Theory of Probability

Aleksandr Lenin

Sample space

Set Ω is called the *sample space* – a collection of *all possible outcomes* of a random experiment.

Requirement on Ω: a given instance of the experiment must produce a result $\omega \in \Omega$ corresponding to *exactly one* of the elements in Ω.

$\Omega = \{H, T\}$ – a coin tossed once may show either heads or tails.

$\Omega = \{HH, HT, TH, TT\}$ – same coin, but tossed twice. Two tosses of a coin correspond to one experiment

$\Omega = \{1, 2, 3, \ldots, 6\}$ – single throw of a dice.

Event

Events are subsets $A \subseteq \Omega$. In our case, all subsets are events.

$A = \{\text{the number of heads} \leqslant 1\} = \{HT, TH, TT\}$.

$B = \{\text{1st toss} = \text{2nd toss}\} = \{HH, TT\}$.

$C = \{\text{the outcome of a die is even}\} = \{2, 4, 6\}$.

An event A happens if $\omega \in A$.

Ω – a certain event (it always happens).

$\emptyset$ – an impossible event (it never happens).

Event Algebra

$\bigcup_{i=1}^{n} A_i$ – a union of n events – a set of elements belonging to at least one of the sets A_i.

$\bigcap_{i=1}^{n} A_i$ – an intersection of n events – a set of elements belonging to all sets A_i.

$\bar{A}$ – complement of A – a set of elements which do not belong to A.

Event Algebra

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

$$A = \{\text{dice outcome is even}\} = \{2, 4, 6\} \ .$$

$$B = \{\text{dice outcome} \geqslant 3\} = \{3, 4, 5, 6\} \ .$$

$$A \cup B = \{2, 3, 4, 5, 6\} \ .$$

$$A \cap B = \{4, 6\} \ .$$

$$\bar{A} = \{1, 3, 5\} \ .$$

$$\bar{B} = \{1, 2\} \ .$$

Events A and B are *mutually exclusive* if $A \cup B = \emptyset$.

Sigma algebra

A sigma-algebra $\mathcal{F}$ is a collection of subsets of Ω, satisfying the following requirements:

1. $\Omega \in \mathcal{F}$
2. $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} A_n \in \mathcal{F}$
3. $A \in \mathcal{F} \Longrightarrow \bar{A} \in \mathcal{F}$

Consequently: $\emptyset \in \mathcal{F}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

In finite Ω we shall be able to take $\mathcal{F}$ as a powerset of Ω, and $\mathcal{F}$ itself is a set of all events.

Any subset $A \subseteq \mathcal{F}$ is called a *measurable* set.

Probability

A probability (measure) is a function $\mathrm{P} : A \to \mathbb{R}$ such that:

- $0 \leqslant \mathrm{P}[A] \leqslant 1$ for every event $A \in \mathcal{F}$
- $\mathrm{P}[\Omega] = 1$
- If $\{A_i\}_{i \in \mathbb{N}}$ are mutually exclusive ($A_i \cap A_j = \emptyset$ for $i \neq j$), then $\mathrm{P}[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathrm{P}[A_i]$.

The triplet $(\Omega, \mathcal{F}, \mathrm{P})$ is called the *probability space*. It is a measure space with the probability function as a measure in this space.

If $\mathcal{F}$ is the powerset of Ω, we omit $\mathcal{F}$ and say that a probability space is a pair (Ω, P).

Learning and Conditional Probability

Somehow we learn that event B (with $\mathrm{P}[B] \neq 0$) happens, i.e. $\omega \in B$.
We may consider *learning* as a process where the probability space (Ω, P) is changing to a new probability space (Ω', P'), where $\Omega' = B$.
We want that there is β, so that $\mathrm{P}'[A] = \beta \cdot \mathrm{P}[A \cap B]$ for any event A.
As in the new space $\mathrm{P}'[B] = \mathrm{P}'[\Omega'] = 1$, we have $\beta = \frac{1}{\mathrm{P}[B \cap B]} = \frac{1}{\mathrm{P}[B]}$, i.e.

$$\mathrm{P}'[A] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]} \quad .$$

The probability $\mathrm{P}'[A]$ is denoted by $\mathrm{P}[A|B]$ and is called the *conditional probability* of A assuming that B happens.

Learning and Conditional Probability

$\Omega = \{1, 2, 3, 4, 5, 6\}$ for a dice.

$A = \{\text{the outcome is even}\}$.

$B = \{\text{the outcome is 2}\}$.

$\mathrm{P}[B] = \frac{1}{6}$.

$\mathrm{P}[A] = \frac{3}{6} = \frac{1}{2}$.

$\mathrm{P}[B|A] = \frac{\mathrm{P}[B \cap A]}{\mathrm{P}[A]} = \frac{\mathrm{P}[B]}{\mathrm{P}[A]} = \frac{1}{3}$.

The probability of outcome 2 given that the outcome was even, is $\frac{1}{3}$.

The Chain rule and the Bayes formula

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \implies P[A \cap B] = P[A|B] \cdot PB$$
$$P[B|A] = \frac{P[A \cap B]}{P[A]} \implies P[A \cap B] = P[B|A] \cdot PA$$

this is known as the *chain rule*. Hence,

$$P[A \cap B] = P[A|B] \cdot P[B] = P[B|A] \cdot PA \quad ,$$

and we obtain the Bayes rule:

$$P[A|B] = \frac{P[B|A] \cdot PA}{P[B]} \quad .$$

Random Variables and Probability Distributions

Random variable X is any function $X : \Omega \Rightarrow R$, where R is called the *range* of X.
For any $x \in R$, we define $X^{-1}(x)$ as the event $\{\omega : X(\omega) = x\}$ and use the notation:

$$\Pr_X[x] = \Pr[X = x] = \Pr[X^{-1}(x)] \ .$$

Note that if $x \neq x'$ then the events $X^{-1}(x)$ and $X^{-1}(x')$ are mutually exclusive.

Random Variables and Probability Distributions

As $\bigcup_{x \in R} X^{-1}(x) = \Omega$, we have:

$$\sum_x \mathrm{P}_X[x] = \mathrm{P}\left[\bigcup_{x \in R} X^{-1}(x)\right] = \mathrm{P}[\Omega] = 1 \quad .$$

If $R = \{x_1, x_2, \ldots, x_n\}$, then the sequence of values $(p_1, p_2, \ldots, p_n)$, where $p_i = \mathrm{P}_X[x_i]$, is called the *probability distribution* of X.

Independent Events and Random Variables

Events A and B are said to be independent if $\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B]$.

If $\mathrm{P}[A] \neq 0 \neq \mathrm{P}[B]$, then independence is equivalent to:

$$\mathrm{P}[A|B] = \mathrm{P}[A] \qquad \mathrm{P}[B|A] = \mathrm{P}[B] \ ,$$

so that the probability of A does not change, if we learn that B happened.

Direct Product of Random Variables

The *direct product* of random variables X and Y on a probability space (Ω, P) is a random variable defined by

$$(XY)(\omega) = (X(\omega),\, Y(\omega)) \ .$$

Independent Events and Random Variables

We say that X and Y are independent random variables if for every $x \in R_X$ and $y \in R_Y$:

$$\begin{aligned} \mathrm{P}[X = x, Y = y] &= \mathrm{P}[X^{-1}(x) \cap Y^{-1}(y)] \\ &= \mathrm{P}[X^{-1}(x)] \cdot \mathrm{P}[Y^{-1}(y)] \\ &= \mathrm{P}[X = x] \cdot \mathrm{P}[Y = y] \ . \end{aligned}$$

This means that the knowledge of the value of Y does not influence the probability distribution of X.

Mean of a Random Variable

By *mean* or *expected value* μ of a random variable X we mean the sum

$$\mu = \sum \omega \cdot X(\omega) \ \ .$$

ω :	1	2	3	4	5	6
$X(\omega)$:	0.1	0.2	0.3	0.2	0.1	0.1

$$\mu = E(X) = 0.1 + 0.4 + 0.9 + 0.8 + 0.5 + 0.6 = 3.3 \ \ .$$

Dispersion of a Random Variable

Dispersion measures the extent to which the distribution is extended or squeezed. Common examples of statistical dispersion are:

The *variance* $Var(X)$ is the sum:

$$Var(X) = E[(X - \mu)^2] = \sum X(\omega)(\omega - \mu)^2 \ .$$

The *standard deviation* is the square root of the variance:

$$\sigma = \sqrt{Var(X)} \ .$$

Ordered samples of size r, with replacement

The number of ordered samples of size r with replacement from n objects

$$n \times n \times \ldots = n^r$$

The number of possible outcomes if 3 dices are thrown is $6 \times 6 \times 6 = 216$.

Ordered samples of size r, without replacement

The number of ordered samples of size r without replacement from n objects is:

$$(n)_r = n \cdot (n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!} \quad ,$$

where $\quad n = 1, 2, \ldots, n.$

The number of 3 digit numbers that can be formed from $1, 2, \ldots, 9$, if no digit can be repeated, is $9 \cdot 8 \cdot 7 = 504$.

Unordered samples of size r, without replacement

The number of unordered samples of size r without replacement from of n objects is:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

How many distinct sequences can we make using 3 letter "A"s and 5 letter "B"s? (AAABBBBB, AABABBBB, etc.)

You have 8 positions total, 3 of them for As, 5 for Bs. The total number of ways is

$$\binom{8}{3} = \binom{8}{5}$$

Unordered samples of size r, with replacement

The number of ways to place r indistinguishable objects into n distinct sells is:

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

10 passengers get on an airport shuttle which stops near 5 hotels, each passengers gets off the shuttle at his/her hotel. How many possibilities exist?

$$\binom{5+10-1}{10} = \binom{5+10-1}{5-1} = \binom{14}{4}$$