equal to zero. We will use the following relations. We write $\boldsymbol{A} \geq \mathbf{0}$ if for all $a_{ij} \geq 0$, $\boldsymbol{A} > \mathbf{0}$ if for all $a_{ij} > 0$ and $\boldsymbol{A} \neq \mathbf{0}$ if there exists a nonzero entry $a_{ij} \neq 0$. Matrix $\boldsymbol{A}$ is said to be *nonnegative* if $\boldsymbol{A} \geq \mathbf{0}$ We say that $\boldsymbol{\rho} = (\rho_j)$ is a *probability vector* if $\rho_i \geq 0$ for all i and $\sum_i \rho_i = 1$.

1 Spectral theory of nonnegative matrices

1.1 Eigenvalues and eigenvectors

Assume that $\boldsymbol{A}$ is an $(l \times l)$ matrix, that $\boldsymbol{x}$ is an l-dimensional vector with at least one component different from zero, and that θ is a real or complex number. A matrix of any dimension all of whose entries are 0 is denoted by $\mathbf{0}$. If

$$\boldsymbol{A}\boldsymbol{x} = \theta\boldsymbol{x}, \tag{1.1}$$

then θ is said to be an *eigenvalue* of $\boldsymbol{A}$ and $\boldsymbol{x}$ is said to be a *right eigenvector* corresponding to θ. Writing (1.1) as $(\boldsymbol{A}-\theta\boldsymbol{I})\boldsymbol{x} = \mathbf{0}$, from the theory of linear algebraic equations we get that the eigenvalues are exactly the solutions to the *characteristic equation*

$$\det(\boldsymbol{A} - \theta\boldsymbol{I}) = 0\,. \tag{1.2}$$

A nonzero vector $\boldsymbol{\xi}$ which is a solution to

$$\boldsymbol{\xi}\boldsymbol{A} = \theta\boldsymbol{\xi} \tag{1.3}$$

is called a *left eigenvector* corresponding to θ. It is easy to see that for each eigenvalue θ, a solution $\boldsymbol{\xi}$ to (1.3) always exists because (1.2) implies that $\det((\boldsymbol{A}-\theta\boldsymbol{I})^\top) = 0$, i.e. there exists a nonzero (column) vector $\boldsymbol{\xi}^\top$ such that $(\boldsymbol{A} - \theta\boldsymbol{I})^\top\boldsymbol{\xi}^\top = \mathbf{0}$, which is equivalent to (1.3).

Note that (1.2) is an algebraic equation of order l, i.e. there are l eigenvalues $\theta_1, \ldots, \theta_l$, which can be complex and some of them can coincide (multiple eigenvalues). We always assume that the eigenvalues $\theta_1, \ldots, \theta_l$ are numbered such that

$$|\theta_1| \geq |\theta_2| \geq \ldots \geq |\theta_l|\,.$$

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l)$ be an $l \times l$ matrix consisting of right (column) eigenvectors,

$$\boldsymbol{\Xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_l \end{pmatrix}$$

an $l \times l$ matrix consisting of left eigenvectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_l$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$ the vector of eigenvalues. There results the equation

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}\,\mathrm{diag}(\boldsymbol{\theta})\,, \tag{1.4}$$

where $\mathrm{diag}(\boldsymbol{\theta})$ denotes the diagonal matrix with diagonal elements $\theta_1, \ldots, \theta_l$ and all other elements equal to zero. A direct consequence of (1.4) is

$$\boldsymbol{A}^n \boldsymbol{X} = \boldsymbol{X}\,\mathrm{diag}(\theta_1^n, \ldots, \theta_l^n)\,. \tag{1.5}$$

We make a number of observations.

- If all eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$ are linearly independent, then $\boldsymbol{X}^{-1}$ exists. In this case, we can put $\Xi = \boldsymbol{X}^{-1}$.
- Under the linear independence assumption $\boldsymbol{A} = \boldsymbol{X}\,\mathrm{diag}(\boldsymbol{\theta})\boldsymbol{X}^{-1} = \boldsymbol{X}\,\mathrm{diag}(\boldsymbol{\theta})\Xi$ and, consequently,

$$\boldsymbol{A}^n = \boldsymbol{X}(\mathrm{diag}(\boldsymbol{\theta}))^n\boldsymbol{X}^{-1} = \boldsymbol{X}(\mathrm{diag}(\boldsymbol{\theta}))^n\Xi. \tag{1.6}$$

- From (1.6) we get

$$\begin{aligned}
\boldsymbol{A}^n &= (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l)(\mathrm{diag}(\boldsymbol{\theta}))^n \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_l \end{pmatrix} \\
&= (\theta_1^n \boldsymbol{x}_1, \ldots, \theta_l^n \boldsymbol{x}_l) \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_l \end{pmatrix},
\end{aligned}$$

  which yields the *spectral representation* of $\boldsymbol{A}^n$, i.e.

$$\boldsymbol{A}^n = \sum_{i=1}^{l} \theta_i^n \boldsymbol{x}_i \boldsymbol{\xi}_i\,. \tag{1.7}$$

We make two remarks. The first is that since $\Xi = \boldsymbol{X}^{-1}$ we have in (1.7) that $\boldsymbol{\xi}_i \boldsymbol{x}_i = 1$. Secondly we stress that the crucial assumption for the validity of (1.7) is that the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$ are linearly independent. The following lemma gives a simple sufficient condition.

Lemma 1.1 *If the eigenvalues $\theta_1, \ldots, \theta_l$ are distinct, then $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$ are linearly independent. Moreover, if the left eigenvectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_l$ are defined via $\Xi = \boldsymbol{X}^{-1}$, then*

$$\boldsymbol{\xi}_i \boldsymbol{x}_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{1.8}$$

We will always assume that if $\boldsymbol{x}_1$ and $\boldsymbol{\xi}_1$ are right and left eigenvectors corresponding to θ_{PF}, then $\boldsymbol{\xi}_1 \boldsymbol{x} = 1$.

1.2 Perron–Frobenius Theorem

Let $\boldsymbol{A}$ be a nonnegative $l \times l$ matrix.

Definition 1.2 We say that a matrix $\boldsymbol{A}$ is *primitive* if for some k, $\boldsymbol{A}^k > \mathbf{0}$.

Recall that $\boldsymbol{A}$ is *irreducible* if $\sum_n \boldsymbol{A}^n > \mathbf{0}$. Let $\boldsymbol{A} = (a_{ij}^{(n)})$. Let d_i be the greatest common divisor of $\{n : a_{ii}^{(n)} > 0\}$. If $d = d_i$ is the common period for all $i = 1, \ldots, l$, then d is said to be the period of $\boldsymbol{A}$. If $d = 1$, then we say $\boldsymbol{A}$ is *aperiodic*. It is known that irreducible $\boldsymbol{A}$ has a common period.
Problem

1.1 Show that $\boldsymbol{A}$ is primitive if and only if it is irreducible and aperiodic.

In many application it is important to know the position of the dominant eigenvalue, and features of the corresponding right and left eigenvectors. These properties are listed in the following important result, called the *Perron–Frobenius theorem.*

Theorem 1.3 *If $\boldsymbol{A}$ is a nonnegative primitive, then*
(a) *θ_1 is strictly positive and of multiplicity 1 and moreover $|\theta_1| > |\theta_i|$ for $i = 2, \ldots, l$;*
(b) *the right and left eigenvectors $\boldsymbol{x}_1, \boldsymbol{\xi}_1$ have all components strictly positive and are unique up to constant multiples;*
(c) *if $\mathbf{0} \le \boldsymbol{B} \le \boldsymbol{A}$ is another nonnegative matrix with an eigenvalue θ', then $|\theta'| \le \theta_{\mathrm{PF}}$.*

The *proof* of Theorem 1.3 can be found, for example, in Chapter 1 of Seneta (1981). The eigenvalue θ_1 of a regular matrix $\boldsymbol{A}$ is called the *Perron–Frobenius eigenvalue*, and therefore we denote it sometimes as θ_{PF}. Another name is *spectral radius.*

From Perron-Frobenius theorem we can learn about asymptotic behaviour of powers $\boldsymbol{A}^n$, for $n \to \infty$. By $(g(n))_n$ we denote a sequence such that for some numbers $0 < \alpha < \beta < \infty$: $\alpha g(n) \le O(g(n)) \le \beta g(n)$, for all n.

Theorem 1.4 *Assume that a nonnegative $\boldsymbol{A}$ is primitive. Then for some $k \in \mathbb{Z}_+$ (which can be detected)*
(a) *for $\theta_2 \neq 0$ and $n \to \infty$,*

$$\boldsymbol{A}^n = \theta_{\mathrm{PF}}^n \boldsymbol{x}_1 \boldsymbol{\xi}_1 + O(n^k |\theta_2|^n)\,; \tag{1.9}$$

(b) *for $\theta_2 = 0$ and $n \ge l - 1$,*

$$\boldsymbol{A}^n = \theta_{\mathrm{PF}}^n \boldsymbol{x}_1 \boldsymbol{\xi}_1\,. \tag{1.10}$$

The *proof* of this theorem can also be found in Chapter 1 of Seneta (1981).

We now consider matrices which are not aperiodic but still the assumption of irreducibility holds. In this case Theorem 1.3 holds with the following modification in point (a): θ_1 is strictly positive and of multiplicity 1 and moreover $|\theta_1| \geq |\theta_i|$ for $i = 2, \ldots, l$.[1]

More specific result we can obtain for $\boldsymbol{A}$ being a cyclic matrix with period d, that is the matrix which after a renumeration of row and columns respectively has the form

$$\begin{pmatrix} \mathbf{0} & \boldsymbol{A}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{A}_d & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Theorem 1.5 *For a cyclic matrix $\boldsymbol{A}$ with period $d > 1$, there are exactly d eigenvalues θ with $|\theta| = \theta_{\mathrm{PF}}$. These values are roots of the equation $\theta^d - (\theta_{\mathrm{PF}})^d = 0$.*

1.3 Applications to Markov chains

We consider here finite state space discrete time Markov chains (DTMC), defined by a *stochastic matrix* $\boldsymbol{P} = (p_{ij})_{i,j=1,\ldots,l}$. Recall that then all the entries $p_{ij} \geq 0$ and that $\sum_{j=1}^{l} p_{ij} = 1$ for all i. We say that $\boldsymbol{P}$ is *substochastic*, if for at least one row, say i_0, $\sum_{j=1}^{l} p_{i_0 j} < 1$. Recall that a probability vector $\boldsymbol{\pi}$, that is with nonnegative entries, summing up to 1, is stationary, if it solves $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$. We will call it *stationary distribution.*

Corollary 1.6 *If $\boldsymbol{P}$ is a primitive stochastic matrix, then*
(a) $\theta_{\mathrm{PF}} = 1, \boldsymbol{x}_1 = \boldsymbol{e}$ *and* $\boldsymbol{\xi}_1 = \boldsymbol{\pi}$*;*
(b) $|\theta_i| < 1$ *for* $i = 2, \ldots, l$.

Proof By inspection we get that $\boldsymbol{P}\boldsymbol{e} = \boldsymbol{e}$. Hence 1 is an eigenvalue of $\boldsymbol{P}$ and $\boldsymbol{e}, \boldsymbol{\pi}$ are right and left eigenvectors for this eigenvalue, respectively. Also, $\theta_1 = 1$, i.e. 1 is the eigenvalue with the largest modulus. □

We will reserve the letter $\boldsymbol{\pi}$ for the Perron-Frobenius left eigenvector, which is a probability function, for the case of primitive stochastic or substochastic matrices. Similarly we reserve $\boldsymbol{h}$ for the Perron-Frobenius right eigenvector.

[1] Seneta, p.22

We denote entries of $\boldsymbol{P}^n = (p_{ij}^{(n)})_{i,j=1,\ldots,l}$. Notice that

$$\boldsymbol{e}\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} \tag{1.11}$$

and such the matrix we denote by $\boldsymbol{\Pi}$. We conclude with the following:

Corollary 1.7 *If $\boldsymbol{P}$ is a stochastic primitive matrix, then for some $k \in \mathbb{Z}_+$*

$$\begin{aligned} |p_{ij}^{(n)} - \pi_j| &= O(n^k |\theta_2|^n) \\ &= o(c^n) \end{aligned} \tag{1.12}$$

as $n \to \infty$, where $c > |\theta_2|$.

1.4 Quasi-stationary distribution

Suppose that $\boldsymbol{P} = (p_{ij})_{i,j=1,\ldots,l}$ is a substochastic matrix. We will assume that $\boldsymbol{P}$ is primitive. Let θ_{PF}, $\boldsymbol{\pi}$ and $\boldsymbol{h}$ are the Perron-Frobienius eigenvalue, left and right Perron-Frobenius vector respectively. We assume that $\boldsymbol{\pi}$ is normalised to be a probability function. If $\boldsymbol{P}$ is primitive, then

$$\boldsymbol{P}^n = \theta_{\mathrm{PF}}^n \boldsymbol{h}\boldsymbol{\pi} + O(n^k |\theta_2|^n). \tag{1.13}$$

Problems

1.1 Show that $0 < \theta_{\mathrm{PF}} < 1$.

1.2 Show that if $\boldsymbol{P}$ is substochastic and primitive, then for $n \to \infty$

$$\boldsymbol{P}^n \sim \theta_{\mathrm{PF}}^n \begin{pmatrix} h_1 \boldsymbol{\pi} \\ \vdots \\ h_l \boldsymbol{\pi} \end{pmatrix} \tag{1.14}$$

$$\sum_{j=1}^{l} p_{ij}^{(n)} \sim h_i \theta_{\mathrm{PF}}^n. \tag{1.15}$$

By adding the 0-th row and column we can extend $\boldsymbol{P}$ to a true stochastic matrix:

$$\boldsymbol{P}^* = \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{p}^\circ & \boldsymbol{P} \end{pmatrix} \tag{1.16}$$

where $\boldsymbol{p}^\circ = \boldsymbol{e} - \boldsymbol{P}\boldsymbol{e}$.

Consider now a DTMC (X_n) with state space $E = \{0, 1, \dots, l\}$, initial distribution $\boldsymbol{\rho}$ (in particular when beginning from i, $\boldsymbol{\rho} = \boldsymbol{\varepsilon}_i$) and transition matrix $\boldsymbol{P}^*$. Such the DTMC is considered on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{\rho}})$ or $(\Omega, \mathcal{F}, \mathbb{P}_i)$ respectively. To avoid trivialities $i \in \{1, \dots, l\}$. Define $\nu = \inf\{n \geq 0 : X_n = 0\}$. The following observation will be important for us.

Problem

1.1 Prove that ν is finite a.s.

Proposition 1.8 *The following sentences are equivalent.*
(i) θ_{PF}, $\boldsymbol{\pi}$ *are the Perron-Frobenius eigenvalue and left eigenvector respectively for* $\boldsymbol{P}$.
(ii)

$$\mathbb{P}_{\boldsymbol{\pi}^*}(X_n = j \mid \nu > n) = \pi_j, \qquad j = 1, \dots, l, \tag{1.17}$$

where $\boldsymbol{\pi}^* = (0, \boldsymbol{\pi})$.

Part (ii) of Proposition 1.8 can serve as the definition for infinite substochastic matrix $\boldsymbol{P} = (p_{ij})_{i,j=1,\dots}$. We can extend $\boldsymbol{P}$ to $\boldsymbol{P}^*$ as above and consider a DTMC (X_n) with state space $E = \{0, 1, \dots\}$.

Definition 1.9 A distribution $\boldsymbol{\pi}$ on $\{1, 2, \dots\}$ is called *quasi-stationary* (QS) if

$$\mathbb{P}_{\boldsymbol{\pi}^*}(X_n = j \mid \nu > n) = \pi_j, \qquad j = 1, \dots, \tag{1.18}$$

where $\boldsymbol{\pi}^* = (0, \boldsymbol{\pi})$ is a probability function on $\{0, 1, \dots\}$.

We can write (1.17) in the vector notation

$$\boldsymbol{\pi}\boldsymbol{P}^n = (\boldsymbol{\pi}\boldsymbol{P}^n\boldsymbol{e})\boldsymbol{\pi}, \qquad n = 0, 1, \dots. \tag{1.19}$$

Proof of Proposition 1.8. (i)$\to$(ii). Iterating the eigenvalue equation we have

$$\boldsymbol{\pi}\boldsymbol{P}^n = \theta_{\mathrm{PF}}^n \boldsymbol{\pi}, \qquad n = 0, 1, \dots$$

which immediately can be written in form (1.19).
(ii)$\to$(i). We have that for some probability vector $\boldsymbol{\rho}$

$$\boldsymbol{\rho}\boldsymbol{P}^n = (\boldsymbol{\rho}\boldsymbol{P}^n\boldsymbol{e})\boldsymbol{\rho},$$

for all n. The above equation for $n = 1$ yields that $\boldsymbol{\rho P e}$ and $\boldsymbol{\rho}$ are an eigenvalue and the corresponding left eigenvector respectively. Then $\boldsymbol{\rho P}^n \boldsymbol{e} = (\boldsymbol{\rho P e})^n$. Assume that $\boldsymbol{\rho} \neq \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the Perron-Frobenius probability vector. Then $\theta = \boldsymbol{\rho P e} < \theta_{\mathrm{PF}}$ and

$$\boldsymbol{\rho P}^n = \theta^n \boldsymbol{\rho}.$$

Applying however (1.14) we have

$$\boldsymbol{\rho P}^n \sim \theta_{\mathrm{PF}}^n \boldsymbol{\pi}$$

which is impossible. Hence $\boldsymbol{\rho} = \boldsymbol{\pi}$. □

Proposition 1.10 *If $\boldsymbol{P}$ is a substochastic primitive matrix with the quasi-stationary distribution $\boldsymbol{\pi}$, then for $i \in \{1, \ldots, l\}$*

$$\lim_{n \to \infty} \mathbb{P}_i(X_n = j \mid \nu > n) = \pi_j, \qquad j = 1, \ldots, l \tag{1.20}$$

Proof We have to prove that for $i, j = 1, \ldots, l$

$$\lim_{n \to \infty} \frac{p_{ij}^{(n)}}{\sum_{k \in E} p_{ik}^{(n)}} = \pi_j \tag{1.21}$$

The proof is immediate from (1.14) and (1.15). □

From (1.13) we also immediately obtain the following result.

Problem

1.1 Show that for primitive substochastic $\boldsymbol{P}$

$$\lim_{n \to \infty} \frac{p_{ij}^{(n+m)}}{p_{i'j'}^{(n)}} = R^{-m} \frac{h_i \pi_j}{h_{i'} \pi_{j'}}, \tag{1.22}$$

where $R = (\theta_{\mathrm{PF}})^{-1}$.

We now state a general definition for the two properties, which trivially hold for finite cases. Suppose that E is denumerable and $\boldsymbol{P} = (p_{ij})_{i,j \in E}$ be a substochastic matrix.

Definition 1.11 It is said that for $\boldsymbol{P}$, the distribution (that is normalised to 1) $\boldsymbol{\pi}$ is the *Yaglom limit* if (1.21) hold. It is said that for $\boldsymbol{P}$ the *strong ratio limit property* (SRLP) holds, if (1.22) hold.

For finite substochastic matrices, the notions of quasi-stationary distribution and Yaglom limit coincide. This is not true in general. For both the considered property, in general cases we have only one side implications.

Problems

1.1 Show that if $\boldsymbol{\pi}$ is a Yaglom limit, then it is a QS distribution. (Hint. In the proof use the following variant of Helly-Brey theorem; let $\boldsymbol{p}_n = (p_{nk})_k$ be probability functions, and for all k the limit $\lim_{n\to\infty} p_{nk} = p_k$ exists and $(p_k)_k$ is a true probability function. Then for any bounded sequence $(a_n)_n$ we have $\lim_{n\to\infty} \sum_k a_k p_{nk} = \sum_k a_k p_k$.)

1.2 Show that if SRLP holds with a true probability function $\boldsymbol{\pi}$, then $\boldsymbol{\pi}$ is a QS distribution.

Comments. A good references for the theory outlined in this chapter is the book of Seneta (1981); see also the book of Asmussen (2003). QS distributions were studied originally by Vere-Jones (1962) and Seneta & Vere-Jones (1966). The key results on QS distributions and SRLP belong to Kesten (1995) and Ferrari *et al* (1995).

Chapter II

Discrete time martingales in applied probability

1 Discrete time martingales

1.1 Filtrations and Stopping Times

Definition 1.1 A family $(\mathcal{F}_n, n \in \mathbb{Z}_+)$ of σ-fields such that $\mathcal{F}_n \subset \mathcal{F}$ and $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \in \mathbb{Z}_+$ is called a *filtration.* We say that the sequence $(X_n, n \in \mathbb{Z}_+)$ is adapted to the filtration $(\mathcal{F}_n)$ if X_n is measurable with respect to $\mathcal{F}_n$ for all $n \in \mathbb{Z}_+$.

Unless state otherwise, we will use *natural filtration*

$$\mathcal{F}_n = \mathcal{F}_n^X = \sigma\{X_0, \ldots, X_n\}, \qquad n \in \mathbb{Z}_+$$

.

Definition 1.2 A random variable ν taking values in $\mathbb{Z}_+ \cup \{\infty\}$ is said to be a *stopping time* with respect to a filtration $(\mathcal{F}_n)$ (or equivalently an $(\mathcal{F}_n)$-stopping time) if the event $\{\nu = n\}$ belongs to $\mathcal{F}_n$, for all $n \in \mathbb{Z}_+$. When filtration is obvious,we say simply *stopping time.*

Problems

1.1 Let (X_n) be a sequence of real-valued random variables. Consider the *first entrance time* ν_B of (X_n) to a Borel set $B \in \mathcal{B}(\mathbb{R})$, i.e

$$\nu_B = \begin{cases} \min\{n : \ X_n \in B\} & \text{if } X_n \in B \text{ for some } n \in \mathbb{Z}_+, \\ \infty & \text{otherwise.} \end{cases}$$

Show that random variable ν^B is a stopping time with respect to $(\mathcal{F}_n^X)$.

We say that a stopping time ν is *bounded* if there exists an $a > 0$ such that $\mathbb{P}(\nu \le a) = 1$.

1.2 Martingales, Sub- and Supermartingales

Definition 1.3 Let $(\mathcal{F}_n)$ be a filtration and let (M_n) be a sequence of random variables adapted to $(\mathcal{F}_n)$ such that $\mathbb{E}\,|M_n| < \infty$ for each $n \in \mathbb{Z}_+$. Then (X_n) is called a *martingale* with respect to $(\mathcal{F}_n)$, if with probability 1

$$\mathbb{E}\,(M_{n+1} \mid \mathcal{F}_n) = M_n\,, \tag{1.1}$$

for all $n \in \mathbb{Z}_+$. Similarly, (M_n) is called a *submartingale* if

$$\mathbb{E}\,(M_{n+1} \mid \mathcal{F}_n) \ge M_n\,, \tag{1.2}$$

and a *supermartingale* if

$$\mathbb{E}\,(M_{n+1} \mid \mathcal{F}_n) \le M_n\,, \tag{1.3}$$

for all $n \in \mathbb{Z}_+$.

Regarding how obvious is the filtration or the underlying probability measure we may use equivalently an $(\mathcal{F}_n)$-martingale, $\mathbb{P}$-martingale or if everything is obvious simply a martingale.

Proposition 1.4 *The following sentences are equivalent:*
(i) (M_n) *is a martingale.*
(ii) *For all* $m \le n$

$$\mathbb{E}\,(M_n \mid \mathcal{F}_m) = M_m\,, \tag{1.4}$$

(ii) *For all* $m \le n$

$$\int_A M_m \mathrm{d}\mathbb{P} = \int_A M_n \mathrm{d}\mathbb{P}, \qquad A \in \mathcal{F}_m. \tag{1.5}$$

Proof Assume (i). Indeed, repeatedly using (1.1) and basic properties of conditional expectation we have

$$\begin{aligned}
\mathbb{E}\,(M_{n+k} \mid \mathcal{F}_n) &= \mathbb{E}\,(\mathbb{E}\,(M_{n+k} \mid \mathcal{F}_{n+k-1}) \mid \mathcal{F}_n) = \mathbb{E}\,(M_{n+k-1} \mid \mathcal{F}_n) \\
&= \mathbb{E}\,(\mathbb{E}\,(M_{n+k-1} \mid \mathcal{F}_{n+k-2}) \mid \mathcal{F}_n) \\
&\;\;\vdots \\
&= \mathbb{E}\,(M_{n+1} \mid \mathcal{F}_n) = M_n\,.
\end{aligned}$$

Analogously, (1.2) and (1.3) imply

$$\mathbb{E}\,(M_{n+k} \mid \mathcal{F}_n) \geq M_n \tag{1.6}$$

and

$$\mathbb{E}\,(M_{n+k} \mid \mathcal{F}_n) \leq M_n \tag{1.7}$$

for all $k, n \in \mathbb{Z}_+$. □

Taking expectations on both sides of (1.4)–(1.7) we get

- for a martingale, $\mathbb{E}\, M_n = \mathbb{E}\, M_0$ for all $n \in \mathbb{Z}_+$,
- for a submartingale, $\mathbb{E}\, M_{n+k} \geq \mathbb{E}\, M_n$ for all $k, n \in \mathbb{Z}_+$,
- for a supermartingale, $\mathbb{E}\, M_{n+k} \leq \mathbb{E}\, M_n$ for all $k, n \in \mathbb{Z}_+$.

Examples

1. Consider a martingale $(W_n, n \in \mathbb{Z}_+)$ with respect to a filtration $(\mathcal{F}_n)$ and an adapted sequence $(Z_n, n = 0, 2, \ldots)$ of random variables. The sequence (X_n) with $X_0 = 0$ and

$$X_n = \sum_{k=1}^{n} Z_{k-1}(W_k - W_{k-1}), \qquad n \in \mathbb{Z}_+\,, \tag{1.8}$$

is a martingale, provided that $\mathbb{E}\,|Z_k(W_k - W_{k-1})| < \infty$ holds for all $k = 1, 2, \ldots$. Indeed,

$$\begin{aligned}
&\mathbb{E}\,(X_{n+1} \mid \mathcal{F}_n) \\
&\quad = \sum_{k=1}^{n} \mathbb{E}\,(Z_{k-1}(W_k - W_{k-1}) \mid \mathcal{F}_n) + \mathbb{E}\,(Z_n(W_{n+1} - W_n) \mid \mathcal{F}_n) \\
&\quad = \sum_{k=1}^{n} Z_{k-1}(W_k - W_{k-1}) + Z_n \mathbb{E}\,(W_{n+1} - W_n \mid \mathcal{F}_n) \\
&\quad = \sum_{k=1}^{n} Z_{k-1}(W_k - W_{k-1}) = X_n\,.
\end{aligned}$$

Note that (1.8) is a discrete analogue to the stochastic integral of a predictable process with respect to a martingale.

2. Consider a DTMC (X_n) with finite state space $E = \{1, \ldots, l\}$ and transition probability matrix $\boldsymbol{P}$. If $\theta \neq 0$ is an eigenvalue of $\boldsymbol{P}$ and $\boldsymbol{x} = (x_1, \ldots, x_\ell)$

the corresponding right eigenvector, then (M_n) with $M_n = \theta^{-n} x_{X_n}$ is a martingale (perhaps a complex valued one). This can be seen as follows. From the Markov property we have $\mathbb{E}\,(M_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\,(M_{n+1} \mid X_n)$. Now, for all $i = 1, \ldots, \ell$, (I.1.1) implies

$$\mathbb{E}\,(M_{n+1} \mid X_n = i) = \theta^{-n}\theta^{-1}\sum_{j=1}^{l} p_{ij}x_j = \theta^{-n}\phi_i\,,$$

from which we have $\mathbb{E}\,(M_{n+1} \mid X_n) = \theta^{-n} x_{Z_n} = M_n$.

Problems

1.1 Assume now the setting of Section I.1.4 and let $\boldsymbol{P}\boldsymbol{x} = \theta\boldsymbol{x}$. Show that

$$M_n = \theta^{-n} x_{X_n} 1(\nu > n)$$

is $\mathbb{P}_i$-martingale, for $i = 1, \ldots, l$.

1.2 Let f and $\tilde{f}$ be density functions on $\mathbb{R}$ such that $f \neq \tilde{f}$. For simplicity assume that the product $f(x)\tilde{f}(x) > 0$ for all $x \in \mathbb{R}$. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, with the common density either f or $\tilde{f}$. The *likelihood ratio sequence* $(M_n, n \in \mathbb{Z}_+)$ is then given by

$$M_n = \begin{cases} \prod_{k=1}^{n} \dfrac{\tilde{f}(X_k)}{f(X_k)} & \text{if } n \geq 1, \\ 1 & \text{if } n = 0. \end{cases}$$

Show that (M_n) is an $(\mathcal{F}_n^X)$-martingale if the X_n have density f. In the alternative situation that X_n has density $\tilde{f}$, and under the additional assumption $\int_{-\infty}^{\infty} \tilde{f}^2(x)/f(x)\,\mathrm{d}x < \infty$ show that (X_n) is a submartingale with respect to $(\mathcal{F}_n^Y)$.

We now state few important for us result on martingales proporties; for proofs see e.g. Rolski *et al* (1999). We will use the following result on the convergence of a supermartingale.

Theorem 1.5 *A nonegative supermartingale converges almost surely to a finite limit.*

A useful tool is also the following *submartingale convergence theorem.*

Theorem 1.6 *If* $(X_n, n \geq 0)$ *is a submartingale such that*

$$\sup_{n\geq 0} \mathbb{E}\,|X_n| < \infty\,, \tag{1.9}$$

then there exists a random variable X_∞ *such that,*

$$\lim_{n\to\infty} X_n = X_\infty \qquad \text{a.s.} \tag{1.10}$$

and $\mathbb{E}\,|X_\infty| < \infty$. *If, additionally,*

$$\sup_{n\geq 0} \mathbb{E}\,X_n^2 < \infty\,, \tag{1.11}$$

then

$$\mathbb{E}\,X_\infty^2 < \infty\,, \qquad \lim_{n\to\infty} \mathbb{E}\,|X_n - X_\infty| = 0\,. \tag{1.12}$$

The next result is known as *optional sampling theorems.*

Theorem 1.7 *Let* (X_n) *be a martingale and* ν *stopping time. Then* $\mathbb{E}\,X_\nu = \mathbb{E}\,X_0$ *holds if one of the following conditions is fulfilled:*
(i) ν *is bounded,*
(ii) ν *is a finite stopping time fulfilling*

$$\mathbb{E}\,|X_\nu| < \infty \tag{1.13}$$

and

$$\lim_{k\to\infty} \mathbb{E}\,[X_k; \nu > k] = 0\,, \tag{1.14}$$

(iii) ν *a stopping time fulfilling*

$$\mathbb{E}\,\nu < \infty \tag{1.15}$$

and, for some constant $c < \infty$,

$$\mathbb{E}\,(|X_{n+1} - X_n| \mid \mathcal{F}_n) \leq c \qquad \text{a.s.} \tag{1.16}$$

for all $n \in \mathbb{Z}_+$.

Theorem 1.7 (iii) can be used to prove *Wald's identity* (III.2.5) for stopping times.

Corollary 1.8 *Consider a random walk* $\{S_n\}$ *with i.i.d increments* $X_1, X_2, \ldots$ *and* $\mathbb{E}\,|X| < \infty$. *If* ν *is a stopping time with respect to the filtration* $\{\mathcal{F}_n\}$ *and if* $\mathbb{E}\,\nu < \infty$, *then*

$$\mathbb{E}\,S_\nu = \mathbb{E}\,\nu \mathbb{E}\,Y. \tag{1.17}$$

We now state *Doob-Kolmogorov inequality* for sub- and supermartingales.

Theorem 1.9 (a) *If* (X_n) *is a nonnegative submartingale, then*

$$\mathbb{P}\Big(\max_{0\le k\le n} X_k \ge x\Big) \le \frac{\mathbb{E}\,X_n}{x}, \qquad x>0, n\in\mathbb{Z}_+ . \tag{1.18}$$

(b) *If* (X_n) *is a nonnegative supermartingale, then*

$$\mathbb{P}\Big(\max_{0\le k\le n} X_k \ge x\Big) \le \frac{\mathbb{E}\,X_0}{x}, \qquad x>0, n\in\mathbb{Z}_+ . \tag{1.19}$$

Example 1.10 Doob-Kolmogorov inequality (1.19) can be used to prove an exponential bound for a compound $\sum_{j=1}^N U_j$,

- where $U_1, U_2, \ldots$ are nonnegative, independent and identically distributed with distribution F_U,
- N be an $\mathbb{Z}_+$-valued random variable with probability function $\{p_k\}$ which is independent of $U_1, U_2, \ldots$.

We suppose that for some $0<\theta<1$,

$$\mathbb{P}(N>n+1 \mid N>n) \le \theta, \qquad n\in\mathbb{Z}_+, \tag{1.20}$$

(i.e. $r_{n+1}\le\theta r_n$ for $n\ge 1$, where $r_n=\sum_{k=n}^{\infty} p_k$). Furthermore, assume that

$$\hat{F}_U(\gamma)=\theta^{-1} \tag{1.21}$$

has the solution $\gamma>0$. We denote the moment generating function of a distribution F by $\hat{F}$; see for definition and properties Appendix I.1. We will show that

$$\mathbb{P}(\sum_{j-1}^{N} U_j > x) \le \frac{1-p_0}{\theta} e^{-\gamma x}, \qquad x\ge 0 . \tag{1.22}$$

For each $n\in\mathbb{N}$, define

$$X_n = \begin{cases} \mathrm{e}^{\gamma S_{n+1}} & \text{if } N>n, \\ 0 & \text{if } N\le n, \end{cases} \tag{1.23}$$

where $S_n = U_1+\ldots+U_n$. Then, $X_{n+1}=Z_{n+1}X_n$ for $n=1,2,\ldots$, where

$$Z_n = \begin{cases} \mathrm{e}^{\gamma U_{n+1}} & \text{if } N>n, \\ 0 & \text{if } N\le n. \end{cases}$$

Consider the filtration $\{\mathcal{F}_n\}$, where $\mathcal{F}_n$ is the σ-algebra generated by the random variables $1(N=0),\ldots,1(N=n),U_1,\ldots,U_{n+1}$. Since we have

$$\mathbb{P}(N>n+1\mid\mathcal{F}_n)\le\theta, \tag{1.24}$$

and hence

$$\begin{aligned}&\mathbb{E}\,(Z_{n+1}\mid\mathcal{F}_n)=\mathbb{E}\,(\mathrm{e}^{\gamma U_{n+2}}1(N>n+1)\mid\mathcal{F}_n)=\\&=\mathbb{E}\,(\mathrm{e}^{\gamma U_{n+2}})\mathbb{P}(N>n+1\mid\mathcal{F}_n)=\theta^{-1}\mathbb{P}(N>n+1\mid\mathcal{F}_n)\le 1\,.\end{aligned}$$

Thus $\mathbb{E}\,(X_{n+1}\mid\mathcal{F}_n)=\mathbb{E}\,(Z_{n+1}X_n\mid\mathcal{F}_n)=\mathbb{E}\,(Z_{n+1}\mid\mathcal{F}_n)X_n\le X_n$, that is $\{X_n,n\in\mathbb{N}\}$ is an $\{\mathcal{F}_n\}$-supermartingale. Then (1.19) gives

$$\begin{aligned}\mathbb{P}\Big(\sum_{j=1}^{N}U_j>x\Big)&=\mathbb{P}\Big(\max_{n\in\mathbb{Z}_+}\{S_{n+1}1(N>n)\}>x\Big)=\mathbb{P}\Big(\max_{n\in\mathbb{Z}_+}X_n>\mathrm{e}^{\gamma x}\Big)\\&=\lim_{m\to\infty}\mathbb{P}\Big(\max_{0\le n\le m}X_n>\mathrm{e}^{\gamma x}\Big)\le\frac{\mathbb{E}\,X_0}{\mathrm{e}^{\gamma x}}=\frac{1-p_0}{\theta}\mathrm{e}^{-\gamma x}.\end{aligned}$$

Problem

1.1 Prove (1.24)

2 Foundation of the change of measure method

Consider the basic a probability space $(\Omega,\mathcal{F},\mathbb{P})$ equipped with a filtration $(\mathcal{F}_n)_{n\ge0}$, where $\mathcal{F}=\bigvee_{n\ge0}\mathcal{F}_n$ is the smallest σ-field generated by all subset from $\mathcal{F}_n$ $(n\ge0)$. We denote by $\mathbb{F}$ the pair $(\mathcal{F},(\mathcal{F}_n)_{n\ge0})$. We call $(\Omega,\mathbb{F})$ a *filtered space* and $(\Omega,\mathbb{F},\mathbb{P})$ a *filtered probability space*. For a probability $\mathbb{P}$ we denote by $\mathbb{P}_{|n}$ its restriction to $\mathcal{F}_n$, that is the probability measure on $(\Omega,\mathcal{F}_n)$ such that $\mathbb{P}_{|n}(A)=\mathbb{P}(A)$ for all $A\in\mathcal{F}_n$. On the filtered space we consider also another probability measure $\tilde{\mathbb{P}}$ and its restriction to $\mathcal{F}_n$ we denote by $\tilde{\mathbb{P}}_{|n}$. In this section we will use the following assumption:

Assumption A: For all $n\in\mathbb{Z}_+$

$$\tilde{\mathbb{P}}_{|n}\ll\mathbb{P}_{|n}$$

From the Radon-Nikodym theorem we may define

$$M_n=\frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}},\qquad n=0,1,\ldots$$

We make now the following simple observation.

Proposition 2.1 *$(M_n)_n$ is $\mathbb{P}$-martingale, such that $\mathbb{E}\, M_n = 1$.*

Proof Since $\mathbb{P}_{|n}$ and $\tilde{\mathbb{P}}_{|n}$ are defined on $\mathcal{F}_n$, the Radon-Nikodym derivative is an $\mathcal{F}_n$ measurable function, which is clearly nonnegative. Next we check:

$$\mathbb{E}\,|M_n| = \mathbb{E}\, M_n = \int_\Omega \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} \mathrm{d}\mathbb{P} = \int_\Omega \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} \mathrm{d}\mathbb{P}_{|n} = \tilde{\mathbb{P}}_{|n}(\Omega) = 1.$$

Let $m \le n$ and suppose that $A \in \mathcal{F}_m$. Then

$$\tilde{\mathbb{P}}_{|m}(A) = \tilde{\mathbb{P}}_{|n}(A)$$

and

$$\begin{aligned}
\tilde{\mathbb{P}}_{|n}(A) &= \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} \mathrm{d}\mathbb{P}_{|n} = \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} \mathrm{d}\mathbb{P} \\
\mathbb{P}_{|m}(A) &= \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{|m}}{\mathrm{d}\mathbb{P}_{|m}} \mathrm{d}\mathbb{P}_{|m} = \int_A \frac{\tilde{\mathbb{P}}_{|m}}{\mathbb{P}_{|m}} \mathrm{d}\mathbb{P}
\end{aligned}$$

Therefore

$$\int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{|m}}{\mathrm{d}\mathbb{P}_{|m}} \mathrm{d}\mathbb{P} = \int_A \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} \mathrm{d}\mathbb{P}$$

which is a martingale condition (iii) (see Proposition 1.4). □

Proposition 2.2 *Assume that $M_n > 0$ $\tilde{\mathbb{P}}$-a.s. for all $n \in \mathbb{Z}_+$. Then $\mathbb{P}_{|n} \ll \tilde{\mathbb{P}}_{|n}$ and (M_n^{-1}) is a $\tilde{\mathbb{P}}$-martingale.*

Proof

As a special case, assume that $X_1, X_2, \ldots$ are independent and identically distributed random variables; under $\mathbb{P}$ or $\tilde{\mathbb{P}}$ random variable X_n has distribution F or $\tilde{F}$ respectively. Take $\mathcal{F}_0$ trivial and $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$ for $n \ge 1$. We leave the reader to demonstrate the following results.

Problem

2.1 We have $\tilde{\mathbb{P}}_{|n} \ll \mathbb{P}_{|n}$ if and only if $\tilde{F} \ll F$. Suppose furthermore that F and $\tilde{F}$ have densities $g(x)$ and $\tilde{g}(x)$ with respect a measure $\mathrm{d}\mu$ respectively. Show that then $\tilde{\mathbb{P}}_{|n} = M_n \mathrm{d}\mathbb{P}_{|n}$, where

$$M_n = \begin{cases} 1 & n = 0 \\ \frac{\tilde{g}(X_1)}{g(X_1)} \frac{\tilde{g}(X_2)}{g(X_2)} \cdots \frac{\tilde{g}(X_n)}{g(X_n)} & n \ge 1 \end{cases}$$

which is a *likelihood ratio sequence.*

Standard setup We now describe the so called *standard setup*. We take $\Omega = \mathbb{R} \times \mathbb{R} \times \ldots$, $\omega = (\omega_n)$, $X_n(\omega) = \omega_n$, $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ and $\mathcal{F} = \vee_n \mathcal{F}_n$ is the σ-field of cylindrical set. Note that $\mathcal{F}_n$ is the family of sets from $\mathcal{F}$ of form $\{(\omega) : (\omega_0, \ldots, \omega_n) \in B^{n+1}\}$, $B^{n+1} \in \mathcal{B}(\mathbb{R}^{n+1})$.

Suppose now that for all n we have $M_n > 0$ $\tilde{\mathbb{P}}$-a.s. Then we have:

Proposition 2.3 *For $A \in \mathcal{F}_n$*

$$\mathbb{P}_{|n}(A) = \tilde{\mathbb{E}}\,[M_n^{-1}; A]. \tag{2.25}$$

Moreover for a stopping time ν the above can be extended to the basic identity*: for $A \in \mathcal{F}_\nu$ such that $A \subset \{\nu < \infty\}$*

$$\mathbb{P}(A) = \tilde{\mathbb{E}}\,[M_\nu^{-1}; A], \qquad A \in \mathcal{F}_\nu. \tag{2.26}$$

Proof Relation (2.25) is standard; see Appendix I.2. Relation (2.26) can be demonstrated as follows. Conditions $A \in \mathcal{F}_\nu$ and $A \subset \{\nu < \infty\}$ yield $A = \bigcup_{n=0}^{\infty} A \cap \{\nu = n\}$ and $A \cap \{\nu = n\} \in \mathcal{F}_n$. Hence

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{n=0}^{\infty} \mathbb{P}(A \cap \{\nu = n\}) \\
&= \sum_{n=0}^{\infty} \tilde{\mathbb{E}}\,[M_n^{-1}; A \cap \{\nu = n\}] \\
&= \sum_{n=0}^{\infty} \tilde{\mathbb{E}}\,[M_n^{-1} 1(\nu = n); A] \\
&= \tilde{\mathbb{E}}\,[\sum_{n=0}^{\infty} M_n^{-1} 1(\nu = n); A] \\
&= \tilde{\mathbb{E}}\,[M_\nu^{-1}; A]
\end{aligned}$$

□

We start off a filtered probability space $(\Omega, \mathbb{F}, \mathbb{P})$ and show a converse method, how to determine a new probability measure $\tilde{\mathbb{P}}$ by a martingale (M_n), such that

- $M_n > 0$ $\mathbb{P}$-a.s.
- $\mathbb{E}\, M_n = 1$.

Define for all $n \in \mathbb{Z}_+$ a probability measure on $(\Omega, \mathcal{F}_n)$ by

$$\mathrm{d}\tilde{\mathbb{P}}_{|n} = M_n \,\mathrm{d}\mathbb{P}_{|n}, \tag{2.27}$$

which means that

$$\tilde{\mathbb{P}}_{|n}(A) = \mathbb{E}\,[M_n; A], \qquad A \in \mathcal{F}_n.$$

Lemma 2.4 $\{\tilde{\mathbb{P}}_{|n}, n = 0, 1, \ldots\}$ *is a consistent family of probability measures in the sense that for all* $m \leq n$

$$\tilde{\mathbb{P}}_{|m}(A) = \tilde{\mathbb{P}}_{|n}(A), \qquad A \in \mathcal{F}_m.$$

Proof This is actually Proposition 2.1.

Proposition 2.5 *Assume that the standard setup holds. There exists a unique probability measure* $\tilde{\mathbb{P}}$ *on* $(\Omega, \mathcal{F})$ *such that its restriction to* $\mathcal{F}_n$ *is* $\tilde{\mathbb{P}}_{|n}$ *for all* n*.*

Proof Use Lemma 2.4 and Kolmogorov consistency theorem. □

To get rid off the standard setup assumption, we now introduce a notion of isomorphism of two filtered probability spaces $(\Omega, \mathbb{F}, \mathbb{P})$ and $(\Omega^{'}, \mathbb{F}^{'}, \mathbb{P}^{'})$; they are said to be isomorphic if there exists a one to one transformation ϕ from $\mathcal{F}$ to $\mathcal{F}^{'}$ such that

- $\phi(A - A') = \phi(A) - \phi(A')$, $\phi(\bigcup_{j=1}^{\infty} A_j) = \bigcup_{n=1}^{\infty} \phi(A_n)$, whenever $A, A', A_j \in \mathcal{F}$
- $\phi : \mathcal{F}_n \to \mathcal{F}_n^{'}$ is a bijection for $n = 0, 1, \ldots$,
- $\mathbb{P}(A) = \mathbb{P}^{'}(\phi(A))$ for all $A \in \mathcal{F}$.

We may now consider a filtered probability space $(\Omega', \mathbb{F}', \mathbb{P}')$ and a sequence (X_n) of random variables. If $\mathcal{F}_n^{'} = \sigma\{X_0, \ldots, X_n\}$ and $\mathcal{F}^{'} = \bigvee_n \mathcal{F}_n^{'}$, then the considered filtered probability space is isomorphic to the standard setting.

Corollary 2.6 *Let* (M_n) *be a positive mean one martingale and* $\mathrm{d}\mathbb{P}_{|n} = M_n \mathrm{d}\mathbb{P}_{|n}$*. There exists a unique probability measure* $\tilde{\mathbb{P}}$ *on* $(\Omega, \mathcal{F})$ *such that its restriction to* $\mathcal{F}_n$ *is* $\tilde{\mathbb{P}}_{|n}$ *for all* n*.*

Definition 2.7 For $(\Omega, \mathbb{F}, \mathbb{P})$ and (X_n) with natural filtration $\mathcal{F}_n$ we say that $\tilde{\mathbb{P}}$ is obtained by the exponential change of measure (ECM) with a positive martingale (M_n) satisfying $\mathbb{E}\, M_n = 1$, if $\tilde{\mathbb{P}}$ is constructed in the way like above.

It is important for subsequent applications to know when a Markov process under the new probability measure is again a Markov process. Without loss of generality we may assume now the standard setup $(\Omega, \mathbb{F}, \mathbb{P}_x)$, where $x \in E$ and $(X_n)_n$ is a DTMP with state spece E. On Ω one introduces the mapping $\vartheta : \Omega \to \Omega$ by

$$\vartheta((\omega_0, \omega_1, \ldots)) = (\omega_1, \omega_2, \ldots)$$

and we denote by $\vartheta^n = \vartheta \circ \ldots \circ \vartheta$ is n-fold composition. Let $(L_n)_n$ be a sequence of random variables. The sequence is said to be a *multiplicative functional* if it is adapted to flitration $(\mathcal{F}_n)_n$, and for all $n, m \in \mathbb{Z}_+$

$$L_{n+m} = L_n \cdot L_m \circ \vartheta^n.$$

Induction principle yields

$$L_n = L_1 \cdot L_1 \circ \vartheta \cdots L_1 \circ \vartheta^{n-1}.$$

Thus, for nonnegative measurable functions $f_1, \ldots, f_n$

$$\begin{aligned}
&\mathbb{E}_x[L_n f(X_1) \ldots f(X_2)] \\
&= \mathbb{E}_x[\prod_{j=0}^{n-1} L_1 \circ \vartheta^j f(X_1 \circ \vartheta^j)] = \mathbb{E}_x \mathbb{E}\,[\prod_{j=0}^{n-1} L_1 \circ \vartheta^j f(X_1 \circ \vartheta^j) | X_1] \\
&= \mathbb{E}_x L_1 f(X_1) \mathbb{E}_{X_1} \mathbb{E}\,[\prod_{j=0}^{n-2} L_1 \circ \vartheta^j f(X_1 \circ \vartheta^j)] = \ldots = \\
&= \mathbb{E}_x[L_1 f(X_1) \mathbb{E}_{X_1} L_1 f(X_1) \ldots \mathbb{E}_{X_{n-1}} L_1 f(X_1)] \qquad (2.28)
\end{aligned}$$

Theorem 2.8 *Suppose that $(L_n)_n$ is a multiplicative functional such that $\mathbb{E}_x L_n = 1$ for all $x \in E$. Then $(L_n)_n$ is a martingale. If moreover $(L_n)_n$ is positive, then under the probabilty measure $\tilde{\boldsymbol{P}}_x$ defined by (M_n), the sequence (X_n) is a DTMP.*

Proof For the martingale we must prove that

$$\begin{aligned}
\mathbb{E}_x[L_{n+m} \mid \mathcal{F}_n] &= = \mathbb{E}_x[L_n \cdot L_m \circ \vartheta_n \mid \mathcal{F}_n] \\
&= L_n \mathbb{E}_x[L_m \circ \vartheta_n \mid \mathcal{F}_n] \\
&= L_n \mathbb{E}_{X_n}[L_m] = L_n.
\end{aligned}$$

The proof that $(X_n)_n$ is a DTMP under $\tilde{\boldsymbol{P}}_x$ is immediate because of (2.28).
□

Problems

2.1 Suppose that (Z_n) be a process adapted to $(\mathcal{F}_n)$. Show that Z_n is a $\tilde{\mathbb{P}}$-martingale if and only if $(Z_n M_n)$ is a $\mathbb{P}$-martingale. [Similar as in Harrison [**?**], Corollary 1.6.4 on page 10.]

2.1 Harmonic functions; Doob h-transforms

We now consider a DTMC with a general denumerable state space $E = \{1, 2, \ldots\}$ and transition probability matrix $\boldsymbol{P} = (p_{ij})_{i,j=1,\ldots}$. In this case a function is a column vector $\boldsymbol{h} = (h_j)_{j=1,\ldots}$. We consider a DTMC (X_n) with transition probability matrix $\boldsymbol{P}$ and initial state defined on $(\Omega, \mathbb{F}, \mathbb{P}_i)$.

Definition 2.9 We say that a function $\boldsymbol{h}$ such that $\sum_{j=1}^{\infty} |h_j| p_{ij} < \infty$ is *harmonic* if $\boldsymbol{Ph} = \boldsymbol{h}$, is *subharmonic* if $\boldsymbol{Ph} \geq \boldsymbol{h}$, and is *superharmonic* if $\boldsymbol{Ph} \leq \boldsymbol{h}$.

Suppose now (h_j) is a harmonic function such that $h_j > 0$ for all $j = 1, 2, \ldots$ and define

$$\tilde{p}_{ij} = \frac{h_j}{h_i} p_{ij}, \qquad i, j = 1, 2, \ldots$$

and let $\tilde{\boldsymbol{P}} = (\tilde{p}_{ij})_{i,j=1,\ldots}$.

Problem

2.1 Show that: (i) if $\boldsymbol{h}$ is a positive harmonic function, then $\tilde{\boldsymbol{P}}$ is a stochastic matrix matrix, (ii) if $\boldsymbol{h}$ is a positive superharmonic function, then $\tilde{\boldsymbol{P}}$ is a substochastic matrix.

Then the following result shows that nontrivial nonnegative superharmonic function exist only for transient DTMCs.

Proposition 2.10 *If a DTMC (X_n) is irreducible and recurrent, then a nonnegative superharmonic functions is constant. Similar statement is true for bounded subharmonic functions.*

Proof Asmussen (2003), p. 20.

Problems

2.1 Prove the following result. If a chain is irreducible with a nonnegative harmonic function $\boldsymbol{h}$, then $\boldsymbol{h} > \boldsymbol{0}$.

2.2 Show that a nonzero function $\boldsymbol{h}$ is harmonic if and only if the sequence of random variables (M_n) defined by $M_n = h_{X_n}/h_i$ is a nonnegative mean one $\mathbb{P}_i$-martingale. State a similar statement for subharmonic and superharmonic functions respectively (of course they cannot be mean one).

Suppose now that $h_j > 0$ for all $j = 1, 2, \ldots$ and define

$$M_n = \frac{h_{X_n}}{h_i}, \qquad n = 0, 1, \ldots$$

Lemma 2.11 *(M_n) is a positive, mean one $\mathbb{P}_i$-martingale with respect to the natural filtration $\mathcal{F}_n^X$.*

Proof Since $(X_n)_n$ is DTMC we have

$$\begin{aligned} \mathbb{E}\,(M_{n+1}|\mathcal{F}_n) &= \mathbb{E}\,(M_{n+1}|X_n) \\ = \sum_i 1(X_n = k)\frac{\sum_j h_j p_{kj}}{h_i} &= \frac{h_{X_n}}{h_i}. \end{aligned}$$

□

We may now define a new probability measure $\tilde{\mathbb{P}}_i$ by the martingale (M_n).

Lemma 2.12 *The sequence of random variables (X_n) on [1] $(\Omega, \mathbb{F}, \tilde{\mathbb{P}}_i)$ is a Markov chain defined by the initial distribution $\boldsymbol{\varepsilon}_i$ and transition probability matrix $\tilde{\boldsymbol{P}}$.*

Proof The result follows from the fact that for $i_1, \ldots, i_n \in E$ we have

$$\begin{aligned} \tilde{\mathbb{P}}_i(X_1 = i_1, \ldots, X_n = i_n) &= \mathbb{E}_i M_n 1(X_1 = i_1, \ldots, X_n) \\ = p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \frac{h_{i_n}}{h_i} &= \tilde{p}_{ii_1} \tilde{p}_{i_1 i_2} \cdots \tilde{p}_{i_{n-1} i_n}. \end{aligned}$$

□

Suppose now that $\boldsymbol{P} = (p_{ij})_{i,j=1,\ldots,l}$ is essentially substochastic and primitive. Let θ_{PF} and $\boldsymbol{h}$, $\boldsymbol{\pi}$ be the Perron-Frobenius eigenvalue and the corresponding right and left eigenvectors respectively. We define a stochastic

[1] filtered space is natural

$(l+1)\times(l+1)$ matrix $\boldsymbol{P}^*$ as in Section I.1.4 and let (X_n) be a Markov chain with state space $\{0,1\ldots,l\}$ with initial state $i\in\{1,\ldots,l\}$ and transition probability matrix $\boldsymbol{P}^*$ defined on $(\Omega,\mathbb{F},\mathbb{P}_i)$. As usual the filtration is natural; $\mathcal{F}_n=\sigma\{X_1,\ldots,X_n\}$ and $\mathcal{F}_0$ is trivial. Let $\nu=\min\{n:X_n=0\}$ be the time to reach 0. We define

$$M_n=(\theta_{\mathrm{PF}})^{-n}\frac{h_{X_n}}{h_i}1(\nu>n)$$

Problems

2.1 Show that (M_n) is a nonnegative mean one martingale.

We now define a new probability measure $\tilde{\mathbb{P}}_i$ by $\mathrm{d}\tilde{\mathbb{P}}_{i|n}=M_n\mathrm{d}\mathbb{P}_{i|n}$ $(n=0,1,\ldots$.

Problems

2.1 Show that (X_n) under $\tilde{\mathbb{P}}_i$ is a DTMC with state space $E=\{1,\ldots,l\}$ and with probability matrix $\tilde{\boldsymbol{P}}=(\tilde{p}_{ij})_{i,j=1,\ldots}$, where

$$\tilde{p}_{ij}=\theta_{\mathrm{PF}}^{-1}\frac{h_j}{h_i}p_{ij},$$

or in matrix notations

$$\tilde{\boldsymbol{P}}=\theta_{\mathrm{PF}}^{-1}\mathrm{diag}(\boldsymbol{h}^{-1})\boldsymbol{P}\mathrm{diag}(\boldsymbol{h}).$$

2.2 Show that there exists the stationary distribution $\boldsymbol{\pi}^*=(\pi_1,\ldots,\pi_l)$ for $\tilde{\boldsymbol{P}}$, that is fulfilling $\boldsymbol{\pi}^*\tilde{\boldsymbol{P}}=\boldsymbol{\pi}^*$, and that $\boldsymbol{\pi}^*\mathrm{diag}(\boldsymbol{h}^{-1})$ is the left Perron-Frobenius eigenvector (that is when normalised the quasi-stationary distribution). This means $\boldsymbol{\pi}$ is the left Perron-Frobenius vector, where

$$\pi_i=Ch_i\pi_i^*,$$

and $C^{-1}=\sum_{i=1}^l h_i\pi_i^*$.

2.3 Show that for $i_1,\ldots,i_m\in E$

$$\lim_{n\to\infty}\mathbb{P}_i(X_1=i_1,\ldots,X_m=i_m|\nu>m+n)=\tilde{\mathbb{P}}_i(X_1=i_1,\ldots,X_m=i_m).$$

(Hint. Write

$$\mathbb{P}_i(X_1=i_1,\ldots,X_m=i_m|\nu>m+n)=\frac{p_{ii_1}\cdots p_{i_{m-1}i_m}\sum_{j=1}^l p_{i_mj}^{(n)}}{\sum_{j=1}^l p_{ij}^{(n+m)}}$$

and use the SRLP.)

The DTMC (X_n) on $\{1,\ldots,l\}$ with transition function $\tilde{\boldsymbol{P}}$ is said to be *never exiting the subspace* $\{1,\ldots,l\}$ (NE) or *conditioned to stay in the subspace* $\{1,\ldots,l\}$.

Comments.

For details about martingales we refer to Rolski *et al* (1999). Theorem 2.8 is a discrete time version of Kunita-Watanabe theorem from Kunita and Watanabe (1963); see also Asmussen (2003), Ch. 13.3.

Chapter III

Random walk in applied probability

In this chapter we will study some aspects of the theory of random. Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables. The sequence (S_n), defined as

$$S_0 = 0, \qquad S_n = X_1 + \dots X_n$$

is called a *random walk*. Let $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\} = \sigma\{S_1, \dots, S_n\}$ for $n = 1, 2, \dots$ and $\mathcal{F}_0$ is the trivial σ-field.

We will quote now the following result, which will be useful later on. We have also quite an important result about fluctuation of the random walk.

Theorem 0.13 *Suppose that* $\mathbb{E}\,|X| < \infty$.
(i) *If* $\mathbb{E}\,X > 0$, *then* $\lim_{n\to\infty} S_n = \infty$ *a.s.*
(ii) *If* $\mathbb{E}\,X < 0$, *then* $\lim_{n\to\infty} S_n = -\infty$ *a.s.*
(iii) *Suppose that* X *is not degenerated at 0. If* $\mathbb{E}\,X = 0$, *then* $\limsup_{n\to\infty} S_n = \infty$ *a.s. and* $\liminf_{n\to\infty} S_n = -\infty$ *a.s.*

Proof (i) From the strong law of large numbers $S_n/n \to \mathbb{E}\,X > 0$ a.s. and hence the result follows.
(ii) Similar to (i).
(iii) Harder; see e.g. Rolski *et al* (1999). □

Problem

0.1 Assume now $\mathbb{E}\,|X| < \infty$. Show that if $\mathbb{E}\,X = 0$, then (S_n) is a martingale. Furthermore, if $\mathbb{E}\,X > 0$ ($\mathbb{E}\,X < 0$), then (S_n) is a submartingale (supermartingale).

A relation of the random to queues is shown in the next section.

1 GI/G/1 queue and random walk

We now define the *waiting time process* in GI/G/1 queue. We first give sample path notations. Customers arrive at instants $T_0, T_0+T_1, T_0+T_1+T_2, \ldots$ where (T_n) are *inter-arrival times*. If there is an arrival at 0, then we set $T_0 = 0$. The customer arriving at $T_0+T_1+\ldots+T_n$ brings its service times U_{n+1}. Let $X_n = U_n - T_n$. Let $T_0 = 0$. We assume that the service is in order of arrivals; such the discipline is said to First Come First Served (FCFS). We define the waiting time sequence (W_n) recursively by $W_{n+1} = (W_n + X_{n+1})_+$, $n = 0, 1, \ldots$.

We set $X_n = U_n - T_n$ and $S_0 = 0$, $S_n = X_1 + \ldots + X_n$.

Lemma 1.1 *For* $n = 1, 2, \ldots$

$$W_n = \max(W_0 + S_n, S_n - S_1, S_n - S_2, \ldots, S_n - S_{n-1}, 0) \tag{1.1}$$

Proof Induction with respect to $n = 1, \ldots$. For $n = 0$ is trivially true. Assume (1.1). Then

$$\begin{aligned} & W_n + X_{n+1} \\ = \ & \max(W_0 + X_{n+1} + S_n, X_{n+1} + S_n - S_1, X_{n+1} + S_n - S_2, \\ & \quad \ldots, X_{n+1} + S_n - S_{n-1}, X_{n+1}). \end{aligned}$$

Hence

$$\begin{aligned} & W_{n+1} = (W_n + X_{n+1})_+ \\ = \ & \max(W_0 + S_{n+1}, S_{n+1} - S_1, S_{n+1} - S_2, \ldots, S_{n+1} - S_n, 0). \end{aligned}$$

□

For the waiting time in GI/G/1 queue we assume

- $T_0 = 0$ a.s.
- $(T_n)_{n=1,\ldots}$ are i.i.d. independent of an i.i.d sequence $(U_n)_{n=1,\ldots}$,
- W_0, (T_n) and (U_n) are independent.

Let $\mathcal{M}_n = \max(0, S_1, \ldots, S_n)$ and $\mathcal{M} = \sup(0, S_1, \ldots)$

In the following proposition we make use of the following fact.

Lemma 1.2 *If $X_1, X_2, \dots$ are i.i.d., then*

$$(X_1+\ldots+X_n, X_2+\ldots+X_n, \ldots, X_n) =_{\mathrm{d}} (X_1+\ldots+X_n, X_1+\ldots+X_{n-1}, \ldots, X_1)$$

Proposition 1.3 *In the GI/G/1 queue, if $\mathbb{E}\, X < 0$, then*

$$W_n \xrightarrow{\mathrm{d}} \mathcal{M}.$$

Furthermore if $\mathcal{M}$ and X are independent, then

$$(\mathcal{M} + X)_+ =_{\mathrm{d}} \mathcal{M}.$$

Proof The condition yields $S_n \to -\infty$ a.s. Now, by Lemma 1.2, and because $S_n \to -\infty$ a.s. we write

$$\begin{aligned} W_n &= \max(W_0 + S_n, S_n - S_1, S_n - S_2, \ldots, S_n - S_{n-1}, 0) \\ &= \max(W_0 + X_1 + \ldots + X_n, X_2 + \ldots + X_n, \ldots, X_n, 0) \\ &=_{\mathrm{d}} \max(W_0 + X_1 + \ldots + X_n, X_1 + \ldots + X_{n-1}, \ldots, X_1, 0) \\ &\uparrow \mathcal{M}. \end{aligned}$$

For the second part we write

$$\begin{aligned} (\mathcal{M} + X)_+ &= (\max(X, X + X_1, X + X_1 + X_2, \ldots))_+ \\ &=_{\mathrm{d}} \mathcal{M}. \end{aligned}$$

□

In the theory of queues the stability condition $\mathbb{E}\, X < 0$ is expressed in terms of *traffic intensity* $\rho = \mathbb{E}\, U / \mathbb{E}\, T$ by $\rho < 1$.

Problems

1.1 Suppose that $F(x)$ is the distribution function of X and $F_{\mathcal{M}}(x)$ is the distribution function of $\mathcal{M}$. Show that $F_{\mathcal{M}}(x)$ fulfills the following integral equation, known under the name of *Wiener-Hopf* integral equation

$$F_{\mathcal{M}}(x) = \begin{cases} 0 & x < 0, \\ \int_{-\infty}^{x} F_{\mathcal{M}}(x-y)\, F(dy) & x \geq 0 \end{cases}$$

We close this section with a classification of GI/G/1 FCFS queues:

- if T is exponentially distributed $\mathrm{Exp}(a)$ and U is exponentially distributed $\mathrm{Exp}(b)$, then the system is denoted by M/M/1 FCFS;
- if T is exponentially distributed $\mathrm{Exp}(a)$ and U has a general distribution B, then the system is denoted by M/G/1 FCFS,
- if T has a general distribution A and U is exponentially distributed $\mathrm{Exp}(b)$, then the system is denoted by GI/M/1 FCFS.

2 Basic facts

In this section we recall needed facts from the theory of random walk.

2.1 Ladder Epochs

Consider the random walk $\{S_n\}$. We first define the generic *(strong) ascending ladder epoch* by

$$\nu^+ = \min\{n > 0 : \ S_n > 0\}, \tag{2.1}$$

setting $\nu^+ = \infty$ if $S_n \le 0$ for all $n \in \mathbb{Z}_+$. Similarly we define half-line $(-\infty, 0]$ by

$$\nu^- = \min\{n > 0 : \ S_n \le 0\}, \tag{2.2}$$

setting $\nu^- = \infty$ if $S_n > 0$ for all $n = 1, 2, \ldots$, and call ν^- the generic *(weak) descending ladder epoch* of $\{S_n\}$. As we will see later, we need to know whether $\mathbb{E}\, X$ is strictly positive, zero or strictly negative, as otherwise we cannot say whether ν^+ or ν^- are proper. We will drop writing *weak* in the sequel.

For each $k = 1, 2, \ldots$, the events

$$\{\nu^+ = k\} = \{S_1 \le 0, S_2 \le 0, \ldots, S_{k-1} \le 0, S_k > 0\} \tag{2.3}$$

and

$$\{\nu^- = k\} = \{S_1 > 0, S_2 > 0, \ldots, S_{k-1} > 0, S_k \le 0\} \tag{2.4}$$

are determined by the first k values of $\{S_n\}$. We have $\{\nu^+ = k\} \in \mathcal{F}_k$ and $\{\nu^- = k\} \in \mathcal{F}_k$ for $k \in \mathbb{Z}_+$. This means that the ladder epochs ν^+ and ν^- are *stopping times* From Corollary II.1.8 proved there we have that, for each stopping time ν with respect to $\{\mathcal{F}_n\}$,

$$\mathbb{E}\, S_\nu = \mathbb{E}\, \nu \mathbb{E}\, X \tag{2.5}$$

provided that $\mathbb{E}\,\nu < \infty$ and $\mathbb{E}\,|X| < \infty$, which is known as *Wald's identity* for stopping times.

Actually, we can recursively define a sequence of ladder epochs $\{\nu_n^+, n \in \mathbb{Z}_+\}$ by

$$\nu_{n+1}^+ = \min\{j > \nu_n^+ : S_j > S_{\nu_n^+}\}, \tag{2.6}$$

where $\nu_0^+ = 0$ and $\nu_1^+ = \nu^+$and call ν_n^+ the n-th (strong ascending) *ladder epoch.* A priori, we cannot exclude the case that, from some random index on, all the ladder epochs are equal to ∞.

In a similar way, we recursively define the sequence $\{\nu_n^-, n \in \mathbb{Z}_+\}$ of consecutive *(weak) descending ladder epochs* by $\nu_0^- = 0$, $\nu_1^- = \nu^-$ and

$$\nu_{n+1}^- = \min\{j > \nu_n^-, S_j \le S_{\nu_n^-}\}\,, \qquad n = 1, 2, \ldots. \tag{2.7}$$

Note that it is possible to define dually: the sequence of weak ascending ladder epochs and strong descending ladder epochs, but we do not need it here.

2.2 Ladder Heights

A basic characteristic of $\{S_n\}$ is then the first ascending ladder epoch ν^+. As one can expect, and we confirm this in Theorem 2.1, the distribution of the random variable ν^+ is defective under the assumption of a negative drift. The *overshoot* X^+ above the zero level is defined by

$$X^+ = \begin{cases} S_{\nu^+} & \text{if } \nu^+ < \infty, \\ \infty & \text{otherwise} \end{cases}$$

and is called the generic *(strong) ascending ladder height.*

We have the following result for $G^+(x) = \mathbb{P}(X^+ \le x)$, the distribution function of X^+ and $G^+(\infty) = \lim_{x\to\infty} G^+(x)$.

Problem

2.1 Show that the following statements are equivalent:
 (a) $\mathbb{E}\,X < 0$,
 (b) $\mathcal{M}$ is finite with probability 1,
 (c) $G^+(\infty) < 1$.

Suppose that $\nu^+ < \infty$. We can then repeat the same argument as above, but now from the point (ν^+, X^+), because of our assumption that the increments $X_1, X_2, \ldots$ of the random walk $\{S_n\}$ are independent and identically distributed. This means in particular, that we can define a new random walk $S_{\nu^++1} - S_{\nu^+}, S_{\nu^++2} - S_{\nu^+}, \ldots$ which can be proved to be an identically distributed copy of the original random walk $\{S_n\}$ and independent of $S_1, S_2, \ldots, S_{\nu^+}$. We leave it to the reader to show this. Iterating this procedure, we can recursively define the sequence $\{\nu_n^+\}$ of consecutive ladder epochs in the same way as this was done in (2.6). The random variable

$$X_n^+ = \begin{cases} S_{\nu_n^+} - S_{\nu_{n-1}^+} & \text{if } \nu_n^+ < \infty, \\ \infty & \text{otherwise} \end{cases}$$

is called the n-th *(strong) ascending ladder height* of $\{S_n\}$. It is not difficult to show that the sequence $\{X_1^+ + \ldots + X_n^+, n = 0, 1, \ldots\}$ forms a terminating renewal process, provided $\mathbb{E}\, X < 0$ and we will assume this condition to be fulfilled for a while. Moreover, for the maximum we have

$$\mathcal{M} = \sum_{i=1}^{N} X_i^+, \tag{2.8}$$

where $N = \max\{n : \nu_n^+ < \infty\}$ is the number of finite ladder epochs. Thus, with the notation $G_0(x) = G^+(x)/G^+(\infty)$, where $G_0(x)$ is a proper (i.e. non-defective) distribution function, we arrive at the following result, saying that $\mathcal{M}$ has a compound geometric distribution.

Theorem 2.1 *If $\mathbb{E}\, X < 0$, then for all $x \geq 0$ and for $p = G^+(\infty)$*

$$\mathbb{P}(\mathcal{M} \leq x) = (1-p) \sum_{k=0}^{\infty} (G^+)^{*k}(x) = \sum_{k=0}^{\infty} (1-p) p^k G_0^{*k}(x)\,. \tag{2.9}$$

Proof Recall that $\{X_1^+ + \ldots + X_n^+, n = 0, 1, \ldots\}$ is a terminating renewal process and the distribution of $N = \max\{n : \nu_n^+ < \infty\}$ is geometric with parameter $p = G^+(\infty)$, i.e. $\mathbb{P}(N = k) = (1-p)p^k$ for $k = 0, 1, \ldots$. Thus, using (2.8) we have

$$\mathbb{P}(\mathcal{M} \leq x) = \mathbb{P}\Big(\sum_{i=1}^{N} X_i^+ \leq x\Big) = \sum_{k=0}^{\infty} (1-p) p^k G_0^{*k}(x) = \sum_{k=0}^{\infty} (1-p)(G^+)^{*k}(x)\,.$$

This completes the proof. □

We now introduce the dual notions of (weak) descending ladder heights. Consider the generic descending ladder epoch ν^-. The *undershoot* X^- below the zero level is defined by $X^- = S_{\nu^-}$ and called the (first) *descending ladder* *height*. The n-th descending ladder height is defined by $X_n^- = S_{\nu_n^-} - S_{\nu_{n-1}^-}$. Since $X_1^-, \ldots, X_n^-$ are independent and identically distributed copies of X^-, it is clear that the sequence $\{-\sum_{i=1}^n X_i^-,\ n \in \mathbb{Z}_+\}$ is a nonterminating renewal process (in the case of the negative drift). Indeed, under our assumption on the negative drift it follows from Theorem 0.13 that all descending ladder epochs and heights are proper random variables.

Comments. The basic references are Feller (1971) and Chung (1974), Resnick (1992).

2.3 The Wiener–Hopf Factorisation; General Representation Formulae

Define the ladder height distribution G^-, concentrated on $\mathbb{R}_-$, by

$$G^-(x) = \mathbb{P}(X^- \le x)\,, \qquad x \in \mathbb{R}\,. \tag{2.10}$$

Thus G^- dualizes the ladder height distribution G^+ which is concentrated on $(0,\infty)$ and is given by

$$G^+(x) = \mathbb{P}(X^+ \le x)\,, \qquad x \in \mathbb{R}\,. \tag{2.11}$$

Let H_0^- be the measure on $\mathbb{R}_-$ given by

$$H_0^-(B) = \sum_{k=0}^{\infty} (G^-)^{*k}(B)\,, \qquad B \in \mathcal{B}(\mathbb{R}_-)\,. \tag{2.12}$$

We also introduce as a dual measure H_0^+ on $\mathbb{R}_+$

$$H_0^+(B) = \sum_{k=0}^{\infty} (G^+)^{*k}(B)\,, \qquad B \in \mathcal{B}(\mathbb{R}_+)\,. \tag{2.13}$$

Moreover, from (2.12) it follows that

$$H_0^- * G^- = H_0^- - \delta_0\,. \tag{2.14}$$

It turns out that H_0^- is equal to the so-called *pre-occupation measure* γ^- given by

$$\gamma^-(B) = \mathbb{E}\Big(\sum_{i=0}^{\nu^+-1} 1(S_i \in B)\Big)$$

for $B \in \mathcal{B}(\mathbb{R})$, where obviously $\gamma^-(B) = 0$ for $B \subset (0,\infty)$.

Lemma 2.2 *For each $B \in \mathcal{B}((-\infty,0])$ and we have $H_0^-(B) = \gamma^-(B)$.*

Proof Note that $(G^-)^{*0}(B) = 1(0 \in B) = \mathbb{P}(S_0 \in B)$ and $(G^-)^{*n}(B) = \mathbb{P}(X_1^- + \ldots + X_n^- \in B) = \mathbb{P}(S_{\nu_n^-} \in B)$ for all $n = 1, 2, \ldots$. Thus

$$\begin{aligned}
H_0^-(B) &= \sum_{n=0}^{\infty}(G^-)^{*n}(B) = 1(0 \in B) + \sum_{n=1}^{\infty}\mathbb{P}(S_{\nu_n^-} \in B) \\
&= 1(0 \in B) + \mathbb{E}\Big(\sum_{n=1}^{\infty} 1(S_{\nu_n^-} \in B)\Big) \\
&= 1(0 \in B) + \mathbb{E}\Big(\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} 1(\nu_n^- = k)1(S_{\nu_n^-} \in B)\Big) \\
&= 1(0 \in B) + \sum_{k=1}^{\infty}\mathbb{E}\Big(\sum_{n=1}^{\infty} 1(\nu_n^- = k, S_{\nu_n^-} \in B)\Big).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{n=1}^{\infty} 1(\nu_n^- = k, S_{\nu_n^-} \in B)\Big) = \mathbb{P}(S_k \le S_i,\ i = 0, \ldots, k-1,\ S_k \in B) \\
&= \mathbb{P}\Big(\sum_{i=1}^{k} X_i \le 0, \sum_{i=2}^{k} X_i \le 0, \ldots, X_k \le 0, \sum_{i=1}^{k} X_i \in B\Big) \\
&= \mathbb{P}\Big(X_1 \le 0, \ldots, \sum_{i=1}^{k} X_i \le 0, \sum_{i=1}^{k} X_i \in B\Big).
\end{aligned}$$

This gives

$$\begin{aligned}
H_0^-(B) &= 1(0 \in B) + \sum_{k=1}^{\infty}\mathbb{P}(S_i \le 0, i = 0, \ldots, k, S_k \in B) \\
&= \mathbb{E}\Big(\sum_{k=0}^{\infty} 1(\nu^+ \ge k+1, S_k \in B)\Big).
\end{aligned}$$

Thus, the proof is complete since

$$\mathbb{E}\Big(\sum_{k=0}^{\infty} 1(\nu^+ \ge k+1, S_k \in B)\Big) = \mathbb{E}\Big(\sum_{k=0}^{\nu^+-1} 1(S_k \in B)\Big).$$

□

Next we show that the distribution F of the increments $X_1, X_2, \ldots$ of the random walk $\{S_n\}$ can be expressed in terms of the ladder height distributions G^+ and G^-. This is the so-called *Wiener–Hopf factorisation* of F.

Theorem 2.3 *The following relationship holds:*

$$F = G^+ + G^- - G^- * G^+. \tag{2.15}$$

Proof We first show that

$$\delta_0 + \gamma^- * F = \gamma^- + G^+. \tag{2.16}$$

Let $B \in \mathcal{B}(\mathbb{R})$ be an arbitrary Borel set. Then,

$$\sum_{n=0}^{\nu^+-1} \mathbf{1}(S_n \in B) + \mathbf{1}(S_{\nu^+} \in B) = \mathbf{1}(0 \in B) + \sum_{n=0}^{\nu^+-1} \mathbf{1}(S_{n+1} \in B)\,. \tag{2.17}$$

Now $\mathbb{E}\left(\sum_{n=0}^{\nu^+-1} \mathbf{1}(S_{n+1} \in B)\right) = \mathbb{E}\left(\sum_{n=0}^{\infty} \mathbf{1}(\nu^+ > n, S_{n+1} \in B)\right)$ and, since the event $\{\nu^+ > n\}$ is independent of X_{n+1}, the above equals

$$\sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}(\nu^+ > n, S_n \in B - y)\, F(\mathrm{d}y) = F * \gamma^-(B)\,.$$

Taking the expected value of both sides of (2.17), we get (2.16). Convoluting both sides of (2.16) with G^- we obtain

$$G^- + G^- * \gamma^- * F = G^- * \gamma^- + G^- * G^+.$$

On the other hand, by (2.14) and Lemma 2.2, we have $\gamma^- * G^- = \gamma^- - \delta_0$. Thus, $G^- + (\gamma^- - \delta_0) * F = \gamma^- - \delta_0 + G^- * G^+$ and, equivalently, $F = G^- - G^- * G^+ + \gamma^- * F - \gamma^- + \delta_0$. Using (2.16) again, this gives (2.15). □

If we want to compute ruin probabilities, we need to determine the ladder height distribution G^+ that appears in Theorem 2.1. The Wiener–Hopf factorisation (2.15) yields the following representation formula for G^+.

Corollary 2.4 *For* $B \in \mathcal{B}((0,\infty))$,

$$G^+(B) = F * H_0^-(B) = \int_{-\infty}^{0} F(B - y)\, \mathrm{d}H_0^-(y)\,, \tag{2.18}$$

while for $B \in \mathcal{B}(\mathbb{R}_-)$

$$G^-(B) = F * H_0^+(B) = \int_0^\infty F(B-y)\,\mathrm{d}H_0^+(y)\,. \tag{2.19}$$

Proof Convoluting both sides of the Wiener–Hopf factorisation (2.15) by G^-, we obtain $F * G^- = G^+ * G^- + G^- * G^- - G^+ * (G^-)^{*2}$. Iterating this procedure we get $F * (G^-)^{*k} = G^+ * (G^-)^{*k} + (G^-)^{*(k+1)} - G^+ * (G^-)^{*(k+1)}$ for each $k = 1, 2, \ldots$. Summing over k from 0 to n we obtain

$$F * \sum_{k=0}^{n} (G^-)^{*k}(B) = G^+(B) - G^+ * (G^-)^{*(n+1)}(B)\,,$$

for $B \in \mathcal{B}((0,\infty))$. This completes the proof, because $\lim_{n\to\infty}(G^-)^{*n}(B) = 0$ for all $B \in \mathcal{B}((0,\infty))$ and hence $\lim_{n\to\infty} G^+ * (G^-)^{*(n+1)}(B) = 0$. The proof of (2.19) is similar. □

2.4 An analytical form of Wiener-Hopf factorisation

Throughout this section we suppose that $\mathbb{E}\,X < 0$. From Theorem 2.1 we can easily draw the following general result.

Corollary 2.5 (a) *For $s \le 0$,*

$$\hat{m}_{\mathcal{M}}(s) = \frac{1 - G^+(\infty)}{1 - \hat{m}_{G^+}(s)}\,. \tag{2.20}$$

The *proof* is obvious because (2.20) directly follow from (2.9). □

Recall that the moment generating functions of the distributions G^+, G^- and F can be defined as functions of a complex variable $z \in \mathbb{C}$. Since G^+ is concentrated on $(0,\infty)$, the moment generating function $\hat{G^+}(z)$ is well-defined on the half-plane $\Re(z) \le 0$. Analogously, since G^- is concentrated on $\mathbb{R}_-$, its moment generating function $\hat{G^-}(z)$ is well-defined on the half-plane $\Re(z) \ge 0$. The moment generating function $\hat{m}_F(z)$ is well-defined at least on the imaginary axis $\Re(z) = 0$ because each point on $\Re(z) = 0$ can be represented as $z = \mathrm{i}t$ for some real t and then $\hat{F}(z) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}F(x)$, which is the characteristic function of F. An immediate consequence of the Wiener–Hopf factorisation (2.15) is the following *analytical factorisation* of the corresponding moment generating functions.

Corollary 2.6 *If for some $z \in \mathbb{C}$ all the moment generating functions $\hat{F}(z)$, $\hat{G^+}(z)$, $\hat{G^-}(z)$ exist, in particular if $\Re(z) = 0$, then*

$$1 - \hat{F}(z) = (1 - \hat{G^+}(z))(1 - \hat{G^-}(z)) . \tag{2.21}$$

We conclude this section with a few comments on a probabilistic solution of the Wiener–Hopf factorisation (2.21). Suppose that for some $\varepsilon > 0$ the function $\hat{F}(z)$ is well-defined in $1 - \varepsilon \leq \Re z \leq 1 + \varepsilon$. Define the functions

$$d^+(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n > 0]}{n}\Big) \tag{2.22}$$

and

$$d^-(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n \leq 0]}{n}\Big). \tag{2.23}$$

It can be shown (see, for example, Prabhu (1980)) that, in $\Re(z) < 1 + \varepsilon$, $d^+(z)$ is analytic, bounded and bounded away from zero and that $d^-(z)$ has the same properties in $\Re(z) > 1 - \varepsilon$. Moreover, $d^+(z)$, $d^-(z)$ are separated from zero and $d^+(z) \to 1$ for $\Im(z) \to \infty$. Show that

$$1 - \hat{F}(z) = d^+(z)d^-(z) , \qquad 1 - \varepsilon \leq |z| \leq 1 + \varepsilon . \tag{2.24}$$

Hence,

$$1 - \hat{G^+}(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n > 0]}{n}\Big) \tag{2.25}$$

and

$$1 - \hat{G^-}(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n \leq 0]}{n}\Big) , \tag{2.26}$$

since the factorisation given in (2.24) is unique within the class of functions satisfying the same conditions as mentioned above for $d^+(z)$ and $d^-(z)$. See e.g. Prabhu [15], p. 50.

Problems

2.1 Let $H_0^+ = \sum_{k=0}^{\infty}(G^+)^{*k}$ and define the pre-occupation measure γ^+ by $\gamma^+(B) = \mathbb{E}\sum_{n=0}^{\nu^- -1} 1(S_n \in B)$ for $B \in \mathcal{B}(\mathbb{R})$. Show that $\gamma^+ = H_0^+$.

2.2 Assume that $\hat{F}(s) < \infty$ for some $s > 0$. Show that then $\hat{m}_{G^+}(s) < \infty$. [Hint. Use Corollary 2.4 and (2.12) showing first that

$$\int_{-\infty}^{\infty} \mathrm{e}^{sy} H_0^-(\mathrm{d}y) = \sum_{k=0}^{\infty}\Big(\int_{-\infty}^{0} \mathrm{e}^{sy} G^-(\mathrm{d}y)\Big)^k < \infty . \quad]$$

2.3 Show that if $F_T = \mathrm{Exp}(\lambda)$ and $F_U = \Gamma(n,\delta)$ for some $n = 2,3,\ldots$, then $G_0 = n^{-1}\sum_{k=0}^{n-1}\Gamma(n-k,\delta)$.

2.5 The overshoot process

In this part we consider random walk with posisitve drift, that is $\mathbb{E}\,X > 0$ and Let $\nu(u) = \min\{n :\ S_n > u\}$. Define the so called *overshoot process* by

$$B(u) = S_{\nu(u)} - u, \quad u \geq 0.$$

Upon an inspection we may write the following representation for $(B(t))$ in terms of the ladder heights; let $X_1^+, X_2^+, \ldots$ be consecutive i.i.d. ladder heights. They are, due to our assumption that $\mathbb{E}\,X > 0$, true nonnegative random variables.

Problem

2.1 Show that

$$\begin{aligned} B(t) &= Y_1^+ \ldots + X_{n+1}^+ - t, \\ &\quad \text{if for some } n = 0,1,\ldots \quad X_1^+ \ldots + X_n^+ \leq t < X_1^+ \ldots + X_{n+1}^+ \end{aligned}$$

Such the process is called a *forward recurrent time* process. We say that distribution F is *lattice* if it is concentrated on a set of form $\{na : n \in \mathbb{Z}\}$ otherwise we say *nonlattice*. We quote now the following result.

Proposition 2.7 *Suppose that F is nonlattice and $\mathbb{E}\,X^+ < \infty$. Then*

$$B(t) \xrightarrow{\mathrm{d}} G^{+,\mathrm{s}}$$

as $t \to \infty$, where

$$G^{+,\mathrm{s}}(t) = \frac{\int_t^\infty (1 - G^+(v)\ \mathrm{d}v}{\mathbb{E}\,X^+}$$

Proof See e.g. Asmussen (2003), Theorem 4.3 on page 155.

3 Wald martingale for random walk and its applications

Consider a random walk (S_n) with $S_n = \sum_{i=1}^n X_i$, where $X_1, X_2, \ldots$ are independent and identically distributed random variables with with the distribution F and $\mathbb{E}|X| < \infty$. The moment generating function of F is $\hat{F}$ and $\mathcal{I}_{\hat{F}} = \{s \in \mathbb{R} : \hat{F}(s) < \infty\}$

An important role in these notes will play the following martingale, called *Wald martingale*:

$$M_n^{(s)} = \frac{e^{sS_n}}{(\hat{F}(s))^n}, \qquad n = 0, 1, \ldots$$

where $s \in \mathcal{I}_{\hat{F}}$. To check that $(M_n^{(s)})$ is indeed a martingale, we notice that $M_n^{(s)}$ is $\mathcal{F}_n$ measurable, for n=1,2,...

$$\mathbb{E}\,|M_n| = \mathbb{E}\,M_n = \mathbb{E}\,\frac{e^{sS_n}}{(\hat{F}(s))^n} = 1$$

(of course $\mathbb{E}\,M_0 = 1$). Next we demonstrate that

$$\begin{aligned} \mathbb{E}\,[M_{n+1}^{(s)}|\mathcal{F}_n] &= \mathbb{E}\left[\frac{e^{s(S_n+X_{n+1})}}{(\hat{F}(s))^{n+1}} \mid \mathcal{F}_n\right] \\ = \frac{e^{sS_n}}{(\hat{F}(s))^n}\mathbb{E}\left[\frac{e^{sX_{n+1}}}{(\hat{F}(s))}\right] &= M_n^{(s)} \end{aligned}$$

We can rewrite the Wald martingale in the form

$$M_n^{(s)} = e^{sS_\nu - n\kappa(s)}$$

where

$$\kappa(s) = \log \hat{F}(s). \tag{3.27}$$

Problems

3.1 Show that $\kappa(s)$ is a strictly convex function in $\mathcal{I}_{\hat{F}}$, such that $\kappa(0) = 0$. Demonstrate that the following scenarios are possible:

1. $\kappa(s) = \infty$ for $s \neq 0$,
2. if $\mathbb{E}\,X \geq 0$, then $\kappa(s) > 0$ for $s > 0$,
3. if $\mathbb{E}\,X < 0$, then

   $\kappa(s) = \infty$ for $s > 0$,

   (i) there exists $\gamma > 0$ such that $\kappa(\gamma) = 0$; in this case the γ is unique.

   (ii) $\kappa(s) < 0$ for $0 < s < s_+$ and $\kappa(s) = \infty$ for $s > s_+$.

Suppose now that for $u > 0$ we define

$$\nu = \nu(u) = \min\{n : S_n > u\}$$

Problems

3.1 Show that $\nu(u)$ is a stopping time. Demonstrate that $\nu(u) < \infty$ a.s. provided $\mathbb{E}\, X \geq 0$.

If $\mathbb{E}\, X \geq 0$ and is nondegenerated at 0, then by Theorem 0.13 we can infer that $\mathbb{P}(\nu < \infty) = 1$, otherwise it turns out that $\mathbb{P}(\nu < \infty)$ is less than 1, which makes the study of this probability, called probability of ruin, interesting.

We choose $s = \gamma$ such that $\hat{F}(\gamma) = 1$ (we tacitly assume that $\gamma > 0$ exists, which is not always the case as we can see from Exercise 3.1). If such the $\gamma > 0$ exists, then it is called *adjustment coefficient.* We make the following important observation. For $s = \gamma$

$$M_n^{(\gamma)} = e^{\gamma S_n}, \qquad n = 0, 1, \ldots$$

is a martingale. Recall that $\mathbb{E}\, M_n^{(\gamma)} = 1$ The following is a version of *Lundberg inequality*

Proposition 3.1 *For all $u > 0$*

$$\mathbb{P}(\sup_{n\geq 0} S_n > u) \leq e^{-\gamma u}.$$

Proof Let $M_n^{(\gamma)} = e^{\gamma S_n}$ Observe that for all $m = 1, 2, \ldots$

$$\{\sup_{0\leq n\leq m} S_n > u\} = \{\sup_{0\leq n\leq m} e^{\gamma S_n} > e^{\gamma u}\} = \{\sup_{0\leq n\leq m} M_n > e^{\gamma u}\}.$$

From Doob's inequality

$$\mathbb{P}(\sup_{0\leq n\leq m} M_n > e^{\gamma u}) \leq \frac{\mathbb{E}\, M_m}{e^{\gamma u}} = e^{-\gamma u}.$$

Since $\{\sup_{0\leq n} S_n > u\} = \bigcup_m \{\sup_{0\leq n\leq m} S_n > u\}$, and $\{\sup_{0\leq n\leq m} S_n > u\}$ is an ascending sequence of events, we may write

$$\begin{aligned}
\mathbb{P}(\sup_{0\leq n} S_n > u) &= \lim_{m\to\infty} \mathbb{P}(\sup_{0\leq n\leq m} S_n > u) \\
\lim_{m\to\infty} \mathbb{P}(\sup_{0\leq n\leq m} M_n > e^{\gamma u}) &\leq \frac{\mathbb{E}\, M_m}{e^{\gamma u}} \\
&= e^{-\gamma u}.
\end{aligned}$$

□

We now make the change of measure $\mathbb{P}$ by the Wald martingale

$$M_n^{(\gamma)} = \frac{e^{\gamma S_n}}{(\hat{F}(\gamma))^n} = \begin{cases} 1 & n = 0 \\ \frac{e^{\gamma X_1}}{\hat{F}(\gamma)} \times \ldots \times \frac{e^{\gamma X_n}}{\hat{F}(\gamma)} & n \geq 1 \end{cases} \tag{3.28}$$

according to formula (II.2.27) to define $\mathbb{P}^{(s)}$. We assume, of course, that the standard setup holds for the sequence (X_n). Define

$$F^{(s)}(B) = \frac{\int_B e^{sx}\, F(dx)}{\hat{F}(s)}$$

which is a proper distribution on $\mathbb{R}$ for such s that the m.g.f. is finite. Sometimes $F^{(s)}$ is said to be an *associated distribution.* We can see that $(M_n^{(s)})$ is a likelihood sequence if $\tilde{\mathbb{P}}$ is defined by $\tilde{F} = F^{(s)}$. In this case we denote $\tilde{\mathbb{P}} = \mathbb{P}^{(s)}$.

Example 3.2 (i) Suppose that $X_i = 1$ with probability p and $X_i = -1$ with probability $1-p$. The m.g.f. of X is

$$\hat{F}(s) = e^s p + e^{-s}(1-p)$$

and let $\mathbb{P}^{(s)}$ be the underlying probability measure after the ECM by the Wald martingale. We have then that

$$\mathbb{P}^{(s)}(X = 1) = \frac{pe^s}{e^s p + e^{-s}(1-p)} = \tilde{p},$$

and

$$\mu^{(s)} = \mathbb{E}^{\,s} X = \frac{e^s p - e^{-s}(1-p)}{e^s p + e^{-s}(1-p)}$$

Hence, for example the random walk S_n under $\mathbb{P}^{(s)}$ is driftless, if $e^s p = e^{-s}(1-p)$, which yields

$$s = \log\left(\frac{1-p}{p}\right)^{1/2}.$$

We now compute the adjustment coefficient γ that is the solution of

$$e^{\gamma} p + e^{-\gamma}(1-p) = 1.$$

Substituting $x = e^\gamma$ the above reduces to $px^2 - x + (1-p) = 0$ which has solutions $x_1 = 1$ and $x_2 = (1-p)/p$. Since it is required $\mathbb{E}\, X < 0$ which is equivalent $p < 1/2$, then $\gamma = \log(1-p) - \log p > 0$.
(ii) Consider a sequence $X_1, \ldots$ and suppose that $X = U - T$, where U and T are independent random variables $U \sim$ Exp(b) and $T \sim$ Exp(a). Change the measure by the Wald martingale $M_n^{(s)} = \exp(sS_n)/(\hat{F}(s))^n$.

Problems

3.1 Under the framework of Example 3.2 (ii), show that (X_i) is again a sequence of i.i.d. random variables with $X = U - T$, where U and T are independent random variables $U \sim$ Exp$(b-s)$ and $T \sim$ Exp$(a+s)$. Moreover $\gamma = b - a$ and $s^* = (b-a)/2$.

3.2 Show that the Wald martingale $M_n^{(s)} = e^{sS_n}/(\hat{F}(s))^n$ is a likelihood ratio martingale.

Let for $s \in \mathcal{I}_{\hat{F}}$

$$\hat{F}^{(s)}(v) = \mathbb{E}^{(s)} e^{vX} = \int_{-\infty}^{\infty} e^{vx}\, F^{(s)}(\mathrm{d}x)$$

be the m.g.f. of $F^{(s)}$.

Lemma 3.3 *We have the following formulas for the moment generating function, mean and variance of* $F^{(s)}$

$$\begin{aligned}
\hat{F}^{(s)}(v) &= \mathbb{E}^{(s)} e^{vX} = \frac{\hat{F}(v+s)}{\hat{F}(s)}, \qquad v \in (\mathcal{I}_{\hat{F}} - s) \\
\mu^{(s)} &= \mathbb{E}^{(s)} X = \frac{\hat{F}'(s)}{\hat{F}(s)}, \\
(\sigma^{(s)})^2 &= \mathrm{Var}^{(s)} X = \frac{\hat{F}''(s)\hat{F}(s) - (\hat{F}'(s))^2}{(\hat{F}(s))^2}.
\end{aligned}$$

Proof We have

$$\begin{aligned}
\hat{F}^{(s)}(v) &= \frac{\int_{-\infty}^{\infty} e^{vx} e^{sx}\, F(\mathrm{d}x)}{\hat{F}(s)} \\
&= \frac{\int_{-\infty}^{\infty} e^{(v+s)x}\, F(\mathrm{d}x)}{\hat{F}(s)} \\
&= \frac{\hat{F}(v+s)}{\hat{F}(s)}.
\end{aligned}$$

Now the mean and variance can be computed by finding the first and second derivative of the logarithmic moment generating function

$$\log(\hat{F}^{(s)}(v)) = \log(\hat{F}(v+s)) - \log(\hat{F}(s)).$$

$\square$

Corollary 3.4 *Let $\mathbb{E}\, X < 0$ and suppose that there exists $\gamma > 0$ such that $\hat{F}(\gamma) = 1$. Function $\mathbb{E}^{(s)} X$ for $0 \le s \le \gamma$ increase continuously from strictly negative to strictly positive value. In particular for $0 < s^* < \gamma$ fulfilling $\hat{F}'(s^*) = 0$ we have $\mathbb{E}^{(s^*)} X = 0$.*

Corollary 3.5 *Suppose, for a given x, there exists the solution of*

$$x = \frac{\hat{F}'(s(x))}{\hat{F}(s(x))}.$$

Then $\mu^{(s(x)} = x$. In particular if for some s^* we have $\hat{F}'(s^*) = 0$, Then $\mathbb{E}^{(s^*)} X = 0$ and so under $\mathbb{P}^{s^*}$ the random walk (S_n) is driftless.

Problems

3.1 Show an example that there exists s^* such that $\mathbb{E}^{(s^*)} X = 0$ although there is no $\gamma > 0$ such that $\hat{F}(\gamma) = 1$.

The basic identity can be now expressed as follows: if ν is a finite $\mathbb{P}^{(s)}$-a.s. stopping time and $A \in \mathcal{F}_\nu$, then

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{E}^{(s)}[\frac{e^{-s(X_1+\ldots+X_\nu)}}{(\hat{F})^{-\nu(s)}}; A] \\ &= \mathbb{E}^{(s)}[e^{-sS_\nu+\nu\kappa(s)}; A], \end{aligned} \tag{3.29}$$

Let $\mu = \mathbb{E}\, X$. We now show a proof of *Chernov bound* for an inequality for $\mathbb{P}(A_n)$, where

$$A_n = \{S_n > (\mu+\epsilon)n\}.$$

Proposition 3.6 *If $\mu > 0$*

$$\mathbb{P}(S_n > (\mu+\epsilon)n) \le e^{-nI}, \tag{3.30}$$

where $I = s_0(\mu+\epsilon) - \kappa(s_0) > 0$, and s_0 is the solution (we assume that it exists) of

$$\kappa^{'}(s_0) = \frac{\hat{F}^{'}(s_0)}{\hat{F}(s_0)} = \mu + \epsilon.$$

Recall that $\kappa(s)$ was defined in (3.27).

Proof The relevant choice of s is fulfilling

$$\mathbb{E}^{(s)}X = \frac{\hat{F}^{'}(s)}{\hat{F}(s)} = \mu + \epsilon.$$

Let $I = s(\mu+\epsilon) - \kappa(s) > 0$. Using the basic identity (3.29) with $\nu = n$ we have for all $s \in \mathcal{I}_{\hat{F}}$

$$\begin{aligned}
\mathbb{P}(S_n > (\mu+\epsilon)n) &= \mathbb{E}^{(s)}[e^{-sS_n + n\kappa(s)}; S_n > (\mu+\epsilon)n] \\
&= e^{-n(s(\mu+\epsilon)-\kappa(s))}\mathbb{E}^{(s)}s[e^{-s(S_n - n(\mu+\epsilon))}; S_n > (\mu+\epsilon)n] \\
&\leq e^{-n(s(\mu+\epsilon)-\kappa(s))}
\end{aligned}$$

We now look for the minimum with respect to s of the right hand side which is achieved at s_0. □

Lemma 3.7 *Suppose that distribution F of the increment X is nonlattice and $\hat{F}(s) < \infty$ for $s < \gamma + \epsilon$, where $\epsilon > 0$. Then under $\mathbb{P}^{(\gamma)}$ we have $B(u) \stackrel{\mathrm{d}}{\to} B(\infty)$, where*

$$\mathbb{P}^{(\gamma)}(B(\infty) > t) = \frac{\int_t^\infty e^{\gamma v}\, G^+(\mathrm{d}v)}{\mathbb{E}^{(\gamma)}X^+}\,.$$

Our aim is to prove the following celebrated theorem *Cramer-Lundberg approximation*, wherein we compute the probability of *ultimate ruin*, that is $\nu(u) < \infty$. Note that the event $\{\sup_{n\geq 0} S_n > u\}$ equals to the event $\{\nu(u) < \infty\}$. Moreover $\{\nu < \infty\} \in \mathcal{F}_\nu$.

Proposition 3.8 *[Cramer-Lundberg approximation] Suppose that F is non-lattice and that $\hat{F}(s) < \infty$ for $s < \gamma + \epsilon$, for some $\epsilon > 0$. Then*

$$\lim_{u\to\infty} \frac{\mathbb{P}(\nu(u) < \infty)}{ce^{-\gamma u}} = 1.$$

where the constant $c \geq 0$ is finite and given by

$$c = \frac{1 - G^+(\infty)}{\gamma \int_0^\infty v e^{\gamma v}\, G^+(\mathrm{d}v)}\,. \tag{3.31}$$

Before the proof we need few preliminaries.

Recall the overshoot process $B(u) = S_{\nu(u)} - u$, which we study under $\mathbb{P}^{(\gamma)}$. Thus we need to know the distribution of X^+ under $\mathbb{P}^{(\gamma)}$. Note that if $\hat{F}(s) < \infty$ for $0 \le s \le \gamma$, then we can demonstrate that $\hat{G}^+(s) < \infty$ for $0 \le s \le \gamma$. $\hat{G^-}(\gamma) < 1$. Thus, $\hat{F}(\gamma) = 1$ and

Lemma 3.9 *If $\hat{F}(s) < \infty$ for $0 \le s \le \gamma$, then the distribution $\tilde{G}^+$ of X^+ under $\mathbb{P}^{(\gamma)}$ is*

$$\mathrm{d}\tilde{G}^+(x) = \mathrm{e}^{\gamma x}\,\mathrm{d}G^+(x)\,, \qquad x \ge 0\,. \tag{3.32}$$

Proof We have $1 - \tilde{G}^+(x) = \sum_{n=1}^{\infty} \mathbb{P}^{(\gamma)}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n > x)$ and $1 - G^+(x) = \sum_{n=1}^{\infty} \mathbb{P}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n > x)$ for all $x \ge 0$. It is therefore sufficient to show that

$$\begin{aligned}
&\mathbb{P}^{(\gamma)}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n > x) \\
&\quad = \int_x^{\infty} \mathrm{e}^{\gamma y}\,\mathbb{P}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n \in \mathrm{d}y)\,,
\end{aligned}$$

for all $n = 1, 2, \ldots$ and $x \ge 0$. Putting $v_n = y_1 + \ldots + y_n$, we have

$$\begin{aligned}
&\mathbb{P}^{(\gamma)}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n > x) \\
&\quad = \int_{-\infty}^{0} \int_{-\infty}^{-v_1} \cdots \int_{-\infty}^{-v_{n-2}} \int_{x-v_{n-1}}^{\infty} \mathbb{P}^{(\gamma)}(X_n \in \mathrm{d}y_n) \\
&\qquad\qquad \times \mathbb{P}^{(\gamma)}(X_{n-1} \in \mathrm{d}y_{n-1}) \ldots \mathbb{P}(X_2 \in \mathrm{d}y_2)\mathbb{P}(X_1 \in \mathrm{d}y_1) \\
&\quad = \int_{-\infty}^{0} \int_{-\infty}^{-v_1} \cdots \int_{-\infty}^{-v_{n-2}} \int_{x-v_{n-1}}^{\infty} \mathrm{e}^{\gamma v_n}\,\mathbb{P}(X_n \in \mathrm{d}y_n) \\
&\qquad\qquad \times \mathbb{P}(X_{n-1} \in \mathrm{d}y_{n-1}) \ldots \mathbb{P}(X_2 \in \mathrm{d}y_2)\mathbb{P}(X_1 \in \mathrm{d}y_1) \\
&\quad = \int_x^{\infty} \mathrm{e}^{\gamma y}\,\mathbb{P}(S_1 \le 0, \ldots, S_{n-1} \le 0, S_n \in \mathrm{d}y)\,.
\end{aligned}$$

□

Proof of Theorem 3.8. From (3.29) we have

$$\mathbb{P}(\nu < \infty) = \mathbb{E}^{(\gamma)}[e^{-\gamma S_\nu}; \nu < \infty] \tag{3.33}$$

and since $\mathbb{P}^{(\gamma)}(\nu < \infty) = 1$ we have

$$\mathbb{P}(\nu < \infty) = \mathbb{E}^{(\gamma)}[e^{-\gamma S_\nu}] = e^{-\gamma u}\mathbb{E}^{(\gamma)}[e^{-\gamma(S_\nu - u)}] \tag{3.34}$$

By Lemma 3.7 the overshoot process $B(u) = S_{\nu(u)} - u$ does converge in distribution to a limit. To complete the proof we compute

$$\begin{aligned} c = \mathbb{E}^{(\gamma)} e^{-\gamma B(\infty)} &= \frac{\int_0^\infty e^{-\gamma x}(1-\tilde{G}^+(x))\,\mathrm{d}x}{\mathbb{E}^{(\gamma)} X^+} \\ = \frac{\int_0^\infty e^{-\gamma x}\int_x^\infty e^{\gamma v}\,G^+(\mathrm{d}v)}{\int_0^\infty v e^{\gamma v}\,G^+(\mathrm{d}v)} &= \frac{1-G^+(\infty)}{\gamma\int_0^\infty v e^{\gamma v}\,G^+(\mathrm{d}v)} < \infty\,. \end{aligned}$$

□

Remark For the lattice case the result of Theorem 3.8 holds in the following form

$$\lim_{\mathbb{Z}_+ \ni u \to \infty} \frac{\mathbb{P}(\nu(u) < \infty)}{Ce^{-\gamma u}} = 1,$$

where $0 < C < \infty$ is a constant different than c from Theorem 3.8.

3.1 Efficient simulation of $\mathbb{P}(\sup_{n\geq 0} S_n > u)$.

We first introduce some general notions from the Monte Carlo theory. We have a family of events $A(u)_{u>0}$ or $(A_n)_{n\in\mathbb{Z}_+}$, like $A(u) = \{\mathcal{M} > u\}$, or $A_n = \{S_n > u\}$. In the following part of this section we will work with a family $(A(u))_u$. Suppose we aim to compute the probability $\mathbb{P}(A(u))$ for a big u by simulation. An ad hoc method, called the *crude Monte Carlo method*, suggests the following procedure. Make k independent experiments of the event $A(u)$. resulting in $Z_i(u) = Z_i = 1$ if in the i-th experiment the event holds, otherwise $Z_i(u) = Z_i = 0$ $(i = 1, 2, \ldots, n)$. To avoid writing subscripts we use the notion of the generic random variable, that is we write Z for the generic one in the i.i.d. sequence $Z_1, \ldots, Z_k$. Formally we have a sequence of independent events $A_1(u), \ldots, A_k(u)$ and $Z_i(x) = 1(A_i(u))$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The theory presented here deals with the so called family $(A(u))_u$ of *rare events* is , that is fulfilling if $z(u) = \mathbb{E}\, Z(u) \to 0$ as $u \to \infty$. We will make a stronger assumption on the sequence of rare events, that

Hyphotesis 1.

$$\lim_{n\to\infty} \frac{1}{u} \log \mathbb{P}(A(u)) = -\gamma$$

for some $\gamma > 0$.

The *crude Monte Carlo estimator* (CMC) estimator is then

$$\hat{z}_k^{\mathrm{CMC}} = \frac{\sum_{i=1}^k Z_i}{k}.$$

The variance of z_k^{CMC} variance $z(u)(1-z(u))/k$ which, for each fixed k, it converges to zero as $u \to \infty$. However, in a moment we will see that this is not a good estimator, but first we need some notions.

Suppose it is possible to generate a random variable $\zeta = \zeta(u)$ on a probability space $(\Omega^{'}, \mathcal{F}^{'}, \mathbb{P}^{'})$, perhaps different than in the original problem, where the family $(A(u)_u$ was defined on $(\Omega, \mathcal{F}, \mathbb{P})$. However we have

$$z(u) = \mathbb{E}^{'}\zeta(u) = \mathbb{P}^{'}(A(u)).$$

Suppose that now we run k independent experiments $\zeta_1, \ldots, \zeta_k$ and define the estimator

$$\hat{z}_k = \frac{\sum_{i=1}^k \zeta_i}{k}.$$

We now recall few definitions. We say that an estimator $\hat{z}_k$ is *unbiased* if $\mathbb{E}^{'}\hat{z}_n = z(x)$ and it is the case of our estimator $\hat{z}_k$. We say that for a given u, the estimator $\hat{z}_k$ is an (ϵ, δ)–*accurate estimator* of $A(u)$ if

$$\mathbb{P}^{'}(\mid \hat{z}_k - z(u) \mid < \epsilon z(u)) > 1 - \delta.$$

Notice that

$$\frac{\hat{z}_k - z(u)}{z(u)}$$

is the relative error, between the estimated value $\hat{z}_k$ and the real value $z(u)$. The problem is to find u such that $\hat{z}_k$ is (ϵ, δ)–accurate for a given $\epsilon > 0, \delta > 0$. For this we may use the Central Limit Theorem approximations:

$$\mathbb{P}^{'}(\mid \hat{z}_k - z(u) \mid < \epsilon z(u)) = \mathbb{P}^{'}(\mid \frac{\sum_{i=1}^n (\zeta_i - \mathbb{E}\,\zeta)}{\sqrt{k}\sqrt{\mathrm{Var}\,\zeta}} \mid < \epsilon \frac{\sqrt{k} z(u)}{\sqrt{\mathrm{Var}\,\zeta}})$$

Since $\frac{\sum_{i=1}^k (\zeta_i - \mathbf{E}^{'}\zeta_i)}{\sqrt{k}\sqrt{\mathrm{Var}^{'}\zeta_i}}$ is approximately N(0,1) normally distributed, we may write, although with some abuse, that

$$\Phi(\epsilon\frac{\sqrt{n} z(u)}{\sqrt{\mathrm{Var}\,\zeta}}) - \Phi(-\epsilon\frac{\sqrt{k} z(u)}{\sqrt{\mathrm{Var}\,\zeta}}) > 1 - \delta.$$

Hence we may approximate that

$$\epsilon\frac{\sqrt{k} z(u)}{\sqrt{\mathrm{Var}\,\zeta(u)}} = \Phi^{-1}(1 - \frac{\delta}{2}),$$

which yields

$$k = k(u) = \frac{(\Phi^{-1}(1-\frac{\delta}{2}))^2}{\epsilon^2}\frac{\mathrm{Var}\,\zeta(u)}{z^2(u)}.$$

We see that the important quantity is the so called *square coefficient of variation*

$$\mathrm{SCV}_{\zeta(u)} = \frac{\mathrm{Var}'\zeta(u)}{(\mathbb{E}'\zeta(u))^2} = \frac{\mathrm{Var}'\zeta(u)}{z^2(u)}.$$

Coming back to the crude estimator $\hat{z}_k^{\mathrm{CMC}}$ we see that

$$k = k(u) = \frac{\Phi^{-1}(1-\frac{\delta}{2})}{\epsilon^2}\frac{\mathrm{Var}'\zeta(u)}{z^2(u)} = \mathrm{const}\frac{z(u)(1-z(u))}{z^2(u)}.$$

Therefore, as $u \to \infty$, the number of experiments growth like $k(u) \sim \mathrm{const} 1/z(u)$. In view our Hyphotesis 1, $k(u)$ growth exponentialy fast as u tends to ∞, which is not a good property.

Of course the ideal situation would be that $n(u)$ is bounded.

Definition 3.10 We say that the estimator $\hat{z}_k$ is an *efficient* one, if the square coefficient of variation $\mathrm{SCV}_{\zeta(\mathrm{u})}$ is a bounded function of u.

Another choice is that $k(u)$ growth in *polynomial time*. Such the estimators are sait to be *polynomial time* estimator. However we choose a little more general definition.

Definition 3.11 We say that the estimator $\hat{z}_k$ is an *weakly efficient* one, if

$$\limsup_{u\to\infty} \frac{\mathrm{Var}'\zeta(u)}{z^{2-\epsilon}(u)} < \infty.$$

Problems

3.1 Show that under Hyphotesis 1 the estimator $\hat{z}_k$ is *weakly efficient* if and only if if for all $\epsilon > 0$

$$\limsup_{x\to\infty} \mathrm{SCV}_{Z(x)} e^{-\epsilon x} = 0$$

3.2 Show that if an estimator $\hat{z}_k$ is a *polynomial time* time one, then it is weakly efficient and that $k(x) = O(|\log z(x)|^p)$, as $x \to \infty$ for some $0 < p < \infty$. Show that if an estimator $\hat{z}_k$ is an *exponential time* one, then $n(x)$ is of order at least $z(x)^{-q}$ for some $q > 0$.

We now stude ta concrete problem. Suppose that for a random walk $(S_n)_n$ with the generic increment X, considered on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define $\nu = \nu(u) = \min\{n : S_n > u\}$. We want to estimate the probability $z(u)$ of the event $A(u) = \{\nu(u) < \infty\}$, where $u > 0$. The CMC estimator $\hat{z}^{\mathrm{CMC}}$ defined on the base of generic r.v. $Z(u) = 1(A(u))$, as we show above is not good becasue $k(u) \sim \mathrm{const}1/z(u)$. We now define a new estimator on the base of generic r.v.

$$\zeta(u) = e^{-\gamma S_{\nu(u)}} 1(A(u)),$$

defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^{(\gamma)}$, where $\hat{F}(\gamma) = 1$. Note that, since $\mathbb{E}^{(\gamma)} X > 0$ we may consider

$$\zeta(u) = e^{-\gamma S_{\nu(u)}}.$$

Let $\zeta_1, \ldots, \zeta_k$ be i.i.d. copies of $\zeta(u)$ we define above

$$\hat{z}_k^{\mathrm{OECM}} = \frac{1}{k} \sum_{j=1}^{k} \zeta_j.$$

From the basic identity (3.29) it follows that $\hat{z}_k^{\mathrm{OECM}}$ is unbiased.

Theorem 3.12 *Suppose there exists $\gamma > 0$ such that $\hat{F}(\gamma) = 1$ and that $\hat{F}(\gamma + \epsilon) < \infty$. The estimator $\hat{z}_k^{\mathrm{OCM}}$ is efficient.*

Proof We have to study the asymptotics of $\mathrm{Var}^{(\gamma)} \zeta(u)$. for this we study $\mathbb{E}^{(\gamma)} \zeta^2(u)$. However

$$\begin{aligned} \mathbb{E}^{(\gamma)} \zeta^2(u) &= \mathbb{E}^{(\gamma)} e^{-2\gamma S_{\nu(u)}} \\ &= e^{-2\gamma u} \mathbb{E}^{(\gamma)} e^{-2\gamma B(u)} \end{aligned}$$

Since $\mathbb{E}^{(\gamma)} e^{-2\gamma B(u)}$ converges to a positive constant c_2 (see Lemma 3.7 and the proof of Theorem 3.8), we have that

$$\frac{\mathrm{Var}^{(\gamma)} \zeta(u)}{z^2(u)} \to \frac{c_2 - c^2}{c^2} > 0,$$

where c was computed in Theorem 3.8. □

A converse statement is also true in the following form. In Theorem 3.12 we used the generic random variable

$$\begin{aligned} \zeta &= \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_1) \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_2) \ldots \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_\nu) 1(\nu < \infty) \\ &= \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_1) \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_2) \ldots \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_\nu) \end{aligned}$$

consider under $\mathbb{P}^{\gamma}$. We may now define for any distribution $\tilde{F}$ equivalent to F the likelihood ration martingale

$$M_k = \frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_1)\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_2)\ldots\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_k)$$

and on the basis of this positive and mean 1 martingale, define the probability measure $\tilde{\mathbb{P}}$. On this probability space we may now define the generic random variable ζ by

$$\zeta = \frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_1)\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_2)\ldots\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_\nu)1(\nu < \infty). \tag{3.35}$$

Again the estimator $\hat{z}_k = (\zeta_1 + \ldots \zeta_k)/k$ defined on the base of ζ from (3.35) is unbiased.

Theorem 3.13 *Suppose that $\tilde{F} \equiv F$. If the estimator $\hat{z}_n$ is weakly efficient, then $\tilde{F} = F_\gamma$.*

Proof Asmussen and Rubinstein (1998), p. 437. or Asmussen (2003).

4 Cramer's large deviation theorem

We consider here a random walk $(S_n)_n$ with the log m.g.f. $\kappa(s)$ and $\mathcal{I}_\kappa = \{s : \kappa(s) < \infty\}$. We denote

$$\hat{S}_n = \frac{S_n}{n}.$$

Assume for simplicity that $\mathbb{E}\, X = \mu$ is finite although the general theory does not require it. For the motivation we propose to solve the following problem

Problem

4.1 Suppose that $(X_n)_n$ is a sequence of i.i.d. random variables with normal distribution $X \sim \mathcal{N}(0, \sigma^2)$. Show that for $a > 0$

$$\mathbb{P}(\hat{S}_n > a) \sim \frac{\sigma e^{-\frac{na^2}{2\sigma^2}}}{\sqrt{2\pi n}a} \tag{4.36}$$

as $n \to \infty$ and

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n > a) = -\frac{a^2}{2\sigma^2}. \tag{4.37}$$

Note that the central limit theorem is not adequate for the problem of estimating of $\mathbb{P}(\hat{S}_n > a)$ or in general $\mathbb{P}(\hat{S}_n \in A)$ for some A.

The *Fenchel-Legendre* transform of $\kappa(s)$ is

$$\kappa^*(x) = \sup_{s \in \mathbb{R}} \{sx - \kappa(s)\}.$$

Our aim is to prove the following theorem.

Theorem 4.1 *[Cramer]*
(i) *For each closed set B*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in B) \leq -\inf_{x\in B} \kappa^*(x),$$

(ii) *For each open set B*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in B) \geq -\inf_{x\in B} \kappa^*(x).$$

In particular, if a Borel set B is such that

$$\inf_{x\in \mathrm{cl(B)}} \kappa^*(x) = \inf_{x\in \mathrm{int(B)}} \kappa^*(x),$$

then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in B) = -\inf_{x\in B} \kappa^*(x).$$

Such the sets are called κ^*–*continuity* sets.

Before we demonstrate a proof of Cramer's theorem we have to study properties of $\kappa^*(x)$. Observe first that for example if $\mathcal{I}_\kappa = \{0\}$, then $\kappa^* \equiv 0$. On the otherside we may have, even for $\mathcal{I}_\kappa = \mathbb{R}$ the domain $\mathcal{I}_{\kappa^*}$ either the whole real line $\mathbb{R}$ or its subinterval. Recall that $\kappa(s)$ is C^∞ in the open interval $\mathrm{int}(\mathcal{I}_\kappa)$ with

$$\kappa'(s) = \frac{\mathbb{E}\, X e^{sX}}{\hat{F}(s)}.$$

Furthermore in this interval $\kappa'(s)$ is strictly increasing. Thus if there exists the solution $\kappa'(s_0) = x$, then $\kappa^*(x) = s_0 x - \kappa(s_0)$.

Lemma 4.2
(i) $\kappa^* : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *is a convex, lower semicontinuous function.*[1] *Furthermore κ^* is strictly convex and C^∞ in the interior of the set* $\{\kappa'(s),\ s \in$

[1] that is such that each level set $\{x : I(x) \leq \alpha\}$ is closed

$\mathrm{int}(\mathcal{I}_\kappa)\}$.
(ii) *For all* $x \geq \mu$

$$\kappa^*(x) = \sup_{s\geq 0}\{sx - \kappa(s)\},$$

and is an increasing function. For all $x \leq \mu$

$$\kappa^*(x) = \sup_{s\leq 0}\{sx - \kappa(s)\},$$

and is a decreasing function.
(iii) $\kappa^*(\mu) = 0$.

Problems

4.1 Suppose that $X \sim \mathrm{B}(p)$ a Bernoulli random variable. Show that

$$\kappa^*(x) = \begin{cases} x\log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right) & x \in [0,1] \\ \infty & \text{otherwise} \end{cases}$$

4.2 Show without using Cramer's theorem that if $S_n = X_1 + \ldots + X_n$, where $X_1, \ldots$ are independent copies of a Bernoulli random variable, then

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n > 1) &= 0 \\ \lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \geq 1) &= \log p. \end{aligned}$$

Compare with Cramer's theorem applied to sets $(1,\infty)$ and $[1,\infty)$ respectively.

In particular we are interested in open intervals (proper or infinite). We have the following result.

Proposition 4.3 *If* $A = (y,\infty)$ *is an open interval such that* $a > \mu$ *and* $a \in \mathrm{Int}\{\kappa^{'}(s) : s \in \mathrm{Int}(\mathcal{I}_\kappa)\}$, *then*

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in A) = -\inf_{x\in A} \kappa^*(x).$$

Proof Dembo and Zeitouni (1993), p. 35.

The situation like in Cramer's theorem turns out to be quite typical, and therefore the following definition is needed. Any function $I : \mathbb{R} \to$

$\mathbb{R} \cup \infty$, which is lower semicontinuous is called a *rate function*. Among rate functions we distinguish the so called *good*, that is such that that each level set $\{x : \kappa^*(x) \leq \alpha\}$ is compact.

Let $(\hat{Z}_n)_n$ be a sequence of random variables. It is said that the sequences fulfills the *large deviation principle* (LDP) with rate function I if
for each closed set B

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{Z}_n \in B) \leq - \inf_{x\in B} I(x),$$

for each open set B

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{Z}_n \in B) \geq - \inf_{x\in B} I(x).$$

The following variant of Varadhan's integral lemma is useful.

Theorem 4.4 *[Varadhan] Suppose that* $(\hat{Z}_n)_n$ *fulfills the* large deviation principle *(LDP) with good rate function* I *and* $f : \mathbb{R} \to \mathbb{R}$ *is a continuous function. If*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\,[e^{nf(\hat{Z}_n)}; f(\hat{Z}_n) \geq M] = -\infty. \tag{4.38}$$

(i) *If* B *is closed, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\,[e^{nf(\hat{Z}_n)}; \hat{Z}_n \in B] \leq \sup_{x\in B}\{f(x) - I(x)\}.$$

(ii) *If* B *is open, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\,[e^{nf(\hat{Z}_n)}; \hat{Z}_n \in B] \geq \sup_{x\in B}\{f(x) - I(x)\}.$$

Proof Dembo and Zeitoni Section 4.3.

Lemma 4.5 *If for some* $c > 1$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\,[e^{cnf(\hat{Z}_n)}] < \infty,$$

then (4.38) holds.

5 Large deviations and efficient simulation

We now consider a sequence of rare events $(A_n)_n$ in the setup of Section 3.1. Recall that for a sequence of events $(A_n)_n$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we have $\lim_{n\to\infty} z(n) = 0$, where $z(n) = \mathbb{P}(A_n)$. The Hyphotesis 1 is fulfilled, that is

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(A_n) = -\gamma$$

for some $\gamma > 0$.

We want to compute $z(n) = \mathbb{P}(A_n)$ by the Monte Carlo simulation with the use of unbiased estimators of the form

$$\hat{z}_k = \frac{1}{k} \sum_{j=1}^{k} \zeta_j(n),$$

where $\zeta_1(n), \zeta_2(n), \ldots, \zeta_k(n)$ are i.i.d. copies of the generic random variable $\zeta(n)$ on $(\Omega', \mathcal{F}', \mathbb{P}')$, such that $\mathbb{E}'\zeta(n) = z(n)$. The asymptotic quality is measured by the behaviour at $n \to \infty$ of

$$\mathrm{SCV}_{\zeta(n)} = \frac{\mathrm{Var}'(\zeta(n))}{(\mathbb{E}'\zeta(n))^2} = \frac{\mathrm{Var}'(\zeta(n))}{z(n)^2},$$

see Section 3.1 for definitions.

For this section the following definition of efficiency is more suitable.

Definition 5.1 We say that the estimator $\hat{z}_k$ is a *logarithmic efficient* one, if the square coefficient of variation

$$\liminf_{n\to\infty} \frac{\log \mathrm{Var}'\zeta(n)}{2\log \mathbb{E}'\zeta(n)} \geq 1.$$

We will assume that

Hyphotesis 2 $\mathrm{Var}'(\zeta(n)) \sim \mathbb{E}'\zeta^2(n)$, for $n \to \infty$.

Problems

5.1 Show that under Hyphotesis 2 the estimator is weakly efficient if and only if for all $\epsilon > 0$

$$\limsup_{n\to\infty} \frac{\mathbb{E}'(\zeta^2(n))}{(\mathbb{E}'\zeta(n))^{2-\epsilon}} = 0.$$

5.2 Show that, under Hyphotesis 1 and 2, the estimator is logarithmic efficient if and only if

$$\liminf_{n\to\infty}(-\frac{1}{n})\log \mathbb{E}^{'}\zeta^2(n) \geq 2\gamma.$$

5.3 Show that, under Hyphotesis 1 and 2, the estimator is logarithmic efficient if and only if

$$\lim_{n\to\infty}(-\frac{1}{n})\log \mathbb{E}^{'}\zeta^2(n) = 2\gamma.$$

From the result of Exercise 5.3 we see that

$$\liminf_{n\to\infty}(-\frac{1}{n})\log \mathbb{E}^{'}\zeta^2(n) \tag{5.39}$$

is an indicator of the quality of simulation. Show that always

$$\liminf_{n\to\infty}(-\frac{1}{n})\log \mathbb{E}^{'}\zeta^2(n) \leq 2\gamma.$$

Lemma 5.2 *If the estimator is logarithmic efficient, then it is weakly efficient.*

Proof By the result of Exercise 5.3 we have

$$\begin{aligned}
&\lim_{n\to\infty}\left(\frac{\log \mathbb{E}^{'}\zeta^2(n)}{2\log \mathbb{E}^{'}\zeta(n)} - \frac{2-\epsilon}{2}\right) 2\log \mathbb{E}^{'}\zeta(n) \\
&= \lim_{n\to\infty}\left(\frac{\log \mathbb{E}^{'}\zeta^2(n)}{2\log \mathbb{E}^{'}\zeta(n)} - \frac{2-\epsilon}{2}\right) 2n\frac{1}{n}\log p_n = -\infty
\end{aligned}$$

Hence

$$\begin{aligned}
-\infty &= \lim_{n\to\infty}(\log \mathbb{E}^{'}\zeta^2(n) - (2-\epsilon)\log \mathbb{E}^{'}\zeta(n)) \\
&= \lim_{n\to\infty}\log\frac{\mathbb{E}^{'}\zeta^2(n)}{(\mathbb{E}^{'}\zeta(n))^{2-\epsilon}}
\end{aligned}$$

which yields the weak efficiency. □

Further on we will consider a specific model. Let $(S_n)_n$ be a random walk with the log m.g.f. of the increment $\kappa(s)$. We assume that

Hyphotesis 3. $\mathcal{I}_\kappa = \mathbb{R}$.

Let $\kappa^*(x)$ be the Fenchel-Legendre transform of $\kappa(s)$. Suppose there is given a Borel set B such that $\mu \neq \mathrm{cl}(B)$. We now consider the following sentence of rare events $A_n = \{\hat{S}_n \in B\}$. In particular we are interested in $B = [b, \infty)$ or $B = (-\infty, a] \cup [b, \infty)$, where $a < \mu < b$. Anyway we will assume that B is a closed set such that

$$\lim_{n\to\infty} (-\frac{1}{n}) \log \mathbb{P}'(\hat{S}_n \in B) = \inf_{x \in B} \kappa^*(x) < \infty.$$

The crude MC method uses the estimator

$$\hat{z}_k^{\mathrm{CMC}} = \frac{1}{k} \sum_{j=1}^{k} \zeta_k(n),$$

where $\zeta_1(n), \zeta_2(n), \ldots$ are i.i.d. copies of the generic random variable $\zeta(n) = 1(A_n)$. Note that in this case

$$\frac{\log \mathbb{E}\,\zeta^2(n)}{\log(\mathbb{E}\,\zeta(n))^2} = \frac{1}{2}.$$

Consider now a family of importance sampling estimator $\hat{z}_k^{(s)}$ $(s \in \mathbb{R})$ defined by generic

$$\zeta(n) = e^{-sS_n + n\kappa(s)} 1(A_n) = e^{nf(\hat{S}_n)} 1(\hat{S}_n \in B), \tag{5.40}$$

where

$$f(x) = \kappa(s) - sx, \tag{5.41}$$

considered on the probability space $(\Omega, \mathbb{F}, \mathbb{P}^{(s)})$, where $\mathbb{P}^{(s)}$ is the probability measure defined by Wald martingale $M_n^{(s)}$; see Section 3. Note that

$$z(n) = \mathbb{E}^{(s)} \zeta(n).$$

To find quality of $\hat{z}_n^{(s)}$, following the result of Exercise 5.3, we have to compute

$$\liminf_{n\to\infty} (-\frac{1}{n}) \log \mathbb{E}^{(s)} \zeta^2(n).$$

In particular for the logarithmic efficiency we need that

$$\lim_{n\to\infty} (-\frac{1}{n}) \log \mathbb{E}^{(s)} \zeta^2(n) = 2\gamma.$$

We denote by $\kappa^{*,s}(x)$ the Fenchel-Legendre transform of $\kappa^{(s)}(v) = \kappa(s + v) - \kappa(v)$.

Lemma 5.3

$$\kappa^{*,s}(x) = \kappa^*(x) + \kappa(s) - sx.$$

Proof From the definition

$$\begin{aligned}&\sup_{v\in\mathbb{R}}\{vx - \kappa^{(s)}(v)\}\\ &\qquad = \sup_{v\in\mathbb{R}}\{(v+s)x - \kappa(v+s) + \kappa(s) - sx\} = \kappa^*(x) + \kappa(s) - sx.\end{aligned}$$

□

Lemma 5.4

$$\begin{aligned}\liminf_{n\to\infty}(-\frac{1}{n})\log \mathbb{E}^{(s)}\zeta^2(n) &\geq -\sup_{x\in B}\{\kappa(s) - \kappa^*(x) - sx\}\\ &= \inf_{x\in B}\{\kappa^*(x) - \kappa(s) + sx\}.\end{aligned}$$

Proof We write

$$\limsup_{n\to\infty}\frac{1}{n}\log \mathbb{E}^{(s)}\zeta^2(n) = \limsup_{n\to\infty}\frac{1}{n}\log \mathbb{E}^{(s)}[e^{n2f(\hat{S}_n)}; \hat{S}_n \in B],$$

where $f(x) = \kappa(s) - sx$. To use the Varadhan integral lemma (Theorem 4.4) we have to check the assumption of Lemma 4.5 holds for $c > 1$:

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log \mathbb{E}^{(s)}[e^{cn2f(\hat{S}_n)}] &= \limsup_{n\to\infty}\frac{1}{n}\log \mathbb{E}^{(s)}e^{-2c(sS_n - n\kappa(s))}\\ &= \kappa(s-2cs) - \kappa(-2cs) + 2c\kappa(-2cs) < \infty.\end{aligned}$$

Now by Theorem 4.4 (i), recalling that we assumed B closed

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log \mathbb{E}^{(s)}[e^{n2f(\hat{S}_n)}; \hat{S}_n \in B] &\leq \sup_{s\in\mathbb{R}}\{2(\kappa(s) - sx) - \kappa^{*,s}(x)\}\\ &= \sup_{s\in\mathbb{R}}\{\kappa(s) - \kappa^*(x) - sx\},\end{aligned}$$

from which the lemma follows. □

We will obtain the possibly optimal change of measure by the use of importance sampling estimator from (5.40) if we find s_0 such that

$$\begin{aligned}&\inf_{x\in B}\{\kappa^*(x) - \kappa(s_0) + s_0 x\}\\ &\qquad = \max_{s\in\mathbb{R}}\inf_{x\in B}\{\kappa^*(x) - \kappa(s) + sx\}\end{aligned} \tag{5.42}$$

For this a necessary condition is that the minimax hyphotesis is fulfilled,

$$\inf_{x\in B}\max_{s\in\mathbb{R}}\{\kappa^*(x)-\kappa(s)+sx\} = 2\inf_{x\in B}\kappa^*(x) = 2\gamma.$$

Note we always have

$$\max_{s\in\mathbb{R}}\inf_{x\in B}\{\kappa^*(x)-\kappa(s)+sx\} \le \inf_{x\in B}\max_{s\in\mathbb{R}}\{\kappa^*(x)-\kappa(s)+sx\}. \tag{5.43}$$

Proposition 5.5 *If $B=[b,\infty)$ for $b>\mu$ and $\kappa^*(b+)<\infty$, then there exists s_0 such that*

$$\max_{s\in\mathbb{R}}\inf_{x\ge b}\{\kappa^*(x)-\kappa(s)+sx\} = \inf_{x\ge b}\{\kappa^*(x)-\kappa(s_0)+s_0x\} = 2\gamma,$$

and so $\hat{z}_n^{(s_0)}$ is logarithmic efficient.

Proof We have

$$\begin{aligned}&\max_{s\in\mathbb{R}}\inf_{x\in B}\{\kappa^*(x)-\kappa(s)+sx\}\\ &\quad\ge \max_{s\in\mathbb{R}_+}\inf_{x\in B}\{\kappa^*(x)-\kappa(s)+sx\} = \max_{s\in\mathbb{R}_+}\{\kappa^*(b)-\kappa(s)+sb\}.\end{aligned}$$

Now using Lemma 4.2 (ii) we obtain

$$\max_{s\in\mathbb{R}_+}\{\kappa^*(b)-\kappa(s)+sb\} = \kappa^*(x).$$

In view of the inequality (5.43) the proof is completed. $\square$

Example 5.6 We consider a random walk with generic increment r.v. X normally distributed $\mathcal{N}(1,1)$. Suppose that $B=[b,\infty)$, where $1<b$. Then $\kappa(s)=s+s^2/2$, $\kappa^*(x)=(x-1)^2/2$ and hence

$$k(s,x)=\kappa^*(x)-\kappa(s)+sx = \frac{(x-1)^2-(s+1)^2+2sx+1}{2}.$$

In fact we have that

$$\begin{aligned}g_1(s) &= \inf_{x\ge b}k(s,x)\\ &= \begin{cases} -\frac{(s+1-b)^2}{2}+(b-1)^2 & s\ge 1-b\\ -s^2 & s<1-b\end{cases}\end{aligned}$$

and $\sup_{s\in\mathbb{R}} g_1(s) = (1-b)^2$ is achieved at $s_0 = b-1$

On the other side we have

$$\gamma = \inf_{x\geq b} \kappa^*(x) = \frac{(b-1)^2}{2}.$$

and hence we conclude that

$$\sup_{s\in\mathbb{R}} \inf_{x>b} k(s,x) = 2\gamma,$$

which yields the logarithmic efficiency, when using $\hat{z}^{(s_0)}$. Using Lemma 3.3 we find that X under $\mathbb{P}^{(b-1)}$ is $\mathcal{N}(b,1)$. We now continue the example, however with $B = (-\infty,-b] \cup [b,\infty)$, where $1 < b$. Then

$$\gamma = \min(\frac{(b-1)^2}{2}, \frac{(b+1)^2}{2}) = \frac{(b-1)^2}{2}.$$

We also have

$$\begin{aligned} g_2(s) &= \inf_{x\leq -b} k(s,x) \\ &= \begin{cases} -\frac{(s+1+b)^2}{2} + (b+1)^2 & s \geq 1+b, \\ -s^2 & s < 1+b. \end{cases} \end{aligned}$$

We have

$$g(s) = \min(g_1(s), g_2(s)) = \inf_{x\in B} k(s,x)$$

Since $\sup_s g(s) \leq (1-b)^2$, we may conclude, that this can be achieved at $s_0 = b-1$, when $g_1(b-1) \leq g_2(b-1)$ that is only for if $b \in (1,2]$. Otherwise $\sup_{s\in\mathbb{R}} g(s) < (1-b)^2$.

Remark There exists a minimax theorem, like the following one: Let $k(s,x)$ be convex and lower semicontinuous in x and concave and upper semicontinuous in s, for $s \in \mathbb{R}$ and $x \in B$. If B is convex and compact, then

$$\inf_{x\in C} \sup_s k(s,x) = \sup_s \inf_{x\in C} k(s,x)$$

(Dembo and Zeitoni, p. 42.)

Comments. For further reading on importance sampling in simulations see e.g. Asmussen and Rubinstein (1985) and Asmussen (1998)

Basic references for the large deviation theory are books of Dembo and Zeitouni (1993) and Bucklew (1990). Logarithmic efficient simulation of $\mathbb{P}(S_n \in B)$ for $\boldsymbol{S}_n$ being a random walk in $\mathbb{R}^d$ is studied recently by Dupuis and Wang (2003) for B being convex or nonconvex. Counterexamples, as e.g. our counterexample 5.6, showing that not always the importance sampling estimators improve the performance were given by Glasserman and Wang (1997). Other reference on simulation, when $\boldsymbol{S}_n$ is a random walk or Markov additive sequence in $\mathbb{R}^d$ are Sadowsky (1996), Collamore (2002).

Chapter IV

Markov additive sequence

In this section we consider a generalization of the random walk, introducing the so called background Markov process, which the state directs next jump. We suggest the reader to prove that the following sentence is an equivalent definition of the random walk (S_n).

1 Definitions and basic properties

Problems

1.1 Suppose that (S_n) is a sequence of random variables, $S_0 = 0$ with natural filtration $\mathcal{F}_n = \sigma\{S_0, S_1, \ldots, S_n\}$. Show that if for $n = 0, 1, \ldots$

$$(S_{n+m} - S_n)_{m=0,1,\ldots}|\mathcal{F}_n =_{\mathrm{d}} (S_m)_{m=0,1,\ldots}, \tag{1.1}$$

then (S_n) is a random walk.

1.2 Consider a random walk $(S_n)_{n=0,1,\ldots}$ on $(\Omega, \mathbb{F}, \mathbb{P})$. We want to consider this random walk beginning at $x \in \mathbb{R}$ and indicate it by saying that a sequence is considered on $(\Omega, \mathbb{F}, \mathbb{P}_x)$, where now $S_0 = x$ and $S_n = x + X_1 + \ldots + X_n$. Show that (S_n) is a Markov sequence, that is that $(S_{n+m})_{m=0,1,\ldots}|\mathcal{F}_n$ under $\mathbb{P}_x$ equals in distribution to $(S_m)_{m=0,1,\ldots}$ under $\mathbb{P}_{S_n}$.

We begin with few simple but motivating examples. The *simple* random walk on $\mathbb{Z}$ is a random walk with generic increment X assuming only two values 1 and -1 with probability p and q respectively. Then (S_n) is a DTMC with transition probability matrix

$$\boldsymbol{P}^{'} = \begin{array}{cc} & \begin{array}{ccccccc} \vdots & k-2 & k-1 & k & k+1 & k+2 & \vdots \end{array} \\ \begin{array}{c} \vdots \\ k-2 \\ k-1 \\ k \\ k+1 \\ \vdots \end{array} & \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & p & 0 & 0 & 0 & \vdots \\ \vdots & q & 0 & p & 0 & 0 & \vdots \\ \vdots & 0 & q & 0 & p & 0 & \vdots \\ \vdots & 0 & 0 & q & 0 & p & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \end{array}$$

Suppose now there is given a DTMC $(J_n)_n$ with state space $E = \{1,2\}$ governed by probability transition matrix

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}.$$

We now define a random sequence such that, if $S_{n-1} = k$, then conditioned on $J_{n-1} = i$, $J_n = j$ we have $S_n = k + X_n$ where

$$\begin{aligned} \mathbb{P}(X_n = 1 | J_{n-1} = i, J_n = j) &= p^{ij} \\ \mathbb{P}(X_n = -1 | J_{n-1} = i, J_n = j) &= q^{ij}, \end{aligned}$$

and $p^{ij} + q^{ij} = 1$ $(i,j = 1,2)$. The bivariate sequence (S_n, J_n) is a DTMC with state space $E^{'} = \mathbb{Z} \times E$ with transition probability matrix

$$\boldsymbol{P}^{'} = \begin{array}{cc} & \begin{array}{cccccc} \cdots & (k-1,1)(k-1,2) & (k,1)(k,2) & (k+1,1)(k+1,2) & (k+2,1)(k+2,2) & \cdots \end{array} \\ \begin{array}{c} \cdots \\ (k-2,1)\ (k-2,2) \\ (k-1,1)\ (k-1,2) \\ (k,1)\ (k,2) \\ (k+1,1)\ (k+1,2) \\ \cdots \end{array} & \left(\begin{array}{c|c|c|c|c|c} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hline \cdots & \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \hline \cdots & \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \hline \cdots & \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{0} & \cdots \\ \hline \cdots & \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{A} & \cdots \\ \hline \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right) \end{array},$$

where

$$\boldsymbol{A} = \begin{pmatrix} p_{11}p^{11} & p_{12}p^{12} \\ p_{21}p^{21} & p_{22}p^{22} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} p_{11}q^{11} & p_{12}q^{12} \\ p_{21}q^{21} & p_{22}q^{22} \end{pmatrix}.$$

We now introduce a general definition of a *Markov additive sequence* with a background DTMC and real valued *additive component*. In these notes we consider only the theory with a denumerable state space E of the background DTMC. Thus we consider a bivariate process $(S_n, J_n)_n$, on the probability space $(\Omega, \mathbb{F}, \mathbb{P}_{x,i})$ where

- (J_n) is a DTMC with state space E with initial state $J_0 = i$ and transition probability matrix $\boldsymbol{P}$,
- $S_n = S_0 + X_1 + \ldots + X_n$, where $X_i \in \mathbb{R}^d$,
- the distribution of X_n depends on J_{n-1} and J_n and given $J_1, \ldots, J_n$, the sequence $X_1, \ldots, X_n$ consists of independent random variables. Unless stated otherwise, $S_0 = x$ in particular $x = 0$. Anyway, if S_0 is random, then some independence of the other variables is required.

In the case when $S_0 = 0$, we write shortly that the underlying probability measure is $\mathbb{P}_i$.

We may reformulate the above definition as follows.

Problem

1.1 Let $\mathcal{F}_n = \sigma\{S_0, J_0, \ldots, S_n, J_n\}$ be the natural filtration. Show that

$$\mathbb{E}_{x,i}[f(S_{n+m} - S_n)g(J_{n+m}) \mid \mathcal{F}_n] = \mathbb{E}_{J_n}[f(S_m)g(J_m)], \mathbb{P}_{x,i} - \text{a.s.} \tag{1.2}$$

Some special cases are:

- Suppose $E = \{1\}$. Then the Markov additive sequence is a random walk.
- If $(J_n)_n$ is DTMC, the $X_n = f(J_n)$ defines a Markov additive sequence.

In the following we will consider Markov additive sequences with $X_n \in \mathbb{R}$. Specifying $f(x) = 1(x \in B)$ and $g(y) = 1(y = j)$ and $m = 1$ at the RHS of (1.2) we have subdistributions $F_{ij}(B)$, where

$$F_{ij}(B) = \mathbb{P}_i(S_1 \in B, J_1 = j).$$

We define $F_{ij}^{(n)}(B)$ recurrently by $F_{ij}^{(1)} = F_{ij}$ and

$$\begin{aligned} F_{ij}^{(n+1)}(B) &= \sum_{k \in E} \int F_{ik}^{(n)}(B - x)\, \mathrm{d}F_{kj}(\mathrm{d}x) \\ &= \sum_{k \in E} \int F_{ik}(B - x)\, \mathrm{d}F_{kj}^{(n)}(\mathrm{d}x). \end{aligned} \tag{1.3}$$

For an abbreviated form of notations, we introduce matrices:

$$\boldsymbol{F}(dx) = (F_{ij}(dx))_{i,j\in E} \quad \text{and} \quad \boldsymbol{F}^{(n)}(\mathrm{d}x) = (F_{ij}^{(n)}(\mathrm{d}x))_{i,j\in E}$$

Note that $F_{ij}(B) \leq 1$. Sometimes it is useful to use, instead of $\boldsymbol{F}$, equivalently the stochastic matrix $\boldsymbol{P} = (p_{ij}) = \boldsymbol{F}(\mathbb{R})$ and $\boldsymbol{F}^{\circ}(\mathrm{d}x) = (F_{ij}^{\circ}(\mathrm{d}x))_{i,j\in E}$, where

$$F_{ij}^{\circ}(\mathrm{d}x) = \frac{F_{ij}(\mathrm{d}x)}{p_{ij}} \qquad \text{if} \quad p_{ij} \neq 0.$$

Notice that $F^{\circ}(\mathbb{R}) = 1$ for i, j.

Problems

1.1 Show that

$$\begin{aligned}\mathbb{P}_{x,i}&(X_1 \in \mathrm{d}x_1, J_1 = j_1, \dots, X_n \in \mathrm{d}x_n, J_n = j_n)\\ &= F_{ij_1}(\mathrm{d}x_1)F_{j_1j_2}(\mathrm{d}x_2)\dots F_{j_{n-1}j_n}(\mathrm{d}x_n).\end{aligned}$$

1.2 Show that

$$\mathbb{P}_i(S_n \in B, J_n = j) = F_{ij}^{(n)}(B).$$

We now show two further examples of Markov additive sequences.

Example 1.1 Consider an M/G/1 FCFS queue. Suppose that $W_0 = 0$ (technical assumption only). Denote by $D_0 = 0$, and D_n the moment of the n-th departure from the system and $L_0 = 0$ and L_n be the number of customers in the system just after the n-the departure. We claim that (D_n, L_n) is a Markov additive process, with

$$\mathbb{P}(D_n \in \mathrm{d}x, L_{n-1} = i, L_n = i+j) = \begin{cases} B(dx)\frac{(ax)^{j+1}}{(j+1)!}e^{-ax}, & i \geq 1 \\ \\ ae^{-ax}\int_0^x \frac{(ax)^j}{j!}B(\mathrm{d}u), & i = 0\,. \end{cases}$$

Example 1.2 The following model is built for the study of M/M/1 queue with the processor sharing discipline. One of the basic characteristic in such the systems is the so called *response time* of the tagged customer. Suppose that at time $t = 0$ there is an arrival of the customer with service requirement of size x. One sometims says that x is an ideal service time of the tagged customer. It depends clearly on m the number of customers the taggged one meets at its start. Therefore we may say (x, m)-customer. This requirement is served with the speed $1/(1+k)$, if there is additionally more k customers in the service. The response time is the time to complete the service of the tagged customer. To model this system consider two independent sequences: an i.i.d. sequence (ξ_i) of Exp(b)–random variables and an i.i.d. sequence (η_i) of Exp(a)–random variables. On the base of these two sequences we define a bivariate sequence $(X_n, J_n)_n$. We proceed recursively as follows. Let $J_0 = m$. If $\xi_n < \eta_n$, then

$$\begin{aligned} J_n &= (J_{n-1} - 1)_+ \ , \\ X_n &= \xi_n/(1 + J_{n-1}) \ , \\ \psi_n &= \xi_n \ , \end{aligned}$$

otherwise if $\xi_n \geq \eta_n$, then

$$\begin{aligned} J_n &= J_{n-1} + 1 \ , \\ X_n &= \eta_n/(1 + J_{n-1}) \ . \\ \psi_n &= \eta_n . \end{aligned}$$

Let $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$. Let $Z_0 = 0$ and $Z_n = \psi_1 + \ldots + \psi_n$. We may consider the process $S(t)$, which is obtained by the linear interpolation between points $S_n, Z_n)_{n\in\mathbb{Z}_+}$. Let $\tau(x) = \inf\{t > 0 : S(t) > x\}$. Notice that $\tau(x)$ has the same distribution as the response time of a (x, m) type tagged customer. It would be interesting to learn the asymptotics of $\mathbb{P}(\tau(x) > t)$. We point out that for the processor sharing queues it is known the asymptotic behaviour of V the tail of the response time distribution. In our setting this is the distribution of (ξ, L)-tagged customer, where J is the generic steady state of the DTMP (J_n) and ξ is independent of J. It was proved that

$$\mathbb{P}(V > t) \sim c_1 t^{5/6} e^{-c_2 t - c_3 t^{1/3}}, \qquad t \to \infty,$$

where constants c_1, c_2, c_3 are known. The proof of this fact is purely analytical, and it would be interesting to find a probabilistic one, which uses the Markov additive structure as above.

Problems

1.1 Show that J_n in Example 1.2 is a DTMC on $\{0, 1, \ldots\}$ with transition matrix $\boldsymbol{P} = (p_{ij})$ given by $p_{i,i+1} = \frac{a}{a+b}$ $(i = 0, 1, \ldots)$, $p_{i,i-1} = \frac{b}{a+b}$ $(i = 1, \ldots)$ and $p_{0,0} = \frac{b}{a+b}$.

1.2 Show that $((S_n, Z_n), J_n)$ in Example 1.2 is a Markov additive sequence.

We now define

$$\hat{F}_{ij}(s) = \int_{-\infty}^{\infty} e^{sx}\, F_{ij}(\mathrm{d}x)$$

and next the *moment generating matrix function* (m.g.m.f.)

$$\hat{\boldsymbol{F}}(s) = (\hat{F}_{ij}(s))_{i,j\in E}.$$

Lemma 1.3 *We have for all $n = 1, 2, \ldots$*

$$(\int_{-\infty}^{\infty} e^{sx}\, F_{ij}^{(n)}(\mathrm{d}x))_{i,j\in E} = (\hat{\boldsymbol{F}}(s))^n.$$

Proof The proof uses the principle of mathematical induction. Clearly it is true for $n = 1$. Asuming it is true for n, using (1.3) we write

$$\begin{aligned}
(\int_{-\infty}^{\infty} e^{sx} F_{ij}^{(n+1)}(\mathrm{d}x))_{ij} &= (\sum_{k\in E} \int_{-\infty}^{\infty} e^{sx} F_{ik}^{(n)}(\mathrm{d}x) \int_{-\infty}^{\infty} e^{sx} F_{kj}(\mathrm{d}x))_{ij} \\
&= (\hat{\boldsymbol{F}}(s))^n \hat{\boldsymbol{F}}(s).
\end{aligned}$$

□

We now study the spectral properties of $\hat{\boldsymbol{F}}(s)$. Let

$$\mathcal{I}_{\hat{\boldsymbol{F}}} = \{s \in \mathbb{R} :\ \hat{F}_{ij}(s) < \infty, i, j \in \mathbf{E}\}.$$

be the interval that all $F_{ij}(s)$ $(i, j \in E)$ are well defined. Remark that it is possible that $\mathcal{I}_{\hat{\boldsymbol{F}}} = \{0\}$. From now on we assume the following

- $E = \{1, \ldots, l\}$,
- $\boldsymbol{P} = \boldsymbol{F}(\infty) = \hat{\boldsymbol{F}}(0)$ is primitive.

From Perron-Frobienius theorem we have that the dominating eigenvalue $\theta_{\mathrm{PF}}(s)$ of $\hat{\boldsymbol{F}}(s)$ is positive and therefore we may define

$$\kappa(s) = \log \theta_{\mathrm{PF}}(s). \tag{1.4}$$

Let $\boldsymbol{\xi}^{(s)}$ and $\boldsymbol{h}^{(s)}$ be the corresponding left and right eigenvectors respectively. The basic properties of $\kappa(s)$ are as follows: if $s = 0$, then $\hat{F}(0) = \boldsymbol{P}$ and so $\kappa(0) = 0$, $\boldsymbol{\xi}^{(0)} = \boldsymbol{\pi}$ - the stationary distribution of $\boldsymbol{P}$ and $(\boldsymbol{h}^{(0)})^\top = (1, \ldots, 1)$. Recall that, according to our conventions

$$\boldsymbol{\xi}^{(s)} \boldsymbol{h}^{(s)} = 1.$$

Moreover from the Perron-Frobieniu theorem we have that $\boldsymbol{h}^{(s)}$ has all components strictly positive and that for $n \to \infty$

$$\mathbb{E}_i[e^{sS_n}; J_n = j] \sim e^{n\kappa(s)} h_i^{(s)} \xi_j^{(s)}. \tag{1.5}$$

Lemma 1.4 *Suppose that* $\mathrm{int}(\mathcal{I}_{\hat{\boldsymbol{F}}})$ *is a nonempty interval. Then functions* $\mathcal{I}_{\hat{\boldsymbol{F}}} \ni s \to \kappa(s)$, $\mathcal{I}_{\hat{\boldsymbol{F}}} \ni s \to \boldsymbol{\nu}^{(s)}$ *and* $\mathcal{I}_{\hat{\boldsymbol{F}}} \ni s \to \boldsymbol{h}^{(s)}$ *are infinite times differentiable.*

Let $\mathcal{I}$ be an open interval. We say that a function $\phi : \mathcal{I} \to \mathbb{R} \cup \{\infty\}$ is *convex* if for all $0 \le \alpha \le 1$, $s_1, s_2 \in \mathcal{I}$

$$\phi(\alpha s_1 + (1-\alpha)s_2) \le \alpha\phi(s_1) + (1-\alpha)\phi(s_2).$$

We say that a function $f : \mathcal{I} \to (0,\infty) \cup \{\infty\}$ is *logconvex* if $\log f(x)$ is a convex function. Recall that a continuous function $\phi(s)$ on $\mathcal{I} = \{s : \phi(s) < \infty\}$ is convex if and only if for $s_1, s_2 \in \mathcal{I}$

$$\phi(\frac{s_1+s_2}{2}) \le \frac{\phi(s_1)+\phi(s_2)}{2}.$$

Lemma 1.5 *A function f on $\mathcal{I}$ is logconvex if and only if it is continuous and for $s_1, s_2 \in \mathcal{I}$*

$$f(\frac{s_1+s_2}{2}) \le (f(s_1)f(s_2))^{1/2}.$$

Lemma 1.6 *Class of logconvex functions on I is closed under the operation of addition, multiplication and passing to the limit.*

Proof [1] It is straightforward that the product of two logconvex functions is logconvex. Since the limit of convex functions is a convex function, perhaps with ∞ value, then we have to prove that the class is closed under addition operation. Hence we have that if $f(s)$ and $g(s)$ are logconvex, then

$$\begin{aligned} f(\frac{s_1+s_2}{2}) &\le (f(s_1)f(s_2))^{1/2} \\ g(\frac{s_1+s_2}{2}) &\le (g(s_1)g(s_2))^{1/2} \end{aligned}$$

Hence

$$f(\frac{s_1+s_2}{2}) + g(\frac{s_1+s_2}{2}) \le (f(s_1)f(s_2))^{1/2} + (g(s_1)g(s_2))^{1/2}.$$

Since the geometric mean is less or equal than the arithmetic mean we write

$$\begin{aligned} &(f(\frac{s_1+s_2}{2}) + g(\frac{s_1+s_2}{2}))^2 \\ \le\ & f(s_1)f(s_2) + g(s_1)g(s_2) + 2(f(s_1)f(s_2)g(s_1)g(s_2))^{1/2} \\ \le\ & f(s_1)f(s_2) + g(s_1)g(s_2) + 2\frac{f(s_1)g(s_2)+f(s_2)g(s_1)}{2} \\ =\ & (f(s_1)+g(s_1))(f(s_2)+g(s_2)). \end{aligned}$$

[1] Presented to the author by Z. Puchała.

This is equivalent to $f+g$ being logconvex. $\square$

Lemma 1.7 *Function* $\hat{F}_{ij}(s)=\int_{-\infty}^{\infty} e^{sx}\, F_{ij}(\mathrm{d}x)$ *is a logconvex.*

Proof For demonstration we use Hölder inequality with $p=1/\alpha$ and $q=1/(1-\alpha)$ such that $0<\alpha<1$

$$\begin{aligned}\hat{F}_{ij}(\alpha s_1+(1-\alpha)s_2) &= \int_{-\infty}^{\infty} e^{\alpha s_1 x} e^{(1-\alpha)s_2 x}\, F_{ij}(\mathrm{d}x)\\ &\geq (\int_{-\infty}^{\infty} (e^{\alpha s_1 x})^{1/\alpha}\, F_{ij}(\mathrm{d}x))^{\alpha} (\int_{-\infty}^{\infty} (e^{\alpha s_2 x})^{1/(1-\alpha)}\, F_{ij}(\mathrm{d}x))^{1-\alpha}\\ &= (\hat{F}_{ij}(s_1))^{\alpha}(\hat{F}_{ij}(s_2))^{1-\alpha}\end{aligned}$$

$\square$

Lemma 1.8 *[Kingman]* $\kappa(s)$ *is a convex function.*

Proof Let $\theta_{\mathrm{PF}}(s), \theta_2(s), \ldots(s), \theta_l(s)$ are eigenvalues of $\hat{\boldsymbol{F}}(s)$. Recall that the trace of a matrix $\boldsymbol{A}=(a_{ij})$ is $\mathrm{tr}(\boldsymbol{A})=\sum_{j=1}^{l} a_{jj}$. We have

$$\theta_{\mathrm{PF}}^n(s)+\theta_2^n(s)+\ldots+\theta_l^n(s)=\mathrm{tr}(\hat{\boldsymbol{F}}^n(s)).$$

However, by Lemma 1.6, $\mathrm{tr}(\hat{\boldsymbol{F}}^n(s))$ is logconvex. Hence $(\theta_{\mathrm{PF}}^n(s)+\theta_2^n(s)+\ldots+\theta_l^n(s))^{1/n}$ is logconvex too. Now passing to the limit, by Perron-Frobenius theorem $(\theta_{\mathrm{PF}}^n(s)+\theta_2^n(s)+\ldots+\theta_l^n(s))^{1/n}\to\theta_{\mathrm{PF}}(s)$. Hence $\theta_{\mathrm{PF}}(s)$ is logconvex and the proof is completed. $\square$

2 Wald martingales and change of measure

We now define the Wald martingale for Markov additive sequences. We assume that the background E is finite and $\boldsymbol{P}$ is primitive. Then $\kappa(s)$ and $\boldsymbol{h}^{(s)}$ are well defined in Section 1. Let

$$M_n^{(s)}=\frac{h_{J_n}^{(s)}}{h_i^{(s)}}e^{sS_n-n\kappa(s)}, \qquad n=0,1,\ldots$$

and $\mathcal{F}_n=\sigma\{S_0,J_0,\ldots,S_n,J_n\}$.

Proposition 2.1 *For $s \in \mathcal{I}_{\hat{\boldsymbol{F}}}$ the sequence $(M_n^{(s)})_n$ is a $\mathbb{P}_{(x,i)}$-martingale, called furthermore* Wald martingale. *It is positive and mean one. In particular*

$$\mathbb{E}_i[e^{sS_n}h_{J_n}^{(s)}] = h_i^{(s)}e^{n\kappa(s)}. \tag{2.6}$$

Proof We have

$$\begin{aligned}
&\mathbb{E}_i[\frac{h_{J_{n+m}}^{(s)}}{h_i^{(s)}}e^{sS_{n+m}-(n+m)\kappa(s)} \mid \mathcal{F}_n] \\
&= \mathbb{E}_i[\frac{h_{J_{n+m}}^{(s)}}{h_i^{(s)}}e^{sS_n-n\kappa(s)+s(S_{n+m}-S_n)-m\kappa(s)} \mid \mathcal{F}_n] \\
&= e^{sS_n-n\kappa(s)}\mathbb{E}_i[e^{s(S_{n+m}-S_n)-m\kappa(s)}\frac{h_{J_{n+m}}^{(s)}}{h_i^{(s)}} \mid \mathcal{F}_n] \\
&= e^{sS_n-n\kappa(s)}\mathbb{E}_{J_n}[e^{sS_m}\frac{h_{J_m}}{h_i^{(s)}}]e^{-m\kappa(s)}.
\end{aligned}$$

Now from the definition of $\kappa(s)$ and $\boldsymbol{h}^{(s)}$ we have

$$\mathbb{E}_{J_n}[e^{sS_m}\frac{h_{J_m}}{h_i^{(s)}}]e^{-m\kappa(s)} = \frac{h_{J_n}^{(s)}}{h_i^{(s)}}$$

which shows the martingale property of $(M_n^{(s)})_n$. Since $\mathbb{E}_i M_0^{(s)} = \mathbb{E}_i M_n^{(s)}$ and $\mathbb{E}_i M_0^{(s)} = 1$ a.s. we have that the martingale is mean 1. □

Differentiating (2.6) twicely we obtain the following formulas.

Lemma 2.2 *For $s \in \mathcal{I}_\kappa$*

$$\mathbb{E}_i[S_n e^{sS_n}h_{J_n}^{(s)}] + \mathbb{E}_i[e^{sS_n}k_{J_n}^{(s)}] = e^{n\kappa(s)}k_i^{(s)} + n\kappa'(s)e^{n\kappa(s)}h_i^{(s)},$$

where $\boldsymbol{k}^{(s)} = \mathrm{d}/\mathrm{d}s\boldsymbol{h}^{(s)}$. In particular, setting $s = 0$ (provided $s = 0$, otherwise a special care is needed in taking one sided derivatives)

$$\mathbb{E}_i S_n + \mathbb{E}_i k_{J_n} = k_i + n\kappa'(0),$$

where $\boldsymbol{k} = \boldsymbol{k}^{(0)}$. Furthermore

$$\begin{aligned}
&\mathbb{E}_i[S_n^2 e^{sS_n}h_{J_n}^{(s)}] + 2\mathbb{E}_i[S_n e^{sS_n}k_{J_n}] + \mathbb{E}_i[e^{sS_n}\frac{\mathrm{d}}{\mathrm{d}s}k_{J_n}^{(s)}] \\
&= e^{n\kappa(s)}[\frac{\mathrm{d}}{\mathrm{d}s}k_i^{(s)} + n\kappa'(s)k_i^{(s)} + (\kappa''(s)h_i^{(s)} + n(\kappa'(s))^2 h_i^{(s)} + \kappa'(s)k_i^{(s)})].
\end{aligned}$$

From Lemma 2.2 we obtain the following proposition.

Proposition 2.3 *For every initial distribution* $\boldsymbol{\mu}$

$$\lim_{n\to\infty} \frac{\mathbb{E}_{\mu} S_n}{n} = \kappa^{'}(0),$$
$$\lim_{n\to\infty} \frac{\mathrm{Var}_{\mu} S_n}{n} = \kappa^{''}(0)$$

where derivatives are well defined finite if $0 \in \mathrm{int}(\mathcal{I}_\kappa)$ *and otherwise we have to take one sided derivatives provided they are finite.*

Looking at Proposition 2.3 the following result is intuitively obvious. However the proof require different techniques, and therefore we refer for it to Asmussen (2003). We say that a Markov additive sequence $(S_n, J_n)_n$ is degenerated if $\sup |S_n| < \infty$ a.s.

Proposition 2.4 *Suppose that a Markov additive sequence* $(S_n, J_n)_n$ *is nondegenerated. Then we have*
(i) *if* $\kappa^{'}(0) < 0$, *then* $\lim_{n\to\infty} S_n = -\infty$ *a.s.*,
(ii) *if* $\kappa^{'}(0) > 0$, *then* $\lim_{n\to\infty} S_n = \infty$ *a.s.*,
(iii) *if* $\kappa^{'}(0) = 0$, *then* $\liminf_{n\to\infty} S_n = -\infty$ *a.s. and* $\limsup_{n\to\infty} S_n = \infty$ *a.s.*

Let $\mathbb{P}^{(s)}$ be the probability measure on $(\Omega, \mathbb{F})$ obtained by the change of measure $\mathrm{d}\mathbb{P}^{(s)}_{(x,i)|n} = M_n^{(s)} \mathrm{d}\mathbb{P}_{(x,i)|n}$.

Theorem 2.5 *Under* $\mathbb{P}^{(s)}_{(x,i)}$ *the sequence* $(S_n, J_n)_n$ *is a Markov additive process with the matrix m.g.f.* $\hat{\boldsymbol{F}}^{(s)}$ *of the kernel* $\boldsymbol{F}^{(s)}$ *given by*

$$\hat{\boldsymbol{F}}^{(s)}(v) = e^{-\kappa(s)}(\mathrm{diag}(\boldsymbol{h}^{(s)}))^{-1} \hat{\boldsymbol{F}}(s+v) \mathrm{diag}(\boldsymbol{h}^{(s)}), \tag{2.7}$$

which yields

$$F_{ij}^{(s)}(\mathrm{d}x) = e^{-\kappa(s)} \frac{h_j^{(s)}}{h_i^{(s)}} e^{sx} F_{ij}(\mathrm{d}x). \tag{2.8}$$

Furthermore

$$\boldsymbol{P}^{(s)} = e^{-\kappa(s)}(\mathrm{diag}(\boldsymbol{h}^{(s)}))^{-1} \hat{\boldsymbol{F}}(s) \mathrm{diag}(\boldsymbol{h}^{(s)}) \tag{2.9}$$

and

$$F_{ij}^{\circ,(s)}(\mathrm{d}x) = \frac{e^{xs}}{\hat{F}_{ij}^{\circ}(s)} F_{ij}^{\circ}(\mathrm{d}x). \tag{2.10}$$

Proof Putting

$$Z_i = \frac{h_{J_i}}{h_{J_{j-1}}} e^{sX_j - \kappa(s)},$$

and $L_n = \prod_{j=1}^n Z_j$, we easily see that $(M_n^{(s)})_n$ is a multiplicative functional. What is left, it is to compute the kernel. We have

$$\begin{aligned}
&\hat{F}_{ij}^{(s)}(v) \\
&= \mathbb{E}_i^{(s)}[e^{sX_1}; J_n] = \mathbb{E}_i[M_n^{(s)} e^{sX_1}; J_n] \\
&= e^{-\kappa(s)} \frac{h_j^{(s)}}{h_i^{(s)}} \mathbb{E}_i[e^{(v+s)} X_1; J_n = j] = e^{-\kappa(s)} \frac{h_j^{(s)}}{h_i^{(s)}} \hat{F}_{ij}(s+v).
\end{aligned}$$

Passing to the matrix notations we obtain (2.7). Now setting $v = 0$ we obtain (2.9). Formula (2.10) follows from (2.8)

$$\begin{aligned}
F^{\circ,(s)}(\mathrm{d}x) &= p_{ij}^{(s)} \frac{F_{ij}^{(s)}(\mathrm{d}v)}{p_{ij}^{(s)}} \\
= p_{ij}^{(s)} \frac{e^{sx} F_{ij}(\mathrm{d}x)}{\hat{F}_{ij}(s)} &= \frac{e^{sx} F_{ij}^{\circ}(\mathrm{d}x)}{\hat{F}_{ij}^{\circ}(s)}
\end{aligned}$$

□

Corollary 2.6 *Under $\mathbb{P}_{(x,i)}^{(s)}$ the function $\kappa^{(s)}(v) = \log \theta_{\mathrm{PF}}$, where θ_{PF} is the Perron-Frobenius eigenvalue of $\hat{\boldsymbol{F}}^{(s)}$, is*

$$\kappa^{(s)}(v) = \kappa(s+v) - \kappa(s).$$

Proof From formula (1.5) we have

$$(\hat{\boldsymbol{F}}^{(s)}(v))^n \sim e^{-n\kappa(s)} (\mathrm{diag}(\boldsymbol{h}^{(s)}))^{-1} (\hat{\boldsymbol{F}}(s+v))^n \mathrm{diag}(\boldsymbol{h}^{(s)}),$$

as $n \to \infty$. Now using the Perron–Frobenius theorem we have

$$(\hat{\boldsymbol{F}}^{(s)}(v))^n \sim e^{-n\kappa(s)} (\mathrm{diag}(\boldsymbol{h}^{(s)}))^{-1} e^{n\kappa(s+v)} \boldsymbol{h}\boldsymbol{\xi} \mathrm{diag}(\boldsymbol{h}^{(s)}),$$

from which the lemma follows. □

Let $\nu(u) = \min\{n : S_n > u\}$. As before, let $\mathcal{M} = \sup_{n \geq 0} S_n$ and we have $\mathcal{M} > u$ equivalent to $\nu(u) < \infty$. Assume that the asymptotic drift $\kappa^{'}(0) < 0$ and so $\mathcal{M}$ is a well defined random variable. We have the following counterpart to the Cramer-Lundberg approximations. As before we define the overshoot process $(B(u))_u$, where $B(u) = S_{\nu(u)} - u$. We have to couple the overshoot process with $(I(u))_u$, where $I(u) = J_{\nu(u)}$. The bivariate process $(B(u), I(u))$ is built on the Markov renewal process defined by the ladder epochs ν_n^+ and ladder heights X_n^+ . Here we have to consider a bivariate sequence of ladder epochs $(\nu_n^+, X_n^+)_n$ coupled with a DTMC $(I_n^+)_n$, where $I_n^+ = J_{\nu_1^+ + \ldots + \nu_n^+}$. The kernel is $\boldsymbol{G}^+ = (G_{ij}^+)_{ij}$, where

$$G_{ij}^+ = \mathbb{P}_i(X^+ \in \mathrm{d}x, I_1^+ = j).$$

Notice that if the drift is positive, then the ladder epochs and the corresponding ladder heights are well defined random variables, and they form a Markov additive sequence

Similarly like for the renewal process we can proof the following lemma.

Lemma 2.7 *If $\hat{\boldsymbol{F}}(s)$ is well defined for some $s > 0$, then $\hat{\boldsymbol{G}}^+(s)$ is well defined too.*

Assuming the drift to infinity, Lemma 2.7, plus some nonlattice type assumptions, which will not be discussed here, imply that the Markov renewal theorem can be used to prove, that

$$(B(u), I(u)) \xrightarrow{\mathrm{d}} (B(\infty), I(\infty)),$$

as $t \to \infty$.

Proposition 2.8 *[Cramer-Lundberg approximation for MAS] Suppose that $\boldsymbol{F}$ is nonlattice, $\kappa^{'}(0) < 0$ and that there exists $\gamma > 0$ such that $\kappa(\gamma) = 0$. Furthermore we assume that $\kappa^{'}(\gamma) < \infty$. Then*

$$\lim_{u \to \infty} \frac{\mathbb{P}_i(\nu(u) < \infty)}{c h_i^{(\gamma)} e^{-\gamma u}} = 1.$$

where the constant $c \geq 0$ is finite and given by

$$c = \mathbb{E}_i^{(\gamma)} \left[\frac{e^{-\gamma B(\infty)}}{h_{I(\infty)}^{(\gamma)}} \right] . \tag{2.11}$$

Proof

$$\begin{aligned}
\mathbb{P}_i(\tau(u) < \infty) &= \mathbb{E}_i^{(\gamma)}[\frac{h_i^{(\gamma)}}{h_{I(u)}^{(\gamma)}} e^{\gamma S_{\tau(u)}}; \tau(u) < \infty] \\
&= e^{-\gamma u} h_i^{(\gamma)} \mathbb{E}_i^{(\gamma)} \frac{e^{-\gamma B(u)}}{h_{I(u)}^{(\gamma)}}
\end{aligned}$$

We now have to study $(B(u), I(u))_u$ under $\mathbb{P}_i^{(\gamma)}$, to prove

$$\mathbb{E}^{(\gamma)} \frac{e^{-\gamma B(u)}}{h_{I(u)}^{(\gamma)}} \to c > 0.$$

□

3 Gärtner-Ellis theorem

The following is an extension of Cramer's theorem for non random walk cases. Suppose there is given a sequence $(Z_n)_n$, with

$$\kappa_n(s) = \log \mathbb{E}\, e^{sZ_n},$$

fulfilling for all $s \in \mathbb{R}$

$$\kappa(s) = \lim_{n\to\infty} \frac{1}{n} \kappa_n(ns) \in \mathbb{R} \cup \{\infty\}.$$

Clearly $\kappa(s)$ is convex as a limit of convex functions. We will assume moreover that $\kappa(s)$ is *essentially smooth* , that is fulfilling

- $\mathrm{int}(\mathcal{I}_\kappa) \neq \{0\}$,
- $\kappa(s)$ is differentiable throughout $\mathrm{int}(\mathcal{I}_\kappa)$,
- $\kappa(s)$ is *steep* it reaches infinity in a continuous way.

Let $\kappa^*(s)$ be the Fenchel–Legendre transform of $\kappa(s)$.

Theorem 3.1 *[Gärtner–Ellis] Assume that $\kappa(s)$ is an essentially smooth function. Then $(Z_n)_n$ fulfills the large deviation principle with a good rate function $\kappa^*(s)$.*

Example 3.2 Suppose that $(S_n, J_n)_n$ is a MAS, with $\hat{\boldsymbol{F}}(s)$ defined for $s \in \mathcal{I}_{\boldsymbol{F}}$, and $\mathcal{I}_{\boldsymbol{F}} \neq \{0\}$. Moreover we assume that E is finite and that $\boldsymbol{P}$ is primitive and so the $\kappa(s)$, where $e^{\kappa}(s)$ is the Perron-Frobenius eigenvalue of $\hat{\boldsymbol{F}}(s)$, is well defined. Set $Z_n = S_n/n$ and

$$\begin{aligned}\kappa_n(s) &= \log \mathbb{E}_i e^{sZ_n} \\ &= = \log \boldsymbol{\varepsilon}_i(\hat{\boldsymbol{F}}(s/n))^n \boldsymbol{e}.\end{aligned}$$

Using Perron–Frobenius theorem we may write

$$\boldsymbol{\varepsilon}_i(\hat{\boldsymbol{F}}(s))^n \boldsymbol{e} = \text{const } e^{n\kappa(s)} + O(n^k|\theta_2|^n),$$

and hence

$$(\boldsymbol{\varepsilon}_i(\hat{\boldsymbol{F}}(s))\boldsymbol{e})^{1/n} = (\text{const } e^{n\kappa(s)} + O(n^k|\theta_2|^n))^{1/n} \to e^{\kappa(s)}.$$

Thus

$$\lim_{n\to\infty} \frac{1}{n} = \log \mathbb{E}_i e^{sZ_n} = \kappa(s).$$

Example 3.3 Suppose that $(X_n)_n$ is a Gaussian sequence, centered, and such that $\sum_{j=1}^{\infty} |R(j)| < \infty$, where

$$R(n) = \mathbb{E}\, X_n X_0$$

is its covariance function. Let $S_n = \sum_{j=1}^{n} X_j$ and $Z_n = S_n/n$. Clearly Z_n has normally distributed with mean zero. To compute $\kappa_n(s)$ we must find the formula for $\operatorname{Var} S_n$. Thus

$$\begin{aligned}\operatorname{Var} Z_n &= \mathbb{E}\,(X_1 + \ldots + X_n)^2 \\ &= \mathbb{E}\,(\sum_{i=1}^{n} X_i^2 + 2\sum_{1\le i<j\le n} X_i X_j) = 2\sum_{j=1}^{n}\sum_{i=1}^{j-1} R(j-i) + nR(0) \\ &= 2\sum_{j=1}^{n}\sum_{i=1}^{j-1} R(i) + nR(0).\end{aligned}$$

Hence

$$\frac{1}{n}\kappa_n(s) = \frac{s^2}{2}\frac{1}{n}(2\sum_{j=1}^{n}\sum_{i=1}^{j-1} R(i) + nR(0)) \to \frac{s^2}{2}(R(0) + 2\sum_{j=1}^{\infty} R(j)).$$

Chapter V

Conditioned DTMCs and exit times asymptotics

Let $E = \{1, 2, \dots\}$ and assume $\boldsymbol{P} = (p_{ij})_{i,j=E})$ is substochastic. Extending $\boldsymbol{P}$ to $\boldsymbol{P}^*$ according to (I.1.16), we may construct a DTMC (X_n) on $(\Omega, \mathcal{F}, \mathbb{P}_i)$ with state space $E = \{0, 1, \dots\}$, and transition matrix $\boldsymbol{P}^*$. Let ν be the time of hitting 0 of the DTMC (X_n).

1 Quasi-stationary distribution for DTMCs

Recall the definition from Chapter I. A distribution $\boldsymbol{\pi}$ on $\{1, 2, \dots\}$ is called *quasi-stationary* (QS) if

$$\mathbb{P}_{\boldsymbol{\pi}^*}(X_n = j \mid \nu > n) = \pi_j, \qquad j = 1, \dots,$$

where $\boldsymbol{\pi}^* = (0, \boldsymbol{\pi})$ is a probability function on $\{0, 1, \dots\}$. In vector notation we may write that $\boldsymbol{\pi}$ is QS if and only if

$$\boldsymbol{\pi} \boldsymbol{P}^n = (\boldsymbol{\pi} \boldsymbol{P}^n \boldsymbol{e}) \boldsymbol{\pi}, \qquad n = 0, 1, \dots$$

We may define another DTMC (Y_n) with state space $E = \{1, 2, , \dots\}$ for a given $\boldsymbol{P}$ substochastic and $\boldsymbol{\rho}$ a true probability function on E as follows. We define a true stochastic matrix $\boldsymbol{P}^\rho$ by

$$p_{ij}^\rho = p_{ij} + (1 - \sum_{k \geq 1} p_{ij}) \rho_j.$$

If $\boldsymbol{P}$ is irreducible and aperiodic, then $\boldsymbol{P}^\rho$ is too. Notice that ν has the same distribution for both the chains. Suppose that

$$\sum_{j \geq 1} \rho_j \mathbb{E}_j \nu < \infty. \tag{1.1}$$

Then there exists the stationary distribution $\boldsymbol{\zeta}^\rho$ for (Y_n), that is

$$\boldsymbol{\zeta}^\rho \boldsymbol{P}^\rho = \boldsymbol{\zeta}^\rho$$

or

$$\boldsymbol{\zeta}^\rho \boldsymbol{P} + (1 - \boldsymbol{\zeta}^\rho \boldsymbol{P}\boldsymbol{e})\boldsymbol{\rho} = \boldsymbol{\zeta}\boldsymbol{\rho}. \tag{1.2}$$

Lemma 1.1 $\boldsymbol{\pi} = \boldsymbol{\rho}$ *is QS if and only if* $\boldsymbol{\zeta}^{\boldsymbol{\pi}} = \boldsymbol{\pi}$. *Hence* $\pi_i > 0$ *for all* i.

Proof Suppose that $\boldsymbol{\pi}$ is QS. Then from (1.2) we write

$$\boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P} + (1 - \boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P}\boldsymbol{e})\boldsymbol{\pi} = \boldsymbol{\zeta}^{\boldsymbol{\pi}},$$

which rewrite in the form

$$\frac{1 - \boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P}\boldsymbol{e}}{1 - \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e}}(1 - \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e})\boldsymbol{\pi} = \boldsymbol{\zeta}^{\boldsymbol{\pi}}(\boldsymbol{I} - \boldsymbol{P}).$$

Since for $\boldsymbol{\pi}$ QS $\boldsymbol{\pi} - (\boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e})\boldsymbol{\pi} = \boldsymbol{\pi} - \boldsymbol{\pi}\boldsymbol{P}$ we obtain

$$\frac{1 - \boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P}\boldsymbol{e}}{1 - \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e}}\boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{\zeta}^{\boldsymbol{\pi}}(\boldsymbol{I} - \boldsymbol{P}).$$

Hence (see Exercise 1.1) we may conclude that

$$(1 - \boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P}\boldsymbol{e})\boldsymbol{\pi} = (1 - \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e})\boldsymbol{\zeta}^{\boldsymbol{\pi}}.$$

Since $\boldsymbol{\pi}$ and $\boldsymbol{\zeta}^{\boldsymbol{\pi}}$ are true probability functions, hence $1 - \boldsymbol{\zeta}^{\boldsymbol{\pi}} \boldsymbol{P}\boldsymbol{e} = 1 - \boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e}$. The converse can be verified by substitution. Now the second claim follows immediately becasue the stationary measures charges all points. □

Problem

1.1 Show that if $\boldsymbol{P}$ is irreducible and substochastic, then for $\boldsymbol{\eta}_1(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{\eta}_2(\boldsymbol{I} - \boldsymbol{P})$ we have $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_2$. (Hint: Use that $\boldsymbol{P}^n \to 0$ and the identity $(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{I} + \boldsymbol{P} + \ldots + \boldsymbol{P}^{n-1} = \boldsymbol{I} - \boldsymbol{P}^n$.)

Let $E = \{1, 2, \ldots\}$ and assume $\boldsymbol{P} = (p_{ij})_{i,j=E})$ is substochastic. Let ν be the time of hitting 0. We assume that ν is finite $\mathbb{P}_i$–a.s. Actually we need even stronger assumptions which yields that $\mathbb{E}_i\nu < \infty$ $(i \in E)$. Our aim is to prove the following theorem.

Theorem 1.2 *Suppose that $\boldsymbol{P}$ is irreducible, such that $\mathbb{P}_i(\nu < \infty) = 1$ for some $i \in E$ (and hence for all i) and*

$$\lim_{i \to \infty} \mathbb{P}_i(\nu < t) = 0$$

for all $t \geq 0$. A necessary and sufficient condition for the existence of a QS distribution is

$$\mathbb{E}_i R^\nu < \infty$$

for some $R > 1$ and some $i \in E$ (and hence for all i).

Proof of the necessary part. Suppose that $\boldsymbol{\pi}$ is QS. Note that

$$\mathbb{P}_{\boldsymbol{\pi}}(\nu > n) = \boldsymbol{\pi}\boldsymbol{P}^n\boldsymbol{e}.$$

Since $\boldsymbol{\pi}$ is QS, we have

$$\boldsymbol{\pi}\boldsymbol{P}^n = (\boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e})^n\boldsymbol{\pi}.$$

Hence $\boldsymbol{\pi}\boldsymbol{P}^n\boldsymbol{e} = (\boldsymbol{\pi}\boldsymbol{P}\boldsymbol{e})^n$. Therefore

$$\mathbb{P}_{\boldsymbol{\pi}}(\nu > n + m) = \mathbb{P}_{\boldsymbol{\pi}}(\nu > n)\mathbb{P}_{\boldsymbol{\pi}}(\nu > m)$$

for all $n, m = 0, 1, \ldots$. The solution is $\mathbb{P}_{\boldsymbol{\pi}}(\nu > n) = r^n$ for some $r < 1$ (must be strictly less than 1 becasue ν is a true random variable). Hence $\mathbb{E}_{\boldsymbol{\pi}} R^\nu < \infty$ for all $R < 1/r$. Since $\mathbb{E}_i R^\nu < (1/\pi_i)\mathbb{E}_{\boldsymbol{\pi}} R^\nu$, we have the proof of the necessary conditon completed. $\square$

Comments. Theorem 1.2 is from Ferrari *et al* (1995).

2 Kesten's strong ratio limit property theorem

Kesten (1995)

3 Asymptotics for ν^+ and other exit times in random walk

Doney (1989)

Appendix A

1 Moment generating function and other transforms

For a random variable X with distribution F, let $\mathcal{I}_{\hat{F}}\{s \in \mathbb{R} : \mathbb{E}\, \mathrm{e}^{sX} < \infty\}$. Note that $\mathcal{I}_{\hat{F}}$ is an interval which can be the whole real line $\mathbb{R}$, a halfline or even the singleton $\{0\}$. We will call $\hat{F}(s) = \mathbb{E}\, \mathrm{e}^{sX}$ by *moment generating function*.

It is somewhat delicate to decide for what arguments s the transform $\hat{F}(s)$ is well-defined. For example, if X is nonnegative, then $\hat{F}(s)$ is well-defined for all $s \leq 0$. There are also examples when $\hat{F}(s)$ may be ∞ for all $s > 0$, while others show that $\hat{F}(s)$ is finite on $(-\infty, a)$ for some $a > 0$.

If the n-th moment of the random variable $|X|$ is finite, then the n-th derivative $\hat{F}(s)$ exist at zero, perhaps one sided, provided $\hat{m}(s)$, is well-defined in a certain (complex) neighbourhood of $s = 0$. In this case, the following equation holds:

$$\mathbb{E}\, X^n = \hat{F}^{(n)}(0)\,. \tag{A.1}$$

If $\hat{F}(s)$ is well-defined only on $(-\infty, 0]$ and $[0, \infty)$, respectively, then the derivatives in (A.1) have to be replaced by one-sided derivatives, i.e.

$$\mathbb{E}\, X^n = \hat{F}^{(n)}(0-)\,. \tag{A.2}$$

Consider a real-valued (not necessarily nonnegative) random variable X with distribution F and moment generating function $\hat{F}(s) = \mathbb{E}\, \mathrm{e}^{sX} = \int_{-\infty}^{\infty} \mathrm{e}^{sx}\, \mathrm{d}F(x)$ for all $s \in \mathbb{R}$ for which this integral is finite. Let $s_F^- = \inf\{s \leq 0 :\ \hat{F}(s) < \infty\}$ and $s_F^+ = \sup\{s \geq 0 :\ \hat{F}(s) < \infty\}$ be the lower and upper *abscissa of convergence*, respectively, of the moment generating function $\hat{F}(s)$. Clearly $s_F^- \leq 0 \leq s_F^+$. Assume now that $\hat{m}_F(s)$ is finite for a value $s \neq 0$. From the definition of the moment generating function we see that $\hat{F}(s)$ is a well-defined, continuous and strictly increasing function of $s \in (s_F^-, s_F^+)$ with value 1 at the origin. Furthermore,

$$\begin{aligned}
\hat{F}(s) - 1 &= \int_{-\infty}^{\infty} (\mathrm{e}^{sx} - 1)\, \mathrm{d}F(x) \\
&= -s \int_{-\infty}^{0} \int_{x}^{0} \mathrm{e}^{sy}\, \mathrm{d}y\, \mathrm{d}F(x) + s \int_{0}^{\infty} \int_{0}^{x} \mathrm{e}^{sy} \mathrm{d}y\, \mathrm{d}F(x) \\
&= -s \int_{-\infty}^{0} F(y) \mathrm{e}^{sy}\, \mathrm{d}y + s \int_{0}^{\infty} \overline{F}(y) \mathrm{e}^{sy}\, \mathrm{d}y. \qquad \text{(A.3)}
\end{aligned}$$

This relation is useful in order to derive a necessary and sufficient condition that $\hat{F}(s) < \infty$ for some $s \neq 0$.

Lemma A.1 *Assume that* $\hat{F}(s_0) < \infty$ *for some* $s_0 > 0$. *Then there exists* $b > 0$ *such that for all* $x \geq 0$

$$1 - F(x) \leq b\, \mathrm{e}^{-s_0 x}\,. \qquad \text{(A.4)}$$

Conversely, if (A.4) *is fulfilled, then* $\hat{F}(s) < \infty$ *for all* $0 \leq s < s_0$. *Analogously, if* $\hat{F}(s_0) < \infty$ *for some* $s_0 < 0$, *then there exists* $b > 0$ *such that for all* $x \leq 0$

$$F(x) \leq b\, \mathrm{e}^{s_0 x}\,. \qquad \text{(A.5)}$$

Conversely, if (A.5) *is fulfilled, then* $\hat{F}(s) < \infty$ *for all* $s_0 < s \leq 0$.

Proof Assume that condition (A.4) is fulfilled. Then (A.3) leads to

$$\frac{\hat{m}_F(s) - 1}{s} \leq \int_0^{\infty} (1 - F(y))\, \mathrm{e}^{sy}\, \mathrm{d}y \leq \frac{b}{s_0 - s}$$

for $0 < s < s_0$, which shows that $\hat{m}_F(s)$ is finite at least for all $0 \leq s < s_0$. Conversely, if $\hat{m}_F(s)$ is finite for a positive value $s = s_0$, then for any $x \geq 0$

$$\begin{aligned}
\infty > \frac{\hat{F}(s_0) - 1}{s_0} &\geq -\int_{-\infty}^{0} F(y) \mathrm{e}^{s_0 y}\, \mathrm{d}y + \int_0^x \mathrm{e}^{s_0 y} (1 - F(y))\, \mathrm{d}y \\
&\geq -\frac{1}{s_0} + (1 - F(x)) \frac{\mathrm{e}^{s_0 x} - 1}{s_0}\,.
\end{aligned}$$

This means that for all $x \geq 0$, $1 - F(x)$ is bounded by $b\, \mathrm{e}^{-s_0 x}$ for some constant b. The second part of the lemma can be proved similarly. $\square$

Theorem A.2 *If* $a^+ = \liminf_{x\to\infty} -x^{-1}\log \overline{F}(x) > 0$, *then*

$$a^+ = s_F^+ \,. \tag{A.6}$$

If $a^- = \limsup_{x\to-\infty} -x^{-1}\log F(x) < 0$, *then*

$$a^- = s_F^- \,. \tag{A.7}$$

Proof We show only (A.6). Let $\varepsilon > 0$ be such that $a^+ - \varepsilon > 0$. Then, there exists $x_0 > 0$ such that $-x^{-1}\log \overline{F}(x) \ge a^+ - \varepsilon$ for $x > x_0$, which is equivalent to $\overline{F}(x) \le \mathrm{e}^{-(a^+-\varepsilon)x}$ for $x > x_0$. Because $\varepsilon > 0$ was arbitrary we conclude from Lemma A.1 that $\hat{F}(s) < \infty$ for all $s < a^+$. Conversely, suppose that $\hat{F}(s_0) < \infty$ for some $s_0 > a^+$. By Lemma A.1, $\overline{F}(x) \le b\mathrm{e}^{-s_0 x}$ for some $b > 0$. Hence $-x^{-1}\log \overline{F}(x) \ge -x^{-1}\log b + s_0$ for $x > 0$, which yields $a^+ \ge s_0$. This contradicts $s_0 > a^+$. Therefore $s_F^+ = a^+$. □

From the above considerations we get $s_F^+ = \liminf_{x\to\infty} -x^{-1}\log \overline{F}(x)$ and $s_F^- = \limsup_{x\to-\infty} x^{-1}\log F(x)$ provided that the limits are nonzero. Note that for nonnegative random variables $s_F^- = -\infty$, and in this case we write $s_F^+ = s_F$.

By the logarithmic moment generating function we call $\hat{\kappa}_F(s) = \log \hat{F}(s)$. The values of consecutive derivatives of the logarithmic moment generating function at zero are known under the name of *semi-invariants*. We propose to check that the first two semi-invariantes are the mean and variance.

Problems

1.1 Show $\kappa_X^{'}(0) = \mathbb{E}\,X$ and $\kappa_X^{''}(0) = \mathrm{Var}\,X$

2 Absolute Continuity of Probability Measures

Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be two probability measures on $(\Omega, \mathcal{F})$.

Definition A.1 It is said $\tilde{\mathbb{P}}$ is *absolutely continuous* with respect $\mathbb{P}$ that $\mathbb{P}(A) = 0$ for $A \in \mathcal{F}$ yields $\tilde{\mathbb{P}}(A) = 0$. We write then $\tilde{\mathbb{P}} \ll \mathbb{P}$. If $\mathbb{P} \ll \tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}} \ll \mathbb{P}$ then we say that the measures are *equivalent* and we write $\tilde{\mathbb{P}} \equiv \mathbb{P}$

Theorem A.2 *[Radon-Nikodym theorem] Suppose that* $\tilde{\mathbb{P}} \ll \mathbb{P}$. *Then there exists a nonnegative random variable* M *such that*

$$\tilde{\mathbb{P}}(A) = \int_A M\,\mathrm{d}P, \qquad A \in \mathcal{F}$$

Actually the theorem is true for $\mathbb{P}$ and $\tilde{\mathbb{P}}$ being a σ-finite measures. The r.v. M from Radon-Nikodym theorem is called *likelihood ratio* or *(Radon-Nikodym) density* and we denote it by $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$. We have the following converese statement.

Proposition A.3 *If* $M = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} > 0$ $\mathbb{P}$*-a.s. then* $\mathbb{P} \ll \tilde{\mathbb{P}}$ *and*

$$\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}} = M^{-1}, \qquad \mathbb{P} - a.s.$$

(see [?], p. 227).

Proposition A.4 *Suppose that for* σ*-finite measures* $\mathbb{P}_1 \ll \mathbb{P}_2$ *and* $\mathbb{P}_2 \ll \mathbb{P}_3$*. Then* $\mathbb{P}_1 \ll \mathbb{P}_3$ *and*

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_3} = \frac{d\mathbb{P}_1}{d\mathbb{P}_2}\frac{d\mathbb{P}_2}{d\mathbb{P}_3}.$$

Jacod, J & Shiryaev, A.N. *Limit Theorems for Stochastic Processes* Springer-Verlag, 1987.

Liptser, R. & Shiryayev, A.N. (1977) *Statistics of Random Processes I; General Theory* Springer-Verlag, New York.

List of Principal Notation

Matrices and vectors

$\boldsymbol{A} = (a_{ij})_{ij}$, $\boldsymbol{X}$, $\boldsymbol{\Xi}$	matrices
$\boldsymbol{a}$,$\boldsymbol{x}$	column vectors
$\boldsymbol{\alpha}$,$\boldsymbol{\xi}$	row vectors
$\boldsymbol{A}^\top$	transposition of matrix $\boldsymbol{A}$
$\boldsymbol{I}$	identity matrix
$\boldsymbol{A}^n = (a_{ij}^{(n)})_{ij}$	n-th power of $\boldsymbol{A}$
$\mathbf{0}$	matrix with all zero elements
$\boldsymbol{\varepsilon} = (1, \ldots, 1)$	row vector with all components equal to 1
$\boldsymbol{\varepsilon}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$	
$\boldsymbol{e} = \boldsymbol{\varepsilon}^\top$	column vector with all components equal to 1
$\boldsymbol{e}_i = \boldsymbol{\varepsilon}_i^\top$	
$\mathrm{diag}(\boldsymbol{\alpha}) = (\alpha_i \delta_{ij})_{ij}$	diagonal matrix

Numbers and spaces

$\mathbb{R}$	real numbers
$\mathbb{R}_+$	nonnegative real numbers
$\mathbb{Z}$	integer numbers
$\mathbb{Z}_+$	nonnegative integer numbers
$\mathcal{B}(\mathbb{R})$, $\mathcal{B}(E)$	Borel σ-algebra on $\mathbb{R}$, E
$1(A)$	indicator of event A
$\mathbb{C}$	complex plane
$\Re z$	real part of z
$\Im z$	imaginary part of z
E	state space
$\emptyset$	empty set
$\mathcal{I}_f = \{s : \ \lvert f(s)\rvert < \infty$	the domain of f

Probability and expectation

Ω	basic space
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
$\mathcal{F}, \mathcal{F}_n, \mathcal{G}$	σ-algebras
$\mathbb{E}$	expectation
$\mathbb{E}(X \mid \mathcal{G})$	conditional expectation
Cov	covariance
Var	variance
F, G	distributions; distribution functions
$F * G$	convolution
F^{*k}	k-th convolution of F; $F^{*1} = F$, $F^{*0} = \delta_0$
F^{s}	integrated tail distribution
$\overline{F}(x)$	tail function
$F^{-1}(y)$	generalized inverse function
$f(t)$	density function
μ_n	n-th moment
ν^-, ν^+	ladder epochs
τ, ν	stopping time

Functions

$\hat{F}(s)$	moment generating function of F
$g_1(x) \sim g_2(x)$	asymptotical equivalence as $x \to \infty$

Other

$\lfloor x \rfloor$	largest integer $\le x$
$\lceil x \rceil$	smallest integer $\ge x$
$x \wedge y = \min\{x, y\}$	
$x \vee y = \max\{x, y\}$	
δ_{ij}	Kronecker delta function
γ	adjustment coefficient
$\sum_{i=1}^{j} a_i = 0$ if $j = 0$	
$\prod_{i=1}^{j} a_i = 1$ if $j = 0$	
$\inf \emptyset = \infty$	

Abreviations

i.i.d.	independent and identically distributed
DTMC	discrete time Markov chain
CTMC	continuous time Marov chain
OECM	optimal exponential change of measure
ECM	exponential change of measure
QS	quasi-stationary

Bibliography

[1] Asmussen S. (1997) *Fundamentals of Ruin Probability Theory.* World Science Publishers, Singapore.

[2] Asmussen S. (1998) *Stochastic Simulation with a view towards stochastic processes.* MaPhySto, University of Aarhus, Denmark.

[3] Asmussen S. (2003) *Applied Probability and Queues.* Springer, New York.

[4] Asmussen S. and Rubinstein R.Y. (1995) Steady-state rare events simulation in queueing models and its complexity properties. *In: J. H. Dshalalov (Ed.), Advances in Queueing,* CRC Press, Boca Raton 429–461.

[5] Chung K.L. (1974) *A Course in Probability Theory.* Academic Press, New York.

[6] Collamore J. (2002) Importance sampling techniques. *Ann. Appl. Prob.* **12**: 382–421.

[7] Dembo A. and Zeitouni O. (1993) *Large Deviations Techniques and Applications.* Jones and Bartlett Publishers, Boston.

[8] Doney R. (1989) On the asymptotic behaviour of first passage times for transient random walk. *Probab. Th. Rel. Fields* **81**: 239–246.

[9] Dupuis P. and Wang H. (2003) Importance sampling, large deviation, and differentional games. *MS* .

[10] Feller W. (1971) *An Introduction to Probability Theory and its Applications, vol 2..* J. Wiley & Sons, New York, 2nd ed.

[11] Glasserman P. and Wang Y. (1997) Counter examples in importance sampling for large deviations probabilites. *Ann. Appl. Prob.* **7**: 731–746.

[12] Kesten H. (1995) A ratio limit theorem for (sub) markov chains on $\{1, 2, \ldots\}$ wit bounde jumps. *Adv. Appl. Prob.* **27**: 652–691.

[13] Kunita H. and Watanabe T. (1963) Notes on transformations of markov processes connected with multiplicative functional. *Memoirs of the Faculity of Science, Kyushu University* **17**: 181–191.

[14] P.A. Ferrari H. Kesten S.M. and Picco P. (1995) Existence of quasi-stationary distributions. a renwal dynamical approach. *Ann Probab.* **23**: 501–521.

[15] Prabhu N.U. (1980) *Stochastic Storage Processes*. Springer-Verlag, New York.

[16] Resnick S.I. (1992) *Adventures in Stochastic Processes*. Birkhäuser, Boston.

[17] Rolski T., Schmidli H.P., Schmidt V., and Teugels J. (1999) *Stochastic Processes for Insurance and Finance*. Wiley, Chichester.

[18] Sadowsky J. (1996) On monte carlo estimation of large deviations probabilities for multidimensional ruin problem for general markov additive sequences of random vectors. *Ann. Appl. Prob.* **6**: 399–422.

[19] Seneta E. (1981) *Non-Negative Matrices and Markov Chains*. Springer-Verlag, New York.

[20] Seneta E. and Vere-Jones D. (1966) On quasi-stationary behaviour in the random walk in discrete time markov chains with a denumerable infinite of states. *J. Appl. Probab.* **3**: 403–434.

[21] Vere-Jones D. (1962) Geometric ergodicity in denumerable markov chains. *Quart. J. Math. Oxford Ser .2* **13**: 7–28.

Index