Large deviations

CMT Journal Club
May 15th, 2015 Thomas Veness

References

- Review and introduction by Hugo Touchette: arXiv:0804.0327, arXiv:1106.4146

1 Introduction

Often when considering a large number of trials, one will invoke either the law of large numbers or the central limit theorem. In a sense, the central limit theorem is a more refined version of the law of large numbers and in the same way, large deviations encapsulate more again.

If we have N independent and identically distributed random variables $\{X_i\}$, $i = 1, \ldots, N$, then we often wish to consider the distribution of the random variable

$$S_N := \frac{1}{N} \sum_{i=1}^{N} X_i \tag{1}$$

What do we already know about the probability distribution $p_{S_N}(s)$? The law of large numbers tells us that for $N \to \infty$ we converge to $E[X_i]$ almost surely. The central limit theorem tells us that $p_{S_N}(s)$ is normally distributed *near* this mean value. However, neither of these generally provide information about the tails of the distributions and this is precisely the remit of large deviations. Let us explore what happens for a couple of common distributions for $p_X(x)$

Where is this useful? Generically, one cares about this in situtations where we care not only about the mean, but also about the extreme cases which can do a lot of damage

1. Actuarial applications
2. General insurance
3. Queuing and networks

We also care about it in physics, as it allows us to go beyond predictions of mean values, which can be trivial. Consider a non-equilibrium scenario with a current flowing down a wire.

1.1 Some examples

1.1.1 Probability distributions and generating functions

We first briefly recap some results from probability theory

Definition 1.1 (Generating function)**.** For a continiuous probability distribution $p(x)$, we define the generating function

$$f(k) = \int \mathrm{d}x e^{ikx} p(x) = \sum_{n=1}^{\infty} \frac{(ik)^n \langle x^n \rangle}{n!} \tag{2}$$

And therefore the moments can be extracted from

$$\langle x^n \rangle = (-i)^n \frac{\mathrm{d}^n f}{\mathrm{d}k^n}|_{k=0} \tag{3}$$

Theorem 1.1. If we wish to find the probability distribution for a sum of random variables, we have that

$$f_{S_N}(k) = \int \mathrm{d}x_1 \dots \mathrm{d}x_N e^{i\frac{k}{N}\sum_{i=1}^{N} x_i} p(x_1) \dots p(x_N) = \left[f_X\left(\frac{k}{N}\right)\right]^N \tag{4}$$

Example 1.1 (Gaussian distribution)**.** If we have that

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

Then we can exactly calculate the distribution for p_{S_N} using our knowledge of Fourier transforms. It is immediately transparent that we end up with a Gaussian and therefore the central limit theorem is, in this case, exact.

Example 1.2 (Bernoulli distribution)**.** In this case we begin with a discrete distribution, but for large N the set of outcomes becomes dense and is appropriately described by a continuous probability distribution.

$$p(m) = q^m (1-q)^{N-m} \left(\begin{array}{c} N \\ m \end{array} \right) \tag{6}$$

Definition 1.2 (Large deviations principle)**.** We note that in both of the previous cases that the probability distributions are of the form

$$\lim_{N\to\infty} p_{S_N}(s) \approx e^{-NI(s)} \tag{7}$$

We say that the probability density function satisfies a *large deviations principle* if the following limit exists

$$\lim_{N\to\infty} -\frac{1}{N} p_{S_N}(s) = I(s) \tag{8}$$

and $I(s)$ is not everywhere 0. Such and $I(s)$ is known as the *rate function.*

2 Gärtner-Ellis theorem

Of course, these two cases are very particular and we don't know: (a) whether many distributions will even satisfy this large-deviations form and (b) how to calculate the rate function for less tractable distributions. The Gärtner-Ellis theorem tells us exactly how to do this

Theorem 2.1 (Gärtner-Ellis theorem)**.**

$$\lambda(k) := \lim_{N\to\infty} \frac{1}{N} \ln E[e^{NkS_N}] \tag{9}$$

If $\lambda(k)$ is differentiable, then S_N satisfies a large deviations principles and $I(s)$ is given by the Legendre-Fenchel transform of $\lambda(k)$ i.e.

$$I(s) = \sup_{k\in\mathbb{R}} (ks - \lambda(k)) \tag{10}$$

The proof of this is quite technical, but we can provide a heuristic proof of Varadhan's theorem

Theorem 2.2 (Varadhan's theorem)**.** If we consider some generating functional

$$W_n[f] = E[e^{nf(S_n)}] = \int \mathrm{d}s p_{S_n}(s) e^{nf(s)} \tag{11}$$

Then, if S_n satisfies an LDP, we have

$$W_n[f] \approx \int \mathrm{d}s e^{n[f(s)-I(s)]} \tag{12}$$

and, for large n, we have the saddle-point approximation that

$$W_n[f] \approx e^{n \sup_{s\in\mathbb{R}}[f(s)-I(s)]} \tag{13}$$

So, if we define

$$\lambda[f] = \lim_{n\to\infty} \frac{1}{n} \ln W_n[f] = \sup_{s\in\mathbb{R}} \{f(s) - I(s)\} \tag{14}$$

If we define $f = ks$, then we have that

$$\lambda(k) = \sup_{s\in\mathbb{R}} \{ks - I(s)\} \tag{15}$$

This transformation is invertible is $\lambda(k)$ is differentiable. This, of course, doesn't tell us a few of the key results of GE i.e. that p_{S_N} satisfies a LDP. This is, however, as close as we can get without touching GE.

Example 2.1 (Applying the Gärtner-Ellis theorem)**.** We can prove the two assertions earlier. For the binomial distribution

$$\lim_{N\to\infty}\frac{1}{N}\langle e^{km}\rangle = \ln(pe^k + 1 - p) \tag{16}$$

$$\Rightarrow I(s) = s\ln\frac{s}{p} + (1-s)\ln\frac{1-s}{1-p} \tag{17}$$

Example 2.2 (Simple application of large deviations)**.** There is an insurance company which settles say, 1 claim a day and also receives p in premiums every day. The size of the claims, however, are random and there is therefore a risk that in some long time T that the total amount collected in premiums will not exceed the claims. This is therefore a problem dealing with small probabilities concerning the sum of a large number of random variables, which is precisely the remit of large deviations!

Using the property that the rate function is convex, the probability is bounded by

$$P(x > p) \sim e^{-TI(p)} \tag{18}$$

If we demand that this is exponentially small, i.e. $= e^{-r}$, then we simply have

$$I(p) = \frac{r}{T} \tag{19}$$

As $I(p)$ is convex we can solve this. If the claims we normally distributed with mean μ and variance σ^2, then, we would have

$$p_{\text{Gauss}} = \mu + \sigma\sqrt{\frac{2r}{T}} \tag{20}$$

Which can be interpreted as the mean plus some safety margin to account for the risk.

3 Large deviations in physics

3.1 Markov processes

Imagine we have now some probability distribution

$$p(X_1,\ldots,X_N) = p(X_1)\prod_{i=1}^{N-1}\pi(X_{i+1}|X_i) \tag{21}$$

with $p(X_1)$ some initial PDF for X_1 and $\pi(X_{i+1}, X_i)$ a transition probability $X_1 \to \ldots X_N$. If we wish to consider the current

$$Q_N = \frac{1}{N}\sum_{i=1}^{N-1} f(x_i, x_{i+1}) \tag{22}$$

With $f(x,x') = 1 - \delta_{x,x'}$, then the Gärtner-Ellis theorem applies:

$$\lambda(k) = \lim_{N\to\infty} \frac{1}{N} \ln\langle e^{Nk\sum_{i=1}^{N-1} f(x_i,x_{i+1})}\rangle \tag{23}$$

Using the Markov property, we can examine

$$\langle e^{Nk\sum_{i=1}^{N-1} f(x_i,x_{i+1})}\rangle = \sum_{\{x_i\}} p(x_1)\pi(x_2|x_1)\pi(x_3|x_2)\dots\pi(x_N|x_{N-1})e^{kf(x_1,x_2)}\dots e^{kf(x_{N-1},x_N)} \tag{24}$$

If we define $\pi_k(x_i|x_j) = e^{kf(x_i,x_j)}$, then we have

$$\langle e^{Nk\sum_{i=1}^{N-1} f(x_i,x_{i+1})}\rangle = \sum_{\{x_i\}} \Pi_k(x_N|x_{N-1})\dots\Pi_k(x_2|x_1)p(x_1) \tag{25}$$

We can read this as an iterated matrix equation

$$\langle e^{Nk\sum_{i=1}^{N-1} f(x_i,x_{i+1})}\rangle = \sum_j \left(\Pi_k^{N-1}p(x_1)\right)_j \tag{26}$$

If this matrix has the appropriate properties, then it will have some spectrum of eigenvalues $\lambda_1,\dots,\lambda_N$ and this expectation value will be give by $\lambda_{\max}^{N-1} + \dots$, where the $\dots$ are exponential corrections. Therefore, we can write that

$$\lambda(k) = \ln\zeta(\tilde{\Pi}_k) \tag{27}$$

Where $\zeta(\tilde{\Pi}_k)$ is the dominant eigenvalue of the "tilted matrix" associate with Q_N, given by the elements

$$\tilde{\Pi}_k(x,x') = \pi(x|x')e^{kq(x,x')} \tag{28}$$

Example 3.1 (Simplest Markov chain)**.** Imagine we have a two-state Markov system with transition matrix

$$\pi(x|x') = \begin{cases} 1-\alpha & x = x' \\ \alpha & x \neq x' \end{cases} \tag{29}$$

Then, if we wish to consider the current

$$Q_N = \frac{1}{N}\sum_{i=1}^{N-1}(1 - \delta_{x_i,x_{i+1}}) \tag{30}$$

Using the result from before, we simply wish to calculate the largest eigenvalue of the matrix

$$\tilde{\Pi}_k = \begin{pmatrix} 1-\alpha & \alpha e^k \\ \alpha e^k & 1-\alpha \end{pmatrix} \tag{31}$$

which are simply $\lambda = 1 - \alpha \pm \alpha e^k$. This gives us that the rate function is

$$I(s) = s\ln\left[\frac{s(1-\alpha)}{\alpha(1-s)}\right] - \ln\left[1 - \alpha + \alpha\frac{s(1-\alpha)}{\alpha(1-s)}\right] \tag{32}$$

This has, intuitively, a minimum at $s = \alpha$.