Large deviation theory

Rajesh Singh

The Institute of Mathematical Sciences, Chennai, India

September 16, 2017

Outline

Characteristic function

Characteristic function of a probability distribution $P(x)$ is

$$\tilde{P}_X(k) = \langle e^{kX} \rangle \tag{1}$$

Expanding e^{kX} in a power series, we have

$$\tilde{P}_X(k) = \sum_n \frac{(k)^n}{n!} \langle X^n \rangle = 1 + k\langle X \rangle + \frac{k^2 \langle X^2 \rangle}{2!} + \frac{k^3 \langle X^3 \rangle}{3!} + \ldots \tag{2}$$

where $\mu_N = \langle X^N \rangle$ is the nth moment. Thus, above is also called the moment generating function.

Cumulant generating function

The logarithm of characteristic function

$$K_X(k) = \ln \tilde{P}_X(k) = \sum_n \frac{(k)^n}{n!} \langle X^N \rangle_c \tag{3}$$

First two cumulants and moments are related as

$$\langle X \rangle_c = \langle X \rangle, \qquad \langle X^2 \rangle_c = \langle X \rangle^2 - \langle X^2 \rangle \tag{4}$$

where $\langle X^N \rangle_c$ is the n-th cumulant.

- To see an analogy in physics note that

$$\langle X \rangle_c^n = \partial_k^n K_X(k)|_{k=0}. \tag{5}$$

Cumulants

- n-th cumulant of random variables $X_1, \ldots, X_n$ measures the interaction of the variables which is genuinely of n-body.
- For Gaussian variables, the fourth moment is

  $$\langle X_1 X_2 X_3 X_4 \rangle = \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle + \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle + \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle$$

  The cumulant, on the other hand is difference of LHS and RHS, and is thus zero for a Gaussian distribution.
- Moments: correlations; cumulants: connected correlation
- $\tilde{P}_X(k) \sim$ partition function, while $K_X \sim$ free energy

Cumulants

Cumulants are additive for independent variables X and Y

$$K_{X+Y}(t) = K_X(t) + K_Y(t). \tag{6}$$

- 1st cumulant - Mean (describes central value)
- 2nd cumulant - Variance (describes dispersion)
- 3rd cumulant - Skewness (describes asymmetry)
- 4th cumulant - Kurtosis (describes peakedness)

Gaussian distribution

- The PDF is

$$P(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in [-\infty;\infty] \quad (7)$$

- mean=μ, variance=σ^2. If $\mu = 0$ and $\sigma = 1$, the distribution is called the standard normal distribution.
- Mode : value that appears most often.
- Median: value separating the higher half from lower half.
- For Gaussian distribution, mean, median and mode coincide!

Law of large numbers

- Sum of random variables

$$Y_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

- It is given that each X_i has mean μ and standard deviation σ.
- As n increases, the average of the sum approaches the mean of the distribution. More is the variation in X, higher is the n for the average of the sum to reach the mean.
- What is $P_Y(X)$?
- when is δY small ?
- statistics of rare events when such fluctuations are "large"

Central limit theorem

- Let $\{X_n\}$ be a sequence of independent identically distributed (iid) random variables.
- CLT says that the sum of iid random variables approaches a Gaussian distribution, as the sample size increases.
- The only condition is that μ and σ do not diverge!
- Proof using characteristic function

Central limit theorem

Choose

$$Z_n = \frac{X_1 + X_2 + ... + X_n - \mu n}{\sigma \sqrt{n}} \tag{8}$$

$$\begin{aligned} \tilde{P}(k) &= \int dz\, P_{Z_n}(z) e^{kz} = \int dz \int \prod_i dX_i P(X_i) e^{kz} \delta(z - Z_n) \\ &= \left[\int dX\, P(X) e^{-kX/\sigma\sqrt{n}} \right]^n e^{k\mu n/\sigma\sqrt{n}} \end{aligned}$$

This can be trivially simplified to

$$\ln \tilde{P}_Y(k) = n \ln \tilde{P}_X(k/\sigma\sqrt{n}) + \frac{kn\mu}{\sigma} \tag{9}$$

Central limit theorem

- We now use the definition of cumulant generating function to obtain

$$\ln \tilde{P}_{Z_n}(k) = n\frac{(k/\sigma\sqrt{n})^2}{2!}\sigma^2 = \frac{k^2}{2!}; \quad \text{as } n \to \infty$$

- Thus, we approach a normal distribution

$$P(Z_n) = \frac{1}{\sqrt{2\pi}} e^{-Z_n^2/2}$$

- More generally, as n gets bigger the distribution of the sum of random variables $Y_n = \frac{1}{n}\sum_i X_i$ will always converge to a Gaussian distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Large deviation theory

- The large deviation theory is a generalization of the CLT
- Given a random variable Y_n, we seek what is the probability density of $P(Y_n = y)$

[1] H. Touchette, Phys Rep, 2009.

Large deviation theory

- The large deviation theory is a generalization of the CLT
- Given a random variable Y_n, we seek what is the probability density of $P(Y_n = y)$
- Large deviation principle [1]

$$P(Y_n = y) \approx e^{-nI(y)}. \tag{10}$$

- What it means is

$$\lim_{n\to\infty} -\frac{1}{n} \ln P(Y_n = y) = I(y). \tag{11}$$

- The rate function $I(y) \geq 0$ is positive definite.
- Agenda: Prove the existence of a LDP and derive $I(y)$

[1] H. Touchette, Phys Rep, 2009.

Sum of Gaussian random number

- Sum of random variables

$$Y_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

- Probability density function (pdf) of Y_n, sum of Gaussian distribution,

$$p(Y_n = y) = \sqrt{\frac{n}{2\pi\sigma}} e^{-\frac{n(y-\mu)^2}{2\sigma^2}} = e^{-nI(l)}. \tag{12}$$

- Scaling law

$$P(Y_n = y) \approx e^{-nI(y)}. \tag{13}$$

$I(y) = \frac{(y-\mu)^2}{2\sigma^2}$ is the rate function.

Sum of Exponential random number

- Sum of exponentially distributed random number
- Large deviation principle says

$$P(Y_n = y) \approx e^{-nI(y)}. \tag{14}$$

$$P(x) \approx \frac{1}{\mu} e^{-x/\mu}. \tag{15}$$

Sum of Exponential random number

- Sum of exponentially distributed random number
- Large deviation principle says

$$P(Y_n = y) \approx e^{-nI(y)}. \tag{14}$$

$$P(x) \approx \frac{1}{\mu} e^{-x/\mu}. \tag{15}$$

- The rate function in this case

$$I(y) = \tfrac{y}{\mu} - 1 - \ln \tfrac{y}{\mu}.$$

Spin system

- Sum of spins is the magnetization

$$M = \frac{1}{n} \sum S_i$$

- Number of states with a magnetization value of m is given as

$$\Omega(m) = \frac{n!}{((1-m)n/2)!((1+m)n/2)!}$$

- Then it can be written as $\Omega(m) = e^{ns(m)}$

$$s(m) = -\frac{1-m}{2} \ln \frac{1-m}{2} - \frac{1+m}{2} \ln \frac{1+m}{2}$$

- Thus, the enropy plays the role of rate function

Quick summary

- Large deviation principle (LDP): $P(Y_n = y) \approx e^{-nI(y)}$.
- Generalization of CLT,
- Applicable for uncorrelated and correlated process
- How to calculate rate functions?

Quick summary

- Large deviation principle (LDP): $P(Y_n = y) \approx e^{-nI(y)}$.
- Generalization of CLT,
- Applicable for uncorrelated and correlated process
- How to calculate rate functions?

The Gärtner–Ellis Theorem

- Define the *scaled cumulant generating function* as

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nkY_n}\rangle \tag{16}$$

- The Gärtner–Ellis Theorem states that, if $\lambda(k)$ exists and is differentiable, then Y_n satisfies a large deviation principle, i.e.,

$$P(Y_n \in dy) \approx e^{-nI(y)} dy. \tag{17}$$

$$I(y) = \sup_{k\in\mathbb{R}} \{ky - \lambda(k)\} \tag{18}$$

Cramér's Theorem

- Apply Gärtner–Ellis Theorem to

$$Y_N = \frac{1}{N}\sum_N X_i. \tag{19}$$

yields Cramér's Theorem, where X is IID.

$$\lambda(k) = \lim_{n\to\infty}\frac{1}{n}\ln\langle e^{k\Sigma X_i}\rangle = \lim_{n\to\infty}\frac{1}{n}\ln\prod\langle e^{kX_i}\rangle = \ln\langle e^{kX}\rangle. \tag{20}$$

Varadhan's Theorem

- If Y_n satisfies a large deviation principle with rate function $I(y)$, then $\lambda(k)$ is the Legendre-fenchel transform of $I(y)$

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nkA_n}\rangle = \sup_a\{ky - I(y)\}. \tag{21}$$

- An arbitrary continuous function f of Y_n yield Varadhan's theorem

$$\lambda(f) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nf(A_n)}\rangle = \sup_a\{f(y) - I(y)\}. \tag{22}$$

Gaussian distribution

- For Gaussian distribution

$$
\begin{aligned}
\langle e^{kX} \rangle &= \int dx\, \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} e^{kx} = e^{k\mu + \sigma^2 k^2/2} \\
\lambda(k) &= k\mu + \frac{\sigma^2 k^2}{2}
\end{aligned}
$$

Then

$$
I(y) = k(y)y - \lambda(k(y)) = \frac{(y-\mu)^2}{\sigma^2},
$$

where $k(y)$ is the maxima of $ky - \lambda(k)$, with $\lambda'(k(y)) = y$.

Exponential distribution

- For Exponential distribution

$$\langle e^{kX} \rangle = \int dx \, \frac{1}{\mu} e^{-x/\mu} e^{kx} = e^{k\mu + \sigma^2 k^2/2}$$
$$\lambda(k) = -\ln(1 - \mu k),$$

this implies

$$I(y) = \tfrac{y}{\mu} - 1 - \ln \tfrac{y}{\mu}.$$

Langevin equation

- Motion of a Brownian colloid

$$\dot{x} = \mu F + \sqrt{2D}\xi$$

  Here ξ is zero mean, unit variance delta-correlated noise.
- The probability distribution is then

$$P(x) \approx e^{-J(x)/4D}, \qquad J(x) = \int_0^t (\dot{x} - \mu F)^2 \, dt$$

- Here J is sometimes called entropy of the path[2] or the action functional[3] or.

---

[2]Donsker and Varadhan 1983
[3]Freidlin, A.D. Wentzell 1984

Ornstein-Uhlenbeck process

- Consider OUP

$$\dot{x} = -\alpha x + \sqrt{2D}\xi$$

- The most probable value of $J(x)$ can obtained using Euler-Lagrange equation such that $\delta J = 0$ to give

$$\partial_t \partial_{\dot{x}} L - \partial_x L = 0, \qquad L = (\dot{x} + \alpha x)^2$$

- Consider $x(-\infty) = 0$ and $x(\tau) = x$, the solution at the maxima is then $x^*(t) = xe^{\alpha(t-\tau)}$. Thus

$$L = 4\alpha x^2$$

Summary

- Large deviation principle

$$P(Y_n = y) \approx e^{-nI(y)}.$$

- It is valid for any value of y and thus considers large fluctuations
- CLT: the rate function is parabolic
- Obtain the rate function using an optimization principle