NCKU 112.1
Probability

Eric Liu

CONTENTS

Chapter 0

Probability Space

0.1 Basic Definition for Background Space

Definition 0.1.1. (Definition of σ-Algebra) We say $(\Omega, \mathcal{G} \subseteq \mathcal{P}(\Omega))$ is a σ-algebra if

$$\Omega \in \mathcal{G}$$

$$X \in \mathcal{G} \implies \Omega \setminus X \in \mathcal{G} \text{ (Closed under complement)}$$

$$\mathcal{A} \subseteq \mathcal{G} \text{ and } |\mathcal{A}| \leq |\mathbb{N}| \implies \bigcup \mathcal{A} \in \mathcal{G} \text{ (Closed under countable union)}$$

From now, we denote $\Omega \setminus X$ by X^c. We say $\mathcal{G} = \{\varnothing, \Omega\}$ is the trivial σ-algebra on Ω.

Theorem 0.1.2. (Basic Property of σ-Algebra) Let $(\Omega, \mathcal{G})$ be a σ-algebra. Then we have

$$\varnothing \in \mathcal{G}$$
$$\mathcal{A} \subseteq \mathcal{G} \implies \bigcap \mathcal{A} \in \mathcal{G}$$
$$A, B \in \mathcal{G} \implies A \setminus B \in \mathcal{G}$$

Proof. Observe $\varnothing = \Omega^c$, and observe $\bigcap \mathcal{A} = (\bigcup_{X \in \mathcal{A}} X^c)^c$, and observe $A \setminus B = A \cup B^c$ ∎

Lemma 0.1.3. (Intersection of σ-Algebras is a σ-Algebra) Let S be a set of σ-algebra over Ω, then $\bigcap S$ is a σ-algebra.

Proof. $\Omega \in \bigcap S$ is trivial. Observe $A \in \bigcap S \implies \forall \mathcal{G} \in S, A \in \mathcal{G} \implies \forall \mathcal{G} \in S, A^c \in \mathcal{G} \implies A^c \in \bigcap S$. Observe $\mathcal{A} \subseteq \bigcap S$ and $|\mathcal{A}| \leq |\mathbb{N}| \implies \forall \mathcal{G} \in S, \mathcal{A} \subseteq \mathcal{G} \implies \forall \mathcal{G} \in S, \bigcup \mathcal{A} \in \mathcal{G} \implies \bigcup A \in \bigcap S$. ∎

Definition 0.1.4. (Definition of Generating σ-Algebra) Let $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. The σ-algebra generated by $\mathcal{F}$ is defined to be the smallest σ-algebra that contain $\mathcal{F}$

We have now gave enough tools to generate σ-algebra. | The f

Definition 0.1.5. (Definition of Measure) Let $\mathcal{G}$ be a σ-algebra over Ω. Function $\mu : \mathcal{G} \to \mathbb{R}$ is called a measure if

$$\forall E \in \mathcal{G}, \mu(E) \geq 0 \text{ (Nonnegative)}$$

$$\mu(\varnothing) = 0$$

$$\mathcal{F} \subseteq \mathcal{G} \text{ and } |\mathcal{F}| \leq |\mathbb{N}| \text{ and } \mathcal{F} \text{ is disjoint} \implies \mu\left(\bigcup \mathcal{F}\right) = \sum_{X \in \mathcal{F}} \mu(X) \text{ (σ additivity)}$$

Theorem 0.1.6. (Intended Property of Measure) We have

$$A \subseteq B \in \mathcal{F} \implies \mu(A) \leq \mu(B)$$

Proof. Observe $\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \implies \mu(B) - \mu(A) = \mu(B \setminus A) \geq 0$ ∎

Definition 0.1.7. (Definition of Probability Measure) A probability measure P is a measure that satisfy $P(\Omega) = 1$

Definition 0.1.8. (Definition of Probability Space) A probability space is a triple $(\Omega, \mathcal{G}, P)$ where

$$\Omega \text{ is a set called } \textit{sample space}$$

$$\mathcal{G} \text{ is a σ-algebra over } \Omega \text{ called event space}$$

$$P : \mathcal{G} \to [0, 1] \text{ is a measure called probability measure}$$

Basically, probability measure "measure" the probability an event will happen. Normally we don't measure the probability a sample will happen, since it is undoable in some cases, for instance, shooting a dart to an infinite set. We only measure the probability that the outcome sample is in a set of certain samples, where the assignment of measure is outside of the system.

A simple example of a σ-algebra is

$$\Omega_2 = \{HH, HT, TH, TT\}, \mathcal{G} = \{\varnothing, X, \{HT, HH\}, \{TH, TT\}\}$$

Notice in this example, Ω is ought to be interpreted as tossing two coins and $\mathcal{G}$ is to observe the first coin is head or tail.

To expand the first example, we have another simple example:

$$\Omega_3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Define

$$A_H := \{HHH, HHT, HTH, HTT\} \text{ and } A_T := \{THH, THT, TTH, TTT\}$$

which is the information of tossing head or tail on first try.

Notice $A_T = A_H^c$. Define

$$A_{HH} := \{HHH, HHT\} \text{ and } A_{HT} := \{HTH, HTT\}$$

$$A_{TH} := \{THH, THT\} \text{ and } A_{TT} := \{TTH, TTT\}$$

so we have

$$A_H = A_{HH} \cup A_{HT} \text{ and } A_T = A_{TH} \cup A_{TT}$$

Then we can define

$$\mathcal{G} := \{\bigcup N : N \in \mathcal{P}(M)\}$$

where $M = \{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$

Notice we can define four σ-algebras by

$$\mathcal{F}_0 = \{\varnothing, \Omega\}, \mathcal{F}_1 = \{\varnothing, \Omega, A_T, A_H\}, \mathcal{F}_2 = \mathcal{G}, \mathcal{F}_3 = \mathcal{P}(\Omega)$$

then we have

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$$

The following concern Borel σ-algebra
$M = \{X^{-1}[x] : x \in R(X)\}$

Definition 0.1.9. (Definition of a Random Variable) We say the function from Ω to $\mathbb{R}$ is a random variable, and often a random variable is defined on a fixed probability space, since the usage of random variable are mostly associated with a probability measure.

We now define 3 random variable for example from the last example of σ-algebra.

Let $S_0, u, d \in \mathbb{R}^+$ and let $d < 1 < u$. We define three random variables S_1, S_2, S_3 on Ω_3

$$S_1(\omega) = \begin{cases} uS_0 & \text{if } \omega \in A_H \\ dS_0 & \text{if } \omega \in A_T \end{cases} \quad S_2(\omega) = \begin{cases} u^2 S_0 & \text{if } \omega \in A_{HH} \\ udS_0 & \text{if } \omega \in A_{HT} \cup A_{TH} \\ d^2 S_0 & \text{if } \omega \in A_{TT} \end{cases}$$

$$S_3(\omega) = \begin{cases} u^3 S_0 & \text{if } \omega \in \{HHH\} \\ u^2 dS_0 & \text{if } \omega \in \{HHT, HTH, THH\} \\ ud^2 S_0 & \text{if } \omega \in \{HTT, THT, TTH\} \\ d^3 S_0 & \text{if } \omega \in \{TTT\} \end{cases}$$

Often, we just use S to denote $S(\omega)$.

Definition 0.1.10. (Definition of Borel-Algebra) The Borel-Algebra on $\mathbb{R}$, which we denote $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by all open interval of $\mathbb{R}$.

Some members of $\mathcal{B}(\mathbb{R})$:
$$(b, a), (a, \infty), \mathbb{R}$$
$$(b, a] = (b, \infty) \setminus (a, \infty)$$
$$[a, \infty) = \mathbb{R} \setminus (-\infty, a)$$
$$[a, b] = [a, \infty) \setminus (b, \infty)$$
$$\{a\} = \mathbb{R} \setminus (-\infty, a) \cup (a, \infty)$$

Theorem 0.1.11. (Construct σ-Algebra with Random Variable) Let X be a random variable on Ω. We define

$$X^{-1}[B] = \{\omega \in \Omega : X(\omega) \in B\}$$

and define the σ-algebra $\sigma(X)$ by

$$\sigma(X) = \{X^{-1}[B] : B \in \mathcal{B}(\mathbb{R})\}$$

We can verify $\sigma(X)$ is a σ-algebra.

Proof. Notice $\Omega = X^{-1}[\mathbb{R}] \in \sigma(X)$. Observe $(X^{-1}[B])^c = X^{-1}[B^c] \in \sigma(X)$. Let $C = \{X^{-1}[B] : B \in \mathcal{C}\} \subseteq \sigma$ be countable. Observe $\bigcup_{B \in C} X^{-1}[B] = X^{-1}[\bigcup_{B \in \mathcal{C}} B] \in \sigma(X)$ ∎

Notice $\sigma(S_1) = \mathcal{F}_1, \sigma(S_2) \neq \mathcal{F}_2, \sigma(S_3) \neq \mathcal{F}_3$.

Normally, we write $P(X = x_i)$ to denote $P(X^{-1}[x_i])$

Theorem 0.1.12. (Induce Measure on Borel Algebra by a Random Variable and a Probability Space) Let $(\Omega, \mathcal{F}, P : \mathcal{F} \to \mathbb{R})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable. We can induce a measure on $\mathcal{B}(\mathbb{R})$ (we say this measure is induced by X and P)

$$\mathcal{L}_X : \mathcal{B}(\mathbb{R}) \to \mathbb{R}, B \mapsto P(X^{-1}[B])$$

Proof. Because P is nonnegative, we know $\mathcal{L}_X$ is nonnegative. Notice that $X^{-1}[\varnothing] = \varnothing$, so we know $\mathcal{L}_X(\varnothing) = P(\varnothing) = 0$. Let $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R})$ and $\mathcal{A}$ be countable and disjoint. We have

$$\begin{aligned}
\mathcal{L}_X(\bigcup \mathcal{A}) &= P(X^{-1}[\bigcup \mathcal{A}]) \\
&= P(\bigcup \mathcal{C}) \text{ where } \mathcal{C} = \{X^{-1}[a] : a \in \mathcal{A}\} \\
&= \sum_{c \in \mathcal{C}} P(c) \\
&= \sum_{a \in \mathcal{A}} P(X^{-1}[a]) = \sum_{a \in \mathcal{A}} \mathcal{L}_X(a)
\end{aligned}$$

Notice that $\mathcal{C}$ is countable because $\mathcal{A}$ is countable, and $\mathcal{C}$ is disjoint because $x \in X^{-1}[a] \cap X^{-1}[b] \implies X(x) \in a \cap b$. ∎

Definition 0.1.13. (Definition of Cumulative Distribution Function) We say $F_X : \mathbb{R} \to \mathbb{R}$ is the cumulative distribution function of X if

$$F_X(x) = \mathcal{L}_X((-\infty, x]) = P(X^{-1}[(-\infty, x]])$$

Distribution function is a function on the image of a random variable X, that map a real number to the probability the random variable X is the real number x_1. How? Probability is the measure of the set that can be mapped to x_1.

In a more formal language, the random variable doesn't specify the probability of each event, but the one does, probability measure is often implicitly given by the random variable.

We now use the random variable $S_3 : \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, P : \mathcal{P}(\Omega) \to \mathbb{R})$ to serve as an example of induced measure and cumulative distribution function.

Notice that the range of S_3 is $\{d^3 S_0 < d^2 u S_0 < du^2 S_0 < u^3 S_0\}$. We have

$$F_{S_3}(d^3 S_0) = \mathcal{L}_{S_3}((-\infty, d^3 S_0]) = P(S_3^{-1}[(-\infty, d^3 S_0]]) = P(\{TTT\})$$

$$F_{S_3}(d^2 u S_0) = \mathcal{L}_{S_3}((-\infty, d^2 u S_0]) = P(S_3^{-1}[(-\infty, d^2 u S_0]]) = P(\{TTT, TTH, THT, HTT\})$$

$$F_{S_3}(du^2 S_0) = \mathcal{L}_{S_3}((-\infty, du^2 S_0]) = P(S_3^{-1}[(-\infty, du^2 S_0]]) = P(\Omega \setminus \{HHH\})$$

$$F_{S_3}(u^3 S_0) = \mathcal{L}_{S_3}((-\infty, u^3 S_0]) = P(S_3^{-1}[(-\infty, u^3 S_0]]) = P(\Omega)$$

With above deduction, we can precisely deduce

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3 S_0 \\ P(\{TTT\}) & \text{if } d^3 S_0 \le x < d^2 u S_0 \\ P(\{TTT, TTH, THT, HTT\}) & \text{if } d^2 u S_0 \le x < du^2 S_0 \\ P(\Omega \setminus \{HHH\}) & \text{if } du^2 S_0 \le x < u^3 S_0 \\ P(\Omega) & \text{if } u^3 S_0 \le x \end{cases}$$

Notice that we can define measure P in multiple ways, and then we will have different induced measures $\mathcal{L}_{S_3}$ and different cumulative distribution function $F_{S_3}(x)$. For instance, one can check that $P : \mathcal{P}(\Omega) = \mathcal{F}_3 \to \mathbb{R}, X \mapsto \frac{|X|}{8}$ is a measure, and $P : \mathcal{F}_3 \to \mathbb{R}, X \mapsto \sum_{x \in X} f(x)$ where f is defined by $\omega_1 \omega_2 \omega_3 \mapsto \frac{2^i}{3^3} : i = |\{\omega_j : \omega_j = T, j = 1, 2, 3\}|$ are all measures, but the former is

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3 S_0 \\ \frac{1}{8} & \text{if } d^3 S_0 \le x < d^2 u S_0 \\ \frac{1}{2} & \text{if } d^2 u S_0 \le x < du^2 S_0 \\ \frac{7}{8} & \text{if } du^2 S_0 \le x < u^3 S_0 \\ 1 & \text{if } u^3 S_0 \le x \end{cases}$$

and the latter is

$$F_{S_3}(x) = \begin{cases} 0 & \text{if } x < d^3 S_0 \\ \frac{8}{27} & \text{if } d^3 S_0 \le x < d^2 u S_0 \\ \frac{8}{27} + \binom{3}{1}\frac{4}{27} & \text{if } d^2 u S_0 \le x < du^2 S_0 \\ \frac{8}{27} + \binom{3}{1}\frac{4}{27} + \binom{3}{2}\frac{2}{27} & \text{if } du^2 S_0 \le x < u^3 S_0 \\ 1 & \text{if } u^3 S_0 \le x \end{cases}$$

0.2 Binomial Asset Pricing Model

Definition 0.2.1. (Definition of Stochastic Process) Given a common probability space $(\Omega, \mathcal{F}, P : \mathcal{F} \to \mathbb{R})$, we say an indexed family of random variables defined on our given probability space is a stochastic process, denoted

$$\{X_t : t \in \Lambda, \omega \in \Omega\}$$

where the index set Λ is called time horizon. So, if someone say X is a stochastic process, we can say X_t is a random variable, and we can say $X(t, \omega) = X_t(\omega) \in \mathbb{R}$.

Additionally, we say Ω is "background space" and $\Lambda \times \mathbb{R}$, where we can draw a trajectory to record the sequence of outcome, is "foreground space" or "configuration space", and we say $\Theta = \{X(t, \omega) : t \in \Lambda, \omega \in \Omega\}$ is the "state space" that contains "state of process"

The "frequency" of a state of the process $X(t, \omega)$ is ?

We now give an example of stochastic process. Let $\Omega = \{H, T\}$ and $\Lambda = [0, 2]$. The stochastic process if the indexed set $\{X_t : t \in \Lambda\}$ of random variables specified by

$$X_t(H) = \sin(2\pi t) \text{ and } X_t(T) = \sin(4\pi t)$$

Then

$$\Theta = [-1, 1]$$

Another example is: Let $\Lambda = \{1, 2, 3, 4, 5\}$ and let $\Omega = \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) : \forall i \in \Lambda, \omega_i \in \{H, T\}\}$. The configuration space is then $\{1, 2, 3, 4, 5\} \times \mathbb{R}$, and we can specify the stochastic process $\{S_t : t \in \Lambda\}$ by

$$S_t((\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)) = u^i d^{t-i} S_0 : i = |\{\omega_j : 1 \leq j \leq t, \omega_j = H\}|$$

Then

$$\Theta = \{u^i d^j S_0 : i, j \in \mathbb{N} \cup \{0\}, 1 \leq i + j \leq 5\}$$

To buy an option from someone means acquiring the right, but not the obligation, to purchase a stock at a predetermined price, K. This comes with an associated cost called the "risk premium" or V_0, which is paid at the time of the transaction.

Notice when we say net earn or profit, we are comparing his money with and without making the deal.

If the stock ends up at price $S_1(\omega)$ lower than K, the buyer won't buy the stock at the price K, so for seller, he profit V_0.

If the stock ends up at price $S_1(\omega)$ higher than K, then the seller net earn: $-(S_1(\omega) - K) + V_0$

Now, we consider the profit we can make for depositing money into bank and get interest. The interest rate is r.

If the stock ends up at price $S_1(\omega)$ lower than K, the buyer won't buy the stock at the price K, so for seller, he profit $(1 + r)V_0$.

If the stock ends up at price $S_1(\omega)$ higher than K, then the seller net earn: $-(S_1(\omega) - K) + (1 + r)V_0$

Now, we wish to lower the risk so we buy Δ_0 share of the stock we short at price S_0 to form a portfolio.

If the stock ends up at price $S_1(\omega)$ lower than K, the buyer won't buy the stock at the price K, so for seller, he profit $(1 + r)V_0 + \Delta_0(S_1(\omega) - (1 + r)S_0)$.

If the stock ends up at price $S_1(\omega)$ higher than K, then the seller net earn: $-(S_1(\omega) - K) + (1 + r)V_0 + \Delta_0(S_1(\omega) - (1 + r)S_0)$

10

From now, we denote the value of the option at time t

$$V_t(\omega) := \max\{S_t(\omega) - K, 0\}$$

So we can say the value of our portfolio $X_t(\omega)$, our action of selling the option and buy the stock, at time 1 is

$$X_1(\omega) = -V_1(\omega) + (1+r)V_0 + \Delta_0(S_1(\omega) - (1+r)S_0)$$

Now we wish, for all ω, we have $X_1(\omega) \geq 0$, by negotiating with the buyer on V_0 and deciding how many share Δ_0 we want to buy.

Notice that if we negotiate V_0 at

$$V_0 := \frac{1}{1+r}\left[\frac{1+r-d}{u-d}V_1(H) + \frac{u-(1+r)}{u-d}V_1(T)\right]$$

and buy share Δ_0

$$\Delta_0 := \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}$$

We see

$$X_1(H) = -V_1(H) + (1+r)\frac{1}{1+r}\left[\frac{1+r-d}{u-d}V_1(H) + \frac{u-(1+r)}{u-d}V_1(T)\right]$$
$$+ \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}(S_1(H) - (1+r)S_0)$$

equals 0 by substituting $S_1(H)$ and $S_1(T)$ with S_0u and S_0d

Chapter 1

Discrete Random Variable

1.1   Basic Discrete Random Variable and Their Expected Values and Variance

Definition 1.1.1. (Definition of Discrete Random Variable and Probability Mass Function) Let $(\Omega, \mathcal{F}, P : \mathcal{F} \to \mathbb{R})$ be a probability space. We say a random variable $X : \Omega \to \mathbb{R}$ is a discrete random variable, if

$$X[\Omega] \text{ if countable}$$

and we define the probability mass function p_X as

$$p_X(r) = P(X = r)$$

In this section, X is always a discrete random variable.

Theorem 1.1.2. (Property of Probability Mass Function) Let $X[\Omega] = \{x_1, \ldots\}$. We have

$$P(a \leq X \leq b) = \sum_{x_i \in [a,b]} p_X(x_i)$$

This reads: The probability X is in $[a, b]$ equals the sum of the probability X is a number in $[a, b]$

Theorem 1.1.3. (Property of Probability Mass Function) We have

$$\sum_{x_i \in \mathbb{R}} p_X(x_i) = 1$$

Now we introduce expected value.

Definition 1.1.4. (Definition of Expected Value) We say the expected value of X is
$$E(X) = \sum x_i p(x_i)$$
If possible, normally we write
$$\mu = E(X)$$

Theorem 1.1.5. (Expected Value) For function $g : \mathbb{R} \to \mathbb{R}$, we have
$$E(g \circ X) = \sum g(x_i) p(x_i)$$

Theorem 1.1.6. (Linearity of Expected Value of Discrete Random Variable) We have
$$E(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 E(X_1) + \alpha_2 E(X_2)$$

We now introduce variance.

Definition 1.1.7. (Definition of Variance) We define variance of X to be
$$\text{Var}(X) := E((X - \mu)^2)$$

If possible, normally we write
$$\sigma(X) := \sqrt{\text{Var}(X)}$$

Theorem 1.1.8. (Calculation of Variance) We have
$$\text{Var}(X) = E(X^2) - \mu^2$$

Proof.

$$
\begin{aligned}
\text{Var}(X) &= E[(X - \mu^2)] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}
$$

$\blacksquare$

Theorem 1.1.9. (Variance) We have
$$\text{Var}[\alpha_1 X_1] = \alpha_1^2 \text{Var}[X_1]$$

Proof.

$$
\begin{aligned}
\text{Var}[\alpha_1 X_1] &= E[(\alpha_1 X_1)^2] - E[\alpha_1 X_1]^2 \\
&= \alpha_1^2 E[X_1^2] - \alpha_1^2 E[X_1]^2 \\
&= \alpha_1^2 (E[X_1^2] - E[X_1]^2) = \alpha_1^2 \text{Var}[X_1]
\end{aligned}
$$

$\blacksquare$

We now introduce Bernoulli random variable.

Definition 1.1.10. (Definition of Bernoulli Random Variable) Let A be an event and $P(A) = p$. We define
$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Theorem 1.1.11. (Properties of Bernoulli Random Variable) We have
$$P(I_A = 1) = p \text{ and } P(I_A = 0) = 1 - p$$
$$E[I_A] = p$$
$$\text{Var}[I_A] = p(1 - p)$$

Proof.
$$\text{Var}[I_A] = E[I_A^2] - E[I_A]^2 = 1^2 p - p^2 = p - p^2 = p(1 - p)$$
∎

Definition 1.1.12. (Definition of Binomial Random Variable) We say the binomial random variable $B_{n,p}$ is the number of success in n independent trials, each of which is a success with probability p. We have

$$P(B_{n,p} = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Theorem 1.1.13. (Properties of Binomial Random Variable) We have
$$E[B_{n,p}] = np \text{ and } \text{Var}[B_{n,p}] = np(1 - p)$$

Proof.
$$E[B_{n,p}] = E[\sum I_A] = \sum E[I_A] = np$$
That of variance is left unproved. ∎

Now we introduce the notion of Poisson random variable.

Definition 1.1.14. (Definition of Poisson Random Variable) For $\lambda > 0$, we define Poisson random variable X_λ by

$$P(X_\lambda = i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

Theorem 1.1.15. (Verification of Poisson Random Variable) Let X be a Poisson random variable with parameter λ. We verify

$$\sum_{k=0}^{\infty} P(X = k) = 1$$

Proof.

$$\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

∎

Theorem 1.1.16. (Expectation Value of Poisson Random Variable) Let X be a Poisson random variable with parameter λ. We have

$$E[X] = \lambda$$

Proof. Compute

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{\infty} k P(x=k) \\
&= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\
&= e^{\lambda} \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{\lambda} \lambda \sum_{u=0}^{\infty} \frac{\lambda^u}{u!} \\
&= \lambda e^{\lambda} e^{-\lambda} = \lambda
\end{aligned}
$$

∎

Theorem 1.1.17. (Variance of Poisson Random Variable) Let X be a Poisson random variable with parameter λ. We have

$$\mathrm{Var}[X] = \lambda$$

Proof. Compute

$$E[X^2] = \sum_{k=0}^{\infty} k^2 P(x = k)$$

$$= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda e^{-\lambda} [\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}]$$

$$= \lambda e^{-\lambda} [\sum_{k=2}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!}]$$

$$= \lambda e^{-\lambda} [\sum_{k=2}^{\infty} \lambda \frac{\lambda^{k-2}}{(k-2)!} + e^{\lambda}]$$

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

$$= \lambda e^{-\lambda} [\lambda e^{\lambda} + e^{\lambda}]$$

$$= \lambda(\lambda + 1)$$

So we have

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

∎

Now we introduce Poisson Approximation of Binomial Random Variable

Theorem 1.1.18. (Poisson Approximation of Binomial Random Variable) Let X be a binomial $B_{n,p}$ and let $\lambda = np$. We have

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

16

Proof. Compute

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} (\frac{\lambda^k}{n^k})(1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k}$$

$$= \frac{\lambda^k}{k!}(1 - \frac{\lambda}{n})^n (\frac{n(n-1)\cdots(n-k+1)}{n^k})(1 - \frac{\lambda}{n})^{-k}$$

So we have

$$\lim_{n\to\infty} P(X = k) = \frac{\lambda^k}{k!} \lim_{n\to\infty} (1 - \frac{\lambda}{n})^n \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \lim_{n\to\infty} (1 - \frac{\lambda}{n})^{-k}$$

$$= \frac{\lambda^k}{k!} e^{-\lambda}(1)(1) = e^{-\lambda}\frac{\lambda^k}{k!}$$

∎

Now we introduce the geometric random variable.

Definition 1.1.19. (Definition of Geometric Random Variable) A geometric random variable counts the number of trials required for the first success in independent trials with success probability p. Let X be a geometric random variable. We have

$$P(X = k) = (1-p)^{k-1}p$$

Theorem 1.1.20. (Verification of Geometric Random Variable) Let X be a geometric random variable with parameter p. We have

$$\sum_{k=1}^{\infty} P(X = k) = 1$$

Proof.

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p$$

$$= p \sum_{k=1}^{\infty} (1-p)^{k-1}$$

$$= p \sum_{k=0}^{\infty} (1-p)^{k}$$

$$= p\left(\frac{1}{p}\right) = 1$$

∎

Theorem 1.1.21. (Expected Value of Geometric Random Variable) Let X be a geometric random variable with parameter p. We have

$$E[X] = \frac{1}{p}$$

Proof. Notice that the trials are independent. We have

$$E[X] = (p)(1) + (1-p)(E[X] + 1)$$

So we have

$$p(E[X]) = p + 1 - p$$

$$\implies E[X] = \frac{1}{p}$$

∎

Theorem 1.1.22. (Variance of Geometric Random Variable) Let X be a geometric random variable with parameter p. We have

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

Proof. Denote $1 - p$ by q. First notice

$$\sum_{k=1}^{\infty} kq^{k-1} p = E[X] = \frac{1}{p}$$

18

Then deduce

$$E[X^2] = \sum_{k=1}^{\infty} k^2 q^{k-1} p$$

$$= \sum_{k=1}^{\infty} (k-1+1)^2 q^{k-1} p$$

$$= \sum_{k=1}^{\infty} (k-1)^2 q^{k-1} p + \sum_{k=1}^{\infty} 2(k-1) q^{k-1} p + \sum_{k=1}^{\infty} q^{k-1} p$$

$$= q \sum_{k=1}^{\infty} (k-1)^2 q^{k-2} p + 2 \sum kq^{k-1} p - \sum_{k=1}^{\infty} q^{k-1} p$$

$$= q \sum_{u=1}^{\infty} u^2 q^{u-1} p + 2E[X] - p \sum_{k=1}^{\infty} q^{k-1}$$

$$= qE[X^2] + 2E[X] - \frac{p}{1-q}$$

Then we can deduce

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

So in summary we have

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

∎

Theorem 1.1.23. (Geometric Random Variable) Let X be a geometric random variable with parameter p. We have

$$P(X > n) = (1-p)^n$$

19

Proof. Let $q = 1 - p$

$$P(X > n) = 1 - \sum_{k=1}^{n} P(X = k)$$
$$= 1 - \sum_{k=1}^{n} q^{k-1} p$$
$$= 1 - (1 - q) \sum_{k=1}^{n} q^{k-1}$$
$$= 1 - \sum_{k=1}^{n} q^{k-1} + \sum_{k=1}^{n} q^{k}$$
$$= 1 - \sum_{k=0}^{n-1} q^{k} + \sum_{k=1}^{n} q^{k}$$
$$= 1 + (q^{n} - 1) = q^{n}$$

∎

Corollary 1.1.24. We have

$$P(X > n + k \text{ while } X > n) = \frac{q^{n+k}}{q^{n}} = q^{k}$$

Theorem 1.1.25. (Summary) We have

A Bernoulli random variable I is a building block
Parameter p is the probability it happens
$E[I] = p$ and $\text{Var}[X] = p - p^{2}$

A Binomial random variable $B_{n,p}$ counts the success of multiple independent
Bernoulli random variable (trials) of parameter p.
Parameter n is the number of trials and parameter p is the success probability.
$E[B_{n,p}] = np$ and $\text{Var}[B_{n,p}] = np(1 - p)$

A Poisson random variable X_λ is a random variable that counts the number of success in a given time unit when it is expected λ amount of success would occur. Parameter λ is the expected amount of success in the given time unit.
$$E[X_\lambda] = \lambda \text{ and } \mathrm{Var}[X_\lambda] = \lambda$$

A geometric random variable X_p is a random variable that counts the counts the number of trials required for the first success. Parameter p is the success rate of each independent trials.
$$E[X_p] = \frac{1}{p} \text{ and } \mathrm{Var}[X_p] = \frac{1-p}{p^2}$$

P measure is specified by the natural condition of experiment (if the coin is fair). X is specified by what do you want to observe in the experiment.

Chapter 2

Continuous Random Variable

2.1   Basic Knowledge

Definition 2.1.1. (Definition of Continuous Random Variable) A continuous random variable $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable such that

$$X[\Omega] \text{ is uncountable}$$

Definition 2.1.2. (Definition of Cumulative Distribution Function) The cumulative distribution function F_X is

$$F_X(x) = P(X \le x)$$

Definition 2.1.3. (Definition of Absolutely Continuous Random Variable) By an absolutely continuous random variable, we mean a continuous random variable $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$F_X \text{ is differentiable everywhere}$$

Definition 2.1.4. (Definition of Probability Density Function) Let X be an absolutely continuous random variable. The probability density function f_X for X is defined to be

$$f_X(x) = F'_X(x)$$

Above is how we begin in our argument, we now develop some simple yet crucial tool. These tool enable us too calculate the probability of X taking value in some Borel set B by taking

$$P(X \in B) = \int_B f_X(s)ds$$

Theorem 2.1.5. (Fundamental Theorem of Calculus) We have

$$P(a \le X \le b) = \int_a^b f_X(s)ds$$

Proof.

$$
\begin{aligned}
P(a \le X \le b) &= P(X \le b) - P(X < a) \\
&= P(X \le b) - P(X \le a) \\
&= F_X(b) - F_X(a) \\
&= \int_a^b F_X'(s)ds \\
&= \int_a^b f_X(x)ds
\end{aligned}
$$

∎

Corollary 2.1.6. (Fundamental Theorem of Calculus) We have

$$P(a \le X) = \int_a^\infty f_X(s)ds$$

Proof. Because measure is countable additive, we have

$$
\begin{aligned}
P(a \le X) &= \lim_{n \to \infty} \sum_{k=0}^n P(a + k \le X < a + k + 1)? \\
&= \lim_{n \to \infty} \sum_{k=0}^n \int_{a+k}^{a+k+1} f_X(s)ds \\
&= \lim_{n \to \infty} \int_a^{a+n+1} f_X(s)ds \\
&= \int_a^\infty f_X(s)ds
\end{aligned}
$$

∎

Corollary 2.1.7. (Fundamental Theorem of Calculus) We have

$$P(X \le a) = \int_{-\infty}^a f_X(s)ds$$

Proof. Because measure is countable additive, we have

$$P(X \leq a) = \lim_{n \to \infty} \sum_{k=0}^{n} P(a - k - 1 < X \leq a - k)$$

$$= \lim_{n \to \infty} \sum_{k=0}^{n} \int_{a-k-1}^{a-k} f_X(s) ds$$

$$= \lim_{n \to \infty} \int_{a-n-1}^{a} f_X(s) ds$$

$$= \int_{-\infty}^{a} f_X(s) ds$$

∎

The definition of expected value and variance of continuous random variable present.

Definition 2.1.8. (Definition of Expected Value) Let X be an absolutely continuous random variable. We define

$$E[X] := \int_{-\infty}^{\infty} s \times f_X(s) ds$$

Definition 2.1.9. (Definition of Variance) Let X be an absolutely continuous random variable. We define

$$\text{Var}[X] = \int_{-\infty}^{\infty} (s - E[X])^2 \times f_X(s) ds$$

Theorem 2.1.10. (Formula for Variance)

$$\text{Var}[X] = E[X^2] - E^2[X]$$

Proof.

$$\text{Var}[X] = \int_{-\infty}^{\infty} (s - E[X])^2 \times f_X(s) ds$$

$$= \int_{-\infty}^{\infty} s^2 \times f_X(s) ds - 2E[X] \int_{-\infty}^{\infty} s \times f_X(s) ds + \int_{-\infty}^{\infty} E^2[X] \times f_X(s) ds$$

$$= E[X^2] - 2E^2[X] + E^2[X]$$

$$= E[X^2] - E^2[X]$$

∎

The below show case when a variable Y depend on a random variable X (thus Y is a random variable): what is the correct way to compute f_Y from f_X. Notice that Theorem 2.1.11 is inspired by Theorem 2.1.12

Theorem 2.1.11. (Example) Define

$$f_X(x) := \begin{cases} 3x^2 & \text{if } x \in [0,1] \\ 0 & \text{if otherwise} \end{cases} \quad \text{and } g(x) := 1 - x^4 \text{ and } Y = g \circ X$$

We don't have

$$f_Y(y) = g^{-1}(f_X(g^{-1}(y)))$$

Proof. We have

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(1 - X^4 \leq y) \\
&= P(1 - y \leq X^4) \\
&= P(X \geq (1 - y)^{\frac{1}{4}}) \\
&= \int_{(1-y)^{\frac{1}{4}}}^{\infty} f_X(s) ds \\
&= \int_{(1-y)^{\frac{1}{4}}}^{1} 3s^2 ds \\
&= \left| s^3 \right|_{s=(1-y)^{\frac{1}{4}}}^{1} \\
&= 1 - (1 - y)^{\frac{3}{4}}
\end{aligned}$$

This give us

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy} 1 - (1 - y)^{\frac{3}{4}} \\
&= \frac{3}{4}(1 - y)^{\frac{-1}{4}}
\end{aligned}$$

While

$$\begin{aligned}
g^{-1}(f_X(g^{-1}(y))) &= g^{-1}(f_X((1 - y)^{\frac{1}{4}})) \\
&= g^{-1}(3(1 - y)^{\frac{1}{2}}) \\
&= (1 - 3(1 - y)^{\frac{1}{2}})^{\frac{1}{4}}
\end{aligned}$$

∎

Theorem 2.1.12. (Example) Define

$$f_X(x) := \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{if elsewhere} \end{cases} \quad \text{and } g(x) := 3x \text{ and } Y = g \circ X$$

Proof. We see

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(3X \le y) \\
&= P(X \le \tfrac{y}{3}) \\
&= F_X(\tfrac{y}{3}) \\
&= \tfrac{y}{3} \text{ for } y \in [0, 3]
\end{aligned}
$$

Then

$$
f_Y(y) = \begin{cases} \frac{1}{3} & \text{if } y \in [0, 3] \\ 0 & \text{if otherwise} \end{cases}
$$

And

$$
g^{-1}(f_X(g^{-1}(y))) = g^{-1} f_X(\tfrac{y}{3})) = g^{-1}(1) = \frac{1}{3}
$$

■

2.2 Law of Unconscious Statistician

Theorem 2.2.1. (Law of Unconscious Statistician) Let $Y = g(X)$. If X is discrete, we have

$$E[Y] = \sum_{x \in X[\Omega]} g(x)p_X(x)$$

If X is continuous, we have

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

Proof. We only prove when X is continuous and g is differentiable and monotone and g^{-1} exists and is differentiable and monotone.

First notice

$$x = g^{-1}(y) \implies \frac{dx}{dy} = \frac{d}{dy}g^{-1}(y) \implies dx = \frac{d}{dy}g^{-1}(y)dy$$

Notice

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(g(X) \leq y) \\
&= P(X \leq g^{-1}(y)) \\
&= F_X(g^{-1}(y))
\end{aligned}$$

Then

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \frac{d}{dy}F_X(g^{-1}(y)) \\
&= \frac{d}{dg^{-1}(y)}F_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} \\
&= f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}
\end{aligned}$$

We now have

$$\begin{aligned}
\int_{-\infty}^{\infty} yf_Y(y)dy &= \int_{-\infty}^{\infty} yf_X(g^{-1}(y))dg^{-1}(y) \\
&= \int_{-\infty}^{\infty} g(x)f_X(x)dx
\end{aligned}$$

∎

2.3 Uniform, Exponential and Normal

We now go into each continuous random variable. Start with the easiest, uniform random variable.

Definition 2.3.1. (Definition of Uniform Random Variable) We say an absolutely continuous random variable $X \sim \text{Uni}(\alpha, \beta)$ is uniform if there exists $\alpha < \beta \in \mathbb{R}$ such that

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{if elsewhere} \end{cases}$$

Theorem 2.3.2. (Expected Value and Variance of Uniform Random Variable) Let $X \sim \text{Uni}(\alpha, \beta)$. We have

$$E[X] = \frac{\alpha + \beta}{2} \text{ and } \text{Var}[X] = \frac{(\beta - \alpha)^2}{12}$$

Proof. First verify for

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dx = \frac{\beta - \alpha}{\beta - \alpha} = 1$$

Then

$$\begin{aligned} E[X] &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{1}{\beta - \alpha} \times \frac{x^2}{2}\Big|_{x=\alpha}^{\beta} \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

Then

$$E[X^2] = \int_\alpha^\beta \frac{x^2}{\beta - \alpha} dx$$
$$= \frac{1}{\beta - \alpha} \times \left.\frac{x^3}{3}\right|_{x=\alpha}^\beta$$
$$= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)}$$
$$= \frac{(\beta - \alpha)(\beta^2 + \beta\alpha + \alpha^2)}{3(\beta - \alpha)}$$
$$= \frac{\beta^2 + \beta\alpha + \alpha^2}{3}$$

Then

$$\mathrm{Var}[X] = E[X^2] - E^2[X]$$
$$= \frac{\beta^2 + \beta\alpha + \alpha^2}{3} - \frac{\beta^2 + 2\beta\alpha + \alpha^2}{4}$$
$$= \frac{\beta^2 - 2\beta\alpha + \alpha^2}{12}$$
$$= \frac{(\beta - \alpha)^2}{12}$$

∎

Second, exponential.

Definition 2.3.3. (Definition of Exponential Random Variable) We say an absolute continuous random variable $X \sim \mathrm{Exp}(\lambda)$ is exponential if there exists $\lambda \in \mathbb{R}^+$ such that

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

Theorem 2.3.4. (Expected Value and Variance of Exponential Random Variable) Let $X \sim \mathrm{Exp}(\lambda)$. We have

Proof. Fist, verify for

$$\int_{-\infty}^\infty f_X(x) dx = \int_0^\infty \lambda e^{-\lambda x} dx$$
$$= \left. -e^{-\lambda x} \right|_{x=0}^\infty$$
$$= -0 - (-1)$$
$$= 1$$

29

We have

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x}dx$$

$$= -xe^{-\lambda x}\Big|_{x=0}^\infty - \int_0^\infty -e^{-\lambda x}dx$$

$$= \int_0^\infty e^{-\lambda x}dx$$

$$= \frac{e^{-\lambda x}}{-\lambda}\Big|_{x=0}^\infty$$

$$= 0 - (\frac{1}{-\lambda})$$

$$= \frac{1}{\lambda}$$

And have

$$E[X^2] = \int_0^\infty x^2\lambda e^{-\lambda x}dx$$

$$= -x^2 e^{-\lambda x}\Big|_{x=0}^\infty - \int_0^\infty -2xe^{-\lambda x}dx$$

$$= 2\int_0^\infty xe^{-\lambda x}dx$$

$$= 2(\frac{-1}{\lambda})xe^{-\lambda x}\Big|_{x=0}^\infty - 2\int_0^\infty (\frac{-1}{\lambda})e^{-\lambda x}dx$$

$$= \frac{2}{\lambda}\int_0^\infty e^{-\lambda x}dx$$

$$= \frac{2}{\lambda}(\frac{-1}{\lambda})e^{-\lambda x}\Big|_{x=0}^\infty$$

$$= (\frac{-2}{\lambda^2})(0-1) = \frac{2}{\lambda^2}$$

Then we have

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

$\blacksquare$

Theorem 2.3.5. (Property of Exponential Random Variable) Let $X \sim \mathrm{Exp}(\lambda)$.
We see

$$P(X \geq a) = e^{-a\lambda}$$

Proof.

$$P(X \geq a) = \int_a^\infty \lambda e^{-\lambda x} dx$$

$$= -e^{-\lambda x}\Big|_{x=a}^\infty$$

$$= e^{-a\lambda}$$

∎

Theorem 2.3.6. (Memory-less Property of Exponential Random Variable) Let $X \sim \text{Exp}(\lambda)$. We see

$$P(X \geq a + b \mid X \geq b) = \frac{e^{-(a+b)\lambda}}{e^{-b\lambda}} = e^{-a\lambda}$$

If $X \sim \text{Exp}(\lambda)$, it means the measurement of waiting time of an occurrence of event that is expected to happen after λ time period.

Before we start to talk about normal distribution. We first have to talk about Gauss Integral.

Theorem 2.3.7. (Gauss Integral)

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}$$

Proof. Let $I = \int_{-\infty}^\infty e^{-x^2} dx$. We have

$$I^2 = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x^2+y^2)} dx dy$$

$$= \int_0^\infty \int_0^{2\pi} r e^{-r^2} d\theta dr$$

$$= 2\pi \int_0^\infty r e^{-r^2} dr$$

$$= 2\pi \frac{1}{-2} e^{-r^2}\Big|_{r=0}^\infty$$

$$= \pi$$

Then

$$I^2 = \pi \implies I = \sqrt{\pi}$$

∎

Theorem 2.3.8. (Gauss Integral)

$$\int_{-\infty}^{\infty} e^{\frac{-(x-a)^2}{b}} dx = \sqrt{b\pi}$$

Proof. Let $x = a + \sqrt{b}y$. We have

$$\frac{dx}{dy} = \sqrt{b}$$

Then we have

$$\int_{-\infty}^{\infty} e^{\frac{-(x-a)^2}{b}} dx = \int_{-\infty}^{\infty} e^{-y^2} \sqrt{b} dy$$
$$= \sqrt{b\pi}$$

∎

Theorem 2.3.9. (Gamma Function) For $z \in \mathbb{R}$, let

$$\Gamma(z) = \int_0^{\infty} t^z e^{-t} dt$$

Now we start to discuss normal distribution.

Definition 2.3.10. (Definition of Normal Distribution) A normal random variable $X \sim N(\mu, \sigma^2)$ is defined by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Theorem 2.3.11. (Expected Value and Variance of Normal Random Variable) Let $X \sim N(\mu, \sigma^2)$. We have

$$E[X] = \mu \text{ and } \mathrm{Var}[X] = \sigma^2$$

Proof. We first verify

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

Observe

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

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

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

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2\sigma^2}} dz + \mu$$

$$= \mu$$

Where

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2\sigma^2}} dz = 0 \text{ because } z e^{\frac{-z^2}{2\sigma^2}} \text{ is odd function}$$

Observe

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

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

$$= \int_{-\infty}^{\infty} z^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-z^2}{2\sigma^2}} dz + \mu^2$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} s^2 e^{-s^2} ds + \mu^2 \text{ where } s = \frac{z}{\sqrt{2}\sigma}$$

$$= 4\frac{\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} s^2 e^{-s^2} ds + \mu^2$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} t^{\frac{1}{2}} e^{-t} dt + \mu^2 \text{ where } t = s^2$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} \Gamma(\frac{3}{2}) + \mu^2$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} (\frac{\sqrt{\pi}}{2}) + \mu^2$$

$$= \sigma^2 + \mu^2$$

Then we have

$$\text{Var}[X] = E[X^2] - E^[X] = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

∎

33

Now, we introduce how to do linear change of variable for normal distribution.

Theorem 2.3.12. (Linear Change of Normal Random Variable) Let

$$X \sim N(\mu, \sigma^2) \text{ and } Y = \alpha X + \beta \text{ where } \alpha > 0$$

We have

$$Y \sim N(\alpha\mu + \beta, (\alpha\sigma)^2)$$

Proof. Observe

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(\alpha X + \beta \leq y) \\
&= P(X \leq \frac{y - \beta}{\alpha}) \\
&= \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt
\end{aligned}
$$

This give us

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d\frac{y-\beta}{\alpha}}{dy} \frac{d}{d\frac{y-\beta}{\alpha}} \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \\
&= \frac{1}{\alpha} \times \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\frac{y-\beta}{\alpha}-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{\frac{-(y-\beta-\alpha\mu)^2}{2\alpha^2\sigma^2}}
\end{aligned}
$$

This give us

$$Y \sim N(\alpha\mu + \beta, (\alpha\sigma)^2)$$

∎

Corollary 2.3.13.

$$X \sim N(\mu, \sigma^2) \text{ and } Z = \frac{X - \mu}{\sigma} \implies Z \sim N(0, 1)$$

Definition 2.3.14. (Definition of Standard Normal Random Variable) Let $Z \sim N(0,1)$. We denote

$$\Phi(z) := F_Z(z)$$

Thus

$$\Phi(z) = F_Z(z) = \int_{-\infty}^{z} \frac{1}{1\sqrt{2\pi}} e^{\frac{-(s-0)^2}{21^2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-s^2}{2}} ds$$

Theorem 2.3.15. (Probability of Normal Variable within Confidence Level) Let $X \sim N(\mu, \sigma^2)$. We have

$$P(X \in [\mu - \sigma, \mu + \sigma]) = 0.3173$$
$$P(X \in [\mu - 2\sigma, \mu + 2\sigma]) = 0.0455$$
$$P(X \in [\mu - 3\sigma, \mu + 3\sigma]) = 0.0027$$

Proof. Let $Z = \frac{X - \mu}{\sigma}$. We have $Z \sim N(0, 1)$. This give us

$$P(X \in [\mu - \sigma, \mu + \sigma]) = P(Z \in [-1, 1]) = 0.3173$$
$$P(X \in [\mu - 2\sigma, \mu + 2\sigma]) = P(Z \in [-2, 2]) = 0.0455$$
$$P(X \in [\mu - \sigma, \mu + \sigma]) = P(Z \in [-3, 3]) = 0.0027$$

∎

Theorem 2.3.16. (Example) Let

$$X \sim N(2, 25)$$

Compute

$$P(1 \leq X \leq 4)$$

Proof.

$$\mu = 2 \text{ and } \sigma = 5$$

∎

2.4 Exercises

Question 1

A "dyadic rational number" is a real number of the form $\frac{m}{2^k}$ where k and m are integers. Suppose we set $p = q = \frac{1}{2}$ in the construction for a probability measure on Ω_∞ and $X(\omega) = \sum_{i=1}^{\infty} \frac{Y_i(\omega)}{2^i}$ is a random variable on Ω_∞.

- Show that, the induced measure $\mathcal{L}_X$ by the random variable X on Ω satisfies that, for any positive integers k and m such that $0 \leq \frac{m-1}{2^k} < \frac{m}{2^k} \leq 1$, we have

$$\mathcal{L}_X \left[\frac{m-1}{2^k}, \frac{m}{2^k} \right] = \frac{1}{2^k}.$$

  In other words, the induced measure $\mathcal{L}_X$ on all intervals in $[0,1]$ whose endpoints are dyadic rational numbers is the same as the Lebesgue measure of these intervals. The only possible way is that $\mathcal{L}_X$ is indeed the Lebesgue measure.

- Show that, in this case ($p = \frac{1}{2}$), the random variable $X(\omega) = \sum_{k=1}^{\infty} \frac{Y_k(\omega)}{2^k}$ is uniformly distributed on $[0,1]$.

Proof. **(Part 1)** First notice

$$\mathcal{L}_X \left[\frac{m-1}{2^k}, \frac{m}{2^k} \right] = P \left(X \in \left[\frac{m-1}{2^k}, \frac{m}{2^k} \right] \right)$$
$$= P \left((2^k) X \in [m-1, m] \right)$$

Then we express X in the form

$$X(\omega) = \sum_{i=1}^{\infty} \frac{Y_i(\omega)}{2^i} = \sum_{i=1}^{k} \frac{Y_i(\omega)}{2^i} + \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i}$$

Notice that because $Y_i(\omega) \in \{0, 1\}$, we can compute

$$\sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in \left[0, \frac{1}{2^k} \right]$$

Then

$$2^k X(\omega) = \sum_{i=1}^{k} \left(2^{k-i}\right) Y_i(\omega) + 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \text{ where } 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in [0,1]$$

$$\text{and } \sum_{i=1}^{k} \left(2^{k-i} Y_i(\omega)\right) \in \mathbb{Z}$$

Notice that we have

$$P\left(\sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} = 0\right) = 0$$

Because of such, we can deduce

$$P\left(\left(2^k\right) X \in [m-1, m]\right) = P\left(\sum_{i=1}^{k} \left(2^{k-i}\right) Y_i(\omega) + 2^k \sum_{i=k+1}^{\infty} \frac{Y_i(\omega)}{2^i} \in [m-1, m]\right)$$

$$= P\left(\sum_{i=1}^{k} \left(2^{k-i}\right) Y_i(\omega) = m-1\right)$$

Let $N = m - 1$. Express N by binary

$$N = \sum_{j=0}^{n} a_j 2^j \text{ where } a_j \in \{0, 1\} \text{ and } a_n = 1$$

That is

$$N = a_n 2^n + a_{n-1} 2^{n-1} + \cdots + a_1 2^1 + a_0 2^0$$

The premise of the problem include $\frac{m}{2^k} \leq 1$, so we have $n < k$. Then we can express N by

$$N = a_{k-1} 2^{k-1} + a_{k-2} 2^{k-2} + \cdots + a_0 2^0$$

Then we see

$$P\left(\sum_{i=1}^{k} \left(2^{k-i}\right) Y_i(\omega) = m-1\right) = P\left(\sum_{i=1}^{k} \left(2^{k-i}\right) Y_i(\omega) = N\right)$$

$$= P\left(\forall i : 1 \leq i \leq k, Y_i(\omega) = a_{k-i}\right)$$

$$= \prod_{i=1}^{k} P\left(Y_i(\omega) = a_{k-i}\right)$$

$$= \prod_{i=1}^{k} \frac{1}{2} = \frac{1}{2^k}$$

37

(Part 2) Arbitrarily select an closed interval $[a, b]$ contained by $[0, 1]$, we only wish to prove

$$\mathcal{L}_X [a, b] = b - a$$

Express a, b by

$$a = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \text{ and } b = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$$

Define

$$A_k = \sum_{n=1}^{k} \frac{a_n}{2^n} \text{ and } B_k = \sum_{n=1}^{k} \frac{b_n}{2^n} \text{ where } \{a_n : n \in \mathbb{N}\} \cup \{b_n : n \in \mathbb{N}\} \subseteq \{0, 1\}$$

We have

$$\lim_{k \to \infty} A_k = a \text{ and } \lim_{k \to \infty} B_k = b$$

Express A_k, B_k by

$$A_k = \frac{\sum_{n=1}^{k} \left(2^{k-n}\right) a_n}{2^k} \text{ and } B_k = \frac{\sum_{n=1}^{k} \left(2^{k-n}\right) b_n}{2^k}$$

Define $A'_k := \sum_{n=1}^{k} \left(2^{k-n}\right) a_n$ and $B'_k = \sum_{n=1}^{k} \left(2^{k-n}\right) b_n$. It is easy to check $A'_k, B'_k \in \mathbb{Z} \cap \left[0, 2^k\right]$. Then by the result in part 1, we have

$$\forall k, P\left(X \in [A_k, B_k]\right) = P\left(X \in \left[\frac{A'_k}{2^k}, \frac{B'_k}{2^k}\right]\right)$$
$$= B_k - A_k$$

Then we see

$$\mathcal{L}_X [a, b] = P\left(X \in [a, b]\right)$$
$$= \lim_{k \to \infty} P\left(X \in [A_k, B_k]\right)$$
$$= \lim_{k \to \infty} B_k - A_k$$
$$= b - a$$

∎

2.5 Central Limit Theorem

Theorem 2.5.1. (Central Limit Theorem) Let X be an absolutely continuous random variable with expectation value μ and variance σ^2. Define

$$Z_n := \frac{X_1 + \cdots + X_n - \mu n}{\sigma \sqrt{n}}$$

We have

$$\forall x, \lim_{n \to \infty} f_{Z_n}(x) = f_X(x)$$

Corollary 2.5.2. (De Moivre-Laplace Central Limit Theorem) Fix p. Suppose $S_n \sim \text{Bin}(n, p)$. Define $\mu_n := E[S_n] = np$ and $\sigma_n := \sqrt{\text{Var}[S_n]} = \sqrt{np(1-p)}$. Define $Z_n := \frac{S_n - \mu_n}{\sigma_n}$. We have

$$\forall k \in \mathbb{N}, \lim_{n \to \infty} \frac{f_{S_n}(k)}{f_{Z_n}(k)} = 1$$

In other words

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{\frac{-(k-np)^2}{2npq}}$$

Corollary 2.5.3. (De Moivre-Laplace Central Limit Theorem) Recall

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

Let

$$S_n \sim \text{Bin}(n, p)$$

Fix p, we have

$$\forall x \in \mathbb{R}, \lim_{n \to \infty} P(\frac{S_n - \mu_n}{\sigma_n} \leq x) = \Phi(x)$$

where

$$\mu_n = E[S_n] = np \text{ and } \sigma_n = \sqrt{\text{Var}[S_n]} = \sqrt{np(1-p)}$$

Then very weakly, we have

$$\lim_{n \to \infty} \frac{S_n - \mu_n}{\sigma_n} \sim N(0, 1)$$

Corollary 2.5.4. (Approximation of Large Binomial R.V.) Let $S_n \sim \text{Bin}(n, p)$. If n is large enough, we can approximate

$$P(m_1 \leq S_n \leq m_2) = P\left(\frac{m_1 - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}}\right)$$

$$\approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$$

$$= \int_{\frac{m_1 - np}{\sqrt{npq}}}^{\frac{m_2 - np}{\sqrt{npq}}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

And

$$P(S_n = m) \approx \frac{1}{\sqrt{2\pi npq}} e^{\frac{-(m - np)^2}{2npq}}$$

> **Question 2**
>
> A die is threw 12000 times. What is the probability that there will be exactly 1800 rolls of 6? What is the probability that the number of 6's lies in the interval $[1950, 2100]$

Proof. We have $n = 12000$ and $p = \frac{1}{6}$. Let $S_n \sim \text{Bin}(n, p)$. We have

$$P(S_n = 1800) \approx \frac{1}{\sqrt{2\pi \times 12000 \times \frac{5}{36}}} e^{\frac{-(1800 - 2000)^2}{2 \times 12000 \times \frac{5}{36}}} \sim 6.004 \times 10^{-8}$$

Notice that the direct computation of $P(S_n = 1800)$ is $\binom{12000}{1800} \left(\frac{1}{6}\right)^{1800} \left(\frac{5}{6}\right)^{10200}$, in which the computation of $\binom{12000}{1800}$ is difficult even for computer.

Again, notice that the Poisson approximation $V \sim \text{Poisson}(2000)$ of $P(V = 1800)$ is $\frac{e^{-2000} 2000^{1800}}{1800!}$, in which the computation of 1800! is difficult even for computer.

Let $\sigma = \sqrt{npq}$ and $\mu = np$. We have

$$P(1950 \leq S_n \leq 2100) \approx \int_{\frac{1950 - \mu}{\sigma}}^{\frac{2100 - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

$$= \Phi\left(\frac{2100 - \mu}{\sigma}\right) - \Phi\left(\frac{1950 - \mu}{\sigma}\right)$$

$$\approx 0.8825$$

$\blacksquare$

Chapter 3

Independence

3.1 Independence of Event

Definition 3.1.1. (Definition of Independence of Events) Let A, B be two events in $\mathcal{F}$. We define

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

We say A is independent of B if

$$P(A|B) = P(A)$$

Theorem 3.1.2. (Independence)

$$A \text{ is independent of } B \iff P(A \cap B) = P(A)P(B)$$
$$\iff B \text{ is independent of } A$$

Proof.

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

$\blacksquare$

Corollary 3.1.3.

$$A, A^c \text{ are independent} \iff P(A) = 0 \text{ or } 1$$
$$\text{Every event is independent of } \Omega \text{ and } \varnothing$$

Corollary 3.1.4. (Independence)

$$
\begin{aligned}
A \text{ is independent of } B \iff & A^c \text{ is independent of } B \\
\iff & A \text{ is independent of } B^c \\
\iff & A^c \text{ is independent of } B^c
\end{aligned}
$$

Proof. We have

$$
\begin{aligned}
P\left(A^c \cap B\right) &= P\left(B \setminus A\right) \\
&= P\left(B\right) - P\left(A \cap B\right)
\end{aligned}
$$

and

$$
P\left(A^c\right) = 1 - P\left(A\right)
$$

Then we have

$$
P\left(A^c\right) P\left(B\right) = P\left(B\right) - P\left(A\right) P\left(B\right)
$$

This tell us

$$
P\left(A^c \cap B\right) = P\left(A^c\right) P\left(B\right) \iff A, B \text{ are independent}
$$

∎

Corollary 3.1.5.

$$
A, B \text{ are independent} \iff \forall A' \in \sigma\left(A\right), \forall B' \in \sigma\left(B\right), A', B' \text{ are independent}
$$

Definition 3.1.6. (Independence of Sub-σ-Algebra) We say two sub-σ-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ is independent if

$$
\forall A \in \mathcal{G}, \forall B \in \mathcal{H}, A, B \text{ are independent}
$$

Corollary 3.1.7. Given two event A, B

$$
A, B \text{ are independent} \iff \sigma\left(A\right) \text{ and } \sigma\left(B\right) \text{ are independent}
$$

Question 3

Toss a coin twice with $\Omega_2 = \{HH, HT, TH, TT\}$ and $\mathcal{F}_2$ (with $|\mathcal{F}_2| = 16$) is the σ-algebra consisting of all information up to time 2.

Let the probability of tossing a H in the first trial be $p_1 \in (0, 1)$ and that for the second trial be $p_2 \in (0, 1)$.

Define the probability P as

$$
P\left(HH\right) = p_1 p_2, P\left(HT\right) = p_1 \left(1 - p_2\right) \text{ vice versa}
$$

Then,
$$(\Omega_2, \mathcal{F}_2, P) \text{ is a probability space}$$

Let
$$\mathcal{G} = \{\varnothing, \Omega_2, \{HH, HT\}, \{TH, TT\}\} \subset \mathcal{F}_2$$

be the sub-σ-algebra consisting of information of the first toss; while
$$\mathcal{H} = \{\varnothing, \Omega_2, \{HH, TH\}, \{HT, TT\}\} \subset \mathcal{F}_2$$

be the sub-σ-algebra consisting of information of the second toss.

Show that
$$p_1 = \frac{1}{2} = p_2 \implies \mathcal{G}, \mathcal{H} \text{ are independent}$$

and that
$$P(HH) = \frac{1}{9}, P(HT) = \frac{2}{9}, P(TH) = \frac{1}{3}, P(TT) = \frac{1}{3} \implies \mathcal{G}, \mathcal{H} \text{ are dependent}$$

Proof. Let $A = \{HH, HT\}$ and $B = \{HH, TH\}$. We see
$$\mathcal{G} = \{\varnothing, \Omega_2, A, A^c\} \text{ and } \mathcal{H} = \{\varnothing, \Omega_2, B, B^c\}$$

And
$$A \cap B = \{HH\}$$

Then if $p_1 = \frac{1}{2} = p_2$, we see
$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A)P(B)$$

If $P(HH) = \frac{1}{9}, P(HT) = \frac{2}{9}, P(TH) = \frac{1}{3}, P(TT) = \frac{1}{3}$, we see
$$P(A \cap B) = \frac{1}{9} \neq \frac{1}{3} \times \frac{4}{9} = P(A)P(B)$$

$\blacksquare$

3.2 Independence of Random Variable

Definition 3.2.1. (σ-Algebra Generated by Random Variable) The σ-algebra $\sigma(X)$ generated by random variable X is defined by

$$\sigma(X) := \{X^{-1}[B] : B \in \mathcal{B}(\mathbb{R})\}$$

Definition 3.2.2. (Definition of Independence of Random Variable) Given two random variable X, Y on the same probability space $(\Omega, \mathcal{F})$, we say X, Y are independent if

$$\sigma(X) \text{ and } \sigma(Y) \text{ are independent}$$

Lemma 3.2.3. Suppose

$$A \text{ is independent of } C \text{ and } B \text{ is independent of } C \text{ and } A \cap B = \emptyset$$

Then
$$A \cup B \text{ is independent of } C$$

Proof.

$$
\begin{aligned}
P((A \cup B) \cap C) &= P((A \cap C) \cup (B \cap C)) \\
&= P(A \cap C) + P(B \cap C) \\
&= P(A)P(C) + P(B)P(C) \\
&= (P(A) + P(B))P(C) \\
&= P(A \cup B)P(C)
\end{aligned}
$$

■

Question 4

Show that in the probability space of the binomial asset pricing model with three independent tosses of the same coin, the two random variable S_2 and $\frac{S_3}{S_2}$ are independent.

Proof. The σ-algebra $\sigma(S_2)$ generated by S_2 is $\mathcal{F}_2$

$$\sigma(S_2) = \{\bigcup B : B \subseteq \{A_{HH}, A_{TT}, A_{HT} \cup A_{TH}\}\}$$

The σ-algebra $\sigma\left(\frac{S_3}{S_2}\right)$ generated by $\frac{S_3}{S_2}$ is

$$\sigma\left(\frac{S_3}{S_2}\right) = \{\emptyset, \Omega, A_{..H}, A_{..T}\}$$

44

Observe

$$P\left(A_{HH} \cap A_{..H}\right) = P\left(\{HHH\}\right) = \frac{1}{p^3} = \frac{1}{p^2} \times \frac{1}{p} = P\left(A_{HH}\right) P\left(A_{..H}\right)$$

and

$$P\left(A_{TT} \cap A_{..H}\right) = P\left(\{TTH\}\right) = \frac{1}{q^2 p} = \frac{1}{q^2} \times \frac{1}{p} = P\left(A_{TT}\right) P\left(A_{..H}\right)$$

Then by corollary 3.1.3 and lemma 3.2.4, we see

$$\sigma\left(S_2\right), \sigma\left(\frac{S_3}{S_2}\right) \text{ are independent}$$

∎

Theorem 3.2.4. (Change of Independent Random Variable) Given

(a) X, Y are independent

(b) g, h are measurable.

Then

$$g\left(X\right) \text{ and } h\left(Y\right) \text{ are independent}$$

Proof. This is an immediate consequence of

$$\sigma\left(g\left(X\right)\right) \subseteq \sigma\left(X\right) \text{ and } \sigma\left(h\left(Y\right)\right) \subseteq \sigma\left(Y\right)$$

∎

An example for Theorem 3.2.4 is $X\left(H\right) = 1$ and $X\left(T\right) = -1$. We see

$$\sigma\left(X\right) = \{\varnothing, \{H\}, \{T\}, \Omega\}$$

and

$$\sigma\left(X^2\right) = \{\varnothing, \Omega\} \subset \sigma\left(X\right)$$

Theorem 3.2.5. (Joint Distribution of Independent Random Variable)

$$X \text{ and } Y \text{ are independent}$$
$$\Longleftrightarrow \forall A, B \in \mathcal{B}(\mathbb{R}), P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$$

Proof. ($\longleftarrow$)

Let $V \in \sigma(X)$ and $W \in \sigma(Y)$. We wish to prove

$$P(V \cap W) = P(V)P(W)$$

45

Because $V \in \sigma(X)$ and $W \in \sigma(Y)$, we know there exists $A, B \in \mathcal{B}(\mathbb{R})$ such that

$$V = X^{-1}[A] \text{ and } W = Y^{-1}[B]$$

Then

$$\begin{aligned}
P(V \cap W) &= P(X^{-1}[A] \cap Y^{-1}[B]) \\
&= P(\{\omega \in \Omega : X \in A \text{ and } Y \in B\}) \\
&= P(X \in A \text{ and } Y \in B) \\
&= P(X \in A)P(Y \in B) \\
&= P(X^{-1}[A])P(Y^{-1}[B]) \\
&= P(V)P(W) \text{ (done)}
\end{aligned}$$

$(\longrightarrow)$

Let $A, B \in \mathcal{B}(\mathbb{R})$. We wish to prove

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$$

In other words, we wish to prove

$$P(\{\omega \in \Omega : X \in A \text{ and } Y \in B\}) = P(\{\omega \in \Omega : X \in A\})P(\{\omega \in \Omega : Y \in B\})$$

Notice

$$\{\omega \in \Omega : X \in A \text{ and } Y \in B\} = X^{-1}[A] \cap Y^{-1}[B]$$

and notice

$$\{\omega \in \Omega : X \in A\} = X^{-1}[A] \in \sigma(X) \text{ and } \{\omega \in \Omega : Y \in B\} = Y^{-1}[B] \in \sigma(Y)$$

Then because X, Y are independent, we now have

$$\begin{aligned}
P(\{\omega \in \Omega : X \in A \text{ and } Y \in B\}) &= P(X^{-1}[A] \cap Y^{-1}[B]) \\
&= P(X^{-1}[A])P(Y^{-1}[B]) \\
&= P(\{\omega \in \Omega : X \in A\})P(\{\omega \in \Omega : Y \in B\}) \text{ (done)}
\end{aligned}$$

$\blacksquare$

Corollary 3.2.6. (Joint Distribution of Independent Random Variable) Given two discrete random variable X, Y, we have

$$X \text{ and } Y \text{ are independent} \iff \forall (x, y) \in R(X) \times R(Y), f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Proof. ($\longrightarrow$)

Notice that

$$f_{X,Y}(x,y) = P(X \in \{x\} \text{ and } Y \in \{y\})$$
$$\text{and } f_X(x) = P(X \in \{x\}) \text{ and } f_Y(y) = P(Y \in \{y\})$$

and we are done.

($\longleftarrow$)

Let $A, B \in \mathcal{B}(\mathbb{R})$. We wish to prove

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$$

$\blacksquare$

Corollary 3.2.7. (Joint Distribution of Independent Random Variable) Given two continuous random

3.3 Bayes' Theorem

> **Question 5**
>
> We have the following knowledge
>
> $$P(\text{ tested } H|\text{ being } H) = 0.9$$
>
> $$P(\text{ tested } T|\text{ being } T) = 0.8$$
>
> $$P(\text{ being } H) = 0.05$$
>
> Compute
>
> $$P(\text{ being } H|\text{ tested } H)$$

Proof. We have

$$\begin{cases} \frac{P(\text{ tested } H\cap \text{ being } H)}{P(\text{ being } H)} = 0.9 \\ P(\text{ being } H) = 0.05 \end{cases}$$

So we have

$$P(\text{ tested } H \cap \text{ being } H) = 0.045$$

Also, the fact $P(\text{ being } H) = 0.05$ tell us $P(\text{ being } T) = 0.95$. Then from

$$\begin{cases} \frac{P(\text{ tested } T\cap \text{ being } T)}{P(\text{ being } T)} = 0.8 \\ P(\text{ being } T) = 0.95 \end{cases}$$

We have

$$P(\text{ tested } T \cap \text{ being } T) = 0.76$$

This give us

$$P(\text{ tested } H \cap \text{ being } T) = P(\text{ being } T) - P(\text{ tested } T \cap \text{ being } T) = 0.19$$

Then we have

$$P(\text{ tested } H) = P(\text{ tested } H \cap \text{ being } H) + P(\text{ tested } H \cap \text{ being } T) = 0.235$$

Then

$$\frac{P(\text{ being } H \cap \text{ tested } H)}{P(\text{ tested } H)} = \frac{0.045}{0.235}$$

∎

Theorem 3.3.1. (Bayes's Theorem)

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

We have the following knowledge

$$P(\text{ tested } H| \text{ being } H) = 0.9$$

$$P(\text{ tested } T| \text{ being } T) = 0.8$$

$$P(\text{ being } H) = 0.05$$

Compute

$$P(\text{ being } H| \text{ tested } H)$$

Proof. Let $B = (\text{being } H)$ and $A = (\text{tested } H)$. By Bayes' Theorem, we have

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{0.9 \times 0.05}{0.235}$$

where $P(A)$ is computed as above. ∎

Theorem 3.3.2. (First Bayes' Theorem) Let $\{F_1, \ldots, F_n\} \subseteq \mathcal{F}$ be a partition of Ω. We have

$$P(A) = \sum_{k=1}^{n} P(A|F_k)P(F_k)$$

Question 7

Consider the situation: We try to put u balls into distinct n barrels. Let $p_{r,k}$ be the probability that there are exactly r barrels filled with balls, when the first $k \le u$ balls is putted.

Find $p_{n,u}$ by giving a recurrence relation.

Proof. We have

$$p_{r,k} = p_{r,k-1}\frac{r}{n} + p_{r-1,k-1}\frac{n-(r-1)}{n}$$

In fact, we have

$$p_{r,k} =$$

$$= \sum_{i=0}^{n} p_{i,k-1}P(r \text{ barrels is filled with the first } k \text{ balls}|i \text{ barrels are filled with the first } k-1 \text{ b}$$

where we define $P(A|B) = 0$ if $P(B) = 0$.

The reason we have the recurrence relation is that the conditional probability is 0 when $i < r - 1$ or $i > r$. ∎

Theorem 3.3.3. (Second Bayes' Theorem) Let $\{F_1, F_2, \ldots, F_n\} \subseteq \mathcal{F}$ be a partition of Ω. We have

$$P(F_j|A) = \frac{P(A|F_j) P(F_j)}{\sum_{i=1}^{n} P(A|F_i) P(F_i)}$$

Proof.

$$P(F_j|A) = \frac{P(F_j \cap A)}{P(A)}$$
$$= \frac{P(A|F_j) P(F_j)}{P(A)}$$
$$= \frac{P(A|F_j) P(F_j)}{P(A \cap F_1) + P(A \cap F_2) + \cdots P(A \cap F_n)}$$
$$= \frac{P(A|F_j) P(F_j)}{\sum_{i=1}^{n} P(A|F_i) P(F_i)}$$

∎

3.4 Independent Disjoint Grouping Theorem

Definition 3.4.1. (Definition of Sub-σ-Algebra Being Independent) In $(\Omega, \mathcal{F}, P)$, we say a finite amount of sub-σ-algebra $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n\}$ is independent if we arbitrarily choose one event G_i from each sub-σ-algebra $\mathcal{G}_i$, we have

$$P\left(\bigcap_{i=1}^{n} G_i\right) = \prod_{i=1}^{n} P(G_i)$$

Definition 3.4.2. (Definition of Set of Random Variable Being Independent) We say a set of random variable $\{x_1, \ldots, x_n\}$ is independent if

$$\{\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)\} \text{ are independent}$$

Theorem 3.4.3. (Example)

$$\{X_1, X_2, X_3, X_4\} \text{ are independent} \implies X_2 X_3 \text{ is independent of } X_1 + \cos(2X_4)$$

Chapter 4

Multiple Random Variable

4.1   Joint Distribution

Definition 4.1.1. (Definition of Random Vector) Given $(\Omega, \mathcal{F}, P)$ and two random variables X, Y on $(\Omega, \mathcal{F})$. Consider the random vector

$$(X, Y) : (\Omega, \mathcal{F}, P) \to (\mathbb{R} \times \mathbb{R}, \sigma(\mathcal{B} \times \mathcal{B}))$$

Notice that $\sigma(\mathcal{B} \times \mathcal{B})$ is the smallest sigma algebra that contain $\mathcal{B} \times \mathcal{B}$. This shows that not all element element in $\sigma(\mathcal{B}, \mathcal{B})$ is rectangular. That is, not all element is of the form

$$A \times B$$

We see

$$[0, 2] \times [1, 2] \text{ and } [1, 2] \times [0, 2] \text{ are both element of } \sigma(\mathcal{B}, \mathcal{B})$$

and their union is not rectangle. The induced measure by the pair of random vector is defined to be

$$\mathcal{L}_{X,Y}(C) = P(\{\omega \in \Omega : (X, Y) \in C\})$$

Notice that given rectangular $C = A \times B$ where $A, B \in \mathcal{B}$. We have

$$\{\omega \in \Omega : (X, Y) \in C\} = \{\omega \in \Omega : X \in A\} \cap \{\omega \in \Omega : X \in B\}$$

This kind of expression can only be given if C is rectangular.

Definition 4.1.2. (Definition of Discrete Joint Distribution) Given $(\Omega, \mathcal{F}, P)$ and two discrete random variable X, Y. The random vector

$$(X, Y) : (\Omega, \mathcal{F}, P) \to (\mathbb{R} \times \mathbb{R}, \sigma(\mathcal{B} \times \mathcal{B}))$$

has joint probability $p_{(X,Y)}$ defined on $R(X) \times R(Y)$ by

$$\begin{aligned} p_{(X,Y)}(x, y) &:= P(X = x, Y = y) \\ &= P(\omega \in \Omega : (X, Y) = (x, y)) \\ &= \mathcal{L}_{(X,Y)}(\{x\}, \{y\}) \end{aligned}$$

Theorem 4.1.3. (Computation of Probability) Let $A \in \sigma(\mathcal{B} \times \mathcal{B})$ be a joint Borel set. We have

$$
\begin{aligned}
P((X,Y) \in A) &= P(\omega \in \Omega : (X,Y) \in A) \\
&= \sum_{(x,y) \in A} P(\{\omega \in \Omega : (X,Y) = (x,y)\}) \\
&= \sum_{(x,y) \in A} p_{(X,Y)}(x,y)
\end{aligned}
$$

Corollary 4.1.4. We have

$$
\begin{aligned}
P(X = x) &= P((X,Y) \in \{x\} \times R(Y)) \\
&= \sum_{y \in R(Y)} p_{(X,Y)}(x,y) \\
&= \sum_{y \in R(Y)} P(X = x, Y = y)
\end{aligned}
$$

Question 8

An urn has 2 red, 5 white and 3 green balls. Select 3 balls at random and let

(a) X be the number of red balls

(b) Y be the number of white balls

We have the following table

$y \backslash x$	0	1	2	$P(Y = y)$
0	$\frac{1}{120}$	$\frac{2 \cdot 3}{120}$	$\frac{3}{120}$	$\frac{10}{120}$
1	$\frac{5 \cdot 3}{120}$	$\frac{2 \cdot 5 \cdot 3}{120}$	$\frac{5}{120}$	$\frac{50}{120}$
2	$\frac{10 \cdot 3}{120}$	$\frac{10 \cdot 2}{120}$	0	$\frac{50}{120}$
3	$\frac{10}{120}$	0	0	$\frac{10}{120}$
$P(X = x)$	$\frac{56}{120}$	$\frac{56}{120}$	$\frac{8}{120}$	1

One should notice that

$$
P(X = x, Y = y) = \frac{\binom{2}{x}\binom{5}{y}\binom{3}{3-x-y}}{\binom{10}{3}} \text{ where } \binom{n}{m} = 0 \text{ if } m < 0
$$

With this table, we can compute

$$
P(X = 2 | X \geq Y) = \frac{8}{45}
$$

Definition 4.1.5. (Definition of Jointly Continuous Random Variable) Let X, Y be two continuous random variable. There exists $f : R(X) \times R(Y) \to \mathbb{R}$ such that

$$P((X,Y) \in S) = \iint_S f(x,y)dxdy$$

4.2 Conditional Probability and Expectation

Theorem 4.2.1. (Conditional Probability of Discrete Random Variable) Given two discrete random variable $X, Y : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, we have

$$P(X = x | Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Theorem 4.2.2. (Conditional Probability of Continuous Random Variable) Given two continuous random variable $X, Y : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, we have

$$P(X \in [a, b] | Y = y) = \frac{\int_a^b f_{X,Y}(x, y) dx}{f_Y(y)}$$

Definition 4.2.3. (Conditional Distribution) Given two discrete random variables, we write

$$p_X(x | Y = y) = P(X = x | Y = y)$$

and two continuous random variables, we write

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

Definition 4.2.4. (Definition of Conditional Expectation) Conditional Expectation of Y given $X = x$ is defined by

$$E[Y | X = x] = \sum_{y \in R(Y)} y P(Y = y | X = x)$$
$$= \int_{\mathbb{R}} y f_Y(y | X = x) dy$$

Theorem 4.2.5. (Tower Property) We have

$$E[Y] = E[E[Y | X = x]]$$

Proof. This is clear as one can deduce

$$E[E[Y|X = x]] = \sum_{x \in R(X)} E[Y|X = x]p_X(x)$$

$$= \sum_{x \in R(X)} \sum_{y \in R(Y)} y p_Y(y|X = x)p_X(x)$$

$$= \sum_{y \in R(Y)} y \sum_{x \in R(X)} p_{X,Y}(x, y)$$

$$= \sum_{y \in R(Y)} y p_Y(y)$$

$$= E[Y]$$

∎

Conjecture: Provided $\forall \omega, P(\{\omega\}) \in \mathbb{R}$, is it true

$$\forall A \in \sigma(E[X|Y]), \sum_{\omega \in A} E[X|Y]P(\{\omega\}) = \sum_{\omega \in A} X P(\{\omega\})$$

?

4.3   On Expectation

Theorem 4.3.1. (Expectation of Real Valued Function of Random Variable)
Given $g : \mathbb{R} \to \mathbb{R}$ and $X : \Omega \to \mathbb{R}$, we have

$$E[g(X)] = \sum_{x \in R(X)} g(x)p_X(x)$$

$$\text{or if continuous } = \int_{\mathbb{R}} g(x)f_X(x)dx$$

Corollary 4.3.2. (Expectation of Linear Change of Random Variable) Given $X : \Omega \to \mathbb{R}$, we have

$$E[aX + b] = aE[X] + b$$

Theorem 4.3.3. (Expectation of Addition of Random Variables) Given random variables $X_1, \ldots, X_n$

$$E[\sum_{k=1}^{n} X_k] = \sum_{k=1}^{n} E[X_k]$$

Theorem 4.3.4. (Expectation of Real Valued Function of Multiple Random Variables) Given $g : \mathbb{R}^2 \to \mathbb{R}$ and $X, Y : \Omega \to \mathbb{R}$, we have

$$E[g(X,Y)] = \sum_{(x,y) \in R(X) \times R(Y)} g(x,y)p_{X,Y}(x,y)$$

$$\text{or if continuous } = \iint_{\mathbb{R}^2} g(x,y)f_{X,Y}(x,y)dxdy$$

Corollary 4.3.5. (Expectation of Product of Independent Random Variables) Given $X, Y : \Omega \to \mathbb{R}$, we have

$$E[g(X)h(Y)] = E[g(X)] \times E[h(Y)]$$

Proof. If X, Y are discrete

$$E[g(X)h(Y)] = \sum_{(x,y)\in R(X)\times R(Y)} g(x)h(y)p_{X,Y}(x,y)$$

$$= \sum_{(x,y)\in R(X)\times R(Y)} g(x)p_X(x)h(y)p_Y(y)$$

$$= \sum_{x\in R(X)}\sum_{y\in R(Y)} g(x)p_X(x)h(y)p_Y(y)$$

$$= \sum_{x\in R(X)} g(x)p_X(x) \sum_{y\in R(Y)} h(y)p_Y(y)$$

$$= E[g(X)] \times E[h(Y)]$$

If X, Y are continuous

$$E[g(X)h(Y)] = \iint_{\mathbb{R}^2} g(x)h(y)f_{X,Y}(x,y)dxdy$$

$$= \iint_{\mathbb{R}^2} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{\mathbb{R}} g(x)f_X(x)dx \int_{\mathbb{R}} h(y)f_Y(y)dy$$

$$= E[g(X)] \times E[h(Y)]$$

∎

Question 9

Let (X, Y) be a random point in the triangle $\{(x, y) : x, y \geq 0, x+y \leq 1\}$. Show that X, Y are not independent and

$$E[XY] \neq E[X] \times E[Y]$$

Proof. Notice the area of the triangle is $\frac{1}{2}$, so we have

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{if } (x,y) \text{ in the triangle} \\ 0 & \text{if else where} \end{cases}$$

Compute the marginal probability

$$f_X(x) = \int_0^{1-x} f_{X,Y}(x,y)dy$$

$$= \int_0^{1-x} 2dy = 2 - 2x$$

Notice the symmetry so we have

$$f_Y(y) = 2 - 2y$$

Then it is quite clear that X, Y are not independent, as we can observe $f_{X,Y}(0,0) = 2 \neq 4 = f_X(0)f_Y(0)$ ∎

Compute

$$
\begin{aligned}
E[X] &= \int_0^1 x f_X(x)dx \\
&= \int_0^1 2x - 2x^2 dx \\
&= x^2 - \frac{2x^3}{3}\Big|_{x=0}^1 \\
&= 1 - \frac{2}{3} = \frac{1}{3}
\end{aligned}
$$

and compute

$$
\begin{aligned}
E[XY] &= \int_0^1 \int_0^{1-x} xy f_{X,Y}(x,y)dydx \\
&= 2\int_0^1 \int_0^{1-x} xydydx \\
&= 2\int_0^1 \frac{xy^2}{2}\Big|_{y=0}^{1-x}dx \\
&= \int_0^1 x(1-x)^2 dx \\
&= \int_0^1 x^3 - 2x^2 + xdx \\
&= \frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}\Big|_{x=0}^1 \\
&= \frac{1}{4} - \frac{2}{3} + \frac{1}{2} = \frac{1}{12} \neq \frac{1}{9} = E[X] \times E[Y]
\end{aligned}
$$

Question 10

Let (X, Y) be a random point in the diamond $\{(x,y) : |x| \leq 1 \text{ and } |y| \leq 1\}$. Show that X, Y are not independent, yet

$$E[XY] = E[X] \times E[Y]$$

Proof. The area is 2. Then the density function is

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{if } (x,y) \text{ in the diamond} \\ 0 & \text{if not} \end{cases}$$

Compute the marginal density

$$f_X(x) = \int_{-|1-x|}^{|1-x|} f_{X,Y}(x,y)dy$$
$$= |1-x|$$

We see X, Y are not independent as

$$f_{X,Y}(0,0) = \frac{1}{2} \neq 1 = f_X(0) \times f_Y(0)$$

Compute

$$E[XY] = \int_0^1 \int_{-|1-x|}^{|1-x|} \frac{xy}{2}dydx$$
$$= \int_0^1 \frac{xy^2}{4}\Big|_{y=-|1-x|}^{|1-x|} dx = 0$$

and also

$$E[Y] = E[X] = \int_{-1}^1 x|1-x|dx = 0 (\because \text{ odd function })$$

∎

Question 11

Assumethat an urn contains 10 black, 7 red, and 5 white balls. Select 5 balls

(a) with and

(b) without replacement

and let X be the number of red balls selected. Compute $E[X]$

Proof. Let $X = \sum_{i=1}^5 X_i$ where

$$p_{X_i}(1) = \frac{7}{22} \text{ and } p_{X_i}(0) = \frac{15}{22}$$

60

We have

$$E[X] = E[\sum_{i=1}^{5} X_i]$$

$$= \sum_{i=1}^{5} E[X_i]$$

$$= \sum_{i=1}^{5} \frac{7}{22} = \frac{35}{22}$$

For (b), we again have $E[X_i] = \frac{7}{22}$. The fact they are dependent doesn't matter, we still have

$$E[X] = \sum_{i=1}^{5} E[X_i] = \frac{35}{22}$$

∎

Question 12

n people bought n gifts. The gifts then are distributed to the n people at random. Let X be the number of people who receive their own gift. Compute $E[X]$

Proof. The question is ambiguous, but the ambiguity doesn't make our question have multiple answer. The question doesn't specified that whether every one get one gift or if one can get multiple. Yet, if we let

$$I_i = I_{\text{person } i \text{ receive own gift}}$$

Former situation leads to the dependence of variables while the latter leads to the independence. However, dependent or not, because

$$X = \sum_{i=1}^{n} I_i$$

we have

$$E[X] = \sum_{i=1}^{n} E[I_i] = n(\frac{1}{n}) = 1$$

∎

Five married couples are seated around a table at random. Let X be the number of wives who sit next to their husbands. What is $E[X]$?

Proof. Let

$$X = \sum_{i=1}^{5} I_i$$

We have $E[I_i] = \frac{2}{9}$, then

$$E[X] = \sum_{i=1}^{5} E[I_i] = \frac{10}{9}$$

$\blacksquare$

Sample from n cards, with replacement, indefinitely. Let N be the number of draws you need to get a collection of all n cards. Compute $E[N]$

Proof. Let Ω be the set of sequences of the card drawn. Let $N_1(\omega)$ be the number of card you need to draw to get the first new card if ω is given. Let $N_2(\omega)$ be the number of additional cards you need to draw after the first new card to get the second new card. Random Variables $N_i(\omega)$ is defined similarly. Then we see

$$N = \sum_{i=1}^{n} N_i$$

It is quite clear that

$$N_i \sim \text{Geo}(\frac{n-i+1}{n})$$

Then we have

$$E[N] = \sum_{i=1}^{n} E[N_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i}$$

$\blacksquare$

Question 15

Assume that an urn contain 10 black, 7 red, and 5 white balls. Select 5 balls

(a) with replacement

(b) without replacement

Let W be the number of whit balls selected, and Y the number of different colors. Compute $E[W], E[Y]$

Proof. In both $(a), (b)$,

$$E[W] = \sum_{i=1}^{5} E[I_i] = \frac{25}{22}$$

Let

$$Y = I_B + I_R + I_W$$

where I_B is the indicator that if black ball is selected in the 5 balls.

In (a), we have

$$E[I_B] = P(I_B = 1) = 1 - (\frac{12}{22})^5$$

$$E[I_R] = 1 - (\frac{15}{22})^5$$

$$E[I_W] = 1 - (\frac{17}{22})^5$$

Then

$$E[Y] = 3 - \frac{12^5 + 15^5 + 17^5}{22^5}$$

In (b), we have

$$E[I_B] = 1 - \frac{\binom{12}{5}}{\binom{22}{5}}$$

$$E[I_R] = 1 - \frac{\binom{15}{5}}{\binom{22}{5}}$$

$$E[I_W] = 1 - \frac{\binom{17}{5}}{\binom{22}{5}}$$

Then

$$E[Y] = 3 - \frac{\binom{12}{5} + \binom{15}{5} + \binom{17}{5}}{\binom{22}{5}}$$

■

4.4   Covariance

Definition 4.4.1. (Definition of Covariance) The covariance of two random variable X, Y is defined by

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

Theorem 4.4.2. (Formula for Covariance) We have

$$\text{Cov}[X, Y] = E[XY] - E[X] \times E[Y]$$

Proof. Theorem 4.3.3 let us deduce

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

■

Theorem 4.4.3. (Independent Implies Zero Covariance)

$$X, Y \text{ are independent} \implies \text{Cov}[X, Y] = 0$$

Proof. See Theorem 4.3.4

■

An example for X, Y be dependent, yet $\text{Cov}[X, Y] = 0$ is question 10 in last section.

4.5 Variance of Sum of Random Variables

At now, one should notice that expectation of sum of random variables is just the sum of expectations of random variables with or without independence.

Also, the expectation of product of independent random variables is the product of expectation of independent variables. Notice that this won't necessarily hold true if variables are independent.

Theorem 4.5.1. (Variance-Covariance Formula) We have

$$E[(\sum_{i=1}^{n} X_i)^2] = \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$$

$$\text{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j]$$

Proof. Notice

$$(\sum_{i=1}^{n} X_i)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j$$

Then the proof follows from the fact expectation is a homomorphism from set of random variables on $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, +)$.

Compute

$$\text{Var}[\sum_{i=1}^{n} X_i] = E[(\sum_{i=1}^{m} X_i)^2] - (E[\sum_{i=1}^{n} X_i])^2$$

$$= \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] - (\sum_{i=1}^{n} E[X_i])^2$$

$$= \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] - (\sum_{i=1}^{n} E^2[X_i] + \sum_{i \neq j} E[X_i]E[X_j])$$

$$= \sum_{i=1}^{n} E[X_i^2] - E^2[X_i] + \sum_{i \neq j} E[X_i X_j] - E[X_i]E[X_j]$$

$$= \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j]$$

∎

Corollary 4.5.2. (Variance of Sum of Mutually Independent Random Variables Is the Sum of Variances of Mutually Independent Random Variables)
If $X_1, \ldots, X_n$ are mutually independent, then

$$\text{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \text{Var}[X_i]$$

Proof. Because $X_1, \ldots, X_n$ are mutually independent, we can deduce

$$\text{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j]$$

$$= \sum_{i=1}^{n} \text{Var}[X_i]$$

$\blacksquare$