SEPTEMBER 15, 2014

LECTURE 1
BASICS OF PROBABILITY

1 Uncertainty or why we need probability and statistics

Many important economic models are concerned with decision making under uncertainty. For example:

- The investor's decision to buy or sell stocks involves the uncertainty about their future prices.
- Airlines' decision whether or not to open a new route involves the uncertainty about the number passengers, the cost of fuel and aircraft maintenance.
- The investor's decision to start production of a new product involves the uncertainty about the product's market price and the cost of inputs.
- A bidder's bid in a first-price sealed-bid auction involves the uncertainty about other bidders' valuations, and often even his own valuation.
- The central bank's decision to lower or raise the interest rate involves the uncertainty about future inflation and unemployment.

We say that an experiment is *deterministic* if its outcome is unique or uniquely determined by the conditions of the experiment. An experiment is *indeterministic* or *random* if its outcome cannot be uniquely determined even if the experiment repeated under the same conditions. There are several sources of uncertainty or randomness. It can arise due to **incomplete information.** Consider flipping a coin experiment. Without knowing the position, velocity and rotation of a coin it would be impossible to predict the outcome of the experiment with certainty even for someone with a very advanced knowledge of physics. Another source of uncertainty is **complexity of the relationships between variables and/or states.** Even with the knowledge of the initial conditions such as position, velocity, and etc, a typical person does not know enough physics to predict the outcome of flipping a coin experiment with certainty.

If the sources of uncertainty were limited only to the two described above, at least theoretically randomness could be eliminated by providing a complete description of an experiment. For example, it was demonstrated that coin flipping can be accurately modeled using the laws of mechanics and is highly predictable under controlled initial conditions. Nevertheless, scientists in different disciplines convinced that there are **fundamentally indeterministic** phenomena such that their randomness cannot be resolved by supplying more information. One prominent example is quantum indeterminacy in physics. In neurosciences, it was found that neurons (cells that are responsible for processing and transmission of information) discharge irregularly adding a *noise* to the information processed by the brain. Thus human decision making is affected by the noise and, as a result, random to some extent.

Instead of trying to find a certain (deterministic) prediction for an experiment, a probability model provides its probabilistic description by characterizing the likeliness or *probabilities* (or *probability distribution*) of potential outcomes. The properties of probability models are governed by strict mathematical laws; those laws are the subject of *Probability Theory.*

The knowledge of probability distributions and the laws of probability are crucial for making optimal decisions in the face of uncertainty. However, the probability distribution of any random variable is typically unknown and has to be evaluated (or estimated) *empirically* using data. Estimation of probability distributions from is the subject of *Statistics*.

The basic components of a probability model are *sample spaces*, *events*, and *probability functions*. They are discussed below.

2 Sample spaces and set theory

Definition 1. Sample space is a *set* (collection or list) of *all* possible *outcomes* for a random experiment.

The sample space is often denoted by Ω, and its elements or outcomes are often denoted by ω. In any random experiment, one and only one $\omega \in \Omega$ is selected. For example:

- In the coin flipping experiment: $\Omega = \{H, T\}$, where H stands for heads and T stands for tails.
- Flipping a coin twice: $\Omega = \{HH, TT, HT, TH\}$.
- Rolling a dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Predicting tomorrow's weather: $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$.
- Temperature measurement in °C: $\Omega = [-273.15, \infty)$.
- AAPL price tomorrow: $\Omega = \mathbb{R}_+$.
- Throwing a dart at a 1 meter × 1 meter board: $\Omega = [0, 1] \times [0, 1]$. In this example, $\omega = (\omega_1, \omega_2)$, where ω_1 are the horizontal and vertical coordinates respectively.

Definition 2. (a) A is a *subset* of Ω, denoted as $A \subset \Omega$, if $\omega \in A$ implies that $\omega \in \Omega$.

(b) An *event* is any subset of the sample space.

We say that an event A occurred if the outcome of the random experiment is some $\omega \in A$. If the outcome of the random experiment is some $\omega \notin A$, we say that the event A did not occur. Since $\Omega \subset \Omega$, the sample space is also a (trivial) event: its meaning is that something happened. For example:

- In the coin flipping experiment, we can define the following events $A_1 = \{H\}$ (heads), $A_2 = \{T\}$ (tails), $A_3 = \{H, T\}$ (heads or tails), $A_4 = \emptyset$ (an empty set containing no elements: the coin hangs in the air).
- Flipping a coin twice: $A_1 = \{HH\}$ (observing two heads), $A_2 = \{TT\}$ (observing two tails), $A_3 = \{HH, TT\}$ (observing the same side in two tosses), etc. Note that $A_4 = \{H\}$ is not a valid event in this case, since $\{H\}$ is not a subset of $\Omega = \{HH, TT, HT, TH\}$.
- Rolling a dice: $A_1 = \{1, 2, 3\}$ (a number less than four rolled), $A_2 = \{1, 3, 5\}$ (an odd number is rolled), $A_3 = \{6\}$ (six is rolled), and etc.
- Tomorrow's weather: $A_1 = \{\text{rain}, \text{snow}\}$ (not a clear weather), $A_2 = \{\text{rain}, \text{clear}\}$ (no snow), and etc.
- AAPL price tomorrow: $A_1 = \{97\}$ (the price is exactly \$97), $A_2 = [97, \infty)$ (the price is at least \$97), $A_3 = [97, 99]$ (the price is between \$97 and \$99).

The basic operations on sets are defined below.

Definition 3. (a) **Union** (A **or** B): $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$.
(b) **Intersection** (A **and** B): $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$.
(c) **Complement** (**not** A)**:** $A^c = \{\omega \in \Omega : \omega \notin A\}$.

The union of A and B is an event, which occurs when A **or** B have occurred. In the experiment with two flips of a coin, if $A_1 = \{HH\}$ and $A_2 = \{TT\}$, then $A_3 = A_1 \cup A_2 = \{HH, TT\}$. The intersection of A and B is an event, which occurs when both A **and** B have occurred. In the AAPL example, if $B_1 = [0, 101)$ and $B_2 = [98, \infty)$, then $B_1 \cap B_2 = [98, 101)$. The complement of A is an event that occurs if A does **not** occur. In the AAPL example, if $A = [95, 101)$, then $A^c = [0, 95) \cup [101, \infty)$. Note that $\Omega^c = \emptyset$.

When two events cannot occur simultaneously, we say that they are mutually exclusive.

Definition 4. Events A and B are said to be *mutually exclusive* if $A \cap B = \emptyset$.

If A and B are mutually exclusive and we know that A occurred, we can conclude that B did not occur.

Some properties of operations on sets are stated in the following theorem.

Theorem 5. *(a) Commutative:* $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$.
(b) Associative: $(A \cup B) \cup C = A \cup (B \cup C)$ *and* $(A \cap B) \cap C = A \cap (B \cap C)$.
(c) Distributive: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ *and* $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.
(d) If $A \subset B$ *then* $A \cup B = B$ *and* $A \cap B = A$.
(e) De Morgan's laws: $(A \cup B)^c = A^c \cap B^c$ *and* $(A \cap B)^c = A^c \cup B^c$.

Proof. We will prove here only some parts of the theorem. Note that two sets S_1 and S_2 are equal if and only if every element of S_1 is also an element of S_2 ($S_1 \subset S_2$) and vice versa ($S_2 \subset S_1$).

To prove the first result in part (d), note that by the definition of the union operation, $B \subset (A \cup B)$. Thus, it suffices to show that $(A \cup B) \subset B$. The last statement can be easily shown by contradiction. Suppose not: i.e. there is $\omega \in (A \cup B)$ such that $\omega \notin B$. Since $(A \cup B)$ consists of the elements of A and B, it must be true that $\omega \in A$. Thus, there is ω such that $\omega \in A$ and $\omega \notin B$. Hence, $A \not\subset B$, which is a contradiction to the assumption made in part (d). We conclude that there is no ω such that $\omega \in (A \cup B)$ and $\omega \notin B$, and consequently $(A \cup B) \subset B$.

To prove the second statement in part (c), suppose that $\omega \in ((A \cap B) \cup C)$. By the definition of the union, $\omega \in (A \cap B)$ or $\omega \in C$. If $\omega \in C$ then $\omega \in (A \cup C)$ and $\omega \in (B \cup C)$, and therefore $\omega \in (A \cup C) \cap (B \cup C)$. On the other hand, if $\omega \in (A \cap B)$ then, by the definition of the intersection, $\omega \in A$ and $\omega \in B$. We obtain again that $\omega \in (A \cup C)$ and $\omega \in (B \cup C)$, and therefore $\omega \in (A \cup C) \cap (B \cup C)$. Hence, $((A \cap B) \cup C) \subset ((A \cup C) \cap (B \cup C))$. Now, suppose there is $\omega \in ((A \cup C) \cap (B \cup C))$ that is not an element of $(A \cap B) \cup C$. From the definition of the union and $\omega \notin ((A \cap B) \cup C)$, it follows that $\omega \notin C$. Next, $\omega \notin C$ and $\omega \in ((A \cup C) \cap (B \cup C))$ together imply that $\omega \in (A \cap B)$. On the other hand, $\omega \notin ((A \cap B) \cup C)$ also implies that $\omega \notin (A \cap B)$. We arrived at a contradiction.

To prove the second statement in part (e), suppose that $\omega \in (A \cap B)^c$ and therefore $\omega \notin (A \cap B)$. By the definition of the intersection, we have that either $\omega \notin A$ or $\omega \notin B$ (or both). In the first case, $\omega \in A^c$. In the second case, $\omega \in B^c$. In either case, $\omega \in (A^c \cup B^c)$. Hence, $(A \cap B)^c \subset (A^c \cup B^c)$. Now, suppose there is $\omega \in (A^c \cup B^c)$ such that $\omega \notin (A \cap B)^c$. Then, $\omega \in (A \cap B)$ and, therefore, $\omega \in A$ and $\omega \in B$. Consequently, $\omega \notin A^c$ and $\omega \notin B^c$, and therefore, $\omega \notin (A^c \cup B^c)$, which is a contradiction. □

The union and intersection operations can be extended to more than two sets:

$$\begin{aligned} A_1 \cup A_2 \cup \ldots \cup A_n &= \cup_{i=1}^n A_i, \\ A_1 \cap A_2 \cap \ldots \cap A_n &= \cap_{i=1}^n A_i. \end{aligned}$$

Moreover, if $\{A_i : i = 1, 2, \ldots\}$ is a *countably* infinite sequence of sets,

$$\begin{aligned} A_1 \cup A_2 \cup \ldots &= \cup_{i=1}^\infty A_i, \\ A_1 \cap A_2 \cap \ldots &= \cap_{i=1}^\infty A_i. \end{aligned}$$

In the first case, $\omega \in \cup_{i=1}^\infty A_i$ if $\omega \in A_i$ for some $i \in \{1, 2, \ldots\}$. In the second case, $\omega \in \cap_{i=1}^\infty A_i$ if $\omega \in A_i$ for all $i = 1, 2, \ldots$.

3 Probability function

Probability measures the likeliness of the occurrence of an event[1] on the scale $[0, 1]$, where zero means that the event never occurs, and one means that the event always occurs. There are two main approaches to interpreting probabilities: frequentist (objective) and subjective. According to the frequentist approach, the probability of an event is the relative frequency of occurrence of the event when the experiment is repeated a "large" number of times. The problem with this definition is that it does not state how many times exactly the experiment must be repeated. Also, there are experiments that cannot be repeated, which would prevent us from applying the definition. According to the subjective approach, the probability of an event is ascribed by one's knowledge and/or beliefs. For example, when one argues that the probability of seeing heads in a coin flipping experiment is $1/2$, the statement is based on his beliefs and knowledge rather than on actual measurements.

Regardless of the interpretation, the mathematical definition of probability is the same.

Definition 6. Given a collection of events defined on the sample space Ω, probability P is a *function*, which assigns a number to every event in the collection according to the following rules (*axioms of probability*):

A1. $P(A) \geq 0$ for every event A.

A2. $P(\Omega) = 1$.

A3. Let $\{A_i : i = 1, 2, \ldots\}$ be a countable collection of mutually exclusive events, i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.[2]

Remark. Here we are trying to avoid some technical details that are beyond the scope of this course and intentionally vague about the *collection* of events. When the sample space is discrete as in the coin flipping, dice rolling and weather examples, this collection can be taken as the collection of all possible subsets of the sample space, i.e. it will contain every possible event on Ω. However, when the sample space is continuous (uncountable) as in the AAPL, temperature, and dart throwing examples, one cannot assign probabilities to every possible event without violating the axioms of probability. In such cases, the collection of events to which probabilities are assigned has to be restricted. For example when $\Omega = [0, 1]$, we can restrict our attention to the collection of all intervals contained in $[0, 1]$ and their unions.

The intuition behind the axioms of probability is very natural. According to A1, the measures of likeliness of an event must be positive. One represents certainty according to Axiom A2, which acts as a normalization. Lastly according to A3, if events cannot occur simultaneously, then the probability of at least one of them occurring is equal to the sum of their individual probabilities. In other words, the probability of a composite event is equal

[1] Hence, probabilities are assigned to events and not outcomes.

[2] Here, $\sum_{i=1}^{n} a_i$ denotes the sum: $\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$.

to the sum of the probabilities of its (mutually exclusive) components. For example, suppose that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. Suppose further that $P(\{\omega_i\}) = 1/n$ for all $i = 1, \ldots, n$. For an event A defined on Ω, let $N(A)$ be the number of ω's contained in A. Then, according to axiom A3, $P(A) = N(A)/n$.

Some of the properties that follow from the axioms of probability as stated in the following theorem.

Theorem 7. *(a)* $P(A^c) = 1 - P(A)$.

(b) $P(\emptyset) = 0$.

(c) $P(A) \leq 1$.

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

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

(f) If $A \subset B$ *then* $P(A) \leq P(B)$.

(g) Boole's Inequality: $P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$.

(h) Bonferroni's Inequality: $P(A \cap B) \geq P(A) + P(B) - 1$.

(i) $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.

Proof. (a) Write $\Omega = A \cup A^c$. By definition, $A \cap A^c = \emptyset$, and therefore by Axiom A3, $P(\Omega) = P(A) + P(A^c)$. By Axiom A2, $1 = P(A) + P(A^c)$, and the result follows.

(b) Since $\emptyset = \Omega^c$, $P(\emptyset) = 1 - P(\Omega) = 0$ by Axiom A2.

(c) Omitted

(d) First, $B = B \cap \Omega = B \cap (A \cup A^c)$. By Theorem 5(c), $B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$. Since ω cannot be simultaneously in A and A^c, the events $B \cap A$ and $B \cap A^c$ are mutually exclusive, and therefore by Axiom A3, $P(B) = P(B \cap A) + P(B \cap A^c)$. Note that in this exercise, $\{A, A^c\}$ is a *partition* of the sample space, since $A \cup A^c = \Omega$ and the parts are mutually exclusive.

(e) We will show first that $A \cup B = A \cup (B \cap A^c)$:

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

where the first equality follows by Theorem 5(a) and (c), the second equality follows by the definition of the complement, and the last equality follows by 5(d). Next, note that A and $B \cap A^c$ are mutually exclusive. Hence, $P(A \cup B) = P(A) + P(B \cap A^c)$. From part (d) we have $P(B \cap A^c) = P(B) - P(B \cap A)$, and the result follows.

(f) Omitted.

(g) Note that if the events are mutually exclusive, the result holds as an equality by Axiom A3. In general case of non-mutually exclusive events, the proof is by induction. Suppose that $n = 2$. Then,

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ &\leq P(A_1) + P(A_2), \end{aligned}$$

where the inequality in the second line holds because $P(A_1 \cap A_2) \geq 0$ by Axiom A1. Next, suppose $n = 3$. Then, by the associative property (Theorem 5(b)),

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P((A_1 \cup A_2) \cup A_3) \\ &\leq P(A_1 \cup A_2) + P(A_3) \\ &\leq P(A_1) + P(A_2) + P(A_3), \end{aligned}$$

where the inequalities in the second and third lines hold by the result for $n = 2$. Now, suppose that the result holds for $(n - 1)$:

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

We need to show that it continues to hold for n. Using the same approach as in the case of $n = 3$,

$$\begin{aligned} P(\cup_{i=1}^{n} A_i) &= P((\cup_{i=1}^{n-1} A_i) \cup A_n) \\ &\leq P(\cup_{i=1}^{n-1} A_i) + P(A_n) \\ &\leq \sum_{i=1}^{n-1} P(A_i) + P(A_n) \\ &= \sum_{i=1}^{n} P(A_i), \end{aligned}$$

where the inequality in the second line holds by the result for $n = 2$, and the inequality in the third line holds by (1).

(h) Omitted.

(i) Omitted. □

Example. Suppose that in a typical year in city V, it rains on 70% of the days, snows on 20% of the days, and rains and snows on 10% of the days. If you pick a day at random, what is the probability that it won't rain or snow? Let R and S be the events denoting rain and snow respectively. Then the event $R \cup S$ represents rain or snow. From Theorem 7(e)

$P(R \cup S) = P(R) + P(S) - P(R \cap S) = 0.70 + 0.20 - 0.10 = 0.80$. Hence, the probability of no rain or snow on a randomly chosen day is 20%.

4 Counting techniques: permutations and combinations

When the sample space Ω has finitely many outcomes that are equally likely to be selected, the probability of an event A can be easily determined by counting the number of elements in A and dividing it by the total number of elements in Ω:

$$P(A) = \frac{N(A)}{N(\Omega)}.$$

Example. Suppose that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{10}\}$, and each ω is equally likely to be selected Let $R = \{\omega_1, \omega_2, \ldots, \omega_7\}$ and $S = \{\omega_7, \omega_8\}$. Then $P(R) = 0.7$, $P(S) = 0.2$, and $P(R \cap S) = P(\{\omega_7\}) = 0.1$. Lastly, $P((R \cup S)^c) = P(\{\omega_{9,}\omega_{10}\}) = 0.2$.

In more complicated examples, the following techniques can be useful.

Definition 8. (Multiplication Principle) Suppose that procedure E_1 has n_1 possible outcomes and procedure E_2 has n_2 possible outcomes. Then an experiment consisting of performing E_1 first and E_2 next has $n_1 \times n_2$ possible outcomes.

Example. Suppose we first flip a coin and then roll a dice. The number of possible outcomes is $2 \times 6 = 12$: $\Omega = \{H1, H2, \ldots, H6, T1, T2, \ldots, T6\}$.

By applying the multiplication principle repeatedly, we can extend this to experiments with more than two procedures.

Example. Suppose we first flip a coin three times and then roll a dice twice. The number of possible outcomes is $2 \times 2 \times 2 \times 6 \times 6 = 288$.

Sometimes experiments involve selecting (without replacement) r objects from a set of n objects and arranging them in a particular order.

Definition 9. An ordered arrangement of r distinct elements selected from a set of n distinct elements, where $n \geq r$, is called a *permutation* and denoted $P(n, r)$.

Theorem 10. $P(n, r) = n!/(n - r)!$

Proof. For the first position, we can choose any of n elements. Since elements are not placed back into the set (selection without replacement), for the second position we can choose from $n-1$ remaining elements and etc. For the r-th position, we can choose from $n-(r-1) = n-r+1$

elements. Hence,[3]

$$P(n,r) = n \times (n-1) \times \ldots \times (n-r+1) = n!/(n-r)!$$

□

Example. (Ice Hockey) (a) Suppose that a hockey team has 12 forwards, 6 defenceman, and two goaltenders. There are 3 forward positions (L, C, R), two defense positions (L and R), and one position in the goal net. In how many different ways a complete line of players can be selected? Answer: Applying the result in Theorem 10 together with the multiplication principle, we obtain: $P(12,3) \times P(6,2) \times P(2,1) = (12!/9!) \times (6!/4!) \times (2!/1!) = 79,200$.

(b) Higgins is a forward and Bieksa is a defenceman playing on the same team that has a composition as described in (a). If the lineup for a shift is selected randomly (but following the rules in (a)), what is the probability that Higgins and Bieksa will play in the same shift? Answer: The number of ways to select the line of forwards with Higgins is given by $P(11,2) \times 3 = 330$. The number of ways to select defencemen with Bieksa is $P(5,1) \times 2 = 10$. Thus, the total number of lineup permutations with Higgins and Bieksa on the ice at the same time is $330 \times 10 \times 2 = 6,600$. Hence, the probability that Higgins and Bieksa will play in the same shift is $6600/79200 \approx 0.083333$.

(c) Continuing with the example in (a) and (b), what is the probability that Higgins and Bieksa will both play L on the same shift? Answer: $P(11,2) \times P(5,1) \times 2/79200 = 110 \times 5 \times 2/79200 \approx 0.013889$.

When the order in which elements are selected from a set is not important, we have combinations.

Definition 11. A selection of r distinct elements from a set of n distinct elements, where $n \geq r$, is called a *combination* and denoted $C(n,r)$.

Theorem 12.

$$C(n,r) = \frac{n!}{(n-r)!r!} \equiv \binom{n}{r}.$$

Proof. Note that once r elements have been selected into a combination, there will be exactly $P(r,r) = r!$ different ways to arrange the chosen elements into permutations, i.e. for every combination of r elements there are $r!$ permutations:

$$P(n,r) = C(n,r) \times P(r,r) = C(n,r) \times r!$$

□

[3] Recall that for a non-negative integer n its *factorial* is defined as $n! = n \times (n-1) \times \ldots \times 1$. Also, we define $0! = 1$.

Example. Continuing with the ice hockey example, if we ignore L, C, and R positions for forwards and L and R positions for defencemen, the number of ways to select a complete lineup is $C(12,3) \times C(6,2) \times C(2,1) = 6,600$. The number of lineups that include Higgins and Bieksa is given by $C(11,2) \times C(5,1) \times C(2,1) = 550$. Hence, the probability that Higgins and Bieksa will be in the same randomly selected shift is $550/6600 \approx 0.083333$, which is the same anwer as in part (b) above.