Lecture 1

Last updated by Serik Sagitov: February 6, 2013

Abstract

A course based on the book Probabilities and Random Processes by Geoffrey Grimmett and David Stirzaker. Chapter 1. Events and their probabilities. Chapter 2. Random variables and their distributions.

1 Probability space

A random experiment is modeled in terms of a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$

- the *sample space* Ω is the set of all possible outcomes of the experiment,
- the *σ-field* (or sigma-algebra) $\mathcal{F}$ is a collection of measurable subsets $A \subset \Omega$ (which are called *random events*) satisfying
  1. $\emptyset \in \mathcal{F}$,
  2. if $A_i \in \mathcal{F}$, $0 = 1, 2, \ldots$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$, countable unions,
  3. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, *complementary event*,
- the probability measure $\mathbb{P}$ is a function on $\mathcal{F}$ satisfying three probability axioms
  1. if $A \in \mathcal{F}$, then $\mathbb{P}(A) \geq 0$,
  2. $\mathbb{P}(\Omega) = 1$,
  3. if $A_i \in \mathcal{F}$, $0 = 1, 2, \ldots$ are all disjoint, then $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

De Morgan's laws

$$\Big(\bigcap_i A_i\Big)^c = \bigcup_i A_i^c, \qquad \Big(\bigcup_i A_i\Big)^c = \bigcap_i A_i^c.$$

Properties derived from the axioms

$$\begin{aligned}\mathbb{P}(\emptyset) &= 0,\\ \mathbb{P}(A^c) &= 1 - \mathbb{P}(A),\\ \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).\end{aligned}$$

Inclusion-exclusion rule

$$\begin{aligned}\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \ldots \\ + (-1)^{n+1} \mathbb{P}(A_1 \cap \ldots \cap A_n).\end{aligned}$$

Continuity of the probability measure

- if $A_1 \subset A_2 \subset \ldots$ and $A = \cup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i$, then $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$,
- if $B_1 \supset B_2 \supset \ldots$ and $B = \cap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i$, then $\mathbb{P}(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$.

2 Conditional probability and independence

If $\mathbb{P}(B) > 0$, then the conditional probability of A given B is

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

The law of total probability and the Bayes formula. Let $B_1, \ldots, B_n$ be a partition of Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i),$$
$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Definition 1 *Events $A_1, \ldots, A_n$ are independent, if for any subset of events $(A_{i_1}, \ldots, A_{i_k})$*

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}).$$

Example 2 *Pairwise independence does not imply independence of three events. Toss two coins and consider three events*

- $A =$\{*heads on the first coin*\},
- $B =$\{*tails on the first coin*\},
- $C =$\{*one head and one tail*\}.

Clearly, $\mathbb{P}(C|A) = \mathbb{P}(C)$ *and* $\mathbb{P}(C|B) = \mathbb{P}(C)$ *but* $\mathbb{P}(C|A \cap B) = 0$.

3 Random variables

A real random variable is a measurable function $X : \Omega \to \mathbb{R}$ so that different outcomes $\omega \in \Omega$ can give different values $X(\omega)$. Measurability of $X(\omega)$:

$$\{\omega : X(\omega) \le x\} \in \mathcal{F} \text{ for any real number } x.$$

Probability distribution $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$ defines a new probability space $(R, \mathcal{B}, \mathbb{P}_X)$, where $\mathcal{B} = \sigma$(all open intervals) is the Borel sigma-algebra.

Definition 3 *Distribution function (cumulative distribution function)*

$$F(x) = F_X(x) = \mathbb{P}_X\{(-\infty, x]\} = \mathbb{P}(X \le x).$$

In terms of the distribution function we get

$$\mathbb{P}(a < X \le b) = F(b) - F(a),$$
$$\mathbb{P}(X < x) = F(x-),$$
$$\mathbb{P}(X = x) = F(x) - F(x-).$$

Any monotone right-continuous function with

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

can be a distribution function.

Definition 4 *The random variable X is called discrete, if for some countable set of possible values*

$$\mathbb{P}(X \in \{x_1, x_2, \ldots\}) = 1.$$

Its distribution is described by the probability mass function $f(x) = \mathbb{P}(X = x)$.

The random variable X is called continuous, if its distribution has a probability density function $f(x)$:

$$F(x) = \int_{-\infty}^{x} f(y)dy, \quad \textit{for all } x,$$

so that $f(x) = F'(x)$ almost everywhere.

Example 5 *The indicator of a random event* $I_A = 1_{\{\omega\in A\}}$ *with* $p = \mathbb{P}(A)$ *has a Bernoulli distribution* $Ber(p)$

$$\mathbb{P}(I_A = 1) = p, \quad \mathbb{P}(I_A = 0) = 1 - p.$$

For several events $S_n = \sum_{i=1}^n I_{A_i}$ *counts the number of events that occurred. If independent events* $A_1, A_2, \ldots$ *have the same probability* $p = \mathbb{P}(A_i)$*, then* S_n *has a binomial distribution* $Bin(n,p)$

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Example 6 *(Cantor distribution) Consider* $(\Omega, \mathcal{F}, \mathbb{P})$ *with* $\Omega = [0,1]$*,* $\mathcal{F} = \mathcal{B}_{[0,1]}$*, and*

$$\begin{aligned}
&\mathbb{P}([0,1]) = 1\\
&\mathbb{P}([0,1/3]) = \mathbb{P}([2/3,1]) = 2^{-1}\\
&\mathbb{P}([0,1/9]) = \mathbb{P}([2/9,1/3]) = \mathbb{P}([2/3,7/9]) = \mathbb{P}([8/9,1]) = 2^{-2}
\end{aligned}$$

and so on. Put $X(\omega) = \omega$*, its distribution, called the Cantor distribution, is neither discrete nor continuous. Its distribution function, called the Cantor function, is continuous but not absolutely continuous.*

4 Random vectors

Definition 7 *The joint distribution of a random vector* $\mathbf{X} = (X_1, \ldots, X_n)$ *is the function*

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(\{X_1 \le x_1\} \cap \ldots \cap \{X_n \le x_n\}).$$

Marginal distributions

$$\begin{aligned}
F_{X_1}(x) &= F_{\mathbf{X}}(x, \infty, \ldots, \infty),\\
F_{X_2}(x) &= F_{\mathbf{X}}(\infty, x, \infty, \ldots, \infty),\\
&\ldots\\
F_{X_n}(x) &= F_{\mathbf{X}}(\infty, \ldots, \infty, x).
\end{aligned}$$

The existence of the joint probability density function $f(x_1, \ldots, x_n)$ *means that the distribution function*

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(y_1, \ldots, y_n) dy_1 \ldots dy_n, \quad \text{for all } (x_1, \ldots, x_n),$$

is absolutely continuous, so that $f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$ *almost everywhere.*

Example 8 *In general, the joint distribution can not be recovered form the marginal distributions. If*

$$F_{X,Y}(x,y) = xy 1_{\{(x,y)\in[0,1]^2\}},$$

then vectors (X,Y) *and* (X,X) *have the same marginal distributions.*

Exersize 2.7.14 b. Consider

$$F(x,y) = \begin{cases} 1 - e^{-x} - xe^{-y} & \text{if } 0 \le x \le y,\\ 1 - e^{-y} - ye^{-y} & \text{if } 0 \le y \le x,\\ 0 & \text{otherwise.} \end{cases}$$

Show that $F(x,y)$ is the joint distribution function of some pair (X,Y). Find the marginal distribution functions and densities.

Solution. Three properties should be satisfied for $F(x,y)$ to be the joint distribution function of some pair (X,Y):

1. $F(x,y)$ is non-decreasing on both variables,
2. $F(x,y) \to 0$ as $x \to -\infty$ and $y \to -\infty$,

3. $F(x,y) \to 1$ as $x \to \infty$ and $y \to \infty$.

Observe that

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = e^{-y} 1_{\{0 \le x \le y\}}$$

is always non-negative. Thus the first property follows from the integral representation:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dudv,$$

which, for $0 \le x \le y$, is verifies as

$$\int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dudv = \int_0^x \Big(\int_u^y e^{-v} dv \Big) du = 1 - e^{-x} - xe^{-y},$$

and for $0 \le y \le x$ as

$$\int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dudv = \int_0^y \Big(\int_u^y e^{-v} dv \Big) du = 1 - e^{-y} - ye^{-y}.$$

The second and third properties are straightforward. We have shown also that $f(x,y)$ is the joint density.

For $x \ge 0$ and $y \ge 0$ we obtain the marginal distributions as limits

$$F_X(x) = \lim_{y\to\infty} F(x,y) = 1 - e^{-x}, \qquad f_X(x) = e^{-x},$$
$$F_Y(y) = \lim_{x\to\infty} F(x,y) = 1 - e^{-y} - ye^{-y}, \qquad f_Y(y) = ye^{-y}.$$

$X \sim \text{Exp}(1)$ and $Y \sim \text{Gamma}(2,1)$.

5 Repeated coin tossing

Let S_n be the number of heads in n independent tossings of a fair coin. Figure 1 shows imbedded partitions $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \mathcal{F}_4 \subset \mathcal{F}_5$ of the sample space generated by S_1, S_2, S_3, S_4, S_5.

Definition 9 *A sequence of sigma-fields $\{\mathcal{F}_n\}_{n=1}^\infty$ such that*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_n \subset \ldots, \quad \mathcal{F}_n \subset \mathcal{F} \textit{ for all } n$$

is called a filtration.

The events representing our knowledge of the first three tossings is given by $\mathcal{F}_3$. From the perspective of $\mathcal{F}_3$ we can not say exactly the value of S_4. Clearly, there is dependence between S_3 and S_4. The joint distribution of S_3 and S_4:

	$S_4=0$	$S_4=1$	$S_4=2$	$S_4=3$	$S_4=4$	Total
$S_3=0$	1/16	1/16	0	0	0	1/8
$S_3=1$	0	3/16	3/16	0	0	3/8
$S_3=2$	0	0	3/16	3/16	0	3/8
$S_3=3$	0	0	0	1/16	1/16	1/8
Total	1/16	1/4	3/8	1/4	1/16	1

The conditional expectation

$$\mathbb{E}(S_4|S_3) = S_3 + 0.5$$

is a discrete random variable with values $0.5, 1.5, 2.5, 3.5$ and probabilities $1/8, 3/8, 3/8, 1/8$.

For finite n the picture is straightforward. For $n = \infty$ it is a non-trivial task to define an overall $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega = (0,1]$. One can use the Lebesgue measure $\mathbb{P}(dx) = dx$ and the sigma-field $\mathcal{F}$ of Lebesgue measurable subsets of $(0,1]$. Not all subsets of $(0,1]$ are Lebesgue measurable. Not all Lebesgue measurable sets are Borel sets.

More generally, if the coin has probability p of heads, we can use $(\Omega, \mathcal{F}, \mathbb{P}_p)$ with the same $(\Omega, \mathcal{F})$ and

$$\mathbb{P}_p(dx) = l(p)x^{l(p)-1}dx, \quad l(p) = -\log_2(p).$$

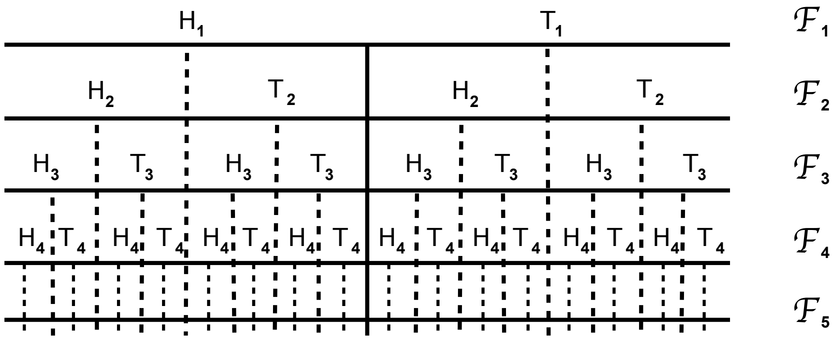

Figure 1: Sigma-fields for four consecutive coin tossings.