Available online at www.sciencedirect.com

SCIENCE DIRECT®

Stochastic Processes and their Applications 116 (2006) 724–756

stochastic processes and their applications

www.elsevier.com/locate/spa

Worst-case large-deviation asymptotics with application to queueing and information theory☆

Charuhas Pandit[a], Sean Meyn[b,*]

[a] *Morgan Stanley and Co., 1585 Broadway, New York, NY 10019, United States*
[b] *Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 1308 W. Main Street, Urbana, IL 61801, United States*

Received 28 September 2004; received in revised form 19 October 2005; accepted 11 November 2005
Available online 12 December 2005

Abstract

An i.i.d. process $\boldsymbol{X}$ is considered on a compact metric space X. Its marginal distribution π is unknown, but is assumed to lie in a moment class of the form,

$$\mathbb{P} = \{\pi : \langle \pi, f_i \rangle = c_i, i = 1, \ldots, n\},$$

where $\{f_i\}$ are real-valued, continuous functions on X, and $\{c_i\}$ are constants. The following conclusions are obtained:

(i) For any probability distribution μ on X, Sanov's rate-function for the empirical distributions of $\boldsymbol{X}$ is equal to the Kullback–Leibler divergence $D(\mu \parallel \pi)$. The worst-case rate-function is identified as

$$L(\mu) := \inf_{\pi \in \mathbb{P}} D(\mu \parallel \pi) = \sup_{\lambda \in R(f,c)} \langle \mu, \log(\lambda^T f) \rangle,$$

where $f = (1, f_1, \ldots, f_n)^T$, and $R(f, c) \subset \mathbb{R}^{n+1}$ is a compact, convex set.

(ii) A stochastic approximation algorithm for computing L is introduced based on samples of the process $\boldsymbol{X}$.

(iii) A solution to the worst-case one-dimensional large-deviation problem is obtained through properties of *extremal distributions*, generalizing Markov's canonical distributions.

☆ This paper is based upon work supported by the National Science Foundation under Award Nos. ECS 02 17836 and ITR 00-85929. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

* Corresponding author. Tel.: +1 217 244 1782; fax: +1 217 244 1653.
E-mail addresses: charuhas.pandit@morganstanley.com (C. Pandit), meyn@uiuc.edu (S. Meyn).
URL: http://decision.csl.uiuc.edu:80/~meyn (S. Meyn).

doi:10.1016/j.spa.2005.11.003

(iv) Applications to robust hypothesis testing and to the theory of buffer overflows in queues are also developed.

MSC: 60F10; 94A17; 62F15; 37M05; 68M20

Keywords: Large deviations; Entropy; Bayesian inference; Simulation; Queueing

1. Introduction and background

Consider an i.i.d. sequence $\boldsymbol{X}$ on a compact metric space X. It is assumed that its marginal distribution π is not known exactly, but belongs to the *moment class* $\mathbb{P}$ defined as follows: A finite set of real-valued continuous functions $\{f_i : i = 1, \ldots, n\}$ and real constants $\{c_i : i = 1, \ldots, n\}$ are given, and

$$\mathbb{P} := \{\pi \in \mathcal{M}_1 : \langle \pi, f_i \rangle = c_i, i = 1, \ldots, n\}, \tag{1}$$

where $\mathcal{M}_1$ is the space of probability distributions on X, and the notation $\langle \pi, f_i \rangle$ is used to denote the mean of the function f_i according to the distribution π.

The motivation for consideration of moment classes comes primarily from the simple observation that the most common approach to partial statistical modeling is through moments, typically mean and correlation. Moment classes have been considered in applications to finance [41]; admission control [10,5,32]; queueing theory [20,19,12]; and other applications.

For a moment class of this form, and a given function $g \in C(\mathsf{X})$, the map $\pi \to \langle \pi, g \rangle$ defines a continuous linear functional on $\mathcal{M}_1$. Consequently, the following maximization may be viewed as a linear program,

$$\max_{\pi \in \mathbb{P}} \langle \pi, g \rangle. \tag{2}$$

The value of this (infinite-dimensional) linear program provides a bound on the mean of g that is uniform over $\pi \in \mathbb{P}$. A.A. Markov, a student of Chebyshev, considered a special case of the linear programs (2) in which the functions $\{f_i\}$ are polynomials. A comprehensive survey by M.G. Kreĭn in 1959 describes many of Markov's original results [26]. Since then, these ideas have been developed in various directions [1,13,44,29,21,8,38,34,36,3,42].

The present paper concerns various large-deviation bounds that are uniform across a moment class. One set of results concerns relaxations of *Chernoff's bound*: For a given function $h \in C(\mathsf{X})$, and any $r \geq \langle \pi, h \rangle$,

$$\mathsf{P}\{S_N \geq r\} \leq \exp(-N I_{\pi,h}(r)), \quad N \geq 1, \tag{3}$$

where $\{S_N = N^{-1} \sum_{j=1}^{N} h(X_j) : N \geq 1\}$, and $I_{\pi,h}$ is the usual one-dimensional large-deviation rate-function under the distribution π. Denoting the log moment-generating function as,

$$M_{\pi,h}(\theta) := \log \langle \pi, \exp(\theta h) \rangle, \quad \theta \in \mathbb{R}, \tag{4}$$

the rate-function is equal to the convex dual,

$$I_{\pi,h}(r) = \sup_{\theta \in \mathbb{R}} \{\theta r - M_{\pi,h}(\theta)\}, \quad r \in \mathbb{R}. \tag{5}$$

Theorem 1.5 and related results in Section 3 contain expressions for the minimum of the rate function $I_{\pi,h}(r)$ over the moment class $\mathbb{P}$. To put these results in context we present some known results in the special case of polynomial constraint functions.