# Lecture-11: Correlated Variables

## 1 Asymptotic equipartition

We can count the number of configurations $X \in \mathcal{X}^N$ with either a given type $q(x)$ or a given empirical average $\bar{f}$ of some observable $f : \mathcal{X}^N \to \mathbb{R}$. We say that 'there are approximately $2^{NH(q)}$ sequences with type $q$'.

**Proposition 1.1.** *The number $\mathcal{N}_{K,N}$ of sequences $X \in \mathcal{X}^N$ which have a type belonging to the compact set $K \subset \mathcal{M}(X)$ behaves as $\mathcal{N}_{K,N} \doteq 2^{NH(q^*)}$, where $q^* = \arg\max\{H(q) : q \in K\}$.*

*Proof.* Let $p(x)$ be the uniform probability distribution on $\mathcal{X}$, then $P_p\{q \in K\} = \frac{\mathcal{N}_{K,N}}{|\mathcal{X}|^N}$. From Sanov's theorem, we know that $P_p\{q \in K\} \doteq e^{-ND(q^* \| p)}$. Since $p$ is uniform over $\mathcal{X}$, we have $D(q\|p) = \mathbb{E}_q \log_2 q(x) + \mathbb{E}_q \log_2 |\mathcal{X}| = -H(q) + \log_2 |\mathcal{X}|$. Combining the above results, we get that

$$\mathcal{N}_{K,N} = \mathbb{P}_p(q \in K) 2^{N \log_2 |\mathcal{X}|} \doteq 2^{NH(q^*)}.$$

□

As a consequence of Sanov's theorem, we know that for any i.i.d. random sequence $X \in \mathcal{X}^N$ of $N$ with the common probability distribution $p(x)$, the most probable type is $p(x)$ itself, and that deviations are exponentially rare in $N$. We expect that almost all the probability will be concentrated into sequences that have a type close to $p(x)$ in some sense. On the other hand, because of the above proposition, the number of such sequences is exponentially smaller than the total number of possible sequences $|\mathcal{X}|^N$.

Let's define what is meant by a sequence having a type 'close to $p(x)$'. Given a sequence $X$, we introduce the quantity empirical entropy defined as

$$r(X) \triangleq -\frac{1}{N} \log_2 P_N(X) = -\frac{1}{N} \sum_{i=1}^{N} \log_2 p(X_i).$$

Clearly, $\mathbb{E}[r(X)] = H(p)$.

**Definition 1.2.** A random sequence $X \in \mathcal{X}^N$ is called **$\epsilon$-typical** iff $|r(X) - H(p)| \leqslant \epsilon$.

**Theorem 1.3.** *Let $T_{N,\epsilon}$ be the set of $\epsilon$-typical sequences, then the following hold.*

*(i)* $\lim_{N\to\infty} P\{X \in T_{N,\epsilon}\} = 1$.

*(ii) For large enough $N$, we have* $2^{N(H(p)-\epsilon)} \leqslant |T_{N,\epsilon}| \leqslant 2^{N(H(p)+\epsilon)}$.

*(iii) For any $x \in T_{N,\epsilon}$, we have* $2^{-N(H(p)+\epsilon)} \leqslant P\{X = x\} \leqslant 2^{-N(H(p)-\epsilon)}$.

*Proof.* From Corollary to the Sanov's theorem, we have $P\{X \notin T_{N,\epsilon}\} \doteq e^{-NI}$, where the exponent

$$I = \min\{D(q\|p) : q \notin K\}, \text{ where } K = \{q \in \mathcal{M}(\mathcal{X}) : |\mathbb{E}_q r(X) - H(p)| \leqslant \epsilon\}.$$

(i) That is, we can write $K = \{q \in \mathcal{M}(\mathcal{X}) : |D(q\|p) + H(q) - H(p)| \leqslant \epsilon\}$. It follows that $p \in K$ and hence $I > 0$. Therefore, $\lim_{N\to\infty} P\{X \notin T_{N,\epsilon}\} = 0$.

(ii) The compact set of types of the sequence $X \in T_{N,\epsilon}$ is given by $K$. Let $q \in K$, then we have $|D(q\|p) + H(q) - H(p)| \leqslant \epsilon$, which implies that $|H(q) - H(p)| \leqslant \epsilon$. From the previous proposition, we have

$$2^{N(H(p)-\epsilon)} \leqslant |T_{N,\epsilon}| \leqslant 2^{N(H(p)+\epsilon)}.$$

(iii) Recall that $T_{N,\epsilon} = \left\{x \in \mathcal{X}^N : |r(x) - H(p)| \leqslant \epsilon\right\}$, and hence for any $x \in T_{N,\epsilon}$ we have $P\{X = x\} = 2^{-Nr(x)}$. The result follows.

□

**Definition 1.4.** The behavior described in the above theorem is called the **asymptotic equipartition property**.

# 2 Correlated variables

For independent random variables in finite spaces, the probability of a large deviation is easily computed by combinatorics. We now present some general result for large deviations of non-independent random variables using Legendre transforms and saddle point methods.

## 2.1 Legendre transformation

Consider the joint distribution of a set of $N$ random variables over the configuration space $\mathcal{X}^N$ given by

$$P_N(x) = P_N(x_1, \dots, x_N),\ x \in \mathcal{X}^N.$$

Let $f : \mathcal{X} \to \mathbb{R}$ be a real valued function, and its empirical average

$$\overline{f}(X) = \frac{1}{N}\sum_{i=1}^{N} f(X_i).$$

We showed that finite fluctuation of $\overline{f}$ is exponentially unlikely for i.i.d. random variables. We will show the same for 'weakly correlated' random variables. In particular, let $P_N(x)$ be the Boltzmann distribution of $N$ particle interacting system, and $f$ be a macroscopic observable. Then, we will show that the relative fluctuation of macroscopic observables is small.

**Assumption 2.1.** The distribution of $\overline{f}$ follows a **large-deviation principle**, meaning that the asymptotic behavior of the distribution at large $N$ is

$$P_N(\overline{f}) \doteq e^{-NI(\overline{f})},$$

with a **rate function** $I(\overline{f}) \geqslant 0$.

In order to determine the rate function, a useful method is to 'tilt' the measure $P_N(\cdot)$ in such a way that the rare events responsible for $O(1)$ fluctuations of $\overline{f}$ become likely.

**Definition 2.2.** The **logarithmic moment-generating function** of $\overline{f}$ is defined as

$$\psi_N(t) \triangleq \frac{1}{N}\log \mathbb{E}\left[e^{Nt\overline{f}(x)}\right],\ t \in \mathbb{R}.$$

When the large-deviation principle holds, we can evaluate the large-$N$ limit of $\psi_N(t)$ using the saddle point method

$$\psi(t) \triangleq \lim_{N\to\infty} \psi_N(t) = \lim_{N\to\infty} \frac{1}{N}\log \int d\overline{f} e^{-NI(\overline{f})} e^{Nt\overline{f}}.$$

It follows that $\psi(t)$ is the Legendre transform of $I(\overline{f})$

$$\psi(t) = \sup\left\{t\overline{f} - I(\overline{f}) : \overline{f} \in \mathbb{R}\right\}.$$

Since $\psi(t)$ is supremum of affine functions in $t$, it is convex in $t$. Therefore, we can invert the Legendre transform as

$$I_{\psi}(\overline{f}) = \sup\left\{t\overline{f} - \psi(t) : t \in \mathbb{R}\right\},$$

where $I_{\psi}(\overline{f})$ is the convex envelope of $I(\overline{f})$. This procedure is useful when computing $\psi(t)$ is easier than the probability distribution $P_N(\overline{f})$. The above method informally captures the essence of Gärtner-Ellis theorem(explained in the following lecture).

**Example 2.3.** Consider the one-dimensional Ising model, with external magnetic field $B = 0$.We have $x_i = \sigma_i \in \{+1, -1\}$, and $P_N(\sigma) = exp[-\beta E(\sigma)]/Z$ the Boltzmann distribution with energy function

$$E(\sigma) = -\sum_{i=1}^{N-1} \sigma_i \sigma_{i+1}$$

We want to compute the large deviation properties of the magnetization.

$$m(\sigma) = \frac{1}{N}\sum_{i=1}^{N} \sigma_i$$

In order to evaluate probability of a large fluctuation of $m$ we can apply the moment generating function of $m$.

$$\begin{aligned}\psi_N(t) &= \frac{1}{N}\log \mathbb{E}\left[e^{Ntm(\sigma)}\right] \\ &= \frac{1}{N}\log \mathbb{E}\left[exp[t\sum_{i=1}^{N}\sigma_i]\right]\end{aligned}$$

The above expectation is taken over Boltzmann distribution. Hence,

$$\psi_N(t) = \frac{1}{N}\log\frac{\sum_{\sigma} exp(\beta\sum_{i=1}^{N-1}\sigma_i\sigma_{i+1} + t\sum_{i=1}^{N}\sigma_i)}{\sum_{\sigma} exp(\beta\sum_{i=1}^{N-1}\sigma_i\sigma_{i+1})}$$

$$= \frac{1}{N}\log\frac{z_n(\beta, \frac{t}{\beta})}{z_n(\beta, 0)}$$

Using the relation between free entropy($\phi(\beta)$) and partition function($z(\beta)$) for limiting value of N, we get

$$\psi(t) = \phi(\beta, \frac{t}{\beta}) - \phi(\beta, 0).$$

In one of the previous lectures we have derived the following formal expression for free entropy in one dimensional Ising model.

$$\phi(\beta, B) = \log\left[e^{\beta}\cosh(\beta B) + \sqrt{e^{2\beta}\sinh^2(\beta B) + e^{-2\beta}}\right]$$

Therefore,

$$\begin{aligned}\psi(t) &= \log\frac{\left[e^{\beta}\cosh t + \sqrt{e^{2\beta}\sinh^2 t + e^{-2\beta}}\right]}{e^{\beta} + e^{-\beta}} \\ &= \log\frac{\left[\cosh t + \sqrt{\sinh^2 t + e^{-4\beta}}\right]}{1 + e^{-2\beta}}\end{aligned}$$

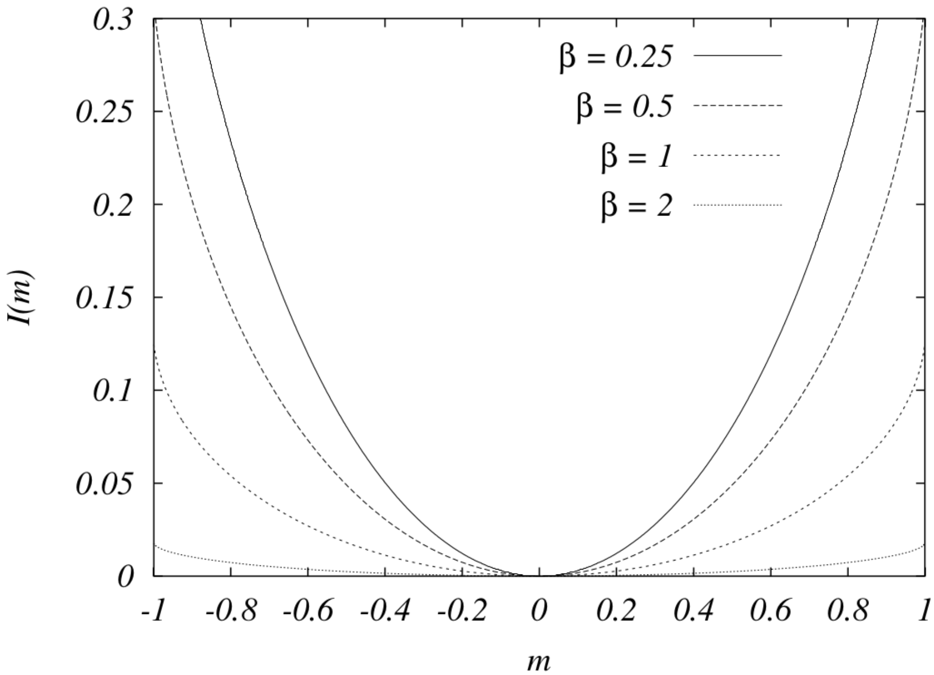


Figure 1: Rate function for the magnetization of the one-dimensional ising model

and rate function,

$$I_{\psi}(m) = \sup_{t \in \mathbb{R}} \{tm - \psi(t)\}$$

**Inference from figure 1:** As $\beta$ increases i.e as temperature decreases the probability of large fluctuations increases.