Math 5530H Notes

Several People

May 20, 2021

1/11/2021

Definition. A probability space is an ordered triple $(\Omega, \mathscr{F}, P)$ such that Ω is the set of possible outcomes, $\mathscr{F}$ is the set of events (an event is a subset of Ω where probability is defined), and P is a probability measure.

We say $\mathscr{F}$ is a σ-field (or σ-algebra) on Ω, meaning $\Omega \in \mathscr{F}$ and $\mathscr{F}$ is closed under complements and countable unions (and hence contains the empty sat and is closed under countable intersection). We also require that $P : \mathscr{F} \longrightarrow [0,1]$ have $P(\Omega) = 1$, $P(\varnothing) = 0$, and that P is countably additive: if $A_1, A_2, \ldots \in \mathscr{F}$ are all disjoint then $P\left(\bigsqcup_n A_n\right) = \sum_n P(A_n)$.

Examples of probability spaces.

(a) Suppose we wish to describe flipping a coin as a probability space. In this case, $\Omega = \{H, T\}$ (heads and tails, respectively), $\mathscr{F} = \{\{H\}, \{T\}, \{H, T\}, \emptyset\}$, and, for any $A \in \mathscr{F}$, we have $P(A) = \frac{|A|}{2}$. What, then, is the probability that a heads is flipped? In this case, $P(\{H\}) = \frac{1}{2}$.

(b) Let's do the above, but flipping a coin twice. $\Omega = \{\{HH\}, \{HT\}, \{TH\}, \{TT\}\}$. As before, we have $\mathscr{F} = \mathcal{P}(\Omega)$ (the power set of Ω). In this case, for any $A \in \mathscr{F}$, we have $P(A) = \frac{|A|}{4}$.

(c) Here's an example of a probability space in which $\mathscr{F} \neq \mathcal{P}(\Omega)$. Let's say we want to pick a random point in $[0,1]$. So, $\Omega = [0,1]$. What is $\mathscr{F}$? Well, it suffices for now to say that it's complicated. Under the axiom of choice, we can deduce that $\mathscr{F} \neq \mathcal{P}(\Omega)$, although there exist set-theoretic models where this does not hold. In this probability space, P is the Lebesgue measure.

Relations, functions, and families of sets.

Definition. A relation R is a set of ordered pairs $(x, y) \in R$, which we write xRy. The domain of R is defined $\operatorname{dom} R = \{x \mid (x, y) \in R \text{ for some } y\}$ and similarly $\operatorname{rng} R = \{y \mid (x, y) \in R \text{ for some } x\}$. The inverse R^{-1} is defined to be $\{(y, x) \mid (x, y) \in R\}$.

If R, S are relations then we define their composition $S \circ R = \{(x, z) \mid \text{ there exists } y \text{ such that } (x, y) \in R \text{ and } (y, z) \in S\}$.

Definition. A function f is a relation where for any x, y_1, y_2, if $(x, y_1) \in f$ and $(x, y_2) \in f$ then $y_1 = y_2$ (i.e. we regard f as being the same as its graph). If f is a function then for any $x \in \operatorname{dom} f$, $f(x)$ denotes the unique y such that $(x, y) \in f$. We say f is *one-to-one* (*injective*) if and only if f^{-1} is a function. We write $f : X \longrightarrow Y$ to mean f is a function with $\operatorname{dom} f = X$ and $\operatorname{rng} f \subseteq Y$.

Definition. A family $(x_\alpha)_{\alpha \in A}$ is a function f such that $\operatorname{dom} f = A$ and for each $\alpha \in A$, we have $f(\alpha) = x_\alpha$. The set $\{x_\alpha \mid \alpha \in A\} = \operatorname{rng} f$ and the family $(x_\alpha)_{\alpha \in A} = \{(\alpha, x_\alpha) \mid \alpha \in A\}$

1/13/2021

Definition. Given a family of sets $\{X_\alpha\}_{\alpha \in A}$ where A is some indexing set, we can define a choice function as some $f : A \to \bigcup_{\alpha \in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for each $\alpha \in A$.

We can further define Cartesian Products of arbitrary families of sets.

Definition. The set $\prod_{\alpha\in A} X_\alpha = \{\{x_\alpha\}_{\alpha\in A} \mid x_\alpha \in X_\alpha \ \forall\ \alpha \in A\}$ is the set of choice functions from A to X.

The Axiom of Choice states that the Cartesian product of arbitrary non-empty sets is non-empty.

Definition. $Y^X = \prod_{x\in X} Y_x = \{\{y_x\}_{x\in X} \mid y_x \in Y \ \forall\ x \in X\} = \{f \mid f : X \to Y\}$ where $Y_x = Y$ for each $x \in X$. Additionally, it is worth noting that $Y^{\mathbb{N}}$ is the set of sequences in Y.

Coin tossing and [0,1]. Let $\Sigma = \{0,1\}^{\mathbb{N}}$ and for each $n \in \mathbb{N}$ let $S_n = \{0,1\}^n$. Now, for each $n \in \mathbb{N}$, $s = (s_1, s_2, \cdots, s_n) \in S_n$ define $\Sigma(s) = \{\tau = (t_1, t_2, t_3, \cdots) \in \Sigma \mid t_k = s_k \text{ for } k = 1, 2, \cdots, n\}$. Define $P(\Sigma(s)) = 2^{-n}$ for $s \in S_n$. So, for each $n \in \mathbb{N}$, $\Sigma = \sqcup_{s\in S_n} \Sigma(s)$. By countable additivity, $P(\Sigma) = 1$. Next, for each $n \in \mathbb{N}$, for each $s = (s_1, \ldots, s_n) \in S_n$, let $I(s) = [a(s), b(s)]$ where $a(s) = \sum_{k=1}^{n} \frac{s_k}{2^{-k}}$ and $b(s) = a(s) + \frac{1}{2^n}$. Thus $I(0) = [0, \frac{1}{2}]$, $I(1) = [\frac{1}{2}, 1]$ $(n = 1)$ $\quad I(0) = [0, 1/2], I(1) = [1/2, 1]$ $(n = 2)$ $\quad I(0,0) = [0, 1/4], I(0,1) = [1/4, 1/2], I(1,0) = [1/2, 3/4], I(1,1) = [3/4, 1]$. Let m be the Lebesgue measure on [0,1]. For each $n \in \mathbb{N}$, for each $s \in S_n$, $I(s)$ is an interval of length $1/2^n$ and so $m(I(s)) = 1/2^n = P(\Sigma(S))$. We notice this nice connection between the coin tossing space and the Lebesgue measure space on [0,1].

Aside on non-uniqueness of binary expansions. $1/2 = (0.1000\ldots)_2 = (0.01111\ldots)_2$. This non-uniqueness only happens for countably many of the elements of [0,1] (specifically those x for which the denominator is a power of 2). For each $x \in [0,1]$, $m(\{x\}) = 0$. For each $\sigma = (s_1, s_2, s_3, \ldots) \in \Sigma, P(\{\sigma\}) \le P(\Sigma(s_1, \ldots, s_n)) = \frac{1}{2^n}$ for each n, so $P(\{\sigma\}) = 0$. Thus for each countable set $A \subseteq \Sigma, P(A) = 0$. Moreover, for each countable set $B \subseteq [0,1]$, $m(B) = 0$.

Correspondence between Σ and $[0,1]$**.** Define f on Σ by $f(\sigma) = \sum_{k=1}^{\infty} s_k/2^k$ for each $\sigma = (s_1, s_2, \ldots) \in \Sigma$. Note that $0 \le f(\sigma) \le \sum_{k=1}^{\infty} 1/2^k = 1$. Thus $f : \Sigma \longrightarrow [0,1]$. Further, for each $\sigma = (s_1, s_2, s_3, \ldots) \in \Sigma$ and for each $n \in \mathbb{N}$, if $s = (s_1, \ldots, s_n)$ then

$$a(s) = \sum_{k=1}^{n} \frac{s_k}{2^k} \le f(\sigma) \le a(s) + \sum_{k=n+1}^{\infty} \frac{1}{2^k} = a(s) + \frac{1}{2^n}$$

so $f(\sigma) \in I(s)$.

Thus for each $\sigma = (s_1, s_2, \ldots) \in \Sigma$, $f(\sigma) \in \bigcap_{n=1}^{\infty} I(s_1, s_2, \ldots, s_n)$. Notice that the former will have length 0 as it is an intersection of intervals of length $1/2^n$ for each, so the length is less than $1/2^n$ for each n. In particular, we have that $\{f(\sigma)\} = \bigcap_{n=1}^{\infty} I(s_1, s_2, \ldots, s_n)$.

Claim. $f[\Sigma] = [0,1]$

Proof. Let $x \in [0,1]$ Define $s_1, s_2, s_3, \ldots$ inductively as follows. Note that $[0,1] = I(0) \cup I(1)$. Let $s_1 = 1$ if $s \in I(1)$, otherwise $s_1 = 0$, so $x \in I(s_1)$. For $n > 1$, note that $I(s_{n-1}) = I(s_{n-1}, 0) \cup I(s_{n-1}, 1)$, so let $s_n = 1$ if $x \in I(s_{n-1}, 1)$, otherwise $s_n = 0$. Let $\sigma = (s_1, s_2, s_3, \ldots)$. Then for each $n \in \mathbb{N}$, $x \in I(s_1, \ldots, s_n)$. Hence $x \in \bigcap_{n\in\mathbb{N}} I(s_1, \ldots, s_n) = \{f(\sigma)\}$, so $x = f(\sigma)$. ☐

But note that f isn't quite one to one. (Though it is up to a measure 0 subset. cf. Aside on non-uniqueness of binary expansions). Let Σ_0 be the set of all $\sigma = (s_1, s_2, s_3, \ldots) \in \Sigma$ such that there exists $n \in \mathbb{N}$ with $s_n = 1$ and for each $k > n$, $s_k = 0$. Also we define Σ_1 to be the set of all $\sigma = (s_1, s_2, s_3, \ldots)$ such that there is some k such that $s_k = 0$ and $s_k = 1$ for all $i > k$ Let $\sigma = (s_1, s_2, s_3, \ldots) \in \Sigma$ and $\tau = (t_1, t_2, t_3, \ldots) \in \Sigma$ be such that $\sigma \ne \tau$ but $f(\sigma) = f(\tau)$.

Claim. Either $\sigma \in \Sigma_1$ and $\tau \in \Sigma_0$, or vice versa.

Proof. Because $\sigma \ne \tau$, we have that the set $\{l \mid s_l \ne t_l\}$ is nonempty, and in particular has a least element n. Either $s_n = 1$ and $t_n = 0$, or vice versa. Without loss of generality, suppose $s_n = 1$.

Thus

$$\sum_{k=n+1}^{\infty} \frac{s_k + (1 - t_k)}{2^k} = \sum_{k=n}^{\infty} \frac{s_k - t_k}{2^k} = f(\sigma) - f(\sigma) = 0$$

where the second step abuses the fact that $\sum_{k=n+1}^{\infty} 1/2^k = 1/2^n$.

Thus for $k > n$, we have $s_k + (1 - t_k) = 0$. So, $s_k = 0$ and $t_k = 1$. Moreover, $\sigma \in \Sigma_0$ and $\tau \in \Sigma_1$. Similarly, if $s_n = 0$ and $t_n = 1$, $\sigma \in \Sigma_1$ and $\tau \in \Sigma_0$. $\square$

1/15/2021

As before, $\Sigma = \{0,1\}^{\mathbb{N}}$. The function $h : \Sigma \to [0,1]$ is defined by $h(\sigma) = \sum_{k=1}^{\infty} \frac{s_k}{2^k}$ for each $\sigma = (s_1, s_2, s_3, \dots) \in \Sigma$ (that is, $h = f$). $h[\Sigma] = [0,1]$ (though it is not one to one up to a countable set). Similarly, for each $n \in \mathbb{N}$, for each $s \in S_n$, we have $h[\Sigma(s)] = I(s)$. But, if $h(\sigma) - h(\tau)$ and $\sigma \neq \tau$ and if n is the least h such that $s_k \neq t_k$, then either $s_n = 1$ and $\forall k > n, s_k = 0$ and $t_k = 1$ or $s_n = 1$ for $k > n$ and $s_k = 1$ and $t_k = 0$. In particular, either $\sigma \in \Sigma_0$ and $\tau \in \Sigma_1$ or vice versa. Let $D = \{\frac{m}{2^n} \mid n \in \mathbb{N} \text{ and } m \in \{1, \dots, 2^n - 1\}\} = \cup_n D_n$ where $D_n = \{\frac{1}{2^n}, \frac{2}{2^n}, \dots, \frac{2^n-1}{2^n}\}$, and notice that $h[\Sigma_0] = D = h[\Sigma_1]$. Define $\Sigma_* = \Sigma \setminus (\Sigma_0 \cup \Sigma_1)$. h is one-to-one on $\Sigma \setminus \Sigma_0$ and also $\Sigma \setminus \Sigma_1$. Hence $h|_{\Sigma_*}$ is a bijection from Σ_* onto $[0,1] \setminus D$, and $h|_{\Sigma_0}$ is a bijection from Σ_0 onto D. Similarly, $h|_{\Sigma_1}$ is a bijection from Σ_1 onto D. Note that Σ_0 and Σ_1 are countably infinite sets and so there exists a bijection u from $\Sigma_0 \cup \Sigma_1$ onto Σ_0. Define $g : \Sigma \to [0,1]$ by

$$g(\sigma) = \begin{cases} h(\sigma) & \text{if } \sigma \in \Sigma_* \\ h(u(\sigma)) & \text{if } \sigma \in \Sigma_0 \cup \Sigma_1 \end{cases}$$

Then g is a bijection from Σ onto $[0,1]$. Note that $\{\sigma \in \Sigma : g(\sigma) \neq h(\sigma)\}$ is a subset of $\Sigma_0 \cup \Sigma_1$, so it is is countable. Thus $|[0,1]| = |\{0,1\}^{\mathbb{N}}|$ where $|A|$ denotes the cardinality $\#(A)$ here in this instance right now at the moment. $|[0,1]^2| = |(\{0,1\}^{\mathbb{N}})^2| = |\{0,1\}^{\mathbb{N}} \times \{0,1\}^{\mathbb{N}}| = |\{0,1\}^{\mathbb{N}}|$ This is because the map $((s_1, s_2, s_3, \dots), (t_1, t_2, t_3, \dots)) \mapsto (s_1, t_1, s_2, t_2, \cdots)$ is a bijection from $\left(\{0,1\}^{\mathbb{N}}\right)^2$ and $\{0,1\}^{\mathbb{N}}$. We can use this interleaving argument to show that $|[0,1]^n| = |[0,1]|$ for any $n \in \mathbb{N}$, and further $|[0,1]^{\mathbb{N}}| = |[0,1]|$ because $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$ and the following interleaving argument:

$$\begin{array}{ccccccccc}
(s_1^1, & \to & s_2^1, & & s_3^1, & \to & s_4^1, & & \dots) \\
& \swarrow & & \nearrow & & \swarrow & & & \\
(s_1^2, & & s_2^2, & & s_3^2, & & s_4^2, & & \dots) \\
\downarrow & \nearrow & & \swarrow & & & & & \\
(s_1^3, & & s_2^3, & & s_3^3, & & s_4^3 & & \dots) \\
& & & & \vdots & & & &
\end{array}$$

$$\mapsto (s_1^1, s_2^1, s_1^2, s_1^3, s_2^2, s_3^1, \dots)$$

Remark. We adopt the following notation: $\approx$ refers to "the same cardinality" or "equinumerous" and $\preceq$ refers to "smaller cardinality" or "subnumerous".

Claim. $|\mathbb{N}^{\mathbb{N}}| \approx |[0,1]|$, or in other words $\mathbb{N}^{\mathbb{N}} \approx [0,1]$.

Proof. $[0,1] \approx \{0,1\}^{\mathbb{N}} \subseteq \mathbb{N}^{\mathbb{N}} \preceq [0,1]^{\mathbb{N}}$ and $[0,1]^{\mathbb{N}} \approx [0,1]$. Thus, $[0,1] \preceq \mathbb{N}^{\mathbb{N}}$ and $\mathbb{N}^{\mathbb{N}} \preceq [0,1]$. Therefore, $[0,1] \approx \mathbb{N}^{\mathbb{N}}$ by the Cantor-Schöder-Bernstein theorem. $\square$

The Schröder-Bernstein theorem.

Theorem 1 (Schröder-Bernstein)**.** Let A and B be sets and suppose there are injections $f : A \to B$ and $g : B \to A$. Then there is a bijection from A onto B.

Proof. Let $A_0 = A$ and $B_0 = B$. For $n = 0, 1, 2, \dots$, define inductively $A_{n+1} = g[B_n]$ and $B_{n+1} = f[A_n]$. Then

$$A = A_0 \supseteq g[B_0] = A_1 \supseteq g[B_1] = A_2 \supseteq g[B_2] = A_3 \supseteq \cdots$$
$$B = B_0 \supseteq f[A_0] = B_1 \supseteq f[A_1] = B_2 \supseteq f[A_2] = B_3 \supseteq \cdots$$

Now, we let $C = \bigcap_{n=0}^{\infty} A_n$ and $D = \bigcap_{n=0}^{\infty} B_n$, and let $X_n = A_n \setminus A_{n+1}$ and similarly $Y_n = B_n \setminus B_{n+1}$ for $n = 0, 1, 2, \dots$ so that $f|X_{2n}$ is a bijection from X_{2n} onto Y_{2n+1} and $g^{-1}|X_{2n+1}$ is a bijection from X_{2n+1} onto Y_{2n+2}. Thus $f|C$ and $g^{-1}|C$ are both bijections from C onto D.

Define h on A by

$$h(x) = \begin{cases} f(x) & \text{if } x \in X_{2k} \\ g^{-1}(x) & \text{if } x \in X_{2k+1} \\ f(x) & \text{if } x \in C \end{cases}$$

so h is a bijection from A onto B. $\square$

Claim. $[1, \infty) \approx (1, \infty)$

Fancy Proof. $(1, \infty) \subseteq [1, \infty) \approx [2, \infty) \subseteq (1, \infty)$ so, $(1, \infty) \preceq [1, \infty)$ and $[1, \infty) \preceq (0, \infty)$, so $(1, \infty) \approx [1, \infty)$ by Schröder-Bernstein Theorem. $\square$

Elementary Proof. Define $f : [1, \infty) \to (1, \infty)$ by

$$f(x) = \begin{cases} x + 1 & \text{if } x \in \mathbb{N} \\ x & \text{otherwise} \end{cases}$$

Then f is a bijection from $[1, \infty)$ onto $(1, \infty)$. $\square$

Similarly, we have $[0, 1] \approx [0, 1) \approx (0, 1] \approx (0, 1)$.

If $-\infty < a < b < \infty$, then $[a, b] \approx [0, 1]$.

$(0, 1) \approx (-1, 1)$ and $(-1, 1) \approx \mathbb{R}$.

1/20/2021

Let $n \in \mathbb{N}$ and let $x \in \mathbb{R}$ such that $x \neq 1$. Let $S_n = 1 + x + \cdots + x^{n-1}$. We have $xS_n = x + x^2 + \cdots + x^n$, so $S_n - xS_n = 1 - x^n$, so we have

$$S_n = \frac{1 - x^n}{1 - x}$$

If $|x| < 1$, then as $n \to \infty$, $x^n \to 0$, so $S_n \to \frac{1}{1-x}$. We write $1 + x + x^2 + x^3 + \cdots = \frac{1}{x}$ for $|x| < 1$.

A careless example: We can write $1 + 2 + 4 + 8 + \cdots = \frac{1}{1-2} = -\frac{1}{2}$. What goes wrong? Well, $|2| > 1$, so the formula does not apply.

Example: Consider a biased coin, with the probability of heads being $p \in (0, 1)$, and the probability of tails being $q = 1 - p$. Consider tossing the coin until it lands on "heads":

$$\Omega = \{H, TH, TTH, TTTH, \dots\}$$

$$\mathscr{F} = \mathcal{P}(\Omega)$$

We see that $P(\{H\}) = p$, $P(\{TH\}) = qp$, and so on.

$$\begin{aligned} P(\Omega) &= p + qp + q^2p + q^3p + \cdots \\ &= p\left(1 + q + q^2 + q^3 + \cdots\right) \\ &= p\left(\frac{1}{1-q}\right) \\ &= 1 \end{aligned}$$

Let A be the even that the first head occurs on an odd toss. So,

$$A = \{H, TTH, TTTTH, \dots\}$$

$$P(A) = p + q^2p + q^4p + \cdots = p(1 + q^2 + q^4 + \dots) = p\left(\frac{1}{1-q^2}\right) = p\left(\frac{1}{2-p}\right)$$

Example. Roll 6 dice. What is the probability that each dice has a different number?

$$\Omega = \{1,2,3,4,5,6\}^6$$

$$\mathscr{F} = \mathcal{P}(\Omega), |\Omega| = 6^6$$

$P(\text{All 6 dice show different numbers}) = \frac{6!}{6^6} = \frac{5!}{6^5} = \frac{20}{6^4} = \frac{5}{3^4 2^2} = \frac{5}{324}$

Let $A, B, A_1, \dots, A_n$ be finite sets. If $A \cap B = \emptyset$, then $|A \cup B| = |A| + |B|$. If $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $|A_1 \cup \dots \cup A_n| = |A_1| + |A_2| + \dots + |A_n|$.

Further, $|A \times B| = |A| \times |B|$, since $A \times B = \dot\bigcup_{\{a\} \in A} \{a\} \times B$ and $|B| = |\{a\} \times B|$, $b \mapsto (a, b)$ is a bijection from B onto $\{0\} \times B$:

$$|\prod_{i=1}^n A_i| = \prod_{i=1}^n |A_i|$$
$$|B^A| = |B|^{|A|}$$

In general, $|A \cup B| = |A| + |B| - |A \cap B|$, because $A \cup B = \underbrace{(A \backslash B) \cup (A \cap B) \cup (B \backslash A)}_{\text{All disjoint}}$

So, $|A \cup B| = |A \backslash B| + |A \cap B| + |B \backslash A|$. Therefore,

$$|A \cup B| + |A \cap B| = |A \backslash B| + |A \cap B| + |B \backslash A| + |A \cap B| = |A| + |B|$$

Example (A more general Binomial Theorem)**.** If $a_1, \dots, a_n$ are numbers, then

$$\begin{aligned}\prod_{i=1}^n (1 + a_i) &= \sum_{\emptyset \neq I \subset \{1,2,\dots,n\}} (-1)^{|I|+1} \left| \bigcap_{i \in I} A_i \right| \\ &= \sum_{i_1} |A_{i_1}| - \sum_{i_1 < i_2} |A_{i_1} \cap A_{i_2}| + \sum_{i_1 < i_2 < i_3} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \pm \dots + (-1)^{-n+1} |A_1 \cap A_2 \cap \dots \cap A_n|\end{aligned}$$

Where for any set A,

$$1_A = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

$$\begin{aligned}1_{\bigcup_{i=1}^n A_i} &= 1_{(\bigcup_{i=1}^n A_i)^{cc}} = 1 - 1_{(\bigcup_{i=1}^n A_i)^c} = 1 - \prod_{i=1}^n 1_{A_i^c} \\ &= 1 - \prod_{i=1}^n (1 - 1_{A_i}) = 1 - \sum_{I \subseteq \{1,2,\dots,n\}} \prod_{i\ inI} (-1_{A_i}) = 1 - \sum_{I \subset \{1,\dots,n\}} (-1)^{|I|} \prod_{i \in I} 1_{A_i} \qquad \text{(by GBT)} \\ &= 1 - \sum_{I \subset \{1,\dots,n\}} (-1)^{|I|} 1_{\bigcap_{i \in I} A_i} = \sum_{\varnothing \neq I \subseteq \{1,2,\dots,n\}} 1_{\cap_{i \in I} A_i}\end{aligned}$$

Thus

$$\begin{aligned}
\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{\omega} 1_{\bigcup_{i=1}^{n} A_i^{(\omega)}} \\
&= \sum_{\omega, \emptyset \neq I \subseteq \{1,\ldots,n\}} 1_{\bigcap_{i=1}^{n} A_i^{(\omega)}} - \sum_{\omega} \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} 1_{\bigcap_{i \in I} A_i}(\omega) \\
&= \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \sum_{\omega} 1_{\bigcap_{i \in I} A_i}(\omega) \\
&= \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} \left|\bigcap_{i \in I} A_i\right|
\end{aligned}$$

This is called Poincaré's Inclusion Exclusion formula.

$(n)_k$ is read as "n permute k", or sometimes called the Pogchamper Symbol. It is defined as follows:

$$(n)_k = n(n-1)\cdots(n-k+1) \text{ for } n = 0,1,2,\ldots, \text{ and for } k = 0,1,\ldots,n$$

Let $\operatorname{inj}(A,B) =$ the set of injections from A to B.

Theorem 2. For $n = 0,1,2,3,\ldots$, for $k = 0,1,2,\ldots,n$, for all sets A, B, if $|A| = k$ and $|B| = n$, then

$$|\operatorname{inj}(A,B)| = (n)_k$$

Proof. By induction on n.

$$|\operatorname{inj}(\varnothing,\varnothing)| = |\{\varnothing\}| = 1 = (0)_0 \text{ so } P(0) \text{ is true.}$$

Let n be a whole number such that $P(n)$ is true. We wish to show that $P(n+1)$ is true. Let $k \in \{0,1,\ldots,n,n+1\}$ and let A, B be sets. Suppose $|A| = k$ and $|B| = n+1$. Either $k = 0$ or $k \geq 1$. If $k = 0$, then $A = \varnothing$, so $\operatorname{inj}(A,B) = |\{\varnothing\}| = (n)_0$.

Suppose $k \geq 1$. Then $A \neq \varnothing$. Let $a \in A$.

$\operatorname{inj}(A,B) = \dot\bigcup_{b \in B}\{f \in \operatorname{inj}(A,B) \mid f(a) = b\}$ For each $b \in B$,

$$|\{f \in \operatorname{inj}(A,B) \mid f(a) = b\}| = |\operatorname{inj}(A \backslash \{a\}, B \backslash \{b\})|$$

because $f \mapsto f|_{(A\backslash\{a\})}$ is a suitable bijection. Hence,

$$\begin{aligned}
|\operatorname{inj}(A,B)| &= (n+1)(n)_{k-1} \\
&= (n+1)(n)(n-1)\cdots(n-(k-1)+1) \\
&= (n+1)(n)(n-1)\cdots((n+1)-k+1) \\
&= (n+1)_k
\end{aligned}$$

Thus, $P(n+1)$ is true by induction. □

1/22/2021

Continue to assume that A is a finite set and $|A| = n$.

For $k = 0,1,\ldots,n$, $\mathcal{P}_k(A) = \{S \subseteq A \mid |S| = k\}$.

The symbol $\binom{n}{k} = |\mathcal{P}_k(A)|$, read n choose k, is the number of k-element subsets of an n-element set, and has

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1} \qquad k = 1,2,\ldots,n$$

The above is true by the following reasoning. We see

$$\mathcal{P}_k(\{1,\dots,n,n+1\}) = \mathcal{P}_k(\{1,\dots,n\}) \dot{\bigcup} \underbrace{\{S \cup \{n+1\} \mid S \in \mathcal{P}_{k-1}(\{1,\dots,n\})\}}_{\text{Equinumerous to } \mathcal{P}_{k-1}(\{1,\dots,n\})}$$

Pascal's Triangle:

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Theorem 3.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proof. Note that

$$\frac{n!}{k!n-k!!} = \frac{(n)_k}{k!}$$

Put $A = \{1,\dots,n\}$ and let $I = \text{inj}(\{1,\dots,k\}, A)$. Then $|I| = (n)_k$. For each $S \in \mathcal{P}_k(A)$, let

$$\begin{aligned} I_S &= \{f \in I \mid \text{rng}(f) = S\} \\ &= \text{inj}(\{1,\dots,k\}, S) \\ |I_S| &= (k)_k = k! \end{aligned}$$

Now $\bigcup_{S\in\mathcal{P}_k(A)} I_S = I$. Thus $\underbrace{\sum_{S\in\mathcal{P}_k(A)} \underbrace{|I_S|}_{k!}}_{\binom{n}{k} \text{ terms}} = \underbrace{|I|}_{(n)_k} = (n)_k$. There are $\binom{n}{k}$ summands on the left hand side, each $k!$, so we have $\binom{n}{k}k! = (n)_k$ and so

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}$$

. □

Theorem 4 (The Generalized Binomial Theorem (Final Hidden Phase))**.**

$$\prod_{i=1}^{n}(a_i + b_i) = \sum_{I\subseteq\{1,\dots,n\}} \left(\prod_{i\in I} a_i\right)\left(\prod_{i\notin I} b_i\right)$$

Theorem 5 (The Binomial Theorem)**.**

$$(a+b)^n = \sum_{k=0}^{\infty} \binom{n}{k} a^k b^{n-k}$$

Proof. Let $a_i = a$ and $b_i = b$ for $i = 1, 2, \ldots, n$. Then,

$$\begin{aligned}(a+b)^n = \prod_{i=1}^{n}(a_1 + b_i) &= \sum_{I\subseteq\{1,\ldots,n\}} \left(\prod_{i\in I} a_i\right)\left(\prod_{i\notin} b_i\right) \\ &= \sum_{\{1,\ldots,n\}} a^{|I|}b^{n-|I|} = \sum_{\{1,\ldots,n\}} a^k b^{n-k} \\ &= \sum_{k=0}^{n} \underbrace{\sum_{I\in\mathcal{P}_k(\{1,\ldots,n\})} a^k b^{n-k}}_{\binom{n}{k}\text{ terms}} = \sum_{k=0}^{\infty}\binom{n}{k}a^k b^{n-k}\end{aligned}$$

□

Friendship ended with combinatorics, now probability is my best friend.

Definition. Two events A and B are said to be independent when

$$P(AB) = P(A)P(B)$$

Example. Toss a coin twice

$$A = \{\text{heads on first toss}\}$$

$$B = \{\text{heads on the second}\}$$

Then $P(A) = P(B) = \frac{1}{2}$.

$$AB = \{\text{heads on both tosses}\}$$

$$P(AB) = 1/4 = 1/2 \cdot 1/2 = P(A)P(B)$$

Thus A and B are independent.

Let $C = \{\text{both tosses give same result}\} = \{HH, TT\}$.

$$P(C) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$A = \{HH, TT\}$, $AC = \{HH\}$

$$P(AC) = \frac{1}{4} = P(A)P(C)$$

Thus, A and C are independent.

$$B = \{HH, TH\}$$

$$BC = \{HH\}$$

$$P(BC) = \frac{1}{4} = P(B)P(C)$$

Thus B and C are independent.

$$ABC = A\{HH\} = \{HH\}$$

Thus $P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$. So A, B, C are pairwise independent, but not independent.

Definition. Three events A, B, and C, are independent when each two of them are independent, and in addition, $P(ABC) = P(A)P(B)P(C)$.

Definition. We say that n events $A_1, \ldots, A_n$ are independent when for each J, if $\varnothing \neq J \subseteq \{1, \ldots, n\}$, then

$$P(\bigcap_{j\in J} A_j) = \prod_{i\in J} P(A_j)$$

Definition. A family $(A_i)_{i\in I}$ is independent when for each I_0, if $\varnothing \neq I_0 \subseteq I$ is finite we have

$$P\left(\bigcap_{i\in I_0} A_i\right) = \prod_{i\in I_0} P(A_i)$$

Definition. If $P(B) \neq 0$, then the conditional probability of A given B is written

$$P(A \mid B) = \frac{P(AB)}{P(B)} = Q(A)$$

If $P(B) \neq 0$, then A and B are independent if and only if $P(A \mid B) = P(A)$.

Example. Suppose we roll two dice.

$$P(\text{sum} \geq 7 \mid \text{sum} \geq 3) = ?$$

k	$P(\text{sum} = k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

$P(\text{sum} \geq 7) = \frac{21}{36}$. $P(\text{sum} \geq 3) = \frac{35}{36}$

$$P(\text{sum} \geq 7 \mid \text{sum} \geq 3) = \frac{P(\{\text{sum} \geq 7\} \cap \{\text{sum} \geq 3\})}{P(\text{sum} \geq 3)} = \frac{\frac{21}{36}}{\frac{35}{36}} = \frac{21}{35} = \frac{3}{5} > \frac{7}{12} = \frac{21}{36} = P(\text{sum} \geq 7)$$

This is because we have reduced the sample space without removing any favorable outcomes.

$P(\text{sum is odd}) = \frac{18}{36} = \frac{1}{2}$.

$$P(\text{sum is odd} \mid \text{sum} \geq 7) = \frac{P(\text{sum} \in \{7, 9, 11\})}{P(\text{sum} \geq 7)} = \frac{\frac{12}{36}}{\frac{21}{36}} = \frac{12}{21} > \frac{1}{2}$$

Example (The Gambler's Rule)**.** Suppose you play a game over and over, each time with the same chance $\frac{1}{N}$ of winning. How many times n must you play to have a better than 50% chance to win?

According to a very old gambler's rule, n is about $\frac{2}{3}N$:

$$\begin{aligned}
&P(\text{at least one win in } n \text{ games}) \\
&\quad = 1 - P(\qquad\qquad\qquad\qquad\qquad \text{no win in } n \text{ games}) \\
&\quad = 1 - \left(1 - \frac{1}{N}\right)^n.
\end{aligned}$$

We seek the least n such that

$$1 - \left(1 - \frac{1}{N}\right)^n < \frac{1}{2}.$$

Taking logarithms,

$$n \ln\left(1 - \frac{1}{N}\right) < \ln\frac{1}{2}.$$

Then find n_*, perhaps not an integer, so that

$$n_* \ln\left(1-\frac{1}{N}\right) = \ln\frac{1}{2}.$$

Then n is the least integer $> n_*$.

$$n_* = \frac{\ln\frac{1}{2}}{\ln\left(1-\frac{1}{N}\right)}.$$

For $N = 1, 2, \ldots, 27$, n_* is so close to $2/3N$ that n is the least integer $> {}^2\!/_3N$. For $N = 28$, this is not true. But, $\frac{n}{\frac{2}{3}N}$ almost $\to 1$ as $N \to \infty$.

Now, $\frac{\ln(1+z)}{z} \to 1$ as $z \to 0$. Applying this with $z = (-\frac{1}{N})$, as $N \to \infty$, $z \to 0$, so

$$n \sim n_* \sim \frac{\ln\frac{1}{2}}{-\frac{1}{N}} = N\ln 2.$$

In other words,

$$\frac{n}{N\ln 2} \to 1$$

as $N \to \infty$. Now, $\ln 2 \approx 0.693\cdots \approx \frac{2}{3}$. Hence, $\frac{n}{2/3N}$ is almost 1 for large N.

1/25/2021

Example (Probability of a flush in poker)**.**

$$\Omega = \{\mathrm{inj}(\{1,\ldots,5\},\{1,\ldots,52\})\}$$

$$|\Omega| = (52)_5 = (52)(51)(50)(49)(48)\text{and}P(A) = \frac{|A|}{|\Omega|}.$$

Say cards 1 through 13 are the spades so $\{\text{spade flush}\} = \mathrm{inj}(\{1,\ldots,5\},\{1,\ldots,13\})$ and

$$P(\{\text{spade flush}\}) = \frac{(13)_5}{(52)_5} = \frac{33}{1360}$$

The probability of a flush of any suit is the same, so the probability of any flush is

$$P(\{\text{flush}\}) = 4P(\{\text{spade flush}\}) = \frac{33}{16660} \approx 0.00198079\ldots$$

Example (The Birthday Problem)**.** Given a class of n students, what are the probability that at least two have the same birthday? (Assume no leap years, uniform probability of being born on a given day). Then let $p_n = P(\text{at least 2 have the same birthday})$ and $q_n = 1 - p_n$ so

$$\begin{aligned}
P(\text{at least 2 have the same birthday}) &= 1 - P(\text{add } n \text{ have different birthdays}) \\
&= 1 - \frac{(365)_n}{365^n} \\
&= 1 - \left(\frac{365}{365} \cdot \frac{365-1}{365} \cdot \dots \cdot \frac{365-n+1}{365}\right) \\
&= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)
\end{aligned}$$

One can check that the least n such that $p_n > \frac{1}{2}$ is $n = 23$. $p_{23} \approx 0.506\ldots$, $p_{45} \approx 0.94\ldots$, $p_{65} \approx 0.998\ldots$.

Note that for $\frac{n-1}{365} << 1$, we have

$$\ln q_n = \ln\left(1 - \frac{1}{365}\right) + \ln\left(1 - \frac{n-1}{365}\right)$$

For $-1 < x \leq 1$, $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - + \cdots$ so

$$\begin{aligned}
&\approx \left(-\frac{1}{365}\right) + \left(-\frac{2}{365}\right) + \cdots + \left(-\frac{n-1}{365}\right) \\
&= -\frac{1}{365}\left(1 + 2 + \ldots + (n-1)\right) \\
&= -\frac{1}{365}\frac{n(n-1)}{2}
\end{aligned}$$

Hence $q_n \approx e^{-n(n-1)/730}$. This approximation is good for all n (because log gets flat). For $n = 23$, we have

$$1 - e^{23(23-1)/730} = 0.5000018\ldots \approx p_n = 0.0506\ldots$$

Estimating the error. For $0 < x < 1$,

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots = -x - x^2 \underbrace{\left(\frac{1}{2} + \frac{x}{3} + \frac{x^2}{4} + \cdots\right)}_{\theta}$$

$\frac{1}{2} < \theta < \frac{1}{2} + \frac{x}{3}(1 + x + x^2 + \cdots = \frac{1}{2} + \frac{x}{3}\frac{1}{1-x} < \frac{1}{2} + \frac{1}{3}$ if $0 < x < \frac{1}{2}$. Thus

$$\prod_{j=1}^{n-1}\left(1 - \frac{j}{365}\right) = e^{\sum_{j=1}^{n-1}\ln\left(1 - \frac{j}{365}\right)} = e^{\sum_{j=1}^{n-1}\left(-\frac{j}{365} - \left(\frac{j}{365}\right)^2\theta_j\right)}$$

etc.

Example. n men put their hats in a sack. Each man draws a hat from the sack at random and keeps it. What is the probability that at least one man draws his own hat?

Let $A_j = \{\text{man } j \text{ draws his own hat}\}$. We want to find $P\left(\bigcup_{j=1}^n A_j\right)$.

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{\varnothing \neq J \subseteq \{1,\ldots,n\}} (-1)^{|J|+1} P\left(\bigcap_{j \in J} A_j\right)$$

$P(A_j) = \frac{1}{n}$ We can see put

$$\Omega = \text{inj}(\{1, \ldots, n\}, \{1, \ldots, n\})$$

so if $\varnothing \neq J \subseteq \{1, \ldots, n\}$ then $\bigcap_{j \in J} A_j = \{\omega \in \Omega : \omega(j) = j \text{ for each } j \in J\}$, which has is in bijection with $\operatorname{inj}(\{1, \ldots, n\} \setminus J, \{1, \ldots, n\} \setminus J)$ by restricting elements of the first set to $\{1, \ldots, n\} \setminus J$. This cardinality is $(n - |J|)!$.

Hence

$$P\left(\bigcap_{j \in J} A_j\right) = \frac{|\bigcap_{j \in J} A_j|}{|\Omega|} = \frac{(n - |J|)!}{n!}$$

so

$$\begin{aligned} p &= \sum_{\varnothing \neq J \subseteq \{1,\ldots,n\}} (-1)^{|J|+1} \frac{(n - |J|)!}{n!} = \sum_{k=1}^{n} \sum_{J \in \mathcal{P}_k(\{1,\ldots,n\})} (-1)^{k+1} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^{n} \binom{n}{k} (-1)^{k+1} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} \\ &= 1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!} \xrightarrow[n \to \infty]{} 1 - e^{-1} \end{aligned}$$

because $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Expressibility.

Definition. Let Ω be a set, and let X and Y be functions on Ω (we use X and Y with the goal of moving towards random variables). To say Y is expressible in terms of X (Y is expressible in terms of X) means for all $\omega_1, \omega_2 \in \Omega$ if $X(\omega_1) = X(\omega_2)$ then $Y(\omega_1) = Y(\omega_2)$.

Proposition 6. The following are equivalent:

(a) Y is expressible in terms of X.

(b) There is a function f such that $\operatorname{rng} X \subseteq \operatorname{dom} f$ and for each $\omega \in \Omega$, $Y(\omega) = f(X(\omega))$ (that is, $Y = f \circ X$ which, through abuse of notation, we may write $Y = f(X)$).

Proof. (b) $\implies$ (a). Let X, Y, f be as described. Let $\omega_1, \omega_2 \in \Omega$ with $X(\omega_1) = X(\omega_2)$. Then $Y(\omega_1) = f(X(\omega_1)) = f(X(\omega_2)) = Y(\omega_2)$. Thus Y is expressible in terms of X.

(a) $\implies$ (b). Suppose Y is expressible in terms of X. Let $f = \{(X(\omega), Y(\omega) : \omega \in \Omega\}$. We claim f is a function.

Let $(x, y_1), (x, y_2) \in f$. We want to show $y_1 = y_2$. Since $(x, y_1) \in f$, there exists $\omega_1 \in \Omega$ such that $(x, y_1) = (X(\omega_1), Y(\omega_1))$. Since $(x, y_2) \in f$, there exists $\omega_2 \in \Omega$ such that $(x, y_2) = (X(\omega_1), Y(\omega_2))$. Then $X(\omega_1) = X(\omega_2) = x$, so since Y is expressibile in terms of X we have $y_1 = Y(\omega_1) = Y(\omega_2) = y_2$, proving the claim.

Now, $\operatorname{dom} f = \{X(\omega) : \omega \in \Omega\} = \operatorname{rng} X$. For each $(x, y) \in f$, $f(x) = y$. For each $\omega \in \Omega$, $(X(\omega), Y(\omega)) \in f$, so $f(X(\omega)) = Y(\omega)$. □

Remark. The statement $(\exists x)P(x)$ may be read "there exists x such that $P(x)$" or "there is x with $P(x)$." The "such that" is built into the symbol $\exists$ and apparently this is important.

Definition. Let $A_1, \ldots, A_n, B \subseteq \Omega$. To say B is expressible in terms of $A_1, \ldots, A_n$ means Y is expressible in terms of X, where $Y = 1_B$ and X is the function on Ω defined by $X(\omega) = (1_{A_1}(\omega), \ldots, 1_A n(\omega))$.

Proposition 7. Let $B \subseteq \Omega$ and let $X : \Omega \longrightarrow \mathbf{X}$. Then the following are equivalent:

(a) 1_B is expressible in terms of X

(b) $B = X^{-1}[C]$ for some $C \subseteq \mathbf{X}$.

Proof. (a) $\implies$ (b). Suppose 1_B is expressible in terms of X. Then there is a function f on $\operatorname{rng} X$ such that $1_B = f \circ X$. Thus for each $\omega \in \Omega$, $1_B(\omega) = f(X(\omega))$. Hence for each $x \in \operatorname{rng} X$, $f(x) \in \{0, 1\}$. Let $C = \{x \in \operatorname{dom} X : f(x) = 1\}$. Then for each $\omega \in \Omega$, we have $\omega \in B$ if and only if $1_B(\omega) = 1$ if and only if $f(X(\omega)) = 1$ if and only if $X(\omega) \in C$. Thus $B = X^{-1}[C]$.

(b) $\implies$ (a). Suppose $B = X^{-1}[C]$ for some $C \subseteq \mathbf{X}$. Let $f = 1_C$. Then for each $\omega \in \Omega$ we have $1_B(\omega) = 1$ if and only if $\omega \in B$ if and only if $\omega \in X^{-1}[C]$ if and only if $X(\omega) \in C$ if and only if $1_C(X(\omega)) = 1$. Hence also for each $\omega \in \Omega$, $1_B(\omega) = 0$ if and only if $1_C(X(\omega)) = 0$. Since 1_B and 1_C assume only the values 0 and 1, we must have $1_B = 1_C = X$. Hence 1_B is expressible in terms of X. □

1/27/21

Proposition 8. Let $X : \Omega \longrightarrow \mathbf{X}$ be a function with $\mathbf{X} = X[\Omega]$ and let $\mathscr{B} = \{B \subseteq \Omega : 1_B$ is expressible in terms of $X\}$. Then

(a) $C \mapsto X^{-1}[C]$ is a bijection from $\mathcal{P}(\mathbf{X})$ onto $\mathscr{B}$.

(b) $|\mathscr{B}| = 2^{|\mathbf{X}|}$.

Proof. (b) follows from (a) because by (a) $|\mathscr{B}| = |\mathcal{P}(\mathbf{X})| := 2^{|\mathbf{X}|}$ (because $|\mathcal{P}(X)| = |\{0,1\}^{\mathbf{X}}|,\ A \mapsto 1_A$).

(a) follows from (b) because by Proposition 7, $B \in \mathscr{B}$ iff $B = X^{-1}[C]$ for some $C \subseteq \mathbf{X}$ ($X^{-1}[S] = X^{-1}[S \cap \operatorname{rng}(X)]$)] Thus $C \mapsto X^{-1}[C]$ maps $\mathcal{P}(\mathbf{X})$ onto $\mathscr{B}$. For each set S, $X[X^{-1}[S]] = S \cap \mathbf{X}$. Thus for each $C \subseteq \mathbf{X}$, $X[X^{-1}[C]] = C$, so $C \mapsto X^{-1}[S]$ is one-to-one on $\mathcal{P}(\mathbf{X})$. □

Proposition 9. Let $(X_i)_{i \in I}$ be a family of functions on Ω. Define X on Ω by $X(\omega) = (X_i(\omega))_{i \in I}$. Let $x = (x_i)_{i \in I}$. Then $X^{-1}[\{x\}] = \bigcap_{i \in I} X_i^{-1}[\{x_i\}]$.

Proof.

$$\begin{aligned} X^{-1}[\{x\}] &= \{\omega \in \Omega : X(\omega) = x\} \\ &= \{\omega \in \Omega : (X_i(\omega))_{i \in I} = (x_i)_{i \in I}\} \\ &= \{\omega \in \Omega : \text{ for each } i \in I, X_i(\omega) = x_i\} \\ &= \bigcap_{i \in I} \{\omega \in \Omega : X_i(\omega) x_i\} \\ &= \bigcap_{i \in I} X_i^{-1}[\{x_i\}] \end{aligned}$$

□

Remark. Let $G \subseteq \Omega$.

(a) Let $x \in \{0, 1\}$. Then;

$$1_G^{-1}[\{x\}] = \begin{cases} G & \text{if } x = 1 \\ G^c & \text{if } x = 0 \end{cases}$$

(b) Let $\omega \in \Omega$. Then

$$1_G^{-1}[\{1_G(\omega)\}] = \begin{cases} G & \text{if } \omega \in G \\ G^c & \text{if } \omega \in G \end{cases}$$

Definition. To say that $\mathscr{F}$ is a field of subsets of Ω means that

(a) $\mathscr{F} \subseteq \mathcal{P}(\Omega)$

(b) $\Omega \in \mathscr{F}$

(c) For each $F \in \mathscr{F}$, $\Omega \setminus F \in \mathscr{F}$

(d) For all $F_1, F_2 \in \mathscr{F}$, $F_1 \cup F_2 \in \mathscr{F}$

Remark. Let $\mathscr{F}$ be a field of subsets of Ω. Then

(b') $\varnothing \in \mathscr{F}$ by (b) and (c)

(d') For all $F_1, F_2 \in \mathscr{F}$, $F_1 \cap F_2 \in \mathscr{F}$ by (c), (d), and DeMorgan's laws.

Theorem 10. Let $A_1, \ldots, A_n \subseteq \Omega$. Let $\mathscr{B} = \{B \subset \Omega : B \text{ is eit of } A_1, \ldots, A_n\}$. Define $X : \Omega \to \{0,1\}^n$ by $X(\omega) = (1_{A_1}(\omega), \ldots, 1_{A_n}(\omega)\}$. Let $\mathbf{X} = X[\Omega]$. Then

(a) $C \mapsto X^{-1}[C]$ is a bijection from $\mathcal{P}(\mathbf{X})$ onto $\mathscr{B}$.

(b) $|\mathscr{B}| = 2^m$ where $m = |\mathbf{X}| \leq 2^n$.

(c) $\mathscr{B}$ is the smallest field of subsets of Ω which contains $A_1, \ldots, A_n$.

Proof. To say that B is expressible in terms of $A_1, \ldots, A_n$ means 1_B is expressible in terms of X. Hence (a) follows from Proposition 8.

(b) By (a), $|\mathscr{B}| = |\mathcal{P}(\mathbf{X})| = 2^{|\mathbf{X}|}$. Now, $\mathbf{X} \subseteq \{0,1\}^n$, so $|\mathbf{X}| \leq |\{0,1\}^n| = 2^n$.

(c) Since $\mathcal{P}(\mathbf{X})$ is a field of subsets of $\mathbf{X}$ and since passage to preimages preserves set operations, $\mathscr{B}$ is a field of subsets of Ω. For $k = 1, \ldots, n$, $1_{A_k} = \pi_k \circ X$, the projection of X onto A_k. Thus 1_{A_k} is expressible in terms of X, so A_k is expressible in terms of $A_1, \ldots, A_n$. Thus $\mathscr{B}$ is a field of subsets of Ω and $A_1, \ldots, A_n \in \mathscr{B}$.

To see that $\mathscr{B}$ is the smallest such field, let $\mathscr{E}$ be a field of subsets of Ω such that $A_1, \ldots, A_n \in \mathscr{E}$. We want to show that $\mathscr{B} \subseteq \mathscr{E}$.

Let $B \in \mathscr{B}$. Then by (a), $B = X^{-1}[C]$ for some $C \subset \mathbf{X}$. Now $C = \bigcup_{x \in C} \{x\}$, so $B = \bigcup_{x \in C} X^{-1}[\{x\}]$.

Since C is finite and $\mathscr{E}$ is closed under finite unions, it suffices to show that for each $x \in C$, $X^{-1}[\{x\}] \in \mathscr{E}$. By Proposition 9, with $X_i = 1_{A_i}$ for $i = 1, \ldots, n$ we have for each $x \in C$ that

$$X^{-1}[\{x\}] = \bigcap_{i=1}^{n} 1_{A_i}^{-1}[\{x_i\}]. \tag{*}$$

By Remark (a), for $i = 1, \ldots, n$,

$$1_{A_i}^{-1}[\{x_i\}] = \begin{cases} A_1 & \text{if } x_i = 1 \\ A_i^c & \text{if } x_i = 0 \end{cases} \in \mathscr{E}.$$

Hence by (*), $X^{-1}[\{x\}] \in \mathscr{E}$. □

Corollary 11. Let $A_1, \ldots, A_n \subseteq \Omega$. Let

$$\operatorname{atoms}(A_1, \ldots, A_n) = \{A : \varnothing \neq A = \bigcap_{i=1}^{n} B_i \text{ where } B_i \in \{A_i, A_i^c\}, i = 1, \ldots, n\}.$$

Then

(a) The maps $x \mapsto X^{-1}[\{x\}]$ is a bijection from $\mathbf{X}$ onto $\operatorname{atoms}(A_1, \ldots, A_n)$. Hence $|\operatorname{atoms}(A_1, \ldots, A_n)| = |\mathbf{X}| = m = 2^n$. Further, any two distinct atoms are disjoint.

(b) $\operatorname{atoms}(A_1, \ldots, A_n)$ is a partition of Ω.

(c) $\operatorname{field}(A_1, \ldots, A_n) = \{\bigcup \mathscr{A} : \mathscr{A} \subseteq \operatorname{atoms}(A_1, \ldots, A_n)\}$.

Proof. (a) $C \mapsto X^{-1}[C]$ is one-to-one on $\mathcal{P}(\mathbf{X})$ and $x \mapsto \{x\}$ is one-to-one from $\mathbf{X}$ into $\mathcal{P}(\mathbf{X})$. Hence $x \mapsto X^{-1}[\{x\}]$ is one-to-one on $\mathbf{X}$. Clearly $\{X^{-1}[\{x\}] : x \in \mathbf{X}\} = \operatorname{atoms}(A_1, \ldots, A_n)$, since $X^{-1}[\{x\}] = \bigcap_{i=1}^{n} B_i$.

(b) $\{\{x\} : x \in \mathbf{X}\}$ is a partition of $\mathbf{X}$ and $\operatorname{atoms}(A_1, \ldots, A_n) = \{X^{-1}[\{x_i\}] : x \in \mathbf{X}\}$ and $X[\Omega] = \mathbf{X}$, so $\operatorname{atoms}(A_1, \ldots, A_n)$ is a partition of Ω.

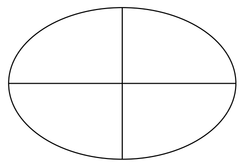
(c) If $B \in \text{field}(A_1, \ldots, A_n)$, then $B = X^{-1}[C]$ for some $C \subseteq \mathbf{X}$, so $B \in \bigcup_{x \in C}\{X^{-1}[\{x\}] = \bigcup \mathscr{A}$, where $\mathscr{A} = \{X^{-1}[\{x\}] : x \in C\}$. $x \mapsto X^{-1}[\{x\}]$ is a bijection from $\mathbf{X}$ onto atoms$(A_1, \ldots, A_n)$, so if $\mathscr{A} \subseteq$ atoms$(A_1, \ldots, A_n)$, then $\mathscr{A} = \{X^{-1}[\{x\}] : x \in C\}$ for some $C \in \mathbf{X}$ and $\bigcup \mathscr{A} = \bigcup_{x \in C} X^{-1}[\{x\}] = X^{-1}[C] \in$ field$(A_1, \ldots, A_n)$. □

1/29/21

Example of previous discussion; $A_1, A_2 \subseteq \Omega$;

Let $A_* = \{A_1A_2, A_1A_2^c, A_1^cA_2, A_1^c, A_2^c\}$

Ω



atoms$(A_1, A_2) = \{A \in A_* : A \neq \varnothing\}$

field$(A_1, A_2) =$ the set of all possible unions of elements of atoms(A_1, A_2).

Next, if $A_1, A_2, A_3, \cdots \subseteq \Omega$, what is field$(A_1, A_2, A_3, \ldots)$?

Lemma 12. Let Φ be an upwards directed set of fields of sets of Ω. Then let $\mathscr{G} = \bigcup \Phi$. Then $\mathscr{G}$ is a field of subsets of Ω.

Proof. To say Φ is upwards directed means for $\mathscr{F}_1, \mathscr{F}_2 \in \Phi$, there exists $\mathscr{F}_3 \in \Phi$ such that $\mathscr{F}_1, \mathscr{F}_2 \subseteq \mathscr{F}_3$. We have $\varnothing, \Omega \in \mathscr{G}$ because there is $\mathscr{F} \in \Phi$ and $\varnothing, \Omega \in \mathscr{F}$.

Let $A \in \mathscr{G}$. Then $A \in \mathscr{F}$ for some $\mathscr{F} \in \Phi$. Now, $\mathscr{F}$ is a field on Ω, so $\Omega \setminus A \in \mathscr{F}$ and hence $\Omega \setminus A \in \mathscr{G}$.

Now let $A_1, A_2 \in \mathscr{G}$. Then for some $\mathscr{F}_1, \mathscr{F}_2 \in \Phi$, we have $A_1 \in \mathscr{F}_1$, $A_2 \in \mathscr{F}_2$.

Since Φ is upwards directed, there exists $\mathscr{F}_3 \in \Phi$ such that $\mathscr{F} \subseteq \mathscr{F}_3$ and $\mathscr{F}_2 \subseteq \mathscr{F}_3$. But $\mathscr{F}_3$ is a field on Ω. Hence, $A_1 \cup A_2 \in \mathscr{F}_3$ and therefore $A_1 \cup A_2 \in \mathscr{G}$.

Thus $\mathscr{G}$ is a field on Ω. □

Proposition 13. Let $(\mathscr{F}_n)_{n \in \mathbb{N}}$ be an increasing sequence of fields on Ω. Let $\mathscr{G} = \bigcup \mathscr{F}$. Then $\mathscr{G}$ is a field on $(\mathscr{F}_n)_{n \in \mathbb{N}}$.

Proof. Apply the lemma with $\Phi = \{\mathscr{F}_n : n \in \mathbb{N}\}$. (Let $n_1, n_2 \in \mathbb{N}$. Let $n_3 = \max\{n_1, n_2\}$. Then $\mathscr{F}_{n_1} \subseteq \mathscr{F}_{n_2}$ and $\mathscr{F}_{n_2} \subseteq \mathscr{F}_{n_3}$.) □

Corollary 14. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of subsets of Ω. Then

$$\text{field}(A_1, A_2, A_3 \ldots) = \bigcup_{n \in \mathbb{N}} \text{field}(A_1, \ldots, A_n).$$

Proof. For each $n \in \mathbb{N}$, field$(A_1, \ldots, A_{n+1})$ contains $A_1, \ldots, A_n$, so

$$\text{field}(A_1, \ldots, A_n) \subseteq \text{field}(A_1, \ldots, A_n, A_{n+1}).$$

Hence $\bigcup_{n \in \mathbb{N}} \text{field}(A_1, \ldots, A_n)$ is a field on Ω that contains $A_1, A_2, A_3, \ldots$ and any field on Ω that contains $A_1, A_2, A_3, \ldots$ must contain $\bigcup_{n \in \mathbb{N}} \text{field}(A_1, \ldots, A_n)$. □

Definition. A field is *complete* if it is closed under arbitrary unions and intersections.

Theorem 15. Let $\mathscr{G} \subseteq \mathcal{P}(\Omega)$. Define $X : \Omega \to \{0,1\}^{\mathscr{G}}$ by $X(\omega) = (1_G(\omega))_{G\in\mathscr{G}}$. Let $\mathscr{B} = \{B \subseteq \Omega : 1_B \text{ expressible in terms of } X\}$. Let $\mathbf{X} = X[\Omega]$. Then

(a) $C \mapsto X^{-1}[C]$ is a bijection from $\mathcal{P}(\mathbf{X})$ onto $\mathscr{B}$.

(b) $|\mathscr{B}| = |\mathcal{P}(\mathbf{X})|$

(c) $\mathscr{B}$ is the smallest complete field on Ω containing $\mathscr{G}$

Remark. (a) This is just proved as in the case where $\mathscr{G}$ is finite.

(b) Since a finite field is a complete field, this theorem includes the earlier one as a special case.

Example. Suppose $\Omega = \mathbb{Z}$ and $\mathscr{G} = \{\{n\} : n \in \mathbb{Z}\}$. Then field$(\mathscr{G}) = \{A \subseteq \mathbb{Z} : A \text{ is finite or } \mathbb{Z} \setminus A \text{ is finite }\}$.

The complete field on $\mathbb{Z}$ generated by $\mathscr{G}$ is $\mathcal{P}(\mathbb{Z})$

Definition. To say $\mathscr{H}$ is a π-system on Ω means $\mathscr{H}$ is a collection of subsets of Ω and for all $A, B \in \mathscr{H}$ we have $A \cap B \in \mathscr{H}$.

Proposition 16. Let $\mathscr{E} \subseteq \mathcal{P}(\Omega)$. Let $\mathscr{H} = \{\cap\mathscr{E}_0 : \varnothing \neq \mathscr{E}_0 \text{ finite } \subseteq \mathscr{E}\}$. Then $\pi(\mathscr{E}) = \mathscr{H}$ is the smallest π-system containing $\mathscr{E}$.

Proof. First, let's show $\mathscr{H}$ is a π-system. Let $A_1, A_2 \in \mathscr{H}$. Then $A_1 = \cap\mathscr{E}_1$, $A_2 = \mathscr{E}_2$ for some nonempty finite $\mathscr{E}_1, \mathscr{E}_2 \subseteq \mathscr{E}$. Then $A_1 \cap A_2 = \cap\mathscr{E}_3$, where $\mathscr{E}_3 = \mathscr{E}_1 \cap \mathscr{E}_2$ (For each ω we have $\omega \in A_2 \cap A_2$ iff $\omega \in A_1$ and $\omega \in A_2$ iff $\omega \in E$ for each $E \in \mathscr{E}_1$ and $\omega \in E$ for each $E \in \mathscr{E}_2$ iff $\omega \in E$ for each $E \in \mathscr{E}_1 \cup \mathscr{E}_2$.) Now $\mathscr{E} \neq \varnothing$ since $\mathscr{E}_1 \neq \mathscr{E}_1$. Also, $\mathscr{E}_3$ is finite, since $\mathscr{E}_1, \mathscr{E}_2$ are both finite. Thus $A_1 \cap A_2 \in \mathscr{H}$, therefore $\mathscr{H}$ is a π-system.

Suppose $\mathscr{H}'$ is a π-system that contains $\mathscr{E}$. We wish to show $\mathscr{H} \subset \mathscr{H}'$. Let $A \in \mathscr{H}$. By the definition of $\mathscr{H}$, $A = \cap\mathscr{E}_0$ for some nonempty finite $\mathscr{E}_0 \subseteq \mathscr{E}$. Since $\mathscr{E} \subseteq \mathscr{H}'$ and $\mathscr{H}'$ is a π-system, $\cap\mathscr{E}_0 \in \mathscr{H}'$ $\square$

2/1/21

Proposition 17. Let $(\mathscr{E}_i)_{i\in I}$ be a family of π-systems. Then

$$\pi\left(\bigcup_{i\in I}\mathscr{E}_i\right) = \left\{\bigcap_{i\in I_0} E_i : \varnothing \neq I_0 \text{ finite } \subseteq I \text{ and } E_i \in \mathscr{E}_i \text{ for } i \in I_0\right\}$$

Proof. Clearly

$$\pi\left(\bigcup_{i\in I}\mathscr{E}_i\right) \supseteq \left\{\bigcap_{i\in I_0} E_i : \varnothing \neq I_0 \text{ finite } \subseteq I \text{ and } E_i \in \mathscr{E}_i \text{ for } i \in I_0\right\}.$$

Hence it suffices to show that the right hand side is in fact a π-system. Let E, F be elements of the right hand side. Now, $E = \bigcap_{i\in I_1} E_i$ for some nonempty finite $I_1 \subseteq I$ and some $E_i \in \mathscr{E}_i, i \in I_1$. Likewise, $F = \bigcap_{i\in I_2} F_i$ for some nonempty finite $I_2 \subseteq I$ and some $F_i \in \mathscr{E}_i, i \in I_2$. Let $I_3 = I_1 \cup I_2$. For $i \in I_3$, let

$$G_i = \begin{cases} E_i & \text{if } i \in I_1 \setminus I_2, \\ F_i & \text{if } i \in I_2 \setminus I_1 \\ E_i \cap F_i & \text{if } i \in I_1 \cap I_2 \end{cases}$$

Then $G_i \in \mathscr{E}_i$ for each $i \in I_3$, since $\mathscr{E}_i$ is a π-system. Then $E \cap F = \bigcap_{i\in I_3} G_i$ is an element of the right hand side. $\square$

Pokemon theory (gotta collect 'em all). If $\mathscr{A} = \{A_i : i \in I\}$, where $I \neq \varnothing$, then $\bigcap\mathscr{A} = \bigcap_{i\in I} A_i$. Any $\mathscr{B}$ is of the form $\{B_j : j \in J\}$ for some $(B_j)_{j\in I'}$. Just let $J = \mathscr{B}$ and let $B_j = j$ for each $j \in J$.

Definition. Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $(\mathscr{E}_i)_{i\in I}$ be a family of π-systems such that for each $i \in I$, $\mathscr{E}_i \subseteq \mathscr{F}$. To say that $(\mathscr{E}_i)_{i\in I}$ is independent means that for each nonempty finite $I_0 \subseteq I$ and for all $E_i \in \mathscr{E}_i$, $i \in I_0$, we have $P(\bigcap_{i\in I_0} E_i) = \prod_{i\in I_0} P(E_i)$.

Example. A family $(A_i)_{i\in I}$ is independent if and only if the family $(\{A_i\})_{i\in I}$ is independent.

Theorem 18 (The Grouping Theorem, Version 1.)**.** Let $(\mathscr{G}_i)_{i\in I}$ be an independent family of π-systems. Let $(I_j)_{j\in J}$ be a disjoint family of nonempty subsets of I. Let

$$\mathscr{H}_j = \pi\left(\bigcup_{i\in I_j} \mathscr{G}_i\right)$$

Then the family of π-systems $(\mathscr{H}_j)_{j\in J}$ is independent.

Proof. Let $\varnothing \neq J_0$ finite $\subseteq J$, and let $H_j \in \mathscr{H}_j$ for each $j \in J_0$. We wish to show that $P\left(\bigcap_{j\in J_0} H_j\right) = \prod_{j\in J_0} P(H_j)$. Since each $\mathscr{G}_i$ is a π-system and $\mathscr{H}_j = \pi\left(\bigcup_{i\in I_j}\mathscr{G}_i\right)$ for each $j \in J_0$, there exists a nonempty finite set $I_j^0 \subseteq I_j$ and sets $G_i \in \mathscr{G}_i, i \in I_j^0$ such that $H_j = \bigcap_{i\in I_j^0} G_i$. Then

$$P\left(\bigcap_{j\in J_0} H_j\right) = P\left(\bigcap_{j\in J_0}\bigcap_{i\in I_j^0} G_i\right) = P\left(\bigcap_{i\in I^0} G_i\right)$$

where $I^0 = \bigcup_{j\in J_0} I_j^0$. By independence of $(\mathscr{G}_i)_{i\in I}$, this is

$$= \prod_{i\in I^0} P(G_i) = \prod_{j\in J_0}\prod_{i\in I_j^0} P(G_i) = \prod_{j\in J_0} P\left(\bigcap_{i\in I_j^0} G_i\right) = \prod_{j\in J_0} P(H_j)$$

as desired. □

Let X be a set. For each $A \subseteq X$, let $A^c = X \setminus A$

Definition. To say $\mathscr{B}$ is a λ-system on X means $\mathscr{B} \subseteq \mathcal{P}(X)$ such that

(a) $X \in \mathscr{B}$

(b) For all $B \in \mathscr{B},\ B^c \in \mathscr{B}$

(c) For each disjoint $(B_n) \in \mathscr{B}^{\mathbb{N}}$, $\dot\bigcup_{n\in\mathbb{N}} B_n \in \mathscr{B}$

Example. Let μ and ν be measures on σ-field $\mathscr{A}$ on X and let $\mathscr{B} = \{B \in \mathscr{A} : \mu(B) = \nu(B)\}$. Then $\mathscr{B}$ need not be closed under countable unions, but it is closed under countable disjoint unions. If in addition $\mu(X) = \nu(X) < \infty$, then $\mathscr{B}$ is a λ system on X.

Definition (belated)**.** Let $\mathscr{A}$ be a σ-field on X. To say μ is a measure on $\mathscr{A}$ means $\mu : \mathscr{A} \to [0,\infty]$, such that;

(a) $\mu(\varnothing) = 0$

(b) For each disjoint sequence $(A_n) \in \mathscr{A}^{\mathbb{N}}$, $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$

Some properties of measures. Let μ be a measure on a σ-field $\mathscr{A}$ on X.

(a) For each $n \in \mathbb{N}$, for all disjoint $A_1, \ldots, A_n \in \mathscr{A}$, $\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k)$

Proof. Let $A_k = \varnothing$ for $k > n$. □

(b) For all $A, B \in \mathscr{A}$, if $A \subseteq B$, then $\mu(A) + \mu(B \setminus A) = \mu(B)$, so $\mu(A) \leq \mu(B)$. If in addition $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

Proof. $A \sqcup (B \setminus A) = B$. □

(c) For all $A, A_1, A_2, \dots \in \mathscr{A}$, if $A \subseteq \bigcup_n A_n$, then $\mu(A) \leq \sum_n \mu(A_n)$

Proof. By (b), it suffices to consider the case where $A = \bigcup_n A_n$. Let $E_n = A_n \setminus \bigcup_{m<n} A_m$. Then $E_1, E_2, \ldots$ are disjoint with $A = \dot\bigcup_n E_n$. Hence $\mu(A) = \sum_n \mu(E_n) \leq \sum_n \mu(A_n)$, again by (b) since $E_n \subseteq A_n$. □

(d) For each increasing sequence $(A_n) \in \mathscr{A}^{\mathbb{N}}$, if $A = \bigcup_u A_n$ then $\mu(A_n) \xrightarrow[n\to\infty]{\text{increases}} \mu(A)$.

Proof. Let $B_1 = A_1$, for $n \geq 2$ let $B_n = A_n \setminus A_{n-1}$. Then $B_1, B_2, \dots$ are disjoint, $\bigcup_{k\in\mathbb{N}} B_k = A$, and for each n $\bigcup_{k=1}^{n} B_k = A_n$. Hence $\mu(A_n) = \sum_{k=1}^{n} \mu(B_k)$ increases to $\sum_{k\in\mathbb{N}} \mu(B_k) = \mu(A)$. $\square$

(e) For each decreasing sequence $(A_n) \in \mathscr{A}^{\mathbb{N}}$, if $\mu(A_1) < \infty$, then $\mu(A_n) \xrightarrow[n\to\infty]{\text{decreases}} \mu(A)$ where $A = \bigcap_n A_n$.

Proof. Let $B_n = A_1 \setminus A_n$. Then $B_1 \subseteq B_2 \subseteq \cdots$ with $B = \bigcup_n B_n A_1 \setminus \bigcap_n A_n$. By (d), $\mu(B_n)$ increases to $\mu(B)$. Hence $\mu(A_1) - \mu(B_n) = \mu(A_1 \setminus B_n) = \mu(A_n)$ decreases to $\mu(A_1) - \mu(B) = \mu(A_1 \setminus B) = \mu(A)$. $\square$

Remark. We need the assumption that $\mu(A_1) < \infty$. For example, define μ on $\mathcal{P}(\mathbb{N})$ by

$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ \infty & \text{if } A \text{ is infinite} \end{cases}$$

and let $A_n = \{n, n+1, \dots\}$. Then $\mu(A_n) = \infty$ for each n with $A_1 \supseteq A_2 \supseteq \cdots$ and $\bigcap_n A_n = \varnothing$. Hence $\mu(A_n) \neq \mu(\bigcap_n A_n)$.

2/3/21

Finite(-ish) λ-systems. Let X be a set.

Definition. To say $\mathscr{B}$ is a $\boldsymbol{\lambda_0}$-system on $\boldsymbol{X}$ means that $\mathscr{B} \subseteq \mathcal{P}(X)$ and

(a) $X \subseteq \mathscr{B}$

(b) For each $B \in \mathscr{B}$, $B^c \in \mathscr{B}$

(c) For all disjoint $A, B \in \mathscr{B}$, $A \sqcup B \in \mathscr{B}$

Proposition 19. Let $\mathscr{B} \subseteq \mathcal{P}(X)$. The following are equivalent:

(a) $\mathscr{B}$ is a λ_0-system on X

(b) $X \in \mathscr{B}$ and for all $A, B \in \mathscr{B}$, if $A \subseteq B$ then $B \setminus A \in \mathscr{B}$.

Proof. (a) $\implies$ (b). Supposeing (a), we have $X \in \mathscr{B}$. Let $A, b \in \mathscr{B}$ and suppose $A \subseteq B$. Then A and B^c are disjoint elements of $\mathscr{B}$, so $A \cup B^c \in \mathscr{B}$, so $(A \cup B^c)^c \in \mathscr{B}$. But $(A \cup B^c)^c = A^c \cap B^{cc} = A^c \cap B = B \setminus A$. Thus $B \setminus A \in \mathscr{B}$ and hence (a) implies (b).

(b) $\implies$ (a). We have $X \in \mathscr{B}$ by assumption. Let $B \in \mathscr{B}$. Then $B, X \in \mathscr{B}$ and $B \subseteq X$, so $B^c = X \setminus B \in \mathscr{B}$. Let $A, B \in \mathscr{B}$ with $A \cap B = \varnothing$. We want to show that $A \cup B \in \mathscr{B}$. Since $A \cap B = \varnothing$, $A \subseteq B^c = X \setminus B$, and of course $B^c \in \mathscr{B}$, so $B^c \setminus A \in \mathscr{B}$, so $(B^c \setminus A)^c = (B^c \cap A^c)^c = B^{cc} \cup A^{cc} = A \cup B \in \mathscr{B}$, as desired. This shows that (b) implies (a). $\square$

Proposition 20. Let $\mathscr{B}$ be a λ_0-system on X and let $A, B \in \mathscr{B}$. Then the following are equivalent:

(a) $A \cap B \in \mathscr{B}$

(b) $A \setminus B \in \mathscr{B}$

(c) $A \cup B \in \mathscr{B}$

(d) $B \setminus A \in \mathscr{B}$

Proof. Suppose $A \cap B \in \mathscr{B}$. Now $A \cap B \subseteq A$, which by the previous proposition means $A \setminus (A \cap B) \in \mathscr{B}$. But $A \setminus (A \cap B) = A \setminus B$, hence $A \setminus B \in \mathscr{B}$ so (a) implies (b).

Now suppose that $A \setminus B \in \mathscr{B}$. Then $A \setminus B$ and B are disjoint elements of $\mathscr{B}$, so $(A \setminus B) \cup B = A \cup B \in \mathscr{B}$. Hence (b) implies (c).

Now suppose $A \cup B \in \mathscr{B}$. Now $A \subseteq A \cup B$ so $(A \cup B) \setminus A = B \setminus A \in \mathscr{B}$ and thus (c) implies (d).

Finally, suppose $B \setminus A \in \mathscr{B}$. Now $B \setminus A \subset B$, so $B \setminus (B \setminus A) = A \cap B \in \mathscr{B}$, so (d) implies (a). $\square$

Lemma 21. Let $\mathscr{B}$ be a λ_0-system on X. Let $A_1, \dots, A_n \subseteq X$ such that $\pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. Let $k \in \{1, \dots, n\}$ and let $B_k = A_k^c$. For $j \neq k$, let $B_j = A_j$. Then $\pi(B_1, \dots, B_n) \subseteq \mathscr{B}$.

Proof. Let $C \in \pi(B_1, \dots, B_n)$. Then $C = \bigcap_{j \in J} B_j$ for some nonempty $J \subseteq \{1, \dots, n\}$. We wish to show that $C \in \mathscr{B}$. If $k \notin J$, then clearly this holds because then $C = \bigcap_{j \in J} A_j \in \pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. Suppose $k \in J$. If $J = \{k\}$ then $C = A_k^c \in \mathscr{B}$. Suppose $J \neq \{k\}$. Then let $J' = J \setminus \{k\}$. Then $\varnothing \neq J' \subseteq \{1, \dots, n\}$. Let $D = \bigcap_{j \in J'} A_j$. Then $D \in \pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. Now, $D \cap A_k \in= \bigcap_{j \in J} A_j \in \pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. Hence by the previous proposition, $B \setminus A_k \in \mathscr{B}$. But $D \setminus A_k = D \cap A_k^c = D \cap B_k = \bigcap_{j \in J} B_j = C$, so $C \in \mathscr{B}$. This holds for each such $C \in \pi(B_1, \dots, B_n)$, hence $\pi(B_1, \dots, B_n) \subseteq \mathscr{B}$. □

Proposition 22. Let $A_1, \dots, A_n \subseteq X$. Suppose $\pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. For $k = 1, \dots, n$, let $B_k \in \{A_k, A_k^c\}$. Then $\pi(B_1, \dots, B_n) \subseteq \mathscr{B}$.

Proof. Let m be the number of $j \in \{1, \dots, n\}$ such that $B_j = A_j^c$. Then the stated result follows by m applications of the lemma. □

Proposition 23. Let $A_1, \dots, A_n \subseteq X$. Let $\mathscr{B}$ be a λ_0-system on X such that $\pi(A_1, \dots, A_n) \subseteq \mathscr{B}$. Then:

(a) $\text{atoms}(A_1, \dots, A_n) \subseteq \mathscr{B}$

(b) $\text{field}(A_1, \dots, A_n) \subseteq \mathscr{B}$

Proof. Let $A \in \text{atoms}(A_1, \dots, A_n)$. Then there exist $B_j \in \{A_j, A_j^c\}$ for $j = 1, \dots, n$ such that $\varnothing \neq A = \bigcap_{j=1}^n B_j$. Then $A \in \pi(B_1, \dots, B_n) \subseteq \mathscr{B}$ by the previous proposition, establishing (a).

We know that $\text{atoms}(A_1, \dots, A_n)$ is a finite (at most cardinality 2^n) partition of X and $\text{field}(A_1, \dots, A_n)$ is the set of all possible (disjoint) unions of the atoms, completing the proof. □

Theorem 24 (π-λ_0 Theorem)**.** Let $\mathscr{H}$ be a π-system on X and let $\mathscr{B}$ be a λ_0-system on X such that $\mathscr{H} \subseteq \mathscr{B}$. Then $\text{field}(\mathscr{H}) \subseteq \mathscr{B}$.

Proof. Let Φ be the set of finite subsets of $\mathscr{H}$. By the previous proposition, for each $\mathscr{H}_0 \in \Phi$, $\text{field}(\mathscr{H}_0) \subseteq \mathscr{B}$. ($\pi(\mathscr{H}_0) \subseteq \pi(\mathscr{H} \subseteq \mathscr{B}$ so $\text{field}(\pi(\mathscr{H}_0)) \subseteq \mathscr{B}$.) But $\text{field}(\pi(\mathscr{H}_0)) = \text{field}(\mathscr{H}_0)$. This holds for each $\mathscr{H}_0 \in \Phi$.

But recall that $\text{field}(\mathscr{H}) = \bigcup\{\text{field}(\mathscr{H}_0) : \mathscr{H}_0 \text{ finite } \subseteq \mathscr{H}\}$, so $\text{field}(\mathscr{H}) \subseteq \mathscr{B}$. □

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

Lemma 25. Let $(\mathscr{H}_i)_{i \in I}$ be an independent family of π-systems. For each $i \in I$, let $\mathscr{F}_i = \text{field}(\mathscr{H}_i)$. Then $(\mathscr{F}_i)_{i \in I}$ is independent.

Proof. Let I_0 be a nonempty finite subset of I and let $F_i \in \mathscr{F}_i$ for $i \in I_0$. Let $F = \bigcap_{i \in I_0} F_i$. we want to show $P(F) = \prod_{i \in I_0} P(F_i)$. If $I_0 = 1$, then we are done.

??? □

2/5/21

Definition. To say $\mathscr{M} \subseteq \mathcal{P}(X)$ is a <u>monotone class</u> on X means that each $(E_n) \in \mathscr{M}^{\mathbb{N}}$,

(a) If $E_n \subseteq E_{n+1}$ for each n, then $\bigcup_n E_n \in \mathscr{M}$.

(b) If $E_n \supseteq E_{n+1}$ for each n, then $\bigcap_n E_n \in \mathscr{M}$.

Proposition 26. Let $\mathscr{E} \subseteq \mathcal{P}(X)$. Then there is a smallest monotone class on X containing $\mathscr{E}$.

Proof. Let $\mathscr{M}$ be the set of all $A \subseteq X$ such that $A \in \mathscr{N}$ for each monotone class $\mathscr{N}$ on X such that $\mathscr{E} \subseteq \mathscr{N}$. Then for each monotone class $\mathscr{N}$ on X such that $\mathscr{E} \subseteq \mathscr{N}$, we have $\mathscr{E} \subseteq \mathscr{M} \subseteq \mathscr{N}$, by definition of $\mathscr{M}$. It remains to show that $\mathscr{M}$ is a monotone class. Let $(E_n) \in \mathscr{M}^{\mathbb{N}}$;

(a) Suppose $E_n \subseteq E_{n+1}$ for each n. Let $E = \bigcup_n E_n$. We wish to show that $E \in \mathscr{M}$. Let $\mathscr{N}$ be a monotone class on X such that $\mathscr{E} \subseteq \mathscr{N}$. Each $E_n \in \mathscr{N}$. Hence $\bigcup_n E_n \in \mathscr{N}$ since $\mathscr{N}$ is a monotone class; in other words $E \in \mathscr{N}$. As this holds for arbitrary such $\mathscr{N}$, then $E \in \mathscr{M}$.

(b) Now suppose $E_n \supseteq E_{n+1}$ for each n. Then by a similar argument to (a) we have $\bigcap_n E_n \in \mathscr{M}$.

Thus $\mathscr{M}$ is a monotone class. $\square$

Proposition 27. Let $\mathscr{B}$ be a λ-system on X. Then $\mathscr{B}$ is a monotone class on X.

Proof. Let $(E_n) \in \mathscr{B}^{\mathbb{N}}$ such that $E_n \subseteq E_{n+1}$ for all n. We want to show $\bigcup_n E_n \in \mathscr{B}$. Let $B_1 = E_1,\ B_2 = E_2 \setminus E_1,\ B_3 = E_3 \setminus E_2$, and so on. Then $B_1,\ B_2,\ B_3, \dots$ are disjoint and $\bigcup_n B_n = E$. Note that each $B_n \in \mathscr{B}$ because $E_{n-1} \subseteq E_n$ for each $n \geq 2$. Thus E is a countable disjoint union of elements of $\mathscr{B}$ and hence an element of $\mathscr{B}$.

Now let $(F_n) \in \mathscr{B}^{\mathbb{N}}$ such that $F_n \supseteq F_{n+1}$ for all n. Let $E_n = F_n^c$. Then $E_n \in \mathscr{B}$ and $E_n \subseteq E_{n+1}$, so by the previous argument $\bigcup_n E_n \in \mathscr{B}$. Hence $\bigcap_n F_n = \bigcap_n E_n^c = (\bigcup_n E_n)^c \in \mathscr{B}$. $\square$

Remark. We see that every λ-system is a monotone class, and every λ_0-system is a λ-system iff it is a monotone class.

Lemma 28. Let $\mathscr{F} \subseteq \mathcal{P}(X)$. Let $\mathscr{M}$ be the monotone class generated by $\mathscr{F}$. Then;

(a) $\bigcup_n E_n \in \mathscr{M}$ for each $(E_n) \in \mathscr{M}^{\mathbb{N}}$ iff $A \cup B \in \mathscr{M}$ for all $A, B \in \mathscr{F}$.

(b) $\bigcap_n E_n \in \mathscr{M}$ for each $(E_n) \in \mathscr{M}^{\mathbb{N}}$ iff $A \cap B \in \mathscr{M}$ for all $A, B \in \mathscr{F}$.

(c) $X \setminus E \in \mathscr{M}$ for each $E \in \mathscr{M}$ iff $X \setminus A \in \mathscr{M}$ for each $A \in \mathscr{F}$.

Proof. (a) ($\Longrightarrow$) is obvious since $\mathscr{F} \subseteq \mathscr{M}$; take $E_1 = A, A_n = B$ for $n \geq 2$.

($\Longleftarrow$) Suppose $A \cup B \in \mathscr{M}$ for all $A, B \in \mathscr{F}$. Let $A \in \mathscr{F}$, let $\mathscr{B} = \{B \subseteq X : A \cup B \in \mathscr{M}\}$. Then $\mathscr{F} \subseteq \mathscr{B}$. Let $(B_n) \in \mathscr{B}^{\mathbb{N}}$.

(i) Suppose $B_n \subseteq B_{n+1}$ for all n. Then $A \cup B_n \in \mathscr{M}$ and $A \cup B_n \subseteq A \cup B_{n+1}$ for all n. Hence $\bigcup_n (A \cup B_n) \in \mathscr{M}$. But $\bigcup_n (A \cup B_n) = A \cup (\bigcup_n B_n) \in \mathscr{B}$ by the definition of $\mathscr{B}$.

(ii) Now suppose $B_n \supseteq B_{n+1}$ for all n. Then $A \cup B_n \in \mathscr{M}$ and $A \cup B_n \supseteq A \cup B_{n+1}$ for all n. Hence $\bigcap_n (A \cup B_n) \in \mathscr{M}$ since $\mathscr{M}$ is a monotone class. But $\bigcup_n (A \cup B_n) = A \cup (\bigcap_n B_n) \in \mathscr{M}$. Thus $\bigcap_n B_n \in \mathscr{B}$ by the definition of $\mathscr{B}$.

We conclude that $\mathscr{B}$ is a monotone class containing $\mathscr{F}$. But $\mathscr{M}$ is the smallest such monotone class, so $\mathscr{M} \subseteq \mathscr{B}$ so for each $B \in \mathscr{M}$, $A \cup B \in \mathscr{M}$. This holds for each $A \in \mathscr{F}$.

Let $B \in \mathscr{M}$. Let $\mathscr{A} = \{A \subseteq X : A \cup B \in \mathscr{M}\}$. Then $\mathscr{F} \subseteq \mathscr{A}$. Also, $\mathscr{A}$ is a monotone class by a similar argument, so $\mathscr{M} \subseteq \mathscr{A}$ and hence for all $A, B \in \mathscr{M}$, $A \cup B \in \mathscr{M}$.

Let $(E_n) \in \mathscr{M}^{\mathbb{N}}$. Let $A_1 = E_1$, $A_2 = E_1 \cup E_2$, $A_3 = E_1 \cup E_2 \cup E_3$ and so on. Then each $A_n \in \mathscr{M}$ and $A_n \subseteq A_{n+1}$ for all n. Hence $\bigcup_n E_n = \bigcup_n A_n \in \mathscr{M}$, since $\mathscr{M}$ is a monotone class.

(b) Similar to (a).

(c) Simpler than (a).

$\square$

Theorem 29 (Monotone Class Theorem)**.** Let $\mathscr{F}$ be a field on X and let $\mathscr{M}$ be a monotone class on X generated by $\mathscr{F}$. Then $\mathscr{M}$ is a σ-field on X. Indeed, $\mathscr{M}$ is the smallest σ-field on X containing $\mathscr{F}$.

Proof. That $\mathscr{M}$ is a σ-field on X follows immediately from the lemma. Let $\mathscr{N}$ be a σ-field on X with $\mathscr{F} \subseteq \mathscr{N}$. Then $\mathscr{N}$ is a monotone class on X, so $\mathscr{M} \subseteq \mathscr{N}$. $\square$

Remark. For any π-system $\mathscr{H}$ on X, there exists a σ-field generated by $\mathscr{H}$, which is the smallest σ-field containing $\mathscr{H}$.

Theorem 30 (The π-λ Theorem)**.** Let $\mathscr{H}$ be a π-system on a set X, let $\mathscr{A}$ be the σ-field on X generated by $\mathscr{H}$, and let $\mathscr{B}$ be a λ-system on X containing $\mathscr{H}$. Then $\mathscr{A} \subseteq \mathscr{B}$

Proof. Let $\mathscr{M}$ be the monotone class on X generated by field($\mathscr{H}$). By the monotone class theorem $\mathscr{M}$ is a σ-field on X. Thus $\mathscr{A} \subseteq \mathscr{M}$. (In fact, $\mathscr{A} = \mathscr{M}$.)

The π-system $\mathscr{H}$ is contained in the λ_0-system $\mathscr{B}$ (since any λ-system is a λ_0-system) so by the $\pi-\lambda_0$ theorem, field($\mathscr{H}$) $\subseteq \mathscr{B}$. But $\mathscr{B}$ is a monotone class, since it is a λ-system, so $\mathscr{M} \subseteq \mathscr{B}$. Thus $\mathscr{A} \subseteq \mathscr{M} \subseteq \mathscr{B}$. □

2/8/21

Theorem 31. Let $(X, \mathscr{A})$ be a measurable space and let $\mathscr{H}$ be a π-system on X which generates the σ-field $\mathscr{A}$. Let μ and ν be finite measures on $\mathscr{A}$ such that $\mu(X) = \nu(X)$ and for each $H \in \mathscr{H}$, $\mu(H) = \nu(H)$. That $\mu = \nu$.

Proof. Let $\mathscr{B} = \{B \in \mathscr{A} : \mu(B) = \nu(B)\}$. Since μ, ν are measures on $\mathscr{A}$ and $\mu(X) = \nu(X)$, $\mathscr{B}$ is a λ-system on X. Now let $\mathscr{H} \subseteq \mathscr{B}$, because $\mu(H) = \nu(H)$ for all $H \in \mathscr{H}$. Since $\mathscr{H}$ is a π-system on X, and $\mathscr{B}$ is a λ-system containing $\mathscr{A} = \sigma(\mathscr{H})$ (notationally, this means the σ-field generated by $\mathscr{H}$), we have $\sigma(\mathscr{H}) = \mathscr{A} \subseteq \mathscr{B}$ by the $\pi - \lambda$ theorem and thus $\mathscr{A} = \mathscr{B}$. Hence, $\mu = \nu$. □

Corollary 32. Let $(X, \mathscr{A})$ be a measurable space. Let $\mathscr{H}$ be a π-system on X such that $\sigma(\mathscr{H}) = \mathscr{A}$. Suppose $(H_n) \in \mathscr{H}^{\mathbb{N}}$ is a sequence of elements such that $\bigcup_n H_n = X$. Let μ, ν be measures on $\mathscr{A}$ such that $\mu(H) = \nu(H)$ $\forall H \in \mathscr{H}$ and assume that $\mu(H_n) = \nu(H_n) < \infty$ for all n. Then $\mu = \nu$.

Proof. For each n, define μ_n, ν_n on $\mathscr{A}$ by $\mu_n(A) = \mu(A \cap H_n)$ and $\nu_n(A) = \nu(A \cap H_n)$ and notice that μ_n and ν_n are finite measures on $\mathscr{A}$ satisfying $\mu_n(X) = \nu_n(X) < \infty$ and $\mu_n(H) := \mu(\underbrace{(H \cap H_n)}_{\in \mathscr{H}}) = \nu(\underbrace{(H \cap H_n)}_{\in \mathscr{H}}) = \nu_n(H)$, thus $\mu_n = \nu_n$ by the previous theorem.

Let $A_n = A \cap H_n$. Then $A = A \cap X = A \cap (\bigcup_n H_n) = \bigcup_n (A \cap H_n) = \bigcup_n (A_n)$. Let $B_n = A_n \setminus \bigcup_{m<n} A_n$. Then (B_n) is a disjoint sequence in $\mathscr{A}$ and $A = \dot{\bigcup}_n B_n$. Now $\mu(B_n) = \mu_n(B_n) = \nu_n(B_n) = \nu(B_n)$.

Hence $\mu(A) = \sum_n \mu(B_n) = \sum_n \nu(B_n) = \nu(A)$. This holds for each $A \in \mathscr{A}$. Hence $\mu = \nu$. □

Remark. If $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ and $b = (b_1, \dots, b_d) \in \mathbb{R}^d$, then $a \leq b$ means $a_k \leq b_k$ for each $k = 1, \dots, d$ and $a < b$ means that $a_k < b_k$ for $k = 1, \dots, d$.

Also $(a, b] = \{x \in \mathbb{R}^d : a < x \leq b\} = \prod_{k=1}^d (a_k, b_k]$.

Corollary 33. Let μ, ν be Borel measures on $\mathbb{R}^d$. Suppose $\mu((a, b]) = \nu((a, b]) < \infty$ for all $a, b \in \mathbb{R}^d$ with $a \leq b$. Then $\mu = \nu$.

Proof. To say μ, ν are Borel measures on $\mathbb{R}^d$ means they are measures on $\mathscr{A} = \sigma(\mathscr{G})$, where $\mathscr{G}$ is the set of open subsets of $\mathbb{R}^d$. Let $\mathscr{H} = \{(a, b] : a, b \in \mathbb{R}^d, a \leq b\}$. Then $\mathscr{H}$ is a π-system.

$\sigma(\mathscr{H}) = \sigma(\mathscr{G})$. Also, for each $n \in \mathbb{N}$, $(-n, n]^d \in \mathscr{H}$ and $\bigcup_n (-n, n]^d = \mathbb{R}^d$. Hence $\mu = \nu$ be the previous corollary. □

Proposition 34. Let $\mathscr{H} \subseteq \mathcal{P}(X)$. Then there is a smallest σ-field $\mathscr{A}$ on X containing $\mathscr{H}$.

Proof. Let $\mathscr{A} = \{A \subseteq X : A \in \mathscr{B}$ for each σ-field $\mathscr{B}$ on X with $\mathscr{H} \subseteq \mathscr{B}\}$. Then $\mathscr{A}$ is a σ-field on X, $\mathscr{H} \subseteq \mathscr{A}$, and for each σ-field $\mathscr{B}$ on X such that $\mathscr{H} \subseteq \mathscr{B}$, we have $\mathscr{A} \subseteq \mathscr{B}$. □

Remark. Let $\mathscr{H} \subseteq \mathcal{P}(X)$. Then $\sigma(\mathscr{H})$ denotes the smallest σ-field $\mathscr{A}$ on X such that $\mathscr{H} \subseteq \mathscr{A}$.

Warning: $\sigma(\mathscr{H})$ depends on X as well as on $\mathscr{H}$.

Proposition 35. Let $\mathscr{G}, \mathscr{H} \subseteq \mathcal{P}(X)$. Then $\sigma(\mathscr{G}) \subseteq \sigma(\mathscr{H})$ if and only if $\mathscr{G} \subseteq \sigma(\mathscr{H})$.

Proof. ($\Longrightarrow$) $\mathscr{G} \subseteq \sigma(\mathscr{G})$.

($\Longleftarrow$) Suppose $\mathscr{G} \subseteq \sigma(\mathscr{H})$. Then $\sigma(\mathscr{H})$ is a σ-field on X such that $\mathscr{G} \subseteq \sigma(\mathscr{H})$. But $\sigma(\mathscr{G})$ is the smallest such σ-field. Hence $\sigma(\mathscr{G}) \subseteq \sigma(\mathscr{H})$. □

Corollary 36. Let $\mathscr{G}, \mathscr{H} \subseteq \mathcal{P}(X)$. Then $\sigma(\mathscr{G}) = \sigma(\mathscr{H})$ if and only if $\mathscr{G} \subseteq \sigma(\mathscr{H})$ and $\mathscr{H} \subseteq \sigma(\mathscr{G})$.

Example. Let $\mathscr{A}$ be the Borel σ-field on $\mathbb{R}^d$ and let $\mathscr{H} = \{(a,b] : a, b \in \mathbb{R}^d, a \leq b\}$. Then $\mathscr{A} = \sigma(\mathscr{H})$.

Proof. By definition, $\mathscr{A} = \sigma(\mathscr{G})$, where $\mathscr{G}$ is the set of open subsets on $\mathbb{R}^d$. By the preceding corollary, to show that $\sigma(\mathscr{G}) = \sigma(\mathscr{H})$, it suffices to show that $\mathscr{G} \subseteq \sigma(\mathscr{H})$ and $\mathscr{H} \subseteq \sigma(\mathscr{G})$.

Let $G \in \mathscr{G}$. Let $\mathscr{H}_0 = \{(a,b] : a, b \in \mathbb{Q}^d, a \leq b\}$. Then $\mathscr{H}_0$ is countable. Let $\mathscr{H}_1 = \{H \in \mathscr{H}_0 : H \subseteq \mathscr{G}\}$. Then $\mathscr{H}_1$ is a countable subset of $\mathscr{H}$ and $\bigcup \mathscr{H}_1 = G$. Then $G \in \sigma(\mathscr{H})$. This holds for each $G \in \mathscr{G}$, hence $\mathscr{G} \subseteq \sigma(\mathscr{H})$, so $\sigma(\mathscr{G}) \subseteq \sigma(\mathscr{H})$.

Now let $H \in \mathscr{H}$. Then $H = (a,b]$ for some $a, b \in \mathbb{R}^d$ with $a \leq b$. We have $a = (a_1, \ldots, a_d)$, $b = (b_1, \ldots, b_d)$ for some $a_1, \ldots, a_d, b_1, \ldots b_d \in \mathbb{R}$ with $a_k \leq b_k$, $k = 1, \ldots, d$. For each $n \in \mathbb{N}$, let $G_n = \prod_{k=1}^{d}(a_k, b_k + \frac{1}{n})$. Then $G_n \in \mathscr{G}$ and $\bigcap_{n \in \mathbb{N}} G_n = H$. Hence $H \in \sigma(\mathscr{G})$. This holds for each $H \in \mathscr{H}$, hence $\mathscr{H} \subseteq \sigma(\mathscr{H})$, so $\sigma(\mathscr{H}) \subseteq \sigma(\mathscr{G})$. □

Random variables. Let $(\Omega, \mathscr{F}, P)$ be a probability space.

Example. Roll 2 dice. Let X_1 be the number from the first die and X_2 the number from the second. X_1 and X_2 are random variables. So is $X_1 + X_2$. The expected value of X_1, is

$$E(X_1) = \sum_{k=1}^{6} kP(X_1 = k) = \frac{1}{6}\sum_{k=1}^{6} k = \frac{1}{6}\frac{6 \cdot 7}{2} = \frac{7}{2}.$$

Similarly, $E(X_2) = 3.5$, and for $X_1 + X_2$ we have

k	$P(X_1 + X_2 = k)$
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

$E(X_1 + X_2) = \sum_{k=2}^{12} kP(X_1 + X_2 = k) = \text{a mess} = 7$, or by linearity $E(X_1 + E_2) = E(X_1) + E(X_2) = 7$.

Example. Roll a die repeatedly. Let X be the number of rolls until a 5 is rolled.

$$\begin{aligned} P(X = 1) &= \frac{1}{6} = p && (q = 1 - p) \\ P(X = 2) &= qp \\ P(X = 3) &= q^2 p \\ &\vdots \\ P(X = n) &= q^{n-1} p && (\text{for } n \in \mathbb{N}) \\ P(X < \infty) &= \sum_{n \in \mathbb{N}} P(X = n) = \sum_{n=1}^{\infty} q^{n-1} p = p \sum_{k=0}^{\infty} q^k = p\frac{1}{1-q} = p\frac{1}{p} = 1. \end{aligned}$$

$$
\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} nP(X=n) = \sum_{n=1}^{\infty} npq^{n-1} \\
qE(X) &= \sum_{n=1}^{\infty} npq^{n} = \sum_{n=2}^{\infty} (n-1)pq^{n-1} \\
&= \sum_{n=1}^{\infty} \underbrace{(n-1)}_{0 \text{ when } n=1} pq^{n-1} \\
\underbrace{E(X) - qE(X)}_{(1-q)E(X)=pE(X)} &= \sum_{n=1}^{\infty} \underbrace{(n-(n-1))}_{1} pq^{n-1} = 1
\end{aligned}
$$

so $E(X) = \frac{1}{p} = 6$.

2/10/21

Example. Toss a biased coin n times. Let p be probability of heads on a given toss and $q = 1 - p$ is the probability of tails on any given toss. Let $A_k = \{\text{heads on toss } k\}$, with $A_1, \ldots, A_n$ independent. $P(A_k) = p$, $P(A_k^c) = q = 1 - p$. Let $X_k = 1_{A_k}$ and let $S_n = \sum_{k=1}^n X_k$, the number of heads in n tosses. Then the possible values of S_n are $0, \ldots, n$. Let $\bar{n} = \{1, \ldots, n\}$

Define a random variable Y_n on Ω by

$$Y_n(\omega) = (X_k(\omega))_{k \in \bar{n}}.$$

Then $Y_n : \Omega \longrightarrow \{0,1\}^{\bar{n}}$. Remember that $I \mapsto 1_I$ is a bijection from $\mathcal{P}(\bar{n})$ onto $\{0,1\}^{\bar{n}}$. For each $I \subseteq \bar{n}$, let

$$A_I = \{\omega \in \Omega : Y_n(\omega) = 1_I\} = \{X_k = 1 \text{ if } k \in I \text{ and } X_k = 0 \text{ if } k \notin I\} = \bigcap_{k=1}^{n} B_k^I$$

where

$$B_k^I = \begin{cases} A_k & \text{if } k \in I \\ A_k^c & \text{if } k \in \bar{n} \setminus I \end{cases}$$

That is, A_I is the event that the k-th throw is heads for all $k \in I$.

Since $\{0,1\}^{\bar{n}} = \bigcup_{I \subseteq \bar{n}} 1_I$, we have $\Omega = Y_n^{-1}[\{0,1\}^{\bar{n}}] = \bigcup_{I \subseteq \bar{n}} Y_n^{-1}[\{1_I\}] = \bigcup_{I \subseteq \bar{n}} A_I$. For each $I \subseteq \bar{n}$, $P(A_I) = P\left(\bigcap_{k=1}^n B_k^I\right) = \prod_{i=1}^n P(B_k^I)$. Now

$$P(B_k^I) = \begin{cases} p & \text{if } k \in I \\ q & \text{if } k \in \bar{n} \setminus I \end{cases}$$

Thus $P(A_I) = \prod_{i=1}^n P(B_k^I) = p^{|I|} q^{n-|I|}$. For each $k \in \{0, \ldots, n\}, \{S_n = k\} = \bigcup_{|I|=k} A_I$ so

$$P(S_n = k) = \sum_{|I|=k} P(A_I) = \underbrace{\sum_{|I|=k} p^k q^{n-k}}_{\text{same thing } \binom{n}{k} \text{ times}} = \binom{n}{k} p^k q^{n-k}$$

We say S_n has a binomial distribution with parameters n and p:

$$\operatorname{law}(S_n) = \operatorname{binom}(n, p).$$

$$E(S_n) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

the summand where $k = 0$ is 0, so

$$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p(p^{k-1} q^{n-k}) = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}$$

let $l = k - 1$, so $n - k = (n-1) - (k-1) = (n-1) - l$

$$= np \underbrace{\sum_{l=0}^{n-1} \binom{n-1}{l} p^l q^{(n-1)-l}}_{\sum_{l=0}^{n-1} P(S_{n-1}=l)=1} = np$$

A better way to do this calculation:

$$\begin{aligned} E(S_n) &= E(X_1 + \cdots + X_n) \\ &= E(X_1) + \cdots + E(X_n) \\ &= \underbrace{p + \cdots + p}_{n \text{ times}} \\ &= np \end{aligned}$$

but the second step ($E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$) requires proof.

Similarly, one can calculate $E[S_n(S_n - 1)]$ and use this to show that

$$\operatorname{var}(S_n) \coloneqq E\left\{[S_n = E(S_n)]^2\right\} = npq$$

or a better way; since $X_1, \ldots, X_n$ are independent,

$$\operatorname{var}(S_n) = \underbrace{\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)}_{n \text{ times, each } pq} = npq$$

Simple functions.

Definition. Let $(X, \mathscr{A})$ be a measurable space. Let Y be a set and let $\varphi : X \longrightarrow Y$. To say $\underline{\varphi \text{ is simple}}$ (or $\mathscr{A}$-simple) means that $\varphi[X]$ is finite and for each $y \in \varphi[X]$, $\varphi^{-1}[\{y\}] \in \mathscr{A}$.

Proposition 37. Let $\varphi : X \longrightarrow Y$ be simple and let $B \subseteq Y$. Then $\varphi^{-1}[B] \in \mathscr{A}$.

Proof.

$$\varphi^{-1}[B] = \varphi^{-1}\left[\bigcup_{y \in B} \{y\}\right] = \bigcup_{y \in B} \underbrace{\varphi^{-1}[\{y\}]}_{\text{empty if } y \notin \varphi[X]} = \bigcup_{y \in B \cap \varphi[X]} \varphi^{-1}[\{y\}] \in \mathscr{A}$$

□

Proposition 38. Let $\varphi : X \longrightarrow Y$ be simple. Let $h : Y \longrightarrow Z$ be any function. Let $\psi = h \circ \varphi$. Then $\psi : X \longrightarrow Z$ is simple.

Proof. $\psi[X] - h[\varphi[X]]$ is finite and for each $z \in \psi[X]$, $\psi^{-1}[\{z\}] = (h \circ \varphi)^{-1}[\{z\}] = \varphi^{-1}[h^{-1}[\{z\}]] \in \mathscr{A}$ by the previous proposition (since $h^{-1}[\{z\}]$ is a subset of Y, which gets pulled back to an element of $\mathscr{A}$). □

Proposition 39. Let $\varphi_k : X \longrightarrow Y_k$ for $k = 1, \ldots, n$. Let $Y = Y_1 \times \cdots \times Y_n$ and define $\varphi : X \longrightarrow Y$ by $\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x))$. Then the following are equivalent:

(a) $\varphi_1, \ldots, \varphi_n$ are simple.

(b) φ is simple.

Proof. Suppose $\varphi_1, \ldots, \varphi_n$ are simple. Let $R_k = \varphi_k[X]$ for $k = 1, \ldots, n$. Let $R = \varphi[X]$. For each $y \in R$, we have $y = \varphi(x)$ for some $x \in X$, so $y = (y_1, \ldots, y_n)$ where $y_k = \varphi_k(x) \in R_k$. Then $R \subseteq R_1 \times \cdots \times R_n$, which is finite, so $R = \varphi[X]$ is finite. Let $y \in R$ and let's show $\varphi^{-1}[\{y\}] \in \mathscr{A}$.

Now $y = (y_1, \ldots, y_n)$ for suitable $y_k \in Y_k$. For each $x \in X$, we have $\varphi(x) = y$ if and only if $\varphi_k(x) = y_k$ for $k = 1, \ldots, n$. Hence $\varphi^{-1}[\{y\}] = \bigcap_{k=1}^n \varphi_k^{-1}[\{y_k\}] \in \mathscr{A}$. Thus φ is simple.

Conversely suppose φ is simple. Then for $k = 1, \ldots, n$, φ_k is simple because $\varphi_k = \pi_k \circ \varphi$. $\square$

Corollary 40. Let $\varphi_k : X \longrightarrow Y_k$ be simple for $k = 1, \ldots, n$. Let $Y = Y_1 \times \cdots \times Y_n$ and let $h : Y \longrightarrow Z$. Define ψ on X by $\psi(x) = h(\varphi_1(x), \ldots, \varphi_n(x))$. Then ψ is simple.

Proof. $\psi = h \circ \varphi$, where φ is defined on X by $\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x))$. By the previous proposition, φ is simple, hence $\psi = h \circ \varphi$ is simple. $\square$

Example. Let $\varphi_1, \varphi_2 : X \longrightarrow \mathbb{R}$ be simple. Then $\varphi_1 + \varphi_2$ is simple because $(\varphi_1 + \varphi_2)(x) = \varphi_1(x) + \varphi_2(x) = h(\varphi_1(x), \varphi_x(x))$ for all $x \in X$, where $h(y_1, y_2) = y_1 + y_2$.

Similarly, $\varphi_1\varphi_2$ is simple and if φ_2 is not 0, then φ_1/φ_2 is simple.

2/12/21: Sane Integration I

Let X be a random variable on $(\mathbf{X}, \mathscr{A})$. law$(X)$ is the probability measure μ on $\mathscr{A}$ defined by $\mu(A) = P(X \in A) = P(\{\omega \in \Omega \mid X(\omega) \in A\}) = P(X^{-1}[A])$. The distribution of X is law(X).

Special Cases.

1. Suppose $\mathbf{X} = \mathbb{R}$ and $\mathscr{A} = \text{Borel}(\mathbb{R})$. Define F on $\mathbb{R}$ by $F(x) = P(X \leq x) = \mu((-\infty, x])$. F is the cumulative distribution function for X.
2. (Discrete Case) Suppose $\mathbf{X}$ is countable and $\mathscr{A} = \mathcal{P}(\mathbf{X})$. Then for each $A \subseteq \mathbf{X}$, $\mu(A) = \sum_{x\in A} f(x)$, since $A = \dot\bigcup_{x\in A}\{x\}$ and A is countable, so $\mu(A) = \sum_{x\in A} \mu(\{x\})$ where $f : \mathbf{X} \longrightarrow [0, 1]$ is defined by $f(x) = P(X = x) = \mu(\{x\})$.
3. Suppose $\mathbf{X} = \mathbb{R}^d$, $\mathscr{A} = \text{Borel}(\mathbb{R}^d)$, and there is a (Borel) function $f : \mathbb{R}^d \longrightarrow [0, \infty]$ such that for each $A \in \mathscr{A}$,
$$P(X \in A) = \int_A f(x)\, dx.$$
Then we say f is a probability density function for X (or for μ).

Note. Let μ, ν be probability measures on Borel$(\mathbb{R})$. Suppose $\mu((-\infty, x]) = \nu((-\infty, x])$ for each $x \in \mathbb{R}$. Then $\mu = \nu$.

Proof. Let $\mathscr{H} = \{(-\infty, x] : x \in \mathbb{R}\}$. $\mathscr{H}$ is a π-system on $\mathbb{R}$. $\sigma(\mathscr{H}) = \text{Borel}(\mathbb{R})$. Hence $\mu = \nu$ by a consequence of the $\pi - \lambda$ theorem. $\square$

$(X, \mathscr{A}, \mu)$ is a measure space. Let $\mathscr{S} = \{\varphi : \varphi \text{ is an } \mathscr{A}\text{-simple function from } X \text{ to } \mathbb{R}\}$, $\mathscr{S}^+ = \{\varphi \in \mathscr{S} : \varphi \geq 0\}$.

Definition. Let $\varphi \in \mathscr{S}^+$. Then $I(\varphi) = \sum_y y\mu(\varphi = y)$. (Think of $I(\varphi)$ as $\int \varphi\, d\mu$.)

Note. $0 \cdot \infty = 0$. Then $0 \cdot \mu(\varphi = 0) = 0$ even if $\mu(\varphi = 0) = \infty$.

Example.
$$\int_{\mathbb{R}} 1_{[0,7]}\, dm = 1m([0, 7]) + 0\, \underbrace{m(\mathbb{R} \setminus [0, 7])}_{\infty} = 7 + 0 = 7,$$
where m is the Lebesgue measure on $\mathbb{R}$.

Proposition 41. Let $\varphi, \psi \in \mathscr{S}^+$. Then $\varphi + \psi \in \mathscr{S}^+$ and $I(\varphi + \psi) = I(\varphi) + I(\psi)$.

Proof. We already know $\varphi + \psi$ is simple and clearly $\varphi + \psi \geq 0$.

$$I(\varphi+\psi) = \sum_w w\mu(\varphi+\psi = w) = \sum_w \sum_{y,z} \mu(\varphi+\psi = w, \varphi = y, \psi = z)$$

$\{\varphi = y, \psi = z\}, (y,z) \in \varphi[X] \times \psi[X]$ are disjoint and their union is X ???

$$\begin{aligned}
&= \sum_{y,z} \sum_w w\mu(\underbrace{\varphi+\psi = w, \varphi = y, \psi = z}_{\text{empty unless } w=y+z}) \\
&= \sum_{y,z} (y+z)\mu(\varphi = y, \psi = z) \\
&= \sum_{y,z} y\mu(\varphi = y, \psi = z) + \sum_{y,z} z\mu(\varphi = y, \psi = z) \\
&= \sum_y y \sum_z \mu(\varphi = y, \psi = z) + \sum_z z \sum_y \mu(\varphi = y, \psi = z) \\
&= \sum_y y\mu(\varphi = y) + \sum_z z\mu(\psi = z) \\
&= I(\varphi) + I(\psi).
\end{aligned}$$

□

Corollary 42. Let $\varphi, \psi \in \mathscr{S}^+$. Suppose $\varphi \leq \psi$. Then $I(\varphi) \leq I(\psi)$.

Proof. $\psi = (\psi - \varphi) + \varphi$ and $\varphi - \psi \in \mathscr{S}^+$ so

$$I(\psi) = \underbrace{I(\psi - \varphi)}_{\geq 0 \text{ since } \mu \geq 0} + I(\varphi) \geq I(\varphi)$$

□

Proposition 43. Let $\varphi \in \mathscr{S}^+$ and let $c \in [0, \infty)$. Then $c\varphi \in \mathscr{S}^+$ and $I(c\varphi) = cI(\varphi)$.

Proof. Easy. □

Measurable functions. Let $(X, \mathscr{A})$ and $(Y, \mathscr{B})$ be measurable spaces. Let $f : X \longrightarrow Y$. To say f is <u>measurable</u> (or $\mathscr{A}/\mathscr{B}$-measurable) means that for each $B \in \mathscr{B}$, we have $f^{-1}[B] \in \mathscr{A}$.

Proposition 44. Suppose $\mathscr{B} = \sigma(\mathscr{H})$. Then f is measurable if and only if for each $H \in \mathscr{H}$, we have $f^{-1}[H] \in \mathscr{A}$.

Proof. The forward direction is obvious because $\mathscr{H} \subseteq \mathscr{B}$. For the backwards direction, suppose $f^{-1}[H] \in \mathscr{A}$ for each $H \in \mathscr{H}$. Let $\widetilde{\mathscr{B}} = \{B \subseteq Y : f^{-1}[B] \in \mathscr{A}\}$ Then $\mathscr{H} \subseteq \widetilde{\mathscr{B}}$. Then $\widetilde{\mathscr{B}}$ is a σ-field on Y (easy exercise). But $\mathscr{B}$ is the smallest σ-field on Y containing $\mathscr{H}$. Thus $\mathscr{B} \subseteq \widetilde{\mathscr{B}}$. In other words, for each $B \in \mathscr{B}$, we have $f^{-1}[B] \in \mathscr{A}$. In other words (again), f is measurable. □

Corollary 45. Suppose Y is a topological space. Let $\mathscr{B} = \text{Borel}(Y) := \sigma(\mathscr{G})$ where $\mathscr{G}$ is the set of open subsets of Y. Let $f : X \longrightarrow Y$. Then f is measurable if and only if for each open set $G \subseteq Y$, $f^{-1}[G] \in \mathscr{A}$.

Corollary 46. Let $f : X \longrightarrow [-\infty, \infty]$. Then the following are equivalent:

(a) f is measurable.

(b) For all $y \in \mathbb{R}$, $\{f > y\} \in \mathscr{A}$.

(c) For all $y \in \mathbb{R}$, $\{f \geq y\} \in \mathscr{A}$.

(d) For all $y \in \mathbb{R}$, $\{f < y\} \in \mathscr{A}$.

(e) For all $y \in \mathbb{R}$, $\{f \le y\} \in \mathscr{A}$.

Proof. Obviously (a) $\implies$ all of (b), ..., (e).

Suppose (b) holds. Let $\mathscr{H} = \{(y, \infty] : y \in \mathbb{R}\}$. For each $y \in \mathbb{R}$, $(y, \infty]$ is open in $[-\infty, \infty]$. Hence $\mathscr{H} \subseteq \mathscr{G} =$ the set of open subsets of $[-\infty, \infty]$. Hence $\sigma(\mathscr{H}) \subseteq \sigma(\mathscr{G}) = \text{Borel}([-\infty, \infty])$. Let G be open in $[-\infty, \infty]$. Note that for each $y \in \mathbb{R}$, $[-\infty, y) = \bigcup_{n \in \mathbb{N}} [-\infty, y - \frac{1}{n}] = \bigcup_{n \in \mathbb{N}} ([-\infty, \infty] \setminus (y - \frac{1}{n}, \infty] \in \sigma(\mathscr{H})$. Continuing, we get $G \in \sigma(\mathscr{H})$. This holds for each $G \in \mathscr{G}$. Hence $\sigma(\mathscr{G}) \subseteq \sigma(\mathscr{H})$. Thus $\sigma(\mathscr{G}) = \sigma(\mathscr{H})$. Thus f is measurable if and only if for all $H \in \mathscr{H}$, $f^{-1}[H] \in \mathscr{A}$.

Similarly, (c) $\implies$ (a), (d) $\implies$ (a), and (e) $\implies$ (a). □

2/15/21: Sane Integration II

Remark. Any simple function is measurable. This is because the image is finite, hence every set in the image is measurable, and by definition the preimage of any set in the image is measurable.

Proposition 47. Let (f_n) be a sequence of measurable functions $f_n : X \to [-\infty, \infty]$. Then;

(a) $\sup_n f_n$ and $\inf_n f_n$ are measurable

Proof. For each $y \in \mathbb{R}$, $\{\sup_n f_n > y\} = \{x \in X : f_n(x) > y \text{ for some } n\} = \bigcup_n \underbrace{\{f_n > y\}}_{\in \mathscr{A}} \in \mathscr{A}$.

Similarly, for each $y \in \mathbb{R}, \{\inf_n f_n \ge y\} = \{x \in X : f_n(x) \ge y \text{ for each } x\} = \bigcap_n \underbrace{\{f_n \ge y\}}_{\in \mathscr{A}} \in \mathscr{A}$. □

(b) $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are measurable functions.

Proof. For each $x \in X$,

$$\limsup_{n\to\infty} f_n(x) = \inf_m \sup_{n \ge m} f_n(x) \text{ and } \liminf_{n\to\infty} f_n(x) = \sup_m \inf_{n \ge m} f_n(x).$$

By (a), for each $m \in \mathbb{N}$, $\sup_{n \ge m} f_n$ and $\inf_{m \ge n} f_n$ are measurable.

Hence by (a) again, $\inf_m \sup_{n \ge m} f_n$ and $\sup_m \inf_{m \ge n} f_n$ are measurable. In other words, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are measurable. □

(c) Suppose $\lim_{n\to\infty} f_n(x)$ exists in $[-\infty, \infty]$ for each $x \in X$. Then $\lim_{n\to\infty} f_n(x)$ is measurable.

Proof. For each $x \in X$, since $\lim_{n\to\infty} f_n(x)$ exists in $[-\infty, \infty]$, we have

$$\lim_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x).$$

Hence $\lim_{n\to\infty} f_n = \limsup_{n\to\infty} f_n$ is measurable by (b). □

Definition. Let $W \subseteq X$ and let $\mathscr{E} \subseteq \mathcal{P}(X)$. Then the *trace of $\mathscr{E}$ on W* is

$$E|W = \{E \cap W : E \in \mathscr{E}\}.$$

Example. If X is a topological space, $\mathscr{G}$ is the set of open subsets of X and $W \subseteq X$, then $\mathscr{G}|W$ is the subspace topology on W inherited from X; i.e. the set of relatively open subsets of W.

Example. If $(X, \mathscr{A})$ is a measurable space and $W \subseteq X$, then $\mathscr{A}|W$ is a σ-field on W and is called the subspace σ-field that W inherits from $(X, \mathscr{A})$.

Proposition 48. Let $(X, \mathscr{A})$ be a measurable space. Let $f, g : X \longrightarrow [-\infty, \infty]$ be measurable. Then the following sets are all measurable:

(a) $\{f < g\}$

(b) $\{f > g\}$

(c) $\{f \ne g\}$

(d) $\{f = g\}$

(e) $\{f \leq g\}$

(f) $\{f \geq g\}$

Proof. $\{f < g\} = \{x \in X : \exists\, r \in \mathbb{Q}, f(x) < r \text{ and } r < g(x)\} = \bigcup_{r \in \mathbb{Q}} (\underbrace{\{f < r\}}_{\in \mathscr{A}} \cap \underbrace{\{r < g\}}_{\in \mathscr{A}}) \in \mathscr{A}$.

Similarly, $\{f > g\} \in \mathscr{A}$. $\{f \neq g\} = \{f < g\} \cup \{f > g\} \in \mathscr{A}$. $\{f = g\} = X \setminus \{f \neq g\}$. $\{f \leq g\} = \{f < g\} \cup \{f = g\}$. $\{f \geq g\} = \{f > g\} \cup \{f = g\}$ ☐

Proposition 49. Let (f_n) be a sequence of measurable functions $f_n : X \to [-\infty, \infty]$. Let $W = \{x \in X : \lim_{n\to\infty} f_n(x) \text{ exists on } [-\infty, \infty]\}$. Define $f : W \to [-\infty, \infty]$ by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

Then $W \in \mathscr{A}$ and f is measurable.

Proof. Let $g = \liminf_{n\to\infty} f_n$ and $h = \limsup_{n\to\infty} f_n$. Then $g, h : X \to [-\infty, \infty]$ are measurable. Now $W = \{x \in X : g(x) = h(x)\} = \{g = h\} \in \mathscr{A}$. For each n, $f_n|W : W \to [-\infty, \infty]$ is measurable. For each $x \in W$, $\lim_{n\to\infty} \underbrace{(f_n|W)(x)}_{f_n(x),\ \text{b.c } x \in W}$ and hence $f = \lim_{n\to\infty}(f_n|W)$ is measurable. ☐

As we've seen, a pointwise limit of measurable functions is measurable. Hence, a pointwise limit of simple functions is measurable.

Theorem 50. Let $f : X \longrightarrow [0, \infty]$ be measurable. Then there is a sequence (φ_n) of simple functions $\varphi_n : X \longrightarrow [0, \infty)$ such that $\varphi_n \uparrow f$. Furthermore, we may choose (φ_n) so that for each $W \subseteq X$, if f is bounded on W, then $\varphi_n \to f$ uniformly on W.

Proof. Let (r_n) be an enumeration of $\mathbb{Q} \cap [0, \infty)$. Let $A_k = \{f \geq r_k\}$ ($\in \mathscr{A}$ since f is measurable), let $\psi_k = r_k 1_{A_k}$, let $\varphi_n = \max\{\psi_1, \ldots, \psi_n\}$. φ_n is simple because $\varphi_n(x) = h_n(\psi_1(x), \ldots, \psi_n(x))$ where $h_n(y_1, \ldots, y_n) = \max\{y_1, \ldots, y_n\}$. $\varphi_n \leq \varphi_{n+1}$. Each $\psi_k \leq f$, because

$$\psi_k = \begin{cases} 0 \leq f(x) & \text{if } x \in X \setminus A_k \\ r_k \leq f(x) & \text{if } x \in A_k \end{cases}$$

If $0 \leq y \leq f(x)$, then there exists k such that $r_k \in (y, f(x))$,so for $n \geq k$, $\varphi_n(x) \geq \psi_k(x) = r_k 1_{A_k} = r_k > y$. Thus $\varphi_n(x) \to f(x)$. Then holds for each $x \in X$. Now let $W \subseteq X$ such that f is bounded on W. Then there exists $b \in [0, \infty)$ such that for all $x \in W$, $f(x) \leq b$. Let $\varepsilon > 0$. Choose $m \in \mathbb{N}$ such that $b/m < \varepsilon$. For $l = 1, \ldots, m$, choose $k_l \in \mathbb{N}$ such that $r_{k_l} \in (\frac{l-1}{m}b, \frac{l}{m}b)$.

Let $n \geq \max\{k_1, \ldots, k_m\}$. Then $k_1, \ldots, k_m \in \{1, \ldots, n\}$ so $\max\{\psi_k, \ldots, \psi_{k_m}\} < \varphi_n$. Let $x \in W$. Then $f(x) \in [0, b)$. Hence $\exists\, l \in \{1, \ldots, m\}$ such that $f(x) \in [\frac{l-1}{m}b, \frac{l}{m}b)$. Then $f(x) \geq \frac{l-1}{m}b$. Then $\varphi_n(x) \geq \psi_{k_l}(x) =$

FIX LATER

☐

Thus nonnegative measurable functions are exactly the pointwise limit of sequences of nonnegative (finite) simple functions.

Integrals of Nonnegative Measurable Functions. Let $(X, \mathscr{A}, \mu)$ be a measure space.

Definition. Let $f : X \longrightarrow [0, \infty]$ be measurable. Then

$$\int f \, d\mu \overset{def}{=} \sup\{I(\varphi) : \varphi \in \mathscr{S}^+, \varphi \leq f\}.$$

As we know, for all $\varphi, \psi \in \mathscr{S}^+$, if $\varphi \leq \psi$, then $I(\varphi) \leq I(\psi)$. Hence, for all $\psi \in \mathscr{S}^+$,

$$\int \psi \, d\mu = I(\psi).$$

For all measurable functions $f, g : X \longrightarrow [0, \infty]$, if $f \leq g$, then $\int f \, d\mu \leq \int g \, d\mu$, because $\{I(\varphi) : \varphi \in \mathscr{S}^+, \varphi \leq f\} \subseteq \{I(\varphi) : \varphi \in \mathscr{S}^+, \varphi \leq g\}$.

Markov's Inequality. Let $f : X \longrightarrow [0, \infty]$ be measurable. Let $y \in (0, \infty)$. Then $\mu(f \geq y) \leq \frac{1}{y} \int f \, d\mu$.

Proof. $y1_{\{f \geq y\}} \leq f$, so $\underbrace{I(y1_{\{f \geq y\}})}_{y\mu(\{f \geq y\})} \leq \int f \, d\mu$ so $\mu(\{f \geq y\}) \leq \frac{1}{y} \int f \, d\mu$. □

Definition. Let $P(x)$ be a property of points $x \in X$. To say P holds almost everywhere (or μ-almost everywhere; abbreviated a.e., μ-a.e.) means that that $\{x \in X : P(x) \text{ is false}\}$ belongs to $\mathscr{A}$ and has μ-measure 0.

Proposition 51. Let $f : X \longrightarrow [0, \infty]$ be measurable. Then $\int f \, d\mu = 0$ if and only if $f = 0$ almost everywhere.

Proof. ($\Longrightarrow$) Suppose $\int f \, d\mu = 0$. Then for each $n \in \mathbb{N}$, $\mu(f \geq \frac{1}{n}) \leq n \int f \, d\mu = 0$ by Markov's inequality. But $\bigcup_n \{f \geq \frac{1}{n}\} = \{f > 0\} = \{f \neq 0\}$ since f is nonnegative. Hence $0 = \mu(f \neq 0) \leq \sum_n \mu(f \geq \frac{1}{n}) = 0$. Thus $\mu(f \neq 0) = 0$, so $f = 0$ μ-almost everywhere.

($\Longleftarrow$) Suppose $f = 0$ almost everywhere. Let $\varphi \in \mathscr{S}^+$ with $\varphi \leq f$. Then for each $y \in \varphi[X] \setminus \{0\}$, $\{\varphi = y\} \subseteq \{f > 0\}$, so $\mu(\varphi = y) = 0$. Hence $I(\varphi) = \sum_y y\mu(\varphi = y) = 0$. Then $\int f \, d\mu = \sup\{I(\varphi) : \varphi \in \mathscr{S}^+, \varphi \leq f\} = \sup\{0\} = 0$. □

2/17/21: Sane Integration II: Episode 1

Remark (Theorem 50.0.1 Patch Notes.)**.** Suppose we have $f : X \to [0, \infty]$, and (r_n) an enumeration of $\mathbb{Q} \cap [0, \infty)$. Let $A_k = \{f \geq r_k\}$, $\psi_k = r_k 1_{A_k}$, $\varphi_n = \max\{\psi_1, \ldots, \psi_n\}$. We have seen that $\varphi_n \uparrow f$ pointwise.

Let $W \subseteq X$ such that $f[W]$ is bounded. We want to show that $\varphi_n \to f$ uniformly on W. Since $f[W]$ is a bounded subset of $[0, \infty)$, there exists $b \in \mathbb{N}$ such that $f < b$ on W. Let $\varepsilon > 0$. Choose $m \in \mathbb{N}$ such that $b/m < \varepsilon$. For $l = 1, \ldots, m$, choose $k_l \in \mathbb{N}$ such that $r_{k_l} = \frac{l-1}{m} b \in \mathbb{Q} \cap [0, \infty)$. Let $n \geq \max\{k_1, \ldots, k_m\}$. Let $x \in W$. Then $0 \leq f(x) < b$, so there exists $l \in \{1, \ldots, m\}$ such that $\frac{l-1}{m} b \leq f(x) < \frac{l}{m} b$. Then

$$\frac{l}{m} b > f(x) \geq \varphi_n(x) \geq \psi_{k_l}(x) = r_{k_l} 1_{A_{k_l}} = \frac{l-1}{m} b 1_{\{f \geq \frac{l-1}{m} b\}}(x) = \frac{l-1}{m} b,$$

so $0 \leq f(x) = \varphi_n(x) < \frac{l-1}{m} b - \frac{l-1}{m} b = b/m < \varepsilon$. We've shown that for each $n \geq \max\{k_{l_1}, \ldots, k_{l_m}\}$, for each $x \in W$, $|f(x) - \varphi_n(x)| < \varepsilon$. Thus $\varphi_n \to f$ uniformly on W.

Definition (Another Loose End.)**.** Let X, Y be topological spaces. Let $f : X \longrightarrow Y$. To say that f is a Borel function means that f is $\mathscr{A}/\mathscr{B}$-measurable, where $\mathscr{A} = \text{Borel}(X)$ and $\mathscr{B} = \text{Borel}(Y)$. If f is continuous, then f is Borel, but the converse may not hold:

Example. $1_{\mathbb{Q}}$ is a Borel function from $\mathbb{R}$ to $\mathbb{R}$ but it is not continuous.

Measurability of functions defined piecewise. Let $(X, \mathscr{A})$ and $(Y, \mathscr{B})$ be measurable spaces. Let Π be a countable partition of X with $\Pi \subseteq \mathscr{A}$. Let (X_n) be an enumeration of Π. For each n, let $f_n : X_n \longrightarrow Y$ be $\mathscr{A}_n/\mathscr{B}$-measurable, where $\mathscr{A}_n = \mathscr{A}|X_n$. Define $f : X \longrightarrow Y$ by $f(x) = f_n(x)$ for each $x \in X_n$, for each n. Then f is $\mathscr{A}/\mathscr{B}$-measurable.

Proof. Let $B \in \mathscr{B}$. Then $f^{-1}[B] = \bigcup_n f_n^{-1}[B] \in \mathscr{A}$, since $f_n^{-1}[B] = A_n \cap X_n$ for some $A_n \in \mathscr{A}$. □

Back to integration of nonnegative measurable functions. Let $(X, \mathscr{A}, \mu)$ be a measure space.

Proposition 52. Let $f : X \longrightarrow [0, \infty]$ be measurable with $\int f \, d\mu < \infty$. Then

(a) $f < \infty$ almost everywhere

(b) $\{f > 0\}$ is of σ-finite μ-measure.

Proof. (a) For each $n \in \mathbb{N}$, $\{f = \infty\} \subseteq \{f \geq n\}$, so by Markov's inequality

$$\mu(f = \infty) < \mu(f \geq n) \leq \frac{1}{n} \underbrace{\int f \, d\mu}_{<\infty} \xrightarrow[n\to\infty]{} 0.$$

Hence $\mu(f = \infty) = 0$.

(b) $\{f > 0\} = \bigcup_{n\in\mathbb{N}} \{f \geq 1/n\}$. Now for each $n \in \mathbb{N}$, $\mu(f \geq 1/n) \leq n \int f \, d\mu < \infty$. Thus $\{f > 0\}$ is a union of countably many measurable sets of finite μ-measure. In other words, $\{f > 0\}$ is of $\underline{\sigma\text{-finite}}$ μ-measure.

□

Proposition 53. Let $f : X \longrightarrow [0, \infty]$ be measurable. Let $c \in [0, \infty]$. Then cf is measurable and $\int cf \, d\mu = c \int f \, d\mu$.

Proof. Either $c = 0$ or $0 < c < \infty$ or $c = \infty$.

Case 1. Suppose $c = 0$. Then $cf(x) = 0$ for each $x \in X$ (even if $f(x) = \infty$) so

$$\int cf \, d\mu = \int 0 \, d\mu = 0 = 0 \int f \, d\mu,$$

even if $\int f \, d\mu = \infty$.

Case 2. Suppose $0 < c < \infty$. Then

$$\begin{aligned}
\int cf \, d\mu &= \sup\{I(\varphi) : \varphi \in \mathscr{S}^+, \varphi \leq cf\} \\
&= \sup\left\{cI\left(\frac{\varphi}{c}\right) : \varphi \in \mathscr{S}^+, \frac{\varphi}{c} \leq f\right\} \\
&= c \sup\{I(\psi) : \psi \in \mathscr{S}^+, \psi \leq f\} \\
&= c \int f \, d\mu
\end{aligned}$$

Case 3. Suppose $c = \infty$. Let $A = \{f > 0\}$. Then $cf = \infty \cdot 1_A$. Suppose $\mu(A) > 0$, then for each $n \in \mathbb{N}$,

$$\int cf \, d\mu = \int \infty 1_A \, d\mu \geq \int n 1_A \, d\mu = n\mu(A) \to \infty \text{ as } n \to \infty.$$

Hence $\int cf \, d\mu = \infty$. Since $\mu(f > 0) > 0, \int f \, d\mu > 0$ too. Now suppose $\mu(A) = 0$. Then $cf = 0$ almost everywhere, so $\int cf \, d\mu = 0$, and $f = 0$ almost everywhere, so $\int f \, d\mu = 0$, so $c \int f \, d\mu = 0$. □

Theorem 54 (Monotone Convergence Theorem)**.** Let (f_n) be an increasing sequence of measurable functions $f_n : X \longrightarrow [0, \infty]$. Let $f = \lim_{n\to\infty} f_n$. Then $\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu$.

Lemma 55. Let $\varphi \in \mathscr{S}^+$. Define ν on $\mathscr{A}$ by $\nu(A) = \int \varphi 1_A \, d\mu$. Then ν is a measure

Proof of MCT. Since $0 \leq f_n \leq f_{n+1} \leq f$, we have $0 \leq \int f_n \, d\mu \leq \int f_{n+1} \, d\mu \leq f \, d\mu$. Hence $L = \lim_{n\to\infty} = \int f_n \, d\mu$ exists and $L = \int f \, d\mu$. We wish to show $L = \int f \, d\mu$.

Suppose not. Then $L < \int f \, d\mu$. Hence by the definition of $\int f \, d\mu$, there exists $\varphi \in \mathscr{S}^+$ such that $\varphi \leq f$ and $L < \int \varphi \, d\mu$. Then there exists $\alpha \in (0, 1)$ such that $L < \alpha \int \varphi \, d\mu$. Let $\psi = \alpha\varphi$. Then $L < \int \psi \, d\mu$. Let $A_n = \{f_n \geq \psi\}$. If $f(x) > 0$, then $f(x) > \alpha f(x) \geq \alpha\varphi(x) = \psi(x)$. Then there exists n such that $f_n(x) > \psi(x)$, so $x \in A_n$. If $f(x) = 0$, then $\varphi(x) = 0$, so $\psi(x) = 0$, so $f_n(x) \geq \psi(x)$ so $x \in A_n$.

Thus $\bigcup_n A_n = X$. Since $f_n \leq f_{n+1}$, $A_n \subseteq A_{n+1}$. Then $A_n \uparrow X$, so $\nu(A_n) \uparrow \nu(X)$, where ν is the measure on $\mathscr{A}$ defined by $\nu(A) = \int \psi 1_A \, d\mu$. Now $\nu(X) = \int \psi 1_X \, d\mu = \int \psi \, d\mu > L$. Hence there is an n such that $\mu(A_n) > L$. But

$$\nu(A_n) = \int \psi 1_{A_n} \, d\mu \leq \int f_n 1_{A_n} \, d\mu \leq \int f_n \, d\mu \leq L,$$

a contradiction. Therefore $L = \int f \, d\mu$, as desired. □

2/19/21: Sane Integration II Episode 2: Consequences of MCT

Lemma 56 (Fatou's lemma)**.** Let (f_n) be a sequence of measurable functions $f_n : X \longrightarrow [0, \infty]$. Let $f = \liminf_{n\to\infty} f_n$. Then

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

Proof. Let $g_m = \inf_{m \leq n} f_n$. Then $g_m \uparrow f$, so $\int g_m \, d\mu \uparrow \int f \, d\mu$. But for each m, $f_m \geq g_m$, so

$$\liminf_{m\to\infty} \int f_m \, d\mu \geq \liminf_{m\to\infty} \int g_m \, d\mu = \lim_{m\to\infty} \int g_m \, d\mu = \int f \, d\mu.$$

□

Corollary 57. Let (f_n) be a sequence of measurable functions $f_m : X \longrightarrow [0, \infty]$. Suppose $f(x) = \lim_{n\to\infty} f_n(x)$ exists in $[0, \infty]$ for each $x \in X$. Suppose also that $f_n \leq f$ for each n. Then $\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu$.

Proof. By Fatou's lemma, $\int f \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu$. But for each n, $\int f_n \, d\mu \leq \int f \, d\mu$. Hence

$$\limsup_{n\to\infty} \int f_n \, d\mu \leq \int f \, d\mu$$

Hence

$$\liminf_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu = \limsup_{n\to\infty} \int f_n \, d\mu$$

so $\int f_n \, d\mu \to \int f \, d\mu$ as $n \to \infty$. □

Corollary 58. Let (f_n) be a sequence of measurable functions $f_n : X \longrightarrow [0, \infty]$ and suppose $f(x) = \lim_{n\to\infty} f_n(x)$ exists for $x \in X$. Suppose also that $L = \lim_{n\to\infty} \int f_n \, d\mu$ exists and is finite. Then $\int f \, d\mu < \infty$.

Proof. By Fatou,

$$\int f \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu = L < \infty.$$

□

Remark. Let $f, g : X \longrightarrow [0, \infty]$ be measurable. Suppose $f \leq g$ almost everywhere. Then $\int f \, d\mu \leq \int g \, d\mu$. Similarly, if $f = g$ almost everywhere, then $\int f \, d\mu = \int g \, d\mu$.

Proof. Suppose $f \leq g$ almost everywhere. Let $\varphi \in \mathscr{S}^+$ with $\varphi \leq f$. Let $A = \{f \leq g\}$ and $B = \{f > g\}$. Then $A, B \in \mathscr{A}$, $A \cap B = \varnothing$, $A \cup B = X$, and $\mu(B) = 0$. Let $\psi = \varphi 1_A$. Then $\psi \in \mathscr{S}^+$, $\psi \leq g$, and $I(\psi) = I(\varphi)$. Thus $I(\varphi) \leq \int g\, d\mu$ for each such φ, hence $\int g\, d\mu$ is an upper bound for $\{I(\varphi) : \varphi \in \mathscr{S}^+$ and $\varphi \leq f\}$. Thus $\int g\, d\mu \geq \int f\, d\mu$. □

A loose end about simple functions. Let $c_1, \ldots, c_n$ be constants, let $A_1, \ldots, A_n$ be measurable sets, and let $\varphi = \sum_{j=1}^n c_j 1_{A_j}$. Then φ is simple. Conversely, let φ be simple. Let $n = |\varphi[X] \setminus \{0\}|$. Let $c_1 \ldots, c_n$ be an enumeration of $\varphi[X] \setminus \{0\}$ Let $A_j = \{\varphi = c_j\}$ for $j = 1, \ldots, n$. Then $\sum_{j=1}^n c_j 1_{A_j} = \varphi$ because $A_1, \ldots, A_n$ are disjoint.

Additivity of the integral. Let $f, g : X \longrightarrow [0, \infty]$ be measurable. Then $f + g$ is measurable (measurable functions are limits of simple functions, a sum of two simple functions is simple, a limit of a sequence of simple functions is measurable). Is $\int f + g\, d\mu = \int f\, d\mu = \int g\, d\mu$?

If $\varphi, \psi \in \mathscr{S}^+$ with $\varphi \leq f$ and $\psi \leq g$, then $\varphi + \psi \in \mathscr{S}^+$ and $\varphi + \psi \leq f + g$. It follows easily from this observation that $\int f\, d\mu + \int g\, d\mu \leq \int f + g\, d\mu$.

What about "$\geq$"? On $[0, 1]$ with μ the Lebesgue measure, let $f(x) = x$ and $g(x) = 1-x$. Then $f(x)+g(x) = 1$. Thus $f + g$ is simple. Hence a simple function $\leq f + g$ cannot necessarily be written as a sum of two simple functions $\leq f, g$, respectively. So to prove "$\geq$", we must approximate somehow.

There is one case where "$\geq$" is easy: $\{f > 0\} \cap \{g > 0\} = \varnothing$. Then given $\varphi \in \mathscr{S}^+$ with $\varphi \leq f + g$, consider $\psi 1_{\{f>0\}}$ and $\varphi 1_{\{g>0\}}$

Now for the general case. Since $f : X \longrightarrow [0, \infty]$ is measurable, there is an increasing sequence (φ_n) of simple functions $\varphi_n : X \longrightarrow [0, \infty)$ such that for each $x \in X$, $\varphi_n(x) \uparrow f(x)$. Similarly, there is an increasing sequence (ψ_n) of simple functions $\psi_n : X \longrightarrow [0, \infty)$ such that for each $x \in X$, $\psi_n(x) \uparrow g(x)$. Then $\varphi_n + \psi_n \uparrow f + g$. Hence

$$\begin{aligned}
\int f + g\, d\mu &= \lim_{n\to\infty} \int \varphi_n + \psi_n\, d\mu && \text{(MCT)} \\
&= \lim_{n\to\infty} \left(\int \varphi_n\, d\mu + \int \psi_m\, d\mu \right) && (\varphi, \psi \text{ are simple}) \\
&= \lim_{n\to\infty} \int \varphi_n\, d\mu + \lim_{n\to\infty} \int \psi_n\, d\mu && \text{(since both limits exist)} \\
&= \int f\, d\mu + \int g\, d\mu. && \text{(MCT)}
\end{aligned}$$

Theorem 59 (Beppo Levi's Theorem)**.** Let (f_k) be a sequence of measurable functions $f_k : X \longrightarrow [0, \infty]$. Then

$$\int \sum_{k=1}^{\infty} f_k\, d\mu = \sum_{k=1}^{\infty} \int f_k\, d\mu.$$

Proof. Let $g_n = \sum_{k=1}^n f_k$ and $g = \sum_{k=1}^{\infty} f_k$. Then $0 \leq g_n \uparrow g$, so $\sum_{k=1}^n \int f_k\, d\mu = \int g_n\, d\mu$ by the additivity of the integral, which increases to $\int g\, d\mu = \int \sum_{k=1}^{\infty} f_k\, d\mu$ by the monotone convergence theorem. Hence $\int \sum_{k=1}^{\infty} f_k\, d\mu = \sum_{k=1}^{\infty} \int f_k\, d\mu$. □

Integration of $\bar{\mathbb{R}}$-valued functions. Define $\bar{\mathbb{R}} = [-\infty, \infty]$.

Let $f : X \longrightarrow \bar{\mathbb{R}}$. Then $f^+, f^- : X \longrightarrow [0, \infty]$ are defined by

$$f^+(x) = \max\{f(x), 0\} \text{ and } f^-(x) = \max\{-f(x), 0\}.$$

We have $f = f^+ - f^-$ (there is no possibility of $\infty - \infty$ since if one is ∞, the other must be zero) and $|f| = f^+ + f^-$. If f is measurable, then f^+ and f^- are measurable, because for each $y \in \mathbb{R}$,

$$\{f^+ > y\} = \begin{cases} \{f > y\} & \text{if } y \geq 0, \\ X & \text{if } y < 0, \end{cases}$$

$-f$ is measurable, and $\{f^- > y\} = \{(-f)^+ > y\}$. Conversely, if f^+ and f^- are measurable, then there are sequences (φ_n) and (ψ_n) of $[0,\infty)$-valued simple functions such that $\varphi_n \uparrow f^+$ and $\psi_n \uparrow f^-$ and $\varphi_n - \psi_n \to f^+ - f^- = f$, so f is measurable as the limit of a sequence of measurable functions.

Definition. Let $f : X \longrightarrow \bar{\mathbb{R}}$ be measurable.

(a) Then $\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu$. In particular, $\int f\,d\mu$ is defined if and only if $\int f^+\,d\mu$ and $\int f^-\,d\mu$ are not both ∞.

(b) To say f is integrable ($\int$ble) means $\int f\,d\mu$ is defined and is finite.

Remark. A measurable function $f : X \longrightarrow \bar{\mathbb{R}}$ is integrable if and only if $\int f^+\,d\mu$ and $\int f^-\,d\mu$ are both finite.

Example. $\int_0^\infty \frac{\sin x}{x}\,dx$ is not defined because $\int_0^\infty \left(\frac{\sin x}{x}\right)^+ dx = \infty$ and $\int_0^\infty \left(\frac{\sin x}{x}\right)^- dx = \infty$. But

$$\int_0^{\to\infty} \frac{\sin x}{x}\,dx = \lim_{b\to\infty} \int_0^b \frac{\sin x}{x}\,dx = \frac{\pi}{2}.$$

Remark. $\int f\,d\mu = \int f(x)\,d\mu(x) = \int f(x)\mu(dx)$ where $\mu(dx)$ is the measure of the infinitesimal something or other? We write $\mu(dx)$ to distinguish $\int f(x)\,M(dx,y)$ (integration in x with respect to a family of measures parameterized by y).

2/22/21: Sane Integration II: Episode 2 - Part i

Let $(X, \mathscr{A}, \mu)$ be a measure space.

Proposition 60 (Triangle Inequality)**.** Let $f : X \longrightarrow \bar{\mathbb{R}}$ be measurable. Suppose $\int f\,d\mu$ is defined. Then $|\int f\,d\mu| \leq \int |f|\,d\mu$.

Proof. $|\int f\,d\mu| = |\int f^+\,d\mu - \int f^-\,d\mu| \leq \int f^+\,d\mu + \int f^-\,d\mu = \int f^+ + f^-\,d\mu = \int |f|\,d\mu$. □

Proposition 61. Let $f : X \longrightarrow \bar{\mathbb{R}}$ be measurable. Then f is integrable if and only if $\int |f|\,d\mu < \infty$.

Proof. f is integrable if and only if $\int f^+\,d\mu < \infty$ and $\int f^-\,d\mu < \infty$ if and only if $\int f^+\,d\mu + \int f^-\,d\mu = \int f^+ + f^-\,d\mu = \int |f|\,d\mu < \infty$. □

Proposition 62. Let $f : X \longrightarrow \bar{\mathbb{R}}$ be measurable and suppose $\int f\,d\mu$ is defined. Then $\int cf\,d\mu = c\int f\,d\mu$ for $c \in \mathbb{R}$. (If $c = \pm\infty$, then $\int cf\,d\mu$ may be undefined.)

Proof. Consider cases, based on what we know about when $f \geq 0$ and $c \geq 0$. □

Proposition 63. Let $f, g : X \longrightarrow (-\infty, \infty]$ be measurable. Suppose $\int f\,d\mu > -\infty$ and $\int g\,d\mu > -\infty$ (and are defined). Then $\int f + g\,d\mu$ is defined and $= \int f\,d\mu + \int g\,d\mu$.

Lemma 64. Let $h = u - v$, where $u : X \longrightarrow [0,\infty]$ and $v : X \longrightarrow [0,\infty)$ are measurable and $\int v\,d\mu < \infty$. Then $\int h\,d\mu$ is defined and $= \int u\,d\mu - \int v\,d\mu$.

Proof of the lemma. $h = h^+ - h^- = u - v$. For each $x \in X$, $h^+(x) = \max\{h(x), 0\} = \max\{u(x) - v(x), 0\} \leq \max\{u(x), 0\} = u(x)$. Similarly $h^-(x) = \max\{-h(x), 0\} = \max\{v(x) - u(x), 0\} \leq \max\{v(x), 0\} = v(x)$. Thus $h^+ \leq u$ and $h^- \leq v$. Since $h^- \leq v$ and $\int u\,d\mu < \infty$, we have $\int h^-\,d\mu$ is finite, so $\int h\,d\mu$ is defined and equal to $\int h^+\,d\mu 0 \int h^-\,d\mu$. Since $h^+ - h^- = u - v$ and h^- and v are never ∞, we have $h^+ + v = u + h^-$ so $\int h^+\,d\mu + \int v\,d\mu = \int u\,d\mu + \int h^-\,d\mu$. Since $\int v\,d\mu$ and $\int h^-\,d\mu$ are both finite we may subtract them from both sides to get $\int h\,d\mu = \int h^+\,d\mu - \int h^-\,d\mu = \int u\,d\mu - \int v\,d\mu$. □

Corollary 65. Let $f, g : X \longrightarrow \mathbb{R}$ be integrable. Then $f + g$ is integrable and $\int f + g\,d\mu = \int f\,d\mu + \int g\,d\mu$.

Proof of the proposition. Let $h = f+g$, $u = f^+ + g^+$, and $v = f^- + g^-$. Then $u : X \longrightarrow [0,\infty]$ and $v : X \longrightarrow [0,\infty)$ are measurable, $\int v\,d\mu = \int f^-\,d\mu + \int g^-\,d\mu$ and $h = f+g = (f^+ - f^-) + (g^+ - g^-) = (f^+ + g^+) - (f^- + g^-) = u - v$ so by the lemma $\int h\,d\mu$ and equal to $\int u\,d\mu - \int v\,d\mu = (\int f^+\,d\mu + \int g^+\,d\mu) - (\int f^-\,d\mu + \int g^-\,d\mu)$, which since $\int f^-\,d\mu$ and $\int g^-\,d\mu$ is finite we can rearrange to get $(\int f^+\,d\mu - \int f^-\,d\mu) + (\int g^+\,d\mu - \int g^-\,d\mu) = \int f\,d\mu + \int g\,d\mu$. ☐

Integrating $\mathbb{R}^d$-valued measurable functions. Let $f : X \longrightarrow \mathbb{R}^d$. Then $f(x) = (f_1(x), \dots, f_d(x))$ for all $x \in X$, where $f_k = \pi_k \circ f$ for $k = 1, \dots, d$. When is f measurable? Exactly when $f_1, \dots, f_d$ are measurable.

Proof. Let $\mathscr{H}_1 = \{(a,b] : a,b \in \mathbb{Q}, a < b\}$. Then $\sigma(\mathscr{H}_1) = \text{Borel}(\mathbb{R})$m because each element of $\mathscr{H}_1$ is a countable intersection of open subsets of $\mathbb{R}$ and each open subset of $\mathbb{R}$ is a countable union of elements of $\mathscr{H}_1$. Let $\mathscr{H}_d = \left\{\prod_{k=1}^d H_k : H_1, \dots, H_d \in \mathscr{H}_1\right\}$. Then $\sigma(\mathscr{H}_d) = \text{Borel}(\mathbb{R}^d)$ by a similar argument. Suppose $f, \dots, f_d$ are measurable. Let $H \in \mathscr{H}_d$. Then $H = \prod_{k=1}^d H_k$ for some $H_1, \dots, H_d \in \mathscr{H}_1$, so $f^{-1}[H] = \bigcap_{k=1}^d f_k^{-1}[H_k] \in \mathscr{A}$. Since $\sigma(\mathscr{H}_d) = \text{Borel}(\mathbb{R})$, it follows that f is measurable.

Conversely, suppose f is measurable. Let $k = \{1, \dots, d\}$. Let U be open in $\mathbb{R}$. For $j = 1, \dots, d$, let

$$G_j = \begin{cases} U & \text{if } j = k, \\ \mathbb{R} & \text{otherwise.} \end{cases}$$

Then $G_1, \dots, G_d$ are open in $\mathbb{R}$. Let $G = \prod_{j=1}^d G_j$. Then G is open in $\mathbb{R}^d$, Hence $f^{-1}[G] \in \mathscr{A}$. But $f^{-1}[G] = \bigcap_{j=1}^d f_j^{-1}[G_j] = f_k^{-1}[U]$. Thus $f_k^{-1}[U] \in \mathscr{A}$. This holds for each U open in $\mathbb{R}$. Hence f_k is measurable. This holds for $k = 1, \dots, d$. ☐

Definition. Assume f is measurable. To say f is integrable means that f_k is integrable for $k = 1, \dots, d$. If f is integrable, then by definition $\int f\,d\mu = \left(\int f_1,\,d\mu, \dots, \int f_d\,d\mu\right)$.

It's clear that if $f, g : X \longrightarrow \mathbb{R}^d$ are integrable and $c \in \mathbb{R}$, then $f+g$ and cf are integrable, $\int f + g\,d\mu = \int f\,d\mu + \int g\,d\mu$, and $\int cf\,d\mu = c \int f\,d\mu$. (Proof is component by component.)

Proposition 66 (Triangle inequality)**.** Let $f : X \longrightarrow \mathbb{R}^d$ be integrable. Then $|\int f\,d\mu| \le \int |f|\,d\mu$.

Proof for Euclidean norm. Let $y = \int f\,d\mu$. If $y = 0$, then $|y| = 0 - \int |f|\,d\mu$. Suppose $y \neq 0$. Then $|y| > 0$. Let $u = y/|y|$. Then $|u| = 1$ and $\langle u \mid y\rangle = \langle y \mid y\rangle/|y| = |y|^2/|y| = |y|$. Thus

$$\begin{aligned} |y| = \langle u \mid y\rangle &= \left\langle u \mid \int f\,d\mu \right\rangle \\ &= \sum_{k=1}^d u_k \int f_k\,d\mu = \int \sum_{k=1}^d u_k f_k(x)\,d\mu(x) = \int \langle u \mid f(x)\rangle\,d\mu(x) \\ &\le \int |u||f(x)|\,d\mu(x) = \int |f(x)|\,d\mu(x) \end{aligned}$$

by Cauchy-Schwartz

$$= \int |f|\,d\mu$$

since $|u| = 1$. ☐

Theorem 67 (Dominated Convergence)**.** Let (f_n) be a particular convergent sequence of measurable functions $f_n : X \longrightarrow \mathbb{R}^d$. Let $f = \lim_{n\to\infty} f_n$ (the pointwise limit). Suppose that there is a measurable function $g : X \longrightarrow [0,\infty]$ such that $\int g\,d\mu < \infty$ and for each n, $|f_n| \le g$. Then

(a) $\int |f - f_n|\,d\mu \to 0$.

(b) $\int f_n\,d\mu \to \int f\,d\mu$.

Proof. (b) follows from (a), because

$$\left| \int f \, d\mu - \int f_n \, d\mu \right| = \left| \int f - f_n \, d\mu \right| \leq \int |f - f_n| \, d\mu.$$

Now let's prove (a). For each $x \in X$, we have $|f_n(x)| \leq g(x)$ for each n and $f_n(x) \to f(x)$ as $n \to \infty$. Hence $|f(x)| \leq g(x)$, so $|f(x) - f_n(x)| \leq |f(x)| + |f_n(x)| \leq 2g(x)$ for each n and each x. Let $h_n = 2g - |f - f_n|$. Then $h_n \to 2g$ pointwise and $h_n \geq 0$. Hence by Fatou,

$$\begin{aligned} \int 2g \, d\mu \leq \liminf_{n\to\infty} \int h_n \, d\mu &= \liminf_{n\to\infty} \left(\int 2g \, d\mu - \int |f - f_n| \, d\mu \right) \\ &= \int 2g \, d\mu + \liminf_{n\to\infty} \left(- \int |f - f_n| \, d\mu \right) \\ &= \int 2g \, d\mu - \limsup_{n\to\infty} \int |f - f_n| \, d\mu \end{aligned}$$

so $0 \leq -\limsup_{n\to\infty} \int |f - f_n| \, d\mu$ so $\limsup_{n\to\infty} \int |f - f_n| \, d\mu \leq 0 \leq \liminf_{n\to\infty} \int |f - f_n| \, d\mu$ so $\lim_{n\to\infty} \int |f - f_n| \, d\mu$ exists and equals 0. $\square$

Integrals on complex-valued functions Think of $\mathbb{C}$ as $\mathbb{R}^d$. Remember to check that $\int cf \, d\mu = c \int f \, d\mu$ even when $c \in \mathbb{C}$.

2/26/21: Sane Integration II: Episode 2 - Part ii

Abstract change of variable formula Let $(X, \mathscr{A}, \mu)$ be a measure space, let $(Y, \mathscr{B})$ be a measurable space, and let $T : X \longrightarrow Y$ be $\mathscr{A}/\mathscr{B}$-measurable. Define ν on $\mathscr{B}$ by $\nu(B) = \mu(T^{-1}[B])$. Then

(a) ν is a measure on $\mathscr{B}$.

(b) For each measurable function $f : Y \longrightarrow [0, \infty]$, $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$.

(c) For each measurable function $f : Y \longrightarrow \bar{\mathbb{R}}$, $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$, in the sense that if either integral is defined then so is the other, and in this case they are equal.

(d) For each measurable function $f : Y \longrightarrow \bar{\mathbb{R}}$, f is ν-integrable if and only if $f \circ T$ is μ-integrable, in which case $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$.

Proof. (a) Clearly $\nu : \mathscr{B} \longrightarrow [0, \infty]$. ($\nu$ is defined on all of $\mathscr{B}$ because T is $\mathscr{A}/\mathscr{B}$-measurable.) $\nu(\varnothing) = \mu(T^{-1}[\varnothing]) = 0$. Let (B_n) be a disjoint sequence on $\mathscr{B}$. Let $B = \bigcup_n B_n$. Then

$$\nu(B) = \mu(T^{-1}[B]) = \mu\left(T^{-1} \left[\dot{\bigcup_n} B_n \right] \right) = \mu\left(\dot{\bigcup_n} T^{-1}[B_n] \right) = \sum_n \mu(T^{-1}[B_n]) = \sum_n \nu(B_n).$$

Thus ν is a measure on $\mathscr{B}$.

(b) Case 1: Let $B \in \mathscr{B}$. Then

$$\int_Y 1_B \, d\nu = \nu(B) = \mu(T^{-1}[B]) = \int_X 1_{T^{-1}[B]}(x) \, d\mu(x) = \int_X 1_B(T(x)) \, d\mu(x) = \int_X 1_B \circ T \, d\mu.$$

Case 2: Let $\varphi : Y \longrightarrow [0, \infty)$ be $\mathscr{B}$-simple. Then $\varphi = \sum_{k=1}^n c_k 1_{B_k}$ for suitable $c_1, \ldots, c_n \in [0, \infty)$, and $B_1, \ldots, B_n \in \mathscr{B}$. Then

$$\begin{aligned} \int_Y \varphi \, d\nu = \int_Y \sum_{k=1}^n c_k 1_{B_k} \, d\nu = \sum_{k=1}^n c_k \int_Y 1_{B_k} \, d\nu = \sum_{k=1}^n c_k \int_X 1_{B_k} \circ T \, d\mu \\ = \sum_{k=1}^n c_k \int_X 1_{B_k}(T(x)) \, d\mu(x) = \int_X \sum_{k=1}^n c_k 1_{B_k}(T(x)) \, d\mu(x) \\ = \int_X \varphi(T(x)) \, d\mu(x) = \int_X \varphi \circ T \, d\mu. \end{aligned}$$

Case 3: Consider any measurable function $f : Y \longrightarrow [0, \infty]$. Then there is an increasing sequence (φ_n) of $\mathscr{B}$-simple functions $\varphi_n : X \longrightarrow [0, \infty)$ such that for each $y \in Y$, $\varphi_n(y) \uparrow f(y)$. Then for each $x \in X$, $\varphi_n(T(x)) \uparrow f(T(x))$. Thus by the monotone convergence theorem,

$$\int_Y \varphi_n \, d\nu \uparrow \int_Y f \, d\nu$$

and

$$\int_X \varphi_n \circ T \, d\mu = \int_X \varphi_n(T(x)) \, d\mu(x) \uparrow \int_X f(T(x)) \, d\mu(x) = \int_X f \circ T \, d\mu.$$

But $\int_Y \varphi_n \, d\nu = \int_X \varphi_n \circ T \, d\mu$ by Case 2, hence $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$. This proves (b).

(c) follows from (b) by considering f^+ and f^-.

(d) follows from (c).

$\square$

Notation. We will write $\nu = \mu \circ T^{-1}$ or better, $\nu = T_*(\mu)$.

Application of the abstract change of variable formula to probability theory Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $(\mathbf{X}, \mathscr{A})$ be a measurable space. Let X be a random variable in $\mathbf{X}$. (This means X is $\mathscr{F}/\mathscr{A}$-measurable function form Ω into $\mathbf{X}$.) Let $\mu = \text{law}(X)$. (This means that μ is the measure on $\mathscr{A}$ defined by $\mu(A) = P(X^{-1}[A])$, or $\mu = X_*(P)$.) Let $f : \mathbf{X} \longrightarrow \bar{\mathbb{R}}$ be $\mathscr{A}$-measurable. Then $E[f(X)] = \int_{\mathbf{X}} f \, d\mu$, in the sense that if either is defined then so is the other and they are equal. $f(X)$ means $f \circ X$. $E(Y)$ means $\int_\Omega Y \, dP$.

Abstract indefinite integrals. Let $(X, \mathscr{A}, \lambda)$ be a measure space. Let $f : X \longrightarrow [0, \infty]$ be measurable. The indefinite integral of f with respect to λ is the function μ of $\mathscr{A}$ defined by $\mu(A) = \int_A f \, d\lambda = \int 1_A f \, d\lambda$. Note that μ is a measure on $\mathscr{A}$.

Proof. $\mu : \mathscr{A} \longrightarrow [0, \infty]$. $\mu(\varnothing) = \int 1_\varnothing f \, d\lambda = \int 0 \, d\lambda = 0$. Let (A_n) be a disjoint sequence in $\mathscr{A}$ and let $A = \bigcup_n A_n$. Then $1_A = \sum_n 1_{A_n}$. Now $\mu(A) = \int 1_A f \, d\lambda = \int \sum_n 1_{A_n} f \, d\lambda = \sum_n \int 1_{A_n} f \, d\lambda = \sum_n \mu(\lambda)$. (Note that we can exchange the sum with the integral since f is nonnegative.) Thus μ is a measure on $\mathscr{A}$. $\square$

Let $g : X \longrightarrow [0, \infty]$ be $\mathscr{A}$-measurable. Then $\int g \, d\mu = \int g f \, d\lambda$.

Proof. Case 1: Let $A \in \mathscr{A}$. Then $\int 1_A \, d\mu = \mu(A) = \int 1_A f \, d\lambda$.

Case 2: Let $\varphi : X \longrightarrow [0, \infty)$ be $\mathscr{A}$-simple. Then $\int \varphi \, d\mu = \int \varphi f \, d\lambda$ by Case 1 and linearity ($\varphi = \sum_{k=1}^n c_k 1_{A_k}$).

Case 3: Since $g : X \longrightarrow [0, \infty]$ is measurable there is an increasing sequence (φ_n) of simple functions $\varphi_n : X \longrightarrow [0, \infty)$ such that $\varphi_n \uparrow g$. Then $\int \varphi_n \, d\mu \uparrow \int g \, d\mu$ and $\varphi_n f \uparrow g f$, so $\int \varphi_n d \, d\lambda \uparrow \int g f \, d\lambda$ (by MCT). But for each n, by Case 2, $\int \varphi_n \, d\mu = \int \varphi_n f \, d\lambda$. Hence, letting $n \to \infty$, we get $\int g \, d\mu = \int g f \, d\lambda$. $\square$

Notation. $d\mu = f \, d\lambda$, $\mu = f \cdot \lambda$. We say f is a density for μ with respect to λ.

3/1/21: Measures in terms of integrals in terms of measures

Definition. Let $(X, \mathscr{A}, \lambda)$ be a measure space.

(a) To say λ is finite means $\lambda(X) < \infty$.

(b) To say λ is σ-finite means $X = \bigcup_{n \in \mathbb{N}} X_n$ for some sequence (X_n) in $\mathscr{A}$ such that $\lambda(X_n) < \infty$ for each n.

(c) To say λ is semifinite means for each $A \in \mathscr{A}$, if $\lambda(A) = \infty$, then there exists $F \in \mathscr{A}$ such that $F \subseteq A$ and $0 < \lambda(F) < \infty$.

Example. (a) The counting measure on $\mathbb{Z}$ is σ-finite.

(b) The Lebesgue measure is σ-finite.

(c) The counting measure on $\mathbb{R}$ is semifinite but not σ-finite.

Remark. Finite $\implies$ σ-finite $\implies$ semifinite.

Proof. The first is obvious. Suppose λ is σ-finite. Let $(X_n) \in \mathscr{A}$ such that $X = \bigcup_n X_n$, $\lambda(X_n) < \infty$ for each n. Let $A \in \mathscr{A}$ such that $\lambda(A) = \infty$. Now $A = A \cap X = A \cap \bigcup_n X_n = \bigcup_n (A \cap X_n)$ so $\infty = \lambda(A) \leq \sum_n \lambda(A \cap X_n)$. Hence $\lambda(A \cap X_n) > 0$ for some n. But $A \cap X_n \subseteq X_n$, so $\lambda(A \cap X_n) \leq \lambda(X_n) < \infty$. Thus we can take $F = A \cap X_n$. $\square$

Remark. Let $(X, \mathscr{A}, \lambda)$ be a measure space. Let $f, g : X \longrightarrow [0, \infty]$ be measurable. Define μ, ν on $\mathscr{A}$ by

$$\mu(A) = \int_A f \, d\lambda \text{ and } \nu(A) = \int_A g \, d\lambda.$$

(a) If $f \leq g$ λ-almost everywhere, then $\mu \leq \nu$.

(b) If $f = g$ λ-almost everywhere, then $\mu = \nu$.

Question. If $\mu \leq \nu$, is $f \leq g$ λ-almost everywhere? (Similar question for "=".)

Answer. Either λ is semifinite or not. Case 1: Suppose λ is semifinite. Suppose $\mu \leq \nu$. We wish to show that $f \leq g$ λ-almost everywhere. Let $A = \{f > g\}$. We wish to show $\lambda(A) = 0$. Suppose not. Then $\lambda(A) > 0$, so since λ is semifinite, there exists $F \in \mathscr{A}$ such that $F \subseteq A$ and $0 < \lambda(F) < \infty$. Now let $F_{r,s} = \{f > r\} \cap \{s > g\} \cap F$ so $F = \bigcup_{r,s \in \mathbb{Q},\, r > s > 0} F_{r,s}$, since for every $x \in F$, $f(x) > g(x)$, so there exist $r, s \in \mathbb{Q}$ such that $f(x) > r > s > g(s)$. Then $0 < \lambda(F) \leq \sum_{r,s \in \mathbb{Q}, r > s > 0} \lambda(F_{r,s})$. Hence there are $r, s \in \mathbb{Q}$, $r > s > 0$, and $\lambda(F_{r,s}) > 0$. Then $\mu(F_{r,s}) \leq \nu(F_{r,s})$ by assumption but

$$\mu(F_{r,s}) = \int_{F_{r,s}} f \, d\lambda \geq \int_{F_{r,s}} r \, d\lambda = r\lambda(F_{r,s}) > s\lambda(F_{r,s})$$

since $0 < \lambda(F_{r,s}) < \infty$)

$$= \int_{F_{r,s}} s \, d\lambda \geq \int_{F_{r,s}} g \, d\lambda = \nu(F_{r,s})$$

a contradiction. Therefore $\lambda(A) = 0$.

Case 2: Suppose λ is not semifinite. Let's make a malicious choice of f and g. Since λ is not semifinite, there exists $B \in \mathscr{A}$ such that $\lambda(B) = \infty$ and for each $F \in \mathscr{A}$, if $F \subseteq B$ then either $\lambda(F) = 0$ or $\lambda(F) = \infty$. Then for each $A \in \mathscr{A}$, $\lambda(A \cap B) = 0$ or ∞. Then for each $A \in \mathscr{A}$,

$$\int_A \underbrace{2 \cdot 1_B}_{f}, d\lambda = 2\lambda(A \cap B) = \lambda(A \cap B) = \int_A \underbrace{1_B}_{g} \, d\lambda.$$

But $\lambda(2 \cdot 1_B > 1_B) = \lambda(B) = \infty$.

Definition. Let λ, μ be measures on a measurable space $(X, \mathscr{A})$. To say μ is absolutely continuous with respect to λ ($\mu \ll \lambda$) means for each $A \in \mathscr{A}$, if $\lambda(A) = 0$ then $\mu(A) = 0$.

Example. Let $(X, \mathscr{A}, \lambda)$ be a measure space. Let $f : X \longrightarrow [0, \infty]$ be measurable. Define μ on $\mathscr{A}$ by $\mu(A) = \int_A f \, d\lambda$. Then $\mu \ll \lambda$.

Question. Suppose λ, μ are measures on a measurable space $(X, \mathscr{A})$ and $\mu \ll \lambda$. Does there exist a measurable function $f : X \longrightarrow [0, \infty]$ such that for each $A \in \mathscr{A}$, $\mu(A) = \int_A f \, d\lambda$?

Answer. Yes, provided λ is σ-finite. This is the Radon-Nikodym theorem. (μ need not be σ-finite, though most books assume it is.) No in general. Consider λ the counting measure on $\mathbb{R}$ (restricted to the Lebesgue σ-algebra) and μ the Lebesgue measure on $\mathbb{R}$.

In Probability Theory. Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let X be a random variable in a measurable space $(\mathbf{X}, \mathscr{A})$. Let λ be a measure on $\mathscr{A}$. Let $\mu = \text{law}(X)$. $(P(X^{-1}[A]) = P(x \in A) = \mu(A).)$ To say f is a probability density for X with respect to λ means for each $A \in \mathscr{A}$, $\mu(A) = \int_A f\, d\lambda$. In this case, for each measurable function $g : \mathbf{X} \longrightarrow \bar{\mathbb{R}}$,

$$\int_\Omega g(X(\omega))\, dP(\omega) = E[g(X)] = \int_{\mathbf{X}} g\, d\mu = \int_{\mathbf{X}} fg\, d\lambda$$

where if any of these integrals exists, then they are all equal. (Because of this equality, it is sometimes written that $\mu = f\, d\lambda$.)

If $\mathbf{X} = \bar{\mathbb{R}}$, then we would usually take λ to be the Lebesgue measure.

We can write μ in terms of its absolutely continuous part μ_a and its singular part μ_s.

$$\mu = \mu_a + \mu_s,\ \mu_a = f \cdot \lambda,\ \mu_s = \mu_d - (\mu_s - \mu_d)$$

Example. Toss a fair coin infinitely many times. Let

$$X_n = \begin{cases} 1 & \text{if we get heads on the } n\text{-th toss} \\ -1 & \text{otherwise} \end{cases}$$

Let $X = 1/2 + \sum_{n=1}^\infty X_n/3^n$. $P(X \in \mathcal{C}) = 1$ (where $\mathcal{C}$ is the Cantor set). Let $\mu = \text{law}(X)$. μ lives on $\mathcal{C}$, so μ is singular with respect to Lebesgue measure. $\mu(\{x\}) = 0$ for each $x \in \mathcal{C}$. (This is, morally, the uniform measure on $\mathcal{C}$, and in fact the bijection from $\mathcal{C}$ to $\{0,1\}^{\mathbb{N}}$ to $[0,1]$ transforms μ into the Lebesgue measure.)

Why do we want to be able to integrate wild functions? or; Insane Integration - Prologue Pick a number at random in $[0,1)$. $\Omega = [0,1)$, $\mathscr{F} = \text{Borel}[0,1)$, $P =$ Borel-Lebesgue measure on $[0,1)$ (i.e. the restriction of Lebesgue measure to the Borel σ-field). For $n = 1, 2, 3, \ldots$ define X_n on Ω by $X_n(\omega) =$ the n-th bit in the standard binary expansion of ω (that is, for diadic fractions which have two binary expansions, pick the one that does not contain $111\ldots$). Then

$$X_1(\omega) = \begin{cases} 0 & \omega \in [0, \frac{1}{2}) \\ 1 & \omega \in [\frac{1}{2}, 1) \end{cases} \qquad X_2(\omega) = \begin{cases} 0 & \omega \in [0, \frac{1}{4}) \cup [\frac{1}{2}, \frac{3}{4}) \\ 1 & \omega \in [\frac{1}{4}, \frac{1}{2}) \cup [\frac{3}{4}, 1) \end{cases}$$

etc. and $X_1(\omega) + \cdots + X_n(\omega)$ is the number of 1's in the first n bits of the standard binary expansion of ω. Define

$$Y_n(\omega) = \frac{X_1(\omega) + \cdots + X_n(\omega)}{n},$$

the proportion of 1's in the first n bits of the standard binary expansion of ω. Let $G = \{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega)$ exists$\}$. Define $Y : G \longrightarrow [0,1]$ by $Y(\omega) = \lim_{n\to\infty} Y_n(\omega)$. What can we say about Y? $P(Y = 1/2) = 1$. (We will prove this eventually.) $E(Y) = 1/2$.

Let $I = (a,b)$ (or $[0,b)$) be a relatively open subinterval of $[0,1)$. Let $n \in \mathbb{N}$ such that for some $n \in \mathbb{N}$, $a \le m/2^n < (m+1)/2^n < b$.

Let $\omega = m/2^n$. Then $\{Y(\omega') : \omega \le \omega' \le \omega + 1/2^n\} = [0,1]$.

3/3/21

Hints for Extra Problem 19 Let $a_n \in (-1, \infty)$ and $b_n \in (0, \infty)$. Suppose $b_n \to \infty$ and $a_n b_n \to c \in \mathbb{R}$. Then $(1 + a_n)^{b_n} \to e^c$.

Proof. As $h \to 0$,

$$(1+h)^{\frac{1}{h}} = \left(e^{\ln(1+h)}\right)^{\frac{1}{h}} = e^{\frac{\ln(1+h)}{h}} = e^{\frac{\ln(1+h) - \ln(1)}{h}} \to e^{\ln'(1)} = e^{\frac{1}{1}} = e.$$

Let $x \in \mathbb{R}$. Then as $h \to 0$, $(1+hx)^{\frac{1}{h}} = \left((1+hx)^{\frac{1}{hx}}\right)$ (provided $x \neq 0$), which since $hx \to 0$ goes to $(e)^x = e^x$. If $x = 0$, then for all h, $(1+hx)^{\frac{1}{h}} = 1^{\frac{1}{h}} = 1 = e^0 = e^x$. Let $G = \{n \in \mathbb{N} : a_n \neq 0\}$. If G is not bounded above, then for $n \in G$, $(1+a_n)^{b_n} = \left((1+a_n)^{\frac{1}{a_n}}\right)^{a_n b_n} \xrightarrow[n\to\infty]{n \in G} (e)^c = e^c$.

Finish this proof to make sure it is complete. $\square$

Remark. Take the definition of e to be the unique real number such that $\frac{d}{dx}e^x = e^x$, or the unique number such that $\ln(e) = 1$.

Back to the density example We have for each $n \in \mathbb{N}$, for each $m \in \{0, 1, \dots, 2^n - 1\}$, for each nonempty closed set $L \subseteq [0,1]$, there exists $w \in [\frac{m}{2^n}, \frac{m+1}{2^n})$ such that the set of cluster points in the sequence $(Y_k(\omega))$ is L.

Back to independence On 4M, 1II, we proved the for a family of independent π-systems.

Theorem (The Grouping Theorem, Version 1.)**.** Let $(\mathscr{G}_i)_{i \in I}$ be an independent family of π-systems. Let $(I_j)_{j \in J}$ be a disjoint family of nonempty subsets of I. Let $\mathscr{H}_j = \pi\left(\bigcup_{i \in I_j} \mathscr{G}_i\right)$ Then the family of π-systems $(\mathscr{H}_j)_{j \in J}$ is independent.

Reminder: Since each $\mathscr{G}$ is a π-system,

$$\pi\left(\bigcup_{i \in I_j} \mathscr{G}_i\right) = \left\{\bigcap_{i \in I^\circ} G_i : \varnothing \neq I^\circ \text{ finite } \subseteq I_j, G_i \in \mathscr{G}_i \text{ for each } i \in I_j\right\}.$$

Lemma 68 (Lemma 1)**.** Let $\mathscr{H}_1, \dots, \mathscr{H}_n$ be independent π-systems. Let $k \in \{1, \dots, n\}$. For $j = 1, \dots, n$, let

$$\mathscr{H}_j' = \begin{cases} \mathscr{H}_j & \text{if } j \neq k \\ \sigma(\mathscr{H}_k) & \text{if } j = k. \end{cases}$$

Then the family of π-systems $\mathscr{H}_1', \dots, \mathscr{H}_n'$ is independent.

Lemma 69 (Lemma 0)**.** Let $A \in \mathscr{F}$. Let $\mathscr{B} = \{B \in \mathscr{F} : A, B \text{ are independent}\}$. Then $\mathscr{B}$ is a λ-system.

Proof. If $P(A) = 0$, then $\mathscr{B} = \mathscr{F}$. Suppose $P(A) \neq 0$. Define Q on $\mathscr{F}$ by

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

Then Q is a probability measure on $\mathscr{F}$. Then for each $B \in \mathscr{F}$, we have A and B are independent if and only if $P(AB) = P(A)P(B)$ if and only if $P(B) = Q(B)$. Then $\mathscr{B} = \{B \in \mathscr{F} : P(B) = Q(B)\}$. Therefore $\mathscr{B}$ is a λ-system. $\square$

Corollary 70 (Corollary of Lemma 0 (cf. 69))**.** Let $A \in \mathscr{F}$. Let $\mathscr{H}$ be a π-system contained in $\mathscr{F}$ such that $\{A\}$ and $\mathscr{H}$ are independent. Then $\{A\}$ and $\sigma(\mathscr{H})$ are independent.

Proof. Let $\mathscr{B} = \{B \in \mathscr{F} : A, B \text{ are independent}\}$. By Lemma 0, $\mathscr{B}$ is a λ-system. By assumption $\mathscr{H} \subseteq \mathscr{B}$. Hence by the π-λ Theorem, $\sigma(\mathscr{H}) \subseteq \mathscr{B}$. $\square$

Proof of Lemma 1 (cf. 68). Let $\varnothing \neq J \subseteq \{1, \dots, n\}$ and let $H_j \in \mathscr{H}_j'$ for each $j \in J$. We wish to show that $P\left(\bigcap_{j \in J} H_j\right) = \prod_{j \in J} P(H_j)$. If $k \notin J$, then this holds because then $H_j \in \mathscr{H}_j$ for each $j \in J$. Suppose $k \in J$. If $J = \{k\}$, then there is nothing to prove. Suppose $J \neq \{k\}$. Then $J' \coloneqq J \setminus \{k\} \neq \varnothing$. Let $A = \bigcap_{j \in J'} H_j$. Then A and $\mathscr{H}_k$ are independent:

Let $H \in \mathscr{H}_k$. Let

$$G_j = \begin{cases} H_j & \text{if } j \in J' \\ H & \text{if } j = k. \end{cases}$$

Then

$$\begin{aligned}
P(AH) &= P\left(\bigcap_{j\in J} G_j\right) = \prod_{j\in J} P(G_j) \\
&= \left(\prod_{j\in J'} P(G_j)\right) P(G_k) = \left(\prod_{j\in J'} P(G_j)\right) P(H) = P\left(\bigcap_{j\in J'} H_j\right) P(H) \\
&= P(A)P(H).
\end{aligned}$$

Then A and $\mathscr{H}_k' = \sigma(\mathscr{H}_k)$ are independent. Thus for each $H_k \in \mathscr{H}_k'$,

$$P\left(\bigcap_{j\in J} H_j\right) = P(AH_k) = P(A)P(H_k) = \left(\prod_{j\in J'} P(H_j)\right) P(H_k) = \prod_{j\in J} P(H_j).$$

□

Lemma 71 (Lemma 2)**.** Let $\mathscr{H}_1, \ldots, \mathscr{H}_n$ be independent π-systems. Then

$$\sigma(\mathscr{H}_1), \ldots, \sigma(\mathscr{H}_n)$$

are independent.

Proof. Apply Lemma 1 (cf. 68) n times. □

Theorem 72. Let $(\mathscr{H})_{j\in J}$ be an independent family if π-systems. Then $(\sigma(\mathscr{H}_j))_{j\in J}$ are independent.

Proof. A family of π-systems is independent if and only if each of its finite subfamilies is independent. Hence this follows from Lemma 2 (cf 71). □

Theorem 73 (The Full Grouping Theorem)**.** Let $(\mathscr{G}_i)_{i\in I}$ be a family of independent π-systems (or σ-fields. Let $(I_j)_{j\in J}$ be a disjoint family of subsets of I. For each $j \in J$, let $\mathscr{F}_j = \sigma\left(\bigcup_{i\in I_j} \mathscr{G}_i\right)$. Then the family $(\mathscr{F}_j)_{j\in J}$ is independent.

Proof. From version 1 of the grouping theorem, we know $(\mathscr{H}_j)_{j\in J}$ is independent, where $\mathscr{H}_j = \pi\left(\bigcup_{i\in I_j} \mathscr{G}_i\right)$. Hence by the previous theorem, $(\sigma(\mathscr{H}_j))_{j\in J}$ is independent. But for each $j \in J$, $\sigma(\mathscr{H}_j) = \sigma\left(\bigcup_{i\in I_j} \mathscr{G}_i\right)$. (Reason: Let $\mathscr{E}$ be any collection of subsets of Ω. Then $\mathscr{E} \subseteq \pi(\mathscr{E}) \subseteq \sigma(\mathscr{E})$ so $\sigma(\mathscr{E}) = \sigma(\pi(\mathscr{E}))$.) □

Notation. Let $(\mathbf{X}, \mathscr{A})$ be a measurable space. Let X be a random variable in $\mathbf{X}$. Then

$$\sigma(X) = \{X^{-1}[A] : A \in \mathscr{A}\}.$$

$\sigma(X)$ is a σ-field on Ω. Furthermore, X is $\sigma(X)/\mathscr{A}$-measurable and for a σ-field $\mathscr{G}$ on Ω, if X is $\mathscr{G}/\mathscr{A}$-measurable. then $\sigma(X) \subseteq \mathscr{G}$. Then $\sigma(X)$ is the smallest σ-field on Ω with respect to which X is measurable. We call $\sigma(X)$ the σ-field generated by X.

3/5/21

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

Definition. Let $((\mathbf{X}_i, \mathscr{A}_i))_{i\in I}$ be a family of measurable spaces. For each $i \in I$, let X_i be a random variable in $(\mathbf{X}_1, \mathscr{A}_i)$. (So X_i is a $\mathscr{F}/\mathscr{A}_i$-measurable function from Ω to $\mathbf{X}_i$.) Then the σ-field generated by the family of random variables $(X_i)_{i\in I}$ is

$$\sigma(X_i : i \in I) = \sigma(\{X_i^{-1}[A] : i \in I, A \in \mathscr{A}_i\})$$

Definition. Let $(X_i)_{i\in I}$ be a family of random variables. To say $(X_i)_{i\in I}$ is independent means that the family of σ-fields $(\sigma(X_i))_{i\in I}$ they generate is independent.

Definition. Let $((\mathbf{X}_i, \mathscr{A}_i))_{i\in I}$ be a family of measurable spaces. Let $\mathbf{X} = \prod_{i\in I} \mathbf{X}_i$. Then $\bigotimes_{i\in I} \mathscr{A}_i = \sigma(\{\pi_i^{-1}[A] : i \in I, A \in \mathscr{A}_i\})$. Also $\bigodot_{i\in I} \mathscr{A}_i = \left\{\bigcap_{i\in I_0} \pi_i^{-1}[A_i] : \varnothing \neq I_0 \text{ finite} \subseteq I, A_i \in \mathscr{A}_i \text{ for each } i \in I_0\right\}$ (i.e. the measurable rectangles in $\mathbf{X}$).

Note that $\bigotimes_{i\in I} \mathscr{A}_i = \sigma\left(\bigodot_{i\in I} \mathscr{A}_i\right)$.

Also, $\bigcap_{i\in I_0} \pi_i^{-1}[A_i] = \prod_{i\in I} B_i$ where

$$B_i = \begin{cases} A_1 & \text{if } i \in I_0 \\ \mathbf{X_i} & \text{if } i \in I \setminus I_0. \end{cases}$$

Fact. $\bigotimes_{i\in I} \mathscr{A}_i$ is the unique σ-field $\mathscr{B}$ and $\mathbf{X}$ such that for each measurable space $(E, \mathscr{E})$, for each map $f : E \longrightarrow \mathbf{X}$, f is $\mathscr{E}/\mathscr{B}$-measurable if and only if f_i is $\mathscr{E}/\mathscr{A}_i$-measurable for all $i \in I$.

Fact. Suppose $\mathbf{X}_i$ is a measurable topological space for each $i \in I$ and $\mathscr{A}_i = \text{Borel}(\mathbf{X}_i)$ for all $i \in I$. Let $\mathscr{A} = \bigotimes_{i\in I} \mathscr{A}_i$ and let $\mathscr{B} = \text{Borel}(\mathbf{X})$. Then $\mathscr{A} \subseteq \mathscr{B}$. If I is countable and each $\mathbf{X}_i$ is second countable, then $\mathscr{A} = \mathscr{B}$. Thus "YES" in $\mathbb{R}^{\mathbb{N}}$ but "NO" in $\mathbb{R}^{\mathbb{R}}$. (For any $x \in \mathbb{R}^{\mathbb{R}}$, $\{x\}$ is closed in $\mathbb{R}^{\mathbb{R}}$, so $\{x\} \in \text{Borel}(\mathbb{R}^{\mathbb{R}})$, but $\{x\} \notin \text{Borel}(\mathbb{R})^{\otimes\mathbb{R}}$.

Fact. For each $A \in \bigotimes_{i\in I} \mathscr{A}_i$, there exists I_1 countable $\subseteq I$ and $B \in \bigotimes_{i\in I_1} \mathscr{A}_i$ such that $A = \pi_{I_i}^{-1}[B]$ where $\pi_{I_i} : \mathbf{X} \longrightarrow \prod_{i\in I_1} \mathbf{X}_i$ is defined by $\pi_{I_1}((x_i)_{i\in I}) = (x_i)_{i\in I_1}$.

Now for each $i \in I$, let X_i be a random variable in $(\mathbf{X}_i, \mathscr{A}_i)$. Define $X : \Omega \longrightarrow \mathbf{X}$ by $X(\omega) = (X_i(\omega))_{i\in I}$. $\sigma(X) = \{X^{-1}[A] : A \in \mathscr{A}\} = \sigma(X_i : i \in I)$ (where $\mathscr{A} = \bigotimes_{i\in I} \mathscr{A}_i$). Checking this is straightforward.

Theorem 74 (The Grouping Theorem for Random Variables)**.** Suppose the family of random variables $(X_i)_{i\in I}$ is independent. Let $(I_j)_{j\in J}$ be a disjoint family of subsets of I. For each $j \in J$, let $\mathbf{X}'_j = \prod_{i\in I_j} \mathbf{X}_i$ and $\mathscr{A}'_j = \bigotimes_{i\in I_j} \mathscr{A}_i$, and define $X'_j : \Omega \longrightarrow \mathbf{X}'_j$ be $X'_j(\omega) = (X_i(\omega))_{i\in I_j}$. Then the family of random variables $(X'_j)_{j\in J}$ is independent.

Proof. By assumption, the family of σ-fields of $(\sigma(X_i) : i \in I)$ is independent. Let $\mathscr{F}_j = \sigma(X_i : i \in I_j)$. Note that $\mathscr{F}_j = \sigma\left(\bigcup_{i\in I_j} \sigma(X_i)\right)$. Hence, by the grouping theorem for σ-fields, the family of σ-fields $(\mathscr{F}_j)_{j\in J}$ is independent. But for each $j \in J$, $\mathscr{F}_j = \sigma(X'_j)$, by the preceding discussion. In other words, the family of random variables $(X'_j)_{j\in J}$ is independent. □

Remark. Random variables $X_i, \ldots, X_n$ are independent if and only if $P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_i) \cdots P(X_n \in A_n)$ for all $A_1 \in \mathscr{A}_1, \ldots, A_n \in \mathscr{A}_n$. We don't have to consider subfamilies of $X_1, \ldots, X_n$ because any A_i can be taken to be $\mathbf{X}_i$.

Proposition 75. Suppose the family of random variables $(X_i)_{i\in I}$ is independent. For each $i \in I$, let $f_i : \mathbf{X}_i \longrightarrow \mathbf{Y}_i$ be $\mathscr{A}_i/\mathscr{B}_i$-measurable, where $(\mathbf{Y}_i, \mathscr{B}_i)$ is another measurable space, and let $Y_i = f_i \circ X_i$. Then the family of random variables $(Y_i)_{i\in I}$ is independent.

Proof. Let $i \in I$ and $E \in \sigma(Y_i)$. Then $E = Y_i^{-1}[B]$ for some $B \in \mathscr{B}_i$. Hence, putting $A = f_i^{-1}[B] \in \mathscr{A}$, $E = (f_i \circ X_i)^{-1}[B] = X_i^{-1}[f_i^{-1}[B]] = X_i^{-1}[A] \in \sigma(X_i)$. □

Example. Let X_1, X_2, X_3 be independent real random variables. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ be Borel functions. Then the two random variables $f(X_1, X_2)$ and $g(X_3)$ are independent.

Kolmogorov's Zero-One Law. Let $X_1, X_2, X_3, \ldots$ be a sequence of independent random variables. Let $\mathscr{G}_n = \sigma(X_{n+1}, X_{n+2}, X_{n+3}, \ldots)$ and let $\mathscr{T} = \bigcap_n \mathscr{G}_n$. ($\mathscr{T}$ is the tail σ-field of $X_1, X_2, X_3, \ldots$). Let $A \in \mathscr{T}$. Then $P(A) = 0$ or 1.

Proof. Let $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$. By the , $\mathscr{F}_n$ and $\mathscr{G}_n$ are independent. But $\mathscr{T} \subseteq \mathscr{G}_n$. Hence $\mathscr{F}_n$ and $\mathscr{T}$ are independent. Let $\mathscr{F}_* = \bigcup_n \mathscr{F}_n$. Notice that $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \mathscr{F}_3 \subseteq \ldots$ Hence $\mathscr{F}_*$ is a field. In particular, $\mathscr{F}_*$ is a π-system. Now $\mathscr{F}_*$ and $\mathscr{T}$ are independent. Hence $\sigma(\mathscr{F}_*)$ and $\mathscr{T}$ are independent.

But $\mathscr{T} \subseteq \mathscr{G}_n = \sigma(X_{n+1}, X_{n+2}, \dots) \subseteq \sigma(X_1, X_2, \dots) = \sigma(\mathscr{F}_*)$. Thus for all $B \in \sigma(X_1, X_2, X_3, \dots)$, A and B are independent. In particular, A and A are independent so $P(A) = P(A \cap A) = P(A)P(A)$. So $P(A)(P(A) - 1) = 0$ so $P(A) = 0$ or $P(A) - 1 = 0$. $\square$

3/8/21

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

Proposition 76. Let $X, Y : \Omega \longrightarrow [0, \infty]$ be independent random variables. Then $E(XY) = E(X)E(Y)$.

Proof. Step 1: Suppose $X = 1_A$ and $Y = 1_B$ where $A, B \in \mathscr{F}$. Then $E(XY) = E(1_A 1_B) = E(1_{AB}) = P(AB) = P(A)P(B) = E(1_A)E(1_B) = E(X)E(Y)$ (since X, Y independent means A, B independent events).

Step 2: Suppose X and Y are simple. Let $x_1, \dots, x_m$ be an enumeration of $X[\Omega]$ and let $y_1, \dots, y_n$ be an enumeration of $Y[\Omega]$. Let $A_j = \{X = x_j\}$ for $j = 1, \dots, m$ and let $B_k = \{X = y_k\}$ for $k = 1, \dots, n$. Then $X = \sum_{j=1}^m x_j 1_{A_j}$ and $Y = \sum_{k=1}^n y_k 1_{B_k}$. Now A_j and B_k are independent for all j and k.

$$\begin{aligned}
E(XY) &= E\left(\left(\sum_{j=1}^m x_j 1_{A_j}\right)\left(\sum_{k=1}^n y_k 1_{B_k}\right)\right) = E\left(\sum_{j=1}^m \sum_{k=1}^n x_j y_k 1_{A_j} 1_{B_k}\right) \\
&= \sum_{j=1}^m \sum_{k=1}^n x_j y_k E(1_{A_j} 1_{B_k}) = \sum_{j=1}^m \sum_{k=1}^n x_j y_k E(1_{A_j}) E(1_{B_k}) \\
&= \sum_{j=1}^m x_j E(1_{A_j}) \sum_{k=1}^n y_k E(1_{B_k}) = \left(\sum_{j=1}^m x_j E(1_{A_j})\right)\left(\sum_{k=1}^n y_k E(1_{B_k})\right) \\
&= E\left(\sum_{j=1}^m x_j 1_{A_j}\right) E\left(\sum_{k=1}^n y_k 1_{B_k}\right) = E(X)E(Y).
\end{aligned}$$

Step 3. The general case. Then there are increasing sequences (X_n), (Y_n) of simple random variables $X_n : \Omega \longrightarrow [0, \infty)$ and $Y_n : \Omega \longrightarrow [0, \infty)$ such that $X_n \uparrow X$ and $Y_n \uparrow Y$ pointwise. Then $X_n Y_n \uparrow XY$. Hence by the monotone convergence theorem and Step 2,

$$E(XY) = \lim_{n\to\infty} E(X_n Y_n) = \lim_{n\to\infty} E(X_n)E(Y_n) = \left(\lim_{n\to\infty} E(X_n)\right)\left(\lim_{n\to\infty} E(Y_n)\right).$$

(X_n and Y_n can be taken to be independent because we may take X_n to be $\sigma(X)$-measurable and Y_n to be $\sigma(Y)$-measurable. The σ-fields $\sigma(X)$ and $\sigma(Y)$ are independent because X and Y are independent.) $\square$

Corollary 77. For each $n \in \mathbb{N}$, for all independent random variables $X_1, \dots, X_n : \Omega \longrightarrow [0, \infty]$ we have $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$.

Proof. Denote by $P(n)$ the sentence "for all independent random variables $X_1, \dots, X_n : \Omega \longrightarrow [0, \infty]$ we have $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$." $P(1)$ is true trivially.

Let $n \in \mathbb{N}$ such that $P(n)$ is true. Let $X_1, \dots, X_{n+1} : \Omega \longrightarrow [0, \infty]$ be independent random variables. Let $X = X_1$ and $Y = X_2 \cdots X_{n+1}$. Then X and Y are independent by the grouping theorem and proposition 75. By the grouping theorem we get that X_1 and $(X_2, \dots, X_n)$ are independent, and then by letting $f_1 : \mathbb{R} \to \mathbb{R}$ be the identity and $f_2 : \mathbb{R}^{n-1} \to \mathbb{R}$ being the map from $(x_2, \dots, x_n) \mapsto x_2 \cdots x_n$. Then $X = f_1(X_1)$ and $Y = f_2(X_2, \dots, X_n)$ are independent. So by the proposition, $E(X_1 \cdots X_{n+1}) = E(XY) = E(X)E(Y)$. But $E(Y) = E(X_2) \cdots E(X_{n+1})$ by the inductive hypothesis. Thus $P(n+1)$ is true too.

Therefore by induction, for each $n \in \mathbb{N}$, $P(n)$ is true. $\square$

Remark (Another kind of bad writing to avoid)**.** Instead of $\implies$, use "so". $P \implies Q$ means if P were true, then Q would be true too; "Q is at least as true as P".

Corollary 78. Let $X_1, \dots, X_n : \Omega \longrightarrow \mathbb{R}$ (or $\mathbb{C}$) be independent integrable random variables. Let $X = X_1, \cdots, X_n$. Then X is integrable and $E(X) = E(X_1) \cdots E(X_n)$.

Proof. $|X| = |X_1| \cdots |X_n|$ and $|X_1|, \dots, |X_n|$ are independent, so $E(|X|) = E(|X_1|) \cdots E(|X_n|) < \infty$ so X is integrable. Now for $k = 1, \dots, n$, $0 \le X_k^+ \le |X_k|$ and $0 \le X_k^- \le |X_k|$, so X_k^+ and X_k^- are integrable.

$$\begin{aligned} E(X_1 \cdots, X_n) &= E((X_1^+ + X_1^-) \cdots (X_n^+ + X_n^-)) \\ &= E\left(\sum_{a \in \{+,-\}^n} \operatorname{sgn}(a) X_1^{a_1} \cdots X_n^{a_n} \right) \\ &= \sum_{a \in \{+,-\}^n} \operatorname{sgn}(a) E(X_1^{a(1)} \cdots X_n^{a(n)}) \\ &= \sum_{a \in \{+,-\}^n} \operatorname{sgn}(a) E(X_1^{a(1)}) \cdots E(X_n^{a(n)}) \\ &= (E(X_1^+) - E(X_1^-)) \cdots (E(X_n^+) - E(X_n^-)) \\ &= E(X_1) \cdots E(X_n). \end{aligned}$$

(You do the complex case). $\square$

Notation. $L^0 = L^0(\Omega)$ denotes the set of all real-valued random variables from Ω. For $0 < p < \infty$, $L^p = \{X \in L^0 : E(|X|^p) < \infty\}$. For now, we shall mainly focus on L^1 and L^2.

If $1 < p < \infty$, then $X \mapsto E(|X|^p)^{1/p}$ is a (semi)norm on L^p. If $0 < p < 1$, then $(X, Y) \mapsto E(|X - Y|^p)^{1/p}$ is a (pseudo)metric on L^p.

Recall that for $X \in L^1$, $E(X) = \int_\Omega X \, dP$.

Definition. If $X \in L^1$, then "X centered" is $X^\circ = X - E(X)$.

Remark. Let $X, Y \in L^1$. Then $E(X^\circ) = 0$ and $(X + Y)^0 = X^\circ + Y^\circ$.

Definition. Let $X \in L^1$. Then variance of X is $\operatorname{var}(X) = E((X^\circ)^2)$.

The standard deviation of X is $\sigma_X = \sqrt{\operatorname{var}(X)} = \|X^\circ\|_2$. (Do not confuse this with $\sigma(X)$.)

Remark. Let $X \in L^1$. Then $\operatorname{var}(X) = E(X^2) - E(X)^2$, because letting $\xi = E(X)$, we have $\operatorname{var}(X) = E((X - \xi)^2) = E(X^2 - 2\xi X + \xi^2) = E(X^2) - 2\xi E(X) + \xi^2 = E(X^2) - 2\xi^2 + \xi^2 = E(X^2) - \xi^2 = E(X^2) - E(X)^2$. $\square$

Remark. Let $X \in L^1$ and $a, b \in \mathbb{R}$. Then $(aX + b)^\circ = a(X^\circ)$, so $\operatorname{var}(aX + b) = E((a(X^\circ))^2) = a^2 E((X^\circ)^2) = a^2 \operatorname{var}(X)$. $\square$

Remark. Let $u, v \in \mathbb{R}$. Then $2uv \le u^2 + v^2$, because $(u - v)^2 \ge 0$. Hence $(u + v)^2 = u^2 + 2uv + v^2 \le 2u^2 + 2v^2$.

Remark. Let $X \in L^0$. Then for all $a, b \in \mathbb{R}$, we have $(X, b)^2 = (X - a + a - b)^2 \le 2(X - a)^2 + 2(a - b)^2$ so if $E((X - a)^2) < \infty$, then $E((X - b)^2) < \infty$. This holds for all $a, b \in \mathbb{R}$. Thus either $E((X - c)^2) < \infty$ for all $c \in \mathbb{R}$ or $E((X - c)^2) = \infty$ for all $c \in \mathbb{R}$. Thus $\operatorname{var}(X) < \infty$ if and only if $X \in L^2$ if and only if $E((X - c)^2) < \infty$ for all $c \in \mathbb{R}$.

Remark. Let $X \in L^2$. Then $X \in L^1$, because $|X| = |X| 1_A + |X| 1_B \le 1_A + |X|^2$ (where $A = \{|X| < 1\}, B = \{|X| \ge 1\}$; $1_A \in L^1$ since $P(\Omega) = 1 < \infty$).

Proposition 79. Let $X \in L^2$. Let $f(x) = E((X - x)^2)$. Then f is minimized exactly when $x = E(X)$.

Proof. Let $\bar{x} = E(X)$. Then

$$\begin{aligned} f(x) &= E((X - \bar{x} + \bar{x} - x)^2) = E((X - x)^2 + 2(X - \bar{x}(\bar{x} - x) + (\bar{x} - x)^2) \\ &= E((X - \bar{x})^2) + 2 \underbrace{E(X - \bar{x})}_{0} (\bar{x} - x) + (\bar{x} - x)^2 \\ &= \operatorname{var}(X) + (\bar{x} - x)^2 \end{aligned}$$

$\square$

3/10/21

If $X \in L^1$ and $\bar{x} = E(X)$, then for each $x \in \mathbb{R}$, $E[(X-x)^2] = E[(X-\bar{x})^2]+(\bar{x}+x)^2$. Hence if $E[(X-x)^2] < \infty$ for some $x \in \mathbb{R}$, then $E[(X-x)^2] < \infty$ for all $x \in \mathbb{R}$, and $x \in L^2$.

Also, if $X \in L^2$, then $E[(X-x)^2]$ is minimized over $x \in \mathbb{R}$ exactly when $x = E(X)$ and $\min_{x\in\mathbb{R}} E[(X-x)^2] = \operatorname{var}(X)$.

Remark. Let $X, Y \in L^1$. Let $\bar{x} = E(X)$ and $\operatorname{var} y = E(Y)$. Thus

$$X^\circ Y^\circ = (X-\bar{x})(Y-\bar{y}) = XY - \underbrace{\bar{x}Y - \bar{y}X + \bar{x}\bar{y}}_{\in L^1}.$$

Thus when $X, Y \in L^1$, then $XY \in L^1$ if and only if $X^\circ Y^\circ \in L^1$.

Definition. Let $X, Y \in L^1$ with $XY \in L^1$. The covariance of X and Y is $\operatorname{cov}(X,Y) = E(X^\circ Y^\circ)$.

(For instance, $\operatorname{cov}(X,X) = \operatorname{var}(X)$.)

Reminder. Let $X, Y \in L^2$. Then $XY \in L^1$ because $|XY| \leq (X^2+Y^2)/2$, which follows from $(|X|-|Y|)^2 \geq 0$. Thus for any $X, Y \in L^2$, we have $XY \in L^1$.

Reminder. If $X, Y \in L^1$ and X and Y are independent, then $E(X^\circ Y^\circ) = E(|X||Y|) - E(|X|)E(|Y|) < \infty$.

Remark. Let $X, Y \in L^1$ with $XY \in L^1$. Then

(a) $\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)$

(b) $\operatorname{var}(X+Y) = \operatorname{var}(X) + 2\operatorname{cov}(X,Y) + \operatorname{var}(Y)$

(c) for all $a, b, c, d \in \mathbb{R}$, $\operatorname{cov}(aX+b, cY+d) = ac\operatorname{cov}(X,Y)$.

Proof. Routine. □

Best Least Squares Estimator of Y by $aX+b$. Let $X, Y \in L^2$. We seek to minimize

$$E\left[(Y-(aX+b)^2)\right].$$

Assume $\operatorname{var}(X) \neq 0$. (So X is almost surely (that is, P-almost everywhere) not constant). Recall that for each $Z \in L^1$, we have $\operatorname{var}(Z) = E(Z^2) - E(Z)^2$, so $E(Z^2) = \operatorname{var}(Z) + E(Z)^2$. Let's apply this to $Z = Y - (aX+b)$, where $a, b \in \mathbb{R}$. We have

$$\begin{aligned} E[(Y-(aX+b))^2] &= \operatorname{var}(Y-aX) + E[(Y-aX)-b]^2 \\ &= \operatorname{var}(Y) - 2a\operatorname{cov}(X,Y) + a^2\operatorname{var}(X) + E[(Y-aX)-b]^2 \\ &= \left[a\sigma_X - \frac{\operatorname{cov}(X,Y)}{2}\right]^2 + \left[\operatorname{var}(Y) - \frac{\operatorname{cov}(X,Y)^2}{\operatorname{var}(X)}\right] + E[(Y-aX)-b]^2 \end{aligned}$$

Thus $E[(Y-(aX+b)^2]$ is minimized exactly when $a = \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}$ and $b = E(Y-aX)$ and its minimum value is $\operatorname{var}(Y) - \frac{\operatorname{cov}(X,Y)^2}{\operatorname{var}(X)}$.

Since $E[(Y-(aX-b))^2] \geq 0$, it follows that $\operatorname{cov}(X,Y)^2 \leq \operatorname{var}(X)\operatorname{var}(Y)$, so

$$|\operatorname{cov}(X,Y)| \leq \sigma_X\sigma_Y. \tag{$*$}$$

($E(X^\circ Y^\circ) = \|X\|_2\|Y\|_2$ by Schwartz's inequality.)

Definition. Let $X \in L^2$ with $\operatorname{var}(X) \neq 0$. Then X standardized is

$$X^* = \frac{X - E(X)}{\sigma_X}.$$

Note that $E(X^*) = 0$ and $\sigma_{X^*} = 1$.

Definition. Let $X, Y \in L^2$ with $\mathrm{var}(X) \neq 0$ and $\mathrm{var}(Y) \neq 0$. The correlation coefficient of X and Y is

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Notice that $\rho(X,Y) = \mathrm{cov}(X^*, Y^*)$. We have $-1 \leq \rho(X,Y) \leq 1$ by $(*)$.

We have $\rho(X,Y) = 1$ if and only if $Y = aX + b$ almost surely for some $a \in (0,\infty)$ and $b \in \mathbb{R}$. $\rho(X,Y) = -1$ if and only if $\rho(X,-y) = 1$.

Remark. If $X, Y \in L^2$ with $\mathrm{var}(X) \neq 0$ and $\mathrm{var}(Y) \neq 0$, then the straight line function of X^* which best approximates Y^* in the least squares sense is $aX^* + b$ where $a = \rho(X,Y)$ and $b = 0$.

Definition. Let $X, Y \in L^1$. To say X and Y are uncorrelated means $XY \in L^1$ and $E(XY) = E(X)E(Y)$ (i.e. $\mathrm{cov}(X,Y) = 0$). Note that if $X, Y \in L^2$, then X and Y are uncorrelated if $X^\circ \perp Y^\circ$ in L^2.

Remark. Let $X, Y \in L^1$. Suppose X and Y are independent. Then X and Y are uncorrelated.

Example. Let $\Omega = [0,1]$, $P =$ Lebesgue measure on $[0,1]$. Defined X and Y on Ω by $X(\omega) = \cos(2\pi\omega)$ and $Y(\omega) = \sin(2\pi\omega)$. Then X and Y are uncorrelated because

$$E(XY) = \int_0^1 \cos(2\pi\omega)\sin(2\pi\omega)\,d\omega = \int_0^1 \frac{1}{2}\sin(4\pi\omega)\,d\omega = 0$$

$$E(X) = 0,\ E(Y) = 0$$

so $E(XY) = E(X)E(Y)$.

But X and Y are not independent, because $\mathrm{law}(X,Y)$ is uniform on $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Then

$$P\left((X,Y) \in \left[-\frac{1}{2}, \frac{1}{2}\right]^2\right) = 0 \text{ but } P\left(X \in \left[-\frac{1}{2}, \frac{1}{2}\right]\right) P\left(Y \in \left[-\frac{1}{2}, \frac{1}{2}\right]\right) \neq 0.$$

Proposition 80. Let $X = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n \in L^1$ with $X_i X_j \in L^1$ for $i \neq j$. Then

$$\mathrm{var}(X) = \sum_{i,j=1}^n \mathrm{cov}(X_i, X_j) = \sum_{i=1}^n \mathrm{var}(X_i) + \sum_{i \neq j} \mathrm{cov}(X_i, X_j).$$

If $X_1, \ldots, X_n$ are pairwise uncorrelated, then $\mathrm{var}(X) = \sum_{i=1}^n \mathrm{var}(X_i)$.

In particular, if $X_1, \ldots, X_n$ are pairwise independent, then $\mathrm{var}(X) = \sum_{i=1}^n \mathrm{var}(X_i)$.

3/12/21

The L^2 Weak Law of Large Numbers, First Version. Let $X_1, X_2, X_3, \ldots, \in L^2$ be independent and identically distributed ($\mathrm{law}(X_j) = \mathrm{law}(X_1)$ for all j). Let $\bar{x} = E(X_j)$ (this is the same for each j because the X_j are identically distributed). Let $S_n = X_1 + \cdots + X_n$. Then for each $\varepsilon > 0$, $P(|S_n/n - \bar{x}| \geq \varepsilon) \to 0$ as $n \to \infty$.

Proposition 81 (Chebyshev's Inequality)**.** Let $Y \in L^2$, let $\bar{y} = E(Y)$, and let $\varepsilon > 0$. Then

$$P(|Y - \bar{y}| \geq \varepsilon) \leq \frac{\mathrm{var}(Y)}{\varepsilon^2}.$$

Proof. By Markov's inequality,

$$P(|Y - \bar{y}| \geq \varepsilon) = P(|Y - \bar{y}|^2 \geq \varepsilon^2) \leq \frac{E(|Y - \bar{y}|^2)}{\varepsilon^2} = \frac{\mathrm{var}(Y)}{\varepsilon^2}.$$

□

Proof of L^2 Law of Large Numbers, Version 1. Let $\varepsilon > 0$. Then by Chebyshev's inequality,

$$P\left(\left|\frac{S_n}{n} - \bar{x}\right| \geq \varepsilon\right) \leq \frac{\operatorname{var}\left(\frac{S_n}{n} - \bar{x}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\operatorname{var}(S_n)}{\varepsilon^2}$$

which since $X_1, \ldots, X_n$ are independent,

$$= \frac{1}{\varepsilon^2 n^2}(\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)) = \frac{n\operatorname{var}(X_1)}{\varepsilon^2 n^2} = \frac{\operatorname{var}(X_1)}{\varepsilon^2 n} \to 0.$$

□

Generalizations.

1. It's enough if $X_1, \ldots, X_n$ are pairwise independent, or even just uncorrelated.
2. $X_1, X_2, X_3, \ldots$ need not be identically distributed. For instance, it would be enough if instead $E(X_j) = \bar{x}$ for each j and $\sup \operatorname{var}(X_j) < \infty$. Further extensions are possible with slight changes in the proof.
3. The conclusion can be strengthened: $\|S_n/n - \bar{x}\|_2 \to 0$. (This is convergence in L^2, as opposed to the convergence in probability in the conclusion of the weak law of large numbers. Convergence in L^p is defined analogously, and in this sense convergence in probability is like "convergence in L^0". Convergence in L^∞ is "almost uniform convergence".)

However, the weak law of large numbers does not tell us that $S_n/n \to \bar{x}$ almost surely. The strong law of large numbers is what tells us that $S_n/n \to \bar{x}$ almost surely.

Reminder. If $\operatorname{law}(X) = \mu$, then $E[f(X)] = \int_\Omega f \circ X \, dP = \int_{\mathbf{X}} f \, d\mu$.

The Typewriter Sequence. In $[0,1)$, let $A_0 = [0,1)$, $A_1 = [0, 1/2)$, $A_2 = [1/2, 1)$, $A_3 = [0, 1/4)$, $A_4 = [1/4, 1/2)$, $A_5 = [1/2, 3/4)$, $A_6 = [3/4, 1)$, and so on.

Let $f_n = 1_{A_n}$. Then $f_n \to 0$ in probability ($P =$ Lebesgue measure on $[0,1)$) because for each $\varepsilon > 0$, $P(|f_n - 0| \geq \varepsilon) \to 0$. In fact, for each $p \in (0.\infty)$, $\int |f_n|^p \, dP = \int 1_{A_n}^p \, dP = \int 1_{A_n} \, dP = P(A_n) \to 0$. Thus $f_n \to 0$ in L^p for each $p \in (0, \infty)$. For each $\omega \in [0,1)$, $\limsup_{n\to\infty} f_n(\omega) = 1$ and $\liminf_{n\to\infty} f_n(\omega) = 0$, so $\lim_{n\to\infty} f_n(\omega)$ does not exist.

Let $g_0 = f_0$, $g_1 = -2f_1$, $g_2 = -2f_2$, $g_3 = 3f_3$, $g_4 = 3f_4$, $g_5 = 3f_5$, $g_6 = 3f_6$, and so on. For each $p \in (0, \infty)$, $\int |g_n|^p \, dP \to 0$. But for each $\omega \in [0,1)$, $\limsup_{n\to\infty} g_n(\omega) = \infty$ and likewise $\liminf_{n\to\infty} g_n(\omega) = -\infty$.

Digression. Let $f : \mathbb{R} \to \mathbb{C}$ be 1-periodic and suitably regular. Let $S_n f = \sum_{k=-n}^n \hat{f}(k) e_k$ where $e_k = e^{2\pi i k x}$ for $k \in \mathbb{Z}$, $x \in \mathbb{R}$ and $\hat{f}(k) = \int_0^1 = \overline{e_k(x)} f(x) \, dx = \langle e_k \mid f \rangle$. If $f \in L^2(\mathbb{T})$, then $\|f - S_n f\|_2 \to 0$.

What about almost everywhere? Lusin (1912) conjectured "yes".

Kolmogorov (late 1920s): there exists $f \in L^1(\mathbb{T})$ such that for all $x \in \mathbb{R}$, $\lim_{n\to\infty} S_n f(x)$ does not exist.

Zygmund (∼1960): "the problem of the existence of a continuous function whose series diverges everywhere remains open".

Carlson (1966): Yes to Lusin's conjecture. For all $f \in L^2(\mathbb{T})$, $S_n f \to f$ almost everywhere. (A similar method can be used to prove this for $f \in L^p$ for any $p > 1$; Hunt did this generalization later).

Theorem ((Sergei) Bernstein's Inequality.)**.** Let $0 < p < 1$. Let $A_1, A_2, A_3, \ldots$ be independent events with $P(A_j) = p$ for all j. Let $X_j = 1_{A_j}$ and $S_n = X_1 + \cdots + X_n$. Let $\varepsilon > 0$. Then

$$P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2}{4} n\right).$$

Proposition 82. Let $Y, Y_1, Y_2, \ldots$ be real random variables. Suppose for all $\varepsilon > 0$, $\sum_n P(|Y - Y_n| \geq \varepsilon) < \infty$. Then $Y_n \to Y$ almost surely.

Proof. For each $m \in \mathbb{N}$, let $B_n^m = \{|Y - Y_m| \geq 1/m\}$ and note that $\sum_n P(B_n^m) < \infty$, so letting $E_n = \{\omega \in \Omega : \omega \in B_n^m \text{ for infinitely many } n\}$ we have $P(E_m) = 0$ by the first Borel-Cantelli lemma. Hence $P\left(\bigcup_m E_m\right) = 0$. Let $G = \Omega \setminus \bigcup_m E_m$. Let $\omega \in G$, let $m \in \mathbb{N}$. Then $\omega \notin E_m$, so $\omega \in B_n^m$ for only finitely any n, so there exists $N_m \in \mathbb{N}$ such that for all $n \geq N_m$, $\omega \notin B_n^m$, so $|Y(\omega) - Y_n(\omega)| < 1/m$.

We've shown that for every $\omega \in G$, for any $m \in \mathbb{N}$, there is $N_m \in \mathbb{N}$ such that for all $n > N_m$, $|Y(\omega) - Y_n(\omega)| < 1/m$. Thus for every $\omega \in G$, $Y_n(\omega) \to Y(\omega)$. Then $G \subseteq \{Y_n \to Y\}$; $\{Y_n \to Y\}$ is measurable, hence $P(Y_n \to Y) = 1$, so $Y_n \to Y$ almost surely. □

3/15/21

Lemma 83. For each $x \in \mathbb{R}$, $e^x \leq x + e^{x^2}$.

Proof. Let $f(x) = x^{x^2} - e^x$ and $g(x) = -x$. We wish to show that $f \geq g$. Next $f(x) = (1 + x^2 + \dots) - (1 + x + x^2/2! + \dots) = -x + x^2/2 + \dots$ so $f(0) = 0$ and $f'(0) = -1$. Thus g is the tangent line to f at $x = 0$. Hence it suffices to show that f is convex (concave up). Now $f'(x) = 2xe^{x^2} - e^x$ and $f''(x) = 2e^{x^2} + (2x)(2xe^{x^2}) - e^x = (4x^2 + 2)e^{x^2} - e^x = (4x^2 + 2 - e^{x - x^2})e^{x^2}$. Now, $4x^2 + 2 \geq 2$ while $e^{x - x^2} = e^{1/4 - (1/2 - x)^2} \leq e^{1/4} \approx 1.28 \dots < 2$. Hence $f''(x) > 0$ for each $x \in \mathbb{R}$. Thus f is strictly convex on $\mathbb{R}$. □

Theorem 84 (Bernstein's inequality)**.** Let $0 < p < 1$. Let $A_1, A_2, A_3, \dots$ be independent events with $P(A_j) = p$ for all j. Let $X_j = 1_{A_j}$ (we say X is a Bernoulli random variable) and $S_n = X_1 + \dots + X_n$. Let $\varepsilon > 0$. Then

$$P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2 n}{4}\right).$$

and

$$P\left(\frac{S_n}{n} \leq p - \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2 n}{4}\right).$$

Proof. Let $m = \lceil n(p + \varepsilon) \rceil$, $q = 1 - p$, and $\lambda > 0$. Then

$$\begin{aligned}
P\left(\frac{X_n}{n} \geq p + \varepsilon\right) &= P(S_n \geq n(p + \varepsilon)) = \sum_{k=m}^{n} P(S_n = k) = \sum_{k=m}^{n} \binom{n}{k} p^k q^{n-k} \\
&\leq \sum_{k=m}^{n} e^{\lambda(k - n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k} = e^{-\lambda \varepsilon n} \sum_{k=m}^{n} e^{\lambda(pk + qk - pn} \binom{n}{k} p^k q^{n-k} \\
&= e^{\lambda \epsilon n} \sum_{k=m}^{n} \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k} \leq e^{-\lambda \varepsilon n} \sum_{k=0}^{n} \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k} \\
&= e^{-\lambda \varepsilon n} (pe^{\lambda q} + qe^{-\lambda p})^n.
\end{aligned}$$

Now $e^x \leq x + e^{x^2}$ for each $x \in \mathbb{R}$, by the lemma. Hence

$$\begin{aligned}
P\left(\frac{S_n}{n} \geq p + \varepsilon\right) &\leq e^{-\lambda \varepsilon n} \Big(p(\underbrace{\lambda q + e^{\lambda^2 q^2}}_{x = \lambda q}) + q(\underbrace{-\lambda p + e^{\lambda^2 p^2}}_{x = -\lambda p})\Big)^n \\
&= e^{-\lambda \varepsilon n} (pe^{\lambda^2 q^2} + qe^{\lambda^2 p^2})^n \\
&\leq e^{-\lambda \varepsilon n} (pe^{\lambda^2} + qe^{\lambda^2})^n \\
&= e^{-\lambda \varepsilon n} (p + q)^n e^{\lambda^2 n} \\
&= e^{(\lambda^2 - \lambda \varepsilon) n} = e^{((\lambda - \varepsilon/2)^2 - \varepsilon^2/4) n}.
\end{aligned}$$

This holds for each $\lambda > 0$. In particular, it holds for $\lambda = \varepsilon/2$. Hence $P(S_n/n \geq p + \varepsilon) \leq e^{-\varepsilon^2 n/4}$. Let $B_j = A_j^c$. Then $B_1, B_2, B_3, \dots$ are independent and $P(B_j) = q$ and $1_{B_j} = 1 - X_j$. Hence by what we've done applied with A_j replaced by B_j,

$$P\left(\frac{S_n}{n} \le p - \varepsilon\right) = P\left(1 - \frac{S_n}{n} \ge 1 - p + \varepsilon\right) = P\left(\frac{(1-X_1) + \cdots + (1-X_n)}{n} \ge q + \varepsilon\right) \le e^{-\varepsilon^2 n/4}.$$

□

Theorem 85 (Bernoulli Strong Law of Large Numbers)**.** Let $0 < p < 1$. Let $A_1, A_2, A_3, \ldots$ be independent events with $P(A_j) = p$ for all j. Let $X_j = 1_{A_j}$ and $S_n = X_1 + \cdots + X_n$. Then $S_n/n \to p$ almost surely as $n \to \infty$.

Proof. By Bernstein's inequalities, for each $\varepsilon > 0$, $P(|S_n/n - p| \ge \varepsilon) \le 2e^{-\varepsilon^2 n/4}$, so $\sum_{n=1}^{\infty} P(|S_n/n - p| \ge \varepsilon) < \infty$. Hence $S_n/n - p \to 0$ almost surely. □

Reminder. Recall Proposition 82.

Proof. Let $Y_n = S_n/n - p$. For all $m, n \in \mathbb{N}$, let $B_{m,n} = \{Y_n \ge 1/m\}$. For each $m \in \mathbb{N}$, let $E_m = \{\omega \in \Omega : \omega \in B_{m,n}$ for infinitely many $n\}$. By the first Borel-Cantelli lemma, for each $m \in \mathbb{N}$, since $\sum_n P(B_{m,n}) < \infty$, we have $P(E_m) = 0$. Let $G = \Omega \setminus \bigcup_m E_m$. For each $\omega \in G$, for each $m \in \mathbb{N}$, we have $\omega \notin E_m$, so there is $N(\omega, m)$ such that for any $n > N(\omega, m)$, $\omega \notin B_{m,n}$, so $Y_n(\omega) < 1/m$, so $Y_n(\omega) \to 0$ as $n \to \infty$. □

Theorem (Kolmogorov's Strong Law of Large Numbers (∼1930).)**.** Let $X_1, X_2, X_2, \ldots$ be independent identically distributed integrable real random variables. Let $S_n = X_1 + \cdots + X_n$ and $\bar{x} = E(X_j)$. Then $S_n/n \to \bar{x}$ almost surely as $n \to \infty$ (and in L^1).

Generalizations.

1. $X_1, X_2, X_3, \ldots$ need not be identically distributed. It is enough if $X_1^\circ, X_2^\circ, X_3^\circ, \ldots$ are L^1-stochastically dominated, which means that there is an integrable random variable Y such that for each j, for each $x \ge 0$, $P(|X_j^\circ| \ge x) \le P(|Y| \ge x)$. If they are identically distributed, $Y = X_1$ would do. Then $E(X_j)$ may depend on j, but we get $(X_1^\circ + \cdots + X_n^\circ)/n \to 0$ almost surely (and in L^1).
2. (Etemadi, c.1981) It suffices that $X_1, X_2, X_3, \ldots$ be pairwise independent.

The Central Limit Theorem, Roughly. Roughly, the standardized sum of a "large" number of "small" independent random variables is "approximately" normal (i.e. a "bell curve" $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$).

Consider coin tossing with a fair coin. Let

$$X_j = \begin{cases} 1 & \text{if we get heads on the } j\text{-th toss.} \\ -1 & \text{if we get tails on the } j\text{-th toss.} \end{cases}$$

let $S_n = X_1 + \cdots + X_n$, the number of heads minus the number of tails in the first n tosses. Then $E(X_j) = 0$, $\mathrm{var}(X_j) = 1$, $E(S_n) = 0$, $\mathrm{var}(S_n) = \sum_{j=1}^{n} \mathrm{var}(X_j) = n$. $\sigma_{S_n} = \sqrt{n}$. Thus $S_n^* = S_n/\sqrt{n}$. $S_n/n \to 0$ almost surely, but $S_n/\sqrt{n} \Rightarrow N(0,1)$ (where $\Rightarrow$ means weak convergence: for each $x \in \mathbb{R}$, $P(S_n/\sqrt{n} \le x) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$; $N(0,1)$ means "normal with mean 0 and variance 1).

This is the De Moivre-Laplace Theorem (De Moivre 1718, 1738, 1756; Laplace 1812).

Reminder (Stirling's Formula.)**.** $n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$ as $n \to \infty$. In fact, for each n,

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} < n! < \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n}}.$$

In fact, the following sharper lower bound holds: for each $n \in \mathbb{N}$,

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}} < n!$$

The hard part is getting the $\sqrt{2\pi}$ in front.

Proof (start). We'll prove this with $\sqrt{2\pi}$ replaced by e^c for some constant c. (Then use this to prove De Moivre-Laplace theorem then with e^{-c} in place of $1/\sqrt{2\pi}$. Since $\int_{\mathbb{R}} e^{-c} e^{-x^2/2}\, dx = 1$, $e^{-c} = 1/\sqrt{2\pi}$.) □

3/17/21

Idea in Proof of De Moivre-Laplace Theorem Suppose $P(X_j = 1) = p$, $P(X_j = 0) = q = 1 - p$, $X_1, X_2, X_3, \ldots$ independent. $S_n = X_1 + \cdots + X_n$ is the number of heads in the first n tosses, $E(S_n) = np$, $\operatorname{var}(S_n) = npq$ (since $\operatorname{var}(X_j) = E(X_j^2) - E(X_j)^2 = p - p^2 = p(1-p) = pq$). Then $\sigma_{S_n} = \sqrt{npq}$ so $S_n^* = (S_n - np)/\sqrt{npq}$. The De Moivre-Laplace theorem says that for each $x \in \mathbb{R}$, $P(S_n^* \leq x) \to \int_{-\infty}^{x} e^{-\xi^2/2}/\sqrt{2\pi}\, d\xi$ as $n \to \infty$.

Digression (Normal Distributions.)**.** Let X be a standard normal random variable. This <u>means</u> X has density $\varphi(x) = ce^{-x^2/2}$ where c is chosen for $\int_{\mathbb{R}} \varphi(x)\, dx = 1$. Let Y be another standard normal random variable. So Y has density φ as well. Suppose X and Y are independent. (One way we could get X and Y is let $\Omega = \mathbb{R}^2, \mathscr{F} = \operatorname{Borel}(\mathbb{R}), P(A) = \iint_A \varphi(x)\varphi(y)\, dxdy = \int_A = \varphi(x)\varphi(y) m_2(dx, dy)$ where $m_2(dx, dy)$ is the Lebesgue measure on $\mathbb{R}^2$ and $X(\omega) = \omega_1$, $Y(\omega) = \omega_2$ for each $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$.) Note that (X, Y) has density $\psi(x, y) = \varphi(x)\varphi(y) = c^2 e^{-(x^2+y^2)/2}$:

$$\begin{aligned} 1 &= \left(\int_{\mathbb{R}} \varphi(x)\, dx\right)\left(\int_{\mathbb{R}} \varphi(y)\, dy\right) = \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x)\, dx \varphi(y)\, dy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x)\varphi(y)\, dxdy = \int_{\mathbb{R}}\int_{\mathbb{R}} c^2 e^{-(x^2+y^2)/2}\, dxdy \end{aligned}$$

passing to polar coordinates,

$$= \int_0^\infty c^2 e^{-r^2/2}\, 2\pi r\, dr = 2\pi c^2 \int_0^\infty e^{-r^2/2} r\, dr$$

and letting $u = r^2/2$ so $du = r\, dr$

$$= 2\pi c^2 \int_0^\infty e^{-u}\, du = 2\pi c^2.$$

Thus $1 = 2\pi c^2$ so $c^2 = 1/(2\pi)$, so $c = 1/\sqrt{2\pi}$. Thus $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$. Clearly $E(X) = 0$ (since $\int_{\mathbb{R}} (x/\sqrt{2\pi}) e^{-x^2/2} = 0$) and $E(Y) = 0$.

Let $R = (X^2 + Y^2)^{1/2}$. For $t \geq 0$,

$$\begin{aligned} P(R^2 > t) = P(R > \sqrt{t} &= \int_0^{2\pi} \int_{\sqrt{t}}^\infty \psi(r\cos\theta, r\sin\theta) r\, dr\, d\theta \\ &= \int_{\sqrt{t}}^\infty e^{-r^2/2} r\, dr = \int_{t/2}^\infty e^{-u}\, du = e^{-t/2}. \end{aligned}$$

Thus R^2 has an exponential distribution with parameter $1/2$. Hence $E(R^2) = \int_0^\infty P(R^2 > t)\, dt$. Hence

$$E(X^2 + Y^2) = E(R^2) = \int_0^\infty P(R^2 > t)\, dt = \int_0^\infty e^{-t/2}\, dt = 2$$

so $E(X^2) = 1 = E(Y^2)$. Thus $\operatorname{var}(X) = 1 = \operatorname{var}(Y)$. Thus a standard normal random variable has mean 0 and variance 1. Now since φ, the density of (X, Y), is constant on each circle centered at the origin, and since rotations preserve area, we have $\operatorname{law}(X_\theta, Y_\theta) = \operatorname{law}(X, Y)$, where (X_θ, Y_θ) is attained by rotating (X, Y) through an angle θ about the origin. Such a rotation maps $e_1 = (1, 0)$ to $u_1 = (\cos\theta, \sin\theta)$ and $e_2 = (0, 1)$ to $u_2 = (-\sin\theta, \cos\theta)$, so it maps $(x, y) = xe_1 + ye_2$ to $(x_\theta, y_\theta) = xu_1 + yu_2 = x(\cos\theta, \sin\theta) + y(-\sin\theta, \cos\theta) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$. Thus $X_\theta = X\cos\theta - Y\sin\theta$ and $Y_\theta = X\sin\theta + Y\cos\theta$. Thus for each angle θ, $\operatorname{law}(Y_\theta) = \operatorname{law}(X) = N(0, 1)$.

Next for any $\alpha, \beta \in \mathbb{R}$ with $\alpha^2 + \beta^2 = 1$, there exists $\theta \in [0, 2\pi)$ such that $(\alpha, \beta) = (\cos\theta, \sin\theta)$ so $\operatorname{law}(\alpha X + \beta Y) = N(0, 1)$.

Now suppose instead that X is normal with mean μ and variance a. This means $X^* = (X - \mu)/\sqrt{a}$ is standard normal. Suppose also Y is normal with mean ν and variance b. This means $Y^* = (Y - \nu)/\sqrt{b}$ is standard normal. Suppose also that X and Y are independent. Then so are X^* and Y^*. Let $\alpha = (a/(a+b))^{1/2}$ and $\beta = (b/(a+b))^{1/2}$. Then $\alpha^2 + \beta^2 = a/(a+b) + b/(a+b) = 1$. Hence $\alpha X^* + \beta Y^*$ is standard normal. But

$$\alpha X^* + \beta Y^* = \frac{\sqrt{a}}{\sqrt{a+b}}\left(\frac{X-\mu}{\sqrt{a}}\right) + \frac{\sqrt{b}}{\sqrt{a+b}}\left(\frac{Y-\nu}{\sqrt{b}}\right) = \frac{(X+Y)-(\mu+\nu)}{\sqrt{a+b}} = (X+Y)^*$$

since $\mu + \nu = E(X+Y)$, $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) = a+b$ so $\sigma_{X,Y} = \sqrt{a+b}$. Thus $\operatorname{law}(X+Y) = N(\mu+\nu, a+b)$.

Back to De Moivre-Laplace

$$\operatorname{law}\left(\frac{S_n - np}{\sqrt{npq}}\right) = \operatorname{law}(S_n^*) \Rightarrow N(0,1).$$

Idea: Let $x_{n,k} = (k-np)/\sqrt{npq}$ for $k = 0, 1, \ldots, n$.

$$P(S_n^* = x_{n,k}) \sim \frac{1}{2\pi npq} e^{-x^2 nk/2}$$

as $n \to \infty$.

3/19/21

Reminder (of Notation.)**.** $0 < p < 1$, $q = 1-p$ $X_1, X_2, X_3, \ldots$ independent, $P(X_j = 1) = p$, $P(X_j = 0) = q$, $S_n = X_1 + \cdots + X_n$. $E(S_n) = np$, $\sigma_{S_n} = \sqrt{npq}$,

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \qquad x_{n,k} = \frac{k-np}{\sqrt{npq}}$$

$k = 0, 1, \ldots, n$.

Theorem 86 (Theorem 1)**.** Let $0 \le A_m < \infty$ and suppose $A_n^3/\sqrt{n} \to 0$ as $n \to \infty$. Let $J_n = \{k \in \{1, \ldots, n\} : |x_{n,k}| \le A_n\}$. Then $P(S_n^* = x_{n,k}) \sim e^{-x_{n,k}^2/2}/\sqrt{2\pi npq}$ uniformly for $k \in J_n$ as $n \to \infty$.

Proof of the De Moivre-Laplace Theorem. Let $a < b \in \mathbb{R}$. We wish to show that $\lim_{n\to\infty} P(a < S_n^* \le b) = \int_a^b e^{-x^2/2}/\sqrt{2\pi}\, dx$. Let $A_n = \max\{|a|, |b|\}$ for all n. Then $[a,b] \subseteq [-A_n, A_n]$ for all n. Hence by theorem 1 (cf. Theorem 86) with e^c in place of $\sqrt{2\pi}$, $P(S_n^* = n_{n,k} \sim (e^{-c}/\sqrt{npq})e^{-x_{n,k}^2/2}$ uniformly for $a < \infty$ as $n \to \infty$. Hence, as $n \to \infty$,

$$P(a < S_n^* \le b) \sim \sum_{k: a < x_{n,k} \le b} e^{-c} e^{-x_{n,k}^2/2} \frac{1}{\sqrt{npq}} = \sum_{k: a < x_{n,k} \le b} e^{-c} e^{-x_{n,k}^2/2} (\underbrace{x_{n,k} - x_{n,k-1}}_{\Delta x_{n,k}}) \to \int_a^b e^{-c} e^{-x^2/2}\, dx.$$

(The sums converge to the integral because the continuous function $x \mapsto e^{-x^2/2}$ is Riemann-integrable over $[a,b]$). Now let's show that $e^c = \sqrt{2\pi}$. By Chebyshev's inequality, for each $b > 0$, $P(|S_n^*| \ge b) \le 1/b^2$ (since $E(S_n^*) = 0$ and $\operatorname{var}(S_n^*) = 1$). Hence $1 - 1/b^2 \le \int_{-b}^b e^{-c} e^{-x^2/2}\, dx \le 1$. Letting $b \to \infty$, we get $\int_{-\infty}^\infty e^{-c} e^{-x^2/2}\, dx = 1$. Hence $e^{-c} = 1/\sqrt{2\pi}$, so $e^c = \sqrt{2\pi}$. $\square$

Here is another result that can be derived from theorem 1 (Theorem 86): Suppose $x_n \to \infty$ but $x_n^3/\sqrt{n} \to 0$. Then $P(S_n^* \ge x_n) \sim \int_{x_n}^\infty e^{-x^2/2}/\sqrt{2\pi}\, dx$ as $n \to \infty$. (See Feller, vol. 1 (3rd ed.) Ch. VII, Section 6.)

Example. Toss a fair coin 100 times. $P(|S_{100}| - 50) \ge 10 \approx ?$ $E(S_{100}) = 50$, $\sigma_{S_{100}} = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5$.

Chebyshev:

$$P(|S_{100} - 50| \ge 10) \le \frac{\operatorname{var}(S_{100})}{10^2} = \frac{25}{100} = \frac{1}{4}.$$

Bernstein:

$$P(|S_{100}-50|\geq 10)=P\left(\left|\frac{1}{100}S_{100}-\frac{1}{2}\right|\geq\frac{1}{10}\right)\leq\underbrace{2e^{-n\delta^2/4}}_{n=100,\delta=\frac{1}{10}}=2e^{-1/4}=1.557>1.$$

De Moivre-Laplace:

$$P(|S_{100}-50|\geq 10)=P(|S^*_{100}|\geq 2)\approx 1-\int_{-2}^{2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx=2\int_{-\infty}^{-2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx\approx 2(0.0228)=0.0456.$$

Theorem 87 (The Independent Identically Distributed (iid) Central Limit Theorem)**.** Let X_1, X_2, X_3 be independent and identically distributed. Let $S_n = X_1 + \cdots + X_n$. Then for all $a, b \in \mathbb{R}$,

$$\lim_{n\to\infty} P(a < S^*_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

Proof. Uses characteristic functions and Lévy's continuity theorem. □

Theorem 88 (Lindenberg's Central Limit Theorem (1922))**.** For $n = 1, 2, 3, \ldots$, let $X_{n,j}, j = 1, \ldots, k_n$ be independent real random variables (note that this means the $X_{n,j}$ are independent for fixed n. In fact, the rows may be on different probability spaces!) satisfying $E(X_{n,j}) = 0$, $\sum_{j=1}^{k_n} E(X^2_{n,j}) = 1$. Let $S_n = \sum_{j=1}^{k_n} X_{n,j}$ (so $E(S_n) = 0$ and $\operatorname{var}(S_n) = 1$. For each $\varepsilon > 0$, let $E_{n,j,\varepsilon} = E(X^2_{n,j}; |X_{n,j}| \geq \varepsilon)$.

Suppose for each $\varepsilon > 0$,

$$\sum_{j=1}^{n_k} E_{n,j,\varepsilon} \to 0 \text{ as } n \to \infty.$$

(this is Lindenberg's condition). Then for all $a, b \in \mathbb{R}$ with $a < b$,

$$\lim_{n\to\infty} P(a < S_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

Theorem 89 (Paul Lévy's Central Limit Theorem (1934))**.** (Reference: Loève.) For $n = 1, 2, 3, \ldots$, let $X_{n,j}$, $j = 1, \ldots, k_n$ be independent real random variables. Let $S_n = \sum_{j=1}^{k_n} X_{n,j}$. Suppose there is a probability measure μ on Borel($\mathbb{R}$) such that $\operatorname{law}(S_n) \Rightarrow \mu$ as $n \to \infty$. Then (a) and (b) below are equivalent:

(a) For each $\varepsilon > 0$, $P(|\max_j |X_{n,j}| \geq \varepsilon) \to 0$ as $n \to \infty$.

(b) μ is normal (possibly degenerate) and for each $\varepsilon > 0$ $\max_j P(|X_{n,j}| \geq \varepsilon) \to 0$ a $n \to \infty$.

3/22/21

Notation.

$$E(X; A) = \int_A X\,dP = E(X 1_A)$$

Remark. In the Do Moivre-Laplace theorem, De Moivre: $p = 1/2$, Laplace: $0 < p < 1$.

Corollary 90 (of Lévy's Central Limit Theorem)**.** Let $(X_t)_{0\leq t<\infty}$ be a family of real random variables (a (continuous) stochastic process in $\mathbb{R}$). Suppose for all $n \in \mathbb{N}$, for all $t_0, t_1, \ldots, t_n$, if $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$, the increments $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent ("independent increments").

Suppose also that for each $\omega \in \Omega$, $t \mapsto X_t(\omega)$ is continuous. Then for $0 \leq s < t < \infty$, $\operatorname{law}(X_t - X_s)$ is normal (possibly degenerate).

Proof. Let $0 \leq s < t < \infty$. For $n = 1, 2, 3, \ldots$, for $j = 1, \ldots, n$, let $X_j^n = X_{t_j^n} - X_{t_{j-1}^n}$, where $t_j^n = s + j(t-s)/n$. For each n, the random variables $X_1^n, \ldots, X_n^n$ are independent. For each ω, $t \mapsto X_t(\omega)$ is continuous on $[0, \infty)$, so uniformly continuous on $[a, b]$, so $\max_{j \in \{1,\ldots,n\}} |X_j^n(\omega)| \to 0$ as $n \to \infty$, and for each $\varepsilon > 0$, $P\left(\max_{j \in \{1,\ldots,n\}} |X_j^n(\omega)| \geq \varepsilon\right) \to 0$ as $n \to \infty$ (pointwise convergence implies convergence in probability).

Let $S_n = \sum_{j=1}^n X_j^n$. Then $S_n = X_b - X_a$. Hence $\operatorname{law}(S_n) \Rightarrow \mu$ where $\mu = \operatorname{law}(X_b - X_a)$. Hence μ is normal. $\square$

Proposition 91. Let Y, Y_1, Y_2, Y_3 be real random variables. Suppose $Y_n \to Y$ almost surely. Then $Y_n \xrightarrow{P} Y$. ($Y_n \xrightarrow{P} Y$ means Y_n converges to Y in probability.)

Proof. We wish to show that for each $\varepsilon > 0$, $P(|Y_n - Y| \geq \varepsilon) \to 0$ as $n \to \infty$. Let $Z_n = \sum_{k \geq n} |Y - Y_k|$. Then $Z_n \downarrow 0$ almost surely. Let $G = \{\omega \in \Omega : Y_n(\omega) \to Y(\omega)\}$. Then $G = \{\omega \in \Omega : Z_n(\omega) \downarrow 0\}$. Let $B = \Omega \setminus G$. By assumption, $P(B) = 0$. Let $\varepsilon > 0$. Let $G_n = \{\omega \in \Omega : Z_n(\omega) < \varepsilon\}$, $B_n = \{\omega \in \Omega : Z_n(\omega) \geq \varepsilon\}$. Now $G_1 \subseteq G_2 \subseteq G_3 \subseteq \ldots$ so $\bigcup_{n=1}^\infty G_n \supseteq G$. Hence $P\left(\bigcup_{n=1}^\infty G_n\right) = 1$. Since $G_1 \subseteq G_2 \subseteq G_3 \subseteq \ldots$, $P(G_n) \uparrow P\left(\bigcup_{n=1}^\infty G_n\right) = 1$. Thus $P(B_n) \downarrow 0$. For each n, $\{|Y - Y_n| \geq \varepsilon\} \subseteq B_n$, because $Z_n \geq |Y - Y_n|$. Hence $0 \leq P(|Y - Y_n| \geq \varepsilon) \leq P(B_n) \to 0$. Hence $P(|Y - Y_n| \geq \varepsilon) \to 0$. $\square$

Remark. The preceding proposition generalizes to random variables taking values in a separable metric space.

Definition. Let $\mu, \mu_1, \mu_2, \mu_3, \ldots$ be Borel probability measures on a topological space $\mathbf{X}$. To say that (μ_n) converges weakly to μ (denoted $\mu_n \Rightarrow \mu$) means that for each bounded continuous function $f : \mathbf{X} \to \mathbb{R}$, $\int f \, d\mu_n \to \int f \, d\mu$.

Example. Let $x, x_1, x_2, x_3, \dots \in \mathbb{R}$ ($\mathbf{X} = \mathbb{R}$). Let $\mu = \delta_x$, $\mu_n = \delta_{x_n}$. Then $\mu_n \Rightarrow \mu$ if and only if $x_n \to x$.

Example. With $\mathbf{X} = [0, 1]$. For $n = 0, 1, 2, \ldots$, let $0 \leq x_0^n < x_1^n \cdots < x_n^n = 1$, let $\Delta_j^n = x_j^n - x_{j-1}^n$, let $c_j^n \in [x_{j-1}^n, x_j^n]$ for $j = 1, \ldots, n$. Let $\mu_n = \sum_{j=1}^n \Delta_j^n \delta_{c_j^n}$. ($\mu_n$ is a convex combination of probability measures). Suppose $\max_{j \in \{1,\ldots,n\}} \Delta_j^n \to 0$. Then $\mu_n \Rightarrow \mu$, where μ is the Borel-Lebesgue measure (the Lebesgue measure restricted to Borel sets) on $[0, 1]$.

For let $f : [0, 1] \to \mathbb{R}$ be continuous. Then $\int f \, d\mu_n = \sum_{j=1}^n f(c_j^n) \Delta_j^n \to \int_0^1 f(x) \, dx$ because f is Riemann integrable and $\max_{j \in \{1,\ldots,n\}} \Delta_j^n \to 0$.

Theorem 92. Let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on $\mathbb{R}$. Then the following are equivalent:

(a) $\mu_n \Rightarrow \mu$.

(b) For each bounded Borel function $f : \mathbb{R} \to \mathbb{R}$, if $\mu(\{x \in \mathbb{R} : f \text{ is not continuous at } x\}) = 0$, $\int f \, d\mu_n \to \int f \, d\mu$.

(c) For each Borel set $C \subseteq \mathbb{R}$, if $\mu(\partial C) = 0$, $\mu_n(C) \to \mu(C)$.

(d) For each nonempty bounded interval I in $\mathbb{R}$ with endpoints a and b, if $\mu(\{a\}) = \mu(\{b\}) = 0$, then $\mu_n(I) \to \mu(I)$.

(e) For each $x \in \mathbb{R}$, if $\mu(\{x\}) = 0$ then $\mu_n((-\infty, x])$. ($F_{\mu_n}(x) \to F_\mu(x)$ for each $x \in \mathbb{R}$ such that F_μ is continuous at x.)

Poisson Distributions. Let $b(k; n, p) = \binom{n}{p} p^k (1-p)^{n-k}$ (the probability of k successful trials out of n with probability of success p on any given trial) for $n \in \{0, 1, 2, \ldots\}$, $k = \{0, 1, \ldots, n\}$, and $p \in [0, 1]$ (the binomial distribution with parameters n and p). Let $\pi(k; \lambda) = e^{-\lambda} \lambda^k / k!$ for $\lambda \in (0, \infty)$ and $k \in \{0, 1, 2, \ldots\}$ (the Poisson distribution with parameter λ).

Lemma 93 (Poisson Limit Lemma.). Let $0 < p_n < 1$, let $0 < \lambda < \infty$, and suppose $np_n \to \lambda$. Then fix k. For $n \geq k$, $\binom{n}{k} p_n^k (1-p_n)^{n-k} \to e^{-\lambda} \lambda^k / k!$.

Proof.

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{p_n}{1-p_n}\right)^k (1-p_n)^n$$
$$= \underbrace{(1)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}_{\to 1}(1-p_n)^n \frac{1}{k!}\underbrace{\left(\frac{np_n}{1-p_n}\right)^k}_{\to \lambda^k}$$

Thus it remains to show that $(1-p_n)^n \to e^{-\lambda}$. As $h \to 0$, $(1+h)^{1/h} = [e^{\ln(1+h)}]^{1/h} = e^{\frac{\ln(1+h)-\ln(1)}{h}} \to e^{\ln'(1)} = e^1 = e$. Hence

$$(1-p_n)^n = \left[(1-p_n)^{1/-p_n}\right]^{-np_n} \to [e]^{-\lambda} = e^{-\lambda}.$$

(We've used the fact that $(a,b) \mapsto a^b = e^{b\ln(a)}$ is continuous on $(0,\infty)\times\mathbb{R}$.) $\square$

Remark. Let $0 < \lambda < \infty$.

$$\sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

(Consequence: if $np_n \to \lambda$, then

$$\sum_{k=0}^{\infty} |\pi(k;\lambda) - \underbrace{b(k;n,p_n)}_{0 \text{ when } k>n}| \to 0.$$

by Scheffe's theorem (cf. Extra Problem 26).)

Suppose $P(X=k) = e^{-\lambda}\lambda^k/k!$ for $k = 0,1,2,\ldots$ Then

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \lambda.$$

Similarly,

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2\sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2$$

so $E(X^2) - E(X) = \lambda^2$ so $E(X^2) = \lambda^2 + \lambda$, so $\operatorname{var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

3/24/21

Definition. Let μ be a Borel probability measure on $\mathbb{R}$. The characteristic function of μ is

$$\varphi_\mu(u) = \int_{\mathbb{R}} e^{-ixu}\, d\mu(x)$$

for $u \in \mathbb{R}$. (φ_μ is uniformly continuous.)

Fact. If $\varphi_{\mu_1} = \varphi_{\mu_2}$, then $\mu_1 = \mu_2$.

Remark. $\varphi_\mu(-2\pi u) = \int_{\mathbb{R}} e^{-2\pi ixu}\, d\mu(x) = \widehat{\mu}(u)$, the Fourier Transform of μ.

Remark. If $\mu_n \Rightarrow \mu$, then for each $u \in \mathbb{R}$, $\varphi_{\mu_n}(u) \to \varphi_\mu(u)$, because $x \mapsto e^{ixu}$ is a bounded continuous function on $\mathbb{R}$.

Fact (Upcoming homework.)**.** If $d\mu(x) = e^{-x^2/2}/\sqrt{2\pi}\, dx$, (i.e. if $\mu = N(0,1)$), then $\varphi_\mu(u) = e^{-u^2/2}$.

(If $f(x) = e^{-\pi x^2}$ for all $x \in \mathbb{R}$, then

$$\widehat{f}(\xi) := \int_{\mathbb{R}} e^{-2\pi ix\xi} f(x)\, dx = e^{-\pi\xi^2}.$$

For $x \in \mathbb{R}^d$, $\widehat{f}(\xi) := \int_{\mathbb{R}} e^{-2\pi ix\cdot\xi} f(x)\, dx = e^{-\pi|\xi|^2}$.)

Theorem 94 (Lévy-Cramer Continuity Theorem (1920s))**.** Let (μ_n) be a sequence of Borel probability measures on $\mathbb{R}$. Suppose $\varphi_{\mu_n} \to \varphi$ pointwise on $\mathbb{R}$ and φ is continuous at 0. Then $\varphi = \varphi_\mu$ for some Borel probability measure μ on $\mathbb{R}$ and $\mu_n \Rightarrow \mu$.

Corollary 95 (Polya, ∼1920.)**.** If $\varphi_{\mu_n}(u) \to e^{-u^2/2}$ for each $u \in \mathbb{R}$, then $\mu_n \Rightarrow N(0,1)$.

Remark. To say $\mu = N(0,1)$ means μ is the Borel probability measure on $\mathbb{R}$ such that for all $A \in \text{Borel}(\mathbb{R})$, $\mu(A) = \int_A e^{-x^2/2}/\sqrt{2\pi}\, dx$.

Poisson Processes

Definition. A simple counting process is a family $(N_t)_{0 \le t < \infty}$ of random variables such that for each $\omega \in \Omega$, there is a set $J(\omega) \subseteq (0,\infty)$ such that for each $t \in [0,\infty)$, $N_t(\omega) = |J(\omega) \cap (0,t]| < \infty$.

Note. Equivalently, $(N_t)_{0 \le t < \infty}$ is a simple counting process if and only if for each $\omega \in \Omega$, $t \mapsto N_t(\omega)$ is a right-continuous function from $[0,\infty) \to \{0,1,2,\dots\}$ such that $N_0(\omega) = 0$ and for each $t \in [0,\infty)$, $N_t(\omega) - N_{t^-}(\omega) \in \{0,1\}$ (where $N_{t^-} = \lim_{s \to t^-} N_s$).

Definition. A Poisson process is a simple counting process $(N_t)_{0 \le t < \infty}$ such that

(A) (N_t) has independent increments; i.e. for each $n \in \mathbb{N}$, for all $t_0, t_1, \dots, t_n$, if $0 \le t_0 < t_1 < \cdots < t_n < \infty$, then $N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}$ are independent.

(B) $(N_t)_{0 \le t < \infty}$ has stationary increments; i.e., for $s,t,u,v \in [0,\infty)$, if $s < t$ and $u < v$ and $t - s = u - v$, then $\text{law}(N_t - N_s) = \text{law}(N_v - N_u)$.

Example. N_t = the number of cosmic rays detected in $(0,t]$.

Questions.

1. Why Poisson?
2. Uniqueness? (up to change of time scale)
3. Existence?

Lemma 96 (Lemma 1.)**.** Suppose (A_n) is a sequence of events such that for each ω, there exists n such that for each $n' \ge n$, we have $\omega \in A_{n'}$. Then $P(A_n^c) \to 0$.

Proof. Let $B_n = \bigcap_{n' \ge n} A_{n'}$. Then for each ω, there exists n such that $\omega \in B_n$. In other words, $\bigcup_n B_n = \Omega$. But also $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots$ Hence $P(B_n) \uparrow 1$. But for each n, $B_n \subseteq A_n$, so $A_n^c \subseteq B_n^c$. Hence

$$0 \le P(A_n^c) \le P(B_n^c) = 1 - P(B_n) \to 0.$$

Hence $P(A_n^c) \to 0$ by the squeeze theorem. □

Corollary 97. Let $X, X_1, X_2, \dots$ be real random variables. Suppose for each $\omega \in \Omega$, there exists n such that for each $n' \ge n$, we have $X_n(\omega) = X(\omega)$. Then $P(X_n \ne X) \to 0$.

Proof. Apply the lemma with $A_n = \{X_n = X\}$. □

Notation. For any sets E and F, the symmetric difference of E and F is written $E \triangle F \coloneqq (E \setminus F) \cup (F \setminus E)$.

Lemma 98 (Lemma 2)**.** Let E and F be events. Then $|P(E) - P(F)| \le P(E \triangle F)$.

Proof 1. $E = (E \cap F) \uplus (E \setminus F)$ and $F = (E \cap F) \uplus (F \setminus E)$. Hence

$$\begin{aligned}
|P(E) - P(F)| &= |P(E \cap F) + P(E \setminus F) - [P(E \cap F) + P(F \setminus E)]| \\
&= |P(E \setminus F) - P(F \setminus E)| \\
&\le |P(E \setminus F)| + |-P(F \setminus E)| \\
&= P(E \setminus F) + P(F \setminus E) = P(E \triangle F).
\end{aligned}$$

□

Proof 2.

$$\begin{aligned}|P(E) - P(F)| &= \left|\int 1_E \, dP - \int 1_F \, dP\right| \\ &= \left|\int 1_E - 1_F \, dP\right| \\ &\le \int |1_E - 1_F| \, dP = \int 1_{E \triangle F} \, dP = P(E \triangle F).\end{aligned}$$

□

Lemma 99 (Lemma 3.)**.** Let $X, X_1, X_2, \ldots$ be real random variables. Suppose $P(X_n \neq X) \to 0$. Then for each Borel set $A \subseteq \mathbb{R}$, $P(X_n \in A) \to P(X \in A)$.

Proof. Consider any such A. Let $F_n = \{X_n \in A\}$ and $F = \{X \in A\}$. For each ω, if $\omega \in F \setminus F_n$, then $X(\omega) \in A$ and $X_n(\omega) \notin A$, so $X(\omega) \neq X_n(\omega)$. Similarly, if $\omega \in F_n \setminus F$, then $X_n(\omega) \in A$ and $X(\omega) \notin A$, so $X_n(\omega) \neq X(\omega)$. Thus $F_n \triangle F \subseteq \{X_n \neq X\}$. Hence $P(F_n \triangle F) \to 0$. By Lemma 2 (cf. Lemma 98), $|P(F) - P(F_n)| \le P(F \triangle F_n)$. Hence $P(F_n) \to P(F)$. In other words, $P(X_n \in A) \to P(X \in A)$. □

3/26/21

Reminder. Suppose $P(X_n \neq X) \to 0$. Then for each Borel set $A \subseteq \mathbb{R}$, $P(X_n \in A) \to P(X \in A)$.

Theorem 100 (Theorem 1)**.** Let (N_t) be a Poisson process. Then there is a number $\lambda \in [0, \infty)$ such that for $0 \le s < t < \infty$, $N_t - N_s$ is Poisson distributed with parameter $\lambda(t - s)$. (We call λ the intensity of (N_t).)

We have $\lambda = E(N_1)$, $P(N_1 = 0) = e^{-\lambda}$, $\lambda = -\ln P(N_1 = 0)$.

Proof of Theorem 1 (100). Instead of (A) and (B) from the definition of a Poisson process, we'll use the following seemingly weaker assumptions:

(A') For each $n \in \mathbb{N}$, for all $t_0, \ldots, t_n$, if $0 < t_0 < t_1 < \cdots < t_n < \infty$, then the events $\{N_{t_1} - N_{t_0} \ge 1\}, \ldots, \{N_{t_n} - N_{t_{n-1}} \ge 1\}$ are independent.

(B') For all $s, t, u, v \in [0, \infty)$, if $s < t$, $u < v$, and $t - s = v - u$, then $P(N_t - N_s \ge 1) = P(N_v - N_u \ge 1)$.

First, let $t \in [0, \infty)$ and let"s show that N_t is Poisson distributed with some parameter $\lambda_t \in [0, \infty)$.For each $n \in \mathbb{N}$, for each $i \in \{1, \ldots, n\}$, let $\Delta_{n,i} = N_{it/n} - N_{(i-1)t/n}$, $A_{n,i} = \{\Delta_{n,i} \ge 1\}$, the event that (N_u) jumps at least once during the time interval $(\frac{i-1}{n}t, \frac{i}{n}t]$.

Let $X_n = \sum_{i=1}^n 1_{A_{n,i}}$. Then $X_n \le N_t$. Consider any $\omega \in \Omega$. Then there is a set $J(\omega) \subseteq [0, \infty)$ such that for each $u \in [0, \infty)$, $N_u(\omega) = |J(\omega) \cap (0, u]| < \infty$. Hence for all u, v, if $0 \le u < v < \infty$, then $N_v(\omega) - N_u(\omega) = |J(\omega) \cap (u, v]|$. Choose $n_0 \in \mathbb{N}$ such that $t/n_0 < \inf\{v - u : u, v \in J(\omega) \cap (0, t], u < v\}$ (min if the set $\neq \varnothing$, $\inf \varnothing = \infty$). Then for each $n \ge n_0$, $X_n(\omega) = N_t(\omega)$. Since this holds for each $\omega \in \Omega$, $P(X_n \neq N_t) \to 0$ for each set $A \subseteq \{0, 1, 2, \ldots\}$, $P(X_n \in A) \to P(N_t \in A)$.

Remark. By a similar argument, (A') $\Longrightarrow$ (A). To see this, approximate each increment by random variables defined analogously to the X_ns.

Now by (B'), for each n, $P(A_{n,i})$ is the same for all i, say $= p_n$. By (A'), the events $A_{n,1}, \ldots, A_{n,n}$ are independent. Hence X_n is binomially distributed, $P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}$ for $k = 0, 1, \ldots, n$. Let $\lambda_t = -\ln P(N_t = 0)$.

Claim. $\lambda_t < \infty$ and $np_n \to \lambda_t$.

Proof of Claim. $P(N_t = 0) = P\left(\bigcap_{i=1}^n A_{n,i}^c\right) = (1 - p_n)^n$. Hence $P(N_t = 0) \neq 0$, for otherwise, for each n, $p_n = 1$, $P(A_{n,i}) = 1$, $P(X_n = n) = 1$, $P(N_t \ge n) = 1$, but $\{N_t \ge n\} \downarrow \{N_t = \infty\} = \varnothing$, so $P(N_t \ge n) \downarrow 0$, a contradiction. Hence $P(N_t = 0) \neq 0$, as stated. Thus $\lambda_t < \infty$.

Next, $-n\ln(1-p_n) = -\ln[(1-p_n)^n] = -\ln P(N_t = 0) = \lambda_t < \infty$ so $\ln(1-p_n) = -\lambda_t/n \to 0$, so $p_n \to 0$. But

$$-n\ln(1-p_n) = -n\left(-p_n - \frac{p_n^2}{2} - \frac{p_n^3}{3} - \cdots\right) = np_n\left(1 + \frac{p_n}{2} + \frac{p_n^2}{3} + \cdots\right).$$

As $n \to \infty$, we have $p_n \to 0$, so $1 + p_n/2 + p_n^2/3 + \cdots \to 1$, so $np_n \to \lambda_t$. □

Thus for each integer $k \geq 0$, as $n \to \infty$, $P(X_n = k) \to e^{-\lambda_t}\lambda_t^k/k!$. But earlier, we saw that for each $A \subseteq \{0, 1, d, \dots\}$, $P(X_n \in A) \to P(N_t \leq A)$. Hence $P(N_t = k) = e^{-\lambda_t}\lambda_t^k/k!$. Hence for all $s, t \in (0, \infty)$,

$$\operatorname{law}(N_{t+s} - N_s) = \operatorname{law}(N_t - N_0) = \operatorname{law}(N_t) = \operatorname{Poisson}(\lambda_t).$$

(Thus (A') and (B') $\implies$ (A) and (B).)

Now,

$$\begin{aligned}
\lambda_{s+t} &= -\ln P(N_{t+s} = 0) = -\ln P[(N_{t+s} - N_t) + N_t = 0] \\
&= -\ln P(N_{t+s} - N_t = 0 \text{ and } N_t \geq 0) \\
&= -\ln[P(N_{t+s} - N_t = 0)P(N_t = 0)] \\
&= -\ln P(N_{t+s} - N_t = 0) - \ln P(N_t = 0) = \lambda_s + \lambda_t.
\end{aligned}$$

Hence $\lambda_t = t\lambda_1$ for $t \in \mathbb{Q} \cap [0, \infty)$. As t increases, $\{N_t = 0\}$ decreases, so $\ln P(N_t = 0)$ decreases, so $\lambda_t = -\ln P(N_t = 0)$ increases. Hence $\lambda_t = t\lambda_1$ for all $t \in [0, \infty)$.

Digression. Let $a \in (1, \infty)$, $x \in (0, \infty)$. For $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, $m/n < \log_a x$ if and only if $m < n\log_a x$ if and only if $m\log_a a < n\log_a x$ if and only if $\log_a a^m < \log_a x^n$ if and only if $a^m < x^n$. So $\log_a x = \sup\{m/n : m \in \mathbb{Z}, n \in \mathbb{N}, a^m < x^n\}$. $x \mapsto \log_a x$ is a strictly increasing function from $(0, \infty)$ onto $\mathbb{R}$, so the inverse function is also increasing.

Hence we can take $\lambda = \lambda_1$. (For $s \in [0, \infty)$, the process $(N_{u+s} - N_s)_{0 \leq u < \infty}$ satisfies (A') and (B') and we have $\lambda_1 = -\ln P(N_{1+s} - N_s = 0)$. Hence $\operatorname{law}(N_{u+s} - N_s) = \operatorname{Poisson}(\lambda u)$. (In particular, (A') and (B') $\implies$ (A) and (B).) □

3/29/21

Something to add the the end of the proof of the last theorem. The λ for $(N_{u+s} - N_s)_{0 \leq u < \infty}$ is the same as the λ for $(N_t)_{0 \leq t < \infty}$ because $\lambda = -\ln P(N_1 = 0) = -\ln P(N_1 - N_0 = 0) = -\ln P(N_{1+s} - N_s = 0)$.

Theorem 101 (Theorem 2. "Uniqueness")**.** Let (M_t) and (N_t) be Poisson processes with rates ("intensities") λ_1 and λ_2, respectively. Assume $\lambda_1 \neq 0$. Let $L_t = M_{\lambda_2 t/\lambda_1}$ for $0 \leq t < \infty$. Then (L_t) and (N_t) have the same finite dimensional joint distributions.

Lemma 102. Let $(\mathbf{X}, \mathscr{A})$ and $(\mathbf{Y}, \mathscr{B})$ be measurable spaces. Let X, X' be random variables in $(\mathbf{X}, \mathscr{A})$ such that $\operatorname{law}(X) = \operatorname{law}(X')$. Let $f : \mathbf{X} \to \mathbf{Y}$ be measurable. Then $\operatorname{law}(f(X)) = \operatorname{law}(f(X'))$.

Remark. X and X' could even be on different probability spaces $(\Omega, \mathscr{F}, P)$ and $(\Omega,', \mathscr{F}', P')$ respectively.

Notation. If Y is a random variable in $(\mathbf{Y}, \mathscr{B})$, then $\operatorname{law}(Y) = \nu$ where ν is the function on $\mathscr{B}$ defined by $\nu(B) = P(Y \in B)$.

Proof. For each $B \in \mathscr{B}$,

$$\begin{aligned}
\operatorname{law}(f(X))(B) &= P(f(X) \in B) = P(X \in f^{-1}[B]) = \operatorname{law}(X)(f^{-1}[B]) \\
&= \operatorname{law}(X')(f^{-1}[B]) = P'(X' \in f^{-1}[B]) = P'(f(X') \in B) = \operatorname{law}(f(X'))(B).
\end{aligned}$$

This holds for all $B \in \mathscr{B}$. Hence $\operatorname{law}(f(X)) = \operatorname{law}(f(X'))$. □

Proof of Theorem 2 (101). Let $n \in \mathbb{N}$ and let $0 \le t_1 < t_2 < \cdots < t_n < \infty$. We wish to show that $\text{law}(N_{t_1}, N_{t_2}, \dots, N_{t_n}) = \text{law}(L_{t_1}, L_{t_2}, \dots, L_{t_n})$.

$$E(L_1) = E(M_{\lambda_2 \cdot 1/\lambda_1}) = \lambda_1 \cdot \frac{\lambda_2 \cdot 1}{\lambda_1} = \lambda_2.$$

(In other words, L and N "have the same λ.") Hence for $0 \le s < t < \infty$, $L_t - L_s$ and $N_t - N_s$ are both Poisson distributions with parameter λ_2. Note that $N_{t_1} = N_{t_1} - N_0$ and $L_{t_1} = L_{t_1} - L_0$. Hence, by independence of increments, for all Borel sets $A_1, \dots, A_n \subseteq \mathbb{R}$,

$$\begin{aligned}
&P(N_{t_1} \in A_1, N_{t_2} - N_{t_1} \in A_2, \dots, N_{t_n} - N_{t_{n-1}} \in A_n) \\
&= P(N_{t_1} \in A_1) P(N_{t_2} - N_{t_1} \in A_2) \cdots P(N_{t_n} - N_{t_{n-1}} \in A_n) \\
&= P(L_{t_1} \in A_1) P(L_{t_2} - L_{t_1} \in A_2) \cdots P(L_{t_n} - L_{t_{n-1}} \in A_n) \\
&= P(L_{t_1} \in A_1, L_{t_2} - L_{t_1} \in A_2, \dots, L_{t_n} - L_{t_{n-1}} \in A_n)
\end{aligned}$$

so in other words, $P((N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}) \in A_1 \times A_2 \times \cdots \times A_n)$ so in still other words,

$$\begin{aligned}
&\text{law}(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}})(A_1 \times A_2 \times \cdots \times A_n) \\
&= \text{law}(L_{t_1}, L_{t_2} - L_{t_1}, \dots, L_{t_n} - L_{t_{n-1}})(A_1 \times A_2 \times \cdots \times A_n).
\end{aligned}$$

Hence, by a corollary of the π-λ theorem,

$$\text{law}(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}) = \text{law}(L_{t_1}, L_{t_2} - L_{t_1}, \dots, L_{t_n} - L_{t_{n-1}}).$$

Define $f : \mathbb{R}^n \to \mathbb{R}^n$ by $f(x_1, x_2, \dots, x_n) = (x_1, x_1 + x_2, \dots, x_1 + x_2 + \cdots + x_n)$. Then f is continuous, hence Borel measurable. We can apply the lemma with $\text{Borel}(\mathbf{X}, \mathscr{A}) = (\mathbf{Y}, \mathscr{B}) = (\mathbb{R}^n, \text{Borel}(\mathbb{R}^n))$, $X = (L_{t_1}, L_{t_2} - L_{t_1}, \dots, L_{t_n} - L_{t_{n-1}})$, and $X' = (N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}})$. We get $\text{law}(f(X)) = \text{law}(f(X'))$. But $f(X) = (L_{t_1}, L_{t_2}, \dots, L_{t_n})$ and $f(X') = (N_{t_1}, N_{t_2}, \dots, N_{t_n})$. ☐

Theorem 103 (Theorem 3. "Existence")**.** (Think of cosmic rays.)

Proof. Let $\lambda \in (0, \infty)$. Let $T_1, T_2, T_3, \dots$ be independent random variables, each exponential distributed with parameter λ. (So $P(T_k > t) = e^{-\lambda t}$ for $0 \le t < \infty$.) Let $S_n = \sum_{k=1}^n T_k$. $E(T_k \wedge 1) = E(T_1 \wedge 1) > 0$ so $\sum_{k=1}^\infty E(T_k \wedge 1) = \infty$ so $\sum_{k=1}^\infty (T_k \wedge 1) = \infty$ almost surely (cf. Extra Problem 31).

Notation. $a \wedge b \coloneqq \min\{a, b\}$.

Hence $\sum_{k=0}^\infty T_k = \infty$ almost surely. Hence, without loss of generality, for each $\omega \in \Omega$, $\sum_{k=0}^\infty T_k(\omega) = \infty$. Also $P\left(\bigcap_{k=1}^\infty \{T_k > 0\}\right) = 1$ so without loss of generality, for each $k \in \mathbb{N}$, for each $\omega \in \Omega$, $T_k(\omega) > 0$. For $0 \le t < \infty$, define N_t on Ω by $N_t(\omega) = |\{n \ge 1 : S_n(\omega) \le t\}| = |J(\omega) \cap (0, t]|$, where $J(\omega) = \{S_1(\omega), S_2(\omega), S_3(\omega), \dots\} \subseteq (0, \infty)$. Then $(N_t)_{0 \le t < \infty}$ is clearly a simple counting process (thanks to the WLOGs).

In about 8 pages, it can be shown that (N_t) is a Poisson process. ☐

Digression. A real random variable T with $P(T > t + s | T > s) = P(T > t)$ is said to be memoryless. In particular, consider $P(T > t) = e^{-\lambda t}$ for some $\lambda > 0$. Also something about the waiting time paradox.

Poisson Random Measures Let $(E, \mathscr{E})$ be a measurable space. Let λ be a measure on $\mathscr{E}$. Suppose $0 < \lambda(E) < \infty$. Let $p - \lambda/\lambda(E)$. Then p is a probability measure on $\mathscr{E}$. Let $N, X_1, X_2, X_3, \dots$ be independent random variables with N taking values in $\{0, 1, 2, \dots\}$, $\text{law}(N) = \text{Poisson}(\lambda(E))$, $X_1, X_2, X_3, \dots$ taking values in E, and $\text{law}(X_j) = p$ for $j = 1, 2, 3, \dots$ For each $\omega \in \Omega$, for each $A \in \mathscr{E}$, let $N(\omega, A) = |\{j \le N(\omega) : X_j(\omega) \in A\}|$ (so $N(\omega, E) = N(\omega)$). For each $\omega \in \Omega$, the function $A \mapsto N(\omega, A) = \sum_{j=1}^{N(\omega)} \delta_{X_j(\omega)}(A)$ is a finite measure on $\mathscr{E}$ taking values in the set $\{0, 1, \dots, N(\omega)\}$ and for each $A \in \mathscr{E}$, the function $\omega \mapsto N(\omega, A)$ is $\mathscr{F}$-measurable. Both of these assertions follow from the equalities

$$N(\omega, A) = \sum_{j=1}^\infty 1_{[j, \infty)}(N(\omega)) \delta_{X_j(\omega)}(A) = \sum_{j=1}^\infty 1_{[j, \infty)}(N(\omega)) 1_A(X_j(\omega)).$$

For each $A \in \mathscr{E}$, let us also write $N(A)$ for the function $\omega \mapsto N(\omega, A)$. Let $A_1, \dots, A_r$ be disjoint sets belonging to $\mathscr{E}$ with $A_1 \cup \cdots \cup A_r = E$. We claim that the random variables $N(A_1), \dots, N(A_r)$ are independent and for $\ell = 1, \dots, r$, $\text{law}(N(A_\ell)) = \text{Poisson}(\lambda(A_\ell))$.

Let $n_1, \dots, n_r \in \{0, 1, 2, \dots\}$. We wish to show that

$$P(N(A_1) = n_1, \dots, N(A_r) = n_r) = \prod_{\ell=1}^{r} e^{-\lambda(A_\ell)} \frac{\lambda(A_\ell)^{n_\ell}}{n_\ell!}.$$

Let $n = n_1 + \cdots + n_r$. For $\ell = 1, \dots, r$, let $Z_\ell = |\{j \le n : X_\ell \in A_\ell\}|$. Observe that for $\ell = 1, \dots, r$, we have $N(A_\ell) = Z_\ell$ on $\{N = n\}$.

4/2/21

Reminder. Let $(E, \mathscr{E})$ be a measurable space and define λ on $\mathscr{E}$ such that $0 < \lambda(E) < \infty$. Let $p = \lambda/\lambda(E)$ and let $N, X_1, X_2, X_3, \dots$ be independent random variables with $\text{law}(N) = \text{Poisson}(\lambda(E))$, $\text{law}(X_j) = p$, and $N(\omega, A) = |\{j < N(\omega) : X_j(\omega) \in A\}|$ (where $\omega \in \Omega$, $A \in \mathscr{E}$, $N(A)(\omega) = N(\omega, A)$ $(N(E) = N)$). Let $A_1, \dots, A_r$ be disjoint elements of $\mathscr{E}$ with $A_1 \cup \cdots \cup A_r = E$. We claim that $N(A_1), \dots, N(A_r)$ are independent and $\text{law}(N(A_\ell)) = \text{Poisson}(\lambda(A_\ell))$ for $\ell = 1, \dots, r$. Let $n_1, \dots, n_r \in \{0, 1, 2, \dots\}$. We wish to show that

$$P(N(A_1) = n_1, \dots, N(A_r) = n_r) = \prod_{\ell=1}^{r} e^{-\lambda(A_\ell)} \lambda(A_\ell)^{n_\ell}/n_\ell!.$$

Let $n = n_1 + \cdots + n_r$. For $\ell = 1, \dots, r$, let $Z_\ell = |\{j \le n : X_j \in A_\ell\}|$. For $\ell = 1, \dots, r$, on $\{N = n\}$, we have $N(A_\ell) = Z_\ell$.

Also, for $\ell = 1, \dots, r$, Z_ℓ is measurable with respect to $\sigma(X_1, \dots, X_n)$, because $Z_\ell = \sum_{j=1}^{n} 1_{A_\ell}(X_j)$. Thus $(Z_1, \dots, Z_n)$ and N are independent. Now

$$P(N(A_1) = n_1, \dots, N(A_r) = n_r) = P(N(A_1) = n_1, \dots, N(A_r) = n_r, N = n)$$

(since $E = \bigcup_{\ell=1}^{r} A_\ell, N = N(E) = \sum_{\ell=1}^{r} N(A_\ell)$)

$$\begin{aligned} &= P(Z_1 = n_1, \dots, Z_r = n_r, N = n) \\ &= P(Z_1 = n_1, \dots, Z_r = n_r) P(N = n) \end{aligned}$$

(since $(Z_1, \dots, Z_r)$ and N are independent).

For $j = 1, \dots, n$, define $Y_j : \Omega \to \{1, \dots, r\}$ by

$$Y_j = \begin{cases} 1 & \text{on } \{X_j \in A_1\} \\ \vdots & \vdots \\ r & \text{on } \{X_j \in A_j\}. \end{cases}$$

$Y_1, \dots, Y_r$ are independent, $P(Y_j = \ell) = P(X_j \in A_\ell) = p(A_\ell)$ Now for $\ell = 1, \dots, r$, $Z_\ell = |\{j \le n : Y_j = \ell\}|$. Thus $(Z_1, \dots, Z_r)$ has a <u>multinomial distribution</u> with parameters $n, r, p(A_1), \dots, p(A_r)$, so

$$P(Z_1 = n_2, \dots, Z_\ell = n_\ell) = \frac{n!}{n_1! \cdots n_r!} p(A_1)^{n_1} \cdots p(A_r)^{n_r} = \frac{n!}{n_1 \cdots n_r} \frac{\lambda(A_1)^{n_1}}{\lambda(E)^{n_1}} \cdots \frac{\lambda(A_r)^{n_r}}{\lambda(E)^{n_r}}.$$

Also,

$$P(N = n) = e^{-\lambda(E)} \frac{\lambda(E)^n}{n!} = e^{-[\lambda(A_1) + \cdots + \lambda(A_r)} \frac{\lambda(E)^{n_1} \cdots \lambda(E)^{n_r}}{n!}.$$

Hence

$$P(N(A_1) = n_1, \dots, N(A_r) = n_r) = e^{-\lambda(A_1)} \frac{\lambda(A_1)^{n_1}}{n_1!} \cdots e^{-\lambda(A_r)} \frac{\lambda(A_r)^{n_r}}{n_r!},$$

as desired.

Definition. The family of random variables $(N(A))_{A\in\mathscr{E}}$ is called a Poisson random measure with mean measure λ.

Now suppose in addition that the measurable space $(E,\mathscr{E})$ is countably separated. (e.g. $\mathscr{E} = \text{Borel}(E)$, where $E = \mathbb{R}$ or $\mathbb{R}^d$ or any separable metric space, or any second countable T_0 space). Note that then for each $E \in \mathscr{E}$, we have $\{x\} \in \mathscr{E}$, because if (H_j) is a sequence in $\mathscr{E}$ which separates points in E, and if we let

$$G_j = \begin{cases} H_j & \text{if } x \in H_j, \\ E \setminus H)j & \text{if } x \notin H_j, \end{cases}$$

then $\{x\} = \bigcap_j G_j$. Assume $\lambda(\{x\}) = 0$ for each $x \in E$. Let $\Delta = \{(x,x) : x \in E\}$. Then $\Delta \in \mathscr{E} \otimes \mathscr{E}$, because if (H_j) is a sequence in $\mathscr{E}$ which separates points in E, then $\Delta = (E \times E) \setminus \bigcup_j [(H_j \times H_j^c) \cup (H_j^c \times H_j)]$. For all $j_1, j_2 \in \{1,2,3,\dots\}$, if $j_1 \neq j_2$, then

$$\begin{aligned} P(X_{j_1} = X_{j_2}) = P[(X_{j_1}, X_{j_2}) \in \Delta] = (p \otimes p)(\Delta) &= \int_E \int_E 1_\Delta(x,y)\, p(dy)p(dx) \\ = \int_E \int_E 1_{\{x\}}(y)\, p(dy)p(dx) = \int_E p(\{x\})\, p(dx) &= \int_E 0\, p(dx) = 0. \end{aligned}$$

Let $\Omega_1 = \{\omega \in \Omega : X_{j_1}(\omega) \neq X_{j_2}(\omega) \text{ for all } j_1 \neq j_2\}$. Then $P(\Omega_1) = 1$. Redefine N to 0 on $\Omega \setminus \Omega_1$ and redefine $N(\omega, A)$ and $N(A)$ accordingly. Then $(N(A))_{A\in\mathscr{E}}$ is still a Poisson random measure with mean measure λ but now for each $\omega \in \Omega$ and for each $A \in \mathscr{E}$, $N(\omega, A) = |\Pi(\omega) \cap A|$, where $\Pi(\omega) = \{X_j(\omega) L j \leq N(\omega)\}$. (When $N(\omega) = 0$, $\Pi(\omega) = \varnothing$.) The random set Π is an example of a Poisson random set with mean measure λ.

Digression (On Extra Problem 32.)**.** Let X, Y be sets, $f : X \to Y$, $\mathscr{H} \subseteq \mathcal{P}(Y)$, $\mathscr{B} = \sigma(\mathscr{H})$ on Y. $\mathscr{G} = \{f^{-1}[H] : H \in \mathscr{H}\}$, $\mathscr{A} = \sigma(\mathscr{G})$ on X, $\mathscr{A}' = \{f^{-1}[B] : B \in \mathscr{B}\}$. Show $\mathscr{A} = \mathscr{A}'$.

Solution. Let $\mathscr{B}' = \{B \subseteq Y : f^{-1}[B] \in \mathscr{A}\}$. Then $\mathscr{B}'$ is a σ-field on Y and $\mathscr{H} \subseteq \mathscr{B}'$. Now $\mathscr{B}$ is the smallest σ-field on Y that contains $\mathscr{H}$. Hence $\mathscr{B} \subseteq \mathscr{B}'$, so $\mathscr{A}' \subseteq \mathscr{A}$.

Now let's show $\mathscr{A} \subseteq \mathscr{A}'$. Since $\mathscr{B}$ is a σ-field on Y and passage to preimages preserves set operations, $\mathscr{A}'$ is a σ-field on X. Let $G \in \mathscr{G}$. Then $G = f^{-1}[H]$ for some $H \in \mathscr{H} \subseteq \mathscr{B}$. Hence $G \in \mathscr{A}'$. Thus $\mathscr{G} \subseteq \mathscr{A}'$. But $\mathscr{A}$ is the smallest σ-field on X that contains $\mathscr{G}$. Hence $\mathscr{A} \subseteq \mathscr{A}'$. We've shown that $\mathscr{A}' \subseteq \mathscr{A}$ and $\mathscr{A} \subseteq \mathscr{A}'$. Hence $\mathscr{A} = \mathscr{A}'$. □

04/05/21

Remarks on Product Measure

Definition. To say that a measure is s-finite means that it is a sum of countably many finite measures.

Remark. We say λ is the image of μ (under a measurable function f) if $\mu(B) = \lambda(f^{-1}[B])$.

1. A σ-finite measure is s-finite.
2. An image of a σ-finite measure need not be σ-finite (e.g. map a space with infinite measure to one point).
3. An image of an s-finite measure is s-finite.
4. Any s-finite measure is an image of some σ-finite measure. (Let μ be s-finite. Then $\mu = \sum_n \mu_n$ where each μ_n is a finite measure. Let $\lambda = \sum_n \mu_n \otimes \delta_n$. Then λ is σ-finite and μ is the image of λ under π_1.)

Theorem (Product measures and Fubini's Theorem (Telegraphic Statement))**.** Let $(X, \mathscr{A}, \lambda)$ and $(Y, \mathscr{B}, \mu)$ be s-finite measure spaces. Then for all $\mathscr{A} \otimes \mathscr{B}$-measurable functions $f : X \times Y \to [0, \infty]$,

$$\int_X \underbrace{\int_Y f(x,y)\, \mu(dy)}_{\mathscr{A}\text{-measurable}}\, \mu(dx) = \int_Y \underbrace{\int_X f(x,y)\, \lambda(dx)}_{\mathscr{B}\text{-measurable}}\, \mu(dy).$$

Telegraphic Proof. Prove equality first for $1_{A\times B}$, then for $f = 1_C$, then for simple functions, then measurable functions. Prove the theorem beginning with λ and μ finite measures, then apply the following fact:

Fact. If $\nu = \sum_{i\in I} \nu_i$, then for any measurable $h \geq 0$, $\int h\, d\nu = \sum_{i\in I} \int h\, d\nu_i$.

Define ν on $\mathscr{A} \otimes \mathscr{B}$ by $\nu(C) = \int_X \int_Y 1_C(x, y)\, \mu(dy)\, \lambda(dx) = \int_Y \int_X 1_C(x, y)\, \lambda(dx)\, \mu(dy)$. We denote ν by $\lambda \otimes \mu$ and we call ν the product of λ and μ. Then for all measurable $\mathscr{A} \otimes \mathscr{B}$-measurable $f : X \times Y \to [0, \infty]$,

$$\int_{X\times Y} f\, d\nu = \int_X \int_Y f(x, y)\, \mu(dy)\, \lambda(dx) = \int_Y \int_X f(x, y)\, \lambda(dx)\, \mu(dy).$$

For all $A \in \mathscr{A}$ and $B \in \mathscr{B}$,

$$\nu(A \times B) = \lambda(A)\mu(B). \tag{$*$}$$

When λ and μ are σ-finite, then ν is the only measure on $\mathscr{A} \otimes \mathscr{B}$ satisfying $(*)$. □

Example. Let m_1^0 be the Borel-Lebesgue measure on $\mathbb{R}$. Then $m_1^0 \otimes m_1^0$ is the Borel-Lebesgue measure on $\mathbb{R}^2$, and $\bigotimes_{i=1}^d m_1^0$ is the Borel-Lebesgue measure on $\mathbb{R}^d$.

Example. Let X and Y be random variables. Then X and Y are independent if and only if $\mathrm{law}(X, Y) = \mathrm{law}(X) \otimes \mathrm{law}(Y)$. Then generalize to random variables $X_1, \ldots, X_d$.

Example (Special Case of Last Example)**.** Let X be a random variable in $\mathbf{X} = \{x_1, \ldots, x_m\}$, x_i distinct. Let Y be a random variable in $\mathbf{Y} = \{y_1, \ldots, y_n\}$, y_j distinct. Let $\nu_{i,j} = P(X = x_i, Y = y_j)$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$. Let $\lambda_i = P(X = x_i)$ for $i = 1, \ldots, m$, $\mu_j = P(Y = y_j)$ for $j = 1, \ldots, n$. Then we have the following matrix $(v_{i,j})$:

$$\begin{pmatrix} \nu_{1,1} & \cdots & \nu_{1,n} \\ \vdots & \ddots & \vdots \\ \nu_{m,1} & \cdots & \nu_{m,n} \end{pmatrix}$$

where the sum over the i-th row is λ_i and the sum over the j-th column is μ_j: (these are the marginal distributions)

$$\lambda_i = \sum_{j=1}^n \nu_{i,j},\ \ \mu_j = \sum_{i=1}^n \nu_{i,j}$$

X and Y are independent if and only if $v_{i,j} = \lambda_i \mu_j$ for all i and j.

Example. Let X and Y be real random variables. Let $\nu = \mathrm{law}(X, Y)$. Let $\lambda = \mathrm{law}(X)$, $\mu = \mathrm{law}(Y)$. Then

$$\nu(C) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_C(x, y)\, \mu_x(dy)\, \lambda(dx) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_C(x, y)\, \lambda_y(dx)\, \mu(dy)$$

where $(\lambda_y)_{y\in\mathbb{R}}$ and $(\mu_x)_{x\in\mathbb{R}}$ are suitable families of Borel probability measures on $\mathbb{R}$. Intuitively, $\lambda_y = \mathrm{law}(X \mid Y = y)$, and $\mu_x = \mathrm{law}(Y \mid X = x)$. It may be that $P(Y = y) = 0$, but we can find a suitable family of measures so that in aggregate, this makes sense (Google "regular conditional probabilities", or "disintegration of measures"). This is still true if X and Y take values in "nice" measurable spaces, but not arbitrary ones.

Random Walks

Definition. A random walk in $\mathbb{R}$ is a sequence $S_0, S_1, S_2, \ldots$ such that $S_0 = 0$ and the increments $S_1 - S_0, S_2 - S_1, S_3 - S_2, \ldots$ are independent and identically distributed. ($S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent and identically distributed.)

We say the walk (S_n) is nondegenerate when $P(S_1 \neq 0) > 0$.

Theorem 104. Let (S_n) be a nondegenerate random walk in $\mathbb{R}$. Let $-\infty < a < 0 < b < \infty$. Let $N = \inf\{n \geq 0 : S_n \notin (a, b)\}$. ($N(\omega) = \inf\{n \geq 0 : S_n(\omega) \notin (a, b)\}$, for each $\omega \in \Omega$.) (N is the first exit time of (S_n) from a, b.) Then $E(N) < \infty$.

Proof. Note that $n = 0, 1, 2, \dots$ $N(\omega) > n$ if and only if $S_n(\omega) \in (a, b)$ for $m = 0, \dots, n$, so

$$\{N > n\} = \bigcap_{m=0}^{n} \{S_m \in (a, b)\} \in \mathscr{F}.$$

Thus N is a random variable. (N takes values in $\{0, 1, 2, \dots, \infty\}$).

Let $X_j = S_j - S_{j-1}$ for $j = 1, 2, 3, \dots$ Since $P(X_1 \neq 0) > 0$ and $\{X_1 \neq 0\} = \{X_1 > 0\} \cup \{X_1 < 0\}$, either $P(X_1 > 0)$ or $P(X_1 < 0) > 0$. The two cases are similar, so let's just consider the case where $P(X_1 > 0) > 0$. $\{X_1 > 0\} = \uparrow \bigcup_{k=1}^{\infty} \{X_1 \geq 1/k\}$, so $P(X_1 \geq 1/k) \uparrow P(X_1 > 0) > 0$. Hence there exists $\varepsilon > 0$ such that $P(X_1 \geq \varepsilon) > 0$. Choose $m \in \mathbb{N}$ such that $m\varepsilon > b - a$. Then for any $x \in (a, b)$,

$$P(x + S_m \notin (a, b)) \geq P(S_m \geq b - a) \geq P(X_1 \geq \varepsilon, \dots, X_m \geq \varepsilon) = P(X_1 \geq \varepsilon)^m,$$

so

$$P(N > m) = P(S_k \in (a, b) \text{ for } k = 1, \dots, n) \leq P(S_m \in (a, b)) = 1 - P(S_m \notin (a, b)) \leq 1 - (X_1 \geq \varepsilon)^m.$$

Now for $n = 1, 2, 3, \dots,$

$$\begin{aligned} P(N > (n+1)m) &= P(N > (n+1)m, N > nm) \leq P(S_{(n+1)m} \in (a, b), N > nm) \\ &= P(N > nm) - P(S_{(n+1)m} \notin (a, b), N > mn) \\ &\leq P(N > nm) - P(X_{nm+1} \geq \varepsilon, \dots, X_{(n+1)m} \geq \varepsilon, N > nm) \end{aligned}$$

($S_{(n+1)m} = S_{nm} + X_{nm+1} + \dots + X_{(n+1)m} \geq \varepsilon$.)

$$= P(N > nm)(1 - P(X_1 \geq \varepsilon)^m).$$

Hence, by induction, $P(N > nm) \leq (1 - P(X_1 \geq \varepsilon^m))^n$. Hence

$$E\left(\frac{N}{m}\right) \leq \sum_{n=0}^{\infty} P\left(\frac{N}{m} > n\right) \leq \frac{1}{P(X_1 \geq \varepsilon)^m} < \infty.$$

□

4/7/21

Definition. A filtration is an increasing sequence $\mathscr{F}_0 \subseteq \mathscr{F}_1, \subseteq \mathscr{F}_2 \subseteq \dots$ of sub-σ-fields of $\mathscr{F}$.

Example. Let (S_n) be a random walk. Let $\mathscr{F}_n = \sigma(S_0, \dots, S_n)$ for $n = 0, 1, 2, \dots$ Then $(\mathscr{F}_n)$ is a filtration called the natural filtration of (S_n). Notice that if $X_j = S_j - S_{j-1}$ for $j = 1, 2, \dots,$ then $S_n = X_1 + \dots + X_n$ $(S_0 = 0)$, so $\sigma(S_0, \dots, S_n) = \sigma(X_1, \dots, X_n)$.

Information in a σ-field.

Theorem 105. Let $(\mathbf{X}, \mathscr{A})$ be a measurable space. Let $X : \Omega \to \mathbf{X}$. Let $\mathscr{F}_X = \sigma(X) = \{f^{-1}[A] : A \in \mathscr{A}\}$. Let $Y : \Omega \to \mathbb{R}$. Then the following are equivalent:

(a) Y is $\mathscr{F}_X$-measurable.

(b) $Y = f(X)$ for some $\mathscr{A}$-measurable function $f : \mathbf{X} \to \mathbb{R}$.

Proof. (b) $\Longrightarrow$ (a). Suppose (b) holds. $f(X)$ means $f \circ X$. Let B be a Borel subset of $\mathbb{R}$. Then $f(X)^{-1}[B] = (f \circ X)^{-1}[B] = X^{-1}[f^{-1}[B]] \in \mathscr{F}_X$, because $f^{-1}[B] \in \mathscr{A}$. (That is, because the composition of measurable functions is measurable).

(a) $\Longrightarrow$ (b). Suppose (a) holds.

(1) Suppose $Y = 1_E$ for some $E \in \mathscr{F}_X$. Then $E = X^{-1}[A]$ for some $A \in \mathscr{A}$. Then $Y = 1_A \circ X$, because for each $\omega \in \Omega$, $Y(\omega) = 1$ if and only if $\omega \in E$ if and only if $X(\omega) \in A$ if and only if $1_A(X(\omega)) = 1$, and because Y and $1_A \circ X$ both map ω into $\{0, 1\}$.

(2) Suppose Y is $\mathscr{F}_X$-measurable. Then $Y[\Omega]$ is a finite subset of $\mathbb{R}$, say $Y[\Omega] = \{y_1, \ldots, y_n\}$, y_j distinct, and for $j = 1, \ldots, n$, $\{Y = y_j\} \in \mathscr{F}_X$, so there exists $B_j \in \mathscr{A}$ such that $\{Y = y_j\} = X^{-1}[B_j]$. Let $A_j = B_j \setminus \bigcup_{i<j} B_i$. Then $A_1, \ldots, A_n$ are disjoint elements of $\mathscr{A}$. Furthermore, for each $\omega \in \{Y = y_j\}$, we have $Y(\omega) = y_i$, so for $i \neq j$, $Y(\omega) \neq y_i$, so $Y(\omega) \notin B_i$. Hence for each $j = 1, \ldots, n$, $\{Y = y_i\} = X^{-1}[A_j]$. Notice that $Y[\Omega] \subseteq \bigcup_{j=1}^n A_j \subseteq \mathbf{X}$. Let $A_0 = \mathbf{X} \setminus \bigcup_{j=1}^n A_j$ and let $y_0 \in \mathbb{R}$. Define $f : \mathbf{X} \to \mathbb{R}$ by

$$f(x) = \begin{cases} y_0 & \text{if } \in A_0 \\ y_1 & \text{if } \in A_1 \\ & \vdots \\ y_n & \text{if } \in A_n. \end{cases}$$

Then f is $\mathscr{A}$-measurable (in fact, $\mathscr{A}$-simple. Finally $Y - f(X)$, because for $j = 1, \ldots, n$, for each $\omega \in \{Y = y_j\}$, $X(\omega) \in A_j$, so $f(X(\omega)) = y_j = Y(\omega)$. (The sets $\{Y = y_1\}, \ldots, \{Y = y_n\}$ form a partition of Ω.)

(3) The general case. Let $Y : \Omega \to \mathbb{R}$ be $\mathscr{F}_X$-measurable. Then there is a sequence (Y_n) of $\mathscr{F}_X$-simple functions $Y_n : \Omega \to \mathbb{R}$ such that for each $\omega \in \Omega$, $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$. By (2), for each n, there is an $\mathscr{A}$-simple function $g_n : \mathbf{X} \to \mathbb{R}$ such that for each $\omega \in \Omega$, $Y_n(\omega) = g_n(Y(\omega))$. Then, for each $\omega \in \Omega$, $\lim_{n\to\infty} g_n(X(\omega)) = \lim_{n\to\infty} Y_n(\omega) = Y(\omega)$. In particular, $X[\Omega] \subseteq C$, where $C = \{x \in \mathbf{X} : \lim_{n\to\infty} g_n(x) \text{ exists in } \mathbb{R}\}$. Note that $X \in \mathscr{A}$. (C $= \{-\infty < \liminf g_n = \limsup g_n < \infty\}$.) (If $\mathbb{R}$ is replaced by a separable metric space $\mathbf{Y}$, we still say that $\{x \in \mathbf{X} : (g_n(x)) \text{ is Cauchy }\} \in \mathscr{A}.$, so if in addition $\mathbf{Y}$is complete, then $C \in \mathscr{A}$.) Let $y_0 \in \mathbb{R}$. Define $f : \mathbf{X} \to \mathbb{R}$ by

$$f(x) = \begin{cases} \lim_{n\to\infty} g_n(x) & \text{if } x \in C, \\ y_0 & \text{if} x \in \mathbf{X} \setminus C. \end{cases}$$

Then for each $\omega \in \Omega$, $Y(\omega) = \lim_{n\to\infty} Y_n = \lim_{n\to\infty} g_n(X(\omega)) = f(X(\omega))$ (because $X[\Omega] \subseteq C$).

□

Corollary 106. Let $(\mathbf{X}_1, \mathscr{A}_1), \ldots, (\mathbf{X}_n, \mathscr{A}_n)$ be measurable spaces. For $j = 1, \ldots, n$, let $X_j : \Omega \to \mathbf{X}_j$. Let $X = \mathbf{X}_1 \times \cdots \times \mathbf{X}_n$ and $\mathscr{A} = \mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_n$. Let $Y : \Omega \to \mathbb{R}$. Then the following are equivalent:

(a) Y is $\sigma(X_1, \ldots, X_n)$-measurable.

(b) There is an $\mathscr{A}$-measurable function $f : \mathbf{X} \to \mathbb{R}$ such that $Y = f(X_1, \ldots, X_n)$.

Proof. Define $X : \Omega \to \mathbf{X}$ be $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$. Then $\sigma(X) = \sigma(X_1, \ldots, X_n)$. ($\sigma(X) = \{X^{-1}[A] : A \in \mathscr{A}\} = \sigma(X_j^{-1}[A] : j \in \{1, \ldots, n\} \text{ and } A \in \mathscr{A}_j)$. $\mathscr{A} = \bigotimes_{j=1}^n \mathscr{A}_j = \sigma(A_1 \times \cdots \times A_n : A_j \in \mathscr{A}_j$ for $j \in \{1, \ldots, n\}) = \bigcap_{j=1}^n \pi_j^{-1}[A_j]$.) □

Example. Let (X_n) be a random walk in $\mathbb{R}$. Let $(\mathscr{F}_n)$ be the natural filtration of (S_n). Then $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and for $n = 1, 2, 3, \ldots$, the following are equivalent for $Y : \Omega \to \mathbb{R}$:

(a) Y is $\mathscr{F}_n$-measurable.

(b) $Y = f(S_0, \ldots, S_n)$ for some Borel function $f : \mathbb{R}^n \to \mathbb{R}$.

(c) $Y = g(X_1, \ldots, X_n)$ for some Borel function $g : \mathbb{R}^n \to \mathbb{R}$, where $X_j = S_j - S_{j-1}$ for $j = 1, 2, 3, \ldots$

04/09/21

Definition. Let $(\mathscr{F}_n)$ be a filtration. To say T is a stopping time (with respect to $\mathscr{F}_n$) means $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ and for each n, $\{T \leq n\} \in \mathscr{F}_n$.

Remark. For discrete time processes, this is equivalent to the event $\{T = n\} \in \mathscr{F}_n$ for each n. The above definition generalizes to continuous time processes.

Example. 1. $T = \inf\{n : A_n \text{ occurs}\}$ is a stopping time of $A_n \in \mathscr{F}_n$ for each n, because $\{T \leq n\} = \{A_m$ occurs for at least one $m \leq n\} = A_0 \cup \cdots \cup A_n$.

2. Suppose (S_n) is adapted to $(\mathscr{F}_n)$ (that is, S_n is $\mathscr{F}_n$-measurable for every n). Let B be a measurable set in the space where the S_n's take values.

   (a) Let $T_1 = \inf\{n : S_n \in B\}$; $\{S_n = B\} = \{A_n \in \mathscr{F}_n\}$. Then T_1 is a stopping time.

   (b) Let T be any stopping time. Let $T_2 = \inf\{n \geq T : S_n \in B\}$. Then $T_2(\omega) = \inf\{n : \omega \in A_n\}$ where $A_n = \{n \geq T\} \cap \{S_n \in B\} \in \mathscr{F}_n$. Then T_2 is a stopping time.

   (c) Let $T_3 = \inf\{n > T : S_n \in B\}$. $T_3(\omega) = \inf\{n : \omega \in A_n\}$, where $A_n = \{n > T\} \cap \{S_n \in B\}$. $\{n > T\} = \varnothing$ if $n = 0$, $\{T \leq n-1\}$ if $n \geq 1$, so $\{n > T\} \in \mathscr{F}_{n-1} \subseteq \mathscr{F}_n$. $\{S_n \in B\} \in \mathscr{F}_n$. Thus T_3 is a stopping time.

Definition. Let $(\mathscr{F}_n)$ be a filtration. To say that (S_n) is a random walk (in $\mathbb{R}$) with respect to $\mathscr{F}_n$ means (S_n) is a sequence of real random variables (with n starting from 0) such that $S_0 = 0$, $\text{law}(S_{n+1} - S_n) = \text{law}(S_1)$, S_n is $\mathscr{F}_n$-measurable, and $S_{n+1} - S_n$ is independent of $\mathscr{F}_n$. for $n = 0, 1, 2, \ldots$

Remark. 1. If (S_n) is a random walk with respect to $\mathscr{F}_n$, then (S_n) is a random walk in the previous sense because if $A_1, \ldots, A_{n+1} \in \text{Borel}(\mathbb{R})$, then

$$\begin{aligned} &P(\underbrace{S_1 - S_0 \in A_1, \ldots, S_n - S_{n-1} \in A_n}_{\in \mathscr{F}_n}, S_{n+1} - S_n \in A_{n+1}) \\ &= P(S_1 - S_0 \in A_1, \ldots, S_n - S_{n-1} \in A_n) P(S_{n+1} - S_n \in A_{n+1}) \end{aligned}$$

etc. (proceed by induction).

2. (S_n) is a random walk with respect to its natural filtration if and only if (S_n) is a random walk in the previous sense.

Theorem 107 (Wald's Equation)**.** Let (S_n) be a real random walk with respect to a filtration $(\mathscr{F}_n)$. Let T be a stopping time with respect to $(\mathscr{F}_n)$. Let $X_n = S_n - S_{n-1}$. Suppose X_1 is integrable (so all X_n are integrable). Suppose also that $E(T) < \infty$. Then S_T is integrable and $E(S_T) = E(X_1)E(T)$.

Proof. Case 1. Suppose $X_1 \geq 0$ almost surely (so all $X_n \geq 0$ almost surely). Then $S_T(\omega) = S_{T(\omega)}(\omega) = \sum_{n \geq 1} X_n(\omega) 1_{\{n \leq T\}}(\omega)$. More briefly, $S_T = \sum_{n \geq 1} X_n 1_{\{n \leq T\}}$ (even on $\{T = \infty\}$). By Beppo-Levi, $E(S_T) = \sum_{n \geq 1} E(X_n 1_{\{n \leq T\}})$. For $n \geq 1$, $\{n \leq T\}^c = \{T < n\} = \{T \leq n-1\}$, so $\{n \leq T\} = \{T \leq n-1\}^c \in \mathscr{F}_{n-1}$, so X_n and $1_{\{n \leq T\}}$ are independent, so $E(X_n 1_{\{n \leq T\}}) = E(X_n)E(1_{\{n \leq T\}}) = E(X_1)E(1_{\{n \leq T\}})$, so

$$E(S_T) = E(X_1) \sum_{n \geq 1} E(1_{\{n \leq T\}}) = E(X_1) E\left(\sum_{n \geq 1} 1_{\{n \leq T\}}\right) = E(X_1)E(T).$$

(Even if $E(X_1) = \infty$ or $E(T) = \infty$, when each $X_n \geq 0$.)

General Case. Let $S'_n = X_1^+ + \cdots + X_n^+$ and $S''_n = X_1^- + \cdots + X_n^-$. By Case 1, $E(S'_T) = E(X_1^+)E(T)$ and $E(S''_T) = E(X_1^-)E(T)$ (even if $E(T) = \infty$). Now assume X_1 is integrable and $E(T) < \infty$. Then $T < \infty$ almost surely and for all $\omega \in \{T < \infty\}$, $S_T = S'_T - S''_T$, so S_T is integrable and $E(S_T) = E(S'_T) - E(S''_T) = E(X_1^+)E(T) - E(X_1^-)E(T) = E(X_1)E(T)$. □

Example. Let (S_n) be a symmetric simple random walk in $\mathbb{R}$. $(P(S_1 = 1) = 1/2 = P(S_1 = -1).)$ Let $a, b \in \mathbb{Z}$ with $a < 0 < b$. Let $T = \inf\{n : S_n \notin (a, b)\}$. (S_n) is nondegenerate, so $E(T) < \infty$. Thus $P(T < \infty) = 1$. For almost every $\omega \in \{T < \infty\}$, we have $S_T(\omega) \in \{a, b\}$. Hence $P(S_T = a$ or $S_T = b) = 1$. Let $\alpha = P(S_T = a)$ and $\beta = P(S_T = b)$. Then $\alpha + \beta = 1$. By Wald's equation, $E(S_T) = E(X_1)E(T) = 0 \cdot E(T) = 0$. That is, $\alpha a + \beta b = 0$ so $\alpha = b/(b-a)$ and $\beta = -a/(b-a)$. $P(S_T = a) = b/(b-a)$, $P(S_T = b) = -a/(b-a)$.

Theorem (Wald's Second Equation (Telegraphic Statement)**.** If $E(X_1) = 0$ and $E(X_1^2) < \infty$ and $E(T) < \infty$, then $E(S_T^2) = E(X_1^2)E(T)$.

04/12/21

Continuation of the example on symmetric simple random walk. Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed with $P(X_j = 1) = 1/2 = P(X_j = -1)$. Let $S_n = X_1 + \cdots + X_n$ ($S_0 = 0$). For $a, b, \in \mathbb{Z}$ with $a < 0 < b$, $T_{a,b} = \inf\{n : S_n \notin (a, b)\}$. Then $E(T_{a,b}) < \infty$, so $P(T_{a,b} < \infty) = 1$. $P(S_{T_{a,b}} \in \{a, b\}) = 1$, $E(S_{T_{a,b}}) = 0$ (by Wald's equation). $P(S_{T_{a,b}} = a) = b/(b-a)$, $P(S_{T_{a,b}} = b) = -a/(b-a)$.

Remark. A clarification. Let $\Omega_1 = \bigcap_{j\in\mathbb{N}}\{X_j = 1 \text{ or } -1\}$. Then $P(\Omega_1) = 1$. For each $\omega \in \Omega_1 \cap \{T_{a,b} < \infty\}$, we have $S_{T_{a,b}}(\omega) = a$ or b.

For $c \in \mathbb{Z}$, let $T_c = \inf\{n : S_n = c\}$. On Ω_1, $T_{a,b} = T_a \wedge T_b$ (recall that we use $\wedge$ to take the minimum). $S_{T_{a,b}} = b$ if and only if $T_a < T_b$, and $S_{T_{a,b}} = b$ in and only if $T_b < T_a$. Thus $P(T_a < T_b) = b/(b-a)$ and $P(T_a < T_b) = -a/(b-a)$. Hold b fixed and let $a \to -\infty$. Then $P(T_b < T_a) = -a/(b-a) \to 1$, $\{T_b < T_a\} \subseteq \{T_b < \infty\}$. For each negative integer a, $-a/(b-a) \le P(T_b < \infty) \le 1$. Letting $a \to -\infty$, we see that $P(T_b < \infty) = 1$. This holds for each integer $b > 0$.

$E(T_b) = \infty$, for if $E(T_b)$ were finite, then by Wald, we would have $0 = E(X_1)E(T_b) - E(S_{T_b}) = E(b) = b$, a contradiction.

Let $b = 1$. Let $Z = \inf\{S_n : n \le T_1\}$. For each integer $a < 0$, $P(Z \le a) = P(T_a < T_1) = 1/(1-a)$. For each integer $z \ge 1$, $P(-Z \ge z) = 1/(1+z)$. Now, $\sum_{z\in\mathbb{N}} 1_{\{-Z\ge z\}} = -Z$, so $E(-Z) = \sum_{z\in\mathbb{N}} P(-Z \ge z) = \sum_{z\in\mathbb{N}} 1/(1+z) = \infty$.

Lemma 108 (Generalized Pythagorean Equation)**.** Let (Y_n) be an orthogonal sequence in L_2. Suppose $\sum_{n=1}^\infty \|Y_n\|_2^2 < \infty$ and $\sum_{n=1}^\infty Y_n$ converges almost surely to Y. Then $Y \in L^2$, $\sum_{n=1}^\infty Y_n$ converges to Y in L^2, and $\sum_{n=1}^\infty \|Y_n\|_2^2 = \|Y\|_2^2$.

Proof. Let $S_n = \sum_{k=1}^n Y_k$ and let $S_0 = 0$. For all $m < n$, we have

$$\int |S_n - S_m|^2\, dP = \|S_n - S_m\|_2^2 = \left\| \sum_{k=m+1}^n Y_k \right\|_2^2 = \sum_{k=m+1}^n \|Y_k\|_2^2$$

so by Fatou,

$$\int |Y - S_m|^2\, dP \le \liminf_{n\to\infty} \int |S_n - S_m|^2\, dP = \sum_{k=m+1}^\infty \|Y_k\|_2^2 \xrightarrow[m\to\infty]{} 0. \qquad (*)$$

Also by Fatou,

$$\int |Y|^2\, dP = \int \lim_{n\to\infty} \left|\sum_{k=1}^\infty Y_k\right|^2 dP = \liminf_{n\to\infty} \int \left|\sum_{k=1}^\infty Y_k\right|^2 dP = \liminf_{n\to\infty} \sum_{k=1}^n \|Y_k\|_2^2 = \sum_{k=1}^n \|Y_k\|_2^2 < \infty$$

so $Y \in L^2$. From $(*)$, we have $\|Y - S_n\|_2^2 \to 0$. Then $S_n \to Y$ in L^2. □

Remark (Optional)**.** Furthermore, for each $A \subseteq \mathbb{N}$, $\left\|Y - \sum_{k\in A} Y_k\right\|_2^2 = \sum_{k\in\mathbb{N}\setminus A} \|Y_k\|_2^2$.

Theorem 109 (Wald's Second Equation)**.** Let (S_n) be a random walk in $\mathbb{R}$ with respect to a filtration $(\mathscr{F}_n)$, let $X_j = S_j - S_{j-1}$ for $j = 1, 2, 3, \ldots$, and suppose $E(X_1^2) < \infty$ and $E(X_1) = 0$. Let T be a stopping time such that $E(T) < \infty$. Then $E(S_T^2) = E(X_1^2)E(T)$.

Proof. Let $Y_n = X_n 1_{T\ge n}$. Then $\sum_{n=1}^\infty Y_n$ converges pointwise to S_T on $\{T < \infty\}$ and hence almost surely. $\|Y_n\|_2^2 = E[(S_n 1_{\{T\ge n\}})^2] = E(X_n^2 1_{\{T\ge n\}}$. Now $\{T \ge n\} = \{T \le n-1\}^2 \in \mathscr{F}_{n-1}$ so X_n and $1_{\{T\ge n\}}$ are independent. Hence $\|Y_n\|_2^2 = E(X_n^2)E(1_{\{T\ge n\}}) = E(X_1^2)P(T \ge n)$. Hence $\sum_{n=1}^\infty \|Y - n\|_2^2 = E(X_1^2)\sum_{n=1}^\infty P(T \ge n) = E(X_1^2)E(T)$ (since T takes values in $\{0, 1, 2, \ldots, \infty\}$).

For $m < n$, $E(X_m 1_{\{T\ge m\}} X_n 1_{\{T\ge n\}}) = E(X_m X_n 1_{\{T\ge n\}})$, and $X_n 1_\{T \ge n\}$ is measurable with respect to $\mathscr{F}_{n-1}$, so X_n and $X_m 1_{\{T\ge n\}}$ are independent, so $E(X_m X_n 1_{\{T\ge n\}}) = E(X_m 1_{\{T\ge n\}})E(X_n) = 0$. Thus (Y_n) is an orthogonal sequence in L^2. Hence by the lemma, $E(S_T^2) = \|S_T\|_2^2 = \sum_{n=1}^\infty \|Y_n\|_2^2 = E(X_1^2)E(T)$. □

Example. Again let (S_n) be a symmetric simple random walk in $\mathbb{R}$. Let $a, b \in \mathbb{Z}$ with $a < 0 < b$. Let $T = \inf\{n : S_n \notin (a,b)\}$. Then $E(T) < \infty$. Also $E(X_j) = 0$, where $X_j = S_j - S_{j-1}$. Thus by Wald's second equation, $E(S_T^2) = E(X_1^2)E(T) = E(T)$. But $P(S_T \in \{a,b\}) = 1$, so $E(S_T^2) = a^2 P(S_T = a) + b^2 P(S_T = b) = a^1 b/(b-a) - b^2 a/(b-a) = ab(a-b)/(b-a) = -ab$. Thus $E(T) = -ab$. If $a = -b$, then $E(T) = b^2$.

Martingales.

Definition. Let $(\mathscr{F}_n)$ be a filtration. To say (M_n) is a martingale with respect to $(\mathscr{F}_n)$ means

(a) For each n, $M_n \in L^1$.

(b) (M_n) is $(\mathscr{F}_n)$-adapted; that is, for each n, M_n is $\mathscr{F}_n$-measurable.

(c) For each n, for each $A \in \mathscr{F}_n$, $E(M_{n+1}; A) = E(M_n; A)$.

Remark. A more advanced statement of (c) is $E(M_{n+1} \mid \mathscr{F}_n) = E(M_n \mid \mathscr{F}_n)$, which is necessary for continuous time processes.

If $A \in \mathscr{F}_n$, then $E(M_{n+2}; A) = E(M_{n+1}; A) = E(M_n; A)$.

Example. Let $X_1, X_2, X_3, \dots \in L^1$ be independent. Let $S_n = \sum_{j \leq n} X_j$ and $\mathscr{F}_n = \sigma(X_j : j \leq n)$ for $n = 0, 1, 2, \dots$ Note that $S_n = 0$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $\mathscr{F}_n = \sigma(S_m : m \leq n)$. (S_n) is an $(\mathscr{F}_n)$-martingale if and only if $E(X_j) = 0$ for $j = 1, 2, 3, \dots$

4/14/2021

Reminder. Given suitable (in particular finite or σ-finite) measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$, the product measure $\mu \otimes \nu : \mathscr{A} \otimes \mathscr{B} \to [0, \infty)$ is given by

$$(\mu \otimes \nu)(C) = \int_X \int_Y 1_C(x,y)\, d\nu(y)\, d\mu(x) = \int_Y \int_X 1_C(x,y)\, d\mu(x)\, d\nu(y)$$

and in particular for $A \in \mathscr{A}$ and $B \in \mathscr{B}$, $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$. (For s-finite measures there is a family of product measures satisfying the multiplication.)

Remark (Hint for Extra Problem 37)**.**

$$E(e^{uX}) = E\left[\sum_{n=0}^{\infty} \frac{(uX)^n}{n!}\right] \overset{\text{Why?}}{=} \sum_{n=0}^{\infty} E\left[\frac{(uX)^n}{n!}\right] = \sum_{n=0} c_n u^n$$

where $c_n = E(X^n)/n!$. By the change of variables $u = s + it$, $s, t \in \mathbb{R}$, $|e^{uX}| = |e^{sX}||e^{itX}| = |e^{sX}|$. Since $v \leq e^v$ for $v \in [0, \infty)$, $|v| \leq e^{|v|}$ for all $v \in \mathbb{C}$.

Remark (Hint for Extra Problem 38)**.** Let $\text{law}(X) = N(0,1)$.

(a) We have

$$E(e^{sX}) = \int_{\mathbb{R}} e^{sx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

Complete the square, so

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x^2 - 2sx)/2}\, dx \\ &= \cdots = e^{s^2/2}. \end{aligned}$$

(b) $E(e^{itX}) = ?$

Definition. Let $(\mathscr{F}_n)_{n=0,1,2,\dots}$ be a filtration. To say $(H_n)_{n=1,2,3,\dots}$ is predictable means H_n is $\mathscr{F}_{n-1}$-measurable for $n = 1, 2, 3, \dots$.

For real valued processes, $(H_n)_{n\geq 1}$ and $(X_n)_{n \geq 0}$, we define a process $H \cdot X$ by $(H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1})$ for $n \geq 1$ and $(H \cdot X)_0 = 0$. Such a predictable process is sometimes called a gambling strategy.

Example. Let X_n = number of heads − number of tails in n tosses. Let H_n = the amount bet on heads on the n-th toss. Then $(H \cdot X)_n$ is the total amount won in n tosses.

Theorem 110. Let $(M_n)_{n\geq 0}$ be a martingale. Let $(H_n)_{n\geq 0}$ be a predictable sequence of bounded real random variables. Then $H \cdot M$ is a martingale.

Proof. $(H \cdot M)_n = \sum_{k=1}^n H_k(M_k - M_{k-1}$ is $\mathscr{F}_n$-measurable. Thus $H \cdot M$ is adapted.

$$E(|(H \cdot M)_n|) \leq \sum_{k=1}^n E(|H_k||M_k - M_{k-1}|) < \infty$$

since $|H_k|$ is bounded and $|M_k - M_{k-1}|$ is integrable, so $(H \cdot M)_n \in L^1$. Let $A \in \mathscr{F}_n$, We wish to show that $E((H \cdot M)_{n+1} 1_A) = E((H \cdot M)_n 1_A)$. Now

$$(H \cdot M)_{n+1} = \sum_{k=1}^{n+1} H_k(M_k - M_{k+1}) = (H \cdot M)_n + H_{n+1}(M_{n+1} - M_n).$$

Thus we want to show that $E[H_{n+1}(M_{n+1} - M_n)1_A] = 0$.

Let $Y = H_{n+1}1_A$. Y is $\mathscr{F}_n$-measurable since (H_n) is predictable. Y is bounded. Let $Z = M_{n+1} - M_n$, and notice $Z \in L^1$. For each $B \in \mathscr{F}_n$, $E(Z1_B) = 0$, because (M_n) is a martingale. Now let $\varphi : \Omega \to \mathbb{R}$ be $\mathscr{F}_n$-simple. Then $E(Z\varphi) = 0$. Y is $\mathscr{F}_n$-measurable, so there is a sequence (φ_j) of $\mathscr{F}_n$-simple functions $\varphi_j : \Omega \to \mathbb{R}$ such that for each $\omega \in \Omega$, $\varphi_j(\omega) \to Y(\omega)$ as $j \to \infty$ and $|\varphi_j(\omega)| \leq |Y(\omega)|$ for every j. Then $|Z\varphi_j| \leq |Z||Y|$, $E(|Z||Y|) < \infty$, and $Z\varphi_j \to ZY$ pointwise on Ω, so $E(Z\varphi_j) \to E(ZY)$ by the (Lebesgue's) dominated convergence theorem. Hence $E(ZY) = 0$, as desired. □

Theorem 111 (Optional (Stopping) Stopping Theorem)**.** Let (M_n) be a martingale and let T be a stopping (optional) time. Then $(M_{T\wedge n})$ is a martingale. ($(T \wedge n) = \min\{T(\omega), n\}$.)

Proof. $M_{T\wedge n} = M_0 + (H \cdot M)_n$, where $H_n = 1_{\{M\leq T\}}$. H is predictable and bounded ($0 \leq H_n \leq 1$). □

Corollary 112. Let (M_n) be an adapted sequence of integrable real random variables. Then the following are equivalent:

(a) (M_n) is a martingale.

(b) For each bounded stopping time T, $E(M_T) = E(M_0)$.

Proof. (a) $\Longrightarrow$ (b). Consider $n = \max R$ in the optional stopping theorem.

(b) $\Longrightarrow$ (a). Think about it. □

Example (Asymmetric random walk in $\mathbb{R}$)**.** Let $1/2 < p < 1$ and $q = 1 - p$. Let $X_1, X_2, X_3, \ldots$ be independent with $P(X_j = 1) = p$ and $P(X_j = -1) = q$. Let $S_n = X_1 + \cdots + X_n$ (and $S_0 = 0$). Let $\mathscr{F}_n = \sigma(X_1, \ldots, X_n) = \sigma(S_1, \ldots, S_n)$.

(a) Define φ on $\mathbb{R}$ by $\varphi(x) = (q/p)^x$. Then $(\varphi(S_n))$ is a martingale.

Proof. Let $A \in \mathscr{F}_n$. $\varphi(S_n) = \varphi(X_{n+1})\varphi(S_n)$. $\varphi(X_{n+1})$ and $\varphi(S_n)1_A$ are independent. Hence

$$E(\varphi(S_{n+1})1_A) = E(\varphi(X_{n+1})\varphi(S_n)1_A) = E(\varphi(X_{n+1}))E(\varphi(S_n)1_A) = E(\varphi(S_n)1_A)$$

because $E(\varphi(X_{n+1})) = \varphi(1)p + \varphi(-1)q = qp/p + pq/q = q + p = 1$. □

(b) For $x \in \mathbb{Z}$, let $T_x = \inf\{n : S_n = x\}$. Then for $a, b \in \mathbb{Z}$, with $a < 0 < b$,

$$P(T_a < T_b) = \frac{\varphi(b) - \varphi(0)}{\varphi(b) - \varphi(a)}.$$

Proof. Next time.

4/16/21

Asymmetric simple random walks on $\mathbb{Z}$. Let $1/2 < p < 1$ and let $q = 1 - p$. Let $X_1, X_2, X_3, \ldots$ be independent $\{-1, 1\}$-valued with $P(X_j = 1) = p$ and $P(X_j = -1) = q$ for each j. Let $S_0 = 0$, and for $n = 1, 2, 3, \ldots$, $S_n = X_1 + \cdots + X_n$. Let $M_n = (q/p)^{S_n}$ ($= \varphi(S_n)$ from last time) for each n. Then

(a) (M_n) is a martingale.

Proof. Last time. $\square$

(b) For each $x \in \mathbb{Z}$, let $T_x = \inf\{n : S_n = x\}$. Let $a, b \in \mathbb{Z}$ with $a < 0 < b$. Then

$$P(T_a < T_b) = \frac{r^b - r^0}{r^b - r^a}$$

where $r = q/p$.

Proof. As $n \to \infty$, $S_n/n \xrightarrow{\text{a.s.}} E(X_1) = p - q > 0$ (by the Bernoulli strong law of large numbers, which follows from Bernstein's inequality) so $S_n \to \infty$ almost surely. Hence $P(T_b < \infty) = 1$.

Remark. Note that $P(T_b < \infty) = 1$ for $0 < b \in \mathbb{Z}$.

Let $N = T_a \wedge T_b$. Then $N < \infty$ almost surely. N is a stopping time, so by the optional stopping theorem, $M_{N \wedge n}$ is a martingale. Since $N < \infty$ almost surely, $M_{N\wedge n} \to M_N$ almost surely as $n \to \infty$. (For each $\omega \in \{N < \infty\}$, for each $n \geq N(\omega)$, $M_{N\wedge n}(\omega) = M_N(\omega)$.) For each n, $a \leq S_{N\wedge n} \leq b$ so $(q/p)^a \geq (q/p)^{M_{N\wedge n}} \geq (q/p)^b$. Then by Lebesgue's dominated convergence theorem, $E(M_{N\wedge n}) \to E(M_N)$. But $E(M_{N\wedge n}) = 1$, so $E(M_N) = 1$.

Now $S_N = a$ on $\{T_a < T_b\}$, $S_N - b$ on $\{T_b < T_a\}$, and $P(T_a = T_b) = 0$ because if $T_a(\omega) = T_b(\omega)$, then both are ∞, and $P(T_b = \infty) = 0$. Hence $1 = E[M_N] = r^a P(T_a < T_b) + r^b P(T_b < T_a)$ and $1 = P(T_a < T_b) + P(T_b < T_a)$. Thus $1 = r^a P(T_a < T_b) + r^b[1 - P(T_a < T_b)]$ so $1 - r^b = (r^a - r^b)P(T_a < T_b)$ so

$$P(T_a < T_b) = \frac{1 - r^b}{r^a - r^b} = \frac{r^b - r^0}{r^b - r^a}.$$

$\square$

(c) Let $0 > a \in \mathbb{Z}$. Then $P(T_a < \infty) = r^{-a}$ $(r = q/p)$.

Proof. $\{T_a < \infty\} =\uparrow \bigcup_{b\in\mathbb{N}}\{T_a < T_b\}$. (If $T_a(\omega) < \infty$ and $b \geq T_a(\omega)$, then $T_a(\omega) \leq T_b(\omega)$.)

$$P(T_a < \infty) = \lim_{b\to\infty} P(T_a < T_b) = \lim_{b\to\infty} \frac{r^b - r^0}{r^b - r^a} = \frac{0 - 1}{0 - r^a} = r^{-a}.$$

$\square$

Remark. Let $0 > a \in \mathbb{Z}$. Then $\{T_a < \infty\} = \{\inf_n S_n \leq a\}$.

Remark. Let $Y = -\inf_n S_n$. Then $E(Y) < \infty$.

Proof. $Y : \Omega \to \{0, 1, 2, \ldots, \infty\}$, so $Y = \sum_{k\in\mathbb{N}} 1_{Y \geq k}$ and

$$E(Y) = \sum_{k\in\mathbb{N}} P(Y \geq k) = \sum_{k\in\mathbb{N}} r^k = r(1 + r + r^2 + \cdots)r\frac{1}{1-r} = \frac{q}{p}\frac{1}{1 - q/p} = \frac{q}{p-q}.$$

$\square$

(d) Let $0 < b \in \mathbb{Z}$. Then $E(T_b) = b/(p - q) = b/E(X_1)$.

Proof. Let $W_n = S_n - (p-q)n$ for $n = 0, 1, 2, \ldots$ Then $W_n = \sum_{m\leq n} X'_m$, where $X'_m = X_m - (p - q)$. Now $X'_1, X'_2, X'_3, \ldots$ are independent and identically distributed, so (W_n) is a random walk. $E(X'_n) = 0$ for $m = 1, 2, 3$ hence (W_n) is a martingale. Thus for $n = 1, 2, 3, \ldots$, $E(W_{T_b\wedge n}) = E(W_0) = 0$. In other words, $(p - q)E(T_b \wedge n) = E(S_{T_b\wedge n})$. Now $\inf_k S_k \leq S_{T_b\wedge n} \leq b$ for all n. Hence $E(\sup_n |S_{T_b\wedge n}|) < \infty$. Since $T_b < \infty$ almost surely, $S_{T_b\wedge n} \to b$ almost surely as $n \to \infty$. Hence be Lebesgue's dominated convergence theorem, $E(S_{T_b\wedge n}) \to E(b) = b$ as $n \to \infty$. Thus $(p - q)E(T_b \wedge n) \to b$. But $T_b \wedge n \uparrow T_b$, so by the monotone convergence theorem, $E(T_b \wedge n) \to E(T_b)$. Hence $(p - q)E(T_b) = b$, so $E(T_b) = b/(p - q)$. $\square$

Arcsine laws for symmetric simple random walks.

Definition. A path is a finite sequence $(k_0, x_0), \ldots (k_n, x_n)$ in $\mathbb{Z}\times\mathbb{Z}$ such that $k_j = 1+k_{j+1}$ and $|x_j - x_{j-1}| = 1$ for $j = 1, \ldots, n$. (Usually k_0 would be ≥ 0.) Such a path is said to be from (k_0, x_0) to (k_n, x_n) and is said to be of length n. Thus the length of the path is the number of segments $(k_{j-1}, x_{j-1}) \to (k_j, x_j)$.

The number of positive steps in such a path is $a = |\{j \in \{1, \ldots, n\} : x_j - x_{j-1} = 1\}|$.

The number of negative steps in such a path is $b = |\{j \in \{1, \ldots, n\} : x_j - x_{j-1} = -1\}|$.

Clearly $a + b = n$ and $a - b = x_n - x_0$, so $a = (n + (x_n - x_0))/2$ and $b = (n - (x_n - x_)))/2$. Since a and b are integers ≥ 0, $-n \leq x_n - x_0 \leq n$ and the integers n and $x_n - x_0$ are either both even or both odd.

Conversely, given integers k_0, n, x_0, x_n such that $-n \leq x_n - x_0 \leq n$ (this implies that $n \geq 0$) and such that n and $x_n - x_0$ are both even or both odd, hen the set of paths from (k_0, x_0) to $(k_0 + n, x_n)$ is in one-to-one correspondence with the set of a-element subsets of $\{1, \ldots, n\}$ ($\varnothing$ if $n = 0$), where $a - (n + (x_n - x_0))/2$, so the number of such paths is $\binom{n}{a}$, which we'll denote bu $N_{n,x_n - x_0}$. Of course this is the same as the number of paths from $(0, 0)$ to (n, x), where $x = x_n - x_0$.

4/19/21

Theorem 113 (The Reflection Principle (Desini André, 1887))**.** Let x and y be integers > 0. Then the number of paths from $(0, x)$ to (n, y) that 0 at some time is equal to the number of possible paths from $(0, -x)$ to (n, y).

Proof. Let $(0, s_0), (1, s_i), \ldots, (n, s_n)$ be a path from $(0, x)$ to (n, y) such that $s_k = 0$ for some $k \in \{0, \ldots, n\}$. Let K be the least such k. Let

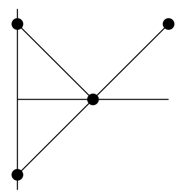
$$s'_k = \begin{cases} -s_k & \text{if } k < K \\ s_k & \text{if } k \geq K \end{cases}$$

for $k = 0, 1, 2 \ldots, n$. Then $(0, s'_0), (1, s'_1), \ldots, (n, s'_n)$ is a path from $(0, -x)$ to (n, y), because each increment $s'_k - s'_{k-1}$, being either $s_k - s_{k-1}$ or $-(s_k - s_{k-1})$, is ± 1. Also, K is the least k such that $s'_k = 0$.

Conversely, let $(0, t_0), (1, t_1), \ldots, (n, t_n)$ be a path from $(0, -x)$ to (n, y). Then t_k must be 0 for some $k \in \{0, 1, \ldots, n\}$. (Actually $k \in \{0, 1, \ldots n - 1\}$.) Let K be the least such k. Let

$$t'_k = \begin{cases} -t_k \text{if } k < K \\ t_k \text{if } k \geq K \end{cases}$$

for $k = 0, 1, \ldots, n$. Then $(0, t'_0), (1, t'_1), \ldots, (n, t'_n)$ is a path from $(0, x)$ to (n, y) and K is the least k such that $t'_k = 0$. The correspondences $s \mapsto s'$ and $t \mapsto t'$ are inverses of one another. $\square$



Let (S_n) be a symmetric simple random walk on $\mathbb{Z}$.

Theorem 114. Let $n \in \mathbb{N}$. Then $P(S_1 \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2n} = 0)$.

Proof. By symmetry, $P(S_1 > 0, \ldots, S_{2n} > 0) = P(S_1 < 0, \ldots, S_{2n} < 0)$. Now (S_n) cannot change sign without passing through 0. Hence $P(S_1 > 0, \ldots, S_{2n} > 0) + P(S_1 < 0, \ldots, S_{2n} < 0) = P(S_1 \neq 0, \ldots, S_{2n} \neq 0)$. Now $P(S_1 > 0, \ldots, S_{2n} > 0) = \sum_{r=1}^{n} P(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r)$. The number of paths from $(0, 0)$ to $(2n, 2r)$ that are never 0 at times ≥ 1 is equal to the number of paths from $(1, 1)$ to $(2n, 2_r)$ that are never 0, and this is equal to $K_r - J_r$ where k_r is the number of paths from $(1, 1)$ to $(2n, 2r)$ and J_r is the number of such paths that are 0 at some time. By the reflection principle, J_r is equal to the number of paths from $(1, -1)$ to $(2n, 2r)$. Thus J_r is the number of paths from $(1, 1)$ to $(2n, 2(r + 1))$. In other words, $J_r = K_{r+1}$. Thus the number of paths from $(0,0)$ to $(2n, 2r)$ that are never 0 at times ≥ 1 is equal to $K_r - K_{r+1}$. Hence $P(S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r) = (K_r - K_{r+1})/2^{2n}$. So

$$P(S_1 > 0, \ldots, S_{2n} > 0) = \sum_{r=1}^{n} \frac{K_r - K_{r+1}}{2^{2n}} = \frac{K_1 - K_{n+1}}{2^{2n}}.$$

But $K_{n+1} = 0$, because $2(n+1) - 1 > 2n - 1$. Also K_1 is the number of paths from $(1,1)$ to $(2n, 2)$ and this is equal to the number of paths from $(0,0)$ to $(2n-1, 1)$, which is $\binom{2n-1}{n}$, since such paths have n steps of $+1$ and $n-1$ steps of -1. Thus

$$P(S_1 > 0, \ldots, S_{2n>0}) = \frac{1}{2}\binom{2n-1}{n}\frac{1}{2^{2n-1}} = \frac{1}{2}P(S_{2n-1} = 1).$$

To finish, observe that $P(S_{2n} = 0) = P(S_{2n-1} = -1, X_{2n} = 1) + P(S_{2n-1} = 1, X_{2n} = -1) = (1/2)P(S_{2n-1} = -1) + (1/2)P(S_{2n-1} = 1)$. □

Corollary 115. Let R be the first time (S_n) returns to 0. In other words, let $R = \inf\{m \geq 1 : S_m = 0\}$. Then $P(R > 2n) \sim 1/\sqrt{\pi n}$ as $n \to \infty$.

Proof.

$$P(R = 2n) = P(S_1 \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2n} = 0)$$

by the theorem

$$= \binom{2n}{n}\frac{1}{2^{2n}} = \frac{(2n)!}{(n!)^2}\frac{1}{2^{2n}} \sim \frac{\sqrt{2\pi 2n}(2n)^{2n}e^{-2n}}{(\sqrt{2\pi n}n^n e^{-n})^2}\frac{1}{2^{2n}}$$

as $n \to \infty$, by Stirling's formula

$$= \frac{\sqrt{2\pi 2n}}{2\pi n}\frac{n^{2n}}{n^{2n}}\frac{2^{2n}}{2^{2n}}\frac{e^{-2n}}{e^{-2n}} = \frac{1}{\sqrt{\pi n}}.$$

□

Remark. In particular, $P(R > 2n) \to 0$ as $n \to \infty$, so $P(R = \infty) = 0$. But in fact,

$$E(|\{n : S_{2n} = 0\}|) = E\left(\sum_{n=0}^{\infty} 1_{\{0\}}(S_{2n}\right) = \sum_{n=0}^{\infty} E(1_{\{0\}}(S_{2n})) = \sum_{n=1}^{\infty} P(S_{2n} = 0) = \infty$$

since $P(S_{2n} = 0) \sim 1/\sqrt{\pi n}$ as $n \to \infty$.

Notation. Let $L_{2n} = \max\{m \leq 2n : S_m = 0\}$ (this set is nonempty, because 0 belongs to it). L_{2n} is the last time the players are all square, if they play for $2n$ tosses.

Proposition 116. Let $k \in \{0, \ldots, n\}$. Then $P(L_{2n} = 2k) = P(S_{2k} = 0) = P(S_{2n-2k} = 0)$.

Proof.

$$P(L_{2n} = 2k) = P(S_{2k} = 0, S_{k+1} \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2k} = 0, \underbrace{S_{2k+1} - S_{2k} \neq 0, \ldots, S_{2n} - S_{2k} \neq 0}_{\text{independent of } S_{2k}})$$
$$= P(S_{2k=0})P(S_{2k+1} - S_{2k} \neq 0, \ldots, S_{2n} - S_{2k} \neq 0)$$
$$= P(S_{2k=0})P(S_1 \neq 0, \ldots, S_{2n-2k} \neq 0) = P(S_{2k} = 0)P(S_{2n-2k} = 0)$$

by the theorem. □

Remark. It follows that the distribution of L_{2n} in $\{0, 2, \ldots, 2n\}$ is symmetric about n. Thus if two people were to bet \$1 on a coin flip each day for a year, then with probability at least 1/2 one of them will be ahead from early July to the end of the year, which might cause the other to complain about his bad luck. (January - June is 181 or 182 days, July - December is 184 days).

Theorem 117 (The Arcsine Law for L_{2n})**.** Let $0 < b < 1$. Then as $n \to \infty$,

$$P\left(a < \frac{L_{2n}}{2n} < b\right) \to \int_a^b \frac{1}{\pi\sqrt{x(1-x)}}\,dx = \frac{2}{\pi}(\arcsin\sqrt{b} - \arcsin\sqrt{a}).$$

Proof. Let $K_n = ¶; : a \le (2k)/(2n) \le b\}$. As $n \to \infty$,

$$P(L_{2n} \le 2k) = P(S_{2k} = 0)P(S_{2n-2k} = 0) \sim \frac{1}{\sqrt{\pi k}}\frac{1}{\sqrt{\pi(n-k)}}$$

(This holds uniformly for $k \in K_n$, because for each k, we have $2k \ge 2na$ and $2n - 2k \ge 2n(1-b)$.) Let $x_{n,k} = k/n$.

$$= \frac{1/n}{\pi\sqrt{k/n(1-k/n)}} = \frac{1}{\pi\sqrt{x_{n,k}(1-x_{n,k})}}(x_{n,k} - x_{n,k-1}).$$

Therefore

$$P\left(a \le \frac{L_{2n}}{2n} \le b\right) \sim \sum_{k \in K_n} \frac{1}{\pi\sqrt{x_{n,k}(1-x_{n,k})}}(x_{n,k} - x_{n,k-1}) \to \int_a^b \frac{1}{\pi\sqrt{x(1-x)}}\,dx$$

Let $y = \sqrt{x}$, so $y^2 = x$, $2y\,dy = dx$; $a \xrightarrow{x} b$, $\sqrt{a} \xrightarrow{y} \sqrt{b}$.

$$= \int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{\pi}\frac{1}{y}\frac{1}{\sqrt{1-y^2}}2y\,dy = \frac{2}{\pi}\int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{1-y^2}\,dy = \frac{2}{\pi}(\arcsin\sqrt{b} - \arcsin\sqrt{a}).$$

□

1 4/21/21

On Midterm 2. Problem 1.

(a)

$$\begin{aligned} P(H = k) &= \sum_{n=k}^{6} P(H = k, D = n) = \sum_{n=k}^{6} P(H = k \mid D = n)P(D = n) \\ &= \sum_{n=k}^{6} \binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}\frac{1}{6} = \sum_{n=k}^{6}\binom{n}{k}\frac{1}{2^n}\frac{1}{6} \end{aligned}$$

(b) If 3 heads are obtained, what is the probability that the die showed n?

$$\begin{aligned} P(D = n \mid H = 3) &= \frac{P(D = n, H = 3)}{P(H = 3)} \\ &= \frac{P(H = 3 \mid D = n)P(D = n)}{P(H = 3)}. \end{aligned}$$

$P(H = 3) = \sum_{n=3}^{6}\binom{n}{3}2^{-n}/6 = \cdots = 1/6 = P(D = n)$. Hence $P(D = n \mid H = 3) = P(H = 3 \mid D = n) = \binom{n}{3}2^{-n}$.

Remark (A False Generalization of the Monotone Convergence Theorem)**.** Let $(X, \mathscr{A}, \mu)$ be a measure space. Let f be a nonnegative measurable function on X. Let (f_n) be a sequence of nonnegative measurable functions on X. Suppose that for all $x \in X$, $f_n(x) \to f(x)$ as $n \to \infty$ there is an $n_0 \in \mathbb{N}$ such that for each $n \ge n_0$, $f_n(x) \le f_{n+1}(x)$. Then $\int f_n\,d\mu \to \int f\,d\mu$.

NO!!!

Example. Let $f_n = 1_{[n,\infty)}$. Then $f_n \to 0$ pointwise on $\mathbb{R}$. Let $x \in \mathbb{R}$. Let $n_0 = \lfloor x \rfloor + 1$. Then for all $n \geq n_0$, $f_n(x) = f_{n+1}(x)$ (because both are 0).

Problem 2. Let X be a random variable and let $\varphi(u) = E(e^{iuX})$.

$$\begin{aligned}\frac{2-\varphi(u)-\varphi(-u)}{u^2} &= E\left(\frac{2-e^{iuX}-e^{-iuX}}{u^2}\right) = E\left(\frac{2-2\cos(uX)}{u^2}\right)\\ &= E\left(\frac{4\sin^2(uX/2)}{u^2}\right) = E\left[\left(\frac{\sin(uX/2)}{u/2}\right)^2\right].\end{aligned}$$

If $v \neq 0$, then $\sin(uv)/u = v\sin(uv)/(uv) \to v \cdot 1 = v$ as $u \to 0$. If $f = 0$, then $\sin(uv)/u - 0 = v \to v$ as $u \to 0$. Thus

$$\left(\frac{\sin(uX/2)}{u/2}\right)^2 \to X^2 \text{ as } u \to 0.$$

Now $|\sin\theta| \leq |\theta|$ for θ real. Thus

$$0 = \left(\frac{\sin(uX/2)}{u/2}\right)^2 \leq \left(\frac{uX/2}{u/2}\right) = X^2.$$

Hence by a corollary of Fatou's lemma,

$$E\left[\left(\frac{\sin(uX/2)}{u/2}\right)^2\right] \to E(X^2).$$

Remark. To be completely correct, one should replace u with a sequence $u_n \to 0$, but the above analysis is necessary; it is not enough, for example, to consider u along $1/n$.

Example. Let

$$f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{1}{n} \text{ where } n \in \mathbb{N} \\ 1 & \text{otherwise.} \end{cases}$$

Then $\lim_{n\to\infty} f(1/n) = 0$ but $\lim_{x\to 0} f(x)$ does not exist.

Example. Let $f = 1$ on $\mathbb{R}$. Let $f_A = 1_A$ for each finite $A \subseteq \mathbb{R}$. Then $f_A \to f$ pointwise as $A \uparrow \mathbb{R}$. But for all finite $A \subseteq \mathbb{R}$, $\int_{\mathbb{R}} f_A(x)\,dx = 0$ so $\int f_A \not\to \int f$ as $A \uparrow \mathbb{R}$.

4/23/21

Remark (On Extra Problem 47)**.** Let $S'_n = (S_n)^\circ$. Then $S'_n = S_n - (p-q)n$, so $S'_{T_b} = S_{T_b} - (p-q)T_b = b - (p-q)T_b$. Also, by Wald's second equation, $E[(S'_{T_b})^2] = E[(X_1^\circ)^2]E(T_b)$ and $E[(X_1^\circ)^2] = E(X_1^2) - E(X_1)^2 = 1-(p-q)^2$.

Note that $0 = (S_{T_b})^\circ \neq (S^\circ)_{T_b} = S'_{T_b}$.

On Handwriting Script Letters, Typeset Edition. Don't write $\mathscr{Q}$, $\mathscr{X}$, or $\mathscr{Z}$.

$\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}, \mathscr{E}, \mathscr{F}, \mathscr{G}, \mathscr{H}, \mathscr{I}, \mathscr{J}, \mathscr{K}, \mathscr{L}, \mathscr{M}, \mathscr{N}, \mathscr{O}, \mathscr{P}$, ?, $\mathscr{R}, \mathscr{S}, \mathscr{T}, \mathscr{U}, \mathscr{V}, \mathscr{W}, \mathscr{X}, \mathscr{Y}$, ??

Brownian Motion in $\mathbb{R}^d$. Let $\Omega = C([0,\infty), \mathbb{R}^d)$ (the set of continuous paths in $\mathbb{R}^d$), $B_t(\omega) = \omega(t)$ (the location of a path at t), $\mathscr{F}^\circ = \sigma(B_t : 0 \leq t < \infty)$. For each $x \in \mathbb{R}^d$, let P^x be the probability distribution of paths starting at x, $P^x(B_0 = x) = 1$. For $t > 0$,

$$P^x(B_t \in dy) = P_t(x, dy) = p_t(x,y) = \left(\frac{1}{2\pi t}\right)^{d/2} e^{-|x-y|^2/(2t)}\,dy.$$

Remark. For all $\varepsilon > 0$, $t^{-1}\sup_x P_t(x : \{y : \mathrm{dist}(x,y) \geq \varepsilon\}) \to 0$ as $t \downarrow 0$. This is why we can construct that P^x's on $C([0,\infty), \mathbb{R}^d)$. Compare with the Cauchy process (for $d = 1$)

$$Q_t(x, dy) = \frac{t/\pi}{(x-y)^2 + t^2}\,dy.$$

Joint distributions for (B_t)**.** For $0 < t_1 < t_2 < \cdots < t_n < \infty$, for $A_1, A_2, \dots, A_n \in \mathrm{Borel}(\mathbb{R}^d)$,

$$P^x(B_{T_1} \in A_1, B_{T_2} \in A_2, \dots, B_{T_n} \in A_n) = \int_{A_1} P^x(B_{t_1} \in dy) P^y(B_{t_2 - t_1} \in A_2, \dots, B_{T_n - t_{n-1}} \in A_n)$$

(this is the Markov property).

$$P^x(B_{t_1} \in dy_1, B_{t_2} \in y_2, \dots, B_{T_n} \in dy_n) = P_{t_1}(x, dy_1) P_{t_2 - t_1}(y_1, dy_2) \cdots P_{t_n - t_{n-1}}(y_{n-1}, dy_n).$$

(B_t) has stationary independent increments:

$$E^x[f_1(B_{t_1} - B_0) f_2(B_{t_2} - B_{t_1}) \cdots f_n(B_{t_n} - B_{t_{n-1}})] = E^0[f_1(B_{t_1})] E^0[f_2(B_{t_2})] \cdots E^0[f_n(B_{t_n - t_{n-1}})].$$

Stochastic Integrals. ($d = 1$ for now)

$$\int_0^1 C_t \, dB_t$$

C_t should be non-anticipating: $C_t(\omega) = f(t, (B_s(\omega))_{0 \le s \le t})$. If not, bad things can happen Say $C_t^n(\omega) = \sum_{k=0}^{2^n - 1} Y_k^n(\omega) 1_{(t_k^n < t \le t_{k+1}^n)}$, where $t_k^n = k/2^n$. Then $\int_0^1 C_t^n \, dB_t$ should be

$$I_n = \sum_{k=0}^{2^n - 1} Y_k^n \cdot \Delta_k^n$$

where $\Delta_k^n = B_{t_{k+1}^n} - B_{t_k^n}$. Consider $Y_k^n = \mathrm{sgn}(\Delta_k^n)/n$. (Then (C_t^n) is anticipating.) Note that $E^0(|B_t|^p) = \int_{\mathbb{R}} (|x|^p/\sqrt{2\pi t}) e^{-x^2/(2t)} \, dx = \gamma_p t^{p/2}$ (for some $\gamma_p \in (0, \infty)$). Then

$$E(|\Delta_k^n|) = \gamma_1 2^{-n/2}$$

$$\mathrm{var}(|\Delta_k^n|) = E(|\Delta_k^n|^2) - E(|\Delta_k^n|)^2 = 2^{-n} - (\gamma_1 2^{-n/2})^2 = c_2 2^{-n}.$$

So,

$$E(I_n) = \frac{2^n}{n} \gamma_1 2^{-n/2} = \gamma_1 \frac{2^{n/2}}{n}.$$

$$\mathrm{var}(I_n) = \frac{1}{n^2} \sum_{k=0}^{2^n - 1} \mathrm{var}(|\Delta_k^n|) = \frac{c_2}{n^2}.$$

Thus

$$E\left[\sum_{n=1}^{\infty} (I_n - E(I_n))^2\right] = \sum_{n=1}^{\infty} \mathrm{var}(I_n) = \sum_{n=1}^{\infty} \frac{c_2}{n^2} < \infty$$

so $\sum_{n=1}^{\infty} (I_n - E(I_n))^2 < \infty$ almost surely so $I_n - E(I_n) \to 0$ almost surely so $I_n \to \infty$ almost surely. Thus, even though $|C_t^n| \le 1/n \to 0$, so $|C_t^n| \to 0$ uniformly on $\Omega \times [0,1]$, we have $\int C_t^n \, dB_t \to \infty$ almost surely.

Actually, $I_n = n^{-1} \sum_{k=0}^{2n-1} |\Delta_k^n|$. Thus $\sum_{k=0}^{2^n - 1} |\Delta_k^n| \to \infty$ almost surely. Thus almost every Brownian path has infinite arc length.

A similar calculation shows that $\sum_{k=0}^{2^n - 1} (B_{t_{k-1}^n} - B_{t_k^n})^2 \to 1$ almost surely. This is related to Taylor's formula: $\int_0^t f'(B_s) \, dB_0 = f(B_t) - f(B_0) + \int_0^t f''(B_s)/2 \, ds$.