(i) F is non-decreasing and right-continuous,

(ii) $\forall x \in \mathrm{Cont}(F): \lim_{i\to\infty} F_{n_i}(x) = F(x)$.

Proof: Ad (i): Obviously F is non-decreasing. For $x \in \mathbb{R}$ and $\varepsilon > 0$ take $\delta_2 > 0$ such that
$$\forall q \in \mathbb{Q} \cap]x, x+\delta_2[: \quad G(q) \leq F(x) + \varepsilon.$$
Thus, for $z \in]x, x+\delta_2[$,
$$F(x) \leq F(z) \leq F(x) + \varepsilon.$$
Ad (ii): If $x \in \mathrm{Cont}(F)$ and $\varepsilon > 0$ take $\delta_1 > 0$ such that
$$F(x) - \varepsilon \leq F(x - \delta_1).$$
Thus, for $q_1, q_2 \in \mathbb{Q}$ with
$$x - \delta_1 < q_1 < x < q_2 < x + \delta_2,$$
we get
$$F(x) - \varepsilon \leq F(x-\delta_1) \leq G(q_1) \leq \liminf_{i\to\infty} F_{n_i}(x) \leq \limsup_{i\to\infty} F_{n_i}(x) \\ \leq G(q_2) \leq F(x) + \varepsilon.$$
Claim:
$$\lim_{x\to-\infty} F(x) = 0 \ \wedge \ \lim_{x\to\infty} F(x) = 1.$$
Proof: For $\varepsilon > 0$ take $m \in \mathbb{Q}$ such that
$$\forall n \in \mathbb{N}: \ P_n(]-m, m]) \geq 1 - \varepsilon.$$
Thus
$$G(m) - G(-m) = \lim_{i\to\infty} \big(F_{n_i}(m) - F_{n_i}(-m)\big) = \lim_{i\to\infty} P_{n_i}(]-m, m]) \geq 1 - \varepsilon.$$
Since $F(m) \geq G(m)$ and $F(-m-1) \leq G(-m)$, we obtain
$$F(m) - F(-m-1) \geq 1 - \varepsilon.$$
It remains to apply Theorems 1.3 and 2. □

4 Uniform Integrability

In the sequel: X_n, X random variables on a common probability space $(\Omega, \mathfrak{A}, P)$.

Definition 1. $(X_n)_{n\in\mathbb{N}}$ *uniformly integrable (u.i.)* if
$$\lim_{\alpha\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|\geq\alpha\}} |X_n| \, dP = 0.$$

Remark 1.

(i) $(X_n)_{n\in\mathbb{N}}$ u.i. $\Rightarrow (\forall\, n \in \mathbb{N}:\ X_n \in \mathfrak{L}^1) \wedge \sup_{n\in\mathbb{N}} \|X_n\|_1 < \infty$.

(ii) $\exists\, Y \in \mathfrak{L}^1\ \forall\, n \in \mathbb{N}:\ |X_n| \le Y \Rightarrow (X_n)_{n\in\mathbb{N}}$ u.i.

(iii) $\exists\, p > 1\ (\forall\, n \in \mathbb{N} : X_n \in \mathfrak{L}^p) \wedge \sup_{n\in\mathbb{N}} \|X_n\|_p < \infty \Rightarrow (X_n)_{n\in\mathbb{N}}$ u.i.
Proof: $\int_{\{|X_n|\ge\alpha\}} |X_n|\, dP = 1/\alpha^{p-1} \cdot \int_{\{|X_n|\ge\alpha\}} \alpha^{p-1}|X_n|\, dP \le 1/\alpha^{p-1} \cdot \|X_n\|_p^p$.

Example 1. For the uniform distribution P on $[0,1]$ and

$$X_n = n \cdot 1_{[0,1/n]}$$

we have $X_n \in \mathfrak{L}^1$ and $\|X_n\|_1 = 1$, but for any $\alpha > 0$ and $n \ge \alpha$

$$\int_{\{|X_n|\ge\alpha\}} |X_n|\, dP = n \cdot P([0,1/n]) = 1,$$

so that $(X_n)_{n\in\mathbb{N}}$ is not u.i.

Lemma 1. $(X_n)_{n\in\mathbb{N}}$ u.i. iff

$$\sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|) < \infty \tag{1}$$

and

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, A \in \mathfrak{A}: \quad P(A) < \delta\ \Rightarrow\ \sup_{n\in\mathbb{N}} \int_A |X_n|\, dP < \varepsilon. \tag{2}$$

Proof. '⇒': For (1), see Remark 1.(i). Moreover,

$$\begin{aligned}\int_A |X_n|\, dP &= \int_{A\cap\{|X_n|\ge\alpha\}} |X_n|\, dP + \int_{A\cap\{|X_n|<\alpha\}} |X_n|\, dP \\ &\le \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP + \alpha \cdot P(A).\end{aligned}$$

For $\varepsilon > 0$ take $\alpha > 0$ with

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP < \varepsilon/2$$

and $\delta = \varepsilon/(2\alpha)$ to obtain (2).
'⇐': Put $M = \sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|)$. Then

$$M \ge \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP \ge \alpha \cdot P(\{|X_n| \ge \alpha\}).$$

Hence $P(\{|X_n| \ge \alpha\}) \le M/\alpha$. Let $\varepsilon > 0$, take $\delta > 0$ according to (2) to obtain for $\alpha > M/\delta$

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP < \varepsilon.$$

□

Theorem 1. Let $1 \leq p < \infty$, and assume $X_n \in \mathfrak{L}^p$ for every $n \in \mathbb{N}$. Then

$$(X_n)_{n\in\mathbb{N}} \text{ converges in } \mathfrak{L}^p$$

iff

$$(X_n)_{n\in\mathbb{N}} \text{ converges in probability } \wedge \ (|X_n|^p)_{n\in\mathbb{N}} \text{ is u.i.}$$

Proof. '$\Rightarrow$': Assume $X_n \xrightarrow{\mathfrak{L}^p} X$. From Remark 2.1 we get $X_n \xrightarrow{P} X$. For every $A \in \mathfrak{A}$

$$\|1_A \cdot X_n\|_p \leq \|1_A \cdot (X_n - X)\|_p + \|1_A \cdot X\|_p.$$

Take $A = \Omega$ to obtain $\sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|^p) < \infty$. Let $\varepsilon > 0$, take $k \in \mathbb{N}$ such that

$$\sup_{n>k} \|X_n - X\|_p < \varepsilon. \tag{3}$$

Put $X_0 = 0$. Note that

$$\sup_{0\leq n\leq k} |X_n - X|^p \leq \sum_{n=0}^{k} |X_n - X|^p \in \mathfrak{L}^1.$$

Hence, by Remark 1.(ii),

$$(|X_1 - X|^p, \dots, |X_k - X|^p, |X|^p, |X|^p, \dots) \text{ u.i.}$$

By Lemma 1

$$P(A) < \delta \quad \Rightarrow \quad \sup_{0\leq n\leq k} \|1_A \cdot (X_n - X)\|_p < \varepsilon.$$

for a suitable $\delta > 0$. Together with (3) this implies

$$P(A) < \delta \quad \Rightarrow \quad \sup_{n\in\mathbb{N}} \|1_A \cdot X_n\|_p < 2 \cdot \varepsilon.$$

'$\Leftarrow$': Let $\varepsilon > 0$, put $A = A_{m,n} = \{|X_m - X_n| > \varepsilon\}$. Then

$$\begin{aligned} \|X_m - X_n\|_p &\leq \|1_A \cdot (X_m - X_n)\|_p + \|1_{A^c} \cdot (X_m - X_n)\|_p \\ &\leq \|1_A \cdot X_m\|_p + \|1_A \cdot X_n\|_p + \varepsilon. \end{aligned}$$

By assumption $X_n \xrightarrow{P} X$ for some $X \in \mathfrak{Z}(\Omega, \mathfrak{A})$. Take $\delta > 0$ according to (2) for $(|X_n|^p)_{n\in\mathbb{N}}$, and note that

$$A_{m,n} \subset \{|X_m - X| > \varepsilon/2\} \cup \{|X_n - X| > \varepsilon/2\}.$$

Hence, for m, n sufficiently large,

$$P(A_{m,n}) < \delta,$$

which implies

$$\|X_m - X_n\|_p \leq 2 \cdot \varepsilon^{1/p} + \varepsilon.$$

Apply Theorem II.6.3. □

Remark 2.

(i) Theorem 1 yields a generalization of Lebesgue's convergence theorem: If $X_n \in \mathfrak{L}^1$ for every $n \in \mathbb{N}$ and $X_n \xrightarrow{P\text{-a.s.}} X$, then

$$(X_n)_{n\in\mathbb{N}} \text{ u.i.} \quad \Leftrightarrow \quad X \in \mathfrak{L}^1 \wedge X_n \xrightarrow{\mathfrak{L}^1} X.$$

(ii) Uniform integrability is a property of the distributions only.

Theorem 2.

$$X_n \xrightarrow{d} X \quad \Rightarrow \quad \mathrm{E}(|X|) \le \liminf_{n\to\infty} \mathrm{E}(|X_n|).$$

Proof. From Skorohod's Theorem 3.4 we get a probability space $(\widetilde{\Omega}, \widetilde{\mathfrak{A}}, \widetilde{P})$ with random variables $\widetilde{X}_n$, $\widetilde{X}$ such that

$$\widetilde{X}_n \xrightarrow{\widetilde{P}\text{-a.s.}} \widetilde{X} \quad \wedge \quad \widetilde{P}_{\widetilde{X}_n} = P_{X_n} \quad \wedge \quad \widetilde{P}_{\widetilde{X}} = P_X.$$

Thus $\mathrm{E}(|X|) = \mathrm{E}(|\widetilde{X}|)$ and $\mathrm{E}(|X_n|) = \mathrm{E}(|\widetilde{X}_n|)$. Apply Fatou's Lemma II.5.2. □

Theorem 3. If

$$X_n \xrightarrow{\mathrm{d}} X \quad \wedge \quad (X_n)_{n\in\mathbb{N}} \text{ u.i.}$$

then

$$X \in \mathfrak{L}^1 \quad \wedge \quad \lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).$$

Proof. Notation as previously. Now $(|\widetilde{X}_n|)_{n\in\mathbb{N}}$ is u.i., see Remark 2.(ii). Hence, by Remark 2.(i), $\widetilde{X} \in \mathfrak{L}^1$ and $\widetilde{X}_n \xrightarrow{\mathfrak{L}^1} \widetilde{X}$. Thus $\mathrm{E}(|X|) < \infty$ and

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \lim_{n\to\infty} \mathrm{E}(\widetilde{X}_n) = \mathrm{E}(\widetilde{X}) = \mathrm{E}(X).$$

□

Example 2. Example 1 continued. With $X = 0$ we have $X_n \xrightarrow{P\text{-a.s.}} X$, and therefore $X_n \xrightarrow{\mathrm{d}} X$. But $\mathrm{E}(X_n) = 1 > 0 = \mathrm{E}(X)$.

5 Independence

'. . . the concept of independence . . . plays a central role in probability theory; it is precisely this concept that distinguishes probability theory from the general theory of measure spaces', see Shiryayev (1984, p. 27).

In the sequel, $(\Omega, \mathfrak{A}, P)$ denotes a probability space and I is a non-empty set.

Definition 1. Let $A_i \in \mathfrak{A}$ for $i \in I$. Then $(A_i)_{i\in I}$ is *independent* if

$$P\Bigl(\bigcap_{i\in S} A_i\Bigr) = \bigtimes_{i\in S} P(A_i) \tag{1}$$

for every $S \in \mathfrak{P}_0(I)$. Elementary case: $|I| = 2$.

In the sequel, $\mathfrak{E}_i \subset \mathfrak{A}$ for $i \in I$.

Definition 2. $(\mathfrak{E}_i)_{i\in I}$ is *independent* if (1) holds for every $S \in \mathfrak{P}_0(I)$ and all $A_i \in \mathfrak{E}_i$ for $i \in S$.

Remark 1.

(i) $(\mathfrak{E}_i)_{i\in I}$ independent $\wedge\ \forall\, i \in I:\ \widetilde{\mathfrak{E}}_i \subset \mathfrak{E}_i \ \Rightarrow\ (\widetilde{\mathfrak{E}}_i)_{i\in I}$ independent.

(ii) $(\mathfrak{E}_i)_{i\in I}$ independent $\ \Leftrightarrow\ \forall\, S \in \mathfrak{P}_0(I):\ (\mathfrak{E}_i)_{i\in S}$ independent.

Lemma 1.

$$(\mathfrak{E}_i)_{i\in I} \text{ independent} \quad \Rightarrow \quad (\delta(\mathfrak{E}_i))_{i\in I} \text{ independent.}$$

Proof. Without loss of generality, $I = \{1, \ldots, n\}$ and $n \geq 2$, see Remark 1.(ii). Put

$$\mathfrak{D}_1 = \{A \in \delta(\mathfrak{E}_1) : (\{A\}, \mathfrak{E}_2, \ldots, \mathfrak{E}_n) \text{ independent}\}.$$

Then $\mathfrak{D}_1$ is a Dynkin class and $\mathfrak{E}_1 \subset \mathfrak{D}_1$, hence $\delta(\mathfrak{E}_1) = \mathfrak{D}_1$. Thus

$$(\delta(\mathfrak{E}_1), \mathfrak{E}_2, \ldots, \mathfrak{E}_n) \text{ independent.}$$

Repeat this step for $2, \ldots, n$. □

Theorem 1. If

$$(\mathfrak{E}_i)_{i\in I} \text{ independent } \wedge\ \forall\, i \in I:\ \mathfrak{E}_i \text{ closed w.r.t. intersections} \tag{2}$$

then

$$(\sigma(\mathfrak{E}_i))_{i\in I} \text{ independent.}$$

Proof. Use Theorem II.1.2 and Lemma 1. □

Corollary 1. Assume that $I = \bigcup_{j\in J} I_j$ for pairwise disjoint sets $I_j \neq \emptyset$. If (2) holds, then

$$\left(\sigma\Big(\bigcup_{i\in I_j} \mathfrak{E}_i\Big)\right)_{j\in J} \text{ independent.}$$

Proof. Let

$$\widetilde{\mathfrak{E}}_j = \Big\{\bigcap_{i\in S} A_i : S \in \mathfrak{P}_0(I_j) \wedge A_i \in \mathfrak{E}_i \text{ for } i \in S\Big\}.$$

Then $\widetilde{\mathfrak{E}}_j$ is closed w.r.t. intersections and $\big(\widetilde{\mathfrak{E}}_j\big)_{j\in J}$ is independent. Finally

$$\sigma\Big(\bigcup_{i\in I_j} \mathfrak{E}_i\Big) = \sigma(\widetilde{\mathfrak{E}}_j).$$

□

In the sequel, $(\Omega_i, \mathfrak{A}_i)$ denotes a measurable space for $i \in I$, and $X_i : \Omega \to \Omega_i$ is $\mathfrak{A}$-$\mathfrak{A}_i$-measurable for $i \in I$.

Definition 3. $(X_i)_{i\in I}$ is *independent* if $(\sigma(X_i))_{i\in I}$ is independent.

Example 1. Actually, the essence of independence. Assume that

$$(\Omega, \mathfrak{A}, P) = \left(\bigtimes_{i\in I} \Omega_i, \bigotimes_{i\in I} \mathfrak{A}_i, \bigtimes_{i\in I} P_i \right)$$

for probability measures P_i on $\mathfrak{A}_i$. Let

$$X_i = \pi_i.$$

Then, for $S \in \mathfrak{P}_0(I)$ and $A_i \in \mathfrak{A}_i$ for $i \in S$

$$P\biggl(\bigcap_{i\in S} \{X_i \in A_i\} \biggr) = P\biggl(\bigtimes_{i\in S} A_i \times \bigtimes_{i\in I\setminus S} \Omega_i \biggr) = \bigtimes_{i\in S} P_i(A_i) = \bigtimes_{i\in S} P(\{X_i \in A_i\}).$$

Hence $(\pi_i)_{i\in I}$ is independent. Furthermore, $P_{X_i} = P_i$.

Recall the question that was posed in the introductory Example I.2.

Theorem 2. Given: probability spaces $(\Omega_i, \mathfrak{A}_i, P_i)$ for $i \in I$. Then there exist

(i) a probability space $(\Omega, \mathfrak{A}, P)$ and

(ii) $\mathfrak{A}$-$\mathfrak{A}_i$-measurable mappings $X_i : \Omega \to \Omega_i$ for $i \in I$

such that

$$(X_i)_{i\in I} \text{ independent} \quad \wedge \quad \forall\, i \in I : \; P_{X_i} = P_i.$$

Proof. See Example 1. □

Theorem 3. Let $\mathfrak{F}_i \subset \mathfrak{A}_i$ for $i \in I$. If

$$\forall\, i \in I : \; \sigma(\mathfrak{F}_i) = \mathfrak{A}_i \;\wedge\; \mathfrak{F}_i \text{ closed w.r.t. intersections}$$

then

$$(X_i)_{i\in I} \text{ independent} \quad \Leftrightarrow \quad (X_i^{-1}(\mathfrak{F}_i))_{i\in I} \text{ independent.}$$

Proof. Recall that $\sigma(X_i) = X_i^{-1}(\mathfrak{A}_i) = \sigma(X_i^{-1}(\mathfrak{F}_i))$. '$\Rightarrow$': See Remark 1.(i). '$\Leftarrow$': Note that $X_i^{-1}(\mathfrak{F}_i)$ is closed w.r.t. intersections. Use Theorem 1. □

Example 2. Independence of a family of random variables X_i, i.e., $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B})$ for $i \in I$. In this case $(X_i)_{i\in I}$ is independent iff

$$\forall\, S \in \mathfrak{P}_0(I)\; \forall\, c_i \in \mathbb{R}, i \in S : P\biggl(\bigcap_{i\in S} \{X_i \le c_i\} \biggr) = \bigtimes_{i\in S} P(\{X_i \le c_i\}).$$

Theorem 4. Let

(i) $I = \bigcup_{j\in J} I_j$ for pairwise disjoint sets $I_j \neq \emptyset$,

(ii) $(\widetilde{\Omega}_j, \widetilde{\mathfrak{A}}_j)$ be measurable spaces for $j \in J$,

(iii) $f_j : \times_{i\in I_j} \Omega_i \to \widetilde{\Omega}_j$ be $\left(\bigotimes_{i\in I_j} \mathfrak{A}_i\right)$-$\widetilde{\mathfrak{A}}_j$ measurable mappings for $j \in J$.

Put

$$Y_j = (X_i)_{i\in I_j} : \Omega \to \bigtimes_{i\in I_j} \Omega_i.$$

Then

$$(X_i)_{i\in I} \text{ independent} \quad \Rightarrow \quad (f_j \circ Y_j)_{j\in J} \text{ independent.}$$

Proof.

$$\begin{aligned}\sigma(f_j \circ Y_j) = Y_j^{-1}(f_j^{-1}(\widetilde{\mathfrak{A}}_j)) &\subset Y_j^{-1}\Bigg(\bigotimes_{i\in I_j} \mathfrak{A}_i\Bigg) \\ &= \sigma(\{X_i : i \in I_j\}) = \sigma\Bigg(\bigcup_{i\in I_j} X_i^{-1}(\mathfrak{A}_i)\Bigg).\end{aligned}$$

Use Corollary 1 and Remark 1.(i). □

Example 3. For an independent sequence $(X_i)_{i\in\mathbb{N}}$ of random variables

$$\Big(\max(X_1, X_2),\ 1_{\mathbb{R}_+}(X_3),\ \limsup_{n\to\infty} 1/n \sum_{i=1}^{n} X_i\Big)$$

are independent.

Remark 2. Consider the mapping

$$X : \Omega \to \bigtimes_{i\in I} \Omega_i : \omega \mapsto (X_i(\omega))_{i\in I}.$$

Clearly X is $\mathfrak{A}$-$\bigotimes_{i\in I} \mathfrak{A}_i$-measurable. By definition, $P_X(A) = P(\{X \in A\})$ for $A \in \bigotimes_{i\in I} \mathfrak{A}_i$. In particular, for measurable rectangles $A \in \bigotimes_{i\in I} \mathfrak{A}_i$, i.e.,

$$A = \bigtimes_{i\in S} A_i \times \bigtimes_{i\in I\setminus S} \Omega_i \tag{3}$$

with $S \in \mathfrak{P}_0(I)$ and $A_i \in \mathfrak{A}_i$,

$$P_X(A) = P\Bigg(\bigcap_{i\in S} \{X_i \in A_i\}\Bigg). \tag{4}$$

Definition 4. P_X is called the *joint distribution* of X_i, $i \in I$.

Example 4. Let $\Omega = \{1, \dots, 6\}^2$ and consider the uniform distribution P on $\mathfrak{A} = \mathfrak{P}(\Omega)$, which is a model for rolling a die twice.
Moreover, let $\Omega_i = \mathbb{N}$ and $\mathfrak{A}_i = \mathfrak{P}(\Omega_i)$ such that $\bigotimes_{i=1}^{2} \mathfrak{A}_i = \mathfrak{P}(\mathbb{N}^2)$. Consider the random variables

$$X_1(\omega_1, \omega_2) = \omega_1, \qquad X_2(\omega_1, \omega_2) = \omega_1 + \omega_2.$$

Then

$$P_X(A) = \frac{|A \cap M|}{36}, \qquad A \subset \mathbb{N}^2,$$

where

$$M = \{(k, \ell) \in \mathbb{N}^2 : 1 \leq k \leq 6 \wedge k+1 \leq \ell \leq k+6\}$$

Claim: (X_1, X_2) are not independent. Proof:

$$P(\{X_1 = 1\} \cap \{X_2 = 3\}) = P_X(\{(1,3)\}) = P(\{(1,2)\}) = 1/36$$

but

$$P(\{X_1 = 1\}) \cdot P(\{X_2 = 3\}) = 1/6 \cdot P(\{(1,2),(2,1)\}) = 1/3 \cdot 1/36.$$

We add that

$$P_{X_1} = \sum_{k=1}^{6} 1/6 \cdot \varepsilon_k, \qquad P_{X_2} = \sum_{\ell=2}^{12} (6 - |\ell - 7|)/36 \cdot \varepsilon_\ell.$$

Theorem 5.

$$(X_i)_{i \in I} \text{ independent} \quad \Leftrightarrow \quad P_X = \mathop{\times}_{i \in I} P_{X_i}.$$

Proof. For A given by (3)

$$\Big(\mathop{\times}_{i \in I} P_{X_i}\Big)(A) = \mathop{\times}_{i \in S} P_{X_i}(A_i) = \mathop{\times}_{i \in S} P(\{X_i \in A_i\}).$$

On the other hand, we have (4). Thus '$\Leftarrow$' hold trivially. Use Theorem II.4.4 to obtain '$\Rightarrow$'. □

In the sequel, we consider random variables X_i, i.e., $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B})$ for $i \in I$.

Theorem 6. Let $I = \{1, \dots, n\}$. If

$$(X_1, \dots, X_n) \text{ independent } \wedge \ \forall\, i \in I: \ X_i \geq 0 \ (X_i \text{ integrable})$$

then ($\times_{i=1}^{n} X_i$ is integrable and)

$$\mathrm{E}\Big(\mathop{\times}_{i=1}^{n} X_i\Big) = \mathop{\times}_{i=1}^{n} \mathrm{E}(X_i).$$

Proof. Use Fubini's Theorem and Theorem 5 to obtain

$$\begin{aligned}
\mathrm{E}\Big(\Big|\mathop{\times}_{i=1}^{n} X_i\Big|\Big) &= \int_{\mathbb{R}^n} |x_1 \cdot \dots \cdot x_n| \, P_{(X_1,\dots,X_n)}(d(x_1,\dots,x_n)) \\
&= \int_{\mathbb{R}^n} |x_1 \cdot \dots \cdot x_n| \, (P_{X_1} \times \dots \times P_{X_n})(d(x_1,\dots,x_n)) \\
&= \mathop{\times}_{i=1}^{n} \int_{\mathbb{R}} |x_i| \, P_{X_i}(dx_i) = \mathop{\times}_{i=1}^{n} \mathrm{E}(|X_i|).
\end{aligned}$$

Drop $|\cdot|$ if the random variables are integrable. □

Definition 5. $X_1, X_2 \in \mathfrak{L}^2$ are *uncorrelated* if

$$\mathrm{E}(X_1 \cdot X_2) = \mathrm{E}(X_1) \cdot \mathrm{E}(X_2).$$

Theorem 7 (Bienaymé)**.** Let $X_1, \dots, X_n \in \mathfrak{L}^2$ be pairwise uncorrelated. Then

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Proof. We have

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) &= \mathrm{E}\left(\sum_{i=1}^n (X_i - \mathrm{E}(X_i))\right)^2 \\ &= \sum_{i=1}^n \mathrm{E}(X_i - \mathrm{E}(X_i))^2 + \sum_{\substack{i,j=1\\ i\neq j}}^n \mathrm{E}((X_i - \mathrm{E}(X_i)) \cdot (X_j - \mathrm{E}(X_j))).\end{aligned}$$

Moreover,

$$\mathrm{E}((X_i - \mathrm{E}(X_i)) \cdot (X_j - \mathrm{E}(X_j))) = \mathrm{E}(X_i \cdot X_j) - \mathrm{E}(X_i) \cdot \mathrm{E}(X_j).$$

(The latter quantity is called the *covariance* between X_i and X_j.) □

Definition 6. The *convolution product* of probability measures $P_1, \dots, P_n$ on $\mathfrak{B}$ is defined by

$$P_1 * \cdots * P_n = s(P_1 \times \cdots \times P_n)$$

where

$$s(x_1, \dots, x_n) = x_1 + \cdots + x_n.$$

Theorem 8. Let $(X_1, \dots, X_n)$ be independent and $S = \sum_{i=1}^n X_i$. Then

$$P_S = P_{X_1} * \cdots * P_{X_n}.$$

Proof. Put $X = (X_1, \dots, X_n)$. Since $S = s \circ (X_1, \dots, X_n)$ we get

$$P_S = s(P_X) = s(P_{X_1} \times \cdots \times P_{X_n}).$$

□

Remark 3. The class of probability measure on $\mathfrak{B}$ forms an abelian semi-group w.r.t. $*$, and $P * \varepsilon_0 = P$.

Theorem 9. For all probability measures P_1, P_2 on $\mathfrak{B}$ and every $P_1 * P_2$-integrable function f

$$\int_{\mathbb{R}} f \, d(P_1 * P_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x+y) \, P_1(dx) \, P_2(dy).$$

If $P_1 = h_1 \cdot \lambda_1$ then $P_1 * P_2 = h \cdot \lambda_1$ with

$$h(x) = \int_{\mathbb{R}} h_1(x-y) \, P_2(dy).$$

If $P_2 = h_2 \cdot \lambda_1$, additionally, then

$$h(x) = \int_{\mathbb{R}} h_1(x-y) \cdot h_2(y) \, \lambda(dy).$$

Proof. Use Fubini's Theorem and the transformation theorem. See Billingsley (1979, p. 230). □

Example 5.

(i) Put $N(\mu, 0) = \varepsilon_\mu$. By Theorem 9

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

for $\mu_i \in \mathbb{R}$ and $\sigma_i \geq 0$.

(ii) Consider n independent Bernoulli trials, i.e., $(X_1, \dots, X_n)$ independent with

$$P_{X_i} = p \cdot \varepsilon_1 + (1-p) \cdot \varepsilon_0$$

for every $i \in \{1, \dots, n\}$, where $p \in [0,1]$. Inductively, we get for $k \in \{1, \dots, n\}$

$$\sum_{i=1}^{k} X_i \sim B(k, p),$$

see also Übung 7.3. Thus, for any $n, m \in \mathbb{N}$,

$$B(n,p) * B(m,p) = B(n+m, p).$$