SOME CONCEPTS IN PROBABILITY

Let us now set up the probability setting by using the theory of measures developed in the previous sections.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An $\mathcal{F}$-measurable function X (complex, or valued in $[-\infty, \infty]$) on Ω is called a random variable. The concept of random variables may be generalized to mappings, which may be useful in discussing probability models. In general, if $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measurable spaces, then a mapping $\Phi : \Omega_1 \to \Omega_2$ is measurable if $\Phi^{-1}(A) \in \mathcal{F}_1$ whenever $A \in \mathcal{F}_2$. Thus a real random variable $X : \Omega \to \mathbb{R}$ is just a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

If X is integrable or non-negative random variable, then its integral $\int_\Omega X(\omega)\mathbb{P}(d\omega)$ is called the *expectation* of X, or the mean value of X, denoted by $\mathbb{E}[X]$. We say the expectation of X exists if X is integrable.

Exercise 6.1 *The* inclusion-exclusion formula *holds:*

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \sum_j \mathbb{P}(A_j) - \sum_{j_1<j_2} \mu\mathbb{P}(A_{j_1}A_{j_2}) + \sum_{j_1<j_2<j_3} \mathbb{P}(A_{j_1}A_{j_2}A_{j_3}) \\ &\quad - \cdots + (-1)^{k-1} \sum_{j_1<\cdots<j_k} \mathbb{P}(A_{j_1}\cdots A_{j_k}) + \cdots\end{aligned}$$

where $A_j \in \mathcal{F}$ *for* $j = 1, 2, \cdots$.

Exercise 6.2 *1) Let* $f : \Omega \to S$. *Show that* $f^{-1}(\cup_\alpha A_\alpha) = \cup_\alpha f^{-1}(A_\alpha)$ *and* $f^{-1}(A^c) = (f^{-1}(A))^c$ *for any sets* A, A_α *where* α *runs over an arbitrary index set.*

2) Let $f : \Omega \to S$ *where* $(\Omega, \mathcal{F})$ *and* (S, Σ) *are two measurable spaces. Show that*

$$f^{-1}(\Sigma) \equiv \{f^{-1}(A) : A \in \Sigma\}$$

is a σ*-algebra on* Ω, *and* f *is measurable (with respect to* $\mathcal{F}$*) if and only if* $f^{-1}(\Sigma) \subset \mathcal{F}$.

Exercise 6.3 *Let* (S, Σ) *be a measurable space, and* $X_\alpha : \Omega \to S$ $(\alpha \in \Lambda)$ *be a family of functions on* Ω *taking values in* S. *Then we use* $\sigma\{X_\alpha : \alpha \in \Lambda\}$ *to denote the smallest* σ*-algebra such that each* X_α *is a measurable map from* $(\Omega, \sigma\{(X_\alpha)_{\alpha\in\Lambda}\})$ *to* (S, Σ).

1) Let $\Sigma_0 = \{X_\alpha^{-1}(A) : A \in \Sigma \text{ and } \alpha \in \Lambda\}$. *Show that*

$$\sigma\{X_\alpha : \alpha \in \Lambda\} = \sigma(\Sigma_0).$$

2) Let $\mathcal{F} \equiv \sigma\{X_\alpha : \alpha \in \Lambda\}$. *Show that, if* $\alpha_j \in \Lambda$ $(j = 1, 2, \cdots)$ *is a countable subset of* Λ *and* $A \in \Sigma$, *then*

$$\{\omega : X_{\alpha_j}(\omega) \in A \text{ for all } j = 1, 2, \cdots\}$$

belongs to $\mathcal{F}$. *The above event is often written as* $\{X_{\alpha_j} \in A \text{ for } j = 1, 2, \cdots\}$.

6.1 Laws, distribution functions

These are basic concepts associated with random variables. Let us begin with the following

Proposition 6.4 *Let $(\Omega, \mathcal{F})$ and (S, Σ) be two measurable spaces, $\mathbb{P}$ a measure on $(\Omega, \mathcal{F})$, and $X : \Omega \to S$ be a measurable map. Define*

$$\begin{aligned} \mu(A) &\equiv \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}\left[X \in A\right] \\ &= \mathbb{P}\left(\{\omega : X(\omega) \in A\}\right) \end{aligned}$$

for every $A \in \Sigma$. Then μ is a measure on (S, Σ), denoted by $\mathbb{P} \circ X^{-1}$, which is called the distribution of X.

In particular, if X is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{R}^n$, then $\mathbb{P} \circ X^{-1}$ is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, called the *law* or called *the distribution* of the random variable X. Sometimes we also use μ_X to denote the distribution of X.

If $X : \Omega \to \mathbb{R}$ is a real-valued random variable, then its *distribution function*

$$\begin{aligned} F(x) &= \mathbb{P}\left(X \leq x\right) \\ &= \mathbb{P}\left(\{\omega : X(\omega) \leq x\}\right) \\ &= \mu_X\left((-\infty, x]\right), \end{aligned}$$

is a non-decreasing function on $\mathbb{R}$ with values in $[0,1]$. Then $0 \leq F \leq 1$; $F \uparrow$; $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$; F is right-continuous:

$$\lim_{x \downarrow x_0} F(x) = F(x_0) \qquad \forall x_0 \in \mathbb{R}.$$

The Lebegue-Stieltjes measure m_F associated with the increasing and right-continuous function F is the unique measure such that

$$m_F((a,b]) = F(b) - F(a) = \mathbb{P}\left(a < X \leq b\right) = \mu_X((a,b])$$

for all $a < b$. Since the collection $\mathcal{C}$ of all $(a,b]$ (where $a < b$ are reals) is a π-system, according to the Uniqueness Lemma 2.2, $m_F = \mu_X$, that is, the distribution (law) of a real random variable X is the Lebesgue-Stieltjes measure associated with the distribution function of X.

6.2 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

1. Independent events. Recall that, if $A, B \in \mathcal{F}$ be two events, then A and B are independent, if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)P(B). \tag{6.1}$$

Let

$$\begin{aligned} \mathcal{F}_A &= \sigma\{A\} = \{\Omega, A, A^c, \emptyset\}, \\ \mathcal{F}_B &= \sigma\{B\} = \{\Omega, B, B^c, \emptyset\}. \end{aligned}$$

Then (6.1) implies that

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F), \quad \forall E \in \mathcal{F}_A, F \in \mathcal{F}_B,$$

and therefore the σ-algebras $\mathcal{F}_A$ and $\mathcal{F}_B$ are *independent.*

Definition 6.5 *1) Let $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ be a collection of sub σ-algebras of $\mathcal{F}$. Then $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ are independent if for any $k \in \mathbb{N}$, and any $\alpha_1, \cdots, \alpha_k \in \wedge$ such that $\alpha_i \neq \alpha_j$ if $i \neq j$, we have*

$$\mathbb{P}(A_1 \cdots A_k) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_k), \quad \forall A_1 \in \mathcal{F}_{\alpha_1}, \cdots, A_k \in \mathcal{F}_{\alpha_k}.$$

2) Let $\{F_\alpha : \alpha \in \wedge\}$ be a family of events: $F_\alpha \in \mathcal{F}$. Then we say $\{F_\alpha : \alpha \in \wedge\}$ are independent if $\{\sigma(F_\alpha) : \alpha \in \wedge\}$ are independent.

3) Let $\{X_\alpha : \alpha \in \wedge\}$ be a family of random variables. Then $\{X_\alpha : \alpha \in \wedge\}$ are independent if the family of σ-algebras $\{\sigma(X_\alpha) : \alpha \in \wedge\}$ are independent.

2. Independence via π-system. In elementary probability theory, we already give a definition of independence for random variables. You should show that the above definition coincides with the one you have learned before. The following Lemma is very useful although it is very simple and follows a simple application of Lemma 2.2.

Lemma 6.6 *Let $\mathcal{F}_\alpha \equiv \sigma\{\mathcal{C}_\alpha\}$ where each $\mathcal{C}_\alpha$ is a π-system in the sense that*

$$A, B \in \mathcal{C}_\alpha \quad \text{implies that } A \cap B \in \mathcal{C}_\alpha.$$

Then $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ are independent if and only if for any $k \in \mathbb{N}$, any $F_1 \in \mathcal{C}_{\alpha_1}, \cdots, F_k \in \mathcal{C}_{\alpha_k}$ where $\alpha_1, \cdots, \alpha_k$ are different, we have

$$\mathbb{P}[F_1 \cap \cdots \cap F_k] = \mathbb{P}[F_1] \cdots \mathbb{P}[F_k].$$

In fact, we can show the equality by induction on k. Consider two measures on $\mathcal{F}_{\alpha_k}$ defined by

$$\mu_1(E) = \mathbb{P}[F_1 \cap \cdots \cap F_{k-1} \cap E]$$

and

$$\mu_2(E) = \mathbb{P}[F_1 \cap \cdots \cap F_{k-1}]\mathbb{P}(E)$$

where F_i as in the lemma, but fixed, and $E \in \mathcal{F}_{\alpha_k}$. The induction assumption and the condition in the lemma implied that $\mu_1 = \mu_2$ on $\mathcal{C}_{\alpha_k}$, hence, by Lemma 2.2, $\mu_1 = \mu_2$ on $\mathcal{F}_{\alpha_k}$ and the proof is complete.

3. Independent random variables.

Theorem 6.7 *Let $X_1, \cdots, X_n, \cdots$ be a sequence of real random variables. Then $X_1, \cdots, X_n, \cdots$ are independent if and only if for any $k \in \mathbb{N}$, and any $x_1, \cdots, x_k \in \mathbb{R}$*

$$\mathbb{P}[X_1 \leq x_1, \cdots, X_k \leq x_k] = \mathbb{P}[X_1 \leq x_1] \cdots \mathbb{P}[X_k \leq x_k].$$

That is, the joint distribution of $X_1, \cdots, X_n$ is the product of the distribution functions of the random variables $X_k, 1 \leq k \leq n$.

This follows from the previous lemma, as $\mathcal{C}_k$ the collection of all subsets $\{X_k \leq x\}$ where x runs through all reals is a π-system, where $k = 1, 2, \cdots$.

Therefore, the *joint* law or distribution of a sequence of independent random variables $(X_1, X_2, \cdots, X_n, \cdots)$ is the product probability measure $\mu_1 \times \cdots \times \mu_n \times \cdots$, where μ_n is the distribution of X_n. In particular, if $\{X_n : n = 1, 2, \cdots\}$ is a sequence of independent real random variables, then its joint law (or called joint distribution) is the product probability measures of the Lebesgue-Stieltjes measure m_{F_n} where $F_n(x) = \mathbb{P}[X_n \leq x]$ is the distribution function of X_n, $n = 1, 2, \cdots$.

Theorem 6.8 *Let X be a random variable (valued in a measurable space) on some probability space. Then there is a sequence of independent identically distributed random variables $\{X_n : n \in \mathbb{N}\}$, each X_n has the same law as that of X.*

Proof. [*The proof is not examinable*] Let X be a random variable taking its values in a measurable space $(S, \mathcal{G})$, and let μ be the distribution of X. Then μ is a probability measure. Let $(S_n, \mathcal{G}_n, \mu_n) = (S, \mathcal{G}, \mu)$ $(n = 1, 2, \cdots)$ and let $\mathbb{P} = \mu_1 \times \cdots \times \mu_n \times \cdots$ be the product probability measure on $\Omega = \prod_{n=1}^{\infty} S_n$. Define $X_n : \Omega \to S$ by $X_n(w) = w_n$ if $w = (w_n) \in \Omega$ for $n = 1, 2, \cdots$. Then X_n are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ (where $\mathcal{F} = \prod_{n=1}^{\infty} \mathcal{G}_n$) and by construction, X_n have the common distribution μ, and (X_n) are independent. ∎

6.3 Borel-Cantelli lemma

1. Limiting events, Borel-Cantelli's first and second lemma. Let $A_n \in \mathcal{F}$ for $n = 1, 2, \cdots$. The event that "A_n's occur infinitely often" (or "infinitely many A_n occur") is given by

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \\ &= \{\omega : \omega \text{ belongs to infinitely many } A_n\}. \end{aligned}$$

The event $\limsup_{n\to\infty} A_n$ is also denoted by $\{A_n : \text{ i.o.}\}$. Similarly, though less important in applications, the event that "A_n take place eventually" is

$$\begin{aligned} \liminf_{n\to\infty} A_n &= \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n \\ &= \{\omega : \exists N(\omega) \text{ s.t. } \omega \in A_n \text{ for all } n \geq N(\omega)\} \\ &= \{\omega : \omega \text{ eventually belongs to } A_n \text{ for large } n\}. \end{aligned}$$

This event is denoted sometimes by $\{A_n : \text{ev.}\}$. By definition, it is easy to see that

$$\limsup_{n\to\infty} A_n = \left\{\sum_{n=\cdot}^{\infty} 1_{A_n} = \infty\right\} = \left\{\limsup_{n\to\infty} 1_{A_n} = 1\right\}$$

while

$$\liminf_{n\to\infty} A_n = \left\{\lim_{n\to\infty} 1_{A_n} = 1\right\}.$$

Theorem 6.9 *Let $A_n \in \mathcal{F}$ (where $n = 1, 2, \cdots$).*

*1) (*Borel-Cantelli Lemma, *first Borel-Cantelli lemma). If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then* $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$.

*2) (*Borel zero-one criterion, *second Borel Cantelli lemma). If the events $\{A_n\}$ are independent, then $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ if and only if* $\mathbb{P}[\limsup_{n\to\infty} A_n] = 1$.

Proof. 1) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\lim_{m\to\infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0$, and therefore

$$\mathbb{P}[A_n: \text{ i.o.}] = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} A_n\right) \leq \lim_{m\to\infty} \sum_{n\geq m} \mathbb{P}(A_n) = 0.$$

2) If A_n are independent, and if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\begin{aligned} \mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n^c\right) &= \lim_{N\to\infty} \prod_{n=m}^{N} \mathbb{P}(A_n^c) = \lim_{N\to\infty} \prod_{n=m}^{N} (1 - \mathbb{P}(A_n)) \\ &\leq \lim_{N\to\infty} \exp\left(-\sum_{n=m}^{N} \mathbb{P}(A_n)\right) \\ &= 0 \end{aligned}$$

for every m, where we have used the elementary inequality: $1 - x \leq e^{-x}$ for $x \in [0, 1]$. Since

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

and every $\bigcap_{n=m}^{\infty} A_n^c$ has probability zero, so that their union $\{A_n: \text{ i.o}\}^c$ over $m = 1, 2, \cdots$ has zero probability too, hence $\mathbb{P}[A_n: \text{ i.o}] = 1$. ■

2. Tail events and tail σ-algebra. The $\limsup A_n$ and $\liminf A_n$ are examples of so-called *tail events* – these events are determined by $\{A_{m+1}, A_{m+1}, \cdots, A_n, \cdots\}$ for every m. For example

$$\limsup_{n\to\infty} A_n = \left\{\sum_{n=m+1}^{\infty} 1_{A_n} = \infty\right\}$$

for any m. From Borel zero-one criterion above, we can deduce the limiting behavior of these tail events by combining with the concept of independence. *If* $X_1, X_2, \cdots, X_n, \cdots$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then the σ-algebra $\mathcal{G}_\infty = \bigcap_{n=1}^{\infty} \sigma\{X_j : j > n\}$ is called the tail σ-algebra of $\{X_k\}_{k\geq 1}$. Any element in $\mathcal{G}_\infty$ is called a *tail event.*

Proposition 6.10 *(*A. Kolomogorov's 0-1 law*) If $\{X_n\}$ is a sequence of* independent *random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}_\infty = \bigcap_{n=1}^{\infty} \sigma\{X_j : j > n\}$. Then $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathcal{G}_\infty$. In particular, if $\{A_n\}$ is a sequence of* independent *events, then* $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$ *or* 1 .

Proof of 0-1 law. Since $\sigma\{X_j : j \leq n\}$ and $\sigma\{X_j : j > n\}$ are independent for any $n = 1, 2, \cdots$, so that $\sigma\{X_j : j \leq n\}$ and $\mathcal{G}_\infty$ for every n are independent. It follows that $\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ and $\mathcal{G}_\infty$ are independent. If $B, C \in \bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$, then $B \cap C \in \bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ as well, so $\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ is a π-system, thus, by Lemma 6.6, the σ-algebra

$$\sigma\left[\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}\right] = \sigma\{X_j : j \geq 1\}$$

and $\mathcal{G}_\infty$ are independent. Since $\mathcal{G}_\infty \subset \sigma\{X_j : j \geq 1\}$, $\mathcal{G}_\infty$ and itself are independent. Therefore, for every $A \in \mathcal{G}_\infty$, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which yields that $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. The last conclusion comes from the fact that $\limsup_{n\to\infty} A_n \in \mathcal{G}_\infty$, so that $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$ or 1.

3. Example. Suppose (X_n) is a sequence of independent random variables (real or complex), and $\mathcal{G}_\infty$ is its tail σ-algebra, and suppose $\{b_n\}$ be an increasing sequence of positive numbers such that $b_n \uparrow \infty$. Then the following events

$$\left\{\lim_{n\to\infty} X_n \text{ exists}\right\}, \left\{\sum_{n=1}^{\infty} X_n \text{ converges}\right\} \text{ and } \left\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{b_n} \text{ exists}\right\}$$

are all tail events, i.e. belong to $\mathcal{G}_\infty$, and thus have probability one or zero.