Random vectors I.

(Random vectors, independence, conditional distribution)

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Probability and Statistics

BIE-PST, WS 2023/24, Lecture 6

Content

- **Probability theory:**
  - Events, probability, conditional probability, Bayes' Theorem, independence of events.
  - Random variables, distribution function, functions of random variables, characteristics of random variables: expected value, variance, moments, generating function, quantiles, critical values, important discrete and continuous distributions.
  - Random vectors, joint and marginal distributions, functions of random vectors, independence of random variables, conditional distribution, conditional expected value, covariance and correlation.
  - Markov's and Chebyshev's inequality, weak law of large numbers, strong law of large numbers, Central limit theorem.
- **Mathematical statistics:**
  - Point estimators, sample mean, sample variance, properties of point estimators, Maximum likelihood method.
  - Interval estimators, hypothesis testing, one-sided vs. two-sided alternatives, linear regression, estimators of regression parameters, testing of linear model.

Recap

- A **random variable** X is a measurable function, which assigns real values to the outcomes of a random experiment.
- The **distribution** of X gives the information of the probabilities of its values and is uniquely given by the **distribution function**:

$$F_X(x) = \mathrm{P}(X \leq x).$$

- There are two major types of random variables:
  - **Discrete**, taking only countably many possible values.
  - **Continuous**, taking uncountably many values from an interval.
- The distribution can be given by:
  - for discrete distributions by the **probabilities** of possible values $\mathrm{P}(X = x_k)$.
  - for continuous distributions by the **density** f_X for which

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

Random vectors

Sometimes we can measure several random variables at once from one result of an experiment.

The individual variables can have different distributions and the values of the variables can be strongly mutually interconnected. It is appropriate to describe their distribution together as the so called **joint distribution**.

Definition

Consider two random variables X and Y defined on the same probability space $(\Omega, \mathcal{F}, \mathrm{P})$. We define their **joint distribution function** $F_{X,Y}(x, y)$ as

$$F_{X,Y}(x, y) = \mathrm{P}(X \leq x \cap Y \leq y).$$

For n random variables $X_1, X_2, \ldots, X_n \overset{\text{denote}}{=} \boldsymbol{X}$ we define the joint distribution function as

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathrm{P}(X_1 \leq x_1 \cap \ldots \cap X_n \leq x_n).$$

The couple (X, Y) or, n-tuple $(X_1, \ldots, X_n)$, is called a **random vector**.

Example – joint distribution

Example

Let X and Y be random variables with a joint discrete distribution given by the following probabilities:

$\mathrm{P}(X = x \cap Y = y)$		$x = 0.5$	$x = 1$	$x = 2$
y	2	0.3	0.06	0.04
	1	0.4	0.15	0.05

Compute the joint distribution function $F_{X,Y}(x,y) = \mathrm{P}(X \leq x \cap Y \leq y)$*:*

Properties of the joint distribution function

The joint distribution function has analogous properties as the distribution function of one variable.

Theorem

The joint distribution function $F_{X,Y}$ of random variables X and Y has following properties:

- **i)** *if $x_1 < x_2$ and $y_1 < y_2$ then $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$.*
- **ii)** $\forall y \in \mathbb{R}, \lim_{x\to-\infty} F_{X,Y}(x, y) = 0$ *and*
  $\forall x \in \mathbb{R}, \lim_{y\to-\infty} F_{X,Y}(x, y) = 0$.
- **iii)** $\forall y \in \mathbb{R}, \lim_{x\to+\infty} F_{X,Y}(x, y) = F_Y(y)$ *and*
  $\forall x \in \mathbb{R}, \lim_{y\to+\infty} F_{X,Y}(x, y) = F_X(x)$.

Proof

Analogously as for the distribution function of one random variable. □

Vectors of discrete random variables

A distribution of random variables X and Y on the same probability space is described by the joint distribution function

$$F_{X,Y}(x, y) = \mathrm{P}(X \leq x \cap Y \leq y).$$

If the variables X and Y are discrete, it is often useful to describe the distribution by the **joint probabilities** of their values.

Definition

The joint probabilities of values of two discrete random variables X and Y is

$$\mathrm{P}(X = x \cap Y = y) = \mathrm{P}(\{X = x\} \cap \{Y = y\}).$$

Taken as a function of x and y, the probabilities are called the **joint probability mass function**.

Joint probabilities and the joint distribution function

The **joint distribution function** of two discrete random variables X and Y is

$$F_{X,Y}(x,y) = \mathrm{P}(X \le x \cap Y \le y) = \sum_{\{i:\, x_i \le x\}} \sum_{\{j:\, y_j \le y\}} \mathrm{P}(X = x_i \cap Y = y_j)$$

From this it follows that $F_{X,Y}(x,y)$ has a stepwise structure.

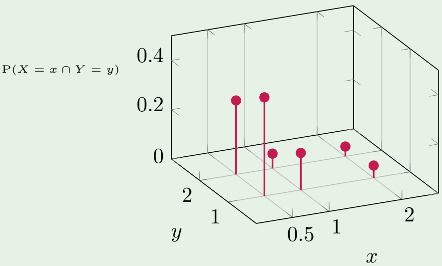
The **normalization condition** follows from the properties of the joint distribution function:

$$\begin{aligned}
&\sum_i \sum_j \mathrm{P}(X = x_i \cap Y = y_j) = \sum_i \mathrm{P}\left(\{X = x_i\} \cap \bigcup\nolimits_j \{Y = y_j\}\right) \\
&= \sum_i \mathrm{P}\left(\{X = x_i\} \cap \{Y \in \mathbb{R}\}\right) = \sum_i \mathrm{P}\left(X = x_i\right) \\
&= \mathrm{P}\left(\bigcup\nolimits_j \{X = x_i\}\right) = \mathrm{P}\left(\{X \in \mathbb{R}\}\right) = \mathrm{P}(\Omega) = 1.
\end{aligned}$$

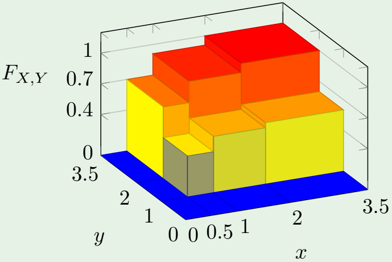
Joint distribution – visualization

Example

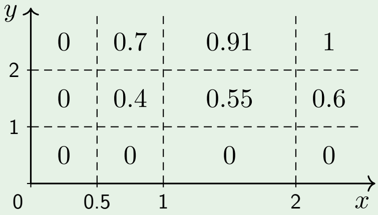
Joint probabilities



Joint distribution function



$P(X = x \cap Y = y)$		x: 0.5	1	2
y	2	0.3	0.06	0.04
	1	0.4	0.15	0.05



Marginal distribution

Sometimes we have the joint distribution of variables X and Y, but we are not interested in the values of Y. From the joint distribution function $F_{X,Y}$ we would then want to obtain only the distribution function F_X of the variable X.

The distribution obtained this way is called the **marginal** distribution of random variable X.

Theorem

*Let $\mathrm{P}(X = x \cap Y = y)$ be the joint probabilities of values of two discrete variables X and Y. The **marginal distribution** (or **marginal probabilities**) of a X is given by*

$$\mathrm{P}(X = x) = \sum_j \mathrm{P}(X = x \cap Y = y_j).$$

Proof

The events $\{Y = y_j\}$ for $j = 1, 2, \ldots$ create a countable partition of Ω. From this follows:

$$\mathrm{P}(X = x) = \mathrm{P}(\{X = x\} \cap \{Y \in \mathbb{R}\}) = \mathrm{P}(\{X = x\} \cap (\textstyle\bigcup_j \{Y = y_j\})) =$$

$$= \mathrm{P}\left(\bigcup_j (\{X = x\} \cap \{Y = y_j\})\right) = \sum_j \mathrm{P}(\{X = x\} \cap \{Y = y_j\}).$$

□

Example – marginal distribution

Example

Let X and Y be two random variables with the following joint distribution:

$\mathrm{P}(X = x \cap Y = y)$		x = 0.5	1	2	$\mathrm{P}(Y = y)$
y	2	0.3	0.06	0.04	0.4
	1	0.4	0.15	0.05	0.6
	$\mathrm{P}(X = x)$	0.7	0.21	0.09	

Find the marginal distribution of X and Y separately (find the marginal probabilities $\mathrm{P}(X = x)$ and $\mathrm{P}(Y = y)$.)

$$\mathrm{P}(Y = y) = \begin{cases} 0.6 & \text{for } y = 1 \\ 0.4 & \text{for } y = 2 \\ 0 & \text{elsewhere} \end{cases} \qquad \mathrm{P}(X = x) = \begin{cases} 0.7 & \text{for } x = 0.5 \\ 0.21 & \text{for } x = 1 \\ 0.09 & \text{for } x = 2 \\ 0 & \text{elsewhere} \end{cases}$$

Independence of random variables

Similarly as with random events, we want to be able to determine, whether the knowledge of one variable changes in some way the distribution of an other one.

Definition

Random variables X and Y are called independent if for all $x,\ y \in \mathbb{R}$ the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent. Equivalently, if it holds that for all $x, y \in \mathbb{R}$

$$\mathrm{P}(X \leq x \cap Y \leq y) = \mathrm{P}(X \leq x) \cdot \mathrm{P}(Y \leq y).$$

Random variables $X_1, \ldots, X_n$ are called independent if for all $\boldsymbol{x} \in \mathbb{R}^n$ it holds that

$$\mathrm{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \prod_{i=1}^{n} \mathrm{P}(X_i \leq x_i).$$

Random variables forming a countable collection $X_1, X_2, \ldots$ are called independent if all finite n-tuples $X_{i_1}, \ldots, X_{i_n}$ are independent.

Independence of discrete random variables

For discrete random variables we can verify the independence by means of the probabilities of values:

Theorem

Discrete random variables X and Y are independent if for all $x,\ y \in \mathbb{R}$ the events $\{X = x\}$ and $\{Y = y\}$ are independent. Equivalently, if it holds that for all $x, y \in \mathbb{R}$

$$\mathrm{P}(X = x \cap Y = y) = \mathrm{P}(X = x) \cdot \mathrm{P}(Y = y).$$

Random variables $X_1, \ldots, X_n$ are independent if for all $\boldsymbol{x} \in \mathbb{R}^n$ it holds that

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i).$$

Proof

If the condition regarding equalities holds, it must hold also for all inequalities, because they can be rewritten as sums of probabilities of disjoint events.

If the condition regarding inequalities hold, it must hold also for all equalities, because the difference of probabilities of inequalities yields probabilities of equalities. □

Example – checking independence of random variables

Example – continuation

Random variables X and Y have the following joint and marginal distributions:

$\mathrm{P}(X = x \cap Y = y)$		x: 0.5	x: 1	x: 2	$\mathrm{P}(Y = y)$
y	2	0.3	0.06	0.04	0.4
	1	0.4	0.15	0.05	0.6
$\mathrm{P}(X = x)$		0.7	0.21	0.09	

Are X and Y independent?

No, they are not independent because e.g. for $x = 0.5$ and $y = 2$ it holds that

$$0.3 = \mathrm{P}(X = 0.5 \cap Y = 2) \neq \mathrm{P}(X = 0.5) \cdot \mathrm{P}(Y = 2) = 0.7 \cdot 0.4 = 0.28.$$

Vectors of continuous random variables

The distribution of random variables X and Y on the same probability space is described by the joint distribution function

$$F_{X,Y}(x,y) = \mathrm{P}(X \leq x \cap Y \leq y).$$

If the variables X and Y are continuous, it is often useful to describe the distribution by the

joint probability density.

Definition

Two random variables X and Y have a **joint (absolutely) continuous** distribution if there exists a **non-negative** function $f_{X,Y} : \mathbb{R}^2 \to [0,+\infty)$ such that for all $x, y \in \mathbb{R}$ it holds

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v) \,\mathrm{d}u \,\mathrm{d}v.$$

The function $f_{X,Y}$ is called the **joint probability density** of the random variables X, Y or of the random vector (X, Y).

Properties of continuous random variables

Similarly as in the one-dimensional case it holds that:

- Where the derivative exists:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y).$$

- The joint distribution function is continuous.
- **Normalization condition**: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y = 1$
- For all $x, y \in \mathbb{R}$ and all Borel sets A, B on $\mathbb{R}$

$$\mathrm{P}(X = x \cap Y \in B) = \mathrm{P}(X \in A \cap Y = y) = \mathrm{P}(X = x \cap Y = y) = 0.$$

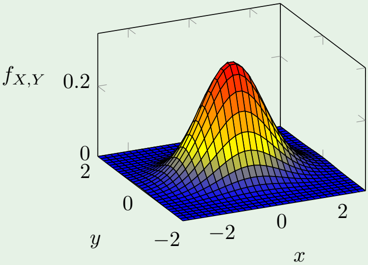
- $\mathrm{P}(\{a < X \leq b\} \cap \{c < Y \leq d\}) = \int_c^d \int_a^b f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y.$
- For all B Borel subset of $\mathbb{R}^2$ (meaning that $\{X \in B\}$ is an event)

$$\mathrm{P}\big((X,Y) \in B\big) = \iint_B f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y.$$

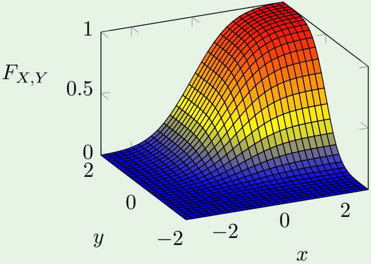
Joint distribution – visualization

Example

Joint density



Joint distribution function



$$f_{X,Y}(x,y) = \frac{1}{\pi} e^{-\frac{x^2}{2} - 2y^2}$$

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} \frac{1}{\pi} e^{-\frac{u^2}{2} - 2v^2} \, \mathrm{d}u \, \mathrm{d}v$$

Marginal distribution

For computing the marginal distribution of two variables X and Y from the joint density we can use a formula analogous to the discrete case:

Theorem

Let X and Y be two random variables having a joint continuous distribution with joint density $f_{X,Y}$. Then X and Y are both continuous too, and the ***marginal densities*** *f_X, f_Y are given by*

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

Proof

We know that:

$$F_X(x) = \mathrm{P}(X \le x) = \mathrm{P}(X \le x \cap Y \in \mathbb{R}) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f_{X,Y}(u,v)\,\mathrm{d}v \right) \mathrm{d}u\,.$$

The statement of the theorem is obtained by differentiating with respect to x, or by comparing this formula to the definition of the distribution function of a continuous random variable. The second part is analogous. □

Independence of continuous random variables

The independence of continuous random variables can be determined by means of densities.

Theorem

Two continuous random variables X and Y are called independent if and only if for all $x,\ y \in \mathbb{R}$ the following equality holds

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

Random variables $X_1, \ldots, X_n$ are called independent if for all $\boldsymbol{x} \in \mathbb{R}^n$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Proof

Two random variables X and Y are independent if

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y).$$

Taking the derivatives of both sides with respect to both x and y yields one implication. Integrating both sides of the equality for densities yields the other direction. □

Independence of continuous random variables

Remark

While verifying the independence of X and Y we can use the following: **Consequence:** If it is possible to decompose $f_{X,Y}$ to

$$f_{X,Y}(x, y) = g(x) \cdot h(y), \quad \forall x, y \in \mathbb{R},$$

where $g(x)$ and $h(y)$ are non-negative functions, then the variables X and Y are independent.

✓ Do the proof yourself by inserting into the formula for marginal densities.

✓ Beware, the functions $g(x)$ and $h(y)$ may not necessarily be the marginal densities $f_X(x)$ and $f_Y(y)$.; they may differ by a multiplicative constant.

The statement of the consequence can be formulated for independence of a general random vector $X_1, \ldots, X_n$ too.

Example – marginal distribution and independence

Example

Let X and Y random variables having the joint probability density

$$f_{X,Y}(x,y) = ye^{-2x} \qquad \text{for } x \in [0,+\infty) \text{ and } y \in [0,2].$$

Are the variables X and Y independent?

Marginal densities:

$$f_X(x) = \int_0^2 ye^{-2x}\,\mathrm{d}y = e^{-2x}\int_0^2 y\,\mathrm{d}y = e^{-2x}\left[\frac{y^2}{2}\right]_0^2 = e^{-2x}\left(\frac{4}{2}-0\right) = 2e^{-2x}.$$

$$f_Y(y) = \int_0^{+\infty} ye^{-2x}\,\mathrm{d}x = y\int_0^{+\infty} e^{-2x}\,\mathrm{d}x = y\left[\frac{e^{-2x}}{-2}\right]_0^{+\infty} = y\left(0-\frac{1}{-2}\right) = \frac{y}{2}.$$

Independence:

$$ye^{-2x} = f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y) = 2e^{-2x}\cdot\frac{y}{2} = ye^{-2x}.$$

Yes, they are independent!

Discrete conditional distribution

Now we will study the distribution of a random variable X under the assumption that we know the value of the variable $Y = y$.

Suppose that we have a partial information about the result of an experiment and we are interested in the change in our prediction.

It is reasonable to introduce the conditional distribution by means of the conditional probability under the condition of the event $\{Y = y\}$.

Definition

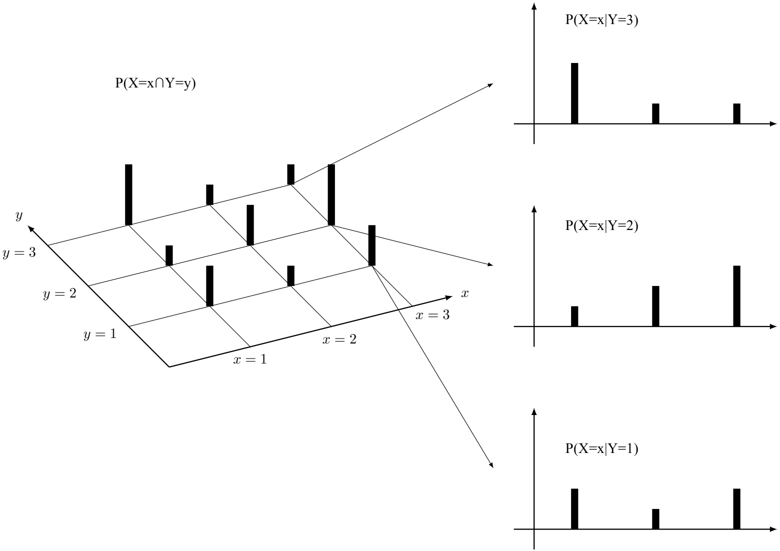
Let $\mathrm{P}(Y = y) > 0$. Then, the **conditional distribution function** $F_{X|Y}(\cdot|y)$ of the variable X given $Y = y$ is defined as

$$F_{X|Y}(x|y) = \mathrm{P}(X \leq x|Y = y).$$

The **conditional probabilities of values** of X given (under the condition of) $Y = y$ are given, analogously, by

$$\mathrm{P}(X = x|Y = y).$$

Illustration of conditional probabilities $\mathrm{P}(X = x|Y = y)$



Conditional expectation of a discrete random variable

From the definition it follows that:

$$\mathrm{P}(X = x|Y = y) = \frac{\mathrm{P}(X = x \cap Y = y)}{\mathrm{P}(Y = y)}.$$

Definition

Let $\mathrm{P}(Y = y) > 0$. The expectation of the variable X with conditional probabilities $\mathrm{P}(X = x|Y = y)$ is called the **conditional expectation** of X given $Y = y$ and is denoted as $\mathrm{E}(X|Y = y)$.

Thus it holds that:

$$\mathrm{E}(X|Y = y) = \sum_i x_i \, \mathrm{P}(X = x_i|Y = y) = \sum_i x_i \frac{\mathrm{P}(X = x_i \cap Y = y)}{\mathrm{P}(Y = y)}.$$

Continuous conditional distribution

When observing two continuous random variables X and Y, it is not possible to use an event $\{Y = y\}$ as a condition, because $\mathrm{P}(Y = y) = 0$.

The conditional distribution can be obtained using a limit approach:
Let $f_{X,Y}$ be joint density of X, Y and it holds $f_Y(y) > 0$. Then for $\Delta y << 1$

$$\mathrm{P}(X \leq x \mid y \leq Y \leq y + \Delta y) = \frac{\mathrm{P}(X \leq x \cap y \leq Y \leq y + \Delta y)}{\mathrm{P}(y \leq Y \leq y + \Delta y)} =$$

$$= \frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta y} f_{X,Y}(u, v) \,\mathrm{d}v \,\mathrm{d}u}{\int_{y}^{y+\Delta y} f_Y(v) \,\mathrm{d}v} \approx \frac{\int_{-\infty}^{x} f_{X,Y}(u, y) \Delta y \,\mathrm{d}u}{f_Y(y) \Delta y} =$$

$$= \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} \,\mathrm{d}u.$$

After taking a limit $\Delta y \to 0$ we intuitively obtain the result as

$$\mathrm{P}(X \leq x | Y = y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} \,\mathrm{d}u.$$

Continuous conditional distribution

the previous inference lead us to the following formal definition:

Definition

The **conditional distribution function** of a variable X given (under the condition of) $Y = y$ is defined as

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u,y)}{f_Y(y)} \mathrm{d}u,$$

for all y such that $f_Y(y) > 0$. We use the notation $\mathrm{P}(X \leq x | Y = y) = F_{X|Y}(x|y)$, too.

The conditional density is defined accordingly:

Definition

The **conditional probability density** of X given (under the condition of) $Y = y$ is given as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

for all y such that $f_Y(y) > 0$.

Conditional expectation of a continuous random variable

Analogously as in the discrete case we define the **conditional expectation** for continuous random variables:

Definition

Let $f_Y(y) > 0$. The expectation of variable X with density $f_{X|Y}(x|y)$ is called the **conditional expectation** of X given $Y = y$ and is denoted as $\mathrm{E}(X|Y = y)$.

We compute the conditional expectation for a given value y as follows:

$$\mathrm{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x = g(y),$$

where g is a function which arises from the integration.

Recap

Joint distribution function of a random vector (X, Y):

$$F_{X,Y}(x, y) = \mathrm{P}(X \leq x \cap Y \leq y).$$

Discrete random variables X and Y	**Continuous random variables X and Y**
Joint probabilities of values: $\mathrm{P}(X = x \cap Y = y)$	Joint density: $f_{X,Y}(x, y)$

Marginal distribution:

$\mathrm{P}(X = x) = \sum_{\text{all } y} \mathrm{P}(X = x \cap Y = y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y$
$\mathrm{P}(Y = y) = \sum_{\text{all } x} \mathrm{P}(X = x \cap Y = y)$	$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x$

Independence of X and Y:

$\mathrm{P}(X = x \cap Y = y) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y)$	$f_{X,Y}(x, y) = f_X(x)\,f_Y(y)$

Conditional probabilities / density of X given $Y = y$:

$\mathrm{P}(X = x \mid Y = y) = \dfrac{\mathrm{P}(X = x \cap Y = y)}{\mathrm{P}(Y = y)}$	$f_{X\mid Y}(x\mid y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$

Conditional expectation of X given $Y = y$:

$\mathrm{E}(X \mid Y = y) = \sum_{x} x\,\mathrm{P}(X = x \mid Y = y)$	$\mathrm{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X\mid Y}(x\mid y)\,\mathrm{d}x$