
Topic 3

Probability Distributions

Contents

3.1	Discrete Random Variables	2
3.1.1	Random Variables and Expected Value	2
3.1.2	The Binomial Probability Distribution	3
3.1.3	The Geometric Probability Distribution	8
3.1.4	The Hypergeometric Distribution	11
3.1.5	The Poisson Distribution	14
3.2	Continuous Random Variables	19
3.2.1	The Uniform Probability Distribution	24
3.2.2	The Gamma Distribution	25
3.2.3	The Chi-Square Distribution	27
3.2.4	The Exponential Probability Distribution	27
3.3	The Normal Distribution	27
3.3.1	Everyday examples of Normal Distribution	27
3.3.2	Drawing the Curve	29
3.3.3	Calculations Using Normal Distribution	30
3.3.4	Properties of Normal Distribution Curves	32
3.3.5	Link with Probability	32
3.3.6	Upper and Lower Bounds	37
3.4	Summary	40

3.1 Discrete Random Variables

Many problems in statistics can be solved by classifying them into particular types. For example, in quality control the probability of finding faulty goods is an important issue. This is a standard situation where we are dealing with success or failure and so there are tried and trusted approaches to tackling a problem like this (in fact it can be dealt with by using the Binomial Distribution). We will consider many different probability distributions, some relevant to discrete random variables and others using the continuous type.

3.1.1 Random Variables and Expected Value

A *random variable* is a function taking numerical values which is defined over a sample space. Such a random variable is called *discrete* if it only takes countably many values.

Example A quality control engineer checks randomly the content of bags, each of which contains 100 resistors. He selects 2 resistors and measures whether they match the specification (exact value plus or minus 10% tolerance).

The number of resistors not matching the specification is a discrete random variable. Another random variable would be the function taking values 0 and 1, for the outcomes that there are faulty resistors in the bag, or not.

The *probability distribution* of a random variable is a table, graph, or formula that gives the probability $p(x)$ for each possible value of the random variable x . The requirements are that

- $0 \leq p(x) \leq 1$, i.e., the probability must lie between 0 and 1, and that
- $\sum_{all} p(x) = 1$.

For a discrete random variable x with probability distribution $p(x)$ the *expected value* (or *mean*) is defined as

$$\mu = E(x) = \sum_{all\ x} x \cdot p(x)$$

Example We consider the random variable which shows the outcome of rolling a die.

Each possible outcome between 1 and 6 is equally likely, so $p(x) = 1/6$ for $x = 1, \dots, 6$. For the expected value we calculate

$$\begin{aligned} E(x) &= \sum_{all\ x} x \cdot p(x) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} \\ &= \frac{21}{6} \\ &= 3.5 \end{aligned}$$

Let x be any discrete random variable with probability distribution $p(x)$, and let g be any function of x . The *expected value of $g(x)$* is defined as

$$E[g(x)] = \sum_{all\ x} g(x) \cdot p(x)$$

The *variance* of a discrete random variable x with probability distribution $p(x)$ is defined as

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

the *standard deviation* is defined as $\sqrt{E \left[(x - \mu)^2 \right]}$.

Example Suppose we are rolling a die for gambling. If we roll 1, 2, or 3 we lose 1 pound, if we roll 4 or 5 we win 1 pound, and if we roll 6 we will win 2 pounds. The function $g(x)$, sending each possible outcome 1, ..., 6 to the win -1, -1, -1, 1, 1, 2 is a function of the random variable x , and the expected value

$$\begin{aligned} E(g(x)) &= \sum_{\text{all } x} g(x) \cdot p(x) \\ &= -1 \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} \\ &= \frac{1}{6} \end{aligned}$$

is the average value that we are going to win (or lose) in each round of the game. (In other words, 17p here)

The following is a list of basic properties of expected values:

- $E(c) = c$, for every constant c ;
- $E(cx) = cE(x)$, for every constant c ;
- $E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]$,
for any two functions g_1, g_2 on x .

It follows the important formula used in many calculations that

$$\sigma^2 = E[x^2] - \mu^2$$

For the proof of this formula note that

$$\begin{aligned} \sigma^2 &= E \left[(x - \mu)^2 \right] = E \left[x^2 - 2\mu x + \mu^2 \right] \\ &= E[x^2] - 2\mu E[x] + \mu^2 E[1] \\ &= E[x^2] - 2\mu\mu + \mu^2 \end{aligned}$$

3.1.2 The Binomial Probability Distribution

The following statistical experiments have all common features:

- Tossing a coin 10 times.
- Asking 100 people on Princes Street in Edinburgh if they know that Madonna's wedding took place in a Scottish castle.
- Checking whether or not batches of transistors contain faulty transistors.

These experiments or observations are all examples of what is called a *binomial experiment* (the corresponding discrete random variable is called a *binomial random variable*). The examples have the following common characteristics:

- The experiment consists of n identical trials.
- In each trial there are exactly two possible outcomes (yes/no, pass/failure, or success/failure).
- The probability of success is usually referred to as p and the probability of failure as q , with $q = 1 - p$.
- The discrete (binomial) random variable x is the number of successes in the n trials.

The *binomial probability distribution* is given by the formula

$$p(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, \dots, n,$$

where

- p is the probability of a success in a single trial, and $q = 1 - p$;
- n is the number of trials; and
- x is the number of successes; and
- $\binom{n}{x}$ is the binomial coefficient given by the number $\frac{n!}{x!(n-x)!}$

The expected value (mean) and standard deviation are given by the formulae $\mu = np$ and $\sigma = \sqrt{npq}$

The following four example graphs Figure 3.1 Figure 3.2 Figure 3.3 Figure 3.4 show binomial distributions for different values of n and p , where p is always shown on the y -axis, the random variable on the x -axis:

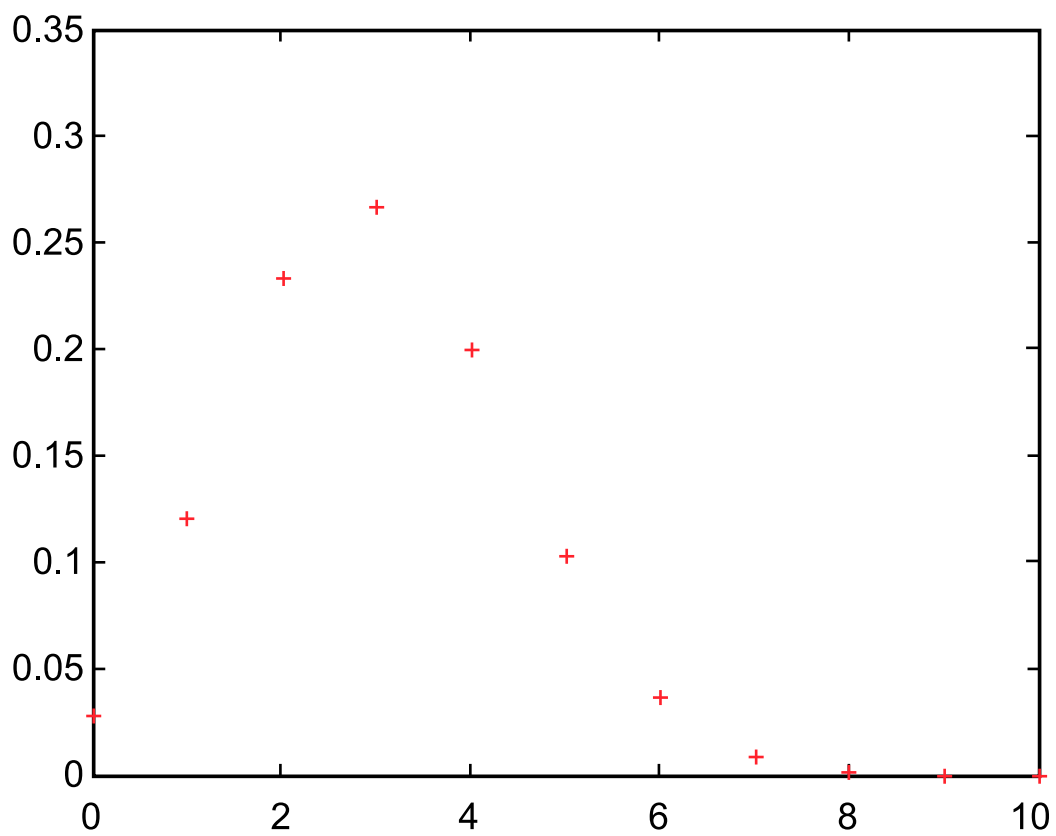


Figure 3.1: Binomial distribution, $n = 10, p = 0.3$,

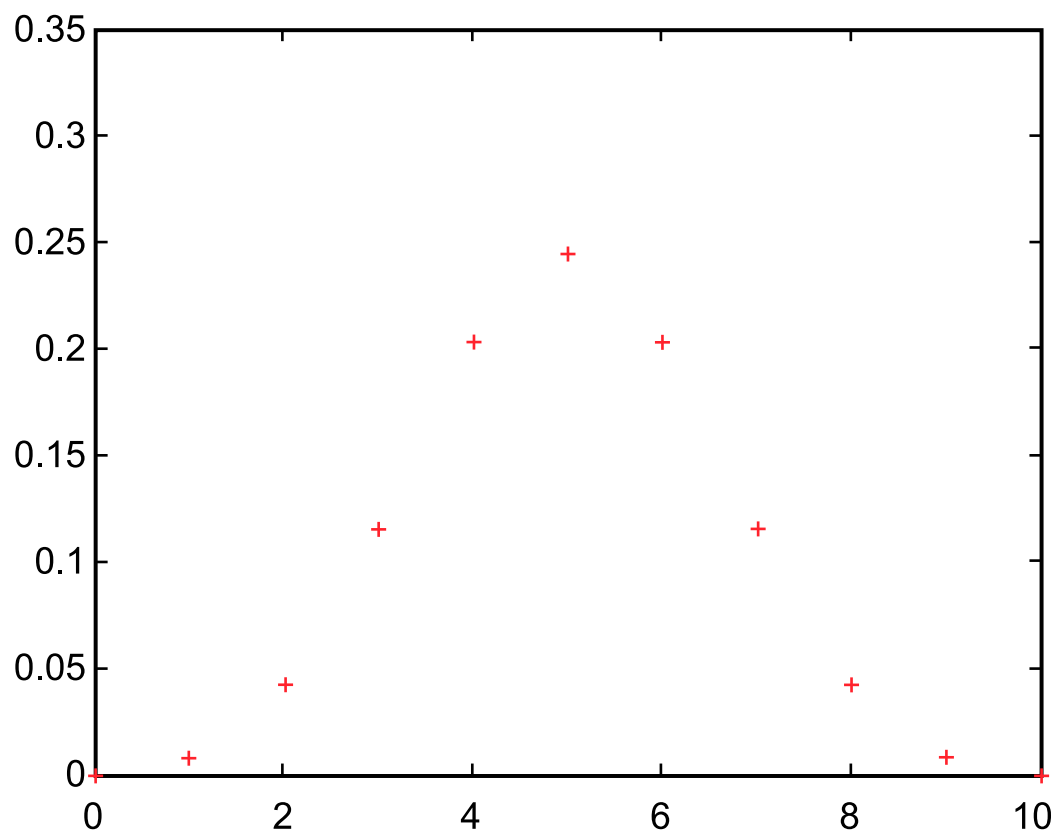


Figure 3.2: Binomial distribution, $n = 10, p = 0.5$

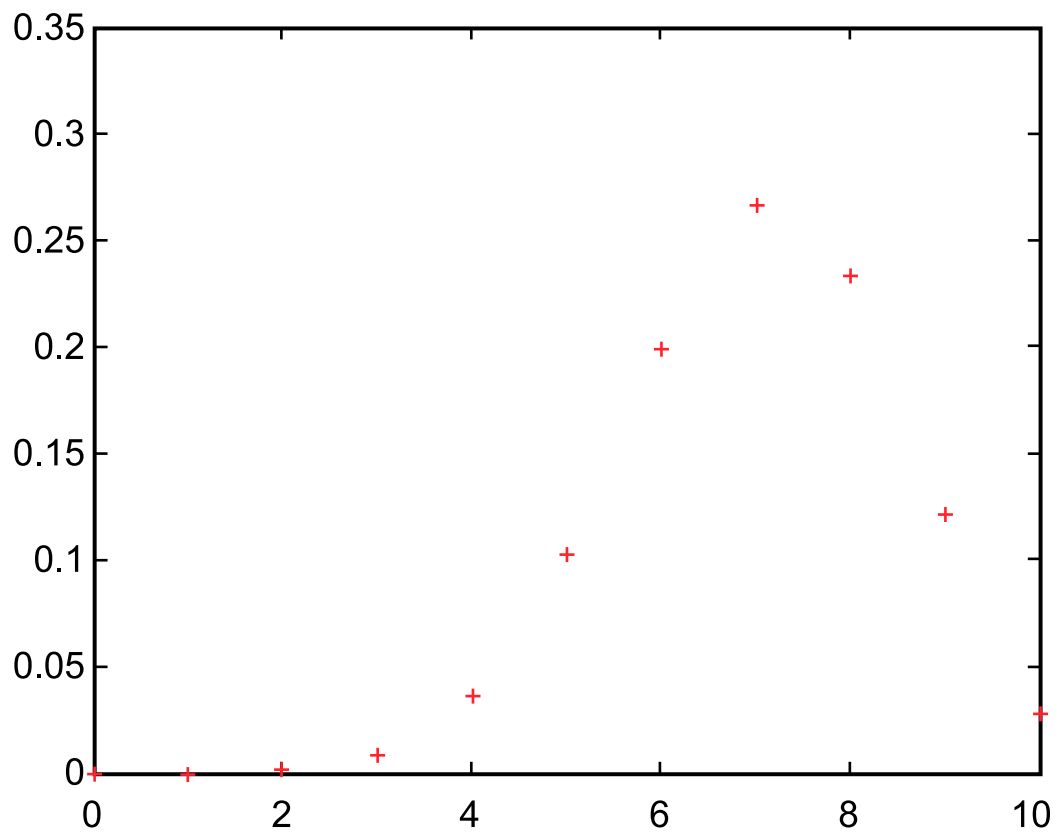


Figure 3.3: Binomial distribution, $n = 10, p = 0.7$

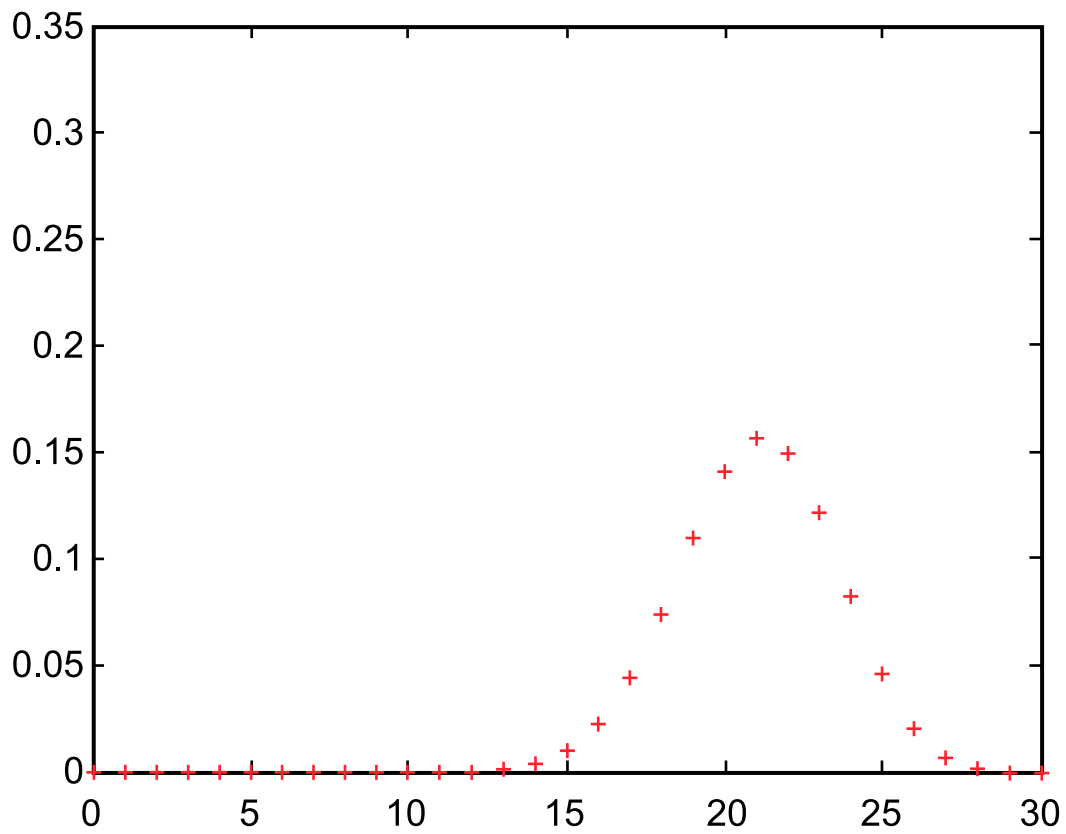


Figure 3.4: Binomial distribution, $n = 30, p = 0.7$

Example Tests show that about 20% of all private wells in some specific region are contaminated. What are the probabilities that in a random sample of 4 wells:

- a) exactly 2
- b) fewer than 2
- c) at least 2

wells are contaminated.

The random variable "number of contaminated wells" is clearly modelled by a binomial random variable. The probability that one particular well is contaminated is independent of the probability that some other well is contaminated, and for each well this probability is 0.2. Thus the parameters for the distribution are $n = 4$ and $p = 0.2$.

1.

$$\begin{aligned}
 P(x = 2) &= \binom{4}{2} 0.2^2 0.8^{4-2} \\
 &= \frac{4!}{2!2!} 0.2^2 0.8^2 \\
 &= \frac{4 \times 3}{2 \times 1} 0.2^2 0.8^2 \\
 &= 6 \cdot 0.2^2 \cdot 0.8^2 \\
 &= 0.1536
 \end{aligned}$$

2.

$$\begin{aligned}
 P(x < 2) &= P(x = 0) + P(x = 1) \\
 &= \binom{4}{0} 0.2^0 0.8^4 + \binom{4}{1} 0.2^1 0.8^3 \\
 &= 0.8192
 \end{aligned}$$

3.

$$\begin{aligned}
 P(x \geq 2) &= P(x = 2) + P(x = 3) + P(x = 4) \\
 &= 0.1536 + \binom{4}{3} 0.2^3 0.8^1 + \binom{4}{0} 0.2^4 0.8^0 \\
 &= 0.1808
 \end{aligned}$$

The following figure shows the graph for the corresponding probability distribution.

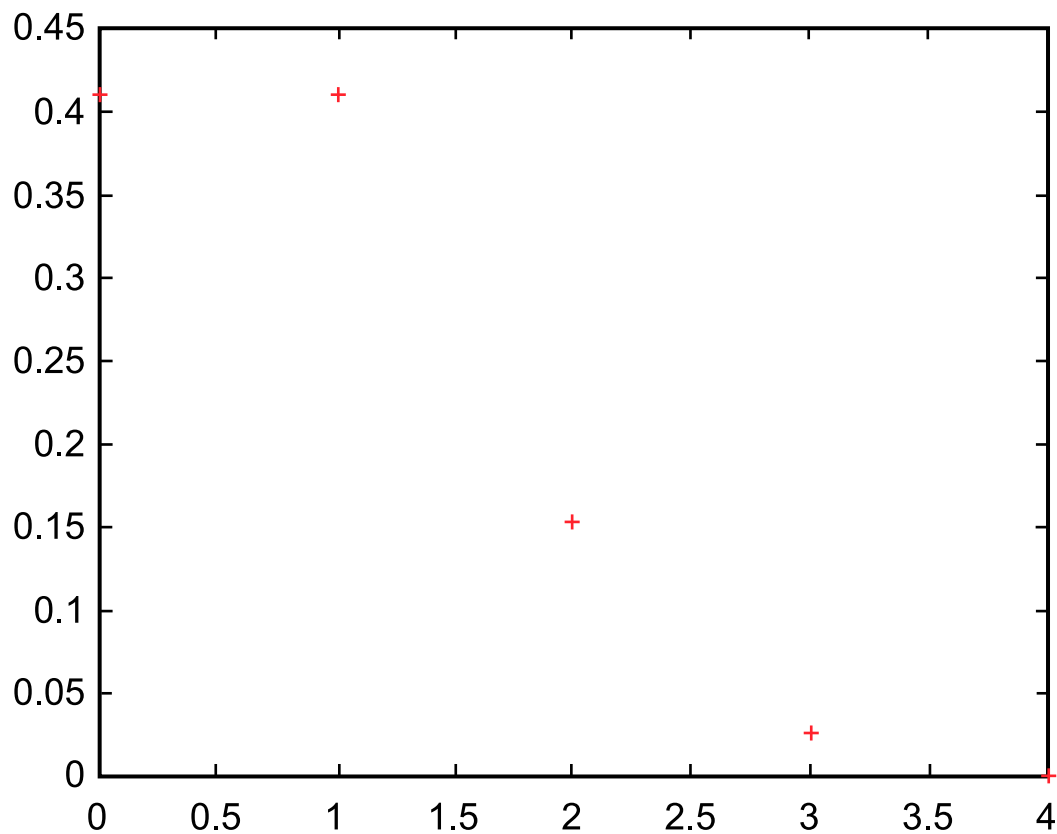


Figure 3.5: Binomial distribution, $n = 4, p = 0.2$

3.1.3 The Geometric Probability Distribution

Again we start with a typical application:

Example Customers wait in line to be served by a bank teller. Per minute the probability that a customer is served is 10%. What is the probability that a customer has to wait 15 minutes before being served?

Such and similar events are modeled by the *geometric probability distribution*. Each time interval we have an 'independent experiment' which can succeed or fail with success probability p (as for the binomial probability distribution). To be successful in the x th try we need $x - 1$ failures (with probability $q = p - 1$) and one success (with probability p).

The data for the geometric probability distribution are

- $P(x) = pq^{x-1}$, for $x = 1, 2, \dots$,
where x is the number of trials until the first success; and
- $\mu = 1/p$, and
-

$$\sigma = \sqrt{\frac{q}{p^2}}$$

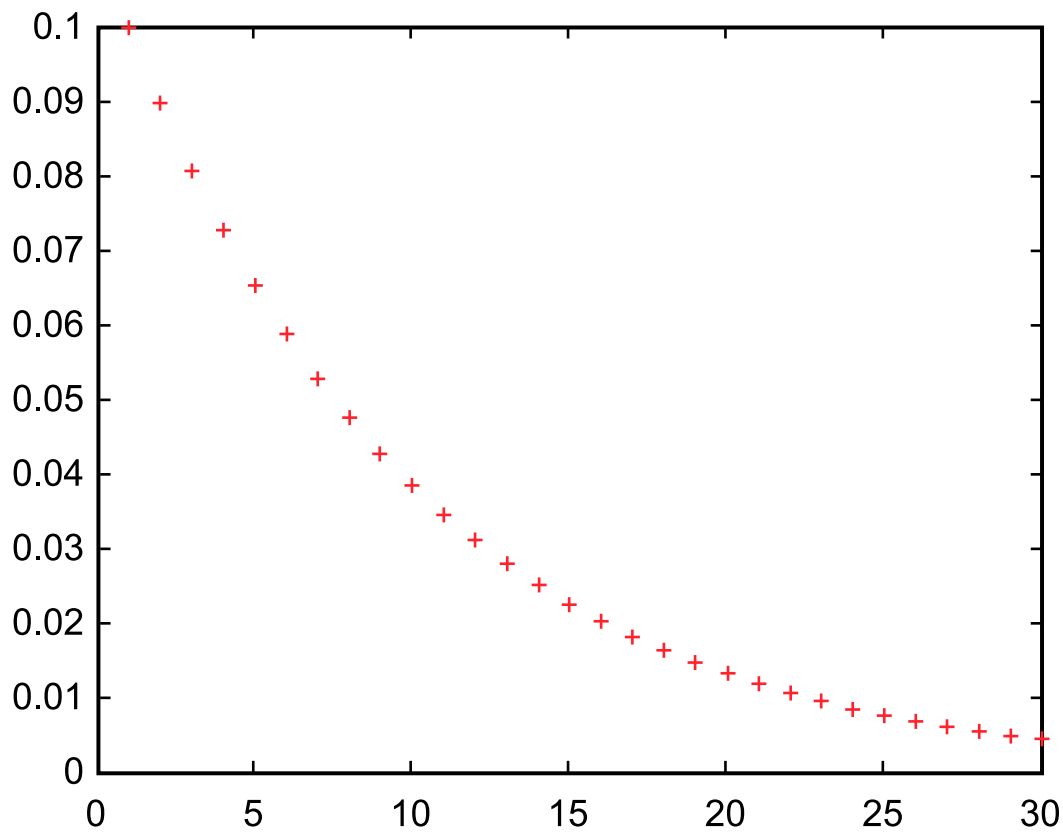


Figure 3.6: Geometric probability distribution, $p = 0.1$

Example The average life expectancy of a fuse is 15 months. What is the probability that the fuse will last exactly 20 months?

:

We have that $\mu = 15$ (months), or $p = 1/15$, which is the probability that a fuse will break.

For $x = 20$ we obtain

$$P(x = 20) = \frac{1}{15} \left(1 - \frac{1}{15}\right)^{20-1}$$

which is approximately 0.018. For σ we find $\sqrt{210} \approx 14.49$. The two graphs Figure 3.7 Figure 3.8 below show the probability distribution and the *cumulative distribution*, which is the function showing for each possible value x of the random variable the function $P(x' \leq x)$.

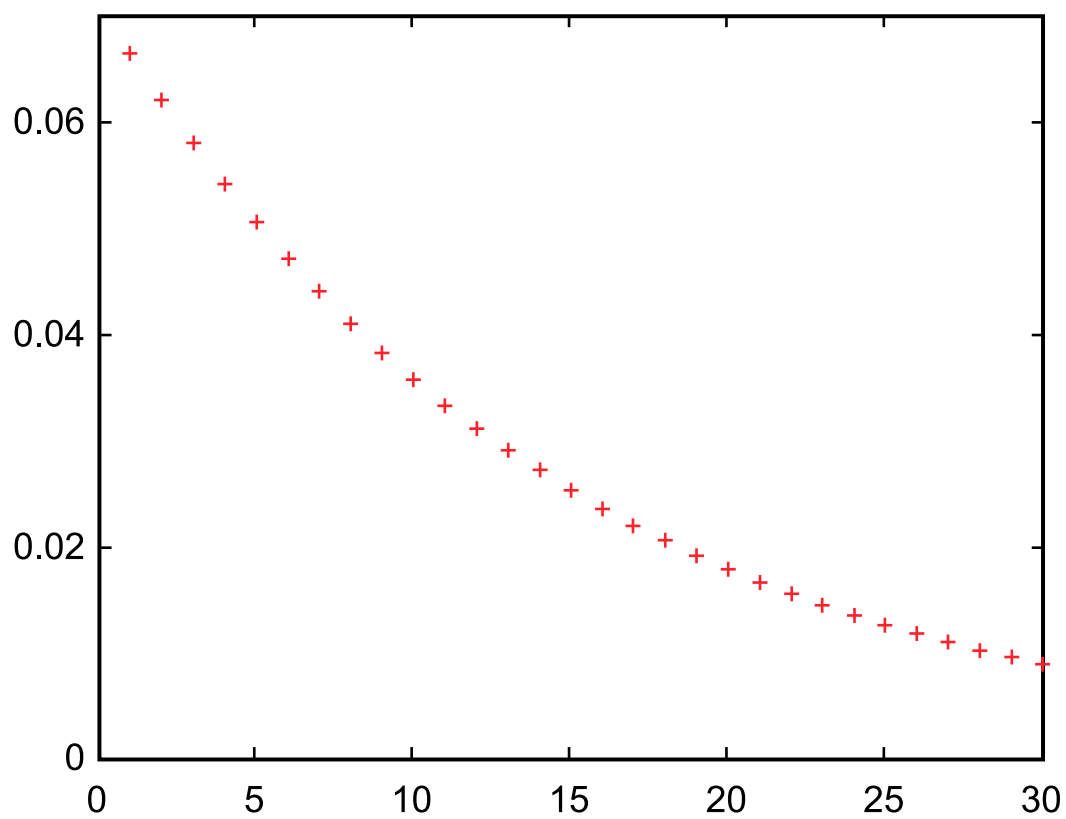


Figure 3.7: Geometric probability distribution, $p = 1/15$

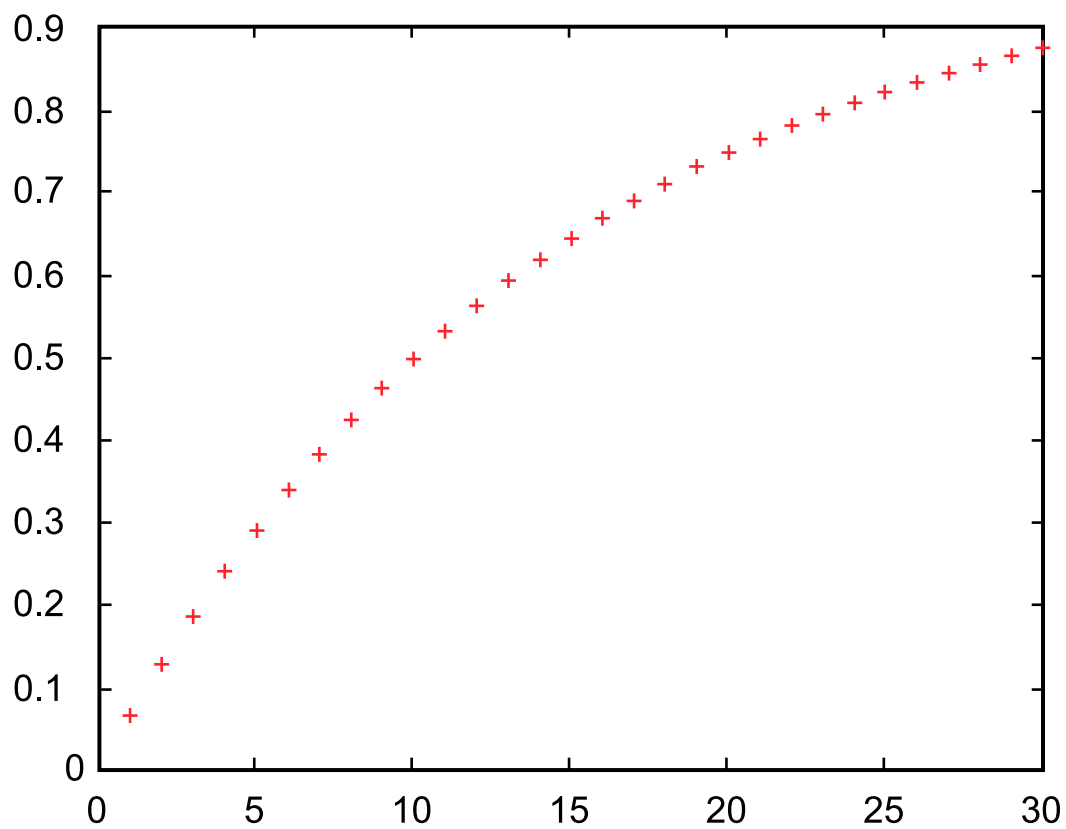


Figure 3.8: Geometric cumulative distribution, $p = 1/15$

3.1.4 The Hypergeometric Distribution

The binomial and the geometric probability distribution are to be applied if, after observing a result, the sample is put back into the population. However, in practice, we often sample without replacement:

- If we test a bag of 1000 resistors whether they meet certain specifications we usually will not put back the tested items.
- Suppose people are randomly selected on Princes Street in Edinburgh to fill in a questionnaire about a new product. When people are approached they are usually first asked whether they have already taken part in this marketing research.
- A big manufacturing company maintains their machines on a regular basis. Suppose that on average 15% of the machines need repair. What is the probability that among the five machines inspected this week, one of them needs repair?
- A box of 1000 fuses is tested one by one until the first defective fuse is found. Supposing that about 5% of the fuses are defective, what is the probability that a defective fuse is among the first 5 fuses tested?

Such and similar random variables have a *hypergeometric probability distribution*: The hypergeometric probability distribution is a discrete distribution that models sampling without replacement.

- The populations consists if N objects.
- The possible outcomes of the experiment are success or failure.
- Each sample of size n is equally likely to be drawn.

The formula for the hypergeometric probability distribution is

$$p(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}, \quad 0, n - N + r \leq x \leq n, r$$

where

- N is the number of elements in the population;
- r is the number in the population for success;
- n is the number of elements drawn; and
- x is the number of successes in the n randomly drawn elements.

The mean and standard deviation are given by

$$\mu = n \frac{r}{N}, \quad \text{and} \quad \sigma = \sqrt{\frac{r(N-r)n(N-n)}{N^2(N-1)}}$$

If we write $p = r / N$ then expected value and standard deviation become $\mu = np$ and

$$\sigma = \sqrt{\frac{N-n}{N-1} np(1-p)}$$

. This shows that the binomial and the hypergeometric distributions have the same expected value, but different standard deviations. The correction factor $\sqrt{\frac{N-n}{N-1}}$ is less than 1, but close to 1 if n is small relative to N . The following two graphs Figure 3.9 Figure 3.10 show the binomial and the hypergeometric distribution for different parameters:

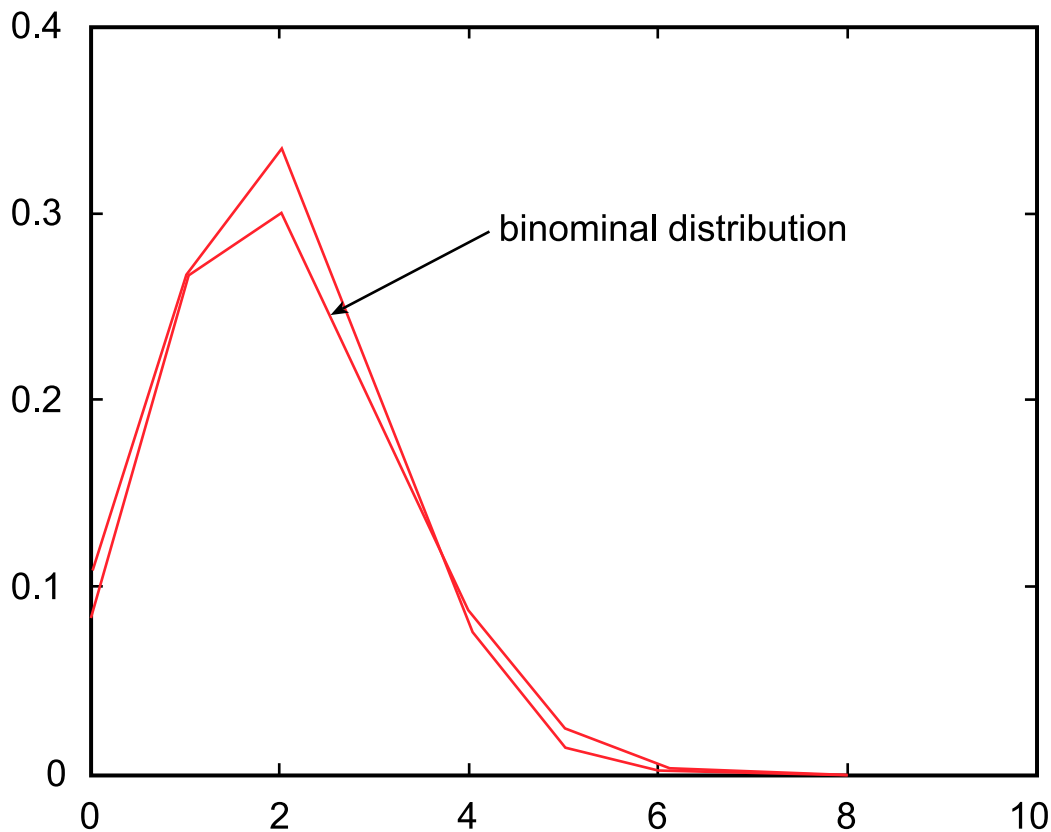


Figure 3.9: Comparing the binomial and the hypergeometric distribution,
 $n = 10, N = 50, r = 10, p = 0.2$

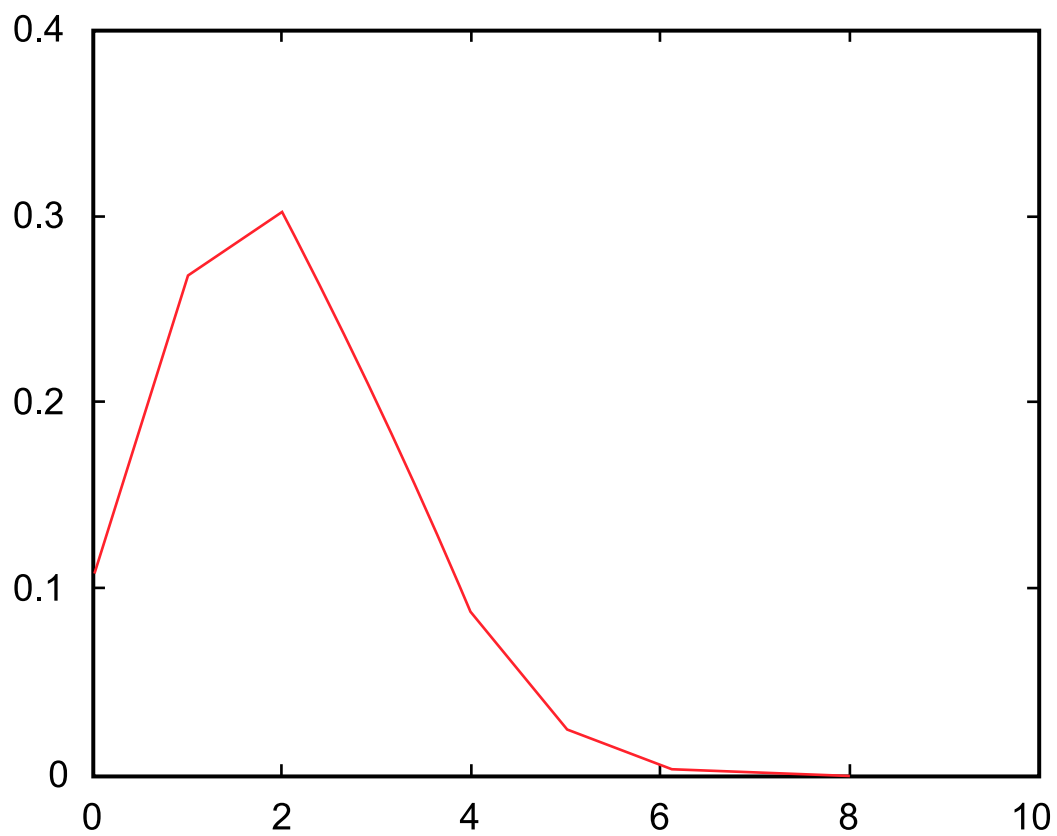


Figure 3.10: Comparing the binomial and the hypergeometric distribution,
 $n = 10, N = 1000, r = 200, p = 0.2$

Examples

1. A retailer sells computers. He buys lots of 10 motherboards from a manufacturer who sells them cheaply, but offers low quality only. Suppose the current lot contains one defective item. If the retailer usually tests 4 items per lot, what is the probability that the lot is accepted?

:

Here $N = 10$, $r = 1$, and $n = 4$, and we are looking for $P(x = 0)$, which is

$$P(x = 0) = \frac{\binom{1}{0} \binom{9}{4}}{\binom{10}{4}} = \frac{1 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} \frac{1 \cdot 2 \cdot 3 \cdot 4}{10 \cdot 9 \cdot 8 \cdot 7} = \frac{6}{10}$$

We would use the same calculation if we would only know that on average 10% of the motherboards are faulty.

2. We test lots of 100 fuses. On average 5% of the fuses are defective. If we test 4 fuses, what is the probability that we accept the current lot?

:

Again, the random variable is hypergeometric, and since $N = 100$ is large we can assume that there are 5 defective fuses in this lot. We find

$$P(x = 0) = \frac{\binom{5}{0} \binom{95}{5}}{\binom{100}{5}} = \frac{\frac{5!}{0!5!} \frac{95!}{5!90!}}{\frac{100!}{5!95!}} = \frac{95 \times 94 \times 93 \times 92 \times 91}{100 \times 99 \times 98 \times 97 \times 96} \approx 0.7696$$

Later we will see how reliable this value is, as we don't know the exact number of faulty fuses in this lot.

3.1.5 The Poisson Distribution

The *Poisson probability distribution* provides a model for the frequency of events, like the number of people arriving at a counter, the number of plane crashes per month, or the number of micro-cracks in steel. (Micro-cracks in steel wheels of a German high-speed train ICE led to a disastrous rail accident in 1998 killing 101 people.) The characteristics of a Poisson random variable are as follows:

- The experiment consists of counting events in a particular unit (time, area, volume, etc.).
- The probability that an event occurs in a given unit is the same for every unit.

- The number of events that occur in one unit is independent of the number of events that occur in other units

The Poisson probability distribution with parameter λ is given by the formula

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad (x = 0, 1, 2, \dots)$$

where e is the mathematical constant 2.71828..... The expected value and standard deviation are $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$

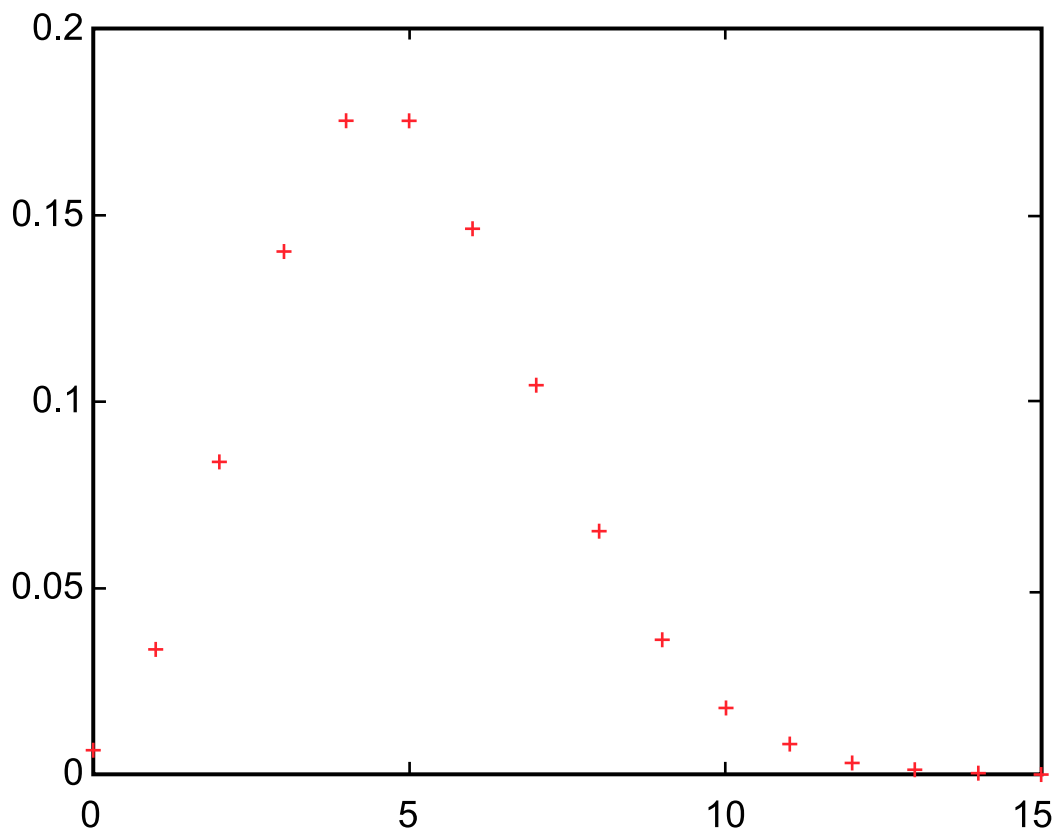


Figure 3.11: Poisson distribution, $\lambda = 5$

Figure 3.11 and Figure 3.12 illustrate the Poisson distribution with values of $\lambda = 5$ and 10 respectively.

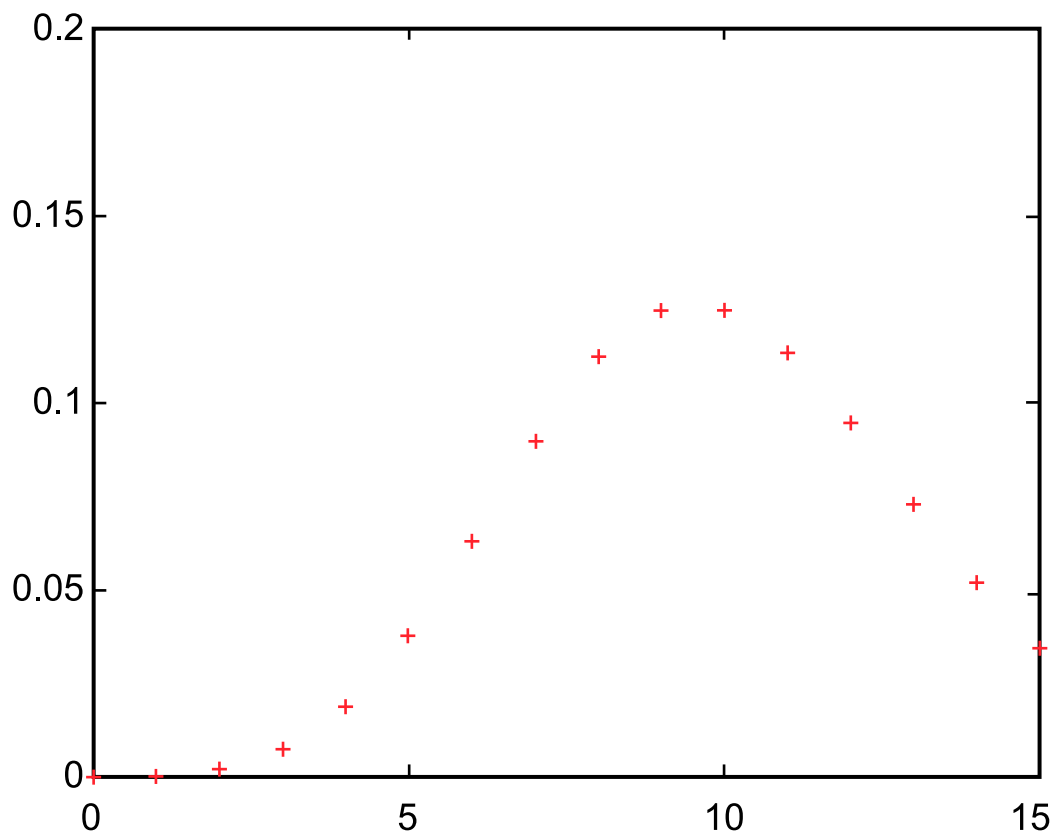


Figure 3.12: Poisson distribution, $\lambda = 3$

Example Suppose customers arrive at a counter at an average rate of 6 per minute, and suppose that the random variable ‘customer arrival’ has a Poisson distribution. What is the probability that in a half-minute interval at most one new customer arrives?

Here $\lambda = 6/2 = 3$ customers per half-minute. So

$$\begin{aligned} P(x \leq 1) &= P(x = 0) + P(x = 1) \\ &= \frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} \\ &= \frac{4}{e^3} \end{aligned}$$

which equals approximately 0.1991.

The following two graphs Figure 3.13 Figure 3.14 show the probability distribution $p(x)$ for this example, and the *cumulative distribution* $F(x) = \sum_{x' \leq x} p(x')$

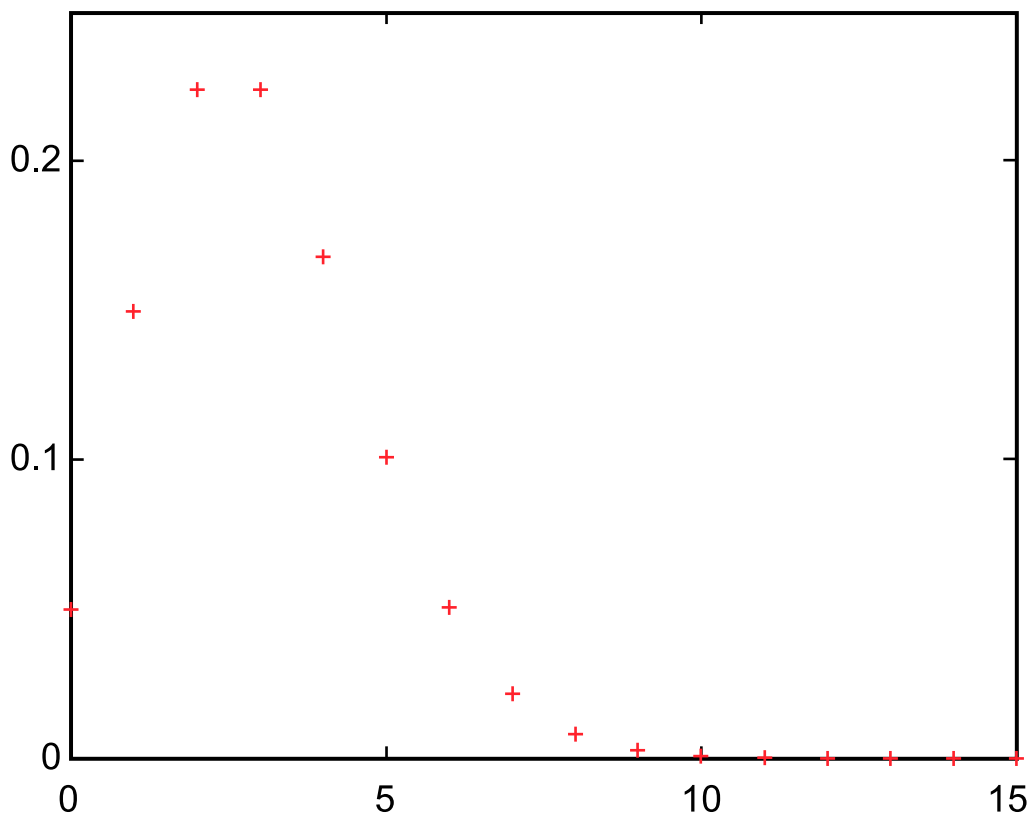


Figure 3.13: Poisson distribution, $\lambda = 3$

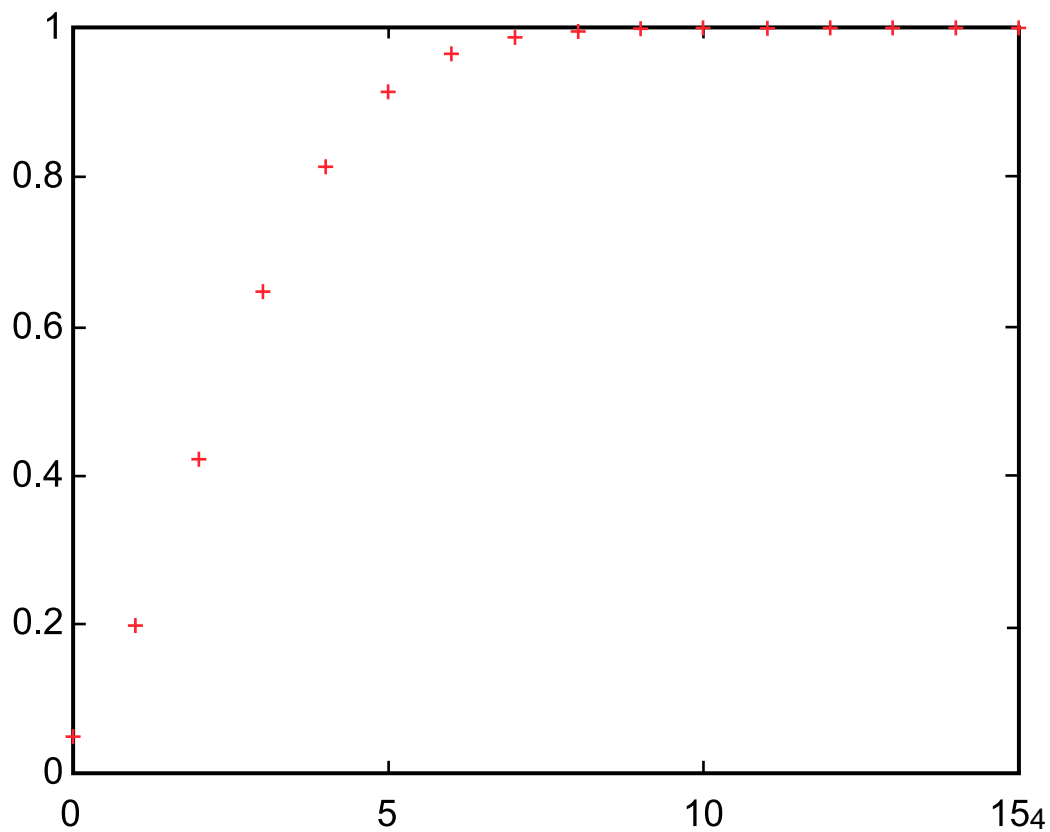


Figure 3.14: Cumulative Poisson distribution, $\lambda = 3$

As an example we will verify that the Poisson probability distribution $p(x)$ really is a distribution, and that the mean is λ . The reader not familiar with infinite sums should skip the rest of this subsection.

We note first that $0 \leq p(x)$ for all values of x . Also, since we have the infinite sum

$$e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

we can calculate $1 = e^{-\lambda} e^\lambda = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=0}^{\infty} p(x)$, which shows that $p(x) \leq 1$ and $\sum_{all\ x} p(x) = 1$.

For the mean we calculate

$$\begin{aligned} E(x) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= 0 + \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} \frac{\lambda e^{-\lambda} \lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \lambda \end{aligned}$$

Here from line 2 to line 3 we cancelled x and took out one λ in the numerator, and in the

next step replaced $x - 1$ (running through $x = 1, 2, \dots$) by x (running through $x = 0, 1, \dots$).

3.2 Continuous Random Variables

Many random variables arising in practice are not discrete. Examples are the strength of a beam, the height of a person, the capacity of a conductor, or time it takes to access memory, etc. Such random variables are called *continuous*.

A practical problem arises, as it is *impossible* to assign finite amounts of probabilities to uncountably many values of the real axis (or some interval) so that the values add up to 1. Thus, continuous probability distributions are usually based on *cumulative distribution functions*.

The cumulative distribution $F(x)$ of a random variable x is the function $F(x) = P(x' \leq x)$.

The following graphs Figure 3.15 Figure 3.16 Figure 3.17 Figure 3.18 show the binomial probability distribution with $n = 10$, $p = 0.3$, and the corresponding cumulative distribution, and as a second example, a Poisson distribution and the corresponding cumulative distribution. Both distributions are, of course, discrete examples.

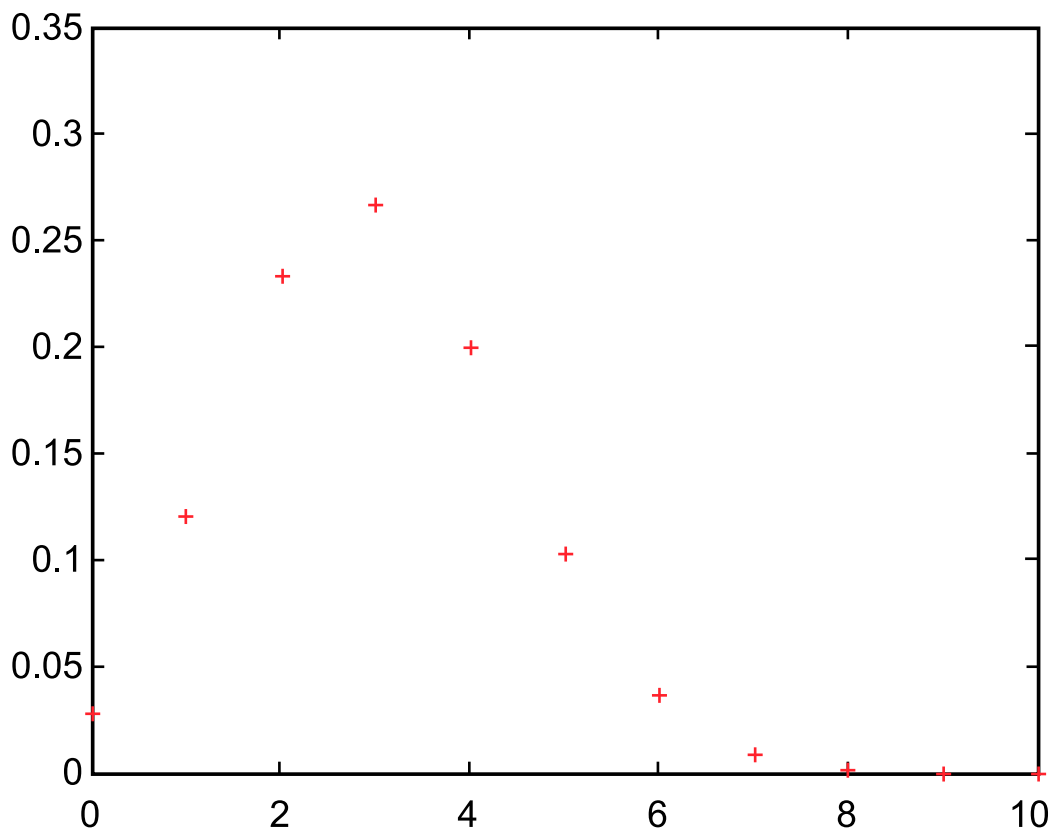


Figure 3.15: Binomial distribution, $n = 10$, $p = 0.3$

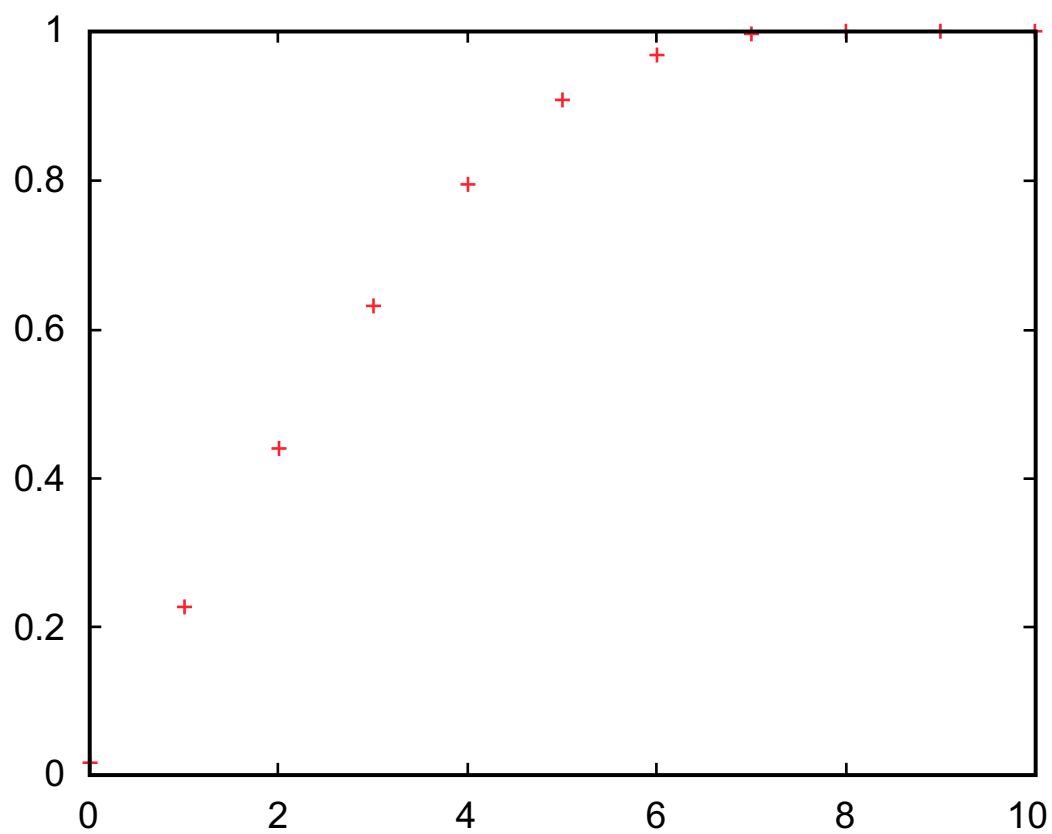


Figure 3.16: Cumulative binomial distribution, $n = 10, p = 0.3$

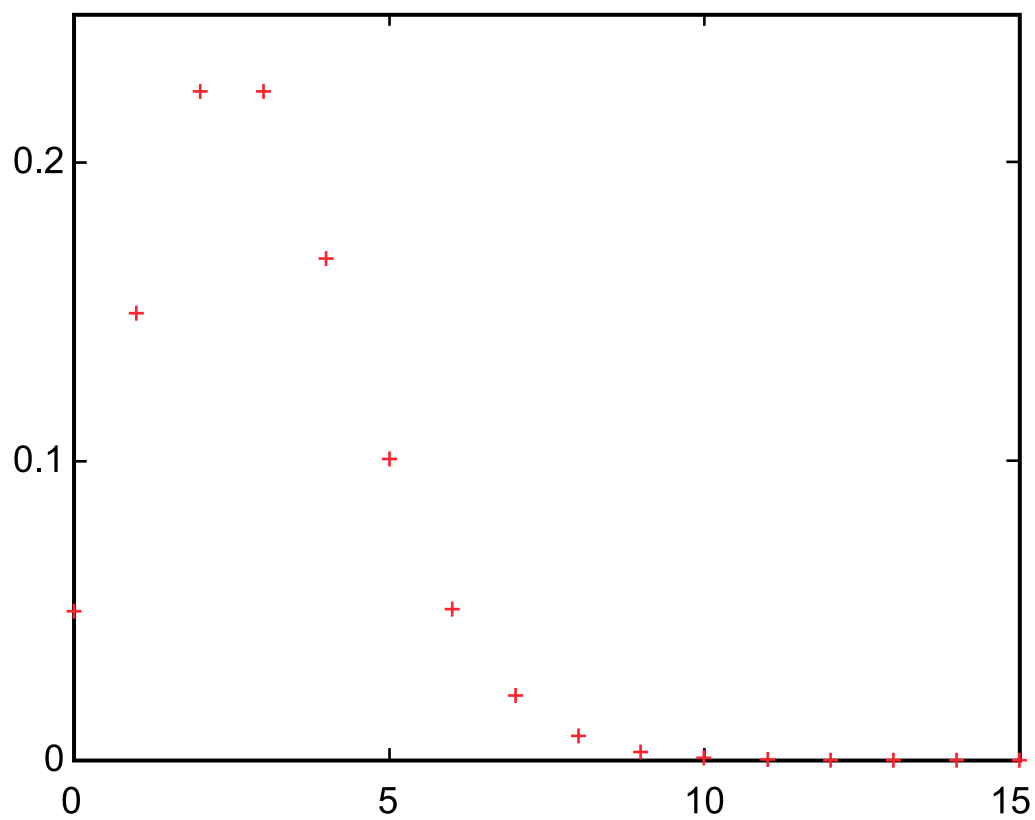


Figure 3.17: Poisson distribution, $\lambda = 3$

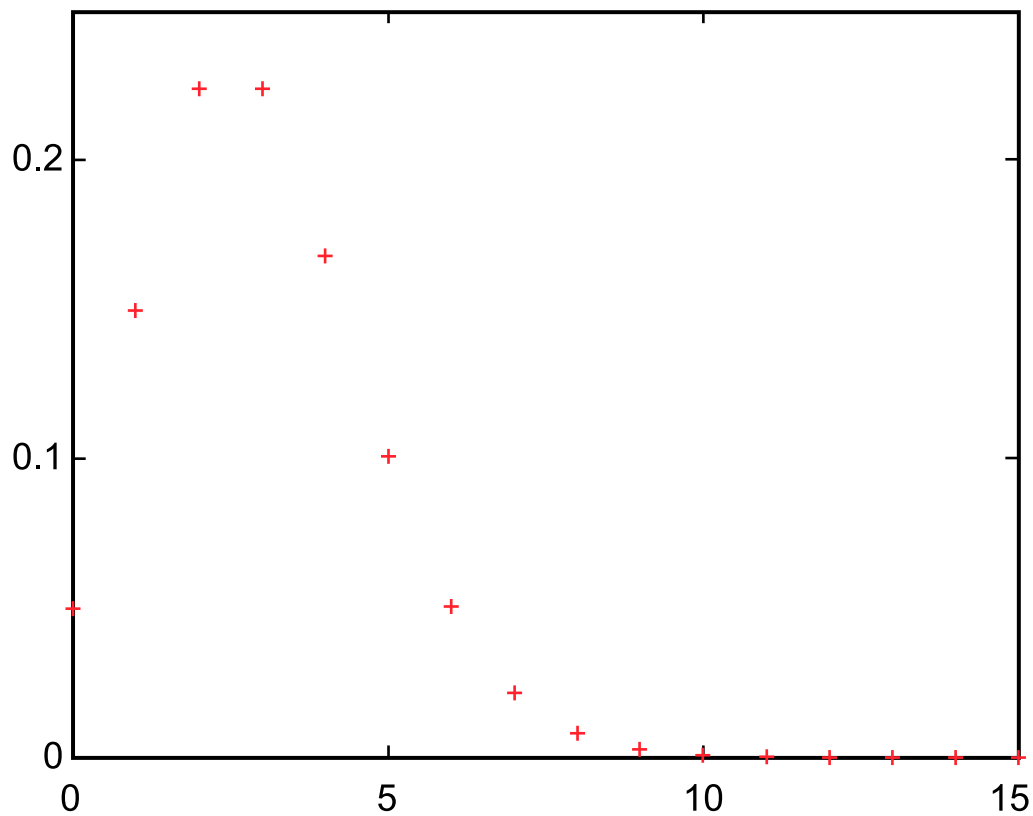


Figure 3.18: Cumulative Poisson distribution, $\lambda = 3$

If F is the cumulative distribution of a continuous random variable x then the *density* function $\rho(x)$ for x is given by $\rho(x) = \frac{dF}{dx}$ (provided that F is differentiable). It follows that

$$F(x) = \int_{-\infty}^x \rho(t) dt$$

and that the density function satisfies $\rho(x) \geq 0$ and

$$\int_{-\infty}^{\infty} \rho(t) dt = 1$$

. The probability between two values a and b is given by

$$\int_a^b \rho(t) dt$$

.

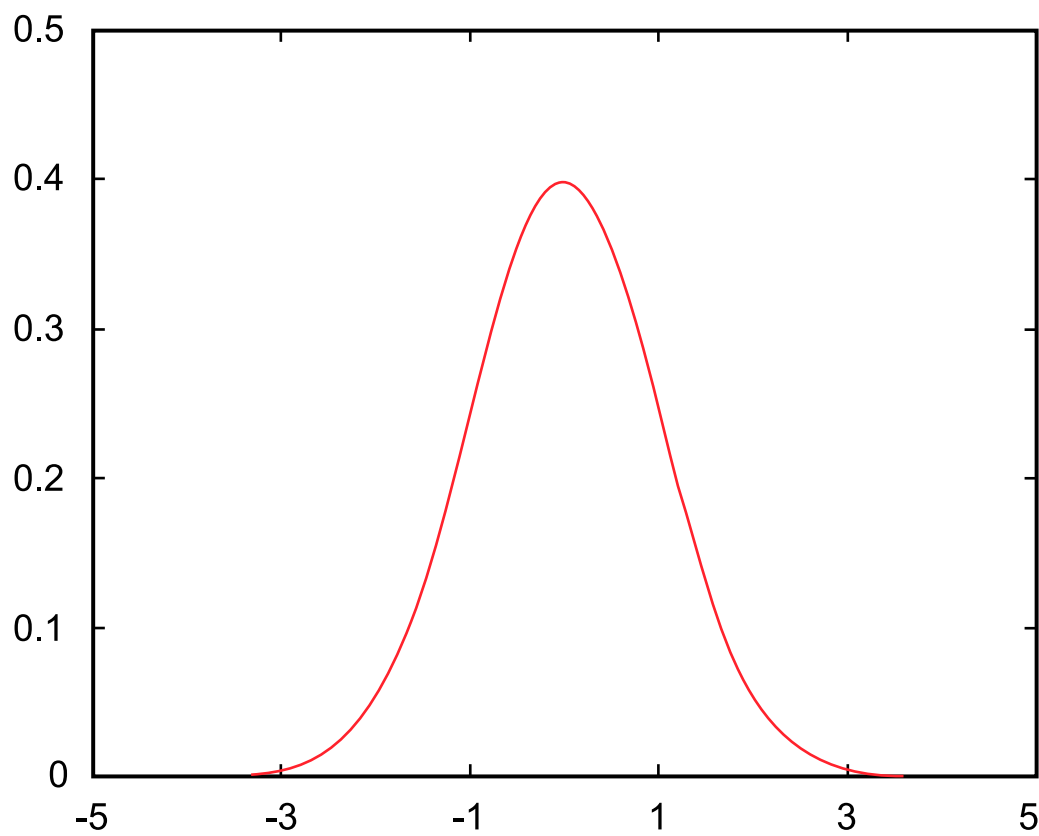


Figure 3.19: Example of a continuous probability distribution

Figure 3.19 illustrates a typical continuous probability distribution and Figure 3.20 its cumulative equivalent

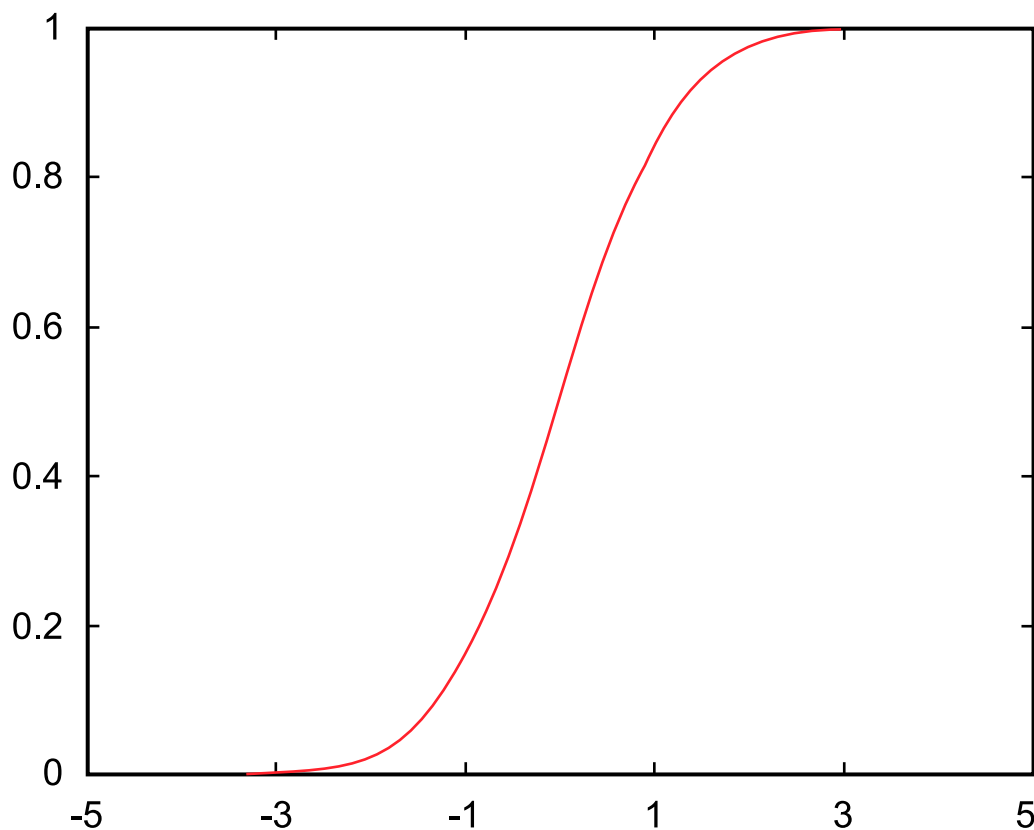


Figure 3.20: The corresponding cumulative distribution

Let us recall from calculus that an integral is a limit process of a summation. Finding

$$F(x_0) = \int_{-\infty}^{x_0} \rho(x) dx$$

for a continuous random variable is analogous to finding

$$F(x_0) = \sum_{x \leq x_0} \rho(x)$$

for a discrete random variable. Thus, we define the *expected value* analogous to the discrete case.

The *expected value* of a continuous random variable x with density function $\rho(x)$ is given by

$$\mu = E(x) = \int_{-\infty}^{\infty} t \rho(t) dt$$

If g is any function we define the *expected value of $g(x)$* as

$$E[g(x)] = \int_{-\infty}^{\infty} g(t) \rho(t) dt$$

provided that these integrals exist. The *standard deviation* is defined as

$$\sigma = \sqrt{E[(x - \mu)^2]}. \text{ Note that}$$

- $E(c) = c$, for every constant c ;

- $E(cx) = cE(x)$, for every constant c ;
- $E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]$,
for any two functions g_1, g_2 on x .
- $\sigma^2 = E[x^2] - \mu^2$.

3.2.1 The Uniform Probability Distribution

If we select randomly a number in the interval $[a, b]$ then the corresponding random variable x is called a *uniform random variable*. Its density function is

$$\rho(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{else} \end{cases}$$

For the mean and standard deviation one finds

$$\mu = \frac{a+b}{2} \text{ and } \sigma = \frac{b-a}{2\sqrt{3}} = \frac{\sqrt{3}}{6}(b-a)$$

Example A manufacturer of wires believes that one of her machines makes wires with diameter uniformly distributed between 0.98 and 1.03 millimeters.

The mean of the thickness is $\frac{1.03+0.98}{2} = 1.005$ millimeters, and the standard deviation is $\sigma = \frac{\sqrt{3}}{6}(1.03 - 0.98) \approx 0.014$ millimeters. The density function for this uniform random variable is $\rho(x) = \frac{1}{.05} = 20$ for $0.98 \leq x \leq 1.03$, and 0 elsewhere. And, for example,

$$\begin{aligned} P(x \leq 1.00) &= \int_{-\infty}^{0.98} \rho(t) dt + \int_{0.98}^{1.00} \rho(t) dt \\ &= \int_{0.98}^{1.00} 20 dt \\ &= 20[1.00 - 0.98] \\ &= 0.4 \end{aligned}$$

The corresponding distribution is shown in uniformdistributionFigure 3.21

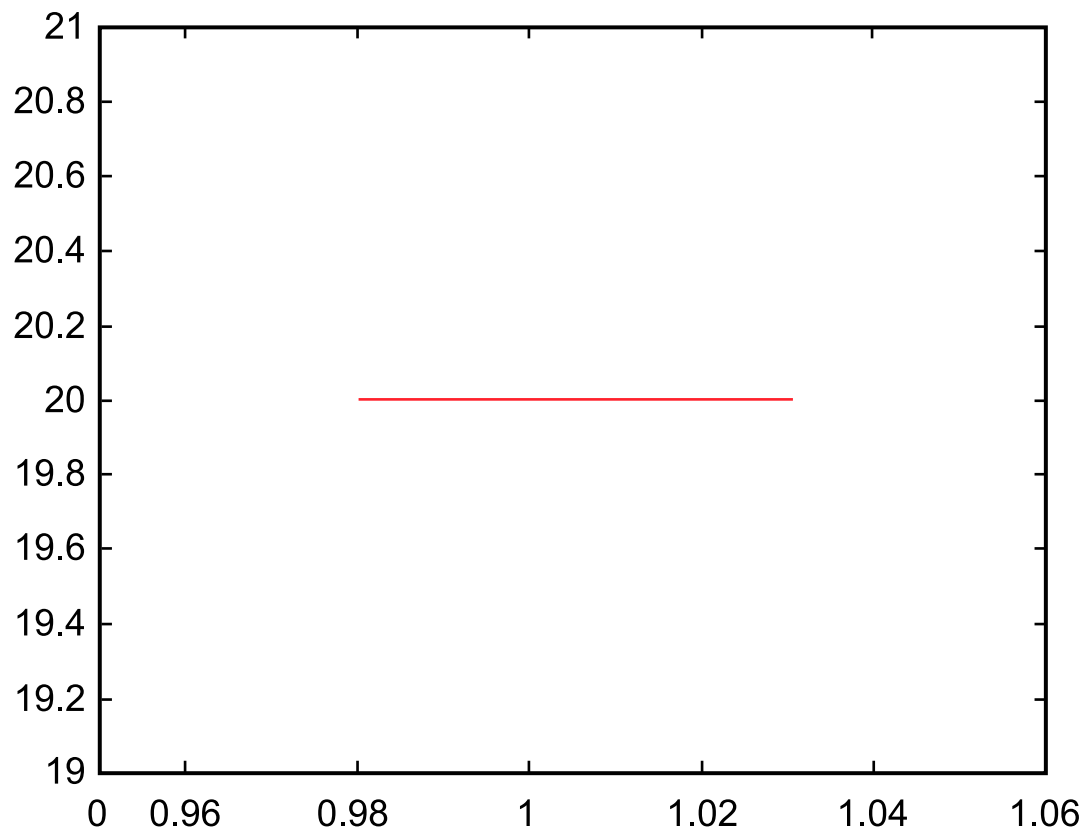


Figure 3.21: Uniform distribution, $[0.98, 1.03]$

3.2.2 The Gamma Distribution

Many continuous random variables can only take positive values, like height, thickness, life expectations of transistors, etc. Such random variables are often modeled by *gamma type random variables*. The corresponding density functions contain two parameters α, β . The first is known as the *shape* parameter, the second as the *scale* parameter.

Figure 3.22 shows Gamma density function with $\alpha = 1, 3, 6$ and $\beta = 1$

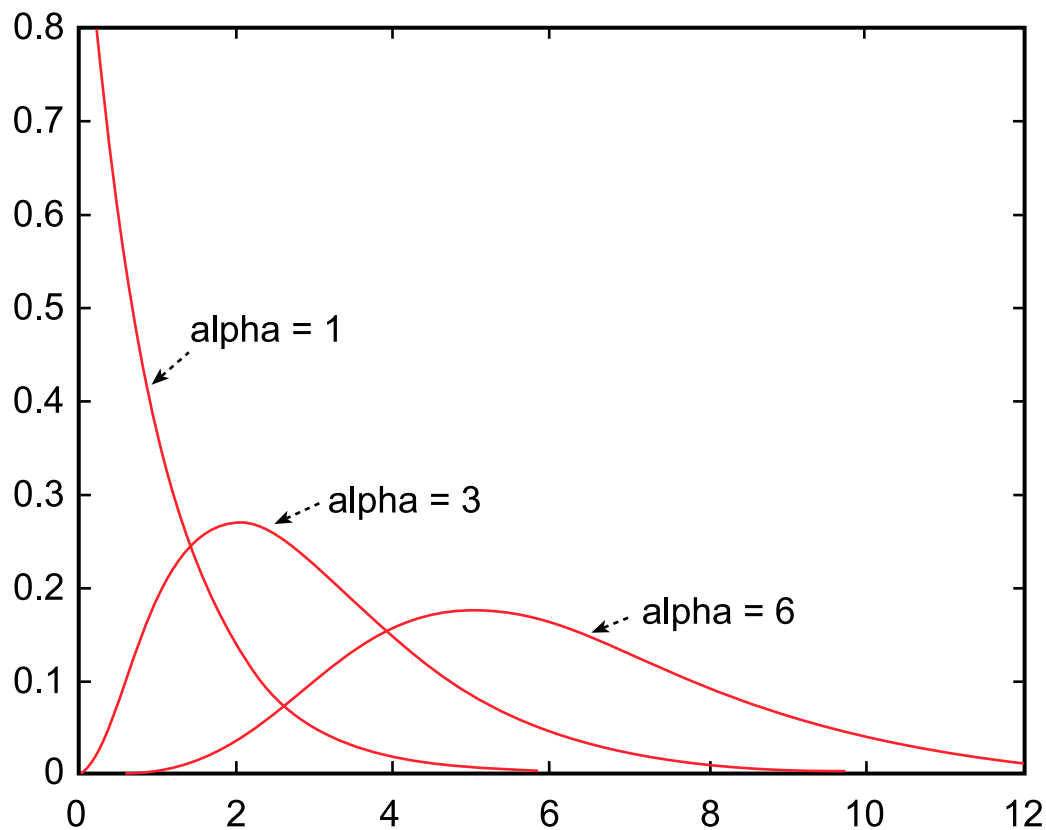


Figure 3.22: Gamma density functions, $\alpha = 1, 3, 6$ and $\beta = 1$

The density function is given by

$$\rho(x) = \begin{cases} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta \alpha \Gamma(\alpha)} & \text{if } 0 \leq x < \infty, \alpha, \beta \geq 0 \\ 0 & \text{else} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ is the *gamma function*, giving the gamma distribution its name. The mean and standard deviation are

$$\mu = \alpha\beta \text{ and } \sigma = \sqrt{\alpha\beta^2}$$

The gamma function plays an important role in mathematics. It holds that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, and $\Gamma(1) = 1$, so that for integer values of α , $\Gamma(\alpha) = \alpha!$. However, in general there is no closed form for the gamma function, and its values are approximated and taken from tables.

Example A manufacturer of CPUs knows that the relative frequency of complaints from customers (in weeks) about total failures is modeled by a gamma distribution with $\alpha=2$ and $\beta=4$. Exactly 12 weeks after the quality control department was restructured the next (first) major complaint arrives. Does this suggest that the restructuring resulted in an improvement of quality control?

We calculate $\mu = \alpha\beta = 8$ and $\sigma = 4\sqrt{2} \approx 5.657$. The value $x=12$ lies well within one standard deviation from the (old) mean, so we would not consider it an exceptional value. Thus there is insufficient evidence to indicate an improvement in quality control given just this data.

3.2.3 The Chi-Square Distribution

The χ^2 (*chi-square*) probability distribution plays an important role in statistics. The distribution is a special case of the gamma distribution for $\alpha = \frac{\nu}{2}$ and $\beta=2$ (ν is called the *number of degrees of freedom*). The density function is

$$\rho(\chi^2) = c(\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

where $c(\chi^2) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$. For mean and standard deviation one finds

$$\mu = \nu \text{ and } \sigma = \sqrt{2\nu}$$

3.2.4 The Exponential Probability Distribution

The exponential density function is a gamma density function with $\alpha=1$,

$$\rho(x) = \frac{e^{-\frac{x}{\beta}}}{\beta}, \quad x \geq 0$$

with mean $\mu=\beta$ and standard deviation $\sigma=\beta$. The corresponding random variable models for example the length of time between events (arrivals at a counter, requests to a CPU, etc) when the probability of an arrival in an interval is independent from arrivals in other intervals. This distribution also models the life expectancy of equipment or products, provided that the probability that the equipment will last t more time intervals is the same as for a new product (this holds for well-maintained equipment).

If the arrival of events follows a Poisson distribution with mean $\frac{1}{\beta}$ (arrivals per unit interval), then the time interval between two successive arrivals is modeled by the exponential distribution with mean β .

Activity

Q1: Emails come in to a university at a rate of 3 per minute. Assuming a Poisson distribution, calculate the probability that in any two-minute period, more than 2 emails arrive.

Q2: In a large school it was reported that 1 out of 3 pupils take school meals regularly. In a random sample of 20 pupils from this school, calculate the probability that exactly 5 do not take school meals regularly.



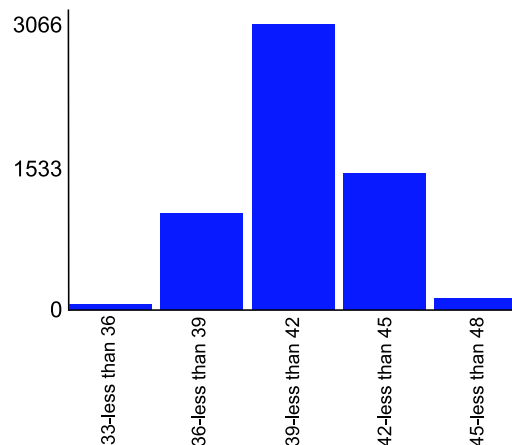
3.3 The Normal Distribution

3.3.1 Everyday examples of Normal Distribution

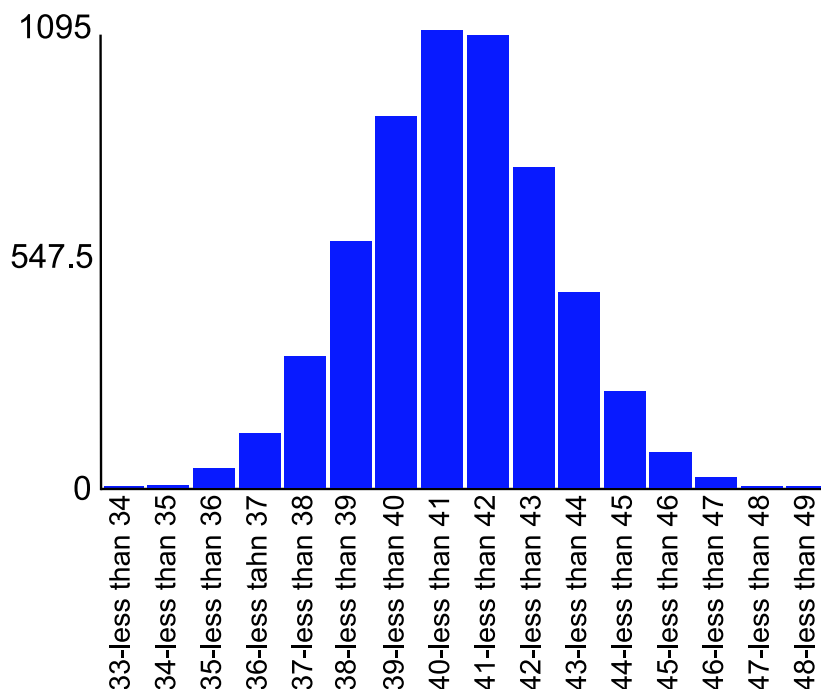
A number of continuous distributions have been defined in the previous section. By far the most common to be used, however, is the Normal distribution and so it is described in detail here

In day to day life much use is made of statistics, in many cases without the person doing so even realising it. If you were to go into a shop and you noticed that everybody

waiting to be served was over 6 and a half feet tall, you would more than likely be a bit surprised. You probably would have expected most people to be around the "average" height, maybe spotting just one or two people in the shop that would be taller than 6 and a half feet. In making this judgement you are actually employing a well used statistical distribution known as the **Normal Distribution**. There are numerous things that display the same characteristic including body temperature, shoe size, IQ score and diameter of trees to name but a few. Recall that in Topic 1, when a histogram was plotted of the chest sizes of Scottish soldiers, the graph had the appearance:



Now consider changing the number of class intervals. The image below will show you what happens to the histogram as the number of class intervals increases.



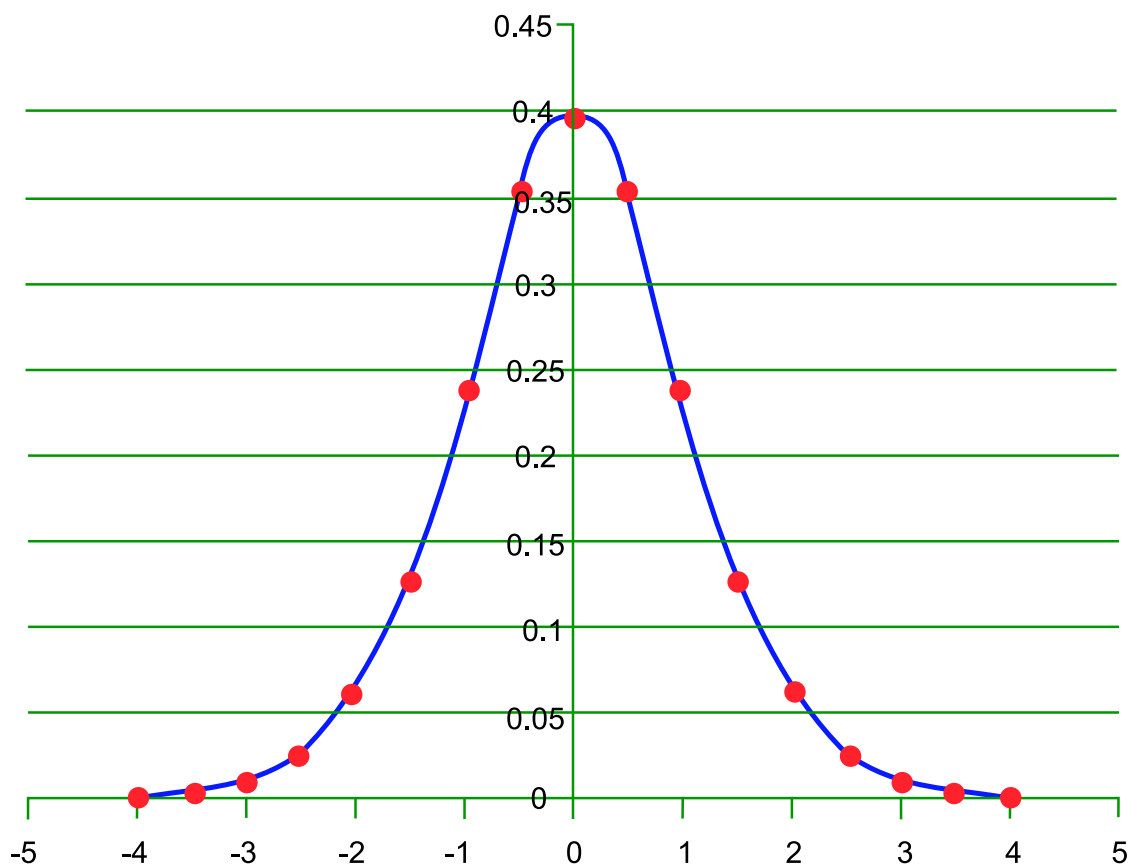
Notice that as the number of class intervals increases the graph begins to take on the shape of a bell. Of course, as more and more class intervals are formed, the size of each class interval will become smaller and smaller. If it were possible to make a class interval of just the value itself, the graph would actually become a smooth curve as shown below.

This will hopefully soon become the very familiar shape to you of the Normal distribution that will be used many times throughout this course. Much work on this topic was carried out by a German mathematician called Johann Carl Friedrich Gauss; indeed the distribution is also sometimes called the Gaussian distribution.

3.3.2 Drawing the Curve

Like some other curves that you may have plotted in the past the Normal distribution curve can be represented by an equation.

This equation is given by $y = \frac{1}{\sqrt{2\pi}}e^{-0.5x^2}$ where e is the number 2.71828... very often used in mathematics and π is the familiar number 3.14159... The following animation calculates the value of y for various values of x and plots the graph. You may wish to plot the graph yourself on paper to get a feel for its shape.



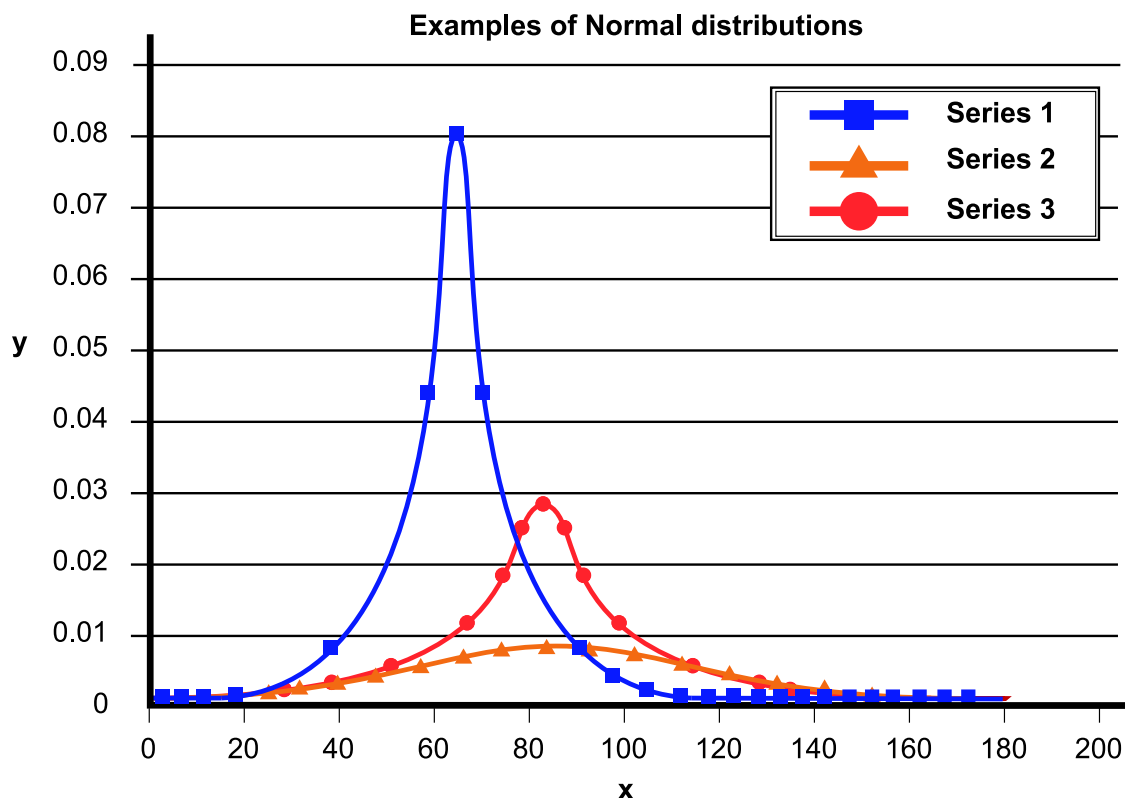
Note that in the equation, The term $\frac{1}{\sqrt{2\pi}}$ is called a normalisation factor and is chosen to make the total area under the curve equal to 1.

It may seem difficult to accept that so many examples from real life can produce diagrams that have this same shape, as clearly there will be major differences; the numbers for heights of humans, for example, will have values like 175, 177 or 169 (centimetres), whilst the volume of liquid in a sample of milk cartons may have measurements like 500, 502 or 498 (millilitres). In addition, some sets of results will be very tightly clustered around the mean whilst other data sets will have a large spread. In fact, the curve drawn above is the Standardised Normal Distribution and relates to a

population with mean=0 and standard deviation=1. By simply using a transformation to scale results, though, it is possible to represent **any** Normal distribution by a curve like the one shown. So it is an important fact, then, that a Normal distribution is dependent on two variables, the mean (μ) and standard deviation (σ).

The general equation of a Normal distribution curve is in fact $y = \frac{1}{\sigma\sqrt{2\pi}}e^{-0.5(x-\mu)^2/\sigma^2}$

A few examples of Normal distribution curves are now drawn on the same diagram. Series 1 represents lifetimes of a certain type of battery which has mean 82 and standard deviation 15. Series 2 represents diameters of leaves on a particular plant that has mean leaf diameter 60mm and standard deviation 5mm. Series 3 is from a population consisting of weights of cement bags with mean 85kg and standard deviation 40kg.



Each series displays a typical Normal distribution shape. It should be becoming clear, then, that by making a transformation on the units of measurement, all three graphs could be redrawn as the standardised normal curve discussed earlier.

It must be mentioned here that you will never have to use the equation for the Normal curves to carry out any calculations so do not be frightened off by the complicated looking formula. The equation is simply mentioned to show you that the normal curve can be drawn in the same way as any other much simpler curve could be (e.g. $y = x^2$).

3.3.3 Calculations Using Normal Distribution

To transform the data for a particular example into values appropriate to the standardised Normal curve requires the use of a formula. This produces what are sometimes called **z-scores**.

The formula is given by $z = \frac{x - \mu}{\sigma}$

Where μ is the population mean, σ is the standard deviation and x represents the result that is to be standardised.

Example The lifetime of a particular type of light-bulb has been shown to follow a Normal distribution with mean lifetime of 1000 hours and standard deviation of 125 hours. Three bulbs are found to last 1250, 980 and 1150 hours. Convert these values to standardised normal scores.

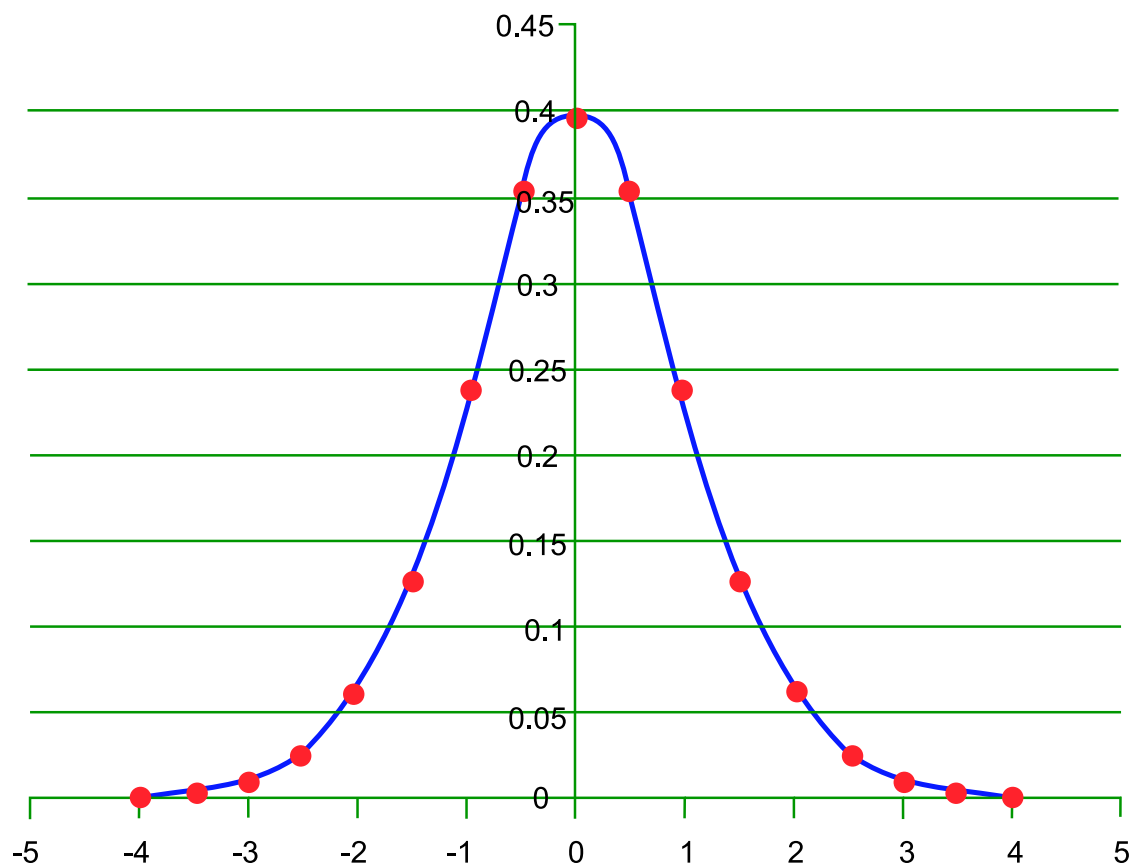
Using the formula $z = \frac{x - \mu}{\sigma}$

1250 converts to $1250 - 1000/125=2$

980 converts to $980 - 1000/125=-0.16$

1150 converts to $1150 - 1000/125=1.2$

This therefore gives equivalencies - in much the same way as temperatures can be converted from Celsius to Fahrenheit. Each x value is equivalent to another z value - the z results simply measure the number of standard deviations away from the mean of the corresponding x result. The important fact is that the converted z scores can be represented by the standardised normal curve.



3.3.4 Properties of Normal Distribution Curves

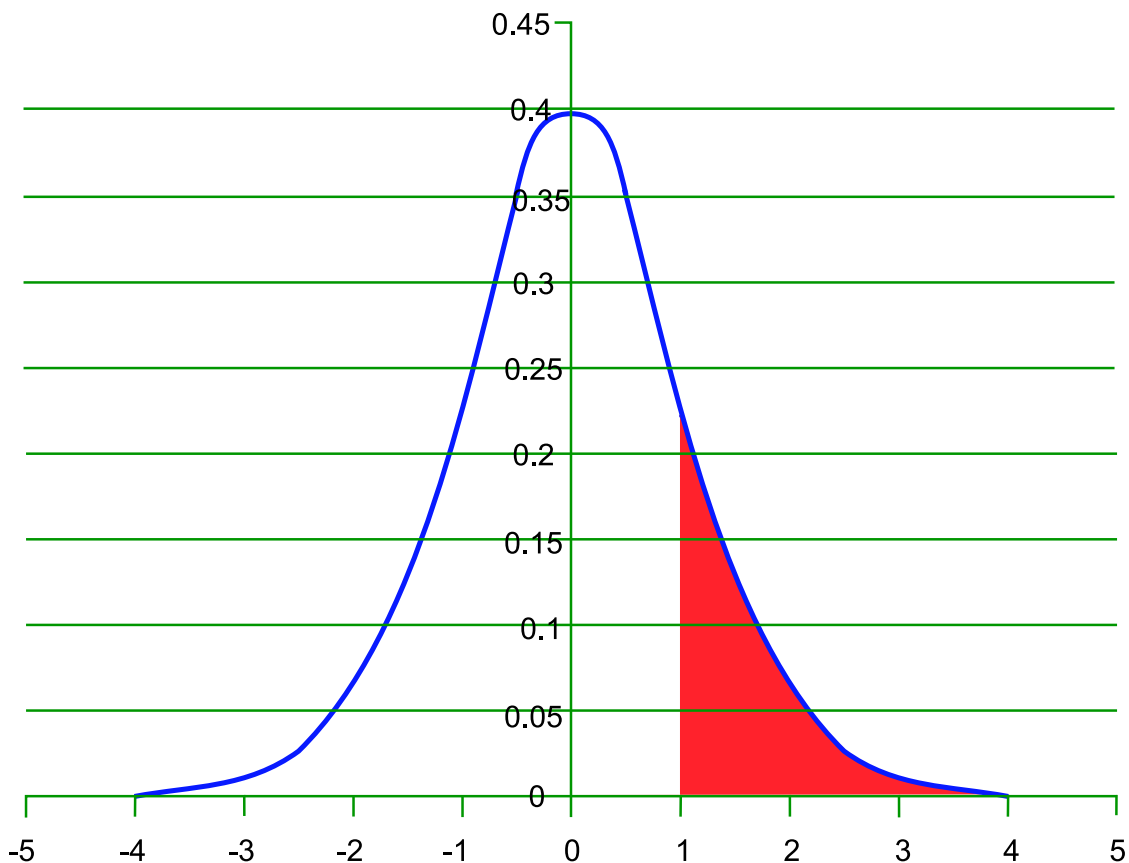
The graph of a Normal distributions has a bell shape with the shape and position being completely determined by the mean, μ , and standard deviation, σ , of the data.

- The curve peaks at the mean.
- The curve is symmetric about the mean.
- Unique to the Normal distribution curve is the property that the mean, median, and mode are the same value.
- The tails of the curve approach the x-axis, but never touch it.
- Although the graph will go on indefinitely, the area under the graph is considered to have a value of 1.
- Because of symmetry, the area of the part of the graph less than the mean is 0.5. The same is true for more than the mean.

3.3.5 Link with Probability

Recall that the Normal distribution graph was first observed by looking at a histogram of results and it was stated then that the size of each "bar" in the histogram was proportional to the probability of that particular outcome occurring. Since measuring size involves an examination of an area, this implies that the area of the bar is equivalent to the probability of that particular outcome happening. This gives one of the most important properties of the Normal distribution, that areas under the curve enable calculations about probabilities to be made. The fact that the total area under the curve is 1 is consistent with saying that the probability of finding values from the very lowest possible z value to the very highest possible z value is 1 (a certainty). It is desirable to calculate the areas between values - this will in turn result in discovering the corresponding probability of an event occurring.

Example For a population of data following the standardised Normal distribution, calculate the probability of finding a result greater than 1



Areas under curves can be found using a mathematical technique called Integration. If you have done any work in this area before you will know that for complicated equations like the one for the standardised Normal curve the process can be lengthy and difficult. Fortunately, statistical tables have been produced to give you the answer without too much hard work. A portion of one such set of tables is shown below.

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170

This extract is taken from tables by J. Murdoch and J. A. Barnes.

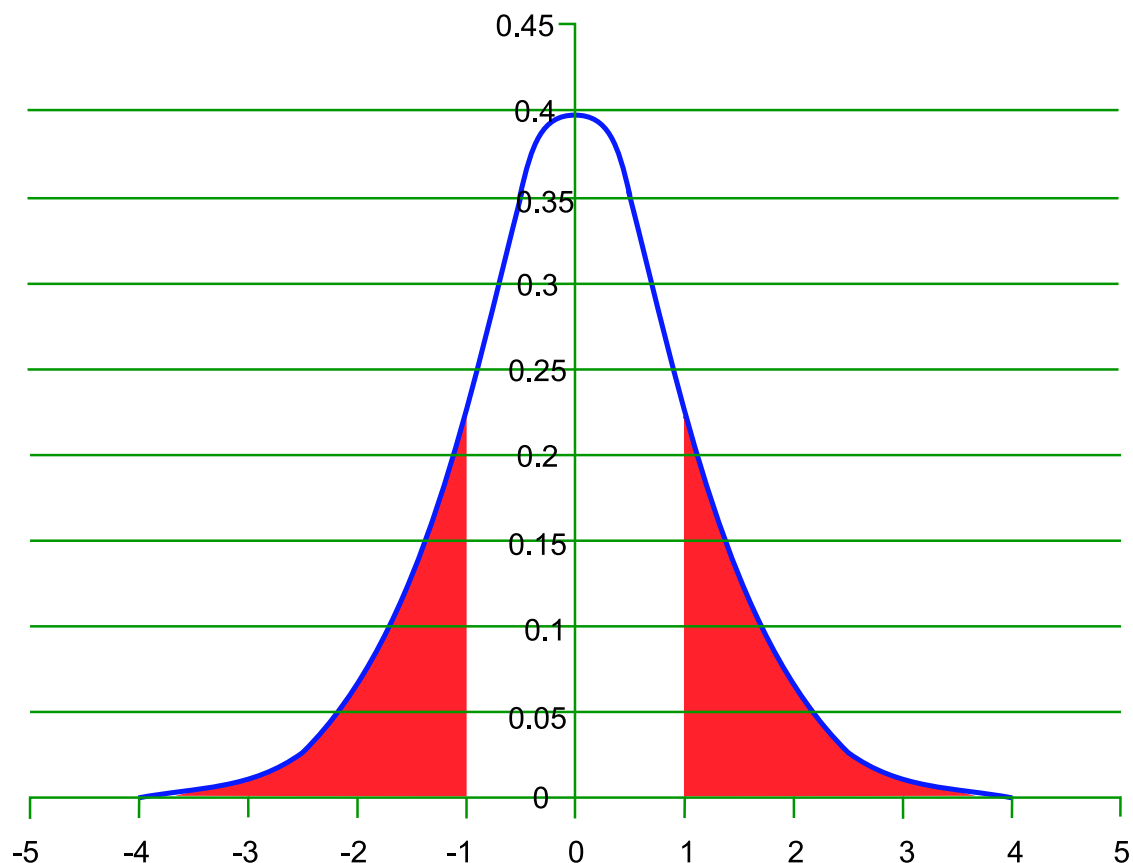
For this example, it is necessary to look up the value 1.00. This is achieved by moving down the first column to the value 1.0 and then moving along to the second column headed .00. The result is clearly 0.1587. Because of the way these tables are compiled, this automatically gives the area to the right of the value 1, as required. So the answer

to the problem is a probability of 0.1587 (or approximately 16%).

Similarly the tables can be used to find the probability of finding a result greater than any other number; if you were asked to find the probability of finding a value greater than, for example, 0.57, the answer would be a probability of 0.2843.

Notice that these particular statistical tables are calculated to give the area to the right of certain values. In other variations, tables give areas between the mean and a certain value. The work carried out in this topic, however, assumes the use of tables like those of Murdoch and Barnes.

Because of the symmetry of the Normal graphs, Murdoch and Barnes tables can be used directly to give the probability of finding results less than a given (negative) value. It can be seen from the graphs below that one is simply the mirror image of the other.

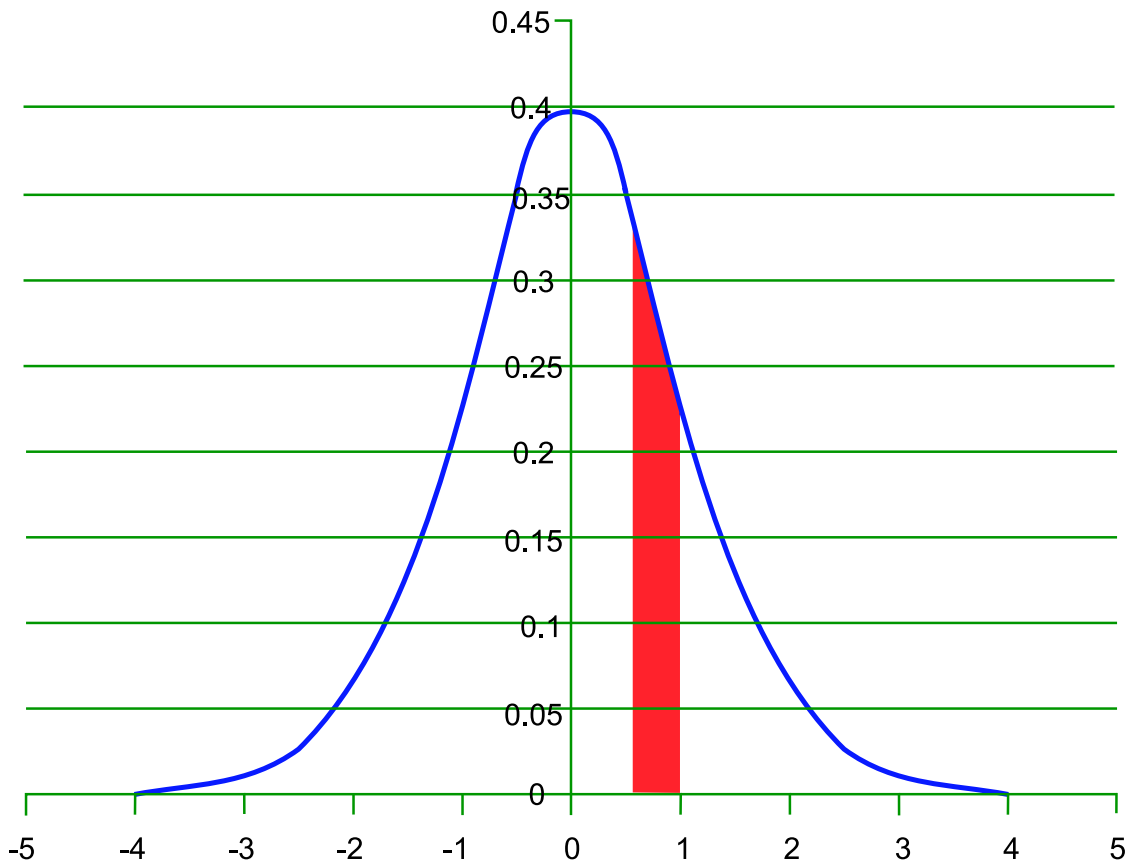


So the probability of finding a result less than -1 is 0.1587, and the probability of finding a value less than -0.57 is 0.2843. Note that probabilities are never negative.

Also, by making more use of symmetry it is possible to obtain the probability of finding a result between ANY two values.

Examples

1. For a population of data following the standardised Normal distribution, calculate the probability of finding a result between 0.57 and 1.



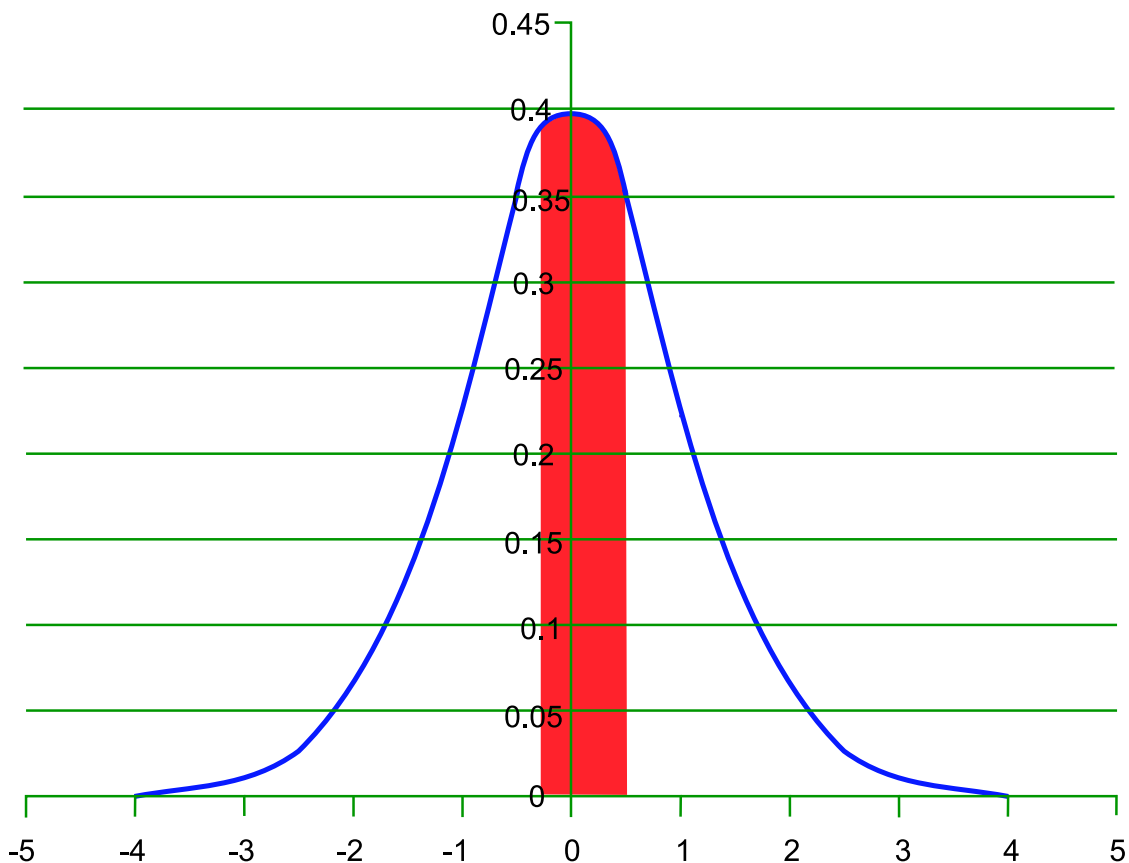
The area to the right of 0.57 is 0.2843.

The area to the right of 1 is 0.1587.

So the area between the two must be $0.2843 - 0.1587 = 0.1256$.

This is therefore the probability of finding a result between 0.57 and 1.

2. For a population of data following the standardised Normal distribution, calculate the probability of finding a result between -0.25 and 0.50.



From tables, the area to the LEFT of -0.25 is 0.4013.

From tables, the area to the RIGHT of 0.50 is 0.3085.

Since the total area under the curve is 1, the area between the two numbers is

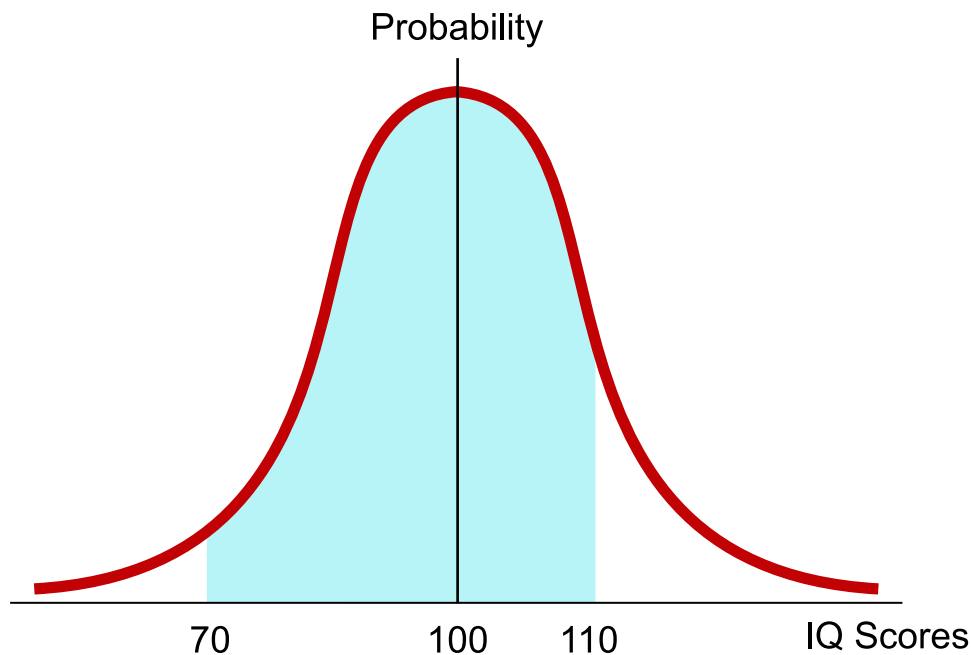
$$1 - (0.4013 + 0.3085) = 0.2902.$$

So the probability of finding a result between -0.25 and 0.50 is 0.2902.

Of course, it is very rare that you will be working with numbers that follow the standardised normal distribution, but the techniques shown here work equally well as long as the appropriate results are converted into z-values.

Example For the earlier example of IQ scores which it has been suggested follow a Normal distribution with mean 100 and standard deviation 15, find the probability that any person chosen at random will have

- a) An IQ greater than 110
- b) An IQ less than 70
- c) An IQ between 70 and 110.



The x values (70 and 110) must first be converted to z values so that the tables can be used.

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

a) $z = 110 - 100/15 = 0.67$

Looking up 0.67 in tables gives 0.2514. This is, therefore, the probability of finding someone with an IQ greater than 110.

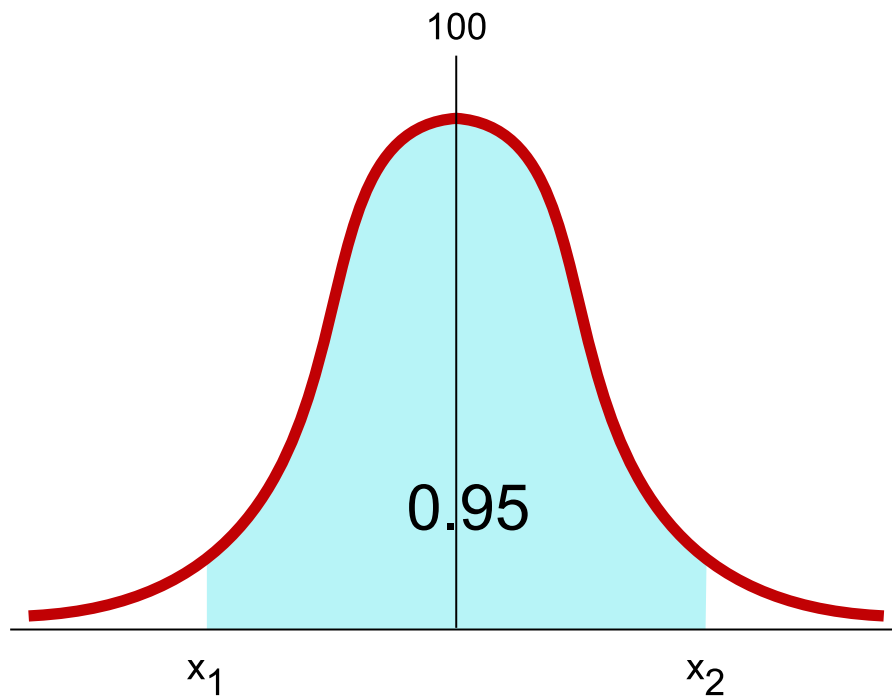
b) $z = 70 - 100/15 = -2.00$

Looking up 2.00 in the tables gives 0.0228 (not shown in the sample tables above). By symmetry this is, therefore, the probability of finding someone with an IQ less than 70 (in other words, approximately 2%).

c) All the work has been done, so the required probability is $1 - (0.2514 + 0.0228) = 0.7258$.

3.3.6 Upper and Lower Bounds

Sometimes it is useful to START with a probability and then work out related z or x values. To do this, the statistical tables can be used in reverse to give upper and lower bounds as to where, say, 95% of all the data will lie. Take the example of the IQ scores just given and use a 95% interval. Now, the area between the two bounded values is known, but it is the corresponding x values that are not.



Since the area to the right of x_2 is 0.025 (this is because the total area under the curve is 1 and by symmetry the remaining 0.05 must be split in two), statistical tables can be used in reverse to find the appropriate z value of the standardised normal distribution that gives a probability of 0.025.

Examination of the tables shows this to be 1.96, but for illustrative purposes, this will be rounded to 2. It was stated earlier that the standardised z distribution measures the number of standard deviations away from the mean so the point x_2 is therefore $100 + 2 \times 15 = 130$.

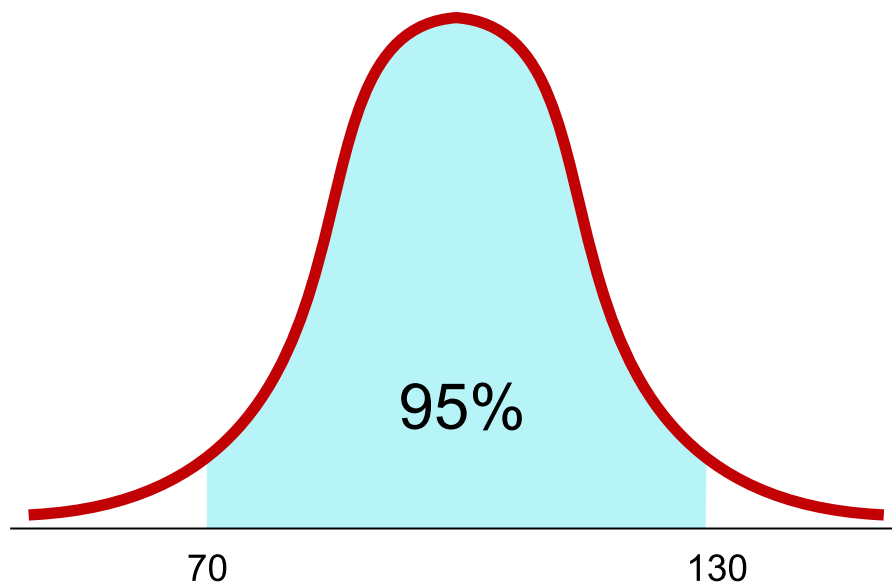
Similarly the value x_1 can be calculated as $100 - 2 \times 15 = 70$

These numbers could also be found by solving the equations

$$z = \frac{x - \mu}{\sigma} = 2$$

So $\frac{x_2 - 100}{15} = 2$ i.e. $x_2 = 130$, and similarly for x_1 .

Thus it is expected that 95% of the population will have IQ values between 70 and 130. The animation below shows the range of values that it would be expected other percentages of the population would lie between.



Notice that the range of values increases as a greater percentage of the population is required.

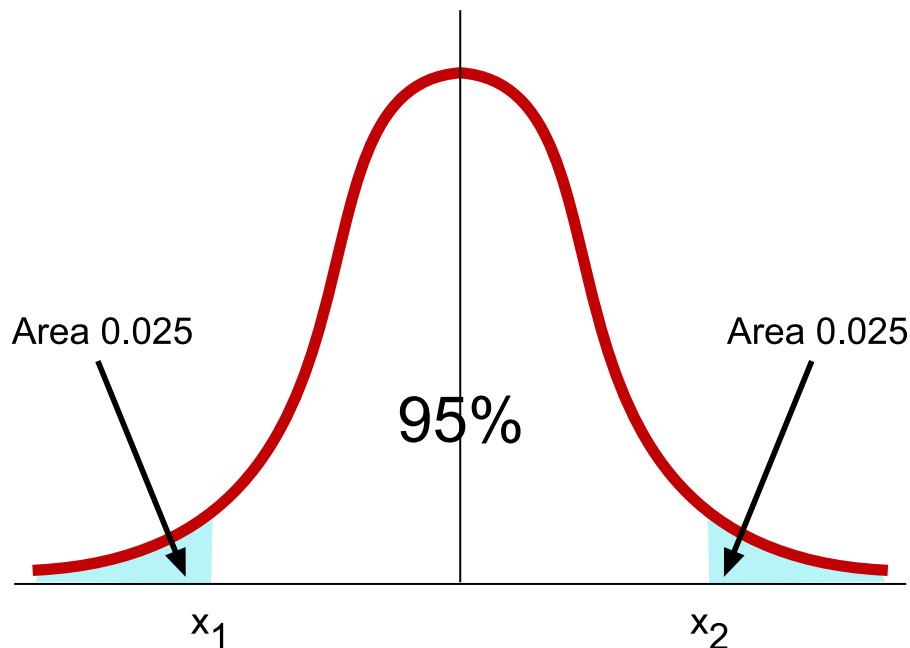
In general, for a Normally distributed data set, an empirical rule states that 68% of the data elements are within one standard deviation of the mean, 95% are within two standard deviations, and 99.7% are within three standard deviations. This rule is often stated simply as 68-95-99.7.

This type of reasoning can be extended to other distributions that do not have the familiar Normal shape. The Russian mathematician Chebyshev (1821-1894) primarily worked on the theory of prime numbers, although his writings covered a wide range of subjects. One of those subjects was probability and he produced a theorem which states that the proportion of any set of data within K standard deviations of the mean is always at least $1 - \frac{1}{K^2}$, where K may be any number greater than 1. Note that this theorem applies to any data set, not only Normally distributed ones.

So for $K=2$, this gives a proportion of $1 - \frac{1}{K^2}$, i.e. $1 - \frac{1}{4} = \frac{3}{4}$. Thus at least 75% of the data must always be within two standard deviations of the mean. It has already been shown that for a Normal distribution the value is 95% (and 75% is clearly less than that).

Similarly, for $K=3$, it can be seen that $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9}$. Thus at least 89% of the data must always be within three standard deviations of the mean.

Example A machine is designed to fill packets with sugar and the mean value over a long period of time has been found to be 1kg. The standard deviation has also been measured and this is given as 0.02kg. What are the upper and lower limits that it would be expected 95% of the bags would lie between? Assume the distribution to be Normal.



Upper limit : $\frac{x_2 - 1}{0.02} = 2$. Thus $x_2 = 1.04$

Lower limit : $\frac{x_1 - 1}{0.02} = -2$. Thus $x_1 = 0.96$

95% of the bags of sugar will lie between 0.96kg and 1.04kg.

Activity

You are given that the mean salary of a UK middle manager is 54.3 thousand pounds, with a standard deviation of 20.1. Assuming a Normal Distribution:



Q3: Calculate the probability that a UK middle manager will earn between 45 and 60 thousand pounds.

Q4: Calculate the salary that only 5% of middle managers will earn more than.

3.4 Summary

- Statistics is about collecting, presenting and characterizing data and assists in data analysis and decision making.
- Statistics is usually about quantitative data. Often, such data is presented in diagrams.
- Basic analysis of data is about the central tendency of data (mean, median, mode), and about the variance of data (variance, standard deviation).
- Random variables are functions assigning numerical values to each simple event of a sample space. We distinguish discrete and continuous random variables.
- The probability distribution of a discrete random variable is a function that gives for each event the probability that the event occurs.

- The expected value $E(x)$ is the mean, the standard deviation the square root of $E[(x - E(x))^2]$.
- Examples of discrete probability distribution are the binomial, geometric, hypergeometric and the Poisson distribution.
- For continuous random variables we have to give the cumulative probability distribution.
- The relative frequency distribution for a population with continuous random variable can be modeled using a density function $\rho(x)$ (usually a smooth curve) such that $\rho(x) \geq 0$ and $\int_{-\infty}^{\infty} \rho(x) dx = 1$
- Examples of continuous distributions are the uniform distribution, normal distribution, gamma distribution, the exponential distribution and the Weibull distribution, of which the normal distribution is the most important.

Glossary

binomial probability distribution

The *binomial probability distribution* is given by the formula

$$p(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, \dots, n,$$

where

- p is the probability of a success in a single trial, and $q = 1 - p$;
- n is the number of trials; and
- x is the number of successes; and
- $\binom{n}{x}$ is the binomial coefficient given by the number $\frac{n!}{x!(n-x)!}$

binomial random variable

A binomial random variable is a discrete random variable with probability distribution the Binomial distribution.

chi-square distribution

The χ^2 (*chi-square*) *probability distribution* plays an important role in statistics. The distribution is a special case of the gamma distribution for $\alpha = \frac{\nu}{2}$ and $\beta=2$ (ν is called the *number of degrees of freedom*).

continuous random variable

A random variable is called continuous if it does not take discrete values.

cumulative distribution of a random variable

The cumulative distribution $F(x)$ of a random variable x is the function $F(x) = P(x' \leq x)$.

density function

If F is the cumulative distribution of a continuous random variable x then the *density* function $\rho(x)$ for x is given by $\rho(x) = \frac{dF}{dx}$ (provided that F is differentiable).

expected value of a continuous random variable

The *expected value* of a continuous random variable x with density function $\rho(x)$ is given by

$$\mu = E(x) = \int_{-\infty}^{\infty} t \rho(t) dt$$

expected value of a discrete random variable

For a discrete random variable x with probability distribution $p(x)$ the *expected value* (or *mean*) is defined as

$$\mu = E(x) = \sum_{\text{all } x} x \cdot p(x)$$

expected value of a function of a random variable

The *expected value* of $g(x)$ is defined as

$$E[g(x)] = \sum_{\text{all } x} g(x) \cdot p(x)$$

exponential density function

The exponential density function is a gamma density function with $\alpha=1$,

$$\rho(x) = \frac{e^{-\frac{x}{\beta}}}{\beta}, \quad x \geq 0$$

with mean $\mu=\beta$ and standard deviation $\sigma=\beta$.

gamma random variable

A gamma random variable is a continuous random variable with density function the gamma distribution.

geometric probability distribution

The geometric probability distribution is a discrete distribution modeling the event of a first success after $x - 1$ failures. The data for the geometric probability distribution are

- $P(x) = pq^{x-1}$, for $x = 1, 2, \dots$,
where x is the number of trials until the first success; and
- $\mu = 1/p$, and
-

$$\sigma = \sqrt{\frac{q}{p^2}}$$

hypergeometric probability distribution

The hypergeometric probability distribution is a discrete distribution that models sampling without replacement.

Poisson probability distribution

The Poisson probability distribution is a discrete probability distribution. It is often used to model frequencies of events.

probability distribution

The *probability distribution* of a random variable is a table, graph, or formula that gives the probability $p(x)$ for each possible value of the random variable x .

random variable

A *random variable* is a function taking numerical values which is defined over a sample space. Such a random variable is called *discrete* if it only takes countably many values.

standard deviation of a continuous random variable

The *standard deviation* is defined as $\sigma = \sqrt{E[(x - \mu)^2]}$.

standard deviation of a discrete random variable

The *variance* of a discrete random variable x with probability distribution $p(x)$ is defined as

$$\sigma^2 = E[(x - \mu)^2]$$

the *standard deviation* is defined as $\sqrt{E[(x - \mu)^2]}$.

uniform random variable

If we select randomly a number in the interval $[a, b]$ then the corresponding random variable x is called a *uniform random variable*.

Answers to questions and activities

3 Probability Distributions

Activity (page 27)

Q1: $p(x) = \frac{e^{-m} m^x}{x!}$ and $m=6$ (in 2 minutes)

Required probability is $p(>2) = [p(0) + p(1) + p(2)]$

Now, $p(0) = \frac{e^{-6} 6^0}{0!} = 0.00248$, $p(1) = \frac{e^{-6} 6^1}{1!} = 0.01487$, $p(2) = \frac{e^{-6} 6^2}{2!} = 0.04462$

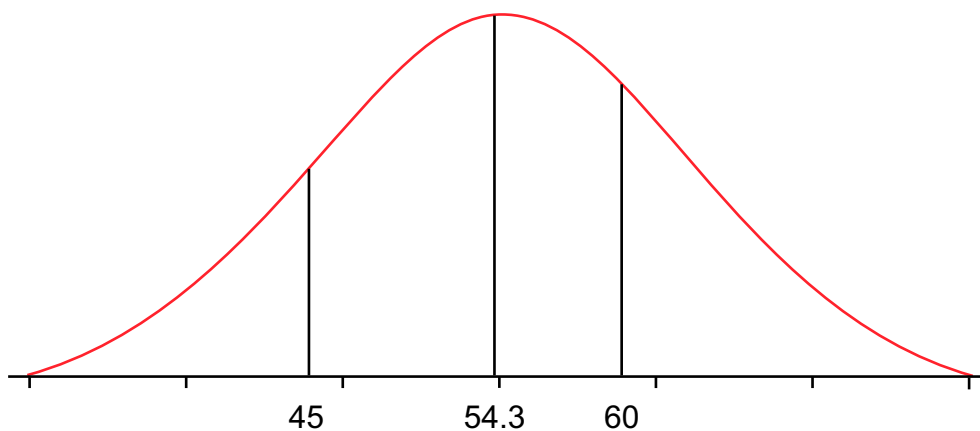
So $p(>2) = 1 - (0.00248 + 0.01487 + 0.04462) = 0.93803$

Q2: Binomial distribution. Probability of success, a , is equal to 0.667. This is the probability of NOT taking school meals regularly. Sample size, n , is equal to 20, so the appropriate formula gives

$$p(x) = \frac{n!}{x!(n-x)!} a^x (1-a)^{n-x} = \frac{20!}{5!15!} 0.667^5 0.333^{15} = 0.0001405$$

Activity (page 40)

Q3: The data follows a Normal distribution as follows:



Area to the right of 60:

$$z = \frac{x - \mu}{\sigma} = \frac{60 - 54.3}{20.1} = 0.28$$

Looking up tables gives a value of 0.3897

Area to the left of 45:

$$z = \frac{x - \mu}{\sigma} = \frac{45 - 54.3}{20.1} = -0.46$$

Looking up tables gives a value of 0.3228

The required area is the "bit in the middle" and this is equal to $1 - (0.3897 + 0.3228)$. The answer is 0.2875.

Q4: Let the point at which the "cut-off" occurs for the top 5% be called t . The area must be 5% (or 0.05) so using the tables in reverse gives a z value of 1.64.

Therefore, $z = \frac{x - \mu}{\sigma}$ becomes $1.64 = \frac{t - 54.3}{20.1} \Rightarrow t = 20.1 \times 1.64 + 54.3 = 87.3$

So only 5% will earn more than 87.3 thousand.