ma5259_hw1

Nguyen Ngoc Khanh - A0275047B

September 2024

1 Lecture Exercises

1.1 Lecture 1 Exercise 1

If $E, F \subseteq \Omega$ are two events, then we can define two random variables $1_E, 1_F : \Omega \to \mathbb{R}$ with $1_E(\omega) = 1_{w \in E}$ and $1_F(\omega) = 1_{w \in F}$. Show the independence of the events E and F is equivalent to the independence of random variables 1_E and 1_F

1.1.1 Independence of E, F implies independence of $1_E, 1_F$

Let $X : \Omega \to \mathbb{R}^2$ be the joint random variable defined by $X(\omega) = (1_E(\omega), 1_F(\omega))$. For any Borel measurable rectangle $A \times B \subseteq \mathbb{R}^2$, we have $A \times B = (A \times \mathbb{R}) \cap (\mathbb{R} \times B)$, therefore

$$X^{-1}(A \times B) = X^{-1}(A \times \mathbb{R}) \cap X^{-1}(\mathbb{R} \times B) = 1_E^{-1}A \cap 1_F^{-1}B$$

Note that, the independence of $E, F \subseteq \Omega$ implies the following independence:

$$\begin{aligned}
P(E \cap F) &= P(E)P(F) \\
P(E \cap F^C) &= P(E)P(F^C) \\
P(E^C \cap F) &= P(E^C)P(F) \\
P(E^C \cap F^C) &= P(E^C)P(F^C)
\end{aligned}$$

As $1_E^{-1}A$ is either E or E^C, $1_F^{-1}A$ is either F or F^C, then the two events $1_E^{-1}A$ and $1_F^{-1}B$ are independent for all Borel measurable sets $A, B \subseteq \mathbb{R}$, therefore

$$P(1_E \in A, 1_F \in B) = P(X^{-1}(A \times B)) = P(1_E^{-1}A \cap 1_F^{-1}B) = P(1_E^{-1}A)P(1_F^{-1}B) = P(1_E \in A)P(1_F \in B)$$

1.1.2 Independence of $1_E, 1_F$ implies independence of E, F

Similarly, the independence of $1_E, 1_F$ implies for all measurable sets A, B

$$P(1_E^{-1}A \cap 1_F^{-1}B) = P(1_E \in A, 1_F \in B) = P(1_E \in A)P(1_F \in B) = P(1_E^{-1}A)P(1_F^{-1}B)$$

Take $A = B = \{1\}$, then

$$P(E \cap F) = P(E)P(F)$$

1.2 Lecture 1 Exercise 2

Let X and Y denote the outcome of two independent fair coin tosses. Let $Z := Head$ if $X = Y$ and $Z := Tail$ if $X \neq Y$. Show that X, Y, Z are pairwise independent but not jointly independent.

1.2.1 X, Y, Z are pairwise independent

X, Y are independent by the premise. We need to prove the independence of the pair X, Z.
Let X, Y, Z be defined on $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ ($\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{H, T\}$). Given any $(x, z) \in \mathcal{X} \times \mathcal{Z}$. Define the two disjoint sets

$$\begin{aligned}
\mathcal{F}_{xz} &= \{y \in \mathcal{Y} : P(Z = z | X = x, Y = y) = 1\} \\
\mathcal{G}_{xz} &= \{y \in \mathcal{Y} : P(Z = z | X = x, Y = y) = 0\}
\end{aligned}$$

By the premise, we have $\mathcal{F}_{xz} \amalg \mathcal{G}_{xz} = \mathcal{Y}$ and $|\mathcal{F}_{xz}| = 1$, therefore

$$\begin{aligned}
P(X=x, Z=z) &= \sum_{y\in\mathcal{Y}} P(X=x, Y=y, Z=z) && \text{(marginalize)}\\
&= \sum_{y\in\mathcal{Y}} P(X=x, Y=y) P(Z=z|X=x, Y=y) && \text{(conditional probability)}\\
&= \sum_{y\in\mathcal{F}_{xz}} P(X=x, Y=y) && \text{(definition of } \mathcal{F}_{xz}, \mathcal{G}_{xz})\\
&= \sum_{y\in\mathcal{F}_{xz}} P(X=x) P(Y=y) && (X, Y \text{ are independent})\\
&= P(X=x)P(Y=y) && (|\mathcal{F}_{xz}| = 1)\\
&= P(X=x)P(Z=z) && (P(Y=y) = P(Z=z) = 1/2 \text{ for all } y, z)
\end{aligned}$$

1.2.2 X, Y, Z are not jointly independent

We have $P(X=H) = P(Y=H) = P(Z=H) = 1/2$, so $P(X=H)P(Y=H)P(Z=H) = 1/8$ but

$$\begin{aligned}
P(X=H, Y=H, Z=H) &= P(Z=H|X=H, Y=H)P(X=H, Y=H)\\
&= 1 \cdot (1/2 \cdot 1/2) = 1/4
\end{aligned}$$

1.3 Lecture 2 Exercise 1

Let (Ω, F, P) be a probability space. Show the following are all equivalent given P is finitely additive

1. P is countably additive on F
2. For any $A_1 \subseteq A_2 \subseteq ... \subseteq \Omega$ with $A_i \in F$,
$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n)$$
3. For any $\Omega \supseteq A_1 \supseteq A_2 \supseteq ...$ with $A_i \in F$,
$$P\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n)$$
4. For any $\Omega \supseteq A_1 \supseteq A_2 \supseteq ...$ with $\bigcap_{n\in\mathbb{N}} A_n = \varnothing$, $P(A_n) \searrow 0$ as $n \to \infty$

1.3.1 $(1 \implies 2)$

Let $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, ..., $B_i = A_i \setminus A_{i-1}$. So that, $\{B_n\}_{n\in\mathbb{N}}$ is a collection of disjoint measurable sets, $A_n = \coprod_{i=1}^n B_i$ and $\bigcup_{n\in\mathbb{N}} A_n = \coprod_{n\in\mathbb{N}} B_n$. We have

$$\begin{aligned}
P\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= P\left(\coprod_{n\in\mathbb{N}} B_n\right)\\
&= \sum_{n=1}^{\infty} P(B_n) && (B_n \text{ disjoint, } P \text{ is countably additive})\\
&= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) && \text{(infinite sum)}\\
&= \lim_{n\to\infty} P\left(\coprod_{i=1}^{n} B_i\right) && (B_i \text{ disjoint, } P \text{ is finite additive})\\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

1.3.2 $(2 \implies 3)$

Let $B_n = \Omega \setminus A_n$ so that $B_1 \subseteq B_2 \subseteq ...$ and $\bigcap_{n\in\mathbb{N}} A_n = \Omega \setminus \bigcup_{n\in\mathbb{N}} B_n$. By finite additivity, $P(A_n) = 1 - P(B_n)$, $P\big(\bigcap_{n\in\mathbb{N}} A_n\big) = 1 - P\big(\bigcup_{n\in\mathbb{N}} B_n\big)$. By (2), $\lim_{n\to\infty} P(B_n) = P\big(\bigcup_{n\in\mathbb{N}} B_n\big)$. Hence, $\lim_{n\to\infty} P(A_n)$ exists and

$$\begin{aligned}
\lim_{n\to\infty} P(A_n) &= 1 - \lim_{n\to\infty} P(B_n) \\
&= 1 - P\left(\bigcup_{n\in\mathbb{N}} B_n\right) && \text{(by (2))} \\
&= P\left(\bigcap_{n\in\mathbb{N}} A_n\right)
\end{aligned}$$

1.3.3 $(3 \implies 4)$

By (3)

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n\in\mathbb{N}} A_n\right) = P(\varnothing) = 0$$

1.3.4 $(4 \implies 1)$

Let $\{B_n\}_{n\in\mathbb{N}}$ be a collection of disjoint measurable sets. Let $A_n = \coprod_{i=n+1}^{\infty} B_i$ so that $\coprod_{n\in\mathbb{N}} B_n = A_0 \supseteq A_1 \supseteq ...$ and $\bigcap_{n\in\mathbb{N}_0} A_n = \varnothing$. By finite additivity, $\sum_{i=1}^{n} P(B_n) = P(\coprod_{i=1}^{n} B_n) = P(A_1 \setminus A_n) = P(A_1) - P(A_n)$. By (4), $\lim_{n\to\infty} P(A_n) = P\big(\bigcap_{n\in\mathbb{N}_0} A_n\big) = P(\varnothing) = 0$, then

$$\sum_{i=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_n) = P(A_1) - \lim_{n\to\infty} P(A_n) = P(A_1) = P\left(\bigcup_{n\in\mathbb{N}} B_n\right)$$

1.4 Lecture 2 Exercise 2

Let Z be a uniform random variable on $[0,1]$. How can one generate from Z discrete random variable X with geometric distribution $P(X=n) = 2^{-n}$ for each $n \in \mathbb{N}$?

Answer. Let $X : [0,1] \to \mathbb{N}$ be defined by

$$X = \begin{cases} n & \text{if } Z \in (2^{-n}, 2\cdot 2^{-n}] \\ 1 & \text{if } Z = 0 \end{cases}$$

So that if $P(X=1) = P(Z \in (1/2, 1]) + P(Z=0) = 2^{-1}$ and if $n > 1$

$$P(X=n) = P(Z \in (2^{-n}, 2\cdot 2^{-n}]) = 2\cdot 2^{-n} - 2^{-n} = 2^{-n}$$

□

1.5 Lecture 3 Exercise 1

Construct a random variable X and Y such that $\mathrm{Cov}(X,Y) = 0$ but X, Y are not independent. However, show that if X, Y are Bernoulli random variables, then $\mathrm{Cov}(X,Y) = 0$ implies that X, Y are independent

1.5.1 Construction of X, Y so that $\mathrm{Cov}(X,Y) = 0$ but X, Y are not independent

Let $\Omega = \{-1, +1\}^2$ with the σ-algebra F consists of all subsets of Ω and a uniform probability measure $P : F \to \mathbb{R}$. Let $X : \Omega \to \mathbb{R}$ be defined by $X(a,b) = a$ and $Y : \Omega \to \mathbb{R}$ be defined by

$$\begin{aligned}
Y : \Omega &\to \mathbb{R} \\
(-1,-1) &\mapsto 0 \\
(-1,+1) &\mapsto 0 \\
(+1,-1) &\mapsto -1 \\
(+1,+1) &\mapsto +1
\end{aligned}$$

So that $\mathbb{E}[X] = 0$, $\mathbb{E}[Y] = 0$, $\text{Cov}(X, Y) = \mathbb{E}[XY]$. XY have the following values

$$\begin{aligned}
XY &: \Omega \to \mathbb{R} \\
(-1, -1) &\mapsto 0 \\
(-1, +1) &\mapsto 0 \\
(+1, -1) &\mapsto (+1)(-1) = -1 \\
(+1, +1) &\mapsto (+1)(+1) = +1
\end{aligned}$$

Hence, $\text{Cov}(X, Y) = \mathbb{E}[XY] = 0$. On the other hand, $P(X = +1) = 1/2$, $P(Y = 0) = 1/2$, but

$$P(X = +1, Y = 0) = P(Y = 0|X = +1)P(X = +1) = 0$$

1.5.2 $\text{Cov}(X, Y) = 0$ implies X, Y are independent given X, Y are Bernoulli random variables

Let X, Y be Bernoulli random variables

$$\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= P(X = 1, Y = 1) - P(X = 1)P(Y = 1)
\end{aligned}$$

Hence, $\text{Cov}(X, Y) = 0$ implies $P(X = 1, Y = 1) = P(X = 1)P(Y = 1)$ implies $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all $x, y \in \{0, 1\}$, that is, X, Y are independent

1.6 Lecture 4 Exercise 1

1. $X \sim N(0, \sigma^2)$ has characteristic function $\phi(t) = e^{-\frac{\sigma^2 t^2}{2}}$. Compute the first 4 moments of X
2. Compute the characteristic of $X \sim Pois(\lambda)$

1.6.1 $X \sim N(0, \sigma^2)$, $\phi(t) = e^{-\frac{\sigma^2 t^2}{2}}$, compute the first 4 moments

We have

$$\begin{aligned}
\phi^{(1)}(t) &= \left(e^{-\frac{\sigma^2 t^2}{2}}\right)\left(-\frac{\sigma^2}{2}\right)2t \\
&= -\sigma^2 t\phi(t) \\
\phi^{(2)}(t) &= -\sigma^2\phi(t) - \sigma^2 t\phi^{(1)}(t) \\
\phi^{(3)}(t) &= -\sigma^2\phi^{(1)}(t) - \left(\sigma^2\phi^{(1)}(t) + \sigma^2 t\phi^{(2)}(t)\right) \\
&= -2\sigma^2\phi^{(1)}(t) - \sigma^2 t\phi^{(2)}(t) \\
\phi^{(4)}(t) &= -2\sigma^2\phi^{(2)}(t) - \left(\sigma^2\phi^{(2)}(t) + \sigma^2 t\phi^{(3)}(t)\right) \\
&= -3\sigma^2\phi^{(2)}(t) - \sigma^2 t\phi^{(3)}(t)
\end{aligned}$$

So,

$$\begin{aligned}
\phi(0) &= 1 \\
\phi^{(1)}(0) &= 0 \\
\phi^{(2)}(0) &= -\sigma^2 \\
\phi^{(3)}(0) &= 0 \\
\phi^{(4)}(0) &= (-3\sigma^2)(-\sigma^2) = 3\sigma^4
\end{aligned}$$

As $\phi^{(k)}(0) = i^k\mathbb{E}[X^k]$,

$$\begin{aligned}
\mathbb{E}[X] &= 0 \\
\mathbb{E}[X^2] &= \sigma^2 \\
\mathbb{E}[X^3] &= 0 \\
\mathbb{E}[X^4] &= 3\sigma^4
\end{aligned}$$

1.6.2 $X \sim Pois(\lambda)$, compute the characteristic function

Poisson distribution

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Characteristic function

$$\begin{aligned}
\phi(t) &= \int_{\mathbb{N}} e^{itk} dp(k) \\
&= \sum_{k=0}^{\infty} e^{itk} p(k) \\
&= \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k e^{-\lambda}}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} \\
&= e^{-\lambda} e^{\lambda e^{it}} \\
&= e^{-\lambda + \lambda e^{it}}
\end{aligned}$$

2 Homework Questions

2.1 Q1

Let X_1, X_2 be two independent exponential random variables with parameters $\lambda_1, \lambda_2 > 0$ respectively. In other words, $P(X_1 > x) = e^{-\lambda_1 x}$ and $P(X_2 > x) = e^{-\lambda_2 x}$ for all $x \geq 0$. Let $X := \min\{X_1, X_2\}$. Show that X is again an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$

Proof.

$$\begin{aligned}
P(X > x) &= P(X_1 > x, X_2 > x) && (\{\min\{X_1, X_2\} > x\} \text{ and } \{X_1 > x\} \cap \{X_2 > x\} \text{ are the same subset in } \Omega) \\
&= P(X_1 > x) P(X_2 > x) && (X_1, X_2 \text{ are independent}) \\
&= e^{-\lambda_1 x} e^{-\lambda_2 x} \\
&= e^{-(\lambda_1 + \lambda_2) x}
\end{aligned}$$

As $P(X > x)$ determines the distribution of X and it matches the exponential distribution with parameter $\lambda = \lambda_1 + \lambda_2$, therefore, X is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$

□

2.2 Q2

12 people stand in a circle. Independently, every pair of neighbours decide to link arms with probability $1/2$. This breaks 12 people into disjoint groups. What is the expected number of groups this results in?

Answer. Let $X_1, ..., X_{12}$ be 12 i.i.d Bernoulli random variables with $p = 1/2$, X_i represents whether there is a missing link between person i and person $i + 1$ (or 12 and 1). Let $X = X_1 + ... + X_{12}$, then X is the number of missing links. Let Y be the number of connected components, then

$$Y = \begin{cases} 1 & \text{if } X = 0, 1 \\ X & \text{otherwise} \end{cases}$$

We have

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{k=1}^{12} kP(Y=k) \\
&= P(Y=1) + \sum_{k=2}^{12} kP(Y=k) \\
&= P(Y=1) + \sum_{k=2}^{12} kP(X=k) && (P(Y=k) = P(X=k) \text{ if } k \geq 2) \\
&= P(Y=1) - P(X=1) + \sum_{k=0}^{12} kP(X=k) \\
&= P(Y=1) - P(X=1) + \mathbb{E}[X] \\
&= (P(X=0) + P(X=1)) - P(X=1) + \mathbb{E}[X] && (P(Y=1) = P(X=0) + P(X=1)) \\
&= P(X=0) + \mathbb{E}[X]
\end{aligned}
$$

We have $P(X=0) = P(X_1 = 0, X_2 = 0, ..., X_{12} = 0) = \frac{1}{2^{12}}$ and

$$
\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_1 + ... + X_{12}] \\
&= \mathbb{E}[X_1] + ... + \mathbb{E}[X_{12}] && \text{(independent)} \\
&= 12\mathbb{E}[X_1] = 6 && \text{(identical)}
\end{aligned}
$$

Hence, $\mathbb{E}[Y] = \frac{1}{2^{12}} + 6$

□

2.3 Q3

We toss a fair coin repeatedly. What is the expected number of coin tosses it takes to observe Head followed by two consecutive Tails?

Answer. Let $X = (X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d Bernoulli random variables of probability $p = 1/2$ each represents a coin toss. Let $Y = (Y_n)_{n\in\mathbb{N}}$ be another sequence such that

$$Y_n = (X_n, X_{n+1}, X_{n+2})$$

Then, Y is a time-homogeneous Markov chain with state space

$$S = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

and transition matrix Π and Y_1 is distributed uniformly on S. For any $n \in \mathbb{N}$, define $g : \mathbb{N} \to \mathbb{R}$ by

$$g(n) = P(Y_1 \neq HTT, Y_2 \neq HTT, ..., Y_n \neq HTT)$$

Then, by defintion of conditional probability

$$g(n) = P(Y_n \neq HTT | Y_1 \neq HTT, ..., Y_{n-1} \neq HTT)g(n-1)$$

We have

$$
\begin{aligned}
&P(Y_n \neq HTT | Y_1 \neq HTT, ..., Y_{n-1} \neq HTT) \\
&= P(Y_n \neq HTT | Y_{n-1} \neq HTT) && \text{(Markov property)} \\
&= 1 - P(Y_n = HTT | Y_{n-1} \neq HTT) && \text{(complement event)} \\
&= 1 - \frac{P(Y_n = HTT, Y_{n-1} \neq HTT)}{P(Y_{n-1} \neq HTT)} && \text{(conditional probability)} \\
&= 1 - \frac{P(Y_n = HTT) - P(Y_n = HTT, Y_{n-1} = HTT)}{1 - P(Y_{n-1} = HTT)} && \text{(complement event)}
\end{aligned}
$$

For all $n \in \mathbb{N}$, $P(Y_n = HTT) = P(X_n = H, X_{n+1} = T, X_{n+2} = T) = 1/8$ and $P(Y_n = HTT, Y_{n-1} = HTT) = 0$, then

$$P(Y_n \neq HTT | Y_1 \neq HTT, ..., Y_{n-1} \neq HTT) = 1 - \frac{1/8}{1 - 1/8} = \frac{6}{7}$$

As $g(1) = P(Y_1 \neq HTT) = 1 - P(Y_1 = HTT) = 7/8$, then, for all $n \in \mathbb{N}$

$$g(n) = \frac{7}{8}\left(\frac{6}{7}\right)^{n-1}$$

Let $f : \mathbb{N} - \{1\} \to \mathbb{R}$ be defined by

$$f(n) = P(Y_1 \neq HTT, Y_2 \neq HTT, ..., Y_{n-1} \neq HTT, Y_n = HTT)$$

Then $f(n) = g(n-1) - g(n)$ for all $n \in \mathbb{N} - \{1\}$. We extend the domain of f by defining $f(1) = P(Y_n = HTT) = \frac{1}{8}$. Now, on the space $\mathrm{Hom}(\mathbb{N}, S)$ of all sequences of realizations of Y_i, we define function $\phi : \mathrm{Hom}(\mathbb{N}, S) \to \mathbb{N}$ by the first time observing $Y_i = HTT$. Then $\mathbb{E}[\phi]$ is the expected time to observe $Y_i = HTT$. We can decompose $\mathrm{Hom}(\mathbb{N}, S)$ into a sequence of disjoint events

$$\mathrm{Hom}(\mathbb{N}, S) = \{Y_n \neq HTT : n \in \mathbb{N}\} \amalg \{Y_1 = HTT\} \amalg \{Y_1 \neq HTT, Y_2 = HTT\} \amalg \amalg \{Y_1 \neq HTT, Y_2 \neq HTT, Y_3 = HTT\} \amalg ...$$

with $P(\{Y_n \neq HTT : n \in \mathbb{N}\}) = 0$. Therefore,

$$\begin{aligned}
\mathbb{E}[\phi] &= \sum_{n=1}^{\infty} n f(n) \\
&= f(1) + \sum_{n=2}^{\infty} n f(n) \\
&= f(1) + \sum_{n=2}^{\infty} n(g(n-1) - g(n)) \\
&= f(1) + \sum_{n=2}^{\infty} n g(n-1) - \sum_{n=2}^{\infty} n g(n) && \text{suppose } \sum_{n=2}^{\infty} n g(n) < \infty \\
&= f(1) + \left(\sum_{n=2}^{\infty} (n-1) g(n-1) + \sum_{n=2}^{\infty} g(n-1)\right) - \sum_{n=2}^{\infty} n g(n) \\
&= f(1) + \sum_{n=1}^{\infty} g(n) + \sum_{n=1}^{\infty} n g(n) - \sum_{n=2}^{\infty} n g(n) \\
&= f(1) + \sum_{n=1}^{\infty} g(n) + g(1) \\
&= f(1) + g(1) + \frac{7}{8} \sum_{n=1}^{\infty} \left(\frac{6}{7}\right)^n \\
&= \frac{1}{8} + \frac{7}{8} + \frac{7}{8}\frac{6/7}{1 - 6/7} = \frac{25}{4} = 6.25
\end{aligned}$$

Expected number of coin tosses $6.25 + 2 = 8.25$

Lemma 1. *Let $\alpha \in (0,1)$, $\sum_{n=1}^{\infty} n\alpha^n$ converges.*

Proof. Let $\alpha < \beta < 1$, then there exists $N \in \mathbb{N}$, such that for all $n > N$, $n < \left(\frac{\beta}{\alpha}\right)^n$. We write

$$\sum_{n=1}^{\infty} n\alpha^n = \sum_{n=1}^{N} n\alpha^n + \sum_{n=N}^{\infty} n\alpha^n < \sum_{n=1}^{N} n\alpha^n + \sum_{n=N}^{\infty} \beta^n < \infty$$

□

□

2.4 Q4

Let G_n be a random graph with n vertices $\{1, ..., n\}$ where every pair of vertices is connected by an edge independently with probability $p \in (0, 1)$. A set of four vertices $\{i, j, k, l\}$ is said to form a square if there are exactly 4 edges among them, forming a square.

1. What is the expected number of squares in G_n?
2. Find a suitable upperbound on variance of N_n as $n \to \infty$ and use it to prove the weak law of large numbers for $\frac{N_n}{\mathbb{E}[N_n]}$. More precisely, show that

$$P\left(\left|\frac{N_n}{\mathbb{E}[N_n]} - 1\right| > \epsilon\right) \to 0$$

as $n \to \infty$ for every ϵ

2.4.1 Expected number of squares

Let $[n] = \{i \in \mathbb{N} : i < n\}$, let $\mathcal{S}_m = \{S \in \mathcal{P}([n]) : |S| = m\}$ be the collection of subsets of size m of n vertices. Let $f : \mathcal{S}_4 \to \{0, 1\}$ denote the function that has value $f(S) = 1$ if S is a square. If $S = \{a, b, c, d\}$ then $f(S) = 1$ if and only if one of the following disjoint events happen ($ab = 1$ denotes $a - b$ being an edge)

- E_1: $ab = bc = cd = da = 1, ac = bd = 0$
- E_2: $ac = cd = db = ba = 1, ad = bc = 0$
- E_3: $ac = cb = bd = da = 1, ab = cd = 0$

Given $S = \{a, b, c, d\} \in \mathcal{S}_4$, then

$$\begin{aligned}
\mathbb{E}[f(S)] &= P(E_1) + P(E_2) + P(E_3) && \text{(disjoint events)} \\
&= 3P(E_1) && \text{(symmetry)} \\
&= 3(P(ab = 1)P(bc = 1)P(cd = 1)P(da = 1)P(ac = 0)P(bd = 0)) && \text{(independent)} \\
&= 3p^4(1 - p)^2
\end{aligned}$$

Let $\mu = \mathbb{E}[f(S)]$, let N_n denote the number of squares in G_n, then $N_n = \sum_{S \in \mathcal{S}_4} f(S)$ and

$$\begin{aligned}
\mathbb{E}[N_n] &= \mathbb{E}\left[\sum_{S \in \mathcal{S}_4} f(S)\right] \\
&= \sum_{S \in \mathcal{S}_4} \mathbb{E}[f(S)] && \text{(linearity of expectation)} \\
&= |\mathcal{S}_4|\mu = \binom{n}{4} 3p^4(1 - p)^2
\end{aligned}$$

2.4.2 Upper bound on variance of N_n as $n \to \infty$

Now, we bound the variance of N_n, for each $S \in \mathcal{S}_4$, let $\mu = \mathbb{E}[f(S)]$ and $\sigma^2 = \text{Var}(f(S))$

$$\begin{aligned}
\text{Var}(N_n) &= \text{Var}\left(\sum_{S\in\mathcal{S}_4} f(S)\right) \\
&= \mathbb{E}\left[\left(\sum_{S\in\mathcal{S}_4} f(S) - \mathbb{E}\left[\sum_{S\in\mathcal{S}_4} f(S)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{S\in\mathcal{S}_4} f(S) - |\mathcal{S}_4|\mu\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{S\in\mathcal{S}_4} (f(S) - \mu)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{P\in\mathcal{S}_4}\sum_{Q\in\mathcal{S}_4} (f(P)-\mu)(f(Q)-\mu)\right] \\
&= \sum_{P\in\mathcal{S}_4}\sum_{Q\in\mathcal{S}_4} \mathbb{E}[(f(P)-\mu)(f(Q)-\mu)] && \text{(linearity of expectation)} \\
&= \sum_{S\in\mathcal{S}_4} \mathbb{E}[(f(S)-\mu)^2] + \sum_{(P,Q)\in\mathcal{S}_4^2:P\neq Q} \mathbb{E}[(f(P)-\mu)(f(Q)-\mu)] \\
&= |\mathcal{S}_4|\sigma^2 + \sum_{(P,Q)\in\mathcal{S}_4^2:P\neq Q} \mathbb{E}[(f(P)-\mu)(f(Q)-\mu)]
\end{aligned}$$

Note that, if n is large, for each $P \in \mathcal{S}_4$, there are $\binom{4}{2}\binom{n-4}{2} + \binom{4}{3}\binom{n-4}{1} = O(n^2)$ possible choices of $Q \in \mathcal{S}_4$ so that $Q \neq P$ and $f(P)$ and $f(Q)$ are dependent (either P and Q share 2 or 3 vertices). Therefore, there are at most $nO(n^2) = O(n^3)$ dependent pairs of P, Q with $P \neq Q$. For each dependent pair P, Q with $P \neq Q$, by Cauchy–Schwarz

$$\mathbb{E}[(f(P)-\mu)(f(Q)-\mu)] \leq \sqrt{\mathbb{E}[(f(P)-\mu)^2]\mathbb{E}[(f(Q)-\mu)^2]} = \sigma^2$$

Then,

$$\sum_{(P,Q)\in\mathcal{S}_4^2:P\neq Q} \mathbb{E}[(f(P)-\mu)(f(Q)-\mu)] = O(n^3)\sigma^2$$

Note that, $|\mathcal{S}_4| = O(n^4)$, then

$$\text{Var}(N_n) = O(n^4)\sigma^2 + O(n^3)\sigma^2 = O(n^4)\sigma^2$$

which is of the same n-order in the independent case. Therefore, it yields the weak law of large numbers for N_n

2.5 Q5

Tom throws a fair die repeatedly. If Tom throws the die n times, and let N_n denote the number of distinct faces of the die that he has seen, the find the mean and variance of N_n. If Y denote the number of times Tom has to throw the die in order to see each face of the die at least once, then find the mean and variance of Y (*variable name was changed from X to Y*)

2.5.1 mean and variance of N_n

$$N_n = \sum_{i=1}^{6} f\left(\bigcup_{m=1}^{n}\{X_m = i\}\right) = 6 - \sum_{i=1}^{6} f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)$$

where $f(A) = 1$ if event A occurs and 0 otherwise. By linearity of expectation

$$\mathbb{E}[N_n] = 6 - \sum_{i=1}^{6}\mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)\right]$$

for each $i = 1, ..., 6$,

$$f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right) = \begin{cases} 1 & \text{with probability } P(X_1 \neq i, ..., X_n \neq i) \\ 0 & \text{with probability } 1 - P(X_1 \neq i, ..., X_n \neq i) \end{cases}$$

As $P(X_1 \neq i, ..., X_n \neq i) = \left(\frac{5}{6}\right)^n$, then $\mathbb{E}[f(\bigcap_{m=1}^n \{X_m \neq i\})] = \left(\frac{5}{6}\right)^n$, then

$$\mathbb{E}[N_n] = 6 - 6\left(\frac{5}{6}\right)^n$$

As $\mathrm{Var}(N_n) = \mathrm{Var}(6 - N_n)$, we have

$$\begin{aligned}
\mathbb{E}[(6-N_n)^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{6} f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{6}\sum_{j=1}^{6} f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right) f\left(\bigcap_{m=1}^{n}\{X_m \neq j\}\right)\right] \\
&= \sum_{i=1}^{6}\mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)^2\right] + \sum_{i\neq j}\mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right) f\left(\bigcap_{m=1}^{n}\{X_m \neq j\}\right)\right]
\end{aligned}$$

Left term:

$$\mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)^2\right] = \mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right)\right] = \left(\frac{5}{6}\right)^n$$

Right term: as

$$f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right) f\left(\bigcap_{m=1}^{n}\{X_m \neq j\}\right) = \begin{cases} 1 & \text{with probability } P(X_1 \neq i,j, ..., X_n \neq i,j) \\ 1 & \text{with probability } 1 - P(X_1 \neq i,j, ..., X_n \neq i,j) \end{cases}$$

As $P(X_1 \neq i,j, ..., X_n \neq i,j) = \left(\frac{4}{6}\right)^n$, then

$$\mathbb{E}\left[f\left(\bigcap_{m=1}^{n}\{X_m \neq i\}\right) f\left(\bigcap_{m=1}^{n}\{X_m \neq j\}\right)\right] = \left(\frac{4}{6}\right)^n$$

Therefore,

$$\mathbb{E}[(6-N_n)^2] = 6\left(\frac{5}{6}\right)^n + \frac{6\cdot 5}{2}\left(\frac{4}{6}\right)^n$$

$$\mathrm{Var}(N_n) = \mathrm{Var}(6-N_n) = \mathbb{E}[(6-N_n)^2] - \mathbb{E}[6-N_n]^2 = \left(\frac{5}{6}\right)^n + \frac{6\cdot 5}{2}\left(\frac{4}{6}\right)^n - 6^2\left(\frac{5}{6}\right)^{2n}$$

2.5.2 mean and variance of Y

We have $\{Y \leq n\}$ is the event where we see 6 faces in the first n throws. As the distribution is uniform, we will use counting to calculate $P(Y \leq n)$. After the first n throws

- number of outcomes: 6^n
- number of ways to see 6 faces is (number of choice of subsets of size 6) x (permutation of 6) (the rest): $\binom{n}{6}6!6^{n-6}$

Then

$$P(Y \leq n) = \frac{6!\binom{n}{6}6^{n-6}}{6^n} = \frac{6!}{6^6}\binom{n}{6}$$

Then

$$P(Y = n) = P(Y \leq n) - P(Y \leq n-1) = \frac{(n-1)...(n-5)}{6^5}$$

2.6 Q6

We toss a biased coin repeatedly, where the probability of seeing Head equals $\frac{1}{n}$ for some $n \in \mathbb{N}$. Let T_n denote the number of coin tosses needed to see Head for the first time, and let N_n denote the number of Heads among the first n coin tosses.

1. Compute the mean and variance of T_n and N_n respectively.
2. Show that as $n \to \infty$, the distribution of $\frac{T_n}{n}$ converges to the exponential distribution with parameter 1. In other words, show that for each $x \in \mathbb{R}$
$$P\left(\frac{T_n}{n} > x\right) \to P(Z > x) = e^x$$
as $n \to \infty$ where Z is an exponential random variable with parameter 1
3. Show that as $n \to \infty$, the distribution of N_n converges to the Poisson distribution with parameter 1. In other words, show that for each $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$$P(N_n = k) \to P(W = k) = e^{-1}\frac{1}{k!}$$
as $n \to \infty$ where W is a Poisson random variable with parameter 1

Let $p = \frac{1}{n}, q = 1 - p$

2.6.1 mean and variance of T_n and N_n

As $P\left(\bigcap_{i\in\mathbb{N}}\{X_i = T\}\right) = 0$

$$\begin{aligned}
P(T_n = k) &= P(X_1 = T, X_2 = T, ..., X_{k-1} = T, X_k = H) \\
&= P(X_1 = T)P(X_2 = T)...P(X_{k-1} = T)P(X_k = H) \\
&= q^{k-1}p
\end{aligned}$$

T_n follows geometric distribution

Lemma 2 (geometric series)**.** *If $x < 1$, then the series below converges*

$$\begin{aligned}
f(x) &= \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \\
f'(x) &= \sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2} \\
f''(x) &= \sum_{k=0}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}
\end{aligned}$$

Furthermore, we can rewrite

$$f''(x) = \sum_{k=1}^{\infty} k^2x^{k-1} + f'(x)$$

We have

$$\begin{aligned}
\mathbb{E}[T_n] &= \sum_{k=1}^{\infty} kP(T_n = k) \\
&= p\sum_{k=1}^{\infty} kq^{k-1} \\
&= pf'(q) = n
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[T_n^2] &= \sum_{k=1}^{\infty} k^2 P(T_n = k) \\
&= p \sum_{k=1}^{\infty} k^2 q^{k-1} \\
&= p(f''(q) - f'(q)) \\
&= 2n^2 - n
\end{aligned}
$$

$$
\mathrm{Var}(T_n) = \mathbb{E}[T_n^2] - \mathbb{E}[T_n]^2 = n^2 - n
$$

We have

$$
P(N_n = k) = \binom{n}{k} p^k q^{n-k}
$$

N_n follows binomial distribution

Lemma 3 (binomial)**.**

$$
\begin{aligned}
g(n,p,q) &= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} && \textit{(change the range of } k\textit{)} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} && \textit{(}k(n,k) = n(n-1,k-1)\textit{)} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \sum_{l=0}^{m} \binom{m}{l} p^l q^{m-l} \\
&= np && \textit{(Binomial theorem)}
\end{aligned}
$$

We have

$$
\mathbb{E}[N_n] = \sum_{k=0}^{n} k P(N_n = k) = g(n,p,q) = np
$$

$$
\begin{aligned}
\mathbb{E}[N_n^2] &= \sum_{k=0}^{n} k^2 P(N_n = k) \\
&= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} && \text{(change the range of } k\text{)} \\
&= \sum_{k=1}^{n} kn \binom{n-1}{k-1} p^k q^{n-k} && (k(n,k) = n(n-1,k-1)) \\
&= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \left(\sum_{k=1}^{n} (k-1) \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} + \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \right) \\
&= np\big(g(n-1,p,q) + (p+q)^{n-1}\big) \\
&= n^2p^2 - np^2 + np
\end{aligned}
$$

$$\mathrm{Var}(N_n) = \mathbb{E}[N_n^2] - \mathbb{E}[N_n]^2 = np(1-p)$$

2.6.2 T_n/n converges to exponential distribution

$$P(T_n > k) = \sum_{l=k+1}^{\infty} P(T_n = l) = \sum_{l=k+1}^{\infty} q^{l-1}p = \frac{q^k}{1-q}p = q^k$$

For each $x \in \mathbb{R}$

$$P\left(\frac{T_n}{n} > x\right) = P(T_n > \lfloor nx \rfloor) = q^{\lfloor nx \rfloor}$$

Let $m = nx$, then we have

$$\left(1 - \frac{x}{m}\right)^m \leq P\left(\frac{T_n}{n} > x\right) \leq \left(1 - \frac{x}{m}\right)^{m+1}$$

As $n \to \infty$, both side converges to e^{-x}. Hence, $P\left(\frac{T_n}{n} > x\right) \to e^{-x}$

2.6.3 N_n converges to Poisson distribution

$$\begin{aligned}
P(N_n = k) &= \binom{n}{k} p^k q^{n-k} \\
&= \frac{n!}{k!(n-k)!} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{n-k} \\
&= \frac{1}{k!}(1 + o(1)) \left(1 - \frac{1}{n}\right)^{n-k} && (o(1) \to 0 \text{ as } n \to \infty)
\end{aligned}$$

Then,

$$\lim_{n\to\infty} P(N_n = k) = \frac{1}{k!}\left(\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n\right)\left(\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^{-k}\right) = \frac{1}{k!}e^{-1}$$

2.7 Q7

A sequence of biased coins are flipped. The r-th coin has probability θ_r of showing Head where θ_r is a random variable in $[0,1]$. Let S_n be the number of Heads among the first n coin tosses.

1. Assuming that $(\theta_r)_{r\in\mathbb{N}}$ are i.i.d uniformly distributed on $[0,1]$. Is there a centering and scaling of S_n such that the Central Limit Theorem holds? If the answer is yes, then find the centering and scaling constants which lead to the standard normal distribution.
2. What if $(\theta_r)_{r\in\mathbb{N}}$ all equal the same uniform variable θ

2.7.1 θ_r are uniformly distributed on $[0,1]$

Let X_r be a Bernoulli random variable of parameter θ_r for the r-th coin toss. As X_r are i.i.d and

$$S_n = \sum_{r=1}^{n} X_r$$

the question is to find the mean and variance of X_r. Consider the pair of random variable $(X_r, \theta_r) : \Omega \to \{0,1\} \times [0,1]$, for each partition $\{0 = x_0 < x_1 < ... < x_n = 1\}$, we have

$$
\begin{aligned}
P(X_r = 1) &= P(X_r = 1, \theta_r \in [0,1]) \\
&= \sum_{i=1}^{n} P(X_r = 1, \theta_r \in [x_{i-1}, x_i]) \\
&= \sum_{i=1}^{n} P(X_r = 1 | \theta_r \in [x_{i-1}, x_i]) P(\theta_r \in [x_{i-1}, x_i])
\end{aligned}
$$

As $x_{i-1} \leq P(X_r = 1|\theta_r \in [x_{i-1}, x_i]) \leq x_i$ and $P(\theta_r \in [x_{i-1}, x_i])$, the sum is a Riemann sum of function $f : [0,1] \to [0,1]$ defined by $f(x) = x$. For each partition, the value of Riemann sum is fixed, therefore, the Riemann sum equals its limit

$$
P(X_r = 1) = \sum_{i=1}^{n} P(X_r = 1|\theta_r \in [x_{i-1}, x_i])P(\theta_r \in [x_{i-1}, x_i]) = \int_0^1 f dx = \frac{1}{2}
$$

Hence, $\mathbb{E}[X_r] = \frac{1}{2}$ and $\mathrm{Var}(X_r) = \frac{1}{4}$. Scaling of S_n is

$$
\frac{S_n - n\mathbb{E}[X_r]}{\sqrt{n \,\mathrm{Var}(X_r)}}
$$

2.7.2 θ_r is the same for all r

Using the same method as above, we know $P(X_1 = 1) = P(X_2 = 1) = \frac{1}{2}$, for each partition $\{0 = x_0 < x_1 < ... < x_n = 1\}$, we have

$$
P(X_1 = 1, X_2 = 1) = P(X_1 = 1, X_2 = 1, \theta \in [0,1]) = \sum_{i=1}^{n} P(X_1 = 1, X_2 = 1|\theta \in [x_{i-1}, x_i])P(\theta \in [x_{i-1}, x_i])
$$

Similarly, we have $x_{i-1}^2 \leq P(X_1 = 1, X_2 = 1|\theta \in [x_{i-1}, x_i]) \leq x_i^2$, then $P(X_1 = 1, X_2 = 1) = \int_0^1 x^2 dx = \frac{1}{3}$. Hence, X_1 and X_2 are not independent. The central limit theorem on S_n no longer holds.

2.8 Q8

In Lecture 5, we compute the probability of Gambler A's run if A and B start with respectively \$$m$ and \$$n$, and they bet on a fair coin. Suppose now that the coin is biased with probability $p \in (0,1)$ of seeing Head, while A still bets on seeing Head and B bets on seeing Tail. What is the probability of A's ruin as a function of m, n, and p?

Answer.

Let X_t denote the total wealth of A at time t, then $X_0 = m$. Let $L = m + n$, let E denote the event for A's ruin, then

$$
P(E|X_0 = m) = \mathbb{E}[P(E|X_0 = m, X_1)] = pP(E|X_1 = m+1) + (1-p)P(E|X_1 = m-1)
$$

Let $f(k) = P(E|X_0 = k)$, we have the recurrence relation for all $1 \leq k \leq L-1$

$$
f(k) = pf(k+1) + (1-p)f(k-1)
$$

And the boundary conditions: $f(0) = 1, f(L) = 0$. We have the characteristic function

$$
r = pr^2 + (1-p)
$$

If $p = 1/2$, the case is reduced to Lecture 5, if $p \neq 1/2$, the characteristic function has two distinct roots

$$
r_1 = \frac{1 + \sqrt{1 - 4p(p-1)}}{2p} \text{ and } r_2 = \frac{1 - \sqrt{1 - 4p(p-1)}}{2p}
$$

$f(k)$ is of the form

$$
f(k) = Ar_1^k + Br_2^k
$$

Using the boundary conditions we have

$$
\begin{aligned}
1 &= A + B \\
0 &= Ar_1^L + Br_2^L
\end{aligned}
$$

Solve for A, B, we have $P(E) = f(m)$ can be written as a function of m, n and p

$\square$

2.9 Q9

Let $X_1, X_2, ...$ be a sequence of i.i.d Bernoulli random variables with parameter 1/2. We can think of $(X_i)_{i\in\mathbb{N}}$ as indicator random variables for showing Head for a sequence of i.i.d fair coin tosses. Let us construct a random variable Y from $X_1, X_2, ...$ by defining

$$Y = \sum_{n\in\mathbb{N}:X_n=1} \frac{1}{2^n}$$

Show that Y is uniformly distributed on $[0, 1]$ if we treat Y as a binary number $0.X_1X_2...$

Proof. We need to show that $P(Y < y) = y$ for all $y \in (0, 1)$. Let $y = 0.x_1x_2...$ be the binary representation of y, let $y_n = 0.x_1x_2..., x_n = \frac{\lfloor 2^n y \rfloor}{2^n}$ be the number from the first n bits of y. We have $\{Y < y\} = \bigcup_{n\in\mathbb{N}}\{Y < y_n\}$ and

$$\{Y < y_1\} \subseteq \{Y < y_2\} \subseteq ... \subseteq \{Y < y\}$$

Therefore

$$P(Y < y) = \lim_{n\to\infty} P(Y < y_n)$$

Now, we calculate $P(Y < y_n)$. Partition the space of all sequences into 2^n disjoint subsets where two sequences are in the same subset if they have the same first n bits. The construction yields 2^n measurable subsets where each subset has the same probability $\frac{1}{2^n}$ and moreover each subset is either less than y_n or greater than or equal y_n, there are exactly $2^n y_n$ subsets that are less than y_n (for example, if $y_n = 0.10$, then all sequences starts with 0.00 or 0.01 are less than y_n), hence

$$P(Y < y_n) = y_n$$

As $y_n \to y$ as $n \to \infty$,

$$P(Y < y) = \lim_{n\to\infty} P(Y < y_n) = \lim_{n\to\infty} y_n = y$$

$\square$