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

$$P\left(\bigcap_{i\in S} A_i\right) = \prod_{i\in S} P(A_i) \tag{1}$$

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, $I = \{1, \dots, n\}$ and $n \geq 2$, see Remark 1.(ii). Put

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

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

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

Repeat this step for $2, \dots, n$. □

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.(i) 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

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

Proof. Let
$$\widetilde{\mathfrak{E}}_j = \left\{ \bigcap_{i \in S} A_i : S \in \mathfrak{P}_0(I_j) \wedge A_i \in \mathfrak{E}_i \text{ for } i \in S \right\}.$$
Then $\widetilde{\mathfrak{E}}_j$ is closed w.r.t. intersections and $\left(\widetilde{\mathfrak{E}}_j\right)_{j \in J}$ is independent. Finally
$$\sigma\left(\bigcup_{i \in I_j} \mathfrak{E}_i \right) = \sigma(\widetilde{\mathfrak{E}}_j).$$
□

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
$$(\Omega, \mathfrak{A}, P) = \left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathfrak{A}_i, \prod_{i \in I} P_i \right)$$
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 $i \in S$
$$P\left(\bigcap_{i \in S} \{X_i \in A_i\} \right) = P\left(\prod_{i \in S} A_i \times \prod_{i \in I \setminus S} \Omega_i \right) = \prod_{i \in S} P_i(A_i) = \prod_{i \in S} P(\{X_i \in A_i\}).$$
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

$$\forall\, S \in \mathfrak{P}_0(I)\ \forall\, c_i \in \mathbb{R}, i \in S : P\left(\bigcap_{i\in S}\{X_i \le c_i\}\right) = \prod_{i\in S} P(\{X_i \le c_i\}).$$

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

Put

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

Then

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

Proof.

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

Use Corollary 1 and Remark 1.(i). □

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

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

are independent.

Remark 2. Consider the mapping

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

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.,

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

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

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

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 (X_1, X_2) are not independent. Indeed,

$$P(\{X_1 = 2\} \cap \{X_2 = 2\}) = P(\emptyset) = 0\ ,$$

but

$$P(\{X_1 = 2\}) \cdot P(\{X_2 = 2\}) \neq 0\ .$$

We add that

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

Theorem 5.

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

Proof. For A given by (3)

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

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

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

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

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

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., $(X_1, \ldots, X_n)$ independent with

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

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

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

see also Übung 8.1. Thus, for any $n, m \in \mathbb{N}$,

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