1 Introduction

The lecture notes cover sections F and G of chapter 2 and appendix Appendix A of [1]. The notes also discuss some extra examples.

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*

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

An alternative notation may help. Observe that the set-theoretic union ($\cup$) operation has the probabilistic meaning of "*at least*". The event $\tilde{F}_n = \cup_{k=n}^{\infty} F_k$ defines a sampling of the tail (starting from n) of the sequence. It occurs if at least one event in the tail occurs. The intersection $\cap_{n=1}^{\infty} \tilde{F}_n$ differs from the empty set if disregarding how long is the tail (how large is n) we can observe the occurrence of an $F_{n'}$ for $n' \geq n$. This is possible only if an infinite number of the events in the sequence indeed occur:

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

Recall also that

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

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

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

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

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

Proof. :

- By definition

$$P(F_n\, i.o.) = P(\cap_{n=0}^{\infty} \cup_{k=n}^{\infty} F_k) = P(\lim_{n\uparrow\infty} \cup_{k=n}^{\infty} F_k) \tag{2.1}$$

Lebesgue's dominated convergence theorem (see e.g. [2] pag. 187), allows us to carry out the limit from the integral

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

whilst the definition of probability measure enforces the inequality

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

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

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

The inequality

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

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

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

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 Characteristic function of a random variable

Definition 3.1 (*Characteristic function*)**.** *Let*

$$\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$$

the expectation value

$$\check{p}_{\boldsymbol{\xi}}(\boldsymbol{q}) := \prec e^{\imath \boldsymbol{\xi}\cdot\boldsymbol{q}} \succ$$

is referred to as the characteristic function of the random variable

Example 3.1 (*Characteristic function of a Gaussian random*)**.** Let

$$\xi : \Omega \to \mathbb{R}$$

distributed with Gaussian PDF. The characteristic function is

$$\check{g}_{\bar{x},\sigma}(q) = \int dx\, e^{\imath q x} g_{\bar{x},\sigma}(x) = e^{\imath q x - \frac{\sigma^2 q^2}{2}}$$

For a Gaussian variable it is also true

$$\check{G}_{\bar{x},\sigma}(q) = \frac{i^n q^n}{\Gamma(n+1)} \prec \xi^n \succ$$

having used the Γ-function representation of the factorial (see appendix B). The remaining expectation value is

$$\prec \xi^n \succ = \frac{\Gamma(2n+1)}{2^n \Gamma(n+1)} = (2n-1)!! \tag{3.1}$$

Formally for a random variable ξ one can write

$$\prec \xi^n \succ = \frac{1}{\imath^n} \frac{d^n}{dq^n} \check{p}_{\xi}(q)\bigg|_{q=0} \tag{3.2}$$

The relation is formal because it may be a relation between infinities.

Example 3.2 (*Lorentz distribution*)**.** Let

$$p : \mathbb{R} \to \mathbb{R}_{+}$$

be

$$p_{y\,\sigma}(x) = \frac{\sigma}{\pi\,\{(x-y)^2 + \sigma^2\}}$$

the Lorentz probability density so that $(\mathbb{R}\,,\mathcal{B}\,,P_{y\,\sigma}(\mathcal{B}))$ a probability space. Note that

$$p_{0\,\sigma}(x) = p_{0\,\sigma}(-x)$$

Using a change of variable and it is straightforward to verify that

$$\int_{\mathbb{R}} dx\, x\, p_{y\,\sigma}(x) = y$$

however

$$\int_{\mathbb{R}} dx\, x^2\, p_{y\,\sigma}(x) = \infty$$

The characteristic function can be computed using Cauchy theorem

$$\begin{aligned} \check{p}_{y\,\sigma}(q) &= e^{\imath q y} \int dx\, e^{\imath q x} p_{0\,\sigma}(x) \\ &= \frac{e^{\imath q y}}{2\,\imath\,\pi} \int_{\mathbb{R}} dx\, e^{\imath q x} \left\{ \frac{1}{x - \imath\,\sigma} - \frac{1}{x + \imath\,\sigma} \right\} = e^{\imath q y} \begin{cases} e^{-q\,\sigma} & \text{if } q > 0 \\ e^{q\,\sigma} & \text{if } q < 0 \end{cases} \end{aligned}$$

The characteristic function develops a cusp for $q = 0$

$$\check{p}_{y\,\sigma}(q) = e^{\imath q y - \sigma |q|}$$

Finally note that

$$\delta(x-y) \overset{w}{=} \lim_{\sigma \downarrow 0} p_{y\,\sigma}(x)$$

Remark 3.1 (*The Fourier representation of the δ-Dirac*)**.** Let $f_i : \mathbb{R}^d \to \mathbb{R}\,, i = 1, 2$ some smooth integrable functions and

$$\check{f}_i(\boldsymbol{q}) = \int_{\mathbb{R}^d} d^d x\, e^{\imath \boldsymbol{q}\cdot\boldsymbol{x}} f_i(\boldsymbol{x}) \qquad i = 1, 2$$

their Fourier transform. The convolution identity

$$(f_1 \star f_2)(\boldsymbol{x}) := \int d^d y\, f_1(\boldsymbol{x} - \boldsymbol{y})\, f_2(\boldsymbol{x}) = \int \frac{d^d q}{(2\pi)^d}\, e^{\imath\,\boldsymbol{q}\cdot\boldsymbol{x}} \check{f}_1(\boldsymbol{q}) \check{f}_2(\boldsymbol{q})$$

maybe thought as a consequence of

$$\delta^{(d)}(\boldsymbol{x} - \boldsymbol{y}) \overset{w}{=} \int \frac{d^d q}{(2\pi)^d}\, e^{\imath\,\boldsymbol{q}\cdot(\boldsymbol{x}-\boldsymbol{y})}$$

The δ-Dirac in such a case could be very formally though as the characteristic function of a ”random variable” uniformly distributed over $\mathbb{R}^d$.

4 Probability density and Dirac-δ

Let

$$\boldsymbol{\xi} \,:\, \Omega \,\to\, \mathbb{R}^d$$

with PDF $p_{\boldsymbol{\xi}}(\cdot)$. From the properties of the δ function we have

$$\prec \delta^{(d)}(\boldsymbol{\xi} - \boldsymbol{x}) \succ = \int_{\mathbb{R}^d} d^d y \, p_{\boldsymbol{\xi}}(\boldsymbol{y}) \delta^{(d)}(\boldsymbol{y} - \boldsymbol{x}) = p_{\boldsymbol{\xi}}(\boldsymbol{x})$$

This relation allows us to derive the relation between the PDF's of functionally dependent random variables

4.1 Derivation in one dimension

Suppose the random variable has the

$$P_{\xi}(x < \xi < x + dx) = p_{\xi}(x)\, dx$$

Functional relation between random variables

$$\phi = f(\xi) \tag{4.1}$$

again

$$P_{\phi}(y < \phi < y + dy) = p_{\phi}(y)\, dy$$

From

$$P_{\xi}(x < \xi < x + dx) = P_{\phi}(y < \phi < y + dy)$$

one gets into

$$p_{\xi}(x) = p_{\phi}(f(x)) \frac{df}{dx}$$

4.2 Multi-dimensional case using the δ-function

Suppose f is one-to-one and write

$$p_{\boldsymbol{\phi}}(\boldsymbol{y}) = \prec \delta^{(d)}(\boldsymbol{\phi} - \boldsymbol{y}) \succ = \prec \delta^{(d)}(f(\boldsymbol{\xi}) - \boldsymbol{y}) \succ$$

It follows

$$p_{\boldsymbol{\phi}}(\boldsymbol{y}) = \int_{\mathbb{R}^d} d^d x \, \delta^{(d)}(f(\boldsymbol{x}) - \boldsymbol{y}) p_{\boldsymbol{\xi}}(\boldsymbol{x}) \equiv \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} d^d x \, \frac{e^{-\frac{\|f(\boldsymbol{x}) - \boldsymbol{y}\|^2}{2\,\sigma^2}}}{(2\,\pi\,\sigma^2)^{\frac{d}{2}}} \, p_{\boldsymbol{\xi}}(\boldsymbol{x})$$

Since f is one-to-one

$$f^{-1}(\boldsymbol{y}) = \boldsymbol{x}_{\star} \tag{4.2}$$

is globally well defined. Taylor-expanding the argument of the exponential we get for the i-th component of $\boldsymbol{y}$

$$y^i = f^i(\boldsymbol{x}_{\star}) + (x^j - x^j_{\star}) \frac{\partial f^i}{\partial x^j}(\boldsymbol{x}_{\star}) + O((x^j - x^j_{\star})^2)$$

Call

$$A_{ij} := \frac{\partial f^i}{\partial x^j}(\boldsymbol{x}_\star)$$
$$z^i = \frac{x^j - x^j_\star}{\sigma}$$

the integral becomes

$$p_{\boldsymbol{\phi}}(\boldsymbol{y}) = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} d^d z \, \frac{e^{-\frac{z^i A_{li} A_{lj} z^j + O(\sigma)}{2}}}{(2\,\pi)^{\frac{d}{2}}} \, p_{\boldsymbol{\xi}}(\boldsymbol{x}_\star + O(\sigma)) = \frac{p_{\boldsymbol{\xi}}(\boldsymbol{x}_\star)}{|\det A|} \tag{4.3}$$

5 Limit theorems for Bernoulli variables

Definition 5.1 (*i.i.d. random variables*)**.** *A sequence of random variables $\{\xi_i\}_{i=1}^{\infty}$ is said identically distributed if*

$$P_{\xi_1}(x) = P_{\xi_2}(x) = \dots = P_{\xi_n}(x) = \dots \tag{5.1}$$

Furthermore if they are mutually independent they are usually referred to with the acronym i.i.d.

Let $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Bernoulli variables i.e. for all i, $\xi_i \stackrel{d}{=} \xi$ (*equality in distribution*) and

$$\xi : \Omega \to \{-x, x\} \qquad \& \qquad P_\xi(x) = p$$

From the characteristic function

$$\prec e^{\imath t \xi} \succ = \cos(x t) + \imath\,(2\,p - 1)\,\sin(x t)$$

we find

$$\prec \xi^n \succ = \begin{cases} x^n & n = 2\,k \\ x^n\,(2\,p - 1) & n = 2\,k + 1 \end{cases}$$

whence

$$\prec \xi \succ = (2\,p - 1)\,x \qquad \& \qquad \prec (\xi - \prec \xi \succ)^2 \succ = 4\,p\,(1 - p)\,x^2$$

and

$$\prec (\xi - \prec \xi \succ)^4 \succ = 16\,p\,(1 - p)\,[1 - 3\,(1 - p)\,p]$$

If we introduce the random variable

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

then

$$\prec S_n \succ = \prec \xi \succ \qquad \& \qquad \prec (S_n - \prec \xi \succ)^2 \succ = \frac{4\,p\,(1 - p)}{n}$$

- We can apply Chebyshev lemma to show

$$P(|S_n - \prec \xi \succ| \geq \varepsilon) \leq \frac{\prec (S_n - \prec \xi \succ)^2 \succ}{\varepsilon^2} = \frac{4\,p\,(1 - p)}{n\,\varepsilon^2} \stackrel{n \uparrow \infty}{\to} 0 \tag{5.2}$$

  This is *law of large numbers* for i.i.d. Bernoulli variables.

- Upon setting

$$\tilde{\xi_i} = \xi_i - \prec \xi \succ \tag{5.3}$$

$$\begin{aligned}
\prec (S_n - \prec \xi \succ)^4 \succ &= \sum_{i=1} \frac{\prec \tilde{\xi}_i^4 \succ}{n^4} + 3 \sum_{ijkl} \delta_{ij}\delta_{kl}(1-\delta_{jk}) \frac{\prec \tilde{\xi}_i^2 \succ^2}{n^4} \\
&= \frac{16\, p\,(1-p)}{n^3}\,[1 - 3\,(1-p)\,p] + 3\,n\,(n-1) \frac{16\, p^2\,(1-p)^2}{n^4} \;\le \frac{C}{n^2}
\end{aligned}$$

  for $n \gg 1$, we also find

$$P(|S_n - \prec \xi \succ| \ge \varepsilon) \le \frac{\prec (S_n - \prec \xi \succ)^4 \succ}{\varepsilon^4} \le \frac{C}{n^2\,\varepsilon^4} \stackrel{n\uparrow\infty}{\to} 0$$

  The new bound tends to zero sufficiently fast at infinity to be useful in order to apply the Borel-Cantelli lemma which entitles us to conclude:

$$S_n \stackrel{n\uparrow\infty}{\to} \prec \xi \succ \qquad a.s.$$

  i.e. we have proved the *strong law of large numbers* for Bernoulli schemes.

- Let us convene that S_0 is zero. Then we can write (see appendix A)

$$P\left(S_n = \frac{m\,x}{n}\right) = \frac{\Gamma(n+1)}{\Gamma\left(\frac{n+m}{2}+1\right)\Gamma\left(\frac{n-m}{2}+1\right)} p^{\frac{n+m}{2}} (1-p)^{\frac{n-m}{2}}$$

  For $n\,,m \gg 1$ we can extricate the asymptotics of the probability using Stirling's formula (see formula B.1 in appendix (B)):

$$P\left(S_n = \frac{m\,x}{n}\right) \simeq \frac{1}{\sqrt{2\,\pi\,\frac{n^2-m^2}{n}}} e^{\frac{n+m}{2}\ln p + \frac{n-m}{2}\ln(1-p) + n(\ln n - 1) - \frac{n+m}{2}\left(\ln\frac{n+m}{2} - 1\right) - \frac{n-m}{2}\left(\ln\frac{n-m}{2} - 1\right)} \tag{5.4}$$

  The strong law of large numbers allows us to relate asymptotically m to the *empirical probability* (i.e. the observed frequency) of a displacement to the right

$$m = (2\,\tilde{p} - 1)$$

  As n tends to infinity the observation allows as to recast (5.4) into the form

$$P\left(S_n = \frac{m\,x}{n}\right) \simeq \frac{e^{-n\,K(\tilde{p}|p) + o(n)}}{\sqrt{2\,\pi\,\frac{4\,\tilde{p}\,(1-\tilde{p})}{n}}}$$

  with

$$K(\tilde{p}|p) = -\left\{\tilde{p}\ln\frac{\tilde{p}}{p} + (1-\tilde{p})\ln\frac{1-\tilde{p}}{1-p}\right\} \tag{5.5}$$

  the *Kullback-Leibler divergence (entropy)* between the empirical and the Bernoulli distribution. This quantity measure the (rate of) discrepancy between two probability measures.

- Let z be an indicator of the discrepancy between p and $\tilde{p}$ i.e.

$$\tilde{p} = p + \frac{z}{2}$$

Taylor expanding (5.5) around z equal zero yields

$$K(\tilde{p}|p) = \frac{z^2}{8\,p(1-p)} + O(z^3) \tag{5.6}$$

We can use this result to estimate the probability with which S_n asymptotically deviates from its expected value

$$P\left(\frac{S_n - \prec \xi \succ}{\prec (S_n - \prec \xi \succ)^2 \succ^{1/2}} = z\right) \overset{n\uparrow\infty}{\simeq} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\,\pi\, \frac{\prec (S_n - \prec \xi \succ)^2 \succ}{x^2}}}$$

This is the content of the *central limit theorem* for Bernoulli variables. It is tempting to interpret the vanishing of the variance in the denominator as n tends to infinity, as a weight in a Riemann sum so to infer

$$P\left(\frac{S_n - \prec \xi \succ}{\prec (S_n - \prec \xi \succ)^2 \succ^{1/2}} \leq z\right) \overset{n\uparrow\infty}{\simeq} \int_{-\infty}^{z} du\, \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\,\pi}}$$

- Consider now the characteristic function

$$\prec e^{\imath S_n\, t} \succ = \prec e^{\imath \xi\, t} \succ^n = \left[\cos\left(\frac{x\,t}{n}\right) + \imath\,(2\,p-1)\,\sin\left(\frac{x\,t}{n}\right)\right]^n$$

The limit

$$\lim_{n\uparrow\infty} \ln \prec e^{\imath S_n\, t} \succ = \imath\,(2\,p-1)\,x\,t - \frac{x^2\,t^2}{2}\frac{4\,p\,(p-1)}{n} + o\left(\frac{1}{n}\right)$$

contains the same type of information of the central limit theorem. Namely we can couch the result into the form

$$\begin{aligned}
&\lim_{n\uparrow\infty} \ln \prec e^{\imath S_n\, t} \succ \\
&\qquad = \imath \prec \xi \succ\; t - \frac{t^2}{2}\frac{\prec (\xi - \prec \xi \succ)^2 \succ}{n} + o\left(\frac{1}{n}\right) \\
&\qquad = \imath\, \prec S_n \succ\; t - \frac{t^2}{2} \prec (S_n - \prec S_n \succ)^2 \succ + o\left(\frac{1}{n}\right)
\end{aligned}$$

The rescaling

$$t \;\to\; \frac{t}{\prec (S_n - \prec S_n \succ)^2 \succ^{1/2}}$$

suggests that the characteristic function

$$\prec e^{\imath \frac{S_n - \prec S_n \succ}{\prec (S_n - \prec S_n \succ)^2 \succ^{1/2}} t} \succ \overset{n\uparrow\infty}{\to} e^{-\frac{t^2}{2}}$$

tends indeed to the characteristic function of the Gaussian distribution.

Appendices

A Random walk

Let

$$R_n = n\, S_n$$

and suppose that of the n steps n_r were to the right and n_l to the left:

$$n = n_l + n_r \tag{A.1}$$

The displacement from the origin in units of x is then given by

$$m = n_r - n_l \tag{A.2}$$

We can solve for n_l, n_r and obtain

$$n_r = \frac{n+m}{2} \qquad \& \qquad n_l = \frac{n-m}{2}$$

The probability of an *individual sequence* of samples of Bernoulli variables such that $R_n(\omega) = m\,x$ is

$$P(\omega) = p^{\frac{n+m}{2}}\,(1-p)^{\frac{n-m}{2}}$$

In order to evaluate the total probability $P(R_n = m\,x)$ we must *count* all possible sequences of samples such that (A.1), (A.2) are verified. This number is equal to the number of ways we can extract n_r out of n *indistinguishable object* (this means that the extraction order does not matter):

$$C_n^{n_r} = \frac{n!}{n_r!(n-n_r)!} \equiv \frac{n!}{n_r! n_l!}$$

The conclusion is

$$P(R_n = m\,x) = \frac{n!}{\left(\frac{n+m}{2}\right)!\left(\frac{n-m}{2}\right)!} p^{\frac{n+m}{2}}\,(1-p)^{\frac{n-m}{2}}$$

Using binomial formula it is straightforward to check the normalization condition for $m = -n, -(n-1), \dots, n-1, n$.

B Gamma function

The Γ function for any $x \in \mathbb{R}_+$ is specified by the integral

$$\Gamma(x) = \int_0^\infty \frac{dy}{y}\, y^x\, e^{-y}$$

For $x \in \mathbb{N}$ the integral can be performed explicitly and it is equal to the factorial:

$$\Gamma(x) = (x-1)! \qquad\qquad x \in \mathbb{N}$$

For $x \in \mathbb{R}_+$, integration by parts yields the identity

$$\Gamma(x+1) = \int_0^\infty \frac{dy}{y}\, y^{x+1}\, e^{-y} = x\,\Gamma(x)$$

which is trivially satisfied by factorials. For $x \gg 1$ the value of the integral is approximated by *Laplace's stationary point* method

$$\Gamma(x+1) \simeq e^{x\,(\ln x - 1)} \int_{\mathbb{R}} dy\, e^{-\frac{y^2}{2\,x}} = \sqrt{2\,\pi\,x}\, e^{x\,(\ln x - 1)} \qquad x \gg 1 \tag{B.1}$$

Such asymptotic estimation is usually referred to as *Stirling formula*.

References

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

[2] A. N. Shiryaev, *Probability*, 2nd Ed. Springer (1996), `http://books.google.com/`