Homework 02: Large Deviations for Log-Likelihood

Problem 1: Large Deviations for Log-Likelihood

Let P and Q be two probability distributions such that $P \ll Q$. Let X_i be i.i.d. random variables under P and Y_i be i.i.d. random variables under Q. Define for each index i

$$W_i = \log \frac{p(Y_i)}{q(Y_i)},$$
$$Z_i = \log \frac{p(X_i)}{q(X_i)}.$$

The purpose of this problem is to prove:

Proposition I. For any $t \geq 0$ and any $n \in \mathbb{N}$, the following large-deviation bound holds:

$$\mathbb{P}\Big[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\Big] \leq \exp\big(-n\big(\alpha + \tfrac{t}{2}\big)\big),$$

where

$$\mathcal{B}(P, Q) = \mathbb{E}_{Y\sim Q}\Big[\sqrt{\tfrac{p(Y)}{q(Y)}}\Big], \quad \alpha = -2\log \mathcal{B}(P, Q).$$

1. (Chernoff bound) Show that for all $t \geq 0$,

   $$\mathbb{P}\Big[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\Big] \leq \exp\big(-n \cdot F(t)\big),$$

   where

   $$\psi_Q(\lambda) = \log \mathbb{E}[e^{\lambda W_i}],$$
   $$\psi_P(\lambda) = \log \mathbb{E}[e^{\lambda Z_i}],$$
   $$F(t) = \sup_{\lambda \geq 0}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}.$$

2. (Value at zero) Show that

   $$F(0) = -\psi_P\big(-\tfrac{1}{2}\big) - \psi_Q\big(\tfrac{1}{2}\big) = \alpha.$$

3. (Linear lower bound) Prove that for all $t \geq 0$,

   $$F(t) \geq F(0) + \tfrac{t}{2},$$

   and conclude Proposition I.

Solution

1. Chernoff Bound

Let

$$S_n = \sum_{i=1}^{n} (W_i - Z_i).$$

For any $\lambda \geq 0$, by Markov's inequality,

$$\mathbb{P}[S_n \geq nt] = \mathbb{P}\big[e^{\lambda S_n} \geq e^{\lambda nt}\big] \leq e^{-\lambda nt}\mathbb{E}[e^{\lambda S_n}] = e^{-\lambda nt}\big(\mathbb{E}[e^{\lambda W_1}]\big)^n\big(\mathbb{E}[e^{-\lambda Z_1}]\big)^n.$$

Define

$$\psi_Q(\lambda) = \log \mathbb{E}[e^{\lambda W_1}], \quad \psi_P(-\lambda) = \log \mathbb{E}[e^{-\lambda Z_1}].$$

Then

$$\mathbb{P}[S_n \geq nt] \leq \exp\big(-n(\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda))\big).$$

Optimizing over $\lambda \geq 0$ yields

$$\mathbb{P}[S_n \geq nt] \leq \exp(-nF(t)), \quad F(t) = \sup_{\lambda \geq 0}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}.$$

2. Value at Zero

We have

$$F(0) = \sup_{\lambda \geq 0}\{-\psi_P(-\lambda) - \psi_Q(\lambda)\}.$$

Observe:

$$\psi_P(-\lambda) = \log \mathbb{E}_P\big[\big(\tfrac{p(X)}{q(X)}\big)^{-\lambda}\big] = \log \int p(x)^{1-\lambda} q(x)^{\lambda}\, dx$$
$$\psi_Q(\lambda) = \log \mathbb{E}_Q\big[\tfrac{p(Y)}{q(Y)}^{\lambda}\big] = \log \int p(y)^{\lambda} q(y)^{1-\lambda}\, dy.$$

Now by cauchy-schwartz (or Hölder with p=q) we have:

$$\left(\int p(x)^{1-\lambda} q(x)^{\lambda}\, dx\right)\left(\int p(y)^{\lambda} q(y)^{1-\lambda}\, dy\right) \geq \left(\int \sqrt(pq)\right)^2 = B^2(P,Q)$$

Thus by monoticity of log we have:

$$F(0) = -\log\left[\left(\int p(x)^{1-\lambda} q(x)^{\lambda}\, dx\right)\left(\int p(y)^{\lambda} q(y)^{1-\lambda}\, dy\right)\right] \leq -\log B^2(P,Q) = -2\log B(P,Q) = \alpha$$

Moreover, if we set $\lambda = \frac{1}{2}$, we have $F(0) = -2\log B^2(P,Q) = \alpha$, so we have equality condition at $\lambda = \frac{1}{2}$, and the proof is complete.

3. Linear Lower Bound

Using the candidate $\lambda = \frac{1}{2}$ in the definition of $F(t)$,

$$F(t) \geq \tfrac{1}{2}t - \psi_P(-\tfrac{1}{2}) - \psi_Q(\tfrac{1}{2}) = F(0) + \tfrac{t}{2}.$$

Therefore,

$$\mathbb{P}[S_n \geq nt] \leq \exp(-nF(t)) \leq \exp\big(-n(\alpha + \tfrac{t}{2})\big),$$

as claimed.

$\square$