Chapter 5

Random Variables

5.1 Student Learning Objective

This section introduces some important examples of random variables. The distributions of these random variables emerge as mathematical models of real-life settings. In two of the examples the sample space is composed of integers. In the other two examples the sample space is made of continuum of values. For random variables of the latter type one may use the density, which is a type of a histogram, in order to describe the distribution.

By the end of the chapter the student should:

- Identify the Binomial, Poisson, Uniform, and Exponential random variables, relate them to real life situations, and memorize their expectations and variances.
- Relate the plot of the density/probability function and the cumulative probability function to the distribution of a random variable.
- Become familiar with the `R` functions that produce the density/probability of these random variables and their cumulative probabilities.
- Plot the density and the cumulative probability function of a random variable and compute probabilities associated with random variables.

5.2 Discrete Random Variables

In the previous chapter we introduced the notion of a random variable. A random variable corresponds to the outcome of an observation or a measurement prior to the actual making of the measurement. In this context one can talk of all the values that the measurement may potentially obtain. This collection of values is called the *sample space*. To each value in the sample space one may associate the *probability* of obtaining this particular value. Probabilities are like relative frequencies. All probabilities are positive and the sum of the probabilities that are associated with all the values in the sample space is equal to one.

A random variable is defined by the identification of its sample space and the probabilities that are associated with the values in the sample space. For each type of random variable we will identify first the sample space — the values it may obtain — and then describe the probabilities of the values. Examples of situations in which each type of random variable may serve as a model of a measurement will be provided. The `R` system provides functions for the computation of probabilities associated with specific types of random variables. We will use these functions in this and in proceeding chapters in order to carry out computations associated with the random variables and in order to plot their distributions.

The distribution of a random variable, just like the distribution of data, can be characterized using numerical summaries. For the latter we used summaries such as the mean and the sample variance and standard deviation. The mean is used to describe the central location of the distribution and the variance and standard deviation are used to characterize the total spread. Parallel summaries are used for random variable. In the case of a random variable the name *expectation* is used for the central location of the distribution and the *variance* and the *standard deviation* (the square root of the variation) are used to summarize the spread. In all the examples of random variables we will identify the expectation and the variance (and, thereby, also the standard deviation).

Random variables are used as probabilistic models of measurements. Theoretical considerations are used in many cases in order to define random variables and their distribution. A random variable for which the values in the sample space are separated from each other, say the values are integers, is called a *discrete random variable*. In this section we introduce two important integer-valued random variables: The *Binomial* and the *Poisson* random variables. These random variables may emerge as models in contexts where the measurement involves counting the number of occurrences of some phenomena.

Many other models, apart from the Binomial and Poisson, exist for discrete random variables. An example of such model, the Negative-Binomial model, will be considered in Section 5.4. Depending on the specific context that involves measurements with discrete values, one may select the Binomial, the Poisson, or one of these other models to serve as a theoretical approximation of the distribution of the measurement.

5.2.1 The Binomial Random Variable

The Binomial random variable is used in settings in which a trial that has two possible outcomes is repeated several times. Let us designate one of the outcomes as "Success" and the other as "Failure". Assume that the probability of success in each trial is given by some number p that is larger than 0 and smaller than 1. Given a number n of repeats of the trial and given the probability of success, the actual number of trials that will produce "Success" as their outcome is a random variable. We call such random variable *Binomial*. The fact that a random variable X has such a distribution is marked by the expression: "$X \sim \text{Binomial}(n, p)$".

As an example consider tossing 10 coins. Designate "Head" as success and "Tail" as failure. For fair coins the probability of "Head" is $1/2$. Consequently, if X is the total number of "Heads" then $X \sim \text{Binomial}(10, 0.5)$, where $n = 10$ is the number of trials and $p = 0.5$ is the probability of success in each trial.

It may happen that all 10 coins turn up "Tail". In this case X is equal to 0. It may also be the case that one of the coins turns up "Head" and the others

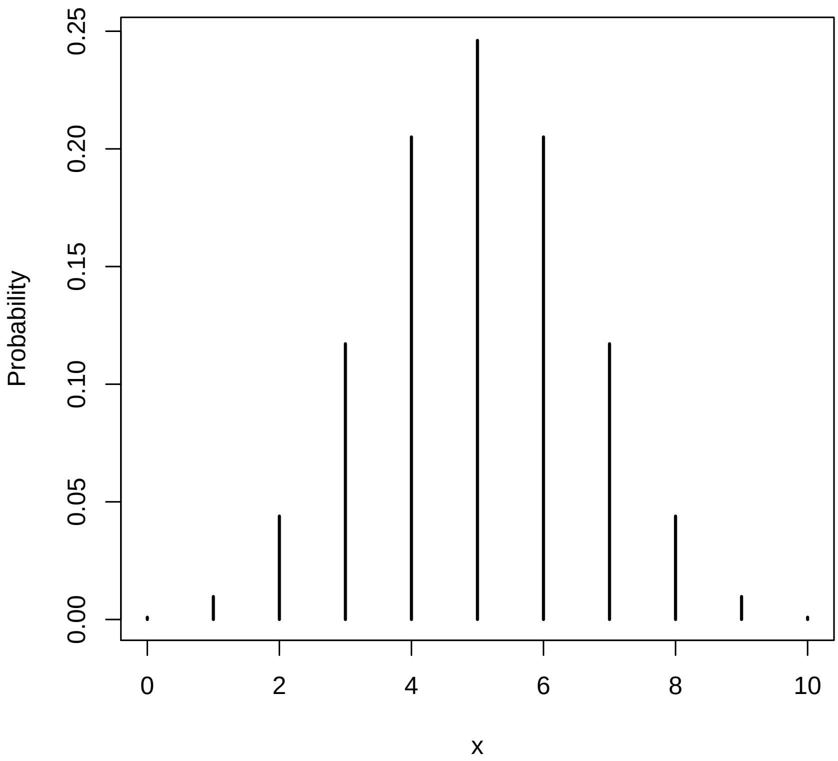

Figure 5.1: The Binomial(10,0.5) Distribution

turn up "Tail". The random variable X will obtain the value 1 in such a case. Likewise, for any integer between 0 and 10 it may be the case that the number of "Heads" that turn up is equal to that integer with the other coins turning up "Tail". Hence, the sample space of X is the set of integers $\{0, 1, 2, \ldots, 10\}$. The probability of each outcome may be computed by an appropriate mathematical formula that will not be discussed here[1].

The probabilities of the various possible values of a Binomial random variable may be computed with the aid of the `R` function "`dbinom`" (that uses the mathematical formula for the computation). The input to this function is a sequence of values, the value of n, and the value of p. The output is the sequence of probabilities associated with each of the values in the first input.

For example, let us use the function in order to compute the probability that the given Binomial obtains an odd value. A sequence that contains the odd values in the Binomial sample space can be created with the expression "`c(1,3,5,7,9)`". This sequence can serve as the input in the first argument of the function "`dbinom`". The other arguments are "`10`" and "`0.5`", respectively:

[1] If $X \sim \text{Binomial}(n, p)$ then $\text{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x = 0, 1, \ldots, n$.

```
> dbinom(c(1,3,5,7,9),10,0.5)
[1] 0.009765625 0.117187500 0.246093750 0.117187500 0.009765625
```

Observe that the output of the function is a sequence of the same length as the first argument. This output contains the Binomial probabilities of the values in the first argument. In order to obtain the probability of the event {X is odd} we should sum up these probabilities, which we can do by applying the function "`sum`" to the output of the function that computes the Binomial probabilities:

```
> sum(dbinom(c(1,3,5,7,9),10,0.5))
[1] 0.5
```

Observe that the probability of obtaining an odd value in this specific case is equal to one half.

Another example is to compute all the probabilities of all the potential values of a Binomial$(10, 0.5)$ random variable:

```
> x <- 0:10
> dbinom(x,10,0.5)
 [1] 0.0009765625 0.0097656250 0.0439453125 0.1171875000
 [5] 0.2050781250 0.2460937500 0.2050781250 0.1171875000
 [9] 0.0439453125 0.0097656250 0.0009765625
```

The expression "`start.value:end.value`" produces a sequence of numbers that initiate with the number "`start.value`" and proceeds in jumps of size one until reaching the number "`end.value`". In this example, "`0:10`" produces the sequence of integers between 0 and 10, which is the sample space of the current Binomial example. Entering this sequence as the first argument to the function "`dbinom`" produces the probabilities of all the values in the sample space.

One may display the distribution of a discrete random variable with a bar plot similar to the one used to describe the distribution of data. In this plot a vertical bar representing the probability is placed above each value of the sample space. The hight of the bar is equal to the probability. A bar plot of the Binomial$(10, 0.5)$ distribution is provided in Figure 5.1.

Another useful function is "`pbinom`", which produces the cumulative probability of the Binomial:

```
> pbinom(x,10,0.5)
 [1] 0.0009765625 0.0107421875 0.0546875000 0.1718750000
 [5] 0.3769531250 0.6230468750 0.8281250000 0.9453125000
 [9] 0.9892578125 0.9990234375 1.0000000000
> cumsum(dbinom(x,10,0.5))
 [1] 0.0009765625 0.0107421875 0.0546875000 0.1718750000
 [5] 0.3769531250 0.6230468750 0.8281250000 0.9453125000
 [9] 0.9892578125 0.9990234375 1.0000000000
```

The output of the function "`pbinom`" is the cumulative probability $\mathrm{P}(X \leq x)$ that the random variable is less than or equal to the input value. Observe that this cumulative probability is obtained by summing all the probabilities associated with values that are less than or equal to the input value. Specifically, the cumulative probability at $x = 3$ is obtained by the summation of the probabilities at $x = 0$, $x = 1$, $x = 2$, and $x = 3$:

$$\mathrm{P}(X \leq 3) = 0.0009765625 + 0.009765625 + 0.0439453125 + 0.1171875 = 0.171875$$

The numbers in the sum are the first 4 values from the output of the function "`dbinom(x,10,0.5)`", which computes the probabilities of the values of the sample space.

In principle, the expectation of the Binomial random variable, like the expectation of any other (discrete) random variable is obtained from the application of the general formulae:

$$\mathrm{E}(X) = \sum_x \big(x \times \mathrm{P}(X = x)\big) \;, \quad \mathrm{V}ar(X) = \sum_x \big((x - \mathrm{E}(X))^2 \times \mathrm{P}(x)\big) \;.$$

However, in the specific case of the Binomial random variable, in which the probability $\mathrm{P}(X = x)$ obeys the specific mathematical formula of the Binomial distribution, the expectation and the variance reduce to the specific formulae:

$$\mathrm{E}(X) = np \;, \quad \mathrm{V}ar(X) = np(1 - p) \;.$$

Hence, the expectation is the product of the number of trials n with the probability of success in each trial p. In the variance the number of trials is multiplied by the product of a probability of success (p) with the probability of a failure $(1 - p)$.

As illustration, let us compute for the given example the expectation and the variance according to the general formulae for the computation of the expectation and variance in random variables and compare the outcome to the specific formulae for the expectation and variance in the Binomial distribution:

```
> X.val <- 0:10
> P.val <- dbinom(X.val,10,0.5)
> EX <- sum(X.val*P.val)
> EX
[1] 5
> sum((X.val-EX)^2*P.val)
[1] 2.5
```

This agrees with the specific formulae for Binomial variables, since $10 \times 0.5 = 5$ and $10 \times 0.5 \times (1 - 0.5) = 2.5$.

Recall that the general formula for the computation of the expectation calls for the multiplication of each value in the sample space with the probability of that value, followed by the summation of all the products. The object "`X.val`" contains all the values of the random variable and the object "`P.val`" contains the probabilities of these values. Hence, the expression "`X.val*P.val`" produces the product of each value of the random variable times the probability of that value. Summation of these products with the function "`sum`" gives the expectation, which is saved in an object that is called "`EX`".

The general formula for the computation of the variance of a random variable involves the product of the squared deviation associated with each value with the probability of that value, followed by the summation of all products. The expression "`(X.val-EX)^2`" produces the sequence of squared deviations from the expectation for all the values of the random variable. Summation of the product of these squared deviations with the probabilities of the values (the outcome of "`(X.val-EX)^2*P.val`") gives the variance.

When the value of p changes (without changing the number of trials n) then the probabilities that are assigned to each of the values of the sample space of the Binomial random variable change, but the sample space itself does not. For example, consider rolling a die 10 times and counting the number of times that the face 3 was obtained. Having the face 3 turning up is a "Success". The probability p of a success in this example is 1/6, since the given face is one out of 6 equally likely faces. The resulting random variable that counts the total number of success in 10 trials has a $\text{Binomial}(10, 1/6)$ distribution. The sample space is yet again equal to the set of integers $\{0, 1, \ldots, 10\}$. However, the probabilities of values are different. These probabilities can again be computes with the aid of the function "`dbinom`":

```
> dbinom(x,10,1/6)
 [1] 1.615056e-01 3.230112e-01 2.907100e-01 1.550454e-01
 [5] 5.426588e-02 1.302381e-02 2.170635e-03 2.480726e-04
 [9] 1.860544e-05 8.269086e-07 1.653817e-08
```

In this case smaller values of the random variable are assigned higher probabilities and larger values are assigned lower probabilities..

In Figure 5.2 the probabilities for $\text{Binomial}(10, 1/6)$, the $\text{Binomial}(10, 1/2)$, and the $\text{Binomial}(10, 0.6)$ distributions are plotted side by side. In all these 3 distributions the sample space is the same, the integers between 0 and 10. However, the probabilities of the different values differ. (Note that all bars should be placed on top of the integers. For clarity of the presentation, the bars associated with the $\text{Binomial}(10, 1/6)$ are shifted slightly to the left and the bars associated with the $\text{Binomial}(10, 0.6)$ are shifted slightly to the right.)

The expectation of the $\text{Binomial}(10, 0.5)$ distribution is equal to $10 \times 0.5 = 5$. Compare this to the expectation of the $\text{Binomial}(10, 1/6)$ distribution, which is $10 \times (1/6) = 1.666667$ and to the expectation of the $\text{Binomial}(10, 0.6)$ distribution which equals $10 \times 0.6 = 6$.

The variance of the $\text{Binomial}(10, 0.5)$ distribution is $10 \times 0.5 \times 0.5 = 2.5$. The variance when $p = 1/6$ is $10 \times (1/6) \times (5/6) = 1.388889$ and the variance when $p = 0.6$ is $10 \times 0.6 \times 0.4 = 2.4$.

Example 5.1. *As an application of the Binomial distribution consider a pre-election poll. A candidate is running for office and is interested in knowing the percentage of support in the general population in its candidacy. Denote the probability of support by p. In order to estimate the percentage a sample of size 300 is selected from the population. Let X be the count of supporters in the sample. A natural model for the distribution of X is the* $\text{Binomial}(300, p)$ *distribution, since each subject in the sample may be a supporter ("Success") or may not be a supporter ("Failure"). The probability that a subject supports the candidate is p and there are $n = 300$ subjects in the sample.*

Example 5.2. *As another example consider the procedure for quality control that is described in Discussion Forum of Chapter 4. According to the procedure 20 items are tested and the number of faulty items is recorded. If p is the probability that an item is identified as faulty then the distribution of the total number of faulty items may be modeled by the* $\text{Binomial}(20, p)$ *distribution.*

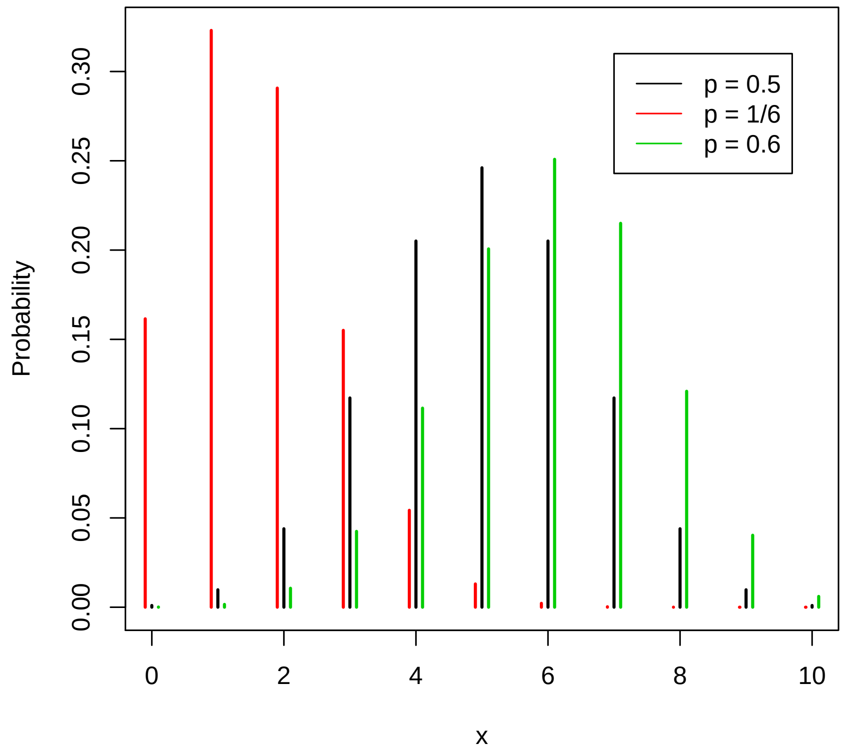

Figure 5.2: The Binomial Distribution for Various Probability of "Success" p

In both examples one may be interested in making statements on the probability p based on the sample. Statistical inference relates the actual count obtained in the sample to the theoretical Binomial distribution in order to make such statements.

5.2.2 The Poisson Random Variable

The *Poisson* distribution is used as an approximation of the total number of occurrences of rare events. Consider, for example, the Binomial setting that involves n trials with p as the probability of success of each trial. Then, if p is small but n is large then the number of successes X has, approximately, the Poisson distribution.

The sample space of the Poisson random variable is the unbounded collection of integers: $\{0, 1, 2, \ldots\}$. Any integer value is assigned a positive probability. Hence, the Poisson random variable is a convenient model when the maximal number of occurrences of the events in a-priori unknown or is very large. For example, one may use the Poisson distribution to model the number of phone calls that enter a switchboard in a given interval of time or the number of

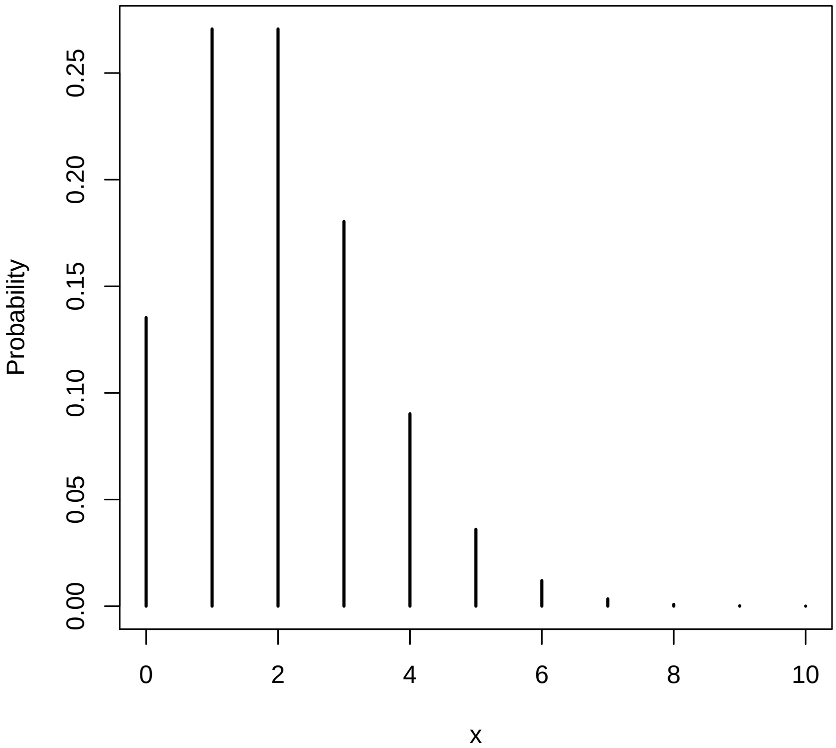

Figure 5.3: The Poisson(2) Distribution

malfunctioning components in a shipment of some product.

The Binomial distribution was specified by the number of trials n and probability of success in each trial p. The Poisson distribution is specified by its expectation, which we denote by λ. The expression "$X \sim \text{Poisson}(\lambda)$" states that the random variable X has a Poisson distribution[2] with expectation $\text{E}(X) = \lambda$. The function "`dpois`" computes the probability, according to the Poisson distribution, of values that are entered as the first argument to the function. The expectation of the distribution is entered in the second argument. The function "`ppois`" computes the cumulative probability. Consequently, we can compute the probabilities and the cumulative probabilities of the values between 0 and 10 for the Poisson(2) distribution via:

```
> x <- 0:10
> dpois(x,2)
 [1] 1.353353e-01 2.706706e-01 2.706706e-01 1.804470e-01
 [5] 9.022352e-02 3.608941e-02 1.202980e-02 3.437087e-03
 [9] 8.592716e-04 1.909493e-04 3.818985e-05
```

[2] If $X \sim \text{Poisson}(\lambda)$ then $\text{P}(X = x) = e^{-\lambda}\lambda^x/x!$, for $x = 0, 1, 2, \ldots$.

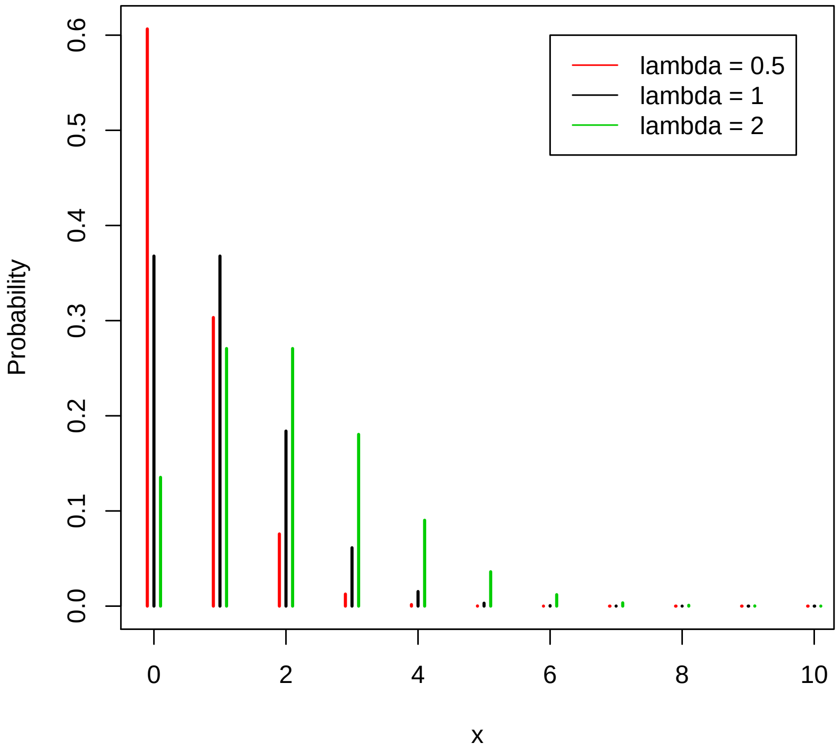

Figure 5.4: The Poisson Distribution for Various Values of λ

```
> ppois(x,2)
 [1] 0.1353353 0.4060058 0.6766764 0.8571235 0.9473470 0.9834364
 [7] 0.9954662 0.9989033 0.9997626 0.9999535 0.9999917
```

The probability function of the Poisson distribution with $\lambda = 2$, in the range between 0 and 10, is plotted in Figure 5.3. Observe that in this example probabilities of the values 8 and beyond are very small. As a matter of fact, the cumulative probability at $x = 7$ (the 8th value in the output of "`ppois(x,2)`") is approximately 0.999, out of the total cumulative probability of 1.000, leaving a total probability of about 0.001 to be distributed among all the values larger than 7.

Let us compute the expectation of the given Poisson distribution:

```
> X.val <- 0:10
> P.val <- dpois(X.val,2)
> sum(X.val*P.val)
[1] 1.999907
```

Observe that the outcome is almost, but not quite, equal to 2.00, which is the actual value of the expectation. The reason for the inaccuracy is the fact that we have based the computation in R on the first 11 values of the distribution only, instead of the infinite sequence of values. A more accurate result may be obtained by the consideration of the first 101 values:

```
> X.val <- 0:100
> P.val <- dpois(X.val,2)
> EX <- sum(X.val*P.val)
> EX
[1] 2
> sum((X.val-EX)^2*P.val)
[1] 2
```

In the last expression we have computed the variance of the Poisson distribution and obtained that it is equal to the expectation. This results can be validated mathematically. For the Poisson distribution it is always the case that the variance is equal to the expectation, namely to λ:

$$\mathrm{E}(X) = \mathrm{V}ar(X) = \lambda \ .$$

In Figure 5.4 you may find the probabilities of the Poisson distribution for $\lambda = 0.5$, $\lambda = 1$ and $\lambda = 2$. Notice once more that the sample space is the same for all the Poisson distributions. What varies when we change the value of λ are the probabilities. Observe that as λ increases then probability of larger values increases as well.

Example 5.3. *A radio active element decays by the release of subatomic particles and energy. The decay activity is measured in terms of the number of decays per second in a unit mass. A typical model for the distribution of the number of decays is the Poisson distribution. Observe that the number of decays in a second is a integer and, in principle, it may obtain any integer value larger or equal to zero. The event of a radio active decay of an atom is a relatively rare event. Therefore, the Poisson model is likely to fit this phenomena*[3].

Example 5.4. *Consider an overhead power line suspended between two utility poles. During rain, drops of water may hit the power line. The total number of drops that hit the line in a one minute period may be modeled by a Poisson random variable.*

5.3 Continuous Random Variable

Many types of measurements, such as height, weight, angle, temperature, etc., may in principle have a continuum of possible values. Continuous random variables are used to model uncertainty regarding future values of such measurements.

The main difference between discrete random variables, which is the type we examined thus far, and continuous random variable, that are added now to the list, is in the sample space, i.e., the collection of possible outcomes. The former

[3]The number of decays may also be considered in the Binomial(n, p) setting. The number n is the total number of atoms in the unit mass and p is the probability that an atom decays within the given second. However, since n is very large and p is very small we get that the Poisson distribution is an appropriate model for the count.

type is used when the possible outcomes are separated from each other as the integers are. The latter type is used when the possible outcomes are the entire line of real numbers or when they form an interval (possibly an open ended one) of real numbers.

The difference between the two types of sample spaces implies differences in the way the distribution of the random variables is being described. For discrete random variables one may list the probability associated with each value in the sample space using a table, a formula, or a bar plot. For continuous random variables, on the other hand, probabilities are assigned to intervals of values, and not to specific values. Thence, densities are used in order to display the distribution.

Densities are similar to histograms, with areas under the plot corresponding to probabilities. We will provide a more detailed description of densities as we discuss the different examples of continuous random variables.

In continuous random variables integration replaces summation and the density replaces the probability in the computation of quantities such as the probability of an event, the expectation, and the variance.

Hence, if the expectation of a discrete random variable is given in the formula $\mathrm{E}(X) = \sum_x \big(x \times \mathrm{P}(x)\big)$, which involves the summation over all values of the product between the value and the probability of the value, then for continuous random variable the definition becomes:

$$\mathrm{E}(X) = \int \big(x \times f(x)\big)dx \;,$$

where $f(x)$ is the density of X at the value x. Therefore, in the expectation of a continuous random variable one multiplies the value by the density at the value. This product is then integrated over the sample space.

Likewise, the formula $\mathrm{Var}(X) = \sum_x \big((x - \mathrm{E}(X))^2 \times \mathrm{P}(x)\big)$ for the variance is replaced by:

$$\mathrm{Var}(X) = \int \big((x - \mathrm{E}(X))^2 \times f(x)\big)dx \;.$$

Nonetheless, the intuitive interpretation of the expectation as the central value of the distribution that identifies the location and the interpretation of the standard deviation (the square root of the variance) as the summary of the total spread of the distribution is still valid.

In this section we will describe two types of continuous random variables: Uniform and Exponential. In the next chapter another example – the Normal distribution – will be introduced.

5.3.1 The Uniform Random Variable

The Uniform distribution is used in order to model measurements that may have values in a given interval, with all values in this interval equally likely to occur.

For example, consider a random variable X with the Uniform distribution over the interval $[3, 7]$, denoted by "$X \sim \mathrm{Uniform}(3, 7)$". The density function at given values may be computed with the aid of the function "`dunif`". For instance let us compute the density of the $\mathrm{Uniform}(3, 7)$ distribution over the integers $\{0, 1, \ldots, 10\}$:

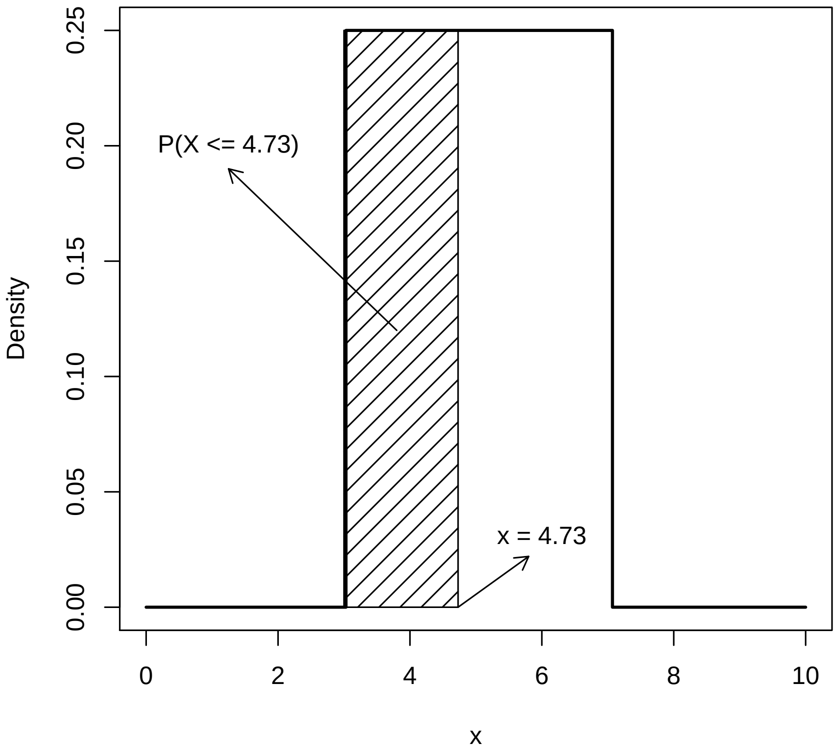

Figure 5.5: The Uniform(3,7) Distribution

```
> dunif(0:10,3,7)
 [1] 0.00 0.00 0.00 0.25 0.25 0.25 0.25 0.25 0.00 0.00 0.00
```

Notice that for the values 0, 1, and 2, and the values 8, 9 and 10 that are outside of the interval the density is equal to zero, indicating that such values cannot occur in the given distribution. The values of the density at integers inside the interval are positive and equal to each other. The density is not restricted to integer values. For example, at the point 4.73 we get that the density is positive and of the same height:

```
> dunif(4.73,3,7)
[1] 0.25
```

A plot of the Uniform$(3,7)$ density is given in Figure 5.5 in the form of a solid line. Observe that the density is positive over the interval $[3,7]$ where its height is 1/4. Area under the curve in the density corresponds to probability. Indeed, the fact that the total probability is one is reflected in the total area under the curve being equal to 1. Over the interval $[3,7]$ the density forms a rectangle. The base of the rectangle is the length of the interval $7-3=4$. The height of the rectangle is thus equal to 1/4 in order to produce a total area of $4 \times (1/4) = 1$.

The function "punif" computes the cumulative probability of the uniform distribution. The probability $\mathrm{P}(X \leq 4.73)$, for $X \sim \mathrm{Uniform}(3, 7)$, is given by:

```
> punif(4.73,3,7)
[1] 0.4325
```

This probability corresponds to the marked area to the left of the point $x = 4.73$ in Figure 5.5. This area of the marked rectangle is equal to the length of the base 4.73 - 3 = 1.73, times the height of the rectangle 1/(7-3) = 1/4. Indeed:

```
> (4.73-3)/(7-3)
[1] 0.4325
```

is the area of the marked rectangle and is equal to the probability.

Let us use R in order to plot the density and the cumulative probability functions of the Uniform distribution. We produce first a large number of points in the region we want to plot. The points are produced with aid of the function "seq". The output of this function is a sequence with equally spaced values. The starting value of the sequence is the first argument in the input of the function and the last value is the second argument in the input. The argument "length=1000" sets the length of the sequence, 1,000 values in this case:

```
> x <- seq(0,10,length=1000)
> den <- dunif(x,3,7)
> plot(x,den)
```

The object "den" is a sequence of length 1,000 that contains the density of the $\mathrm{Uniform}(3, 7)$ evaluated over the values of "x". When we apply the function "plot" to the two sequences we get a scatter plot of the 1,000 points that is presented in the upper panel of Figure 5.6.

A scatter plot is a plot of points. Each point in the scatter plot is identify by its horizontal location on the plot (its "x" value) and by its vertical location on the plot (its y value). The horizontal value of each point in the plot is determined by the first argument to the function "plot" and the vertical value is determined by the second argument. For example, the first value in the sequence "x" is 0. The value of the Uniform density at this point is 0. Hence, the first value of the sequence "den" is also 0. A point that corresponds to these values is produced in the plot. The horizontal value of the point is 0 and the vertical value is 0. In a similar way the other 999 points are plotted. The last point to be plotted has a horizontal value of 10 and a vertical value of 0.

The number of points that are plotted is large and they overlap each other in the graph and thus produce an impression of a continuum. In order to obtain nicer looking plots we may choose to connect the points to each other with segments and use smaller points. This may be achieved by the addition of the argument "type="l"", with the letter "l" for line, to the plotting function:

```
> plot(x,den,type="l")
```

The output of the function is presented in the second panel of Figure 5.6. In the last panel the cumulative probability of the $\mathrm{Uniform}(3, 7)$ is presented. This function is produced by the code:

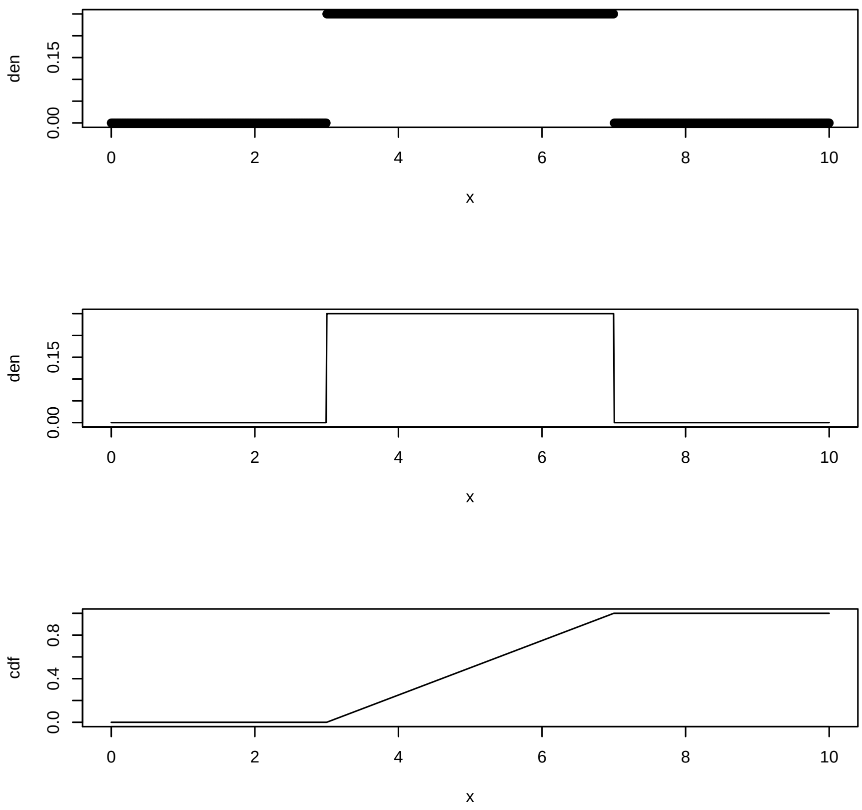

Figure 5.6: The Density and Cumulative Probability of Uniform(3,7)

```
> cdf <- punif(x,3,7)
> plot(x,cdf,type="l")
```

One can think of the density of the Uniform as an histogram[4]. The expectation of a Uniform random variable is the middle point of it's histogram. Hence, if $X \sim \mathrm{Uniform}(a,b)$ then:

$$\mathrm{E}(X) = \frac{a+b}{2} \ .$$

For the $X \sim \mathrm{Uniform}(3,7)$ distribution the expectation is $\mathrm{E}(X) = (3+7)/2 = 5$. Observe that 5 is the center of the Uniform density in Plot 5.5.

It can be shown that the variance of the $\mathrm{Uniform}(a,b)$ is equal to

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12} \ ,$$

with the standard deviation being the square root of this value. Specifically, for $X \sim \mathrm{Uniform}(3,7)$ we get that $\mathrm{Var}(X) = (7-3)^2/12 = 1.333333$. The standard deviation is equal to $\sqrt{1.333333} = 1.154701$.

[4]If $X \sim \mathrm{Uniform}(a,b)$ then the density is $f(x) = 1/(b-a)$, for $a \leq x \leq b$, and it is equal to 0 for other values of x.

Example 5.5. *In Example 5.4 we considered rain drops that hit an overhead power line suspended between two utility poles. The* **number** *of drops that hit the line can be modeled using the Poisson distribution. The* **position** *between the two poles where a rain drop hits the line can be modeled by the Uniform distribution. The rain drop can hit any position between the two utility poles. Hitting one position along the line is as likely as hitting any other position.*

Example 5.6. *Meiosis is the process in which a diploid cell that contains two copies of the genetic material produces an haploid cell with only one copy (sperms or eggs, depending on the sex). The resulting molecule of genetic material is linear molecule (chromosome) that is composed of consecutive segments: a segment that originated from one of the two copies followed by a segment from the other copy and vice versa. The border points between segments are called points of crossover. The Haldane model for crossovers states that the position of a crossover between two given loci on the chromosome corresponds to the Uniform distribution and the total number of crossovers between these two loci corresponds to the Poisson distribution.*

5.3.2 The Exponential Random Variable

The Exponential distribution is frequently used to model times between events. For example, times between incoming phone calls, the time until a component becomes malfunction, etc. We denote the Exponential distribution via "$X \sim \mathrm{Exponential}(\lambda)$", where λ is a parameter that characterizes the distribution and is called the *rate* of the distribution. The overlap between the parameter used to characterize the Exponential distribution and the one used for the Poisson distribution is deliberate. The two distributions are tightly interconnected. As a matter of fact, it can be shown that if the distribution between occurrences of a phenomena has the Exponential distribution with rate λ then the total number of the occurrences of the phenomena within a unit interval of time has a $\mathrm{Poisson}(\lambda)$ distribution.

The sample space of an Exponential random variable contains all non-negative numbers. Consider, for example, $X \sim \mathrm{Exponential}(0.5)$. The density of the distribution in the range between 0 and 10 is presented in Figure 5.7. Observe that in the Exponential distribution smaller values are more likely to occur in comparison to larger values. This is indicated by the density being larger at the vicinity of 0. The density of the exponential distribution given in the plot is positive, but hardly so, for values larger than 10.

The density of the Exponential distribution can be computed with the aid of the function "`dexp`"[5]. The cumulative probability can be computed with the function "`pexp`". For illustration, assume $X \sim \mathrm{Exponential}(0.5)$. Say one is interested in the computation of the probability $\mathrm{P}(2 < X \leq 6)$ that the random variable obtains a value that belongs to the interval $(2, 6]$. The required probability is indicated as the marked area in Figure 5.7. This area can be computed as the difference between the probability $\mathrm{P}(X \leq 6)$, the area to the left of 6, and the probability $\mathrm{P}(X \leq 2)$, the area to the left of 2:

```
> pexp(6,0.5) - pexp(2,0.5)
[1] 0.3180924
```

[5]If $X \sim \mathrm{Exponential}(\lambda)$ then the density is $f(x) = \lambda e^{-\lambda x}$, for $0 \leq x$, and it is equal to 0 for $x < 0$.

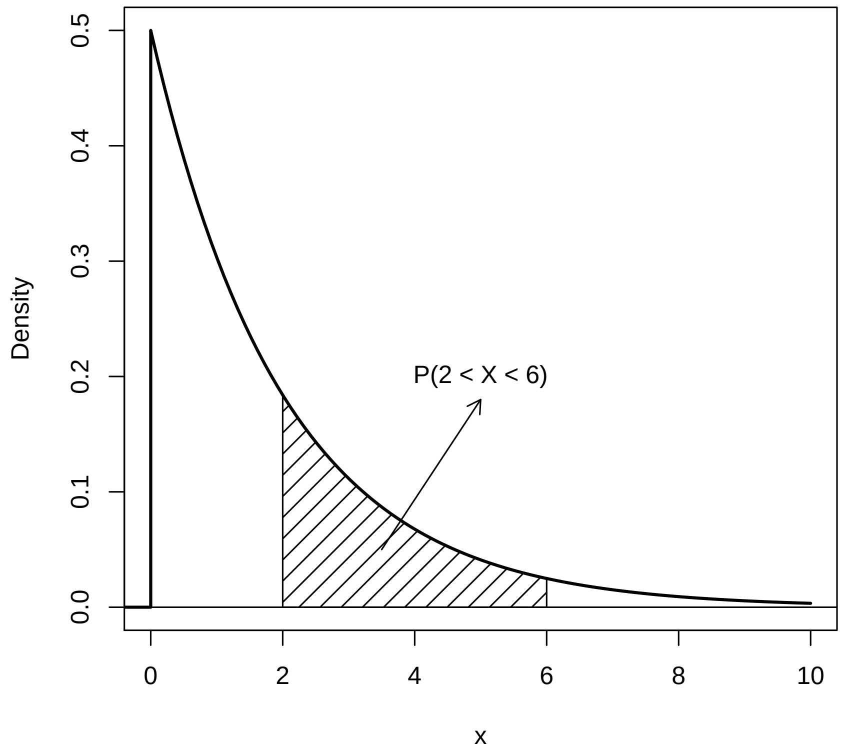

Figure 5.7: The Exponential(0.5) Distribution

The difference is the probability of belonging to the interval, namely the area marked in the plot.

The expectation of X, when $X \sim \text{Exponential}(\lambda)$, is given by the equation:

$$\mathrm{E}(X) = 1/\lambda \ ,$$

and the variance is given by:

$$\mathrm{V}ar(X) = 1/\lambda^2 \ .$$

The standard deviation is the square root of the variance, namely $1/\lambda$. Observe that the larger is the rate the smaller are the expectation and the standard deviation.

In Figure 5.8 the densities of the Exponential distribution are plotted for $\lambda = 0.5$, $\lambda = 1$, and $\lambda = 2$. Notice that with the increase in the value of the parameter then the values of the random variable tends to become smaller. This inverse relation makes sense in connection to the Poisson distribution. Recall that the Poisson distribution corresponds to the total number of occurrences in a unit interval of time when the time between occurrences has an Exponential

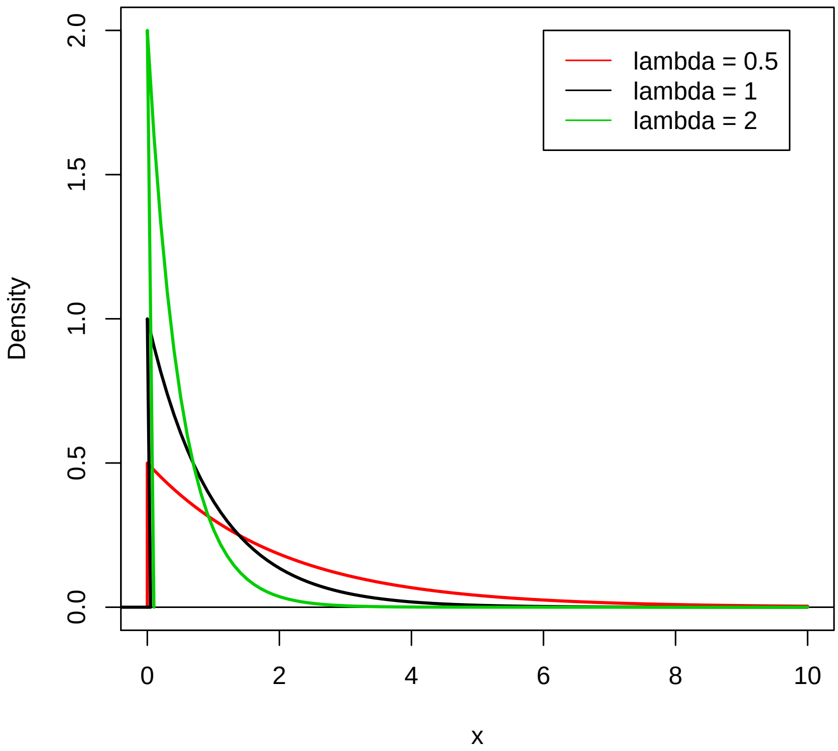

Figure 5.8: The Exponential Distribution for Various Values of λ

distribution. A larger expectation λ of the Poisson corresponds to a larger number of occurrences that are likely to take place during the unit interval of time. The larger is the number of occurrences the smaller are the time intervals between occurrences.

Example 5.7. *Consider Examples 5.4 and 5.5 that deal with rain dropping on a power line. The times between consecutive hits of the line may be modeled by the Exponential distribution. Hence, the time to the first hit has an Exponential distribution. The time between the first and the second hit is also Exponentially distributed, and so on.*

Example 5.8. *Return to Example 5.3 that deals with the radio activity of some element. The total count of decays per second is model by the Poisson distribution. The times between radio active decays is modeled according to the Exponential distribution. The rate λ of that Exponential distribution is equal to the expectation of the total count of decays in one second, i.e. the expectation of the Poisson distribution.*

5.4 Solved Exercises

Question 5.1. A particular measles vaccine produces a reaction (a fever higher that 102 Fahrenheit) in each vaccinee with probability of 0.09. A clinic vaccinates 500 people each day.

1. What is the expected number of people that will develop a reaction each day?
2. What is the standard deviation of the number of people that will develop a reaction each day?
3. In a given day, what is the probability that more than 40 people will develop a reaction?
4. In a given day, what is the probability that the number of people that will develop a reaction is between 50 and 45 (inclusive)?

Solution (to Question 5.1.1): The Binomial distribution is a reasonable model for the number of people that develop high fever as result of the vaccination. Let X be the number of people that do so in a give day. Hence, $X \sim \text{Binomial}(500, 0.09)$. According to the formula for the expectation in the Binomial distribution, since $n = 500$ and $p = 0.09$, we get that:

$$\text{E}(X) = np = 500 \times 0.09 = 45 \ .$$

Solution (to Question 5.1.2): Let $X \sim \text{Binomial}(500, 0.09)$. Using the formula for the variance for the Binomial distribution we get that:

$$\text{V}ar(X) = np(1-p) = 500 \times 0.09 \times 0.91 = 40.95 \ .$$

Hence, since $\sqrt{Var(X)} = \sqrt{40.95} = 6.3992$, the standard deviation is 6.3992.

Solution (to Question 5.1.3): Let $X \sim \text{Binomial}(500, 0.09)$. The probability that more than 40 people will develop a reaction may be computed as the difference between 1 and the probability that 40 people or less will develop a reaction:

$$\text{P}(X > 40) = 1 - \text{P}(X \leq 40) \ .$$

The probability can be computes with the aid of the function "`pbinom`" that produces the cumulative probability of the Binomial distribution:

```
> 1 - pbinom(40,500,0.09)
[1] 0.7556474
```

Solution (to Question 5.1.4): The probability that the number of people that will develop a reaction is between 50 and 45 (inclusive) is the difference between $\text{P}(X \leq 50)$ and $\text{P}(X < 45) = \text{P}(X \leq 44)$. Apply the function "`pbinom`" to get:

```
> pbinom(50,500,0.09) - pbinom(44,500,0.09)
[1] 0.3292321
```

Question 5.2. The Negative-Binomial distribution is yet another example of a discrete, integer valued, random variable. The sample space of the distribution are all non-negative integers $\{0, 1, 2, \ldots\}$. The fact that a random variable X has this distribution is marked by "$X \sim$ Negative-Binomial(r, p)", where r and p are parameters that specify the distribution.

Consider 3 random variables from the Negative-Binomial distribution:

- $X_1 \sim$ Negative-Binomial$(2, 0.5)$
- $X_2 \sim$ Negative-Binomial$(4, 0.5)$
- $X_3 \sim$ Negative-Binomial$(8, 0.8)$

The bar plots of these random variables are presented in Figure 5.9, re-organizer in a random order.

1. Produce bar plots of the distributions of the random variables X_1, X_2, X_3 in the range of integers between 0 and 15 and thereby identify the pair of parameters that produced each one of the plots in Figure 5.9. Notice that the bar plots can be produced with the aid of the function "`plot`" and the function "`dnbinom(x,r,p)`", where "`x`" is a sequence of integers and "`r`" and "`p`" are the parameters of the distribution. Pay attention to the fact that you should use the argument "`type = "h"`" in the function "`plot`" in order to produce the horizontal bars.

2. Below is a list of pairs that includes an expectation and a variance. Each of the pairs is associated with one of the random variables X_1, X_2, and X_3:

   (a) $\mathrm{E}(X) = 4$, $\mathrm{Var}(X) = 8$.

   (b) $\mathrm{E}(X) = 2$, $\mathrm{Var}(X) = 4$.

   (c) $\mathrm{E}(X) = 2$, $\mathrm{Var}(X) = 2.5$.

   Use Figure 5.9 in order to match the random variable with its associated pair. Do not use numerical computations or formulae for the expectation and the variance in the Negative-Binomial distribution in order to carry out the matching[6]. Use, instead, the structure of the bar-plots.

Solution (to Question 5.2.1): The plots can be produced with the following code, which should be run one line at a time:

```
> x <- 0:15
> plot(x,dnbinom(x,2,0.5),type="h")
> plot(x,dnbinom(x,4,0.5),type="h")
> plot(x,dnbinom(x,8,0.8),type="h")
```

The first plot, that corresponds to $X_1 \sim$ Negative-Binomial$(2, 0.5)$, fits Barplot 3. Notice that the distribution tends to obtain smaller values and that the probability of the value "0" is equal to the probability of the value "1".

The second plot, the one that corresponds to $X_2 \sim$ Negative-Binomial$(4, 0.5)$, is associated with Barplot 1. Notice that the distribution tends to obtain larger

[6] It can be shown, or else found on the web, that if $X \sim$ Negative-Binomial(r, p) then $\mathrm{E}(X) = r(1 - p)/p$ and $\mathrm{Var}(X) = r(1 - p)/p^2$.

values. For example, the probability of the value "10" is substantially larger than zero, where for the other two plots this is not the case.

The third plot, the one that corresponds to $X_3 \sim$ Negative-Binomial$(8, 0.8)$, matches Barplot 2. Observe that this distribution tends to produce smaller probabilities for the small values as well as for the larger values. Overall, it is more concentrated than the other two.

Solution (to Question 5.2.2): Barplot 1 corresponds to a distribution that tends to obtain larger values than the other two distributions. Consequently, the expectation of this distribution should be larger. The conclusion is that the pair $\mathrm{E}(X) = 4$, $\mathrm{V}ar(X) = 8$ should be associated with this distribution.

Barplot 2 describes a distribution that produce smaller probabilities for the small values as well as for the larger values and is more concentrated than the other two. The expectations of the two remaining distributions are equal to each other and the variance of the pair $\mathrm{E}(X) = 2$, $\mathrm{V}ar(X) = 2.5$ is smaller. Consequently, this is the pair that should be matched with this box plot.

This leaves only Barplot 3, that should be matched with the pair $\mathrm{E}(X) = 2$, $\mathrm{V}ar(X) = 4$.

5.5 Summary

Glossary

Binomial Random Variable: The number of successes among n repeats of independent trials with a probability p of success in each trial. The distribution is marked as Binomial(n, p).

Poisson Random Variable: An approximation to the number of occurrences of a rare event, when the expected number of events is λ. The distribution is marked as Poisson(λ).

Density: Histogram that describes the distribution of a continuous random variable. The area under the curve corresponds to probability.

Uniform Random Variable: A model for a measurement with equally likely outcomes over an interval $[a, b]$. The distribution is marked as Uniform(a, b).

Exponential Random Variable: A model for times between events. The distribution is marked as Exponential(λ).

Discuss in the Forum

This unit deals with two types of discrete random variables, the Binomial and the Poisson, and two types of continuous random variables, the Uniform and the Exponential. Depending on the context, these types of random variables may serve as theoretical models of the uncertainty associated with the outcome of a measurement.

In your opinion, is it or is it not useful to have a theoretical model for a situation that occurs in real life?

When forming your answer to this question you may give an example of a situation from you own field of interest for which a random variable, possibly from one of the types that are presented in this unit, can serve as a model. Discuss the importance (or lack thereof) of having a theoretical model for the situation.

For example, the Exponential distribution may serve as a model for the time until an atom of a radio active element decays by the release of subatomic particles and energy. The decay activity is measured in terms of the number of decays per second. This number is modeled as having a Poisson distribution. Its expectation is the rate of the Exponential distribution. For the radioactive element Carbon-14 (^{14}C) the decay rate is 3.8394×10^{-12} particles per second. Computations that are based on the Exponential model may be used in order to date ancient specimens.

Summary of Formulas

Discrete Random Variable:

$$\mathrm{E}(X) = \sum_x \big(x \times \mathrm{P}(x)\big)$$
$$\mathrm{Var}(X) = \sum_x \big((x - \mathrm{E}(X))^2 \times \mathrm{P}(x)\big)$$

Continuous Random Variable:

$$\mathrm{E}(X) = \int \big(x \times f(x)\big)dx$$
$$\mathrm{Var}(X) = \int \big((x - \mathrm{E}(X))^2 \times f(x)\big)dx$$

Binomial: $\mathrm{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$

Poisson: $\mathrm{E}(X) = \lambda$, $\mathrm{Var}(X) = \lambda$

Uniform: $\mathrm{E}(X) = (a+b)/2$, $\mathrm{Var}(X) = (b-a)^2/12$

Exponential: $\mathrm{E}(X) = 1/\lambda$, $\mathrm{Var}(X) = 1/\lambda^2$

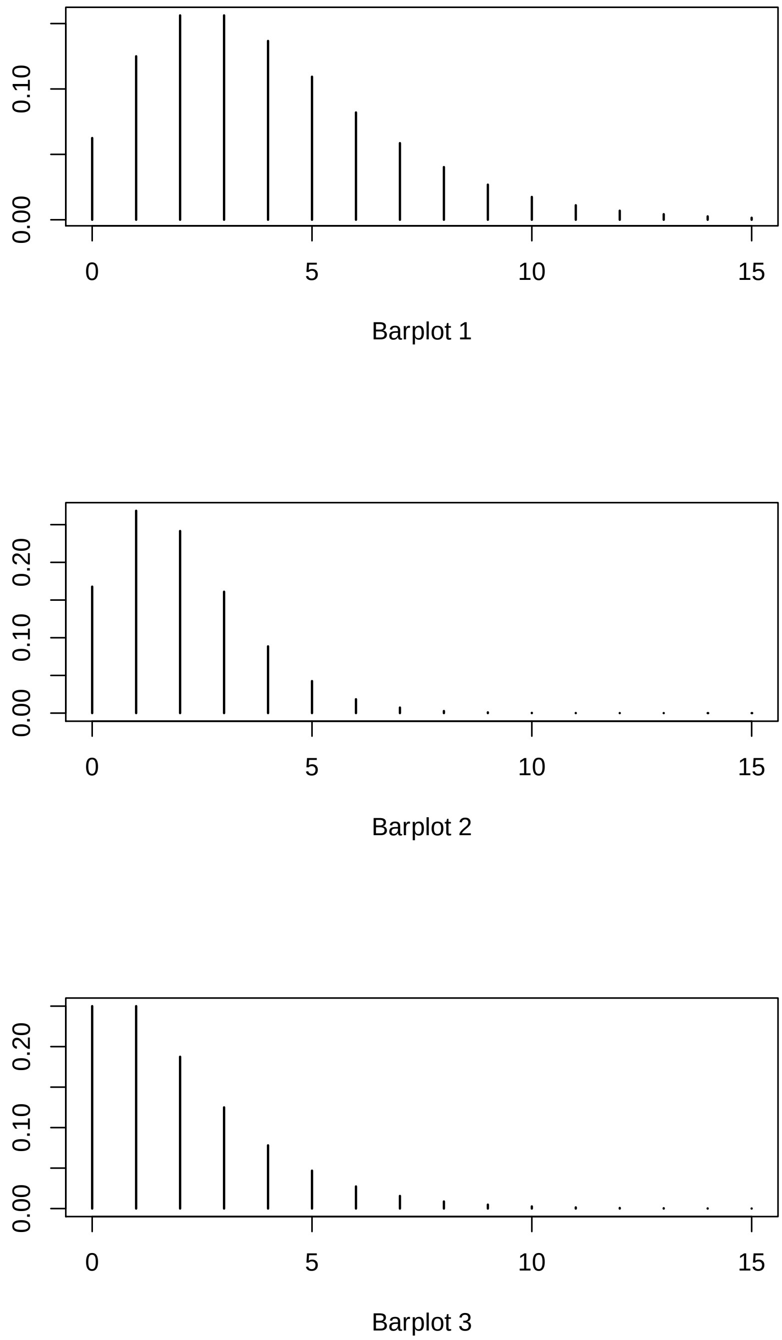

Figure 5.9: Bar Plots of the Negative-Binomial Distribution

Chapter 6

The Normal Random Variable

6.1 Student Learning Objective

This chapter introduces a very important bell-shaped distribution known as the Normal distribution. Computations associated with this distribution are discussed, including the percentiles of the distribution and the identification of intervals of subscribed probability. The Normal distribution may serve as an approximation to other distributions. We demonstrate this property by showing that under appropriate conditions the Binomial distribution can be approximated by the Normal distribution. This property of the Normal distribution will be picked up in the next chapter where the mathematical theory that establishes the Normal approximation is demonstrated. By the end of this chapter, the student should be able to:

- Recognize the Normal density and apply `R` functions for computing Normal probabilities and percentiles.

- Associate the distribution of a Normal random variable with that of its standardized counterpart, which is obtained by centering and re-scaling.

- Use the Normal distribution to approximate the Binomial distribution.

6.2 The Normal Random Variable

The Normal distribution is the most important of all distributions that are used in statistics. In many cases it serves as a generic model for the distribution of a measurement. Moreover, even in cases where the measurement is modeled by other distributions (i.e. Binomial, Poisson, Uniform, Exponential, etc.) the Normal distribution emerges as an approximation of the distribution of numerical characteristics of the data produced by such measurements.

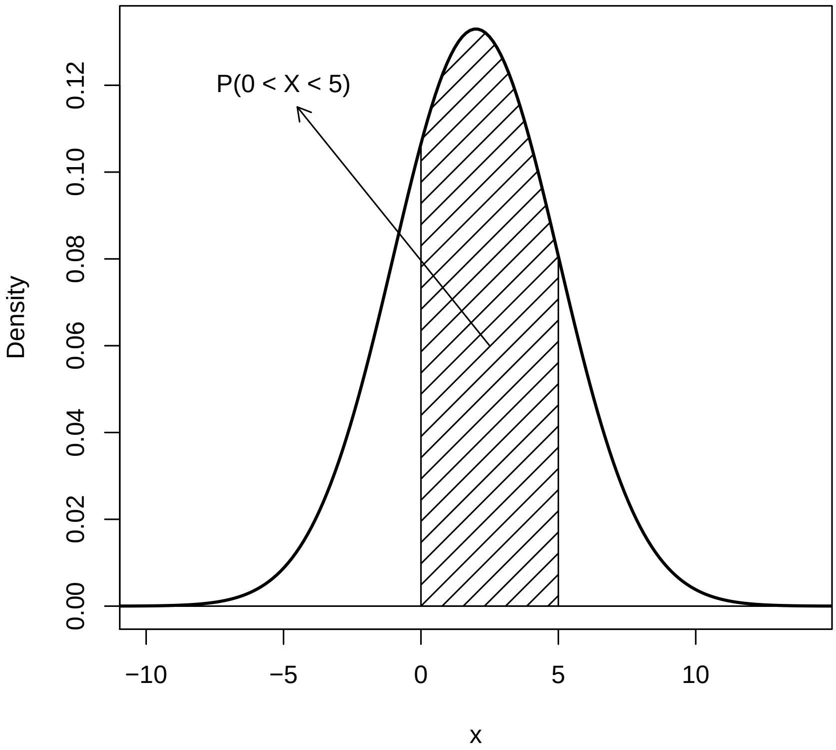

Figure 6.1: The Normal(2,9) Distribution

6.2.1 The Normal Distribution

A Normal random variable has a continuous distribution over the sample space of all numbers, negative or positive. We denote the Normal distribution via "$X \sim \text{Normal}(\mu, \sigma^2)$", where $\mu = \text{E}(X)$ is the expectation of the random variable and $\sigma^2 = \text{Var}(X)$ is it's variance[1].

Consider, for example, $X \sim \text{Normal}(2, 9)$. The density of the distribution is presented in Figure 6.1. Observe that the distribution is symmetric about the expectation 2. The random variable is more likely to obtain its value in the vicinity of the expectation. Values much larger or much smaller than the expectation are substantially less likely.

The density of the Normal distribution can be computed with the aid of the function "`dnorm`". The cumulative probability can be computed with the function "`pnorm`". For illustrating the use of the latter function, assume that $X \sim \text{Normal}(2, 9)$. Say one is interested in the computation of the probability $\text{P}(0 < X \leq 5)$ that the random variable obtains a value that belongs to the

[1]If $X \sim \text{Normal}(\mu, \sigma^2)$ then the density of X is given by the formula $f(x) = \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}/\sqrt{2\pi\sigma^2}$, for all x.

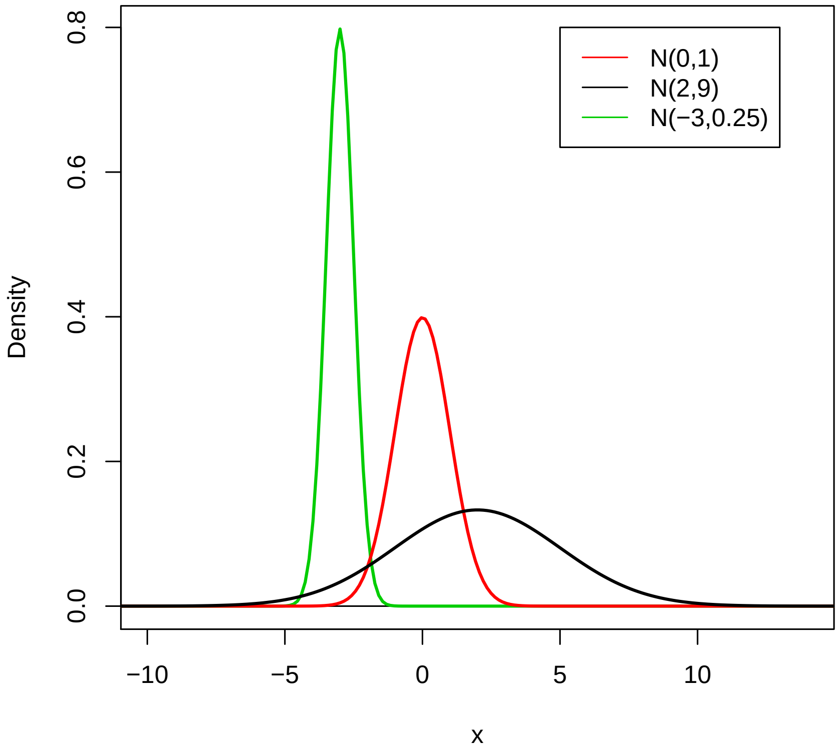

Figure 6.2: The Normal Distribution for Various Values of μ and σ^2

interval $(0, 5]$. The required probability is indicated by the marked area in Figure 6.1. This area can be computed as the difference between the probability $\mathrm{P}(X \leq 5)$, the area to the left of 5, and the probability $\mathrm{P}(X \leq 0)$, the area to the left of 0:

```
> pnorm(5,2,3) - pnorm(0,2,3)
[1] 0.5888522
```

The difference is the indicated area that corresponds to the probability of being inside the interval, which turns out to be approximately equal to 0.589. Notice that the expectation μ of the Normal distribution is entered as the second argument to the function. The third argument to the function is the standard deviation, i.e. the square root of the variance. In this example, the standard deviation is $\sqrt{9} = 3$.

Figure 6.2 displays the densities of the Normal distribution for the combinations $\mu = 0$, $\sigma^2 = 1$ (the *red* line); $\mu = 2$, $\sigma^2 = 9$ (the *black* line); and $\mu = -3$, $\sigma^2 = 1/4$ (the *green* line). Observe that the smaller the variance the more concentrated is the distribution of the random variable about the expectation.

Example 6.1. *IQ tests are a popular (and controversial) mean for measuring intelligence. They are produced as (weighted) average of a response to a long list of questions, designed to test different abilities. The score of the test across the entire population is set to be equal to 100 and the standard deviation is set to 15. The distribution of the score is Normal. Hence, if X is the IQ score of a random subject then $X \sim \mathrm{Normal}(100, 15^2)$.*

Example 6.2. *Any measurement that is produced as a result of the combination of many independent influencing factors is likely to poses the Normal distribution. For example, the hight of a person is influenced both by genetics and by the environment in which that person grew up. Both the genetic and the environmental influences are a combination of many factors. Thereby, it should not come as a surprise that the heights of people in a population tend to follow the Normal distribution.*

6.2.2 The Standard Normal Distribution

The standard normal distribution is a normal distribution of standardized values, which are called z-scores. A z-score is the original measurement measured in units of the standard deviation from the expectation. For example, if the expectation of a Normal distribution is 2 and the standard deviation is $3 = \sqrt{9}$, then the value of 0 is 2/3 standard deviations smaller than (or to the left of) the expectation. Hence, the z-score of the value 0 is -2/3. The calculation of the z-score emerges from the equation:

$$(0 =)\ x = \mu + z \cdot \sigma\ (= 2 + z \cdot 3)$$

The z-score is obtained by solving the equation

$$0 = 2 + z \cdot 3 \quad \Longrightarrow \quad z = (0 - 2)/3 = -2/3\ .$$

In a similar way, the z-score of the value $x = 5$ is equal to 1, following the solution of the equation $5 = 2 + z \cdot 3$, which leads to $z = (5 - 2)/3 = 1$.

The standard Normal distribution is the distribution of a standardized Normal measurement. The expectation for the standard Normal distribution is 0 and the variance is 1. When $X \sim N(\mu, \sigma^2)$ has a Normal distribution with expectation μ and variance σ^2 then the transformed random variable $Z = (X - \mu)/\sigma$ produces the standard Normal distribution $Z \sim N(0, 1)$. The transformation corresponds to the reexpression of the original measurement in terms of a new "zero" and a new unit of measurement. The new "zero" is the expectation of the original measurement and the new unit is the standard deviation of the original measurement.

Computation of probabilities associated with a Normal random variable X can be carried out with the aid of the standard Normal distribution. For example, consider the computation of the probability $\mathrm{P}(0 < X \leq 5)$ for $X \sim N(2, 9)$, that has expectation $\mu = 2$ and standard deviation $\sigma = 3$. Consider X's standardized values: $Z = (X - 2)/3$. The boundaries of the interval $[0, 5]$, namely 0 and 5, have standardized z-scores of $(0 - 2)/3 = -2/3$ and $(5 - 2)/3 = 1$, respectively. Clearly, the original measurement X falls between the original boundaries $(0 < X \leq 5)$ if, and only if, the standardized measurement Z falls

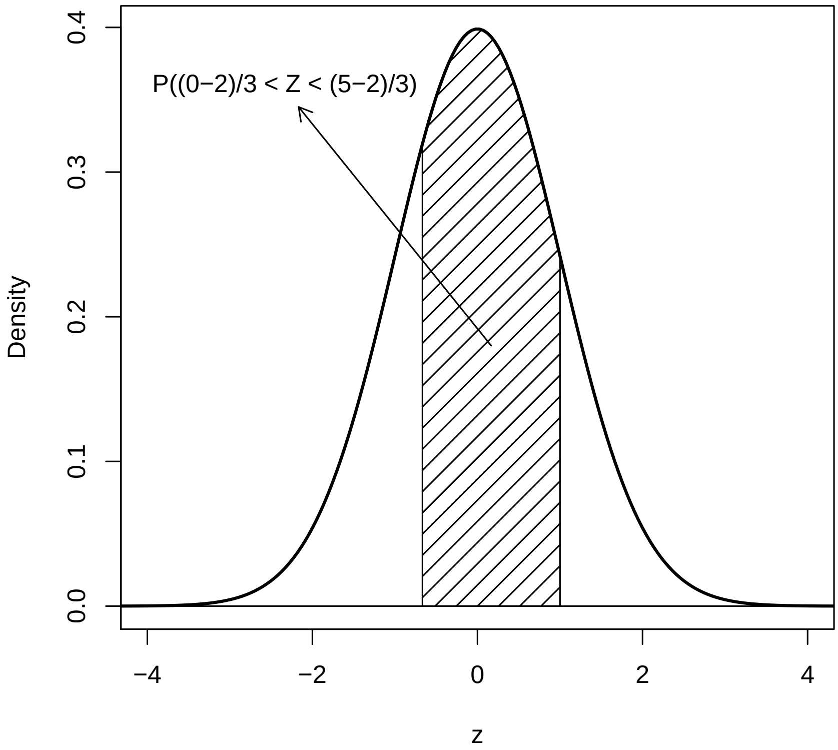

Figure 6.3: The Standard Normal Distribution

between the standardized boundaries $(-2/3 < Z \leq 1)$. Therefore, the probability that X obtains a value in the range $[0, 5]$ is equal to the probability that Z obtains a value in the range $[-2/3, 1]$.

The function "`pnorm`" was used in the previous subsection in order to compute that probability that X obtains values between 0 and 5. The computation produced the probability 0.5888522. We can repeat the computation by the application of the same function to the standardized values:

```
> pnorm((5-2)/3) - pnorm((0-2)/3)
[1] 0.5888522
```

The value that is being computed, the area under the graph for the standard Normal distribution, is presented in Figure 6.3. Recall that 3 arguments where specified in the previous application of the function "`pnorm`": the x value, the expectation, and the standard deviation. In the given application we did not specify the last two arguments, only the first one. (Notice that the output of the expression "`(5-2)/3`" is a single number and, likewise, the output of the expression "`(0-2)/3`" is also a single number.)

Most `R` function have many arguments that enables flexible application in a wide range of settings. For convenience, however, default values are set to most of these arguments. These default values are used unless an alternative value for the argument is set when the function is called. The default value of the second argument of the function "`pnorm`" that specifies the expectation is "`mean=0`", and the default value of the third argument that specifies the standard deviation is "`sd=1`". Therefore, if no other value is set for these arguments the function computes the cumulative distribution function of the standard Normal distribution.

6.2.3 Computing Percentiles

Consider the issue of determining the range that contains 95% of the probability for a Normal random variable. We start with the standard Normal distribution. Consult Figure 6.4. The figure displays the standard Normal distribution with the central region shaded. The area of the shaded region is 0.95.

We may find the z-values of the boundaries of the region, denoted in the figure as z_0 and z_1 by the investigation of the cumulative distribution function. Indeed, in order to have 95% of the distribution in the central region one should leave out 2.5% of the distribution in each of the two tails. That is, 0.025 should be the area of the unshaded region to the right of z_1 and, likewise, 0.025 should be the area of the unshaded region to the left of z_0. In other words, the cumulative probability up to z_0 should be 0.025 and the cumulative distribution up to z_1 should be 0.975.

In general, given a random variable X and given a percent p, the x value with the property that the cumulative distribution up to x is equal to the probability p is called the *p-percentile* of the distribution. Here we seek the 2.5%-percentile and the 97.5%-percentile of the standard Normal distribution.

The percentiles of the Normal distribution are computed by the function "`qnorm`". The first argument to the function is a probability (or a sequence of probabilities), the second and third arguments are the expectation and the standard deviations of the normal distribution. The default values to these arguments are set to 0 and 1, respectively. Hence if these arguments are not provided the function computes the percentiles of the standard Normal distribution. Let us apply the function in order to compute z_1 and z_0:

```
> qnorm(0.975)
[1] 1.959964
> qnorm(0.025)
[1] -1.959964
```

Observe that z_1 is practically equal to 1.96 and $z_0 = -1.96 = -z_1$. The fact that z_0 is the negative of z_1 results from the symmetry of the standard Normal distribution about 0. As a conclusion we get that for the standard Normal distribution 95% of the probability is concentrated in the range $[-1.96, 1.96]$.

The problem of determining the central range that contains 95% of the distribution can be addresses in the context of the original measurement X (See Figure 6.5). We seek in this case an interval centered at the expectation 2, which is the center of the distribution of X, unlike 0 which was the center of the standardized values Z. One way of solving the problem is via the application of the function "`qnorm`" with the appropriate values for the expectation and the standard deviation:

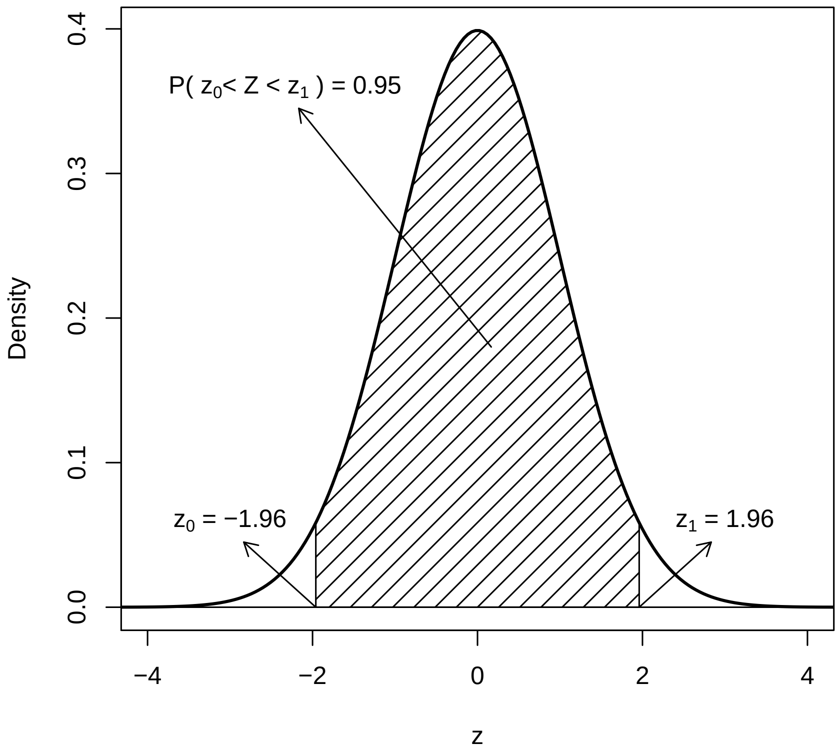

Figure 6.4: Central 95% of the Standard Normal Distribution

```
> qnorm(0.975,2,3)
[1] 7.879892
> qnorm(0.025,2,3)
[1] -3.879892
```

Hence, we get that $x_0 = -3.88$ has the property that the total probability to its left is 0.025 and $x_1 = 7.88$ has the property that the total probability to its right is 0.025. The total probability in the range $[-3.88, 7.88]$ is 0.95.

An alternative approach for obtaining the given interval exploits the interval that was obtained for the standardized values. An interval $[-1.96, 1.96]$ of standardized z-values corresponds to an interval $[2 - 1.96 \cdot 3, 2 + 1.96 \cdot 3]$ of the original x-values:

```
> 2 + qnorm(0.975)*3
[1] 7.879892
> 2 + qnorm(0.025)*3
[1] -3.879892
```

Hence, we again produce the interval $[-3.88, 7.88]$, the interval that was obtained before as the central interval that contains 95% of the distribution of the

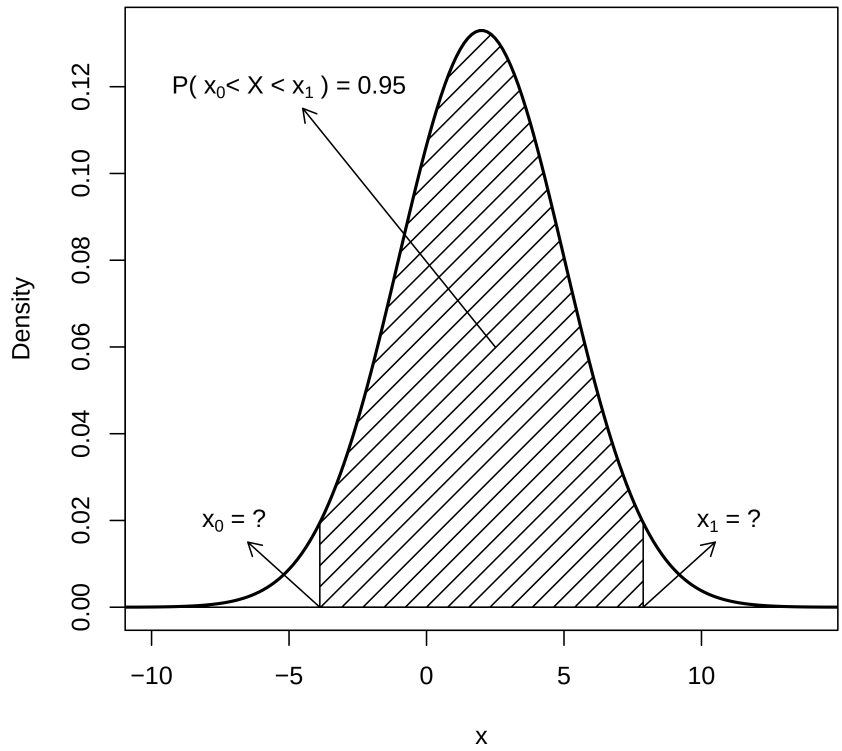

Figure 6.5: Central 95% of the Normal(2,9) Distribution

Normal$(2, 9)$ random variable.

In general, if $X \sim N(\mu, \sigma)$ is a Normal random variable then the interval $[\mu - 1.96 \cdot \sigma, \mu + 1.96 \cdot \sigma]$ contains 95% of the distribution of the random variable. Frequently one uses the notation $\mu \pm 1.96 \cdot \sigma$ to describe such an interval.

6.2.4 Outliers and the Normal Distribution

Consider, next, the computation of the interquartile range in the Normal distribution. Recall that the interquartile range is the length of the central interval that contains 50% of the distribution. This interval starts at the first quartile (Q1), the value that splits the distribution so that 25% of the distribution is to the left of the value and 75% is to the right of it. The interval ends at the third quartile (Q3) where 75% of the distribution is to the left and 25% is to the right.

For the standard Normal the third and first quartiles can be computed with the aid of the function "`qnorm`":

```
> qnorm(0.75)
```

```
[1] 0.6744898
> qnorm(0.25)
[1] -0.6744898
```

Observe that for the standard Normal distribution one has that 75% of the distribution is to the left of the value 0.6744898, which is the third quartile of this distribution. Likewise, 25% of the standard Normal distribution are to the left of the value -0.6744898, which is the first quartile. the interquartile range is the length of the interval between the third and the first quartiles. In the case of the standard Normal distribution this length is equal to $0.6744898-(-0.6744898) = 1.348980$.

In Chapter 3 we considered box plots as a mean for the graphical display of numerical data. The box plot includes a vertical rectangle that initiates at the first quartile and ends at the third quartile, with the median marked within the box. The rectangle contains 50% of the data. Whiskers extends from the ends of this rectangle to the smallest and to the largest data values that are not outliers. Outliers are values that lie outside of the normal range of the data. Outliers are identified as values that are more then 1.5 times the interquartile range away from the ends of the central rectangle. Hence, a value is an outlier if it is larger than the third quartile plus 1.5 times the interquartile range or if it is less than the first quartile minus 1.5 times the interquartile range.

How likely is it to obtain an outlier value when the measurement has the standard Normal distribution? We obtained that the third quartile of the standard Normal distribution is equal to 0.6744898 and the first quartile is minus this value. The interquartile range is the difference between the third and first quartiles. The upper and lower thresholds for the defining outliers are:

```
> qnorm(0.75) + 1.5*(qnorm(0.75)-qnorm(0.25))
[1] 2.697959
> qnorm(0.25) - 1.5*(qnorm(0.75)-qnorm(0.25))
[1] -2.697959
```

Hence, a value larger than 2.697959 or smaller than -2.697959 would be identified as an outlier.

The probability of being less than the upper threshold 2.697959 in the standard Normal distribution is computed with the expression "`pnorm(2.697959)`". The probability of being above the threshold is 1 minus that probability, which is the outcome of the expression "`1-pnorm(2.697959)`".

By the symmetry of the standard Normal distribution we get that the probability of being below the lower threshold -2.697959 is equal to the probability of being above the upper threshold. Consequently, the probability of obtaining an outlier is equal to twice the probability of being above the upper threshold:

```
> 2*(1-pnorm(2.697959))
[1] 0.006976603
```

We get that for the standard Normal distribution the probability of an outlier is approximately 0.7%.

6.3 Approximation of the Binomial Distribution

The Normal distribution emerges frequently as an approximation of the distribution of data characteristics. The probability theory that mathematically establishes such approximation is called the Central Limit Theorem and is the subject of the next chapter. In this section we demonstrate the Normal approximation in the context of the Binomial distribution.

6.3.1 Approximate Binomial Probabilities and Percentiles

Consider, for example, the probability of obtaining between 1940 and 2060 heads when tossing 4,000 fair coins. Let X be the total number of heads. The tossing of a coin is a trial with two possible outcomes: "Head" and "Tail." The probability of a "Head" is 0.5 and there are 4,000 trials. Let us call obtaining a "Head" in a trial a "Success". Observe that the random variable X counts the total number of successes. Hence, $X \sim \text{Binomial}(4000, 0.5)$.

The probability $\text{P}(1940 \leq X \leq 2060)$ can be computed as the difference between the probability $\text{P}(X \leq 2060)$ of being less or equal to 2060 and the probability $\text{P}(X < 1940)$ of being strictly less than 1940. However, 1939 is the largest integer that is still strictly less than the integer 1940. As a result we get that $\text{P}(X < 1940) = \text{P}(X \leq 1939)$. Consequently, $\text{P}(1940 \leq X \leq 2060) = \text{P}(X \leq 2060) - \text{P}(X \leq 1939)$.

Applying the function "`pbinom`" for the computation of the Binomial cumulative probability, namely the probability of being less or equal to a given value, we get that the probability in the range between 1940 and 2060 is equal to

```
> pbinom(2060,4000,0.5) - pbinom(1939,4000,0.5)
[1] 0.9442883
```

This is an exact computation. The Normal approximation produces an approximate evaluation, not an exact computation. The Normal approximation replaces Binomial computations by computations carried out for the Normal distribution. The computation of a probability for a Binomial random variable is replaced by computation of probability for a Normal random variable that has the same expectation and standard deviation as the Binomial random variable.

Notice that if $X \sim \text{Binomial}(4000, 0.5)$ then the expectation is $\text{E}(X) = 4,000 \times 0.5 = 2,000$ and the variance is $\text{V}ar(X) = 4,000 \times 0.5 \times 0.5 = 1,000$, with the standard deviation being the square root of the variance. Repeating the same computation that we conducted for the Binomial random variable, but this time with the function "`pnorm`" that is used for the computation of the Normal cumulative probability, we get:

```
> mu <- 4000*0.5
> sig <- sqrt(4000*0.5*0.5)
> pnorm(2060,mu,sig) - pnorm(1939,mu,sig)
[1] 0.9442441
```

Observe that in this example the Normal approximation of the probability (0.9442441) agrees with the Binomial computation of the probability (0.9442883) up to 3 significant digits.

Normal computations may also be applied in order to find approximate percentiles of the Binomial distribution. For example, let us identify the central region that contains for a Binomial$(4000, 0.5)$ random variable (approximately) 95% of the distribution. Towards that end we can identify the boundaries of the region for the Normal distribution with the same expectation and standard deviation as that of the target Binomial distribution:

```
> qnorm(0.975,mu,sig)
[1] 2061.980
> qnorm(0.025,mu,sig)
[1] 1938.020
```

After rounding to the nearest integer we get the interval $[1938, 2062]$ as a proposed central region.

In order to validate the proposed region we may repeat the computation under the actual Binomial distribution:

```
> qbinom(0.975,4000,0.5)
[1] 2062
> qbinom(0.025,4000,0.5)
[1] 1938
```

Again, we get the interval $[1938, 2062]$ as the central region, in agreement with the one proposed by the Normal approximation. Notice that the function "`qbinom`" produces the percentiles of the Binomial distribution. It may not come as a surprise to learn that "`qpois`", "`qunif`", "`qexp`" compute the percentiles of the Poisson, Uniform and Exponential distributions, respectively.

The ability to approximate one distribution by the other, when computation tools for both distributions are handy, seems to be of questionable importance. Indeed, the significance of the Normal approximation is not so much in its ability to approximate the Binomial distribution as such. Rather, the important point is that the Normal distribution may serve as an approximation to a wide class of distributions, with the Binomial distribution being only one example. Computations that are based on the Normal approximation will be valid for all members in the class of distributions, including cases where we don't have the computational tools at our disposal or even in cases where we do not know what the exact distribution of the member is! As promised, a more detailed discussion of the Normal approximation in a wider context will be presented in the next chapter.

On the other hand, one need not assume that any distribution is well approximated by the Normal distribution. For example, the distribution of wealth in the population tends to be skewed, with more than 50% of the people possessing less than 50% of the wealth and small percentage of the people possessing the majority of the wealth. The Normal distribution is not a good model for such distribution. The Exponential distribution, or distributions similar to it, may be more appropriate.

6.3.2 Continuity Corrections

In order to complete this section let us look more carefully at the Normal approximations of the Binomial distribution.

In principle, the Normal approximation is valid when n, the number of independent trials in the Binomial distribution, is large. When n is relatively small

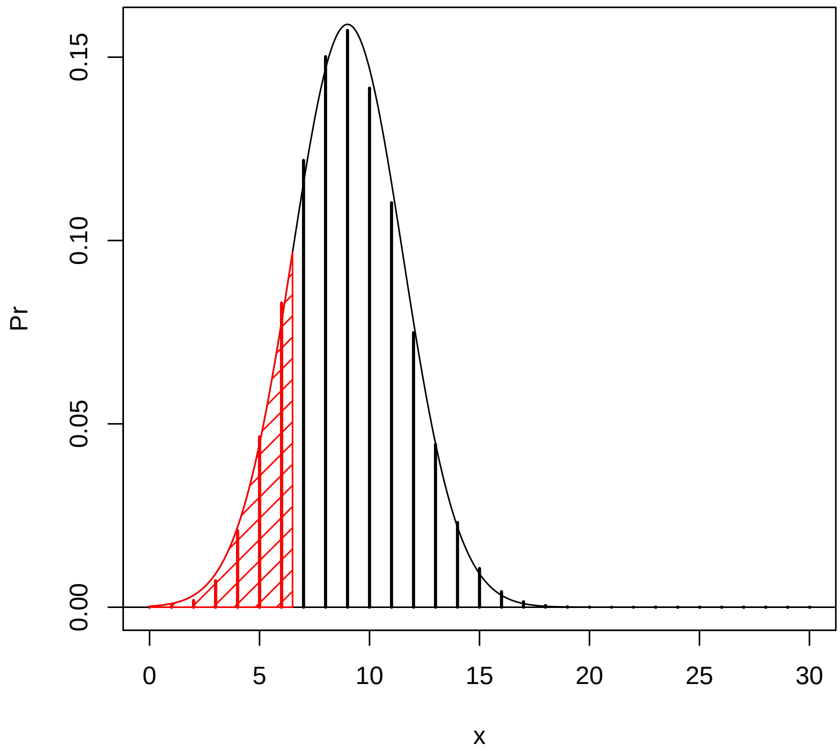

Figure 6.6: Normal Approximation of the Binomial Distribution

the approximation may not be so good. Indeed, take $X \sim \text{Binomial}(30, 0.3)$ and consider the probability $\text{P}(X \leq 6)$. Compare the actual probability to the Normal approximation:

```
> pbinom(6,30,0.3)
[1] 0.1595230
> pnorm(6,30*0.3,sqrt(30*0.3*0.7))
[1] 0.1159989
```

The Normal approximation, which is equal to 0.1159989, is not too close to the actual probability, which is equal to 0.1595230.

A naïve application of the Normal approximation for the $\text{Binomial}(n, p)$ distribution may not be so good when the number of trials n is small. Yet, a small modification of the approximation may produce much better results. In order to explain the modification consult Figure 6.6 where you will find the bar plot of the Binomial distribution with the density of the approximating Normal distribution superimposed on top of it. The target probability is the sum of heights of the bars that are painted in *red*. In the naïve application of the Normal approximation we used the area under the normal density which is to the left of the bar associated with the value $x = 6$.

Alternatively, you may associate with each bar located at x the area under the normal density over the interval $[x - 0.5, x + 0.5]$. The resulting correction to the approximation will use the Normal probability of the event $\{X \leq 6.5\}$, which is the area shaded in *red*. The application of this approximation, which is called *continuity correction* produces:

```
> pnorm(6.5,30*0.3,sqrt(30*0.3*0.7))
[1] 0.1596193
```

Observe that the corrected approximation is much closer to the target probability, which is 0.1595230, and is substantially better that the uncorrected approximation which was 0.1159989. Generally, it is recommended to apply the continuity correction to the Normal approximation of a discrete distribution.

Consider the Binomial(n, p) distribution. Another situation where the Normal approximation may fail is when p, the probability of "Success" in the Binomial distribution, is too close to 0 (or too close to 1). Recall, that for large n the Poisson distribution emerged as an approximation of the Binomial distribution in such a setting. One may expect that when n is large and p is small then the Poisson distribution may produce a better approximation of a Binomial probability. When the Poisson distribution is used for the approximation we call it a *Poisson Approximation*.

Let us consider an example. Let us analyze 3 Binomial distributions. The expectation in all the distributions is equal to 2 but the number of trials, n, vary. In the first case $n = 20$ (and hence $p = 0.1$), in the second $n = 200$ (and $p = 0.01$), and in the third $n = 2,000$ (and $p = 0.001$. In all three cases we will be interested in the probability of obtaining a value less or equal to 3.

The Poisson approximation replaces computations conducted under the Binomial distribution with Poisson computations, with a Poisson distribution that has the same expectation as the Binomial. Since in all three cases the expectation is equal to 2 we get that the same Poisson approximation is used to the three probabilities:

```
> ppois(3,2)
[1] 0.8571235
```

The actual Binomial probability in the first case ($n = 20$, $p = 0.1$) and a Normal approximation thereof are:

```
> pbinom(3,20,0.1)
[1] 0.8670467
> pnorm(3.5,2,sqrt(20*0.1*0.9))
[1] 0.8682238
```

Observe that the Normal approximation (with a continuity correction) is better than the Poisson approximation in this case.

In the second case ($n = 200$, $p = 0.01$) the actual Binomial probability and the Normal approximation of the probability are:

```
> pbinom(3,200,0.01)
[1] 0.858034
> pnorm(3.5,2,sqrt(200*0.01*0.99))
[1] 0.856789
```

Observe that the Poisson approximation that produces 0.8571235 is slightly closer to the target than the Normal approximation. The greater accuracy of the Poisson approximation for the case where n is large and p is small is more pronounced in the final case ($n = 2000$, $p = 0.001$) where the target probability and the Normal approximation are:

```
> pbinom(3,2000,0.001)
[1] 0.8572138
> pnorm(3.5,2,sqrt(2000*0.001*0.999))
[1] 0.8556984
```

Compare the actual Binomial probability, which is equal to 0.8572138, to the Poisson approximation that produced 0.8571235. The Normal approximation, 0.8556984, is slightly off, but is still acceptable.

6.4 Solved Exercises

Question 6.1. Consider the problem of establishing regulations concerning the maximum number of people who can occupy a lift. In particular, we would like to assess the probability of exceeding maximal weight when 8 people are allowed to use the lift simultaneously and compare that to the probability of allowing 9 people into the lift.

Assume that the total weight of 8 people chosen at random follows a normal distribution with a mean of 560kg and a standard deviation of 57kg. Assume that the total weight of 9 people chosen at random follows a normal distribution with a mean of 630kg and a standard deviation of 61kg.

1. What is the probability that the total weight of 8 people exceeds 650kg?
2. What is the probability that the total weight of 9 people exceeds 650kg?
3. What is the central region that contains 80% of distribution of the total weight of 8 people?
4. What is the central region that contains 80% of distribution of the total weight of 9 people?

Solution (to Question 6.1.1): Let X be the total weight of 8 people. By the assumption, $X \sim \mathrm{Normal}(560, 57^2)$. We are interested in the probability $\mathrm{P}(X > 650)$. This probability is equal to the difference between 1 and the probability $\mathrm{P}(X \leq 650)$. We use the function "`pnorm`" in order to carry out the computation:

```
> 1 - pnorm(650,560,57)
[1] 0.05717406
```

We get that the probability that the total weight of 8 people exceeds 650kg is equal to 0.05717406.

Solution (to Question 6.1.2): Let Y be the total weight of 9 people. By the assumption, $Y \sim \mathrm{Normal}(630, 61^2)$. We are interested in the probability $\mathrm{P}(Y > 650)$. This probability is equal to the difference between 1 and the probability $\mathrm{P}(Y \leq 650)$. We use again the function "`pnorm`" in order to carry out the computation:

```
> 1 - pnorm(650,630,61)
[1] 0.3715054
```

We get that the probability that the total weight of 9 people exceeds 650kg is much higher and is equal to 0.3715054.

Solution (to Question 6.1.3): Again, $X \sim \text{Normal}(560, 57^2)$, where X is the total weight of 8 people. In order to find the central region that contains 80% of the distribution we need to identify the 10%-percentile and the 90%-percentile of X. We use the function "`qnorm`" in the code:

```
> qnorm(0.1,560,57)
[1] 486.9516
> qnorm(0.9,560,57)
[1] 633.0484
```

The requested region is the interval [486.9516, 633.0484].

Solution (to Question 6.1.4): As before, $Y \sim \text{Normal}(630, 61^2)$, where Y is the total weight of 9 people. In order to find the central region that contains 80% of the distribution we need to identify the 10%-percentile and the 90%-percentile of Y. The computation this time produces:

```
> qnorm(0.1,630,61)
[1] 551.8254
> qnorm(0.9,630,61)
[1] 708.1746
```

and the region is [551.8254, 708.1746].

Question 6.2. Assume $X \sim \text{Binomial}(27, 0.32)$. We are interested in the probability $\text{P}(X > 11)$.

1. Compute the (exact) value of this probability.
2. Compute a Normal approximation to this probability, without a continuity correction.
3. Compute a Normal approximation to this probability, with a continuity correction.
4. Compute a Poisson approximation to this probability.

Solution (to Question 6.2.1): The probability $\text{P}(X > 11)$ can be computed as the difference between 1 and the probability $\text{P}(X \leq 11)$. The latter probability can be computed with the function "`pbinom`":

```
> 1 - pbinom(11,27,0.32)
[1] 0.1203926
```

Therefore, $\text{P}(X > 11) = 0.1203926$.

Solution (to Question 6.2.2): Refer again to the probability $\text{P}(X > 11)$. A formal application of the Normal approximation replaces in the computation the Binomial distribution by the Normal distribution with the same mean and variance. Since $\mathrm{E}(X) = n \cdot p = 27 \cdot 0.32 = 8.64$ and $\mathrm{Var}(X) = n \cdot p \cdot (1 - p) = 27 \cdot 0.32 \cdot 0.68 = 5.8752$. If we take $X \sim \mathrm{Normal}(8.64, 5.8752)$ and use the function "pnorm" we get:

```
> 1 - pnorm(11,27*0.32,sqrt(27*0.32*0.68))
[1] 0.1651164
```

Therefore, the current Normal approximation proposes $\mathrm{P}(X > 11) \approx 0.1651164$.

Solution (to Question 6.2.3): The continuity correction, that consider interval of range 0.5 about each value, replace $\mathrm{P}(X > 11)$, that involves the values $\{12, 13, \ldots, 27\}$, by the event $\mathrm{P}(X > 11.5)$. The Normal approximation uses the Normal distribution with the same mean and variance. Since $\mathrm{E}(X) = 8.64$ and $\mathrm{Var}(X) = 5.8752$. If we take $X \sim \mathrm{Normal}(8.64, 5.8752)$ and use the function "pnorm" we get:

```
> 1 - pnorm(11.5,27*0.32,sqrt(27*0.32*0.68))
[1] 0.1190149
```

The Normal approximation with continuity correction proposes $\mathrm{P}(X > 11) \approx 0.1190149$.

Solution (to Question 6.2.4): The Poisson approximation replaces the Binomial distribution by the Poisson distribution with the same expectation. The expectation is $\mathrm{E}(X) = n \cdot p = 27 \cdot 0.32 = 8.64$. If we take $X \sim \mathrm{Poisson}(8.64)$ and use the function "ppois" we get:

```
> 1 - ppois(11,27*0.32)
[1] 0.1635232
```

Therefore, the Poisson approximation proposes $\mathrm{P}(X > 11) \approx 0.1651164$.

6.5 Summary

Glossary

Normal Random Variable: A bell-shaped distribution that is frequently used to model a measurement. The distribution is marked with $\mathrm{Normal}(\mu, \sigma^2)$.

Standard Normal Distribution: The $\mathrm{Normal}(0, 1)$. The distribution of standardized Normal measurement.

Percentile: Given a percent $p \cdot 100\%$ (or a probability p), the value x is the percentile of a random variable X if it satisfies the equation $\mathrm{P}(X \leq x) = p$.

Normal Approximation of the Binomial: Approximate computations associated with the Binomial distribution with parallel computations that use the Normal distribution with the same expectation and standard deviation as the Binomial.

Poisson Approximation of the Binomial: Approximate computations associated with the Binomial distribution with parallel computations that use the Poisson distribution with the same expectation as the Binomial.

Discuss in the Forum

Mathematical models are used as tools to describe reality. These models are supposed to characterize the important features of the analyzed phenomena and provide insight. Random variables are mathematical models of measurements. Some people claim that there should be a perfect match between the mathematical characteristics of a random variable and the properties of the measurement it models. Other claim that a partial match is sufficient. What is your opinion?

When forming your answer to this question you may give an example of a situation from you own field of interest for which a random variable can serve as a model. Identify discrepancies between the theoretical model and actual properties of the measurement. Discuss the appropriateness of using the model in light of these discrepancies.

Consider, for example, testing IQ. The score of many IQ tests are modeled as having a Normal distribution with an expectation of 100 and a standard deviation of 15. The sample space of the Normal distribution is the entire line of real numbers, including the negative numbers. In reality, IQ tests produce only positive values.