FAMILIES OF RANDOM VARIABLES AND RANDOM VECTORS

VÍTĚZSLAV VESELÝ

Department of Statistics and Operations Research
University of Malta [1]
Email: vvese01@um.edu.mt

CONTENTS

Date: 27-th January 2002.

[1] Visiting lecturer, 2-nd semester from 4-th February to 12-th July 2002

1.1. Some general notation and abbreviations.

$s := v$ or $v =: s$ ... denoting expression v by symbol s.
iff stands for *if and only if*.

Sets and mappings:

- $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$ ... natural numbers, integers, real and complex numbers, respectively.
- $\mathbb{Z}_N := \{0, 1, \dots, N-1\}$ ... residuals modulo $N \in \mathbb{N}$.
- $\mathbb{R}^+$ ... the set of all non-negative real numbers.
- $\exp X$ ... class of all subsets of the set X.
- $\operatorname{card} M$ ... cardinality of a set M.
- $(\cdot)^+ : \mathbb{R} \to \mathbb{R}^+$ ... mapping defined as $(x)^+ = \max(0, x)$.
- (a, b), $[a, b]$, $(a, b]$, $[a, b)$ ... intervals on real line.
- $\mathfrak{I}(a, b) = \{x \mid \min(a, b) < x < \max(a, b)\}$ $\mathfrak{I}[a, b] = \{x \mid \min(a, b) \leq x \leq \max(a, b)\}$.
- $f(A) := \{y \in Y \mid y = f(x), x \in A \subseteq X\}$ ... range (image) of set A under mapping $f : X \to Y$.
- $f^{-1}(B) := \{x \in X \mid f(x) \in B\} \subseteq X$ ... inverse image of set $B \subseteq Y$ under mapping $f : X \to Y$.
- I_A ... indicator function of set $A \subseteq X$: $I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases}$.
- $A_n \uparrow$ ... increasing or non-decreasing sequence of numbers or sets.
- $A_n \downarrow$ ... decreasing or non-increasing sequence of numbers or sets.
- $\sum_{i=1}^n A_i := \bigcup_{i=1}^n A_i$ ... union of a family of sets which are pairwise disjoint.
- $A^c := X - A$ ... complement of set $A \subseteq X$ in X where X is a priori known from the context.
- $\underline{A} := \liminf_{n\to\infty} A_n := \bigcup_{n=1}^\infty \bigcap_{j=n}^\infty A_j$ ... inferior limit of a sequence of sets.

- $\overline{A} := \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j$ ...superior limit of a sequence of sets.
- $A = \lim_{n\to\infty} A_n$ iff $\underline{A} = \overline{A}$, clearly
  $A_n \uparrow A$ implies $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$ and
  $A_n \downarrow A$ implies $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Vectors and matrices:

- $\boldsymbol{x} := [x_1, \dots, x_n]^T$ ...vector of numbers (by default column vector if not stated otherwise).
- $\boldsymbol{x} + h := [x_1 + h, \dots, x_n + h]^T,\ h \in \mathbb{C}$.
- $\boldsymbol{x_t} := [x_{t_1}, \dots, x_{t_k}]^T \in \mathbb{C}^k$ where $\boldsymbol{t} = [t_1, \dots, t_k]^T \in \mathbb{N}^k$, $t_i \in \{1, \dots, n\}$ for $i = 1, \dots, k$.
- $\boldsymbol{x}(i) := [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]^T$ for any $1 \le i \le n$.
- $f(\boldsymbol{x}) := f(x_1, \dots, x_n),\ d\boldsymbol{x} := dx_1 \dots dx_n$.
- $\mathbf{0}, \mathbf{0}_{n\times 1}$ ...vector of n zero entries.
- $\boldsymbol{A}$, $\boldsymbol{A}_{m\times n} := [a_{ij}] = [A(i,j)]$ ...matrix of size $m \times n$.
- $\mathcal{R}(\boldsymbol{A}) := \{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}$ ...range space of matrix operator $\boldsymbol{A}$.
- $\mathcal{N}(A) := \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = 0\}$ ...null space (kernel) of matrix operator $\boldsymbol{A}$.
- $\boldsymbol{A}^T := [a_{ji}]$ ...matrix transpose.
- $\boldsymbol{A}^* := [\bar{a}_{ji}]$ ...matrix adjoint.
- $\boldsymbol{I}$, $\boldsymbol{I}_n := \boldsymbol{I}_{n\times n} = [\delta_{ij}]$ ...identity matrix of order n.
- $\det \boldsymbol{A}$ ...determinant of a square matrix $\boldsymbol{A}$.
- $\mathbf{0}$, $\mathbf{0}_{m\times n}$ ...zero matrix of size $m \times n$.
- $\operatorname{diag}(\boldsymbol{x}) := \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & x_n \end{bmatrix}$ ...diagonal matrix.
- $\boldsymbol{A}(i,:) := [a_{i1}, \dots, a_{in}]$ ...i-th row of matrix $\boldsymbol{A}$ using MATLAB style.
- $\boldsymbol{A}(:,j) := [a_{1j}, \dots, a_{mj}]^T$ ...j-th column of matrix $\boldsymbol{A}$ using MATLAB style.

- $\boldsymbol{A} := [r_1; \dots; r_m] = [s_1, \dots, s_n]$ ... forming matrix $\boldsymbol{A}$ row-by-row or columnwise using MATLAB style.
- $\boldsymbol{A} > 0$ (or $\boldsymbol{A} \geq 0$) ... positively (semi)definite (non-negatively definite) matrix.
- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{n} x_i \bar{y}_i = \boldsymbol{y}^* \boldsymbol{x}$ ... scalar (inner) product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
- $\|\boldsymbol{x}\| := \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ ... Euclidean norm of vector $\boldsymbol{x}$.

1.2. **Measurable spaces and measurable mappings.**

Measurable spaces:

Definition 1.1. The pair $(\Omega, \mathcal{A})$, $\Omega \neq \emptyset$ is called **measurable space** if $\mathcal{A} \subseteq \exp \Omega$ is a **σ-field (σ-algebra)** of so-called **measurable sets on** Ω satisfying:

(M1) $\mathcal{A} \neq \emptyset$
(M2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
(M3) $A_i \in \mathcal{A}\ \forall\, i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
(M4) $A_i \in \mathcal{A}\ \forall\, i \in \mathbb{N} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$
(M5) $\emptyset \in \mathcal{A}$
(M6) $\Omega \in \mathcal{A}$
(M7) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$
(M8) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cap A_2 \in \mathcal{A}$
(M9) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 - A_2 \in \mathcal{A}$
(M10) $A_i \in \mathcal{A}\ \forall\, i \in \mathbb{N} \Rightarrow \limsup_{n\to\infty} A_n \in \mathcal{A}$
(M11) $A_i \in \mathcal{A}\ \forall\, i \in \mathbb{N} \Rightarrow \liminf_{n\to\infty} A_n \in \mathcal{A}$
(M12) $A_i \in \mathcal{A}\ \forall\, i \in \mathbb{N} \Rightarrow \lim_{n\to\infty} A_n \in \mathcal{A}$ if such limit exists.

Properties (M1)-(M3) are axioms (typeset boldfaced), the remaining ones are their easy consequences in that order. Observe that this choice is not unique as there can be found other subsets of properties which may play the role of axioms.

Theorem 1.2. *Let $\mathcal{A}_i \subseteq \exp \Omega$ be a family of σ-fields $i \in I \neq \emptyset$, then $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-field as well.*

Corollary 1.3. *Let $C \subseteq \exp \Omega$ be arbitrary class of subsets of Ω then $\sigma(C) := \bigcap\{\mathcal{C} \mid C \subseteq \mathcal{C} \subseteq \exp \Omega,\ \mathcal{C}$ a σ-field $\}$ is the unique minimal σ-field on Ω containing C. We say also that $\sigma(C)$ is a σ-field on Ω **generated by** C.*

Definition 1.4. Let $(\Omega, \mathcal{A})$ be a measurable space. A mapping $\mu : \mathcal{A} \to [0, \infty]$ is called **measure on** $(\Omega, \mathcal{A})$ if

$$\mu(\emptyset) = 0 \text{ and } \mu\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i)$$

holds for any sequence $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, of pairwise disjoint measurable sets. In such a case we write $(\Omega, \mathcal{A}, \mu)$ instead of $(\Omega, \mathcal{A})$.

Definition 1.5. If P is a measure on $(\Omega, \mathcal{A})$ with additional property $P(\Omega) = 1$ then the triple $(\Omega, \mathcal{A}, P)$ is called **probability space**, P **probability measure on it** and Ω **sample space** of **sample points** $\omega \in \Omega$ **(outcomes of random experiments)**. Measurable sets $A \in \mathcal{A}$ are called **events**. To some events are given specific names:

$\emptyset$	...	**impossible event**,
$\{\omega\}$	...	**simple event**,
$A \in \mathcal{A}$, card $A > 1$	...	**composite event** and
Ω	...	**sure (certain) event**.

Theorem 1.6 (Additive Theorem)**.** *For any finite number of events A_j, we have*

$$P\Big(\bigcup_{j=1}^{n} A_j\Big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le j_1 < \cdots < j_k \le n} P(A_{j_1} \cap \cdots \cap A_{j_k})$$

.

Theorem 1.7 (Summary of measure properties)**.** *Let μ and P be measures introduced in 1.4 and 1.5. They satisfy properties summarized in the following table where $(\bullet)$ denotes axiom, $(\circ)$ consequence of axioms and $(-)$ property not valid in general.*

No.	*Description of the property*	μ	$\mu = P$
(P1)	$\mu(A) \geq 0$ *for any* $A \in \mathcal{A}$	•	•
(P2)	$\mu(\Omega) = 1$	–	•
(P3)	$\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$	•	•
(P4)	$\mu(A_1 + A_2) = \mu(A_1) + \mu(A_2)$	◦	◦
(P5)	$A_1 \subseteq A_2 \Rightarrow \mu(A_2) = \mu(A_1) + \mu(A_2 - A_1)$	◦	◦
(P6)	$A_1 \subseteq A_2 \Rightarrow \mu(A_1) \leq \mu(A_2)$	◦	◦
(P7)	$\mu(A) \leq 1$	–	◦
(P8)	$A_1 \subseteq A_2 \Rightarrow \mu(A_2 - A_1) = \mu(A_2) - \mu(A_1)$	–	◦
(P9)	$\mu(A^c) = 1 - \mu(A)$	–	◦
(P10)	$\mu(\emptyset) = 0$	•	◦
(P11)	$\mu(A_1 \cup A_2) \leq \mu(A_1) + \mu(A_2)$	◦	◦
(P12)	$\mu(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$	◦	◦
(P13)	$A_n \uparrow A$ *or* $A_n \downarrow A$, $\mu(A) < \infty \Rightarrow$ $\mu(A) = \lim_{n\to\infty} \mu(A_n)$	◦	◦
(P14)	*Additive theorem 1.6*	–	◦

Proof. The properties are listed in the table in the order allowing us to prove each property as a logical consequence of one or more preceding ones. The derivation of (P13) and (P14) is a little laborious. In case of $\mu = P$ the details can be found in [Rou97, p.17–20]. Let us also note that (P13) might be stated in a more general form, namely assuming $A = \lim_{n\to\infty} A_n$ provided that such limit exists and $\mu(A) < \infty$. The latter assumption may be omitted in case of $\mu = P$ in view of (P7). □

Lemma 1.8. *Let $(\Omega, \mathcal{A})$ be a measurable space and μ a measure on it. Then for any $A \in \mathcal{A}$, $\mu(A) > 0$, is $\mathcal{A}_A := \{C \,|\, C = A \cap B, B \in \mathcal{A}\}$ a σ-field on A yielding measurable space $(A, \mathcal{A}_A)$ on which* ***induced measure*** *μ_A may be introduced by $\mu_A(C) := \frac{\mu(C)}{\mu(A)}$, $C \in \mathcal{A}_A$. If $\mu = P$ is a probability measure then $P_A := \mu_A$ is a probability measure as well.*

Proof. It is straightforward to verify (M1)–(M3) for $\mathcal{A}_A$ and (P1), (P3), (P10) for μ_A or (P1)–(P3) for P_A. □

Definition 1.9. If $(\Omega, \mathcal{A}, P)$ is a probability space and $A \in \mathcal{A}$, $P(A) > 0$, arbitrary event then $(A, \mathcal{A}_A, P_A)$ is in view of 1.8 a probability space as well. We call it **conditional probability space given** A and $P(B|A) := P_A(A \cap B) := \frac{P(A \cap B)}{P(A)}$ **conditional probability of** B **given** A.

Measurable mappings:

Definition 1.10. Let $(\Omega_1, \mathcal{A})$ and $(\Omega_2, \mathcal{B})$ be two measurable spaces. The mapping $X : \Omega_1 \to \Omega_2$ is called **measurable** if $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. We write $X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$.

Lemma 1.11. *If $\mathcal{B}$ is a σ-field on Ω_2 and $X : \Omega_1 \to \Omega_2$ arbitrary mapping then $X^{-1}(\mathcal{B}) := \{A \,|\, A = X^{-1}(B) \text{ for some } B \in \mathcal{B}\}$ is a σ-field on Ω_1.*

Proof. Inverse image X^{-1} preserves (may be interchanged with) any of $\bigcup$, $\sum$, $\bigcap$ and ^{c}. Along with evident identities $X^{-1}(\Omega_2) = \Omega_1$ and $X^{-1}(\emptyset) = \emptyset$ we get immediately the properties (M1)-(M3) for $X^{-1}(\mathcal{B})$. □

A similar argumentation gives a converse of the above lemma:

Lemma 1.12. *If $\mathcal{A}$ is a σ-field on Ω_1 and $X : \Omega_1 \to \Omega_2$ arbitrary mapping then $C^* := \{B \subseteq \Omega_2 \,|\, X^{-1}(B) \in \mathcal{A}\}$ is a σ-field on Ω_2.*

Corollary 1.13. *Let $\mathcal{B} = \sigma(C)$ for some subclass $C \subseteq \exp \Omega_2$. Then mapping $X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ is measurable iff $X^{-1}(B) \in \mathcal{A}$ for each $B \in C$.*

Proof. Let $X^{-1}(B) \in \mathcal{A}$ for each $B \in C$. Clearly $C \subseteq C^*$ in view of 1.12. As C^* is σ-field on Ω_2 and $\mathcal{B} = \sigma(C)$ is a minimal one containing C, we have $\mathcal{B} \subseteq C^*$ and consequently $X^{-1}(\mathcal{B}) \subseteq X^{-1}(C^*) \subseteq \mathcal{A}$ which proves inverse implication. The converse is immediate by the definition of measurability. □

Theorem 1.14. *Let $X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ be a measurable mapping, and μ a measure on $(\Omega_1, \mathcal{A})$. Then $\mu_X(B) := \mu(X^{-1}(B))$,*

$B \in \mathcal{B}$, defines a measure on $(\Omega_2, \mathcal{B})$ which is called ***measure induced by*** *X. If $\mu = P$ is a probability measure then $P_X := \mu_X$ is a probability measure as well.*

Proof. We have to verify (P1), (P3) and (P10) for μ_X, and in addition (P2) for P_X. (P1), (P2) and (P10) follow immediately using the same properties for μ and the identities $X^{-1}(\Omega_2) = \Omega_1$ and $X^{-1}(\emptyset) = \emptyset$. $\mu_X(\sum_{i=1}^{\infty} A_i) = \mu(X^{-1}(\sum_{i=1}^{\infty} A_i)) = \mu(\sum_{i=1}^{\infty} X^{-1}(A_i)) \overset{(P3)}{=} \sum_{i=1}^{\infty} \mu(X^{-1}(A_i)) = \sum_{i=1}^{\infty} \mu_X(A_i)$ gives (P3) for μ_X. □

2. Multivariate Probabilistic Concepts

2.1. Borel σ-fields and Borel functions.

Definition 2.1. Putting $\mathcal{I}_n := \{J_1 \times \cdots \times J_n \mid J_i \subseteq \mathbb{R}$ subinterval for $i = 1, 2, \ldots, n\}$, $n \in \mathbb{N}$, then σ-field $\mathcal{B}_n := \sigma(\mathcal{I}_n)$ generated by all rectangular regions from $\mathcal{I}_n \subseteq \mathbb{R}^n$ is called **n-dimensional Borel σ-field** and its elements **Borel sets**. In particular $\mathcal{B} := \mathcal{B}_1 = \sigma(\mathcal{I})$ stands for the **Borel σ-field over real line** $\mathbb{R}$ which is generated by all subintervals $\mathcal{I} := \mathcal{I}_1 \subseteq \mathbb{R}$.

Remark 2.2. As $(-\infty, b] = \bigcap_{m=1}^{\infty}(-\infty, b_m)$ for any $b_m \downarrow b$, $b_m > b$ and conversely $(-\infty, b) = \bigcup_{m=1}^{\infty}(-\infty, b_m]$ for any $b_m \uparrow b$, $b_m < b$. We see that any interval from $\mathcal{I}$ can be obtained either from $\{(-\infty, b]\}_{b\in\mathbb{R}}$ or from $\{(-\infty, b)\}_{b\in\mathbb{R}}$ applying unions, intersections and complementing. Indeed, $\mathcal{B}$ is closed under such operations in view of (M2)-(M4) and, consequently, it may be generated by either of those restricted families of intervals. There are a lot of other generating subfamilies, for example complements to those above mentioned, or the family of all open intervals, to mention a few. The same reasoning may apply to the n-dimensional case: subfamilies of rectangular regions from $\mathcal{I}_n$ either of type $\mathbb{R}^{i-1} \times (-\infty, b] \times \mathbb{R}^{n-i}$ or $\mathbb{R}^{i-1} \times (-\infty, b) \times \mathbb{R}^{n-i}$ are sufficient to generate $\mathcal{B}_n$ for all $i = 1, 2, \ldots, k$ and $b \in \mathbb{R}$. Later on the former family will be of interest for us. Let us therefore state the following lemma.

Lemma 2.3. $\mathcal{B}_n = \sigma(\mathcal{I}_{n,0})$ *where* $\mathcal{I}_{n,0} := \{\mathbb{R}^{i-1}\times(-\infty,b]\times\mathbb{R}^{n-i} \mid i = 1,2,\dots,n; b \in \mathbb{R}\}$. *In particular* $\mathcal{B} = \sigma(\mathcal{I}_0)$ *where* $\mathcal{I}_0 := \mathcal{I}_{1,0} = \{(-\infty,b] \mid b \in \mathbb{R}\}$.

Definition 2.4. Every measurable mapping $\varphi : (\mathbb{R}^k, \mathcal{B}_k) \to (\mathbb{R}^n, \mathcal{B}_n)$, $k, n \in \mathbb{N}$, is called **Borel function**. We write explicitly $\varphi(\cdot) = [\varphi_1(\cdot), \varphi_2(\cdot), \dots, \varphi_n(\cdot)]^T$ where $\varphi_i(\cdot)$ are the respective component mappings $\mathbb{R}^k \to \mathbb{R}$.

2.2. **Random variables, random vectors and their distribution.**

Definition 2.5. Every measurable mapping $\mathbb{X} : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathcal{B}_n)$, $n \in \mathbb{N}$, is called **n-dimensional random vector on** $(\Omega, \mathcal{A}, P)$. In particular with $n = 1$ we obtain **random variable** as a measurable mapping $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ and with $n = 2$ the **bivariate random variable** as a measurable mapping $\mathbb{X} : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^2, \mathcal{B}_2)$. We write explicitly $\mathbb{X} = [X_1, X_2, \dots, X_n]^T$ where X_i are the respective component mappings $\Omega \to \mathbb{R}$. For fixed $\omega \in \Omega$ we get **sample vector** $\boldsymbol{x} = [X_1(\omega), X_2(\omega), \dots, X_n(\omega)]^T$ as an outcome of a random experiment.
Probability measure $P_{\mathbb{X}}$ induced on $(\mathbb{R}^n, \mathcal{B}_n)$ by $\mathbb{X}$ is called **probability distribution of** $\mathbb{X}$ and will be denoted later on as $\mathcal{L}(\mathbb{X})$.

Remark 2.6. Clearly, Borel function of 2.4 may be considered as a special case of random variable or random vector provided that there has been defined a suitable probability measure on $(\mathbb{R}^k, \mathcal{B}_k)$, for example probability distribution $\mathcal{L}(\mathbb{X})$ of some k-dimensional random vector $\mathbb{X}$. It is a matter of interpreting $\mathbb{R}^k$ as a sample space, and Borel sets from $\mathcal{B}_k$ as events.

Theorem 2.7. *A mapping* $\mathbb{X} : \Omega \to \mathbb{R}^n$ *is a random vector on* $(\Omega, \mathcal{A}, P)$ *iff* $\mathbb{X}^{-1}(J) \in \mathcal{A}$ *for each* $J \in \mathcal{I}_{n,0}$. *In particular (cf. remark 2.6) a mapping* $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ *is a Borel function iff* $\varphi^{-1}(J) \in \mathcal{B}_k$ *for each* $J \in \mathcal{I}_{n,0}$.

Proof. The statement is a consequence of Corollary 1.13. ☐

Corollary 2.8. *A mapping $\mathbb{X} : \Omega \to \mathbb{R}^n$ is a random vector on $(\Omega, \mathcal{A}, P)$ iff each component mapping $X_i : \Omega \to \mathbb{R}$ is a random variable on $(\Omega, \mathcal{A}, P)$. In particular a mapping $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ is a Borel function iff each component mapping $\varphi_i : \mathbb{R}^k \to \mathbb{R}$ is a Borel function.*

Proof. For arbitrary $J = \mathbb{R}^{i-1} \times (-\infty, b] \times \mathbb{R}^{n-i} \in \mathcal{I}_{n,0}$ clearly holds $X_i^{-1}((-\infty, b]) = \mathbb{X}^{-1}(J)$ which confirms the statement in view of 2.7. ☐

Corollary 2.9. *A mapping $\mathbb{X} : \Omega \to \mathbb{R}^n$ is a random vector on $(\Omega, \mathcal{A}, P)$ iff for each $\{t_1, t_2, \ldots, t_m\} \subseteq \{1, 2, \ldots, n\}$ the marginal mapping $\mathbb{X}_{\boldsymbol{t}} := [X_{t_1}, X_{t_2}, \ldots, X_{t_m}]^T : \Omega \to \mathbb{R}^m$ is a random vector on $(\Omega, \mathcal{A}, P)$. In such a case $\mathbb{X}_{\boldsymbol{t}}$ is called **marginal random vector of** $\mathbb{X}$. In particular a mapping $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ is a Borel function iff the marginal mapping $\varphi_{\boldsymbol{t}} : \mathbb{R}^k \to \mathbb{R}^m$ is a Borel function for each $\boldsymbol{t}$.*

Corollary 2.10. *Every continuous function $\mathbb{R}^k \to \mathbb{R}^n$ is a Borel function.*

Proof. $\mathcal{B}_n$ is generated by open rectangular regions in view of remark 2.2. Inverse images of them under continuous mappings are open sets in $\mathbb{R}^k$. Every open set belongs to $\mathcal{B}_k$ because it is a union of at most countably many open rectangular regions which belong to $\mathcal{B}_k$ as well. ☐

Theorem 2.11. *If $\mathbb{X}$ is a k-dimensional random vector on $(\Omega, \mathcal{A}, P)$ and $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ then $\varphi(\mathbb{X}) : \Omega \to \mathbb{R}^n$ is an n-dimensional random vector on the same probability space.*

Proof. Let $J \in \mathcal{I}_{n,0}$ be arbitrary. Then $(\varphi\mathbb{X})^{-1}(J) = \mathbb{X}^{-1}(\varphi^{-1}(J)) \in \mathcal{A}$ when applying theorem 2.7 in succession to Borel mapping φ and random vector $\mathbb{X}$. ☐

2.3. **Conditional Probability and Stochastic Independence.**

Consider fixed probability space $(\Omega, \mathcal{A}, P)$ in this subsection and all events belonging to the σ-field $\mathcal{A}$ by default.
By definition 1.9 conditional probability of event B given event A, $P(A) > 0$, is defined as follows:

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}.} \tag{2.1}$$

Example 2.12.

(1) Casting a die with odd-numbered sides painted white and even-numbered sides painted black [Rou97, p.21].
(2) Gender of children in two-children families [Rou97, p.22].

Theorem 2.13 (Multiplicative theorem)**.** *Let A_j, $j = 1, 2, \dots, n$, $n > 1$, be events such that $P\left(\bigcap_{j=1}^{n} A_j\right) > 0$. Then*

$$P\left(\bigcap_{j=1}^{n} A_j\right) = P(A_n \mid A_1 \cap A_2 \cap \dots A_{n-1}) \times$$
$$\times P(A_{n-1} \mid A_1 \cap A_2 \cap \dots A_{n-2}) \dots P(A_2 \mid A_1) P(A_1).$$

Example 2.14. Drawing balls painted to various colours from an urn without replacement [Rou97, p.23].

Theorem 2.15 (Total probability theorem)**.** *Let $\Omega = \sum_j A_j$ be a partition of Ω with $P(A_j) > 0$ for all j. Then for any event B we have $P(B) = \sum_j P(B \mid A_j) P(A_j)$.*

Corollary 2.16 (Bayes formula)**.** *Let $\Omega = \sum_j A_j$ be a partition of Ω with $P(A_j) > 0$ for all j. Then for any event B, $P(B) > 0$, we have*

$$\boxed{P(A_j \mid B) = \frac{P(B \mid A_j) P(A_j)}{\sum_i P(B \mid A_i) P(A_i)}.}$$

Example 2.17. An example illustrating that with $P(A) > 0$ all situations out of $P(B \mid A) > P(B)$, $P(B \mid A) < P(B)$ and $P(B \mid A) = P(B)$ may occur [Rou97, p.27].
The last case $P(B \mid A) = P(B)$ is a good basis for saying that event B is *stochastically independent* of the event A because additional knowledge of the event A provides no additional information when calculating the probability of the event B.
Another example illustrating that case [Rou97, p.28].

Remark 2.18. If both $P(A) > 0$ and $P(B) > 0$ holds, and B is independent of A then also the symmetric statement A *is independent of* B is valid. Indeed, applying (2.1) symmetrically twice we obtain

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B)P(A)}{P(B)} = P(A).$$

Thus *independence* is a symmetric property quite in accordance with our expectation.

Due to (2.1) the assumption of independence leads to the equation $P(B) = \frac{P(A \cap B)}{P(A)}$ and hence to the following definition of stochastic independence which is valid also with $P(A) = 0$ and/or $P(B) = 0$. Indeed, it is because $A \cap B \subseteq A, B$ gives $0 \leq P(A \cap B) \leq P(A), P(B)$ in view of (P1) and (P6). Consequently, $P(A) = 0$ or $P(B) = 0$ implies $P(A \cap B) = 0$ confirming the statement.

Definition 2.19. The events A and B are said to be **stochastically** (or **statistically** or **in the probability sense**) **independent** if $P(A \cap B) = P(A)P(B)$.

Definition 2.20. The events A_j, $j = 1, 2, \dots, n$, $n > 1$, are said to be (**mutually** or **completely**) **independent** if the following relationships

$$P(A_{j_1} \cap \dots \cap A_{j_k}) = P(A_{j_1}) \dots P(A_{j_k})$$

hold for any $k = 2, \dots, n$ and any integers $1 \leq j_1 < \dots < j_k \leq n$. These events are said to be **pairwise independent** if the above identities hold only with $k = 2$.

Theorem 2.21 (Basic statements on independence)**.** *Given events A_j, $j = 1, 2, \ldots, n$, $n > 1$, then the following statements are true:*

(1) *If $A_1, \ldots, A_n$ are independent, so are the events $A'_1, \ldots, A'_n$ where A'_j is either A_j or A_j^c.*
(2) *Any subset of independent events $A_1, \ldots, A_n$ is independent as well.*
(3) *If $A_1, \ldots, A_n$ are independent, $1 \leq i_1 < \cdots < i_m \leq n$ and $1 \leq j_1 < \cdots < j_k \leq n$ two nonempty disjoint index subsets, then the intersections $A_{i_1} \cap \cdots \cap A_{i_m}$ and $A_{j_1} \cap \cdots \cap A_{j_k}$ are independent as well, i.e. $P(A_{i_1} \cap \cdots \cap A_{i_m} \cap A_{j_1} \cap \cdots \cap A_{j_k}) = P(A_{i_1} \cap \cdots \cap A_{i_m}) P(A_{j_1} \cap \cdots \cap A_{j_k})$.*
(4) *If $A_i \cap A_j = \emptyset$ for some $i \neq j$ and $P(A_k) > 0$ for all $k = 1, \ldots, n$, then $A_1, \ldots, A_n$ cannot be independent.*
(5) *Events A and Ω are independent for each $A \in \mathcal{A}$.*
(6) *Events A and $\emptyset$ are independent for each $A \in \mathcal{A}$.*

Notation .
Given a random vector $\mathbb{X} = [X_1, \ldots, X_n]^T$ on the probability space $(\Omega, \mathcal{A}, P)$, we introduce a simplified notation for probability $P_{\mathbb{X}}$ induced by $\mathbb{X}$ on $B \in \mathcal{B}_n$ and $B_i \in \mathcal{B}$:
• $P(\mathbb{X} \in B) := P(\{\omega \in \Omega \,|\, \mathbb{X}(\omega) \in B\}) = P(\mathbb{X}^{-1}(B)) = P_{\mathbb{X}}(B)$.
• $P(X_1 \in B_1, \ldots, X_n \in B_n) :=$
$P(\{\omega \in \Omega \,|\, X_1(\omega) \in B_1 \& \ldots \& X_n(\omega) \in B_n) = P(\cap_{i=1}^n X_i^{-1}(B_i)) = P_{\mathbb{X}}(\cap_{i=1}^n (\mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{n-i}))$.
If B_i is an interval, for example $B_i = J_i = (a_i, b_i]$ we prefer writing $a_i < X_i \leq B_i$ rather than $X_i \in J_i$ or simply $X_i \leq b_i$ if $a_i = -\infty$ or $a_i < X_i$ if $b_i = \infty$.
• $P(\boldsymbol{a} < \mathbb{X} \leq \boldsymbol{b}) := P(a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n)$ and similarly for other relations.

3.1. The Distribution Function of a Random Vector and Its Properties.

Definition 3.1. Let $\mathbb{X} = [X_1, \ldots, X_n]^T$ be a random vector. The function $F_{\mathbb{X}}(\boldsymbol{x}) := P(\mathbb{X} \leq \boldsymbol{x})$ (or more explicitly $F_{\mathbb{X}}(x_1, \ldots, x_n) := P(X_1 \leq x_1, \ldots, X_n \leq x_n)$) is called **(cumulative) distribution function of** $\mathbb{X}$ (c.d.f.), or for $n > 1$ sometimes also **joint distribution function of** $\mathbb{X}$ to emphasize dimension $n > 1$. We obtain as special cases the distribution function $F_X(x) := P(X \leq x)$ of a random variable X with $n = 1$, and $F_{\mathbb{X}}(x_1, x_2) := P(X_1 \leq x_1, X_2 \leq x_2)$ of the bivariate random variable $\mathbb{X} = [X_1, X_2]^T$ with $n = 2$. The subscript may be omitted if there is no danger of misunderstanding.

Theorem 3.2. *Let $\mathbb{X} = [X_1, \ldots, X_n]^T$ be a random vector and F its distribution function. For any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{a} \leq \boldsymbol{b}$, we define for each $i = 1, \ldots, n$ an operator $\Delta^i_{\boldsymbol{a},\boldsymbol{b}}$ by the formula*

$$\Delta^i_{\boldsymbol{a},\boldsymbol{b}} F(x_1, \ldots, x_n) := F(x_1, \ldots, x_{i-1}, b_i, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, a_i, x_{i+1}, \ldots, x_n).$$

Then

$$\boxed{P(\boldsymbol{a} < \mathbb{X} \leq \boldsymbol{b}) = \Delta^n_{\boldsymbol{a},\boldsymbol{b}} \dots \Delta^1_{\boldsymbol{a},\boldsymbol{b}} F(x_1, \dots, x_n)}$$

where the operators $\Delta^i_{\boldsymbol{a},\boldsymbol{b}}$ may be applied in any than natural order as well.

Corollary 3.3.
(1) If X is a random variable then $P(a < X \leq b) = F(b) - F(a)$ holds for all $a, b \in \mathbb{R}$, $a \leq b$.
(2) If $\mathbb{X} = [X_1, X_2]^T$ is a bivariate random variable then

$$\begin{aligned} P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F(a_1, a_2) + F(b_1, b_2) \\ - F(a_1, b_2) - F(b_1, a_2) \end{aligned}$$

holds for all $a_1, a_2, b_1, b_2 \in \mathbb{R}$, $a_1 \leq b_1$ and $a_2 \leq b_2$.

Theorem 3.4. *Distribution function of each random vector $\mathbb{X} = [X_1, \dots, X_n]^T$ has the following properties:*
(1) $\Delta^n_{\boldsymbol{a},\boldsymbol{b}} \dots \Delta^1_{\boldsymbol{a},\boldsymbol{b}} F(x_1, \dots, x_n) \geq 0$ for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{a} \leq \boldsymbol{b}$;
(2) F is continuous from the right in the sense that
$\lim_{\boldsymbol{y} \downarrow \boldsymbol{x}} F(\boldsymbol{y}) = F(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$
($\boldsymbol{y} \downarrow \boldsymbol{x}$ stands for $y_i \downarrow x_i$ for all $i = 1, \dots, n$);
(3) $\lim_{x_1 \to \infty, \dots, x_n \to \infty} F(\boldsymbol{x}) = 1$;
(4) $\lim_{x_i \to -\infty} F(\boldsymbol{x}) = 0$ for any $i \in \{1, \dots, n\}$.

Corollary 3.5. *Distribution function of each random variable X has the following properties:*
(1) F is a non-decreasing function;
(2) F is continuous from the right:
$\lim_{y \downarrow x} F(y) = F(x)$ *for all* $x \in \mathbb{R}$;
(3) $\lim_{x \to \infty} F(x) = 1$;
(4) $\lim_{x \to -\infty} F(x) = 0$.

Theorem 3.6. *The probability distribution of each random vector (variable) is fully determined by its distribution function.*

Remark 3.7.
(1) Every function with properties (1)–(4) is a distribution function of a random vector (variable) defined on a suitable probability space (Kolmogorov-type theorems).
(2) Frequently distribution functions are introduced alternatively using sharp inequality
$F_{\mathbb{X}}(x_1,\dots,x_n) := P(\mathbb{X} < \boldsymbol{x}) = P(X_1 < x_1,\dots,X_n < x_n)$.
When using this definition, all properties remain valid except (2) where continuity from the right is to be replaced by continuity from the left.
(3) Simplified notation for limits similar to those of 3.4(3)(4):
$F(\infty,\dots,\infty) := \lim_{x_1\to\infty,\dots,x_n\to\infty} F(x_1,\dots,x_n)$ or
$F(x_1,\dots,x_{i-1},\pm\infty,x_{i+1},\dots,x_n) := \lim_{x_i\to\pm\infty} F(x_1,\dots,x_n)$. Analogical shortcuts may be used if $x_i \to \pm\infty$ for any subset of subscripts $i \in I \subseteq \{1,\dots,n\}$.
(4) We denote the respective limits from the left and right at $\boldsymbol{x}$ by
$F(\boldsymbol{x}^-) := \lim_{\boldsymbol{y}\uparrow\boldsymbol{x}} F(\boldsymbol{y})$ if $y_i < x_i$ for some i, and
$F(\boldsymbol{x}^+) := \lim_{\boldsymbol{y}\downarrow\boldsymbol{x}} F(\boldsymbol{y})$ if $y_i > x_i$ for some i.
The property (2) of 3.4 may be then rewritten as $F(\boldsymbol{x}^+) = F(\boldsymbol{x})$.

Theorem 3.8. *Let $\mathbb{X} = [X_1,\dots,X_n]^T$ be n-dimensional random vector. Then $P(\mathbb{X} = \boldsymbol{x}) = F_{\mathbb{X}}(\boldsymbol{x}) - F_{\mathbb{X}}(\boldsymbol{x}^-)$ holds for each $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. $\boldsymbol{x}_n \uparrow \boldsymbol{x}$, $\boldsymbol{x}_n < \boldsymbol{x} \Rightarrow \lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}}(\boldsymbol{x}_n,\boldsymbol{x}] = \cap_{n=1}^{\infty}(\boldsymbol{x}_n,\boldsymbol{x}] = \{\boldsymbol{x}\} \Rightarrow$
$P_{\mathbb{X}}(\{\boldsymbol{x}\}) = P_{\mathbb{X}}(\lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}}(\boldsymbol{x}_n,\boldsymbol{x}]) \overset{(P13)}{=} \lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}} P_{\mathbb{X}}((\boldsymbol{x}_n,\boldsymbol{x}]) =$
$\lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}} P_{\mathbb{X}}((-\infty,\boldsymbol{x}] - (-\infty,\boldsymbol{x}_n]) \overset{(P8)}{=}$
$\lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}}(P_{\mathbb{X}}((-\infty,\boldsymbol{x}]) - P_{\mathbb{X}}((-\infty,\boldsymbol{x}_n])) =$
$P_{\mathbb{X}}((-\infty,\boldsymbol{x}]) - \lim_{\boldsymbol{x}_n\uparrow\boldsymbol{x}} P_{\mathbb{X}}((-\infty,\boldsymbol{x}_n]) = F_{\mathbb{X}}(\boldsymbol{x}) - F_{\mathbb{X}}(\boldsymbol{x}^-)$. □

Corollary 3.9. *If $F_{\mathbb{X}}$ is continuous at $\boldsymbol{x}$ then $P(\mathbb{X} = \boldsymbol{x}) = 0$.*

3.2. **The Probability Density Function of a Random Vector and Its Properties.**

Definition 3.10.
The random vector $\mathbb{X} = [X_1, \dots, X_n]^T$ is said to be:
(1) **(absolutely) continuous** if $F_{\mathbb{X}}(\boldsymbol{x})$ is an absolutely continuous function on $\mathbb{R}^n$; i.e. there exists a function $f_{\mathbb{X}}(\boldsymbol{x}) := f_{\mathbb{X}}(x_1, \dots, x_n)$: $F_{\mathbb{X}}(\boldsymbol{y}) = \int_{-\infty}^{\boldsymbol{y}} f_{\mathbb{X}}(\boldsymbol{x})\, d\boldsymbol{x} := \int_{-\infty}^{y_n} \dots \int_{-\infty}^{y_1} f_{\mathbb{X}}(x_1, \dots, x_n)\, dx_1 \dots dx_n$;
(2) **discrete** if there exist at most countably many isolated points $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots \in \mathbb{R}^n$ such that $f_{\mathbb{X}}(\boldsymbol{x}) := P(\mathbb{X} = \boldsymbol{x}) > 0$ for each $\boldsymbol{x} = \boldsymbol{\xi}_i$, $P(\mathbb{X} = \boldsymbol{x}) = 0$ elsewhere, and $F_{\mathbb{X}}(\boldsymbol{x}) = \sum_{\boldsymbol{\xi}_i \le \boldsymbol{x}} f_{\mathbb{X}}(\boldsymbol{\xi}_i) = \sum_{\boldsymbol{x}' \le \boldsymbol{x}} f_{\mathbb{X}}(\boldsymbol{x}')$;
(3) **of mixed type** if neither (1) nor (2) holds for $F_{\mathbb{X}}$.

Definition 3.11. Let $\mathbb{X} = [X_1, \dots, X_n]^T$ be a random vector. The function $f_{\mathbb{X}}(\boldsymbol{x})$ of definition 3.10 is called **probability density function of** $\mathbb{X}$ (p.d.f.), or for $n > 1$ sometimes also **joint probability density function of** $\mathbb{X}$ to emphasize dimension $n > 1$. We obtain as special cases the p.d.f. $f_X(x)$ of a random variable X with $n = 1$, and $f_{\mathbb{X}}(x_1, x_2)$ of the bivariate random variable $\mathbb{X} = [X_1, X_2]^T$ with $n = 2$. The subscript may be omitted if there is no danger of misunderstanding.

Theorem 3.12 (Properties of p.d.f.)**.**
Probability density $f(\boldsymbol{x})$ of each continuous or discrete n-dimensional random vector $\mathbb{X}$ has the following properties:
(1) $f(\boldsymbol{x}) \ge 0$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$*;*
• $\mathbb{X}$ *continuous:*
(2a) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\boldsymbol{x})\, d\boldsymbol{x} = 1$*;*
(3a) $P(\mathbb{X} = \boldsymbol{x}) = 0$ *for each* $\boldsymbol{x} \in \mathbb{R}^n$*;*
(4a) $P(\boldsymbol{a} < \mathbb{X} \le \boldsymbol{b}) = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(\boldsymbol{x})\, d\boldsymbol{x}$*;*
(5a) For each $B \in \mathcal{B}_n$*:* $P(\mathbb{X} \in B) = \int_B f(\boldsymbol{x})\, d\boldsymbol{x}$*;*
(6a) The distribution function $F(\boldsymbol{x})$ has the n-th order partial derivatives at continuity points of f where $\frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n) = f(\boldsymbol{x})$.
• $\mathbb{X}$ *discrete with positive probabilities concentrated at* $\boldsymbol{\xi}_i$*:*

(2b) $\sum_{\boldsymbol{\xi}_i} f(\boldsymbol{\xi}_i) = \sum_{\boldsymbol{x}'} f(\boldsymbol{x}') = 1$;
(3b) $f_{\mathbb{X}}(\boldsymbol{x}) := P(\mathbb{X} = \boldsymbol{x}) = F_{\mathbb{X}}(\boldsymbol{x}) - F_{\mathbb{X}}(\boldsymbol{x}^-)$ *is nonzero iff* $\boldsymbol{x} = \boldsymbol{\xi}_i$;
(4b) $P(\boldsymbol{a} < \mathbb{X} \leq \boldsymbol{b}) = \sum_{\boldsymbol{a} < \boldsymbol{\xi}_i \leq \boldsymbol{b}} f(\boldsymbol{\xi}_i) = \sum_{\boldsymbol{a} < \boldsymbol{x}' \leq \boldsymbol{b}} f(\boldsymbol{x}')$;
(5b) For each $B \in \mathcal{B}_n$: $P(\mathbb{X} \in B) = \sum_{\boldsymbol{\xi}_i \in B} f(\boldsymbol{\xi}_i) = \sum_{\boldsymbol{x}' \in B} f(\boldsymbol{x}')$;
(6b) The distribution function $F(\boldsymbol{x})$ *is a non-decreasing rectangular staircase function. The position of each 'stair' is given by its 'angle' at* $\boldsymbol{\xi}_i$ *and 'height'* $f(\boldsymbol{\xi}_i)$.

Remark 3.13.
(1) The probability of a continuous random vector is 'continuously' spread over $\mathbb{R}^n$ with eventually varying intensity (see (5a)). In contrast with that, the probability of a discrete random vector is concentrated at finite or countable set of isolated points (see (3b) and (5b)).
The mixed type lies somewhere in-between: c.d.f. is usually a sum of absolutely continuous component and discrete component. With $n > 1$ there may be semi-discrete components with probability concentrated on suitable subsets of dimension less than n: on lines and curves with $n = 2$ or even on surfaces with $n = 3$, or on more complex formations with $n > 3$.
(2) Exceptionally c.d.f. may contain singular component which is continuous but unfortunately not absolutely, not allowing us to construct p.d.f. because n-th order derivative of F does not exist on a countable subset dense in $\mathbb{R}^n$, typically on $\mathbb{Q}^n$ ($\mathbb{Q}$...set of rational numbers).
(3) Similarly to remark 3.7(1) every function with properties (1)(2a) or (1)(2b) is p.d.f of a suitable random vector (variable).

Example 3.14.
(1) Continuous type: Ball falling down on a flat interval or rectangle with uniform probability distribution.
(2) Discrete type: Similar as (1) but interval or rectangle fully covered with triangular or pyramidal hollows.
(3) Mixed type: Similar as (2) but with only partial covering with triangular or pyramidal hollows or grooves.

3.3. **Marginal and Conditional Distributions.**

Definition 3.15. Let $\mathbb{X} = [X_1, \dots, X_n]^T$ be a random vector with c.d.f. F, and $\mathbb{X}_{\boldsymbol{t}} := [X_{t_1}, X_{t_2}, \dots, X_{t_m}]^T$ its marginal vector as of corollary 2.9, $m < n$. Its distribution function $F_{\boldsymbol{t}} := F_{\mathbb{X}_{\boldsymbol{t}}}$ is called **(joint) marginal distribution function of F at $\boldsymbol{t}$**. In particular for all $t \in \{1, \dots, n\}$ we obtain the univariate marginal distributions $F_t := F_{X_t}$ as distributions of the component random variables X_t.

Theorem 3.16.
For the marginal distribution function of definition 3.15 it holds

$$F_{\boldsymbol{t}}(\boldsymbol{x_t}) := F_{\boldsymbol{t}}(x_{t_1}, \dots, x_{t_m}) = F(\infty, \dots, \infty, x_{t_1}, \infty, \dots, \infty, x_{t_m}, \infty, \dots, \infty).$$

In particular for $t \in \{1, \dots, n\}$ we have

$$F_t(x_t) = F(\infty, \dots, \infty, x_t, \infty, \dots, \infty).$$

Proof. $F_{\boldsymbol{t}}(x_{t_1}, \dots, x_{t_m}) = P(\{\omega \mid X_i(\omega) \le x_i \text{ for } i \in \{t_1, \dots, t_m\}\}) = P(\{\omega \mid X_i(\omega) \le x_i \text{ for } i \in \{t_1, \dots, t_m\} \text{ and } X_i(\omega) \in \mathbb{R} \text{ otherwise }\}) = F(\infty, \dots, \infty, x_{t_1}, \infty, \dots, \infty, x_{t_m}, \infty, \dots, \infty)$. □

Corollary 3.17. *If there exists p.d.f. f of $\mathbb{X}$ then p.d.f. $f_{\boldsymbol{t}} := f_{\mathbb{X}_t}$ exists for each marginal random vector $\mathbb{X}_{\boldsymbol{t}}$, and is called* **(joint) marginal p.d.f. of f at $\boldsymbol{t}$**. *Putting $\boldsymbol{s} := [s_1, \dots, s_{n-m}]$ where $\{s_1, \dots, s_{n-m}\} := \{1, \dots, n\} - \{t_1, \dots, t_m\}$, it holds:*

- *$\mathbb{X}$ continuous:*

$$f_{\boldsymbol{t}}(\boldsymbol{x_t}) := f_{\boldsymbol{t}}(x_{t_1}, \dots, x_{t_m}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_{s_1} \dots dx_{s_{n-m}} =: \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x})\, d\boldsymbol{x_s}.$$

In particular for $t \in \{1, \dots, n\}$ we have

$$f_t(x_t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_{x_1} \dots dx_{t-1}\, dx_{t+1} \dots dx_n =: \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x})\, d\boldsymbol{x}(t).$$

- *$\mathbb{X}$ discrete:*

$$f_{\boldsymbol{t}}(\boldsymbol{x_t}) := f_{\boldsymbol{t}}(x_{t_1}, \dots, x_{t_m}) = \sum_{x_{s_1}} \cdots \sum_{x_{s_{n-m}}} f(x_1, \dots, x_n) =: \sum_{\boldsymbol{x_s}} f(\boldsymbol{x}).$$

In particular for $t \in \{1, \dots, n\}$ we have

$$f_t(x_t) = \sum_{x_1} \cdots \sum_{x_{t-1}} \sum_{x_{t+1}} \cdots \sum_{x_n} f(x_1, \dots, x_n) =: \sum_{\boldsymbol{x}(t)} f(\boldsymbol{x}).$$

Proof. Using theorem 3.16 we get:
1. $\mathbb{X}$ continuous: When interchanging the integration order in definition 3.10(1) by Fubini's theorem (due to $f(\boldsymbol{x}) \geq 0$), we get:

$$F_{\boldsymbol{t}}(\boldsymbol{y_t}) = \int_{-\infty}^{y_{t_1}} \cdots \int_{-\infty}^{y_{t_m}} \underbrace{\Big(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, d\boldsymbol{x_s} \Big)}_{f_{\boldsymbol{t}}(\boldsymbol{x_t})} d\boldsymbol{x_t}. \tag{3.1a}$$

2. $\mathbb{X}$ discrete: We have by definition 3.10(2) $F(\boldsymbol{y}) = \sum_{\boldsymbol{x} \leq \boldsymbol{y}} f(\boldsymbol{x})$. When interchanging the summation order due to absolute convergence ($f(\boldsymbol{x}) \geq 0$), we get:

$$F_{\boldsymbol{t}}(\boldsymbol{y_t}) = \sum_{\boldsymbol{x_t} \leq \boldsymbol{y_t}} \underbrace{\Big(\sum_{\boldsymbol{x_s} \in \mathbb{R}^{n-m}} f(x_1, \ldots, x_n) \Big)}_{f_{\boldsymbol{t}}(\boldsymbol{x_t})}. \tag{3.1b}$$

It is straightforward to see from equations (3.1a) and (3.1b) that $f_{\boldsymbol{t}}(\boldsymbol{x_t})$ are probability densities of the marginal random vector $\mathbb{X}_{\boldsymbol{t}}$ in the sense of definitions 3.10 and 3.11, preserving the type (continuous or discrete) of the original random vector $\mathbb{X}$, and satisfying the necessary conditions (1),(2a) or (1),(2b) of theorem 3.12. Clearly, the remaining properties are their consequence in view of remark 3.13(3). ☐

Remark 3.18 (Conditional distribution — introductory remarks). For simplicity we shall start the discussion with the bivariate random vector $\mathbb{X} = [X_1, X_2]^T$ defined on the probability space $(\Omega, \mathcal{A}, P)$, having joint p.d.f. $f(x_1, x_2)$ and marginal p.d.f. $f_1(x_1)$ of X_1 and $f_2(x_2)$ of X_2. Given x_2 arbitrary but fixed, our goal is to develop a concept allowing us to evaluate for each Borel set $B_1 \in \mathcal{B}$ the conditional probability $P(X_1 \in B_1 \,|\, X_2 = x_2)$, or more exactly $P(X_1^{-1}(B_1) \,|\, X_2^{-1}(\{x_2\}) = \frac{P(X_1^{-1}(B_1) \cap X_2^{-1}(\{x_2\}))}{P(X_2^{-1}(\{x_2\}))}$ in the sense of eq. (2.1). In accordance with definition 1.9, the fraction defines conditional probability P_2 on a new probability space $(A_2, \mathcal{A}_2, P_2)$ where

$A_2 = X_2^{-1}(\{x_2\}) \subseteq \Omega$, $P(A_2) > 0$, is a restricted sample space and $\mathcal{A}_2$ σ-field of restricted events obtained as intersections of the original ones with A_2. Thus restriction of the random variable X_2 on A_2 yields measurable mapping $X_1' : (A_2, \mathcal{A}_2, P_2) \to (\mathbb{R}, \mathcal{B})$ because $X_1'^{-1}(B_1) = A_2 \cap X_1^{-1}(B_1) \in \mathcal{A}_2$ for each $B_1 \in \mathcal{B}$. Then $P(X_1 \in B_1 \,|\, X_2 = x_2) = P_2(X_1' \in B_1)$ and it is sufficient to find p.d.f. and c.d.f. of X_1', hereafter denoted as $f(\cdot \,|\, x_2)$ and $F(\cdot \,|\, x_2)$, respectively.

1. <u>$\mathbb{X}$ discrete</u>: By 3.17 both marginal random variables X_1 and X_2 are discrete as well, so as is the restriction X_1' of X_1. In view of 3.12(3b) $f(x_1 \,|\, x_2) = P_2(X_1' = x_1) = P(X_1 = x_1 \,|\, X_2 = x_2) = \frac{P(X_1=x_1, X_2=x_2)}{P(X_2=x_2)} = \frac{f(x_1,x_2)}{f_2(x_2)}$ provided that $f_2(x_2) = P(X_2 = x_2) = P(A_2) > 0$. By definition 3.10(2) $F(x_1 \,|\, x_2) = \sum_{y_1 \le x_1} f(y_1 \,|\, x_2)$ and $P(X_1 \in B_1 \,|\, X_2 = x_2) = P_2(X_1' \in B_1) = \sum_{y_1 \in B_1} f(y_1 \,|\, x_2)$ for arbitrary $B_1 \in \mathcal{B}$.

2. <u>$\mathbb{X}$ continuous</u>: By 3.17 both marginal random variables X_1 and X_2 are continuous as well. The procedure of discrete case cannot be applied because neither $f(x_1, x_2)$ nor $f_2(x_2)$ are probabilities of some events and $P(X_1 = x_1, X_2 = x_2) = P(X_2 = x_2) = 0$ by 3.12(3a). They are true probability densities by 3.12(6a) at every continuity point of $f(x_1, x_2)$ and $f_2(x_2)$:
$F(y_1 \,|\, x_2) = P(X_1 \le y_1 \,|\, X_2 = x_2) = \frac{P(X_1 \le y_1, X_2 = x_2)}{P(X_2 = x_2)} = \lim_{h_2 \downarrow 0} \frac{P(X_1 \le y_1, x_2 \le X_2 \le x_2 + h_2)}{P(x_2 \le X_2 \le x_2 + h_2)} = \lim_{h_2 \downarrow 0} \frac{F(y_1, x_2+h_2) - F(y_1, x_2)}{F_2(x_2+h_2) - F_2(x_2)} = \lim_{h_2 \downarrow 0} \frac{(F(y_1, x_2+h_2) - F(y_1, x_2))/h_2}{(F_2(x_2+h_2) - F_2(x_2))/h_2} = \frac{\partial F(y_1, x_2)/\partial x_2}{\partial F_2(x_2)/\partial x_2} = \frac{\int_{-\infty}^{y_1} f(x_1, x_2)\, dx_1}{f_2(x_2)} = \int_{-\infty}^{y_1} \frac{f(x_1,x_2)}{f_2(x_2)}\, dx_1$ provided that $f_2(x_2) > 0$. Then X_1' is continuous as well with p.d.f. $f(x_1 \,|\, x_2) = \frac{f(x_1,x_2)}{f_2(x_2)}$.

It is straightforward to extend the procedure of remark 3.18 to any n-dimensional random vector $\mathbb{X}$ and any pair of its marginal random vectors $\mathbb{X}_{\boldsymbol{t}}$ and $\mathbb{X}_{\boldsymbol{s}}$ at complementary subscript vectors $\boldsymbol{t}$ and $\boldsymbol{s}$ as they were introduced in corollary 3.17. This leads to the general concept of conditional multivariate p.d.f. and c.d.f. as follows.

Definition 3.19. Let $\mathbb{X} = [X_1, \ldots, X_n]^T$ be a discrete or continuous random vector defined on the probability space $(\Omega, \mathcal{A}, P)$, having c.d.f. F and p.d.f. f. Let $\mathbb{X}_{\boldsymbol{t}} := [X_{t_1}, X_{t_2}, \ldots, X_{t_m}]^T$ and $\mathbb{X}_{\boldsymbol{s}} := [X_{s_1}, X_{s_2}, \ldots, X_{s_{n-m}}]^T$ be its marginal random vectors with marginal p.d.f.'s $f_{\boldsymbol{t}}(\boldsymbol{x_t})$ and $f_{\boldsymbol{s}}(\boldsymbol{x_s})$ at complementary subscript vectors $\boldsymbol{t}$ and $\boldsymbol{s}$ in the sense of corollary 3.17, $1 \leq m < n$. Given $\boldsymbol{x_s} \in \mathbb{R}^{n-m}$ arbitrary but fixed such that $f_{\boldsymbol{s}}(\boldsymbol{x_s}) > 0$, then

$$f(\boldsymbol{x_t} \,|\, \boldsymbol{x_s}) := \frac{f(\boldsymbol{x})}{f_{\boldsymbol{s}}(\boldsymbol{x_s})}$$

is called **conditional probability density function of $\mathbb{X}_{\boldsymbol{t}}$, given** $\mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}$. The corresponding distribution function (see definition 3.10)

$$F(\boldsymbol{x_t} \,|\, \boldsymbol{x_s}) = \begin{cases} \sum_{\boldsymbol{x}'_{\boldsymbol{t}} \leq \boldsymbol{x_t}} f(\boldsymbol{x}'_{\boldsymbol{t}} \,|\, \boldsymbol{x_s}) & \text{for discrete } \mathbb{X} \\ \int_{-\infty}^{x_{t_1}} \cdots \int_{-\infty}^{x_{t_m}} f(\boldsymbol{x}'_{\boldsymbol{t}} \,|\, \boldsymbol{x_s})\, d\boldsymbol{x}'_{\boldsymbol{t}} & \text{for continuous } \mathbb{X} \end{cases}$$

is called **conditional distribution function of $\mathbb{X}_{\boldsymbol{t}}$, given** $\mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}$.

Theorem 3.20. *The conditional p.d.f. and c.d.f. of 3.19 are true p.d.f. and c.d.f. of the restriction of the marginal random vector $\mathbb{X}_{\boldsymbol{t}}$ to the conditional probability space, given event $\mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}$ (c.f. definition 1.9) allowing us to compute conditional probability $P(\mathbb{X}_{\boldsymbol{t}} \in B \,|\, \mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}) = \frac{P(\mathbb{X}_{\boldsymbol{t}} \in B, \mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s})}{P(\mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s})}$ for any Borel set $B \in \mathcal{B}_m$ using formula (see theorem 3.12(5))*

$$P(\mathbb{X}_{\boldsymbol{t}} \in B \,|\, \mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}) = \begin{cases} \sum_{\boldsymbol{x_t} \in B} f(\boldsymbol{x_t} \,|\, \boldsymbol{x_s}) & \textit{for discrete } \mathbb{X} \\ \int_B f(\boldsymbol{x_t} \,|\, \boldsymbol{x_s})\, d\boldsymbol{x_t} & \textit{for continuous } \mathbb{X}. \end{cases} \tag{3.2}$$

Corollary 3.21. *If $\mathbb{X}$ is discrete we get by theorem 3.12(3b):*

$$P(\mathbb{X}_{\boldsymbol{t}} = \boldsymbol{x_t} \,|\, \mathbb{X}_{\boldsymbol{s}} = \boldsymbol{x_s}) = f(\boldsymbol{x_t} \,|\, \boldsymbol{x_s}). \tag{3.3}$$

Example 3.22. See [Rou97, p.95,96] for the construction of marginal and conditional p.d.f. of the discrete n-dimensional random vector having multinomial distribution. For example $f(x_1, \dots, x_6) := P(X_1 = x_1, ..., X_6 = x_6)$ is p.d.f. of such 6-dimensional random vector describing probability that out of k casts there will be 1 spot x_1-times, 2 spots x_2-times, ..., 6 spots x_6-times ($k = x_1 + \dots + x_6$). Putting $\boldsymbol{t} := [1, 3, 5]$ we have complementary subscripts $\boldsymbol{s} := [2, 4, 6]$ and using marginal p.d.f. $f_{\boldsymbol{t}}(x_1, x_3, x_5)$ we may confine ourselves to probabilities of events related to sides with odd number of spots, eventually to conditional probabilities of such events, when we know for example that after a series of k casts 2 spots appeared once, 4 spots four times and 6 spots five times: $\boldsymbol{x_s} = [1, 4, 5]$ and $f_{\boldsymbol{t}}(x_1, x_3, x_5 \,|\, 1, 4, 5)$ is the corresponding conditional p.d.f.

3.4. **Independence of random variables and vectors.**

Definition 3.23.
The random variables $X_1, \dots, X_n$ defined on the same probability space $(\Omega, \mathcal{A}, P)$ are said to be **independent** if for every choice of Borel sets $B_1, \dots, B_n \in \mathcal{B}$ the events $X_1 \in B_1, \dots, X_n \in B_n$ are independent in the sense of definition 2.20.

Theorem 3.24.
Given random vector $\mathbb{X} = [X_1, \dots, X_n]^T$ with c.d.f. $F_{\mathbb{X}}$ and p.d.f. $f_{\mathbb{X}}$ then the following statements are equivalent:

(1) *The random variables $X_1, \dots, X_n$ are independent.*
(2) *For every choice of Borel sets $B_1, \dots, B_n \in \mathcal{B}$ it holds: $P(X_1 \in B_1, \dots, X_n \in B_n) = \Pi_{i=1}^n P(X_i \in B_i)$.*
(3) *$F_{\mathbb{X}}(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n)$ for all $x_1, \dots, x_n \in \mathbb{R}$.*
(4) *$f_{\mathbb{X}}(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$ for all $x_1, \dots, x_n \in \mathbb{R}$.*

Proof.
(1)⇒(2): is the direct consequence of the probability product rule of independence by definition 2.20.

(2)⇒(1): Given $B_i \in \mathcal{B}$, $i = 1, \ldots, n$, arbitrary but fixed, and $1 \le j_1 < \cdots < j_k \le n$ any subset of subscripts, we introduce new Borel sets B_i' as $B_i' := B_i$ for $i \in \{j_1, \ldots, j_k\}$ and $B_i := \mathbb{R}$ otherwise. In the latter case $P(X_i \in B_i') = 1$ and applying (2) on B_i' we have $P(X_{j_1} \in B_{j_1}, \ldots, X_{j_k} \in B_{j_k}) = P(X_1 \in B_1', \ldots, X_n \in B_n') = \Pi_{i=1}^n P(X_i \in B_i') = \Pi_{m=1}^k P(X_{j_m} \in B_{j_m})$ confirming the probability product rule for any subset of events $X_{j_1} \in B_{j_1}, \ldots, X_{j_k} \in B_{j_k}$, and thus independence in the sequel.

(2)⇒(3): Putting $B_i := (-\infty, x_i]$ for $i = 1, \ldots, n$ we have by definition of c.d.f. $F(x_1, \ldots, x_n) = P(X_1 \in (-\infty, x_1], \ldots, X_n \in (-\infty, x_n]) \overset{(2)}{=} \Pi_{i=1}^n P(X_i \in (-\infty, x_i]) = \Pi_{i=1}^n F_i(x_i)$.

$\mathbb{X}$ discrete:

(3)⇒(4): For $\boldsymbol{a} < \boldsymbol{b}$, $\boldsymbol{a} \uparrow \boldsymbol{b}$ we have $f(b_1, \ldots, b_n) = P(\mathbb{X} = \boldsymbol{b}) = \lim_{\boldsymbol{a} \uparrow \boldsymbol{b}} P(\boldsymbol{a} < \mathbb{X} \le \boldsymbol{b}) =$
$\lim_{a_n \uparrow b_n} \Delta^n_{\boldsymbol{a},\boldsymbol{b}} \ldots \lim_{a_1 \uparrow b_1} \Delta^1_{\boldsymbol{a},\boldsymbol{b}} F_1(b_1) \ldots F_n(b_n) =$
$$\lim_{a_n \uparrow b_n} \Delta^n_{\boldsymbol{a},\boldsymbol{b}} \ldots \lim_{a_2 \uparrow b_2} \Delta^2_{\boldsymbol{a},\boldsymbol{b}} \underbrace{\lim_{a_1 \uparrow b_1} (F_1(b_1) - F_1(a_1))}_{P(X_1 = b_1) \text{ by theorem 3.8}} F_2(b_2) \ldots F_n(b_n)$$
$= \cdots = f_1(b_1) \ldots f_n(b_n)$.

(4)⇒(2): $P(X_1 \in B_1, \ldots, X_n \in B_n) = \sum_{B_1 \times \cdots \times B_n} f(x_1, \ldots, x_n) \overset{(*)}{=} \sum_{B_1} \cdots \sum_{B_n} f_1(x_1) \ldots f_n(x_n) = (\sum_{B_1} f_1(x_1)) \ldots (\sum_{B_n} f_n(x_n)) = P(X_1 \in B_1) \ldots P(X_n \in B_n)$ in view of theorem 3.12(5b) where identity $(*)$ is by absolute convergence due to nonnegative addents.

$\mathbb{X}$ continuous:

(3)⇒(4): $f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_1(x_1) \ldots F_n(x_n) = \frac{\partial}{\partial x_1} F_1(x_1) \ldots \frac{\partial}{\partial x_n} F_n(x_n) = f_1(x_1) \ldots f_n(x_n)$ at continuity points of f in view of theorem 3.12(6a).

(4)⇒(2): $P(X_1 \in B_1, \ldots, X_n \in B_n) = \int_{B_1 \times \cdots \times B_n} f(x_1, \ldots, x_n)\, d\boldsymbol{x} \overset{(*)}{=} \int_{B_1} \cdots \int_{B_n} f_1(x_1) \ldots f_n(x_n)\, dx_1 \ldots dx_n = (\int_{B_1} f_1(x_1)\, dx_1) \ldots (\int_{B_n} f_n(x_n)\, dx_n) = P(X_1 \in B_1) \ldots P(X_n \in B_n)$

in view of theorem 3.12(5a) where identity $(*)$ is by Fubini's theorem due to nonnegative integrand. $\square$

Corollary 3.25. *Any subset of independent random variables is independent as well.*

Proof. The statement is a consequence of theorem 2.21(2). $\square$

Corollary 3.26.
*If $X_1, \ldots, X_n$ are independent random variables with c.d.f.'s F_i and p.d.f.'s f_i, $i = 1, \ldots, n$, then $X_{(1)}(\omega) := \min(X_1(\omega), \ldots, X_n(\omega))$ and $X_{(n)}(\omega) := \max(X_1(\omega), \ldots, X_n(\omega))$, $\omega \in \Omega$, are random variables with c.d.f.'s $F_{(1)}(x) = 1 - \Pi_{i=1}^n (1 - F_i(x))$, $F_{(n)}(x) = \Pi_{i=1}^n F_i(x)$ and p.d.f.'s $f_{(1)}(x) = \sum_{j=1}^n f_j(x) \Pi_{i=1, i \neq j}^n (1 - F_i(x))$,
$f_{(n)}(x) = \sum_{j=1}^n f_j(x) \Pi_{i=1, i \neq j}^n F_i(x)$ (at continuity points of f_i when $\mathbb{X}$ is continuous), respectively.*

Theorem 3.27. *Given independent variables $X_1, \ldots, X_n$ and Borel functions $\varphi_i : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$, $i = 1, \ldots, n$, then the transformed random variables $\varphi_1(X_1), \ldots, \varphi_n(X_n)$ are independent as well.*

Proof. Let $B_1, \ldots, B_n \in \mathcal{B}$ be arbitrary Borel sets.
Clearly $\varphi_1(X_1) \in B_1, \ldots, \varphi_n(X_n) \in B_n$ are the same events as $X_1 \in \varphi_1^{-1}(B_1), \ldots, X_n \in \varphi_n^{-1}(B_n)$ where $\varphi_i^{-1}(B_i)$ are Borel sets due to measurability of φ_i for $i = 1, \ldots, n$. They must be independent because $X_1, \ldots, X_n$ are independent random variables. $\square$

Remark 3.28. Properties (3) and (4) of theorem 3.24 may be formulated in stronger form due to remarks 3.7(1) and 3.13(3):
(3') If $F(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n)$ for all $x_i \in \mathbb{R}$ is c.d.f. of some $\mathbb{X} = [X_1, \ldots, X_n]$ then X_i are independent random variables. Conversely: if F_i are c.d.f.'s of some random variables X_i then F is c.d.f. of the random vector $\mathbb{X} = [X_1, \ldots, X_n]$ where X_i are independent with marginal c.d.f.'s F_i for all $i = 1, \ldots, n$.
(4') If $f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n)$ for all $x_i \in \mathbb{R}$ is p.d.f. of some $\mathbb{X} = [X_1, \ldots, X_n]$ then X_i are independent random variables.

Conversely: if f_i are p.d.f.'s of some random variables X_i then f is p.d.f. of the random vector $\mathbb{X} = [X_1, \dots, X_n]$ where X_i are independent with marginal p.d.f.'s f_i for all $i = 1, \dots, n$.

Remark 3.29 (Independence of random vectors).
The above concept of independence of random variables is easy to extend to random vectors as follows:
• Random variables X_i are to be replaced by random vectors $\mathbb{X}_i$ of dimension m_i and Borel sets $B_i \in \mathcal{B}_{m_i}$ for $i = 1, \dots, n$.
• In Theorem 3.24 the random vectors $\mathbb{X}_i$ are merged into a random vector $\mathbb{X}$ of dimension $m_1 + \cdots + m_n$, thus being its marginal subvectors with marginal (joint) c.d.f.'s F_i and marginal p.d.f.'s f_i.
• In Theorem 3.27 the Borel functions are arbitrary $\varphi_i : (\mathbb{R}^{m_i}, \mathcal{B}_{m_i}) \to (\mathbb{R}^{k_i}, \mathcal{B}_{k_i})$, the transformed random vectors $\varphi_i(\mathbb{X}_i)$ being then of dimension k_i for $i = 1, \dots, n$.

Detailed discussion of basic theoretical concepts given above can be found along with numerous illustrative examples and exercises primarily in [Rou97]. Additional information is available for example in [HC78], or in a lot of other introductory textbooks on mathematical statistics.

References

[HC78] Robert V. Hogg and Allen T. Craig, *Introduction to mathematical statistics*, 4-th ed., Macmillan Publishing Co., Inc., New York, 1978.

[Rou97] Georg G. Roussas, *A course in mathematical statistics*, 2-nd ed., Academic Press, San Diego, 1997.

Doc. RNDr. Vítězslav Veselý, CSc., Department of Applied Mathematics, Masaryk University of Brno, Janáčkovo nám. 2a, 662 95 Brno, Czech republic.

tel.: +420-5-41321251/35, fax: +420-5-41210337

E-mail address: vesely@math.muni.cz, *URL:* http://www.math.muni.cz/~vesely