MTH102 Engineering Mathematics II

Revision

Term: 2024

Probability theory

- Random experiment: the outcome of an experiment is not predictable with certainty.
- Sample space S: the set of all possible outcomes of an experiment.
- Event E: any subset of the sample space, i.e. a set consisting of possible outcomes of the experiment.
- $E \cup F$: the **union** of E and F, i.e. either E or F occur.
- EF ($E \cap F$): the **intersection** of E and F, i.e. both E and F occur.
- E^c: the **complement** of E, i.e. E does not occur.

Axioms of probability

Consider an experiment whose sample space is S. For each event E of S, we assume that a number $P(E)$ is defined and satisfies the following three axioms:

1. $$0 \leq P(E) \leq 1.$$
2. $$P(S) = 1.$$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i E_j = \emptyset$ when $i \neq j$),
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

We refer to $P(E)$ as the probability of the event E.

Basic propositions of probability

- **The complementation rule**:

$$P(E^c) = 1 - P(E).$$

- **Law of total probability**: for any event A and B, it holds that

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

- **Addition rule**: for any event A and B, it holds that

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

Equally likely models

- The sample space S consists of finite outcomes:

$$S = \{x_1, x_2, \ldots, x_n\}.$$

- All the outcomes are equally likely to occur, i.e.

$$P(\{x_1\}) = P(\{x_2\}) = \cdots = P(\{x_n\}) = \frac{1}{n}.$$

- The probability of an event A is

$$P(A) = \frac{\text{number of outcomes in A}}{\text{number of outcomes in S}} = \frac{|A|}{|S|},$$

i.e. $P(A)$ equals the proportion of outcomes in S that are contained in A.

Permutations and combinations

- Given n objects, there are
$$n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1 = n!$$
different permutations of the n objects.
- Given a set of n objects of which certain are indistinguishable from each other. There are
$$\frac{n!}{n_1! n_2! \cdots n_r!}$$
different permutations of n objects, of which n_1 are alike, n_2 are alike, $\ldots$, n_r are alike.
- The number of different groups of r items that could be formed from a set of n items is
$$\binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

Conditional probability

Let A and B be events of a sample space S and $P(B) \neq 0$. The probability that an event A occurs given that an event B occurs is called the *conditional probability of A given B* and it is denoted by

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

Similarly if $P(A) \neq 0$, the conditional probability of B given A is defined as

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

The multiplication rule: if A and B are events of a sample space S with $P(A) \neq 0$ and $P(B) \neq 0$, then

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

Law of total probability

Let the events $B_1, B_2, \ldots, B_n$ be mutually exclusive and

$$S = B_1 \cup B_2 \cup \cdots \cup B_n.$$

We assume that $P(B_i) > 0$ for $i = 1, 2, \ldots, n$. Then for any event A, A can be represented as the union of mutually exclusive events, i.e.

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

Therefore,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i),$$

which is called **law of total probability**.

Bayes' rule

Let the events $B_1, B_2, \dots, B_n$ be mutually exclusive with $P(B_i) > 0$ and

$$S = B_1 \cup B_2 \cup \cdots \cup B_n.$$

For any event A, it follows by law of total probability that

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

Therefore for $i = 1, 2, \dots, n$

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)},$$

which is called *Bayes' Rule*.

Independent events

Definition

Two events A and B are said to be independent if

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

Two events A and B that are not independent are said to be dependent.

Remark. Pay attention, A and B are independent does not mean that $A \cap B = \emptyset$!

Property. If A and B are independent, then

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B).$$

The probability of A does not depend on the occurrence or nonoccurrence of B, and conversely.

Discrete random variable

- A random variable that can take on at most a countable number of possible values is said to be **discrete random variable**.
- For a discrete random variable X and any value x, the probability $P(X = x)$ is frequently denoted by $p(x)$, called the **probability mass function**. It is hereafter abbreviated **pmf**.

X	x_1	x_2	$\cdots$	x_i	$\cdots$
$P(X = x)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_i)$	$\cdots$

- We call the function defined by

$$F(x) = P(X \leq x), \ -\infty < x < \infty,$$

the **cumulative distribution function** and abbreviate it as **cdf**.

Mean and variance

- If X is a discrete random variable that takes on the values $x_1, x_2, \ldots$, with the pmf $p(x)$, then the *mean*, or the *expectation*, of X, denoted by $E(X)$, is defined by

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i).$$

- If X is a discrete random variable that takes on the values $x_1, x_2, \ldots$, with the pmf $p(x)$, then for any real-valued function $g : \mathbb{R} \to \mathbb{R}$,

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p(x_i).$$

- The *variance* of a random variable X, denoted by $Var(X)$, is defined as the mean of the function of X with $g(X) = (X - E[X])^2$, i.e.

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Discrete distributions

- **Binomial distribution**:

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, \ldots, n,$$

$$E(X) = np, \ Var(X) = np(1-p).$$

- **Geometric distribution**:

$$p(k) = P(X = k) = (1-p)^{k-1} p, \ k = 1, 2, \ldots,$$

$$E(X) = \frac{1}{p}, \ Var(X) = \frac{1-p}{p^2}.$$

- **Poisson distribution**:

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, \ldots,$$

Probability density function

Definition

Let X be a random variable over the outcome space S which is an interval or union of intervals. X is called a **continuous random variable** *if there exists an integrable function $f(x)$ satisfying the following:*

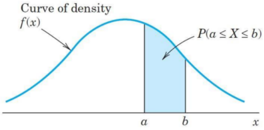
(a) $f(x) \geq 0, \ x \in \mathbb{R}$;

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$;

(c) *If $a \leq b$, then*

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

$f(x)$ *is called the* **probability density function (pdf)** *of X.*



Cumulative distribution function

The **cumulative distribution function (cdf)** of a continuous random variable X with the pdf $f(x)$ is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}.$$

Properties:

- $F(x)$ is continuous, and for x values for which the derivative $F'(x)$ exists,

$$F'(x) = f(x).$$

- For any $(a, b) \subseteq S$,

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

Mean and variance

- If X is a continuous random variable having a probability density function $f(x)$, then the **mean**, or the **expectation**, of X, denoted by $E(X)$, is defined by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

- If X is a continuous random variable having a probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

- The *variance* of a random variable X, denoted by $Var(X)$, is defined as the mean of the function of X with $g(X) = (X - E[X])^2$, i.e.

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - (E[X])^2.$$

Uniform random variable

A random variable X is said to be *uniformly* distributed over the interval (a, b) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

The cdf of X is

$$F(x) = \begin{cases} 0 & x \leq a, \\ \frac{x-a}{b-a} & a < x < b, \\ 1 & x \geq b. \end{cases}$$

The mean and variance of X are

$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}.$$

Exponential distribution

A continuous random variable X is said to be an *exponential* random variable (or is said to be *exponentially distributed*) if the pdf is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

The cdf of X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

The mean and variance of X are

$$E(X) = \frac{1}{\lambda}, \; Var(X) = \frac{1}{\lambda^2}.$$

Normal distribution

- The random variable X has a **normal distribution** if its pdf is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R},$$

where $\mu, \sigma \in \mathbb{R}$ are parameters. Briefly, we say that X is $N(\mu, \sigma^2)$.

- The mean and variance of X are

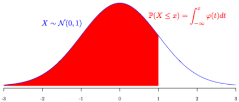
$$E(X) = \mu, \quad Var(X) = \sigma^2.$$

- If X is $N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma$ is $N(0, 1)$. Therefore

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

where Φ is the cdf of standard normal distribution $N(0, 1)$.

Table of Standard Normal distribution



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Joint probability mass function

Definition

Let X and Y be two random variables defined on a discrete space. Let S denote the corresponding two-dimensional space of X and Y, the two random variables of the discrete type. The probability that $X = x$ and $Y = y$ is denoted by

$$f(x, y) = P(X = x, Y = y).$$

The function $f(x, y)$ is called the **joint probability mass function (joint pmf)** *of X and Y and has the following probabilities:*

(a) $0 \le f(x, y) \le 1$.

(b) $\sum\sum_{(x,y)\in S} f(x, y) = 1$.

(c) *If A is a subset of S, then $P((X, Y) \in A) = \sum\sum_{(x,y)\in A} f(x, y)$.*

Marginal probability mass function

Definition

Let X and Y have the joint pmf $f(x, y)$ with space S. The pmf of X alone, which is called the **marginal probability mass function (marginal pmf)** *of X, is defined by*

$$f_X(x) = \sum_y f(x, y) = P(X = x), \ x \in S_X,$$

where the summation is taken over all possible y values for each given x in the x space S_X. That is , the summation is over all (x, y) in S with a given x value. Similarly, the **marginal pmf** *of Y is defined by*

$$f_Y(y) = \sum_x f(x, y) = P(Y = y), \ y \in S_Y,$$

where the summation is taken over all possible x values for each given y in the y space S_Y.

Joint probability density function

Definition

Let X and Y be two random variables of the continuous type. The **joint probability density function (joint pdf)** *of X and Y is an integrable function $f(x, y)$ with the following properties:*

(a) $f(x, y) \geq 0$, *where* $f(x, y) = 0$ *when* (x, y) *is not in the space* S *of* X *and* Y.

(b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$.

(c) $P[(X, Y) \in A] = \iint_A f(x, y) dxdy$, *where* $\{(X, Y) \in A\}$ *is an event defined in the plane.*

Property (c) implies that $P[(X, Y) \in A]$ is the volume of the solid over the region A in the xy-plane and bounded by the surface $z = f(x, y)$.

Marginal probability density function

The respective **marginal pdfs** of continuous-type random variables X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \ x \in S_X,$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \ y \in S_Y,$$

where S_X and S_Y are the respective spaces of X and Y. The mathematical expectation of $g(X,Y)$ is given by

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy,$$

which applies to define the means and variances of X and Y.

Independence of random variables

Let X and Y have the joint pdf $f(x, y)$ with space S. $f_X(x)$ and $f_Y(y)$ are the marginal pdf of X and Y respectively. The random variables X and Y are **independent** if and only if, for every $x \in S_X$ and every $y \in S_Y$,

$$f(x, y) = f_X(x)f_Y(y);$$

otherwise, X and Y are said to be **dependent**.

Covariance and correlation coefficient

Definition

Let X and Y be random variables.

(a) *If $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, then*

$$E[g(X,Y)] = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = Cov(X,Y)$$

is called the **covariance** *of X and Y.*

(b) *If the standard deviations σ_X and σ_Y are positive, then*

$$\rho = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

is called the **correlation coefficient** *of X and Y.*

Properties of sample mean, sample variance and sample covariance

Let X and Y be two random variables. The following holds.

- $$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

- $$Var(aX + b) = a^2 Var(X).$$

- $$Cov(X, Y) = E(XY) - E(X)E(Y).$$

- $$Cov(aX, bY) = abCov(X, Y).$$

- $$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Properties of independence

Let X and Y be two independent random variables. The following holds.

- For any function $u(X)$ and $v(Y)$,

$$E[u(X)v(Y)] = E[u(X)]E[v(Y)].$$

In particular,

$$E(XY) = E(X)E(Y).$$

- $$Cov(X, Y) = 0, \text{ and } \rho = 0.$$

- $$Var(X + Y) = Var(X) + Var(Y).$$

Chebyshev's inequality

Proposition

If X is a random variable with mean μ and variance σ^2, then for any value $k > 0$,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Or equivalently,

$$P(|X - \mu| < k) \geq 1 - \frac{\sigma^2}{k^2}.$$

Law of large numbers

Theorem

Let $X_1, X_2, \dots$ *be a sequence of independent and identically distributed (i.i.d.) random variables, each having mean* μ *and variance* σ^2*. Then for any* $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq \varepsilon \right) = 0.$$

Central limit theorem

Theorem

Let $X_1, X_2, \ldots$ *be a sequence of independent and identically distributed (i.i.d.) random variables, each having mean* μ *and variance* σ^2*. Then the distribution of*

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \left(= \frac{X_1 + \cdots + X_n - n\mu}{\sigma \sqrt{n}}\right)$$

tends to the standard normal distribution $N(0, 1)$ *as* $n \to \infty$*. That is, for any* $a \in \mathbb{R}$,

$$\lim_{n \to \infty} P\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq a\right) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx.$$

Data analysis

Consider a sample of 16 data: $1, 3, 4, 5, 5, 6, 8, 10, 15, 15, 16, 18, 18, 20, 20, 26$.

- Mean: 11.875.
- Median: $\frac{x_8+x_9}{2} = \frac{10+15}{2} = 12.5$.
- Q_1 is the median of $x_1, \ldots, x_8$: $\frac{5+5}{2} = 5$.
- Q_3 is the median of $x_9, \ldots, x_{16}$: $\frac{18+18}{2} = 18$.
- Range: $26 - 1 = 25$.
- IQR: $Q_3 - Q_1 = 18 - 5 = 13$.
- $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 56.65$.