Theorems & Definitions

Akiko Iwamizu

8/25/2021

Contents

1 Probability **1**

2 Random Variables **3**

1 Probability

Definition 1.1 (Event Space)**.** A set S of subsets of Ω is an *event space* if it satisfies the following:

- *Nonempty*: $S \neq \emptyset$.
- *Closed under complements*: if $A \in S$, then $A^c \in S$.
- *Closed under countable unions*: if A_1, A_2, A_3,... $\in S$, then $A_1 \cup A_2 \cup A_3 \cup \ldots \in S$.

Definition 1.2 (Kolmogorov Axioms)**.** Let Ω be a *sample space*, S be an *event space*, and P be a *probability measure*. Then (Ω,S,P) is a *probability space* if it satifies the following:

- *Non-negativity*: $\forall A \in S$, $P(A) \geq 0$ where $P(A)$ is finite and real.
- *Unitary*: $P(\Omega) = 1$.
- *Countable additivity*: if A_1, A_2, A_3,... $\in S$ are pairwise disjoint, then

$$P(A_1 \cup A_2 \cup A_3 \cup ...) = P(A_1) + P(A_2) + P(A_3) + ... = \sum P(A_i).$$

The intuition behind these axioms is as follow: The first axiom states that the probability of any event is a non-negative number; there cannot be a less-than-zero chance of an event occuring. The second axiom states that the probability measure of the entire sample space is one. In other words, it is certain that some outcome will occur. Finally, the third axiom states that, given any number of *mutually exclusive events*, the probability that one of those events will occur is the sum of their individual probabilities.

Definition 1.3 (Pairwise Disjoint)**.** Recall that sets A and B are disjoint if $A \cap B = \emptyset$. We say that A_1, A_2, A_3,... are pairwise disjoint if each of them is disjoint from every other, that is, $\forall i \neq j, A_i \cap A_j = \emptyset$.

Theorem 1.1 (Basic Properties of Probability)**.** *Let (Ω,S,P) be a probability space. Then*

- Monotonicity*: $\forall A, B \in S$, if $A \subseteq B$, then $P(A) \leq P(B)$.*
  - *Monotonicity implies that, if one event is a subset of another (so that the former always occurs whenever the latter does), then the probability of the former occurring is no greather than that of the latter.*
- Subtraction rule*: $\forall A, B \in S$, if $A \subseteq B$, then $P(B \setminus A) = P(B) - P(A)$.*
  - *The subtraction rule implies that the probability that the second event occurs but not the first is equal to the probability of the second event minus the probability of the first event.*

- Zero probability of the empty set*:* $P(\emptyset) = 0$*.*
  - *Zero probability of the empty set means that some event in our event space must occur, and probability bounds mean that each of these events has some probability of occuring between zero and one.*
- Probability bounds*:* $\forall A \in S,\ 0 \leq P(A) \leq 1$*.*
  - *Monotonicity and unitarity (and non-negativity) imply the probability bounds since* $A \subseteq \Omega$*.*
- Complement rule*:* $\forall A \in S,\ P(A^c) = 1 - P(A)$*.*
  - *The complement rule implies that the probability of any of these events not occurring is one minus the probability of the event occurring - so that the probability that a given event either occurs or does not occur is one.*

Definition 1.4 (Joint Probability)**.** For $A, B \in S$, the *joint probability* of A and B is $P(A \cap B)$.

In other words, the joint probability of two events A and B is the probability of the intersection of A and B (which is itself an event in S), that is, the set of all states of the world in which both A and B occur.

Theorem 1.2 (Addition Rule)**.** *For* $A, B \in S$*,*

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

Definition 1.5 (Conditional Probability)**.** For $A, B \in S$ with $P(B) > 0$, the *conditional probability* of A given B is

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

Theorem 1.3 (Multiplicative Law of Probability)**.** *For* $A, B \in S$ *with* $P(B) > 0$*,*

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

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

Theorem 1.4 (Bayes' Rule)**.** *For* $A, B \in S$ *with* $P(A) > 0$ *and* $P(B) > 0$*,*

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

Definition 1.6 (Partition Rule 1)**.** If $A_1,\ A_2,\ A_3, \ldots \in S$ are nonempty and pairwise disjoint, and $\Omega = A_1 \cup A_2 \cup A_3 \cup ...$, then $\{A_1,\ A_2,\ A_3, \ldots\}$ is a *partition* of Ω.

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

Definition 1.7 (Partition Rule 2)**.** If $A_1,\ A_2,\ A_3, \ldots \in S$ are nonempty and pairwise disjoint, and $\Omega = A_1 \cup A_2 \cup A_3 \cup ...$, then $\{A_1,\ A_2,\ A_3, \ldots\}$ is a *partition* of Ω.

$$A \cup B = [(A \setminus B) \cup (A \cap B)] \cup [(A \cap B) \cup (B \setminus A)] = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

Theorem 1.5 (Law of Total Probability)**.** *If* $\{A_1, A_2, A_3, \ldots\}$ *is a* partition *of* Ω *and* $B \in S$*, then*

$$P(B) = \sum P(B \cap A_i).$$

If we also have $P(A_i) > 0$ *for* $i = 1, 2, 3, \ldots$*, then this can also be stated as*

$$P(B) = \sum P(B \mid A_i)P(A_i).$$

Theorem 1.6 (Alternative Forms of Bayes' Rule)**.** *If* $\{A_1, A_2, A_3, \ldots\}$ *is a* partition *of* Ω *with* $P(A_i) > 0$ *for* $i = 1, 2, 3, \ldots$*, and* $B \in S$*, then apply the Law of Total Probability to the denominator to get*

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum P(B \cap A_i)}.$$
$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum P(B \mid A_i)P(A_i)}.$$
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}.$$

Definition 1.8 (Independence of Events)**.** Events $A, B \in S$ are *independent* if

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

Theorem 1.7 (Conditional Probability and Independence)**.** *For* $A, B \in S$ *with* $P(B) > 0$*,* A *and* B *are independent if and only if*

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

2 Random Variables

Definition 2.1 (Random Variable)**.** A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that

$$\forall r \in \mathbb{R}, \{\omega \in \Omega : X(\omega) \leq r\} \in S.$$

Where each $\omega \in \Omega$ denotes a state of the world, which may be represented by anything: numbers, letters, words, etc. to describe all the distinct possible outcomes that could occur. A *random variable* maps each of these states of the world to a real number. Thus, it is often remarked that, a *random variable* is neither random nor a variable, as it is merely a *function.* So when the state of the world is ω, the random variable takes on the value $X(\omega)$. For example, the event $\{X = 1\}$ should be understood to mean the set of states $\{\omega \in \Omega : X(\omega) = 1\}$.

Definition 2.2 (Function of a Random Variable)**.** Let $g : U \to \mathbb{R}$ be a function, where $X(\Omega) \subseteq U \subseteq \mathbb{R}$. Then, if $g \circ X : \Omega \to \mathbb{R}$ is a random variable, we say that g is a *function of* X, and write $g(X)$ to denote the randome variable $g \circ X$. This general definition allows us to formally work with transformations of random variables as random variables in their own right.

Definition 2.3 (Operator on a Random Variable)**.** An *operator* A on a random variable maps the function $X(\cdot)$ to a real number, denoted by $A[X]$.

Example 2.1 (Example of a Random Variable)**.** We can define events in S in terms of a random variable X. For example we could let

$$A = \{\omega \in \Omega : X(\omega) = 1\}.B = \{\omega \in \Omega : X(\omega) \geq 0\}.C = \{\omega \in \Omega : X(\omega)^2 \leq 10, X(\omega) \neq 3\}. \tag{1}$$