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 probabilitiy 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_1, X_2,... be i.i.d. random variables with unknown distribution that belongs to a statistical model $(P_\theta : \theta \in \Theta)$, where Θ 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 (1988) studies the rate of convergence of posterior distributions in the neighborhood of the mode. In the nonparametric Bayesian framework, Shen and Wasserman (2001) studies the rate at which the posterior distribution concentrates around the true parameter, and Ganesh and O'Connell (1999) 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 $[\theta,\infty)$ given $X_1,\ldots,X_n$ 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 (Theorem A.3) together with the conditional Markov inequality.

2 Constructing the model

Let $(\Theta,\mathscr{U})$ be a measurable space. A stochastic kernel from $(\Theta,\mathscr{U})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, where $\mathscr{B}(\mathbb{R}^n)$ is the Borel σ-algebra of $\mathbb{R}^n$ $(n=1, 2,\ldots, \infty)$, is a family $(P_\theta : \theta\in\Theta)$ of probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ indexed by $\theta\in\Theta$ such that for each $A\in\mathscr{A}$, $\theta\in\Theta\mapsto P_\theta(A)\in[0,1]$ is measurable. As is usual, $(P_\theta : \theta\in\Theta)$ is referred to as a statistical model. If $P_\theta^{(n)}$ is the n dimensional product measure $P_\theta\times\cdots\times P_\theta$, the infinite product probability measure $P_\theta^{(\infty)}=P_\theta\times P_\theta\times\cdots$, $\theta\in\Theta$ is the unique probability measure on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ such that

$$\begin{aligned} P_\theta^{(\infty)}(A_1\times\cdots\times A_n\times\mathbb{R}\times\mathbb{R}\times\cdots) &= P_\theta(A_1)\cdots P_\theta(A_n) \\ &= P_\theta^{(n)}(A_1\times\cdots\times A_n) \end{aligned}$$

for all $n\geqslant 1$ and $A_1,\ldots, A_n\in\mathscr{B}(\mathbb{R})$.

Lemma 1. *For each* $n=1, 2,\ldots, \infty$, *the family* $(P_\theta^{(n)} : \theta\in\Theta)$ *is a stochastic kernel from* $(\Theta,\mathscr{U})$ *to* $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$.

Proof. We only show that $(P_\theta^{(\infty)} : \theta\in\Theta)$ is a stochastic kernel, since $(P_\theta^{(n)} : \theta\in\Theta)$, $1\leqslant n<\infty$ will be shown to be stochastic kernels in the same manner.

If we define

$$\mathscr{L}=\{B\in\mathscr{B}(\mathbb{R}^\infty):\theta\mapsto P_\theta^{(\infty)}(B)\text{ is measurable}\},$$

then $\mathscr{L}$ is a λ-class containing the π-class

$$\mathscr{D}=\{A_1\times\cdots\times A_n\times\mathbb{R}\times\mathbb{R}\times\cdots:n\geqslant 1,\ A_1,\ldots,A_n\in\mathscr{B}(\mathbb{R})\}.$$

It follows that $\mathscr{B}(\mathbb{R}^\infty)=\sigma(\mathscr{D})\subset\mathscr{L}$. □

For a prior distribution π on $(\Theta,\mathscr{U})$, define $\mathbb{P}$ to be the probability measure on $(\Omega,\mathscr{F})\overset{\triangle}{=}(\Theta\times\mathbb{R}^\infty,\ \mathscr{U}\times\mathscr{B}(\mathbb{R}^\infty))$ satisfying

$$\mathbb{P}(U\times B)=\int_U P_\theta^{(\infty)}(B)\pi(d\theta). \tag{1}$$

for every $U\in\mathscr{U}$ and $B\in\mathscr{B}(\mathbb{R}^\infty)$. It is not difficult to show the existence and uniqueness of $\mathbb{P}$. Now let us introduce the coordinate mappings ϑ, X and ξ_i defined by

$$\vartheta(\omega)=\vartheta(\theta,x)=\theta, \tag{2}$$
$$X(\omega)=X(\theta,x)=x, \tag{3}$$
$$\xi_i(x)=x_i\ \ (i\geqslant 1)$$

for $\omega=(\theta,x)\in\Omega$ and $x=(x_i)\in\mathbb{R}^\infty$. A random element X is a sequence of random variables X_1, $\mathrm{X}_2,\ldots$, where $X_i\overset{\triangle}{=}\xi_i(X)$. We think of ϑ as the unknown parameter, $X=(X_1, X_2,\ldots)$ a date, where the distribution of X_i is specified by ϑ. By (2), (3) and (1), the parameter ϑ has π as its distribution:

$$\mathbb{P}(\vartheta\in U)=\pi(U);$$

the distribution $\mathbb{P}(X\in dx)$ of X is given by the mixture

$$\int_{\Theta}\mathbb{P}_{\theta}^{\infty}(B)\pi(d\theta),\ \ B\in\mathscr{B}(\mathbb{R}^{\infty});\tag{4}$$

the distribution $\mathbb{P}((X_1,\ldots, X_n)\in(dx_1,\ldots, dx_n))$ is given by the mixture

$$\int_{\Theta}P_{\theta}^{(n)}(B_n)\pi(d\theta),\ \ B_n\in\mathscr{B}(\mathbb{R}^{\mathrm{n}}),\tag{5}$$

and the distribution $\mathbb{P}(X_i\in dx_i)$ of X_i is given by the mixture

$$\int_{\Theta}P_{\theta}(A)\pi(d\theta),\ \ A\in\mathscr{B}(\mathbb{R}),\tag{6}$$

In particular, $X_1, X_2,\ldots$ are identically distributed (but not independent in general) under $\mathbb{P}$. Distributions defined by (4), (5) and (6) are called prior predictive distributions of X, $(X_1,\ldots, X_n)$ and X_i, respectively.

Lemma 2. *The function $P^{(\infty)}_{\vartheta(\omega)}(B)$, defined on $\Omega\times\mathscr{B}(\mathbb{R}^{\infty})$, is a regular conditional distribution for $X=(X_1, X_2,\ldots)$ given ϑ. For each $n<\infty$, the function $P^{(n)}_{\vartheta(\omega)}(B_n)$, defined for $(\omega,B_n)\in\Omega\times\mathscr{B}(\mathbb{R}^{n})$, is a regular conditional distribution of $(X_1,\ldots, X_n)$ given ϑ. Moreover, $P_{\vartheta(\omega)}(A)$, defined for $(\omega,A)\in\Omega\times\mathscr{B}(\mathbb{R})$, is a regular conditional distribution of X_i given ϑ for every $i\geqslant1$.*

Proof. For each $\omega\in\Omega$, $P^{(\infty)}_{\vartheta(\omega)}$ is a probability measure on $(\mathbb{R}^{\infty}, \mathscr{B}(\mathbb{R}^{\infty}))$. If $B\in\mathscr{B}(\mathbb{R}^{\infty})$

$$\begin{aligned}\int_{\vartheta\in U}P^{(\infty)}_{\vartheta(\omega)}(B)\,\mathbb{P}(d\omega)&=\int_{U}P_{\theta}^{(\infty)}(B)\pi(d\theta)\\&=\mathbb{P}(U\times B)\\&=\mathbb{P}(\vartheta\in U,X\in B).\end{aligned}$$

Thus, $P^{(\infty)}_{\vartheta(\omega)}(B)$ is a version of $\mathbb{P}(X\in B|\upsilon)(\omega)$, because $P^{(\infty)}_{\vartheta(\omega)}(B)$ is $\sigma(\vartheta)$-measurable as a function of ω for each B.

Likewise, $P^{(n)}_{\vartheta(\omega)}(B_n)$ and $P_{\vartheta(\omega)}(A)$ are regular conditional distributions for $(X_1,\ldots, X_n)$ and $X_i(i=1, 2,\ldots)$, respectively given ϑ, since they are $\sigma(\vartheta)$-measurable and almost surely

$$\begin{aligned}\mathbb{P}((X_1,\ldots, X_n)\in B_n\,|\,\vartheta)(\omega) &= \mathbb{P}(X\in B_n\times\mathbb{R}\times\mathbb{R}\times\cdots\,|\,\vartheta)(\omega)\\ &= P^{(\infty)}_{\vartheta(\omega)}(B_n\times\mathbb{R}\times\mathbb{R}\times\cdots)\\ &= P^{(n)}_{\vartheta(\omega)}(B_n),\quad B_n\in\mathscr{B}(\mathbb{R}^n),\\ \mathbb{P}(X_i\in A\,|\,\vartheta)(\omega) &= \mathbb{P}(X\in\mathbb{R}\times\cdots\times\mathbb{R}\times A\times\mathbb{R}\times\cdots\,|\,\vartheta)(\omega)\\ &= P^{(\infty)}_{\vartheta(\omega)}(\mathbb{R}\times\cdots\times\mathbb{R}\times A\times\mathbb{R}\times\cdots)\\ &= P_{\vartheta(\omega)}(A),\quad A\in\mathscr{B}(\mathbb{R}).\end{aligned}$$

□

Lemma 3. *The random variables X_1, X_2,... are conditionally i.i.d. given ϑ.*

Proof. For all $n\geqslant 1$ and all A$_1$,..., A$_n\in\mathscr{B}(\mathbb{R})$

$$\begin{aligned}\mathbb{P}(X_1\in A_1,\ldots, X_n\in A_n\,|\,\vartheta)(\omega) &= P^{(n)}_{\vartheta(\omega)}(A_1\times\cdots\times A_n)\\ &= P_{\vartheta(\omega)}(A_1)\cdots P_{\vartheta(\omega)}(A_n)\\ &= \mathbb{P}(X_1\in A_1\,|\,\vartheta)(\omega)\cdots\mathbb{P}(X_n\in A_n\,|\,\vartheta)(\omega)\ \textit{a.s.},\end{aligned}$$

where the first and third equalities follow from Lemma 2. Thus, X$_1$, X$_2$,... are conditionally independent given ϑ. Since $\mathbb{P}(X_i\in A\,|\,\vartheta)(\omega)=P_{\vartheta(\omega)}(A)=\mathbb{P}(X_1\in A\,|\,\vartheta)(\omega)$ a.s. for all $i\geqslant 1$, X_1, X_2,... are conditionally identically distributed.

Rea-valued random variables Y_1, Y_2,... are exchangeable if for all $n\geqslant 1$ and all permutations τ of $\{1,\ldots, n\}$

$$(Y_1,\ldots, Y_n)\overset{d}{=}(Y_{\tau(1)},\ldots, Y_{\tau(n)}). \tag{7}$$

Here $\overset{d}{=}$ stands for equality in distribution. de Finetti's theorem claims that random variables Y_1, Y_2,... are conditionally i.i.d. given some sub σ-algebra if and only if they are exchangeable. Lemma 3 tells us that X$_1$, X$_2$,... are exchangeable random variables. See Aldous (1982) for an abstract version of de Finetti's theorem.

In what follows, we assume that Θ is a complete seperable metric space, which is referred to as a Polish space. Accordingly, there exists a regular conditional distribution of ϑ given $X_1,\ldots, X_n$ for all $n\geqslant 1$, which is termed a posterior distribution of ϑ given $X_1,\ldots, X_n$ and denoted by $\pi_n^\omega(U)$, $(\omega, U)\in\Omega\times\mathscr{U}$. More precisely, there exists a function $\pi_n^\omega(U)$ on $\Omega\times\mathscr{U}$ such that

(a) for each $\omega\in\Omega$, $\pi_n^\omega(\cdot)$ is a probability measure on (Θ, U);
(b) for each $U\in\mathscr{U}$, $\pi_n^\omega(U)$ is a variant of $\mathbb{P}(\vartheta\in U\,|\,X_1,\ldots, X_n)(\omega)$.

Suppose that the statistical model $(P_\theta : \theta\in\Theta)$ is dominated by a σ-finite measure ν on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with density function $f(x\,|\,\theta)$, $x\in\mathbb{R}$. We assume that $f(x\,|\,\theta)$ is measurable as a function of $(\theta, x)\in\Theta\times\mathbb{R}$. The marginal distribution $\mathbb{P}((X_1,\ldots, X_n)\in(dx_1,\ldots, dx_n))$ of $(X_1,\ldots, X_n)$ has the marginal density function

$$f_n(x_1,\ldots, x_n)=\int_\Theta\prod_{i=1}^n f(x_i\,|\,\theta)\pi(d\theta)$$

with respect to $\nu^{(\mathrm{n})}$ (the n-fold measure of ν), i.e.,

$$\mathbb{P}((X_1,\ldots, X_n)\in B_n)=\int_{B_n} f_n(x_1,\ldots, x_n)\nu^{(n)}(d(x_1,\ldots, dx_n)).$$

This can be seen from

$$\begin{aligned}\mathbb{P}(X_1\in A_1,\ldots, X_n\in A_n) &= \int_\Theta P_\theta^{(n)}(A_1\times\cdots\times A_n)\pi(d\theta)\\ &= \int_\Theta P_\theta(A_1)\cdots P_\theta(A_n)\pi(d\theta)\\ &= \int_\Theta\left[\int_{A_1} f(x_1\,|\,\theta)\nu(dx_1)\cdots\int_{A_1} f(x_n\,|\,\theta)\nu(dx_n)\right]\pi(d\theta)\\ &= \int_\Theta\int_{A_1\times\cdots\times A_n}\prod_{i=1}^n f(x_i\,|\,\theta)\nu^{(n)}(d(x_1,\ldots, x_n))\pi(d\theta)\\ &= \int_{A_1\times\cdots\times A_n}\int_\Theta\prod_{i=1}^n f(x_i\,|\,\theta)\pi(d\theta)\nu^{(n)}(d(x_1,\ldots, x_n))\\ &= \int_{A_1\times\cdots\times A_n} f_n(x_1,\ldots, x_n)\nu^{(n)}(d(x_1,\ldots, x_n)).\end{aligned}$$

Note that $\mathbb{P}(f_n(X_1,\ldots, X_n)=0)=0$.

Lemma 4. *If the statistical model* $(P_\theta:\theta\in\Theta)$ *is dominated by a* σ*-finite measure* ν *on* $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ *with density* $f(x\,|\,\theta)$, *a measurable function on* $\Theta\times\mathbb{R}$, *then*

$$\pi_n^\omega(U)\triangleq\Big[\int_U\frac{\Pi_{i=1}^n f(x_i\,|\,\theta)}{f_n(X_1,\ldots, X_n)}\pi(d\theta)\Big]1_{\{f_n>0\}}(X_1,\ldots, X_n)$$
$$+\pi(U)1_{\{f_n=0\}}(X_1,\ldots, X_n)$$

is a posterior distribution of ϑ *given* $X_1,\ldots, X_n$.

Proof. It is easily seen that for each ω, $\pi_{\mathrm{n}}^\omega(\cdot)$ is a probability measure on $(\Theta, \mathscr{U})$ and that for each $\mathrm{U}\in\mathscr{U}$, $\pi_n^\omega(U)$ *is* $\sigma(X_1,\ldots, X_n)$-measurable. Thus it suffices to show that $\pi_n^\omega(U)=\mathbb{P}(\vartheta\in U\,|\,X_1,\ldots, X_n)(\omega)$ a.s. and this can be shown in the following way:

$$\int_{\{(X_1,\ldots, X_n)\in B_n\}}\pi_n^\omega(U)d\mathbb{P}=\int_{\substack{\{(X_1,\ldots, X_n)\in B_n\}\\ f_n(X_1,\ldots, X_n)>0}}\Big[\int_U\frac{\Pi_{i=1}^n f(X_i\,|\,\theta)}{f_n(X_1,\ldots, X_n)}\pi(d\theta)\Big]d\mathbb{P}$$

$$=\int_U\Big[\int_{\substack{\{(X_1,\ldots, X_n)\in B_n\}\\ f_n(X_1,\ldots, X_n)>0}}\frac{\Pi_{i=1}^n f(X_i\,|\,\theta)}{f_n(X_1,\ldots, X_n)}d\mathbb{P}\Big]\pi(d\theta)$$

$$=\int_U\Big[\int_{B_n\cap\{f_n>0\}}\frac{\Pi_{i=1}^n f(x_i\,|\,\theta)}{f_n(x_1,\ldots, x_n)}f_n(x_1,\ldots, x_n)\nu^{(n)}(d(x_1,\ldots, x_n))\Big]\pi(d\theta)$$

$$=\int_U\Big[\int_{B_n\cap\{f_n>0\}}\prod_{i=1}^n f(x_i\,|\,\theta)\nu^{(n)}(d(x_1,\ldots, x_n))\Big]\pi(d\theta)$$

$$=\int_U P_\theta^{(n)}(B_n\cap\{f_n>0\})\pi(d\theta)$$
$$=\mathbb{P}(\vartheta\in U,\ (X_1,\ldots, X_n)\in B_n,\ f_n(X_1,\ldots, X_n)>0)$$
$$=\mathbb{P}(\vartheta\in U,\ (X_1,\ldots, X_n)\in B_n,\ f_n(X_1,\ldots, X_n)>0)$$
$$+\mathbb{P}(\vartheta\in U,\ (X_1,\ldots, X_n)\in B_n,\ f_n(X_1,\ldots, X_n)=0)$$
$$=\mathbb{P}(\vartheta\in U,\ (X_1,\ldots, X_n)\in B_n).$$

3. The large deviation principle

Let S be a Polish space equipped with the Borel σ-algebra $\mathscr{B}(S)$. A function $I : S \to [0, \infty]$ is a rate function if for each $M < \infty$ the level set $\{x \in S : I(x) \leqslant M\}$ 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 $F \subset S$

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

and for each open $G \subset S$

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

Large deviation theory focuses on probability measures Q_n for which $Q_n(A)$ converges to 0 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 (1998), Deuschel and Stroock (2000).

In analogous way, let us define the large deviation principle for regular conditional distributions. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(\mathscr{F}_n)$ a filtration of sub σ-algebras. We define a function $I : \Omega \times S \to [0, \infty]$ to be a rate function if for each $\omega \in \Omega$, $\mathrm{I}(\omega, \cdot)$ is a rate function on S.

Definition 5. Suppose that $Q_n^\omega(B)$, $n \geqslant 1$ is a family of regular conditional distributions for a random variable taking values in S given $\mathscr{F}_n$. We say that $Q_n^\omega(B)$, $n \geqslant 1$ satisfies the large deviation principle if for each closed set F of S

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n^{\omega}(F) \leqslant -\inf_{x\in F} I(\omega, x) \quad \text{a.s.} \tag{8}$$

and for each open set G of S

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

In this paper we restrict ourselves to the analysis on the large deviation upper bound (8) for the posterior distributions of ϑ given $X_1,\ldots, X_n$. We will examine the posterior distributions π_n^{ω} given $X_1,\ldots, X_n$ in the normal, Poisson and exponential cases and give a large deviation upper bound (8) explicitly for the posterior probability of the closed set $[\theta, \infty)$ in each case.

4．The normal case

Suppose that

$$P_{\theta}(dx) = f(x\,|\,\theta)dx \overset{\triangle}{=} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right)dx, \quad \theta\in\Theta\overset{\triangle}{=}\mathbb{R}$$

and assume that the prior distribution for the normal mean ϑ is a conjugate distribution

$$\pi(d\theta) \overset{\triangle}{=} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right)d\theta, \quad \sigma>0, \quad \mu\in\mathbb{R}$$

It follows from Lemma 4 that the posterior distribution of ϑ given $X_1,\ldots, X_n$ is given by

$$\begin{aligned}\pi_n^{\omega}(d\theta) &= \frac{\Pi_{i=1}^n f(X_i\,|\,\theta)}{f_n(X_1,\ldots, X_n)}\pi(d\theta)\\ &= \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(\mu_n(X_1,\ldots, X_n)-\mu)^2}{2\sigma_n^2}\right] d\theta,\end{aligned}$$

where $\mu_n = \mu_n(x_1,\ldots, x_n)$ and σ_n^2 are defined by

$$\mu_n(x_1,\ldots, x_n) = \left(\frac{1}{1+n\sigma^2}\right)\mu + \left(\frac{n\sigma^2}{1+n\sigma^2}\right)\bar{x}_n, \quad \bar{x}_n = \frac{x_1+\cdots+x_n}{n},$$

$$\sigma_{\mathrm{n}}^2=\frac{\sigma^2}{1+\mathrm{n}\sigma^2}$$

Theorem 6. *For each* $\theta\in\Theta$

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-\frac{(\theta-\vartheta(\omega))^2}{2}\qquad \textit{on}\ \{\omega:\theta>\vartheta(\omega)\}\quad \textit{a.s.}$$

Proof. By Markov's inequality for conditional expectations, for all $t>0$

$$\begin{aligned}\pi_n^{\omega}[\theta,\infty)&=\mathbb{P}(\{\omega':\vartheta(\omega')\geqslant\theta\}\,|\,X_1,\dots,X_n)(\omega)\\&=\mathbb{P}(e^{nt\vartheta(\omega')}:\geqslant e^{nt\theta}\,|\,X_1,\dots,X_n)(\omega)\\&\leqslant e^{-nt\theta}\,\mathbb{E}(e^{nt\vartheta}\,|\,X_1,\dots,X_n)(\omega)\\&=e^{-nt\theta}\exp\left[\mu_n(X_1,\dots,X_n)nt+\frac{\sigma_n^2n^2t^2}{2}\right]\quad\text{a.s.,}\end{aligned}$$

so that

$$\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-t\theta+\mu_n(X_1,\dots,X_n)t+\frac{\sigma_n^2nt^2}{2}.$$

Since $\mu_n(X_1,\dots,X_n)\to\mathbb{E}(X_1\,|\,\vartheta)=\vartheta$ a.s. by Theorem A.3 and Lemma A.1, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-t\theta+\vartheta(\omega)t+\frac{t^2}{2}.$$

Since $t>0$ is arbitrary

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)&\leqslant\inf_{t>0}\left[-t\theta+\vartheta(\omega)t+\frac{t^2}{2}\right]\\&=-\frac{(\theta-\vartheta(\omega))^2}{2}\quad\text{on}\ \{\omega:\theta>\vartheta(\omega)\}\ \text{a.s.}\qquad(9)\end{aligned}$$

□

In the same manner, it follows that

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}(-\infty,\theta]\leqslant-\frac{(\theta-\vartheta(\omega))^2}{2}\quad\text{on}\ \{\omega:\theta<\vartheta(\omega)\}\ \text{a.s.}$$

In Theorem 6 the rate function $I(\omega,\theta')$, $(\omega,\theta')\in\Omega\times\Theta$ is

$$I(\omega,\theta')=\frac{(\theta'-\vartheta(\omega))^2}{2}=K(\vartheta(\omega),\theta'),$$

where $K(\theta_1,\theta_2)$ is the Kullback-Leibler distance

$$K(\theta_1,\theta_2)=\int_{-\infty}^{\infty}\log\frac{f(x\,|\,\theta_1)}{f(x\,|\,\theta_2)}f(x\,|\,\theta_1)dx=\frac{(\theta_1-\theta_2)^2}{2}.$$

If $\theta>\vartheta(\omega)$, then

$$\frac{(\theta-\vartheta(\omega))^2}{2}=\inf_{\theta\geqslant\theta}I(\theta',\omega).$$

and so the large deviation upper bound inequality (9) is rewritten by using the rate function $I(\omega,\theta')$ as

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-\inf_{\theta\geqslant\theta}I(\omega,\theta')\quad\text{on }\{\omega:\theta>\vartheta(\omega)\}\quad\text{a.s.}$$

We now turn to the case where the samples are observed from the normal distribution with mean 0 and unknown precision. A precision is the reciprocal of the variance. Accordingly, we assume that

$$P_\theta(dx)\triangleq\left(\frac{\theta}{2\pi}\right)^{1/2}\exp\left(-\frac{\theta x^2}{2}\right)dx,\quad\theta\in\Theta\triangleq(0;\infty).$$

If the prior distribution π is specified by

$$\pi(d\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}1_{(0;\infty)}d\theta,\quad\alpha>0,\ \beta>0,$$

which is a gamma distribution with parameters α and β ($\alpha>0$, $\beta>0$), then the posterior distribution of ϑ given $X_1,\dots,X_n$ is a gamma distribution with parameters

$$\alpha_n=\alpha+\frac{n}{2}\quad\text{and}\quad\beta_n=\beta_n(X_1,\dots,X_n)=\beta+\frac{1}{2}\sum_{i=1}^{n}X_i^2.$$

Theorem A.3 together with Lemma A.1 entails the convergence

$$\frac{\beta_n}{n}\to\frac{1}{2}\mathbb{E}(X_1^2\,|\,\upsilon)=\frac{1}{2\vartheta}\quad\text{a.s.}$$

Theorem 7. *For each* $\theta>1$

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-\frac{1}{2\vartheta(\omega)}(\theta-1-\vartheta(\omega)\log\theta)$$

on $\{\omega:\theta>\vartheta(\omega)\}$ *a.s.*

Proof. For almost all $\omega\in\{\theta>\vartheta\}$ and $t\in(0,1/2\vartheta(\omega))$, there is an n_0 such that $\beta_n/(\beta_n-nt)>0$ for all $n\geqslant n_0$, since

$$\frac{\beta_n}{\beta_n-nt}=\frac{\beta_n/n}{\beta_n/n-t}\to\frac{1/(2\vartheta(\omega))}{1/(2\vartheta(\omega))-t}=\frac{1}{1-2\vartheta(\omega)t}.$$

By Markov's inequality

$$\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-t\theta+\log\mathbb{E}(e^{nt\theta}\mid X_1,\ldots,X_n)(\omega)$$
$$=-t\theta+\frac{\alpha n}{n}\log\left(\frac{\beta_n(X_1,\ldots,X_n)}{\beta_n(X_1,\ldots,X_n)-nt}\right).$$

It follows that

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-t\theta+\frac{1}{2}\log\left(\frac{1}{1-2\vartheta(\omega)t}\right)$$

for almost all $\omega\in\{\theta>\vartheta\}$ and $t\in(0,1/2\vartheta(\omega))$. Now we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant\inf_{0<t<1/2\vartheta(\omega)}\left[-t\theta+\frac{1}{2}\log\left(\frac{1}{1-2\vartheta(\omega)t}\right)\right]$$
$$=-\frac{1}{2\vartheta(\omega)}(\theta-1-\vartheta(\omega)\log\theta)$$

on $\{\theta>\vartheta\}$ a.s.

5. The Poisson case

Let ν_0 be the counting measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ and define $\nu(A)=\nu_0(A\cap\{0,1,\ldots\})$, $A\in\mathscr{B}(\mathbb{R})$. Then ν is a σ-finite measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$. If

$$P_\theta(dx)=f(x\,|\,\theta)\,\nu(dx)\overset{\triangle}{=}\frac{e^{-\theta}\theta^{x}}{x!}\nu(dx),\quad \theta\in\Theta=(0,\infty)$$

and the prior distribution π is a gamma distribution with parameters α and β, then the posterior distribution of ϑ given $X_1,\dots, X_n$ is given by a gamma distribution with parameters $\alpha_n=\alpha_n(X_1,\dots, X_n)$, β_n. Here we define

$$\alpha_n=\alpha_n(x_1,\dots, x_n)=\alpha+\sum_{i=1}^{n}x_i,\quad \beta_n=\beta+n.$$

Theorem 8. *For each* $\theta\in\Theta$

$$\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)\leqslant-(\theta-\vartheta(\omega))+\vartheta(\omega)\log\frac{\theta}{\vartheta(\omega)}$$

on $\{\omega:\theta>\vartheta(\omega)\}$ *a.s.*

Proof. For all $t\in(0,1)$, Markov's inequality yields

$$\begin{aligned}\pi_n^{\omega}[\theta,\infty)&=\mathbb{P}(\{\omega:\vartheta(\omega)\geqslant\theta\,|X_1,\dots, X_n)(\omega)\\&\leqslant e^{-nt\theta}\,\mathbb{E}(e^{nt\vartheta}\,|\,X_1,\dots, X_n)(\omega)\\&\leqslant e^{-nt\theta}\left(\frac{\beta_n}{\beta_n-nt}\right)^{\alpha_n(X_1,\dots, X_n)}\text{a.s.,}\end{aligned}$$

and hence for all $t\in(0,1)$

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)&\leqslant-\theta t+\lim_{n\to\infty}\frac{\alpha_n(X_1,\dots, X_n)}{n}\log\left(\frac{\beta_n}{\beta_n-nt}\right)\\&=-\theta t+\mathbb{E}(X_1\,|\,\vartheta)(\omega)\log\left(\frac{1}{1-t}\right)\\&=-\theta t+\vartheta(\omega)\log\left(\frac{1}{1-t}\right)\quad\text{a.s.}\end{aligned}$$

Thus on $\{\omega:\theta>\vartheta(\omega)\}$

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\pi_n^{\omega}[\theta,\infty)&\leqslant\inf_{0<t<1}\left[-\theta t+\vartheta(\omega)\log\left(\frac{1}{1-t}\right)\right]\\&=-(\theta-\vartheta(\omega))+\vartheta(\omega)\log\frac{\theta}{\vartheta(\omega)}\quad\text{a.s.}\end{aligned}$$

□

6. The exponential case

Suppose that $\Theta=(0,\infty)$ and that for each $\theta\in\Theta$

$$P_\theta(dx)=\theta e^{-\theta x}1_{(0,\infty)}dx.$$

If the prior distribution π is a gamma distribution with parameters α and β, then the posterior distribution given $X_1,\ldots, X_n$ is a gamma distribution with parameters α_n and $\beta_n=\beta_n(X_1,\ldots, X_n)$, where

$$\alpha_n=\alpha+n,\ \ \beta_n=\beta_n(x_1,\ldots, x_n)=\beta+\sum_{i=1}^n x_i.$$

Theorem 9. *For each* $\theta\in\Theta$

$$\limsup_{n\to\infty}\frac{1}{n}\pi_n^\omega[\theta,\infty)\leqslant 1-\theta\vartheta(\omega)+\log\,(\theta\vartheta(\omega))$$

on $\{\omega:\theta>\vartheta(\omega)\}$ *a.s.*

Proof. For almost all $\omega\in\{\theta>\vartheta\}$ and $t\in(0,\vartheta(\omega))$, there is an n_0 such that $\dfrac{\beta_n(X_1,\ldots, X_n)}{\beta_n(X_1,\ldots, X_n)-nt}>0$ for all $n\geqslant n_0$, since

$$\frac{\beta_n(X_1,\ldots, X_n)}{\beta_n(X_1,\ldots, X_n)-nt}\to\frac{\mathbb{E}(X_1\,|\,\vartheta)(\omega)}{\mathbb{E}(X_1\,|\,\vartheta)(\omega)-t}=\frac{\vartheta(\omega)}{\vartheta(\omega)-t}>0.$$

Thus for almost all $\omega\in\{\theta>\vartheta\}$ and all $t\in(0,\vartheta(\omega))$

$$\frac{1}{n}\log\,\pi_n^\omega[\theta,\infty)\leqslant-\theta t+\frac{\alpha_n}{n}\log\left(\frac{\beta_n(X_1,\ldots, X_n)}{\beta_n(X_1,\ldots, X_n)-nt}\right)$$

for all $n\geqslant n_0$, so that for $\omega\in\{\theta>\vartheta\}$ and $t\in(0,\vartheta(\omega))$

$$\limsup_{n\to\infty}\frac{1}{n}\log\,\pi_n^\omega[\theta,\infty)\leqslant-\theta t+\log\left(\frac{\vartheta(\omega)}{\vartheta(\omega)-\mathrm{t}}\right).$$

Consequently

$$\limsup_{n\to\infty}\frac{1}{n}\log\,\pi_n^\omega[\theta,\infty)\leqslant\inf_{0<t<\vartheta(\omega)}\left[-\theta t+\log\left(\frac{\vartheta(\omega)}{\vartheta(\omega)-t}\right)\right]$$

$$=1-\theta\upsilon(\omega)+\log\,(\theta\upsilon(\omega)).$$

□

Appendix

Lemma A.1. Let Y_1 and Y_2 be random variables on $(\Omega,\mathscr{F},\mathbb{P})$ with values in measurable spaces $(E_1,\mathscr{E}_1)$ and $(E_2,\mathscr{E}_2)$, respectively, and $\mathscr{G}$ a sub-σ-algebra with respect to which Y_2 is measurable. If μ is a regular conditional distribution for Y_1 given $\mathscr{G}$, then for every measurable function $f : E_1\times E_2 \to \mathbb{R}$ such that $h(Y_1, Y_2)\in L^1(\Omega,\mathscr{F},\mathbb{P})$,

$$\int_{E_1} h(y_1, Y_2(\omega))\,\mu(\omega,dy_1) \tag{A.1}$$

is $\mathscr{G}$-measurable and

$$\mathbb{E}(h(Y_1, Y_2)\,|\mathscr{G})(\omega)=\int_{E_1} h(y_1, Y_2(\omega))\,\mu(\omega,dy_1)\quad\text{a.s.} \tag{A.2}$$

In other words, (A.1) is a version of $\mathbb{E}(h(Y_1, Y_2)\,|\mathscr{G})$.

Proof. If $h=1_{A_1\times A_2}$, $A_i\in\mathscr{E}_i$, then (A.1) is $\mathscr{G}$-measurable and (A.2) holds. Since

$$\mathscr{H}=\left\{A\in\mathscr{E}_1\times\mathscr{E}_2 : \int_{E_1}1_A(y_1, Y_2(\omega))\,\mu(\omega,dy_1)\text{ is a version of }\mathbb{E}(1_A(Y_1; Y_2)\,|\mathscr{G})(\omega)\right\}$$

is a λ-class and $\mathscr{H}$ contains the π-class

$$\mathscr{D}=\{A_1\times A_2 : A_i\in\mathscr{E}_i, i=1,2\},$$

$\mathscr{E}_1\times\mathscr{E}_2\subset\mathscr{H}$. Thus (A.1) is a version of $\mathbb{E}(h(Y_1, Y_2)\,|\mathscr{G})$ whenever h is an indicator function. By linearity, (A.1) is a version of $\mathbb{E}(h(Y_1, Y_2)\,|\mathscr{G})$ 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. □

Let $Y_1, Y_2,\ldots$ be real-valued random variables defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and $\mathscr{G}$ a sub σ-algebra. If for all $n\geqslant 1$ and $A_1,\ldots, A_n\in\mathscr{B}(\mathbb{R})$

$$\mathbb{P}(Y_1\in A_1,\ldots, Y_n\in A_n\,|\,\mathscr{G})=\prod_{i=1}^{n}\mathbb{P}(X_i\in A_i\,|\,\mathscr{G})\quad \text{a.s.,}$$

$Y_1, Y_2,\ldots$ are declared conditionally independent given $\mathscr{G}$. If $\mathscr{G}=\sigma(\eta)$ for some random element η, $Y_1, Y_2,\ldots$ are called conditionally independent given η. In addition to the conditional independence, if for all $i\geqslant 1$ $\mathbb{P}(Y_i\in A\,|\,\mathscr{G})=\mathbb{P}(Y_1\in A\,|\,\mathscr{G})$ a.s., $Y_1, Y_2,\ldots$ are defined to be conditionally independent and identically distributed (abbreviated to conditionally i.i.d.) given $\mathscr{G}$. If $Y_1, Y_2,\ldots$ are conditionally i.i.d. and φ is a measurable function, then $\varphi(Y_1), \varphi(Y_2),\ldots$ are conditionally i.i.d.

Lemma A. 2. If $Y_1, Y_2,\ldots$ are conditionally i.i.d. given $\mathscr{G}$, there exists a regular conditional distribution $\mu(\omega,B), (\omega,B)\in\Omega\times\mathscr{B}(\mathbb{R}^\infty)$ for $Y=(Y_1, Y_2,\ldots)$ given $\mathscr{G}$ such that for each $\omega\in\Omega$ the coordinate functions $\xi_1, \xi_2,\ldots$ on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty), \mu(\omega,\cdot))$ are i.i.d. Moreover, if Y_1 is integrable, then $\xi_1, \xi_2,\ldots$ are integrable with respect to $\mu(\omega,\cdot)$ for almost all $\omega\in\Omega$.

Proof. Since $\mathbb{R}^\infty$ is a Borel space, there is a regular conditional distribution $\nu_0(\omega,\mathrm{B})$ for $Y=(Y_1, Y_2,\ldots)$ given $\mathscr{G}$. For each $i\geqslant 1$ and each $r\in\mathbb{Q}$ there is a null set $N_{i,r}\in\mathscr{G}$ such that for each $\omega\notin N_{i,r}$

$$\begin{aligned}\nu_0(\omega,\xi_i\leqslant r)&=\nu_0(\omega,\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,r]\times\mathbb{R}\times\cdots)\\&=\mathbb{P}(Y\in\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,r]\times\mathbb{R}\times\cdots\,|\,\mathscr{G})(\omega)\\&=\mathbb{P}(Y_i\leqslant r\,|\,\mathscr{G})(\omega)=\mathbb{P}(Y_1\leqslant r\,|\,\mathscr{G})(\omega)\\&=\nu_0(\omega,\xi_1\leqslant r),\end{aligned}$$

and hence for all $\omega\notin\mathrm{N}\triangleq\bigcup_{i\geqslant 1,r\in\mathbb{Q}}N_{i,r}$ and for all $i\geqslant 1, r\in\mathbb{Q}$, we have

$$\nu_0(\omega,\xi_i\leqslant r)=\nu_0(\omega,\xi_1\leqslant r).$$

Since the sets of the form $(-\infty, r], r\in\mathbb{Q}$ form a π-class generating $\mathscr{B}(\mathbb{R})$, it follows that for each $\omega\notin N$, $\nu_0(\omega, \xi_i\in\cdot)$ and $\nu_0(\omega, \xi_1\in\cdot)$ agree as probability measures on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$. For each ω define a measure ν^ω by

$$\nu^\omega(\cdot)=\begin{cases}\nu_0(\omega, \xi_1\in\cdot), & \omega\notin N\\ \nu(\cdot), & \omega\in N,\end{cases}$$

where ν is any probability measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$. Now we define a probability measure

$$\mu(\omega, \cdot)=(\nu^\omega\times\nu^\omega\times\cdots)(\cdot)$$

for each $\omega\in\Omega$ on $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$. We will show that μ is a regular conditional distribution given $\mathscr{G}$ that satisfies the requirement of the theorem. Since $\mu(\omega, \cdot)$ is the infinite-dimensional product measure of ν^ω with itself, the coordinate functions $\xi_1, \xi_2,\ldots$ are necessarily i.i.d. random variables on $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty), \mu(\omega, \cdot))$ for each ω with distribution

$$\begin{aligned}\mu(\omega, \xi_i\in A)&=\nu^\omega(A)\\ &=\begin{cases}\nu_0(\omega, \xi_1\in A), & \omega\notin N\\ \nu(A), & \omega\in N,\end{cases}\quad A\in\mathscr{B}(\mathbb{R}).\end{aligned}$$

To show that $\mu(\omega, \mathrm{B})$ is a regular conditional distribution for $Y=(Y_1, Y_2,\ldots)$ given $\mathscr{G}$, it suffices to verify that $\mu(\cdot, \mathrm{B})$ is a version of $\mathbb{P}(Y\in B\,|\,\mathscr{G})$ for each $B\in\mathscr{B}(\mathbb{R}^\infty)$, since $\mu(\omega, \cdot)$ is a probability measure by definition. If $A_1,\ldots, A_n\in\mathscr{B}(\mathbb{R}), n\geqslant 1$, then

$$\begin{aligned}\mu(\omega, A_1\times\cdots\times A_n\times\mathbb{R}\times\cdots)&=\nu^\omega(A_1)\cdots\nu^\omega(A_n)1_{N^c}+\nu^\omega(A_1)\cdots\nu^\omega(A_n)1_N\\ &=\nu_0(\omega, \xi_1\in A_1)\cdots\nu_0(\omega, \xi_1\in A_n)1_{N^c}+\nu(A_1)\cdots\nu(A_n)1_N,\end{aligned}$$

and therefore $\mu(\omega, A_1\times\cdots\times A_n\times\mathbb{R}\times\cdots)$ is $\mathscr{G}$-measurable. Besides outside the $\mathscr{G}$-null set N

$$\begin{aligned}\mu(\omega, A_1\times\cdots\times A_n\times\mathbb{R}\times\cdots) &= \nu_0(\omega, \xi_1\in A_1)\cdots\nu_0(\omega, \xi_1\in A_n)\\ &= \nu_0(\omega, \xi_1\in A_1)\cdots\nu_0(\omega, \xi_n\in A_n)\\ &= \mathbb{P}(Y_1\in A_1\,|\,\mathscr{G})(\omega)\cdots\mathbb{P}(Y_n\in A_n\,|\,\mathscr{G})(\omega)\\ &= \mathbb{P}(Y_1\in A_1,\ldots, Y_n\in A_n\,|\,\mathscr{G})(\omega)\\ &= \mathbb{P}(Y\in A_1\times\cdots A_n\times\mathbb{R}\times\cdots\,|\,\mathscr{G})(\omega) \text{ a.s.}\end{aligned}$$

Therefore $\mu(\cdot, A_1\times\cdots\times A_n\times\mathbb{R}\times\cdots)$ is a version of $\mathbb{P}(Y\in A_1\times\cdots A_n\times\mathbb{R}\times\cdots\,|\,\mathscr{G})$. Note that

$$\mathscr{D}=\{A_1\times\cdots A_n\times\mathbb{R}\times\cdots : n\geqslant 1, A_i\in\mathscr{B}(\mathbb{R}), i=1,\ldots, n\}$$

is a π-class that generates $\mathscr{B}(\mathbb{R}^\infty)$. Since

$$\mathscr{H}=\{B\in\mathscr{B}(\mathbb{R}^\infty) : \mu(\cdot, B) \text{ is a version of } \mathbb{P}(Y\in B\,|\,\mathscr{G})\}$$

is a λ-class with $\mathscr{D}\subset\mathscr{H}$, $\mathscr{B}(\mathbb{R}^\infty)\subset\mathscr{H}$. This implies that $\mu(\cdot, B)$ is a version of $\mathbb{P}(Y\in B\,|\,\mathscr{G})$ for each $B\in\mathscr{B}(\mathbb{R}^\infty)$.

Finally by Lemma A.1

$$\begin{aligned}\int_{\mathbb{R}^\infty} |\xi_i(y)|\,\mu(\omega, dy) &= \int_{\mathbb{R}^\infty} |\xi_1(y)|\,\mu(\omega, dy)\\ &= \mathbb{E}(|\xi_1(Y)|\,|\,\mathscr{G})(\omega) = \mathbb{E}(|Y_1|\,|\,\mathscr{G})(\omega) \quad \text{a.s.}\end{aligned}$$

The integrability of Y_1 entails $\mathbb{E}(|Y_1|\,|\,\mathscr{G})(\omega)<\infty$ a.s., and hence the claims follows. This completes the proof. □

Theorem A.3. If Y_1, Y_2,... are conditionally i.i.d. random variables given a sub σ-algebra $\mathscr{G}$ and if Y_1 is integrable, then

$$\bar{Y}_n=\frac{Y_1+\cdots+Y_n}{n}\to\mathbb{E}(Y_1\,|\,\mathscr{G}) \quad \text{a.s. } (n\to\infty).$$

Proof. Let $\mu^\omega(B)=\mu(\omega, B)$, $(\omega, B)\in\Omega\times\mathscr{B}(\mathbb{R}^\infty)$ be a regular conditional distribution for $Y=(Y_1, Y_2,\ldots)$ given $\mathscr{G}$ such that the coordinate functions ξ_1, $\xi_2\ldots$ are i.i.d. random variables on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty), \mu^\omega)$ for each ω. We will show that

$$\mathbb{P}\left(\sup_{n\geqslant m} |\bar{Y}_n - \mathbb{E}(Y_1|\mathscr{G})| > \varepsilon\right) \to 0 \quad (m\to\infty), \tag{A.3}$$

which is equivalent to the convergence $\bar{Y}_n \to \mathbb{E}(Y_1|\mathscr{G})$ a.s. as $n\to\infty$. For all $\varepsilon>0$

$$\begin{aligned}\mathbb{P}\left(\sup_{n\geqslant m} |\bar{Y}_n - \mathbb{E}(Y_1|\mathscr{G})| > \varepsilon\right) &= \mathbb{E}\left[\mathbb{P}\left(\sup_{n\geqslant m} |\bar{Y}_n - \mathbb{E}(Y_1|\mathscr{G})| > \varepsilon \,|\mathscr{G}\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(\sup_{n\geqslant m} \left|\frac{1}{n}\sum_{i=1}^n \xi_i(Y) - \mathbb{E}(Y_1|\mathscr{G})\right| > \varepsilon \,|\mathscr{G}\right)\right] \\ &= \mathbb{E}\left[\mu^\omega\left\{y\in\mathbb{R}^\infty : \sup_{n\geqslant m}\left|\frac{1}{n}\sum_{i=1}^n \xi_i(y) - \mathbb{E}(Y_1|\mathscr{G})(\omega)\right| > \varepsilon\right\}\right].\end{aligned}$$

The last equation follows from Lemma A.1. Since Y_1 is assumed to be integrable, Lemma A.2 shows that ξ_1, $\xi_2,\ldots$ are i.i.d. integrable random variables on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty), \mu^\omega)$ for almost all ω. It follows by the strong law of large numbers and Lemma A.1 that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \xi_i \to \int_{\mathbb{R}^\infty} \xi_1 d\mu^\omega &= \mathbb{E}(\xi_1(Y)\,|\mathscr{G})(\omega) \\ &= \mathbb{E}(Y_1|\mathscr{G})(\omega) \quad \mu^\omega\text{-a.s.}\end{aligned}$$

for almost all ω. It follows that

$$\mu^\omega\left\{y\in\mathbb{R}^\infty : \sup_{n\geqslant m}\left|\frac{1}{n}\sum_{i=1}^n \xi_i(y) - \mathbb{E}(Y_1|\mathscr{G})(\omega)\right| > \varepsilon\right\} \to 0$$

for almost all ω. And now (A.3) is obtained by the dominated convergence theorem.

References

[1] Aldous, D. J. (1982). On exchangeabilitiy and conditional independence, In Koch, G. and Spizzichino, F., editors, *Exchangeabilitiy in Probability and Statistics*, 165-170,

North-Holland, Amsterdam.

[2] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed., Springer-Verlag, New York.

[3] Deuschel, J. D. and Stroock, D. W. (2000). *Large Deviations*, AMS Chelsea Publishing, Amer. Math. Soc.

[4] Fu, J. C. and Kass, R. E. (1988). The exponential rate of convergence of posterior distributions, *Ann. Inst. Statist. Math.*, **40**, 683-691.

[5] Ganesh, A. and O'Connell, N. (1999). An inverse of Sanov's theorem, *Statist. Probab. Lett.* **42**,201-206.

[6] Shen, X. and Wasserman, L. (1998). Rates of convergence of posterior distributions, *Ann. Probab.*, **29**, 687-714.