Large Deviations for Randomly Weighted Sum of Random Measures

Shui Feng
McMaster University

The 16th Workshop on Markov Processes and Related Topics

Central South University.

July 12-16, 2021

- Problem
- Motivation
- Result
- Application

Problem

Let $\mathcal{Z}_n, n = 1, 2, ...$ be sequence of i.i.d. random probability measures on space S with common law Π, and, independently, let $\mathbf{W} = \{W_i : i \geq 1\}$ be i.i.d. positive random variables.

Set

$$X_{nk} = \frac{W_k}{\sum_{i=1}^{n} W_i}, n \geq 1$$

and

$$\mathcal{W}_n(\Pi, \mathbf{W}) = \sum_{k=1}^{n} X_{nk}\mathcal{Z}_k$$

Assume that

1 S is compact Polish space.

2 The mean measure of Π is ν_0.

3 $\lim_{n\to\infty} \mathbb{E}[\sum_{k=1}^{n} X_{nk}^2] = 0$

Then

$$\mathcal{W}_n(\Pi, \mathbf{W}) \to \nu_0 \text{ in probability as } n \to \infty.$$

Problem: Large deviation principle for $\{\mathcal{W}_n(\Pi, \mathbf{W}) : n \geq 1\}$.

Example 1(Empirical Distribution)

$$\mathcal{L}_{n,\nu_0} = \frac{1}{n}\sum_{k=1}^{n} \delta_{\xi_k}$$

where $\xi_k, k = 1, 2, \ldots$ are i.i.d. with common distribution ν_0.

Example 2(Finite Dirichlet Weight)

$$\mathcal{W}_{n,\nu_0} = \sum_{k=1}^{n} \frac{W_k}{\sum_{i=1}^{n} W_i}\delta_{\xi_k}$$

where $W_i, i = 1, 2, \ldots$ are i.i.d. exponential with parameter one, $\xi_k, k = 1, 2, \ldots$ are i.i.d. with common distribution ν_0.

Example 3(Dirichlet Process)

For each $n \geq 1$, let $\{J_i : i \geq 1\}$ denote the set of jump sizes of a gamma subordinator over the interval $[0, n]$. For $1 \leq k \leq n$, the set $\{J_{ki} : i \geq 1\}$ consists of all jump sizes over $[k-1, k]$. The sequence $\xi_k, k = 1, 2, \ldots$ is i.i.d. with common distribution ν_0.

$$\begin{aligned}
\mathcal{V}_{n,\nu_0} &= \sum_{k=1}^{\infty} \frac{J_k}{\sum_{i=1}^{\infty} J_i} \delta_{\xi_k} \\
&= \sum_{k=1}^{n} \frac{\sum_{i=1}^{\infty} J_{ki}}{\sum_{i=1}^{\infty} J_i} \sum_{l=1}^{\infty} \frac{J_{kl}}{\sum_{i=1}^{\infty} J_{ki}} \delta_{\xi_l} \\
&= \sum_{k=1}^{n} \frac{\sum_{i=1}^{\infty} J_{ki}}{\sum_{i=1}^{\infty} J_i} \mathcal{Z}_k
\end{aligned}$$

where $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$ are i.i.d. with the same distribution as $\mathcal{V}_{1,\nu_0}$.

Motivation

Bayesian Basics: Consider a random variable ξ with distribution depending on parameter θ. In Bayesian statistics, the parameter is modelled as a random variable Θ. This reflects the basic principle of Bayesian statistics: all forms of uncertainty should be expressed as randomness.

The distribution Q of Θ is called the *prior distribution*.

Under a Bayesian model, data is generated in two stages, as

$$\Theta \sim Q$$

$$\xi_1, \xi_2, \ldots | \Theta \sim \text{iid with common distribution } P_\Theta.$$

Here the sequence $\xi_1, \xi_2, \ldots$ is conditionally i.i.d. or exchangeable.

The main objective is then to determine the *posterior distribution*, the conditional distribution of Θ given the data,

$$Q\{\Theta \in \cdot | \xi_i, i = 1, \ldots, n\}.$$

This corresponds to parameter estimation in the classical approach.

A *Bayesian nonparametric model* is a Bayesian model with infinite dimensional parametric space.

Examples include the space of probability measures and the space of probability density functions.

Let Θ follow the **Dirichlet process** with base measure ν_0 and concentration parameter β. Then the posterior distribution of Θ given the sample $\xi_1, \ldots, \xi_n$ is

$$U\mathcal{V}_{\beta,\nu_0} + (1-U)\sum_{k=1}^{n} X_{nk}\delta_{\xi_k}$$

where U is $Beta(\beta, n)$, $W_1, \ldots, W_n$ i.i.d. exponential(1), and

$$X_{nk} = \frac{W_k}{\sum_{i=1}^{n} W_i}.$$

If

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{\xi_i} \to \nu_0,$$

then

$$U\mathcal{V}_{\beta,\nu_0} + (1-U)\sum_{k=1}^{n} X_{nk}\delta_{\xi_k} \to \nu_0.$$

The corresponding large deviation principle is shown in **Ganesh and O'Connell** [5] to have rate function $H(\nu_0|\mu)$.

Our large deviation result can thus be viewed as the annealed large deviation associated with the Dirichlet posterior.

Result

Theorem 1. (**F.** [2])*Large deviation principle for* $\{\mathcal{W}_n(\Pi, \mathbf{W}) : n \geq 1\}$ *holds if some local estimates hold for projections on a large class of partitions.*

Remark 1. *The i.i.d. assumption for* $\{\mathcal{Z}_n : n \geq 1\}$ *can be relaxed. But the local estimates will be difficult to obtain. A successful situation is found in* **Gamboa and Rouault** [3] *and* **Gamboa, Nagel and Rouault** [4] *where* $\{\mathcal{Z}_n : n \geq 1\}$ *are the Dirac measures of eigenvalues of random matrices.*

Theorem 2. (**F.** [2]) *Large deviation principle for* $\{\mathcal{W}_{n,\nu_0} : n \geq 1\}$ *holds with rate function*

$$J(\mu, \nu_0) = \inf\{H(\nu|\mu) + H(\nu|\nu_0) : \nu \in M_1(S)\}.$$

Comparison of Rate Functions

1 Rate function for empirical distribution or Sanov's theorem (**Sanov** [10]):

$H(\mu|\nu_0)$ corresponding to

$$\mathcal{L}_{n,\nu_0} \to \nu_0$$

2 Rate function for Dirichlet process (**Lynch and Sethuraman** [7], **Dawson and F.** [1]): $H(\nu_0|\mu)$ corresponding to

$$\mathcal{V}_{n,\nu_0} \to \nu_0$$

3 Rate function for the finite Dirichlet weighted sum: $J(\mu, \nu_0)$ corresponding to

$$\mathcal{W}_{n,\nu_0} \to \nu_0$$

Inequality

$$\boxed{J(\mu, \nu_0) \leq \min\{H(\mu|\nu_0), H(\nu_0|\mu)\}}$$

- **The inequality can be strict!**
- $\mathcal{W}_{n,\nu_0}$ **is more "**$random$**" and "**$far\ from\ \nu_0$**" than both** $\mathcal{L}_{n,\nu_0}$ **and** $\mathcal{V}_{n,\nu_0}$

Random Means and LDP

Given a random probability of the form

$$\sum_{i=1}^{\infty} P_i \delta_{X_i},$$

the random mean is

$$\sum_{i=1}^{\infty} X_i P_i.$$

The comprehensive coverages on random means can be found in **Lijoi and Prünster** [6] and **Pitman** [8].

Sample and Dirichlet Means

Sample Mean

$$M_n^1 = \langle \mathcal{L}_{n,\nu_0}, x \rangle = \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

Finte Dirichlet Mean

$$M_n^2 = \langle \mathcal{W}_{n,\nu_0}, x \rangle = \sum_{i=1}^{n} \frac{W_i}{\sum_{k=1}^{n} W_k} \xi_i$$

Dirichlet Mean

$$M_n^3 = \langle \mathcal{V}_{n,\nu_0}, x \rangle = \sum_{i=1}^{\infty} V_i \xi_i$$

where $V_i = (1 - U_1) \ldots (1 - U_{i-1}) U_i$, U_i is i.i.d. $Beta(1, n)$.

LDPs

LDPs hold for $M_n^i, i = 1, 2, 3$ by contraction principle.

Rate Functions

Sample Mean (forward projection)

$$I_1(u, \nu_0) = \inf\{H(\mu|\nu_0) : \langle \mu, x \rangle = u\}$$

Finite Dirichlet Mean

$$I_2(u, \nu_0) = \inf\{J(\mu, \nu_0) : \langle \mu, x \rangle = u\}$$

Dirichlet Mean (reversed projection)

$$I_3(u, \nu_0) = \inf\{H(\nu_0|\mu) : \langle \mu, x \rangle = u\}$$

Relations

In general, we have

$$I_2(u,\nu_0) \le \min\{I_1(u,\nu_0), I_3(u,\nu_0)\}.$$

If $\nu_0(d\,x) = d\,x$, then

$$\begin{aligned} I_1(u,\nu_0) &= I_3(u,\nu_0) \\ &= I_2(u,\nu_1) \end{aligned}$$

where (see **Regazzini, Guglielmi, and Di Nunno** [9])

$$\begin{aligned} \nu_1 &= \text{law of } \; M_1^3 = \langle \mathcal{V}_{1,\nu_0}, x\rangle \\ \frac{d\,\nu_1}{d\,\nu_0} &= \frac{1}{\pi} e^{-\int_0^1 \log|x-y|\,d\,y} \sin(\pi x). \end{aligned}$$

The density of the law of M_n^3 has Radon-Nikodym derivative with respect to ν_0

$$\frac{n-1}{\pi}\int_0^x (x-y)^{n-2} e^{-n\int_0^1 \log|y-z| d\,z} \sin(n\pi y) d\,y, n \geq 2, 0 < x < 1.$$

References

[1] D.A. Dawson and S. Feng (2001). Large deviations for the Fleming–Viot process with neutral mutation and selection, II. *Stoch. Proc. Appl.* **92**, 131–162.

[2] S. Feng (2021). Large deviations for randomly weighted sum of random measures. arXiv: 2106.12493v1.

[3] F. Gamboa and A. Rouault (2010). Canonical moments and random spectral measures. *J. Theor. Probab.*, **23**, 1015–1038.

[4] F. Gamboa, J. Nagel and A. Rouault (2016). Sum rules via large deviations. *J. Funct. Anal.*, **270**, 509–559.

[5] A.J. Ganesh and N. O'Connell (2000). A large-deviation principle for Dirichlet posteriors. *Bernoulli* **6**, No. 6, 1021–1034.

[6] A. Lijoi and I. Prünster (2009). Distributional properties of means of random probability measures. *Stat. Surv.*, 3, 47–95.

[7] J. Lynch and J. Sethuraman (1987). Large deviations for processes with independent increments. *Ann. Probab.* **15**, 610–627.

[8] J. Pitman (2018). Random weighted averages, partition structures and generalized arcsine laws. arXiv:1804.07896v1.

[9] E. Regazzini, A. Guglielmi, and G. Di Nunno (2002). Theory and numerical analysis for exact distribution of functionals of a Dirichlet process. *Ann. Statist.*, **30**, 1376–1411.

[10] I.N. Sanov (1957). On the probability of large deviations of random variables. (Russian)*Mat. Sb. N. S.* 42 (84), 11–44. (English translation in *Select. Transl. Math. Statist. and Probability*, Vol. 1 pp. 213–244, 1961.)

Thank You!