1.1. Introduction

Example 1.1.1. The production manager of a bulb manufacturing company wishes to study the effect of a new manufacturing process on the lifetimes of bulbs produced through it.

Here the population under study is

\[ P : \text{Collection of lifetimes of all electric bulbs produced using new manufacturing process} \]

In most practical situations, \( P \) is generally large (e.g. Collection of lifetimes of all electric bulbs that would be produced using new manufacturing process) and it is not possible to get complete information about \( P \). Thus a representative sample (a sample that is a certain amount in true representative of the population) is taken from \( P \) and using this representative sample inferences regarding various population characteristics of \( P \) (such as population mean, population variance etc.) are made. Note that the samples contain only partial information about \( P \) and the goal is to make inferences about various population characteristics based on partial information in the sample drawn from \( P \).

\[ X : \text{lifetime of a typical electric bulb manufactured using new manufacturing process (a typical element of} \ P \)\]

\( X \) is random (called random variable) and its value varies across \( P \) according to some law

Probability Theory: A mathematical tool for modelling uncertainty (e.g. to describe the law according to which values of \( X \) vary across \( P \))
Statistics

Concern with procedures for analyzing data (sample) and drawing inferences about various characteristics of the population.

For understanding of Statistics, one must have a sound background in probability theory.

The only way to elicit information about any random phenomenon is to perform experiments (e.g., selecting a set of bulbs manufactured by the new manufacturing process and putting them on test for measuring their lifetimes) and putting them on test for measuring their lifetimes. Each experiment terminates in an outcome which cannot be predicted in advance, prior to the performance of the experiment (e.g., lifetime of the bulbs but on test can not be predicted before they are put on test).

**Definition 1.1.1 (Random Experiment).** A random experiment is an experiment in which

(a) all possible outcomes of the experiment are known in advance;

(b) outcome of a particular performance (trial) of the experiment can not be predicted in advance;

(c) the experiment can be repeated under identical conditions.

We will generally denote a random experiment by $E$.

**Definition 1.1.2 (Sample Space).** The collection of all possible outcomes of a random experiment is called its sample space. A sample space will usually be denoted by $S$.

**Example 1.1.2**

(a) $E$: Tossing a coin once

Sample Space $S = \{H, T\}$,

Where

$H$: Heads

$T$: Tails

[Page 2]
(b) E: Rolling one red die and one white die

Sample Space:
\[ S = \{(x, w) :\text{number of } x\text{ on the red die and } w\text{ on the number of } y\text{ on the white die}\} \]
\[ = \{(1, 1), (1, 2), \ldots, (1, 6), (2, 1), \ldots, (2, 6), \ldots, (6, 1), \ldots, (6, 6)\} \]
\[ = \{(x, y) : 1 \leq x \leq 6 \} \times \{(1, 2, \ldots, 6)\} \]
\[ \rightarrow \text{has 36 elements} \]

(c) E: Putting two electric bulbs produced by the new manufacturing process into test and measuring their lifetimes

Sample Space:
\[ S = \{(x, y) : x \geq 0, y \geq 0\} \]
\[ = \{(x, y) : x, y \in \mathbb{N}\} \]

**Definition 1.1.3 (Event)**

If the outcome of a random experiment is a member of the set \( E \subseteq S \), we say that event \( E \) has occurred.

Generally, we are interested in specific subsets of \( S \), called events. Thus, event \( S \) is the sample space (events under consideration) \( E \) is a subset of the power set of \( S \). In many situations, \( E = \mathcal{P}(S) \), the power set of \( S \). For all practical purposes \( S = \mathcal{P}(\mathbb{N}) \).

**Example 1.1.3.** In Example 1.1.2 (b)

\[ A = \{(1, 5), (6, 2), (2, 2)\} \]

is an event.

In Example 1.1.2 (c)

\[ A = \{(x, y) : x \geq 0, y \geq 0\} \]
\[ \rightarrow \{(0, 6) \times (0, 0)\} \]

may be an event.
The algebra of set theory is applicable in probability theory. Probability is a measure of uncertainty. We are interested in quantifying uncertainties associated with various outcomes of a random experiment by assigning probabilities to these outcomes.

Here we will not discuss how probabilities are assigned (which is a part of probability modelling). Rather we will discuss properties of probability as a measure.

1.2. Probability Function (or Probability Measure)

\[ \mathbb{E}: \text{a random experiment} \]
\[ \mathbb{S}: \text{Sample Space } \mathbb{E} \]
\[ \mathbb{F}: \text{Event Space} \]

For all practical purposes one may take \( \mathbb{F} = \mathbb{P}(\mathbb{S}) \).

A net function in a function whose domain is a collection of sets (called domain).

**Definition 1.2.1. (Probability Function or Probability Measure)**

A probability function (or probability measure) is a real-valued net function defined on event space \( \mathbb{F} \), satisfying the following axioms:

(a) \( \mathbb{P}(\mathbb{S}) = 1 \);
(b) \( \mathbb{P}(\mathbb{E}) \geq 0 \), \( \forall \mathbb{E} \in \mathbb{F} \);
(c) If \( \mathbb{E}_1, \mathbb{E}_2 \in \mathbb{F} \) are mutually exclusive/disjoint (i.e., \( \mathbb{E}_1 \cap \mathbb{E}_2 = \emptyset \); the empty set), then

\[ \mathbb{P}(\mathbb{E}_1 \cup \mathbb{E}_2) = \mathbb{P}(\mathbb{E}_1) + \mathbb{P}(\mathbb{E}_2). \]

(See property 3.2 in the previous).

Now, generally \( \bigoplus \{ \mathbb{E}_n \} \) is a sequence of mutually exclusive/disjoint sets in \( \mathbb{F} \) (\( \mathbb{E}_i \cap \mathbb{E}_j = \emptyset \) \( \forall i \neq j \)), then

\[ \mathbb{P}\left( \bigcup_{i=1}^{\infty} \mathbb{E}_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(\mathbb{E}_i). \] (Countable additivity)

The triplet \((\mathbb{S}, \mathbb{F}, \mathbb{P})\) is called probability triple.
**Remark 1.2.1.** Axioms (b) and (c1) are derivable for any measure (such as area, volume, probability, etc.). Since the sample space $\Omega$ consists of all possible outcomes in occurrence in certain (100%) chance of occurrence and therefore Axiom (a) ($P(\Omega) = 1$) is also reasonable.

---

**Elementary Properties of Probability Function / Measure**

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

**P.1**

$P(\emptyset) = 0$.

**Proof.** Let $E_1 = \Omega$ and $E_i = \emptyset, i = 2, \ldots$. Then $E_i \cap E_j = \emptyset \forall i, j$, and $\Omega = \bigcup E_i$. Therefore

\[
P(\Omega) = P\left(\bigcup_{i=1}^{n} E_i\right)
\]

\[
= \sum_{i=1}^{n} P(E_i) \tag{Axiom (a) and (c1)}
\]

\[
= \lim_{n \to \infty} \sum_{i=1}^{n} P(E_i)
\]

\[
= \lim_{n \to \infty} \left( P(\Omega) + (n-1) P(\emptyset) \right)
\]

\[
= 1 + \lim_{n \to \infty} (n-1) P(\emptyset)
\]

\[
= 1 + \lim_{n \to \infty} (n-1) \cdot 0
\]

\[
= 1 + 0 = 1
\]

\[
P(\emptyset) = 0
\]

**P.2.** For some natural number $n$, let $E_1, E_2, \ldots, E_n \in \mathcal{F}$ be mutually exclusive. Then

\[
P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).
\]

**Proof.** Let $E_1 \cap E_2 = \cdots = \emptyset$. Then $E_i \cap E_j = \emptyset \forall i, j$ and

\[
\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} E_i
\]

\[
P\left(\bigcup_{i=1}^{n} E_i\right) = P\left(\bigcup_{i=1}^{n} E_i\right)
\]

\[
= \sum_{i=1}^{n} P(E_i) \tag{Axiom (e1)}
\]

\[
= \sum_{i=1}^{n} P(E_i) \tag{P(E_i) = 1(\emptyset) = 0 \forall i \in \mathcal{F}}
\]
0 ≤ P(E) ≤ 1 \\quad \forall E \in \mathcal{F}
\quad \text{and}
\quad P(E^c) = 1 - P(E) \quad \forall E \in \mathcal{F}.

\text{S1: } E \cup E^c = \mathcal{F} \quad \text{and} \quad E \cap E^c = \emptyset. \quad \text{Therefore}
\quad P(E) = P(E \cup E^c)
\quad \Rightarrow 1 = P(E) + P(E^c) \geq P(E) \quad \text{(using Axioms (a) and (2))}
\quad \Rightarrow 0 < P(E) ≤ 1 \quad \text{and}
\quad \Rightarrow 1 - P(E) = P(E^c) = 1 - P(E).

P.4
Let \( E_1, E_2 \in \mathcal{F} \) be such that \( E_1 \subseteq E_2 \). Then
\[ P(E_2 - E_1) = P(E_2) - P(E_1). \]

P.5
Let \( E_1, E_2 \in \mathcal{F} \) be such that \( E_1 \subseteq E_2 \). Then \( P(E_1) \leq P(E_2) \) (i.e., P.1 is monotone).

\[ P(E_2 - E_1) = P(E_2) - P(E_1) \]

By Axiom (a) we have \( P(E_2 - E_1) > 0 \), which implies that
\[ P(E_2) - P(E_1) > 0 \]
\[ \Rightarrow P(E_1) \leq P(E_2). \]
Since we have \( \emptyset \in E \subseteq \mathcal{G} \), using \( P.5 \) we get
\[
P(E) \leq P(E) \leq P(D) \quad \forall E \in \mathcal{G}.
\]

Let \( E_1, E_2 \in \mathcal{G} \). Then
\[
P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad \text{(Inclusion-Exclusion Principle for two events)}
\]

\( E_1 \cup E_2 = E_1 \cup (E_2 - E_1) \) and \( E_1 \cap (E_2 - E_1) = \emptyset \)

\[
\Rightarrow P(E_1 \cup E_2) = P(E_1 \cup (E_2 - E_1)) = P(E_1) + P(E_2 - E_1) \quad \text{(using P.1)}
\]

Also
\[
E_2 = (E_1 \cap E_2) + (E_2 - E_1) \quad \text{and} \quad (E_1 \cap E_2) \cap (E_2 - E_1) = \emptyset.
\]

Therefore
\[
P(E_2) = P(E_1 \cap E_2) + (E_2 - E_1)
\]

\[
= P(E_1 \cap E_2) + P(E_2 - E_1)
\]

\[
\Rightarrow P(E_2 - E_1) = P(E_2) - P(E_1 \cap E_2)
\]

Using this in equation (1.2.1) we get
\[
P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)
\]
Remark 1.2.1
(a) If \( P(A) = 0 \) and \( B \subseteq A \), then \( P(B) = 0 \) (using P.5 and Axiom (B)).

Similarly if \( P(A) = 1 \) and \( C \subseteq D \), then \( P(D) = 1 \) (using P.5 and P.6).

(b) If \( P(D) = 1 \), then
\[
P(A) = P(A \cap D) \quad \text{for all} \quad A \in \mathcal{G}.
\]

Similarly if \( P(D) = 0 \), then
\[
P(A) = P(A \cap D^c) \quad \text{for all} \quad A \in \mathcal{G}.
\]

(c) Let \( E_1, E_2 \in \mathcal{G} \). Then, using P.7 and Axiom (b)
\[
P(E_1 \cup E_2) \leq P(E_1) + P(E_2) \quad \text{(Double inequality for two events)}
\]

(Boole's inequality for two events)

(d) Let \( E_1, E_2 \in \mathcal{G} \). Then using P.7 and P.6 we have
\[
P(E_1 \cap E_2) > \max \left\{ 0, P(E_1) + P(E_2) - 1 \right\}
\]

(Bayes's inequality for two events)

Theorem 1.2.1. (Inclusion-Exclusion Principle)

For \( E_1, E_2, \ldots, E_k \in \mathcal{G} \),

\[
P_{\infty} = P(E_1 \cup \cdots \cup E_k) = \sum_{i=1}^{k} P(E_i) - \sum_{1 \leq i < j \leq k} P(E_i \cap E_j) + \cdots + (-1)^{k-1} P(E_1 \cap \cdots \cap E_k).
\]

(Num of probabilities of all possible intersections involving \( k \) events out of \( k \) events)

Then
\[
P(\bigcup_{i=1}^{k} E_i) = P_{\infty} - P_{2} + P_{3} - P_{4} + \cdots + (-1)^{k-1} P_{k}.
\]
Note that, for \( k = 2 \),

\[
\begin{align*}
\phi_1 &= P(E_1) + P(E_2) \\
\phi_2 &= P(E_1 \cup E_2)
\end{align*}
\]

and

\[
\begin{align*}
P(E_1 \cup E_2) &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\
&= \phi_1 - \phi_2.
\end{align*}
\]

Thus the result is true for \( k = 2 \).

Now suppose that the result is true for \( k = 3, 4, \ldots, n \), i.e.,

\[
P(U_k) = \sum_{j=1}^{k-1} \phi_j + \cdots + (-1)^{k-1} \phi_k, \quad k \geq 2, 3, \ldots, n
\]

Then

\[
P(U_k) = P(U_k) + P(E_{k+1}) - P(U_k \cap E_{k+1})
\]

(using result for \( k = 2 \))

\[
= \sum_{j=1}^{k-1} \phi_j + P(E_{k+1}) - P(U_k \cap E_{k+1})
\]

(using the result for \( k = 3, 4, \ldots, n \))

\[
= \sum_{j=1}^{k-1} \phi_j + P(E_{k+1}) - \sum_{j=1}^{k-1} (-1)^{j-1} t_{j+1}
\]

(using the result for \( k = 3, 4, \ldots, n \))

Where

\[
\begin{align*}
t_{j+1} &= \sum_{i=1}^{j} P(E_i \cap E_{j+1}) \\
t_3 &= \sum_{i=1}^{2} \sum_{j=1}^{3} P(E_i \cap E_j \cap E_{j+1}) \\
&\vdots \\
t_k &= \sum_{i=1}^{k-1} \cdots \sum_{j=1}^{k} P(E_i \cap E_j \cdots \cap E_{j+k}) \\
t_{k+1} &= P(E_k \cap E_{k+1} \cdots \cap E_n \cap E_{k+1})
\end{align*}
\]
Therefore

\[ p(E_1 E_i) = (p_{21} + p_1 E_{21}) - (p_{31} + p_1 E_{31}) + (p_{32} + p_2 E_{32}) \]

\[ \vdots \]

\[ + (-1)^{n-1} (p_{2n} + p_1 E_{2n}) + (-1)^n p_{n1} E_{n1}, \]

\[ = p_{11} E_{11} - p_{21} E_{21} + p_{31} E_{31} - \ldots - (-1)^{n-1} E_{2n} + (-1)^n E_{n1}, \]

and

\[ p_{11} + p_1 E_{11} = \sum_{i=1}^{\infty} p_i(E_j) + p_i(E_{11}) = p_{11} E_{11}, \]

\[ p_{21} + p_1 E_{21} = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} p_i(E_j E_i) = p_{21} E_{21}, \]

\[ p_{31} + p_1 E_{31} = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=i}^{\infty} p_i E_{j E_k E_i} = p_{31} E_{31}, \]

\[ \vdots \]

\[ p_{n1} + p_1 E_{n1} = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=j}^{\infty} \ldots \sum_{l=n}^{\infty} p_i E_{j E_k E_l \ldots E_i} = p_{n1} E_{n1}. \]

The result now follows by induction.

Remark 1.2.3.

Let \( E_1, E_2, E_3 \in \mathcal{E}. \) Then

\[ p(E_1 E_2 \cup E_3) = p_{13} - p_{23} + p_{33} \]

\[ = (p_i E_1 + p_i E_2 + p_i E_3) \]

\[ - (p_i E_1 E_2 + p_i E_1 E_3 + p_i E_2 E_3) \]

\[ + p_i E_1 E_2 E_3. \]
Theorem 1.2.2. For some positive integer \( k \geq 2 \), let \( E_1, E_2, \ldots, E_n \in \mathcal{G} \).

Then
\[
p_{2k} - p_{2k} \leq p\left( \bigcup_{i=1}^{k} E_i \right) \leq p_{2k}
\]

Proof. Note that for \( k = 2 \)
\[
p_{2} = p(E_1 + p(E_2))
\]
\[
p_{2} = p(E_1 E_2)
\]
\[
p\left( \bigcup_{i=1}^{2} E_i \right) = p(E_1 + p(E_2) - p(E_1 E_2) \leq p(E_1) + p(E_2)
\]

Thus the result is true for \( k = 2 \).

Now suppose that for some positive integer \( m > 2 \)
\[
p_{2} \leq p\left( \bigcup_{i=1}^{k} E_i \right) \leq p_{2k}, \quad k = 2, 3, \ldots, m
\]

Then
\[
p\left( \bigcup_{i=1}^{m+1} E_i \right) = p\left( \left( \bigcup_{i=1}^{m} E_i \right) + E_{m+1} \right)
\]
\[
\leq p\left( \bigcup_{i=1}^{m} E_i \right) + p(E_{m+1})
\]
\[
\leq p_{2m} + p(E_{m+1})
\]
\[
\geq p_{2m+1}
\]

... \[1\] 2.2

Also using the result for \( k = m \) we get
\[
p\left( \bigcup_{i=1}^{m} E_i \right) \geq p_{2m} - p_{2m+1}
\]

and
\[
p\left( \bigcup_{i=1}^{m} E_i \right) \leq \sum_{i=1}^{m} p(E_i + E_{m+1})
\]

Thus
\[
p\left( \bigcup_{i=1}^{m+1} E_i \right) = p\left( \left( \bigcup_{i=1}^{m} E_i \right) + E_{m+1} \right)
\]
\[
= p\left( \bigcup_{i=1}^{m} E_i \right) + p(E_{m+1}) - p\left( \left( \bigcup_{i=1}^{m} E_i \right) E_{m+1} \right)
\]
\[
\geq p_{2m} - p_{2m+1} + p(E_{m+1}) - \sum_{i=1}^{m} p(E_i + E_{m+1})
\]

\[\Box\]
\[= (\beta_{\text{min}} + \sum_{i=1}^{n} \beta_i) - \sum_{i=1}^{n} \beta_i \]

Using (1.2.2) and (1.2.3) we get

\[\beta_{\text{min}} - \beta_{\text{max}} \leq \sum_{i=1}^{n} \beta_i \leq \beta_{\text{max}} \]

and the result follows using principle of mathematical induction.

**Remark 1.2.4**

One can also show that

\[\beta_{\text{min}} - \beta_{\text{max}} \leq \sum_{i=1}^{n} \beta_i \leq \beta_{\text{max}} - \sum_{i=1}^{n} \beta_i \]

\[\sum_{i=1}^{n} \beta_i \leq \sum_{i=1}^{n} \beta_i \leq \sum_{i=1}^{n} \beta_i \]

\[\sum_{i=1}^{n} \beta_i \leq \sum_{i=1}^{n} \beta_i \]

**Theorem 1.2.3.** *Bonferroni's Inequality*

Let \( E_1, E_2, \ldots, E_k \).

Then

\[ P(\bigcap_{i=1}^{k} E_i) \geq \prod_{i=1}^{k} P(E_i) - (k-1) P(E) \]

**Proof.**

We have

\[ P(\bigcap_{i=1}^{k} E_i) = P(\bigcup_{i=1}^{k} E_i^c) \]

\[ \geq 1 - P(\bigcup_{i=1}^{k} E_i) \]

\[ \geq 1 - \sum_{i=1}^{k} P(E_i^c) \]

\[ \geq 1 - \sum_{i=1}^{k} [1 - P(E_i)] \]

\[ \geq \sum_{i=1}^{k} P(E_i) - (k-1) \]

Also

\[ P(\bigcap_{i=1}^{k} E_i) > 0 \]

Combining (1.2.4) and (1.2.5) we get

\[ P(\bigcap_{i=1}^{k} E_i) \geq \max \left( \sum_{i=1}^{k} P(E_i) - (k-1) P(E) \right) \text{ for } \]

12/1
Example 12.1  Random experiment: Casting a red and a white die

Sample space: \( \Omega = \{(i,j): \ i \in \{1,2,3,4,5,6\}, \ j \in \{1,2,3,4,5,6\}\} \)

For \((i, j) \in \Omega \)

- \(i\): number of spots up on the red die.
- \(j\): number of spots up on the white die.

Event space \( \mathcal{F} \) is the set of all \( \mathcal{F} \).

For \( E \in \mathcal{F} \), define \( Q: \mathcal{F} \rightarrow \mathbb{R} \) as

\[
Q(E) = \frac{|E|}{36}, \quad \text{where} \quad |E| = \text{# of elements in } E.
\]

Then

(a) \( Q(\Omega) = \frac{|\Omega|}{36} = \frac{36}{36} = 1 \)

(b) \( Q(E) = \frac{|E|}{36} \geq 0 \quad \forall \ E \in \mathcal{F} \)

(c) For mutually exclusive events \( E_1, E_2, \ldots \),

\[
Q\left(\bigcup_{i=1}^{g} E_i \right) = \frac{|\bigcup_{i=1}^{g} E_i|}{36} = \frac{\sum_{i=1}^{g} |E_i|}{36} \quad \text{(since union of disjoint sets)}
\]

\[
= \frac{g}{36} \sum_{i=1}^{g} |E_i| = \sum_{i=1}^{g} Q(E_i)
\]

Thus \((\Omega, Q, \mathcal{F})\) is a probability space.
Suppose that the sample space
\[ \mathcal{S} = \{ \omega_1, \omega_2, \ldots, \omega_k \} \]
is finite (has \( k \) elements). Here \( \omega_i \) are called elementary events and
\[ \mathcal{S} = \bigcup_{i=1}^{k} \{ \omega_i \}. \]

Suppose that
\[ P(\{\omega_i\}) = \frac{1}{k}, \quad i = 1, \ldots, k \]
(each elementary event is equally likely).

For any event \( E \subseteq \mathcal{S} \), we have
\[ E = \bigcup_{i \in I} \{ \omega_i \} \]
for some \( I \subseteq \{1, \ldots, k\} \). Then
\[ P(E) = \sum_{i \in I} P(\{\omega_i\}) \]
\[ = \frac{|I|}{k} \]
\[ = \frac{\text{# of ways favorable to event } E}{\text{total number of ways in which the random experiment can terminate}} \]
\[ \text{here the assumption of equally likely} \quad (P(\{\omega_i\}) = \frac{1}{k}, i = 1, \ldots, k) \]
in a part of probability modelling.
In a random experiment with finite sample space, whenever we say that the experiment has been performed at random, it means that all the outcomes in the sample space are equally likely.

**Example 1.2.2. (Birthday Problem)** Suppose that a college has \( n \) students, including you. Each of them were born on non-leap years.

(a) Find the probability that at least two of them have the same birthday. For what values of \( n \) is this probability more than 0.5, 0.8, 0.95? You will find someone with birthday on \( \frac{1}{2} \).

(b) For what value of \( n \) the probability that someone who has your birthday is \( \frac{1}{2} \).

**Solution**

(a) Required probability = \( 1 - P(\text{all of them have different birthdays}) \)

\[
= 1 - \frac{365 \times 364 \times \cdots \times (365-1+1)}{365^n}
\]

(b) Required probability = \( 1 - P(\text{no one shares the same birthday as mine}) \)

\[
= 1 - \frac{364^n}{365^n}
\]

For \( \frac{1 - 364^n}{365^n} \approx 0.5 \)

\( n \approx 253 \)
Example 12.3. Five cards are drawn at random and without replacement from a deck of 52 cards. Find the probability that:

(i) Each card is a heart (event $E_1$);

(ii) At least one card is a heart (event $E_2$);

(iii) Exactly two cards are King and two cards are Queen (event $E_3$);

(iv) Exactly two Kings, two Queens and one Jack are drawn.

Solution

(i) $P(E_1) = \frac{\binom{13}{3}}{\binom{52}{5}}$

(ii) $P(E_2) = 1 - P(E_2^c)$
    
    $= 1 - P(\text{no card is a heart})$
    
    $= 1 - \frac{\binom{39}{5}}{\binom{52}{5}}$

(iii) $P(E_3) = \frac{\binom{4}{2}\binom{4}{2}}{\binom{52}{5}}$

(iv) $P(E_4) = \frac{\binom{4}{2}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}$
Example 1.2.4. (Capture/Recapture Method for Estimating Population Size)

In a wildlife population suppose that the population size \( n \) is unknown. To estimate the population size \( n \), 20 animals are captured, tagged, and then released back. Thereafter 40 animals are captured at random and it is found that 8 of them are tagged. Find an estimate of the population size \( n \) based on the given data.

**Solution**

We have:

\[
\begin{align*}
\text{total \# of animals} &= n \\
\text{\# of tagged animals in the population} &= 20 \\
\text{\# of untagged animals in the population} &= n - 20
\end{align*}
\]

Sample of 40 animals yielded:

\[
\begin{align*}
\text{\# of tagged animals} &= 8 \\
\text{\# of untagged animals} &= 32
\end{align*}
\]

The probability of obtaining this data is:

\[
\ell(n) = \frac{(20\choose 8) (n-20\choose 32)}{(40\choose 40)}
\]

\[
= \frac{(20\choose 8) (n-20\choose 32)}{(40\choose 40)}
\]

\[
\ell(n+1) > \ell(n) \implies \frac{(n+1-19\choose 32)}{(n+1\choose 40)} > \frac{(n-20\choose 32)}{(n\choose 40)}
\]

\[
\implies \frac{h-19}{32} > \frac{h}{40} \implies \frac{(h-19)}{(h-51)} > \frac{h-20}{32} \implies \frac{h-19}{(h-51)(h+1)} > 1
\]

\[
h < 99 (h-51)(h+1)
\]
Similarly,
\[ \lim_{n \to \infty} l(n) \geq n \]

This is maximized at \( n = 99 \), i.e., for \( n = 99 \) the observed data (among 96 captured animals 8 are tagged and 91 are untagged) is most probable.

Thus an estimate of \( n \) is
\[ \hat{n} = 99 \quad \text{(Maximum likelihood estimator)} \]

1.3. Conditional Probability

Consider a probability space \( (\Omega, \mathcal{F}, P) \), where \( \Omega = \{\omega_1, \ldots, \omega_n\} \) is finite and
\[ P(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, \ldots, n \quad \text{(Equally likely probability model)} \]

Then, for any \( C \subseteq \Omega \)
\[ P(C) = \frac{\# \text{ of cases favorable to } C}{\text{total # of cases}} \]
\[ = \frac{|C|}{|\Omega|} = \frac{|C|}{n} \]

Now suppose it is known a priori that event \( A \) has occurred (i.e., outcome of the experiment is an element of \( A \)), where \( |A| \geq 1 \). (No that \( P(A) > \frac{|A|}{n} \).

Given this prior information (that the event \( A \) has occurred) we want to define probability, a function on event \( \Omega \). A natural way to define \( P(B|A) \) is
\[ P(B|A) = \frac{|\text{AND}|}{|A|} = \frac{|\text{AND}|}{|A|/n} = \frac{P(\text{AND})}{P(A)}, \text{ for } B \subseteq \Omega \]

\[ 18/1 \]
**Definition 1.3.1**

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A} \in \mathcal{F}$ be non-empty that $P(\mathcal{A}) > 0$. Then

$$P(B|A) = \frac{P(A \cap B)}{P(\mathcal{A})}, \quad B \in \mathcal{F}.$$  

This is called the **conditional probability of event B given the event A**.

**Remark 1.3.1.**

(a) In the above definition the event $A$ (with $P(\mathcal{A}) \neq 0$) is fixed and for two fixed $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, $P(\cdot \mid \mathcal{A})$ is a set function defined on $\mathcal{F}$.

(b) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$, $A \in \mathcal{F}$.

**Theorem 1.3.1.**

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A} \in \mathcal{F}$ with $P(\mathcal{A}) > 0$ be fixed. Then $P(\cdot \mid \mathcal{A}) : \mathcal{F} \to \mathbb{R}$ is a probability function (called the conditional probability function) on $\mathcal{F}$ (where $(\Omega, P(\cdot \mid \mathcal{A}))$ is a prob. space).

**Proof.**

Clearly

$$P(B|A) = \frac{P(A \cap B)}{P(\mathcal{A})} > 0, \quad A \in \mathcal{F}.$$  

Also

$$P(B|A) = \frac{P(A \cap B)}{P(\mathcal{A})} \geq 1.$$  

Let $\{B_n\}_{n \geq 1}$ be a sequence of disjoint events in $\mathcal{F}$. Then

$$P\left(\bigcup_{n \geq 1} B_n \mid \mathcal{A}\right) = \frac{P\left((\bigcup_{n \geq 1} B_n) \cap \mathcal{A}\right)}{P(\mathcal{A})} = \frac{P\left(\bigcup_{n \geq 1} (B_n \cap \mathcal{A})\right)}{P(\mathcal{A})} = \sum_{n \geq 1} P(B_n \cap \mathcal{A}) \geq \sum_{n \geq 1} \frac{P(B_n \cap \mathcal{A})}{P(\mathcal{A})}.$$  

Since $\{B_n\}_{n \geq 1}$ are disjoint, then the union of $\{B_n \cap \mathcal{A}\}_{n \geq 1}$ are also disjoint. Since $P(\cdot)$ is a prob. measure, we get

$$P\left(\bigcup_{n \geq 1} B_n \mid \mathcal{A}\right) \geq \sum_{n \geq 1} \frac{P(B_n \cap \mathcal{A})}{P(\mathcal{A})}.$$  

19/1
$$= \sum_{k} P(D_k|A)$$

It follows that \( P(A) \) is a prob. function on \( \mathcal{F} \) for any fixed \( A \in \mathcal{F} \) with \( P(A) \neq 0 \).

**Example 1.3.1.** Five cards are dealt at random (without replacement) from a deck of 52 cards. Define events:

- \( B \): all \( \spadesuit \) cards in hand
- \( A \): at least 4 \( \spadesuit \) cards in hand

Find \( P(B|A) \).

**Solution** We have

\[
P(B|A) = \frac{P(A \land B)}{P(A)}
\]

\[
= \frac{P(D)}{P(A)} \quad \text{(since \( B \subseteq A \))}
\]

\[
= \frac{(\begin{pmatrix} 13 \end{pmatrix}/\begin{pmatrix} 52 \end{pmatrix})}{\begin{pmatrix} 13 \end{pmatrix}/\begin{pmatrix} 52 \end{pmatrix}} = 0.424.
\]

**Remark 1.3.2.** *(Multiplication Law)*

(i) \( P(A \land B) = P(A) P(B|A) \), cb \( P(A) > 0 \)

(ii) \( P(A \land B \land C) = P(A \land B) P(C|A \land B) \),

\[
P(A \land B \land C) = P(A) P(B|A) P(C|A \land B)
\]

provided \( P(A \land B) > 0 \) (which ensures that \( P(A) \neq 0 \))

(iii) Using principle of mathematical induction, we have

\[
P(A_1 \land A_2 \land \ldots \land A_n) = P(A_1 \land A_2 \land \ldots \land A_{n-1}) P(A_n|A_1 \land A_2 \land \ldots \land A_{n-1})
\]

\[
= P(A_1 \land A_2 \land \ldots \land A_{n-2}) P(A_{n-1}|A_1 \land A_2 \land \ldots \land A_{n-2}) P(A_n|A_1 \land A_2 \land \ldots \land A_{n-2})
\]

\[
= \cdots
\]

\[
P(A_1 \land A_2 \land \ldots \land A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 \land A_2) \ldots P(A_n|A_1 \land A_2 \land \ldots \land A_{n-1}),
\]

provided \( P(A_1 \land A_2 \land \ldots \land A_{n-1}) > 0 \) (which also ensures that \( P(A_1 \land A_2 \land \ldots \land A_{n-1}) > 0 \)).
Due to symmetry, if \((x_1, \ldots, x_n)\) is a permutation of \((1, \ldots, n)\), then

\[
P(\bigwedge_{i=1}^{n} C_i) = P(C_{x_1} \land C_{x_2} \land \ldots \land C_{x_n})
\]

\[
= P(C_{x_1}) P(C_{x_2} | C_{x_1}) P(C_{x_3} | C_{x_2}, C_{x_1}) \ldots P(C_{x_n} | C_{x_{n-1}}, \ldots, C_{x_1})
\]

provided \(P(C_{x_1} \land C_{x_2} \land \ldots \land C_{x_n}) > 0\) (which also ensures that \(P(C_{x_1} \land C_{x_2} \land \ldots \land C_{x_n}) > 0\) for \(i = 1, \ldots, n-1\)).

**Example 1.3.2.** A bowl contains 3 red and 5 blue chips. All chips that are of the same colour are identical. Two chips are drawn successively at random and without replacement. Define events

- \(A\): First draw resulted in a red chip.
- \(B\): Second draw resulted in a blue chip.

**Solution**

- \(P(A) = \frac{3}{8}\)
- \(P(B | A) = \frac{5}{7}\)
- \(P(B) = \frac{5}{8} \times \frac{1}{1} = \frac{5}{8}\)

\(P(A \land B) = P(A) P(B | A) = \frac{3}{8} \times \frac{5}{7} = 0.2679\)

**Theorem 1.3.2.** For a countable set \(\mathcal{A}\) (that is elements of \(\mathcal{A}\) can either be put in 1-1 correspondence with \(\mathbb{N} = \{1, 2, \ldots\}\) or with \(\mathbb{R}^+\) for some real number), let \(\{E_x : x \in \mathcal{A}\}\) be a countable collection of mutually exclusive (i.e. \(E_x \cap E_y = \emptyset\) if \(x \neq y\)) and exhaustive (i.e. \(\bigcup_{x \in \mathcal{A}} E_x = \Omega\)) events. Then, for any \(E \in \Omega\),

\[
P(E) = \sum_{x \in \mathcal{A}} P(E | E_x) P(E_x)
\]
Proof. Since \( P(\bigvee_{\alpha \in \Delta} E_{\alpha}) = 1 \), we have
\[
P(\text{E}) = P(\bigwedge_{\alpha \in \Delta} (E_{\alpha} \cup U(\text{E}_{\alpha})))
\]
\[
\geq P(\bigcup_{\alpha \in \Delta} U(\text{E}_{\alpha}))
\]
\[
= \sum_{\alpha \in \Delta} P(U(\text{E}_{\alpha})) (\text{E}_{\alpha}'s \text{ are disjoint} \Rightarrow \text{their unions})
\]
\[
= \sum_{\alpha \in \Delta} P(E_{\alpha}) (\text{E}_{\alpha} \cap \text{E}_{\alpha}' = 0 \Rightarrow P(E_{\alpha} \cap \text{E}_{\alpha}')) = 0)
\]
\[
= \sum_{\alpha \in \Delta} P(E_{\alpha} | \text{E}_{\alpha}) P(\text{E}_{\alpha})
\]
\[
= \sum_{\alpha \in \Delta} P(E_{\alpha}) > 0
\]

Example 1.3.3. A population consists of 40% female and 60% male. Suppose that 15% of females and 20% of males are smokers. A person is selected at random from the population.

(a) Find the probability that he/she is a smoker.
(b) Given that the selected person is a smoker, find the probability that he is male.

Solution. Define the events:
- \( M \): Selected person is a male
- \( F = M^c \): Selected person is a female
- \( S \): Selected person is a smoker
- \( T = S^c \): Selected person is a non-smoker.

We have:
\[
P(F) = 0.4, \quad P(M) = 0.6, \quad P(F \cup M) = P(F) + P(M) = 1.0
\]
\[
P(S | F) = 0.15, \quad P(T | F) = 0.85
\]
\[
P(S | M) = 0.30, \quad P(T | M) = 0.70
\]
(a) By Theorem of total probability
\[ P(S) = P(S\mid F) + P(S\mid N) \]
\[ = P(S\mid F) P(F) + P(S\mid N) P(N) \]
\[ = 0.15 \times 0.4 + 0.30 \times 0.6 \]
\[ = 0.06 + 0.18 = 0.24 \]

(b) \[ P(N\mid S) = \frac{P(N\mid S)}{P(S)} = \frac{P(S\mid N) P(N)}{P(S)} = \frac{0.30 \times 0.60}{0.24} \]
\[ = \frac{0.18}{0.24} = \frac{3}{4} \]

**Theorem 1.3.3. (Bayes' Theorem)**

Let \( \mathcal{E} = \{E_1, E_2, \ldots\} \) be a countable collection of mutually exclusive and exhaustive events and let \( E \) be any event with \( P(E) > 0 \). Then, for any \( E \in \mathcal{E} \) with \( P(E) > 0 \),
\[
P(E_j \mid E) = \frac{P(E_j \cap E)}{P(E)} = \frac{P(E_j) P(E \mid E_j)}{\sum_{\alpha \in \mathcal{E}} P(E_\alpha) P(E \mid E_\alpha)}
\]

**Proof.**

For \( j \in \mathcal{E} \),
\[
P(E_j \mid E) = \frac{P(E_j \cap E)}{P(E)} = \frac{P(E_j) P(E \mid E_j)}{\sum_{\alpha \in \mathcal{E}} P(E_\alpha) P(E \mid E_\alpha)}
\]

**Remark 1.3.3.**

(a) Suppose that occurrence of any of the mutually exclusive and exhaustive events \( \mathcal{E} = \{E_1, E_2, \ldots\} \) (where \( \mathcal{E} \) is a countable set) may cause the occurrence of an event \( E \). Given that the event \( E \) has occurred (i.e., given the effect), Bayes' Theorem provides the conditional probability that the event \( E \) (effect) is caused by occurrence of event \( E_j \).
(b) In Bayes' Theorem,

\{P(E_i)\}; i \in A \text{ are called prior probabilities.}

and

\{P(E_i|A)\}; i \in A \text{ are called posterior probabilities.}

**Example 1.3.4.**

Bowl C_1 contains 3 red and 7 blue chips.

Bowl C_2 contains 9 red and 2 blue chips.

Bowl C_3 contains 5 red and 5 blue chips.

All chips of the same color are identical.

All chips of the same color are identical.

A die is cast and a bowl is selected at random for the following.

Let event:

- Bowl C_1 is selected if 5 or 6 appear.
- Bowl C_2 is selected if 2, 3, or 4 appear.
- Bowl C_3 is selected if 6 appear.

A die is cast and a bowl is selected at random for the following.

Let event:

- Bowl C_1 is selected if 5 or 6 appear.
- Bowl C_2 is selected if 2, 3, or 4 appear.
- Bowl C_3 is selected if 6 appear.

The selected bowl is handed over to another person who
draws two chips at random from that bowl. Find the
probability that:

two red chips are drawn.

(a) given that drawn chips are both red, find the
probability that it came from bowl C_3.

**Solution**

Define the events

- A_i: Selected bowl is C_i (i=1,2,3)
- R: The chips drawn from the selected bowl are both red.

Then

\[ P(A_1) = \frac{2}{6} = \frac{1}{3}; \quad P(A_2) = \frac{2}{6} = \frac{1}{3}; \quad P(A_3) = \frac{1}{6} \]

\[ \{A_1, A_2, A_3\} \text{ are mutually exclusive and exhaustive.} \]
(a) \[
P(R) = P(R | A_1) P(A_1) + P(R | A_2) P(A_2) + P(R | A_3) P(A_3)
\]
\[
= \frac{3}{10} \cdot \frac{1}{2} + \frac{8}{10} \cdot \frac{1}{2} + \frac{5}{10} \cdot \frac{1}{6}
\]
\[
= \frac{1}{6} \left[ \frac{2 \times 3}{45} + \frac{2 \times 25}{45} + \frac{10}{45} \right]
\]
\[
= \frac{1}{6} \times \frac{100}{45} = \frac{10}{27}
\]

(b) \[
P(A_3 | R) = \frac{P(R | A_3) P(A_3)}{P(R)} = \frac{\left(\frac{5}{10}\right) \times \frac{1}{6}}{\frac{10}{27}}
\]
\[
= \frac{10}{45} \times \frac{1}{6} \times \frac{27}{10} = \frac{1}{10}
\]

Remark 1.3.4.

In the above example,

\[
P(A_1 | R) \Rightarrow P(R | A_1) = \frac{\left(\frac{3}{10}\right) \times \frac{1}{3}}{\frac{10}{27}} = \frac{3}{45} \times \frac{1}{3} \times \frac{27}{10} = \frac{3}{50}
\]

\[
P(A_2 | R) = \frac{\left(\frac{8}{10}\right) \times \frac{1}{2}}{\frac{10}{27}} = \frac{28}{45} \times \frac{1}{2} \times \frac{27}{10} = \frac{21}{25}
\]

\[
P(A_1 | R) = \frac{3}{50} < \frac{1}{3} = P(A_1) \iff P(A_1 \cap R) < P(A_1) P(R)
\]
\[\iff \text{R has negative information about } A_1\]

\[
P(A_2 | R) = \frac{21}{25} > \frac{1}{2} = P(A_2) \iff P(A_2 \cap R) > P(A_2) P(R)
\]
\[\iff \text{R has positive information about } A_2\]

\[
P(A_3 | R) = \frac{1}{10} < \frac{1}{6} = P(A_3) \iff P(A_3 \cap R) < P(A_3) P(R)
\]
\[\iff \text{R has negative information about } A_3\]

Note that:

\[\text{Proportion of red chips in } C_1 > \text{Proportion of red chips in } C_2, \quad C_2 \subseteq C_1\]
Definition 1.3.2
Let $\{E_j : j \in J\}$ be a collection of events.

(a) Events $\{E_j : j \in J\}$ are said to be pairwise independent if for any pair of events $E_i$ and $E_j$ ($i, j \in J$, $i \neq j$) in the collection $\{E_j : j \in J\}$ we have

$$P(E_i \cap E_j) = P(E_i)P(E_j)$$

(b) Events $\{E_j : j \in J\}$ are said to be independent if for any subcollection $\{E_{j_1}, E_{j_2}, \ldots, E_{j_k}\}$ of $\{E_j : j \in J\}$

$$P\left(\bigcap_{j=1}^{k} E_{j} \right) = \prod_{j=1}^{k} P(E_{j})$$

(c) Let $\Lambda$ be an arbitrary subset of $\mathbb{R}$ such that $\{E_j : j \in \Lambda\}$ is an arbitrary collection of events. Events $\{E_j : j \in \Lambda\}$ are said to be independent if any finite subcollection of events in $\{E_j : j \in \Lambda\}$ forms a collection of independent events.

Theorem 1.3.4
Let $E_1, E_2, \ldots$ be a collection of independent events. Then

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P(A_k)$$

Proof
Let

$$B_n = \bigcap_{k=1}^{n} A_k, \quad n \geq 2, \ldots$$

Then $B_n \in \sigma(A_{1,2,\ldots,n})$. From Problem 1.4, Assignment I,

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \to \infty} P(B_n)$$

But $\bigcap_{n=1}^{\infty} B_n = A_1$ and $P(B_n) = P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P(A_k)$. Thus

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \lim_{n \to \infty} \prod_{k=1}^{n} P(A_k) = \prod_{k=1}^{n} P(A_k).$$
Remark 1.3.5. (a) To verify that \( n \) events \( E_1, \ldots, E_n \) are independent one must verify
\[
\binom{2^n}{2} + \binom{2^n}{3} + \cdots + \binom{2^n}{n} = 2^{n-1}
\]
Conditions. For example to conclude that three events \( E_1, E_2, \) and \( E_3 \) are independent the following 4 (2^3-3-1) conditions must be verified:
\[
\begin{align*}
P(E_1 \cap E_2) &= P(E_1 \cap E_2) ; \quad P(E_1 \cap E_3) = P(E_1) P(E_3) \\
P(E_2 \cap E_3) &= P(E_2 \cap E_3) ; \quad P(E_1 \cap E_2 \cap E_3) = P(E_1) P(E_2) P(E_3)
\end{align*}
\]
(b) Any \( n \)-fold collection of independent events \( A_1 \) of independent events.
In particular the independence of a collection of events implies their pairwise independence.
(c) If \( E_1 \) and \( E_2 \) are independent events \( P(E_1) P(E_2) \), then
\[
P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1) P(E_2)}{P(E_2)} = P(E_1)
\]
(c) i.e. \( P(E_1 | E_2) = P(E_1) \) (Conditional probability of \( E_1 \) given \( E_2 \) is the same as unconditional probability of \( E_1 \)).
Similarly if \( E_1, E_2, \) and \( E_3 \) are independent events then
\[
P(E_1 | E_2 \cap E_3) = P(E_1)
\]
Example 1.3.5. Consider the probability space \((\Omega, \mathcal{F}, P)\) with
\( \Omega = \{1, 2, 3, 4\} \) and \( P(\{1\}) = \frac{1}{4}, \quad P(\{2\}) = \frac{1}{4}, \quad P(\{3\}) = \frac{1}{4}, \quad P(\{4\}) = \frac{1}{4} \). Then \( A, B, \) and \( C \) are pairwise independent but not independent.

Solution. We have
\[
P(A) = P(B) = P(C) = \frac{1}{2}.
\]
\[
P(A \cap B) = P(A \cap C) = P(B \cap C) = P(\{1\}) = \frac{1}{4}.
\]
Thus
\[
P(A \cap B) \neq P(A) P(B) ; \quad P(A \cap C) = P(A) P(C) ; \quad P(B \cap C) = P(B) P(C).
\]

29/3
Implying that $A$, $B$ and $C$ are pairwise independent.

However

$P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\bar{A}) \neq \frac{1}{9} = P(A) P(B) P(C)$

= $A$, $B$ and $C$ are not independent although they are pairwise independent.

**Example 13.6** Let $E_1, E_2, \ldots, E_n$ be a collection of independent events. Show that,

(a) for any permutation $(x_1, \ldots, x_n)$ of $(1, \ldots, n)$, $E_{x_1}, E_{x_2}, \ldots, E_{x_n}$ are independent;

(b) $E_1, E_2, E_3, \ldots, E_n$ are independent for any $k \in \{1, \ldots, n\}$

(c) $E_1^c, E_2 \cup E_3 \cup E_5$ are independent

(d) $E_1 \cup E_3^c, E_2^c$ and $E_4 \cup E_5^c$ are independent.

**Remark 13.6.** When we say that the two random experiments are performed independently, it means that the events associated with two random experiments are independent.