ESE 520 Probability and Stochastic Processes

Lecture 3

"Probability measures on Borel sets. Random variables and their probability distributions: general theory".

Let $\Omega = \mathbb{R}$ (or any set from $\mathbb{R}$) and let us look at some ways how we can construct a *probability measure* on Ω.

A. (a particular but important case) Assume $f : \mathbb{R} \to \mathbb{R}$ be any non-negative function satisfying $\int_{\mathbb{R}} f(x)dx = 1$. For any interval I from $\mathbb{R}$ we define

$$P(I) := \int_I f(x)dx, \tag{1}$$

where the integral in (1) is a Riemann integral.

It follows then from the properties of the Riemann integral and those of the function f that the set function P in (1) will be a probability measure well-defined on all Borel sets.

Exercise: Prove that P satisfies all three properties from the definition of a probability measure.

Actually, in this case we have that P is defined for any subset from $\mathbb{R}$, not only a Borel subset.

B. (general case) Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing, right-continuous function satisfying $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. For any interval $I = (a,b]$, define

$$P(I) := F(b) - F(a). \tag{2}$$

It is not hard to verify that the set function P defined in (2) will be *a finite additive probability measure* when operating on the collection of intervals of the form $(a,b]$ and all sets that can be obtained from them using finite number of unions, intersections, and complements of such intervals.

It is then a little bit more difficult task to show that the relation (2) will define the unique probability measure acting on the σ-algebra of Borel sets

(that result is called the *Caratheodory extension theorem*). The later result is a cornerstone of the entire *measure theory*. In fact, the converse is then also true: any probability measure P defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ has the representation (2).

Remark 1. It follows from (2) that for $I = (-\infty, b]$, one has

$$P(I) = F(b).$$

Comparing (1) and (2), we see that the case A is a particular case of the general case B when

$$F(b) = \int_{-\infty}^{b} f(x)dx,$$

or, equivalently,

$$F'(b) = f(b)$$

meaning that the function F is differentiable and f is the derivative of F.

Actually, the case A above is usually referred in the probability theory to the case when P is *absolute continuous.*

Remark 2. Another particular case of the function F is the case when F is a piece-wise constant function with a *countable number of jumps*. It follows then from (2) that the measure P must be concentrated only on the countable number of points at which those jumps of F occur. The size of jump at a particular point of jump b is then $F(b) - F(b-)$.

An alternative way to describe the corresponding probability measure P would be the following: let D be the countable set of points from $\mathbb{R}$ representing the points of jump of F. For any $x \in D$, let $p(x) := F(x) - F(x-)$ so that $0 \le p(x) \le 1$ and $\sum_{x \in D} p(x) = 1$. Then, for any subset $I \subset D$, we define a set function P as

$$P(I) := \sum_{x \in I} p(x). \tag{3}$$

Clearly, P will be a probability measure (we discussed it in Lecture 1) called in this case a *discrete probability measure* corresponding to F.

Remark 3. The cases of absolute continuous and discrete probability measures are the most important ones though there is one more class

of probability measures called *singular probability measures.* However, we won't discuss in this course singular probability measures (or singular probability distributions). We only mention that, as can be proven, any probability measure P defined on a Borel σ-algebra of $\mathbb{R}$ can be represented in a form

$$P = \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3,$$

where P_1 is an absolute continuous probability measure, P_2 is a discrete probability measure, and P_3 is a singular probability measure, and the real numbers $\alpha_i, i = 1, 2, 3$ are such that

$$\alpha_1 + \alpha_2 + \alpha_3 = 1.$$

Now, after we discussed how a probability measure on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ can be defined, we have to discuss what are the *random variables* in the probability theory and how to deal with them mathematically. After all, we have to understand what is meant if we read somewhere in a book "...let X be a random variable that has a geometric probability distribution" (or any other specified probability distribution as a matter of fact).

It should be clear that a random variable X must be a particular map $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where P is a probability measure on the sets from $\mathcal{F}$. What properties the map X should satisfy? It turns out that not too many because the most important property on X must be the one that we should be able to calculate probabilities of the form $P(w : X(w) \in B)$, say for any Borel set from $\mathbb{R}$. This leads to the following general

Definition 1. A map $X : \Omega \to \mathbb{R}^n$ is said to be a *random variable* if the inverse image of any Borel set $B \in \mathcal{B}(\mathbb{R}^n)$ is an event in $\mathcal{F}$, i.e.

$$X^{-1}(B) := \{w : X(w) \in B\} \in \mathcal{F} \tag{4}$$

for any $B \in \mathcal{B}(\mathbb{R}^n)$.

Therefore, a random variable X is understood as a map between two measurable spaces $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying the condition (4). X

is then also called to be a *measurable map*. If $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then X is a *measurable function*.

To study measurable maps acting between two measurable spaces, everything what one needs is in fact the definition of the inverse image of a set as defined in (4).

Example 1. Let D be a subset from $\mathbb{R}$ and consider a map $f : D \to \mathbb{R}$ meaning only that, for any $x \in D$, it holds that $f(x) \in \mathbb{R}$. Now, for a subset $A \subset D$, we define

$$f(x) = \mathbf{1}_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Clearly, $\mathbf{1}_A(x)$ is a map between D and $\mathbb{R}$. Let us find the inverse image of $\mathbf{1}_A^{-1}(B)$, where B is any Borel set from $\mathbb{R}$.

By the definition of the inverse image, we have that

$$f^{-1}(B) = \mathbf{1}_A^{-1}(B) := \begin{cases} D, & 0, 1 \in B \\ A^c, & 0 \in B, 1 \notin B \\ A, & 0 \notin B, 1 \in B \\ \emptyset, & 0, 1 \notin B. \end{cases}$$

The following properties of inverse images can be proved directly by definition and are useful when working with measurable maps as random variables:

i) $X^{-1}(\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty} X^{-1}(B_i)$;

ii) $X^{-1}(\cap_{i=1}^{\infty} B_i) = \cap_{i=1}^{\infty} X^{-1}(B_i)$

iii) $X^{-1}(B^c) = (X^{-1}(B))^c$,

where X is any measurable map and B, B_i are arbitrary Borel sets.

Remark 4. It follows from the above three properties that the collection of subsets from $\mathcal{F}$ defined as

$$\mathcal{F}^X := \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\}$$

will be a σ-algebra called sometimes *the σ-algebra generated by X*. Clearly, $\mathcal{F}^X \subset \mathcal{F}$.

Example 2. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ corresponding to the experiment if we roll a die once.

a) Consider a σ-algebra

$$\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

Define a map $X(w) = w$ (identity map). We see that

$$X^{-1}(\{3\}) = \{w \in \Omega : X(w) = 3\} = \{3\} \notin \mathcal{F}$$

so that X is not $\mathcal{F}$-measurable random variable.

b) Define

$$X(w) = \begin{cases} 0, & w = 1, 3, 5 \\ 1, & w = 2, 4, 6. \end{cases}$$

We can easily see that

$$\mathcal{F}^X = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

The value of X only tells us whether the result of the toss was even or odd but the σ-algebra generated by X exactly tells us WHEN it happened.

Now, formally we are ready to define what is called the *probability distribution* of a random variable X.

Definition 2. The set function P_X defined on Borel sets $B \in \mathcal{B}(\mathbb{R}^n)$ as

$$P_X(B) := P(w \in \mathcal{F}^X : X(w) \in B\}) \tag{5}$$

is said to be the *probability distribution of X*.

We see from (5) that P_X is the image of the probability measure P restricted to the sets in $\mathcal{F}^X$ in the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

As indicated before, it is enough in fact to know $P_X(B)$ *not for all Borel sets B* but for those from a *generating collection of sets $\mathcal{A}$*. Then, instead of $P_X(B)$ one has a function called the *cumulative distribution function* (cdf).

If $n = 1$, then one defines the cdf as

$$F_X(x) := P(w : X(w) \le x) = P(w : X^{-1}((-\infty, x]) \tag{6}$$

for any $x \in \mathbb{R}$.

For $n \ge 2$, one defines

$$F_X(x_1, x_2, ..., x_n) := P(w : X_i(w) \le x_i, i = 1, 2, ..., n) \tag{7}$$

for all $x_i \in \mathbb{R}, i = 1, 2, ..., n$.

Proposition. Let F be a function defined as in (6). Then F is a non-decreasing, right-continuous function satisfying

$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1.$$

Proof. We choose $a < b$ so that $(-\infty, a] \subset (-\infty, b]$ and since P is an increasing function

$$F(a) = P(X \le a) \le P(X \le b) = F(b)$$

thus F is non-decreasing.

Now, choose a sequence $a_n \downarrow a$ so that $(-\infty, a_n] \downarrow \cap_{n=1}^{\infty}(-\infty, a_n] = (-\infty, a]$. Using the property of "continuity from below" of P discussed earlier we obtain that

$$F(a_n) = P(X \le a_n) \to P(X \le a) = F(a)$$

meaning that F is continuous from the right.

Choose $a_n \uparrow +\infty$ so that $\cup_{n=1}^{\infty}(-\infty, a_n] = \mathbb{R}$. By similar argument,

$$\lim_{n \to \infty} F(a_n) = \lim_{n \to \infty} P(X \le a_n) = P(X \in \mathbb{R}) = 1.$$

Choose $a_n \downarrow -\infty$ so that $\cap_{n=1}^{\infty}(-\infty, a_n] = \emptyset$. By similar argument,

$$\lim_{n \to \infty} F(a_n) = P(\emptyset) = 0.$$

We also note that $F(x)$ is *not necessarily left-continuous*. To see that, we choose $a_n \uparrow a$ so that $\cup_{n=1}^{\infty}(-\infty, a_n] = (-\infty, a)$. Then,

$$P(X \le a_n) \to P(X < a) = F(a) - P(X = a)$$

so that

$$\lim_{n \to \infty} F(a_n) = F(a) - P(X = a).$$

It means that F will be left-continuous only if $P(X = a) = 0$ for any $a \in \mathbb{R}$.

Theorem. (Caratheodory extension theorem) Any function $F : \mathbb{R} \to \mathbb{R}$ being a non-decreasing, right-continuous function with properties $\lim_{n \to -\infty} F(x) = 0$ and $\lim_{n \to \infty} F(x) = 1$ is the cdf of a random variable X.

In fact, one can construct then such X as $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, where

$$X(w) = w, \Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}(\mathbb{R}), P = P_X$$

and

$$P_X((a, b]) = F(b) - F(a)$$

for all $(a, b], a < b$.

It can then be shown (not so trivial!) that P_X defined above is indeed a probability measure defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.