Econ 325

Notes on Probability[1]

By Hiro Kasahara

Properties of Probability

In statistics, we consider **random experiments**, experiments for which the outcome is random, i.e., cannot be predicted with certainty. The possible outcomes of a random experiment are called the **basic outcomes** and the collection of all possible basic outcomes is called **the sample space** denoted by S. Let A be a part of the collection of outcomes in S. Then, A is called an **event**. If the outcome of the experiment is in A, then we say that **event** A **has occurred**.

Here is the terminology and notations we use:

- $\emptyset$ is the **empty** set. No element is in the empty set $\emptyset$. For example, if A and B does not share any common element, then $A \cap B = \emptyset$.
- $A \subset B$ means A is a **subset** of B, i.e., any element in A is also in B.
- $A \cup B$ is the **union** of A and B. $A \cup B$ is a set that consists of the elements that are contained in either A or B.
- $A \cap B$ is the **intersection** of A and B. $A \cap B$ is a set that consists of the elements that are contained in both A and B.
- $\bar{A}$ is the **complement** of A, i.e., all elements in S that are not in A.
- A_1, A_2,..., A_k are **mutually exclusive** events if $A_i \cap A_j = \emptyset$ for all i and j such that $i \neq j$; that is no pair of sets A_i and A_j share common elements so that A_1, A_2,..., A_k are disjoint sets.
- A_1, A_2,..., A_k are **collectively exhaustive** events if $A_1 \cup A_2 \cup ... \cup A_k = S$; that is, A_1, A_2,..., A_k collectively cover all elements in S.

Example 1 (Rolling a die) *Consider a random experiment of rolling a die. In this case, the sample space is* $S = \{1, 2, 3, 4, 5, 6\}$ *and there are six basic outcomes, i.e.,* $\{1\}$, $\{2\}$, ..., $\{6\}$. *We define an event* $A = \{1, 3, 5\}$, *i.e., odd numbers. If we roll a die and if any one of odd numbers has been realized, then we say that event* A *has occurred. Let* $B = \{2, 4, 6\}$ *and* $C = \{1, 3\}$. *Then,* $A \cap B = \emptyset$ *so that* A *and* B *are mutually exclusive. Also,* $A \cup B = S$ *and, therefore,* A *and* B *are collectively exhaustive. In fact,* B *is the complement of* A *and* B *is the complement of* A. *Note that* $C \subset A$ *because all elements in* C *are in* A. *When* $C \subset A$, *we have* $C \cup A = A$ *and* $A \cap C = C$.

Set operations satisfy the following properties:

(Commutative Laws) $\quad A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A$

(Associative Laws) $\quad (A \cup B) \cup C = A \cup (B \cup C) \quad \text{and} \quad (A \cap B) \cap C = A \cap (B \cap C)$

(Distributive Laws) $\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(De Morgan's Laws) $\quad \overline{(A \cup B)} = \bar{A} \cap \bar{B} \quad \text{and} \quad \overline{(A \cap B)} = \bar{A} \cup \bar{B}$

Given event A defined as a subset of the sample space S, we may consider the probability of event A as the relative frequency that event A happens when we repeat the random experiment an infinitely many times.

Formally, we define **probability** as a function from the space of sets to the space of real values between 0 and 1 as follows.

Definition 1 (Probability) ***Probability*** *is a real-valued set function P that assigns, to each event A in the sample space S, a number $P(A)$ such that the following three properties are satisfied:*

1. $P(A) \geq 0$
2. $P(S) = 1$
3. *if A_1, A_2, ..., A_k are events that are mutually exclusive, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$, then*

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

for any positive integer k.

In the above definition, we consider an infinite countable number of sequence of events by letting $k \to \infty$.

For any events A and B in the sample space, probability satisfies the following properties:

1. **Complimentary rule**: For each event A, $P(A) = 1 - P(\bar{A})$.
2. **Addition Rule**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $P(\emptyset) = 0$. This follows from $P(S) = 1$ and $\bar{S} = \emptyset$ so that $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$.
4. If $A \subset B$, then $P(A) \leq P(B)$. This is intuitive because A is contained in B.
5. $P(A \cap B) \geq P(A) + P(B) - 1$.
6. $P(A \cap \bar{A}) = 0$ because $A \cap \bar{A} = \emptyset$.
7. $P(A \cup \bar{A}) = 1$ because $A \cup \bar{A} = S$.

We may formally prove "4. If $A \subset B$, then $P(A) \leq P(B)$' ' as follows. Because $A \cap B$ and $A \cap \bar{B}$ are mutually exclusive (i.e., their interaction is an empty set), writing $B = (A \cap B) \cup (A \cap \bar{B})$ gives $P(B) = P((A \cap B) \cup (A \cap \bar{B})) = P(A \cap B) + P(A \cap \bar{B})$, where the last equality uses the third property in the definition of probability, . Because $A \cap B = A$ when A is a subset of B, we have $P(A \cap B) = P(A)$ and hence the previous equation becomes $P(B) = P(A) + P(A \cap \bar{B})$. Finally, because $P(A \cap \bar{B}) \geq 0$ by the first property of probability in the definition, we have $P(B) \geq P(A)$.

Suppose that A and B are mutually exclusive, i.e., $A \cap B = \emptyset$. In this case, $P(A \cap B) = P(\emptyset) = 0$, and it follows from **Addition Rule** that $P(A \cup B) = P(A) + P(B) - 0 = P(A) + P(B)$.

Factorial, Permutations, and Combinations

The factorial of a positive integer n, denoted by $n!$, represents the number of ways n distinct objects can be put into a sequence is represented by

$$n! = n(n-1) \cdots 2 \cdot 1.$$

For example, consider a marathon race by 8 runners: A, B, C, D, E, F, G, and H. In the first place, we may put either A, or B, or C, ..., or H so that there are 8 ways. If we pick A in the first place, we can choose B, or C, or D, ..., or H as the second place so that there are $8-1=7$ ways to choose once the first place is fixed. Continuing this reasoning, there are 6 ways for the third place once the first and the second places are fixed, and so on. As a result, there are $8 \cdot 7 \cdot 6 \cdots 2 \cdot 1 = 8!$ possible ways to put 8 different runners into a sequence.

Now suppose that we are only interested in the first, the second, and the third places (who gets gold, silver, and bronze medals) while we don't care about who gets the fourth, the fifth, the sixth, the seventh, and the eighth. In this case, we may compute the number of possible ways for picking the first, the second, and the third places as:

$$8 \cdot 7 \cdot 6 = \frac{8!}{5!}.$$

This is an example of permutation formula:

$$P_k^n = \frac{n!}{(n-k)!}.$$

If we are only interested in who gets a medal without worrying about the colour of medals (i.e., ordering does not matter), how can we compute the number of ways to pick 3 runners out of 8 runners? This can be computed by dividing $\frac{8!}{5!}$ by $3!$. To understand this, suppose that A, B, and C are placed within the third place so that they got medals. There are $3! = 6$ possible orders for who gets which medals among A, B, and C (i.e., (A,B,C), (A,C,B), (B,A,C), (B,C,A), (C,A,B), and (C,B,A)). When we do not care about the order, all of these 6 sequences are treated as the same. Therefore, the number of ways to choose 3 runners regardless of their orders is given by $(\frac{8!}{5!})/3!$. This is an example of the computation formula.

The number of (unordered) ways to choose k objects from n different objects can be computed as

$$C_k^n = \frac{n!}{k!(n-k)!}.$$

This is a combination of n objects take k at a time.

Example 2 *When the sample space is $S = \{A, B, C\}$, there are $n = 3$ objects. We have three possible unordered combinations of choosing $k = 2$ objects out of $n = 3$ objects: $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$. In fact, $C_2^3 = \frac{3!}{2!1!} = \frac{6}{2} = 3$.*

Example 3 *Suppose a personnel officer has 5 candidates to fill 2 positions. Among 5 candidates, 3 candidates are men and 2 candidates are women. If every candidate is equally likely to be chosen, what is the probability that no women will be hired?*

To be concrete, denote the five candidates by $(M_1, M_2, M_3, W_1, W_2)$, where (M_1, M_2, M_3) are men and (W_1, W_2) are women. Basic outcomes can be classified into three cases: (i) one man and one women, (M_i, W_j) for $i = 1, 2, 3$ and $j = 1, 2$, (ii) both are women, (W_1, W_2), and (iii) both are men, (M_1, M_2), (M_1, M_3), and (M_2, M_3).

We may count that there are $C_2^5 = \frac{5!}{2!(5-3)!} = 10$ possible combinations of choosing 2 candidates out of 5 candidates while there are $C_2^3 = 3$ cases in which both candidates are men so that the answer is $3/10$.

Conditional Probability

Definition 2 *The **conditional probability** of an event A given that event B has occurred is defined by*

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

provided that $P(B) > 0$*.*

Example 4 *Rolling two dice, what is the probability that at least one die is equal to 2 when the sum of two numbers is less than or equal to 3?*

We can formally follow the definition of the conditional probability. The sample space is all unordered combinations of two numbers and there are $6 \times 6 = 36$ *basic outcomes. Let the event* A *be "at least one die is equal to 2" and let the event* B *be "the sum of two numbers is less than or equal to 3." Then,* $A = \{(1,2),(2,1),(2,2),(2,3),...,(2,6),(6,2)\}$ *and* $B = \{(1,1),(1,2),(2,1)\}$*. Also,* $A \cap B = \{(1,2),(2,1)\}$*. Therefore,* $P(A \cap B) = 2/36$ *and* $P(B) = 3/36$ *and it follows from the definition of conditional probability that* $P(A|B) = P(A \cap B)/P(B) = 2/3$*.*

One way to think of conditional probability is to view the conditioning event B as the "new sample space" and calculate the probability that part of A contained in B. In this case, part of A contained in B is expressed as $A \cap B = \{(1,2),(2,1)\}$. Given the equal probability of each unordered pair, the probability of $A \cap B = \{(1,2),(2,1)\}$ contained in $B = \{(1,1),(1,2),(2,1)\}$ is 2/3 because there are two elements in $A \cap B$ while there are three elements in B.

By multiplying both sides in the definition of conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$, we have $P(A|B)P(B) = P(A \cap B)$.

Definition 3 *The probability of both* A *and* B *occur is given by the* ***multiplication rule***

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$

provided $P(A), P(B) > 0$*.*

Statistical Independence

Definition 4 *Events* A *and* B *are* ***statistically independent*** *if and only if*

$$P(A \cap B) = P(A)P(B).$$

In some cases, we have a good intuition for independence. When we draw a card from a randomly shuffled deck, the event of "drawing Ace" is independent of the event of drawing "Black." When we flip two coins, the event of "Head in the first coin" is independent of "Head in the second coin." In other cases, however, whether two events are independent or not is not so obvious—in order to check if two events are independent or not, we need to check the formal condition $P(A \cap B) = P(A)P(B)$.

Example 5 *Roll a fair die once, define* $A = \{2,4,6\}$ *and* $B = \{1,2,3,4\}$*. Are the events* A *and* B *independent? To answer this, we need to check the formal definition of statistical independence, i.e., whether* $P(A \cap B) = P(A)P(B)$ *holds or not. Because* $A \cap B = \{2,4\}$*, we have* $P(A \cap B) = 2/6 = 1/3$ *while* $P(A)P(B) = (1/2)(4/6) = 1/3$ *and, therefore,* $P(A \cap B) = P(A)P(B)$ *holds. That is,* A *and* B *are statistically independent.*

By multiplication rule, we can write $P(A \cap B) = P(B|A)P(A)$ if $P(A) > 0$. On the other hand, when A and B are independent, we have $P(A \cap B) = P(A)P(B)$. Comparing these two equations, we conclude that,

$$\text{if } A \text{ and } B \text{ are statistically independent, then } P(B|A) = P(B)$$

when $P(A) > 0$. This is intuitive: if A and B are independent, then knowing A happened will not provide any information regarding whether B happened or not.

Law of Total Probability

Let B_1, B_2, ..., B_k be mutually exclusive and exhaustive events so that

$$B_1 \cup B_2 \cup ... \cup B_k = S \quad \text{and} \quad B_i \cap B_j = \emptyset \quad \text{for } i \neq j.$$

Now, using B_1, B_2, ..., B_k, we may partition a set A into k mutually exclusive and exhaustive events as $(A \cap B_1)$, $(A \cap B_2)$, ... , and $(A \cap B_k)$ such that

$$A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k) \quad \text{and} \quad (A \cap B_i) \cap (A \cap B_j) = \emptyset \quad \text{for } i \neq j.$$

Therefore,

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k) = \sum_{i=1}^{k} P(A \cap B_i).$$

Substituting $P(A \cap B_i) = P(A|B_i)P(B_i)$ (multiplication rule) into the left side of the above equation, we have $P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$. This is called **Law of Total Probability**.

Theorem 1 (Law of Total Probability) *Suppose that B_1, B_2, ..., B_k be mutually exclusive and exhaustive events such that $P(B_i) > 0$ for $i = 1, ..., k$. Then,*

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i) \quad \text{for any event } A.$$

Bayes' Theorem

Consider two events A and B. By multiplication rule, we have

$$P(A \cap B) = P(A|B)P(B). \tag{1}$$

Because B and $\bar{B}$ are mutually exclusive and exhaustive, we have

$$(A \cap B) \cap (A \cap \bar{B}) = \emptyset \quad \text{and} \quad B = (A \cap B) \cup (A \cap \bar{B}).$$

Therefore,

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}), \tag{2}$$

where the second equality uses multiplication rule. Using (2)-(1), we may express the conditional probability of B given A as

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

The generalization of this to any partition of S gives the **Bayes theorem**.

Theorem 2 *Let B_1, B_2, ..., B_k be mutually exclusive and exhaustive events so that*

$$B_1 \cup B_2 \cup ... \cup B_k = S \quad \text{and} \quad B_i \cap B_j = \emptyset \quad \text{for } i \neq j.$$

Then, for any event B_i with $P(B_i) > 0$, we have

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

Table 1: Joint distribution of A and B in %

	$\bar{B}$	B	Marginal Prob.
$\bar{A}$	89.1	0.05	89.15
A	9.9	0.95	10.85
Marginal Prob	99.0	1.0	1.00

Proof: By **Law of Total Probability**, we have

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i). \tag{3}$$

Substituting $P(A \cap B_i) = P(A|B_i)P(B_i)$ and $P(A)$ in equation (3) to the right hand side of $P(B_i|A) = \frac{P(A \cap B_i)}{P(A)}$ (the definition of conditional probability), we have

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

■

Example 6 *Suppose you took a blood test for cancer diagnosis and your blood test was positive. This blood test is positive with probability 95 percent if you indeed have a cancer. This blood test is positive with probability 10 percent if you dont have a cancer. In population, it is known that 1 percent of people have cancer. What is the probability of your having a cancer given that your blood test is positive?*

Define $A = \{$positive blood test$\}$ and $B = \{$having a cancer$\}$. Then,

$$P(A|B) = 0.95, \quad P(A|\bar{B}) = 0.10, \quad P(B) = 0.1.$$

Note also that $P(\bar{B}) = 1 - P(B) = 0.9$.

By apply the Bayes' theorem, we can compute the probability of having a cancer (B) given that the blood test is positive (A) as

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} \\ &= \frac{0.95 \times 0.1}{0.95 \times 0.1 + 0.1 \times 0.9} = 0.0876. \end{aligned}$$

With the information given, we may compute the joint and the marginal distribution of A and B as in Table 6. These numbers are computed as follows. $P(B) = 0.99$ and $P(\bar{B}) = 0.01$ gives the marginal probability of B and $\bar{B}$. Looking at the column of B, we can compute $P(A \cap B)$ as $P(A|B)P(B) = 0.95 \times 0.01 = 0.0095$ (0.95 percent) and $P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.01 - 0.0095 = 0.0005$ (0.05 percent). Similarly, looking at the column of $\bar{B}$, we can compute $P(A \cap \bar{B}) = P(A|\bar{B})P(\bar{B}) = 0.10 \times 0.99 = 0.099$ (9.9 percent) and $P(\bar{A} \cap \bar{B}) = P(\bar{B}) - P(A \cap \bar{B}) = 0.99 - 0.099 = 0.891$ (89.1 percent).