



**HAL**  
open science

# Large deviations for estimators of unknown probabilities, with applications in risk theory

Claudio Macci

► **To cite this version:**

Claudio Macci. Large deviations for estimators of unknown probabilities, with applications in risk theory. *Statistics and Probability Letters*, 2010, 81 (1), pp.16. 10.1016/j.spl.2010.09.013 . hal-00698869

**HAL Id: hal-00698869**

**<https://hal.science/hal-00698869>**

Submitted on 18 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Large deviations for estimators of unknown probabilities, with applications in risk theory

Claudio Macci

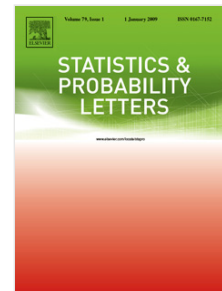
PII: S0167-7152(10)00263-4  
DOI: [10.1016/j.spl.2010.09.013](https://doi.org/10.1016/j.spl.2010.09.013)  
Reference: STAPRO 5799

To appear in: *Statistics and Probability Letters*

Received date: 12 July 2010  
Revised date: 16 September 2010  
Accepted date: 16 September 2010

Please cite this article as: Macci, C., Large deviations for estimators of unknown probabilities, with applications in risk theory. *Statistics and Probability Letters* (2010), doi:10.1016/j.spl.2010.09.013

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Large deviations for estimators of unknown probabilities, with applications in risk theory

Claudio Macci\*

## Abstract

We present large deviation results for estimators of unknown probabilities which satisfy a suitable exponential decay condition. These results provide some extensions of the large deviation estimates in Macci and Petrella (2006). Furthermore we propose a classical approach which is different from the one presented in Ganesh et al. (1998) and we cannot say that the Bayesian approach is more conservative as in that paper.

*Key words:* Large deviations, Varadhan's Lemma, level crossing probability, compound Poisson process, Brownian motion.

## 1 Introduction

In this paper we prove asymptotic results for estimators of unknown probabilities and, motivated by potential applications in risk theory, we present several examples where these probabilities concern the occurrence of level crossing events.

The unknown probabilities are  $(\rho_n(\underline{q}|\theta))$ , where  $\underline{q} = (q_1, \dots, q_d) \in (0, \infty)^d$  for some  $d \geq 1$  and the unknown parameter  $\theta$  belongs to some parameter space  $\Theta$ . One may wish to assess the unknown probabilities based on experience. Several estimation methods separate the point estimation of the unknown parameter  $\theta$  and the estimation of the unknown probabilities but, as pointed out in Ganesh et al. (1998) for the gambler's ruin probability, these methods can lead to a very misleading inference. In this paper we propose two alternative approaches which provide a wide full statistical analysis for the unknown probabilities, capturing all the uncertainties on the unknown parameter  $\theta$ : the Bayesian approach proposed in Ganesh et al. (1998) (see also Ganesh and O'Connell, 1999 and 2000) based on estimates of the *predictive probabilities*  $\int_{\Theta} \rho_n(\underline{q}|\theta) \pi_n(d\theta)$ , where  $\pi_n$  is the posterior distribution given the  $n$ -sample; the classical approach based on estimates of the expected values  $\mathbb{E}_{\theta}[\rho_n(\underline{q}|\hat{\Theta}_n)]$ , where  $\hat{\Theta}_n$  is an empirical estimator based on the  $n$ -sample. The latter is different from the *frequentist* approach presented in Ganesh et al. (1998) and here we cannot say that the Bayesian approach is more conservative as in that paper. Typically  $(\hat{\Theta}_n)$  will be the sequence of the maximum likelihood estimators (MLEs from now on).

In this paper we refer to theory of large deviations which gives an asymptotic computation of small probabilities on exponential scale; see e.g. Dembo and Zeitouni (1993) as a reference on this topic. We provide some extensions of the results in Macci and Petrella (2006): the large deviation principles for posterior distributions are proved without considering finite mixtures of conjugate prior distributions; we consider a further statistical model based on underlying Inverse Gaussian distributed samples concerning independent replications of Brownian motions; we consider level crossing probabilities concerning some bivariate risk models, i.e. the model in Avram et al. (2008) and a version of it based on the Brownian motion.

---

\*Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, I-00133 Rome, Italy.  
e-mail: macci@mat.uniroma2.it

We also remark that we present a unique result (we mean Proposition 3.2 for the classical approach and Proposition 3.3 for the Bayesian approach) which works well for all the level crossing probabilities presented in this paper. This unique result is formulated in terms of sequences of probabilities  $((\rho_n(q|\theta)) : \theta \in \Theta)$  which satisfy the *exponential decay condition* in Definition 3.1 below; this condition allows to obtain large deviation estimates with standard applications of the Varadhan's Lemma.

We conclude with the outline of the paper. Section 2 is devoted to recall some preliminaries. In section 3 we present large deviation results for estimators of unknown probabilities with exponential decay; moreover we discuss the comparison between classical and Bayesian approaches. The applications in risk theory are presented in section 4 (standard risk models) and in section 5 (bivariate risk models).

## 2 Preliminaries

We start with some preliminaries on large deviations and we refer to Dembo and Zeitouni (1993). A rate function is a lower semi-continuous function  $I : \Omega \rightarrow [0, \infty]$ , where  $\Omega$  is a Hausdorff topological space with Borel  $\sigma$ -algebra  $\mathcal{B}_\Omega$ . A sequence of probability measures  $(\nu_n)$  on  $(\Omega, \mathcal{B}_\Omega)$  satisfies the large deviation principle (LDP from now on) with rate function  $I$  if

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \nu_n(F) \leq - \inf_{\omega \in F} I(\omega) \quad \text{for all closed sets } F \subset \Omega$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \nu_n(G) \geq - \inf_{\omega \in G} I(\omega) \quad \text{for all open sets } G \subset \Omega.$$

A rate function  $I$  is *good* if all the level sets  $\{\{\omega \in \Omega : I(\omega) \leq \eta\} : \eta \geq 0\}$  are compact. A sequence of  $\Omega$  valued random variables  $(Z_n)$  satisfies the LDP if  $(\nu_n)$  does with  $\nu_n = P(Z_n \in \cdot)$ .

We refer to the definition for random variables when we consider the classical approach, where  $(Z_n)$  is a sequence of estimators. We refer to the definition for probability measures when we consider the Bayesian approach, where  $(\nu_n)$  is the sequence of posterior distributions; in this case we have a family of random probability measures which satisfies the LDP almost surely. In some sense we should say that we have a *conditional* LDP, indeed this LDP holds assuming that a sequence of sufficient statistics converges to some limit value (which determines the true value of the parameter).

Throughout this paper we consider the following well known large deviation results in the literature, and we refer again to Dembo and Zeitouni (1993): the principle of the largest term (Lemma 1.2.15), the contraction principle (Theorem 4.2.1), the Varadhan's Lemma (Theorem 4.3.1) and the Sanov's Theorem (Theorem 6.2.10).

All the rate functions in this paper can be expressed in terms of the relative entropy between two probability measures (also called Kullback-Leibler divergence). Given two probability measures  $\nu_1$  and  $\nu_2$  on the same measurable space  $(\Omega, \mathcal{B}_\Omega)$ , the relative entropy of  $\nu_1$  with respect to  $\nu_2$  is

$$H(\nu_1|\nu_2) = \begin{cases} \int_{\Omega} \log \left( \frac{d\nu_1}{d\nu_2}(\omega) \right) \nu_1(d\omega) & \text{if } \nu_1 \ll \nu_2 \\ \infty & \text{otherwise,} \end{cases}$$

where  $\nu_1 \ll \nu_2$  means that  $\nu_1$  is absolutely continuous with respect to  $\nu_2$ , and  $\frac{d\nu_1}{d\nu_2}$  is the corresponding density.

We recall the following formulas. If  $N[\mu, \sigma^2]$  is the Normal distribution with mean  $\mu$  and variance  $\sigma^2$ , we have  $H(N[\mu_1, \sigma^2]|N[\mu_2, \sigma^2]) = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2}$ ; if  $G[\alpha, \beta]$  is the Gamma distribution with continuous density  $g_{\alpha, \beta}(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} 1_{(0, \infty)}(y)$  for  $\alpha, \beta > 0$ , we have  $H(G[\alpha, \beta_1]|G[\alpha, \beta_2]) =$

$\alpha\left(\frac{\beta_2}{\beta_1} - 1 - \log\left(\frac{\beta_2}{\beta_1}\right)\right)$ ; if  $P[\lambda]$  is the Poisson distribution with discrete density  $g_\lambda(y) = \frac{\lambda^y}{y!} e^{-\lambda} \mathbf{1}_{\mathbb{N} \cup \{0\}}$  for  $\lambda > 0$ , we have  $H(P[\lambda_1]|P[\lambda_2]) = \lambda_1 \log\left(\frac{\lambda_1}{\lambda_2}\right) - \lambda_1 + \lambda_2$ . Note that, if we consider the law of the constant 0 as  $P[0]$  and the equality  $0 \log 0 = 0$ , the latter formula also holds with  $\lambda_1 = 0$ .

In view of the applications in risk theory it is useful to recall some known results: firstly the LDPs for MLEs (Lemmas 2.1-2.2-2.3), secondly the LDPs for posterior distributions (Lemmas 2.4-2.5-2.6). We cite some old references for the results for MLEs: see e.g. Bahadur and Ranga Rao (1960) and Bahadur and Zabell (1979). A more recent reference for the results for both MLEs and posterior distributions is Macci (2010, section 4) which deals with exponential families: the mean parameter  $\gamma$  coincides with the parameter  $\theta$  for Normal and Poisson samples, whereas for Gamma samples we have to consider a change of variable (via the contraction principle) to transfer the LDP to the family of posterior distributions on  $\theta = \frac{\alpha}{\beta\gamma}$ . We always use the standard notation  $\bar{Z}_n = \frac{1}{n} \sum_{k=1}^n Z_k$ .

**Lemma 2.1 (Normal samples)** *Let  $(Z_n)$  be a sequence of i.i.d.  $N[\theta, 1/r]$  distributed random variables, where  $\theta \in \mathbb{R}$  is unknown. Then the sequence of MLEs  $(\hat{\Theta}_n)$  coincides with  $(\bar{Z}_n)$  and satisfies the LDP with good rate function  $I_\theta$  defined by  $I_\theta(\hat{\theta}) = H(N[\hat{\theta}, 1/r]|N[\theta, 1/r])$ .*

**Lemma 2.2 (Gamma samples)** *Let  $(Z_n)$  be a sequence of i.i.d.  $G[\alpha, \beta\theta]$  distributed random variables, where  $\theta > 0$  is unknown. Then the sequence of MLEs  $(\hat{\Theta}_n)$  coincides with  $(\alpha/(\beta\bar{Z}_n))$  and satisfies the LDP with good rate function  $I_\theta$  defined by  $I_\theta(\hat{\theta}) = H(G[\alpha, \hat{\theta}]|G[\alpha, \theta])$  if  $\hat{\theta} > 0$ , and  $I_\theta(\hat{\theta}) = \infty$  otherwise.*

**Lemma 2.3 (Poisson samples)** *Let  $(Z_n)$  be a sequence of i.i.d.  $P[\theta]$  distributed random variables, where  $\theta > 0$  is unknown. Then the sequence of MLEs  $(\hat{\Theta}_n)$  coincides with  $(\bar{Z}_n)$  and satisfies the LDP with good rate function  $I_\theta$  defined by  $I_\theta(\hat{\theta}) = H(P[\hat{\theta}]|P[\theta])$  if  $\hat{\theta} \geq 0$ , and  $I_\theta(\hat{\theta}) = \infty$  otherwise.*

In view of Lemmas 2.4-2.5-2.6 presented below we need some further preliminaries. We denote the posterior distribution given  $(Z_1, \dots, Z_n) = (z_1, \dots, z_n)$  by  $\pi_n$ , namely we set  $\pi_n = \pi_0(\cdot|Z_1 = z_1, \dots, Z_n = z_n)$  where  $\pi_0$  is the prior distribution. Moreover we denote the support of the prior distribution  $\pi_0$  by  $\mathbb{S}(\pi_0)$ ; obviously  $\mathbb{S}(\pi_0)$  is a closed subset of the parameter space  $\Theta$ .

In each lemma we shall assume the following condition **(C)** on  $\hat{\theta}_n$ , where  $\hat{\theta}_n$  is the sampled value of  $\hat{\Theta}_n$  as in Lemmas 2.1-2.2-2.3 above:

$$\boxed{\text{(C): } \hat{\theta}_n \rightarrow \hat{\theta} \text{ as } n \rightarrow \infty \text{ for some } \hat{\theta} \in \mathbb{S}(\pi_0).}$$

Then  $\hat{\theta}$  can be actually interpreted as the true value of the parameter.

**Lemma 2.4 (Normal samples)** *Consider the same situation of Lemma 2.1 and assume that **(C)** holds. Then  $(\pi_n)$  satisfies the LDP with good rate function  $I(\cdot|\hat{\theta})$  defined by  $I(\theta|\hat{\theta}) = H(N[\hat{\theta}, 1/r]|N[\theta, 1/r])$  if  $\theta \in \mathbb{S}(\pi_0)$ , and  $I(\theta|\hat{\theta}) = \infty$  otherwise.*

**Lemma 2.5 (Gamma samples)** *Consider the same situation of Lemma 2.2 and assume that **(C)** holds. Then  $(\pi_n)$  satisfies the LDP with good rate function  $I(\cdot|\hat{\theta})$  defined by  $I(\theta|\hat{\theta}) = H(G[\alpha, \hat{\theta}]|G[\alpha, \theta])$  if  $\theta \in \mathbb{S}(\pi_0)$ , and  $I(\theta|\hat{\theta}) = \infty$  otherwise.*

**Lemma 2.6 (Poisson samples)** *Consider the same situation of Lemma 2.3 and assume that **(C)** holds. Then  $(\pi_n)$  satisfies the LDP with good rate function  $I(\cdot|\hat{\theta})$  defined by  $I(\theta|\hat{\theta}) = H(P[\hat{\theta}]|P[\theta])$  if  $\theta \in \mathbb{S}(\pi_0)$ , and  $I(\theta|\hat{\theta}) = \infty$  otherwise.*

### 3 Sequences of probabilities with exponential decay

We start with the definition of sequence of probabilities  $((\rho_n(\underline{q}|\theta)) : \theta \in \Theta)$  with an exponential decay (as  $n \rightarrow \infty$ ).

**Definition 3.1** *The sequences of probabilities  $((\rho_n(\underline{q}|\theta)) : \theta \in \Theta)$  have exponential decay if, for some  $d \geq 1$ , there exist continuous functions  $w_1, \dots, w_d : \Theta \rightarrow [0, \infty)$  and numbers  $c_1, \dots, c_d \in (0, 1]$  and  $M \in [0, \infty)$  such that*

$$c_i e^{-(nq_i + M)w_i(\theta)} \leq \rho_n(\underline{q}|\theta) \leq \sum_{j=1}^d e^{-nq_j w_j(\theta)} \quad (1)$$

for all  $i \in \{1, \dots, d\}$ ,  $\theta \in \Theta$ ,  $\underline{q} = (q_1, \dots, q_d) \in (0, \infty)^d$  and  $n \geq 1$ .

Now we prove the large deviation estimates for any sequence of probabilities with an exponential decay. We start with the classical approach.

**Proposition 3.2 (Classical approach)** *Let  $((\rho_n(\underline{q}|\theta)) : \theta \in \Theta)$  be sequences of probabilities with exponential decay. Moreover let  $\bar{w}_i^\theta(q_i)$  be defined by*

$$\bar{w}_i^\theta(q_i) := \inf_{\hat{\theta} \in \Theta} \{q_i w_i(\hat{\theta}) + I_\theta(\hat{\theta})\}.$$

Let  $(\hat{\Theta}_n)$  be a sequence of random variables which satisfies the LDP with good rate function  $I_\theta$ , which uniquely vanishes at  $\theta$  (the true value of the parameter). Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[\rho_n(\underline{q}|\hat{\Theta}_n)] = -\min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\}.$$

PROOF. We have to check the upper bound

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[\rho_n(\underline{q}|\hat{\Theta}_n)] \leq -\min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\}$$

and the lower bound

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[\rho_n(\underline{q}|\hat{\Theta}_n)] \geq -\min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\}.$$

*Proof of the upper bound.* For each fixed  $i \in \{1, \dots, d\}$  we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[e^{-nq_i w_i(\hat{\Theta}_n)}] = \sup_{\hat{\theta} \in \Theta} \{-q_i w_i(\hat{\theta}) - I_\theta(\hat{\theta})\} = -\bar{w}_i^\theta(q_i) \quad (2)$$

by the Varadhan's Lemma; thus we obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[\rho_n(\underline{q}|\hat{\Theta}_n)] &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \sum_{j=1}^d \mathbb{E}_\theta[e^{-nq_j w_j(\hat{\Theta}_n)}] \right) = \\ &= \max \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[e^{-nq_i w_i(\hat{\Theta}_n)}] : i \in \{1, \dots, d\} \right\} = -\min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\} \end{aligned}$$

by the second inequality in (1), the principle of the largest term and (2).

*Proof of the lower bound.* Let  $i \in \{1, \dots, d\}$  and  $\varepsilon > 0$  be arbitrarily fixed. Then we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[\rho_n(\underline{q}|\hat{\Theta}_n)] \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta[c_i e^{-(nq_i + M)w_i(\hat{\Theta}_n)}] =$$

$$= \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta [e^{-n(q_i + M/n)w_i(\hat{\Theta}_n)}] \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta [e^{-n(q_i + \varepsilon)w_i(\hat{\Theta}_n)}] = -\bar{w}_i^\theta(q_i + \varepsilon)$$

by the first inequality in (1) and the Varadhan's Lemma. We notice that  $-\bar{w}_i^\theta(\cdot)$  is continuous on  $[0, \infty)$ ; indeed it is easy to check that it is convex on the open set  $(0, \infty)$  (and therefore continuous on  $(0, \infty)$ ; see e.g. Theorem 3.2 in Rudin, 1986), and  $\lim_{q_i \downarrow 0} -\bar{w}_i^\theta(q_i) = -\bar{w}_i^\theta(0) = 0$  since  $0 \geq -\bar{w}_i^\theta(q_i) \geq -q_i w_i(\theta)$  (by taking into account the equality  $I_\theta(\theta) = 0$ ). Then we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta [\rho_n(\underline{q} | \hat{\Theta}_n)] \geq -\bar{w}_i^\theta(q_i)$$

letting  $\varepsilon$  go to zero, and we obtain the lower bound taking  $\max\{-\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\}$  at the right hand side.  $\square$

We can also prove a Bayesian version of Proposition 3.2. We have the same proof (we have to replace the classical approach items with the analogous items of the Bayesian approach) and therefore we omit it. We shall denote again the sampled value of  $\hat{\Theta}_n$  by  $\hat{\theta}_n$ .

**Proposition 3.3 (Bayesian approach)** *Let  $((\rho_n(\underline{q} | \theta)) : \theta \in \Theta)$  be sequences of probabilities with exponential decay. Consider a prior distribution  $\pi_0$  on  $\theta$  and let  $\pi_n$  be the posterior distribution given  $n$  sampled values. Assume that there exists a sequence of sufficient statistics  $(\hat{\Theta}_n)$  such that, if  $\hat{\theta}_n \rightarrow \hat{\theta}$  as  $n \rightarrow \infty$  for some  $\hat{\theta} \in \Theta$ , the sequence  $(\pi_n)$  satisfies the LDP with good rate function  $I(\cdot | \hat{\theta})$ , which uniquely vanishes at  $\hat{\theta}$  (the true value of the parameter). Moreover let  $\bar{w}_i(q_i | \hat{\theta})$  be defined by*

$$\bar{w}_i(q_i | \hat{\theta}) := \inf_{\theta \in \Theta} \{q_i w_i(\theta) + I(\theta | \hat{\theta})\}. \quad (3)$$

Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \int_{\Theta} \rho_n(\underline{q} | \theta) \pi_n(d\theta) = -\min\{\bar{w}_i(q_i | \hat{\theta}) : i \in \{1, \dots, d\}\}.$$

**Comparison between classical and Bayesian approaches.** In the next sections the sequences of probabilities with exponential decay are suitable level crossing probabilities. Thus, motivated by potential applications in risk theory, now we investigate whether the Bayesian approach is more or less conservative than the classical approach. More precisely, by taking into account Propositions 3.2 and 3.3, we can say the Bayesian approach is more (less) conservative than the classical approach if  $\min\{\bar{w}_i(q_i | \hat{\theta}) : i \in \{1, \dots, d\}\}$  is smaller (larger) than  $\min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\}$  for each  $\hat{\theta} \in \Theta$ .

In all the examples in this paper we can check the equality

$$I_\theta(\hat{\theta}) = I(\theta | \hat{\theta}) \quad \text{for all } \theta, \hat{\theta} \in \Theta \quad (4)$$

if  $\mathbb{S}(\pi_0) = \Theta$ ; this is what typically happens, even if there exists a counterexample (see Proposition 6.1 in Macci and Petrella, 2009). Then, if we restrict the attention on the case  $\mathbb{S}(\pi_0) = \Theta$ , (4) yields

$$\bar{w}_i^\theta(q_i) = \inf_{\theta \in \Theta} \{q_i w_i(\theta) + I_{\hat{\theta}}(\theta)\} = \inf_{\theta \in \Theta} \{q_i w_i(\theta) + I(\hat{\theta} | \theta)\},$$

and we cannot say if it is larger or smaller than  $\bar{w}_i(q_i | \hat{\theta})$  in (3). Thus, in general, we cannot compare

$$\min\{\bar{w}_i(q_i | \hat{\theta}) : i \in \{1, \dots, d\}\} \quad \text{and} \quad \min\{\bar{w}_i^\theta(q_i) : i \in \{1, \dots, d\}\},$$

but they coincide if  $I(\theta | \hat{\theta}) = I(\hat{\theta} | \theta)$  for all  $\theta, \hat{\theta} \in \Theta$ . For instance this happens for the statistical models based on the Brownian motion with unknown drift (and known variance parameter), where the rate functions are expressed in terms of the relative entropy between two Normal distributions with the same variance (indeed we have  $H(N[\mu_1, \sigma^2] | N[\mu_2, \sigma^2]) = H(N[\mu_2, \sigma^2] | N[\mu_1, \sigma^2])$ ).

A different comparison between the classical approach and the Bayesian approach has been also considered in Ganesh et al. (1998). The difference concerns the classical approach which is based on the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \rho_n(\underline{q}|\hat{\theta}) = -\min\{q_i w_i(\hat{\theta}) : i \in \{1, \dots, d\}\};$$

thus, in that case, the Bayesian approach is more conservative than the classical one since the inequality  $\min\{\bar{w}_i(q_i|\hat{\theta}) : i \in \{1, \dots, d\}\} \leq \min\{q_i w_i(\hat{\theta}) : i \in \{1, \dots, d\}\}$  holds. In our opinion the classical approach presented in this paper captures all the uncertainty about the unknown parameter (as happens for the Bayesian approach), and therefore seems to be more appropriate than the one in Ganesh et al. (1998).

## 4 Applications in risk theory (standard models)

In this section we consider two different sequences  $((\rho_n(q|\theta)) : \theta \in \Theta)$  which concern level crossing probabilities for compound Poisson processes and Brownian motions. More precisely we deal with the statistical models in Macci and Petrella (2006) (where only Bayesian approach is discussed) and, for each one, we study the two following sequences of level crossing probabilities: firstly the level goes to infinity (we set  $Q = nq$  for some fixed  $q > 0$  and  $n \rightarrow \infty$ ), secondly we consider a suitable scaling called slow Markov walk limit (we refer to the presentation in Asmussen and Nielsen, 1995; see also Bucklew, 1990). For each statistical model we check the hypotheses of Propositions 3.2 and 3.3: firstly we check the exponential decay condition; secondly we provide a LDP as required in their statements.

**Preliminaries on compound Poisson process.** A compound Poisson process  $(S(t))$  is defined by  $S(t) = \sum_{k=1}^{N(t)} B_k$ , where:  $(N(t))$  is a Poisson process with intensity  $\lambda$ , i.e.  $N(t) = \sum_{n \geq 1} 1_{T_1 + \dots + T_n \leq t}$ , and  $(T_k)$  are i.i.d. exponentially distributed with mean  $1/\lambda$ ;  $(B_k)$  are i.i.d. and positive random variables with common law  $\ell$ , and independent of  $(T_k)$ . This model is often used in insurance where  $Q$  is the initial capital,  $p$  the premium rate,  $(N(t))$  is the claim number process,  $(B_k)$  are the claims and  $(S(t))$  is the aggregate claims process. The infinite horizon level crossing probability  $\Psi_{(\lambda, \ell)}^{(p)}(Q) = P(\{\exists t \geq 0 : S(t) \geq Q + pt\})$  represents the ruin probability.

We have upper and lower bounds for  $\Psi_{(\lambda, \ell)}^{(p)}(Q)$  if  $\ell$  is concentrated on a bounded set. Here we consider the set  $[0, M]$  where  $M$  is the maximum claim size and we have the following refinement of the Lundberg's inequality, where  $w_1(\lambda, \ell)$  is defined below in eq. (5):

$$e^{(Q+M)w_1(\lambda, \ell)} \leq \Psi_{(\lambda, \ell)}^{(p)}(Q) \leq e^{Qw_1(\lambda, \ell)}.$$

It is also known that, if  $\ell = G[1, \beta]$  (i.e.  $\ell$  is the exponential distribution with mean  $1/\beta$ ), we have  $\Psi_{(\lambda, G[1, \beta])}^{(p)}(Q) = \left(1 - \frac{\max\{\beta p - \lambda, 0\}}{\beta p}\right) e^{-\frac{\max\{\beta p - \lambda, 0\}}{p} Q}$  (see e.g. equation (5.3.8) in Rolski et al., 1999).

The slow Markov walk limit scaling consists of the sequence of parameters  $(\lambda_n, \ell_n) := (n\lambda, \ell[n])$ , where  $\ell[n]$  is the law of the random variables  $\{B_k/n : k \geq 1\}$ . In particular, if  $\ell = G[1, \beta]$ , we have  $\ell[n] = G[1, n\beta]$ .

**Preliminaries on Brownian motion.** We recall that  $(S(t))$  is a Brownian motion with drift  $\mu$  and variance parameter  $1/r$  if, for each fixed  $h \geq 1$  and  $0 < t_1 < \dots < t_h$ , the random variables  $S(t_1), S(t_2) - S(t_1), \dots, S(t_h) - S(t_{h-1})$  are independent and normally distributed, with means  $\mu t_1, \mu(t_2 - t_1), \dots, \mu(t_h - t_{h-1})$  and variances  $t_1/r, (t_2 - t_1)/r, \dots, (t_h - t_{h-1})/r$ . If we consider the infinite horizon level crossing probability  $\Psi_{(\mu, r)}(Q) = \inf\{t \geq 0 : S(t) \geq Q\}$ , we



have  $\Psi_{(\mu,r)}(Q) = e^{-\max\{-2\mu r, 0\}Q}$  (see e.g. Corollary 1.6 in Asmussen, 2000); note that there is a minor error in the expression presented in Macci and Petrella (2006).

The slow Markov walk limit scaling consists of the sequence of parameters  $(\mu_n, r_n) := (\mu, nr)$ , i.e. the variance parameter simply goes to zero.

#### 4.1 Compound Poisson process with bounded jumps

Here we have  $\theta = (\lambda, \ell)$  and  $\Theta = (0, \infty) \times M_1[0, M]$ , where  $M_1[0, M]$  is the space of probability measures on  $[0, M]$  for some  $M \in (0, \infty)$  equipped with the metric of weak convergence. Moreover we think to have  $\rho_n(q|\lambda, \ell) = \Psi_{(\lambda, \ell)}^{(p)}(nq)$  or  $\rho_n(q|\lambda, \ell) = \Psi_{(n\lambda, \ell[n])}^{(p)}(q)$ . Note that we could consider two simplified statistical models in which only  $\lambda$  or  $\ell$  is unknown, and the other parameter is known. In these cases the results can be obtained with minor changes of the results presented below and therefore we omit the details.

It is easy to check (for details see equation (4.3) and its consequence in Macci and Petrella, 2006) that, for both the choices of  $\rho_n(q|\lambda, \ell)$ , (1) holds with  $d = 1$ ,  $c_1 = 1$ , and

$$w_1(\lambda, \ell) = \sup\{\gamma \geq 0 : \lambda(\mathcal{G}_\ell(\gamma) - 1) - p\gamma \leq 0\}, \quad (5)$$

where  $\mathcal{G}_\ell(\gamma) = \int_0^M e^{\gamma x} \ell(dx)$  is the moment generating function of concerning  $\ell$ ; in particular see Lemma 4.2 in Macci and Petrella (2006) for the continuity of  $w_1(\lambda, \ell)$ .

We consider the  $n$ -sample  $(Z_1, \dots, Z_n) = ((T_1, B_1), \dots, (T_n, B_n))$ .

*The LDP in the statement of Proposition 3.2.* Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = (\bar{T}_n^{-1}, \bar{\ell}_n)$ , where:  $\bar{T}_n = \frac{1}{n} \sum_{k=1}^n T_k$  is the empirical mean of  $T_1, \dots, T_n$ ;  $\bar{\ell}_n = \frac{1}{n} \sum_{k=1}^n \delta_{B_k}$  is the empirical law of  $B_1, \dots, B_n$ , where  $\delta_x$  is the concentrated unit measure concentrated at  $x$ . Then the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_{(\lambda, \ell)}$  defined by

$$I_{(\lambda, \ell)}(\hat{\lambda}, \hat{\ell}) = \begin{cases} H(G[1, \hat{\lambda}]|G[1, \lambda]) + H(\hat{\ell}|\ell) & \text{if } (\hat{\lambda}, \hat{\ell}) \in \Theta \\ \infty & \text{otherwise} \end{cases}$$

by the LDP of  $(\bar{T}_n^{-1})$  provided by Lemma 2.2 (with  $\alpha = \beta = 1$  and  $(T_1, \dots, T_n)$  in place of  $(Z_1, \dots, Z_n)$ ), by the LDP of  $(\bar{\ell}_n)$  provided by the Sanov's Theorem and by Corollary 2.9 in Lynch and Sethuraman (1987) which provides the LDP for sequences of multivariate random variables with independent components.

*The LDP in the statement of Proposition 3.3.* We consider the prior distribution  $\pi_0 = \nu_0 \otimes D_{\eta_0}$  on  $(\lambda, \ell)$  where  $\nu_0$  is the prior distribution on  $\lambda$  and  $D_{\eta_0}$  is the Dirichlet Process prior on  $\ell$  with hyperparameter the finite measure  $\eta_0$  on  $[0, M]$  (see e.g. Ferguson, 1973). Then the posterior distribution on  $(\lambda, \ell)$  given the  $n$ -sample is  $\pi_n = \nu_n \otimes D_{\eta_n} := \nu_0(\cdot|T_1, \dots, T_n) \otimes D_{\eta_0}(\cdot|B_1, \dots, B_n)$ , where  $D_{\eta_0}(\cdot|B_1, \dots, B_n) = D_{\eta_0 + \sum_{i=1}^n \delta_{B_i}}$  by Theorem 1 in Ferguson (1973). Then, if  $\hat{\theta}_n \rightarrow (\hat{\lambda}, \hat{\ell})$  as  $n \rightarrow \infty$  for some  $(\hat{\lambda}, \hat{\ell}) \in \mathbb{S}(\nu_0) \times \mathbb{S}(D_{\eta_0})$ , the sequence of posterior distributions  $(\pi_n)$  on  $(\lambda, \ell)$  satisfies the LDP with good rate function  $I(\cdot|\hat{\lambda}, \hat{\ell})$  defined by

$$I(\lambda, \ell|\hat{\lambda}, \hat{\ell}) = \begin{cases} H(G[1, \hat{\lambda}]|G[1, \lambda]) + H(\hat{\ell}|\ell) & \text{if } (\lambda, \ell) \in \mathbb{S}(\nu_0) \times \mathbb{S}(D_{\eta_0}) \\ \infty & \text{otherwise} \end{cases}$$

by the LDP of  $(\nu_n)$  provided by Lemma 2.5, by the LDP of  $(D_{\eta_n})$  provided by Theorem 1 in Ganesh and O'Connell (2000) and by Corollary 2.9 in Lynch and Sethuraman (1987) which provides the LDP for sequences of product measures.

**Another statistical model where  $\lambda$  is unknown and  $\ell$  is known.** Here  $\theta = \lambda$  and  $\Theta = (0, \infty)$ . We consider a different  $n$ -sample based on the claim numbers in consecutive unit time intervals, i.e.  $(Z_1, \dots, Z_n) = (N(1), \dots, N(n) - N(n-1))$ .

The LDP in the statement of Proposition 3.2. Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = \bar{Z}_n = \frac{N(n)}{n}$ . Then, by Lemma 2.3, the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_\lambda$  defined by  $I_\lambda(\hat{\lambda}) = H(P[\hat{\lambda}]|P[\lambda])$  if  $\hat{\lambda} \geq 0$ , and  $I_\lambda(\hat{\lambda}) = \infty$  otherwise.

The LDP in the statement of Proposition 3.3. If  $\hat{\theta}_n \rightarrow \hat{\lambda}$  as  $n \rightarrow \infty$  for some  $\hat{\lambda} \in \mathbb{S}(\pi_0)$ , by Lemma 2.6 the sequence of posterior distributions  $(\pi_n)$  on  $\lambda$  satisfies the LDP with good rate function  $I(\cdot|\hat{\lambda})$  defined by  $I(\lambda|\hat{\lambda}) = H(P[\hat{\lambda}]|P[\lambda])$  if  $\lambda \in \mathbb{S}(\pi_0)$ , and  $I(\lambda|\hat{\lambda}) = \infty$  otherwise.

## 4.2 Compound Poisson process with exponential jumps

One could consider  $\ell = G[1, \beta]$  and therefore the inference problem with  $\theta = (\lambda, \beta)$  and  $\Theta = (0, \infty) \times (0, \infty)$ . In such a case we should think to have  $\rho_n(q|\lambda, \beta) = \Psi_{(\lambda, G[1, \beta])}^{(p)}(nq)$  or  $\rho_n(q|\lambda, \beta) = \Psi_{(n\lambda, G[1, n\beta])}^{(p)}(q)$ . Unfortunately in this case we cannot prove the limits in Propositions 3.2-3.3 because we cannot check the lower bound in (1) (an explanation can be found in subsection 4.3 in Macci and Petrella, 2006).

We do not have this problem assuming to have  $\beta = k\lambda$  for some known  $k > 0$  (as in subsection 4.3 in Macci and Petrella, 2006). Then here we have  $\theta = \lambda$  and  $\Theta = (0, \infty)$ . Moreover we think to have  $\rho_n(q|\lambda) = \Psi_{(\lambda, G[1, k\lambda])}^{(p)}(nq)$  or  $\rho_n(q|\lambda) = \Psi_{(n\lambda, G[1, kn\lambda])}^{(p)}(q)$ .

It is easy to check (for details see equation (4.5) and the first part of subsection 4.3 in Macci and Petrella, 2006) that, for both the choices of  $\rho_n(q|\lambda)$ , (1) holds with  $d = 1$ ,  $c_1 = \min\{(kp)^{-1}, 1\}$ ,  $M = 0$  and  $w_1(\lambda) = k\lambda \max\{1 - (kp)^{-1}, 0\}$ .

We consider the  $n$ -sample  $(Z_1, \dots, Z_n) = (kB_1 + T_1, \dots, kB_n + T_n)$ .

The LDP in the statement of Proposition 3.2. Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = 2\bar{Z}_n^{-1} = 2(k\bar{B}_n + \bar{T}_n)^{-1}$ , where  $\bar{T}_n = \frac{1}{n} \sum_{k=1}^n T_k$  is the empirical mean of  $T_1, \dots, T_n$  and  $\bar{B}_n = \frac{1}{n} \sum_{k=1}^n B_k$  is the empirical mean of  $B_1, \dots, B_n$ . Then, by Lemma 2.2 (with  $\alpha = 2$  and  $\beta = 1$ ), the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_\lambda$  defined by  $I_\lambda(\hat{\lambda}) = H(G[2, \hat{\lambda}]|G[2, \lambda])$  if  $\hat{\lambda} > 0$ , and  $I_\lambda(\hat{\lambda}) = \infty$  otherwise.

The LDP in the statement of Proposition 3.3. If  $\hat{\theta}_n \rightarrow \hat{\lambda}$  as  $n \rightarrow \infty$  for some  $\hat{\lambda} \in \mathbb{S}(\pi_0)$ , by Lemma 2.5 the sequence of posterior distributions  $(\pi_n)$  on  $\lambda$  satisfies the LDP with good rate function  $I(\cdot|\hat{\lambda})$  defined by  $I(\lambda|\hat{\lambda}) = H(G[2, \hat{\lambda}]|G[2, \lambda])$  if  $\lambda \in \mathbb{S}(\pi_0)$ , and  $I(\lambda|\hat{\lambda}) = \infty$  otherwise.

## 4.3 Brownian motion

In this subsection we consider two cases:  $\mu$  is unknown and  $r$  is known;  $\mu$  is known and  $r$  is unknown. We present a statistical model for the first case, and two statistical models for the second one.

**A statistical model where  $\mu$  is unknown and  $r$  is known.** Here we have  $\theta = \mu$  and  $\Theta = \mathbb{R}$  (thus  $r$  is known). Moreover we think to have one of the two following cases:  $\rho_n(q|\mu) = \Psi_{(\mu, r)}(nq)$  or  $\rho_n(q|\mu) = \Psi_{(\mu, nr)}(q)$ .

It is easy to check that, for both the choices of  $\rho_n(q|\mu)$ , (1) holds with  $d = 1$ ,  $c_1 = 1$ ,  $M = 0$  and  $w_1(\mu) = \max\{-2\mu r, 0\}$ .

We consider the  $n$ -sample  $(Z_1, \dots, Z_n) = (S(1), \dots, S(n) - S(n-1))$ .

The LDP in the statement of Proposition 3.2. Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = \bar{Z}_n = \frac{S(n)}{n}$ . Then, by Lemma 2.1, the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_\mu$  defined by  $I_\mu(\hat{\mu}) = H(N[\hat{\mu}, 1/r]|N[\mu, 1/r])$ .

The LDP in the statement of Proposition 3.3. If  $\hat{\theta}_n \rightarrow \hat{\mu}$  as  $n \rightarrow \infty$  for some  $\hat{\mu} \in \mathbb{S}(\pi_0)$ , by Lemma 2.4 the sequence of posterior distributions  $(\pi_n)$  on  $\mu$  satisfies the LDP with good rate function  $I(\cdot|\hat{\mu})$  defined by  $I(\mu|\hat{\mu}) = H(N[\hat{\mu}, 1/r]|N[\mu, 1/r])$  if  $\mu \in \mathbb{S}(\pi_0)$ , and  $I(\mu|\hat{\mu}) = \infty$  otherwise.

**A statistical model where  $\mu$  is known and  $r$  is unknown.** Here we have  $\theta = r$  and  $\Theta = (0, \infty)$ ; thus  $\mu$  is known and, in order to avoid the trivial situation in which the level crossing occurs with probability 1, we should consider  $\mu < 0$ . Moreover we think to have one of the two following cases:  $\rho_n(q|r) = \Psi_{(\mu,r)}(nq)$  or  $\rho_n(q|r) = \Psi_{(\mu,nr)}(q)$ .

It is easy to check that, for both the choices of  $\rho_n(q|r)$ , (1) holds with  $d = 1$ ,  $c_1 = 1$ ,  $M = 0$  and  $w_1(r) = \max\{-2\mu r, 0\}$ .

We consider the  $n$ -sample  $(Z_1, \dots, Z_n) = ((S(1) - \mu)^2, \dots, (S(n) - S(n-1) - \mu)^2)$ .

*The LDP in the statement of Proposition 3.2.* Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = \bar{Z}_n^{-1}$ . Then, by Lemma 2.2 (with  $\alpha = \beta = 1/2$ ), the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_r$  defined by  $I_r(\hat{r}) = H(G[1/2, \hat{r}]|G[1/2, r])$  if  $\hat{r} > 0$ , and  $I_r(\hat{r}) = \infty$  otherwise.

*The LDP in the statement of Proposition 3.3.* If  $\hat{\theta}_n \rightarrow \hat{r}$  as  $n \rightarrow \infty$  for some  $\hat{r} \in \mathbb{S}(\pi_0)$ , by Lemma 2.5 the sequence of posterior distributions  $(\pi_n)$  on  $r$  satisfies the LDP with good rate function  $I(\cdot|\hat{r})$  defined by  $I(r|\hat{r}) = H(G[1/2, \hat{r}]|G[1/2, r])$  if  $r \in \mathbb{S}(\pi_0)$ , and  $I(r|\hat{r}) = \infty$  otherwise.

**Another statistical model where  $\mu$  is known and  $r$  is unknown.** Here we have  $\theta = r$  and  $\Theta = (0, \infty)$  as before, and  $\mu < 0$  is known. We consider a different  $n$ -sample based on the i.i.d. first passage times  $V_1, \dots, V_n$  to a barrier  $a < 0$  concerning  $n$  independent replications of the Brownian motion; since  $\mu < 0$ , the negativeness of  $a$  ensures that each passage time occurs within a finite time with probability 1. Then the random variables  $V_1, \dots, V_n$  are  $IG[a/\mu, a^2r]$  distributed, where  $IG[a/\mu, a^2r]$  is a suitable Inverse Gaussian distribution (see equation (1.2) and the beginning of section 3.2 in Seshadri, 1998); in other words these random variables have continuous density

$$f(v) = \frac{|a|\sqrt{r}}{\sqrt{2\pi}v^{3/2}} \exp\left(-\frac{r\mu^2(v-a/\mu)^2}{2v}\right) 1_{(0,\infty)}(v).$$

Furthermore it is known that the i.i.d. random variables  $\left\{\frac{(V_i - a/\mu)^2}{V_i} : i \in \{1, \dots, n\}\right\}$  are  $G[1/2, \mu^2 r/2]$  distributed (indeed each random variable  $\bar{Q} = a^2 r \frac{(V_i - a/\mu)^2}{(a/\mu)^2 V_i}$  is  $\chi_1^2 = G[1/2, 1/2]$  distributed by Proposition 1.3 in Seshadri, 1998).

We consider the  $n$ -sample  $(Z_1, \dots, Z_n) = \left(\frac{(V_1 - a/\mu)^2}{V_1}, \dots, \frac{(V_n - a/\mu)^2}{V_n}\right)$ .

*The LDP in the statement of Proposition 3.2.* Let  $(\hat{\Theta}_n)$  be the estimators defined by  $\hat{\Theta}_n = 1/(\mu^2 \bar{Z}_n)$ . Then, by Lemma 2.2 (with  $\alpha = 1/2$  and  $\beta = \mu^2/2$ ), the sequence  $(\hat{\Theta}_n)$  satisfies the LDP with good rate function  $I_r$  defined by  $I_r(\hat{r}) = H(G[1/2, \hat{r}]|G[1/2, r])$  if  $\hat{r} > 0$ , and  $I_r(\hat{r}) = \infty$  otherwise.

*The LDP in the statement of Proposition 3.3.* If  $\hat{\theta}_n \rightarrow \hat{r}$  as  $n \rightarrow \infty$  for some  $\hat{r} \in \mathbb{S}(\pi_0)$ , by Lemma 2.5 the sequence of posterior distributions  $(\pi_n)$  on  $r$  satisfies the LDP with good rate function  $I(\cdot|\hat{r})$  defined by  $I(r|\hat{r}) = H(G[1/2, \hat{r}]|G[1/2, r])$  if  $r \in \mathbb{S}(\pi_0)$ ,  $I(r|\hat{r}) = \infty$  otherwise.

## 5 Applications in risk theory (bivariate models)

In this section we consider sequences of probabilities  $(\rho_n(q|\theta)) : \theta \in \Theta$  having interest in risk theory and, as we shall see, the exponential decay condition holds with  $d = 2$ . More precisely we deal with the level crossing probabilities of a (possibly piecewise) linear barrier

$$b(t; Q_1, p_1, Q_2, p_2) = \min\{Q_1 + p_1 t, Q_2 + p_2 t\} \quad (6)$$

for  $Q_1, p_1, Q_2, p_2 > 0$ . These probabilities appear in the bivariate risk model in Avram et al. (2008) and can be reduced to one-dimensional level crossing probabilities (some details are recalled below in subsection 5.3); in Avram et al. (2008) only the compound Poisson process is considered and here we also present a version of the model based on the Brownian motion. We have to consider

again some items presented in the previous section; we have the same preliminaries, the same sequences of estimators  $(\widehat{\Theta}_n)$  and the same sequences of posterior distributions  $(\pi_n)$ . Here we only check the exponential decay condition with  $d = 2$  for the sequences  $((\rho_n(\underline{q}|\theta)) : \theta \in \Theta)$  concerning each statistical model.

### 5.1 Compound Poisson process

Let  $(S(t))$  be the compound Poisson process in the previous section and let  $\overline{\Psi}_{(\lambda,\ell)}(Q_1, Q_2)$  be the level crossing probabilities defined by

$$\overline{\Psi}_{(\lambda,\ell)}(Q_1, Q_2) = P(\{\exists t \geq 0 : S(t) \geq b(t; Q_1, p_1, Q_2, p_2)\});$$

then, if we consider the level crossing probabilities  $\Psi_{(\lambda,\ell)}^{(p)}(Q)$  in the previous section, we have

$$\Psi_{(\lambda,\ell)}^{(p_i)}(Q_i) \leq \overline{\Psi}_{(\lambda,\ell)}(Q_1, Q_2) \leq \Psi_{(\lambda,\ell)}^{(p_1)}(Q_1) + \Psi_{(\lambda,\ell)}^{(p_2)}(Q_2) \quad \text{for } i \in \{1, 2\}. \quad (7)$$

**Case with bounded jumps.** We consider  $\rho_n(q_1, q_2|\lambda, \ell) = \overline{\Psi}_{(\lambda,\ell)}(nq_1, nq_2)$  or  $\rho_n(q_1, q_2|\lambda, \ell) = \overline{\Psi}_{(n\lambda, \ell[n])}(q_1, q_2)$ . Arguing as in subsection 4.1 and by (7), one can check the exponential decay condition for both the choices of  $\rho_n(q_1, q_2|\lambda, \ell)$  with  $d = 2$ ,  $c_i = 1$  for  $i \in \{1, 2\}$ , and

$$w_i(\lambda, \ell) = \sup\{\gamma \geq 0 : \lambda(\mathcal{G}_\ell(\gamma) - 1) - p_i\gamma \leq 0\} \quad \text{for } i \in \{1, 2\}.$$

**Case with exponential jumps.** We consider  $\rho_n(q_1, q_2|\lambda) = \overline{\Psi}_{(\lambda, G[1, k\lambda])}(nq_1, nq_2)$  or  $\rho_n(q_1, q_2|\lambda) = \overline{\Psi}_{(n\lambda, G[1, kn\lambda])}(q_1, q_2)$ . Arguing as in subsection 4.2 and by (7), one can check the exponential decay condition for both the choices of  $\rho_n(q_1, q_2|\lambda)$  with  $d = 2$ ,  $c_i = \min\{(kp_i)^{-1}, 1\}$  for  $i \in \{1, 2\}$ ,  $M = 0$  and  $w_i(\lambda) = k\lambda \max\{1 - (kp_i)^{-1}, 0\}$  for  $i \in \{1, 2\}$ .

### 5.2 Brownian motion

Let  $(S(t))$  be the Brownian motion in the previous section and we use the notation

$$\Psi_{(\mu,r)}^{(p)}(Q) = P(\{\exists t \geq 0 : S(t) \geq Q + pt\});$$

then we consider the level crossing probabilities

$$\overline{\Psi}_{(\mu,r)}(Q_1, Q_2) = P(\{\exists t \geq 0 : S(t) \geq b(t; Q_1, p_1, Q_2, p_2)\})$$

and we have

$$\Psi_{(\mu,r)}^{(p_i)}(Q_i) \leq \overline{\Psi}_{(\mu,r)}(Q_1, Q_2) \leq \Psi_{(\mu,r)}^{(p_1)}(Q_1) + \Psi_{(\mu,r)}^{(p_2)}(Q_2) \quad \text{for } i \in \{1, 2\}. \quad (8)$$

**The statistical model where  $\mu$  is unknown and  $r$  is known.** We consider  $\rho_n(q_1, q_2|\mu) = \overline{\Psi}_{(\mu,r)}(nq_1, nq_2)$  or  $\rho_n(q_1, q_2|\mu) = \overline{\Psi}_{(\mu, nr)}(q_1, q_2)$ . Arguing as in subsection 4.3 and by (8), one can check the exponential decay condition for both the choices of  $\rho_n(q_1, q_2|\mu)$  with  $d = 2$ ,  $c_i = 1$  for  $i \in \{1, 2\}$ ,  $M = 0$  and  $w_i(\mu) = \max\{-2(\mu - p_i)r, 0\}$  for  $i \in \{1, 2\}$ .

**The two statistical models where  $\mu$  is known and  $r$  is unknown.** We think to have  $\rho_n(q_1, q_2|r) = \overline{\Psi}_{(\mu,r)}(nq_1, nq_2)$  or  $\rho_n(q_1, q_2|r) = \overline{\Psi}_{(\mu, nr)}(q_1, q_2)$ . Arguing as in subsection 4.3 and by (8), one can check the exponential decay condition for both the choices of  $\rho_n(q_1, q_2|r)$  with  $d = 2$ ,  $c_i = 1$  for  $i \in \{1, 2\}$ ,  $M = 0$  and  $w_i(r) = \max\{-2(\mu - p_i)r, 0\}$  for  $i \in \{1, 2\}$ .

In order to avoid the trivial case where the level crossing occurs with probability 1, we should have  $\mu < p_1, p_2$  (which is indeed equivalent to  $w_1(r), w_2(r) > 0$ ). Anyway we remark that the condition  $\mu < 0$  is needed if we consider the  $n$ -sample  $Z_1, \dots, Z_n$  in terms of the first passage times  $V_1, \dots, V_n$  to a barrier  $a < 0$  as in subsection 4.3.

### 5.3 Some connections with the geometric considerations in Avram et al. (2008)

In this subsection we illustrate the relationship between the geometric considerations in Avram et al. (2008) and some inequalities between the exponential decay rates in this paper.

Following the description of the model in Avram et al. (2008), we assume to have two companies that divide between them both claims and premia in some specified proportions  $\delta_1, \delta_2 \in (0, 1)$  with  $\delta_1 + \delta_2 = 1$ ; in detail

$$R_i(t) = u_i + v_i t - \delta_i S(t) \quad \text{for } i \in \{1, 2\},$$

where  $u_1, u_2 > 0$  are the initial capitals and  $v_1, v_2 > 0$  are the premium rates. Thus, if we think to the choices of  $(S(t))$  in this paper, the compound Poisson process is more appropriate than the Brownian motion. We note that we have  $R_1(t) \leq 0$  or  $R_2(t) \leq 0$  for some  $t \geq 0$  if and only if we have  $S(t) \geq b(t; Q_1, p_1, Q_2, p_2)$  for some  $t \geq 0$ , where  $b(t; Q_1, p_1, Q_2, p_2)$  is defined by (6) and  $Q_i = u_i/\delta_i$  and  $p_i = v_i/\delta_i$  for  $i \in \{1, 2\}$ . We also note that, if  $p_1 > p_2$ , the two-dimensional ruin problem is equivalent to the ruin problem for the second company if  $Q_2 \leq Q_1$  (indeed  $b(t; Q_1, p_1, Q_2, p_2)$  reduces to the linear barrier  $Q_2 + p_2 t$ ). In other words, if  $p_1 > p_2$  and if we consider the general notation

$$\bar{\Psi}_\theta(Q_1, Q_2) = P(\{\exists t \geq 0 : S(t) \geq b(t; Q_1, p_1, Q_2, p_2)\})$$

with  $\theta \in \Theta$  (instead of the specific notation of each example), we have  $\bar{\Psi}_\theta(Q_1, Q_2) = \Psi_\theta^{(p_2)}(Q_2)$  if  $Q_2 \leq Q_1$ . This fact has the following geometric interpretation: if  $p_1 > p_2$ , the two-dimensional ruin problem is equivalent to the ruin problem for the second company if  $u_2 \leq \frac{\delta_2}{\delta_1} u_1$ , i.e. if the initial capitals of the two companies is situated below of the line  $u_2 = \frac{\delta_2}{\delta_1} u_1$  (in the positive quadrant).

In conclusion we illustrate how these geometric considerations in Avram et al. (2008) can be related with some inequalities between  $\bar{w}_1^\theta(q_1)$  and  $\bar{w}_2^\theta(q_2)$ , and between  $\bar{w}_1(q_1|\hat{\theta})$  and  $\bar{w}_2(q_2|\hat{\theta})$ . For all the examples presented in this section assume  $p_1 \geq p_2$ ; then, for all  $\theta \in \Theta$ , we have  $w_1(\theta) \geq w_2(\theta)$  whence we obtain  $q_2 w_2(\theta) \leq q_1 w_1(\theta)$  if  $q_2 \leq q_1$ . In conclusion, if  $p_1 \geq p_2$  and  $q_2 \leq q_1$ , we have:

$$\begin{aligned} \min\{\bar{w}_1^\theta(q_1), \bar{w}_2^\theta(q_2)\} &= \bar{w}_2^\theta(q_2) \quad (\text{or } \bar{w}_2^\theta(q_2) \leq \bar{w}_1^\theta(q_1)) \quad \text{for all } \theta \in \Theta; \\ \min\{\bar{w}_1(q_1|\hat{\theta}), \bar{w}_2(q_2|\hat{\theta})\} &= \bar{w}_2(q_2|\hat{\theta}) \quad (\text{or } \bar{w}_2(q_2|\hat{\theta}) \leq \bar{w}_1(q_1|\hat{\theta})) \quad \text{for all } \hat{\theta} \in \Theta. \end{aligned}$$

## Acknowledgements

I thank Mauro Piccioni for inspiring the classical approach in this paper and Florin Avram for telling me about the model in his paper. I also thank the referee for some useful comments and for pointing out an inexactness in the previous version of the paper. The financial support of the Research Grant PRIN 2008 *Probability and Finance* is gratefully acknowledged.

## References

- [1] Asmussen S., *Ruin Probabilities*. World Scientific, Singapore 2000.
- [2] Asmussen S. and Nielsen H.M., *Ruin probabilities via local adjustment coefficients*. J. Appl. Probab. **33** (1995) 736–755.
- [3] Avram F., Palmowski Z., and Pistorius M., *A two dimensional ruin problem on the positive quadrant*. Insurance Math. Econom. **42** (2008) 227–234.
- [4] Bahadur R.R. and Ranga Rao R., *On deviations of the sample mean*. Ann. Math. Statist. **31** (1960) 1015–1027.

- [5] Bahadur R.R. and Zabell S.L., *Large deviations of the sample mean in general vector space*. Ann. Probab. **7** (1979) 587–621.
- [6] Bucklew J.A., *Large Deviations Techniques in Decision, Simulation and Estimation*. John Wiley and Sons, New York 1990.
- [7] Dembo A. and Zeitouni O., *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston 1993.
- [8] Ferguson T.S., *A Bayesian analysis of some nonparametric problems*. Ann. Statist. **1** (1973) 209–230.
- [9] Ganesh A., Green P., O’Connell N. and Pitts S., *Bayesian network management*. Queueing Systems Theory Appl. **28** (1998) 267–282.
- [10] Ganesh A. and O’Connell N., *An inverse of Sanov’s Theorem*. Statist. Probab. Lett. **42** (1999) 201–206.
- [11] Ganesh A. and O’Connell N., *A large deviation principle for Dirichlet posteriors*. Bernoulli **6** (2000) 1021–1034.
- [12] Lynch J. and Sethuraman J., *Large deviations for processes with independent increments*. Ann. Probab. **15** (1987) 610–627.
- [13] Macci C., *Large deviations for estimators of some threshold parameters*. Stat. Methods Appl. **19** (2010) 63–77.
- [14] Macci C. and Petrella L., *Mixture of conjugate prior distributions and large deviations for level crossing probabilities*. Sankhyā **68** (2006) 61–89.
- [15] Macci C. and Petrella L., *Censored exponential data: large deviations for MLEs and posterior distributions*. Comm. Statist. Theory Methods **38** (2009) 2435–2452.
- [16] Rolski T., Schmidli H., Schmidt V. and Teugels J.L., *Stochastic Processes for Insurance and Finance*. John Wiley and Sons, Chichester 1999.
- [17] Rudin W., *Real and Complex Analysis* (3rd Edition), McGraw-Hill, New York 1986.
- [18] Seshadri V., *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Lecture Notes in Statistics **137**, Springer, New York 1998.