ESE 520 Probability and Stochastic Processes

Lecture 2

"Probability space $(\Omega, \mathcal{F}, P)$: continued".

Let us start with a simple example of a probability measure on an arbitrary sample space Ω.

Example 1. Let Ω be any non-empty set (not necessarily a set of numbers) and let $w_0 \in \Omega$ be a fixed point. For any subset A from Ω, we define

$$\mu(A) := \begin{cases} 1, & w_0 \in A \\ 0, & w_0 \notin A. \end{cases}$$

By definition, $\mu : \mathcal{P}^{\Omega} \to \{0, 1\} \subset \mathbb{R}$ is a set function defined on the largest σ-algebra on Ω, its power σ-algebra.

Let us show that μ is a probability measure (such a measure is often called an *atom measure* since it is concentrated only at one point).

Solution: By definition, we have to show three properties:

1) For any subset A: $\mu(A) \geq 0$ which is trivial;

2) $\mu(\Omega) = 1$ which is also trivially satisfied;

3) σ-additivity of μ: we choose any sequence of pairwise disjoint subsets $A_1, A_2, ...$ from Ω and have to show that

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i). \tag{1}$$

There are two cases here: if $w_0 \in \cup_{i=1}^{\infty} A_i$, then there is m such that $w_0 \in A_m$ and $w_0 \notin A_n$ for any $n \neq m$. It follows then that $\mu(A_m) = 1$ and $\mu(A_n) = 0$ for all $n \neq m$. Therefore, $\mu(\cup_{i=1}^{\infty} A_i) = 1$ and $\sum_{i=1}^{\infty} \mu(A_i) = 1$ so that (1) is true.

The second case is when $w_0 \notin \cup_{i=1}^{\infty} A_i$. Then $w_0 \notin A_i$ for all $i = 1, 2, ...$ so that $\mu(\cup_{i=1}^{\infty} A_i) = 0$ and $\sum_{i=1}^{\infty} \mu(A_i) = 0$ yielding (1).

Example 2. Let P be a probability measure defined on a measurable space $(\Omega, \mathcal{F})$. Show that, for any sequence $A_i, i = 1, 2, ..., n$ from $\mathcal{F}$, it holds

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

Solution: By the addition formula,

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

which implies that

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$

since $P(A_1 \cap A_2) \geq 0$.

The general statement follows then by the mathematical induction argument:

$$P(\cup_{i=1}^{n} A_i) = P(\cup_{i=1}^{n-1} A_i + A_n) \leq P(\cup_{i=1}^{n-1} A_i) + P(A_n) \leq$$

$$\sum_{i=1}^{n-1} P(A_i) + P(A_n) = \sum_{i=1}^{n} P(A_i).$$

$\circ$ In applications, we often are interested in finding the probability of an event B given we know that a particular event A has occured. This leads to the concept of *conditional probability* of B given A.

Definition 1. Assume $A, B \in \mathcal{F}$ and $P(A) > 0$. Then, the conditional probability of the event B given A is defined to be the number

$$P(B|A) := \frac{P(A \cap B)}{P(A)}. \tag{2}$$

Clearly, if $A = \Omega$, then $P(B|\Omega) = P(B)$ so that the conditional probability coincides with the unconditional one and, in general, $P(\bullet|A)$ is nothing but a generalization of the unconditional probability. The function $P(\bullet|A)$ preserves then all the properties of the function $P(\bullet)$.

Another very important concept in probability theory is the concept of independence.

Definition 2. Assume that $A \cap B \neq \emptyset$. Then, the events A and B are said to be *independent* if and only if

$$P(A \cap B) = P(A)P(B). \tag{3}$$

It is important to notice that the concept of two independent events should not be confused with the concept of two disjoint events as for two events to be independent or not, they must have the chance to occur at the same time!

When calculating conditional probabilities, the following two simple formulas might be useful.

Definition 3. A sequence of events $A_1, A_2, ..., A_n$ is called a *partition* of Ω if

i) $A_i \cap A_j = \emptyset$ for all $i \neq j$ (pairwise disjoint events);

ii) $\cup_{i=1}^n A_i = \Omega$.

Then, for any event B, it holds

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{4}$$

called the *formula of total probability*, and

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \tag{5}$$

for any $j = 1, 2, ..., n$, called the *Bayes' formula.*

The proof of (4) and (5) is elementary (exercise).

Example 4. If a patient has a particular disease, a test performed has 99% accuracy. If he does not have the disease, then the accuracy is 95%. Assume that 1 in 10,000 has the disease and find how good the test is if a patient has disease.

Solution: Let A="patient has disease" and "$+$"="test is positive". Using the Bayes formula and formula of total probability, we find:

$$P(A|+) = \frac{P(+|A)P(A)}{P(+|A)P(A) + P(+|A^c)P(A^c)} =$$

$$\frac{(.99)(10^{-5})}{(.99)(10^{-5}) + (.05)(1 - 10^{-5})} = .002.$$

∘ The independence of events and later of random variables is often convenient to characterize in terms of the corresponding σ-algebras generated by the events or random variables. The reason for that lies, in particular, in the fact that σ-algebras are interpreted as the *amounts of information* generated by the events or random variables (collection of events that can happen).

Definition 4. Two σ-algebras $\mathcal{G}$ and $\mathcal{F}$ defined on Ω are said to be *independent* if and only if

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

for all $A \in \mathcal{G}$ and $B \in \mathcal{F}$.

∘ The above definition can be then generalized accordingly for more than two σ-algebras.

∘ We recall: For any $A \in \mathcal{F}$, the σ-algebra $\mathcal{F}_A := \{\emptyset, \Omega, A, A^c\}$ is the smallest σ-algebra containing A and often is called the σ-algebra *generated by the event* A.

Proposition. Two events A and B are independent if and only if the σ-algebras $\mathcal{F}_A$ and $\mathcal{F}_B$ are independent.

Indeed:

$$P(A^c)P(B) = [1-P(A)]P(B) = P(B)-P(A)P(B) = P(B)-P(A\cap B) =$$

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

The proof of independence for other pairs (A, B^c) and (A^c, B^c) is similar.

There are several useful facts about σ-algebras but we mention here additionally only one that should help to understand the construction procedure of probability measures on σ-algebras.

Proposition. Let $\mathcal{A}$ by any collection of subsets from Ω. Then there exists the smallest σ-algebra which contains all elements from $\mathcal{A}$. Such σ-algebra is typically denoted by $\sigma(\mathcal{A})$.

Proof. We notice first that the set of σ-algebras containing $\mathcal{A}$ is not empty since there is always at least one σ-algebra containing all elements from $\mathcal{A}$ and it is the power σ-algebra $\mathcal{P}_\Omega$.

Let $\mathcal{F}_i, i \in I$ be all such σ-algebras so that $\mathcal{A} \subset \mathcal{F}_i, i \in I$.

Define $\mathcal{F} := \cap_{i\in I}\mathcal{F}_i$.

It is then easy to see (verify it as an exercise) that $\mathcal{F}$ will be indeed a σ-algebra. By construction, it must be the smallest σ-algebra containing $\mathcal{A}$ so that $\mathcal{F} = \sigma(\mathcal{A})$.

∘ Now we choose $\Omega = \mathbb{R}^n$, where $n = 1, 2, ...$ is arbitrary but fixed (or Ω is any uncountable subset from $\mathbb{R}^n$).

Definition 5. Let $\mathcal{A}$ be the collection of all open sets in $\mathbb{R}^n$. Then, the σ-algebra $\sigma(\mathcal{A})$, that is the smallest σ-algebra which contains all open sets in $\mathbb{R}^n$, is called *the Borel σ-algebra on* $\mathbb{R}^n$ or σ-algebra of *Borel sets* in $\mathbb{R}^n$.

Examples of Borel sets in $\mathbb{R}$: all open sets in $\mathbb{R}$, all closed sets in $\mathbb{R}$, all half-closed intervals in $\mathbb{R}$ etc...

$$[a, b) = \cap_{n=1}^{\infty}(a - 1/n, b)$$

$$\{b\} = \cap_{n=1}^{\infty}[b, b + 1/n),$$

$$(a, b] = (a, b) \cup \{b\}$$

.....

Remark 1. a) Often the Borel σ-algebra on $\mathbb{R}^n$ is also denoted by $\mathcal{B}(\mathbb{R}^n)$.

b) The set $\mathcal{A}$ in the definition of the Borel σ-algebra is called the *generating set of subsets* from $\mathbb{R}^n$. In other words, the Borel sets are the sets which can be obtained from the open sets when using maximal countable number of unions, intersections, and complements on such sets and the sets obtained from open sets.

c) It is not hard to see that we could also use in $\mathcal{A}$ instead of "all open sets" some other generating collections of sets, for example, "all closed sets" or "all open intervals from left" etc..

Some more examples.

Example 5. Let $\Omega \neq \emptyset$ and $\mathcal{F}, \mathcal{G}$ are two σ-algebras on Ω. Is $\mathcal{F} \cup \mathcal{G}$ a σ-algebra?

Solution: First of all, under the union of two collections of subsets from Ω, one understands a collection of subsets which are either in one of two collections or in both of them.

A union of two σ-algebras does not need to be a σ-algebra as the following counterexample should illustrate.

Let $\Omega = \{1, 2, 3, 4\}$ and we choose $\mathcal{F} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$ and $\mathcal{G} = \{\emptyset, \Omega, \{1, 3\}, \{2, 4\}\}$.

Then

$$\mathcal{F} \cup \mathcal{G} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}, \{1, 3\}, \{2, 4\}\}$$

so that $\{1, 2\} \cap \{1, 3\} = \{1\} \notin \mathcal{F} \cup \mathcal{G}$.

Example 6. We are given that:

i) events A and B are independent: $P(A \cap B) = P(A)P(B)$;

ii) events A and C are independent: $P(A \cap C) = P(A)P(C)$;

iii) $C \subset B$;

Are the events A and $B \setminus C$ independent?

Solution: Condition iii) implies that $B = C \cup B \setminus C$ and using the addition formula we can write

$$P(B) = P(C) + P(B \setminus C).$$

Using i) and the last relation, we obtain

$$P(A \cap B) = P(A)P(B) = P(A)[P(C) + P(B \setminus C)] = P(A \cap C) + P(A)P(B \setminus C).$$

From another side,

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

Comparing the last two relations, we get that

$$P(A)P(B \setminus C) = P(A \cap B \setminus C)$$

proving the independence.