THE UNIVERSITY OF WESTERN ONTARIO
LONDON CANADA

Paul Klein
Office: SSC 4028
Phone: 661-2111 ext. 85227
Email: paul.klein@uwo.ca
URL: www.ssc.uwo.ca/economics/faculty/klein/

# Probability theory and Lebesgue integration

## 1 Outcomes, events, expectations

**Definition 1.** A non-empty set $\Omega$ is called a *sample space*.

**Definition 2.** An element $\omega \in \Omega$ is called an *outcome*.

**Definition 3.** A *$\sigma$-algebra on a set* $\Omega$ is a collection $\mathcal{F}$ of subsets of $\Omega$ such that

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
3. If a countable collection $\{A_n\}_{n=1}^{\infty}$ satisfies $A_n \in \mathcal{F}$ for each $n = 1, 2, \ldots$, then $\left(\bigcup_{n=1}^{\infty} A_n\right) \in \mathcal{F}$

**Definition 4.** If $\Omega$ is a sample space and $\mathcal{F}$ is a $\sigma$-algebra of subsets of $\Omega$, then a set $A \in \mathcal{F}$ is called an *event*.

**Exercise 1.** Verify that if $\mathcal{F}$ is a $\sigma$-algebra then $\varnothing \in \mathcal{F}$ and if the countable collection $\{A_n\}$ satisfies $A_n \in \mathcal{F}$ for each $n = 1, 2, \ldots$ then $(\bigcap_{n=1}^{\infty} A_n) \in \mathcal{F}$.

**Exercise 2.** Verify that if $\mathcal{F}$ and $\mathcal{G}$ are $\sigma$-algebras, then $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ is a $\sigma$-algebra.

**Remark**. This result can be extended to intersections of arbitrary (not necessarily countable) intersections of $\sigma$-algebras.

**Warning.** If $\mathcal{F}$ and $\mathcal{G}$ are $\sigma$-algebras, $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is not necessarily a $\sigma$-algebra unless of course $\mathcal{G} \subset \mathcal{F}$ or vice versa. Indeed, even if $\{\mathcal{F}_n\}$ is a countable collection of $\sigma$-algebras satisfying $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for each $n = 1, 2, \ldots$, the union $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ is not necessarily a $\sigma$-algebra. For example, let $\Omega = \mathbb{N}$ and let $\mathcal{F}_n$ be the smallest $\sigma$-algebra containing $\{1\}, \{2\}, \ldots, \{n\}$. (This is the power set of $\{1, \ldots, n\}$ and all their complements in $\mathbb{N}$.) Then all $\{2n\}$ are in $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ but their union is not in any $\mathcal{F}_n$ so not in the union either.

**Proposition 1.** If $\mathbb{G}$ is an arbitrary collection of subsets of a set $\Omega$, then there exists a unique smallest extension to a $\sigma$-algebra, i.e. a set $\mathcal{G}$ such that

1. $\mathbb{G} \subset \mathcal{G}$
2. $\mathcal{G}$ is a $\sigma$-algebra
3. If $\mathcal{H}$ is a $\sigma$-algebra and $\mathbb{G} \subset \mathcal{H}$, then $\mathcal{G} \subset \mathcal{H}$.

In this case, we write $\mathcal{G} = \sigma(\mathbb{G})$.

**Proof.** Take the set of $\sigma$-algebras $\{\mathcal{F}_\alpha\}_{\alpha \in I}$ such that $\mathbb{G} \subset \mathcal{F}_\alpha$ for each $\alpha \in I$. This set is not empty since $2^\Omega$ is a member. Now define $\mathcal{G} = \bigcap_{\alpha \in I} \mathcal{F}_\alpha$. ■

**Example 1.** Consider $\mathbb{R}$ (or any set) with the Euclidean topology (or any other topology).

Then the smallest $\sigma$-algebra containing all the open sets is called the *Borel* $\sigma$-algebra and we denote it by $\mathcal{B}$.

**Definition 5.** Let $\mathcal{G}$ and $\mathcal{H}$ be $\sigma$-algebras. Then $\mathcal{G} \vee \mathcal{H} = \sigma\left(\mathcal{G} \cup \mathcal{H}\right)$.

Of course this definition can be extended to arbitrary unions, not just pairwise unions.

**Definition 6.** Let $\mathcal{F}$ be a $\sigma$-algebra. A (positive) *measure* is a function $\mu : \mathcal{F} \to \mathbb{R}_+ \cup \{+\infty\}$ such that

1. $\mu(\varnothing) = 0$
2. If $\{A_n\}$ is a countable collection of pairwise disjoint members of $\mathcal{F}$, then
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

**Definition 7.** A *measurable space* is a pair $(\Omega, \mathcal{F})$ where $\Omega$ is a non-empty set and $\mathcal{F}$ is a $\sigma$-algebra of subsets of $\Omega$.

**Definition 8.** A *measure space* is a triple $(\Omega, \mathcal{F}, \mu)$ where $\Omega$ is a non-empty set, $\mathcal{F}$ is a $\sigma$-algebra of subsets of $\Omega$ and $\mu : \mathcal{F} \to \mathbb{R}_+ \cup \{+\infty\}$ is a measure.

**Definition 9.** A *probability space* is a measure space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{P}(\Omega) = 1$. A set $A \in \mathcal{F}$ is called an *event*. An event $A$ is said to occur $\mathsf{P}$-*almost surely* or $\mathsf{P}$-*a.s.* if $\mathsf{P}(A) = 1$.

**Exercise 3.** Let $\{A_k\}_{k=1}^{\infty}$ be a seqence of events such that $A_k \subset A_{k+1}$ and define $A = \bigcup_{k=1}^{\infty} A_k$. (In this situation, we write $A_k \uparrow A$.) Show that
$$\mathsf{P}(A) = \lim_{k \to \infty} \mathsf{P}(A_k).$$

**Definition 10.** A *random variable* is a mapping $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable, i.e. such that $X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$ for each $a \in \mathbb{R}$.

**Definition 11.** A $\mathcal{G}$-measurable random variable (where $\mathcal{G} \subset \mathcal{F}$ is a $\sigma$-algebra) is a mapping $X : \Omega \to \mathbb{R}$ that is $\mathcal{G}$-measurable, i.e. such that $X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{G}$ for each $a \in \mathbb{R}$.

**Exercise 4.** Let $(\Omega, \mathcal{F})$ be a measurable space and let $X : \Omega \to \mathbb{R}$ be a function. Verify that the set $\mathbb{B}$ defined via

$$\mathbb{B} = \{B \subset \mathbb{R} : X^{-1}(B) \in \mathcal{F}\}$$

is a $\sigma$-algebra.

**Remark.** It follows from this exercise that we can define measurability in a number of equivalent ways. $X$ is $\mathcal{G}$-measurable just in case any of the following is true.

- $X^{-1}([a, b]) \in \mathcal{G}$ for all $a, b \in \mathbb{R}$
- $X^{-1}([a, b)) \in \mathcal{G}$ for all $a, b \in \mathbb{R}$
- $X^{-1}((a, b]) \in \mathcal{G}$ for all $a, b \in \mathbb{R}$
- $X^{-1}((a, b)) \in \mathcal{G}$ for all $a, b \in \mathbb{R}$
- $X^{-1}(B) \in \mathcal{G}$ for all Borel subsets $B$ of $\mathbb{R}$

**Remark.** We sometimes write the event $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ as $\{X \in B\}$ or sometimes just $X \in B$.

**Proposition 2.** If $X$ and $Y$ are random variables, then so is $Z = X + Y$.

**Proof.** Omitted. ■

**Proposition 3.** If $X$ is a random variable and $\alpha \in \mathbb{R}$, then so is $Z = \alpha X$.

**Proof.** Exercise. ■

**Proposition 4.** If $X$ and $Y$ are random variables, then so is $Z = XY$.

**Proof.** Omitted. ■

**Proposition 5.** Let $\{X_n\}$ be a sequence of random variables. Then $\overline{X}(\omega) = \limsup_{n\to\infty} X_n(\omega)$ is a random variable and so is $\underline{X}(\omega) = \liminf_{n\to\infty} X_n(\omega)$.

**Proof.** Omitted. ■

**Definition 12.** The *law* or *distribution* of a random variable $X$ is a probability measure on the Borel $\sigma$-algebra $\mathcal{B}$ on $\mathbb{R}$ defined via

$$\mu_X(B) = \mathsf{P}(X^{-1}(B)).$$

The most elementary random variable is called an indicator, defined as follows.

**Definition 13.** Let $A \in \mathcal{F}$. Then the random variable $I_A(\omega)$ is defined via

$$I_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}$$

**Definition 14.** A random variable with finite range is called *simple*.

**Proposition 6.** A simple $\mathcal{G}$-measurable random variable $X$ has the representation

$$X(\omega) = \sum_{k=1}^{n} a_k I_{A_k}(\omega) \tag{1}$$

where $a_k \in \mathbb{R}$ and $A_k \in \mathcal{G}$.

**Proof.** Let $a_1, a_2, \ldots, a_n$ be the finitely many values that $X$ takes and define $A_k = X^{-1}(\{a_k\})$. ■

**Exercise 5.** While it is obvious that a $\mathcal{G}$-measurable random variable with finite range has the representation (1), it is not obvious that any function defined via (1) with $A_k \in \mathcal{G}$ is $\mathcal{G}$-measurable. Nevertheless it is true. Prove it.

**Definition 15.** Given a random variable $X$, we denote by $\sigma(X)$ the smallest $\sigma$-algebra $\mathcal{G}$ such that $X$ is $\mathcal{G}$-measurable. By the result of Exercise 4, $\sigma(X)$ is simply the set of sets that can be written $X^{-1}(B)$ with $B$ a Borel subset of $\mathbb{R}$.

**Proposition 7.** Let $X$ be a $\mathcal{G}$-measurable random variable. Then there exists a sequence $\{X_n\}$ of $\mathcal{G}$-measurable simple functions such that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. If $X(\omega) \geq 0$ for all $\omega \in \Omega$ then the convergence can be made monotone, i.e. $X(\omega) \geq X_{n+1}(\omega) \geq X_n(\omega)$ for all $\omega$ and all $n$. In this case we write $X_n \uparrow X$.

**Proof.** Define a sequence of quantizer functions $q_n : \mathbb{R} \to \mathbb{R}$ via

$$q_n(x) = \begin{cases} n \text{ if } x \geq n \\ (k-1)2^{-n} \text{ if } (k-1)2^{-n} \leq x < k2^{-n};\ k = 1, 2, \ldots, n2^n \\ -(k-1)2^{-n} \text{ if } -k2^{-n} \leq x < -(k-1)2^{-n};\ k = 1, 2, \ldots, n2^n \\ -n \text{ if } x < -n \end{cases}$$

and define $X_n(\omega) = q_n(X(\omega))$. ■

**Definition 16.** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $X$ be a non-negative simple random variable with the representation

$$X = \sum_{k=1}^{n} a_k I_{A_k}.$$

Then its *expectation* is defined via

$$\mathsf{E}_{\mathsf{P}}[X] = \sum_{k=1}^{n} a_k \mathsf{P}(A_k)$$

where we usually suppress the subscript $\mathsf{P}$ where the choice of measure is clear from the context.

**Exercise 6.** The definition of the expected value of a non-negative simple random variable apparently depends on its precise representation. Show that this is appearance only, i.e. that if, for all $\omega \in \Omega$,

$$X(\omega) = \sum_{k=1}^{n} a_k I_{A_k}(\omega) = \sum_{k=1}^{m} b_k I_{B_k}(\omega)$$

then

$$\sum_{k=1}^{n} a_k \mathsf{P}(A_k) = \sum_{k=1}^{m} b_k \mathsf{P}(B_k).$$

**Definition 17.** Let $X$ be a non-negative random variable. Then its expectation is defined as follows. Let $F$ denote the set of $\mathcal{F}$-measurable simple random variables $\varphi$ such that $\varphi \leq X$. Then

$$\mathsf{E}[X] = \sup_{\varphi \in F} \mathsf{E}[\varphi]$$

where on the right hand side we invoke Definition 16.

**Remark.** Notice that Definitions 16 and 17 are equivalent whenever they both apply.

**Definition 18.** Let $X$ be a random variable and suppose $\mathsf{E}[X^+] < \infty$ and $\mathsf{E}[X^-] < \infty$.[1]

[1] By definition, $X^+(\omega) = \max\{X(\omega), 0\}$ and $X^-(\omega) = \max\{-X(\omega), 0\}$.

Then we say that $X$ is *integrable* and we define

$$\mathsf{E}[X] = \mathsf{E}[X^+] - \mathsf{E}[X^-].$$

**Remark.** This is just the definition of the Lebesgue integral, i.e.

$$\mathsf{E}[X] = \int_\Omega X(\omega) d\mathsf{P}(\omega).$$

and occasionally we will use this notation. But when we don't we will write $\mathsf{E}[X; A] = \mathsf{E}[I_A \cdot X]$ instead of the more conventional

$$\int_A X d\mathsf{P}.$$

When we integrate with respect to measures that are not necessarily probability measures, however, we will always use the more conventional notation.

A fundamental property of the expectation is that it is a linear operator so that

$$\mathsf{E}[\alpha X + \beta Y] = \alpha \mathsf{E}[X] + \beta \mathsf{E}[Y].$$

This property follows almost immediately from the definition, though it is interesting to note that it may fail if $X$ and/or $Y$ are not $\mathcal{F}$-measurable. In that case, one or the other of $X$ and $Y$ are not well approximated by $\mathcal{F}$-simple functions and then very strange things can happen.

**Exercise 7.** Let $\mu_X$ be the law of an integrable random variable. Show that

$$E[X] = \int_{-\infty}^{\infty} x d\mu_X.$$

We end this Section by recalling two fundamental facts about Lebesgue integrals.

**Proposition 8.** [Monotone convergence.] Let $X$ be a random variable and let $\{X_n\}$ be a sequence of non-negative random variables such that $X_n \uparrow X$ with probability 1. Then

$$\lim_{n\to\infty} \mathsf{E}[X_n] = \mathsf{E}[X].$$

**Remark.** The limit may be infinite, in which case $\mathsf{E}[X] = +\infty$ as well.

**Proof.** Omitted.

**Proposition 9.** [Dominated convergence.] Let $Y$ be an integrable random variable and let $\{X_n\}$ be a sequence of random variables such that $|X_n| \le Y$ and suppose $X_n$ converges to the random variable $X$ with probability one. Then

$$\lim_{n\to\infty} \mathsf{E}[X_n] = \mathsf{E}[X].$$

**Proof.** Omitted. ■

**Exercise 8.** Let $X$ be either integrable or non-negative. Suppose $\{A_n\}$ is a sequence of events such that $A_n \uparrow A$. Show that

$$\lim_{n\to\infty} \mathsf{E}[X; A_n] = \mathsf{E}[X; A].$$

**Exercise 9.** Let $X$ be integrable. Show that

$$\lim_{n\to\infty} \mathsf{E}[|X|; |X| > n] = 0.$$

**Definition 19.** If $p = 1, 2, \ldots$, then we denote by $\mathcal{L}^p(\Omega, \mathcal{F}, \mathsf{P})$ the set of $\mathcal{F}$-measurable random variables such that $\mathsf{E}[|X|^p] < \infty$ together with the norm

$$\|X\|_p = \mathsf{E}[|X|^p]^{1/p}.$$

**Exercise 10.** Suppose $X$ is an integrable random variable and that $\{Y_n\}$ is a sequence of uniformly bounded random variables, i.e. there is an $M \geq 0$ such that $|Y_n| \leq M$ for all $n = 1, 2, \ldots$. Suppose the event

$$\lim_{n\to\infty} Y_n(\omega) = X(\omega)$$

has probability 1. Show that

$$\lim_{n\to\infty} \mathsf{E}[|X - Y_n|] = 0$$

i.e. that $Y_n \to X$ in $\mathcal{L}^1$.

**Proposition 10.** In any measure space $(\Omega, \mathcal{F}, \mu)$ such that $\mu(\Omega) < \infty$, $\mathcal{L}^2 \subset \mathcal{L}^1$.

**Proof.** Let $X : \Omega \to \mathbb{R}$ be such that

$$\int_\Omega |X|^2 d\mu = \int_\Omega X^2 d\mu = M < \infty.$$

Apparently

$$\int_\Omega |X(\omega)| d\mu = \int_{|X|\leq 1} |X| d\mu + \int_{|X|>1} |X| d\mu \leq$$

$$\int_{|X|\leq 1} 1 d\mu + \int_{|X|>1} |X|^2 d\mu \leq \mu(|X| \leq 1) + M < \infty.$$

Another way to see the same thing is to help oneself to the Cauchy-Schwarz inequality, and simply note that

$$\int_\Omega |X| d\mu = \int_\Omega |X| \cdot 1 d\mu \leq \|X\| \cdot \sqrt{\mu(\Omega)} = \left(\int_\Omega |X|^2 d\mu\right)^{1/2} \cdot \sqrt{\mu(\Omega)}.$$

where the norm $\|\cdot\|$ is the norm on $\mathcal{L}^2(\Omega, \mathcal{F}, \mu)$.

**Exercise 11.** Verify that $\mathcal{L}^2$ is dense in $\mathcal{L}^1$.

**Definition 20.** Two events $A$ and $B$ are said to be *independent* if $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$.

**Definition 21.** Two $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$ are said to be independent if $\mathsf{P}(F \cap G) = \mathsf{P}(F)\mathsf{P}(G)$ for all $F \in \mathcal{F}$ and $\mathcal{G}$.

**Definition 22.** Two random variables $X$ and $Y$ are said to be independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

**Exercise 12.** Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathsf{P})$ be independent. Show that $\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$.

**Proposition 11.** [Chebyshev's inequality.] Let $X$ be a non-negative random variable and let $\varphi\colon \mathbb{R} \to \mathbb{R}_+$ be a non-decreasing function with $\varphi(x) > 0$ whenever $x > 0$ such that $\varphi(X)$ is integrable. Then, for each $\varepsilon > 0$,

$$\mathsf{P}(\{X(\omega) \geq \varepsilon\}) \leq \frac{1}{\varphi(\varepsilon)}\mathsf{E}[\varphi(X)].$$

**Proof.**

$$\mathsf{E}[\varphi(X)] \geq \mathsf{E}[\varphi(X); X \geq \varepsilon] \geq$$

$$\mathsf{E}[\varphi(\varepsilon); X \geq \varepsilon] = \varphi(\varepsilon)\mathsf{P}(X \geq \varepsilon).$$

■

# 2 Conditional expectations

## 2.1 Conditioning on an event

Suppose we know that the event $A$ has occurred and we want to know what to expect of a random variable $X$ given this information.

**Definition 23.** Let $X$ be an integrable random variable and $A$ an event such that $\mathsf{P}(A) > 0$. Then we define the number $\mathsf{E}[X|A]$ via

$$\frac{\mathsf{E}[X;A]}{\mathsf{P}(A)}.$$

If, on the other hand, $\mathsf{P}(A) = 0$ we leave $\mathsf{E}[X|A]$ undefined.

## 2.2 Conditioning on a measurable partition

Now suppose that we have a whole collection of sets that we know whether (or not) they have occurred. We want to define the conditional expectation as a *rule* whose value (prediction) is contingent on which of these known events occurred. To begin with, let this collection be a *measurable partition* of $\Omega$.

**Definition 24.** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. A *measurable partition* $\mathbb{P}$ of $\Omega$ is a finite collection of sets $\{A_1, A_2, \ldots, A_n\}$ such that

1. $A_k \in \mathcal{F}$ for all $k$
2. $A_j \cap A_k = \varnothing$ if $j \neq k$
3. $\bigcup_{k=1}^n A_k = \Omega$.

**Definition 25.** Let $X$ be a random variable and let $\mathbb{P}$ be a measurable partition of $\Omega$. Then $X$ is said to be $\mathbb{P}$-measurable if it is $\sigma(\mathbb{P})$-measurable.

**Exercise 13.** Let $X$ be a random variable and let $\mathbb{P}$ be a measurable partition of $\Omega$. Verify that $X$ is $\mathbb{P}$-measurable just in case it is constant on each element of the partition, i.e. if and only if $X(\omega) = X(\omega')$ whenever there is an $A \in \mathbb{P}$ such that $\{\omega, \omega'\} \subset A$.

**Definition 26.** Let $\mathbb{P} = \{A_1, A_2, \ldots, A_n\}$ be a measurable partition of $\Omega$ and let $X$ be an integrable random variable. Then we define the conditional expectation given $\mathbb{P}$ via

$$\mathsf{E}[X|\mathbb{P}] = \sum_{k=1}^{n} I_{A_k} \mathsf{E}[X|A_k].$$

**Remark 1** *If $\mathsf{P}(A_k) = 0$ for some values of $k$, then $\mathsf{E}[X|\mathbb{P}]$ is left undefined on those sets.*

**Exercise 14.** Let $X$ be an integrable random variable and $\mathbb{P}$ be a measurable partition of $\Omega$. Define $Z = \mathsf{E}[X|\mathbb{P}]$. Verify that $Z$ is $\mathbb{P}$-measurable and that for each $A \in \mathbb{P}$, we have

$$\mathsf{E}[X; A] = \mathsf{E}[Z; A].$$

## 2.3 Conditioning on a $\sigma$-algebra

Inspired by Exercise 14, we would like to define the conditional expectation of an integrable random variable $X$ given the $\sigma$-algebra $\mathcal{G}$ as a $\mathcal{G}$-measurable random variable $Z$ such that $\mathsf{E}[Z; G] = \mathsf{E}[X; G]$ for all $G \in \mathcal{G}$. However, at this stage we have no guarantee that such a random variable exists, so a digression on three key theorems is necessary: the Hilbert space projection theorem, the Riesz representation theorem and the Radon-Nikodym theorem. Before we start that endeavor, however, let's establish the basic concept by considering a measurable partition $\mathbb{P} = \{A_k\}_{k=1}^{n}$. A measure $\mu$ on $\mathbb{P}$, or for that matter on the $\sigma$-algebra generated by $\mathbb{P}$, is defined by the $n$ numbers

$$\mu_k = \mu(A_k).$$

Now let there be another measure $\lambda$. We now want to translate back and forth between these two measures. Might there exist a $\mathbb{P}$-simple function

$$f(\omega) = \sum_{k=1}^{n} a_k I_{A_k}$$

such that

$$\lambda(A_k) = a_k \mu(A_k) \tag{2}$$

for $k = 1, 2, \ldots, n$? Well, let's try to construct such a function. Define

$$a_k = \frac{\lambda(A_k)}{\mu(A_k)}.$$

This of course goes wrong if $\mu(A_k) = 0$, but even then things are not so bad if $\lambda(A_k) = 0$ also; we could then define $a_k$ arbitrarily, and Equation (2) would still hold. So if $\lambda(A_k) = 0$ whenever $\mu(A_k) = 0$ we say that $\lambda \ll \mu$ and declare that the rescaling function $f$ exists, is $\mathbb{P}$-measurable and is defined uniquely almost everywhere ($\mu$). We call this function the Radon-Nikodym derivative $\dfrac{d\lambda}{d\mu}$.

In one set of cases, *every* $\sigma$-algebra is generated by a measurable partition. This is when $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite set and $\mathcal{F}$ is its power set. The probability measure $\mathsf{P}$ is defined by the point masses $\mathsf{P}(\{\omega_k\}) = p_k$.

**Exercise 15.** Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite set and $\mathcal{F}$ be its power set. Let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra. Let $X$ be a random variable. Let the point masses be denoted by $p_k$. Describe the conditional expectation $\mathsf{E}[X|\mathcal{G}]$ as explicitly as possible and establish the connection to the Radon-Nikodym derivative.

A Hilbert space $(\mathscr{H}, (\cdot, \cdot))$ is a vector space associated with an inner product that is complete in the norm generated by this inner product. The details of the definition can be found in many textbooks.

**Proposition 12.** The space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathsf{P})$ with the inner product

$$(X, Y) = \mathsf{E}[X \cdot Y]$$

is a Hilbert space.

**Proof.** Omitted. ■

**Definition 27.** Let $\mathscr{H}$ be a Hilbert space and let $G \subset \mathscr{H}$ be a set. Then

$$G^{\perp} = \{y \in \mathscr{H} : (x, y) = 0 \text{ for all } x \in G\}.$$

**Theorem 1.** [The projection theorem] Let $\mathscr{H}$ be a Hilbert space and let $\mathscr{G} \subset \mathscr{H}$ be another Hilbert space. Then there are unique linear mappings $P : \mathscr{H} \to \mathscr{G}$ and $Q : \mathscr{H} \to \mathscr{G}^{\perp}$ such that $x = Px + Qx$ and $\|x - Px\| = \inf_{y \in \mathscr{G}} \|x - y\|$ for all $x \in \mathscr{H}$.

**Proof.** Omitted. ■

**Theorem 2.** [Riesz representation] Let $(\mathscr{H}, (\cdot, \cdot))$ be a Hilbert space and let $f : \mathscr{H} \to \mathbb{R}$ be linear and continuous ("a continuous linear functional"). Then there is a unique $y \in \mathscr{H}$ such that $f(x) = (x, y)$ for all $x \in \mathscr{H}$.

**Proof.** Define $M = \{x \in \mathscr{H} : f(x) = 0\}$ be the nullspace of $f$ and let $M^{\perp} = \{x \in \mathscr{H} : (x, y) = 0 \text{ for all } y \in M\}$. By the linearity of $f$, $M$ is a vector space. By the continuity of $f$, $M$ is closed. Hence $M$ is a Hilbert space. By the Hilbert space projection theorem, every $x \in \mathscr{H}$ can be written as $x = w + z$ where $w \in M$ and $z \in M^{\perp}$. Evidently (why?) $M^{\perp}$ is at most one-dimensional. If $M^{\perp} = \{0\}$ then $M = \mathscr{H}$ and $y = 0$. Otherwise let $y_0 \neq 0$ be a member of $M^{\perp}$. Every other $z \in M^{\perp}$ can be written as $z = \alpha y_0$ for some $\alpha \in \mathbb{R}$. In particular, $y = \alpha_0 y_0$. We want $f(y_0) = (y_0, y) = (y_0, \alpha_0 y_0) = \alpha_0 \|y_0\|^2$. So we choose

$$\alpha_0 = \frac{f(y_0)}{\|y_0\|^2}$$

i.e. choose

$$y = \frac{f(y_0)}{\|y_0\|^2} y_0.$$

The remaining details of the proof are left to the reader. ∎

**Definition 28.** A linear functional is said to be *bounded* if there is an $M > 0$ such that $|f(x)| \leq M\|x\|$ for all $x \in \mathscr{H}$.

**Proposition 13.** A linear functional is continuous if and only if it is bounded.

**Proof.** Exercise. ∎

**Definition 29.** Let $\lambda$ and $\mu$ be two measures with domain $\mathcal{F}$. We write $\lambda \ll \mu$ ($\lambda$ is absolutely continuous with respect to $\mu$) if $\lambda(A) = 0$ whenever $\mu(A) = 0$.

**Definition 30.** Let $(\Omega, \mathcal{F})$ be a measurable space. A mapping $\mu$ from $\mathcal{F}$ into $\mathbb{R} \cup \{+\infty\}$ or $\mathbb{R} \cup \{-\infty\}$ is called *signed measure* if

1. $\mu(\varnothing) = 0$
2. If $\{A_n\}$ is a countable collection of pairwise disjoint members of $\mathcal{F}$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Notice that $\mu$ attains at most one of the values $+\infty$ and $-\infty$.

**Theorem 3.** (Hahn decomposition) Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mu$ be a signed measure. Then there exist two sets $P, N \in \mathcal{F}$ such that

1. $P \cap N = \varnothing$

2. $P \cup N = \Omega$

3. For each $E \in \mathcal{F}$ such that $E \subset P$, $\mu(E) \geq 0$

4. For each $E \in \mathcal{F}$ such that $E \subset N$, $\mu(E) \leq 0$

**Proof.** Omitted. ∎

**Definition 31.** [Hahn-Jordan decomposition] Let $(\Omega, \mathcal{F})$ be a measurable space, let $\mu$ be a signed measure and let $P, N \in \mathcal{F}$ be a Hahn decomposition for $\mu$. Then we define, for each $E \in \mathcal{F}$,

$$\mu^+(E) = \mu(E \cap P)$$

and

$$\mu^-(E) = -\mu(E \cap N).$$

**Remark 2** *Notice that $\mu^+$ and $\mu^-$ are both positive measures and that $\mu = \mu^+ - \mu^-$.*

**Theorem 4.** [Radon-Nikodym, version 1] Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mu$ and $\lambda$ be finite positive measures such that $\lambda \ll \mu$. Then there exists an a.s. $(\mu)$ unique non-negative function $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ such that

$$\lambda(A) = \int_A f d\mu$$

for all $A \in \mathcal{F}$ and we write

$$f = \frac{d\lambda}{d\mu}.$$

**Lemma 1.** Let $(\Omega, \mathcal{F})$ be a measurable space, Let $\mu$ be a finite measure, let $f$, $g$ be measurable, non-negative real-valued functions, let $\lambda$ be a finite measure and suppose $f$, $g$ and $\lambda$ are such that

$$\int_A f d\lambda = \int_A g d\mu$$

for each $A \in \mathcal{F}$. Then

$$\int_A fh d\lambda = \int_A gh d\mu$$

for each $A \in \mathcal{F}$ and each measurable, non-negative real-valued function $h$.

**Proof** (of the lemma). Exercise. ■

**Proof** (of the theorem). Define a new measure via $\nu(A) = \mu(A) + \lambda(A)$. For any $g \in \mathcal{L}^2(\Omega, \mathcal{F}, \nu)$, we can define the linear functional

$$\Phi(g) = \int_\Omega g d\lambda.$$

By the triangle and Cauchy-Schwartz inequalities, we have

$$|\Phi(g)| \leq \left| \int_\Omega g d\lambda \right| \leq \int_\Omega |g| d\lambda \leq \int_\Omega |g| d\nu \leq \sqrt{\nu(\Omega)} \cdot \|g\|$$

where the norm of $g$ is that associated with the space $\mathcal{L}^2(\Omega, \mathcal{F}, \nu)$ so that $\Phi$ is bounded and hence continuous by Proposition 13. Hence by Theorem 2 there is an $h \in \mathcal{L}^2(\Omega, \mathcal{F}, \nu)$ such that

$$\int_\Omega g d\lambda = \int_\Omega gh d\nu \tag{3}$$

for all $g \in \mathcal{L}^2(\Omega, \mathcal{F}, \nu)$. By setting $g = I_A$ for an arbitrary $A \in \mathcal{F}$ and using the fact that $0 \leq \lambda(A) \leq \nu(A)$, we see that $0 \leq h \leq 1$. Now rewrite Equation (3) as

$$\int_\Omega g d\lambda = \int_\Omega gh d\lambda + \int_\Omega gh d\mu,$$

i.e.

$$\int_\Omega g(1-h) d\lambda = \int_\Omega gh d\mu \tag{4}$$

for all $g \in \mathcal{L}^2(\Omega, \mathcal{F}, \nu)$. In particular, it holds for all indicator functions. But then by Lemma 1 we have

$$\int_A d\lambda = \int_A \frac{h}{1-h}\, d\mu$$

for every $A \in \mathcal{F}$, provided $1/(1-h)$ is well-defined a.e. $(\lambda)$ and $(\mu)$. So we proceed to show that $h \neq 1$ a.e. $(\mu)$ and hence also $(\lambda)$. For that purpose, define $A = \{\omega \in \Omega : h(\omega) = 1\}$ and set $g = I_A$. From Equation (4), we obtain

$$\int_A h d\mu = \int_A (1-h) d\lambda$$

which implies that $\mu(A) = 0$. Since $\lambda \ll \mu$, it follows that $\lambda(A) = 0$ also. We can then, with a good conscience, define

$$f = \frac{h}{1-h}$$

and this function is non-negative since $0 \leq h \leq 1$ as we have seen. For integrability, notice that

$$\lambda(\Omega) = \int_\Omega f d\mu < \infty$$

by the finiteness of $\lambda$. ■

**Definition 32.** A measure $\mu$ on $(\Omega, \mathcal{F})$ is said to be *$\sigma$-finite* if there exists a countable collection $\{B_n\}$ of members of $\mathcal{F}$ such that

1. $|\mu(B_n)| < \infty$ for all $n$
2. $\bigcup_n B_n = \Omega$

**Theorem 5.** [Radon-Nikodym, version 2] Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mu$ be a $\sigma$-finite measure and let $\lambda$ be a finite[2] measure such that $\lambda \ll \mu$. Then there exists an

[2] If $\lambda$ is merely $\sigma$-finite, then we may lose integrability of $f$, but we still have existence and $\mathcal{F}$-measurability. This result is omitted only because the proof is a bit more complicated.

a.s. $(\mu)$ unique $\mathcal{F}$-measurable function $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ such that

$$\lambda(A) = \int_A f d\mu$$

for all $A \in \mathcal{F}$.

**Proof.** Let $\{B_n\}$ be a countable measurable covering of $\Omega$ such that each $B_n$ has finite measure under $\mu$. Define $\lambda_n(A) = \lambda(A \cap B_n)$ for each $A \in \mathcal{F}$. On each $B_n$, define $f_n$ as the Radon-Nikodym derivative $\frac{d\lambda_n}{d\mu}$. Define $f = f_n$ on $B_n$ for each $n$ and the proof is done. ■

**Exercise 16.** Verify that $f$ in the previous proof is measurable. What if $\{B_n\}$ is uncountable?

**Example 1** *Let $\Omega = [0, 1]$ and let $\mathcal{F} = \mathcal{B}$ be the Borel $\sigma$-algebra generated by the Euclidean topology. Let $\mu$ be the counting measure and $m$ be the Lebesgue measure. Apparently $m \ll \mu$ but there is no $f$ such that $dm = f d\mu$.*

**Theorem 6.** [Radon-Nikodym, version 3] Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mu$ be a $\sigma$-finite measure and let $\lambda$ be a finite signed measure (both the negative and the positive parts are finite) such that $\lambda \ll \mu$. Then there exists an a.s. $(\mu)$ unique $\mathcal{F}$-measurable function $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ such that

$$\lambda(A) = \int_A f d\mu$$

for all $A \in \mathcal{F}$.

**Proof.** Take the Hahn-Jordan decomposition $\lambda = \lambda^+ - \lambda^-$ and apply Theorem 4 to $\lambda^+$ and to $\lambda^-$, yielding two Radon-Nikodym derivatives; call them (without abuse of notation!) $f^+$ and $f^-$. For integrability, notice that

$$\lambda^+(\Omega) = \int_\Omega f^+ d\mu < \infty$$

and

$$\lambda^-(\Omega) = \int_\Omega f^- d\mu < \infty$$

by assumption. ■

With the Radon-Nikodym theorem in hand, we can define the conditional expectation via the following recipe.

**Proposition 14.** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathsf{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra. Then there exists a $Z \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ such that $\mathsf{E}[Z; G] = \mathsf{E}[X; G]$ for all $G \in \mathcal{G}$. This $Z$ is a.s. *(*$\mathsf{P}$*)* unique and we denote it by $\mathsf{E}[X|\mathcal{G}]$.

**Proof.** Apparently $(\Omega, \mathcal{G})$ is a measurable space and $\mathsf{P}_\mathcal{G}$, the restriction of $\mathsf{P}$ to $\mathcal{G}$, is a finite measure; from now on, abusing the notation somewhat, we will call it $\mathsf{P}$. Now define the finite signed measure $\mu\colon \mathcal{G} \to \mathbb{R}$ via

$$\mu(G) = \mathsf{E}[X; G]$$

and apparently $\mu \ll \mathsf{P}$ on $(\Omega, \mathcal{G})$. By the Radon-Nikodym theorem, there exists an essentially unique $Z \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ such that

$$\mu(G) = \mathsf{E}[Z; G].$$

■

**Definition 33.** Let $\mathcal{G}$ be a $\sigma$-algebra and let $A \in \mathcal{F}$ be an event. Its *conditional probability* is defined via

$$\mathsf{P}[A|\mathcal{G}] = \mathsf{E}[I_A|\mathcal{G}].$$

**Exercise 17.** Let $\mathcal{G} \subset \mathcal{H}$ be two $\sigma$-algebras and let $X$ be an integrable random variable. Verify the law of iterated expectations, i.e. that

$$\mathsf{E}\left[\mathsf{E}[X|\mathcal{H}]|\mathcal{G}\right] = \mathsf{E}\left[X|\mathcal{G}\right].$$

**Exercise 18.** Let $X$ and $Y$ be square integrable, let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra and suppose $Y$ is $\mathcal{G}$-measurable. Then

$$\mathsf{E}[XY|\mathcal{G}] = Y\mathsf{E}[X|\mathcal{G}].$$

**Exercise 19.** Let $X$ be an integrable random variable and suppose the $\sigma$-algebras $\mathcal{G}$ and $\sigma(X)$ are independent. Show that $\mathsf{E}[X|\mathcal{G}] = \mathsf{E}[X]$. Hence (or otherwise) verify that $\mathsf{E}[X|\{\varnothing, \Omega\}] = \mathsf{E}[X]$.

## 2.4 Conditioning on a random variable

**Definition 34.** Let $Z$ be an integrable random variable and let $X$ be an arbitrary random variable. Then we define

$$\mathsf{E}[Z|X] = \mathsf{E}[Z|\sigma(X)].$$

Preferably, though, we would like to give precise meaning to the following expression: $\mathsf{E}[Z|X = x]$. For this we need the following proposition.

**Proposition 15.** Let $X$ be a random variable and let $Y$ be a $\sigma(X)$-measurable random variable. Then there exists a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $Y(\omega) = f(X(\omega))$ for all $\omega \in \Omega$.

**Proof.** Let $\{Y_n\}$ be a sequence of $\sigma(X)$-measurable simple random variables such that

$\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$ for each $\omega \in \Omega$. Fix $n$ and let $\{a_1, a_2, \ldots, a_N\}$ be the range of $Y_n$, where without loss of generality we assume that $a_j \neq a_k$ whenever $j \neq k$. Form the sets $A_k = Y_n^{-1}(\{a_k\})$ and the sets $B_k = X(A_k)$. By the $\sigma(X)$-measurability of $Y_n$ and the distinctness of the $a_k$:s, the $B_k$:s are pairwise disjoint. Hence we can define $f_n(x) = a_k$ on $B_k$ and zero elsewhere. Finally, define $f(x) = \lim_{n\to\infty} f_n(x)$ wherever the limit exists and 0 elsewhere. ■

Using this result, we define $Y = \mathsf{E}[Z|X]$ and define $\mathsf{E}[Z|X = x] = f(x)$.

**Exercise 20.** Suppose $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathsf{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra. Without using the Hilbert space projection theorem, show that $Z^* = \mathsf{E}[X|\mathcal{G}]$ solves

$$\min_{Z\in\mathcal{L}^2(\Omega,\mathcal{G},\mathsf{P})} \mathsf{E}\left[(X - Z)^2\right].$$

Hint: Start by showing that $\mathsf{E}[Z(X - Z^*)] = 0$ for each $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathsf{P})$.

## 2.5 Alternative definition of the conditional expectation

The material so far suggests that there is an alternative approach to defining the conditional expectation.

**Definition 35.** Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathsf{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra. Then $\mathsf{E}[X|\mathcal{G}]$ is the projection of $X$ on $\mathcal{L}^2(\Omega, \mathcal{G}, \mathsf{P})$.

**Exercise 21.** If $X \notin \mathcal{L}^2(\Omega, \mathcal{F}, \mathsf{P})$ but $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathsf{P})$ then let $\{X_n\}$ be a sequence in $\mathcal{L}^2$ that converges to $X$ in $\mathcal{L}^1$. (Such a sequence exists by Exercise 11.) Now define the sequence $Z_n = \mathsf{E}[X_n|\mathcal{G}]$. Verify that this sequence converges to a limit $Z$ in $\mathcal{L}^1(\Omega, \mathcal{G}, \mathsf{P})$. (This is of course our definition of $\mathsf{E}[X|\mathcal{G}]$.)

# 3 The normal distribution

**Definition 36.** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $X : \Omega \to \mathbb{R}^n$ be a random vector. This vector is said to be *normally distributed* with mean $\mu$ and (non–singular) variance matrix $\Sigma$ if, for each Borel set $A \subset \mathbb{R}^n$,

$$\mathsf{P}(X^{-1}(A)) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \int_A \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} dm(x)$$

where $m$ is Lebesgue measure on $\mathbb{R}^n$. If $\Sigma$ is singular, then, with probability 1, $X$ is confined to a subspace.

A nice thing about normal vectors is that the conditional expectation function is linear in the following sense. Suppose

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}$$

is a normal vector with mean

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

and variance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}.$$

Then the conditional expectation function is linear, i.e. there exists a matrix $M$ such that

$$\mathsf{E}[Y|X] = \mu_y + M(X - \mu_x).$$

We can use the (Hilbert space) projection theorem to compute $M$. Setting the prediction error orthogonal to all the elements of $X$, we get

$$\mathsf{E}[(X-\mu_x)(Y-\mu_y-M(X-\mu_x))^T] = 0$$

which implies

$$\Sigma_{xy} = \Sigma_{xx} M^T$$

and it follows that, if $\Sigma_{xx}$ is invertible,

$$M = \Sigma_{xy}^T \Sigma_{xx}^{-1}.$$

Thus

$$\mathsf{E}[Y|X] = \mu_y + \Sigma_{xy}^T \Sigma_{xx}^{-1}(X - \mu_x).$$

Incidentally, this formula gives the best (in a mean square error sense) linear predictor even if $Z$ is not normal. This is also a consequence of the Hilbert space projection theorem.

# 4 Further reading

See Rudin (1987) and Chung (2001).

# References

Chung, K. L. (2001). *A Course in Probability Theory, Third Edition.* Academic Press.

Rudin, W. (1987). *Real and Complex Analysis.* McGraw-Hill.