THE METHOD OF TYPES

YU-LIANG WU

Table of Contents

1. The Method of Types

The *method of types* refers to a set of combinatorial arguments used to establish Large Deviation Principles (LDP) for i.i.d. random variables taking values in a finite alphabet $\Sigma = \{a_1, a_2, ..., a_{|\Sigma|}\}$.

To demonstrate the argument, we consider i.i.d. random variables $(Y_i)_{i=1}^{\infty}$ taking values in Σ. Let $M_1(\Sigma)$ (which can be identified as a subset of $\mathbb{R}^{|\Sigma|}$) denote the space of all probability measures (laws) on Σ. Assuming the law of each Y_i is $\mu \in M_1(\Sigma)$, we denote by P_μ the joint law of $(Y_i)_{i=1}^{\infty}$ on the product space $\Sigma^{\mathbb{N}}$.

Definition 1.1. The *empirical measure* $L_n^{\mathbf{y}}$ of a finite sequence $\mathbf{y} = (y_1, y_2, ..., y_n) \in \Sigma^n$ is the measure in $M_1(\Sigma)$ that records the frequency of occurrences of each a_i; namely, $L_n^{\mathbf{y}} = \left(L_n^{\mathbf{y}}(a_1), ..., L_n^{\mathbf{y}}\left(a_{|\Sigma|}\right)\right)$ with

$$L_n^{\mathbf{y}}(a_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{a_i\}}(y_j).$$

Below are some related notions:

- (Support) $\Sigma_\nu = \{a_i \in \Sigma : \nu(a_i) > 0\}$ for a measure ν.
- (Empirical distribution) $\mathcal{L}_n = \{L_n^{\mathbf{y}} : \mathbf{y} \in \Sigma^n\}$.
- (Types) $T_n(\nu) = \{\mathbf{y} \in \Sigma^n : L_n^{\mathbf{y}} = \nu\}$ for $\nu \in M_1(\Sigma)$.
- (Variational distance) For $\nu, \nu' \in M_1(\Sigma)$, $d_V(\nu, \nu') := \frac{1}{2}\sum_{a\in\Sigma}|\nu(a) - \nu'(a)|$.

We have the following fundamental estimates of the size of $T_n(\nu)$ and $\mathcal{L}_n$.

Lemma 1.2.
(1) $|T_n(\nu)| = \binom{n}{n\nu(a_1), n\nu(a_2), ..., n\nu\left(a_{|\Sigma|}\right)}$.
(2) $1 \le |\mathcal{L}_n| \le (n+1)^{|\Sigma|}$.

Proof. (1) It follows from a combinatorial argument.

(2) Note that

$$n \cdot \mathcal{L}_n = \left\{(k_a)_{a\in\Sigma} : k_a \in \mathbb{Z}_+, \sum_{a\in\Sigma} k_a = n\right\},$$

where the latter set has cardinality

$$\binom{n + |\Sigma| - 1}{|\Sigma| - 1} = \prod_{i=1}^{|\Sigma|-1} \frac{n+i}{i} \le (n+1)^{|\Sigma|},$$

as desired. □

In the following, we estimate the large deviation probabilities.

Definition 1.3 (entropy)**.** Let $\nu, \eta \in M_1(\Sigma)$.

- The *entropy* of $\nu \in M_1(\Sigma)$ is
$$H(\nu) = -\sum_{a\in\Sigma} \nu(a)\log\nu(a).$$
- The *relative entropy* (or Kullback–Leibler divergence) of ν with respect to η is
$$H(\nu|\eta) = \sum_{a\in\Sigma} \nu(a)\log\frac{\nu(a)}{\eta(a)}.$$

We adopt the convention that $0\log 0 = 0$ and $0\log\left(\frac{0}{0}\right) = 0$.

Remark 1.4. *Let $M_1(\Sigma)$ be endowed with the variational distance. Then,*

(1) $M_1(\Sigma)$ is compact.
(2) $H(\cdot)$ is continuous on $M_1(\Sigma)$
(3) $H(\cdot\,|\mu)$ is continuous on the compact set $\left\{\nu \in M_1(\Sigma) : \Sigma_\nu \subset \Sigma_\mu\right\}$.

Lemma 1.5. *If $\mathbf{y} \in T_n(\nu)$, then the following hold.*

(1) $P_\mu((Y_1, ..., Y_n) = \mathbf{y}) = e^{-n(H(\nu)+H(\nu|\mu))}$.
(2) $(n+1)^{-|\Sigma|}e^{nH(\nu)} \le |T_n(\nu)| \le e^{nH(\nu)}$.
(3) $(n+1)^{-|\Sigma|}e^{-nH(\nu|\mu)} \le P_{\mu(L_n^{\mathbf{Y}}=\nu)} \le e^{-nH(\nu|\mu)}$.

Proof. (1) The case $\Sigma_\nu \not\subseteq \Sigma_\mu$ is trivial, as $P_\mu((Y_1, ..., Y_n) = \mathbf{y}) = 0$ and $H(\nu|\mu) = \infty$. Assuming $\Sigma_\nu \subseteq \Sigma_\mu$, we have by definition
$$P_\mu((Y_1, ..., Y_n) = \mathbf{y}) = \prod_{a\in\Sigma} \mu(a)^{n\nu(a)} = e^{-n(H(\nu)+H(\nu|\mu))}.$$

(2) The second inequality follows immediately from
$$1 \ge P_\nu(L_n^{\mathbf{Y}} = \nu) = e^{-nH(\nu)}\ |T_n(\nu)|.$$

To prove the first, we claim that if $\Sigma_{\nu'} \subset \Sigma_\nu$, then
$$\begin{aligned}\frac{P_\nu(L_n^{\mathbf{Y}} = \nu)}{P_\nu(L_n^{\mathbf{Y}} = \nu')} &= \prod_{a\in\Sigma} \frac{|T_n(\nu)|\ \nu(a)^{n\nu(a)}}{|T_n(\nu')|\ \nu(a)^{n\nu'(a)}} \\ &= \prod_{a\in\Sigma} \frac{(n\nu'(a))!}{(n\nu(a))!}\nu(a)^{n\nu(a)-n\nu'(a)} \ge 1.\end{aligned}$$

Indeed, the last expression consists of terms of the form $\frac{m!}{\ell!}\left(\frac{\ell}{n}\right)^{\ell-m}$ where the inequality is validated using the fact $\frac{m!}{\ell!} \ge \ell^{m-\ell}$:
$$\frac{P_\nu(L_n^{\mathbf{Y}} \in T_n(\nu))}{P_{\nu'}(L_n^{\mathbf{Y}} \in T_n(\nu))} \ge n^{n\left[\sum_{a\in\Sigma}\nu'(a) - \sum_{a\in\Sigma}\nu(a)\right]} = 1.$$

To prove the desired inequality, observe that
$$1 = \sum_{\nu'\in\mathcal{L}_n} P_\nu(L_n^{\mathbf{Y}} = \nu') \le |\mathcal{L}_n|\ P_\nu(L_n^{\mathbf{Y}} = \nu) = |\mathcal{L}_n|\ e^{-nH(\nu)}\ |T_n(\nu)|.$$

The proof is concluded by applying Lemma 1.2.

(3) Combining (1) and (2), we arrive at the estimate

$$P_\mu(L_n^{\mathbf{Y}} = \nu) = |T_n(\nu)|\, P_\mu((Y_1, ..., Y_n) = \mathbf{y}) \le |T_n(\nu)|\; e^{-n(H(\nu)+H(\nu|\mu))} \le e^{-n(H(\nu|\mu))},$$

whereas the other inequality follows from the lower bound on $|T_n(\nu)|$. □

Theorem 1.6 (Sanov)**.** *For every set $\Gamma \subseteq M_1(\Sigma)$ of probability measures,*

$$\begin{aligned} -\inf_{\nu\in\mathring{\Gamma}} H(\nu|\mu) &\le \liminf_{n\to\infty} \frac{1}{n} \log P_\mu(L_n^{\mathbf{Y}} \in \Gamma) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log P_\mu(L_n^{\mathbf{Y}} \in \Gamma) \le -\inf_{\nu\in\overline{\Gamma}} H(\nu|\mu). \end{aligned}$$

Proof. By Lemma 1.5, we obtain the upper bound:

$$\begin{aligned} P_\mu(L_n^{\mathbf{Y}} \in \Gamma) &= \sum_{\nu\in\Gamma\cap\mathcal{L}_n} P_\mu(L_n^{\mathbf{Y}} = \nu) \le |\Gamma \cap \mathcal{L}_n|\; e^{-n \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)} \\ &\le (n+1)^{|\Sigma|} e^{-n \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}, \end{aligned}$$

which yields

$$\limsup_{n\to\infty} \frac{1}{n} \log P_\mu(L_n^{\mathbf{Y}} \in \Gamma) \le -\liminf_{n\to\infty} \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu) \le -\inf_{\nu\in\overline{\Gamma}} H(\nu|\mu).$$

To prove the lower bound, we begin with an analogous estimate

$$P_\mu(L_n^{\mathbf{Y}} \in \Gamma) = \sum_{\nu\in\Gamma\cap\mathcal{L}_n} P_\mu(L_n^{\mathbf{Y}} = \nu) \ge (n+1)^{-|\Sigma|} e^{-n \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}.$$

Since $\sup_{\nu\in M_1(\Sigma)} d_V(\nu, \mathcal{L}_n) \le \frac{|\Sigma|}{2n}$, for every $\nu \in \mathring{\Gamma}$ with $\Sigma_\nu \subset \Sigma_\mu$, one can find for every sufficiently large n some $\nu_n \in \mathring{\Gamma} \cap \mathcal{L}_n$ such that $\nu_n \to \nu$. By the continuity of the relative entropy on the restricted domain,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\mu(L_n^{\mathbf{Y}} \in \Gamma) \ge -\limsup_{n\to\infty} H(\nu_n|\mu) = -H(\nu|\mu).$$

The theorem follows since $\nu \in \mathring{\Gamma}$ is arbitrary as long as $\Sigma_\nu \subset \Sigma_\mu$, and $H(\nu|\mu) = \infty$ whenever $\Sigma_\nu \not\subseteq \Sigma_\mu$. □

As an application of Sanov's theorem, we prove a version of Cramér's theorem concerning Large Deviations for the empirical mean. Let $f : \Sigma \to \mathbb{R}$ be a function. We identify f with the vector $\mathbf{f} \in \mathbb{R}^{|\Sigma|}$ so that the empirical average $\hat{S}_n = \frac{1}{n}\sum_{i=1}^n f(Y_i)$ can be written as the inner product $\hat{S}_n = \langle L_n^{\mathbf{Y}}, \mathbf{f}\rangle$. Sanov's theorem then implies the following result.

Theorem 1.7 (Cramér)**.** *For any set $A \subseteq \mathbb{R}$,*

$$\begin{aligned} -\inf_{x\in\mathring{A}} I(x) &\le \liminf_{n\to\infty} \frac{1}{n} \log P_\mu\big(\hat{S}_n \in A\big) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log P_\mu\big(\hat{S}_n \in A\big) \le -\inf_{x\in\overline{A}} I(x), \end{aligned}$$

where $I(x) = \inf_{\nu:\langle\nu,\mathbf{f}\rangle=x} H(\nu|\mu)$. The rate function I is continuous on the interval $[\min_{a\in\Sigma_\mu} f(a), \max_{a\in\Sigma_\mu} f(a)]$ and satisfies

$$I(x) = \sup_{\lambda\in\mathbb{R}}[\lambda x - \Lambda(\lambda)],$$

where

$$\Lambda(\lambda) = \log \sum_{a\in\Sigma} \mu(a) e^{\lambda f(a)}.$$

Proof. We provide two proofs of the theorem.

(1) The formula

$$I(x) = \inf_{\nu:\langle \nu, \mathbf{f}\rangle = x} H(\nu|\mu) = \inf_{\nu\in M_1(\Sigma)} \sup_{\lambda\in\mathbb{R}} \left[\lambda x - \sum_{a\in\Sigma} \nu(a) \log\left(\frac{\mu(a)e^{\lambda f(a)}}{\nu(a)}\right)\right]$$

follows from that $\hat{S}_n = \langle L_n^{\mathbf{Y}}, \mathbf{f}\rangle$ and that $\nu \mapsto \langle \nu, \mathbf{f}\rangle$ is a continuous function. Moreover, in view of the formula, I is clearly continuous on its effective domain $\left[\min_{a\in\Sigma_\mu} f(a), \max_{a\in\Sigma_\mu} f(a)\right]$. We note that the optimization problem here is a classical problem in convex optimization where strong duality holds (see, for example, [1, Section 5]); specifically, one can swap inf and sup without changing the value of the expression:

$$I(x) = \sup_{\lambda\in\mathbb{R}} \inf_{\nu\in M_1(\Sigma)} \left[\lambda x - \sum_{a\in\Sigma} \nu(a) \log\left(\frac{\mu(a)e^{\lambda f(a)}}{\nu(a)}\right)\right],$$

The inner minimization can then be solved using Gibbs inequality (or Jensen's inequality applied to $x \mapsto x\log x$), yielding

$$I(x) = \sup_{\lambda\in\mathbb{R}} \left[\lambda x - \log \sum_{a\in\Sigma} \mu(a) e^{\lambda f(a)}\right].$$

This establishes the Fenchel–Legendre transform representation.

(2) Alternatively, this result is a special case of Cramér's theorem for real-valued random variables, and its rate function $I(x) = \Lambda^*(x)$, a fortiori, is the Fenchel–Legendre transform of the logarithmic moment generating function

$$\Lambda(\lambda) = \log \mathbb{E}\left[e^{\lambda f(Y_1)}\right] = \log \sum_{a\in\Sigma} \mu(a) e^{\lambda f(a)}.$$

It remains to show that $\Lambda^*(x) = \inf_{\nu:\langle \nu, \mathbf{f}\rangle = x} H(\nu|\mu)$, for, once established, the continuity of I follows from that of $H(\cdot\,|\mu)$ immediately. To this end, it suffices to show that $\Lambda^*(x) \geq \inf_{\nu:\langle \nu, \mathbf{f}\rangle = x} H(\nu|\mu)$, since by Gibbs' inequality,

$$\lambda\langle \mathbf{f}, \nu\rangle - H(\nu|\mu) = \sum_{a\in\Sigma} \nu(a) \log\left(\frac{\mu(a)e^{\lambda f(a)}}{\nu(a)}\right) \leq \Lambda(\lambda)$$

for all $\nu \in M_1(\Sigma)$ and $\lambda \in \mathbb{R}$, implying $\Lambda^*(x) \leq \inf_{\nu:\langle \nu, \mathbf{f}\rangle = x} H(\nu|\mu)$.

Recall that Λ is convex and differentiable in the interior of its effective domain. By asymptotic analysis of the derivative $\Lambda'(\lambda)$, for every $x \in \left(\min_{a\in\Sigma_\mu} f(a), \max_{a\in\Sigma_\mu} f(a)\right)$, there exists $\lambda \in \mathbb{R}$ such that

$$x = \Lambda'(\lambda) = \frac{\sum_{a\in\Sigma} f(a)\mu(a)e^{\lambda f(a)}}{\sum_{a\in\Sigma} \mu(a)e^{\lambda f(a)}} \quad \Rightarrow \quad \Lambda^*(x) = \lambda x - \Lambda(\lambda).$$

Choosing $\eta_a = \frac{\mu(a)e^{\lambda f(a)}}{\sum_b \mu(b)e^{\lambda f(b)}}$, we have $\langle \eta, \mathbf{f}\rangle = x$ and

$$\Lambda^*(x) = \lambda x - \Lambda(\lambda) = H(\eta|\mu) \geq \inf_{\nu:\langle\nu,\mathbf{f}\rangle=x} H(\nu|\mu)$$

as desired. For the boundary point $x = \max_{a\in\Sigma_\mu} f(a)$ (the case for min is similar), we have that

$$\inf_{\nu:\langle\nu,\mathbf{f}\rangle=x} H(\nu|\mu) = -\log\mu(\{a \in \Sigma : f(a) = x\}).$$

and, through direct evaluation, that

$$\Lambda^*(x) = \lim_{\lambda\to\infty}[\lambda x - \Lambda(\lambda)] = -\log\mu(\{a \in \Sigma : f(a) = x\}).$$

Combining these results proves $\Lambda^*(x) = \inf_{\nu:\langle\nu,\mathbf{f}\rangle=x} H(\nu|\mu)$. Finally, for $x \notin \left[\min_{a\in\Sigma_\mu} f(a), \max_{a\in\Sigma_\mu} f(a)\right]$, one can show $\Lambda^*(x) = \infty$. This concludes the proof. □

References

[1] Boyd, Stephen, and Vandenberghe, Lieven, *Convex Optimization*, Cambridge University Press, 2004.

Research Unit of Mathematical Sciences, University of Oulu, P.O. Box 3000, 90014 University of Oulu, Finland.
Email address: yu-liang.wu@oulu.fi
URL: s92077.github.io