

5. Convergence and Independence

5.1 Convergence

Definition 5.1 (Almost sure convergence). *We say $X_n \rightarrow X$ almost surely (a.s.) or with probability 1 if*

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Recall that in the original MCT in Theorem 4.2, we suppose that $0 \leq X_n(\omega) \leq X_{n+1}(\omega)$ and $X_n(\omega) \rightarrow X(\omega)$ for all $\omega \in \Omega$. Here, we show that we can replace the condition “for all $\omega \in \Omega$ ” with “almost surely”. Let

$$A = \{\omega : X_n(\omega) \rightarrow X(\omega), 0 \leq X_n(\omega) \leq X_{n+1}(\omega), n \geq 1\}$$

and suppose $\mathbb{P}(A) = 1$. We have $0 \leq X_n \mathbf{1}_A(\omega) \leq X_{n+1} \mathbf{1}_A(\omega)$, and $X_n \mathbf{1}_A(\omega) \rightarrow X \mathbf{1}_A(\omega)$ $\forall \omega \in \Omega$. Then from the MCT in Theorem 4.2, we have

$$\mathbb{E}[X_n \mathbf{1}_A] \rightarrow \mathbb{E}[X \mathbf{1}_A].$$

For any $Y \geq 0$, let $Y \wedge n = \min(Y, n)$ for $n \geq 1$. Observe that

$$\begin{aligned} \mathbb{E}[Y \mathbf{1}_{A^c}] &= \mathbb{E}\left[\lim_{n \rightarrow \infty} (Y \wedge n) \mathbf{1}_{A^c}\right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E}[(Y \wedge n) \mathbf{1}_{A^c}] \\ &\leq \lim_{n \rightarrow \infty} n \mathbb{P}(A^c) \\ &= 0, \end{aligned} \tag{1}$$

where we have use the MCT in (1). Therefore, $\mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}X_n$ and $\mathbb{E}[X \mathbf{1}_A] = \mathbb{E}X$, and the MCT holds if “ $\forall \omega \in \Omega$ ” is replaced with “almost surely” in its theorem statement. The same thing applies for Fatou’s Lemma and the DCT.

Definition 5.2 (Convergence in probability). *If $\forall \epsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0,$$

we say that X_n converges to X in probability and write $X_n \xrightarrow{P} X$.

Lemma 5.1. *If $X_n \rightarrow X$ a.s., then $X_n \xrightarrow{P} X$.*

Proof. Suppose $X_n \rightarrow X$ a.s., then $\mathbf{1}_{\{|X_n - X| \geq \epsilon\}} \rightarrow 0$ a.s. From DCT (since probability measure is finite), we have $\mathbb{P}(|X_n - X| \geq \epsilon) \rightarrow 0$. \square

The converse is not true and here is an example.

Example 5.1. *Suppose $\Omega = [0, 1]$. Consider the random variables X_n shown in Fig. 1, where X_n follows a similar pattern for $n \geq 5$. We have $X_n \xrightarrow{P} 0$, but clearly $X_n \not\rightarrow 0$ a.s. as $X_n(\omega) = 1$ for infinitely many values of n for every $\omega \in \Omega$.*

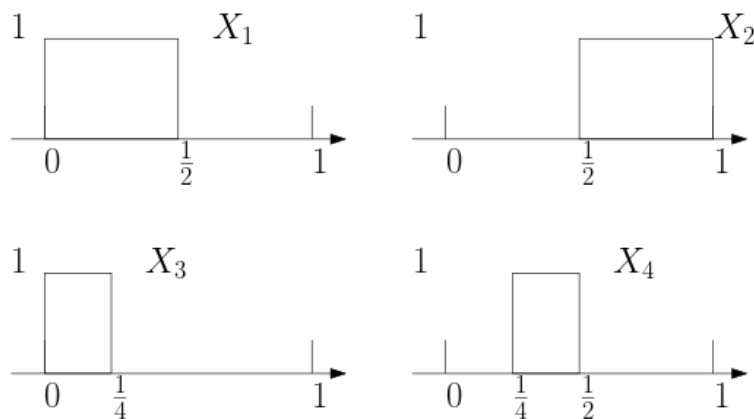


Fig. 1: X_n converges in probability but not almost surely.

5.2 Weak Law of Large Numbers

Markov's inequality: Suppose $a > 0$ and $X \geq 0$. Then,

$$a\mathbf{1}_{\{X \geq a\}}(\omega) \leq X\mathbf{1}_{\{X \geq a\}}(\omega) \leq X.$$

Taking expectations, we have

$$\begin{aligned} a\mathbb{P}(X \geq a) &\leq \mathbb{E}X, \\ \mathbb{P}(X \geq a) &\leq \frac{1}{a}\mathbb{E}X. \end{aligned}$$

Chebyshev's inequality follows from Markov's inequality by replacing X with $|X - \mathbb{E}X|^2$ and setting $a = \epsilon^2$: for $\epsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \epsilon) \leq \frac{1}{\epsilon^2} \text{var}(X).$$

Theorem 5.1 (Weak Law of Large Numbers (WLLN)). *Suppose X_1, X_2, \dots are such that $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = \sigma^2 < \infty$ and $\mathbb{E}[X_i X_j] \leq 0$ for $i \neq j$. Then,*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Proof. We have

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right] &= \frac{1}{n^2} \mathbb{E} \left[\sum_i X_i^2 + 2 \sum_{i < j} X_i X_j \right] \\ &\leq \frac{1}{n^2} \sum_i \mathbb{E} X_i^2 = \frac{\sigma^2}{n}. \end{aligned}$$

From Chebyshev's inequality, we then have for any $\epsilon > 0$,

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq \epsilon \right) \leq \frac{1}{\epsilon^2} \frac{\sigma^2}{n} \rightarrow 0,$$

as $n \rightarrow \infty$. □

5.3 Product Measures

Given two measure spaces $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$, where both μ_1 and μ_2 are σ -finite, we can define a new measurable space (Ω, \mathcal{F}) , where $\Omega = \Omega_1 \times \Omega_2$ and

$$\mathcal{F} = \sigma \{A \times B : A \in \mathcal{A}_1, B \in \mathcal{A}_2\}.$$

We call $A \times B$ a rectangle. A natural measure μ for this measurable space satisfies

$$\mu(A \times B) = \mu_1(A) \mu_2(B) \tag{2}$$

for all rectangles $A \times B$. It can be checked that the collection of finite disjoint unions of rectangles is an algebra (exercise). To extend this measure μ to \mathcal{F} , we make use of Caratheodory's Extension Theorem (Theorem 3.1): we show that for $A \times B = \bigcup_{i \geq 1} A_i \times B_i$ a disjoint union of rectangles, we have $\mu(A \times B) = \sum_{i \geq 1} \mu_1(A_i) \mu_2(B_i)$.

For $x \in A$, let $I(x) = \{i : x \in A_i\}$ and $B = \bigcup_{i \in I(x)} B_i$ a disjoint union. We have

$$\begin{aligned} \mathbf{1}_A(x) \mu_2(B) &= \mathbf{1}_A(x) \mu_2 \left(\bigcup_{i \in I(x)} B_i \right) \\ &= \sum_{i \in I(x)} \mathbf{1}_A(x) \mu_2(B_i) \\ &= \sum_{i \geq 1} \mathbf{1}_{A_i}(x) \mu_2(B_i). \end{aligned}$$

Integrating w.r.t. μ_1 , we have

$$\begin{aligned}
\mu(A \times B) &= \int \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{1}_{A_i}(x) \mu_2(B_i) d\mu_1 \\
&\stackrel{\text{MCT}}{=} \lim_{n \rightarrow \infty} \int \sum_{i=1}^n \mathbf{1}_{A_i}(x) \mu_2(B_i) d\mu_1 \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^n \int \mathbf{1}_{A_i}(x) \mu_2(B_i) d\mu_1 \\
&= \sum_{i \geq 1} \mu_1(A_i) \mu_2(B_i),
\end{aligned}$$

where the interchange of the sum and integral in the penultimate equality holds because the number of terms is finite. Therefore, Caratheodory's Extension Theorem shows that there is a unique extension of μ defined by (2) to \mathcal{F} . This is called a *product measure*. Notationally, we write $\mu = \mu_1 \times \mu_2$.

The next theorem tells us when we can interchange integrals in general.

Theorem 5.2 (Fubini or Fubini-Tonelli). *Suppose $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ are σ -finite and $\mu = \mu_1 \times \mu_2$ is the product measure. Consider a measurable function $f : \Omega = \Omega_1 \times \Omega_2 \mapsto \mathbb{R}$. If $f \geq 0$ or $\int |f| d\mu < \infty$, then*

$$\int_{\Omega_1} \int_{\Omega_2} f(x, y) d\mu_2 d\mu_1 = \int_{\Omega} f d\mu = \int_{\Omega_2} \int_{\Omega_1} f(x, y) d\mu_1 d\mu_2.$$

Proof. Note that implicit in the theorem statement are the following that we need to prove:

- (i) For each $x \in \Omega_1$, $y \mapsto f(x, y)$ is \mathcal{A}_2 -measurable.
- (ii) $x \mapsto \int_{\Omega_2} f(x, y) d\mu_2$ is \mathcal{A}_1 -measurable.
- (iii) $\int_{\Omega_1} \int_{\Omega_2} f(x, y) d\mu_2 d\mu_1 = \int_{\Omega} f d\mu$.

Without loss of generality, we may assume $\mu_1, \mu_2 < \infty$ as the same proof is valid on each partition of Ω and then we can apply the MCT.

The proof follows the steps discussed at the end of Section 4.2. We first prove the theorem for the simplest case where $f = \mathbf{1}_E$, where $E \in \mathcal{F}$, the product σ -algebra. Let $E_x = \{y : (x, y) \in E\}$.

(i): Fix x , then $y \mapsto f(x, y) = \mathbf{1}_{E_x}(y)$. We need to show $E_x \in \mathcal{A}_2$. Let $\mathcal{E} = \{E \in \mathcal{F} : E_x \in \mathcal{A}_2\}$. We have $(E^c)_x = (E_x)^c$ since $y \in (E^c)_x \Leftrightarrow (x, y) \in E^c \Leftrightarrow y \in (E_x)^c$, and $(\bigcup_{i \geq 1} E_i)_x = \bigcup_{i \geq 1} (E_i)_x$. Therefore, \mathcal{E} is a σ -algebra and it contains all rectangles $A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$, which implies that $\mathcal{F} \subset \mathcal{E}$.

(ii) & (iii): Let $\mathcal{L} = \{E \in \mathcal{F} : f = \mathbf{1}_E \text{ satisfies (ii) \& (iii)}\}$.

- $\Omega \in \mathcal{L}$.
- If $E \in \mathcal{L}$, $\mu_2((E^c)_x) = \mu_2((E_x)^c) = \mu_2(\Omega_2) - \mu_2(E_x)$. Since $\mu_2(\Omega_2) < \infty$ and $\mu_2(E_x)$ is \mathcal{A}_1 -measurable, $\mu_2((E^c)_x)$ is \mathcal{A}_1 -measurable. We also have

$$\begin{aligned} \int \mu_2((E^c)_x) d\mu_1 &= \mu_2(\Omega_2)\mu_1(\Omega_1) - \int \mu_2(E_x) d\mu_1 \\ &= \mu(\Omega) - \mu(E) \\ &= \mu(E^c). \end{aligned}$$

Therefore, $E^c \in \mathcal{L}$.

- If $E_i \in \mathcal{L}$, $i \geq 1$, are disjoint, then

$$\begin{aligned} \mu_2\left(\left(\bigcup_{i \geq 1} E_i\right)_x\right) &= \mu_2\left(\bigcup_{i \geq 1} (E_i)_x\right) \\ &= \sum_{i \geq 1} \mu_2((E_i)_x) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu_2((E_i)_x), \end{aligned}$$

which is \mathcal{A}_1 -measurable from Lemma 4.2 since each $\mu_2((E_i)_x)$ is \mathcal{A}_1 -measurable. From the MCT, we have

$$\begin{aligned} \int \mu_2\left(\left(\bigcup_{i \geq 1} E_i\right)_x\right) d\mu_1 &= \lim_{n \rightarrow \infty} \int \sum_{i=1}^n \mu_2((E_i)_x) d\mu_1 \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \int \mu_2((E_i)_x) d\mu_1 \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(E_i) \\ &= \mu\left(\bigcup_{i \geq 1} E_i\right). \end{aligned}$$

Therefore, \mathcal{L} is a λ -system containing the collection of rectangles, which is a π -system. From the π - λ Theorem, we obtain $\mathcal{F} \subset \mathcal{L}$. We have now shown that Fubini's Theorem holds for $f = \mathbf{1}_E$, $E \in \mathcal{F}$.

From linearity of integrals, the theorem holds for all simple functions f .

For $f \geq 0$, \exists simple $f_i \uparrow f$. Applying MCT gives Fubini's theorem for non-negative f .

Finally, for general $f = f^+ - f^-$, we note that $\int |f| d\mu < \infty$ implies $\int f^+ d\mu, \int f^- d\mu < \infty$, and

$$\begin{aligned}\int f^+ d\mu &= \int \int f^+(x, y) d\mu_1 d\mu_2 \implies \int f^+(x, y) d\mu_1 < \infty \text{ } \mu_2\text{-a.e.} \\ &= \int \int f^+(x, y) d\mu_2 d\mu_1 \implies \int f^+(x, y) d\mu_2 < \infty \text{ } \mu_1\text{-a.e.}\end{aligned}$$

Similarly for f^- so that

$$\int \int f^+(x, y) d\mu_1 d\mu_2 - \int \int f^-(x, y) d\mu_1 d\mu_2 = \int \int f(x, y) d\mu_1 d\mu_2.$$

The proof is now complete. □

Example 5.2.

$$\begin{array}{c} x \longrightarrow \\ y \uparrow \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \dots \\ 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 1 & -1 & \dots \\ 0 & 1 & -1 & 0 & \dots \\ 1 & -1 & 0 & 0 & \dots \end{pmatrix} \end{array} \quad (3)$$

We have

$$\begin{aligned}\sum_y \sum_x f(x, y) &= \sum_y 0 = 0. \\ \sum_x \sum_y f(x, y) &= 1 + 0 + 0 + \dots = 1.\end{aligned}$$

This example shows that the conditions in Fubini's Theorem are essentially necessary.

Example 5.3. Suppose $((0, 1), \mathcal{B}(0, 1), \lambda) \times ((0, 1), 2^{(0, 1)}, \nu)$, where $\nu(A) = |A|$ is the counting measure, which is not σ -finite. Let

$$f(x, y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Then we have

$$\begin{aligned}\int f(x, y) d\nu &= 1 \text{ for each } x \implies \int \int f(x, y) d\nu d\lambda = 1, \\ \int f(x, y) d\lambda &= 0 \text{ for each } y \implies \int \int f(x, y) d\lambda d\nu = 0.\end{aligned}$$

This example shows that σ -finiteness of the measures is necessary.

5.4 Independence

Throughout this section, we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 5.3. Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. We write $A \perp\!\!\!\perp B$.

If $A \perp\!\!\!\perp B$, then it is easy to verify the following:

- (i) $A^c \perp\!\!\!\perp B$.
- (ii) $A \perp\!\!\!\perp B^c$.
- (iii) $A^c \perp\!\!\!\perp B^c$.

I.e., the two σ -algebras $\{\emptyset, \Omega, A, A^c\}$ and $\{\emptyset, \Omega, B, B^c\}$ are “independent”.

Definition 5.4. The sub- σ -algebra $\mathcal{A}_i \subset \mathcal{A}$, $i = 1, 2, \dots, n$, are independent if $\forall A_i \in \mathcal{A}_i$,

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i).$$

They are said to be pairwise independent if $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \forall i \neq j$.

We say that the random variables X_1, X_2, \dots, X_n are independent if $\sigma(X_i)$, $i = 1, \dots, n$, are independent, i.e., $\forall B_i \in \mathcal{B}$, $i = 1, \dots, n$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i).$$

Note that independence \implies pairwise independence but the converse is false. Here are two counter examples to show that pairwise independence does not imply independence.

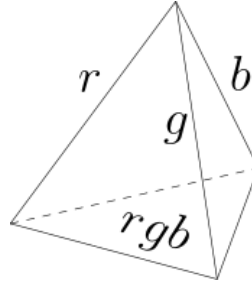


Fig. 2: Dice example: pairwise independence does not imply independence.

Example 5.4. Consider a fair tetrahedron dice that has one red edge, one green edge and one blue edge as shown in Fig. 2. The bottom has all three colors. Let r be the event that the

dice when tossed lands on a face with a red boundary. The events b , g and rgb are defined similarly. Then, by symmetry, we have

$$\begin{aligned}\mathbb{P}(r) &= \mathbb{P}(g) = \mathbb{P}(b) = \frac{1}{2}, \\ \mathbb{P}(rgb) &= \mathbb{P}(rg) = \mathbb{P}(gb) = \mathbb{P}(rb) = \frac{1}{4}.\end{aligned}$$

Therefore, these events are pairwise independent but not independent.

Example 5.5. Consider two six-sided fair dice. Let

$$\begin{aligned}A_1 &= \{1st\ dice\ is\ odd\}, \\ A_2 &= \{2nd\ dice\ is\ odd\}, \\ A_{sum} &= \{sum\ of\ the\ two\ dice\ is\ odd\}.\end{aligned}$$

We have

$$\begin{aligned}\mathbb{P}(A_1) &= \mathbb{P}(A_2) = \mathbb{P}(A_{sum}) = \frac{1}{2}, \\ \mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1 \cap A_{sum}) = \mathbb{P}(A_2 \cap A_{sum}) = \frac{1}{4}, \\ \mathbb{P}(A_1 \cap A_2 \cap A_{sum}) &= 0.\end{aligned}$$

Therefore, the events are pairwise independent but not independent. In particular, $\sigma(A_1)$ is not independent with $\sigma(\{A_2, A_{sum}\})$.

Lemma 5.2. Suppose the two collections of subsets \mathcal{E} and \mathcal{C} are π -systems and $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, $\forall B \in \mathcal{E}, C \in \mathcal{C}$. Then $\sigma(\mathcal{E})$ and $\sigma(\mathcal{C})$ are independent.

Proof. Let $\mathcal{D}_1 = \{D \in \mathcal{A} : \mathbb{P}(D \cap C) = \mathbb{P}(D)\mathbb{P}(C), \forall C \in \mathcal{C}\}$. As an exercise, one can check that \mathcal{D}_1 is a λ -system. Since $\mathcal{E} \subset \mathcal{D}_1$, from the π - λ theorem we have $\sigma(\mathcal{E}) \subset \mathcal{D}_1$.

Let $\mathcal{D}_2 = \{D \in \mathcal{A} : \mathbb{P}(B \cap D) = \mathbb{P}(B)\mathbb{P}(D), \forall B \in \sigma(\mathcal{E})\}$. Similarly \mathcal{D}_2 is a λ -system. From above, $\mathcal{C} \subset \mathcal{D}_2$ and by the π - λ theorem, we have $\sigma(\mathcal{C}) \subset \mathcal{D}_2$. Therefore, $\sigma(\mathcal{E}) \perp\!\!\!\perp \sigma(\mathcal{C})$. \square

By induction, if \mathcal{B}_i for $i = 1, \dots, n$ are π -systems and are independent, then $\sigma(\mathcal{B}_i)$ are independent.

Corollary 5.1. The random variables X_1, X_2, \dots, X_n are independent if

$$\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq t_i).$$

Proof. Since $\{(-\infty, t] : t \in \mathbb{R}\}$ is a π -system that generates $\mathcal{B}(\mathbb{R})$, the corollary follows from Lemma 5.2. \square

Lemma 5.3. Suppose that each random variable X_i has pdf f_i , $i = 1, \dots, n$. Then X_1, \dots, X_n are independent iff \exists a joint pdf $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(X_i)$.

Proof. ‘ \Leftarrow ’:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) &= \int_{A_1 \times \dots \times A_n} f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{A_1 \times \dots \times A_n} \prod_{i=1}^n f_i(x_i) dx_1 \cdots dx_n \\ &= \prod_{i=1}^n \int_{A_i} f_i(x_i) dx_i \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i). \end{aligned}$$

‘ \Rightarrow ’: Let $X = (X_1, \dots, X_n)$. For $A_i \in \mathcal{B}$, $i = 1, \dots, n$, we are given

$$\begin{aligned} \mathbb{P}(X \in A_1 \times \dots \times A_n) &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \\ &= \prod_{i=1}^n \int_{A_i} f_i(x_i) dx_i \\ &= \int_{A_1 \times \dots \times A_n} \prod_{i=1}^n f_i(x_i) dx_1 \cdots dx_n. \end{aligned} \tag{5}$$

We want to show that

$$\mathbb{P}(X \in A) = \int_A \prod_{i=1}^n f_i(x_i) dx_1 \cdots dx_n$$

for all A in the product σ -algebra $\mathcal{B}(\mathbb{R}^n) = \sigma\{A_1 \times \dots \times A_n : A_i \in \mathcal{B}\}$.

Let $\mathcal{L} = \{A \in \mathcal{E} : \mathbb{P}(X \in A) = \int_A \prod_{i=1}^n f_i(x_i) dx_1 \cdots dx_n\}$. It can be checked that \mathcal{L} is a λ -system and $\mathcal{P} = \{A_1 \times \dots \times A_n : A_i \in \mathcal{B}\}$ is a π -system and generates $\mathcal{B}(\mathbb{R}^n)$. Since $\mathcal{P} \subset \mathcal{L}$ from (5), by the π - λ Theorem, we have $\sigma(\mathcal{P}) \subset \mathcal{L}$ and the proof is complete. \square

Lemma 5.4. If $X \perp\!\!\!\perp Y$ and are integrable, then $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

Proof. The distribution of (X, Y) on \mathbb{R}^2 is the product measure $\mathbb{P}_X \times \mathbb{P}_Y$. From Fubini's theorem, we have

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy d\mathbb{P}_X \times \mathbb{P}_Y(x, y) = \int_{\mathbb{R}} x d\mathbb{P}_X(x) \int_{\mathbb{R}} y d\mathbb{P}_Y(y) = \mathbb{E}X\mathbb{E}Y.$$

\square

Suppose $X \perp\!\!\!\perp Y$. For any measurable functions f and g , $f(X) \perp\!\!\!\perp g(Y)$ since $\sigma(f(X)) \subset \sigma(X)$.