6 Distributions

SUMMARY

6a Dimension one

A *distribution* is a Borel probability measure.

The correspondence $P \longleftrightarrow F_P$ between distributions and cumulative distribution functions is bijective. Notation: $P = dF$; $\int \varphi \, dP = \int \varphi \, dF = \int \varphi(x) \, dF(x)$.

Every random variable X has its distribution P_X; $P_X(B) = \mathbb{P}\big(X \in B \big)$; $P_X = dF_X$. Two random variables X, Y are identically distributed if and only if $P_X = P_Y$, which is equivalent to $F_X = F_Y$.

The following three random variables are always identically distributed:

$$
\begin{aligned}
&(\Omega, \mathcal{F}, P) \xrightarrow{X} \mathbb{R}\,; && \text{(the original r.v.)} \\
&\big((0,1), \mathcal{B}|_{(0,1)}, \operatorname{mes}|_{(0,1)}\big) \xrightarrow{X^*} \mathbb{R}\,; && \text{(its quantile function)} \\
&(\mathbb{R}, \mathcal{B}, P_X) \xrightarrow{X_*} \mathbb{R}\,; && X_*(x) = x\,.
\end{aligned}
$$

Accordingly,

$$
\underbrace{\int_\Omega \varphi\big(X(\omega)\big)\, dP(\omega)}_{\mathbb{E}\,\varphi(X)} = \underbrace{\int_0^1 \varphi\big(X^*(p)\big)\, dp}_{\mathbb{E}\,\varphi(X^*)} = \underbrace{\int_{-\infty}^{+\infty} \varphi(x)\, dF_X(x)}_{\mathbb{E}\,\varphi(X_*)}\,.
$$

A *density* is a Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $f(\cdot) \ge 0$ almost everywhere, and $\int_{-\infty}^{+\infty} f(x)\, dx = 1$ (Lebesgue integral, in general). Each density determines a distribution P_f by $P_f(B) = \int_{-\infty}^{+\infty} \mathbf{1}_B(x) f(x)\, dx$, and $(P_f = P_g) \iff (f = g$ almost everywhere $)$.

A random variable X has a density f_X, if $P_X = P_{f_X}$, that is, $\mathbb{P}\big(X \in B \big) = \int_B f_X(x)\, dx$ or equivalently, $F_X(x) = \int_{-\infty}^{x} f_X(x_1)\, dx_1$. Such f_X is determined up to equivalence (if exists).

6b Dimension two

A *two-dimensional distribution* is a Borel probability measure on $\mathbb{R}^2$.

Each pair of random variables $X, Y : \Omega \to \mathbb{R}$ is a two-dimensional random variable $(X, Y) : \Omega \to \mathbb{R}^2$ having its two-dimensional distribution (called also the *joint distribution*), $P_{X,Y}(B) = \mathbb{P}\big((X,Y) \in B \big)$ for Borel sets $B \subset \mathbb{R}^2$. Corresponding *marginal distributions* are $P_X(B) = P_{X,Y}(B \times \mathbb{R})$, $P_Y(B) = P_{X,Y}(\mathbb{R} \times B)$. Cumulative distribution functions: $F_{X,Y}(x,y) = \mathbb{P}\big(X \le x \,\&\, Y \le y \big)$; marginals: $F_X(x) = F_{X,Y}(x, +\infty)$, $F_Y(y) = F_{X,Y}(+\infty, y)$. Note that $P_{X,Y}$ is not determined by P_X, P_Y, that is, $F_{X,Y}$ is not determined by F_X, F_Y.

Random variables $X, Y : \Omega \to \mathbb{R}$ are *independent,* if they satisfy any of the following equivalent conditions:

- $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$ for all $x, y \in \mathbb{R}$;
- $\mathbb{P}\big(X \in A \;\;\&\;\; Y \in B \big) = \mathbb{P}\big(X \in A \big) \cdot \mathbb{P}\big(Y \in B \big)$ for all Borel sets $A, B \subset \mathbb{R}$;
- $P_{X,Y}(A \times B) = P_X(A) \cdot P_Y(B)$ for all Borel sets $A, B \subset \mathbb{R}$;

such $P_{X,Y}$ is called a *product measure.* It is uniquely determined by its marginals. Notation:

$$P_{X,Y} = P_X \otimes P_Y\,.$$

There exist independent random variables X, Y distributed uniformly on $(0,1)$, defined on the probability space $\big((0,1), \mathcal{B}|_{(0,1)}, \operatorname{mes}|_{(0,1)}\big)$. For example,

$$\begin{aligned} X(\omega) &= \big(0.\alpha_1\alpha_3\alpha_5\dots\big)_2 = \frac{\alpha_1}{2^1} + \frac{\alpha_3}{2^2} + \frac{\alpha_5}{2^3} + \dots\,, \\ Y(\omega) &= \big(0.\alpha_2\alpha_4\alpha_6\dots\big)_2 = \frac{\alpha_2}{2^1} + \frac{\alpha_4}{2^2} + \frac{\alpha_6}{2^3} + \dots\,, \\ \text{when } \omega &= \big(0.\alpha_1\alpha_2\alpha_3\dots\big)_2 = \frac{\alpha_1}{2^1} + \frac{\alpha_2}{2^2} + \frac{\alpha_3}{2^3} + \dots \end{aligned}$$

Their joint distribution $P_{X,Y} = U(0,1) \otimes U(0,1)$ is the uniform distribution on the unit square $(0,1) \times (0,1)$, that is, $P_{X,Y}\big((a,b) \times (c,d)\big) = (b-a)(d-c)$ whenever $(a,b) \subset (0,1)$, $(c,d) \subset (0,1)$.

Two-dimensional Lebesgue measure:

$$\operatorname{mes}_2\big((a,b) \times (c,d)\big) = (b-a)(d-c)$$

for all $(a,b) \subset \mathbb{R}$, $(c,d) \subset \mathbb{R}$. Its existence and uniqueness.

Two-dimensional Lebesgue integral: for a Borel function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\underbrace{\iint_{\mathbb{R}^2} f(x,y)\,dxdy}_{\text{Lebesgue}} = \underbrace{\int_0^\infty \big(\operatorname{mes}\{(x,y) : f(x,y) > z\} - \operatorname{mes}\{(x,y) : f(x,y) < -z\}\big)\,dz}_{\text{improper Riemann}}\,;$$

two-dimensional Lebesgue integrability: $\iint_{\mathbb{R}^2} |f(x,y)|\,dxdy = \int_0^\infty \big(\operatorname{mes}\{(x,y) : |f(x,y)| > z\}\,dz < \infty$. All general properties of Lebesgue integral hold for $\iint_{\mathbb{R}^2}$.

Fubini theorem: for every Borel function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\text{(a)} \qquad \iint\limits_{\mathbb{R}^2} |f(x,y)|\,dxdy = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{+\infty} |f(x,y)|\,dy\Big)dx \in [0,\infty]\,,$$

the internal integral being a Borel function $\mathbb{R} \to [0,+\infty]$ (of x); if the integral (a) is finite, then

$$\text{(b)} \qquad \iint\limits_{\mathbb{R}^2} f(x,y)\,dxdy = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{+\infty} f(x,y)\,dy\Big)dx \in \mathbb{R}\,,$$

the internal integral coinciding almost everywhere with a Borel function $\mathbb{R} \to \mathbb{R}$ (of x).

A two-dimensional density is a Borel function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\cdot,\cdot) \ge 0$ almost everywhere, and $\iint_{\mathbb{R}^2} f(x,y)\,dxdy = 1$ (Lebesgue integral, in general). Each 2-dim density f determines a 2-dim distribution P_f by $P_f(B) = \iint_{\mathbb{R}^2} 1_B(x,y) f(x,y)\,dxdy$; and $(P_f = P_g) \iff (f = g$ almost everywhere).

Random variables $X, Y : \Omega \to \mathbb{R}$ have a joint density $f_{X,Y}$, if $P_{X,Y} = P_{f_{X,Y}}$, that is, $\mathbb{P}\big((X,Y) \in B \big) = \iint_B f_{X,Y}(x,y)\,dxdy$ or equivalently, $F_{X,Y}(x,y) = \int_{-\infty}^x dx_1 \int_{-\infty}^y dy_1\, f_{X,Y}(x_1,y_1)$. Such $f_{X,Y}$ is determined up to equivalence (if exists).

If a 2-dim distribution $P_{X,Y}$ has a density $f_{X,Y}$, then marginal distributions P_X, P_Y have (1-dim) densities

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy\,, \qquad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,.$$

Random variables $X, Y : \Omega \to \mathbb{R}$ having a joint density $f_{X,Y}$ are independent if and only if $f_{X,Y}$ is of the form $f_{X,Y}(x,y) = g(x)h(y)$ (almost everywhere), which is equivalent to

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \text{(a.e.)}$$

6c Discrete, singular, and absolutely continuous distributions

Atomic (or discrete) part of a distribution is its maximal part concentrated on a finite or countable set. A nonatomic (or continuous) distribution does not charge finite and countable sets.

Singular part of a nonatomic distribution is its maximal part concentrated on a set of zero Lebesgue measure. An absolutely continuous distribution does not charge sets of zero Lebesgue measure.

If P_X and P_Y are discrete then $P_X \otimes P_Y$ is discrete.

If P_X is discrete and P_Y is continuous then $P_X \otimes P_Y$ is singular.

If P_X and P_Y are absolutely continuous then $P_X \otimes P_Y$ is absolutely continuous.

In general, a distribution has a discrete part, a singular part, and an absolutely continuous part:

$$P_X = p_{X,\mathrm{d}} P_{X,\mathrm{d}} + p_{X,\mathrm{s}} P_{X,\mathrm{s}} + p_{X,\mathrm{ac}} P_{X,\mathrm{ac}}\,;$$
$$P_{X,Y} = p_{X,Y,\mathrm{d}} P_{X,Y,\mathrm{d}} + p_{X,Y,\mathrm{s}} P_{X,Y,\mathrm{s}} + p_{X,Y,\mathrm{ac}} P_{X,Y,\mathrm{ac}}\,;$$

$p_{X,\mathrm{d}} = \mathbb{P}\big(X \in A_X \big)$; $p_{X,\mathrm{s}} = \mathbb{P}\big(X \in S_X \big)$; $p_{X,\mathrm{ac}} = \mathbb{P}\big(X \in \mathbb{R} \setminus S_X \setminus A_X \big)$; also $p_{X,Y,\mathrm{d}} = \mathbb{P}\big((X,Y) \in A_{X,Y} \big)$, etc. The same in terms of cumulative distribution functions: $F_X(x) = p_{X,\mathrm{d}} F_{X,\mathrm{d}}(x) + p_{X,\mathrm{s}} F_{X,\mathrm{s}}(x) + p_{X,\mathrm{ac}} F_{X,\mathrm{ac}}(x)$; also for $F_{X,Y}$.

If X has a density, then P_X is absolutely continuous. If X, Y have a joint density, then $P_{X,Y}$ is absolutely continuous. The converse is also true (Radon-Nikodym theorem):

$$\text{(existence of density)} \quad \Longleftrightarrow \quad \text{(absolute continuity)}\,.$$

Expectation:

$$\mathbb{E}\varphi(X) = p_{\mathrm{d}} \underbrace{\mathbb{E}\,\big(\varphi(X) \bigm| X \in A_X \big)}_{\int \varphi\, dP_{X,\mathrm{d}}} + p_{\mathrm{s}} \underbrace{\mathbb{E}\,\big(\varphi(X) \bigm| X \in S_X \big)}_{\int \varphi\, dP_{X,\mathrm{s}}} + p_{\mathrm{ac}} \underbrace{\mathbb{E}\,\big(\varphi(X) \bigm| X \in \mathbb{R} \setminus A_X \setminus S_X \big)}_{\int \varphi\, dP_{X,\mathrm{ac}}};$$

the same for dimension two:

$$\begin{aligned}\mathbb{E}\varphi(X,Y) = p_{\mathrm{d}}\, \mathbb{E}\,\big(\varphi(X,Y) \bigm| (X,Y) \in A_{X,Y} \big) \; + \\ + p_{\mathrm{s}}\, \mathbb{E}\,\big(\varphi(X,Y) \bigm| (X,Y) \in S_{X,Y} \big) + p_{\mathrm{ac}}\, \mathbb{E}\,\big(\varphi(X,Y) \bigm| (X,Y) \in \mathbb{R}^2 \setminus A_{X,Y} \setminus S_{X,Y} \big)\;.\end{aligned}$$

Discrete case:

$$\mathbb{E}\,\varphi(X) = \sum_{x\in A} \varphi(x)\mathbb{P}\,\big(\,X = x\,\big)\,;$$
$$\mathbb{E}\,\varphi(X,Y) = \sum_{(x,y)\in A} \varphi(x,y)\mathbb{P}\,\big(\,X = x\,\&\,Y = y\,\big)\,.$$

Absolutely continuous case:

$$\mathbb{E}\,\varphi(X) = \int \varphi(x) f_X(x)\,dx\,;$$
$$\mathbb{E}\,\varphi(X,Y) = \iint \varphi(x,y) f_{X,Y}(x,y)\,dxdy\,.$$

6d Regression and correlation

All the same as in discrete probability:

$$\mathbb{E}\,(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y) \quad \Longleftarrow \quad (X,Y \text{ are independent}) \quad \Longleftrightarrow$$
$$\Longleftrightarrow \quad \forall \varphi,\psi \;\; \mathbb{E}\,\big(\varphi(X)\psi(Y)\big) = \big(\mathbb{E}\,\varphi(X)\big)\big(\mathbb{E}\,\varphi(Y)\big)\,;$$
$$\operatorname{Var}(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = \mathbb{E}\,\big(X - \mathbb{E}\,X\big)^2 = \min_{a\in\mathbb{R}} \mathbb{E}\,\big((X-a)^2\big)\,; \quad \sigma_X = \sqrt{\operatorname{Var}(X)}\,;$$
$$\operatorname{Cov}(X,Y) = \mathbb{E}\,(XY) - (\mathbb{E}\,X)(\mathbb{E}\,Y) = \mathbb{E}\,\big((X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y)\big)\,; \qquad \rho_{X,Y} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}\,.$$

Linear regression, or (optimal) linear prediction:

$$\mathbb{E}\left(\frac{Y - \mathbb{E}\,Y}{\sigma_Y} - \rho\frac{X - \mathbb{E}\,X}{\sigma_X}\right)^2 = \min_{a\in\mathbb{R}} \mathbb{E}\left(\frac{Y - \mathbb{E}\,Y}{\sigma_Y} - a\frac{X - \mathbb{E}\,X}{\sigma_X}\right)^2 = 1 - \rho^2\,.$$

6e Transformations

An open set $D \subset \mathbb{R}^2$ and a map $T : D \to \mathbb{R}^2$; $T(x,y) = \big(u(x,y), v(x,y)\big)$; its Jacobian:

$$J_T(x,y) = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}\,.$$

Call T *smooth,* if it is one-one (from D onto $T(D)$) and $J_T(x,y) \ne 0$ on D. If T is smooth, then $T(D)$ is open, $T^{-1} : T(D) \to D$ is smooth, and

$$\frac{\partial(u,v)}{\partial(x,y)} \cdot \frac{\partial(x,y)}{\partial(u,v)} = 1\,.$$

Change of variables: for a smooth T,

$$\iint_D f\big(T(x,y)\big)\,|J_T(x,y)|\,dxdy = \iint_{T(D)} f(u,v)\,dudv$$

(Lebesgue integral, in general); either both are integrable, or both are not.

Let $X, Y, U, V : \Omega \to \mathbb{R}$ be random variables, $(X, Y) \in D$ a.s., $T : D \to \mathbb{R}^2$ smooth, and $T(X, Y) = (U, V)$ a.s. Then the densities $f_{X,Y}$, $f_{U,V}$ either both exist or both do not exist, and if they exist,

$$(u, v) = T(x, y) \quad \Longrightarrow \quad \left| \frac{\partial(u, v)}{\partial(x, y)} \right| f_{U,V}(u, v) = f_{X,Y}(x, y) \, .$$

That is the 2-dim counterpart of the 1-dim formula

$$y = \varphi(x) \quad \Longrightarrow \quad \left| \frac{dy}{dx} \right| f_Y(y) = f_X(x) \, .$$

Examples:

$$\begin{matrix} U = X + Y \\ V = X \end{matrix} \quad \Longrightarrow \quad f_{U,V}(u, v) = f_{X,Y}(x, y) \, ;$$

$$\begin{matrix} X = R \cos \Phi \\ Y = R \sin \Phi \end{matrix} \quad \Longrightarrow \quad r f_{X,Y}(x, y) = f_{R,\Phi}(r, \varphi) \, .$$

Implication:

$$f_{X+Y}(u) = \int_{-\infty}^{+\infty} f_{X,Y}(x, u - x) \, dx \, ;$$

$$f_{\sqrt{X^2+Y^2}}(r) = r \int_0^{2\pi} f_{X,Y}(r \cos \varphi, r \sin \varphi) \, d\varphi \, ;$$

$$f_\Phi(\varphi) = \int_0^{\infty} r f_{X,Y}(r \cos \varphi, r \sin \varphi) \, dr \, .$$