RANDOM VARIABLES

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

September 16, 2019

Outline

Definition

In Ross the definitions are introduced in Chapter 4.

We have been using the concept already.

New notation for the sample space: Greek Ω instead of S.

(We need the letter S for other purposes.)

Definition A **random variable** is a function defined on Ω.

Notation: X or Y or other capital letters (e.g. S).

Values: alphanumeric (strings of characters). Now: mostly numeric.

Examples

(1) Flip a coin and denote the outcome by “H” or “T”. Now:

Let $X = 0$ *or* $X = 1$.

(2) Sample 3 balls from an urn containing red and green balls and denote the outcomes by “1 red” or “2 red” or “3 red”. Now:

R *is the number of red balls,* $R = r$*, where* $r = 0, 1, 2, 3$.

(3) Break a stick of unit length at a random point U. That is,

U *is a uniform random variable with values in* $[0, 1]$.

Structured values

A random point on the plane can be written as (X_1, X_2), in the n-space as $\mathbf{X} = (X_1, \ldots, X_n)$. We may still call it "random variable" or use the term "random vector" or "random sequence". A random sequence may be even infinite. A sequence is but a function.

So, why not admit random functions, as well?

Some semantic purists restrict the usage of the term "random variable" to real values (Ross does). Then in all other case they use a generic name, e.g., "random element".

Distribution of RV

Definition. A **probability distribution** of a random variable X is any system of information that yields all relevant probabilities

$$\mathrm{P}(X \in A) \quad - \quad \text{probability that } X \text{ takes values in a set } A.$$

For example, $\mathrm{P}(X = x)$ or $\mathrm{P}(X \leq x)$, etc.

In this course we distinguish two major types of random variables:

discrete $\mathrm{P}(X = x) > 0$ for all x of interest,

continuous $\mathrm{P}(X = x) = 0$ for all x and a density (next) exists.

There are others but we hardly encounter them here.

Probability mass and density functions

Discrete RV, **pmf**:

Let $f(x) \geq 0$ and $\sum_x f(x) = 1$. $f(x) = \mathrm{P}(X = x)$.

Then $\mathrm{P}(X \in A) = \sum_{x \in A} f(x)$.

Continuous RV: **pdf**:

Let $f(x) \geq 0$ be a function on $\mathbb{R}$ such that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Then $\mathrm{P}(X \in A) = \int_A f(x)\,dx$,

Note the connections to physics (probability perceived as mass).

Structured RVs

The **joint distribution** of a sequence $\mathbf{X} = (X_1, \ldots, X_n)$ of RVs:

$$\mathrm{P}(\mathbf{X} \in E), \quad E \subset \mathbb{R}^n, \quad \text{in particular:}$$

$$\mathrm{P}(X_1 \in A_1, \ldots, X_n \in A_n) \quad \text{(the commas are read as "and").}$$

We prefer even simple forms, for $\mathbf{x} = (x_1, \ldots, x_n)$:

- **discrete**, **joint pmf** $f(\mathbf{x}) = \mathrm{P}(X = x_1, \ldots, X_n = x_n)$,
- **continuous**: **joint pdf** $f(\mathbf{x})$ so $\mathrm{P}(\mathbf{X} \in E) = \underset{E}{\int \cdots \int} f(\mathbf{x})\, d\mathbf{x}$.

Ross: Chapter 6

Domain vs. support

Previously, we wrote "$\mathrm{P}(X = x) > 0$ for all x of interest".

Example. Roll a standard die. Denote the outcome by X.

What is $\mathrm{P}(X = 7)$? We simply say that $f(7) = \mathrm{P}(X = 7) = 0$.

The value "7" is not of interest but we may include it in the domain of the pmf $f(x)$, and even take the entire $\mathbb{R}$.

Definition. For a pmf or pdf, the set $\{x : f(x) > 0\}$ is called the **support** of f or X.

Example. The pdf $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$ has the domain $(-\infty, \infty)$ and the support $[0, \infty)$.

Notation

We may write

$$f(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad \text{or} \quad f(x) = e^{-x},\ x \geq 0.$$

We may use the Boolean **indicator function**

$$\mathbb{1}_{\{x \in A\}} = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A; \end{cases} = \mathbb{1}_A(x).$$

So,

$$f(x) = e^{-x}\mathbb{1}_{\{x \geq 0\}} \quad \text{or} \quad f(x) = e^{-x}\mathbb{1}_{[0,\infty)}(x).$$

RVs vs probability distribution

Every RV has a probability distribution (PD).

In fact, there are (infinitely many) RVs with the same distribution.

Given a PD, we can define a corresponding RV in many ways, e.g.,

- discrete case: given $(\mathbf{x}, \mathbf{p}) = (x_n,\, p_n)$ such that $p_n \geq 0$ and $\sum_n p_n = 1$, we put $\Omega = \{x_n\}$ and $\mathrm{P}(x_n) = p_n$;
- continuous case: given $f(x)$ such that $f(x) \geq 0$ and $\int_{\mathbb{R}} f(x)\, dx = 1$, we put $\Omega = \mathbb{R}$ and $\mathrm{P}(A) = \int_A f(x)\, dx$;

Then we define $X(\omega) = \omega$ by tautology.

Although such "construction" is rigorous but it's abstract.

A "physical" construction may be desired instead.

Making examples

The quickest example: take any $g(x) \geq 0$.

- if discrete, then summing up, put $m = \sum_x g(x)$,
- if continuous, then integrating, put $m = \int_{\mathbb{R}} g(x)$.

Then, with $c = \dfrac{1}{m}$, $f(x) = c\, g(x)$ becomes a pmf or pdf.

A discrete RV with infinitely many values entails an infinite series.

A continuous RV may yield an improper integral.

That's why we need Calculus 2 as the prerequisite.

Example

Consider discrete $g(x) = \dfrac{1}{x^p}$, $x = 1, 2, \ldots$.
For what values of p we can turn $g(x)$ into the pmf?

Answer: The series $m = \sum_x \dfrac{1}{x^p}$ must converge.
It happens iff $p > 1$.

Q: Do we need to know the value m of the series?

A: Not really, we can use the constant $c = \frac{1}{m}$ without knowing it.

The continuous $g(x) = \dfrac{1}{x^p}$, $x \geq 1$ is much simpler because we can easily compute the integral $m = \displaystyle\int_1^\infty \frac{1}{x^p}\, dx = \frac{1}{p+1}$ for $p > 1$.

Definition

We say that $X_1, \ldots, X_n$, with pmf's or pdf's $f_1, ..., f_n$ are **independent** if the joint pmf or pdf satisfies

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Notice that in the discrete case the equation means

$$\mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathrm{P}(X_1 = x_1) \cdots \mathrm{P}(X_n = x_n)$$

and so the definition extends the notion of independent events.

Extension

We can deduce even more general property:

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

- **discrete**: by summation
- **continuous**: by integration

Preservation of independence

Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be independent. Create new random variables using arbitrary functions ϕ and ψ:

$$X = \phi(X_1, \ldots, X_m), \quad Y = \psi(Y_1, \ldots, Y_n).$$

Theorem

X and Y are independent.

Proof. (for discrete RVs only).

We'll use the conditioning upon the supports of **X** and **Y**.

The underlying conditional probabilities are either 1 or 0, so we can use the Boolean notation $\mathbb{1}_{\{\ldots\}}$.

Calculations

Also, $\mathbb{1}_{A\cap B} = \mathbb{1}_A \, \mathbb{1}_B$. Then, by basic arithmetics,

$$
\begin{aligned}
&\mathrm{P}(X = x,\ Y = y) \\
&= \sum_{\mathbf{x}} \sum_{\mathbf{y}} \mathrm{P}\Big(\phi(\mathbf{X}) = x,\ \psi(\mathbf{Y}) = y | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}\Big)\, \mathrm{P}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}\Big) \\
&= \sum_{\mathbf{x}} \sum_{\mathbf{y}} \mathbb{1}_{\{\phi(\mathbf{x})=x,\, \psi(\mathbf{y})=y\}} \mathrm{P}(\mathbf{X} = \mathbf{x})\, \mathrm{P}(\mathbf{Y} = \mathbf{y}) \\
&= \left(\sum_{\mathbf{x}} \mathbb{1}_{\{\phi(\mathbf{x})=x\}} \mathrm{P}(\mathbf{X} = \mathbf{x}) \right) \left(\sum_{\mathbf{y}} \mathbb{1}_{\{\phi(\mathbf{y})=y\}} \mathrm{P}(\mathbf{Y} = \mathbf{y}) \right) \\
&= \mathrm{P}(X = x)\mathrm{P}(Y = y)
\end{aligned}
$$

Corollary

Thus we have obtained an easy proof of the Theorem on preservation of independence of events (`BP_03_conditioning`, Slide 6).

Indeed, events $A_1, \dots, A_n$ are independent iff RVs $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ are independent.

The case of continuous distributions is much more complicated and requires Measure Theory or Advanced Calculus.

Definition

The **cumulative distribution function** is defined as:

$$F(x) = \mathrm{P}(X \leq x),$$

which also makes sense in the discrete case, but then the pmf is much more convenient.

In the continuous case: $F(x) = \displaystyle\int_{-\infty}^{x} f(u)\, du.$

Then $\mathrm{P}(X \leq x) = \mathrm{P}(X < x)$, which may be false for a discrete RV, because both probabilities differ by $\mathrm{P}(X = x)$.

Properties of cdf

1. $\mathrm{P}(a < X \leq b) = F(b) - F(a)$.
2. $\mathrm{P}(X > x) = 1 - F(x)$.
3. In the continuous case, the pdf can be recovered from cdf:
   $f(x) = F'(x)$.
   (In the discrete case, the recovery of the pmf is possible.)
4. This is a typical cdf:

$$F(-\infty) = 0, \quad F(\infty) = 1, \quad F(x) \text{ is non-decreasing}$$

$$F(x) \text{ is } \textbf{rcll} \text{ (right continuous with left limits)}$$