1 Introduction

The lecture notes cover and integrate section G of chapter 2 of [1].

2 The main results of classical probability theory

The two theorems below give a mathematical model to describe the statistical properties of the outcomes of repeated identical experiments. The first theorem gives information about the average outcome, the second about the typical size of the fluctuations around the average.

Theorem 2.1 (*The strong law of large numbers*)**.** *Let $\{\boldsymbol{\xi}_i\}_{i=1}^n$ a sequence of independent identically distributed integrable random variables defined over the same probability space. Then if $\xi_i \stackrel{d}{=} \xi$ we have*

$$P\left(\lim_{n\uparrow\infty}\frac{\sum_{i=1}^n \boldsymbol{\xi}_i}{n} = \prec \boldsymbol{\xi} \succ\right) = 1$$

Proof. see e.g. [1] □

Theorem 2.2 (*The central limit theorem*)**.** *Let $\{\xi_i\}_{i=1}^n$ a sequence of independent identically distributed real-valued integrable random variables defined over the same probability space. Assume $\xi_i \stackrel{d}{=} \xi$ and*

$$\begin{aligned}&\prec \xi \succ = y\\ &\prec (\xi - \prec \xi \succ)^2 \succ = \sigma^2 \, > \, 0\end{aligned}$$

Set

$$S_n = \frac{1}{n}\sum_{i=1}^n \xi_i$$

Then for all $-\infty < a < b < \infty$ the limit holds

$$\lim_{n\uparrow\infty} P\left(a < \frac{S_n - y}{\prec (S_n - y)^2 \succ^{1/2}} < b\right) = \int_a^b dx\, g_{0\,1}(x) \tag{2.1}$$

where

$$g_{0\,1}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}}$$

Sketch of the proof. Consider the characteristic function

$$\prec e^{\imath\, t\, \frac{S_n - y}{\prec (S_n - y)^2 \succ}} \succ = \prec e^{\imath\, t\, \frac{\sum_{i=1}^n \xi_i - n\, y}{\sqrt{n}\sigma}} \succ = \prod_{i=1}^n \prec e^{\imath\, t\, \frac{\xi_i - y}{\sqrt{n}\sigma}} \succ = \left[\int_{\mathbb{R}} dx\, e^{\imath\, t\, \frac{x-y}{\sqrt{n}\sigma}}\, p_{\xi_1}(x)\right]^n$$

As n increases to infinity one expects the characteristic function for "small values" of q to be well approximated by a Taylor expansion of the exponential

$$\prec e^{\imath\, t\, \frac{S_n - n\, y}{\sqrt{n}\sigma}} \succ = \left[1 - \frac{t^2}{2\,n}\int_{\mathbb{R}} dx\, \frac{(x-y)^2}{\sigma^2}\, p_{\xi_1}(x) + O(\frac{1}{n^{3/2}})\right]^n \stackrel{n\uparrow\infty}{\to} e^{-\frac{t^2}{2}}$$

Thus the small wave number behavior of the characteristic function is approximated by the characteristic function of the Gaussian distribution. □

2.1 Some observations on the central limit theorem and its generalizations

The central limit theorem is often invoked in applications as it describes *universal* properties of a physical system. This means properties which depend only on a coarse characterization of the phenomena (e.g. finiteness of the fourth moment) rather than on its fine details.

2.1.1 The role of the Fourier transform

In the sketch of the proof we made use of the relation between the PDF of the a random variable and its characteristic function. Such relation becomes particularly useful when dealing with sums of random variables. Namely let

$$\zeta = \xi_1 + \xi_2$$

then

$$p_\zeta(x) = \int_{\mathbb{R}} dy_1 dy_2\, \delta(x - y_1 - y_2)\, p_{\xi_1}(y_1)\, p_{\xi_2}(y_2) = \int_{\mathbb{R}} dy\, p_{\xi_1}(x-y)\, p_{\xi_2}(y)$$

From the general properties of the Fourier transform, we know that

$$p_\zeta(x) = \int_{\mathbb{R}} \frac{dt}{2\pi}\, e^{-\imath\, t\, x} \check{p}_{\xi_1}(t) \check{p}_{\xi_2}(t)$$

Thus dealing with characteristic functions in the proof of limit theorems it is helpful because it replaces convolutions with products of Fourier transforms.

2.1.2 Domain of validity

It is important to understand that the central limit theorem is a statement concerning the *bulk* of the asymptotic distribution of

$$\zeta_n := \frac{S_n - y}{\prec (S_n - y)^2 \succ^{1/2}} \qquad n \gg 1$$

This means that we can use the predicted Gaussian distribution only to evaluate the first moments of ζ_n but not to sample the behavior of the tails of the distribution. The situation is illustrated by the following example.

- Let $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. *positive definite* random variables with density

  $$p_\xi(x) = \frac{e^{-\frac{x}{\bar{x}}}}{\bar{x}} \qquad \xi_i \stackrel{d}{=} \xi \;\; \forall\, i$$

  From this sequence we can construct the products

  $$\eta_n = \prod_{i=1}^{n} \xi_i = \bar{x}\, e^{\sum_{i=1}^{n} \psi_i} \qquad \& \qquad \psi \stackrel{d}{=} \psi_i \stackrel{d}{:=} \ln \frac{\xi}{\bar{x}}$$

  The change of variables $x = \bar{x}\, y$ then yields

  $$\begin{aligned} m &:= \prec \psi \succ = \int_0^\infty dy\, \ln y\, p_\xi(y) \,<\, \infty \\ \sigma^2 &:= \prec (\psi - m)^2 \succ = \int_0^\infty dy\, \ln^2 y\, p_\xi(y) - m^2 \,<\, \infty \end{aligned}$$

so we can apply the central limit theorem to

$$S_n[\psi] := \frac{\sum_{i=1}^n \psi_i}{n} \equiv \frac{\sum_{i=1}^n \ln \frac{\xi_i}{\bar{x}}}{n}$$

and write for the density of this latter variable

$$p_{S_n[\psi]}(x) \stackrel{n\uparrow\infty}{\to} \frac{e^{-\frac{n\,(x-m)^2}{2\,\sigma^2}}}{\sqrt{\frac{2\,\pi\,\sigma^2}{n}}} \tag{2.2}$$

We can use (2.2) to tentatively compute moments of arbitrary order of

$$\eta_n = \bar{x}^n\, e^{n\, S_n[\psi]}$$

using

$$\prec \eta_n^k \succ \stackrel{n\uparrow\infty}{\simeq} \bar{x}^{k\,n} \int_{-\infty}^{\infty} dx\, e^{n\,k\,x} \frac{e^{-\frac{n\,(x-m)^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\frac{\sigma^2}{n}}} = \bar{x}^{n\,k} e^{n\left(k\,m+\frac{k^2\,\sigma^2}{2}\right)} \tag{2.3}$$

The same quantity can be, however, computed *directly* from its very definition:

$$\prec \eta_n^k \succ = \prod_{i=1}^n \prec \xi_i^k \succ = \prec \xi^k \succ^n = e^{n\,\ln \prec \xi^k \succ} = e^{n\,(k\,\ln \bar{x} + \ln\Gamma(k+1))} \tag{2.4}$$

From Stirling formula we know that

$$\ln\Gamma(k+1) \stackrel{k\uparrow\infty}{\to} k\,(\ln k - 1) + o(k)$$

which, for k sufficiently large, disproves (2.3). On the other hand, for small k we have

$$\ln \prec \xi^k \succ = \ln \prec 1 + k \ln \xi + \frac{k^2}{2} \ln^2 \xi + \cdots \succ = k\,m + \frac{\sigma^2\,k^2}{2} + \dots$$

which coincides with the central limit prediction.

- An alternative way to phrase the content of the above example is the following: when computing expectation values of random variables which take large values with small probability contributions from such values *cannot be* neglected. The product of something big by something small can still be big. A systematic way to tackle the problem is provided by the theory of *large deviations* (see e.g. [2]).

In applications, a qualitative estimate of the *bulk* of the asymptotic distribution is provided by the variance of ζ_n

$$S_n - \prec \xi \succ \sim O\left(\frac{\sigma}{\sqrt{n}}\right)$$

References

[1] L.C. Evans, *An Introduction to Stochastic Differential Equations*, lecture notes, http://math.berkeley.edu/ evans/

[2] S.R.S. Varadhan, *Large deviations*, The Annals of Probability 2008, Vol. 36, No. 2, 397419. http://www.math.nyu.edu/faculty/varadhan/wald.pdf