APMA 1655: Compilation of Assignments

APMA 1655: Final Exam Review

Contents

1 Probability

1.1 Definitions

Random Event: an event with more than one possible outcome of varying likelihoods where the true outcome is a priori unknown

Sample Space Ω: the set of all possible outcomes of an experiment

Event: Each subset E of Ω

Impossible Event: the empty set $\emptyset$

1.2 Set Operations

Suppose Ω is a sample space and A and B are events $\{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$

Intersection $(A \cap B)$: A and B

Union $(A \cup B)$: A or B

Complement (A^c): not A

De Morgan's Laws:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Infinite Sets

Infinite Intersection: the collection of events that are in all the sets A_1, ..., A_n

$$\bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n,\ \forall n = 1, 2, ...\}$$

Infinite union: the collection of events in at least one of the sets ("at least one of these events happens")

$$\bigcup_{n=1}^{\infty} A_n = \{\omega \in \Omega : \exists i \ |\omega \in A_i\}$$

1.3 Probability Space

Definitions

Disjoint: $A \cap B = \emptyset$

Mutually disjoint: all pairwise intersections of A_1, ..., A_n are empty: $A_n \cup A_m = \emptyset \quad n \neq m$

Probability $\mathbb{P}$: a real-valued function $\mathbb{P} : \{\text{subsets of } \Omega\} \to \mathbb{R}$. This is often defined as

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} \qquad A \subset \Omega$$

Probability space: The pair $(\Omega, \mathbb{P})$ if $\mathbb{P}$ satisfies the following three axioms:

1. $\mathbb{P}(A) \geq 0 \qquad \forall A \subset \Omega$
2. $\mathbb{P}(\Omega) = 1$
3. For any sequence of disjoint subsets $\{A_i\}_{i=1}^{\infty}$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

1.4 Properties of Probability

1. $\mathbb{P}(\emptyset) = 0$
2. if $E_1 \cap E_2 = \emptyset$

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$$

3. if $A, B \subset \Omega$ and $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. $0 \leq \mathbb{P}(A) \leq 1$
5. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
7.

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

1.5 Conditional Probability

If $\mathbb{P}(B) > 0$,

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

Theorem: If $(\Omega, \mathbb{P})$ is a probability space, $B \subset \Omega$, $\mathbb{P}(B) > 0$, then $(\mathbb{P}(A|B), \Omega)$ is also a probability space

Multiplication Law

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$$

Partition: $B_1, ..., B_n \subset \Omega$ if they are mutually disjoint and $\bigcup_{i=1}^{n} B_i = \Omega$

The Law of Total Probability: If $B_1, ..., B_n$ provide a partition of Ω

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

Corollary of the Law of Total Probability: If $0 < \mathbb{P}(B) < 1$,

$$\mathbb{P}(A) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)$$

Bayes' Rule: Suppose $B_1, ..., B_n$ partition Ω and $\mathbb{P}(B_i), \mathbb{P}(A) > 0 \quad (i \in [1, n])$. Then

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)} \qquad i = 1, 2, ..., n$$

Independence: events that do not affect each other's outcomes:

$$\begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A) \\ \mathbb{P}(B|A) = \mathbb{P}(B) \end{cases}$$

For independent events,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Mutually independent: if $\mathbb{P}(A_m \cap A_n) = \mathbb{P}(A_m) \cdot \mathbb{P}(A_n) \quad m \neq n$

1.6 Random Variable

Definition: On a probability space $(\Omega, \mathbb{P})$, a real valued function $X : \Omega \to \mathbb{R}$ is a random variable

Continuous Random Variable: a random variable with a continuous CDF

Discrete Random Variable: a random variable with a discrete CDF

Independent Random Vairbales: Y, Z on $(\Omega, \mathbb{P})$ are independent if

$$\mathbb{P}((Y \in A) \cap (Z \in B)) = \mathbb{P}(Y \in A) \cdot \mathbb{P}(Z \in B)$$

for any subsets $A, B \in \mathbb{R}$

1.7 Cumulative Distribution Functions (CDF)

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

Bernoulli Distribution: $X \sim \text{Bernoulli}(p)$ if

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Properties of the CDF F_X

1. $F_X(x_1) \leq F_X(x_2) \qquad x_1 \leq x_2$
2. $$\begin{cases} \lim_{x \to -\infty} F_X(x) = 0 \\ \lim_{x \to \infty} F_X(x) = 1 \end{cases}$$
3. F_X is right continuous ($F_X(x_0) = \lim_{x \to x_0^+} F_X(x)$)
4. $\mathbb{P}(X = x_0) = F_X(x_0) - \lim_{x \to x_0^-} F_X(x)$ Note that this is zero if the CDF is continuous

Grand Theorem: Given a CDF, there exist a corresponding probability space and a random variable

1.8 Continuous Random Variable

For a continuous random variable and a real number x_0

$$\mathbb{P}(X = x_0) = 0$$

Theorem: if F_X is a CDF, it is piecewise differentiable

Probability density function (PDF): $p_X(x) = F'_X(x)$

For a continuous random variable X,

$$F_X(x) = \int_{-\infty}^{x} p_X(t)\ dt$$

Theorem: For X a continuous random variable with PDF p_X,

$$\int_{-\infty}^{\infty} p_X(t)\ dt = 1$$

Normal Distribution: $X \sim N(\mu, \sigma^2)$ (Normal distribution with mean μ and variance σ^2) if

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\ dt$$

1.9 Discrete Random Variables

For a discrete random variable X, its CDF is

$$F_X(x) = \sum_{k=1}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$$

where $\{p_k\}_{k=1}^{K}$ is the **probability mass function** and is the probability $\mathbb{P}(X = x_k)$

Example: $X \sim \text{Bernoulli}(p)$

$$F_X(x) = (1-p) \cdot \mathbf{1}_{[0,\infty)}(x) + p \cdot \mathbf{1}_{[1,\infty)}$$

Poisson Distribution: $X \sim \text{Pois}(\lambda)$ if

$$F_X(x) = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \cdot \mathbf{1}_{[k,\infty)}(x)$$

1.10 Expected Value (Mean)

Discrete Version: IF X is a discrete RV and $\sum_{k=0}^{K} |x_t| \cdot p_k < \infty$ (i.e. if the sum if absolutely convergent), then

$$\mathbb{E}(X) = \sum_{k=0}^{K} x_k \cdot p_k$$

Continuous version: IF X is a continuous RV then if $\int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx < \infty$ then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx$$

For a permutation σ and an absolutely convergent series,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} x_k \cdot p_k = \sum_{k=0}^{\infty} x_{\sigma(k)} \cdot p_{\sigma(k)}$$

In other words, the order of summation does not matter.

1.11 Transformations of RV

For a real-valued function $g(x)$, $g(X)$ is also a random variable.

Assuming the expected value exists,

- If X is discrete,
$$\mathbb{E}[g(X)] = \sum_{k=0}^{K} g(x_k) \cdot p_k$$
- If X is continuous,
$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot p_X(x)\, dx$$

Properties of Expected Values

- For a constant c, $\mathbb{E}(c) = c$
- For constants a, b,
$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}(X) + b$$

- For $g_1(x)$, ..., $g_J(x)$ as functions where $\mathbb{E}[g_k(x)]$ exists for all $k = 1, ..., J$,

$$\mathbb{E}[g_1(X) + \ldots + g_J(x)] = \mathbb{E}\left[\sum_{k=1}^{J} g_k(x)\right] = \sum_{k=1}^{J} \mathbb{E}[g_k(X)]$$

(the expected value is linear)

1.12 Variance

For a RV X that follows some distribution and generates numbers X_1, ..., X_n

Sample average:

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \approx \mathbb{E}X$$

Variance: For a RV X whose expected value exists,

$$\mathrm{Var}\, X = \mathbb{E}[(x - \mathbb{E}X)^2]$$

this can also be written

$$\mathrm{Var}\, X = \sum_{k=0}^{K} (x_k - \mathbb{E}X)^2 \cdot p_k \qquad \text{(X discrete)}$$

$$\mathrm{Var}\, X = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 \cdot p_X(x)\, dx \qquad \text{(X continuous)}$$

Properties of Variance

1. $\mathrm{Var}\, X = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
2. $\mathrm{Var}\,(aX + b) = a^2 \mathrm{Var}\,(X)$
3. For any constant c, $\mathrm{Var}\, c = 0$
4. If $\mathrm{Var}\, X = 0$ then there exists a c such that $\mathbb{P}(X = c) = 1$

1.13 The Law of Large Numbers (LLN)

Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of independently and identically distributed RVs defined on $(\Omega, \mathbb{P})$, then

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} \frac{X_1(\omega) + X_2(\omega) + \ ... + X_n(\omega)}{n} = \mathbb{E}X_1\}\right) = 1$$

where $\mathbb{E}X_1 = \mathbb{E}X_2 = ...$ because the CDFs are equal (by identical distribution)

Generalized Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be iid RVs defined on $(\Omega, \mathbb{P})$. If $\mathbb{E}[g(X_1)]$ exists, then

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} \frac{g(X_1(\omega)) + g(X_2(\omega)) + \ ... + g(X_n(\omega))}{n} = \mathbb{E}[g(X_1)]\}\right) = 1$$

1.14 Monte Carlo Integration

If we seek to solve a very hard integral, e.g.,

$$I = \int_0^1 \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}x)}{1 + 2\cos(\frac{\pi}{2}x)}\right)\, dx$$

We let $U \sim \text{Unif}(0,1)$ whose PDF is $\mathbf{1}_{[0,1)}(x)$. Denote the integrand $g(x)$. Then

$$I = \mathbb{E}[g(U)]$$

We can generate $X_1(\omega)$, $X_2(\omega), ... \overset{iid}{\sim} \text{Unif}(0,1)$ and with enough random values

$$\overline{g(X_n)} = \frac{g(X_1) + g(X_2) + \ ... + g(X_n)}{n} \approx \mathbb{E}[g(X_1)] = \int_0^1 \cos^{-1}\left(\frac{\cos(\frac{\pi}{2}x)}{1 + 2\cos(\frac{\pi}{2}x)}\right)\, dx$$

For integrals with bounds (a, b) rather than $(0, 1)$ we can use the same method but define a new random variable from $U \sim \text{Unif}(0,1)$ where

$$X = a + (b - a) \cdot U \sim \text{Unif}(a, b)$$

1.15 Law of the Iterated Logarithm

Error: $e_n(\omega) = \overline{g(X_n)} - \mathbb{E}[g(X_1)]$

Theorem: Let X_1, X_2, ... be iid RVs on $(\Omega, \mathbb{P})$ with EX_1 and $\operatorname{Var} X_1$ existing. Then (heuristically)

$$\mathbb{P}(\{\omega \in \Omega : |e_n(\omega)| \leq \sqrt{\operatorname{Var} X_i \cdot \frac{2\log(\log n)}{n}}\}) \approx 1$$
$$\mathbb{P}(\{\omega \in \Omega : |e_n(\omega)| > \sqrt{\operatorname{Var} X_i \cdot \frac{2\log(\log n)}{n}}\}) \approx 0$$

1.16 Central Limit Theorem

Theorem: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid RVs on $(\Omega, \mathbb{P})$. Suppose $\mathbb{E}X_i$ and $\operatorname{Var} X_i$ exist. Define a sequence of random variables $\{G_n\}_{n=1}^{\infty}$ so that

$$G_n(\omega) = \sqrt{n} \cdot e_n(\omega) = \sqrt{n} \cdot (\overline{X}_n(\omega) - \mathbb{E}X_1)$$

Then the CDf of G_n converges to the CDF of $N(0, \operatorname{Var} X_1)$ as $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{P}(G_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi \cdot \operatorname{Var} X_1}} \cdot \exp\left(-\frac{t^2}{2\operatorname{Var} X_1}\right)\, dt$$

Corollary: Under the same conditions,

$$\frac{G_n(\omega)}{\sqrt{\operatorname{Var} X_1}} = \sqrt{n} \cdot \frac{\overline{X}_n(\omega) - \mathbb{E}X_1}{\sqrt{\operatorname{Var} X - 1}} \dot{\sim} N(0,1)$$

Proof of the CLT

Weak Convergence $\mathbf{G_n \xrightarrow{w} G}$: A sequence $G_1, ..., G_n$ of RV converge weakly to a continuous RV G if

$$\lim_{n\to\infty} G_{G_n}(x) = F_G(x)$$

Strong Convergence: if

$$\lim_{n\to\infty} G_n(\omega) = G(\omega) \quad \forall \omega \in \Omega$$

Moment Generating Functions: For a RV X,

$$M_X(t) = \mathbb{E}[e^{tX}]$$

is the moment generating function.

k-th Moment of X:

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}X^k$$

For a sequence of RVs G_1, ..., G_n with continuous RV G,

$$\lim_{n\to\infty} M_{G_n}(t) = M_G(t) \Longrightarrow G_n \xrightarrow{w} G$$

Then if $G \sim N(0, \operatorname{Var} X_1)$, proof of the CLT simply depends on proving the convergence of the moment-generating functions.

Some Lemmas:

1. For $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$ with X_i RVs, if X_1, ..., X_n are independent

$$M_{S_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

2. If the same RVs X_1, ..., X_n are also identically distributed

$$M_{S_n}(t) = (M_{X_1}(t))^n$$

3. For $\{C_n\}_{n=1}^{\infty}$ being a sequence of real-numbers for which $\lim_{n\to\infty} C_n = 0$, if $\lim_{n\to\infty} n \cdot C_n = \lambda$, then

$$\lim_{n\to\infty} (1 + C_n)^n = e^{\lambda}$$

4. If $G \sim N(0, \sigma^2)$ then

$$M_G(t) = \exp(\frac{t^2\sigma^2}{2})$$

Finally the proof:

$$\begin{aligned}
G_n &= \sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mathbb{E}X_1\right] \\
&= \sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{1}{n}\cdot n \cdot \mathbb{E}X_1\right] \\
&= \frac{\sqrt{n}}{n}\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbb{E}X_1\right] \\
&= \frac{\sqrt{n}}{n}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}X_1)\right]
\end{aligned}$$

Then

$$\begin{aligned}
M_{G_n}(t) &= \mathbb{E}[e^{tG_n}] = \mathbb{E}[\exp(\frac{t}{\sqrt{n}})\sum_{i=1}^{n}(X_i - \mathbb{E}X_1)] \\
&= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right] \\
&= \mathbb{E}\left[prod_{i=1}^{n}\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right] \quad \text{(by iid)} \\
&= \left(\mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)\right)\right]\right)^n \\
&= \left(\mathbb{E}\left[1 + \frac{t}{\sqrt{n}}(X_1 - \mathbb{E}X_1) + \frac{t^2}{2n}(X_1 - \mathbb{E}X_1)^2 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{k/2}}(X_1 - \mathbb{E}X_1)^k\right]\right)^n \\
&= \left(1 + \underbrace{\frac{t^2}{2n}\mathrm{Var}\,X_1 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{k/2}}(X_1 - \mathbb{E}X_1)^k}_{C_n}\right)^n \qquad \text{(because } \mathbb{E}[X_1 - \mathbb{E}X_1] = 0\text{)}
\end{aligned}$$

Using the lemmas above $\lim_{n\to\infty} C_n = 0$ and

$$n \cdot C_n = \lambda = \frac{t^2}{2}\mathrm{Var}\,X_1 + \sum_{k=3}^{\infty}\frac{t^k}{k!n^{\frac{k}{2}-1}}\mathbb{E}[(...)^k]$$

But when $k \geq 3$, $(\frac{k}{2} - 1) > 0$ so

$$\lim_{n\to\infty} n \cdot C_n = \frac{t^2}{\operatorname{Var} X_1} := \lambda$$

Then again by the lemmas

$$M_{G_n}(t) = (1 + C_n)^n \overset{n\to\infty}{\longrightarrow} e^{\lambda} = \exp(\frac{t^2}{2}\operatorname{Var} X_1)$$

But from the final lemma, the MGF of $N(0, \operatorname{Var} X_1)$ is

$$M_G(t) = \exp(\frac{t^2}{2}\operatorname{Var} X_1)$$

Thus,

$$M_{G_n}(t) \overset{n\to\infty}{M}_G(t)$$

and

$$G_n \overset{w}{\to} G \sim N(0, \operatorname{Var} X_1) \quad \blacksquare$$

1.17 Error Bounds

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid RVs on $(\Omega, \mathbb{P})$. Suppose $\mathbb{E}X_i$ and $\operatorname{Var} X_i$ exist.

From the law of the iterated logarithm, $|e_n(\omega) \leq \sqrt{2\log(\log n)} \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}$ with probability around 100%.

$$\begin{aligned}
&\mathbb{P}\left(|e_n(\omega)| \leq z \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}\right) \\
&= \mathbb{P}(-z \leq \sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq z) \\
&= \mathbb{P}\left(\sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq z\right) - \mathbb{P}\left(\sqrt{n} \cdot \frac{e_n(\omega)}{\sqrt{\operatorname{Var} X_1}} \leq -z\right) \\
&\approx \Phi(z) - \Phi(-z) \qquad (CLT) \\
&= 2\Phi(x) - 1
\end{aligned}$$

Where Φ is the CDF of $N(0, 1)$.

Now let z^* denote the positive real number such that $\Phi(z^*) = 0.975$ so

$$\mathbb{P}\left(|e_n(\omega) \leq z^* \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}\right) \approx 2\Phi(z^*) - 1 = 0.95$$

Generally, you can choose z^* such that $\Phi(z^*) = 1 - \frac{\alpha}{2}$ so $2\Phi(z^*) - 1 = 1 - \alpha$. Then z^* os the "$1 - \alpha/2$ quantile of $N(0,1)$."

All together, this gives

$$|e_n(\omega)| \leq z^* \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}} \approx 1.96 \cdot \sqrt{\frac{\operatorname{Var} X_1}{n}}$$

Conclusion: using the CLT we can establish much tighter error bounds than the LIL approach at the cost of only 5% confidence.

1.18 Random Vectors

Random Vector: a column vector $\vec{X} = (X_1, X_2 ..., X_n)^T$ defined on $(\Omega, \mathbb{P})$ if each of its components is a RV.

CDF of a Random Vector: an n-variable function

$$F_{\vec{X}}(x_1, x_2, ..., x_n) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \leq x_i\}\right)$$

Continuous Random Vector: a random vector $\vec{X}$ if $F_{\vec{X}}$ is differentiable

The PDF of a Random Vector:

$$p_{\vec{X}}(x_1, ..., x_n) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} ... \frac{\partial}{\partial x_n} F_{\vec{X}}(x_1, x_2, ..., x_n)$$

Expected value of a Random vector: If $\vec{X} = (X_1, ..., X_n)^T$ is a continuous random vector with PDF $p_{\vec{X}}$, $g(\vec{x})$ is an n-variable function, $\int_{\mathbb{R}^n} |g(x_1 ..., x_n)| \cdot p_{\vec{X}}(x_1 ..., x_n)\, dx_1...dx_n < \infty$ then

$$\mathbb{E}[g(\vec{X})] = \underbrace{\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty}}_{n \text{ integrals}} g(x_1 ..., x_n) \cdot p_{\vec{X}}(x_1, ..., x_n)\, dx_1...dx_n$$

2 Statistics

2.1 Statistical Models

Sample Data: a collection $\{x_i\}_{i=1}^n = X_i(\omega^*)$ of deterministic numbers for some fixed $\omega^* \in \Omega$

Sample size: n in the definition of data

$\mathfrak{F}$-based model: Let $\mathfrak{F} = \{F_\theta\}_{\theta\in\Theta}$ be some family of real-valued functions satisfying the CDF properties. Then the $\mathfrak{F}$-based model is the assumption that there exists some "true" $\theta^* \in \Theta$ for which

$$X_1, X_2, \ ..., X_n \overset{iid}{\sim} F_{\theta^*}(x) = \mathbb{P}(X_1 \leq x)$$

Parameter Space: $\{F_\theta\}_{\theta\in\Theta}$, the family of functions from which we are selecting

Parametric Model: a model for which Θ is a subset of a finite-dimensional space

Non-parametric Model: a model for which Θ is a subset of an infinite-dimensional space

Unspecified model: a model for which the underlying assumption is incorrect (i.e. no true θ^* exists)

Statistical inference: the process of combining probability theory and data to infer the value of θ^*

2.2 Hypothesis Testing

We assume the $\{F_\theta\}_{\theta\in\Theta}$-based model is correct. We then let $\Theta = \Theta_0 \cup \Theta_1$, giving us two hypthoses (Θ_0 and Θ_1 partition Θ). Either:

1. **The Null Hypothesis:** $H_0 : \theta^* \in \Theta_0$
2. **The Alternative Hypothesis:** $H_1 : \theta^* \in \Theta_1$

Test: For sample size n, a test is any function $T : \mathbb{R}^n \to \{0, 1\}$. If $T(\vec{x}) = 1$ we reject H_0. If it is 0, we accept H_0. As T outputs in $\{0, 1\}$,

$$T(X_1(\omega), \ ..., X_n(\omega)) = R(\omega) \sim \text{Bernoulli}(r)$$

where $r = \mathbb{P}(R = 1) = \mathbb{E}R$

Type 1 Error: the null hypothesis is true ($\theta^* \in \Theta_0$) but we reject it ($T = 1$)

Type 2 Error: the null hypothesis is false ($\theta^* \in \Theta_1$) but we fail to reject it ($T = 0$)

Criteria for a Good Test

We define a function

$$\beta_T(\theta) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{n} T(\xi_1,\, \dots, \xi_n) \cdot \prod_{i=1}^{n} p(\xi_i|\theta)\; d\xi_1 \dots d\xi_n$$

where $p(\xi_i|\theta) = F'_\theta(\xi_i)$ so $\beta_T(\theta^*) = \mathbb{E}[T(\vec{X})]$

This function can be interpreted "if F_θ is the true CDF, the probability of rejecting H_0 through T is $\beta_T(\theta)$"

Minimize Type 1 Error: make $\sup_{\theta \in \Theta}$ ("the significance of T") small

Minimize Type 2 Error: make $\beta_T(\theta)$ large for every $\theta \in \Theta_1$

2.3 Uniformly Most Powerful Test (UMP Test)

Definition: Let $\alpha \in (0, 1)$ be pre-specified. Suppose T^* is a test with significance α ($\sup_{\theta \in \Theta_0} \beta_{T^*}(\theta) = \alpha$). Then T^* is said to be a UMP test with significance α if for all T for which $\sup_{\theta \in \Theta_0} \beta_T(\theta) = \alpha$

$$\beta_T(\theta) \leq \beta_{T^*}(\theta) \quad \forall \theta \in \Theta_1$$

Neyman-Pearson Lemma: With $\Theta = \{\theta_0\,, \theta_1\}$, $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$. Let $p(\xi|\theta) = F'_\theta(\xi)$ for all $\theta \in \Theta$. For any $\alpha \in (0, 1)$, the UMP test with significance alpha is

$$T_\alpha(\xi_1,\, \dots, \xi_n) = \mathbf{1}\left(\frac{\prod_{i=1}^n p(\xi_i|\theta_1)}{\prod_{i=1}^n p(\xi_i|\theta_0)} > C_\alpha\right)$$

where C_α is the solution to $\beta_{T_{NP,\,\alpha}}(\theta_0) = \alpha$

2.4 The Maximum Likelihood Estimator

Point Estimating: the process of estimating θ^* in the $\mathfrak{F}$-based model, usually via MLE, the method of moments, or mean squared estimation.

The Likelihood Function: We assume the $\{F_\theta\}_{\theta\in\Theta}$-based model is correct. We assume that F_θ is piecewise differentiable for all θ and we have a collection of given, fixed, deterministic data $D = \{x_i\}_{i=1}^n$. Then

$$L(\theta|D) = \prod_{i=1}^n p(x_i|\theta)$$

The MLE: We select the θ which maximized the likelihood function

$$\widehat{\theta}_{MLE} = \arg\max_{\theta\in\Theta} L(\theta|D)$$

which by the consistency property is approximately equal to θ^*

Loss function: $-L(\theta|D)$ as

$$\arg\max_{\theta\in\Theta} L(\theta|D) = \arg\min_{\theta\in\Theta}\ -L(\theta|D)$$

Log-likelihood function:

$$l(\theta|D) = \log L(\theta|D) = \sum_{i=1}^n \log p(x_i|\theta)$$

Calculating the MLE

Example: $\Theta = \mathbb{R}$ and F_θ is the CDF of $N(0,1)$.

$$\begin{aligned}
p(\xi|\theta) &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{(\xi-\theta)^2}{2}) \\
l(\theta|D) &= \sum_{i=1}^n \log p(x_i|\theta) \\
&= \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi}}\exp(-\frac{(x_i-\theta)^2}{2})\right) \\
&= \sum_{i=1}^n \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i-\theta)^2\right) \\
&= -\frac{n}{2}\log(2\pi) - \sum_{i=1}^n \frac{1}{2}(x_i-\theta)^2
\end{aligned}$$

The max will occur at the critical point so

$$\begin{aligned}
\widehat{\theta} &= \frac{\partial}{\partial\theta} l(\theta|D) = 0 \\
&= \frac{\partial}{\partial\theta}\left(-\frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{1}{2}(x_i-\theta)^2\right) \qquad\qquad = \sum_{i=1}^{N}(x_i-\theta) \\
&= -n\cdot\theta + \sum_{i=1}^{n} x_i = 0 \\
&\implies \theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n
\end{aligned}$$

From the second derivative

$$\frac{\partial^2}{\partial\theta^2} l(\theta|D) = -n < 0$$

so this is the maximum

$$\widehat{\theta}_{MLE} = \overline{x}_n$$

Consistency of the MLE: Supposing that on $(\Omega, \mathbb{P})$, the $\{F_\theta\}_{\theta\in\Theta}$-based model is correct and the "regularity conditions" apply,

$$\forall\varepsilon > 0, \quad \lim_{n\to\infty}\mathbb{P}(\{\omega\in\Omega : |\widehat{\theta}(\omega) - \theta^*| < \varepsilon\}) = 1$$

where $\widehat{\theta}_n(\omega) = \widehat{\theta}_{MLE}(X_1(\omega), \ldots, X_n(\omega))$

Asymptotic normality of the MLE: under the same conditions as consistency,

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) \dot{\sim} N(0, \frac{1}{I(\theta^*)})$$

where the "Fisher Information" $I(\theta)$ is

$$I(\theta) = \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log p(\xi|\theta)^2\right)^2 \cdot p(\xi|\theta)\; d\xi$$

APMA 1655 Honors Statistical Inference I May 11, 2023

Homework 1

Name: Milan Capoor *Due: 11 pm, February 10*

Collaborators: NA

- You are strongly encouraged to work in groups, but solutions must be written independently.

Please feel free to use the following laws without proving them:
Let A, B, and C be events. Then, we have

- (Commutative Law) $A \cup B = B \cup A$,
- (Commutative Law) $A \cap B = B \cap A$,
- (Associative Law) $(A \cup B) \cup B = A \cup (B \cup C)$,
- (Associative Law) $(A \cap B) \cap C = A \cap (B \cap C)$,
- (Distributive law) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,
- (Distributive law) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.
- Let $\{A_1, A_2, \ldots, A_n, \ldots\}$ be a sequence of events, then we have

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c, \quad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Problem 1 (Set theory)

Suppose we are interested in a sample space Ω. Please review the following definitions

$$\bigcup_{n=1}^{\infty} A_n = \left\{\omega \in \Omega : \text{ there exists at least one } n' \text{ such that } \omega \in A_{n'}\right\},$$
$$\bigcap_{n=1}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for all } n = 1, 2, 3, \ldots\}$$

1. (0.5 points) We define a sequence $\{A_n\}_{n=1}^{\infty} = \{A_1, A_2, \ldots, A_n, \ldots\}$ of events as the following:

$$A_1 = \Omega,$$
$$A_n = \emptyset, \quad \text{for all } n = 2, 3, \ldots.$$

Please prove the following:

$$\Omega = \bigcup_{n=1}^{\infty} A_n. \tag{1}$$

$$\begin{aligned}
\bigcup_{n=1}^{\infty} A_n &= A_1 \cup \bigcup_{n=2}^{\infty} A_n \\
&= \Omega \cup \bigcup_{n=2}^{\infty} A_n \\
&= \Omega \cup A_2 \cup \bigcup_{n=3}^{\infty} A_n = \Omega \cup \emptyset \cup \bigcup_{n=3}^{\infty} A_n \\
&= \Omega \cup \bigcup_{n=3}^{\infty} A_n
\end{aligned}$$

Correction : induction wrong

Hence, by induction,

$$\Omega \cup \bigcup_{n=i}^{\infty} A_n = \Omega \cup \bigcup_{n=i+1}^{\infty} A_n \quad (i \geq 2)$$

So,

$$\bigcup_{n=2}^{\infty} A_n = \emptyset$$

And,

$$\bigcup_{n=1}^{\infty} A_n = \Omega \cup \emptyset = \Omega = \Omega$$

2. Let E_1 and E_2 be two events with $E_1 \cap E_2 = \emptyset$. We define a sequence $\{A_n\}_{n=1}^{\infty} = \{A_1, A_2, \ldots, A_n, \ldots\}$ of events as the following:

$$\begin{aligned} A_1 &= E_1, \\ A_2 &= E_2, \\ A_n &= \emptyset, \quad \text{for all } n = 3, 4, \ldots. \end{aligned} \tag{2}$$

Please prove the following:

(a) (0.5 points) The sequence $\{A_n\}_{n=1}^{\infty} = \{A_1, A_2, \ldots, A_n, \ldots\}$ defined in Eq. (2) is mutually disjoint.

i. By definition, a sequence is mutually disjoint if all pairwise intersections of events are $\emptyset$

ii. $A_1 \cap A_2 = \emptyset$ because $E_1 \cap E_2 = \emptyset$ (given)

iii. $\begin{cases} A_1 \cap A_{n>2} = A_1 \cap \emptyset = \emptyset \\ A_2 \cap A_{n>2} = A_2 \cap \emptyset = \emptyset \end{cases}$ by definition of $\emptyset$

iv. $A_i \cap A_j = \emptyset \cap \emptyset = \emptyset \quad \{i \neq j \quad i, j > 2\}$

v. Therefore, all possible pairwise intersections are $\emptyset$ and the sequence is mutually disjoint by definition

(b) (0.5 points) We have the following identity

$$E_1 \cup E_2 = \bigcup_{n=1}^{\infty} A_n,$$

where $A_1, A_2, \ldots$ are defined in Eq. (2).

By definition, the RHS can be taken as

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \bigcup_{n=3}^{\infty} A_n$$

But, because

$$\bigcup_{n=3}^{\infty} A_n = \emptyset \cup \bigcup_{n=4}^{\infty} A_n = \bigcup_{n=4}^{\infty} A_n$$

we have the new identity

$$\bigcup_{n=i}^{\infty} A_n = \bigcup_{n=i+1}^{\infty} A_n = \emptyset \quad (i > 2)$$

Thus,

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup A_3 = E_1 \cup E_2 \cup \emptyset = E_1 \cup E_2 = LHS$$

3. (1 points) Let $\Omega = \mathbb{R}$ = the collection of all real numbers. We define a sequence of events as follows

$$A_n = \left[0, 1 + \frac{1}{n}\right), \qquad \text{for all } n = 1, 2, 3, \ldots. \tag{3}$$

Please prove the following identity

$$[0, 1] = \bigcap_{n=1}^{\infty} A_n,$$

where $A_1, A_2, A_3, \ldots$ are defined in Eq. (3).

$$\bigcap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \bigcap_{n=3}^{\infty} A_n$$
$$= [0, 2) \cap [0, \frac{3}{2}) \cap \bigcap_{n=3}^{\infty} A_n$$

But because $A_2 \subset A_1$,

$$[0, 2) \cap [0, \frac{3}{2}) = [0, \frac{3}{2})$$

so

$$A_1 \cap A_2 \cap \bigcap_{n=3}^{\infty} A_n = A_2 \cap \bigcap_{n=3}^{\infty} A_n$$

and because it is always true that

$$1 + \frac{1}{n} > 1 + \frac{1}{n+1}$$

we know that

$$A_{i+1} \subset A_i$$

so

$$\bigcap_{n=1}^{\infty} A_n = A_i \cap A_{i+1} \cap \bigcap_{n=i+2}^{\infty} A_i = A_{i+1} \cap \bigcap_{n=i+2}^{\infty} A_i$$

But as i tends to ∞, the intersection of A_i and A_{i+1} shortens to the interval

$$\left[0, \lim_{n \to \infty} 1 + \frac{1}{n}\right) = [0, 1]$$

(Note that because the upper bound of the interval is $1 + \frac{1}{n}$ and the value of $1/n$ is strictly positive for all $n = 1, 2, 3, \ldots$, 1 itself must be on the interval thus justifying the closed bracket)
As a result, it is clear that

$$\bigcap_{n=1}^{\infty} A_n = [0, 1]$$

Problem 2 (Definition of Probability Spaces)

(1 point) Suppose n is a fixed positive integer. We define the pair $(\Omega, \mathbb{P})$ as follows

- $\Omega = \{1, 2, \cdots, n\}$.
- For any $A \subset \Omega$, we define $\mathbb{P}(A) = \frac{\#A}{n}$, where $\#A$ denotes the number of elements in A.

Please prove that the pair $(\Omega, \mathbb{P})$ defined herein is a probability space.

By definition, for $(\Omega, \mathbb{P})$ to be a probability space, $\mathbb{P}$ must satisfy three axioms:

1. $\mathbb{P}(A) \geq 0 \quad A \subset \Omega$

   Because A is a subset of Ω, $0 \leq \#A \leq n$. Hence,

$$\frac{0}{n} \leq \mathbb{P}(A) \leq \frac{n}{n} \implies 0 \leq \mathbb{P}(A) \leq 1$$

2. $\mathbb{P}(\Omega) = 1$

   Similarly,

$$\mathbb{P}(A) := \frac{\#A}{n} \implies \mathbb{P}(\Omega) = \frac{\#\Omega}{n} = \frac{n}{n} = 1$$

3. For any sequence of disjoint subsets, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

   Define a mutually disjoint sequence of m events A in Ω such that

$$A = \bigcup_{i=1}^{m} A_i$$

   Then clearly, $\#A = m$ and

$$\mathbb{P}(\bigcup_{i=1}^{m} A_i) = \mathbb{P}(A) = \frac{\#A}{n} = \frac{m}{n}$$

   But as m is necessarily a positive integer,

$$\frac{m}{n} = \sum_{i=1}^{m} \frac{1}{n}$$

   And $1/n$ is simply the probability of a single event in A so

$$\frac{m}{n} = \sum_{i=1}^{m} \mathbb{P}(A_i)$$

   Clearly then,

$$\mathbb{P}(\bigcup_{i=1}^{m} A_i) = \sum_{i=1}^{m} \mathbb{P}(A_i)$$

   But as this is true so long as the sequence is mutually disjoint and in Ω, the equality can be rendered

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Problem 3 (Properties of $\mathbb{P}$)

Let $(\Omega, \mathbb{P})$ be a probability space. Then, we have the following properties

1. (0 point) $\mathbb{P}(\emptyset) = 0$, i.e., the probability of the impossible event is zero;

2. (0 point) if two events E_1 and E_2 satisfy $E_1 \cap E_2 = \emptyset$, we have $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$;

3. (0.5 points) suppose $A, B \subset \Omega$. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$;[1]

   Because $A \subset B$, we know that
   $$B = (B \cap A^c) \cup A$$
   and that $(B \cap A^c)$ and A are disjoint. But then from property two above,
   $$\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(A)$$
   But because of the axiom that $\mathbb{P}(A) \geq 0$ for any event A in Ω (by axiom) we know that $\mathbb{P}(B \cap A^c) \geq 0$. But if that value is positive, then
   $$\mathbb{P}(A) \leq \mathbb{P}(B)$$

4. (0.5 points) $0 \leq \mathbb{P}\{A\} \leq 1$ for any subsets $A \subset \Omega$;

   By axiom,
   $$\mathbb{P}(A) \geq 0$$
   But, as $A \subset \Omega$,
   $$\mathbb{P}(A) \leq \mathbb{P}(\Omega)$$
   from property 3 above. And by axiom, $\mathbb{P}(\Omega) = 1$ so clearly
   $$0 \leq \mathbb{P}(A) \leq 1 \quad (A \subset \Omega)$$

5. (0.5 points) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

   By definition,
   $$A^c \cup A = \Omega$$
   and A^c and A are disjoint. Thus, by property 2,
   $$\mathbb{P}(A^c) + \mathbb{P}(A) = \mathbb{P}(\Omega)$$
   But axiomatically, $\mathbb{P}(\Omega) = 1$. So,
   $$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

[1] Hint: If $A \subset B$, we have $B = (B \cap A^c) \cup A$; furthermore, $(B \cap A^c)$ and A are disjoint.

6. (1 point) for any $A, B \subset \Omega$, we have $\mathbb{P}\{A \cup B\} = \mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \cap B\}$;

   Because $A, B \subset \Omega$, we know that $(B \cap A^c)$ and A are disjoint so from axiom three and property two acting on a sequence such that

$$A_1 = A, \quad A_2 = B \cap A^c, \quad A_n = \emptyset \ (n \geq 3)$$

   we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$$

   But additionally note that

$$\begin{aligned}(B \cap A^c) \cap (B \cap A) &= (B \cap (B \cap A)) \cap (A^c \cap (B \cap A)) \\ &= ((B \cap B) \cap (B \cap A)) \cap ((A^c \cap B) \cap (A^c \cap A)) \\ &= (B \cap (B \cap A)) \cap ((A^c \cap B) \cap \emptyset) \\ &= (B \cap (B \cap A)) \cap \emptyset \\ &= \emptyset\end{aligned}$$

   Proving that $(B \cap A^c)$ and $(A \cap B)$ are disjoint. Thus from property two:

$$\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(A \cap B)$$

   Rearranging,

$$\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

   So substituting this back into the equation above we finally have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

7. (1 point) for any sequence of subsets $\{A_n\}_{n=1}^{\infty}$, we have $\mathbb{P}\{\bigcup_{n=1}^{\infty} A_n\} \leq \sum_{n=1}^{\infty} \mathbb{P}\{A_n\}$. [2]

   For the case where $\{A_n\}_{n=1}^{\infty}$ is mutually disjoint, the equality follows trivially from axiom three. In the case where the sequence is not mutually disjoint, however, the inequality arises from property 6 above:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

   implies that for any two events $A, B \subset \Omega$,

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

   (because $\mathbb{P}(A \cap B) \geq 0$ by axiom)

[2] More precisely, we have the following:

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k-1} \left\{ \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mathbb{P}\left(\bigcap_{l=1}^{k} A_{i_l}\right) \right\};$$

However, the identity above is not usually used in applications. You do not need to prove this precise identity in HW 1.

But by extending that same property to three events:

$$\begin{aligned}\mathbb{P}((A \cup B) \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cap B) \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}((A \cap B) \cap C) \\ &= \sum_{n=1}^{3} \mathbb{P}(A_n) - (\mathbb{P}(A \cap B) + \mathbb{P}((A \cap B) \cap C)) \quad (A_1 = A, A_2 = B, A_3 = C)\end{aligned}$$

And as the quantity $(\mathbb{P}(A \cap B) + \mathbb{P}((A \cap B) \cap C)) \geq 0$ (as both terms must themselves be non-negative by axiom), we now have

$$\mathbb{P}(A \cup B \cup C) \leq \sum_{n=1}^{3} \mathbb{P}(A_n) \quad (A_1 = A, A_2 = B, A_3 = C)$$

Then inductively,

Correction: Induction wrong

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Properties 1 and 2 were proved in class, and you do not need to prove them in HW 1.
Please prove Properties 3-7 above.

Problem 4 (Application of the Probability Properties)

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

1. (1 point) Let A and B are two events. Suppose $B \subset A$. Please prove the following:

$$\mathbb{P}(A^c) \leq \mathbb{P}(B^c).$$

Assume the proposition $\mathbb{P}(A^c) \leq \mathbb{P}(B^c)$ is false such that $\mathbb{P}(B^c) < \mathbb{P}(A^c)$. Then, by property 5 above,

$$\mathbb{P}(B^c) < \mathbb{P}(A^c) = 1 - \mathbb{P}(B) < 1 - \mathbb{P}(A) \implies \mathbb{P}(A) < \mathbb{P}(B)$$

However, because $B \subset A$,

$$\mathbb{P}(B) \leq \mathbb{P}(A)$$

by property 6 above. But this is a contradiction so the proposition $\mathbb{P}(A^c) \leq \mathbb{P}(B^c)$ must be true.

2. (1 point) Let A and B are two events. If $\mathbb{P}(A) = 0.7$ and $\mathbb{P}(B) = 0.6$, what is the smallest possible value of $\mathbb{P}(A \cup B)$? What is the largest possible value of $\mathbb{P}(A \cup B)$?

From property 6 above,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.7 + 0.6 - \mathbb{P}(A \cap B) = 1.3 - \mathbb{P}(A \cap B)$$

Then from the work below,

$$0.3 \leq \mathbb{P}(A \cap B) \leq 0.6$$

Then clearly,

$$1.3 - 0.6 \leq \mathbb{P}(A \cup B) \leq 1.3 - 0.3$$

or

$$0.7 \leq \mathbb{P}(A \cup B) \leq 1$$

3. (1 point) Let A and B are two events. If $\mathbb{P}(A) = 0.7$ and $\mathbb{P}(B) = 0.6$, what is the smallest possible value of $P(A \cap B)$? What is the largest possible value of $P(A \cap B)$?

From property 6 above,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.7 + 0.6 - \mathbb{P}(A \cap B) = 1.3 - \mathbb{P}(A \cap B)$$

And from property 4 above, $0 \leq \mathbb{P}(A) \leq 1 \quad (A \subset \Omega)$ so because $\mathbb{P}(A \cup B)$ is itself an event in Ω,

$$0 \leq 1.3 - \mathbb{P}(A \cap B) \leq 1$$

This gives a lower bound for $\mathbb{P}(A \cap B)$ of 0.3. And because the maximum value of $\mathbb{P}(A \cap B)$ would occur if $B \subset A$ such that

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) = 0.6$$

we can see that

$$0.3 \leq \mathbb{P}(A \cap B) \leq 0.6$$

APMA 1655 Honors Statistical Inference I February 15, 2023

Homework 2

Name: Milan Capoor *Due: 11 pm, February 17*

Collaborators:

- You are strongly encouraged to work in groups, but solutions must be written independently.

Please feel free to use all the results in the Appendix of HW 2 without proving them.

1 Problem Set

1. Suppose $(\Omega, \mathbb{P})$ is a probability space, and B is a event with $\mathbb{P}(B) > 0$. We define a function $\tilde{\mathbb{P}}$ of subsets of Ω by the following

$$\tilde{\mathbb{P}}(A) \stackrel{\text{def}}{=} \mathbb{P}(A \,|\, B), \quad \text{for all } A \subset \Omega.$$

Please prove that $\tilde{\mathbb{P}}$ is a probability, i.e., $(\Omega, \tilde{\mathbb{P}})$ is a probability space as well.

By definition, for $(\Omega, \tilde{\mathbb{P}})$ to be a probability space, $\tilde{\mathbb{P}}$ must satisfy three axioms:

(a) $\tilde{\mathbb{P}}(A) \geq 0 \quad A \subset \Omega$

By definition of conditional probability,

$$\tilde{\mathbb{P}}(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

But because $(\Omega, \mathbb{P})$ is a probability space,

$$0 \leq \mathbb{P}(A') \leq 1 \quad A' \subset \Omega$$

so as $(A \cap B)$ is itself an event in Ω,

$$0 \leq \mathbb{P}(A \cap B) \leq 1$$

And hence the quotient $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ must be non-negative as its numerator and denominator are each themselves non-negative (and the denominator is not 0). Thus,

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0 \Longrightarrow \tilde{\mathbb{P}}(A) \geq 0$$

(b) $\tilde{\mathbb{P}}(\Omega) = 1$ By definition,

$$\tilde{\mathbb{P}}(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

(c) For any sequence of disjoint subsets, $\tilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \tilde{\mathbb{P}}(A_i)$

$$\tilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\mathbb{P}(B \cap \bigcup_{i=1}^{\infty} A_i)}{\mathbb{P}(B)}$$

But because

$$B \cap \left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} (B \cap A_n)$$

(given in Appendix 2), the above reduces to

$$\frac{\mathbb{P}(\bigcup_{i=1}^{\infty}(B \cap A_i))}{\mathbb{P}(B)}$$

And because $(\Omega, \mathbb{P})$ is itself a probability space,

$$\mathbb{P}(\bigcup_{i=1}^{\infty}(B \cap A_i)) = \sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i)$$

by this same third axiom.

Then we have

$$\tilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i)}{\mathbb{P}(B)}$$

which by the rules of fraction addition is

$$\sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}$$

which is simply the same as

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i | B) = \tilde{\mathbb{P}}(A_i)$$

so clearly

$$\tilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \tilde{\mathbb{P}}(A_i)$$

2. Let $(\Omega, \mathbb{P})$ be a probability space and n be a positive integer. $B_1, B_2, \ldots, B_n$ are events and provide a partition of Ω, i.e.,

   - $\bigcup_{i=1}^n B_i = \Omega$,
   - $B_1, B_2, \ldots, B_n$ are mutually disjoint.

   Let A be any event. **Please prove that** $A \cap B_1, A \cap B_2, A \cap B_3, \ldots, A \cap B_n$ **are mutually disjoint**, i.e.,

$$(A \cap B_i) \cap (A \cap B_j) = \emptyset, \quad \text{if } i \neq j.$$

Note that

$$\bigcap_{i=1}^{n} (A \cap B_n) = A \cap \bigcap_{i=1}^{n} B_n$$

But because the sequence $B_1, ..., B_n$ is mutually disjoint,

$$A \cap \bigcap_{i=1}^{n} B_n = A \cap \emptyset$$

So

$$\bigcap_{i=1}^{n} (A \cap B_n) = \emptyset$$

and the sequence $A \cap B_1, A \cap B_2, A \cap B_3, \ldots, A \cap B_n$ is mutually disjoint

Correction: incorrect definition of mutually disjoint

3. A box contains w white balls and b black balls. A ball is chosen at random.

   - If the chosen ball is white, we add d white balls to the box, that is, now there are $w + d$ white balls and b black balls.
   - If the chosen ball is black, we add d black balls to the box, that is, now there are w white balls and $b + d$ black balls.

   After adding the d balls, another ball is drawn at random from the box. **Show that the probability that the second chosen ball is white does not depend on d.** Hint: Use the law of total probability (LTP).

   The law of total probability says that for a series $B_1, ...B_n$ which partitions Ω and for which $\mathbb{P}(B_i) > 0$,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

   As the box contains only white and black balls, no ball can be both, and one ball must be chosen, the events "a white ball is chosen first" (call this B_1 so $\mathbb{P}(B_1) = w/(w + b)$) and "a black ball is chosen first" (B_2 so $\mathbb{P}(B_2) = b/(w + b)$) partition Ω, the space of outcomes of the first drawing. Thus, if A is the probability that the second ball is white, the law of total probability gives

$$\mathbb{P}(A) = \mathbb{P}(A|B_1) \cdot \mathbb{P}(B_1) + \mathbb{P}(A|B_2) \cdot \mathbb{P}(B_2)$$

   But as d white balls are added if the first ball is white and black balls conversely for black first,

$$\begin{cases} \mathbb{P}(A|B_1) = \frac{w+d}{w+d+b} \\ \mathbb{P}(A|B_2) = \frac{w}{w+d+b} \end{cases}$$

   thus

$$\begin{aligned} \mathbb{P}(A) &= \frac{w + d}{w + d + b} \cdot \frac{w}{w + b} + \frac{w}{w + d + b} \cdot \frac{b}{w + b} \\ &= \frac{w^2 + dw + wb}{(w + d + b)(w + b)} \\ &= \frac{w(w + d + b)}{(w + b)(w + d + b)} \\ &= \frac{w}{w + b} \end{aligned}$$

   Which does not depend on d.

4. Suppose the underlying probability space is $(\Omega, \mathbb{P})$. Let G and H be events such that $0 < \mathbb{P}(G) < 1$ and $0 < \mathbb{P}(H) < 1$. **Give a formula for $\mathbb{P}(G|H^c)$ in terms of $\mathbb{P}(G)$, $\mathbb{P}(H)$ and $\mathbb{P}(G \cap H)$ only.**

   By the definition of conditional probability,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad A, B \subset \Omega$$

   so,

$$\mathbb{P}(G|H^c) = \frac{\mathbb{P}(G \cap H^c)}{\mathbb{P}(H^c)}$$

   But note that $(G \cap H^c)$ and $(G \cap H)$ are disjoint so

$$\mathbb{P}(G) = \mathbb{P}(G \cap H^c) + \mathbb{P}(G \cap H)$$

   (see HW 1 for proof), allowing us to say

$$\frac{\mathbb{P}(G \cap H^c)}{\mathbb{P}(H^c)} = \frac{\mathbb{P}(G) - \mathbb{P}(G \cap H)}{\mathbb{P}(H^c)}$$

   then from the properties of complements, $\mathbb{P}(H^c) = 1 - \mathbb{P}(H)$ and we thus have

$$\mathbb{P}(G|H^c) = \frac{\mathbb{P}(G) - \mathbb{P}(G \cap H)}{1 - \mathbb{P}(H)}$$

5. Suppose we have the following

$$\begin{aligned}&\mathbb{P}(\text{“snow today”}) = 30\%,\\&\mathbb{P}(\text{“snow tomorrow”}) = 60\%,\\&\mathbb{P}(\text{“snow today and tomorrow”}) = 25\%.\end{aligned}$$

Given that it snows today, what is the probability that it will snow tomorrow?

Let $\mathbb{P}(A)$ be the probability it snows today. Then $\mathbb{P}(B)$ is the probability it will snow tomorrow. By the definition of conditional probability,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{0.25}{0.3} = \frac{5}{6} \approx \boxed{83.3\%}$$

6. Let $(\Omega, \mathbb{P})$ be a probability space. Suppose we have two events A and B such that $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. **Please prove that the following three equations are equivalent.**

   (a) $\mathbb{P}(A \mid B) = \mathbb{P}(A)$,

   (b) $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$,

   (c) $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

   As $\mathbb{P}(B) > 0$,

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

   but by (a), $\mathbb{P}(A|B) = \mathbb{P}(A)$ so

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \implies \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

   giving (b). But equivalently,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \overset{\text{by (c)}}{\implies} \mathbb{P}(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \implies \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A \cap B)$$

   As this also gives (b), it is clear that the three statements are equivalent.

2 Appendix

Please feel free to use all the results in the appendix without proving them.

2.1 Appendix 1

Let A, B, and C be events. Then, we have

- (Commutative Law) $A \cup B = B \cup A$,
- (Commutative Law) $A \cap B = B \cap A$,
- (Associative Law) $(A \cup B) \cup B = A \cup (B \cup C)$,
- (Associative Law) $(A \cap B) \cap C = A \cap (B \cap C)$,
- (Distributive law) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,
- (Distributive law) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$,
- $(A \cup B)^c = A^c \cap B^c$,
- $(A \cap B)^c = A^c \cup B^c$.

2.2 Appendix 2

Let $A_1, A_2, \dots$ be any sequence of events and B be an event. We have the following

$$\begin{aligned}
\left(\bigcup_{n=1}^{\infty} A_n \right)^c &= \bigcap_{n=1}^{\infty} A_n^c, \\
\left(\bigcap_{n=1}^{\infty} A_n \right)^c &= \bigcup_{n=1}^{\infty} A_n^c, \\
B \cap \left(\bigcup_{n=1}^{\infty} A_n \right) &= \bigcup_{n=1}^{\infty} (B \cap A_n), \\
B \cup \left(\bigcap_{n=1}^{\infty} A_n \right) &= \bigcap_{n=1}^{\infty} (B \cup A_n).
\end{aligned}$$

APMA 1655 Honors Statistical Inference I February 24, 2023

Homework 3

Name: Milan Capoor *Due: 11 pm, February 24*

Collaborators: Garv Gaur

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

To help you better answer the questions in HW 3, we review the example of Bernoulli distributions as follows:

- The experiment of interest is flipping a fair coin;
- the sample space corresponding to this experiment is $\Omega = \{\texttt{heads}, \texttt{tails}\}$;
- the probability $\mathbb{P}$ is defined by $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$, i.e., $\mathbb{P}(\{\texttt{heads}\}) = \mathbb{P}(\{\texttt{tails}\}) = \frac{1}{2}$;
- the random variable X is defined by

$$X(\texttt{heads}) = 1, \quad X(\texttt{tails}) = 0.$$

The CDF of X is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{1}{2}, & \text{if } 0 \le x < 1, \\ 1, & \text{if } x \ge 1. \end{cases} \tag{1}$$

Proof:

1. When $x < 0$, we have $A_x = \{\omega \in \Omega : X(\omega) \le x\} = \emptyset$; then, $F_X(x) = \mathbb{P}(A_x) = \mathbb{P}(\emptyset) = 0$.
2. When $0 \le x < 1$, we have $A_x = \{\omega \in \Omega : X(\omega) \le x\} = \{\texttt{tails}\}$; then, $F_X(x) = \mathbb{P}(A_x) = \mathbb{P}(\{\texttt{tails}\}) = \frac{1}{2}$.
3. When $x \ge 1$, we have $A_x = \{\omega \in \Omega : X(\omega) \le x\} = \Omega$; then, $F_X(x) = \mathbb{P}(A_x) = \mathbb{P}(\Omega) = 1$.

The proof is completed. $\square$

In addition, the Wikipedia page on random variables is nice material for learning the concept of random variables.

2 Problem Set

1. Let $(\Omega, \mathbb{P})$ be a probability space. Suppose B is an event and $0 < \mathbb{P}(B) < 1$. **Please prove the following:**

   (a) (1 point) If A and B are independent, then A and B^c are also independent.
   Assume A and B^c are not independent. Then

   $$\begin{aligned}\mathbb{P}(A \cap B^c) &\neq \mathbb{P}(A) \cdot \mathbb{P}(B^c)\\ &\neq \mathbb{P}(A) \cdot (1 - \mathbb{P}(B))\\ &\neq \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B)\end{aligned}$$

   But as A and B are independent,

   $$\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$$

   so

   $$\mathbb{P}(A \cap B^c) \neq \mathbb{P}(A) - \mathbb{P}(A \cap B)$$

   or

   $$\mathbb{P}(A) \neq \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B)$$

   However, as $A \cap B$ and $A \cap B^c$ are disjoint,

   $$\mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A)$$

   so we have

   $$\mathbb{P}(A) \neq \mathbb{P}(A)$$

   but this is a contradiction so A and B^c must be independent, completing the proof.

   (b) (1 point) $\mathbb{P}(A|B) + \mathbb{P}(A^c|B) = 1$.
   From the conditional probability definition,

   $$\begin{aligned}\mathbb{P}(A|B) + \mathbb{P}(A^c|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)}{\mathbb{P}(B)}\end{aligned}$$

   But $(A \cap B)$ and $(A^c \cap B)$ are disjoint so

   $$\begin{aligned}\mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) &= \mathbb{P}((A \cap B) \cup (A^c \cap B))\\ &= \mathbb{P}((A \cup A^c) \cap (A \cup B) \cap (B \cup A^c) \cap (B \cup B))\\ &= \mathbb{P}(\Omega \cap (A \cup B) \cap (B \cup A^c) \cap B)\\ &= \mathbb{P}(B)\end{aligned}$$

   Combining this with the above, we have

   $$\mathbb{P}(A|B) + \mathbb{P}(A^c|B) = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1 \quad \blacksquare$$

2. (2 points) Let n be a positive integer, and $\Omega \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$. Suppose $\mathbb{P}$ is a function of subsets of Ω defined as follows

$$\mathbb{P}(A) \stackrel{\text{def}}{=} \frac{\#A}{\#\Omega}, \quad \text{for all } A \subset \Omega.$$

You have proved in HW 1 that $(\Omega, \mathbb{P})$ is a probability space.

We define a random variable X as follows

$$X(\omega) = \omega, \quad \text{for all } \omega \in \Omega = \{1, 2, \ldots, n\}.$$

Please derive the CDF of the random variable X defined above. Please present your answer using a formula like the one in Eq. (1).

Clearly, $\mathbb{P}(X \leq 1) = 0$ and $\mathbb{P}(X > n) = 1$. Then as X is a discrete random variable defined on $1 \leq x \leq n$ with $x, n \in \mathbb{Z}^+$,

$$\mathbb{P}(X \leq x) = \sum_{i=1}^{x} \frac{1}{n} = \frac{x}{n}$$

which gives the simple formula

$$\boxed{F_X(x) = \begin{cases} 0 & x < 1 \\ x/n & 1 \leq x < n \\ 1 & x \geq n \end{cases}}$$

Correction: should be floor(x/n

3. (2 points) Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$. Suppose X satisfies the following

x	1	2	3	4	5
$\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$	1/2	1/4	1/8	1/16	1/16

Please derive the CDF of the random variable X. Please present your answer using a formula like the one in Eq. (1).

Note that for $x < 1$ and $x > 5$, $F_X(x)$ is 0 and 1 respectively. Then because X is a discrete random variable,

$$\mathbb{P}(X \leq x) = \sum_{i=1}^{x} \mathbb{P}(X = x) = \sum_{i=1}^{x} \frac{1}{2^i} \quad (1 \leq x < 5)$$

$$\boxed{F_X(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{2^n} & 1 \leq x < 5 \\ 1 & x \geq 5 \end{cases}}$$

Correction: include nonintegers

4. (2 points) Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$. Suppose X satisfies the following

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = 0\}) = 1.$$

Please derive the CDF of the random variable X. Please present your answer using a formula like the one in Eq. (1).

Note that

$$\mathbb{P}(x = 0) = F_X(0) - \lim_{x \to 0^-} F_X(x)$$

so as we are given the value of $\mathbb{P}(x = 0)$,

$$F_x(0) = 1 + \lim_{x \to 0^-} F_X(x)$$

but as F_X is itself defined as a probability, it must range on the interval $[0, 1]$ so

$$\lim_{x \to 0^-} F_X(x) = 0$$

and as the CDF is non-decreasing, this implies that

$$F_X(x) = 0 \quad x < 0$$

similarly, as the CDF is non-decreasing, $F_X(0) = 1$ and $\lim_{x \to \infty} F_X(x) = 1$, then

$$F_X(x) = 1 \quad x \geq 1$$

giving us the CDF for the entire domain

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

5. (2 points) Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$. Suppose the CDF of X is the following

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1; \\ \log x, & \text{if } 1 \leq x < e; \\ 1, & \text{if } e \leq x. \end{cases}$$

Please compute the values of the following:

(a) $\mathbb{P}(\{\omega \in \Omega : X(\omega) < 2\})$;

$$\begin{aligned} \mathbb{P}(X < 2) &= F_X(2) - \mathbb{P}(X = 2) \\ &= F_X(2) - (F_X(2) - \lim_{x \to 2^-} F_X(x)) \\ &= \lim_{x \to 2^-} F_X(x) \\ &= \boxed{\log 2 \approx 0.693} \end{aligned}$$

(b) $\mathbb{P}(\{\omega \in \Omega : 0 < X(\omega) \leq 3\})$;

$$\mathbb{P}(X \leq 3) = F_X(3) = \boxed{1}$$

(c) $\mathbb{P}(\{\omega \in \Omega : 2 < X(\omega) < 2.5\})$.

$$\begin{aligned} \mathbb{P}(2 < X < 2.5) &= \mathbb{P}(X \leq 2.5) - \mathbb{P}(X = 2.5) - \mathbb{P}(X < 2) \\ &= F_X(2.5) - \mathbb{P}(X = 2.5) - \log 2 \\ &= \log(2.5) - \log(2) - (F_X(2.5) - \lim_{x \to 2.5^-} F_X(x)) \\ &= \log(2.5) - \log(2) - \log(2.5) + \log(2.5) \\ &= \boxed{\log 1.25 \approx 0.223} \end{aligned}$$

Remark: For simplicity, many textbooks suppress the ω and represent $\mathbb{P}(\{\omega \in \Omega : X(\omega) < 2\})$, $\mathbb{P}(\{\omega \in \Omega : 0 < X(\omega) \leq 3\})$, and $\mathbb{P}(\{\omega \in \Omega : 2 < X(\omega) < 2.5\})$ as $\mathbb{P}(X < 2)$, $\mathbb{P}(0 < X < 3)$, and $\mathbb{P}(2 < X < 2.5)$, respectively. When you read those textbooks, this remark helps you understand what they mean.

APMA 1655 Honors Statistical Inference I

Homework 4

Name: Milan Capoor *Due: 11 pm, March 10*

Collaborators: Garv Gaur

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

1.1 Definition of Discrete Random Variables

Let $(\Omega, \mathbb{P})$ be a probability space. Suppose X is a random variable defined on Ω, and F_X is the CDF of X.

1. We say X is a **discrete random variable** if its CDF F_X is of the following form

$$\boxed{F_X(x) = \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_i, +\infty)}(x),} \tag{1}$$

where $p_k \geq 0$ for all $k = 1, 2, \ldots, K$ and $\sum_{k=0}^{K} p_k = 1$; the K is allowed to be ∞.

2. If X is a discrete random variable whose CDF is of the form in Eq. (1), we call the ordered sequence $\{p_k\}_{k=0}^{K}$ as the **probability mass function (PMF)**$^{\dagger 1}$ of X.

1.2 Independence between Events

Let $(\Omega, \mathbb{P})$ be a probability space. Suppose A and B are two events. We say A and B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

1.3 Independence between Random Variables — Version I

Let Y and Z be two random variables defined on the probability space $(\Omega, \mathbb{P})$. We say that Y and Z are independent if they satisfy the following **for any** subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$

$$\boxed{\mathbb{P}\left(\{\omega \in \Omega \,:\, Y(\omega) \in A \text{ and } Z(\omega) \in B\}\right) = \mathbb{P}\left(\{\omega \in \Omega \,:\, Y(\omega) \in A\}\right) \cdot \mathbb{P}\left(\{\omega \in \Omega \,:\, Z(\omega) \in B\}\right).}$$

[1]†: The ordered sequence $\{p_k\}_{k=0}^{K}$ is conventionally called as a function. You may view the map $k \mapsto p_k$ as a function. I think the reason $\{p_k\}_{k=0}^{K}$ is called a function is to make the names "PMF" and "PDF" look similar. In addition, if you are comfortable with the concept of vectors, you may view the ordered sequence $\{p_k\}_{k=0}^{K}$ as a vector $(p_0, p_1, \ldots, p_K)$; if $K = \infty$, this vector is infinitely long.

1.4 Independence between Random Variables — Version II

Let Y and Z be two random variables defined on the probability space $(\Omega, \mathbb{P})$. We say that Y and Z are independent if the following is true: **for any** subsets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, the following two events are independent

$$\tilde{A} = \{\omega \in \Omega : Y(\omega) \in A\}, \qquad \tilde{B} = \{\omega \in \Omega : Z(\omega) \in B\}.$$

2 Problem Set

1. (2 points) Let n be a positive integer, and $\Omega \stackrel{\text{def}}{=} \{1, 2, \dots, n\}$. Suppose $\mathbb{P}$ is a function of subsets of Ω defined as follows

$$\mathbb{P}(A) \stackrel{\text{def}}{=} \frac{\#A}{\#\Omega}, \quad \text{for all } A \subset \Omega.$$

We define a random variable X as follows

$$X(\omega) = \omega, \qquad \text{for all } \omega \in \Omega = \{1, 2, \dots, n\}.$$

Suppose you have done the following

- You have proved that $(\Omega, \mathbb{P})$ is a probability space (see HW 1).
- You have derived the CDF F_X of X (see HW 3).

Please represent the CDF F_X in the form in Eq. (1). Specifically, please show what the K, $\{p_k\}_{k=0}^K$, and $\{x_k\}_{k=0}^K$ in Eq. (1) should be.

From HW 3,

$$F_X(x) = \begin{cases} 0 & x < 1 \\ x/n & 1 \le x < n \\ 1 & x \ge n \end{cases}$$

which in summation notation is

$$F_X(x) = \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_i, +\infty)}(x)$$

where

$$\begin{cases} \{p_k\}_{k=0}^K = \{\frac{1}{K}\} \\ \{x_k\}_{k=0}^K = \{k\} \\ K = n \end{cases}$$

or

$$\boxed{F_X(x) = \frac{1}{n} \sum_{k=0}^{n} \mathbf{1}_{[k, +\infty)}}$$

Correction: incorrect limit

2. (2 points) Let Y and Z be random variables defined on the probability space $(\Omega, \mathbb{P})$; the distribution of the random variable X defined as follows

$$\begin{aligned}
&X(\omega) \stackrel{\text{def}}{=} Y(\omega) + (1 - Y(\omega)) \cdot Z(\omega), \quad \text{for all } \omega \in \Omega, \quad \text{where} \\
&Y \sim \text{Bernoulli}\left(\frac{1}{2}\right), \quad Z \sim N(0,1), \\
&Y \text{ and } Z \text{ are independent.}
\end{aligned} \tag{2}$$

Then, we claim that the CDF of the random variable X defined in Eq. (2) is the following

$$\begin{aligned}
F_X(x) &= \frac{1}{2} \cdot \mathbf{1}_{[1,+\infty)}(x) + \frac{1}{2} \cdot F_Z(x) \\
&= \frac{1}{2} \cdot \mathbf{1}_{[1,+\infty)}(x) + \frac{1}{2} \cdot \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt,
\end{aligned} \tag{3}$$

where F_X denotes the CDF of X, and F_Z denotes the CDF of Z (i.e., the CDF of $N(0,1)$).

Please prove the formula in Eq. (3).

First let

$$\begin{cases} A_X = \{\omega \in \Omega : X(\omega \leq x)\} \\ B = \{\omega \in \Omega : Y(\omega) = 1\} \end{cases}$$

Then by definition of CDF and the law of total probability

$$\begin{aligned}
F_X(x) = \mathbb{P}(A_x) &= \mathbb{P}(A_X|B) \cdot \mathbb{P}(B) + \mathbb{P}(A_X|B^c) \cdot \mathbb{P}(B^c) \\
&= \frac{\mathbb{P}((X \leq x) \cap (Y = 0))}{\mathbb{P}(Y = 0)} \cdot \mathbb{P}(Y = 0) + \frac{\mathbb{P}((X \leq x) \cap (Y = 1))}{\mathbb{P}(Y = 1)} \cdot \mathbb{P}(Y = 1) \\
&= \mathbb{P}((X \leq x) \cap (Y = 0)) + \mathbb{P}((X \leq x) \cap (Y = 1))
\end{aligned}$$

Looking at the first term,

$$\begin{aligned}
\mathbb{P}((X \leq x) \cap (Y = 0)) &= \mathbb{P}(\{\omega \in \Omega : (Y(\omega) + (1 - Y(\omega)) \cdot Z(\omega) \leq x) \cap Y(\omega) = 0\}) \\
&= \mathbb{P}(\{\omega \in \Omega : (0 + 1 \cdot Z(\omega) \leq x) \cap Y(\omega) = 0\}) \\
&= \mathbb{P}(\{\omega \in \Omega : (Z(\omega) \leq x) \cap Y(\omega) = 0\}) \\
&= \mathbb{P}(\{\omega \in \Omega : Y(\omega) = 0\}) \cdot \mathbb{P}(\{\omega \in \Omega : Z(\omega) \leq x\}) \quad \text{(by independence of Y and Z)} \\
&= \frac{1}{2} F_Z(x)
\end{aligned}$$

And similarly,

$$\begin{aligned}
\mathbb{P}((X \leq x) \cap (Y = 1)) &= \mathbb{P}(\{\omega \in \Omega : (Y(\omega) + (1 - Y(\omega)) \cdot Z(\omega) \leq x) \cap Y(\omega) = 1\}) \\
&= \mathbb{P}(\{\omega \in \Omega : (1 + 0 \cdot Z(\omega) \leq x) \cap Y(\omega) = 1\}) \\
&= \mathbb{P}(\{\omega \in \Omega : (1 \leq x) \cap Y(\omega) = 1\}) \\
&= \mathbb{P}(\{\omega \in \Omega : Y(\omega) = 1\}) \cdot \mathbb{P}(\{\omega \in \Omega : 1 \leq x\}) \quad \text{(by independence of Y and Z)} \\
&= \frac{1}{2} \mathbf{1}_{[1,+\infty)}
\end{aligned}$$

so

$$F_X(x) = \frac{1}{2} \mathbf{1}_{[1,+\infty)} + \frac{1}{2} F_Z(x) \quad \blacksquare$$

3. (2 points) Let Y, Z, and W be random variables defined on the probability space $(\Omega, \mathbb{P})$. Suppose

   - $Y \sim \text{Bernoulli}(p)$;
   - the CDFs of Z and W are F_Z and F_W, respectively;
   - Y, Z, and W are mutually independent, i.e., Y and Z are independent, Y and W are independent, Z and W are independent.

   We define a new random variable X by $X(\omega) \stackrel{\text{def}}{=} Y(\omega) \cdot Z(\omega) + (1 - Y(\omega)) \cdot W(\omega)$ for all $\omega \in \Omega$. **Please prove that the CDF of X is the following**

$$F_X(x) = p \cdot F_Z(x) + (1-p) \cdot F_W(\omega).$$

From the law of total probability,

$$F_X(x) = \mathbb{P}(A_x) = \mathbb{P}(A_X|B) \cdot \mathbb{P}(B) + \mathbb{P}(A_X|B^c) \cdot B^c$$

where

$$\begin{aligned}
A_X &= \{\omega \in \Omega : X(\omega) \leq x\} = \{\{\omega \in \Omega : Y(\omega) \cdot Z(\omega) + (1 - Y(\omega)) \cdot W(\omega) \leq x\}\} \\
B &= \{\omega \in \Omega : Y(\omega) = 1\} \implies \mathbb{P}(B) = p \\
B^c &= \{\omega \in \Omega : Y(\omega) = 0\} \implies \mathbb{P}(B^c) = 1 - p
\end{aligned}$$

Then,

$$\begin{aligned}
F_X(x) &= \frac{\mathbb{P}(A_X \cap B)}{\mathbb{P}(B)} \cdot \mathbb{P}(B) + \frac{\mathbb{P}(A_X \cap B^c)}{\mathbb{P}(B^c)} \\
&= \mathbb{P}(A_X \cap B) + \mathbb{P}(A_X \cap B^c) \\
&= \mathbb{P}((Y \cdot Z + (1-Y) \cdot W \leq x) \cap (Y = 1)) + \mathbb{P}((Y \cdot Z + (1-Y) \cdot W \leq x) \cap (Y = 0)) \\
&= \mathbb{P}((Z \leq x) \cap (Y = 1)) + \mathbb{P}((W \leq x) \cap (Y = 0)) \\
&= F_Z(x) \cdot \mathbb{P}(Y = 1) + F_W(x) \cdot \mathbb{P}(Y = 0) \quad \text{(by independence of Y and Z)} \\
&= p \cdot F_Z(x) + (1-p) \cdot F_W(x) \quad \blacksquare
\end{aligned}$$

4. (2 points) Let Y, Z, and W be random variables defined on the probability space $(\Omega, \mathbb{P})$. Suppose

- $Y \sim \text{Bernoulli}(1/3)$;
- $Z \sim \text{Pois}(1)$;
- $W \sim N(0,1)$;
- Y, Z, and W are mutually independent, i.e., Y and Z are independent, Y and W are independent, Z and W are independent.

We define a new random variable X by $X(\omega) \overset{\text{def}}{=} Y(\omega) \cdot Z(\omega) + (1 - Y(\omega)) \cdot W(\omega)$ for all $\omega \in \Omega$. Let F_X denote the CDF of X. **Please draw the graph of $F_X(x)$ for $-1 \le x \le 5.5$, i.e.,**

$$\{(x, F_X(x)) : -1 \le x \le 5.5\}.$$

From question 3, the CDF of $X(\omega) = Y(\omega) \cdot Z(\omega) + (1 - Y(\omega)) \cdot W(\omega)$ where $Y \sim \text{Bernoulli}(p)$ is

$$F_X(x) = pF_Z(x) + (1-p)F_W(x)$$

Then,

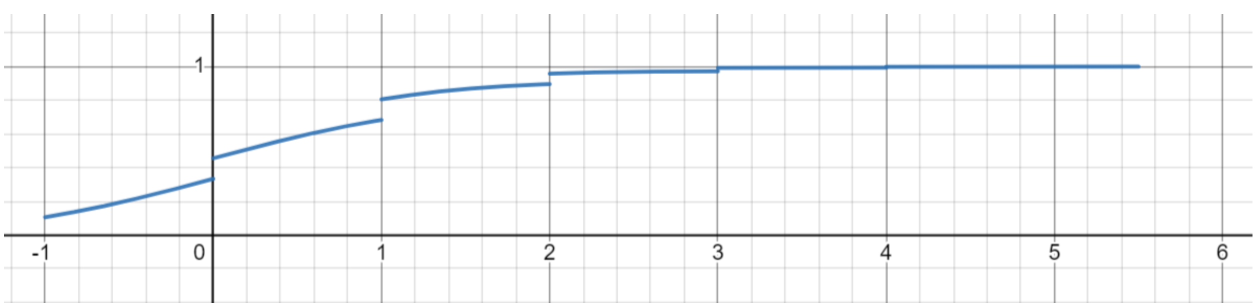
$$Z \sim \text{Pois}(1) \implies F_Z(x) = \sum_{k=0}^{x} \frac{1^k e^{-1}}{k!} \cdot \mathbf{1}_{[x_k, +\infty)}(x) = \frac{1}{e} \sum_{k=0}^{x} \frac{1}{k!} \cdot \mathbf{1}_{[k,+\infty)}$$

and

$$W \sim N(0,1) \implies F_Z(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt$$

$$\begin{aligned} F_X(x) &= \frac{1}{3} F_Z(x) + \frac{2}{3} F_W(x) \\ &= \frac{1}{3e} \sum_{k=0}^{x} \frac{1}{k!} \cdot \mathbf{1}_{[k,+\infty)} + \frac{2}{3} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}\, dt \end{aligned}$$

Using Desmos,



5. (2 points) Let $p_k = \frac{\lambda^k e^{-\lambda}}{k!}$ for all $k = 1, 2, \ldots$. **Please prove the following identity**

$$\sum_{k=0}^{\infty} k \cdot p_k = \lambda. \tag{4}$$

Remark: Eq. (4) shows that the "expected value" of Pois(λ). We will discuss the concept of expected values in Chapter 3 of my lecture notes.

$$\sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

Then because $e^{-\lambda}$ does not depend on k, we have

$$e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!}$$

but by the Taylor series expansion of the exponential function, we have

$$\sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = \lambda e^{\lambda}$$

so

$$\sum_{k=0}^{\infty} k \cdot p_k = e^{-\lambda} \cdot e^{\lambda} \cdot \lambda = 1 \cdot \lambda = \lambda \quad \blacksquare$$

APMA 1655 Honors Statistical Inference I

Homework 5

Name: Milan Capoor *Due: 11 pm, March 17*

Collaborators: Garv Gaur

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

All the materials in the Review section are from my lecture notes. Please read the Review section before going to the problem set.

1.1 Series

Please take the following results for granted:

- $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$ (see Basel problem);
- $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \log(2)$, where log is the natural logarithm; this series is usually referred to as the alternating harmonic series;
- $\sum_{k=1}^{\infty} \frac{1}{k} = +\infty$, which is usually referred to as the harmonic series.

1.2 Definition of CDFs

Definition 1.1 *Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$. For each real number x, we define an event $A_x \overset{\text{def}}{=} \{\omega : \omega \in \Omega \text{ and } X(\omega) \leq x\}$. The probability of A_x, i.e., $\mathbb{P}(A_x)$, is conventionally denoted as $\mathbb{P}(X \leq x)$. We call the function F_X defined as follows as the* ***cumulative distribution function (CDF)*** *of X*

$$F_X : \quad \mathbb{R} \to [0, 1],$$
$$x \mapsto \mathbb{P}(X \leq x).$$

Long story short, $F_X(x) = \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$.

Takeaway: Every CDF is a function defined on $\mathbb{R}$ = the collection of all real numbers.

1.3 CDFs Produce Random Variables

Theorem 1.1 *Suppose a function $F : \mathbb{R} \to \mathbb{R}$ satisfies the following three properties*

i) F is non-decreasing, i.e., $F(x_1) \leq F(x_2)$ if $x_1 \leq x_2$;

ii) we have the following limits

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1;$$

iii) F is right-continuous, i.e.,[1]

$$\lim_{x\to x_0+} F(x) = F(x_0).$$

Then, there exist a probability space $(\Omega, \mathbb{P})$ and a random variable X such that the CDF of X is the given function F.

1.4 Definition of Expected Values

Definition 1.2 *Let X be a discrete random variable whose CDF is $\sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$. If $\sum_{k=0}^{K} |x_k| \cdot p_k < \infty$, then the following sum is called the mean/expected value of X and denoted as $\mathbb{E}X$*

$$\boxed{\mathbb{E}X \overset{\text{def}}{=} \sum_{k=0}^{K} x_k \cdot p_k.} \tag{1.1}$$

If $\sum_{k=0}^{K} |x_k| \cdot p_k = \infty$, we say the expected value of x does not exist.

Definition 1.3 *Let X be a continuous random variable with PDF $p_X(x)$. If $\int_{-\infty}^{+\infty} |x| \cdot p_X(x)dx < \infty$, we call the following integral as the mean/expected value of X and denote it as $\mathbb{E}X$*

$$\boxed{\mathbb{E}X \overset{\text{def}}{=} \int_{-\infty}^{+\infty} x \cdot p_X(x)dx.} \tag{1.2}$$

If $\int_{-\infty}^{+\infty} |x| \cdot p_X(x)dx = \infty$, we say the expected value of X does not exist.

1.5 Definition of $\mathbb{E}[g(X)]$

Definition 1.4 *Suppose g is a real-valued function defined on $\mathbb{R}$.*

i) Let X be a discrete random variable whose CDF is $\sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$. If $\sum_{k=0}^{K} |g(x_k)| \cdot p_k < \infty$, then the following sum is called the mean/expected value of $g(X)$ and denoted as $\mathbb{E}[g(X)]$

$$\boxed{\mathbb{E}\left[g(X)\right] \overset{\text{def}}{=} \sum_{k=0}^{K} g(x_k) \cdot p_k.}$$

If $\sum_{k=0}^{K} |g(x_k)| \cdot p_k = \infty$, we say the expected value of $g(X)$ does not exist.

[1] The sign "$\lim_{x\to x_0+}$" denotes a right-sided limit.

ii) Let X be a continuous random variable with PDF $p_X(x)$. If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx < \infty$, we call the following integral as the mean/expected value of $g(X)$ and denote it as $\mathbb{E}[g(X)]$

$$\boxed{\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} g(x) \cdot p_X(x)dx.}$$

If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx = \infty$, we say the expected value of X does not exist.

Remark: In graduate-level probability theory courses, you will learn that Definition 1.4 is actually a theorem called the law of the unconscious statistician (LOTUS).[2]

[2] "The naming is sometimes attributed to Sheldon Ross' textbook Introduction to Probability Models, although he removed the reference in later editions."

2 Problem Set

Problem 1. (2 points) Suppose we have a function $F(x)$ defined by the following

$$
\begin{aligned}
&F(x) = \sum_{k=0}^{+\infty} p_k \cdot \mathbf{1}_{[x_k,+\infty)}(x), \quad \text{where} \\
&x_k = (-1)^{k+1} \cdot k, \quad \text{for all } k = 0, 1, 2, \ldots, \\
&p_0 = 0 \text{ and } p_k = \frac{6}{\pi^2} \cdot \frac{1}{k^2}, \quad \text{for all } k = 1, 2, 3, \ldots
\end{aligned}
\tag{2.1}
$$

Please prove the following: There exist a probability space $(\Omega, \mathbb{P})$ and a random variable X such that the CDF of X is the function F defined in Eq. (2.1).
(Hint: Apply Theorem 1.1 in the Review section.)
By theorem 1.1, there exist a probability space $(\Omega, \mathbb{P})$ and a random variable X such that the CDF of X is the given function F if F satisfies the three following properties:

i) F is non-decreasing ($F(x_1) \leq F(x_2) \quad \text{if} \quad x_1 \leq x_2$)

ii) The following limits are true:

$$\lim_{x \to -\infty} F(x) = 0$$
$$\lim_{x \to \infty} F(x) = 1$$

iii) F is right-continuous ($\lim_{x \to x_0^+} F(x) = F(x_0)$)

1) The first property is obviously true as F is defined as a summation and all the terms are zero (from the indicator function product) or positive (as $p_k = \frac{6}{(\pi k)^2}$ is positive for all $k > 0$).

2) Then as x goes to $-\infty$,

$$\lim_{x \to -\infty} F(x) = \sum_{k=0}^{\infty} p_k \cdot \lim_{x \to -\infty} \mathbf{1}_{[x_k,\infty)}(x)$$

but the minimum value of $x_k = 0$ so

$$\lim_{x \to -\infty} \mathbf{1}_{[x_k,\infty)}(x) = 0$$

and clearly

$$\lim_{x \to -\infty} F(x) = \sum_{k=0}^{\infty} 0 = 0.$$

Similarly, for $\lim_{x \to \infty} F(x)$, we must maximize x_k:

$$\lim_{k \to \infty} x_k = \lim_{k \to \infty} (-1)^{k+1} \cdot k = \infty$$

so

$$\lim_{x \to \infty} \mathbf{1}_{[x_k,\infty)}(x) = \lim_{x \to \infty} \mathbf{1}_{[\infty,\infty)}(x) = 1$$

then $\lim_{x\to\infty} F(x)$ reduces to

$$\lim_{x\to\infty} F(x) = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{6}{\pi^2} \cdot \frac{1}{k^2} = \frac{6}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{k^2} = \frac{6}{\pi^2} \cdot \frac{\pi^2}{6} = 1$$

showing that the second property also holds.

3) Finally, to prove that F is right continuous, notice that it is the sum of a product of two functions. Note that the second function,

$$\mathbf{1}_{[x_k,\infty]}(x) = \begin{cases} 0 & x < x_k \\ 1 & x \geq x_k \end{cases}$$

which is a stepwise constant function that is right continuous because of the $\geq$ sign. Then, the only time the first function p_k is discontinuous is at $k = 0$. However, because of the infinite summation, this zero term does not actually change the value of F. Then for all other k, p_k is just a constant which acts as a scaling factor for the stepwise function which is right continuous. Therefore, F is right-continuous. Then all three properties hold and by theorem 1.1, there exist a probability spave and a random variable X such that F is the CDF of X. ■

Problem 2. (2 points) Let X be a continuous random variable, and the PDF of X is the following[3]

$$p_X(x) = \frac{1}{\pi\left(1+x^2\right)}, \quad \text{for all } x \in \mathbb{R}. \tag{2.2}$$

i) For each $\varepsilon > 0$, please calculate the following integral

$$\int_{-\varepsilon}^{\varepsilon} x \cdot p_X(x) dx.$$

$$\begin{aligned}
\int_{-\varepsilon}^{\varepsilon} x \cdot p_X(x) dx &= \frac{1}{\pi} \int_{-\varepsilon}^{\varepsilon} \frac{x}{1+x^2}\, dx \\
&\overset{u=1+x^2}{=} \frac{1}{\pi} \int_{-\varepsilon}^{\varepsilon} \frac{1}{2u}\, du \\
&= \frac{1}{\pi}\left[\frac{1}{2}\ln(x^2+1)\right]_{-\varepsilon}^{\varepsilon} \\
&= \frac{1}{2\pi}\ln(\varepsilon^2+1) - \frac{1}{2\pi}\ln((-\varepsilon)^2+1) \\
&= \boxed{0}
\end{aligned}$$

ii) Please calculate the following limit

$$\lim_{\varepsilon \to +\infty} \int_{-\varepsilon}^{\varepsilon} x \cdot p_X(x) dx.$$

$$\lim_{\varepsilon \to +\infty} \int_{-\varepsilon}^{\varepsilon} x \cdot p_X(x) dx = \lim_{\varepsilon \to +\infty} 0 = \boxed{0}$$

iii) Does the expected value of X exist? Please prove your answer.

The expected value exists if

$$\int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx < \infty$$

Here, that is

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{|x|}{1+x^2}\, dx < \infty$$

And because both $|x|$ and $\frac{1}{x^2}$ are even, this is equal to

$$\begin{aligned}
\frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}\, dx &= \frac{2}{\pi}\lim_{t\to\infty}\left[\frac{1}{2}\ln(x^2+1)\right]_0^t \\
&= \frac{2}{\pi}\lim_{t\to\infty}\frac{1}{2}\ln(t^2+1) \\
&= \infty
\end{aligned}$$

So the expected value does not exist ■

[3]The PDF corresponds to a Cauchy-Lorentz distribution.

Problem 3. (2 points) The following integral is called the Gaussian integral, and we take it for granted

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}. \tag{2.3}$$

Suppose μ and σ are two real numbers and given. Let X be a continuous random variable, and $X \sim N(\mu, \sigma^2)$, i.e., the PDF of X is the following

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Please calculate the expected value $\mathbb{E}X$. Please provide the details of your calculation.
(Hint: Apply the formula in Eq. (2.3).)
First confirming that the expected value exists:

$$\int_{-\infty}^{\infty} |x| \cdot p_X(x)\,dx \overset{?}{<} \infty$$

$$\begin{aligned}
\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} &= -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx \\
&= \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^u\,du - \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} e^u\,du \\
&= \frac{1}{\sqrt{2\pi}}\left(\left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right]_{-\infty}^{0} - \left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right]_{0}^{\infty}\right) \\
&= \frac{1}{\sqrt{2\pi}}\left((e^{-\frac{\mu^2}{2\sigma^2}} - 1) - (1 - e^{-\frac{\mu^2}{2\sigma^2}})\right) \\
&= \frac{1}{\sqrt{2\pi}}(2e^{-\frac{\mu^2}{2\sigma^2}} - 2) < \infty \quad \checkmark
\end{aligned}$$

So the expected value exists and is equal to

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x \cdot p_X(x)\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2}\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left(\mu \int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2}\,dx - \int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2}\,dx\right)
\end{aligned}$$

Let

$$u = \frac{x-\mu}{\sigma\sqrt{2}} \Longrightarrow dx = \sigma\sqrt{2}\,du$$

so

$$
\begin{aligned}
\mathbb{E}X &= \frac{1}{\sigma\sqrt{2\pi}}\left(x\sigma\sqrt{2}\int_{-\infty}^{\infty} e^{-u^2}\ du - \sigma\sqrt{2}\int_{-\infty}^{\infty} e^{-u^2}\ du\right) \\
&= \frac{1}{\sigma\sqrt{2\pi}}(\mu\sigma\sqrt{2}\sqrt{\pi} - \sigma\sqrt{2}\sqrt{\pi}) \\
&= \boxed{\mu}
\end{aligned}
$$

Problem 4. (2 points) Suppose λ is a given positive number. Let X be a continuous random variable, and the PDF of X is the following[4]

$$p_X(x) = \lambda \cdot e^{-\lambda x} \cdot \mathbf{1}_{[0,+\infty)}(x), \quad \text{for all } x \in \mathbb{R}.$$

Please calculate the expected value $\mathbb{E}X$. Please provide the details of your calculation.

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x \cdot \lambda \cdot e^{-\lambda x} \cdot \mathbf{1}_{[0,+\infty)}(x)\, dx \\
&= \int_{-\infty}^{0} x\lambda e^{-\lambda x} \cdot 0\, dx + \int_{0}^{\infty} x\lambda e^{-\lambda x} \cdot 1\, dx \\
&= \int_{0}^{\infty} x\lambda e^{-\lambda x}\, dx \\
&= \lambda \int_{0}^{\infty} x e^{-\lambda x}\, dx \\
&= \lambda \left[x \cdot -\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} - \int_0^{\infty} e^{-\lambda x}\, dx \\
&= 0 - \left[\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} \\
&= -(0 - \frac{1}{\lambda} e^{0}) \\
&= \boxed{\frac{1}{\lambda}}
\end{aligned}$$

[4]The PDF corresponds to the exponential distribution denoted as $\text{Exp}(\lambda)$

Problem 5. (2 points) Let X be a random variable defined on the probability space $(\Omega, \mathbb{P})$. Suppose $X \sim N(0,1)$. We define a new random variable Y by the following

$$Y(\omega) \stackrel{\text{def}}{=} |X(\omega)|, \quad \text{for all } \omega \in \Omega.$$

Please calculate the expected value $\mathbb{E}Y$ (see Definition 1.4). Please provide the details of your calculation.

For a continuous random random variable

$$\begin{aligned}
\mathbb{E}[\,|X|\,] &= \int_{-\infty}^{\infty} |x| \cdot p_X(x)\, dx \\
&= \int_{-\infty}^{\infty} |x| \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \\
&= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x e^{-\frac{1}{2}x^2}\, dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x e^{-\frac{1}{2}x^2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{u}\, du - \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{u}\, du \\
&= \frac{1}{\sqrt{2\pi}} \left(\left[e^{-\frac{1}{2}x^2} \right]_{-\infty}^{0} - \left[e^{-\frac{1}{2}x^2} \right]_{0}^{\infty} \right) \\
&= \frac{1}{\sqrt{2\pi}} ((1-0) - (0-1)) \\
&= \frac{2}{\sqrt{2\pi}} = \frac{\sqrt{2}}{\sqrt{\pi}} = \boxed{\sqrt{\frac{2}{\pi}}}
\end{aligned}$$

APMA 1655 Honors Statistical Inference I

Homework 6

Name: Milan Capoor *Due: 11 pm, March 24*

Collaborators:

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

`All the materials in the Review section are from my lecture notes. Please read the Review section before going to the problem set. - Mike`

1.1 Definition of $\mathbb{E}[g(X)]$

Definition 1.1 *Suppose g is a real-valued function defined on $\mathbb{R}$.*

i) Let X be a discrete random variable whose CDF is $\sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$. If $\sum_{k=0}^{K} |g(x_k)| \cdot p_k < \infty$, then the following sum is called the mean/expected value of $g(X)$ and denoted as $\mathbb{E}[g(X)]$

$$\boxed{\mathbb{E}\left[g(X)\right] \stackrel{\text{def}}{=} \sum_{k=0}^{K} g(x_k) \cdot p_k.}$$

If $\sum_{k=0}^{K} |g(x_k)| \cdot p_k = \infty$, we say the expected value of $g(X)$ does not exist.

ii) Let X be a continuous random variable with PDF $p_X(x)$. If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx < \infty$, we call the following integral as the mean/expected value of $g(X)$ and denote it as $\mathbb{E}[g(X)]$

$$\boxed{\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} g(x) \cdot p_X(x)dx.}$$

If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx = \infty$, we say the expected value of X does not exist.

Remark: In graduate-level probability theory courses, you will learn that Definition 1.1 is actually a theorem called the law of the unconscious statistician (LOTUS).[1]

[1] "The naming is sometimes attributed to Sheldon Ross' textbook Introduction to Probability Models, although he removed the reference in later editions."

1.2 Definition of Variance

Definition 1.2 *Let X be a random variable whose expected value $\mathbb{E}X$ exists. Define the function $g(x) = (x - \mathbb{E}X)^2$. Then, the expected value of $g(X)$ is called the variance of X if it exists; we denote the variance of X as* $\mathrm{Var}(X)$. *Briefly,*

$$\boxed{\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]}. \tag{1.1}$$

Remark 1.1 *i) If the* $\mathrm{Var}(X) = +\infty$, *we say the variance of X does not exist.*

ii) The variance of X is interpreted as the "expected squared deviation of X from its mean."

iii) Let X be a discrete random variable whose CDF is $\sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$ with $K \leq +\infty$. Then, $\mathrm{Var}(X)$ *is represented as follows if it exists*

$$\boxed{\mathrm{Var}(X) = \sum_{k=0}^{K} (x_k - \mathbb{E}X)^2 \cdot p_k, \qquad \textit{where } \mathbb{E}X = \sum_{k=0}^{K} x_k \cdot p_k.}$$

iv) Let X be a continuous random variable with PDF $p_X(x)$, the variance $\mathrm{Var}(X)$ *is represented as follows if it exists*

$$\boxed{\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}X)^2 \cdot p_X(x)dx, \qquad \textit{where } \mathbb{E}X = \int_{-\infty}^{+\infty} x \cdot p_X(x)dx.}$$

1.3 Properties of Variance

Theorem 1.1 *Let X be either a continuous or discrete random variable.*[2] *We have the following*

i) $\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2}$

ii) For any constants a and b, we have $\boxed{\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X).}$

iii) The variance of any constant is zero, i.e., $\mathrm{Var}(c) = 0$ *for all $c \in \mathbb{R}$.*

iv) If $\mathrm{Var}(X) = 0$, *there exists a constant c such that $\mathbb{P}(\{\omega \in \Omega : X(\omega) = c\}) = 1$. That is, "random variable has zero variance if and only if this random variable is 'almost surely' constant."*

[2] In fact, Theorem 1.1 is true for all random variables.

2 Problem Set

Problem 1 (2 points) We consider the probability space $(\Omega, \mathbb{P})$ and random variable X defined as follows

- $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \le x \le 1\}$ = the collection of real numbers between 0 and 1.
- For any $A \subset \Omega$, we define $\mathbb{P}(A)$ = the "length" of A; for example, if $A = (0.3, 0.5)$, then $\mathbb{P}(A)$ = the length of the interval $(0.3, 0.5)$, which is equal to $0.5 - 0.3 = 0.2$.
- Define X as a function on the sample space $\Omega = [0,1]$ by

$$X(\omega) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \omega \neq 0.5, \\ 0 & \text{if } \omega = 0.5. \end{cases}$$

  Since Ω is a sample space, the function X is a random variable defined on this sample space.

Please answer the following questions by "yes" or "no," and you do not need to provide any explanations.

i) Is the expected value $\mathbb{E}X$ equal to 1? [Circle Yes or No.]

   $\boxed{\text{Yes}}$

ii) Is the expected value $\mathbb{E}(X^2)$ equal to 1? Precisely, we define the function $g(x) = x^2$; is $\mathbb{E}[g(X)]$ equal to 1? [Circle Yes or No.]

   $\boxed{\text{Yes}}$

iii) Is the variance $\mathrm{Var}(X)$ equal to 0? [Circle Yes or No.]

   $\boxed{\text{Yes}}$

iv) Is there a constant c such that $X(\omega) = c$ for all $\omega \in \Omega$? [Circle Yes or No.]

   $\boxed{\text{No}}$

Problem 2. (2 points) Let X be a discrete random variable whose CDF is

$$F_X(x) = \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,+\infty)}(x),$$

where $K < +\infty$. We define the following function of ξ

$$V(\xi) \stackrel{\text{def}}{=} \sum_{k=0}^{K} (x_k - \xi)^2 \cdot p_k.$$

Please find the minimum of the function $V(\xi)$.

Because V is a finite sum of positive terms, we can take the derivative of each term individually, so:

$$V'(\xi) = \sum_{k=0}^{K} -2(x_k - \xi) \cdot p_k$$

Then, the extrema shall occur where $V'(\xi) = 0$:

$$\sum_{k=0}^{K} -2(x_k - \xi) \cdot p_k = 0$$

$$\sum_{k=0}^{K} -2x_k p_k + 2\xi p_k = 0$$

$$\xi \cdot \sum_{k=0}^{K} p_k = \sum_{k=0}^{K} x_k \cdot p_k$$

However, as

$$F_X(x) = \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x)$$

is the CDF of X, we know that

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(x) = \sum_{k=0}^{K} p_k \cdot \mathbf{1}_{[x_k,\infty)}(\infty) = \sum_{k=0}^{K} p_k = 1$$

so

$$\xi = \sum_{k=0}^{K} x_k \cdot p_k = \mathbb{E}X$$

then because

$$V''(\xi) = \sum_{k=0}^{K} 2p_k = 2 > 0$$

for all ξ we know that their is only one minimum which was found above. Then at last, the minimum value of the function is

$$\boxed{V(\mathbb{E}X) = \sum_{k=0}^{K} (x_k - \mathbb{E}X)^2 \cdot p_k = \text{Var}(X)}$$

Problem 3. (2 points)

i) We define the following function of z

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt, \quad \text{for } z > 0,$$

which is conventionally referred to as the Gamma function. **Please prove the following formula**

$$\Gamma(z+1) = z \cdot \Gamma(z), \quad \text{for } z > 0.$$

Hint: Use integration by parts.

$$\begin{aligned}
\Gamma(z+1) &= \int_0^{\infty} t^z e^{-t}\, dt \\
&\overset{IBP}{=} \left[t^z(-e^{-t})\right]_0^{\infty} - \int_0^{\infty} z t^{z-1}(-e^{-t})\, dt \\
&= (\infty(0) - 0(1)) + z \int_0^{\infty} t^{z-1} e^{-t}\, dt \\
&= z \cdot \Gamma(z) \quad \blacksquare
\end{aligned}$$

ii) You can take the following result as granted and apply it without proving it

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{2.1}$$

Let $X \sim N(0,1)$. **Please calculate the variance** $\mathrm{Var}(X)$**. Please provide the details of your calculation.**
Hint: You may use the result in part i) and Eq. (2.1).

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

so first we calculate the expected value of X:

$$\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx$$

where

$$X \sim N(0,1) \implies p_X = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

so

$$\begin{aligned}
\mathbb{E}X &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{x^2}{2}}\, dx \\
&\overset{u=x^2/2}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u}\, du \\
&= \frac{1}{\sqrt{2\pi}} \left[-e^{-x^2/2}\right]_{-\infty}^{\infty} \\
&= \frac{1}{\sqrt{2\pi}}(0-0) = 0
\end{aligned}$$

Then $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{x^2}{2}} \, dx$$

Then, as the integrand is a product of squares, the function is odd and

$$\mathbb{E}(X^2) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} x^2 \cdot e^{-\frac{x^2}{2}} \, dx$$

then with the substitution $t = x^2/2$ we have

$$\mathbb{E}(X^2) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} \sqrt{2t} \cdot e^{-t} \, dt = \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} t^{\frac{1}{2}} \cdot e^{-t} \, dt$$

But that integral is just $\Gamma(\frac{3}{2})$! So

$$\mathbb{E}(X^2) = \frac{2}{\sqrt{\pi}}\Gamma(\frac{3}{2}) = \frac{2}{\sqrt{\pi}}(\frac{1}{2} \cdot \Gamma(\frac{1}{2})) = \frac{2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = 1$$

and at long last,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1 - 0^2 = \boxed{1}$$

Problem 4. (2 points) Let X be either a continuous or discrete random variable. Let $g_j(x)$, for $j \in \{1, 2, \ldots, J\}$ with $J < \infty$, be functions. **Please prove the following:** If each $\mathbb{E}[g_j(X)]$ exists, then $\mathbb{E}[\sum_{j=1}^{J} g_j(X)]$ exists and $\mathbb{E}[\sum_{j=1}^{J} g_j(X)] = \sum_{j=1}^{J} \mathbb{E}[g_j(X)]$.

First note that for $\mathbb{E}[\sum_{j=1}^{J} g_j(X)]$ to exist means that for

$$f(X) = \sum_{j=1}^{J} g_j(X),$$

$$\sum_{k=0}^{K} |f(X)| \cdot p_k < \infty$$

However, as we know that each $\mathbb{E}[g_j(X)]$ exists and $J < \infty$, we know that $f(X)$ is finite so the sum must also be finite (assuming $K < \infty$). Thus, $\mathbb{E}[\sum_{j=1}^{J} g_j(X)]$ exists. Then as each $\mathbb{E}[g_j(X)]$ exists, we also know that for each g_j, $\sum_{k=0}^{K} g_j(X) \cdot p_k$ is absolutely convergent. Therefore, in the expression

$$\mathbb{E}[\sum_{j=1}^{J} g_j(X)] = \sum_{k=0}^{K} \sum_{j=1}^{J} g_j(X) \cdot p_k$$

we can interchange the limits by the properties of double summations such that

$$\sum_{k=0}^{K} \sum_{j=1}^{J} g_j(X) \cdot p_k = \sum_{j=1}^{J} \sum_{k=0}^{K} g_j(X) \cdot p_k = \sum_{j=1}^{J} \mathbb{E}[g_j(X)]$$

giving us the desired result that

$$\mathbb{E}[\sum_{j=1}^{J} g_j(X)] = \sum_{j=1}^{J} \mathbb{E}[g_j(X)] \quad \blacksquare$$

Problem 5 (2 points)

Problem 5 (2 points) Let X be either a continuous or discrete random variable. **Please prove the following:** For any constants a and b, we have $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$.

$$\begin{aligned}
\mathrm{Var}(aX + b) &= \mathbb{E}[(aX + b)^2] - (\mathbb{E}[aX + b])^2 \\
&= \mathbb{E}(a^2X^2 + 2abX + b^2) - (\mathbb{E}[aX + b])(\mathbb{E}[aX + b]) \\
&= a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2 - (a\mathbb{E}(X) + b)(a\mathbb{E}(X) + b) \\
&= a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2 - (a^2(\mathbb{E}[X])^2 + 2ab\mathbb{E}(X) + b^2) \\
&= a^2\mathbb{E}(X^2) - a^2(\mathbb{E}X)^2 \\
&= a^2(\mathbb{E}(X^2) - (\mathbb{E}X)^2) \\
&= a^2\mathrm{Var}(X) \quad \blacksquare
\end{aligned}$$

APMA 1655 Honors Statistical Inference I

Homework 7

Name: Milan Capoor *Due: 11 pm, April 7*

Collaborators:

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

```
All the materials in the Review section are from my lecture notes. Please read the
Review section before going to the problem set. - Mike
```

1.1 iid-ness

Definition 1.1 *Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of random variables defined on the probability space $(\Omega, \mathbb{P})$. We say random variables $X_1, X_2, \ldots$ are independent if*

$$\mathbb{P}\left(\{\omega \in \Omega : X_i(\omega) \in A_i \text{ for all } i = 1, 2, \ldots, n\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(\{\omega \in \Omega : X_i(\omega) \in A_i\})$$

for any positive integer n and any subsets $A_1, A_2, \ldots, A_n \subset \mathbb{R}$.

Definition 1.2 *Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of random variables defined on the probability space $(\Omega, \mathbb{P})$. If the random variables $X_1, X_2, \ldots, X_n, \ldots$ share the same CDF, i.e., $F_{X_1} = F_{X_2} = \cdots = F_{X_n} = \cdots$, we say the random variables $X_1, X_2, \ldots, X_n, \ldots$ are identically distributed. Furthermore, if $X_1, X_2, \ldots, X_n, \ldots$ are independent, we say they are* ***independent and identically distributed (iid)****.*

1.2 LLN

Theorem 1.1 *(**Strong LLN**) Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of random variables defined on the probability space $(\Omega, \mathbb{P})$. Suppose the random variables $X_1, X_2, \ldots, X_n, \ldots$ are independent and identically distributed (iid) and the expected value $\mathbb{E}X_1$ exists. Then, we have the following*

$$\begin{aligned} &\mathbb{P}(A) = 1, \quad \text{where} \\ &A \overset{\text{def}}{=} \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)}{n} = \mathbb{E}X_1\right\}. \end{aligned} \tag{1.1}$$

1.3 Mean and Variance of $g(X)$

Let X be a continuous random variable with PDF $p_X(x)$, and g is a real-valued function. Then,

- $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot p_X(x)dx$, if the integral $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx < \infty$;

- $\mathrm{Var}[g(X)] = \mathbb{E}\left[(g(X) - \mathbb{E}[g(X)])^2\right] = \int_{-\infty}^{+\infty} (g(x) - \mathbb{E}[g(X)])^2 \cdot p_X(x)dx$, if the expected value exists.

The scenario for discrete random variables is similar.

1.4 Generalized LLN

Theorem 1.2 *Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of random variables defined on the probability space $(\Omega, \mathbb{P})$. Suppose the random variables $X_1, X_2, \ldots, X_n, \ldots$ are iid, and $g(x)$ is a real-valued function. If the expected value $\mathbb{E}[g(X_1)]$ exists, we have the following*

$$\begin{aligned}
&\mathbb{P}(A) = 1, \quad \textit{where} \\
&A \overset{\text{def}}{=} \left\{\omega \in \Omega \,:\, \lim_{n\to\infty} \frac{g\left(X_1(\omega)\right) + g\left(X_2(\omega)\right) + \cdots + g\left(X_n(\omega)\right)}{n} = \mathbb{E}[g(X_1)]\right\}.
\end{aligned}$$

1.5 Monte Carlo Integration

Based on the generalized LLN (Theorem 1.2), the Monte Carlo integration works for all integrals of the following form

$$\int_{-\infty}^{+\infty} g(x) \cdot p_X(x)dx,$$

where $p_X(x)$ is a PDF and $g(x)$ is a continuous function such that $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x)dx < +\infty$. We conclude the Monte Carlo integration method by the following algorithm

---

Algorithm 1 : Monte Carlo Integration

Input: (i) Function $g(x)$; (ii) PDF $p_X(x)$; (iii) sample size n.
Output: An estimator $\widehat{v}_n$ of the integral $v = \int_{-\infty}^{\infty} g(x) \cdot p_X(x)dx$.
1: Generate random numbers $X_1, \ldots, X_n$ iid from a distribution whose PDF is $p_X(x)$.
2: $\widehat{v}_n \leftarrow \frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

---

2 Problem Set

Problem 1. (6 points) We consider the following:

- The experiment: We flip a fair coin infinitely many times.
- Since there are infinitely many flips, one possible outcome ω is an infinitely long sequence of H and T, e.g.,

$$\omega = (H, H, T, H, H, \ldots).$$

  Then, the sample space Ω of this experiment is a collection of all such infinitely long sequences of H and T, i.e.,

$$\Omega \stackrel{\text{def}}{=} \left\{ \omega = (\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \ldots) \,\middle|\, \omega^{(i)} \text{ is either } H \text{ or } T, \text{ for all } i = 1, 2, 3, \ldots \right\}.$$

- For each fixed positive integer i, we define the following

$$X_i(\omega) = \begin{cases} 1 & \text{if } \omega^{(i)} = H, \text{ i.e., the i-th flip gives } H, \\ 0 & \text{if } \omega^{(i)} = T, \text{ i.e., the i-th flip gives } T. \end{cases}$$

  That is, all the $X_1, X_2, X_3, \ldots$ are defined on the sample space Ω. Hence, they are random variables.

- For each positive integer k, we define the following event,

$$A_k \stackrel{\text{def}}{=} \left\{ \omega \in \Omega \,:\, \lim_{n\to\infty} \frac{X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)}{n} = \frac{1}{k} \right\} \tag{2.1}$$

Suppose there exists a probability $\mathbb{P}$ defined on Ω satisfying the following: for any positive integer n and $v_1, v_2, \ldots, v_n \in \{0, 1\}$, we have

$$\mathbb{P}\left(\{\omega \in \Omega \,:\, X_i(\omega) = v_i \text{ for all } i = 1, \ldots, n\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{\omega \in \Omega \,:\, X_i(\omega) = v_i\}\right) = \frac{1}{2^n}.$$

(We take the existence of such $\mathbb{P}$ for granted.[1])

Questions:

i) Please prove that $X_1, X_2, X_3, \ldots$ are iid.

Lemma 1: $X_1, X_2, X_3, \ldots$ are independent.

Proof: Let S_i be a subset of Ω. By definition, it is a sequence of H and T. Then, as X_i are random variables defined on Ω which map $\{H, T\} \to \{1, 0\}$, $v_1, ..., v_n \in \{0, 1\}$ are also in S_i, which is itself a subset of Ω.

Then as $v_i = \{0, 1\}$, $S_i \subset \Omega$.

[1]The proof of the existence needs the Kolmogorov extension theorem, which is far beyond the scope of APMA 1655 and is omitted.

From the given this gives us,

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{\omega \in \Omega \,:\, X_i(\omega) = v_i\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{n}\{\omega \in \Omega \,:\, X_i(\omega) \in S_i\}\right) = \frac{1}{2^n} = \prod_{n=1}^{n}\frac{1}{2}$$

Note though that as X_i correspond to the toss of a *fair* coin, the probability of a $X_i(\omega) = v_i$ for any particular value of v_i is $\frac{1}{2}$. Thus,

$$\prod_{n=1}^{n}\frac{1}{2} = \prod_{n=1}^{n}\mathbb{P}(X_i = v_i)$$

which together with the above reduces to

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{\omega \in \Omega \,:\, X_i(\omega) \in S_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(\{\omega \in \Omega \,:\, X_i(\omega) \in S_i\})$$

showing that $X_1, ...X_n$ are independent. $\square$

Lemma 2: $X_1, X_2, X_3, \ldots$ are identically distributed.

Proof: Then, as each X_i is determined by the outcome of a single independent coin flip, the CDF of any particular X_i is

$$F_{X_i}(x) = \frac{1}{2}\mathbf{1}_{[0,\infty)}(x) + \frac{1}{2}\mathbf{1}_{[1,\infty)}(x)$$

and thus all $X_1, X_2, ...$ share the same CDF so they are identically distributed. $\square$

Conclusion: $X_1, X_2, X_3, \ldots$ are iid. $\square$

ii) Please calculate the probability $\mathbb{P}(A_1 \cup A_2)$ and provide the reasoning for your calculation, where A_1 and A_2 are defined in Eq. (2.1).

Let $\bar{X} := \lim_{n\to\infty}\bar{X}_n$.Notice that $A_1 = \{\bar{X} = 1\}$ and $A_2 = \{\bar{X} = \frac{1}{2}\}$ are disjoint as $\bar{X}$ is a single real number so cannot have two values. Thus,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Then because $X_1, X_2, ...$ are iid, they share the same CDF,

$$F_{X_i}(x) = \frac{1}{2}\mathbf{1}_{[0,\infty)} + \frac{1}{2}\mathbf{1}_{[1,\infty)}$$

which has expected value

$$\mathbb{E}X_i = \frac{1}{2}(0) + \frac{1}{2}(1) = \frac{1}{2}.$$

And by the Law of Large Numbers,

$$\mathbb{P}(\bar{X} = \mathbb{E}X_1) = 1$$

so using the calculated value of $\mathbb{E}X_1$,

$$\mathbb{P}(A_2) = \mathbb{P}(\bar{X} = \frac{1}{2}) = \mathbb{P}(\bar{X} = \mathbb{E}X_1) = 1$$

Then from the probability properties, $\mathbb{P}(A_1) = 0$ and

$$\boxed{\mathbb{P}(A_1 \cup A_2) = 1}$$

iii) Please calculate the following probability and provide the reasoning for your calculation

$$\mathbb{P}\left(\bigcup_{k=3}^{\infty} A_k\right).$$

For the same reason as above, note that $A_k \quad (k \geq 1)$ are disjoint and

$$\mathbb{P}\left(\bigcup_{k=3}^{\infty} A_k\right) = \sum_{k=3}^{\infty} \mathbb{P}(A_k)$$

(Proof: HW 1)

And

$$\mathbb{P}\left(A_1 \cup A_2 \cup \bigcup_{k=3}^{\infty} A_k\right) = 1$$

$$\implies \mathbb{P}(A_1 \cup A_2) + \mathbb{P}\left(\bigcup_{k=3}^{\infty} A_k\right) = 1$$

$$\implies \mathbb{P}\left(\bigcup_{k=3}^{\infty} A_k\right) = 1 - \mathbb{P}(A_1 \cup A_2) = \boxed{0}$$

Problem 2. (4 points) Suppose we want to calculate the approximate value of the following integral

$$\int_{-\infty}^{+\infty} \left(\sum_{k=0}^{100} x^k\right) \cdot e^{-\frac{x^2}{2}}\, dx.$$

Please design an approach to estimating the integral above and justify your approach.

First recognize that

$$e^{-\frac{x^2}{2}} = \sqrt{2\pi} p_X(x)$$

where p_X is the PDF of the random variable X corresponding to the standard normal distribution. Then let

$$g(x) = \sum_{k=0}^{100} x^k$$

which will be a continuous polynomial with 100 terms. Together, these facts reduce the given integral to one of the form

$$\sqrt{2\pi} \int_{-\infty}^{\infty} g(x) \cdot p_X(x)$$

where $g(x)$ is continuous and $p_X(x)$ is a PDF, allowing us to use Monte-Carlo Integration to determine the value. First generate a large number of independently and identically random values X_i from the standard normal distribution. Then for each X_i calculate $g(X_i)$ and add the value to a running sum. After calculating

$$\sum_{i=1}^{n} g(X_i)$$

divide the sum by n. By the Law of Large numbers, this value will be approximately equal to

$$\int_{-\infty}^{+\infty} \left(\sum_{k=0}^{100} x^k\right) \cdot e^{-\frac{x^2}{2}}\, dx$$

given a large enough n.

APMA 1655 Honors Statistical Inference I

Homework 8

Name: Milan Capoor *Due: 11 pm, April 14*

Collaborators:

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

```
All the materials in the Review section are from my lecture notes. Please read the
Review section before going to the problem set. - Mike
```

1.1 PDFs

Definition 1.1 *Let F_X be a CDF of a continuous random variable X. We define the following function p_X as the* ***probability density function (PDF)*** *of the continuous random variable X*

$$\begin{aligned} p_X : \quad & \mathbb{R} \to \mathbb{R}, \\ & x \mapsto p_X(x) = F_X'(x) = \frac{d}{dx} F_X(x). \end{aligned}$$

1.2 Calculation of $\mathbb{E}[g(X)]$

Let X be a continuous random variable with PDF $p_X(x)$, and g is a real-valued function. Then,

- $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot p_X(x) dx$, if the integral $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x) dx < \infty$;
- $\text{Var}[g(X)] = \mathbb{E}\left[(g(X) - \mathbb{E}[g(X)])^2\right] = \int_{-\infty}^{+\infty} (g(x) - \mathbb{E}[g(X)])^2 \cdot p_X(x) dx$, if the expected value exists.

The scenario for discrete random variables is similar.

1.3 CLT

Theorem 1.1 *(CLT) Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of iid random variables defined on $(\Omega, \mathbb{P})$. Suppose their expected values and variances exist. We define a sequence of random variables $\{G_n\}_{n=1}^{\infty}$ by the following*

$$G_n(\omega) := \sqrt{n} \cdot \left(\overline{X}_n(\omega) - \mathbb{E}X_1\right) = \sqrt{n} \cdot e_n(\omega). \tag{1.1}$$

i) (Heuristic version.) When n is large, G_n looks like a random variable following the normal distribution $N(0, \text{Var}(X_1))$.

ii) (Rigorous version.) The CDF of G_n converges to the CDF of $N(0, \text{Var}(X_1))$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : G_n(\omega) \le x\}) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi \cdot \text{Var}(X_1)}} \cdot \exp\left(-\frac{t^2}{2 \cdot \text{Var}(X_1)}\right) dt.$$

1.4 MGFs

Definition 1.2 *Let X be a random variable. The MGF $M_X(t)$ of X is defined by the following*

$$\boxed{M_X(t) := \mathbb{E}\left[e^{tX}\right]}$$

provided this expected value exists for all $t \in \mathbb{R}$.

Remark: The definition of MGFs is motivated by the Fourier transform.

Analyzing the limit of MGFs is easier than analyzing the limit of CDFs in the proof of CLT because of the following

Theorem 1.2 *Let $X_1, X_2, \ldots, X_n$ be independent random variables defined on $(\Omega, \mathbb{P})$. We define $S_n(\omega) = \sum_{i=1}^n X_i(\omega)$. Then, we have*

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t).$$

Furthermore, if $X_1, X_2, \ldots, X_n$ are not just independent, but also identically distributed, then

$$M_{S_n}(t) = \Big(M_{X_1}(t)\Big)^n.$$

To prove Theorem 1.2, we need the following lemma

Lemma 1.1 *Let $X_1, X_2, \ldots, X_n$ are random variables. Suppose $f_1, f_2, \ldots, f_n$ are functions such that the following expected values exist: $\mathbb{E}[f_1(X_1) \cdot f_2(X_2) \cdot \cdots \cdot f_n(X_n)], \mathbb{E}[f_1(X_1)], \mathbb{E}[f_2(X_2)], \ldots, \mathbb{E}[f_n(X_n)]$. If $X_1, X_2, \ldots, X_n$ are* ***independent****, we have*

$$\boxed{\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(X_i)\right].}$$

Proof of the lemma involves some real analysis concepts, hence, omitted. The idea of the proof is very similar to that in Problem 1 of HW 7 but much more complicated.

Proof of Theorem 1.2: Because $e^{t\,S_n(\omega)} = e^{t\,\sum_{i=1}^n X_i(\omega)} = \prod_{i=1}^n e^{t\,X_i(\omega)}$, we have

$$M_{S_n}(t) = \mathbb{E}[e^{t\,S_n}] = \mathbb{E}\left[\prod_{i=1}^n e^{t\,X_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{t\,X_i}\right] = \prod_{i=1}^n M_{X_i}(t),$$

where the third equal sign comes from Lemma 1.1. $\square$

2 Problem Set

Problem 1. (3 points) Let X be a continuous random variable defined on the probability space $(\Omega, \mathbb{P})$ and $p_X(x)$ the PDF of X. Let $g(x)$ be a continuous function defined on $\mathbb{R}$. We assume the following

- the PDF $p_X(x)$ is continuous, and $p_X(x) > 0$ for all $x \in \mathbb{R}$;
- $g(x)$ is strictly increasing; furthermore,

$$\lim_{x \to -\infty} g(x) = -\infty \quad \text{and} \quad \lim_{x \to +\infty} g(x) = +\infty.$$

- the inverse function $g^{-1}(x)$ has a continuous derivative $\frac{d}{dx} g^{-1}(x)$.

We define a random variable Y by the following

$$Y(\omega) \stackrel{\text{def}}{=} g(X(\omega)), \quad \text{for all } \omega \in \Omega.$$

Please prove that the PDF $p_Y(x)$ of Y is the following

$$p_Y(x) = p_X\left(g^{-1}(x)\right) \cdot \frac{d}{dx} g^{-1}(x).$$

Let F_Y be the CDF of Y. Then by definition of the CDF and Y,

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(g(X) \leq x)$$

Then as g is strictly increasing and its inverse has a continuous derivative, its inverse is continuous. Thus,

$$\mathbb{P}(g(X) \leq x) = \mathbb{P}(X \leq g^{-1}(x)) = F_X(g^{-1}(x))$$

Taking the derivatives,

$$F'_Y(x) = F'_X(g^{-1}(x))$$

$$p_Y(x) = p_X(g^{-1}(x)) \cdot \frac{d}{dx} g^{-1}(x) \quad \blacksquare$$

Problem 2. (3 points) Suppose we are under the condition of the CLT (Theorem 1.1). **Please prove the following:** The CDF of $\frac{G_n(\omega)}{\sqrt{\mathrm{Var}(X_1)}}$ converges to the CDF of $N(0,1)$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega : \frac{G_n(\omega)}{\sqrt{\mathrm{Var}(X_1)}} \le x\right\}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

(Hint: Apply the CLT.)

For $\{X_i\}_{i=1}^{\infty}$ an infinitely long sequence of iid RVs on $(\Omega, \mathbb{P})$, the sequence $\{G_n\}_{n=1}^{\infty}$ is defined by

$$G_n(\omega) = \sqrt{n} \cdot (\bar{X}_n(\omega) - \mathbb{E}X_1)$$

Then to prove that by the CLT,

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega : \frac{G_n(\omega)}{\sqrt{\mathrm{Var}(X_1)}} \le x\right\}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

it suffices to show that the moment generating function of $\frac{G_n(\omega)}{\sqrt{\mathrm{Var}(X_1)}}$ converges to the moment generating function of the standard normal distribution. From the definition of G_n,

$$\begin{aligned}
\tilde{G}_n(\omega) = \frac{G_n(\omega)}{\sqrt{\mathrm{Var}\, X_1}} &= \frac{\sqrt{n}(\bar{X}_n(\omega) - \mathbb{E}X_1)}{\sqrt{\mathrm{Var}\, X_1}} \\
&= \frac{\sqrt{n}}{\sqrt{\mathrm{Var}\, X_1}} \left(\frac{1}{n}\sum_{n=1}^{\infty} X_n - \mathbb{E}X_1\right) \\
&= \frac{\sqrt{n}}{\sqrt{\mathrm{Var}\, X_1}} \cdot \frac{1}{n}\sum_{i=1}^{n} (X_i - \mathbb{E}X_1) \\
&= \frac{1}{\sqrt{n \cdot \mathrm{Var}\, X_1}} \sum_{i=1}^{n} (X_i - \mathbb{E}X_1)
\end{aligned}$$

Then

$$\begin{aligned}
M_{\tilde{G}_n}(t) &= \mathbb{E}[e^{t\tilde{G}_n}] \\
&= \mathbb{E}[e^{\sum_{i=1}^{n}\left(\frac{t}{\sqrt{n\cdot \mathrm{Var}\, X_1}}(X_i - \mathbb{E}X_1)\right)}] \\
&= \left(\mathbb{E}[e^{\frac{t}{\sqrt{n\cdot \mathrm{Var}\, X_1}}}(X_1 - \mathbb{E}X_1)]\right)^n \\
&= \left(\mathbb{E}\left[1 + \frac{t}{\sqrt{n \cdot \mathrm{Var}\, X_1}}(X_1 - \mathbb{E}X_1) + \frac{t^2}{2(n\cdot \mathrm{Var}\, X_1)}(X_1 - \mathbb{E}X_1)^2 + \sum_{k=3}^{\infty} \frac{t^k}{k \cdot (n \cdot \mathrm{Var}\, X_1)^{k/2}}(X_1 - \mathbb{E}X)^k\right]\right)^n \\
&\approx (1 + \frac{t^2}{2(n \cdot \mathrm{Var}\, X_1)} \mathrm{Var}\, X_1)^n = (1 + \frac{t^2}{2n})^n
\end{aligned}$$

Then, denoting $C_n = \frac{t^2}{2n}$,

$$\lim_{n\to\infty} C_n = 0$$

and

$$\lim_{n\to\infty} n \cdot C_n = \frac{t^2}{2} = \lambda$$

we have

$$\lim_{n\to\infty} M_{\tilde{G}_n}(t) = (1 + c_n)^n = e^{\lambda} = \exp(\frac{t^2}{2}) = M_{\tilde{G}}(t)$$

Then examining the moment generating function of $Z \sim N(0,1)$,

$$\begin{aligned}
\mathbb{E}[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tx} \cdot p_Z(x)\, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(tx - \frac{x^2}{2})\, dx \\
&= \frac{e^{\frac{1}{2}t^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x-t}{\sqrt{2}}\right)^2}\, dx \\
&= e^{\frac{1}{2}t^2}
\end{aligned}$$

Which equals $M_{\tilde{G}}(t)$ from above. Hence, the moment generating function of $G_n(\omega)/\sqrt{\operatorname{Var} X_1}$ converges to the moment generating function of $N(0,1)$ and

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega : \frac{G_n(\omega)}{\sqrt{\operatorname{Var}(X_1)}} \leq x\right\}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \quad \blacksquare$$

Problem 3. (3 points) Suppose $X \sim N(\mu, \sigma^2)$, where μ is a real number and $\sigma > 0$. Please show that the moment-generating function of X is the following

$$M_X(t) = \exp(\mu t + \sigma^2 t^2/2).$$

The moment generating function of X is

$$\begin{aligned}
M_X(t) = \mathbb{E}[e^{tX}] &= \int_{-\infty}^{\infty} e^{tx} \cdot p_X(x)\, dx \\
&= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(tx - \left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(tx - \frac{1}{2\sigma^2}\left(x^2 - 2\mu x + \mu^2\right)\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{1}{2\sigma^2}\left(-x^2 + (2\mu - 2\sigma^2 t)x - \mu^2\right)\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(\frac{x^2 - (2\sigma^2 t + 2\mu)x + \mu^2}{2\sigma^2}\right)\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(\sqrt{\frac{x^2 - (2\sigma^2 t + 2\mu)x + \mu^2}{\sigma\sqrt{2}}}\right)^2\right)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left[\sqrt{2\pi}|\sigma| \exp(\frac{\sigma^2 t^2}{2 + \mu t})\right] \\
&= \exp(\frac{\sigma^2 t^2}{2} + \mu t) \quad \blacksquare
\end{aligned}$$

Correction: this derivation is incorrect

Problem 4. (1 point) Read Section 1.4 of this problem set and understand it.

APMA 1655 Honors Statistical Inference I

Homework 9

Name: Milan Capoor *Due: 11 pm, April 21*

Collaborators:

- You are strongly encouraged to work in groups, but solutions must be written independently.

1 Review

`All the materials in the Review section are from my lecture notes. Please read the Review section before going to the problem set. - Mike`

1.1 Random Vectors

Definition 1.1 *i) A (column) vector $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is called a **random vector** defined on $(\Omega, \mathbb{P})$ if each of its components is a random variable defined on $(\Omega, \mathbb{P})$.*

ii) The CDF of random vector $\boldsymbol{X}$ is an n-variate function defined as follows

$$\begin{aligned} F_{\boldsymbol{X}}(x_1, \ldots, x_n) &:= \mathbb{P}\left(\omega \in \Omega : X_1(\omega) \le x_1, \ldots, X_n(\omega) \le x_n\right) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \le x_i\}\right), \quad \text{for } x \in \mathbb{R}. \end{aligned}$$

iii) If $F_{\boldsymbol{X}}(x_1, \ldots, x_n)$ is differentiable, $\boldsymbol{X}$ is said to be a continuous random vector; in addition, the following n-variate function is called the PDF of $\boldsymbol{X}$

$$\boxed{p_{\boldsymbol{X}}(x_1, \ldots, x_n) := \underbrace{\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n}}_{n \text{ partial derivatives}} F_{\boldsymbol{X}}(x_1, \ldots, x_n) = \frac{\partial}{\partial x_1} \cdots \left(\frac{\partial}{\partial x_{n-1}} \left(\frac{\partial}{\partial x_n} F_{\boldsymbol{X}}(x_1, \ldots, x_n)\right)\right)}$$

Definition 1.2 *Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a continuous random vector with PDF $p_{\boldsymbol{X}}(x_1, \ldots, x_n)$ and $g(\boldsymbol{x}) = g(x_1, \ldots, x_n)$ an n-variate function. Then, the expected value of $g(\boldsymbol{X})$ is defined as follows*

$$\boxed{\mathbb{E}[g(\boldsymbol{X})] = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ integrals}} g(x_1, \ldots, x_n) \cdot p_{\boldsymbol{X}}(x_1, \ldots, x_n) \underbrace{dx_1 \cdots dx_n}_{n\ dx_i}}$$

provided $\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ integrals}} |g(x_1, \ldots, x_n)| \cdot p_{\boldsymbol{X}}(x_1, \ldots, x_n) \underbrace{dx_1 \cdots dx_n}_{n\ dx_i} < \infty.$

Theorem 1.1 *Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a random vector. If $X_1, X_2, \ldots, X_n$ are independent, we have*

$$\boxed{F_{\boldsymbol{X}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)}$$

where F_{X_i} is the CDF of X_i.

Proof of the theorem is quite complicated and omitted. The theorem directly implies the following corollary

Corollary 1.1 *Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a continuous random vector. If $X_1, X_2, \ldots, X_n$ are independent, we have*

$$\boxed{p_{\boldsymbol{X}}(x_1, \ldots, x_n) := \prod_{i=1}^{n} p_{X_i}(x_i),}$$

where p_{X_i} is the PDF of X_i.

1.2 MGFs

Definition 1.3 *Let X be a random variable. The MGF $M_X(t)$ of X is defined by the following*

$$\boxed{M_X(t) := \mathbb{E}\left[e^{tX}\right]}$$

provided this expected value exists for all $t \in \mathbb{R}$.

Function $M_X(t)$ is called the moment-generating function of X because of the following formula

$$\boxed{\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = \mathbb{E}[X^n],} \tag{1.1}$$

where $\mathbb{E}[X^n]$ is called the n-th moment of X.

Analyzing the limit of MGFs is easier than analyzing the limit of CDFs in the proof of CLT because of the following

Theorem 1.2 *Let $X_1, X_2, \ldots, X_n$ be independent random variables defined on $(\Omega, \mathbb{P})$. We define $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$. Then, we have*

$$M_{S_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Furthermore, if $X_1, X_2, \ldots, X_n$ are not just independent, but also identically distributed, then

$$M_{S_n}(t) = \left(M_{X_1}(t)\right)^n.$$

1.3 CLT

MGFs are needed in the proof of CLT because of the following theorem

Theorem 1.3 *Let $G, G_1, G_2, \ldots$ be random variables whose MGFs exist . If $\lim_{n\to\infty} M_{G_n}(t) = M_G(t)$ for all t, then G_n converges weakly to G (i.e., $\lim_{n\to\infty} F_{G_n}(x) = F_G(x)$ for all x).*

Lemma 1.1 *Let $\{c_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying $\lim_{n\to\infty} c_n = 0$. If $\lim_{n\to\infty} n \cdot c_n = \lambda \in \mathbb{R}$, we have*

$$\lim_{n\to\infty} (1+c_n)^n = e^{\lambda}.$$

Theorem 1.4 *(CLT) Let $\{X_i\}_{i=1}^{\infty}$ be an infinitely long sequence of iid random variables defined on $(\Omega, \mathbb{P})$. Suppose their expected values and variances exist. We define a sequence of random variables $\{G_n\}_{n=1}^{\infty}$ by the following*

$$G_n(\omega) := \sqrt{n} \cdot \left(\overline{X}_n(\omega) - \mathbb{E}X_1\right) = \sqrt{n} \cdot e_n(\omega). \tag{1.2}$$

i) (Heuristic version.) When n is large, G_n looks like a random variable following the normal distribution $N(0, \operatorname{Var}(X_1))$.

ii) (Rigorous version.) The CDF of G_n converges to the CDF of $N(0, \operatorname{Var}(X_1))$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : G_n(\omega) \le x\}) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi \cdot \operatorname{Var}(X_1)}} \cdot \exp\left(-\frac{t^2}{2 \cdot \operatorname{Var}(X_1)}\right) dt.$$

1.4 Proof of the CLT

Recall $G_n = \sqrt{n}\,(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n}(X_i - \mathbb{E}X_1)$. Then, Theorem 1.2 implies

$$\begin{aligned} M_{G_n}(t) &= \mathbb{E}[e^{t\,G_n}] \\ &= \mathbb{E}\left[e^{\sum_{i=1}^{n} \frac{t}{\sqrt{n}}(X_i - \mathbb{E}X_1)}\right] \\ &= \left(\mathbb{E}\left[e^{\frac{t}{\sqrt{n}}(X_1 - \mathbb{E}X_1)}\right]\right)^n. \end{aligned}$$

Furthermore, the Taylor expansion implies

$$\begin{aligned} M_{G_n}(t) &= \left(1 + \frac{t}{\sqrt{n}}\mathbb{E}[X_1 - \mathbb{E}X_1] + \frac{t^2/2}{n}\mathbb{E}[(X_1 - \mathbb{E}X_1)^2] + \sum_{k=3}^{\infty} \frac{t^k}{k!\,n^{k/2}}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k]\right)^n \\ &= \left(1 + \frac{t^2/2}{n}\operatorname{Var}(X_1) + \sum_{k=3}^{\infty} \frac{t^k}{k!\,n^{k/2}}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k]\right)^n. \end{aligned}$$

We denote $c_n := \frac{t^2/2}{n}\operatorname{Var}(X_1) + \sum_{k=3}^{\infty} \frac{t^k}{k!\,n^{k/2}}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k]$. Then, we have

$$n \cdot c_n = \frac{t^2}{2}\operatorname{Var}(X_1) + \sum_{k=3}^{\infty} \frac{t^k}{k!\,n^{k/2-1}}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k] \to \frac{t^2}{2}\operatorname{Var}(X_1), \quad \text{as } n \to \infty.$$

Therefore, Lemma 1.1 implies

$$\lim_{n\to\infty} M_{G_n}(t) = e^{\frac{t^2}{2}\cdot\operatorname{Var}(X_1)} = \text{ the MGF of } N(0,\, \operatorname{Var}(X_1)).$$

The desired result follows from Theorem 1.3. ■

2 Problem Set

Problem 1. (2 points) Prove Corollary 1.1 using Theorem 1.1.

Corollary: Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a continuous random vector. If $X_1, X_2, \ldots, X_n$ are independent, we have

$$p_{\boldsymbol{X}}(x_1, \ldots, x_n) := \prod_{i=1}^{n} p_{X_i}(x_i)$$

where p_{X_i} is the PDF of X_i.

Proof:

From theorem 1.1, if X_1, ..., X_n are independent and entries in a random vector $\mathbf{X}$, then the CDF of $\mathbf{X}$ is

$$F_{\mathbf{X}}(x, ..., x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

we may take n partial derivatives to find the PDF of $F_{\mathbf{X}}$ such that

$$\begin{aligned} p_{\mathbf{X}}(x_1, ..., x_n) &= \frac{\partial}{\partial x_1} ... \frac{\partial}{\partial x_n} \prod_{i=1}^{n} F_{X_i}(x_i) \\ &= \prod_{i=1}^{n} \frac{\partial}{\partial x_i} F_{X_i}(x_i) \quad \text{(by independence)} \\ &= \prod_{i=1}^{n} p_{X_i}(x_i) \quad \text{(by PDF definition)} \quad \blacksquare \end{aligned}$$

Problem 2. (2 points) Let $\boldsymbol{X} = (X_1, X_2)^T$ be a continuous random vector whose PDF is the following

$$p_{\boldsymbol{X}}(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2 + (x_2-\mu_2)^2\right]\right), \quad \text{for all } -\infty < x_1, x_2 < +\infty,$$

where μ_1 and μ_2 are two prespecified constants. Let $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ be real-valued functions defined as follows

$$g_1(x_1, x_2) = x_1, \quad g_2(x_1, x_2) = x_2.$$

Please calculate the expected values $\mathbb{E}[g_1(\boldsymbol{X})]$ and $\mathbb{E}[g_2(\boldsymbol{X})]$.

Assuming they exist,

$$\begin{aligned}
\mathbb{E}[g_1(X_1, X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1, x_2)\cdot p_{\mathbf{X}}(x_1, x_2)\, dx_1\, dx_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 \cdot \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2 + (x_2-\mu_2)^2\right]\right)\, dx_1\, dx_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mu_1 \cdot \sqrt{2\pi}\cdot \exp\left(-\frac{1}{2}(x_2-\mu_2)^2\right)\, dx_2 \\
&= \frac{\mu_1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(\frac{x_2-\mu_2}{\sqrt{2}})^2}\, dx_2 \\
&= \boxed{\mu_1}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[g_2(X_1, X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1, x_2)\cdot p_{\mathbf{X}}(x_1, x_2)\, dx_1\, dx_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 \cdot \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2 + (x_2-\mu_2)^2\right]\right)\, dx_1\, dx_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x_1-\mu_1)^2\right)\int_{-\infty}^{\infty} x_2\cdot \exp\left(-\frac{1}{2}(x_2-\mu_2)^2\right)\, dx_2\, dx_1 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x_1-\mu_1)^2\right)\cdot(\sqrt{2\pi}\cdot\mu_2)\, dx_1 \\
&= \frac{1}{2\pi}\cdot(\sqrt{2\pi}\cdot\mu_2)\cdot\sqrt{2\pi} \\
&= \boxed{\mu_2}
\end{aligned}$$

Problem 3. (4 points) Suppose we are under the condition of the CLT (Theorem 1.4). For every $\alpha > 0$, we define a random variable $G_{n,\alpha}$ as follows

$$G_{n,\alpha}(\omega) := n^{\alpha} \cdot \Big(\overline{X}_n(\omega) - \mathbb{E}X_1\Big).$$

Let $M_{G_{n,\alpha}}(t)$ denote the MGF of $G_{n,\alpha}$.

i) Suppose $\alpha < 0.5$. Please calculate the following limit for every $t \in \mathbb{R}$

$$\lim_{n\to\infty} M_{G_{n,\alpha}}(t).$$

(Hint: Apply Lemma 1.1.)

$$\begin{aligned}
G_{n,a} &= n^a \cdot \left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mathbb{E}X_1\right] \\
&= n^{a-1} \cdot \sum_{i=1}^{n}(X_i - \mathbb{E}X_1)
\end{aligned}$$

$$\begin{aligned}
M_{G_{n,a}}(t) &= \mathbb{E}[e^{tG_{n,a}}] = \mathbb{E}[\exp(tn^{a-1}\sum_{i=1}^{n}(X_i - \mathbb{E}X_1))] \\
&= \mathbb{E}\left[\exp(\sum_{i=1}^{n} tn^{a-1}(X_i - \mathbb{E}X_1))\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} \exp\left(tn^{a-1}(X_i - \mathbb{E}X_1)\right)\right] \quad \text{(by iid)} \\
&= \left(\mathbb{E}\left[\exp\left(tn^{a-1}(X_i - \mathbb{E}X_1)\right)\right]\right)^n \\
&= \left(\mathbb{E}\left[1 + tn^{a-1}(X_1 - \mathbb{E}X_1) + \frac{t^2n^{2a-2}}{2}(X_1 - \mathbb{E}X_1)^2 + \sum_{k=3}^{\infty}\frac{(tn^{a-1})^k}{k!}(X_1 - \mathbb{E}X_1)^k\right]\right)^n \\
&= \left(1 + \underbrace{\frac{t^2n^{2a-2}}{2}\operatorname{Var} X_1 + \sum_{k=3}^{\infty}\frac{(tn^{a-1})^k}{k!}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k]}_{C_n}\right)^n \quad \text{(because } \mathbb{E}[X_1 - \mathbb{E}X_1] = 0\text{)}
\end{aligned}$$

Then because $a < 0.5$, $k(a-1) < 0$ and

$$\lim_{n\to\infty} C_n = 0.$$

Thus by lemma 1.1

$$\lim_{n\to\infty}(1 + c_n)^n = \exp(\lim_{n\to\infty} n \cdot C_n)$$

$$n \cdot C_n = \frac{t^2n^{2n-1}}{2}\operatorname{Var} X_1 + \sum_{k=3}^{\infty}\frac{t^k}{k!}n^{ka-k+1}\mathbb{E}[(X_1 - \mathbb{E}X_1)^k]$$

But again because $a < 0.5$, $ka - k + 1 < 0$ for $k \geq 2$. Hence

$$\lim_{n\to\infty} n \cdot C_n = tn^{a-1}\mathbb{E}[X_1 - \mathbb{E}X_1] + \sum_{k=2}^{\infty} 0 = 0$$

So

$$\lim_{n\to\infty} (1 + c_n)^n = e^0 = 1$$

which means that

$$\boxed{\lim_{n\to\infty} M_{G_{n,a}}(t) = 1}$$

ii) If $\alpha > 0.5$, does the limit $\lim_{n\to\infty} M_{G_{n,\alpha}}(t)$ exist for all $t \in \mathbb{R}$? Please prove your answer. (Hint: Mimic Section 1.4 of the problem set.)

In the case where $a > 0.5$, the moment generating function is the same as above:

$$M_{G_{n,a}}(t) = \left(1 + \underbrace{\sum_{k=2}^{\infty} \frac{(tn^{a-1})^k}{k!} \mathbb{E}[(X_1 - \mathbb{E}X_1)^k]}_{C_n} \right)^n$$

But as $a > 0.5$, $(a-1)k < 0$ only when $0.5 < a < 1$ so for $a > 1$,

$$\lim_{t\to\infty} \lim_{k\to\infty} (tn^{a-1})^k > \lim_{k\to\infty} k!$$

so

$$\lim_{n\to\infty} C_n = \infty.$$

Thus

$$1 + C_n > 1 \implies \lim_{n\to\infty} (1 + C_n)^n = \infty$$

showing that the $\lim_{n\to\infty} M_{G_{m,a}}(t)$ does not exist. ■

Problem 4. (2 points) Let $X \sim N(\mu, \sigma^2)$. In HW 8, you proved that the MGF of X is the following

$$M_X(t) = \exp(\mu t + \sigma^2 t^2/2).$$

Please calculate $\mathbb{E}X$, $\mathbb{E}[X^2]$, and $\mathbb{E}[X^3]$ using Eq. (1.1).

From equation 1.1,

$$\frac{d^n}{dt^n} M_X(0) = \mathbb{E}[X^n]$$

Then

$$\begin{aligned}
\mathbb{E}X &= \frac{d}{dt} M_X(0) \\
&= \frac{d}{dt} \left[\exp(\mu t + \sigma^2 t^2/2)\right]_{t=0} \\
&= \left[(\mu + \sigma^2 t) \exp(\mu t + \sigma^2 t^2/2)\right]_{t=0} \\
&= \boxed{\mu}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^2] &= \frac{d^2}{dt^2} M_X(0) \\
&= \frac{d}{dt} \left[(\mu + \sigma^2 t) \exp(\mu t + \sigma^2 t^2/2)\right]_{t=0} \\
&= \left[(\mu + \sigma^2 t)^2 \exp(\mu t + \sigma^2 t^2/2) + \sigma^2 \exp(\mu t + \sigma^2 t^2/2)\right]_{t=0} \\
&= \boxed{\mu^2 + \sigma^2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^3] &= \frac{d^3}{dt^3} M_X(0) \\
&= \left[(\mu + \sigma^2 t)^2 \exp(\mu t + \sigma^2 t^2/2) + \sigma^2 \exp(\mu t + \sigma^2 t^2/2)\right]_{t=0} \\
&= \left[(\mu + \sigma^2 t)^3 \exp(\mu t + \sigma^2 t^2/2) + 2(\mu + \sigma^2 t)(\sigma^2) \exp(\mu t + \sigma^2 t^2/2) + \sigma^2(\mu + \sigma^2 t)\right]_{t=0} \\
&= \mu^3 + 2\mu\sigma^2 + \sigma^2\mu \\
&= \boxed{\mu^3 + 3\mu\sigma^2}
\end{aligned}$$

APMA 1655: Quiz 1

Milan Capoor

24 March 2023

Problem 1:

You have a fair, four-sided die with 1, 2, 3, and 4 on its four faces. Roll it two times, and record the two numbers on the upper face. Let X be the absolute value of the difference between the two numbers.

1. Please write down the CDF of X.

   The table of all possible values of the absolute difference is given as follows:

		Roll 1			
		1	2	3	4
Roll 2	1	0	1	2	3
	2	1	0	1	2
	3	2	1	0	1
	4	3	2	1	0

   So where $\mathbb{P}(A) := \frac{\#A}{16}$, the probabilities are:

$$\begin{cases} \mathbb{P}(X = 0) = \frac{4}{16} \\ \mathbb{P}(X = 1) = \frac{6}{16} \\ \mathbb{P}(X = 2) = \frac{4}{16} \\ \mathbb{P}(X = 3) = \frac{2}{16} \end{cases}$$

   so the CDF of X is

$$\boxed{F_X(x) = \frac{1}{4} \cdot \mathbf{1}_{[0,\infty)} + \frac{3}{8} \cdot \mathbf{1}_{[1,\infty)} + \frac{1}{4} \cdot \mathbf{1}_{[2,\infty)} + \frac{1}{8} \cdot \mathbf{1}_{[3,\infty)}}$$

2. Please calculate the expected value of X and provide the details of your calculation.

   The expected value of a discrete random variable is

   $$\mathbb{E}X = \sum_{k=0}^{K} x_k \cdot p_k$$

   so

   $$\mathbb{E}X = \frac{4}{16}(0) + \frac{6}{16}(1) + \frac{4}{16}(2) + \frac{2}{16}(3) = \frac{6+8+6}{16} = \frac{20}{16} = \boxed{1.25}$$

Problem 2:

Let $(\Omega, \mathbb{P})$ be a probability space; A and B are two events. Suppose we know the following

- $\mathbb{P}(A) = \frac{1}{4}$
- $\mathbb{P}(B|A) = \frac{1}{3}$
- $\mathbb{P}(A|B) = \frac{1}{2}$

Please calculate $\mathbb{P}(A \cup B)$ and provide the details of your calculation

By the definition of conditional probability

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{1}{3} \implies \mathbb{P}(A \cap B) = \frac{1}{12}$$
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2} \implies \mathbb{P}(A \cap B) = \frac{1}{2}\mathbb{P}(B)$$

So

$$\frac{1}{2}\mathbb{P}(B) = \frac{1}{12} \implies \mathbb{P}(B) = \frac{2}{12}$$

Then, as

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
$$\mathbb{P}(A \cup B) = \frac{3}{12} + \frac{2}{12} - \frac{1}{12} = \frac{4}{12} = \boxed{\frac{1}{3}}$$

Problem 3:

X is a continuous random variable defined on the probability space $(\Omega, \mathbb{P})$, and its PDF is $p_X(x)$. Let $F_X(x)$ denote the CDF of X. Suppose $p_X(x)$ is an even function. Please prove the following three formulas

1. $F_X(-a) = 1 - F_X(a)$ for any real number a

   As X is a continuous random variable,

   $$F_X(x) = \int_{-\infty}^{x} p_X(t)\, dt$$

   so

   $$F_X(-a) = \int_{-\infty}^{-a} p_X(t)\, dt$$

   Then as p_X is even $p_X(-t) = p_X(t)$ and

   $$\int_{-\infty}^{-a} p_X(t)\, dt = \int_{a}^{\infty} p_X(t)\, dt$$

   Correction: Justify this

   But $p_X = F_X'$ so by the fundamental theorem of calculus,

   $$\int_{a}^{\infty} p_X(t)\, dt = [F_X(t)]_a^{\infty} = \lim_{t\to\infty} F_X(t) - F_X(a)$$

   However, as F_X is a CDF $\lim_{t\to\infty} F_X(t) = 1$ and

   $$F_X(-a) = 1 - F_X(a) \quad \blacksquare$$

2. $\mathbb{P}(\{\omega \in \Omega : |X(\omega)| < a\}) = 2F_X(a) - 1$ for any positive a

   $$\begin{aligned}\mathbb{P}(\{\omega \in \Omega : |X(\omega)| < a\}) &= \mathbb{P}((X < a) \cap (X > -a)) \\ &= \mathbb{P}(X < a) + \mathbb{P}(X > -a) - \mathbb{P}((X < a) \cup (X > -a)) \\ &= (F_X(a) - \mathbb{P}(X = a)) + (1 - F_X(-a)) - 1 \\ &= F_X(a) - \mathbb{P}(X = a) - F_X(-a)\end{aligned}$$

   Then from part 1, $F_X(-a) = 1 - F_X(a)$ so

   $$\begin{aligned}F_X(a) - \mathbb{P}(X = a) - F_X(-a) &= F_X(a) - \mathbb{P}(X = a) - 1 + F_X(a) \\ &= 2F_X(a) - \mathbb{P}(X = a) - 1\end{aligned}$$

And X is continuous so $\mathbb{P}(X = a) = 0$, giving

$$\mathbb{P}(|X| < a) = 2F_X(x) - 1 \quad \blacksquare$$

3. $\mathbb{P}(\{\omega \in \Omega : |X(\omega)| > a\}) = 2(1 - F_X(a))$ for any positive a

   Similarly,

$$\mathbb{P}(|X| > a) = \mathbb{P}((X > a) \cap (X < -a))$$

   but these are disjoint so

$$\begin{aligned}
\mathbb{P}(|X| > a) &= \mathbb{P}(X > a) + \mathbb{P}(X < -a) \\
&= (1 - F_X(a)) + (F_X(-a) - \mathbb{P}(X = -a)) \\
&= 1 - F_X(a) + 1 - F_X(a) - \mathbb{P}(X = -a) \\
&= 2 - 2F_X(a) - 0 \\
&= 2(1 - F_X(a)) \quad \blacksquare
\end{aligned}$$

Problem 4:

Let X be a discrete random variable defined on the probability space $(\Omega, \mathbb{P})$. The CDF of X is the following

$$F_X(x) = \sum_{k=0}^{\infty} p_k \cdot \mathbf{1}_{[k,\infty)}(x)$$

Suppose the expected value $\mathbb{E}X = \sum_{k=0}^{\infty} k \cdot p_k$ exists. Please prove the following formula

$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(\{\omega \in \Omega : X(\omega) \geq k\})$$

Note that because k is defined on the integers,

$$\mathbb{P}(X \geq k) = \sum_{j=k}^{\infty} \mathbb{P}(X = j)$$

Then

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{P}(\{\omega \in \Omega : X(\omega) \geq k\}) &= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} \mathbb{P}(X = j) \\
&= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} p_j \\
&= \sum_{j=1}^{\infty} p_j + \sum_{j=2}^{\infty} p_j + \sum_{j=3}^{\infty} p_j + \dots \\
&= \left(p_1 + \sum_{j=2}^{\infty} \right) + \sum_{j=2}^{\infty} p_j + \sum_{j=3}^{\infty} p_j + \dots \\
&= \left(p_1 + p_2 + \sum_{j=3}^{\infty} \right) + \left(p_2 + \sum_{j=3}^{\infty} p_j \right) + \sum_{j=3}^{\infty} p_j + \dots \\
&= (1)p_1 + (2)p_2 + (3)p_3 + \dots \\
&= \sum_{k=1}^{\infty} k \cdot p_k \\
&= \mathbb{E}X \quad \blacksquare
\end{aligned}$$

APMA 1655: Quiz 2

Milan Capoor

26 April 2023

Problem 1:

Suppose $\{X_1, X_2, ..., X_{10000}\}$ are independently and identically distributed (iid) random variables defined on the probability space $(\Omega, \mathbb{P})$. In addition, $X_1 \sim \text{Bernoulli}(\frac{1}{2})$. We define a new random variable S by the following

$$S(\omega) := X_1(\omega) + X_2(\omega) + \ ... + X_{10000}(\omega) \quad \forall \omega \in \Omega$$

Please use the central limit theorem (CLT) to estimate the following probability approximately

$$\mathbb{P}(\{\omega \in \Omega : 4950 < S(\omega) \leq 5100\})$$

Please express your result in terms of the CDF F_G of the random variable $G \sim N(0,1)$. (No need to find the numerical value of probability.)

$$\begin{aligned} \mathbb{P}(4950 < S \leq 5100) &= \mathbb{P}(S \leq 5100) - \mathbb{P}(S \leq 4950) \\ &= F_S(5100) - F_S(4950) \end{aligned}$$

where $F_S(x)$ is the CDF of $S(\omega)$. By the corollary of the CLT, the CDF of

$$\frac{\sqrt{n} \cdot (\frac{S(\omega)}{n} - \mathbb{E}X_1)}{\sqrt{\text{Var}\, X_1}}$$

converges to F_G, the CDF of $N(0,1)$. Thus,

$$F_G(x) \approx \mathbb{P}(\frac{\sqrt{n} \cdot (\frac{S(\omega)}{n} - \mathbb{E}X_1)}{\sqrt{\text{Var}\, X_1}} \leq x)$$

Because $X_1 \sim \text{Bernoulli}(\frac{1}{2})$, $\mathbb{E}X_1 = \frac{1}{2}$ and $n = 10000$ Calculating $\text{Var}\, X_1$:

$$\begin{aligned}
\text{Var}\, X_1 &= \sum_{k=0}^{K} (x_k - \mathbb{E}X_1)^2 \cdot p_k \\
&= \frac{1}{2}(0 - \mathbb{E}X_1)^2 + \frac{1}{2}(1 - \mathbb{E}X_1)^2 \\
&= \frac{1}{2} \cdot (-\frac{1}{2})^2 + \frac{1}{2} \cdot (\frac{1}{2})^2 \\
&= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}
\end{aligned}$$

Thus $\sqrt{\text{Var}\, X_1} = \frac{1}{2}$ So

$$F_G(x) \approx \mathbb{P}(\frac{100(\frac{S(\omega)}{10000} - \frac{1}{2})}{\frac{1}{2}} \leq x) = \mathbb{P}(\frac{2}{100}S(\omega) - 100 \leq x) = \mathbb{P}(S \leq 50(x+100)) = F_S(50(x+100))$$

Back to the original problem,

$$\mathbb{P}(4950 < S \leq 5100) = F_S(5100) - F_S(4949)$$

So

$$\begin{cases} 5100 = 50(x + 100) \implies x = 2 \\ 4950 = 50(x + 100) \implies x = -1 \end{cases}$$

so

$$\boxed{\mathbb{P}(4950 < S \leq 5100) \approx F_G(2) - F_G(-1)}$$

Problem 2:

Suppose $\{X_1, X_2, ..., X_{10000}\}$ are independently and identically distributed (iid) random variables defined on the probability space $(\Omega, \mathbb{P})$. In addition, $X_1 \sim \text{Unif}(0,1)$, i.e., the CDF of X_1 is the following

$$F_{X_1}(x) = \begin{cases} 0 & \text{if } \ x \leq 0 \\ x & \text{if } \ 0 < x \leq 1 \\ 1 & \text{if } 1 < x \end{cases}$$

We define a new random variable T by the following

$$\begin{aligned} T(\omega) &:= \min\{X_1(\omega), X_2(\omega), \ ..., X_{10000}(\omega)\} \\ &= \text{the minimal value among } X_1(\omega), \ ..., X_{10000}(\omega) \qquad \forall \omega \in \Omega \end{aligned}$$

Please derive the CDF of the random variable T.

$$F_T(x) = \mathbb{P}(T \leq x) = 1 - \mathbb{P}(T > x)$$

However, by the definition of T, for it to be larger than a given x means that every X_i must be larger than x:

$$F_T(x) = 1 - \mathbb{P}((X_1 > x) \cap (X_2 > x) \cap ... \cap (X_{10000} > x))$$

by independence, this is the same as

$$F_T(x) = 1 - \prod_{i=1}^{10000} \mathbb{P}(X_i > x) = 1 - \prod_{i=1}^{10000} (1 - F_{X_i}(x))$$

But as the random variables X_i are also identically distributed, all the CDFs F_{X_i} are equal so

$$F_T(x) = 1 - (1 - F_{X_1}(x))^{10000}$$

Then by the definition of F_{X_1},

$$\boxed{F_T(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - (1-x)^{10000} & \text{if } 0 < x \leq 1 \\ 1 & \text{if } 1 < x \end{cases}}$$

Problem 3:

Let $\vec{X} = (X, Y)^T$ be a continuous random vector whose PDF is the following

$$p_{\vec{X}}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}[x^2 + y^2]\right) \qquad -\infty < x,\, y < \infty$$

Let $g(x, y)$ be a real-valued function defined by the following

$$g(x, y) = x^2 + y^2 \qquad -\infty < x\,, y < \infty$$

Please calculate the expected value $E[g(X)]$

$$\begin{aligned}
\mathbb{E}[g(\vec{X})] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) p_{\vec{X}}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2 + y^2) \cdot \frac{1}{2\pi} \exp(-\frac{1}{2}[x^2 + y^2])\, dx\, dy \\
&= \frac{1}{2\pi} \int_{0}^{\infty}\int_{0}^{2\pi} r^3 \exp(-\frac{r^2}{2})\, d\theta\, dr \\
&= \int_{0}^{\infty} r^3 e^{-\frac{r^2}{2}}\, dr \\
&= \frac{1}{2}\int_{0}^{\infty} u e^{-\frac{u}{2}}\, du \qquad (u = r^2) \\
&\overset{IBP}{=} \frac{1}{2}\left[-2ue^{-u/2}\right]_0^{\infty} - \frac{1}{2}\int_{0}^{\infty} -2e^{-u/2}\, du \\
&= \frac{1}{2}(0 - 0) + [-2e^{-u/2}]_0^{\infty} \\
&= 0 - (-2) = \boxed{2}
\end{aligned}$$