Lecture 1
Probability Theory

1. Probability

Definition [Random Experiment]: A random experiment is a mechanism which has at least two possible outcomes, and which outcome to occur is unknown in advance. In other words, a random experiment is a mechanism for which the outcome cannot be predicted with certainty.

Definition [Sample Space]: A random experiment is an experiment leading to at least two possible outcomes with uncertainty as to which will occur. The possible outcomes of the random experiment are called "basic outcomes", and the set of all basic outcomes is called the "sample space", denoted by S.

Example: [A coin is thrown] Two possible outcomes: "head", and "tail". So the sample space is $S = \{H, T\}$.
Example: [Direction of Changes] Let $Y = 1$ if the U.S. Gross Domestic Product (GDP) growth rate $X > 0$ and $Y = 0$ if the U.S. GDP growth rate $X < 0$. Then we can make inference of asymmetric business cycles using information of Y with $S = \{0, 1\}$.
Example: [A die is rolled] The basic outcomes are the numbers $1, 2, 3, 4, 5, 6$. The sample space $S = \{1, 2, 3, 4, 5, 6\}$.
Example: Two coins are thrown. How many possible outcomes? $S = \{(H, H), (H, T), (T, H), (T, T)\}$.

A sample space S can be countable or uncountable. The distinction between a countable sample space and an uncountable sample space dictates the ways in which probabilities will be assigned.

Example: T_0 is the lowest temperature of an area, and T_1 is the highest temperature of the area. Let T denotes possible temperature of the area. The sample space $S = \{T : T_0 \leq T \leq T_1\}$.

We next define the concept of events that will allow us to investigate what we may be interested in.

Definition [Event]: An event A is a collection of basic outcomes from the sample space S that share certain common features. The event A is said to occur if the random experiment gives rise to one (and only one) of the constituent basic outcomes in A. That is, an event occurs if any of its basic outcomes has occurred (or equivalently if the outcome of the random experiment is an element of event A).

Example: [A die is rolled] Event A is defined as "number resulting is even". Event B is "number resulting is at least 4". Then $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$.

Remark: Distinguish the sample space, a basic outcome and an event.

basic outcome $\in$ event $\subset$ sample space.

Definition [Sigma Algebra]: A sigma (σ-) algebra, denoted by $\mathcal{F}$, is a collection of subsets (events) of S with
(i) $\emptyset \in \mathcal{F}$ (the empty set is contained in $\mathcal{F}$).
(ii) If $A \in \mathcal{F}$; then $A^c \in \mathcal{F}$ ($\mathcal{F}$ is closed under countable complements).
(iii) If $A_1, A_2, \ldots \in \mathcal{F}$; then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ ($\mathcal{F}$ is closed under countable unions).

Remarks:

- A σ-field $\mathcal{F}$ is the domain of a probability function on which we will define our probability. In probability theory, the event space is a σ-field.
- (i) and (ii) imply that $S \in \mathcal{F}$.
- (ii) and (iii) imply that $\mathcal{F}$ is also closed countable intersection.
- For each sample space S, we can have many different σ-fields.

Example: The trivial σ-field: Let $\mathcal{F} = \{\emptyset, S\}$.
(i) $\emptyset \in \{\emptyset, S\}$. Thus $\emptyset \in \mathcal{F}$.
(ii) $\emptyset^c = S \in \{\emptyset, S\}$ and $S^c = \emptyset \in \mathcal{F}$.
(iii) $\emptyset \cup S = S \in \mathcal{F}$.
Example: $S = \{1, 2, 3\}$, then a set containing the following subsets $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$, and $\emptyset$ is a σ-field.
$\mathcal{F} = \{S, \emptyset\}$
$\mathcal{F} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \emptyset\}$
Remarks: Define σ-field $\mathcal{F}$ to be the collection of all possible subsets in S.
A σ-field is a collection of events.
The pair $(S, \mathcal{F})$ is called a measur*able* space.

Next, we define a (probability) measure of a measuable space $(S, \mathcal{F})$.

Definition [Probability Function]: Suppose a random experiment has a sample space S and an associated σ-field $\mathcal{F}$: The probability function $P : \mathcal{F} \to \mathbb{R}^+$ is a mapping that satisfies the following properties:

(i) $0 \leq P(A) \leq 1$ for any event A in $\mathcal{F}$.

(ii) $P(S) = 1$.

(iii) If $A_1, A_2, \ldots \in \mathcal{F}$ are mutually exclusive, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P\left(A_i\right).$

Definition [Probability Space]: A probability space is a triple $(S, \mathcal{F}, P)$ where:
(a) S is the sample space corresponding to outcomes of the underlying random experiment.
(b) $\mathcal{F}$ is the σ-field of subsets of S. These subsets are called events.
(c) P is a probability measure.

Remark: A probability space $(S, \mathcal{F}, P)$ completes describes a random experiment associated with sample space S.

Theorem 1: If $\emptyset$ denotes the empty set, then $P(\emptyset) = 0$.
Proof: $S = S \cup \emptyset, P(S) = P(S \cup \emptyset)$. $P(S) = P(S) + P(\emptyset)$.

Theorem 2: $P(A) = 1 - P(A^c)$.
Proof: Observe $S = A \cup A^c$. Then $P(S) = P(A \cup A^c)$ and $1 = P(A) + P(A^c)$.

Question for you: Suppose X denotes the outcome of some random experiment. The following is the probability distribution for X:

$$\begin{array}{cccccccc} X & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\ P & \frac{1}{2^1} & \frac{1}{2^2} & \frac{1}{2^3} & \frac{1}{2^4} & \frac{1}{2^5} & \frac{1}{2^6} & \cdots \end{array}$$

Find the probability that X is greater than 3 using Theorem 2.

Theorem 3: If A and B are two events in S, and $A \subseteq B$, then $P(A) \leq P(B)$.
Proof: Observe

$$\begin{aligned} B &= S \cap B = (A \cup A^c) \cap B \\ &= (A \cap B) \cup (A^c \cap B) \\ &= A \cup (A^c \cap B); \end{aligned}$$

where the last equality follows from $A \subseteq B$. Hence, we have

$$P(B) = P(A) + P(A^c \cap B) \geq P(A).$$

Corollary: For any event $\emptyset \subseteq A \subseteq S$, $0 \leq P(A) \leq 1$.

Theorem 4: For any two events A and B in $\mathcal{F}$,

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

Proof: Since $A \cup B = A \cup (A^c \cap B)$ and A and $A^c \cap B$ are mutually exclusive, we have

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

On the other hand, because $B = S \cap B = (A \cap B) \cup (A^c \cap B)$ and both $A \cap B$ and $A^c \cap B$ are mutually exclusive, we have

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

Adding both equations yields

$$\begin{aligned} & P(A \cup B) + P(A \cap B) + P(A^c \cap B) \\ = \;& P(A) + P(A^c \cap B) + P(B). \end{aligned}$$

This delivers the desired result.

Theorem 4 gives:

Bonferroni's Inequality:

$$
\begin{aligned}
P(A \cup B) &\geq P(A) + P(B) - 1. \\
P(A \cap B) &\geq P(A) + P(B) - 1.
\end{aligned}
$$

Theorem 5: Suppose P is a probability function.
(a) [Rule of Total Probability] If $C_1, C_2, \ldots$ are mutually exclusive and collectively exhaustive, then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

(b) [Subadditivity] For any sequence of events $\{A_i\}$,

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Remark:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$$

$$P(A_1 \cup A_2 \cup A_3) = P(A_1 \cup B) \leq P(A_1) + P(B), \qquad B = A_2 \cup A_3.$$

Example: If $A = \{$students in our school whose scores > 90 points$\}$ and $C_i = \{$students from class $i\}$, then $A \cap C_i = \{$students from class i whose scores are > 90 points$\}$.

2. Conditional Probability

Examples

- *Volatility spillover:* Shocks that originate in the asset market can have impact on output fluctuations, and vice versa.
- *Financial Contagion:* A large drop of the price in one market can cause a large drop of the price in another market, given the speculations and reactions of market participants. This can occur regardless of market fundamentals.
- *Wage gap:* Earnings may vary depending on education, gender, race, etc.
- *Boston HMDA Data:* Mortgage Lending in Boston, Munnell et al (1996, *AER*)

Definition [Conditional Probability]: Let A and B be two events in probability space $(S, \mathcal{F}, P)$. Then the conditional probability of event A given event B, denoted as $P(A|B)$, is defined as

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

provided that $P(B) > 0$. Similarly, the conditional probability of event B given A is defined as

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

provided $P(A) > 0$.

Remarks:

- Consider B and A as a new sample space in each case. That is, treat B as a new sample space when we consider $P(A|B)$; the intuition is that our original sample space S has been updated to B: All further occurrences are then calibrated with respect to their relation to B: Similarly, treat A as a new sample space when we consider $P(B|A)$.
- Event B becomes a new sample space when considering the conditional probability $P(A|B)$. In this case, $(S \cap B, \mathcal{F} \cap B, P(\cdot|B))$ is a probability space associated with the conditional probability function $P(A|B)$.

Question for you: Is it true that $P(A) = P(A|S)$? Yes.
(a) Intuitively, S represents the "least knowledge" among all possible events, or equivalently the highest degree of uncertainty. Therefore, it is the same as not conditioning on any useful information.
(b) We consider only the case $P(B) > 0$. This is because $P(B) = 0$ implies that B is impossible to happen, and conditioning on an impossible event is meaningless.
(c) The whole sample space S describes the largest degree of uncertainty. When some event B has occurred, this uncertainty is then reduced from S to B. In other words, the new sample space is now equal to B.
(d) $P(A|B)$ satisfies all probability laws defined on the sample space B. That is, conditional probability can be regarded as a reduction/updating of sample space.

Question for you: Is it true that $P(A^c|B) = 1 - P(A|B)$? Yes.

Given $(A^c \cap B) \cup (A \cap B) = B$, we have

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

It follows that

$$\begin{aligned} P(A^c|B) &= P(A^c \cap B)/P(B) \\ &= 1 - P(A \cap B)/P(B) \\ &= 1 - P(A|B). \end{aligned}$$

Multiplication Rules:
(i) $P(A \cap B) = P(A|B)P(B)$
(ii) $P(A \cap B) = P(B|A)P(A)$.
This formula can be used to compute the joint probability.

Example: Selecting two balls. Suppose two balls are to be selected, without replacement, from a box containing r red balls and b blue balls. What is the probability that the first is red and the second is blue?**.**
$A = \{\text{the first ball is red}\}$, $B = \{\text{the second ball is blue}\}$. Then

$$\begin{aligned} P(A) &= \frac{r}{r+b} \\ P(B|A) &= \frac{b}{(r+b-1)} \\ P(A \cap B) &= P(B|A)P(A) = \frac{rb}{(r+b)(r+b-1)}. \end{aligned}$$

Question for you: Let A and B be disjoint and $P(B) > 0$.

1. Is it true that $P(A|B) = 0$? Yes.
2. Does it mean that $P(A|B) = P(A)$? No.
3. Provide an example to demonstrate #1 and #2.

The multiplication rule can be repeatedly used to obtain the joint probability of multiple events.

Example [Computation of Joint Probabilities]:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 \cap B) = P(A_3|B)P(B) \\ &= P(A_3|A_2 \cap A_1)P(A_2 \cap A_1) \\ &= P(A_3|A_2 \cap A_1)P(A_2|A_1)P(A_1). \end{aligned}$$

Remark: Joint probabilities are important for the so-called maximum likelihood estimation.

Rule of Total Probability: Let $\{B_i\}_{i=1}^{\infty}$ be a partition (i.e., mutually exclusive and collectively exhaustive) of S, $P(B_i) > 0$ for $i \geq 1$. For event A in S,

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

This follows from

$$\begin{aligned} A &= A \cap S \\ &= A \cap (B_1 \cup B_2 \cup B_3 \ldots) \\ &= \cup_{i=1}^{\infty} (A \cap B_i). \end{aligned}$$

Remark: This is also called the rule of elimination.

Bayes' Theorem:

$$\begin{aligned} P(A|B) &= \frac{P(A\cap B)}{P(B)} \\ &= \frac{P(A\cap B)}{P(B|A)P(A)+P(B|A^c)P(A^c)} \end{aligned}$$

Remarks: In words, Bayes' theorem shows how conditional probabilities of the form $P(B|A)$ may be combined with initial probability of A to obtain the final probability $P(A|B)$.
(i) $P(A) =$ prior probability about the event A.
(ii) $P(A|B) =$ posterior probability about the event A given that B has occurred.

Bayes' Theorem: Suppose $B_1, \ldots, B_n$ be n mutually exclusive and collectively exhaustive events in the sample space S. The conditional probability of B_i given event A is

$$\begin{aligned} P(B_i|A) &= \frac{P(A \cap B_i)}{P(A)} \\ &= \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}, \qquad i = 1, \ldots, n. \end{aligned}$$

Interpretation: Our interest is the probability about B_i given event A has occurred. When event A has occurred, we may have better knowledge about the occurrence of B_i. Event A provides useful information from our updating knowledge.

Example [Auto-insurance]: B_1, B_2, B_3 are the events that a customer is a type 1, 2, 3 customer. $A =$ Event that the customer has had received one speeding ticket in the last 12 months. What is $P(B_i|A)$?

Suppose an insurance company has three types of customers – high risk, medium risk and low risk. From the company's consumer database, it is known that 25% of its customers are high risk, 25% are medium risk, and 50% are low risk. Also, the database shows that the probability that a customer has at least one accident in the current year is 0.25 for high risk, 0.16 for medium risk, and 0.10 for low risk. Now a new customer wants to be insured and he has had one accident this year. What is the probability that he is high risk, given that he has had one accident during the current year?

Solution: We denote $H = \{\text{high risk}\}$, $M = \{\text{medium risk}\}$, $L = \{\text{low risk}\}$, and $A = \{\text{received a speeding ticket}\}$. Then $P(H) = 0.25, P(M) = 0.25, P(L) = 0.50,$
$P(A|H) = 0.25, P(A|M) = 0.16, P(A|L) = 0.10$. Find $P(H|A)$.
[Answer. 0.41].

Remark: Similar question – Credit default risk

3. Independence

Definition [Independence]: Both A and B are said to be statistically independent if $P(A \cap B) = P(A)P(B)$.

Interpretation: By this definition,

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

Therefore, the knowledge of B does not help in predicting A, because the occurrence of B does not affect the probability of the occurrence of A. Similarly, we have $P(B|A) = P(B)$, i.e. the occurrence of A has no effect on the occurrence or probability of B.

Example [Random Walk Hypothesis, Fama 1970]: If a stock market is informationally fully efficient, then the stock price P_t will follow a random walk; that is, $P_t = P_{t-1} + X_t$ where the stock price change $\{X_t = P_t - P_{t-1}\}$ is independent across different periods. If $\{X_t\}$ is serially independent across observations, then X_t is not predictable using past information.

Theorem: Let A and B are two independent events. Then
(a) A and B^c,
(b) A^c and B,
(c) A^c and B^c are all independent.
Remark: Use intuition first: A and B^c should be independent because if not, we would be able to predict B^c from A, and thus predict B.
Proof: *Question for you:* . (See CB 1.40.)

We now provide a definition of independence for more than two events.

Definition [Independence Among Several Events]: k events $A_1, A_2, \ldots, A_k$ are mutually independent if, for every subset $A_{i_1}, \ldots, A_{i_j}$ of j of those events $(j = 2, 3, \ldots, k)$,

$$P(A_{i_1} \cap \cdots \cap A_{i_j}) = P(A_{i_1}) \cdots P(A_{i_j}).$$

Note: We need $(2^k - 1 - k)$ conditions to characterize independence among several events because $\sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) = 2^k$, $\left(\begin{array}{c} k \\ 0 \end{array} \right) = 1$, $\left(\begin{array}{c} k \\ 1 \end{array} \right) = k$.

Example: Three events A, B, and C are independent, if

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(A \cap C) &= P(A)P(C), \\ P(B \cap C) &= P(B)P(C), \end{aligned} \tag{1}$$

and

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{2}$$

Question for you:
Show that checking just (2) is not enough. See CB Example 1.3.10.
Show that checking just (1) is not enough either. See CB Example 1.3.11.

Remarks:

(a) For three or more events, independence is called mutual independence or joint independence. If there is no possibility of misunderstanding, independence is often used without "mutual" or "joint" when considering several events.

(b) A collection of events are mutually independent if the joint probability of *any* subcollection of the events is equal to the product of the individual probabilities.

(c) Joint independence implies pairwise independence, but not vice versa.

(d) *Question for you:* It is possible to find that three events are pairwise independent but not jointly independent. The implication is that if one uses A_2 or A_3 to predict A_1; then A_2 or A_3 are not helpful. But if one uses A_2 and A_3 simultaneously to predict A_1, then A_1 is predictable. [This can be an interesting empirical research topic!]

Question for you: Suppose two events A_1 and A_2 have a common element, i.e., $P(A_1 \cap A_2) > 0$. Can they be independent?

Example: Four cards are numbered 1, 2, 3, 4. The experiment is to select one card randomly.
$A_1 = \{1 \text{ or } 2\}$, $A_2 = \{1 \text{ or } 3\}$
$P(A_1) = \frac{1}{2} = P(A_2)$.
$A_1 \cap A_2 = \{1\}$. $P(A_1 \cap A_2) = \frac{1}{4}$.
A_1 and A_2 are independent, although they have a common element.

4. Random Variables and Distribution Functions

Random variable

Definition [Measurable Function]: A function $g : S \to \mathbb{R}$ is measurable-$\mathcal{F}$ if for every $a \in \mathbb{R}$, the set $\{s \in S : g(s) \leq a\}$ belongs to $\mathcal{F}$.

Definition [Random Variable]: Let $\{S, \mathcal{F}, P\}$ be a probability space and $X(\cdot)$ is a real-valued function on S which assigns one and only one real number $x = X(s)$ to each $s \in S$. Then X is a random variable if and only if for each $x \in \mathbb{R}$, $A \equiv \{s \in S : X(s) \leq x\} \in \mathcal{F}$.

Remark: Let $A = \{s \in S : X(s) \in B\}$, where $B = (-\infty, x]$. Since $A \in \mathcal{F}$, the probability measure $P(A)$ is the probability of the event A. We now need a probability measure to measure the probability of the event B. Define $\mu(B) = \Pr(X \in B) = \Pr(X \leq x) = P(A)$. Thus to each random variable X, there corresponds a probability space $\{\mathbb{R}, \mathcal{B}, \mu\}$. This probability space is said to be induced by

$$X : \{S, \mathcal{F}, P\} \xrightarrow{X} \{\mathbb{R}, \mathcal{B}, \mu\}.$$

Distribution function

Given $(S, \mathcal{F}, P)$, let $X(s)$ be a real function on $s \in S$. Then X is a random variable iff for $\forall x \in \mathbb{R}$, $A = \{s \in S | X(s) \leq x\} \in \mathcal{F}$. $B = (-\infty, x] \in \mathcal{B}$ is called a Borel set.

$$\begin{aligned} P(A) &= \Pr(s \in A) \\ &= \Pr\left[(s \in \{s \in S | X(s) \leq x\}\right] \\ &= \Pr(X(s) \leq x) \\ &= \Pr(X(s) \in B) \\ &= \mu(B). \end{aligned}$$

Write this in a new notation

$$\mu(B) = \Pr(X \leq x) \equiv F_X(x),$$

which we call the (cumulative) distribution function (CDF) of $X(s)$.

Theorem: The distribution function $F(x)$ has the following properties:

1. $F(x)$ is monotonically non-decreasing: $F(x) \geq F(y)$ if $x \geq y$
2. $F(x)$ is right continuous: $\lim_{y \downarrow x} F(y) = F(x)$
3. $F(-\infty) = \lim_{x \downarrow -\infty} F(x) = 0$,
   $F(+\infty) = \lim_{x \uparrow +\infty} F(x) = 1$.

Example [CB 1.5.2] Verify the three conditions of the above theorem for the distribution function of a random experiment of tossing three coins with X being the number of heads. Write down $F(x) = \Pr(X \leq x)$ and show that it is a cdf.
Consider a random experiment of tossing a fair coin: $S = \{H, T\}$.
Let $X(s = H) = 1$ and $X(s = T) = 0$.

$$F(x) = \Pr(X \leq x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Example [CB 1.5.4] Verify the three conditions of the above theorem for the *geometric distribution*. Consider a random experiment of tossing a coin until a head appears.

$$\Pr(X = x) = (1-p)^{x-1}p, \qquad 0 < p < 1,\ x = 1, 2, \ldots$$

$$F(x) = \Pr(X \le x) = \sum_{i=1}^{x} \Pr(X = i) = \sum_{i=1}^{x} (1-p)^{i-1}p = 1-(1-p)^{x}.$$

Example [CB 1.5.5] Verify the three conditions of the above theorem for the *logistic distribution*

$$F(x) = \Pr(X \le x) = \frac{1}{1+e^{-x}}.$$

Definition [CB 1.5.7] A random variable X is *continuous* if $F(x)$ is a continuous function of x. A random variable X is *discrete* if $F(x)$ is a step function of x.

Remark: A probability measure on $\mathcal{B}$ determines a distribution function. The converse is also true: each distribution function F determines a unique probability measure on $\mathcal{B}$. The probability measure and the distribution function are uniquely related. Therefore, the distribution of X (in terms of probability measure μ) is completely determined by the distribution function F.

Definition [Probability Mass Function, CB 1.6.1] The probability mass function (pmf) of a discrete random variable X is given by

$$f(x) = \Pr(X = x) \qquad \text{for all } x.$$

Definition [Probability Density Function, CB 1.6.3] The probability density function (pdf) of a continuous random variable X is given by $f(x)$ satisfying

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \text{for all } x.$$

Example [CB 1.6.2] The pmf for the geometric distribution for a random experiment of tossing a coin until a head appears is

$$f(x) = \Pr(X = x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases}$$

Example [CB 1.6.3] The pdf for the logistic distribution is

$$f(x) = \frac{d}{dx}F(x) = \frac{e^{-x}}{(1+e^{-x})^2},$$

where $F(x) = \Pr(X \leq x) = \frac{1}{1+e^{-x}}$.

Definition [CB 1.5.8, CB 1.5.10] The random variables X and Y are *identically distributed* if

$$\Pr(X \in (-\infty, b]) = \Pr(Y \in (-\infty, b]),$$

for all $b \in \mathbb{R}$.

Question for you: The definition does not say that $X = Y$. See CB Example 1.5.9.

Random vector

Definition: Let $\{S, \mathcal{F}, P\}$ be a probability space. A random vector $X \in \mathbb{R}^k$ defined on $\{S, \mathcal{F}, P\}$ is a $\mathbb{R}^k$-valued function $X(\cdot)$ on S which assigns one and only one $x = X(s)$ to each $s \in S$, and is such that for each $x_1, \cdots, x_k \in \mathbb{R}$,
$A \equiv \{s \in S : X_1(s) \leq x_1, \cdots, X_k(s) \leq x_k\} \in \mathcal{F}$.

Remark: The above definition implies that the components of X are random variables, $X = (X_1, X_2, \cdots, X_k)'$. A vector of random variables is a random vector. Similarly to the univariate case, each random vector induces a probability space $\left\{\mathbb{R}^k, \mathcal{B}, \mu\right\}$, where μ is defined as $\mu(B) = \Pr(X \in B) = \Pr(A)$, where
$B = \times_{i=1}^{k}(-\infty, x] \in \mathcal{B}$.

Multivariate distribution function

We can similarly define a multivariate distribution function:

$$F(x_1, \cdots, x_k) = \mu\left(\times_{j=1}^{k}(-\infty, x_j]\right) = \Pr(X_1 \le x_1, \cdots, X_k \le x_k),$$

which has the following properties:

1. $F(x_1, \cdots, x_k)$ is monotonically non-decreasing in each of its arguments:

$$F(x_1, \cdots, x_k) \ge F(y_1, \cdots, y_k) \quad \text{if} \quad x_j \ge y_j, j = 1, \cdots, k.$$

2. $F(x_1, \cdots, x_k)$ is right continuous in each of its arguments:

$$\lim_{h \downarrow 0} F(x_1, \cdots, x_i + h, \cdots x_k) = F(x_1, \cdots, x_k).$$

3. $\lim_{\min x_j \to -\infty} F(x_1, \cdots, x_k) = 0$,
   $\lim_{\min x_j \to +\infty} F(x_1, \cdots, x_k) = 1$.

Let $X = (X_1', X_2')'$ be a random vector in $\mathbb{R}^{k_1+k_2}$ with $X_1 \in \mathbb{R}^{k_1}$ and $X_2 \in \mathbb{R}^{k_2}$. Let F be the (joint) distribution function of X and F_1, F_2 be the distribution functions of X_1, X_2, respectively. The distribution functions F_1 and F_2 are called marginal distribution functions of F. They can be constructed from F by setting the arguments of F corresponding to X_1 and X_2, respectively, equal to ∞. Thus, $F_1\left(x_1, \cdots, x_{k_1}\right) = F\left(x_1, \cdots, x_{k_1}, \infty, \cdots, \infty\right)$, $F_2\left(x_{k_1+1}, \cdots, x_{k_1+k_2}\right) = F\left(\infty, \cdots, \infty, x_{k_1+1}, \cdots, x_{k_1+k_2}\right)$.

Multivariate density or mass function

Definition: A distribution function F on $\mathbb{R}^k$ is continuous if there exists a non-negative function f on $\mathbb{R}^k$ such that for $x_1, \cdots, x_k \in \mathbb{R}$,

$$F\left(x_1, x_2, \cdots, x_k\right) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_k} f\left(y_1, y_2, \cdots, y_k\right) dy_1 dy_2 \cdots dy_k.$$

This function f is called the density (probability density function, PDF) of F.

Definition: A distribution function F on $\mathbb{R}^k$ is discrete if there exists a countable subset X of $\mathbb{R}^k$ and a positive function f on $\mathbb{R}^k$ such that for $x_1, \cdots, x_k \in \mathbb{R}$,

$$F\left(x_1, x_2, \cdots, x_k\right) = \sum_{\substack{(y_1, y_2, \cdots, y_k)' \in X \\ y_j \leq x_j,\ j=1,\cdots,k}} f\left(y_1, y_2, \cdots, y_k\right).$$

This f is also called the mass (probability mass function, PMF) of F.

Independence of random variables

Definition: Let $X_1, X_2, X_3, \cdots$ be random variables defined on a common probability space $\{S, \mathcal{F}, P\}$. Then X_j, $j = 1, 2, \cdots, n$ (n may be ∞) are independent if $F(x_1, \cdots, x_n) = \Pi_{j=1}^{n} F_j(x_j)$, where $F(x_1, \cdots, x_n)$ is the joint CDF and $F_j(x_j)$, $j = 1, 2, \cdots, n$, are the marginal CDFs.

Theorem: If X_j, $j = 1, 2, \cdots, n$ (n may be ∞) are independent, then

$$f(x_1, \cdots, x_n) = \Pi_{j=1}^{n} f_j(x_j).$$

5. Appendix: Set Theory

Let A and B be two events in the sample space S. Then we will have the following definitions.

Definition [Intersection]: Intersection of A and B, denoted $A \cap B$, is the set of basic outcomes in S that belong to both A and B. The intersection occurs if and only if both events A and B occur.

Definition [Exclusiveness]: If A and B have no common basic outcomes, they are called mutually exclusive and their intersection is empty set $\emptyset$;, i.e., $A \cap B = \emptyset$.

Remarks: (a) Any mutually exclusive events cannot occur simultaneously. (b) Any pair of the basic outcomes are mutually exclusive. (c) Mutually exclusive events are also called disjoint.

Example: Suppose the events A and B are disjoint. Under what condition are A^c and B^c disjoint?

Solution: If and only if $A \cup B = S$, because
$A^c \cap B^c = \emptyset \Longrightarrow (A \cup B)^c = \emptyset \Longrightarrow A \cup B = S.$

Definition [Union]: The union of A and B; $A \cup B$, is the set of all basic outcomes in S that belong to either A or B. The union of A and B occurs if and only if either A or B (or both) occurs.
Definition [Collectively Exhaustive]: Suppose $A_1, A_2, \ldots, A_n$ are n events in the sample space S, where n is any positive integer. If $\cup_{i=1}^{n} A_i = S$, then these n events are said to be collectively exhaustive.
Remark: The set of all basic outcomes in S is collectively exhaustive.
Definition [Complement]: The complement of A is the set of basic outcomes of a random experiment belonging to S but not to A.
Definition [Difference]: The difference of A and B, $A - B = A \cap B^c$, is the set of basic outcomes in S that belong to A but not B.

Laws of Set Operations:

For any three events, A, B, C defined on a sample space S,

1. *Complementation:*

$$
\begin{aligned}
(A^c)^c &= A, \\
\emptyset^c &= S, \\
S^c &= \emptyset.
\end{aligned}
$$

Note: The empty set $\emptyset$ is a subset of any set.

2. *Commutativity* of union and intersection:

$$
\begin{aligned}
A \cup B &= B \cup A, \\
A \cap B &= B \cap A.
\end{aligned}
$$

3. *Associativity* of union and intersection:

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

4. *Distributive Laws*:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

More generally, for any $n \geq 1$,

$$B \cap (\cup_{i=1}^{n} A_i) = \cup_{i=1}^{n} (B \cap A_i),$$
$$B \cup (\cap_{i=1}^{n} A_i) = \cap_{i=1}^{n} (B \cup A_i).$$

5. *De Morgan's laws:*

$$(A \cup B)^c = A^c \cap B^c,$$
$$(A \cap B)^c = A^c \cup B^c.$$

More generally, for any $n \geq 1$,

$$(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} A_i^c,$$
$$(\cap_{i=1}^{n} A_i)^c = \cup_{i=1}^{n} A_i^c.$$

Example: Let A and B be two events in S. Then
(i) Are $A \cap B$ and $A^c \cap B$ mutually exclusive?
(ii) Is $(A \cap B) \cup (A^c \cap B) = B$?

$$(A \cup A^c) \cap B = S \cap B = B.$$

(iii) Are A and $A^c \cap B$ mutually exclusive?
(iv) $A \cup (A^c \cap B) = A \cap B$?

$$(A \cup A^c) \cap (A \cup B) = S \cap (A \cup B) = A \cup B.$$

Example: Let A_i, $i = 1, 2, 3, \ldots, n$, be mutually exclusive and collectively exhaustive, and A is an event. Then
(i) Are $A_1 \cap A$, $A_2 \cap A, \ldots,$ $A_n \cap A$, mutually exclusive?
(ii) Is the union of $A_i \cap A$, $i = 1, 2, 3, \ldots, n$, equal to A?

$$\cup_{i=1}^{n}(A_i \cap A) = A?$$

$n = 2:$

$$\begin{aligned}(A_1 \cap A) \cup (A_2 \cap A) &= (A_1 \cup A_2) \cap A \\ &= S \cap A \\ &= A\end{aligned}$$

Remark: A sequence of collectively exhaustive and mutually exclusive events forms a *partition* of sample space S.