Random variables

David J. H. Shih

Intended learning outcomes

- Recognize and define events and random variables
- Implement code to sample values for a random variable

Event

Definition

Given a sample space $\mathcal{S} = \{s_1, \ldots, s_J\}$, an event $\mathcal{E}$ is a subset of the sample space $\mathcal{S}$: $\mathcal{E} \subseteq \mathcal{S}$.

$\mathcal{E}$ occurs if any $e_j \in \mathcal{E}$ is observed.
Given a probability function P,

$$
\begin{aligned}
P(\mathcal{E}) &= P\left(\bigcup_j^{\infty} e_j\right) \quad \text{(expansion)} \\
&= \sum_j^{\infty} P(e_j) \quad (\sigma\text{-additivity})
\end{aligned}
$$

Event

Further, if $P\left(s_i\right) = P\left(s_j\right)$ for all $s_i, s_j \in \mathcal{S}$, then

$$P\left(\mathcal{E}\right) = \frac{\left|\mathcal{E}\right|}{\left|\mathcal{S}\right|},$$

where $\left|\mathcal{A}\right|$ is the number of elements in a set $\mathcal{A}$.

Example

Sequence: Toss a coin 2 times.

Define event $\mathcal{E}$ as the occurrence of one head:

$$\mathcal{E} = \{HT, TH\}$$

Derive $P(\mathcal{E})$.

Randomness

Apparent or actual lack of definite pattern in information.

Compatible with a deterministic or stochastic process.

Examples

- flipping a coin
- drawing cards from a shuffled deck
- atmospheric noise (random.org)
- somatic mutations

Random variable

Definition

Given a sample space $\mathcal{S} = \{s_1, ..., s_J\}$ and a probability function P, a **random variable** X is a function that maps from $\mathcal{S}$ onto real numbers with domain $\mathcal{X} = \{x_1, ..., x_N\}$.

Illustration

Random variable

Each realization $x_i \in \mathcal{X}$ corresponds to an event $\mathcal{E}_i \subseteq \mathcal{S}$ such that

$$P_X\left(X = x_i\right) \;=\; P\left(\mathcal{E}_i\right),$$

where $\mathcal{E}_i = \{s_j \in S : X(s_j) = x_i\}$ and P_X is the *induced* probability function on $\mathcal{X}$.

Remark: We often abbreviate $P_X\left(X = x_i\right)$ as $P_X\left(x_i\right)$ or $P\left(x_i\right)$.

Example 1

Sequence: Toss a coin 1 time.

Define random variable X as the number of heads.

What event does each possible value of X correspond to?

x_i	$\mathcal{E}_i = \{s_j \in S : X(s_j) = x_i\}$

Example 2

Sequence: Toss a coin 2 times.

Define random variable X as the number of heads.

What event does each possible value of X correspond to?

x_i	$\mathcal{E}_i = \{s_j \in S : X(s_j) = x_i\}$

Random variable

Previously, we assumed that sample space $\mathcal{S} = \{s_1, \dots, s_J\}$, i.e. $\mathcal{S}$ is (countable and) finite.

We also assumed that $\mathcal{X} = \{x_1, \dots, x_N\}$ is finite.

If $\mathcal{X}$ is uncountable, we can define the induced probability function for some set $\mathcal{A} \subset \mathcal{X}$ as

$$P_X\left(X \in \mathcal{A}\right) \;=\; P\left(\{s \in \mathcal{S} : X(s) \in \mathcal{A}\}\right).$$

Cumulative distribution function

Definition

The cumulative distribution function (cdf) of a random variable X is defined by

$$F_X(x) \triangleq P_X(X \leq x), \quad \text{for all } x.$$

Theorem

A function $F(x)$ is a cdf if and only if

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

b. $F(x_1) \leq F(x_2)$ for all $x_1 \leq x_2$ (non-decreasing).

c. $\lim_{x \to x_0^+} F(x) = F(x_0)$ (right-continuous).

Examples of cdf

continuous function

step function

Probability mass function

Definition

The probability mass function (pmf) of a discrete random variable X is given by

$$f_X(x) \triangleq P_X(X = x) \quad \text{for all } x.$$

Example

Sequence: Toss a fair coin 3 independent times.

Define random variable X as the number of heads.

Calculate $f_X(x_i) = P_X\left(X = x_i\right)$ for all $x_i \in X$.

x_i	$\mathcal{E}_i$	$f_X(x_i)$

$f_X(x)$ for continuous random variable?

For a continuous random variable X, we might try to define

$$f_X(x) = P_X(X = x).$$

However, $P_X(X = x) = 0$ for all $x \in \mathcal{X}$ here.

So, we need to define $f_X(x)$ differently for continuous random variables.

$P_X(X = x)$ for continuous random variable

For a continuous random variable X, $P_X(X = x) \ = \ 0, \forall\, x \in \mathcal{X}$.

Illustration

$P_X(X = x)$ for continuous random variable

For a continuous random variable X, $P_X(X = x) = 0, \forall\, x \in \mathcal{X}$.

Proof

Recall from set theory that

$$\begin{aligned} &(1) \quad A \subseteq B \;\Rightarrow\; P(A) \le P(B) \\ &(2) \quad P(A \cap B) = P(B) - P(B \cap A^c) \end{aligned}$$

For any $\epsilon > 0$, we have

$$\{r \in \mathcal{X} : r = x\} \subseteq \{r \in \mathcal{X} : x - \epsilon < r \le x\}.$$

As short-hand, we will write

$$\{X = x\} \;\subseteq\; \{x - \epsilon < X \le x\}.$$

By (1),

$$P\left(X = x\right) \;\le\; P\left(x - \epsilon < X \le x\right).$$

By definition,

$$\begin{aligned} P\left(x-\epsilon < X \leq x\right) &= P\left(\{r \in \mathcal{X} : x-\epsilon < r\} \cap \{r \in \mathcal{X} : r \leq x\}\right) \\ &= P\left(\{x-\epsilon < X\} \cap \{X \leq x\}\right) \end{aligned}$$

By (2),

$$\begin{aligned} & P\left(\{x-\epsilon < X\} \cap \{X \leq x\}\right) \\ &= P\left(\{X \leq x\}\right) - P\left(\{X \leq x\} \cap \{X \leq x-\epsilon\}\right) \\ &= P\left(\{X \leq x\}\right) - P\left(\{X \leq x-\epsilon\}\right) \\ &= P\left(X \leq x\right) - P\left(X \leq x-\epsilon\right) \\ &= F_X(x) - F_X(x-\epsilon) \end{aligned}$$

Therefore,

$$0 \leq P(X = x) \leq \lim_{\epsilon \to 0+} F_X(x) - F_X(x-\epsilon) = 0. \quad \blacksquare$$

Probability density function

Definition

The probability density function (pdf) of a continuous random variable is a function $f_X(x)$ that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \text{for all } x.$$

Further, if $f_X(x)$ is continuous, by the Fundamental Theorem of Calculus,

$$\frac{d}{dx}F_X(x) = f_X(x).$$

Notation for sampling

If a random variable X has a cdf $F_X(x)$, we write

$$X \sim F_X(x),$$

where the $\sim$ (tilde) symbol means "is distributed according to".

The right-hand-side of the $\sim$ operator can be anything that help defines a probability distribution.

For instance, if X has a pmf or pdf $f_X(x)$, we can also write

$$X \sim f_X(x).$$

If randomv variables X and Y have the same distribution, we write

$$X \sim Y.$$

Summary

"Math is not magic." - High school math teacher

Casella & Berger 2002, sections 1.4, 1.5, 1.6.

Intended learning outcomes

- Recognize and define events and random variables
- Implement code to sample values for a random variable