## Measurable sets and functions

**Definition 1 (\sigma-algebras)** Let  $\Omega$  be a set and  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  be a collection of subsets of  $\Omega$ : i. A is an algebra if  $\varnothing \in A$  and for  $A, B \in A$ ,  $A^c = \Omega \setminus A \in A$  and  $A \cup B \in A$ . ii. A is a  $\sigma$ -algebra if  $\varnothing \in A$ , for  $A \in A$ ,  $A^c \in A$ , and for  $(A_n)$  a sequence of sets in A,  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$ 

A collection of sets is an algebra subject to being closed under finite applications of the basic operators. The  $\sigma$ -algebra concept extends this slightly to infinite ones. Consider where this distinction is relevant?

Note that if we have  $\{\mathcal{F}_i : i \in I\}$  are  $\sigma$ -algebras, then

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$$

is a  $\sigma$ -algebra. This allows us to consider the notion of a smallest  $\sigma$ -algebra containing a set (the  $\sigma$ -algebra 'generated' by a set). We write the  $\sigma$ -algebra generated by a collection of collections of sets  $\mathfrak{A}$  as  $\sigma(\mathfrak{A})$ .

**Definition 2 (Borel**  $\sigma$ -algebra) Let  $(E, \mathcal{T})$  be a topological space. The  $\sigma$ -algebra generated by the open sets in E is called the Borel  $\sigma$ -algebra on E and is denoted  $\mathcal{B}(E) = \sigma(\mathcal{T})$ .

**Definition 3** Suppose  $(\Omega_i, \mathcal{F}_i)_{i \in I}$  are measurable spaces. With  $\Omega = \prod_{i \in I} \Omega_i$ ,  $\mathcal{F}$  the  $\sigma$ -algebra generated by  $A = \prod_{i \in I} A_i$  where  $A_i \in \mathcal{F}_i$  for all  $i \in I$  and for all but finitely many  $i \in I$ ,  $A_i = \Omega_i$ :  $(\Omega, \mathcal{F})$  is the product space.

This space is measurable, and  $\mathcal{F}$  is a  $\sigma$ -algebra.

**Definition 4 (\pi and \lambda-systems)** A collection of sets A is called a  $\pi$ -system if it is closed under intersections.

A collection of sets  $\mathcal{M}$  is called a  $\lambda$ -system if  $\Omega \in \mathcal{M}$ , if  $A, B \in \mathcal{M}$ ,  $A \subseteq B$ , then  $B \setminus A \in \mathcal{M}$ , and if  $(A_n) \subseteq \mathcal{M}$  with  $A_n \subseteq A_{n+1}$  increasing then  $\bigcup_{n>1} A_n \in \mathcal{M}$ .

A collection of sets is a  $\sigma$ -algebra if and only if it is both a  $\pi$ -system and a  $\lambda$ -system.

**Lemma 1** ( $\pi$ - $\lambda$  systems lemma) Let  $\mathcal{A}$  be a  $\pi$ -system and  $\mathcal{M}$  a  $\lambda$ -system. Then if  $\mathcal{A} \subseteq \mathcal{M}$ then  $\sigma(\mathcal{A}) \subseteq \mathcal{M}$ .

We can use this with a convenient  $\pi$ -system to show that our  $\lambda$ -system contains more than is immediately obvious.

Let  $\lambda(\mathcal{A})$  be the smallest  $\lambda$ -system containing  $\mathcal{A}$ . This is a subset of  $\mathcal{M}$  and  $\sigma(\mathcal{A})$ , so we just need to show that  $\lambda(\mathcal{A})$  is a  $\sigma$ -algebra (for which we just have to show that it is a  $\pi$ -system).

**Definition 5 (Random variables)** With measurable spaces  $(\Omega, \mathcal{F})$ ,  $(E, \mathcal{E})$ , a function  $f: \Omega \to \mathcal{F}$ E is said to be an E-valued random variable (or a measurable function) if for all  $A \in \mathcal{E}$ ,  $f^{-1}(A) \in \mathcal{F}$ .

We get immediately that random variables can be composed as one would expect. We can also use random variables to define new  $\sigma$ -algebras. Note that  $(\Omega, \{f^{-1}(A) : A \in \mathcal{E}\})$  is a  $\sigma$ -algebra.

**Definition 6** With  $\{f_i: i \in I\}$  a family of functions  $\Omega \to E$ ,  $\sigma(f_i: i \in I)$  is the smallest  $\sigma$ -algebra on  $\Omega$  for which all  $f_i$  are measurable.

This is initially a slightly intimidating definition, but the intuition is just that we need our  $\sigma(f_i: i \in I) = \sigma(f_i^{-1}(A): A \in \mathcal{E}, i \in I).$ 

**Theorem 2 (Monotone Class Theorem)** Let  $\mathcal{H}$  be a class of bounded functions from  $\Omega \to \mathbb{R}$ such that

•  $\mathcal{H}$  is a vector space over  $\mathbb{R}$ , • the constant function  $1 \in \mathcal{H}$ ,

• if  $(f_n) \subseteq \mathcal{H}$ ,  $f_n \to f$  monotonically increasing, then  $f \in \mathcal{H}$ ,

then if  $\mathcal{C} \subseteq \mathcal{H}$ , and  $\mathcal{C}$  is closed under pointwise multiplication, then all bounded  $\sigma(\mathcal{C})$ measurable functions are in  $\mathcal{H}$ .

To get an intuition for this, note that any  $f \in \mathcal{C}$  is necessarily bounded and  $\sigma(\mathcal{C})$ -measurable, but the converse is not immediate. Thus we essentially get a statement of the  $\lambda$ - $\pi$  systems lemma but for functions on analogous systems.

We can firstly see that  $\mathcal{H}$  is closed in  $\mathcal{F}_b(\Omega)$ . Then, we can prove the statement for the special case of  $\mathcal{C} = \{\chi_A : A \in \mathcal{A}\}$  for a  $\pi$ -system  $\mathcal{A}$ , then adding 1 to  $\mathcal{C}$  without loss of generality we can make the proof more general (concretely, because  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{C} \cup \{1\})$ ).

It may allow this theorem to make more sense to note that  $\lambda$ -systems are sometimes referred to as 'monotone classes'. Thus the  $\pi$ - $\lambda$  systems lemma can be seen as saying that for  $\mathcal{A}$  a  $\pi$ -system, the (smallest) monotone class generated by  $\mathcal{A}$  is  $\sigma(\mathcal{A})$ .

We can use the monotone class theorem to demonstrate that for  $f:\Omega_1\times\Omega_2\to\mathbb{R}$  is measurable, then fixing  $\omega_1 \in \Omega_1$ ,  $\omega_2 \mapsto f(\omega_1, \omega_2)$  is measurable.

## Conditional Probability

Up until presently, we've considered the notion of event A conditioned on event B as having a fixed probability. This doesn't entirely capture what a conditional is however – we're conditioning on the amount of information we have, and therefore we want the conditional probability to change as a function of our information. In particular, we want our conditional probability to be a function of  $\omega \in \Omega$ , and in order to reflect conditioning as a reflection of information, we want to condition over events in a  $\sigma$ -algebra, rather than individual events.

Doing more algebra, we see that expectation is a more fitting operator, leading us to the following definition:

**Definition 7 (Conditional expectation)** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,  $\mathcal{G} \subseteq \mathcal{F}$  a  $\sigma$ -algebra. A random variable  $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$  is (a version of) the conditional expectation of X given  $\mathcal{G}$  if for  $G \in \mathcal{G}$ ,

 $\mathbb{E}[Y\chi_G] = \mathbb{E}[X\chi_G].$ The key aspect of this statement can be rewritten as

 $\int_C \mathbb{E}[X \mid \mathcal{G}] d\mathbb{P} = \int_C X d\mathbb{P},$ 

which allows us to carry over all of our normal integration properties to conditional expectations.

**Theorem 3** The conditional expectation of X given  $\mathcal{G}$  exists, denoted  $\mathbb{E}[X | \mathcal{G}]$ , and if Z is also the conditional expectation of X given  $\mathcal{G}$ , then

 $Z = \mathbb{E}[X \mid \mathcal{G}]$  a.s.

Come back to the proof of this.

Note importantly how conditional expectations behave with respect to measurability. If we have X a  $\mathcal{G}$ -measurable random variable, then

 $\mathbb{E}[X | \mathcal{G}] \stackrel{\text{a.s.}}{=} X.$ Meanwhile if  $\sigma(X)$  and  $\mathcal{G}$  are independent, then

 $\mathbb{E}[X \mid \mathcal{G}] \stackrel{\text{a.s.}}{=} \mathbb{E}[X].$ 

**Lemma 4 (Tower property)** Take  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,  $\mathcal{F}_1$ ,  $\mathcal{F}_2$  both  $\sigma$ -algebras, satisfying  $\mathcal{F}_1 \subseteq$  $\mathcal{F}_2 \subseteq \mathcal{F}$ . Then

 $\mathbb{E}\Big[\mathbb{E}ig[X\,|\,\mathcal{F}_2ig]\,|\,\mathcal{F}_1\Big] \stackrel{ ext{a.s.}}{=} \mathbb{E}ig[X\,|\,\mathcal{F}_1ig].$ 

This should be relatively intuitive –  $\mathbb{E}[X | \mathcal{F}_2]$  contains more information than can be represented in  $\mathcal{F}_1$ , but is fundamentally still expressing a reduced form of X, which can be reduced more to give  $\mathbb{E}[X | \mathcal{F}_1]$ .

One can also consider this as a commutativity statement: as  $\mathbb{E}[X | \mathcal{F}_1]$  is  $\mathcal{F}_2$ -measurable, thus  $\mathbb{E}[X | \mathcal{F}_1] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_1] | \mathcal{F}_2]$ , so the tower property is stating that with  $\mathcal{F}_1 \subseteq \mathcal{F}_2$ :  $\mathbb{E}\Big[\mathbb{E}\big[X\,|\,\mathcal{F}_1\big]\,|\,\mathcal{F}_2\Big]\stackrel{\mathrm{a.s.}}{=}\mathbb{E}\Big[\mathbb{E}\big[X\,|\,\mathcal{F}_2\big]\,|\,\mathcal{F}_1\Big].$ 

I'm tempted to claim the more general statement, that for  $\mathcal{F}_1$ ,  $\mathcal{F}_2$  both  $\sigma$ -algebras in  $\mathcal{F}$ :

 $\mathbb{E}\left|\mathbb{E}\left[X\,|\,\mathcal{F}_1
ight]\,|\,\mathcal{F}_2
ight|\overset{\mathrm{a.s.}}{=}\mathbb{E}\left[X\,|\,\mathcal{F}_1\cap\mathcal{F}_2
ight].$ This statement is true if  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are independent, because then both sides are equal to  $\mathbb{E}[X]$ , but it seems that there could be a 'middle-ground' between independence and containment for which commutativity stops holding.

On the other side, attempting to prove this statement, the tripping point is that it's unclear that the LHS is  $\mathcal{F}_1$ -measurable (although clearly it is  $\mathcal{F}_2$ -measurable).

**Lemma 5** Take X, Y random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$  with X, Y, and XY integrable. Then  $\mathbb{E}[XY \mid \sigma(Y)] \stackrel{\text{a.s.}}{=} Y \mathbb{E}[X \mid \sigma(Y)].$ 

Ensure for yourself that it's clear why this implies the same holding for  $\mathcal{G} \supset \sigma(Y)$  instead of  $\sigma(Y)$ .

## $\hspace{0.1in} \textbf{Measures on} \hspace{0.1in} \mathbb{R}$

**Definition 8** A measure space is a triple  $(\Omega, \mathcal{F}, \mu)$  such that  $\Omega$  is a set,  $\mathcal{F}$  is a  $\sigma$ -algebra on  $\Omega$ , and  $\mu: \mathcal{F} \to [0, \infty]$  is countably additive ( $\mu$  is then a measure on  $(\Omega, \mathcal{F})$ ).

**Definition 9** Let  $\mu$  be a probability measure on  $\mathcal{B}(\mathbb{R})$ . The distribution function of  $\mu$  is  $F_{\mu}(x) = \mu(-\infty, x]$ , where we require that  $F_{\mu}$  is non-decreasing, tends to 0 as  $x \to -\infty$ , to 1 as  $x \to \infty$ , and is right continuous.

**Definition 10** We say that  $\nu$  is absolutely continuous with respect to  $\mu$ ,  $\nu \ll \mu$ , if for any  $A \in \mathcal{F}, \ \mu(A) = 0 \ implies \ that \ \nu(A) = 0.$  Further, we say that  $\mu$  and  $\nu$  are equivalent,  $\mu \sim \nu$ , if  $\mu \ll \nu$  and  $\nu \ll \mu$ .

Extensions

For the most part, it's difficult to characterise a measure explicitly, due to  $\sigma$ -algebras being incredibly large in all but countable  $\Omega$ . We therefore wish to characterise them in terms of their value on algebras.

**Theorem 6 (Uniqueness of extension)** Let  $\mu_1$  and  $\mu_2$  be measures on a space  $(\Omega, \mathcal{F})$ , and  $\mathcal{A} \subseteq \mathcal{F}$  is a  $\pi$ -system with  $\sigma(\mathcal{A}) = \mathcal{F}$ . Then if  $\mu_1(\Omega) = \mu_2(\Omega) < \infty$  and  $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}}$ , then  $\mu_1 = \mu_2$ .

This follows immediately via the  $\lambda$ - $\pi$  systems lemma.

**Theorem 7 (Carathéodory Extension theorem)** Let  $\Omega$  be a set and A an algebra on  $\Omega$ , then with  $\mu_0: \mathcal{A} \to [0,\infty]$  a countably additive set function, there exists a measure  $\mu:$  $\sigma(\mathcal{A}) \to [0, \infty] \text{ such that } \mu|_{\Lambda} = \mu_0.$ 

One can derive this from defining the outer measure  $\mu^*$  in terms of  $\mu_0$ , and claiming that a set is measurable iff for all  $E \subseteq \Omega$ ,  $\mu^*(E) = \mu^*(E \cap B) + \mu^*(E \setminus B)$ . We can then prove that this gives the smallest  $\sigma$ -algebra containing  $\mathcal{A}$ .

**Definition 11 (Distribution function)** If a function  $F : \mathbb{R} \to [0, 1]$  satisfies:

i. F is non-decreasing; ii.  $F(x) \to 0$  as  $x \to -\infty$ ,  $F(x) \to 1$  as  $x \to \infty$ ; and

iii. F is continuous from the right,

then F is a distribution function.

**Theorem 8** Let F be a distribution function. Then there exists a unique Borel probability measure  $\mu$  on  $\mathbb{R}$  such that  $\mu(-\infty,x]=F(x)$ . Further, every Borel probability measure on  $\mathbb{R}$ defines a distribution function.

A corollary of this is that there is a unique Borel measure such that for all  $a < b \in \mathbb{R}$ ,  $\mu(a, b) = b - a$ .

This result demonstrates that there is a bijection between measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and distribution functions. In particular, we call these measures the Lebesgue-Stieltjes measures.

The proof of this theorem follows using both of the extension theorems. In particular, we use the algebra of left open right closed intervals.

**Definition 12 (Pushforward measure)** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $X : \Omega \to E$ , then for  $A \in \mathcal{E}$ ,

 $\mathbb{Q}(E) := \mathbb{P}(X^{-1}(E))$ 

Thus we have a (very non-injective) map from random variables in E to probability measures on E. This is useful, on the basis that many properties of random variables will just be properties of the pushforward measure rather than the random variable itself.

**Theorem 9** For  $\{(\Omega_i, \mathcal{F}_i, \mathbb{P}_i) : i \in \{1, \dots, n\}\}$  a set of probability measures, there is a unique measure  $\mathbb{P}$  on  $\left(\prod_{i=1}^n \Omega_i, \underset{i=1}{\overset{n}{\times}} \mathcal{F}_i\right)$  such that for  $E_i \in \mathcal{F}_i$  with  $i \in \{1, \ldots, n\}$ ,

 $\mathbb{P}\left(\prod^n E_i\right) = \prod^n \mathbb{P}_i(E_i).$ 

It's hopefully natural here that one should aim an induction proof.

The theorem here then allows us to extend the matter to infinite products, although at this point we require that what we're dealing with are probability measures (to keep each term in [0,1] for convergence reasons), rather than just finite measures as could work with the previous statement of the theorem.

## Independence

**Definition 13 (Independence)** With  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space,  $(\mathcal{G}_i)_{i=1}^n$  a collection of  $\sigma$ algebras, these  $\sigma$ -algebras are independent if for  $E_i \in \mathcal{G}_i$  for  $i \in \{1, \ldots, n\}$ 

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

Further, an arbitrary collection  $(G_i)_{i\in I}$  of  $\sigma$ -algebras is independent if any finite subset of the collection is independent.

Note that this means  $\{\emptyset, \Omega\}$  is independent of anything else.

Additionally, we say that a set  $(X_i)_{i\in I}$  of random variables is independent iff  $(\sigma(X_i))_{i\in I}$  is independent.

This definition requires a bit of work to deal with properly. One of the best general results we can attain quickly gives a fairly applicable result:

**Theorem 10** With  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space,  $(\mathcal{A}_i)_{i \in I}$  an arbitrary collection of  $\pi$ -systems, then  $(\sigma(A_i))_{i\in I}$  are independent iff for any finite  $J\subseteq I$ ,  $A_i\in A_i$  for  $i\in J$ :

$$\mathbb{P}\left(\bigcap A_i\right) = \prod \mathbb{P}(A_i)$$

We also have the result that for any independent set of  $\sigma$ -algebras, any subset is also independent.

It takes a small bit of proving, but from the above results we get the lemma: **Lemma 11** With  $(\Omega, \mathcal{F}, \mathbb{P})$ , a family of independent random variables  $X_i : \Omega \to E_i$ , measurable functions  $f_i: E_i \to \mathbb{R}$  for  $i \in I$ , then  $(f(X_i))_{i \in I}$  are independent.

Tail events **Definition 14** For a sequence of random variables  $(X_n)$ , the tail  $\sigma$ -algebra is defined as

uence of random variables 
$$(X_n)$$
, the tail  $\sigma$ -algebra is defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\{X_k : k > n\})$$

The intuition here is that all events in the tail  $\sigma$ -algebra contain sample information distinguishing the results of functions of infinitely many sequence elements.

**Theorem 12 (Kolmogorov's 0-1 Law)** Let  $(X_n)$  be a sequence of independent random variables. Then the tail  $\sigma$ -algebra of  $(X_n)$  contains only events with probability 0 or 1.

To see this, we demonstrate that  $\mathcal{T}$  is independent of a  $\sigma$ -algebra containing it, and therefore

Borel-Cantelli lemmas **Definition 15** With  $(A_n)$  a sequence of sets from  $\mathcal{F}$ :

that all of its events are independent with themselves.

$$\limsup_{n \to \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{m \ge n} A_m$$

 $= \{ \omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \}$ 

 $= \{A_n \text{ infinitely often}\}$ and  $\liminf_{n\to\infty} A_n = \bigcup \bigcap A_m$ 

 $= \{ \omega \in \Omega : \omega \in A_n \ eventually \}$  $= \{A_n \ eventually\}$ 

**Lemma 13 (Fatou and Reverse Fatou for sets)** With  $(A_n)$  a sequence of sets in  $\mathcal{F}$ ,

 $\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n)$ 

 $\mathbb{P}(\limsup A_n) \ge \limsup \mathbb{P}(A_n).$ 

**Lemma 14 (First Borel-Cantelli lemma)** For  $(A_n)$  a sequence of events in  $\mathcal{F}$ , if  $\sum \mathbb{P}(A_n) < \infty,$ 

then  $\mathbb{P}(A_n \ i.o.) = 0$ .

**Lemma 15 (Second Borel-Cantelli lemma)** For  $(A_n)$  a sequence of independent events in

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty,$$

then  $\mathbb{P}(A_n \ i.o.) = 1$ .

 $\mathcal{F}$ , if

By their nature, the BC lemmas are only informative in relation to almost sure events. While this may seem incredibly limited, by Kolmogorov's 0-1 Law, it turns out that many events of interest are in fact almost sure events.

Integration

As already covered in Part A Integration, we define integration as normal: **Definition 16 (Integral on simple functions)** For a measure space  $(\Omega, \mathcal{F}, \mu)$ ,  $f: \Omega \to [0, \infty]$ a non-negative simple function taking values  $\{a_1, \ldots, a_n\} \subseteq \mathbb{R}$ :

$$\int f \, d\mu = \sum_{i=1}^{n} a_i \mu \left( f^{-1} \left( \{ a_i \} \right) \right).$$

**Definition 17 (Integral on non-negative functions)** For  $f: \Omega \to [0, \infty]$  a non-negative measurable function:

$$\int f = \sup \left\{ \int g \, \mathrm{d}\mu : g \, simple \,, \, 0 \le g \le f \right\}$$

**Definition 18 (Integral)** For  $f: \Omega \to \mathbb{R}$  a measurable function and  $\int |f| d\mu < \infty$ , we write

$$: \Omega \to \mathbb{R} \ a \ measurable \ function$$
), and

 $f^+ = \max(f, 0), f^- = -\min(f, 0), and$ 

$$\int f \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu$$

**Theorem 16 (Monotone convergence theorem)** For  $(f_n)$  a sequence of non-negative functions measurable on  $(\Omega, \mathcal{F}, \mu)$ , such that  $f_n \to f$  monotonically. Then  $\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu.$ 

**Theorem 17 (Fatou's lemma)** For  $(f_n)$  a sequence of non-negative functions measurable on  $(\Omega, \mathcal{F}, \mu),$ 

$$\int \liminf_{n \to \infty} f_n \, \mathrm{d}\mu \le \liminf_{n \to \infty} \int f_n \, \mathrm{d}\mu$$

**Lemma 18 (Reverse Fatou's lemma)** For  $(f_n)$  a sequence of non-negative functions measurable on  $(\Omega, \mathcal{F}, \mu)$ , assume that there is an integrable function g such that  $f_n \leq g$  for  $n \geq 1$ . Then

$$\int \limsup_{n \to \infty} f_n \, \mathrm{d}\mu \ge \limsup_{n \to \infty} \int f_n \, \mathrm{d}\mu$$

**Theorem 19 (Dominated convergence theorem)** For  $(f_n)$  a sequence of functions measurable on  $(\Omega, \mathcal{F}, \mu)$  with  $f_n \to f$  pointwise. Assume that there is an integrable function g such that  $|f_n| \leq g$  for  $n \geq 1$ . Then

$$\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu.$$

**Theorem 20** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $X : \Omega \to E$ ,  $g : E \to \mathbb{R}$  measurable on their respective spaces. Then g is  $(\mathbb{P} \circ X^{-1})$ -integrable iff  $g \circ X$  is  $\mathbb{P}$ -integrable. Further,

 $\int_{\mathbb{T}} g(x) d(\mathbb{P} \circ X^{-1})(x) = \int_{\mathbb{R}} g(X(\omega)) d\mathbb{P}(\omega).$ **Definition 19 (Expectation)** For  $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ,

We say that X admits an nth moment if  $X \in \mathcal{L}^n(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definition 20 (Variance)** For  $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ ,

 $\operatorname{Var}(X) = \mathbb{E}\left[ (X - \mathbb{E}[X])^2 \right].$ 

$$\mathbb{E}\left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}}\right)^{n}\right]$$

**Theorem 21 (Fubini-Tonelli)** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the product of probability spaces  $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$  for

**Definition 21 (Standardised moment)** If  $X \in \mathcal{L}^n(\Omega, \mathcal{F}, \mathbb{P})$ , the nth standardised moment of

$$i \in \{1, 2\}, \ and \ f : \Omega \to \mathbb{R} \ is \ a \ bounded \ measurable \ function. \ Then \ both$$
 
$$x \mapsto \int_{\Omega_2} f(x, y) \, \mathrm{d}\mathbb{P}_2(y)$$
 
$$and \quad y \mapsto \int f(x, y) \, \mathrm{d}\mathbb{P}_1(x)$$

are measurable (respectively in  $\mathcal{F}_1$  and  $\mathcal{F}_2$ ).

X is

If either  $f \geq 0$  or f is  $\mathbb{P}$ -integrable over  $\Omega$ , then  $\int_{\Omega} f \, d\mathbb{P} = \int_{\Omega} \int_{\Omega} f(x, y) \, d\mathbb{P}_{1}(x) \, d\mathbb{P}_{2}(y) = \int_{\Omega} \int_{\Omega} f(x, y) \, d\mathbb{P}_{2}(y) \, d\mathbb{P}_{1}(x)$ 

ntegral of a fixed non-negative measurable function over the set being measured.

Radon-Nikodym theorem ntegration as we've defined it gives a canonical method of defining a measure on a space: as the

Concretely: with a measure space  $(\Omega, \mathcal{F}, \mu)$ , a measurable function  $f: \Omega \to [0, \infty], A \in \mathcal{F}$ ,  $\nu(A) := \int_A f \,\mathrm{d}\mu$ 

is a measure on  $\mathcal{F}$  (via MCT). We therefore want to characterise how often a measure can be characterised in this way (in

for which we have a wealth of tools). **Theorem 22 (Radon-Nikodym theorem)** Let  $\mu$ ,  $\nu$  be two probability measures on a  $\sigma$ algebra  $(\Omega, \mathcal{F})$ . Then  $\nu \ll \mu$  if and only if there is a measurable function  $f: \Omega \to [0, \infty]$ 

particular, whether we can get this the case with respect to leb or the counting measure, both

$$\nu(A) = \int_A f \,\mathrm{d}\mu.$$

Further,  $\nu \sim \mu$  if and only if  $\mu(f^{-1}(\{0\})) = \nu(f^{-1}(\{0\})) = 0$ .

We call f the radon-nikodym derivative of  $\nu$  with respect to  $\mu$ ,  $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ . If  $\nu \sim \mu$ , then  $\frac{1}{f} = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ .

This means that providing leb(A) = 0 implies that  $\nu(A) = 0$ , we can construct  $\nu$  in this way. Using this theorem, we can define for  $A, B \in \mathcal{F}$  the conditional distribution  $\mathbb{P}(A \mid B)$ , provided  $\mathbb{P}(B) > 0$ . If  $\mathbb{P}(A) = 0$ , then  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(\mathbb{R})} = 0$ , so there is some  $f_B : \Omega \to [0, \infty]$  measurable such

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \int f_B(\omega) \, \mathrm{d}\mathbb{P}(\omega).$$

We now consider modes of convergence of random variables using our results in integration.

verges to X:

as  $n \to \infty$ .

 $\varphi(\lambda) < \infty$  we have

i. If  $X_n \stackrel{\text{a.s.}}{\to} X$  then  $X_n \stackrel{\mathbb{P}}{\to} X$ .

such that for  $A \in \mathcal{F}$ ,

**Definition 22 (** $\mathcal{L}^p$  **spaces)** For  $p \geq 0$ ,  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R} \text{ measurable s.t. } \mathbb{E}[|X|^p] < \infty\}$ In particular,  $\mathcal{L}^0$  is the space of all random variables, and  $\mathcal{L}^{\infty}$  is the space of random variables

which are bounded almost surely. Note that for  $0 \le p < 1$ ,  $\mathcal{L}^p$  is not a normed space. **Definition 23** For a sequence  $(X_n)$  of random variables over  $(\Omega, \mathcal{F}, \mathbb{P})$ , we say that  $X_n$  con-

i. almost surely  $(X_n \stackrel{\text{a.s.}}{\rightarrow} X \text{ or } X_n \rightarrow X \text{ a.s.})$  if  $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1.$ 

ii. in probability 
$$(X_n \stackrel{\mathbb{P}}{\to} X)$$
 if for all  $\varepsilon > 0$ 

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0$$

as  $n \to \infty$ .

iii. in 
$$\mathcal{L}^p$$
  $(X_n \stackrel{\mathcal{L}^p}{\to} X)$  if  $X_n \in \mathcal{L}^p$  for  $n \ge 1$  and 
$$\mathbb{E}[|X_n - X|^p] \to 0$$

$$as \ n \to \infty.$$
 $iv. \ weakly \ in \ \mathcal{L}^1 \ if \ X_n \in \mathcal{L}^1 \ for \ n \ge 1 \ and \ for \ all \ Y \in \mathcal{L}^\infty$ 

$$\mathbb{E}[X_nY] o \mathbb{E}[XY]$$

$$as\ n \to \infty.$$
 $v.\ in\ distribution\ (X_n \stackrel{d}{\to} X)\ if\ for\ x \in \mathbb{R}\ such\ that\ F_X\ is\ continuous,$ 

**Theorem 23** For a sequence  $(X_n)$  of random variables,

ii. If  $X_n \stackrel{\mathbb{P}}{\to} X$  then there is a subsequence  $(X_{n_k})$  such that  $X_{n_k} \stackrel{\text{a.s.}}{\to} X$ .

**Lemma 24 (Markov's inequality)** Let 
$$(\Omega, \mathcal{F}, \mathbb{P})$$
 be a probability space,  $X$  a non-negative random variable. Then for  $\lambda > 0$ ,

 $F_{X_n}(x) \to F_X(x)$ 

 $\mathbb{P}(X \ge \lambda) \le \frac{\mathbb{E}[X]}{\lambda}$ Corollary 1 (General Chebyshev's Inequality) For a measurable set  $A \subseteq \mathbb{R}, X : \Omega \to A$ a random variable,  $\varphi:A\to [0,\infty]$  an increasing measurable function. For  $\lambda\in A$  with

$$\mathbb{P}(X \ge \lambda) \le \frac{\mathbb{E}[\varphi(X)]}{\varphi(X)}.$$

This allows us to then demonstrate that for p > 0,  $X_n \stackrel{\mathcal{L}^p}{\to} X$  implies  $X_n \stackrel{\mathbb{P}}{\to} X$ .

Further, we can also show the weak law of large numbers: Corollary 2 For  $(X_n)$  be a sequence of i.i.d. random variables with mean  $\mu$ , variance  $\sigma^2$ , then

$$\frac{1}{n} \sum_{k=1}^{n} X_k \to \mu$$

as  $n \to \infty$ .

**Theorem 25 (Jensen's inequality)** Let  $f: I \to \mathbb{R}$  be a convex function on an interval  $I \subseteq \mathbb{R}$ . If  $X:\Omega\to I$  is an integrable random variable then  $\mathbb{E}[f(X)] \ge f(\mathbb{E}[X]).$ 

For considering the  $\mathcal{L}^p$  spaces, we define  $\|\cdot\|_p := (\mathbb{E}[|X|^p])^{1/p}$ . We can note immediately that for  $0 \leq p \leq q, \mathcal{L}^q \subseteq \mathcal{L}^p$ .

Aside from this, note all the standard results regarding  $\mathcal{L}^p$  spaces, in particular Hölder's inequality. A particular application of use is to note that for  $1 < p, q < \infty$  with 1/p + 1/q = 1, and for x > 0,  $x\mathbb{P}(X \ge x) \le \mathbb{E}[Y\chi_{X \ge x}],$ 

then  $||X||_p \le q||Y||_p$ .

**Definition 24 (Uniform integrability)** A collection C of random variables is called uniformly integrable (UI) if