Sample Space (Ω): Set of all possible outcomes

Events: a set of outcomes / a subset of Ω

Countably infinite: if exists 1-1 correspondence $f: \Omega \to \mathbb{N}$

Disjoint: $S \cap T = \emptyset$

Mutually exclusive: $S_i \cap S_j = \emptyset,\ \forall i, j,\ i \neq j$

Complement: $S^c = \{x \in \Omega : x \notin S\}$

Subset: $S \subseteq T \iff$ For every $x \in S$, $x \in T$

Equal: $S = T \iff S \subseteq T$ and $T \subseteq S$

$\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} S_n = \{x \in \Omega : x$ appears in infinitely many $S_n\}$

$\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} S_n = \{x \in \Omega : x$ appears in all S_n after some $k\}$

$\bigcup$ increasing $\downarrow$ $A_1 \cap A_2 \cap A_3 \ldots \subseteq A_1 \subseteq A_1 \cup A_2 \cup A_3 \ldots$ $\downarrow$ decreasing $\bigcap$

$A_2 \cap A_3 \ldots \subseteq A_2 \subseteq A_2 \cup A_3 \ldots$

$$\bigcup_{n=1}^{\infty} S_n = \sup S_n \qquad \bigcap_{n=1}^{\infty} S_n = \inf S_n \quad \Rightarrow \quad \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} S_n = \lim_{n\to\infty} \sup S_n \qquad \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} S_n = \lim_{n\to\infty} \inf S_n$$

pf. $\underbrace{\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} S_n}_{(S)} = \underbrace{\{x \mid x \in S_k,\ \forall k \text{ except finitely many } k\}}_{(T)}$

($S \subseteq T$): Let x be an element of S. i.e. $x \in S$. Then there must exist some k such that $x \in \bigcap_{n=k}^{\infty} S_n$, then $x \in S_k, S_{k+1}, S_{k+2}, \ldots \Rightarrow x \in T$

($T \subseteq S$): pick some $y \in T$. so there must exist some $m \in \mathbb{N}$ s.t $y \in S_i\ \forall i \geq m \Rightarrow y \in \bigcap_{n=i}^{\infty} S_n\ \forall i \geq m \Rightarrow y \in \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} S_n$

pf. $\underbrace{\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} S_n}_{(S)} = \underbrace{\{x \mid x \in S_k \text{ for infinitly many } k\}}_{(T)}$

($S \subseteq T$): Let x be an element of S. i.e. $x \in S$. then $x \in \bigcup_{n=k}^{\infty} S_n, \forall k$. This implies that $x \in T$:

(assume $x \notin T \Rightarrow \exists m$ s.t $x \in S_{i_1}, S_{i_2}, \ldots, S_{i_m}$, $i_1 < i_2 < \cdots < i_m$, so $x \notin \bigcup_{n=i_m+1}^{\infty} S_n$. $\rightarrow\!\leftarrow$)

($T \subseteq S$): Let y be an element of T, i.e. $y \in T$, then there must exist a countably infinite seq. $\{i_m\}$ s.t $y \in S_m\ \forall m \in \mathbb{N}$. This implies that $y \in \bigcup_{n=k}^{\infty} S_n$. $\forall k \in \mathbb{N}$. $\Rightarrow y \in \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} S_n$

De Morgan's Laws

$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c \qquad \left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$$

3 Axioms of Probability

A probability assignment is valid if

1. $P(A) \geq 0$ for any event A
2. $P(\Omega) = 1$
3. $A_1, A_2 \ldots$ are mutually exclusive, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

$P(\emptyset) = 0 \quad (A_1 = \Omega,\ A_2 = A_3 \ldots = \emptyset)$

$A_1, \ldots, A_n$ are disjoint events, then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

$P(A) \leq 1$ for any A ($A \cup A^c = \Omega$, A and A^c are disjoint)

$P(A^c) = 1 - P(A) \quad (A_1 = A,\ A_2 = A^c,\ A_3 = A_4 \ldots = \emptyset)$

$P(A) = P(A - B) + P(A \cap B) \quad (A_1 = A - B,\ A_2 = A \cap B,\ A_3 = A_4 \ldots = \emptyset)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \left(\begin{array}{l} P(A \cup B) = P(A) + P(B - A) \\ P(B) = P(B - A) + P(A \cap B) \end{array}\right)$

Union Bound

$P(\bigcup_{n=1}^{N} A_n) \leq \sum_{n=1}^{N} P(A_n)$

for any events $A_1, \ldots, A_N$

(pf) step 1. $N = 2$ $P(A_1 \cup A_2) = P(A_1 \cup (A_2 - A_1))$ $= P(A_1) + P(A_2 - A_1) \leq P(A_1) + P(A_2)$

step 2. assume $N = k$ is true. then when $N = k+1$.

$P(\bigcup_{n}^{k+1} A_n) = P(\bigcup_{n}^{k} A_n \cup A_{k+1}) \leq P(\bigcup_{n}^{k} A_n) + P(A_{k+1})$ (step 1.)

$\leq \sum_{n=1}^{k} P(A_n) + P(A_{k+1})$ (assumption)

Continuity of Probability Function

Probability function $P(\cdot)$ is a function of events (needs to satisfy the 3 axioms)

$E_1, E_2 \ldots$ is increasing if $E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq E_{n+1} \subseteq \ldots$; decreasing if $E_1 \supseteq E_2 \supseteq \ldots \supseteq E_n \supseteq E_{n+1} \supseteq \ldots$

Theorem For any increasing / decreasing sequence of events $E_1, E_2 \ldots$, we have

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

(pf) $P(\lim_{n\to\infty} E_n) = P(\bigcup_{i=1}^{\infty} G_i) \overset{\text{Axiom\#3}}{=} \lim_{n\to\infty} \sum_{i=1}^{n} P(G_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} G_i)$

$= \lim_{n\to\infty} P(E_n)$ ($G_n = E_n - E_{n-1}$, mutually exclusive)

$B_k = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$: decreasing, $C_k = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$: increasing

Conditional Probability

$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$ = conditional probability of event A given B

Theorem (Reduction of Sample Space) Ω: sample space, $P(B) > 0$

1. $P(A \mid B) \geq 0$ for any event A
2. $P(\Omega \mid B) = 1$
3. $A_1, A_2 \ldots$ are mutually exclusive, then $P(\bigcup_{i=1}^{\infty} A_i \mid B) = \sum_{i=1}^{\infty} P(A_i \mid B)$

Tool #1: Multiplication Rule

$P(\bigcap_{i=1}^{n} A_i) = P(A_1) P(A_2 \mid A_1) \ldots P(A_n \mid A_1 \cap A_2 \ldots A_{n-1})$

Tool #2: Total Probability Theorem

$A_1, \ldots, A_n$ are partition of Ω and mutually exclusive, $P(A_i) > 0$

$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B) = P(A_1) P(B \mid A_1) + \ldots + P(A_n) P(B \mid A_n)$

Tool #3: Bayes' Rule

$P(A_i \mid B) = \dfrac{P(A_i) P(B \mid A_i)}{P(B)} = \dfrac{P(A_i) P(B \mid A_i)}{P(A_1) P(B \mid A_1) + \ldots + P(A_n) P(B \mid A_n)}$

Independence

$P(A \cap B) = P(A) P(B)$, if $P(B) > 0$, then $P(A \mid B) = P(A)$

If A, B are independent, then A, B^c are also independent

(pf) $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A) P(B) = P(A) \underbrace{(1 - P(B))}_{P(B^c)}$

$A_1, \ldots, A_n$ are independent if $P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$ for every $S \subseteq \{1, \ldots, n\}$

Conditionally independent if $P(\bigcap_{i \in S} A_i \mid B) = \prod_{i \in S} P(A_i \mid B)$

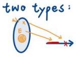
Random Variable

A function that maps each outcome to a real number

Noted as " $X: S \to S_X$, $S_X = \{X(\omega) \mid \omega \in S, X \in R\}$

two types: 1. discreate: take values over a discreate range

2. continuous: take values over a continuous range



*(range: $[0, 1]$)

CDF of X: $F_X(t) = P(X \leq t),\ \forall t \in \mathbb{R}$, $F_X(t) \in [0, 1]$

Cumulative Distribution Function

- $X \leq a$: $F_X(a)$ · $X > a$: $1 - F_X(a)$ · $X < a$: $F_X(a^-)$ or $\lim_{t\to a^-} F_X(t)$ or $F_X(a) - p(X=a)$
- $X \geq a$: $1 - F_X(a^-)$ · $X = a$: $F_X(a) - F_X(a^-)$

Property: 1. F_X is non-decreasing 2. $\lim_{t\to\infty} F_X(t) = 1$ 3. $\lim_{t\to-\infty} F_X(t) = 0$.

4. F_X is right continuous: $F_X(t^+) = F_X(t)$

$a < X \leq b$ $\quad F_X(b) - F_X(a)$

$a < X < b$ $\quad F_X(b^-) - F_X(a)$

$a \leq X \leq b$ $\quad F_X(b) - F_X(a^-)$

$a \leq X \leq b$ $\quad F_X(b^-) - F_X(a^-)$

PMF of X: $p(\cdot)$ of X is the function that satisfies

Probability Mass Function

1. $P(x_i) = P(X = x_i)$
2. $p(x) = 0$ if $x \notin \{x_1, x_2, x_3\}$
3. $\sum_{i=1}^{\infty} p(x_i) = 1$

Special Discrete Random Variables

1. Bernoulli Random Variables ($X \sim \text{Bernoulli}(p)$)

PMF : $P(X=k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \\ 0 & \text{otherwise} \end{cases}$ ($X \sim \text{Binomial}(1,p)$)

$E[X] = p$, $Var[X] = p(1-p)$

1 experiment trial with 2 possible outcomes

2. Binomial Random Variables ($X \sim \text{Binomial}(n,p)$)

PMF : $P(X=k) = \begin{cases} C_k^n p^k (1-p)^{n-k} & \text{if } k=0,1,2,\dots,n \\ 0 & \text{otherwise} \end{cases}$

$E[X] = np$, $Var[X] = np(1-p)$

n repetitions of the same Bernoulli experiment

Q (for fixed k, under $p=?$ is $P(X=k)$ is max) : $P(X=k) = C_k^N p^k (1-p)^{N-k} = f(p)$

$\Rightarrow \ln f(p) = \ln C_k^N + k \cdot \ln p + (N-k) \cdot \ln(1-p)$

$\Rightarrow \frac{d}{dp} \ln f(p) = \frac{k}{p} - \frac{N-k}{1-p} = 0 \Leftrightarrow k(1-p) = p(N-k) \Leftrightarrow \boxed{p = \frac{k}{N}}$

3. Poisson Random Variables ($X \sim \text{Poisson}(\lambda, T)$)

λ: Average Rate; T: Duration of observation window

PMF : $P(X=n) = \frac{e^{-\lambda T} (\lambda T)^n}{n!}$, $n = 0, 1, \dots$ ($X \sim \text{Binomial}(n, \frac{\lambda}{n})$)

$E[X] = Var[X] = \lambda T$

$\Rightarrow P(X=k) = C_k^n \cdot (\frac{\lambda}{n})^k \cdot (1-\frac{\lambda}{n})^{n-k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} (\frac{\lambda}{n})^k \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}$

$[\, n \to \infty : \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} \to 1, \ (1-\frac{\lambda}{n})^n = \left((1-\frac{\lambda}{n})^{-\frac{n}{\lambda}}\right)^{-\lambda} = e^{-\lambda}, \ (1-\frac{\lambda}{n})^k \to 1 \,]$

as $n \to \infty$, $P(X=k) \to \frac{e^{-\lambda} \lambda^k}{k!}$

$P(\Omega) = \sum_{n=0}^{\infty} P(X=n) = e^{-\lambda T} \sum \frac{(\lambda T)^n}{n!} = 1.$

Average rate is known and static

$\text{Poisson}(\lambda_1, T) + \text{Poisson}(\lambda_2, T) = \text{Poisson}(\lambda_1 + \lambda_2, T)$

4. Geometric Random Variables ($X \sim \text{Geometric}(p)$)

PMF : $P(X=k) = \begin{cases} (1-p)^{k-1} p & \text{if } k = 1,2,3\dots \\ 0 & \text{otherwise} \end{cases}$

Repetitions of the same Bernoulli experiment

$E[X] = \frac{1}{p}$, $E[X^2] = \frac{2-p}{p^2}$, $Var[X] = \frac{1-p}{p^2}$

$P(X = n+m \mid X > m) = P(X=n)$
$P(X > n+m \mid X > m) = P(X > n)$ (Memoryless Property)

5. Discrete Uniform Random Variables

PMF : $P(X=k) = \frac{1}{b-a+1}$, $k = a, a+1, \dots, b$

$E[X] = \frac{a+b}{2}$, $Var[X] = \frac{(b-a+1)^2 - 1}{12}$

Shannon Entropy

$H(X) := -\sum_i p(x_i) \log p(x_i) \equiv \mathbb{E}[-\log p(X)]$ base 2 : bits; base e : nats

Expected Value ($E[X] \equiv \mu_x$)

$E[X] := \sum_{x \in S} x \cdot \underbrace{p_X(x)}_{\text{PMF}} = \sum_{i=1}^{\infty} (x_i - x_{i-1})(1 - F_X(x_i^-))$ $\qquad E[\alpha X + \beta] = \alpha E[X] + \beta$

$E[g(X)] := \sum_{x \in S} g(x) p_X(x)$ (LOTUS) $\qquad E[g(X) + h(X)] = E[g(X)] + E[h(X)]$

Moments

$E[X^n]$: n-th moment $\quad E[e^{tX}]$: moment generating function

$E[(X - \mu_X)^n]$: n-th central moment

Variance (2nd central moment)

$Var[X] := E[(X-\mu_X)^2] = \sum_{x \in S} (x - \mu_X)^2 \cdot p_X(x) = E[X^2] - (E[X])^2$

Property:

1. $Var[X+c] = Var[X]$
2. $Var[aX] = a^2 Var[X]$
3. $Var[|X|] \neq Var[X]$
4. $E[X^2] \geq (E[X])^2$

Riemann Rearrangement Theorem: let $\{a_n\}$ be a sequence of number if 1. $\sum_{n=1}^{\infty} a_n$ converges (收敛). 2. $\sum_{n=1}^{\infty} |a_n| = \infty$. then for any $B \in \mathbb{R} \cup \{\infty\}$ there exists a rearrangement $\{b_n\}$ of $\{a_n\}$ such that $\sum_{n=1}^{\infty} b_n = B$.

Existence of Moments: if $E[|X^n|] < \infty$ then $E[X^n]$ exists

- if $E[|X|^{n+1}] < \infty$, then $E[|X|^n] < \infty$

Moment Generating Functions

$M_X(t) = E[e^{tX}]$, $E[X^n] = M_X^{(n)}(0)$

Ex $X \sim N(\mu, \sigma^2)$, $M_X(t) = e^{(\mu t + \frac{\sigma^2 t^2}{2})}$

$X \sim \text{Exp}(\lambda)$, $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$

Borel-Cantelli Lemma : if $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = 0$.

Borel Zero-one law : Let $A_1, A_2, \dots$, be a countably infinite sequence of events. Then, $P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = \begin{cases} 0, & \text{if } \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1, & \text{if } \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}$

if $X_1, \dots, X_n$ are "independent geometry r.v. then $Y = \min(X_1, \dots, X_n)$ is also a geometry r.v. with $p = 1-(1-p)^n$

if $L(\alpha) = C_k^N \cdot \alpha^k \cdot (1-\alpha)^{N-k}$, which α can maximize $L(\alpha)$ is solved by calculate $L'(\alpha) = 0$.

Probability of Winning if Switch Door

$P(E_{M,N}) = \frac{N-\underline{M}}{N} \frac{N-M-1}{N-2} + \frac{M}{N} \frac{N-M}{N-2}$ (Empty)

$H(X) = -\sum_{i=1}^{n} p_i \ln p_i = \sum_{i=1}^{n} p_i \cdot \ln \frac{1}{p_i} = \sum_{i=1}^{n} \ln (\frac{1}{p_i})^{p_i} = \ln (\prod_{i=1}^{n} (\frac{1}{p_i})^{p_i})$

and by weighted inequality of arithmetic and geometric means, we have

$\frac{\frac{1}{p_1} \cdot p_1 + \frac{1}{p_2} \cdot p_2 + \dots + \frac{1}{p_n} \cdot p_n}{\sum_{i=1}^{n} p_i} \geq \left[\prod_{i=1}^{n} (\frac{1}{p_i})^{p_i}\right]^{\frac{1}{\sum_{i=1}^{n} p_i}} \Rightarrow n \geq \prod_{i=1}^{n} (\frac{1}{p_i})^{p_i}$

$\because \ln$ is an increasing function $\therefore \ln n \geq \ln \prod_{i=1}^{n} (\frac{1}{p_i})^{p_i} = H(X)$

and we also know that equality holds if $(\frac{1}{p_1})^{p_1} = (\frac{1}{p_2})^{p_2} = \dots = (\frac{1}{p_n})^{p_n}$

$\Rightarrow p_1 = p_2 = \dots = p_n = \frac{1}{n}$. In conclusion, when the PMF $\{p_i\}_{i=1}^n$: $p(x) = \begin{cases} 1/n, & x = 1, 2, 3, \dots, n \\ 0, & \text{otherwise} \end{cases}$

$H(X)$ has maximum value $\ln n$

Continuous Random Variables

PDF of X : For $f_X(x)$, $P(X \in B) = \int_B f_X(x) dx$. check vaild :

1. $P(X \in \mathbb{R}) = 1 \Leftrightarrow \int_{-\infty}^{\infty} \underbrace{f_X(x)}_{\text{PDF}} dx$ 2. $P(X \in A) \geq 0 \ \forall A \Leftrightarrow \int_A f_X(x) dx \geq 0 \ \forall A$

3. $P(X \in \bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} P(X \in A_i) \Leftrightarrow \int_{\bigcup_{i \geq 1} A_i} f_X(x) dx = \sum_{i \geq 1} \int_{A_i} f_X(x) dx$

(don't need to check, $\because$ hold by the definition of integration)

CDF $\Leftrightarrow$ PDF : $F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x) dx$. $f_X(x) = F_X'(x)$

$\Rightarrow$ if $f_X(\cdot)$ is continuous at x_0, then $F_X'(x_0) = f_X(x_0)$

Property: 1. $f_X(x)$ could be ≥ 1 because it doesn't have meaning as a single point 2. $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f_X(x) dx$

$E[X] = \int_0^{\infty} (1 - F_X(t)) dt - \int_{-\infty}^{0} \underbrace{F_X(t)}_{\text{CDF}} dt$

$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$

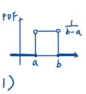
$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$

$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x) dx$

Special Continuous Random Variables

1. Continuous Uniform Random Variables ($X \sim \text{Unif}(a,b)$)

PDF of X : $f_X(x) = \frac{1}{b-a} (a < x < b)$, 0 (otherwise)



CDF of X: $F_X(X \leq t) = t$

- if we take CDF of X : $F_X(X)$ as r.v. then it's a kind of uniform r.v. : $F_X(X) \sim \text{Unif}(0,1)$

Inverse Transform Sampling (ITS) : given CDF: $F(t)$, generate r.v.

1. choose $U \sim \text{Unif}(0,1)$ 2. $X = F^{-1}(U)$, where $F^{-1}(u) := \inf\{z \mid : F(z) \geq u\}$

$P[F^{-1}(U) \leq x] \overset{F(x) \text{ is non-decreasing}}{=} P[F(F^{-1}(U)) \leq F(x)] = P[U \leq F(x)] \overset{\text{def. of unif}}{=} F(x)$

2. Standard Normal Random Variables ($X \sim \mathcal{N}(0,1)$)

PDF : $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $E[X] = 0$, $Var[X] = 1$

CDF : $\Phi(t) := P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, $\Phi(\infty) = 1$, $\Phi(0) = \frac{1}{2}$

From Standard Normal to Normal

$Y = aX + b$ CDF : $F_Y(t) = P(Y \leq t) = P(aX + b \leq t) = \begin{cases} P(X \leq \frac{t-b}{a}) & \text{if } a > 0 \quad F_X(\frac{t-b}{a}) \\ P(X \geq \frac{t-b}{a}) & \text{if } a < 0 \quad 1 - F_X(\frac{t-b}{a}) \end{cases}$

PDF : $\frac{d}{dt} F_Y(t) = \begin{cases} \frac{1}{a} f_X(\frac{t-b}{a}) & \text{if } a > 0 \\ \frac{1}{a}(-f_X(\frac{t-b}{a})) & \text{if } a < 0 \end{cases}$

Normal Random Variables : $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(\frac{-(x-\mu)}{2\sigma^2})}$ $\quad X \sim \mathcal{N}(\mu, \sigma^2)$

3. Exponential Random Variables ($X \sim \text{Exp}(\lambda)$)

PDF : $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Under a larger λ, X is more likely to be smaller

CDF : $t \leq 0$: $F_X(t) = 0$
$t > 0$: $F_X(t) = P(X \leq t) = \int_0^t f_X(x) dx = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$

$P(X > s+t \mid X > t) = \frac{P(X > s+t)}{P(X > t)} = \frac{\int_{s+t}^{\infty} f_X(x) dx}{\int_t^{\infty} f_X(x) dx} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$

$F_X(t) = P(\{w : \min(X_1(w), X_2(w)) \leq t\}) = 1 - P(X_1 > t) P(X_2 > t) = 1 - e^{-(\lambda_1 + \lambda_2) t}$

$X_1 \sim \text{Exp}(\lambda_1)$, $X_2 \sim \text{Exp}(\lambda_2)$, $X = \min(X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2)$

Binomial Expansion

$(x+y)^n = \sum_{i=0}^{n} C_i^n x^{(n-i)} y^i \qquad C_0^n + \dots + C_n^n = 2^n$

$e^x = \lim_{n \to \infty} (1 + \frac{x}{n})^n = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} \dots$

$\Gamma(n) = \int_0^{\infty} u^{n-1} e^{-u} du \quad \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \quad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^2}{90} \quad \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \quad \ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}$

Expected Value via CDF

X: continuous r.v. with CDF $F_X(t)$, $E[X] = \overset{A}{\int_0^\infty (1-F_X(t))\,dt} - \overset{B}{\int_0^\infty F_X(-t)\,dt}$

$^A\ \int_0^\infty \int_t^\infty f_X(x)\,dx\,dt = \int_0^\infty \int_0^x f_X(x)\,dt\,dx = \int_0^\infty x f_X(x)\,dx$ $\quad ^B\ \int_{-\infty}^0 \int_0^{-x} f_X(x)\,dt\,dx = \int_{-\infty}^0 -x f_X(x)\,dx$

Joint CDF

($0 \le F_{XY}(t,u) \le 1$. $t_1 \le t_2$, $u_1 \le u_2$ then $F_{XY}(t_1,u_1) \le F_{XY}(t_2,u_2)$)

X, Y be 2 r.v. on same Ω. $F_{XY}(t,u) = \underbrace{P(X \le t, Y \le u)}_{\text{Joint PMF}}$, $\forall t, u \in \mathbb{R}$

Marginal CDF

$F_X(t) = P(X \le t) = F_{XY}(t, \infty)$. $F_Y(t) = P(Y \le t) = F_{XY}(\infty, t)$

$P(t_1 < X \le t_2) = F_X(t_2) - F_X(t_1) = F_{XY}(t_2, \infty) - F_{XY}(t_1, \infty)$

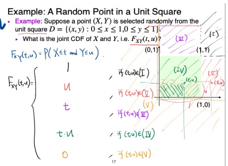
$P(t_1 < X \le t_2, u_1 < Y \le u_2) = F_{XY}(t_2,u_2) - F_{XY}(t_1,u_2) - F_{XY}(t_2,u_1) + F_{XY}(t_1,u_1)$

$P(t_1 < X < t_2, u_1 < Y \le u_2) = F_{XY}(t_2^-,u_2) - F_{XY}(t_1,u_2) - F_{XY}(t_2^-,u_1) + F_{XY}(t_1,u_1)$

Joint PMF

X, Y be 2 discrete r.v. on same Ω

$p_{XY}(x,y) = P(X=x, Y=y)$



Marginal PMF

$P(X=x) = P(X=x, Y \in S_Y) = \sum_{y \in S_Y} p_{XY}(x,y)$

$p_{XY}(x,y) = \begin{cases} \frac{1}{25}(x^2+y^2) & \text{if } x = 1,2,\ y = 0,1,2 \\ 0 & \text{otherwise} \end{cases}$

$P(X=x) = \begin{cases} \frac{3x^2}{25} + \frac{1}{5} & \text{if } x = 1 \text{ or } 2 \quad \left(\sum_{y=0,1,2} \frac{1}{25}(x^2+y^2)\right) \\ 0 & \text{otherwise} \end{cases}$

Example: Bernoulli and Poisson

- Example: Consider $X \sim$ Bernoulli(p), $Z_0 \sim$ Poisson$(\lambda_0 T)$ and $Z_1 \sim$ Poisson$(\lambda_1 T)$ (all are independent). Suppose $Y = Z_0$ if $X = 0$, and $Y = Z_1$ if $X = 1$.
  - What is the joint PMF of X and Y?
  - What is the marginal PMF of Y?

$Y(\omega) = \begin{cases} Z_0(\omega), & \text{if } X(\omega) = 0 \\ Z_1(\omega), & \text{if } X(\omega) = 1 \end{cases}$

Joint PMF of X and Y:

$p_{XY}(x,y) = P(X = x \text{ and } Y = y) = \begin{cases} \frac{e^{-\lambda_0 T}(\lambda_0 T)^y}{y!}(1-p) & \text{if } X=0 \text{ and } y \in \mathbb{N} \cup \{0\} \\ \frac{e^{-\lambda_1 T}(\lambda_1 T)^y}{y!} p, & \text{if } X=1 \text{ and } y \in \mathbb{N} \cup \{0\} \\ 0, & \text{otherwise} \end{cases}$

PMF of Y: $P(Y=y) = \sum_{x=0,1} p_{XY}(x,y)$

22

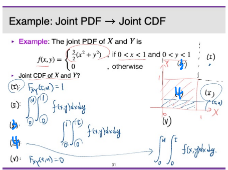
Joint PDF

X, Y be 2 continuous r.v. for B of $\mathbb{R}^2$

$P((X,Y) \in B) = \iint_B f_{XY}(x,y)\,dx\,dy$ $\quad P(X \le t) = \int_{-\infty}^{+\infty}\int_{-\infty}^t f_{XY}(x,y)\,dx\,dy$

Marginal PDF

$f_X(x) = \int_{-\infty}^\infty f_{XY}(x,y)\,dy$



Joint PDF → Marginal PDF ($f_X(x)$, $f_Y(y)$)

$f_X(x) = \int_{-\infty}^\infty f(x,y)\,dy = \begin{cases} \int_0^1 \frac{3}{2}(x^2+y^2)\,dy = \frac{3}{2}x^2 + \frac{1}{2} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

Joint PDF ($f_{XY}(x,y)$) → Marginal CDF ($F_X(t)$, $F_Y(u)$)

$F_X(t) = P(X \le t, Y \in \mathbb{R}) = \int_{-\infty}^\infty \int_{-\infty}^t f_{XY}(x,y)\,dx\,dy$

Partial Derivative of Joint CDF is Joint PDF

$f_{XY}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{XY}(x,y)$ ← Operate with (not unique)

X, Y be 2 continuous r.v. Assume partial derivatives of $F_{XY}(x,y)$ exist, one valid choice

Joint PDF: Interpret Density using limits

$f_{XY}(x,y) \equiv \lim_{\delta x, \delta y \to 0} \frac{P(x < X \le x+\delta x,\ y < Y \le y + \delta y)}{\delta x\,\delta y}$

$= \lim_{\delta y \to 0} \frac{1}{\delta y} \lim_{\delta x \to 0} \frac{F_{XY}(x+\delta x, y+\delta y) - F_{XY}(x, y+\delta y) - (F_{XY}(x+\delta x, y) - F_{XY}(x,y))}{\delta x} = \lim_{\delta y \to 0} \frac{1}{\delta y} \lim_{\delta x \to 0} \frac{\partial}{\partial x} F_{XY}(x, y+\delta y) - \frac{\partial}{\partial x} F_{XY}(x,y) =$

Independence

Independence $\equiv$ Joint CDF is the product of two marginal CDFs $F_{XY}(t,u) = F_X(t) \cdot F_Y(u)$ both discrete / continuous (Independence if and only if ↑)

If 2 r.v. X, Y are independent, ① $p_{XY}(x,y) = p_X(x) \cdot p_Y(y)$; ② PDF $f_{XY}(t,u) = f_X(t) \cdot f_Y(u)$ $\quad (E[XY] = \iint xy f_{XY}(t,u) = E[X]E[Y])$

① By independence of X, Y, $\underbrace{P(X \in A, Y \in B)}_{p_{XY}(x,y)} = \underbrace{P(X \in A)}_{p_X(x)} \cdot \underbrace{P(Y \in B)}_{p_Y(y)}$ $\quad$ ② $f_{XY}(t,u) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)\Big|_{x=t, y=u} = f_X(x)|_{x=t} \cdot f_Y(y)|_{y=u}$

Convolution Thm.

X, Y 2 independent discrete r.v. / continuous independent r.v. $Z = X+Y$, PMF $p_Z(z) = P(Z = z) = \sum_x p_X(x)\, p_Y(z-x)$; PDF $f_Z(z) = \int_{-\infty}^\infty f_1(x) f_2(z-x)\,dx$

MGF

r.v. X, $M_X(t) = E[e^{tX}]$, $t \in \mathbb{R}$ If exists $\delta > 0$ s.t. $M_X(t) < \infty$ for all $t \in (-\delta, \delta)$

$M_X(t) = \underbrace{\sum_x p_X(x) e^{tx}}_{\text{discrete}} = \underbrace{\int_{-\infty}^\infty f_X(x) e^{tx}\,dx}_{\text{continuous}}$

Example: Find MGF of Normal Random Variables

- Example: Let $Z \sim \mathcal{N}(\mu, \sigma^2)$
- Question: What is the MGF of Z, $M_Z(t)$ = ?

PDF of Z: $f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$

$M_Z(t) = E[e^{tZ}] = \int_{-\infty}^{+\infty} e^{tz} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) dz$

$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2 - (2\mu + 2\sigma^2 t)z + \mu^2}{2\sigma^2}\right) dz$

$= \int_{-\infty}^{+\infty} \boxed{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-(\mu+\sigma^2 t))^2}{2\sigma^2}\right)} \exp\left(\frac{\mu^2 - (\mu+\sigma^2 t)^2}{2\sigma^2}\right) dz = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$

PDF of $\mathcal{N}(\mu + \sigma^2 t, \sigma^2)$

24

Property 1 MGF uniqueness Thm. X_1, X_2. 2 r.v. (same distribution)

If $M_{X_1}(t) = M_{X_2}(t)$ for all t in $(-\alpha, \alpha)$, then $P(X_1 \le u) = P(X_2 \le u)$ for all $u \in \mathbb{R}$ (same CDF $F_{X_1}(u) = F_{X_2}(u)$)

Property 2 2 independent r.v., $M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

$= E[e^{tX_1} e^{tX_2}] = \sum_{x_1, x_2} p_{X_1 X_2}(x_1,x_2) e^{tx_1} e^{tx_2} = \sum_{x_1} p_{X_1}(x_1) e^{tx_1} + \ldots$

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ $\quad M_{X_1}(t) = e^{t\mu_1 + \frac{1}{2}\sigma_1^2 t^2}$ $\quad M_{X_1+X_2}(t) = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$

$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. $\quad M_{X_2}(t) = e^{t\mu_2 + \frac{1}{2}\sigma_2^2 t^2}$ $\quad (X_1+X_2) \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

$M_X(t) = \frac{1}{6}e^{-2t} + \frac{1}{6}e^{-t} + \frac{1}{4}e^{t} + \frac{1}{4}e^{2t}$

Property 3 X r.v., for every $n \in \mathbb{N}$, $E[X^n] = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$

$P(|X| \le 1) = \frac{5}{12}$ $\left(p_X(k) = \begin{cases} \frac{1}{6} & k = -2 \\ \frac{1}{3} & k = -1 \\ \frac{1}{4} & k = 1, 2 \end{cases} \right)$

$\frac{d^n}{dt^n} E[e^{tX}]\Big|_{t=0} = \sum_x p_X(x) \cdot \frac{d^n}{dt^n} e^{tx}\Big|_{t=0} = \sum_x p_X(x)\, x^n = E[X^n]$

Markov's Inequality

X nonnegative r.v. For any $t > 0$, $P(X \ge t) \le \frac{E[X]}{t}$

$Y(\omega) = \begin{cases} t & \text{if } X(\omega) \ge t \\ 0 & \text{otherwise} \end{cases}$ $\quad E[Y] = t \cdot P(X \ge t) \le E[X] \Rightarrow P(X \ge t) \le \frac{E[X]}{t}$

Chebyshev's Inequality

X r.v. mean $\mu < \infty$, variance $\sigma^2 < \infty$. For any $t > 0$ $P(|X - \mu| \ge t) \le \frac{\sigma^2}{t^2}$

$P(|X-\mu| \ge t) = P(\underbrace{|X-\mu|^2}_{Z} \ge t^2) \le \frac{E[Z]}{t^2} = \frac{E[|X-\mu|^2]}{t^2} = \frac{\sigma^2}{t^2}$ $\quad Var[\frac{1}{n}(X_1 + \ldots + X_n)] = \frac{1}{n^2}(\sigma_1^2 + \ldots + \sigma_n^2)$

$E[X_1 + \ldots + X_n] = \mu_1 + \ldots + \mu_n$; $E[\frac{1}{n}(X_1 + \ldots + X_n)] = \frac{1}{n}(\mu_1 + \ldots + \mu_n)$; $Var[X_1 + \ldots + X_n] = \sigma_1^2 + \ldots + \sigma_n^2$ $= E[((X_1+X_2) - (\mu_1+\mu_2))^2] = E[(X_1-\mu_1)^2] + E[(X_2-\mu_2)^2]$

Chebyshev's Inequality and Sample Mean

- Example: Tossing moon blocks $X_1 \cdots X_n$
  - Each toss X_i is Bernoulli with P(outcome is "Yes") $= p$
  - Each toss is independent from other tosses
  - Question: Can we say anything about the sample mean of n tosses $\frac{1}{n}(X_1 + \cdots + X_n)$?

Chebyshev's inequality: $P(|X-\mu| \ge t) \le \frac{\sigma^2}{t^2}$

$X \equiv \frac{1}{n}(X_1 + \cdots + X_n)$ $\quad \Rightarrow P(|\frac{1}{n}(X_1 + \cdots + X_n) - p| \ge t) \le \frac{p(1-p)}{nt^2}$

$\mu \equiv E[X] = p$

$\sigma^2 \equiv Var[X] = \frac{1}{n^2}(\underbrace{p(1-p) + p(1-p) + \cdots + p(1-p)}_{n \text{ times}}) = \frac{p(1-p)}{n}$

$X_1 \ldots X_n$ i.i.d. r.v. mean μ variance σ, $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$, for any $\varepsilon > 0$

$P(|\bar{X} - \mu| \ge \varepsilon) \le \frac{\sigma^2}{\varepsilon^2 n}$ $\quad S_n = (X_1 + \ldots + X_n)$ $\quad \lim_{n \to \infty} P(|\frac{S_n}{n} - \mu| \ge \varepsilon) = 0$

Chernoff Bound

X r.v. MGF $M_X(t)$ exists for all t in set S. For any $t > 0$, $t \in S$, $a \in \mathbb{R}$

$P(X \ge a) \le e^{-ta} \cdot M_X(t)$

$P(X \ge a) = P(\{\omega : X(\omega) \ge a\}) = P(\{\omega : e^{t \cdot X(\omega)} \ge e^{ta}\}) = P(\underbrace{e^{tX}}_{\text{non-negative}} \ge e^{ta}) \le \frac{E[e^{tX}]}{e^{ta}} = e^{-ta} \cdot M_X(t)$

Optimizing the Chernoff Bound

$P(X \ge a) \le e^{-\phi(a)}$, $\phi(a) = \max_{t > 0, t \in S}(ta - \ln M_X(t))$

$P(X \ge a) \le \min_{t > 0, t \in S} e^{-ta} \cdot M_X(t) = e^{-\max_{t>0, t \in S}(ta - \ln M_X(t))}$

$X \sim$ Bernoulli(p) $\quad P(X = k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \\ 0, & \text{otherwise} \end{cases}$ $\quad$ ① $M_X(t) = E[e^{tX}] = p \cdot e^{t \cdot 1} + (1-p)e^{t \cdot 0} = pe^t + (1-p)$, $\forall t \in \mathbb{R}$

② Chernoff bound for X: $P(X \ge a) \le \underbrace{e^{-ta}(pe^t + (1-p))}_{f(t)}$ $\quad$ ③ Optimize $\frac{d}{dt} f(t) = \frac{d}{dt}(pe^{t(1-a)} + (1-p)e^{-ta})$

$= \underbrace{p(1-a)e^{t(1-a)} - a(1-p)e^{-ta}}_{(*)}$. $\quad (*) = 0 \Leftrightarrow t = \ln \frac{a(1-p)}{p(1-a)}$

Hoeffding's Inequality (Bernoulli)

$X_1 \ldots X_n$ i.i.d. Bernoulli(p) r.v. $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ For any $\varepsilon > 0$,

$P(|\bar{X} - p| \ge \varepsilon) \le 2e^{(-n\varepsilon^2)}$ $\quad \Big| \quad E[e^{tZ}] \le e^{\left(\frac{t^2(b-a)^2}{8}\right)}$

Proof of Hoeffding's Inequality (Positive Part)

$P(\bar{X} - p \ge \varepsilon) \le \exp(-2n\varepsilon^2)$

- [Hint] Chernoff bound:

$P(\bar{X} - p \ge \varepsilon) \le P(\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - p \ge \varepsilon)$

$= P((X_1 + X_2 + \cdots + X_n) - np \ge n\varepsilon)$

$= P(\underbrace{(X_1-p) + (X_2-p) + \cdots + (X_n-p)}_{Z} \ge n\varepsilon) \le M_Z(t) \cdot e^{-tn\varepsilon}$ (by Chernoff bound)

$M_Z(t) = E[e^{t((X_1-p)+(X_2-p)+\cdots+(X_n-p))}]$ $= E[e^{t(X_1-p)}]^n e^{-tn\varepsilon}$

by independence $= E[e^{t(X_1-p)}] \cdot E[e^{t(X_2-p)}] \cdots E[e^{t(X_n-p)}]$ $= e^{\frac{t^2 n}{8}} \cdot e^{-tn\varepsilon}$

$= (E[e^{t(X_1-p)}])^n$ $\quad t = 4\varepsilon$

11

Hoeffding's Lemma

Z r.v. $E[Z] = 0$, $Z \in [a,b]$ with $p = 1$. For any $t > 0$

Hoeffding's Inequality

$X_1 \ldots X_n$ i.i.d. r.v. $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$, $\mu := E[X_1]$

For any $\varepsilon > 0$, $P(|\bar{X} - \mu| \ge \varepsilon) \le 2e^{\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)}$ $\quad f_U(u) = \begin{cases} 1, & 0 < u < 1 \\ 0, & \text{otherwise} \end{cases}$

Monte-Carlo Method

$U \sim$ Unif$(0,1)$, $U_1 \ldots U_k$ i.i.d. r.v. ① Rewrite $\int_0^1 e^{-x^3} dx = E[e^{-U^3}]$ ② $E[e^{-U^3}] \approx \frac{1}{k}\sum_{i=1}^k e^{-U_i^3}$

WLLN (Khinchin's Law)

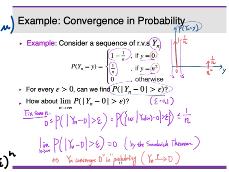
without assumption of finite σ^2

$X_1 \ldots X_n$ i.i.d. r.v. mean μ, $S_n = X_1 + \ldots + X_n$ For every $\varepsilon > 0$, $P(|\frac{S_n}{n} - \mu| \ge \varepsilon) \to 0$ as $n \to \infty$

Convergence to a Scalar in Probability ($Y_n \xrightarrow{P} a$)

$Y_1 \ldots Y_n$ r.v. $a \in \mathbb{R}$ For every $\varepsilon > 0$, $\lim_{n \to \infty} P(\{\omega : |Y_n - a| > \varepsilon\}) = 0$

More Formally $\lim_{n \to \infty} P(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \ge \varepsilon\}) = 0$ $(\frac{S_n}{n} \xrightarrow{P} \mu)$



Convergence in Probability ($Y_n \xrightarrow{P} Y$) to a R.V.

$\lim_{n \to \infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$

Example: Convergence in Probability

- Example: Consider a random variable Y ($Y \sim$ Bernoulli$(p = \frac{1}{2})$)

$P(Y = y) = \begin{cases} 1/2, & \text{if } y = 0 \\ 1/2, & \text{if } y = 1 \\ 0, & \text{otherwise} \end{cases}$ $\quad Y_1 = (1 + \frac{1}{1})Y$, $Y_2 = (1 + \frac{1}{2})Y$

- For every $n \in \mathbb{N}$, define $Y_n = (1 + \frac{1}{n})Y$ $\quad Y_1(\omega) = (1+\frac{1}{1})Y(\omega)$, for all $\omega \in \Omega$
- Do we have $Y_n \xrightarrow{P} Y$ (i.e. $\boxed{\lim_{n \to \infty} P(|Y_n - Y| > \varepsilon) = 0}$?

$n = 1$: $P(\{\omega : |Y_1(\omega) - Y(\omega)| > \varepsilon\}) = P(\{\omega : |Y(\omega)| > \varepsilon\}) = \begin{cases} 0, & \text{if } \varepsilon \ge 1 \\ \frac{1}{2}, & \text{if } 0 < \varepsilon < 1 \end{cases}$

n: $P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = P(\{\omega : |(1+\frac{1}{n})Y(\omega) - Y(\omega)| > \varepsilon\}) = \begin{cases} 0, & \text{if } n\varepsilon \ge 1 \\ \frac{1}{2}, & \text{if } 0 < n\varepsilon < 1 \end{cases}$

Therefore, we have $\lim_{n \to \infty} P(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$ (i.e., $Y_n \xrightarrow{P} Y$)

6

$X_n \sim$ Unif$(0,1)$

$Y_n = \min\{X_1, X_2, \ldots X_n\}$

$P(\{\omega : Y_n(\omega) > \varepsilon\})$

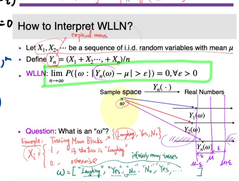
$= P(\{\omega : X_1(\omega) > \varepsilon\}) \times \ldots = (1-\varepsilon)^n$

$\Rightarrow \lim_{n \to \infty} P(\{\omega : Y_n(\omega) > \varepsilon\}) = 0$ $\quad |Y_n(\omega) - 0| > \varepsilon$

$P(\{\omega : \lim_{n \to \infty} Y_n(\omega) > \varepsilon\})$ $\quad (1-\varepsilon)^m$

$< P(\{\omega : Y_m(\omega) > \varepsilon\})$

$\Rightarrow P(\{\omega : \lim_{n \to \infty} Y_n(\omega) > \varepsilon\}) = 0$



Distribution	Moment-generating function $M_X(t)$
Degenerate δ_a	e^{ta}
Bernoulli $P(X=1) = p$	$1 - p + pe^t$
Geometric $(1-p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t}$ $\forall t < -\ln(1-p)$
Binomial $B(n,p)$	$(1 - p + pe^t)^n$
Negative Binomial $NB(r,p)$	$\frac{(1-p)^r}{(1-pe^t)^r}$
Poisson $Pois(\lambda)$	$e^{\lambda(e^t - 1)}$
Uniform (continuous) $U(a,b)$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Uniform (discrete) $DU(a,b)$	$\frac{e^{at} - e^{(b+1)t}}{(b-a+1)(1-e^t)}$
Laplace $L(\mu, b)$	$\frac{e^{t\mu}}{1-b^2t^2}$, $\lvert t \rvert < 1/b$
Normal $N(\mu, \sigma^2)$	$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

Example: Moments of Exp(λ)

- Example: Suppose $X \sim$ Exp(λ) (rate)
  - What is the MGF of X?
  - Use MGF to verify that $E[X] = \frac{1}{\lambda}$ and $Var[X] = \frac{1}{\lambda^2}$?

$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$

$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\,dx = \int_0^\infty \lambda \cdot e^{(t-\lambda)x}\,dx = \frac{\lambda}{t-\lambda} e^{(t-\lambda)x}\Big|_0^\infty = \frac{\lambda}{\lambda - t}$ (if $t < \lambda$)

$E[X] = \frac{d}{dt} M_X(t)\Big|_{t=0} = \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} = \frac{1}{\lambda}$ $\quad (\lambda(\lambda-t)^{-2})$

$E[X^2] = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$

$Var[X] = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$

10

$M_X(t) = \frac{1}{2} + \frac{1}{2}e^t$, $X \sim$ Bernoulli$(\frac{1}{2})$

$M_Z(t) = e^{2t^2 - t}$, $Z \sim \mathcal{N}(\mu = -1, \sigma^2 = 4)$ (unique solution)

Convergence to a R.V. almost surely ($Y_n \xrightarrow{a.s.} Y$) or ($Y_n \to Y$ w.p. 1) $Y_1 \ldots Y_n$ r.v. $P(\{\omega : \lim_{n\to\infty} |Y_n(\omega) - Y(\omega)| = 0\}) = 1$

※ Almost-Sure Convergence ⇒ Convergence in Probability ($\cancel{\Leftarrow}$) $X_1 = \mathbb{1}\{x \in [0,1]\}$ $X_1(\omega) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$ $X_2 = \mathbb{1}\{x \in [0,\frac{1}{2}]\}$ ⋮

If $X_n \xrightarrow{a.s.} X$, then for any $\varepsilon > 0$, $P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ infinitely often$) = 0$ —①

For each $n \in \mathbb{N}$, define event $A_n = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$, ① can be $P(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n) = 0$ —② | Hence we have ③ $= 0$, $X_k \xrightarrow{P} X$.

$0 = P(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n) = P(\lim_{k\to\infty} \bigcap_{m=1}^{k}\bigcup_{n=m}^{\infty} A_n) = \lim_{k\to\infty} P(\bigcap_{m=1}^{k}\bigcup_{n=m}^{\infty} A_n) = \lim_{k\to\infty} P(\bigcup_{n=k}^{\infty} A_n) \geq \lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} P[\{\omega : |X_k(\omega) - X(\omega)| > \varepsilon\}]$ —③

SLLN $X_1 \ldots X_n$ i.i.d. r.v. mean μ, $S_n = X_1 + \ldots + X_n$ $P(\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\}) = 1$ $(\frac{S_n}{n} \xrightarrow{a.s.} \mu)$

Equivalent Definition of Almost-Sure Convergence $\forall \varepsilon > 0$, $P(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} \underbrace{\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}}_{A_n}) = 0$ (ω appears in infinitely many A_n)

A bound For 4th Moment $X_1 \ldots X_n$ iid r.v., $E[X^4] < \infty$, $S_n = (X_1 + \ldots + X_n)$, exist K s.t. $E[(S_n - n\mu)^4] \leq Kn^2$ $E[(\frac{S_n}{n} - \mu)^4] = \frac{Kn^2}{n^4}$

※ SLLN says about $S_n(\omega) = n\mu + \underbrace{f(n)}_{o(n)}$ for large n (gap) $(\lim_{n\to\infty} \frac{f(n)}{n} = 0)$

Put Everything Together: Proof of SLLN $P(\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\}) = 1$

- SLLN: $P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{\omega : \left|\frac{S_n(\omega)}{n} - \mu\right| > \varepsilon\}\right) = 0, \forall \varepsilon > 0$ 4-th central moment of $\frac{S_n}{n}$
- Proof:

Step 1: $P\left(\left\{\left|\frac{S_n}{n} - \mu\right| \geq n^{-\gamma}\right\}\right) = P\left(\left|\frac{S_n}{n} - \mu\right|^4 \geq n^{-4\gamma}\right) \leq \frac{E\left[\left|\frac{S_n}{n} - \mu\right|^4\right]}{n^{-4\gamma}}$ true mean, empirical mean, A_n, Z

Step 2: $\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} \frac{K}{n^{2-4\gamma}} < \infty$ if $\gamma \in (0, \frac{1}{4})$ $\leq \frac{\frac{K}{n^2}}{n^{-4\gamma}} = \frac{K}{n^{2-4\gamma}}$

Step 3: $\underbrace{P(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n) = 0}_{\text{by Borel-Cantelli Lemma}} \Longleftrightarrow P(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq n^{-\gamma}\}) = 0$

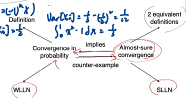
Central Limit Theorem $X_1 \ldots X_n$ i.i.d. R.V. $S_n = (X_1 + \ldots + X_n)$ μ, variance σ^2

$\lim_{n\to\infty} P\left(\frac{S_n - \overset{E[S_n]}{n\mu}}{\underset{\sqrt{Var[S_n]}}{\sigma\sqrt{n}}} \leq z\right) = \underbrace{\Phi(z)}_{\text{CDF of standard normal}}$, $\forall z \in \mathbb{R}$

Convergence in Distribution $(Y_n \Rightarrow Y)$ $(F_n \Rightarrow F)$ (weak convergence) $\lim_{n\to\infty} \underbrace{F_n(t)}_{\text{CDF}} = F(t)$ CLT: $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1)$

※ Convergence in Distribution $\cancel{\Rightarrow}$ Convergence in Probability, Almost-Sure Convergence $(X \sim N(0,1), X_n = (-1)^n X)$

Approx via CLT $X_1 \ldots X_{20}$ continuous uniform $(0,1)$ r.v.s, $P(\sum_{i=1}^{20} X_i \leq 8) \underset{CLT}{\approx} \Phi\left(\frac{8-10}{\sqrt{\frac{20}{12}}}\right)$ $E[X_i] = \frac{1}{2}$ $Var[X_i] = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$ $\int_0^1 x^2 \cdot 1\, dx = \frac{1}{3}$



Converge in the mean square $\lim_{n\to\infty} E[(X_n - c)^2] = 0$

Hoeffding's Inequality (Negative Part)

1. Start with:
$$P(\overline{X} \leq p - \epsilon) = P(p - \overline{X} \geq \epsilon).$$
2. For any $t > 0$:
$$P(\overline{X} \leq p - \epsilon) = P(e^{-t(\overline{X}-p)} \geq e^{t\epsilon}).$$
3. By Markov's inequality:
$$P(\overline{X} \leq p - \epsilon) \leq \frac{\mathbb{E}[e^{-t(\overline{X}-p)}]}{e^{t\epsilon}}.$$
4. Since $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:
$$\mathbb{E}[e^{-t(\overline{X}-p)}] = e^{tp}\mathbb{E}\left[e^{-\frac{t}{n}X_1}\right]^n.$$
5. For $X_1 \sim$ Bernoulli(p):
$$\mathbb{E}[e^{-\frac{t}{n}X_1}] = pe^{-\frac{t}{n}} + (1-p).$$
$\frac{t}{n}$
6. Thus:
$$\mathbb{E}[e^{-t(\overline{X}-p)}] = e^{tp}[pe^{-\frac{t}{n}} + (1-p)]^n.$$
7. Substitute back:
$$P(\overline{X} \leq p - \epsilon) \leq \frac{e^{tp}(pe^{-\frac{t}{n}} + (1-p))^n}{e^{t\epsilon}}.$$
8. Rewrite the exponent:
$$P(\overline{X} \leq p - \epsilon) \leq \exp(tp - t\epsilon + n\ln[pe^{-\frac{t}{n}} + (1-p)]).$$
9. Define:
$$f(t) := tp - t\epsilon + n\ln[(1-p) + pe^{-\frac{t}{n}}].$$
10. To minimize w.r.t. t:
$$f'(t) = p - \epsilon + n\frac{p(-\frac{1}{n})e^{-\frac{t}{n}}}{(1-p) + pe^{-\frac{t}{n}}} = p - \epsilon - \frac{pe^{-\frac{t}{n}}}{(1-p) + pe^{-\frac{t}{n}}}.$$
11. Set $f'(t) = 0$:
$$p - \epsilon = \frac{pe^{-\frac{t}{n}}}{(1-p) + pe^{-\frac{t}{n}}}.$$
12. Solve for $e^{-\frac{t}{n}}$:
$$(p-\epsilon)((1-p) + pe^{-\frac{t}{n}}) = pe^{-\frac{t}{n}}$$
$$(p-\epsilon)(1-p) + (p-\epsilon)pe^{-\frac{t}{n}} = pe^{-\frac{t}{n}}$$
$$(p-\epsilon)(1-p) = pe^{-\frac{t}{n}} - (p-\epsilon)pe^{-\frac{t}{n}}$$
$$(p-\epsilon)(1-p) = pe^{-\frac{t}{n}}[1 - (p-\epsilon)]$$
$$e^{-\frac{t}{n}} = \frac{(p-\epsilon)(1-p)}{p[1-(p-\epsilon)]} = \frac{(p-\epsilon)(1-p)}{p(\epsilon - (p-p))} = \frac{(p-\epsilon)(1-p)}{p(\epsilon - (0))} = \frac{(p-\epsilon)(1-p)}{p\epsilon}.$$
13. Substitute back into $f(t)$ with optimal t and simplify (using standard bounds):
$$P(\overline{X} \leq p - \epsilon) \leq \exp(-2n\epsilon^2).$$

Final:
$$P(\overline{X} \leq p - \epsilon) \leq e^{-2n\epsilon^2}.$$

X, Y are independent ⇒ $E[XY] = E[X]E[Y]$ ($\cancel{\Leftarrow}$) $(P(X=1) = P(X=-1) = P(X=0) = \frac{1}{3})$ $Y = |X|$

Cauchy-Schwarz Inequality

$E[X^2] \cdot E[Y^2] \geq (E[XY])^2$

$E[(tX+Y)^2] = E[X^2]t^2 + 2E[XY]t + E[Y^2] \geq 0$

$\Rightarrow (2E[XY])^2 - 4E[X^2]E[Y^2] \leq 0$

$\sum_{\text{all } x} \frac{P_{XY}(x,y)}{P_Y(y)}$ $\sum_x P_{X|Y}(x|y) = \frac{P_Y(y)}{P_Y(y)} = 1$

Conditional PMF joint PMF $P_{XY}(x,y)$ When $P(Y=y) > 0$ the conditional PMF of X given $Y=y$ is $P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)}$ $= P(X=x|Y=y)$

Conditional Expectation $E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y) = \sum_x x P_{X|Y}(x|y)$

Law of Iterated Expectation (LIE) $E[\overset{\text{over } X}{E[X|Y]}] = E[X]$ over Y — both discrete, continuous

Conditional PDF $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$

$f(x,y) = \begin{cases} 2, & 0 < y < x < 1 \\ 0, & \text{otherwise} \end{cases}$ $f_{X|Y}(x|y) = \begin{cases} \frac{2}{2(1-y)} & \text{if } (x,y) \in \triangle \\ 0 & \text{otherwise} \end{cases}$

$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_y^1 2\,dx = 2(1-y)$ if $y \in (0,1)$, 0 if $y \geq 1$ or $y \leq 0$

Example: Conditioning and Sum of Poisson

Let N_1, N_2 be the # of bits transmitted by AP₁, AP₂ in a time interval T
- N_1 and N_2 are Poisson with rates λ_1 and λ_2, respectively.
- Moreover, N_1 and N_2 are independent
- Define $M = N_1 + N_2$
- Question: Conditional PMF $p_{N_1|M}(n|m) = ?$

~Binomial$(m, \frac{\lambda_1}{\lambda_1+\lambda_2})$

$N_1 \sim$ Poisson(λ_1, T), $N_2 \sim$ Poisson(λ_2, T) let $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$

We already know $M \sim$ Poisson$(\lambda_1+\lambda_2, T)$ since N_1, N_2 are independent

$P_{N_1|M}(n|m) = \frac{P_{N_1,M}(n,m)}{P_M(m)} = \frac{P_{N_1,N_2}(n,m-n)}{P_M(m)} = \frac{P_{N_1}(n)P_{N_2}(m-n)}{P_M(m)}$

By plugging the Poisson PMFs, we have

$P_{N_1|M}(n|m) = \frac{\frac{e^{-\lambda_1 T}(\lambda_1 T)^n}{n!} \cdot \frac{e^{-\lambda_2 T}(\lambda_2 T)^{m-n}}{(m-n)!}}{\frac{e^{-(\lambda_1+\lambda_2)T}((\lambda_1+\lambda_2)T)^m}{m!}} = \frac{m!}{n!(m-n)!}\frac{(\lambda_1)^n(\lambda_2)^{m-n}}{(\lambda_1+\lambda_2)^m}$ $C_n^m p^n(1-p)^{m-n}$

2 Independent Normal: Matrix Form

- Joint PDF of 2 Independent Normal:
$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right]$$

$\Sigma = \begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix}\frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2}\end{bmatrix}$ $[x_1-\mu_1 \;\; x_2-\mu_2]\begin{bmatrix}\frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2}\end{bmatrix}\begin{bmatrix}x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}$

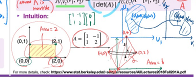
$x = \begin{bmatrix}x_1 \\ x_2\end{bmatrix}, \mu = \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}$ $\det(\Sigma) = \sigma_1^2\sigma_2^2$

- Joint PDF of 2 Independent Normal:
$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{|\det(\Sigma)|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

Linear Transformation of 2 Random Variables

$V_1(\omega) = a \cdot U_1(\omega) + b \cdot U_2(\omega)$, for all $\omega \in \Omega$

- Theorem: Let U_1, U_2, V_1, V_2 be random variables that satisfy $V_1 = aU_1 + bU_2$ and $V_2 = cU_1 + dU_2$. Define the matrix $A = \begin{bmatrix}a & b \\ c & d\end{bmatrix}$. Then, we have (assume A is invertible)
$$f_{V_1V_2}(v_1,v_2) = \frac{1}{|\det(A)|}f_{U_1U_2}(A^{-1}[v_1,v_2]^T)$$
$\begin{bmatrix}V_1 \\ V_2\end{bmatrix} = \begin{bmatrix}a & b \\ c & d\end{bmatrix}\begin{bmatrix}U_1 \\ U_2\end{bmatrix}$
- Intuition:



For more details, check: https://www.stat.berkeley.edu/~aditya/resources/AllLectures2018Fall201A.pdf

Bivariate Normal and Linear Transformation

$X_1 = Z$, $X_2 = \rho Z + \sqrt{1-\rho^2}W$ $f_{X_1X_2}(x_1,x_2) = \frac{1}{|\det(A)|}f_{ZW}(A^{-1}[x_1,x_2]^T)$

$\begin{bmatrix}X_1 \\ X_2\end{bmatrix} = \underbrace{\begin{bmatrix}1 & 0 \\ \rho & \sqrt{1-\rho^2}\end{bmatrix}}_{A}\begin{bmatrix}Z \\ W\end{bmatrix}$ $A^{-1} = \frac{1}{\det(A)}\begin{bmatrix}\sqrt{1-\rho^2} & 0 \\ -\rho & 1\end{bmatrix}$ $\det(A) = \sqrt{1-\rho^2}$ $f_{ZW} = \frac{1}{2\pi}\exp(-\frac{1}{2}(z^2+w^2))$

$f_{X_1X_2}(x_1,x_2) = \frac{1}{\underset{\sqrt{1-\rho^2}}{|\det(A)|}} \cdot \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(x_1^2 + \left(\frac{-\rho x_1}{\sqrt{1-\rho^2}} + \frac{x_2}{\sqrt{1-\rho^2}}\right)^2\right)\right)$

$x_1^2 + \frac{\rho^2 x_1^2}{1-\rho^2} + \frac{-2\rho x_1 x_2}{1-\rho^2} + \frac{x_2^2}{1-\rho^2}$ $= \frac{x_1^2}{1-\rho^2}$

$f_{X_1X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{(1-\rho)^2}}\exp\left(-\frac{x_1^2 - 2\rho x_1x_2 + x_2^2}{2(1-\rho^2)}\right)$

Bivariate Normal R.V.s (Formally)

- Bivariate Normal: X_1 and X_2 are said to be bivariate normal random variables if the joint PDF of X_1, X_2 is
$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)}\right]$$
The joint PDF can be written in matrix form as
$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{|\det(\Sigma)|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$
where covariance matrix $\Sigma = \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$, $x = \begin{bmatrix}x_1 \\ x_2\end{bmatrix}$, $\mu = \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}$ mean vector
- Notation for bivariate normal: $\begin{bmatrix}X_1 \\ X_2\end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$

properties
1. Marginal: $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$
2. $X_2 | X_1 = x_1 \sim N(\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2)$
3. If $\rho = 0$, X_1, X_2 are independent

※ X_1, X_2 Normal $\cancel{\Rightarrow}$ X_1, X_2 Bivariate Normal

- Covariance: Let X, Y be two random variables. Then, the covariance of X and Y is defined as
$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$
- $\text{Cov}(X,X) =$ $Var[X]$
- $\text{Cov}(X,Y) = 0$: X, Y are said to be uncorrelated
- $\text{Cov}(X,Y) > 0$: X, Y are said to be positively correlated
- $\text{Cov}(X,Y) < 0$: X, Y are said to be "negatively" correlated

Independent ⇒ $Cov(X,Y) = 0$ ($\cancel{\Leftarrow}$) $(Var[aX+bY] = a^2Var[X] + b^2Var[Y] + 2abCov(X,Y))$

3-1 $Z \sim$ Bernoulli(p), $Y \sim$ Poisson$(\lambda, 1)$ independent X: $Z=0, X=0$; $Z=1, X=Y$ (a) PMF of X (b) $\tilde{X} = (1-I) \cdot Y$, $I \sim$ Bernoulli$(1-p)$. $X, \tilde{X}$ same PMF (c) Show $E[UV] = E[U] \cdot E[V]$ (all $< \infty$)

3-2 (a) $X, Y \sim$ Possion$(\lambda=2, T=1)$. use MGF test if $3X+4Y$ is also poisson r.v. (b) $P_Y(k) = \begin{cases}\frac{6}{\pi^2 k^2} & k = \mathbb{N} \\ 0 & \text{otherwise}\end{cases}$ show MGF D.N.E.

3-3 $F_{XY}(t,u) = \begin{cases}1 - e^{-t} - e^{-u} + e^{-(t+u+\theta tu)} & \text{if } t>0, u>0 \\ 0 & \text{otherwise}\end{cases}$ $\theta \in [0,1]$ (a) Find marginal CDF of X, Y. $\theta = ?$ independent (b) Find joint PDF of X, Y

3-5 $U_1 \ldots U_n$ i.i.d. $[0,1]$. $S_n = \sum_{i=1}^{n} X_i$, N: smallest integer s.t. $\sum_{i=1}^{N} U_i > 1$, $N \in$ r.v. (a) Show $E[N] = e$ (b) How to apply Monte-Carlo

Problem 1

(a) For $x = 0$, $P(X = x) = P(Z = 0) + P(Z = 1) \times P(Y = 0) = (1-p) + pe^{-\lambda}$
For $x = 1, 2, 3, \ldots$, $P(X = x) = P(Z = 1) \times P(Y = x) = p \times \frac{e^{-\lambda}\lambda^x}{x!}$

Therefore, $P(X = x) = \begin{cases} 1 - p + pe^{-\lambda} & x = 0 \\ p \times \frac{e^{-\lambda}\lambda^x}{x!} & x \in \mathbb{N} \\ 0 & otherwise \end{cases}$

(b) For $x = 0$, $P(\tilde{X} = x) = P(1 - I = 0) + P(1 - I = 1) \times P(Y = 0) = P(I = 1) + P(I = 0) \times P(Y = 0) = (1-p) + pe^{-\lambda}$
For $x = 1, 2, 3, \ldots$, $P(\tilde{X} = x) = P(1 - I = 1) \times P(Y = x) = P(I = 0) \times P(Y = x) = p \times \frac{e^{-\lambda}\lambda^x}{x!}$

Therefore, $P(\tilde{X} = x) = \begin{cases} 1 - p + pe^{-\lambda} & x = 0 \\ p \times \frac{e^{-\lambda}\lambda^x}{x!} & x \in \mathbb{N} \\ 0 & otherwise \end{cases}$, which is the same as $P(X = x)$.

(c) Let $p_U(u) = P(U = u)$, $p_V(v) = P(V = v)$. Since U and V are independent, $p_{UV}(u,v) = p_U(u) \times p_V(v)$. Therefore, $E[UV] = \sum_{u\in S_U}\sum_{v\in S_V} uv \times p_{UV}(u,v) = \sum_{u\in S_U}\sum_{v\in S_V} uv \times p_U(u) \times p_V(v) = \sum_{u\in S_U} u \times p_U(u) \times \sum_{v\in S_V} v \times p_V(v) = E[U] \times E[V]$.

Problem 2

(a) The MGF of a Poisson random variable is $e^{\lambda(e^t-1)}$, so we can know $M_{3X}(t) = M_X(3t) = e^{\lambda(e^{3t}-1)}$, $M_{4Y}(t) = M_Y(4t) = e^{\lambda(e^{4t}-1)}$. We also know that $M_{3X+4Y}(t) = M_{3X}(t) \times M_{4Y}(t) = e^{\lambda(e^{3t}-1)} \times e^{\lambda(e^{4t}-1)} = e^{\lambda(e^{3t}+e^{4t}-2)}$, which is not the MGF of a Poisson random variable. Therefore, $3X + 4Y$ is not a Poisson random variable.

(b) $M_Y(t) = \sum_{k=1}^{\infty} e^{tk} \cdot \frac{6}{\pi^2k^2} = \frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{e^{tk}}{k^2}$. For $t > 0$, the term $\frac{e^{tk}}{k^2}$ grows exponentially, which makes the series $\sum_{k=1}^{\infty}\frac{e^{tk}}{k^2}$ diverges. Therefore, $M_Y(t)$ does not exist since it is not finite for any $t > 0$.

Problem 3

(a) Marginal CDF of X: $F_X(t) = \lim_{u\to\infty} F_{XY}(t,u) = \lim_{u\to\infty}(1 - e^{-t} - e^{-u} + e^{-(t+u+\theta tu)}) = 1 - e^{-t}$
Marginal CDF of Y: $F_Y(t) = \lim_{t\to\infty} F_{XY}(t,u) = \lim_{t\to\infty}(1 - e^{-t} - e^{-u} + e^{-(t+u+\theta tu)}) = 1 - e^{-u}$
If X and Y are independent, $F_{XY}(t,u) = F_X(t) \times F_Y(u) = (1 - e^{-t}) \times (1 - e^{-u}) = 1 - e^{-t} - e^{-u} + e^{-(t+u)}$, which means $e^{-(t+u+\theta tu)} = e^{-(t+u)}$. Therefore, X and Y are independent only if $\theta = 0$.

(b) $\frac{\partial F_{X,Y}(t,u)}{\partial t} = \frac{\partial}{\partial t}(1 - e^{-t} - e^{-u} + e^{-(t+u+\theta tu)}) = e^{-t} + (1 + \theta u)e^{-(t+u+\theta tu)}$
Joint PDF of X and Y: $f_{X,Y}(t,u) = \frac{\partial}{\partial u}(e^{-t} + (1 + \theta u)e^{-(t+u+\theta tu)}) = \theta e^{-(t+u+\theta tu)} + (1 + \theta u)(1 + \theta t)e^{-(t+u+\theta tu)} = [1 + \theta(1 + u + t) + \theta^2 ut]e^{-(t+u+\theta tu)}$

Problem 4

(a) Since $R(T_a) > R(T_b)$, we can know that each labeler votes for T_a with probability $p = \sigma(R(T_a) - R(T_b)) > 0.5$. We can define the indicator random variable X_i for the i-th labeler, where $X_i = 1$ if the labeler votes for T_a and $X_i = 0$ if the labeler votes for T_b. Therefore, $X_i \sim Bernoulli(p)$.
Define the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the event E that T_a gets more votes than T_b is equivalent to $E = \{\bar{X} > \frac{1}{2}\}$. Using Hoeffding's inequality for bounded independent random variables, $P(|\bar{X} - \mu| \geq \epsilon) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}$, where $a = 0$ and $b = 1$ in this case, so $P(E) = P(\bar{X} > \frac{1}{2}) = 1 - P(\bar{X} \leq \frac{1}{2}) = 1 - P(\bar{X} - \mu \leq -(\mu - 0.5)) \geq 1 - e^{-2n(\mu-0.5)^2}$. Therefore, $P(E) \geq 1 - e^{-2n(p-\frac{1}{2})^2}$.

(b)
1. We should ensure that $P(E) = P(\bar{X} > \frac{1}{2}) \geq 1 - \delta$, which means $P(\bar{X} \leq \frac{1}{2}) \leq \delta$. Applying Chebyshev's inequality, $P(\bar{X} \leq \frac{1}{2}) = P(|\bar{X} - \mu| \geq \mu - \frac{1}{2}) \leq \frac{p(1-p)}{n(\mu-\frac{1}{2})^2}$. Therefore, we require $\frac{p(1-p)}{n(\mu-\frac{1}{2})^2} \leq \delta$, which means $n \geq \frac{p(1-p)}{\delta(p-\frac{1}{2})^2}$.
2. $P(E) \geq 1 - e^{-2n(p-\frac{1}{2})^2} \geq 1 - \delta$. This requires $e^{-2n(p-\frac{1}{2})^2} \leq \delta$, which can be simplified to $2n(p - \frac{1}{2})^2 \geq \ln(\frac{1}{\delta})$. Therefore, $n \geq \frac{1}{2(p-\frac{1}{2})^2}\ln(\frac{1}{\delta})$.

(c) We need to find an upper bound on $P(S > 90)$, where $S = \sum_{j=1}^{100} I_j$. For each I_j, $\mathbb{E}[e^{tI_j}] = (1 - p_j)e^{t\cdot 0} + p_je^t = 1 + p_j(e^t - 1)$. Since I_j are independent, for any $t > 0$, we have $P(S \geq a) \leq e^{-ta} \cdot M_X(t) = e^{-ta}\prod_{j=2}^{100}(1 + p_j(e^t - 1))$, where $a = 91$ since $S > 90$. Using the inequality $1 + x \leq e^x$, we can simplify the expression to $e^{-ta}\prod_{j=2}^{100} e^{p_j(e^t-1)} = e^{-ta} \cdot e^{\sum_{j=2}^{100} p_j(e^t-1)}$. Now we have $P(S \geq a) \leq e^{-ta+\mu(e^t-1)}$, where $\mu = \sum_{j=2}^{100} p_j$. To find the tightest bound, set $\phi(t) = -ta + \mu(e^t - 1)$, then $\phi'(t) = -a + \mu e^t$. Set $\phi'(t) = 0$, we can find the optimal $t = \ln\frac{a}{\mu}$. Therefore, $P(S \geq a) \leq e^{-a\ln\frac{a}{\mu}+a-\mu} = e^{-91\ln\frac{91}{\mu}+91-\mu}$.

Problem 5

(a)
1. Since N is the smallest interget such that $S_N > 1$, therefore $N = n$ occurs when $S_{n-1} < 1$ and $S_n > 1$. So $P(N = n) = P(S_{n-1} < 1$ and $S_n > 1) = P(S_{n-1} < 1) - P(S_n \leq 1)$. Since $P(S_n = 1) = 0$, $P(N = n) = P(S_{n-1} < 1) - P(S_n < 1)$.
2. Proof $P(S_k < v) = \frac{v^k}{k!}$ by induction:
Base case: $P(S_1 < 1) = 1 = \frac{1^1}{1!}$
Inductive step: Assume $n = k$, $P(S_k < 1) = \frac{v^k}{k!}$. Then for $n = k+1$, $P(S_{k+1} < 1) = \int_{u_{k+1}=0}^{1} P(S_k < 1 - u_{k+1})\,du_{k+1}$. Let $v = 1 - u_{k+1}$, then $P(S_{k+1} < 1) = \int_{v=0}^{1} P(S_k < v)\,dv = \int_{v=0}^{1}\frac{v^k}{k!}\,dv = \frac{1}{k!}\int_0^1 v^k\,dv = \frac{1}{k!}\left[\frac{v^{k+1}}{k+1}\right]_{v=0}^{1} = \frac{1}{k!} \cdot \frac{1}{k+1} = \frac{1}{(k+1)!}$.
Therefore, $P(S_k < 1) = \frac{1^k}{k!} = \frac{1}{k!}$.
3. $P(N = n) = P(S_{n-1} < 1) - P(S_n < 1) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n}{n!} - \frac{1}{n!} = \frac{n-1}{n!}$.
$E[N] = \sum_{n=1}^{\infty} n \times P(N = n) = \sum_{n=1}^{\infty}\frac{n(n-1)}{n!} = \frac{1\cdot 0}{1!} + \sum_{n=2}^{\infty}\frac{1}{(n-2)!} = \sum_{m=0}^{\infty}\frac{1}{m!} = e$

(b) Procedure of simulation:
1. Define the number of trials N_{trials} of each simulation run.
2. Generate a sequence of i.i.d. $Uniform(0,1)$ random variables $U_1, U_2, \ldots, U_{N_{trials}}$.
3. For each $i = 1, 2, \ldots, N_{trials}$, record the value n such that $S_{n-1} < 1$ and $S_n > 1$.
4. After all trials, compute the average of all recorded N_i values. This average is an estimate of e.

Problem 1

(a) By Markov's inequality, $0 \leq P(|X_n = c| \geq \varepsilon) = P((X_n = c)^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[(X_n-c)^2]}{\varepsilon^2}$. Take the limit as $n \to \infty$, we have $0 \leq \lim_{n\to\infty}((X_n = c)^2 \geq \varepsilon^2) \leq \lim_{n\to\infty}(\frac{\mathbb{E}[(X_n-c)^2]}{\varepsilon^2}) = 0$. Therefore, by the squeeze theorem, $\lim_{n\to\infty} P(|X_n = c| \geq \varepsilon) = 0$, which implies convergence in probability.

(b) Consider the random variable
$$X_n = \begin{cases} 1 - \frac{1}{n} & \text{if } x = 0 \\ \frac{1}{n} & \text{if } x = n \\ 0 & \text{otherwise} \end{cases}$$
Then this random variable converges in probability since $0 \leq \lim_{n\to\infty}(P(|Y_n - 0| > \varepsilon)) \leq \lim_{n\to\infty}(\frac{1}{n}) = 0$. However, $\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[x_n^2] = 0^2 \cdot (1 - \frac{1}{n}) + n^2 \cdot \frac{1}{n} = n$, which implies $\lim_{n\to\infty}\mathbb{E}[x_n^2] = \infty$. Therefore, convergence in probability does not imply convergence in the mean square.

Problem 2

$|X_nY_n - ab| = |X_nY_n - aY_n + aY_n - ab| = |Y_n||X_n - a| + |a||Y_n - b|$. By triangle inequality, we know that $|X_nY_n - ab| \leq |Y_n||X_n - a| + |a||Y_n - b|$. To have $|X_nY_n - ab| < \varepsilon$, we can set $\varepsilon_1 = \frac{\varepsilon}{2(|b|+1)}$ and $\varepsilon_2 = \frac{\varepsilon}{2|a|}$, so that $|X_n - a| < \varepsilon_1$ and $|Y_n - b| < \varepsilon_2$, which ensures $|X_n - a|(|b| + 1) < \frac{\varepsilon}{2}$ and $|Y_n - b||a| < \frac{\varepsilon}{2}$. Using the above bounds, we have
$$P(|X_nY_n - ab| \geq \varepsilon) \leq P\left(|X_n - a| \geq \frac{\varepsilon}{2(|b|+1)}\right) + P\left(|Y_n - b| \geq \frac{\varepsilon}{2|a|}\right)$$
Since $X_n \xrightarrow{P} a$ and $Y_n \xrightarrow{P} b$, for any $\delta > 0$, $\lim_{n\to\infty} P(|X_n - a| \geq \delta) = 0$ and $\lim_{n\to\infty} P(|Y_n - b| \geq \delta) = 0$. By choosing $\delta = \frac{\varepsilon}{2(|b|+1)}$ and $\delta = \frac{\varepsilon}{2|a|}$ respectively, we ensure
$$\lim_{n\to\infty} P\left(|X_n - a| \geq \frac{\varepsilon}{2(|b|+1)}\right) = 0 \quad \text{and} \quad \lim_{n\to\infty} P\left(|Y_n - b| \geq \frac{\varepsilon}{2|a|}\right) = 0$$
Therefore, $\lim_{n\to\infty} P(|X_nY_n - ab| \geq \varepsilon) \leq 0 + 0 = 0$.

4-2 Prove: $\{(X_n \cdot Y_n)\}_{n=1}^{\infty}$ converges to ab in probability

Problem 3

(a) By definition, $C_{n+1} = C_n \cdot X_{n+1}$, then $\mathbb{E}[C_{n+1}|C_n] = \mathbb{E}[C_n \cdot X_{n+1}|C_n] = C_n \cdot \mathbb{E}[X_{n+1}|C_n]$. Since the random variable x_{n+1} is independent of C_n, we can know that $\mathbb{E}[X_{n+1}|C_n] = \mathbb{E}[X_{n+1}] = 1$. Therefore, $\mathbb{E}[C_{n+1}|C_n] = C_n \cdot 1 = C_n$.

(b) Since $C_n = \prod_{k=1}^{n} X_k$, we have $\ln C_n = \sum_{k=1}^{n}\ln X_k$. We can define $Y_k = \ln X_k$, then $\ln C_n = \sum_{k=1}^{n} Y_k$. Since the logarithm is a concave function, we can apply Jensen's inequality to get $\mathbb{E}[\ln X_k] \leq \ln\mathbb{E}[X_k] = \ln 1 = 0$, hence $\mathbb{E}[Y_k] \leq 0$. Given $P(X_k = 1) < 1$, X_k is not almost surely 1. Therefore, $\mathbb{E}[\ln X_k] < \ln\mathbb{E}[X_k] = 0$, so $\mu := \mathbb{E}[Y_k] = \mathbb{E}[\ln X_k] < 0$. Applying SLLN we have $\frac{1}{n}\ln C_n = \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{a.s.} \mu < 0$ as $n \to \infty$. Therefore, $\ln C_n \xrightarrow{a.s.} n\mu$ as $n \to \infty$. Since $\mu < 0$, $n\mu \to -\infty$ as $n \to \infty$, then $\ln C_n \xrightarrow{a.s.} -\infty$. Thus $C_n = e^{\ln C_n} \xrightarrow{a.s.} e^{-\infty} = 0$. $C_n \xrightarrow{a.s.} 0$, $n \to \infty$

Problem 4

(a) Each X_i is a Bernoulli random variable with $\mathbb{E}[X_i] = p$ and $\text{Var}(X_i) = p(1-p)$. So the mean of S_n is $\mathbb{E}[S_n] = \sum_{i=1}^{n}\mathbb{E}[X_i] = np$, and the variance of S_n is $\text{Var}(S_n) = \sum_{i=1}^{n}\text{Var}(X_i) = np(1-p)$. To apply the Central Limit Theorem, we standardize S_n to $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$. Then by CLT, as $n \to \infty$, $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$. Therefore, $S_n \sim N(np, np(1-p))$.

(b) $Q_n(x) = P(S_n = x) = \binom{n}{x}p^x(1-p)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$. For fixed x and large n, $\binom{n}{x} = \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} = \frac{n^x}{x!}\prod_{k=0}^{x-1}\left(1 - \frac{k}{n}\right) \approx \frac{n^x}{x!}$. Therefore, $Q_n(x) \approx \frac{n^x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^{n-x}$. Now we can analyze the limit of $Q_n(x)$ as $n \to \infty$. We have $\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$ and $\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{-x} \approx \lim_{n\to\infty}\left(1 + \frac{x\lambda}{n}\right) = 1$. As a result, $\lim_{n\to\infty} Q_n(x) = \frac{\lambda^x}{x!}e^{-\lambda}$, which is the probability mass function of a Poisson distribution with parameter λ.

(c) The Central Limit Theorem (CLT) assumes that $\mathbb{E}[S_n] = np$, and $\text{Var}(S_n) = np(1-p)$, scale appropriately with n. However, when $p = \frac{\lambda}{n}$, $\mathbb{E}[S_n] = np = \lambda$ (constant), and $\text{Var}(S_n) = np(1-p) \approx \lambda(1 - \frac{\lambda}{n}) \to \lambda$ (constant as $n \to \infty$). Since the mean and variance of S_n are constant, this violates the CLT requirement that the variance grows with n. As a result, the distribution does not spread out enough to approach a normal distribution, making CLT inapplicable.

$\frac{n^x}{x!}\prod_{k=0}^{x-1}(1 - \frac{k}{n}) \approx \frac{n^x}{x!}$

$P(\ldots | M=m) = \frac{C_{m-2}^{2n-2}}{C_m^{2n}} = \frac{m(m-1)}{2(2n-1)}$

Problem 5

(a) The total number of ways to choose m students out of $2n$ is $\binom{2n}{m}$. The number of ways to choose m students that include two specific students is $\binom{2n-2}{m-2}$ because we have fixed 2 students and need to choose the remaining $m-2$ students from the remaining $2n-2$. Thus, the probability that both members of a given pair are chosen is $P(\text{both chosen} \mid M = m) = \frac{\binom{2n-2}{m-2}}{\binom{2n}{m}} = \frac{\frac{(2n-2)!}{(m-2)!(2n-m)!}}{\frac{(2n)!}{m!(2n-m)!}} = \frac{(2n-2)!(m!)}{(m-2)!(2n)!} = \frac{m(m-1)}{(2n)(2n-1)}$.
Since there are n independent pairs, and for each pair the probability of both being chosen is the same, we have $E[X \mid M = m] = n \cdot P(\text{both chosen} \mid M = m) = \frac{m(m-1)}{2(2n-1)}$.

(b) From law of total expectation, $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid M]] = \sum_{m=0}^{2n}\frac{m(m-1)}{2(2n-1)}P(M = m) = \frac{1}{2(2n-1)}\mathbb{E}[M(M-1)]$. Since $M \sim \text{Binomial}(2n, p)$, $\mathbb{E}[M(M-1)] = 2n(2n-1)p^2$. Therefore, $\mathbb{E}[X] = \frac{1}{2(2n-1)}\mathbb{E}[M(M-1)] = \frac{1}{2(2n-1)}2n(2n-1)p^2 = np^2$.

Problem 4 (Sum of Independent Random Variables and Chernoff Bounds) (8+8+8=24 points)

In this problem, from a probabilistic perspective, we study the fine-tuning method behind ChatGPT – *Reinforcement Learning from Human Feedback* (RLHF). Specifically, under RLHF, ChatGPT is fine-tuned by repeatedly applying the following *pairwise text comparison by human labelers* :

- Each piece of text, denoted by $\mathcal{T}$ has some underlying true score $R(\mathcal{T})$, which is a scalar for specifying the quality of the text (the higher the score, the better the quality).
- Define the Sigmoid function as $\sigma(x) := 1/(1 + \exp(-x))$. It is easy to verify that (i) $\sigma(x) \in (0,1)$ for all $x \in \mathbb{R}$ and (ii) $\sigma(x) = 1 - \sigma(-x)$ for all $x \in \mathbb{R}$.
- At each trial, two pieces of texts $\mathcal{T}_a, \mathcal{T}_b$ (with true scores $R(\mathcal{T}_a)$ and $R(\mathcal{T}_b)$, respectively) are provided to a human labeler for comparison. Under the RLHF model, the human labeler would output one of the following two responses: (i) The labeler would vote for "$\mathcal{T}_a$ is better than $\mathcal{T}_b$", with probability $\sigma(R(\mathcal{T}_a) - R(\mathcal{T}_b))$. (ii) The labeler would vote for "$\mathcal{T}_b$ is better than $\mathcal{T}_a$", with probability $\sigma(R(\mathcal{T}_b) - R(\mathcal{T}_a))$.
- All the comparison trials by the human labelers are assumed to be independent.

(a) Suppose our CS department would like to train a new GPT model, called NYCU-GPT, by using RLHF. We hire n human labelers (n is odd) and provide the same two pieces of texts (denoted by $R(\mathcal{T}_a)$ and $R(\mathcal{T}_b)$) to these n labelers. Suppose $R(\mathcal{T}_a) > R(\mathcal{T}_b)$. Define an event $E := \{\mathcal{T}_a$ get more votes than $\mathcal{T}_b\}$. Then, by using Hoeffding's inequality, what could we say about $P(E)$? (Please provide a lower bound for $P(E)$).

(b) Based on the result in (a), suppose we would like to ensure that $P(E) \geq 1 - \delta$, where δ is a small positive constant. Then, by Chebyshev's inequality, how many human labelers do we need to achieve this? Similarly, suppose we use the Hoeffding's inequality for the analysis instead. How many human labelers do we need to ensure that $P(E) \geq 1 - \delta$?

(c) For further training, we proceed to prepare a collection of 100 pieces of texts, denoted by $\mathcal{T}_1, \cdots, \mathcal{T}_{100}$. Suppose we already know that the scores satisfy that $R(\mathcal{T}_1) > R(\mathcal{T}_2) > \cdots > R(\mathcal{T}_{100})$. Then, we ask a human labeler (whose name is Sam Altmann), to finish pairwise text comparison for all the pairs of these 100 pieces of texts. Define random variables $I_j := \mathbb{I}\{$Sam thinks $R(\mathcal{T}_j)$ is better than $R(\mathcal{T}_1)\}$ (where $\mathbb{I}\{A\}$ is the indicator function of an event A, i.e., $\mathbb{I}\{A\}$ would output 1 if the event occurs and output 0 otherwise.) Please use the Chernoff technique to find an upper bound of $P(I_2 + I_3 + \cdots + I_{100} > 90)$. (Hint: Note that we could NOT directly apply Hoeffding's inequality here since $\{I_j\}$'s are independent but NOT identically distributed. Despite this, we may still use the Chernoff technique to find an upper bound by leveraging independence)