Chapter 6

Independence of random variables

We want to speak of independence of random variables, but all we have is the notion of independence of events. A function $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ is a random variable (measurable function) if $\forall B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{F}$, or equivalently, from a notational perspective $X^{-1}(\mathcal{B}) \subset \mathcal{F}$. In addition, recall from Example 1.1.4 that $X^{-1}(\mathcal{B})$ is a σ-algebra associated with Ω. Hence, we will say that $X^{-1}(\mathcal{B})$ is a sub-σ-algebra of $\mathcal{F}$, as $X^{-1}(\mathcal{B}) \subset \mathcal{F}$. In particular, we will say that $X^{-1}(\mathcal{B})$ is the σ-algebra generated by X and it is common to write $\sigma(X) := X^{-1}(\mathcal{B})$.

Before we proceed with the notion of independence of random variables we establish the following theorem.

Theorem 6.1. *Let $X : (\Omega, \mathcal{F}, P) \to (T, \mathcal{T})$ and $\mathcal{C}$ a class of subsets of T. Then,*

$$X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})).$$

Proof. From Example 1.1.4 $X^{-1}(\sigma(\mathcal{C}))$ is a σ-algebra associated with Ω. Since $\mathcal{C} \subset \sigma(\mathcal{C})$, $X^{-1}(\mathcal{C}) \subset X^{-1}(\sigma(\mathcal{C}))$ and consequently $\sigma(X^{-1}(\mathcal{C})) \subset X^{-1}(\sigma(\mathcal{C}))$.

Now, as in Theorem 3.1, $\mathcal{U} = \{U \in 2^T : X^{-1}(U) \in \sigma(X^{-1}(\mathcal{C}))\}$ is a σ-algebra. By definition of $\mathcal{U}$

$$X^{-1}(\mathcal{U}) \subset \sigma(X^{-1}(\mathcal{C})).$$

Also, $\mathcal{C} \subset \mathcal{U}$ since $X^{-1}(\mathcal{C}) \subset X^{-1}(\mathcal{U}) \subset \sigma(X^{-1}(\mathcal{C}))$. Since $\mathcal{U}$ is a σ-algebra we have that

$\sigma(\mathcal{C}) \subset \mathcal{U}$. So,

$$X^{-1}(\sigma(\mathcal{C})) \subset X^{-1}(\mathcal{U}) \subset \sigma(X^{-1}(\mathcal{C})).$$

The last set containment combined with the reverse obtained on the last paragraph completes the proof. ■

Earlier, we defined a finite collection of events $\{E_i\}_{i=1}^n$ and $n \geq 2$ as being independent if

$$P\left(\underset{j\in J}{\cap} E_j\right) = \prod_{j\in J} P(E_j) \text{ for any } J \subset I = \{1, \cdots, n\}. \tag{6.1}$$

We will use this definition to speak of independence of sub-σ-algebras and associated random variables.

Definition 6.1. *Let $n \in \mathbb{N}$, $n \geq 2$ and $\{\mathcal{C}_i\}_{i=1}^n$ be a collection of classes of events. That is, each $\mathcal{C}_i$ contains events associated with the probability space $(\Omega, \mathcal{F}, P)$. The collection $\{\mathcal{C}_i\}_{i=1}^n$ is said to be independent if for any $E_i \in \mathcal{C}_i$ we have that $\{E_i\}_{i=1}^n$ is an independent collection of events.*

This definition motivates the following:

Definition 6.2. *(Independence of σ-algebras). Let $I = \{1, \cdots, n\}$, $n \in \mathbb{N}$ with $n \geq 2$ and $(\Omega, \mathcal{F}, P)$ be a probability space. Then,*

(a) Sub-σ-algebras $\mathcal{F}_i$ of $\mathcal{F}$ with $i \in I$ are independent if for every $J \subset I$ and all $E_i \in \mathcal{F}_i$

$$P\left(\underset{j\in J}{\cap} E_j\right) = \prod_{j\in J} P(E_j),$$

(b) Random variables $X_i : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ for $i \in I$ are independent if the sub σ-algebras $X_i^{-1}(\mathcal{B})$ are independent.

Remark 6.1. *Recall that by definition $X^{-1}(B) = \{\omega : X(\omega) \in B\}$. Hence, when we write $P(X \in B)$ we mean $P(X^{-1}(B))$, for $B \in \mathcal{B}$.*

The following theorem provides a criterion for establishing the independence of σ-algebras.

Theorem 6.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. For each $i \in I = \{1, \cdots, n\}$ let $\mathcal{C}_i$ be a non-empty collection of events satisfying*

1. *$\mathcal{C}_i$ is a π-system,*

2. *$\{\mathcal{C}_i\}_{i \in I}$ is an independent collection.*

Then, $\{\sigma(\mathcal{C}_i)\}_{i \in I}$ is an independent collection.

Proof. First, let $n = 2$. In this case we need to consider $\mathcal{C}_1$ and $\mathcal{C}_2$. Choose an arbitrary $A_2 \in \mathcal{C}_2$ and let $L = \{A \in \mathcal{F} : P(A \cap A_2) = P(A)P(A_2)\}$. L is the collection of P-measurable sets (events) that are independent of A_2. Now, note that:

1. $P(\Omega \cap A_2) = P(A_2) = P(\Omega)P(A_2)$ since $P(\Omega) = 1$. Thus, $\Omega \in L$.

2. Suppose $A \in L$. Note that

$$\begin{aligned} P(A^c \cap A_2) &= P((\Omega - A) \cap A_2)) = P(A_2 - (A \cap A_2)) = P(A_2) - P(A \cap A_2) \\ &= P(A_2) - P(A)P(A_2) \text{ since } A \in L \\ &= P(A_2)(1 - P(A)) = P(A_2)P(A^c). \end{aligned}$$

   Thus, if $A \in L$ we have that $A^c \in L$.

3. If $\{A_n\}_{n=1,\cdots} \in L$ is a pairwise disjoint collection

$$\begin{aligned} P(\cup_{n=1}^{\infty} A_n \cap A_2) &= P(\cup_{n=1}^{\infty}(A_n \cap A_2)) \\ &= \sum_{n=1}^{\infty} P(A_n \cap A_2) \text{ since the sets in the union are disjoint} \\ &= \sum_{n=1}^{\infty} P(A_n)P(A_2) = P(A_2)P(\cup_{n=1}^{\infty} A_n) \text{ since the sets } A_n \text{ are in } L. \end{aligned}$$

   Thus, if $\{A_n\}_{n=1,\cdots} \in L$ is a pairwise disjoint collection, we have that $\cup_{n=1}^{\infty} A_n \in L$.

Since 1-3 are the defining properties of a Dynkin system, we conclude that L is a Dynkin system. Note also that, by assumption, $\mathcal{C}_1$ is independent of $\mathcal{C}_2$, every $A_1 \in \mathcal{C}_1$ is in L. Thus, $\mathcal{C}_1 \subset L$. By Theorem 2.3, since $\mathcal{C}_1$ is a π-system $L \supseteq \delta(\mathcal{C}_1) = \sigma(\mathcal{C}_1)$.[1] Thus, all the events in $\sigma(\mathcal{C}_1)$ are in L and we can conclude that $\sigma(\mathcal{C}_1)$ is independent of $\mathcal{C}_2$. We can also conclude, by the symmetry of the argument, that $\sigma(\mathcal{C}_2)$ is independent of $\mathcal{C}_1$.

Now, repeat the argument above by choosing an arbitrary $A_2 \in \sigma(\mathcal{C}_2)$. Then, L is a Dynkin system, and by the fact that $\sigma(\mathcal{C}_2)$ is independent of $\mathcal{C}_1$ we have that $\mathcal{C}_1 \subset L$ and, as above, $\sigma(\mathcal{C}_1) \subset L$. Consequently, $\sigma(\mathcal{C}_1)$ is independent of $\sigma(\mathcal{C}_2)$. Finally, use induction to establish that this is true for all n finite. ■

Definition 6.3. *Let $\mathcal{I}$ be an arbitrary index set (not necessarily finite or even countable). The collection $\{\mathcal{C}_i\}_{i\in\mathcal{I}}$ is independent if for each finite $I \subset \mathcal{I}$, the collection $\{\mathcal{C}_i\}_{i\in I}$ is independent.*

The following is a corollary to Theorem 6.2.

Corollary 6.1. *Let $\{\mathcal{C}_i\}_{i\in\mathcal{I}}$ be a collection of non-empty independent π-systems. Then, $\{\sigma(\mathcal{C}_i)\}_{i\in\mathcal{I}}$ is an independent collection.*

Definition 6.2 can be naturally expanded in accordance to Definition 6.3 to accommodate an arbitrarily indexed collection of random variables. We now provide some characterizations for independence of random variables.

Definition 6.4. *Let $\{X_i\}_{i\in\mathcal{I}}$ be a collection of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. The finite dimensional distribution functions (fddf) are the collection*

$$F_I(x_i,\, i \in I) = P(\{\omega : X_i(\omega) \le x_i,\, i \in I\}) \text{ for all } x_i \in \mathbb{R}. \tag{6.2}$$

for all finite subsets $I \subset \mathcal{I}$.

[1] $\delta(\mathcal{C}_1)$ is the smallest Dynkin system generated by $\mathcal{C}_1$.

Theorem 6.3. *The collection* $\{X_i\}_{i\in\mathcal{I}}$ *of random variables defined on the probability space* $(\Omega, \mathcal{F}, P)$ *is independent if, and only if, for all finite subsets* $I \subset \mathcal{I}$ *we have*

$$F_I(x_i,\, i \in I) = \prod_{i\in I} P(\{\omega : X_i(\omega) \le x_i\}) \text{ for all } x_i \in \mathbb{R}. \tag{6.3}$$

Proof. From definition 6.3 it suffices to show that for an arbitrary finite $I \subset \mathcal{I}$ the collection $\{X_i\}_{i\in I}$ is independent if, and only if, equation (6.3) holds.

($\Longleftarrow$) Let $\mathcal{C}_i = \{\{\omega : X_i(\omega) \le x\},\, x \in \mathbb{R}\} := \{X_i^{-1}((-\infty, x]),\, x \in \mathbb{R}\}$ and note that these are subsets of Ω. Furthermore,

1. $\mathcal{C}_i$ is a π-system since

$$\{\omega : X_i(\omega) \le x\} \cap \{\omega : X_i(\omega) \le y\} = \{\omega : X_i(\omega) \le \min\{x, y\}\}.$$

2. Recall that $\sigma(\{(-\infty, x],\, x \in \mathbb{R}\}) = \mathcal{B}$. By Theorem 3.1 X_i is a random variable ($X_i^{-1}(\mathcal{B}) \subset \mathcal{F}$) if, and only if,

$$\{X_i^{-1}((-\infty, x]),\, x \in \mathbb{R}\} = \mathcal{C}_i \subset \mathcal{F}.$$

Hence,

$$\begin{aligned} \sigma(\mathcal{C}_i) &= \sigma(\{X_i^{-1}((-\infty, x]),\, x \in \mathbb{R}\}) \\ &= X_i^{-1}(\sigma(\{(-\infty, x],\, x \in \mathbb{R}\})) \text{ by Theorem 6.1} \\ &= X_i^{-1}(\mathcal{B}) = \sigma(X_i). \end{aligned}$$

Now, equation (6.3) implies that $\{\mathcal{C}_i\}_{i\in I}$ is independent collection and, therefore, by Theorem 6.2, the collection $\{\sigma(\mathcal{C}_i) = \sigma(X_i)\}_{i\in I}$ is independent. Consequently, by definition, $\{X_i\}_{i\in I}$ are forms an independent collection of random variables.

($\Longrightarrow$) This follows directly from the definition of independence. ■

Remark 6.2. *1. It follows directly from Theorem 6.3 that a finite collection of random variables $\{X_i\}_{i=1}^m$ is independent if, and only if,*

$$P(\cap_{i\in J}\{\omega : X_i(\omega) \le x_i\}) = \prod_{i\in J} P(\{\omega : X_i(\omega) \le x_i\}), \text{ for all } J \subset \{1, \cdots, m\}.$$

2. If X_i has a density $\{X_i\}_{i=1}^m$ are independent if, and only if,

$$P(\cap_{i\in J}\{\omega : X_i(\omega) \le x_i\}) = \prod_{i\in J} \int_{(-\infty,x_i]} f_{X_i} d\lambda.$$

6.1 Random elements

The most common cases where we deal with random elements occur when the co-domain of the element is endowed with a metric, so that the co-domain is a metric space.

Definition 6.5. *Let $X : (\Omega, \mathcal{F}, P) \to (T, \mathcal{T} = \sigma(\mathcal{O}))$, where $\mathcal{O}$ are the open sets in T. Then, X is a random element if*

$$X^{-1}(B) \in \mathcal{F} \text{ for all } B \in \mathcal{T}.$$

In this definition, $\mathcal{T}$ is the collection of Borel sets of T and we write $\mathcal{B}(T)$. The following examples include definitions.

Example 6.1. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ where $k \in \mathbb{N}$. Then X is a random vector if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^k)$. Now, define $d_E : \mathbb{R}^k \times \mathbb{R}^k \to [0, \infty)$ as $d_E(x, y) = \left(\sum_{i=1}^k (x_i - y_i)^2\right)^{1/2}$. It can be easily verified that d_E is a metric on $\mathbb{R}^k$.*

Example 6.2. *Let $m : \mathbb{R}^2 \to \mathbb{R}$ be given by $m(x_1, x_2) = \frac{|x_1-x_2|}{1+|x_1-x_2|}$. Clearly, from the definition of m, $m \ge 0$, $m = 0$ if, and only if, $x_1 = x_2$ and $m(x_1, x_2) = m(x_2, x_1)$. To verify that $m(x_1, x_2) \le m(x_1, z) + m(z, x_2)$ we note that $|x_1 - x_2| = \frac{m(x_1,x_2)}{1-m(x_1,x_2)}$. Since $|x_1 - x_2| = |x_1 - z + z - x_2| \le |x_1 - z| + |z - x_2|$, we have*

$$\frac{m(x_1, x_2)}{1 - m(x_1, x_2)} \le \frac{m(x_1, z)}{1 - m(x_1, z)} + \frac{m(z, x_2)}{1 - m(z, x_2)}.$$

Let $c = m(x_1, x_2)$, $a = m(x_1, z)$ *and* $b = m(z, x_2)$. *Then,* $\frac{c}{1-c} \leq \frac{a}{1-a} + \frac{b}{1-b} = \frac{a+b-2ab}{(1-a)(1-b)}$ *and*

$$a + b \geq \frac{c}{1-c}(1-a)(1-b) + 2ab = -\frac{c}{1-c}(a+b) + \frac{c}{1-c} + \frac{1}{1-c}(2ab - abc).$$

Then,

$$\frac{a+b}{1-c} \geq \frac{c}{1-c} + \frac{1}{1-c}(2ab - abc) \iff a + b \geq c + ab(2-c).$$

Since $0 \leq m \leq 1$, $ab(2-c) \geq 0$ *and* $c \leq a+b$. *Hence,* $m(x_1, x_2) \leq m(x_1, z) + m(z, x_2)$. *This shows that* m *is a metric on* $\mathbb{R}$.

Now, consider a space of sequences $\{x_i\}_{i\in\mathbb{N}}$ *where* $x_i \in \mathbb{R}$ *for all* i *and define* $m_\infty : \mathbb{R}^\infty \times \mathbb{R}^\infty \to \mathbb{R}$ *as* $m_\infty(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}) = \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{2^j} m(x_j, y_j) = \lim_{n\to\infty} S_n$. *Since* $0 \leq S_1 \leq S_2 \leq \cdots$ *is a monotonic sequence, it converges if, and only if, it is bounded. Boundedness follows from the fact that* $|S_n| \leq \sum_{j=1}^{n} \frac{1}{2^j} m(x_j, y_j) \leq \sum_{j=1}^{n} \frac{1}{2^j} \leq \sum_{j=1}^{\infty} \frac{1}{2^j} = 1$. *Hence, the limit in the definition of* m_∞ *exists and* $0 \leq m_\infty \leq 1$. *If* $m_\infty(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}) = 0$ *then it must be that* $m(x_j, y_j) = 0$ *for all* j, *which implies* $x_j = y_j$ *for all* j. *Clearly, if* $x_j = y_j$ *for all* j *we have* $m_\infty(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}) = 0$.

Since $m(x_j, y_j) \leq m(x_j, z_j) + m(z_j, y_j)$ *we have*

$$\sum_{j=1}^{n} 2^{-j} m(x_j, y_j) \leq \sum_{j=1}^{n} 2^{-j} m(x_j, z_j) + \sum_{j=1}^{n} 2^{-j} m(z_j, y_j).$$

Taking limits on both sides as $n \to \infty$ *gives* $m_\infty(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}) \leq m_\infty(\{x_i\}_{i\in\mathbb{N}}, \{z_i\}_{i\in\mathbb{N}}) + m_\infty(\{z_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}})$. *Hence,* m_∞ *is a metric in the space of infinite sequences.*

Alternatively, we can define $\mu_\infty : \mathbb{R}^\infty \times \mathbb{R}^\infty \to \mathbb{R}$ *as*

$$\begin{aligned}\mu_\infty(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}) &= \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{2^j} \frac{\sum_{i=1}^{j} |x_i - y_i|}{1 + \sum_{i=1}^{j} |x_i - y_i|} = \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{2^j} u_j(\{x_i\}_{i=1}^{j}, \{y_i\}_{i=1}^{j}) \\ &= \lim_{n\to\infty} S_n.\end{aligned}$$

As in the case of m_∞, $0 \leq S_1 \leq S_2 \leq \cdots$ *and* $|S_n| \leq 1$. *Hence,* $0 \leq \mu_\infty \leq 1$ *and* $\mu_\infty = 0$ *if, and only if,* $x_i = y_i$ *for all* i. *Now, write* $S_1(x_1, y_1) = \frac{1}{2} m(x_1, y_1)$ *and since*

$m(x_1, y_1) \leq m(x_1, z_1) + m(z_1, y_1)$ we have that $S_1(x_1, y_1) \leq S_1(x_1, z_1) + S_1(z_1, y_1)$. Now, suppose

$$S_n\left(\{x_i\}_{i=1}^n, \{y_i\}_{i=1}^n\right) \leq S_n(\{x_i\}_{i=1}^n, \{z_i\}_{i=1}^n) + S_n(\{z_i\}_{i=1}^n, \{y_i\}_{i=1}^n).$$

Then,

$$\begin{aligned} S_{n+1}\left(\{x_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}\right) &= \sum_{j=1}^{n+1} \frac{1}{2^j} u_j(\{x_i\}_{i=1}^{j}, \{y_i\}_{i=1}^{j}) = S_n\left(\{x_i\}_{i=1}^n, \{y_i\}_{i=1}^n\right) \\ &+ \frac{1}{2^{n+1}} u_{n+1}(\{x_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}) \\ &\leq S_n\left(\{x_i\}_{i=1}^n, \{z_i\}_{i=1}^n\right) + S_n\left(\{z_i\}_{i=1}^n, \{y_i\}_{i=1}^n\right) \\ &+ \frac{1}{2^{n+1}} u_{n+1}(\{x_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}) \end{aligned}$$

Following the same arguments used for m, we have

$$\begin{aligned} \frac{1}{2^{n+1}} u_{n+1}(\{x_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}) &\leq \frac{1}{2^{n+1}} u_{n+1}(\{x_i\}_{i=1}^{n+1}, \{z_i\}_{i=1}^{n+1}) + \frac{1}{2^{n+1}} u_{n+1}(\{z_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}) \\ &= \frac{1}{2^{n+1}} \frac{\sum_{i=1}^{n+1} |x_i - z_i|}{1 + \sum_{i=1}^{n+1} |x_i - z_i|} + \frac{1}{2^{n+1}} \frac{\sum_{i=1}^{n+1} |z_i - y_i|}{1 + \sum_{i=1}^{n+1} |z_i - y_i|}. \end{aligned}$$

Hence,

$$\begin{aligned} S_{n+1}\left(\{x_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}\right) &\leq S_n\left(\{x_i\}_{i=1}^n, \{z_i\}_{i=1}^n\right) + S_n\left(\{z_i\}_{i=1}^n, \{y_i\}_{i=1}^n\right) \\ &+ \frac{1}{2^{n+1}} \frac{\sum_{i=1}^{n+1} |x_i - z_i|}{1 + \sum_{i=1}^{n+1} |x_i - z_i|} + \frac{1}{2^{n+1}} \frac{\sum_{i=1}^{n+1} |z_i - y_i|}{1 + \sum_{i=1}^{n+1} |z_i - y_i|} \\ &= S_{n+1}(\{x_i\}_{i=1}^{n+1}, \{z_i\}_{i=1}^{n+1}) + S_{n+1}(\{z_i\}_{i=1}^{n+1}, \{y_i\}_{i=1}^{n+1}). \end{aligned}$$

Hence, by induction, and taking limits we have $\mu_\infty\left(\{x_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}\right) \leq \mu_\infty\left(\{x_i\}_{i\in\mathbb{N}}, \{z_i\}_{i\in\mathbb{N}}\right) + \mu_\infty\left(\{z_i\}_{i\in\mathbb{N}}, \{y_i\}_{i\in\mathbb{N}}\right)$.

Example 6.3. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ where $\mathbb{R}^\infty = \times_{n=1}^\infty \mathbb{R}$ and $\mathcal{B}(\mathbb{R}^\infty) = \sigma(\mathcal{C})$ with $\mathcal{C} = \{C : C = \theta_i^{-1}(B),\ B \in \mathcal{B}^i,\ \theta_i(x) = (X_1, \cdots, X_i) : \mathbb{R}^\infty \to \mathbb{R}^i,\ i \in \mathbb{N}\}$. Then X is a random sequence if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R}^\infty)$ and $d : \mathbb{R}^\infty \times \mathbb{R}^\infty \to [0, \infty)$ is $d(x, y) = \sum_{i=1}^\infty \frac{1}{2^i} \left(\frac{\sum_{j=1}^i |x_j - y_j|}{1 + \sum_{j=1}^i |x_j - y_j|} \right)^{1/2}$ is the metric on $\mathbb{R}^\infty$.*

Remark 6.3. *1. Let $X \in \mathbb{R}^k$ be a random vector and $f : \mathbb{R}^k \to \mathbb{R}$ be measurable. Then, $h : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ with $h(\omega) = f(X(\omega)) = (f \circ X)(\omega)$ is a random variable since compositions of measurable functions are measurable by Theorem 3.3. In particular the result follows if f is continuous. That is, real valued continuous functions of random vectors are random variables.*

2. In 1, if $f(X) = \pi_i(X) = X_i$ and X is random vector then X_i is a random variable for $i = 1, \cdots, k$.

Theorem 6.4. *$X \in \mathbb{R}^k$ is a random vector $\iff$ X_i is a random variable, where X_i is the ith component of X.*

Proof. ($\Longleftarrow$) Suppose X_i is a random variable for $i = 1, \cdots, k$. Let $R_k = I_1 \times \cdots \times I_k$, where $I_i = [a_i, b_i)$ are intervals in $\mathbb{R}$. Then,

$$\begin{aligned} X^{-1}(R_k) &= \{\omega : X_i(\omega) \in [a_i, b_i) \,\forall\, i\} \\ &= \{\omega : X_i^{-1}([a_i, b_i)) \,\forall\, i\} = \cap_{i=1}^k X_i^{-1}(I_i). \end{aligned}$$

Since X_i is a random variable, $X_i^{-1}(I_i) \in \mathcal{F}$. Furthermore, since $\mathcal{F}$ is a σ-algebra, it is closed under intersections, and $X^{-1}(R_k) \in \mathcal{F}$. The other direction of the equivalence follows from the previous remark. ■

Remark 6.4. *1. Theorem 6.4 extends to $X = \{X_1, X_2, \cdots\}$. That is, X is a random sequence if, and only if, each X_i is a random variable. Furthermore, X is a random sequence if, and only if, $(X_1 \cdots X_k)$ is random vector for any k.*
2. $X^{-1}((-\infty, a_1] \times \cdots \times (-\infty, a_k]) \in \mathcal{F}$ and we write $P(X^{-1}((-\infty, a_1] \times \cdots \times (-\infty, a_k])) = P \circ X^{-1}(\times_{i=1}^k(-\infty, a_i]) = P_X(\times_{i=1}^k(-\infty, a_i])$.
Also, if there exists a non-negative Borel measurable function $f_X : \mathbb{R}^k \to \mathbb{R}$ that satisfies

$$P_X(\times_{i=1}^k(-\infty, a_i]) = \int_{C(a)} f_X d\lambda^k,$$

where $C(a) = \times_{i=1}^{k}(-\infty, a_i]$ *and* $a = (a_1 \cdots a_k)^T$*, we call* f_X *the "joint density" of* X*. Naturally, the joint distribution function associated with* X *is*

$$F_X(a) : \mathbb{R}^k \to [0, 1],$$

where $F_X(a) = P(C(a))$ *for* $a \in \mathbb{R}^k$*. We can write* $C(a) = \cap_{i=1}^{k}\{\omega : X_i(\omega) \leq a_i\}$*. That* $\{\omega : X_i(\omega) \leq a_i\}$ *is an element of* $\mathcal{F}$ *follows from Theorem 6.4.*

Theorem 6.5. *Consider two random variables* $X_1, X_2 : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$*.* X_1 *and* X_2 *are independent if, and only if, one of the following holds:*

a) $P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) := P(X \in A_1, X \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)$*, for all* A_1*,* $A_2 \in \mathcal{B}$*,*

b) $P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)$*, for all* $A_1 \in \mathcal{A}_1$*,* $A_2 \in \mathcal{A}_2$*, where* $\mathcal{A}_1, \mathcal{A}_2$ *are* π *systems which generate* $\mathcal{B}$*,*

c) $f(X_1)$ *and* $g(X_2)$ *are independent for each pair* (f, g) *of measurable functions,*

d) $E(f(X_1), g(X_2)) = E(f(X_1))E(g(X_2))$ *for each pair of* (f, g) *of bounded measurable (or non-negative measurable) functions.*

Proof. First, note that X_1 and X_2 independent means that $\sigma(X_1) = X_1^{-1}(\mathcal{B})$ and $\sigma(X_2) = X_2^{-1}(\mathcal{B})$ are independent. That is, for all $A_1, A_2 \in \mathcal{B}$,

$$\begin{aligned} P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) &= P(X_1^{-1}(A_1))P(X_2^{-1}(A_2)) \\ &\iff P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2). \end{aligned}$$

$[a) \implies b)]$ Since $\mathcal{A}_1$ generates $\mathcal{B}$ and $\mathcal{A}_2$ generates $\mathcal{B}$, $\mathcal{A}_1 \subset \mathcal{B}$ and $\mathcal{A}_2 \subset \mathcal{B}$, and if a) is true for all $A_1 \in \mathcal{B}$, $A_2 \in \mathcal{B}$, then b) is true.

$[b) \implies a)]$ Let $C_1 = \{A \in \mathcal{B} : P(X_1 \in A, X_2 \in A_2) = P(X_1 \in A)P(X_2 \in A_2)$ for a given $A_2 \in \mathcal{A}_2\}$. From the proof of Theorem 6.2, C_1 is a Dynkin system. $\mathcal{A}_1 \subset C_1$ and

$\delta(\mathcal{A}_1) = \sigma(\mathcal{A}_1) = \mathcal{B} \subset C_1$. Analogously, $C_2 = \{A \in \mathcal{B}_2 : P(X_1 \in A_1, X_2 \in A) = P(X_1 \in A_1)P(X_2 \in A)$ for a given $A_1 \in \mathcal{A}_1\}$ is such that $\delta(\mathcal{A}_2) = \sigma(\mathcal{A}_2) = \mathcal{B} \subset C_2$. Consequently, $b) \implies a)$.

$[c) \implies a)]$ The identity function is measurable, therefore take $f(x) = g(x) = x$

$[a) \implies c)]$ For concreteness, let $f : (\mathbb{R}, \mathcal{B}) \to (M_f, \mathcal{M}_f)$ and $g : (\mathbb{R}, \mathcal{B}) \to (M_g, \mathcal{M}_g)$. f measurable implies that for all $M \in \mathcal{M}_f$, $f^{-1}(M) \in \mathcal{B}$. But X_1 a random variable implies that $X_1^{-1}(f^{-1}(M)) \in \mathcal{F}$ which we can write as $(X_1^{-1} \circ f^{-1})(M) \in \mathcal{F}$. In addition, $X_1^{-1}(f^{-1}(M)) := (X_1^{-1} \circ f^{-1})(M) \in X_1^{-1}(\mathcal{B})$. Analogously, $X_2^{-1}(g^{-1}(M')) = X_2^{-1} \circ g^{-1}(M') \in X_2^{-1}(\mathcal{B})$, for all $M' \in \mathcal{M}_g$. But by a) $X_1^{-1}(\mathcal{B})$ and $X_2^{-1}(\mathcal{B})$ are independent. Therefore $f(X_1)$ and $g(X_2)$ are independent.

$[d) \implies a)]$ Let $f = I_{A_1}$ and $g = I_{A_2}$. Then,

$$f(X_1) = \begin{cases} 1 & \text{if } X_1 \in A_1 \\ 0 & \text{if } X_1 \notin A_1 \end{cases} \text{ and } g(X_2) = \begin{cases} 1 & \text{if } X_2 \in A_2 \\ 0 & \text{if } X_2 \notin A_2. \end{cases}$$

with $E(f(X_1)) = P(X_1 \in A_1)$ and $E(g(X_2)) = P(X_2 \in A_2)$. By d)

$$E(f(X_1)g(X_2)) = P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = P(X_1 \in A_1)P(X_2 \in A_2).$$

Hence, $d) \implies a)$.

$[a) \implies d)]$ From the implication $[d) \implies a)]$ we see that if f, g are indicator functions in d) $E(f(X_1)g(X_2)) = P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\})$, which by independence a) is $P(X_1 \in A_1)P(X_2 \in A_2) = E(f(X_1))E(g(X_2))$.

Now, suppose f and g are simple functions of X_1 and X_2. Then,

$$\begin{aligned} f(X_1) &= \sum_{i=0}^{k_f} a_i^f I_{\{X_1 \in A_i^f)\}} \text{ and } E(f(X_1)) = \sum_{i=0}^{k_f} a_i^f P(X_1 \in A_i^f), \\ g(X_2) &= \sum_{i=0}^{k_g} a_i^g I_{\{X_2 \in A_i^g)\}} \text{ and } E(g(X_2)) = \sum_{i=0}^{k_g} a_i^g P(X_2 \in A_i^g) \end{aligned}$$

Consequently,

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E\left(\sum_{i=0}^{k_f}\sum_{j=0}^{k_g} a_i^f a_j^g I_{\{X_1\in A_i^f\}\cap\{X_2\in A_j^g\}}\right)\\
&= \sum_{i=0}^{k_f}\sum_{j=0}^{k_g} a_i^f a_j^g P(X_1\in A_i^f)P(X_2\in A_j^g) \text{ by independence}\\
&= E(f(X_1))E(g(X_2)) \qquad (6.4)
\end{aligned}$$

Now, let f be a measurable non-negative function such that $\{f_n\}_{n\in\mathbb{N}}$ are simple functions increasing to f and g is non-negative and simple. Then,

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E\left(\lim_{n\to\infty} f_n(X_1)g(X_2)\right)\\
&= \lim_{n\to\infty} E(f_n(X_1)g(X_2)) \text{ by Lebesgue's Monotone Convergence Theorem}\\
&= \lim_{n\to\infty} E(f_n(X_1))E(g(X_2)) \text{ by equation (6.4)}\\
&= E(f(X_1))E(g(X_2)) \text{ by Lebesgue's Monotone Convergence Theorem} \qquad (6.5)
\end{aligned}$$

Now, let f be non-negative and let $\{g_n\}_{n\in\mathbb{N}}$ be non-negative simple functions increasing to g measurable and non-negative. Then,

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E\left(f(X_1)\lim_{n\to\infty} g_n(X_2)\right)\\
&= \lim_{n\to\infty} E(f(X_1)g_n(X_2))\\
&= \lim_{n\to\infty} E(f(X_1))E(g_n(X_2)) \text{ by equation (6.5)}\\
&= E(f(X_1))E(g(X_2))
\end{aligned}$$

Finally, let $f = f^+ - f^-$ be bounded and measurable and g bounded and non-negative.

$$\begin{aligned}
E(f(X_1)g(X_2)) &= E([f^+(X_1) - f^-(X_1)]g(X_2))\\
&= E(f^+(X_1)g(X_2)) - E(f^-(X_1)g(X_2))\\
&= E(f^+(X_1))E(g(X_2)) - E(f^-(X_1))E(g(X_2))\\
&= E(f(X_1))E(g(X_2)).
\end{aligned}$$

To complete the proof, repeat the last argument for $g = g^+ - g^-$. ■