1 Introduction

These notes shortly recall some basic concepts in classical probability. The main reference are sections from A to F of chapter two of [1] integrated with some extra examples, to be discussed in the exercise session.

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

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

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 $(\mathbb{R}^d, \mathcal{B}, P)$.

Example 3.4 (*Dirac mass and Dirac* δ*-function*). Let $\boldsymbol{y}$ be the coordinate of a point in $\boldsymbol{R}^d$. Define for any $B \in \mathcal{B}$

$$P_{\boldsymbol{y}}(B) = \begin{cases} 1 & \text{if } \boldsymbol{y} \in B \\ 0 & \text{if } \boldsymbol{y} \notin B \end{cases} \tag{3.5}$$

then $(\mathbb{R}^d, \mathcal{B}, P)$ is a probability space. The probability P is the Dirac mass concentrated at $\boldsymbol{x}$. The "density" associated to P is the Dirac δ-function (distribution). A possible definition of the Dirac δ-function on $\mathbb{R}$

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

The definition must be understood in weak sense. Namely, let

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

a bounded Lebesgue measurable test function then

$$\int_{\mathbb{R}} dx\, \delta(x-y)\, f(x) = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}} dx\, g_{y\,\sigma}(x) f(x) = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}} dx\, g_{0\,1}(x) f(\sigma x + y) = f(y)$$

The above chain of equalities show that the Dirac δ is not a density with respect to the standard Lebesgue measure as it has support on a set of zero Lebesgue measure. A consequence is that indefinite integral

$$H_y(x) = \int_{-\infty}^{x} dz\, \delta(z-y) = \frac{1 + \operatorname{sgn}(x-y)}{2}$$

yields

$$H_y(y) = \begin{cases} 1 & x > y \\ * & x = y \\ 0 & x < y \end{cases}$$

meaning that the result is not defined on the zero measure set $x = y$. The result may be interpreted in weak sense as the definition of the *Heaviside distribution*.

Properties of the Dirac δ distribution

In weak sense (i.e. applied to suitable test functions), the Dirac δ over $\mathbb{R}$ satisfies

i localization of the integral:

$$\int_{y-\varepsilon}^{y+\varepsilon} dx\, \delta(x-y)\, f(x) = f(y)$$

ii derivative of the Dirac δ:

$$\int_{y-\varepsilon}^{y+\varepsilon} dx\, \frac{d}{dx}\delta(x-y)\, f(x) = -\frac{df}{dy}(y) \qquad \Rightarrow \qquad f(x)\frac{d\delta}{dx}(x-y) \overset{w}{=} -\frac{df}{dx}(x)\delta(x-y)$$

iii for $h(x)$ having a simple zero $x = x_\star$ and otherwise non-vanishing and smooth in $(x_\star - \varepsilon\,, x_\star + \varepsilon)$ with $\varepsilon > 0$

$$\int_{x_\star-\varepsilon}^{x_\star+\varepsilon} dx\, f(x)\delta(h(x)) = \frac{f(x_\star)}{\left|\frac{dh}{dx}(x_\star)\right|} \qquad \Rightarrow \qquad \delta(h(x)) \overset{w}{=} \frac{\delta(x-x_\star)}{\left|\frac{dh}{dx}(x_\star)\right|}$$

iv The d-dimensional Dirac-δ

$$\delta^{(d)}(\boldsymbol{x}-\boldsymbol{y}) = \prod_{i=1}^{d}\delta(x_i - y_i) \tag{3.6}$$

maybe defined by repeating the limiting procedure on each variable e.g.

$$\delta^{(d)}(\boldsymbol{x}-\boldsymbol{y}) \overset{w}{=} \prod_{i=1}^{d}\lim_{\sigma\downarrow 0} g_{y_i\,\sigma}(x_i) \tag{3.7}$$

v Let

$$h\,:\,\mathbb{R}^d \to \mathbb{R} \tag{3.8}$$

such that

$$h(\boldsymbol{x}) = 0 \tag{3.9}$$

describes a smooth $d-1$-dimensional hyper-surface Σ embedded in $\mathbb{R}^d$, then

$$\int_{\mathbb{R}^d} d^d x\, \delta(h(\boldsymbol{x})) = \int d\Sigma\, \frac{f(\boldsymbol{x})}{\|\nabla h\|} \tag{3.10}$$

4 Random variables

Definition 4.1 (*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 $B \in \mathcal{B}$ 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.

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*)**.**

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

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

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

we define

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

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

$$\min\{\prec \eta_+ \succ, \prec \eta_- \succ\} < \infty$$

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.

Proposition 5.1 (*Chebyshev's inequality*). *If $\boldsymbol{\xi}$ is a random variable and $1 \leq n < \infty$, then*

$$P(\|\xi\| \geq x) \leq \frac{1}{x^n} \prec \|\xi\|^n \succ \qquad \forall\, n$$

Proof.

$$\prec \|\xi\|^n \succ = \int_\Omega dP\, \|\xi\|^n \geq x^n \int_{\|\xi\| \geq x} dP\, \|\xi\|^n \equiv x^n\, P(\|\xi\| \geq x)$$

□

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

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

Definition 6.2 (*PDF of a random variable*). *Let $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ be a random variable and $P_{\boldsymbol{\xi}}$ 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

$$\prec y \succ \equiv E\{y\} = \int d^d x\, p_{\boldsymbol{\xi}}(\boldsymbol{x}) f(\boldsymbol{x})$$

In particular

$$\prec \xi^i \succ = \int d^d x\, p_{\boldsymbol{\xi}}(\boldsymbol{x})\, x^i \qquad \textit{average or mean value}$$

and

$$\prec (\xi^i - \prec \xi^i \succ)^2 \succ = \int d^d x\, p_{\boldsymbol{\xi}}(\boldsymbol{x})\, {x^i}^2 - \prec \xi^i \succ^2 \qquad \textit{variance}$$

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

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

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

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

$$\prec \xi^n \succ = \int_{\Omega} dP\, \xi^n$$

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:

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

  As

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

  we find

$$\prec \xi \succ = \bar{x}$$

- Variance

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

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

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

Namely

$$Z(\jmath) = e^{\frac{\jmath^2}{2}} \int_{\mathbb{R}^2} \prod_{i=1}^{2} dx_i\, \frac{e^{-\frac{x_1^2+x_2^2}{2}}}{2\,\pi} = e^{\frac{\jmath^2}{2}} \int_0^{\infty} dr\, r\, e^{-\frac{r^2}{2}} = e^{\frac{\jmath^2}{2}}$$

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

7 Independence

Definition 7.1 (*Conditional probability*)**.** *Let $(\Omega\,, \mathcal{F}\,, P)$ a probability space and $F_1\,, F_2$ two events in $\mathcal{F}$. Suppose*

$$P(F_1) > 0$$

Then the probability of the event F_2 **given** *the occurrence of F_1 is*

$$P(F_2|F_1) = \frac{P(F_2 \cap F_1)}{P(F_2)}$$

A clear interpretation of this definition see [1] pag. 17.

Definition 7.2 (*Independence*)**.** *F_2 is said to be independent of F_1 if*

$$P(F_2|F_1) = P(F_2) \qquad \Longleftrightarrow \qquad P(F_2 \cap F_1) = P(F_1)\, P(F_2)$$

Definition 7.3 (*Independence of random variables*)**.** *The random variables*

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

$i = 1, \dots$ are said to be independent if for all integers $1 \le k_1 < k_2 < k_m$ and all choices of Borel sets $\{B_{k_i}\}_{i=1}^m \subset \mathbb{R}^d$ the factorisation property

$$P(\boldsymbol{\xi}_{k_1} \in B_{k_1}\,, \boldsymbol{\xi}_{k_2} \in B_{k_2}\,, \dots, \boldsymbol{\xi}_{k_m} \in B_{k_m}) = \prod_{i=1}^{m} P(\boldsymbol{\xi}_{k_i} \in B_{k_i})$$

holds true.

The definition implies that if there exists a PDF

$$p_{\boldsymbol{\xi}_{k_1} \dots \boldsymbol{\xi}_{k_m}} : \underbrace{\mathbb{R}^d \times \mathbb{R}^d}_{m\,\text{times}} \to \mathbb{R}_+ \tag{7.1}$$

such that

$$P(\boldsymbol{\xi}_{k_1} \in B_{k_1}, \boldsymbol{\xi}_{k_2} \in B_{k_2}, \ldots, \boldsymbol{\xi}_{k_m} \in B_{k_m}) = \int_{B_{k_1} \times B_{k_m}} \prod_{i=1}^{m} d^d x_{k_i} p_{\boldsymbol{\xi}_{k_1} \cdots \boldsymbol{\xi}_{k_m}}(\boldsymbol{x}_{k_1}, \ldots, \boldsymbol{x}_{k_m})$$

then

$$p_{\boldsymbol{\xi}_{k_1} \cdots \boldsymbol{\xi}_{k_m}}(\boldsymbol{x}_{k_1}, \ldots, \boldsymbol{x}_{k_m}) = \prod_{i=1}^{m} p_{\boldsymbol{\xi}_{k_i}}(\boldsymbol{x}_{k_i})$$

Furthermore the characteristic function of m-independent random variables is equal to the product of the characteristic functions.

References

[1] L.C. Evans, *An Introduction to Stochastic Differential Equations*, lecture notes, `http://math.berkeley.edu/~evans/`. 1, 8