# Chapter 7. Continuous Random Variables

Suppose again we have a random experiment with sample space $S$ and probability measure P.

Recall our definition of a random variable as a mapping $X : S \to \mathbb{R}$ from the sample space $S$ to the real numbers inducing a probability measure $P_X(B) = P\{X^{-1}(B)\}, B \subseteq \mathbb{R}$.

**Definition 7.0.1.** *A random variable X is (absolutely)* **continuous** *if* $\exists f_X : \mathbb{R} \to \mathbb{R}$ *(measurable) such that*

$$P_X(B) = \int_{x \in B} f_X(x) dx, \qquad B \subseteq \mathbb{R},$$

*in which case* $f_X$ *is referred to as the* **probability density function**, *or* **pdf**, *of X.*

### 7.0.6 Continuous Cumulative Distribution Function

**Definition 7.0.2.** *The* **cumulative distribution function** *of* **CDF**, $F_X$ *of a continuous random variable X is defined as*

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

***Note*** From now one, when we speak of a continuous random variable, we will implicitly assume the absolutely continuous case.

### 7.0.7 Properties of Continuous $F_X$ and $f_X$

By analogy with the discrete case, let $\mathbb{X}$ be the range of $X$, so that $\mathbb{X} = \{x : f_X(x) > 0\}$.

i) For the cdf of a continuous random variable,

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1.$$

ii) At values of $x$ where $F_X$ is differentiable

$$f_X(x) = \frac{d}{dt} F_X(t) \bigg|_{t=x} \equiv F'_X(x).$$

iii) If $X$ is continuous,

$$f_X(x) \neq P(X = x) = \lim_{h \to 0^+} [P(X \leq x) - P(X \leq x - h)] = \lim_{h \to 0^+} [F_X(x) - F_X(x - h)] = 0$$

**Warning!** People usually forget, that $P(X = x) = 0$ for all $x$, when $X$ is a continuous random variable.

iv) The pdf $f_X(x)$ is not itself a probability, then unlike the pmf of a discrete random variable we do not require $f_X(x) \leq 1$.

v) For $a < b$,

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

vi) From Definition 7.0.1 it is clear that the pdf of a continuous random variable $X$ completely characterises its distribution, so we often just specify $f_X$.

It follows that a function $f_X$ is a pdf for a continuous random variable $X$ if and only if

i) $f_X(x) \geq 0$,

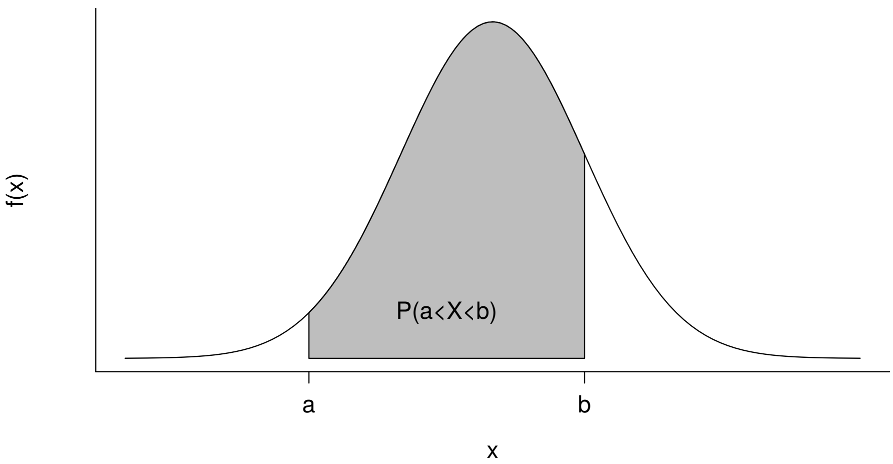
ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

This result follows direct from the definitions and properties of $F_X$.

Suppose we are interested in whether a continuous random variable $X$ lies in an interval $(a, b]$. Well, $P_X(a < X \leq b) = P_X(X \leq b) - P_X(X \leq a)$, which in terms of the cdf and pdf gives

$$\begin{aligned} P_X(a < X \leq b) &= F_X(b) - F_X(a) \\ &= \int_a^b f_X(x)dx. \end{aligned}$$

That is, the area under the pdf between $a$ and $b$.




**Example** Consider an experiment to measure the length of time that an electrical component functions before failure. The sample space of outcomes of the experiment, $S$ is $\mathbb{R}^+$ and if $A_x$ is the event that the component functions longer than $x > 0$ time units, suppose that $P(A_x) = \exp\{-x^2\}$.

Define continuous random variable $X : S \to \mathbb{R}^+$, by $X(s) = x \iff$ component fails at time $x$. Then, if $x > 0$

$$F_X(x) = P(X \leq x) = 1 - P(A_x) = 1 - \exp\{-x^2\}$$

and $F_X(x) = 0$ if $x \leq 0$. Hence if $x > 0$,

$$f_X(x) = \frac{d}{dt} F_X(t) \bigg|_{t=x} = 2x \exp\{-x^2\}$$

and zero otherwise.

Figure 7.1 displays the probability density function (left) and cumulative distribution function (right). Note that both the PDF and CDF are defined for all real values of $x$, and that both are continuous functions.

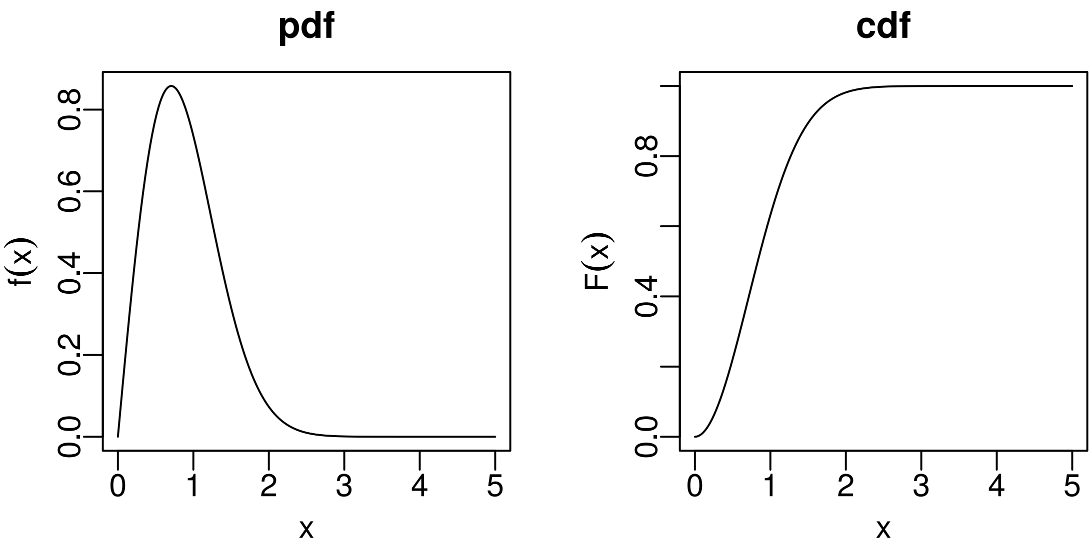



Figure 7.1: PDF $f_X(x) = 2x\exp\{-x^2\}$, $x > 0$, and CDF $F_X(x) = 1 - \exp\{-x^2\}$.

Also note that here

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} f_X(t)dt$$

as $f_X(x) = 0$ for $x \leq 0$, and also that

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{0}^{\infty} f_X(x)dx = 1$$

■

**Example** Suppose we have a continuous random variable $X$ with probability density function given by

$$f_X(x) = \begin{cases} cx^2, & 0 < x < 3 \\ 0, & \text{otherwise} \end{cases}$$

for some unknown constant $c$.

*Questions*

Q1) Determine $c$.

Q2) Find the cdf of $X$.

Q3) Calculate $\mathrm{P}(1 < X < 2)$.

*Solutions*

S1) We must have

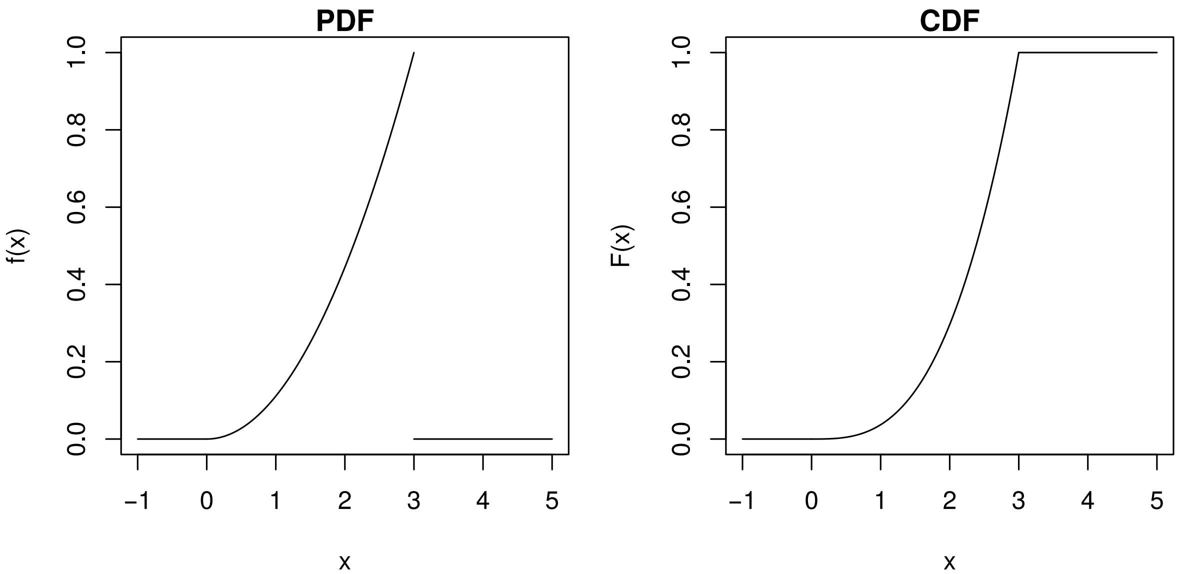
$$1 = \int_0^3 cx^2 dx = c\left[\frac{x^3}{3}\right]_0^3 = 9c$$
$$\implies c = \frac{1}{9}.$$

S2)

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_{-\infty}^{x} f(u)du = \int_0^x \frac{u^2}{9}du = \frac{x^3}{27} & 0 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

S3)

$$\mathrm{P}(1 < X < 2) = F(2) - F(1) = \frac{8}{27} - \frac{1}{27} = \frac{7}{27} = 0.2593.$$




■

### 7.0.8 Transformations

Suppose that $X$ is a continuous random variable $X$ with pdf $f_X$ and cdf $F_X$. Let $Y = g(X)$ be a function of $X$ for some (measurable) function $g : \mathbb{R} \to \mathbb{R}$ s.t. $g$ is continuous and strictly monotonic (so $g^{-1}$ exists) . We call $Y = g(X)$ a transformation of $X$.

Suppose $g$ is monotonic increasing. We can compute the pdf and cdf of $Y = g(X)$ as follows:
The cdf of $Y$ is given by

$$F_Y(y) = \mathrm{P}_Y(Y \leq y) = \mathrm{P}_Y(g(X) \leq y) = \mathrm{P}_X(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

The pdf of $Y$ is given by using the chain rule of differentiation:

$$f_Y(y) = F_Y'(y) = f_X\{g^{-1}(y)\}g^{-1\prime}(y)$$

Note $g^{-1\prime}(y) = \dfrac{d}{dy}g^{-1}(y)$ is positive since we assumed $g$ was increasing.

If $g$ monotonic decreasing, we have that

$$F_Y(y) = \mathrm{P}_Y(Y \leq y) = \mathrm{P}_Y(g(X) \leq y) = \mathrm{P}_X(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

By comparison with before, we would have

$$f_Y(y) = F_Y'(y) = -f_X\{g^{-1}(y)\}g^{-1\prime}(y)$$

with $g^{-1\prime}(y)$ always negative.

Therefore, for $Y = g(X)$ we have

$$f_Y(y) = f_X\{g^{-1}(y)\}|g^{-1\prime}(y)|. \tag{7.1}$$

**Example** Let $f_X(x) = e^{-x}$ for $x > 0$. Hence, $F_X(x) = \int_0^x f_X(u)du = 1 - e^{-x}$. Let $Y = g(X) = \log(X)$. Then

$$g^{-1}(y) = e^y \qquad \text{and} \qquad g^{-1\prime}(y) = e^y$$

Then, using (7.1), the pdf of $Y$ is

$$f_Y(y) = e^{-e^y}e^y \quad \text{for } y \in \mathbb{R}.$$

■

## 7.1 Mean, Variance and Quantiles

### 7.1.1 Expectation

**Definition 7.1.1.** *For a continuous random variable $X$ we define the* **mean** *or* **expectation** *of $X$,*

$$\mu_X \text{ or } E_X(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

**Extension:** More generally, for a (measurable) *function of interest* of the random variable $g : \mathbb{R} \to \mathbb{R}$ we have

$$\mathrm{E}_X\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Properties of Expectations

Clearly, for continuous random variables we again have linearity of expectation

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b, \qquad \forall a, b \in \mathbb{R},$$

and that for two functions $g, h : \mathbb{R} \to \mathbb{R}$, we have additivity of expectation

$$\mathrm{E}\{g(X) + h(X)\} = \mathrm{E}\{g(X)\} + \mathrm{E}\{h(X)\}.$$

### 7.1.2 Variance

**Definition 7.1.2.** *The* **variance** *of a continuous random variable $X$ is given by*

$$\sigma_X^2 \text{ or } Var_X(X) = E\{(X - \mu_X)^2\} = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx.$$

and again it is easy to show that

$$\mathrm{Var}_X(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \mu_X^2 = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2.$$

For a linear transformation $aX + b$ we again have

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X), \qquad \forall a, b \in \mathbb{R}.$$

### 7.1.3 Quantiles

Recall we defined the lower and upper quartiles and median of a sample of data as points (1/4,3/4,1/2)-way through the ordered sample. This idea can be generalised as follows:

**Definition 7.1.3.** *For a (continuous) random variable $X$ we define the $\alpha$-**quantile** $Q_X(\alpha)$, $0 \leq \alpha \leq 1$ to satisfy* $\mathrm{P}(X \leq Q_X(\alpha)) = \alpha$,

$$Q_X(\alpha) = F_X^{-1}(\alpha).$$

In particular the **median** of a random variable $X$ is $F_X^{-1}\left(\frac{1}{2}\right)$. That is, the solution to the equation $F_X(x) = \frac{1}{2}$.

**Example** Suppose we have a continuous random variable $X$ with probability density function given by

$$f_X(x) = \begin{cases} x^2/9, & 0 < x < 3 \\ 0, & \text{otherwise.} \end{cases}$$

*Questions*

Q1) Calculate $\mathrm{E}(X)$.

Q2) Calculate $\mathrm{Var}(X)$.

Q3) Calculate the median of $X$.

*Solutions*

S1)

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^3 x.\frac{x^2}{9} dx = \left.\frac{x^4}{36}\right|_0^3 = \frac{3^4}{36} = 2.25$$

S2)

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) = \int_0^3 x^2.\frac{x^2}{9} = \left.\frac{x^5}{45}\right|_0^3 = \frac{3^5}{45} = 5.4$$

So $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = 5.4 - 2.25^2 = 0.3375$

S3) From earlier, $F(x) = \frac{x^3}{27}$, for $0 < x < 3$.

Setting $F(x) = \frac{1}{2}$ and solving, we get

$$\frac{x^3}{27} = \frac{1}{2} \iff x = \sqrt[3]{\frac{27}{2}} = \frac{3}{\sqrt[3]{2}} \approx 2.3811$$

for the median.

■

## 7.2 Some Important Continuous Random Variables

### 7.2.1 Continuous Uniform Distribution

Suppose $X$ is a continuous random variable with probability density function

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

Then $X$ is said to follow a uniform distribution on the interval $(a, b)$ and we write $X \sim \mathrm{U}(a, b)$.

***Notes***

- The cdf is

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

- The case $a = 0$ and $b = 1$ is referred to as the **Standard uniform**.

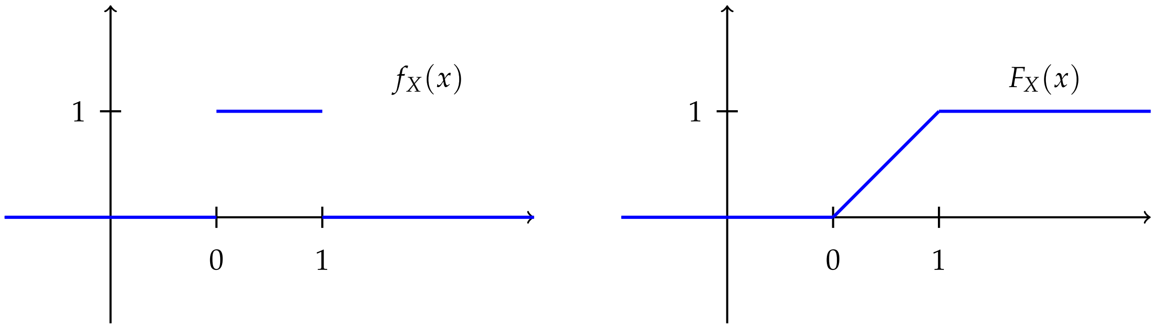



Figure 7.2: PDF and CDF of a standard uniform distribution.

- Suppose $X \sim \mathrm{U}(0, 1)$, so $F_X(x) = x$, $0 \leq x \leq 1$. We wish to map the interval $(0, 1)$ to the general interval $(a, b)$, where $a < b \in \mathbb{R}$. So we define a new random variable $Y = a + (b - a)X$, so $a < Y < b$.

  We first observe that for any $y \in (a, b)$,

$$Y \leq y \iff a + (b-a)X \leq y \iff X \leq \frac{y-a}{b-a}.$$

  From this we find $Y \sim \mathrm{U}(a, b)$, since

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}\left(X \leq \frac{y-a}{b-a}\right) = F_X\left(\frac{y-a}{b-a}\right) = \frac{y-a}{b-a}.$$

- To find the mean of $X \sim \mathrm{U}(a, b)$,

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x.\frac{1}{b-a} dx = \left[\frac{x^2}{2(b-a)}\right]_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

Similarly we get $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{(b-a)^2}{12}$, so

$$\mu = \frac{a+b}{2}, \qquad \sigma^2 = \frac{(b-a)^2}{12}.$$

### 7.2.2 Exponential Distribution

Suppose now $X$ is a random variable taking value on $\mathbb{R}^+ = [0, \infty)$ with pdf

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0,$$

for some $\lambda > 0$.

Then $X$ is said to follow an exponential distribution with *rate* parameter $\lambda$ and we write $X \sim \mathrm{Exp}(\lambda)$.

***Notes***

- The cdf is
$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0.$$
- An alternative representation uses $\theta = 1/\lambda$ as the parameter of the distribution. This is sometimes uses because the expectation and variance of the Exponential distributions are
$$\mathrm{E}(X) = \frac{1}{\lambda} = \theta, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$
- If $X \sim \mathrm{Exp}(\lambda)$, then, for all $x, t > 0$,
$$\mathrm{P}(X > x+t | X > t) = \frac{\mathrm{P}(X > x+t \cap X > t)}{\mathrm{P}(X > t)} = \frac{\mathrm{P}(X > x+t)}{\mathrm{P}(X > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathrm{P}(X > x).$$
Thus, for all $x, t > 0$, $\mathrm{P}(X > x+t | X > t) = \mathrm{P}(X > x)$ — this is known as the **Lack of Memory Property**, and is unique to the exponential distribution amongst continuous distributions.

  *Interpretation:* So if we think of the exponential variable as the time to an event, then knowledge that we have waited time $s$ for the event tells us nothing about how much longer we will have to wait – the process has *no memory*.
- Exponential random variables are often used to model the time until occurrence of a random event where there is an assumed constant risk ($\lambda$) of the event happening over time, and so are frequently used as a simplest model, for example, in reliability analysis. So examples include:
  - the time to failure of a component in a system;
  - the time until we find the next mistake on my slides;
  - the distance we travel along a road until we find the next pothole;

– the time until the next jobs arrives at a database server;

Notice the duality between some of the exponential random variable examples and those we saw for a Poisson distribution. In each case, "number of events" has been replaced with "time between events".

*Claim:*

If events in a random process occur according to a Poisson distribution with rate $\lambda$ then the time between events has an Exponential distribution with rate parameter $\lambda$.

*Proof.* Suppose we have some random event process such that $\forall x > 0$, the number of events occurring in $[0, x]$, $N_x$, follows a Poisson distribution with rate parameter $\lambda$, so $N_x \sim \text{Poisson}(\lambda x)$. Such a process is known as an *homogeneous Poisson process*. Let $X$ be the time until the first event of this process arrives.

Then we notice that

$$\begin{aligned} P(X > x) &\equiv P(N_x = 0) \\ &= \frac{(\lambda x)^0 e^{-\lambda x}}{0!} \\ &= e^{-\lambda x}. \end{aligned}$$

and hence $X \sim \text{Exp}(\lambda)$. The same argument applies for all subsequent inter-arrival times. □

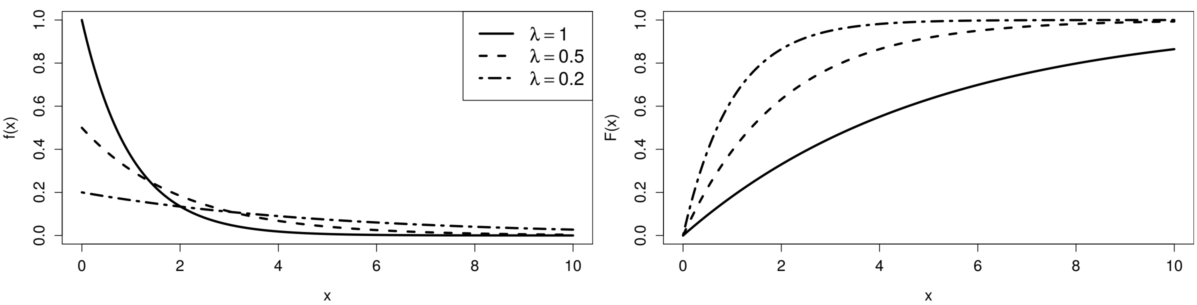



Figure 7.3: PDFs and CDFs for Exponential distribution with different rate parameters.

### 7.2.3 Normal Distribution

Suppose $X$ is a random variable taking value on $\mathbb{R}$ with pdf

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

for some $\mu \in \mathbb{R}, \sigma > 0$. Then $X$ is said to follow a Gaussian or normal distribution with mean $\mu$ and variance $\sigma^2$, and we write $X \sim N(\mu, \sigma^2)$.

***Notes***

- The cdf of $X \sim \mathrm{N}(\mu, \sigma^2)$ is not analytically tractable for any $(\mu, \sigma)$, so we can only write

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt.$$

- Special Case: If $\mu = 0$ and $\sigma^2 = 1$, then $X$ has a **standard** or **unit** normal distribution. The pdf of the standard normal distribution is written as $\phi(x)$ and simplifies to

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

  Also, the cdf of the standard normal distribution is written as $\Phi(x)$. Again, for the cdf, we can only write

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

- If $X \sim N(0, 1)$, and

$$Y = \sigma X + \mu$$

  then $Y \sim N(\mu, \sigma^2)$. Re-expressing this result: if $X \sim N(\mu, \sigma^2)$ and $Y = (X - \mu)/\sigma$, then $Y \sim N(0, 1)$. This is an important result as it allows us to write the cdf of any normal distribution in terms of $\Phi$: If $X \sim \mathrm{N}(\mu, \sigma^2)$ and we set $Y = \dfrac{X-\mu}{\sigma}$, then since $\sigma > 0$ we can first observe that for any $x \in \mathbb{R}$,

$$\begin{aligned} X \le x &\iff \frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \\ &\iff Y \le \frac{x-\mu}{\sigma}. \end{aligned}$$

  Therefore we can write the cdf of $X$ in terms of $\Phi$,

$$\begin{aligned} F_X(x) = \mathrm{P}(X \le x) &= \mathrm{P}\left(Y \le \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

- Since the cdf, and therefore any probabilities, associated with a normal distribution are not analytically available, numerical integration procedures are used to find approximate probabilities. In particular, statistical tables contain values of the standard normal cdf $\Phi(z)$ for a range of values $z \in \mathbb{R}$, and the quantiles $\Phi^{-1}(\alpha)$ for a range of values $\alpha \in (0, 1)$. Linear interpolation is used for approximation between the tabulated values. As seen in the point above, all normal distribution probabilities can be related back to probabilities from a standard normal distribution.

- Table 7.1 is an example of a statistical table for the standard normal distribution.

| $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ |
|---|---|---|---|---|---|---|---|
| 0 | 0.5 | 0.9 | 0.816 | 1.8 | 0.964 | 2.8 | 0.997 |
| 0.1 | 0.540 | 1.0 | 0.841 | 1.9 | 0.971 | 3.0 | 0.998 |
| 0.2 | 0.579 | 1.1 | 0.864 | 2.0 | 0.977 | 3.5 | 0.9998 |
| 0.3 | 0.618 | 1.2 | 0.885 | 2.1 | 0.982 | 1.282 | 0.9 |
| 0.4 | 0.655 | 1.3 | 0.903 | 2.2 | 0.986 | 1.645 | 0.95 |
| 0.5 | 0.691 | 1.4 | 0.919 | 2.3 | 0.989 | 1.96 | 0.975 |
| 0.6 | 0.726 | 1.5 | 0.933 | 2.4 | 0.992 | 2.326 | 0.99 |
| 0.7 | 0.758 | 1.6 | 0.945 | 2.5 | 0.994 | 2.576 | 0.995 |
| 0.8 | 0.788 | 1.7 | 0.955 | 2.6 | 0.995 | 3.09 | 0.999 |

Table 7.1

Notice that $\Phi(z)$ has been tabulated for $z > 0$.

This is because the standard normal pdf $\phi$ is *symmetric* about 0, that is, $\phi(-z) = \phi(z)$. For the cdf $\Phi$, this means

$$\Phi(z) = 1 - \Phi(-z).$$

So, for example, $\Phi(-1.2) = 1 - \Phi(1.2) \approx 1 - 0.885 = 0.115$.

Similarly, if $Z \sim \mathrm{N}(0, 1)$ and we want, for example, $\mathrm{P}(Z > 1.5) = 1 - \mathrm{P}(Z \leq 1.5) = 1 - \Phi(1.5)(= \Phi(-1.5))$.

So more generally we have

$$\mathrm{P}(Z > z) = \Phi(-z).$$

We will often have cause to use the 97.5% and 99.5% quantiles of $\mathrm{N}(0, 1)$, given by $\Phi^{-1}(0.975)$ and $\Phi^{-1}(0.995)$.

$$\Phi(1.96) \approx 97.5\%.$$

So with 95% probability an $\mathrm{N}(0, 1)$ random variable will lie in $[-1.96, 1.96]$ $(\approx [-2, 2])$.

$$\Phi(2.58) = 99.5\%.$$

So with 99% probability an $\mathrm{N}(0, 1)$ random variable will lie in $[-2.58, 2.58]$.

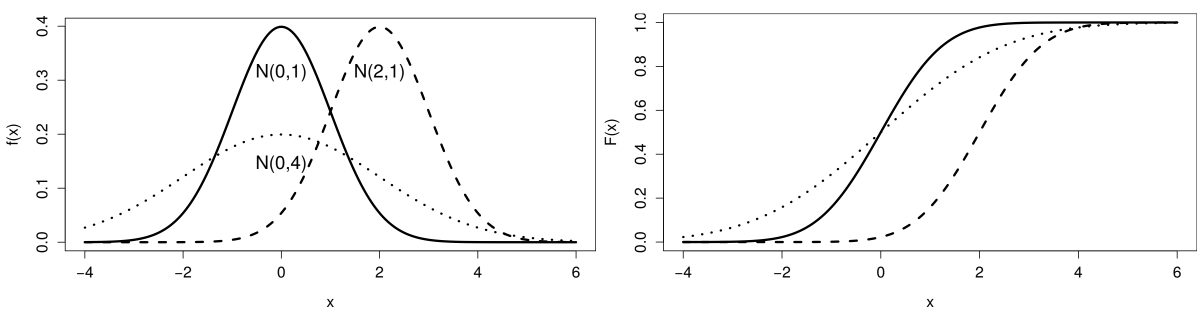



Figure 7.4: PDFs and CDFs of normal distributions with different means and variances.

**Example** An analogue signal received at a detector (measured in microvolts) may be modelled as a Gaussian random variable $X \sim \mathrm{N}(200, 256)$.

*Questions*

Q1) What is the probability that the signal will exceed $240\mu V$?

Q2) What is the probability that the signal is larger than $240\mu V$ given that it is greater than $210\mu V$?

*Solutions*

S1)

$$\mathrm{P}(X > 240) = 1 - \mathrm{P}(X \leq 240) = 1 - \Phi\left(\frac{240 - 200}{\sqrt{256}}\right) = 1 - \Phi(2.5) \approx 0.00621.$$

S2)

$$\mathrm{P}(X > 240 | X > 210) = \frac{\mathrm{P}(X > 240)}{\mathrm{P}(X > 210)} = \frac{1 - \Phi\left(\frac{240-200}{\sqrt{256}}\right)}{1 - \Phi\left(\frac{210-200}{\sqrt{256}}\right)} \approx 0.02335.$$

■

Let $X_1, X_2, \ldots, X_n$ be $n$ independent and identically distributed (i.i.d.) random variables from **any** probability distribution, each with mean $\mu$ and variance $\sigma^2$.

From before we know

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = n\mu, \quad \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2.$$

First notice

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i - n\mu\right) = 0, \quad \mathrm{Var}\left(\sum_{i=1}^{n} X_i - n\mu\right) = n\sigma^2.$$

Dividing by $\sqrt{n}\sigma$, we obtain

$$\mathrm{E}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\right) = 0, \quad \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}\right) = 1.$$

**Theorem 7.3** (Central Limit Theorem or CLT)**.**

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim \Phi.$$

This can also be written as

$$\lim_{n\to\infty} \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \Phi, \quad \text{where} \quad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Or finally, for large $n$ we have approximately

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

or

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right).$$

We note here that although all these approximate distributional results hold irrespective of the distribution of the $\{X_i\}$, in the special case where $X_i \sim \mathrm{N}(\mu, \sigma^2)$ these distributional results are, in fact, exact. This is because the sum of independent normally distributed random variables is also normally distributed.

**Example** Consider the most simple example, that $X_1, X_2, \ldots$ are i.i.d. Bernoulli$(p)$ discrete random variables taking value 0 or 1.

Then the $\{X_i\}$ have mean $\mu = p$ and variance $\sigma^2 = p(1-p)$. Then, by definition, we know that for any $n$ we have

$$\sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p).$$

which has mean $np$ and variance $np(1-p)$.

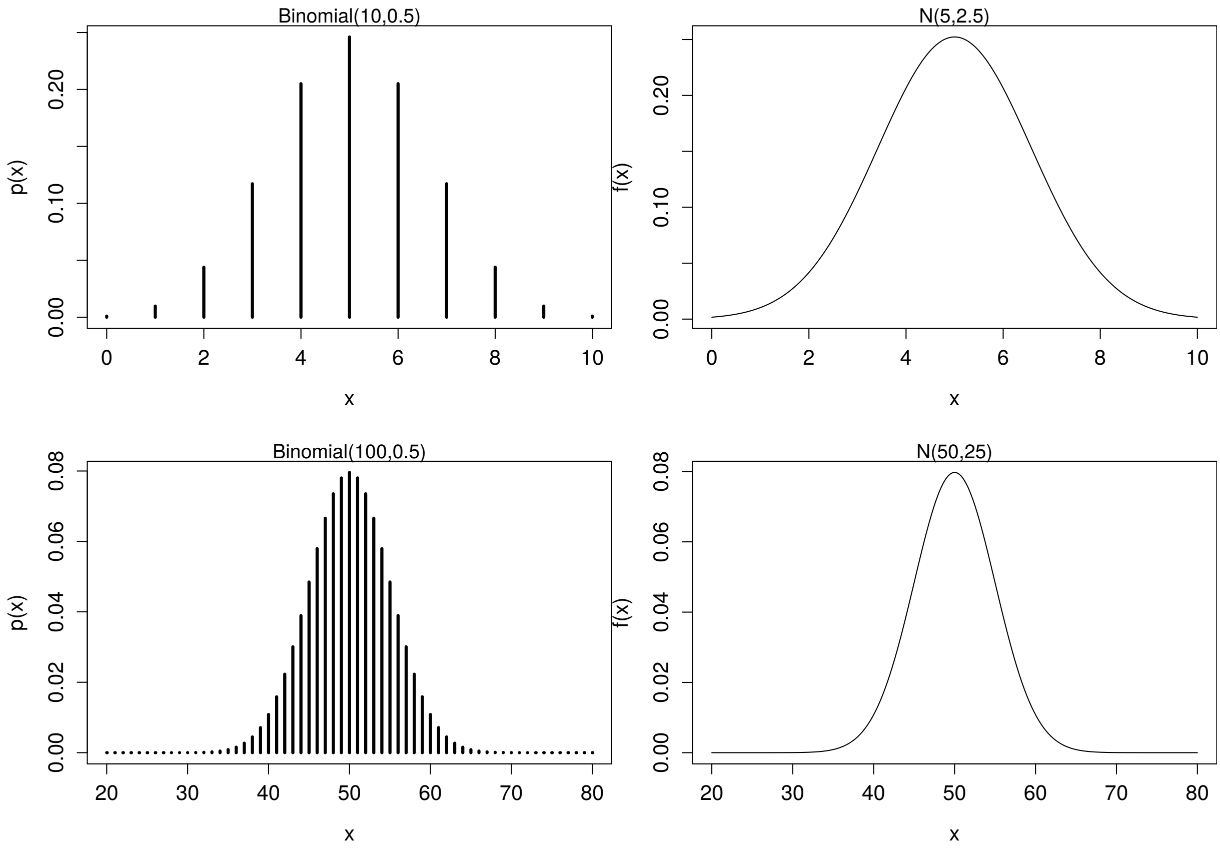
But now, by the Central Limit Theorem (CLT), we also have for large $n$ that approximately:

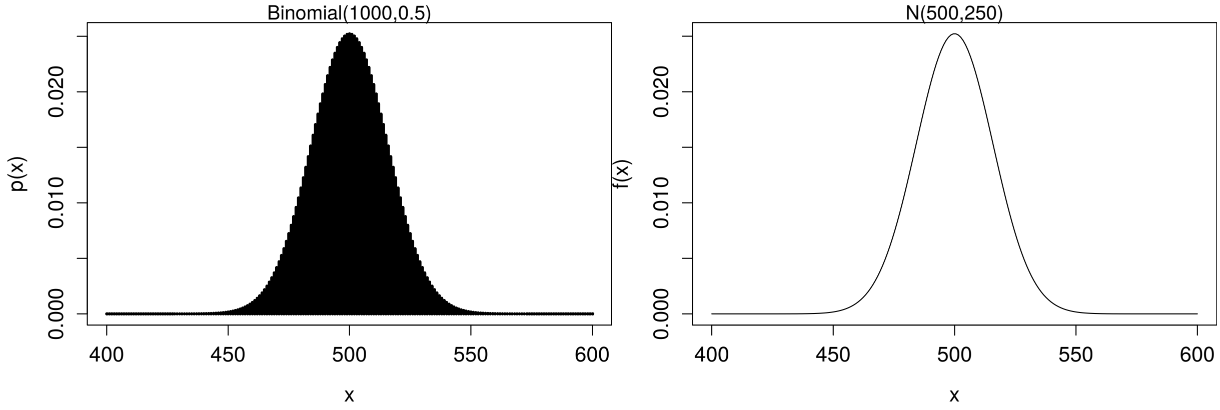
$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right) \equiv \mathrm{N}(np, np(1-p)).$$

So for large n

$$\text{Binomial}(n, p) \approx \mathrm{N}(np, np(1-p)).$$

Notice that the LHS is a discrete distribution, and the RHS is a continuous distribution.






■

**Example** Suppose $X$ was the number of heads found on 1000 tosses of a fair coin, and we were interested in $\mathrm{P}(X \leq 490)$.

Using the binomial distribution pmf, we would need to calculate

$$\mathrm{P}(X \leq 490) = p_X(0) + p_X(1) + p_X(2) + \ldots + p_X(490) (\approx 0.27).$$

However, using the CLT we have approximately $X \sim \mathrm{N}(500, 250)$ and so

$$\mathrm{P}(X \leq 490) \approx \Phi\left(\frac{490 - 500}{\sqrt{250}}\right) = \Phi(-0.632) = 1 - \Phi(0.632) \approx 0.26.$$

■

**Example** Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$, and consider the transformation $Y = e^X$.

Then if $g(x) = e^x$, $g^{-1}(y) = \log(y)$ and $g^{-1\prime}(y) = \dfrac{1}{y}$.

Then by (7.1) we have

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left[-\frac{\{\log(y) - \mu\}^2}{2\sigma^2}\right], \quad y > 0,$$

and we say $Y$ follows a **log-normal** distribution. ■