Probability 564

Taught by: Dr. Russell Lyons
Written by: Thanic Nur Samin

Class 1: 01/09

23: Poisson Processes

Poisson approximation, law of small numbers.
Take $\text{Bin}(n, p_n) = \sum_{k=1}^{n} \mathbb{1}_{[\text{trial k is a success}]}$
Where $p_n = \frac{\lambda}{n}, \lambda \in (0, \infty)$
Then we have,

Proposition 1. $\text{Bin}(n, \frac{\lambda}{n}) \to \text{Pois}(\lambda)$

Where the convergence is convergence in distribution, or it converges weakly. It means that the cdf converges pointwise.
Let $F_n(x)$ be the cdf of $\text{Bin}(n, p_n)$ and let $F(x)$ be the cdf of $\text{Pois}(\lambda)$.
Then, $F_n(x) \to F(x)$ for every x where F is continuous.
The definition of cdf tells us that,
$F_X(x) := \Pr(X \leq x)$
The cdf only changes at the 'atoms'.
Suppose $x \in (k, k-1)$. Then,
$F_n(x) = \sum_{j=0}^{k} \Pr[X_n = j]$
$F(x) = \sum_{j=0}^{k} \Pr[X = j]$
Thus, We only need to show that,
$\Pr[\text{Bin}(n, \frac{\lambda}{n}) = k] \to \Pr[\text{Pois}(\lambda) = k]$ for all $k \in \mathbb{N}$

Proof. We need to show that,

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\Longleftarrow \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \to e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\Longleftarrow \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \to e^{-\lambda}$$

Which is obvious.

∎

Theorem 1 (23.2 Law of Rare Events)**.** Suppose that $\forall n, \langle z_{n,k}; k \leq r_n \rangle$ are independent indicator r.v.s. If $\lim_{n\to\infty} \sum_{k=1}^{r_n} \Pr[z_{n,k} = 1] = \lambda \in [0, \infty)$ and $\lim_{n\to\infty} \max_{1\leq k\leq r_n} \Pr[z_{n,k} = 1] = 0$, then $\sum_{k=1}^{r_n} z_{n,k} \to \text{Pois}(\lambda)$

Proof. Set $p_{n,k} \coloneqq \Pr[z_{n,k} = 1]$ and $\lambda_n \coloneqq \sum_{k=1}^{r_n} p_{n,k}$. Since $\lambda_n \to \lambda$, we have $\mathrm{Pois}(\lambda_n) \to \mathrm{Pois}(\lambda)$.
So, it suffices to show that $\Pr\left[\sum_{k=1}^{r_n} z_{n,k} = i\right] - e^{-\lambda_n}\frac{\lambda_n^i}{i!} \to 0$
We do this by finding r.v.s V_n, W_n on a common probability space such that $V_n = \sum_{k=1}^{r_n} z_{n,k}$, $W_n \sim \mathrm{Pois}(\lambda_n)$ and $\Pr[V_n \neq W_n] \to 0$.
[Rest of proof not clear from images. Need to rewrite]

□

Class 2: 01/11
[Insert Picture for finishing proof of law of rare events]
Basically $V \cup W \sim \mathrm{Pois}(p)$
$U_k \sim U(0,1)$ independent
We have $V_{n,k}, W_{n,k}$
We have $V_n \coloneqq \sum_{k=1}^{r_n} V_{n,k}$ and $W_n \coloneqq \sum_{k=1}^{r_n} Z_{n,k}$
Since $\forall n, \langle V_{n,k} : 1 \le k \le r_n \rangle \overset{\mathcal{D}}{=} \langle Z_{n,k} : 1 \le k \le r_n \rangle$
Thus $\sum_{k=1}^{r_n} V_{n,k} \overset{\mathcal{D}}{=} \sum_{k=1}^{r_n} Z_{n,k}$
Recall that for random variable $X : (\Omega, \Pr, \mathcal{F}) \to (E, \mathcal{E})$ the distribution is given by the pushforward $\Pr \circ X^{-1} = X_* P$
So, if we have a composition $X \to f(X)$
$(\Omega, \Pr, \mathcal{F}) \xrightarrow{X} (E, \mathcal{E}) \xrightarrow{f} (\mathbb{R}, \mathcal{R})$
So we have the pushforward $f_* X_* \Pr$ for composition. [Don't understand this properly]
Now, since W_n is just sum of independent Poisson r.v.s, $W_n \sim \mathrm{Pois}(\sum_{k=1}^{r_n} p_{n,k}) = \mathrm{Pois}(\lambda_n)$
We want to show that $\Pr[V_n \neq W_n] \to 0$.
$\Pr[V_n \neq W_n] \le \Pr[\exists k : V_{n,k} \neq W_{n,k}]$
From picture, for each k, the RHS probability is $\le \sum_{k=1}^{r_n} \Pr[V_{n,k} \neq W_{n,k}]$ [Union Bound]
$= \sum_{k=1}^{r_n} = \sum_{k=1}^{r_n} \Pr[V_{n,k} = 1] - \Pr[W_{n,k} = 1] = \sum_{k=1}^{r_n} (p_{n,k} - e^{-p_{n,k} p_{n,k}} = \sum_{k=1}^{r_n} p_{n,k}(1 - e^{-p_{n,k}}))$
$\le \max p_{n,k} \cdot \sum_k (1 - e^{-p_{n,k}} \le \max p_{n,k} \cdot \sum_k p_{n,k}) \to 0$
We finally start studying Poisson Processes.
There's also Poisson Point Processes. FIrst we look at examples
Suppose you're manufacturing, and you have surface of a tablet. You don't want defects, but defects are random. You can model that with a poisson point process.
Also suppose you're raising dough for baking cookies. Raise isn't predictable, it can be modeled as a poisson point process.
Typos in location of book is a poisson point process.
Before people understood what stars were, they were assumed to be randomly distributed, poisson point process.
First of all, there are stochastic processes. This course isn't explicitly about them, but we have studied them in Markov Processes. We can consider random variables indexed by time, or we can consider indexed by sets in the space case.
Time t is positive real variable. For each t we can define $N(t)$
$\langle N(t); t \ge 0 \rangle$
Before Poisson Process, we're going to talk about something more general: A counting process. Since we're counting particles or something in space or in time. These values are going to be non-negative integers. $N(t)$ is the (finite) member of "events" that occur in $(0, t]$, if something happens that is called an event. [This is not related to measurable subsets of the probability space.]
Thus, for $s \le t$, we have, $N(t) - N(s)$ counting the number of events in $(s, t]$
Formal definition:

Definition 1. Counting Process: For $\forall t, N(t)$ is an $\mathbb{N}$-valued r.v., so that $\forall s < t, N(s) \leq N(t)$, and $N(\cdot)$ is right-continuous a.s.

Right continuity is equivalent to taking the time interval to be closed on the right.

Definition 2. The increments of $N(\cdot)$ are $\langle N(t) - N(s); 0 \leq s < t \rangle$

Definition 3. We say $N(\cdot)$ has <u>independent increments</u> if suppose we have an independent finite sequence of times $0 = t_0, t_1, \ldots, t_n$ implies $\langle N(t_{i+1} - N(t_i)); 0 \leq t < n \rangle$ are independent.

Definition 4. We say $N(\cdot)$ has <u>stationary increments</u> if the distribution of $N(t) - N(s)$ where $s < t$ depends only on $t - s$.

Note that these conditions require uncountably many things!
Poisson processes are Counting processes that satisfy both. Added stipulation, assume that two events cannot happen at the same time. These are simple counting processes.
First we think of them as discrete times. Then time is indexed by $\mathbb{N}$ and we have bernoulli process in each time slot.
Poisson Processes are essentially limit of Bernoulli processes.

Theorem 2. Suppose that $N(\cdot)$ is a counting process with independent stationary increments that never jumps by > 1. Suppose $N(0) = 0$ but $N \not\equiv 0$. Then $\exists \lambda \in (0, \infty)$ so that $\forall t, N(t) \sim \text{Pois}(\lambda t)$.

Definition 5. A process satisfying those hypotheses is called a <u>Poisson process with rate λ</u>.

Proof. We use theorem 23.2 (law of rare events) as extended in the HW.
[draw picture from 0 to t with x where the event occurs]
[divide the picture into n equal parts with length t/n]
[at each interval we have increments]
[number of events can be anything. but if we take small enough intervals there can only be at most 1 event in any interval]
[it is almost an indicator function]
[increments are nearly indicators, independent and have similar probability, which means they're bernoulli.]
Fix $t > 0$ and let $X_{n,i} := N(\frac{it}{n}) - N(\frac{(i-1)t}{n})$ for $1 \leq i \leq n$. Thus $N(t) = \sum_{j=1}^{n} X_{n,i}$.
Because $N(\cdot)$ is simple (it never jumps by > 1), we have $N(t) = \sum_{r=1}^{n} \mathbb{1}_{[X_{n,i} \geq 1]}$ for large enough n.
Let $p_n := \Pr[X_{n,1} = 1] = \Pr[X_{n,i} = 1]$
Take a subsequence $\langle n_k; k \geq 1 \rangle$ so that $n_k p_{n_k}$ converges in $[0, \infty]$. Call its limit $g(t)$.
The Homework [Insert Problem from HW1] implies, $N(t) \sim \text{Pois}(g(t))$
Thus, $g(t)$ must also be finite, and $g(t) = E[N(t)]$.
So, $g(t)$ does not depend on $\langle n_k \rangle$. So the limit must exist.
Now let t vary. Then $g(s + t) = E[N(s + t)] = E[N(s) + (N(s + t) - N(s))] = E[N(s)] + E[N(t)] = g(s) + g(t)$
Since g is non-decreasing (since N is non-decreasing), we can use Cauchy's Functional Equation to conclude that $g(t) = \lambda t$.
So, $N(t) \sim \text{Pois}(\lambda t)$

□

Class 3: 01/16
Correction: Suppose you have a counting process of some qualiative stuff. Then the points of incerement are distributed via poisson.
[Picture of interval [0,t] divided into parts]

Then, $N(t) = \displaystyle\sum_{i=1}^{n} \mathbb{1}_{[X_{n,i} \geq 1]}$ for large n. Basically, there is a n for each ω.

That is, $N(t) = \displaystyle\lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{1}_{[X_{n,i} \geq 1]}$

So, $N(t)$ is the weak limit of the $\sum$.
Now, if we set $p_n := P[X_{n,i} \geq 1]$
Then $\sum_{i=1}^{n} \mathbb{1}_{[X_{n,i}\geq 1]} \sim Bin(n, p_n)$
n_k so that $np_{n_k} \to \lambda \in [0, \infty]$
[Insert Picture ——x–t-x—-x——-x—¿]
We have arrival times, and inbetween we have waiting times. A lot of the terminology comes from the original application, queueing theory.
Let $X_1 :=$ time of the first arrival.
$P[X_1 \leq t] = 1 - e^{-\lambda t}$
Since $P[X_1 > t] = P[N(t) = 0] = e^{-\lambda t}$
Since, $[N(t) = 0] = [X_1 > t]$
Thus we have, $X_1 \sim Exp(\lambda)$
Note, exponential random variables are important because of its 'memorylessness', aka $P[X_1 > s + t | X_1 > s] = P[X_1 > t]$. One discrete analogue is $geom(p)$ random variables, since that is also memoryless.
We can assume wait times are i.i.d. exponential random.
Now time to prove it rigorously.

Theorem 3. $\forall \lambda \in (0, \infty) \exists$ Poisson process of rate λ.

Proof. Let $X_n \sim Exp(\lambda)$ be independent. Now, define $S_n := \sum_{i=1}^{n} X_i$
In particular, by the strong law of large numbers, $S_n \to \infty$ a.s.
Also, $X_i > 0$ a.s.
Assume those hold always. Define $N(t) = \max_n \{n; S_n \leq t\}$
Note that $N(t)$ is a random variable since firstly it's defined in the same probability space and measurable?
$[N(t) \geq n] = [S_n \leq t]$ and $[N(t) = n] = [S_n \leq t < S_{n+1}]$
In particular, N is indeed a random variable.
Also, it is clearly a counting process with $N(0) = 0$, never jumps by more than 1, and $N \not\equiv 0$.
Fix t. The waiting times after t are $X_1^{(t)} := S_{N(t)+1} - t$
$X_2^{(t)} := X_{N(t)+2}, X_3^{(t)} := X_{N(t)+3}, \ldots$
The counting process $\langle N(t+s) - N(t); s \geq 0\rangle$ is defined via waiting times exactly as $\langle N(t), t \geq 0\rangle$ defined via $\langle X_n; n \geq 1\rangle$
The memory loss property is responsible for $\langle N(t+s) - N(t), s \geq 0\rangle$ being independent of $N(t)$ and with the same law as $N(\cdot)$. This gies that the increments of $N(\cdot)$ are independent and stationary.
To establish this rigorously, we can prove the following:
(i) $\forall j \geq 0, n \geq 0$,
$P[S_n \leq t < S_{n+1}, S_{n+1} - t > y] = e^{-\lambda y} P[S_n \leq t < S_{n+1}]$
(ii) $\forall n \geq 0 \forall j \geq 0$ $P[S_n \leq t < S_{n+1}, S_{n+1} - t > y_1, X_{n+2} > \vec{y_2} \ldots, X_{n+j} > y_j] = P[S_n \leq t < S_{n+1}] \cdot e^{-\lambda y_1} \cdots e^{-\lambda y_j}$
(iii) $\forall n \geq 0 \forall j \geq 1, \forall H \in \mathcal{R}^0, P[N(t) = n, (X_1^{(t)}, \ldots, X_j^{(t)}) \in H] = P[N(t) = n], P[(X_1, \cdots, X_j) \in H]$
(iv) $\forall u \geq 1 \forall m_1 \geq 0 \forall n \geq 0 \forall 0 < s_1 < s_2 < \cdots < s_u$,
$P[N(t) = n, \forall i \in [1, n] N(t + s_i) - N(t) = m_i] = P[N(t) = n] P[\forall i \in [1, n] N(s_i) = m_i]$
(v) $\forall k \geq 1 \forall n_i \geq 0 \forall 0 = t_0 < t_1 < \cdots < t_k$,
$P[\forall i \in [1, k], N(t_i) - N(t_{i-1}) = n_i] = \prod_{i=1}^{k} P[N(t_i - t_{i-1}) = n_i]$
Proof of i:
$P[S_n \leq t < S_{n+1}, S_{n+1} - t > y] = P[S_n = t, X_{n+1} > t + y - S_n]$
$= \int_{x \leq t} P[X_{n+1} > t + y - x] \, \mathrm{d}F_{S_n}(x) = e^{-\lambda y} P[S_n \leq t < S_{n+1}] = e^{-\lambda y} P[X_{n+1} > t - x]$
Proof of ii: First, $X_{n+2}, \ldots, X_{n+j}$ are independent of $[S_n \leq t < S_{n+1}] \cap [S_{n+1} - ty_1]$ so we an take out $X_{n+i} > y_i$ and use (i).

Proof of iii: if $H = (y_1, \infty) \times (y_2, \infty) \times \ldots (y_j, \infty)$ then this is the same as ii. $\pi - \lambda$ theorem gives this to us. [Theorem 10.4]
(iv) Use (3) with $j = \sum_{i=1}^{u} w_i \star 1$ and
$H \coloneqq \left\{(x_1, \ldots x_j) \in \mathbb{R}^j : \forall i \in [1, u] X_1 + \cdots + X_{m_i} \leq S_i < X_1 + \cdots + X_{m_i+1}\right\}$

□

Class 4: 01/18
iv: $\forall u \geq 1, \forall m_i \geq 0, \forall 0 < s_1 < s_2 < \cdots < s_n$, $P[N(t) = n, \forall i \in [1, n], N(t + s_i) - N(t) = m_i] = P[N(t) = n] P[\forall i \in [1, n], N(s_i) = m_i]$
v: $\forall n_i \geq 0, \forall k \geq 1, \forall 0 = t_0 < t_1 < \cdots < t_k$, $P[\forall i \in [1, k], N(t_i) - N(t_{i-1}) = n_i] = \prod_{i=1}^{k} P[(N(t_i - t_{i-1}) = n_i)]$
How to get iv $\implies$ v?
The increments are not the same. The increments in v are successive. But the first one we can. In v, set $t_1 = t$, then first one is the same. Note, in iv, $N(t + s_i) - N(t)$ got changed to $N(s_i)$ so starting time became 0. We keep going, and by induction on k, the number of terms, we get v from iv.
Basically,
$n \coloneqq n_1, u \coloneqq k - 1$
$m_i = n_2 + \cdots + n_{i+1}$
$s_1 = t_2 - t_1, s_i = t_{i+1} - t_1$
Now we need to show that λ is indeed the correct parameter for the poisson process.
Now, $P[N(1) = 0] = P[X_1 > 1] = e^{-\lambda \cdot 1}$
This tells us λ is indeed the right variable for this distribution.

Theorem 4. If $N(\cdot)$ is a Poisson process with rate λ then there exists independent $X_k \sim Exp(\lambda)$ so that almost surely (a.s.) $\forall t, N(t) = \max\{n : \sum_{k=1}^{n} X_k \leq t\}$

This is the same relationship we used to construct N from X in the previous theorem.

Proof. Define $S_n \coloneqq \inf\{t; N(t) \geq n\}$ for $n \geq 0$
We use infimum instead of minimum since a priori we don't know it exists. It is a.s. a minimum.
Then $X_n \coloneqq S_n - S_{n-1}$ for $n \geq 1$
First, we show that these are actually random variables.
So,$[S_n \leq t]$ has to be measureable.
$[S_n \leq t] = [N(t) \geq n]$
Since $N(t)$ is a random variable the latter set must be measurable, so the former set is measureable so S_n is indeed a random variable.
Now, $P[X_1 > t] = P[S_1 > t] = P[N(t) = 0] = e^{-\lambda t}$ which means $X_1 \sim Exp(\lambda)$
Intuition behind why X_k is independent: memorylessness!!!
Suppose we know the value of X_1, then after that $X_2, X_3, \ldots$ must also be random similarly. Pursuing this argument is a bit difficult since X_1 is random so we can't actually set $t = X_1$
Our proof will be to show that $(S_1, \ldots, S_k)$ has density $\lambda^k e^{-\lambda y_k}$
on $\left\{y \in \mathbb{R}^k; 0 < y_1 < \cdots < y_k\right\}$ and 0 elsewhere.
Then deduce that $(X_1, \ldots, X_k)$ has density $\prod_{i=1}^{k} (\lambda e^{-\lambda x_i})$ on $x \in \mathbb{R}^k; \forall i x_i > 0$ and 0 elsewhere.
The second stepfollows from 20.20, using the linear map $x_i \coloneqq y_i - y_{i-1}$ with the Jacobian = 1
Consier $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_k < t_k$
0—s1—-S1—-t1——s2—-S2—t2
Then $P[s_i < S_i \leq t_i \text{ for } 1 \leq i \leq k] = P[N(s_1) = 0, N(t_1) - N(s_1) = 1, N(s_2) - N(t_1) = 0, \ldots, N(t_k) - N(s_k) \geq 1]$
We have expressed the probability over disjoint intervals of N so we can just multiply to get the probability
$= e^{-\lambda s_1} e^{-\lambda(t_1 - s_1)} \lambda(t_1 - s_1) e^{-\lambda(s_2 - t_1)} e^{-\lambda(t_2 - s_2)} \lambda(t_2 - s_2) \cdots e^{-\lambda(s_k - t_{k-1})} (1 - e^{-\lambda(t_k - s_k)})$
$= \lambda^{k-1} e^{-\lambda s_k} (t_1 - s_1)(t_2 - s_2) \ldots (t_{k-1} - s_{k-1})(1 - e^{-\lambda(t_k - s_k)})$
$= \lambda^{k-1} (\prod_{i=1}^{k-1} (t_i - s_i))(e^{-\lambda s_k} - e^{-\lambda t_k})$
$= \int_A \lambda^k e^{-\lambda y_k} \, \mathrm{d}y$ for $A = (s_1, t_1] \times \cdots \times (s_k, t_k]$

So, the densities actually give us probabilities.
That is, if A is a rectangle contained in $G = \{y; 0 < y_1 < y_2 \cdots < y_k\}$ we have $P[(S_1, \ldots, S_k) \in A] = \int_A \lambda^k e^{-\lambda y_k}\, \mathrm{d}y$
These rectangles form a π-system that generates the sigma algebra that has all the borel sets contained in G which is $R^k \cap G$
So this holds for all A in $R^k \cap G$. This gives us the density we wanted.
Since S_n strictly increases to ∞ almost surely we get the desired relation.

□

Let $N(\cdot)$ be a poisson process with parameter λ. Modify it with the deterministic function $f(t) \coloneqq t\mathbb{1}_{\mathbb{Q}}(t)$ and $M(t) \coloneqq N(t) + f(t + X_1)$
Suppose t is fixed. Then $t + X_1$ is irrational a.s.
So we're adding 0 a.s.
So $M(t) = N(t)$ a.s.
So, finite dimensional distributions of M and N are the same - a poisson process
But M is not a counting process.
So knowing finite dimensional distributions only is not enough. This makes Brownian Motion very difficult.

Class 05: 01/23
For HW1 P1, the simplest solution is just using Scheffé's theorem.
Today we work with central limit theorem.
This is called central because it's central to so many things.
Chapter 5: Convergence of Distributions.
Section 25: Weak Convergence
Weak convergence is denoted by $\implies$. It is the same thing as convergence in distribution.
We can look at:
$F_n \implies F$
$\mu_n \implies \mu$
$X_n \implies X$
Note that for weak convergence, X_n, X don't HAVE to be on the same space.
Section 25 is about the interplays between these 3.
Theorem 25.1 we omit.

Theorem 5 (25.2:)**.** Suppose that X_n, X are real valued random variables on the same probability space.
Then, $X_n \overset{a.s.}{\to} X$ implies $X_n \overset{P}{\to} X$ implies $X_n \implies X$

Proof. We just prove $X_n \overset{P}{\to} X$ implies $X_n \implies X$
Let x be a point where F is continuous and $\epsilon > 0$. We are interested in comparing probabilities of $[X \leq x]$ and $[X_n \leq x]$.
Consider the event $[|X_n - X| \geq \epsilon]$. This goes to 0 as $n \to \infty$
Then,
$[X \leq x - \epsilon] \subseteq [X_n \leq x] \cup [|X_n - X| \geq \epsilon]$
Because if $X \leq x - \epsilon$ then $X_n \leq x$ or $|X_n - X| \geq \epsilon$
Also,
$[X_n \leq x] \subseteq [X \leq x + \epsilon] \cup [|X_n - X| \geq \epsilon]$
Because if $X_n \leq x$ then $X \leq x + \epsilon$ or $|X_n - X| \geq \epsilon$
Taking probability,
$P[X \leq x - \epsilon] \leq P[X_n \leq x] + P[|X_n - X| \geq \epsilon]$
$P[X_n \leq x] \leq P[X \leq x + \epsilon] + P[|X_n - X| \geq \epsilon]$
Let $n \to \infty$. Since we'renot sure the limits exist we take limsup/liminf as it suits us.
$P[X \leq x - \epsilon] \leq \liminf_{n\to\infty} P[X_n \leq x]$
$\limsup_{n\to\infty} P[X_n \leq x] \leq P[X \leq x + \epsilon]$
Now let $\epsilon \to 0$
Then,
$P[X \leq x] \leq \liminf_{n\to\infty} P[X_n \leq x] \leq \limsup_{n\to\infty} P[X_n \leq x] \leq P[X \leq x]$

So we're done.

□

At this point, we depart from the order of the book. We prove a later theorem and use that to prove some intermediate theorem.
We have $(X_n \overset{a.s.}{\to} X) \implies (X_n \overset{P}{\to} X) \implies (X_n \implies X)$.
We can also go from last to first under some kind of conditions and stuff. This is theorem 25.6, due to Skorohod.

Proposition 2. If $X_n \implies X$, then $\exists Y_n, Y$ on a common probability space with $Y_n \overset{\mathcal{D}}{=} X_n$ and $X \overset{\mathcal{D}}{=} Y$, and $Y_n \to Y$ pointwise.

Proof. Remember the proof of Poisson? We put them in the same probability space. This is also a similar idea.
We use Lebesuge Measure on $(0,1)$ and let Y_n, Y be the 'inverses' of the cdf's of X_n, X
[Insert Picture here]
The graphs converge in the Lévy metric as defined in Problem 14.5:
$F(x-\epsilon) - \epsilon \leq F_n(x) \leq F(x+\epsilon) + \epsilon$
We have a problem in places where cdf is constant. But there can only be countably many such places, so the problematic stuff has lebesgue measure 0. So we can just not care about it.

□

Now we use this to prove theorem 25.3:

Proposition 3. If $a \in \mathbb{R}$ and $X_n \implies a$, Then $X_n \overset{\text{“}P''}{\to} a$ in the sense that $\forall \epsilon > 0$ we have $P[|X_n - a| \geq \epsilon] \to 0$ as $n \to \infty$ although X_n may not be defined in the same probability space.

Proof. We can just use 25.6 to change them in variables so that they converge almost surely, and then just use 25.2. □

Now consider theorem 25.4:

Proposition 4. Let X_n, Z_n be defined on the same probability space for each n seperately. $[X_j, Z_J$ on the same space, not necessary X_j, X_k or Z_j, Z_K $]$. Suppose $X_n \implies X, Z_n \implies 0$. Then $X_n + Z_n \implies X$

We can't directly use Skorohood because even though we can send X_n to something, we can't do it with $X_n + Z_n$
The proof is similar to the proof of theorem 25.2
Convergence in Distribution we can have some leeway, in our ϵ

Proof. Consider x so that $P[X = x] = 0$
Then we may choose $x' < x < x''$ so that $P[X \in (x', x'']]$ is arbitrarily small and such that $P[X \in \{x', x''\}] = 0$. Now,
$[X_n \leq x'] \subseteq [X_n + Z_n \leq x] \cup [|Z_n| \geq x - x']$
$[X_n + Z_n \leq x] \subseteq [X_n \leq x''] \cup [|Z_n| \geq x'' - x]$
Taking $n \to \infty$
$P[X \leq x'] \leq \liminf_{n\to\infty} P[X_n + Z_n \leq x]$
$\limsup_{n\to\infty} P[X_n + Z_n \leq x] \leq P[X \leq x'']$
This gives us
$P[X_n + Z_n \leq x] \to P[X \leq x]$

□

This is Slutsky's theorem.
Omit theorem 25.5
In HW, the first problems we can already do with what we have.

Class 06: 01/25
Note about HW: you will have more mathematical power if you do less calculation and understand more.

Theorem 6 (25.7, Mapping Theorem)**.** Let $h : \mathbb{R} \to \mathbb{R}$ be Borel, $X_n \implies X$, $P[h \text{ discontinuous at } X] = 0$. Then $h \circ X_n \implies h \circ X$

Proof. We use Skorohood. Let $Y_n \stackrel{\mathcal{D}}{=} X, Y \stackrel{\mathcal{D}}{=} X, Y_n \to Y$ pointwise. Then $h \circ Y_n \to h \circ y$ when h is continuous at Y. Note that $P[h \text{ discontinuous on } X] = 0$ statement only depends on the distribution of X since distribution is a probability measure on the values X can take, so $h \circ Y_n \to h \circ Y$ on the set of probability 1.
Formally, the hypothesis is $(X_* P(\text{discont. set of } h) = 0)$
Therefore, $h \circ Y_n \implies h \circ Y$ and thus $h \circ X_n \implies h \circ X_n$

□

Some notation:

Definition 6.

$$D_h := \{x \in \mathbb{R}; h \text{ is discontinuous set of } x\}$$

So, the hypothesis was: $P[X \in D_h] = 0$
We need D_h to be a borel set for $[X \in D_h]$ to be an event.
Claim: D_h is a borel set for all function h [not necessarily borel h].

Proof.

$$D_h = \bigcup_{\epsilon > 0} \bigcap_{\delta > 0} A_h(\epsilon, \delta)$$

Where $A_h(\epsilon, \delta) := \{x; \exists y, z \in (x - \delta, x + \delta), |h(y) - h(z)| \geq \epsilon\}$ is open.
This is not countable union and intersection but we can make it countable using $\frac{1}{n}$

□

If μ is the law of X then $\mu \circ h^{-1} = h_* \mu$ is the law of $h \circ X$

Theorem 7 (25.8, Portmonteau Theorem)**.** Note: a word blending the sounds and combining the meanings of two others, for example motel (from 'motor' and 'hotel') or brunch (from 'breakfast' and 'lunch').
Let μ_n, μ be probabilities on $\mathbb{R}$. The Following Are Equivalent (TFAE):

1. $\mu_n \implies \mu$
2. $\int f \, \mathrm{d}\mu_n = \int f \, \mathrm{d}\mu, \forall f \in C_b(\mathbb{R})$ [bounded, continuous]
3. $\int f \, \mathrm{d}\mu_n = \int f \, \mathrm{d}\mu$ if f is bounded, borel and $\mu(D_f) = 0$
4. $\mu_n(A) \to \mu(A)$ if A is borel and $\mu(\partial A) = 0$

Proof. $3 \implies 2, 4$ clearly. $4 \implies 1$ by taking $A := (-\infty, x]$
For $1 \implies 3$: Let $Y_n \sim \mu_n, Y \sim \mu, Y_n \to Y$ pointwise.
If $\mu(D_f) = 0$ then $f \circ Y_n \to f \circ Y$ by mapping theorem and so $E[f \circ Y_n] \to E[f \circ Y]$ which gives us 3. [eqn 21.1] by dominated convergence since f is bounded.
For $2 \implies 1$,
Let $f = \mathbb{1}_{(-\infty, x]}$
draw picture:
f is not necessarily continuous. But we can bound f by continuous functions that converge to f :
$g_k \leq \mathbb{1}_{(0, -\infty]} \leq f_k$

$$\lim_{n \to \infty} \int g_k d\mu \leq \mu(-\infty, x] \leq \lim_{n \to \infty} \int f_k d\mu$$

Note: on real numbers, sometimes ineqalities are enough. This is an important trick, might be on exams.

□

For definition of Weak* topologies look at the note in canvas.
This is useful because of the Riesz Representation Theorem, identifying $M(\mathbb{R})$ [the banach space of finite, signed measures on $\mathbb{R}$ wit norm being the total variation] asthe dual of $C_0(\mathbb{R})$ [the space of continuous functions on $\mathbb{R}$ that vanish - tend to 0- at ∞, with the sup norm]. Kakutani extended this from $\mathbb{R}$ to every locally compact Hausdorff space, such as $\mathbb{R}^d$. Recall some basic facts from functional analysis
Just read the note. Won't write the rest.
Basically, $\mu_n \xrightarrow{w^\star} \mu$ if and only if $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ for all $f \in C_0(\mathbb{R})$ [if we required it for $f \in C_b(\mathbb{R})$ then by portmanteau theorem, this is the same as weak convergence for probability measures.]
Uniform Integrability: [Look at notes]
Let $\mathcal{X}$ be a class of real valued random variables. $\mathcal{X}$ is <u>uniformly integrable</u> (UI) if:

$$\lim_{\alpha\to 0} \sup_{X\in\mathcal{X}} E[|X| : |X| \geq \alpha] = 0$$

If $\phi : [0,\infty) \to [0,\infty)$ is Borel and $\lim_{x\to\infty} \phi(x)/x = \infty$ then for all $M < \infty$ we have $\{X; E[\phi(|X|) < M]\}$ is UI.
Also, UI $\iff$ $\sup\{E[|X|]; X \in \mathcal{X}\} < \infty$ and

$$\forall\epsilon > 0 \exists\delta > 0 : P(A) < \delta \implies \sup\{E[|X|; A]; X \in \mathcal{X}\} < \epsilon$$

Thus if $\mathcal{X}$ is UI so is its convex hull.
Corollary: Let X, X_n be integrable random variables with $X_n \to X$ a.s. Then TFAE:

1. $\{X_n\}$ is UI
2. $E[|X_n - X|] \to 0$
3. $E[|X_n|] \to E[|X|]$

We could have a sequence of bounded measure that converges to a not measure. Note that bounded would mean we can't have infinite mass so we can't do the trick in assignment.
A collection $\mathcal{M}$ of probability measures on $\mathbb{R}$ is <u>tight</u> [masses can't run off to infinity] if $\lim_{a\to\infty} \sup_{\mu\in\mathcal{M}} \mu((-a,a)^c) = 0$
Equivalently, $\lim_{a\to\infty} \inf_{\mu\in\mathcal{M}} \mu(-a,a) = 1$. Here we're talking about probability measures, the first one is true for any measures.
Fatou's lemma for weak convergence:

$$X_n \implies X \implies E[|X|] \leq \liminf_{n\to\infty} E[|X_n|]$$

Characteristic Functions [Section 26]

They are basically Fourier Transforms. So alternative name for this section is Fourier Analysis.

Definition 7. The <u>characteristic function</u> or <u>fourier transform</u> of a probability measure μ (or of a random variable X with law μ or distribution F of μ) is the function $\phi : \mathbb{R} \to \mathbb{C}$ given by

$$\boxed{\phi(t) := \hat{\mu}(t) := \int_{-\infty}^{\infty} e^{itx} \,\mathrm{d}\mu(x) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \,\mathrm{d}F(x)}$$

Note that as long as μ is a finite measure this is well-defined.

Alternate definitions: $e^{\pm itx}$, $\frac{1}{\sqrt{2\pi}}e^{\pm itx}$, $\frac{1}{2\pi}e^{\pm itx}$
Note that we write $\hat{\mu}$ but not $\hat{X}$ or $\hat{F}$. But if F has a density, f, then we can write $\hat{f}$.
Note that this is defined since the integrand e^{itx} has absolute value 1. So $|\hat{\mu}| \leq 1$.
Note that we have $\hat{\mu}(s) - \hat{\mu}(t) = \int_{-\infty}^{\infty}(e^{isx} - e^{itx})\,\mathrm{d}\mu(x)$
If $s_n \to t$ then $\hat{\mu}(s_n) \to \hat{\mu}(t)$ by Bounded Convergence Theorem (BCT).
Thus, $\hat{\mu} \in C_b(\mathbb{R})$
The reason this is useful to us:

1. Independence: if X, Y are independent then $E[e^{it(X+Y)}] = E[e^{itX}]E[e^{itY}]$
2. Uniqueness: [to be proved] $\mu \mapsto \hat{\mu}$ is 1-1 [injective]. So if we can find the characteristic function we know what the measure is.
3. Weak Convergence: [to be proved] $\mu_n \implies \mu$ iff $\hat{\mu}_n \to \hat{\mu}$ pointwise.

$|e^{ix} - 1| \leq |x|$ for $x \in \mathbb{R}$
To see this, just draw unit circle.
Recall [Example 18.4]:

$$\lim_{T\to\infty}\int_0^T \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}$$

Let $S(T) \coloneqq \int_0^T \frac{\sin x}{x}\,\mathrm{d}x$
Then for real $\theta \neq 0$,

$$\int_{-T}^T \frac{e^{it\theta}}{2it}\,\mathrm{d}t$$

This is not Lebesuge integrable. We can write $e^{it\theta} = \cos(t\theta) + i\sin(t\theta)$, since cos is even and divided by odd it is odd and goes away. So we have:

$$\int_0^T \frac{\sin(t\theta)}{t}\,\mathrm{d}t = \int_{t=0}^T \frac{\sin(t\theta)}{t\theta}\,\mathrm{d}(t\theta) = \operatorname{sgn}\theta S(T|\theta|)$$

Also, given $\hat{\mu}$ we can find μ
Inversion and Uniquness:

Theorem 8 (26.2, Inversion Formula)**.** If $a \leq b$ and $\mu\{a, b\} = 0$ then

$$\mu(a,b] = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it}\hat{\mu}(t)\,\mathrm{d}t$$

The map $\mu \mapsto \hat{\mu}$ is $1-1$

Last piece of motivation [vague]: we can think of $\mu(a, b]$ as $\mathbb{1}_{[a,b)}$. Think of its fourier transform:

$$\hat{\mathbb{1}}_{(a,b]}(t) = \int_{-\infty}^{\infty} e^{itx}\mathbb{1}_{(a,b]}(x)\,\mathrm{d}x = \int_a^b e^{itx}\,\mathrm{d}x = \left.\frac{e^{itx}}{it}\right|_a^b = \frac{e^{itb} - e^{ita}}{it}$$

Look at the similarity with Inversion Formula. Proof next time.

Class 08: 02/01
$|e^{ix} - 1| \leq |x|$ for $x \in \mathbb{R}$

$$S(T) \coloneqq \int_0^T \frac{\sin x}{x}\,\mathrm{d}x$$

$$\int_{-T}^T \frac{e^{it\theta}}{2it}\,\mathrm{d}t = \operatorname{sgn}(T|\theta|) \text{ when } \theta \neq 0$$

We also had theorem 26.2 above.

Proof. Note that the integrand is bounded by $b - a$ in absolute value.
Let's talk about uniqueness first.
We have $\mu(a, b] = F(b) - F(a)$ if F is the distribution function of μ.
We can let $a \to -\infty$, and we don't get in trouble since $\mu(\{a\})$ is non-zero for only countably many, and $\mu(b)$ is zero because we define it that way.
Since this is increasing, we can say that F is the distribution function of μ when $\mu(\{b\}) = 0$. Since F is right continuous we use that on both sides of b to see that F is actually the distribution function everywhere. That gives us uniqueness.
To prove the inversion formula, calculate the RHS.

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \hat{\mu} \, \mathrm{d}t$$

$$= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \int_{-\infty}^{\infty} e^{itx} \, \mathrm{d}\mu(x) \, \mathrm{d}t$$

The integrand is bounded (in (x, t)) and te product measure is finite so we can apply Fubini's Theorem to get:

$$= \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \, \mathrm{d}t \, \mathrm{d}\mu(x)$$

$$= \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{-it(x-a)} - e^{-it(x-b)}}{it} \, \mathrm{d}t \, \mathrm{d}\mu(x)$$

$$= \int_{-\infty}^{\infty} 2[\operatorname{sgn}(x-a) \cdot S(T|x-a|) - \operatorname{sgn}(x-b) \cdot S(T|x-b|)] \, \mathrm{d}\mu(x)$$

We want to take the limit $T \to \infty$ inside the integral. We can do this because of the bounded convergence theorem.
[Taking limit $T \to \infty$]

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} (\dots) \, \mathrm{d}\mu(x)$$

Note that, when $x \neq a, b$, as $T \to \infty$ we have $S(T|x-a|) \to \frac{\pi}{2}$, same for $x - b$. Using this,

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} 2[\mathbb{1}_{(-\infty,a)}(x)[0] + \mathbb{1}_{(a,b)}(x)[\pi] + \mathbb{1}_{(b,\infty)}(x)[0]] \, \mathrm{d}\mu(x)$$

$$= \int_{a}^{b} \mathrm{d}\mu(x) = \mu(a, b]$$

Which was what we wanted.

□

Finally, Continuity Theorem

Theorem 9. [26.3, Continuity Theorem] Let μ_n, μ be probabilities. Then $\mu_n \implies \mu \iff \hat{\mu_n} \to \hat{\mu}$ pointwise. In fact, [Stronger Condition] if μ_n are probabilities with $\hat{\mu_n} \to g$ pointwise then there exist probability μ such that $g = \hat{\mu}$ and $\mu_n \implies \mu$.

Proof. $\implies$ is the easy one, weak convergence implies pointwise transform of the characteristic function.
Look at the formula:

$$\hat{\mu}(t) = \int_{-\infty}^{\infty} e^{itx} \, \mathrm{d}\mu(x)$$

Note that $f := e^{itx}$ is a bounded continuous function.
So, if $\mu_n \implies \mu$ then using theorem 25.8 with $x \mapsto e^{itx}$ being the function in $C_b(\mathbb{R})$ we directly get the result.
For $\impliedby$:For the converse: we use a "mysterious" calculation.

A key concept is tightness of μ_n so the 'mass' outside big intervals $(-a,a)$ goes to 0. We want to get that from pointwise convergence of $\hat{\mu}$
$x \mapsto e^{itx}$ is a function of f and it's frequency of oscillation is determined by how big t is. If we have a lot of oscillation we're more likely to have cancellations.
So, we'll try to bound the oscillation for small t
Calculation: $\forall u > 0$ we look at the integral:

$$\frac{1}{u}\int_{-u}^{u}(1-\hat{\mu_n}(t))\,\mathrm{d}t$$

$$= \int_{-\infty}^{\infty}\frac{1}{u}\int_{-u}^{u}(1-e^{itx})\,\mathrm{d}t\,\mathrm{d}\mu_n(x)$$

$$= \int_{-\infty}^{\infty}\left(2-\frac{e^{itx}}{uix}\Big|_{-u}^{u}\right)\mathrm{d}\mu_n(x)$$

$$= \int_{-\infty}^{\infty}\left(2-\frac{e^{ixu}-e^{-ixu}}{uix}\right)\mathrm{d}\mu_n(x)$$

$$= \int_{-\infty}^{\infty}\left(2-\frac{2\sin(xu)}{xu}\right)\mathrm{d}\mu_n(x)$$

$$= 2\int_{-\infty}^{\infty}\left(1-\frac{\sin(ux)}{ux}\right)\mathrm{d}\mu_n(x)$$

$$\geq 2\int_{|x|\geq\frac{2}{u}}\left(1-\frac{1}{|ux|}\right)\mathrm{d}\mu_n(x)$$

$$\geq 2\cdot\frac{1}{2}\mu_n\left(\left[-\frac{2}{u},\frac{2}{u}\right]^c\right) = \mu_n\left(\left[-\frac{2}{u},\frac{2}{u}\right]^c\right)$$

Now, $g(0) = \lim_{n\to\infty}\hat{\mu_n}(0) = 1$. So, $1-g(t)\approx 0$ for $t\approx 0$. That is, given $\epsilon > 0$ choose u so that $\frac{1}{u}int_{-u}^{u}(1-g(t))\,\mathrm{d}t < \epsilon$.
By Bounded Convergence Theore, there exists n_0 so that $n\geq n_0$ such that $\frac{1}{u}\int_{-u}^{u}(1-\hat{\mu_n(t)})\,\mathrm{d}t < \epsilon$
Thus $n\geq n_0 \implies \mu_n[-\frac{2}{u},\frac{2}{u}]^c < \epsilon$
If we choose $a\geq\frac{2}{u}$ such that $n < n_0 \implies \mu_n[-a,a]^c < \epsilon$ then for all n we have $\mu_n[-a,a]^c < \epsilon$.
So we have tightness.
Use the corollary from the weak* note.
So we have a subsequential weak limit.
Then we only have to show that the only subsequential weak limit has g as characteristic function by theorem 26.2.
This follows from our first part.

□

Note that $||\hat{\mu}||_\infty \leq ||\mu||_{M(\mathbb{R})}$.
In particular, $\|\hat{f}\|_\infty \leq \|f\|_{L^1(\mathbb{R})}$
Where $\mu = f\cdot\lambda$

Theorem 10 (26.1, Riemann-Lebesgue Lemma)**.** If $f\in L^1(\mathbb{R})$ then $\hat{f}\in C_0(\mathbb{R})$

Lots of ways to prove this.
Note that step functions are finite linear combinations of indicator functions of intervals, and thus are $C_0(\mathbb{R})$. But any $f\in L^1(\mathbb{R})$ can be approximated by step functions!
Then we just use the inequality.
Now we have everything needed to to HW.

Class 09: 02/06
Today we prove:
$\left|\frac{e^{-ita}-e^{-itb}}{it}\right| \leq b-a$

We prove that using the fact $\|\hat{f}\|_\infty \leq \|f\|_{L^1(\mathbb{R})}$
See that $\hat{\mathbb{1}}_{(a,b]}(t) = \frac{e^{itb}-e^{ita}}{it}$, taking the L1 norm it's $b-a$ so we have the inequality.
For $x \in \mathbb{R}$,
We prove $|e^{ix} - 1| \leq |x|$
For $x \in \mathbb{R}$,
$|e^{ix} - (1 + ix - \frac{x^2}{2})| \leq \min\{\frac{|x|^3}{6}, |x|^2\}$
To prove this: [not geometric, taylor series with remainder]
Integrate by parts to get, for $n \geq 0$,

$$\int_0^x (x-s)^n e^{is}\,\mathrm{d}s = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1} e^{is}\,\mathrm{d}s$$

This gives, for any $n \geq 0$,

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}\,\mathrm{d}s$$

Use $n = 2$ case, we get upper bound:

$$\left|\frac{i^3}{2!}\int_0^x (x-s)^2 e^{is}\,\mathrm{d}s\right| \leq \frac{1}{2}\left|\frac{x^3}{3}\right| = \frac{|x|^3}{6}$$

For $n = 1$ case, we get,

$$|e^{ix} - (1 + ix)| \leq \left|\frac{i^2}{1!}\int_0^x (x-s)^1 e^{is}\,\mathrm{d}s\right| \leq \frac{x^2}{2}$$

Triangle inequality gives us the other bound since $\frac{x^2}{2} + \frac{x^2}{2} = x^2$

Central Limit Theorem (CLT) [Section 27]

Theorem 11 (Standard CLT, the plain version, 27.1, Lindenberg - Lévy)**.** This applies for an infinite i.i.d. sequence of random variables.
Let $\langle X_n; n \geq 1\rangle$ be i.i.d. with mean c and standard deviation $\sigma \in (0, \infty)$
If $S_n \coloneqq \sum_{k=1}^n X_k$,
Question: What is the approximate law of S_n as n is big?

$$\frac{S_n - nc}{\sigma\sqrt{n}} \implies N(0,1)$$

We're going to only see the idea of this, and see the proof of a more general version.
Idea: Weak convergence is the same as pointwise convergence of characteristic function. For ease of writing, take $c = 0, \sigma = 1$. Then, we have,

$$E[e^{itS_n/\sqrt{n}}] = (E[e^{itX/\sqrt{n}}])^n$$

$$\text{[we have to justify this]} \approx \left(E\left[1 + \frac{itX}{\sqrt{n}} - \frac{t^2X^2}{2n}\right]\right)^n$$

$$= \left(1 - \frac{t^2}{2n}\right)^n \to e^{-t^2/2}$$

The generality we will work with:
We will assume independence, but do not assume i.i.d.
Second generality:
This is just an statement about distribution. So, they don't have to be in the same probability space.
For each n, we can have independent random variables:
$X_{n,1}, \ldots, X_{n,r_n}$ [not necessarily identically distributed]

Theorem 12 (27.2, Lindeberg)**.** Let $\langle X_{n,k}; 1 \leq k \leq r_n \rangle$ be independent for each n and let $E[X_{n,k}] = 0$ and $\sigma_{n,k}^2 = E[X_{n,k}^2], s_n^2 = \sum_{k=1}^{r_n} \sigma_{n,k}^2 \in (0, \infty)$.
If we have the condition [Lindeberg Condition, 27.8]:

$$\forall \epsilon > 0, \lim_{n \to \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{n,k}| \geq \epsilon S_n} X_{n,k}^2 \, \mathrm{d}P = 0$$

[In words, we can't have any set of random variables dominating the others. So, if we compute the normalized total variance in the space where all the random variables are somewhat big, that must be 0]
$S_n \coloneqq \sum_{k=1}^{r_n} X_{n,k}$
Then,

$$\frac{S_n}{s_n} \implies N(0,1)$$

For intuition of the fact that no one is dominating:

$$\max_{1 \leq k \leq r_n} \frac{\sigma_{n,k}^2}{s_n^2} = \max_{1 \leq k \leq r_n} \frac{1}{s_n^2} \left(\int_{|X_{nk} < \epsilon s_n|} X_{nk}^2 \, \mathrm{d}P + \int_{|X_{nk}| \geq \epsilon s_n} X_{nk}^2 \, \mathrm{d}P \right)$$

as $n \to \infty$ this is $\leq \epsilon^2 + o(1)$ [use the fact that max is ¡ sum]
Example: Theorem 27.1 follows, we only need to verify the Lindeberg condition. That becomes: $\forall \epsilon > 0$,

$$\lim_{n \to \infty} \sum_{k=1}^{n} \frac{1}{n\sigma^2} \int_{|X| > \epsilon\sigma\sqrt{n}} X^2 \, \mathrm{d}P = \lim_{n \to \infty} \frac{1}{\sigma^2} \int_{|X| > \epsilon\sigma\sqrt{n}} X^2 \, \mathrm{d}P \to 0$$

Which must be true since $\int X^2 \, \mathrm{d}P$ is finite.
Lindeberg condition is complicated, so sometimes we might want a sufficient easier condition.
Example: Lindeberg condition holds <u>if</u> for some $\delta > 0$,

$$\lim_{n \to \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E[|X_{n,k}|^{2+\delta}] = 0$$

This is called Lyapounov's Condition.
Suppose $\delta = 1$ then we ae talking about the third moment.

Proof.

$$\sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{n,k}| > \epsilon s_n} X_{n,k}^2 \, \mathrm{d}P \leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{n,k}| \geq \epsilon s_n} \frac{|X_{n,k}|^{2+\delta}}{(\epsilon s_n)^\delta} \, \mathrm{d}P$$
$$\leq \frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E[|X_{n,k}|^{2+\delta}] \to 0$$

□

<u>Example 27.4</u> (27.8) holds if $\sup_{n,k} \|X_{n,k}\|_\infty < \infty$ and $s_n \to \infty$

Proof.

$$\sum_{k=1}^{r_n} \int_{|X_{n,k}| \geq \epsilon s_n} X_{n,k}^2 \, \mathrm{d}P = 0$$

□

Since for big enough n, ϵs_n is big enough so that we're integrating where $X_{n,k}$ can't be that big.

Proof. We start today, but we won't finish today.
To prove 27.2, we use two easy estimates.
$e^x - (1+x) = o(x)$ as $x \to 0$
If $|z_i|, |w_i| \leq 1$ then,

$$\left|\prod_{i=1}^m z_i - \prod_{i=1}^m w_i\right| \leq \sum_{i=1}^m |z_i - w_i|$$

Proof by induction.

$$\left|\prod_{i=1}^m z_i - \prod_{i=1}^m w_i\right| = \left|\prod_{i=1}^{m-1} z_i(z_m - w_m) + w_m \left(\prod_{i=1}^{m-1} z_i - \prod_{i=1}^{m-1} w_i\right)\right| \leq |z_m - w_m| + [\dots]$$

□

Class 10: 02/08

Billingsley 27.2. WLOG, $s_n^2 = 1$.

$$\left|E[e^{itX_{n,k}}] - \left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right)\right|$$

$$= \left|E\left[e^{itX_{n,k}} - \left(1 + itX_{n,k} + \frac{(itX_{n,k})^2}{2}\right)\right]\right|$$

$$\leq E\left[\left|e^{itX_{n,k}} - \left(1 + itX_{n,k} + \frac{(itX_{n,k})^2}{2}\right)\right|\right]$$

$$\leq E\left[\min\left\{\frac{|tX_{n,k}|^3}{6}, |tX_{n,k}|^2\right\}\right]$$

When random variable is small, first inequality is better. When random variable is big, second inequality is better. For all $\epsilon > 0$,

$$\leq \int_{|X_{n,k}|<\epsilon} \frac{1}{6}|tX_{n,k}|^3 \,\mathrm{d}P + \int_{|X_{n,k}|\geq\epsilon} |tX_{n,k}|^2 \,\mathrm{d}P$$

For first integral, bounding by ϵ^3 isn't good enough since it'll become $r_n\epsilon^3$. But since $\sum |X_{n,k}^2|$ is bounded, we can do a trick:

$$\leq \frac{\epsilon|t|^3}{6} \int_{|X_{n,k}|<\epsilon} X_{n,k}^2 \,\mathrm{d}P + t^2 \int_{|X_{n,k}|\geq\epsilon} X_{n,k}^2 \,\mathrm{d}P$$

Adding all this over k, for every ϵ we have,

$$\sum_{k=1}^{r_n} \left|E[e^{itX_{n,k}}] - \left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right)\right| \leq \frac{\epsilon|t|^3}{6} + t^2 \sum_{k=1}^{r_n} \int_{|X_{n,k}|\geq\epsilon} X_{n,k}^2 \,\mathrm{d}P$$

Thus, taking limit,

$$\limsup_{n\to\infty} \sum_{k=1}^{r_n} \left|E[e^{itX_{n,k}}] - \left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right)\right| \leq \epsilon\frac{|t|^3}{6}$$

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \left|E[e^{itX_{n,k}}] - \left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right)\right| = 0$$

Now consider the characteristic function of normals.

$$\sum_{k=1}^{r_n} \left|\left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right) - e^{-t^2\sigma_{n,k}^2/2}\right|$$

Recall that $\max_k \frac{\sigma_{n,k}^2}{s_n^2} \to 0$. Using that,

$$= \sum_{k=1}^{r_n} o\left(\frac{t^2\sigma_{n,k}^2}{2}\right) = o\left(\frac{t^2}{2}\right)$$

Therefore,

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \left| \left(1 - \frac{1}{2}t^2\sigma_{n,k}^2\right) - e^{-t^2\sigma_{n,k}^2/2} \right| \to \infty$$

Therefore, for all t,

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \left| E[e^{itX_{n,k}}] - e^{-t^2\sigma_{n,k}^2/2} \right| \to 0$$

From yesterday's lemma,

$$\lim_{n\to\infty} \left| E[e^{itS_n}] - e^{-t^2/2} \right| \to 0$$

Therefore, $S_n \implies N(0,1)$

$\square$

Estimating the parameter of Exponential Distribution [Useful for, lets say, radioactive decay]

Example 27.2:

Suppose we want to estimate the parameter of an exponential disribution. If i.i.d. $X_k \sim Exp(\alpha)$ then $\overline{X}_n \coloneqq \frac{1}{n}X_k \to \frac{1}{\alpha}$ a.s. Furthermore,

$$\frac{S_n - \frac{n}{\alpha}}{\frac{1}{\alpha}\sqrt{n}} \implies N(0,1)$$

$$\implies \alpha\sqrt{n}\left(\overline{X}_n - \frac{1}{\alpha}\right) \implies N(0,1)$$

$$\implies |X_n| \approx N\left(\frac{1}{\alpha}, \frac{1}{n\alpha^2}\right)$$

How good of an estimate of α is $\overline{X}_n^{-1}$?

First, we're working with weak convergence. Apply Skorohod to get $Z_n \overset{\mathcal{D}}{=} \alpha\sqrt{n}(\overline{X}_n - \frac{1}{\alpha})$

So $Y_n \overset{\mathcal{D}}{=} \overline{X}_n$ and $Y_n^{-1} \overset{\mathcal{D}}{=} \overline{X}_n^{-1}$

Moreover,

$$\frac{\sqrt{n}}{\alpha}(Y_n^{-1} - \alpha) = \frac{\sqrt{n}}{Y_n}\left(\frac{1}{\alpha} - Y_n\right) = -\frac{Z_n}{\alpha Y_n}$$

Note that Z_n is going to Z and $\alpha Y_n \to 1$. So, this goes to $-Z \sim N(0,1)$

Thus, $\frac{\sqrt{n}}{\alpha}(\overline{X}_n^{-1} - \alpha) \implies N(0,1)$

We can say $\overline{X}_n \approx N(\alpha, \alpha^2/n)$

Weak convergence is a precise statement, but $\approx$ is not a precise mathematical statement.

Caution: α is NOT $E[\overline{X}_n^{-1}]$ and $\frac{\alpha^2}{n}$ is not $Var[(\overline{X}_n^{-1})]$. In fact, expectation can be infinite, and convergence statements can be true.

We omit page 363-367, which is CLT for approximate independence, not true independence.

Something that is in the book but we do not prove: CLT says it converges, but how should we tell the 'rate' of convergence?

Theorem 13 (Berry-Esseen)**.** Suppose that X_k are i.i.d. of mean 0, variance σ^2, and $E[|X_k|^3] = \rho < \infty$.
If F_n is the distribution function of $\frac{S_n}{\sigma\sqrt{n}}$ and Φ is the distribution functionalion of $N(0,1)$ then $\|F_n - \Phi\|_\infty \leq \frac{3\rho}{\sigma^3\sqrt{n}}$

We skip section 28 and go to section 29, which is CLT in $\mathbb{R}^k$
First we do weak convergence. First we recall distribution functions in higher dimension.
Suppose X is an $\mathbb{R}^k$- valued random variable.
Its distribution function F is defined by:

$$F(x) \coloneqq P[X \leq x]$$

for $x \in \mathbb{R}^k$. $\leq$ in $\mathbb{R}^k$ means $\leq$ in <u>every</u> coordinate. For 'boxes' A the total mass in them can be calculated by principle of inclusion and exclusion, and that is called $\Delta_A F$.
$F_n \implies F$ means $F_n(x) \to F(x)$ when x is a continuity point of F.
That also gives us the notion of $\mu_n \implies \mu$ and $X_n \implies X$
First we do Portmonteau theorem.

Theorem 14 (Portmonteau Theorem in $\mathbb{R}^k$)**.** Let μ_n, μ be probability measures on $\mathbb{R}^k$. Then TFAE:

1. $\mu_n \implies \mu$
2. $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu, \forall f \in C_b(\mathbb{R}^k)$
3. $\mu_n(A) \to \mu(A)$ whenever A is borel and $\mu(\partial A) = 0$
4. $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ for all closed C
5. $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ for all open G

1,2,3 are what we expected from 1 dimensional version. In 4,5 we have inequalities instead of equality.

Proof. 1 $\implies$ 2:
We look at boxes (open in left, down etc) even though 2 doesn't talk about boxes. The boxes aren't closed or open.
We have $\mu_n(A) = \Delta_A F \to \Delta_A F = \mu(A)$ whenever A is a box all of whose corners are continuity points of F.
Given $f \in C_b(\mathbb{R}^k)$, $\epsilon > 0$, let A be a bounded box with corners at F-continuity points and $\mu(A^c) < \epsilon$. We can choose because we have countable discontinuous <u>hyperplanes</u>. $\square$

Class 11: 02/13
We finish the proof from yesterday.

Proof. 1 $\implies$ 2:
$\mu_n(A) = \Delta_A F_n \to \Delta_A F = \mu(A)$ when A is a box with F-continuous corners. Given $f \in C_b(\mathbb{R}^k)$ and $\epsilon > 0$ let A be a bounded box with F-continuous corners and $\mu(A^c) < \epsilon$. We can choose because we have countably many hyperplanes with positive measure.
Another way to see this is: fix a line, and consider the family of hyperplanes perpendicular to that line. We can 'project' to get a measure on the real line, by passing the measure of the half space to that of the 'left' of the line.
[insert picture]
Claim: for n large, $\mu_n(A^c) < \epsilon$.
To see this, note that $\mu_n(A) \to \mu(A)$ so $\mu_n(A^c) \to \mu(A^c)$ and thus $\mu_n(A^c) < \epsilon$ for large enough n.
What happens inside A?

Note that f is uniformly continuous in A.
Notation: $f \restriction A$ means f restricted to A.
We may partition A into boxes $A_1, \dots, A_r$ with corners F-continuous and $\sup f \restriction A_i - \inf f \restriction A_i < \epsilon$. Then,

$$\begin{aligned}
\limsup_{n\to\infty} \int f \, \mathrm{d}\mu_n &\le \limsup_{n\to\infty} \left(\epsilon \|f\|_\infty + \sum_{i=1}^{r} \int_{A_i} f \, \mathrm{d}\mu_n \right) \\
&\le \epsilon \|f\|_\infty + \limsup_{n\to\infty} \sum_{i=1}^{r} (\max f \restriction A_i) \mu_n(A_i) \\
&= \epsilon \|f\|_\infty + \sum_{i=1}^{r} (\max f \restriction A_i) \mu(A_i) \\
&\le 2\epsilon \|f\|_\infty + \int f \, \mathrm{d}\mu + \epsilon
\end{aligned}$$

Likewise,

$$\liminf_{n\to\infty} \int f \, \mathrm{d}\mu_n \ge -2\epsilon \|f\|_\infty + \int f \, \mathrm{d}\mu - \epsilon$$

Hence $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$
Now, $2 \implies 4$:
We want to take $C_b(\mathbb{R}^k) \ni f \ge \mathbb{1}_C$ with close integrals w.r.t. μ.
Construction: For $x \in C$, set $f(x) = 1$
For $x \notin C$ do linear interpolation with $dist(x, C)$ from 1 to 0 over a distance ϵ. Then,

$$\limsup_{n\to\infty} \mu_n(C) = \limsup_{n\to\infty} \int \mathbb{1}_C \, \mathrm{d}\mu_n \le \limsup_{n\to\infty} \int f \, \mathrm{d}\mu_n = \int f \, \mathrm{d}\mu = \mu(C + B_\epsilon(0))$$

Where $C + B_\epsilon(0)$ is 'adding' the ball of radius ϵ to every point in C, which means it contains stuff ϵ distance away. Since C is closed, $\mu(C + B_\epsilon(0))$ has limit $\mu(C)$ when $\epsilon \to 0$. That gives us 4.
$4 \iff 5$ by taking complement, $\mu_n(\mathbb{R}^k) = 1 = \mu_n(G) + \mu_n(G^c)$
$5 \implies 3$: Since $4 \iff 5$ we can use both 4 and 5.
Since $\mu(\partial A) = 0$, we have:

$$\begin{aligned}
\mu(A) = \mu(A^o) \le \liminf_{n\to\infty} \mu_n(A^0) \le \liminf_{n\to\infty} \mu_n(A) \le \limsup_{n\to\infty} \mu_n(A) \\
\le \limsup_{n\to\infty} \mu_n(\overline{A}) \le \mu(\overline{A}) = \mu(A)
\end{aligned}$$

Finally, we prove that $3 \implies 1$ to finish the proof.
1 is just a special case of 3. Take A to be a box $\{y : y \le x\}$ with F continuous at x. From definition of continuity means $\mu(\partial A) = 0$. This gives us the result.

□

Definition 8. Tightness in $\mathbb{R}^k$: A collection $\mathcal{M}$ of probabilities on $\mathbb{R}^k$ is tight if $\lim_{a\to\infty} \sup_{\mu\in\mathcal{M}} \mu\left(\left((-a, a]^k \right)^c \right) = 0$

Theorem 15 (29.3)**.** Every tight sequence has a weakly convergent subsequence.

This is sequential compactness of measures. This proof is different from the 1d case.

Proof. Let $\langle \mu_n; n \ge 1 \rangle$ be tight.
Regardless of tightness, there is a subsequence that converges in weak*.
Take a subsequence $\langle \mu_{n_j}; k \ge 1 \rangle$ that converges weak* [by Banach-Alaoglu] to some μ.
We will show that $\mu_{n_j} \implies \mu$.

We are going to use part 2 of Portmonteau Theorem. Consider $f \in C_b(\mathbb{R}^k), \epsilon > 0$. Choose $a > 0$ such that $\mu((-a,a]^k) < \epsilon$ [since μ is finite] and for all n, $\mu_n(((-a,a]^k)^c) < \epsilon$ [since tight].
We want to approximate $f \in C_b(\mathbb{R}^k)$ with $g \in C_0(\mathbb{R}^k)$.
Define $g \in C_0(\mathbb{R}^k)$ so that $f = g$ on $(-a,a]^k$ and $\|g\|_\infty \leq \|f\|_\infty$. Why can we do something like this?
Take $A = (-a,a]^k$, take $h \in C_0$ that goes to 0 by linear interpolation and 1 in A, and take $g = fh$.
Other construction: Instead of A take a ball containing A, make $g = f$ in A, and go down linearly on the rays. That gives us a g.
We can also make make g have compact support.
Then,

$$\left|\int f \,\mathrm{d}\mu_{n_j} - \int f \,\mathrm{d}\mu\right| \leq \int |f-g| \,\mathrm{d}\mu_{n_j} + \left|\int g \,\mathrm{d}\mu_{n_j} - \int g \,\mathrm{d}\mu\right| + \int |g-f| \,\mathrm{d}\mu$$

Define the integrals to be I_1, I_2, I_3
Then, $I_1 \leq 2\epsilon\|f\|_\infty$
$I_2 \to 0$ because weak*.
$I_3 \leq 2\epsilon\|f\|_\infty$
Thus, $I_2 + I_2 + I_3 \leq 4\epsilon\|f\|_\infty$.
Thus $\mu_n \implies \mu$.
Now we need to prove that μ is a probability measure. That is given by taking $f \equiv 1$.
□

It follows that $\langle \mu_n; n \geq 1\rangle$ is tight if and only if every subsequence contains a further weakly convergent subsequence.
Why is the reverse direction true? Recall the definition of tightness. If not tight, then we can find ϵ so that for some μ in the sequence $\sup \mu(((-a,a]^k)^c)$, so by varying a we can find a subsequence so $\mu_{n_j}(((-j,j]^k)^c) > \epsilon$. We can choose $\mu_{n_{j_l}} \implies \mu$. Thus, $\mu(((-j,j]^k)^c) > \epsilon$ for all j. So μ is not a probability measure $\mu(\mathbb{R}^k) < 1 - \epsilon$.

Class 12: 02/20
Today we talked about Exam 1 solutions.
Before the exam, we were doing weak convergence in $\mathbb{R}^k$ in order to do CLT in $\mathbb{R}^k$.
We also had theorem 29.3: every tight sequence has a weakly convergent subsequence.
Lemma: Let $\langle \mu_n; n \geq 1\rangle$ be a tight sequence of probability measures on $\mathbb{R}^k$ Let μ be a sub-probability measure. Then $\mu_n \implies \mu \iff \mu_n \xrightarrow{w^*} \mu$.
Proof is in the lecture notes.
We also had corollary: If $\langle \mu_n, n \leq 1\rangle$ is a tight sequence with ≤ 1 weak limit point, then it has a weak limit.

Proof. By theorem 29.3 the sequence has some a limit point. Let it be μ. Thus, it has exactly one weak limit point.
[if we had a metric then it would be over. but we might not so we go the weak * route]
For every subsequence of μ_n that converges weakly to μ, it converges weakstarly to μ. Since we have a metric in weakstar, the sequence converges to μ in weakstar and thus it converges weakly to μ.
Since the sequence has exactly one weak limit point, it only has one weak* limit point.
By the metrizability of the weak* topology, $\mu_n \xrightarrow{w^*} \mu$. Thus, $\mu_n \implies \mu$.
□

Characteristic Function on $\mathbb{R}^k$

Definition 9. For a $\mathbb{R}^k$-valued random variable X, $t \in \mathbb{R}^k$, we define the characteristic function:

$$t \mapsto E[e^{it\cdot X}]$$

Note that instead of multiplying we have the dot product. We write it as $\hat{u}(t)$ if the law of X is μ.

The proof that $\mu \mapsto \hat{\mu}$ is $1-1$: in $\mathbb{R}$ was by the 'inversion formula'. That extends to $\mathbb{R}^k$ as follows:
If $A = (a_1, b_1] \times \cdots (a_k, b_k]$ is a bounded rectangle such that $\mu(\partial A) = 0$, then:

$$\mu(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^k} \int_{[-T,T]^k} \prod_{j=1}^{k} (e^{-it_j a_j} - e^{-it_j b_j}) \cdot \hat{\mu}(t)\,\mathrm{d}t$$

Indeed, $\hat{\mu}(t) = \int_{\mathbb{R}^k} \prod_{j=1}^{k} e^{it_j x_j}\,\mathrm{d}\mu(x)$
The integrand in the inversion formula is product of two things that are uniformly bounded so we can apply Fubini. Lebesgue measure is a product measure. Thus,

$$\lim_{T\to\infty} \int_{\mathbb{R}^k} \prod_{j=1}^{k} \left[\frac{1}{\pi} sgn(x_j - a_j) S(T|x_j - a_j|) - \frac{1}{\pi} sgn(x_j - b_j) S(T|x_j - b_j|) \right] \mathrm{d}\mu(x)$$

Applying Bounded Convergence Theorem

$$= \int_{\mathbb{R}^k} \prod_{j=1}^{k} \mathbb{1}_{(a_j, b_j]}(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}^k} \mathbb{1}_A(x)\,\mathrm{d}\mu(x) = \mu(A)$$

So, μ is detemined uniquely on bounded rectangles. This determines the distribution function F_μ where it is continuous. Since F_μ is right continuous and continuity points are dense, F_μ is determined everywhere. So μ is also determined.
Write $h_t(x) \coloneqq t \cdot x$. If t is a unit vector, then $t \cdot x$ is the length of the orthogonal projection. So it is the length of orthogonal projection scaled by the length of t.
So, inverse image of any point is a hyperplane.
Suppose $\alpha \in \mathbb{R}$. We look at $h_t^{-1}(-\infty, \alpha]$. This gives us one side of the hyperplane, that is a half space. Also, for $s \in \mathbb{R}$,

$$\widehat{\mu \circ h_t^{-1}}(s) = \int_{\mathbb{R}} e^{isy}\,\mathrm{d}(\mu h_t^{-1})(y)$$

Using theorem 16.13,

$$= \int_{\mathbb{R}^k} e^{ish_t(x)}\,\mathrm{d}\mu(x)$$

$$= \int_{\mathbb{R}^k} e^{ist\cdot x}\,\mathrm{d}\mu(x) = \hat{\mu}(st)$$

Thus, the values of μ on half spaces determine all measures μh_t^{-1}, hence $\widehat{\mu h_t^{-1}}$ hence all $\hat{\mu}$ and hence μ.

Theorem 16 (29.4, Cramér-Wold device)**.** $X_n \implies Y \iff \forall t \in \mathbb{R}^k, t\cdot X_n \implies t\cdot Y$

Class 13: 02/22
We try to construct a counterexample to the case $X_n^{(i)} \implies Y^{(i)}$ but $X_n \not\Longrightarrow Y$.
We use dependence.
Suppose all the $X_n^{(j)}$ have same distribution as $Y^{(j)}$. But we can use dependence to show that X_n might not converges to Y.

Proof. Easy direction: Suppose $X_n \implies Y$. Then, for all $f \in C_b(\mathbb{R})$ we need to prove that $E[f(t \cdot X_n)] \implies E[f(t \cdot Y)]$.
Note that $E[f(t \cdot X_n)] = E[(f \circ h_t)(X_n)] \implies E[(f \circ h_t)(Y_n)] = E[f(t \cdot Y)]$
Thus $t \cdot X_n \implies t \cdot Y$.
Hard direction: If for all t we have $t \cdot X_n \implies t \cdot Y$,
$\phi_{t \cdot X_n}(1) = E[e^{it \cdot X_n}]$
Since $y \mapsto e^{iy} \in C_b(\mathbb{R})$ we have:
$E[e^{it \cdot X_n}] \to E[e^{it \cdot Y}]$
Thus, $\phi_{X_n}(t) \to \phi_Y(t)$
Note that $\{X_n; n \geq 1\}$ is tight by HW [29.3b].
So, there is at least one weak limit.
Recall that weak convergence implies convergence in characteristic functions. The above shows that every weak limit has the same characteristic function as Y, so by the uniqueness of characteristic functions, $\langle X_n \rangle$ has ≤ 1 weak limit. Thus, it converges to Y.

□

Thus, $X_n \implies Y$ if and only if $\phi_{X_n} \to \phi_Y$

Multivariate Normal Distribution

Key: all we need are means of the coordinates and covariances of the coordinates.
Notation: for $x \in \mathbb{R}^k$, we will use $|x| \coloneqq \sqrt{x \cdot x}$.
The function:

$$\boxed{x \mapsto (2\pi)^{-\frac{k}{2}} e^{-|x|^2/2}}$$

is a density, called the standard normal density.
If $X = (X_1, \cdots, X_k)$ has this density, then $X_1, \cdots, X_k$ are independent $N(0,1)$ random variables by section 20 and conversely.

$$E[e^{it \cdot X}] = \prod_k E[e^{it_k X_k}] = \prod_k \phi_{X_k}(t_k) = \prod_k e^{-t_k^2/2} = e^{-\sum_k t_k^2/2} = \boxed{e^{-|t|^2/2}}$$

Now we need to discuss the Covariance Matrix.
Let X be a n-dimensional random variable with mean 0, that is $E[X] = 0$, the covariance matrix has i,j-th entry given by $E[X_i X_j]$.
We write it in matrix form. If s, t are column vectors, then $s \cdot t = s^\top t$. But in probability, transpose is denoted often by $'$. So we write $s't$ instead of $s^\top t$.
Thus, if X is a column vector random variable, the covariance matrix is given by:

$$\boxed{\Sigma \coloneqq E[XX']}$$

Σ is non-negative definite/positive semidefinite, meaning for al $x \in \mathbb{R}^k$ [column vector], we have $x'\Sigma x \geq 0$.
$x'\Sigma x = x'E[XX']x = E[x'XX'x] = E[(X'x)'X'x] = E[|X'x|^2] \geq 0$

Proposition 5. For every symmetric p.s.d. Σ, there is a normal distribution $N(0, \Sigma)$ with covariance matrix Σ.

Proof. Because Σ is symmetric, we can diagonalize it by an orthogonal matrix U. Because it is p.s.d., the eigenvalues must be ≥ 0. Thus $\Sigma = UDU'$ where D is diagonal. Columns of U are the eigenvectors.
Notation: positive semidefinite is also written as $\Sigma \geq 0$. That tells us $D \geq 0$.
Let $A \coloneqq U\sqrt{D}$. Then $AA' = U\sqrt{D}\sqrt{D}U' = UDU' = \Sigma$.
Now, if $X \sim N(0, I)$ [standard normal in k variables], Define:

$$\boxed{Y \coloneqq AX}$$

We have $E[YY'] = E[AXX'A'] = AE[XX']A' = AIA' = AA' = \Sigma$.

□

Lets calculate the characteristic function of $Y = AX$.

$$\phi_Y(t) = E[e^{it\cdot Y}] = E[e^{it'AX}] = E[e^{i(A't)\cdot X}] = \phi_X(A't) = e^{-|A't|^2/2}$$

$$= e^{-t'AA't/2} = \boxed{e^{-t'\Sigma t/2}}$$

Thus the characteristic function only depends on Σ.
This shows that the distribution depends only on Σ.
If Σ is singular, then so is D since one of the eigenvalues must be 0. So A must be singular as well.
We defined $Y := AX$. Since A is singular, A must have a null space. So, Y does not have a density. Canonical example: $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, then $Y = \begin{pmatrix} X_1 \\ 0 \end{pmatrix}$ which does not have a density in $\mathbb{R}^2$.
If Σ is invertible then so is A. We claim that in this case Y does have a density.
Let f be the density of X.
Let λ_k be the lebesgue measure on $\mathbb{R}^k$.
Then Y has law $(f\lambda_k) \circ A^{-1} = (f \circ A^{-1})(\lambda_k \circ A^{-1})$
Since $\int hd(f\lambda_k) \circ A^{-1} = \int h \circ Ad(f\lambda_k) = \int (h \circ A) f d\lambda_k$
And $\int h(f \circ A^{-1})d(\lambda_k \circ A^{-1}) = \int (h \circ A) f d\lambda_k$.
Now, $(f \circ A^{-1})(x) = (2\pi)^{-k/2} e^{-|A^{-1}x|^2/2} = (2\pi)^{-k/2} e^{-x'A'^{-1}A^{-1}x/2}$
$= (2\pi)^{-k/2} e^{-x'\Sigma^{-1}x/2}$
And $\lambda_k \circ A^{-1} = |\det A^{-1}|\lambda_k = |\det A|^{-1}\lambda_k = (\det\Sigma)^{\frac{1}{2}}\lambda_k$
Therefore, Y has density:

$$(2\pi)^{-k/2}(\det\Sigma)^{-1/2} e^{-x'\Sigma^{-1}x/2}$$

There are several properties of multivariate normal random variables.
If two random variables are independent then they are uncorrelated. The converse is not true: random variables can be uncorrelated and dependent.
We construct a counterexample. Suppose X, Y have mean 0, indepdendent [so $E[XY] = 0$] and $E[Y^3] = 0$
Take $X - aY^2, Y$.
Then $E[(X - aY^2)Y] = 0 - aE[Y^3] = 0$
So they are uncorrelated for all a.
Uncorrelatedness is a one-dimensional phenomenon, independence is more general and thus more restrictive.
We want to show that for multivariate/joint normal, uncorrelatedness is equivalent to independence.

Class 14: 02/27
Recall:
Uncorrelation is a one dimensional concept, Independence is infinite dimensional. So we do not expect uncorrelated to imply indepndent, but that happens for joint normal. Today we see why.

Proposition 6. Suppose $(X_1, \ldots, X_k)$ has a normal distribution. Then so does $X_1, \ldots, X_j$ is normal for all $j < k$

Proof. There is a covariance matrix for $X = (X_1, \cdots, X_k)$. Say that is A. Then $X = AY$ where Y is a standard normal.
Let π be the projection on the first j coordinates. If $Z = (X_1, \cdots, X_j)$ then we have: $Z = \pi(X) = \pi(AY) = \pi \circ A(Y)$
Then, $Z' = (X_1, \cdots, X_j, 0, \cdots, 0) \in \mathbb{R}^k$ is normal because of the covariance matrix $\pi \circ A$.
We are going nowhere.
X has a covariance matrix, so does Z. Let them be Σ_1, Σ_2
$E[e^{it\cdot X}] = e^{-t'\Sigma_1 t/2}$
$E[e^{is\cdot Z}] = e^{-t'\Sigma_2 t/2}$

for each $s \in \mathbb{R}^j$ we have $t \in \mathbb{R}^k$ so that $s \cdot Z = t \cdot X$
So we have the characteristic function, so it is normal.

$\square$

Suppose all coordinates of X are uncorrelated. We want to show they are independent.
Since all coordinates of X are uncorrelated, Σ is diagonal.
Then, the characteristic function is $t \mapsto e^{-\sum_i \sigma_i^2 t_i^2/2}$ where $\sigma_i^2 = Var(X_i)$.
We can write it as $\prod_i e^{-\sigma_i^2 t_i^2/2}$
This is the same characteristic function we would have if they were independent, therefore they must be independent.
Characteristic functions are product if and only if random variables are independnet.
General statement: if $Y \in \mathbb{R}^k$ is normal and M is a linear transformation $\mathbb{R}^k \to \mathbb{R}^j$, then $MY \in \mathbb{R}^j$ is also normal.
Note that this also means normality doesn't depend on the basis, it is a property of the space.
If the covariance matrix of Y is Σ, then,

$$E[e^{it'MY}] = E[e^{i(M't)'Y/2}] = e^{-t'M\Sigma M't/2} = e^{-t'(M\Sigma M')t/2}$$

Note that $M\Sigma M'$ is symmetric and positive semidefinite, therefore it is a covariance matrix, and as a result this is the characteristic function of a normal random variable. Thus this is a normal random variable.

Theorem 17 (29.5, IID)**.** Let X_n be i.i.d. $\mathbb{R}^k$ valued random variables with all components having finite 2nd moment. Let $c = E[X_n]$ and $\Sigma = E[(X_n - c)(X_n - c)']$.
Let $S_n := \sum_{k=1}^n X_k$.
Then, $\frac{S_n - nc}{\sqrt{n}} \implies N(0, \Sigma)$.
If Σ is invertible, $\sqrt{\Sigma}^{-1} \frac{S_n - nc}{\sqrt{n}} \implies N(0, I)$.

Proof. Recall Cramér-Wold, we only need to show one dimensional.
Let $Y \sim N(0, \Sigma)$. By 29.4, it suffices to prove that for all $t \in \mathbb{R}^k$,
$t'(S_n - nc)/\sqrt{n} \implies t'Y$.
Note that, $t'Y \sim N(0, t'\Sigma t)$.
$t' \frac{S_n - c}{\sqrt{n}} = \frac{\sum_{k=1}^n t'(X_k - c)}{\sqrt{n}}$
So, we are summing i.i.d. random variables in the numerator with mean $E[t'(X_k - c)] = t'E[X_k - c] = t'0 = 0$ and variance $E[(t'(X_k - c))^2] = E[t'(X_k - c)(X_k - c)'t] = t'E[(X_k - c)(X_k - c)']t = t'\Sigma t$.
If $t'\Sigma t = 0$ then these random variables have mean 0 and variance 0, so the random variables are trivially 0.
If not, we can divide by $t'\Sigma t$ [this is just a number], and since $t'(X_k - c)$ are i.i.d. we have our result by the Lindeber Lévy theorem.

$\square$

Now we go to chapter 6.

Chapter 6: Derivatives and Conditional Probability

Section 32: The Radon-Nikodym Theorem.
Recall that if μ, ν are signed or complex measures on the same measurable space, we call ν <u>absolutely continuous</u> with respect to μ, written $\nu \ll \mu$ if every μ-null set is a ν-null set. Recall that, a null set is a set such that every subset of that set has measure 0.
If ν is finite, this is equivalent to: $\forall \epsilon > 0 \exists \delta > 0 \forall$ measurable E, $|\mu|E < \delta \implies |\nu| < \epsilon$.
If there is a μ-null set whose complement is ν-null, then we call μ and ν <u>singular</u>, written $\mu \perp \nu$

Theorem 18 (Lebesgue-Radon-Nikodym Theorem)**.** Let μ, ν be σ-finite [the whole set can be written as a countable union of sets with finite measure] signed or complex measures on $(\Omega, \mathcal{F})$. Then there are unique signed or complex measures ν_a and ν_s on $\Omega, \mathcal{F}$ such that:
$\nu = \nu_a + \nu_s$
$\nu_a \ll \mu$
$\nu_s \perp \mu$
This is a lebesgue decomposition.
There is a function $f : \Omega \to \mathbb{C}$, unique upto ν-null modifications, such that $\nu_a = f\mu$. This means any two are equal μ-almost everywhere. We denote f by $\frac{\mathrm{d}\nu_a}{\mathrm{d}\nu}$. Also, ν_a is finite if and only if $f \in L^1(|\mu|)$.

33. Conditional Probabilities

Instead of conditioning on an event, we condition on a σ-field.
How do we define $P(A|\mathcal{G})$, for a σ-field $\mathcal{G} \subseteq \mathcal{F}$?
Typically, $\mathcal{G}$ is generated by a (or some) random variables. Consider the simplest case: $\mathcal{G} = \{\varnothing, \Omega\}$.
This doesn't give us any information, so it should be that $P(A|\{\varnothing, \Omega\}) = P(A)$.
Next, suppose $\mathcal{G} = \{\varnothing, B, B^c, \Omega\}$.
This gives us information on whether B happened or not. So, this gives us different values, and thus this is actually a random variable!
$P(A|\mathcal{G}) = P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}$
Similarly, if $\mathcal{G}$ is generated by a countable partition $\mathcal{P}$, then,
$P(A|\mathcal{G}) = \sum_{B \in \mathcal{P}} P(A|B)\mathbb{1}_B$
To generalize, note that,
$P(A|\mathcal{G})$ is a random variable. It is measureable with respect to $\mathcal{F}$, but it is also measureable with respect to $\mathcal{G}$. We write it as $P(A|\mathcal{G}) \in \mathcal{G}$.
$\forall G \in \mathcal{G}$, $\int_G P(A|\mathcal{G})\,\mathrm{d}P$. Since we're integrating on G, we're integrating over all possible sets. Thus, we have
$\int_G P(A|\mathcal{G})\,\mathrm{d}P = \sum_{B \in \mathcal{P}} \int_G P(A|B)\mathbb{1}_B\,\mathrm{d}P = \sum_{B \in \mathcal{P}} P(A|B)P(B \cap G)$
Note that, $B \cap G = B$ or $\varnothing$
$= \sum_{B \in \mathcal{P}, B \subseteq G} P(A|B)P(B) = \sum_{B \in \mathcal{P}, B \subseteq G} P(A \cap B) = P(A \cap G)$.

Class 15: 02/29
Last time, we saw conditional probability given a σ-field. It had two properties:

Definition 10. Any random variable $P(A|\mathcal{G})$ such that:
i: $P(A|\mathcal{G})$ is $\mathcal{G}$-measurable and integrable random variable
ii: For any $G \in \mathcal{G}$ if we integrate the random variable on this set, $\int_G P(A|\mathcal{G})\,\mathrm{d}P = P(A \cap G)$
is called a version of the probability of A given $\mathcal{G}$

Recall from real analysis: if we have a measurable function on some measure, and we know the integral of that function on every set of the sigma field, then we know the function except on a set of measure 0. That implies this.
Another way: think of this as a function times a measure instead of a fucntion, where the measure is defined by the formula $\int_G P(A|\mathcal{G})\,\mathrm{d}P$. Then this tells us what is the measure of the sigma field. Question: why is it absolutely continuous w.r.t. P? because it's a function times P.

Theorem 19. $\forall A \in \mathscr{F}$, $\forall \sigma$ field $\mathscr{G} \subset \mathscr{F}$ there is a version of $P(A|\mathscr{G})$. It is unique up to P-null modifications.

This theorem is due to Kolmogorov.

Proof. This will be a radon nikodym derivative.
Fix $A, \mathscr{G}$. Let $P_{\mathscr{G}} := P \upharpoonright \mathscr{G}$, P restricted to $\mathscr{G}$
Define $\nu_A : G \to P(A \cap G)$ on $\mathscr{G}$
These are measures in $\mathscr{G}$.

We have $\nu_A \ll P_{\mathscr{G}}$.
By Radon-Nikodym theorem, we may set $P(A|\mathscr{G}) := \frac{\mathrm{d}\nu_A}{\mathrm{d}P_{\mathscr{G}}}$
This gives us what we wanted.

$\square$

Example: $P(A|\mathscr{F}) = \mathbb{1}_A$ almost surely.
Example: if $A \in \mathscr{G}$ we have
$P(A|\mathscr{G}) = \mathbb{1}_A$
Example 33.6: If A is independent of $\mathscr{G}$, then $P(A|\mathscr{G}) = P(A)$ almost surely.
If $\mathscr{X}$ is a collection of random variables, we write $P(A|\mathscr{X})$ for $P(A|\sigma(\mathscr{X}))$. If $\mathscr{X} = \{X\}$ we just write $P(A|X)$
$P(A|X)$ must be measurable with respect to X aka $\sigma(X)$
By theorem 20.1, if $\mathscr{X} = \{X\}$ then $P(A|X)$ is a function of X. This is because random variables are measurable with respect to a random variable, it is a function of that random variable.

Conditioning on a random varible with density

Example 33.5: Suppose that $(X,Y) \sim f\lambda_2$ [f is density, λ_2 is lebesgue measure, so f is lebesgue measurable].
What is $P(Y \in \cdot|X)$?
The classical formula says that it has a density:
When $X = x$, it is $y \mapsto \frac{f(x,y)}{\int_{\mathbb{R}} f(x,t)\,\mathrm{d}t}$
In other words, we want a version of $P(Y \in C|X)$ to be:

$$\frac{\int_{y\in C} f(X,y)\,\mathrm{d}y}{\int_{\mathbb{R}} f(X,t)\,\mathrm{d}t}$$

This is a measurable function of X by theorem 18.3 [fubini].
So, this is $\sigma(X)$ measurable.
Now, for every $G \in \sigma(X)$, we have $G = [X \in H]$ for some borel set $H \in \mathscr{R}^1$, hence

$$\int_G \frac{\int_{t\in C} f(X,t)\,\mathrm{d}t}{\int_{t\in\mathbb{R}} f(X,t)\,\mathrm{d}t}\,\mathrm{d}P = \int_{[X\in H]} \frac{\int_{t\in C} f(x,t)\,\mathrm{d}t}{\int_{t\in\mathbb{R}} f(X,t)\,\mathrm{d}t}\,\mathrm{d}P$$

Change of varibles fn of (X,Y)

$$\int_{x\in H, y\in\mathbb{R}} \frac{\int_{t\in C} f(x,t)\,\mathrm{d}t}{\int_{t\in\mathbb{R}} f(x,t)\,\mathrm{d}t} f(x,y)\,\mathrm{d}\lambda_2(x,y)$$

By Tonelli:

$$= \int_{x\in H}\int_{y\in\mathbb{R}} \frac{\int_{t\in C} f(x,t)\,\mathrm{d}t}{\int_{t\in\mathbb{R}} f(x,t)\,\mathrm{d}t} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

$$= \int_{x\in H}\int_{t\in C} f(x,t)\,\mathrm{d}t\,\mathrm{d}x$$

$$= (f\lambda_2)(H \times C) = P[(X,Y) \in H \times C] = P[X \in H, Y \in C]$$

$$= P[G \cap [Y \in C]]$$

So we have verified property 2. It is indeed a version.
Now we prove some general properties of conditional probability theory.
A key idea in undergrad probability is: if we condition on an event, we are reducing the probability space.

Theorem 20 (33.2)**.** :
i: If $P(A) = 0$ then $P(A|\mathscr{G}) = 0$ a.s.
ii: If $P(A) = 1$ then $P(A|\mathscr{G}) = 1$ a,s,
iii: If $\langle A_n; n \geq 1\rangle$ are disjoint, then $P(\bigcup_n A_n|\mathscr{G}) = \sum_n P(A_n|\mathscr{G})$
iv: $0 \leq P(A|\mathscr{G}) \leq 1$ a.s.
v: If $A \subseteq B$ then $P(A|\mathscr{G}) \leq P(B|\mathscr{G})$
vi: If $A_n \uparrow A$ or $A_n \downarrow A$ then $P(A_n|\mathscr{G}) \to P(A|\mathscr{G})$ a.s.

Proof. i: just check definition
ii: just check definition
iii: Let $A := \cup A_n$ then $\int_G \sum_n P(A_n|\mathscr{G})\,\mathrm{d}P = \sum_n \int_G P(A_n|\mathscr{G})\,\mathrm{d}P = \sum_n P(A_n \cap G) = P(A \cap G)$
iv: Let $G := [P(A|\mathscr{G}) < 0] \in \mathscr{G}$
$\int_G P(A|\mathscr{G})\,\mathrm{d}P = P(A \cap G) \geq 0$
Therefore $P(A|\mathscr{G}) \geq 0$ a.s.
For ≤ 1 just take $G := [P(A|\mathscr{G}) > 1]$ and do the same.
v: part iii with part iv
vi: by part v, the limit exists by v. We want that to be a version of A.
Check that it is a version of $P(A|\mathscr{G})$ by MCT or BCT and continuity of probability. □

Theorem 21 (33.1)**.** Let $\mathscr{G} = \sigma(\mathscr{P})$ where $\mathscr{P}$ is a π system, Ω a countable union of $\mathscr{P}$-sets. An integrable $f \in \mathscr{G}$ is a version of $P(A|\mathscr{G})$ if for all $G \in \mathscr{P}$ we have:

$$\int_G f\,\mathrm{d}P = P(A \cap G)$$

idea: we have a finite measure on each side and they agree on $\mathscr{P}$. Since they agree on $\mathscr{P}$ they agree on $\mathscr{G}$ on theorem 10.4.

Class 16: 03/05
The following theorem we state without proof.
We use for motivation an example. Suppose two random variable have joint density. Classical formula of density of y given x we can interpet as given any set of Y we integrate over that set for density of X.

Theorem 22 (33.3)**.** Suppose that $(\Omega, \mathscr{F}, P)$ is a probability space, and we have $\mathscr{G} \subseteq \mathscr{F}$ a sub σ field and $X : (\Omega, \mathscr{F}) \to (T, \mathscr{T})$ with $\mathscr{T}$ is the collection of borel sets of a complete seperable metric space with $\mathscr{T}$ the borel sets of complete, seperable metric space T. Then there exists $\mu : \sigma(X) \times \Omega \to [0, 1]$ we have:

1. $\forall \omega \in \Omega, \mu(\cdot, \omega)$ is a probability measure of $\sigma(X)_i$
2. $\forall A \in \sigma(X)$, $\mu(A, \cdot)$ is a version of $P(A|\mathscr{G})$.

We call such a μ a regular conditional distribution of X given $\mathscr{G}$
Example 33.5 was such an example with $T = \mathbb{R}$ for the law of y given x.
See ex. 33.12 for more on this.
If $(\Omega, \mathscr{F}) = (T, \mathscr{T})$ and $X = id$ then μ is called a regular conditional probability.

Proof. We oit the proof. It is done first for $T = \mathbb{R}$ as in Billingsley, and then extended by mapping T surjectively to $[0, 1]$. See Durrett. □

Conditional Expectation

Note that when $\mathscr{G} = \sigma(\mathscr{P})$ with $\mathscr{P}$ a countable partition, the formula:

$$P(A) = \sum_{B \in \mathscr{P}} P(A|B)P(B) = \int_\Omega P(A|\mathscr{G})\,\mathrm{d}P = E[P(A|\mathscr{G})]$$

This is a way to compute $P(A)$.

Likewise, we can compute expectation:

$$E[X] = E\left[\sum_{B\in\mathscr{P}} \mathbb{1}_B\right] = \sum_{B\in\mathscr{P}} E[X\mathbb{1}_B] = \sum_{B\in\mathscr{P}} E[X|B]P(B) = \int_\Omega E[X|\mathscr{G}]\,\mathrm{d}P = E[E[X|\mathscr{G}]]$$

Similarly, for $G \in \mathscr{G}$,

$$E[X\mathbb{1}_G] = E[X;G] = \int_G X\,\mathrm{d}P = \int_G E[X|\mathscr{G}]\,\mathrm{d}P$$

The case $X = \mathbb{1}_A$ is conditional probability.

Definition 11. Let $X \in L^1(\Omega, \mathscr{F}, \mathscr{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ be a σ field. A random variable $E[X|\mathscr{G}]$ such that:

1. $E[X|\mathscr{G}] \in L^1(\Omega, \mathscr{G}, P \restriction \mathscr{G})$
2. $\forall G \in \mathscr{G}$

   $$\int_G E[X|\mathscr{G}]\,\mathrm{d}P = \int_G X\,\mathrm{d}P$$

   is called a <u>version</u> <u>of the conditional expectation of X given $\mathscr{G}$</u>.

For $G = \Omega$ ii becomes $E[E[X|\mathscr{G}]] = E[X]$, In general, ii says:

$$E[X|\mathscr{G}](P \restriction \mathscr{G}) = (XP) \restriction \mathscr{G}$$

Theorem 23. Conditional expectations exist and are unique up to P-null modifications.

Proof. $E[X|\mathscr{G}]$ is the radon nikodym derivative of $(XP) \restriction \mathscr{G}$ w.r.t. $(P \restriction \mathscr{G})$. We prove that the conditions are satisfied.
Set $P_{\mathscr{G}} := P \restriction \mathscr{G}$
$\nu_X : G \mapsto \int_G X\,\mathrm{d}P$. Then $\nu_X \ll P_{\mathscr{G}}$. So the conditions are satisfied, and thus we're done. □

Example: $E[X|\{\varnothing, \Omega\}] = E[X]$
$E[X|\mathscr{F}] = X$
We can think about expectation to be the best guess of what X is. That is, it minimizes the square error.
Checking part ii might be complicated. It suffices to check on some special sets.

Theorem 24 (34.1)**.** Let $\mathscr{P} \subseteq \mathscr{F}$ be a π-system, $\mathscr{G} = \sigma(\mathscr{P})$ and $X \in L^1(\Omega, \mathscr{F}, \mathscr{P})$, $f \in L^1(\Omega, \mathscr{G}, P \restriction \mathscr{G})$. Then, $f = E[X|\mathscr{G}]$ almost surely if and only if for all $G \in \mathscr{P}$ we have $\int_G f\,\mathrm{d}P = \int_G X\,\mathrm{d}P$

Proof. Use 16.10 □

Theorem 25 (34.2)**.** Let $X, Y, X_n \in L^1(\Omega, \mathscr{F}, \mathscr{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ sigma field. We have the following properties:

1. $\forall a \in \mathbb{E}$ we have $E[a|\mathscr{G}] = a$ a.s.
2. $\forall a, b \in \mathbb{R}$ we have $E[aX + bY|\mathscr{G}] = aE[X|\mathscr{G}] + bE[Y|\mathscr{G}]$ a.s.
3. $X \leq Y$ a.s. implies $E[X|\mathscr{G}] \leq E[Y|\mathscr{G}]$
4. $|E[X|\mathscr{G}]| \leq E[|X||\mathscr{G}]$ a.s.
5. c-LDCT: $X_n \to X$ a.s., $|X_n| \leq Y$ a.s. implies $E[X_n|\mathscr{G}] \to E[X|\mathscr{G}]$a.s.

Proof. 1. obvious

2. check definition
3. $E[Y - X|\mathscr{G}] \geq 0$ by the positivity of the radon nikodym derivative.
4. We want $-E[|X||\mathscr{G}] \leq E[X|\mathscr{G}] \leq E[|X||\mathscr{G}]$ a.s., which follows from iii, ii and $-|X| \leq X \leq |X|$
5. Convert it to easier problem: define $Z_n := \sup_{k \geq n} |X_k - X|$, then Z_n goes to 0 a.s. monotonically. Since $|Z_n| \leq 2Y$ a.s., if we proved that $E[Z_n|\mathscr{G}] \to 0$ almost surelythen we may deduce $E[X_n|\mathscr{G}] \to E[X|\mathscr{G}]$ a.s. by iv. To prove the former, note that Z_n decreases as n increases, so by iii $E[Z_n|\mathscr{G}]$ is also decreasing. So, by MCT, we have some limit $Z \geq 0$. We want to show $Z = 0$ a.s. Consider $E[Z] \leq E[E[Z_n|\mathscr{G}]] = E[Z_n]$ and $E[Z_n] \to 0$ since $Z_n \to 0$ and $|Z_n| < 2Y$ so we can use dominated convergence.

□

Theorem 26 (34.3)**.** If X is $\mathscr{G}$-measurable and $Y, XY \in L^1(\Omega, \mathscr{F}, \mathscr{P})$ then $E[XY|\mathscr{G}] = XE[Y|\mathscr{G}]$ a.s.

Intuition: we know everything $\mathscr{G}$ tells us. Since X is $\mathscr{G}$ measurable, we treat X as a constant.

Proof. If $X = \mathbb{1}_H$ for $H \in \mathscr{G}$ then we must check that for all $G \in \mathscr{G}$

$$\int_G \mathbb{1}_H Y \, \mathrm{d}P = \int_G \mathbb{1}_H E[Y|\mathscr{G}] \, \mathrm{d}P$$

Which means

$$\int_{G \cap H} Y \, \mathrm{d}P = \int_{G \cap H} E[Y|\mathscr{G}] \, \mathrm{d}P$$

Which is true by the definition of conditional expectation.

□

Class 17: 03/07
We complete the proof.

Proof. We have proved for indicator function.
By theorem 34.2(ii), our equation holds whenever X is simple.
In general, take simple $X_n \in \mathscr{G}$ such that $|X_n| \leq |X|$ and $X_n \to X$. Then,

$$E[X_n Y|\mathscr{G}] = X_n \cdot E[Y|\mathscr{G}]$$

Now, by LDCT, [34.2(v)]

$$E[XY|\mathscr{G}] \overset{a.s.}{=} \lim_{n \to \infty} E[X_n Y|\mathscr{G}] \overset{a.s.}{=} \lim_{n \to \infty} X_n E[Y|\mathscr{G}] \overset{a.s.}{=} X \cdot E[Y|\mathscr{G}]$$

□

Theorem 27 (34.4)**.** Let $X \in L^1(\Omega, \mathscr{F}, \mathscr{P})$
Suppose $\mathscr{G}_1 \subseteq \mathscr{G}_2 \subseteq \mathscr{F}$ are finite σ-fields. Then,

$$E[E[X|\mathscr{G}_2]|\mathscr{G}_1] = E[X|\mathscr{G}_1] = E[E[X|\mathscr{G}_1]|\mathscr{G}_2]$$

a.s.
In the trivial sigma field this is just law of total expectation.

Proof. We prove the second equality first. Intuitively, since $\mathscr{G}_1 \subseteq \mathscr{G}_2$, if X is $\mathscr{G}_1$ measureable then it is $\mathscr{G}_2$ measureable. So we don't get any extra info, so expectation remains the same.

$$E[X|\mathscr{G}_1] \in \mathscr{G}_1 \subseteq \mathscr{G}_2$$

So it doesn't change when we compute its conditional expectation w.r.t. $\mathscr{G}_2$
Also,
$E[E[X|\mathscr{G}_2]|\mathscr{G}_1](P \restriction \mathscr{G}_1) = (E[X|\mathscr{G}_2]P) \restriction \mathscr{G}_1$
$= (E[X|\mathscr{G}_2](P \restriction \mathscr{G}_2) \restriction \mathscr{G}_1)$
$= ((XP) \restriction \mathscr{G}_2) \restriction \mathscr{G}_1$
$= (XP) \restriction \mathscr{G}_1$
$= E[X|\mathscr{G}_1](P \restriction \mathscr{G}_1)$
Therefore,

$$E[E[X|\mathscr{G}_2]|\mathscr{G}_1] = E[X|\mathscr{G}_1]$$

This is very commonly used. This is called the 'Tower Property'.

□

Theorem 28 (34.5)**.** Suppose that μ is a regular conditional probability for $P[\cdot|\mathscr{G}]$. Then, for $X \in L^1(P)$ for P-a.e.ω,

$$E[X|\mathscr{G}](\omega) = \int_{\omega'} X(\omega')\,\mathrm{d}\mu(\omega', \omega)$$

Proof. Suppose $X = \mathbb{1}_A$. Then $E[X|\mathscr{G}](\omega) = P(A|\mathscr{G})(\omega)$ and

$$\int X(\omega')\,\mathrm{d}\mu(\omega', \omega) = \mu(A, \omega)$$

Thus, these are equal by hypothesis.
Build from there by using theorem 34.2(ii),(v).

□

<u>Jensen's Inequality for Conditional Expectation</u>:

Theorem 29. Let $X \in L^1(P)$ and let ϕ be convex on an interval containing the range of X and $\phi \circ X \in L^1(P)$.
Then, for all $\mathscr{G} \subseteq \mathscr{F}$

$$\phi(E[X|\mathscr{G}]) \leq E[\phi \circ X|\mathscr{G}]$$

a.s.

Proof. φ is the pointwise sup of all linear functions that are $\leq \varphi$.
We can choose a countable class of such linear functions for which their sup is still φ
For linear function, the inequality we want is actually an equality.
Hence,

$$\varphi(E[X|\mathscr{G}]) = \sup_l l(E[X|\mathscr{G}]) \overset{a.s.}{=} \sup_l E[l(X)|\mathscr{G}] \leq E[\varphi(X)|\mathscr{G}]$$

□

Omit page 450 to end of section.

Martingales

We use notes.
Suppose we are gambling, and $X_n :=$ our fortune after n plays.
$\mathscr{F}_n := \sigma(X_1, \cdots, X_n)$ or more.
Suppose we have a fair game. Then, our expected fortune after the next gamble would be equal to our current fortune:
$X_n = E[X_{n+1}|\mathscr{F}_n]$ fair.
$X_n \geq E[X_{n+1}|\mathscr{F}_n]$ unfavorable
$X_n \leq E[X_{n+1}|\mathscr{F}_n]$ favorable
One way to be favorable: before we gamble we get handed some amount of money, and then we make a fair bet. Example: Casino gives you free chips.
The amount of money doesn't need to be definite, it can be random with positive expectation.

Definition 12. Let $X_n \in L^1(\Omega, \mathscr{F}, P)$ and $\mathscr{F}_n \subseteq \mathscr{F}$ be σ fields. We get more and more information, since the σ fields are monotonically increasing.
We call $\langle \mathscr{F}_n; n \geq 1\rangle$ a filtration if $\forall n$, $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$
We call $\langle X_n; n \geq 1\rangle$ adaped to $\langle \mathscr{F}_n\rangle$ if $\forall n, X_n \in \mathscr{F}_n$
We call $\langle (X_n, \mathscr{F}_n); n \geq 1\rangle$ a martingale (submartingale, supermartingale) if:
$\langle X_n\rangle$ is adapted to the filtration $\langle \mathscr{F}_n\rangle$ and for all n , $X_n = E[X_{n+1}|\mathscr{F}_n]$ a.s. $(\leq; \geq)$
So, we go up in submartingales and down in supermartingales.
The reason for this is submartingales correspond to subharmonic and supermartingales correspond to superharmonic functions.
We call $\langle X_n; n \geq 1\rangle$ a martingale if there exists some filtration w.r.t. it is a martingale.
If there exists such a filtration, $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ always works.
Suppose $X_n \in \mathscr{F}'_n$ where $\langle \mathscr{F}'_n\rangle$ is a filtration and $X_n = E[X_{n+1}|\mathscr{F}'_n]$ a.s.
We know $X_n \in \mathscr{F}_n \subseteq \mathscr{F}'_n$ thus,

$$E[X_{n+1}|\mathscr{F}_n] \overset{a.s.}{=} E[E[X_{n+1}|\mathscr{F}'_n]|\mathscr{F}_n] \overset{a.s.}{=} E[X_n|\mathscr{F}_n] \overset{a.s.}{=} X_n$$

In general, if $\langle (X_n, \mathscr{F}_n)\rangle$ is a martingale, then using tower property,

$$E[X_m|\mathscr{F}_n] = X_{m\wedge n} = X_{\min(m,n)}$$

Also, $E[X_n]$ is the same for all n.

Class 18: 03/19
Today we do examples of martingales.
A popular example is: sums of independent random variable.
Example 35.1 Suppose Y_n are independent random variables, $E[|Y_n|] < \infty$.
Then Y_n is the change in fortune after the n'th game [we're gambling], and we're looking at the cumulative change, aka partial sums.

$$X_n := \sum_{k=1}^{n} Y_k$$

These are neither martingales or submartingales or supermartingales without additional assumptions.
If $E[Y_n] = 0 \forall n$ then $E[X_{n+1}|\sigma(X_1, \cdots, X_n)]$
$= E[X_{n+1}|\sigma(Y_1, \cdots, Y_n)]$
$= E[X_n + Y_{n+1}|\sigma(Y_1, \cdots, Y_n)]$
$= X_n + E[Y_{n+1}|\sigma(Y_1, \cdots, Y_n)]$
They are independent and all have expectation 0. Thus this equals X_n a.s.
So this is a martingale.
Correspondingly, if the expectations were all non-negative, we would have:
$X_n + E[Y_{n+1}|\sigma(Y_1, \cdots, Y_n)] \geq X_n$ so this is a submartingale.
Correspondingly, non-positive would give us a supermartingale.
Another different example. Significant for the theory.
Example 35.5 If $Z \in L^1(\Omega, \mathscr{F}, P)$ and $\langle \mathscr{F}_n\rangle$ is a filtration. Then,

$$\langle (E[Z|\mathscr{F}_n], \mathscr{F}_n)\rangle$$

is a martingale by the tower property:

$$E[E[Z|\mathscr{F}_{n+1}]|\mathscr{F}_n] = E[Z|\mathscr{F}_n]$$

Example 35.2 On $(\Omega, \mathscr{F})$, let P be a probability measure and ν be a finite, signed measure. Let $\langle \mathscr{F}_n \rangle$ be a filtration.
We can only talk about Radon-Nikodym derivative if we have absolute continuity. Suppose that $\forall n, \nu \restriction \mathscr{F}_n \ll P \restriction \mathscr{F}_n$. Let

$$X_n := \frac{\mathrm{d}(\nu \restriction \mathscr{F}_n)}{\mathrm{d}(P \restriction \mathscr{F}_n)}$$

Then $\langle (X_n, \mathscr{F}_n)\rangle$ is a martingale:
$E[X_{n+1}|\mathscr{F}_n](P \restriction \mathscr{F}_n) = (X_{n+1}P) \restriction \mathscr{F}_n = (X_{n+1}P \restriction \mathscr{F}_{n+1}) \restriction \mathscr{F}_n$
$= (\nu \restriction \mathscr{F}_{n+1}) \restriction \mathscr{F}_n = \nu \restriction \mathscr{F}_n = X_n(P \restriction \mathscr{F}_n)$
Exercise 35.3 Let P be Lebesgue measure on $(0, 1]$ and $\mathscr{F}_n := \sigma(I_k^{(n)}; 0 \le k \le 2^n)$ where $I_k^{(n)} := (k2^{-n}, (k+1)2^{-n}]$. Since $(A \in \mathscr{F}_n, P(A) = 0) \implies A = \varnothing$ we have for all $\nu, \nu \restriction \mathscr{F}_n \ll P \restriction \mathscr{F}_n$. In this case,

$$X_n = \sum_{k=0}^{2^n-1} \mathbb{1}_{I_k^{(n)}} \frac{\nu(I_k^{(n)})}{P(I_k^{(n)})} = \sum_{k=0}^{2^n-1} \mathbb{1}_{I_k^{(n)}} \nu(I_k^{(n)}) 2^n$$

Exercise 35.8 If $\langle (X_n, \mathscr{F}_n)\rangle$ is a martingle, then $\langle |X_n|, \mathscr{F}_n\rangle$ is a submartingale.

Theorem 30 (35.1)**.** Let $\langle X_n \rangle \subseteq L^1(P)$ adapted to a filtration $\langle \mathscr{F}_n \rangle$, and let φ be convex on an interval containing all the ranges of X_n, and let $\varphi \circ X_n \in L^1(P)$. Then $\langle (\phi \circ X_n, \mathscr{F}_n)\rangle$ is a submartingale if either:

1. $\langle (X_n, F_n)\rangle$ is a martingale
2. is a submartingale and φ is increasing
3. is a supermartingale and φ is decreasing

Proof. In all cases, we have:

$$\varphi(X_n) \le \varphi(E[X_{n+1}|\mathscr{F}_n])$$

by using monotonicity and convexity.
By Jensen, this is $\le E[\varphi(X_{n+1})|\mathscr{F}_n]$ $\square$

Now we depart from the book.

Wald's Equation

This is about expectation of adding up random variables but the number of random variables is itself random.

Theorem 31 (Wald's Equation)**.** Let $Z_n \in L^1(P)$ for $n \ge 1$ and let τ be an $\mathbb{N}$-valued random variable, and $\mu \in \mathbb{R}$.
Suppose that,

1. $\forall n \ge 1, P[\tau \ge n] > 0 \implies E[Z_n|\tau \ge n] = \mu$ [or $\le$ or $\ge$]
2. one of the following holds:
   (a) $\forall n, Z_n \ge 0$
   (b) $\sup_{n; P[\tau \ge n] > 0} E[|Z_n||\tau \ge n] < \infty$ and $E[\tau] < \infty$
   (c) $E[|\sum_{n=1}^{\tau} Z_n|] < \infty$ and $\lim_{n\to\infty} E[\sum_{k=1}^{n} Z_k \mathbb{1}_{[\tau > n]}] = 0$

then $E[\sum_{n=1}^{\tau} Z_n] = \mu E[\tau]$ [or $\leq$ or $\geq$]
where $0 \cdot \infty = 0$

Proof. In case a, we have
$E[\sum_{n=1}^{\tau} Z_n] = E[\sum_{n=1}^{\infty} Z_n \mathbb{1}_{[\tau \geq n]}] \overset{\text{tonelli/MCT}}{=} \sum_{n=1}^{\infty} E[Z_n \mathbb{1}_{[\tau \geq n]}] = \sum_{n=1}^{\infty} E[Z_n | \tau \geq n] P[\tau \geq n] = \sum_{n=1}^{\infty} \mu P[\tau \geq n] = \mu E[\tau]$.
This is equation N1.
In case b, we have
$E[\sum_{n=1}^{\infty} |Z_n| \mathbb{1}_{[\tau \geq n]}] = \sum_{n=1}^{\infty} E[|Z_n| | \tau \geq n] P[\tau \geq n] \leq \sup E[\cdots] \cdot \sum P[\tau \geq n] = \sup E[\cdots] \cdot E[\tau] < \infty$
Now we do the previous calculation with Fubini.
In case c, we have
$E[\sum_{k=1}^{\tau} Z_k] \overset{\text{LDCT}}{=} \lim_{n\to\infty} E[\sum_{k=1}^{\tau} Z_k \mathbb{1}_{[\tau \leq n]}]$
$= \lim_{n\to\infty} E[\sum_{k=1}^{n} Z_k \mathbb{1}_{k \leq \tau \leq n}]$
$= \lim_{n\to\infty} E[\sum_{k=1}^{n} Z_k (\mathbb{1}_{[\tau \geq k]} - \mathbb{1}_{[\tau > n]})]$
by the second condition of part c,
$= \lim_{n\to\infty} E[\sum_{k=1}^{n} Z_k \mathbb{1}_{[\tau \geq k]}] = \lim_{n\to\infty} \sum_{k=1}^{n} \mu P[\tau \geq k] = \mu E[\tau]$

□

Stopping times will be important.
Make sure you are familiar with homework!!!

Class 18 and 19 skipped

Class 20: 04/02
We are betting on whether a card is red. We have 52 cards, 26 red, 26 black and we keep seeing one by one. Assume uniform shuffle.
$A_k :=$ event that the k'th card is red.
$\tau :=$ the time k we bet.
$\mathscr{F}_k := \sigma(A_1, \cdots, A_{k-1})$
τ is a stopping time.
$[\tau = k] \in \mathscr{F}_k$
At time k the chance of winning would be:
$P(A_k \mid \mathscr{F}_k) =: X_k$
X_k is between 0 and 1
The chance of winning for stopping time τ is X_τ
X_τ is random. Our actual chance is $E[X_\tau]$
We check if X_k is a martingale:
$E[X_{k+1} \mid \mathscr{F}_k] = E[P(A_{k+1} \mid \mathscr{F}_{k+1}) \mid \mathscr{F}_k]$
By tower property, $= P(A_{k+1} \mid \mathscr{F}_k)$
We have information about $k-1$ cards. We want to know if card number $k+1$ is red. It is the same as the k'th card being red. So this is equal to X_k. For $k \leq 51$.
This is a finite martingale!
$\langle X_k; 1 \leq k \leq 52 \rangle$ is a martingale.
So, $E[X_\tau] = E[X_1] = \frac{1}{2}$
Just read notes for this section.

Class 21: 04/04
...
Class 22: 04/09
We did Doob's Maximal Inequality.

Chapter 7: Stochastic Processes

Recall a stochastic process is a collection of random variables on the same space indexed by some set, T.
For example, we can think about $N((s,t])$ where T contains intervals.

In higher dimension, we can have half open boxes, or even the set of borel sets as the index set.

Section 36: Kolmogorov's Existence/Consistency Theorem

$P[(X_{k_1}, \cdots, X_{t_k}) \in H]$ for $t_1, \cdots, t_k \in T, H \in \mathcal{R}^k$ are the finite dimensional distributions/marginals (f.d.d.) of the process.
$\langle X_k; t \in T\rangle \in \mathbb{R}^T =^T \mathbb{R}$
For $t \in T$ write $Z_t = \mathbb{R}^t \to \mathbb{R}$ by $x \mapsto x(t) = x_k$
For $S \subseteq T$ set $\mathcal{R}_T^S \coloneqq \sigma(Z_t : t \in S)$. This is not a sigma field on $\mathcal{R}^S$, rather it is a sigma field on $\mathcal{R}^T$
These are sub-σ-fields of $\mathcal{R}_T^T$
The f.d.d.'s are the probability measures μ_F on $\mathcal{R}_T^F$ for finite $F \subseteq T$.
We can think of $\mathcal{R}_T^F$ as

$$\{\underbrace{A \times \mathbb{R}^{T\setminus F}}_{\in \mathcal{R}_T^T}; A \in \mathcal{R}_F^F\}$$

These satisfy the consistency condition:

$$F_1 \subseteq F_2 \implies \mu_{F_2} \restriction \mathcal{R}_T^{F_1} = \mu_{F_1} \quad (*)$$

Thus, $\langle Z_k; t \in T\rangle$ is a stochasting process w.f.d.d.'s $\langle \mu_F; F \subseteq T \text{ finite}\rangle$

Proof. Define $\mathcal{R}_0^T \coloneqq \bigcup\{\mathcal{R}_T^F; F \subseteq T, F \text{ finite}\}$.
Sets in $\mathcal{R}_0^T$ are called finite-dimensional or cylinders. This is a field since given F_1 and F_2

$$\mathcal{R}_T^{F_1} \cup \mathcal{R}_T^{F_2} \subseteq \mathcal{R}_T^{F_1 \cup F_2} \quad (**)$$

Thus, the plan is to use theorem 3.1:
Define $\mathcal{P}$ on $\mathcal{R}_0^T$ and show that $\mathcal{P}$ is countably additive there.
To define $\mathcal{P}$ on $\mathcal{R}_0^T$, set $P(A) \coloneqq \mu_F(A)$ for any F with $A \in \mathcal{R}_T^F$. By $(*)$ this is well defined. Also, since μ_F is a probability measure, P is finitely additive on $\mathcal{R}_T^F$, hence on $\mathcal{R}_0^T$. by $(**)$
To show countable additivity, recall from example 2.10 that it suffices to show for $A_n \in \mathcal{R}_0^T$ with $A_n \downarrow \varnothing$, we have $P(A_n) \to 0$.
Equivalently, if $A_n \downarrow A$ and $P(A_n) \geq \epsilon > 0$ then $A \neq \varnothing$. Now, eah $A_n \in \mathcal{R}_T^{F_n}$ for some finite F_n. By regularity (thm 12.3) there exists compact $K_n \subseteq \mathbb{R}^{F_n}$ such that $K_n \times \mathbb{R}^{T\setminus F_n} \subseteq A_n$ and $\mu_{F_n}(A_n \setminus (K_n \times \mathbb{R}^{T\setminus F_n})) < \epsilon/2^{n+1}$. Then $A \supseteq \bigcap_n (K_n \times \mathbb{R}^{T\setminus F_n})$.
We claim that this is non $\varnothing$.
FIrst, note that $A_n \setminus (K_n \times \mathbb{R}^{T\setminus F_n})$
$= \bigcup_{n\leq N}(A_N \setminus (K_n \times \mathbb{R}^{T\setminus F_n}))$
$\leq \bigcup_{n\leq N}(A_n \setminus (K_n \times \mathbb{R}^{T\setminus F_n}))$
has probability $< \sum_{n\leq N} \epsilon/2^{n+1} < \epsilon/2$
whence $P(\bigcap(K_n \times \mathbb{R}^{T\setminus F_n})) > \epsilon/2$
Thus, $\forall N \geq 1 \exists x^{(N)} \restriction F_n \in K_n$ for $1 \leq n \leq N$. Let $\langle N_1(j); j \geq 1\rangle$ be a subsequence such that $x^{(N,(j))} \restriction F$ covergence. Recursively, choose $\langle N_{m+1}(j) l j \geq 1\rangle$ to e a subsequence of $\langle N_m(j)\rangle$ such that $x^{N_{m+1}(j)} \restriction F_{m+1}$ converges. Then $x^{(N_m(m))} \restriction F_n$ converges for all n.
Define $x(t) \coloneqq x^{N_m(m)}(t)$ if $t \in \cup_n F_n$ and 0 otherwise. Then $x \in A$ so $A \neq \varnothing$.

□

Class 23: 04/11
If $X : \Omega \to \Omega'$ and $\mathscr{A} \subseteq 2^{\Omega'}$ then $\sigma(X^{-1}\mathscr{A}) = X^{-1}(\sigma(\mathscr{A}))$
The relevant operations of σ fields (intersection union complements) commutes with pre-image so this is expected.

Thus, given $X_t : \Omega \to \mathbb{R}$ $[t \in T]$ stochastic process, let $X : \Omega \to \mathbb{R}^T$ be $\langle X_t; t \in T\rangle$
Claim: $\sigma(X) = \sigma(\{X_t; t \in T\})$
To see this, note first that:

$$\begin{tikzcd} \Omega \arrow[r, "X"] \arrow[dr, "X_t"'] & \mathbb{R}^T \arrow[d, "Z_t"] \\ & \mathbb{R} \end{tikzcd}$$

So $X_t = Z_t \circ X$
Thus, $\sigma(\{X_t; t \in T\}) = \sigma(\{X_t^{-1}\mathcal{R}; t \in T\})$
$= \sigma(X^{-1}(\{Z_t^{-1}\mathcal{R}\}))$
$= X^{-1}(\sigma(\{Z_t^{-1}\mathcal{R}\}))$
$= X^{-1}\mathcal{R}_T^T$ by definition of $\mathcal{R}_T^T$
$= \sigma(X)$
Note also that $\bigcup\{\sigma(\{X_t; t \in S\}); S \subseteq T, S \text{ countable}\}$ is a σ-field.
So, it equals $\sigma(X)$
That is, every set in $\sigma(X)$ depends on only countably many coordinates!
Consider $T = [0, \infty), \Omega = \mathbb{R}^T, X_t \coloneqq Z_t$.
Consider the class of continuous functions $C(T)$.
$C(T)$ can't depend on only a countable set since knowing the values on a countable set doesn't tell us whether something is continuous.
So, $C(T)$ is not measurable here!
This means this space is not good enough to model brownian motion.

37 - Brownian Motion

We consider brownian motion on only one dimension. For bigger dimension we can take independent brownian motion in perpendicular direction.
Consider $\langle W_t; t \in [0, \infty)\rangle$ a stochastic process that has independent stationary increments and continuous sample paths.
W stands for Wiener.
W_t is uniformly continuous on $[0, 1]$ so:

$$H_n \coloneqq \sup_{1 \le k \le n} \left| W\left(\frac{k}{n}\right) - W\left(\frac{k-1}{n}\right)\right|$$

Which $\to 0$ as $n \to \infty$
Hence $\forall \delta > 0, P[H_n \ge \delta] \to 0$
Now, $P[H_n \ge \delta] = 1 - P[H_n < \delta]$
$= 1 - \prod_{k=1}^{n} P[|W(\frac{k}{n}) - W(\frac{k-1}{n})| < \delta] = 1 - P[|W(\frac{1}{n}) - W(0)| < \delta]^n$
$= 1 - \{1 - P[|W(\frac{1}{n}) - W(0)| \ge \delta]\}^n \ge 1 - e^{-n \cdot P[|W(\frac{1}{n}) - W(0)| \ge \delta]}$
≥ 0.
So, $1 - e^{-n \cdot P[|W(\frac{1}{n}) - W(0)| \ge \delta]} \to 0$
Which implies $P[|W(\frac{1}{n}) - W(0)| \ge \delta] \to 0$
$(*)$ $\forall \delta > 0, \lim_{h \downarrow 0} \frac{P[|W(h) - W(0)| \ge \delta]}{h} = 0$
This implies that $(**) \exists \mu \in \mathbb{R} \exists \sigma \ge 0 \forall t \ge 0$,
$W(t) - W(0) \sim \mathcal{N}(\mu t, \sigma^2 t)$

Theorem 32. For a stochastic process with independent stationary increment, $(*) \iff (**)$

Why should such a process exist?
Let $\langle Y_n\rangle$ be symmetric ± 1 i.i.d. steps, $\Delta x > 0, \theta > 0$

$$D(t) \coloneqq \sum_{k=1}^{\lfloor t/\Delta t \rfloor} \delta x \cdot Y_k$$

Then $Var(D(t)) = (\Delta x)^2 \lfloor \frac{t}{\Delta t} \rfloor$

So, if $D(t)$ 'converges' to W_t we should have $\frac{(\Delta x)^2}{\Delta t} \to 1$
So take $\Delta t := \frac{1}{n}, \Delta x = \frac{1}{\sqrt{n}}$
Let $D_n(t)$ be the corresponding process of partial sums:

$$D_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} Y_k$$

So, $D_n(t)$ converges weakly to normal.

Class 24: 04/16

$$D(t) = \sum_{k=1}^{\lfloor t/\Delta t \rfloor} \Delta x \cdot y_k$$

$$\frac{(\Delta x)^2}{\Delta t} \to \sigma^2$$

Definition 13. A <u>Brownian Motion</u> (BM) or <u>Wiener Process</u> with <u>drift</u> μ and <u>variance parameter</u> σ^2 is a stochastic process $\langle W_t; t \geq 0 \rangle$ such that:
i: If $0 \leq t_0 < t_1 < \cdots < t_k$ then $\langle W_{t_i} - W_{t_{i-1}}, 1 \leq i \leq k \rangle \sim \mathcal{N}(\mu \langle t_i - t_{i-1}; 1 \leq i \leq k \rangle, \sigma^2 \cdot \operatorname{diag}(\langle t_i - t_{i-1}; 1 \leq i \leq k \rangle))$
And W_0 is independent of $\sigma(W_t - W_0; t \geq 0)$
ii: For every $\omega, t \mapsto W_t(\omega)$ is continuous.
If $\mu = 0, \sigma = 1, W_0 \equiv 0$ then the process is <u>standard</u> <u>B.M.</u> or just <u>B.M.</u>

Existence by Kolmogorov
Question: What are the f.d.d.'s?
Assume $W_0 \equiv 0$ and $t_0 = 0$
Let M b the linear transformation that takes $\langle y_1, y_2 - y_1, \cdots, y_k - y_{k-1} \rangle$ to $\langle y_1, y_2, \cdots, y_k \rangle$

$$M = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

Recall that if $Y \sim \mathcal{N}(c, \Sigma)$ then $MY \sim \mathcal{N}(Mc, M\Sigma M')$
Use this with $Y = \langle W_{t_1} - W_{t_0}, \cdots, W_{t_k} - W_{t_{k-1}} \rangle$ and $c = \mu \langle t_1 - t_0, \cdots, t_k - t_{k-1} \rangle$
$\Sigma = \sigma^2 \operatorname{diag}(\langle t_1 - t_0, \cdots, t_k - t_{k-1} \rangle)$
Then $MY = \langle W_{t_1}, \cdots, W_{t_k} \rangle$
$Mc = \mu \langle t_1, \cdots, t_k \rangle$
And the covariances are: for $0 \leq s \leq t$
$\operatorname{Cov}(W_s, W_t) = E[(W_s - \mu s)(W_t - \mu t)]$
We use independent increments to calculate it. $W_t - \mu t = W_s - \mu s + (W_t - W_s + \mu(t-s))$
So, $E[(W_s - \mu s)(W_t - \mu t)] = \operatorname{Var}(W_s) + 0 = \sigma^2 s$
Thus, $M\Sigma M' = \sigma^2 (\min\{t_i, t_j\})_{1 \leq i,j \leq k}$
These f.d.d.'s are consistent so Kolmogorov's theorem gives a process that satisfies (i).
If $W_0 \not\equiv 0$ then take X independent of the process $\langle W_t; t \geq 0 \rangle$ constructed by Kolmogorov's theorem with X having law of W_0 and use a new process $\langle X + \tilde{W}_t; t \geq 0 \rangle$
We now modify the process given by Kolmogorov's theorem to ensure continuity. We'll do this for standard brownian motion and then show how to
Let $D := \{k 2^{-n}; n, k \in \mathbb{N}\}$
Claim: $\forall \delta > 0, \forall \alpha > 0$

$$P[\sup\{|W(r\delta)|; r \in [0,1] \cap D\} > \alpha] \leq \frac{3\delta^2}{\alpha^4}$$

It suffices to prove this with D replaced by $D_n := \{k 2^{-n}; k \in \mathbb{N}\}$ since these eents are increasing in n

Now, $\langle W(r\delta); r \in [0,1] \cap D_n \rangle$
Forms the partial sums of i.i.d. mean 0 random variables, hence, a martingale.
Therefore, $\langle W(r\delta)^4 \rangle$ is a submartingale.
Applying Doob's inequality (thm 35.3),

$$P[\sup_{r\in[0,1]\cap D_n} |W(r\delta)| > \alpha] = P[\sup W(r\delta)^4 > \alpha^4] \leq \frac{E[W(\delta)^4]}{\alpha^4}$$

Note, $W(\delta) \sim \mathcal{N}(0,\delta) = \sqrt{\delta} N(0,1)$
So, fourth moment is 3
Thus, our probability is bounded by $\frac{3\delta^2}{\alpha^4}$
Let $I_{n,k} \coloneqq [k2^{-n}, (k+1)2^{-n}]$
$M_{n,k} = \sup\{|W(r) - W(k2^{-n})|; r \in I_{n,k} \cap D\}$,
$M_n \coloneqq \max\{M_{n,k}; 0 \leq k < n2^n\}$
Claim: $P[M_n > \frac{1}{n}] \leq \frac{3n^5}{2^n}$
For we have by stationarity of moments that for $\delta = 2^{-n}$,

$$P[M_{n,k} > \frac{1}{n}] = P[\sup_{r\in[0,1]\cap D} |W(r\delta)| > \frac{1}{n}] \leq \frac{3n^4}{2^{2n}}$$

THerefore,

$$P[M_{n,k} > \frac{1}{n}] \leq \sum_{k=0}^{n2^n-1} P[M_{nk} > \frac{1}{n} > \frac{1}{n}] \leq n2^n \cdot \frac{3n^4}{2^{2n}} = \frac{3n^5}{2^n}$$

Therefore, $\sum_n P[M_n > \frac{1}{n}] < \infty$
So we can use Borel Cantelli
Let $B \coloneqq \{\omega; M_n(\omega) > \frac{1}{n} i.o.\}$
Then $P(B) = 0$ by Borel Cantelli
Claim: $\forall t > 0, \forall \omega \notin B, W(r,\omega)$ is uniformly continuous in $r \in [0,t] \cap D$
For let $t > 0, \omega \notin B, \epsilon > 0$. Choose n such that $n > t, n > \frac{3}{\epsilon}, M_n(\omega) \leq \frac{1}{n}$. Set $\delta \coloneqq 2^{-n}$.
Let $r, r' \in [0,t] \cap D$ with $|r - r'| < \delta$.
Then $\exists k \in [0, n2^n]$ such that $r \in I_{n,k}$ and $r' \in I_{n,k+1}$ so that

$$|W(r,\omega) - W(r',\omega)| \leq |W(r,\omega) - W(k2^{-n},\omega)| + |W(k2^{-n},\omega) - W((k+1)2^{-n},\omega)|$$

$$+|W((k+1)2^{-n} - W(r',\omega))| \leq 2M_{nk}(\omega) + M_{n,k+1}(\omega) \leq 3M_n(\omega) \leq \frac{3}{n} < \epsilon$$

Class 25: 04/18
$W(t), t \in D$
unifomrly continuous on $[0,t] \cap D$ for all $t > 0$
off $B, P(B) = 0$
Now, define $W'_t(\omega)$ to be 0 if $\omega \in B$ and to be $\lim W_r(\omega)$ as $r \to t$ with $r \in D$ for $\omega \in B$:
This is because of the cauchy property. This holds by uniform contunity.
Also, $W'_t(\omega)$ is continuous int for all ω (if t, t' are close then $\exists r$ close to t and r' close t' such that $W_r(\omega)$ is close to $W_t(\omega)$ and $W_{r'}(\omega)$ is close to $w_{t'}(\omega)$ but r, r' are close so $W_r(\omega), W_{r'}(\omega)$ are close) and $W'_t(\omega) = W_t(\omega)$ for $t \in D$ for $\omega \notin B$
Finally, we claim that $\langle W'_t; t \geq 0\rangle$ has the same fdd's as $\langle W_t; t \geq 0\rangle$. Given $t_1, \cdots, t_k \in [0,\infty)$ let $r_i^{(n)} \to t_i, r_i^{(n)} \in D$
Then $\langle W_{r_i(n)} \rangle \to \langle W'_{t_i} \rangle$ everewhere.
Since we have pointwise convergence we also have weak convergence.
The covariance matrices $[\min\{r_i^{(n)}, r_j^{(n)}\}]_{i,j}$ converges to $[\min\{t_i, t_j\}]_{i,j}$

Theorem 33 (37.1)**.** Standard B.M. exists.

$C([0,\infty),\mathbb{R})$
To get a general B.M. with any W_0, μ, σ let $\langle W'_t\rangle$ be a std B.M. and let X be a random variable independent of $\langle W'_t\rangle$ with X having the law of the desired W_0. Set $W_t := X + \mu t + \sigma W'_t$.
$W_t - W_s = \mu(t-s) + \sigma(W'_t - W'_s) \sim \mathcal{N}(\mu(t-s), \sigma^2(t-s))$

Symmetries of BM, or new BMs from old

Let $\langle W_t\rangle$ be a std BM. Then so is $\langle -W_t; t \geq 0\rangle$ (space reversal). Also, given $t_0 > 0, \langle W_{t_0-t} - W_{t_0}\rangle_{0\leq t\leq t_0}$ has the law of BM on $[0,t_0]$ [time reversal].
In addition, $\langle W_{t_0+t} - W_{t_0}\rangle_{t\geq 0}$ is BM [time transition]
If $c > 0$ then $\langle W'_t := \frac{1}{c} W_{c^2 t}\rangle_{t\geq 0}$ is B.M. [scale invariance]
$\frac{1}{c^2}\min\{c^2 s, c^2 t\}$
Suppose $\frac{W(t_0+\Delta t_0) - W(t_0)}{\Delta t_0} > \epsilon$
$(c\{W'\left(\frac{t_0}{c^2} + \frac{\Delta t_0}{c^2}\right) - W'(\frac{t_0}{c^2})\}/\Delta t_0)$
Let $t_1 := t_0/c^2, \Delta t_1 = \Delta t_0/c^2$

$$\frac{W'(t_1+\Delta t_1) - W'(t_1)}{\Delta t_1} > c\epsilon$$

Taking c large, we see that there are chords of arbitrarily large slope arbitrarily close to 0

Theorem 34 (37.3)**.** On a set of probability 1, $\forall t$,

$$\limsup_{s\downarrow t}\left|\frac{W_s - W_t}{s-t}\right| = +\infty$$

$$\limsup_{s\uparrow t}\left|\frac{W_s - W_t}{s-t}\right| = +\infty$$

Proof. By symmetry enough to prove the first for $t \in [0,1)$
Fix $c > 0$ and let

$$A_n := \left[\exists t \in [0,1)\forall s \in (t, t+\frac{4}{n}), \left|\frac{W_s - W_t}{s-t} \leq c\right|\right]$$

It suffices to show there exists event A'_n containing A_n and $P(A'_n) \implies 0$
If $\omega \in A_n$ and t is a witness (that $\omega \in A_n$), let k be such that $\frac{k-1}{n} \leq t < \frac{k}{n}$
We compare $W\left(\frac{k+j}{n}\right)$ to $W(t)$ for $j = 0,1,2,3$ to get $|W(\frac{k+j}{n}) - W(t)| \leq C\left|\frac{k+j}{n} - t\right| \leq \frac{4C}{n}$
whence $\left|\frac{W(k+j+1)}{n} - W(\frac{k+j}{n})\right| \leq \frac{8C}{n}$ for $j = 0,1,2$
On the other hand,
$P[|W(\frac{k+j+1}{n}) - W(\frac{k+j}{n})| \leq \frac{8C}{n}] = P[|W(\frac{1}{n})| \leq \frac{8C}{n}]$
$= P[|W(1)| \leq \frac{8C}{\sqrt{n}}] \leq \frac{16C}{\sqrt{2\pi n}}$
since the standard normal density is $\leq \frac{1}{\sqrt{2\pi}}$
Therefore, $A_n \subseteq A'_n := \left[\exists k \in [1,n] s.t. \forall j = 0,1,2, |W(\frac{k+j+1}{n}) - W(\frac{k+j}{n})| \leq \frac{8C}{n}\right]$
with $P(A'_n) \leq n\left(\frac{16C}{\sqrt{2\pi n}}\right)^3 \to 0$

□

Class 26: 04/23
Let $X(t) = \frac{W(t)}{t}$ $[t > 0]$
Then $\langle X(t); t > 0\rangle$ is a <u>Gaussian Process</u> (i.e. all f.d.d.s are multivariate normal).
Covariances: $E[X(s)X(t)] = \frac{1}{st}E[W(s)W(t)] = \frac{1}{st}(s \wedge t) = \frac{1}{t} \wedge \frac{1}{s}$
Thus, if $W''(t) := X(\frac{1}{t})$ then $\langle W''(t); t > 0\rangle$ has f.d.d.'s of B.M.
<u>Claim</u>: If we define $W''(0) := 0$ then W'' has the law of B.M.

IE we want $\lim_{t\downarrow 0} W''(t) = 0$ a.s.
To see this, note that W'' has same f.d.d.s as BM on $[0,\infty)$. In particular, on dyadic rationals.
Therefore, W'' is uniformly continuous on the dyadic rationals a.s.
So if we complete W'' by continuity, we get B.M. But W'' is already continuous on $(0,\infty)$ and has the right value at 0
Note that, $W''(t) := \begin{cases} tW(\frac{t}{t}), & \text{if } t > 0; \\ 0, & \text{if } t = 0; \end{cases}$
this is called time inversion
Claim:

$$\limsup_{n\to\infty} W_n = \infty$$

$$\liminf_{n\to\infty} W_n = -\infty$$

Both a.s.
To see this, note that $W_n = \sum_{k=1}^{n}(W_k - W_{k-1})$ with $W_k - W_{k-1}$ being i.i.d. $N(0,1)$.
Each of the events in question belong to the tail σ-field of $\langle W_k - W_{k-1}; k \geq 1\rangle$
By Kolmogorov's 0-1 law, each has probability 0 or 1
Both probabilities are equal by space symmetry.
Case 1: both 0. Means $\limsup_{n\to\infty}, \liminf_{n\to\infty}$ both finite. Meaning W_n is bounded above and below. Then $|W_n|$ is bounded.
But $P[|W_n| < n^{\frac{1}{4}}] = P[\sqrt{n}|W_1| \leq n^{\frac{1}{4}}] = P[|W_1| \leq n^{-\frac{1}{4}}] \to 0$ as $n \to \infty$
So, probability is actually 1
Thus, $\langle W_t\rangle$ changes sign i.o. as $t \to \infty$
Time inversion tells us $\langle W_t\rangle$ changes sign i.o. as $t \downarrow 0$
In particular, the zero set $\{t; W_t = 0\}$ has 0 as a limit point.

Theorem 35 (37.4)**.** The zero set of B.M. is almost surely perfect [every point of it is a limit point], has lebesgue measure 0 and is unbounded.

Proof. Let $\mathscr{Z}(\omega) := \{t; W_t(\omega) = 0\}$
Let λ be lebesgue measure. Then,

$$\int \lambda(\mathscr{Z}(\omega))\,\mathrm{d}P(\omega) = \int\int \mathbb{1}_A(t,\omega)\,\mathrm{d}\lambda(t)\,\mathrm{d}P(\omega)$$

Where $A := \{(t,\omega); W_t(\omega) = 0\}$
If we show that A is measurable ($\mathcal{R}^1 \times \mathscr{F}$) then we can use Fubini to get:

$$\int\int \mathbb{1}_A(t,\omega)\,\mathrm{d}P(\omega)\,\mathrm{d}\lambda(t)$$

$$= \int_0^\infty \underbrace{P\{\omega; W_t(\omega) = 0\}}_{=0}\,\mathrm{d}\lambda(t) = 0$$

So we only nead to show measurability of A
In fact, we will show (thm 37.2) that $(t,\omega) \mapsto W_t(\omega)$ is measurable.
Now we show that $\mathscr{Z}(\omega)$ is perfect.
The idea is to treat $t \in \mathscr{Z}(\omega)$ the start of a 'new' brownian motion.
We know about time 0.
Now this can't be exactly right, since $\exists t \in \mathscr{Z}(\omega)$ such that $\forall\epsilon > 0$ with $(t, t+\epsilon) \cap \mathscr{Z}(\omega) = \varnothing$. But, apparently, such t are limits of points $< t$ in $\mathscr{Z}(\omega)$.
Indeed, when t is NOT a limit point from below, $\exists r \in \mathbb{Q}^+$ such that $(r,t)\cap\mathscr{Z}(\omega) = \varnothing$ i.e. t is the first zero after r.
So, for $r \in \mathbb{Q}^+$ let $\tau(\omega) := \tau_r(\omega) := \inf\{t \geq r; W_t(\omega) = 0\}$
τ is a random variable since we may write $\{\omega; \tau(\omega) \leq t\} = \{\omega; \inf_{s\in[r,t]\cap\mathbb{Q}} |W_s(\omega)| = 0\}$ $\in \sigma(W_s; s \in \mathbb{Q}, s \leq t) \subseteq \sigma(W_s; s \leq t)$
Thus, τ is a stopping time ($\tau \geq 0$ and $[\tau \leq t] \in \sigma(W_s; s \leq t)$)

Define $W_t^*(\omega) := W_{\tau(\omega)+t}(\omega) - W_{\tau(\omega)}(\omega) = W_{\tau(\omega)+t}(\omega)$ for $t \geq 0$
We will show (theorem 37.5) that $\langle W_t^*; t \geq 0\rangle$ is a B.M.
Hence τ_r is a.s. a limit point of $\mathscr{Z}$
Now, every point of $\mathscr{Z}$ that is not a limit point from below in $\mathscr{Z}$ is equal to τ_r for some $r \in \mathbb{Q}^+$
Hence it is a limit of $\mathscr{Z}$ too.
Let B_r be a set of pr. 0 such that $\forall \omega \notin B_r$, $\tau_r(\omega)$ is a limit point of $\mathscr{Z}(\omega)$.
Then $\forall \omega \notin \bigcup_{r\in\mathbb{Q}^+} B_r, \forall r \in \mathbb{Q}^+$,
$\tau_r(\omega)$ is a limit point of $\mathscr{Z}(\omega)$ and $P(\bigcup_r B_r) = 0$

□

Theorem 36 (37.2)**.** B.M is measurable $\mathcal{R}^1 \times \mathscr{F}$

Proof. Set $W^{(n)}(t,\omega) := W(\lfloor nt\rfloor/n, \omega)$
HW: show $W^{(n)}$ is measurable $\mathcal{R}^1 \times \mathscr{F}$
By continuity of simple paths,
$W^{(n)} \to W$ everywhere as $n \to \infty$ so W is measurable. □

Class 27: 04/25
HW Exercise 37.2:
$W\left(\frac{2k+1}{2^{n+1}}\right) = \cdots$ [rank n]
—
Let $\mathscr{F}_t := \sigma(W_s; s \leq t)$.
τ is an $\langle \mathscr{F}_t; t \geq 0\rangle$ stopping time if $\forall t$, $[\tau \leq t] \in \mathscr{F}_t$
<u>Fix τ</u> a.s. finite.
$W_t^*(\omega) := W_{\tau(\omega)+t}(\omega) - W_{\tau(\omega)}(\omega)$
$W_t^* := W_{\tau+t} - W_\tau$
$\mathscr{F}_{\tau := \{A; \forall t, A\cap[\tau\leq t]\in\mathscr{F}_t\}}$
$\mathscr{F}_\tau^* := \sigma(W_t^*; t \geq 0)$

Theorem 37 (37.5)**.** (Strong Markov Property):
If τ is an a.s. finite stopping time, then $\langle W_t^*; t \geq 0\rangle$ is a std B.M. and $\mathscr{F}_\tau^*$ is independent of $\mathscr{F}_\tau$

$\tau_a := \inf\{t; W_t = a\}, (a \neq 0)$
This is finite a.s. since $\overline{\lim} W_t = +\infty$ a.s. and $\underline{\lim} W_t = -\infty$ a.s.
Also, τ is a r.v. and is a stopping time.
$[\tau_a \leq t] = [\inf_{s\in[0,t]\cap\mathbb{Q}} |W_s - a| = 0]$
<u>The reflection Principle</u>
Look at the first time BM hits a, and from that point on reflect the brownian motion. Then what we get is also a brownian motion.
$$W'_t := \begin{cases} W_t, & \text{if } t \leq \tau_a \\ W_{\tau_a} - (W_t - W_{\tau_a}) = 2W_{\tau_a} - W_t = 2a - W_t, & \text{if } t \geq \tau_a \end{cases}$$

Theorem 38. $\langle W'_t; t \geq 0\rangle$ is a B.M. In fact, this holds for any a.s. finite stopping time instead of τ_a [though we might not have the simplification $2a - W_t$].

Proof. Note that $W_t - W_\tau$ is W_t^*
So, we can think of W' as the pair $(\langle W_t; t \leq \tau\rangle, \langle W_t^*; t \geq 0\rangle)$.
Conversely, from the pair we can get W'
Think of W as $(\langle W_t; t \leq \tau\rangle, \langle W_t^*; t \geq 0\rangle)$
Similarly, think of W' as $(\langle W'_t; t \leq \tau\rangle, \langle (W'_t)^*; t \geq 0\rangle)$
$\langle W_t; t \leq \tau\rangle = \langle W'_t; t \leq \tau\rangle$
$\langle W_t^*; t \geq 0\rangle = \langle -(W')_t^*; t \geq 0\rangle$
By theorem 37.5, in both cases, the first element of the path is independent of the second element. Hence the theorem is an instance of the general lemma:
If X, Y, Z are r.v.'s with X, Y independent and X, Z independent and $Y \stackrel{\mathscr{D}}{=} Z$ then $(X, Y) \stackrel{\mathscr{D}}{=} (X, Z)$

□

We use the reflection principle on τ_a.
Let $M_t \coloneqq \sup_{s \leq t} W_s$
M_t is a non-negative random variable, non-decreasing in t.
What is the law [cdf]of M_t? Easier to calculate tail probability.
Note that $[M_t \geq a] = [\tau_a \leq t]$ so this is indeed an event.
This is disjoint union of $[W_t \geq a]$ and $[M_t \geq a, W_t < a]$
Note: $[M_t \geq a, W_t < a] = [W_t' > a]$
It is a disjoint union. So, we can calculate the probability:
$P[M_t \geq a] = P[W_t \geq a] + P[W_t' > a] = 2P[W_t \geq a]$
W_t is normal with mean 0, variance t
$= \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{t}}^{\infty} e^{-\frac{y^2}{2}} \, \mathrm{d}y$
Or we can just write: $P[|W_t| \geq a]$
So, $M_t \overset{\mathscr{D}}{=} |W_t|$
This may also remind you of maximal inequality.
<u>Corollary</u>: $\forall a \neq 0, \tau_a \overset{\mathscr{D}}{=} \frac{a^2}{W_1^2}$ and $E[\sqrt{\tau_a}] = \infty$

Proof. By symmetry, we may assume $a > 0$
For $t > 0$ $P[\tau_a \leq t] = P[M_t \geq a] = P[|W_t| \geq a] = P[W_t^2 \geq a^2] = P[tW_1^2 \geq a^2] = P[\frac{a^2}{W_1^2} \leq t]$ which proves the first part.
$E[\sqrt{\tau_a}] = E[\frac{a}{|W_1|}] = aE[\frac{1}{|W_1|}] = a \int_{-\infty}^{\infty} \frac{1}{|t|} \phi(t) \, \mathrm{d}t$
ϕ is density of standard normal.
This is essentially integrating $\frac{1}{|t|}$ for t near 0 so this is infinity which gives us the second part. $\square$

Also interesting: $E[\tau_a^{-1}] = E[\frac{W_1^2}{a^2}] = \frac{1}{a^2}$