Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 1, pp. 61-89

Mixtures of Conjugate Prior Distributions and Large Deviations for Level Crossing Probabilities

Claudio Macci
University of Rome, Tor Vergata, Italy
Lea Petrella
University of Rome, La Sapienza, Italy

Abstract

In this paper we present asymptotic estimates of level crossing probabilities from a Bayesian point of view, based on large deviations. For the Bayesian analysis we choose a finite mixture of conjugate prior distributions to model the uncertainty on the unknown parameters of the two classes of stochastic processes considered: the Brownian motion and the compound Poisson process with upward jumps and negative drift. The estimates of level crossing probabilities are derived as a consequence of large deviation principles for posterior distributions.

AMS (2000) *subject classification.* Primary 62C10, 60F10.
Keywords and phrases. Large deviations, Bayes asymptotics, mixture of conjugate families, level crossing probabilities.

1 Introduction

Level crossing probabilities estimation has become important in several fields like mathematical risk theory connected with insurance problems, queueing theory and many others. In this paper we estimate the probabilities that some real valued stochastic processes cross a positive level from a Bayesian point of view.

More precisely we assume $\left(Z_t^\theta\right)$ to be a real valued stochastic process starting at zero with unknown parameter $\theta \in \Theta$ and we define $T_Q(\theta)$ as the first time at which $\left(Z_t^\theta\right)$ reaches a positive level Q i.e.

$$T_Q(\theta) = \inf\left\{t \geq 0 : Z_t^\theta \geq Q\right\}.$$

Since the true value of θ is unknown, one may wish to asses the probability to cross a positive level based on experience. This leads us to consider the so called *predictive level crossing probability*

$$\int_{\Theta} p(Q,\theta)\pi(d\theta|[\text{data}]_n), \tag{1.1}$$

where $p(Q,\theta) = P(T_Q(\theta) < \infty)$ is the probability that the process crosses the positive level Q when the parameter $\theta \in \Theta$ and $\pi(\cdot|[\text{data}]_n)$ is the posterior distribution of θ given an i.i.d. n-sample from $\left(Z_t^{\theta}\right)$.

From a gambler's point of view, if a gambler has initial capital Q, then $p(Q,\theta)$ is the probability of ultimate ruin and (1.1) is the predictive probability of ruin.

For the estimation of (1.1) we choose a finite mixture of conjugate prior distributions to model the uncertainty on the unknown parameter θ of the two classes of stochastic processes considered: the Brownian motion and the compound Poisson process with upward jumps and negative drift. This approach can lead to a full statistical analysis, capturing all the uncertainties related to the crossing probability estimation; other approaches based on estimation of $p(Q,\widehat{\theta})$, where $\widehat{\theta}$ is an estimate of θ, can give a very misleading inference (Ganesh et al., 1998).

The two classes of stochastic processes considered in this paper are widely used in the literature; in particular in insurance theory, the compound Poisson process with upward jumps and negative drift is the claim surplus process for the so called Cramér-Lundberg model (Embrechts et al., 1997) or compound Poisson model (Asmussen, 1987).

In what follows we present an asymptotic Bayesian analysis based on large deviations (Dembo and Zeitouni, 1993); in particular we show the large deviations for the posterior distributions as the sample n goes to infinity under the assumption of the convergence of some suitable sufficient statistics to some limit value. We refer to Theorem 1 of Ganesh and O'Connell (2000) when we deal with the compound Poisson process and we choose a Dirichlet process prior (Ferguson, 1973) for the unknown common law of the jumps. In all of the other cases we apply Gärtner Ellis Theorem (Dembo and Zeitouni, 1993).

Starting from the large deviations of posterior distributions we derive two kinds of large deviation estimates of predictive level crossing probabilities using Varadhan's Lemma (Dembo and Zeitouni, 1993): one when the level

Q goes to infinity setting $p(Q,\theta) = p(qn,\theta)$ for some fixed $q > 0$, the other inspired by the so called *slow Markov walk limit* (Asmussen and Nielsen, 1995, Example 1 and Theorem 2) setting $p(Q,\theta) = p(q,\theta(n))$ for some fixed $q > 0$ and $\theta(n)$ chosen in a suitable way. The slow Markov walk limit for the Brownian motion consists of letting the variance parameter go to zero; for the compound Poisson process the intensity of the underlying Poisson process diverges at the same rate at which the jumps go to zero.

Some of the results presented here are a generalization of those presented in Macci (2004). Other papers on large deviations in Bayesian setting with a similar approach are Ganesh and O'Connell (1998), Ganesh and O'Connell (2000), Paschalidis and Vassiralas (2001) and Eichelsbacher and Ganesh (2002). Fu and Kass (1988) presented some results on the same topic with a different approach. Some asymptotic results for posterior distribution under mixture of conjugate priors can be found in Brunner and Lo (1996). A non-asymptotic Bayesian analysis for the gambler's ruin problem has been considered by Tsay and Tsao (2003).

We conclude with the outline of the paper. In section 2 we consider some preliminaries on large deviations and we discuss the choice of the prior distribution. Results on Brownian motion are presented in section 3 where large deviation principles for posterior distributions are stated for the multivariate cases when both drift and precision are unknown; we restrict our attention on the univariate cases when we estimate the predictive level crossing probabilities. Results on compound Poisson process are presented in section 4 where we consider different possibilities for the common law ℓ of the jumps. Some concluding remarks on the differences between the Bayesian approach and the frequentist approach are presented in section 5. The appendix at the end of the paper is devoted to recall the statements of Gärtner Ellis Theorem and Varadhan's Lemma.

2 Preliminaries

2.1. Preliminaries on large deviations. Let Ω be a Hausdorff topological space with Borel σ-algebra $\mathcal{B}_\Omega$ and let a lower semicontinuous function $I : \Omega \to [0,\infty]$ be called a rate function; then a sequence of probability measures (ν_n) on $(\Omega, \mathcal{B}_\Omega)$ satisfies the large deviation principle (LDP hereafter) with rate function I if

$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n(F) \leq -\inf_{\omega \in F} I(\omega) \;\; (\forall F \text{ closed})$$

and

$$\liminf_{n\to\infty}\frac{1}{n}\log\nu_n(G)\geq-\inf_{\omega\in G}I(\omega)\ (\forall G\text{ open}).$$

A rate function I is *good* if all the level sets

$$\{\omega\in\Omega:I(\omega)\leq\alpha\}\quad(\alpha>0)$$

are compact sets.
A sequence of Ω-valued random variables (Y_n) satisfies the LDP if the inequalities above hold with $\nu_n=P(Y_n\in\cdot)$. When dealing with posterior distributions, we have a family of random probability measures which satisfies the LDP almost surely.

Since in this paper the rate functions can be expressed in terms of the relative entropy, let us recall some definitions and some useful formulae concerning Normal, Gamma and Poisson distributions.

Let ν_1 and ν_2 be two probability measures on the same measurable space $(\Omega,\mathcal{B}_\Omega)$. The relative entropy of ν_1 with respect to ν_2 is

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

It is known that $H(\nu_1|\nu_2)$ is nonnegative and is equal to zero if and only if $\nu_1=\nu_2$ (Kullback, 1959).

Let $N_{\boldsymbol{m},\Sigma}$ be a d-variate normal distribution with mean vector $\boldsymbol{m}$ and precision matrix $R=\Sigma^{-1}$. Then, if we denote the transpose of $\boldsymbol{x}$ by $\boldsymbol{x}'$, we have

$$\begin{aligned}H\big(N_{\boldsymbol{m}_1,R^{-1}}|N_{\boldsymbol{m}_2,R^{-1}}\big)=&\\ \int_{\mathbb{R}^d}\log&\left[\frac{(\det R/(2\pi))^{\frac{d}{2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_1)'R(\boldsymbol{x}-\boldsymbol{m}_1))}{(\det R/(2\pi))^{\frac{d}{2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_2)'R(\boldsymbol{x}-\boldsymbol{m}_2))}\right]\\ &\times(\det R/(2\pi))^{\frac{d}{2}}\exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_1)'R(\boldsymbol{x}-\boldsymbol{m}_1)\Big)d\boldsymbol{x},\end{aligned}$$

and one can check that

$$H\left(N_{\boldsymbol{m}_1,R^{-1}}|N_{\boldsymbol{m}_2,R^{-1}}\right)=\frac{1}{2}(\boldsymbol{m}_1-\boldsymbol{m}_2)'R(\boldsymbol{m}_1-\boldsymbol{m}_2).\tag{2.1}$$

The relative entropy between two Gamma distributions G_{α,β_1} and G_{α,β_2} is

$$\begin{aligned} H\left(G_{\alpha,\beta_1}|G_{\alpha,\beta_2}\right) =& \int_0^\infty \log\left(\frac{\frac{\beta_1^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_1\theta}}{\frac{\beta_2^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_2\theta}}\right)\frac{\beta_1^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_1\theta}d\theta \\ =&\alpha\left(\frac{\beta_2}{\beta_1}-1-\log\left(\frac{\beta_2}{\beta_1}\right)\right), \end{aligned}$$

and the relative entropy between two Poisson distributions P_{λ_1} and P_{λ_2} is

$$H(P_{\lambda_1}|P_{\lambda_2}) = \sum_{k\geq 0}\log\left(\frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}}{\frac{\lambda_2^k}{k!}e^{-\lambda_2}}\right)\frac{\lambda_1^k}{k!}e^{-\lambda_1} = \lambda_1\log\left(\frac{\lambda_1}{\lambda_2}\right) - \lambda_1 + \lambda_2.$$

In view of what follows it is useful to recall the following formulae. The first one relates the relative entropies between Gamma distributions to the one between Poisson distributions:

$$H(P_{\widehat{\lambda}}|P_\lambda) = \widehat{\lambda}H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right)\ (\widehat{\lambda},\lambda>0). \tag{2.2}$$

The second one is useful when we apply Gärtner Ellis Theorem with Gamma distributions: For $\alpha,\beta>0$,

$$\sup_{\gamma<\beta}\left[\gamma\theta-\alpha\log\left(\frac{\beta}{\beta-\gamma}\right)\right]=\begin{cases}\alpha\left(\beta\frac{\theta}{\alpha}-1-\log\left(\beta\frac{\theta}{\alpha}\right)\right)=H\left(G_{\alpha,\frac{\alpha}{\beta}}|G_{\alpha,\theta}\right) & \text{if } \theta>0\\ \infty & \text{if } \theta\leq 0\end{cases} \tag{2.3}$$

2.2. The choice of prior distribution on the unknown parameters. In Bayesian inference the role played by the prior distribution is crucial so its careful specification is of great importance. The use of conjugate families for prior distributions has been criticized as being too restrictive. On the other hand, in problems related with large deviations, some hypotheses on prior distribution seem to be necessary: Ganesh and O'Connell (1998, Introduction) pointed out that the large deviation results can be stated without additional assumptions on prior distributions only when the sample space is finite.

In our case, the choice of a mixture of conjugate prior distributions to model uncertainty on the parameter θ of $\left(Z_t^\theta\right)$ can be a good compromise. They are flexible enough to represent a wide variety of prior beliefs; moreover Diaconis and Ylvisaker (1985) and Dalal and Hall (1983) prove that they can approximate any distribution arbitrarily accurately. Although the accuracy of approximation may require the use of many distributions in the mixture, a small number of them can capture a wide variety of forms and characteristics of prior beliefs.

Let θ be the unknown parameter of the process $\left(Z_t^\theta\right)$. Consider π as the prior distribution on θ which can be written as a finite mixture of conjugate prior distributions π_i:

$$\pi = \sum_{i=1}^{k} p_i \pi_i$$

where $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^{k} p_i = 1$, are known weights.
In this paper we show that the application of Gärtner Ellis Theorem to derive the LDP for posterior distributions, is preserved under finite mixtures of conjugate priors. This result is a consequence of the following Lemma 2.1 which is proved as a more general result concerning any finite mixture of prior distributions not necessarily conjugate.

Lemma 2.1. *Let us consider $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^{k} p_i = 1$, some prior distributions $\pi_1, \ldots, \pi_k$ and their mixture $\pi = \sum_{i=1}^{k} p_i \pi_i$ with weights $p_1, \ldots, p_k$. Let us denote the posterior distributions concerning $\pi_1, \ldots, \pi_k, \pi$ by $\pi_1(\cdot|[\mathrm{data}]_n), \ldots, \pi_k(\cdot|[\mathrm{data}]_n), \pi(\cdot|[\mathrm{data}]_n)$ respectively. Furthermore let f be a measurable function. Assume that there exists the limit*

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n} \log \int e^{nf(\theta)} \pi_i(d\theta|[\mathrm{data}]_n),$$

with $\Lambda(f) \in (-\infty, \infty]$ which does not depend on $i \in \{1, \ldots, k\}$. Then

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{nf(\theta)} \pi(d\theta|[\mathrm{data}]_n). \tag{2.4}$$

Proof. It is known (O'Hagan and Forster, 2004) that the posterior distribution $\pi(\cdot|[\mathrm{data}]_n)$ can be expressed as a suitable finite mixture of the posterior distributions $\pi_1(\cdot|[\mathrm{data}]_n), \ldots, \pi_k(\cdot|[\mathrm{data}]_n)$ with weights depending on $p_1, \ldots, p_k$ and on $[\mathrm{data}]_n$; more precisely we have

$$\pi(\cdot|[\mathrm{data}]_n) = \sum_{i=1}^{k} p_i([\mathrm{data}]_n) \pi_i(\cdot|[\mathrm{data}]_n),$$

where $p_1([\mathrm{data}]_n), \ldots, p_k([\mathrm{data}]_n) \geq 0$ such that $\sum_{i=1}^k p_i([\mathrm{data}]_n) = 1$. Thus we have

$$\int e^{nf(\theta)} \pi(d\theta|[\mathrm{data}]_n) = \sum_{i=1}^{k} p_i([\mathrm{data}]_n) \int e^{nf(\theta)} \pi_i(d\theta|[\mathrm{data}]_n).$$

Then we consider two separate cases.

Case $\Lambda(f) < \infty$. For all $\varepsilon > 0$ there exists $\overline{n}$ such that for all $n \geq \overline{n}$ we have

$$e^{n(\Lambda(f)-\varepsilon)} < \int e^{nf(\theta)} \pi_i(d\theta|[\mathrm{data}]_n) < e^{n(\Lambda(f)+\varepsilon)} \ (\forall i \in \{1, \ldots, k\});$$

then, by multiplying by $p_i([\mathrm{data}]_n)$ and by taking the sum over $i \in \{1, \ldots, k\}$, we obtain

$$e^{n(\Lambda(f)-\varepsilon)} < \int e^{nf(\theta)} \pi(d\theta|[\mathrm{data}]_n) < e^{n(\Lambda(f)+\varepsilon)},$$

and (2.4) holds.

Case $\Lambda(f) = \infty$. For all $M > 0$ there exists $\overline{n}$ such that for all $n \geq \overline{n}$ we have

$$\int e^{nf(\theta)} \pi_i(d\theta|[\mathrm{data}]_n) > e^{nM} \ (\forall i \in \{1, \ldots, k\});$$

then, by multiplying by $p_i([\mathrm{data}]_n)$ and by taking the sum over $i \in \{1, \ldots, k\}$, we obtain

$$\int e^{nf(\theta)} \pi(d\theta|[\mathrm{data}]_n) > e^{nM},$$

and (2.4) holds. □

3 Bayesian analysis for Brownian motion

Let (Z_t^θ) be a 1-variate Brownian motion starting at zero, with $\theta = (m, r^{-1})$, where $m \in \mathbb{R}$ is the drift and $r > 0$ is the precision. In this case the level crossing probabilities have the following simple expression:

$$P(T_Q(m, r^{-1}) < \infty) = e^{-w(m,r^{-1})Q} \text{ where } w(m, r^{-1}) = \max\Big\{0, -2mr^{-1}\Big\}. \tag{3.1}$$

Furthermore $\theta(n) = (m, (nr)^{-1})$ is the choice for the slow Markov walk limit.

In the next subsections we prove Bayesian LDPs for a d-variate Brownian motion $(Z_t^{(\boldsymbol{m},R^{-1})})$ starting at the origin, where $\boldsymbol{m} \in \mathbb{R}^d$ and R are respectively the drift vector and the precision matrix. We state the results for three cases: when the drift is unknown and the precision is known (subsection 3.1), when the precision is unknown and the drift is known (subsection 3.2) and when both the drift and precision are unknown (subsection 3.3). In each case the prior distribution of the corresponding unknown parameter is modelled with mixture of conjugate prior distributions when the data is

$$[\mathrm{data}]_n = \left(Z_1^{(\boldsymbol{m},R^{-1})}, Z_2^{(\boldsymbol{m},R^{-1})} - Z_1^{(\boldsymbol{m},R^{-1})}, \ldots, Z_n^{(\boldsymbol{m},R^{-1})} - Z_{n-1}^{(\boldsymbol{m},R^{-1})}\right).$$

Once the LDP is proved, we obtain the large deviation estimates of the predictive level crossing probabilities in the univariate case.

3.1. First case: unknown drift and known precision. Let $\left(Z_t^{(\boldsymbol{m},R^{-1})}\right)$ be a d-variate Brownian motion starting at zero with unknown drift $\boldsymbol{m} \in \mathbb{R}^d$ and known precision matrix R. Let $\overline{X}_n$ be defined as:

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n}\left(Z_k^{(\boldsymbol{m},R^{-1})} - Z_{k-1}^{(\boldsymbol{m},R^{-1})}\right) = \frac{Z_n^{(\boldsymbol{m},R^{-1})}}{n}. \tag{3.2}$$

It is known (DeGroot, 1970) that, given any d-variate Normal prior distribution $N_{\boldsymbol{m}_0,R_0^{-1}}$ on $\boldsymbol{m}$, the posterior distribution is

$$N_{\boldsymbol{m}_0,R_0^{-1}}(\cdot|[\mathrm{data}]_n) = N_{\boldsymbol{m}_n,R_n^{-1}},$$

where $\boldsymbol{m}_n = (R_0 + nR)^{-1}(R_0\boldsymbol{m}_0 + nR\overline{X}_n)$ and $R_n = R_0 + nR$.

LEMMA 3.1. *Let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i N_{\boldsymbol{m}_0(i),R_0^{-1}(i)}$ on $\boldsymbol{m}$. Assume $\overline{X}_n \to \widehat{\boldsymbol{m}}$ as $n \to \infty$. Then $(\pi(\cdot|[\mathrm{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\boldsymbol{m}}}$ defined by $I_{\widehat{\boldsymbol{m}}}(\boldsymbol{m}) = H\left(N_{\widehat{\boldsymbol{m}},R^{-1}}|N_{\boldsymbol{m},R^{-1}}\right)$ (with $\boldsymbol{m} \in \mathbb{R}^d$).*

PROOF. Let $\boldsymbol{\gamma} \in \mathbb{R}^d$ be arbitrarily fixed. Then, since $\overline{X}_n \to \widehat{\boldsymbol{m}}$ as $n \to \infty$, for any conjugate prior distribution $N_{\boldsymbol{m}_0, R_0^{-1}}$ we have

$$\begin{aligned}
\Lambda_{\widehat{\boldsymbol{m}}}(\boldsymbol{\gamma}) &= \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{n\boldsymbol{\gamma}'\boldsymbol{m}} N_{\boldsymbol{m}_0, R_0^{-1}}(d\boldsymbol{m}|[\text{data}]_n) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \exp\Big(n\boldsymbol{\gamma}'\boldsymbol{m}_n + \frac{1}{2}(n\boldsymbol{\gamma})' R_n^{-1}(n\boldsymbol{\gamma})\Big) \\
&= \lim_{n\to\infty} \boldsymbol{\gamma}' \left(\frac{R_0}{n} + R\right)^{-1} \left(\frac{R_0}{n}\boldsymbol{m}_0 + R\overline{X}_n\right) + \frac{1}{2}\boldsymbol{\gamma}' \left(\frac{R_0}{n} + R\right)^{-1} \boldsymbol{\gamma} \\
&= \boldsymbol{\gamma}'\widehat{\boldsymbol{m}} + \frac{1}{2}\boldsymbol{\gamma}' R^{-1} \boldsymbol{\gamma}.
\end{aligned}$$

Furthermore

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{n\boldsymbol{\gamma}'\boldsymbol{m}} \pi(d\boldsymbol{m}|[\text{data}]_n) = \Lambda_{\widehat{\boldsymbol{m}}}(\boldsymbol{\gamma})$$

by Lemma 2.1. Then, by Gärtner Ellis Theorem, $(\pi(\cdot|[\text{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\boldsymbol{m}}} = \Lambda^*_{\widehat{\boldsymbol{m}}}$ defined by

$$\Lambda^*_{\widehat{\boldsymbol{m}}}(\boldsymbol{m}) = \sup_{\boldsymbol{\gamma}\in\mathbb{R}^d} [\boldsymbol{\gamma}'\boldsymbol{m} - \Lambda_{\widehat{\boldsymbol{m}}}(\boldsymbol{\gamma})] = H\left(N_{\widehat{\boldsymbol{m}}, R^{-1}} | N_{\boldsymbol{m}, R^{-1}}\right),$$

where the latter equality follows from (2.1). □

Let us consider now $d = 1$ i.e. the univariate Brownian motion case and investigate the estimation of level crossing probabilities. For the next Proposition, the following function (of $q > 0$) is useful:

$$-[\overline{w}(\widehat{m}, r^{-1})](q) = \sup_{m\in\mathbb{R}} \left[-qw(m, r^{-1}) - H\left(N_{\widehat{m}, r^{-1}} | N_{m, r^{-1}}\right)\right].$$

PROPOSITION 3.1. *Let $d = 1$. Furthermore let $k \geq 1$, $p_1, \dots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and consider the mixture of prior distributions $\pi = \sum_{i=1}^k p_i N_{m_0(i), r_0^{-1}(i)}$. Assume $\overline{X}_n \to \widehat{m}$ as $n \to \infty$. Then, for all $q > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}} P(T_{qn}(m, r^{-1}) < \infty) \pi(dm|[\text{data}]_n) = -[\overline{w}(\widehat{m}, r^{-1})](q)$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}} P(T_q(m, (nr)^{-1}) < \infty) \pi(dm|[\text{data}]_n) = -[\overline{w}(\widehat{m}, r^{-1})](q).$$

PROOF. It is an immediate consequence of Lemma 3.1 with $d = 1$ and of Varadhan's Lemma which can be applied since the function $m \mapsto e^{-qw(m, r^{-1})}$ is continuous. □

3.2. Second case: the drift is known and the precision is unknown. Let $\left(Z_t^{(\boldsymbol{m},R^{-1})}\right)$ be a d-variate Brownian motion starting at zero with known drift $\boldsymbol{m} \in \mathbb{R}^d$ and precision matrix θR, where $\theta > 0$ is an unknown parameter and R is a known matrix. Let $\overline{T}_n$ be defined as follows:

$$\overline{T}_n = \frac{1}{n}\sum_{k=1}^{n}\Big(Z_k^{(\boldsymbol{m},R^{-1})} - Z_{k-1}^{(\boldsymbol{m},R^{-1})} - \boldsymbol{m}\Big)' R\Big(Z_k^{(\boldsymbol{m},R^{-1})} - Z_{k-1}^{(\boldsymbol{m},R^{-1})} - \boldsymbol{m}\Big).$$

It is known (DeGroot, 1970) that, given any Gamma prior distribution $G_{\alpha,\beta}$ on θ, the posterior distribution is

$$G_{\alpha,\beta}(\cdot|[\mathrm{data}]_n) = G_{\alpha_n,\beta_n},$$

where $\alpha_n = \alpha + \frac{nd}{2}$ and $\beta_n = \beta + \frac{n}{2}\overline{T}_n$.

LEMMA 3.2. *Let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$ on θ. Assume $\overline{T}_n \to \frac{d}{2\widehat{\theta}}$ as $n \to \infty$. Then $(\pi(\cdot|[\mathrm{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\theta}}$ defined by*

$$I_{\widehat{\theta}}(\theta) = \begin{cases} H\left(G_{\frac{d}{2},\widehat{\theta}}|G_{\frac{d}{2},\theta}\right) & \text{if } \theta > 0, \\ \infty & \text{if } \theta \leq 0. \end{cases}$$

PROOF. Let $\gamma \in \mathbb{R}$ be arbitrarily fixed. Then, since $\overline{T}_n \to \frac{d}{2\widehat{\theta}}$ as $n \to \infty$, for any conjugate prior distribution $G_{\alpha,\beta}$ we have

$$\Lambda_{\widehat{\theta}}(\gamma) = \lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\theta}G_{\alpha,\beta}(d\theta|[\mathrm{data}]_n) =$$

$$= \lim_{n\to\infty}\begin{cases} \frac{\alpha_n}{n}\log\left(\frac{\beta_n}{\beta_n - n\gamma}\right) & \text{if } n\gamma < \beta_n \\ \infty & \text{if } n\gamma \geq \beta_n \end{cases} = \begin{cases} \frac{d}{2}\log\left(\frac{\frac{d}{2\widehat{\theta}}}{\frac{d}{2\widehat{\theta}} - \gamma}\right) & \text{if } \gamma < \frac{d}{2\widehat{\theta}} \\ \infty & \text{if } \gamma \geq \frac{d}{2\widehat{\theta}} \end{cases}.$$

Furthermore

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\theta}\pi(d\theta|[\mathrm{data}]_n) = \Lambda_{\widehat{\theta}}(\gamma)$$

by Lemma 2.1. Then, by Gärtner Ellis Theorem, $(\pi(\cdot|[\mathrm{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\theta}} = \Lambda^*_{\widehat{\theta}}$ defined by

$$\Lambda^*_{\widehat{\theta}}(\theta) = \sup_{\gamma\in\mathbb{R}}[\gamma\theta - \Lambda_{\widehat{\theta}}(\gamma)] = \begin{cases} H\left(G_{\frac{d}{2},\widehat{\theta}}|G_{\frac{d}{2},\theta}\right) & \text{if } \theta > 0, \\ \infty & \text{if } \theta \leq 0, \end{cases}$$

where the latter equality follows from (2.3) with $\alpha = \frac{d}{2}$ and $\beta = \frac{d}{2\widehat{\theta}}\cdot$ □

To estimate the level crossing probabilities in the univariate case it is useful to consider the following function (of $q > 0$):

$$-[\overline{w}(m, \widehat{\theta}^{-1})](q) = \sup_{\theta > 0} \left[-qw\left(m, \widehat{\theta}^{-1}\right) - H\left(G_{\frac{1}{2},\widehat{\theta}} | G_{\frac{1}{2},\theta}\right)\right].$$

Proposition 3.2. *Let $d = 1$. Furthermore let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$. Assume $\overline{T}_n \to \frac{1}{2\widehat{\theta}}$ as $n \to \infty$. Then, for all $q > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_{qn}(m, \theta^{-1}) < \infty)\pi(d\theta|[\text{data}]_n) = -[\overline{w}(m, \widehat{\theta}^{-1})](q)$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_q(m, (n\theta)^{-1}) < \infty)\pi(d\theta|[\text{data}]_n) = -[\overline{w}(m, \widehat{\theta}^{-1})](q).$$

Proof. It is an immediate consequence of Lemma 3.2 with $d = 1$ and of Varadhan's Lemma which can be applied since the function $\theta \mapsto e^{-qw(m,\theta^{-1})}$ is continuous. □

3.3. Third case: unknown drift and precision. Let $(Z_t^{(\boldsymbol{m},R^{-1})})$ be a d-variate Brownian motion starting at zero where both the drift $\boldsymbol{m} \in \mathbb{R}^d$ and $\theta > 0$ of the precision matrix θR are unknown (we assume R to be known). Let $\overline{X}_n$ be as in (3.2) and let $\overline{S}_n$ be defined as follows:

$$\overline{S}_n = \frac{1}{n} \sum_{k=1}^n \Big(Z_k^{(\boldsymbol{m},R^{-1})} - Z_{k-1}^{(\boldsymbol{m},R^{-1})} - \overline{X}_n\Big)' R \Big(Z_k^{(\boldsymbol{m},R^{-1})} - Z_{k-1}^{(\boldsymbol{m},R^{-1})} - \overline{X}_n\Big).$$

It is known (DeGroot, 1970) that, given any d-variate Normal-Gamma prior distribution $NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}$ on $(\boldsymbol{m}, \theta)$ (we recall that $\alpha, \beta > 0$, τ is a $d \times d$ symmetric positive definite matrix and $\boldsymbol{\mu} \in \mathbb{R}^d$), the posterior distribution is

$$NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}(\cdot|[\text{data}]_n) = NG_{\alpha_n,\beta_n,\tau_n,\boldsymbol{\mu}_n},$$

where $\boldsymbol{\mu}_n = (nR+\tau)^{-1}(\tau\boldsymbol{\mu} + nR\overline{X}_n)$, $\alpha_n = \alpha + \frac{nd}{2}$, $\beta_n = \beta + \frac{n}{2}\overline{S}_n + \frac{1}{2}(\boldsymbol{\mu}_n - \boldsymbol{\mu})'\tau(\overline{X}_n - \boldsymbol{\mu})$ and $\tau_n = nR + \tau$.

In the next Proposition we show that we can apply Gärtner Ellis Theorem to the sequence of marginal posterior distributions on θ but not to the sequence of posterior distributions on $(\boldsymbol{m}, \theta)$.

PROPOSITION 3.3. *Let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i NG_{\alpha(i),\beta(i),\tau(i),\boldsymbol{\mu}(i)}$ on $(\boldsymbol{m}, \theta)$. Assume $\overline{X}_n \to \widehat{\boldsymbol{m}}$ and $\overline{S}_n \to \frac{d}{2\widehat{\theta}}$ as $n \to \infty$. Then the sequence of marginal distributions $(\pi(\mathbb{R}^d \times \cdot|[\mathrm{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\boldsymbol{m}},\widehat{\theta}}$ defined by*

$$I_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\theta) = \begin{cases} H\left(G_{\frac{d}{2},\widehat{\theta}}|G_{\frac{d}{2},\theta}\right) & \text{if } \theta > 0, \\ \infty & \text{if } \theta \leq 0. \end{cases}$$

PROOF. Let $(\boldsymbol{\gamma}_1, \gamma_2) \in \mathbb{R}^d \times \mathbb{R}$ be arbitrarily fixed. Let us consider an arbitrary conjugate prior distribution $NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}$. Then we have

$$\begin{aligned}
&\int_{\mathbb{R}^d \times (0,\infty)} e^{n[\boldsymbol{\gamma}_1'\boldsymbol{m}+\gamma_2\theta]} NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}(d\boldsymbol{m}, d\theta|[\mathrm{data}]_n) \\
&= \int_0^\infty G_{\alpha_n,\beta_n}(d\theta) e^{n\gamma_2\theta} \int_{\mathbb{R}^d} N_{\boldsymbol{\mu}_n,(\theta\tau_n)^{-1}}(d\boldsymbol{m}) e^{n\boldsymbol{\gamma}_1'\boldsymbol{m}} \\
&= \int_0^\infty G_{\alpha_n,\beta_n}(d\theta) e^{n\gamma_2\theta} e^{n\boldsymbol{\gamma}_1'\boldsymbol{\mu}_n + \frac{1}{2\theta}(n\boldsymbol{\gamma}_1)'\tau_n(n\boldsymbol{\gamma}_1)} \\
&= e^{n\boldsymbol{\gamma}_1'\boldsymbol{\mu}_n} \int_0^\infty d\theta \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \theta^{\alpha_n-1} e^{-(\beta_n-n\gamma_2)\theta} e^{\frac{1}{2\theta}(n\boldsymbol{\gamma}_1)'\tau_n(n\boldsymbol{\gamma}_1)} \\
&= \begin{cases} e^{n\boldsymbol{\gamma}_1'\boldsymbol{\mu}_n} \left(\frac{\beta_n}{\beta_n - n\gamma_2}\right)^{\alpha_n} \times & \\ \quad \int_0^\infty d\theta \frac{(\beta_n - n\gamma_2)^{\alpha_n}}{\Gamma(\alpha_n)} \theta^{\alpha_n-1} e^{-(\beta_n-n\gamma_2)\theta} e^{\frac{1}{2\theta}(n\boldsymbol{\gamma}_1)'\tau_n(n\boldsymbol{\gamma}_1)} & \text{if } n\gamma_2 < \beta_n, \\ \infty & \text{if } n\gamma_2 \geq \beta_n. \end{cases}
\end{aligned}$$

Let us recall that

$$\int_0^\infty d\theta \frac{(\beta_n - n\gamma_2)^{\alpha_n}}{\Gamma(\alpha_n)} \theta^{\alpha_n-1} e^{-(\beta_n-n\gamma_2)\theta} e^{\frac{1}{2\theta}(n\boldsymbol{\gamma}_1)'\tau_n(n\boldsymbol{\gamma}_1)}$$

is the moment generating function of an Inverse Gamma distributed random variable with parameters (α_n, β_n) computed at $\gamma_2 = \frac{1}{2}(n\boldsymbol{\gamma}_1)'\tau_n(n\boldsymbol{\gamma}_1)$. Then, since the k-th moment is finite if and only if $k < \alpha_n$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d \times (0,\infty)} e^{n[\boldsymbol{\gamma}_1'\boldsymbol{m}+\gamma_2\theta]} NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}(d\boldsymbol{m}, d\theta|[\mathrm{data}]_n) \\
&= \begin{cases} \left(\frac{\beta_n}{\beta_n - n\gamma_2}\right)^{\alpha_n} & \text{if } \boldsymbol{\gamma}_1 = \mathbf{0} \text{ and } n\gamma_2 < \beta_n, \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

The infinite value for $\boldsymbol{\gamma}_1 \neq \mathbf{0}$ is not surprising, since it is known that $NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}(\cdot \times (0,\infty)|[\text{data}]_n)$ is a t distribution, which has heavy tail.

In conclusion, we can define the function $\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}$ by the following limit:

$$\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\boldsymbol{\gamma}_1,\gamma_2) = \lim_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{R}^d\times(0,\infty)} e^{n[\gamma_1'\boldsymbol{m}+\gamma_2\theta]} NG_{\alpha,\beta,\tau,\boldsymbol{\mu}}(d\boldsymbol{m}\times d\theta|[\text{data}]_n)$$
$$= \begin{cases} \frac{d}{2}\log\left(\frac{\frac{d}{2\widehat{\theta}}}{\frac{d}{2\widehat{\theta}}-\gamma_2}\right) & \text{if } \boldsymbol{\gamma}_1 = \mathbf{0} \text{ and } \gamma_2 < \frac{d}{2\widehat{\theta}}, \\ \infty & \text{otherwise.} \end{cases}$$

Furthermore

$$\lim_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{R}^d\times(0,\infty)} e^{n[\gamma_1'\boldsymbol{m}+\gamma_2\theta]}\pi(d\boldsymbol{m}\times d\theta|[\text{data}]_n) = \Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\boldsymbol{\gamma}_1,\gamma_2)$$

by Lemma 2.1. In this situation, if we want to prove the LDP for the sequence $(\pi(\cdot|[\text{data}]_n))$, we cannot apply Gärtner Ellis Theorem because $\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}$ is not finite in a neighborhood of the origin $(\boldsymbol{\gamma}_1,\gamma_2) = (\mathbf{0},0)$. On the other hand it is possible to apply Gärtner Ellis Theorem to the sequence of marginal distributions $(\pi(\mathbb{R}^d\times\cdot|[\text{data}]_n))$ since $\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\mathbf{0},\cdot)$ is a finite function in a neighborhood of the origin $\gamma_2 = 0$. For this reason we can conclude that $(\pi(\mathbb{R}^d\times\cdot|[\text{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\boldsymbol{m}},\widehat{\theta}} = [\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\mathbf{0},\cdot)]^*$, where

$$[\Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\mathbf{0},\cdot)]^*(\theta) = \sup_{\gamma_2\in\mathbb{R}}[\gamma_2\theta - \Lambda_{\widehat{\boldsymbol{m}},\widehat{\theta}}(\mathbf{0},\gamma_2)] = \begin{cases} H\left(G_{\frac{d}{2},\widehat{\theta}}|G_{\frac{d}{2},\theta}\right) & \text{if } \theta > 0, \\ \infty & \text{if } \theta \leq 0, \end{cases}$$

and the latter equality follows from (2.3) with $\alpha = \frac{d}{2}$ and $\beta = \frac{d}{2\widehat{\theta}}\cdot$ □

The rate function $I_{\widehat{\boldsymbol{m}},\widehat{\theta}}$ in Proposition 3.3 coincides with the rate function $I_{\widehat{\theta}}$ in Lemma 3.2 concerning the case with mean $\boldsymbol{m}$ known; this fact seems to be not surprising because $I_{\widehat{\boldsymbol{m}},\widehat{\theta}}$ does not depend on the limit $\widehat{\boldsymbol{m}}$ of $\overline{X}_n$.

4 Bayesian Analysis for Compound Poisson Process

Let (Z_t^θ) be a compound Poisson process with upwards jumps and negative drift; more precisely let us write $Z_t^\theta = \sum_{k=1}^{N_t} B_k - ct$, where: (N_t) is a homogeneous Poisson Process with intensity λ; (B_k) is a sequence of i.i.d. positive random variables and independent of (N_t); $c > 0$ is a constant (we avoid the trivial case $c \leq 0$ since (Z_t^θ) crosses any positive level with probability 1).

The common law and the common moment generating function of the random variables (B_k) are denoted by ℓ and $\mathcal{G}_\ell$ respectively. Furthermore we represent N_t as $N_t = \sum_{k\geq 1} 1_{T_1+\cdots+T_k\leq t}$, where (T_k) is a sequence of i.i.d. exponentially distributed random variables with failure rate λ.

In the compound Poisson process case the unknown parameter is $\theta = (\lambda, \ell)$ and, denoting the common law of the random variables $\left(\frac{B_k}{n}\right)$ by $\frac{1}{n}\ell$, we choose $\theta(n) = \left(n\lambda, \frac{1}{n}\ell\right)$ for the slow Markov walk limit.
Before showing the LDP for the posterior distributions and evaluating the corresponding level crossing probabilities, let us give some preliminary results and notation useful for the following sections.

Consider $w(\lambda, \ell)$ defined as follows

$$w(\lambda, \ell) = \sup\Big\{\gamma \geq 0 : \lambda(\mathcal{G}_\ell(\gamma) - 1) - c\gamma \leq 0\Big\},$$

for which it is true that

$$w\Big(n\lambda, \frac{1}{n}\ell\Big) = nw(\lambda, \ell). \tag{4.1}$$

In the next sections we focus on estimation of level crossing probability when the law ℓ is concentrated on $[0, M]$ for some $M > 0$ (we always consider $M < \infty$) and when $\ell = \exp(\beta)$ i.e. ℓ is the exponential distribution with failure rate β. In all of the cases we have the Lundberg's inequality

$$P(T_Q(\lambda, \ell) < \infty) \leq e^{-w(\lambda,\ell)Q} \tag{4.2}$$

(Asmussen, 1987).

When ℓ is concentrated on $[0, M]$ we can consider a refinement of the Lundberg's inequality

$$e^{-w(\lambda,\ell)(Q+M)} \leq P(T_Q(\lambda, \ell) < \infty) \leq e^{-w(\lambda,\ell)Q}. \tag{4.3}$$

Using (4.3) we have

$$e^{-nw(\lambda,\ell)(q+\frac{M}{n})} \leq P(T_{qn}(\lambda, \ell) < \infty), P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big) \leq e^{-nw(\lambda,\ell)q}$$

which is trivial for $P(T_{qn}(\lambda, \ell) < \infty)$, while for $P\left(T_q\left(n\lambda, \frac{1}{n}\ell\right) < \infty\right)$ we have to take into account (4.1) and (4.3) with $\frac{1}{n}\ell$ and $\frac{M}{n}$ in place of ℓ and M

respectively. Thus, for all $\varepsilon > 0$, there exists $n_\varepsilon \geq 1$ such that for all $n \geq n_\varepsilon$:

$$e^{-nw(\lambda,\ell)(q+\varepsilon)} \leq P(T_{qn}(\lambda,\ell) < \infty), P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big) \leq e^{-nw(\lambda,\ell)q}. \tag{4.4}$$

It is important to remark that, if ℓ is not concentrated on a bounded set, we cannot have the lower bounds in (4.3) and (4.4).

When we assume $\ell = \exp(\beta)$ it is easy to show that

$$w(\lambda, \exp(\beta)) = \frac{\max\{\beta c - \lambda, 0\}}{c} = -\frac{\min\{\lambda - \beta c, 0\}}{c}$$

and we have the closed formula

$$P(T_Q(\lambda, \exp(\beta)) < \infty) = \Big(1 - \frac{w(\lambda, \exp(\beta))}{\beta}\Big) e^{-w(\lambda,\exp(\beta))Q} \tag{4.5}$$

(Embrechts et al., 1997).

In this section we prove the LDP for the posterior distributions of θ and evaluate the predictive level crossing probability in the following cases: the distribution ℓ is unknown and concentrated on $[0, M]$ for some $M > 0$; the distribution ℓ is known and concentrated on $[0, M]$ for some $M > 0$; the distribution $\ell = \exp(\beta)$ and we assume $\beta = k\lambda$ where λ is unknown and $k > 0$ is known.

4.1. The distribution ℓ is unknown and concentrated on $[0, M]$ *for some* $M > 0$. Let $(Z_t^{(\lambda,\ell)})$ be the compound Poisson process presented above and assume the common distribution ℓ of the random variables (B_k) concentrated on $[0, M]$. We denote the family of all probability measures on $[0, M]$ by $M_1[0, M]$, where we assume $M_1[0, M]$ is equipped with the topology of the weak convergence, and by $M_+[0, M]$ the family of all finite nonnegative measures on $[0, M]$.

We choose $\pi \otimes D_\mu$ as the prior distribution for $\theta = (\lambda, \ell)$ where

$$\pi = \sum_{i=1}^{k} p_i G_{\alpha(i),\beta(i)}$$

is a finite mixture of Gamma conjugate priors on λ and D_μ is the Dirichlet process prior for ℓ with parameter $\mu \in M_+[0, M]$ (Ferguson, 1973). The support of D_μ will be denoted by $\mathrm{supp}(D_\mu)$.

For the large deviation Bayesian analysis the n-sample is

$$[\text{data}]_n = ((B_1, T_1), \dots, (B_n, T_n))$$

and the sufficient statistics $(\overline{\ell}_n, \overline{T}_n)$ is defined by the empirical law $\overline{\ell}_n = \frac{1}{n}\sum_{k=1}^n \delta_{B_k}$ of $\{B_1, \dots, B_n\}$ and the empirical mean $\overline{T}_n = \frac{1}{n}\sum_{k=1}^n T_k$ of $\{T_1, \dots, T_n\}$.

The next Lemma provides the LDP for the posterior distributions on the parameter θ.

LEMMA 4.1. *Let $k \geq 1$, $p_1, \dots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$. Let us consider the prior distribution $\pi \otimes D_\mu$ on (λ, ℓ) , where $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$. Assume $\overline{T}_n \to \frac{1}{\widehat{\lambda}}$ and $\overline{\ell}_n \to \widehat{\ell} \in \operatorname{supp}(D_\mu)$ as $n \to \infty$. Then $(\pi \otimes D_\mu(\cdot|[\text{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\lambda},\widehat{\ell}}$ defined by*

$$I_{\widehat{\lambda},\widehat{\ell}}(\lambda, \ell) = \begin{cases} H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right) + H(\widehat{\ell}|\ell) & \text{if } \lambda > 0 \text{ and } \ell \in \operatorname{supp}(D_\mu), \\ \infty & \text{otherwise.} \end{cases}$$

PROOF. Since the sequences (B_k) and (T_k) are independent of each other, we have $\pi \otimes D_\mu(\cdot|[\text{data}]_n)$ is the product measure between $\pi(\cdot|T_1, \dots, T_n)$ and $D_\mu(\cdot|B_1, \dots, B_n)$.

It is known (Ferguson, 1973) that $D_\mu(\cdot|B_1, \dots, B_n) = D_{\mu_n}$, where $\mu_n = \mu + n\overline{\ell}_n$; moreover (D_{μ_n}) satisfies the LDP with good rate function $J^{(2)}_{\widehat{\ell}}$ defined by

$$J^{(2)}_{\widehat{\ell}}(\ell) = \begin{cases} H(\widehat{\ell}|\ell) & \text{if } \ell \in \operatorname{supp}(D_\mu) \\ \infty & \text{otherwise} \end{cases}$$

(Ganesh and O'Connell, 2000, Theorem 1).

To complete the proof we shall show that $(\pi(\cdot|T_1, \dots, T_n))$ satisfies the LDP with good rate function $J^{(1)}_{\widehat{\lambda}}$ defined by

$$J^{(1)}_{\widehat{\lambda}}(\lambda) = \begin{cases} H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right) & \text{if } \lambda > 0, \\ \infty & \text{if } \lambda \leq 0; \end{cases}$$

indeed, in such a case, $\pi \otimes D_\mu(\cdot|[\text{data}]_n) = \pi(\cdot|T_1, \dots, T_n) \otimes D_\mu(\cdot|B_1, \dots, B_n)$ satisfies the LDP with good rate function $I_{\widehat{\lambda},\widehat{\ell}}$ defined by

$$I_{\widehat{\lambda},\widehat{\ell}}(\lambda, \ell) = J^{(1)}_{\widehat{\lambda}}(\lambda) + J^{(2)}_{\widehat{\ell}}(\ell).$$

In order to prove the LDP of $(\pi(\cdot|T_1,\dots,T_n))$, we recall (DeGroot, 1970) that choosing the conjugate prior distribution $G_{\alpha,\beta}$ on λ, we have $G_{\alpha,\beta}(\cdot|T_1,\dots,T_n) = G_{\alpha_n,\beta_n}$, where $\alpha_n = \alpha + n$, $\beta_n = \beta + n\overline{T}_n$. Then

$$\begin{aligned}\Lambda_{\widehat{\lambda}}(\gamma) &= \lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\lambda}G_{\alpha,\beta}(d\lambda|T_1,\dots,T_n)\\ &= \lim_{n\to\infty}\begin{cases}\frac{\alpha_n}{n}\log\left(\frac{\beta_n}{\beta_n-n\gamma}\right) & \text{if } n\gamma<\beta_n\\ \infty & \text{if } n\gamma\ge\beta_n\end{cases}\\ &= \begin{cases}\log\left(\frac{\frac{1}{\widehat{\lambda}}}{\frac{1}{\widehat{\lambda}}-\gamma}\right) & \text{if } \gamma<\frac{1}{\widehat{\lambda}},\\ \infty & \text{if } \gamma\ge\frac{1}{\widehat{\lambda}}.\end{cases}\end{aligned}$$

Furthermore

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\lambda}\pi(d\lambda|T_1,\dots,T_n) = \Lambda_{\widehat{\lambda}}(\gamma)$$

by Lemma 2.1. Then, by Gärtner Ellis Theorem, $(\pi(\cdot|T_1,\dots,T_n))$ satisfies the LDP with good rate function $J^{(1)}_{\widehat{\lambda}} = \Lambda^*_{\widehat{\lambda}}$ defined by

$$\Lambda^*_{\widehat{\lambda}}(\lambda) = \sup_{\lambda\in\mathbb{R}}[\gamma\lambda - \Lambda_{\widehat{\lambda}}(\gamma)] = \begin{cases} H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right) & \text{if } \lambda>0,\\ \infty & \text{if } \lambda\le 0,\end{cases}$$

where the latter equality follows from (2.3) with $\alpha = 1$ and $\beta = \frac{1}{\widehat{\lambda}}$. □

In order to evaluate the large deviation predictive level crossing probabilities stated in the following Proposition 4.1 we need a preliminary result given in Lemma 4.2 and the following function (of $q > 0$):

$$-[\overline{w}(\widehat{\lambda},\widehat{\ell})](q) = \sup_{\lambda>0,\ell\in\mathrm{supp}(D_\mu)}[-qw(\lambda,\ell) - \{H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right) + H(\widehat{\ell}|\ell)\}].$$

LEMMA 4.2. *The function $(\lambda,\ell)\mapsto w(\lambda,\ell)$ is continuous.*

PROOF. To prove the statement we show that for any $(\lambda,\ell)\in(0,\infty)\times M_1[0,M]$, $\lim_{n\to\infty} w(\lambda_n,\ell_n) = w(\lambda,\ell)$ for any sequence $((\lambda_n,\ell_n))$ such that $\lim_{n\to\infty}(\lambda_n,\ell_n) = (\lambda,\ell)$.

For notational convenience we simply write w_n in place of $w(\lambda_n,\ell_n)$ for any $n\ge 1$; moreover we consider the functions H and $(H_n : n\ge 1)$ defined by

$$H(\gamma) = \lambda(\mathcal{G}_\ell(\gamma)-1) - c\gamma \text{ and } H_n(\gamma) = \lambda_n(\mathcal{G}_{\ell_n}(\gamma)-1) - c\gamma.$$

We recall that, since $(\ell_n) \subset M_1[0,M]$ and $[0,M]$ is a bounded set, for any $\gamma \in \mathbb{R}$ we have $\lim_{n\to\infty} \mathcal{G}_{\ell_n}(\gamma) = \mathcal{G}_{\ell}(\gamma)$ and consequently $\lim_{n\to\infty} H_n(\gamma) = H(\gamma)$. Furthermore we have $H_n(w_n) = 0$ for any $n \geq 1$; indeed each function H_n is continuous, assumes finite values and $H_n(0) = 0$.

We can say that there exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha < w(\lambda,\ell) < \beta$ (where α and β depend on λ and ℓ) such that $H(\alpha) < 0$ and $H(\beta) > 0$; thus there exists $\overline{n} \geq 1$ such that $\alpha \leq w_n \leq \beta$ for all $n \geq \overline{n}$. Moreover there exists a subsequence $(w_{n(k)})$ of (w_n) such that $\lim_{k\to\infty} w_{n(k)} = \eta$ for some $\eta \in [\alpha,\beta]$.

If $\eta < w(\lambda,\ell)$ there exists $\eta_- \in (\eta, w(\lambda,\ell))$ such that $w_{n(k)} < \eta_-$ eventually, and we have

$$0 = H_{n(k)}(w_{n(k)}) < H_{n(k)}(\eta_-) \to_{k\to\infty} H(\eta_-) < H(w(\lambda,\ell)) = 0$$

which is impossible; similarly, if $\eta > w(\lambda,\ell)$ there exists $\eta_+ \in (w(\lambda,\ell),\eta)$ such that $w_{n(k)} > \eta_+$ eventually, and we have

$$0 = H_{n(k)}(w_{n(k)}) > H_{n(k)}(\eta_+) \to_{k\to\infty} H(\eta_+) > H(w(\lambda,\ell)) = 0$$

which is impossible. In conclusion any convergent subsequence of (w_n) converges to $w(\lambda,\ell)$ and this concludes the proof. Indeed, if $\lim_{n\to\infty} w_n = w(\lambda,\ell)$ is false, there exists $\varepsilon_0 > 0$ and a subsequence $(w_{n(k)})$ of (w_n) such that $|w_{n(k)} - w(\lambda,\ell)| \geq \varepsilon_0$; such a subsequence has a convergent subsequence because $(w_{n(k)}) \subset [\alpha,\beta]$, $(w_{n(k)})$ converges to $w(\lambda,\ell)$ for the ressoning above and this is a contradiction. $\square$

PROPOSITION 4.1. *Let $k \geq 1$, $p_1,\dots,p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi \otimes D_\mu$ where $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$. Assume $\overline{T}_n \to \frac{1}{\widehat{\lambda}}$ and $\overline{\ell}_n \to \widehat{\ell} \in \operatorname{supp}(D_\mu)$ as $n \to \infty$. Then, for all $q > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P(T_{qn}(\lambda,\ell) < \infty)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) = -[\overline{w}(\widehat{\lambda},\widehat{\ell})](q)$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) = -[\overline{w}(\widehat{\lambda},\widehat{\ell})](q).$$

PROOF. As pointed out at the beginning of section 4 we know that, for all $\varepsilon > 0$, there exists $n_\varepsilon \geq 1$ such that (4.4) holds for all $n \geq n_\varepsilon$. Then, by Lemma 4.1 and by Varadhan's Lemma applied to the continuous function $(\lambda, \ell) \mapsto w(\lambda, \ell)$, we obtain

$$\begin{aligned}
&- [\overline{w}(\widehat{\lambda}, \widehat{\ell})](q + \varepsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P(T_{qn}(\lambda, \ell) < \infty)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P(T_{qn}(\lambda, \ell) < \infty)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) \\
&\leq - [\overline{w}(\widehat{\lambda}, \widehat{\ell})](q)
\end{aligned}$$

and

$$\begin{aligned}
&- [\overline{w}(\widehat{\lambda}, \widehat{\ell})](q + \varepsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,M]} P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big)\pi \otimes D_\mu(d\lambda, d\ell|[\text{data}]_n) \\
&\leq - [\overline{w}(\widehat{\lambda}, \widehat{\ell})](q).
\end{aligned}$$

We conclude the proof by showing that $-[\overline{w}(\widehat{\lambda}, \widehat{\ell})](q)$ is a continuous function of $q > 0$. It is easy to check that $-[\overline{w}(\widehat{\lambda}, \widehat{\ell})](q)$ is a convex function of $q > 0$; moreover this function assumes finite values since $0 \geq -[\overline{w}(\widehat{\lambda}, \widehat{\ell})](q) \geq -qw(\widehat{\lambda}, \widehat{\ell})$ for all $q > 0$. Then the continuity of $-[\overline{w}(\widehat{\lambda}, \widehat{\ell})](q)$ follows from a well known property of convex functions (Rudin, 1986). □

4.2. The distribution ℓ is known and concentrated on $[0, M]$ for some $M > 0$. In this subsection we give results for the predictive level crossing probability in the particular case where ℓ is known and concentrated on $[0, M]$, λ is unknown, the n-sample is

$$[\text{data}]_n = (T_1, \dots, T_n) \tag{4.6}$$

and we assume π as prior distribution for λ as stated in the previous section. Adapting the proof of Proposition 4.3 and considering the function

$$-[\overline{w}^\ell(\widehat{\lambda})](q) = \sup_{\lambda>0} \Big[-qw(\lambda, \ell) - H\Big(G_{1,\widehat{\lambda}}|G_{1,\lambda}\Big)\Big]$$

(of $q>0$), it is easy to check that:

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P(T_{qn}(\lambda,\ell)<\infty)\pi(d\lambda|[\text{data}]_n)=-[\overline{w}^\ell(\widehat{\lambda})](q),$$

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P\Big(T_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big)\pi(d\lambda|[\text{data}]_n)=-[\overline{w}^\ell(\widehat{\lambda})](q).$$

Instead of data (4.6) we can consider the sample

$$[\text{alternative data}]_n=(N_1,N_2-N_1,\dots,N_n-N_{n-1})$$

as in Macci (2004), where as before $N_t=\sum_{k\geq 1}1_{T_1+\dots+T_k\leq t}$. Consider the function (of $q>0$):

$$-[\overline{w}_\ell(\widehat{\lambda})](q)=\sup_{\lambda>0}[-qw(\lambda,\ell)-H(P_{\widehat{\lambda}}|P_\lambda)],$$

where $H(P_{\widehat{\lambda}}|P_\lambda)$ is the relative entropy between two Poisson distributions $P_{\widehat{\lambda}}$ and P_λ stated in section 2. Then we can extend the proof of Proposition 3 of Macci (2004) where the law ℓ coincided with the law δ_1 of a constant random variable equal to 1. Here ℓ is concentrated on a bounded set and the prior distribution on λ is a finite mixture of Gamma distributions $\pi=\sum_{i=1}^k p_iG_{\alpha(i),\beta(i)}$; then we have

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P(T_{qn}(\lambda,\ell)<\infty)\pi(d\lambda|[\text{alternative data}]_n)=-[\overline{w}_\ell(\widehat{\lambda})](q)$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P\Big(T_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big)\pi(d\lambda|[\text{alternative data}]_n)=-[\overline{w}_\ell(\widehat{\lambda})](q)$$

as $\frac{N_n}{n}\to\widehat{\lambda}$ when $n\to\infty$.

Finally it is interesting to compare $-[\overline{w}_\ell(\widehat{\lambda})](q)$ and $-[\overline{w}^\ell(\widehat{\lambda})](q)$ obtained above.
By (2.2) we have

$$-[\overline{w}_\ell(\widehat{\lambda})](q)=\sup_{\lambda>0}\left[-qw(\lambda,\ell)-\widehat{\lambda}H\left(G_{1,\widehat{\lambda}}|G_{1,\lambda}\right)\right],$$

so that $-[\overline{w}_\ell(\widehat{\lambda})](q)$ and $-[\overline{w}^\ell(\widehat{\lambda})](q)$ coincides when $\widehat{\lambda}=1$. Moreover we have $-[\overline{w}_\ell(\widehat{\lambda})](q)\leq-[\overline{w}^\ell(\widehat{\lambda})](q)$ when $\widehat{\lambda}\geq 1$ and $-[\overline{w}_\ell(\widehat{\lambda})](q)\geq-[\overline{w}^\ell(\widehat{\lambda})](q)$ when $\widehat{\lambda}\leq 1$.

4.3. The case $\ell = \exp(\beta)$. Suppose that $\ell = \exp(\beta)$, i.e. ℓ is the exponential distribution with failure rate β. In this case jumps are not bounded with a positive constant M and we cannot adapt the technicalities presented in the proof of Proposition 4.1. More precisely, if $(\pi(\cdot|[\text{data}]_n))$ satisfies the LDP, by the Lundberg's inequality (4.2) we can derive upper bounds for

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,\infty)} P(T_{qn}(\lambda,\ell) < \infty)\pi(d\lambda, d\ell|[\text{data}]_n)$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,\infty)} P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big)\pi(d\lambda, d\ell|[\text{data}]_n)$$

by Varadhan's Lemma and contraction principle. On the other hand we cannot derive lower bounds for

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,\infty)} P(T_{qn}(\lambda,\ell) < \infty)\pi(d\lambda, d\ell|[\text{data}]_n)$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{(0,\infty)\times M_1[0,\infty)} P\Big(T_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big)\pi(d\lambda, d\ell|[\text{data}]_n).$$

We recall that (4.5) provides a closed form expression for the ruin probability $P(T_Q(\lambda,\ell) < \infty)$ when $\ell = \exp(\beta)$. On the other hand, when it is not possible to find a constant $\widetilde{M} > 0$ such that $1 - \frac{w(\lambda,\exp(\beta))}{\beta} \geq \widetilde{M}$ for all (λ,β) in a set of probability 1 with respect to the prior distribution, the lower bounds cannot be derived. For instance the natural (conjugate) choice of products of Gamma distributions on (λ,β) is supported on $(0,\infty)\times(0,\infty)$, but it cannot solve the problem since $\lim_{\beta\to\infty} \frac{w(\lambda,\exp(\beta))}{\beta} = 1$ (for each fixed $\lambda > 0$) and $\lim_{\lambda\to 0} \frac{w(\lambda,\exp(\beta))}{\beta} = 1$ (for each fixed $\beta > 0$).

One way to overcome the problem is to assume $\beta = k\lambda$ for some $k > 0$ in a way that

$$1 - \frac{w(\lambda,\exp(\beta))}{\beta} = 1 - \frac{w(\lambda,\exp(k\lambda))}{k\lambda} = \min\Big\{\frac{1}{kc}, 1\Big\} > 0$$

which does not depend on λ; then the inequality $1 - \frac{w(\lambda,\exp(\beta))}{\beta} \geq \widetilde{M}$ holds as an equality with $\widetilde{M} = \min\Big\{\frac{1}{kc}, 1\Big\}$.

Assuming $k > 0$ to be known, λ becomes the only unknown parameter. Considering the n-sample

$$[\text{data}]_n = ((B_1, T_1), \ldots, (B_n, T_n))$$

the likelihood becomes

$$L(\lambda|[\text{data}]_n) = k\lambda^{-k\lambda B_1}\lambda e^{-\lambda T_1} \cdots k\lambda^{-k\lambda B_n}\lambda e^{-\lambda T_n} = k^n\lambda^{2n}e^{-\lambda n[k\overline{B}_n + \overline{T}_n]},$$

where $\overline{B}_n = \frac{1}{n}\sum_{k=1}^n B_k$ and $\overline{T}_n = \frac{1}{n}\sum_{k=1}^n T_k$ are the empirical means of $\{B_1, \ldots, B_n\}$ and of $\{T_1, \ldots, T_n\}$ respectively; thus $k\overline{B}_n + \overline{T}_n$ is a sufficient statistics.
With the conjugate $G_{\alpha,\beta}$ prior distribution on λ the posterior distribution becomes

$$G_{\alpha,\beta}(\cdot|[\text{data}]_n) = G_{\alpha_n,\beta_n},$$

where $\alpha_n = \alpha + 2n$ and $\beta_n = \beta + n(k\overline{B}_n + \overline{T}_n)$.

Let us prove the LDP for posterior distributions.

LEMMA 4.3. *Let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$ on λ. Assume $k\overline{B}_n + \overline{T}_n \to \frac{2}{\widehat{\lambda}}$ as $n \to \infty$. Then $(\pi(\cdot|[\text{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\lambda}}$ defined by*

$$I^*_{\widehat{\lambda}}(\lambda) = \sup_{\gamma\in\mathbb{R}}[\gamma\lambda - \Lambda_{\widehat{\lambda}}(\gamma)] = \begin{cases} H\left(G_{2,\widehat{\lambda}}|G_{2,\lambda}\right) & \text{if } \lambda > 0, \\ \infty & \text{if } \lambda \leq 0. \end{cases}$$

PROOF. Let $\gamma \in \mathbb{R}$ be arbitrarily fixed. Then, since $k\overline{B}_n + \overline{T}_n \to \frac{2}{\widehat{\lambda}}$ as $n \to \infty$, for any conjugate prior distribution $G_{\alpha,\beta}$ we have

$$\Lambda_{\widehat{\lambda}}(\gamma) = \lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\lambda}G_{\alpha,\beta}(d\lambda|[\text{data}]_n) =$$

$$= \lim_{n\to\infty}\begin{cases} \frac{\alpha+2n}{n}\log\left(\frac{\beta+n(k\overline{B}_n+\overline{T}_n)}{\beta+n(k\overline{B}_n+\overline{T}_n)-n\gamma}\right) & \text{if } n\gamma < \beta + n(k\overline{B}_n + \overline{T}_n) \\ \infty & \text{if } n\gamma \geq \beta + n(k\overline{B}_n + \overline{T}_n) \end{cases}$$

$$= \begin{cases} 2\left(\frac{\frac{2}{\widehat{\lambda}}}{\frac{2}{\widehat{\lambda}}-\gamma}\right) & \text{if } \gamma < \frac{2}{\widehat{\lambda}}, \\ \infty & \text{if } \gamma \geq \frac{2}{\widehat{\lambda}}. \end{cases}$$

Furthermore

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty e^{n\gamma\lambda}\pi(d\lambda|[\text{data}]_n) = \Lambda_{\widehat{\lambda}}(\gamma)$$

by Lemma 2.1. Then, by Gärtner Ellis Theorem, $(\pi(\cdot|[\mathrm{data}]_n))$ satisfies the LDP with good rate function $I_{\widehat{\lambda}} = \Lambda^*_{\widehat{\lambda}}$ defined by

$$\Lambda^*_{\widehat{\lambda}}(\lambda) = \sup_{\gamma\in\mathbb{R}}[\gamma\lambda - \Lambda_{\widehat{\lambda}}(\gamma)] = \begin{cases} H\left(G_{2,\widehat{\lambda}}|G_{2,\lambda}\right) & \text{if } \lambda > 0, \\ \infty & \text{if } \lambda \leq 0, \end{cases}$$

where the latter equality follows from (2.3) with $\alpha = 2$ and $\beta = \frac{2}{\widehat{\lambda}}$. □

We conclude this subsection with the estimation of level crossing probabilities for which the following function (of $q > 0$) is useful:

$$-[\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q) = \sup_{\lambda>0}\left[-qw(\lambda, \exp(k\lambda)) - H\left(G_{2,\widehat{\lambda}}|G_{2,\lambda}\right)\right].$$

PROPOSITION 4.2. *Let $k \geq 1$, $p_1, \ldots, p_k \geq 0$ such that $\sum_{i=1}^k p_i = 1$ and let us consider the prior distribution $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$. Assume $k\overline{B}_n + \overline{T}_n \to \frac{2}{\widehat{\lambda}}$ as $n \to \infty$. Then, for all $q > 0$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P(T_{qn}(\lambda, \exp(k\lambda)) < \infty)\pi(d\lambda|[\mathrm{data}]_n) = -[\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q)$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^\infty P(T_q(n\lambda, \exp(nk\lambda)) < \infty)\pi(d\lambda|[\mathrm{data}]_n) = -[\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q).$$

PROOF. Set $\widetilde{M} = \min\left\{\frac{1}{kc}, 1\right\}$ and let $n \geq 1$ be arbitrarily fixed. Then, by (4.5), we have

$$\widetilde{M}e^{-w(\lambda,\exp(k\lambda))qn} \leq P(T_{qn}(\lambda, \exp(k\lambda)) < \infty) \leq e^{-w(\lambda,\exp(k\lambda))qn}$$

and

$$\widetilde{M}e^{-nw(\lambda,\exp(k\lambda))q} \leq P(T_q(n\lambda, \exp(nk\lambda)) < \infty) \leq e^{-nw(\lambda,\exp(k\lambda))q};$$

for the latter we take into account $w(n\lambda, \exp(nk\lambda)) = nw(\lambda, \exp(k\lambda))$.
In conclusion we obtain

$$\begin{aligned}
&- [\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q) \\
&\quad \leq \liminf_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_{qn}(\lambda, \exp(k\lambda)) < \infty)\pi(d\lambda|[\text{data}]_n) \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_{qn}(\lambda, \exp(k\lambda)) < \infty)\pi(d\lambda, |[\text{data}]_n) \\
&\quad \leq - [\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q)
\end{aligned}$$

and

$$\begin{aligned}
&- [\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q) \\
&\quad \leq \liminf_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_q(n\lambda, \exp(nk\lambda)) < \infty)\pi(d\lambda|[\text{data}]_n) \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{n} \log \int_0^\infty P(T_q(n\lambda, \exp(nk\lambda)) < \infty)\pi(d\lambda, |[\text{data}]_n) \\
&\quad \leq - [\overline{w}(\widehat{\lambda}, \exp(k\widehat{\lambda}))](q)
\end{aligned}$$

by Lemma 4.3 and by Varadhan's Lemma applied to the continuous function $\lambda \mapsto w(\lambda, \exp(k\lambda))$. □

5 Bayesian Versus Frequentist: Some Typical Features

In this section we underline some typical features that distinguish Bayesian from frequentist results when LDP and estimation of level crossing probabilities are concerned.

Starting from the LDP it is worth noting that the rate functions for posterior distributions can be expressed in terms of the same relative entropy $H(\cdot|\cdot)$ used for the rate function concerning a suitable sequence of sufficient statistics $(T_n([\text{data}]_n))$. In the latter case, however, the roles played by the arguments of $H(\cdot|\cdot)$ are interchanged. This result can be explained with the different role played by the parameter θ and by the data in the Bayesian and frequentist approaches with respect to a suitable statistical model $(X, \mathcal{B}_X, (F(\theta) : \theta \in \Theta))$. In fact, in the LDP of the sufficient statistics we ask how likely it is for the sufficient statistics to be close to some $\widehat{\theta}$ when the true value of the parameter is θ and the rate function is of the form $\widehat{\theta} \mapsto I_\theta(\widehat{\theta}) = H(F(\widehat{\theta})|F(\theta))$. On the other hand, in the LDP of posterior distributions we ask how likely it is for the true value of the parameter to be close to θ given that we observe the sufficient statistics close to some $\widehat{\theta}$ and the rate function is of the form $\theta \mapsto I_{\widehat{\theta}}(\theta) = H(F(\widehat{\theta})|F(\theta))$.

Let us denote the interior and the closure of C by C° and $\overline{C}$ respectively. Then we have

$$\lim_{n\to\infty} P_\theta(T_n([\text{data}]_n) \in C) = 0 \quad \text{if} \quad \theta \notin \overline{C}$$

(where P_θ corresponds to the distribution function $F(\theta)$) and

$$\lim_{n\to\infty} \pi(C|[\text{data}]_n) = 0 \quad \text{if} \quad \widehat{\theta} \notin \overline{C}.$$

By the LDPs of sufficient statistics and of posterior distributions we can say that $P_\theta(T_n([\text{data}]_n) \in C)$ and $\pi(C|[\text{data}]_n)$ go to zero exponentially with n (as $n \to \infty$) under suitable hypotheses.

- Assume $I_\theta(C) = \inf_{\widehat{\theta}\in C} I_\theta(\widehat{\theta})$ and $I_\theta(C) = I_\theta(C^\circ) = I_\theta(\overline{C}) \in (0,\infty)$ (we remark that $I_\theta(C) > 0$ since $\theta \notin \overline{C}$). Then, for all $\varepsilon > 0$, there exists $n_\varepsilon \geq 1$ such that for all $n \geq n_\varepsilon$:

$$e^{-n(I_\theta(C)+\varepsilon)} < P_\theta(T_n([\text{data}]_n) \in C) < e^{-n(I_\theta(C)-\varepsilon)}.$$

- Assume $I_{\widehat{\theta}}(C) = \inf_{\theta\in C} I_{\widehat{\theta}}(\theta)$ and $I_{\widehat{\theta}}(C) = I_{\widehat{\theta}}(C^\circ) = I_{\widehat{\theta}}(\overline{C}) \in (0,\infty)$ (we remark that $I_{\widehat{\theta}}(C) > 0$ since $\widehat{\theta} \notin \overline{C}$). Then, for all $\varepsilon > 0$, there exists $n_\varepsilon \geq 1$ such that for all $n \geq n_\varepsilon$:

$$e^{-n(I_{\widehat{\theta}}(C)+\varepsilon)} < \pi(C|[\text{data}]_n) < e^{-n(I_{\widehat{\theta}}(C)-\varepsilon)}.$$

The second difference is related to the level crossing probabilities estimation. To see this, it is important to recall the following limits as $n \to \infty$, necessary to 'define' the frequentist estimation of level crossing probabilities. Let $q > 0$ be arbitrarily fixed, then:

- for the Brownian motion case we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log P(T_{qn}(m, r^{-1}) < \infty) &= \lim_{n\to\infty} \frac{1}{n} \log P(T_q(m, (nr)^{-1}) < \infty)\\ &= -qw(m, r^{-1})\end{aligned}$$

  by (3.1);

- for the compound Poisson process case, with ℓ concentrated on $[0, M]$ for some $M > 0$ we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log P(T_{qn}(\lambda,\ell)<\infty) &= \lim_{n\to\infty}\frac{1}{n}\log P\Big(T_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big)\\ &= -qw(\lambda,\ell)\end{aligned}$$

  by (4.4), since $\varepsilon > 0$ is arbitrary;

- for the compound Poisson process case, with $\ell = \exp(\beta)$ we have

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P(T_{qn}(\lambda,\exp(\beta))<\infty)\\ &=\lim_{n\to\infty}\frac{1}{n}\log P\Big(T_q\Big(n\lambda,\exp(n\beta)\Big)<\infty\Big)\\ &=-\,qw(\lambda,\exp(\beta))\end{aligned}$$

  by (4.5), where for the second limit we consider the identity

$$w(n\lambda,\exp(nk\lambda)) = nw(\lambda,\exp(k\lambda)).$$

It is worth noting that in all the cases considered the limits have the same expression $-qw(\theta)$, with

$$w(\theta)=\sup\Big\{\gamma\geq 0:\log\mathbb{E}\Big[e^{\gamma Z_1^\theta}\Big]\leq 0\Big\},$$

being a linear function of $q > 0$ for each fixed possible value of the parameter $\theta \in \Theta$.

The frequentist estimates of the above limits are $-qw(\widehat{\theta})$ (when we plug $\widehat{\theta}$ into $-qw(\theta)$) where $\widehat{\theta}$ is the almost sure limit of the sufficient statistics. In the Bayesian framework considered in sections 3 and 4, the estimates are

$$-[\overline{w}(\widehat{\theta})](q)=\sup_{\theta\in\Theta}[-qw(\theta)-H(F(\widehat{\theta})|F(\theta))], \tag{5.1}$$

where $(X, \mathcal{B}_X, (F(\theta) : \theta \in \Theta))$ is the above statistical model.

Comparing $-[\overline{w}(\widehat{\theta})](q)$ with $-qw(\widehat{\theta})$ it is easy to check that $-[\overline{w}(\widehat{\theta})](q) - (-qw(\widehat{\theta}))$ is nonnegative by (5.1) and is nondecreasing function of $q > 0$ (see Proposition 4 of Macci, 2004). This consideration leads to point out that the Bayesian estimates of level crossing probabilities are asymptotically guaranteed to be more conservative to a degree which becomes more pronounced as q increases.

Appendix: Gärtner Ellis Theorem and Varadhan's Lemma

As in subsection 2.1 let Ω be a Hausdorff topological space with Borel σ-algebra $\mathcal{B}_\Omega$ and let (ν_n) be sequence of probability measures on $(\Omega, \mathcal{B}_\Omega)$.

GÄRTNER ELLIS THEOREM. *Let us consider* $(\Omega, \mathcal{B}_\Omega) = (\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ *and assume that, for all* $\boldsymbol{\gamma} \in \mathbb{R}^m$, *there exists the limit*

$$\Lambda(\boldsymbol{\gamma}) = \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^m} e^{n\boldsymbol{\gamma}'\boldsymbol{x}} \nu_n(d\boldsymbol{x})$$

as an extended real number. Then, if Λ *is an essentially smooth and lower semicontinuous function, then the LDP holds with the good rate function* $\Lambda^* : \mathbb{R}^m \to [0, \infty]$ *defined by*

$$\Lambda^*(\boldsymbol{x}) = \sup_{\boldsymbol{\gamma}\in\mathbb{R}^m} [\boldsymbol{\gamma}'\boldsymbol{x} - \Lambda(\boldsymbol{\gamma})].$$

For completeness we also recall the definition of essentially smooth function. Let $\Lambda : \mathbb{R}^m \to (-\infty, \infty]$ be a convex function, let $\mathcal{D}_\Lambda$ be the set

$$\mathcal{D}_\Lambda = \{\boldsymbol{\gamma} \in \mathbb{R}^m : \Lambda(\boldsymbol{\gamma}) < \infty\}$$

and let $\mathcal{D}_\Lambda^\circ$ be the interior of $\mathcal{D}_\Lambda$. Then Λ is an essentially smooth function if: (i) $\mathcal{D}_\Lambda^\circ$ is non-empty; (ii) Λ is differentiable throughout $\mathcal{D}_\Lambda^\circ$; the function Λ is steep, i.e. $\lim_{n\to\infty} |\nabla\Lambda(\boldsymbol{\gamma}_n)| = \infty$ whenever $(\boldsymbol{\gamma}_n)$ in $\mathcal{D}_\Lambda^\circ$ converging to a boundary point of $\mathcal{D}_\Lambda^\circ$.

LEMMA 5.1 (VARADHAN'S LEMMA). *Assume* (ν_n) *satisfies the LDP with the good rate function* $I : \Omega \to [0, \infty]$ *and let* $\phi : \Omega \to \mathbb{R}$ *be a continuous function. Assume further either the tail condition*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \int_\Omega e^{n\phi(\omega)} 1_{\{\phi(\omega)\geq M\}} \nu_n(d\omega) = -\infty$$

or the following moment condition for some $\gamma > 1$

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_\Omega e^{\gamma n\phi(\omega)} \nu_n(d\omega) < \infty.$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \log \int_\Omega e^{n\phi(\omega)} \nu_n(d\omega) = \sup_{\omega\in\Omega} [\phi(\omega) - I(\omega)].$$

In our applications of Varadhan's Lemma the moment condition holds trivially; indeed ϕ is non-positive and, for all $\gamma \geq 0$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_\Omega e^{\gamma n \phi(\omega)} \nu_n(d\omega) \leq 0.$$

Acknowledgements. We thank A.J. Ganesh for some useful discussion on the proof of Lemma 4.2, G.L. Torrisi for leading us to consider the case with exponentially distributed jumps and Pier Luigi Conti for useful comments. This work has been partially supported by Murst Project "Metodi Stocastici in Finanza Matematica".

References

Asmussen, S. (1987). *Applied Probability and Queues.* John Wiley and Sons, New York.

Asmussen, S. and Nielsen, H.M. (1995). Ruin probabilities via local adjustment coefficients. *J. Appl. Probab.*, **33**, 736–755.

Brunner, L.J. and Lo, A.Y. (1996). Limiting posterior distributions under mixture of conjugate priors. *Statist. Sinica*, **33**, 187–197.

Dalal, S. and Hall, W.J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. Roy. Statist. Soc. Ser. B.*, **45**, 278–286.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* Mc Graw-Hill, New York.

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

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In *Bayesian Statistics 2*, (J.M. Bernardo et al., eds.), 133–156. North-Holland: Amsterdam.

Eichelsbacher, P. and Ganesh, A. (2002). Bayesian inference for Markov chains. *J. Appl. Probab.*, **39**, 91–99.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events.* Springer, Berlin Heidelberg.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

Fu, J.C. and Kass, R.E. (1988). The exponential rates of convergence of posterior distributions. *Ann. Inst. Statist. Math.*, **40**, 683–691.

Ganesh, A., Green, P., O'Connell, N. and Pitts, S. (1998). Bayesian network management. *Queueing Systems Theory Appl.*, **28**, 267–282.

Ganesh, A. and O'Connell, N. (1999). An inverse of Sanov's Theorem. *Statist. Probab. Lett.*, **42**, 201–206.

Ganesh, A. and O'Connell, N. (2000). A large deviation principle for Dirichlet posteriors. *Bernoulli*, **6**, 1021–1034.

Grandell, J. (1991). *Aspects of Risk Theory.* Springer, New York.

Kullback, S. (1959). *Information Theory and Statistics.* John Wiley and Sons, New York.

MACCI, C. (2004). Large deviations estimates for some level crossing probabilities in Bayesian setting. *Rend. Sem. Mat. Univ. Pol. Torino*, **62**, 179–188.

O'HAGAN, A. and FORSTER, J.J. (2004). *Bayesian Inference.* 2nd edition. Volume 2B of "Kendall's Advanced Theory of Statistics". Arnold, London.

PASCHALIDIS, I.C. and VASSILARAS, S. (2001). On the estimation of buffer overflow probabilities from measurements. *IEEE Trans. Inform. Theory*, **47**, 178–191.

RUDIN, W. (1986). *Real and Complex Analysis.* 3rd Edition. McGraw-Hill, New York.

TSAY, J.J. and TSAO, S.A. (2003). Statistical gambler's ruin problem. *Comm. Statist. Theory Methods*, **32**, 1337–1359.

CLAUDIO MACCI
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI ROMA "TOR VERGATA"
VIA DELLA RICERCA SCIENTIFICA
I-00133 ROME, ITALY
E-mail: macci@mat.uniroma2.it

LEA PETRELLA
DIPARTIMENTO DI STUDI GEOECONOMICI
LINGUISTICI, STATISTICI,
STORICI PER L'ANALISI REGIONALE
UNIVERSITÀ DI ROMA "LA SAPIENZA"
VIA DEL CASTRO LAURENZIANO 9
I-00161 ROME, ITALY
E-mail: lea.petrella@uniroma1.it

Paper received May 2005; revised November 2005.