Emory University
MATH 361 Mathematical Statistics I
Learning Notes

Jiuru Lyu

December 3, 2023

Contents

1 Prerequisites

Definition 1.0.1 (Geometric Series). A geometric series has the form

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + \cdots$$

If $|r| < 1$, then the series converges to $\dfrac{a}{1-r}$. Otherwise, it diverges.

Example 1.0.2 Does the series $\sum_{n=1}^{\infty} 2^{2n}3^{1-n}$ converge or divers?

Solution 1.

Note that

$$2^{2n}3^{1-n} = \left(2^2\right)^n 3^{1-n} = 4^n\left(\frac{1}{3}\right)^{n-1} = 4 \cdot 4^{n-1}\left(\frac{1}{3}\right)^{n-1} = 4\left(\frac{4}{3}\right)^{n-1}.$$

So,

$$\sum_{n=1}^{\infty} 2^{2n}3^{1-n} = \sum_{n=1}^{\infty} 4\left(\frac{4}{3}\right)^{n-1}$$

is a geometric series, with $a = 4$ and $r = \dfrac{4}{3}$.

Since $|r| = \left|\dfrac{4}{3}\right| = \dfrac{4}{3} > 1$, the series diverges. □

Definition 1.0.3 (Taylor Series). The Taylor series expanded about a of a differentiable function f is

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots.$$

Definition 1.0.4 (Maclaurin Series). The Taylor series expanded about $a = 0$.

Remark. *The Maclurin Series of* e^x *is given by* $e^x = \sum_{n=0}^{\infty} \dfrac{x^n}{n!}$.

Theorem 1.0.5 Binomial Expansion

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

where $\binom{n}{k}$ is read as "n choose k" and can also be written as nCk.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Theorem 1.0.6 Integration by Parts

$$\int u \, \mathrm{d}v = uv - \int v \, \mathrm{d}u.$$

Example 1.0.7 Evaluate $\displaystyle\int xe^{-x} \, \mathrm{d}x$.

Solution 2.

Let $u = x, \ \mathrm{d}v = e^{-x} \, \mathrm{d}x$. So, $\mathrm{d}u = \mathrm{d}x$ and $v = \displaystyle\int e^{-x} \, \mathrm{d}x = -e^{-x}$. Then,

$$\int xe^{-x} \, \mathrm{d}x = -xe^{-x} - \int -e^{-x} \, \mathrm{d}x = -xe^{-x} - e^{-x} + C.$$

□

Definition 1.0.8 (Type I Improper Integral). If $\displaystyle\int_a^t f(x) \, \mathrm{d}x$ exists for all $t > 0$, then

$$\int_a^\infty f(x) \, \mathrm{d}x = \lim_{t \to \infty} \int_a^t f(x) \, \mathrm{d}x.$$

Example 1.0.9 Evaluate $\displaystyle\int_0^\infty xe^{-x} \, \mathrm{d}x$.

Solution 3.

$$\begin{aligned}
\int_0^\infty xe^{-x} \, \mathrm{d}x = \lim_{t \to \infty} \int_0^t xe^{-x} \, \mathrm{d}x &= \lim_{t \to \infty} \left[-xe^{-x} - e^{-x}\right]_0^t \\
&= \lim_{t \to \infty} \left(-te^{-t} - e^{-t} + 1\right) \\
&= -\lim_{t \to \infty} \left(\frac{t}{e^t}\right) - \lim_{t \to \infty} e^{-t} + 1 \\
&= -\lim_{t \to \infty} \left(\frac{1}{e^t}\right) - 0 + 1 = -0 - 0 + 1 = 1.
\end{aligned}$$

□

Example 1.0.10 Double Integrals over Irregular Domains.
Consider

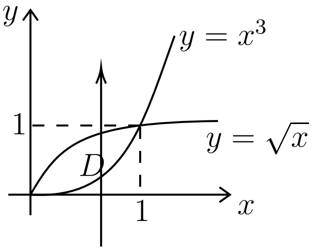
$$\iint_D 4xy - y^4 \, \mathrm{d}A,$$

where D is the region bounded between $y = \sqrt{x}$ and $y = x^3$.

Evaluate this double integral over D.

Solution 4.

Firstly, we draw the diagram representing D as follows:



$$
\begin{aligned}
\iint_D 4xy - y^3 \,\mathrm{d}A = \int_0^1 \int_{x^3}^{\sqrt{x}} 4xy - y^3 \,\mathrm{d}y\mathrm{d}x &= \int_0^1 \left[2xy^2 - \frac{1}{4}y^4\right]_{x^3}^{\sqrt{x}} \mathrm{d}x \\
&= \int_0^1 2x\big(x - x^6\big) - \frac{1}{4}\big(x^2 - x^{12}\big) \,\mathrm{d}x \\
&= \int_0^1 2x^2 - 2x^7 - \frac{1}{4}x^2 + \frac{1}{4}x^{12} \,\mathrm{d}x \\
&= \left[\frac{2}{3}x^3 - \frac{1}{4}x^8 - \frac{1}{12}x^3 + \frac{1}{52}x^{13}\right]_0^1 \\
&= \frac{2}{3} - \frac{1}{4} - \frac{1}{12} + \frac{1}{52} = \frac{55}{156}.
\end{aligned}
$$

$\square$

2 Probability

2.1 Sample Space and Probability

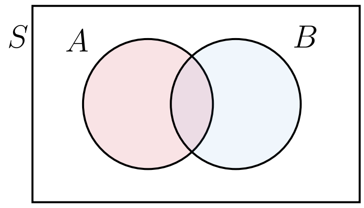
Definition 2.1.1 (Experiment). An *experiment* is a procedure with well-defined outcome.

Definition 2.1.2 (Sample Space/S). The *sample space*, denoted as S is the set of all possible outcomes of an experiment.

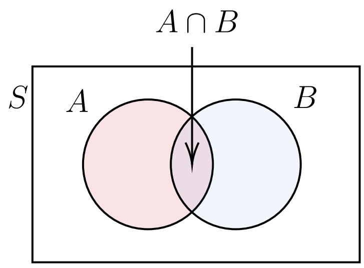
Definition 2.1.3 (Event). An *event* is a collection of outcomes.

> **Example 2.1.4** Consider flipping two coins. Use H to represent heads and T to represent tails. Then, $S = \{HH, HT, TH, TT\}$. Event "one heads"$= \{HT, TH\}$, and the event "at least one heads"$= \{HT, TH, HH\}$.

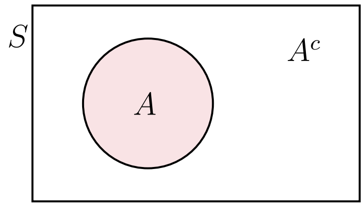
Definition 2.1.5 (Union/$\cup$). $A \cup B$ is the *union* of A and B, meaning everything in A and everything in B.



Definition 2.1.6 (Intersection/$\cap$). $A \cap B$ is the *intersection* of A and B, everything in both A and B



Definition 2.1.7 (Complement/A^c). A^c denotes the *complement* of A, meaning everything in$S$ that is not in A.



Corollary 2.1.8 $A \cap A^c = \{\} = \varnothing$.

Definition 2.1.9 (Mutually Exclusive). Two sets A and B over the same sample space are *mutually exclusive* if they have no outcomes in common. i.e., $A \cap B = \varnothing$.

Remark. *A and A^c are mutually exclusive, but not all sets mutually exclusive are complements of each other.*

Definition 2.1.10 (Probability Function). Let A be an event over a sample space S. Then, $\mathbf{P}(A)$ denotes the *probability* of A and P is the *probability function*. The probability function P assigns a number $\mathbf{P}(A)$ for each event $A \subseteq S$.

Axiom 2.1.11 Kolmogorov Axioms

1. Let A be an event in S, then $\mathbf{P}(A) \geq 0$.
2. $\mathbf{P}(S) = 1$.
3. If A and B are mutually exclusive, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.
4. If $A_1, \ldots, A_n, \ldots$ are mutually exclusive sets, then

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

Proposition 2.1.12 $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

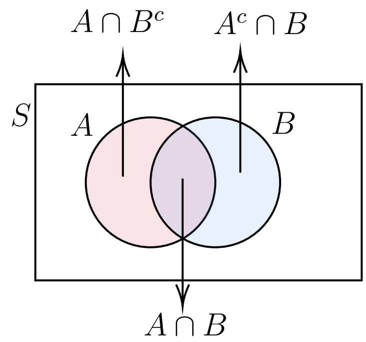
Proof 1. Note that $\mathbf{P}(S) = 1$. Since $A^c \cup A = S$, we have $\mathbf{P}(A \cap A^c) = 1$. Since A and A^c are mutually exclusive, $\mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c) = 1$. So, $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$. ■

Proposition 2.1.13 $\mathbf{P}(\varnothing) = 0$.

Proof 2. Note that $\mathbf{P}(S) = 1$. Then, $\mathbf{P}(S^c) = 1 - \mathbf{P}(S)$. By definition, we know $S^c = \varnothing$. So, $\mathbf{P}(\varnothing) = 1 - \mathbf{P}(S) = 1 - 1 = 0$. ■

Proposition 2.1.14 $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

Proof 3. Consider the following Venn diagram:



Note that $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^c)$ and $\mathbf{P}(B) = \mathbf{P}(A \cap B) + \mathbf{P}(A^c \cap B)$. So, we have

$$\mathbf{P}(A) + \mathbf{P}(B) = \boxed{\mathbf{P}(A \cap B^c) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A \cap B)} + \mathbf{P}(A \cap B). \tag{1}$$

From the Venn diagram, we notice that $\mathbf{P}(A \cap B^c) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A \cap B)$ is exactly $\mathbf{P}(A \cup B)$. So, Eq. (1) becomes $\mathbf{P}(A) + \mathbf{P}(B) = \mathbf{P}(A \cup B) + \mathbf{P}(A \cap B)$. That is exactly what is required: $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$. ■

Definition 2.1.15 (Classical Probability). In a discrete and finite case, S is finite and all outcomes are equally likely, and the probability function is defined as

$$\mathbf{P}(A) = \frac{|A|}{|S|},$$

where $|A|$ is the cardinality of A and $|S|$ is the cardinality of S.

Example 2.1.16 Despite the definition of classical probability (probability function defined for a discrete and finite case), there are other definitions of probability functions:

1. Discrete and Countably Infinite:

   Let $S = \mathbb{N}$ be the set of natural numbers. Then,

   $$\mathbf{P}(k) = \frac{1}{2^k}.$$

   It can also be verified that

   $$\mathbf{P}(S) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

2. Continuous and Uncountably Infinite:

   Let $S = [0, 1]$. Suppose E is a subset of $[0, 1]$ such that $\int_E \mathrm{d}x$ is defined. Then,

   $$\mathbf{P}(E) = \int_E \mathrm{d}x,$$

   and it can also be verified that $\mathbf{P}(S) = 1$.

2.2 Conditional Probability and Independence

Definition 2.2.1 (Conditional Probability). We read $\mathbf{P}(A|B)$ as the probability of A given B. Knowing B occurs, we create a new sample space, in which the probability of A occurs changes:

$$\mathbf{P}(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{1/|S|}{1/|S|} = \frac{|A \cap B|/|S|}{|B|/|S|} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Corollary 2.2.2 $\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B)$

Example 2.2.3 Find the probability of dealing A first, 2 second, and 3 third.

Solution 1.

$$\begin{aligned}\mathbf{P}(\text{dealing } A, 2, 3) &= \mathbf{P}(A \text{ first})\mathbf{P}(2 \text{ second}|A \text{ first})\mathbf{P}(3 \text{ third}|A \text{ first} \cap 2 \text{ second}) \\ &= \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50}\end{aligned}$$

□

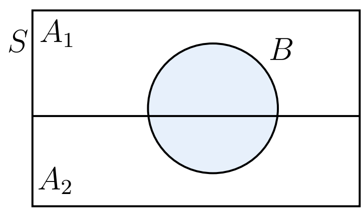
Corollary 2.2.4 $\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_2 \cap A_1)$.

Theorem 2.2.5 The Law of Total Probability
Suppose the sample space $S = A_1 \cup A_2 \cup \cdots \cup A_n$, with $A_i \cap A_j = \varnothing \quad \forall i \neq j$. Then,

$$\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \cdots + \mathbf{P}(B \cap A_n).$$

Remark. *This theorem gives us a nice way to partition the sample space.*

Example 2.2.6



As represented in the diagram above, $\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2)$.

Theorem 2.2.7 Bayes Theorem

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)}.$$

Example 2.2.8 Coronary Artery Disease (CAD)

The probability of someone having CAD is 60%. In a study of 101 patients, 37 of them are known to NOT have CAD and 64 are known to have CAD. Of the 37 patients without CAD, 34 had negative tests while 3 had positive tests. Of the 64 with CAD, 54 had positive tests and 10 had negative tests. Find the probability of a patient has CAD given positive test.

Solution 2.

Let $T+$ be positive test, $T-$ be negative test, $D+$ be presence of CAD, and $D-$ be absence of CAD. Then, from the problem, we have

$$\mathbf{P}(D+) = 0.6; \quad \mathbf{P}(D-) = 1 - \mathbf{P}(D+) = 0.4$$

and

$$\mathbf{P}(T+|D+) = \frac{54}{64} \approx 0.84; \quad \mathbf{P}(T-|D-) = \frac{34}{37} \approx 0.92; \quad \mathbf{P}(T+|D-) = \frac{3}{37} \approx 0.08.$$

Then, by Bayes Theorem,

$$\begin{aligned}\mathbf{P}(D+|T+) &= \frac{\mathbf{P}(T+|D+)\mathbf{P}(D+)}{\mathbf{P}(T+|D+)\mathbf{P}(D+) + \mathbf{P}(T+|D-)\mathbf{P}(D-)} \\ &= \frac{0.84 \times 0.6}{0.84 \times 0.6 + 0.08 \times 0.4} \approx \boxed{0.94}.\end{aligned}$$

□

Definition 2.2.9 (Independence). Events A and B are *independent* if $\mathbf{P}(A|B) = \mathbf{P}(A)$, meaning the occurrence of B does note affect the occurrence of A.

Corollary 2.2.10 If A and B are independent, then $\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B) = \mathbf{P}(A)\mathbf{P}(B)$.

Example 2.2.11 Draw a card from 52 card deck

Let A: The card is an Ace and H: The card is a hearts. Then,

$$\mathbf{P}(A \cap H) = \mathbf{P}(\text{The card is an Ace of hearts}) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = \mathbf{P}(H)\mathbf{P}(A).$$

So, ranks and suits are independent.

Example 2.2.12 Mutually Exclusive v.s. Independence

A coin is flipped twice: $S = \{HH, TH, HT, TT\}$. Let A = The first flip is H = $\{HH, HT\}$ and B = The second flip is $T = \{HT, TT\}$.

- A and B are independent: $A \cap B = \{HT\}$. So, $\mathbf{P}(A \cap B) = \frac{1}{4}$. Since $\mathbf{P}(A \cap B)\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}(A)\mathbf{P}(B)$, we know A and B are independent.

- A and B are not mutually exclusive because $\mathbf{P}(A \cap B) = \dfrac{1}{4} \neq 0$.

Definition 2.2.13 (Repeated Trials). A sequence of events $A_1, \ldots, A_n$ is called independent if for any combination

$$\mathbf{P}(A_{i1} \cap A_{i2} \cap \cdots \cap A_{ik}) = \mathbf{P}(A_{i1})\mathbf{P}(A_{i2}) \cdots \mathbf{P}(A_{ik}).$$

In this case, each individual event is called a *trial*.

Example 2.2.14 Roll a fair die repeatedly. What is the probability that the first 6 appears on the roll k? If I win when 6 is rolled, what is the probability that I win?

Solution 3.

Let $A_j =$ the first 6 is rolled on roll j.

$$\begin{aligned}
j = 1 \quad & \mathbf{P}(A_1) = \frac{1}{6} \\
j = 2 \quad & \mathbf{P}(A_2) = \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) \\
j = 3 \quad & \mathbf{P}(A_3) = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) \\
j = 4 \quad & \mathbf{P}(A_4) = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) \\
\vdots \\
j = k \quad & \mathbf{P}(A_k) = \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)
\end{aligned}$$

So,

$$\begin{aligned}
\mathbf{P}(\text{I win}) &= \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots + \mathbf{P}(A_k) + \cdots \\
&= \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \cdots + \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right) + \cdots \\
&= \left(\frac{1}{6}\right)\left(1 + \left(\frac{5}{6}\right) + \cdots + \left(\frac{5}{6}\right)^{k-1} + \cdots\right) \\
&= \left(\frac{1}{6}\right)\sum_{i=0}^{\infty}\left(\frac{5}{6}\right)^i = \frac{1}{6} \cdot \frac{1}{1 - \frac{5}{6}} = \frac{1}{6} \cdot 6 = \boxed{1}.
\end{aligned}$$

□

Example 2.2.15 Three people A, B and C take turn to flip a coin. Whoever gets a heads wins. Find the probability of each individual winning.

Solution 4.
First consider the case when Player A wins. Let A_j = Player A wins on the j-th turn.

$$j=1 \quad \mathbf{P}(A_1)=\frac{1}{2}$$

$$j=2 \quad \mathbf{P}(A_2)=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

$$j=3 \quad \mathbf{P}(A_3)=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)\left(\frac{1}{2}\right)=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^2\left(\frac{1}{6}\right)$$

$$j=4 \quad \mathbf{P}(A_4)=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^3\left(\frac{1}{2}\right)$$

$$\vdots$$

$$j=k \quad \mathbf{P}(A_k)=\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right)=\left(\frac{1}{8}\right)^{k-1}\left(\frac{1}{2}\right)$$

So,

$$\begin{aligned}\mathbf{P}(A\text{ wins})=\sum_{j=1}^{\infty}\mathbf{P}(A_j)&=\frac{1}{2}+\left(\frac{1}{8}\right)\left(\frac{1}{2}\right)+\cdots+\left(\frac{1}{8}\right)^{k-1}\left(\frac{1}{2}\right)+\cdots\\&=\frac{1}{2}\sum_{i=0}^{\infty}\left(\frac{1}{8}\right)^i\\&=\frac{1}{2}\cdot\frac{1}{1-\frac{1}{8}}=\frac{1}{2}\cdot\frac{8}{7}=\boxed{\frac{4}{7}}.\end{aligned}$$

Similarly, we can get the probability of player B wins to be $\mathbf{P}(B\text{ wins})=\frac{2}{7}$. Finally, we can compute the probability of player C wins by

$$\mathbf{P}(C\text{ wins})=1-\mathbf{P}(A\text{ wins})-\mathbf{P}(B\text{ wins})=1-\frac{4}{7}-\frac{2}{7}=\frac{1}{7}.$$

□

2.3 Combinatorics

Theorem 2.3.1 Multiplication Rule
If operation A can be performed in n ways and operation B in m ways, then the sequence (operation A, operation B) can be performed in $n\times m$ ways.

Corollary 2.3.2 Ordered Sequence Consider a set A and $|A|=n$. Then, an *ordered sequence* of A, $(x_1,x_2,\ldots,x_k)$ $s.t.$ $x_i\in A$, is picked with replacement of elements. Then,

$$|(x_1,x_2,\ldots,x_k)|=n^k.$$

Remark. *In this situation, repetition is allowed.*

Definition 2.3.3 (Permutation). *Permutation* is an ordered sequence without replacement of elements. That is, $(x_1, x_2, \ldots, x_k)\ s.t.\ x_i \in A$ and $x_i \neq x_j \forall i \neq j$. Then,

$$|(x_1, x_2, \ldots, x_k)| = n(n-1)\cdots(n-k+1).$$

It is also written as ${}_nP_k = \dfrac{n!}{(n-k)!}$.

Definition 2.3.4 (Combination). *Combination* is an unordered permutation (no order, no replacement of elements). So, we have

$$\begin{aligned}\text{permutation} &= \text{combination} \times \text{orderings} \\ {}_nP_k &= {}_nC_k \times k! \\ {}_nC_k &= \frac{{}_nP_k}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}\end{aligned}$$

Remark. *People are always distinct. Letter or coins are not usually distinct.*

Example 2.3.5 How many ways can we scramble the letters in STATISTICS?

Solution 1.

If the letter are distinct, then $10!$ ways to scramble the word. However, they are not distinct:

Non-distinct Letters	Ways to Scramble
$S-3$	$3!$
$T-3$	$3!$
$I-2$	$2!$

So, ways to scramble the word N satisfies

$$\begin{aligned}10! &= N \cdot 3! \cdot 3! \cdot 2! \\ N &= \frac{10!}{3! \cdot 3! \cdot 2!} \quad \text{Multinomial Coefficient}\end{aligned}$$

□

Definition 2.3.6 (Multinomial Coefficient). The *multinomial coefficient* is the number of ways that n objects with n_j of type j, where $j = 1, \ldots, r$, can be distinctly ordered. So,

$$\sum_{j=1}^{r} n_j = n$$

and

$$\text{Multinomial Coefficient} = \frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_r!}$$

Remark. *Tips for Counting:*

1. *Draw a picture of the structure*
2. *Construct a smaller problem when there are large numbers or variables.*
3. *If the structure of the problem falls into different categories, then add instead of multiple.*

2.4 Combinatorial Probabilities

Remark. *Probability Tips*

1. *Avoid multiplying probabilities. Always set up quotient.*
2. *Keep track of order. If we have order in the sample space, we will need order in the event.*
3. *Know some basic sample spaces:*
   - *Rolling n fair die:* $|S| = 6^n$ *(ordered).*
   - *Flipping n coins:* $|S| = 2^n$
   - *Dealing a hand of n cards:* $\binom{52}{n}$

Example 2.4.1 Roll 5 Fair Die. What is the size of the sample space? What is the probability that the first three have one face and the last two another? What is the probability that two faces show up exactly twice?

Solution 1.

Size of the sample space: $|S| = 6^5$.

Let A = the probability that the first three have one face and the last two another. $|A| = (6 \times 1 \times 1) \times (5 \times 1) = 30$. So, $\mathbf{P}(A) = \frac{30}{6^5} = \frac{5}{6^4}$.

Let B = the probability that two faces show up exactly twice. Note that we use $\binom{6}{2}$ to give the faces of the pairs. $\binom{4}{1}$ to the last one. $\frac{5!}{2! \cdot 2! \cdot 1!}$ ways to order the faces. So, $|B| = \binom{6}{2}\binom{4}{1}\frac{5!}{2! \cdot 2! \cdot 1!}$. Then, $\mathbf{P}(B) = \frac{\binom{6}{2}\binom{4}{1}\frac{5!}{2!\cdot 2!\cdot 1!}}{6^5}$. □

3 Random Variables

3.1 Discrete RV: Binomial & Hypergeometric

Definition 3.1.1 (Random Variable). A *random variable* is a number determined by the outcome of an experiment, $X : S \to \mathbb{R}$. Usually, we are not particularly interested in S but in the distribution of the outcomes for X. We want to describe the probability associated with different values of X.

Example 3.1.2 Flip three coins. Count the number of heads (H):

$$S = \{HHH, HHT, HTH, TTHTTHTHT, HTT, TTT\}$$

$$\widetilde{S} = \{3H, 2H, 1H, 0H\}.$$

Note in $\widetilde{S}$, not every outcome is equally likely. We would need to define a function for or a list of the values for each outcome in $\widetilde{S}$. This is an example of a *discrete random variable.*

Example 3.1.3 Pick a student. Let the random variable Y = height of the students in cm. Then, Y is an example of a *continuous random variable.* Continuous random variables can take on an interval of values.

Notation 3.1.4. X is a random variable and has a distribution (it is still abstract and unrealized). x is a number and a realized random variable X.

Definition 3.1.5 (Discrete Random Variable). A *discrete random variable* is a random variable whose range is finite of countable.

Definition 3.1.6 (Probability Density (Mass) Function / pdf).

$$\mathbf{P}_X(x) = \mathbf{P}(X = x) = \mathbf{P}(\{s \in S \mid X(s) = x\}).$$

Definition 3.1.7 (Cumulative Density Function / cdf).

$$F_X(x) = \mathbf{P}(X \leq x).$$

Remark. *CDFs and PDFs can be represented by functions, graphs, or tables.*

Example 3.1.8 Roll three fair die. Let X be the largest value of the three die. Find the pdf.

Solution 1.

Note the pdf

$$\mathbf{P}_X(x) = \mathbf{P}(X = x) = \mathbf{P}(X \leq x) - \mathbf{P}(X \leq x - 1).$$

Find the cdf of x. The die that take on at most the value x, so each die have x possible outcomes, and considering order, we know

$$F_X(x) = \frac{x^3}{6^3}.$$

Therefore,

$$\mathbf{P}_X(x) = \mathbf{P}(X \le x) - \mathbf{P}(X \le x-1) = F_X(x) - F_X(x-1) = \frac{x^3}{6^3} - \frac{(x-1)^3}{6^3}.$$

$\square$

Definition 3.1.9 (Bernoulli Distribution). The *Bernoulli distribution* is the classic "flip one coin," where X is the number of heads. Let $X \sim$ Bernoulli(p), where p stands for the probability of success. $x = 1$ for success and $x = 0$ for failure. The pdf of Bernoulli distribution is

$$\mathbf{P}_X(x) = p^x(1-p)^{1-x}$$

So,

$$\mathbf{P}_X(1) = p; \qquad \mathbf{P}_X(0) = (1-p).$$

Definition 3.1.10 (Binomial Distribution). The *binomial distribution* is adding Bernoulli trials together. Let $Y = X_1 + \cdots + X_n$ be the number of success with $X \sim$ Bernoulli(p) and $Y \sim$ Binomial(n, p). n is the number of trials and p is the probability of success. The pdf of binomial distribution is

$$\mathbf{P}_Y(y) = \binom{n}{y} p^y (1-p)^y.$$

Definition 3.1.11 (Hypergeometric Distribution). Suppose we have a bag of red (r) and white (w) chips and $r + w = N$. Let $X =$ the number of red chips when choosing n chips *without replacement*. Then, $X \sim$ Hypergeometric(r, w, n), and the pdf is given by

$$\mathbf{P}_X(x) = \mathbf{P}(X = x) = \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{r+w}{n}},$$

where $\binom{r}{x}$ is the number of ways to get x red, $\binom{w}{n-x}$ is the number of ways to get $n - x$ white, and $\binom{r+w}{n}$ is the total number of picking a size of x.

Remark. *If we choose sequence (ordered choose without replacement), we should get the exact same answer:* ${}_rP_x$ *is the number of ways the red chips can be selected in order,* ${}_wP_{n-x}$ *is the ways to choose the white chips,* $\binom{n}{x}$ *is the locations of the red chips, and* ${}_NP_n$ *is the* *orders of n chips. Then,*

$$\mathbf{P}_X(x) = \mathbf{P}(X = x) = \frac{({}_rP_x)\cdot({}_wP_{n-x})\cdot\binom{n}{x}}{{}_NP_n} = \frac{\dfrac{r!}{(r-x)!}\cdot\dfrac{w!}{(w-(n-x))!}\cdot\dfrac{n!}{(n-x)!x!}}{\dfrac{N!}{(N-n)!}}$$
$$= \frac{\dfrac{r!}{(r-x)!x!}\cdot\dfrac{w!}{(w-(n-x))!(n-x)!}}{\dfrac{N!}{(N-n)!n!}}$$
$$= \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{N}{n}}.$$

3.2 Continuous Random Variables

Definition 3.2.1 (Random Variable). A *random variable* is an outcome of an experiment mapped to a number: $X : S \to \mathbb{R}$. We are most interested describing the probability associated with different values of X.

> **Example 3.2.2** In a chemistry experiment that depends on temperature. Let $Y =$ the temperature measured for experiment, then the unit of temperature will/may lead to different "looking" results.

Definition 3.2.3 (Continuous Random Variables). A *continuous random variable* is a function from a sample space S to the real numbers $s.t.$ for any values a, b with $a < b$, there exists a function $f_Y(y)$ $s.t.$

$$\mathbf{P}(a < Y < b) = \int_a^b f_Y(y)\,\mathrm{d}y,$$

where $f_Y(y)$ is called the *pdf (probability density function)*, and (1) $f_Y(y) \geq 0 \quad \forall y \in \mathbb{R}$; (2) $\displaystyle\int_{-\infty}^{\infty} f_Y(y)\,\mathrm{d}y = 1$. Meanwhile, the *cdf (cumulative distribution function)* is defined as

$$\mathbf{P}(Y \leq y) = F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t.$$

The *q-th quantile* c can be defined as $F(c) = q \quad \forall q \in (0, 1)$.

Remark. *The probability for $a < Y < b$ can be regarded as the area under $f_Y(y)$ over the interval (a, b). Therefore, $\mathbf{P}(Y = a) = 0$.*

Definition 3.2.4 (Uniform Distribution). Suppose Y is a continuous random variable, and $Y \sim$ Uniform(a, b), then

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & y \in [a, b]; \\ 0 & \text{otherwise}. \end{cases}$$

Remark. *Y is continuous does not imply $f_Y(y)$ is also continuous.*

Definition 3.2.5 (Exponential Distribution). Suppose Y is a continuous random variable, and $Y \sim \text{Exponential}(\lambda)$ with $\lambda > 0$ is defined as

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y \geq 0.$$

Theorem 3.2.6 Fundamental Theorem of Calculus

- $\dfrac{\mathrm{d}}{\mathrm{d}x} \displaystyle\int_a^x f(t)\,\mathrm{d}t = F(x)$; and
- $\displaystyle\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a) \ s.t. \ \frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$.

Example 3.2.7 The Temperature Example - Cont'd.

Let X = temperature in °F and Y = temperature in °C. Given $f_X(x)$, $F_X(x)$, and $Y = \frac{5}{9}(X - 32)$. Use the cdf→pdf method to find $f_Y(y)$.

Solution 1.

$$\begin{aligned} F_Y(y) = \mathbf{P}(Y \leq y) &= \mathbf{P}\left(Y = \frac{5}{9}(X - 32) \leq y\right) \\ &= \mathbf{P}\left(X \leq \frac{9}{5}y + 32\right) \\ &= F_X\left(\frac{9}{5}y + 32\right). \end{aligned}$$

The derivative of cdf gives pdf:

$$\begin{aligned} f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_X\left(\frac{9}{5}y + 32\right) \\ &= f_X\left(\frac{9}{5}y + 32\right) \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{9}{5}y + 32\right) && \left[\textit{Chain Rule}\right] \\ &= \frac{9}{5} f_X\left(\frac{9}{5} + 32\right). \end{aligned}$$

□

3.3 Expected Values and Variances

Definition 3.3.1 (Expected Values). For discrete random variables,

$$\mathbf{E}(X) = \sum_{\text{all values of } x} x \mathbf{P}_X(x) = \sum_{\text{all values of } x} x \mathbf{P}(X = x).$$

For continuous random variables,

$$\mathbf{E}(Y) = \int_{-\infty}^{\infty} y f(y)\,\mathrm{d}y$$

Remark. $\mathrm{E}(X) = \mu_X = \mu$ *is the balancing point of the distribution, also known as the* ***first moment****. Another center of the distribution is the* ***median****,* m*, such that*

$$\mathbf{P}(X \geq m) = \mathbf{P}(X < m) = \frac{1}{2}.$$

Theorem 3.3.2 Properties of Expected Values

- Let $Y = g(X)$. Then,

$$\mathbf{E}(Y) = \mathbf{E}(g(X)) = \begin{cases} \displaystyle\sum_{\text{all } x} g(x)\mathbf{P}_X(x), & \text{discrete} \\ \displaystyle\int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x, & \text{continuous} \end{cases}$$

- Special Case: $Y = aX + b$.

$$\mathbf{E}(Y) = \mathbf{E}(aX + b) = a\mathbf{E}(X) + b.$$

This special case also indicates that $\mathbf{E}(X)$ is linear.

Definition 3.3.3 (Variance). The width of a distribution can be described by the variance. The *variance* is the *second centered moment*:

$$\mathbf{Var}(X) = \mathbf{E}\big((X - \mu)^2\big)$$

Another way to write variance is

$$\mathbf{Var}(X) = \mathbf{E}\big(X^2\big) - \mathbf{E}(X)^2$$

Theorem 3.3.4 Properties of Variance

- Variance is not linear.
- Variance is translation invariant:

$$\mathbf{Var}(X + b) = \mathbf{Var}(X).$$

Remark. *Variance of a line:*

$$\mathbf{Var}(aX + b) = a^2\mathbf{Var}(X)$$

Proof 1.

$$\begin{aligned}
\mathbf{Var}(aX+b) &= \mathbf{E}\big((aX+b)^2\big) - \mathbf{E}(aX+b)^2 \\
&= \mathbf{E}\big(a^2X^2 + b^2 + 2abX\big) - (a\mathbf{E}(X)+b)^2 \\
&= a^2\mathbf{E}\big(X^2\big) + b^2 + 2ab\mathbf{E}(X) - a^2\mathbf{E}(X)^2 - b^2 - 2ab\mathbf{E}(X) \\
&= a^2\big(\mathbf{E}\big(X^2\big) - \mathbf{E}(X)^2\big) \\
&= a^2\mathbf{Var}(X)
\end{aligned}$$

■

3.4 Joint Densities

Definition 3.4.1 (Joint pdf of X, Y).

- Discrete: $p_{X,Y}(k_1, k_2) = \mathbf{P}(X = k_1, Y = k_2)$.
- Continuous: $f_{X,Y} = \mathbf{P}(X, Y \in R) = \iint_R f_{x,Y}(x,y)\,\mathrm{d}A$.

Definition 3.4.2 (Marginal pdf). We can recover the single variable pdf from the joint pdf. This is called the *marginal pdf.*

- Discrete: $p_X(x) = \mathbf{P}(X = x, Y = \text{anything})$.
- Continuous: $\mathbf{P}(a \le X \le b) = \int_a^b \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y\mathrm{d}x. \implies f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y$.

Theorem 3.4.3 Independence
Two random variable X, Y are *independent* if

- Discrete: $p_{X,Y}(x,y) = \mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y) = p_X(x)p_Y(y)$.
- Continuous: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Proof 1. We will prove the continuous case using the cdf $\to$ pdf method. If X and Y are independent, the events $X \in [a,b]$ and $Y \in [c,d]$ are independent $\forall a, b, c, d$. Then,

$$\begin{aligned}
F_{X,Y}(x,y) = \mathbf{P}(X \le x, Y \le y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\,\mathrm{d}t\mathrm{d}s \\
&= \int_{-\infty}^{x} f_X(s) \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t\mathrm{d}s \qquad \big[Independence\big]
\end{aligned}$$

So,

$$f_{x,y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad \big[Fund.\ Thm.\ of\ Calculus\big]$$

■

Corollary 3.4.4 Independent random variables must have a rectangle domain.

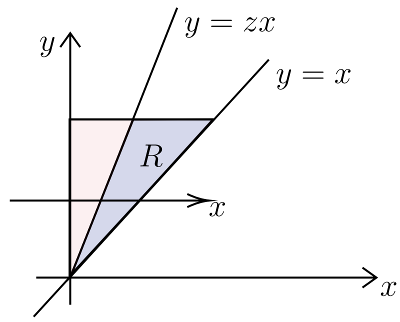
Example 3.4.5 Given $X < 0$, $Y < 0$, there's a chance that they can be independent. However, given $0 < X \leq Y \leq 1$, there is no chance that they will be independent.

Example 3.4.6 Suppose X and Y have the joint pdf

$$f_{X,Y}(x,y) = 2 \quad 0 < x \leq y \leq 1$$

Find the pdf of $Z = \dfrac{Y}{X}$.

Solution 2.



$$\mathbf{P}_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}\left(\frac{Y}{X} \leq z\right) = \mathbf{P}(Y \leq zX) = \iint_R f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y$$
$$= \int_0^1 \int_{y/z}^{y} 2\,\mathrm{d}x\mathrm{d}y$$
$$= \int_0^1 2\left(y - \frac{y}{z}\right)\mathrm{d}y$$
$$= \int_0^1 2\left(1 - \frac{1}{z}\right) y\,\mathrm{d}y$$
$$= 2 \cdot \left(1 - \frac{1}{z}\right) \cdot \frac{1}{2}\left[y^2\right]_0^1$$
$$= 1 - \frac{1}{2}$$

Therefore,

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(1 - \frac{1}{z}\right) = z^{-2}, \quad z > 1.$$

□

3.5 Combining Independent Random Variables

Remark. *First check for dependencies of the domain: We need rectangular relationship between X and Y.*

Theorem 3.5.1 Review

- Joint cdf: $F_{X,Y}(x,y) = \mathbf{P}(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\,\mathrm{d}t\mathrm{d}s$
- If $Y = aX + b$, then
$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Theorem 3.5.2 Strategies for Finding New pdfs

- Discrete pdf: compute the probability directly
- Continuous pdf: use the cdf $\rightarrow$ pdf method.

Example 3.5.3 Let $X \sim \text{Binomial}(n,p)$ and $Y \sim \text{Binomial}(m,p)$. Let X and Y be independent. Find the pdf for $W = X + Y$.

Solution 1.

$$\mathbf{P}_X(x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad \mathbf{P}_Y(y) = \binom{m}{y} p^y (1-p)^{m-y}$$

$$\begin{aligned}
\mathbf{P}_W(w) = \mathbf{P}(W = w) &= \sum_x \underbrace{\mathbf{P}_X(x)\mathbf{P}_Y(w-x)}_{\text{Joint pdf}\mathbf{P}_{X,Y}(x,y),\ w=x+y} \\
&= \sum_x \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{w-x} p^{w-x}(1-p)^{m-(w-x)} \\
&= \sum_x \binom{n}{x}\binom{m}{w-x} p^{x+w-x}(1-p)^{n-x+m-w+x} \\
&= \sum_x \binom{n}{x}\binom{m}{w-x} p^{w}(1-p)^{n+m-w} \\
&= p^w(1-p)^{n+m-w}\left(\sum_x \binom{n}{x}\binom{m}{w-x}\right) \\
&= \binom{m+n}{w} p^w (1-p)^{n+m-w}
\end{aligned}$$

So, $W = X + Y \sim \text{Binomial}(n+m, p)$ $\square$

Theorem 3.5.4 Sum of Continuous Independent Random Variable, $W = X + Y$

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\,\mathrm{d}x$$

Proof 2. By independence, we have

$$\begin{aligned} F_W(w) &= \mathbf{P}(W \leq w) \\ &= \mathbf{P}(X + Y \leq w) \\ &= \mathbf{P}(Y \leq w - X) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_X(x) f_Y(y) \, \mathrm{d}y \mathrm{d}x \end{aligned}$$

Then, by Fundamental Theorem of Calculus, we have

$$\begin{aligned} f_W(w) &= \frac{\mathrm{d}}{\mathrm{d}w} \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_X(x) f_Y(y) \, \mathrm{d}y \mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_X(x) \frac{\mathrm{d}}{\mathrm{d}w} \int_{-\infty}^{w-x} f_Y(y) \, \mathrm{d}y \, \mathrm{d}x \\ \text{convolution} &= \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)(1) \, \mathrm{d}x \end{aligned}$$

■

Theorem 3.5.5 Quotient of Continuous Independent Random Variables, $W = \dfrac{Y}{X}$

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx) \, \mathrm{d}x; \quad \frac{\mathrm{d}}{\mathrm{d}w}(wx) = |x|$$

Theorem 3.5.6 Product of Continuous Independent Random Variables, $W = XY$

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{w}{x}\right) \mathrm{d}x, \quad x \neq 0; \quad \frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{w}{x}\right) = \frac{1}{|x|}.$$

3.6 Further Properties of Mean and Variance

Lemma 3.6.1 Assume X, Y are continuous random variable with joint pdf $f_{X,Y}(x, y)$ and let $g(X, Y)$ be a function of random variable X, Y, then

$$\mathbf{E}(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) \, \mathrm{d}x \mathrm{d}y.$$

Theorem 3.6.2 Linearity of Expected Values

$$\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$$

Proof 1. Suppose X and Y are two continuous random variables, we want to examine the sum of their expected values.

$$\begin{aligned}
\mathbf{E}(X+Y) &= \iint_{\mathbb{R}^2} (x+y) f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \iint_{\mathbb{R}^2} \left[x f_{X,Y}(x,y) + y f_{X,Y}(x,y)\right] \mathrm{d}x\mathrm{d}y \\
&= \iint_{\mathbb{R}^2} x f_{X,Y}(x,y)\,\mathrm{d}y\mathrm{d}x + \iint_{\mathbb{R}^2} y f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x + int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} y f_Y(x)\,\mathrm{d}y = \mathbf{E}(X) + \mathbf{E}(Y).
\end{aligned}$$

■

Conjecture 3.6.3 We hope the expected value of a product to be $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.

Disproof 2. Take an urn with two chips numbered 1 and 2. Draw 2 chips without replacement. Let $X_1 =$ the value on draw 1 and $X_2 =$ the value on draw 2.

$$\begin{aligned}
\mathbf{E}(X_1) &= \sum_{i=1}^{2} i\mathbf{P}(X=i) = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) = \frac{3}{2} \\
\mathbf{P}(X_2 = 1) &= \mathbf{P}(X_2 = 1 \mid X_1 = 1)\mathbf{P}(X_1 = 1) + \mathbf{P}(X_2 = 1 \mid X_1 = 2)\mathbf{P}(X_1 = 2) = 0 + 1\left(\frac{1}{2}\right) = \frac{1}{2} \\
\mathbf{P}(X_2 = 2) &= \mathbf{P}(X_2 = 2 \mid X_1 = 1)\mathbf{P}(X_1 = 1) + \mathbf{P}(X_2 = 2 \mid X_1 = 2)\mathbf{P}(X_1 = 2) = 1\left(\frac{1}{2}\right) + 0 = \frac{1}{2} \\
\mathbf{E}(X_2) &= \sum_{i=1}^{2} i\mathbf{P}(X=i) = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) = \frac{3}{2}
\end{aligned}$$

$$\mathbf{E}(X_1)\mathbf{E}(X_2) = \left(\frac{3}{2}\right)\left(\frac{3}{2}\right) = \frac{9}{4}; \quad \mathbf{E}(X_1 X_2) = \mathbf{E}(2) = 2.$$

Therefore, $\mathbf{E}(X_1)\mathbf{E}(X_2) \neq \mathbf{E}(X_1 X_2)$. ■

Theorem 3.6.4 Expected Value of a Product
If X, Y are independent, then

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y).$$

Proof 3. Suppose X and Y are independent. Then,

$$\begin{aligned}
\mathbf{E}(X,Y) &= \iint_{\mathbb{R}^2} xy f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \iint_{\mathbb{R}^2} xy f_X(x) f_Y(y)\,\mathrm{d}x\mathrm{d}y && \textit{Independence} \\
&= \left(\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y\right) = \mathbf{E}(X)\mathbf{E}(Y).
\end{aligned}$$

■

Theorem 3.6.5 Variance of a Sum

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{Cov}(X,Y),$$

where

$$\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y).$$

Specially, if X and Y are independent,

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y),$$

and

$$\mathbf{Cov}(X,Y) = 0.$$

Proof 4. Recall $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$. Then,

$$\begin{aligned}
\mathbf{Var}(X+Y) &= \mathbf{E}\big((X+Y)^2\big) - \mathbf{E}(X+Y)^2 \\
&= \mathbf{E}\big(X^2 + Y^2 + 2XY\big) - [\mathbf{E}(X) + \mathbf{E}(Y)]^2 \\
&= \mathbf{E}\big(X^2\big) + \mathbf{E}\big(Y^2\big) + 2\mathbf{E}(XY) - \mathbf{E}(X)^2 - \mathbf{E}(Y)^2 - 2\mathbf{E}(X)\mathbf{E}(Y) \\
&= \big[\mathbf{E}\big(X^2\big) - \mathbf{E}(X)^2\big] + \big[\mathbf{E}\big(Y^2\big) - \mathbf{E}(Y)^2\big] + 2[\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)] \\
&= \mathbf{Var}(X) + \mathbf{Var}(Y) + 2[\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)] \\
&= \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{Cov}(X,Y)
\end{aligned}$$

When X and Y are independent, $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$, so $\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0$. Then,

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y).$$

■

Remark. $\mathbf{Cov}(X,Y)$ *does not imply* X *and* Y *are independent. When* $\mathbf{E}(X)$ *or* $\mathbf{E}(Y)$ *is* 0*, we would have* $\mathbf{Cov}(X,Y) = 0$ *as well.*

Remark. *Summary of Properties:*

- *Always true:*
  - $\mathbf{E}(aX+b) = a\mathbf{E}(X) + b$
  - $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$
  - $\mathbf{Var}(aX+b) = a^2\mathbf{Var}(X)$
  - $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$
- *When* X *and* Y *are independent:*
  - $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$
  - $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$

3.7 Order Statistics

Definition 3.7.1 (Median). The point where half data lies above and half below.

Example 3.7.2 When all the GPAs are ordered, a number is assigned by placement in the list

Definition 3.7.3 (Max/Min). The largest and smallest values.
Definition 3.7.4 (Percentiles). A score that tells you the percent of people that you scored better than.
Definition 3.7.5 (i-th Order Statistics). Let Y be a continuous random variable for which we have drawn a *random sample* (independent and identically distributed/ *i.i.d.*), say we have $y_1, y_2, \ldots, y_n$. We re-order them from smallest to largest:

$$y_{\text{min}} = y_1' \leq y_2' \leq \cdots \leq y_n' = y_{\text{max}}$$

Define a new random variable Y_i', and Y_i' is called the i*-th order statistics*. Given a random sample $Y_1, \ldots, Y_n$, we define

$$Y_{\text{min}} = \min{(Y_1, \ldots, Y_n)}; \quad Y_{\text{max}} = \max{(Y_1, \ldots, Y_n)}.$$

Definition 3.7.6 (Percentiles). For any value p between 0 and 1, the $100p$-th percentile is the observation such that np observations are less than that value, and $n(1-p)$ are greater.

Example 3.7.7 Suppose a random sample $y_1 = 3.2,\ y_2 = 4,\ y_3 = 1.1,\ y_4 = 0$. Then,

$$0 \leq 1.1 \leq 3.2 \leq 4.$$

So, $Y_{\text{min}} = 0 = y_{\text{min}}$, $y_2' = 1.1$, $y_3' = 3.2$, and $y_{\text{max}} = 4 = Y_{\text{max}}$.

Example 3.7.8 Let $Y_1, \ldots, Y_n$ be a random sample from $Y \sim$ Uniform$(0, a)$ where are do not know a. We can estimate a by different methods.

$\boxed{\text{Method } 1}$ Use the sample mean:

$$\mathbf{E}(Y) = \mu = \frac{a}{2}$$

So we solve for $\hat{a} \approx 2\mu \approx 2\overline{Y}$. Then, we can estimate a with $\overline{Y}$. However, we might have $y_{\text{max}} \geq 2\overline{Y}$ which leads to a better method.

$\boxed{\text{Method } 2}$ Use observed y_{max}. We know the following inequality must hold:

$$a \geq Y_{\text{max}}.$$

To find $\mathbf{E}(Y_{\text{max}})$, we first need to find the pdf of Y_{max}. We consider the cdf→pdf method:

$$\begin{aligned} F_{Y_{\text{max}}}(y) = \mathbf{P}(Y_{\text{max}} \leq y) &= \mathbf{P}(Y_1 \leq y_1, Y_2 \leq y_2, Y_3 \leq y_3, \ldots, Y_n \leq y_n) \\ &= \mathbf{P}(Y_1 \leq y_1)\mathbf{P}(Y_2 \leq y_2) \cdots \mathbf{P}(Y_n \leq y_n) \\ &= \left(\frac{y}{a}\right)^n \end{aligned}$$

So, we know

$$f_{Y_{\text{max}}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{Y_{\text{max}}}(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left[\left(\frac{y}{a}\right)^n\right] = \frac{n}{a}\left(\frac{y}{a}\right)^{n-1}.$$

So, we can find the expected value:

$$\begin{aligned} \mathbf{E}(Y_{\text{max}}) = \int_0^a y f_{Y_{\text{max}}}(y)\,\mathrm{d}y &= \int_0^a y \frac{n}{a}\left(\frac{y}{a}\right)^{n-1} \mathrm{d}y \\ &= \frac{n}{a^n}\int_0^a y^n\,\mathrm{d}y \\ &= \frac{n}{a^n}\left[\frac{1}{n+1}y^{n+1}\right]_0^a \\ &= \frac{n}{a^n} \cdot \frac{1}{n+1} a^{n+1} \\ &= \frac{n}{n+1} a \end{aligned}$$

Therefore, we can estimate $y_{\text{max}} = \dfrac{n}{n+1} a$. So we get $a \approx \dfrac{n+1}{n} y_{\text{max}}$.

Theorem 3.7.9 Order Statistics

$$Y_{\text{max}}: \ F_{Y_{\text{max}}}(y) = F_y(y)^n \implies f_{Y_{\text{max}}}(y) = nF_Y(y)^{n-1} f_Y(y).$$

$$Y_{\text{min}}: \ F_{Y_{\text{min}}}(y) = 1 - [1 - F_y(y)]^n \implies f_{Y_{\text{min}}}(y) = n[1 - F_Y(y)]^{n-1} f_Y(y).$$

Proof 1. The formula of Y_{min} will proven here.

$$\begin{aligned} F_{Y_{\text{min}}}(y) = \mathbf{P}(Y_{\text{min}} \leq y) &= 1 - \mathbf{P}(Y_{\text{min}} > y) \\ &= 1 - \mathbf{P}(Y_1 > y, Y_2 > y, \ldots, Y_n > y) \\ &= 1 - \mathbf{P}(Y_1 > y)\mathbf{P}(Y_2 > y) \cdots \mathbf{P}(Y_n > y) \\ &= 1 - [1 - F_Y(y)][1 - F_Y(y)] \cdots [1 - F_Y(y)] \\ &= 1 - [1 - F_Y(y)]^n \end{aligned}$$

So, we would also have $f_{Y_{\text{min}}}(y) = n[1 - F_Y(y)]^{n-1} f_Y(y)$. ■

Theorem 3.7.10 i-th Order Statistics

$$Y_i': \ f_{Y_i'}(y) = \frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1 - F_Y(y)]^{n-i} f_Y(y), \quad 1 \leq i \leq n.$$

3.8 Conditional Densities

Definition 3.8.1 (Conditional Probability). Let X, Y be discrete random variables with joint pdf $p_{X,Y}(x,y)$, then the conditional probability

$$p_{Y|X}(y) = \mathbf{P}(Y = y \mid X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)}.$$

Example 3.8.2 A bag with 5 fair coins and one 2 headed coin. Choose a coin and flip it n times. Let $X = \begin{cases} 1 & \text{if the coin is fiar} \\ 2 & \text{if the two headed coin is selected} \end{cases}$ and $Y =$ the number of heads in n tosses. Find $\mathbf{P}_Y(y)$.

Solution 1.

Our plan: find $p_X(x) \implies p_{Y|X=1}(y),\ p_{Y|X=2}(y) \implies p_{X,Y}(x,y) = p_{Y|X}(y)p_X(x) \implies$ sum over all values of x to get the $p_Y(y)$.

$$p_X(x) = \begin{cases} \frac{5}{6} & X = 1 \\ \\ \frac{1}{6} & X = 2 \end{cases}$$

$$p_{Y|1}(y) = \binom{n}{y}\left(\frac{1}{2}\right)^y\left(1 - \frac{1}{2}\right)^{n-y} = \binom{n}{y}\left(\frac{1}{2}\right)^n$$

$$p_{Y|2}(y) = \begin{cases} 0 & y < n \\ 1 & y = n \end{cases}.$$

Therefore, the joint pdf

	$Y = 0$	$Y = 1$	$Y = k$	$Y = n$
$X = 1$	$\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$\binom{n}{1}\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$\binom{n}{k}\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$1\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$
$X = 2$	0	0	0	$1\left(\frac{1}{6}\right)$
$p_Y(y)$	$\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$\binom{n}{1}\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$\binom{n}{k}\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right)$	$\left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)$

In other words,

$$p_Y(y) = \begin{cases} \left(\frac{5}{6}\right)\binom{n}{y}\left(\frac{1}{2}\right)^n & 0 \le y < n \\ \\ \left(\frac{1}{2}\right)^n\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right) & y = n \end{cases}$$

□

Remark. *If we define the conditional density in the same way as in the discrete case, we shall have*

$$\frac{\mathbf{P}(X=x,Y=y)}{\mathbf{P}(X=x)}.$$

However, $\mathbf{P}(X=x)=0$ *when* X *is continuous. Thus, we need an alternative definition when* X *is continuous.*

Definition 3.8.3 (Conditional Probability - Continuous). Let X and Y be two continuous random variables. Then, the conditional probability

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

Proof 2. We will use the cdf→pdf method. Define the cdf of the conditional probability to be

$$\begin{aligned}
F_{Y|X}(y) &= \lim_{h\to 0}\mathbf{P}(Y\le y \mid X\in[x,x+h]) \\
&= \lim_{h\to 0}\frac{\displaystyle\int_x^{x+h}\int_{-\infty}^{y}\mathbf{P}_{X,Y}(s,t)\,\mathrm{d}t\mathrm{d}s}{\displaystyle\int_x^{x+h} f_X(s)\,\mathrm{d}x} && \frac{0}{0}\implies L'Hopitals' \\
&= \lim_{h\to 0}\frac{\displaystyle\frac{\mathrm{d}}{\mathrm{d}h}\int_x^{x+h}\int_{-\infty}^{y}\mathbf{P}_{X,Y}(s,t)\,\mathrm{d}t\mathrm{d}s}{\displaystyle\frac{\mathrm{d}}{\mathrm{d}h}\int_x^{x+h} f_X(s)\,\mathrm{d}s} \\
&= \lim_{h\to 0}\frac{\displaystyle\int_{-\infty}^{y}\mathbf{P}_{X,Y}(x+h,t)\,\mathrm{d}t}{f_X(x+h)} \\
&= \frac{\displaystyle\int_{-\infty}^{y}\mathbf{P}_{X,Y}(x,t)\,\mathrm{d}t}{f_X(x)}
\end{aligned}$$

So, we know

$$\begin{aligned}
f_{Y|X}(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\frac{\displaystyle\int_{-\infty}^{y}\mathbf{P}_{X,Y}(x,t)\,\mathrm{d}t}{f_X(x)} \\
&= \frac{f_{X,Y}(x,y)}{f_X(x)}.
\end{aligned}$$

■

Example 3.8.4 A stick of unit length, and break it at a random point. Let $X=$ length of the larger piece. Then, break the larger piece, and let $Y=$ the length of the larger piece after the second break. Find $\mathbf{E}(Y)$.

Solution 3.

Find the pdf of X: let $W =$ the breaking point. Assume $W \sim \text{Uniform}(0,1)$. Then, $X = \max(W, 1-W)$. So we have

$$\begin{aligned}
F_X(x) = \mathbf{P}(X \le x) &= \mathbf{P}(W \le x \text{ and } 1-x \le W) \\
&= \mathbf{P}(W \in [1-x, x]) && X \in \left[\frac{1}{2}, 1\right] \\
&= \int_{1-x}^{x} 1 \,\mathrm{d}w = 2x - 1 \\
f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x}[2x-1] = 2.
\end{aligned}$$

Find the joint pdf: repeat the same process for $Y \mid X \sim \text{Uniform}\left(\frac{X}{2}, X\right)$. We have

$$\begin{aligned}
f_{Y|X}(y) &= \frac{1}{x - \frac{x}{2}} = \frac{2}{x} && y \in \left[\frac{x}{2}, x\right] \\
f_{X,Y}(x,y) &= f_{Y|X}(y) f_X(x) \\
&= \frac{2}{x} \cdot 2 = \frac{4}{x}. && x \in \left[\frac{1}{2}, 1\right],\ y \in \left[\frac{x}{2}, x\right]
\end{aligned}$$

Now, we can find the expected value:

$$\begin{aligned}
\mathbf{E}(Y) &= \int_{1/2}^{1} \int_{x/2}^{x} y f_{X,Y}(x,y) \,\mathrm{d}y\mathrm{d}x \\
&= \int_{1/2}^{1} \int_{x/2}^{x} y \frac{4}{x} \,\mathrm{d}y\mathrm{d}x \\
&= \int_{1/2}^{1} \frac{4}{x} \left[\frac{1}{2} y^2\right]_{x/2}^{x} \mathrm{d}x \\
&= \int_{1/2}^{1} \frac{2}{x} \left(x^2 - \frac{x^2}{4}\right) \mathrm{d}x \\
&= \int_{1/2}^{1} \frac{3}{2} x \,\mathrm{d}x \\
&= \frac{3}{2} \left[\frac{1}{2} x^2\right]_{1/2}^{1} \\
&= \frac{3}{4}\left(1 - \frac{1}{4}\right) = \frac{9}{16}.
\end{aligned}$$

□

3.9 Moment Generating Functions (mgf)

Definition 3.9.1 (Moment). The moments of the distribution of X are

$$\mathbf{E}(X^k) \quad \text{for } k = 1, 2, \ldots$$

if they exists. The mean μ is the first moment.

Definition 3.9.2 (Moment Generating Functions/mgf).

$$\mathbf{M}_X(t) = \mathbf{E}\big(e^{tX}\big) = \begin{cases} \displaystyle\sum_k e^{tk} p_X(k) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

Remark. *Why we would use* $\mathbf{E}\big(e^{tX}\big)$*?*

Recall the Macluarin Series expansion for

$$e^{tX} = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$$

Then, we have

$$\begin{aligned} \mathbf{M}_X(t) = \mathbf{E}\big(e^{tX}\big) = \mathbf{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) &= \sum_{n=0}^{\infty} \mathbf{E}\left(\frac{t^n x^n}{n!}\right) \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{E}(X^n) \\ &= 1 + t\mathbf{E}(X) + \frac{t^2}{2!}\mathbf{E}\big(X^2\big) + \frac{t^3}{3!}\mathbf{E}\big(X^3\big) + \cdots \\ &= 1 + tm_1 + \frac{t^2}{2!}m_2 + \frac{t^3}{3!}m_3 + \cdots \\ &= 1 + m_1 t + \frac{m_2}{2!}t^2 + \frac{m_3}{3!}t^3 + \cdots \end{aligned}$$

Example 3.9.3 Recover m_1 from the moment generating function.

$\boxed{\text{Step 1}}$ Differentiate $\mathbf{M}_X(t)$ with respect to t.

$$\mathbf{M}_X(t) = 1 + m_1 t + \frac{m_2}{2!}t^2 + \frac{m_3}{3!}t^3 + \cdots$$

$$\mathbf{M}_X^{(1)}(t) = 0 + m_1 + \frac{2m_2}{2!}t + \frac{3m_3}{3!}t^2 + \cdots = m_1 + m_2 t + \frac{m_3}{2!}t^2 + \cdots$$

$\boxed{\text{Step 2}}$ Evaluate $t = 0$: Higher order terms drop out when $t = 0$.

$$\mathbf{M}_X^{(1)}(0) = m_1$$

Theorem 3.9.4

$$\mathbf{M}_Y^{(n)}(0) = \mathbf{E}(Y^n) \quad \text{as long as } \mathbf{E}(Y^n) < \infty.$$

Example 3.9.5 Recover m_2:

$$\mathbf{M}_X^{(2)}(t) = 0 + m_2 + \frac{2m_3}{2!}t + \cdots = m_2 + m_3 t + \cdots \implies \mathbf{M}_X^{(2)}(0) = m_2.$$

Theorem 3.9.6
If Y_1 and Y_2 have $\mathbf{M}_{Y_1}(t) = \mathbf{M}_{Y_2}(t)$ for an open interval of t about 0, then the pdf's are identical. i.e., $f_{Y_1}(y) = f_{Y_2(y)}$.

Theorem 3.9.7 Properties of mgf

- If $Y = aX + b$, then
$$\mathbf{M}_Y(t) = e^{bt}\mathbf{M}_X(at)$$
- If $W = X + Y$ and X, Y are independent, then
$$\mathbf{M}_W(t) = \mathbf{M}_X(t) \cdot \mathbf{M}_Y(t).$$
This is much easier than computing $f_W(w) = f_X \cdot f_Y$.

Example 3.9.8 Let $Y \sim N(0, 1)$ (Normal with $\mu = 0$ and $\sigma^2 = 1$). Find the mgf for $W \sim N(\mu, \sigma^2)$.

Solution 1.
We have that $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}, \quad y \in (-\infty, \infty)$.

$$\begin{aligned}
\mathbf{M}_Y(t) = \mathbf{E}\big(e^{tY}\big) = \int_{-\infty}^{\infty} e^{ty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,\mathrm{d}y &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ty-y^2/2}\,\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2+t^2/2}\,\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2}e^{t^2/2}\,\mathrm{d}y \\
&= e^{t^2/2}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2}\,\mathrm{d}y}_{\text{pdf of } Y \text{ shift by } t \text{ unit}=1} \\
&= e^{t^2/2}
\end{aligned}$$

$W = \sigma Y + \mu$, then by the property of mgf, we have

$$\mathbf{M}_W(t) = e^{\mu t}\mathbf{M}_Y(\sigma t) = e^{\mu t}e^{(\sigma t)^2/2} = e^{\frac{1}{2}\sigma^2 t^2 + \mu t}.$$

□

4 Special Distributions

4.1 Poisson Distribution

Example 4.1.1 A publisher estimates that their books have an average 1 typo every 4 pages. Let X = number of typos in a 20 page chapter. Assume: (1) the typos are equally likely to occur anywhere, (2) different typos are independent, and (3) average rate of the typos is $\frac{1}{4}$ per page.

a. Let X_j = # of typos on page j.

Solution 1.

Expect $\frac{1}{4}$ per page $\implies p_X(k) = \begin{cases} 1/4 & k=1 \\ 3/4 & k=0 \end{cases}$.

Then, $X_j \sim$ Binomial$\left(n=20, p=\frac{1}{4}\right)$. So,

$$\mathbf{P}(X=k) = \binom{20}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{20-k}$$

Therefore,

$$\mathbf{P}(X=5) = \binom{20}{5}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^{15} \approx 20.2\%$$

Problem: maximum # of typo is 20, but we can have more typos per page! □

b. Subdivide each page into 40 lines. $\implies n = 800$ total lines and $p = \frac{1/4}{40} = \frac{1}{160}$.

Solution 2.

Let X_j = # of typos on line j. Then, $X_j \sim$ Binomial$\left(n=800, p=\frac{!}{160}\right)$. So,

$$p_{X_j}(k) = \binom{800}{k}\left(\frac{1}{160}\right)^k\left(\frac{159}{160}\right)^{800-k}, \quad k=0,\dots,800$$

$$\mathbf{P}(X_j=5) = 17.6\%$$

□

c. Subdivide each line into quarter lines. That is, $n = 3200$ quarter lines and $p = \frac{1}{640}$.

Solution 3.

X_j = a typo on quarter line j. So, $X_j \sim$ Binomial$\left(3200, \frac{1}{640}\right)$.

$$p_{X_j}(k) = \binom{3200}{k}\left(\frac{1}{640}\right)^k\left(\frac{639}{640}\right)^{3200-k}.$$

□

d. Does this approach a limit? Let $\lambda =$ total # expected, $n =$ # of subdivisions, $p = \frac{\lambda}{n}$ probability within one division. Then, $p_X(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$. Fix k and λ. Find $\lim_{n\to\infty} p_X(k)$.

Solution 4.

$$
\begin{aligned}
\lim_{n\to\infty} p_X(k) &= \lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{(n-k)!k!}\cdot\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!}\cdot\frac{1}{n^k\left(1-\frac{\lambda}{n}\right)^k}\cdot\left(1-\frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)(n-k)!}{(n-k)!(n-\lambda)^k}\cdot\left(1-\frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!}\lim_{n\to\infty}\underbrace{\frac{n(n-1)\cdots(n-k+1)}{(n-\lambda)^k}}_{\sim\lim_{n\to\infty}\frac{n^k}{n^k}=1}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{=\lim_{n\to\infty}e^{-\lambda}} \\
&= \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}
$$

$\square$

Definition 4.1.2 (Poisson Distribution). For $X \in \mathbb{N}$ and $\lambda > 0$, the pdf of *poisson distribution* is given by

$$p_X(k) = P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad \text{for } k = 0, 1, 2, \ldots$$

Definition 4.1.3 (Poisson Model). Assume the events occur with the following assumptions:

- Can subdivide into intervals small enough that $\mathbf{P}(2 \text{ events in one interval}) = 0$.
- Occurrences in different intervals are independent.
- Probability of event is constant.

Then, set $\lambda =$ expected # of occurrence per unit and $X =$ actual # of occurrence per unit. Then, we have the *poisson model*:

$$\mathbf{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Example 4.1.4 Application of Poisson Distribution

1. Radioactive decay (or other discrete reactions).
2. Errors/flaws/accidents injuries
3. Outbreaks of non-contagious disease.

Example 4.1.5 Wait time Between Poisson Distribution

Suppose we use a geiger counter to measure the radiation coming off a radioactive sample. The number of ticks in 1 second of the counter has a poisson distribution, where λ is the expected number of ticks, i.e., $\lambda =$ average number of ticks per second. Let $Y =$ the time interval between two consecutive ticks. What is the distribution of Y?

Solution 5.

The cdf of Y: $F_Y(y) = \mathbf{P}(Y \leq y)$.

Let $t = 0$ be the time of the first click. Then, $Y \leq y$ can be interpreted as at least 1 click in $[0, y]$. So, we have

$$F_Y(y) = 1 - \mathbf{P}(\text{No clicks in } [0, y]).$$

The expected number of clicks is $\lambda y \implies \mathbf{P}(\text{none}) = e^{-\lambda y}$. Thus,

$$F_Y(y) = 1 - e^{-\lambda y}.$$

So, the pdf of Y is given by

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \lambda e^{-\lambda y}, \quad y \geq 0$$

□

Theorem 4.1.6 Wait Time Between Poisson Distribution

Suppose a series of events satisfying the poisson distribution model are occurring at the rate of λ units per time. Let random variable Y denote the interval between consecutive events. Then, Y has the exponential distribution: $Y \sim \text{Exponential}(\lambda)$, and

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0.$$

4.2 Normal Distribution

Definition 4.2.1 (The Standard Normal). A continuous random variable $Z \sim N(0,1) = \text{Normal}(\mu = 0, \sigma^2 = 1)$ is defined to be the standard normal distribution with pdf

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in (-\infty, \infty)$$

and the moment generating function

$$\mathbf{M}_Z(t) = e^{t^2/2}, \quad t \in \mathbb{R}$$

The standard normal distribution is also known as the Gaussian distribution or the bell curve.

Definition 4.2.2 (The General Normal Distribution). We define the general normal distribution $X \sim N(\mu, \sigma^2)$ by using the transformation $X = \sigma Z + \mu$. Any normal distribution is a linear transformation of the standard normal, and the pdf is given by

$$f_X(x) = \frac{1}{\sqrt{a\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}$$

Proof 1. Note that the cdf of $X = \sigma Z + \mu$ is given by

$$\begin{aligned} F_X(x) = \mathbf{P}(X \leq x) &= \mathbf{P}(\sigma Z + \mu \leq x) \\ &= \mathbf{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

So, the pdf of X is

$$\begin{aligned} f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_Z\left(\frac{x-\mu}{\sigma}\right) \\ &= f_Z\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{\sigma} - \frac{\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2 \cdot \frac{1}{2}} \left(\frac{1}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R} \end{aligned}$$

■

Remark. *Using the MGF that if* $\text{Normal} + \text{Normal} = \text{Normal}$.

Example 4.2.3 Applications of Normal Distribution

- Sample means
- Population features
- Lab instrument measures

Example 4.2.4 Why $\dfrac{1}{\sqrt{2\pi}}$?

Solution 2.

We want to compute $\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x$. Instead of finding it directly, we can find

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x \int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\mathrm{d}y.\end{aligned}$$

Use the polar coordinate: $x = r\cos\theta$ and $y = r\sin\theta$. So, we have $\mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta$. So,

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)^2 &= \int_0^{2\pi}\int_0^{\infty} e^{-\left(r^2\cos^2\theta+r^2\sin^2\theta\right)/2} r\,\mathrm{d}r\mathrm{d}\theta\\ &= \int_0^{2\pi}\int_0^{\infty} e^{-\frac{1}{2}r^2\left(\cos^2\theta+\sin^2\theta\right)} r\,\mathrm{d}r\mathrm{d}\theta\\ &= \int_0^{2\pi}\int_0^{\infty} re^{-\frac{1}{2}r^2}\,\mathrm{d}r\mathrm{d}\theta && u=\frac{1}{2}r^2,\quad \mathrm{d}u = r\mathrm{d}r\\ &= \int_0^{\pi}\int_0^{\infty} e^{-u}\,\mathrm{d}u\mathrm{d}\theta\\ &= \int_0^{2\pi}\lim_{t\to\infty}\left[-e^{-u}\right]_0^t\,\mathrm{d}\theta\\ &= \int_0^{2\pi}\lim_{t\to\infty}\left(-e^{-t}+e^{-0}\right)\mathrm{d}\theta\\ &= \int_0^{2\pi}\mathrm{d}\theta\\ &= \Big[\theta\Big]_0^{2\pi} = 2\pi\end{aligned}$$

Therefore, $\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}$. Thus, to ensure the area under the pdf goes to 1, we need the factor $\dfrac{1}{\sqrt{2\pi}}$. □

Theorem 4.2.5 Central Limit Theorem

Let $Y_1,\ldots,Y_n$ be a random sample with $\mu = \mathbf{E}(Y_i)$ and $\sigma^2 = \mathbf{Var}(Y_i)$, both finite. Then, if

$$Z_n = \frac{1}{\sqrt{n}\sigma}(Y_1+\cdots+Y_n-n\mu),$$

we have

$$\mathbf{P}(a\le Z_n\le b) = \int_a^b \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z \quad \text{as } n\to\infty$$

That is, Z_n approaches the standard normal as $n\to\infty$.

Proof 3. Tools used in this proof: MGF determines the distribution. If $\mathbf{M}_{Z_n}(t)\to\mathbf{M}_Z(t)$, then $\mathbf{P}(a \leq Z_n \leq b) \to \mathbf{P}(a \leq Z \leq b)$.

Given $Y_1, Y_2, \ldots, Y_n$ a random sample as above. Let's define

$$W_i = \frac{1}{\sigma}(Y_i - \mu).$$

Then,

$$\mathbf{E}(W_i) = 0 \qquad \text{and} \qquad \mathbf{Var}(W_i) = 1$$

Assume the MGF of W_i is well-defined. Without loss of generality, consider the MGF of W_i:

$$\begin{aligned}
\mathbf{M}_{W_i}(t) &= e^{\varphi(t)} \\
\mathbf{M}_{W_i}(0) &= e^{\varphi(0)} = 1 && \textit{by well defined} \\
\mathbf{M}'_{W_i}(t) &= e^{\varphi(t)}\varphi'(t) \\
\mathbf{M}'_{W_i}(t) &= e^{\varphi(0)}\varphi'(0) = \mathbf{E}(W_i) = 0 \\
1 \cdot \varphi'(0) &= 0 \implies \varphi'(0) = 0 \\
\mathbf{M}''_{W_i}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left[e^{\varphi(t)}\varphi'(t)\right] \\
&= e^{\varphi(t)}\varphi'(t)\varphi'(t) + e^{\varphi(t)}\varphi''(t) \\
&= e^{\varphi(t)}\left[(\varphi'(t))^2 + \varphi''(t)\right] \\
\mathbf{M}''_{W_i}(0) &=^{\varphi(0)}\left[(\varphi'(0))^2 + \varphi''(0)\right] \\
&= 1 \cdot (0 + \varphi''(0)) = \varphi''(0)
\end{aligned}$$

So,

$$\mathbf{Var}(W_i) = \mathbf{E}\big(W_i^2\big) - \mathbf{E}(W_i)^2 = \mathbf{E}\big(W_i^2\big) - 0 = \mathbf{E}\big(W_i^2\big) = \mathbf{M}''_{W_i}(0) = 1$$
$$\implies \varphi''(0) = 1.$$

Now, define

$$Z_n = \frac{1}{\sqrt{n}}(W_1 + W_2 + \cdots + W_n).$$

By the MGF property that if X and Y are independent, then

$$\mathbf{M}_{aX}(t) = \mathbf{M}_X(at) \quad \text{and} \quad \mathbf{M}_{X+Y}(t) = \mathbf{M}_X(t)\mathbf{M}_Y(t)$$

We have

$$\mathbf{M}_{Z_n}(t) = \prod_{i=1}^{n} \mathbf{M}_{W_i}\big(t/\sqrt{n}\big) = \prod_{i=1}^{n} e^{\varphi\left(t/\sqrt{n}\right)} = \left(e^{\varphi\left(t/\sqrt{n}\right)}\right)^n = e^{n\varphi\left(t/\sqrt{n}\right)}$$

With $x = \dfrac{1}{\sqrt{n}}$, consider

$$\lim_{n\to\infty} n\varphi\big(t/\sqrt{n}\big) = \lim_{n\to\infty} \frac{\varphi(xt)}{x^2} \quad \left(= \frac{0}{0}\right) = \lim_{n\to\infty} \frac{\varphi'(xt)t}{2x} \quad \left(= \frac{0}{0}\right) = \lim_{n\to\infty} \frac{\varphi''(xt)t^2}{2} = \frac{t^2}{2}.$$

That is $\lim\limits_{n\to\infty} \mathbf{M}_{Z_n}(t) = e^{t^2/2} = \mathbf{M}_Z(t)$, the standard normal MGF. ■

Example 4.2.6 Binomial Distribution and the Central Limit Theorem
Consider $X \sim \text{Binomial}(n, p)$. Note that

$$X = X_1 + X_2 + \cdots + X_n, \quad \text{where } X_i \sim \text{Bernoulli}(p).$$

So,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

Example 4.2.7 Poisson Distribution and the Central Limit Theorem
Consider $X \sim \text{Poisson}(\lambda)$. Suppose

$$X = X_1 + X_2 + \cdots + X_n, \quad \text{where } X_i \sim \text{Poisson}\left(\frac{\lambda}{n}\right)$$

Then,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0,1).$$

Remark. *What we are doing here is to approximate discrete distributions (e.g. binomial/poisson) with a continuous distribution (normal). There will be some errors. So we need* continuity correction, *adding or subtracting* $\frac{1}{2}$ *to account for approximations of a discrete distribution with a continuous distribution.*

Example 4.2.8 Binomial Distribution Continuity Correction

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

$$\mathbf{P}(a \le Z \le b) \approx \mathbf{P}\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}} \le Z \le \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right).$$

4.3 CDF Tricks

Remark. (Key idea) *The key idea here is that we can use a cdf to generate a random sample.*

Example 4.3.1 We can let $U \sim \text{Uniform}(0,1)$ and compute $X = F_X^{-1}(U)$. Here we note that

$$F_U(u) = P(U \le u) = u.$$

Then, we have the following:

$$\begin{aligned} F_X(x) = P(X \le x) = P\big(F_X^{-1}(u) \le x\big) &= P\big(F_X\big(F_X^{-1}(U)\big) \le F_X(x)\big) \\ &= P(U \le F_X(x)) \\ &= F_U(F_X(x)) \\ &= F_X(x) \end{aligned}$$

Theorem 4.3.2
We can use a uniform random sample Uniform$(0, 1)$ to simulate a random variable with cdf $F_X(x)$ by setting $X = F_X^{-1}(U)$.

Example 4.3.3 Simulate from the density $f_Y(y) = \dfrac{2\theta^2}{y^3}, \quad y \ge \theta.$

Solution 1.

$\boxed{\text{Step } 1}$ Find the cdf.

$$\begin{aligned} F_Y(y) = P(Y \le y) = \int_\theta^y f_Y(t)\,\mathrm{d}t = \int_\theta^y \frac{2\theta^2}{t^3}\,\mathrm{d}t &= 2\theta^2 \int_\theta^y t^{-3}\,\mathrm{d}t \\ &= 2\theta^2 \left[-\frac{1}{2}t^{-2}\right]_\theta^y \\ &= \theta^2\big(-y^{-2} + \theta^{-2}\big) \\ &= 1 - \frac{\theta^2}{y^2}, \quad y \ge \theta \end{aligned}$$

$\boxed{\text{Step } 2}$ Find the inverse of the cdf:

$$\begin{aligned} u = F_Y(y) \implies u &= 1 - \frac{\theta^2}{y^2} \\ \frac{\theta^2}{y^2} &= 1 - u \\ \frac{1}{y^2} &= \frac{1-u}{\theta^2} \\ y^2 &= \frac{\theta^2}{1-u} \\ y &= \sqrt{\frac{\theta^2}{1-u}} = \frac{\theta}{\sqrt{1-u}} \end{aligned}$$

Now, evaluating the above with a value of u will give us a value of y.

$\boxed{\text{Step } 3}$ Turn it into `R` code. □

4.4 Geometric and Negative Binomial Distributions

Definition 4.4.1 (Geometric Distribution). $X =$ the trial on which the first success occurs. $X \sim$ Geometric(p), and

$$p_X(k) = \underbrace{(1-p)^{k-1}}_{\text{\# of failure}} \underbrace{p}_{\text{success}}, \quad k \geq 1$$

Definition 4.4.2 (Negative Binomial Distribution). $X =$ the number of trials needed for r success. $X \sim$ Negative Binomial(r, p), and

$$p_X(k) = \underbrace{\binom{k-1}{r-1}}_{\text{\# of ways the successes can occur}} \underbrace{p^r}_{r \text{ successes}} \underbrace{(1-p)^{k-r}}_{k-r \text{ failure}}, \quad k \geq r$$

Remark. *Suppose $X =$ trials for r successes. i.e., $x \sim$ Negative Binomial(r, p). Then, we have $X = X_1 + X_2 + \cdots + X_r$, where X_j's are independent and each $X_j \sim$ Geometric(p).*

Theorem 4.4.3 Geometric Expectation
If $X \sim$ Geometric(p), then

$$\mathbf{E}(x) = \frac{1}{p.}$$

Proof 1. Recall $\displaystyle\sum_{k=1}^{\infty} ka^k = \frac{a}{(1-a)^2}$ and $\displaystyle\sum_{k=1}^{\infty} k^2 a^k = \frac{a+a^2}{(1-a)^3}$. So,

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p.$$

Redefine $\widetilde{X} =$ the failures needed to obtain 1 success. Then $\widetilde{X} = X - 1$. Then,

$$\begin{aligned}
\mathbf{E}(\widetilde{X}) = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} k(1-p)^k p = \underbrace{0}_{k=0} + \sum_{k=1}^{\infty} k(1-p)^k \cdot p = p \sum_{k=1}^{\infty} k \underbrace{(1-p)^k}_{a} \\
= p \cdot \frac{(1-p)}{(1-(1-p))^2} \\
= \frac{p(1-p)}{(1-1+p)^2} = \frac{1-p}{p}.
\end{aligned}$$

So,

$$\frac{1-p}{p} = \mathbf{E}(X) - 1 \implies \mathbf{E}(X) = \frac{1-p}{p} + 1 = \frac{1-p+p}{p} = \frac{1}{p}.$$

■

Theorem 4.4.4 Geometric Variance
If $X \sim$ Geometric(p), then

$$\mathbf{Var}(X) = \frac{1-p}{p^2}.$$

Proof 2. Use the same definition of $\widetilde{X}$ as above. Then,

$$\begin{aligned}\mathbf{E}(\widetilde{X}^2) = \sum_{k=0}^{\infty} k^2 p_X(k) = \sum_{k=1}^{\infty} k^2(1-p)^k \cdot p = p\sum_{k=1}^{\infty} k^2(1-p)^k \\ &= p\frac{(1-p+(1-p)^2)}{(1-(1-p))^3} \\ &= \frac{p(1-p+1+p^2-2p)}{(1-1+p)^3} \\ &= \frac{2-3p+p^2}{p^3}\end{aligned}$$

Then,

$$\begin{aligned}\mathbf{Var}(\widetilde{X}) = \mathbf{E}(\widetilde{X}^2) - \mathbf{E}(\widetilde{X})^2 &= \frac{2-3p+p^2}{p^3} - \frac{(1-p)^2}{p^2} \\ &= \frac{2-3p+p^2-1-p^2+2p}{p^3} \\ &= \frac{1-p}{p^2}.\end{aligned}$$

As shifting from X to $\widetilde{X} = X - 1$ does not change the value of variance,

$$\mathbf{Var}(X) = \frac{1-p}{p^2}.$$

■

Theorem 4.4.5 Negative Binomial Expectation and Variance
If $X \sim$ Negative Binomial(r, p), then

$$\mathbf{E}(X) = \frac{r}{p} \quad \text{and} \quad \mathbf{Var}(X) = \frac{r(1-p)}{p^2}.$$

Proof 3. As we have shown $X = X_1 + X_2 + \cdots + X_r$, where X_j's are independent and $X_j \sim$ Geometric(p). Then,

$$\mathbf{E}(X) = \mathbf{E}\left(\sum_{j=1}^{r} X_j\right) = \sum_{j=1}^{r} \mathbf{E}(X_j) = \sum_{j=1}^{r} \frac{1}{p} = \frac{r}{p},$$

and

$$\mathbf{Var}(X) = \mathbf{Var}\left(\sum_{j=1}^{r} X_j\right) = \sum_{j=1}^{r} \mathbf{Var}(X_j) = \sum_{j=1}^{r} \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}.$$

■

Theorem 4.4.6 The MGF of Geometric Distribution
If $X \sim$ Geometric(p), then

$$\mathbf{M}_X(t) = \frac{p}{e^{-t} - 1 + p}.$$

Proof 4.

$$\begin{aligned}
\mathbf{M}_X(t) = \mathbf{E}\left(e^{tX}\right) &= \sum_{k=1}^{\infty} e^{kt}(1-p)^{k-1}p \\
&= \sum_{k=1}^{\infty} e^{t}e^{t(k-1)}(1-p)^{k-1}p \\
&= e^{t}p\sum_{k=1}^{\infty}\left(e^{t}(1-p)\right)^{k-1} \\
&= \frac{e^{t}p}{1-e^{t}(1-p)}\cdot\frac{e^{-t}}{e^{-t}} \\
&= \frac{e^{-t}e^{t}p}{e^{-t}-e^{-t}e^{t}(1-p)} \\
&= \frac{p}{e^{-t}-1+p}.
\end{aligned}$$

■

Theorem 4.4.7 The MGF of Negative Binomial Distribution
If $X \sim$ Negative Binomial(r,p), then

$$\mathbf{M}_X(t) = \mathbf{M}_{\sum X_j}(t) = \prod_{j=1}^{r}\mathbf{M}_{X_j}(t) = \left(\frac{p}{e^{-t}-1+p}\right)^{r}$$

4.5 Gamma Distribution & Gamma Function

Definition 4.5.1 (Gamma Distribution). Let Y be a random variable that denotes the time between consecutive *Poisson* events ($X \sim$ Poission(λ)). So, $Y \sim$ Exponential(λ), and

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0.$$

Also, though of as the waiting time until the 1^{st} Poisson event. Now, let's consider Y as the waiting time to the $n-$th event, where we have Poisson event occurring at a rate of λ. In order to find the pdf of the random variable Y, we will use the cdf $\to$ pdf method. Let's consider the following distirbution.

$$\begin{aligned}
F_Y(y) &= \mathbf{P}(Y \le y) \\
&= \mathbf{P}(\text{at least } n \text{ events occur in } [0,y]) \\
&= 1-\mathbf{P}(\text{fewer than } n \text{ events occur in } [0,y]) \\
&= 1-\sum_{k=0}^{n-1} e^{-\lambda y}\frac{(\lambda y)^k}{k!}
\end{aligned}$$

Remark. *We expect λy events in $[0,y]$ since Poisson(λ) events.*

So,

$$\begin{aligned}
f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left[1 - \sum_{k=0}^{n-1} e^{-\lambda y}\frac{(\lambda y)^k}{k!}\right] \\
&= 0 - \left(\sum_{k=0}^{n-1} -\lambda e^{-\lambda y} \cdot \frac{(\lambda y)^k}{k!} + \sum_{k=1}^{n-1} e^{-\lambda y}\frac{\lambda k(\lambda y)^{k-1}}{k(k-1)!}\right) \\
&= \lambda e^{-\lambda y}\left[\sum_{k=0}^{n-1}\frac{(\lambda y)^k}{k!} - \sum_{k=1}^{n-1}\frac{(\lambda y)^{k-1}}{(k-1)!}\right] \\
&= \lambda e^{-\lambda y}\left[\sum_{k=0}^{n-1}\frac{(\lambda y)^k}{k!} - \sum_{k=0}^{n-2}\frac{(\lambda y)^k}{k!}\right] \\
&= \lambda e^{-\lambda y}\frac{(\lambda y)^{n-1}}{(n-1)!} \\
&= \frac{\lambda^n}{(n-1)!}y^{n-1}e^{-\lambda y}.
\end{aligned}$$

The pdf $f_Y(y) = \dfrac{\lambda^n}{(n-1)!}y^{n-1}e^{-\lambda y}$ is the pdf for the *Gamma Distribution* with $Y \sim \text{Gamma}(n, \lambda)$.

Remark. *Interpretation:*

$$\underbrace{\textit{Poisson}(\lambda)}_{\textit{countable evenets "per unit"}} \longrightarrow \underbrace{\textit{Exponential}(\lambda)}_{\textit{Wait time between two events (one time interval)}} \longrightarrow \underbrace{\textit{Gamma}(r, \lambda)}_{\textit{wait time for } r \textit{ events (} r \textit{ time invertval)}}$$

Theorem 4.5.2 Relationship between Exponential(λ) and Gamma(r, λ)
Suppose $Y_i \sim \text{Exponential}(\lambda)$ are independent. If

$$Y = Y_1 + Y_2 + \cdots + Y_r,$$

then $Y \sim \text{Gamma}(r, \lambda)$.

Definition 4.5.3 (Gamma Function). For $x \geq 0$, the *gamma function* of x, $\Gamma(x)$ is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1}e^t \,\mathrm{d}t.$$

Theorem 4.5.4 Properties of Gamma Function

- $$\sin(\pi x) = \frac{\pi}{\Gamma(x)\Gamma(1-x)}.$$

- Recursive Relationship:

$$\Gamma(1) = \int_0^\infty e^{-t}\,\mathrm{d}t = 1;\ \Gamma(2) = \int_0^\infty te^{-t}\,\mathrm{d}t = 1 \implies \Gamma(x+1) = x\Gamma(x) \quad \text{for } x \geq 2.$$

Theorem 4.5.5 Alternative Definition of $\Gamma(n)$ for $n \in \mathbb{N}$ and $Y \sim$ Gamma(n, λ)

$$\begin{aligned}\Gamma(n) = (n-1)\Gamma(n-1) &= (n-1)(n-2)\Gamma(n-2) \\ &= (n-1)(n-2)\cdots 2\Gamma(2) \\ &= (n-1)(n-2)\cdots 2\cdot 1\cdot 1 \\ &= (n-1)!\end{aligned}$$

So, suppose $Y \sim \text{Gamma}(n, \lambda)$, we then have

$$f_Y(y) = \frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y} = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}.$$

Theorem 4.5.6 Γ-Integral

$$\int_0^\infty t^{\alpha-1} e^{-\beta t}\,\mathrm{d}t \xrightarrow[\mathrm{d}x=\beta\,\mathrm{d}t]{x=\beta t} \beta^\alpha \underbrace{\int_0^\infty x^{\alpha-1} e^{-x}\,\mathrm{d}x}_{\Gamma(\alpha)} = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

Theorem 4.5.7 Expectation of $Y \sim$ Gamma(r, λ)
If $Y \sim \text{Gamma}(r, \lambda)$, then

$$\mathbf{E}(Y) = \frac{r}{x}.$$

Proof 1.

$$\begin{aligned}\mathbf{E}(Y) = \int_0^\infty y f_Y(y)\,\mathrm{d}y &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty y\cdot y^{r-1} e^{-\lambda y}\,\mathrm{d}y \\ &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty y^{r+1-1} e^{-\lambda y}\,\mathrm{d}y && [x = \lambda y] \\ &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty \lambda^{-(r+1)} x^{r+1-1} e^{-x}\,\mathrm{d}x \\ &= \frac{\lambda^r}{\Gamma(r)\lambda^{r+1}} \underbrace{\int_0^\infty x^{(r+1)-1} e^{-x}\,\mathrm{d}x}_{\Gamma(r+1)} \\ &= \frac{\lambda^r}{\Gamma(r)\lambda^r\cdot\lambda}\cdot\Gamma(r+1) \\ &= \frac{r\Gamma(r)}{\Gamma(r)\cdot\lambda} = \frac{r}{\lambda}.\end{aligned}$$

■

Theorem 4.5.8 Higher Moments of $Y \sim$ Gamma(r, λ)
In general,

$$\mathbf{E}(Y^n) = \int_0^\infty y^n f_Y(y)\,\mathrm{d}y = \lambda^n \frac{\Gamma(r+n)}{\Gamma(r)}.$$

Definition 4.5.9 (χ^2, Chi-Squared Distribution). Given $Z \sim N(0,1)$, then $Z^2 \sim \chi^2$. Let $Y = Z^n$, then

$$F_Y(y) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-z^2/2} \, \mathrm{d}z$$

and thus

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{1}{\sqrt{2\pi}} \underbrace{y^{-1/2} e^{-y/2}}_{\text{basis of Gamma pdf}} .$$

So, in fact, $\chi^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$.

Remark. *We have*

$$\frac{\lambda^r}{\Gamma(r)} = \frac{1}{\sqrt{2\pi}} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\sqrt{\pi}} \implies \Gamma(r) = \sqrt{\pi}.$$