Chapter 1

Question

There are many random phenomena (example?) in our real life. What is the language/mathematical structure that we use to depict them?

Outline

- sample space
- event
- probability measure
  - conditional probability
  - independence
- three theorems
  - multiplication law
  - law of total probability
  - Bayes’ rule

Website of My Probability Course

Definition (sample space, TBp. 2)

A sample space $\Omega$ is the set of all possible outcomes in a random phenomenon.

Example 1.1 (throw a coin 3 times, TBp. 35)

$\Omega = \{ hhh, hht, hth, thh, htt, tht, tth, ttt \}$  $h$: head  $t$: tail

$\Omega$ is a finite set

Example 1.2 (number of jobs in a print queue, Ex. B, TBp. 2)

$\Omega = \{ 0, 1, 2, \ldots \}$

$\Omega$ is an infinite, but countable, set

Example 1.3 (length of time between successive earthquakes, Ex. C, TBp. 2)

$\Omega = \{ t \mid t \geq 0 \} = [0, \infty)$

$\Omega$ is an infinite, but uncountable, set

Question

What are the differences between the $\Omega$ in these examples?
**Definition (event, TBp. 2)**

A particular subset of $\Omega$ is called an **event**.

**Example 1.4 (cont. Ex. 1.1)**

Let $A$ be the event that total number of heads equals 2, then $A = \{hht, hth, thh\}$.

**Example 1.5 (cont. Ex. 1.2)**

Let $A$ be the event that fewer than 5 jobs in the print queue, then $A = \{0, 1, 2, 3, 4\}$.

- **union.** $C = A \cup B \Rightarrow C$: at least one of $A$ and $B$ occur.
- **intersection.** $C = A \cap B \Rightarrow C$: both $A$ and $B$ occur.
- **complement.** $C = A^c \Rightarrow C$: $A$ does not occur.
- **disjoint.** $A \cap B = \emptyset \Rightarrow A$ and $B$ have no outcomes in common.

**Definition (probability measure, TBp. 4)**

A **probability measure** on $\Omega$ is a function $P$ from subsets of $\Omega$ to the real numbers that satisfies the following axioms:

1. $P(\Omega) = 1$. ← total prob. = 1
2. If $A \subset \Omega$, then $P(A) \geq 0$. ← non-negativity
3. If $A_1$ and $A_2$ are disjoint, then ← additivity

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

More generally, if $A_1, A_2, \ldots$ are mutually disjoint, then

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

**Example 1.6 (cont. Ex. 1.1)**

Suppose the coin is fair. For every outcome $\omega \in \Omega$, $P(\omega) = \frac{1}{8}$.

$$\Omega = \{h\text{hh}, h\text{ht}, h\text{th}, \text{thh}, h\text{tt}, \text{tht}, \text{tht}, \text{ttt}\} \quad \sum P(\Omega) = [0, 1]$$
Property A. \( P(A^C) = 1 - P(A). \)

Property B. \( P(\emptyset) = 0. \)

Property C. If \( A \subset B, \) then \( P(A) \leq P(B). \)

Property D. \( P(A \cup B) = P(A) + P(B) - P(A \cap B). \)

**Definition** (conditional probability, TBp. 17)

Let \( A \) and \( B \) be two events with \( P(B) > 0. \) The **conditional probability** of \( A \) given \( B \) is defined to be

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

**Example 1.7** (Ex. B, TBp. 18)

Suppose that the first throw is \( h. \) What is the probability that we can get exact two \( h \)'s in the three trials?

\[
\Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}
\]

\[
B = \{hhh, hht, hth, htt\}
\]

\[
A = \{hht, htt, thh\}
\]

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4} = \frac{6}{8}
\]

**Theorem** (Multiplication Law, TBp. 17)

Let \( A \) and \( B \) be events and assume \( P(B) > 0. \) Then

\[
P(A \cap B) = P(A|B)P(B).
\]

**Example 1.7** (Ex. B, TBp. 18)

Suppose if it is cloudy \((B)\), the probability that it is raining \((A)\) is 0.3, and that the probability that it is cloudy is \( P(B) = 0.2. \)

The probability that it is cloudy and raining is

\[
P(A \cap B) = P(A|B)P(B) = 0.3 \times 0.2 = 0.06.
\]
**Theorem (Law of Total Probability, TBp. 18)**
Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all $i$. Then, for any event $A$, $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$.

**Theorem (Bayes’ Rule, TBp. 20)**
Let $A$ and $B_1, \ldots, B_n$ be events where the $B_i$ are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$ and $P(B_i) > 0$ for all $i$. Then $P(A \cap B_j) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$.

**Definition (independence, TBp. 24)**
Two events $A$ and $B$ are said to be independent if $P(A \cap B) = P(A)P(B)$. A collection of events $A_1, A_2, \ldots, A_n$ are said to be mutually independent if for any subcollection, $A_{i_1}, \ldots, A_{i_m}$, $P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$.

When $A$ and $B$ are independent, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$, and $P(A^c|B) = P(A^c)$. Furthermore, $P(A|B^c) = P(A)$ and $P(A^c|B^c) = P(A^c)$.

**Reading**: textbook, Sections 1.1, 1.2, 1.3, 1.5, 1.6, 1.7

**Further Reading**: Roussas, Chapters 1 and 2

made by Shao-Wei Cheng (NTHU, Taiwan)
**Chapters 2 and 3**

**Outline**

- random variables
- distribution
  - discrete and continuous
  - univariate and multivariate
  - cdf, pmf, pdf
- conditional distribution
- independent random variables
- function of random variables
  - distribution of transformed r.v.
  - extrema and order statistics

### random variable

**Definition 2.1** (random variable, TBp. 33)

A random variable is a function from $\Omega$ to the real numbers.

$\Omega \rightarrow \mathbb{R}$

- statistical modeling usually starts from here.

### Example 2.1 (cont. Ex. 1.1)

1. $X_1$ = the total number of heads
2. $X_2$ = the number of heads on the first toss
3. $X_3$ = the number of heads minus the number of tails

$\Omega = \{hhh, hht, hth, thh, htt, ttt\}$

- \[ x_1 : \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \]
- \[ x_2 : \frac{1}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8} \]
- \[ x_3 : \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \]

**Question 2.1**

Why statisticians need random variables? Why they map to real line?

We need random variable because...
**Question 2.2**

A random variable have a sample space on real line. Does it bring some special ways to characterize its probability measure?

<table>
<thead>
<tr>
<th>/univariate r.v.</th>
<th>discrete</th>
<th>continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>pmf</td>
<td></td>
<td>pdf</td>
</tr>
<tr>
<td>cdf</td>
<td></td>
<td>cdf</td>
</tr>
<tr>
<td>mgf/chf</td>
<td></td>
<td>mgf/chf</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/multivariate r.v.'s</th>
<th>discrete</th>
<th>continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>joint pmf</td>
<td></td>
<td>joint pdf</td>
</tr>
<tr>
<td>joint cdf</td>
<td></td>
<td>joint cdf</td>
</tr>
<tr>
<td>joint mgf/chf</td>
<td></td>
<td>joint mgf/chf</td>
</tr>
</tbody>
</table>

pmf: probability mass function, pdf: probability density function, cdf: cumulative distribution function

mgf (moment generating function) and chf (characteristic function) will be defined in Chapter 4

**Definition 2.2** (discrete and continuous random variables, TBp. 35 and 47)

A discrete random variable can take on only a finite or at most a countably infinite number of values. A continuous random variable can take on a continuum of values.

**e.g.**

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X \in {0,1,2,3}$</td>
<td>$X \in [0,1]$</td>
</tr>
<tr>
<td>$X \in \mathbb{Z}_+$</td>
<td>$X \in (-\infty, \infty)$</td>
</tr>
</tbody>
</table>

**Definition 2.3** (cumulative distribution function, TBp. 36)

A function $F$ is called the cumulative distribution function (cdf) of a random variable $X$ if

$$F(x) = P(X \leq x), \ x \in \mathbb{R}.$$
**Definition 2.4** (probability mass function/frequency function, TBP. 36)

A function $p(x)$ is called a **probability mass function** (pmf) or a **frequency function** if and only if (1) $p(x) \geq 0$ for all $x \in X$, and (2) $\sum_{x \in X} p(x) = 1$.

For a **discrete** random variable $X$ with pmf $p(x)$, we have:

$$P(X = x) = p(x),$$

and

$$P(X \in A) = \sum_{x \in A} p(x).$$

The **cumulative distribution function** is defined as:

$$F(x) = \sum_{t \leq x} P(X = t) = \sum_{t \leq x} p(t).$$

Also, for any $a < b$,

$$p(x) = P(X = x) = F(x) - F(x-) = \lim_{x \uparrow a} F(t).$$