Lecture-10: Many Random Variables

1 Qualitative view of Random Variables

Physical systems of particles can be modelled by assigning a random variable to state of each particles. Here we assume the particles have no interaction or interaction is minimal. Thus for N particles we have an N length vector $x = (x_1, x_2, \ldots, x_N)$ with probability

$$P_N(x) = P_N(X_1, \ldots, x_N). \tag{1}$$

An instance of this would be Boltzmann distribution for a physical system with N degrees of freedom. Shannon Entropy of this distribution is

$$H_N = -\mathbb{E} log P_N(x) \tag{2}$$

Typically H_N grows linearly with N for large N. So entropy per variable $h_n = H_N/N$ has a finite limit;so we define a quantity h.

Definition 1.1. $h : \lim_{N\to\infty} h_N$

We will be interested in defining the quantity

$$r(x) = 1/N log \left[1/P_N(x)\right] \tag{3}$$

to characterize any particular realization $\underline{x}$. We would like to know how close is $r(x)$ to h_N. In most cases $r(x)$ is peaked around $r = h_N$. Often probability distribution of $r(x)$ behave exponentially with

$$P\{r(x) \approx \rho\} = e^{-NI(\rho)} \tag{4}$$

where $I(\rho)$ has minimum at $\rho = h$. and $I(h) = 0$ From this observation and (3) we have

$$P_N(x) = e^{-Nh} \tag{5}$$

Total probability of realization x with $r(x) \approx h$ is 1. This will imply that number of configurations with major contribution to probability is e^{Nh}.This in general is a small number when compared to total number of configurations $\mathcal{X}^N$.

$$\text{Number of typical configurations} = e^{Nh} << \mathcal{X}^N \tag{6}$$

This has a lot of implications.For example if we want to estimate an observable $\mathcal{O}$ we can sample configuration and it will give good estimates because major part of total probability is distributed among typical configurations. But it takes orderof $e^{N(\log|\mathcal{X}|)-h}$ time to generate a sample. Monte Carlo methods provide better solution for this problem

2 Large deviations for independent variables

Theorem 2.1. *(Sanov) Let $s1, \ldots s_N \in \mathcal{X}$ be N i.i.d random variables drawn from the probability distribution $p(x)$, and let $K \subset \mathcal{M}(\mathcal{X})$ be a compact set of probability distributions over $\mathcal{X}$. If q is the type of $(s1, \ldots, s_N)$ then*

$$\mathbb{P}\left[q \in K\right] = e^{-ND(q^* \| p)} \tag{7}$$

where $q^* = arg\ min_{q\in K} D(q \parallel p)$, and $D(q \parallel p)$ is the Kullback-Leibler divergence.

Proof. Now for a particular q'

$$\begin{aligned} (P(q_s = q')) &= \prod_{x \in \mathcal{X}} P(Nq'(x)\text{times } x \text{ appears in s}) \\ &= \binom{N}{Nq'(x_1), Nq'(x_2), \ldots, Nq'(x_{|\mathcal{X}|})} \prod_{x \in \mathcal{X}} p(x)^{Nq'(x)} \\ &= \binom{N!}{Nq'(x_1)!, Nq'(x_2)!, \ldots, Nq'(x_{|\mathcal{X}|})!} \prod_{x \in \mathcal{X}} p(x)^{Nq'(x)} \end{aligned} \tag{8}$$

Using Stirlings equation

$$\log_2 n! = n \log_2 n \tag{9}$$

$$n! = 2^{n \log_2 n} \tag{10}$$

$$\frac{N!}{(Nq'(x_1))!(Nq'(x_2))!\ldots(Nq'(x_{|\mathcal{X}|}))!} = 2^{N \log N - Nq_1 \log Nq_1 - Nq_2 \log Nq_2 \ldots Nq_{|\mathcal{X}|} \log Nq_{|\mathcal{X}|}} \tag{11}$$

$$\prod_{x \in \mathcal{X}} p(x)^{Nq'(x)} = \prod_{x \in \mathcal{X}} 2^{Nq'(x) \log p(x)} \tag{12}$$

We have KL Divergence D as

$$D(q \| p) = \sum_x q(x) \log \frac{q(x)}{p(x)} \tag{13}$$

From 8, 10, 11, 12 we have

$$P(q_s = q') = \prod_{x \in \mathcal{X}} 2^{-ND(q \| p)} \tag{14}$$

By approximating with leading term in exponential we have

$$P(q_s = q') = e^{-ND(q* \| p)} \tag{15}$$

□

Example 2.2. A simple model of a column of the atmosphere is obtained by considering N particles in the earth's gravitational field.The state of particle $i \in \{1, 2, \ldots, N\}$ is given by a single coordinate $z_i \geq 0$ which measures its height with respect to level. To simplify , we make assumption z_i is integer.

$$E = \sum_{i=1} N z_i \tag{16}$$

Consider a configuration : $z_1, \ldots, z_N$. Its type can be interpreted as:

$$\rho(z) = 1/N \sum_{i=1}^{N} \mathbb{1}_{\{z = z_i\}} \tag{17}$$

$$\rho_{eq}(z) = \langle \rho(z) \rangle = (1 - e^{-\beta}) e^{-\beta z} \tag{18}$$

Using Boltzmann probability distribution; expected density profile can be computed as follows.

$$\text{Partition function } Z(\beta) = \sum_{x \in \mathfrak{X}} e^{-E(x)} = \sum_{z=0}^{\infty} e^{-\beta z} = \frac{1}{1 - e^{-\beta}}. \tag{19}$$

So probability distribution follows

$$\mu_\beta(x) = \frac{1}{Z(\beta)} e^{-\beta E(x)} = (1 - e^{-\beta}) e^{-\beta z} \tag{20}$$

We compute the probability of getting a general exponential density profile with parameter λ:

$$\rho_\lambda(z) = (1 - e^{-\lambda})e^{-\lambda z} \tag{21}$$

For that we find the KL divergence between ρ_λ and ρ_{eq}

$$D(\rho_\lambda \| \rho_{eq}) = \sum_z \rho_\lambda(z) \log \frac{\rho_\lambda(z)}{\rho_{eq}(z)} = \log \frac{1 - e^{-\lambda}}{1 - e^{-\beta}} + \frac{\beta - \lambda}{e^\lambda - 1} \tag{22}$$

We now plot KL divergence as a function of λ in figure 1.

$$I_\beta(\lambda) = D(\rho_\lambda \| \rho_{eq}) \tag{23}$$

It can be noted from the figure 1 that small values of λ are very rare.

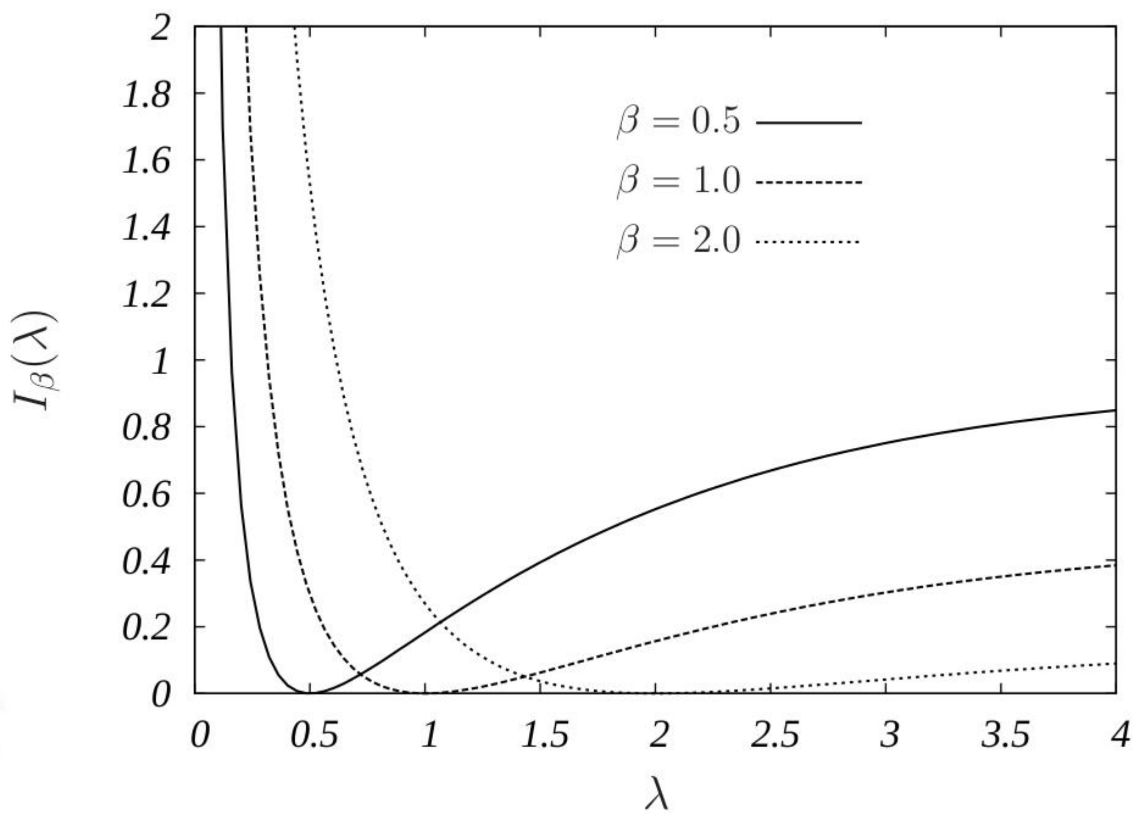

Figure 1: Equilbrium density profile

3 How typical is an empirical average?

The empirical average of a measurement is given by:

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(s_i) \tag{24}$$

Corollary 3.1. *Let* $s1, \ldots, s_N$ *be* N *i.i.d random variables drawn from a probability distribution* $p(.)$. *Let* $f : \mathcal{X} \to \mathbb{R}$ *be a real-valued function and let* $\bar{f}$ *be its empirical average. If* $A \subset R$ *is a closed interval of the real axis, then*

$$\mathbb{P}[\bar{f} \in A] = e^{[-NI(A)]}, \tag{25}$$

where

$$I(A) = \min_q \left[D(q\|p)| \sum_{x\in\mathcal{X}} q(x)f(x) \in A\right] \tag{26}$$

Proof. We note that $\bar{f}$ is related to the type of sequence $x_1, x_2, \ldots x_n$ as $\bar{f} = \sum_x q(x)f(x)$. Keeping in mind Sanov theorem , we define a set as follows

$$K = \left\{q \in \mathcal{M}(\mathcal{X})| \sum_{x\in\mathcal{X}} q(x)f(x) \in A\right\} \tag{27}$$

Then the result follows directly from Sanov by camparing K with set in Sanov and $I(A)$ with $D(q^*\|p)$. □

Example 3.2. Let $s_1, ..., s_N$ be N i.i.d random variablesdrawn from a probability distribution $p(.)$ with bounded support. Show that, to leading exponential order,

$$\mathbb{P}\{s_1 + \cdots + s_N \leq= 0\} = \{\inf_{z\geq 0} \mathbb{E}e^{-zs_1}\}^N \tag{28}$$

Example 3.3. We look at N particles in a gravitational field, and consider the average height of the particles

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i. \tag{29}$$

Expected value of this quantity is

$$\mathbb{E}\bar{z} = z_{eq} = (e^\beta - 1)^{-1}. \tag{30}$$

The probability of a fluctuation in $\bar{z}$ is easily computed using the obove corollary.For $z > z_{eq}$ we obtain

$$\mathbb{P}(z > \bar{z}) = e^{-NI(z)}, \tag{31}$$

where

$$I(z) = \big(1+z\big)\log\big(\frac{1+z_{eq}}{1+z}\big) + z\log\big(\frac{z}{z_{eq}}\big) \tag{32}$$