1 Introduction

These notes shortly recall some basic concepts in classical probability. The material presented here is more leisurely expounded in a physics style language in chapter two of [2]. For a more mathematics style presentation see chapter one of [3] or chapter two of [1].

2 Measure theoretic definitions

Let Ω a non-empty set.

Definition 2.1 (σ-*algebra*)**.** *A σ-algebra is a collection $\mathcal{F}$ of subsets of Ω with these properties*

1. $\emptyset\,,\Omega\,\in\,\mathcal{F}$.
2. *if* $F\,\in\,\mathcal{F}$ *then* $F^{c}\,\in\,\mathcal{F}$ *for* $F^{c}:=\Omega\setminus F$ *the complement of* F.
3. *if* $\{F_k\}_{k=1}^{\infty}\,\in\,\mathcal{F}$ *then*

$$\cap_{k=1}^{\infty}F_k\,,\cup_{k=1}^{\infty}F_k\,\in\,\mathcal{F}$$

Definition 2.2 (*Probability measure*)**.** *Let $\mathcal{F}$ be a σ-algebra of subsets of Ω. We call*

$$P\,:\,\mathcal{F}\,\to\,[0,1]$$

a probability measure provided:

1. $P(\emptyset)=0\,,\,P(\Omega)=1$
2. *if* $\{F_k\}_{k=1}^{\infty}$ *then*

$$P(\cup_{k=1}^{\infty}F_k)\,\leq\,\sum_{k=1}^{\infty}P(F_k)$$

3. *if* $\{F_k\}_{k=1}^{\infty}$ *are* **disjoint** *sets*

$$P(\cup_{k=1}^{\infty}F_k)\,=\,\sum_{k=1}^{\infty}P(F_k) \tag{2.1}$$

It follows that if $F_1\,,F_2\,\in\,\mathcal{F}$

$$F_1\,\subset\,F_2\qquad\Rightarrow\qquad P(F_1)\,\leq\,P(F_2)$$

Definition 2.3 (*Borel σ-algebra*)**.** *The smallest σ-algebra containing all the* **open** *subsets of $\mathbb{R}^d$ is called the Borel σ-algebra, denoted by $\mathcal{B}$*

The **Borel subsets** of $\mathbb{R}^d$ i.e. the content of $\mathcal{B}$ may be thought as the collection of all the well-behaved subsets of $\mathbb{R}^d$ for which Lebesgue measure theory applies.

3 Probability Space

Definition 3.1 (*Probability space*)**.** *A triple*

$$(\Omega\,,\mathcal{F}\,,P)$$

is called a probability space *provided*

1. *Ω is any set*
2. *$\mathcal{F}$ is a σ-algebra of subsets of Ω*
3. *P is a probability measure on $\mathcal{F}$*

- Points $\omega \in \Omega$ are sample (outcome) points.
- A set $F \in \mathcal{F}$ is called an event.
- $P(F)$ is the probability of the event F.
- A property which holds true but for events of probability zero is said to hold almost surely (usually abbreviated "a.s.").

Example 3.1 (*Single unbiased coin tossing*)**.** :

- outcomes: head, tail
- $\Omega = \{head, tail\}$.
- σ-algebra $\mathcal{F}$: it comprises $|\mathcal{F}| = 2^{|\Omega|} = 4$ events
  1 T=tail
  2 H=head
  3 $\emptyset$=neither head nor tail
  4 $T \vee H$=head or tail
- Probability measure:

$$P(T) = P(H) = \frac{1}{2} \qquad \& \qquad P(\emptyset) = 0 \qquad \& \qquad P(T \vee H) = 1 \tag{3.1}$$

Example 3.2 (*Uniform distribution*)**.** :

- $\Omega = [0, 1]$.
- $\mathcal{F}$: the σ-algebra of all Borel subsets of $[0, 1]$.
- P: the Lebesgue measure on $[0, 1]$. (Note: as $0 \cup 1$ has zero measure $[0, 1] \sim (0, 1)$.)

Definition 3.2 (*Probability density on* $\mathbb{R}^d$)**.** *Let* p *be a* **non-negative**, *integrable function, such that*

$$\int_{\mathbb{R}^d} d^d x \, p(\boldsymbol{x}) = 1 \tag{3.2}$$

then to each $B \in \mathcal{B}$ *(Borel* σ*-algebra) is possible to associate a probability*

$$P(B) = \int_B d^d x \, p(\boldsymbol{x}) \tag{3.3}$$

so that $(\mathbb{R}^d, \mathcal{B}, P)$ *is a probability space. The function* p *is called the density of the probability* P*.*

Example 3.3 (*Gaussian distribution*)**.** The function

$$\begin{aligned} g_{\bar{x}\,\sigma} &: \mathbb{R} \to \mathbb{R}_+ \\ g_{\bar{x}\,\sigma}(x) &= \frac{e^{-\frac{(x-\bar{x})^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}} \end{aligned} \tag{3.4}$$

is a probability density on $(\mathbb{R}^d, \mathcal{B}, P)$.

4 Random variables

Definition 4.1 ($\mathbb{R}^d$*-valued random variable*)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. A mapping*

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

is called an d*-dimensional* **random variable** *if for each* $B \in \mathcal{B}$ *(*$\mathcal{B}$ *is the Borel* σ*-algebra over* $\mathbb{R}^d$*) one has*

$$\boldsymbol{\xi}^{-1}(B) \in \mathcal{F}$$

i.e. if $\boldsymbol{\xi}$ *is* $\mathcal{F}$*-measurable.*

The definition associates to each event a Borel subset. More generally if $\mathbb{S}$ is a set and $\mathcal{S}$ a σ-algebra on it such that the pair $(\mathbb{S}, \mathcal{S})$ is a measurable space then we have

Definition 4.2 ($\mathbb{S}$*-valued random variable*)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. A mapping*

$$\boldsymbol{\xi} : \Omega \to \mathbb{S}$$

is an $\mathbb{S}$*-valued random variable if for each* $S \in \mathcal{S}$ *one has*

$$\boldsymbol{\xi}^{-1}(S) \in \mathcal{F}$$

See [3] **1.3** for further details.

Example 4.1 (*Indicator function*)**.** Let $F \in \mathcal{F}$. The indicator function of F is

$$\chi_F(\omega) = \begin{cases} 1 & \text{if } \omega \in F \\ 0 & \text{if } \omega \notin F \end{cases}$$

Example 4.2 (*Simple function*)**.** Let $\{F_i\}_{i=1}^m \in \mathcal{F}$ are disjoint (i.e. $F_i \cap F_j = \emptyset$) and form a partition of Ω (i.e. $\cup_{i=1}^m F_i = \Omega$) and $\{x_i\}_{i=1}^m \in \mathbb{R}$ then

$$\xi = \sum_{i=1}^m x_i \chi_{F_i}(\omega)$$

is a random variable, called a **simple function**.

Lemma 4.1. *Let*

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

be a random variable. Then

$$\mathcal{F}(\boldsymbol{\xi}) = \{\boldsymbol{\xi}^{-1}(B) \,|\, B \in \mathcal{B}\}$$

is a σ-algebra, called the σ-algebra generated by $\boldsymbol{\xi}$. This is the smallest sub σ-algebra of $\mathcal{F}$ with respect to which $\boldsymbol{\xi}$ is measurable.

Proof. It suffices to verify that $\mathcal{F}(\boldsymbol{\xi})$ is a σ-algebra. □

Remark 4.1 (*Meaning of measurability*)**.** : The σ-algebra $\mathcal{F}(\xi)$ encodes all the information described by the random variable $\boldsymbol{\xi}$. This means that if $\boldsymbol{\zeta}$ is a second random variable, the statement

- $\boldsymbol{\zeta} = \boldsymbol{f}(\boldsymbol{\xi})$ for some mapping $\boldsymbol{f}$ implies that $\boldsymbol{\zeta}$ is $\mathcal{F}(\boldsymbol{\xi})$-measurable.
- $\boldsymbol{\zeta}$ is $\mathcal{F}(\boldsymbol{\xi})$-measurable, implies that there exists a mapping $\boldsymbol{f}$ such that $\boldsymbol{\zeta} = \boldsymbol{f}(\boldsymbol{\xi})$.

5 Expectation values

Expectation values of generic random variables are defined following the same steps taken to define the Lebesgue integral of measurable functions. Let $(\Omega, \mathcal{F}, P)$ a probability space and ξ a simple 1-dimensional random variable

$$\xi = \sum_{i=1}^{n} x_i \chi_{F_i}$$

Definition 5.1 (*Expectation value (integral) of a* **simple** *random variable*)**.**

$$\mathrm{E}\xi := \int_\Omega dP\, \xi = \sum_{i=1}^{n} x_i P(F_i)$$

Definition 5.2 (*Expectation value (integral) of a* **non-negative** *random variable η*)**.** *For*

$$\eta : \Omega \to \mathbb{R}_+$$

we define

$$\mathrm{E}\,\eta \equiv \int_\Omega dP\, \eta := \sup_{\substack{\xi \leq \eta \\ \xi = \text{simple}}} \int_\Omega dP\, \xi$$

Definition 5.3 (*Expectation value a random variable η*)**.** *For*

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

we define

$$\eta_+ := \max\{\eta, 0\} \qquad \& \qquad \eta_- := \max\{-\eta, 0\}$$

If

$$\min\{\mathrm{E}\,\eta_+, \mathrm{E}\,\eta_-\} < \infty$$

we define the expectation variable of

$$\eta \equiv \eta_+ - \eta_-$$

as

$$\int_\Omega dP\,\eta := \int_\Omega dP\,\eta_+ - \int_\Omega dP\,\eta_-$$

With these definitions all the standard rules of Lebesgue integrals apply to expectation values.

6 Moments of a random variable

Definition 6.1 (*Distribution function*)**.** *The distribution function of a random variable* $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ *is the function*

$$\tilde{P}_{\boldsymbol{\xi}} : \mathbb{R}^d \to \mathbb{R}_+$$

such that

$$\tilde{P}_{\boldsymbol{\xi}}(\boldsymbol{x}) = P_{\boldsymbol{\xi}}(\xi_1 \leq x_1, \dots, \xi_d \leq x_d)$$

Definition 6.2 (*PDF of a random variable*)**.** *Let* $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ *be a random variable and* $\tilde{P}_{\boldsymbol{\xi}}$ *its distribution function. If there exists a* **non-negative, integrable** *function*

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

such that

$$\tilde{P}_{\boldsymbol{\xi}}(\boldsymbol{x}) = \prod_{i=1}^{d} \int_{-\infty}^{x_i} dy_i\, p_{\boldsymbol{\xi}}(\boldsymbol{y})$$

then $p_{\boldsymbol{\xi}}$ *specifies the probability density function of* $\boldsymbol{\xi}$ *(PDF).*

Lemma 6.1. *Let*

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

be a random variable, with statistics described by PDF $p_{\boldsymbol{\xi}}$*. Suppose*

$$f : \mathbb{R}^d \to \mathbb{R}$$

and

$$y = f(\boldsymbol{x})$$

Then

$$\mathrm{E} f(\boldsymbol{\xi}) = \int d^d x\, p_{\boldsymbol{\xi}}(\boldsymbol{x}) f(\boldsymbol{x})$$

Proof. Suppose first f is a simple function on $\mathbb{R}^d$. Then

$$\mathrm{E}\, f(\boldsymbol{\xi}) = \sum_{i=1}^{n} f_i \int \chi_{B_i}(\boldsymbol{\xi}) dP = \sum_{i=1}^{n} f_i P(B_i) = \sum_{i=1}^{n} \int_{B_i} p_{\boldsymbol{\xi}}(\boldsymbol{x})\, f(\boldsymbol{x})$$

Consequently the formula holds for all simple functions g and, by approximation, it holds therefore for general functions g. □

In particular for a $\mathbb{R}^d$-valued random variable we have

$$\mathrm{E}\boldsymbol{\xi} = \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, \boldsymbol{x} \qquad \text{average or mean value}$$

meaning that if $\{\boldsymbol{e}_i\}_{i=1}^d$ is the collection of the unit vectors spanning the canonical basis of $\mathbb{R}^d$ i.e.

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \qquad etc. \tag{6.1}$$

then

$$\boldsymbol{e}_i \cdot \mathrm{E}\boldsymbol{\xi} = \mathrm{E}\boldsymbol{e}_i \cdot \boldsymbol{\xi} \equiv \mathrm{E}\xi^i = \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, x^i \tag{6.2}$$

Similarly

$$\mathrm{E}(\boldsymbol{\xi} - \mathrm{E}\boldsymbol{\xi}) \otimes (\boldsymbol{\xi} - \mathrm{E}\boldsymbol{\xi}) = \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, (\boldsymbol{x} - \mathrm{E}\boldsymbol{\xi}) \otimes (\boldsymbol{x} - \mathrm{E}\boldsymbol{\xi}) \qquad \text{(co-)variance tensor}$$

stands for the rank two tensor with components

$$\begin{aligned} \mathrm{E}(\xi^i - \mathrm{E}\xi^i)(\xi^j - \mathrm{E}\xi^j) &= \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, (x^i - \mathrm{E}\, x^i)(x^j - \mathrm{E}\, x^j) \\ = \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, (x^i x^j - x^j \mathrm{E}\, x^i - x^i \mathrm{E}\, x^j) + \mathrm{E}\, x^i \mathrm{E}\, x^j &= \int d^d x \, p_{\boldsymbol{\xi}}(\boldsymbol{x}) \, x^i \, x^j - \mathrm{E}\, x^i \mathrm{E}\, x^j \end{aligned} \tag{6.3}$$

Note that for any (non-random) vector $\boldsymbol{v}$ the inequality

$$\boldsymbol{v} \cdot \mathrm{E}(\boldsymbol{\xi} - \mathrm{E}\boldsymbol{\xi}) \otimes (\boldsymbol{\xi} - \mathrm{E}\boldsymbol{\xi}) \cdot \boldsymbol{v} \equiv \mathrm{E}[(\boldsymbol{\xi} - \mathrm{E}\boldsymbol{\xi}) \cdot \boldsymbol{v}]^2 \geq 0 \tag{6.4}$$

holds true. The covariance tensor also referred to as correlation tensor is therefore positive definite and in particular strictly positive definite for any random variable $\boldsymbol{\xi}$ whose probability distribution does not degenerate on a single deterministic value.

Definition 6.3 (*Moments of a random variable*)**.** *Let*

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

we call the expectation value of the n-th power of ξ

$$\mathrm{E}\xi^n = \int_\Omega dP \, \xi^n$$

the moment of order n of ξ.

The lower order moments are those most recurrent in applications and as such are given special names such as the average and the variance.

Example 6.1. (*Average and variance of a Gaussian variable*)

- Average:

$$\mathrm{E}\,\xi = \int_{\mathbb{R}} dx\, x\, g_{\bar{x}\,\sigma}(x) = \int_{\mathbb{R}} dx\, (\bar{x} + \sigma\, x)\, g_{0\,1}(x)$$

  As

$$g_{0\,1}(x) = g_{0\,1}(-x)$$

  we find

$$\mathrm{E}\,\xi = \bar{x}$$

- Variance

$$\mathrm{E}\,(\xi - \mathrm{E}\,\xi)^2 = \sigma^2 \int_{\mathbb{R}} dx\, x^2\, g_{0\,1}(x)$$

  The remaining integral I can be evaluated for example using the identity

$$\int_{\mathbb{R}} dx\, x^2\, g_{0\,1}(x) = \left.\frac{d^2}{d\jmath^2}\right|_{\jmath=0} Z(\jmath)$$

$$Z(\jmath) := \int_{\mathbb{R}} dx\, g_{0\,1}(x)\, e^{\jmath x} = \int_{\mathbb{R}} dx\, \frac{e^{-\frac{(x-\jmath)^2}{2}}}{\sqrt{2\,\pi}} e^{\frac{\jmath^2}{2}}$$

  Performing the change of variable $x \mapsto x + \jmath$ we can therefore write

$$Z(\jmath) = e^{\frac{\jmath^2}{2}} \int_{\mathbb{R}} dx\, \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}} = e^{\frac{\jmath^2}{2}}$$

  Finally in order to prove that $g_{0\,1}(x)$ is indeed normalized to the unity we observe that

$$\left[\int_{\mathbb{R}} dx\, \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}}\right]^2 = \int_{\mathbb{R}^2} \prod_{i=1}^{2} dx_i\, \frac{e^{-\frac{x_1^2+x_2^2}{2}}}{2\,\pi} = \int_0^\infty dr\, r\, e^{-\frac{r^2}{2}} = 1$$

  The statistical properties of a Gaussian variable are therefore fully specified by its first two moments.

7 Characteristic function of a random variable

Definition 7.1 (*Characteristic function*)**.** *Let*

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

the expectation value

$$\check{p}_{\boldsymbol{\xi}}(\boldsymbol{q}) := \mathrm{E}\, e^{\imath \boldsymbol{\xi}\cdot\boldsymbol{q}}$$

is referred to as the characteristic function of the random variable

Example 7.1 (*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

$$\check{g}_{\bar{x}\,,\sigma}(q) = \sum_{n=0}^{\infty} \frac{i^n\, q^n}{\Gamma\,(n+1)} \prec \xi^n \succ$$

having used the Γ-function representation of the factorial (see appendix A). If $\bar{x} = 0$ odd moments vanish whilst for even moments we get into

$$\mathrm{E}\,\xi^{2\,n} = \sigma^{2\,n} \frac{\Gamma\,(2\,n+1)}{2^n\,\Gamma\,(n+1)} = (2\,n-1)!!\sigma^{2\,n} \tag{7.1}$$

Formally for a random variable ξ one can write

$$\mathrm{E}\,\xi^n = \frac{1}{\imath^n} \frac{d^n}{dq^n} \check{p}_\xi(q)\bigg|_{q=0} \tag{7.2}$$

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

Example 7.2 (*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 $q = 0$

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

Appendices

A 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

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

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

References

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

[2] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*, volume 13 of *Springer series in synergetics*. Springer, 2 edition, 1994.

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