July 2

Notes

- **Definition 1.1.3:** Event $E \subseteq \Omega$ can be empty.
- **Note:** Probability is a function that takes sets as inputs.
- **Definition 1.1.4:** Not testable. σ-algebra is used to define the domain of the probability function and allows the function to have properties that we desire.
- **Definition 1.1.5:** Not testable.
- **Definition 1.1.6:** i) should be $\forall E \in \mathcal{F}$ instead of $\forall E \in \Omega$. ii) should be $\mathbb{P}(\Omega) = 1$ instead of $\mathbb{P}(\varnothing) = 0$. The latter condition is insufficient for the definition since examples of $\mathbb{P}$ can be constructed such that $\mathbb{P}(\varnothing) = 0$ yet $\mathbb{P}'(\Omega) < 1$.
- **Note:** The Vitali set $\in \mathbb{R}$ does not satisfy the 3 axioms in 1.1.6 for it to be a probability measure.
- **Proposition 1.1.10:** WTS $E \subseteq F \implies \mathbb{P}(E) \leq \mathbb{P}(F)$. Notice $E \subseteq F \implies F = E \cup (F \setminus E) \implies \mathbb{P}(F) = \mathbb{P}(E \cup (F \setminus E)) = \mathbb{P}(E) + \mathbb{P}(F \setminus E) \geq \mathbb{P}(E)$ since $\mathbb{P}(F \setminus E) \geq 0$.
- **Lemma 1.1.11:** Define $F_1 = E_1, F_2 = E_2 \cap F_1^c, F_3 = E_3 \cap (F_1 \cup F_2)^c, \ldots, F_i = E_i \cap (\cup_{k=1}^{i-1} F_k)^c$. Notice $\cup_i^\infty F_i = \cup_i^\infty E_i$. Then, $\mathbb{P}(\cup_i E_i) = \mathbb{P}(\cup_i F_i) = \sum_i \mathbb{P}(F_i) \leq \sum_i \mathbb{P}(E_i)$ since $F_i \subseteq E_i \implies \mathbb{P}(F_i) \leq \mathbb{P}(E_i)$.
- **Lemma 1.1.12:** Define $E = \{x_i\}$ and fix $\varepsilon > 0$. For convenience, define $E_i(\varepsilon) = [x_i - 2^{-i}\varepsilon, x_i + 2^{-i}\varepsilon)$. Since $E \subseteq \cup_i E_i(\varepsilon), \mathbb{P}(E) \leq \mathbb{P}(\cup_i E_i(\varepsilon)) \leq \sum_i \mathbb{P}(E_i(\varepsilon)) = \sum_i (x_i + 2^{-i}\varepsilon - x_i + 2^{-i}\varepsilon) = 2\sum_i 2^{-i}\varepsilon = 2\varepsilon$. Thus, $\mathbb{P}(E) \leq 0 \implies \mathbb{P}(E) = 0$.
- **Proposition 1.1.13:** Prove by contradiction.
- **Lemma 1.1.14:** Prove by induction.
- **Proposition 1.1.17:** Suppose $A_0 = \varnothing$ and $A_n \nearrow A = \lim_{n\to\infty} A_n$ and define $B_1 = A_1, \ldots, B_n = A_n \cap A_{n-1}^c = A_n - A_{n-1}$. Notice $\cup_m^n B_m = \cup_m^n (A_m \cap A_{m-1}^c) = \cup_m^n A_m \cap \cup_m^n A_{m-1}^c$ by distributing the union $= A_n \cap A_0^c = A_n$. Then, $\mathbb{P}(\lim_{n\to\infty} A_n) = \mathbb{P}(A) = \mathbb{P}(\cup_m B_m) = \sum_m^\infty \mathbb{P}(B_m) = \lim_{n\to\infty} \sum_m^n \mathbb{P}(B_m) = \lim_{n\to\infty} \mathbb{P}(\cup_m^n B_m) = \lim_{n\to\infty} \mathbb{P}(A_n)$. For $A_n \searrow A$, notice it implies $A_n^c \nearrow A^c$, so $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c) = 1 - \lim_{n\to\infty} (1 - \mathbb{P}(A_n)) = \lim_{n\to\infty} \mathbb{P}(A_n)$.
- **Example 1.1.19:** Does not have a limit.
- **Definition 1.2.1:** Replace Ω with $\mathcal{F}$. Also, note that any subset of the events $E_1, \ldots, E_n$ must be independent as well.
- **Proposition 1.2.2:** Fix $I \subseteq [n] = \{1, \ldots, n\}$. If $1 \notin I, \mathbb{P}(\cap_{i\in I} E_i) = \Pi_{i\in I}\mathbb{P}(E_i)$. If $1 \in I$, let $I' = I \setminus \{1\}$. Then, $\mathbb{P}((\cap_{i\in I'} E_i) \cap E_1^c) = \mathbb{P}((\cap_{i\in I'} E_i) \setminus E_1) = \mathbb{P}(\cap_{i\in I'} E_i) - \Pi_{i\in I}\mathbb{P}(E_i)$ by property 6 in Proposition 1.1.10 since $(\cap_{i\in I'} E_i) \cap (E_1) = \cap_{i\in I} E_i$. Continuing on, this equals $\Pi_{i\in I'}\mathbb{P}(E_i) - \Pi_{i\in I}\mathbb{P}(E_i) = (1 - \mathbb{P}(E_1))\Pi_{i\in I'}\mathbb{P}(E_i) = \mathbb{P}(E_1^c)\Pi_{i\in I}\mathbb{P}(E_i)$.
- **Definition 1.2.3:** $\{E_\alpha : \alpha \in \mathcal{I}\}$ does not need to be countable, which hints at Definition 2.1.5 later on.

- **Lemma 1.2.6:** Prove that the measure satisfies the 3 axioms in 1.1.6.
- **Proposition 1.2.7:** $E_i \cap E_j = \varnothing \implies (A \cap E_i) \cap (A \cap E_j) = \varnothing$. Then, $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap \cup_i E_i) = \mathbb{P}(\cup_i (A \cap E_i)) = \sum_i \mathbb{P}(A \cap E_i) = \sum_i \mathbb{P}(A|E_i)\mathbb{P}(E_i)$.
- **Proposition 1.2.10:** Replace Ω with $\mathcal{F}$.

Questions

- **1.1:**
  1. $1 = \mathbb{P}(\Omega) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c) \implies \mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
  2. $1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \varnothing) = \mathbb{P}(\Omega) + \mathbb{P}(\varnothing) \implies \mathbb{P}(\varnothing) = 0$.
  3. $E \subseteq F \implies F = E \cup (F \setminus E) \implies \mathbb{P}(F) = \mathbb{P}(E \cup (F \setminus E)) = \mathbb{P}(E) + \mathbb{P}(F \setminus E) \geq \mathbb{P}(E)$.
  4. $\mathbb{P}(E \cup F) = \mathbb{P}((E \setminus F) \cup (F \setminus E) \cup (E \cap F)) = \mathbb{P}(E \setminus F) + \mathbb{P}(F \setminus E) + \mathbb{P}(E \cap F) = \mathbb{P}(E) - \mathbb{P}(E \cap F) + \mathbb{P}(F) - \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$ since $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(F \cap E)$.
  5. Rearrange $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.
  6. $F = (F \cap E) \cup (F \setminus E) \implies \mathbb{P}(F) = \mathbb{P}(F \cap E) + \mathbb{P}(F \setminus E)$ since $(F \cap E) \cup (F \setminus E) = (F \cap E) \cap (F \cap E^c) = (F \cap F) \cap (E \cap E^c) = F \cap \varnothing = \varnothing$.
- **1.5:** Define A_i as the event where the ith person gets their chair. Then, the desired result is $\mathbb{P}(\cap_{i=1}^n A_i^c = 1 - \mathbb{P}(\cup_{i=1}^n A_i) = 1 - \mathbb{P}[\Sigma_{i=1}^n \mathbb{P}(A_i) - \Sigma_{i<j}\mathbb{P}(A_i \cap A_j) + \Sigma_{i<j<k}\mathbb{P}(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n=1}\mathbb{P}(\cap_{i=1}^n A_i)] = 1 - \Sigma_{i=1}^n \frac{(-1)^i}{i!}$. Notice that $\lim_{n\to\infty} \Sigma_{i=1}^\infty \frac{(-1)^i}{i!} = e^{-1}$, so the probability approaches $1 - e^{-1}$ as $n \to \infty$.
- **1.8:** Let $\Omega = \{1, 2, 3\}$, $A = \varnothing$, $B = C = \{1\}$. Then $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 0$, yet $\mathbb{P}(B \cap C) = \frac{1}{3} \neq \frac{1}{9} = \mathbb{P}(B)\mathbb{P}(C)$.
- **1.9:** $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1$ since $A \cap B \subseteq B \implies \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$. $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$. $\mathbb{P}_B(\cup_{i=1}^\infty E_i) = \frac{\mathbb{P}((\cup_{i=1}^\infty E_i) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{i=1}^\infty (E_i \cap B))}{\mathbb{P}(B)} = \frac{\Sigma_{i=1}^\infty \mathbb{P}(E \cap B)}{\mathbb{P}(B)} = \Sigma_{i=1}^\infty \mathbb{P}_B(E_i)$ since $(E_i \cap B) \cap (E_j \cap B) = (E_i \cap E_j) \cap (B \cap B) = \varnothing$ for $i \neq j$.

July 4

Notes

- **Definition 2.1.1:** The argument of a random variable is random, not the variable itself.
- **Lemma 2.2.1:** WTS μ is a probability measure. Proof: i) $\forall A \in \mathcal{B}(\mathbb{R}), \mu(A) = \mathbb{P}(x \in A) \in [0, 1]$. ii) $\mu(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\{w \in \mathbb{R} : X(w) \in \mathbb{R}\}) = 1$. iii) Take $E_i = \{w \in \Omega : X(w) \in A\} \in \mathcal{F}$, so $\mathbb{P}(E_i)$ is defined. Since X is a function, it cannot have different outputs for the same input. Thus, the E_i's are disjoint. Then, $\mu(\cup_i^\infty A_i) = \mathbb{P}(X(w) \in \cup_i^\infty A_i) = \mathbb{P}(\cup_i^\infty \{X(w) \in A_i\}) = \mathbb{P}(\cup_i^\infty E_i) = \sum_i^\infty \mathbb{P}(E_i) = \sum_i^\infty \mu(A_i)$.
- **Theorem 2.2.5:**
  - i) $x \leq y \implies \{X \leq x\} \subseteq \{X \leq y\} \implies \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$.
  - ii) Define $\{x_n\} \to \infty$ such that $A_n = \{w \in \mathbb{R} : X(w) \leq X_n\} \nearrow A = \{w \in \Omega : X(w) \leq \infty\} = \Omega$. By the continuity of probability, $F(X_n) = \mathbb{P}(A_n) \to \mathbb{P}(A) = \mathbb{P}(\Omega) = 1$, so $\lim_{x\to\infty} F(x) = 1$. For the case $\lim_{x\to-\infty} F(x) = 0$, define $\{x_n\} \to -\infty$ to yield $A_n \searrow A = \{w \in \Omega : X(w) \leq -\infty\} = \varnothing$.

iii) WTS $\lim\limits_{x \to a^+} F(x) = F(a)$. Let $a \in \mathbb{R}$. Define $\{x_n\} \to a^+$ such that $A_n = \{w \in \mathbb{R} : X(w) \leq X_n\} \searrow A = \{w \in \Omega : X(w) \leq a\}$. By the continuity of probability, $F(x_n) = \mathbb{P}(A_n) \to \mathbb{P}(A) = F(a)$.

- **Theorem 2.2.7:** Let $U \sim \text{Unif}[0,1]$, define $Y(w) = F^{-1}(U(w))$, and choose $x, t \in [0,1]$. Case $F^{-1}(t) > x$: $\sup\{y : F(y) < t\} > x \implies F(x) < t$ since F is non-decreasing. Case $F^{-1}(t) \leq x$: $\forall \delta > 0, F(x+\delta) \geq t \implies F(x) \geq t$ since F is right continuous. Thus, $\{t : F^{-1}(t) \leq x\} = \{t : t \leq F(x)\}$, so $\mathbb{P}(Y \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$ since U is uniform.

- **Note:** Always write the support of PMFs and PDFs.

- **Proposition 2.2.11:** Prove that the CDF is uniquely defined for a PMF, and that the random variable is uniquely defined for the CDF.

- **Lemma 2.2.18:** $P(X = x) = \lim\limits_{\delta \to 0} \mathbb{P}(x - \delta \leq X \leq x + \delta) = \lim\limits_{\delta \to 0} \int_{x-\delta}^{x+\delta} f(y)dy = 0$.

- **Theorem 2.2.25:** WLOG, suppose g is strictly increasing. Then, $\mathbb{P}(Y \leq y) = \mathbb{P}(g(x) \leq y) = \mathbb{P}(x \leq g^{-1}(y)) = F_X(g^{-1}(y))$, so $f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$ by the chain rule. For the other case, $\frac{d}{dy} g^{-1}(y)$ is negative.

- **Theorem 2.2.28:** Prove for $n = 2$ only.

- **Definition 2.3.5:** $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ and $f_Y(y) = \int f(x,y)dx = \int f_{Y|X}(y|x) f_X(x)dx$. Also, $f_Y(y)dy$ in the denominator of equation 2.31 should be $f_X(x)dx$.

- **Example 2.3.6:** Sketch of proof: $f(y|x) \propto f(x|y)f(y) \propto \exp(-\frac{(x-y)^2}{2\sigma^2}) \exp(-\frac{(y-\mu)^2}{2\tau^2})$.

Questions

- **2.1:**
  - Proposition 2.2.11: Let X, Y have the same PMF p and let F_X, F_Y be their respective CDFs. Then, $F_X(x) - F_Y(x) = \Sigma_{t \in D: t \leq x} p(t) - \Sigma_{t \in D: t \leq x} p(t) = \Sigma_{t \in D: t \leq x} (p(t) - p(t)) = \Sigma_{t \in D: t \leq x} 0 = 0 \implies F_X = F_Y$. Thus, by Theorem 2.2.3, X, Y have the same measure μ and thus the distribution.
  - Proposition 2.2.19: Let X, Y have the same PDF f and let F_X, F_Y be their respective CDFs. Then, $F_X(x) - F_Y(x) = \int_{-\infty}^x f(t)dt - \int_{-\infty}^x f(t)dt = \int_{-\infty}^x (f(t) - f(t))dt = \int_{-\infty}^x 0 dt = 0 \implies F_X = F_Y$. Thus, by Theorem 2.2.3, X, Y have the same measure μ and thus the distribution.
  - Theorem 2.2.25: Suppose g is strictly decreasing. Then, $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$. Thus, $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(1 - F_X(g^{-1}(y))) = (-1) f_X(g^{-1}(y))(g^{-1}(y))' = f_X(g^{-1}(y)) |g^{-1}(y))'|$ since $g'(y) < 0 \implies (g^{-1})'(y) < 0$.

- **2.2:** Sketch of proof: Only do this for $n = 2$.

- **2.3:** $f_{Y|X}(y|x) \propto \mathbb{P}_{X|Y}(x|y) f_Y(y) \propto \binom{n}{x} y^x (1-y)^{n-x} y^{a-1} (1-y)^{b-1} \propto y^{x+a-1}(1-y)^{n-x+b-1}$, which means that $Y|X \sim \text{Beta}(x+a, n-x+b)$.

- **2.5:** Sketch of proof: Notice that $f(x_1) = 1$ and $f(x_2|x_1) = \frac{1}{1-x_1}$ for $x_2 \in (x_1, 1)$, so $f(x_2) = \int_0^{x_2} f(x_2|x_1) f(x_1) dx_1 = \frac{1}{1-x_1}\Big|_0^{x_2} = -\ln(1-x_2)$ for $x_2 \in (0,1)$. Similarly, $f(x_3|x_2) = \frac{1}{1-x_2}$ for $x_3 \in (x_2, 1)$, so $f(x_3) = \frac{1}{2}(\ln(1-x_3))^2$ for $x_3 \in (0,1)$. Use induction to yield that $f(x_n) = \frac{(-1)^{n-1}}{(n-1)!}(\ln(1-x_n))^{n-1}$ for $x_n \in (0,1)$.

- **2.10:** If $a \le b < c \le d$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \iff 0 = (b-a)(d-c) \iff a = b$ or $c = d$. If $a \le c \le d \le b$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \iff d-c = (b-a)(d-c) \iff (d-c)(b-a-1) = 0 \iff c = d$ or $a = 0, b = 1$. If $a \le c \le b \le d$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \iff b - c = (b-a)(d-c)$.

July 9

Notes

- **Note:** A corollary of Theorem 2.1.4 is that for a PMF p and PDF f, $p(x_1, \dots, x_n) = \Pi_i p(x_i)$ and $f(x_1, \dots, x_n) = \Pi_i f(x_i)$.

- **Note:** $\forall w \in \Omega$, if $X(w) = Y(w)$, they are pointwise equal, and if $\mathbb{P}(\{w \in \Omega : X(w) = Y(w)\}) = 1$, they are equal a.s.

- **Note:** $\limsup\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty}(\sup\limits_{m\ge n} x_m)$ and $\liminf\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty}(\inf\limits_{m\ge n} x_m)$. Since $\sup\limits_{m\ge n} x_m \ge \sup\limits_{m\ge n+1} x_m$, $\{\sup\limits_{m\ge n} x_m\}_n$ is monotonic and the former limit always exists. Similarly, since $\inf\limits_{m\ge n} x_m \le \inf\limits_{m\ge n+1} x_m$, $\{\inf\limits_{m\ge n} x_m\}_n$ is monotonic and the latter limit always exists.

- **Definition 3.1.1:** $\{A_n$ infinitely often$\} = \{\forall n \in \mathbb{N}^+, \exists k \ge n$ s.t. A_k occurs$\}$ and $\{A_n$ almost always$\} = \{\exists n \in \mathbb{N}^+, \forall k \ge n$ s.t. A_k occurs$\}$. For both lim sup and lim inf, the equalities hold since the $\forall$ corresponds with $\cap$ and the $\exists$ corresponds with $\cup$.

- **Corollary 3.1.2:** The proof directly follows from De Morgan's laws.

- **Proposition 3.1.3:** Proof of left inequality: $\cap_{k=n}^{\infty} A_k \subseteq \cap_{k=n+1}^{\infty} A_k$, so $\mathbb{P}(A_n \text{ a.a.}) = \mathbb{P}(\cup_n \cap_{k=n}^{\infty} A_k) = \mathbb{P}(\lim\limits_{n\to\infty} \cap_{k=n}^{\infty} A_k) = \lim\limits_{n\to\infty} \mathbb{P}(\cap_{k=n}^{\infty} A_k)$ by the continuity of probability $= \liminf\limits_{n\to\infty} \mathbb{P}(\cap_{k=n}^{\infty} A_k)$ since the limit is equal to its infimum if it exists $\le \liminf\limits_{n\to\infty} \mathbb{P}(A_n)$. The middle inequality holds since inf $\le$ sup by definition.

- **Theorem 3.1.4:**
  - i) Note that $\cup_{k=n+1} A_k \subseteq \cup_{k=n} A_k$. Hence, $\mathbb{P}(A_n \text{ i.o}) = \mathbb{P}(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k) = \mathbb{P}(\lim\limits_{n\to\infty} \cup_{k=n}^{\infty} A_k) = \lim\limits_{n\to\infty} \mathbb{P}(\cup_{k=n}^{\infty} A_k)$ by the continuity of probability $\le \lim\limits_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0$ since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.
  - ii) Note that $\cap_{k=n} A_k^c \subseteq \cap_{k=n+1} A_k^c$. Hence, using the trick that $1 - x \le e^{-x}\ \forall x \in \mathbb{R}$, $1 - \mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}((A_n \text{ i.o.})^c) = \mathbb{P}(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c)$ by De Morgan's laws $= \mathbb{P}(\lim\limits_{n\to\infty} \cap_{k=n}^{\infty} A_k^c) = \lim\limits_{n\to\infty} \mathbb{P}(\cap_{k=n}^{\infty} A_k^c)$ by the continuity of probability $= \lim\limits_{n\to\infty} \Pi_{k=n}^{\infty}(1 - \mathbb{P}(A_k))$ since the $\{A_n\}$ are independent $\le \lim\limits_{n\to\infty} \Pi_{k=n}^{\infty} \exp\{-\mathbb{P}(A_k)\} = \lim\limits_{n\to\infty} \exp\{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)\} = e^{-\infty} = 0$.

- **Definition 3.2.1:** This is equivalent to $\mathbb{P}(\{w \in \Omega : \lim\limits_{n\to\infty} X_n(w) = X(w)\}) = 1$.

- **Proposition 3.2.2:** $\lim\limits_{n\to\infty} X_n = X \iff \forall \varepsilon > 0, |X_n - X| < \varepsilon$ for all but finitely many n (almost always). Then, $\mathbb{P}(\lim\limits_{n\to\infty} X_n = X) = \mathbb{P}(\forall \varepsilon > 0, |X_n - X| < \varepsilon \text{ a.a.}) = 1 - \mathbb{P}(\exists \varepsilon > 0 \text{ s.t. } |X_n - X| \ge \varepsilon \text{ i.o.})$. Notice that $\exists \varepsilon > 0$ s.t. $|X_n - X| \ge \varepsilon$ i.o. $\implies \exists \varepsilon \in \mathbb{Q}^+$ s.t. $|X_n - X| \ge \varepsilon$ i.o. Thus, $\mathbb{P}(\exists \varepsilon > 0 \text{ s.t. } |X_n - X| \ge \varepsilon \text{ i.o.}) \le \mathbb{P}(\exists \varepsilon \in \mathbb{Q}^+ \text{ s.t. } |X_n - X| \ge \varepsilon \text{ i.o}) \le \sum_{\varepsilon \in \mathbb{Q}^+} \mathbb{P}(|X_n - X| \ge \varepsilon \text{ i.o}) = 0$ by assumption and since $\mathbb{Q}^+$ is countable.

- **Corollary 3.2.3:** The proof follows from Borel-Cantelli and Proposition 3.2.2.

- **Definition 3.2.4:** This is equivalent to $\lim\limits_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

- **Proposition 3.2.5:** $\forall \varepsilon > 0$, define $E_n = \{w \in \Omega : \exists m \geq n \text{ s.t. } |X_m(w) - X(w)| > \varepsilon\}\ \forall n \in \mathbb{N}^+$. Notice that $E_{n+1} \subseteq E_n$ and that $w \in \cap_{n=1}^{\infty} E_n \implies X_n \nrightarrow X$. Hence, $\lim\limits_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq \lim\limits_{n\to\infty} \mathbb{P}(E_n) = \mathbb{P}(\lim\limits_{n\to\infty} E_n)$ by the continuity of probability $= \mathbb{P}(\cap_{n=1}^{\infty} E_n) \leq \mathbb{P}(X_n \nrightarrow X) = 1 - \mathbb{P}(X_n \to X) = 0$.

- **Theorem 3.2.7:** Convergence in probability $\iff \forall k \in \mathbb{N}, \exists n_k$ s.t. $\forall n \geq n_k, \mathbb{P}(|X_n - X| > 2^{-k}) \leq 2^{-k}$ (2^{-k} is the ε here). Choose a subsequence s.t. $n_{k+1} > n_k$ and define $A_k = \{w \in \Omega : |X_{n_k}(w) - X(w)| > 2^{-k}\}$. Notice that $\sum_{k=1}^{\infty} \mathbb{P}(A_k) \leq \sum_{k=1}^{\infty} 2^{-k} = 1 < \infty$. By Borel-Cantelli, $\mathbb{P}(A_k$ i.o.$) = 0$, so $1 = \mathbb{P}((A_k \text{ i.o.})^c) = \mathbb{P}(\{|X_{n_k}(w) - X(w)| > 2^{-k} \text{ finitely many times}\}) \leq \mathbb{P}(X_{n_k} \to X)$.

- **Theorem 3.2.8:** The proof of i) follows from the definition of a continuous function.

- **Proposition 4.1.7:** iii) only holds for a finite number of random variables.

Questions

- **3.1:** Theorem 3.2.8: ii) Since f is continuous, $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $|X_n - X| \leq \delta \implies |f(X_n) - f(X)| \leq \varepsilon$. By assumption, $1 = \lim\limits_{n\to\infty} \mathbb{P}(|X_n - X| \leq \delta) \leq \lim\limits_{n\to\infty} \mathbb{P}(|f(X_n) - f(X)| \leq \varepsilon)$.

- **3.5:** Observe that $\forall \varepsilon > 0$, $\lim\limits_{n\to\infty} X_n = X \implies \lim\limits_{n\to\infty} |X_n - X| = 0 \implies \lim\limits_{n\to\infty} |X_n - X| < \varepsilon \implies \{\exists k \text{ s.t. } \forall n \geq k, |X_n - X| < \varepsilon\} \implies |X_n - X| < \varepsilon$ a.a. Taking the probability of these yields $\mathbb{P}(|X_n - X| < \varepsilon$ a.a.$) \geq 1$.

- **3.10:** $\{w \in \Omega : \Sigma_{i=1}^{\infty} X_i(w) = \infty\} \iff (\exists k \in \mathbb{N} \text{ s.t. } \forall n \in \mathbb{N}, \exists i \geq n \text{ s.t. } X_i \geq \frac{1}{k}) \iff \cup_{k=1}^{\infty} \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} \{X_i \geq \frac{1}{k}\} = A$. For K such that $\delta \geq \frac{1}{K}$, define $A_K = \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} \{X_i(w) \geq \frac{1}{K}\} \subseteq A$. Notice that $\mathbb{P}(X_i \geq \frac{1}{K}) \geq \mathbb{P}(X_i \geq \delta) \geq \varepsilon \implies \Sigma_{i=1}^{\infty} \mathbb{P}(X_i \geq \frac{1}{K}) \geq \Sigma_{i=1}^{\infty} \varepsilon = \infty$, so by Borel-Cantelli, $1 = \mathbb{P}(X_i \geq \frac{1}{K} \text{ i.o.}) = \mathbb{P}(A_K) \leq \mathbb{P}(A)$. Alternatively, notice that $X_i \geq \delta$ i.o. $\implies \Sigma_{i=1}^{\infty} X_i = \infty$.

July 11

Notes

- **Note:** $\mathbb{E}(X|Y)$ can be informally thought of as a random variable with Y as input: $\mathbb{E}(X|Y) = \mathbb{E}(X|Y(w)) = \mathbb{E}(X|Y)(w)$.

- **Theorem 4.2.3:** Proof for the continuous case: $\mathbb{E}(\mathbb{E}(X|Y)) = \int_{\Omega_y} \mathbb{E}(X|Y = y) f(y) dy$ by definition $= \int_{\Omega_y} \int_{\Omega_x} x \frac{f(x,y)}{f(y)} dx f(y) dy = \int_{\Omega_y} \int_{\Omega_x} x \frac{f(x,y)}{f(y)} f(y) dx dy = \int_{\Omega_x} \int_{\Omega_y} x f(x,y) dy dx = \int_{\Omega_x} x \int_{\Omega_y} f(x,y) dy dx = \int_{\Omega_x} x f(x) dx = \mathbb{E}(X)$.

- **Proposition 4.2.5:** $\mathbb{P}(X + Y \leq z) = \mathbb{E}(\mathbb{I}(X + Y \leq z)) = \mathbb{E}(\mathbb{E}(\mathbb{I}(X \leq z - Y)|Y))$ by Theorem 4.2.3 $= \int_{\Omega_Y} \mathbb{E}(\mathbb{I}(X \leq z - y)|Y = y) f_Y(y) dy = \int_{\Omega_Y} \mathbb{E}(\mathbb{I}(X \leq z - y)) f_Y(y) dy$ since X and Y are independent $= \int_{\Omega_Y} \mathbb{P}(X \leq z - y) f_Y(y) dy = \int_{\Omega_Y} F_X(z - y) f(y) dy$. Alternatively, $\mathbb{P}(X + Y \leq z) = \mathbb{P}(X \leq z - Y) = \int_{\Omega_Y} \mathbb{P}(X \leq z - y | Y = y) f_Y(y) dy = \int_{\Omega_Y} \mathbb{P}(X \leq z - y) f_Y(y) dy$ since X and Y are independent $= \int_{\Omega_Y} F_X(z - y) f(y) dy$.

- **Note:** An identity related to Proposition 4.2.5 is $f(z) = \int_{\Omega_Y} f_X(z - y) f_Y(y) dy$ for $Z = X + Y$. Proof: $f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \mathbb{P}(X + Y \leq z) = \frac{d}{dz} \int_{\Omega_Y} F_X(z - y) f_Y(y) dy$ by Proposition 4.2.5 $= \int_{\Omega_Y} \frac{d}{dz} F_X(z - y) f(y) dy = \int_{\Omega_Y} f_X(z - y) f(y) dy$.

- **Example 4.3.1:** $\lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} n\mathbb{P}(U(w) \in [0, \frac{1}{n}]) = n\frac{1}{n} = 1$.

- **Lemma 4.3.4:** Define $X_n = X \wedge n$ such that $X_n \leq X_{n+1} \leq X \implies \mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1}) \leq \mathbb{E}(X) \implies \lim_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X)$. Let Y be a bounded random variable such that $0 \leq Y \leq X$ a.s. Thus, $\mathbb{E}(X_n) \geq \mathbb{E}(Y)$ for large enough n, so $\lim_{n\to\infty} \mathbb{E}(X_n) \geq \sup\{\mathbb{E}(Y) : Y \text{ is bounded}, 0 \leq Y \leq X \text{ a.s.}\}$ by Proposition 4.3.3 $= E(X)$. Thus, $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

- **Theorem 4.3.5:** Let $\varepsilon > 0$ and define $G_n = \{|X_n - X| > \varepsilon\}$. Notice that $|\mathbb{E}(X_n) - \mathbb{E}(X)| = |\mathbb{E}(X_n - X)| \leq \mathbb{E}(|X_n - X|)$ by property vi) of Lemma 4.1.8 $= \mathbb{E}(|X_n - X|\mathbb{I}_{G_n}) + \mathbb{E}(|X_n - X|\mathbb{I}_{G_n^c})$. As $\varepsilon \to 0$, $\mathbb{E}(|X_n - X|\mathbb{I}_{G_n}) \leq \mathbb{E}((|X_n| + |X|)\mathbb{I}_{G_n}) \leq 2M\mathbb{E}(\mathbb{I}_{G_n}) = 2M\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ since $X_n \xrightarrow{P} X$, while $\mathbb{E}(|X_n - X|\mathbb{I}_{G_n^c}) \leq \varepsilon\mathbb{E}(\mathbb{I}_{G_n^c}) = \varepsilon\mathbb{P}(|X_n - X| \leq \varepsilon) \to 0$. Thus, $|\mathbb{E}(X_n) - \mathbb{E}(X)| \to 0$ as $\varepsilon \to 0$ and the result follows.

- **Theorem 4.3.6:** Define $Y_n = \inf_{m\geq n} X_m$ so that $X_n \geq Y_n$ a.s. $\implies \liminf_{n\to\infty} \mathbb{E}(X_n) \geq \liminf_{n\to\infty} \mathbb{E}(Y_n)$ and $Y_n \uparrow Y = \liminf_{n\to\infty} X_n$ a.s. Let $M \in \mathbb{R}$. Since $|Y_n \wedge M| \leq M$ and $(Y_n \wedge M) \to (Y \wedge M)$ a.s., by the BCT, $\liminf_{n\to\infty} \mathbb{E}(Y_n) \geq \lim_{n\to\infty} \mathbb{E}(Y_n \wedge M) = \mathbb{E}(Y \wedge M)$. As $M \to \infty$, by Lemma 4.3.4, $\mathbb{E}(Y \wedge M) \to \mathbb{E}(Y)$, so $\liminf_{n\to\infty} \mathbb{E}(X_n) \geq \liminf_{n\to\infty} \mathbb{E}(Y_n) \geq \mathbb{E}(Y) = \mathbb{E}(\liminf_{n\to\infty} X_n)$.

- **Theorem 4.3.7:** $X_n \uparrow X \implies \mathbb{E}(X_n) \leq \mathbb{E}(X) \implies \lim_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X)$. Since $X_n \geq 0$ a.s., by Fatou's lemma, $\lim_{n\to\infty} \mathbb{E}(X_n) = \liminf_{n\to\infty} \mathbb{E}(X_n) \geq \mathbb{E}(\liminf_{n\to\infty} X_n) = \mathbb{E}(X)$. Thus, $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

- **Example 4.3.8:** Let $X_n \geq 0$ a.s. and define $Y_n = \Sigma_{i=1}^n X_i \nearrow Y = \Sigma_{i=1}^\infty X_i$. Since $Y_n \geq 0$ a.s., by the MCT, $\lim_{n\to\infty} \mathbb{E}(Y_n) = \mathbb{E}(Y) \implies \lim_{n\to\infty} \mathbb{E}(\Sigma_{i=1}^n X_i) = \mathbb{E}(\Sigma_{i=1}^\infty X_i) \implies \lim_{n\to\infty} \Sigma_{i=1}^n \mathbb{E}(X_i) = \mathbb{E}(\Sigma_{i=1}^\infty X_n)$ since the sum is finite $\implies \Sigma_{i=1}^\infty \mathbb{E}(X_i) = \mathbb{E}(\Sigma_{i=1}^\infty X_i)$.

- **Theorem 4.3.9:** $|X_n| \leq Y$ a.s. $\implies |X_n| \leq |Y|$ a.s. $\implies X_n + Y \geq 0$ a.s. and $Y - X_n \geq 0$ a.s. Using $X_n + Y \geq 0$ a.s., by Fatou's lemma, $\liminf_{n\to\infty} \mathbb{E}(X_n + Y) \geq \mathbb{E}(\liminf_{n\to\infty}(X_n + Y)) \implies \liminf_{n\to\infty} \mathbb{E}(X_n) + \mathbb{E}(Y) \geq \mathbb{E}(\liminf_{n\to\infty} X_n) + \mathbb{E}(Y) \implies \liminf_{n\to\infty} \mathbb{E}(X_n) \geq \mathbb{E}(\liminf_{n\to\infty} X_n)$ since Y is integrable $= \mathbb{E}(X)$. Using $Y - X_n \geq 0$ a.s., by Fatou's lemma, $\liminf_{n\to\infty} \mathbb{E}(Y - X_n) \geq \mathbb{E}(\liminf_{n\to\infty}(Y - X_n)) \implies \mathbb{E}(Y) + \liminf_{n\to\infty} \mathbb{E}(-X_n) \geq \mathbb{E}(Y) + \mathbb{E}(\liminf_{n\to\infty}(-X_n)) \implies \liminf_{n\to\infty} \mathbb{E}(-X_n) \geq \mathbb{E}(\liminf_{n\to\infty}(-X_n))$ since Y is integrable $\implies -\limsup_{n\to\infty} \mathbb{E}(X_n) \geq \mathbb{E}(-X) \implies \limsup_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X)$. Altogether, $\limsup_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$ implies that they are all equal, yielding the result.

- **Theorem 4.4.2:** Let $\lambda \in (0, \delta)$. Notice that $e^{\lambda|X|} \leq e^{\lambda X} + e^{-\lambda X}$, so $\mathbb{E}(e^{\lambda|X|}) \leq M_X(\lambda) + M_X(-\lambda) < \infty$. Also, the Taylor expansion of $e^{\lambda|X|} = \Sigma_{n=0}^\infty \frac{\lambda^n |X|^n}{n!}$, so define $S_k = \Sigma_{n=0}^k \frac{\lambda^n |X|^n}{n!} \geq 0$ as the partial sum. Since $S_k \nearrow e^{\lambda|X|}$ and $|S_k| = S_k \leq e^{\lambda|X|}$, by either the MCT or DCT, $\mathbb{E}(e^{\lambda|X|}) = \lim_{k\to\infty} \mathbb{E}(S_k) = \lim_{k\to\infty} \mathbb{E}(\Sigma_{n=0}^k \frac{\lambda^n |X|^n}{n!}) = \lim_{k\to\infty} \Sigma_{n=0}^k \mathbb{E}(\frac{\lambda^n |X|^n}{n!})$ by property iii) of Proposition 4.1.7 since the sum is finite $= \Sigma_{n=0}^\infty \frac{\lambda^n \mathbb{E}(|X|^n)}{n!} < \infty$. Thus, $\mathbb{E}(|X^n|) < \infty$. Moving on, notice that $\mathbb{E}(X^n) \leq |\mathbb{E}(X^n)| \leq \mathbb{E}(|X^n|) < \infty$ by property vi) of Lemma 4.1.8, implying that $\Sigma_{n=0}^\infty \frac{\lambda^n \mathbb{E}(X^n)}{n!}$ is absolutely convergent for $\lambda \in (-\delta, \delta)$. Then, $|M_X(\lambda) - \mathbb{E}(\Sigma_{n=0}^k \frac{\lambda^n X^n}{n!})| \leq \Sigma_{n=k+1}^\infty \frac{\lambda^n \mathbb{E}(|X|^n)}{n!} \to 0$ as $k \to \infty$ since $\Sigma_{n=0}^\infty \frac{\lambda^n \mathbb{E}(|X|^n)}{n!} < \infty$. Thus, $M_X(\lambda) = \Sigma_{n=0}^\infty \frac{\lambda^n \mathbb{E}(X^n)}{n!}$ and $M_X^{(n)}(0) = \mathbb{E}(X^n)$.

Questions

- **4.1:**

- Lemma 4.1.8: iv) $Y - X \geq 0$ a.s. $\implies \mathbb{E}(Y - X) \geq 0 \implies \mathbb{E}(Y + (-X)) = \mathbb{E}(Y) + \mathbb{E}(-X) = \mathbb{E}(Y) - \mathbb{E}(X) \geq 0$. v) $\mathbb{E}(Y - X) = \mathbb{E}(Y + (-X)) = \mathbb{E}(Y) + \mathbb{E}(-X) = \mathbb{E}(Y) - \mathbb{E}(X) = \mathbb{E}(X) - \mathbb{E}(X) = 0$. vi) $\mathbb{E}(X) \leq \mathbb{E}(|X|)$ since $X \leq |X|$ and $\mathbb{E}(X) \geq \mathbb{E}(-|X|) = -\mathbb{E}(|X|)$ since $X \geq -|X|$.
- Lemma 4.4.4: $M_S(\lambda) = \mathbb{E}(\exp\{\lambda \Sigma_{i=1}^n X_i\}) = \mathbb{E}(\Pi_{i=1}^n \exp\{\lambda X_i\}) = \Pi_{i=1}^n \mathbb{E}(\exp\{\lambda X_i\})$ by Theorem 4.1.12 since $M_{X_i}(\lambda) < \infty$ and functions of independent random variables are independent.

- **4.5:** The MGF of the Poisson distribution is $M_X(t) = \mathbb{E}(e^{tX}) = \exp(\lambda(e^t - 1))$ for some $X \sim \text{Poisson}(\lambda)$. Notice that $\forall t, M_X(t) < \infty$. By Lemma 4.4.4, for $S = \Sigma_{i=1}^n X_i$, $M_S(t) = \Pi_{i=1}^n M_{X_i}(t) = \Pi_{i=1}^n \exp(\lambda(e^t - 1)) = \exp(\Sigma_{i=1}^n \lambda(e^t - 1)) = \exp(n\lambda(e^t - 1))$, which is the MGF for some $Y \sim \text{Poisson}(n\lambda)$. By Theorem 4.4.3, S and Y have the same distribution.

- **4.6:** The MGF of the exponential distribution is $M_X(t) = \mathbb{E}(e^{tX}) = \frac{\lambda}{\lambda - t}$ for some $X \sim \text{Exp}(\lambda)$. Notice that $\forall t < \lambda, M_X(t) < \infty$. By Lemma 4.4.4, for $S = \Sigma_{i=1}^n X_i$, $M_S(t) = \Pi_{i=1}^n M_{X_i}(t) = \Pi_{i=1}^n \frac{\lambda}{\lambda - t} = (\frac{\lambda}{\lambda - t})^n = (\frac{\lambda - t}{\lambda})^{-n} = (1 - \frac{t}{\lambda})^{-n}$, which is the MGF for some $Y \sim \text{Expo}(n, \lambda)$. Additionally, $\forall t < \lambda, M_S(t) < \infty$. By Theorem 4.4.3, S and Y have the same distribution.

- **4.8:** Define $Y = X^2$, so $F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \implies f_Y(y) = \frac{d}{dy}[\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx - \int_0^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx] = \frac{1}{\sqrt{2\pi y}} e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{(1/2)-1} e^{-(1/2)y}$. By Proposition 2.2.19, $Y \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$, and its MGF is $M_Y(\lambda) = \mathbb{E}(e^{\lambda Y}) = (1 - \frac{\lambda}{1/2})^{-1/2} < \infty$ for $\lambda < \frac{1}{2}$. By Lemma 4.4.4, $M_{\chi^2}(\lambda) = \Pi_{i=1}^n M_{Y_i}(\lambda) = (1 - \frac{\lambda}{1/2})^{-n/2}$, which is the MGF for some $Z \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$. Additionally, $\forall \lambda < \frac{1}{2}, MGF_{\chi^2}(\lambda) < \infty$. By Theorem 4.4.3, Z and χ^2 have the same distribution.

- **4.9:**

  i) $(X - \mathbb{E}X)^2 \geq 0$ a.s. $\implies \mathbb{E}(X - \mathbb{E}X)^2 \geq 0$.

  ii) $\mathbb{E}(cX - \mathbb{E}(cX))^2 = \mathbb{E}(c(X - \mathbb{E}X))^2 = c^2 \mathbb{E}(X - \mathbb{E}X)^2$.

  iii) $\mathbb{E}(X + Y - \mathbb{E}(X + Y))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2\mathbb{E}X\mathbb{E}Y - 2\mathbb{E}X\mathbb{E}Y$.

  iv) $\mathbb{E}[(aX + bY - \mathbb{E}(aX + bY))(Z - \mathbb{E}Z)] = a\mathbb{E}(XZ - Z\mathbb{E}X - X\mathbb{E}Z + \mathbb{E}X\mathbb{E}Z) + b\mathbb{E}(YZ - Z\mathbb{E}Y - Y\mathbb{E}Z + \mathbb{E}Y\mathbb{E}Z) = a\mathbb{E}[(X - \mathbb{E}X)(Z - \mathbb{E}Z)] + b\mathbb{E}[(Y - \mathbb{E}Y)(Z - \mathbb{E}Z)]$.

- **4.11:** By question 4.10, $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 \iff (X - \mathbb{E}(X))^2 = 0$ a.s. $\iff X - \mathbb{E}(X) = 0$ a.s. $\iff X = \mathbb{E}(X)$ a.s. The result follows if $a = \mathbb{E}(X)$.

- **4.12:** $\mu = \mathbb{E}(X) = \mathbb{E}(X\mathbb{I}(X < \mu)) + \mathbb{E}(X\mathbb{I}(X \geq \mu) < \mathbb{E}(\mu\mathbb{X} < \mu) + \mathbb{E}(X\mathbb{X} \geq \mu) \leq \mu\mathbb{P}(X < \mu) + \mathbb{E}(X\mathbb{X} \geq \mu) \leq \mu + \mathbb{E}(X\mathbb{I})$. Notice that $\mathbb{E}(\mathbb{I}(X \geq \mu)) \implies \mathbb{I}(X \geq \mu) = 0$ a.s. by question 4.10 $\implies \mathbb{E}(X\mathbb{I}(X \geq \mu) = 0$, so $\mu < \mu$, which is a contradiction. Alternatively, $X < \mathbb{E}(X)$ a.s. $\implies \mathbb{E}(X) < \mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$, a contradiction.

- **4.16:** Define $X_n = \Sigma_{k=1}^n \mathbb{I}(X \geq k) \nearrow X$. Since $X_n \geq 0$ a.s., by the MCT, $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$. Notice that $\lim\limits_{n\to\infty} \mathbb{E}(X_n) = \lim\limits_{n\to\infty} \mathbb{E}(\Sigma_{k=1}^n \mathbb{I}(X \geq k)) = \lim\limits_{n\to\infty} \Sigma_{k=1}^n \mathbb{E}(\mathbb{I}(X \geq k))$ by property iii) of Proposition 4.1.7 since the sum is finite $= \lim\limits_{n\to\infty} \Sigma_{k=1}^n \mathbb{P}(X \geq k) = \Sigma_{k=1}^\infty \mathbb{P}(X \geq k)$. Thus, $\mathbb{E}(X) = \Sigma_{k=1}^\infty \mathbb{P}(X \geq k)$.

- **4.17:** Assuming X has a PDF f_X, we have $\int_0^\infty px^{p-1}\mathbb{P}(X \geq x)dx = \int_0^\infty px^{p-1}\mathbb{E}[\mathbb{I}(X \geq x)]dx = \int_0^\infty px^{p-1} \int_{-\infty}^\infty \mathbb{I}(y \geq x) f_X(y) dy dx = \int_0^\infty \int_{-\infty}^\infty px^{p-1}\mathbb{I}(y \geq x) f_X(y) dy dx = \int_0^\infty \int_x^\infty px^{p-1} f_X(y) dy dx = \int_0^\infty \int_0^y px^{p-1} f_X(y) dx dy$ by Fubini's theorem $= \int_0^\infty [\int_0^y px^{p-1} dx] f_X(y) dy = \int_0^\infty y^p f_X(y) dy = \mathbb{E}(X^p)$.

July 16

Notes

- **Note:** $\{a_n \to a\}$ for a sequence $\{a_n\}_n$ is a deterministic event: the probability of it occurring is either 0 or 1. In particular, $a_n \to a \implies \mathbb{P}(E \cap \{a_n \to a\}) = \mathbb{P}(E \cap \Omega)$ for $E \in \Omega$.
- **Theorem 5.1.1:** $X \geq a\mathbb{I}(X \geq a)$ a.s. $\implies \mathbb{E}(X) \geq \mathbb{E}(a\mathbb{I}(X \geq a)) = a\mathbb{P}(X \geq a)$.
- **Corollary 5.1.2:** The proof follows by using Markov's inequality on $(X - \mathbb{E}(X))^2$. Note that the variance must exist for Chebyshev's inequality to hold.
- **Note:** Variance is finite $\implies$ expectation is finite.
- **Corollary 5.1.3:** $\forall \lambda > 0, \mathbb{P}(X - \mathbb{E}(X) \geq t) = \mathbb{P}(e^{\lambda(X - \mathbb{E}(X))} \geq e^{\lambda t})$ since e^x is monotonic $\leq M_{X-\mathbb{E}(X)}(\lambda)e^{-\lambda t}$ by Markov's inequality $\implies \mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \inf\limits_{\lambda > 0}\{M_{X-\mathbb{E}(X)}(\lambda)e^{-\lambda t}\}$. Note that $M_{X-\mathbb{E}(X)}(\lambda)$ can be infinite.
- **Lemma 5.1.5:** Since $Y \in [a, b]$ a.s., define $\alpha = \frac{b-Y}{b-a}$ such that $Y = \alpha a + (1-\alpha)b$. Notice that $e^{\lambda Y}$ is convex with respect to Y, so $e^{\lambda Y} = \exp(\lambda[\alpha a + (1-\alpha)b]) \leq \alpha e^{\lambda a} + (1-\alpha)e^{\lambda b} = \frac{b-Y}{b-a}e^{\lambda a} + \frac{Y-a}{b-a}e^{\lambda b} \implies \mathbb{E}(e^{\lambda Y}) \leq \frac{b}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b}$ since $\mathbb{E}(Y) = 0$. Next, define $P = \frac{b}{b-a}$ and $u = \lambda(b-a)$. Consider $\varphi(u) = \log(pe^{\lambda a} + (1-p)e^{\lambda b}) = \lambda a + \log(p + (1-p)e^{\lambda(b-a)}) = u(p-1) + \log(p + (1-p)e^u)$. Its second order Taylor approximation is $\varphi(0) + \varphi'(0)u + \frac{1}{2}\varphi''(\xi)u^2 \leq \frac{1}{8}u^2$ for $\xi \in (0, u)$ since $\varphi(0) = 0$, $\varphi'(u) = (p-1) + \frac{(1-p)e^u}{p+(1-p)e^u} \implies \varphi'(0) = 0$, and $\varphi''(u) = \frac{p(1-p)e^u}{(p+(1-p)e^u)^2} \implies \varphi''(u) \leq \frac{1}{4}$. Thus, $\varphi(u) \leq \frac{u^2}{8} \implies \mathbb{E}(e^{\lambda Y}) \leq e^{\varphi(u)} \leq e^{\frac{u^2}{8}}$.
- **Theorem 5.1.6:** First, by Chernoff's inequality, $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \inf\limits_{\lambda \geq 0} e^{-\lambda t}M_{S_n - \mathbb{E}(S_n)}(\lambda) \leq \inf\limits_{\lambda \geq 0} \exp(-\lambda t + \frac{\lambda^2}{8}\Sigma_{i=1}^n(b_i - a_i)^2)$ since $M_{S_n-\mathbb{E}(S_n)}(\lambda) = \Pi_{i=1}^n M_{X_i - \mathbb{E}(X_i)}(\lambda) \leq \Pi_{i=1}^n \exp(\frac{\lambda^2}{8}(b_i - a_i)^2) = \exp(\frac{\lambda^2}{8}\Sigma_{i=1}^n(b_i - a_i)^2)$ by Lemma 5.1.5. Notice that the minimum of the set is achieved when $\lambda = 4t[\Sigma_{i=1}^n(b_i - a_i)^2]^{-1}$, so plugging in this λ yields $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp(-2t^2[\Sigma_{i=1}^n(b_i - a_i)^2]^{-1})$. Repeat the argument with $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq -t)$ applied to $-X_1, \ldots, -X_n$.
- **Proposition 5.2.1:** f is convex $\implies$ there exists $g(x) = ax + b$ s.t. $g \leq f$ and $g(\mathbb{E}(X)) = f(\mathbb{E}(X))$. In other words, g is the tangent of f at $\mathbb{E}(X)$. Then, $\mathbb{E}(f(X)) \geq \mathbb{E}(g(X)) = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = g(\mathbb{E}(X)) = f(\mathbb{E}(X))$.
- **Lemma 5.2.3:** $q/p > 1 \implies x^{\frac{q}{p}}$ is convex $\implies (\mathbb{E}|X|^p)^{\frac{q}{p}} \leq \mathbb{E}((|X|^p)^{\frac{q}{p}}) = \mathbb{E}(|X|^q)$ by Jensen's inequality.

Questions

- **5.2:** $\Sigma_{n=1}^\infty \mathbb{P}(X_n \geq n) \leq \Sigma_{n=1}^\infty \mathbb{P}(|X_n| \geq n) \leq \Sigma_{n=1}^\infty \frac{\mathrm{Var}(X_n)}{n^2}$ by Chebyshev's inequality $= \Sigma_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6} < \infty$. By Borel-Cantelli, $\mathbb{P}(X_n \geq n \text{ i.o.}) = 0$.
- **5.4:** Sketch of proof: Show that $\log(\lambda x + (1-\lambda)y) \geq \lambda\log(x) + (1-\lambda)\log(y)$.
- **5.5:** It suffices to show that $f(x) = \max\{x, 0\}$ is convex since $\max\{x, a\} = \max\{x - a, 0\}$. For $x, y \leq 0$, $f(\lambda x + (1-\lambda)y) = 0 \leq \lambda f(x) + (1-\lambda)f(y) = 0$. For $x, y > 0$, $f(\lambda x + (1-\lambda)y) = \lambda x + (1-\lambda)y \leq f(\lambda x) + f((1-\lambda)y) = \lambda x + (1-\lambda)y$. For $x \leq 0, y > 0$, either $\lambda x + (1-\lambda)y \leq 0$, in which case $f(\lambda x + (1-\lambda)y) = 0 \leq \lambda f(x) + (1-\lambda)f(y) = (1-\lambda)y$ since $f(x) = 0$ and $f(y) = y$, or $\lambda x + (1-\lambda)y > 0$, in which case $f(\lambda x + (1-\lambda)y) = \lambda x + (1-\lambda)y \leq (1-\lambda)y$ since $x < 0$. The result directly follows from Jensen's inequality.

July 25

Notes

- **Proposition 5.2.5:** Either $(\mathbb{E}|X|^p)^{\frac{1}{p}} = 0$ or $(\mathbb{E}|X|^q)^{\frac{1}{q}} = 0 \implies |X| = 0$ a.s. by question 4.10 $\implies |XY| = 0$ a.s. $\implies \mathbb{E}(XY) = 0 \leq 0$. If $(\mathbb{E}|X|^p)^{\frac{1}{p}} > 0$ and $(\mathbb{E}|X|^q)^{\frac{1}{q}} > 0$, consider the function $f(x) = \frac{1}{p}x^p + \frac{1}{q}y^q - xy$ for $x, y \geq 0$. Notice that $f'(x) = x^{p-1} - y$ and $f''(x) = (p-1)x^{p-2} \geq 0$ since $p > 1$ and $x \geq 0$, so f is convex and achieves its minimum at $x = y^{\frac{1}{p-1}}$. Then, $f(y^{\frac{1}{p-1}}) = \frac{1}{p}y^{\frac{p}{p-1}} + \frac{1}{q}y^q - y^{\frac{1}{p-1}+1} = y^q(\frac{1}{p} + \frac{1}{q}) - y^q = 0 \implies xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q$ for $x, y \geq 0$ since $\frac{1}{p} + \frac{1}{q} = 1$. Thus, $\frac{|X|}{(\mathbb{E}|X|^p)^{1/p}}\frac{|Y|}{(\mathbb{E}|Y|^q)^{1/q}} \leq \frac{1}{p}(\frac{|X|}{(\mathbb{E}|X|^p)^{1/p}})^p + \frac{1}{q}(\frac{|Y|}{(\mathbb{E}|Y|^q)^{1/q}})^q \implies \frac{\mathbb{E}(XY)}{(\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q}} \leq \frac{1}{p}\mathbb{E}(\frac{|X|}{(\mathbb{E}|X|^p)^{1/p}})^p + \frac{1}{q}\mathbb{E}(\frac{|Y|}{(\mathbb{E}|Y|^p)^{1/p}}) = \frac{1}{p} + \frac{1}{q} = 1$.

- **Proposition 5.2.7:** Let $q = \frac{p}{p-1}$ so that $\frac{1}{p} + \frac{1}{q} = \frac{1}{p} + \frac{p-1}{p} = 1$. By Holder's inequality, $\mathbb{E}(|X||X + Y|^{p-1}) \leq (\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^{q(p-1)})^{\frac{1}{q}} = (\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^p)^{\frac{p-1}{p}}$. Thus, $\mathbb{E}|X+Y|^p = \mathbb{E}(|X+Y||X + Y|^{p-1}) \leq \mathbb{E}(|X||X+Y|^{p-1}+|Y||X+Y|^{p-1})$ by the triangle inequality $\leq ((\mathbb{E}|X|^p)^{\frac{1}{p}}+(\mathbb{E}|Y|^p)^{\frac{1}{p}})(\mathbb{E}|X+Y|^p)^{\frac{p-1}{p}}$ by the previous result.

- **Proposition 5.3.2:** $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p}\mathbb{E}|X_n - X|^p$ by Markov's inequality $\to 0$ by assumption. For a counterexample for the converse, define $U \sim \text{Unif}(0, 1)$ and $X_n = 2^n\mathbb{I}(U \in [0, \frac{1}{n}])$. Notice that $X_n \xrightarrow{P} 0$, but $\mathbb{E}|X_n|^p = \frac{2^{np}}{n} \to \infty$ as $n \to \infty$ for any $p > 0$.

- **Lemma 5.3.3:** Let $p \geq 1$. By the reverse triangle inequality, $|\mathbb{E}|X_n| - \mathbb{E}|X||^p \leq (\mathbb{E}|X_n - X|)^p \leq E|X_n - X|^p$ by Jensen's inequality since x^p is convex $\to 0 \implies \mathbb{E}|X_n| - \mathbb{E}|X| \to 0$.

- **Proposition 5.3.4:** By Holder's inequality, $||X||_p = ||X \cdot 1||_p \leq ||1||_{\frac{1}{1/p-1/q}}||X||_q = ||X||_q$.

- **Theorem 6.1.1:** $\mathbb{E}(\frac{1}{n}S_n) = \mu$ since n is finite and $\text{Var}(\frac{1}{n}S_n) \leq \frac{\sigma^2}{n}$. By Chebyshev's inequality, $\mathbb{P}(|\frac{1}{n}S_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$.

- **Theorem 6.1.2:** WLOG, assume $\mu = 0$. Notice that $\mathbb{E}(X_i - \mu)^2 = \mathbb{E}[(X_i - \mu)^2\mathbb{I}((X_i - \mu)^2 \geq 1)]+\mathbb{E}[(X_i-\mu)^2\mathbb{I}((X_i-\mu)^2 < 1)] \leq \mathbb{E}(X_i-\mu)^2+1 \implies \mathbb{E}(X_i-\mu)^2 \leq a+1$. Also, $\mathbb{E}(S^4) = \mathbb{E}(\Sigma_{i=1}^n X_i^4 + k_1\Sigma_{i=1}^n\Sigma_{j\neq i}X_i^3X_j+k_2\Sigma_{i=1}^n\Sigma_{j\neq i}X_i^2X_j^2+k_3\Sigma_{i=1}^n\Sigma_{j\neq i}\Sigma_{k\neq i,j}X_i^2X_jX_k+k_4\Sigma_{i=1}^n\Sigma_{j\neq i}\Sigma_{k\neq i,j}\Sigma_{\ell\neq i,j,k}X_iX_jX_kX_\ell) = \mathbb{E}(\Sigma_{i=1}^n X_i^4 + k_2\Sigma_{i=1}^n\Sigma_{j\neq i}X_i^2X_j^2)$ since $\mathbb{E}(X_i) = \mu = 0$ and the X_i's are independent $= \Sigma_{i=1}^n\mathbb{E}(X_i^4) + k_2n(n-1)\mathbb{E}(X_i^2)\mathbb{E}(X_j^2) \leq na + k_2n(n-1)(a+1)^2 \leq Kn^2$ for some constants k_1, k_2, k_3, k_4, and K. Thus, $\forall\varepsilon > 0, \mathbb{P}(|\frac{1}{n}S_n| \geq \varepsilon) = \mathbb{P}(S_n^4 \geq n^4\varepsilon^4) \leq \frac{1}{n^4\varepsilon^4}\mathbb{E}(S_n^4)$ by Markov's inequality $\leq \frac{Kn^2}{n^4\varepsilon^4} < \infty \implies \mathbb{P}(|\frac{1}{n}S_n| \geq \varepsilon \text{ i.o.}) = 0$ by Borel-Cantelli $\implies \frac{1}{n}S_n \to 0$ a.s.

- **Theorem 6.2.2:** Since $\mathbb{E}(\frac{1}{n}S_n) = \mu$, $\mathbb{E}(\frac{1}{n}S_n - \mu)^2 = \text{Var}(\frac{1}{n}S_n) \leq \frac{\sigma^2}{n} \to 0$ as $n \to \infty$. Note that this is equivalent to $\frac{1}{n}S_n \xrightarrow{L^2} \mu$, which implies $\frac{1}{n}S_n \xrightarrow{P} \mu$.

- **Theorem 6.2.3:** Let $\varepsilon > 0$. Define $\bar{X}_k^{(n)} = X_k\mathbb{I}(|X_n| \leq n)$ and $\bar{S}_n = \Sigma_{k=1}^n\bar{X}_k^{(n)}$. First, $\mathbb{P}(S_n \neq \bar{S}_n) \leq \mathbb{P}(\cup_{k=1}^n\{\bar{X}_k^{(n)} \neq X_k\}) \leq \Sigma_{k=1}^n\mathbb{P}(|X_k| > n) = n\mathbb{P}(|X_1| > n) \to 0$. Second, $\mathbb{E}(\bar{S}_n) = \Sigma_{k=1}^n\mathbb{E}(X_k\mathbb{I}(|X_k| \leq n)) = n\mu_n \implies \mathbb{P}(|\frac{1}{n}\bar{S}_n - \mu_n| > \frac{\varepsilon}{2}) \leq \frac{4}{n^2\varepsilon^2}\text{Var}(\bar{S}_n)$ by Chebyshev's inequality $= \frac{4}{n^2\varepsilon^2}\Sigma_{k=1}^n\text{Var}(\bar{X}_k^{(n)})$ by independence $= \frac{4n}{n^2\varepsilon^2}\text{Var}(\bar{X}_1^{(n)}) = \frac{4}{n\varepsilon^2}\mathbb{E}(X_1\mathbb{I}(|X_1| \leq n))^2$. Since $\mathbb{E}(X^p) = \int_0^\infty px^{p-1}\mathbb{P}(X \geq x)dx$ for $X \leq 0$ by question 4.17, $\mathbb{E}(X_1\mathbb{I}(|X_1| \leq n))^2 = \int_0^\infty 2x\mathbb{P}(|\bar{X}_1^{(n)}| \geq x)dx = \int_0^n 2x\mathbb{P}(|X_1| \geq x)dx$ due to the indicator function. Notice that $x\mathbb{P}(|X_1| \geq x) \in [0, x]$ and $x\mathbb{P}(|X_1| \geq x) \to 0 \implies \sup_x x\mathbb{P}(|X_1| \geq x) < \infty \implies \lim_{n\to\infty}\frac{1}{n}\int_0^n x\mathbb{P}(|X_1| \geq x)dx = \lim_{n\to\infty}\int_0^1 ny\mathbb{P}(|X_1| \geq ny)dy$ by a change of variable $= \int_0^1 \lim_{n\to\infty} ny\mathbb{P}(|X_1| \geq ny)dy = 0$ by bounded convergence. Thus, $\mathbb{P}(|\frac{1}{n}\bar{S}_n - \mu_n| > \frac{\varepsilon}{2}) \to 0$. Finally, notice that $\mathbb{P}(|\frac{1}{n}S_n - \mu_n| > \varepsilon) = \mathbb{P}(|\frac{1}{n}S_n - \mu_n| > \varepsilon, S_n \neq \bar{S}_n) + \mathbb{P}(|\frac{1}{n}S_n - \mu_n| > \varepsilon, S_n = \bar{S}_n) \leq \mathbb{P}(S_n \neq \bar{S}_n) + \mathbb{P}(|\frac{1}{n}\bar{S}_n - \mu_n| > \frac{\varepsilon}{2}) \to 0$ by the previous results.

- **Proposition 6.4.1:** Let $x \in [0,1]$ and $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim}$ Bernoulli(x). Define $S_n = \Sigma_{i=1}^n X_i \sim$ Binomial(n,x). Notice that $\mathbb{E}(X_i) = x, \text{Var}(X_i) = x(1-x), \mathbb{P}(S_n = m) = \binom{n}{m}x^m(1-x)^{n-m}$, and $\mathbb{E}(f(\frac{S_n}{n})) = \Sigma_{m=0}^n f(\frac{m}{n})\mathbb{P}(S_n = m)$ by the definition of expected value $= \Sigma_{m=0}^n \binom{n}{m}x^m(1-x)^{n-m}f(\frac{n}{m}) = f_n(x)$. Next, let $\varepsilon > 0$. Since $[0,1]$ is bounded and compact, f is uniformly continuous on $[0,1]$, meaning that $\exists \delta > 0$ s.t. $|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$ for any $x, y \in [0,1]$. Then, $|f_n(x) - f(x)| = |\mathbb{E}(f(\frac{S_n}{n})) - f(x)| \leq \mathbb{E}|f(\frac{S_n}{n}) - f(x)| = \mathbb{E}(|f(\frac{S_n}{n}) - f(x)|\mathbb{I}(|\frac{S_n}{n} - x| < \delta)) + \mathbb{E}(|f(\frac{S_n}{n}) - f(x)|\mathbb{I}(|\frac{S_n}{n} - x| \geq \delta)) \leq \varepsilon + 2M\mathbb{P}(|\frac{S_n}{n} - x| \geq \delta)$ where M is some upper bound of $|f|$ on the interval $\leq \varepsilon + \frac{2M}{\delta^2}\text{Var}(\frac{S_n}{n})$ by Chebyshev's inequality $= \varepsilon + \frac{2Mx(1-x)}{n\delta^2} \leq \varepsilon + \frac{2M}{4n\delta^2}$ since the maximum of $x(1-x)$ on $[0,1]$ is $\frac{1}{4}$. As $n \to \infty$ and since ε is arbitrary, $\sup_{x\in[0,1]} |f_n(x) - f(x)| \to 0$, as needed.

Questions

- **5.6:** $|\text{Cov}(X,Y)| = |\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)| \leq \mathbb{E}|(X - \mathbb{E}X)(Y - \mathbb{E}Y)| = \sqrt{\mathbb{E}(X - \mathbb{E}X)^2\mathbb{E}(Y - \mathbb{E}Y)^2}$ by Cauchy-Schwarz $= \sqrt{\text{Var}(X)\text{Var}(Y)}$.

- **5.8:** Not testable. For the first part, if $a = b = 0, (0^p + 0^p)^2 = 0 = (0^2 + 0^2)^p$. If $a > 0$ or $b > 0, \frac{2}{p} \geq 1 \implies (\frac{a^p}{a^p+b^p})^{2/p} \geq \frac{a^p}{a^p+b^p}$ and $(\frac{b^p}{a^p+b^p})^{2/p} \geq \frac{b^p}{a^p+b^p}$ since $\frac{a^p}{a^p+b^p}, \frac{b^p}{a^p+b^p} \in (0,1]$. Then, $(\frac{a^p}{a^p+b^p})^{2/p} + (\frac{b^p}{a^p+b^p})^{2/p} \leq \frac{a^p}{a^p+b^p} + \frac{b^p}{a^p+b^p} = 1$.

- **5.9:** Not testable.

- **5.11:** Define $X_n = n$ with probability $\frac{1}{n}$ and 0 with probability $1 - \frac{1}{n}$. $\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|X_n/n| \geq \varepsilon) = \lim_{n\to\infty} \mathbb{P}(X_n/n = 1) = \lim_{n\to\infty} \frac{1}{n} = 0 \implies X_n/n \xrightarrow{p} 0$. Also, $\mathbb{P}(\lim_{n\to\infty} X_n/n^2 \leq \lim_{n\to\infty} \frac{n}{n^2} = 0) = 1 \implies X_n/n^2 \to 0$ a.s. However, $\forall \varepsilon > 0, \Sigma_{n=1}^\infty \mathbb{P}(X_n/n > \varepsilon) = \Sigma_{n=1}^\infty \mathbb{P}(X_n/n = 1) = \Sigma_{n=1}^\infty \frac{1}{n} = \infty \implies \mathbb{P}(X_n/n > \varepsilon \text{ i.o.}) = 1$ by Borel-Cantelli since the X_n's are independent $\implies \mathbb{P}(X_n/n \to 0) = 0 < 1$.

- **6.3:** By question 4.17, $\mathbb{E}|X_1| = \int_0^\infty \mathbb{P}(|X_1| \geq x)dx \geq \int_e^\infty \mathbb{P}(X_1 > x)dx = \int_e^\infty \frac{e}{x\log x}dx = \infty$. Since X can only take values in $[e,\infty)$, $\lim_{x\to\infty} x\mathbb{P}(|X_1| > x) = \lim_{x\to\infty} x\mathbb{P}(X_1 > x) = \lim_{x\to\infty} \frac{e}{\log x} = 0$, so the result follows by Theorem 6.2.3.

July 30

Notes

- **Theorem 6.4.2:** It is also true that $F_n(x) \to F(x)$ a.s. since $\mathbb{E}(\mathbb{I}(X_i \leq x)) = F(x) \implies \mathbb{I}(X_i \leq x) \sim$ Bernoulli$(F(x)) \implies \frac{1}{n}\Sigma_{i=1}^n \mathbb{I}(X_i \leq x) \to F(x)$ a.s. by SLLN.

- **Definition 7.2.2:** X can be thought of as $X(w) = a(w) + ib(w)$ for $w \in \mathcal{F}$.

- **Proposition 7.2.4:** The proof is not testable. iii) holds since $(\mathbb{E}\cos(tX))^2 + (\mathbb{E}\sin(tX))^2 \leq \mathbb{E}\cos^2(tX) + \mathbb{E}\sin^2(tX) = \mathbb{E}(\cos^2(tX) + \sin^2(tX)) = \mathbb{E}(1) = 1$ by Jensen's inequality since x^2 is convex.

Questions

- **7.1:** Not testable.

- **7.3:** For $x^+ = \max(x, 0), x^- = \max(-x, 0)$, notice that $x^+ + x^- = |x|, x^+ - x^- = x$, and $(-x)^+ = x^-$. For any continuity point y, $|F_n(y) - F(y)| = |\Sigma_{-\infty}^{y}(p_n(x) - p(x))| \leq \Sigma_{-\infty}^{y}|p_n(x) - p(x)| \leq \Sigma_{-\infty}^{\infty}|p_n(x) - p(x)| = \Sigma_{-\infty}^{\infty}(p_n(x) - p(x))^+ + \Sigma_{-\infty}^{\infty}(p_n(x) - p(x))^- = \Sigma_{-\infty}^{\infty}[p_n(x) - p(x) + (p_n(x) - p(x))^-] + \Sigma_{-\infty}^{\infty}(p_n(x) - p(x))^- = 2\Sigma_{-\infty}^{\infty}(p_n(x) - p(x))^- = \Sigma_{-\infty}^{\infty}(p(x) - p_n(x))^+$. Since $(p(x) - p_n(x))^+ \leq (p(x))^+ = p(x)$, by DCT, $\lim_{n\to\infty}(p(x) - p_n(x))^+ = 0$.

- **7.4:** Define $S_1 = [0, 1], S_2 = [0, \frac{1}{2}], S_3 = [\frac{1}{2}, 1], S_4 = [0, \frac{1}{3}], S_5 = [\frac{1}{3}, \frac{2}{3}]$, and so on. Define X_n such that f_n is uniform on $[0, 1] \setminus S_n$. Since $\mu(S_n) \to 0$ as $n \to \infty$, $X_n \xrightarrow{D} \text{Unif}(0, 1)$. However, for any $x \in [0, 1]$, there are infinitely many n such that $x \in S_n \iff f_n(x) = 0$.

- **7.5:** Not testable.

- **7.9:**

  - Bernoulli: $\mathbb{E}e^{itx} = pe^{it} + (1 - p)e^0$.
  - Poisson: $\mathbb{E}e^{itx} = \Sigma_{x=0}^{\infty} e^{itx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\Sigma_{x=0}^{\infty}\frac{(\lambda e^{it})^x}{x!} = \exp(-\lambda + \lambda e^{it})$.
  - Exponential: $\mathbb{E}e^{itx} = \int_0^\infty e^{itx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{(it-\lambda)x}dx = \frac{\lambda}{it-\lambda}e^{(it-\lambda)x}|_0^\infty = \frac{\lambda}{it-\lambda}e^{-\infty} - \frac{\lambda}{it-\lambda}e^0$ since $e^{it} = \sin(t) + i\cos(t)$ is bounded.

- **7.10:** Not testable.

August 1

Notes

- **Note:** Convergence in distribution might not hold for discontinuous points. For instance, define $\{X_n\}_n$ with $F_n(x) = (1 + e^{-nx})^{-1}$ and X with $F(x) = (1 + e^{-x})^{-1}$. Then, $\lim_{n\to\infty} F_n(0) = 1 \neq \frac{1}{2} = F(0)$.

- **Theorem 7.3.2:** The general case is when $\mathbb{E}X_n = \mu$ and can be written as $\frac{1}{\sqrt{n}}\Sigma_{i=1}^{n}(X_i - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$, or $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0, 1)$. Proof: Define $Y_n = \Sigma_{i=1}^{n}(X_i - \mu)$. Since $\text{Var}(X_i) = \mathbb{E}|X_i - \mu|^2 < \infty$, by Lemma 7.3.1, $\varphi_{X_i-\mu}(t) = 1 + it\mathbb{E}(X_i - \mu) + \frac{(it)^2}{2}\mathbb{E}(X_i - \mu)^2 + o(|t|^2) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$. Then, $\varphi_{Y_n}(t) = (\varphi_{X_i-\mu}(\frac{t}{\sqrt{n}}))^n = (1 - \frac{\sigma^2 t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{-\sigma^2 t^2/2}$ as $n \to \infty$, which is the characteristic function of $\mathcal{N}(0, \sigma^2)$. The result follows by Theorem 7.2.9.

- **Theorem 8.1.7:** Not testable.