

Chapter 1: Basic Concepts of Probability

DEFINITION 1 (EXPERIMENT, SAMPLE SPACE, EVENT)

A **statistical experiment** is any procedure that produces data or observations.

The **sample space**, denoted by S , is the set of **all possible outcomes of a statistical experiment**. The sample space depends on the problem of interest!

A **sample point** is an **outcome (element)** in the sample space.

An **event** is a **subset of the sample space**.

Union: $A \cup B = x : x \in A \vee x \in B$

Union of n events: $\cup_{i=1}^n A_i = x : x \in A_1 \vee x \in A_2 \dots$

Union: $A \cap B = x : x \in A \wedge x \in B$

Intersection of n events: $\cap_{i=1}^n A_i = x : x \in A_1 \wedge x \in A_2 \dots$

Complement: $A' = x : x \in S \wedge x \notin A$

Mutually Exclusive: $A \cap B = \emptyset$

MORE EVENT OPERATIONS

(a) $A \cap A' = \emptyset$ (b) $A \cap \emptyset = \emptyset$

(c) $A \cup A' = S$ (d) $(A')' = A$

(e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(g) $A \cup B = A \cup (B \cap A')$

(h) $A = (A \cap B) \cup (A \cap B')$

DE MORGAN'S LAW

For any n events A_1, A_2, \dots, A_n ,

(i) $(A_1 \cup A_2 \cup \dots \cup A_n)' = A_1' \cap A_2' \cap \dots \cap A_n'$

A special case: $(A \cup B)' = A' \cap B'$

(j) $(A_1 \cap A_2 \cap \dots \cap A_n)' = A_1' \cup A_2' \cup \dots \cup A_n'$

A special case: $(A \cap B)' = A' \cup B'$

Multiplication

Sequential experiments
 $n_1 * n_2 * \dots * n_r$
outcomes

Addition

Independent experiments
 $n_1 + n_2 + \dots + n_r$
outcomes

Permutation: $P_r^n = \frac{n!}{(n-r)!} = n(n-1)\dots(n-(r-1))$

Selection and arrangement of r objects out of n . Order is considered.

Combination: $C_r^n = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\dots(n-(r-1))}{r!}$

Selection of r objects out of n . Order is not considered.

$C_r^n = C_{n-r}^n$

Axioms and Properties of Probability

1 $0 \leq P(A) \leq 1$

2 $P(S) = 1$

3 $P(A \cup B) = P(A) + P(B)$ if A and B are **mutex** (not to be confused with independence)

4 $P(\emptyset) = 0$

5 $P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$

6 $P(A') = 1 - P(A)$

7 $P(A \cup B) = P(A) + P(B) - P(AB)$

8 $P(A) = P(A \cap B) + P(A \cap B')$

9 $A \subset B \rightarrow P(A) \leq P(B)$

10 $P(A_1) = P(A_2) = \dots = P(A_k) \rightarrow \text{for } B \subset$

$$S, P(B) = \frac{|B|}{|S|}$$

Conditional Probability (B occurs given A):

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

Multiplication rules

$P(A \cap B) = P(A)P(B|A)$ if $P(A) \neq 0$

$P(A \cap B) = P(B)P(A|B)$ if $P(B) \neq 0$

Inverse Probability Formula: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

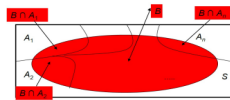
Independence $A \perp B$ if $P(A \cap B) = P(A)P(B)$ This implies $P(A|B) = P(A)$ and $P(B|A) = P(B)$

Partition A_1, A_2, \dots, A_r are mutually exclusive and $\sum_{i=1}^n A_i = S$

THEOREM 11 (LAW OF TOTAL PROBABILITY)

Suppose A_1, A_2, \dots, A_n is a partition of S . Then for any event B , we have

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i).$$



$$P(B) = P(A)P(B|A) + P(A')P(B|A')$$

Bayes' theorem

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B|A)P(A) + P(B|A')P(A')}$$

$$\text{Bayes' Theorem}(n=2) P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

Denom is $P(B)$

$$P(A|B) = 1 - P(A'|B)$$

L2: Random Variables

Random variable X is a function from S to R

Uppercase letters denote random variables

Lowercase letters denote observed values of random variables

DEFINITION 3 (PROBABILITY MASS FUNCTION)

For a discrete random variable X , define

$$f(x) = \begin{cases} P(X=x), & \text{for } x \in R_X; \\ 0, & \text{for } x \notin R_X. \end{cases}$$

Then $f(x)$ is known as the **probability function (pf)**, or **probability mass function (pmf)** of X .

The collection of pairs $(x_i, f(x_i)), i = 1, 2, 3, \dots$, is called the **probability distribution** of X .

PROPERTIES OF THE PROBABILITY MASS FUNCTION

The probability mass function $f(x)$ of a discrete random variable **must** satisfy:

(1) $f(x_i) \geq 0$ for all $x_i \in R_X$;

(2) $f(x) = 0$ for all $x \notin R_X$;

(3) $\sum_{i=1}^{\infty} f(x_i) = 1$, or $\sum_{x_i \in R_X} f(x_i) = 1$.

For any set $B \subset R$, we have

$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i).$$

DEFINITION 4 (PROBABILITY DENSITY FUNCTION)

The **probability density function** of a continuous random variable X , denoted by $f(x)$, is a function that satisfies:

(1) $f(x) \geq 0$ for all $x \in R_X$; and $f(x) = 0$ for $x \notin R_X$;

(2) $\int_{R_X} f(x) dx = 1$;

(3) For any a and b such that $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

note: $P(a < X < b) = P(a \leq X \leq b)$

To check that a function is a PDF, check conditions 1 and 2

Cumulative Distribution Function: $F(x) = P(X \leq x)$

and

$$P(a \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a-)$$

Deriving CDF from probability distribution

x	0	1	2
$f(x)$	1/4	1/2	1/4

We have

$$F(0) = f(0) = 1/4, F(1) = f(0) + f(1) = 3/4, F(2) = f(0) + f(1) + f(2) = 1.$$

Deriving probability distribution from cdf

$$f(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

$$\begin{aligned} \bullet f(1) &= P(X=0) = F(0) - F(0-) = \frac{1}{4} - 0 = \frac{1}{4} \\ \bullet f(2) &= P(X=1) = F(1) - F(1-) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2} \\ \bullet f(3) &= P(X=2) = F(2) - F(2-) = 1 - \frac{3}{4} = \frac{1}{4} \end{aligned}$$

CDF - Continuous Random Variable

- $F(X)$ assumes different values in R_x when $F(X) \neq 0$
- $F(X)$ is always non-decreasing (for discrete variable too)
 $x_1 < x_2 \rightarrow F(x_1) < F(x_2)$
- Probability function and CDF is one-to-one and uniquely determined
- $0 < F(x) < 1$
- (discrete), $0 < f(x) < 1$
- (continuous), $f(x) \geq 0$ but $f(x)$ is not necessarily ≤ 1 . While the PDF itself cannot exceed 1, there are points within its range that are greater than 1

CDF: CONTINUOUS RANDOM VARIABLE

If X is a **continuous random variable**,

$$F(x) = \int_{-\infty}^x f(t) dt,$$

and

$$f(x) = \frac{dF(x)}{dx}.$$

Further

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a).$$

Find CDF of a continuous random variable

$$f(x) = \begin{cases} \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, & \text{for } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \int_0^x \frac{t}{\theta^2} e^{-t^2/(2\theta^2)} dt = \left[-e^{-t^2/(2\theta^2)} \right]_{t=0}^x = 1 - e^{-x^2/(2\theta^2)}.$$

Discrete: Expectation (mean) and Variance

$$\bullet E(X) = \sum_{x \in R_X} x_i f(x_i) = \mu_x$$

Random: Expectation (mean) and Variance

$$\bullet E(X) = \mu_x = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x f(x) dx$$

• The mean of X is not necessarily a possible value

Properties of Expectation

$$1 E(aX + b) = aE(X) + b$$

$$2 E(X + Y) = E(X) + E(Y)$$

$$3 (\text{discrete}) E|g(x)| = \sum_{x \in R_X} g(x) f(x)$$

$$3 (\text{continuous}) E|g(x)| = \int_{x \in R_X} g(x) f(x) dx$$

• Using properties 1, 2 we have

$$E(a_1 X_1 + a_2 X_2 + \dots + a_k X + k) = a_1 E(X_1) + \dots + a_k E(X_k)$$

Variance

$$\bullet Var(X) = \sigma_x^2 = E(X - \mu_x)^2 = E(X^2) - E(X)^2$$

$$\bullet (\text{discrete}) Var(X) = \sum_{x \in R_X} (x - \mu_x)^2 f(x)$$

$$\bullet (\text{continuous}) Var(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

• $V(X) \leq 0$ for any X , $V(X) = 0$ when X is a constant

Normal Distribution

$$\bullet f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\bullet f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

• Shape is identical if two curves share the same σ

• Curve flattens as σ increases

• $\Phi(x)$ is the CDF of the normal distribution

$$\bullet \Phi(x) = P(x < 2), \Phi(-x) = 1 - \Phi(x)$$

$$\bullet P(|x| < 2) = 2\Phi(x) - 1 = p(-2 < x < 2)$$

$$\bullet P(x_1 < x < x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

$$\bullet z = \frac{x - \mu}{\sigma}$$

$$\bullet \phi(z) = f_Z(z)$$

$$\bullet \Phi(z) = \int_{-\infty}^z (\phi(t)) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z (e^{-\frac{t^2}{2}}) dt$$

Central Limit Theorem

• Under CLT, sample mean approaches population mean as n increases

• Law of large numbers: $P(|\bar{x} - \mu_x| > \epsilon) \rightarrow 0$ as n increases

• Symmetric no outliers: 15-20

• Skewed: 30-50, heavily skewed: > 1000

• $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \rightarrow$ large n , higher accuracy

Margin of Error

DIFFERENT CASES

	Population	σ	n	Statistic	E	n for desired E_0 and α
I	Normal	known	any	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$
II	any	known	large	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot \sigma}{E_0}\right)^2$
III	Normal	unknown	small	$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	$t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}}$	$\left(\frac{t_{n-1, \alpha/2} \cdot s}{E_0}\right)^2$
IV	any	unknown	large	$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	$z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$	$\left(\frac{z_{\alpha/2} \cdot s}{E_0}\right)^2$

Confidence Interval

CONFIDENCE INTERVALS FOR THE MEAN

The table below gives the $(1 - \alpha)$ confidence interval (formulas) for the population mean.

Case	Population	σ	n	Confidence Interval
I	Normal	known	any	$\bar{x} \pm z_{\alpha/2} \cdot \sigma / \sqrt{n}$
II	any	known	large	$\bar{x} \pm z_{\alpha/2} \cdot \sigma / \sqrt{n}$
III	Normal	unknown	small	$\bar{x} \pm t_{n-1, \alpha/2} \cdot s / \sqrt{n}$
IV	any	unknown	large	$\bar{x} \pm z_{\alpha/2} \cdot s / \sqrt{n}$

Note that n is considered large when $n \geq 30$.

Alternative pooled variance:

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

Equal Variance Assumption: $\frac{1}{2} < \frac{s_1}{s_2} < 2$

Assumptions

- Normal population (esp for small samples)
- Random and independent sampling

	Do not reject H_0	Reject H_0
H_0 is true	Correct Decision	type I error
H_0 is false	Type II error	Correct Decision

DEFINITION 1 (TYPE I vs TYPE II ERROR)

The rejection of H_0 when H_0 is true is called a **Type I error**.

Not rejecting H_0 when H_0 is false is called a **Type II error**.

DEFINITION 2 (SIGNIFICANCE LEVEL VS POWER)

The probability of making a **Type I** error is called the **level of significance**, denoted by α . That is,

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true}).$$

Let

$$\beta = P(\text{Type II error}) = P(\text{Do not reject } H_0 \mid H_0 \text{ is false}).$$

We define $1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ is false})$ to be the **power of the test**.