<table>
<thead>
<tr>
<th>項目</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>タイトル</td>
<td>Large deviations for the posterior distributions under conjugate prior distributions</td>
</tr>
<tr>
<td>作者</td>
<td>Shikimi, Takuhisa</td>
</tr>
<tr>
<td>引用</td>
<td>経営と経済, 91(1-2), pp.17-36; 2011</td>
</tr>
<tr>
<td>発行日</td>
<td>2011-09-26</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/28488">http://hdl.handle.net/10069/28488</a></td>
</tr>
</tbody>
</table>

NAOSITE: Nagasaki University’s Academic Output SITE
Large Deviations for the Posterior Distributions under Conjugate Prior Distributions

Takuhisa Shikimi

Abstract
This paper takes up three parametric cases - the normal, Poisson, exponential cases - in order to study a large deviation upper bound for some posterior probability of the unknown parameter when in each case the prior distribution is assumed to be in a conjugate family. The upper bound will be given explicitly in each case.

Keywords: large deviations □posterior distributions □exchangeability □

1 Introduction
Let \( X_0, X_1, \ldots \) be i.i.d. random variables with unknown distribution that belongs to a statistical model \( \mathcal{P}_0 \) is a parameter space, where \( \theta \) is a parameter space. In this paper, we focus on exponential rates of convergence of the posterior distributions in three parametric models - the normal, Poisson and exponential statistical models - when in each case the prior distribution is assumed to be in a conjugate family. There is comparatively little literature on the exponential rate of convergence of posterior distribution. Fu and Kass studies the rate of convergence of posterior distributions in the neighborhood of the mode. In the nonparametric Bayesian framework, Shen and Wasserman studies the rate at which the posterior distribution concentrates
around the true parameter, and Ganesh and O’Connell proves the large deviation principle for posterior distributions given i.i.d. random variables taking values in a finite set.

We will give a large deviation upper bound in an explicit form for posterior probabilities of the event given \( X_0, \ldots, X_n \) in each of the three parametric cases. In all cases, the basic tool to derive the results is the law of large numbers for exchangeable random variables together with the conditional Markov inequality.

2 Constructing the model

Let \( \mathcal{D} \) be a measurable space. A stochastic kernel from \( \mathcal{D} \) to \( \mathbb{R} \), \( \mathcal{D} \subseteq \mathbb{R} \), where \( \mathcal{D} \subseteq \mathbb{R}^n \) is the Borel \( \sigma \)-algebra of \( \mathbb{R}^n \), is a family \( \{ P_0 \} \) of probability measures on \( \mathbb{R} \), indexed by \( \mathcal{D} \), such that for each \( A \subseteq \mathcal{D} \), \( P_0(A) \) is measurable. As is usual, \( \{ P_0 \} \) is referred to as a statistical model. If \( P_0 \) is the \( n \) dimensional product measure \( P_0 \) of \( P_0 \), the infinite product probability measure \( P_0 \) is the unique probability measure on \( \mathcal{D} \) such that

\[
P_0(A_0 \times \cdots \times A_n) = P_0(A_0) \times \cdots \times P_0(A_n)
\]

for all \( n \geq 0 \) and \( A_0, \ldots, A_n \).

Lemma. For each \( n = 0, 1, \ldots \), the family \( \{ P_0 \} \) is a stochastic kernel from \( \mathcal{D} \) to \( \mathcal{D} \).

Proof. We only show that \( \{ P_0 \} \) is a stochastic kernel, since \( \{ P_0 \} \) will be shown to be stochastic kernels in the same manner.
If we define

$$\mathcal{L} = \{ B \in \mathcal{B} \mid \text{if } B \text{ is measurable,} \}
$$

then \( \mathcal{L} \) is a \( \mathcal{D} \)-class containing the \( \mathcal{A} \)-class

$$\mathcal{D} = \{ A_0 \cup \cdots \cup A_n \in \mathcal{A} \cup \mathcal{B} : n \geq 0, \ A_0, \ A_n \in \mathcal{A} \}.$$ 

It follows that \( \mathcal{B} \cup \mathcal{D} \) defines a \( \mathcal{D} \)-class.

For a prior distribution \( \mathcal{P} \) on \( \mathcal{B} \cup \mathcal{B} \), define \( \mathbb{P} \) to be the probability measure on \( \mathcal{B} \cup \mathcal{B} \cup \mathcal{B} \cup \mathcal{B} \cup \mathcal{B} \) satisfying

$$\mathbb{P} \ (U \cup B) = \int_U P(U \cup B) d\mathbb{P}$$

for every \( U \in \mathcal{B} \) and \( B \in \mathcal{B} \cup \mathcal{B} \cup \mathcal{B} \). It is not difficult to show the existence and uniqueness of \( \mathbb{P} \). Now let us introduce the coordinate mappings \( \vartheta \), \( X \) and \( \mathbb{I}_i \) defined by

$$\vartheta \ (x) = \vartheta \ (x) = x,$$

$$X \ (x) = X \ (x) = x,$$

$$\mathbb{I}_i \ (x) = x_i \ (x \geq 0)$$

for \( x = x \cup x \cup x \cup x \cup x \). A random element \( X \) is a sequence of random variables \( X_0 \cup X_0 \cup \cdots \cup X \cup X \cup \cdots \cup X \cup X \) and \( x = x \cup x \cup x \cup x \cup x \). We think of \( \vartheta \) as the unknown parameter, \( X = X \cup X \cup \cdots \cup X \cup X \cup \cdots \cup X \cup X \) a date, where the distribution of \( X_i \) is specified by \( \vartheta \) and \( \vartheta \). By \( \vartheta \cup \vartheta \), \( \vartheta \cup \vartheta \) and \( \vartheta \cup \vartheta \), the parameter \( \vartheta \) has \( \vartheta \) as its distribution

$$\mathbb{P} \ (\vartheta \cup U \cup U) = \mathbb{P} \ (\vartheta \cup U \cup U)$$

the distribution \( \mathbb{P} \cup X \cup dx \cup of X \) is given by the mixture.
the distribution \( \mathbb{P} \otimes X_{0}, \ldots, X_{n} \otimes dx_{0}, \ldots, dx_{n} \) is given by the mixture
\[
\int_{\mathcal{B}} \mathbb{P} \otimes \mathcal{B} \otimes dx_{0}, \ldots, dx_{n}, \mathcal{B} \otimes \mathbb{R}^{n}, \mathcal{B},
\]
and the distribution \( \mathbb{P} \otimes X_{i} \otimes dx_{i} \) of \( X_{i} \) is given by the mixture
\[
\int_{\mathcal{A}} \mathbb{P} \otimes \mathcal{A} \otimes dx_{i}, \mathcal{A} \otimes \mathbb{R}^{n}, \mathcal{A},
\]
In particular, \( X_{0}, X_{1}, \ldots \) are identically distributed but not independent in general under \( \mathbb{P} \). Distributions defined by \( \mathbb{P} \otimes \mathcal{B} \otimes \mathbb{R}^{n} \) and \( \mathbb{P} \otimes \mathcal{A} \otimes \mathbb{R}^{n} \) are called prior predictive distributions of \( X, X_{0}, \ldots, X_{n} \) and \( X_{i} \), respectively.

**Lemma.** The function \( P_{\otimes \mathcal{B}} \), defined on \( \mathbb{R} \otimes \mathcal{B} \otimes \mathbb{R}^{n} \), is a regular conditional distribution for \( X = \otimes X_{0}, X_{1}, \ldots \) given \( \mathcal{G} \). For each \( n \in \mathbb{N} \), the function \( P_{\otimes \mathcal{B}_{n}} \), defined for \( \otimes \mathcal{B}_{n} \otimes \mathbb{R}^{n} \), is a regular conditional distribution of \( \otimes X_{0}, \ldots, X_{n} \) given \( \mathcal{G} \). Moreover, \( P_{\otimes \mathcal{A}} \), defined for \( \otimes \mathcal{A} \) is a regular conditional distribution of \( X_{i} \) given \( \mathcal{G} \) for every \( i \geq n \).

**Proof.** For each \( \mathcal{G} \), \( P_{\otimes \mathcal{B}} \) is a probability measure on \( \mathbb{R} \otimes \mathcal{B} \otimes \mathbb{R}^{n} \). If \( B \otimes \mathcal{B} \otimes \mathbb{R}^{n} \)
\[
\int_{\mathcal{G}} P_{\otimes \mathcal{B}} \otimes dx_{i} = \int_{\mathcal{G}} P_{\otimes \mathcal{B}_{n}} \otimes dx_{i} = \mathbb{P} \otimes dx_{i} \otimes B \otimes \mathcal{B} \otimes \mathcal{B}
\]
Thus, \( P_{\otimes \mathcal{B}} \) is a version of \( \mathbb{P} \otimes X \otimes B \otimes \mathcal{G} \otimes \mathcal{G} \), because \( P_{\otimes \mathcal{B}} \) is \( \mathcal{G} \otimes \mathcal{G} \) measurable as a function of \( \mathcal{G} \) for each \( B \).

Likewise, \( P_{\otimes \mathcal{B}_{n}} \) and \( P_{\otimes \mathcal{A}} \) are regular conditional distributions for \( \otimes X_{0}, \ldots, X_{n} \) and \( X_{i} \) given \( \mathcal{G} \) respectively, since they are \( \mathcal{G} \otimes \mathcal{G} \) measurable and almost surely.
Lemma. The random variables $X_\varnothing, X_\varnothing, \ldots$ are conditionally i.i.d. given $\varnothing$.

Proof. For all $n \geq \varnothing$ and all $A_\varnothing, A_\varnothing, \ldots$,

\[
\mathbb{P} \left[ X_\varnothing \in A_\varnothing, \ldots, X_n \in A_n \mid \varnothing \right] = P_{\varnothing, A_\varnothing} \cdots P_{\varnothing, A_n} \mathbb{P} \left[ A_\varnothing \right] = \cdots = P_{\varnothing, A_n} \mathbb{P} \left[ A_\varnothing \right] \quad \text{a.s.,}
\]

where the first and third equalities follow from Lemma 2. Thus, $X_\varnothing, X_\varnothing, \ldots$ are conditionally independent given $\varnothing$. Since $\mathbb{P} \left[ X_i \in A \mid \varnothing \right] = P_{\varnothing, A} \mathbb{P} \left[ A \right]$ for all $i \geq \varnothing$, $X_\varnothing, X_\varnothing, \ldots$ are conditionally identically distributed.

Rea-valued random variables $Y_\varnothing, Y_\varnothing, \ldots$ are exchangeable if for all $n \geq \varnothing$ and all permutations $\varnothing$ of $\{ \varnothing, \ldots, n \}$

\[
\varnothing Y_\varnothing, \ldots, Y_n \varnothing \overset{\text{d}}{=} \varnothing Y_\varnothing, \ldots, Y_\varnothing, \varnothing.
\]

Here $\overset{\text{d}}{=}$ stands for equality in distribution. de Finetti’s theorem claims that random variables $Y_\varnothing, Y_\varnothing, \ldots$ are conditionally i.i.d. given some sub $\varnothing$-algebra if and only if they are exchangeable. Lemma 2 tells us that $X_\varnothing, X_\varnothing, \ldots$ are exchangeable random variables. See Aldous for an abstract version of de Finetti’s theorem.

In what follows, we assume that $\varnothing$ is a complete separable metric space,
which is referred to as a Polish space. Accordingly, there exists a regular conditional distribution of $\theta$ given $X_0, \ldots, X_n$ for all $n \geq 0$, which is termed a posterior distribution of $\theta$ given $X_0, \ldots, X_n$ and denoted by $\mathbb{P}_\theta(U)$. More precisely, there exists a function $\mathbb{P}_\theta(U)$ on $\Omega$ such that

(1) for each $U \subseteq \Omega$, $\mathbb{P}_\theta(U)$ is a variant of $\mathbb{P}_\theta(U) | X_0, \ldots, X_n$.

Suppose that the statistical model $\mathbb{P}_{\theta_0}(U)$ is dominated by a $\mu$-finite measure $\mu$ on $\mathbb{R}$, $\mathbb{P}_{\theta_0}(U)$ with density function $f_{\theta_0}(x)$, $x \in \mathbb{R}$. We assume that $f_{\theta_0}(x)$ is measurable as a function of $\theta_0$, $x \in \mathbb{R}$. The marginal distribution $\mathbb{P}_{\theta_0}(X_0, \ldots, X_n) = \mu(dx_0, \ldots, dx_n)$ of $X_0, \ldots, X_n$ has the marginal density function

$$f_{n}(x_0, \ldots, x_n) = \prod_{i=0}^{n} f_{\theta_0}(x_i)$$

with respect to $\mu$, the $n$-fold measure of $\mu$, i.e.,

$$\mathbb{P}_{\theta_0}(X_0, \ldots, X_n) \in \mathbb{R}$$

This can be seen from

$$\mathbb{P}_{\theta_0}(X_0 \in A_0, \ldots, X_n \in A_n) = \prod_{A_0} \mathbb{P}_{\theta_0}(A_0) \cdots \prod_{A_n} \mathbb{P}_{\theta_0}(A_n)$$

$$= \prod_{A_0} \mathbb{P}_{\theta_0}(A_0) \cdots \prod_{A_n} \mathbb{P}_{\theta_0}(A_n)$$

$$= \prod_{A_0} \int_{A_0} f_{\theta_0}(x_0) \, dx_0 \cdots \prod_{A_n} \int_{A_n} f_{\theta_0}(x_n) \, dx_n$$

$$= \prod_{A_0} \int_{A_0} \cdots \int_{A_n} f_{\theta_0}(x_0, \ldots, x_n) \, dx_0 \cdots dx_n$$

$$= \int_{A_0} \cdots \int_{A_n} f_{\theta_0}(x_0, \ldots, x_n) \, dx_0 \cdots dx_n$$

$$= \int_{A_0} \cdots \int_{A_n} f_{n}(x_0, \ldots, x_n) \, dx_0 \cdots dx_n.$$
Note that \( \mathbb{P} \{ \mathcal{G}_n \cap \mathcal{B}_n \} = 0 \)

**Lemma.** If the statistical model \( \mathcal{P} \triangleq \{ \mathcal{P}_n \} \) is dominated by a \( \mathcal{M} \)-finite measure \( \mathcal{Z} \) on \( \mathcal{M} \), with density \( f \) on \( \mathcal{M} \), a measurable function on \( \mathcal{M} \), then

\[
\mathcal{G}_n \int \mathcal{U} = \nabla f \frac{\mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}}{f_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}
\]

is a posterior distribution of \( \mathcal{G} \) given \( \mathcal{G}_n \), \( \mathcal{B}_n \)

**Proof.** It is easily seen that for each \( \mathcal{G} \), \( \mathcal{G}_n \cap \mathcal{B}_n \) is a probability measure on \( \mathcal{M} \) and that for each \( \mathcal{U} \), \( \mathcal{G}_n \cap \mathcal{B}_n \) is \( \mathcal{M} \)-measurable. Thus it suffices to show that \( \mathcal{G}_n \int \mathcal{U} = \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \) a.s. and this can be shown in the following way

\[
\mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} = \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}
\]

\[
= \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}
\]

\[
= \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}
\]

\[
= \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \int \mathcal{U} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \} \mathcal{P}_n \{ \mathcal{G}_n \cap \mathcal{B}_n \}
\]
3. The large deviation principle

Let $S$ be a Polish space equipped with the Borel $\mathcal{B}$-algebra $\mathcal{B}(S)$. A function $I : S \to \mathbb{R}$ is a rate function if for each $M < \infty$ the level set $\{x \in S : I(x) \leq M\}$ is a compact subset of $S$. A rate function is necessarily a lower semicontinuous function, a function with closed level sets. A family $\{Q_n\}$ of probability measures on $S$ is defined to satisfy the large deviation principle with rate function $I$ if for each closed $F \subseteq S$

$$\limsup_{n \to \infty} \frac{1}{n} \log Q_n[F] \leq -\inf_{x \in F} I(x)$$

and for each open $G \subseteq S$

$$\liminf_{n \to \infty} \frac{1}{n} \log Q_n[G] \geq -\inf_{x \in G} I(x)$$

Large deviation theory focuses on probability measures $Q_n$ for which $Q_n[A]$ converges to exponentially fast for a class of events $A$. The exponential decay of $Q_n[A]$ is characterized in terms of a rate function defined above. General treatments of the theory of large deviations and a wide variety of applications may be found in Dembo and Zeitouni \cite{ref}, Deuschel and Stroock \cite{ref}.

In analogous way, let us define the large deviation principle for regular conditional distributions. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathbb{F}_n$ a filtration of sub $\mathcal{B}$-algebras. We define a function $I : \Omega \to S$ to be a rate function if for each $\omega \in \Omega$, $I(\omega) : S \to \mathbb{R}$ is a rate function on $S$. [Definition]

Suppose that $Q_n[\mathbb{B}]$, $n \geq 1$ is a family of regular conditional distributions for a random variable taking values in $S$ given $\mathcal{F}_n$. We say that $Q_n[\mathbb{B}]$, $n \geq 1$ satisfies the large deviation principle if for each closed set $F$ of $S$
Large Deviations for the Posterior Distributions under Conjugate Prior Distributions

\[
\limsup_{n \to \infty} \frac{1}{n} \log Q_n^{\mathcal{F}} - \inf_{x \in \mathcal{F}} I(x) \quad \text{a.s.}
\]

and for each open set \( G \) of \( S \)

\[
\liminf_{n \to \infty} \frac{1}{n} \log Q_n^G - \inf_{x \in G} I(x) \quad \text{a.s.}
\]

In this paper we restrict ourselves to the analysis on the large deviation upper bound \( \limsup \) for the posterior distributions of \( \vartheta \) given \( X_0, \ldots, X_n \). We will examine the posterior distributions \( Q_n^{\vartheta} \) given \( X_0, \ldots, X_n \) in the normal, Poisson and exponential cases and give a large deviation upper bound \( \limsup \) explicitly for the posterior probability of the closed set \( \vartheta \) in each case.

4. The normal case

Suppose that

\[
P_\vartheta \mathcal{U}x = f \mathcal{U}x \frac{1}{\sqrt{2\pi}} \exp \left( - \frac{1}{2} \frac{x - \mu}{\sigma} \right) \mathcal{U}x = \mathcal{U}x, \quad \sigma > 0, \quad \mu, \sigma \in \mathbb{R}
\]

and assume that the prior distribution for the normal mean \( \vartheta \) is a conjugate distribution

\[
P_\vartheta \mathcal{U}x = \frac{1}{\sqrt{2\pi} \sigma} \exp \left( - \frac{1}{2} \frac{x - \mu}{\sigma} \right) \mathcal{U}x, \quad \sigma > 0, \quad \mu, \sigma \in \mathbb{R}
\]

It follows from Lemma \( \Box \) that the posterior distribution of \( \vartheta \) given \( X_0, \ldots, X_n \) is given by

\[
P_n^{\vartheta} \mathcal{U}x = \frac{1}{\sqrt{2\pi} \sigma_n} \exp \left( - \frac{1}{2} \frac{x - \mu_n}{\sigma_n} \right) \mathcal{U}x
\]

where \( \sigma_n = \sigma_n \mathcal{U}x, x_n \mathcal{U} \) and \( \sigma_n \mathcal{U} \) are defined by

\[
x_n = \frac{x_0 + \cdots + x_n}{n},
\]
\[ I_n \leq \frac{1}{n} \] is the rate function.

**Theorem 1.** For each \( n \) \( i \in \mathbb{N} \)

\[
\limsup_{n \to \infty} - \frac{1}{n} \log I_n \leq - \frac{1}{n} - \frac{1}{n} \quad \text{on} \quad \{ I_n > 9 \} \quad \text{a.s.}
\]

**Proof.** By Markov’s inequality for conditional expectations, for all \( t > 0 \)

\[
\mathbb{P} \left( \frac{1}{n} \log I_n \geq t \right) \leq e^{-nt} \mathbb{E} \left( \frac{1}{n} \log I_n \right) \leq e^{-nt} \mathbb{E} \left( \frac{1}{n} \log I_n \right) \]

so that

\[
\limsup_{n \to \infty} - \frac{1}{n} \log I_n \leq - t + \mathbb{E} \left( \frac{1}{n} \log I_n \right). \]

Since \( \mathbb{E} \left( \frac{1}{n} \log I_n \right) \leq \mathbb{E} \left( \frac{1}{n} \log I_n \right) \) \( \text{a.s.} \) by Theorem A and Lemma A, we have

\[
\limsup_{n \to \infty} - \frac{1}{n} \log I_n \leq - t + \mathbb{E} \left( \frac{1}{n} \log I_n \right) \]

Since \( t > 0 \) is arbitrary

\[
\limsup_{n \to \infty} - \frac{1}{n} \log I_n \leq \inf_{t > 0} \left[ - t + \mathbb{E} \left( \frac{1}{n} \log I_n \right) \right] = - \frac{1}{n} - \frac{1}{n} \quad \text{on} \quad \{ I_n > 9 \} \quad \text{a.s.}
\]

In the same manner, it follows that

\[
\limsup_{n \to \infty} - \frac{1}{n} \log I_n \leq - \frac{1}{n} - \frac{1}{n} \quad \text{on} \quad \{ I_n < 9 \} \quad \text{a.s.}
\]

In Theorem 1 the rate function \( I \) \( i \in \mathbb{N} \), \( \mathbb{N} \) \( i \in \mathbb{N} \) is
Large Deviations for the Posterior Distributions under Conjugate Prior Distributions

\[ I[\theta, \phi] = \frac{e^{\theta \phi - f(\|x\|;\phi)} }{\int e^{\theta \phi - f(\|x\|;\phi)} \, dx} = K[\theta, \phi, \phi], \]

where \( K[\theta, \phi, \phi] \) is the Kullback-Leibler distance

\[ K[\theta, \phi, \phi] = \int \log \frac{f(x; \theta, \phi)}{f(x; \theta, \phi)} \, dx = \frac{1}{\theta} - \frac{1}{\phi}. \]

If \( \theta > \phi, \) then

\[ \inf_{\theta > \phi} I[\theta, \phi, \phi] = \inf_{\theta > \phi} I[\theta, \phi, \phi]. \]

and so the large deviation upper bound inequality \( \inf \) is rewritten by using the rate function \( I[\theta, \phi, \phi] \) as

\[ \lim sup_{n \to \infty} \frac{1}{n} \log N \sum_{i=1}^{n} \inf_{\theta > \phi} I[\theta, \phi, \phi] \quad \text{on} \quad \{ \theta > \phi \} \quad \text{a.s.} \]

We now turn to the case where the samples are observed from the normal distribution with mean \( \mu \) and unknown precision. A precision is the reciprocal of the variance. Accordingly, we assume that

\[ P[dx] = \frac{e^{-\frac{1}{2n} (x - \bar{x})^2}}{\sqrt{2\pi n}} \exp \left[ -\frac{1}{\theta} \frac{1}{\phi} \frac{1}{\theta} dx \right], \quad \theta > \phi > 0, \]

If the prior distribution \( \phi \) is specified by

\[ \phi[dx] = \frac{e^{-\frac{1}{\theta} (x - \phi)^2}}{\sqrt{2\pi \theta}} \exp \left[ -\frac{1}{\theta} \frac{1}{\phi} \frac{1}{\theta} dx \right], \quad \theta > \phi > 0, \]

which is a gamma distribution with parameters \( \theta \) and \( \phi > \phi > \phi, \) then the posterior distribution of \( \phi \) given \( X_\theta, \ldots, X_n \) is a gamma distribution with parameters

\[ \phi_n = \phi + \frac{n}{\theta} \quad \text{and} \quad \phi_n = \phi_n \phi X_\theta, \ldots, X_n \phi = \phi + \frac{1}{\theta} \sum_{i=1}^{n} X_i. \]

Theorem A \( \theta \) together with Lemma A \( \theta \) entails the convergence

\[ \frac{\phi_n}{n} \Rightarrow \mathbb{E} X^2 \quad \text{a.s.} \]
Theorem. For each $\square > \square$

$$\limsup_{n \to \infty} \frac{n}{\log n} \sum_{i=1}^{n} \sum_{j=1}^{n} \leq - \frac{1}{\log n} - 9 \log \log n$$
on $\{ \square > \square \}$ a.s.

Proof. For almost all $\square \in \{ \square > \square \}$ and $t \in \mathbb{R}^{+}$, there is an $n_0$ such that $\square_n \sum_{i=1}^{n} - nt \triangleright \square$ for all $n \geq n_0$, since

$$\frac{\square_n}{n - nt} \leq \frac{\square_n}{n} - t \leq \frac{\square_n}{n} \cdot - t = \frac{\square}{n} \cdot - \frac{\square}{t} \cdot .$$

By Markov’s inequality

$$\frac{n}{n} \log \frac{n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \leq - t \cdot + \frac{n}{n} \log \frac{\square_n X_{1}, \ldots, X_n}{\square_n - nt} .$$

It follows that

$$\limsup_{n \to \infty} \frac{n}{n} \log \frac{n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \leq - t \cdot + \frac{n}{n} \log \frac{\square_n X_{1}, \ldots, X_n}{\square_n - nt} .$$

for almost all $\square \in \{ \square > \square \}$ and $t \in \mathbb{R}^{+}$. Now we obtain

$$\limsup_{n \to \infty} \frac{n}{n} \log \frac{n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \leq \inf_{t < \sum_{i=1}^{n} \sum_{j=1}^{n}} - t \cdot + \frac{n}{n} \log \frac{\square_n X_{1}, \ldots, X_n}{\square_n - nt} .$$

$$= - \frac{1}{\log n} - 9 \log \log n$$
on $\{ \square > \square \}$ a.s.

5. The Poisson case

Let $\square$ be the counting measure on $\mathbb{R} \supset \mathbb{Q}$ and define $\square \mathbb{A} \supset \square \mathbb{A} \notin \{ \square \mathbb{A} \cup \mathbb{R} \supset \} \supset$, $A \cup \mathbb{Q} \cup \mathbb{A}$. Then $\square$ is a $\square$-finite measure on $\mathbb{R}$, $\mathbb{Q} \supset \mathbb{A}$. If
Large Deviations for the Posterior Distributions under Conjugate Prior Distributions

\[
P_{\theta} \, dx = f_{\theta}(x) \, dx = \frac{e^{\theta x}}{x^{n}} \, dx, \quad x > 0
\]

and the prior distribution \( \theta \) is a gamma distribution with parameters \( \alpha \) and \( \beta \), then the posterior distribution of \( \theta \) given \( X_{1}, \ldots, X_{n} \) is given by a gamma distribution with parameters \( \alpha_{n} = \alpha_{0} + n \sigma_{0}^{2} / \tau^{2}, \beta_{n} = \beta_{0} + n \). Here we define

\[
\alpha_{n} = \alpha_{0} + n \sigma_{0}^{2} / \tau^{2}, \quad \beta_{n} = \beta_{0} + n.
\]

**Theorem**. For each \( \theta \in \Theta \)

\[
\limsup_{n \to \infty} \frac{1}{n} \log \, \frac{\alpha_{n} \beta_{n}^{n}}{\Gamma(n)} \leq - \theta - \theta \log \theta + \theta \log \frac{\theta}{\beta_{n}}
\]
on \{ \theta > \beta \} \ a.s.

**Proof.** For all \( t > \beta \), Markov’s inequality yields

\[
\frac{1}{n} \log \, \frac{\alpha_{n} \beta_{n}^{n}}{\Gamma(n)} \leq e^{nt} \mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^{n} X_{i} > \beta \right]
\]

\[
\leq e^{nt} \mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^{n} X_{i} > t \right] \text{ a.s.}
\]

and hence for all \( t > \beta \)

\[
\limsup_{n \to \infty} \frac{1}{n} \log \, \frac{\alpha_{n} \beta_{n}^{n}}{\Gamma(n)} \leq - t + \lim_{n \to \infty} \frac{\beta_{n} \log \frac{\beta_{n}}{t}}{n}
\]

\[
= - t + \mathbb{E} \left[ \sum_{i=1}^{n} X_{i} \right] \log \frac{\beta_{n}}{t} \text{ a.s.}
\]

Thus on \{ \theta > \beta \}

\[
\limsup_{n \to \infty} \frac{1}{n} \log \, \frac{\alpha_{n} \beta_{n}^{n}}{\Gamma(n)} \leq \inf_{t < \beta} \left( - t + \theta \log \frac{\theta}{\beta_{n}} \right)
\]

\[
= - \beta - \theta \log \theta + \theta \log \frac{\theta}{\beta_{n}} \text{ a.s.}
\]

\[\square\]
6. The exponential case

Suppose that $\mathcal{B}_1 = \mathbb{R}_+$ and that for each $\mathcal{B}_1$

$$P_{\theta} d\xi = \theta e^{-\theta x} dx.$$

If the prior distribution $\mathcal{B}$ is a gamma distribution with parameters $\mathcal{B}$ and $\mathcal{B}_1$, then the posterior distribution given $X_1, \ldots, X_n$ is a gamma distribution with parameters $\mathcal{B}_n$ and $\mathcal{B}_n = \mathcal{B}_n [X_1, \ldots, X_n]$, where

$$\mathcal{B}_n = \mathcal{B} + n, \quad \mathcal{B}_n = \mathcal{B}_n [X_1, \ldots, X_n], \quad x_n \mathcal{B} = \mathcal{B} + \sum_{i=1}^{n} x_i.$$

**Theorem** $\mathfrak{B}$. For each $\mathcal{B}_1$

$$\limsup_{n \to \infty} \frac{\mathcal{B}_n}{\mathcal{B}_n} \leq \mathcal{B} \mathcal{B}_1 + \log \mathcal{B} \mathcal{B}_1 
$$

on $\{ \mathcal{B}_1 > \mathcal{B}_1 \} \text{ a.s.}$

**Proof.** For almost all $\mathcal{B}_1 \mathcal{B}_1 > \mathcal{B}_1 \mathcal{B}_1$ and $t \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1$, there is an $n_0$ such that

$$\mathcal{B}_n [X_1, \ldots, X_n] > \mathcal{B}_n \text{ for all } n \geq n_0,$$

since

$$\mathcal{B}_n [X_1, \ldots, X_n] - nt > \mathcal{B}_n \text{ for all } n \geq n_0.$$

Thus for almost all $\mathcal{B}_1 \mathcal{B}_1 > \mathcal{B}_1 \mathcal{B}_1$ and all $t \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1$

$$\frac{1}{n} \log \mathcal{B}_n [X_1, \ldots, X_n] \leq -t + \frac{1}{n} \log \mathcal{B}_n [X_1, \ldots, X_n] - nt$$

for all $n \geq n_0$, so that for $\mathcal{B}_1 \mathcal{B}_1 > \mathcal{B}_1 \mathcal{B}_1$ and $t \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1$

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathcal{B}_n [X_1, \ldots, X_n] \leq -t + \log \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1 \mathcal{B}_1.$$

Consequently

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathcal{B}_n [X_1, \ldots, X_n] \leq \inf_{t < \mathcal{B}_1 \mathcal{B}_1} -t + \log \mathcal{B}_1 \mathcal{B}_1.$$
\[
= \mathbb{E}[\log|x|] + \log \mathbb{E}|x|.
\]

\[\Box\]

**Appendix**

**Lemma A**. Let \( Y_\omega \) and \( Y_\omega' \) be random variables on \( \Omega \) with values in measurable spaces \( \mathcal{E}_\omega \) and \( \mathcal{E}_{\omega'} \), respectively, and \( \mathcal{G} \) a sub-\( \sigma \)-algebra with respect to which \( Y_\omega \) is measurable. If \( \Pi \) is a regular conditional distribution for \( Y_\omega \) given \( \mathcal{G} \), then for every measurable function \( f : \mathcal{E}_\omega \circ \mathcal{G} \to \mathbb{R} \) such that \( h(Y_\omega \circ \Pi_{\omega} \circ f) \) is \( \mathcal{G} \)-measurable and

\[
\mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] = \int_{\mathcal{E}_\omega} h(Y_\omega \circ \Pi_{\omega} \circ f) \, d\Pi_{\omega} \quad \text{a.s.}
\]

is \( \mathcal{G} \)-measurable and

\[
\mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] = \int_{\mathcal{E}_{\omega'}} h(Y_{\omega'} \circ \Pi_{\omega'} \circ f) \, d\Pi_{\omega'} \quad \text{a.s.}
\]

In other words, \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) is a version of \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \).

**Proof.** If \( h = \mathbb{1}_{A \cap \mathcal{G}} \), \( A \in \mathcal{G} \), then \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) is \( \mathcal{G} \)-measurable and \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) holds. Since

\[
\mathcal{H} = \left\{ A \cap \mathcal{G} : \int_{\mathcal{E}_\omega} h(Y_\omega \circ \Pi_{\omega} \circ f) \, d\Pi_{\omega} \text{ is a version of } \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \right\}
\]

is a \( \mathcal{G} \)-class and \( \mathcal{H} \) contains the \( \mathcal{G} \)-class

\[
\mathcal{D} = \left\{ A \circ \mathcal{G} : A \in \mathcal{G} \right\} = \mathcal{G} \cap \mathcal{G},
\]

\( \mathcal{G} \circ \mathcal{G} \). Thus \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) is a version of \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) whenever \( h \) is an indicator function. By linearity, \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) is a version of \( \mathbb{E}[h(Y_\omega \circ \Pi_{\omega} \circ f)] \) for all simple functions \( h \), and hence for all nonnegative functions by the monotone convergence theorem. For the general case, the result follows by splitting the function into positive and negative parts. \[\Box\]
Let $Y_0, Y_\beta, ...$ be real-valued random variables defined on a probability space $\Omega \ni \mathcal{P}$ and $\mathcal{A}$ a sub $\sigma$-algebra. If for all $n \geq 0$ and $A_0, ..., A_n \ni \mathcal{B} \lor \mathcal{R} \lor \mathcal{G}$

$$\mathbb{P} \ni Y_0 \ni A_0, ..., Y_n \ni A_n \ | \mathcal{G} = \prod_{i=0}^{n} \mathbb{P} \ni Y_i \ni A_i \ | \mathcal{G} \text{ a.s.,}$$

$Y_0, Y_\beta, ...$ are declared conditionally independent given $\mathcal{G}$. If $\mathcal{G} = \mathcal{A} \ni \mathcal{B}$ for some random element $\mathcal{G}, Y_0, Y_\beta, ...$ are called conditionally independent given $\mathcal{A}$. In addition to the conditional independence, if for all $i \geq 0 \mathbb{P} \ni Y_i \ni A \ | \mathcal{G} = \mathbb{P} \ni Y_0 \ni A \ | \mathcal{G}$ a.s., $Y_0, Y_\beta, ...$ are defined to be conditionally independent and identically distributed (abbreviated to conditionally i.i.d. given $\mathcal{G}$). If $Y_0, Y_\beta, ...$ are conditionally i.i.d. and $\varphi$ is a measurable function, then $\varphi \ni Y_0 \lor \mathcal{G}, \varphi \ni Y_\beta \lor \mathcal{G}...$ are conditionally i.i.d.

**Lemma A.** $\mathcal{B}$. If $Y_0, Y_\beta, ...$ are conditionally i.i.d. given $\mathcal{G}$, there exists a regular conditional distribution $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$ such that for each $\mathcal{A} \ni \mathcal{B}$ the coordinate functions $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$ are i.i.d. Moreover, if $Y_0$ is integrable, then $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$ are integrable with respect to $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$ for almost all $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$.

**Proof.** Since $\mathbb{R}^0$ is a Borel space, there is a regular conditional distribution $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$ such that for each $\mathcal{A} \ni \mathcal{B}$ and each $r \ni \mathcal{Q}$ there is a null set $N_i, \mathcal{G}$ such that for each $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G}$

$$\mathbb{P} \ni Y_i \ni r \ni \mathcal{G} \ni \mathbb{P} \ni Y_0 \ni r \ni \mathcal{G},$$

and hence for all $\mathcal{A} \ni \mathcal{B} \ni \mathcal{G} = \bigcup_{i \geq 0} N_i, \mathcal{G}$ and for all $i \geq 0 \mathcal{B} \ni \mathcal{Q}$, we have

$$\mathbb{P} \ni Y_i \ni r \ni \mathcal{G} = \mathbb{P} \ni Y_0 \ni r \ni \mathcal{G}.$$
Large Deviations for the Posterior Distributions under Conjugate Prior Distributions

Since the sets of the form - form a ω-class generating ω, it follows that for each R, and agree as probability measures on R. For each define a measure by

where is any probability measure on R. Now we define a probability measure

for each on R. We will show that is a regular conditional distribution given that satisfies the requirement of the theorem. Since is the infinite-dimensional product measure of with itself, the coordinate functions are necessarily i.i.d. random variables on R for each with distribution

To show that is a regular conditional distribution for given , it suffices to verify that is a version of for each B, since is a probability measure by definition. If A, then

and therefore . Besides outside
Therefore $\emptyset \in \mathcal{A}_0 \circ \cdots \circ A_n \circ R \circ \cdots \circ \emptyset$ is a version of $\mathbb{P} \circ Y \circ A_0 \circ \cdots \circ A_n \circ R \circ \cdots \circ \mathcal{G}$. Note that

$$\mathcal{G} = \mathcal{A}_0 \circ \cdots \circ A_n \circ R \circ \cdots \circ \emptyset \cong \mathcal{A}_0 \circ \mathcal{B} \circ R \circ \emptyset = \emptyset, \ldots, n \emptyset$$

is a $\emptyset$-class that generates $\mathcal{B} \circ R \circ \emptyset$. Since

$$\mathcal{H} = \mathcal{B} \circ \mathcal{B} \circ R \circ \cdots \circ \emptyset \circ \mathcal{B}$$

is a $\emptyset$-class with $\mathcal{G} \circ \mathcal{H}, \mathcal{B} \circ R \circ \cdots \circ \mathcal{H}$. This implies that $\emptyset \in \mathcal{B}$ is a version of $\mathbb{P} \circ Y \circ B \circ \mathcal{G}$ for each $B \in \mathcal{B} \circ R \circ \emptyset$.

Finally by Lemma A

$$\int_{\mathbb{R}^n} |\emptyset, \mathcal{G}, y| \in \mathcal{G}, dy = \int_{\mathbb{R}^n} |\emptyset, \emptyset, y| \in \mathcal{G}, dy \emptyset$$

is $\mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}$ a.s.

The integrability of $Y_0$ entails $\mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}, \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G}$ a.s., and hence the claims follows. This completes the proof.

**Theorem A**. If $Y_0, Y_1, \ldots$ are conditionally i.i.d. random variables given a sub $\emptyset$-algebra $\mathcal{G}$ and if $Y_0$ is integrable, then

$$\frac{Y_0 + \cdots + Y_n}{n} \in \mathbb{E} \circ |\emptyset, Y_0| \circ \mathcal{G} \text{ a.s.}$$

**Proof.** Let $\emptyset \in \mathcal{B} \in \mathcal{B} \circ \mathcal{B} \circ R \circ \cdots \circ \emptyset$ be a regular conditional dis-
tribution for $Y = \bigotimes_{\beta} Y_{\beta} \ldots$ given $\mathcal{G}$ such that the coordinate functions $\otimes_{\beta} Y_{\beta} \ldots$ are i.i.d. random variables on $\otimes_{\beta} \otimes \otimes$ for each $\beta$. We will show that

$$\mathbb{P} \left[ \sup_{n \geq m} | \mathcal{Y}_n - \mathbb{E} \bigotimes_{\beta} Y_{\beta} | \mathcal{G} \right] > 0 \quad \text{if} \quad m \to \infty \quad \text{for all} \quad \beta \to \beta$$

which is equivalent to the convergence $\mathcal{Y}_n \bigotimes_{\beta} \mathbb{E} \bigotimes_{\beta} Y_{\beta} \mathcal{G}$ a.s. as $n \to \infty$. For all $\beta > \beta$

$$\mathbb{P} \left[ \sup_{n \geq m} | \mathcal{Y}_n - \mathbb{E} \bigotimes_{\beta} Y_{\beta} | \mathcal{G} \right] > 0 \quad \text{implies} \quad \mathbb{E} \left[ \sup_{n \geq m} | \mathcal{Y}_n - \mathbb{E} \bigotimes_{\beta} Y_{\beta} | \mathcal{G} \right] > 0 \quad \text{a.s.}$$

The last equation follows from Lemma A. Since $Y_{\beta}$ is assumed to be integrable, Lemma A shows that $\otimes_{\beta} \otimes \otimes \ldots$ are i.i.d. integrable random variables on $\otimes_{\beta} \otimes \otimes \otimes$ for almost all $\beta$. It follows by the strong law of large numbers and Lemma A that

$$\mathbb{E} \otimes_{\beta} \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes = \mathbb{E} \otimes_{\beta} \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes$$

for almost all $\beta$. It follows that

$$\mathbb{E} \otimes_{\beta} \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes \otimes$$

for almost all $\beta$. And now Lemma A is obtained by the dominated convergence theorem.

References

North-Holland, Amsterdam.


