Statistics for Data Science

MSc Data Science WiSe 2019/20

Prof. Dr. Dirk Ostwald

(3) Joint distributions

Joint distributions

- Joint distributions
- Marginal distributions
- Independent random variables
- Conditional distributions
- Multivariate distributions

Joint distributions

- **Joint distributions**
- Marginal distributions
- Independent random variables
- Conditional distributions
- Multivariate distributions

Definition (Joint distribution)

Let X and Y denote two random variables with outcome spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. The *joint distribution* (or *bivariate distribution*) of X and Y is the probability measure

$$\mathbb{P}_{X,Y}\left((X,Y) \in A\right) \text{ for all } A \subset \mathcal{X} \times \mathcal{Y} \tag{1}$$

such that $\{(X,Y) \in A\}$ is an event.

Remarks

- Technically, defining joint distributions requires the theory of *product measures*.
- We omit the ensuing theory of *product probability spaces*.
- We focus on the specification of joint distributions using PMFs and PDFs.
- We will omit the X, Y subscript of $\mathbb{P}_{X,Y}$ in the following.

Definition (Discrete joint distributions)

Let X and Y be random variables and consider the ordered pair (X, Y). If there are only finitely or at most countably many different possible values (x, y) for the pair (X, Y), then X and Y have a *discrete joint distribution*.

Remark

- X and Y have a discrete joint distribution, if both X and Y are discrete.
- It is conventional to write, e.g., $\{X = x, Y = y\}$ for $\{(X, Y) = (x, y)\}$.

Definition (Joint probability mass function)

Let X and Y be discrete random variables with outcome spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then a *joint probability mass function (PMF)* (or *bivariate PMF*) of X and Y is defined as

$$p : \mathcal{X} \times \mathcal{Y} \to [0,1], (x,y) \mapsto p(x,y) := \mathbb{P}(X = x, Y = y). \tag{2}$$

Remarks

- $\{(X = x, Y = y)\}$ means that $X = x$ AND $Y = y$.
- A joint PMF is non-negative: $p(x,y) \geq 0$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.
- A joint PMF is normalized: $\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) = 1$.

Example (Joint probability mass function)

Let $\mathcal{X} = \{1, 2, 3\}$ and $\mathcal{Y} = \{1, 2, 3, 4\}$. Then a joint PMF of X and Y is given by

$$p : \{1, 2, 3\} \times \{1, 2, 3, 4\} \to [0, 1], (x, y) \mapsto p(x, y) \tag{3}$$

with $p(x, y)$ specified according to the table below:

$p(x, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$
$x = 1$	0.1	0	0.2	0.1
$x = 2$	0.1	0.2	0	0
$x = 3$	0.0	0.1	0.1	0.1

Note that $\sum_{x=1}^{3} \sum_{y=1}^{4} p(x, y) = 1$.

Definition (Continuous joint distribution)

Two random variables X and Y have a *continuous joint distribution*, if there exists a non-negative function $p : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ such that for every subset $A \subseteq \mathbb{R}^2$

$$\mathbb{P}((X,Y) \in A) = \iint_A p(x,y)\,dx\,dy. \tag{4}$$

Remark

- The function $p : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ is central.
- The integral is formed over the set A.
- If $A = \{(x,y)\}$ for $(x,y) \in \mathbb{R}^2$, then

$$\mathbb{P}((X,Y) \in \{(x,y)\}) = \int_x^x \int_y^y p(x,y)\,dx\,dy = 0- \tag{5}$$

- Hence $\mathbb{P}(X = x, Y = y) = 0$ for all $(x,y) \in \mathbb{R}^2$.

Definition (Joint probability density function)

Let X and Y be continuous random variables with joint distribution $\mathbb{P}$. Then a function

$$p : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}, (x, y) \mapsto p(x, y) \tag{6}$$

is called a *joint probability density function (PDF)* for (X, Y), if

- $p(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$,
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y)\, dx\, dy = 1$, and
- for any set $A \subseteq \mathbb{R}^2$ it holds that $\mathbb{P}((X, Y) \in A) = \iint_A p(x, y)\, dx\, dy$.

Example (The bivariate Gaussian distribution)

Let (X, Y) have a continuous joint distribution with joint probability density function

$$p : \mathbb{R}^2 \to \mathbb{R}_{>0}, (x, y) \mapsto$$

$$p(x, y) := \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right)$$

Then (X, Y) are said to be distributed according to a *bivariate* (or *two-dimensional*) Gaussian distribution with expectation parameters μ_1, μ_2, variance parameters $\sigma_X^2 > 0, \sigma_Y^2 > 0$ and correlation parameter $-1 < \rho < 1$.

Remarks

- The parameters μ_1, μ_2 specify the location of highest probability density in $\mathbb{R}^2$.
- The parameters σ_X^2 and σ_Y^2 specify the width of the distribution w.r.t. X and Y.
- The parameter ρ specifies the covariation of X and Y.
- The term $\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$ is the normalization constant for the exponential term.

Definition (Joint cumulative distribution function)

The *joint cumulative distribution function* of two random variables X and Y is defined as the function

$$P : \mathcal{X} \to \mathcal{Y} \to [0,1], (x,y) \mapsto P(X,Y) := \mathbb{P}\left(X \leq x, Y \leq y\right) \tag{7}$$

Remark

- For each fixed $x \in \mathcal{X}$, P is monotonically increasing in y
- For each fixed $y \in \mathcal{X}$, P is monotonically increasing in x

Joint distributions

- Joint distributions
- **Marginal distributions**
- Independent random variables
- Conditional distributions
- Multivariate distributions

Definition (Marginal probability mass and probability density functions)

If the discrete random variables X and Y have a joint distribution with joint PMF $p_{X,Y}$, then the marginal mass functions of X and Y are defined by

$$p_X(x) := \sum_{y \in \mathcal{Y}} p_{X,Y}(x, y) \text{ and } p_Y(y) := \sum_{x \in \mathcal{X}} p_{X,Y}(x, y), \tag{8}$$

respectively. Similarly, if the continuous random variables X and Y have a joint distribution with joint PDF $p_{X,Y}$, then the marginal density functions of X and Y are defined as

$$p_X(x) := \int_{-\infty}^{\infty} p_{X,Y}(x, y)\, dy \text{ and } p_Y(y) := \int_{-\infty}^{\infty} p_{X,Y}(x, y)\, dx, \tag{9}$$

respectively.

Remarks

- We omit the formal definition of a marginal distribution.
- The X, Y subscripts at $p_{X,Y}, p_X, p_Y$ are commonly omitted.

Example (Marginal probability mass function)

Consider the earlier example of a joint PMF. The marginal PMFs of this joint distribution evaluate as documented below.

$p_{X,Y}(x, y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$p_X(x)$
$x = 1$	0.1	0.0	0.2	0.1	0.4
$x = 2$	0.1	0.2	0.0	0.0	0.3
$x = 3$	0.0	0.1	0.1	0.1	0.3
$p_Y(y)$	0.2	0.3	0.3	0.2	

Note that $\sum_{x=1}^{3} p_X(x) = 1$ and $\sum_{y=1}^{3} p_Y(y) = 1$.

Example (Marginal probability density function)

Let (X, Y) be distributed according to a bivariate Gaussian distribution with expectation parameters μ_X, μ_Y, variance parameters σ_X^2, σ_Y^2 and correlation parameter ρ. Then X and Y have the marginal probability density functions

$$p_X : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto p_X(x) := \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\right)$$
$$p_Y : \mathbb{R} \to \mathbb{R}_{>0}, y \mapsto p_Y(y) := \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{1}{2\sigma_Y^2}(x - \mu_Y)^2\right)$$

The marginal distributions of bivariate Gaussian distributions are thus univariate Gaussian distributions.

Joint distributions

- Joint distributions
- Marginal distributions
- **Independent random variables**
- Conditional distributions
- Multivariate distributions

Definition (Independent random variables)

Two random variables X and Y with outcome spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, are independent, if for every $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \tag{10}$$

Remark

- The definition implies that $\{X \in A\}$ and $\{Y \in B\}$ are independent events.
- Event independence implies that $\mathbb{P}(\{X \in A\}|\{Y \in B\}) = \mathbb{P}(\{X \in A\})$.
- This means that knowing $\{Y \in B\}$ does not change the probability of $\{X \in A\}$.

Theorem (Independent discrete and continuous random variables)

Let X and Y be discrete random variables with joint PMF $p_{X,Y}$ and marginal PMFs p_X and p_Y, respectively. Then X and Y are independent, if and only if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \text{ for all } (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{11}$$

Similarly, let X and Y be continuous random variables with joint PDF $p_{X,Y}$ and marginal PDFs p_X and p_Y, respectively. Then X and Y are independent, if and only if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \text{ for all } (x,y) \in \mathbb{R}^2. \tag{12}$$

Proof

DeGroot and Shervish, 2012 Corollary 3.5.1.

Remarks

- The property $p(x,y) = p(x)p(y)$ is referred to as “factorization”.
- Independence is equivalent to the factorization of joint PMFs/PDFs.

Example (Independent discrete random variables)

Consider the earlier example of a joint PMF and its associated marginal PMFs.

$p_{X,Y}(x,y)$	$y=1$	$y=2$	$y=3$	$y=4$	$p_X(x)$
$x=1$	0.1	0.0	0.2	0.1	0.4
$x=2$	0.1	0.2	0.0	0.0	0.3
$x=3$	0.0	0.1	0.1	0.1	0.3
$p_Y(y)$	0.2	0.3	0.3	0.2	

As

$$p_{X,Y}(1,1) = 0.1 \neq 0.08 = p_X(1)p_Y(1) \tag{13}$$

the random variables X, Y are not independent.

For the same marginal PMFs a, joint PMFs of independent random variables X, Y is

$p_{X,Y}(x,y)$	$y=1$	$y=2$	$y=3$	$y=4$	$p_X(x)$
$x=1$	0.08	0.12	0.12	0.08	0.4
$x=2$	0.06	0.09	0.09	0.06	0.3
$x=3$	0.06	0.09	0.09	0.06	0.3
$p_Y(y)$	0.2	0.3	0.3	0.2	

Example (Independent continuous random variables)

Let (X, Y) be distributed according to a bivariate Gaussian distribution with arbitrary expectation parameters μ_X, μ_Y, arbitrary variance parameters σ_X^2, σ_Y^2, and correlation parameter $\rho := 0$. Then X and Y are independent random variables.

Proof

Substitution in the joint PDF of X and Y yields

$$
\begin{aligned}
p_{X,Y}(x, y) \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-0^2}} \exp\left(-\frac{1}{2(1-0^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\cdot 0\cdot(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right) \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{1}{2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{1}{2\sigma_X^2}(x-\mu_X) - \frac{1}{2\sigma_Y^2}(y-\mu_Y)^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{1}{2\sigma_X^2}(x-\mu_X)\right) \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{1}{2\sigma_Y^2}(y-\mu_Y)\right) \\
&= p_X(x)p_Y(y).
\end{aligned}
$$

□

Joint distributions

- Joint distributions
- Marginal distributions
- Independent random variables
- **Conditional distributions**
- Multivariate distributions

Definition (Conditional PMF and conditional discrete distribution)

For a joint distribution of two discrete random variables with joint PMF $p_{X,Y}$, the conditional probability mass function of X given $Y = y$ is

$$p_{X|Y}(\cdot|y) : \mathcal{X} \to [0,1],\ x \mapsto p_{X|Y}(x|y) := \frac{p_{X,Y}(x,y)}{p_Y(y)} \text{ for } p_Y(y) > 0. \tag{14}$$

The discrete distribution with PMF $p_{X|Y}(\cdot|Y = y)$ is called the the conditional distribution of X given that $Y = y$.

Example (Discrete conditional distributions)

Consider the earlier example of a joint PMF and its associated marginal PMFs.

$p_{X,Y}(x,y)$	$y=1$	$y=2$	$y=3$	$y=4$	$p_X(x)$
$x=1$	0.1	0.0	0.2	0.1	0.4
$x=2$	0.1	0.2	0.0	0.0	0.3
$x=3$	0.0	0.1	0.1	0.1	0.3
$p_Y(y)$	0.2	0.3	0.3	0.2	

Then the conditional distributions of Y given X are given by

$p_{Y\|X}(y\|x)$	$y=1$	$y=2$	$y=3$	$y=4$	
$p_{X\|Y}(y\|x=1)$	$\frac{0.1}{0.4}=\frac{1}{4}$	$\frac{0.0}{0.4}=0$	$\frac{0.2}{0.4}=\frac{1}{2}$	$\frac{0.1}{0.4}=\frac{1}{4}$	$\sum_{y=1}^{4} p_{Y\|X}(y\|x)=1$
$p_{X\|Y}(y\|x=2)$	$\frac{0.1}{0.3}=\frac{1}{3}$	$\frac{0.2}{0.3}=\frac{2}{3}$	$\frac{0.0}{0.3}=0$	$\frac{0.0}{0.3}=0$	$\sum_{y=1}^{4} p_{Y\|X}(y\|x)=1$
$p_{X\|Y}(y\|x=3)$	$\frac{0.0}{0.3}=0$	$\frac{0.1}{0.3}=\frac{1}{3}$	$\frac{0.1}{0.3}=\frac{1}{3}$	$\frac{0.1}{0.3}=\frac{1}{3}$	$\sum_{y=1}^{4} p_{Y\|X}(y\|x)=1$

Note that qualitative similarity of $p_{X,Y}(x,y)$ and $p_{Y|X}(y|x)$

Definition (Conditional PDF and conditional continuous distribution)

For a joint distribution of two continuous random variables with joint PDF $f_{X,Y}$, the conditional probability density function of X given Y is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \text{ for } p_Y(y) > 0. \tag{15}$$

Remarks

- $p_Y(y) > 0$ refers to a probability density, not a probability.
- Relating conditional PDFs to conditional probabilities is technically demanding.
- A rigorous framework is afforded by the theory of *Markov kernels*.

Joint distributions

- Joint distributions
- Marginal distributions
- Independent random variables
- Conditional distributions
- **Multivariate distributions**

Definition (Joint cumulative distribution function)

The joint cumulative distribution function of n random variables $X_1, X_2, ..., X_n$ is

$$P : \mathbb{R}^n \to [0,1], (x_1, x_2, ..., x_n) \mapsto \mathbb{P}\left(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\right) \tag{16}$$

Remarks

- It is convenient to use the vector notation $X = (X_1, X_2, ..., X_n)$.
- An X thus defined is referred to as *random vector*.
- It is also convenient to use the vector notation $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$.
- We are concerned with distributions of X with joint CDF values $P(x)$.

Definition (Joint discrete and continuous multivariate distributions)

n random variables $X_1, ..., X_n$ have a *discrete multivariate distribution*, if the random vector $(X_1, ..., X_n)$ can have only a finite number or an infinite sequence of different possible values $(x_1, ..., x_n) \in \mathbb{R}^n$. The joint PMF of $X_1, ..., X_n$ is then defined as

$$p : \mathbb{R}^n \to [0, 1], (x_1, ..., x_n) \mapsto p(x_1, ..., x_n) := \mathbb{P}(X_1 = x_1, ..., X_n = x_n), \tag{17}$$

or in vector notation as

$$p : \mathbb{R}^n \to [0, 1], x \mapsto p(x) := \mathbb{P}(X = x). \tag{18}$$

Similarly, n continuous random variables $X_1, ..., X_n$ have a *continuous multivariate distribution*, if there is a non-negative function $p : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, such that for every $A \subseteq \mathbb{R}^n$

$$\mathbb{P}((X_1, ..., X_n) \in A) = \int \underset{A}{\cdots} \int p(x_1, ..., x_n) \, dx_1 ... dx_n, \tag{19}$$

or in vector notation

$$\mathbb{P}(X \in A) = \int \cdots \int p(x) \, dx. \tag{20}$$

p is called the joint PDF of $X_1, ..., X_n$.

Example (Multivariate Gaussian distribution)

Let X be an n-dimensional random vector with outcome set $\mathbb{R}^n$ and (joint) PDF

$$p : \mathbb{R}^n \to \mathbb{R},\ x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right). \tag{21}$$

Then X is said to be distributed according to a *multivariate (or n-dimensional) Gaussian distribution* with *expectation parameter* $\mu \in \mathbb{R}^n$ and positive-definite *covariance matrix parameter* $\Sigma \in \mathbb{R}^{n \times n}$, for which we write $X \sim N(\mu, \Sigma)$. We abbreviate the PDF of a multivariate Gaussian distribution by

$$N(x; \mu, \Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right). \tag{22}$$

Remarks

- The parameter $\mu \in \mathbb{R}^n$ specifies the location of highest probability density in $\mathbb{R}^n$.
- The diagonal elements of Σ specify the width of the distribution w.r.t. $X_1, ..., X_n$.
- The i, jth off-diagonal element of Σ specifies the covariation of X_i and X_j.
- The term $(2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}$ is the normalization constant for the exponential term.

Marginal distributions

The marginal distribution of each $X_i, i = 1, ..., n$ can be derived from the joint distribution of $X_1, ..., X_n$ by summation/integration over the remaining $n - 1$ variables for PMFs/PDFs, respectively.

For example, if an $X = (X_1, ..., X_n)$ has PMF p_X, then the marginal PMF of X_i is evaluated as

$$p_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p_X(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n) \tag{23}$$

Similarly, if an $X = (X_1, ..., X_n)$ has PDF p_X, then the marginal PDF of X_i is evaluated as

$$p_{X_i}(x_i) = \int \cdots \iint \cdots \int p_X(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)\, dx_1 ... dx_i\, dx_{i+1} ..., dx_n \tag{24}$$

Marginal distributions

More generally, the marginal joint distribution of any k of the $X_i, i = 1, ..., n$ can be derived from the joint distribution of $X_1, ..., X_n$ by summation/integration over the remaining $n - k$ variables for PMFs/PDFs, respectively.

For example, if $X = (X_1, X_2, X_3, X_4)$ has PMF p_X, then the marginal PMF of X_2, X_4 is evaluated as

$$p_{X_2,X_3}(x_2, x_3) = \sum_{x_1} \sum_{x_4} p_X(x_1, x_2, x_3, x_4). \tag{25}$$

Similarly, if an $X = (X_1, X_2, X_3, X_4)$ has PDF p_X, then the marginal PDF of X_1, X_2 is evaluated as

$$p_{X_1,X_2}(x_1, x_2) = \iint p_X(x_1, x_2, x_3, x_4)\, dx_3\, dx_4. \tag{26}$$

Example (Marginal Gaussian distributions)

Let $X = (X_1, ..., X_n)$ be distributed according to an n-dimensional Gaussian distribution, the expectation and covariance matrix parameters of which partition for $n = k+m$ according to

$$\mu = \begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix} \in \mathbb{R}^n, \tag{27}$$

where $\mu_y \in \mathbb{R}^k$ and $\mu_z \in \mathbb{R}^m$ and

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix} \in \mathbb{R}^{n \times n}, \tag{28}$$

where $\Sigma_{yy} \in \mathbb{R}^{k \times k}$, $\Sigma_{yz} \in \mathbb{R}^{k \times m}$, $\Sigma_{zy} \in \mathbb{R}^{m \times k}$, and $\Sigma_{zz} \in \mathbb{R}^{m \times m}$.

Then $Y := (X_1, ..., X_k)$ and $Z := (X_{k+1}, ..., X_n)$ are distributed as

$$Y \sim N(\mu_y, \Sigma_{yy}) \text{ and } Z \sim N(\mu_z, \Sigma_{xx}). \tag{29}$$

Definition (Multivariate conditional PMFs/PDFs)

Suppose that the random vector $X = (X_1, ..., X_n)$ is divided into two subvectors Y and Z, where Y is a k-dimensional random vector comprising k of the n random variables in X and Z is an $(n-k)$-dimensional random vector comprising the remaining $(n-k)$ random variables in X. Suppose also that the n-dimensional joint PMF/PDF of (Y, Z) is $p_{Y,Z}$ and that the marginal $(n-k)$-dimensional PMF/PDF of Z is p_Z. Then for every $z \in \mathbb{R}^{n-k}$ such that $p_Z(z) > 0$, the conditional *k-dimensional PMF/PDF* of Y given $Z = z$ is defined as

$$p_{Y|Z} : \mathbb{R}^k \to \mathbb{R}_{\geq 0}, y \mapsto p_{Y|Z}(y|z) := \frac{p_{Y,Z}(y,z)}{p_Z(z)}. \tag{30}$$

Remarks

- The definition of $p_{Y|Z=z}(y)$ may be rewritten as $p_{Y,Z}(y,z) = p_{Y|Z}(y|z)p_Z(z)$.
- This allows for constructing joint distributions from conditional distributions and marginal distributions.

Theorem (Multivariate Law of Total Probability and Bayes Theorem)

With the conditions and notations used in the definition of multivariate conditional PMFs/PDFs, and a continuous X, the marginal PDF of Y is given by

$$p_Y : \mathbb{R}^k \to \mathbb{R}_{\geq 0}, y \mapsto \int_{\mathbb{R}^k} p_{Y|Z}(y|z) p_Z(z)\, dz \tag{31}$$

and the conditional PDF of Z given $Y = y$ is given by

$$p_{Z|Y} : \mathbb{R}^{n-k} \to \mathbb{R}_{\geq 0}, z \mapsto p_{Z|Y}(z|y) = \frac{p_Y(y|z) p_Z(z)}{p_Y(y)}. \tag{32}$$

For discrete X, the multiple integral must be replaced by the equivalent multiple summation.

Example (Joint Gaussian distributions)

Given an m-dimensional random vector X distributed according to a Gaussian distribution with PDF

$$p_X : \mathbb{R}^m \to \mathbb{R}_{>0},\ x \mapsto p_X(x) := N(x; \mu_x, \Sigma_{xx}) \text{ for } \mu_x \in \mathbb{R}^m, \Sigma_{xx} \in \mathbb{R}^{m \times m} \tag{33}$$

a matrix $A \in \mathbb{R}^{n \times m}$, a vector $b \in \mathbb{R}^n$, and a n-dimensional random vector Y conditionally distributed according to a Gaussian distribution with conditional PDF

$$p_{y|X}(\cdot|x) : \mathbb{R}^n \to \mathbb{R}_{>0},\ y \mapsto p_{Y|X}(y|x) := N(y; AX + b, \Sigma_{yy}) \text{ for } \Sigma_{yy} \in \mathbb{R}^{n \times n} \tag{34}$$

the $m + n$-dimensional random vector (X, Y) is distributed according to a Gaussian distribution with joint PDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0},\ (x, y) \mapsto p_{X,Y}(x, y) = N\left((x, y); \mu_{x,y}, \Sigma_{x,y}\right), \tag{35}$$

where $\mu_{x,y} \in \mathbb{R}^{n+m}$ and $\Sigma_{x,y} \in \mathbb{R}^{m+n \times m+n}$, and in particular

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ A\mu_x + b \end{pmatrix} \text{ and } \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}A^T \\ A\Sigma_{xx} & \Sigma_{yy} + A\Sigma_{xx}A^T \end{pmatrix}. \tag{36}$$

Note that the parameters of the Gaussian joint distribution can be computed from the parameters of the Gaussian marginal and Gaussian conditional distributions.

Example (Conditional Gaussian distributions)

Given an $m+n$-dimensional random vector (X, Y) distributed according to a Gaussian distribution with PDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0}, (x, y) \mapsto p_{X,Y}(x, y) := N\left((x, y); \mu_{x,y}, \Sigma_{x,y}\right), \tag{37}$$

where

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{38}$$

for $x, \mu_x \in \mathbb{R}^m, y, \mu_y \in \mathbb{R}^n$ and $\Sigma_{xx} \in \mathbb{R}^{m\times m}, \Sigma_{xy} \in \mathbb{R}^{m\times n}, \Sigma_{yy} \in \mathbb{R}^{n\times n}$, the distribution of X given Y has an m-dimensional conditional PDF

$$p_{X|Y}(\cdot|y) : \mathbb{R}^m \to \mathbb{R}_{>0}, x \mapsto p_{X|Y}(x|y) := N(x; \mu_{x|y}, \Sigma_{x|y}), \tag{39}$$

where

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \mu_y) \in \mathbb{R}^m \tag{40}$$

and

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \in \mathbb{R}^{m\times m}. \tag{41}$$

Note that with the previous statements, the parameters of conditional and marginal Gaussian PDFs can be computed from the parameters of complementary marginal and conditional Gaussian PDFs.

Definition (n independent random variables)

The random variables $X_1, ..., X_n$ are independent, if for every $A_1, ..., A_n$

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i). \tag{42}$$

If the random variables have a joint PMF/PDF p_X, then independence holds if

$$p_X(x_1, ..., x_n) = \prod_{i=1}^{n} p_{X_i}(x_i). \tag{43}$$

Example (n independent Gaussian random variables)

For $i = 1, ..., n$ let $N(x_i; \mu_i, \sigma_i^2)$ denote the PDF of n independent univariate Gaussian random variables $X_1, ..., X_n$ with $\mu_1, ..., \mu_n \in \mathbb{R}$ and $\sigma_1^2, ..., \sigma_n^2 > 0$. Further, let $N(x; \mu, \Lambda)$ denote the probability density function of an n-variate random vector X with $\mu := (\mu_1, ..., \mu_n)$ and diagonal covariance matrix $\Lambda \in \mathbb{R}^{n \times n}$ with diagonal elements $\sigma_1^2, ..., \sigma_n^2 > 0$. Then it holds that

$$p_X(x) = p_{X_1,...,X_n}(x_1, ..., s_n) \tag{44}$$

and in particular that

$$N(x; \mu, \Lambda) = \prod_{i=1}^{n} N(x_i; \mu_i, \sigma^2). \tag{45}$$

Example (n independent Gaussian random variables cont.)

Proof

We have

$$\begin{aligned} N(x;\mu,\Lambda) &= (2\pi)^{-\frac{n}{2}}\,|\Lambda|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Lambda^{-1}(x-\mu)\right) \\ &= \left(\prod_{i=1}^{n} 2\pi^{-\frac{1}{2}}\right)\left(\prod_{i=1}^{n}\sigma_i^2\right)^{-\frac{1}{2}}\exp\left(\left(\sum_{i=1}^{n}-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right)\right) \\ &= \prod_{i=1}^{n}\left(2\pi\sigma_i^2\right)^{-\frac{1}{2}}\prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right) \\ &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right) \\ &= \prod_{i=1}^{n} N\left(x_i;\mu_i,\sigma_i^2\right). \end{aligned} \tag{46}$$

□

Definition (Independent and identically distributed random variables)

The random variables $X_1, ..., X_n$ are called independent and identically distributed (i.i.d.), if

(1) $X_1, ..., X_n$ are independent random variables, and

(2) each X_i has the same marginal distribution.

Remarks

- If the marginal distributions have CDF P, we write $X_1, ..., X_n \sim P$.
- If P has an associated PMF/PDF p, we write $X_1, ..., X_n \sim p$.
- $X_1, ..., X_n$ is also called a *random sample* of size n from p.

Example (n i.i.d. Gaussian random variables)

For $i = 1, ..., n$ let $N(x_i; \mu, \sigma^2)$ denote the PDF of n i.i.d. univariate Gaussian random variables $X_1, ..., X_n$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Further, let $N(x; \mu 1_n, \sigma^2 I_n)$ denote the PDF of an n-variate random vector X, where $1_n \in \mathbb{R}^n$ denotes a vector of all ones and $I_n \in \mathbb{R}^{n \times n}$ denotes the n-dimensional identity matrixl. Then it holds that

$$p_X(x) = p_{X_1,...,X_n}(x_1, ..., s_n) \tag{47}$$

and in particular that

$$N\left(x; \mu 1_n, \sigma^2 I_n\right) = \prod_{i=1}^{n} N(x_i; \mu, \sigma^2). \tag{48}$$

<u>Proof</u>

The statement follows directly from the proof of the multivariate/univariate identity of n independent Gaussian random variables.

□