Probability and Statistics - Elementary Probability Theory Random Variables

Giuliano Casale
Department of Computing, Imperial College London

Definition

- A probability space $(S, \mathcal{F}, \mathrm{P})$ is a triplet that models our random experiment by means of a probability measure $\mathrm{P}(E)$ defined on subsets $E \subseteq S$ of the sample space S belonging to the σ-algebra $\mathcal{F}$.
- Within this space, we may want to study quantities of interest that are function of randomly occurring events, e.g., inflation, temperature, exchange rates, job response times, ...
- Random variables provide a formalism to map these variables of interest to numerical values.

Definition

- A Random Variable (r.v.) is a mapping from the sample space to the real numbers. So if X is a random variable, $X : S \to \mathbb{R}$.
- Each element of the sample space $s \in S$ is therefore assigned by X a numerical value $X(s)$.
- If we denote the generic (unknown) outcome of the random experiment as s, then the corresponding outcome of the random variable $X(s)$ will be generically referred to as X.

Examples of Random Variables

Consider once again the experiment of rolling a single fair die.

- Then $S = \{\odot, \odot, \odot, : \mathfrak{E}, : \mathfrak{E}\}$ and for any $s \in S, \mathrm{P}(\{s\}) = \frac{1}{6}$.
- An obvious random variable to define on S is $X : S \to \mathbb{R}$, s.t.

$$
\begin{aligned}
X(\odot) &= 1, \\
X(\odot) &= 2, \\
&\vdots \\
X(:) &= 6.
\end{aligned}
$$

- Then e.g. $\mathrm{P}_X(1 < X \leq 5) = \mathrm{P}(\{\odot, \circledast, \circledast, \circledast\}) = \frac{4}{6} = \frac{2}{3}$
- and $\mathrm{P}_X(X \in \{2, 4, 6\}) = \mathrm{P}(\{\odot, \circledast, \circledast\}) = \frac{1}{2}$.
- Alternatively, we could define a random variable $Y : S \to \mathbb{R}$, s.t.

$$
\begin{aligned}
Y(\odot) = Y(\odot) = Y(:\mathrm{B}) = 0, \\
Y(\odot) = Y(:\mathrm{B}) = Y(\ \mathrm{B}) = 1.
\end{aligned}
$$

- Then clearly

$$
\mathrm{P}_Y(Y = 0) = \mathrm{P}\left(\{\odot, \odot, \mathscr{H}^\circ\}\right) = \frac{1}{2}
$$

and

Random variable with a finite set of possible outcomes are called simple. In general, they may also be countable, in which case they are called discrete, or they may be continuous.

Induced Probability

Induced Probability

- How can we formalize in general the probability that assumes a specific value x ?
- Using the probability measure P already defined on S, we may obtain a new probability function P_X on the random variable X in $\mathbb{R}$ with the following procedure.
- For each $x \in \mathbb{R}$, let $S_x \subseteq S$ be the set containing just those elements of S which are mapped by X to numbers no greater than x, i.e., $S_x = \{s \in S \mid X(s) \leq x\}$.
- Then we write

$$
\mathrm{P}_X(X \leq x) \equiv \mathrm{P}\left(S_x\right)
$$

Support of a RV

- The image of S under X is called the support of X :

$$\text{supp}(X) \equiv X(S) = \{x \in \mathbb{R} \mid \exists s \in S \text{ s.t. } X(s) = x\}$$

- So as S contains all the possible outcomes of the experiment, $\text{supp}(X)$ contains all the possible outcomes for the random variable X.

- Thus, $\mathrm{P}_X(X \le x)$ is defined for all $x \in \text{supp}(X)$.

Example

Consider the experiment of tossing a fair coin, with sample space $\{\mathrm{H}, \mathrm{T}\}$ and probability measure $\mathrm{P}(\{\mathrm{H}\}) = \mathrm{P}(\{\mathrm{T}\}) = \frac{1}{2}$.

Suppose that we play a betting game where we win 1£ if we get heads, or we lose it otherwise.

- We can define a random variable $X : \{\mathrm{H}, \mathrm{T}\} \to \mathbb{R}$ taking values, say,

$$\begin{aligned} X(\mathrm{T}) &= -1, \\ X(\mathrm{H}) &= 1. \end{aligned}$$

- What does S_x look like for some $x \in \mathbb{R}$?

- The set S_x is defined by:

$$S_x = \begin{cases} \emptyset & \text{if } x < -1 \\ \{\mathrm{T}\} & \text{if } -1 \le x < 1 \\ \{\mathrm{H}, \mathrm{T}\} & \text{if } x \ge 1 \end{cases}$$

- This induces probabilities P_X on $\mathbb{R}$

$$\mathrm{P}_X(X \le x) = \mathrm{P}(S_x) = \begin{cases} \mathrm{P}(\emptyset) = 0 & \text{if } x < -1; \\ \mathrm{P}(\{\mathrm{T}\}) = \frac{1}{2} & \text{if } -1 \le x < 1; \\ \mathrm{P}(\{\mathrm{H}, \mathrm{T}\}) = 1 & \text{if } x \ge 1 \end{cases}$$

The key point for the theory is that with our definitions we can now associate a probability to every interval of $\mathbb{R}$, i.e., a random variable X has the intervals $(-\infty, x]$ as its events.

Cumulative Distribution Function

Cumulative distribution function

The cumulative distribution function (cdf) of a random variable X, written $F_X(x)$ (or just $F(x)$) is the probability that X takes a value less than or equal to x, i.e.,

$$F_X(x) = \mathrm{P}_X(X \leq x)$$

A cdf offers an alternative way to describe the probability measure P_X for a random variable X. It enables a unified treatment of discrete and continuous random variables.

For any random variable X, F_X is right-continuous, meaning if a decreasing sequence of real numbers $x_1, x_2, \ldots \rightarrow x$, then $F_X\left(x_1\right), F_X\left(x_2\right), \ldots \rightarrow F_X(x)$.

Properties of the cdf

To check a given function, $F_X(x)$, is a valid cdf, we need to verify the following conditions:
(1) Monotonicity: $\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F_X\left(x_1\right) \leq F_X\left(x_2\right)$;
(2) $F_X(-\infty) = 0, F_X(\infty) = 1$.
(3) F_X is right-continuous.

Note that the first two conditions imply $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$.

For finite intervals $(a, b] \subseteq \mathbb{R}$, it is possible to check that

$$\mathrm{P}_X(a < X \leq b) = F_X(b) - F_X(a)$$

This can be done after noting that the event $E = \{X \leq b\}$ may be written as the union $E = H \cup G$ of the disjoint events:

- $H = (-\infty, a]$
- $G = (a, b]$
  and the result follows from Axiom 3.

Comments

- A random variable is simply a numeric relabelling of our underlying sample space, and all probabilities are derived from the associated underlying probability measure.
- Unless there is any ambiguity, we generally suppress the subscript of $\mathrm{P}_X(\cdot)$ in our notation and just write $\mathrm{P}(\cdot)$ for the probability measure associated with a random variable.

- That is, we forget about the underlying sample space and just think about the random variable and its probabilities.
- Essentially, the support of X becomes our sample space S.
- Events for random variables are frequently of the kind $X = x$, $X > a, X \leq b, a \leq X \leq b, \ldots$