1 Independence

Definition 1.1 (*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 1.2 (*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 1.3 (*Independence of random variables*). *The random variables*

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

$i = 1, \ldots$ *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 factorization property*

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

holds true.

The definition implies that if there exists a PDF

$$p_{\boldsymbol{\xi}_{k_1} \ldots \boldsymbol{\xi}_{k_m}} : \underbrace{\mathbb{R}^d \times \mathbb{R}^d}_{m\,\text{times}} \to \mathbb{R}_+ \tag{1.1}$$

such that

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

then

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

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

2 Čebyšev's inequality

Proposition 2.1 (*Čebyšev's inequality*). *If* $\boldsymbol{\xi}$ *is a random variable and* $1 \le n < \infty$, *then*

$$P(\|\xi\| \ge x) \le \frac{1}{x^n} \mathrm{E}\, \|\xi\|^n \qquad \forall\, n$$

Proof.

$$\mathrm{E}\, \|\xi\|^n = \int_\Omega dP\ \|\xi\|^n \ge x^n \int_{\|\xi\| \ge x} dP\, \equiv x^n\, P(\|\xi\| \ge x)$$

□

3 Excursus: Dirac-δ and its uses

Example 3.1 (*Dirac mass and Dirac δ-function*)**.** Let $\boldsymbol{x}$ be the coordinate of a point in $\boldsymbol{R}^d$. Define for any $B \in \mathbb{B}$

$$P_{\boldsymbol{x}}(B) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in B \\ 0 & \text{if } \boldsymbol{x} \notin B \end{cases} \tag{3.1}$$

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

$$\delta^{(d)}(\boldsymbol{x} - \boldsymbol{y}) = \prod_{i=1}^{d} \delta(x_i - y_i) \tag{3.2}$$

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

$$\delta^{(d)}(\boldsymbol{x} - \boldsymbol{y}) \overset{w}{=} \prod_{i=1}^{d} \lim_{\sigma \downarrow 0} g_{y_i\,\sigma}(x_i) \tag{3.3}$$

v Let

$$h\,:\,\mathbb{R}^d \to \mathbb{R} \tag{3.4}$$

such that

$$h(\boldsymbol{x}) = 0 \tag{3.5}$$

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

$$\int_{\mathbb{R}^d} d^d x\, \delta(h(\boldsymbol{x})) = \int d\Sigma\, \frac{f(\boldsymbol{x})}{\|\nabla h\|} \tag{3.6}$$

3.1 Expectation of the Dirac distribution and probability density

Let

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

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

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

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

3.2 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

$$\phi = f(\xi) \tag{3.7}$$

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

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

3.3 Multi-dimensional case using the δ-function

Suppose f is one-to-one and write

$$p_{\boldsymbol{\phi}}(\boldsymbol{y}) = \mathrm{E}\, \delta^{(d)}(\boldsymbol{\phi} - \boldsymbol{y}) = \mathrm{E}\, \delta^{(d)}(f(\boldsymbol{\xi}) - \boldsymbol{y})$$

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

$$f^{-1}(\boldsymbol{y}) = \boldsymbol{x}_{\star} \tag{3.8}$$

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

$$p_{\boldsymbol{\phi}}(\boldsymbol{y}) = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} d^d z\, \frac{e^{-\frac{z^i A_{li} A_{lj} z^j + O(\sigma)}{2}}}{(2\,\pi)^{\frac{d}{2}}}\, p_{\boldsymbol{\xi}}(\boldsymbol{x}_{\star} + O(\sigma)) = \frac{p_{\boldsymbol{\xi}}(\boldsymbol{x}_{\star})}{|\det A|} \tag{3.9}$$

References

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

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

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