Probability, Statistics

Probability theory provides models of probability distributions (theoretical models of the observable reality involving chance effects) to be tested by statistical methods, and it will also supply the mathematical foundation of these methods.

Modern **mathematical statistics** has various engineering applications, for instance, in testing materials, control of production processes, quality control of production outputs, performance tests of systems, robotics, and automatization in general, production planning, marketing analysis, and so on.

Although these applications are very heterogeneous, we will see that most statistical methods are universal in the sense that each of them can be applied in various fields.

Probability Theory

If some numerical data are influenced by "chance," by factors whose effect we cannot predict exactly (e.g., weather data, stock prices, life spans of tires, etc.), we have to rely on **probability theory**.

It gives mathematical models of chance processes called *random experiments* or, briefly, **experiments**. In such an experiment we observe **a random variable** **X**, that is, a function whose values in a **trial** (a performance of an experiment) occur "by chance" according to a **probability distribution** that gives the individual probabilities with which possible values of **X** may occur in the long run.

Experiments, Outcomes, Events

An **experiment** is a process of measurement or observation, in a laboratory, in a factory, on the street, in nature, or wherever; so “experiment” is used in a rather general sense. Our interest is in experiments that involve **randomness**, chance effects, so that we cannot predict a result exactly.

A **trial** is a single performance of an experiment. Its result is called an **outcome** or a **sample point**. **n** trials then give a **sample** of **size** **n** consisting of **n** sample points.

The **sample space** **S** of an experiment is the set of all possible outcomes. The subsets of **S** are called **events** and the outcomes **simple events**. If, in a trial, an outcome a happens and **$a \in A$** (**a** is an **element** of **A**), we say that **A** happens.

For instance, if a die turns up a 3, the event A: Odd number happens

Unions, Intersections, Complements of Events

The **union** $A \cup B$ of A and B consists of all points in A or B or both.

The **intersection** $A \cap B$ of A and B consists of all points that are in both A and B.

If A and B have no points in common, we write

$$A \cap B = \emptyset$$

where $\emptyset$ is the **empty set** (set with no elements) and we call A and B mutually **exclusive** (or **disjoint**) because, in a trial, the occurrence of A excludes that of B (and conversely).

If your die turns up an odd number, it cannot turn up an even number in the same trial. Similarly, a coin cannot turn up Head and Tail at the same time.

Unions, Intersections, Complements of Events

Complement $\mathbf{A^c}$ of **A**. This is the set of all the points of **S** ***not*** in **A**. Thus

$$\mathbf{A \cap A^c = \emptyset}, \qquad \mathbf{A \cup A^c = C}$$

Another notation for the complement of **A** is $\overline{\mathbf{A}}$ (instead of $\mathbf{A^c}$) but we will not use this because in set theory is used to denote the closure of **A**.

Unions and intersections of more events are defined similarly.

Unions, Intersections, Complements of Events

The **union**

$$\bigcup_{j=1}^{m} A_j = A_1 \cup A_2 \cup \cdots \cup A_m$$

of events $A_1, A_2 \ldots A_m$ consists of all points that are in at least one A_j. Similarly for the union $A_1 \cup A_2 \cup \cdots$ of infinitely many subsets $A_1, A_2 \ldots A_m$ of an infinite sample space S (that is, S consists of infinitely many points).

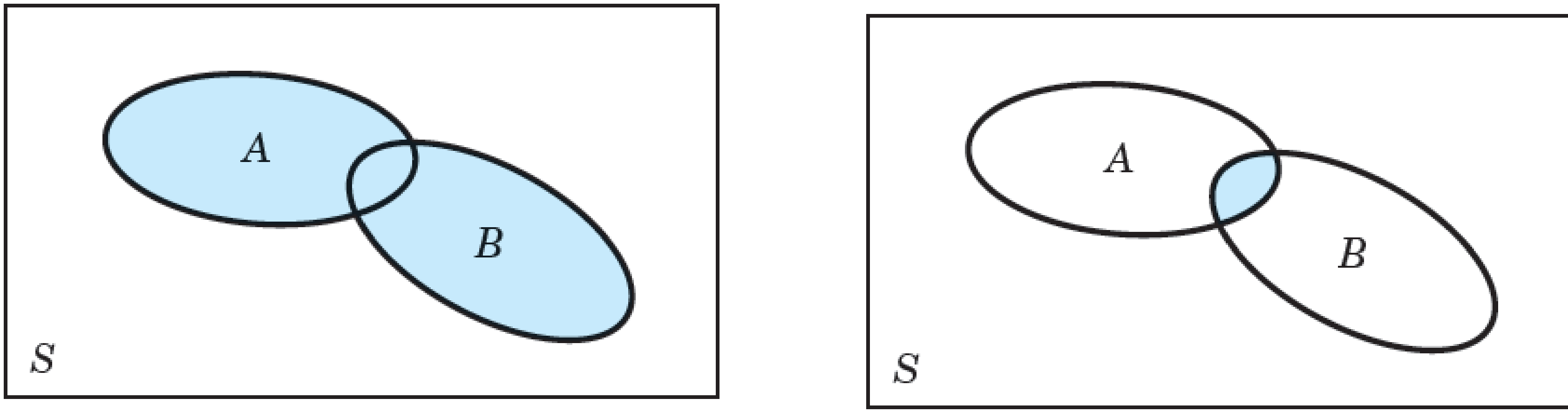
The **intersection**

$$\bigcap_{j=1}^{m} A_j = A_1 \cap A_2 \cap \cdots \cap A_m$$

of $A_1, A_2 \ldots A_m$ consists the points of S that are in each of these events. Similarly for the intersection $A_1 \cap A_2 \cap \cdots$ of infinitely many subsets of S.

Unions, Intersections, Complements of Events

Working with events can be illustrated and facilitated by **Venn diagrams** for showing unions, intersections, and complements, as in Figs., which are typical examples



Union $A \cup B$ Intersection $A \cap B$

Fig. 1. Venn diagrams showing two events **A** and **B** in a sample space **S** and their union **A ∪ B** (colored) and intersection **A ∩ B** (colored)

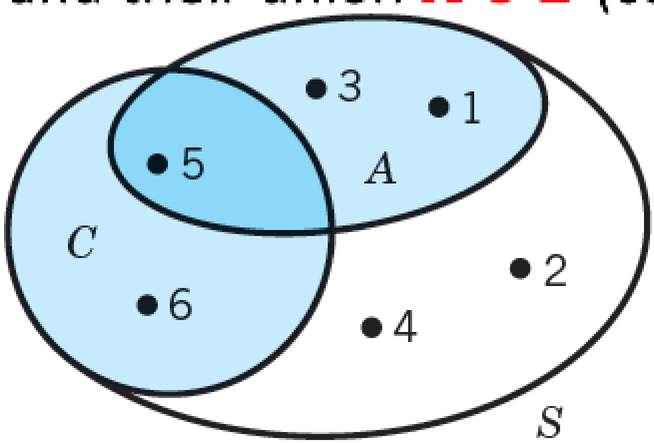


Fig. 2. Venn diagram for the experiment of rolling a die, showing **S**, **A** = {1, 3, 5}, **C** = {5, 6}, **A ∪ C** {1, 3, 5, 6}, **A ∩ C** {5}

Probability

The "**probability**" of an event **A** in an experiment is supposed to measure <u>how frequently **A** is about to occur</u> if we make many trials.

*If we flip a coin, then heads H and tails T will appear about equally often—we say that H and T are "**equally likely**."*

*Similarly, for a regularly shaped die of homogeneous material ("**fair die**") each of the six outcomes will be equally likely.*

Definition of Probability

If the sample space **S** of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability of an event **A** is

$$\mathbf{P(A) = \frac{Number\ of\ point\ in\ A}{Number\ of\ point\ in\ S}}$$

Axioms of probability

Given a sample space **S**, with each event **A** of **S** (subset of **S**) there is associated a number called the probability of **A**, such that the following axioms of probability are satisfied.

1. For every **A** in **S**,

$$\mathbf{0 \le P(A) \le 1.}$$

2. The entire sample space **S** has the probability

$$\mathbf{P(S) = 1.}$$

3. For mutually exclusive events **A** and **B** ($\mathbf{A \cap B = \emptyset}$)

$$\mathbf{P(A \cup B) = P(A) + P(B)}$$

Basic Theorems of Probability

T H E O R E M 1. Complementation Rule

For an event **A** and its complement in a sample space **S**,

$$\mathbf{P(A^c) = 1 - P(A).}$$

T H E O R E M 2. Addition Rule for Mutually Exclusive Events

For mutually exclusive events $\mathbf{A_1, \ldots, A_m}$ in a sample space **S**,

$$\mathbf{P(A_1 \cup A_2 \cup \cdots A_m) = P(A_1) + P(A_2) + \cdots + P(A_m).}$$

T H E O R E M 3. Addition Rule for Arbitrary Events

For events **A** and **B** in a sample space **S**,

$$\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B).}$$

Conditional Probability

Often it is required to find the probability of an event **B** under the condition that an event **A** occurs. This probability is called the **conditional probability** of **B** given **A** and is denoted by **P(B|A)**. In this case **A** serves as a new (reduced) sample space, and that probability is the fraction **P(A)** of which corresponds to $\mathbf{A \cap B}$. Thus

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

Similarly, the conditional probability of **A** given **B** is

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

T H E O R E M 4. Multiplication Rule. If **A** and **B** are events in a sample space **S** and $\mathbf{P(A) \neq 0, P(B) \neq 0}$, then

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

Independent Events

Independent Events. If events B and A are such that

$$P(A \cap B) = P(A)P(B)$$

they are called **independent events**. Assuming $P(A) \neq 0, P(B) \neq 0$,

$$P(A|B) = P(A), \qquad P(B|A) = P(B).$$

This means that the probability of A does not depend on the occurrence or nonoccurrence of B, and conversely. This justifies the term "independent."

Independence of m Events. Similarly, m events $A_1, \dots, A_m$ are called **independent** if

$$P(A_1 \cap \cdots \cap A_m) = P(A_1) \dots P(A_m).$$

Random Variables. Probability Distributions

A **probability distribution** or, briefly, a **distribution**, shows the probabilities of events in an experiment. The quantity that we observe in an experiment will be denoted by **X** and called a **random variable** (or **stochastic variable**) because the value it will assume in the next trial depends on chance, on **randomness**.

If you roll a die, you get one of the numbers from 1 to 6, but you don't know which one will show up next. Thus Number a die turns up is a random variable.

Random Variables. Probability Distributions

If we ***count*** (cars on a road, defective screws in a production, tosses until a die shows the first Six), we have a **discrete random variable and distribution**.
If we ***measure*** (electric voltage, rainfall, hardness of steel), we have a **continuous random variable and distribution**. Precise definitions follow. In both cases the distribution of **X** is determined by the **distribution function** (or **cumulative distribution function**)

$$\mathbf{F(x) = P(X \leq x);}$$

this is the probability that in a trial, **X** will assume any value not exceeding **x**.

Random Variable

A **random variable** **X** is a function defined on the sample space **S** of an experiment. Its values are real numbers. For every number **a** the probability

$$\mathbf{P(X = a)}$$

with which **X** assumes **a** is defined. Similarly, for any interval **I** the probability

$$\mathbf{P(X \in I)}$$

with which **X** assumes any value in **I** is defined.

The fundamental formula for the probability corresponding to an interval $\mathbf{a < x \le b}$

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

Discrete Random Variables and Distributions

By definition, a random variable $\mathbf{X}$ and its distribution are **discrete** if $\mathbf{X}$ assumes only finitely many or at most countably many values $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \ldots$, called the **possible values** of $\mathbf{X}$, with positive probabilities

$\mathbf{p_1 = P(X = x_1), p_2 = P(X = x_2), p_3 = P(X = x_3), \ldots}$,

whereas the probability $\mathbf{P(X \in I)}$ is zero for any interval $\mathbf{I}$ containing no possible value.

Discrete Random Variables and Distributions

Clearly, the discrete distribution of $\mathbf{X}$ is also determined by the **probability function** $\mathbf{f(x)}$ of $\mathbf{X}$, defined by

$$\mathbf{f(x)} = \begin{cases} \mathbf{p_j} & \mathbf{if\ x = x_j} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \quad (\mathbf{j = 1, 2, \dots})$$

From this we get the values of the **distribution function** by taking sums,

$$\mathbf{F(x)} = \sum_{\mathbf{x_j \le x}} \mathbf{f(x_j)} = \sum_{\mathbf{x_j \le x}} \mathbf{p_j}$$

where for any given $\mathbf{x}$ we sum all the probabilities $\mathbf{p_j}$ for which $\mathbf{x_j}$ is smaller than or equal to that of $\mathbf{x}$. This is a **step function** with upward jumps of size $\mathbf{p_j}$ at the possible values $\mathbf{x_j}$ of $\mathbf{X}$ and constant in between.

Discrete Random Variables and Distributions

For the probability corresponding to intervals we have

$$\mathbf{P(a < X \leq b) = F(b) - F(a) = \sum_{a<x_j\leq b} p_j}\,.$$

This is the sum of all probabilities $\mathbf{p_j}$ for which $\mathbf{x_j}$ satisfies $\mathbf{a < x_j \leq b}$.

From this and **P(S)=1** we obtain the following formula (sum of all probabilities)

$$\mathbf{\sum_j p_j = 1}.$$

Continuous Random Variables and Distributions

Discrete random variables appear in experiments in which we **count** (defectives in a production, days of sunshine in Chicago, customers standing in a line, etc.). Continuous random variables appear in experiments in which we **measure** (lengths of screws, voltage in a power line, Brinell hardness of steel, etc.). By definition, a random variable X and its distribution are of continuous type or, briefly, **continuous**, if its **distribution function** can be given by an integral

$$\mathbf{F(x)} = \int_{-\infty}^{\mathbf{x}} \mathbf{f}(v)\,dv$$

whose integrand called the **density** of the distribution, is nonnegative, and is continuous, perhaps except for finitely many **x**-values. Differentiation gives the relation of **f** to **F** as

$$\mathbf{f(x) = F'(x)}$$

for every **x** at which **f(x)** is continuous.

Discrete Random Variables and Distributions

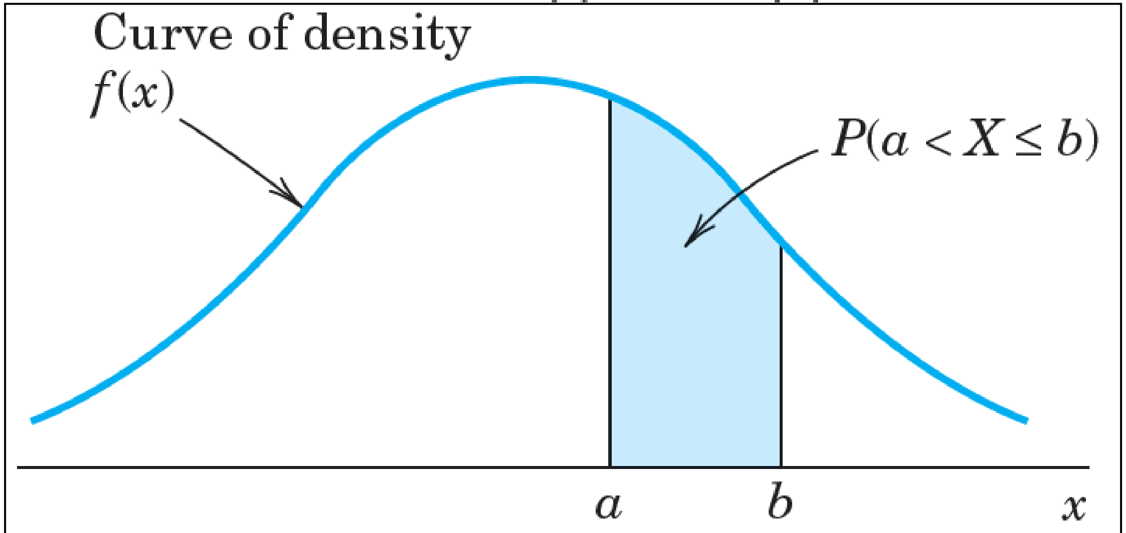
For the probability corresponding to intervals we have

$$\mathbf{P(a < X \le b) = F(b) - F(a) = \int_a^b f(v)\,dv}.$$

From this and **P(S)=1** we obtain the following formula (sum of all probabilities)

$$\int_{-\infty}^{+\infty} \mathbf{f(v)\,dv = 1}.$$

The next example illustrates notations and typical applications of our present formulas.



Discrete Random Variables and Distributions

The **mean μ** characterizes the central location and the variance **σ** the spread (the variability) of the distribution. The mean **μ** (mu) is defined by

$$\mu = \sum_j \mathbf{x_j f(x_j)} \qquad (for\ discrete\ distribution)$$

$$\mu = \int_{-\infty}^{+\infty} \mathbf{x\, f}(x)\,dx \qquad (for\ continuous\ distribution)$$

and the **variance** (sigma square) by

$$\boldsymbol{\sigma}^2 = \sum_j (\mathbf{x_j} - \mu)^2 \mathbf{f(x_j)} \qquad (for\ discrete\ distribution)$$

$$\boldsymbol{\sigma}^2 = \int_{-\infty}^{+\infty} (\mathbf{x} - \mu)^2 \mathbf{f}(x)\,dx \qquad (for\ continuous\ distribution)$$

Discrete Random Variables and Distributions

σ (the positive square root of **σ^2**) is called the **standard deviation** of **X** and its distribution.

The mean **μ** is also denoted by **$E(X)$** and is called the **expectation** *of* **X** because it gives the average value of **X** to be expected in many trials. Quantities such as **μ** and **σ^2** that measure certain properties of a distribution are called **parameters**. **μ** and **σ^2** are the two most important ones.

$$\sigma^2 > 0$$

(except for a discrete “distribution” with only one possible value, so that $\sigma^2 = 0$). We assume that **μ** and **σ^2** exist (are finite), as is the case for practically all distributions that are useful in applications.

Theorems of Mean and Variance

T H E O R E M 1. Mean of a Symmetric Distribution

If a distribution is **symmetric** with respect to **$x = c$**, that is, **$f(c-x) = f(c+x)$**, then **$\mu = c$**

T H E O R E M 2. Transformation of Mean and Variance

If a random variable X has mean **μ** and variance **σ^2**, then the random variable

$$\mathbf{X^* = a_1 + a_2 X \qquad (a_2 > 0).}$$

has the mean **μ^*** and variance **σ^{*2}**, where

$$\mu^* = a_1 + a_2\mu \qquad \text{and} \qquad \sigma^{*2} = a_2{}^2\sigma^2.$$

In particular, the **standardized random variable** Z corresponding to X, given by

$$Z = \frac{X - \mu}{\sigma}$$

has the mean 0 and the variance 1.

Expectation, Moments

If **g(x)** is nonconstant and continuous for all **x**, then **g(X)** is a random variable. Hence its *mathematical expectation* or, briefly, its **expectation E(g(x))** is the value of g(x) to be expected on the average, defined

$$\mathbf{E(g(x))} = \sum_{j} \mathbf{g(x_j)\, f(x_j)} \quad \text{or} \quad \boldsymbol{\mu} = \int_{-\infty}^{+\infty} \mathbf{g(x)\, f(x)\, dx}$$

In the first formula, **f** is the probability function of the discrete random variable **X**. In the second formula, **f** is the density of the continuous random variable **X**. Important special cases are the **k-th moment** of **X** (where **k** = 1, 2, ...)

$$\mathbf{E(X^k)} = \sum_{j} x^k{}_j\, \mathbf{f(x_j)} \quad \text{or} \quad \boldsymbol{\mu} = \int_{-\infty}^{+\infty} x^k \mathbf{f(x)\, dx}$$

Expectation, Moments

and the **k-th central moment** of **X** (**k = 1, 2, …)**

$$\mathbf{E}\left([\mathbf{X}-\boldsymbol{\mu}]^{\mathbf{k}}\right)=\sum_{\mathbf{j}}\left(\mathbf{x_j}-\boldsymbol{\mu}\right)^{k}\mathbf{f}(\mathbf{x_j})\ \text{or}\ \boldsymbol{\mu}=\int_{-\infty}^{+\infty}(x-\boldsymbol{\mu})^{k}\mathbf{f}(\mathbf{x})\mathbf{dx}$$

This includes the first moment, the **mean** of **X**

$$\boldsymbol{\mu}=\mathbf{E}(\mathbf{X})$$

It also includes the second central moment, the **variance** of **X**

$$\boldsymbol{\sigma}^{2}=\mathbf{E}\left([\mathbf{X}-\boldsymbol{\mu}]^{2}\right)$$

END OF LECTURE-1