MATH 587 - ADVANCED PROBABILITY THEORY 1

GREG TAM

Contents

1. September 6th, 2012

Course Webpage: http://www.math.mcgill.ca/louigi/ and click on "Teaching".

1.1. Random Walk (Symmetric, Simple).

Let $X_n, n \geq 1$ be independent random variables,

$$X_i = \begin{cases} +1 & \text{with prob } 1/2 \\ -1 & \text{with prob } 1/2 \end{cases} \qquad \text{for all } i$$

Set $S_0 = 0$, $S_i = X_1 + \ldots + X_i$ for $i \geq 1$.
Question: Is there a $n > 0$ s.t. $S_n = 0$? What is $\mathbb{P}(\exists n > 0 : S_n = 0)$?

$$\mathbb{P}(\exists n > 0 : S_n = 0) \begin{cases} = 1 & \text{means random walk is recurrent} \\ < 1 & \text{means random walk is transient} \end{cases}$$

We may also consider

$$X_i = \begin{cases} +1 & \text{with prob } p \\ -1 & \text{with prob } 1-p \end{cases} \qquad \text{for all } i, 0 < p < 1$$

Theorem 1. *We have*

$$\begin{aligned} \mathbb{P}(\exists n > 0 \textit{ s.t. } S_n = 0) &= 1 - |2p-1| \\ &= \begin{cases} 1 & \textit{if } p = 1/2 \\ < 1 & \textit{if } p \neq 1/2 \end{cases} \end{aligned}$$

Proof. Summing up $z_n = \mathbb{P}(S_n = 0)$ doesn't work since they are not disjoint. Instead summing up $f_n = \mathbb{P}(S_1 \neq 0, \dots, S_{n-1} \neq 0, S_n = 0)$ yields $\mathbb{P}(\exists n > 0 \text{ s.t. } S_n = 0)$.

$$F(s) = \sum_{n \geq 1} f_n s^n \qquad P(s) = \sum_{n \geq 0} z_n s^n$$

Claim. $P(s) = 1 + P(s)F(s)$

Proof.

$$z_n = \sum_{k=1}^{n} f_k z_{n-k}$$

$$\begin{aligned} z_n &= \mathbb{P}(S_n = 0) \\ &= \mathbb{P}\left(\bigcup_{k=1}^{n} \{S_n = 0 \text{ and first return at time } k\} \right) \\ &= \sum_{k=1}^{n} P(S_n = 0, \text{ first return at time } k) \end{aligned}$$

Then

$$z_n s^n = \sum_{k=1}^{n} (f_k s^k)(z_{n-k} s^{n-k})$$

Sum over $n \geq 0$

$$\begin{aligned} P(S) = \sum_{n \geq 0} z_n s^n &= 1 + \sum_{n \geq 1} \sum_{k=1}^{n} (z_{n-k} s^{n-k})(f_k s^k) \\ &= 1 + F(s)P(s) \end{aligned}$$

Since the $n = 0$ term is 1. ☐

Corollary.

$$P(s) = \frac{1}{\sqrt{1 - 4p(1-p)s^2}} \qquad F(s) = 1 - \sqrt{1 - 4p(1-p)s^2}$$

Proof Sketch:

$$z_{2n} = \binom{2n}{n} p^n (1-p)^n$$

$$P(s) = \sum_{m \geq 0} \binom{2m}{m} p^m (1-p)^m s^{2m}$$

Then we look at $F(1)$

$$\begin{aligned} F(1) &= 1 - \sqrt{1 - 4p(1-p)} \\ &= 1 - \sqrt{(2p-1)^2} \end{aligned}$$

☐

Meta-Question: Is $\mathbb{P}(\exists n > 0 : S_n = 0)$ well-defined?
$\{-1, 1\}^{\mathbb{N}}$:=All sequences of $+1$'s and -1's
$(x_1, x_2, \dots)$ with $x_i \in \{-1, +1\}$
Then there is $A \subset \{-1, 1\}^{\mathbb{N}}$ s.t. $\exists n > 0 : S_n = 0 \Leftrightarrow (X_1, X_2, \dots) \subset A$

Theorem 2. *Assuming the axiom of choice, there is no way to assign probabilities to all subsets of $\{-1,+1\}^{\mathbb{N}}$ in such a way that the axioms of probability are satisfied.*

Supplemental reading: "A non-measurable set from coin flips"

1.2. **Branching Processes.** Start from a single individual (the root). The root has some random number B of children where B has some distribution μ. Independently, each child has a random number of children with distribution μ, and so on.

Let Z_n be the size of generation n. The family $(Z_n, n \geq 0)$ is some sequence of random variables.

$$\begin{aligned} E[Z_1] &= E[B] \\ E[Z_2 | Z_1 = 3] &= 3E[B] \\ E[Z_2|Z_1] &= Z_1 E[B] \leftarrow \text{ conditional expectation} \end{aligned}$$

$\mathbb{E}[Z_2|Z_1 = \pi]$ doesn't make much sense since $\mathbb{P}(Z_1 = \pi) = 0$.

$$\mathbb{E}[Z_2] = \sum_{i \geq 0} \underbrace{\mathbb{E}[Z_2|Z_1 = i]}_{i\mathbb{E}[B]} \mathbb{P}(B = i) = (\mathbb{E}[B])^2$$

and so we have

$$\mathbb{E}[Z_n] = (\mathbb{E}[B])^n$$

by induction. Let $W_n = \dfrac{Z_n}{(\mathbb{E}[B])^n}$ such that $\mathbb{E}[W_n] = 1\ \forall n$.

Definition 1 (Martingale Property)**.**

$$\mathbb{E}[W_n | W_0, \ldots, W_{n-1}] = W_{n-1}$$

Theorem 3 (Non-negative martingale convergence theorem)**.** *There is some random variable W such that $W_n \to W$ almost surely.*

Question: Is $\mathbb{E}[W] = 1$?

1.3. **Fatou's Lemma.** For any sequence $(X_n, n \geq 1)$ of random variables,

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$
$$\mathbb{E}[W] \leq \liminf_{n\to\infty} \mathbb{E}[W_n] = 1$$

Theorem 4 (Fundamental theorem of branching processes)**.**

$$\mathbb{P}(\exists n : Z_n = 0) \begin{cases} = 1 & \textit{if } \mathbb{E}[B] \leq 1 \Rightarrow \mathbb{P}(W = 0) = 1 \\ < 1 & \textit{otherwise} \end{cases}$$

Now suppose

$$B = \begin{cases} 2 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

S_n = number of unexplored individuals at step m

$$\mathbb{P}(\text{Extinction}) = \mathbb{P}(\text{Exploration stops}) = \mathbb{P}(\exists n > 0 : S_n = 0) = 1 - |1-2p|$$

2. September 18th, 2012

2.1. **What is a probability space? Last Class:** We can't assign probabilities to every possible event without contradiction.

Definition 2. *An algebra over a set S is a collection $\mathcal{F}$ of subsets of S such that*

1. $\emptyset \in \mathcal{F}$, $S \in \mathcal{F}$
2. *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$*
3. *If $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$*

Remark:

- $A \cap B = (A^c \cup B^c)^c$ so $\mathcal{F}$ is closed under intersection

- If $A_1, A_2, \dots, A_n \in \mathcal{F}$ then $\bigcap_{i=1}^{n} A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$
- This does not imply $A_i \in \mathcal{F}, i \geq 1 \not\Rightarrow \bigcup_{i=1}^{\infty} \in \mathcal{F}$ $(\star)$

Example 1 (Algebras over a set S)**.**

- $\{\emptyset, S\}$
- 2^S = *set of all subsets*
- $\mathcal{F} = \{A \subset S : |A| < \infty$ *or* $|A^c| < \infty\}$ *is an algebra but does not satisfy property* $\star$

Definition 3. *An algebra $\mathcal{F}$ over S is called a σ-algebra if $\star$ holds. A measurable space is a pair $(S, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra over S.*

Definition 4. *Given a set S and $\mathcal{F}_0 \subset 2^S$, then the σ-algebra generated by $\mathcal{F}_0$ is*

$$\sigma(\mathcal{F}_0) := \bigcap_{\substack{\mathcal{F} \supset \mathcal{F}_0 \\ \mathcal{F} \text{ a } \sigma\text{-algebra}}} \mathcal{F} \subseteq 2^S$$

Easy Fact: If $\mathcal{F} \supset \mathcal{F}_0$ is a σ-algebra, then $\sigma(\mathcal{F}_0) \subset \mathcal{F}$

Prop. *$\sigma(\mathcal{F}_0)$ is a σ-algebra*

Proof. If $A_i \in \bigcap \mathcal{F}$ for each $i \geq 1$ then fix any $\mathcal{F}$ in the intersection. Then $A_i \in \mathcal{F}_0$, $i \geq 1$. So by $\star$, $\bigcup_{i \geq 1} A_i \in \mathcal{F}$. so

$$\bigcup_{i \geq 1} A_i \in \bigcap \mathcal{F} = \sigma(\mathcal{F}_0)$$

$\square$

Note: If $\mathcal{F}_0$ is a σ-algebra, then $\sigma(\mathcal{F}_0) = \mathcal{F}_0$

Example 2. $\mathcal{F}_0 = \{$*all sets of size 1 in* $S\}$
$\sigma(\mathcal{F}_0) = 2^S$ *if S is countable*
$\sigma(\mathcal{F}_0) = \{A \subset S : A$ *is countable or* A^c *is countable*$\}$
S is a metric space.
$\mathcal{F}_0 = \{$*open sets in* $S\}$
$\sigma(\mathcal{F}_0)$ is called the Borel sets of S and denoted $B(S)$.

Prop. $\mathfrak{B}(\mathbb{R}) = \sigma(\{(-\infty, x], x \in \mathbb{R}\})$

Proof. "$\supseteq$":

$$(-\infty, x] = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right) \in \mathfrak{B}(\mathbb{R})$$

So

$$\{(-\infty, x], x \in \mathbb{R}\} \subset \mathfrak{B}(\mathbb{R})$$

So

$$\sigma(\{(-\infty, x], x \in \mathbb{R}\}) \subset \mathfrak{B}(\mathbb{R})$$

"$\subseteq$": Fact: Any open set in $\mathbb{R}$ can be written as

$$\bigcup_{i \geq 1} (a_i, b_i)$$

where the intervals (a_i, b_i) are disjoint. So it suffices to show that any interval (a, b) is a member of $\sigma(\{(-\infty, x], x \in \mathbb{R}\})$. To see this, write

$$\begin{aligned}(a, b) &= \bigcup_{n \geq 1} \left(a, b - \frac{1}{n}\right] \\ &= \bigcup_{n \geq 1} \left(-\infty, b - \frac{1}{n}\right] \cap ((-\infty, a]^c)\end{aligned}$$

$\square$

2.2. **Measures:** A measure μ on a measurable space $(S, \mathcal{F})$ is a function

$$\mu : \mathcal{F} \to [0, \infty]$$

such that

- $\mu(\emptyset) = 0$
- If A_i, $i \geq 1$ are disjoint elements of $\mathcal{F}$, then

$$\mu\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mu(A_i)$$

(Less important) definitions:
Given an algebra $\mathcal{F}_0$ over S, a function $\mu : \mathcal{F}_0 \to [0, \infty]$ is additive if

- $\mu(\emptyset) = 0$
- If $A, B \in \mathcal{F}_0$, $A \cap B = \emptyset$ then $\mu(A \cup B) = \mu(A) + \mu(B)$

We say μ is a pre-measure if for any sequence A_i, $i \geq 1$ of disjoint elements of $\mathcal{F}_0$ if $\bigcup_{i\geq 1} A_i \in \mathcal{F}_0$, then

$$\mu\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mu(A_i)$$

Example 3.

(1) *If S is any set, $\mathcal{F} = 2^S$,*

$$\mu(A) = |A| = \begin{cases} i & \text{if } |A| = i < \infty \\ \infty & \text{otherwise} \end{cases}$$

(2) *$S = \mathbb{R}$, $\mathcal{F} = B(\mathbb{R})$, $\mu =$Lebesgue measure defined by*

$$\mu((a, b]) = b - a$$

Question: is this a measure?

We write $\lambda = \mu$ for Lebesgue measure. Take

$$\mathcal{F}_0 = \left\{\bigcup_{i=1}^{n} (a_i, b_i] : -\infty \leq a_1 \leq b_1 \leq a_2 \leq b_2 \ldots a_n \leq b_n \leq \infty\right\}$$

Need to check that this is an algebra first. We want to show: λ is a pre-measure on $\mathcal{F}_0$.

- $\lambda(\emptyset) = b - b = 0$
- $A, B \in \mathcal{F}_0$, $A \cap B = \emptyset$, then $\lambda(A \cup B) = \lambda(A) + \lambda(B)$

Check: λ is a pre-measure

2.3. **Carathéodory Extension Theorem.**

Theorem 5 (Carathéodory Extension Theorem)**.** *Given an algebra $\mathcal{F}_0$ over a set S and a pre-measure μ_0 on $(S, \mathcal{F}_0)$, then there exists a unique measure μ on $(S, \sigma(\mathcal{F}_0))$ such that $\mu(A) = \mu_0(A)$ for all $A \in \mathcal{F}_0$.*

Prop. *If A_i, $i \geq 1$ are disjoint elements of $\mathcal{F}_0$, $A_i \subset [0, 1]$ for all i and $\bigcup_{i\geq 1} A_i \in \mathcal{F}_0$ then*

$$\lambda\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \lambda(A_i)$$

Assuming the proposition, we now prove that λ is a pre-measure. Fix A_i, $i \geq 1$, disjoint elements of $\mathcal{F}_0$ such that $\bigcup_{i\geq 1} A_i \in \mathcal{F}_0$. If $\bigcup_{i\geq 1} A_i$ is unbounded, we need to show that $\sum_{i\geq 1} \lambda(A_i) = \infty$. If $\bigcup_{i\geq 1} A_i$ is bounded, then for $n \in \mathbb{Z}$, let $A_{i,n} = A_i \cap (n, n + 1] \in \mathcal{F}_0$, so

$$\bigcup_{i\geq 1} A_i = \bigcup_{\substack{i\geq 1 \\ n\in\mathbb{Z}}} A_{i,n}$$

Then

$$\lambda\left(\bigcup_{i\geq 1} A_i\right) = \lambda\left(\bigcup_{i\geq 1} \bigcup_{n\in\mathbb{Z}} A_{i,n}\right)$$

$$
\begin{aligned}
&= \lambda\left(\bigcup_{n\in\mathbb{Z}}\bigcup_{i\geq 1} A_{i,n}\right) \text{ finite disjoint union}\\
&= \sum_{n\in\mathbb{Z}} \lambda\left(\bigcup_{i\geq 1} A_{i,n}\right)\\
&= \sum_{n\in\mathbb{Z}}\sum_{i\geq 1} \lambda\left(A_{i,n}\right)\\
&= \sum_{i\geq 1}\sum_{n\in\mathbb{Z}} \lambda\left(A_{i,n}\right)\\
&= \sum_{i\geq 1} \lambda\left(\bigcup_{n\in\mathbb{Z}} A_{i,n}\right)\\
&= \sum_{i\geq 1} \lambda(A_i)
\end{aligned}
$$

3. September 20th, 2012

Note: Write $B_n = \bigcup_{i=1}^n A_i$. Then $B_n \subset B_{n+1} \subset \dots$ and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i = \lim_{n\to\infty} B_n = B_\infty$$

If

$$\lambda(B_\infty) = \lim_{n\to\infty} \lambda(B_n)$$

then

$$
\begin{aligned}
\lambda\left(\bigcup_{i=1}^{\infty} B_i\right) &= \lim_{n\to\infty} \lambda\left(\bigcup_{i=1}^{n} A_i\right)\\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda(A_i)\\
&= \sum_{i=1}^{\infty} \lambda(A_i)
\end{aligned}
$$

So to prove the proposition, it suffices to show:
If B_n, $n \geq 1$ are an increasing sequence of elements of $\mathcal{F}_0$ and $B_\infty = \lim_{n\to\infty} B_n \in \mathcal{F}_0$, then

$$\lambda(B_\infty) = \lim_{n\to\infty} B_n$$

Prop. *If $H_i \subset [0,1]$, $i \geq 1$ is a decreasing sequence of elements of $\mathcal{F}_0$, and $\lambda(H_n) \geq 2\varepsilon > 0$ for all n, then*

$$\bigcap_{n\geq 1} H_n \neq \emptyset$$

Assume this proposition holds. Then write $H_n = B_\infty \setminus B_n$. Then H_n is decreasing and decreases to $\emptyset$, so by the proposition, $\forall \varepsilon > 0$, $\exists n$ such that $\lambda(H_n) < \varepsilon$

$$
\begin{aligned}
\lambda(H_n) &= \lambda(B_\infty \setminus B_n)\\
&= \lambda(B_\infty) - \lambda(B_n)\\
&< \varepsilon
\end{aligned}
$$

So $\lambda(B_\infty) \leq \lim_{n\to\infty} \lambda(B_n)$. On the other hand, $\lambda(B_\infty) \geq \lambda(B_n)$ for all n since $B_n \subset B_\infty$, so $\lambda(B_\infty) = \lim_{n\to\infty} \lambda(B_n)$.

Proof. For $k \geq 1$, pick $J_k \subset H_k$ with $\overline{J_k} \subset H_k$, with

$$\lambda(H_k \setminus J_k) \leq \frac{\varepsilon}{2^k}$$

and so

$$\lambda\left(H_n \setminus \bigcap_{k\le n} J_k\right) \le \lambda\left\{H_n \cap \left(\bigcup_{k\le n} J_k^c\right)\right\} \le \sum_{k\le n} \lambda(H_k \setminus J_k) \le \sum_{k\le n} \frac{\varepsilon}{2^k} < \varepsilon$$

and since we have

$$2\varepsilon - \lambda\left(\bigcap_{k\le n} J_k\right) < \lambda(H_n) - \lambda\left(\bigcap_{k\le n} J_k\right) < \varepsilon$$

then

$$\lambda\left(\bigcap_{k\le n} J_k\right) > \varepsilon$$

Let $K_n = \bigcap_{k\le n} \overline{J_k} \neq \emptyset$, so we can choose $x_n \in K_n$. The sequence $\{x_n\}_{n\ge 1} \subset [0,1]$, so it has a convergent subsequence $\{x_{n_j}\}_{n\ge 1}$ where $x_{n_j} \to x$ as $j \to \infty$. For $n_j > n$, $x_{n_j} \to x$ inside K_n which is closed, so $x \in K_n$. But n was arbitrary so

$$x \in \bigcap_{n\ge 1} K_n \subset \bigcap_{n\ge 1} H_n$$

□

Definition 5. *Given an algebra $\mathcal{F}_0$ over S and $\mu : \mathcal{F}_0 \to [0,\infty]$, $\mu(\emptyset) = 0$, the λ-sets of $\mathcal{F}_0$ are those sets $L \in \mathcal{F}_0$, for which*

$$\mu(F) = \mu(L \cap F) + \mu(L^c \cap F) \text{ for all } F \in \mathcal{F}_0$$

Let

$$\mathcal{L} = \{\lambda\text{-sets in } \mathcal{F}_0\} \subset \mathcal{F}_0$$

Lemma. *$\mathcal{L}$ is an algebra and μ is additive on $\mathcal{L}$.*

Proof.

$$\begin{aligned}\mu(\emptyset) &= \mu(\emptyset \cap F) + \mu(\emptyset \cap F^c) \quad \forall\, F \in \mathcal{F}_0\\ &= \mu(\emptyset) + \mu(\emptyset)\\ &= 0\end{aligned}$$

If $A, B \subset \mathcal{L}$, we want that $A \cup B \in \mathcal{L}$. Let $C = A \cap B \subset \mathcal{F}$.

$$\begin{aligned}\mu(C^c \cap F) &= \mu(A \cap C^c \cap F) + \mu(A^c \cap C^c \cap F)\\ \mu(F) &= \mu(A \cap F) + \mu(A^c \cap F)\\ \mu(A \cap F) &= \mu(B \cap A \cap F) + \mu(B^c \cap A \cap F)\end{aligned}$$

The first two lines use that A is a λ-set and the third line uses that B is a λ-set.

$$\mu(F) = \mu((A \cup B) \cap F) + \mu(A^c \cap B^c \cap F)$$

If $L \in \mathcal{L}$,

$$\begin{aligned}\mu(F) &= \mu(L \cap F) + \mu(L^c \cap F)\\ &= \mu(L^c \cap F) + \mu((L^c)^c \cap F)\end{aligned}$$

□

Lemma. *For all $L_1, \ldots, L_n \in \mathcal{L}_0$ disjoint,*

$$\mu(L_1 \cap F) + \mu(L_2 \cap F) + \ldots + \mu(L_n \cap F) = \mu\left(\bigcup_{i=1}^{n} L_i \cap F\right)$$

for all $F \in \mathcal{F}_0$.

Definition 6 (λ-sets)**.** *S, $\mathcal{F} \subset 2^S$, $\mu : \mathcal{F} \to [0,\infty]$, $\mu(\emptyset) = 0$.*
$L \in \mathcal{F}$ is a λ-set if $\forall F \in \mathcal{F}$

$$\mu(F) = \mu(L \cap F) + \mu(L^c \cap F)$$

Idea for extension: (motivated by Riemann integral)

Definition 7. *For any $A \subset S$, let*

$$\mu(A) = \inf \sum_{B_1, B_2, \ldots} \mu_0(B_i)$$

where the B_i are a countable covering of A by elements of $\mathcal{F}_0$. (The collection $\{B_i\}_{i \geq 1}$ covers A if $A \subset \bigcup_{i \geq 1} B_i$)

The function μ is defined on 2^S.

Fact: If $T \subset U$, $T, U \subset S$, then

$$\mu(T) \leq \mu(U)$$

(μ is increasing)

Prop. *μ is countable sub-additive: $\forall\, T_i$, $i \geq 1$ subsets of S, then*

$$\mu\left(\bigcup_{i \geq 1} T_i\right) \leq \sum_{i \geq 1} \mu(T_i)$$

Proof. **Idea:** Cover T_i using sets from $\mathcal{F}_0$, to a tolerance of $\frac{\varepsilon}{2^i}$.
We know that

$$\mu(T_i) = \inf_{\text{coverings}} \sum_{B_{i,1}, B_{i,2}, \ldots} \mu(B_{i,j})$$

choose a particular $B_{i,1}, B_{i,2}, \ldots$ that cover T_i so that

$$\sum_{j \geq 1} \mu(B_{i,j}) \leq \mu(T_i) + \frac{\varepsilon}{2^i}$$

Then $\{B_{i,j}, i, j \geq 1\}$ is a countable cover of $\bigcup_{i \geq 1} T_i$ and

$$\sum_{i \geq 1} \underbrace{\sum_{j \geq 1} \mu(B_{i,j})}_{\leq \mu(T_i) + \frac{\varepsilon}{2^i}} \leq \underbrace{\sum_{i \geq 1} \left\{\mu(T_i) + \frac{\varepsilon}{2^i}\right\}}_{\sum_{i \geq 1} \mu(T_i) + \varepsilon}$$

So

$$\mu\left(\bigcup_{i \geq 1} T_i\right) = \inf_{\text{coverings}} \sum_{i \geq 1} \mu(C_i) \leq \left\{\sum_{i \geq 1} \mu(T_i)\right\} + \varepsilon$$

Since $\varepsilon > 0$ was arbitrary, this proves the proposition ∎

2^S is too big to prove more nice properties about μ, so we restrict our attention to $\mathcal{L} \subset 2^S$, $\mathcal{L} = \{\lambda\text{-sets in } 2^S\}$

Lemma. *$\forall F$, $\forall L_1, \ldots, L_n \in \mathcal{L}$ where the L's are disjoint, then*

$$\mu\left(F \cap \bigcup_{i=1}^n L_i\right) = \sum_{i=1}^n \mu(F \cap L_i)$$

Proof.

$$\begin{aligned}
&\mu((L_1 \cup L_2) \cap F) \\
&= \mu(L_1 \cap ((L_1 \cup L_2) \cap F)) \\
&+ \mu(L_1^c \cap ((L_1 \cup L_2) \cap F)) \text{ since } L_1 \text{ is a } \lambda\text{-set} \\
&= \mu(L_1 \cap F) + \mu(L_2 \cap F)
\end{aligned}$$

This is the case $n = 2$, then use induction on L_1 and $L_2' = \bigcup_{i=2}^n L_i$ ∎

We should really check that μ extends μ_0.

Claim: If $F \in \mathcal{F}_0$, then $\mu_0(F) = \mu(F)$.

Proof. F is itself a covering of F, so $\mu(F) \leq \mu_0(F)$. To prove $\mu(F) \geq \mu_0(F)$, fix any covering T_i, $i \geq 1$ of F with elements of $\mathcal{F}_0$. It suffices to show that

$$\sum_{i\geq 1} \mu_0(T_i) \geq \mu_0(F)$$

For $n \geq 1$, let $U_n = T_n \setminus \left(\bigcup_{i<n} T_i\right) \subset T_n \in \mathcal{F}_0$ and U_i, $i \geq 1$ covers F. We have

$$\begin{aligned}
\sum_{i\geq 1} \mu_0(T_i) &\geq \sum_{i\geq 1} \mu_0(U_i) \\
&\geq \sum_{i\geq 1} \mu_0(\underbrace{U_i \cap F}_{\in \mathcal{F}_0}) \\
&= \mu_0\left(\sum_{i\geq 1}(U_i \cap F))\right) \\
&= \mu_0(F)
\end{aligned}$$

since the $U_i \cap F$ partition F and μ_0 is a pre-measure and not just an additive function. □

Claim: $\mathcal{F}_0 \subset \mathcal{L}$

Proof. **Want:** $\forall L \subset \mathcal{F}_0,\ \forall\ A \subset S$,

$$\mu(A) = \mu(L \cap A) + \mu(L^c \cap A)$$

We have $\mu(A) \leq \mu(L \cap A) + \mu(L^c \cap A)$ by sub-additivity. For the other direction, fix $\varepsilon > 0$ and a covering T_i, $i \geq 1$ of A from $\mathcal{F}_0$, with

$$\sum_{i\geq 1} \mu_0(T_i) \leq \mu(A) + \varepsilon$$

The family $T_i \cap L, i \geq 1$ covers $A \cap L$, so

$$\mu(A \cap L) \leq \sum_{i\geq 1} \mu_0(T_i \cap L)$$

Likewise,

$$\mu(A \cap L^c) \leq \sum_{i\geq 1} \mu_0(T_i \cap L^c)$$

and so

$$\begin{aligned}
\mu(A \cap L) + \mu(A \cap L^c) &\leq \sum_{i\geq 1} \mu_0(T_i \cap L) + \mu_0(T_i \cap L^c) \\
&= \sum_{i\geq 1} \mu_0(T_i) \\
&\leq \mu(A) + \varepsilon
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this proves the claim. □

Claim: $\mathcal{L}$ is a σ-algebra

Proof. We just need to show that if L_i, $i \geq 1$ are elements of $\mathcal{L}$, then $L = \bigcup_{i\geq 1} L_i \in \mathcal{L}$. Assume the L_i are disjoint, by the same argument as before since we are in an algebra. Fix $T \subset S$: we need that $\mu(T) = \mu(T \cap L) + \mu(T \cap L^c)$. The left hand side is less than the larger by sub-additivity.

It remains to prove "$\geq$". For $n \in \mathbb{N}$, let

$$M_n = \bigcup_{i=1}^{n} L_i$$

$\mathcal{L}$ is an algebra, so $M_n \in \mathcal{L}$, so

$$\begin{aligned}
\mu(T) &= \mu(M_n \cap T) + \mu(M_n^c \cap T) \\
&\geq \mu(M_n \cap T) + \mu(L^c \cap T) \quad \text{since } \mu \text{ is increasing and } M_n^c \supset L^c
\end{aligned}$$

$$= \sum_{i=1}^{n} \mu(L_i \cap T) + \mu(L^c \cap T)$$

Let $n \to \infty$.

$$\begin{aligned}\mu(T) &\geq \sum_{i=1}^{\infty} \mu(L^c \cap T) + \mu(L_i^c \cap T) \\ &\geq \mu(L \cap T) + \mu(L^c \cap T) \quad \text{by countable subadditivity}\end{aligned}$$

□

5. September 27th, 2012

We prove: $\forall\, F \subset S$, $\forall$ disjoint elements $L_1, L_2, \ldots$ of $\mathcal{L}$

$$\mu(F \cap L) = \sum_{i \geq 1} \mu(F \cap L_i) \qquad L = \bigcup_{i \geq 1} L_i$$

Proof.

$$\mu(F \cap L) \leq \sum_{i \geq 1} \mu(F \cap L_i) \quad \text{is subadditivity}$$

For any $n \geq 1$,

$$\bigcup_{i=1}^{n} F \cap L_i \subset F \cap L$$

so

$$\begin{aligned}\mu(F \cap L) &\geq \mu\left(\bigcup_{i=1}^{n} F \cap L_i\right) \\ &= \sum_{i=1}^{n} \mu(F \cap L_i)\end{aligned}$$

Let $n \to \infty$, to obtain

$$\mu(F \cap L) \geq \sum_{i \geq 1} \mu(F \cap L_i)$$

□

Example 4. $\mu_0 \to \mu(F) = \inf_{A_i \text{ cover } F} \sum_{i \geq 1} \mu_0(A_i)$
We can call this function the Carathéodory extension of μ_0.
Fact: μ *is a measure on* $\sigma(\mathcal{F}_0)$.
Let $S = \mathbb{R}$, $\mathcal{F}_0 =$ *"Finite disjoint unions* $(a_i, b_i] \cup \ldots \cup (a_r, b_r]$
Let

$$\mu_0(\underbrace{F}_{\in \mathcal{F}_0}) = \begin{cases} 0 & \text{if } F = \emptyset \\ \infty & \text{if } F \neq \emptyset \end{cases}$$

So the Carathéodory extension is

$$\mu(F) = \begin{cases} 0 & \text{if } F = \emptyset \\ \infty & \text{if } F \neq \emptyset \end{cases}$$

Let $\widehat{\mu} : 2^S \to [0, \infty]$ *be* $\widehat{\mu}(F) = |F|$. *Then we have a contradiction since it is not unique!*

5.1. Dynkin's Uniqueness Lemma.

Theorem 6 (Dynkin's Uniqueness Lemma)**.** *The Carathéodory extension is the only extension, if*

$$\mu_0(S) < \infty$$

Definition 8. *A measure space* $(S, \mathcal{F}, \mu)$ *is* σ*-finite if you can write* $S = \bigcup_{i \geq 1} L_i$, $L_i \in \mathcal{F}$ *disjoint with*

$$\mu(L_i) < \infty \quad \forall\, i$$

If the Carathéodory extension μ is σ-finite, then it is the only σ-finite extension of μ_0.

Want to show: If μ_1 and μ_2 are two extensions of μ_0 to $\sigma(\mathcal{F}_0)$, then $\mu_1 \equiv \mu_2$.
Let $\mathcal{D} = \{F \in \sigma(\mathcal{F}_0) : \mu_1(F) = \mu_2(F)\}$. We know that $\mathcal{F}_0 \subset \mathcal{D}$.

5.2. **Difference Set.** If we have the following 2 claims, then we say $\mathcal{D}$ is a difference set **Claim 1:** If $F_n \in \mathcal{D}$, $n \geq 1$, $F_n \uparrow F \in \sigma(\mathcal{F}_0)$, then $F \in \mathcal{D}$. ($F \in \sigma(\mathcal{F}_0)$ is automatic.)

Proof.
$$\mu_1(F) \underbrace{=}_{\mu_1 \text{ measure}} \lim_{n\to\infty} \mu_1(F_n) \underbrace{=}_{F_n \in \mathcal{D}} \lim_{n\to\infty} \mu_2(F_n) \underbrace{=}_{\mu_2 \text{ measure}} \mu_2(F)$$
□

Claim 2: If $A, B \in \mathcal{D}$, $A \subset B$, then $B \setminus A \in \mathcal{D}$

Proof.
$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A)$$
□

A consequence of this is that this is true for complements. If we take $B = S$, then $B \setminus A = A^c$. What about $\cup$ or $\cap$? This turns out to be harder.
Key Idea: Build up from $\mathcal{F}_0$.

(1) Which sets are "closed under $\cap$ with $\mathcal{F}_0$"
$$\mathcal{D}_1 = \{A \in \mathcal{D} : \forall\, B \in \mathcal{F}_0,\ \mu_1(A \cap B) = \mu_2(A \cap B)\}$$
An easy consequence from this is that $\mathcal{F}_0 \subset \mathcal{D}$.

Claim: $\mathcal{D}_1$ is a difference set

Proof. If $A \in \mathcal{D}_1$, $B \in \mathcal{D}_1$, $A \subset B$, we need that $\forall F \in \mathcal{F}_0$,
$$\mu_1((B \setminus A) \cap F) = \mu_2((B \setminus A) \cap F)$$
We have

$$\begin{aligned}\mu_1((B \setminus A) \cap F) &= \mu_1(B \cap F) - \mu_1(A \cap F)\\ &= \mu_2(B \cap F) - \mu_2(A \cap F)\\ &= \mu_2((B \setminus A) \cap F)\end{aligned}$$
So $B \setminus A \in \mathcal{D}_1$. □

Similarly, $A_n \uparrow A$, $A_n \in \mathcal{D}_1$, then $\forall F \in \mathcal{F}_0$
$$\begin{aligned}\mu_1(A \cap F) &= \lim_{n\to\infty} \mu_1(A_n \cap F)\\ &= \lim_{n\to\infty} \mu_2(A_n \cap F)\\ &= \mu_2(A \cap F)\end{aligned}$$
Now play the same game starting from $\mathcal{D}_1$ instead of $\mathcal{F}_0$.
More precisely: $\mathcal{D}_2 = \{A \in \mathcal{D} : \forall\, B \in \mathcal{D}_1, \mu_1(A \cap B) = \mu_2(A \cap B)\}$
Exercise: Check $\mathcal{D}_2$ is a difference set.
So $\mathcal{D}_2$ is closed under intersections and a difference set, so it is a σ-field. We have that
$$\mathcal{D}_2 \subset \mathcal{D}_1 \subset \mathcal{D} \subset \sigma(\mathcal{F}_0)$$
But $\mathcal{D}_2 = \sigma(\mathcal{F}_0)$ since $\mathcal{D}_2$ is a σ-algebra, so the entire thing collapses on itself and
$$\mathcal{D}_2 = \mathcal{D}_1 = \mathcal{D} = \sigma(\mathcal{F}_0)$$
Finally, for any $A \in \sigma(\mathcal{F}_0) = \mathcal{D} = \mathcal{D}_1 = \mathcal{D}_2$,
$$\mu_1(A) = \mu_1(A \cap S) = \mu_2(A \cap S) = \mu_2(A)$$
so $\mu_1 \equiv \mu_2$ on $\sigma(\mathcal{F}_0)$.

Example 5.

- *One-dimensional Lebesgue measure*
- *Lebesgue measure on* $\mathbb{R}^d$

  Pre-measure: $\mu_0((a_1,b_1]\times\ldots\times(a_d,b_d]) = \prod_{i=1}^{d}(b_i - a_i)$

Let $F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-x^2/2}\,\mathrm{d}x$. Let $\mu_0((a_i,b_i]\cup\ldots\cup(a_r,b_r]) = \sum_{i=1}^{r}(F(b_i)-F(a_i))$
$\mu_0 \to \mu_1 \leftarrow$ Law of standard Normal random variable.
Exercise: Think about the measure corresponding to a non-decreasing function $F:\mathbb{R}\to[0,\infty)$ which may have discontinuities.
Hint: An increasing function has at most countably many discontinuities.

6. October 2nd, 2012

Example 6. $([0,1],\mathfrak{B}([0,1]),\lambda)$ *where* λ *is the Lebesgue measure.*
Fix $A \in B([0,1])$*,* $\lambda(A) = 0$*. Let* $B \subset A$*. What is the measure of* B*?*

6.1. Completion.

Given measure space $(S,\mathcal{F},\mu)$, if $B \subset A \in \mathcal{F}$ and $\mu(A) = 0$, then we should have $\mu(B) = 0$. Define a new measure μ^* on a new σ-algebra $\mathcal{F}^*$.

$$\mathcal{F}^* = \{F : \exists E, G \in \mathcal{F}, E \subset F \subset G, \mu(E) = \mu(G)\}$$

and as expected we should have that $\mu^*(F) = \mu(E) = \mu(G)$ for $F \in \mathcal{F}$.

6.2. Events and Probability Spaces.

A measure space $(S,\mathcal{F},\mu)$ is a probability space if $\mu(S) = 1$. Standard notation is to write $(\Omega,\mathcal{F},\mathbb{P})$

Example 7.

(1) *Die Roll:* $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = 2^{\Omega}$, $\mathbb{P}(A) = \frac{|A|}{6}$

(2) $\Omega = [0,1]$, $\mathcal{F} = \mathfrak{B}([0,1])$, $\mathbb{P} =$*Lebesgue measure*
We have a point $\omega \in [0,1]$,

$$\omega = 0.\omega_1\omega_2\omega_3\ldots$$
$$\omega^{(1)} = 0.\omega_1\omega_3\omega_5\ldots$$
$$\omega^{(2)} = 0.\omega_2\omega_4\omega_6\ldots$$

What is $\mathbb{P}(\{\omega : \omega_1 = 1\})$*. This is just equal to* $\mathbb{P}\left(\left[\frac{1}{2},1\right]\right)$*. Similarly,* $\mathbb{P}(\{\omega : \omega_3 = 1\}) = \frac{1}{2}$ *since we are looking at 4 subintervals of length* $\frac{1}{8}$*.*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

$$\mathbb{P}(\omega_w = 1|\omega_1 = 1) = \frac{1/4}{1/2} = \frac{1}{2}$$

(3) *Infinite sequence of coin tosses.*
$\Omega = \{H,T\}^{\mathbb{N}}$, $\omega \in \Omega$, $\omega = \omega_1\omega_2\omega_3\ldots$ *where each* $\omega_i \in \{H,T\}$

$$\mathcal{F} = \sigma(\{\underbrace{\{\omega : \omega_i = H\}}_{\text{event that ith coin is heads}},\ i \in \mathbb{N}\})$$

This is the σ*-algebra generated by the cylinder sets.*

$$\mathbb{P}_{1/2}\left(\{\omega_1 = H\}\cap\{\omega_5 = T\}\cap\{\omega_8 = H\}\right) = \frac{1}{8}$$
$$\mathbb{P}_p\left(\{\omega_1 = H\}\cap\{\omega_5 = T\}\cap\{\omega_8 = H\}\right) = p^2(1-p)$$
$$\mathbb{P}_p\left(\text{10 heads, 8 tails}\right) = p^{10}(1-p)^8$$

Events: $\{H,T\}^{\mathbb{N}}$, $\mathcal{F} = \sigma(\textit{cylinder sets})$, $\mathbb{P} = \mathbb{P}_{1/2}$.
$E = \{\textit{Only heads}\} = \{H,H,H,H,H,\ldots\}$ $\quad\mathbb{P}(E) = 0$
$E = \{\textit{Infinitely many heads}\}$
$E_i = \{\omega_i = H\}$*, and we can write it as* $E = \bigcap_{m\ge1}\bigcup_{n>m} E_n =: \limsup E_n$
Recall: x_n*,* $n \ge 1$ *sequence of numbers.*

$$\limsup_n x_n = \lim_{n\to\infty}\sup_{n\ge m} x_n$$

$$\limsup_n x_n \geq x \Leftrightarrow \text{ sequence exceeds } x \text{ infinitely many times}$$

We also write "E_n occurs infinitely often" or just "E_n i.o" instead of $\limsup E_n$.

Definition 9. *Given a sequence of events E_n,*

$$\liminf_{n\to\infty} E_n = \bigcup_{m\geq 1}\bigcap_{n\geq m} E_n$$

This is the same as saying "E_n occurs all but finitely many times."

Observation:

$$\left(\limsup_{n\to\infty} E_n\right)^c = \liminf_{n\to\infty}(E_n)^c$$

Now let

$$E = \left\{\omega : \lim_{n\to\infty} \underbrace{\frac{\#\{i \leq n : \omega_i = H\}}{n}}_{\frac{S_n}{n} \to \mathbb{E}[X_1]} = \frac{1}{2}\right\}$$

This is the event that the proportion of the number of tosses tends to a half. We should have that $\mathbb{P}(E) = 1$, but is $E \in \mathcal{F}$?

Definition 10. *We say an event $E \in \mathcal{F}$ occurs almost surely if $\mathbb{P}(E) = 1$.*

6.3. **Reverse Fatou's Lemma.** $\mathbb{P}(\limsup_{n\to\infty} E_n) \geq \limsup_{n\to\infty} \mathbb{P}(E_n)$
Fatou's Lemma: $\mathbb{P}(\liminf_{n\to\infty} E_n) \leq \liminf_{n\to\infty} \mathbb{P}(E_n)$

Proof of Reverse Fatou's Lemma.

$$\underbrace{\limsup_{n\to\infty} E_n}_{E} = \bigcap_{m\geq 1}\underbrace{\bigcup_{n\geq m} E_n}_{G_m}$$

We have $G_m \supset G_{m+1} \supset \ldots$ and

$$\lim_{m\to\infty} G_m = \bigcap_{m\geq 1} G_m = E$$

Thus

$$\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}\Big(\lim_{m\to\infty} G_m\Big) \\
&= \lim_{m\to\infty} \mathbb{P}(G_m) \quad \text{by monotonicity} \\
&= \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} E_n\right) \\
&\geq \lim_{m\to\infty} \sup_{n\geq m} \mathbb{P}(E_n) \\
&= \limsup_{n\to\infty} \mathbb{P}(E_n)
\end{aligned}$$

□

Note: $\mathbb{P}(E_n) \to 0$ does not mean that $\mathbb{P}(E_n \text{ i.o.}) = 0$
Basic Fact: If $\mathbb{P}(E_n) \to 0$ sufficiently quickly, then $\mathbb{P}(E_n \text{ i.o.}) = 0$.

Example 8. *Consider* $\mathbb{P}(E_n) = \frac{1}{3^n}$.

6.4. **(First) Borel-Cantelli Lemma.** If

$$\sum_{n\geq 1} \mathbb{P}(E_n) < \infty$$

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

Proof.
$$\sum_{n\geq 1}\mathbb{P}(E_n)<\infty \Leftrightarrow \forall\varepsilon\ \exists k:\sum_{n\geq k}\mathbb{P}(E_n)<\varepsilon$$
Then
$$\begin{aligned}\mathbb{P}(E_n \text{ i.o.}) &= \mathbb{P}\left(\bigcap_{n\geq 1}\bigcup_{n\geq m}E_n\right)\\ &\leq \min_{m\geq 1}\mathbb{P}\left(\bigcup_{n\geq m}E_n\right)\\ &<\varepsilon\end{aligned}$$
for any $\varepsilon>0$ □

Now consider
$$\begin{aligned}E_n &= \left\{\geq \frac{n}{2}\text{ heads in the first } n \text{ tosses}\right\}\\ &= \left\{S_n\geq \frac{n}{2}\right\}\\ &= \{S_n\geq \mathbb{E}[S_n]\}\end{aligned}$$
so $\liminf_{n\to\infty}E_n=\{S_n \text{ stays above its expected values “for all time”}\}$

6.5. **Random Variables.** Given measurable spaces $(\Omega,\mathcal{F})$ and (S,G), a measurable map from Ω to S is a function $f:\Omega\to S$ such that
$$\forall E\in G,\ f^{-1}(E)\in\mathcal{F}$$
If $(S,G)=(\mathbb{R},\mathfrak{B}(\mathbb{R}))$, then f is called a random variable. If $(S,G)=(\mathbb{R}^*,\mathfrak{B}(\mathbb{R}))$, then f is called an extended real random variable.
$$\mathbb{R}^*=\mathbb{R}\cup\{-\infty,\infty\}$$

7. October 4th, 2012

If $\sum_{n\geq 1}\mathbb{P}(E_n)<\infty$, then $\mathbb{P}(E_n \text{ i.o.})=0$. However, it is not necessarily true that if $\sum_{n\geq 1}\mathbb{P}(E_n)=\infty$, then $\mathbb{P}(E_n \text{ i.o.})>0$. To see this, let $E_n=[0,1/n]$.

E occurs almost surely if $\mathbb{P}(E)=1$. If E_n, $n\geq 1$ each occuring almost surely, then $\bigcap_{n\geq 1}E_n$ occurs almost surely. If $E_x=[0,1]\setminus\{x\}$, then $\mathbb{P}(E_x)=1$, but $\mathbb{P}\left(\bigcap_{x\in[0,1]}E_x\right)=0$.

7.1. **Measurable Maps, Random Variables.** $(\Omega,\mathcal{F})=([0,1),\mathfrak{B}([0,1)))$
We take $\omega\in\Omega$ and define $X(\omega)=\#$ of ones in the binary expansion of ω before the first zero.
$$X^{-1}(1)=\left[\frac{1}{2},\frac{3}{4}\right)\qquad X^{-1}(2)=\left[\frac{3}{4},\frac{7}{8}\right)$$
More generally, for $S\in\mathfrak{B}(\mathbb{R})$,
$$X^{-1}(S)=\bigcup_{n\in S\cap\mathbb{N}}\left[1-\frac{1}{2^n},1-\frac{1}{2^{n+1}}\right)\in\mathfrak{B}([0,1])=\mathcal{F}$$
So
$$\begin{aligned}\mathbb{P}(X\in S) &:= \mathbb{P}(\{\omega:X(\omega)\in S\})\\ &= \mathbb{P}\big(X^{-1}(S)\big)\\ &= \sum_{n\in S\cap\mathbb{N}}\frac{1}{2^{n+1}}\end{aligned}$$

Definition 11. *$X:\Omega\to T$ is $(\Omega,\mathcal{F})$-(T,τ) measurable if $\forall F\in\tau$, $X^{-1}(F)\in\mathcal{F}$*

We will typically use
$$X:\ \left.\begin{matrix}\Omega\to & \mathbb{R}\\ \mathcal{F} & \mathfrak{B}(\mathbb{R})\end{matrix}\right\}\text{Random variable}$$

Theorem 7. *Given $(\Omega,\mathcal{F})$ and $(S,\mathcal{S})$ measurable spaces and $X:\Omega\to S$, and $\mathcal{A}\subset\mathcal{S}$, if $X^{-1}(A)\in\mathcal{F}$ for all $A\in\mathcal{A}$ and $\sigma(\mathcal{A})=\mathcal{S}$ then X is measurable.*

Proof. We show that $G := \{E \subset S : X^{-1}(E) \in \mathcal{F}\}$ is a σ-algebra. Assuming this, then we know that since $\mathcal{A} \subset G$, so $\sigma(\mathcal{A}) \subset G$. However, $\sigma(\mathcal{A}) = \mathcal{S}$, so X is measurable.

- $S \in G$ since $X^{-1}(S) = \Omega \in \mathcal{F}$.
- $A \in G$, then $X^{-1}(A) \in \mathcal{F}$ and $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$.
- For countable unions, $X^{-1}\left(\bigcup_{n\geq 1} A_n\right) = \bigcup_{n\geq 1} X^{-1}(A_n)$ so if $A_n \in G\ \forall\ n \geq 1$, then the RHS $\in \mathcal{F}$

□

Example 9. $X : \Omega \to \mathbb{R}$
$\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ *generates* $\mathfrak{B}(\mathbb{R})$*. So to prove* X *is a random variable we just need to show that* $\forall x \in \mathbb{R}$*, "*$\{X \leq x\}$*"* $\in \mathcal{F}$*. Formally, we mean* $\{\omega : X(\omega) \leq x\} = X^{-1}((-\infty, x])$*.*

Similarly, $X : \Omega \to \mathbb{R}^d$*, to prove that* X *is a random vector, it suffices to show that "*$\{X \in [a_1, b_1] \times \ldots \times [a_d, b_d]\}$*"* $\in \mathcal{F}$ *for any* $a_1 < b_1, \ldots, a_d < b_d$*.*

Example 10.

$$E = \left\{\omega : \frac{\#\{k \leq n : \omega_k = 1\}}{n}\right\} \to \frac{1}{2}$$

$\Omega = \{0, 1\}^{\mathbb{N}}$*. Let* $X_n(\omega) = \sum_{i=1}^n \omega_i/n$*. Now we can rewrite* E *as*

$$E = \left\{\lim_{n\to\infty} X_n \textit{ exists and equals } \frac{1}{2}\right\}$$

Theorem 8 (Composition of measurable maps)**.** *Given three measurable spaces,* $(\Omega, \mathcal{F})$*,*$(S, \mathcal{S})$*, and* $(T, \mathcal{T})$*.*
If $f : \Omega \to S$ *is measurable and* $g : S \to T$ *is measurable, then*

$$g \circ f : \Omega \to T$$

is measurable.

Proof. If $E \in \mathcal{T}$ then $g^{-1}(E) \in \mathcal{S}$. So $f^{-1}(g^{-1}(E)) = (g \circ f)^{-1}(E) \in \mathcal{F}$. □

Corollary. *If* $f : \Omega \to \mathbb{R}^n$ *is a random vector and* $g : \mathbb{R}^n \to \mathbb{R}$ *is continuous, then* $X = g \circ f : \Omega \to \mathbb{R}$ *is a random variable*

Proof. We just need to check that g is measurable. We know that if

$$\mathcal{A} = \{\text{Open sets in } \mathbb{R}\}$$

then $\forall A \in \mathcal{A}$,

$$g^{-1}(A) \in \mathfrak{B}(\mathbb{R}^n)$$

But $\sigma(\mathcal{A}) = \mathfrak{B}(\mathbb{R})$ so we are done by an earlier theorem. □

Corollary. *If* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *are random variables, then*

$$\begin{aligned} \overrightarrow{X} : \quad \Omega \to & \quad \mathbb{R}^n \\ \omega \to & \quad (X_1(\omega), \ldots, X_n(\omega)) \end{aligned}$$

is a random vector

Proof. It suffices to consider rectangles $\prod_{i=1}^n [a_i, b_i] \subset \mathbb{R}^n$.

$$\begin{aligned} \{\overrightarrow{X} \in [a_1, b_1] \times \ldots \times [a_n, b_n]\} &= \bigcap_{i=1}^n \{X_i \in [a_i, b_i]\} \\ &= \bigcap_{i=1}^n X_i^{-1}([a_i, b_i]) \in \mathcal{F} \end{aligned}$$

□

Corollary. *If* $X_1, \ldots, X_n$ *are random variables, then the following are also random variables*

- $-X_1$
- $X_1 + X_2 + \ldots + X_n$
- $X_1 X_2 \ldots X_n$
- $\max_{i\in[1,n]} X_i$

Note: In all the corollaries, we could have used $\mathbb{R}^*$ and $(\mathbb{R}^*)^n$ instead. (Then we get extended real random variables and extended real random vectors.)

Theorem 9. *If X_n, $n \geq 1$ are extended real random variables, then $\inf_{n\geq 1} X_n$, $\sup_{n\geq 1} X_n$, $\limsup_{n\geq 1} X_n$, $\liminf_{n\geq 1} X_n$ are too. These are all functions from Ω to $\mathbb{R}^*$ defined pointwise.*

Proof. Consider $\inf_{n\geq 1} X_n(\omega)$. For $a \in \mathbb{R}^*$, this is at most a iff $\exists n \geq 1$ such that $X_n(\omega) \leq a$. So

$$\left\{\inf_{n\geq 1} X_n \leq a\right\} = \bigcup_{n\geq 1} \underbrace{\{X_n \leq a\}}_{X_n^{-1}((-\infty,a])} \in \mathcal{F}$$

Likewise, the same proof holds for $\sup_{n\geq 1} X_n = \inf_{n\geq 1}(-X_n)$ is a random variable.

$$\limsup_{n\to\infty} X_n = \lim_{m\to\infty} \sup_{n\geq m} X_n = \inf_{m\geq 1} \underbrace{\sup_{n\geq m} X_n}_{\text{R.V.}}$$

is also a random variable and so is $\liminf_{n\to\infty} X_n$. □

Back to Example: $I_n(\omega) = \omega_n$. $\Omega = \{0,1\}^{\mathbb{N}}$. $\mathcal{F} = \sigma(\{\omega_n = 1\}, n \geq 1)$
Exercise: $I_n : \Omega \to \mathbb{R}$ is a random variable.

$$X_n = \frac{1}{n}\sum_{k=1}^{n} I_k$$

Let $L^+ = \limsup_{n\to\infty} X_n$ and $L^- = \liminf_{n\to\infty} X_n$ and $L = L^+ - L^-$. Then

$$E = \{L = 0\} \cap \left\{L^+ \leq \frac{1}{2}\right\} \cap \left\{L^- \geq \frac{1}{2}\right\}$$

We believe that $\mathbb{P}(E) = 1$. (Strong law of large numbers)

Definition 12. Ω, $(S, \mathcal{S})$
Given a function $X : \Omega \to S$, let

$$\sigma(X) = \bigcap_{\substack{\mathcal{F},\ \sigma-algebra\ over\ \Omega \\ which\ make\ X\ measurable}} \mathcal{F}$$

What is clear is that $\forall E \in \mathcal{S}$, $X^{-1}(E) \in \sigma(X)$. Check that $\sigma(X)$ is a σ-algebra, called the σ-algebra generated by X.

More generally, $\sigma(X_i, i \in I)$ is the smallest σ-algebra which makes all the X_i measurable.

Example 11.
- $\Omega = \{0,1\}^{\mathbb{N}}$, *then* $\sigma(I_1) = \{\emptyset, \{\omega_1 = 1\}, \{\omega_1 = 0\}, \Omega\}$
- $\sigma(I_n, n \geq 1) = \mathcal{F}$ *whcih is the σ-algebra generated by cylinder sets.*

Exercise: If $X : (0,1) \to \mathbb{R}$ is increasing and continuous and $X(\omega) \to -\infty$ as $\omega \downarrow 0$, $X(\omega) \to \infty$ as $X(\omega) \uparrow 1$, then $\sigma(X) = \mathfrak{B}([0,1])$.

8. October 9th, 2012

8.1. **Generated σ-algebras.** Suppose we had a measurable map $f : \Omega \to S$ where $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ are measurable spaces, then

$$\sigma(f) = \bigcap_{\substack{\mathcal{G}\subset 2^{\Omega}, \mathcal{G} \text{ a } \sigma\text{-alg} \\ f \text{ is } (\Omega,\mathcal{G})\to(S,\mathcal{S}) \text{ measurable}}} \mathcal{G}$$

If $X : \Omega \to \mathbb{R}$, then

$$\sigma(X) = \{X^{-1}(B) : B \in \mathfrak{B}(\mathbb{R})\}$$

Clearly,

$$\{X^{-1}(B) : B \in \mathfrak{B}(\mathbb{R})\} \subset \sigma(X)$$

and the LHS is a σ-algebra. This is true more generally but not all the time.

Now let us imagine that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$.

Theorem 10. *The value $X(\omega)$ is equivalent to the collection of values*

$$\{I_F(\omega) : F \in \sigma(X)\}$$

in that either can be deduced from the other. Here we have

$$I_F(\omega) = \begin{cases} 1 & \textit{if } \omega \in F \\ 0 & \textit{otherwise} \end{cases}$$

Proof. Fix $x \in \mathbb{R}$. Then if $X(\omega) = x$, then $\omega \in X^{-1}(x)$. For any $F \in \sigma(X)$, we have $F = X^{-1}(B)$, for some $B \in \mathfrak{B}(\mathbb{R})$. If $x \in B$, then $X^{-1}(x) \in F$, so $I_F(\omega) = 1$. Conversely, if $x \notin B$, then $\omega \notin F = X^{-1}(B)$ so $I_F(\omega) = 0$.

Conversely, for a given $\{I_F(\omega), F \in \sigma(X)\}$. Then

$$X(\omega) = \inf\{x \in \mathbb{R} : I_{(-\infty,x]}(\omega) = 1\}$$

$\square$

We can think of $\{I_F(\omega) : F \in \sigma(X)\}$ as a function $f_\omega : \sigma(X) \to \{0, 1\}$. More generally, given a collection $\{X_i, i \in I\}$:

Theorem 11. *The collection $\{X_i(\omega) : i \in I\}$ can be deduced from*

$$\{I_F(\omega), F \in \sigma(X_i,\ i \in I)\}$$

and vice-versa. Let

$$\begin{aligned} A &= \{X_i(\omega) : i \in I\} \\ B &= \{I_F(\omega), F \in \sigma(X_i,\ i \in I)\} \end{aligned}$$

$B \Rightarrow A$ is easy. $I_{[X_i \le x]}$ is in the collection.
$A \Rightarrow B$. Idea: Let $E = \bigcap_{i \in I} X_i^{-1}(x_i)$. This does not work if I is uncountable.

Prop. *Given a random variable $X : \Omega \to \mathbb{R}$, a random variable $Y : \Omega \to \mathbb{R}$ is $\sigma(X)$-measurable iff $\exists$ measurable $f : \mathbb{R} \to \mathbb{R}$ such that $Y = f(X)$.*

$$\begin{tikzcd} \Omega \arrow[d, "X"'] \arrow[dr, "Y"] & \\ \mathbb{R} \arrow[r, "f"'] & \mathbb{R} \end{tikzcd}$$

Proof. ($\Leftarrow$) Easy: composition of measurable maps

$$(\Omega, \sigma(X)) \xrightarrow{X} (\mathbb{R}, \mathfrak{B}(\mathbb{R})) \xrightarrow{f} (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$$

($\Rightarrow$)

Step 1. If we can do it for bounded functions, we can do it for general functions since if we took $Z = \tan^{-1}(T)$, this is bounded on $[-\pi/2, \pi/2]$ and Y is our general function. Then we have Z is $\sigma(X)$-$\mathfrak{B}(\mathbb{R})$ measurable, so $\exists g : \mathbb{R} \to \mathbb{R}$ measurable such that $Z = g(X)$. Let $f = \tan(g) : \mathbb{R} \to \mathbb{R}$. Then

$$\begin{aligned} Y &= \tan Z \\ &= \tan(g(X)) \\ &= f(X) \end{aligned}$$

Step 2. Assume $|Y(\omega)| \le \pi/2$ for all ω. Fix $q \in \mathbb{Q} \cap [-\pi/2, \pi/2]$. Let $E_q = Y^{-1}((q, \infty])) \in \sigma(X)$. Then we know that $\exists B_q \in \mathfrak{B}(\mathbb{R})$ such that $E_q = X^{-1}(B_q)$. Now let

$$f_q(x) = \begin{cases} \frac{\pi}{2} & \text{if } x \in B_q \\ q & \text{otherwise} \end{cases}$$

Then $f_q(X) \ge Y$. Let $f = \inf_{q \in \mathbb{Q}} f_q$. Then $f : \mathbb{R} \to \mathbb{R}$ is measurable and $f(X) = Y$.

$\square$

8.2. **Distribution Functions $(\Omega, \mathcal{F}, \mathbb{P})$, X a random variable.** The distribution function of X is

$$\begin{aligned} F_X : &\mathbb{R} \to [0,1] \\ &x \to \mathbb{P}(X \leq x) \end{aligned}$$

Example 12. $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathfrak{B}([0,1]), \mathbb{P})$ *where $\mathbb{P}$ is the Lebesgue measure.*

- *X is uniform on $[0,1]$:*

$$F_X(x) = x$$

  But we could have $X(\omega) = \omega$ or $X(\omega) = 1 - \omega$.
- *X is normally distributed:*

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \,\mathrm{d}y$$

  Let $X(\omega) = F_X^{-1}(\omega)$.

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(\{\omega : X(\omega) \leq x\}) \\ &= \mathbb{P}\big(\{\omega : F^{-1}(\omega) \leq x\}\big) \\ &= F_X(x) \end{aligned}$$

8.3. **Properties of Distribution Functions.** F_x is the distribution function of some random variable X.

- F_X is increasing/non-decreasing
- $\lim_{y \to \infty} F_X(y) = 1$, $\lim_{y \to -\infty} F_X(y) = 0$
- F_X is right-continuous:

$$\mathbb{P}(X \leq y) = \lim_{z \downarrow y} \mathbb{P}(X \leq z)$$

  This follows from monotonicity.

8.4. **Skorokhod Representation of F.** This gives a random variable on $([0,1], \mathfrak{B}([0,1]), \mathbb{P})$ with distribution F.

DIAGRAMS

Formally,

$$\begin{aligned} X^+(\omega) &= \sup\{x : F(x) \leq \omega\} \\ X^-(\omega) &= \sup\{x : F(x) < \omega\} \end{aligned}$$

Note: $\mathbb{P}(X^- = X^+) = 1$, since $\{\omega : X^-(\omega) \neq X^+(\omega)\}$ is in bijection with the flats of F and the latter is countable.

9. October 11th, 2012

Definition 13 (Law of a random variable)**.** *If we have a space $(\Omega, \mathcal{F}, \mathbb{P})$, $X : \Omega \to \mathbb{R}$ is a random variable. The law of X is a measure μ_X on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$.*

$$\underbrace{\mu_X}_{\mathbb{P}(X)}(B) = \mathbb{P}(X \in B) = \mathbb{P}\big(X^{-1}(B)\big)$$

Remarks:

- We should check that μ_X is indeed a probability measure.
- $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mu_X)$ and we consider $f : \mathbb{R} \to \mathbb{R}$ which takes r to r (aka the identity map). f has the same distribution as X.

$$\mu_X(f \in B) = \mu_X(B) = \mathbb{P}(X \in B)$$

- μ_X is determined by F_X and vice-versa.
  $\Rightarrow$ is obvious.
  $\Leftarrow$ is because $\{(-\infty, x],\ x \in \mathbb{R}\}$ generates $\mathfrak{B}(\mathbb{R})$

9.1. **Independence.** $(\Omega, \mathcal{F}, \mathbb{P})$

- Events $E, F \in \mathcal{F}$ are independent if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\,\mathbb{P}(F)$$

- Random variables $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ are independent if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B) \quad \forall A, B \in \mathfrak{B}(\mathbb{R})$$

- Equivalently, X, Y are independent if $\forall E \in \sigma(X)$, $\forall F \in \sigma(Y)$,

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\,\mathbb{P}(F)$$

  More generally, sub-σ-algebras $\mathcal{G}, \mathcal{H}$ of $\mathcal{F}$ are independent if $\forall G \in \mathcal{G}, H \in \mathcal{H}$

$$\mathbb{P}(G \cap H) = \mathbb{P}(G)\,\mathbb{P}(H)$$

Note: If $F \in \mathcal{G} \cap \mathcal{H}$, then

$$\mathbb{P}(F) = \{\mathbb{P}(F)\}^2$$

so $\mathbb{P}(F) = 0$ or $\mathbb{P}(F) = 1$

Example 13. $([0,1), \mathfrak{B}([0,1)), \mathbb{P})$

- *Let X be the number of zeroes before the first 1 and Y be the number of ones in the first run of ones.*

$$\mathbb{P}(X = i) = \frac{1}{2^{i+1}} \qquad \mathbb{P}(Y = j) = \frac{1}{2^j}$$

$$\mathbb{P}(X = i, Y = j) = \frac{1}{2^{i+j+1}}$$

  we write $X \perp\!\!\!\perp Y$.
- *Now let X be Bernoulli with $p = 1/2$ and Y be Bernoulli with $p = 1/2$, where $X \perp\!\!\!\perp Y$. If*

$$Z = X + Y \mod 2$$

  then $X \perp\!\!\!\perp Z$, $Y \perp\!\!\!\perp Z$. But knowing two of these determines the third, so X, Y, Z are pairwise but not mutually independent.

Definition 14. $(\Omega, \mathcal{F}, \mathcal{P})$*, sub-$\sigma$-algebras $(\mathcal{F}, n \geq 1)$ of $\mathcal{F}$ are (mutually) independent if $\forall n \geq 1$, $\forall E_i \in \mathcal{F}_i$, $1 \leq i \leq n$,*

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

i.e. $E_1, \ldots, E_n$ are mutually independent.

Example 14. *If $E_1, \ldots, E_n$ are independent, then $E_1, \ldots, E_{n-1}, E_n^c$ are also independent.*

Exercise: Given $1 \leq n_1 < n_2 < \ldots$, $E_1 \in \mathcal{F}_{n_1}$, $E_2 \in \mathcal{F}_{n_2}, \ldots$

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i \geq 1} E_i\right) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=1}^n E_n\right) \\
&= \lim_{n\to\infty} \mathbb{P}(E_i) \\
&= \prod_{i=1}^n \mathbb{P}(E_i)
\end{aligned}$$

Definition 15. $(\Omega, \mathcal{F}, \mathbb{P})$*, $X_i : \Omega \to \mathbb{R}$, $i \geq 1$ random variables are independent if $(\sigma(X_i), i \geq 1)$ are independent.*

9.2. **π-System Lemma.** Given Ω, a π-system $\mathcal{P}$ over Ω is a subset of 2^Ω that is closed under finite intersections.

Uniqueness Lemma: Given $\mathcal{P}$ a π system over Ω, $\mathcal{F} = \sigma(\mathcal{P})$. If μ_1, μ_2 are measures on $(\Omega, \mathcal{F})$, $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, and $\mu_1|_\mathcal{P} \equiv \mu_2|_\mathcal{P}$ then $\mu_1 \equiv \mu_2$.

Recall: $\mathcal{D} = \{A \in \mathcal{F}, \mu_1(A) = \mu_2(A)\} \subset \mathcal{F}$. We showed eventually that $\mathcal{D} = \mathcal{F}$.

Lemma (π-system lemma)**.** *If $\mathcal{G}_n$, $n \geq 1$ are σ-algebras over Ω, $\mathcal{P}_n \subset \mathcal{G}_n$ is a π-system for each $n \geq 1$, with $\sigma(\mathcal{P}_n) = \mathcal{G}_n$, and $(\mathcal{P}_n, n \geq 1)$ are independent, then $(\mathcal{G}_n, n \geq 1)$ are independent.*

Example 15. *To show $(X_n, n \geq 1)$ are independent, it suffices to prove*

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i)$$

for all $x_1, \ldots, x_n \in \mathbb{R}$.

Proof. $(n = 2)$
Let $\mathcal{G}, \mathcal{H}$ be σ-algebras and $\mathcal{P} \subset \mathcal{G}, \mathcal{Q} \subset \mathcal{H}$ be π-systems.

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

Fix $E \in \mathcal{P}$ and define measures μ, ν on $\mathcal{H}$ by

$$\mu(F) = \mathbb{P}(E \cap F)$$

and

$$\nu(F) = \mathbb{P}(E)\,\mathbb{P}(F)$$

Note: $\mu|_\mathcal{Q} = \nu|_\mathcal{Q}$ and $\mu(\Omega) = \nu(\Omega)$ so $\mu \equiv \nu$. This means that $\forall E \in \mathcal{P}, F \in \mathcal{H}$

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\,\mathbb{P}(F) \quad (\star)$$

Once again, fix $F \in \mathcal{H}$. Let $\widehat{\mu}(E) = \mathbb{P}(E \cap F)$ for $E \in \mathcal{G}$ and let $\widehat{\nu}(E) = \mathbb{P}(E)\,\mathbb{P}(F)$ for $E \in \mathcal{G}$. By $\star$, $\widehat{\mu}|_\mathcal{P} = \widehat{\nu}|_\mathcal{P}$, so $\widehat{\mu} \equiv \widehat{\nu}$. So $\forall E \in \mathcal{G}, F \in \mathcal{H}$,

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\,\mathbb{P}(F)$$

□

Exercise: If $\mathcal{G}_i, i \geq 1$ are independent σ-algebras and $N_i, i \geq 1$ partitions of $\mathbb{N}$, then letting $\mathcal{H}_i = \sigma\left(\bigcup_{j \in N_i} \mathcal{G}_j\right)$, we have that $\mathcal{H}_i, i \geq 1)$ are independent. (Hint: Intersections of the form $\bigcap_{j \in N_i} E_j$, where $E_j \in \mathcal{G}_j$ for each j, is a π system that generates $\mathcal{H}_i$.)

Corollary. *$X_{i,j}$, $i, j \geq 1$ are independent and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \geq 1$ are random vectors, then by setting*

$$Y_i = f_i(X_{i,1}, \ldots, X_{i,n}) : \Omega \to \mathbb{R}$$

we have that $(Y_i, i \geq 1)$ are independent.

Now suppose I want a sequence of iid $\mathcal{N}(0,1)$ random variables.

(1) Consider $([0,1), \mathfrak{B}((0,1]), \mathcal{P})$,

$$U_1(\underbrace{\omega}_{0.\omega_1\omega_2\omega_3\omega_4\ldots}) = 0.\omega_2\omega_4\omega_8\omega_{16}\omega_{32}\ldots$$

(Exercise: $U_1(\omega)$ is uniform on $[0,1)$)

$$U_2(\omega) = 0.\omega_3\omega_9\omega_{27}\ldots$$

and in general

$$U_n(\omega) = 0.\omega_{p_n}\omega_{p_n^2}\omega_{p_n^3}\ldots$$

where p_n is the nth prime.
(Exercise: $(U_n, n \geq 1)$ are independent)

(2) Let $X_i = F^{-1}(U_i)$, where F is the cumulative distribution function of a $\mathcal{N}(0,1)$ random variable. By the corollary these $(X_i, i \geq 1)$ are independent. We saw last class that they are normally distributed.

9.3. **Borel-Cantelli 2.** If $E_i, i \geq 1$ are independent events and $\sum_{i\geq 1} \mathbb{P}(E_i) = \infty$, then $\mathbb{P}(E_i \text{ i.o.}) = 1$. (This is equal to $\mathbb{P}\Big(\bigcap_{m\geq 1} \bigcup_{n\geq m} E_n\Big)$)

Proof. We will show that the complement $\mathbb{P}\Big(\bigcup_{m\geq 1} \bigcap_{n\geq m} E_n^c\Big) = 0$. It suffices to show that $\forall m \geq 1$, $\mathbb{P}\Big(\bigcap_{n\geq m} E_n^c) = 0\Big)$. Well,

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{n\geq m} E_n^c\right) &= \prod_{n\geq m} \mathbb{P}(E_n^c) \\
&= \prod_{n\geq m} \{1 - \mathbb{P}(E_n)\} \\
&\leq \prod_{n\geq m} e^{-\mathbb{P}(E_n)} \qquad \text{since } 1 - x \leq e^{-x} \\
&= \exp\left\{-\sum_{n\geq m} \mathbb{P}(E_n)\right\} \\
&= 0 \qquad \text{since the sum is infinite}
\end{aligned}$$

□

10. October 16th, 2012

10.1. **Independence. Recall:** $\mathcal{F}, \mathcal{G}$, σ-algebras are independent if $\forall F \in \mathcal{F}, G \in \mathcal{G}$, we have $\mathbb{P}(F \cap G) = \mathbb{P}(F)\,\mathbb{P}(G)$.

To check $\mathcal{F} \perp\!\!\!\perp \mathcal{G}$ independent, it suffices to check that $P \perp\!\!\!\perp Q$ where P, Q are π-systems and

$$\sigma(P) \supset \mathcal{F} \qquad \sigma(Q) \supset \mathcal{F}$$

Borel-Cantelli II: If $(E_n, n \geq 1)$ are independent events, with

$$\sum_{n\geq 1} \mathbb{P}(E_n) = \infty$$

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

Example 16. *$(X_n, n \geq 1)$ iid* $\exp(1)$. *$\mathbb{P}(X_1 \geq t) = e^{-t}, t \geq 0$.*
How does $M_n = \max_{1\leq i\leq n} X_i$ grow?

$$\mathbb{P}(M_n > t) = \mathbb{P}\left(\bigcup_{i=1}^n \{X_i > t\}\right) \leq \sum_{i=1}^n \mathbb{P}(X_i > t) = ne^{-t}$$

If $t < \log n$, then this bound is useless. If $t \geq (1+\varepsilon)\log n$, we get $e^{-t} \leq \frac{1}{n^{1+\varepsilon}}$, so

$$\mathbb{P}(M_n > t) \leq \frac{1}{n^\varepsilon}$$

If $X_n \geq \log n$ infinitely often, then $M_n \geq \log n$ infinitely often.

$$\mathbb{P}(M_n \geq \log n \text{ i.o.}) \geq \mathbb{P}(X_n \geq \log n \text{ i.o.})$$

$$\sum_{n\geq 1} \mathbb{P}(E_n) = \sum_{n\geq 1} \frac{1}{n} = \infty$$

so by Borel Cantelli 2, $\mathbb{P}(X_n \geq \log n \text{ i.o.}) = 1$.

To prove a corresponding upper bound, note:
Given a non-decreasing function $f : \mathbb{N} \to \mathbb{R}$ with $f(n) \to \infty$ as $n \to \infty$. Then a sequence $(a_n, n \geq 1)$ has $a_n > f(n)$ infinitely often iff $m_n = \max_{1\leq i\leq n} > f(n)$ infinitely often.

Proof. $\Rightarrow$ is obvious.
$\Leftarrow$ happens if there exists a subsequence $n_1, n_2, \ldots \to \infty$ such that $m_{n_k} > f(n_k)$ for all k. Choose the sequence $n_1, n_2, \ldots$ such that for all k,

$$f(n_{k+1}) > m_{n_k}$$

This can be done since $f(n) \to \infty$. We have $m_{n_{k+1}} > f(n_{k+1}) > m_{n_k}$. So there is p_k, with $n_k < p_k \leq n_{k+1}$ such that

$$a_{p_{k+1}} = m_{n_{k+1}} > f(n_{k+1}) \geq f(p_{k+1})$$

Then $a_{p_i} \geq f(p_i)$ for all i. $\square$

$$M_n = \max(X_1, \dots X_n), \qquad X_i \sim \exp(1)$$

then

$$\limsup_{n\to\infty} \frac{M_n}{\log n} \geq 1$$

Now fix $\varepsilon > 0$. The note says that

$$M_n \geq (1+\varepsilon)\log n \text{ i.o.} \Leftrightarrow X_n \geq (1+\varepsilon)\log n \text{ i.o.}$$

so

$$\mathbb{P}(M_n \geq (1+\varepsilon)\log n \text{ i.o.}) = \mathbb{P}\left(\underbrace{X_n \geq (1+\varepsilon)\log n}_{E_n} \text{ i.o.}\right)$$

and we have

$$\mathbb{P}(E_n) = \frac{1}{n^{1+\varepsilon}}$$

so

$$\sum_{n\geq 1} \mathbb{P}(E_n) = \sum_{n\geq 1} \frac{1}{n^{1+\varepsilon}} < \infty$$

so $\mathbb{P}(M_n \geq (1+\varepsilon)\log n \text{ i.o.}) = 0$ by Borel-Cantelli.

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{M_n}{\log n} \geq 1\right) = 1$$
$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{M_n}{\log n} \leq 1+\varepsilon\right) = 1$$
$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{M_n}{\log n} < 1-\varepsilon\right) = 0$$

10.2. **St. Petersburg Paradox.**
$X = 2^{\#\text{heads before first tail}}$. We have $B_1, B_2, \dots \overset{iid}{\sim} \mathrm{Ber}(\frac{1}{2})$, so

$$X = 2^{\min\{i : B_i = 0\} - 1}$$

We have that

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k\geq 0} \frac{1}{2^{k+1}} \cdot \overbrace{2^k}^{\text{winnings if } B_{k+1} \text{ is the first } 0} \\ &= \sum_{k\geq 0} \frac{1}{2} \\ &= \infty\end{aligned}$$

Let X_i, $i \geq 1$ be iid as before.

$$M_n = \max_{1\leq i\leq n} X_i$$

and so

$$\mathbb{P}(M_n \geq t) = \mathbb{P}\left(\bigcup_{i=1}^n X_i > t\right)$$

We have

$$\mathbb{P}(X_i > t) = \frac{1}{2^{\lceil \log_2 t\rceil}} \in \left(\frac{1}{2t}, \frac{1}{t}\right]$$

So we get

$$\mathbb{P}(M_n \geq n\log n \text{ i.o.}) = 1$$

and

$$\mathbb{P}\big(M_n \geq n(\log n)^{1+\varepsilon} \text{ i.o.}\big) = 0$$

10.3. **Strengthening of Borel-Cantelli 2.**

Theorem 12. *If E_i, $i \geq 1$ are events, (not necessarily independent), and $\sum_{i\geq 1} \mathbb{P}(E_i) = \infty$, then*

$$\mathbb{P}(E_i \ i.o.) \geq \limsup_{n\to\infty} \frac{\left(\sum_{i=1}^{n} \mathbb{P}(E_i)\right)^2}{\sum_{i\leq i,j\leq n} \mathbb{P}(E_i \cap E_j)}$$

This is typically referred to as the Chung-Erdős inequality.

10.4. **Kolmogorov 0-1 Law.** $(X_n, n \geq 1)$ independent random variables.
Let our tail,

$$\mathcal{T}_n = \sigma(X_n, X_{n+1}, \dots)$$

then let

$$\mathcal{T} = \bigcap_{n\geq 1} \mathcal{T}_n$$

This is a σ-algebra and we call it the tail σ-algebra.
Exercises: The following events are all in τ

- $\lim_{n\to\infty} X_n$ exists
- $\sum_{k\geq 1} X_k$ converges ($\lim_{h\to\infty} \sum_{j=1}^{h} X_j$ exists)
- $\lim\limits_{k\to\infty} \dfrac{X_1 + \dots + X_k}{k}$ exists

Theorem 13 (Kolmogorov's 0-1 Law)**.** *For any $E \in \mathcal{T}$, $\mathbb{P}(E) \in \{0, 1\}$.*

Proof. We'll show that $\mathcal{T}$ is independent of $\mathcal{T}$. Let

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n)$$

Then $\mathcal{F}_n$ and $\mathcal{T}_{n+1}$ are independent. So $\mathcal{F}_n$ and $\mathcal{T}$ are independent. Let

$$\mathcal{P} = \lim_{n\to\infty} \mathcal{F}_n = \bigcup_{n\geq 1} \mathcal{F}_n$$

Then $\mathcal{P}$ and $\mathcal{T}$ are independent. $\mathcal{P}$ is closed under intersections, so it is a π-system. So by the π-system lemma, $\sigma(\mathcal{P})$ and $\mathcal{T}$ are independent.

We are done since

$$\begin{aligned}\sigma(\mathcal{P}) &= \sigma\left(\bigcup_{n\geq 1} \sigma(X_1, \dots, X_n)\right)\\ &\supset \sigma\left(\bigcup_{n\geq 1} \sigma(X_n)\right)\\ &= \sigma(\{X_n\}_{n\geq 1})\end{aligned}$$

So $\sigma(\{X_n\}_{n\geq 1})$ and $\mathcal{T}$ are independent, but $\mathcal{T}_n \subset \sigma(\{X_k\}_{k\geq 1})$ for all n, so

$$\bigcap_{n\geq 1} \mathcal{T}_n = \mathcal{T} \subset \sigma(\{X_k\}_{k\geq 1})$$

so $\mathcal{T}$ and $\mathcal{T}$ are independent. □

Corollary. *If Y is a random variable with $\sigma(Y) \subset \mathcal{T}$ then $\exists c \in [-\infty, \infty]$ such that $\mathbb{P}(Y = c) = 1$.*

Proof. Let $x = \sup\{y : \mathbb{P}(Y \leq y) = 0\}$. For any $z > x$, $\mathbb{P}(Y \leq z) = 1$ since we are in the tail. So $\mathbb{P}(Y = x) = 1$. □

In our branching process example, then

$$\frac{Z_n}{(\mathbb{E}[B])^n} \xrightarrow{\text{a.s.}} W$$

where W is not a constant.

11. October 18th, 2012

(1) Kolmogorov Extension Theorem
If $(X_n, n \geq 1)$ are iid random variables and
$$S_n = X_1 + \dots X_n \qquad (S_n, n \geq 1)$$
with
$$M_n = \max\{S_i,\ 1 \leq i \leq n\} \qquad (M_n, n \geq 1)$$

11.1. Integration and Expectation.

DIAGRAM

When looking at integration, let us start simple, by looking at constant functions. Say we have
$$f(\omega) = 7\ \forall \omega$$
Then we have
$$\{\omega : f(\omega) \in B\} = \begin{cases} \emptyset & B \in 7 \\ \Omega & B \notin 7 \end{cases}$$
So we have
$$\sigma(f) = \{\emptyset, \Omega\}$$
We also have
$$\sigma(\mathbb{1}_E) = \{\emptyset, \Omega, E, E^c\}$$

Definition 16. *Given a measure space $(S, \mathcal{F}, \mu)$, for $E \in \mathcal{F}$ and a constant $c > 0$, let*
$$\int c\mathbb{1}_E \,\mathrm{d}\mu = c\mu(E)$$
More generally, given $E_1, \dots, E_k \in \mathcal{F}$ and $c_1, \dots, c_k > 0$, let
$$\begin{aligned} \int \sum_{i=1}^k c_i \mathbb{1}_{E_i} \,\mathrm{d}\mu &= \sum_{i=1}^k \int c_i \mathbb{1}_{E_i} \,\mathrm{d}\mu \\ &= \sum_{i=1}^k c_i \mu(E_i) \end{aligned}$$
where a linear combination of indicators
$$\sum_{i=1}^k c_i \mathbb{1}_{E_i}$$
is called a simple function. We allow sets of size infinity, so according to this, we make it convention that $0 \cdot \infty = 0$.

Exercise: If
$$f = \sum_{i=1}^k c_i \mathbb{1}_{E_i} = \sum_{j=1}^l b_i \mathbb{1}_{D_i} = g$$
then
$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$$
i.e.
$$\sum_{i=1}^k c_i \mu(E_i) = \sum_{j=1}^l b_j \mu(D_j)$$
We show this by writing
$$E_i = E_i \cap \left(\bigcup_{j=1}^l D_j \right)$$

Prop. *If f, g are simple functions, then*
(1) *If $\mu(f \neq g) = 0$ then*
$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$$

(2) *If $c \geq 0$, then*

$$\int (cf+g)\,\mathrm{d}\mu = c\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$$

This can also be written as $\mu(cf+g)$

(3) *If $f \leq g$, then*

$$\int f\,\mathrm{d}\mu \leq \int g\,\mathrm{d}\mu$$

(4) *$f \vee g$, and $f \wedge g$ are again simple.*

$$f \vee g = \max\{f, g\}$$
$$f \wedge g = \min\{f, g\}$$

Definition 17. *If $f : S \to \mathbb{R}$ is non-negative and measurable then let*

$$\int f\,\mathrm{d}\mu = \sup\left\{\int g\,\mathrm{d}\mu,\ g \text{ simple}, g \leq f\right\}$$

We note that we could take functions g such that $\mu(g > f) = 0$ insetad.

Theorem 14 (Monotone Convergence Theorem)**.** *If f_n, f are non-negative measurable functions and $f_n \uparrow f$ as $n \to \infty$, then*

$$\lim_{n\to\infty}\int f_n\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$$

Lemma: For each function, we will have a change of increasing simple functions. We have $f_1, f_2, \ldots, f_n, f_{n+1}, \ldots, f$. and $f_n^{(1)}, f_n^{(2)}, \ldots f_n^{(k)}, \ldots f_n$ to approximate f_n.

Lemma (1)**.** *If $(a_n^{(k)})_{n,k\geq 1}$ is increasing in n and in k., then*

$$\lim_{n\to\infty}\underbrace{\lim_{k\to\infty} a_n^{(k)}}_{a_n^{(\infty)}} = \lim_{k\to\infty}\underbrace{\lim_{n\to\infty} a_n^{(k)}}_{a_\infty^{(k)}}$$

Fix an n.

$$\begin{array}{cccc} \ldots & a_n^{(k)} & \ldots & a_n^{(\infty)} \\ & a_{n+1}^{(k)} & & \\ & \vdots & & \\ & a_\infty^{(k)} & & \end{array}$$

For each k, all n, $a_\infty^{(k)} \geq a_n^{(k)}$, so

$$\lim_{k\to\infty}\underbrace{a_\infty^{(k)}}_{\lim_{n\to\infty} a_n^{(k)}} \geq \lim_{n\to\infty}\lim_{k\to\infty} a_n^{(k)}$$

By symmetry the reverse inequality holds and this completes the proof.

Lemma (2)**.** *If $E \in \mathcal{F}$, f_n are simple functions, and $f_n \uparrow \mathbb{1}_E$, then*

$$\int f_n\,\mathrm{d}\mu \uparrow \int \mathbb{1}_E\,\mathrm{d}\mu = \mu(E)$$

Proof. $\Leftarrow$ is proven by property (iii).
Now fix $\varepsilon > 0$ and let $E_n = \mathbb{1}\{f_n > 1-\varepsilon\}$. Since $f_n \uparrow f$, $E_n \uparrow E$. Then

$$f_n \geq (1-\varepsilon)\mathbb{1}_{E_n}$$

so

$$\begin{aligned}\lim_{n\to\infty}\int f_n\,\mathrm{d}\mu &\geq (1-\varepsilon)\lim_{n\to\infty}\int \mathbb{1}_{E_n}\,\mathrm{d}\mu \\ &= (1-\varepsilon)\lim_{n\to\infty}\mu(E_n) \\ &= (1-\varepsilon)\mu(E)\end{aligned}$$

$\varepsilon > 0$ was arbitrary so we are done. □

Lemma (3)**.** *If g is a simple function, and $f_n \uparrow g$, then*

$$\int f_n \,\mathrm{d}\mu \uparrow \int g \,\mathrm{d}\mu$$

Proof. Write

$$g = \sum_{i=1}^{k} c_i \mathbb{1}_{E_i},\ E_1, \dots, E_k \text{ disjoint } \bigcup_{i=1}^{k} E_i = S$$

Then

$$f_n = \sum_{i=1}^{n} f_n \cdot \mathbb{1}_{E_i}$$

so

$$\int f_n \,\mathrm{d}\mu = \sum_{i=1}^{k} \int f_n \mathbb{1}_{E_i} \,\mathrm{d}\mu$$

Then we have

$$\begin{aligned}
\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu &= \lim_{n\to\infty} \sum_{i=1}^{k} \int f_n \mathbb{1}_{E_i} \,\mathrm{d}\mu \\
&= \sum_{i=1}^{k} \lim_{n\to\infty} \int f_n \mathbb{1}_{E_i} \,\mathrm{d}\mu \\
&= \sum_{i=1}^{k} \lim_{n\to\infty} c_i \int \underbrace{\frac{1}{c_i} f_n \mathbb{1}_{E_i}}_{\text{simple } \uparrow \mathbb{1}_{E_i}} \,\mathrm{d}\mu \\
&= \sum_{i=1}^{\infty} c_i \mu(E_i) \qquad \text{by Lemma 2} \\
&= \int g \,\mathrm{d}\mu
\end{aligned}$$

□

Lemma (4)**.** *If $f \geq 0$ is measurable, f_n, g_n are simple, and $f_n \uparrow f$, $g_n \uparrow f$, then*

$$\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$$

Proof. Let $h_n^{(k)} = \min(f_n, g_k)$. We know that $h_n^{(k)}$ is simple and increasing in n and in k.

$$\lim_{n\to\infty} h_n^{(k)} = g_k$$

so by Lemma 3,

$$\int g_k \,\mathrm{d}\mu = \int \lim_{n\to\infty} h_n^{(k)} \,\mathrm{d}\mu$$

Similarly,

$$\lim_{k\to\infty} h_n^{(k)} = f_n$$

so

$$\int f_n \,\mathrm{d}\mu = \int \lim_{k\to\infty} h_n^{(k)} \,\mathrm{d}\mu$$

Finally,

$$\begin{aligned}
\lim_{k\to\infty} \int g_k \,\mathrm{d}\mu &= \lim_{k\to\infty} \int \lim_{n\to\infty} h_n^{(k)} \,\mathrm{d}\mu \\
&= \lim_{k\to\infty} \lim_{n\to\infty} \int h_n^{(k)} \,\mathrm{d}\mu \\
&= \lim_{n\to\infty} \lim_{k\to\infty} \int h_n^{(k)} \,\mathrm{d}\mu \qquad \text{by Lemma 1} \\
&= \lim_{n\to\infty} \int \lim_{k\to\infty} h_n^{(k)} \,\mathrm{d}\mu \qquad \text{by Lemma 3}
\end{aligned}$$

$$= \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu$$

□

Lemma (5). *If f_n are simple and $f_n \uparrow f$, where f is measurable, then*

$$\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$$

$\Leftarrow$ is obvious
$\Rightarrow$ Pick a sequence g_n of simple functions such that

$$\lim_{n\to\infty} \int g_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$$

Let $h_n = \max(g_1, \dots g_n, f_1, \dots f_n) \geq g_n$ but it is also still $\leq f$. Then

$$\lim_{n\to\infty} \int h_n \,\mathrm{d}\mu \geq \lim_{n\to\infty} \int g_n \,\mathrm{d}\mu$$

12. October 23rd, 2012

12.1. Assignment Stuff.

Question 2.16

$$\mathcal{F}_3 = \sigma(\{\omega_1 = 1\}, \{\omega_2 = 1\}, \{\omega_3 = 1\})$$

If we have disjoint cylinders

$$A_1, \dots, A_r$$

then

$$\mathbb{P}(A_1 \cup \ldots \cup A_r) = \sum_{i=1}^{r} \frac{1}{2^{\mathrm{rank}(A_i)}}$$

$$\bigcup \mathcal{F}_i = \sigma(\text{cylinders of rank } i)$$

which ends up being

$$\sigma\left(\bigcup_{i\geq 1} \sigma(\{\omega_1 = 1\}, \dots, \{\omega_i = 1\})\right)$$

2.11: σ-algebras are never countable.
Observation: If a σ-algebra $\mathcal{F}$ contains an infinite sequence, $(E_i, i \geq 1)$ of pairwise disjoint sets, then it is uncountable.

Proof. The uncountable set

$$\left\{\bigcup_{i\in S} : S \subset \mathbb{N}\right\} \subset \mathcal{F}$$

Suppose that $\mathcal{F}$ is a countably infinite σ-algebra. For each $x \in \mathcal{F}$, let

$$E_x = \bigcap_{\substack{E \supset x \\ E \in \mathcal{F}}} E \in \mathcal{F}$$

Now we have that for $x, y \in \Omega$, E_x and E_y are either identical or pairwise disjoint. A good <u>exercise</u> is to show that if $\{E_x : x \in \Omega\}$ is finite, then $\mathcal{F}$ is finite. If $\{E_x : x \in \Omega\}$ is ∞, then $\mathcal{F}$ is uncountable. □

2.10c: We look at

$$\mathcal{F}_1 = \mathfrak{B}(\mathbb{R})$$

and

$$\begin{aligned}
\mathcal{F}_2 &= \{\text{sets that are countable or co-countable}\} \\
&\stackrel{?}{=} \sigma(A_i, i \geq 1) \qquad \text{WLOG all } A_i \text{ countable} \\
&\subseteq \sigma\left(\{x\} : x \in \bigcup_{i\geq 1} A_i\right) \\
&= \{S : \text{either } S \subset \bigcup A_i \text{ or } S^c \subset \bigcup A_i\}
\end{aligned}$$

Now any countable set

$$B \subset \left(\bigcup_{i\geq 1} A_i\right)^c$$

is not in $\sigma(A_i, i \geq 1)$.

12.2. **Monotone Convergence Theorem.** If $f_n \geq 0$ are measurable and $f_n \uparrow f$ where f is also measurable, then

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$$

Lemma (5)**.** *If f_n are simple and $f_n \uparrow f$, where f is measurable, then*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$$

Proof. $\Leftarrow$ is obvious
$\Rightarrow$ We use the fact that

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

Pick a sequence g_n of simple functions such that

$$\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$$

Let $h_n = \max(g_1, \ldots g_n, f_1, \ldots f_n) \geq g_n$ but it is also still $\leq f$. Then

$$\lim_{n\to\infty} \int h_n \, \mathrm{d}\mu \geq \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu$$

We have that $h_n \uparrow f$ so

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int h_n \, \mathrm{d}\mu$$

Finally, $h_n \leq f$ for all n, so

$$\int h_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$$

for all n, so

$$\lim_{n\to\infty} \int h_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$$

Also, $h_n \geq g_n$ for all n, so

$$\int h_n \, \mathrm{d}\mu \geq \int g_n \, \mathrm{d}\mu$$

for all n, so

$$\lim_{n\to\infty} \int h_n \, \mathrm{d}\mu \geq \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$$

□

12.3. **Proof of Monotone Convergence Theorem.**

Proof. Let $\alpha^{(k)} : [0, \infty) \to [0, \infty)$ be

$$a^{(k)}(x) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{i}{2^k} & \text{if } x = \left(\frac{i}{2^k}, \frac{i+1}{2^k}\right], \ i \leq k2^k \\ k & \text{if } x > k \end{cases}$$

For any measurable function $g \geq 0$, $\alpha^{(k)}(g)$ is simple and $\alpha^{(k)}(g) \uparrow g$ as $k \uparrow \infty$. We have $\alpha^{(k)}(f_n) \uparrow \alpha^{(k)}(f)$ as $n \to \infty$ and $\alpha^{(k)}(f_n) \uparrow f_n$ as $k \to \infty$. So Lemma 3 says that

$$\lim_{n\to\infty} \int \alpha^{(k)}(f_n) \, \mathrm{d}\mu = \int \alpha^{(k)}(f) \, \mathrm{d}\mu$$

and Lemma 5 says

$$\lim_{k\to\infty} \int \alpha^{(k)}(f_n) \, \mathrm{d}\mu = \int f_n \, \mathrm{d}\mu$$

taking limits on both sides with respect to n, we get

$$
\begin{aligned}
\lim_{n\to\infty}\int f_n \,\mathrm{d}\mu &= \lim_{n\to\infty}\lim_{k\to\infty}\int \alpha^{(k)}(f_n)\,\mathrm{d}\mu \\
&= \lim_{k\to\infty}\lim_{n\to\infty}\int \alpha^{(k)}(f_n)\,\mathrm{d}\mu \qquad \text{by Lemma 1} \\
&= \lim_{k\to\infty}\int \alpha^{(k)}(f)\,\mathrm{d}\mu \\
&= \int f\,\mathrm{d}\mu \qquad \text{by Lemma 5}
\end{aligned}
$$

□

Definition 18. *We say that measurable functions $f, g : \Omega \to \mathbb{R}$ are almost everywhere equal if*

$$\mu(\{f \neq g\}) = 0$$

Exercise:

$$0 \underset{a.s.}{\leq} f_1 \underset{a.s.}{\leq} f_2 \underset{a.s.}{\leq} \dots \underset{a.s.}{\leq} f$$

and

$$\lim_{n\to\infty} f_n \overset{a.e.}{=} f$$

then

$$\int f_n \,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$$

12.4. **Fatou's Lemma for Functions.** If $f_n \geq 0$ are all measurable, then

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

Proof. Let

$$g_n = \inf_{m\geq n} f_m$$

So

$$\int \liminf_{n\to\infty} f_n\,\mathrm{d}\mu = \int \lim_{n\to\infty} g_n\,\mathrm{d}\mu = \lim_{n\to\infty}\int g_n\,\mathrm{d}\mu$$

Now we use the fact that

$$\int \min(h_1, h_2)\,\mathrm{d}\mu \leq \min\left(\int h_1\,\mathrm{d}\mu, \int h_2\,\mathrm{d}\mu\right)$$

so

$$
\begin{aligned}
\int \liminf_{n\to\infty} f_n\,\mathrm{d}\mu &= \lim_{n\to\infty}\int g_n\,\mathrm{d}\mu \\
&= \lim_{n\to\infty}\int \inf_{m\geq n} f_m\,\mathrm{d}\mu \\
&\leq \lim_{n\to\infty}\inf_{m\geq n}\int f_m\,\mathrm{d}\mu \\
&= \liminf_{n\to\infty}\int f_n\,\mathrm{d}\mu
\end{aligned}
$$

□

12.5. **Reverse Fatou.** If $f_n \geq 0$ are measurable and $g \geq 0$ are measurable such that $f_n \leq g$ $\forall n$, then

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

Idea: Apply Fatou to $f_n', n \geq 1$ where $f_n' = g - f_n$. Then use linearity of integration.

13. October 30th, 2012

13.1. **General Integral and L^1.** $(S, \mathcal{S}, \mu)$, $f \geq 0$ with $\int f \, \mathrm{d}\mu$ defined.
For any function $f : S \to \mathbb{R}$ measurable, we have that

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

are defined where

$$\begin{aligned} f^+ &= \max(f, 0) \\ f^- &= -\min(f, 0) \end{aligned}$$

This gives us

$$\begin{aligned} f &= f^+ - f^- \\ |f| &= f^+ + f^- \end{aligned}$$

We saw that $f \in L^1(S, \mathcal{S}, \mathrm{d}\mu)$ or that f is integrable if

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

In this case, we let

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

Check:

- Well-defined
- Monotone
- Linear
- Etc...

Let us check linearity. If we have $f, g \in L^1(S, \mathcal{S}, \mu)$, and

$$\int (cf + g) \, \mathrm{d}\mu = \int (cf + g)^+ \, \mathrm{d}\mu + \int (cf + g)^- \, \mathrm{d}\mu$$

Let $E = \{cf + g \geq 0\}$ so we have

$$\begin{aligned} \int (cf + g) \, \mathrm{d}\mu &= \int (cf + g) \mathbb{1}_E \, \mathrm{d}\mu - \int -(cf + g) \mathbb{1}_{E^c} \, \mathrm{d}\mu \\ &= c \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu \end{aligned}$$

The two integrals actually end up expanding into 8 integrals and combine back to get to the final line.

Basic Question: Given functions f_n, f, when does

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

(Usually we ask this when $f_n \overset{a.e.}{\to} f$)
Observation: $f_n \overset{a.e.}{\to} f$ does not imply $\int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$

Example 17. $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Leb)$
Take

$$f_n = \frac{1}{n} \mathbb{1}_{[-n,n]}$$

and take $f \equiv 0$. *Then we have* $f_n \overset{a.e.}{\to} f$ *But*

$$\int f_n \, \mathrm{d}\mu = 2$$

for all n, *but*

$$\int f \, \mathrm{d}\mu = 0$$

13.2. **Dominated Convergence Theorem.** If $f_n \overset{a.e.}{\to} f$ and $\exists g \in L^1$ such that $|f_n| \le g$ for all n then

$$\int |f_n - f| \, \mathrm{d}\mu \to 0$$

as $n \to \infty$.

Note:

$$\int f_n \, \mathrm{d}\mu - \int f \, \mathrm{d}\mu = \int (f_n - f) \, \mathrm{d}\mu \le \int |f_n - f| \, \mathrm{d}\mu$$

which implies if

$$\int |f_n - f| \, \mathrm{d}\mu \to 0$$

then

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

Exercise: If $\int |f_n - f| \, \mathrm{d}\mu \to 0$, then $\int |f_n| \, \mathrm{d}\mu \to \int |f| \, \mathrm{d}\mu$.

Definition 19. *For measurable functions f_n, f, we say $f_n \to f$ in $L^1(S, \mathcal{S}, \mu)$, or $f_n \overset{L^1}{\to} f$ if*

$$\int |f_n - f| \, \mathrm{d}\mu \to 0$$

Proof of Dominated Convergence Theorem. Apply Reverse Fatou's Lemma to

$$\int |f_n - f| \, \mathrm{d}\mu$$

Note that

$$|f_n - f| \le |f_n| + |f| \le \underbrace{|g| + |f|}_{\le 2|g|} \in L^1(S, \mathcal{S}, \mu)$$

So

$$\begin{aligned} \limsup_{n\to\infty} \int |f_n - f| \, \mathrm{d}\mu &\le \int \limsup_{n\to\infty} |f_n - f| \, \mathrm{d}\mu \\ &= \int 0 \, \mathrm{d}\mu \\ &= 0 \end{aligned}$$

so we have

$$\lim_{n\to\infty} \int |f_n - f| \, \mathrm{d}\mu = 0$$

□

Observation 2: Together, the two properties from the first observation, that is $f_n \overset{a.e.}{\to} f$ and $\int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$ implies

$$f_n \overset{L^1}{\to} f$$

Theorem 15 (Scheffé). *If $f_n, f \in L^1(S, \mathcal{S}, \mu)$ and $f_n \overset{a.e.}{\to} f$, then*

$$\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu \Leftrightarrow \int |f_n - f| \, \mathrm{d}\mu \to 0$$

Proof. Need to prove only "⇒" as the other direction was shown above. So let us assume that $\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu$. We want to show that

$$\int |f_n - f| \, \mathrm{d}\mu = \int (f_n - f)^+ \, \mathrm{d}\mu + \int (f_n - f)^- \, \mathrm{d}\mu \to 0$$

We have that

$$(f_n - f)^- = -\min(f_n - f, 0) = \begin{cases} 0 & f_n \ge f \\ f - f_n & f_n < f \end{cases}$$

So if $f_n \ge 0$, then $(f_n - f)^- \le f$. By dominated convergence theorem,

$$\int (f_n - f)^- \, \mathrm{d}\mu \to \int 0 \, \mathrm{d}\mu = 0$$

Also,

$$\begin{aligned}\int (f_n - f)^+ \, \mathrm{d}\mu &= \int (f_n - f)\mathbb{1}_{\{f_n \geq f\}} \, \mathrm{d}\mu \\ &= \int (f_n - f)(1 - \mathbb{1}_{\{f_n < f\}}) \, \mathrm{d}\mu \\ &= \int (f_n - f) \, \mathrm{d}\mu - \int (f_n - f)\mathbb{1}_{\{f_n < f\}} \, \mathrm{d}\mu \\ &= \int f_n \, \mathrm{d}\mu - \int f \, \mathrm{d}\mu + \int (f_n - f)^- \, \mathrm{d}\mu\end{aligned}$$

which goes to 0, so $f_n \xrightarrow{L^1} f$. □

13.3. **Moving Bump.** $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Leb)$
Let us define a function

$$f_m = \underbrace{f_{2^n+k}}_{0 \leq k \leq 2^{n-1}} := \mathbb{1}\left\{\frac{k}{2^n}, \frac{k+1}{2^n}\right\}$$

Let $f \equiv 0$. Then $f_n, f \in L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Leb)$ that is they are all integrable.

$$\int |f_m - f| \, \mathrm{d}\mu = \int |f_m| \, \mathrm{d}\mu = \int f_m \, \mathrm{d}\mu = \frac{1}{2^n} \leq \frac{2}{m} \to 0$$

when $m \in [2^n, 2^{n+1})$. Hence $f_m \xrightarrow{L^1} f$. But $\forall x \in [0,1]$, $\limsup_{n\to\infty} f_n(x) = 1$ so $f_n \stackrel{a.e.}{\not\to} f$.

13.4. **Integrals Over Subsets. Recall:** Given a measurable space $(S, \mathcal{S}, \mu)$ and $A \in \mathcal{S}$, we can define a new measure μ_A on $(A, \mathcal{S}_A)$ where

$$\mathcal{S}_A = \{E \cap A : E \in \mathcal{S}\} = \{E \in S : E : E \subseteq A\}$$

by

$$\mu_A(E) = \mu(E)$$

Check: μ_A is a measure on $(A, \mathcal{S}_A)$. Now, given $f \in L^1(S, \mathcal{S}, \mu)$, let

$$\int_A f \, \mathrm{d}\mu = \int f \cdot \mathbb{1}_A \, \mathrm{d}\mu$$

In particular, $\int_S f \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$. Defining $f_A : A \to \mathbb{R}$ by $f_A(x) = f(x)$, then we should have

$$\int_A f \, \mathrm{d}\mu = \int f_A \, \mathrm{d}\mu_A$$

14. November 1st, 2012

Proof. This is referred to as the *standard machine.*

(1) Indicator functions
If $f = \mathbb{1}_B$, then since $f_A : A \to \mathbb{R}$, $f_A = \mathbb{1}_{A\cap B}$. Then

$$\begin{aligned}\int_A f \, \mathrm{d}\mu &= \int f \mathbb{1}_A \, \mathrm{d}\mu \\ &= \int \mathbb{1}_{A\cap B} \, \mathrm{d}\mu \\ &= \mu(A \cap B) \\ &= \mu_A(A \cap B) \\ &= \int \mathbb{1}_{A\cap B} \, \mathrm{d}\mu_A\end{aligned}$$

(2) Simple Functions
Suppose

$$f = \sum_{i=1}^{n} c_i \mathbb{1}_{B_i}$$

Then

$$\begin{aligned}
\int_A f \,\mathrm{d}\mu &= \int \left(\sum_{i=1}^{n} c_i \mathbb{1}_{B_i} \right) \mathbb{1}_A \,\mathrm{d}\mu \\
&= \int \sum_{i=1}^{n} c_i \mathbb{1}_{A\cap B_i} \,\mathrm{d}\mu \\
&= \sum_{i=1}^{n} c_i \int \mathbb{1}_{A\cap B_i} \,\mathrm{d}\mu \\
&= \sum_{i=1}^{n} c_i \int \mathbb{1}_{A\cap B_i} \,\mathrm{d}\mu_A \\
&= \int \underbrace{\left(\sum_{i=1}^{n} c_i \mathbb{1}_{B_i} \right) \mathbb{1}_A}_{f_A} \,\mathrm{d}\mu_A \\
&= \int f_A \,\mathrm{d}\mu_A
\end{aligned}$$

One thing to be careful with is when we define the indicator, since $\omega \notin A$ means $\omega \in S \setminus A$, but with respect to $\mathbb{1}_{A\cap B}$ if $\omega \notin A \cap B$, we have a different S as our whole set. In this case, it is A.

(3) $f \geq 0$ Let $f_n \geq 0$, $f_n \uparrow f$, f_n simple.
Then

$$f_n \mathbb{1}_A \uparrow f \mathbb{1}_A$$

so

$$\begin{array}{ccc}
\int_A f_n \,\mathrm{d}\mu & \longrightarrow & \int f \,\mathrm{d}\mu \\
\Vert & & \\
\int f_{n,A} \,\mathrm{d}\mu & \xrightarrow[f_{n,A}\uparrow f_A]{} & \int f_A \,\mathrm{d}\mu_A
\end{array}$$

(4) $f \in L^1$, $f = f^+ - f^-$
Then

$$\begin{aligned}
\int_A f \,\mathrm{d}\mu &= \int (f^+ - f^-) \mathbb{1}_A \,\mathrm{d}\mu \\
&= \int_A f^+ \,\mathrm{d}\mu - \int_A f^- \,\mathrm{d}\mu \\
&= \int f_A^+ \,\mathrm{d}\mu_A - \int f_A^- \,\mathrm{d}\mu_A \\
&= \int f_A^+ - f_A^- \,\mathrm{d}\mu_A \\
&= \int f_A \,\mathrm{d}\mu_A
\end{aligned}$$

□

14.1. **Change of Measure.** $(S, \mathcal{S}, \mu)$ $f \geq 0,\ f \in L^1$
Then define a measure ν on $(S, \mathcal{S})$ by

$$\nu(E) = \int_E f \,\mathrm{d}\mu$$

Bayesian Idea: Updating belief in light of evidence. Is this a measure though? Let us check that. We need that if $E_i, i \geq 1$ are disjoint, then

$$\nu\left(\bigcup_{i\geq 1} E_i \right) = \sum_{i\geq 1} \nu(E_i)$$

or equivalently, if $F_n \uparrow F$, then

$$\nu(F_n) \to \nu(F)$$

We have

$$f\mathbb{1}_{F_n} \to f\mathbb{1}_F$$

so monotone convergence theorem says that

$$\underbrace{\int_{E_n} f \, \mathrm{d}\mu}_{=\nu(E_n)} \to \underbrace{\int_E f \, \mathrm{d}\mu}_{=\nu(E)}$$

We write $f = \frac{d\nu}{d\mu}$ or $\mathrm{d}\nu = f \, \mathrm{d}\mu$, because

Prop. *If $g \geq 0$, which is measurable then*

$$\int g \, \mathrm{d}\nu = \int gf \, \mathrm{d}\mu$$

Proof. Standard machine again: If $g = \mathbb{1}_E$, then

$$\int g \, \mathrm{d}\nu = \nu(E) = \int_E f \, \mathrm{d}\mu = \int f\mathbb{1}_E \, \mathrm{d}\mu = \int fg \, \mathrm{d}\mu$$

We are only showing this for indicators then the remaining steps follow the same as before. ☐

14.2. **Expectation.** $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. If $X \geq 0$ is a random variable or $X \subset L^1(\mathbb{P})$, then let

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

Example 18.

$$X = \begin{cases} 1 & \textit{with prob. } \frac{1}{2} \\ 0 & \textit{with prob. } \frac{1}{2} \end{cases}$$

that is $\exists E, \mathbb{P}(E) = \frac{1}{2}$ such that $X = \mathbb{1}_E$.

$$\mathbb{E}[X] = \int \mathbb{1}_E \, \mathrm{d}\mathbb{P} = \mathbb{P}(E) = \frac{1}{2}$$

Theorem 16 (Bounded Convergence Theorem)**.** $(\Omega, \mathcal{F}, \mathbb{P})$
If X_n, $n \geq 0$ is a sequence of random variables and $X_n \xrightarrow{a.s.} X$ where X is a random variable and $\exists K > 0$ such that $|X_n| < K$ for all n, then

$$\mathbb{E}[|X_n - X|] \to 0$$

This is just the Dominated Convergence Theorem where we take $g = K\mathbb{1}_\Omega$ and this works since we are in a space of finite measure.

14.3. **Jensen's Inequality.**

Theorem 17 (Jensen's Inequality)**.** *We have X is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is convex. If $X \in L^1(\mathbb{P})$ and $\varphi(X) \in L^1(\mathbb{P})$, then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Proof. (Convex functions are continuous) Using this, we know that φ is a measurable function and so $\varphi(X)$ is a valid random variable.
Claim: $\forall x \in \mathbb{R}$, $\exists a, b$ such that $l(y) = ay + b$ satisfies $l(x) = \varphi(x)$ and $l(y) \leq \phi(y) \; \forall y$.

Proof. **Idea:** φ has left and right derivatives at x and $\varphi'(x^-) \leq \varphi'(x^+)$. By convexity,

$$\varphi(x+h) - \varphi(x) \geq \varphi(x) - \varphi(x-h)$$

$(\varphi(x+h) - \varphi(x-h) \geq 2\varphi(x))$ so

$$\frac{\varphi(x+h) - \varphi(x)}{h} \geq \frac{\varphi(x) - \varphi(x-h)}{h}$$

The left-hand side is decreasing in h and the right-hand side is inreasing in h.

$$\varphi'(x^+) = \lim_{h \downarrow 0} \frac{\varphi(x+h) - \varphi(x)}{h} \qquad \varphi'(x^-) = \lim_{h \downarrow 0} \frac{\varphi(x) - \varphi(x-h)}{h}$$

The same proof shows that $\forall c > 0$,

$$\varphi(x + c) - \phi(x) \geq c\varphi'(x^+)$$

and

$$\varphi(x) - \varphi(x - c) \leq c\varphi'(x^-)$$

$\square$

Let a be any constant with $\varphi'(x^-) \leq a \leq \varphi'(x^+)$. Then we have

$$l(y) = ay + (\varphi(x) - ax)$$

This is the line passing through $(x, \varphi(x))$ with slope a. We can see this since

$$\frac{l(y) - \varphi(x)}{y - x} = a$$

We also have

$$l(x) = \varphi(x)$$

Now let $\varphi(\mathbb{E}[X]) = l(\mathbb{E}[X])$, with $l \leq f$ and l is of the form $l(x) = ax+b$. Then by linearity of expectation,

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[l(X)] = a\mathbb{E}[X] + b = l(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$$

$\square$

Question: Where in the proof did I use that this is a probability space?

Theorem 18 (Markov's Inequality). *If $X \geq 0$ is a random variable with $\mathbb{E}[X] < \infty$. Then $\forall t > 0$,*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

Proof.

$$\begin{aligned} \mathbb{E}[X] &\geq \mathbb{E}[X\mathbb{1}_{X \geq t}] \\ &\geq \mathbb{E}[t\mathbb{1}_{X \geq t}] \\ &= t\mathbb{E}[\mathbb{1}_{X \geq t}] \\ &= t\mathbb{P}(X \geq t) \end{aligned}$$

$\square$

Corollary (1.). *If $\mathbb{E}\left[X^2\right]$ is finite then*

$$\mathbb{P}(|X - \mathbb{E}[X]\,| \geq t) \leq \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}{t^2}$$

Proof. Let $Y = (X - \mathbb{E}[X])^2$. Then

$$\mathbb{P}(|X - \mathbb{E}[X]\,| \geq t) = \mathbb{P}\big(Y \geq t^2\big) \leq \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}{t^2}$$

However we do not know that $\mathbb{E}[Y] < \infty$. Using the fact that

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

One more thing is missing though since we do not know that $\mathbb{E}[X] < \infty$. $\square$

Corollary (2.). *If $X \geq 0$ is a random variable and $g : \mathbb{R} \to [0, \infty)$ is increasing and $\mathbb{E}[g(X)] < \infty$, then*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[g(X)]}{g(t)}$$

We can see this as an easy application of Markov's inequality to $g(X)$ as

$$X \geq t \Leftrightarrow g(X) \geq g(t)$$

Definition 20. *For random variable X and $p \in [1,\infty)$, we write*

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$$

which is the L^p-norm. If $\|X\|_p < \infty$, we say that $X \in L^p(\mathbb{P})$

Definition 21. *The essential supremum of X is*

$$\text{esssup}X := \sup(r : \mathbb{P}(X \geq r) > 0)$$

Exercise: $\|X\|_\infty := \lim_{p\to\infty} \|X\|_p = \text{esssup}|X|$

Prop. *If $1 \leq p \leq r$ then $\|X\|_p \leq \|X\|_r$.*

Proof. Idea: Jensen: $\phi : \mathbb{R} \to \mathbb{R}$ where $\phi(x) = x^{r/p}$

$$\begin{aligned}\|X\|_r &= \mathbb{E}[|X|^r]^{1/r}\\ &= \mathbb{E}[\underbrace{(|X|^p)^{r/p}}_{\phi(|X|^p)}]^{1/r}\\ &\geq \phi(\mathbb{E}[|X|^p])^{1/r}\\ &= \mathbb{E}[|X|^p]^{\frac{r}{p}\frac{1}{r}}\\ &= \|X\|_p\end{aligned}$$

But there is a caveat since we are using the fact that $\mathbb{E}[|X|^p] < \infty$ in the proof. To deal with this, we truncate. We write

$$|X|^{\leq N} = |X| \cdot \mathbb{1}_{|X|\leq N}$$

Then $|X|^{\leq N}$ is bounded and $|X|^{\leq N} \uparrow |X|$. So

$$\|X\|_r \overset{MCT}{=} \lim_{N\to\infty} \||X|^{\leq N}\|_r \geq \lim_{N\to\infty} \||X|^{\leq N}\|_p \overset{MCT}{=} \|X\|_p$$

□

15.1. Bounding L^p-norm of X, Y given norm bound on X, Y.

Prop. *If $X, Y \in L^1$ and $X \perp\!\!\!\perp Y$ then $X \cdot Y \in L^1$ and $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.*

Proof. It suffices to prove it when $X, Y \geq 0$ since then

a. If $X \perp\!\!\!\perp Y$ then $|X| \perp\!\!\!\perp |Y|$ so $\mathbb{E}[|XY|] < \infty$ and

$$\mathbb{E}[XY] = \mathbb{E}\big[(X^+ - X^-)(Y^+ - Y^-)\big] \tag{1}$$
$$= \mathbb{E}\big[X^+Y^+\big] - \mathbb{E}\big[X^+Y^-\big] - \mathbb{E}\big[X^-Y^+\big] + \mathbb{E}\big[X^-Y^-\big] \tag{2}$$
$$= \mathbb{E}\big[X^+\big]\,\mathbb{E}\big[Y^+\big] - \mathbb{E}\big[X^+\big]\,\mathbb{E}\big[Y^-\big] - \mathbb{E}\big[X^-\big]\,\mathbb{E}\big[Y^+\big] + \mathbb{E}\big[X^-\big]\,\mathbb{E}\big[Y^-\big] \tag{3}$$
$$= (\mathbb{E}\big[X^+\big] - \mathbb{E}\big[X^-\big])(\mathbb{E}\big[Y^+\big] - \mathbb{E}\big[Y^-\big]) \tag{4}$$
$$= \mathbb{E}[X]\,\mathbb{E}[Y] \tag{5}$$

where all the expectations in 3 are finite since X^+, X^-, Y^+, Y^- are all random variables. Why is this true? $X^+ < |X|$,$X^- < |X|$,$Y^+ < |Y|$, and $Y^- < |Y|$.

b. Simple to Non-negative:
We let

$$X_k = \alpha^{(k)}(X)$$
$$Y_k = \alpha^{(k)}(Y)$$

so $X_k \perp\!\!\!\perp Y_k$ $\forall k$ and $X_k \uparrow X$, $Y_k \uparrow Y$, $X_kY_k \uparrow XY$, so

$$\begin{aligned}\mathbb{E}[XY] &= \lim_{k\to\infty} \mathbb{E}[X_kY_k]\\ &= \lim_{k\to\infty} \mathbb{E}[X_k]\,\mathbb{E}[Y_k]\\ &\overset{MCT}{=} \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

c. Indicator to Simple:
This just done by linearity. It is done as in part a)

d. Indicators:
If $X = \mathbb{1}_A$ and $Y = \mathbb{1}_B$, then $X \perp\!\!\!\perp Y$. Then

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] \\ &= \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A)\,\mathbb{P}(B) \\ &= \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

□

The trick used in this proof is very important/useful! Should remember this for the exam!

Prop. *If $X, Y \in L^2$, then $XY \in L^1$.*

Proof. As before it suffices to prove when $X, Y \geq 0$. In this case,

$$\begin{aligned}\mathbb{E}[|XY|] &= \mathbb{E}[XY] \\ &\leq \mathbb{E}\big[\max(X,Y)^2\big] \\ &\leq \mathbb{E}\big[X^2 + Y^2\big] \\ &< \infty \qquad \text{by assumption}\end{aligned}$$

□

Theorem 19 (Hölder's Inequality)**.** *$p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$.*
If $\|X\|_p < \infty$, $\|Y\|_q < \infty$, then $\|XY\| < \infty$ and

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p}\,\mathbb{E}[|Y|^q]^{1/q}$$

Lemma (Young's inequality)**.** *If $a, b > 0$ then*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Proof. This says that

$$\log(ab) \leq \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right)$$

DIAGRAMS

By concavity,

$$\begin{aligned}\log\left(\frac{a^p}{p} + \frac{b^q}{q}\right) &\geq \frac{1}{p}\log(a^p) + \frac{1}{q}\log(b^q) \\ &= \log a + \log b \\ &= \log ab\end{aligned}$$

□

Proof of Hölder's Inequality.
Claim: We can assume that X and Y are non-negative.
Claim: We can assume $\mathbb{E}[|X|^p] = 1$ and $\mathbb{E}[|X|^q] = 1$.
To prove the second claim, let

$$X' = \frac{X}{\|X\|_p} \qquad Y' = \frac{Y}{\|Y\|_q}$$

then $\|X'\|_p = 1 = \|Y'\|_q$, so

$$\frac{\mathbb{E}[|XY|]}{\|X\|_p \|Y\|_q} = \mathbb{E}[|X'Y'|] \leq \|X'\|_p \cdot \|Y'\|_q = 1$$

Finally, if $\mathbb{E}[|X|^p] = 1 = \mathbb{E}[|X|^q]$, then

$$XY \leq \frac{X^p}{p} + \frac{Y^q}{q}$$

and taking expectations, we get

$$\mathbb{E}[XY] \leq \frac{\mathbb{E}[X^p]}{p} + \frac{\mathbb{E}[Y^q]}{q}$$

$$
\begin{aligned}
&= \frac{1}{p} + \frac{1}{q} \\
&= 1 \\
&= \mathbb{E}[X^p]^{1/p}\,\mathbb{E}[Y^q]^{1/q}
\end{aligned}
$$

$\square$

Corollary (Cauchy-Schwarz)**.** *If $X, Y \in L^2$ then*

$$\mathbb{E}[|XY|] \le (\mathbb{E}\big[X^2\big])^{1/2}(\mathbb{E}\big[Y^2\big])^{1/2} < \infty$$

Definition 22. *If $X, Y \in L^2(\mathbb{P})$, then the covariance between X and Y is*

$$
\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}
$$

If $\operatorname{Cov}(X,Y) = 0$, we say X, Y are uncorrelated.
Exercise: Find an example where X, Y are uncorrelated but not independent.

15.2. **Computations with Random Variables.** Let N be a Normal$(0,1)$ variable, so

$$
\begin{aligned}
\mathbb{P}(N \le x) &= F_N(x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t \\
&=: \frac{1}{\sqrt{2\pi}} \int_{(-\infty,x]} e^{-t^2/2}\,\mathrm{d}Leb
\end{aligned}
$$

Then N has density

$$f_N(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

with respect to Lebesgue measure.

From basic undergraduate probability,

$$\mathbb{E}\big[N^2\big] = \int x^2 f_N(x)\,\mathrm{d}x$$

Prop (1)**.** $(S, \mathcal{S}, \mu)$
Let $f : \mathbb{R} \to [0,\infty)$, $h : \mathbb{R} \to \mathbb{R}$ measurable. Let $\nu = \mu(f)$.

$$\nu(E) = \int_E f\,\mathrm{d}\mu$$

Then

$$h \in L^1(\nu) \Leftrightarrow h \cdot f \in L^1(\mu)$$

and in this case

$$\int h\,\mathrm{d}\nu = \int h \cdot f\,\mathrm{d}\mu$$

Prop (2)**.** *Let X be a random variable with law Λ. Let $h : \mathbb{R} \to \mathbb{R}$ be measurable. Then*

$$\int h(x)\,\mathrm{d}\mathbb{P} = \mathbb{E}[|h(X)|] < \infty \Leftrightarrow \int |h|\,\mathrm{d}\Lambda < \infty$$

and in this case

$$\mathbb{E}[h(X)] = \int h\,\mathrm{d}\Lambda$$

Corollary. *If X has density f, then for any $h : \mathbb{R} \to \mathbb{R}$ measurable,*

$$\mathbb{E}[|h(X)|] < \infty \Leftrightarrow \int_{\mathbb{R}} |h(x)f(x)|\,\mathrm{d}x < \infty$$

and in this case

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)\,\mathrm{d}x$$

16. November 8th, 2012 (Makeup)

16.1. **Chebyshev for Sums.**

Theorem 20 (Chebyshev for sums). *If $Z \in L^2(\mathbb{P})$, then $\mathbb{P}(|Z - \mathbb{E}[Z]\,| \geq t) \leq \frac{\mathrm{Var}(Z)}{t^2}$. Now suppose $Z = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are pairwise independent with $\mathbb{E}[X_i] = 0$, $\mathbb{E}\big[X_i^2\big] = 1$. Then Chebyshev tells us that*

$$\mathbb{P}(|Z| \geq t) \leq \frac{\mathrm{Var}(Z)}{t^2}$$

We have that

$$\begin{aligned}\mathrm{Var}(Z) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^n \mathbb{E}\big[X_i^2\big] \\ &= n\end{aligned}$$

So

$$\mathbb{P}(|Z| \geq t) \leq \frac{n}{t^2}$$

16.2. **Weak Law of Large Numbers for Finite Variance.**

Theorem 21 (WLLN for finite variance). *If $X_i, i \geq 1$ are pairwise independent with $\mathbb{E}[X_i] = \mu$ for all i and $\sup_{i \geq 1} \mathrm{Var}(X_i) = K < \infty$, then for all $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \to 0 \qquad \text{as } n \to \infty$$

where $S_n = X_1 + \ldots + X_n$.

Proof. By replacing X_i by $X_i - \mu$, we can assume $\mu = 0$. We use the same argument as Chebyshev for sums. This gives us

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) &= \mathbb{P}(|S_n| > \varepsilon n) \\ &\leq \sum_{i=1}^n \frac{\mathbb{E}\big[X_i^2\big]}{\varepsilon^2 n^2} \\ &\leq \frac{Kn}{\varepsilon^2 n^2} \\ &= \frac{(K/\varepsilon)^2}{n} \\ &\longrightarrow 0 \text{ as } n \to \infty\end{aligned}$$

$\square$

16.3. **Weierstrass approximation.**

Let $\mathcal{C}([0,1]) = \{\text{continuous functions } f : [0,1] \to \mathbb{R}\}$. For $f, g \in \mathcal{C}([0,1])$, we write

$$\begin{aligned}\|f - g\| &= \|f - g\|_\infty \\ &= \mathrm{esssup}(f - g) \\ &= \sup_{x \in [0,1]} |f(x) - g(x)|\end{aligned}$$

Prop. *For any $f \in \mathcal{C}([0,1])$ and $\varepsilon > 0$, $\exists$ a polynomial $p : [0,1] \to \mathbb{R}$ such that $\|f - p\| \leq \varepsilon$.*

Proof. For $0 \leq p \leq 1$, let $X_1, \ldots, X_n$ be Bernoulli(p) and let $S_n = \sum_{i=1}^n X_i = \text{Bin}(n,p)$. Now let $B_n : [0,1] \to \mathbb{R}$,

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}$$

In other words, $B_n(p) = \mathbb{E}\big[f\left(\frac{S_n}{n}\right)\big]$.
Intuition: $\mathbb{E}[S_n] = np$ and by the WLLN, we have that for any $\delta > 0$, for n large,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \delta\right)$$

is small.

To make this rigorous, note:

$$|B_n(p) - f(p)| = \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right) - f(p)\right]\right|$$

Let E_n be an event such that

$$E_n = \left\{\left|\frac{S_n}{n} - p\right| < \delta\right\}$$

where δ is chosen such that for $|x-y| < \delta$,

$$|f(x) - f(y)| < \frac{\varepsilon}{2}$$

so now we have

$$\begin{aligned}
|B_n(p) - f(p)| &= \left|\mathbb{E}\left[\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\mathbb{1}_{E_n}\right] + \mathbb{E}\left[\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\mathbb{1}_{E_n^c}\right]\right| \\
&\leq \mathbb{E}\left[\left|\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\right|\mathbb{1}_{E_n}\right] + \mathbb{E}\left[\left|\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\right|\mathbb{1}_{E_n^c}\right] \\
&\leq \frac{\varepsilon}{2}\mathbb{P}(E_n) + \mathbb{E}\left[\left|\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\right|\mathbb{1}_{E_n^c}\right]
\end{aligned}$$

We will use the fact that

$$\begin{aligned}
\left|f\left(\frac{S_n}{n}\right) - f(p)\right| &\leq \left|f\left(\frac{S_n}{n}\right)\right| + |f(p)| \\
&\leq 2\|f\|_\infty
\end{aligned}$$

and so

$$\begin{aligned}
|B_n(p) - f(p)| &\leq \frac{\varepsilon}{2} + 2\|f\|_\infty \mathbb{P}(E_n^c) \\
&= \frac{\varepsilon}{2} + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right)
\end{aligned}$$

For n large enough, $\mathbb{P}\big(\big|\frac{S_n}{n} - p\big| \geq \delta\big) < \frac{\varepsilon}{4\|f\|_\infty}$, so $|B_n(p) - f(p)| < \varepsilon$. Since p was arbitrary (as this was uniformly continuous), we are done. ∎

Theorem 22 (Weak Law of Large Numbers). *If $X_i, i \geq 1$ are pairwise independent and identically distributed with $\mathbb{E}[|X|] < \infty$ then $\forall \varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \to 0 \qquad \textit{as } n \to \infty$$

Proof. Fix $N > 0$ and let $X_i^{\leq N} = X_i \mathbb{1}_{|X_i| \leq N}$ and $X_i^{>N} = X_i \mathbb{1}_{|X_i| > N}$. Now let us define

$$S_n^{\leq N} = \sum_{i=1}^n X_i^{\leq N} \qquad S_n^{>N} = \sum_{i=1}^n X_i^{>N}$$

Write

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{S_n^{\leq N}}{n} - \mathbb{E}\left[\frac{S_n^{\leq N}}{n}\right]\right| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\left|\frac{S_n^{>N}}{n} - \mathbb{E}\left[\frac{S_n^{>N}}{n}\right]\right| \geq \frac{\varepsilon}{2}\right)$$

$$= \mathbb{P}\left(\left|\frac{S_n^{\le N}}{n} - \mathbb{E}\left[X_1^{\le N}\right]\right| \ge \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\left|\frac{S_n^{>N}}{n} - \mathbb{E}\left[X_1^{>N}\right]\right| \ge \frac{\varepsilon}{2}\right)$$

Fact: If X is a random variable and $a \le X \le b$, then

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &\le \mathbb{E}\left[(b-a)^2\right] \\ &= (b-a)^2 \end{aligned}$$

Exercise: In fact, $\mathrm{Var}(X) \le \dfrac{(b-a)^2}{4}$.

By the exercise, $\mathrm{Var}(X_1^{\le N}) \le N^2$, so

$$\begin{aligned} \mathbb{P}\left(\left|\frac{S_n^{\le N}}{n} - \mathbb{E}\left[X_1^{\le N}\right]\right| \ge \frac{\varepsilon}{2}\right) &= \mathbb{P}\left(|S_n^{\le N} - \mathbb{E}\left[S_n^{\le N}\right]| > \frac{\varepsilon n}{2}\right) \\ &\le \frac{N^2 n}{(\varepsilon n/2)^2} \\ &= \left(\frac{4N^2}{\varepsilon^2}\right)\frac{1}{n} \end{aligned}$$

and this tends to 0 as $n \to \infty$.
Note: $|X_1^{\le N}| \uparrow |X_1|$ as $N \to \infty$, so by Monotone Convergence Theorem,

$$\mathbb{E}\left[|X_1^{\le N}|\right] \to \mathbb{E}[|X_1|]$$

But $|X_1| = |X_1^{\le N}| + |X_1^{>N}|$.

$$\mathbb{E}\left[|X_1^{>N}|\right] \to 0 \qquad \text{as } N \to \infty$$

We take N large enough so that $\mathbb{E}\left[X_1^{>N}\right] < \frac{\varepsilon^2}{8}$. So this gives us that

$$\mathbb{P}\left(\left|\frac{S_n^{\le N}}{n} - \mathbb{E}\left[X_1^{\le N}\right]\right| \ge \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{2}$$

for $n > \frac{8N^2}{\varepsilon^3}$.

To complete the proof, we now show

$$\mathbb{P}\left(\left|\frac{S_n^{>N}}{n} - \mathbb{E}\left[X_1^{>N}\right]\right| \ge \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{2}$$

This expression can be written as

$$\begin{aligned} \mathbb{P}\left(|S_n^{>N} - \mathbb{E}\left[S_n^{>N}\right]| \ge \frac{\varepsilon n}{2}\right) &\le \frac{\mathbb{E}\left[|S_n^{>N} - \mathbb{E}\left[S_n^{>N}\right]\right]}{\varepsilon n/2} \\ &\le \frac{\mathbb{E}\left[|S_n^{>N}|\right] + |\mathbb{E}\left[S_n^{>N}\right]|}{\varepsilon n/2} \\ &\le \frac{(4/\varepsilon)}{n}\mathbb{E}\left[\underbrace{|S_n^{>N}|}_{\le |X_1^{>N}| + \dots + |X_N^{>N}|}\right] \\ &\le \frac{(4/\varepsilon)}{n} n\mathbb{E}\left[|X_1^{>N}|\right] \\ &< \left(\frac{4}{\varepsilon}\right)\left(\frac{\varepsilon^2}{8}\right) \\ &= \frac{\varepsilon}{2} \end{aligned}$$

We showed that for all $\varepsilon > 0$ and all n sufficiently large, then

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| > \frac{\varepsilon}{2}\right) < \varepsilon$$

□

17. November 8th, 2012

$X_i, i \geq 1$ i.i.d., $\mathbb{E}[X_i] < \infty$, $S_n = X_1 + \ldots + X_n$.
WLLN: $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) = 0$$

SLLN:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} \text{ exists, equals } \mathbb{E}[X_1]\right) = 1$$

which is equivalent to

$$\mathbb{P}\left(\forall\varepsilon > 0,\ \limsup_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1] + \varepsilon,\ \liminf_{n\to\infty} \frac{S_n}{n} > \mathbb{E}[X_1] - \varepsilon\right) = 1$$

Note that we have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{n} > \mathbb{E}[X_1] + \varepsilon\right) \geq \limsup_{n\to\infty} \mathbb{P}\left(\frac{S_n}{n} > \mathbb{E}[X_1] + \varepsilon\right)$$

and

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1] - \varepsilon\right) \leq \liminf_{n\to\infty} \mathbb{P}\left(\frac{S_n}{n} < \mathbb{E}[X_1] - \varepsilon\right)$$

which means that SLLN $\Rightarrow$ WLLN.

Steps:

(1) Assume $\mathbb{E}[X_1] = 0$

(2) Claim: It suffices to prove when $X_1 \geq 0$. Suppose we know the SLLN in the case $X_1 \geq 0$. For general X_1, write $X_i = X_i^+ - X_i^-$. Then

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n^+}{n} = \mathbb{E}\big[X_1^+\big]\right) = 1 = \mathbb{P}\left(\lim_{n\to\infty} \frac{S_n^-}{n} = \mathbb{E}\big[X_1^-\big]\right)$$

We have

$$\begin{aligned} S_n^+ &= X_1^+ + \ldots + X_n^+ \\ S_n^- &= X_1^- + \ldots + X_n^- \end{aligned}$$

If

$$\lim_{n\to\infty} \frac{S_n^+}{n} = \mathbb{E}\big[X_1^+\big] \qquad \text{and} \qquad \lim_{n\to\infty} \frac{S_n^-}{n} = \mathbb{E}\big[X_1^-\big]$$

then

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \left(\frac{S_n^+}{n} - \frac{S_n^-}{n}\right) = \mathbb{E}\big[X_1^+\big] - \mathbb{E}\big[X_1^-\big] = \mathbb{E}[X_1]$$

so

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X_1]\right) \geq \mathbb{P}\left(\lim_{n\to\infty} \frac{S_n^+}{n} = \mathbb{E}[X_1]\,,\ \lim_{n\to\infty} \frac{S_n^-}{n} = \mathbb{E}[X_1]\right) = 1$$

(3) Reduce the number of "times" n at which we must sample S_n.

Definition 23. *A sequence $n_i, i \geq 1$ is lacunary if $\exists c > 1$ such that for all i sufficiently large, $n_{i+1} \geq c n_i$.*

Theorem 23 (Lacunary SLLN). *If $X_i \geq 0$, $\mathbb{E}[X_i] < \infty$, then for any lacunary sequence $n_i, i \geq 1$,*

$$\mathbb{P}\left(\lim_{i\to\infty} \frac{S_{n_i}}{n_i} = \mathbb{E}[X_1]\right) = 1$$

Proof of "non-negative SLLN" from Lacunary SLLN. It suffices to prove that $\forall\varepsilon > 0$,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1] + \varepsilon\right) = 1 = \mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{n} > \mathbb{E}[X_1] - \varepsilon\right)$$

Since then we would have

$$\mathbb{P}\left(\bigcap_{k\geq 1} \left\{\limsup_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1] + \frac{1}{k}\right\} \cap \left\{\liminf_{n\to\infty} \frac{S_n}{n} > \mathbb{E}[X_1] - \frac{1}{k}\right\}\right) = 1$$

DIAGRAMS

Let us fix ε. Let $n_i = \left(1+\frac{\varepsilon}{3}\right)^i$. If

$$\limsup_{n\to\infty} \frac{S_n}{n} \geq \mathbb{E}[X_1]\,(1+\varepsilon)$$

then $\exists i_k, k \geq 1$ such that $\forall k \geq 1$, for some $n \in [n_{i_k}, n_{i_k+1}]$,

$$\frac{S_n}{n} > \mathbb{E}[X_1]\,(1+\varepsilon)$$

Then

$$\frac{S_{n_{i_k+1}}}{n_{i_k+1}} \geq \frac{S_n}{n_{i_k+1}} \geq \frac{S_n}{n\left(1+\frac{\varepsilon}{3}\right)} \geq \mathbb{E}[X_1]\,\frac{1+\varepsilon}{\left(1+\frac{\varepsilon}{3}\right)} > \mathbb{E}[X_1]\left(1+\frac{\varepsilon}{2}\right)$$

so

$$\limsup_{i\to\infty} \frac{S_{n_i}}{n_i} \geq \mathbb{E}[X_1]\left(1+\frac{\varepsilon}{2}\right)$$

It follows that

$$\mathbb{P}\left(\limsup_{i\to\infty} \frac{S_{n_i}}{n_i} > \mathbb{E}[X_1]\left(1+\frac{\varepsilon}{2}\right)\right) \geq \mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{n} > \mathbb{E}[X_1]\,(1+\varepsilon)\right)$$

The left-hand side is zero based on lacunary SLLN.

Similarly, if

$$\liminf_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1]\,(1-\varepsilon)$$

then let $i_k, k \geq 1$ be such that for all k,$\exists n \in [n_{i_k}, n_{i_k+1})$ for which $\frac{S_n}{n} < \mathbb{E}[X_1]\,(1-\varepsilon)$. Then

$$\frac{S_{n_{i_k}}}{n_{i_k}} \leq \frac{S_n}{n_{i_k}} \leq \frac{S_n}{n}\left(1+\frac{\varepsilon}{3}\right) < \mathbb{E}[X_1]\,(1-\varepsilon)\left(1+\frac{\varepsilon}{3}\right) < \mathbb{E}[X_1]\left(1-\frac{\varepsilon}{2}\right)$$

so

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{n} < \mathbb{E}[X_1]\,(1-\varepsilon)\right) = 0$$

$\square$

Now all that we have that's left to prove is the lacunary SLLN. Fix a lacunary sequence $(n_i, i \geq 1)$ such that $n_{i+1} > cn_i$ for i sufficiently large. It suffices to prove that $\forall \varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_{n_i}}{n_i} - \mathbb{E}[X_1]\right| > \varepsilon \text{ i.o.}\right) = 0$$

We will show that

$$\sum_{i\geq 1} \mathbb{P}\left(\left|\frac{S_{n_i}}{n_i} - \mathbb{E}[X_1]\right| > \varepsilon\right) < \infty$$

Then the result follows by Borel-Cantelli. Since we are looking at summability of the series, we may start the sum at i_0, where i_0 is chosen such that $n_{i+1} \geq cn_i$ for $i \geq i_0$. Write

$$S_{n_i}^{\leq n_i} = \sum_{j=1}^{n_i} X_j^{\leq n_i} = \sum_{j=1}^{n_i} X_j \mathbb{1}_{|X_j| \leq n_i}$$

and

$$S_n^{>n_i} = S_{n_i} - S_{n_i}^{\leq n_i}$$

Then we have

$$\mathbb{P}\left(\left|\frac{S_{n_i}}{n_i}\right| > 2\varepsilon\right) \leq \mathbb{P}\left(\left|\frac{S_{n_i}^{\leq n_i}}{n_i} - \mathbb{E}\left[X_1^{\leq n_i}\right]\right| > \varepsilon\right) + \mathbb{P}\left(S_{n_i} \neq S_{n_i}^{\leq n_i}\right)$$

for i large enough that $|\mathbb{E}\left[X_1^{\leq n_i}\right] - \mathbb{E}[X_1]\,| < \varepsilon$. This is because if

$$\left|\frac{S_{n_i}^{\leq n_i}}{n_i} - \mathbb{E}\left[X_1^{\leq n_i}\right]\right| \leq \varepsilon$$

and $S_{n_i} = S_{n_i}^{\leq n_i}$ then

$$\left|\frac{S_{n_i}}{n_i} - \mathbb{E}\left[X_1^{\leq n_i}\right]\right| \leq \varepsilon$$

so

$$\left|\frac{S_{n_i}}{n_i} - \mathbb{E}[X_1]\right| \leq 2\varepsilon$$

First:

$$\mathbb{P}\left(\left|\frac{S_{n_i}^{\le n_i}}{n_i} - \mathbb{E}\left[X_1^{\le n_i}\right]\right| > \varepsilon\right) = \mathbb{P}\left(|S_{n_i}^{\le n_i} - n_i\mathbb{E}\left[X_1^{\le n_i}\right]| > \varepsilon n_i\right)$$
$$\le \frac{\mathrm{Var}(S_{n_i}^{\le n_i})}{(\varepsilon n_i)^2}$$
$$\le \frac{n_i\mathbb{E}\left[(X_1^{\le n_i})^2\right]}{(\varepsilon n_i)^2}$$
$$= \frac{\mathbb{E}\left[(X_1^{\le n_i})^2\right]}{\varepsilon^2 n_i}$$

Now this bound does not look very good but it'll end up working well! Now let us bound the second term.
Second:

$$\mathbb{P}\big(S_{n_i} \ne S_{n_i}^{\le n_i}\big) \le n_i\mathbb{P}(X_1 \ge n_i)$$

In the end what we want to do is sum these bounds so show that the sum of our probabilities is finite. Let $J = \min\{i : n_i \ge X_1\}$.

Summing the first bound, we get

$$\sum_{i=i_0}^{\infty}\mathbb{P}\left(\left|\frac{S_{n_i}^{\le n_i}}{n_i} - \mathbb{E}\left[X_1^{\le n_i}\right]\right| > \varepsilon\right) = \mathbb{E}\left[\sum_{i=i_0}^{\infty}\frac{(X_1^{\le n_i})^2}{n_i}\right]$$
$$= \mathbb{E}\left[\sum_{i=\max(i_0,J)}\frac{X_1^2}{n_i}\right]$$

But $X_1 \le n_J$, so $X_1^2 \le n_J X_1$, and so

$$\sum_{i=i_0}^{\infty}\mathbb{P}\left(\left|\frac{S_{n_i}^{\le n_i}}{n_i} - \mathbb{E}\left[X_1^{\le n_i}\right]\right| > \varepsilon\right) \le \mathbb{E}\left[\sum_{i=\max(i_0,J)}\underbrace{\frac{n_J}{n_i}}_{\le \frac{1}{1-c}} X_1\right]$$
$$\le \mathbb{E}[X_1]\,\frac{1}{1-c}$$
$$< \infty$$

18. November 13th, 2012

Recall: Lacunary SLLN: $S_n = X_1 + \ldots + X_n$, X_i iid $\in L^1(\mathbb{P})$, $n_i, i \ge 1$ such that $n_{i+1} \ge cn_i$ for all $i \ge i_0$, some $c > 1$, then

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{S_{n_i}}{n_i} = \mathbb{E}[X_i]\right) = 1$$

We had reduced the proof to showing two sums were finite.

$$\sum_{i=i_0}^{\infty}\frac{\mathbb{E}\left[(X_i^{\le n_i})^\alpha\right]}{n_i}$$

and

$$\sum_{i=i_0}^{\infty} n_i\mathbb{P}(X_i > n_i)$$

We had that $J = \min\{i : X_i \ge X_1\}$ We have since the sequence is iid, the second sum is

$$\sum_{i=i_0}^{\infty} n_i\mathbb{P}(X_1 \ge n_i) = \sum_{i=i_0}^{\infty} n_i\mathbb{E}\left[\mathbb{1}_{\{X_1 > n_i\}}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{i=i_0}^{\infty} n_i \mathbb{1}_{\{X_1 > n_i\}}\right] \\
&= \mathbb{E}\left[\sum_{i=i_0}^{J-1} n_i\right]
\end{aligned}
$$

We have

$$
n_{J-1-k} \leq \frac{n_{J-1}}{c^k} \leq \frac{X_1}{c^k}
$$

and so

$$
\begin{aligned}
\sum_{i=i_0}^{\infty} n_i \mathbb{P}(X_1 \geq n_i) &\leq \mathbb{E}\left[X_1 \sum_{k=0}^{J-1-i_0} \frac{1}{c^k}\right] \\
&= \frac{1}{c-1}\mathbb{E}[X_1] \\
&< \infty
\end{aligned}
$$

18.1. **Product Spaces.** Let X, Y be independent $\mathcal{U}[0,1]$ random variables.

- What is $\mathbb{E}[XY]$?
  By independence, this is $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{1}{4}$
- What is $\mathbb{E}\left[e^{X+Y}\right]$?
  We can rewrite this as $\mathbb{E}\left[e^X e^Y\right] = \mathbb{E}\left[e^X\right]\mathbb{E}\left[e^Y\right] = (\mathbb{E}\left[e^X\right])^2$. We integrate this to give us

$$
\left(\int_0^1 e^x \,\mathrm{d}x\right)^2 = (e-1)^2
$$

- What is $\mathbb{E}\left[e^{XY}\right]$?
  We will do this in two steps, holding X a constant first, then Y as a contant.

$$
\begin{aligned}
\mathbb{E}\left[e^{XY}\right] &= \mathbb{E}\left[\int_0^1 e^{Xy} \,\mathrm{d}y\right] \\
&= \int_0^1 \int_0^1 e^{xy} \,\mathrm{d}y \,\mathrm{d}x \\
&= \int_0^1 \frac{e^x - 1}{x} \,\mathrm{d}x \\
&= \int_0^1 \sum_{i \geq 1} \frac{x^{i-1}}{i!} \,\mathrm{d}x \\
&= \sum_{i \geq 1} \int_0^1 \frac{x^{i-1}}{i!} \,\mathrm{d}x \\
&= \sum_{i \geq 1} \frac{1}{i \cdot i!}
\end{aligned}
$$

Questions:

- What is a multiple integral?
- What is a joint density function?
- Writing $Z = (X, Y)$, do we have

$$
\mathbb{E}\left[e^{XY}\right] = \int_{[0,1]^2} f \,\mathrm{d}Leb([0,1]^2)
$$

  where $f : [0,1]^2 \to \mathbb{R}$ and $f(x,y) = e^{xy}$.

18.2. **Product σ-algebras.** Given measurable spaces $(\Omega_i, \mathcal{F}_i)$, $i \in I$, form the product

$$
\begin{aligned}
\Omega_I &= \prod_{i \in I} \Omega_i \\
&= \{(\omega_i, i \in I) : \omega_i \in \Omega_i \ \forall i \in I\}
\end{aligned}
$$

Example 19.
$\Omega_i = \{0,1\}$, $I = \mathbb{N}$
$\mathcal{F} = \sigma\{\mathbb{1}_{\{\omega_i=1\}, i\in\mathbb{N}}\}$

For $i \in I$, let

$$\begin{aligned} \rho_i &: \Omega_I \to \Omega_i \\ &(\omega_j, j \in I) \to \omega_i \end{aligned}$$

Let

$$\mathcal{F}_I = \sigma(\rho_i, i \in I)$$

To get a a feel for this, what is $\sigma(\rho_i)$? For $E_i \in \mathcal{F}_i$, then

$$\rho_i^{-1}(E_i) = \{\omega : \omega_i \in E_i\} = E_i \times \prod_{j\neq i} \Omega_j$$

so $\sigma(\rho_i) = \left\{E_i \times \prod_{j\neq i} \Omega_j\right\}$ and $\mathcal{F}_I$ contains this. If $I = \{1,2\}$, then

$$\begin{aligned} \mathcal{F}_I &= \sigma(\{A \times \Omega_2 : A \in \mathcal{F}_1\} \cup \{\Omega_1 \times B : B \in \Omega_2\}) \\ &= \sigma(\underbrace{\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}}_{\text{Rectangles}}) \end{aligned}$$

These are called rectangles and they form a π-system. In general, if I is countable, then

$$\mathcal{F}_I = \sigma\left(\left\{\prod_{i\in I} A_i : A_i \in \mathcal{F}_i\right\}\right)$$

If I is uncountable, then

$$\mathcal{F}_I = \sigma\left(\left\{\prod_{i\in S} A_i \times \prod_{i\in I\setminus S} \Omega_i : A_i \in \mathcal{F}_i, S \subset I, S \text{ countable}\right\}\right)$$

Alternatively, for $S \subset I$, let

$$\mathcal{F}_I^S = \left\{E \times \prod_{i\in I\setminus S} \Omega_i, E \in \mathcal{F}_S\right\}$$

Example 20. *If $I = \{1,2,3\}$ and $S = \{1,2\}$ with $(\Omega_i, \mathcal{F}_i) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, then*

$$\mathcal{F}_S = (\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$$

so in particular, we could take

$$E = \{x : |x| < 1\}$$

For any S countable, $\mathcal{F}_I^S$ is a σ-algebra and

$$\mathcal{F}_I = \bigcup_{\substack{S\subset I \\ S \text{ countable}}} \mathcal{F}_I^S$$

We have that

$$\mathcal{F}_I \supset \bigcup_{\substack{S\subset I \\ S \text{ countable}}} \mathcal{F}_I^S$$

is easy by the definition, and

$$\mathcal{F}_I \subset \bigcup_{\substack{S\subset I \\ S \text{ countable}}} \mathcal{F}_I^S$$

is true if we can show the right-hand side is a σ-algebra.

In summary, $\Omega_I = \prod_{i\in I} \Omega_i$.

- If I is countable, then $\mathcal{F}_I$ is generated by rectangles
- If I is uncountable, then $\mathcal{F}_I$ is generated by "countable cylinders", and in particular only ever restricts countably many coordinates.

Let us consider $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ and form the product space $\Omega = \Omega_1 \times \Omega_2$. It is common convention to write $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, but this is an abuse of notation since it is really the σ-algebra generated by the product. However, we refer to this as the product σ-algebra.

For a measurable function, $f : \Omega \to \mathbb{R}$, and any $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, let $f_{(\omega_1, \bullet)} : \Omega_2 \to \mathbb{R}$ which sends $\omega_2 \to f(\omega_1, \omega_2)$ and let $f_{(\bullet, \omega_2)} : \Omega_1 \to \mathbb{R}$ which sends $\omega_1 \to f(\omega_1, \omega_2)$.

Prop. *For all $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, $f_{(\omega_1, \bullet)}$ is $\mathcal{F}_2$-measurable and $f_{(\bullet, \omega_2)}$ is $\mathcal{F}_1$-measurable.*

Proof. Write $f = f^+ - f^-$. Then $(f^+)_{(\omega_1, \bullet)} = (f_{(\omega_1, \bullet)})^+$ and $(f^-)_{(\omega_1, \bullet)} = (f_{(\omega_1, \bullet)})^-$. So it suffices to prove the proposition for non-negative functions. However going down to simple functions and then indicators is not as trivial as we would like it to be for the standard machine.

Instead, for $f \geq 0$, we have that $f = \lim_{n \to \infty} f \wedge n$ (min) and

$$f_{(\omega_1, \bullet)} = \lim_{n \to \infty} (f_{(\omega_1, \bullet)}) \wedge n = \lim_{n \to \infty} (f \wedge n)_{(\omega_1, \bullet)}$$

so it suffices to prove the proposition for bounded random variables. To prove this for bounded functions, we use the Monotone Class Theorem

Theorem 24 (Monotone Class Theorem)**.** *Suppose we have a measurable space $(\Omega, \mathcal{F})$, and $\mathcal{P} \subset \mathcal{F}$ a π-system, $\sigma(\mathcal{P}) = \mathcal{F}$. Let H be a collection of measurable functions $f : \Omega \to \mathbb{R}$. If*

(1) *$\mathbb{1}_E \in H$ for $E \in \mathcal{P}$.*
(2) *If $f, g \in H$, then $cf + g \in H$.*
(3) *If $f_n \in H$, $f_n \uparrow f$, $f_n \geq 0$ (and f bounded), then $f \in H$.*

then H contains all non-negative (bounded) measurable functions $f : \Omega \to \mathbb{R}$.

Assuming the Monotone Class Theorem, take

$$H = \{f : \Omega \to \mathbb{R} : \forall \omega_1 \in \Omega_1, \omega_2 \in \Omega_2, f_{(\omega_1, \bullet)} \text{ and } f_{(\bullet, \omega_2)} \text{ are measurable}\}$$

where

$$\mathcal{P} = \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$$

that is $\mathcal{P}$ is the set of rectangles.

(1) If $f = \mathbb{1}_{A \times B}$, then

$$f_{(\omega_1, \bullet)} = \begin{cases} \mathbb{1}_B & \text{if } \omega_1 \in A \\ 0 & \text{if } \omega_1 \notin A \end{cases}$$

so $f \in H$. The same thing holds for $f_{(\bullet, \omega_2)}$.
(2) If $f, g \in H$, then $(cf + g)_{(\omega_1, \bullet)} = cf_{(\omega_1, \bullet)} + g_{(\omega_1, \bullet)}$ is $\mathcal{F}_2$-measurable and similarly, $(cf + g)_{(\bullet, \omega_2)}$ is $\mathcal{F}_1$-measurable, so $cf + g \in H$.
(3) Suppose $f_n \uparrow f$, $f_n \in H$, then $(f_n)_{(\omega_1, \bullet)} \uparrow f_{(\omega_1, \bullet)}$ so $f_{(\omega_1, \bullet)}$ is measurable. Likewise, $f_{(\bullet, \omega_2)}$ is measurable, so $f \in H$.

The monotone class theorem completes the proposition. ∎

19. November 15th, 2012 (Makeup)

We defined a product σ-algebra, $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ where

$$\mathcal{F}_1 \times \mathcal{F}_2 = \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$$

Prop. *If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1 \times \mathcal{F}_2$ then $f_{(\omega_1, \bullet)}, f_{(\bullet, \omega_2)}$ are measurable for any $\omega_1 \in \Omega_2, \omega_2 \in \Omega_2$.*

Theorem 25 (Monotone Class Theorem)**.** *$(\Omega, \mathcal{F})$, $\mathcal{P} \subset \mathcal{F}$ a π-system with $\sigma(\mathcal{P}) = \mathcal{F}$. If H is a class of functions from $\Omega \to \mathbb{R}$ such that*

- *$\mathbb{1}_E \in H$ for all $E \in \mathcal{P}$*
- *If $f, g \in H$ then $cf + g \in H$. ($c \in \mathbb{R}$).*
- *If $0 \leq f_n \in H$, $f_n \uparrow f$ (f bounded), then $f \in H$*

Then we have that H contains all (bounded, non-negative) measurable functions from $\Omega \to \mathbb{R}$.

Proof.

(1) **Step 1:**
Verify that $\mathbb{1}_E \in H$, $\forall E \in \mathcal{F}$.
Claim: $\mathcal{D} = \{E \in \mathcal{F} : \mathbb{1}_E \in H\}$ is a d-system (difference set) Assuming the claim, since $\mathcal{P} \subset \{E \in \mathcal{F}, \mathbb{1}_E \in H\}$, we get $\sigma(\mathcal{P}) \subset \{E \in \mathcal{F}, \mathbb{1}_E \in H\}$ and step 1 would be finished.

Proof of claim. If $A, B \in \mathcal{D}$, $A \subset B$, then

$$\mathbb{1}_{B \setminus A} = \mathbb{1}_B - \mathbb{1}_A \in H$$

by (2) with $c = -1$ so $B \setminus A \in \mathcal{D}$. If $A_n \in \mathcal{D}, n \in \mathbb{N}, A_n \uparrow A$, then $\mathbb{1}_{A_n} \uparrow \mathbb{1}_A$, so $\mathbb{1}_A \in H$ by (3) so $A \in \mathcal{D}$. Thus $\mathcal{D}$ is a d-system. □

(2) **Step 2:**
Claim: H contains all simple functions.

Proof. This is immediate from Step 1 and (2) □

(3) **Step 3:**
Claim: If $f \geq 0$, f measurable then $f \in H$

Proof. For $k \in \mathbb{N}$, let $f_k = \alpha^{(k)}(f)$. Then f_k are simple, $f_k \uparrow f$ so $f \in H$ by step 2 and (3). □

(4) **Step 4:**
For general f, write $f = f^+ - f^-$, and apply step 3 and (2)

□

Prop (2). *If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is measurable and $f \geq 0$, and μ_1, μ_2 are σ-finite measures on $(\Omega_1, \mathcal{F}_1)$,$(\Omega_2, \mathcal{F}_2)$ then*

$$I_f^1 = \int f_{(\omega_1, \bullet)} \, \mathrm{d}\mu_2$$

is $\mathcal{F}_1$-measurable and

$$I_f^2 = \int f_{(\bullet, \omega_2)} \, \mathrm{d}\mu_1$$

is $\mathcal{F}_2$-measurable, and

$$\int I_f^1 \, \mathrm{d}\mu_1 = \int I_f^2 \, \mathrm{d}\mu_2$$

Proof. Let H be the class of functions $f : \Omega_1 \times \Omega_2 \to [0, \infty)$ satisfying the conditions of proposition 2.

- **Step 0:**
Claim: $\forall A \in \mathcal{F}_1, B \in \mathcal{F}_2$ we have $f = \mathbb{1}_{A \times B} \in H$.

Proof.

$$I_f^1(\omega) = \int f_{(\omega_1, \bullet)} \, \mathrm{d}\mu_2 = \mu_2(B) \mathbb{1}_A$$

since

$$f_{(\omega_1, \bullet)} = \begin{cases} \mathbb{1}_B & \text{if } \omega_1 \in A \\ 0 & \text{otherwise} \end{cases}$$

DIAGRAMS

so I_f^1 is measurable. So we have

$$\int I_f^1 \, \mathrm{d}\mu_1 = \mu_2(B) \int \mathbb{1}_A \, \mathrm{d}\mu_1 = \mu_2(B) \mu_1(A)$$

Similarly, I_f^2 is measurable and $\int I_f^2 \, \mathrm{d}\mu_2 = \mu_1(A)\mu_2(B)$. So $f = \mathbb{1}_{A \times B} \in H$. □

- **Step 1:**
Claim: if $f \in H$, $g \in H$, then $cf + g \in H$.

Proof. If $f, g \in H$, then

$$I^1_{cf} = cI^1_f$$

is measurable. So

$$I^1_{cf+g} = cI^1_f + I^1_g$$

is measurable. Likewise, I^2_{cf+g} is measurable and

$$\begin{aligned}\int I^1_{cf+g}\,\mathrm{d}\mu_1 &= c\int I^1_f\,\mathrm{d}\mu_1 + \int I^1_g\,\mathrm{d}\mu_1 \\ &= c\int I^2_f\,\mathrm{d}\mu_2 + \int I^2_g\,\mathrm{d}\mu_2 \qquad \text{since } f, g \in H \\ &= \int I^2_{cf+g}\,\mathrm{d}\mu_2\end{aligned}$$

so $cf + g \in H$. □

- **Step 2:**
  Claim: If $f_n \in H$, $f_n \geq 0$, $f_n \uparrow f$, then $f \in H$.

  Proof. Monotone convergence theorem implies that I^1_f, I^2_f are measurable and then

$$\begin{aligned}\int I^1_f\,\mathrm{d}\mu_1 &= \lim_{n\to\infty}\int I^1_{f_n}\,\mathrm{d}\mu_1 \\ &= \lim_{n\to\infty}\int I^2_{f_n}\,\mathrm{d}\mu_2 \\ &= \int I^2_f\,\mathrm{d}\mu_2\end{aligned}$$

  so $f \in H$. □

By the monotone class theorem, if follows that H contains all non-negative measurable functions from $\Omega_1 \times \Omega_2 \to \mathbb{R}$. □

Example 21 (Warning examples)**.** $(\Omega_1, \mathcal{F}_1, \mu_1) = (\Omega_2, \mathcal{F}_2, \mu_2) = (\mathbb{N}, 2^{\mathbb{N}}, \textit{Counting measure})$

$$\Omega_1 \times \Omega_2 = \mathbb{N}^2$$

$$\begin{array}{c|ccccc} & 0 & 0 & 0 & 0 & \cdots \\ & 0 & 0 & 0 & 0 & \cdots \\ & 0 & 0 & 0 & -1 & \cdots \\ m & 0 & 0 & -1 & 1 & \cdots \\ & 0 & -1 & 1 & 0 & \cdots \\ & -1 & 1 & 0 & 0 & \cdots \\ & 1 & 0 & 0 & 0 & \cdots \\ \hline & & & n & & \end{array} \tag{6}$$

$$f(m,n) = \begin{cases} 1 & \textit{if } m = n \\ -1 & \textit{if } m = n+1 \\ 0 & \textit{otherwise} \end{cases}$$

We have

$$\int I^1_f = 1 + 0 + 0 + \cdots = 1$$

and

$$\int I^2_f = 0 + 0 + 0 + \cdots = 0$$

If we define

$$\mathbb{E}\big[e^{XY}\big] = \iint e^{xy}\,\mathrm{d}x\,\mathrm{d}y$$

Definition 24. *Given $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ σ-finite measure spaces, for $E \in \mathcal{F}_1 \times \mathcal{F}_2$, set*

$$\mu(E) = (\mu_1 \times \mu_2)(E) = \iint \mathbb{1}_E \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2$$

Prop. *μ is a measure*

Proof.

- $\mu(\emptyset) = 0$ is trivial
- $\mu \geq 0$ is also trivial as we are integrating a non-negative function (the indicator function)
- If $E_n, n \geq 1$ are disjoint, $E = \bigcup_n E_n$, then

$$\begin{aligned}
\mu(E) &= \iint \mathbb{1}_E \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 \\
&= \iint \sum_{i \geq 1} \mathbb{1}_{E_i} \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 \\
&= \int \left(\sum_{i \geq 1} \int \mathbb{1}_{E_i} \, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 \\
&= \sum_{i \geq 1} \underbrace{\iint \mathbb{1}_{E_i} \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2}_{\mu(E_i)} \\
&= \sum_{i \geq 1} \mu(E_i)
\end{aligned}$$

□

Claim. *μ is σ-finite and so by the uniqueness lemma it is the unique σ-finite extension of its restriction to*

$$\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$$

Proof. There exist $A_n \in \mathcal{F}_1$, $A_n \uparrow \Omega$, $B_n \in \mathcal{F}_2$, $B_n \uparrow \Omega_2$. with $\mu_1(A_n) < \infty$ and $\mu_2(B_n) < \infty$. Then $A_n \times B_n \uparrow \Omega_1 \times \Omega_2$ and $\mu(A_n \times B_n) < \infty$ □

Example 22 (Example from last class)**.** *$X \sim \mathcal{U}[0,1], Y \sim \mathcal{U}[0,1]$. Let μ_1 be the law of X =*Leb$([0,1])$ *and μ_2 be the law of Y =*Leb$([0,1])$*.*

Let $\mu = \mu_1 \times \mu_2$. Then for any $A, B \in \mathfrak{B}([0,1])$,

$$\begin{aligned}
\mu(A \times B) &= \mu_1(A) \cdot \mu_2(B) \\
&= \mathrm{Leb}(A) \cdot \mathrm{Leb}(B) \\
&= \mathrm{Leb}(A \times B)
\end{aligned}$$

so

$$\iint e^{xy} \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 = \underbrace{\int e^{xy} \, \mathrm{d}\mu}_{\mathbb{E}[e^{XY}]} = \iint e^{xy} \, \mathrm{d}\mu_2 \, \mathrm{d}\mu_1$$

20. November 15th, 2012

Prop. *$(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ σ-finite, $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ measurable, $f \geq 0$, then $\int f_{(\omega_1, \bullet)} \, \mathrm{d}\mu_2$ and $\int f_{(\bullet, \omega_2)} \, \mathrm{d}\mu_1$ are measurable and*

$$\iint f \, \mathrm{d}\mu_1 \, \mathrm{d}\mu_2 = \iint f \, \mathrm{d}\mu_2 \, \mathrm{d}\mu_1$$

The proof of this from this morning works if $\mu_1(\Omega_1) < \infty$, $\mu_2(\Omega_2) < \infty$, and f is bounded.

If $f \geq 0$, spaces are σ-finite, then let $A_n \in \mathcal{F}$, $\mu_1(A_n) < \infty$, $A_n \uparrow \Omega_1$, $B_n \in \mathcal{F}_2$, $\mu_2(B_n) < \infty$, $B_n \uparrow \Omega_2$. Let $f_n = f \wedge n$. Then μ_{1,A_n} and μ_{2,B_n}, which are measures that restrict the σ-algebra to A_n and B_n respectively are finite measures. Then

$$f = \lim_{n \to \infty} f_n = \lim_{n \to \infty} \lim_{m \to \infty} f_n \mathbb{1}_{A_m \times B_m}$$

so

$$\iint f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 = \lim_{n\to\infty}\lim_{m\to\infty}\iint f_n \underbrace{\mathbb{1}_{A_m\times B_m}}_{\mathbb{1}_{A_m}\cdot\mathbb{1}_{B_m}}\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2$$

by four applications of monotone convergence theorem. This becomes

$$\begin{aligned}\iint f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 &= \lim_{n\to\infty}\lim_{m\to\infty}\int \mathbb{1}_{B_m}\int f_n\mathbb{1}_{A_m}\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2\\ &= \lim_{n\to\infty}\lim_{m\to\infty}\int \mathbb{1}_{B_m}\int f_n\,\mathrm{d}\mu_{1,A_m}\,\mathrm{d}\mu_2\\ &= \lim_{n\to\infty}\lim_{m\to\infty}\iint f_n\,\mathrm{d}\mu_{1,A_m}\,\mathrm{d}\mu_{2,B_m}\end{aligned}$$

Now these are bounded, so we can apply the result we used earlier, to give us

$$\begin{aligned}\iint f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 &= \iint f_n\,\mathrm{d}\mu_{2,B_m}\,\mathrm{d}\mu_{1,A_m}\\ &\vdots\\ &= \iint f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1\end{aligned}$$

where we reverse what we had just shown.

Example 23 (Warning example)**.** *Take* $(\Omega_1,\mathcal{F}_1)=(\Omega_2,\mathcal{F}_2)=([0,1],\mathfrak{B}([0,1]))$*. Let* μ_1 *be the Lebesgue measure and* μ_2 *the counting measure. Let*

$$D=\{(x,x):0\le x\le 1\}$$

and $f=\mathbb{1}_D$*. On one hand, we have*

$$\iint f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 = \int 0\,\mathrm{d}\mu_2 = 0$$

On the other, we have

$$\iint f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int 1\,\mathrm{d}\mu_1 = 1$$

We defined $\mu=\mu_1\times\mu_2$ by setting

$$\mu(E)=\iint \mathbb{1}_E\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 = \iint \mathbb{1}_E\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1$$

In particular, $\mu(A\times B)=\mu_1(A)\mu_2(B)$.

Theorem 26 (Tonelli's Theorem)**.** *If* $f\ge 0$*,* $f:\Omega_1\times\Omega_2\to\mathbb{R}$ *is measurable then*

$$\iint f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int f\,\mathrm{d}\mu = \iint f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2$$

Given random variables $X:\Omega\to\mathbb{R}$ and $Y:\Omega\to\mathbb{R}$, the joint law of X and Y is the measure Λ on $(\mathbb{R}^2,\mathfrak{B}(sR^2))$ with $\Lambda(E)=\mathbb{P}((X,Y)\in E)$.

Prop. *If* X,Y *are independent, then*

$$\Lambda=\Lambda_X\times\Lambda_Y$$

where Λ_X *is the law of* X *and* Λ_Y *is the law of* Y*.*

Proof. By independence, for any rectangle $A\times B$,

$$\Lambda(A\times B)=\Lambda_X(A)\times\Lambda_Y(B)$$

and the result follows from the uniqueness lemma because rectangles generate the product σ-algebra. □

Now, we have

$$
\begin{aligned}
\mathbb{E}\big[e^{XY}\big] &= \int \underbrace{f}_{g(X,Y)} \mathrm{d}\mathbb{P} \\
&= \int_{\mathbb{R}^2} g \overbrace{\mathrm{d}\Lambda}^{\text{Joint law of }(X,Y)} \\
&= \iint g \,\mathrm{d}\Lambda_X \,\mathrm{d}\Lambda_Y \\
&= \iint e^{xy} \,\mathrm{d}x \,\mathrm{d}y
\end{aligned}
$$

where $(X,Y) : \Omega \to \mathbb{R}^2$, $g : \mathbb{R}^2 \to \mathbb{R}$ and $g(x,y) = e^{xy}$.

If X and Y are independent and have joint density $f : \mathbb{R}^2 \to [0,\infty)$, then

$$\mathbb{P}((X,Y) \in E) = \int_E f \,\mathrm{dLeb} = \iint_E f \,\mathrm{d}x \,\mathrm{d}y$$

Then for $x, y \in \mathbb{R}$, set

$$f_X(x) = \int_{\mathbb{R}} f(x,y) \,\mathrm{d}y \qquad f_Y(y) = \int_{\mathbb{R}} f(x,y) \,\mathrm{d}x$$

Then f_X and f_Y are densities for X, Y respectively and $f \overset{a.e.}{=} f_X \cdot f_Y$.

- We take $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$
- Define $\mathcal{F}_1 \times \mathcal{F}_2$ (generated by rectangles or countable cylinders) and check that sectional functions and sectional integrals are well-defined and that order of integration does not matter.
- We used the monotone class theorem plus monotone convergence.
- This all works for n-fold products.
- Define the product measure using double integrals. (For an n-fold product, we use an n-fold integral and the uniqueness lemma says this is well-defined.)
- Order of integration is unimportant for non-negative functions and all orders give $\int f \,\mathrm{d}(\mu_1 \times \mu_2)$ (or $\int f \,\mathrm{d}(\mu_1 \times \cdots \times \mu_n)$).

Theorem 27 (Fubini's Theorem)**.** *If we have $(\Omega_1, \mathcal{F}_1, \mu_1)$, $(\Omega_2, \mathcal{F}_2, \mu_2)$ are σ-finite and $\mu = \mu_1 \times \mu_2$. If $f \in L^1(\mu)$, then*

$$\iint f \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2 = \int f \,\mathrm{d}\mu = \iint f \,\mathrm{d}\mu_2 \,\mathrm{d}\mu_1$$

(and all these integrals exist)

Proof. Write $f = f^+ - f^-$. By Tonelli, we have

$$\int f^+ \,\mathrm{d}\mu = \iint f^+ \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2$$

and

$$\int f^- \,\mathrm{d}\mu = \iint f^- \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2$$

so

$$
\begin{aligned}
\iint f \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2 &= \iint f^+ \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2 - \iint f^- \,\mathrm{d}\mu_1 \,\mathrm{d}\mu_2 \\
&= \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \\
&= \int f \,\mathrm{d}\mu
\end{aligned}
$$

□

20.1. **Infinite Product Spaces.** $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i \geq 1$ where we have probability spaces. Let $\mathcal{C}_n = \{$Degree n cylinder sets$\}$ and

$$\mathcal{G} = \{\text{Finite disjoint unions of cylinders}\}$$

Define, for $C = \bigcup_{i=1}^n C_i \in \mathcal{G}$, $C_i = E_{i,1} \times E_{i,2}$. where

$$C_i = E_{i,1} \times E_{i,2} \times \cdots \times E_{i,n_i} \times \Omega_{n_i+1} \times \Omega_{n_2+1} \times \cdots$$

We set

$$\mu(C) = \sum_{i=1}^{n} \prod_{i=1}^{n_i} \mu_i(E_{i,j})$$

However, what needs to be checked is

Prop. *$\mathcal{G}$ is an algebra, μ is well-defined and is a pre-measure on $\mathcal{G}$.*

Prop.

$$\sigma(\mathcal{G}) = \prod_{i \geq 1} \mathcal{F}_i$$

Thus, the product measure μ on $\prod_{i\geq 1} \mathcal{F}_i$ is uniquely defined. If $\mu_i, i \geq 1$ are the laws of $X_i, i \geq 1$. (so $\Omega_i = \mathbb{R}$ and $\mathcal{F}_i = \mathfrak{B}(\mathbb{R})$). However, we have that

$$\prod_{i\geq 1} \mathcal{F}_i \neq \mathfrak{B}(\mathbb{R}^\infty)$$

as we had before. Now let

$$\widehat{X}_j : \prod_{i=1}^{\infty} : \Omega_i \to \mathbb{R}$$

which sends $(\omega_i, i \geq 1)$ to ω_j. Then $(\widehat{X}_j, j \geq 1)$ are independent and $\widehat{X}_j$ has law μ_j.

20.2. **Stochastic Processes.** Brownian motion $(B_t, t \geq 0) : \Omega \to \mathbb{R}^{[0,\infty)}$ is a "random function" such that $B_0 = 0$ and

(1) For all $s, t \geq 0$,
$$B_{t+s} - B_t \sim \mathcal{N}(0, s)$$
(2) For all $0 \leq s_1 \leq t_1 \leq t_2 \leq \cdots \leq s_k \leq t_k$,
$$B_{t_1} - B_{s_1}, B_{t_2} - B_{s_2}, \ldots, B_{t_k} - B_{s_k}$$
are all independent.
(3) $(B_t, t \geq 0)$ is continuous with probability 1.

We'll see how to define a measure μ on $(\mathbb{R}^{[0,\infty)}, \mathfrak{B}(\mathbb{R}^{[0,\infty)})$ such that the identity map $B : \mathbb{R}^{[0,\infty)} \to \mathbb{R}^{[0,\infty)}$ satisfies (1) and (2).

21. November 20th, 2012

Recall: If X has law Λ, then in the probability space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \Lambda)$, the identity map $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$ has the same distribution as X, i.e. has law Λ. We want a probability measure $\mathbb{P}$ on $(\mathbb{R}^{[0,\infty)}, \prod_{t\geq 0} \mathfrak{B}(\mathbb{R}))$ so that the identity map from $\mathbb{R}^{[0,\infty)} \to \mathbb{R}^{[0,\infty)}$ has the distribution of Brownian motion (ignoring the continuity requirement.)

The requirements we needed to satisfy for Brownian motion were

(1) for all $s, t \geq 0$, $B_{s+t} - B_s \stackrel{d}{=} \mathcal{N}(0, t)$
(2) $B_{t_1 - s_1}, \ldots, B_{t_k - s_k}$ are independent if $s_! \leq t_1 \leq s_2 \leq t_2 \leq \ldots \leq s_k \leq t_k$. In other words, for $t_1 < t_2 < \ldots < t_k$,
$$(B_{t_1}, B_{t_2}, \ldots, B_{t_k}) \stackrel{d}{=} \left(N_1, N_1 + N_2, \ldots, \sum_{i=1}^{k} N_i \right)$$
where $N_1, \ldots, N_k$ are independent and $N_i \sim \mathcal{N}(0, t_i - t_{i-1})$ and $t_0 = 0$.

21.1. **Poisson Processes.**

Example 24 (Poisson Processes). ***Idea:*** *Modelling "arrivals" in a system where in each tiny time interval* $\mathrm{d}t$ *an arrival occurs with "infinitesimal probability"* $f(t)\,\mathrm{d}t$ *for some non-negative measurable function* $f \in L^1((0,\infty])$*, and this is independent for distinct infinitesimal intervals. Let* $\mathrm{d}t = 1$ *second and* $f(t) \equiv 1$*.*

$$\begin{aligned}\mathbb{P}(\textit{no busses in 4 minutes}) &= (1-\mathrm{d}t)^{240} \\ &= \left(1 - \frac{1}{600}\right)^{240} \\ &= \left(1 - \frac{1}{n}\right)^{cn} \\ &\approx e^{-c}\end{aligned}$$

Done carefully, this yields that the time to first arrival is distributed as $\exp(1)$*. Likewise, the time between first and second arrivals is* $\exp(1)$*. The number of arrivals by time* t *is distributed as* $\mathrm{Poisson}(t)$*. A poisson process on* $[0,\infty)$ *with rate* f *is a process* $(<_t, t \geq 0)$ *that satisfies*

(1) *For all* $0 \leq 2 \leq t < \infty$,

$$N_t - N_s \stackrel{d}{=} \mathrm{Poisson}\left(\int_s^t f(x)\,\mathrm{d}x\right)$$

(2) *Independent increments*

Definition 25. *Suppose we are given probability measures*

$$\{\mu_{t_1,\ldots,t_k}, k \geq 1 : (t_1,\ldots,t_k) \in \mathbb{R}^k\}$$

We say the family is consistent if

(1) *For any permutation* $\pi : [k] \to [k]$ *and any* $E_1,\ldots,E_k \in \mathfrak{B}(\mathbb{R})$,

$$\mu_{t_1,\ldots,t_k}(E_1 \times \cdots \times E_k) = \mu_{t_{\pi(1)},\ldots,t_{\pi(k)}}(E_{\pi(1)} \times \cdots \times E_{\pi(k)})$$

(2)

$$\mu_{t_1,\ldots,t_k}(E_1 \times \cdots \times E_k) = \mu_{t_1,\ldots,t_k,t_{k+1}}(E_1 \times \cdots \times E_k \times \mathbb{R})$$

21.2. **Kolmogorov Extension Theorem.**

Theorem 28 (Kolmogorov Extension Theorem). *For any consistent family of measures,* $\{\mu_{t_1,\ldots,t_k}, k \geq 1 : (t_1,\ldots,t_k) \in \mathbb{R}^k\}$*, then there exists a probability measure* $\mathbb{P}$ *on* $(\mathbb{R}^{\mathbb{R}}, \prod_{r\in\mathbb{R}} \mathfrak{B}(\mathbb{R}))$ *such that if* $X : \mathbb{R}^{\mathbb{R}} \to \mathbb{R}^{\mathbb{R}}$ *is the identity map, then* $\forall (t_1,\ldots,t_k) \in \mathbb{R}^k$*,* $(X(t_1),\ldots,X(t_k))$ *has law* $\mu_{t_1,\ldots,t_k}$*.*

Proof idea. Define $\mathbb{P}_0$ on $\mathcal{G}$ = finite disjoint unions of cylinders, where a cylinder is a set of the form $E_{t_1} \times \cdots \times E_{t_k}$, by setting

$$\mathbb{P}_0\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n \mu_{t_{i,1},\ldots,t_{i,k_i}}(C_i)$$

where $C_i = E_{t_{i,1}} \times E_{t_{i,2}} \times \cdots \times E_{t_{i,k_i}}$. Consistency conditions ensure this is well-defined and additive. Showing that $\mathbb{P}_0$ is very similar to the argument for the Lebesgue measure. Carathéodory extension then gives $\mathbb{P}$ with the desired properties. $\mathbb{P}$ is then the unique measure on $(\mathbb{R}^{\mathbb{R}}, \prod_{r\in\mathbb{R}} \mathfrak{B}(\mathbb{R}))$ with these finite-dimensional distributions. □

22. November 22nd, 2012

22.1. **Conditional Expectation. Idea:** $\mathbb{E}[X|Y]$ ="partial averaging". We pretend to know Y and average over the rest. Similarly, $\mathbb{E}[X|\mathcal{G}]$ = partial average where we pretend we know the information in $\mathcal{G}$ and average over the rest. Informally, for each $y \in \mathbb{R}$, each $\omega \in Y^{-1}(y)$, then we should have

$$\mathbb{E}[X|Y]\,(\omega) = \mathbb{E}[X|Y=y]$$

This suggests that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$, which is called the tower law.

Example 25.

(1) *X, Y are independent Uniform$[0,1]$ random variables. What is $\mathbb{E}\left[e^{XY}|Y\right]$? We first have*

$$\mathbb{E}\left[e^{Xy}\right] = \int_0^1 e^{xy}\,\mathrm{d}x = \frac{e^y - 1}{y}$$

So we should have

$$\mathbb{E}\left[e^{XY}|Y\right] = \frac{e^Y - 1}{Y}$$

(2) *What is $\mathbb{E}[X|X]$? $\mathbb{E}[X|X] = X$. More generally, if $\sigma(X) \subset \mathcal{G}$, then*

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

("$\mathbb{E}[X|\mathcal{G}]$ gets "more random" as $\mathcal{G}$ grows")

(3) *If X, Y are independent, then $\mathbb{E}[X|Y] \equiv \mathbb{E}[X]$.*

(4) *Let $X_i, i \geq 1$ be a sequence of iid random variables, $\mathbb{E}[X_1] = \mu$.*

$$S_n = X_1 + X_2 + \cdots + X_n$$

Then

$$\begin{aligned}
\mathbb{E}[S_{n+1}|S_n] &= \mathbb{E}[S_n + X_{n+1}|S_n] \\
&= \mathbb{E}[S_n|S_n] + \mathbb{E}[X_{n+1}|S_n] \\
&= S_n + \mu
\end{aligned}$$

Observation: *We seem to always have $\mathbb{E}[X|\mathcal{G}]$ is a measurable a function with respect to $\mathcal{G}$.*

(5) *X, Y are independent with laws μ, ν and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ such that $\mathbb{E}[|\varphi(X,Y)|] < \infty$. What is $\mathbb{E}[\varphi(X,Y)|X]$?*

Let $g : \mathbb{R} \to \mathbb{R}$ be such that

$$g(x) = \int_{\mathbb{R}} \varphi(x,y)\,\mathrm{d}\nu(y)$$

Then we should have $\mathbb{E}[\varphi(X,Y)|X] = g(X)$.

(6) *If X, Y have joint density $f_{X,Y}(x,y)$ What should $\mathbb{E}[X|Y=y]$ be?*

$$\mathbb{E}[X|Y=y] = \frac{\int x f_{X,Y}(x,y)\,\mathrm{d}x}{\int f_{X,Y}(x,y)\,\mathrm{d}x}$$

More strongly, given that $Y = y$, the conditional density of X should be

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

Then we should have $\mathbb{E}[X|Y] = \int x f_{X|Y}(x|Y)\,\mathrm{d}x$.

(7) *Let Ω be finite or countable and $\mathcal{F} = 2^\Omega$ and $\mathbb{P}$ is some probability. Let $(\Omega_i, i \geq 1)$ be a partition of Ω. Let $\mathcal{G} = \sigma(\Omega_i, i \geq 1)\sigma(Y)$, where $Y(\omega) = i$ if $\omega \in \Omega_i$. Then $\forall X : \Omega \to \mathbb{R}$, if $\omega \in \Omega_i$, then*

$$\begin{aligned}
\mathbb{E}[X|Y](\omega) &= \mathbb{E}[X|\omega \in \Omega_i] \\
&= \frac{\mathbb{E}[X\mathbb{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}
\end{aligned}$$

We would also get that

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_{i\geq 1} \mathbb{E}[\mathbb{E}[X|Y]\,|\mathbb{1}_{\Omega_i}] \\
&= \sum_{i \geq 1} \mathbb{E}\left[\frac{\mathbb{E}[X\mathbb{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)} \cdot \mathbb{1}_{\Omega_i}\right] \\
&= \sum_{i\geq 1} \mathbb{E}[X\mathbb{1}_{\Omega_i}] \\
&= \mathbb{E}[X]
\end{aligned}$$

Definition 26. *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. We say a random variable C is a **version** of $\mathbb{E}[X|\mathcal{G}]$ if*

(a) *C is $\mathcal{G}$-measurable*

(b) *For all $E \subset \mathcal{G}$, then*

$$\mathbb{E}[X\mathbb{1}_E] = \mathbb{E}[C\mathbb{1}_E]$$

Let us go back up to our examples and check that in (5), $g(X)$ is a version of $\mathbb{E}[\varphi(X,Y)|X] = \mathbb{E}[\varphi(X,Y)|\sigma(X)]$. For any $A \in \sigma(X)$, we have $A = X^{-1}(B)$ for some $B \in \mathfrak{B}(\mathbb{R})$.

$$A = \{\omega : (X(\omega), Y(\omega)) \in B \times \mathbb{R}\}$$

Then to check condition (b),

$$\begin{aligned}
\mathbb{E}[\varphi(X,Y)\mathbb{1}_A] &= \int \varphi(X,Y)\mathbb{1}_A \,\mathrm{d}\mathbb{P} \\
&= \int \varphi(X,Y)\mathbb{1}_{B\times\mathbb{R}}(x,y)\,\mathrm{d}(\mu(x)\times\nu(y)) \\
&= \int_B \underbrace{\int_{\mathbb{R}} \varphi(x,y)\,\mathrm{d}\nu(y)}_{g(x)}\,\mathrm{d}\mu(x) \quad \text{by Fubini} \\
&= \int_B g(x)\,\mathrm{d}\mu(x) \\
&= \int g(x)\mathbb{1}_A \,\mathrm{d}\mathbb{P} \\
&= \mathbb{E}[g(X)\mathbb{1}_A]
\end{aligned}$$

22.2. **Uniqueness of Conditional Expectation.** Fix C, D, two versions of $\mathbb{E}[X|\mathcal{G}]$. Suppose $\mathbb{P}(C > D) > 0$. Then $E := \{C > D\} \in \mathcal{G}, E_n := \{C > D + \frac{1}{n}\} \in \mathcal{G}$. We have $E_n \uparrow E$, so $\mathbb{P}(E_n) \to \mathbb{P}(E)$, so $\exists n$ such that $\mathbb{P}(E_n) > 0$. Then by (b),

$$\begin{aligned}
\mathbb{E}[X\mathbb{1}_{E_n}] &= \mathbb{E}[C\mathbb{1}_{E_n}] \\
&\geq \mathbb{E}\left[\left(D + \frac{1}{n}\right)\mathbb{1}_{E_n}\right] \\
&= \mathbb{E}[D\mathbb{1}_{E_n}] + \frac{\mathbb{P}(E_n)}{n} \\
&> \mathbb{E}[D\mathbb{1}_{E_n}]
\end{aligned}$$

so D is not a version of X, so we have a contradiction, meaning our assumption that $\mathbb{P}(C > D) > 0$ was false. So

$$\mathbb{P}(C \neq D) = \mathbb{P}(C > D) + \mathbb{P}(D > C) = 0$$

Lemma. *If $X \overset{a.s.}{\leq} Y$, then $\mathbb{E}[X|\mathcal{G}] \overset{a.s.}{\leq} \mathbb{E}[Y|\mathcal{G}]$.*

Proof. Suppose that $\mathbb{E}[X|\mathcal{G}]$ is not almost surely at most $\mathbb{E}[Y|\mathcal{G}]$. Then $\exists n$ such that $E_n = \{\mathbb{E}[X|\mathcal{G}] > \mathbb{E}[Y|\mathcal{G}] + \frac{1}{n}\}$ has $\mathbb{P}(E_n) > 0$. Then

$$\begin{aligned}
0 &\leq \mathbb{E}[(Y - X)\mathbb{1}_{E_n}] \\
&= \mathbb{E}[Y\mathbb{1}_{E_n}] - \mathbb{E}[X\mathbb{1}_{E_n}] \\
&= \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]\,\mathbb{1}_{E_n}] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathbb{1}_{E_n}] \\
&< \mathbb{E}\left[-\frac{1}{n}\mathbb{1}_{E_n}\right] \\
&= -\frac{\mathbb{P}(E_n)}{n} \\
&< 0
\end{aligned}$$

and so we have a contradiction. □

22.3. **Existence of Conditional Expectation.**

Theorem 29 (Conditional Expectation Existence)**.** *For all $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and any sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, there exists a version of $\mathbb{E}[X|\mathcal{G}]$.*

Key step: The theorem holds if $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. We project from the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto $L^1(\Omega, \mathcal{G}, \mathbb{P})$ with an inner product defined by

$$\langle U, V\rangle = \mathbb{E}[UV]$$

Then $\mathbb{E}[(X-Y)Z] = 0 \ \forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. We have $\mathbb{E}[XZ] = \mathbb{E}[YZ]$. Take $Z = \mathbb{1}_E$ for $E \in \mathcal{G}$, so

$$\mathbb{E}[X\mathbb{1}_E] = \mathbb{E}[Y\mathbb{1}_E]$$

23. November 27th, 2012

23.1. **Conditional Expectation.** $C = \mathbb{E}[X|\mathcal{G}]$ is such that

- C is $\mathcal{G}$-measurable
- $\forall E \in \mathcal{G}$, $\mathbb{E}[C\mathbb{1}_E] = \mathbb{E}[X\mathbb{1}_E]$.

Prop. *Conditional expectations exist.*

The key step is to show that if $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$, then there exists a version of $\mathbb{E}[X|\mathcal{G}]$.

Proof of prop from key step. We saw that if $X \overset{a.s.}{\leq} Y$, then $\mathbb{E}[X|\mathcal{G}] \overset{a.s.}{\leq} \mathbb{E}[Y|\mathcal{G}]$. Suppose that $X \geq 0$ and $\mathbb{E}[X] < \infty$. Let $X_n = \min(X, n)$ and let $C_n = \mathbb{E}[X_n|\mathcal{G}]$. Then $C_n \overset{a.s.}{\leq} C_{n+1} \overset{a.s.}{\leq} \dots$ so let

$$C = \limsup_{n\to\infty} C_n \overset{a.s.}{=} \lim_{n\to\infty} C_n$$

We have $C_n \xrightarrow{a.s.} C$.
Claim: C is a version of $\mathbb{E}[X|\mathcal{G}]$.

Proof.

- C is $\mathcal{G}$-measurable
- Fix $E \in \mathcal{G}$. Then

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_E] &= \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{1}_E] \\ &\overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}[C_n\mathbb{1}_E] \\ &= \mathbb{E}[C\mathbb{1}_E]\end{aligned}$$

□

For $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, let $X = X^+ - X^-$ and let $C^+ = \mathbb{E}[X^+|\mathcal{G}]$, $C^- = \mathbb{E}[X^-|\mathcal{G}]$. Then for any $E \subset \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_E] &= \mathbb{E}\big[X^+\mathbb{1}_E\big] - \mathbb{E}\big[X^-\mathbb{1}_E\big] \\ &= \mathbb{E}\big[C^+\mathbb{1}_E\big] - \mathbb{E}\big[C^-\mathbb{1}_E\big] \\ &= \mathbb{E}\big[(C^+ - C^-)\mathbb{1}_E\big]\end{aligned}$$

and $C = C^+ - C^-$ is a version of $\mathbb{E}[X|\mathcal{G}]$. □

Basic properties of conditional expectation:

- $\forall X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}[aX + Y|\mathcal{G}] \overset{a.s.}{=} a\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[Y|\mathcal{G}]$$

- If $X \overset{a.s.}{\leq} Y$, then

$$\mathbb{E}[X|\mathcal{G}] \overset{a.s.}{\leq} \mathbb{E}[Y|\mathcal{G}]$$

- If $X \uparrow \lim_{n\to\infty} X_n \overset{a.s.}{=} X$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \overset{a.s.}{=} \mathbb{E}[X|\mathcal{G}]$$

Proof of key step.

DIAGRAMS

We want to give $L^2(\Omega, \mathcal{F}, \mathbb{P})$ an inner product structure. We set

$$\langle X, Y \rangle = \frac{\mathbb{E}[XY]}{\|X\|_2 \|Y\|_2}$$

which we can interpret as "$\cos\theta$" assuming X and Y have mean 0. The norm of X is

$$\sqrt{\langle X, X \rangle} = \|X\|_2$$

For this to define a normed vector space, work with equivalence classes where $X \sim Y$ if

$$\|X - Y\|_2 = 0 \Leftrightarrow \mathbb{E}\big[(X-Y)^2\big] = 0 \Leftrightarrow X \overset{a.s.}{=} Y$$

Check: If $\|X_n - X\|_2 \to 0$, then $\|X_n - Y\|_2 \to 0$, then $X \sim Y$. ($[X_n] \to [X]$) This gives us an inner product space ($L^2\langle \bullet, \bullet \rangle$). More generally, $L^p(\Omega, \mathcal{F}, \mathbb{P}, \| \bullet \|_p)$ is a normed vector space.

Prop. *$L^p(\Omega, \mathcal{F}, \mathbb{P})$ is complete.*

Proof. We need to show: If $X_n, n \geq 1 \in L^p$, $\forall \varepsilon > 0$, $\exists n_0$ such that $\forall n, m \geq n_0$, $\|X_n - X_m\|_p < \varepsilon$, then $\exists X \in L^p$ such that $\|X_n - X\|_p \to 0$ as $n \to \infty$.

Let us first show that there exists a subsequental limit. For each $k \geq 1$, let n_k be such that $\forall n, m \geq n_k$, $\|X_n - X_m\|_p \leq \frac{1}{2^k}$. Then

$$\begin{aligned} \sum_{k \geq 1} \|X_{n_{k+1}} - X_{n_k}\|_p &\leq \sum_{k \geq 1} \frac{1}{2^k} \\ &= 1 \end{aligned}$$

We also have

$$\begin{aligned} \mathbb{E}\left[\sum_{k\geq 1} |X_{n_{k+1}} - X_{n_k}|\right] &= \sum_{k \geq 1} \mathbb{E}\big[|X_{n_{k+1}} - X_{n_k}|\big] \\ &\leq \sum_{k\geq 1} \|X_{n_{k+1}} - X_{n_k}\|_p \\ &= 1 \end{aligned}$$

If $a_k, k \geq 1$ is such that $\sum_{k \geq 1} |a_{k+1} - a_k| < \infty$ then (a_n) converges. Since $\mathbb{E}\Big[\sum_{k\geq 1} |X_{n_{k+1}} - X_{n_k}|\Big] \leq 1$, then

$$\sum_{k \geq 1} |X_{n_{k+1}} - X_{n_k}| < \infty$$

almost surely so X_{n_k} converges almost surely as $k \to \infty$. Let $X = \limsup_{k \to \infty} X_{n_k}$. Then $X_{n_k} \xrightarrow{a.s.} X$ as $k \to \infty$. It remains to show $X \in L^p$, $\|X_n - X\|_p \to 0$.

For any k and any $n \geq n_k$, $\|X_n - X_{n_k}\|_p \leq \frac{1}{2^k}$. By Fatou,

$$\begin{aligned} \liminf_{i \to \infty} \mathbb{E}[|X_n - X_{n_i}|^p] &\geq \mathbb{E}\Big[\liminf_{i\to\infty} |X_n - X_{n_i}|^p\Big] \\ &= \mathbb{E}[|X_n - X|^p] \end{aligned}$$

but we have that

$$\liminf_{i \to \infty} \mathbb{E}[|X_n - X_{n_i}|^p] \leq \frac{1}{2^{kp}}$$

so for all $n \geq n_k$, $\|X_n - X\|_p \leq \frac{1}{2^k}$. Letting $kc \to \infty$, we get $\|X_n - X\|_p \to 0$.

Finally,

$$\begin{aligned} \|X\|_p &\leq \|X - X_{n_k}\|_p + \|X_{n_k}\|_p \\ &< \infty \end{aligned}$$

□

Now let $\Delta = \inf\{\|X - Z\|_2 : Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})\}$. Let $Y_n, n \geq 1$ be any sequence of elements of the $L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that $\|Y_n - X\|_2 \to \Delta$.
Claim: Y_n is Cauchy.

DIAGRAMS

Proof. For $\varepsilon > 0$, let n_0 be such that $\forall n \geq n_0$, $\|X - Y_n\|_2^2 \leq \Delta^2 + \varepsilon$. Then for all $n, m \geq n_0$,

$$\begin{aligned} 2(\Delta^2 + \varepsilon) &\geq \|X - Y_n\|_2^2 + \|X - Y_m\|_2^2 \\ &= \mathbb{E}\big[(X - Y_n)^2\big] + \mathbb{E}\big[(X - Y_m)^2\big] \\ &= 2\mathbb{E}\left[\left(X - \frac{Y_n + Y_m}{2}\right)^2\right] + 2\mathbb{E}\left[\left(\frac{Y_n + Y_m}{2}\right)^2\right] \\ &\geq 2\Delta^2 + \frac{1}{2}\|Y_n - Y_m\|_2^2 \end{aligned}$$

so

$$2\varepsilon \geq \frac{1}{2}\|Y_n - Y_m\|_2^2$$

or in other words

$$\|Y_n - Y_m\| \leq 2\sqrt{\varepsilon}$$

$\square$

Since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is complete, let $Y \overset{a.s.}{=} \lim_{n\to\infty} Y_n$.
Claim: $\langle X - Y, Z\rangle = \mathbb{E}[(X - Y), Z] = 0$ for all $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Proof. Fix any $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then for all $t \in \mathbb{R}$

$$\mathbb{E}\big[(X - (Y + tZ))^2\big] \geq \mathbb{E}\big[(X - Y)^2\big]$$

This gives

$$t^2\mathbb{E}\big[Z^2\big] - 2t\mathbb{E}[Z(X - Y)] \geq 0$$

If $\mathbb{E}[Z(X - Y)] \neq 0$, then take

$$t = \pm\frac{\mathbb{E}[Z(X - Y)]}{\mathbb{E}[Z^2]}$$

and we get

$$t^2\mathbb{E}\big[Z^2\big] - 2t\mathbb{E}[Z(X - Y)] < 0$$

so we have a contradiction. $\square$

Corollary. *Y is a version of* $\mathbb{E}[X|\mathcal{G}]$.

Proof. $\forall E \in \mathcal{G}$, $\mathbb{1}_E \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ so

$$\mathbb{E}[(X - Y)\mathbb{1}_E] = 0$$

and

$$\mathbb{E}[X\mathbb{1}_E] = \mathbb{E}[Y\mathbb{1}_E]$$

$\square$

$\square$

24. November 29th, 2012

24.1. **Conditional Expectation.** Idea: "Hyperplane"(sub-subspace) $\to$ Dense subspace ($L^2(\Omega, \mathcal{F}, \mathbb{P})$) $\to$ Space of random variables (L^1)

Prop.

(1) $\mathbb{E}[aX + Y|\mathcal{G}] \overset{a.s.}{=} a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$
(2) $X \overset{a.s.}{\leq} Y \Rightarrow \mathbb{E}[X|\mathcal{G}] \overset{a.s.}{\leq} \mathbb{E}[Y|\mathcal{G}]$
(3) $0 \leq X_n \underset{a.s.}{\uparrow} X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *then* $\mathbb{E}[X_n|\mathcal{G}] \underset{a.s.}{\uparrow} \mathbb{E}[X|\mathcal{G}]$

Exercise: Use (3) to prove conditional versions of dominated convergence theorem and Fatou's lemma.

Prop (2: More properties of conditional expectation)**.** $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\mathcal{G} \subset \mathcal{F}$

(1) *If* $X \perp\!\!\!\perp Y$ *then* $\mathbb{E}[X|\sigma(Y)] \overset{a.s.}{=} \mathbb{E}[X]$.

(2) *(Tower Law). If $\mathcal{H} \subset \mathcal{G}$ is a sub-σ-algebra, then*

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,|\mathcal{H}] &\overset{a.s.}{=} \mathbb{E}[X|\mathcal{H}] \\ &\overset{a.s.}{=} \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\,|\mathcal{G}]\end{aligned}$$

Proof of (ii). **Recall:** If Z is $\mathcal{G}$-measurable then $\mathbb{E}[Z|\mathcal{G}] = Z$. But $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{G}$-measurable, and the second equality follows. Now let $W = \mathbb{E}[X|\mathcal{G}]$. We need to check

(1) $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$-measurable
(2) For all $E \in \mathcal{H}$,

$$\mathbb{E}[W\mathbb{1}_E] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\,\mathbb{1}_E]$$

For (2), since $W = \mathbb{E}[X|\mathcal{G}]$ and $E \in \mathcal{G}$, then

$$\begin{aligned}\mathbb{E}[W\mathbb{1}_E] &= \mathbb{E}[X\mathbb{1}_E] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\,\mathbb{1}_E]\end{aligned}$$

where the first equality comes from the defining property applied to W and the second equality comes from the defining property applied to $\mathbb{E}[X|\mathcal{H}]$. ☐

Prop (3)**.** *If Y is $\mathcal{G}$-measurable and either $X \in L^p$, $Y \in L^q$ for some $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$ or $X, Y \geq 0$ and $\mathbb{E}[X] < \infty$, $\mathbb{E}[XY] < \infty$, then*

$$\mathbb{E}[YX|\mathcal{G}] \overset{a.s.}{=} Y\mathbb{E}[X|\mathcal{G}]$$

Proof. Suppose that $Y = \mathbb{1}_E$, $E \in \mathcal{G}$. Then for any $F \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]\,\mathbb{1}_F] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathbb{1}_{E\cap F}] \\ &= \mathbb{E}[X\mathbb{1}_{E\cap F}] \\ &= \mathbb{E}[XY\mathbb{1}_F]\end{aligned}$$

By the defining property, $Y\mathbb{E}[X|\mathcal{G}] \overset{a.s.}{=} \mathbb{E}[XY|\mathcal{G}]$. Linearity yields the same result for simple functions. Next suppose that $X \geq 0$, $Y \geq 0$, and $\mathbb{E}[X] < \infty$, $\mathbb{E}[XY] < \infty$. Let $Y_n \uparrow Y$ where $Y_n \geq 0$ are simple functions. Then

$$\begin{aligned}\mathbb{E}[XY|\mathcal{G}] &\overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}[XY_n|\mathcal{G}] \\ &\overset{a.s.}{=} \lim_{n\to\infty} Y_n\mathbb{E}[X|\mathcal{G}] \\ &\overset{a.s.}{=} \lim_{n\to\infty} Y\mathbb{E}[X|\mathcal{G}]\end{aligned}$$

so we are done the second case.

Under the first case, write

$$XY = X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-$$

and apply linear of conditional expectation plus the second case to deduce the result. ☐

2 more examples:

(1) Let $\mathcal{G}, \mathcal{H}$, $\mathcal{H} \perp\!\!\!\perp \sigma(\mathcal{G}, \sigma(X))$. For example, we could have X, Y, Z, with $Z \perp\!\!\!\perp \sigma(X, Y)$. Then

$$\mathbb{E}[X|\mathcal{G}, \mathcal{H}] \overset{a.s.}{=} \mathbb{E}[X|\mathcal{G}]$$

Proof Idea. It suffices to consider $X \geq 0$. Let

$$\mathcal{P} = \{G \cap H : G \in \mathcal{G}, H \in \mathcal{H}\}$$

Claim: For all $F \in \mathcal{P}$,

$$\mu_X(F) = \mathbb{E}[X\mathbb{1}_F] = \mathbb{E}[\underbrace{\mathbb{E}[X|\mathcal{G}]}_{Y}\,\mathbb{1}_F] = \mu_Y(F)$$

If this held for all $F \in \sigma(\mathcal{G}, \mathcal{H})$, then it would precisely say that Y is a version of $\mathbb{E}[X|\mathcal{G}, \mathcal{H}]$, which is the claim. The uniqueness lemma proves this indeed holds for all $F \in \sigma(\mathcal{G}, \mathcal{H})$, assuming the claim.

Proof of Claim. We can write this in the form

$$\begin{aligned}
\mathbb{E}[X\mathbb{1}_F] &= \mathbb{E}[X\mathbb{1}_G\mathbb{1}_H] \\
&= \mathbb{E}[X\mathbb{1}_G]\,\mathbb{P}(H) \\
&= \mathbb{E}[Y\mathbb{1}_G]\,\mathbb{P}(H) \\
&= \mathbb{E}[Y\mathbb{1}_G\mathbb{1}_H] \\
&= \mathbb{E}[Y\mathbb{1}_F]
\end{aligned}$$

□

□

(2) If $X_1, \dots, X_n$ are independent and $h : \mathbb{R}^n \to \mathbb{R}$ such that $h(X_1, \dots, X_n) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then setting $\gamma : \mathbb{R} \to \mathbb{R}$ with

$$\gamma(x) = \mathbb{E}[h(x, X_2, X_3, \dots, X_n)]$$

we have

$$\gamma(X_1) = \mathbb{E}[h(X_1, \dots, X_n)|X_1]$$

This is proved by Fubini's theorem.

24.2. **Conditional Jensen's Inequality.** $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|\varphi(X)|] < \infty$
Then we have that

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]$$

Why is it that the proof of regular Jensen's inequality fails in this case? We had that $l(x) = ax + b$ and $l(\mathbb{E}[X]) = \varphi(\mathbb{E}[X])$, so $l(y) \leq \varphi(y)$ for all y. Then

$$\begin{aligned}
\mathbb{E}[\varphi(X)|\mathcal{G}] &\overset{a.s.}{\geq} \mathbb{E}[l(X)|\mathcal{G}] \\
&= \mathbb{E}[aX + b|\mathcal{G}] \\
&\overset{a.s.}{=} a\mathbb{E}[X|\mathcal{G}] + b \\
&= l(\mathbb{E}[X|\mathcal{G}]) \\
&\overset{?}{=} \varphi(\mathbb{E}[X|\mathcal{G}])
\end{aligned}$$

Proof. Let

$$S = \{(a, b) \in \mathbb{Q}^2 : ax + b \leq \varphi(x),\ \forall x \in \mathbb{R}\}$$

Then $\forall x$, $\varphi(x) = \sup\{ax + b : (a, b) \in S\}$. For any $(a, b) \in S$, $\varphi(X) \geq aX + b$ so

$$\mathbb{E}[\phi(X)|\mathcal{G}] \overset{a.s.}{\geq} a\mathbb{E}[X|\mathcal{G}] + b$$

Taking a supremum over $a, b \in S$, we get

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \overset{a.s.}{\geq} \underbrace{\sup\{a\mathbb{E}[X|\mathcal{G}] + b : (a, b) \in S\}}_{= \varphi(\mathbb{E}[X|\mathcal{G}]) \text{ for every } \omega}$$

□

24.3. **Conditional Hölder.** $\frac{1}{p} + \frac{1}{q} = 1$, $X \in L^p$, $Y \in L$. Then

$$\mathbb{E}[|XY| \mid \mathcal{G}] \leq \mathbb{E}[|X|^p \mid \mathcal{G}]^{1/p}\,\mathbb{E}[|Y|^q \mid \mathcal{G}]^{1/q}$$

Corollary (Corollary of Jensen)**.**

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]\,|^p] \leq \mathbb{E}[|X|^p]$$

Proof. Apply conditional jensen on $\varphi(x) = |x|^p$.

$$|\mathbb{E}[X|\mathcal{G}]\,|^p \overset{a.s.}{\leq} \mathbb{E}[|X|^p|\mathcal{G}]$$

Taking expected values, we get

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

where the right hand side is by the tower law. □

Assignment 4:

$$\mathbb{P}(|S_n| \geq t) \leq \left(\mathbb{E}\Big[e^{\lambda(X-\frac{t}{n})}\Big]\right)^n$$

and optimizing this bound over λ. Let λ^* be the optimal one. The last question on the assignment had to do with

$$\mathbb{P}(S_n = cn) \geq \frac{c}{\sqrt{n}}\left(\mathbb{E}\Big[e^{\lambda^*(X-\frac{t}{n})}\Big]\right)^n$$

25.1. **Changes of Variables.** $(\Omega, \mathcal{F}, \mathbb{P})$.
We have $X \in L^1$ is a random variable with law μ and cumulative distribution function F and density f. We have $g : \mathbb{R} \to \mathbb{R}$ with $g(X) \in L^1$. Then $\mathbb{E}[g(X)] = \int g \, \mathrm{d}\mu \overset{book}{=} \mu(g)$. Other notation for this is $\int g(x) \, \mathrm{d}\mu(x)$, $\int g(x)\mu(\,\mathrm{d}x)$ or $\int g \, \mathrm{d}F$ or $\int g(x)f(x) \, \mathrm{d}x$ or $\int g(x) \, \mathrm{d}F(x)$.

Having a density f means that for all $E \in \mathfrak{B}(\mathbb{R})$,

$$\begin{aligned}\mathbb{P}(X \in E) &= \int_E f \, \mathrm{d}Leb(\mathbb{R}) \\ &= \int_E f(x) \, \mathrm{d}x\end{aligned}$$

Example 26.

(1) $X \sim \exp(1)$, $f(x) = e^{-x}\mathbb{1}(x \geq 0)$

$$\begin{aligned}\mathbb{P}(X \geq t) &= \mathbb{E}\Bigg[\underbrace{\mathbb{1}(X \geq t)}_{g=\mathbb{1}_{[t,\infty)}}\Bigg] \\ &= \int \mathbb{1}_{[t,\infty)} \, \mathrm{d}\mu \\ &= \int_{[t,\infty)} e^{-x} \, \mathrm{d}x\end{aligned}$$

Now let us calculate $\mathbb{E}\big[X^3\big]$.

$$\begin{aligned}\mathbb{E}\big[X^3\big] &= \int_{[0,\infty)} x^3 e^{-x} \, \mathrm{d}x \\ &= \big[-x^3 e^{-x}\big]_0^\infty - 3\int_{[0,\infty)} x^2 e^{-x} \, \mathrm{d}x \\ &= 6\int_0^\infty e^{-x} \, \mathrm{d}x \\ &= 6\end{aligned}$$

We can continue doing this and we would actually get that

$$\mathbb{E}[X^n] = n!$$

(2) *If we have* $X_i, i \geq 1$ *independent and exp(1), then*

$$P_T = \max\left\{i : \sum_{j=1}^{i} X_j \leq t\right\} = Poisson(t)$$

We also have

$$G_k = \sum_{j=1}^{k} X_j = Gamma(k, l)$$

and

$$G_k > t \Leftrightarrow P_t < k$$

We have $X \overset{d}{=} Poisson(\lambda)$.

$$\mu \equiv \sum_{n \geq 0} \frac{\lambda^n e^{-\lambda}}{n!}\mathbb{1}_{\{n\}}$$

which gives us

$$\mathbb{P}(X \geq k) = \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!}$$

This is the same as

$$\begin{aligned}\mathbb{P}(X \geq k) &= \int \mathbb{1}_{[k,\infty)} \, \mathrm{d}\mu \\ &= \int_{[k,\infty)} 1 \, \mathrm{d}\mu \\ &= \mu([k,\infty))\end{aligned}$$

The expected value is given by

$$\begin{aligned}\mathbb{E}[X] &= \int x \, \mathrm{d}\mu \\ &= \sum_{n\geq 0} \frac{n\lambda^n e^{-\lambda}}{n!} \\ &= \lambda \sum_{n\geq 1} \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[e^{tX}\right] &= \int e^{tx} \, \mathrm{d}\mu(x) \\ &= \sum_{n\geq 0} \frac{e^{tn}\lambda^n e^{-\lambda}}{n!} \\ &= e^{-\lambda} \sum_{n\geq 0} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{(\lambda e^t)} \\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

(3) $X \stackrel{d}{=} Gamma(n, \lambda)$,

$$f(x) = \frac{\mathbb{1}_{(x\geq 0)}}{(n-1)!} \frac{x^{n-1} e^{-x/\lambda}}{\lambda^n}$$

$$\begin{aligned}\mathbb{E}[X] &= \int_{\mathbb{R}} x f(x) \, \mathrm{d}x \\ &= \frac{1}{(n-1)!} \int_0^{\infty} \frac{x^n e^{-x/\lambda}}{\lambda^n} \, \mathrm{d}x \\ &= \frac{1}{(n-1)!} \int_0^{\infty} \left(\frac{x}{\lambda}\right)^n e^{-x/\lambda} \, \mathrm{d}x\end{aligned}$$

Now set $y = x/\lambda$. This gives us

$$\begin{aligned}\mathbb{E}[X] &= \frac{\lambda}{(n-1)!} \int_0^{\infty} y^n e^{-y} \, \mathrm{d}y \\ &= \frac{\lambda}{(n-1)!} n! \\ &= \lambda n\end{aligned}$$

For higher moments, we have

$$\begin{aligned}\mathbb{E}\left[X^k\right] &= \int \frac{x^{n+k-1} e^{-x}}{\lambda^n (n-1)!} \, \mathrm{d}x \\ &= \frac{\lambda^{k-1}}{(n-1)!} \int \left(\frac{x}{\lambda}\right)^{n+k-1} e^{-x/\lambda} \, \mathrm{d}x\end{aligned}$$

$$= \frac{\lambda^k}{(n-1)!}(n+k-1)!$$

25.2. **Joint Law.** (X,Y) has joint law μ on $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$ with joint CDF F such that $F : \mathbb{R}^2 \to [0,1]$, where $F(x,y) = \mathbb{P}(X \le x, Y \le y)$. We have $g : \mathbb{R}^2 \to \mathbb{R}$, $g(X,Y) \in L^1$.

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \int_{\mathbb{R}^2} g \,\mathrm{d}\mu \\
&= \int g(x,y) f(x,y) \,\mathrm{d}(x \times y) \\
&= \int g(x,y) \,\mathrm{d}\mu(x,y) \\
&= \int g(x,y) \mu(\mathrm{d}(x \times y))
\end{aligned}$$

If there is a density then

$$\int g(x,y) \,\mathrm{d}\mu(x,y) = \int gf \,\mathrm{d}Leb(\mathbb{R}^2)$$

and

$$\begin{aligned}
\int g(x,y)\mu(\mathrm{d}(x \times y)) &= \iint g(x,y) f(x,y) \,\mathrm{d}x \,\mathrm{d}y \\
&= \iint g(x,y) f(x,y) \,\mathrm{d}y \,\mathrm{d}x
\end{aligned}$$

If (X,Y) has density f, then

$$\mathbb{P}((X,Y) \in E) = \int_E f \,\mathrm{d}Leb(\mathbb{R}^2)$$

Example 27. *(X_1, X_2) are independent $\mathcal{N}(0,1)$.*

$$H = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and we have $a^2 + c^2 = 1$, $b^2 + d^2 = 1$ with $ab + cd = 0$.
Claim:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = H^{-1} X$$

are independent $\mathcal{N}(0,1)$.
***NB:** Uncorrelated Normals need not be independent! If $N \sim \mathcal{N}(0,1)$ and*

$$W = \begin{cases} 1 & \text{with prob. } 1/2 \\ -1 & \text{with prob. } 1/2 \end{cases}$$

and $M = NW$. Then

$$\begin{aligned}
\mathbb{E}[MN] &= \mathbb{E}\big[N^2 W\big] \\
&= 0
\end{aligned}$$

But M and N are not independent.

Proof of Claim. We compute (for fixed $E \in \mathfrak{B}(\mathbb{R}^2)$)

$$\mathbb{P}((Y_1, Y_2) \in E)$$

We'll show this equals $\mathbb{P}((X_1, X_2) \in E)$. Denote $X = (X_1, X_2)$.

$$\begin{aligned}
\mathbb{P}((Y_1, Y_2) \in E) &= \mathbb{P}\big(H^{-1} X \in E\big) \\
&= \mathbb{P}(X \in H(E)) \\
&= \int \mathbb{1}_{[(x,y) \in H(E)]} f_{X_1}(x) f_{X_2}(y) \,\mathrm{d}(x \times y) \\
&= \int_{H(E)} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \,\mathrm{d}(x \times y) \\
&= \frac{1}{2\pi} \int_{H(E)} e^{-\frac{x^2+y^2}{2}} \,\mathrm{d}(x \times y)
\end{aligned}$$

$$= \frac{1}{2\pi} \int_{H(E)} e^{-\frac{1}{2}\|(x,y)\|_2^2} \, \mathrm{d}(x \times y)$$

We have that

$$\|(x,y)\|_2 = \|H(x,y)\|_2$$

since a rotation does not change where the origin is, so the distance must be the same. So we can actually write this as

$$\begin{aligned}
\mathbb{P}((Y_1, Y_2) \in E) &= \frac{1}{2\pi} \int_{H(E)} e^{-\frac{1}{2}\|H(x,y)\|_2^2} \, \mathrm{d}(x \times y) \\
&= \frac{1}{2\pi} \int_E e^{-\frac{1}{2}\|(x,y)\|_2^2} \, \mathrm{d}(x \times y) \\
&= \int_E f_{X_1}(x) f_{X_2}(y) \, \mathrm{d}(x \times y) \\
&= \mathbb{P}((X_1, X_2) \in E)
\end{aligned}$$

□

25.3. **Differentiation Under the Integral Sign.**
Recall:

$$L(\lambda) = \int_0^\infty e^{-\lambda x} f(x) \, \mathrm{d}x$$

We want

$$\begin{aligned}
L'(\lambda) &= \int_0^\infty \frac{\partial}{\partial \lambda} \left(e^{-\lambda x} f(x) \right) \, \mathrm{d}x \\
&= \int_0^\infty -x e^{-\lambda x} f(x) \, \mathrm{d}x
\end{aligned}$$

Proof. **FIX THIS!!** By definition,

$$L'(\lambda) = \lim_{h \to 0} \frac{L(\lambda + h) - L(\lambda)}{h}$$

We have

$$\begin{aligned}
L(\lambda + h) - L(\lambda) &= \int_{[0,\infty)} (e^{-(\lambda+h)x} - e^{-\lambda x}) f(x) \, \mathrm{d}x \\
&= \int_{[0,\infty)} \left(\int_{[0,h)} (-x) e^{-\lambda t} \, \mathrm{d}t \right) e^{-\lambda x} f(x) \, \mathrm{d}x \\
&= \int_{[0,\infty)} \int_{[0,h)} (-\lambda) e^{-\lambda(t+x)} f(x) \, \mathrm{d}t \, \mathrm{d}x
\end{aligned}$$

Write $I(\lambda, x) = e^{-\lambda x} f(x)$, then $(-\lambda) e^{-\lambda(t+x)} f(x+t) = \frac{\partial}{\partial \lambda} I(\lambda, x+t)$. K this all messed up. What we'll end up getting is that

$$\int_{[0,h]} \int_{[0,\infty)} \frac{\partial}{\partial \lambda} I(\lambda + t, x) \, \mathrm{d}x = h \frac{\partial}{\partial \lambda} I(\lambda, x)$$

and we divide by h. We get this by Fubini and continuity.

□

Not responsible for poisson processes, kolomogorov extension, brownian motion on the exam.

There will be 6 questions worth 66 questions. He will mark it out of 60, not 66. Probably of the same level of difficulty as the midterm.