Seminar 5

High-dimensional Probability and Statistics

HSE University, spring 2023

The entropy method

Definition 1 (entropy)**.** Let X be a random element and let f be a function taking positive values on the support of X. The entropy of f is defined as

$$\mathrm{Ent}_X(f) = \mathbb{E} f(X) \log f(X) - \mathbb{E} f(X) \log \mathbb{E} f(X).$$

The entropy method is a useful technique to derive concentration inequalities. The idea is as follows. Let X be a random element and f be a function defined on the support of X with finite exponential moments. Suppose that we are interested in concentration of $f(X)$ around its expectation. Fix $\lambda > 0$ and consider

$$H(\lambda) = \frac{1}{\lambda} \log \mathbb{E} e^{\lambda f(X)}, \quad H(0) = \mathbb{E} f(X).$$

It is easy to observe that

$$H'(\lambda) = \frac{\mathbb{E} f e^{\lambda f(X)}}{\lambda\, \mathbb{E} e^{\lambda f(X)}} - \frac{\log \mathbb{E} e^{\lambda f(X)}}{\lambda^2} = \frac{\mathrm{Ent}_X\left(e^{\lambda f}\right)}{\lambda^2\, \mathbb{E} e^{\lambda f(X)}}.$$

Hence, an upper bound on $\mathrm{Ent}_X\left(e^{\lambda f}\right)$ yields an upper bound on $H'(\lambda)$, which, in its turn, implies a bound on the exponential moment $\mathbb{E} e^{\lambda f(X)}$.

Properties of the entropy

Problem 1 (entropy tenzorization). Let $X_1, \dots, X_n$ be independent random variables, $X = (X_1, \dots, X_n)$, and let f be a positive function on $\mathbb{R}^n$. Prove that

$$\mathrm{Ent}_X(f) \leqslant \sum_{i=1}^n \mathbb{E}\, \mathrm{Ent}_{X_i}(f),$$

where $\mathrm{Ent}_{X_i}(f)$ denotes the entropy of f with respect to X_i with all other variables frozen.

Hint. Use the *duality formula*: $\mathrm{Ent}_X(f) = \sup\left\{\mathbb{E} f(X) g(X) : \mathbb{E} e^{g(X)} \leqslant 1\right\}$.

Problem 2. Prove that $\mathrm{Ent}_X(f) = \inf\limits_{c>0} \mathbb{E}\left[f(X)\big(\log f(X) - \log c\big) - \big(f(X) - c\big)\right]$ for any positive function f and any random element X.

Application: concentration of the supremum of an empirical process

Let $X_1, \dots, X_n$ be independent random variables and let $\mathcal{F}$ be a separable family of univariate functions, taking its values in $[0,1]$. We are interested in large deviation inequalities for

$$Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i).$$

Applying the conclusions of Problems 1 and 2, we obtain that, for any $\lambda \geqslant 0$,

$$\mathrm{Ent}\left(e^{\lambda Z}\right) \leqslant \mathbb{E} \sum_{i=1}^{n} \mathrm{Ent}_{X_i}\left(e^{\lambda Z}\right) \leqslant \mathbb{E} \sum_{i=1}^{n} \left[e^{\lambda Z}\big(\lambda Z - \lambda h_i\big) - \big(e^{\lambda Z} - e^{\lambda h_i}\big)\right],$$

where, $h_1, \dots, h_n$ are arbitrary functions such that, for any $i \in \{1, \dots, n\}$, the function h_i depends on all variables but X_i. Taking

$$h_i = \sup_{f \in \mathcal{F}} \sum_{j \neq i} f(X_i),$$

we obtain that

$$\sum_{i=1}^{n} \left[e^{\lambda Z}\big(\lambda Z - \lambda h_i\big) - \big(e^{\lambda Z} - e^{\lambda h_i}\big)\right] \leqslant \frac{\lambda^2}{2} \cdot Z e^{\lambda Z}.$$

This yields that

$$H'(\lambda) \leqslant \frac{\lambda^2}{2} \cdot \left(\frac{\mathbb{E} Z e^{\lambda Z}}{\mathbb{E} e^{\lambda Z}}\right) = \frac{\lambda^2}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\mathbb{E} e^{\lambda Z}\right) = \frac{\lambda^2}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\lambda H(\lambda)\right) \quad \text{for all } \lambda > 0. \tag{1}$$

Problem 3 (Talagrand's inequality). Using the inequality (1), show that $H(\lambda) \leqslant 2\,\mathbb{E}Z/(2-\lambda)$ for all $\lambda \in (0,2)$ and, hence,

$$\mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leqslant \exp\left\{\frac{\lambda^2/2}{1-\lambda/2}\,\mathbb{E}Z\right\} \quad \text{for any } \lambda \in (0,2). \tag{2}$$

Remark 2. Fix an arbitrary $t > 0$. Applying the Markov inequality and (2) with $\lambda > 0$, satisfying the equality

$$1 - \frac{\lambda}{2} = \frac{\mathbb{E}Z}{\mathbb{E}Z + t/2},$$

we obtain that

$$\mathbb{P}\left(Z - \mathbb{E}Z > t\right) \leqslant \exp\left\{-\frac{t^2/2}{\mathbb{E}Z + t/2}\right\}.$$

This implies that, for any $\delta \in (0,1)$, one has

$$Z - \mathbb{E}Z \leqslant \sqrt{2\mathbb{E}Z \log(1/\delta)} + \log(1/\delta)$$

with probability at least $1 - \delta$. In other words, the Talagrand inequality leads to sharper bounds on the supremum, than McDiarmid's inequality.