Probability Cheat Sheet

Measure Theory Theorems

1. **Monotone Convergence:** For measurable $f_n \uparrow f$, we have $\int f_n \, d\mu \to \int f \, d\mu$

2. **Fatou:** For measurable $f_n : X \to [0, \infty]$, $\int \liminf_{n \to \infty} f_n \, d\mu \le \liminf_{n \to \infty} \int f_n \, d\mu$.

3. **Dominated Convergence:** For measurable $f_n, g : X \to [-\infty, \infty]$ with $f_n \to f$ a.e. and $|f_n|, |f| \le g$ and $g \in L^1(X)$ a.e. then f is integrable and $\int f_n \, d\mu \to \int f \, d\mu$.

Preliminaries

1. **Distribution Functions:** A function $F : \mathbb{R} \to [0, 1]$ is cumulative distribution function if F is **non-decreasing**, tends to 0 on the left and 1 on the right and is right continuous. And given a CDF, we can find a unique probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F(x) = P((\infty, x])$

2. **Markov and Chebyshev** For $X \ge 0$

$$P(X \ge a) \le E(X)/a \implies P(|X|^k \ge a) \le E(|X|^k)/a \implies P(|X - \mu| \ge k) \le \sigma^2/k^2$$

Independence

1. **Independence:** Events $A_i \in \mathcal{F}$ are independent if $P(\cap_i A_i) = \prod_i P(A_i)$. Random variables $\{X_i\}$ are independent if $P(\cap_i X_i \in B_i) = \prod_i P(X_i \in B_i)$ for any $B_i \in \mathcal{B}(\mathbb{R})$. A collection of events $F_i \subset \mathcal{F}$ are independent if for all $I \subset [1, N]$, events $A_i \in F_i$ are independent.

2. **Dynkin:**

   a. π system: Π such that it is closed under intersections $A, B \in \Pi \implies A \cap B \in \Pi$

   b. λ system: Λ such that (1) $\Omega \in \Lambda$, (2) $A, B \in \Lambda \wedge B \subset A \implies A \setminus B \in \Lambda$, (3) $A_n \in \Lambda : A_n \subset A_{n+1} \implies \cup_i A_i \in \Lambda$

   c. Theorem: $\Pi \subset \Lambda \implies \sigma(\Pi) \subset \Lambda$.

   Useful for showing sigma algebras generated by independent π systems in $\mathcal{F}$ are independent. **Proof:** Use the fact that the events independent of all but one pi system is a lambda system and so the sigma algebra generated by the one pi system is contained in this lambda system as the pi system itself is independent of all the other pi systems by definition and so the sigma algebra generated by this pi system is independent because it is contained in the independent lambda system. And repeat.

3. **Equivalent Characterisation:** A family of R.V.s X_n are independent iff $E(f_1(X_1) \dots f_n(X_n)) = E(f_1(X_1)) \dots E(f_n(X_n))$ for all bounded borel measurable functions.

Convergence

1. **Probability**: A sequence of random variables $(X_n)_{n \in \mathbb{N}}$ convergence in probability to a random variable X if for all $\epsilon > 0$

$$P(|X_n - X| \le \epsilon) \to 1$$

   This is metrizable and we get $X_n \xrightarrow{p} X$ iff $d(X_n, X) \to)$ where d is the following metric

$$d(X, Y) = E\left(\frac{|X - Y|}{1 + |X - Y|} \right)$$

   We also have $X_n \xrightarrow{p} X \implies f(X_n) \to f(X)$ for continuous f and $E(f(X_n)) \to E(f(X))$ if f is bounded aswell.

2. **A.S.**: Pointwise convergence P-almost everywhere.

$$P(X_n \to X) = 1$$

   implies convergence in probability. **Proof:** Key step is that $X_n \to X$ a.s. $\implies P(|X_n - X| > \epsilon \ i.o.) = 0$ and use continuity of measure from above.

   We also have $X_n \xrightarrow{a.s.} X \implies f(X_n) \xrightarrow{a.s.} f(X)$ for continuous f by dominated convergence.

3. L^p: A sequence of random variables $(X_n)_{n \in \mathbb{N}} \in L^p(\Omega)$ converges in L^p if

$$E(|X_n - X|^p) \to 0$$

   implies convergence in probability because $P(|X_n - X| > \epsilon) \le P(|X_n - X| \ge \epsilon) \le \frac{E(|X - X^n|^p)}{\epsilon^p} \to 0$

4. **Important Topological Theorem** ($*$): For $(y_n)_{n \in \mathbb{N}}$ in a topological space, $y_n \to y$ as $n \to \infty$ iff every subsequence $(y_{n(m)})_{m \in \mathbb{N}}$ has a further subsequence $y_{n(m(k))_{k \in \mathbb{N}}} \to y$ as $k \to \infty$.

Borel Cantelli

1. **liminf and limsup:**

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n}^{\infty} A_n = \{A_n \ i.o.\} \quad \text{and} \quad \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n}^{\infty} A_n = \{A_n \ a.a.\}$$

2. **BC 1:** For events $A_n \in \mathcal{F}$, $\sum P(A_n) < \infty \implies P(A_n \ i.o.) = 0$.

   Proof: Create a counter random variable, N, which is the sum of the indicators for A_k then argue that $E(N) < \infty \implies N < \infty$ a.s. where we use a fubini to swap the infinite sum and the expectation.

3. **BC 2:** A_n independent $\implies \sum A_n = \infty \implies P(A_n \ i.o.) = 1$. Note that independence is required. Counter example: We can take $\Omega = [0,1], \mathcal{F} = \mathcal{B}(\Omega), P = \lambda$ and $A_n = [0, 1/n]$ and $\sum P(A_n) = \infty$ but $P(A_n \ i.o.) = P(\bigcup_{n=1}^{\infty} \bigcap_{m \geq n}^{\infty} A_n) = P(\emptyset) = 0$.

   Proof: Can prove this by showing $P(\cup_{m \geq n} A_n) = 0$ by showing $P((\cup_{m \geq n} A_n)^C) = 0$ by using continuity from above at 0.

Law of Large Numbers:

1. L^2 **WLLN:** For a sequence of identically distributed and uncorrelated random variables with finite first and second moments, we have $\bar{X}_n \xrightarrow{p} E(X_1)$. **Proof:** $E((\bar{X}_n - \mu)^2) = \text{Var}(\bar{X}_n) \to 0 \implies L^2$ convergence $\implies$ convergence in probability.

2. **WLLN w Truncation:** Long technical theorem which is pretty much reverse engineered to meet the requirements of WLLN 2. Truncation is what we introduce in (3) and controls large values of X_n. If

   a. For sequence of i.i.d. X_n

   b. $b_n > 0$ with $b_n \to 0$

   c. $\bar{X}_{n,k} = X_k \mathbf{1}\{|X_k| \leq b_n\}$

   d. $nP(|X_1| > b_n) \to 0$ as $n \to \infty$ which is saying the mass of the sum of the tails goes to 0.

   e. $\frac{n}{b_n^2} E(\bar{X}_{1,n}^2) \to 0$ as $n \to \infty$ we want this so that we can use chebyshev to control the the bad set.

   then $b_n^{-1}(S_n - nE(\bar{X}_{n,1})) \xrightarrow{p} 0$. The **proof** of this uses a union bound to split

   $$P\left(\left|\frac{S_n - n\mu}{b_n}\right| > \epsilon\right) \leq P(S_n \neq \bar{S}_n) + P\left(\left|\frac{\bar{S}_n - n\mu}{b_n}\right| > \epsilon\right) \overset{\text{union bound}}{\leq} nP(|X_1| > b_n) + \overset{\text{chebyshev}}{\frac{n}{b_n^2 \epsilon^2} E(\bar{X}_{1,n}^2)} \to 0 \text{ by assumption}$$

3. **WLLN 2:** For i.i.d. X_n with finite first moment we have $\bar{X}_n \xrightarrow{p} E(X_1)$.

   Proof: Use the above theorem with $b_n = n \to \infty$, we can show $nP(|X_n| > n) \leq E(|X_n|\mathbf{1}_{\{|X_n|>n\}}) \to 0$ using the dominating function X_n which is integrable by assumption. Next we use $E(Y^p) = \int_0^{\infty} py^{p-1}P(Y > y)dy$ (use indicator and fubini) for non negative Y, which we use to show

   $$\frac{1}{n}E(\bar{X}_{1,n}^2) = \frac{1}{n}\int_0^{\infty} 2yP(|\bar{X}_{1,n}| > y) \, dy = \frac{1}{n}\int_0^n 2yP(|X_n| > y) \, dy \to 0$$

4. **4th Moment SLLN:** For i.i.d random variables X_n with $E(X_n) = \mu$ with $E(X^4) < \infty$ then $\bar{X}_n \xrightarrow{a.s.} \mu$.

   Proof: Assuming $E(X) = 0$ wlog, use chebyshev to bound $P(|S_n/n| > \epsilon)$ and show $\sum P(|S_n/n| > \epsilon) < \infty \implies P(|S_n/n| > \epsilon \ i.o.) = 0 \implies S_n/n \to 0$ a.s. by BC1.

Kolmogorov SLLN

1. **Tail**: For R.V.s (X_n), $\mathcal{T} = \cap_i \sigma(X_i, X_{i+1}, \dots)$

2. **0-1:** If X_n are independent then $A \in \mathcal{T} \implies P(A) \in \{0, 1\}$.
   Proof: Basically show that $\mathcal{T}$ is independent of itself $\implies P(A)^2 = P(A) \implies P(A) \in \{0, 1\}$.

3. **Kolmogorov Maximal Inequality:** For independent randomv variables X_n with 0 expectation and finite second moment, for $a > 0$

   $$P\left(\max_{1 \leq k \leq n} |S_k| \geq a\right) \leq \frac{1}{a^2} E(S_n^2)$$

4. **2 series:** This is a weaker version of Three series in Durrett which only requires the below convergence of truncated X and that the mass of the clippings is finite. For independent X_n

$$\sum E(X_n) < \infty \quad \wedge \quad \sum \mathrm{Var}(X_n) < \infty \implies \sum X_n \text{ converges a.s.}$$

Proof: Assume $E(X_n) = 0$ wlog which we can because $\sum EX_n < \infty$ then we need to show

$$P(S_n(w) \text{ is cauchy}) = P(\sup_{m,n \geq M} |S_m - S_m| \overset{M \to \infty}{\to} 0) = 1$$

Showing $\sup_{m,n \geq M} |S_m - S_m| \overset{p}{\to} 0$ is enough because $\sup_{m,n \geq M} |S_m - S_m|$ is a sequence of non increasing random variables and convergence in probability implies convergence a.s. Next because

$$\sup_{m,n \geq M} |S_m - S_m| \leq 2 \sup_{n \geq M} |S_n - S_M| \implies P(\sup_{m,n \geq M} |S_m - S_m| > \epsilon) \leq P(\sup_{n \geq M} |S_n - S_M| > \epsilon/2)$$

and because

$$\{\max_{M \leq k \leq N} |S_k - S_M| \geq \epsilon/2\} \overset{N \uparrow \infty}{\uparrow} \{\sup_{m \geq M} |S_m - S_M| \geq \epsilon/2\}$$

we can bound the left using the maximal inequality to show that it goes to 0 using $\sum \mathrm{Var}(X_n) < \infty$ hence we have $\sup_{m,n \geq M} |S_m - S_m| \overset{p}{\to} 0$ which shoes convergence a.s.

5. **Second Moment SLLN:** For $b_n \to \infty$, $\sum \mathrm{Var}(X_n)/b_n^2 < \infty \implies \frac{1}{b_n}(S_n - E(S_n)) \to 0$ a.s.

   Proof: $\frac{1}{b_n}(S_n - E(S_n)) = \frac{1}{b_n}\sum(X_n - E(X_n))$ converges a.s. using kroneckers lemma and two series where the sum of the expectations is 0 and the sum of the variance is bounded by assumption.

6. **Kolmogorov SLLN:** For i.i.d random variables with finite first moment and $E(X_i) = \mu$ we have $\bar{X}_n \to \mu$.

   Proof: First truncate to show that $\bar{S}_n \to \mu$ a.s. $\implies S_n \to \mu$ a.s. by using borel cantelli to show $P(X_n \neq \bar{X}_n \ i.o.) = 0$ where $\bar{X}_n = X_n \mathbf{1}_{\{|X_n|<n\}}$. Then show that $\sum \mathrm{Var}(\bar{X}_n)/n^2 < \infty$ using similar technique to WLLN 2. Then use the second moment SLLN to conclude that $\bar{S}_n \to \mu$ a.s. which implies $S_n \to \mu$ a.s..

Central Limit Theorems

1. **DeMoivre-Laplace Theorem:** For i.i.d. random variables $(X_n)_{n \in \mathbb{N}}$ with $P(X_n = \pm 1) = 1/2$, for S_n and $a < b$, we have

$$P(a \leq S_n/\sqrt{n} \leq b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

2. **Weak Convergence/ Convergence in Dist:** For distribution functions $(F_n)_{n \in \mathbb{N}}$, $F_n \Rightarrow F$ if we have pointwise convergence for all points of continuity. Random variables can converge in distribution even if the random variables are defined on different probability spaces. Furthermore, if F is not continuous we have at most **countably many discontinuities**.

3. **Quantile Function:** For probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}(0, 1)$ and $P = \lambda$, the lebesgue measure, we can construct a random variable $X(\omega)$ on this space with distribution function F by setting

$$X(\omega) = \sup\{y \in \mathbb{R} : F(y) < \omega\} = \min\{y \in \mathbb{R} : F(y) \geq \omega\} := F^{-1}(\omega)$$

where F^{-1} is called the quantile function. The quantile function is **non-decreasing and left continuous** and $F^{-1}(\omega)$ is continuous at ω iff $w \in \Omega_c = \{\omega \in \Omega : \sup\{y \in \mathbb{R} : F(y) < \omega\} = \inf \sup\{y \in \mathbb{R} : F(y) > \omega\}\}$. Furthermore, $(0, 1) \setminus \Omega_c$ is countable.

4. **Weak Convergence to a.s:** If $F_n \Rightarrow F$, then we can create random variables on the same probability space with $F_{X_n}(x) = P(X_n \leq x)$ and $F_X(x) = P(X \leq x)$ such that $X_n \overset{a.s}{\to} X$.

   Proof: Create $Y_n = F^{-1} := \sup\{y : F(y) \leq \omega\}$ on $((0, 1), \mathcal{B}(0, 1), \lambda)$ then show $\limsup F_n^{-1} \leq F^{-1}$ and $\liminf F_n^{-1} \geq F^{-1}$ at all points of continuity then we have pointwise convergence of random variables F_n^{-1} to F^{-1}.

5. **Weak Convergence Equivalent:** $X_n \Rightarrow X \iff \forall g \in C_b(\mathbb{R}), E(g(X_n)) \to E(g(X))$. With this theorem we can show a corollary that $X_n \Rightarrow X$ then $g(X_n) \Rightarrow g(X)$ for any $g \in C(\mathbb{R})$ by composing with $h \in C_b(\mathbb{R})$.

   Proof: For left, create a sequence of a.s. converging random variables Y_n using the previous theorem. Then we know that $Y_n \to Y$ a.s. $\implies g(Y_n) \to g(Y)$ a.s. then we use bounded convergence theorem to get $\lim_{n \to \infty} E(g(Y_n)) = E(\lim_{n \to \infty} g(Y_n)) = E(g(Y))$. For the other direction use piecewise continuous function $g_{x,\epsilon}(y) = 1$ when $y \leq x$ and 0 when $y \geq x + \epsilon$ and connects inbetween. Then $E(g_{x,\epsilon}(X_n)) = P(X_n \leq x) + f(\epsilon) \to F_n(x)$ and the right goes to $F(x)$ as $\epsilon \to 0$ and $n \to \infty$.

6. **Helly Selection Theorem:** For a sequence of distribution functions $(F_n)_{n \in \mathbb{N}}$, there is a subsequence and a right continuous non-increasing function F such that at every point of continuity of F, $F_{n_k}(x) \to F(x)$.

   Proof: basically enumerates $\mathbb{Q} = \{q_i\}_{i \in \mathbb{N}}$ then builds a subsequence of (F_n) which converges at q_1 using Bolzano Weierstrass then a sub sub sequence which converges at q_2 and so on. Then for each rational we have a sequence which converges for it and the ones before and so we pick the diagonal sequence and this converges for all rationals. Then set $F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$.

7. **Tightness:** However, this is not good enough because F is not necessarily a distribution function, so we introduce a condition called **tightness** on the probability measures to ensure that F is. Probability measures, $(P_i)_{i \in I}$, on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are tight if for all $\epsilon > 0$ there exists $K_c \in \mathcal{B}(\mathbb{R})$ such that the measures of K_c^C are uniformly bounded by ϵ. Naturally, a $(F_i)_{i \in \mathbb{N}}$ are tight if the probability measures they correspond to are. **Equivalently**, tightness just means that no mass is lost to the side, ie for all $\epsilon > 0$, there is a $t \in \mathbb{R}$ such that $P_i(|X_i| \geq t) \leq \epsilon$.

   Now, for a sequence of **tight** distribution functions, $(F_n)_{n \in \mathbb{N}}$, we have that there exists a subsequence and another distribution function F such that $F_{n_k} \Rightarrow F$.

8. **Levy Metric:** For two distribution functions F, G, we can define the metric

$$d(F, G) = \inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon \text{ for all } x \in \mathbb{R}\}$$

   and show that $F_n \Rightarrow F \iff d(F_n, F) \to 0$. See 4

9. **Characteristic Function and MGF:** The characteristic function is defined as the fourier transform of the power law and is given by

$$\phi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} \, dP_X(x) = E(e^{i\theta X})$$

   and it has the following properties

   a. $\phi_X(0) = 0$ and $\phi_X(-\theta) = \overline{\phi_X(\theta)}$ and $|\phi_X(\theta)| \leq 1$

   b. ϕ is uniformly continuous

   c. If X has finite n-th moment then ϕ_X is n times differentiable and $\phi_x^{(n)}(\theta) = E((iX)^n e^{iX\theta})$ is uniformly continuous.

   d. If X, Y independent and $\alpha \in \mathbb{R}$ then $\phi_{\alpha X + Y} = \phi_X(\alpha\theta) + \phi_Y(\theta)$.

   e. $E_Y(e^{-iYt}\phi_X(Y)) = E_X(\phi_Y(X - t))$.

   and closely related is the **MGF**, which is $M_X(t) = E(e^{tX})$ for $t \in \mathbb{R}$ which may not exists for all random variables but the characteristic function does.

   Additionally we have **uniqueness of characteristic functions** where $\phi_\nu = \phi_\mu \iff \mu = \nu$

   Proof: Given two characteristic functions $\phi_\nu = \phi_\mu$, we have that $F\mu(x) = F_\nu(x)$ which we get from the inversion formula below. Next we build the distribution function $F(x)$ by setting $F(x) := \lim_{x_n \downarrow x} F_\nu(x_n)$ where x_n are points of continuity (which we can do because there are countably many points of discontinuity. Next we have that the distribution functions agree and hence the measures must be the same.

10. **Gaussian CHF**: For a gaussian X, $\phi_X(\theta) = e^{i\mu\theta - \sigma^2\theta^2/2}$

11. **Inversion Formula:** The inversion formula allows us to recover the distribution from the characteristic function and it says that for any point of continuity of F_X, $x \in \mathbb{R}$,

$$F_X(x) = \lim_{a \to \infty} \int_{-\infty}^{x} \left(\frac{1}{2\pi} \int_{\infty}^{\infty} \exp(ist)\phi_X(s) \exp\left(\frac{-s^2}{2a^2}\right) ds \right) dt$$

   Proof: Use gaussian approximation $X_a = X + Z/a \to X$ a.s. $\implies X_a \Rightarrow X \implies F_{X_a} \to F_X$ at all points of continuity.

12. **Inversion when Density exists** Above is a more general case for when X might not have a density but for probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we have

$$\int_{\mathbb{R}} |\phi_\mu(\theta)| \, d\theta < \infty \implies \mu << \lambda \text{ and } f_\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(i\theta x)\phi_\mu(\theta) \, d\theta$$

   which is more convenient and is just a fourier inversion.

13. **Levy Continuity Theorem:** This theorem allows use to go from convergence of characteristic functions to weak convergence and vice versa. For measures μ_n on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we have

   a. $\mu_n \Rightarrow \mu \implies \phi_{\mu_n}(\theta) \to \phi_\mu(\theta)$ for all $\theta \in \mathbb{R}$

   b. If $\phi_{\mu_n}(\theta) \to \phi(\theta)$ for all $\theta \in \mathbb{R}$ and ϕ continuous at 0, then ϕ is the characteristic function for some probability measure μ and $\mu_n \Rightarrow \mu$.

Importantly note the extra condition when going backwards from pointwise convergence of characteristic functions to weak convergence/ convergence in distribution.

Proof: For the first one we have as $\mu_n \Rightarrow \mu$, $g(x) = e^{i\theta x} \in C_b(\mathbb{R})$ we have $E_{\mu_n}(g(X)) \to E_\mu(g(X))$. For the second statement, we can control the mass of the tails by using

$$P(|X| > 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} 1 - \phi_X(\theta) \, d\theta$$

then using continuity at zero we can control $P(|X| > 2/\delta)$ to show the measures are tight and by helly selection theorem we can find a subsequence which converges to a valid probability measure and because the characteristic functions converge, the characteristic functions of this subsequence also converge to $\phi(\theta)$ and so each of these measures that each subsequence converges to is the same by the correspondence of measures and characteristic functions. Because weak convergence is metrizable we have the result.

14. **Central Limit Theorem:** For a sequence of i.i.d. random variables $(X_n)_{n \in \mathbb{N}}$ with finite second moment and $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$, we have

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}} \Rightarrow N(0, \sigma^2)$$

Proof: Assume $\mu = 0$ wlog. Then by Levy we just have to show that $\phi_{Z_n} \to e^{-\sigma^2 \theta^2/2}$ where the right hand side is trivially continuous at 0. We can do this because

$$\phi_{Z_n} = \prod_{k=1}^{n} \phi_{X_j/\sqrt{k}} = \left(\phi_{X_j/\sqrt{k}}(\theta)\right)^n = \left(\phi_{X_j}(\theta/\sqrt{k})\right)^n \stackrel{\text{taylor}}{=} \left(1 - \frac{\theta^2 \sigma^2}{2n} + R_X(\theta/\sqrt{n})\right)^n \to e^{\sigma^2 \theta^2/2}$$

where we get the taylor expansion collapsing because of the finite second moment assumption.

Discrete Time Martingales

1. **Conditional Expectation:** For probability space $(\Omega, \mathcal{F}, P)$ and sub sigma algebra $\mathcal{G} \subset \mathcal{F}$ and random variable X with $E(|X|) < \infty$, $Y := E(Y|\mathcal{G})$ is the conditional expectation of X given $\mathcal{G}$ if Y is $\mathcal{G}$-measurable and for all $A \in \mathcal{G}$, $E(X \cdot \mathbf{1}_A) = E(Y \cdot \mathbf{1}_A)$. A simple lemma is that $E(|Y|) \leq E(|X|)$. This conditional expectation exists and is unique (up to a.s, equality).

   Proof: We can prove uniqueness by showing if Y, Z are condition probabilities $E(X|G)$ and $A_n = \{Y - Z \geq 1/n\}$ then $0 = E((Y - Z)\mathbf{1}_{A_n}) \geq P(A_n)/n$ and $P(Y > Z) = \cup_n P(A_n) = 0$ and vice versa so $Y = Z$ a.s..

2. **Properties**

   a. **Linearity:** $E(\alpha X + Y|\mathcal{F}) = \alpha E(X|F) + E(Y|F)$

   b. **Monotonicity:** $X \leq Y$ (a.s.) $\implies E(X|F) \leq E(Y|F)$ (a.s.)

   c. **Monotone Convergence:** $X_n \geq 0$ and $X_n \uparrow X$ with $EX < \infty$ then $E(X_n|F) \uparrow E(X|F)$

   d. **Fatou:** If $X_n \geq 0$ (a.s.) then $E(\liminf_{n \to \infty} X_n|F) \leq \liminf_{n \to \infty} E(X_n|F)$.

   e. **Dominated Convergence:** If $|X_n| \leq Y$ (a.s.) and $X_n \to X$ (a.s.), then $E(X_n|F) \to E(X|F)$ (a.s.).

   f. **Jensen:** If ϕ is a convex (implies measurablility) function from $\mathbb{R} \to \mathbb{R}$ such that $\phi(X) \in L^1(\Omega)$, then $\phi(E(X|F)) \leq E(\phi(X)|F)$ (a.s.).

   g. **Independent Condition:** If Y is independent of G then $E(Y|G) = E(Y)$.

   h. **Tower:** If $F \subset G$ and $E(X|G) \in F$, then $E(X|F) = E(X|G)$, i.e. given a larger set of data with redundancies, you can refine it into a smaller set of data with the same information. Futhermore, for $F_1 \subset F_2$, $E(E(X|F_1)|F_2) = E(E(X|F_2)|F_1) = E(X|F_1)$.
   Proof: We use the fact that sigma algebras are subsets of each other and point 2 in the definition of condition expectation.

   i. **Factorisation:** If $X \in F$, $E(|Y|)$ and $E(|XY|) < \infty$, then $E(XY|F) = XE(X|F)$.
   Proof: Prove for when X is an indicator function then we can show that we can approximate any random variable with indicator functions and so the integrals will match because of dominated convergence theorem.

3. **Filtrations:** $\mathcal{F}_n$ is called a filtration if it is a sequence of increasing $\sigma-$algebras (ie. $F_1 \subset F_2 \subset \cdots \subset F_n \subset \cdots \subset F$. X_n is **adapted** to F_n if $X_n \in F_n$ for all $n \in \mathbb{N}$.

4. **Martingales:** A martingale is a sequence of random variables, X_n such that $E(X_n) < \infty$, X_n is adapted to filtration F_n and $E(X_{n+1}|F_n) = X_n$. If instead, $E(X_{n+1}|F_n) \leq X_n$ or $E(X_{n+1}|F_n) \geq X_n$ then X_n is a super or sub martingale.

5. **Examples of Martingales:**

   a. **Partial Sum:** For a seuqnce of i.i.d. random variables X_n with finite first moment and 0 mean and the natural filtrations associated with them $\mathcal{F}_n$, the sequence S_n is a martingale. Furthermore, if $E(X^2) = \sigma^2 \in [0, \infty)$ then $M = S_n^2 - n\sigma^2$ is also a martingale w.r.t. $\mathcal{F}_n$.

b. **Exponential Martingale:** For $Y_n \geq 0$ with $E(Y_n) = 1$, $M_n = \prod_{m \leq n} Y_m$ (we have $E(M_n) = 1$) is a martingale on the natural filtration of Y_n.

6. **Martingale Transform:** A sequence of random variables H_n is called **previsible** if H_n is $\mathcal{F}_{n-1}$-measurable for a filtration $\mathcal{F}_n$. Furthermore, for an adapted X_n, the martingale transform of X by H is given by

$$(H \bullet X)_n = \sum H_m(X_m - X_{m-1})$$

This can be thought of as the winnings at time n based on a betting strategy H, the amount of money bet in round n, where X is the net amount of money made after the game in round n if you made a bet of \$1 each round.

The martingale transform of a super matringale or sub martingale by a non-negative previsible sequence is also a super or sub martingale. Additionally, the transform of a martingale by a real value previsible sequence is a martingale.

7. **Stopping Time:** Random variable N is a stopping time if $\{N = n\} \in F_n$ for all n. The stopped process can be seen as a martingale transform of X_n by $H_n = \mathbf{1}_{N \geq n}$ which is previsible because $\{N \geq n\} = \{N \leq n-1\}^c \in F_{n-1}$. Furthermore, since $(\mathbf{1}_{N \geq n} \bullet X)_n = X_{N \wedge n} - X_0$ is F_n measurable, we have that $X_{N \wedge n}$ is also F_n measurable. Furthermore, if X_n is a super martingale then $X_{N \wedge n}$ is a super martingale.

8. **Upcrossing:** For $a < b \in \mathbb{R}$, we define the upcrossing times

$$\tau_{2k-1} = \inf\{m > \tau_{2k-2} : X_m \leq a\} \quad \text{and} \quad \tau_{2k} = \inf\{m > \tau_{2k-2} : X_m \geq b\}$$

and intuitively, the τ_{2k} represents the time at which X_n goes above b for the k-th time and τ_{2k-1} represents the time at which X_n goes below a for the k-th time. Additionally, $U_n = \sup\{k : \tau_{2k} \leq n\}$, ie. the number of times we've had an upcrossing.

For a submartingale $(X_n)_{n \in \mathbb{N}}$ on filtration $(F_n)_{n \in \mathbb{N}}$ and $a < b \in \mathbb{R}$, we have

$$(b-a)E(U_n) \leq E[(X_n - a)^+] - E[(X_0 - a)^+]$$

9. **Martingale Convergence Theorem:** For submartingale $(X_n)_{n \in \mathbb{N}} : \sup E(X_n^+) < \infty$, $X_n \to X$ (a.s) where $E(X) < \infty$.

Proof: $U_n(a,b) \uparrow U(a,b) \in \mathbb{N} \cup \{\infty\}$ and claim $U(a,b)$ is finite because $(b-a)E(U_n(a,b)) \leq E((X_n-a)^+) \leq E(X_n^+) + |a| \to L \in \mathbb{R}$ so by monotone convergence we have $U(a,b) = L$.