Queues and Simulations

Advanced Methods

Tomasz Rolski

Mathematical Institute, Wrocław University

22 April 2004, Wrocław 2004

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. $\boldsymbol{A} = (a_{ij})_{i,j=1,\ldots}$ 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 $\boldsymbol{\varepsilon} = (1, \ldots, l)$, 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 $\boldsymbol{e} = (1, \ldots, l)$, 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 $a_{ij} \neq 0$.

We say that $\boldsymbol{\gamma} = (\gamma_j)$ is a probability vector if $\gamma_i \geq 0$ for all i and $\sum_i \gamma_i = 1$.

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. 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 $\operatorname{diag}(\boldsymbol{\theta})$ 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

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

We make a number of observations.

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

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

- From (1.6) we get

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

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

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

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*

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

Proof We show the asserted independence property by induction. Because the eigenvector $\boldsymbol{x}_1$ has at least one component different from 0, the only solution to $a_1\boldsymbol{x}_1 = \mathbf{0}$ is $a_1 = 0$. Assume now that $\theta_1, \ldots, \theta_l$ are all distinct and that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1}$ are linearly independent for some $k \le l$. In order to prove that also the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are linearly independent, we have to show that

$$\sum_{j=1}^{k} a_j \boldsymbol{x}_j = \mathbf{0} \tag{1.9}$$

implies $a_1 = \ldots = a_k = 0$. If (1.9) holds, then

$$\mathbf{0} = \boldsymbol{A}\mathbf{0} = \sum_{j=1}^{k} a_j \boldsymbol{A}\boldsymbol{x}_j = \sum_{j=1}^{k} a_j \theta_j \boldsymbol{x}_j .$$

On the other hand, $\mathbf{0} = \theta_k \mathbf{0} = \theta_k \sum_{j=1}^{k} a_j \boldsymbol{x}_j = \sum_{j=1}^{k} \theta_k a_j \boldsymbol{x}_j$. This gives $\mathbf{0} = \sum_{j=1}^{k-1} (\theta_k - \theta_j) a_j \boldsymbol{x}_j$ and, consequently,

$$(\theta_k - \theta_1)a_1 = (\theta_k - \theta_2)a_2 = \ldots = (\theta_k - \theta_{k-1})a_{k-1} = 0 .$$

Hence $a_1 = a_2 = \ldots = a_{k-1} = 0$, because $\theta_k \ne \theta_j$ for $1 \le j \le k-1$. This implies $a_k = 0$ by (1.9), and so (1.8) is a direct consequence of $\boldsymbol{\Xi} = \boldsymbol{X}^{-1}$. □

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, $\boldsymbol{A} > \mathbf{0}$.

Recall that $\boldsymbol{A}$ is irreducible if $\sum_n \boldsymbol{A}^n > \mathbf{0}$. Let $\boldsymbol{A} = (a_{ij}^{(n)}$. We now define a period of i if Let d_i be the greatest common divisor of $\{a_{ii}^{(1)}, \ldots, a_{ii}^{n}$ and $d_i = \lim_{n\to\infty}$. It is said that d is the period of i. 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.

Problems

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, 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 $\mathbf{0} \leq \boldsymbol{B} \leq \boldsymbol{A}$ is another nonnegative matrix with an eigenvalue θ', then $|\theta'| \leq \theta_{\mathrm{PF}}$.*

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 behavior of powers $\boldsymbol{A}^n$, for $n \to \infty$. By $(g(n))$ we denote a sequence such that for some numbers $0 < \alpha < \beta < \infty$: $\alpha g(n) \leq O(f(n)) \leq \beta g(n)$, 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 $n \to \infty$,*

$$\boldsymbol{A}^n = \theta_1^n \boldsymbol{x}_1 \boldsymbol{\xi}_1 + O(n^k |\theta_2|^n)\,; \tag{1.10}$$

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

$$\boldsymbol{A}^n = \theta_1^n \boldsymbol{x}_1 \boldsymbol{\xi}_1\,. \tag{1.11}$$

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 1.4 holds with the following modification in point (a): θ_1 is strictly positive and of multiplicity 1 and moreover $|\theta_1| > |\theta_i|$ 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

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

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] Seneta, p.22

1.3 Applications to Markov chains

We consider here finite state space discrete time Markov chains (DTMC), defined be 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 $\boldsymbol{\pi P} = \boldsymbol{\pi}$. We will call it *stationary distribution.*

Corollary 1.6 *If $\boldsymbol{P}$ is a primitive stochastic matrix, then*
(a) $\theta_1 = 1, \boldsymbol{x}_1 = \boldsymbol{e}$ *and* $\boldsymbol{\xi}_1 = \boldsymbol{\pi}$;
(b) $|\theta_i| < 1$ *for* $i = 2, \ldots, l$.

Proof By inspection we get that $\boldsymbol{Pe} = \boldsymbol{e}$. 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. Namely, let θ be some eigenvalue of $\boldsymbol{P}$, and $\boldsymbol{x} = (\phi_1, \ldots, \phi_l)$ the corresponding right eigenvector. Then, (1.1) gives $|\theta|\,|\phi_i| \leq \sum_{j=1}^{l} p_{ij}|\phi_j| \leq \max_{j\in E}|\phi_j|$ for each $i \in E$. Hence $|\theta| \leq 1$. Thus, Theorem 1.3 gives that $|\theta_i| < 1$ for $i = 2, \ldots, l$. □

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.

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

$$\boldsymbol{e\pi} = \begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} \tag{1.12}$$

and such the matrix we denote by $\boldsymbol{\Pi}$.

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

$$\begin{aligned} |p_{ij}^{(n)} - \pi_j| &= O(n^k|\theta_2|^n) \\ &= O(c^n) \end{aligned} \tag{1.13}$$

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

Proof

1.4 Quasi-stationary distribution

Suppose that $\boldsymbol{P} = (p_{ij})_{i,j=1,\dots}$ 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-Frobieniu vector. Recall that according to our convention, $\boldsymbol{\pi}$ is normalized to be a probability function. If $\boldsymbol{P}$ is primitive, then

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

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:

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

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 $\boldsymbol{\varepsilon}_i$ and transition matrix $\boldsymbol{P}'$. To avoid trivialities $i \in \{1, \dots, l\}$. Define $\tau = \inf\{n \ge 0 : X_n = 0\}$. The following observation will be important for us.

Problems

1.1 Prove that τ 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.*
(ii)

$$\mathbb{P}_{\boldsymbol{\pi}}(X_n = j \mid \tau > n) = \pi_j, \qquad j = 1, \dots, l \tag{1.16}$$

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

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

$$\mathbb{P}_{\boldsymbol{\pi}}(X_n = j \mid \tau > n) = \pi_j, \qquad j = 1, \ldots \tag{1.17}$$

We can write (1.16) in the vector notation

$$\boldsymbol{\pi}\boldsymbol{P}^n = \boldsymbol{\pi}\boldsymbol{P}^n\boldsymbol{e}\boldsymbol{\pi}, \qquad n = 0, 1, \ldots. \tag{1.18}$$

Proof of Proposition 1.8. (i)→(ii). Then iterating the eigenvalue equation we have

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

which immediately can be written in form (1.18).
(ii)→(i). We have that for some probability vector $\boldsymbol{\gamma}$

$$\boldsymbol{\gamma}\boldsymbol{P}^n = \boldsymbol{\gamma}\boldsymbol{P}^n\boldsymbol{e}\boldsymbol{\gamma},$$

for all n. The above equation for $n = 1$ yields that $\boldsymbol{\gamma}\boldsymbol{P}\boldsymbol{e}$ and $\boldsymbol{\gamma}$ are an eigenvalue and the corresponding left eigenvector respectively. Then $\boldsymbol{\gamma}\boldsymbol{P}^n\boldsymbol{e} = (\boldsymbol{\gamma}\boldsymbol{P}\boldsymbol{e})^n$. Assume that $\boldsymbol{\gamma} \neq \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the Perron-Frobenius probability vector. Then $\boldsymbol{\gamma}\boldsymbol{P}\boldsymbol{e} < \theta_{\mathrm{PF}}$. Applying (1.14) we have

$$\boldsymbol{\gamma}\boldsymbol{P}^n \sim \theta_{\mathrm{PF}}^n(\boldsymbol{\gamma}\boldsymbol{h})\boldsymbol{\pi}$$

or

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

which yields $\boldsymbol{\gamma} = \boldsymbol{\pi}$. □

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

$$\lim_{n\to\infty} \mathbb{P}_i(X_n = j \mid \tau > n) = \pi_j, \qquad j = 1, \ldots, l \tag{1.19}$$

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

$$\lim_{n\to\infty} \frac{p_{ij}^{(n)}}{\sum_{k=1}^{l} p_{ik}^{(n)}} = \pi_j \tag{1.20}$$

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

The limit (1.19) is called *Yaglom limit.*

Comments.

Seneta, E. (1981) *Non-negative Matrices and Markov Chains* Second Ed. Springer-Verla, New York.

Asmussen, S. (2003) *Applied Probability and Queues* Springer,

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

.

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

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 $(\mathcal{F}_n^X)$ because

$$\{\nu^B = n\} = \{X_0 \notin B, \ldots, X_{n-1} \notin B, X_n \in B\} \in \mathcal{F}_n^X.$$

We say that a stopping time ν is *bounded* if there exists an $n_0 \in \mathbb{Z}_+$ such that $\mathbb{P}(\nu \leq n_0) = 1$.

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

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

and a *supermartingale* if

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

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* $m \leq n$

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

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

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

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

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

Analogously, (1.2) and (1.3) imply

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

and

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

for all $k, n \in \mathbb{Z}_+$. □

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

- for a martingale, $\mathbb{E}\,X_n = \mathbb{E}\,X_0$ for all $n \in \mathbb{Z}_+$,
- for a submartingale, $\mathbb{E}\,X_{n+k} \geq \mathbb{E}\,X_n$ for all $k, n \in \mathbb{Z}_+$,
- for a supermartingale, $\mathbb{E}\,X_{n+k} \leq \mathbb{E}\,X_n$ 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

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

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

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

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 $M_n = \theta^{-n} x_{Z_n}$ 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

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

from which we have $\mathbb{E}(X_{n+1} \mid Z_n) = \theta^{-n} x_{Z_n} = X_n$.
Assume now the setting of Section I.1.4 and let $\boldsymbol{P}\boldsymbol{x} = \theta\boldsymbol{x}$. Then

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

is $\mathbb{P}_i$-martingale, for $i = 1, \ldots, l$.
3. Suppose $X_1, X_2, \ldots$ are strictly positive, independent and identically distributed with $\mathbb{E}\, X = 1$. Then the sequence (M_n) given by

$$M_n = \begin{cases} 1 & \text{if } n = 0, \\ X_1 X_2 \ldots X_n & \text{if } n \geq 1 \end{cases}$$

is a martingale. Here we take filtration generated by the sequence $X_1, \ldots$. Indeed, we have

$$\begin{aligned} \mathbb{E}(M_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}(X_1 X_2 \ldots X_{n+1} \mid \mathcal{F}_n) = X_1 X_2 \ldots x_n \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \\ &= X_1 X_2 \ldots X_n \mathbb{E}\, X_{n+1} = M_n. \end{aligned}$$

Problems

1.1 Prove that (M_n) from Example 3 is a martingale.

4. 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 $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables, with the common density either f or $\tilde{f}$. The *likelihood ratio sequence* $(X_n, n \in \mathbb{Z}_+)$ is then given by

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

We show that (X_n) is an $(\mathcal{F}_n^Y)$-martingale if the Y_n have density f. Indeed,

$$\mathbb{E}\,(X_{n+1} \mid \mathcal{F}_n^Y) = \mathbb{E}\Big(\prod_{k=1}^{n+1} \frac{\tilde{f}(Y_k)}{f(Y_k)} \mid \mathcal{F}_n^Y\Big) = \prod_{k=1}^{n} \frac{\tilde{f}(Y_k)}{f(Y_k)} \mathbb{E}\Big(\frac{\tilde{f}(Y_{n+1})}{f(Y_{n+1})}\Big) = X_n\,,$$

because $\mathbb{E}\,(\tilde{f}(Y_{n+1})/f(Y_{n+1})) = \int_{-\infty}^{\infty} \tilde{f}(x)\,\mathrm{d}x = 1$. In the alternative situation that Y_n has density $\tilde{f}$, the additional assumption $\int_{-\infty}^{\infty} \tilde{f}^2(x)/f(x)\,\mathrm{d}x < \infty$ turns (X_n) into a submartingale with respect to $(\mathcal{F}_n^Y)$. Indeed, in this case

$$\mathbb{E}\Big(\frac{\tilde{f}(Y_{n+1})}{f(Y_{n+1})}\Big) = \int_{-\infty}^{\infty} \frac{\big(\tilde{f}(x)\big)^2}{f(x)}\,\mathrm{d}x = \mathbb{E}\Big(\frac{\tilde{f}(Z)}{f(Z)}\Big)^2 \geq \Big(\mathbb{E}\Big(\frac{\tilde{f}(Z)}{f(Z)}\Big)\Big)^2 = 1\,,$$

where Z is a random variable with density f.

We will use the following result on the convergence of a supermartingale.

Theorem 1.5 *A nonnegative 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 condtions 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 $c < \infty$*,*

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

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

We now show *Doob's inequality* for sub- and supermartingales.

Theorem 1.8 (a) *If* (X_n) *is a nonnegative submartingale, then*

$$\mathbb{P}\Big(\max_{0\le k\le n} X_k \ge x\Big) \le \frac{\mathbb{E}\,X_n}{x}\,, \qquad x > 0, n \in \mathbb{Z}_+\,. \tag{1.17}$$

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

$$\mathbb{P}\Big(\max_{0\le k\le n} X_k \ge x\Big) \le \frac{\mathbb{E}\,X_0}{x}\,, \qquad x > 0, n \in \mathbb{Z}_+\,. \tag{1.18}$$

Doob's inequality (1.18) 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 and N be an $\mathbb{Z}_+$-valued random variable with probability function $\{p_k\}$ which is independent of $U_1, U_2, \ldots$. Assume that, for some $0 < \theta < 1$,

$$\mathbb{P}(N > n+1 \mid N > n) \le \theta\,, \qquad n \in \mathbb{Z}_+\,, \tag{1.19}$$

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

$$\hat{m}_{F_U}(\gamma) = \theta^{-1} \tag{1.20}$$

has the solution $\gamma > 0$. We will show that

$$\mathbb{P}(\sum_{j-1}^{N} U_j > x) \le \frac{1-p_0}{\theta} e^{-\gamma x}, \qquad x \ge 0\,. \tag{1.21}$$

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

$$X_n = \begin{cases} e^{\gamma S_{n+1}} & \text{if } N > n, \\ 0 & \text{if } N \le n, \end{cases} \tag{1.22}$$

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

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

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

$$\mathbb{E}\,(Z_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\,(e^{\gamma U_{n+2}} 1(N > n+1) \mid \mathcal{F}_n) =$$

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

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 (1.18) gives

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

2 Foundation of the change of measure method

Consider the basic a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_n)_{n \ge 0}$, where $\mathcal{F} = \bigvee_{n \ge 0} \mathcal{F}_n$ is the smallest σ-field generated by all subset from $\mathcal{F}_n$ $(n \ge 0)$. We denote by $\mathbb{F}$ the pair $(\mathcal{F}, \{\mathcal{F}_n)_{n \ge 0})$. We call $(\Omega, \mathbb{F})$ a *filtered space* and $(\Omega, \mathbb{F}, \mathbb{P})$ a *filtered probability space.* For a probality $\mathbb{P}$ we denote by $\mathbb{P}_{|n}$ its restriction to $\mathcal{F}_t$, 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 $\mathcal{F}_n$ 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) 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:

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

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

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

and

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

Therefore

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

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.*

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. In this case we take $\mathcal{F}_0$ being trivual and $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$ for $n \ge 1$. We leave the reader to demonstrate the following result.

Problems

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

$$M_n = \begin{cases} 1 & n = 0 \\ \frac{\mathrm{d}\tilde{\mathbb{P}}_{|n}}{\mathrm{d}\mathbb{P}_{|n}} = \frac{\tilde{g}(X_1)}{g(X_1)}\frac{\tilde{g}(X_2)}{g(X_2)}\cdots\frac{\tilde{g}(X_n)}{g(X_n)} & n \geq 1 \end{cases}$$

which is called *likelihood ratio sequence.*

2.2 Show that $\tilde{F} \ll F$ if and only if

$$\mu\{x : g(x) = 0, \tilde{g}(x) > 0\} = 0.$$

Standard setup We now describe the so called standard setup. We take $\Omega = \mathbb{R}^d \times \mathbb{R}^d \times \ldots$, $\omega = (\omega_n)$ $X_n(\omega) = \omega_n$, $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ 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

$$\mathbb{P}_{|n}(A) = \tilde{\mathbb{E}}\,[M_n^{-1}; A], \qquad A \in \mathcal{F}_n. \tag{2.23}$$

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

$$\mathbb{P}(A) = \tilde{\mathbb{E}}\,[M_\nu^{-1}; A], \qquad A \in \mathcal{F}_\nu. \tag{2.24}$$

Proof Relation (2.23) is standard; see Appendix I.2. Relation 2.24 is proved 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

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

□

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

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

We now define for all n a probability measure on $(\Omega, \mathcal{F}_n)$ by

$$\mathrm{d}\tilde{\mathbb{P}}_{|n} = M_n \, \mathrm{d}\mathbb{P}_{|n}, \tag{2.25}$$

which means that

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

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

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

Proof

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* $\mathbb{P}_{|n}$ *for all* n.

Proof Use Lemma 2.4 and Kolmogorov consistency theorem. □

Definition 2.6 Under the standard setup scheme for $(\Omega, \mathbb{F}, \mathbb{P})$ and (X_n) we say that $\tilde{\mathbb{P}}$ is obtained by the exponential change of measure (ECM) by a positive martingale (M_n) satysfying $\mathbb{E}\, M_n = 1$, if $\tilde{\mathbb{P}}$ is constructed in the way like above.

We will write shortly *Let* $\tilde{\mathbb{P}}$ *be the probability measure obtained by ECM by the use of martingale* (M_n).

We now introduce a notion of isomorphism of two filtered probability spaces $(\Omega, \mathbb{F}, \mathbb{P})$ and $(\Omega^{'}, \mathbb{F}^{'}, \mathbb{P}^{'})$, 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 $\mathcal{F}_n^{'} = \sigma\{X_1, \ldots, X_n\}$, $\mathcal{F}_0$ is trivial and $\mathcal{F}^{'} = \bigvee_n \mathcal{F}_n^{'}$, then the considered filtered probabability space is isomporphic to the standard setting.

Corollary 2.7 *xxx*

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.]

Example 2.8 (i) Suppose that $X_i = 1$ with probability p and $X_i = 0$ with probability $1 - p$. The m.g.f. of X is

$$\hat{m}_X(s) = e^{\theta} p + e^{-s}(1 - p)$$

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

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

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 dritless, if $e^s p = e^{-s} p$, which yields

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

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 $M_n^{(s)} = \exp(sS_n/(\hat{m}_X)^n$.

Problems

2.1 Show that (X_i) is again a sequence of i.i.d. random variables with $T = U - T$, where U and T are independent random variables $U \sim$ Exp$(b-s)$ and $T \sim$ Exp$(a+s)$. Moreover $\gamma = b-a$ and $s^* = (b-a)/2$.

2.1 Harmonic functions; Doob h-transforms

We now consider a DTMC with 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 assume a standard setup for $(\Omega, \mathbb{F}, \mathbb{P}_i)$ and (X_n) where $\mathbb{P}_i$ denotes that the DTMC is defined by initial distribution $\boldsymbol{\epsilon}_i$ and transition matrix $\boldsymbol{P}$.

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

Lemma 2.10 *$\tilde{\boldsymbol{P}}$ is a transition probability matrix.*

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

Proposition 2.11 *If a DTMC (X_n) is irreducible and recurrent, then a nonnegative superharmonic functions is contant. Similar statement is true for bounded subharmonic functions.*

Proof Asmussen, p.

Problems

2.1 Prove the following result. if a chain is irreducible with a nonnzero harmonic function $\boldsymbol{h}$ such that $\boldsymbol{h} \geq \boldsymbol{0}$, then $\boldsymbol{h} > \boldsymbol{0}$.

2.2 Show that a 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 2.12 *(M_n) is a positive, mean 1 martingale.*

Proof □

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

Lemma 2.13 *The sequence of random variables* (X_n) *on* [1] $\mathbb{F}$ $(\Omega, \mathbb{F}, \tilde{\mathbb{P}}_i)$ *is a Markov chain defined by the initial distribution* $\boldsymbol{\epsilon}_i$ *and transition probability matrix* $\tilde{\boldsymbol{P}}$.

Suppose now that $\boldsymbol{P} = (p_{ij})_{i,j=1,\ldots,l}$ is issentially substochastic and primitive. Let θ_{PF} and $\boldsymbol{h}_{\mathrm{PF}}$ be the Perron Frobenius eigenvalue and the corresponding right eigenvector. We define $\boldsymbol{P}'$ as in Section I.1.4 and let (X_n) be a Markov chain with state space $\{0, 1 \ldots, l\}$ with initial state $i \in \{1, \ldots, l\}$ governed by $\boldsymbol{P}'$. Recall that $\nu = \min\{n : X_n = 0\}$. Let

$$M_n = (\theta_{\mathrm{PF}})^{-n} \frac{h_j p_{ij}}{h \quad i} 1(\nu > n)$$

In the following result $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$.

Problems

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

We can 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 $\{1, \ldots, l\}$ and with probability matrix $\tilde{\boldsymbol{P}} = (\tilde{p}_{ij})_{i,j=1,\ldots}$, where

$$\tilde{p}_{ij} =$$

Furhermore if $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_l)$ is the stationary distribution for $\boldsymbol{P}$, then $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_l)$ defined by

$$\tilde{\pi}_i =$$

is stationary for $\tilde{\boldsymbol{P}}$. The DTMC (X_n) is said to be *conditioned to stay in the subspace* $\{1, \ldots, l\}$.

[1] filtered space is natural

3 Wald martingale for random walk and its applications

We start off the important definition. Let $X_0, X_2, \ldots$ be a sequence of independent random variables. The sequence (S_n), defined as

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

is called a *random walk.*

Let $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\} = \sigma\{S_n\}$ for $n = 1, 2, \ldots$ and $\mathcal{F}_0$ is the trivial σ-field.

Problems

3.1 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).

We also quite an important result about fluctuation of the random walk.

Theorem 3.1 *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.*
(i) *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). □

Theorem 1.7 (iii) can be used to prove *Wald's identity.*

Corollary 3.2 *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* $\mathbb{E}\,|X| < \infty$. *If* ν *is a stopping time with respect to the filtration* $(\mathcal{F}_n)$ *and if* $\mathbb{E}\,\nu < \infty$, *then*

$$\mathbb{E}\,S_\nu = \mathbb{E}\,\nu \mathbb{E}\,X. \tag{3.26}$$

Proof

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

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

where s is such that the m.g.f. $\hat{F}_X(s) < \infty$. 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,...

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

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

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

We can rewrite the Wald martingale in the form

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

where

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

Problems

3.1 Show that $\kappa(s)$ is a strictly convex function such that $\kappa(0) = 0$. Demonstrate that the following scenarios are possible:

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

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

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

  $\kappa(s) < 0$ 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 $\mathbb{E}\, X \geq 0$, then by Theorem 3.1 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 of interest.

We choose $s = \gamma$ such that $\hat{F}_X(\gamma) = 1$ (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 $\mathbb{E}\, M_n^{(\gamma)} = 1$

Proposition 3.3 *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

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

according to formula (2.25) 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 $\tilde{\mathbb{P}} = \mathbb{P}^{(s)}$.

Problems

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

Let for $s \in I_F$

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

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

Lemma 3.4 *We have the following formulas for the moment generating function, mean and variance of* $F^{(s)}$

$$\begin{aligned}
\hat{F}^{(s)}(v) &= \mathbb{E}^{s} e^{vX} = \frac{\hat{F}_X(v+s)}{\hat{F}_X(s)}, \qquad v \in (I_F - s) \\
\mu^{(s)} &= \mathbb{E}^{s} X_1 = \frac{\hat{F}'_X(s)}{\hat{F}_X(s)}, \\
(\sigma^{(s)})^2 &= \mathrm{Var}^{(s)} X_1 = \frac{\hat{F}''_X(s)\hat{F}_X(s) - (\hat{F}'_X(s))^2}{(\hat{F}_X(s))^2}.
\end{aligned}$$

Proof We have

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

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

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

□

Corollary 3.5 *Let* $\mathbb{E}\, X_1 < 0$ *and suppose that there exists* $\gamma > 0$ *such that* $\hat{F}_X(\gamma) = 1$. *Function* $\mathbb{E}\,^s X_1$ *for* $0 \le s \le \gamma$ *increase continously from strictly negative to strictly positive value. In particular for* $0 < s^* < \gamma$ *fulfilling* $\hat{F}'_X(s^*) = 0$ *we have* $\mathbb{E}\,^{s^*} X_1 = 0$.

Note that under $\mathbb{P}^{s^*}$ the random walk (S_n) is driftless.

Problems

3.1 Show an example that there exists s^* such htat $\mathbb{E}\,^{s^*} X_1 = 0$ although there is no $\gamma > 0$ such that $\hat{F}_X(\gamma) = 1$.

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

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

Let $\mu = \mathbb{E}\, X_1$ 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*

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

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

Proof The relevant choice of s is fulfilling

$$\mathbb{E}^{(s)}X_1 = \frac{\hat{F}'_X(s)}{\hat{F}_X(s)} = \mu + \epsilon.$$

Let $I = s(\mu+\epsilon) - \kappa(s) > 0$. Using the basic identity (3.29) with $\nu = n$ we have

$$\begin{aligned} \mathbb{P}(S_n > (\mu+\epsilon)n) &= \mathbb{E}^{(s_0)}[e^{-s_0 S_n + n\kappa(s_0)}; S_n > (\mu+\epsilon)n] \\ &= e^{-nI}\mathbb{E}^{(s_0)}s[e^{-s_0(S_n - n(\mu+\epsilon))}; S_n > (\mu+\epsilon)n] \\ &\le e^{-nI}. \end{aligned}$$

□

Note that the event $\{\sup_{n\ge 0} S_n > u\}$ equals to the event $\{\nu(u) < \infty\}$. Moreover $\{\nu < \infty\} \in \mathcal{F}_\nu$.

We will state and start proving of the following theorem, however the final points of the proof will be possible to complete later in [**????**]. Therefore we separate the unproven here part in a form of a lemma. Define the overshoot process $B(u) = S_{\nu(u)-u}$.

Lemma 3.7 *Suppose that F_X is nonlattice. If $\hat{F}_X(s) < \infty$ for $s < \gamma + \epsilon$, for some $\epsilon > 0$, then $B(u) \xrightarrow{d} B(\infty)$.*

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

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

Proof From (3.29) we have

$$\mathbb{P}(\nu < \infty) = \mathbb{E}_\gamma[e^{-\gamma S_\nu}; \nu < \infty] \tag{3.31}$$

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

$$\mathbb{P}(\nu < \infty) = \mathbb{E}_\gamma[e^{-\gamma S_\nu}] = e^{-\gamma u}\mathbb{E}_\gamma[e^{-\gamma(S_\nu - u)}] \tag{3.32}$$

By Lemma 3.7 the overshoot process $B(u) = S_\nu(u) - u$ does converge in distribution to a limit. □

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 theory Monte Carlo theory. Suppose we aim to compute a quantity $\mathbb{P}(A(x))$ by simulation. The so called crude Monte Carlo metods says that we should make n independent experiments resulting in $Z_i(x) = Z_i = 1$ if in the i-th experiment the event holds, otherwise $Z_i = 0$ $(i = 1, 2, \dots, n)$. To avoid writing subscripts we use the notion of generic random variable that is we write Z for a generic in the i.i.d. sequence $Z_1, \dots, Z_n$.

Formally we have a sequence of independent events $A_1(x), \dots, A_n(x)$ and $Z_i(x) = 1(A_i(x))$. We say that the family of events $A(x)$ $x > 0$ is *rare* if $z(x) = \mathbb{E}\, Z(x) \to 0$ as $x \to \infty$. The *crude Monte Carlo estimator* (CMC) estimator is then

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

We may compute its variance $z(x)(1 - z(x))/n$ which converges to zero as $x \to \infty$.

In the Monte Carlo theory we admit different estimators $\hat{z}_n$ of $z(x)$. We now recall few definitions. We say that an estimator $\hat{z}_n$ is *unbiased* if $\mathbb{E}\, \hat{z}_n = z(x)$. An important notion is as follows. We say that for a given x, $\hat{z}_n$ is an (ϵ, δ)–*accurate estimator* of $A(x)$ if

$$\mathbb{P}(\mid \hat{z}_n - z(x) \mid < \epsilon z(x)) > 1 - \delta.$$

Notice that

$$\frac{\hat{z}_n - z(x)}{z(x)}$$

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

$$\mathbb{P}(\mid \hat{z}_n - z(x) \mid < \epsilon z(x)) = \mathbb{P}(\mid \frac{\sum_{i=1}^n (Z_i - \mathbb{E}\, Z_i)}{\sqrt{n}\sqrt{\mathrm{Var}\, Z_i}} \mid < \epsilon \frac{\sqrt{n} z(x)}{\sqrt{\mathrm{Var}\, Z}})$$

Since $\frac{\sum_{i=1}^n (Z_i - \mathbb{E}\, Z_i)}{\sqrt{n}\sqrt{\mathrm{Var}\, Z_i}}$ is approximatly N(0,1) normally distributed, we may write, although with some abuse, that

$$\Phi^{-1}(\frac{\sqrt{n} z(x)}{\sqrt{\mathrm{Var}\, Z_1}}) - \Phi^{-1}(-\frac{\sqrt{n} z(x)}{\sqrt{\mathrm{Var}\, Z_i}}) > 1 - \delta.$$

Hence

$$\frac{\sqrt{n} z(x)}{\sqrt{\mathrm{Var}\, Z_1}} = \Phi^{-1}(1 - \frac{\delta}{2}),$$

which yields

$$n = n(x) = \frac{\Phi^{-1}(1-\frac{\delta}{2})}{\epsilon^2} \frac{\mathrm{Var}\, Z_1(x)}{z^2(x)}.$$

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

$$\mathrm{SCV}_{Z(x)} = \frac{\mathrm{Var}\, Z(x)}{\mathbb{E}\,(Z(x))^2} = \frac{\mathrm{Var}\, Z(x)}{z^2(x)}.$$

Coming now to the crude estimator $\hat{z}_n^{\mathrm{CMC}}$ we have

$$n = n(x) = \frac{\Phi^{-1}(1-\frac{\delta}{2})}{\epsilon^2} \frac{\mathrm{Var}\, Z_1(x)}{z^2(x)} = \mathrm{const} \frac{z(x)(1-z(x))}{z^(x)}.$$

Therefore as $x \to \infty$ the number of experiments growth like $n(x) \sim \mathrm{const} 1/z(x)$.

Therefore we can postulate to look for another estimator, say $\hat{z}_n(x)$, defined by independent experiments $Y_1, \ldots, Y_n$, such that $\mathbb{E}\, Y_1 = z(x)$, where

$$\hat{z}_n(x) = \frac{\sum_{i=1}^n Y_i}{n}.$$

We may compute now the number $n = N(x)$ of experiments to assume (ϵ, δ)-accuraccy as above. Of course the ideal situation would be that $n(x)$ is bounded.

Definition 3.9 We say that the estimator $\hat{z}_n$ is an *efficient* one, if the square coefficient of variation

$$\mathrm{SCV}_{Z(x)}(\mathbb{E}\, Z(x))$$

is a bounded function.

However this frequently is difficult to fulfill. Therefore we can postulate to find estimators for which $n(x)$ growth as slow as possible. In this context we have the following definition.

Definition 3.10 We say that the estimator $\hat{z}_n$ is *weakly efficient*, if for all $\epsilon > 0$

$$\mathrm{SCV}_{Z(x)}(\mathbb{E}\, Z(x)) e^{-\epsilon x} = 0$$

Definition 3.11 We say that the estimator $\hat{z}_n$ is a *logarithmic efficient* one, if the square coefficient of variation

$$\liminf_{x\to\infty} \frac{\log \mathrm{Var}\, Z(x)}{2 \log \mathbb{E}\, Z(x)} \geq 1.$$

Problems

3.1 We say that the estimator $\hat{z}_n$ is a *polynomial time* one, $n(x) = O(|\log z(x)|^p)$, as $x \to \infty$ for some $0 < p < \infty$. We say that the estimator $\hat{z}_n$ is a *exponential time* one, if $n(x)$ is of order at least $z(x)^{-q}$ for some $q > 0$. Study the relation between logarithmic efficient and polynomial time estimators.

Our problem is 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 $n(u) \sim \mathrm{const}1/z(u)$. We now define a new estimator on the base of generic r.v.

$$Y(u) = e^{-\gamma S_{\nu(u)}},$$

that is for i.i.d. copies $Y_1, \ldots, Y_n$ of $Y(u)$ we define

$$\hat{z}_n^{\mathrm{OCM}} = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

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

Theorem 3.12 *The estimator $\hat{z}_n^{\mathrm{OCM}}$ is efficient as $n \to \infty$.*

Proof
A converse statement is also true in the following form. Note that

$$Y = \frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_1)\frac{\mathrm{d}F}{ddF_\gamma}(X_2)\ldots\frac{\mathrm{d}F}{\mathrm{d}F_\gamma}(X_\nu)1(\nu < \infty)$$

becasue, under $\mathbb{P}^\gamma$ the event $A(u)$ holds a.s. We may now define

$$Y = \frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_1)\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_2)\ldots\frac{\mathrm{d}F}{\mathrm{d}\tilde{F}}(X_\nu)1(\nu < \infty) \tag{3.33}$$

for any distribution $\tilde{F}$ equivalent to F. The underlying probability measure is dentoted then by $\tilde{\mathbb{P}}$. Again the estimator $\hat{z}_n = (Y_1 + \ldots Y_n)/n$ defined on the base of Y from (3.33) 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 p. 437.

Comments.

Asmussen, S. and Rubinstein, R.Y. (1995) *Steady-state rare events simulation in queueing theory and its complexity properties* in *Advances in Queueing* Ed. J.H. Dshalalow, CRC Press, 429–462.

Asmussen, S. *Ruin Theory*

Halmos Measure theory.

Chapter III

From random walk to Markov additive sequences

In this chapter we will study a more detailed the theory of random walk, which was introduced in Section II.3. We will also introduce a generalization of the random walk to a Markov additive sequence.

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 $T_0, T_0 + T_1, T_1 + T_2, \ldots$ 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* $n = 1, 2, \ldots$

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

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

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

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 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 $\mathcal{M} = \max(0, S_1, \ldots)$

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

Lemma 1.2 *If $X_1, X_2, \ldots$ 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*

$$W_n \stackrel{\mathrm{d}}{\to} \mathcal{M}.$$

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 $F_X(x)$ 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

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

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

- if T is exponentially distributed Exp(a) and U is exponentially distributed Exp(b), then the system is denoted by M/M/1 FCFS;
- if If T is exponentially distributed 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 Exp(b), then the system is denoted by GI/M/1 FCFS.

2 Ladder epochs and heights

2.1 Ladder Epochs

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

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

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

$$\nu^- = \min\{n > 0 : \ S_n \leq 0\}\,, \tag{2.3}$$

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

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

and

$$\{\nu^- = k\} = \{S_1 > 0, S_2 > 0, \ldots, S_{k-1} > 0, S_k \le 0\} \tag{2.5}$$

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.3.2 proved there we have that, for each stopping time ν with respect to $\{\mathcal{F}_n\}$,

$$\mathbb{E}\, S_\nu = \mathbb{E}\, \nu \mathbb{E}\, X \tag{2.6}$$

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

$$\nu_{n+1}^+ = \min\{j > \nu_n^+ : S_j > S_{\nu_n^+}\}, \tag{2.7}$$

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

$$\nu_{n+1}^- = \min\{j > \nu_n^-, S_j \le S_{\nu_n^-}\}, \qquad n = 1, 2, \ldots. \tag{2.8}$$

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 *(stron) ascending ladder height.*

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

Problems

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

The *proof* of this theorem is easy and is left to the reader.

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 (2.7). 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

$$\mathcal{M} = \sum_{i=1}^{N} X_i^+, \tag{2.9}$$

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* $x \ge 0$ *and for* $p = G^+(\infty)$

$$\mathbb{P}(\mathcal{M} \le x) = (1-p)\sum_{k=0}^{\infty}(G^+)^{*k}(x) = \sum_{k=0}^{\infty}(1-p)p^k G_0^{*k}(x)\,. \tag{2.10}$$

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 (2.9) we have

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

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 II.3.1 that all descending ladder epochs and heights are proper random variables.

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

3 The Wiener–Hopf Factorization

3.1 General Representation Formulae

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

$$G^-(x) = \mathbb{P}(X^- \le x), \qquad x \in \mathbb{R}. \tag{3.11}$$

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

$$G^+(x) = \mathbb{P}(X^+ \le x), \qquad x \in \mathbb{R}. \tag{3.12}$$

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

$$H_0^-(B) = \sum_{k=0}^{\infty}(G^-)^{*k}(B), \qquad B \in \mathcal{B}(\mathbb{R}_-). \tag{3.13}$$

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

$$H_0^+(B) = \sum_{k=0}^{\infty} (G^+)^{*k}(B)\,, \qquad B \in \mathcal{B}(\mathbb{R}_+)\,. \tag{3.14}$$

Moreover, from (3.13) it follows that

$$H_0^- * G^- = H_0^- - \delta_0\,. \tag{3.15}$$

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 3.1 *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

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

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 $X_1, X_2, \dots$ 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 factorization* of F_X.

Theorem 3.2 *The following relationship holds:*

$$F = G^+ + G^- - G^- * G^+. \tag{3.16}$$

Proof We first show that

$$\delta_0 + \gamma^- * F = \gamma^- + G^+. \tag{3.17}$$

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

$$\sum_{n=0}^{\nu^+-1} 1(S_n \in B) + 1(S_{\nu^+} \in B) = 1(0 \in B) + \sum_{n=0}^{\nu^+-1} 1(S_{n+1} \in B)\,. \tag{3.18}$$

Now $\mathbb{E}\big(\sum_{n=0}^{\nu^+-1} 1(S_{n+1} \in B)\big) = \mathbb{E}\big(\sum_{n=0}^{\infty} 1(\nu^+ > n, S_{n+1} \in B)\big)$ 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 (3.18), we get (3.17). Convoluting both sides of (3.17) with G^- we obtain

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

On the other hand, by (3.15) and Lemma 3.1, 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 (3.17) again, this gives (3.16). □

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 factorization (3.16) yields the following representation formula for G^+.

Corollary 3.3 *For $B \in \mathcal{B}((0,\infty))$,*

$$G^+(B) = F * H_0^-(B) = \int_{-\infty}^{0} F(B-y)\,\mathrm{d}H_0^-(y)\,, \tag{3.19}$$

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

$$G^-(B) = F * H_0^+(B) = \int_{0}^{\infty} F(B-y)\,\mathrm{d}H_0^+(y)\,. \tag{3.20}$$

Proof Convoluting both sides of the Wiener–Hopf factorization (3.16) 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 (3.20) is similar. □

3.2 An analitical form of Wiener-Hopf facrorization

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

Corollary 3.4 (a) *For $s \le 0$,*

$$\hat{m}_{\mathcal{M}}(s) = \frac{1 - G^+(\infty)}{1 - \hat{m}_{G^+}(s)}\,. \tag{3.21}$$

The *proof* is obvious because (3.21) directly follow from (2.10). □

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 $z = it$ for some real t and then $\hat{F}(z) = \int_{-\infty}^{\infty} \mathrm{e}^{itx}\,\mathrm{d}F(x)$, which is the characteristic function of F. An immediate consequence of the Wiener–Hopf factorization (3.16) is the following *analytical factorization* of the corresponding moment generating functions.

Corollary 3.5 *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*

$$1 - \hat{F}(z) = (1 - \hat{G^+}(z))(1 - \hat{G^-}(z))\,. \tag{3.22}$$

We conclude this section with a few comments on a probabilistic solution of the Wiener–Hopf factorization (3.22). Suppose that for some $\varepsilon > 0$ the function $\hat{g}_F(z)$ is well-defined in $1-\varepsilon \le \Re z \le 1+\varepsilon$. Define the functions

$$d^+(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n > 0]}{n}\Big) \tag{3.23}$$

and

$$d^-(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n \le 0]}{n}\Big). \tag{3.24}$$

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

$$1 - \hat{F}(z) = d^+(z)d^-(z)\,, \qquad 1-\varepsilon \le |z| \le 1+\varepsilon\,. \tag{3.25}$$

Hence,

$$1 - \hat{G^+}(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n > 0]}{n}\Big) \tag{3.26}$$

and

$$1 - \hat{G^-}(z) = \exp\Big(-\sum_{n=1}^{\infty} \frac{\mathbb{E}\,[e^{zS_n}; S_n \le 0]}{n}\Big)\,, \tag{3.27}$$

since the factorization given in (3.25) is unique within the class of functions satisfying the same conditions as mentioned above for $d^+(z)$ and $d^-(z)$. See e.g. Prabhu [**?**], p. 50.

Problems

3.1 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 $\gamma^+ = H_0^+$.

3.2 Assume that $\hat{m}_X(s) < \infty$ for some $s > 0$. Show that then $\hat{m}_{G^+}(s) < \infty$. [Hint. Use Corollary 3.3 and (3.13) showing first that

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

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

3.4 Assume that $F_T = \mathrm{Exp}(\lambda)$ and $F_U = \sum_{k=1}^{n} p_k\,\mathrm{Exp}(\delta_k)$ for some $n = 2,3,\ldots$ and some probability function $\{p_1,\ldots,p_n\}$. Show that

$$G_0 = q_n \sum_{k=1}^{n} \frac{p_k}{\delta_k}\,\mathrm{Exp}(\delta_k),$$

where $q_n = \prod_{i=1}^{n}\delta_i(\sum_{j=1}^{n}\delta_j)^{-1}$.

3.5 Let $F_T = \Gamma(2,2\lambda)$ and $F_U = \mathrm{Exp}(\delta)$, where $\rho = \lambda(\delta\beta)^{-1} < 1$. Show that then $p = \frac{1}{2}(1+\rho-\sqrt{1+2\rho})$. [Hint. Show first that (6.4.37) leads to a cubic equation with root $s = 1$.]

3.6 Assume that $F_T = \sum_{k=1}^{2} p_k\,\mathrm{Exp}(\lambda_k)$ and $F_U = \mathrm{Exp}(\delta)$ for some probability function $\{p_1,p_2\}$ such that $\delta\beta(p_1\lambda_1^{-1} + p_2\lambda_2^{-1}) > 1$. Show that $\delta\beta(1-p)$ is a solution to the quadratic equation

$$x^2 + (\lambda_1+\lambda_2-\delta\beta)x + \lambda_1\lambda_2 - \delta\beta(\lambda_1(1-p_1)+\lambda_2(1-p_2)) = 0.$$

[Hint. Proceed in the same way as in Exercise 3.5.]

4 Two important identities

Theorem 4.1 *Suppose that* $|r| < 1$ *and* $\Re(z) = 0$. *Then*

$$1 - \mathbb{E}\left[r^{\nu^+} e^{zX^+}; \nu^+ < \infty\right] = \exp\left(-\sum_{n=1}^{\infty}\frac{r^n}{n}\mathbb{E}\left[e^{zS_n}; S_n > 0\right]\right). \qquad (4.28)$$

Corollary 4.2

$$1 - \mathbb{E}\left[r^{\nu^+}; \nu^+ < \infty\right] = \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{P}(S_n > 0)\right) \tag{4.29}$$

and

$$G^+(\infty) = 1 - \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}(S_n > 0)\right)$$

Problems

4.1 Show that $G^+(\infty) = 1$ if and only if $\sum n = 1^\infty (1/n)\mathbb{P}(S_n > 0) = \infty$.

Corollary 4.3 *Suppose that $\nu^+ < \infty$ a.s. and $|r| < 1$ Then*

$$\sum_{n=1}^{\infty} r^n \mathbb{P}(\nu^+ > n) = \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}(S_n \le 0)\right)$$

Hence $\mathbb{E}\,\nu^+ < \infty$ if and only if $\sum n = 1^\infty (1/n)\mathbb{P}(S_n \le 0) < \infty$.

We now state the *Spitzer identity* For $|r| < 1$ and $\Re(z) \le 0$

$$\sum_{n=0}^{\infty} \mathbb{E}\, e^{z\mathcal{M}_n} = \exp\left(\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[e^{zS_n}; S_n > 0\right]\right).$$

Moreover if $\mathcal{M} < \infty$ a.s.

$$\mathbb{E}\, e^{z\mathcal{M}} = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n}\left(\mathbb{E}\left[e^{zS_n}; S_n > 0\right] - 1\right)\right).$$

5 Markov additive sequence

We suggest the reader to prove that the following sentence is an equivalent defienition of the random walk (S_n).

Problems

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

$$(S_{n+m} - S_n)_{m=0,1,\ldots} | \mathcal{F}_n =_{\mathrm{d}} (S_m)_{m=0,1,\ldots}, \tag{5.30}$$

then (S_n) is a random walk.

5.2 Consider a random walk $(S_n)_{m=0,1,\ldots}$ on $(\Omega, \mathbb{F}, \mathbb{P})$. We want to consider this random walk begining 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 distibution to $(S_m)_{m=0,1,\ldots}$ under $\mathbb{P}_{S_n}$.

In this section we consider a generalization of the random walk we begin with a simpe but motivating example. 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

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

Suppose now there is given a DTMC (J_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 $J_{n-1} = i$, $J_n = j$, $S_n = k + X_n$ where $\mathbb{P}(X_n = 1|J_{n-1} = i, J_n = j) = p^{ij}$ $\mathbb{P}(X_n = -1|J_{n-1} = i, J_n = j) = q^{ij}$, and $p^{ij} + q^{ij} = 1$. The bivariate sequence (S_n, J_n) is a DTMC with state space $E' = \mathbb{Z} \times E$ with transition probability matrix

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

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 this 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), 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 $X_1, \ldots, X_n$ is independent. 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 $\mathbb{P}_{x,i}$. Let $\mathcal{F}_n = \sigma\{S_0, J_0, \ldots, S_n, J_n\}$ be the natural filtration. We have then

$$
\mathbb{E}_{x,i}[f(S_{n+m})g(J_{n+m}) \mid \mathcal{F}_n] = \mathbb{E}_{J_n}[f(S_m)g(J_m)], \mathbb{P}_{x,i} - \text{a.s.} \tag{5.31}
$$

Some special cases are:

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

In the following we will consider Markov arrivals sequences with $X_n \in \mathbb{R}$. Specifying $f(x) = 1(x \in B)$ and $g(y) = 1(y = j)$ and $m = 1$ in the RHS of (5.31) we have subdistributions $F_{ij}(B)$, where

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

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

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

For an abbreviated form of notations, we introduce matrices:

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

Note that $F_{ij}(B) \le 1$. Sometimes it is useful to use equivalently $\boldsymbol{P} = (p_{ij}) = \boldsymbol{F}(\mathbb{R})$ and

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

Problems

5.1 Show that

$$\begin{aligned} &\mathbb{P}_{x,i}(X_1 \in \mathrm{d}x_1, J_1 = j_1, \ldots, X_n \in \mathrm{d}x_n, J_n = j_n) \\ &\quad = F_{jj_1}(\mathrm{d}x_1) F_{j_1 j_2}(\mathrm{d}x_2) \ldots F_{j_{n-1} j_n}(\mathrm{d}x_n). \end{aligned}$$

5.2 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 5.1 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

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

Example 5.2 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 customer with service requirement of size x. 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 $\mathrm{Exp}(b)$ random variables and an i.i.d. sequence (η_i) of $\mathrm{Exp}(a)$ random variables. On the base of these two sequences we define a bivariate sequences (X_n, J_n) as follows: if $\xi_n < \eta_n$, then

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

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

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

Let $S_n = X_1 + \ldots + X_n$ and $S_0 = 0$. Let $Z_n = \psi_1 + \ldots + \psi_n$ and $Z_0 = 0$. We may consider a process $S(t)$, which is a linearization between points (S_n, Z_n). Let $\nu(x) = \inf\{t > 0 : S(t) > x\}$. Notice that $\nu(x)$ has the same distribution as the response time of a tagged customer with service requirement x. We may consider $\nu^+(x) = \min\{n : S_n > x\}$.

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.) $\hat{\boldsymbol{F}}(s) = (\hat{F}_{ij}(s))_{i,j \in E}$. 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 $F_{ij}(\theta)$ $(i, j \in E)$ 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 $\theta_{\rm PF}(s)$ of $\hat{\boldsymbol{F}}(s)$ is positive and therefor we may define $\kappa(s) = \log\theta_{\rm PF}(s)$. Let $\boldsymbol{\nu}^{(s)}$ and $\boldsymbol{h}^{(s)}$ be the corresponding left and right eigenvectors respectively. We now state a sequence of auxiliary lemmas.

Lemma 5.3 *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 5.4 *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 5.5 *Class of logconvex functions on I is closed under the operation of addition, multiplication and passing to the limit. Furthermore $\hat{F}(s)$ is a logconvex function.*

Proof [1] It is straighforward that the product of two logconvex functions is logcovex. 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

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

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

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

$$(f(\frac{s_1+s_2}{2})+g(\frac{s_1+s_2}{2}))^2$$

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

$$\le \; f(s_1)f(s_2)+g(s_1)g(s_2)+2\frac{f(s_1)g(s_2)+f(s_2}{2}$$

$$(f(s_1)+g(s_1))(f(s_2)+g(s_2)).$$

This is equivalent to $f+g$ being logconvex. □

Lemma 5.6 *[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

$$\theta^n_{\mathrm{PF}}(s)+\theta^n_2(s)+\ldots+\theta^n_l(s)=\mathrm{tr}(\hat{\boldsymbol{F}}^n(s)).$$

However, by Lemma 5.5, $\mathrm{tr}(\hat{\boldsymbol{A}}^n)$ is logconvex. Hence $(\theta^n_{\mathrm{PF}}(s)+\theta^n_2(s)+\ldots+\theta^n_l(s))^{1/n}$ is logconvex too. Now passing to the limit, by Perron-Frobenius theorem $(\theta^n_{\mathrm{PF}}(s)+\theta^n_2(s)+\ldots+\theta^n_l(s))^{1/n}\to\theta_{\mathrm{PF}}(s)$. Hence $\theta_{\mathrm{PF}}(s)$ is logconvex and the proof is copleted. □

Proposition 5.7 *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.

We say that a Markov additive sequence $(S_n,J_n)_n$ is degenerated if $\sup|S_n|<\infty$ a.s.

Proposition 5.8 *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.*

5.1 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. 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 5.9 *For $s \in \mathcal{I}_{\hat{\boldsymbol{F}}}$ the sequence $(M_n^{(s)})$ is a $\mathbb{P}_{(x,i)}$-martingale.*

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

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

Furthermore

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

and

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

6 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

Problems

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

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

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 6.1 *[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

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

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

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

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

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

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 6.2
(i) $\kappa^* : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *is a convex, lower semicontiuous function. Furthermore* κ^* *is strictly convex and* C^∞ *in the interior of the set* $\{\kappa'(s),\ s \in \text{int}(\mathcal{I}_\kappa)\}$.
(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 deccreasing function.
(iii) $\kappa^*(\mu) = 0$.

For the later use we also need the following result.

Corollary 6.3 *Suppose that* $0 \in \text{int}(\mathcal{I}_\kappa)$. *For any* $b \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n([b,\infty)) = -\inf_{x \geq b} \kappa^*(x).$$

Proof Dembo and Zeitoni p. 34.

Problems

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

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

6.2 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 6.4 *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, 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 [2] is called a *rate function.* Among rate functions we distinguish the so called *good*, that is such that that each lavel set $\{x : \ \kappa^*(x) \le \alpha\}$ 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

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

foror each open set B

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

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

Theorem 6.5 *[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*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\,[e^{nf(\hat{Z}_n)}; f(\hat{Z}_n) \ge M] = -\infty. \tag{6.32}$$

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

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

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

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

[2] that is suchn that each lavel set $\{x : I(x) \le \alpha\}$ is closed

Proof Dembo and Zeitoni Section 4.3.

Lemma 6.6 *If for some $c > 1$*

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

then (6.32) holds.

7 Large deviations and efficient simulation

We call a sequence of events $(A_n)_n$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *rare* if $\lim_{n\to\infty} p_n = 0$, where $p_n = \mathbb{P}(A_n)$ We will suppose that

Hyphotesis 1.

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

for some $\gamma > 0$.

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

$$\hat{p}_n = \frac{1}{k} \sum_{j=1}^{k} \xi_j(n),$$

where $\xi_1(n), \xi_2(n), \ldots, \xi_k(n)$ are i.i.d. copies of the generic random variable $\xi(n)$ such that $\mathbb{E}\,\xi(n) = p_n$. The asymptotic quality is measured by the behaviour at ∞ of

$$\mathrm{SCV}_{\xi(n)} = \frac{\mathrm{Var}\,(\xi(n))}{(\mathbb{E}\,\xi(n))^2} = \frac{\mathrm{Var}\,(\xi(n))}{p_n^2}$$

see Section II.3.1 for definitions. We will assume that

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

Problems

7.1 The estimator is weakly efficient if and only if for all $\epsilon > 0$

$$\limsup_{n\to\infty} \frac{\mathrm{Var}\,(\xi(n))}{(\mathbb{E}\,\xi(n))^{2-\epsilon}} = 0$$

or under Hyphotesis 2

$$\limsup_{n\to\infty} \frac{\mathbb{E}\,(\xi^2(n))}{(\mathbb{E}\,\xi(n))^{2-\epsilon}} = 0$$

7.2 Under Hyphotesis 1 and 2, the estimator is logarithmic efficient if and only if

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

7.3 Under Hyphotesis 1 and 2, the estimator is logarithmic efficient if and only if

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

From the result of Exercise 7.3 we see that

$$\liminf_{n\to\infty}(-\frac{1}{n})\log\mathbb{E}\,\xi^2(n) \tag{7.33}$$

is an indicator of the quality of simulation.

Lemma 7.1 *If the estimator is logarithmic efficient, then it is weakly efficient.*

Proof By the result of Exercise 7.3 we have

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

Hence

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

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

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

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

The crude MC method uses the estimator

$$\hat{p}_n^{\mathrm{CMC}} = \frac{1}{k} \sum_{j-1}^{k} \xi_k(n),$$

where $\xi_1(n), \xi_2(n), \ldots$ are i.i.d. copies of the generic random variable $\xi(n) = 1(A_n)$. Note that

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

Consider now importance sampling estimator $\hat{p}_n^{(s)}$ defined by generic

$$\xi(n) = e^{-sS_n + n\kappa(s)} 1(A_n) = e^{nf(\hat{S}_n)} 1(\hat{S}_n \in B), \tag{7.34}$$

where $s \in \mathbb{R}$ and

$$f(x) = \kappa(s) - sx, \tag{7.35}$$

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 II.3. Note that

$$p_n = \mathbb{E}^{(s)} \xi(n).$$

To find quality of $\hat{p}_n^{(s)}$, following the result of Exercise 7.3, we have compute

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

In particular for logarithmic efficiency we need that

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

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

Lemma 7.2

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

Proof From the definition

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

□

Lemma 7.3

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

Proof Recalling definition f from (7.35) we write

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

We now check whether the assumption of Lemma 6.6 holds for $c>1$:

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

Now we use Theorem 6.32 (ii). □

We obtain the possibly optimal change of measure by the use of importance sampling estimator from (7.34) if we find s_0 such that

$$\begin{aligned}&\inf_{x\in B}\{\kappa^*(x)-\kappa(s_0)+s_0x\}\\&\quad=\max_{s\in\mathbb{R}}\inf_{x\in B}\{\kappa^*(x)-\kappa(s)+sx\} \qquad (7.36)\\&\quad\leq\inf_{x\in B}\max_{s\in\mathbb{R}}\{\kappa^*(x)-\kappa(s)+sx\}. \qquad (7.37)\end{aligned}$$

If there is the equality in (7.37), that is the minimax hyphotesis is fulfilled, then

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

and then we would have the logarithmic efficiency.

Proposition 7.4 *If $B = [b, \infty)$ for $b > \mu$ and $\kappa^*(b+) < \infty$, then there exists s_0 such that*

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

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

Example 7.5 We consider a random walk with generic increment r.v. X normally distributed $\mathcal{N}(1,1)$. Suppose that $B = [a, \infty)$, where $1 < a$ Then $\kappa(s) = s + s^2/2$, $\kappa^*(x) = (x-1)^2/2$ and hence

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

We have

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

Standard calculations show that the supremu over $s > 0$ of

$$\inf_{x > a} k(s,x) = (b-1)^2,$$

is achieved at $s = b - 1$ and that $\inf_{x>a} k(b-1, x) = (b-1)^2$. Hence we conclude that

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

which yields the logarithmic efficiency. Using Lemma II.3.4 we find that X under $\mathbb{P}^{(b-1)}$ is $\mathcal{N}(b,1)$. We now continue the with $B = (-\infty, -b] \cup [b, \infty)$, $1 < b$. Then

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

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.)

8 Gärtner-Ellis theorem

Comments.

S. Asmussen, S. (2003) *Applied Probability and Queues* Springer, New York.

S. Asmussen (1999) *Stochastic simulation with a view towards stochastic processes.* MaPhySto, University of Aarhus.

A. Dembo and O. Zeitouni (1993) *Large Deviations Techniques and Applications.* Jones and Bartlett Publishers, Boston.

P. Dupuis and H. Wang (2003) Importance sampling, large deviation, and differentional games. MS.

J.A. Bucklew (1990) *Large Deviation Techniques in Decision, Simulation, and Estimation* Wiley, New York.

J.S. Sadowsky (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.

J.F. Collamore (2002) Importance sampling techniques *Ann. Appl. Prob.* 12 382–421.

Appendix A

1 Moment generating function and other transforms

For a random variable X with distribution F, let $I_F = I_X = \{s \in \mathbb{R} : \mathbb{E}\,\mathrm{e}^{sX} < \infty\}$. Note that I_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 $s \le 0$. There are alos 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 $s_F^- = \inf\{s \le 0 : \ \hat{F}(s) < \infty\}$ and $s_F^+ = \sup\{s \ge 0 : \ \hat{F}(s) < \infty\}$ be the lower and upper *abscissa of convergence*, respectively, of the moment generating function $\hat{F}(s)$. Clearly $s_F^- \le 0 \le s_F^+$. 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. □

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 $\mathbb{P} \ll \tilde{\mathbb{P}}$. 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* $M = \frac{d\tilde{\mathbb{P}}}{\mathbb{P}} > 0$ $\mathbb{P}$*-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.

Index