2. Method of types

The *method of types* refers to a combinatorial argument addressing random variables taking values in a finite alphabet $\Sigma = \left\{a_1, a_2, ..., a_{\{|\Sigma|\}}\right\}$, which made its debut in Example 1.1 and will be revisited in this section in a systematic and more purely combinatorial fashion.

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

Definition 2.1. The *empirical measure (law)* $L_n^{\mathbf{y}}$ of a finite sequence $\mathbf{y} = (y_1, y_2, ..., y_n) \in \Sigma^n$ is the measure in $M_1(\Sigma)$ that catalogs the proportion of occurrences of a_i's; namely, $L_n^{\mathbf{y}} = (L_n^{\mathbf{y}}(a_1), ..., L_n^{\mathbf{y}}(a_N))$ with

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

Below are some related notions.

- (support of measure ν) $\Sigma_\nu = \{a_i \in \Sigma : \nu(a_i) > 0\}$.
- (empirical distribution associated with $\mathbf{y}$) $\mathcal{L}_n = \{L_n^{\mathbf{y}} : \mathbf{y} \in \Sigma^n\}$.
- (sequences associated law ν) $T_n(\nu) = \{\mathbf{y} : L_n^{\mathbf{y}} = \nu\}$ 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 2.2.

(1) $|T_n(\nu)| = \binom{n}{n\nu(a_1), n\nu(a_2), ..., n\nu(a_N)}$.
(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

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

as desired. □

In the following, we estimate the large deviation probabilities.

Definition 2.3 (entropy)**.** Let $\nu, \eta \in M_1(\Sigma)$.

- The *entropy* of $\nu \in M_1(\Sigma)$ is defined as

$$H(\nu) = -\sum_{a\in\Sigma} \nu(a) \log \nu(a).$$

- The *relative entropy* of ν with respect to η is defined as

$$H(\nu|\eta) = \sum_{a\in\Sigma} \nu(a) \log \frac{\nu(a)}{\eta(a)}.$$

We adopt the convention that $0 \log 0 = 0$ and $0 \log \frac{0}{0} = 0$.

Remark 2.4. *Let $M_1(\Sigma)$ be endowed with the variational distance. Then,*
 (1) $M_1(\mu)$ is compact.
 (2) $H(\cdot)$ is continuous on $M_1(\Sigma)$
 (3) $H(\cdot\,|\mu)$ is continuous on the compact set $\{\nu \in M_1(\Sigma) : \Sigma_\nu \subset \Sigma_\mu\}$.

Lemma 2.5. *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) $(n+1)^{-|\Sigma|} e^{nH(\nu)} \leq T_n(\nu) \leq e^{nH(\nu)}$.
 (3) $(n+1)^{-|\Sigma|} e^{-nH(\nu|\mu)} \leq P_\mu(L_n^{\mathbf{Y}} = \nu) \leq e^{-nH(\nu|\mu)}$.

Proof. (1) The case $\Sigma_\nu \subseteq \Sigma_\mu$ 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 that

$$P_\mu((Y_1, ..., Y_n) = \mathbf{y}) = \prod_{a\in\Sigma} \mu(a)^{n\nu(a)} = e^{-n(H(\nu)+H(\nu|\mu))}.$$

(2) The second inequality follows immediately from the following:

$$1 \geq P_\nu((Y_1, ..., Y_n) \in T_n(\nu)) = e^{-nH(\nu)}\ |T_n(\nu)|.$$

To prove the first, we claim that if $\Sigma_{\nu'} \subset \Sigma_\nu$, then

$$\frac{P_\nu(L_n^{\mathbf{Y}} \in T_n(\nu))}{P_\nu(L_n^{\mathbf{Y}} \in T_n(\nu'))} = \prod_{a\in\Sigma} \frac{|T_n(\nu)| \nu(a)^{n\nu(a)}}{|T_n(\nu')| \nu(a)^{n\nu'(a)}} = \prod_{a\in\Sigma} \frac{(n\nu'(a))!}{(n\nu(a))!} \nu(a)^{n\nu(a)-n\nu'(a)} \geq 1.$$

Indeed, the second last expression is of the form $\frac{m!}{\ell!}\left(\frac{\ell}{n}\right)^{\ell-m}$ and the inequality is validated using the fact $\frac{m!}{\ell!} \geq \ell^{m-\ell}$:

$$\frac{P_\nu(L_n^{\mathbf{Y}} \in T_n(\nu))}{P_{\nu'}(L_n^{\mathbf{Y}} \in T_n(\nu))} \geq n^{n\left[\sum_{a\in\Sigma} \nu'(a) - \sum_{a\in\Sigma} \nu(a)\right]} = 1.$$

To prove the the desired inequality, observe that

$$1 = \sum_{\nu'\in\mathcal{L}_n} P_\nu(L_n^{\mathbf{Y}} = \nu') \leq |\mathcal{L}_n|\ P_\nu(L_n^{\mathbf{Y}} = \nu) = |\mathcal{L}_n|\ e^{-nH(\nu)}\ |T_n(\nu)|.$$

The proof is finished by applying Lemma 2.2.

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

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

whereas the other inequality follows naturally. □

Theorem 2.6 (Sanov)**.** *For every set of probability vectors in $M_1(\Sigma)$,*

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

Proof. By Lemma 2.5, we obtain the upper bound for the probability in question:

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

yielding immediately the proposed upper bound:

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

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

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

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

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

This proves the theorem, for $\nu \in \mathring{\Gamma}$ is arbitrary as long as $\Sigma_\nu \subset \Sigma_\mu$ and $H(\nu|\mu) = \infty$ whenever $\Sigma_\nu \not\subset \Sigma_\mu$. □

As an application of Sanov's theorem, a version of Cramér's theorem concerning the large deviations of the empirical mean of i.i.d. random variables is proved. Let $f : \Sigma \to \mathbb{R}$ be a function and consider the large deviation principle of the empirical average $\hat{S}_n = \sum_{i=1}^n f(Y_i)$. For convenience, we identify f as $\mathbf{f} \in \mathbb{R}^\Sigma$ so we may write $\hat{S}_n = \langle L_n^{\mathbf{Y}}, \mathbf{f} \rangle$ as an inner product. Sanov's theorem then imply the following theorem.

Theorem 2.7 (Cramér)**.** *For any set $A \subseteq \mathbb{R}$,*

$$- \inf_{\nu \in \mathring{A}} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log P_\mu\left(\hat{S}_n \in A\right)$$
$$\leq \limsup_{n \to \infty} \frac{1}{n} \log P_\mu\left(\hat{S}_n \in A\right) \leq - \inf_{\nu \in \Gamma} I(x),$$

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

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

where

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