Large deviations for smoothed empirical distributions

Shikimi Takuhisa

Abstract

This paper deals with the large deviations for the empirical distribution of kernel type (smoothed empirical distribution)

$$\hat{F}^n(x) \triangleq \frac{1}{n}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right), \tag{1}$$

where K is the kernel and h_n is a smoothing parameter. We limit ourselves to the i.i.d. case. One of the purposes of this paper is to establish the large deviation principle (LDP) for the smoothed empirical distribution. This result is proved to be not only a smoothed version of Sanov's theorem, i.e., the LDP for the standard empirical distribution, but a generalization of the theorem. In this paper we have considered the weak topology rather than the τ-topology, namely, we endow the space of distributions on $\mathbb{R}^d$ with the weak topology. After proving it, we consider several applications : a smooth version of Dvoretzky-Kiefer-Wolfowitz's large deviation inequality, a limiting behavior of the kernel density estimator and the LDP for the smoothed bootstrap empirical distribution.

We consider two approaches in order to prove the LDP. First, it is proved by the weak convergence approach under the condition that the underlying distribution is absolutely continuous with respect to Lebesugue measure on $\mathbb{R}^d$. This approach, summerized in Dupuis and Ellis (1997), is based on interpreting large deviation problems to the weak convergence problems of associated controlled processes and focuses on the key role of the variational formula of the Kullback-Leibler information. The weak convergence approach is useful and applicable to a wide range of problems, whereas we can also prove the LDP for $\hat{F}^n$ by another approach, which is more simple since it uses Sanov's theorem.

Second, alternative approach is based on introducing a metric in the space of distributions and then using Sanov's theorem. Let us denote by ϕ_G the characteristic function of $G \in \mathcal{F}$, where $\mathcal{F}$ is the space of distributions on $\mathbb{R}^d$. For G, $H \in \mathcal{F}$ we define a metric in $\mathcal{F}$ by

$$\rho(G, H) \triangleq \sum_{k=1}^{\infty} \frac{1}{2^k} \sup_{t\in[-k,k]} |\, \phi_G(t) - \phi_H(t)\,|, \tag{2}$$

where $[-k, k] = [-k, k] \times \cdots \times [-k, k] \subset \mathbb{R}^d$. It is trivial to show that ρ is one of the metrics that are compatible with the weak topoloty. In addition, the space $\mathcal{F}$ is complete with respect to ρ. Let $\{H_n, n \in \mathbb{N}\}$ be a sequence in $\mathcal{F}$ converging weakly to δ_0, where δ_0 is the distribution with unit mass at 0. The following property of ρ is crucial in this approach.

$$\sup_{G\in\mathcal{F}} \rho(G * H_n, G) \longrightarrow 0, \tag{3}$$

where $*$ stands for the convolution operator. This property shows that δ_0 can be regarded as an approximate g identity h with respect to ρ. Let F^n be the standard empirical distribution based on the segment $X_1, \dots, X_n$ of i.i.d. d-dimensional random variables with distribution F. Utilizing (3) and Sanov's theorem we can prove that $\{F^n * H_n, n \in \mathbb{N}\}$ satisfies the LDP on $\mathcal{F}$ with good rate function $R(\,\cdot\,|F)$, where $R(G\,|F)$ denotes the Kullback- Leibler information for G and F in $\mathcal{F}$. This theorem reduces to Sanov's theorem if we choose $H_n = \delta_0$, and thus it can be regarded as an extension of Sanov's theorem. As a corollary of this theorem it can be proved that the smoothed empirical distribution (1) satisfies the LDP with the same good rate function above.

We consider several applications of the LDP for $\hat{F}^n$. We can apply the LDP for $\hat{F}^n$ to obtain the following large deviation inequality : for every $\varepsilon > 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\sup_x |\,\hat{F}^n(x) - F(x)\,| \geqslant \varepsilon\right) \leqslant -\varepsilon^2,$$

where F and K are assumed to be continuous.

Suppose that F has the density f and let f^n be the usual kernel density estimator with density kernel κ. Denote by ϕ_g the characteristic function of a density g. We assume that $\phi_f(t)$ is integrable. It can be shown that $h_n^{c+1} \sup_x |f^n(x) - f(x)|$ converges to 0 superexponentially fast if

$$\int |\phi_\kappa(t)| dt < \infty \tag{4}$$

$$\lim_{k\to\infty} \sup_n h_n^c \int_{|t|>kh_n} |\phi_\kappa(t)| dt = 0 \tag{5}$$

for a constant $c > 0$. Some symmetric kernels, e.g., Gaussian, double exponential, triangular kernels, satisfy (4) and (5) with $c = 1$.

The smoothed bootstrap empirical distribution $\hat{F}^n_*$ is defined as the smoothed empirical distribution of the bootstrap sample drawn from $\hat{F}^n$. It can be also proved that $\hat{F}^n_*$ obeys the LDP with the same good rate function $R(\,\cdot\,|F)$ almost surely.