ESE 520 Probability and Stochastic Processes

Lecture 1

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

∘ What is Probability theory as a part of mathematics? It is dealing with randomness as such and, in particular, helps us to rigorously understand how to calculate the chance of occurring of a particular event that might happen as a result of a "random experiment".

Example 1. Toss a coin n times ($n = 1, 2, ...$). It is then easy to see how to calculate the related probability of a particular event. The experiment has *finitely many* possible elementary outcomes. More formally, the sample space of the experiment denoted here and later on by Ω consists of

$$\Omega = \{w = (w_1, w_2, ..., w_n)\},$$

where any w_i can have only two possible values, "heads" or "tails". For simplicity, we can use the value 1 for "heads" and the value 0 for "tails" so that $w_i \in \{0, 1\}$ for all $i = 1, 2, ..., n$. If we assume that on any single toss the probability of having "heads" is p $(0 < p < 1)$ so that the probability of having "tails" is $1 - p$ (or $q := 1 - p$) and all tosses are assumed to be *independent*, we can find that the probability of any elementary outcome $w \in \Omega$ of the experiment is given by

$$p(w) = p^{\sum_{i=1}^n w_i}(1 - p)^{n - \sum_{i=1}^n w_i}. \tag{1}$$

It is then not hard to see that

1) $0 \le p(w) \le 1, w \in \Omega$,

2) $\sum_{w \in \Omega} p(w) = 1$.

For any subset $A \subset \Omega$ (called then an *event*), we set

$$P(A) := \sum_{w \in A} p(w), \tag{2}$$

called the chance of occurring of event A.

Formula (2) defines then a measure on Ω called a probability that allows us to calculate the chance of occurring of any event A as a subset of Ω. The measure P in the example is concentrated on finitely many points $w \in \Omega$ since Ω consists of exactly 2^n points (number of all possible vectors $(w_1, w_2, ..., w_n)$, where any $w_i \in \{0, 1\}$.

∘ The above example can be generalized to the case of any random experiment with infinitely many but *countable!* outcomes.

Recall: a countable set Ω is a set of points that has as many elements as the set $\{1, 2, 3, ...\}$ so that, without loss of generality, we can assume that

$$\Omega = \{w_1, w_2, ...\},$$

where any w_i denotes an elementary outcome of the experiment.

In that case, any sequence of numbers $p(w_i), i = 1, 2, ...$ satisfying two conditions

1) $0 \leq p(w_i) \leq 1, i = 1, 2, ...$

2) $\sum_{i=1} p(w_i) = 1,$

will define a probability measure P on the subsets of Ω:

$$P(A) = \sum_{w_i \in A} p(w_i),$$

where A is *any subset* of Ω.

∘ Can the approach above be generalized to the case of a random experiment when Ω has infinitely but *uncountable many* elementary outcomes? Some examples of uncountable sets would be $[0, 1], [0, \infty), (-\infty, \infty)$ etc.

Example 2. "Tossing a coin *infinitely* many times independently. The sample space Ω of the experiment is then given as

$$\Omega = \{w = (w_1, w_2, ...), w_i \in \{0, 1\}, i = 1, 2, ...\}.$$

In other words, Ω consists of infinite sequences $(w_1, w_2, ...)$, where any component w_i has the value 1 or 0.

∘ How many such sequences are there in Ω?

It turns out "too many".

It is known from the number theory that any real number $w \in [0, 1)$ can be uniquely represented in the form (called the dyadic representation of w)

$$w = \frac{w_1}{2} + \frac{w_2}{2^2} + \frac{w_3}{2^3} + ...,$$

where any $w_i \in \{0, 1\}$. From that fact, it follows that there is $1 : 1$ relation

between the set Ω of the experiment and the interval of real numbers $[0, 1)$. However, the set $[0, 1)$ has uncountable many points so must Ω.

∘ The consequence from that when trying to construct a *probability measure* P on Ω is the following:

We cannot prescribe probabilities $p(w) > 0$ for all *single* $w \in \Omega$ (or at least uncountable many of them from Ω) since it must hold then

$$\sum_{w \in [0,1)} p(w) = 1.$$

However, the last relation would be impossible unless $p(w) = 0$ for uncountable many values $w \in [0, 1)$!

In other words, it is then mathematically impossible to build an adequate probability measure P on an uncountable sample space Ω prescribing initial probabilities $p(w)$ to single elementary outcomes w from Ω. In that case, Ω is "too large" to follow a "countable" path of constructing P.

∘ What is the solution in the uncountable case of Ω? One prescribes then the "initial probabilities" not to single points from Ω but to some "natural" sets from Ω. Then, a probability measure P defined initially on those "natural sets" is *extended* to a bigger collection of subsets from Ω all of which can be obtained from those "natural sets" using some standard operations on sets. This typical procedure of construction of a probability measure (in fact, of any measure, not necessarily a probability measure) leads to the important concept of a σ-algebra as a collection of events (subsets of Ω) on which a probability measure P is defined.

Definition 1. Let $\Omega \neq \emptyset$. Then, a collection $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* if

i) $\emptyset \in \mathcal{F}$;

ii) if $A \in \mathcal{F}$, then $A^c := \{w \in \Omega : w \notin A\} \in \mathcal{F}$;

iii) if $A_i \in \mathcal{F}, i = 1, 2, ...$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$;

Example 3. a) There exists always the trivial σ-algebra $\mathcal{F}_0 := \{\emptyset, \Omega\}$.

b) For any subset $A \subset \Omega$, define $\mathcal{F}_A := \{\emptyset, \Omega, A, A^c\}$. Clearly, $\mathcal{F}_A$ will be a σ-algebra called the σ-algebra generated by the set A.

c) The exist always so-called *power σ-algebra* $\mathcal{P}^{\Omega}$ as the collection of all possible subsets from Ω. Clearly, $\mathcal{P}^{\Omega}$ will the largest σ-algebra on Ω.

Remark 1. a) Any σ-algebra $\mathcal{F}$ is *closed* under all finite or countable infinite set-theoretical operations. Clearly, we could replace in the Definition 1 the operation of union by the operation on intersection.

b) If in the Definition 1 we replace $\cup_{i=1}^{\infty}$ by $\cup_{i=1}^{n}$ for any $n = 1, 2, ...$, then $\mathcal{F}$ is called *an algebra.*

c) We also recall some additional operations on sets that might be useful:

$$A \setminus B := \{w \in \Omega : w \in A, w \notin B\} = A \cap B^c$$

called the difference of A and B;

$$A \triangle B = A \setminus B \cup B \setminus A$$

called the symmetric difference of A and B;

$$(\cap_{i=1}^{\infty} A_i)^c = (\cup_{i=1}^{\infty} A_i^c)$$

called the De Morgan's law (Exercise!).

$\circ$ Now, a pair $(\Omega, \mathcal{F})$ where Ω is the sample space of a particular random experiment and $\mathcal{F}$ is a σ-algebra on Ω as a collection of subsets from Ω

also called the events is typically called a *measurable space* (in measure theory). What is missing now is the component P called a probability measure (or a probability function or simply probability). What kind of

"function" should be called a probability function?

Definition 2. (due to A.A. Kolmogoroff, 1933) Let $\Omega \neq \emptyset$ and $\mathcal{F}$ is a σ-algebra on Ω. A set function $P : \mathcal{F} \to [0,1]$ is called a *probability (measure)* on $\mathcal{F}$ if it satisfies the following axioms:
i) $P(\Omega) = 1$ (normalization condition);
ii) if $A_i \in \mathcal{F}, i = 1, 2, ...$ are pairwise disjoint (do not have points in common), then

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

(σ-additivity condition) $\circ$ If in the condition ii) we replace $\cup_{i=1}^{\infty} A_i$ by the

finite union $\cup_{i=1}^{n} A_i$, then P is called an additive function.

Remark 2. Clearly, the probability measure P we constructed in the case of a finite or infinite but countable sample space Ω, satisfies the properties i) and ii) from Definition 2. In general, when Ω is uncountable, our goal is then to construct a measure P which will satisfy those properties i) and ii) so that we can use that measure P as a probability measure.

$\circ$ Before we address some details of construction of a probability measure P, let us mention some simple but useful properties/formulas of any probability function:

a) For any $A \in \mathcal{F}$, it holds: $P(A^c) = 1 - P(A)$ (complement rule); - Use: $\Omega = A \cup A^c$;
b) For any sets $A \subset B$, it holds: $P(A) \leq P(B)$ (monotonicity of probability function); Use: $B = A \cup B \setminus A$;
c) For any sets $A, B \in \mathcal{F}$, it holds:

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

(addition rule)

The next property is less trivial and we prove it as a statement:

d) A normalized finitely additive non-negative set function μ is a probability function if and only if it satisfies the following property called the "continuity from below":

$$\lim_{n\to\infty} \mu(A_n) = \mu(A) \tag{3}$$

for any sequence of subsets $A_1, A_2, ...$ such that $A_n \subset A_{n+1}$ where $A = \cup_{n=1}^{\infty} A_n$.

Proof. We have to show two directions. Assume first that μ is a probability function (as defined in Definition 2) and the sequence $A_n, n = 1, 2, ...$ is as specified with $A = \cup_{n=1}^{\infty} A_n$.

Define $B_1 := A_1, B_n := A_n \setminus A_{n-1}, n = 1, 2, ..., A_0 = \emptyset$.

We see that $A = \cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n$, where the sets $B_1, B_2, ...$ are pairwise disjoint (by construction) and $B_n \in \mathcal{F}, n = 1, 2, ...$

Since P is a probability function thus σ-additive, tt follows that

$$P(A) = \sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} [P(A_n) - P(A_{n-1})] =$$

$$\lim_{k\to\infty} \sum_{n=1}^{k} [P(A_n) - P(A_{n-1})] = \lim_{k\to\infty} P(A_k).$$

For the opposite direction, assume that μ is finitely additive and the condition (3) holds.

We choose an arbitrary sequence of pairwise disjoint sets $B_1, B_2, ...$ and define the new sequence of sets $A_1, A_2, ...$ as $A_n := \cup_{i=1}^{n} B_i$. Clearly, by construction,

$$A_n \subset A_{n+1}$$

so that $A := \cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n$.

We have then that

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu(\cup_{i=1}^{n} B_i) = \lim_{n\to\infty} (\sum_{i=1}^{n} \mu(B_i)) = \sum_{i=1}^{\infty} \mu(B_i).$$