Chapter 4: Special Probability Distributions

1 Discrete Distributions

- Recall that for a discrete random variable X, the number of possible values (i.e., R_X) is **finite** or **countable**.
- The elements of R_X can be listed as $x_1, x_2, x_3, \ldots$.
- In this section, we study some classes of discrete random variables.

Discrete Uniform Distribution

DEFINITION 1

- *If RV X assumes the values $x_1, x_2, \ldots, x_k$ with equal probability, then X follows a* ***discrete uniform distribution****.*

- *The p.f. for X is given by*

$$f_X(x) = \frac{1}{k}, \qquad x = x_1, x_2, \ldots, x_k,$$

and 0 otherwise.

THEOREM 2

Suppose X follows the discrete uniform distribution with $R_X = \{x_1, x_2, \ldots, x_k\}$, we have

- *The expectation is given by*

$$\mu_X = E(X) = \sum_{i=1}^{k} x_i f_X(x_i) = \frac{1}{k} \sum_{i=1}^{k} x_i.$$

- *The variance is given by*

$$\sigma_X^2 = V(X) = E(X^2) - (E(X))^2 = \frac{1}{k} \sum_{i=1}^{k} x_i^2 - \mu_X^2.$$

Example 4.1

- A bulb is selected at random from a box that contains a 40-watt bulb, a 60-watt bulb, an 80-watt bulb, and a 100-watt bulb.
- Each bulb has $1/4$ probability of being selected.
- Let $X =$ the watts of the bulb being selected. Then X follows a uniform distribution, and

$$R_X = \{40, 60, 80, 100\}.$$

$$f_X(x) = 1/4, \qquad \text{for } x = 40, 60, 80, 100,$$

and 0 otherwise.

- We can compute the expectation:

$$E(X) = \sum_i x_i f_X(x_i) = 40 \cdot (1/4) + 60 \cdot (1/4) + 80 \cdot (1/4) + 100 \cdot (1/4) = 70$$

- Variance can also be computed:

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= 40^2 \cdot (1/4) + 60^2 \cdot (1/4) + 80^2 \cdot (1/4) + 100^2 (1/4) - 70^2 \\ &= 500. \end{aligned}$$

Bernoulli Trial, Bernoulli Random Variable and Bernoulli Process

DEFINITION 3 (BERNOULLI TRIAL)

- *A **Bernoulli Trial** is a random experiment with only two possible outcomes.*
- *One is called a "success", and the other a "failure".*
- *We code the two outcomes as "1" (success) and "0" (failure).*

DEFINITION 4 (BERNOULLI RANDOM VARIABLE)

- *Let $X =$ number of success in a Bernoulli trial; then X has only two possible values: 1 or 0, and is called a* ***Bernoulli random variable****.*

- *Denote by p $(0 \le p \le 1)$ the probability of success of the Bernoulli trial. Then X has the p.f.:*

$$f_X(x) = P(X = x) = \begin{cases} p & x = 1 \\ (1-p) & x = 0 \end{cases},$$

  and $= 0$ for other values of x.

- *This p.f. can also be written by*

$$f_X(x) = p^x(1-p)^{1-x}, \quad \text{for } x = 0 \text{ or } 1.$$

- We often denote $X \sim \text{Bernoulli}(p)$, and denote $q = 1 - p$. Then the p.f. becomes $f_X(1) = p$ and $f_X(0) = q$.

THEOREM 5

For a Bernoulli RV defined above, we have

$$\begin{aligned} \mu_X &= E(X) = p \\ \sigma_X^2 &= V(X) = p(1-p) = pq. \end{aligned}$$

Remark (Parameters)

- In occasions, $f_X(x)$ may rely on one or more unknown quantities; different values of the quantities lead to different probability distributions.

- Such a quantity is called a **parameter** of the distribution.

- p is the parameter in the Bernoulli distribution.

- The collection of the distributions that are determined by one or more unknown parameters is called a **family of probability distributions**.

- So the aforementioned Bernoulli distributions determined by the parameter p is a family of probability distributions.

Example 4.2 The following are all examples of Bernoulli trials:

- A coin toss
  Say we want heads, then H="heads" is success, and T="tails" is failure.

- Rolling a die
  Say we only care about rolling a 6. The outcome space is binarized to "success"$= \{6\}$ and "failure" $= \{1, 2, 3, 4, 5\}$.

- Polls
  Choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.

Example 4.3

- A box contains 4 blue and 6 red balls.
- Draw a ball from the box at random.
- What is the probability that a blue ball is chosen?

Solution:

- Let $X = 1$ if a blue ball is drawn; and $X = 0$ otherwise.
- Then X is a Bernoulli random variable.
- $P(X = 1) = 4/10 = 0.4$.
- Furthermore, the p.f. for X is given by

$$f_X(x) = \begin{cases} 0.4 & x = 1 \\ 0.6 & x = 0 \end{cases}.$$

DEFINITION 6 (BERNOULLI PROCESS)

- *A **Bernoulli process** consists of a sequence of repeatedly performed **independent and identical** Bernoulli trials.*

- *Correspondingly, a Bernoulli process generates a sequence of **independent and identically distributed, i.i.d.** Bernoulli random variables:* $X_1, X_2, X_3, \ldots$.

We are able to define several useful distributions based on Bernoulli trial and Bernoulli process. These distributions include:

- **Binomial distribution**;
- **Negative Binomial distribution**; **Geometric distribution**;
- **Poisson distribution**.

Binomial Distribution

If we have several (say n) i.i.d. Bernoulli trials, we can establish the binomial distribution to address some interesting questions. For example,

- A student randomly guesses at 5 multiple-choice questions. What is the number of questions the student guessed correctly?
- Randomly pick a family with 4 kids. What is the number of girls amongst the kids?
- Urn has 4 black balls and 3 white balls, draw 5 balls with replacement. How many black balls will there be?

Definition 7 (Binomial random variable)

*A **Binomial random variable** counts the number of successes in n trials in a Bernoulli Process. That is, suppose we have n trials where*

- *the probability of success for each trial is the same p,*
- *the trials are independent.*

Then the number of successes, denoted by X, in the n trials is a Binomial random variable.

We say X has a binomial distribution and write it as $X \sim B(n, p)$.

> *The probability of getting exactly x successes is given as*
>
> $$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \ \textit{for } x = 0, 1, 2, \ldots, n.$$
>
> *It can be shown that $E(X) = np$, and $V(X) = np(1-p)$.*

The theoretical development for Binomial distribution will be given in a lecture meeting.

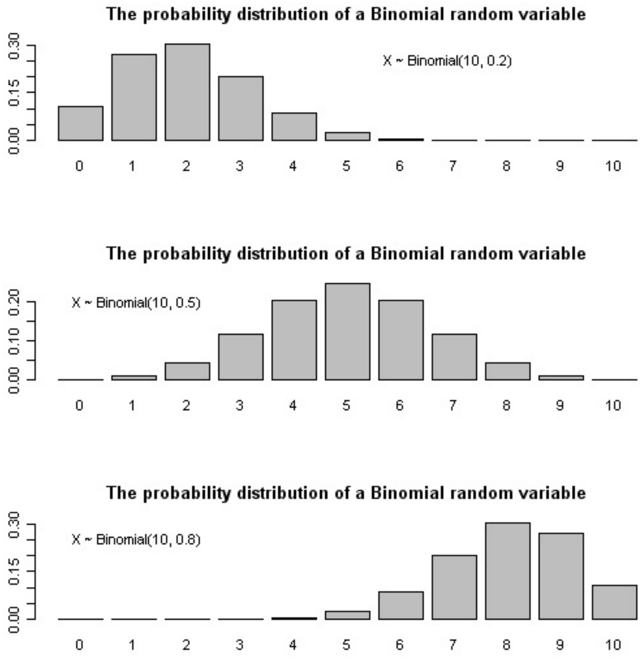
REMARK

- When $n = 1$, the p.f. for the binomial RV X is reduced to

$$f_X(x) = p^x(1-p)^{1-x}, \quad \text{for } x = 0, 1.$$

- It is the p.f. for the Bernoulli distribution. Therefore Bernoulli distribution is a special case of the binomial distribution.

The p.f. for $B(10, 0.2)$, $B(10, 0.5)$, and $B(10, 0.8)$ are compared below.



Example 4.4

- Flip a fair coin 10 independent times.
- What is the probability of observing exactly 6 heads?

Solution:

- Let $X =$ number of heads in 10 flips.
- Each flip of the coin can be observed as a Bernoulli trial, with probability of getting head (success) $p = 0.5$.
- Then X is the number success out of 10 Bernoulli trials; so $X \sim B(10, 0.5)$.
- We can compute

$$P(X = 6) = \binom{10}{6}(0.5)^6(1-0.5)^{10-6} = 0.205.$$

Negative Binomial Distribution

- Consider a Bernoulli process, where the Bernoulli experiments can be repeated an arbitrary number of times.

- The interest could be how many trials are needed so that a certain number of successes occur.

- Set $X =$ number of trials until the kth success occurs. Then X follows a **negative binomial distribution**; denoted by $X \sim NB(k, p)$, where p is probability of success for each Bernoulli trial.

- In comparison with binomial distribution: the random variable "X" is the number of successes out of a fixed number n of trials.

Definition 8 (Negative Binomial Distribution)

- *X = number of i.i.d. Bernoulli(p) trials until the kth success occurs; then X follows a **negative binomial distribution**, denoted by $X \sim NB(k, p)$.*

- *The p.f. of X is given by*

$$f_X(x) = P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k},$$

  for $x = k, k+1, k+2, \ldots$

- *It can be shown that $E(X) = k/p$ and $V(X) = (1-p)k/p^2$.*

Example 4.5

- Keep rolling a fair die, until the 6th time we get the number 6.
- What is the probability that we need to roll the die 10 times?

Solution:

- Let $X =$ number of rolls to get the 6th number 6. $X \sim NB(6, 1/6)$.
- Using the p.f. of negative binomial distribution:

$$P(X = 10) = \binom{10-1}{6-1}(1/6)^6(1-1/6)^4 = 0.001302.$$

Geometric Distribution

Geometric distribution is a special case of the negative binomial distribution.

DEFINITION 9 (GEOMETRIC DISTRIBUTION)

- *X = number of i.i.d. Bernoulli(p) trials until the first success occurs; then X follows a* ***geometric distribution****, denote by $X \sim G(p)$.*

- *The p.f. of X is given by*

$$f_X(x) = P(X = x) = (1 - p)^{x-1} p.$$

- *We have $E(X) = 1/p$ and $V(X) = (1 - p)/p^2$.*

Poisson Distribution

Definition 10 (Poisson Random Variable)
The ***Poisson random variable*** *X denotes the number of events occurring in a* ***fixed period of time or fixed region****.*

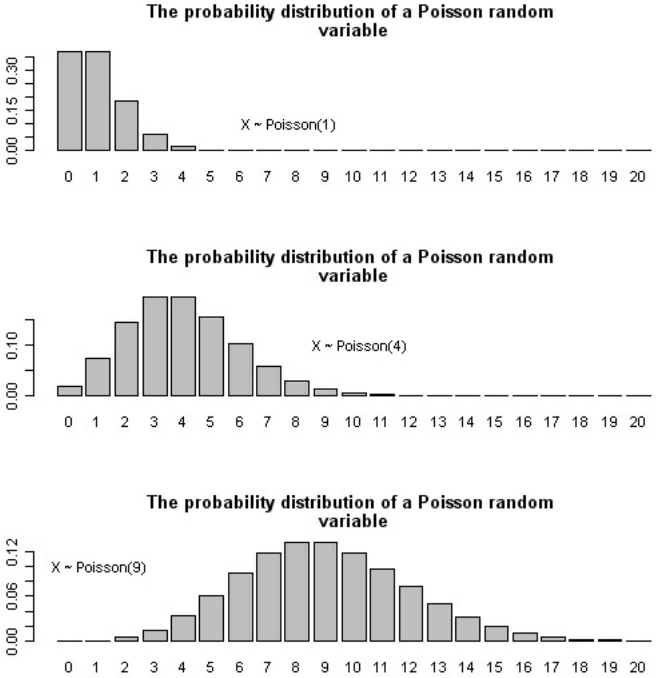
We denote $X \sim \text{Poisson}(\lambda)$ where parameter $\lambda > 0$ is the expected number of occurrences during the given period/region; its p.m.f. is given by

$$f_X(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where $k = 0, 1, \ldots$ is the number of occurrences of events.

It can be shown that $E(X) = \lambda$, and $V(X) = \lambda$.

The p.f. for Poisson(1), Poisson(4), and Poisson(9) are compared below.



Example 4.6 The “fixed period of time or fixed region" given in the definition can be time period of any length, e.g., a minute, a day, a week, a month etc., and region of any size.

Examples of events that may be modeled by the Poisson Distribution:

(a) The number of spelling mistakes one makes while typing a single page.

(b) The number of times a web server is accessed per minute.

(c) The number of road kill (animals killed) found per unit length of road.

(d) The number of mutations in a given stretch of DNA after a certain amount of radiation exposure.

(e) The number of unstable atomic nuclei that decayed within a given period of time in a piece of radioactive substance.

(f) The distribution of visual receptor cells in the retina of the human eye.

(g) The number of light bulbs that burn out in a certain amount of time.

DEFINITION 11 (POISSON PROCESS)
*The **Poisson Process** is a continuous time process. We count the number of occurrences within some interval of time. The defining properties of a Poisson Process with rate parameter α are*

- *the expected number of occurrences in an interval of length T is αT;*
- *there are no simultaneous occurrences;*
- *the number of occurrences in disjoint time intervals are independent.*

The number of occurrences in any interval T of a Poisson Process follows a $\text{Poisson}(\alpha T)$ *distribution.*

Example 4.7

- The average number of robberies in a day is four in a certain big city.
- What is the probability that six robberies occurring in two days?

Solution:

- Let X_1 = number of robberies in one day. Then $X_1 \sim \text{Poisson}(4)$ from the condition.
- Let X = number of robberies in two days. Then $X \sim \text{Poisson}(2 \times 4) = \text{Poisson}(8)$.
- We have

$$P(X = 6) = \frac{e^{-8} 8^6}{6!} = 0.1222.$$

PROPOSITION 12 (POISSON APPROX. OF BINOMIAL DISTRIBUTION) *Let $X \sim B(n, p)$. Suppose that $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains a constant. Then approximately, $X \sim \text{Poisson}(np)$. That is*

$$\lim_{p \to 0; n \to \infty} P(X = x) = \frac{e^{-np}(np)^x}{x!}.$$

The approximation is good when $n \geq 20$ and $p \leq 0.05$, or if $n \geq 100$ and $np \leq 10$.

Example 4.8

- The probability, p, of an individual car having an accident at a junction is 0.0001.
- If there are 1000 cars passing through the junction during certain period of a day, what is the probability of two or more accidents occurring during that period?

Solution:

- Let $X =$ number of accidents among the 1000 cars.
- Then $X \sim B(1000, 0.0001)$. If we compute using binomial distribution,

$$P(X \geq 2) = \sum_{x=2}^{1000} \binom{1000}{x} 0.0001^x 0.9999^{1000-x}.$$

- Computing these numbers is not easy.

- We solve the question using Poisson approximation.
- $n = 1000$ and $p = 0.0001$, hence, $np = \lambda = 0.1$.
- Thus

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-0.1} - e^{-0.1}(0.1)^1/1! \\ &= 0.0047. \end{aligned}$$

2 Continuous Distribution

- For a continuous random variable X, its range R_X is an interval or a collection of multiple intervals.

- In this section, we study some classes of continuous random variables.

Continuous Uniform Distribution

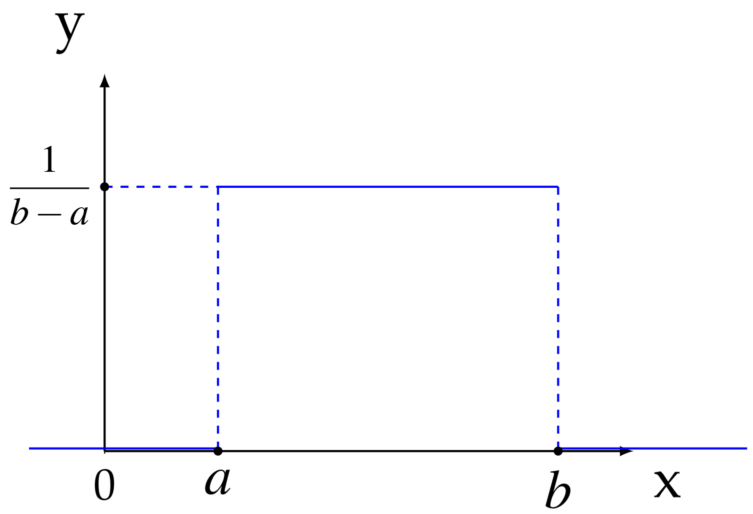
Definition 13 (Continuous Uniform Distribution) *A random variable X is said to follow a **uniform distribution** over the interval (a,b) if its probability density function is given by*

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \textit{otherwise} \end{cases}.$$

We denote this by $X \sim U(a,b)$.

It can be shown that $E(X) = \dfrac{a+b}{2}$ and $V(X) = \dfrac{(b-a)^2}{12}$.

The p.d.f. for the continuous uniform distribution can be drawn as a figure below.



The c.d.f. for the continuous uniform distribution is given by

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Example 4.9

- A point is chosen at random on the line segment $[0, 2]$.
- What is the probability that the chosen point lies between 1 and $3/2$?

Solution:

- Let X = position of the point. $X \sim U(0,2)$.
- We have

$$f_X(x) = \frac{1}{2}, \quad \text{for } 0 \leq x \leq 2,$$

and 0 otherwise.

$$P\left(1 \leq X \leq \frac{3}{2}\right) = \int_1^{3/2} \frac{1}{2} dx = \frac{1}{2} x\Big|_1^{3/2} = 1/4.$$

Exponential Distribution

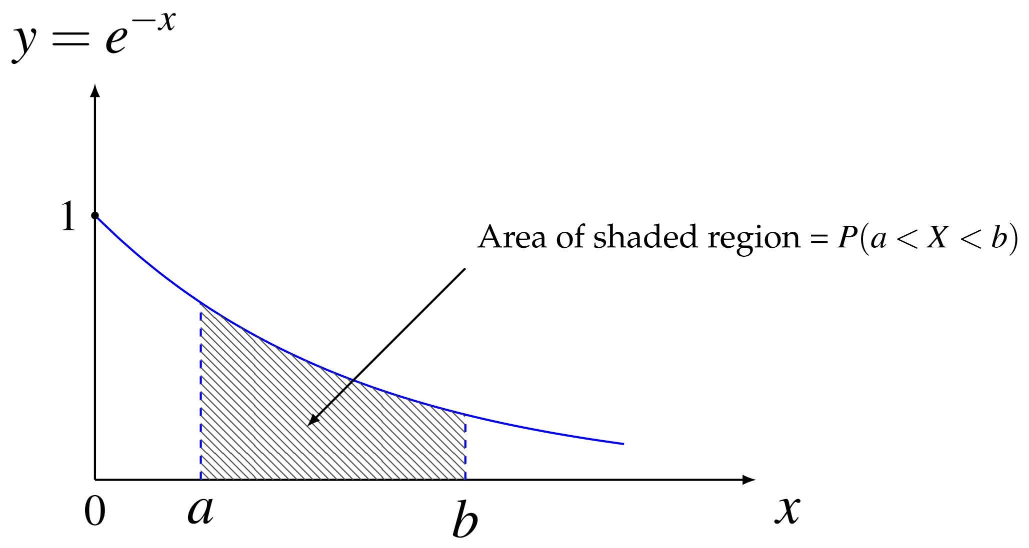
Definition 14 (Exponential Distribution)
A continuous RV X is said to follow an ***exponential distribution*** *with parameter $\lambda > 0$ if its p.d.f. is given by*

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

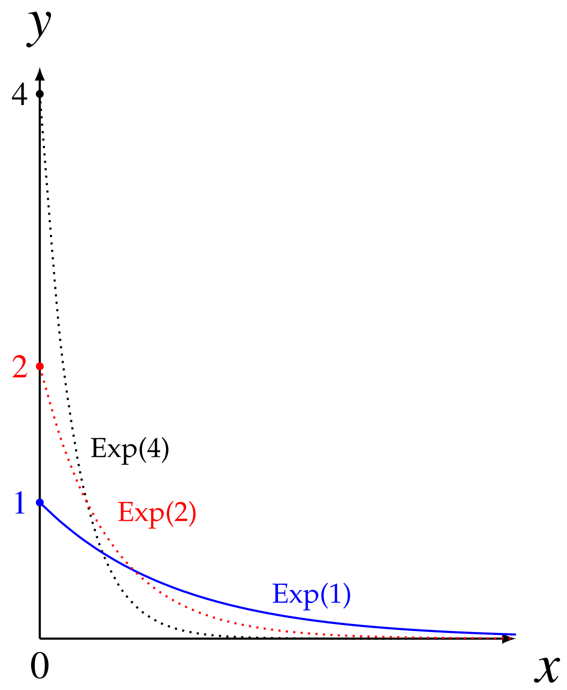
We denote $X \sim \mathrm{Exp}(\lambda)$.

It can be shown that $E(X) = \dfrac{1}{\lambda}$ and $V(X) = \dfrac{1}{\lambda^2}$.

The exponential p.d.f. with $\lambda = 1$ is shown below.



The shapes of the p.d.f.s of $\mathrm{Exp}(\lambda)$ for $\lambda = 1, 2, 4$.



The c.d.f. of $X \sim \text{Exp}(\lambda)$ is given by

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

REMARK

- The p.d.f. can be written in the form

$$f_X(x) = \frac{1}{\mu} e^{-x/\mu}, \quad \text{for } x > 0,$$

  and 0 elsewhere.

- The parameters have the relationship $\mu = 1/\lambda$.

- We have

$$E(X) = \mu, \quad V(X) = \mu^2, \quad \text{and} \quad F_X(x) = 1 - e^{-x/\mu} \quad \text{for } x > 0.$$

Example 4.10

- Suppose that the failure time, T, of a system is exponentially distributed, with a mean of 5 years.
- What is the probability that at least two out of five of these systems are still functioning at the end of 8 years?

Solution:

- Since $E(T) = 5$, therefore $\lambda = 1/5$.
- We have $T \sim \text{Exp}(1/5)$,

$$P(T > 8) = 1 - P(T \leq 8) = 1 - F_X(8) = e^{-(1/5)\times 8} = e^{-1.6} \approx 0.2.$$

- Let X = # of systems out of 5 that are still functioning after 8 years.
- Then $X \sim B(5, 0.2)$. Hence,

$$P(X \geq 2) = 0.2627.$$

Theorem 15

Suppose that X has an exponential distribution with parameter $\lambda > 0$. Then for any two positive numbers s and t, we have

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

REMARK

The above theorem states that the exponential distribution has "**no memory**" in the sense:

- Let X denote the life length of a bulb.
- Given that the bulb has lasted s time units, i.e., $X > s$,
- the probability that it will last for the next t units, i.e., $X > s+t$, is the same as the probability that it will last for the first t units as brand new.

Normal Distribution

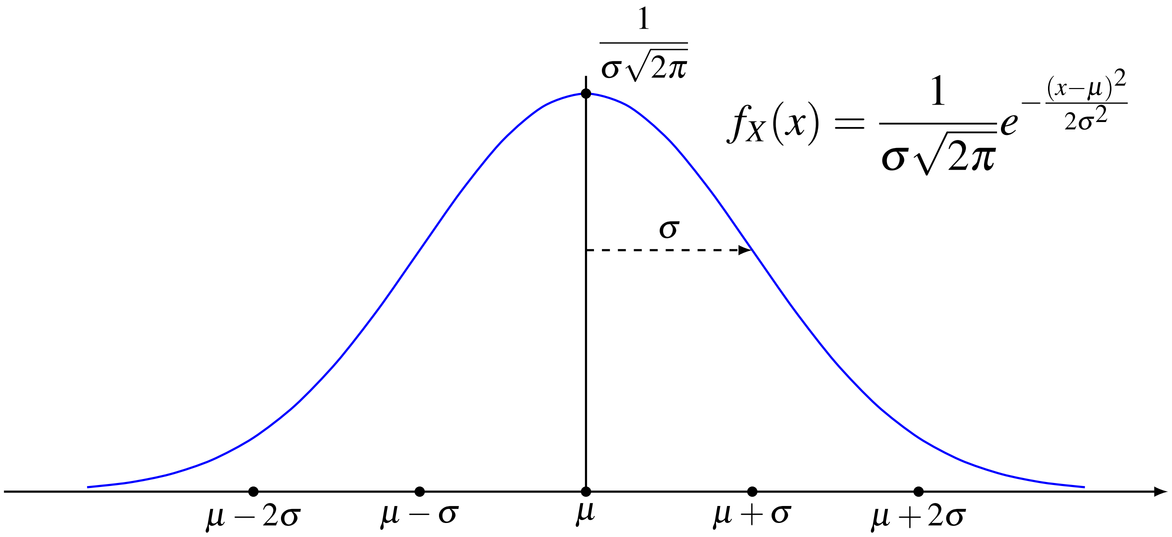
DEFINITION 16 (NORMAL DISTRIBUTION)
A random variable X is said to follow a ***normal distribution*** *with parameters μ and σ^2 if its probability density function is given by*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty.$$

We denote $X \sim N(\mu, \sigma^2)$.

It can be shown that $E(X) = \mu$ and $V(X) = \sigma^2$.

The p.d.f. of normal distribution is positive over the whole real line, symmetric about $x = \mu$, and bell-shaped; see below.



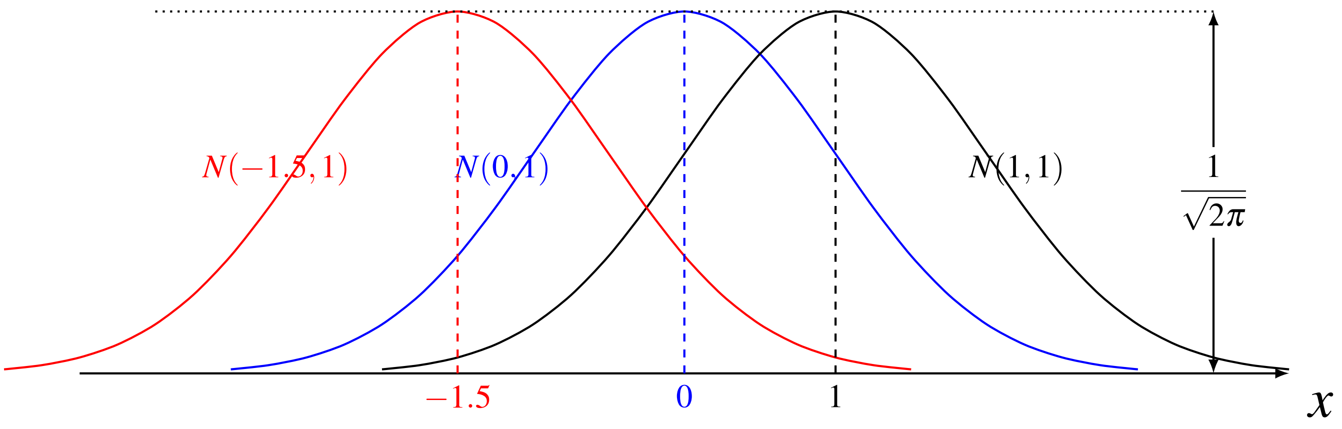
We give some properties of normal distribution.

(1) The total area under the curve and above the horizontal axis is equal to 1.

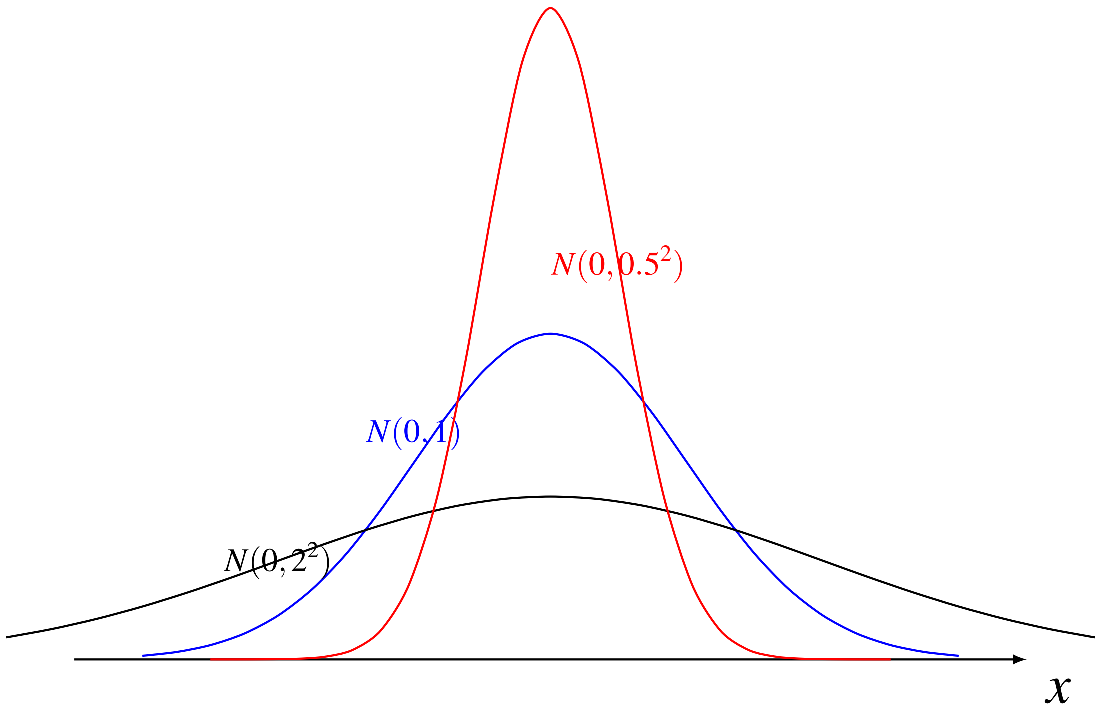
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1.$$

This validates that $f_X(\cdot)$ is a p.d.f.

(2) Two normal curves are identical in shape if they have the same σ^2. But they are centered at different positions when their means are different.



(3) As σ increases, the curve flattens; and vice versa.



(4) If $X \sim N(\mu, \sigma^2)$ and let

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

then Z follows the $N(0,1)$ distribution. Thus $E(Z) = 0$ and $V(Z) = 1$. We say that Z has a standardized normal distribution; the p.d.f. of Z is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

Remark

- The importance of the standardized normal distribution is that it can be tabulated.

- Consider $X \sim N(\mu, \sigma^2)$; if we are to compute $P(x_1 < X < x_2)$ for any real values x_1, x_2, we can use the transformation $Z = (X - \mu)/\sigma$. In particular,

$$x_1 < X < x_2 \Longleftrightarrow \frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}$$

- Let $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$; then

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2).$$

- By convention, we use $\phi(\cdot)$ and $\Phi(\cdot)$ to denote the p.d.f. and c.d.f. of the standard normal distribution respectively. That is,

$$\begin{aligned}\phi(z) &= f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\ \Phi(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \phi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.\end{aligned}$$

- Therefore, for $X \sim N(\mu, \sigma^2)$ and any real numbers x_1, x_2,

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

- However, calculating the probabilities for the normal probabilities is challenging because
  - there is no close formula for $\Phi(z)$;
  - so the computation relies on numerical integration.
- Instead, $\Phi(z)$ can be tabulated, or computed based on some statistical software.

- The standard normal distribution has the following properties:
  - $\star$ $P(Z \geq 0) = P(Z \leq 0) = \Phi(0) = 0.5$;
  - $\star$ For any z, $\Phi(z) = P(Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z)$;
  - $\star$ $-Z \sim N(0,1)$;
  - $\star$ If $Z \sim N(0,1)$, then $\sigma Z + \mu \sim N(\mu, \sigma^2)$.

Example 4.11 Given $X \sim N(50, 100)$, find $P(45 < X < 62)$.

<u>Solution</u>: We have $\mu = 50$, $\sigma = 10$.

$$\begin{aligned} P(45 < X < 62) &= P\left(\frac{45-50}{10} < \frac{X-50}{10} < \frac{62-50}{10}\right) \\ &= P(-0.5 < Z < 1.2) \\ &= P(Z < 1.2) - P(Z \leq -0.5) \\ &= \Phi(1.2) - \Phi(-0.5), \end{aligned}$$

where $\Phi(1.2)$ and $\Phi(-0.5)$ can either be computed from some statistical software or obtained from a table.

Definition 17 (Quantile)
The αth (upper) quantile ($0 \leq \alpha \leq 1$) of the RV X is the number x_α that satisfies

$$P(X \geq x_\alpha) = \alpha.$$

- Specifically, we denote by z_α the αth (upper) quantile (or 100α percentage point) of $Z \sim N(0,1)$. That is

$$P(Z \geq z_\alpha) = \alpha.$$

- For example, $z_{0.05} = 1.645$, $z_{0.01} = 2.326$.

- Since the p.d.f. of Z, i.e., $\phi(z)$, is symmetrical about 0, therefore

$$P(Z \geq z_\alpha) = P(Z \leq -z_\alpha) = \alpha.$$

Example 4.12 Find z such that

(a) $P(Z < z) = 0.95$;

(b) $P(|Z| \leq z) = 0.98$.

Solution:

(a) We need z such that

$$P(Z > z) = 1 - P(Z < z) = 0.05;$$

therefore $z = z_{0.05} = 1.645$.

(b) We have

$$\begin{aligned} 0.98 &= P(|Z| \le z) = 1 - P(|Z| > z) \\ &= 1 - P(Z > z) - P(Z < -z) = 1 - 2P(Z > z), \end{aligned}$$

which implies $P(Z > z) = 0.01$; therefore $z = z_{0.01} = 2.326$.

- Recall that when $n \to \infty$, $p \to 0$, and np remains a constant, we can use **Poisson distribution to approximate the binomial distribution**.

- When $n \to \infty$, but p remains a constant (practically, p is not very close to 0 or 1), we can use **normal distribution to approximate the binomial distribution**.

- A good rule of thumb is to use the normal approximation only when

$$np > 5 \quad \text{and} \quad n(1-p) > 5.$$

PROPOSITION 18 (NORMAL APPROX. TO BINOMIAL DISTRIBUTION) *Let $X \sim B(n, p)$; so that $E(X) = np$ and $V(X) = np(1-p)$. Then as $n \to \infty$,*

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}} \text{ is approximately } \sim N(0,1).$$

Normal Approximation to the Binomial Distribution