The oral exam problems are useful short qual problems or parts of qual problems.

Problems correspond to those in Olive, D.J. (2023) *Probability and Measure*, online notes at (http://parker.ad.siu.edu/Olive/probbook.pdf).

A) Probability and Measure

1.30. Prove the following theorem.

Theorem 1.3. Properties of P: Let A, B, A_i, A_n, A_k be $\mathcal{F}$ sets.
i) Finite additivity: If $A_1, ..., A_n$ are disjoint, then $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
ii) P is monotone: $A \subseteq B \Rightarrow P(A) \leq P(B)$.
iii) If $A \subseteq B$, then $P(B - A) = P(B) - P(A)$.
iv) Complement rule: $P(A^c) = 1 - P(A)$.
v) Finite subadditivity: $P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$.

Proof. i) Let $A_i = \emptyset$ for $i \geq n+1$. Then $A_1, A_2, ...,$ are disjoint, and $P(\cup_{i=1}^n A_i) = P(\biguplus_{i=1}^n A_i) = P(\biguplus_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i) = \sum_{i=1}^n P(A_i)$ by P3).

ii) and iii) If $A \subseteq B$, then $B = A \biguplus (B - A)$ where this notation means A and $B - A$ are disjoint. Hence $P(B) = P(A) + P(B - A) \geq P(A)$ by i), and $P(B - A) = P(B) - P(A)$.

iv) Take $B = \Omega = A \biguplus A^c$. Then $P(\Omega) = 1 = P(A) + P(A^c)$.

v) Let the disjoint sets $B_1 = A_1$ and

$$B_k = A_k \cap A_1^c \cap \cdots \cap A_{k-1}^c = A_k \cap \left[\cup_{i=1}^{k-1} A_i\right]^c$$

for $k = 2, ..., n$. Then a) $B_j \subseteq A_j$ for $j = 1, ..., n$, b) $B_k \subseteq A_j^c$ for $j < k$, and c) $\cup_{i=1}^n A_i = \biguplus_{i=1}^n B_i$. Thus $P(\cup_{i=1}^n A_i) = P(\biguplus_{i=1}^n B_i) = \sum_{i=1}^n P(B_i) \leq \sum_{i=1}^n P(A_i)$.

1.31. Prove the following theorem.

Theorem 1.3. Properties of P: Let A, B, A_i, A_n, A_k be $\mathcal{F}$ sets.

vi) continuity from below: If $A_n \uparrow A$ then $P(A_n) \uparrow P(A)$.
vii) continuity from above: If $A_n \downarrow A$ then $P(A_n) \downarrow P(A)$.
viii) countable subadditivity: $P(\cup_{k=1}^\infty A_k) \leq \sum_{k=1}^\infty P(A_k)$.

Proof. vi) Let the disjoint sets $B_1 = A_1$ and $B_k = A_k - A_{k-1}$ for $k > 1$. Then $A_n = \cup_{k=1}^n A_k = \cup_{k=1}^n B_k$, and $A = \cup_{k=1}^\infty A_k = \cup_{k=1}^\infty B_k$. Hence

$$P(A) = \sum_{k=1}^\infty B_k = \lim_{n\to\infty} \sum_{k=1}^n P(B_k) = \lim_{n\to\infty} P(A_n).$$

Thus

$$P(A_n) = \sum_{k=1}^n P(B_k) \uparrow P(A).$$

vii) $A_n \downarrow A \Rightarrow A_n^c \uparrow A^c$. Hence

$$P(A_n^c) = [1 - P(A_n)] \uparrow [1 - P(A)] = P(A^c)$$

by vi). Thus $P(A_n) \downarrow P(A)$.

viii) Let $B_n = \cup_{k=1}^n A_k$. Then

$$P(B_n) = P(\cup_{k=1}^n A_k) \leq \sum_{k=1}^n P(A_k) \leq \sum_{k=1}^\infty P(A_k)$$

for any positive integer n. Now $B_n \uparrow B = \cup_{k=1}^{\infty} A_k$, and thus $P(B_n) \uparrow P(B) = P(\cup_{k=1}^{\infty} A_k)$ by vi). Hence $P(B)$ is the least upper bound on the sequence $P(B_n)$ while $\sum_{k=1}^{\infty} P(A_k)$ is an upper bound on the $P(B_n)$. Thus $P(\cup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$. $\square$

1.32. Prove the following theorem.

Theorem 1.5. For each sequence $\{A_n\}$ of $\mathcal{F}$ sets,
i) $P(liminf_n\ A_n) \leq liminf_n\ P(A_n) \leq limsup_n\ P(A_n) \leq P(limsup_n\ A_n)$
ii) Continuity of probability: If $A_n \to A$, then $P(A_n) \to P(A)$.

Proof. i) We need to show a) $P(liminf_n\ A_n) \leq liminf_n\ P(A_n)$ and b) $limsup_n\ P(A_n) \leq P(limsup_n\ A_n)$. Let $B_n = \cap_{k=n}^{\infty} A_k \uparrow liminf_n\ A_n$, and $C_n = \cup_{k=n}^{\infty} A_k \downarrow limsup_n\ A_n$. Then $P(A_n) \geq P(B_n) \to P(liminf_n\ A_n)$. (We can't take limits on both sides of the inequality since we do not know if $\lim_n P(A_n)$ exists. Note that $\lim_n P(B_n) = P(liminf_n\ A_n)$ by monotone continuity.) Taking $liminf$ of both sides gives $liminf_n\ P(A_n) \geq liminf_n P(B_n) = P(liminf_n\ A_n)$, proving a).

Similarly, $P(A_n) \leq P(C_n) \to P(limsup_n\ A_n)$. Taking $limsup$ of both sides of the inequality gives $limsup_n\ P(A_n) \leq limsup_n\ P(C_n) = P(limsup_n\ A_n)$, proving b).

ii) Follows from i) since $P(A_n) \to P(A)$ iff $\overline{lim}\ P(A_n) = \underline{\lim}\ P(A_n) = P(A)$. $\square$

1.33. State and prove the First Borel Cantelli Lemma.

Let $(\Omega, \mathcal{F}, P)$ be fixed and A_n events.
If $\sum_{n=1}^{\infty} P(A_n) < \infty$ (the sum converges), then $P(limsup_n\ A_n) = 0$.

Proof: Since $limsup_n\ A_n \subseteq \bigcup_{k=m}^{\infty} A_k$ for any positive integer m, $P(limsup_n\ A_n) \leq P(\bigcup_{k=m}^{\infty} A_k) \leq \sum_{k=m}^{\infty} P(A_k) \leq \epsilon$ for $m \geq m(\epsilon)$ by definition of a convergent sum. Since $\epsilon > 0$ is arbitrary, $P(limsup_n\ A_n) = 0$.

1.34. State and prove the Second Borel Cantelli Lemma.

If the A_n are independent events and $\sum_{n=1}^{\infty} P(A_n) = \infty$ (the sum diverges), then $P(limsup_n\ A_n) = 1$.

The proof uses the fact that if $P(A_i) = 0$ for all i, then $P(\cup_{i=1}^{\infty} A_i) = 0$.

Proof. The result holds if $0 = P[(limsup_n\ A_n)^c] = P(liminf_n\ A_n^c) =$
$P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c) = 0$ which is true if $P(\cap_{k=n}^{\infty} A_k^c) = 0$ for each positive integer n by the above fact. Since $1 - x \leq e^{-x}$,

$$P(\cap_{k=n}^{n+j} A_k^c) = \prod_{k=n}^{n+j} [1 - P(A_k)] \leq \prod_{k=n}^{n+j} \exp[-P(A_k)] = \exp\left[-\sum_{k=n}^{n+j} P(A_k)\right].$$

Since $\sum_{k=n}^{\infty} P(A_k)$ diverges for each positive integer n, the last term converges to 0 as $j \to \infty$. Thus

$$0 \leq P(\cap_{k=n}^{\infty} A_k^c) = \lim_{j\to\infty} P(\cap_{k=n}^{n+j} A_k^c) \leq \lim_{j\to\infty} \exp\left[-\sum_{k=n}^{n+j} P(A_k)\right] = 0$$

(where the first limit exists since $\cap_{k=n}^{n+j} A_k^c \downarrow \cap_{k=n}^{\infty} A_k^c$). $\square$

1.35. Suppose $A_1, A_2, ...$ are independent. There are 4 cases for the convergence and/or divergence of $\sum_n P(A_n)$ and $\sum_n P(A_n^c)$. One case is impossible.

a) Suppose that $\sum_n P(A_n) < \infty$ and $\sum_n P(A_n^c) < \infty$. If possible, find $P(limsup_n\ A_n)$, find $P(liminf_n\ A_n)$, and if $P(A_n) \to c$, find c.

b) Suppose that $\sum_n P(A_n) = \infty$ and $\sum_n P(A_n^c) = \infty$. If possible, find $P(limsup_n\ A_n)$, and find $P(liminf_n\ A_n)$. Does $\lim_n A_n = A$ exist?

c) Suppose that $\sum_n P(A_n) < \infty$ and $\sum_n P(A_n^c) = \infty$. If possible, find $P(limsup_n\ A_n)$, find $P(liminf_n\ A_n)$, and if $P(A_n) \to c$, find c. Was independence needed?

d) Suppose that $\sum_n P(A_n) = \infty$ and $\sum_n P(A_n^c) < \infty$. If possible, find $P(limsup_n\ A_n)$, find $P(liminf_n\ A_n)$, and if $P(A_n) \to c$, find c.

Solution. a) If $\sum_n P(A_n) < \infty$, then by the first Borel Cantelli lemma, $P(A_n) \to 0$ which implies that $P(A_n^c) \to 1$ which implies that $\sum_n P(A_n^c) = \infty$. Hence this case is impossible.

b) By the 2nd Borel Cantelli lemma, $P(limsup_n\ A_n) = 1$ and $1 = P(limsup_n\ A_n^c) = P[(liminf\ A_n)^c]$. Thus $P(limsup_n\ A_n) = 1$ and $P(liminf\ A_n) = 0$. Thus $\lim_n A_n$ does not exist.

c) By the 1st Borel Cantelli lemma, $P(limsup_n\ A_n) = 0$ Thus $P[liminf\ A_n] = 0$ and $c = 0$. Independence was not needed since the 1st Borel Cantelli lemma was used.

d) By the 1st Borel Cantelli lemma, $P(limsup_n\ A_n^c) = P[(liminf\ A_n)^c] = 0$. Hence $P[liminf\ A_n] = 1 \le P(limsup_n\ A_n) = 1 = c$. Thus $P(A_n) \to 1$.

(The 2nd Borel Cantelli lemma also gives $P(limsup_n\ A_n) = 1$, but is not needed.)

B) Random Variables and Random Vectors

The following theorem is used to prove the next problem.

Theorem 2.3. Fix $(\Omega, \mathcal{F}, P)$. Let $X : \Omega \to \mathbb{R}$. X is a measurable function iff X is a random variable iff any one of the following conditions holds.

i) $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall\ B \in \mathcal{B}(\mathbb{R})$.

ii) $X^{-1}((-\infty, t]) = \{X \le t\} = \{\omega \in \Omega : X(\omega) \le t\} \in \mathcal{F}\ \forall t \in \mathbb{R}$.

iii) $X^{-1}((-\infty, t)) = \{X < t\} = \{\omega \in \Omega : X(\omega) < t\} \in \mathcal{F}\ \forall t \in \mathbb{R}$.

iv) $X^{-1}([t, \infty)) = \{X \ge t\} = \{\omega \in \Omega : X(\omega) \ge t\} \in \mathcal{F}\ \forall t \in \mathbb{R}$.

v) $X^{-1}((t, \infty)) = \{X > t\} = \{\omega \in \Omega : X(\omega) > t\} \in \mathcal{F}\ \forall t \in \mathbb{R}$.

2.20 Prove the following theorem.

Theorem 2.4. Let X, Y, and X_i be RVs on $(\Omega, \mathcal{F}, P)$.

a) aX is a RV for any $a \in \mathbb{R}$.

b) $aX + bY$ is a RV for any $a, b \in \mathbb{R}$. Hence $\sum_{i=1}^n X_i$ is a RV.

c) $\max(X, Y)$ is a RV. Hence $\max(X_1, ..., X_n)$ is a RV.

d) $\min(X, Y)$ is a RV. Hence $\min(X_1, ..., X_n)$ is a RV.

e) XY is a RV. Hence $X_1 \cdots X_n$ is a RV.

f) X/Y is a RV if $Y(\omega) \ne 0\ \forall\ \omega \in \Omega$.

g) $sup_n\ X_n$ is a RV.

h) $inf_n\ X_n$ is a RV.

i) $limsup_n\ X_n$ is a RV.

j) $liminf_n\ X_n$ is a RV.

k) If $lim_n X_n = X$, then X is a RV.

l) If $\lim_m \sum_{n=1}^m X_n = \sum_{n=1}^\infty X_n = X$, then X is a RV.

Proof. a) If $a > 0$, then $\{aX \le t\} = \{X \le t/a\} \in \mathcal{F}$.

If $a < 0$, then $\{aX \le t\} = \{X \ge t/a\} \in \mathcal{F}$.
If $a = 0$, then $aX \equiv 0$ is a constant, and a constant is a random variable.
Thus aX is a random variable if X is a random variable.

b) For each t,

$$\{X + Y < t\} = \cup_{r \in \mathbb{Q}} \left[\{X < r\} \cap \{Y < t - r\} \right] \in \mathcal{F}$$

since the union is countable. Thus a sum of two random variables is a random variable, and by induction, a finite sum of random variables is a random variable.

c) For each t, $\{\max(X, Y) \le t\} = \{X \le t\} \cap \{Y \le t\} \in \mathcal{F}$
(since $\max(X, Y) \le t$ iff both $X \le t$ and $Y \le t$).

d) For each t, $\{\min(X, Y) \le t\} = \{X \le t\} \cup \{Y \le t\} \in \mathcal{F}$
(since $\min(X, Y) \le t$ iff at least one of the following holds i) $X \le t$ or ii) $Y \le t$).

e) First show that X^2 is a random variable if X is a random variable. For any $t \ge 0$, $\{X^2 \le t\} = \{-\sqrt{t} \le X \le \sqrt{t}\} = \{x \le \sqrt{t}\} - \{X < -\sqrt{t}\} \in \mathcal{F}$, while for any $t < 0$, $\{X^2 \le t\} = \emptyset \in \mathcal{F}$. Thus X^2 is a random variable. Then $XY = 0.5[(X+Y)^2 - X^2 - Y^2]$ is a random variable by b).

f) First show $1/Y$ is a random variable. Then the result follows by e). Now

$$\left\{ \frac{1}{Y} \le t \right\} = \begin{cases} \{Y \ge 1/t\} \cup \{Y \le 0\}, & t \ge 0 \\ \{Y \ge 1/t\}, & t < 0. \end{cases}$$

(Note that for $t = 0$, then $1/Y \le 0$ iff $Y \le 0$ since $Y(\omega) \ne 0\ \forall \omega$.)

g) For each t, $\{\sup_n X_n \le t\} = \cap_{n=1}^\infty \{X_n \le t\} \in \mathcal{F}$.

h) For each t, $\{\inf_n X_n \ge t\} = \cap_{n=1}^\infty \{X_n \ge t\} \in \mathcal{F}$.

i) $limsup_n\ X_n = \inf_k\ \sup_{m \ge k} X_m = \inf_k Y_k$ is a RV by h).

j) $liminf_n\ X_n = \sup_k\ \inf_{m \ge k} X_m = \sup_k W_k$ is a RV by g).

k) $X = \limsup_n\ X_n = \liminf_n\ X_n$ is a RV by i) and j).

l) By induction and b), $Y_m = \sum_{n=1}^m X_n$ is a RV. Thus $\lim_m Y_m = \lim_m \sum_{n=1}^m X_n = \sum_{n=1}^\infty X_n = X$ is a RV by j).
(Note that $\lim_m \sum_{n=1}^m X_n = \sum_{n=1}^\infty X_n = X$ means that $\lim_m \sum_{n=1}^m X_n(\omega) = \sum_{n=1}^\infty X_n(\omega) = X(\omega)\ \forall\ \omega$.) $\square$

2.21. Fix $(\Omega, \mathcal{F}, P)$. For a random variable X, prove that the induced probability $P_X(B) = P[X^{-1}(B)]$ for $B \in \mathcal{B}(\mathbb{R})$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. You may use without proof i) $X^{-1}(\mathbb{R}) = \Omega$, ii) $X^{-1}(\emptyset) = \emptyset$, iii) $X^{-1}(\bigcup_{i=1}^\infty B_i) = \bigcup_{i=1}^\infty X^{-1}(B_i)$, and iv) if A and C are disjoint, then $X^{-1}(A)$ and $X^{-1}(C)$ are disjoint.

Solution: Proof: P1) Let $B \in \mathcal{B}(\mathbb{R})$. Then $P_X(B) = P(X^{-1}(B))$ implies that $0 \le P_X(B) \le 1$.

P2: $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$
$P_X(\emptyset) = P(X^{-1}(\emptyset)) = P(\emptyset) = 1$

P3: Let $\{B_i\}_{i=1}^\infty$ be disjoint $\mathcal{B}(\mathbb{R})$ sets. Then

$$P_X(\cup_{i=1}^\infty B_i) = P[X^{-1}(\cup_{i=1}^\infty B_i)] = P[\cup_{i=1}^\infty X^{-1}(B_i)] = \sum_{i=1}^\infty P[X^{-1}(B_i)] = \sum_{i=1}^\infty P_X(B_i).$$

2.22. Let $\sigma(X) =$ the collection $\{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$. Prove that $\sigma(X)$ is a σ-field. You may use without proof i) $X^{-1}(\mathbb{R}) = \Omega$, ii) $X^{-1}(\emptyset) = \emptyset$, iii) $X^{-1}(\bigcup_{i=1}^\infty B_i) =$

$\bigcup_{i=1}^{\infty} X^{-1}(B_i)$, iv) if A and C are disjoint, then $X^{-1}(A)$ and $X^{-1}(C)$ are disjoint, v) $[X^{-1}(B)]^c = X^{-1}(B^c)$, and vi) $X^{-1}(C) \cap X^{-1}(D) = X^{-1}(C \cap D)$.

Solution: Proof: $\sigma 1)$ $X^{-1}(\mathbb{R}) = \Omega \in \sigma(X)$.
$\sigma 2)$ Let $A \in \sigma(X)$. Then $A = X^{-1}(B)$ for some $B \in \mathcal{B}(\mathbb{R})$. Thus $A^c = [X^{-1}(B)]^c = X^{-1}(B^c)$ by Theorem 2.1 v), where $B^c \in \mathcal{B}(\mathbb{R})$. Hence $A^c \in \sigma(X)$.
$\sigma 3)$ Let $\{A_i\}_{i=1}^{\infty} \in \sigma(X)$. Then $A_i = X^{-1}(B_i)$ for some $B_i \in \mathcal{B}(\mathbb{R})$. Thus $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} X^{-1}(B_i) = X^{-1}(\cup_{i=1}^{\infty} B_i)$ by Theorem 2.1 iii). Thus $\cup_{i=1}^{\infty} A_i \in \sigma(X)$. $\square$

C) Integration and Expected Value

3.30. Suppose events $A_1, ..., A_n$ are disjoint and $\biguplus_{i=1}^{n} A_i = \Omega$. Let simple random variable (SRV) $X = \sum_{i=1}^{n} x_i I_{A_i}$. Then the **expected value** of X is

$$E(X) = \sum_{i=1}^{n} x_i P(A_i) = \sum_{x} x P(X = x). \tag{1}$$

Prove the existence and uniqueness of Equation (1).

Proof. *Existence*: Suppose SRV X takes on distinct values $x_1, ..., x_m$ where m need not equal n. Then $X = \sum_{i=1}^{m} x_i I_{B_i}$ where the $B_i = \{X = x_i\} = \{\omega : X(\omega) = x_i\}$ are disjoint with $\biguplus_{i=1}^{n} B_i = \Omega$. Thus

$$E(X) = \sum_{i=1}^{m} x_i P(B_i) = \sum_{i=1}^{m} x_i P(X = x_i).$$

Uniqueness:

$$\sum_{i=1}^{n} x_i P(A_i) = \sum_{x} \sum_{i: x_i = x} x_i P(A_i) = \sum_{x} x P(\cup_{i: x_i = x} A_i) = \sum_{x} P(X = x).$$

$\square$

3.31. Prove the following theorem.

Theorem 3.11. Let X_n, X, and Y be SRVs.
a) $-\infty < E(X) < \infty$
b) linearity: $E(aX + bY) = aE(X) + bE(Y)$
c) If SRV $X = \sum_{i=1}^{n} x_i I_{A_i}$ where the A_i are not necessarily disjoint, then $E(X) = \sum_{i=1}^{n} x_i P(A_i)$.
d) monotonicity: If $X \leq Y$, then $E(X) \leq E(Y)$
f) If t is a real valued function, then $E[t(X)] = \sum_x t(x) P(X = x)$
h) If $X \perp\!\!\!\perp Y$, then $E(X) = E(X)E(Y)$.

Proof. a) $E(X) = \sum_x x P(X = x)$ where the x are bounded since X has finite range $x_1, ..., x_m$ and $P(X = x) \in [0, 1]$. Hence
$\min(x_i) \leq E(X) \leq \max(x_i)$.

b) Let $X = \sum_i x_i I_{A_i}$ and $Y = \sum_j y_j I_{B_j}$ where the A_i partition Ω and the B_j partition Ω. Then the $A_i \cap B_j$ partition Ω, and $aX + bY = ax_i + by_j$ for $\omega \in A_i \cap B_j$. Thus

$$aX + bY = \sum_{i} \sum_{j} (ax_i + by_j) I_{A_i \cap B_j}$$

is a SRV with

$$E(aX+bY)=\sum_i\sum_j(ax_i+by_j)P(A_i\cap B_j)=$$

$$\sum_i ax_i\sum_j P(A_i\cap B_j)+\sum_j by_j\sum_i P(A_i\cap B_j)=$$

$$a\sum_i x_iP(A_i)+b\sum_j y_jP(B_j)=aE(X)+bE(Y).$$

c) Since I_{A_i} is a SRV with $E(I_{A_i})=P(A_i)$ by Example 3.3, by linearity and induction,

$$E(X)=\sum_{i=1}^n x_iP(A_i).$$

d) Let $W=Y-X\geq 0$. Then $E(W)=\sum_w wP(W=w)\geq 0$ since each distinct value of $w\geq 0$. By linearity, $0\leq E(Y-X)=E(Y)-E(X)$, or $E(X)\leq E(Y)$.

f) If $X=\sum_{i=1}^n x_iI_{A_i}$ then $W=t(X)=\sum_{i=1}^n t(x_i)I_{A_i}$ shows W is a SRV. Thus $E(W)=E[t(X)]=\sum_w wP(W=w)=\sum_{i=1}^n t(x_i)P(A_i)$ by c).

h)

$$XY=\sum_i x_iI_{A_i}\ \sum_j y_jI_{B_j}=\sum_i\sum_j x_iy_jI_{A_i\cap B_j}$$

is a SRV. Thus

$$E(XY)=\sum_i\sum_j x_iy_jP(A_i\cap B_j)\stackrel{\text{ind}}{=}\sum_i\sum_j x_iy_jP(A_i)P(B_j)=$$

$$\sum_i x_iP(A_i)\ \sum_j y_jP(B_j)=E(X)E(Y).\ \square$$

3.32. Let $X\geq 0$ be a nonnegative RV. Then

$$E(X)=lim_{n\to\infty}E(X_n)=\int XdP\leq\infty \tag{2}$$

where the X_n are nonnegative SRVs with $0\leq X_n\uparrow X$. Prove the existence Equation (2).

Proof. $0\leq E(X_1)\leq E(X_2)\leq ....$ So $\{E(X_n)\}$ is a monotone sequence and $\lim_{n\to\infty}E(X_n)$ exists in $[0,\infty]$.

3.33. Prove the following theorem.

Theorem 3.13. Let X,Y be nonnegative random variables.
a) "restricted linearity:" For $X,Y\geq 0$ and $a,b\geq 0$,
$E(aX+bY)=aE(X)+bE(Y)$.
b) "monotonicity:" If $X\leq Y$ ae, then $E(X)\leq E(Y)$.

Proof. a) For SRVs $0\leq X_n\uparrow X$ and $0\leq Y_n\uparrow Y$, the RVs aX_n+bY_n are SRVs and $aX_n+bY_n\uparrow aX+bY$, which is nonnegative. Thus

$$E(aX+bY)=\lim_{n\to\infty}E(aX_n+bY_n)=\lim_{n\to\infty}(aE[X_n]+bE(Y_n])=$$

$$a\lim_{n\to\infty}E[X_n]+b\lim_{n\to\infty}E(Y_n)=aE(X)+bE(Y).$$

The first and last equalities holds by the definition of expected value for nonnegative RVs. The second inequality holds by linearity for SRVs. The third inequality holds since $\lim\ (a_n + b_n) = \lim\ a_n + \lim\ b_n$ if the RHS exists.

b) Let $W = Y - X \geq 0$. Since $E(Z) \geq 0$ when $Z \geq 0$, $E(Y - X) \geq 0$. Using a) gives

$$E(Y) = E(Y - X + X) = E(Y - X) + E(X).$$

Hence $E(Y) = \infty$ if $E(X) = \infty$. If $E(X) < \infty$, then

$$E(Y) - E(X) = E(Y - X) \geq 0$$

where $E(Y) - E(X)$ exists since $0 \leq E(X) < \infty$. $\square$

3.34. Prove the following theorem. In your proof of iii) and iv), you may use ii) **linearity**: If X and Y are integrable and $a, b \in \mathbb{R}$, then $aX + bY$ is integrable with $E(aX + bY) = aE(X) + bE(Y)$.

Theorem 3.14. i) X is integrable iff both $E[X^+]$ and $E[X^-]$ are finite.

iii) **monotonicity**: If X and Y are integrable and $X \leq Y$ ae, then $E(X) \leq E(Y)$.
iv) $|E(X)| \leq E(\ |X|\ )$.

Proof. i) If X is integrable, then $E[|X|] = E[X^+] + E[X^-]$ by Theorem 3.13 a). Since $E[X^+] \geq 0$, $E[X^-] \geq 0$, and the sum is finite, both terms are finite. If both $E[X^+]$ and $E[X^-]$ are finite, then $E[|X|] = E[X^+] + E[X^-]$ is finite.

iii) By ii) $0 \leq E(Y - X) = E(Y) - E(X)$. Thus $E(Y) \geq E(X)$.

iv) Since $-|X| \leq X \leq |X|$, iii) implies that $E(X) \leq E(|X|)$ and $-E(|X|) \leq E(X)$. Thus $-E(X) \leq E(|X|)$. Hence $|E(X)| \leq E(\ |X|\ )$. $\square$

3.35. State and prove the Monotone Convergence Theorem (for RVs). Ignore "ae" in the proof.

Theorem 3.16: Monotone Convergence Theorem (MCT): If $0 \leq X_n \uparrow X$ ae, then $E(X_n) \uparrow E(X)$.

Proof. The proof is for when the convergence is everywhere. Then $X_n \uparrow X$ implies $E(X_n) \leq E(X)$ for all n using monotonicity of nonnegative RVs. Thus $limsup_n\ E(X_n) \leq E(X)$. By Fatou's lemma:

$$E(X) = E[\lim\ X_n] = E[liminf\ X_n] \leq liminf\ E[X_n] \leq limsup\ E(X_n) \leq E(X).$$

Thus $\lim E(X_n) = E(X)$. Since $X_n \uparrow X$, $E(X_n) \leq E(X_{n+k})$ for $k \geq 0$. Thus $E(X_n) \uparrow E(X)$. $\square$

3.36. State and prove the Lebesgue Dominate Convergence Theorem (for RVs). Ignore "ae" in the proof.

LDCT: If the $|X_n| \leq Y$ ae where Y is integrable, and if $X_n \to X$ ae, then X and X_n are integrable and $E(X_n) \to E(X)$.

Proof. Since $limsup\ |X_n| = liminf|X_n| = lim\ |X_n| = |X| \leq Y$, X_n and X are integrable. Using the nonnegativity of $Y - X_n$ and $Y + X_n$,

$$E(Y) - E(X) = E(Y - X) = E[liminf\ (Y - X_n)] \leq$$

$$liminf E(Y - X_n) = E(Y) - limsup\ E(X_n)$$

where the first and third equalities follow by linearity, the second inequality holds since $lim\ (Y - X_n) = liminf\ (Y - X_n) = Y - X$, and the inequality holds by Fatou's lemma. Since $E(Y)$ is finite by integrability,

$$-E(X) \leq -limsup\ E(X_n).$$

Thus $E(X) \geq limsup\ E(X_n)$. Similarly,

$$E(Y) + E(X) = E(Y + X) = E[liminf\ (Y + X_n)] \leq$$

$$liminf E(Y + X_n) = E(Y) + liminf\ E(X_n).$$

Hence

$$E(X) \leq liminf\ E(X_n) \leq limsup\ E(X_n) \leq E(X).$$

Thus $E(X) = \lim_{n\to\infty} E(X_n)$. $\square$

D) Large Sample Theory:

4.120. a) Suppose that $X_n \sim U(-1/n, 1/n)$. Prove whether or not X_n converges in distribution to a random variable X.

b) Suppose $Y_n \sim U(0, n)$. Prove whether or not X_n converges in distribution to a random variable X.

Solution. a) The cdf $F_n(x)$ of X_n is

$$F_n(x) = \begin{cases} 0, & x \leq \frac{-1}{n} \\ \frac{nx}{2} + \frac{1}{2}, & \frac{-1}{n} \leq x \leq \frac{1}{n} \\ 1, & x \geq \frac{1}{n}. \end{cases}$$

Sketching $F_n(x)$ shows that it has a line segment rising from 0 at $x = -1/n$ to 1 at $x = 1/n$ and that $F_n(0) = 0.5$ for all $n \geq 1$. Examining the cases $x < 0$, $x = 0$ and $x > 0$ shows that as $n \to \infty$,

$$F_n(x) \to \begin{cases} 0, & x < 0 \\ \frac{1}{2} & x = 0 \\ 1, & x > 0. \end{cases}$$

Notice that if X is a random variable such that $P(X = 0) = 1$, then X has cdf

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases}$$

Since $x = 0$ is the only discontinuity point of $F_X(x)$ and since $F_n(x) \to F_X(x)$ for all continuity points of $F_X(x)$ (i.e. for $x \neq 0$),

$$X_n \xrightarrow{D} X.$$

b) $F_n(t) = t/n$ for $0 < t \leq n$ and $F_n(t) = 0$ for $t \leq 0$. Hence $\lim_{n\to\infty} F_n(t) = 0$ for $t \leq 0$. If $t > 0$ and $n > t$, then $F_n(t) = t/n \to 0$ as $n \to \infty$. Thus $\lim_{n\to\infty} F_n(t) = H(t) = 0$ for all t, and Y_n does not converge in distribution to any random variable Y since $H(t) \equiv 0$ is a continuous function but not a cdf.

4.121. State and prove Generalized Chebyshev's Inequality = Generalized Markov's Inequality.

Solution:

Generalized Chebyshev's Inequality or Generalized Markov's Inequality: Let $u : \mathbb{R} \to [0, \infty)$ be a nonnegative function. If $E[u(Y)]$ exists then for any $c > 0$,

$$P[u(Y) \geq c] \leq \frac{E[u(Y)]}{c}.$$

Proof. The proof is given for pdfs. For pmfs, replace the integrals by sums. Now

$$E[u(Y)] = \int_{\mathbb{R}} u(y)f(y)dy = \int_{\{y:u(y)\geq c\}} u(y)f(y)dy + \int_{\{y:u(y)<c\}} u(y)f(y)dy$$

$$\geq \int_{\{y:u(y)\geq c\}} u(y)f(y)dy$$

since the integrand $u(y)f(y) \geq 0$. Hence

$$E[u(Y)] \geq c \int_{\{y:u(y)\geq c\}} f(y)dy = cP[u(Y) \geq c]. \quad \square$$

4.122. State a) the SLLN and b) the WLLN. c) Prove the WLLN for the special case where $V(Y_i) = \sigma^2$.

Solution. Let Y_n be a sequence of iid random variables with $E(Y_i) = \mu$. Then

a) SLLN: $\overline{Y}_n \xrightarrow{wp1} \mu$, and

b) WLLN: $\overline{Y}_n \xrightarrow{P} \mu$.

Proof of WLLN when $V(Y_i) = \sigma^2$**:** By Chebyshev's inequality, for every $\epsilon > 0$,

$$P(|\overline{Y}_n - \mu| \geq \epsilon) \leq \frac{V(\overline{Y}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

as $n \to \infty$. $\square$

4.123. Prove the following theorem.

Theorem 4.6: If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{k} X$ where $0 < k < r$.

Solution. **Proof.** Let $U_n = |X_n - X|^r$ and $W_n = |X_n - X|^k$. then $U_n = W_n^t$ where $t = r/k > 1$. The function $g(x) = x^t$ is convex on $[0, \infty)$. By Jensen's inequality,

$$E[|X_n - X|^r] = E[U_n] = E[W_n^t] \geq (E[W_n])^t = (E[|X_n - X|^k])^{r/k}$$

for $r > k$. Thus $\lim_{n\to\infty} E[|X_n - X|^r = 0$ implies that $\lim_{n\to\infty} E[|X_n - X|^k = 0$ for $0 < k < r$. $\square$

4.124. Prove the following theorem.

Theorem 4.7. If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{P} X$.

Solution. **Proof.** For $\epsilon > 0$,

$$P[|X_n - X| \geq \epsilon] = P[|X_n - X|^r \geq \epsilon^r] \leq \frac{E[|X_n - X|^r]}{\epsilon^r} \to 0$$

as $n \to \infty$ by the Generalized Chebyshev Inequality. $\square$

4.125. State and prove the Central Limit Theorem.

Solution. **CLT:** Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Let the sample mean $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Then

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Proof. Let Z_n be the Z-score of $\overline{Y}_n$. Then the characteristic function

$$c_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + o(t^2/n)\right]^n =$$

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

for all t. Thus $Z_n \xrightarrow{D} Z \sim N(0,1)$ and $\sigma Z_n = \sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$. $\square$

4.126. State and prove the Continuous Mapping theorem.

Solution. **Continuous Mapping Theorem:** If $X_n \xrightarrow{D} X$ and the function g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Proof: If g is real and continuous, then $\cos[tg(x)]$ and $\sin[tg(x)]$ are bounded real continuous functions. Hence by the Helly-Bray-Pormanteau theorem, for each real t, the characteristic function

$$c_{g(X_n)}(t) = E[e^{itg(X_n)}] = E(\cos[tg(X_n)]) + iE(\sin[tg(X_n)]) \to$$

$$E(\cos[tg(X)]) + iE(\sin[tg(X)]) = E[e^{itg(X)}] = c_{g(X)}(t).$$

Thus $g(X_n) \xrightarrow{D} g(X)$ by the continuity theorem. $\square$

4.127. State and prove the Cramér Wold Device.

Solution: Cramér Wold Device. Let $\boldsymbol{X}_n$ be a sequence of $k \times 1$ random vectors and let $\boldsymbol{X}$ be a $k \times 1$ random vector. Then

$$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \ \text{ iff } \ \boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{n}} \xrightarrow{\mathrm{D}} \boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}$$

for all $\boldsymbol{t} \in \mathbb{R}^k$.

Proof. Let $W_n = \boldsymbol{t}^T\boldsymbol{X}_n$ and $W = \boldsymbol{t}^T\boldsymbol{X}$. Note that

$$c_{W_n}(y) = c_{\boldsymbol{t}^T\boldsymbol{X}_n}(y) = E\left[e^{iy\boldsymbol{t}^T\boldsymbol{X}_n}\right] = c_{\boldsymbol{X}_n}(y\boldsymbol{t})$$

where $y \in \mathbb{R}$, and similarly

$$c_W(y) = c_{\boldsymbol{t}^T\boldsymbol{X}}(y) = c_{\boldsymbol{X}}(y\boldsymbol{t})$$

where $y \in \mathbb{R}$.

If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, then $c_{\boldsymbol{X}_n}(\boldsymbol{t}) \to c_{\boldsymbol{X}}(\boldsymbol{t})$ $\forall$ $\boldsymbol{t} \in \mathbb{R}^k$. Fix $\boldsymbol{t}$. Then $c_{\boldsymbol{X}_n}(y\boldsymbol{t}) \to c_{\boldsymbol{X}}(y\boldsymbol{t})$ $\forall$ $y \in \mathbb{R}$. Thus $\boldsymbol{t}^T\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{t}^T\boldsymbol{X}$.

Now assume $\boldsymbol{t}^T \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{t}^T \boldsymbol{X}$ $\forall$ $\boldsymbol{t} \in \mathbb{R}^k$. Then $c_{\boldsymbol{X}_n}(y\boldsymbol{t}) \rightarrow c_{\boldsymbol{X}}(y\boldsymbol{t})$ $\forall$ $y \in \mathbb{R}$ and $\forall$ $\boldsymbol{t} \in \mathbb{R}^k$. Take $y = 1$ to get $c_{\boldsymbol{X}_n}(\boldsymbol{t}) \rightarrow c_{\boldsymbol{X}}(\boldsymbol{t})$ $\forall$ $\boldsymbol{t} \in \mathbb{R}^k$. Hence $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ by the Continuity Theorem. $\square$

4.128. State and prove the multivariate central limit theorem.

Solution. MCLT: If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid $k \times 1$ random vectors with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where the sample mean

$$\overline{\boldsymbol{X}}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i.$$

Proof of the MCLT: Note that for fixed $\boldsymbol{t}$, the $\boldsymbol{t}^T \boldsymbol{X}_i$ are iid random variables with mean $\boldsymbol{t}^T \boldsymbol{\mu}$ and variance $\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}$. Hence by the CLT, $\boldsymbol{t}^T \sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} N(0, \boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t})$. The right hand side has distribution $\boldsymbol{t}^T \boldsymbol{X}$ where $\boldsymbol{X} \sim N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. Hence by the Cramér Wold Device, $\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. $\square$

4.129. Suppose that $\boldsymbol{x}_n \perp\!\!\!\perp \boldsymbol{y}_n$ for $n = 1, 2, ...$ Suppose $\boldsymbol{x}_n \xrightarrow{D} \boldsymbol{x}$, and $\boldsymbol{y}_n \xrightarrow{D} \boldsymbol{y}$ where $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y}$. Prove that

$$\begin{bmatrix} \boldsymbol{x}_n \\ \boldsymbol{y}_n \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}.$$

Solution. Let $\boldsymbol{t} = (\boldsymbol{t}_1^T, \boldsymbol{t}_2^T)^T$, $\boldsymbol{z}_n = (\boldsymbol{x}_n^T, \boldsymbol{y}_n^T)^T$, and $\boldsymbol{z} = (\boldsymbol{x}^T, \boldsymbol{y}^T)^T$. Since $\boldsymbol{x}_n \perp\!\!\!\perp \boldsymbol{y}_n$ and $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y}$, the characteristic function

$$\phi_{\boldsymbol{z}_n}(\boldsymbol{t}) = \phi_{\boldsymbol{x}_n}(\boldsymbol{t}_1)\phi_{\boldsymbol{y}_n}(\boldsymbol{t}_2) \rightarrow \phi_{\boldsymbol{x}}(\boldsymbol{t}_1)\phi_{\boldsymbol{y}}(\boldsymbol{t}_2) = \phi_{\boldsymbol{z}}(\boldsymbol{t}).$$

Hence $\boldsymbol{z}_n \xrightarrow{D} \boldsymbol{z}$. $\square$

4.130. Prove whether the following sequences of random variables X_n converge in distribution to some random variable X. If $X_n \xrightarrow{D} X$, find the distribution of X (for example, find $F_X(t)$ or note that $P(X = c) = 1$, so X has the point mass distribution at c).

a) $X_n \sim U(-n-1, -n)$
b) $X_n \sim U(n, n+1)$
c) $X_n \sim U(a_n, b_n)$ where $a_n \rightarrow a < b$ and $b_n \rightarrow b$.
d) $X_n \sim U(a_n, b_n)$ where $a_n \rightarrow c$ and $b_n \rightarrow c$.
e) $X_n \sim U(-n, n)$
f) $X_n \sim U(c - 1/n, c + 1/n)$

Solution. If $X_n \sim U(a_n, b_n)$ with $a_n < b_n$, then

$$F_{X_n}(t) = \frac{t - a_n}{b_n - a_n}$$

for $a_n \leq t \leq b_n$, $F_{X_n}(t) = 0$ for $t \leq a_n$ and $F_{X_n}(t) = 1$ for $t \geq b_n$. On $[a_n, b_n]$, $F_{X_n}(t)$ is a line segment from $(a_n, 0)$ to $(b_n, 1)$ with slope $\dfrac{1}{b_n - a_n}$.

a) $F_{X_n}(t) \to H(t) \equiv 1 \ \ \forall t \in \mathbb{R}$ since $F_{X_n}(t) = 1$ for $t \geq -n$. Since $H(t)$ is continuous but not a cdf, X_n does not converge in distribution to any RV X.

b) $F_{X_n}(t) \to H(t) \equiv 0 \ \ \forall t \in \mathbb{R}$ since $F_{X_n}(t) = 0$ for $t < n$. Since $H(t)$ is continuous but not a cdf, X_n does not converge in distribution to any RV X.

c)

$$F_{X_n}(t) \to F_X(t) = \begin{cases} 0 & t \leq a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & t \geq b. \end{cases}$$

Hence $X_n \xrightarrow{D} X \sim U(a,b)$.

d)

$$F_{X_n}(t) \to \begin{cases} 0 & t < c \\ 1 & t > c. \end{cases}$$

Hence $X_n \xrightarrow{D} X$ where $P(X = c) = 1$. Hence X has a point mass distribution at c. (The behavior of $\lim_{n\to\infty} F_{X_n}(c)$ is not important, even if the limit does not exist.)

e)

$$F_{X_n}(t) = \frac{t+n}{2n} = \frac{1}{2} + \frac{t}{2n}$$

for $-n \leq t \leq n$. Thus $F_{X_n}(t) \to H(t) \equiv 0.5 \ \ \forall t \in \mathbb{R}$. Since $H(t)$ is continuous but not a cdf, X_n does not converge in distribution to any RV X.

f)

$$F_{X_n}(t) = \frac{t - c + \frac{1}{n}}{\frac{2}{n}} = \frac{1}{2} + \frac{n}{2}(t-c)$$

for $c - 1/n \leq t \leq c + 1/n$. Thus

$$F_{X_n}(t) \to H(t) = \begin{cases} 0 & t < c \\ 1/2 & t = c \\ 1 & t > c. \end{cases}$$

If X has the point mass at c, then

$$F_X(t) = \begin{cases} 0 & t < c \\ 1 & t \geq c. \end{cases}$$

Hence $t = c$ is the only discontinuity point of $F_X(t)$, and $H(t) = F_X(t)$ at all continuity points of $F_X(t)$. Thus $X_n \xrightarrow{D} X$ where $P(X = c) = 1$.

4.131. a) Let $P(X_n = n) = 1/n$ and $P(X_n = 0) = 1 - 1/n$.

i) Determine whether $X_n \xrightarrow{1} 0$.

ii) Determine whether $X_n \xrightarrow{P} 0$.

iii) Determine whether $X_n \xrightarrow{D} 0$.

b) Let $P(X_n = 0) = 1 - \dfrac{1}{n}$ and $P(X_n = 1) = 1/n$.

i) Determine whether $X_n \xrightarrow{2} 0$.

ii) Determine whether $X_n \xrightarrow{1} 0$.
iii) Determine whether $X_n \xrightarrow{P} 0$.
iv) Determine whether $X_n \xrightarrow{D} 0$.

Solution. a) i) X_n is discrete and takes on two values with $E(X_n) = n\frac{1}{n}$ for all positive integers n. Hence $E[|X_n - 0|] = E(X_n) = 1 \ \ \forall n$ and X_n **does not satisfy** $X_n \xrightarrow{1} 0$.

ii) Let $\epsilon > 0$. Then

$$P[|X_n - 0| \geq \epsilon] \leq P(X_n = n) = \frac{1}{n} \to 0$$

as $n \to \infty$. Hence $X_n \xrightarrow{P} 0$.

iii) By ii) $X_n \xrightarrow{D} 0$.

b) i) X_n is discrete and takes on two values with

$$E[(X_n - 0)^2] = E(X_n^2) = \sum x^2 P(X_n = x) = 0^2(1 - \frac{1}{n}) + 1^2\frac{1}{n} = \frac{1}{n} \to 0$$

as $n \to \infty$. Hence $X_n \xrightarrow{2} 0$.

Since i) holds, so do ii), iii) and iv).

(Also note that

$$E[|X_n - 0|] = E(X_n) = \frac{1}{n} \to 0 \ \forall n.$$

Hence $X_n \xrightarrow{1} 0$.)

4.132. Prove the following theorem.

Theorem 4.3. a) Suppose X_n and X are RVs with the same probability space. If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$.

b) $X_n \xrightarrow{P} \tau(\theta)$ **iff** $X_n \xrightarrow{D} \tau(\theta)$.

Solution. **Proof.** a) Assume $X_n \xrightarrow{P} X$, and let $\epsilon > 0$. Then $F_n(x) = P(X_n \leq x) =$

$$P(X_n \leq x, X > x + \epsilon) + P(X_n \leq x, X \leq x + \epsilon) \leq P(|X_n - X| \geq \epsilon) + P(X \leq x + \epsilon)$$

$$= P(|X_n - X| \geq \epsilon) + F_X(x + \epsilon)$$

where the second equality holds because the events for a partition. $P(X_n \leq x, X > x + \epsilon) \leq P(|X_n - X| \geq \epsilon)$ by the following diagram with $e = \epsilon$.

```
  X_n                      X
--------------------------
       x         x+e
```

Note that $P(X_n \leq x, X \leq x + \epsilon) \leq P(X \leq x + \epsilon)$ since $P(A \cap B) \leq P(B)$.

Similarly,

$$F_X(x - \epsilon) = P(X \leq x - \epsilon) = P(X \leq x - \epsilon, X_n > x) + P(X \leq x - \epsilon, X_n \leq x)$$

$$\leq P(|X_n - X| \geq \epsilon) + P(X_n \leq x) = P(|X_n - X| \geq \epsilon) + F_n(x)$$

where the second equality holds because the events for a partition. $P(X \leq x - \epsilon, X_n > x) \leq P(|X_n - X| \geq \epsilon)$ by the following diagram with $e = \epsilon$.

```
  X                    X_n
--------------------------
       x-e        x
```

Thus

$$F_X(x - \epsilon) - P(|X_n - X| \geq \epsilon) \leq F_n(x) \leq P(|X_n - X| \geq \epsilon) + F_X(x + \epsilon).$$

Since $X_n \xrightarrow{P} X$, it follows that $P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$. If $F_X(x)$ is continuous at x, then $F_X(x - \epsilon) \to F_X(x)$ and $F_X(x + \epsilon) \to F_X(x)$ as $\epsilon \to 0$. Taking $liminf$ and $limsup$ gives

$$F_X(x - \epsilon) \leq \liminf_n \ F_n(x) \leq \limsup_n \ F_n(x) \leq F_X(x + \epsilon).$$

Thus $F_n(x) \to F_X(x)$ as $n \to \infty$ if $F_X(x)$ is continuous at x. Thus $X_n \xrightarrow{D} X$.

b) Let $c = \tau(\theta)$. If $X_n \xrightarrow{P} c$, then $X_n \xrightarrow{D} c$ by a). Assume $X_n \xrightarrow{D} c$ and $\epsilon > 0$. Then

$$P[|X_n - c| \geq \epsilon] = P(X_n \geq c + \epsilon) + P[X_n \leq c - \epsilon] =$$

$$1 - P(X_n < c + \epsilon) + P(X_n \leq c - \epsilon) = RHS.$$

Now

$$P(X_n < c + \epsilon) \geq P\left(X_n \leq c + \frac{\epsilon}{2}\right).$$

Thus $P[|X_n - c| \geq \epsilon] = RHS \leq$

$$1 - P\left(X_n \leq c + \frac{\epsilon}{2}\right) + P(X_n \leq c - \epsilon) = 1 - F_n\left(c + \frac{\epsilon}{2}\right) + F_n(c - \epsilon) \to 0$$

as $n \to \infty$ since $F_n(t) \to F_X(t)$ as $n \to \infty$ for $t \neq c$ where

$$F_X(t) = \begin{cases} 0, & t < c \\ 1, & t \geq c. \end{cases}$$

Thus $P[|X_n - c| \geq \epsilon] \to 1 - 1 + 0 = 0$ as $n \to \infty$, and $X_n \xrightarrow{P} c$. $\square$

4.133. a) Let $X_n \sim$ Binomial(n, p) where the positive integer n is large and $0 < p < 1$.

Find the limiting distribution of $\sqrt{n} \ \left(\dfrac{X_n}{n} - \ p \right)$.

b) Let $X_1, ..., X_n$ be iid with cdf $F(x) = P(X \leq x)$. Let $Y_i = I(X_i \leq x)$ where the indicator equals 1 if $X_i \leq x$ and 0, otherwise.

i) Find $E(Y_i)$.

ii) Find VAR(Y_i).

iii) Let $\hat{F}_n(x) = \dfrac{1}{n}\sum_{i=1}^n I(X_i \le x)$ for some fixed real number x. Find the limiting distribution of $\sqrt{n}\ \left(\hat{F}_n(x)\ -\ c_x\right)$ for an appropriate constant c_x.

c) Suppose $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid $p \times 1$ random vectors from a multivariate t-distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with d degrees of freedom. Then $E(\boldsymbol{X}_i) = \boldsymbol{\mu}$ and Cov$(\boldsymbol{X})$ $= \dfrac{d}{d-2}\boldsymbol{\Sigma}$ for $d > 2$. Assuming $d > 2$, find the limiting distribution of $\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{c})$ for appropriate vector $\boldsymbol{c}$.

d) Let $Y_1, ..., Y_n$ be iid with $E(Y^r) = \exp(r\mu + r^2\sigma^2/2)$ for any real r. Find the limiting distribution of $\sqrt{n}(\overline{Y}_n - c)$ for appropriate constant c.

Solution. a) $X_n \sim \sum_{i=1}^n Y_i$ where the Y_i are iid bin$(n = 1, p)$ random variables with $E(Y_i) = p$ and $V(Y_i) = p(1-p)$. Thus

$$\sqrt{n}\ \left(\ \frac{X_n}{n} -\ p\ \right) \stackrel{D}{=} \sqrt{n}\ \left(\ \overline{Y} -\ p\ \right) \stackrel{D}{\to} N[0, p(1-p)]$$

by the CLT.

b) $Y_i = I(X_i \le x) \sim$ bin$(n = 1, F(x))$ for fixed x.

i) $E(Y_i) = P(X_i \le x) = F(x)$

ii) $V(Y_i) = F(x)(1 - F(x))$

iii) $\sqrt{n}\ (\hat{F}_n(x)\ -\ F(x)\ ) \stackrel{D}{\to} N[0, F(x)(1 - F(x)]$ by the CLT.

$$c)\ \sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \stackrel{D}{\to} N_p\left(\boldsymbol{0}, \frac{d}{d-2}\boldsymbol{\Sigma}\right)$$

by the MCLT.

d) $E(Y) = \exp(\mu + \sigma^2/2)$ using $r = 1$, and $E(Y^2) = \exp(2\mu + 2\sigma^2)$ using $r = 2$. $V(Y) = E(Y^2) - [E(Y)]^2$. Thus

$$\sqrt{n}(\overline{Y}_n - E(Y)) = \sqrt{n}(\overline{Y}_n - \exp(\mu + \sigma^2/2)) \stackrel{D}{\to} N(0, V(Y))$$

by the CLT.

4.134. For each $n \in \mathbb{N}$, let W_{nk} be independent with $E(W_{nk}) = 0$, $V(W_{nk}) = \sigma^2_{nk}$, and $s_n^2 = \sum_{k=1}^{r_n} \sigma^2_{nk}$. Suppose $|W_{nk}| \le M_n$ wp1 and $M_n/s_n \to 0$. Verify that Lyapounov's condition holds.

Hint: $|W_{nk}|^{2+\delta} \le M_n^\delta W_{nk}^2$ wp1 for $\delta > 0$. Take expectations of both sides.

Solution: Proof: Let $\delta > 0$. Then

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^{2+\delta}}{s_n^{2+\delta}} \le \lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{M_n^2\ E[|W_{nk}|^2}{s_n^{2+\delta}} = \lim_{n\to\infty} \left(\frac{M_n}{s_n}\right)^\delta \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^2}{s_n^2} =$$

$$\lim_{n\to\infty} \left(\frac{M_n}{s_n}\right)^\delta = 0$$

using $s_n^2 = \sum_{k=1}^{r_n} E[|W_{nk}|^2$. $\square$

4.135. For each $n \in \mathbb{N}$, let W_{nk} be independent with $E(W_{nk}) = 0$, $V(W_{nk}) = \sigma^2_{nk}$, and $s_n^2 = \sum_{k=1}^{r_n} \sigma^2_{nk}$. Suppose $|W_{nk}| \le M_n$ wp1 and $M_n/s_n \to 0$. Verify that Lindeberg's condition holds. Show directly: do not use the fact that if Lyapounov's condition holds, then Lindeberg's condition holds.

Solution: Proof: Let $\epsilon > 0$. Then

$$\frac{1}{s_n^2} E[W_{nk}^2 I(|W_{nk}| \geq \epsilon s_n)] \leq \frac{1}{s_n^2} E[M_n^2 I(|W_{nk}| \geq \epsilon s_n)] = \frac{M_n^2}{s_n^2} P(|W_{nk}| \geq \epsilon s_n) \leq$$

$$\frac{M_n^2}{s_n^2} \frac{E(W_{nk}^2)}{\epsilon^2 s_n^2} = \left(\frac{M_n}{s_n}\right)^2 \frac{\sigma_{nk}^2}{s_n^2 \epsilon^2}$$

where the last inequality holds by Chebyshev's inequality. So

$$\sum_{k=1}^{r_n} \frac{1}{s_n^2} E[W_{nk}^2 I(|W_{nk}| \geq \epsilon s_n)] \leq \left(\frac{M_n}{s_n}\right)^2 \frac{1}{s_n^2 \epsilon^2} \sum_{k=1}^{r_n} \sigma_{nk}^2 = \left(\frac{M_n}{s_n}\right)^2 \frac{1}{\epsilon^2} \to 0$$

using $\sum_{k=1}^{r_n} \sigma_{nk}^2 = s_n^2$. $\square$

4.136. Suppose the X_i are independent Ber$(p_i) \sim$ bin$(m = 1, p_i)$ random variables with $E(X_i) = p_i$, $V(X_i) = p_i q_i$, $q_i = 1 - p_i$, and $\sum_{i=1}^\infty p_i q_i = \infty$. Prove that

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n p_i}{(\sum_{i=1}^n p_i q_i)^{1/2}} \xrightarrow{D} N(0,1)$$

as $n \to \infty$.

Proof. Let $Y_i = |W_i| = |X_i - p_i|$. Then $P(Y_i = 1 - p_i) = p_i$ and $P(Y_i = q_i) = q_i$. Thus

$$E[|X_i - p_i|^3] = E[|W_i|^3] = \sum_y y^3 f(y) = (1-p_i)^3 p_i + p_i^3 q_i = q_i^3 p_i + p_i^3 q_i$$

$$= p_i q_i (p_i^2 + q_i^2) \leq p_i q_i$$

since $p_i^2 + q_i^2 \leq (P_i + q_i)^2 = 1$. Thus $\sum_{i=1}^n E[|X_i - p_i|^3] \leq \sum_{i=1}^n p_i q_i$. Dividing both sides by $(\sum_{i=1}^n p_i q_i)^{3/2}$ gives

$$\frac{\sum_{i=1}^n E[|X_i - p_i|^3]}{(\sum_{i=1}^n p_i q_i)^{3/2}} \leq \frac{1}{(\sum_{i=1}^n p_i q_i)^{1/2}} \to 0$$

as $n \to \infty$. Hence the special case of Lyapounov's condition

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E[|X_i - \mu_i|^3]}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0.$$

holds with $\mu_i = p_i$ and $\sigma_i^2 = p_i q_i$. Thus

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{(\sum_{i=1}^n \sigma_i^2)^{1/2}} \xrightarrow{D} N(0,1).$$

$\square$

4.137. Prove Lyapounov's CLT.

Solution: Proof. Need to show Lyapounov's condition implies Lindeberg's condition. Note that

$$\sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{\{|W_{nk}| \geq \epsilon s_n\}} W_{nk}^2 \, dP \leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{\{|W_{nk}| \geq \epsilon s_n\}} \frac{|W_{nk}|^{2+\delta}}{\epsilon^\delta s_n^\delta} \, dP = RHS$$

since $|W_{nk}|^\delta \geq \epsilon^\delta s_n^\delta$ on the integral set. So

$$\frac{|W_{nk}|^\delta}{\epsilon^\delta s_n^\delta} > 1$$

on the integral set. Thus $RHS \leq$

$$\frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E[|W_{nk}|^{2+\delta}] \to 0$$

for any $\epsilon > 0$ if Lyapounov's condition holds. Thus Lindeberg's condition holds. Note that the above inequality holds since $|W_{nk}|^{2+\delta} \geq 0$. Hence

$$\int_A |W_{nk}|^{2+\delta} \, dP \leq \int_\Omega |W_{nk}|^{2+\delta} \, dP = E[|W_{nk}|^{2+\delta}]$$

using $\Omega = A \cup A^c$ and $\int_\Omega |f| \, dP = \int_A |f| \, dP + \int_{A^c} |f| \, dP$. □

4.138. Let $r_n = n$ and $W_{nk} = W_k$. If there is a constant $c > 0$ such that $P(|W_k| < c) = 1 \ \forall k$, and if $s_n \to \infty$ as $n \to \infty$, prove that Lindeberg's CLT holds.

Solution: Proof: Once n is large enough so that $\epsilon s_n > c$ (which occurs since $s_n \to \infty$), $I[|W_k| \geq \epsilon s_n] = 0$. Hence Lindeberg's condition holds. □

4.139. Let $r_n = n$ and let the $W_{nk} = W_k$ be **iid** with $E(W_k) = 0$, and $V(W_k) = \sigma^2 \in (0, \infty)$. Prove that Lindeberg's CLT holds. (Taking $W_i = X_i - \mu$ proves the usual CLT with the Lindeberg CLT.)

Solution: Proof: Need to show that Lindeberg's condition holds. Now $s_n^2 = n\sigma^2$ and the $W_k^2 I[|W_k| \geq \epsilon s_n]$ are iid for given n. Thus

$$\frac{1}{s_n^2} \sum_{k=1}^{n} E(W_k^2 \ I[|W_k| \geq \epsilon s_n]) = \frac{1}{\sigma^2} E(W_1^2 I[|W_1| \geq \epsilon \sigma \sqrt{n}])$$

$$= \frac{1}{\sigma^2} \int_{|W_1| \geq \epsilon \sigma \sqrt{n}} W_1^2 \, dP \to 0$$

as $n \to \infty$ since $P(|W_1| \geq \epsilon \sigma \sqrt{n}) \downarrow 0$ as $n \to \infty$. Or $Y_n = W_1^2 I[|W_1| \geq \epsilon \sigma \sqrt{n}]$ satisfies $Y_n \leq W_1^2$ and $Y_n \downarrow Y = 0$ as $n \to \infty$. Thus $E(Y_n) \to E(Y) = 0$ by Lebesgue's Dominated Convergence Theorem. Thus Equation (4.15) holds and $Z_n \xrightarrow{D} N(0,1)$. If the $W_i = X_i - \mu$, then

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma \sqrt{n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0,1).$$

Thus $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$. □

E) Conditional Probability and Conditional Expectation

5.9. Suppose X is an integrable random variable on $(\Omega, \mathcal{F}, P)$ and that σ-field $\mathbb{G} \subseteq \mathcal{F}$, $G \in \mathbb{G}$, and $A \in \mathcal{F}$. Then $E(X) = \int X dP = \int_\Omega X dP$. Use the definitions of $E(X|\mathbb{G})$ and $P(A|\mathbb{G})$ to find the following integrals.

a) $\int_G E(X|\mathbb{G}) dP =$
b) $E[E(X|\mathbb{G})] = \int_\Omega E(X|\mathbb{G}) dP =$
c) $\int_G E(I_A|\mathbb{G}) dP =$
d) $\int_G P(A|\mathbb{G}) dP =$
e) $\int_\Omega P(A|\mathbb{G}) dP =$

Solution:
a) $\int_G E(X|\mathbb{G}) dP = \int_G X dP \ (= E[X I_G])$
b) $E[E(X|\mathbb{G})] = \int_\Omega E(X|\mathbb{G}) dP = \int_\Omega X dP = E[X]$
c) $\int_G E(I_A|\mathbb{G}) dP = \int_G I_A \ dP = \int I_A I_G \ dP = \int I_{A \cap G} \ dP = P(A \cap G)$
d) $\int_G P(A|\mathbb{G}) dP = P(A \cap G)$
e) $\int_\Omega P(A|\mathbb{G}) dP = P(A \cap \Omega) = P(A)$