Probability Note

1 Fundamental Mathematics

1.1 Review of Calculus

Definition. (Taylor and Maclaurin series) If $f(x)$ has derivatives of all orders at $x = c$, which means $f^{(k)}(c)$ exists for $k = 0,\ 1,\ 2, \ldots$, then the series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots$$

is called the **Taylor series**(泰勒級數) of f about c.
If $c = 0$, the term **Maclaurin series**(馬克勞林級數) is usually used in place of Taylor series as

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k.$$

Theorem. Suppose the series

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n = a_0 + a_1(x-c) + a_2(x-c)^2 + a_3(x-c)^3 + \cdots$$

converges to $f(x)$ for $c - R < x < c + R$, where $R > 0$, then

$$a_k = \frac{f^{(k)}(c)}{k!}, \text{ for } k = 0,\ 1,\ 2,\ \ldots.$$

Example. What is the Maclaurin series for below functions?

(a) e^x
Let $f(x) = e^x$, $f^{(k)}(x) = e^x$ and $f^{(k)}(0) = e^0 = 1$ for $k = 0, 1, 2, \ldots$. Thus

$$\begin{aligned} e^x &= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} x^k \\ &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots. \end{aligned}$$

(b) $\sin x$
Let $f(x) = \sin x$, then

$$\begin{aligned} f'(x) &= \cos x, & f'(0) &= 1, \\ f''(x) &= -\sin x, & f''(0) &= 0, \\ f^{(3)}(x) &= -\cos x, & f^{(3)}(0) &= 1, \\ f^{(4)}(x) &= \sin x, & f^{(4)}(0) &= 0. \end{aligned}$$

Thus

$$\begin{aligned}
\sin x &= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \\
&= \frac{0}{0!}x^0 + \frac{1}{1!}x - \frac{0}{2!}x^2 - \frac{1}{3!}x^3 + \frac{0}{4!}x^4 + \cdots \\
&= 0 + \frac{1}{1!}x - 0 - \frac{1}{3!}x^3 + 0 + \cdots \\
&= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{(2n+1)}.
\end{aligned}$$

(c) $\dfrac{1}{1-x}$

Let $f(x) = \sin x$, since $\dfrac{d}{dx}\left(\dfrac{1}{1-x}\right) = \dfrac{1}{(1-x)^2}$ and $\dfrac{d^2}{dx^2}\left(\dfrac{1}{1-x}\right) = \dfrac{2}{(1-x)^3}$, we obtain

$$\frac{d^k}{dx^k} f(x) = \frac{k!}{(1-x)^k}.$$

Thus we have

$$f^{(k)}(0) = \left.\frac{d^k f}{dx^k}\right|_{x=0} = k!$$

Hence

$$\begin{aligned}
\frac{1}{1-x} &= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \\
&= \sum_{k=0}^{\infty} \frac{k!}{k!} x^k \\
&= \sum_{k=0}^{\infty} x^k \\
&= 1 + x + x^2 + \cdots .
\end{aligned}$$

1.2 Probability

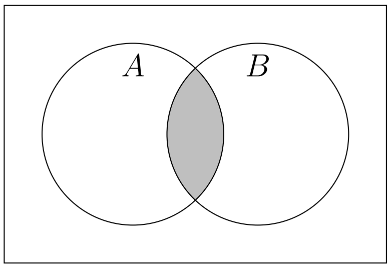
1.2.1 Introduction

- Random Experiment(隨機試驗)
  Data: different outcome every time

- Sample Space(S): The set of all possible outcomes.

$$\text{discrete: } \sum_{k=1}^{n} f(k)$$
$$\text{continuous: } \int_{-\infty}^{\infty} f(x)dx$$

- Events(事件): A subset of the sample space of a random experiment.
  Venn Diagram(文氏圖):

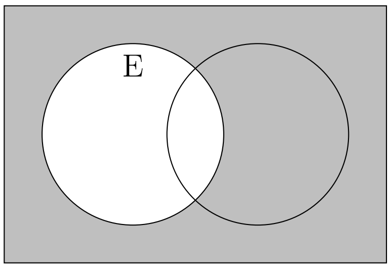


$A \cap B$

- Mutually Exclusive(互斥) Events: Two subset E_1 and E_2, the union are empty. (i.e. $E_1 \cap E_2 = \emptyset$)

1.2.2 Law of Events

1. Complement: E' or $\bar{E}$ and $(E')' = E$



2. Distribute Laws:

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

3. De-Morgan's Laws:

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

4. Commutative:

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

1.2.3 Principle of Fundamental Counting

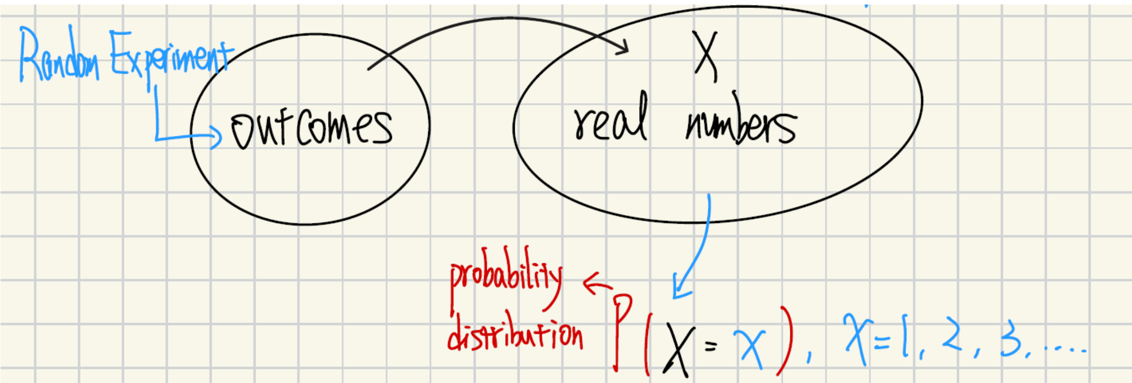
- Rule of Product: $n_1 \times n_2 \times \cdots \times n_k$
- Permutation: $n!$
- Combinations: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$
- Permutation of subset:
  1. A set of n different elements.
  2. Subset of r elements selected from the set: $P_r^n = n \cdot (n-1) \cdot \cdots = \frac{n!}{(n-r)!}$.
- Permutation of similar objects:
  1. There are $n_1 + n_2 + \cdots + n_k = n$ objects, type k has n_k objects.
  2. $N = \frac{n!}{n_1! n_2! \cdots n_k!}$

1.2.4 Probability

2 Discrete Random Variables and Probability Distributions

2.1 Random Variable (RV)

Random Variables(X): A function (mapping) that assign a real number to each outcome in the sample space of a random experiment.



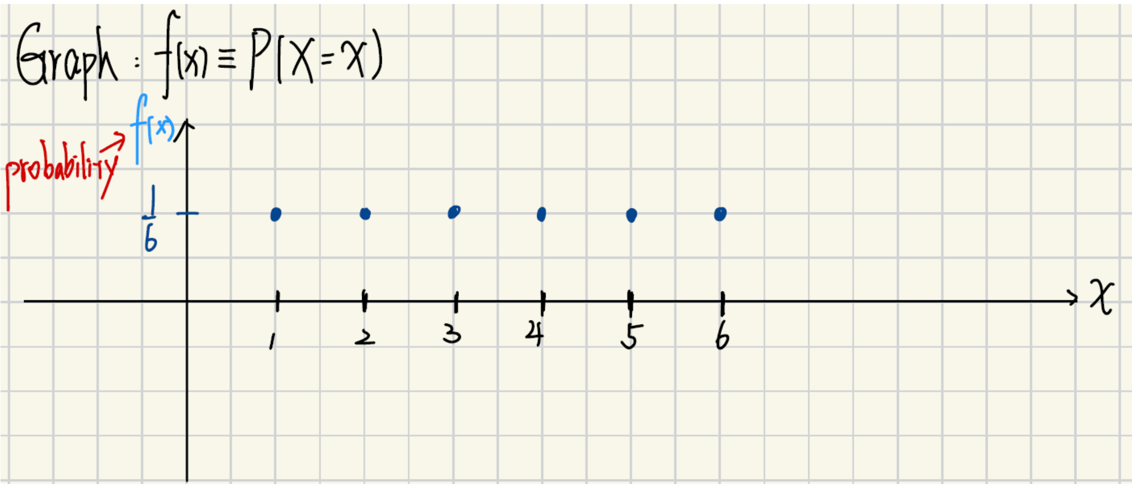
2.1.1 Probability Distribution

1. Define on RVs X.
2. A description of the probability associated with all possible
3. Notation: $f(x) \equiv P(X = x)$

Example.

$$P(X = x) = \begin{cases} \frac{1}{6} & ,1 \leq x \leq 6,\ x \in \mathbb{N} \\ 0 & ,\text{otherwise} \end{cases}$$

The probability distribution is $f(1) = \frac{1}{6}$, $f(2) = \frac{1}{6}$, ... The graph of $f(x)$ is



2.2 Probability and Cumulative Distribution Function

2.2.1 Probability Distribution Function

Definition. For discrete case, probability distribution function(pdf) is called **probability mass function**(pmf).
The pdf(or pmf), $f(x) = P(X = x)$ satisfies $f(x) \geq 0$ for all x and $\sum_{x \in X} f(x) = 1$.

Example. If $f(x) = \begin{cases} \left(\frac{1}{2}\right)^x & , x = 1,\ 2,\ 3, \cdots \\ 0 & ,\ \text{otherwise} \end{cases}$, then show that $f(x)$ is qualified as a probability mass function.

Proof. Since $f(x) = \left(\frac{1}{2}\right)^x \geq 0$, for $x \in \mathbb{N}$ and

$$\sum_{x=1}^{\infty} f(x) = \frac{1/2}{1 - 1/2} = 1.$$

So $f(x)$ is qualified as a probability mass function. □

2.2.2 Cumulative Distribution Function

Definition. Cumulative distribution function (or cdf), is defined as

$$F(x) \equiv P(X \leq x) = \sum_{k=1}^{x} f(k)$$

Remark. If $x \leq y$, then $F(x) \leq F(y)$.

Example. Let $f(x) = \left(\frac{1}{2}\right)^x$. Find $F(y_0)$, where y_0 is a real number.

Sol.

$$F(y_0) = \sum_{k=1}^{[y_0]} f(k) = 1 - \left(\frac{1}{2}\right)^{[y_0]}$$

2.3 Mean and Variance

2.3.1 Mean

Definition. The mean(期望值), or expectation value, is defined on random variable X as

$$E(X) = \sum_{x_k \in X} x_k f(x_k).$$

Sometime it denotes μ or μ_x.

2.3.2 Variance

Definition. The variance is baseline on the mean and define

$$\sigma^2 = Var(X) = E\left[(x - E(x))^2\right] = E(X^2) - \mu^2.$$

The standard deviation(標準差) is $\sqrt{Var(X)} = \sqrt{\sigma^2} = \sigma$.

Remark. If we extension mean as the transformation H, then

$$E[H(x)] = \sum_{x \in X} H(x) f(x)$$

Property. Here is the rule for $E(X)$:

(1) $E(c) = c$, $c \in \mathbb{R}$.

(2) $E(cX) = cE(X)$.

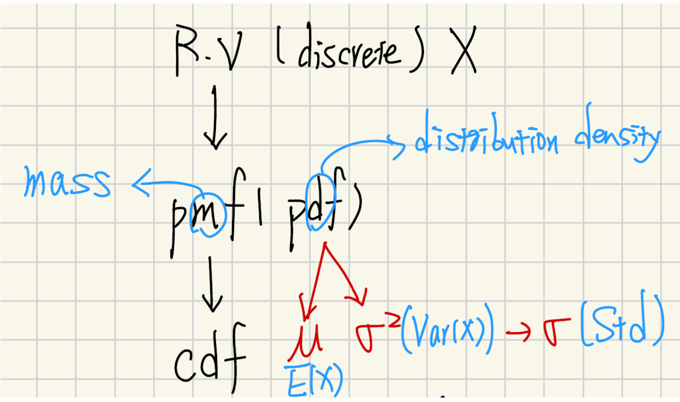
(3) $E(X + Y) = E(X) + E(Y)$.

Property. Here is the rule for $Var(X)$:

(1) $Var(c) = 0$, $c \in \mathbb{R}$.

(2) $Var(cX) = c^2 Var(X)$.

(3) If X and Y are independent, then $Var(X + Y) = Var(X) + Var(Y)$.



2.4 Discrete Uniform Distribution

1. RV X: the possible values $x_1,\ x_2,\ \cdots,\ x_n$

2. pmf: $f(x) = f(x_k) = \dfrac{1}{n},\ k = 1,\ 2,\ \ldots,\ n.$

Theorem. Let X be discrete uniform distributed random variable ranged from $a,\ a+1,\ a+2,\ \cdots,\ b$ for $a \le b$. Then the expected value is $E(X) = \dfrac{a+b}{2}$ and the variance is $Var(x) = \dfrac{(b-a+1)^2-1}{12}$.

Proof. $n = b - a + 1$. Since it is uniform, the pmf if $f(x) = \dfrac{1}{n} = \dfrac{1}{b-a+1}$. Thus

$$\begin{aligned} E(X) &= \sum_{k=a}^{b} x_k \frac{1}{b-a+1} \\ &= \frac{(a+b)(b-a+1)}{2} \cdot \frac{1}{b-a+1} = \frac{a+b}{2} \end{aligned}$$

and

$$\begin{aligned} Var(X) &= \sum_{k=a}^{b} \left(x_k - \frac{a+b}{2} \right)^2 f(x_k) \\ &= \sum_{k=a}^{b} \left(x_k^2 - (a+b)x_k + \frac{(a+b)^2}{4} \right) f(x_k) \\ &\vdots \\ &= \frac{(a+b-1)^2-1}{12}. \end{aligned}$$

□

2.5 Binomial Distribution(二項式分布)

1. Experiment: Do n trials
   Ask: the trial is "**success**" the number among n trials. (Bernoulli trial)

2. Define RV X: X is the number of successes in n **trials**.

3. Define pmf: We define

   (1) Probability of success: p (called parameter of pmf)

   (2) Probability of failure: $1 - p$

   and

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0,\ 1,\ 2,\ \cdots,\ n.$$

Remark. $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1.$

Remark. The Binomial Theorem is $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

2.5.1 Mean and Variance

Theorem. The mean and variance of Binomial Distribution are

1. $E(X) = np$
2. $Var(x) = np(1-p)$

2.5.2 Moment Generating Function (M.G.F)

Definition. Let RV be X, then the k**th moment** of RV X is defined as $E(x^k)$.

Remark. The first moment and second moment:

(1) If $k = 1$, then the first moment $E(x^k) = E(x)$ is mean.

(2) If $k = 2$, $E(x^k) = E(x^2)$ and we can know that $Var(X) = E(x^2) - (E(x))^2$.

Definition. Let RV be X. The moment generating function (M.G.F) of X is

$$m_x(t) = E(e^{tx})$$

Theorem. We can use M.G.F to generate $E(x^k)$, that is

$$\left.\frac{d^k m_x(t)}{dt^k}\right|_{t=0} = E(x^k).$$

Proof. Let $m_x(t) = E(e^{tx})$, using the Maclaurin series, we obtain

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots$$

Thus

$$\begin{aligned} \left.\frac{d^k m_x(t)}{dt^k}\right|_{t=0} &= \left.\frac{d^k}{dt^k}\left(E(e^{tx})\right)\right|_{t=0} \\ &= E\left(\left.\frac{d^k}{dt^k} e^{tx}\right|_{t=0}\right) \\ &= E\left(\left.\frac{d^k}{dt^k}\left(1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots\right)\right|_{t=0}\right) \\ &= E\left(\left.\left(\frac{k!}{k!}x^k + \frac{(k+1)!/1}{(k+1)!} tx^{k+1} + \cdots\right)\right|_{t=0}\right) \\ &= E\left(x^k\right). \end{aligned}$$

□

Remark. The variance, σ^2 or $Var(X)$, is

$$\sigma^2 = \sum_{x \in X} (x-\mu)^2 f(x) = E(x^2) - (E(x))^2$$

We will prove the variance is equal to second moment minus the square of first moment.

Proof.

$$\begin{aligned}
\sigma^2 &= \sum_{x \in X} (x-\mu)^2 f(x) \\
&= \sum_{x \in X} (x^2 - 2\mu x + \mu^2) f(x) \\
&= \sum_{x \in X} x^2 f(x) - 2\mu \sum_{x \in X} x f(x) + \mu^2 \sum_{x \in X} f(x) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2(E(X))^2 + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}$$

□

Example. Let RV X be the number of radar signals properly identified in a 30-minute time period in which 10 signals are received. Find the probability that at most seven signals will be identified correctly.

Sol. Let RV be the number of radar signals properly identified in a 30-minute time period in which 10 signals.

$$f(x) = \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \binom{10}{x} \left(\frac{1}{2}\right)^{10}$$

Then at most seven signals is

$$P(X \leq 7) = \sum_{x=0}^{7} \binom{10}{x} \left(\frac{1}{2}\right)^{10}.$$

2.6 Geometric and Negative Binomial Distribution

2.6.1 Geometric Distribution(幾何分布)

1. Experiment: A series of trials.
   outcome: "success"(s) or "failure"(f). (Bernoulli trial)

2. Assume the trials are **identical and independent**.

3. Define RV: X is the number of "trials" needed to obtain the first success.
   e.g. $X = x$, $x = 1,\ 2,\ 3,\ \cdots$, then $x = 1 \iff \{s\}$, $x = 2 \iff \{f,\ s\}$, $x = 3 \iff \{f,\ f,\ s\}, \ldots$.
   Let

   (1) $P(s) = p$

   (2) $P(f) = 1 - p$

4. The pmf is defined as

$$f(x) = (1-p)^{x-1}p,\ x = 1,\ 2,\ 3,\ 4,\ \cdots.$$

Definition. The cdf of geometric distribution is

$$F(x) = \sum_{k=1}^{x} f(k) = \sum_{k=1}^{x}(1-p)^{k-1}p = 1 - (1-p)^{[x]}$$

Theorem. The mean and variance of geometric distribution are

(1) $E(X) = \mu = \dfrac{1}{p}$

(2) $\sigma^2 = Var(X) = \dfrac{1-p}{p^2}$

Example. The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analyzed before a large particle is detected?

Sol. The probability is

$$P(X = 125) = (0.99)^{124} \times (0.01).$$

Example. The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume that the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the first error. Find the mean and standard deviation.

Sol. Let $p = 0.1$. Then the mean is

$$E(X) = \frac{1}{p} = \frac{1}{0.1} = 10,$$

and the standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{0.9}{0.01}} = \sqrt{90}.$$

Example. In last Example, the probability that a bit is transmitted in error is 0.1. Suppose 50 bits have been transmitted. What is mean number of bits transmitted until the next error

Sol. Let $p = 0.1$. Then

$$\mu = \frac{1}{p} = 10.$$

2.6.2 Negative Binomial Distribution(負二項分布)

1. Experiment: Bernoulli trial.
2. Outcome: The trials are observed until exactly r successes.
3. Define RV X: The number of trials needed to obtain the r successes.
4. pmf $f(x)$:

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)} \cdot p,$$

for $x = r,\ r+1,\ r+2,\ \cdots$ and $r = 1,\ 2,\ 3, \cdots$.

Remark. For $f(x) = \binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)} \cdot p$,

(1) $P(x = r) = p \cdot p \cdot p \cdot \cdots \cdot p = p^r$.

(2) $P(x = r+1) = P(x = r) \cdot p \cdot \binom{r}{r-1}$.

Theorem. The mean and variance of negative binomial distribution are

(1) $\mu = E(X) = \dfrac{r}{p}$.

(2) $\sigma^2 = Var(X) = \dfrac{r(1-p)}{p^2}$.

2.7 Hypergeometric Distribution

- Experiment
  (1) A random sample of size n without replacement(取後不放回) and without regard to order from a collection of N objects. That is $\binom{N}{n}$.
  (2) of the N objects, k have a trait(特徵) of interest to us; the other $N - K$ do not have the trait.

- Define RV: The number of objects in the sample(n) with the trait.
  (1) the upper bound(上界) of x: $x \leq k$ and $x \leq n \implies x \leq \min(k,\ n)$
  (2) the lower bound(下界) of x: $x \geq 0$ and $x \geq n - (N - k) \implies x \geq \max(0,\ n - (N - k))$

  So the range of x is $\max(0,\ n - (N - k)) \leq x \leq \min(k,\ n)$.

- pmf:
  (1) All: $\binom{N}{n}$
  (2) x have trait: $\binom{k}{x}$
  (3) Others do not have trait: $\binom{N-k}{n-x}$

  The pmf is defined as

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

Example. A batch of parts contains 100 from a local supplier of circuit boards and 200 from a supplier in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Sol.

$$p = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

Remark. If $\frac{n}{N} \leq 0.05$, then hypergeometric distribution can binomial distribution to **approximate**. Let $p = \frac{K}{N}$ to $\binom{n}{x}p^x(1-p)^{n-x}$.

2.8 Poisson Distribution(卜瓦松分布)

Definition. (Poisson process) Involve observing discrete events in a continuous interval of time, length, or space.

- RV X: Poisson Random Variable (RV)
  In a Poisson Process, the variable of interest X is the number of occurrences of the event in an interval of length s units.
- pmf is defined as
  $$f(x) = P(X = x) = \frac{e^{-k} \cdot k^x}{x!}$$
  for $x = 0,\ 1,\ 2,\ 3,\ \cdots$ and $k = \lambda \cdot s$, λ is average number per unit(每單位平均個數)

Theorem. The mean and variance of geometric distribution are

(1) $E(X) = \mu = k$

(2) $\sigma^2 = Var(X) = k$

Example. For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter. Find the probability of exactly 2 flaws in 1 millimeter of wire.

Sol. Let $f(x) = \dfrac{e^{-k} \cdot k^x}{x!}$, $\lambda = 2.3$, $s = 1$. Then $k = \lambda \cdot s = 2.3$. Thus

$$f(2) = \frac{e^{-2.3} \cdot 2.3^2}{2!}.$$

Example. Continue last example.

(a) Determine the probability of 10 flaws in 5 millimeters of wire.

(b) Determine the probability of at least one flaw in 2 millimeters of wire.

Sol. Let $f(x) = \dfrac{e^{-k} \cdot k^x}{x!}$, $\lambda = 2.3$.

(a) Let $s = 5$, $k = 2.3 \times 5 = 11.5$. Then

$$P(X = 10) = \frac{e^{-11.5} \cdot (11.5)^{10}}{10!}.$$

(b) Let $s = 2$, $k = 2.3 \times 2 = 4.6$. Then

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{e^{-4.6} \cdot (4.6)^0}{0!} = 1 - e^{-4.6}. \end{aligned}$$

3 Continuous Random Variables and Probability Distributions

3.1 Probability Density Functions

- **Random Variables(X):** Any value in some interval of real numbers.
- **Probability Density Functions(pdf):** $f(x)$ is the pdf with satisfying

  (1) $f(x) \geq 0, \forall x \in \mathbb{R}$.

  (2) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$.

  (3) $\displaystyle P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x)dx$

3.2 Cumulative Distribution Function

Definition. The continuous cumulative distribution function is defined as

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

Remark. Since

$$\begin{aligned} F(b) &= P(X \leq b) = \int_{-\infty}^{b} f(t)dt \\ F(a) &= P(X \leq a) = \int_{-\infty}^{a} f(t)dt, \end{aligned}$$

we obtain

$$F(b) - F(a) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt$$

By the Fundamental Theorem of Calculus, we have

$$F(b) - F(a) = \int_a^b f(t)dt = P(a \leq X \leq b).$$

Hence

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

3.3 Mean and Variance

Definition. Suppose that X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

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

The **variance** of X, denoted as $Var(X)$ or σ^2, is

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$. In particular, if X is a continuous random variable with probability density function $f(x)$, then

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

3.3.1 Moment Generating Function for Continuous RV

Definition. The moment generating function for continuous is

$$m_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

Then,

(1) The first moment(1st moment) is

$$E(X) = \left. \frac{dm_x(t)}{dt} \right|_{t=0}$$

(2) The second moment(2nd moment) is

$$E(X^2) = \left. \frac{d^2 m_x(t)}{dt^2} \right|_{t=0}$$

Example. Let RV X be continuous and pdf is $f(x) = e^{-x}$, $x > 0$. Find the moment generating function of X.

Sol.

$$\begin{aligned} m_x(t) = E(e^{tx}) &= \int_0^\infty e^{tx} e^{-x} dx \\ &= \int_0^\infty e^{(t-1)x} dx \\ &= \lim_{b\to\infty} \int_0^b e^{(t-1)x} dx \\ &= \lim_{b\to\infty} \left[\frac{1}{t-1} e^{(t-1)x} \right]_0^b \\ &= \lim_{b\to\infty} \left[\frac{1}{t-1} (e^{(t-1)b} - e^{(t-1)\cdot 0}) \right] \\ &= \lim_{b\to\infty} \frac{1}{t-1} (e^{(t-1)b} - 1) \end{aligned}$$

If $t-1 > 0$, it will be diverged, so the M.G.F converges if $t-1 < 0$. Hence when $t < 1$, $m_x(t) = \dfrac{1}{1-t}$.

Example. (Electric Current) Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the probability density function of X is $f(x) = 5$ for $4.9 \le x \le 5.1$. What is the probability that a current measurement is less than 5 milliamperes?

Sol.

$$P(x < 5) = \int_{4.9}^5 5dx = 0.5.$$

Example. For the copper current measurement in last Example, find the cumulative distribution function of the random variable X consists of three expressions.

Sol. If $x < 4.9$, $F(x) = f(x) = 0$.
If $4.9 \le x \le 5.1$, then

$$F(x) = \int_{4.9}^x f(t)dt = \int_{4.9}^x 5dt = 5x - 24.5.$$

If $x \ge 5.1$, then

$$F(x) = \int_{4.9}^x f(t)dt = 1.$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 4.9 \\ 5x - 24.5 & 4.9 \le x \le 5.1 \\ 1 & x \ge 5.1 \end{cases}$$

Example. (Reaction Time) The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-0.01x} & x \geq 0 \end{cases}$$

Determine the probability density function of X.

Sol. For $x \geq 0$, since

$$f(x) = F'(x) = 0.01e^{-0.01x}$$

Hence

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & x \geq 0 \end{cases}$$

Example. For the copper wire current measurement, the pdf is $f(x) = 0.05$ for $0 \leq x \leq 20$. Find the mean and variance.

Sol. Given $f(x) = 0.05$ for $0 \leq x \leq 20$. Then

$$\begin{aligned} \mu &= \int_0^{20} x \cdot 0.05dx \\ &= \left.\frac{0.05x^2}{2}\right|_0^{20} \\ &= 0.05 \times 200 = 10 \end{aligned}$$

and

$$\begin{aligned} \sigma^2 &= \int_0^{20} (x-10)^2 \cdot 0.05dx \\ &= \frac{5}{100} \int_{-10}^{10} u^2 du && \text{Let } u = x - 10 \\ &= \frac{5}{100} \left[\frac{u^3}{3}\right]_{-10}^{10} \\ &= \frac{5}{100} \cdot \frac{2000}{3} \\ &= \frac{100}{3}. \end{aligned}$$

3.4 Continuous Uniform Distribution

- RV X: $a \leq x \leq b$.
- pdf: $f(x) = \dfrac{1}{b-a}$
- cdf:

$$F(x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt$$

Theorem. The mean and variance of continuous uniform distribution are

$$\mu = E(X) = \int_a^b x \cdot \frac{1}{b-a}dx = \frac{a+b}{2}$$

and

$$\sigma^2 = Var(X) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a}dx = \frac{(b-a)^2}{12}.$$

3.5 Exponential Distribution(指數分布)

3.5.1 Gamma Function

Definition. The gamma function $\Gamma : (0, \infty) \to \mathbb{R}$ is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx,$$

for $\alpha > 0$.

Property. There are some properties for gamma function.

(1) $\Gamma(1) = \displaystyle\int_0^\infty x^0 e^{-x}dx = 1$.

(2) For $\alpha > 1$,

$$\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1).$$

(3) For $\alpha > 1$ and α is an integer,

$$\Gamma(\alpha) = (\alpha - 1)!.$$

3.5.2 Exponential Distribution(指數分布)

- Define RV X: The time of the occurrence(出發) of the first event.
  (1) Time: continuous RV
  (2) Event number: discrete RV
- RVs:
  (1) X: Exponential, $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.
  (2) N: Poisson, $f(n) = \dfrac{e^{-k}k^n}{n!}$, $n = 1,\ 2,\ \cdots,\ k = \lambda s$.
- pdf

Proof. Let N be the number of occurrence of event in time interval $[0,\, x]$.

$$P(N = 0) = P(X > x) = 1 - p(X \leq x)$$

Then $P(N = 0) = 1 - F(x)$ Thus

$$\begin{aligned} 1 - F(x) = P(N = 0) = f(0) = \frac{e^{-k}k^0}{0!} = e^{-k} = e^{-\lambda x} \\ \Longrightarrow 1 - F(x) = e^{-\lambda x} \\ \Longrightarrow F(x) = 1 - e^{-\lambda x} \\ \Longrightarrow f(x) = F'(x) = \lambda e^{-\lambda x} \end{aligned}$$

□

So the pdf is $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$.

Theorem. The mean and variance of exponential distribution are

$$\mu = E(X) = \int_a^b x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

and

$$\sigma^2 = Var(X) = \int_a^b \left(x - \frac{1}{\lambda}\right)^2 \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

Example. In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in an interval of 6 minutes(0.1 hours)?

Sol. Let X be exponential RV. And $\lambda = 25$, thus we have

$$f(x) = \lambda e^{-\lambda x} = 25e^{-25x}, \ x \geq 0$$

Thus

$$\begin{aligned} P(X > 0.1) &= \int_{0.1}^{\infty} f(x)dx \\ &= \int_{0.1}^{\infty} 25e^{-25x}dx \\ &= e^{-25\times 0.1} = e^{-2.5}. \end{aligned}$$

Example. (Cont. last example) What is the probability that the time until the next log-on is between 2 and 3 minutes?

Sol.

$$\begin{aligned} P\left(\frac{2}{60} \leq x \leq \frac{3}{60}\right) &= \int_{1/30}^{1/20} 25e^{-25x}dx \\ &= e^{-25x}\Big|_{1/30}^{1/20} \\ &= e^{-5/4} - e^{-5/6}. \end{aligned}$$

Example. Continue last example, determine the interval of time such that the probability that no log-on occurs in the interval is 0.90.

Sol.

$$\begin{aligned} P(X > x) = e^{-25x} &= 0.90 \\ -25x &= \ln(0.90) \\ x &= 0.25 \text{ minute}. \end{aligned}$$

Example. Continue last example, what is the mean and standard deviation of the time until the next log-in?

Sol.

$$\mu = \frac{1}{\lambda} = \frac{1}{25}$$

and

$$\sigma = \sqrt{\frac{1}{k^2}} = \frac{1}{k} = \frac{1}{25}.$$

3.6 Erlang and Gamma Distribution

3.6.1 Erlang Distribution(愛爾朗分布)

- RV X: The time of occurrence of the rth moment.
- the pmf is defined as

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$$

  for $x > 0$ and $r = 1,\ 2,\ 3,\ \cdots$.

3.6.2 Gamma Distribution

Definition. The random variable X with pdf:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)},\ \text{for } x > 0$$

is a gamma random variable with parameters $\lambda > 0$ and $r > 0$. If r is an integer, then X has an Erlang distribution.

Theorem. The mean and variance of gamma distribution are

$$\mu = E(X) = \int_a^b x \cdot \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} dx = \frac{k}{\lambda}$$

and

$$\sigma^2 = Var(X) = \int_a^b \left(x - \frac{k}{\lambda}\right)^2 \cdot \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} dx = \frac{k}{\lambda^2}.$$

3.7 Weibull Distribution

Definition. The random variable X with pdf

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^{\beta}}, \text{ for } x > 0$$

is a Weibull random variable with scale parameter $\delta > 0$ and shape parameter $\beta > 0$. The cdf is

$$F(x) = 1 - e^{-(x/\delta)^{\beta}}$$

Theorem. The mean and variance of Weibull distribution is given by

$$\mu = E(X) = \delta \cdot \Gamma\left(1 + \frac{1}{\beta}\right)$$

and

$$\sigma^2 = Var(X) = \delta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right)\right] - \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]$$

Example. (Bearing Wear) The time to failure (in hours) of a bearing in a mechanical shaft(軸) is satisfactorily modeled as a Weibull random variable with $\beta = 2$ and $\delta = 5000$ hours.

(a) Determine the mean time until failure.

(b) Determine the probability that a bearing lasts at least 6000 hours.

Sol. (a)

$$\begin{aligned} E(X) &= 5000 \times \Gamma\left(1 + \frac{1}{2}\right) \\ &= 5000\Gamma(1.5) \\ &= 5000 \times 0.5\sqrt{\pi} \\ &= 2500\sqrt{\pi}. \\ &= 4431.1 \qquad \text{using Excel } \texttt{=5000*EXP(GAMMALN(1.5))} \end{aligned}$$

(b)

$$\begin{aligned} P(X > 6000) &= 1 - F(6000) \\ &= e^{-\left(\frac{6000}{5000}\right)^2} \\ &= e^{-1.2^2} = e^{-1.44} \\ &= 0.237 \qquad \text{using Excel } \texttt{=1-WEIBULL(6000, 2, 5000, TRUE)} \end{aligned}$$

3.8 Lognormal Distribution(對數常態分布)

Definition. Let W denote a normal random variable with mean θ and variance ω^2, then $X = e^W$ is a **lognormal random variabl**e with probability density function

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} e^{-\left[\frac{(\ln(x)-\theta)^2}{2\omega^2}\right]}$$

for $0 < x < \infty$.
The mean and variance of X are

$$E(X) = e^{\theta+\omega^2/2}$$

and

$$Var(X) = e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right)$$

Remark. The cumulative distribution function for X is[1]

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(e^W \leq x) \\ &= P(W \leq \ln(x)) \\ &= P\left[Z \leq \frac{\ln(x)-\theta}{\omega}\right] \\ &= \Phi\left[\frac{\ln(x)-\theta}{\omega}\right]. \end{aligned}$$

Example. (Semiconductor Laser) The lifetime (in hours) of a semiconductor laser has a lognormal distribution with $\theta = 10$ and $\omega = 1.5$. What is the probability that the lifetime exceeds 10,000 hours?

Sol.

$$\begin{aligned} P(X > 10000) &= 1 - P[e^W \leq 10000] \\ &= 1 - P[W \leq \ln(10000)] \\ &= 1 - \Phi\left[\frac{\ln(10000)-10}{1.5}\right] \\ &= 1 - \Phi(-0.5264) \\ &= 1 - 0.30 = 0.70 \end{aligned}$$

or using Excel `=1-NORMDIST(LN(10000), 10, 1.5, TRUE)`

[1] I suggest you read 3.10 at first for Φ.

Example. Continue last example, what lifetime is exceeded by 99% of lasers?

Sol.

$$P(X > x) = 1 - \Phi\left[\frac{\ln(x) - 10}{1.5}\right] = 0.99.$$

Using the table or Excel, we have:

Variable	Value	Formula in Excel
$\frac{\ln(x) - 10}{1.5}$	-2.3263	`=NORMSINV(0.99)`
$\ln(x)$	6.5105	`=-2.3263*1.5+10`
x	672.15	`=EXP(6.5105)`

Example. Continue last example, determine the mean and standard deviation of lifetime.

Sol.

$$\begin{aligned} E(X) &= e^{\theta+\omega^2/2} \\ &= e^{(10+2.25/2)} \\ &= e^{11.125} = 67,846.3 \end{aligned}$$

and

$$\begin{aligned} Var(X) &= e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right) \\ &= e^{20+2.25}(e^{2.25} - 1) \\ &= 39,070,059,886.6. \end{aligned}$$

Thus $\sigma = \sqrt{e^{20+2.25}(e^{2.25} - 1)} = 197,661.5.$

3.9 Beta Distribution

Definition. The random variable X with probability density function

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{for } x \in [0,\ 1]$$

is a **beta random variable** with parameters $\alpha > 0$ and $\beta > 0$.

Example. The service of a constant-velocity joint in an automobile requires disassembly, boot replacement, and assembly. Suppose that the proportion of the total service time for disassembly follows a beta distribution with $\alpha = 2.5$ and $\beta = 1$. What is the probability that a disassembly proportion exceeds 0.7?

Sol. Let X denote the proportion of service time for disassembly.

$$\begin{aligned} P(X > 0.7) &= \int_{0.7}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \int_{0.7}^{1} \frac{\Gamma(3.5)}{\Gamma(2.5)\Gamma(1)} x^{1.5} \\ &= \frac{2.5\times 1.5\times 0.5\sqrt{\pi}}{1.5\times 0.5\sqrt{\pi}} \frac{x^{2.5}}{2.5}\Big|_{0.7}^{1} \\ &= 1 - 0.7^{2.5} = 0.59 \end{aligned}$$

Definition. If X has a beta distribution with parameters α and β,

$$\mu = E(X) = \frac{\alpha}{\alpha+\beta}$$

and

$$\sigma^2 = Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Example. In above example, $\alpha = 2.5$ and $\beta = 1$. What are the mean and variance of this distribution?

Sol.

$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{2.5}{2.5+1} = \frac{5}{7}$$

and

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{2.5}{3.5^2\times 4.5} = 0.045.$$

3.9.1 Mode and Median

Definition. The mode(眾數) of a continuous pdf is the value at which its pdf has its maximum value.
i.e. peak 發生的位置.

Definition. (The mode of beta distribution) If $\alpha > 1$ and $\beta > 1$, the mode (peak of the density) is in the interior of [0, 1] and equals

$$\text{mode} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Example. For the above example, $\alpha = 2.5$ and $\beta = 1$. the mode is

$$\begin{aligned}\text{mode} &= \frac{\alpha - 1}{\alpha + \beta - 2} \\ &= \frac{2.5 - 1}{2.5 + 1 - 2} \\ &= \frac{1.5}{1.5} = 1\end{aligned}$$

Definition. If x is the median of random variable, then it is equivalent to

$$P(X < x) = P(X > x).$$

3.10 Normal Distribution(常態分布)

3.10.1 Normal Distribution (Gaussian Distribution)

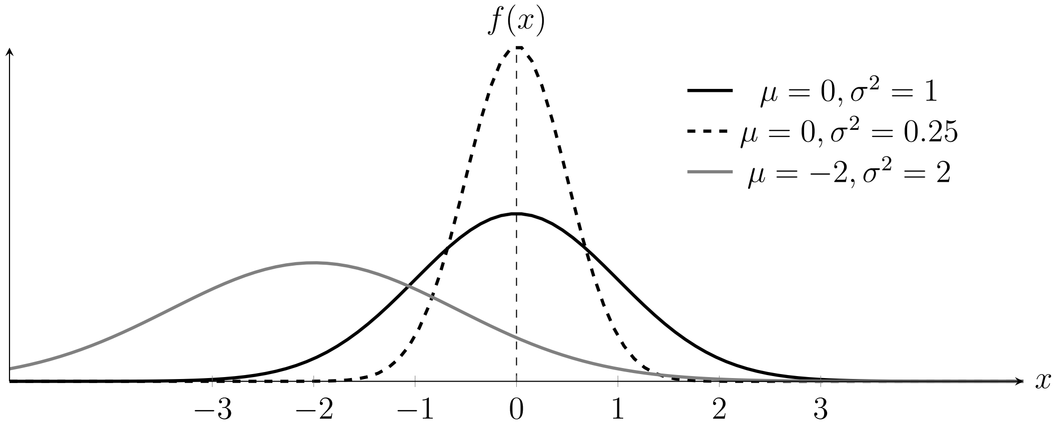
Definition. A random variable X with probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$ and $\sigma > 0$. Also,

$$E(X) = \mu \text{ and } Var(X) = \sigma^2$$

and the notation $N(\mu,\ \sigma^2)$ is used to denote the distribution.



Normal probability density functions for selected values of the parameters μ and σ^2

Theorem. The mean and variance of normal distribution is equal to μ and σ^2.

Proof. Let $u = \frac{x-\mu}{\sqrt{2}\sigma}$, then $du = \frac{1}{\sqrt{2}\sigma}dx$ and implies $\sqrt{2}\sigma du = dx$.

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\left(\dfrac{x-\mu}{\sqrt{2}\sigma}\right)^2} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma u + \mu) e^{-u^2} \sqrt{2}\sigma du \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma u + \mu) e^{-u^2} du \\
&= \frac{1}{\sqrt{\pi}} \left(\int_{-\infty}^{\infty} \sqrt{2}\sigma u e^{-u^2} du + \int_{-\infty}^{\infty} \mu e^{-u^2} du \right)
\end{aligned}$$

Let $\theta = u^2$, then $d\theta = 2udu$. Since $\theta \to \infty$ as $u \to \infty$ and $u \to -\infty$, we obtain

$$\int_{-\infty}^{\infty} \sqrt{2}\sigma u e^{-u^2} du = \int_{\infty}^{\infty} \frac{\sqrt{2}}{2}\sigma e^{-\theta} d\theta = 0.$$

Thus

$$\begin{aligned} E(X) = \frac{1}{\sqrt{\pi}}\left(\int_{-\infty}^{\infty} \sqrt{2}\sigma u e^{-u^2} du + \int_{-\infty}^{\infty} \mu e^{-u^2} du\right) &= \frac{1}{\sqrt{\pi}}(0 + \mu\sqrt{\pi}) \\ &= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} \\ &= \mu. \end{aligned}$$

Next, we prove that the variance is σ^2. By definition, $Var(X) = E[(X-\mu)^2]$.

$$\begin{aligned} Var(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

Let $z = \dfrac{x-\mu}{\sigma}$, then $dz = \dfrac{1}{\sigma}dx$, which implies $dx = \sigma dz$. Substituting these into the integral, we get:

$$\begin{aligned} Var(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{-\dfrac{z^2}{2}} \sigma dz \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\dfrac{z^2}{2}} dz \end{aligned}$$

Using integration by parts for $\displaystyle\int z^2 e^{-\dfrac{z^2}{2}} dz$. Let $w = z$ and $dv = ze^{-\dfrac{z^2}{2}} dz$. Then $dw = dz$ and $v = -e^{-\dfrac{z^2}{2}}$.

$$\begin{aligned} \int_{-\infty}^{\infty} z^2 e^{-\dfrac{z^2}{2}} dz &= \left[-ze^{-\dfrac{z^2}{2}}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-e^{-\dfrac{z^2}{2}}\right) dz \\ &= 0 + \int_{-\infty}^{\infty} e^{-\dfrac{z^2}{2}} dz \end{aligned}$$

We know that the total area under the standard normal distribution (see Section 3.10.2) curve is 1, meaning $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\dfrac{z^2}{2}} dz = 1$, which implies $\displaystyle\int_{-\infty}^{\infty} e^{-\dfrac{z^2}{2}} dz = \sqrt{2\pi}$.

Substituting this back, we obtain:

$$Var(X) = \frac{\sigma^2}{\sqrt{2\pi}}\left(\sqrt{2\pi}\right)$$
$$= \sigma^2.$$

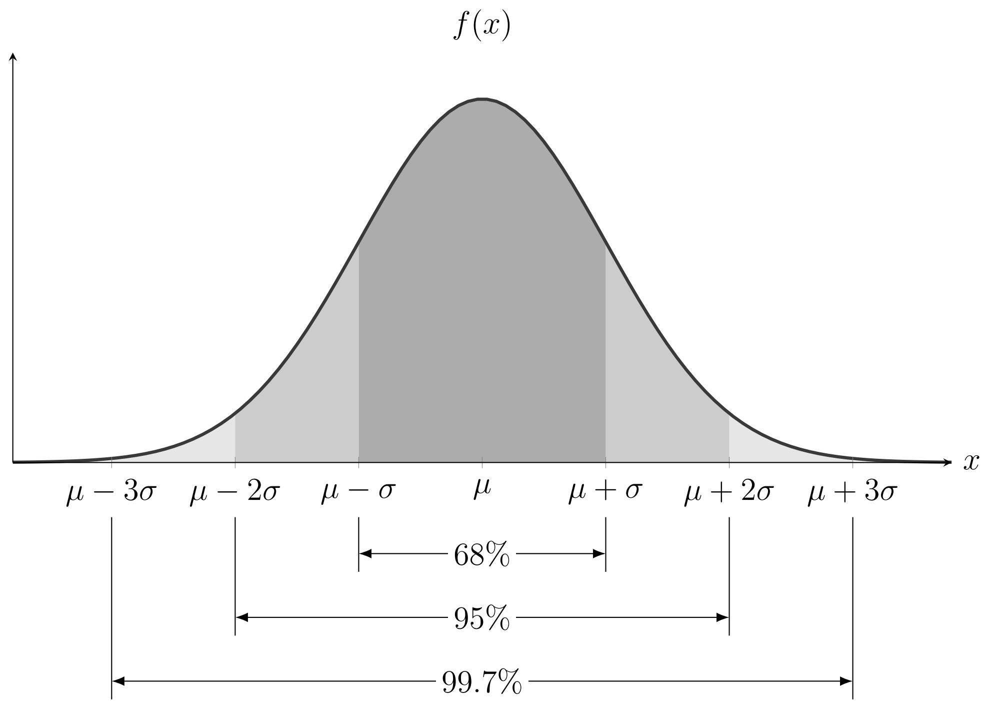
□

Property. For any normal distribution,

(1) $P(\mu - \sigma < X < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0.6827 = 68.27\%$

(2) $P(\mu - 2\sigma < X < \mu + 2\sigma) = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx = 0.9545 = 95.4\%$

(3) $P(\mu - 3\sigma < X < \mu + 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0.9973 = 99.7\%$



3.10.2 Standard Normal Random Variable(標準化常態分佈)

Definition. A normal random variable with

$$\mu = 0 \text{ and } \sigma^2 = 1$$

is called **a standard normal random variable** and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = P(Z \leq z).$$

$$\begin{array}{c} \text{Common Normal } X \sim N(\mu,\ \sigma^2) \\ \Big\downarrow Z = \dfrac{X-\mu}{\sigma} \\ \text{Standard Normal } Z \sim N(0,\ 1) \end{array}$$

Example. Let $X \sim N(5,\ 100)$. Find $P(X \leq 3)$.

Sol. (Method 1: Integral compute)
Let $f(x) = \dfrac{1}{10\sqrt{2\pi}} e^{-\dfrac{(x-5)^2}{200}}$.
Then

$$\begin{aligned} P(X \leq 3) &= \int_{-\infty}^{3} \frac{1}{10\sqrt{2\pi}} e^{-\dfrac{(x-5)^2}{200}} dx \\ &= 0.4207 \end{aligned}$$

(Method 2: Normalize and use table or Excel)
Let $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-5}{10}$.
Then

$$\begin{aligned} P(X \leq 3) &= P(Z \leq \frac{3-5}{10}) = P(Z \leq -0.2) \\ &= 0.4207 \end{aligned}$$

Excel: `=NORM.DIST(3, 5, 10, TRUE)` or `=NORM.S.DIST(-0.2, TRUE)`

3.11 Normal Approximation(常態近似)

3.11.1 Normal Approximation to the Binomial Distribution

Definition. If X is a binomial random variable with parameters n and p, and

(1) $np > 5$

(2) $n(1-p) > 5$,

then

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.

3.11.2 Normal Approximation to the Poisson Distribution

Definition. If X is a Poisson random variable with $E(X) = Var(X) = \lambda$, and $\lambda > 5$, then

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variable.

3.11.3 Normal Approximation to the Hypergeometric Distribution

Definition. If X is a hypergeometric random variable with parameters n, N and k with $\frac{n}{N} < 0.1$. Let $p = \frac{k}{N}$, using the binomial with normal approximation, we can find Z is approximately a standard normal random variable.

4 Some Additional Property

4.1 Lack of Memory Property

Definition. (Memoryless) A random variable X is **memoryless** if

$$P(X > s + t | X > s) = P(X > t).$$

Example. Geometric distribution $X \sim G(p)$ counts trials until first success. We have

$$P(X > k) = (1 - p)^k.$$

Show that $P(X > m + n | X > m) = P(X > n)$.

Proof.

$$\begin{aligned} P(X > m + n | X > m) &= \frac{(1-p)^{m+n}}{(1-p)^m} \\ &= \frac{(1-p)^m (1-p)^n}{(1-p)^m} \\ &= (1-p)^n = P(X > n). \end{aligned}$$

□

Example. Exponential distribution $X \sim E(\lambda)$ with $P(X > t) = e^{-\lambda t}$, prove that Exponential RV is memoryless.

Proof.

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P(X > t). \end{aligned}$$

□

Example. Let X denote the time between detections of a particle with a Geiger counter. Assume X has an exponential distribution with $E(X) = 1.4$ minutes. What is the probability that a particle is detected in the next 30 seconds?

Sol.

$$P(X < 0.5) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

Or use Excel `=EXPONDIST(0.5, 1/1.4, TRUE)`.

Example. Continue last example, no particle has been detected in the last 3 minutes. Will the probability increase since it is "due"?

Sol.

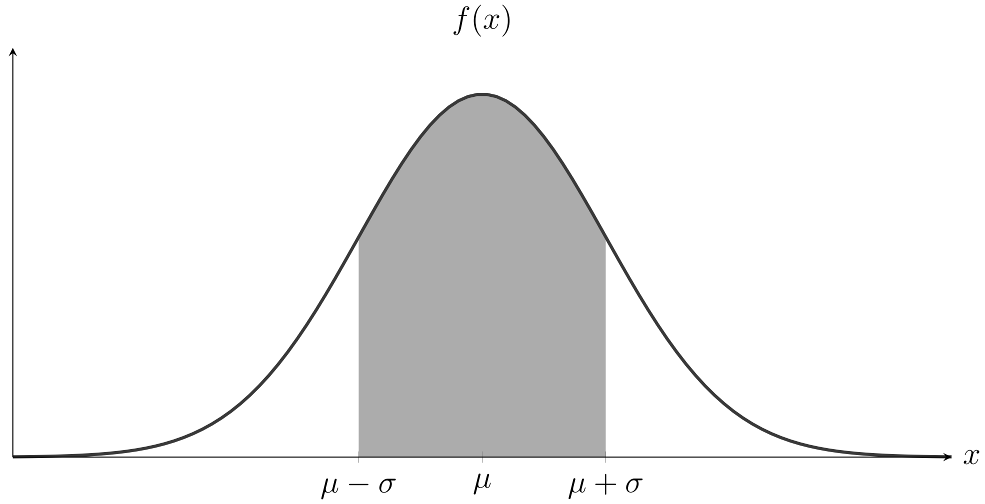
$$P(X < 3.5 | X > 3) = \frac{P(3 < X < 3.5)}{P(X > 3)} = \frac{\displaystyle\int_3^{3.5} \lambda e^{-\lambda x} dx}{\displaystyle\int_3^{\infty} \lambda e^{-\lambda x} dx} = 0.3.$$

4.2 Chebyshev's Inequality (柴比雪夫不等式)

Theorem. (Chebyshev's inequality) Let X be a (unknown) random variable with a finite non-zero variance σ^2 and mean μ. Then for any $k > 0$,

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

Example. Let $k = 1$.



We find that

$$P(|X - \mu| < 1 \cdot \sigma) = P(|X - \mu| < \sigma) \geq 1 - \frac{1}{1^2} = 0.$$

Example. Suppose RV. M is the total staffing-hours worked without a serious accident. If the mean of M is 2 million hours and the standard deviation is 0.1 million hours. If a serious accident has just occurred. Using the Chebyshev's inequality, would it be unusual for the next serious accident within the next 1.6 million staffing-hours?

Sol. k is unknown and

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

we want to estimate $P(M \leq 1.6)$. Since $|X - \mu| = |1.6 - 2| = 0.4$ and $\sigma = 0.1$, then $k = \dfrac{0.4}{0.1} = 4$. Using the Chebyshev's inequality,

$$P(|M - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$
$$P(|M - 2| < 0.4) \geq 1 - \frac{1}{4^2} = 1 - \frac{1}{16} = 1 - 0.0625$$

Hence $P(M \leq 1.6) = 0.0625$.

4.3 Central Limit Theorem (中央極限定理)

Theorem. (Central Limit Theorem) Let $x_1,\ x_2,\ \ldots,\ x_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, RV $\bar{X}$[2] is approximately normal distribution with mean μ and variance $\dfrac{\sigma^2}{n}$.

Example.

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right)$$

$$\Downarrow \text{Standarize}$$

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim Z(0,\ 1)$$

[2] $\bar{X} = \text{Sample mean} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$.