1 Introduction

The Borel–Cantelli lemma is nicely discussed in paragraph 1.2 of [4]. The presentation of the strong law of large numbers and the central limit theorem is drawn from [1].

2 Borel -Cantelli lemma

Let $\{F_k\}_{k=1}^{\infty}$ a sequence of events in a probability space.

Definition 2.1 (*F_n infinitely often*)**.** *The event specified by the simultaneous occurrence an infinite number of the events in the sequence $\{F_k\}_{k=1}^{\infty}$ is called* "F_n infinitely often" *and denoted " F_n i.o.". In formulae*

$$F_n\, i.o. := \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} F_k = \{\omega \in \Omega \,|\, \omega \, \textit{belongs to infinitely many of the} F_n\}$$

Let us analyze the meaning of the individual events intervening in (2.1)

- $\tilde{F}_n := \cup_{k=n}^{\infty} F_k$: for any given n *at least one* of the $\{F_k\}_{k=n}^{\infty}$ occurs.
- $\cap_{n=1}^{\infty} \tilde{F}_n$: this event differ from the empty set if and only if all the $\tilde{F}_n$'s have a non-trivial intersection.
- By definition we have $\tilde{F}_{n+1} \subseteq \tilde{F}_n$ whence it follows

$$\cap_{n=1}^{\infty} \tilde{F}_n \equiv \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} F_k = \lim_{n \uparrow \infty} \sup_{k \geq n} F_k \tag{2.1}$$

An alternative, equivalent notation is

$$F_n\, i.o. = \lim_{n \uparrow \infty} \cup_{k \geq n} F_k = \lim_{n \uparrow \infty} \sup_{k \geq n} F_k$$

Recall also that

$$\mathrm{P}(F_k) = \int dP\, \chi_{F_k}(\omega)$$

for χ_{F_k} the characteristic function of the event χ and that

$$\lim_{n \uparrow \infty} \sup_{k \geq n} \chi_{F_k}(\omega) = \chi_{\lim_{n \uparrow \infty} \sup_{k \geq n} F_k}(\omega)$$

Finally let us observe that the following proposition holds

Proposition 2.1. *Let $\{A_k\}_{k=1}^{\infty}$ an increasing (decreasing) sequence of telescopic events such that $A_1 \subseteq A_2 \subseteq \dots$ ($A_1 \supseteq A_2 \supseteq \dots$). If*

$$A = \lim_{n \uparrow \infty} A_n \tag{2.2}$$

then we have

$$\lim_{n \uparrow \infty} \mathrm{P}(A_n) = \mathrm{P}(A) \tag{2.3}$$

Proof. The axioms of probability require that if $A_n \subseteq A_{n+1}$ then $\mathrm{P}(A_n) \leq \mathrm{P}(A_{n+1})$. Since $\mathrm{P}(A_n) \leq P(A) \leq 1$ for all n the sequence $\{\mathrm{P}(A_n)\}_{n=1}^{\infty}$ is monotonic and bounded. It must therefore have a limit

$$\lim_{n\uparrow\infty} \mathrm{P}(A_n) = \mathrm{P}(A') \ \leq \ \mathrm{P}(A) \tag{2.4}$$

Suppose $A' \subset A$ then by hypothesis for n sufficiently large $A' \subset A_n$ and therefore $\mathrm{P}(A') \ < \ \mathrm{P}(A)$ which contradicts (2.4). □

By virtue of the above considerations we can state and prove the Borel–Cantelli lemma.

Lemma 2.1 (*Borel-Cantelli*)**.** *The following claims hold:*

- *if* $\sum_n \mathrm{P}(F_n) \ < \ \infty$ *then* $\mathrm{P}(F_n\, i.o.) = 0$
- *if* $\sum_n \mathrm{P}(F_n) \ = \ \infty$ *and* $\{F_n\}_{n=1}^{\infty}$ *consists of independent events* $\mathrm{P}(F_n\, i.o.) = 1$

Proof. :

- By definition

$$\mathrm{P}(F_n\, i.o.) \ = \ \mathrm{P}(\cap_{n=0}^{\infty} \cup_{k=n}^{\infty} F_k) = \mathrm{P}(\lim_{n\uparrow\infty} \cup_{k=n}^{\infty} F_k) \tag{2.5}$$

Lebesgue's dominated convergence theorem (see e.g. [3] pag. 187), allows us to carry out the limit from the integral

$$\mathrm{P}(F_n\, i.o.) \ = \ \lim_{n\uparrow\infty} \mathrm{P}(\cup_{k=n}^{\infty} F_k)$$

whilst the definition of probability measure enforces the inequality

$$\mathrm{P}(F_n\, i.o.) \ \leq \ \lim_{n\uparrow\infty} \sum_{k=n}^{\infty} \mathrm{P}(F_k) = 0$$

The proof of the first statement follows from the hypothesized convergence of the series.

- We can turn to the complementary event:

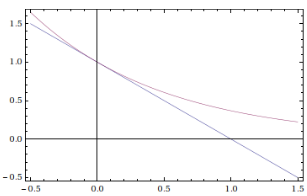
$$(\cup_{k\geq n} F_k)^c = \cap_{k\geq n} F_k^c$$

and use independence

$$\mathrm{P}(\cap_{k\geq n} F_k^c) = \prod_{k=n}^{\infty} \mathrm{P}(F_k^c) = \prod_{k=n}^{\infty} [1 - \mathrm{P}(F_k)]$$

The inequality

$$1 - x \leq e^{-x} \tag{2.6}$$



then provides us with an upper bound for each factor in the product

$$\mathrm{P}(\cap_{k\geq n} F_k^c) \leq \prod_{k=n}^{\infty} e^{-\mathrm{P}(F_k)} = e^{-\sum_{k=n}^{\infty} \mathrm{P}(F_k)}$$

whence the claim follows if the series diverges.

□

The Borel-Cantelli lemma provides an extremely useful tool to prove asymptotic results about random sequences holding *almost surely* (acronym: *a.s.*). This mean that such results hold true but for events of zero probability. An obvious synonym for *a.s.* is then *with probability one*.

3 Law of Large Numbers

The law of large number theorem gives information about the average outcome of repeated sampling of the same random variable performed one independently of the other.

Theorem 3.1 (*The strong law of large numbers*). *Let* $\{\boldsymbol{\xi}_i\}_{i=1}^n$ *a sequence of independent identically distributed (i.e. we suppose* $\xi_i \stackrel{d}{=} \xi$ *for all* $i \in \mathbb{N}$*) random variables defined over the same probability space. Then if* $\mathrm{E}\xi^4 < \infty$ *we have*

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

Proof. Let

$$\Xi_n = \sum_{i=1}^n \frac{\xi_i}{n} \tag{3.1}$$

In order to prove the claim, by Borel-Cantelli lemma it is sufficient to show that

$$\sum_{n=1}^{\infty} \mathrm{P}\left(|\Xi_n - \mathrm{E}\xi| \geq \varepsilon\right) < \infty \tag{3.2}$$

From Čebyšev's lemma we know that

$$\mathrm{P}\left(|\Xi_n - \mathrm{E}\xi| \geq \varepsilon\right) \leq \frac{\mathrm{E}(\Xi_n - \mathrm{E}\xi)^4}{\varepsilon^4} \tag{3.3}$$

Let us observe first that

$$\Xi_n - \mathrm{E}\xi = \sum_{i=1}^n \frac{\xi_i}{n} - \mathrm{E}\,\xi = \sum_{i=1}^n \frac{\xi_i - \mathrm{E}\,\xi}{n} = \sum_{i=1}^n \frac{\tilde{\xi}_i}{n} := \tilde{\Xi}_n \tag{3.4}$$

where $\left\{\tilde{\xi}_i\right\}_{i=1}^n$ is a sequence of *i.i.d.* random variables $\tilde{\xi}_i \stackrel{d}{=} \tilde{\xi}$ such that $\mathrm{E}\tilde{\xi} = 0$. We have then that

$$\mathrm{E}\tilde{\Xi}_n = \sum_{ijlk=1}^n \frac{\mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k}{n} \tag{3.5}$$

where

i if at least one of the labels, say i, differs from all the others

$$\mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k = (\mathrm{E}\,\tilde{\xi}_i)\mathrm{E}\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k \tag{3.6}$$

as the random variable in the sequence are mutually independent; in such a case

$$\mathrm{E}\,\tilde{\xi}_i = 0 \quad \Rightarrow \quad \mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k = 0 \tag{3.7}$$

ii if labels are equal in pairs, say $i = j$ and $l = k$ then

$$\mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k = (\mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j)\mathrm{E}\tilde{\xi}_l\tilde{\xi}_k = (\mathrm{E}\tilde{\xi}^2)^2 \tag{3.8}$$

iii if all labels are equal

$$\mathrm{E}\,\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k = (\mathrm{E}\tilde{\xi}^2)^4 \tag{3.9}$$

We conclude that

$$\sum_{ijlk=1}^{n} \mathrm{E}\tilde{\xi}_i\tilde{\xi}_j\tilde{\xi}_l\tilde{\xi}_k = n\,\mathrm{E}\tilde{\xi}^4 + (\mathrm{E}\tilde{\xi}^2)^2(\sum_{ijlk}\delta^{ij}\delta^{lk} + \delta^{il}\delta^{jk} + \delta^{ik}\delta^{jl} - n) = n\,\mathrm{E}\tilde{\xi}^4 + 3n\,(n-1)(\mathrm{E}\tilde{\xi}^2)^2 \tag{3.10}$$

Jensen's inequality implies that $\mathrm{E}\tilde{\xi}^4 < \infty$ implies $\mathrm{E}\tilde{\xi}^2 < \infty$ thus for n sufficiently large there exists a positive constant $K < \infty$ such that

$$\mathrm{E}\tilde{\Xi}_n^4 \leq \frac{K}{n^2} \tag{3.11}$$

From this latter result we have

$$\sum_{n=1}^{\infty} \mathrm{P}\,(|\Xi_n - \mathrm{E}\xi| \geq \varepsilon) \leq \sum_{n=1}^{\infty} \frac{K}{n^2} < \infty \tag{3.12}$$

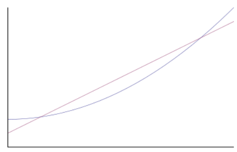
which proves the claim. □

In the proof of the strong law of large numbers we mentioned Jensen's inequality. Jensen's inequality holds for *convex* functions i.e. (Borel measurable) functions

$$f : \Omega \mapsto \mathbb{R} \tag{3.13}$$

such that for any two points x_1 and x_2 in its domain of definition Ω and any $t \in [0, 1]$

$$f(t\,x_1 + (1-t)\,x_2) \leq t\,f(x_1) + (1-t)\,f(x_2)$$



Proposition 3.1 (*Jensen's inequality*). *Let the Borel function $f(x)$ be downward convex and ξ a random variable with absolutely convergent first moment. Then*

$$f(\mathrm{E}\,\xi) \leq \mathrm{E}\,f(\xi)$$

Proof. Using the definition of convex function for each $x_o \in \mathbb{R}$ we can find a number $g(x_o)$ such that

$$f(x) \geq f\,(x_o) + g(x_o)\,(x - x_o)$$

The identifications $x = \xi$ and $x_o = \mathrm{E}\,\xi$ yield the proof of the inequality □

In particular upon setting

$$f(x^2) = x^4 \tag{3.14}$$

we get into

$$\mathrm{E}\,\xi^4 \geq (\mathrm{E}\,\xi^2)^2 \tag{3.15}$$

A further useful consequence of Jensen's inequality is

Proposition 3.2 (*Lyapunov's inequality*)**.** *Let* $0 < s < t$ *then*

$$(\mathrm{E}\,|\xi|^s)^{1/s} \leq (\mathrm{E}\,|\xi|^t)^{1/t}$$

Proof. Define $r = t/s$ and

$$\eta = |\xi|^s \tag{3.16}$$

Since $r > 1$ the function $f(x) = x^r$ is convex. By Jensen's inequality

$$(\mathrm{E}\,\eta)^r \equiv f(\mathrm{E}\,\eta) \leq \mathrm{E}\,f(\eta) \equiv \mathrm{E}\,\eta^r$$

whence the claim follows upon inserting (3.16) into the inequality. □

4 Central Limit theorem

The central limit theorem gives information about the deviation from the average outcome of repeated sampling of the same random variable performed one independently of the other.

Theorem 4.1 (*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* $\xi_i \overset{d}{=} \xi$ *and*

$$\mathrm{E}\,\xi = m$$
$$\mathrm{E}\,(\xi - \mathrm{E}\,\xi)^2 = \sigma^2 > 0$$

Set

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

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

$$\lim_{n \uparrow \infty} P\left(a < \frac{\Xi_n - m}{[\mathrm{E}(\Xi_n - m)^2]^{1/2}} < b\right) = \int_a^b dx\, g_{0\,1}(x) \tag{4.1}$$

where

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

Sketch of the proof. Consider the characteristic function

$$\mathrm{E} e^{\imath t \frac{\Xi_n - m}{[\mathrm{E}(\Xi_n - m)^2]^{1/2}}} = \mathrm{E}\, e^{\imath t \frac{\sum_{i=1}^{n} \xi_i - n\, m}{\sqrt{n}\sigma}} = \prod_{i=1}^{n} \mathrm{E}\, e^{\imath t \frac{\xi_i - m}{\sqrt{n}\sigma}} = \left[\int_{\mathbb{R}} dx\, e^{\imath t \frac{x-m}{\sqrt{n}\sigma}}\, p_{\xi}(x)\right]^n$$

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

$$\mathrm{E}\, e^{\imath t \frac{\Xi_n - n\, m}{\sqrt{n}\sigma}} = \left[1 - \frac{t^2}{2\, n} \int_{\mathbb{R}} dx\, \frac{(x-m)^2}{\sigma^2}\, p_{\xi}(x) + O(\frac{1}{n^{3/2}})\right]^n \overset{n\uparrow\infty}{\to} e^{-\frac{t^2}{2}}$$

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

4.1 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.

4.1.1 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.

4.1.2 Domain of validity

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

$$\zeta_n := \frac{\Xi_n - m}{[\mathrm{E}(\Xi_n - m)^2]^{1/2}} \qquad n \gg 1$$

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 over $\mathbb{R}_+$

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

From this sequence we can construct the products

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

By hypothesis

$$m := \mathrm{E}\psi = \int_0^\infty dx\, \ln \frac{x}{\bar{x}}\, p_\xi(x)\ <\ \infty$$
$$\sigma^2 := \mathrm{E}(\psi - m)^2 = \int_0^\infty dx\, \ln^2 \frac{x}{\bar{x}}\, p_\xi(x) - m^2\ <\ \infty$$

We can explicitly compute mean and variance as derivatives of the Γ function. After passing to nondimensional variable

$$y = \frac{x}{\bar{x}} \tag{4.2}$$

we have

$$m = \lim_{\varepsilon \downarrow 0} \int_0^\infty dy\, \frac{y^\varepsilon - 1}{\varepsilon}\, p_\xi(y) = \lim_{\varepsilon \downarrow 0} \frac{\Gamma(1+\varepsilon) - 1}{\varepsilon} = -\gamma \tag{4.3}$$

where γ stands for the Euler constant

$$\gamma = 0.577\ldots \tag{4.4}$$

Similarly we obtain

$$\sigma^2 = \lim_{\varepsilon \downarrow 0} \frac{\Gamma(1+2\,\varepsilon) + 1 - 2\,\Gamma(1+\varepsilon)}{\varepsilon^2} - \gamma^2 = \frac{\pi^2}{6} \tag{4.5}$$

Since the variance is finite, we can apply the central limit theorem to

$$\Psi_n := \frac{\sum_{i=1}^{n} \psi_i}{n}$$

and write for the density of this latter variable

$$p_{\Psi_n}(x) \stackrel{n\uparrow\infty}{\to} \frac{e^{-\frac{n\,(x-m)^2}{2\,\sigma^2}}}{\sqrt{\frac{2\,\pi\,\sigma^2}{n}}} \tag{4.6}$$

We can use (4.6) to *tentatively* compute moments of arbitrary order of

$$\eta_n = \bar{x}^n\, e^{n\,\Psi_n}$$

using

$$\mathrm{E}\,\eta_n^k \stackrel{n\uparrow\infty}{\simeq} \bar{x}^{n\,k} \int_{-\infty}^{\infty} dx\, e^{n\,k\,x}\, \frac{e^{-\frac{n\,(x-m)^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\frac{\sigma^2}{n}}} = \bar{x}^{n\,k}\, e^{n\left(k\,m + \frac{k^2\,\sigma^2}{2}\right)} \tag{4.7}$$

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

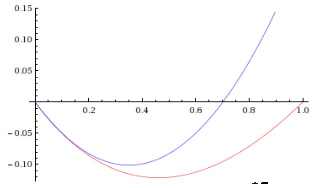
$$\mathrm{E}\,\eta_n^k = \prod_{i=1}^{n} \mathrm{E}\,\xi_i^k = (\mathrm{E}\,\xi^k)^n = \bar{x}^{k\,n}\, e^{n\,\ln \mathrm{E}\left(\frac{\xi}{\bar{x}}\right)^k} = e^{n\,[k\,\ln \bar{x} + \ln \Gamma(k+1)]} \tag{4.8}$$

From Stirling's formula we know that

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

which, for k sufficiently large, disproves (4.7). On the other hand, for small k we have

$$\ln \mathrm{E}\,\xi^k = \ln \mathrm{E}\left(1 + k \ln \xi + \frac{k^2}{2} \ln^2 \xi + \dots\right) = k\,m + \frac{\sigma^2\,k^2}{2} + \dots \tag{4.9}$$



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. [5]).

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

$$\Xi_n - \mathrm{E}\,\xi \sim O\left(\frac{\sigma}{\sqrt{n}}\right)$$

References

[1] L. C. Evans. An introduction to stochastic differential equations. UC Berkeley, Lecture Notes, 2006.

[2] A. I. Khinchin. *Mathematical foundations of statistical mechanics*. Dover series in mathematics and physics. Dover Publications, 1949.

[3] A. N. Shiryaev. *Probability*, volume 95 of *Graduate texts in mathematics*. Springer, 2 edition, 1996.

[4] R. van Handel. Stochastic calculus and stochastic control. Lecture Notes, Caltech, 2007.

[5] S. Varadhan. Large deviations. *The Annals of Probability*, 36(2):397419, 2008.