LARGE DEVIATIONS in a NUTSHELL

We briefly discuss the basic aspects of the Large Deviations theory. Roughly speaking, one can say that the Large Deviations theory is a generalization of the two most important limit theorems of probability theory, i.e. the law of large numbers, and the central limit theorem.

Consider a simple example: a sequence of independent tosses of an unfair coin. The possible outcomes are head (+1) or tail (−1). Denote the possible result of the n-th toss by x_n, where head has probability α, and tail has probability $1 - \alpha$. Let Y_N be the mean value after N tosses,

$$Y_N = \frac{1}{N} \sum_{n=1}^{N} x_n . \tag{1}$$

If $N \gg 1$, a straightforward application of the central limit theorem gives

$$Prob(Y < Y_N < Y + dY) \simeq P_N(Y) dY \sim e^{-\frac{N(Y - \langle Y \rangle)^2}{2\sigma^2}} dY \ , \tag{2}$$

where $\langle Y \rangle = 2\alpha - 1$, $\sigma^2 = 4\alpha(1 - \alpha)$, $P_N(Y)$ is the pdf of the random variable (1). Since Y_N lies between −1 and 1, it is easy to realize that (2) is accurate only for small deviations of Y_N from the mean (namely $|Y_N - \langle Y \rangle| < O(1/\sqrt{N})$).

A natural way to introduce the large deviations and show their deep relation with the concept of entropy is to perform a combinatorial computation. The number of ways in which K heads occur in N tosses is $N!/[K!(N - K)!]$, therefore, the exact binomial distribution yields

$$Prob(Y_N = \frac{2K}{N} - 1) = \frac{N!}{K!(N - K)!} \alpha^K (1 - \alpha)^{N-K} \ . \tag{3}$$

Using Stirling's approximation and writing $K = pN$ and $N - K = (1 - p)N$ one obtains

$$P_N(Y = 2p - 1) \sim e^{-NI(\alpha,p)} \ , \tag{4}$$

where

$$I(\alpha, p) = p \ln \frac{p}{\alpha} + (1 - p) \ln \frac{1 - p}{1 - \alpha} \ . \tag{5}$$

Note that $I(\alpha, p)$ is called "relative entropy" (or Kullback-Leibler divergence), and $I(\alpha, p) = 0$ for $\alpha = p$, while $I(\alpha, p) > 0$ for $\alpha \neq p$. It is easy to repeat the argument for the multinomial case, where $x_1, ..., x_N$ are independent variables that take m possible different values $a_1, a_2, ..., a_m$ with probabilities $\pi_1, \pi_2, ..., \pi_m$. In the limit $N \gg 1$, the probability of observing the frequencies $f_1, f_2, ..., f_m$ is

$$Prob_N(\{f_j\} \simeq \{p_j\}) \sim e^{-NI(\{p\},\{\pi\})}$$

where

$$I(\{p\}, \{\pi\}) = \sum_{j=1}^{m} p_j \ln \frac{p_j}{\pi_j} \ ,$$

is called "relative entropy" of the probability $\{p\}$, with respect to the probability $\{\pi\}$. Such a quantity measures the discrepancy between $\{p\}$ and $\{\pi\}$ in the sense that $I(\{p\}, \{\pi\}) = 0$ if and only if $\{p\} = \{\pi\}$, and $I(\{p\}, \{\pi\}) > 0$ if $\{p\} \neq \{\pi\}$.

From the above computation one undestands that it is possible to go beyond the central limit theory, and to estimate the statistical features of extreme (or tail) events, as the number of observations grows without bounds. Writing $I(p, \alpha)$ in terms of $Y = 2p - 1$, Eq. (4) becomes

$$P_N(Y) \sim e^{-NS(Y)} \ , \qquad (6)$$

with

$$S(Y) = \frac{1+Y}{2} \ln \frac{1+Y}{2\alpha} + \frac{1-Y}{2} \ln \frac{1-Y}{2(1-\alpha)} \ .$$

The term $S(Y)$ is called Cramer's function. Of course, for p close to α, i.e. $Y \simeq \langle Y \rangle$, a Taylor expansion reproduces the central limit theorem (2).

However Eq. (6) is more general, and can be derived in different ways, which show how the shape of $S(Y)$ is related to the behaviour of the moments of the variable x. In particular it is possible to see that $S(Y)$ can be expressed in terms of a Legendre transform:

$$S(Y) = \sup_q \Big[qY - L(q) \Big] \ , \qquad (7)$$

where $L(q)$ is the "Cumulants Generating Function" given by

$$L(q) = \ln \langle e^{qx} \rangle \ . \qquad (8)$$

Let us sketch the argument. Consider the quantities $\langle e^{qNY_N} \rangle$ which can be written in two ways:

$$\langle e^{qNY_N} \rangle = \langle e^{qx} \rangle^N = e^{NL(q)}$$

$$\langle e^{qNY_N}\rangle = \int e^{qNY_N} P_N(Y_N)dY_N \sim \int e^{[qY-S(Y)]N}dY$$

whose identification leads to

$$\int e^{[qY-S(Y)]N}dY \sim e^{NL(q)} \tag{9}$$

and, in the limit of large N, using the steepest descent method, one has

$$L(q) = \sup_Y \Big[qY - S(Y)\Big] \ , \tag{10}$$

which is the inverse of ((7)). Since it is possible to show that $S(Y)$ is convex, Equations (7) and (10) are fully equivalent.

For the more general and interesting case of dependent variables, $L(q)$ is defined as

$$L(q) = \lim_{N\to\infty} \frac{1}{N} \ln\langle e^{q\sum_{n=1}^{N} x_n}\rangle \ ,$$

and (7) is exact if $S(Y)$ is convex, otherwise Eq. (7) gives the convex envelop of the correct $S(Y)$.

Let us note that the Cramer function must obey some constraints:

a) $S(Y) > 0$ for $Y \neq \langle Y\rangle$;

b) $S(Y) = 0$ for $Y = \langle Y\rangle$;

c) $S(Y) \simeq (Y - \langle Y\rangle)^2/(2\sigma^2)$, where $\sigma^2 = \langle(x - \langle x\rangle)^2\rangle$, if Y is close to $\langle Y\rangle$.

Of course a) and b) are consequences of the law of large numbers, and c) is nothing but the central limit theorem.