# 2 Random variables, independence, integration and conditioning

## 2.1 Measurable functions, products and measure pushforward

Let $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ be two measurable spaces. When $\Omega'$ is a topological space, we consider it per default endowed with the Borel $\sigma$-algebra. A function $X : \Omega \to \Omega'$ is called *measurable* if

$$X^{-1}(B) \in \mathcal{F}, \text{ for all } B \in \mathcal{F}', \tag{1}$$

where

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}.$$

When such a function is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we call $X$ a random variable (with values in $\Omega'$).

It is enough to require (1) to hold for $B$ running over some set of generators of $\mathcal{F}'$. For instance, for $\mathbb{R}$-valued $X$, measurability (1) holds if (1) holds for every $B = (-\infty, x]$ with $x$ running over the set of rational numbers.

A function $X : \Omega \to \mathbb{R}$ obtained by algebraic or analytic manipulations with a countable family $(X_n)$ of measurable $\mathbb{R}$-valued functions is again a measurable function. For instance $\limsup X_n$ is measurable (in general, as function into extended real line $\mathbb{R} \cup \{\infty\}$).

**Example** The indicator function of $A \in \mathcal{F}$

$$1_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \end{cases} .$$

is measurable, and so for $A_j \in \mathcal{F}$ are the *simple* functions of the form

$$X(\omega) = \sum_{j=1}^{n} y_j 1_{A_j}(\omega), \quad y_j \in \mathbb{R}\,.$$

**Definition 2.1.** Let $(X_t,\ t \in T)$ be a family of measurable functions $X_t : \Omega \to \Omega'$. The smallest sub-$\sigma$-algebra of $\mathcal{F}$ which makes all $X_t$'s measurable is called the *$\sigma$-algebra generated by* $(X_t,\ t \in T)$ and is denoted $\sigma(X_t,\ t \in T)$.

**Example** Let $\Omega = \{0,1\}^\infty$ be the coin-tossing space, $X_n(\omega) = \omega_n$ for $\omega = (\omega_1, \omega_2 \cdots)$. Then $\sigma(X_1, X_2, \dots)$ is the $\sigma$-algebra having the cylinder sets $A(\epsilon_1, \cdots, \epsilon_n), n \in \mathbb{N}$, as generators.

**Example** Generalising the example of the coin-tossing space, for $((\Omega_t, \mathcal{F}_t),\ t \in T)$ a family of measurable spaces, consider the Cartesian product

$$\Omega := \prod_{t \in T} \Omega_t = \{(\omega_t,\ t \in T) : \omega_t \in \Omega_t\},$$

Define $X_t$ to be the $t$th coordinate of $\omega \in \Omega$. The *product $\sigma$-algebra* is generated by the family $(X_t,\ t \in T)$ and is denoted $\bigotimes_{t \in T} \mathcal{F}_t$; this has the set of generators of the form

$$A_t \times \prod_{s \neq t} \Omega_s, \ \ A_t \in \mathcal{F}_t.$$

For two measure spaces $(\Omega, \mathcal{F}, \mu)$ and $(\Omega', \mathcal{F}', \mu')$ define a function on the family of rectanges

$$\nu(B \times B') := \mu(B)\mu(B'), \quad B \in \mathcal{F},\ B' \in \mathcal{F}'. \tag{2}$$

**Theorem 2.2.** *If $\mu$ and $\mu'$ are $\sigma$-finite measures, the function $\nu$ defined by* (2) *has a unique extension to a measure on the $\sigma$-algebra $\mathcal{F} \otimes \mathcal{F}'$.*

The extension is called *the product measure* and is denoted $\mu \times \mu'$, and the triple $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mu \times \mu')$ is called *the product measure space*.

Under measurable mapping the measure is transported from the source to the target space.

**Definition 2.3.** Let $X : \Omega \to \Omega'$ be a measurable function on a measure space $(\Omega, \mathcal{F}, \mu)$. The *image (or pushforward) measure* is defined as

$$\mu'(B') = \mu(X^{-1}(B')).$$

Sometimes notation $\mu_X$ for $\mu'$ is used.

**Example** For simple random variable

$$X = \sum_{j=1}^{n} y_j 1_{A_j}$$

the image measure on $\mathbb{R}$ is discrete,

$$\mu_X = \sum_{j=1}^{n} \mu(A_j)\delta_{y_j},$$

charging point $y_j$ with mass $\mu(A_j)$.

For $X$ a real random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the image probability measure measure is uniquely determined by the function

$$F_X(x) = \mathbb{P}(X \le x), \quad x \in \mathbb{R},$$

known as the cumulative distribution function of $X$.

For $\mathbb{R}^n$-valued random variable $X = (X_1, \cdots, X_n)$ (random vector) defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the image measure on $\mathbb{R}^n$ is called the *probability distribution* of $X$, or the *joint* probability distribution of $X_1, \cdots, X_n$. Let $i_1 < \cdots < i_m$ be a subset of $\{1, \cdots, n\}$ and consider the projection $(x_1, \cdots, x_n) \mapsto (x_{i_1}, \cdots, x_{i_m})$ which removes the entries outside the index set $\{i_1, \cdots, i_m\}$. Under such projection, the joint distribution of $(X_1, \cdots, X_n)$ is mapped to the joint distribution of subvector $(X_{i_1}, \cdots, X_{i_m})$ called an $m$-dimensional *marginal* distribution of vector $X$.

## 2.2 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Events $(A_t, t \in T) \subset \mathcal{F}$ are called independent if for every selection of distinct $t_1, \dots, t_k \in I$

$$\mathbb{P}(A_{t_1} \cap \cdots \cap A_{t_k}) = \mathbb{P}(A_{t_1}) \cdots \mathbb{P}(A_{t_k}).$$

Let $(\mathcal{F}_t, t \in T)$ be sub-$\sigma$-algebras of $\mathcal{F}$. They are called independent if for any choice of distinct indices $t_1, \dots, t_k$ any events $A_{t_1} \in \mathcal{F}_{t_1}, \dots, A_{t_k} \in \mathcal{F}_{t_k}$ are independent.

Independence of random variables $X_i$ is defined as independence of their generated $\sigma$-algebras $\sigma(X_i)$.

For every family $(P_t, t \in T)$ of probability measures on $\mathbb{R}$ there exists a family of independent random variables $(X_t, t \in T)$ with $X_t$ having distribution $P$. This follows from the construction of the product measure.

### 2.3 Tail events

Let $A_i \in \mathcal{F}$ be events, $i \in \mathbb{N}$. Consider the event '$A_n$ occurs inifnitely often' (more precisely, 'infinitely many of $A_n$'s occur')

$$\{A_n \text{ i.o.}\} := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

**Theorem.** (Borel-Cantelli Lemma)

(a) *If* $\sum_n \mathbb{P}(A_n) < \infty$ *then* $\mathbb{P}(A_n \text{ i.o.}) = 0$,

(b) *If* $A_1, A_2, \ldots$ *are independent and* $\sum_n \mathbb{P}(A_n) = \infty$ *then* $\mathbb{P}(A_n \text{ i.o.}) = 1$.

*Proof.* Part (a) is an exercise from Lecture 1. We focus on (b). We have

$$\{A_n \text{ i.o.}\}^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c.$$

Clearly

$$\bigcap_{k=1}^{\infty} A_k^c \subset \bigcap_{k=2}^{\infty} A_k^c \subset \cdots,$$

hence

$$\mathbb{P}(\{A_n \text{ i.o.}\}^c) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{m} A_k^c\right) =$$

using independence and that $\sum_n \mathbb{P}(A_n) = \infty$

$$= \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} (1 - \mathbb{P}(A_k)) = 0.$$

□

**Example** Let $X_1, X_2, \ldots$ be independent $\mathcal{N}(0,1)$-distributed random variables (any other continuous distribution would also work). We say that there is a record at index $n$ if $X_n = \max(X_1, \ldots, X_n)$; denote this event $A_n$. One can check that $\mathbb{P}(A_n) = 1/n$ and that the events are independent. Since $\sum_n 1/n = \infty$ the number of records is infinite with probability 1.

Suppose the occurence/not occurence of event $A_n$ becomes known to an observer at time $n$. The Borel-Cantelli Lemma exemplifies situation where probability of some related 'distant' event may assume only values 0 and 1. Results of the kind are known as 'zero-one laws, which we discuss next.

Let $\mathcal{F}_j, j \in \mathbb{N}$, be $\sigma$-algebras (sub-$\sigma$-algebras of $\mathcal{F}$). We define the *tail $\sigma$-algebra* as

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \mathcal{F}_k\right).$$

Each $A \in \mathcal{T}$ is called *tail event*.

**Example** In the coin-tossing space, let $\mathcal{F}_n$ be the $\sigma$-algebra generated by outcomes in $n$ first trials. The event 'the pattern 1011101 occurs infinitely many times in the sequence' is a tail event.

**Theorem.** (Kolmogorov's $0-1$ law) *If* $\mathcal{F}_1, \mathcal{F}_2, \cdots$ *are independent, then* $\mathcal{T}$ *is trivial in the sense that* $\mathbb{P}(A) = 0$ *or* $1$ *for each* $A \in \mathcal{T}$.

*Proof.* Suppose $A$ is a tail event, since $A \in \sigma(\bigcup_{k=n}^{\infty} \mathcal{F}_k)$, we have that $A$ is independent of $\mathcal{F}_1, \ldots, \mathcal{F}_{n-1}$. Since this holds for every $n$, $A$ is independent of $\sigma(\bigcup_{k=1}^{\infty} \mathcal{F}_k)$ and thus independent of smaller $\sigma$-algebra $\mathcal{T}$. In particular, $A$ is independent of itself, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\,\mathbb{P}(A)$, which is only possible when $\mathbb{P}(A)$ is 0 or 1. □

**Example** Let $X_1, X_2, \dots$ be independent random variables, generating $\sigma$-algebras $\sigma(X_j),\ j \in \mathbb{N}$. The event

$$A = \{\omega \in \Omega : \sum_{n=1}^{\infty} X_n \quad \text{converges}\}$$

is a tail event, therefore can only have probability $0$ or $1$.

**Theorem.** (Kolmogorov's Three Series Theorem) *Series $\sum_{n=1}^{\infty} X_n$ of independent random variable converges alsmost surely if and only if the following conditions hold with some constant $c > 0$*

(i) $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c) < \infty$,

(ii) $\sum_{n=1}^{\infty} \mathbb{E}(X_n 1_{\{|X_n| \le c\}}) < \infty$,

(iii) $\sum_{n=1}^{\infty} \mathrm{Var}(X_n 1_{\{|X_n| \le c\}}) < \infty$.

**Example** For independent normal random variables $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ convergence of the series $\sum_n X_n$ holds if and only if $\sum_n m_n$ and $\sum_n \sigma_n^2$ both converge.

### 2.4 Lebesgue integral and expectation

The expectation of discrete random variable $X$ with values $x_1, x_2, \dots$ is

$$\mathbb{E}\, X = \sum_j x_j\, \mathbb{P}(X = x_j),$$

and if $X$ has a density $f_X$

$$\mathbb{E}\, X = \int_{-\infty}^{\infty} x f_X(x) dx.$$

These are unified by the general concept of Lebesgue integral.

For measurable function $g : \mathbb{R} \to \mathbb{R}$ and a measure $\mu$ on $\mathbb{R}$ we wish to define

$$\int_{\mathbb{R}} g(x) d\mu(x) \ \text{(also written as } \int_{\mathbb{R}} g(x)\mu(dx))$$

Suppose first that $g$ is nonnegative. For simple

$$g(x) = \sum_{j=1}^{k} y_j 1_{A_j}(x), \quad A_j \in \mathcal{B}(\mathbb{R})$$

we set

$$\int_{\mathbb{R}} g(x) d\mu(x) = \sum_{j=1}^{k} y_j \mu(A_j).$$

For the general $f \ge 0$, consider sets

$$A_{jk} = \begin{cases} \{x : \frac{k}{2^j} \le f(x) < \frac{k+1}{2^j}\}, & k = 0, 1, \dots, j2^j - 1, \\ \{x : f(x) \ge j\}, \ k = j2^j, \end{cases}$$

and simple functions

$$g_j(x) = \sum_{k=0}^{j2^j} \frac{k}{2^j}\ 1_{A_{jk}}(x),$$

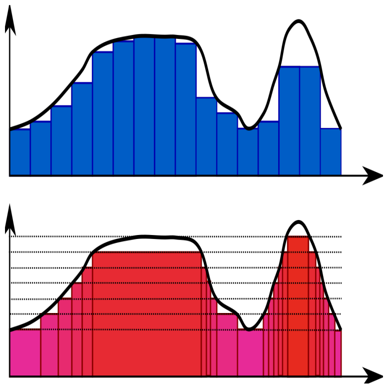

Figure 1: Lower Riemann integral sum and integral sum for Lebesgue integral.

so

$$\int_\Omega g_j(x)d\mu(x) = \sum_{k=0}^{j2^j} \frac{k}{2^j}\ \mu(A_{jk}),$$

which we consider as a lower approximation for Lebesgue integral. The *Lebesgue integral* of $g$ is defined as the limit

$$\int_\Omega g(x)d\mu(x) := \lim_{j\to\infty} \int_\Omega g_j(x)d\mu(x).$$

**Example** Let $g(x) = 1_{[0,1]\setminus\mathbb{Q}}$ be the indicator function of irrational numbers on $[0,1]$. The Riemann integral over $[0,1]$ does not exist, because every upper integral sum is 1, and every lower is 0. The Lebesgue integral is

$$\int_{[0,1]} g(x)dx = 1\cdot\lambda([0,1]\setminus\mathbb{Q}) + 0\cdot\lambda([0,1]\cap\mathbb{Q}) = 1.$$

Note that here $dx$ means the same as $d\lambda(x)$.

For the general $g : \Omega \to \mathbb{R}$ let $g_+(x) = \max(g(x), 0), g_-(x) = \max(-g(x), 0)$ be positive and negative parts, then $g(x) = g_+(x) - g_-(x)$. If

$$\int_\Omega |g(x)|d\mu(x) < \infty$$

we say that $g$ is integrable and we define the Lebesgue integral of $f$ as

$$\int_\Omega g(x)d\mu(x) = \int_\Omega g_+(x)d\mu(x) - \int_\Omega g_-(x)d\mu(x).$$

Note that $\int_0^\infty x^{-1}(\sin x)dx = \lim_{a\to\infty}\int_0^a x^{-1}(\sin x)dx = \pi/2$ exists as improper Riemann integral over $\mathbb{R}_+$, but not as Lebesgue integral because $\int_0^\infty |(\sin x)/x|dx = \infty$.

The definition of Lebesgue integral for random variable $X$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is analogous to the inegral over $\mathbb{R}$. These are related as

$$\mathbb{E}X := \int_\Omega X(\omega)d\,\mathbb{P}(\omega) = \int_\mathbb{R} x dP_X(x),$$

where $P_X$ is the distribution of $X$.

### 2.5 Absolute continuity of measures

Let $p_j, j \in \mathbb{N}$ be positive numbers, with sum 1. We may treat the identity function on $\mathbb{N}$ as a random variable $X : \mathbb{N} \to \mathbb{N}$ with the probability mass function $P = (p_j, j \in \mathbb{N})$. For any function $g$ we have the expectation of $g(X)$ calculated as

$$\mathbb{E}_P\,[g(X)] = \sum_{j \in \mathbb{N}} g(j) p_j$$

If $Q = (Q_j, j \in \mathbb{N})$ is some other probability mass function, the corresponding expectation is

$$\mathbb{E}_Q\,[g(X)] = \sum_{j \in \mathbb{N}} g(j) q_j.$$

To write the $Q$-expectation in terms of $P$, let $\xi(j) = q_j/p_j$, then

$$\mathbb{E}_Q\, g(X) = \sum_{j \in \mathbb{N}} g(j) \xi_j p_j = \mathbb{E}_P\,[\xi g(X)].$$

The random variable $\xi$ is an instance of the Radon-Nikodym derivative (or density).

In full generality, let $\mu, \nu$ be two measures on $(\Omega, \mathcal{F})$. Call $\nu$ *absolutely continuous* with respect to $\mu$, written as $\mu \gg \nu$ if

$$A \in \mathcal{F},\ \mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0.$$

The measures are called *equivalent*, denoted $\mu \sim \nu$, if

$$\mu(A) = 0 \quad \Leftrightarrow \quad \nu(A) = 0,$$

which means that the measures have the same null-sets.

**Theorem.** (Radon-Nikodym theorem.) *If $\mu \gg \nu$ and $\mu$ is $\sigma$-finite then there exists a nonnegative measurable function $\xi$ on $\Omega$ such that for any measurable $g : \Omega \to \mathbb{R}$*

$$\int_\Omega f(x) d\nu(x) = \int_\Omega f(x) \xi(x) d\mu(x),$$

*provided one of the integrals exists.*

In particular, $\nu(A) = \int_A \xi(x) d\mu(x)$. We write $\xi = \frac{d\nu}{d\mu}$ and call $\xi$ the Radon-Nikodym deritative of $\nu$ with respect to $\mu$. Such $\xi$ is unique up to values on a set of $\mu$-measure $0$.

**Example** For $\lambda$ the lebesgue measure, $\nu$ the normal $\mathcal{N}(0, 1)$ distribution, the Radon-Nikodym derivative is the normal density

$$\xi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi.}}$$

### 2.6 Conditional expectation

For two random variables, $X, Y$, recall that the conditional expectation $\mathbb{E}[X|Y]$ is defined as follows. Calculate the function $h(y) = \mathbb{E}[X|Y = y]$, in case of discrete random variables as

$$\mathbb{E}[X|Y = y_j] = \sum_j x_i\, \mathbb{P}(X = x_i | Y = y_j],$$

or when $(X, Y)$ have joint density as

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y=y}(x) dx,$$

where

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

is the conditional density. Then define $\mathbb{E}[X|Y] = h(Y)$ by subsituting random variable $Y$ for dummy variable $y$.

Intuitively, $\mathbb{E}[X|X] = X$, and in the discrete case this is easily checked. When $X$ has density, this is still true but we cannot use the above formula with $Y = X$, because $(X, X)$ has no *joint* density function.

We wish to introduce more general conditional expectation $\mathbb{E}[X|\mathcal{G}]$ given sigma-algebra $\mathcal{G} \subset \mathcal{F}$. Suppose first $X \geq 0$. Let

$$\mathbb{Q}(A) := \mathbb{E}[X \cdot 1_A] = \int_A X d\,\mathbb{P}\,.$$

For disjoint sets $A_n \in \mathcal{G}$

$$\int_{\cup_n A_n} X d\,\mathbb{P} = \sum_n \int_{A_n} X d\,\mathbb{P},$$

which entails that $\mathbb{Q}$ is a measure, and $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$. By the Radon-Nikodym theorem there exists a $\mathcal{G}$-measurable random variable $\xi$ such that

$$\mathbb{Q}(A) = \int_A \xi d\,\mathbb{P}\,.$$

We denote this variable as

$$\xi = \mathbb{E}[X|\mathcal{G}],$$

and call it the conditional expectattion of $X$ given $\mathcal{G}$. The defining property is

$$\int_A X d\,\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}] d\,\mathbb{P}, \qquad A \in \mathcal{G}.$$

For any $X$, we write $X = X_+ - X_-$ and define the conditional expectation by

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X_+|\mathcal{G}] - \mathbb{E}[X_-|\mathcal{G}],$$

which exists if $X$ is integrable.

The following rules will be used in the sequel:

(i) $\mathbb{E}[X|\{\varnothing, \Omega\}) = \mathbb{E}\,X$,

(ii) $\mathbb{E}[aX + bY|\mathcal{G}] = a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]$,

(iii) $\mathbb{E}[1|\mathcal{G}] = 1$,

(iv) taking out what is known: if $Y$ is $\mathcal{G}$-measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = Y \cdot \mathbb{E}[X|\mathcal{G}],$$

(v) tower property: for $\mathcal{G}_1 \subset \mathcal{G}_2$

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1],$$

in particular $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}\,X$.

**Exercises**

1. Let $Y_1, Y_2, \dots$ be independent, exponentially distributed random variables with $\mathbb{E}\, Y_i = 1$. Show that $\mathbb{P}(Y_n > \log n \text{ i.o.}) = 1$.

2. Show that condition (i) in the three series theorem is necessary for convergence of the series.

3. Suppose rv's $X_1, \cdots, X_n$ independent, rv's $Y_1, \cdots, Y_m$ independent, and random vectors $(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_m)$ are independent. Show that the $(n+m)$ random variables $X_1, \cdots, X_n, Y_1, \cdots, Y_m$ are independent.

4. Let $X_1, X_2, \dots$ be arbitrary random variables. Prove that if $\sum_{j=1}^{\infty} \mathbb{E}|X_j| < \infty$ then the series $\sum_{j=1}^{\infty} X_j$ converges absolutely with probability one.

5. Suppose $\mathbb{E}X$ exists. Argue that for every $\epsilon$ there exists $\delta$ such that $\mathbb{P}(A) < \delta$ implies
$$\mathbb{E}(|X| \cdot 1_A) < \epsilon$$
(where $1_A$ indicator of event $A$).

6. Show that $\mathbb{E}[XY] = \mathbb{E}X\, \mathbb{E}Y$ if the rv's are independent.

7. For three measures suppose $\mu \gg \nu \gg \rho$ and that $\mu, \nu, \rho$ are $\sigma$-finite. Prove the chain rule for the Radon-Nikodým derivative:
$$\frac{d\rho}{d\mu} = \frac{d\nu}{d\mu}\frac{d\rho}{d\nu}.$$

8. Let $\mu$ be a normal distribution $\mathcal{N}(m, \sigma^2)$, and $\nu$ the exponential distribution with parameter $\beta$. Argue that $\mu \gg \nu$ and find the Radon-Nikodym derivative $d\nu/d\mu$.

9. Let $A_{i,j}$ be a system of disjoint events, with $\cup_{i,j} A_{i,j} = \Omega$. Let $A_i = \cup_j A_{i,j}$. Let $\mathcal{G}_2$ be generated by all $A_{i,j}$'s, and let $\mathcal{G}_1$ be generated by $A_i$'s. Describe as precise as you can the random variables $\mathbb{E}[X|\mathcal{G}_1], \mathbb{E}[X|\mathcal{G}_2]$. Assuming $\mathbb{P}(A_{i,j}) > 0$, prove the tower property in this example.

**Literature**


1. S. Resnick, A probability path, Springer 2003.

2. R. Schilling, Measures, integrals and martingales, CUP 2005.

3. A. Shiryaev, Probability, Springer, 1996.