Dr Isabella Gollini

Probability and Statistics

6 Continuous Random Variables

6.1 Probability density functions

In the previous sections we have considered discrete r.v.s, which can take only a finite or countable set of values, usually a subset of the non-negative integers. The probability distributions of such r.v.s are specified by sequences (p_r) $(r = 0, 1, 2, \ldots)$. We now consider *continuous* r.v.s, which can take values in a continuous range, such as measurements of height, weight, temperature.

Example

Suppose that we have taken a random sample of size 10,000 from the adult males of a large human population and measured their heights, which could in principle be done to any degree of accuracy, for example, to the nearest inch, the nearest centimetre, the nearest tenth of an inch, etc. We may then construct a histogram using some appropriate class interval. Figure 1 shows such a histogram using simulated data and a class interval of 1 inch. We could readily sketch a smooth curve to represent the shape of the histogram.

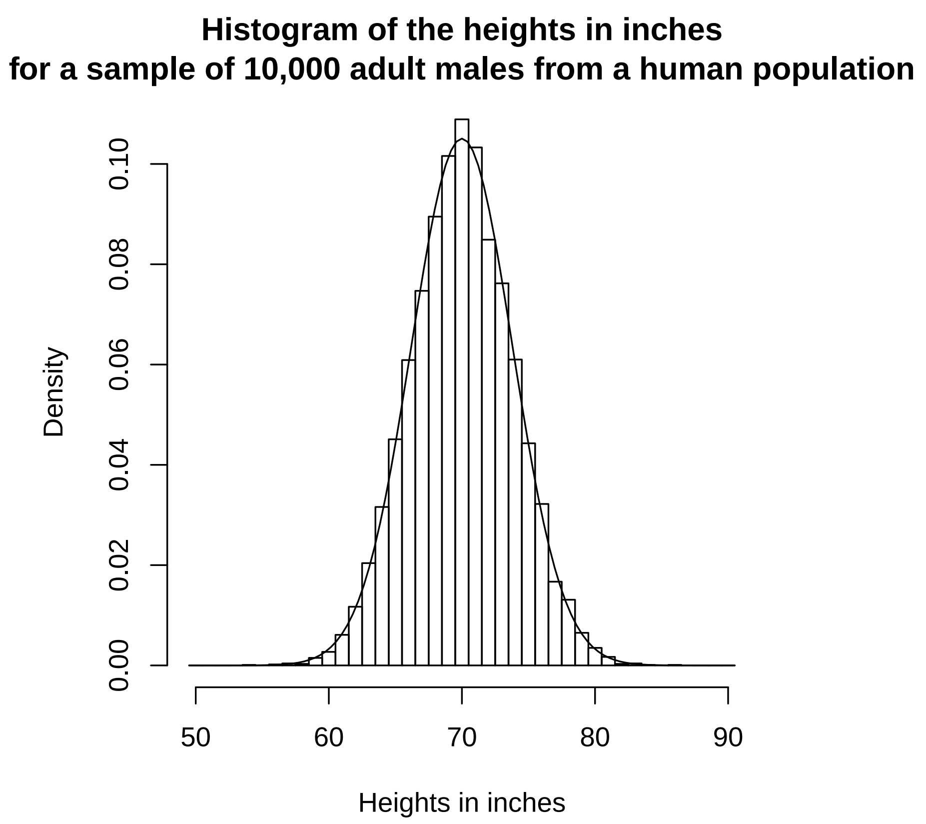

Figure 1: Example of a histogram of human heights (simulated data)

Let the r.v. X denote the height of a randomly chosen individual. Suppose that we let the size n of the random sample of heights $x_1, x_2, \ldots, x_n$ get larger and larger. Suitably normalizing the vertical scale of the histogram so as to keep the total area equal to 1, and making the class intervals smaller and smaller, as $n \to \infty$ the shape of the histogram approaches more and more closely the shape of what is known as the *probability density function* of X.

More generally, we can think of the probability density function of a continuous r.v. X as the limiting shape of the histogram as we take more and more observations. To put it more formally, the probability distribution of a continuous r.v. X is specified by a *probability density function* (p.d.f.) $f(x)$ $(-\infty < x < \infty)$, where $f(x) \geq 0$ $(-\infty < x < \infty)$ and

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

(This corresponds to the condition $\sum_{r=0}^{\infty} p_r = 1$ for a discrete distribution.)

For any constants a and b with $a < b$,

$$\Pr(a < X < b) = \int_a^b f(x)dx.$$

Thus the probability that the continuous r.v. X takes a value in the interval (a, b) is given by the area under the p.d.f between a and b. Note that if X is a continuous r.v. then

- for any constant a, $\Pr(X = a) = 0$,
- for any constants a and b with $a < b$, $\Pr(a < X < b) = \Pr(a \leq X \leq b)$.

Figure 2 gives an example of a p.d.f. of a *positive* continuous r.v., i.e., a r.v. X which can take only positive values, so that $f(x) = 0$ $(x < 0)$. The shaded area under the curve represents $\Pr(5 < X < 10)$, i.e., the case $a = 5, b = 10$.

The *(cumulative) distribution function* (c.d.f.) $F(x)$ is given by

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)du \qquad (-\infty < x < \infty). \tag{1}$$

Note that

- We use the upper case [capital] letter X to denote a r.v. and the lower case [small] letter x to denote a particular value that X might take.
- $0 \leq F(x) \leq 1$ $(-\infty < x < \infty)$.
- $F(x)$ is a non-decreasing function of x such that $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$.
- For $a < b$ we may write $\Pr(a < X < b) = F(b) - F(a)$.
- For a <u>positive</u> continuous r.v., $F(x) = 0$ $(x \leq 0)$ and $F(x) = \int_0^x f(u)du$ $(x \geq 0)$.

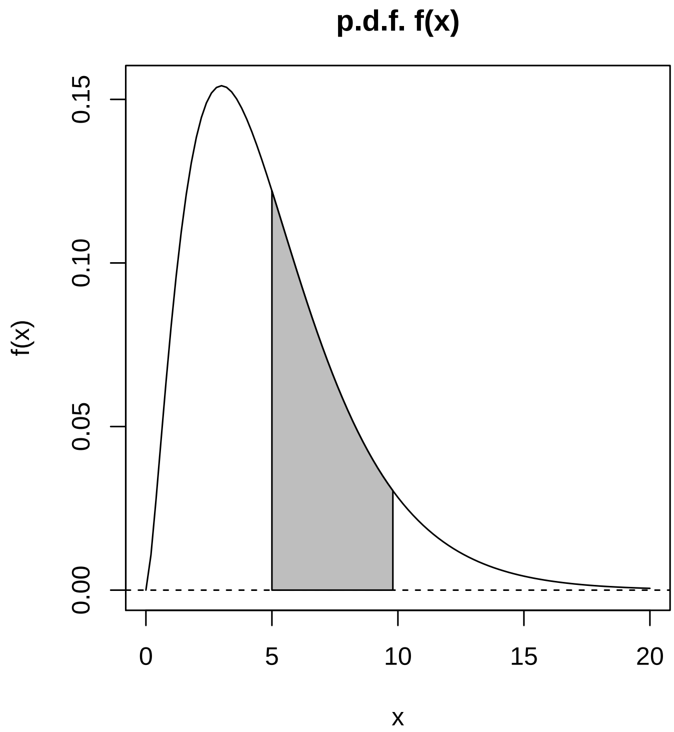

Figure 2: Shaded area: $\Pr(5 < X < 10)$

Differentiating Equation (1) we find the inverse relationship between the p.d.f. and c.d.f.,

$$f(x) = F'(x) \qquad (-\infty < x < \infty). \tag{2}$$

As for discrete r.v.s, we can define the mean, the expected value of a function $g(X)$, and the variance of a continuous r.v. X, where in the definitions the summation that was used in the discrete case is replaced by an integral. Thus

Definition

The *mean* μ or *expected value* or *expectation* $E(X)$ of the r.v. X, a measure of location, is given by

$$\mu \equiv E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{3}$$

The expected value of $g(X)$ is given by

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx. \tag{4}$$

Definition

The *variance* σ^2 or $\mathrm{var}(X)$ of the r.v. X, a measure of spread, is given by

$$\sigma^2 \equiv \mathrm{var}(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx. \tag{5}$$

By an argument similar to the one used in the discrete case, an alternative formula for the variance is given by

$$\sigma^2 \equiv \mathrm{var}(X) = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2. \tag{6}$$

- We may also refer to μ and σ^2 as the mean and variance, respectively, of the continuous probability distribution with p.d.f. $f(x)$ $(-\infty < x < \infty)$.
- σ, the square root of the variance, is the *standard deviation* of the r.v. and its distribution.

6.2 Example — the exponential distributions

Lemma 1 *For any integer n $(n \geq 0)$,*

$$\int_0^\infty x^n e^{-x} dx = n!$$

Proof — by induction on n. For all integers $n \geq 0$ define

$$I_n = \int_0^\infty x^n e^{-x} dx.$$

Note first that

$$I_0 = \int_0^\infty e^{-x} dx = [-e^{-x}]_0^\infty = 1 = 0!$$

For any $n \geq 1$, suppose that the result holds for $n-1$, i.e., $I_{n-1} = (n-1)!$ Integrating by parts,

$$\begin{aligned} I_n \equiv \int_0^\infty x^n e^{-x} dx &= [-x^n e^{-x}]_0^\infty + n \int_0^\infty x^{n-1} e^{-x} dx \\ &= 0 + n(n-1)! = n! \end{aligned}$$

Corollary 2 *For any integer n $(n \geq 0)$ and any $\lambda > 0$,*

$$\int_0^\infty x^n \lambda e^{-\lambda x} dx = \frac{n!}{\lambda^n} \ .$$

Proof. Making the substitution $y = \lambda x$, so that $dy = \lambda dx$,

$$\int_0^\infty x^n \lambda e^{-\lambda x} dx = \int_0^\infty \frac{y^n}{\lambda^n} e^{-y} dy = \frac{n!}{\lambda^n} \ ,$$

using the result of Lemma 1.

A positive continuous r.v. X is said to have an *exponential distribution* with parameter λ, where $\lambda > 0$, if it has p.d.f. $f(x)$ given by

$$f(x) = \lambda e^{-\lambda x} \qquad (x \geq 0).$$

Using the result of Corollary 2 with $n = 0$, we see that

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda e^{-\lambda x} dx = 1,$$

so that $f(x)$ as defined really is a p.d.f. Applying Equation (1), the corresponding c.d.f. $F(x)$ is given by

$$F(x) = \begin{cases} 0 & (x < 0) \\ \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x} & (x \geq 0). \end{cases}$$

Both the p.d.f. and c.d.f. are illustrated in Figure 3 for the cases $\lambda = 1$, $\lambda = 2$, $\lambda = 0.3$. Applying Equation (3) and using the result of Corollary 2 with $n = 1$, the mean μ of the

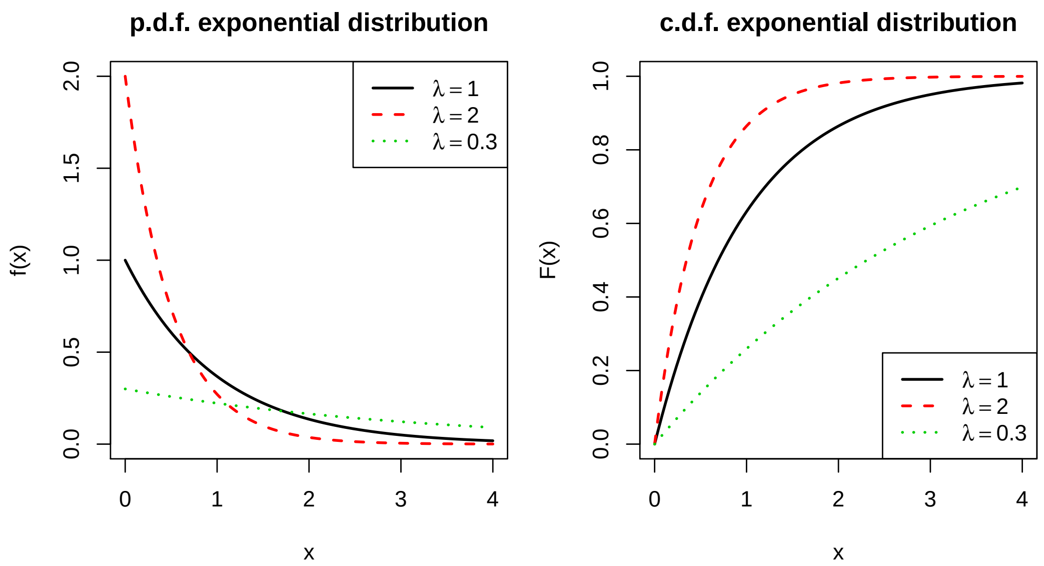

Figure 3: p.d.f. $f(x)$ (left) and c.d.f. F(x) (right) for an exponential distributions with parameter $\lambda = 1$, $\lambda = 2$, $\lambda = 0.3$

exponential distribution is given by

$$\mu = \int_{0}^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}\ .$$

Applying Equation (6) and using the result of Corollary 2 with $n = 2$, the variance σ^2 is given by

$$\sigma^2 = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x} dx - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\ .$$

So, for the exponential distribution with parameter λ,

$$\mu = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2}\ .$$

Furthermore, using the result of Corollary 2, what is known as the *nth moment*, $E(X^n)$, is given by

$$E(X^n) = \int_{0}^{\infty} x^n \lambda e^{-\lambda x} dx = \frac{n!}{\lambda^n}\ .$$

6.3 The normal distributions

The normal distributions play a central role in statistical theory and practice. Many naturally occurring variables, as with the example of human heights in Section 6.1, have normal distributions. We begin with the *standard normal distribution*, which has p.d.f. $f(x)$ given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad (-\infty < x < \infty) \tag{7}$$

and is denoted by $N(0,1)$. For a r.v. X that has this distribution, we write $X \sim N(0,1)$. The p.d.f., which is often described as "bell-shaped", is illustrated in Figure 4. The standard normal distribution is symmetric about the vertical axis $x = 0$ and has mean 0 and variance 1.

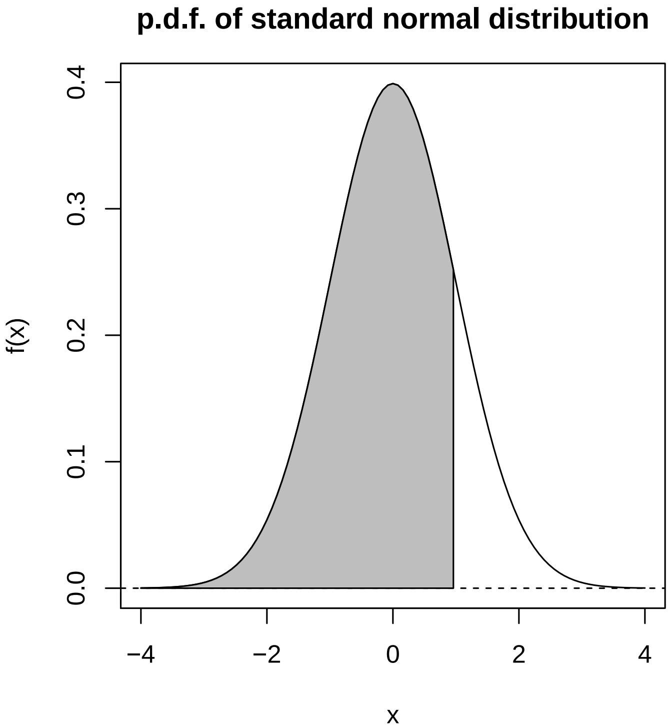

Figure 4: The shaded area is $\Phi(1) \equiv \Pr(X \leq 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1} \exp\left(-\frac{1}{2}u^2\right) du$

We do not have an explicit formula for the indefinite integral $\int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx$, but it does turn out to be the case that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = 1 \ ,$$

so that $f(x)$ as defined in Equation (7) really does represent a p.d.f.

The fact that $E(X) = 0$ follows from the symmetry about $x = 0$ of the standard normal p.d.f. However, we may check this directly:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \left[-\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)\right]_{-\infty}^{\infty} \;=\; 0. \end{aligned}$$

Because of its special role in statistics, we use a special notation, $\Phi(x)$, for the c.d.f. of the standard normal distribution:

$$\Phi(x) \equiv \Pr(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}u^2\right) du \qquad (-\infty < x < \infty).$$

Because there is no explicit formula for it, $\Phi(x)$ has to be computed numerically for particular values of x. Table 4 in the *New Cambridge Statistical Tables* gives values of $\Phi(x)$ for $x = 0$ to $x = 3.3$ in steps of 0.01. Because of the symmetry of the standard normal distribution, to evaluate $\Phi(x)$ for negative values of x we may use the formula

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

Example

1. If $X \sim N(0,1)$, what is $\Pr(-1.13 < X < 1.75)$?

$$\begin{aligned} \Pr(-1.13 < X < 1.75) &= \Phi(1.75) - \Phi(-1.13) \\ &= \Phi(1.75) - (1 - \Phi(1.13)) \\ &= \Phi(1.75) + \Phi(1.13) - 1 \\ &= 0.9599 + 0.8708 - 1 \;=\; 0.8307 \\ &= 0.831 \text{ to 3 d.p.} \end{aligned}$$

2. If $X \sim N(0,1)$, what is $\Pr(-1.96 < X < 1.96)$?

$$\begin{aligned} \Pr(-1.96 < X < 1.96) &= \Phi(1.96) - \Phi(-1.96) \\ &= \Phi(1.96) - (1 - \Phi(1.96)) \\ &= 2\Phi(1.96) - 1 \\ &= (2 \times 0.9750) - 1 \;=\; 0.9500 \\ &= 0.950 \text{ to 3 d.p.} \end{aligned}$$

So for a r.v. with a standard normal distribution the probability that it lies in the interval $(-1.96, 1.96)$ is 0.95.

More generally we shall be interested in a normal distribution with mean μ and variance σ^2, a $N(\mu, \sigma^2)$ distribution, which has p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (-\infty < x < \infty).$$

Thus the family of normal distributions, the $N(\mu, \sigma^2)$ distributions, has two parameters, μ and σ, where $\sigma > 0$. (We sometimes refer to the standard deviation σ as the parameter and sometimes to the variance σ^2 as the parameter.) μ is a *location parameter*: it specifies the value on which the distribution is centred. σ is a *scale parameter*: it specifies how spread out the distribution is (see Figure 5).

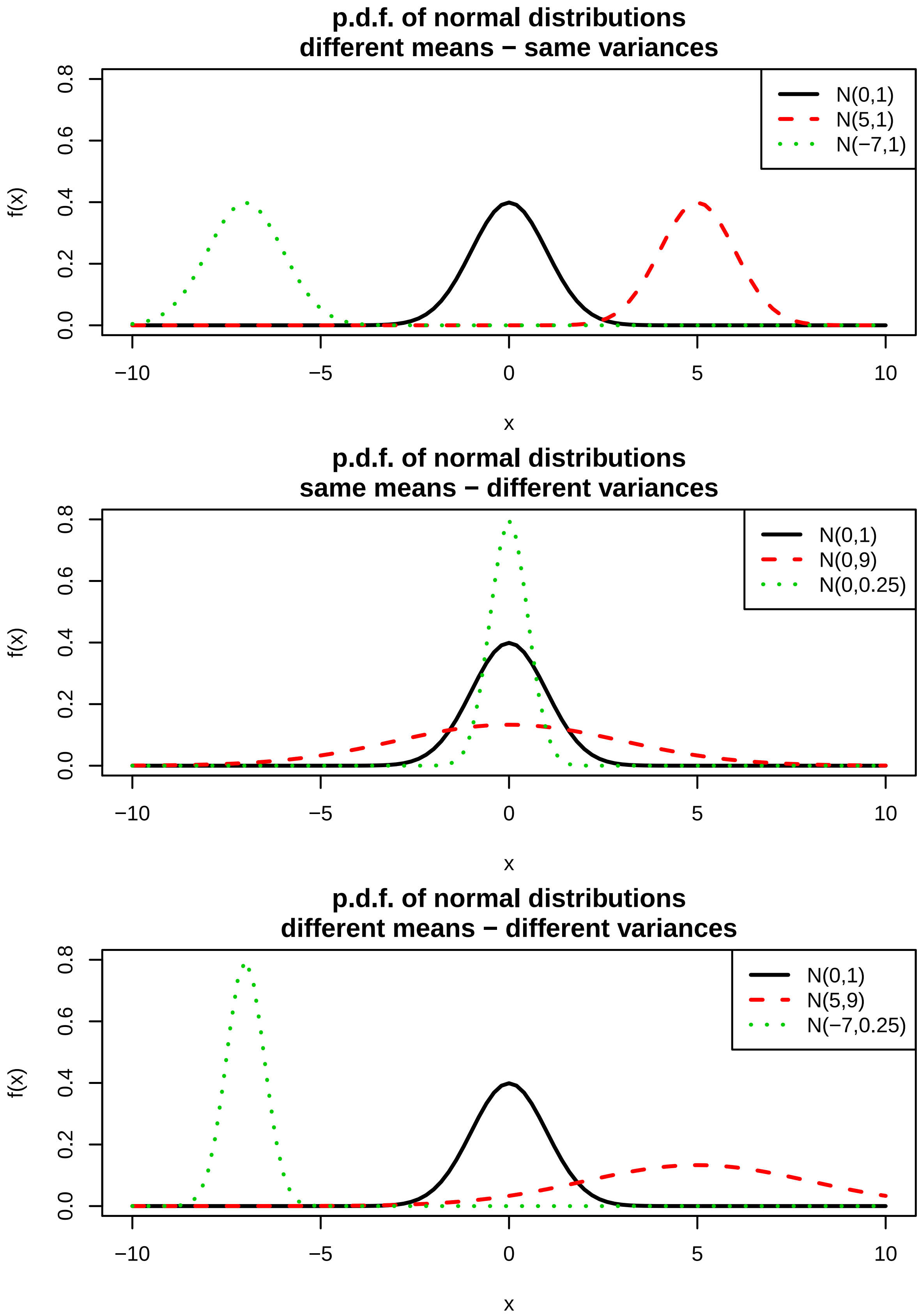

Figure 5: Comparison of p.d.f. of normal distributions

It is a property of the family of normal distributions that if $X \sim N(\mu, \sigma^2)$ then

$$Z \equiv \frac{X - \mu}{\sigma} \tag{8}$$

has the standard normal distribution, i.e., $Z \sim N(0, 1)$.

If x is a value from a $N(\mu, \sigma^2)$, then the z-score of x is given by:

$$z = \frac{x - \mu}{\sigma} \tag{9}$$

This is an important property in that we can use the standardization of Equation (8) in order to calculate probabilities for arbitrary normal distributions from the standard normal distribution tables. If $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$:

$$\begin{aligned} \Pr(X \leq x) &= \Pr\left(Z \leq \frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right) . \end{aligned}$$

Example

Let the r.v. X denote the height in inches of an adult male drawn at random from some population, where X has a normal distribution with mean 70 and standard deviation 4, i.e., $X \sim N(70, 4^2)$. It follows that the standardized r.v. is

$$Z \equiv \frac{X - 70}{4} \sim N(0, 1) .$$

The probability that the height of a randomly selected individual exceeds 72 inches is given by

$$\begin{aligned} \Pr(X > 72) &= \Pr\left(\frac{X - 70}{4} > \frac{72 - 70}{4}\right) \\ &= \Pr(Z > 0.5) \\ &= 1 - \Phi(0.5) \\ &= 1 - 0.6915 \;=\; 0.3085 . \end{aligned}$$

The probability that the height of a randomly selected individual lies between 60 and 66 inches is given by

$$\begin{aligned} \Pr(60 < X < 66) &= \Pr\left(\frac{60 - 70}{4} < \frac{X - 70}{4} < \frac{66 - 70}{4}\right) \\ &= \Pr(-2.5 < Z < -1) \\ &= \Phi(-1) - \Phi(-2.5) \;=\; [1 - \Phi(1)] - [1 - \Phi(2.5)] \\ &= \Phi(2.5) - \Phi(1) \;=\; 0.99379 - 0.8413 \;=\; 0.1525 . \end{aligned}$$

To calculate normal probabilities using `R`, you can use the function `pnorm` to get the cumulative distribution function or `dnorm` to calculated terms from the probability density function. The values you have to specify in the function are (in this order) the value (x), the mean (μ), and the standard deviation (σ). Type `?pnorm` to learn more about this function.

Using `R` (or the `NORMDIST` function in Excel), the process of standardization is not required. For example, the probability calculations in the present case are carried out in the `R` session below.

$X \sim N(70, 4)$

```
mu <- 70
sigma <- 4
```

$\Pr(X > 72)$:

```
1 - pnorm(72, mu, sigma)

## [1] 0.3085375
```

$\Pr(60 < X < 66)$:

```
cdf66 <-  pnorm(66, mu, sigma)
cdf60 <-  pnorm(60, mu, sigma)

cdf66 - cdf60

## [1] 0.1524456
```

If the `pnorm` command is used without specifying the mean and the standard deviation, `R` assumes that the c.d.f. Φ of the standard normal distribution is intended. So, for example:

```
pnorm(1.75)

## [1] 0.9599408
```

Extra Examples

Example – the exponential distribution

The lifetime of a certain electronic component is known to be exponentially distributed with a mean lifetime of 100 hours. What proportion of such components will fail before 50 hours?

Example – the uniform distributions

a) Let X be a random variable with p.d.f.:

$$f(x) = \begin{cases} C & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

(i) Find the value of C.
In general the constant C is called normalising constant and is needed to normalise the density function to have total probability $= 1$.

(ii) Write and do the graph of the p.d.f. $f(x)$.

(iii) Write and do the graph of the c.d.f. $F(x)$.

(iv) Find the mean of the distribution.

(v) Find the variance of the distribution.

This distribution is called uniform distribution in the interval (a, b). And can be written as $X \sim U(a, b)$.

b) Let $X \sim U(2, 5)$:

(i) Find the mean and the variance of the distribution.

(ii) What is $\Pr(X > 3)$?

(iii) What is $\Pr(X < 6)$?

Example – IQ test

IQ test are constructed to have the scores that follow approximately a normal distribution with mean $= 100$ and standard deviation $= 15$.

(i) Identify the random variable and the distribution using the correct notation.

(ii) Find the probability that a randomly chosen person is classified as "very superior" (get a score of more than 130).

(iii) Find the probability that a randomly chosen person is classified as "average" (get a score between 90 and 110).

(iv) Find the probability that a randomly chosen person has a score of exactly 110.

(v) Find the probability that a randomly chosen person has a score of approximately 110 with an approximation to the nearest integer value.