1

General Probability Theory

Probability theory is a branch of mathematics that deals with mathematical models of trials whose outcomes depend on chance. Within the context of mathematical finance, we will review some basic concepts of probability theory that are needed to begin solving stochastic calculus problems. The topics covered in this chapter are by no means exhaustive but are sufficient to be utilised in the following chapters and in later volumes. However, in order to fully grasp the concepts, an undergraduate level of mathematics and probability theory is generally required from the reader (see Appendices A and B for a quick review of some basic mathematics and probability theory). In addition, the reader is also advised to refer to the notation section (pages 369–372) on set theory, mathematical and probability symbols used in this book.

1.1 INTRODUCTION

We consider an *experiment* or a *trial* whose result (*outcome*) is not predictable with certainty. The set of all possible outcomes of an experiment is called the *sample space* and we denote it by Ω. Any subset A of the sample space is known as an *event*, where an event is a set consisting of possible outcomes of the experiment.

The collection of events can be defined as a subcollection $\mathscr{F}$ of the set of all subsets of Ω and we define any collection $\mathscr{F}$ of subsets of Ω as a *field* if it satisfies the following.

Definition 1.1 *The sample space Ω is the set of all possible outcomes of an experiment or random trial. A field is a collection (or family) $\mathscr{F}$ of subsets of Ω with the following conditions:*

(a) $\emptyset \in \mathscr{F}$ where $\emptyset$ is the empty set;
(b) if $A \in \mathscr{F}$ then $A^c \in \mathscr{F}$ where A^c is the complement of A in Ω;
(c) if $A_1, A_2, \ldots, A_n \in \mathscr{F}$, $n \geq 2$ then $\bigcup_{i=1}^{n} A_i \in \mathscr{F}$ – that is to say, $\mathscr{F}$ is closed under finite unions.

It should be noted in the definition of a field that $\mathscr{F}$ is closed under finite unions (as well as under finite intersections). As for the case of a collection of events closed under countable unions (as well as under countable intersections), any collection of subsets of Ω with such properties is called a σ-algebra.

Definition 1.2 *If Ω is a given sample space, then a σ-algebra (or σ-field) $\mathscr{F}$ on Ω is a family (or collection) $\mathscr{F}$ of subsets of Ω with the following properties:*

(a) $\emptyset \in \mathscr{F}$;
(b) if $A \in \mathscr{F}$ then $A^c \in \mathscr{F}$ where A^c is the complement of A in Ω;
(c) if $A_1, A_2, \ldots \in \mathscr{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ – that is to say, $\mathscr{F}$ is closed under countable unions.

We next outline an approach to probability which is a branch of *measure theory*. The reason for taking a measure-theoretic path is that it leads to a unified treatment of both discrete and continuous random variables, as well as a general definition of *conditional expectation*.

Definition 1.3 *The pair* $(\Omega, \mathscr{F})$ *is called a measurable space. A probability measure* $\mathbb{P}$ *on a measurable space* $(\Omega, \mathscr{F})$ *is a function* $\mathbb{P} : \mathscr{F} \mapsto [0, 1]$ *such that:*

(a) $\mathbb{P}(\emptyset) = 0$*;*
(b) $\mathbb{P}(\Omega) = 1$*;*
(c) if $A_1, A_2, \ldots \in \mathscr{F}$ *and* $\{A_i\}_{i=1}^{\infty}$ *is disjoint such that* $A_i \cap A_j = \emptyset$*,* $i \neq j$ *then* $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

The triple $(\Omega, \mathscr{F}, \mathbb{P})$ *is called a probability space. It is called a complete probability space if* $\mathscr{F}$ *also contains subsets* B *of* Ω *with* $\mathbb{P}$*-outer measure zero, that is* $\mathbb{P}^*(B) = \inf\{\mathbb{P}(A) : A \in \mathscr{F}, B \subset A\} = 0$.

By treating σ-algebras as a record of information, we have the following definition of a *filtration*.

Definition 1.4 *Let* Ω *be a non-empty sample space and let* T *be a fixed positive number, and assume for each* $t \in [0, T]$ *there is a* σ*-algebra* $\mathscr{F}_t$*. In addition, we assume that if* $s \leq t$*, then every set in* $\mathscr{F}_s$ *is also in* $\mathscr{F}_t$*. We call the collection of* σ*-algebras* $\mathscr{F}_t$*,* $0 \leq t \leq T$*, a filtration.*

Below we look into the definition of a real-valued random variable, which is a function that maps a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to a measurable space $\mathbb{R}$.

Definition 1.5 *Let* Ω *be a non-empty sample space and let* $\mathscr{F}$ *be a* σ*-algebra of subsets of* Ω*. A real-valued random variable* X *is a function* $X : \Omega \mapsto \mathbb{R}$ *such that* $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{F}$ *for each* $x \in \mathbb{R}$ *and we say* X *is* $\mathscr{F}$ *measurable.*

In the study of stochastic processes, an *adapted stochastic process* is one that cannot "see into the future" and in mathematical finance we assume that asset prices and portfolio positions taken at time t are all adapted to a filtration $\mathscr{F}_t$, which we regard as the flow of information up to time t. Therefore, these values must be $\mathscr{F}_t$ measurable (i.e., depend only on information available to investors at time t). The following is the precise definition of an adapted stochastic process.

Definition 1.6 *Let* Ω *be a non-empty sample space with a filtration* $\mathscr{F}_t$*,* $t \in [0, T]$ *and let* X_t *be a collection of random variables indexed by* $t \in [0, T]$*. We therefore say that this collection of random variables is an adapted stochastic process if, for each* t*, the random variable* X_t *is* $\mathscr{F}_t$ *measurable.*

Finally, we consider the concept of conditional expectation, which is extremely important in probability theory and also for its wide application in mathematical finance such as pricing options and other derivative products. Conceptually, we consider a random variable X defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). Here X can represent a quantity we want to estimate, say the price of a stock in the future, while

$\mathscr{G}$ contains limited information about X such as the stock price up to and including the current time. Thus, $\mathbb{E}(X|\mathscr{G})$ constitutes the best estimation we can make about X given the limited knowledge $\mathscr{G}$. The following is a formal definition of a conditional expectation.

Definition 1.7 (Conditional Expectation) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). Let X be an integrable (i.e., $\mathbb{E}(|X|) < \infty$) and non-negative random variable. Then the conditional expectation of X given $\mathscr{G}$, denoted $\mathbb{E}(X|\mathscr{G})$, is any random variable that satisfies:*

(a) $\mathbb{E}(X|\mathscr{G})$ is $\mathscr{G}$ measurable;
(b) for every set $A \in \mathscr{G}$, we have the partial averaging property

$$\int_A \mathbb{E}(X|\mathscr{G})\, d\mathbb{P} = \int_A X \;\, d\mathbb{P}.$$

From the above definition, we can list the following properties of conditional expectation. Here $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, $\mathscr{G}$ is a sub-σ-algebra of $\mathscr{F}$ and X is an integrable random variable.

- *Conditional probability*. If $\mathbb{I}_A$ is an indicator random variable for an event A then

$$\mathbb{E}(\mathbb{I}_A|\mathscr{G}) = \mathbb{P}(A|\mathscr{G}).$$

- *Linearity*. If $X_1, X_2, \ldots, X_n$ are integrable random variables and $c_1, c_2, \ldots, c_n$ are constants then

$$\mathbb{E}(c_1X_1 + c_2X_2 + \ldots + c_nX_n|\mathscr{G}) = c_1\mathbb{E}(X_1|\mathscr{G}) + c_2\mathbb{E}(X_2|\mathscr{G}) + \ldots + c_n\mathbb{E}(X_n|\mathscr{G}).$$

- *Positivity*. If $X \geq 0$ almost surely then $\mathbb{E}(X|\mathscr{G}) \geq 0$ almost surely.
- *Monotonicity*. If X and Y are integrable random variables and $X \leq Y$ almost surely then

$$\mathbb{E}(X|\mathscr{G}) \leq \mathbb{E}(Y|\mathscr{G}).$$

- *Computing expectations by conditioning*. $\mathbb{E}[\mathbb{E}(X|\mathscr{G})] = \mathbb{E}(X)$.
- *Taking out what is known*. If X and Y are integrable random variables and X is $\mathscr{G}$ measurable then

$$\mathbb{E}(XY|\mathscr{G}) = X \cdot \mathbb{E}(Y|\mathscr{G}).$$

- *Tower property*. If $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$ then

$$\mathbb{E}[\mathbb{E}(X|\mathscr{G})|\mathscr{H}] = \mathbb{E}(X|\mathscr{H}).$$

- *Measurability*. If X is $\mathscr{G}$ measurable then $\mathbb{E}(X|\mathscr{G}) = X$.
- *Independence*. If X is independent of $\mathscr{G}$ then $\mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X)$.
- *Conditional Jensen's inequality*. If $\varphi : \mathbb{R} \mapsto \mathbb{R}$ is a convex function then

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq \varphi[\mathbb{E}(X|\mathscr{G})].$$

1.2 PROBLEMS AND SOLUTIONS

1.2.1 Probability Spaces

1. *De Morgan's Law*. Let A_i, $i \in I$ where I is some, possibly uncountable, indexing set. Show that

   (a) $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c$.
   (b) $\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$.

   Solution:

   (a) Let $a \in \left(\bigcup_{i\in I} A_i\right)^c$ which implies $a \notin \bigcup_{i\in I} A_i$, so that $a \in A_i^c$ for all $i \in I$. Therefore,

   $$\left(\bigcup_{i\in I} A_i\right)^c \subseteq \bigcap_{i\in I} A_i^c.$$

   On the contrary, if we let $a \in \bigcap_{i\in I} A_i^c$ then $a \notin A_i$ for all $i \in I$ or $a \in \left(\bigcup_{i\in I} A_i\right)^c$ and hence

   $$\bigcap_{i\in I} A_i^c \subseteq \left(\bigcup_{i\in I} A_i\right)^c.$$

   Therefore, $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c$.

   (b) From (a), we can write

   $$\left(\bigcup_{i\in I} A_i^c\right)^c = \bigcap_{i\in I} \left(A_i^c\right)^c = \bigcap_{i\in I} A_i.$$

   Taking complements on both sides gives

   $$\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

   □

2. Let $\mathscr{F}$ be a σ-algebra of subsets of the sample space Ω. Show that if $A_1, A_2, \ldots \in \mathscr{F}$ then $\bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$.

   Solution: Given that $\mathscr{F}$ is a σ-algebra then $A_1^c, A_2^c, \ldots \in \mathscr{F}$ and $\bigcup_{i=1}^{\infty} A_i^c \in \mathscr{F}$. Furthermore, the complement of $\bigcup_{i=1}^{\infty} A_i^c$ is $\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathscr{F}$.
   Thus, from De Morgan's law (see Problem 1.2.1.1, page 4) we have $\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} \left(A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$.

   □

3. Show that if $\mathscr{F}$ is a σ-algebra of subsets of Ω then $\{\emptyset, \Omega\} \in \mathscr{F}$.

   Solution: $\mathscr{F}$ is a σ-algebra of subsets of Ω, hence if $A \in \mathscr{F}$ then $A^c \in \mathscr{F}$.
   Since $\emptyset \in \mathscr{F}$ then $\emptyset^c = \Omega \in \mathscr{F}$. Thus, $\{\emptyset, \Omega\} \in \mathscr{F}$.

   □

4. Show that if $A \subseteq \Omega$ then $\mathscr{F} = \{\emptyset, \Omega, A, A^c\}$ is a σ-algebra of subsets of Ω.

   Solution: $\mathscr{F} = \{\emptyset, \Omega, A, A^c\}$ is a σ-algebra of subsets of Ω since

   (i) $\emptyset \in \mathscr{F}$.
   (ii) For $\emptyset \in \mathscr{F}$ then $\emptyset^c = \Omega \in \mathscr{F}$. For $\Omega \in \mathscr{F}$ then $\Omega^c = \emptyset \in \mathscr{F}$. In addition, for $A \in \mathscr{F}$ then $A^c \in \mathscr{F}$. Finally, for $A^c \in \mathscr{F}$ then $(A^c)^c = A \in \mathscr{F}$.
   (iii) $\emptyset \cup \Omega = \Omega \in \mathscr{F}$, $\emptyset \cup A = A \in \mathscr{F}$, $\emptyset \cup A^c = A^c \in \mathscr{F}$, $\Omega \cup A = \Omega \in \mathscr{F}$, $\Omega \cup A^c = \Omega \in \mathscr{F}$, $\emptyset \cup \Omega \cup A = \Omega \in \mathscr{F}$, $\emptyset \cup \Omega \cup A^c = \Omega \in \mathscr{F}$ and $\Omega \cup A \cup A^c = \Omega \in \mathscr{F}$.

   □

5. Let $\{\mathscr{F}_i\}_{i \in I}$, $I \neq \emptyset$ be a family of σ-algebras of subsets of the sample space Ω. Show that $\mathscr{F} = \bigcap_{i \in I} \mathscr{F}_i$ is also a σ-algebra of subsets of Ω.

   Solution: $\mathscr{F} = \bigcap_{i \in I} \mathscr{F}_i$ is a σ-algebra by taking note that

   (a) Since $\emptyset \in \mathscr{F}_i$, $i \in I$ therefore $\emptyset \in \mathscr{F}$ as well.
   (b) If $A \in \mathscr{F}_i$ for all $i \in I$ then $A^c \in \mathscr{F}_i$, $i \in I$. Therefore, $A \in \mathscr{F}$ and hence $A^c \in \mathscr{F}$.
   (c) If $A_1, A_2, \ldots \in \mathscr{F}_i$ for all $i \in I$ then $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}_i$, $i \in I$ and hence $A_1, A_2, \ldots \in \mathscr{F}$ and $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}$.

   From the results of (a)–(c) we have shown $\mathscr{F} = \bigcap_{i \in I} \mathscr{F}_i$ is a σ-algebra of Ω.

   □

6. Let $\Omega = \{\alpha, \beta, \gamma\}$ and let

   $$\mathscr{F}_1 = \{\emptyset, \Omega, \{\alpha\}, \{\beta, \gamma\}\} \quad \text{and} \quad \mathscr{F}_2 = \{\emptyset, \Omega, \{\alpha, \beta\}, \{\gamma\}\}.$$

   Show that $\mathscr{F}_1$ and $\mathscr{F}_2$ are σ-algebras of subsets of Ω.
   Is $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2$ also a σ-algebra of subsets of Ω?

   Solution: Following the steps given in Problem 1.2.1.4 (page 5) we can easily show $\mathscr{F}_1$ and $\mathscr{F}_2$ are σ-algebras of subsets of Ω.
   By setting $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2 = \{\emptyset, \Omega, \{\alpha\}, \{\gamma\}, \{\alpha, \beta\}, \{\beta, \gamma\}\}$, and since $\{\alpha\} \in \mathscr{F}$ and $\{\gamma\} \in \mathscr{F}$, but $\{\alpha\} \cup \{\gamma\} = \{\alpha, \gamma\} \notin \mathscr{F}$, then $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2$ is not a σ-algebra of subsets of Ω.

   □

7. Let $\mathscr{F}$ be a σ-algebra of subsets of Ω and suppose $\mathbb{P} : \mathscr{F} \mapsto [0, 1]$ so that $\mathbb{P}(\Omega) = 1$. Show that $\mathbb{P}(\emptyset) = 0$.

   Solution: Given that $\emptyset$ and Ω are mutually exclusive we therefore have

   $$\emptyset \cap \Omega = \emptyset \text{ and } \emptyset \cup \Omega = \Omega.$$

   Thus, we can express

   $$\mathbb{P}(\emptyset \cup \Omega) = \mathbb{P}(\emptyset) + \mathbb{P}(\Omega) - \mathbb{P}(\emptyset \cap \Omega) = 1.$$

   Since $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset \cap \Omega) = 0$ therefore $\mathbb{P}(\emptyset) = 0$.

   □

8. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathbb{Q} : \mathscr{F} \mapsto [0,1]$ be defined by $\mathbb{Q}(A) = \mathbb{P}(A|B)$ where $B \in \mathscr{F}$ such that $\mathbb{P}(B) > 0$. Show that $(\Omega, \mathscr{F}, \mathbb{Q})$ is also a probability space.

Solution: To show that $(\Omega, \mathscr{F}, \mathbb{Q})$ is a probability space we note that

(a) $\mathbb{Q}(\emptyset) = \mathbb{P}(\emptyset|B) = \dfrac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0.$

(b) $\mathbb{Q}(\Omega) = \mathbb{P}(\Omega|B) = \dfrac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$

(c) Let $A_1, A_2, \ldots$ be disjoint members of $\mathscr{F}$ and hence we can imply $A_1 \cap B, A_2 \cap B, \ldots$ are also disjoint members of $\mathscr{F}$. Therefore,

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

Based on the results of (a)–(c), we have shown that $(\Omega, \mathscr{F}, \mathbb{Q})$ is also a probability space.

$\square$

9. *Boole's Inequality*. Suppose $\{A_i\}_{i \in I}$ is a countable collection of events. Show that

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbb{P}\left(A_i\right).$$

Solution: Without loss of generality we assume that $I = \{1, 2, \ldots\}$ and define $B_1 = A_1$, $B_i = A_i \backslash (A_1 \cup A_2 \cup \ldots \cup A_{i-1})$, $i \in \{2, 3, \ldots\}$ such that $\{B_1, B_2, \ldots\}$ are pairwise disjoint and

$$\bigcup_{i \in I} A_i = \bigcup_{i \in I} B_i.$$

Because $B_i \cap B_j = \emptyset$, $i \neq j$, $i, j \in I$ we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i \in I} A_i\right) &= \mathbb{P}\left(\bigcup_{i \in I} B_i\right) \\
&= \sum_{i \in I} \mathbb{P}(B_i) \\
&= \sum_{i \in I} \mathbb{P}(A_i \backslash (A_1 \cup A_2 \cup \ldots \cup A_{i-1})) \\
&= \sum_{i \in I} \left\{\mathbb{P}(A_i) - \mathbb{P}\left(A_i \cap (A_1 \cup A_2 \cup \ldots \cup A_{i-1})\right)\right\} \\
&\leq \sum_{i \in I} \mathbb{P}(A_i).
\end{aligned}$$

$\square$

10. *Bonferroni's Inequality*. Suppose $\{A_i\}_{i\in I}$ is a countable collection of events. Show that

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) \geq 1 - \sum_{i\in I} \mathbb{P}\left(A_i^c\right).$$

Solution: From De Morgan's law (see Problem 1.2.1.1, page 4) we can write

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i\in I} A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i\in I} A_i^c\right).$$

By applying Boole's inequality (see Problem 1.2.1.9, page 6) we will have

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) \geq 1 - \sum_{i\in I} \mathbb{P}\left(A_i^c\right)$$

since $\mathbb{P}\left(\bigcup_{i\in I} A_i^c\right) \leq \sum_{i\in I} \mathbb{P}\left(A_i^c\right)$.

□

11. *Bayes' Formula*. Let $A_1, A_2, \ldots, A_n$ be a partition of Ω, where $\bigcup_{i=1}^{n} A_i = \Omega$, $A_i \cap A_j = \emptyset$, $i \neq j$ and each A_i, $i, j = 1, 2, \ldots, n$ has positive probability. Show that

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

Solution: From the definition of conditional probability, for $i = 1, 2, \ldots, n$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}\left(\bigcup_{j=1}^{n}(B \cap A_j)\right)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \cap A_j)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

□

12. *Principle of Inclusion and Exclusion for Probability*. Let $A_1, A_2, \ldots, A_n$, $n \geq 2$ be a collection of events. Show that

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

From the above result show that

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3)\\ - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Hence, using mathematical induction show that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{P}(A_i \cap A_j) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} \mathbb{P}(A_i \cap A_j \cap A_k)$$

$$-\ \ldots\ +(-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \ \ldots\ \cap A_n).$$

Finally, deduce that

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \sum_{i=1}^{n}\mathbb{P}(A_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{P}(A_i \cup A_j) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}\mathbb{P}(A_i \cup A_j \cup A_k)$$
$$-\ \ldots\ +(-1)^{n+1}\mathbb{P}(A_1 \cup A_2 \cup \ \ldots\ \cup A_n).$$

Solution: For $n = 2$, $A_1 \cup A_2$ can be written as a union of two disjoint sets

$$A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1) = A_1 \cup (A_2 \cap A_1^c).$$

Therefore,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \cap A_1^c)$$

and since $\mathbb{P}(A_2) = \mathbb{P}(A_2 \cap A_1) + \mathbb{P}(A_2 \cap A_1^c)$ we will have

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

For $n = 3$, and using the above results, we can write

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) &= \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}[(A_1 \cup A_2) \cap A_3]\\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}[(A_1 \cup A_2) \cap A_3].\end{aligned}$$

Since $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$ therefore

$$\begin{aligned}\mathbb{P}[(A_1 \cup A_2) \cap A_3] &= \mathbb{P}[(A_1 \cap A_3) \cup (A_2 \cap A_3)]\\ &= \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3) - \mathbb{P}[(A_1 \cap A_3) \cap (A_2 \cap A_3)]\\ &= \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Thus,

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3)$$
$$-\mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3).$$

Suppose the result is true for $n = m$, where $m \geq 2$. For $n = m + 1$, we have

$$\begin{aligned}\mathbb{P}\left(\left(\bigcup_{i=1}^{m} A_i\right) \cup A_{m+1}\right) &= \mathbb{P}\left(\bigcup_{i=1}^{m} A_i\right) + \mathbb{P}\left(A_{m+1}\right) - \mathbb{P}\left(\left(\bigcup_{i=1}^{m} A_i\right) \cap A_{m+1}\right)\\ &= \mathbb{P}\left(\bigcup_{i=1}^{m} A_i\right) + \mathbb{P}\left(A_{m+1}\right) - \mathbb{P}\left(\bigcup_{i=1}^{m} \left(A_i \cap A_{m+1}\right)\right).\end{aligned}$$

By expanding the terms we have

$$
\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{m+1} A_i\right) &= \sum_{i=1}^{m} \mathbb{P}(A_i) - \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \mathbb{P}(A_i \cap A_j) + \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{k=j+1}^{m} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad - \ldots + (-1)^{m+1}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_m) \\
&\quad + \mathbb{P}\left(A_{m+1}\right) - \mathbb{P}\left((A_1 \cap A_{m+1}) \cup (A_2 \cap A_{m+1}) \ldots \cup (A_m \cap A_{m+1})\right) \\
&= \sum_{i=1}^{m+1} \mathbb{P}(A_i) - \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \mathbb{P}(A_i \cap A_j) + \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{k=j+1}^{m} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad - \ldots + (-1)^{m+1}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_m) \\
&\quad - \sum_{i=1}^{m} \mathbb{P}(A_i \cap A_{m+1}) + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \mathbb{P}(A_i \cap A_j \cap A_{m+1}) \\
&\quad + \ldots - (-1)^{m+1}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{m+1}) \\
&= \sum_{i=1}^{m+1} \mathbb{P}(A_i) - \sum_{i=1}^{m}\sum_{j=i+1}^{m+1} \mathbb{P}(A_i \cap A_j) + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\sum_{k=j+1}^{m+1} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad - \ldots + (-1)^{m+2}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_{m+1}).
\end{aligned}
$$

Therefore, the result is also true for $n = m + 1$. Thus, from mathematical induction we have shown for $n \geq 2$,

$$
\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{P}(A_i \cap A_j) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad - \ldots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n).
\end{aligned}
$$

From Problem 1.2.1.1 (page 4) we can write

$$
\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i^c\right).
$$

Thus, we can expand

$$
\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) &= 1 - \sum_{i=1}^{n} \mathbb{P}(A_i^c) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{P}(A_i^c \cap A_j^c) - \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} \mathbb{P}(A_i^c \cap A_j^c \cap A_k^c) \\
&\quad + \ldots - (-1)^{n+1}\mathbb{P}(A_1^c \cap A_2^c \cap \ldots \cap A_n^c) \\
&= 1 - \sum_{i=1}^{n}(1 - \mathbb{P}(A_i)) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(1 - \mathbb{P}(A_i \cup A_j)) \\
&\quad - \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}(1 - \mathbb{P}(A_i \cup A_j \cup A_k))
\end{aligned}
$$

$$+ \ldots - (-1)^{n+1}(1 - \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n))$$
$$= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{P}(A_i \cup A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \mathbb{P}(A_i \cup A_j \cup A_k)$$
$$- \ldots + (-1)^{n+1}\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n).$$

□

13. *Borel–Cantelli Lemma*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $A_1, A_2, \ldots$ be sets in $\mathscr{F}$. Show that

$$\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k \subseteq \bigcup_{k=m}^{\infty} A_k$$

and hence prove that

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) = \begin{cases} 1 & \text{if } A_i \cap A_j = \emptyset, i \neq j \text{ and } \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k) = \infty \\ \\ 0 & \text{if } \sum\limits_{k=1}^{\infty} \mathbb{P}(A_k) < \infty. \end{cases}$$

Solution: Let $B_m = \bigcup\limits_{k=m}^{\infty} A_k$ and since $\bigcap\limits_{m=1}^{\infty} B_m \subseteq B_m$ therefore we have

$$\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k \subseteq \bigcup_{k=m}^{\infty} A_k.$$

From Problem 1.2.1.9 (page 6) we can deduce that

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) \leq \mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \sum_{k=m}^{\infty} \mathbb{P}(A_k).$$

For the case $\sum\limits_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$ and given it is a convergent series, then $\lim\limits_{m\to\infty} \sum\limits_{k=m}^{\infty} \mathbb{P}(A_k) = 0$ and hence

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) = 0$$

if $\sum\limits_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$.

For the case $A_i \cap A_j = \emptyset$, $i \neq j$ and $\sum\limits_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$, since $\mathbb{P}\left(\bigcup\limits_{k=m}^{\infty} A_k\right) + \mathbb{P}\left(\left(\bigcup\limits_{k=m}^{\infty} A_k\right)^c\right) = 1$ therefore from Problem 1.2.1.1 (page 4)

$$\mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) = 1 - \mathbb{P}\left(\left(\bigcup_{k=m}^{\infty} A_k\right)^c\right) = 1 - \mathbb{P}\left(\bigcap_{k=m}^{\infty} A_k^c\right).$$

From independence and because $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$ we can express

$$\mathbb{P}\left(\bigcap_{k=m}^{\infty} A_k^c\right) = \prod_{k=m}^{\infty} \mathbb{P}\left(A_k^c\right) = \prod_{k=m}^{\infty} \left(1 - \mathbb{P}\left(A_k\right)\right) \leq \prod_{k=m}^{\infty} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=m}^{\infty} \mathbb{P}(A_k)} = 0$$

for all m and hence

$$\mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) = 1 - \mathbb{P}\left(\bigcap_{k=m}^{\infty} A_k^c\right) = 1$$

for all m. Taking the limit $m \to \infty$,

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) = 1$$

for the case $A_i \cap A_j = \emptyset$, $i \neq j$ and $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$.

□

1.2.2 Discrete and Continuous Random Variables

1. *Bernoulli Distribution*. Let X be a Bernoulli random variable, $X \sim \text{Bernoulli}(p)$, $p \in [0, 1]$ with probability mass function

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

Show that $\mathbb{E}(X) = p$ and $\text{Var}(X) = p(1 - p)$.

Solution: If $X \sim \text{Bernoulli}(p)$ then we can write

$$\mathbb{P}(X = x) = p^x(1 - p)^{1-x}, \quad x \in \{0, 1\}$$

and by definition

$$\mathbb{E}(X) = \sum_{x=0}^{1} x\mathbb{P}(X = x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$\mathbb{E}(X^2) = \sum_{x=0}^{1} x^2\mathbb{P}(X = x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

and hence

$$\mathbb{E}(X) = p, \quad \text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p(1 - p).$$

□

2. *Binomial Distribution*. Let $\{X_i\}_{i=1}^n$ be a sequence of independent Bernoulli random variables each with probability mass function

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p, \quad p \in [0, 1]$$

and let

$$X = \sum_{i=1}^{n} X_i.$$

Show that X follows a binomial distribution, $X \sim \text{Binomial}(n, p)$ with probability mass function

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

such that $\mathbb{E}(X) = np$ and $\text{Var}(X) = np(1-p)$.
Using the central limit theorem show that X is approximately normally distributed, $X \dot{\sim} \mathcal{N}(np, np(1-p))$ as $n \to \infty$.

Solution: The random variable X counts the number of Bernoulli variables $X_1, \ldots, X_n$ that are equal to 1, i.e., the number of successes in the n independent trials. Clearly X takes values in the set $N = \{0, 1, 2, \ldots, n\}$. To calculate the probability that $X = k$, where $k \in N$ is the number of successes we let E be the event such that $X_{i_1} = X_{i_2} = \ldots = X_{i_k} = 1$ and $X_j = 0$ for all $j \in N \backslash S$ where $S = \{i_1, i_2, \ldots, i_k\}$. Then, because the Bernoulli variables are independent and identically distributed,

$$\mathbb{P}(E) = \prod_{j \in S} \mathbb{P}(X_j = 1) \prod_{j \in N \backslash S} \mathbb{P}(X_j = 0) = p^k (1-p)^{n-k}.$$

However, as there are $\binom{n}{k}$ combinations to select sets of indices $i_1, \ldots, i_k$ from N, which are mutually exclusive events, so

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

From the definition of the moment generating function of discrete random variables (see Appendix B),

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \sum_x e^{tx} \mathbb{P}(X = x)$$

where $t \in \mathbb{R}$ and substituting $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ we have

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (1 - p + pe^t)^n.$$

By differentiating $M_X(t)$ with respect to t twice we have

$$M'_X(t) = npe^t(1 - p + pe^t)^{n-1}$$
$$M''_X(t) = npe^t(1 - p + pe^t)^{n-1} + n(n-1)p^2e^{2t}(1 - p + pe^t)^{n-2}$$

and hence

$$\mathbb{E}(X) = M'_X(0) = np$$
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = M''_X(0) - M'_X(0)^2 = np(1-p).$$

Given the sequence $X_i \sim \text{Bernoulli}(p)$, $i = 1, 2, \ldots, n$ are independent and identically distributed, each having expectation $\mu = p$ and variance $\sigma^2 = p(1-p)$, then as $n \to \infty$, from the central limit theorem

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, 1)$$

or

$$\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{D} \mathcal{N}(0, 1).$$

Thus, as $n \to \infty$, $X \dot{\sim} \mathcal{N}(np, np(1-p))$.

$\square$

3. *Poisson Distribution*. A discrete Poisson distribution, Poisson(λ) with parameter $\lambda > 0$ has the following probability mass function:

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

Show that $\mathbb{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.
For a random variable following a binomial distribution, Binomial(n, p), $0 \le p \le 1$ show that as $n \to \infty$ and with $p = \lambda/n$, the binomial distribution tends to the Poisson distribution with parameter λ.

Solution: From the definition of the moment generating function

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \sum_x e^{tx}\mathbb{P}(X = x)$$

where $t \in \mathbb{R}$ and substituting $\mathbb{P}(X = x) = \dfrac{\lambda^x}{x!}\,e^{-\lambda}$ we have

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x}{x!}\,e^{-\lambda} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{\lambda(e^t-1)}.$$

By differentiating $M_X(t)$ with respect to t twice

$$M'_X(t) = \lambda e^t e^{\lambda(e^t-1)}, \quad M''_X(t) = (\lambda e^t + 1)\lambda e^t e^{\lambda(e^t-1)}$$

we have

$$\mathbb{E}(X) = M'_X(0) = \lambda$$
$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = M''_X(0) - M'_X(0)^2 = \lambda.$$

If $X \sim \text{Binomial}(n, p)$ then we can write

$$\begin{aligned}
\mathbb{P}(X = k) &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k+1)(n-k)!}{k!(n-k)!} \\
&\quad \times p^k \left(1 - (n-k)p + \frac{(n-k)(n-k-1)}{2!} p^2 + \ldots\right) \\
&= \frac{n(n-1)\ldots(n-k+1)}{k!} p^k \left(1 - (n-k)p + \frac{(n-k)(n-k-1)}{2!} p^2 + \ldots\right).
\end{aligned}$$

For the case when $n \to \infty$ so that $n \gg k$ we have

$$\begin{aligned}
\mathbb{P}(X = k) &\approx \frac{n^k}{k!} p^k \left(1 - np + \frac{(np)^2}{2!} + \ldots\right) \\
&= \frac{n^k}{k!} p^k e^{-np}.
\end{aligned}$$

By setting $p = \lambda/n$, we have $\mathbb{P}(X = k) \approx \dfrac{\lambda^k}{k!} e^{-\lambda}$.

□

4. *Exponential Distribution*. Consider a continuous random variable X following an exponential distribution, $X \sim \text{Exp}(\lambda)$ with probability density function

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

where the parameter $\lambda > 0$. Show that $\mathbb{E}(X) = \dfrac{1}{\lambda}$ and $\text{Var}(X) = \dfrac{1}{\lambda^2}$.
Prove that $X \sim \text{Exp}(\lambda)$ has a memory less property given as

$$\mathbb{P}(X > s + x | X > s) = \mathbb{P}(X > x) = e^{-\lambda x}, \quad x, s \geq 0.$$

For a sequence of Bernoulli trials drawn from a Bernoulli distribution, Bernoulli(p), $0 \leq p \leq 1$ performed at time Δt, $2\Delta t$, $\ldots$ where $\Delta t > 0$ and if Y is the waiting time for the first success, show that as $\Delta t \to 0$ and $p \to 0$ such that $p/\Delta t$ approaches a constant $\lambda > 0$, then $Y \sim \text{Exp}(\lambda)$.

Solution: For $t < \lambda$, the moment generating function for a random variable $X \sim \text{Exp}(\lambda)$ is

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \int_0^\infty e^{tu} \lambda e^{-\lambda u} du = \lambda \int_0^\infty e^{-(\lambda - t)u} du = \frac{\lambda}{\lambda - t}.$$

Differentiation of $M_X(t)$ with respect to t twice yields

$$M_X'(t) = \frac{\lambda}{(\lambda - t)^2}, \quad M_X''(t) = \frac{2\lambda}{(\lambda - t)^3}.$$

Therefore,

$$\mathbb{E}(X) = M_X'(0) = \frac{1}{\lambda}, \quad \mathbb{E}(X^2) = M_X''(0) = \frac{2}{\lambda^2}$$

and the variance of X is

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

By definition

$$\mathbb{P}(X > x) = \int_x^\infty \lambda e^{-\lambda u} du = e^{-\lambda x}$$

and

$$\mathbb{P}(X > s + x | X > s) = \frac{\mathbb{P}(X > s + x, X > s)}{\mathbb{P}(X > s)} = \frac{\mathbb{P}(X > s + x)}{\mathbb{P}(X > s)} = \frac{\int_{s+x}^\infty \lambda e^{-\lambda u} du}{\int_s^\infty \lambda e^{-\lambda v} dv} = e^{-\lambda x}.$$

Thus,

$$\mathbb{P}(X > s + x | X > s) = \mathbb{P}(X > x).$$

If Y is the waiting time for the first success then for $k = 1, 2, \ldots$

$$\mathbb{P}(Y > k\Delta t) = (1 - p)^k.$$

By setting $y = k\Delta t$, and in the limit $\Delta t \to 0$ and assuming that $p \to 0$ so that $p/\Delta t \to \lambda$, for some positive constant λ,

$$\begin{aligned}
\mathbb{P}(Y > y) &= \mathbb{P}\left(Y > \left(\frac{x}{\Delta t}\right)\Delta t\right) \\
&\approx (1 - \lambda\Delta t)^{y/\Delta t} \\
&= 1 - \lambda y + \frac{\left(\frac{y}{\Delta t}\right)\left(\frac{y}{\Delta t} - 1\right)}{2!}(\lambda\Delta t)^2 + \ldots \\
&\approx 1 - \lambda y + \frac{(\lambda y)^2}{2!} + \ldots \\
&= e^{-\lambda x}.
\end{aligned}$$

In the limit $\Delta t \to 0$ and $p \to 0$,

$$\mathbb{P}(Y \le y) = 1 - \mathbb{P}(Y > y) \approx 1 - e^{-\lambda y}$$

and the probability density function is therefore

$$f_Y(y) = \frac{d}{dy}\mathbb{P}(Y \le y) \approx \lambda e^{-\lambda y}, y \ge 0.$$

$\square$

5. *Gamma Distribution*. Let U and V be continuous independent random variables and let $W = U + V$. Show that the probability density function of W can be written as

$$f_W(w) = \int_{-\infty}^{\infty} f_V(w-u)f_U(u)\,du = \int_{-\infty}^{\infty} f_U(w-v)f_V(v)\,dv$$

where $f_U(u)$ and $f_V(v)$ are the density functions of U and V, respectively.
Let $X_1, X_2, \ldots, X_n \sim \text{Exp}(\lambda)$ be a sequence of independent and identically distributed random variables, each following an exponential distribution with common parameter $\lambda > 0$. Prove that if

$$Y = \sum_{i=1}^{n} X_i$$

then Y follows a gamma distribution, $Y \sim \text{Gamma}\,(n, \lambda)$ with the following probability density function:

$$f_Y(y) = \frac{(\lambda y)^{n-1}}{(n-1)!}\lambda e^{-\lambda y}, \quad y \geq 0.$$

Show also that $\mathbb{E}(Y) = \dfrac{n}{\lambda}$ and $\text{Var}(Y) = \dfrac{n}{\lambda^2}$.

Solution: From the definition of the cumulative distribution function of $W = U + V$ we obtain

$$F_W(w) = \mathbb{P}(W \leq w) = \mathbb{P}(U + V \leq w) = \iint\limits_{u+v\leq w} f_{UV}(u, v)\,dudv$$

where $f_{UV}(u, v)$ is the joint probability density function of (U, V). Since $U \perp\!\!\!\perp V$ therefore $f_{UV}(u, v) = f_U(u)f_V(v)$ and hence

$$\begin{aligned}
F_W(w) &= \iint\limits_{u+v\leq w} f_{UV}(u, v)\,dudv \\
&= \iint\limits_{u+v\leq w} f_U(u)f_V(v)\,dudv \\
&= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{w-u} f_V(v)\,dv \right\} f_U(u)\,du \\
&= \int_{-\infty}^{\infty} F_V(w-u)f_U(u)\,du.
\end{aligned}$$

By differentiating $F_W(w)$ with respect to w, we have the probability density function $f_W(w)$ given as

$$f_W(w) = \frac{d}{dw}\int_{-\infty}^{\infty} F_V(w-u)f_U(u)\,du = \int_{-\infty}^{\infty} f_V(w-u)f_U(u)\,du.$$

Using the same steps we can also obtain

$$f_W(w) = \int_{-\infty}^{\infty} f_U(w-v) f_V(v)\, dv.$$

To show that $Y = \sum_{i=1}^{n} X_i \sim \text{Gamma}\,(n, \lambda)$ where $X_1, X_2, \ldots, X_n \sim \text{Exp}(\lambda)$, we will prove the result via mathematical induction.
For $n = 1$, we have $Y = X_1 \sim \text{Exp}(\lambda)$ and the gamma density $f_Y(y)$ becomes

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

Therefore, the result is true for $n = 1$.
Let us assume that the result holds for $n = k$ and we now wish to compute the density for the case $n = k + 1$. Since $X_1, X_2, \ldots, X_{k+1}$ are all mutually independent and identically distributed, by setting $U = \sum_{i=1}^{k} X_i$ and $V = X_{k+1}$, and since $U \geq 0$, $V \geq 0$, the density of $Y = \sum_{i=1}^{k} X_i + X_{k+1}$ can be expressed as

$$\begin{aligned}
f_Y(y) &= \int_0^y f_V(y-u) f_U(u)\, du \\
&= \int_0^y \lambda e^{-\lambda(y-u)} \cdot \frac{(\lambda u)^{k-1}}{(k-1)!} \lambda e^{-\lambda u}\, du \\
&= \frac{\lambda^{k+1} e^{-\lambda y}}{(k-1)!} \int_0^y u^{k-1}\, du \\
&= \frac{(\lambda y)^k}{k!} \lambda e^{-\lambda y}
\end{aligned}$$

which shows the result is also true for $n = k + 1$. Thus, $Y = \sum_{i=1}^{n} X_i \sim \text{Gamma}\,(n, \lambda)$.
Given that $X_1, X_2, \ldots, X_n \sim \text{Exp}(\lambda)$ are independent and identically distributed with common mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$, therefore

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{n}{\lambda}$$

and

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = \frac{n}{\lambda^2}.$$

□

6. *Normal Distribution Property I.* Show that for constants a, L and U such that $t > 0$ and $L < U$,

$$\frac{1}{\sqrt{2\pi t}} \int_L^U e^{aw - \frac{1}{2}\left(\frac{w}{\sqrt{t}}\right)^2} dw = e^{\frac{1}{2}a^2 t} \left[\Phi\left(\frac{U - at}{\sqrt{t}}\right) - \Phi\left(\frac{L - at}{\sqrt{t}}\right)\right]$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.

Solution: Simplifying the integrand we have

$$\begin{aligned}
\frac{1}{\sqrt{2\pi t}}\int_L^U e^{aw-\frac{1}{2}\left(\frac{w}{\sqrt{t}}\right)^2}\,dw &= \frac{1}{\sqrt{2\pi t}}\int_L^U e^{-\frac{1}{2}\left(\frac{w^2-2awt}{t}\right)}dw\\
&= \frac{1}{\sqrt{2\pi t}}\int_L^U e^{-\frac{1}{2}\left[\left(\frac{w-at}{\sqrt{t}}\right)^2-a^2t\right]}dw\\
&= \frac{e^{\frac{1}{2}a^2t}}{\sqrt{2\pi t}}\int_L^U e^{-\frac{1}{2}\left(\frac{w-at}{\sqrt{t}}\right)^2}\,dw.
\end{aligned}$$

By setting $x=\dfrac{w-at}{\sqrt{t}}$ we can write

$$\int_L^U \frac{1}{\sqrt{2\pi t}}\,e^{-\frac{1}{2}\left(\frac{w-at}{\sqrt{t}}\right)^2}\,dw = \int_{\frac{L-at}{\sqrt{t}}}^{\frac{U-at}{\sqrt{t}}}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}dx = \Phi\left(\frac{U-at}{\sqrt{t}}\right)-\Phi\left(\frac{L-at}{\sqrt{t}}\right).$$

Thus, $\dfrac{1}{\sqrt{2\pi t}}\displaystyle\int_L^U e^{aw-\frac{1}{2}\left(\frac{w}{\sqrt{t}}\right)^2}\,dw = e^{\frac{1}{2}a^2t}\left[\Phi\left(\frac{U-at}{\sqrt{t}}\right)-\Phi\left(\frac{L-at}{\sqrt{t}}\right)\right].$

□

7. *Normal Distribution Property II*. Show that if $X\sim\mathcal{N}(\mu,\sigma^2)$ then for $\delta\in\{-1,1\}$,

$$\mathbb{E}[\max\{\delta(e^X-K),0\}] = \delta e^{\mu+\frac{1}{2}\sigma^2}\Phi\left(\frac{\delta(\mu+\sigma^2-\log K)}{\sigma}\right)-\delta K\Phi\left(\frac{\delta(\mu-\log K)}{\sigma}\right)$$

where $K>0$ and $\Phi(\cdot)$ denotes the cumulative standard normal distribution function.

Solution: We first let $\delta=1$,

$$\begin{aligned}
\mathbb{E}[\max\{e^X-K,0\}] &= \int_{\log K}^{\infty}(e^x-K)\ f_X(x)\ dx\\
&= \int_{\log K}^{\infty}(e^x-K)\,\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx\\
&= \int_{\log K}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2+x}\,dx - K\int_{\log K}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx.
\end{aligned}$$

By setting $w=\dfrac{x-\mu}{\sigma}$ and $z=w-\sigma$ we have

$$\mathbb{E}[\max\{e^X-K,0\}] = \int_{\frac{\log K-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}w^2+\sigma w+\mu}dw - K\int_{\frac{\log K-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}w^2}dw$$

$$
\begin{aligned}
&= e^{\mu+\frac{1}{2}\sigma^2}\int_{\frac{\log K-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}(w-\sigma)^2}dw - K\int_{\frac{\log K-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}w^2}dw\\
&= e^{\mu+\frac{1}{2}\sigma^2}\int_{\frac{\log K-\mu-\sigma^2}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz - K\int_{\frac{\log K-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}w^2}dw\\
&= e^{\mu+\frac{1}{2}\sigma^2}\Phi\left(\frac{\mu+\sigma^2-\log K}{\sigma}\right) - K\Phi\left(\frac{\mu-\log K}{\sigma}\right).
\end{aligned}
$$

Using similar steps for the case $\delta = -1$ we can also show that

$$
\mathbb{E}[\max\{K-e^X,0\}] = K\Phi\left(\frac{\log K-\mu}{\sigma}\right) - e^{\mu+\frac{1}{2}\sigma^2}\Phi\left(\frac{\log K-(\mu+\sigma^2)}{\sigma}\right).
$$

□

8. For $x > 0$ show that

$$
\left(\frac{1}{x}-\frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2} \le \int_x^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz \le \frac{1}{x\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}.
$$

Solution: Solving $\int_x^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz$ using integration by parts, we let $u = \frac{1}{z}$, $\frac{dv}{dz} = \frac{z}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}$ so that $\frac{du}{dz} = -\frac{1}{z^2}$ and $v = -\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}$. Therefore,

$$
\begin{aligned}
\int_x^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz &= -\left.\frac{1}{z\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\right|_x^{\infty} - \int_x^{\infty}\frac{1}{z^2\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz\\
&= \frac{1}{x\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2} - \int_x^{\infty}\frac{1}{z^2\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz\\
&\le \frac{1}{x\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}
\end{aligned}
$$

since $\int_x^{\infty}\frac{1}{z^2\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz > 0$.

To obtain the lower bound, we integrate $\int_x^{\infty}\frac{1}{z^2\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz$ by parts where we let $u = \frac{1}{z^3}$, $\frac{dv}{dz} = \frac{z}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}$ so that $\frac{du}{dz} = -\frac{3}{z^4}$ and $v = -\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}$ and hence

$$
\begin{aligned}
\int_x^{\infty}\frac{1}{z^2\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz &= -\left.\frac{1}{z^3\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\right|_x^{\infty} - \int_x^{\infty}\frac{3}{z^4\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz\\
&= \frac{1}{x^3\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2} - \int_x^{\infty}\frac{3}{z^4\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz.
\end{aligned}
$$

Therefore,

$$\begin{aligned}\int_x^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz &= \frac{1}{x\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} - \frac{1}{x^3\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} + \int_x^\infty \frac{3}{z^4\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz \\ &\geq \left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\end{aligned}$$

since $\displaystyle\int_x^\infty \frac{3}{z^4\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz > 0$. By taking into account both the lower and upper bounds we have

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \leq \int_x^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz \leq \frac{1}{x\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}.$$

□

9. *Lognormal Distribution I*. Let $Z \sim \mathcal{N}(0, 1)$, show that the moment generating function of a standard normal distribution is

$$\mathbb{E}\left(e^{\theta Z}\right) = e^{\frac{1}{2}\theta^2}$$

for a constant θ.
Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Y = e^X$ follows a lognormal distribution, $Y = e^X \sim$ log-$\mathcal{N}(\mu, \sigma^2)$ with probability density function

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}$$

with mean $\mathbb{E}(Y) = e^{\mu + \frac{1}{2}\sigma^2}$ and variance $\text{Var}(Y) = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2}$.

Solution: By definition

$$\begin{aligned}\mathbb{E}\left(e^{\theta Z}\right) &= \int_{-\infty}^\infty e^{\theta z} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(z-\theta)^2 + \frac{1}{2}\theta^2}\, dz \\ &= e^{\frac{1}{2}\theta^2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(z-\theta)^2}\, dz \\ &= e^{\frac{1}{2}\theta^2}.\end{aligned}$$

For $y > 0$, by definition

$$\mathbb{P}\left(e^X < y\right) = \mathbb{P}(X < \log y) = \int_{-\infty}^{\log y} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

and hence

$$f_Y(y) = \frac{d}{dy}\mathbb{P}\left(e^X < y\right) = \frac{1}{y\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}.$$

Given that $\log\ Y \sim \mathcal{N}(\mu, \sigma^2)$ we can write $\log\ Y = \mu + \sigma Z$, $Z \sim \mathcal{N}(0, 1)$ and hence

$$\mathbb{E}(Y) = \mathbb{E}\left(e^{\mu+\sigma Z}\right) = e^{\mu}\mathbb{E}\left(e^{\sigma Z}\right) = e^{\mu+\frac{1}{2}\sigma^2}$$

since $\mathbb{E}\left(e^{\theta Z}\right) = e^{\frac{1}{2}\theta^2}$ is the moment generating function of a standard normal distribution. Taking second moments,

$$\mathbb{E}\left(Y^2\right) = \mathbb{E}\left(e^{2\mu+2\sigma Z}\right) = e^{2\mu}\mathbb{E}\left(e^{2\sigma Z}\right) = e^{2\mu+2\sigma^2}.$$

Therefore,

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = e^{2\mu+2\sigma^2} - \left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2 = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2}.$$

□

10. *Lognormal Distribution II*. Let $X \sim \log\text{-}\mathcal{N}(\mu, \sigma^2)$, show that for $n \in \mathbb{N}$

$$\mathbb{E}\left(X^n\right) = e^{n\mu+\frac{1}{2}n^2\sigma^2}.$$

Solution: Given $X \sim \log\text{-}\mathcal{N}(\mu, \sigma^2)$ the density function is

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2}, \quad x > 0$$

and for $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}(X^n) &= \int_0^\infty x^n f_X(x)\ dx\\ &= \int_0^\infty \frac{1}{x\sigma\sqrt{2\pi}} x^n e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2} dx.\end{aligned}$$

By substituting $z = \dfrac{\log x - \mu}{\sigma}$ so that $x = e^{\sigma z+\mu}$ and $\dfrac{dz}{dx} = \dfrac{1}{x\sigma}$,

$$\begin{aligned}\mathbb{E}(X^n) &= \int_{-\infty}^\infty \frac{1}{x\sigma\sqrt{2\pi}}\ e^{n(\sigma z+\mu)}e^{-\frac{1}{2}z^2}x\sigma\ dz\\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}z^2+n(\sigma z+\mu)}\ dz\\ &= e^{n\mu+\frac{1}{2}n^2\sigma^2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(z-n\sigma)^2}\ dz\end{aligned}$$

$$= e^{n\mu+\frac{1}{2}n^2\sigma^2}$$

since $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(z-n\sigma)^2}\, dz = 1.$

□

11. *Folded Normal Distribution*. Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Y = |X|$ follows a folded normal distribution, $Y = |X| \sim \mathcal{N}_f(\mu, \sigma^2)$ with probability density function

$$f_Y(y) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y^2+\mu^2}{\sigma^2}\right)} \cosh\left(\frac{\mu y}{\sigma^2}\right)$$

with mean

$$\mathbb{E}(Y) = \sigma\sqrt{\frac{2}{\pi}} \quad e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right]$$

and variance

$$\mathrm{Var}(Y) = \mu^2 + \sigma^2 - \left\{\sigma\sqrt{\frac{2}{\pi}} \quad e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right]\right\}^2$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.

Solution: For $y > 0$, by definition

$$\mathbb{P}(|X| < y) = \mathbb{P}(-y < X < y) = \int_{-y}^{y} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

and hence

$$\begin{aligned} f_Y(y) = \frac{d}{dy}\mathbb{P}(|X| < y) &= \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{y+\mu}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\right] \\ &= \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y^2+\mu^2}{\sigma^2}\right)} \cosh\left(\frac{\mu y}{\sigma^2}\right). \end{aligned}$$

By definition

$$\begin{aligned} \mathbb{E}(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= \int_0^{\infty} \frac{y}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y+\mu}{\sigma}\right)^2} dy + \int_0^{\infty} \frac{y}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy. \end{aligned}$$

By setting $z = (y + \mu)/\sigma$ and $w = (y - \mu)/\sigma$ we have

$$\mathbb{E}(Y) = \int_{\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}(\sigma z - \mu)e^{-\frac{1}{2}z^2} dz + \int_{-\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}(\sigma w + \mu)e^{-\frac{1}{2}w^2} dw$$

$$
\begin{aligned}
&= \frac{\sigma}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} - \mu\left[1 - \Phi\left(\frac{\mu}{\sigma}\right)\right] + \frac{\sigma}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} - \mu\left[1 - \Phi\left(-\frac{\mu}{\sigma}\right)\right] \\
&= \frac{2\sigma}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \mu\left[\Phi\left(\frac{\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right] \\
&= \sigma\sqrt{\frac{2}{\pi}}\quad e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right].
\end{aligned}
$$

To evaluate $\mathbb{E}\left(Y^2\right)$ by definition,

$$
\begin{aligned}
\mathbb{E}\left(Y^2\right) &= \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy \\
&= \int_0^{\infty} \frac{y^2}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y+\mu}{\sigma}\right)^2} dy + \int_0^{\infty} \frac{y^2}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy.
\end{aligned}
$$

By setting $z = (y + \mu)/\sigma$ and $w = (y - \mu)/\sigma$ we have

$$
\begin{aligned}
\mathbb{E}\left(Y^2\right) &= \int_{\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}(\sigma z - \mu)^2 e^{-\frac{1}{2}z^2} dz + \int_{-\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}(\sigma w + \mu)^2 e^{-\frac{1}{2}w^2} dw \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mu/\sigma}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz - \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{\mu/\sigma}^{\infty} z e^{-\frac{1}{2}z^2} dz + \mu^2 \int_{\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} dz \\
&\quad + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\mu/\sigma}^{\infty} w^2 e^{-\frac{1}{2}w^2} dw + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\mu/\sigma}^{\infty} w e^{-\frac{1}{2}w^2} dw \\
&\quad + \mu^2 \int_{-\mu/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}w^2} dw \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[\left(\frac{\mu}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \sqrt{2\pi}\left(1 - \Phi\left(\frac{\mu}{\sigma}\right)\right)\right] - \frac{2\mu\sigma}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} \\
&\quad + \mu^2\left[1 - \Phi\left(\frac{\mu}{\sigma}\right)\right] \\
&\quad + \frac{\sigma^2}{\sqrt{2\pi}} \left[\left(-\frac{\mu}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \sqrt{2\pi}\left(1 - \Phi\left(-\frac{\mu}{\sigma}\right)\right)\right] + \frac{2\mu\sigma}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} \\
&\quad + \mu^2\left[1 - \Phi\left(-\frac{\mu}{\sigma}\right)\right] \\
&= \left(\mu^2 + \sigma^2\right)\left[2 - \Phi\left(\frac{\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right] \\
&= \mu^2 + \sigma^2.
\end{aligned}
$$

Therefore,

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = \mu^2 + \sigma^2 - \left\{ \sigma\sqrt{\frac{2}{\pi}} \quad e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right] \right\}^2.$$

□

12. *Chi-Square Distribution*. Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Y = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.
Let $Z_1, Z_2, \ldots, Z_n \sim \mathcal{N}(0, 1)$ be a sequence of independent and identically distributed random variables each following a standard normal distribution. Using mathematical induction show that

$$Z = Z_1^2 + Z_2^2 + \ldots + Z_n^2 \sim \chi^2(n)$$

where Z has a probability density function

$$f_Z(z) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}, \quad z \geq 0$$

such that

$$\Gamma\left(\frac{n}{2}\right) = \int_0^\infty e^{-x} x^{\frac{n}{2}-1} dx.$$

Finally, show that $\mathbb{E}(Z) = n$ and $\mathrm{Var}(Z) = 2n$.

Solution: By setting $Y = \dfrac{X-\mu}{\sigma}$ then

$$\mathbb{P}(Y \leq y) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq y\right) = \mathbb{P}(X \leq \mu + \sigma y).$$

Differentiating with respect to y,

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}\mathbb{P}(Y \leq y) \\
&= \frac{d}{dy}\int_{-\infty}^{\mu+\sigma y} \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{\mu+\sigma y-\mu}{\sigma}\right)^2} \sigma \\
&= \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}y^2}
\end{aligned}$$

which is a probability density function of $\mathcal{N}(0, 1)$.
For $Z = Z_1^2$ and given $Z \geq 0$, by definition

$$\mathbb{P}(Z \leq z) = \mathbb{P}\left(Z_1^2 \leq z\right) = \mathbb{P}\left(-\sqrt{z} < Z_1 < \sqrt{z}\right) = 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}z_1^2} \; dz_1$$

and hence

$$f_Z(z) = \frac{d}{dz}\left[2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z_1^2}\, dz_1\right] = \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}}, \quad z \ge 0$$

which is the probability density function of $\chi^2(1)$.
For $Z = Z_1^2 + Z_2^2$ such that $Z \ge 0$ we have

$$\mathbb{P}(Z \le z) = \mathbb{P}(Z_1^2 + Z_2^2 \le z) = \iint\limits_{z_1^2+z_2^2 \le z} \frac{1}{2\pi}\, e^{-\frac{1}{2}(z_1^2+z_2^2)}\, dz_1 dz_2.$$

Changing to polar coordinates (r, θ) such that $z_1 = r\cos\theta$ and $z_2 = r\sin\theta$ with the Jacobian determinant

$$|J| = \left|\frac{\partial(z_1, z_2)}{\partial(r, \theta)}\right| = \begin{vmatrix} \dfrac{\partial z_1}{\partial r} & \dfrac{\partial z_1}{\partial \theta} \\ \dfrac{\partial z_2}{\partial r} & \dfrac{\partial z_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

then

$$\iint\limits_{z_1^2+z_2^2 \le z} \frac{1}{2\pi}\, e^{-\frac{1}{2}(z_1^2+z_2^2)}\, dz_1 dz_2 = \int_{\theta=0}^{2\pi}\int_{r=0}^{\sqrt{z}} \frac{1}{2\pi}\, e^{-\frac{1}{2}r^2} r\, dr d\theta = 1 - e^{-\frac{1}{2}z}.$$

Thus,

$$\begin{aligned} f_Z(z) &= \frac{d}{dz}\left[\iint\limits_{z_1^2+z_2^2 \le z} \frac{1}{2\pi}\, e^{-\frac{1}{2}(z_1^2+z_2^2)}\, dz_1 dz_2\right] \\ &= \frac{1}{2}\, e^{-\frac{1}{2}z}, \quad z \ge 0 \end{aligned}$$

which is the probability density function of $\chi^2(2)$.
Assume the result is true for $n = k$ such that

$$U = Z_1^2 + Z_2^2 + \ldots + Z_k^2 \sim \chi^2(k)$$

and knowing that

$$V = Z_{k+1}^2 \sim \chi^2(1)$$

then, because $U \ge 0$ and $V \ge 0$ are independent, using the convolution formula the density of $Z = U + V = \sum_{i=1}^{k+1} Z_i^2$ can be written as

$$\begin{aligned} f_Z(z) &= \int_0^z f_V(z-u) f_U(u)\, du \\ &= \int_0^z \left\{\frac{1}{\sqrt{2\pi}}(z-u)^{-\frac{1}{2}} e^{-\frac{1}{2}(z-u)}\right\} \cdot \left\{\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} u^{\frac{1}{2}k-1} e^{-\frac{1}{2}u}\right\} du \end{aligned}$$

$$= \frac{e^{-\frac{1}{2}z}}{\sqrt{2\pi}2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} \int_0^z (z-u)^{-\frac{1}{2}} u^{\frac{k}{2}-1}\, du.$$

By setting $v = \dfrac{u}{z}$ we have

$$\begin{aligned}\int_0^z (z-u)^{-\frac{1}{2}} u^{\frac{k}{2}-1}\, du &= \int_0^1 (z-vz)^{-\frac{1}{2}} (vz)^{\frac{k}{2}-1} z\, dv \\ &= z^{\frac{k+1}{2}-1} \int_0^1 (1-v)^{-\frac{1}{2}} v^{\frac{k}{2}-1}\, dv\end{aligned}$$

and because $\displaystyle\int_0^1 (1-v)^{-\frac{1}{2}} v^{\frac{k}{2}-1}\, dv = B\left(\frac{1}{2}, \frac{k}{2}\right) = \frac{\sqrt{\pi}\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)}$ (see Appendix A) therefore

$$f_Z(z) = \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k+1}{2}\right)} z^{\frac{k+1}{2}-1} e^{-\frac{1}{2}z}, \quad z \geq 0$$

which is the probability density function of $\chi^2(k+1)$ and hence the result is also true for $n = k+1$. By mathematical induction we have shown

$$Z = Z_1^2 + Z_2^2 + \ldots + Z_n^2 \sim \chi^2(n).$$

By computing the moment generation of Z,

$$M_Z(t) = \mathbb{E}\left(e^{tZ}\right) = \int_0^\infty e^{tz} f_Z(z)\, dz = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty z^{\frac{n}{2}-1} e^{-\frac{1}{2}(1-2t)}\, dz$$

and by setting $w = \frac{1}{2}(1-2t)z$ we have

$$M_Z(t) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \left(\frac{2}{1-2t}\right)^{\frac{n}{2}} \int_0^\infty w^{\frac{n}{2}-1} e^{-w} dw = (1-2t)^{-\frac{n}{2}}, \quad t \in \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Thus,

$$M_Z'(t) = n(1-2t)^{-\left(\frac{n}{2}+1\right)}, \quad M_Z''(t) = 2n\left(\frac{n}{2}+1\right)(1-2t)^{-\left(\frac{n}{2}+2\right)}$$

such that

$$\mathbb{E}\,(Z) = M_Z'(0) = n, \quad \mathbb{E}\left(Z^2\right) = M_Z''(0) = 2n\left(\frac{n}{2}+1\right)$$

and

$$\mathrm{Var}\,(Z) = \mathbb{E}\left(Z^2\right) - [\mathbb{E}\,(Z)]^2 = 2n.$$

□

13. *Marginal Distributions of Bivariate Normal Distribution*. Let X and Y be jointly normally distributed with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient $\rho_{xy} \in (-1, 1)$ such that the joint density function is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}.$$

Show that $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$.

Solution: By definition,

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \\
&\quad \times e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[(1-\rho_{xy}^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \rho_{xy}^2\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy \\
&= \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} g(x, y)\, dy
\end{aligned}$$

where

$$g(x, y) = \frac{1}{\sqrt{1-\rho_{xy}^2}\,\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)\sigma_y^2}\left[y - \left(\mu_y + \rho_{xy}\sigma_y\left(\frac{x-\mu_x}{\sigma_x}\right)\right)\right]^2}$$

is the probability density function for $\mathcal{N}\left(\mu_y + \rho_{xy}\sigma_y\left(x-\mu_x\right)/\sigma_x, (1-\rho_{xy}^2)\sigma_y^2\right)$. Therefore,

$$\int_{-\infty}^{\infty} g(x, y)\, dy = 1$$

and hence

$$f_X(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}.$$

Thus, $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$. Using the same steps we can also show that $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$.

$\square$

14. *Covariance of Bivariate Normal Distribution*. Let X and Y be jointly normally distributed with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient $\rho_{xy} \in (-1, 1)$ such that the joint density function is

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}.$$

Show that the covariance of X and Y, $\mathrm{Cov}(X, Y) = \rho_{xy}\sigma_x\sigma_y$ and hence show that X and Y are independent if and only if $\rho_{xy} = 0$.

Solution: By definition, the covariance of X and Y is

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mu_x)(Y-\mu_y)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)\, f_{XY}(x,y)\, dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy - \mu_y \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{XY}(x,y)\, dx\,dy \\
&\quad -\mu_x \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf_{XY}(x,y)\, dx\,dy + \mu_x\mu_y \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy - \mu_y \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f_{XY}(x,y)\, dy\right] dx \\
&\quad -\mu_x \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\right] dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu_x\mu_y f_{XY}(x,y)\, dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy - \mu_y \int_{-\infty}^{\infty} xf_X(x)\, dx - \mu_x \int_{-\infty}^{\infty} yf_Y(y)\, dy \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu_x\mu_y f_{XY}(x,y)\, dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy - \mu_x\mu_y - \mu_x\mu_y + \mu_x\mu_y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy - \mu_x\mu_y
\end{aligned}$$

where $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx$. Using the result of Problem 1.2.2.13 (page 27) we can deduce that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\, dx\,dy = \int_{-\infty}^{\infty} \frac{x}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \left(\int_{-\infty}^{\infty} yg(x,y)\, dy\right) dx$$

where

$$g(x,y) = \frac{1}{\sqrt{1-\rho_{xy}^2}\,\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)\sigma_y^2}\left[y - \left(\mu_y + \rho_{xy}\sigma_y\left(\frac{x-\mu_x}{\sigma_x}\right)\right)\right]^2}$$

is the probability density function for $\mathcal{N}\left(\mu_y + \rho_{xy}\sigma_y\left(x-\mu_x\right)/\sigma_x, (1-\rho_{xy}^2)\sigma_y^2\right)$. Therefore,

$$\int_{-\infty}^{\infty} yg(x,y)\,dy = \mu_y + \rho_{xy}\sigma_y\left(\frac{x-\mu_x}{\sigma_x}\right).$$

Thus,

$$\begin{aligned}
\text{Cov}(X,Y) &= \int_{-\infty}^{\infty} \frac{x}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\left[\mu_y + \rho_{xy}\sigma_y\left(\frac{x-\mu_x}{\sigma_x}\right)\right] dx - \mu_x\mu_y \\
&= \mu_x\mu_y + \frac{\rho_{xy}\sigma_y}{\sigma_x}\int_{-\infty}^{\infty}\frac{x^2}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dx - \frac{\rho_{xy}\sigma_y\mu_x^2}{\sigma_x} - \mu_x\mu_y \\
&= \mu_x\mu_y + \frac{\rho_{xy}\sigma_y}{\sigma_x}\left(\sigma_x^2 + \mu_x^2\right) - \frac{\rho_{xy}\sigma_y\mu_x^2}{\sigma_x} - \mu_x\mu_y \\
&= \rho_{xy}\sigma_x\sigma_y
\end{aligned}$$

where

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty}\frac{x^2}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dx = \sigma_x^2 + \mu_x^2.$$

To show that X and Y are independent if and only if $\rho_{xy}=0$ we note that if $X \perp\!\!\!\perp Y$ then $\text{Cov}(X,Y)=0$, which implies $\rho_{xy}=0$. On the contrary, if $\rho_{xy}=0$ then from the joint density of (X,Y) we can express it as

$$f_{XY}(x,y) = f_X(x)\,f_Y(y)$$

where $f_X(x) = \displaystyle\int_{-\infty}^{\infty}\frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dx$ and $f_Y(y) = \displaystyle\int_{-\infty}^{\infty}\frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} dy$ and so $X \perp\!\!\!\perp Y$.

Thus, if the pair X and Y has a bivariate normal distribution with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation ρ_{xy} then $X \perp\!\!\!\perp Y$ if and only if $\rho_{xy}=0$.

□

15. *Minimum and Maximum of Two Correlated Normal Distributions*. Let X and Y be jointly normally distributed with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient $\rho_{xy} \in (-1,1)$ such that the joint density function is

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}.$$

Show that the distribution of $U=\min\{X,Y\}$ is

$$f_U(u) = \Phi\left(\frac{-u+\mu_y+\frac{\rho_{xy}\sigma_y}{\sigma_x}(u-\mu_x)}{\sigma_y\sqrt{1-\rho_{xy}^2}}\right) f_X(u) + \Phi\left(\frac{-u+\mu_x+\frac{\rho_{xy}\sigma_x}{\sigma_y}(u-\mu_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) f_Y(u).$$

and deduce that the distribution of $V = \max\{X, Y\}$ is

$$f_V(v) = \Phi\left(\frac{v - \mu_y - \frac{\rho_{xy}\sigma_y}{\sigma_x}(v - \mu_x)}{\sigma_y\sqrt{1 - \rho_{xy}^2}}\right) f_X(v) + \Phi\left(\frac{v - \mu_x - \frac{\rho_{xy}\sigma_x}{\sigma_y}(u - \mu_y)}{\sigma_x\sqrt{1 - \rho_{xy}^2}}\right) f_Y(v)$$

where $f_X(z) = \dfrac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{z-\mu_x}{\sigma_x}\right)^2}$, $f_Y(z) = \dfrac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{z-\mu_y}{\sigma_y}\right)^2}$ and $\Phi(\cdot)$ denotes the cumulative standard normal distribution function (cdf).

Solution: From Problems 1.2.2.13 (page 27) and 1.2.2.14 (page 28) we can show that $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, $\mathrm{Cov}(X, Y) = \rho_{xy}\sigma_x\sigma_y$ such that $\rho_{xy} \in (-1, 1)$.
For $U = \min\{X, Y\}$ then by definition the cumulative distribution function (cdf) of U is

$$\begin{aligned}
\mathbb{P}(U \le u) &= \mathbb{P}\left(\min\{X, Y\} \le u\right)\\
&= 1 - \mathbb{P}\left(\min\{X, Y\} > u\right)\\
&= 1 - \mathbb{P}\left(X > u, Y > u\right).
\end{aligned}$$

To derive the probability density function (pdf) of U we have

$$\begin{aligned}
f_U(u) &= \frac{d}{du}\mathbb{P}(U \le u)\\
&= -\frac{d}{du}\mathbb{P}\left(X > u, Y > u\right)\\
&= -\frac{d}{du}\int_{x=u}^{\infty}\int_{y=u}^{\infty}\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}}\\
&\quad \times e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy\, dx\\
&= g(u) + h(u)
\end{aligned}$$

where

$$g(u) = \int_{y=u}^{\infty}\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{u-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{u-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy$$

and

$$h(u) = \int_{x=u}^{\infty}\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{u-\mu_y}{\sigma_y}\right) + \left(\frac{u-\mu_y}{\sigma_y}\right)^2\right]}\, dx.$$

By focusing on

$$g(u) = \int_{y=u}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{u-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{u-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy$$

$$= \int_{y=u}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left\{\left[\left(\frac{y-\mu_y}{\sigma_y}\right)-\rho_{xy}\left(\frac{u-\mu_x}{\sigma_x}\right)\right]^2+\left(\frac{u-\mu_x}{\sigma_x}\right)^2(1-\rho_{xy}^2)\right\}}\, dy$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{u-\mu_x}{\sigma_x}\right)^2} \int_{y=u}^{\infty} \frac{1}{\sigma_y\sqrt{2\pi(1-\rho_{xy}^2)}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{y-\mu_y}{\sigma_y}\right)-\rho_{xy}\left(\frac{u-\mu_x}{\sigma_x}\right)\right]^2}\, dy$$

and letting $w = \dfrac{\dfrac{y-\mu_y}{\sigma_y} - \rho_{xy}\left(\dfrac{u-\mu_x}{\sigma_x}\right)}{\sqrt{1-\rho_{xy}^2}}$ we have

$$g(u) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{u-\mu_x}{\sigma_x}\right)^2} \int_{w=\frac{\frac{u-\mu_y}{\sigma_y}-\rho_{xy}\left(\frac{u-\mu_x}{\sigma_x}\right)}{\sqrt{1-\rho_{xy}^2}}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}w^2} dw$$

$$= \Phi\left(\frac{-u+\mu_y+\frac{\rho_{xy}\sigma_y}{\sigma_x}(u-\mu_x)}{\sigma_y\sqrt{1-\rho_{xy}^2}}\right) f_X(u).$$

In a similar vein we can also show

$$h(u) = \int_{x=u}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{u-\mu_y}{\sigma_y}\right)+\left(\frac{u-\mu_y}{\sigma_y}\right)^2\right]} dx$$

$$= \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{u-\mu_y}{\sigma_y}\right)^2} \int_{w=\frac{\frac{u-\mu_x}{\sigma_x}-\rho_{xy}\left(\frac{u-\mu_y}{\sigma_y}\right)}{\sqrt{1-\rho_{xy}^2}}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}w^2} dw$$

$$= \Phi\left(\frac{-u+\mu_x+\frac{\rho_{xy}\sigma_x}{\sigma_y}(u-\mu_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) f_Y(u).$$

Therefore,

$$f_U(u) = \Phi\left(\frac{-u+\mu_y+\frac{\rho_{xy}\sigma_y}{\sigma_x}(u-\mu_x)}{\sigma_y\sqrt{1-\rho_{xy}^2}}\right) f_X(u) + \Phi\left(\frac{-u+\mu_x+\frac{\rho_{xy}\sigma_x}{\sigma_y}(u-\mu_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) f_Y(u).$$

As for the case $V = \max\{X, Y\}$, then by definition the cdf of V is

$$\begin{aligned}\mathbb{P}(V \le v) &= \mathbb{P}(\max\{X, Y\} \le v)\\ &= \mathbb{P}(X \le v, Y \le v).\end{aligned}$$

The pdf of V is

$$\begin{aligned}
f_V(v) &= \frac{d}{dv}\mathbb{P}(V \le v)\\
&= \frac{d}{dv}\mathbb{P}(X \le v, Y \le v)\\
&= \frac{d}{dv}\int_{-\infty}^{x=v}\int_{-\infty}^{y=v} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\\
&\quad \times e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy\, dx\\
&= \int_{-\infty}^{y=v} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{v-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{v-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}\, dy\\
&\quad + \int_{-\infty}^{x=v} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{v-\mu_y}{\sigma_y}\right) + \left(\frac{v-\mu_y}{\sigma_y}\right)^2\right]}\, dx.
\end{aligned}$$

Following the same steps as described above we can write

$$f_V(v) = \Phi\left(\frac{v - \mu_y - \frac{\rho_{xy}\sigma_y}{\sigma_x}(v - \mu_x)}{\sigma_y\sqrt{1-\rho_{xy}^2}}\right) f_X(v) + \Phi\left(\frac{v - \mu_x - \frac{\rho_{xy}\sigma_x}{\sigma_y}(v - \mu_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) f_Y(v).$$

□

16. *Bivariate Standard Normal Distribution*. Let $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ be jointly normally distributed with correlation coefficient $\rho_{xy} \in (-1, 1)$ where the joint cumulative distribution function is

$$\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) = \int_{-\infty}^{\beta}\int_{-\infty}^{\alpha} \frac{1}{2\pi\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2}\left(\frac{x^2 - 2\rho_{xy}xy + y^2}{1-\rho_{xy}^2}\right)} dx\, dy.$$

By using the change of variables $Y = \rho_{xy}X + \sqrt{1-\rho_{xy}^2}Z$, $Z \sim \mathcal{N}(0, 1)$, $X \perp\!\!\!\perp Z$ show that

$$\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) = \int_{-\infty}^{\alpha} f_X(x)\Phi\left(\frac{\beta - \rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) dx$$

where $f_X(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.
Finally, deduce that $\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) + \boldsymbol{\Phi}(\alpha, -\beta, -\rho_{xy}) = \Phi(\alpha)$.

Solution: Let $y = \rho_{xy}x + \sqrt{1-\rho_{xy}^2}z$. Differentiating y with respect to z we have $\dfrac{dy}{dz} = \sqrt{1-\rho_{xy}^2}$, and hence

$$\begin{aligned}
\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) &= \int_{-\infty}^{\beta}\int_{-\infty}^{\alpha} \frac{1}{2\pi\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2}\left(\frac{x^2-2\rho_{xy}xy+y^2}{1-\rho_{xy}^2}\right)} dx\, dy \\
&= \int_{-\infty}^{\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}}\int_{-\infty}^{\alpha} \frac{1}{2\pi\sqrt{1-\rho_{xy}^2}} \\
&\quad \times e^{-\frac{1}{2}\left(\frac{x^2-2\rho_{xy}x(\rho_{xy}x+\sqrt{1-\rho_{xy}^2}z)+(\rho_{xy}x+\sqrt{1-\rho_{xy}^2}z)^2}{1-\rho_{xy}^2}\right)}\sqrt{1-\rho_{xy}^2}\, dx\, dz \\
&= \int_{-\infty}^{\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}}\int_{-\infty}^{\alpha} \frac{1}{2\pi}\, e^{-\frac{1}{2}(x^2+z^2)} dx\, dz \\
&= \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\left[\int_{-\infty}^{\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz\right] dx \\
&= \int_{-\infty}^{\alpha} f_X(x)\Phi\left(\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) dx.
\end{aligned}$$

Finally,

$$\begin{aligned}
\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) + \boldsymbol{\Phi}(\alpha, -\beta, -\rho_{xy}) &= \int_{-\infty}^{\alpha} f_X(x)\Phi\left(\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) dx \\
&\quad + \int_{-\infty}^{\alpha} f_X(x)\Phi\left(\frac{-\beta+\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) dx \\
&= \int_{-\infty}^{\alpha} f_X(x)\left[\Phi\left(\frac{\beta-\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) + \Phi\left(\frac{-\beta+\rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right)\right] dx
\end{aligned}$$

$$= \int_{-\infty}^{\alpha} f_X(x)\, dx$$
$$= \Phi(\alpha)$$

since $\Phi\left(\dfrac{\beta - \rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) + \Phi\left(\dfrac{-\beta + \rho_{xy}x}{\sqrt{1-\rho_{xy}^2}}\right) = 1.$

N.B. Similarly we can also show that

$$\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) = \int_{-\infty}^{\beta} f_Y(y)\Phi\left(\frac{\alpha - \rho_{xy}y}{\sqrt{1-\rho_{xy}^2}}\right) dy$$

where $f_Y(y) = \dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}$ and $\boldsymbol{\Phi}(\alpha, \beta, \rho_{xy}) + \boldsymbol{\Phi}(-\alpha, \beta, -\rho_{xy}) = \Phi(\beta)$.

□

17. *Bivariate Normal Distribution Property*. Let X and Y be jointly normally distributed with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient $\rho_{xy} \in (-1, 1)$ such that the joint density function is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}.$$

Show that

$$\begin{aligned}\mathbb{E}\left[\max\{e^X - e^Y, 0\}\right] &= e^{\mu_x + \frac{1}{2}\sigma_x^2}\Phi\left(\frac{\mu_x - \mu_y + \sigma_x(\sigma_x - \rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2}}\right)\\ &\quad - e^{\mu_y + \frac{1}{2}\sigma_y^2}\Phi\left(\frac{\mu_x - \mu_y - \sigma_y(\sigma_y - \rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2}}\right).\end{aligned}$$

Solution: By definition

$$\begin{aligned}\mathbb{E}\left[\max\{e^X - e^Y, 0\}\right] &= \int_{y=-\infty}^{y=\infty}\int_{x=y}^{x=\infty} (e^x - e^y)\, f_{XY}(x, y)\, dx dy\\ &= I_1 - I_2\end{aligned}$$

where $I_1 = \displaystyle\int_{y=-\infty}^{y=\infty}\int_{x=y}^{x=\infty} e^x f_{XY}(x, y)\, dx dy$ and $I_2 = \displaystyle\int_{y=-\infty}^{y=\infty}\int_{x=y}^{x=\infty} e^y f_{XY}(x, y)\, dx dy$.

For the case $I_1 = \int_{y=-\infty}^{y=\infty} \int_{x=y}^{x=\infty} e^x f_{XY}(x,y)\ dx dy$ we have

$$I_1 = \int_{y=-\infty}^{y=\infty} \int_{x=y}^{x=\infty} \frac{e^x}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\ e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \int_{y=-\infty}^{y=\infty} \frac{1}{\sigma_y\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

$$\times\left[\int_{x=y}^{x=\infty} \frac{e^x}{\sigma_x\sqrt{2\pi(1-\rho_{xy}^2)}}\ e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[x-\left(\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right]^2} dx\right] dy$$

$$= \int_{y=-\infty}^{y=\infty} \frac{1}{\sigma_y\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \left[\int_{x=y}^{x=\infty} e^x g(x,y)\ dx\right]\ dy$$

where $g(x,y) = \dfrac{1}{\sigma_x\sqrt{2\pi(1-\rho_{xy}^2)}}\ e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[x-\left(\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right]^2}$ which is the probability density function of $\mathcal{N}\left[\mu_x + \rho_{xy}\sigma_x\left(\dfrac{y-\mu_y}{\sigma_y}\right), (1-\rho_{xy})^2\sigma_x^2\right]$. Thus, from Problem 1.2.2.7 (page 18) we can deduce

$$\int_{x=y}^{x=\infty} e^x g(x,y)\ dx = e^{\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)+\frac{1}{2}(1-\rho_{xy}^2)\sigma_x^2}$$

$$\times\Phi\left(\frac{\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)+(1-\rho_{xy})^2\sigma_x^2-y}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right).$$

Thus, we can write

$$I_1 = e^{\mu_x+\frac{1}{2}(1-\rho_{xy}^2)\sigma_x^2}$$

$$\times\int_{y=-\infty}^{y=\infty} \frac{1}{\sigma_y\sqrt{2\pi}}\ e^{-\frac{1}{2}\left[\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

$$\times\Phi\left(\frac{\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)+(1-\rho_{xy})^2\sigma_x^2-y}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)\ dy$$

$$= e^{\mu_x+\frac{1}{2}\sigma_x^2} \int_{y=-\infty}^{y=\infty} \frac{1}{\sigma_y\sqrt{2\pi}}\; e^{-\frac{1}{2}\left(\frac{y-\mu_y-\rho_{xy}\sigma_x\sigma_y}{\sigma_y}\right)^2}$$

$$\times\, \Phi\left(\frac{\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)+(1-\rho_{xy})^2\sigma_x^2-y}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) dy.$$

Let $u = \dfrac{y-\mu_y-\rho_{xy}\sigma_x\sigma_y}{\sigma_y}$ then from the change of variables

$$I_1 = e^{\mu_x+\frac{1}{2}\sigma_x^2} \int_{u=-\infty}^{u=\infty} \frac{1}{\sqrt{2\pi}}\; e^{-\frac{1}{2}u^2}\Phi\left(\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)-u(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) du$$

$$= e^{\mu_x+\frac{1}{2}\sigma_x^2} \int_{u=-\infty}^{u=\infty} \int_{v=-\infty}^{v=\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}-\frac{u(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi}\; e^{-\frac{1}{2}(u^2+v^2)}\, dvdu.$$

By setting $w = v + \dfrac{u(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}$ and $\sigma^2 = \sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2$,

$$\begin{aligned}
I_1 &= e^{\mu_x+\frac{1}{2}\sigma_x^2} \\
&\quad \times \int_{u=-\infty}^{u=\infty} \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi} \\
&\quad \times e^{-\frac{1}{2}\left[w^2-2uw\left(\frac{\sigma_y-\rho_{xy}\sigma_x}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)+\left(1+\frac{(\sigma_y-\rho_{xy}\sigma_x)^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)u^2\right]} dwdu \\
&= e^{\mu_x+\frac{1}{2}\sigma_x^2} \\
&\quad \times \int_{u=-\infty}^{u=\infty} \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi} \\
&\quad \times e^{-\frac{1}{2}\left(\frac{\sigma^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)\left[\frac{w^2}{\left(\frac{\sigma^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)}-2uw\frac{\sigma_x(\sigma_y-\rho_{xy}\sigma_x)\sqrt{1-\rho_{xy}^2}}{\sigma^2}+u^2\right]} dwdu.
\end{aligned}$$

Finally, by setting $\overline{w} = \dfrac{w}{\left(\dfrac{\sigma}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)}$,

$$I_1 = e^{\mu_x+\frac{1}{2}\sigma_x^2} \int_{u=-\infty}^{u=\infty} \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sigma}} \frac{1}{2\pi\sqrt{1-\overline{\rho}_{xy}^2}} e^{-\frac{1}{2(1-\overline{\rho}_{xy}^2)}(\overline{w}^2-2\overline{\rho}_{xy}u\overline{w}+u^2)} d\overline{w}du$$

where $\overline{\rho}_{xy} = \dfrac{\sigma_y - \rho_{xy}\sigma_x}{\sigma}$.
Therefore,

$$\begin{aligned} I_1 &= e^{\mu_x+\frac{1}{2}\sigma_x^2}\boldsymbol{\Phi}\left(\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2-2\rho_{xy}\sigma_x\sigma_y+\sigma_y^2}}, \infty, \overline{\rho}_{xy}\right) \\ &= e^{\mu_x+\frac{1}{2}\sigma_x^2}\Phi\left(\frac{\mu_x-\mu_y+\sigma_x(\sigma_x-\rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2-2\rho_{xy}\sigma_x\sigma_y+\sigma_y^2}}\right) \end{aligned}$$

where $\boldsymbol{\Phi}(\alpha, \beta, \rho) = \int_{-\infty}^{\beta}\int_{-\infty}^{\alpha} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right)} dxdy$ is the cumulative distribution function of a standard bivariate normal.
Using similar steps for the case $I_2 = \int_{y=-\infty}^{y=\infty}\int_{x=y}^{x=\infty} e^y f_{XY}(x,y)\, dxdy$ we have

$$\begin{aligned} I_2 &= \int_{y=-\infty}^{y=\infty}\int_{x=y}^{x=\infty} \frac{e^y}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]} \\ &= \int_{y=-\infty}^{y=\infty} \frac{e^y}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \\ &\quad \times\left[\int_{x=y}^{x=\infty} \frac{1}{\sigma_x\sqrt{2\pi(1-\rho_{xy}^2)}} e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[x-\left(\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right]^2} dx\right] dy \\ &= \int_{y=-\infty}^{y=\infty} \frac{e^y}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}\left[\int_{x=y}^{x=\infty} g(x,y)\, dx\right] dy \end{aligned}$$

where $g(x,y) = \dfrac{1}{\sigma_x\sqrt{2\pi(1-\rho_{xy}^2)}}\, e^{-\frac{1}{2(1-\rho_{xy}^2)}\left[x-\left(\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right]^2}$ which is the probability density function of $\mathcal{N}\left[\mu_x + \rho_{xy}\sigma_x\left(\dfrac{y-\mu_y}{\sigma_y}\right), (1-\rho_{xy})^2\sigma_x^2\right]$.

Thus,

$$I_2 = \int_{y=-\infty}^{y=\infty} \frac{e^y}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \left[\int_{x=y}^{x=\infty} \frac{1}{\sigma_x\sqrt{2\pi(1-\rho_{xy}^2)}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x-\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)^2} dx\right] dy$$

and by setting $z = \dfrac{x-\mu_x-\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)}{\sigma_x\sqrt{1-\rho_{xy}^2}}$,

$$\begin{aligned}
I_2 &= \int_{y=-\infty}^{y=\infty} \frac{e^y}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \left[\int_{z=\frac{y-\mu_x-\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)}{\sigma_x\sqrt{1-\rho_{xy}^2}}}^{z=\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz\right] dy \\
&= \int_{y=-\infty}^{y=\infty} \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left[\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2y\right]} \Phi\left(\frac{-y+\mu_x+\rho_{xy}\sigma_x\left(\frac{y-\mu_y}{\sigma_y}\right)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) dy.
\end{aligned}$$

By setting $\overline{z} = \dfrac{y-\mu_y}{\sigma_y}$ therefore

$$I_2 = e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{z=-\infty}^{z=\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(\overline{z}-\sigma_y)^2} \Phi\left(\frac{\mu_x-\mu_y-\overline{z}(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) d\overline{z}$$

and substituting $u = \overline{z}-\sigma_y$,

$$\begin{aligned}
I_2 &= e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{u=-\infty}^{u=\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2} \Phi\left(\frac{\mu_x-\mu_y-(u+\sigma_y)(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) du \\
&= e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{u=-\infty}^{u=\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2} \Phi\left(\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}} - \frac{u(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right) du
\end{aligned}$$

$$= e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{u=-\infty}^{u=\infty} \int_{v=-\infty}^{v=\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}-\frac{u(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi}\ e^{-\frac{1}{2}(u^2+v^2)}\ dvdu.$$

Using the same steps as described before, we let $w = v + \dfrac{u(\sigma_y - \rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}$ and $\sigma^2 = \sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2$ so that

$$\begin{aligned}
I_2 &= e^{\mu_y+\frac{1}{2}\sigma_y^2} \\
&\quad \times \int_{u=-\infty}^{u=\infty} \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi} \\
&\quad \times e^{-\frac{1}{2}\left[w^2-2uw\left(\frac{\sigma_y-\rho_{xy}\sigma_x}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)+\left(1+\frac{(\sigma_y-\rho_{xy}\sigma_x)^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)u^2\right]} dwdu \\
&= e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{u=-\infty}^{u=\infty} \\
&\quad \times \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sigma_x\sqrt{1-\rho_{xy}^2}}} \frac{1}{2\pi} \\
&\quad \times e^{-\frac{1}{2}\left(\frac{\sigma^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)\left[\frac{w^2}{\left(\frac{\sigma^2}{\sigma_x^2(1-\rho_{xy}^2)}\right)}-2uw\frac{\sigma_x(\sigma_y-\rho_{xy}\sigma_x)\sqrt{1-\rho_{xy}^2}}{\sigma^2}+u^2\right]} dwdu.
\end{aligned}$$

By setting $\overline{w} = \dfrac{w}{\left(\dfrac{\sigma}{\sigma_x\sqrt{1-\rho_{xy}^2}}\right)}$,

$$I_2 = e^{\mu_y+\frac{1}{2}\sigma_y^2} \int_{u=-\infty}^{u=\infty} \int_{w=-\infty}^{w=\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sigma}} \frac{1}{2\pi\sqrt{1-\overline{\rho}_{xy}^2}}\ e^{-\frac{1}{2(1-\overline{\rho}_{xy}^2)}(\overline{w}^2-2\overline{\rho}_{xy}u\overline{w}+u^2)}\ d\overline{w}du$$

where $\overline{\rho}_{xy} = \dfrac{\sigma_y - \rho_{xy}\sigma_x}{\sigma}$, thus

$$\begin{aligned}
I_2 &= e^{\mu_y+\frac{1}{2}\sigma_y^2}\boldsymbol{\Phi}\left(\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sqrt{\sigma_x^2-2\rho_{xy}\sigma_x\sigma_y+\sigma_y^2}}, \infty, \overline{\rho}_{xy}\right) \\
&= e^{\mu_y+\frac{1}{2}\sigma_y^2}\Phi\left(\frac{\mu_x-\mu_y-\sigma_y(\sigma_y-\rho_{xy}\sigma_x)}{\sqrt{\sigma_x^2-2\rho_{xy}\sigma_x\sigma_y+\sigma_y^2}}\right).
\end{aligned}$$

By substituting I_1 and I_2 back into $\mathbb{E}[\max\{e^X - e^Y, 0\}]$ we have

$$\mathbb{E}\left[\max\left\{e^X - e^Y, 0\right\}\right] = e^{\mu_x + \frac{1}{2}\sigma_x^2}\Phi\left(\frac{\mu_x - \mu_y + \sigma_x(\sigma_x - \rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2}}\right)$$

$$- e^{\mu_y + \frac{1}{2}\sigma_y^2}\Phi\left(\frac{\mu_x - \mu_y - \sigma_y(\sigma_y - \rho_{xy}\sigma_y)}{\sqrt{\sigma_x^2 - 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2}}\right).$$

□

18. *Markov's Inequality*. Let X be a non-negative random variable with mean μ. For $\alpha > 0$, show that

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mu}{\alpha}.$$

Solution: Since $\alpha > 0$, we can write

$$\mathbb{1}_{\{X\geq\alpha\}} = \begin{cases} 1 & X \geq \alpha \\ 0 & \text{otherwise} \end{cases}$$

and since $X \geq 0$, we can deduce that

$$\mathbb{1}_{\{X\geq\alpha\}} \leq \frac{X}{\alpha}.$$

Taking expectations

$$\mathbb{E}\left(\mathbb{1}_{\{X\geq\alpha\}}\right) \leq \frac{\mathbb{E}(X)}{\alpha}$$

and since $\mathbb{E}\left(\mathbb{1}_{\{X\geq\alpha\}}\right) = 1 \cdot \mathbb{P}(X \geq \alpha) + 0 \cdot \mathbb{P}(X \leq \alpha) = \mathbb{P}(X \geq \alpha)$ and $\mathbb{E}(X) = \mu$, we have

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mu}{\alpha}.$$

N.B. Alternatively, we can also show the result as

$$\mathbb{E}(X) = \int_0^\infty \left(1 - F_x(u)\right)\, du \geq \int_0^\alpha \left(1 - F_x(u)\right)\, du \geq \alpha\left(1 - F_x(\alpha)\right)$$

and hence it follows that $\mathbb{P}(X \geq \alpha) = 1 - F_x(\alpha) \leq \dfrac{\mathbb{E}(X)}{\alpha}$.

□

19. *Chebyshev's Inequality*. Let X be a random variable with mean μ and variance σ^2. Then for $k > 0$, show that

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Solution: Take note that $|X - \mu| \geq k$ if and only if $(X - \mu)^2 \geq k^2$. Because $(X - \mu)^2 \geq 0$, and by applying Markov's inequality (see Problem 1.2.2.18, page 40) we have

$$\mathbb{P}\left((X - \mu)^2 \geq k^2\right) \leq \frac{\mathbb{E}\left[(X - \mu)^2\right]}{k^2} = \frac{\sigma^2}{k^2}$$

and hence the above inequality is equivalent to

$$\mathbb{P}\left(|X-\mu| \geq k\right) \leq \frac{\sigma^2}{k^2}.$$

□

1.2.3 Properties of Expectations

1. Show that if X is a random variable taking non-negative values then

$$\mathbb{E}\left(X\right) = \begin{cases} \displaystyle\sum_{x=0}^{\infty} \mathbb{P}(X > x) & \text{if } X \text{ is a discrete random variable} \\ \displaystyle\int_0^{\infty} \mathbb{P}(X \geq x)\ dx & \text{if } X \text{ is a continuous random variable.} \end{cases}$$

Solution: We first show the result when X takes non-negative integer values only. By definition

$$\begin{aligned} \mathbb{E}(X) &= \sum_{y=0}^{\infty} y\mathbb{P}(X=y) \\ &= \sum_{y=0}^{\infty}\sum_{x=0}^{y} \mathbb{P}(X=y) \\ &= \sum_{x=0}^{\infty}\sum_{y=x+1}^{\infty} \mathbb{P}(X=y) \\ &= \sum_{x=0}^{\infty} \mathbb{P}(X>x). \end{aligned}$$

For the case when X is a continuous random variable taking non-negative values we have

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} yf_X(y)\ dy \\ &= \int_0^{\infty}\left\{\int_0^{y} f_X(y)\ dx\right\} dy \\ &= \int_0^{\infty}\left\{\int_x^{\infty} f_X(y)\ dy\right\} dx \\ &= \int_0^{\infty} \mathbb{P}(X \geq x)\ dx. \end{aligned}$$

□

2. *Hölder's Inequality*. Let $\alpha, \beta \geq 0$ and for $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ show that the following inequality:

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}$$

holds. Finally, if X and Y are a pair of jointly continuous variables, show that

$$\mathbb{E}(|XY|) \leq \{\mathbb{E}(|X^p|)\}^{1/p}\{\mathbb{E}(|Y^q|)\}^{1/q}.$$

Solution: The inequality certainly holds for $\alpha = 0$ and $\beta = 0$. Let $\alpha = e^{x/p}$ and $\beta = e^{y/q}$ where $x, y \in \mathbb{R}$. By substituting $\lambda = 1/p$ and $1 - \lambda = 1/q$,

$$e^{\lambda x + (1-\lambda)y} \leq \lambda e^x + (1-\lambda)e^y$$

holds true since the exponential function is a convex function and hence

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}.$$

By setting

$$\alpha = \frac{|X|}{\{\mathbb{E}(|X^p|)\}^{1/p}}, \quad \beta = \frac{|Y|}{\{\mathbb{E}(|Y^q|)\}^{1/q}}$$

hence

$$\frac{|XY|}{\{\mathbb{E}(|X^p|)\}^{1/p}\{\mathbb{E}(|Y^q|)\}^{1/q}} \leq \frac{|X|^p}{p\,\mathbb{E}(|X^p|)} + \frac{|Y|^q}{q\,\mathbb{E}(|Y^q|)}.$$

Taking expectations we obtain

$$\mathbb{E}(|XY|) \leq \{\mathbb{E}(|X^p|)\}^{1/p}\{\mathbb{E}(|Y^q|)\}^{1/q}.$$

□

3. *Minkowski's Inequality*. Let X and Y be a pair of jointly continuous variables, show that if $p \geq 1$ then

$$\left\{\mathbb{E}\left(|X+Y|^p\right)\right\}^{1/p} \leq \{\mathbb{E}(|X^p|)\}^{1/p} + \{\mathbb{E}(|Y^p|)\}^{1/p}.$$

Solution: Since $\mathbb{E}(|X+Y|) \leq \mathbb{E}(|X|) + \mathbb{E}(|Y|)$ and using Hölder's inequality we can write

$$\begin{aligned}
\mathbb{E}(|X+Y|^p) &= \mathbb{E}\left(|X+Y||X+Y|^{p-1}\right) \\
&\leq \mathbb{E}\left(|X||X+Y|^{p-1}\right) + \mathbb{E}\left(|Y||X+Y|^{p-1}\right) \\
&\leq \{\mathbb{E}(|X^p|)\}^{1/p}\{\mathbb{E}\left(|X+Y|^{(p-1)q}\right)\}^{1/q} + \{\mathbb{E}(|Y^p|)\}^{1/p}\{\mathbb{E}\left(|X+Y|^{(p-1)q}\right)\}^{1/q} \\
&= \{\mathbb{E}(|X^p|)\}^{1/p}\{\mathbb{E}(|X+Y|^p)\}^{1/q} + \{\mathbb{E}(|Y^p|)\}^{1/p}\{\mathbb{E}(|X+Y|^p)\}^{1/q}
\end{aligned}$$

since $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

Dividing the inequality by $\{\mathbb{E}(|X+Y|^p)\}^{1/q}$ we get

$$\{\mathbb{E}\left(|X+Y|^p\right)\}^{1/p} \leq \{\mathbb{E}(|X^p|)\}^{1/p} + \{\mathbb{E}(|Y^p|)\}^{1/p}.$$

□

4. *Change of Measure*. Let Ω be a probability space and let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on Ω. Let $Z(\omega)$ be the Radon–Nikodým derivative defined as

$$Z(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}$$

such that $\mathbb{P}(Z > 0) = 1$. By denoting $\mathbb{E}^{\mathbb{P}}$ and $\mathbb{E}^{\mathbb{Q}}$ as expectations under the measure $\mathbb{P}$ and $\mathbb{Q}$, respectively, show that for any random variable X,

$$\mathbb{E}^{\mathbb{Q}}(X) = \mathbb{E}^{\mathbb{P}}(XZ), \quad \mathbb{E}^{\mathbb{P}}(X) = \mathbb{E}^{\mathbb{Q}}\left(\frac{X}{Z}\right).$$

Solution: By definition

$$\mathbb{E}^{\mathbb{Q}}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{Q}(\omega) = \sum_{\omega\in\Omega} X(\omega)Z(\omega)\mathbb{P}(\omega) = \mathbb{E}^{\mathbb{P}}(XZ).$$

Similarly

$$\mathbb{E}^{\mathbb{P}}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega) = \sum_{\omega\in\Omega} \frac{X(\omega)}{Z(\omega)}\mathbb{Q}(\omega) = \mathbb{E}^{\mathbb{Q}}\left(\frac{X}{Z}\right).$$

□

5. *Conditional Probability*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If $\mathbb{I}_A$ is an indicator random variable for an event A defined as

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

show that

$$\mathbb{E}(\mathbb{I}_A|\mathscr{G}) = \mathbb{P}(A|\mathscr{G}).$$

Solution: Since $\mathbb{E}(\mathbb{I}_A|\mathscr{G})$ is $\mathscr{G}$ measurable we need to show that the following partial averaging property:

$$\int_B \mathbb{E}(\mathbb{I}_A|\mathscr{G})\, d\mathbb{P} = \int_B \mathbb{I}_A\, d\mathbb{P} = \int_B \mathbb{P}(A|\mathscr{G})\, d\mathbb{P}$$

is satisfied for $B \in \mathscr{G}$. Setting

$$\mathbb{I}_B(\omega) = \begin{cases} 1 & \text{if } \omega \in B \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbb{I}_{A\cap B}(\omega) = \begin{cases} 1 & \text{if } \omega \in A \cap B \\ 0 & \text{otherwise} \end{cases}$$

and expanding $\int_B \mathbb{P}(A|\mathscr{G})\, d\mathbb{P}$ we have

$$\int_B \mathbb{P}(A|\mathscr{G})\, d\mathbb{P} = \mathbb{P}(A \cap B) = \int_\Omega \mathbb{I}_{A\cap B}\, d\mathbb{P} = \int_\Omega \mathbb{I}_A \cdot \mathbb{I}_B\, d\mathbb{P} = \int_B \mathbb{I}_A\, d\mathbb{P}.$$

Since $\mathbb{E}(\mathbb{I}_A|\mathscr{G})$ is $\mathscr{G}$ measurable we have

$$\int_B \mathbb{E}(\mathbb{I}_A|\mathscr{G})\, d\mathbb{P} = \int_B \mathbb{I}_A\, d\mathbb{P}$$

and hence $\mathbb{E}(\mathbb{I}_A|\mathscr{G}) = \mathbb{P}(A|\mathscr{G})$.

□

6. *Linearity*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If $X_1, X_2, \ldots, X_n$ are integrable random variables and $c_1, c_2, \ldots, c_n$ are constants, show that

$$\mathbb{E}\left(c_1X_1 + c_2X_2 + \ldots + c_nX_n|\mathscr{G}\right) = c_1\mathbb{E}(X_1|\mathscr{G}) + c_2\mathbb{E}(X_2|\mathscr{G}) + \ldots + c_n\mathbb{E}(X_n|\mathscr{G}).$$

Solution: Given $\mathbb{E}\left(c_1X_1 + c_2X_2 + \ldots + c_nX_n|\mathscr{G}\right)$ is $\mathscr{G}$ measurable, and for any $A \in \mathscr{G}$,

$$\begin{aligned}\int_A \mathbb{E}\left(c_1X_1 + c_2X_2 + \ldots + c_nX_n|\mathscr{G}\right)\, d\mathbb{P} &= \int_A (c_1X_1 + c_2X_2 + \ldots + c_nX_n)\, d\mathbb{P}\\ &= c_1\int_A X_1\, d\mathbb{P} + c_2\int_A X_2\, d\mathbb{P}\\ &\quad + \ldots + c_n\int_A X_n\, d\mathbb{P}.\end{aligned}$$

Since $\displaystyle\int_A X_i\, d\mathbb{P} = \int_A \mathbb{E}(X_i|\mathscr{G})\, d\mathbb{P}$ for $i = 1, 2, \ldots, n$ therefore $\mathbb{E}\left(c_1X_1 + c_2X_2 + \ldots + c_nX_n|\mathscr{G}\right) = c_1\mathbb{E}(X_1|\mathscr{G}) + c_2\mathbb{E}(X_2|\mathscr{G}) + \ldots + c_n\mathbb{E}(X_n|\mathscr{G})$.

□

7. *Positivity*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If X is an integrable random variable such that $X \geq 0$ almost surely then show that

$$\mathbb{E}(X|\mathscr{G}) \geq 0$$

almost surely.

Solution: Let $A = \{w \in \Omega : \mathbb{E}(X|\mathscr{G}) < 0\}$ and since $\mathbb{E}(X|\mathscr{G})$ is $\mathscr{G}$ measurable therefore $A \in \mathscr{G}$. Thus, from the partial averaging property we have

$$\int_A \mathbb{E}(X|\mathscr{G})\, d\mathbb{P} = \int_A X\, d\mathbb{P}.$$

Since $X \geq 0$ almost surely therefore $\displaystyle\int_A X\, d\mathbb{P} \geq 0$ but $\displaystyle\int_A \mathbb{E}(X|\mathscr{G})\, d\mathbb{P} < 0$, which is a contradiction. Thus, $\mathbb{P}(A) = 0$, which implies $\mathbb{E}(X|\mathscr{G}) \geq 0$ almost surely.

□

8. *Monotonicity*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If X and Y are integrable random variables such that $X \leq Y$ almost surely then show that

$$\mathbb{E}(X|\mathscr{G}) \leq \mathbb{E}(Y|\mathscr{G}).$$

Solution: Since $\mathbb{E}(X - Y|\mathscr{G})$ is $\mathscr{G}$ measurable, for $A \in \mathscr{G}$ we can write

$$\int_A \mathbb{E}\left(X - Y|\mathscr{G}\right)\ d\mathbb{P} = \int_A (X - Y)\ d\mathbb{P}$$

and since $X \leq Y$, from Problem 1.2.3.7 (page 44) we can deduce that

$$\int_A \mathbb{E}\left(X - Y|\mathscr{G}\right)\ d\mathbb{P} \leq 0$$

and hence

$$\mathbb{E}\left(X - Y|\mathscr{G}\right) \leq 0.$$

Using the linearity of conditional expectation (see Problem 1.2.3.6, page 44)

$$\mathbb{E}\left(X - Y|\mathscr{G}\right) = \mathbb{E}(X|\mathscr{G}) - \mathbb{E}(Y|\mathscr{G}) \leq 0$$

and therefore $\mathbb{E}(X|\mathscr{G}) \leq \mathbb{E}(Y|\mathscr{G})$.

□

9. *Computing Expectations by Conditioning*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). Show that

$$\mathbb{E}[\mathbb{E}(X|\mathscr{G})] = \mathbb{E}(X).$$

Solution: From the partial averaging property we have, for $A \in \mathscr{G}$,

$$\int_A \mathbb{E}(X|\mathscr{G})\ d\mathbb{P} = \int_A X\ d\mathbb{P}$$

or

$$\mathbb{E}[\,\mathbb{I}_A \cdot \mathbb{E}(X|\mathscr{G})] = \mathbb{E}(\mathbb{I}_A \cdot X)$$

where

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

is a $\mathscr{G}$ measurable random variable. By setting $A = \Omega$ we obtain $\mathbb{E}[\mathbb{E}(X|\mathscr{G})] = \mathbb{E}(X)$.

□

10. *Taking Out What is Known*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If X and Y are integrable random variables and if X is $\mathscr{G}$ measurable show that

$$\mathbb{E}(XY|\mathscr{G}) = X \cdot \mathbb{E}(Y|\mathscr{G}).$$

Solution: Since X and $\mathbb{E}(Y|\mathscr{G})$ are $\mathscr{G}$ measurable therefore $X \cdot \mathbb{E}(Y|\mathscr{G})$ is also $\mathscr{G}$ measurable and it satisfies the first property of conditional expectation. By calculating the partial averaging of $X \cdot \mathbb{E}(Y|\mathscr{G})$ over a set $A \in \mathscr{G}$ and by defining

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

such that $\mathbb{I}_A$ is a $\mathscr{G}$ measurable random variable we have

$$\begin{aligned} \int_A X \cdot \mathbb{E}(Y|\mathscr{G})\ d\mathbb{P} &= \mathbb{E}[\mathbb{I}_A \cdot X\mathbb{E}(Y|\mathscr{G})] \\ &= \mathbb{E}[\mathbb{I}_A \cdot XY] \\ &= \int_A XY\ d\mathbb{P}. \end{aligned}$$

Thus, $X \cdot \mathbb{E}(Y|\mathscr{G})$ satisfies the partial averaging property by setting $\int_A XY\ d\mathbb{P} = \int_A \mathbb{E}(XY|\mathscr{G})\ d\mathbb{P}$. Therefore, $\mathbb{E}(XY|\mathscr{G}) = X \cdot \mathbb{E}(Y|\mathscr{G})$.

□

11. *Tower Property*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$ (i.e., sets in $\mathscr{H}$ are also in $\mathscr{G}$) and X is an integrable random variable, show that

$$\mathbb{E}[\mathbb{E}\,(X|\mathscr{G})|\mathscr{H}] = \mathbb{E}(X|\mathscr{H}).$$

Solution: For an integrable random variable Y, by definition we know that $\mathbb{E}(Y|\mathscr{H})$ is $\mathscr{H}$ measurable, and hence by setting $Y = \mathbb{E}\,(X|\mathscr{G})$, and for $A \in \mathscr{H}$, the partial averaging property of $\mathbb{E}[\mathbb{E}\,(X|\mathscr{G})|\mathscr{H}]$ is

$$\int_A \mathbb{E}[\mathbb{E}\,(X|\mathscr{G})|\mathscr{H}]\ d\mathbb{P} = \int_A \mathbb{E}\,(X|\mathscr{G})\ d\mathbb{P}.$$

Since $A \in \mathscr{H}$ and $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$, $A \in \mathscr{G}$. Therefore,

$$\int_A \mathbb{E}(X|\mathscr{H})\ d\mathbb{P} = \int_A X\ d\mathbb{P} = \int_A \mathbb{E}(X|\mathscr{G})\ d\mathbb{P}.$$

This shows that $\mathbb{E}(X|\mathscr{H})$ satisfies the partial averaging property of $\mathbb{E}[\mathbb{E}\,(X|\mathscr{G})|\mathscr{H}]$, and hence $\mathbb{E}[\mathbb{E}\,(X|\mathscr{G})|\mathscr{H}] = \mathbb{E}(X|\mathscr{H})$.

□

12. *Measurability*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If the random variable X is $\mathscr{G}$ measurable then show that

$$\mathbb{E}(X|\mathscr{G}) = X.$$

Solution: From the partial averaging property, for $A \in \Omega$,

$$\int_A \mathbb{E}(X|\mathscr{G})\, d\mathbb{P} = \int_A X\, d\mathbb{P}$$

and if X is $\mathscr{G}$ measurable then it satisfies

$$\mathbb{E}(X|\mathscr{G}) = X.$$

□

13. *Independence*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If $X = \mathbb{I}_B$ such that

$$\mathbb{I}_B(\omega) = \begin{cases} 1 & \text{if } \omega \in B \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbb{I}_B$ is independent of $\mathscr{G}$ show that

$$\mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X).$$

Solution: Since $\mathbb{E}(X)$ is non-random then $\mathbb{E}(X)$ is $\mathscr{G}$ measurable. Therefore, we now need to check that the following partial averaging property:

$$\int_A \mathbb{E}(X)\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \int_A \mathbb{E}(X|\mathscr{G})\, d\mathbb{P}$$

is satisfied for $A \in \mathscr{G}$.
Let $X = \mathbb{I}_B$ such that

$$\mathbb{I}_B(\omega) = \begin{cases} 1 & \text{if } \omega \in B \\ 0 & \text{otherwise} \end{cases}$$

and the random variable $\mathbb{I}_B$ is independent of $\mathscr{G}$. In addition, we also define

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbb{I}_A$ is $\mathscr{G}$ measurable. For all $A \in \mathscr{G}$ we have

$$\int_A X\, d\mathbb{P} = \int_A \mathbb{I}_B\, d\mathbb{P} = \int_A \mathbb{P}(B)\, d\mathbb{P} = \mathbb{P}(A)\mathbb{P}(B).$$

Furthermore, since the sets A and B are independent we can also write

$$\int_A X\, d\mathbb{P} = \int_A \mathbb{I}_B\, d\mathbb{P} = \int_\Omega \mathbb{I}_A \mathbb{I}_B\, d\mathbb{P} = \int_\Omega \mathbb{I}_{A\cap B}\, d\mathbb{P} = \mathbb{P}(A \cap B)$$

where

$$\mathbb{I}_{A\cap B}(\omega) = \begin{cases} 1 & \text{if } \omega \in A \cap B \\ 0 & \text{otherwise} \end{cases}$$

and hence

$$\int_A X \ d\mathbb{P} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{E}(X) = \int_A \mathbb{E}(X) \ d\mathbb{P}.$$

Thus, we have $\mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X)$.

□

14. *Conditional Jensen's Inequality*. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If $\varphi : \mathbb{R} \mapsto \mathbb{R}$ is a convex function and X is an integrable random variable show that

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq \varphi[\mathbb{E}(X|\mathscr{G})].$$

Deduce that if X is independent of $\mathscr{G}$ then the above inequality becomes

$$\mathbb{E}[\varphi(X)] \geq \varphi[\mathbb{E}(X)].$$

Solution: Given that φ is a convex function,

$$\varphi(x) \geq \varphi(y) + \varphi'(y)(y - x).$$

By setting $x = X$ and $y = \mathbb{E}(X|\mathscr{G})$ we have

$$\varphi(X) \geq \varphi[\mathbb{E}(X|\mathscr{G})] + \varphi'[\mathbb{E}(X|\mathscr{G})][\mathbb{E}(X|\mathscr{G}) - X]$$

and taking conditional expectations,

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq \varphi[\mathbb{E}(X|\mathscr{G})].$$

If X is independent of $\mathscr{G}$ then from Problem 1.2.3.13 (page 47) we can set $y = \mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X)$. Using the same steps as described above we have

$$\varphi(X) \geq \varphi[\mathbb{E}(X)] + \varphi'[\mathbb{E}(X)][\mathbb{E}(X) - X]$$

and taking expectations we finally have

$$\mathbb{E}[\varphi(X)] \geq \varphi[\mathbb{E}(X)].$$

□

15. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$ (i.e., sets in $\mathscr{G}$ are also in $\mathscr{F}$). If X is an integrable random variable and $\mathbb{E}(X^2) < \infty$ show that

$$\mathbb{E}\left[\mathbb{E}(X|\mathscr{G})^2\right] \leq \mathbb{E}\left(X^2\right).$$

Solution: From the conditional Jensen's inequality (see Problem 1.2.3.14, page 48) we set $\varphi(x) = x^2$ which is a convex function. By substituting $x = \mathbb{E}(X|\mathscr{G})$ we have

$$\mathbb{E}(X|\mathscr{G})^2 \leq \mathbb{E}\left(X^2|\mathscr{G}\right).$$

Taking expectations

$$\mathbb{E}\left[\mathbb{E}(X|\mathscr{G})^2\right] \leq \mathbb{E}\left[\mathbb{E}\left(X^2|\mathscr{G}\right)\right]$$

and from the tower property (see Problem 1.2.3.11, page 46)

$$\mathbb{E}\left[\mathbb{E}\left(X^2|\mathscr{G}\right)\right] = \mathbb{E}\left(X^2\right).$$

Thus, $\mathbb{E}\left[\mathbb{E}(X|\mathscr{G})^2\right] \leq \mathbb{E}\left(X^2\right)$.

□