Measurable sets and functions **Definition 1 (\sigma-algebras)** Let Ω be a set and $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a collection of subsets of Ω : 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. \mathcal{A} is a σ -algebra if $\emptyset \in \mathcal{A}$, for $A \in \mathcal{A}$, $A^c \in \mathcal{A}$, and for (A_n) a sequence of sets in \mathcal{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 σ -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 σ -algebras, then

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

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

Definition 3 (Product space) Suppose $(\Omega_i, \mathcal{F}_i)_{i \in I}$ are measurable spaces. With $\Omega =$ $\prod_{i\in I}\Omega_i$, \mathcal{F} the σ -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 $i \in I$, $A_i = \Omega_i$: (Ω, \mathcal{F}) is the product space.

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

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

A collection of sets \mathcal{M} is called a λ -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 σ -algebra if and only if it is both a π -system and a λ -system.

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

We can use this with a convenient π -system to show that our λ -system contains more than is immediately obvious.

Let $\lambda(\mathcal{A})$ be the smallest λ -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 σ -algebra (for which we just have to show that it is a π -system).

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

Note that this statement is equivalent to taking f as a random variable in $(f(\Omega), \{A \cap f(\Omega) :$

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

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

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

The following result is critical for our understanding of the measurability of random variables on these σ -algebras.

Theorem 2 ($\sigma(X)$ -measurability as functional dependence) With $X : (\Omega, \mathcal{F}) \rightarrow$ $(E,\mathcal{E}), Y: (\Omega,\mathcal{F}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ random variables, Y is $\sigma(X)$ -measurable if and only if $Y = f \circ X$ for some random variable $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

To construct this we use a standard method of beginning with simple functions, and then using the density of \mathcal{E} -simple functions in \mathcal{E} -measurable functions under the pointwise topology, we can construct f in general.

Noting that if X_1, \ldots, X_n are measurable on $(\Omega, \mathcal{F}), (X_1, \ldots, X_n)$ is also measurable (start with showing the measurability property on the normal π -system generating $\mathcal{B}(\mathbb{R}^n)$, then use π - λ systems lemma), this statement generalises much further than this, to include sequences of random variables as well.

Theorem 3 (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 λ - π 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 π -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 λ -systems are sometimes referred to as 'monotone classes'. Thus the π - λ systems lemma can be seen as saying that for \mathcal{A} a π -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 σ -algebra, rather than individual events.

Doing more algebra, we see that expectation is a more fitting operator, leading us to the **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 $G \in \mathcal{G}$,

 $\mathbb{E}[Y\chi_G] = \mathbb{E}[X\chi_G].$

Both aspects of this statement are key. \mathcal{G} -measurability entails that our conditional expectation is a random variable whose value provides no more information about \mathcal{F} than is precisely contained in \mathcal{G} . That this random variable actually represents the expectation of X subject to additional information is then represented by the equality of expectations on events in \mathcal{G} , often rewritten as:

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

which allows us to carry over all of our normal integration properties to conditional expecta-

In particular, note that for random variables Y_1, \ldots, Y_n , $\mathbb{E}[X \mid \sigma(Y_1, \ldots, Y_n)]$ = $F(Y_1,\ldots,Y_n)$ for some measurable $F:\mathbb{R}^n\to\mathbb{R}$, by representing measurability as functional dependence. In order to retain our normal definition, one can see that for an event $A \in \mathcal{F}$, $\mathbb{E}[X \mid \sigma(A)]$

may be modified on a null set to be constant on the sets $\{A, A^c\}$. This allows us, in applied settings, to select arbitrarily for each $A \in \mathcal{F}$ some $\omega \in A$ so that $\mathbb{E}[X \mid A] = \mathbb{E}[X \mid \sigma(A)](\omega).$

Theorem 4 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. While the uniqueness proof is relatively straightforward, the existence proof requires some methods from functional analysis. In particular, we prove the statement for $L^2(\Omega, \mathcal{F}, \mathbb{P})$, on noting that this space is Hilbert, and $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace, allowing us to write any $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ as X = Y + Z with $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})^{\perp}$, $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. We

then extend this proof to L^1 by clamping the positive and negative parts of X by n, and demonstrating by extending the MCT to conditionals that we can then define Z.

Note importantly how conditional expectations behave with respect to measurability. If we have X a \mathcal{G} -measurable random variable, then $\mathbb{E}\big[X\,|\,\mathcal{G}\big]\stackrel{\mathrm{a.s.}}{=} X.$

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

ontainment for which commutativity stops holding.

Lemma 5 (Tower property) Take $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, \mathcal{F}_1 , \mathcal{F}_2 both σ -algebras, satisfying $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}$. Then

 $\mathbb{E}\left[\mathbb{E}\left[X\,|\,\mathcal{F}_2
ight]|\,\mathcal{F}_1
ight]\stackrel{ ext{a.s.}}{=}\mathbb{E}\left[X\,|\,\mathcal{F}_1
ight].$ 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 $\mathbb{E}[X \mid \mathcal{F}_1]$ 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 σ -algebras in \mathcal{F} : $\mathbb{E}\left[\mathbb{E}\left[X\,|\,\mathcal{F}_1
ight]\,|\,\mathcal{F}_2
ight]\stackrel{\mathrm{a.s.}}{=}\mathbb{E}\left[X\,|\,\mathcal{F}_1\cap\mathcal{F}_2
ight].$ 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

that the LHS is \mathcal{F}_1 -measurable (although clearly it is \mathcal{F}_2 -measurable).

On the other side, attempting to prove this statement, the tripping point is that it's unclear **Lemma 6** Take X, Y random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with X, Y, and XY integrable.

 $\mathbb{E}\big[XY \mid \sigma(Y)\big] \stackrel{\text{a.s.}}{=} Y \mathbb{E}\big[X \mid \sigma(Y)\big].$ Ensure for yourself that it's clear why this implies the same holding for $\mathcal{G} \supseteq \sigma(Y)$ instead of $\blacksquare \quad \mathbf{Measures \ on} \ \mathbb{R}$

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

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

Definition 10 (Absolute continuity and equivalence) We say that ν is absolutely continuous with respect to μ , $\nu \ll \mu$, if for any $A \in \mathcal{F}$, $\mu(A) = 0$ implies that $\nu(A) = 0$. Further, we say that μ and ν 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 σ -algebras being incredibly large in all but countable Ω . We therefore wish to characterise them in terms of their value on algebras.

Theorem 7 (Uniqueness of extension) Let μ_1 and μ_2 be measures on a space (Ω, \mathcal{F}) , and $\mathcal{A} \subseteq \mathcal{F}$ is a π -system with $\sigma(\mathcal{A}) = \mathcal{F}$. Then if $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ and $\mu_1|_{\mathcal{A}} = |\mathcal{A}|_{\mathcal{A}}$ $|\mu_2|_A$, then $\mu_1 = \mu_2$.

This follows immediately via the λ - π systems lemma.

Theorem 8 (Carathéodory Extension theorem) Let Ω be a set and A an algebra on Ω , then with $\mu_0: \mathcal{A} \to [0, \infty]$ a countably additive set function, there exists a measure $\mu: \sigma(\mathcal{A}) \to [0, \infty] \text{ such that } \mu|_{\mathcal{A}} = \mu_0.$

One can derive this from defining the outer measure μ^* in terms of μ_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 σ -algebra containing \mathcal{A} .

Note additionally that the restriction of any measure to a smaller σ -algebra remains a measure **Lemma 9** Let (Ω, \mathcal{F}) be a measurable space and (μ_n) a sequence of probability measures on \mathcal{F} . With a sequence of positive numbers (a_n) such that $\sum_{n=1}^{\infty} a_n = 1$,

is a probability measure on (Ω, \mathcal{F}) .

iii. F is continuous from the right,

then F is a distribution function.

This requires Fubini's theorem, which is given later on.

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

Theorem 10 Let F be a distribution function. Then there exists a unique Borel probability measure μ 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, $X : \Omega \to E$,

 $\mathbb{Q}(E) := \mathbb{P}(X^{-1}(E))$ Thus we have a (very non-injective) map from random variables in E to probability measures

If $either\ f \geq 0$ or f is \mathbb{P} -integrable over Ω , then 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 11** 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 $\left(\prod \Omega_i, \times \mathcal{F}_i\right)$ such that for $E_i \in \mathcal{F}_i$ with $i \in \{1, \ldots, n\}$,

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

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

then for $A \in \mathcal{E}$,

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

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

Note that this means $\{\varnothing, \Omega\}$ 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

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

We also have the result that for any independent set of σ -algebras, any subset is also

It takes a small bit of proving, but from the above results we get the lemma: **Lemma 13** With $(\Omega, \mathcal{F}, \mathbb{P})$, a family of independent random variables $X_i : \Omega \to E$ 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 σ -algebra is defined $\mathcal{T} = \bigcap \sigma(\{X_k : k > n\})$

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

Theorem 14 (Kolmogorov's 0-1 Law) Let (X_n) be a sequence of independent random variables. Then the tail σ -algebra of (X_n) contains only events with probability 0 or

To see this, we demonstrate that \mathcal{T} is independent of a σ -algebra containing it, and therefore that all of its events are independent with themselves.

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

 $\limsup A_n = \bigcap \bigcup A_m$ $= \{ \omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \}$ $= \{A_n \text{ infinitely often}\}$ and $\liminf A_n = \bigcup \bigcap A_m$

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

Lemma 15 (Fatou and Reverse Fatou for sets) With (A_n) a sequence of sets in \mathcal{F} , $\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n)$ $\mathbb{P}(\limsup A_n) \ge \limsup \mathbb{P}(A_n).$

 $= \{A_n \ eventually\}$

Lemma 16 (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 17 (Second Borel-Cantelli lemma) For (A_n) a sequence of independent

 $\sum \mathbb{P}(A_n) = \infty,$

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

events of interest are in fact almost sure events.

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

Integration

As already covered in Part A Integration, we define integration as normal: **Definition 16 (Integral on simple functions)** For a measure space $(\Omega, \mathcal{F}, \mu)$, $f: \Omega \rightarrow$ $[0,\infty]$ a non-negative simple function taking values $\{a_1,\ldots,a_n\}\subseteq\mathbb{R}$: $\int f \,\mathrm{d}\mu = \sum_{i} a_i \mu \left(f^{-1} \left(\{a_i\} \right) \right).$

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

 $\int f \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu$ **Theorem 18 (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 $\int f_n d\mu \to \int f d\mu$.

Theorem 19 (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 20 (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

 $\lim \sup f_n \, \mathrm{d}\mu \ge \lim \sup \int f_n \, \mathrm{d}\mu$

Theorem 21 (Dominated convergence theorem) For (f_n) a sequence of functions mea-

surable on $(\Omega, \mathcal{F}, \mu)$ with $f_n \to f$ pointwise. Assume that there is an integrable function

Theorem 22 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X : \Omega \to E$, $g : E \to \mathbb{R}$ measurable on

 $\mathbb{E}[X] = \int X \, \mathrm{d}\mathbb{P}.$

g such that $|f_n| \leq g$ for $n \geq 1$. Then $f_n d\mu \rightarrow \int f d\mu$.

their respective spaces. Then g is $(\mathbb{P} \circ X^{-1})$ -integrable iff $g \circ X$ is \mathbb{P} -integrable. Further, $g(x) d(\mathbb{P} \circ X^{-1})(x) = \int g(X(\omega)) d\mathbb{P}(\omega).$ **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 $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$,

 $ment\ of\ X\ is$

 $\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2 \right].$ **Definition 21 (Standardised moment)** If $X \in \mathcal{L}^n(\Omega, \mathcal{F}, \mathbb{P})$, the nth standardised mo-

 $\mathbb{E}\left[\left(\frac{X-\mathbb{E}[X]}{\sqrt{1+|X|}}\right)^n\right]$

Theorem 23 (Fubini-Tonelli) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the product of probability spaces $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ for $i \in \{1, 2\}$, and $f : \Omega \to \mathbb{R}$ is a bounded measurable function. Then

 $x \mapsto \int f(x,y) d\mathbb{P}_2(y)$

and $y \mapsto \int f(x,y) d\mathbb{P}_1(x)$ are measurable (respectively in \mathcal{F}_1 and \mathcal{F}_2).

 $\int f \, d\mathbb{P} = \int \int f(x,y) \, d\mathbb{P}_1(x) \, d\mathbb{P}_2(y) = \int \int f(x,y) \, d\mathbb{P}_2(y) \, d\mathbb{P}_1(x)$

Radon-Nikodym theorem Integration as we've defined it gives a canonical method of defining a measure on a space: as the integral of a fixed non-negative measurable function over the set being measured. Concretely: with a measure space $(\Omega, \mathcal{F}, \mu)$, a measurable function $f: \Omega \to [0, \infty], A \in \mathcal{F}$,

 $\nu(A) := \int f \,\mathrm{d}\mu$ is a measure on \mathcal{F} (via MCT).

We therefore want to characterise how often a measure can be characterised in this way (in particular, whether we can get this the case with respect to leb or the counting measure, both for which we have a wealth of tools).

Theorem 24 (Radon-Nikodym theorem) Let μ , ν be two probability measures on a σ -algebra (Ω, \mathcal{F}) . Then $\nu \ll \mu$ if and only if there is a measurable function $f: \Omega \to \mathbb{R}$ $[0,\infty]$ such that for $A \in \mathcal{F}$, $\nu(A) = \int f \, \mathrm{d}\mu.$

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 ν with respect to μ , $f = \frac{d\nu}{1}$. If $\nu \sim \mu$, then weaker property than convergence in probability.

This means that providing leb(A) = 0 implies that $\nu(A) = 0$, we can construct ν 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 $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0$, so there is some $f_B : \Omega \to [0, \infty]$

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

Uniform Integrability

measurable such that

We've now introduced the notions of convergence in \mathcal{L}^1 , and seen that it implies convergence in probability. We'd quite like to go the other direction however, and find a sufficient

condition for which the two are the same. **Definition 22 (Uniform integrability)** A collection C of random variables is called uniformly integrable (UI) if

 $\lim_{N \to \infty} \sup \mathbb{E}[|X|\chi_{|X|>N}] = 0$ Importantly, this is a property of collections, rather than individual random variables. The larger our collection is, the less likely it is to hold (and conversely, if \mathcal{C} is UI, then $\mathcal{D} \subseteq \mathcal{C}$ is also UI). We can see immediately therefore that this is a property that can only hold for

collections of random variables in \mathcal{L}^1 , by considering the singleton sets which can be UI. If we have a $Y \in \mathcal{L}^1$ such that for $X \in \mathcal{C}$, $|X| \leq Y$, then \mathcal{C} is uniformly integrable. This is one of the most common methods for demonstrating a collection is UI.

Another useful characterisation is found by reframing $|X|\chi_{|X|>N}$. From the below fact, we can see that we can replace it with $(|X|-N)^+$ in the definition of uniform integrability without effect: $0 \le (|X| - N)^{+} \le |X| \chi_{|X| \ge N} \le 2(|X| - N/2)^{+}.$

common formulation that allows us to avoid having to determine the sets on which $|X| \geq N$ is given below: **Lemma 25** Let \mathcal{C} be a family of random variables. Then \mathcal{C} is UI iff $\sup \mathbb{E}[|X|] < \infty$

 $\sup \left\{ \mathbb{E} \left[|X|\chi_A \right] : X \in \mathcal{C}, A \in \mathcal{F}, \mathbb{P}(A) \leq \delta \right\} \to 0$

iii. $X \in \mathcal{L}^1$ and $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$ as $n \to \infty$.

with convergence in probability.

Theorem 26 (Vitali's convergence theorem) Take (X_n) a sequence of integrable random variables which converge in probability to a random variable X. The following are i. $\{X_n : n \geq 1\}$ is uniformly integrable. $ii. X \in \mathcal{L}^1 \text{ and } \mathbb{E}[|X_n - X|] \to 0 \text{ as } n \to \infty.$

It's worth noting that (ii) is the definition of convergence in \mathcal{L}^1 , from which we get (iii) just then $||X||_p \leq q||Y||_p$. via the reverse triangle inequality. The difficult part of this proof is demonstrating that (iii) entails (i) which requires that characterisation of UI in terms of $(|X| - N)^+$ in conjunction

By a non-examinable result (Dunford-Pettis), a collection is UI iff its closure is compact in $\sigma(L^1, L^{\infty})$, which is the weak topology on L^1 .

Theorem 27 Take $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $\{\mathcal{F}_i : i \in I\}$ a family of σ -algebras in \mathcal{F} . Then

 $\{\mathbb{E}[X | \mathcal{F}_i] : i \in I\}$ is uniformly integrable. This result follows via Jensen's inequality, and provides a critical example of UI families – particularly useful for considering backwards martingales later on.

Filtrations

Definition 23 (Filtrations) For $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, a filtration (\mathcal{F}_n) is an sequence of σ -algebras ($\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for $n \geq 1$). We then call $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ a filtered probability space.

Definition 24 (Adapted stochastic process) $Take(X_n)$ a sequence of random variables,

 (\mathcal{F}_n) a filtration. If for $n \geq 1$, X_n is \mathcal{F}_n measurable, then (X_n) is said to be adapted to

Note that for a stochastic process (X_n) , we get a natural filtration $\mathcal{F}_n := \sigma\left(\left\{X_k : k \le n\right\}\right).$

This is the smallest filtration possible. It's worth noting that with $\mathcal{F}_{\infty} = \bigcup_{n>1} \mathcal{F}_n$, we don't necessarily have that $\mathcal{F}_{\infty} = \mathcal{F}$ Nonetheless, once we're working primarily within the filtered space, this shouldn't make a significant amount of difference, because the remainder of events will be inaccessible to any

adapted process. **Definition 25 (Stopping time)** Take $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ a filtered probability space. A ran-

dom variable $\tau:\Omega\to\mathbb{N}\cup\{\infty\}$ is a stopping time with respect to (\mathcal{F}_n) if for $n\geq 1$, $\tau^{-1}(\{n\}) \in \mathcal{F}_n$.

Immediate examples of stopping times can be given. Any constant is a stopping time, as

are the maxima or minima of any two stopping times. Further, the first hitting time for an

adapted stochastic process on any measurable set is a stopping time. **Definition 26** Let τ be a stopping time on $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$. The σ -algebra of information

at time τ is $\mathcal{F}_{\tau} = \{ A \in \mathcal{F}_{\infty} : for \ n \geq 1, \ A \cap \{ \tau = n \} \in \mathcal{F}_n \}.$ This definition isn't entirely straightforward to understand, but the key aspect is this: if we have the value of τ (say we know that $\omega \in \{\tau = n\}$), then we shouldn't be able to infer from $\omega \in A \in \mathcal{F}_{\tau}$ that $\omega \in B \in \mathcal{F}_{\infty} \setminus \mathcal{F}_{n}$. The above definition should then be clear as containing the events which are possible to have observed by time τ ('time' being defined in

We can then get our intuition that if for stopping times τ and ρ , $\tau \leq \rho$, then $\mathcal{F}_{\tau} \subseteq \mathcal{F}_{\rho}$ (we get more information if we have longer to discover it).

Note now that for (X_n) an adapted process, τ a stopping time, $X^\tau := (X_{\min(n,\tau)})$ is a 'stopped process', and is adapted to both the filtration $(\mathcal{F}_{\min(n,\tau)})$ and hence (\mathcal{F}_n) .

The filtration $(\mathcal{F}_{\min(n,\tau)})$ comes up often, and can be conceived essentially as the information be. This is what we mean by τ being a stopping time, in that it stops the observations

Convergence

converges to X:

as $n \to \infty$.

continuing.

accordance to steps in the filtration).

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

Definition 27 (\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}^{∞} is the space of random

variables which are bounded almost surely. In these notes I'll refer almost exclusively to \mathcal{L}^p , although it's worth noting that there's a nuance here: the well-behaved space we generally want to refer to for useful results is not \mathcal{L}^p , but rather $L^p := \mathcal{L}^p/\mathcal{N}$, where $\mathcal{N} := \{X \in \mathcal{L}^0 : \mathbb{E}[|X|] = 0\}$. At the same time, in this course it's not particularly desirable to be working with $X + \mathcal{N}$ constantly (not least

therefore we principally use \mathcal{L}^p . An exception to the comments above comes for $0 \le p < 1$. L^p for these values is almost entirely useless, as it is not a normed space. \mathcal{L}^p does have some use however, in particular

because there are some instances where a property being 'just' almost sure is relevant), and

for p=0, in that it allows us to specify the set of measurable functions. **Definition 28** For a sequence (X_n) of random variables over $(\Omega, \mathcal{F}, \mathbb{P})$, we say that X_n

 $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1.$ ii. in probability $(X_n \stackrel{\mathbb{P}}{\to} X)$ if for all $\varepsilon > 0$

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

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

i. almost surely $(X_n \stackrel{\text{a.s.}}{\rightarrow} X \text{ or } X_n \rightarrow X \text{ a.s.})$ if

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

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

as $n \to \infty$. Of these, notion (v) of convergence in distribution is the odd one out, as it is independent of any particular instance of a random variable with the same distribution. This is a strictly

both than weak convergence in \mathcal{L}^1 and convergence in probability.

Convergence in \mathcal{L}^p is an identical notion to that in functional analysis, and (in a mathematical sense) leans very much into the measure-based notion of random variables. It is stronger

Theorem 28 For a sequence (X_n) of random variables, i. If $X_n \stackrel{\text{a.s.}}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$. ii. If $X_n \stackrel{\mathbb{P}}{\to} X$ then there is a subsequence (X_{n_k}) such that $X_{n_k} \stackrel{\text{a.s.}}{\to} X$.

Useful results **Lemma 29 (Markov's inequality)** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a nonnegative random variable. Then for $\lambda > 0$, $\mathbb{P}(X \ge \lambda) \le \frac{\mathbb{E}[X]}{\lambda}$

 $\Omega \to A$ a random variable, $\varphi: A \to [0,\infty]$ an increasing measurable function. For $\lambda \in A \text{ with } \varphi(\lambda) < \infty \text{ we have}$

Corollary 1 (General Chebyshev's Inequality) For a measurable set $A \subseteq \mathbb{R}$, X

 $\mathbb{P}(X \ge \lambda) \le \frac{\mathbb{E}[\varphi(X)]}{\varphi(X)}.$ 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 μ , variance $\frac{1}{n}\sum X_k \to \mu$

Theorem 30 (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 $\mathbb{E}[f(X)] \ge f(\mathbb{E}[X]).$

The proof of this utilises the relation for $x, m, y \in I$, x < m < y, $f(m) \le \frac{y-m}{y-x}f(x) + \frac{m-x}{y-x}f(y),$

 $\frac{f(m) - f(x)}{m - x} \le \frac{f(y) - f(m)}{m - y}.$ Thus we get an analogous property to $f''(x) \geq 0$ (if f were twice-differentiable), that there is $a \in \mathbb{R}$ such that $f(x) \geq f(m) + a(x - m)$. Thus $f(X) \geq f(\mathbb{E}[X]) + a(X - \mathbb{E}[X])$, so

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,

 $x\mathbb{P}(X \ge x) \le \mathbb{E}[Y\chi_{X \ge x}],$

and for x > 0,

In \mathcal{L}^2 , we are able to introduce some geometry to the space, defining an inner product $\langle X,Y\rangle:=\mathbb{E}[XY]$. This allows us to consider \mathcal{L}^2 as a Hilbert space (modulus random variables equal to 0 almost everywhere). Consequently we get all of the results that we normally have for Hilbert spaces within \mathcal{L}^2 . These spaces are well-studied in functional analysis, although more typically work with the Lebesgue measure rather than a probability measure.

Martingales

We often consider notions of random walks in probability theory. The most interesting of these are random walks where the movement at each step has expectation 0, and this is a concept generalised by the notion of martingales. In this course we cover only the discrete case of martingales, although there is a rich theory concerning continuous martingales.

Definition 29 (Martingales) Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space. An inte-

grable, (\mathcal{F}_n) -adapted stochastic process (X_n) is called

i. a martingale if for $n \geq 0$, $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$ a.s.; ii. a submartingale if for $n \geq 0$, $\mathbb{E}|X_{n+1}|\mathcal{F}_n| \geq X_n$ a.s.; and iii. a supermartingale if for $n \geq 0$, $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$ a.s..

satisfying $\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = 0$ as 'martingale difference sequences'.

We can note immediately that the difference between a super- and submartingale is almost purely aesthetic – negate one and you get the other. Additionally, we can see that a sequence of random variables is a martingale iff it is both a super- and submartingale. Consequently, statements about submartingales are preferable for their generality to statements about

 $\mathbb{E}\left[X_n \,|\, \mathcal{F}_m
ight] \geq X_{\min(m,n)},$ and the analogous relation follows for martingales and supermartingales. What this says is that a martingale is a sequence of random variables for which we expect no change on average. Meanwhile, a submartingale is a sequence which we expect to increase, and a supermartingale one we expect to decrease. To reflect this, we often refer to sequences (Y_n)

If a submartingale is adapted to a smaller filtration, then it is also a submartingale with respect to that filtration. This is because the only information a submartingale is using from the filtration is that in the natural filtration, and consequently the extra information is unnecessary. Note that this doesn't go the other direction though – it's straightforward to construct a large filtration (e.g. the constant filtration (\mathcal{F})) such that (X_n) is not a

submartingale, despite potentially being a martingale in a smaller filtration. A very general example of a martingale can be given just as a sum of independent random variables (Y_n) each with mean 0, with respect to the natural filtration. Another cute example

of a martingale involves taking an integrable random variable X and an arbitrary filtration,

then defining $X_n := \mathbb{E}[X | \mathcal{F}_n]$. **Lemma 31** Let (X_n) be a martingale with respect to (\mathcal{F}_n) , and $f: \mathbb{R} \to \mathbb{R}$ convex. Then

provided $(f(X_n))$ is a sequence of integrable random variables, then it is a submartingale. See this as an application of Jensen's inequality, and note that this has some very widereaching applications.

Definition 30 (Predictable process) A sequence (V_n) of random variables is predictable with respect to (\mathcal{F}_n) if $\sigma(V_n) \subseteq \mathcal{F}_{n-1}$ (V_n is \mathcal{F}_{n-1} -measurable) for $n \geq 1$.

Theorem 32 For $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ a filtered probability space, (Y_n) a martingale, (V_n) a

 $X_n = \sum V_k(Y_k - Y_{k-1}),$

bility space, (X_n) an integrable adapted process.

and in particular, $\mathbb{E}[X_{\rho}] = \mathbb{E}[X_{\tau}] = \mathbb{E}[X_{0}]$.

long as the time is bounded).

if each X_n is integrable then (X_n) is a martingale with respect to (\mathcal{F}_n) .

predictable process, both with respect to (\mathcal{F}_n) , then defining for $n \geq 0$

Note that the sequence (X_n) is a martingale transform, sometimes denoted $((V \circ Y)_n)$. The definition of a submartingale prompts some questions about how exactly the sequence varies from being a martingale, and how exactly we might correct it back to a martingale.

Theorem 33 (Doob's Decomposition theorem) Let $(\Omega, \mathcal{F}, (\mathcal{F})_n, \mathbb{P})$ be a filtered proba-

i. (X_n) has a Doob decomposition $X_n = X_0 + M_n + A_n$ where (M_n) is a martingale, (A_n) is predictable with respect to (\mathbb{F}_n) , and $M_0 = A_0 = 0$. ii. For any (M_n) a martingale, (\widetilde{A}_n) predictable such that $X_n = X_0 + \widetilde{M}_n + \widetilde{A}_n$, $M_n = \widetilde{M}_n$ and $A_n = A_n$ for all $n \ge 0$ almost surely. $iii. (X_n)$ is a sub(super)martingale iff (A_n) is non-decreasing (non-increasing) almost

An important consequence of this is known as the angle bracket process. For (M_n) a martingale of random variables in \mathcal{L}^2 , we have that (M_n^2) is a submartingale. We write the Doob decomposition as

 $M_n^2 = M_0^2 + N_n + \langle M \rangle_n$

where (N_n) is a martingale, $(\langle M \rangle_n)$ an increasing predictable process. Using predictability, we see that $\langle M \rangle_{n+1} - \langle M \rangle_n = \mathbb{E} \left[M_{n+1}^2 - M_n^2 \,|\, \mathcal{F}_n \right] = \mathbb{E} \left[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n \right].$ This is the conditional variance of $(M_{n+1} - M_n)$.

a finite stopping time. Then $(X_{\min(n,\tau)})$ is a martingale with respect to (\mathcal{F}_n) and hence This can be proven by noting that $(\chi_{n<\tau})$ is predictable, and thus we can write $(X_{\min(n,\tau)})$ in

terms of both this and (X_n) . **Theorem 35 (Doob's Optional Sampling theorem)** Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space, (X_n) a martingale, and τ , ρ two finite stopping times, $\tau \leq \rho$. Then $\mathbb{E}[X_{\rho} | \mathcal{F}_{ au}] \stackrel{\mathrm{a.s.}}{=} X_{ au},$

Similarly, if (X_n) is a submartingale, then $\mathbb{E}[X_\rho | \mathcal{F}_\tau] \geq X_\tau$ a.s..

We can prove this in two steps. Firstly, prove the case for ρ constant, and then consider the martingale $(X_{\min(n,\rho)} - X_{\min(n,\tau)})$ stopped at τ . This theorem is quite an important one, as we essentially just expand the definition of a martingale (and, via decomposition, super- and submartingales). The intuition should be that if (X_n) is a process stopping at ρ , but we only observe it up to τ , the properties of the

martingale should allow us to calculate the expected value of X_{ρ} as X_{τ} . Importantly, we also

generalise that a martingale's expectation is constant, even if stopping at a random time (as

We can also get a generalisation beyond just bounded stopping times: Corollary 3 (Variants of Doob's optional stopping theorem) Let (X_n) be a martingale, τ an a.s. finite stopping time. If either $\{X_n: n \geq 0\}$ is UI, or $\mathbb{E}|\tau| < \infty$ and the sequence $(\mathbb{E}[|X_{n+1}-X_n||\mathcal{F}_n])$ is bounded, then

 $\mathbb{E}\left[X_{\tau}\chi_{\tau<\infty}\right] = \mathbb{E}\left[X_{0}\right].$

 $\lambda \mathbb{P}\left(\max_{k \leq n} X_k \geq \lambda\right) \leq \mathbb{E}\left[X_n \chi_{\{\max_{k \leq n} X_k \geq \lambda\}}\right] \leq \mathbb{E}\left[|X_n|\right].$

We would like to keep characterising martingales further – in particular wishing to bound **Theorem 36 (Doob's maximal inequality)** Let (X_n) be a submartingale. Then, for $\lambda > 1$ $Y_n^{\lambda} := (X_n - \lambda) \chi_{\{\max_{k \le n} X_k \ge \lambda\}}$ is a submartingale, and for $n \geq 1$,

Essentially, we expect (X_n) to grow, but only so quickly. To see this, we construct Y_n^{λ} in terms of $(X_n - X_\tau)\chi_{\tau > n}$ and $(X_\tau - \lambda)\chi_{\tau > n}$, where τ is the hitting time of $[\lambda, \infty)$. We can then take expectations and use usual properties.

A useful corollary is that for (X_n) a martingale in \mathcal{L}^p , by applying the above to $(|X_n|^p)$, $\mathbb{P}\left(\max_{k \le n} |X_k| \ge \lambda\right) \le \frac{\mathbb{E}\left[|X_n|^p\right]}{\lambda^p}.$ **Theorem 37 (Doob's** L^p inequality) Let (X_n) be a non-negative submartingale in \mathcal{L}^p

 $\mathbb{E}[X_n^p] \le \mathbb{E}\left[\max_{k \le n} X_k^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_n^p].$

This holds by recalling that for U, V positive random variables, if $u\mathbb{P}(U \geq u) \leq \mathbb{E}[V\chi_{U>u}]$,

 $\lambda \mathbb{P}\left(\bigcup_{k=0}^{\infty}\{|X_k| \geq \lambda\}\right) \leq \mathbb{E}[X_0] + 2\mathbb{E}[X_n^-].$

then with p > 1, $\mathbb{E}[U^p] \leq (p/(p-1))^p \mathbb{E}[V^p]$. As a weaker result, we can also obtain for a supermartingale that

for $p \geq 1$. Then $\max X_k \in \mathcal{L}^p$ and

Martingale Convergence

define the integer sequences (ρ_n) , (τ_n) such that $\rho_n = \inf\{m \ge \tau_{n-1} : x_m \le a\}$ $\tau_n = \inf\{m \ge \rho_n : x_m \ge b\}.$

We can then define the sequence $(U_n([a,b],(x_n)))$ as the number of upcrossings of [a,b]by time n, $U([a,b],(x_n))$ the total number of upcrossings, by $U_n([a, b], (x_n)) = \sup\{k \ge 0 : \tau_k \le n\}$ $U([a, b], (x_n)) = \sup\{k \ge 0 : \tau_k < \infty\}.$

Definition 31 (Upcrossings) With a real sequence (x_n) , $a,b \in \mathbb{R}$ with a < b, we can

Lemma 38 (Doob's upcrossing lemma) Let (X_n) be a supermartingale, $a, b \in \mathbb{R}$ with a < b. For $n \ge 0$,

 $\mathbb{E}\left[U_n([a,b],(X_n))\right] \le \frac{\mathbb{E}\left[(X_n-a)^-\right]}{b-a}.$ To prove this, it's firstly useful to check that each ρ_n , τ_n as defined above are stopping times.

We then define the predictable quantity

by time n if one employs a strategy of buying at each ρ_k , and selling at each τ_k . We can then take expectations of this quantity in order to derive the lemma. Assuming one follows this hypothetical strategy, one can consider $(X_n - a)^-$ as the loss on

the initial investment. Abnormally for this course, we now derive a lemma from real analysis:

 $a, b \in \mathbb{Q}$ with a < b. To make this easier when dealing with extended reals, one should work with the lim sup and

Theorem 40 (Doob's forward convergence theorem) Let (X_n) be a super- or submartingale. If (X_n) is bounded in \mathcal{L}^1 ($\mathbb{E}[|X_n|]$ is bounded), then (X_n) converges a.s. to

Additionally, we can extend the result to \mathcal{L}^p using Hölder's inequality. This is one of the most critical theorems in relation to martingales, which in many ways justifies the attention paid to them. A martingale with bounded expectation will converge

Definition 32 (Galton-Watson branching process) Let $(X_{n,r})$ be an array of i.i.d. random variables. (Z_n) is a Galton-Watson branching process if $Z_0 = 1$, and for $n \ge 1$,

If each $X_{n,r}$ is integrable, and $\mathbb{E}[X_{0,0}] = \mu \neq 0$, we can write $M_n = Z_n/\mu^n$ to get a martingale (M_n) with respect to $\mathcal{F}_n = \sigma(\{X_{k,r} : r \geq 1, k \leq n\}), (Z_n)$ a sub-, super-, or normal martingale depending on whether μ is greater than, less than, or equal to 1. By the forward-convergence theorem, we see that (Z_n) can either converge or blow up.

i. (X_n) is uniformly integrable. ii. There is some \mathcal{F}_{∞} -measurable random variable X such that $X_n \to X$ almost surely and in \mathcal{L}^1 . iii. There is an integrable \mathcal{F}_{∞} -measurable random variable X such that $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$

Further, if $X \in \mathcal{L}^p$ for $p \geq 1$, then the convergence $X_n \to X$ also holds in \mathcal{L}^p .

Note that these conditions require slightly more than the forward convergence theorem,

convergence theorem to prove (part of) this statement however, in conjunction with Vitali's convergence theorem. This theorem is essentially an extension of the forward convergence theorem to convergence in \mathcal{L}^p (as well as a.s. convergence to an element of \mathcal{L}^p , which is already given).

We find that the optional sampling theorem, as well as the maximal and \mathcal{L}^p inequalities all have stronger variants for UI martingales:

Backwards Martingales **Definition 33 (Backwards martingales)** Let (\mathcal{F}_{-n}) with $\mathcal{F}_{-(n+1)} \subseteq \mathcal{F}_{-n} \subseteq \mathcal{F}$ for $n \geq 0$. A backwards martingale with respect to (\mathcal{F}_{-n}) is a sequence (X_{-n}) of integrable random variables such that X_{-n} is \mathcal{F}_{-n} -measurable, and

Theorem 43 Let (X_{-n}) be a backwards martingale with respect to (\mathcal{F}_{-n}) . Then X_{-n} converges a.s. and in \mathcal{L}^1 as $n \to \infty$ to $X_{-\infty} = \mathbb{E}[X_0 | \mathcal{F}_{-\infty}]$.

For exchangeable random variables X_1, \ldots, X_n , we can define for $\varphi \in S_n$ a probability measure \mathbb{P}' such that for $A \in \mathcal{F}$, $\mathbb{P}_{\varphi}(A) = \mathbb{P}(\boldsymbol{X}^{-1}(\varphi \boldsymbol{X}(A))).$

 $k \in \{1, \ldots, n\}$ by

is Doob's backward martingale. To show that this is a backwards martingale (ending at n, rather than 0) on $\mathcal{F}_k =$

 $Z_k = \frac{S_{n+1-k}}{I}$

 $\mathcal{A} = \left\{ \bigcap S_r^{-1}(E_r) : E_r \in \mathcal{B}(\mathbb{R}) \right\}$ We have that for $A = \bigcap_{r=n+1-k}^n S_r^{-1}(E_r) \in \mathcal{A}, i, j \leq n+1-k$, $\mathbb{E}[X_i \chi_A] = \mathbb{E} \left| X_i \prod_{r=n+1-k}^n \chi_{E_r}(S_r) \right|$

 $= \mathbb{E}_{(i\,j)} \left| X_i \prod_{r=n+1-k}^n \chi_{E_r} \left(\sum_{s=1}^r X_{s(i\,j)} \right) \right|$

Consequently, we then see the following:

 $=\mathbb{E}\big[X_1\,|\,\mathcal{F}_kig]$

which is 1 if at time n, (X_n) has begun but not finished an upcrossing, 0 otherwise. We can then define $(V \circ X)$ to represent, understanding (X_n) as the price of a stock, the profit made

the final buy. The upcrossing lemma indicates that in order to maintain the supermartingale property, for every upcrossing, we must spend a significant amount of time losing money on

Lemma 39 A real sequence (x_n) converges in $[-\infty, \infty]$ iff $U([a, b], (x_n)) < \infty$ for all

with probability 1. A specific case utilising this (often used in birth-death process models) is

following are equivalent:

Theorem 42 Let (X_n) be a UI martingale. Then for any stopping times $\tau \leq \rho$

a.s. for $n \geq 1$.

 $\lambda \mathbb{P} \left[\bigcup \{ |X_n| \ge \lambda \} \right] \le \mathbb{E} \left[|X| \chi_{\bigcup_{n \ge 1} \{ |X_n| \ge \lambda \}} \right].$ Finally, if $X \in \mathcal{L}^p$ for some p > 1, then with 1/p + 1/q = 1,

 $\mathbb{E}[X_{-n+1} | \mathcal{F}_{-n}] = X_{-n}.$ Since X_0 is integrable, we see that (X_{-n}) is automatically UI. We then need to adapt

Using this, we can prove the strong law of large numbers. Theorem 44 (Kolmogorov's strong law of large numbers) Let (X_n) be a sequence of

This theorem rests on defining S_n/n as a backwards martingale with respect to the backwards filtration $\sigma(S_{-n}, S_{-n+1}, \dots) = \sigma(S_{-n}, X_{-n+1}, \dots)$, and using this to encode the independence of S_n/n on the permutation of $\{X_1,\ldots,X_n\}$. We formalise this more

(with apologies for overloaded notation).

Definition 35 (Doob's backward martingale) With (X_1, \ldots, X_n) the results of n successive samples without replacement from a set $A \subseteq \mathbb{R}$ with $|A| \geq n$, (Z_k) defined for

 $\sigma(S_{n+1-k},\ldots,S_n)$ for $k\in\{1,\ldots,n\}$ requires some work. We need to demonstrate that for $i, j \leq n+1-k$, $\mathbb{E}[X_i | \mathcal{F}_k] = \mathbb{E}[X_j | \mathcal{F}_k]$. To do this, we just need to show the relation

backwards martingale), we can capture the exchangeability property to define a martingale.

lim inf to prove this.

We can see this from applying the upcrossing lemma in conjunction with our previous lemma.

Theorem 41 Let (X_n) be a martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$. The

because boundedness in \mathcal{L}^1 doesn't require that the X_n themselves have a uniform bound (only that their absolute expectation does). In the other direction, we can use the forward

Theorem 34 Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space, (X_n) a martingale, and τ $\mathbb{E}ig[X_
ho\,|\,\mathcal{F}_ auig]\stackrel{ ext{a.s.}}{=} X_ au$ and in particular $\mathbb{E}[X_{\tau}] = \mathbb{E}[X_0]$.

as $n \to \infty$ almost surely and in \mathcal{L}^1 .

 $||X||_p \le ||\max_{1} |X_n||_p \le q||X||_p.$

certain definitions such as with respect to upcrossings, but restore all of our normal results as expected, and in particular receive the following:

i.i.d. random variables, each integrable with mean μ .

Definition 34 (Exchangeability) The random variables X_1, \ldots, X_n are exchangeable if $(X_{\varphi(1)},\ldots,X_{\varphi(n)})$ has the same probability distribution for each permutation $\varphi\in S_n$

We see that this is a measure, and notably that by exchangeability $\mathbb{P}_{\varphi} \geq \mathbb{P}$ generally, and

on the π -system:

 $= \mathbb{E} \left[X_i \prod_{r=n+1-k}^n \chi_{E_r} \left(\sum_{s=1}^r X_s \right) \right]$

 $\mathbb{E}ig[Z_{k+1}\,|\,\mathcal{F}_kig] = rac{1}{n-k}\sum_{k=1}^{n-k}\mathbb{E}ig[X_i\,|\,\mathcal{F}_kig]$

The key aspect of this is that by defining the filtration backwards (and thus getting a