2. Random Variables and Distributions

2.1. Random objects and random variables

- **Definition.** A random object is a measurable function $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$, where Ω is a sample space and $\mathcal{F}$ is the σ-algebra of events.
- **Definition.** A (real-valued) random variable is a measurable function $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$, where Ω is a sample space and $\mathcal{F}$ is the σ-algebra of events.

- Similarly, when Ω is a sample space and $\mathcal{F}$ is the σ-algebra of events,
  - $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow \left(\overline{\mathbb{R}}, \mathcal{B}\right)$ is an "extended (real-valued)" random variable.
  - $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ is a "(real-valued) random vector" or a "(real-valued) multivariate random variable".
  - $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow \left(\overline{\mathbb{R}}^n, \mathcal{B}\right)$ is an "extended (real-valued) random vector" or a "extended (real-valued) multivariate random variable".
- A random vector is just a vector of random variables:

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n) \quad \text{or} \quad \tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)^\top$$

2.2. Probability distributions

- Let $(\Omega, \mathcal{F}, P)$ be a probability space.
- **Definition.** The probability distribution (or distribution) of a random object $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ is a probability measure $P_{\tilde{x}}$ on $(\Omega', \mathcal{F}')$ defined by

$$P_{\tilde{x}}(B) = P\left(\tilde{x}^{-1}(B)\right) \quad \text{for all } B \in \mathcal{F}'$$

or

$$P_{\tilde{x}}(B) = P\left\{\omega \in \Omega \mid \tilde{x}(\omega) \in B\right\} = P\left\{\tilde{x} \in B\right\} \quad \text{for all } B \in \mathcal{F}'.$$

- Obviously,

$$P_{\tilde{x}}(B) = \int_B 1 dP_{\tilde{x}} \equiv \int_B dP_{\tilde{x}} \equiv \int_{\Omega'} \mathbb{I}_B(x) dP_{\tilde{x}}(x) \quad \text{for all } B \in \mathcal{F}'$$

or

$$P_{\tilde{x}}(B) = \int\limits_{\tilde{x}^{-1}(B)} 1 dP \equiv \int\limits_{\tilde{x}^{-1}(B)} dP \equiv \int_{\Omega} \mathbb{I}_{\tilde{x}^{-1}(B)}(\omega) dP(\omega) \text{ for all } B \in \mathcal{F}'.$$

- **Example:** We roll a balanced dice, $\Omega = \{1,2,3,4,5,6\}$, and consider the random variable $\tilde{x} : \left(\Omega, 2^{\Omega}, P\right) \longrightarrow (\mathbb{R}, \mathcal{B})$ defined as

$$\tilde{x}(\omega) = \begin{cases} 1 & \text{if } \omega = 1,2,3,4 \\ \\ 7 & \text{if } \omega = 5,6. \end{cases}$$

- The induced probability $P_{\tilde{x}}$ on $(\mathbb{R}, \mathcal{B})$ (or distribution of $\tilde{x}$) satisfies

$$\begin{array}{ll} P_{\tilde{x}}\{1\} = P\{1,2,3,4\} = 2/3, & P_{\tilde{x}}\{7\} = P\{5,6\} = 1/3, \\ P_{\tilde{x}}\{12\} = P(\emptyset) = 0, & P_{\tilde{x}}(-3,1) = P(\emptyset) = 0, \\ P_{\tilde{x}}[-3,1] = P\{1,2,3,4\} = 2/3, & P_{\tilde{x}}[5,8] = P\{7\} = 1/3, \\ P_{\tilde{x}}\left[\pi, \sqrt{13}\right] = P(\emptyset) = 0, & P_{\tilde{x}}(-\infty,12] = P(\Omega) = 1, \\ P_{\tilde{x}}[10,\infty) = P(\emptyset) = 0, & P_{\tilde{x}}(1,\infty) = P\{5,6\} = 1/3, \\ P_{\tilde{x}}(-\infty,2] = P\{1,2,3,4\} = 2/3, & \text{etc.} \end{array}$$

- Moreover, using the properties of the probability, we obtain the distribution for all Borel sets in $\mathbb{R}$.

- **Definition.** The support $supp\left(P_{\tilde{x}}\right)$ of the distribution of the random vector $\tilde{x}:(\Omega,\mathcal{F})\longrightarrow(\mathbb{R}^n,\mathcal{B})$ is the smallest closed subset of $\mathbb{R}^n$ whose complement has zero probability distribution, $P_{\tilde{x}}\left\{\left[supp\left(P_{\tilde{x}}\right)\right]^c\right\}=0$.

- **Definition.** Two random objects $\tilde{x}$ and $\tilde{y}$ defined on $(\Omega,\mathcal{F},P)$ and taking values on $(\Omega',\mathcal{F}')$ are equivalent (or equal) in distribution $(\tilde{x}\overset{d}{=}\tilde{y})$ if they have the same distribution, $P_{\tilde{x}}=P_{\tilde{y}}$.

- **Example:** We toss a balanced coin and consider $\tilde{x}:\left(\Omega,2^{\Omega},P\right)\longrightarrow(\mathbb{R},\mathcal{B})$, and $\tilde{y}:\left(\Omega,2^{\Omega},P\right)\longrightarrow(\mathbb{R},\mathcal{B})$ defined as

$$\tilde{x}(\omega)=\begin{cases}-1 & \text{if } \omega=H\\ \\ 1 & \text{if } \omega=T\end{cases}\qquad\text{and}\qquad \tilde{y}(\omega)=\begin{cases}-1 & \text{if } \omega=T\\ \\ 1 & \text{if } \omega=H.\end{cases}$$

Thus, $\tilde{x}\overset{d}{=}\tilde{y}$.

- An event A is sure if $A = \Omega$.
- An event A is almost sure (a.s.) if $P(A) = 1$.
- An event A is negligible if $P(A) = 0$.

- **Definition.** We say that two random objects defined on $(\Omega, \mathcal{F}, P)$ and taking values on $(\Omega', \mathcal{F}')$ are equal, $\tilde{x} = \tilde{y}$, if $\tilde{x}(\omega) = \tilde{y}(\omega)$ for all $\omega \in \Omega$.

- **Definition.** We say that two random objects defined on $(\Omega, \mathcal{F}, P)$ and taking values on $(\Omega', \mathcal{F}')$ are equal almost surely (a.s.), $\tilde{x} \overset{a.s.}{=} \tilde{y}$, if

$$P\{\tilde{x} = \tilde{y}\} = P\{\omega \in \Omega \mid \tilde{x}(\omega) = \tilde{y}(\omega)\} = 1 \ ,$$

or, equivalently, if

$$P\{\tilde{x} \neq \tilde{y}\} = P\{\omega \in \Omega \mid \tilde{x}(\omega) \neq \tilde{y}(\omega)\} = 0.$$

- Note that the concept of "a.s." is the same as that of "a.e." The only difference is that "a.e." applies to functions defined on measure spaces, whereas "a.s." applies to random objects defined on probability spaces.

- Obviously, $\tilde{x} = \tilde{y} \Longrightarrow \tilde{x} \overset{a.s.}{=} \tilde{y}$. Moreover, $\tilde{x} \overset{a.s.}{=} \tilde{y} \Longrightarrow \tilde{x} \overset{d}{=} \tilde{y}$ but the converse is not true (see the previous example of tossing a balanced coin where $\tilde{x} \overset{d}{=} \tilde{y}$ but $\tilde{x} \overset{a.s.}{\neq} \tilde{y}$ since $P\{\tilde{x} \neq \tilde{y}\} = 1$).

2.3. Distribution function and quantile function of a random variable

- Note that the distribution $P_{\tilde{x}}$ of a random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a probability measure on $(\mathbb{R}, \mathcal{B})$ and, thus, is a finite measure.

- Therefore, the distribution $P_{\tilde{x}}$ of a random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a Lebesgue-Stieltjes measure on $\mathbb{R}$ satisfying $P_{\tilde{x}}(\mathbb{R}) = 1$.

- **Definition.** The (cumulative) distribution function (cdf) $F_{\tilde{x}} : \mathbb{R} \longrightarrow \mathbb{R}$ of a random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ (or of its distribution) is the distribution function associated with the distribution $P_{\tilde{x}}$, i.e.,

$$P_{\tilde{x}}\,(a, b] = P\,\{a < \tilde{x} \leq b\} = F_{\tilde{x}}(b) - F_{\tilde{x}}(a),$$

where we make the normalization $\lim_{x \to -\infty} F_{\tilde{x}}(x) = 0$.

- Therefore,

$$P_{\tilde{x}}\,(-\infty, x] = P\,\{\tilde{x} \leq x\} = F_{\tilde{x}}(x) - \lim_{x \to -\infty} F_{\tilde{x}}(x) = F_{\tilde{x}}(x).$$

- Moreover,

$$\lim_{x \to \infty} F_{\tilde{x}}(x) = P_{\tilde{x}}\,(-\infty, \infty) = P\,\{\tilde{x} \in \mathbb{R}\} = 1.$$

- Thus, the distribution function of a random variable $\tilde{x}$ is (weakly) increasing, right-continuous, and satisfies $\lim_{x \to -\infty} F_{\tilde{x}}(x) = 0$ and $\lim_{x \to \infty} F_{\tilde{x}}(x) = 1$. Therefore, we can restrict the range of $F_{\tilde{x}}$ so that $F_{\tilde{x}} : \mathbb{R} \longrightarrow [0, 1]$ without loss of generality.

- Assume that the distribution function $F_{\tilde{x}}$ of the random variable $\tilde{x}$ is strictly increasing and continuous.

- **Definition.** The quantile function (or percentile function, percent-point function, inverse cdf or inverse distribution function) $Q_{\tilde{x}} : (0,1) \longrightarrow \mathbb{R}$ of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ (or of its distribution function) maps the probability value $p \in (0,1)$ to the value $x \in \mathbb{R}$ such that

$$F_{\tilde{x}}(x) = P\{\tilde{x} \leq x\} = p.$$

- Thus, the quantile function $Q_{\tilde{x}}$ assigns to the probability p a threshold value x so that the probability of $\tilde{x}$ being less or equal than x is p.

- Then,

$$Q_{\tilde{x}}(p) = F_{\tilde{x}}^{-1}(p) \text{ for all } p \in (0,1),$$

so that the quantile function $Q_{\tilde{x}}$ is strictly increasing and continuous.

- In the general case of cdf's that are increasing (but not necessarily strictly increasing) and right-continuous (but not necessarily continuous), the quantile function is usually defined as

$$Q_{\tilde{x}}(p) = \inf \{ x \,|\, F_{\tilde{x}}(x) \geq p \} \quad \text{for all } p \in (0,1) \,. \tag{1}$$

- Note that the quantile function is thus in general (weakly) increasing and left-continuous and satisfies the following:

$$Q_{\tilde{x}}(p) \leq x \;\text{ if and only if }\; F_{\tilde{x}}(x) \geq p.$$

2.4. Discrete random variables

- **Definition.** A random variable $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ is discrete if its range (i.e., the image of Ω) is countable or discrete (either finite or infinite). That is, a discrete random variable may take on only a countable number of distinct values.
- If Ω is discrete then $\tilde{x}$ is discrete. The converse is not true.
- Let $\{x_1, x_2, ...\}$ be the range $\tilde{x}(\Omega)$ of the discrete random variable $\tilde{x}$.

- If $\tilde{x}$ is discrete there is a countable partition $\mathcal{A} = \{A_1, A_2, ...\}$ of Ω with

$$A_i = \{\omega \in \Omega \mid \tilde{x}(\omega) = x_i\}, \quad \text{for all } x_i \in \tilde{x}(\Omega).$$

- Therefore, $A_i = \tilde{x}^{-1}(x_i)$, for all $x_i \in \tilde{x}(\Omega)$.
- The σ-algebra $\sigma(\mathcal{A})$ generated by the partition $\mathcal{A}$ is the smallest σ-algebra that makes the random variable $\tilde{x}$ measurable.
- The distribution of a discrete random variable $\tilde{x}$ (which is said to have a discrete distribution) satisfies:

$$P_{\tilde{x}}\{x_i\} = P\{\tilde{x} = x_i\} = P(A_i), \text{ for all } x_i \in \tilde{x}(\Omega).$$

- **Definition.** The probability mass function (pmf) - or just probability function -, $f_{\tilde{x}} : \tilde{x}(\Omega) \longrightarrow [0, 1]$, of a discrete random variable $\tilde{x}$ (or of a discrete distribution $P_{\tilde{x}}$) is given by

$$f_{\tilde{x}}(x) = P\{\tilde{x} = x\} = P_{\tilde{x}}\{x\}, \quad \text{for all } x \in \tilde{x}(\Omega).$$

- **Properties of the probability and distribution functions of a discrete random variable:**
- **1.**
$$\sum_{x \in \tilde{x}(\Omega)} f_{\tilde{x}}(x) = 1.$$
- **2.** Any function $f : \tilde{x}(\Omega) \longrightarrow [0, 1]$, where $\tilde{x}(\Omega)$ is countable, satisfying $\sum_{x \in \tilde{x}(\Omega)} f(x) = 1$ can serve as a probability function of a discrete distribution.
- **3.**
$$F_{\tilde{x}}(x) = \sum_{t \leq x} f_{\tilde{x}}(t), \quad \text{with } t \in \tilde{x}(\Omega).$$

- **4.**

$$P_{\tilde{x}}(B) = P\{\tilde{x} \in B\} = \sum_{x \in B} f_{\tilde{x}}(x), \quad \text{for all } B \in \mathcal{B}, \quad \text{with } x \in \tilde{x}(\Omega).$$

- **5.**

$$f_{\tilde{x}}(x) = F_{\tilde{x}}(x) - \lim_{t \to x^-} F_{\tilde{x}}(t), \text{ for } x \in \tilde{x}(\Omega).$$

In particular, if the range of $\tilde{x}$ can be ordered so that $x_1 < x_2 < \ldots < x_{i-1} < x_i < x_{i+1} < \ldots$, then $f_{\tilde{x}}(x_1) = F_{\tilde{x}}(x_1)$ and $f_{\tilde{x}}(x_i) = F_{\tilde{x}}(x_i) - F_{\tilde{x}}(x_{i-1})$ for $i = 2, 3, \ldots$

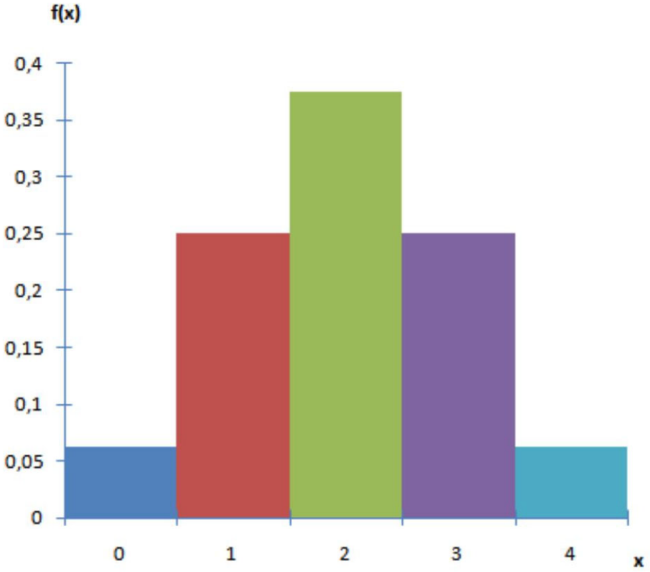
- **Example:** Let $\tilde{x}$ be the number of heads when tossing 4 coins.

$$f_{\tilde{x}}(x) = \begin{cases} 1/16 & \text{for } x = 0 \\ 4/16 & \text{for } x = 1 \\ 6/16 & \text{for } x = 2 \\ 4/16 & \text{for } x = 3 \\ 1/16 & \text{for } x = 4, \end{cases}$$

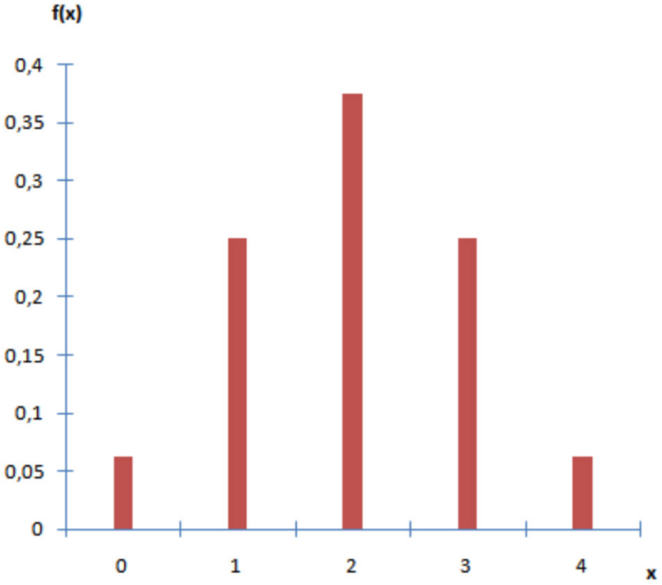
or

$$f_{\tilde{x}}(x) = \frac{1}{16}\binom{4}{x}, \quad \text{for } x = \underbrace{0, 1, 2, 3, 4}_{\tilde{x}(\Omega)}.$$

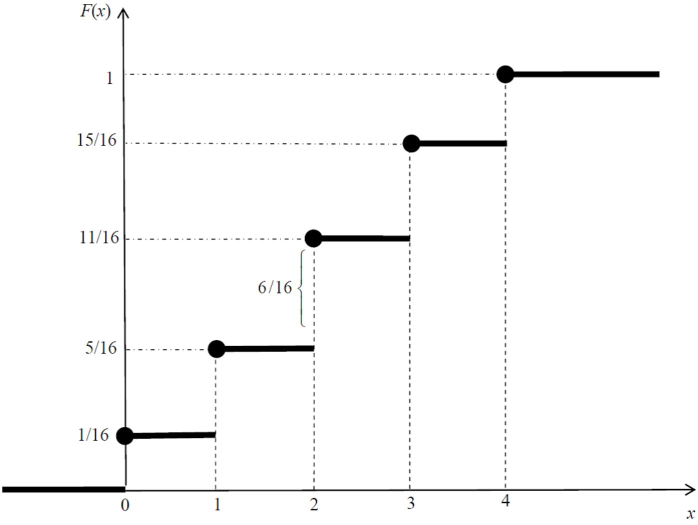
- Probability Histogram:



- Probability Bar Chart:



- Distribution function:



- **Exercise:** Draw the quantile function of this distribution and check that it is weakly increasing and left-continuous.

2.5. Continuous and absolutely continuous random variables

- **Definition 1.** A random variable $\tilde{x}$ is continuous if its range $\tilde{x}(\Omega)$ is continuous.
- **Definition 2.** A random variable $\tilde{x}$ is continuous if its distribution function $F_{\tilde{x}}$ is continuous, that is, if $P_{\tilde{x}}\{x\} = P\{\tilde{x} = x\} = 0$ for all $x \in \mathbb{R}$.
- Continuity according to Definition 2 implies continuity according to Definition 1.
- **Definition.** A random variable $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ is absolutely continuous if its distribution function $F_{\tilde{x}}$ is absolutely continuous, i.e., if there exists a Borel measurable function $f_{\tilde{x}} : (\mathbb{R}, \mathcal{B}) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$ that is integrable with respect to Lebesgue measure such that

$$F_{\tilde{x}}(x) - F_{\tilde{x}}(a) = \int_{[a,x]} f_{\tilde{x}}(t)dt, \text{ for all } a \in \mathbb{R},\ x \in \mathbb{R},\ \text{with } a \leq x.$$

- Absolute continuity implies continuity.
- Random variables that are neither discrete nor absolutely continuous are called "mixed".
- **Equivalent definition:** A random variable $\tilde{x}$ is absolutely continuous if its distribution $P_{\tilde{x}}$ is absolutely continuous with respect to Lebesgue measure.
- Therefore, thanks to the Radon-Nikodym theorem, there exists a Borel measurable function $f_{\tilde{x}} : (\mathbb{R}, \mathcal{B}) \longrightarrow \left(\overline{\mathbb{R}}, \mathcal{B}\right)$ such that

$$P_{\tilde{x}}(B) = \int_B f_{\tilde{x}}(x)dx, \text{ for all } B \in \mathcal{B}.$$

2.6. Density

- The Borel measurable function $f_{\tilde{x}} : (\mathbb{R}, \mathcal{B}) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$ such that

$$P_{\tilde{x}}(B) = \int_B f_{\tilde{x}}(x)dx, \text{ for all } B \in \mathcal{B},$$

  is called the probability density function (pdf) - or density function or just "density" - of the random variable $\tilde{x}$ (or of the distribution $P_{\tilde{x}}$).

- Since $P_{\tilde{x}}(\mathbb{R}) = 1$, the density function $f_{\tilde{x}}$ is integrable with respect to Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

- Moreover, the density $f_{\tilde{x}}$ is finite a.e. with respect to Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

- The density function $f_{\tilde{x}}$ of the random variable $\tilde{x}$ is the Radon-Nikodym derivative of its distribution with respect to Lebesgue measure, $f_{\tilde{x}} = dP_{\tilde{x}} / dx$.

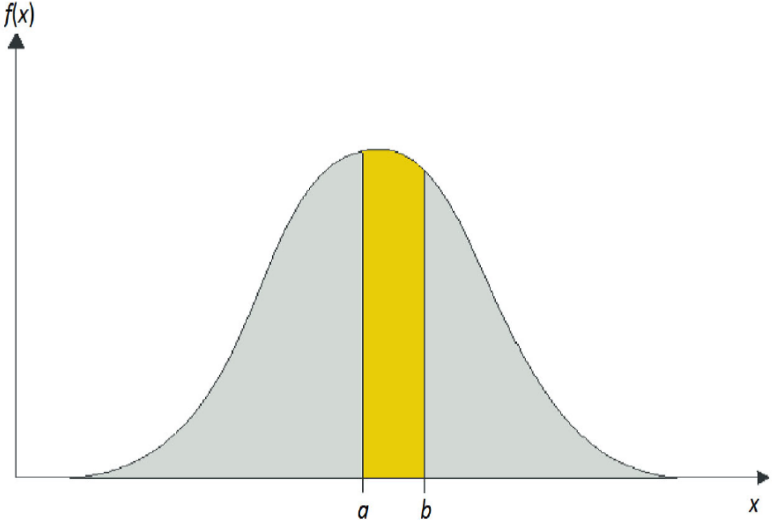
- *Note:* If $\tilde{x}$ is absolutely continuous, then

$$P_{\tilde{x}}(a,b] = P_{\tilde{x}}(a,b) = P_{\tilde{x}}[a,b] = P_{\tilde{x}}[a,b) =$$

$$F_{\tilde{x}}(b) - F_{\tilde{x}}(a) = \int_{[a,b]} f_{\tilde{x}}(x)dx.$$

- **Notation:** If the random variable $\tilde{x}$ has the distribution $P_{\tilde{x}}$, we write $\tilde{x} \sim P_{\tilde{x}}$, $\tilde{x} \sim F_{\tilde{x}}$, or $\tilde{x} \sim f_{\tilde{x}}$, where $F_{\tilde{x}}$ is the corresponding distribution function and $f_{\tilde{x}}$ is the corresponding probability or density function.

$P_{\tilde{x}}\,[a, b]$ is given by the area of the yellow region



- **Properties of the density:**
- **1.**

$$\int_{\mathbb{R}} f_{\tilde{x}}(x)dx = 1.$$

- **2.**

$$F_{\tilde{x}}(x) = \int_{(-\infty,x]} f_{\tilde{x}}(t)dt.$$

- **3.** Any non-negative (a.e. w.r.t. Lebesgue measure) Borel measurable function $f : (\mathbb{R}, \mathcal{B}) \longrightarrow \left(\overline{\mathbb{R}}, \mathcal{B}\right)$ satisfying $\int_{\mathbb{R}} f(x)dx = 1$ can serve as a density of an absolutely continuous distribution on $(\mathbb{R}, \mathcal{B})$.
- **4.** If $\tilde{x}$ is absolutely continuous, then $f_{\tilde{x}} = F'_{\tilde{x}}$ when the derivative of $F_{\tilde{x}}$ exists. Moreover, the derivative $F'_{\tilde{x}}$ exists a.e. w.r.t. Lebesgue measure. If $f_{\tilde{x}}$ is continuous at x then $F_{\tilde{x}}$ is differentiable at x and $f_{\tilde{x}}(x) = F'_{\tilde{x}}(x)$.

2.7. Random vectors

- $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}^n, \mathcal{B})$.
- $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ or $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)^{\mathsf{T}}$.
- $\tilde{x}_i = p_i(\tilde{x})$, where $p_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ is the projection to the ith coordinate.
- The distribution of the random vector $\tilde{x}$ is a probability measure on $(\mathbb{R}^n, \mathcal{B})$ given by

$$P_{\tilde{x}}(B) = P\left(\tilde{x}^{-1}(B)\right) \text{ for all } B \in \mathcal{B}(\mathbb{R}^n).$$

- The distribution function (cdf) of the random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$, $F_{\tilde{x}} : \mathbb{R}^n \longrightarrow \mathbb{R}$, is given by

$$F_{\tilde{x}}\underbrace{(x_1, x_2, ..., x_n)}_{x \in \mathbb{R}^n} = P\Big\{\underbrace{\tilde{x}_i \leq x_i, \text{ for } i = 1, 2, ..., n}_{\tilde{x} \leqq x}\Big\}.$$

- The distribution function of a random vector $\tilde{x}$ is a¡ (i) (weakly) **increasing**,...
- (Increasing: $a < b \Longrightarrow F(a) \leq F(b)$, where a and b belong to $\mathbb{R}^n$.)
- (ii) **right-continuous**,..
- (Right-continuous at x_0: $\lim\limits_{x \to x_0^+} F(x) \equiv F\left(x_0^+\right) = F\left(x_0\right)$, where $x > x_0 \in \mathbb{R}^n$)
- (iii) $F_{\tilde{x}}(x) \rightarrow 0$ if at least one of the components x_i of $x \in \mathbb{R}^n$ tends to $-\infty$, and
- (iv) $F_{\tilde{x}}(x) \rightarrow 1$ if all the components x_i, $i = 1, ..., n$, of $x \in \mathbb{R}^n$ tend to ∞.

- The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is discrete if its range $\tilde{x}(\Omega)$ is countable (or discrete).

- The probability function (pmf), $f_{\tilde{x}} : \tilde{x}_1(\Omega) \times \tilde{x}_2(\Omega) \times ... \times \tilde{x}_n(\Omega) \longrightarrow [0, 1]$, of a discrete random vector $\tilde{x}$ is given by:

$$f_{\tilde{x}}(x) = P_{\tilde{x}}\{x\} = P\left\{\underbrace{(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)}_{\tilde{x}} = \underbrace{(x_1, x_2, ..., x_n)}_{x \in \mathbb{R}^n}\right\} =$$

$$P\{\tilde{x}_i = x_i, \text{ for } i = 1, 2, ..., n\}, \text{ for all } x \in \tilde{x}_1(\Omega) \times \tilde{x}_2(\Omega) \times ... \times \tilde{x}_n(\Omega).$$

- *Note:* $\tilde{x}(\Omega) \subset \tilde{x}_1(\Omega) \times \tilde{x}_2(\Omega) \times ... \times \tilde{x}_n(\Omega)$.

- **Properties of the probability and distribution functions of a discrete random vector:**
- **1.**

$$\sum_{x\in\tilde{x}(\Omega)} f_{\tilde{x}}(x) = 1 \quad \text{or} \quad \sum_{x\in\tilde{x}_1(\Omega)\times\tilde{x}_2(\Omega)\times\ldots\times\tilde{x}_n(\Omega)} f_{\tilde{x}}(x) = 1.$$

- **2.**

$$F_{\tilde{x}}(x) = \sum_{t\leqq x} f_{\tilde{x}}(t), \quad \text{with } t = (t_1, t_2, ..., t_n) \in \tilde{x}(\Omega),$$

where $t \leqq x$ means that $t_i \leq x_i$ for $i = 1, 2, ..., n$.

- **3.**

$$P_{\tilde{x}}(B) = P\{\tilde{x} \in B\} = \sum_{x\in B} f_{\tilde{x}}(x), \quad \text{for all } B \in \mathcal{B}(\mathbb{R}^n).$$

- The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ (or its distribution) is absolutely continuous if there exists a Borel measurable function $f_{\tilde{x}} : (\mathbb{R}^n, \mathcal{B}) \longrightarrow \left(\overline{\mathbb{R}}, \mathcal{B}\right)$, called the density (pdf), that is integrable with respect to Lebesgue measure on $(\mathbb{R}^n, \mathcal{B})$, such that

$$P_{\tilde{x}}(B) = \int_B f_{\tilde{x}}\left(x_1, x_2, ..., x_n\right) d\left(x_1, x_2, ..., x_n\right), \text{ for all } B \in \mathcal{B}\left(\mathbb{R}^n\right).$$

- **Properties of the density of a random vector**:

- **1.**

$$\int_{\mathbb{R}^n} f_{\tilde{x}}(x)dx = \int_{\mathbb{R}} ... \int_{\mathbb{R}} f_{\tilde{x}}(\underbrace{x_1, ..., x_n}_{x \in \mathbb{R}^n})dx_1...dx_n = 1.$$

- **2.**

$$F_{\tilde{x}}(x) = \int_{(-\infty, x_n]} \int_{(-\infty, x_{n-1}]} ... \int_{(-\infty, x_1]} f_{\tilde{x}}(t_1, t_2, ..., t_n)dt_1 dt_2...dt_n.$$

- **3.** Any non-negative (a.e. w.r.t. Lebesgue measure) Borel measurable function $f : (\mathbb{R}^n, \mathcal{B}) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$ satisfying

$$\int_{\mathbb{R}} ... \int_{\mathbb{R}} f(x_1, x_2, ..., x_n) dx_1 dx_2, ..., dx_n = 1$$

  can serve as a density of an absolutely continuous distribution on $(\mathbb{R}^n, \mathcal{B})$.

- **4.** If the random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ is absolutely continuous, then

$$f_{\tilde{x}}(x_1, x_2, ..., x_n) = \frac{\partial^n F_{\tilde{x}}(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}.$$

  when this nth crossed partial derivative of $F_{\tilde{x}}$ exists. Moreover, this derivative exists a.e. w.r.t. Lebesgue measure on $(\mathbb{R}^n, \mathcal{B})$.

2.8. Marginal distributions

- **Definition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random vector with distribution $P_{\tilde{x}}$. The marginal distribution of $\tilde{x}_i$, for $i = 1, ..., n$, is given by

$$P_{\tilde{x}_i}(B) = P_{\tilde{x}}(\mathbb{R} \times ... \times \underset{\substack{\uparrow \\ i}}{B} \times ... \times \mathbb{R}), \quad \text{for all } B \in \mathcal{B}(\mathbb{R}).$$

- **Definition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a discrete random vector with the probability function $f_{\tilde{x}}$, the marginal probability function of $\tilde{x}_i$, for $i = 1, ..., n$, is given by

$$f_{\tilde{x}_i}(x_i) =$$

$$\sum_{x_1 \in \tilde{x}_1(\Omega)} ... \sum_{x_{i-1} \in \tilde{x}_{i-1}(\Omega)} \sum_{x_{i+1} \in \tilde{x}_{i+1}(\Omega)} ... \sum_{x_n \in \tilde{x}_n(\Omega)} f_{\tilde{x}} \underbrace{(x_1, ..., x_{i-1}, x_i, x_{i+1} ..., x_n)}_{x}.$$

for all $x_i \in \tilde{x}_i(\Omega)$.

- **Definition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be an absolutely continuous random vector with the density $f_{\tilde{x}}$, the marginal density of $\tilde{x}_i$, for $i = 1, ..., n$, is given by

$$f_{\tilde{x}_i}(x_i) = \int_{\mathbb{R}} ... \int_{\mathbb{R}} f_{\tilde{x}} \underbrace{(x_1, ..., x_{i-1}, x_i, x_{i+1} ..., x_n)}_{x} dx_1 ... dx_{i-1} dx_{i+1} ... dx_n,$$

for all $x_i \in \mathbb{R}$.

- *Note:* From the marginal probability or density functions we can construct the marginal distributions in the usual way.

- **Example 1:** The discrete random vector $(\tilde{x}, \tilde{y})$, where $\tilde{x}$ is the number of points when rolling a dice and $\tilde{y}$ is the number of heads when tossing a coin has a probability function $f_{\tilde{x},\tilde{y}}(x, y)$ summarized in the following table:

$y \setminus x$	1	2	3	4	5	6	$f_{\tilde{y}}(y)$
0	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/2$
1	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/2$
$f_{\tilde{x}}(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

- The marginal probability functions of $\tilde{x}$ and $\tilde{y}$ are summarized in the "margins".

Example 2: The absolutely continuous random vector $(\tilde{x}, \tilde{y})$ has the following density:

$$f_{\tilde{x},\tilde{y}}(x, y) = \begin{cases} \dfrac{2}{3}(x + 2y) & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

- Marginal densities:

$$f_{\tilde{x}}(x) = \int_{-\infty}^{\infty} f_{\tilde{x},\tilde{y}}(x,y)dy = \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}\left[xy+y^2\right]_0^1$$

$$= \frac{2}{3}(x+1)\,, \quad \text{for } 0 < x < 1.$$

- Therefore,

$$f_{\tilde{x}}(x) = \begin{cases} \frac{2}{3}(x+1) & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

- Similarly,

$$f_{\tilde{y}}(y) = \int_{-\infty}^{\infty} f_{\tilde{x},\tilde{y}}(x,y)dx = \int_0^1 \frac{2}{3}(x+2y)\,dx = \frac{2}{3}\left[\frac{x^2}{2}+2xy\right]_0^1$$

$$= \frac{2}{3}\left(\frac{1}{2}+2y\right) = \frac{1}{3}(1+4y)\,, \quad \text{for } 0<y<1.$$

- Therefore,

$$f_{\tilde{y}}(y) = \begin{cases} \dfrac{1}{3}(1+4y) & \text{for } 0<y<1 \\ 0 & \text{otherwise.} \end{cases}$$

- **Definition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random vector defined on $(\Omega, \mathcal{F}, P)$ with the distribution $P_{\tilde{x}}$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are said to be independent if, for all collections of sets $B_1, B_{2,}..., B_n$ belonging to $\mathcal{B}(\mathbb{R})$, we have

$$P\{\tilde{x}_1 \in B_1, ..., \tilde{x}_n \in B_n\} = P\{\tilde{x}_1 \in B_1\} \cdot ... \cdot P\{\tilde{x}_n \in B_n\}$$

or, equivalently, if the distribution of the random vector $\tilde{x}$ is equal to the product measure of the marginal distributions,

$$P_{\tilde{x}} = P_{\tilde{x}_1} \times P_{\tilde{x}_2} \times ... \times P_{\tilde{x}_n} \equiv \prod_{i=1}^{n} P_{\tilde{x}_i},$$

that is,

$$P_{\tilde{x}}\{B_1 \times B_2 ... \times B_n\} = P_{\tilde{x}_1}(B_1) \cdot P_{\tilde{x}_2}(B_2) \cdot ... \cdot P_{\tilde{x}_n}(B_n).$$

- **Definition.** Let $\tilde{x}_i : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_i, \mathcal{F}_i)$, for $i = 1, ..., n$. The random objects $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are said to be independent if, for all sets $B_1 \in \mathcal{F}_1, ..., B_n \in \mathcal{F}_n$,

$$P\left\{\tilde{x}_1 \in B_1, ..., \tilde{x}_n \in B_n\right\} = P\left\{\tilde{x}_1 \in B_1\right\} \cdot ... \cdot P\left\{\tilde{x}_n \in B_n\right\}.$$

- **Equivalent definition.** Let $\tilde{x}_i : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_i, \mathcal{F}_i)$, for $i = 1, ..., n$. The collection of random objects $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are said to be independent if the joint distribution

$$P_{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n} : \bigotimes_{i=1}^{n} \mathcal{F}_i \longrightarrow [0, 1]$$

of these n random objects is equal to the product measure of the marginal distributions,

$$P_{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n} = \prod_{i=1}^{n} P_{\tilde{x}_i},$$

where $P_{\tilde{x}_i} : \mathcal{F}_i \longrightarrow [0, 1]$ is the marginal distribution of the random object $\tilde{x}_i$, $i = 1, ..., n$, and $\bigotimes_{i=1}^{n} \mathcal{F}_i$ is the product $\sigma-$algebra.

- **Proposition.** Let $\tilde{x}_i : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_i, \mathcal{F}_i)$, for $i = 1, ..., n$, be a collection of independent random objects and $g_i : (\Omega_i, \mathcal{F}_i) \longrightarrow (\Omega'_i, \mathcal{F}'_i)$, for $i = 1, ..., n$, be measurable functions. Then, the random objects $g_i(\tilde{x}_i) : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega'_i, \mathcal{F}'_i)$, for $i = 1, ..., n$, are independent.

Proof. If

$$P\left\{\tilde{x}_1 \in B_1, ..., \tilde{x}_n \in B_n\right\} = P\left\{\tilde{x}_1 \in B_1\right\} \cdot ... \cdot P\left\{\tilde{x}_n \in B_n\right\},$$

for all sets $B_1 \in \mathcal{F}_1, ..., B_n \in \mathcal{F}_n$, then

$$P\left\{\tilde{x}_1 \in g_1^{-1}\left(B_1'\right), ..., \tilde{x}_n \in g_n^{-1}\left(B_n'\right)\right\}$$
$$= P\left\{\tilde{x}_1 \in g_1^{-1}\left(B_1'\right)\right\} \cdot ... \cdot P\left\{\tilde{x}_n \in g_n^{-1}\left(B_n'\right)\right\},$$

for all sets $B_1' \in \mathcal{F}_1', ..., B_n' \in \mathcal{F}_n'$, since $g_1^{-1}\left(B_1'\right) \in \mathcal{F}_1, ..., g_n^{-1}\left(B_n'\right) \in \mathcal{F}_n$ due to the measurability of g_i, for $i = 1, ..., n$. Therefore,

$$P\left\{g_1\left(\tilde{x}_1\right) \in B_1', ..., g_n\left(\tilde{x}_n\right) \in B_n'\right\} = P\left\{g_1\left(\tilde{x}_1\right) \in B_1'\right\} \cdot ... \cdot P\left\{g_n\left(\tilde{x}_n\right) \in B_n'\right\},$$

for all sets $B_1' \in \mathcal{F}_1', ..., B_n' \in \mathcal{F}_n'$, which proves the independency of the random objects $g_i\left(\tilde{x}_i\right) : (\Omega, \mathcal{F}, P) \longrightarrow \left(\Omega_i', \mathcal{F}_i'\right)$, for $i = 1, ..., n$. $Q.E.D.$

- **Proposition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a random vector with the distribution function $F_{\tilde{x}} : \mathbb{R}^n \longrightarrow [0, 1]$ and let $F_i : \mathbb{R} \longrightarrow [0, 1]$ be the marginal distribution function of $\tilde{x}_i$, for $i = 1, ..., n$. Then, the random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are independent if and only if

$$F_{\tilde{x}}(x_1, ..., x_n) = F_1(x_1) \cdot F_2(x_2) \cdot ... \cdot F_n(x_n),$$

for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$.

- **Proposition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be a discrete random vector with the probability function $f_{\tilde{x}} : \tilde{x}_1(\Omega) \times ... \times \tilde{x}_n(\Omega) \longrightarrow [0,1]$ and let $f_i : \tilde{x}_i(\Omega) \longrightarrow [0,1]$ be the marginal probability function of $\tilde{x}_i$, for $i = 1, ..., n$. Then, the random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are independent if and only if

$$f_{\tilde{x}}(x_1, ..., x_n) = f_1(x_1) \cdot f_2(x_2) \cdot ... \cdot f_n(x_n),$$

$$\text{for all } x = (x_1, ..., x_n) \in \tilde{x}_1(\Omega) \times ... \times \tilde{x}_n(\Omega).$$

- **Proposition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)$ be an absolutely continuous random vector with the density function $f_{\tilde{x}} : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ and let $f_i : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$ be the marginal density function of $\tilde{x}_i$, for $i = 1, ..., n$. Then, the random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are independent if and only if

$$f_{\tilde{x}}(x_1, ..., x_n) = f_1(x_1) \cdot f_2(x_2) \cdot ... \cdot f_n(x_n),$$

$$\text{for all } x = (x_1, ..., x_n) \in \mathbb{R}^n.$$

2.10. Generalized conditional probability

- Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ and let us fix the event $B \in \mathcal{F}$. From the Radon-Nikodym theorem we know that there exists a Borel measurable function $g : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})$ such that

$$\underbrace{P\left(\{\tilde{x} \in A\} \cap B\right)}_{\lambda(A)} = \int_A g(x) dP_{\tilde{x}}(x), \text{ for all } A \in \mathcal{F}',$$

  since $\lambda \ll P_{\tilde{x}}$. The function g is called the conditional probability of B given $\tilde{x} = x$ and is written as $g(x) = P(B \,|\tilde{x} = x\,)$. The conditional probability is essentially unique for a given $B \in \mathcal{F}$ (i.e., if there exists another such function h, then $g = h$ a.e. $[P_{\tilde{x}}]$).

- Therefore,

$$P\left(\{\tilde{x} \in A\} \cap B\right) = \int_A P(B \,|\tilde{x} = x\,) dP_{\tilde{x}}(x),$$

  with $P(B \,|\tilde{x} = \cdot\,) : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})\,.$

- However, sometimes the conditional probability is viewed as a measure on $(\Omega, \mathcal{F})$,

$$P(\cdot \,|\tilde{x} = x\,) : \mathcal{F} \longrightarrow \mathbb{R}.$$

- Moreover, if $A = \Omega'$, then

$$P\left(\{\tilde{x} \in \Omega'\} \cap B\right) = P\left(\Omega \cap B\right) = P(B) = \int_{\Omega'} P(B\,|\tilde{x} = x\,)dP_{\tilde{x}}(x),$$

which is a generalization of the theorem of total probability.

- Note that, if $\tilde{x}$ is an absolutely continuous random variable, then $P(B\,|\tilde{x} = x\,)$ is a conditional probability given an event $(\{\tilde{x} = x\})$ that has zero probability!

2.11. Conditional distributions

- **Definition.** Let $(\tilde{x}, \tilde{y})$ be a vector of two random objects $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_x, \mathcal{F}_x)$ and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_y, \mathcal{F}_y)$, and let $C \in \mathcal{F}_y$ be a fixed measurable set. The conditional distribution of $\tilde{y}$ given $\tilde{x} = x$ is the Borel measurable function $P_{\tilde{y}|\tilde{x}}(C\,|x) : (\Omega_x, \mathcal{F}_x) \longrightarrow (\mathbb{R}, \mathcal{B})$ given by

$$P_{\tilde{y}|\tilde{x}}(C\,|x) = P\{\tilde{y} \in C\,|\tilde{x} = x\}, \text{ for all } x \in \Omega_x,$$

  which is essentially unique w.r.t. $P_{\tilde{x}}$.

- However, sometimes the conditional distribution is viewed as a measure on $(\Omega_y, \mathcal{F}_y)$,

$$P_{\tilde{y}|\tilde{x}}(\cdot\,|x) : \mathcal{F}_y \longrightarrow \mathbb{R}.$$

- Assume that the random vector $(\tilde{x}, \tilde{y})$ is discrete with the probability function $f_{\tilde{x},\tilde{y}} : \tilde{x}(\Omega) \times \tilde{y}(\Omega) \longrightarrow [0,1]$. Then, the conditional distribution $P_{\tilde{y}|\tilde{x}}(y\,|x) = P\{\tilde{y} = y\,|\tilde{x} = x\}$ must satisfy

$$P\{\tilde{x} \in A, \tilde{y} \in C\} = \sum_{x\in A} P_{\tilde{y}|\tilde{x}}(C\,|x)\underbrace{P\{\tilde{x} = x\}}_{f_{\tilde{x}}(x)}$$

$$= \sum_{x\in A} \underbrace{\sum_{y\in C} P_{\tilde{y}|\tilde{x}}(y\,|x)}_{P_{\tilde{y}|\tilde{x}}(C|x)} f_{\tilde{x}}(x), \text{ for all } A \in \mathcal{B}(\mathbb{R}) \text{ and } C \in \mathcal{B}(\mathbb{R}). \quad (*)$$

- Let us define the conditional distribution $P_{\tilde{y}|\tilde{x}}(y\,|x)$ as follows:

$$P_{\tilde{y}|\tilde{x}}(y\,|x) = \frac{P\{\tilde{x} = x, \tilde{y} = y\}}{P\{\tilde{x} = x\}} = \frac{f_{\tilde{x},\tilde{y}}(x,y)}{f_{\tilde{x}}(x)} \equiv f_{\tilde{y}|\tilde{x}}(y\,|x),$$

for all $(x,y) \in \tilde{x}(\Omega) \times \tilde{y}(\Omega)$ with $f_{\tilde{x}}(x) > 0$.

- The function $f_{\tilde{y}|\tilde{x}}\left(\cdot\,|x\right) : \tilde{y}(\Omega) \longrightarrow [0,1]$, for all $x \in \tilde{x}(\Omega)$ such that $f_{\tilde{x}}(x) > 0$, is the conditional probability function of $\tilde{y}$ given $\tilde{x} = x$.
- The previous definition of the conditional probability function (or conditional distribution) of $\tilde{y}$ given $\tilde{x} = x$ is the right one since the expression $(*)$ becomes

$$P\left\{\tilde{x} \in A, \tilde{y} \in C\right\} = \sum_{x \in A} \sum_{y \in C} f_{\tilde{y}|\tilde{x}}\left(y\,|x\right) f_{\tilde{x}}(x)$$

$$= \sum_{x \in A} \sum_{y \in C} \frac{f_{\tilde{x},\tilde{y}}(x,y)}{f_{\tilde{x}}(x)} f_{\tilde{x}}(x) = \sum_{x \in A} \sum_{y \in C} f_{\tilde{x},\tilde{y}}(x,y),$$

$$\text{for all } A \in \mathcal{B}(\mathbb{R}) \text{ and } C \in \mathcal{B}(\mathbb{R}).$$

- Assume that the random vector $(\tilde{x}, \tilde{y})$ is absolutely continuous with the density $f_{\tilde{x},\tilde{y}} : \mathbb{R}^2 \longrightarrow \overline{\mathbb{R}}$. Then, we would like to have an expression like this:

$$P\{\tilde{x} \in A, \tilde{y} \in C\} = \int_A P_{\tilde{y}|\tilde{x}}(C\,|x)\,dP_{\tilde{x}}(x)$$

$$= \int_A P_{\tilde{y}|\tilde{x}}(C\,|x) f_{\tilde{x}}(x)dx = \int_A \underbrace{\left[\int_C f_{\tilde{y}|\tilde{x}}\,(y\,|x)\,dy\right]}_{P_{\tilde{y}|\tilde{x}}(C|x)} f_{\tilde{x}}(x)dx, \qquad (**)$$

for all $A \in \mathcal{B}(\mathbb{R})$ and $C \in \mathcal{B}(\mathbb{R})$.

- Let us define the conditional density of $\tilde{y}$ given $\tilde{x} = x$, $f_{\tilde{y}|\tilde{x}}\left(\cdot\,|x\right) : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$, for all $x \in \mathbb{R}$ such that $f_{\tilde{x}}(x) > 0$, as follows:

$$f_{\tilde{y}|\tilde{x}}\left(y\,|x\right) = \frac{f_{\tilde{x},\tilde{y}}(x,y)}{f_{\tilde{x}}(x)}, \quad \text{for all } (x,y) \in \mathbb{R}^2 \text{ with } f_{\tilde{x}}(x) > 0.$$

- The previous definition of the conditional density of $\tilde{y}$ given $\tilde{x} = x$ is the right one since the expression $(**)$ becomes

$$P\left\{\tilde{x} \in A, \tilde{y} \in C\right\} = \int_A \int_C f_{\tilde{y}|\tilde{x}}\left(y\,|x\right) f_{\tilde{x}}(x) dydx$$

$$= \int_A \int_C \frac{f_{\tilde{x},\tilde{y}}(x,y)}{f_{\tilde{x}}(x)} f_{\tilde{x}}(x) dydx = \int_A \int_C f_{\tilde{x},\tilde{y}}(x,y) dydx,$$

$$\text{for all } A \in \mathcal{B}\left(\mathbb{R}\right) \text{ and } C \in \mathcal{B}\left(\mathbb{R}\right).$$

- If the discrete (absolutely continuous) random variables $\tilde{x}$ and $\tilde{y}$ are independent then

$$f_{\tilde{y}|\tilde{x}}\left(y\,|x\right) = \frac{f_{\tilde{x},\tilde{y}}(x,y)}{f_{\tilde{x}}(x)} = \frac{f_{\tilde{x}}(x)\cdot f_{\tilde{y}}\left(y\right)}{f_{\tilde{x}}(x)} = f_{\tilde{y}}\left(y\right), \quad \text{for } f_{\tilde{x}}(x) > 0.$$

  That is, the conditional probability function (density function) is equal to the corresponding "unconditional" probability function (density function).

- In general, if the two random objects $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_x, \mathcal{F}_x)$ and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_y, \mathcal{F}_y)$ are independent then

$$P_{\tilde{y}|\tilde{x}}(C\,|x) = P_{\tilde{y}}(C), \text{ for all } C \in \mathcal{F}_y \text{ a.s. } [P_{\tilde{x}}].$$

- Note that from the conditional probability and density functions we can obtain the conditional distribution in the usual way, namely,

$$P_{\tilde{y}|\tilde{x}}(C|x) = P\{\tilde{y} \in C|\tilde{x} = x\} = \sum_{y \in C} f_{\tilde{y}|\tilde{x}}(y|x), \text{ for all } C \in \mathcal{B},$$

or

$$P_{\tilde{y}|\tilde{x}}(C|x) = P\{\tilde{y} \in C|\tilde{x} = x\} = \int_C f_{\tilde{y}|\tilde{x}}(y|x)\,dy, \text{ for all } C \in \mathcal{B},$$

where

$$P_{\tilde{y}|\tilde{x}}(C|\cdot) : (\mathbb{R}, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B})$$

or, sometimes,

$$P_{\tilde{y}|\tilde{x}}(\cdot|x) : \mathcal{B}(\mathbb{R}) \longrightarrow \mathbb{R}.$$

- Note again that, if $\tilde{x}$ is an absolutely continuous random variable, then $P_{\tilde{y}|\tilde{x}}(C|x)$ is a conditional distribution (and, hence, a conditional probability $P\{\tilde{y} \in C|\tilde{x} = x\}$) given the event $\{\tilde{x} = x\}$, which has zero probability! This conditional distribution is well defined when the marginal density of $\tilde{x}$ evaluated at x, $f_{\tilde{x}}(x)$, is strictly positive.

- **Example:** The absolutely continuous random vector $(\tilde{x}, \tilde{y})$ has the following density:

$$f_{\tilde{x},\tilde{y}}(x, y) = \begin{cases} \dfrac{2}{3}(x + 2y) & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\[2ex] 0 & \text{otherwise.} \end{cases}$$

- We have already proved that the marginal density of the random variable $\tilde{y}$ is

$$f_{\tilde{y}}(y) = \begin{cases} \dfrac{1}{3}(1 + 4y) & \text{for } 0 < y < 1 \\[2ex] 0 & \text{otherwise.} \end{cases}$$

- Therefore, the conditional density of $\tilde{x}$ given $\tilde{y} = y$ is

$$f_{\tilde{x}|\tilde{y}}\left(x\,|y\right) = \begin{cases} \dfrac{\frac{2}{3}\left(x+2y\right)}{\frac{1}{3}\left(1+4y\right)} = \dfrac{2x+4y}{1+4y} & \text{for } 0 < x < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

for $0 < y < 1$.

- Thus, the conditional density of $\tilde{x}$ given $\tilde{y} = 1/4$ is

$$f_{\tilde{x}|\tilde{y}}\left(x\,\middle|\,\frac{1}{4}\right) = \begin{cases} \dfrac{2x+1}{2} = \dfrac{1}{2}\left(2x+1\right) & \text{for } 0 < x < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$P_{\tilde{x}|\tilde{y}}\left(\left(-\infty,\frac{1}{3}\right]\middle|\frac{1}{4}\right)=P\left\{\tilde{x}\leq\frac{1}{3}\middle|\tilde{y}=\frac{1}{4}\right\}=\int_{-\infty}^{1/3}f_{\tilde{x}|\tilde{y}}\left(x\middle|\frac{1}{4}\right)dx$$

$$=\int_{0}^{1/3}\frac{1}{2}(2x+1)\,dx=\frac{1}{2}\left[x^2+x\right]_0^{1/3}=\frac{1}{2}\left(\frac{1}{9}+\frac{1}{3}\right)=\frac{1}{2}\cdot\frac{4}{9}=\frac{4}{18}=\frac{2}{9},$$

while

$$P_{\tilde{x}}\left(-\infty,\frac{1}{3}\right]=\int_{-\infty}^{1/3}\underbrace{\int_{-\infty}^{\infty}f_{\tilde{x},\tilde{y}}(x,y)dy}_{f_{\tilde{x}}(x)}dx$$

$$=\int_{0}^{1/3}\underbrace{\int_{0}^{1}\frac{2}{3}(x+2y)\,dy}_{\frac{2}{3}(x+1)}dx=\frac{7}{27}.$$

- If we have more than 2 random variables, we can generalize the previous conditional probability and density functions.
- **Example:**

$$f_{\tilde{x}_1,\tilde{x}_3|\tilde{x}_2,\tilde{x}_4}(x_1, x_3 | x_2, x_4) = \frac{f_{\tilde{x}_1,\tilde{x}_2,\tilde{x}_3,\tilde{x}_4}(x_1, x_2, x_3, x_4)}{f_{\tilde{x}_2,\tilde{x}_4}(x_2, x_4)},$$

where

$$f_{\tilde{x}_2,\tilde{x}_4}(x_2, x_4) = \sum_{x_1 \in \tilde{x}_1(\Omega)} \sum_{x_3 \in \tilde{x}_3(\Omega)} f_{\tilde{x}_1,\tilde{x}_2,\tilde{x}_3,\tilde{x}_4}(x_1, x_2, x_3, x_4) > 0,$$

or

$$f_{\tilde{x}_2,\tilde{x}_4}(x_2, x_4) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{\tilde{x}_1,\tilde{x}_2,\tilde{x}_3,\tilde{x}_4}(x_1, x_2, x_3, x_4)\, dx_1 dx_3 > 0.$$

Exercises. Probability and Statistics. IDEA.
2. Random Variables and Distributions

1. Verify that $f(x) = \dfrac{2x}{k(k+1)}$ for $x = 1, 2, 3, ..., k$ can serve as the probability function for a random variable.

2. For what values of k can

$$f(x) = (1-k)\,k^x \quad \text{for } x = 0, 1, 2, ...$$

serve as the probability function of a random variable?

3. If the density of the random variable $\tilde{z}$ is given by

$$f(z) = \begin{cases} kze^{-z^2} & \text{for } z > 0 \\ 0 & \text{for } z \leq 0, \end{cases}$$

(a) Find the value of k;

(b) Find the distribution function of this random variable;

(c) Sketch the graphs of the density and the distribution functions;

(d) Find the quantile function of this random variable and sketch its graph.

4. If the density of the random variable $\tilde{z}$ is given by

$$f(z) = \begin{cases} -kz & \text{for } -1 < z < 0 \\ kz & \text{for } 0 \leq z < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

(a) Find the value of k;

(b) Find the distribution function of this random variable and sketch its graph;

(c) Compute $P\left\{-\frac{1}{2} < \tilde{z} < \frac{1}{2}\right\}$;

(d) Find the quantile function of this random variable and sketch its graph.

5. In a certain city the daily consumption of water (in millions of liters) is a random variable whose density is given by

$$f(x) = \begin{cases} \frac{1}{9}xe^{-x/3} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

What are the probabilities that on a given day

(a) the water consumption in this city is no more than 6 million liters;

(b) the water supply is inadequate if the daily capacity of this city is 9 million liters?

6. If the joint density of $\tilde{x}$ and $\tilde{y}$ is given by

$$f(x, y) = \begin{cases} \dfrac{1}{y} & \text{for } 0 < x < y,\ 0 < y < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

find the probability that the sum of the values taken on by the two random variables will exceed $1/2$.

7. If $\widetilde{p}$, the price of a certain commodity (in dollars), and $\widetilde{s}$, total sales (in 10 000 units), are random variables whose distribution can be approximated with the joint density

$$f(p, s) = \begin{cases} 5pe^{-ps} & \text{for } 0.20 < p < 0.40,\ s > 0 \\ 0 & \text{elsewhere,} \end{cases}$$

find the probabilities that

(a) the price will be less than 30 cents and sales will exceed 20 000 units;

(b) the price will be between 25 cents and 30 cents and sales will be less than 10 000 units.

8. Given the joint probability function

$$f(x, y, z) = \frac{xyz}{108} \quad \text{for } x = 1, 2, 3;\ y = 1, 2, 3;\ z = 1, 2,$$

find

(a) the joint marginal probability function of $\tilde{x}$ and $\tilde{y}$;

(b) the joint marginal probability function of $\tilde{x}$ and $\tilde{z}$;

(c) the marginal probability function of $\tilde{x}$;

(d) the conditional probability function of $\tilde{z}$ given $\tilde{x} = 1$ and $\tilde{y} = 2$;

(e) the joint conditional probability function of $\tilde{y}$ and $\tilde{z}$ given $\tilde{x} = 3$.

9. Check whether the random variables $\tilde{x}$ and $\tilde{y}$ are independent, if their joint probability function is given by

(a) $f(x, y) = 1/4$ for $x = -1$ and $y = -1$, $x = -1$ and $y = 1$, $x = 1$ and $y = -1$, and $x = 1$ and $y = 1$;

(b) $f(x, y) = 1/3$ for $x = 0$ and $y = 0$, $x = 0$ and $y = 1$, and $x = 1$ and $y = 1$.

10. If the joint density of $\tilde{x}$ and $\tilde{y}$ is given by

$$f(x,y) = \begin{cases} \frac{1}{4}(2x+y) & \text{for } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{elsewhere,} \end{cases}$$

find

(a) the marginal density of $\tilde{x}$;

(b) the marginal density of $\tilde{y}$;

(c) the conditional density of $\tilde{x}$ given $\tilde{y} = 1$;

(d) the conditional density of $\tilde{y}$ given $\tilde{x} = 1/4$.

11. If $\tilde{x}$ is the amount (in dollars) a salesperson spends on gasoline during a day and $\tilde{y}$ is the amount (in dollars) for which the salesperson is reimbursed, and the joint density of these random variables is given by

$$f(x,y) = \begin{cases} \frac{20-x}{25x} & \text{for } 10 < x < 20,\ \frac{x}{2} < y < x \\ 0 & \text{elsewhere,} \end{cases}$$

find

(a) the marginal densities of $\tilde{x}$ and $\tilde{y}$;

(b) the conditional density of $\tilde{y}$ given $\tilde{x} = 12$;

(c) the probability that the salesperson will be reimbursed at least \$8 when spending \$12.

12. Give an example of random variables $\tilde{x}$ and $\tilde{y}$ (on the same probability space) such that $\tilde{x}$ and $\tilde{y}$ each have densities, but the random vector $(\tilde{x}, \tilde{y})$ does not.

13. (a) Let $\tilde{x}$ be an absolutely continuous random variable with density $f_{\tilde{x}}(x)$. If $\tilde{x} = x$, let $\tilde{y}$ be discrete, with the probability function $f_{\tilde{y}|\tilde{x}}(y\,|x) = P\{\tilde{y} = y \mid \tilde{x} = x\}$ specified. Show that there is a conditional density of $\tilde{x}$ given $\tilde{y}$, namely,

$$f_{\tilde{x}|\tilde{y}}(x\,|y) = \frac{f_{\tilde{x}}(x) f_{\tilde{y}|\tilde{x}}(y\,|x)}{f_{\tilde{y}}(y)},$$

where

$$f_{\tilde{y}}(y) = P\{\tilde{y} = y\} = \int_{\mathbb{R}} f_{\tilde{x}}(x) f_{\tilde{y}|\tilde{x}}(y\,|x)\,dx > 0.$$

(b) Let $\tilde{x}$ be a discrete random variable with the probability function $f_{\tilde{x}}(x)$. If $\tilde{x} = x$, let $\tilde{y}$ be absolutely continuous, with the density function $f_{\tilde{y}|\tilde{x}}(y\,|x)$ specified. Show that there is a conditional probability function of $\tilde{x}$ given $\tilde{y}$, namely,

$$f_{\tilde{x}|\tilde{y}}(x\,|y) = \frac{f_{\tilde{x}}(x) f_{\tilde{y}|\tilde{x}}(y\,|x)}{f_{\tilde{y}}(y)},$$

where

$$f_{\tilde{y}}(y) = \sum_{x \in \tilde{x}(\Omega)} f_{\tilde{x}}(x) f_{\tilde{y}|\tilde{x}}(y\,|x) > 0.$$

14. If $\tilde{x}$ is a random vector with density f, and $A = \{\tilde{x} \in B_0\}$, $B_0 \in \mathcal{B}(\mathbb{R}^n)$, show that there is a conditional density for $\tilde{x}$ given A, namely,

$$f_{\tilde{x}|A}(x\,|A) = \begin{cases} \dfrac{f(x)}{P(A)} & \text{if } x \in B_0 \\ 0 & \text{if } x \notin B_0. \end{cases}$$

Note: The interpretation of the conditional density is that

$$P\{\tilde{x} \in B\,|A\} = \int_B f_{\tilde{x}|A}(x\,|A)\,dx, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

15. Assume that the distribution of the random variable $\tilde{x}$ has the following density:

$$f(x) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

(a) Use the definition of conditional probability given in Chapter 1 (Probability) to compute

$P\{\tilde{x} < 1/3 \mid \tilde{x} < 1/2\}$, $P\{\tilde{x} < 3/4 \mid \tilde{x} < 1/2\}$, and $P\{1/3 < \tilde{x} < 3/4 \mid \tilde{x} < 1/2\}$.

(b) Use the conditional density found in Exercise 14 of this list to compute the same conditional probabilities of part (a).

(c) Consider the random variable $\tilde{x}$ having the distribution $P_{\tilde{x}}$. Find the conditional probability $P\{\tilde{x} \in B \mid \tilde{x} = x\}$ and check that it satisfies the theorem of total probability,

$$P\{\tilde{x} \in B\} = \int_{\mathbb{R}} P\{\tilde{x} \in B \mid \tilde{x} = x\}\,dP_{\tilde{x}}(x).$$

16. If the joint density of $\tilde{x}$ and $\tilde{y}$ is given by

$$f(x,y) = \begin{cases} x + y & \text{for } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

find the corresponding joint distribution function $F(x,y)$. Check that

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y) \quad \text{a.e. with respect to Lebesgue measure.}$$

17. The random vector $(\tilde{x}, \tilde{y})$ is distributed according to the following joint density function:

$$f(x,y) = \begin{cases} c(x^2 - y) & \text{if } x \in (1,2)\,,\ y \in (0,1) \\ \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

(b) Find the marginal density function of $\tilde{x}$, $f_{\tilde{x}}(x)$, and the marginal distribution function of $\tilde{y}$, $F_{\tilde{y}}(y)$.

(c) Are $\tilde{x}$ and $\tilde{y}$ independent random variables?

(d) Find the conditional density function of $\tilde{x}$ given $\tilde{y} = 1/4$, $f_{\tilde{x}|\tilde{y}}\left(x \,\middle|\, \frac{1}{4}\right)$.

18. The joint and marginal distribution functions of the random variables $\tilde{x}$ and $\tilde{y}$ are $F(x,y)$, $F_{\tilde{x}}(x)$ and $F_{\tilde{y}}(y)$. Let us consider the following functions:

$$F^+(x,y) = \min\left\{F_{\tilde{x}}(x), F_{\tilde{y}}(y)\right\},$$

and

$$F^-(x,y) = \max\left\{F_{\tilde{x}}(x) + F_{\tilde{y}}(y) - 1, 0\right\}.$$

(a) Prove that both $F^+(x,y)$ and $F^-(x,y)$ are distribution functions associated with probability measures on the $\sigma-$algebra $\mathcal{B}\left(\mathbb{R}^2\right)$ of the Borel sets of $\mathbb{R}^2$.

(b) Prove that the marginal distribution functions of both $F^+(x,y)$ and $F^-(x,y)$ are also $F_{\tilde{x}}(x)$ and $F_{\tilde{y}}(y)$.

(c) Prove that $F^-(x,y) \leq F(x,y) \leq F^+(x,y)$ for all $(x,y) \in \mathbb{R}^2$.