Probability Distributions 8

We introduced the concept of probability density and probability mass functions of random variables in the previous chapter. In this chapter, we are introducing some common standard discrete and continuous probability distributions which are widely used for either practical applications or constructing statistical methods described later in this book. Suppose we are interested in determining the probability of a certain event. The determination of probabilities depends upon the nature of the study and various prevailing conditions which affect it. For example, the determination of the probability of a head when tossing a coin is different from the determination of the probability of rain in the afternoon. One can speculate that some mathematical functions can be defined which depict the behaviour of probabilities under different situations. Such functions have special properties and describe how probabilities are distributed under different conditions. We have already learned that they are called probability distribution functions. The form of such functions may be simple or complicated depending upon the nature and complexity of the phenomenon under consideration. Let us first recall and extend the definition of independent and identically distributed random variables:

Definition 8.0.1 The random variables $X_1, X_2, \ldots, X_n$ are called independent and identically distributed (i.i.d) if the $X_i\,(i = 1, 2, \ldots, n)$ have the same marginal cumulative distribution function $F(x)$ and if they are mutually independent.

Example 8.0.1 Suppose a researcher plans a survey on the weight of newborn babies in a country. The researcher randomly contacts 10 hospitals with a maternity ward and asks them to randomly select 20 of the newborn babies (no twins) born in the last 6 months and records their weights. The sample therefore consists of $10 \times 20 = 200$ baby weights. Since the hospitals and the babies are randomly selected, the babies' weights are therefore not known beforehand. The 200 weights can be denoted by the random variables $X_1, X_2, \ldots, X_{200}$. Note that the weights X_i are random variables

C. Heumann et al., *Introduction to Statistics and Data Analysis*,
DOI 10.1007/978-3-319-46162-5_8

because, depending on the size of the population, different samples consisting of 200 babies can be randomly selected. Also, the babies' weights can be seen as stochastically independent (an example of stochastically dependent weights would be the weights of twins if they are included in the sample). After collecting the weights of 200 babies, the researcher has a sample of 200 realized values (i.e. the weights in grams). The values are now known and denoted by $x_1, x_2, \ldots, x_{200}$.

8.1 Standard Discrete Distributions

First, we discuss some standard distributions for discrete random variables.

8.1.1 Discrete Uniform Distribution

The discrete uniform distribution assumes that all possible outcomes have equal probability of occurrence. A more formal definition is given as follows:

Definition 8.1.1 A discrete random variable X with k possible outcomes $x_1, x_2, \ldots, x_k$ is said to follow a discrete **uniform** distribution if the probability mass function (PMF) of X is given by

$$P(X = x_i) = \frac{1}{k}, \quad \forall i = 1, 2, \ldots, k. \tag{8.1}$$

If the outcomes are the natural numbers $x_i = i$ $(i = 1, 2, \ldots, k)$, the mean and variance of X are obtained as

$$\mathrm{E}(X) = \frac{k+1}{2}, \tag{8.2}$$

$$\mathrm{Var}(X) = \frac{1}{12}(k^2 - 1). \tag{8.3}$$

Example 8.1.1 If we roll a fair die, the outcomes "1", "2", ..., "6" have equal probability of occurring, and hence, the random variable X "number of dots observed on the upper surface of the die" has a uniform discrete distribution with PMF

$$P(X = i) = \frac{1}{6}, \quad \text{for all} \quad i = 1, 2, \ldots, 6.$$

The mean and variance of X are

$$\mathrm{E}(X) = \frac{6+1}{2} = 3.5,$$

$$\mathrm{Var}(X) = \frac{1}{12}(6^2 - 1) = 35/12.$$

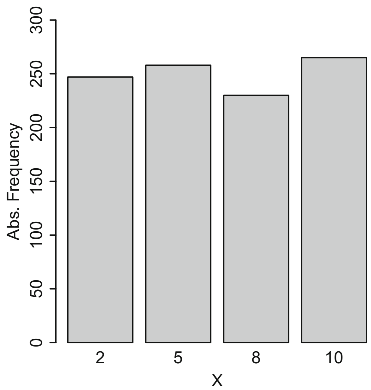


Fig. 8.1 Frequency distribution of 1000 generated discrete uniform random numbers with possible outcomes (2, 5, 8, 10)

Using the function `sample()` in *R*, it is easy to generate random numbers from a discrete uniform distribution. The following command generates a random sample of size 1000 from a uniform distribution with the four possible outcomes 2, 5, 8, 10 and draws a bar chart of the observed numbers. The use of the `set.seed()` function allows to reproduce the generated random numbers at any time. It is necessary to use the option `replace=TRUE` to simulate draws with replacement, i.e. to guarantee that a value can occur more than once.

```
set.seed(123789)
x <- sample(x=c(2,5,8,10), size=1000, replace=T,
prob=c(1/4,1/4,1/4,1/4))
barchart(table(x), ylim=c(0,300))
```

A bar chart of the frequency distribution of the 1000 sampled numbers with the possible outcomes (2, 5, 8, 10) using the discrete uniform distribution is given in Fig. 8.1. We see that the 1000 generated random numbers are not exactly uniformly distributed, e.g. the numbers 5 and 10 occur more often than the numbers 2 and 8. In fact, they are only approximately uniform. We expect that the deviance from a perfect uniform distribution is getting smaller as we generate more and more random numbers but will probably never be zero for a finite number of draws. The random numbers reflect the practical situation that a sample distribution is only an approximation to the theoretical distribution from which the sample was drawn. More details on how to work with random variables in *R* are given in Appendix A.3.

8.1.2 Degenerate Distribution

Definition 8.1.2 A random variable X has a **degenerate distribution** at a, if a is the only possible outcome with $P(X = a) = 1$. The CDF in such a case is given by

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ 1 & \text{if } x \geq a. \end{cases}$$

Further, $\mathrm{E}(X) = a$ and $\mathrm{Var}(X) = 0$.

The degenerate distribution indicates that there is only one possible fixed outcome, and therefore, no randomness is involved. It follows that we need at least two different possible outcomes to have randomness in the observations of a random variable or random experiment. The Bernoulli distribution is such a distribution where there are only two outcomes, e.g. success and failure or male and female. These outcomes are usually denoted by the values "0" and "1".

8.1.3 Bernoulli Distribution

Definition 8.1.3 A random variable X has a Bernoulli distribution if the PMF of X is given as

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0. \end{cases}$$

The cumulative distribution function (CDF) of X is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

The mean (expectation) and variance of a Bernoulli random variable are calculated as

$$\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p \tag{8.4}$$

and

$$\mathrm{Var}(X) = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p), \tag{8.5}$$

respectively.

A Bernoulli distribution is useful when there are only two possible outcomes, and our interest lies in any of the two outcomes, e.g. whether a customer buys a certain product or not, or whether a hurricane hits an island or not. The outcome of an event A is usually coded as 1 which occurs with probability p. If the event of interest does not occur, i.e. the complementary event $\bar{A}$ occurs, the outcome is coded as 0 which occurs with probability $1 - p$. So p is the probability that the event of interest A occurs.

Example 8.1.2 A company organizes a raffle at an end-of-year function. There are 300 lottery tickets in total, and 50 of them are marked as winning tickets. The event A of interest is "ticket wins" (coded as $X = 1$), and the probability p of having a winning ticket is *a priori* (i.e. before any lottery ticket has been drawn)

$$P(X = 1) = \frac{50}{300} = \frac{1}{6} = p \quad \text{and} \quad P(X = 0) = \frac{250}{300} = \frac{5}{6} = 1 - p.$$

According to (8.4) and (8.5), the mean (expectation) and variance of X are

$$E(X) = \frac{1}{6} \quad \text{and} \quad \text{Var}(X) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36} \text{ respectively.}$$

8.1.4 Binomial Distribution

Consider n independent trials or repetitions of a Bernoulli experiment. In each trial or repetition, we may observe either A or $\bar{A}$. At the end of the experiment, we have thus observed A between 0 and n times. Suppose we are interested in the probability of A occurring k times, then the binomial distribution is useful.

Example 8.1.3 Consider a coin tossing experiment where a coin is tossed ten times and the event of interest is A = "head". The random variable X "number of heads in 10 experiments" has the possible outcomes $k = 0, 1, \ldots, 10$. A question of interest may be: What is the probability that a head occurs in 7 out of 10 trials; or in 5 out of 10 trials? We assume that the order in which heads (and tails) appear is not of interest, only the total number of heads is of interest.

Questions of this kind are answered by the binomial distribution. This distribution can either be motivated as a repetition of n Bernoulli experiments (as in the above coin tossing example) or by the urn model (see Chap. 5): assume there are M white and $N - M$ black balls in the urn. Suppose n balls are drawn randomly from the urn, the colour of the ball is recorded and the ball is placed back into the urn (sampling with replacement). Let A be the event of interest that a white ball is drawn from the urn. The probability of A is $p = M/N$ (the probability of drawing a black ball is $1 - p = (N - M)/N$). Since the balls are drawn with replacement, these probabilities do not change from draw to draw. Further, let X be the random variable counting the number of white balls drawn from the urn in the n experiments. Since the order of the resulting colours of balls is not of interest in this case, there are $\binom{n}{k}$ combinations where k balls are white and $n - k$ balls are black. Since the balls are drawn with replacement, every outcome of the n experiments is independent of all others. The probability that $X = k, k = 0, 1, \ldots, n$, can therefore be calculated as

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (k = 0, 1, \ldots, n). \tag{8.6}$$

Please note that we can use the product $p^k(1 - p)^{n-k}$ because the draws are independent. The binomial coefficient $\binom{n}{k}$ is necessary to count the number of possible orders of the black and white balls.

Definition 8.1.4 A discrete random variable X is said to follow a binomial distribution with parameters n and p if its PMF is given by (8.6). We also write $X \sim B(n; p)$. The mean and variance of a binomial random variable X are given by

$$\mathrm{E}(X) = np, \tag{8.7}$$

$$\mathrm{Var}(X) = np(1-p). \tag{8.8}$$

Remark 8.1.1 A Bernoulli random variable is therefore $B(1; p)$ distributed.

Example 8.1.4 Consider an unfair coin where the probability of observing a tail (T) is $p(T) = 0.6$. Let us denote tails by "1" and heads by "0". Suppose the coin is tossed three times. In total, there are the $2^3 = 8$ following possible outcomes:

Outcome	$X = x$
1 1 1	3
1 1 0	2
1 0 1	2
0 1 1	2
1 0 0	1
0 1 0	1
0 0 1	1
0 0 0	0

Note that the first outcome, viz. (1, 1, 1) leads to $x = 3$, the next 3 outcomes, viz., (1, 1, 0), (1, 0, 1), (0, 1, 1) obtained by ($= \binom{3}{2}$) lead to $x = 2$, the next 3 outcomes, viz., (1, 0, 0), ((0, 1, 0), (0, 0, 1) obtained by ($= \binom{3}{1}$) lead to $x = 1$, and the last outcome, viz. (0, 0, 0) obtained by ($= \binom{3}{0}$) leads to $x = 0$. We can, for example, calculate

$$P(X = 2) = \binom{3}{2} 0.6^2 (1 - 0.6)^1 = 0.432 \quad (\text{or} \quad 43.2\,\%).$$

Further, the mean and variance of X are

$$\mathrm{E}(X) = np = 3 \cdot 0.6 = 1.8, \quad \text{and} \quad \mathrm{Var}(X) = np(1-p) = 3 \cdot 0.6 \cdot 0.4 = 0.72.$$

Functions for the binomial distribution, as well as many other distributions, are implemented in *R*. For each of these distributions, we can easily determine the density function (PMF, PDF) for given values and parameters, determine the CDF, calculate quantiles and draw random numbers. Appendix A.3 gives more details. Nevertheless, we illustrate the concept of dealing with distributions in *R* in the following example.

Example 8.1.5 Suppose we roll an unfair die 50 times with the probability of a tail $p_{\text{tail}} = 0.6$. We thus deal with a $B(50, 0.6)$ distribution which can be plotted using the `dbinom` command. The prefix `d` stands for "density".

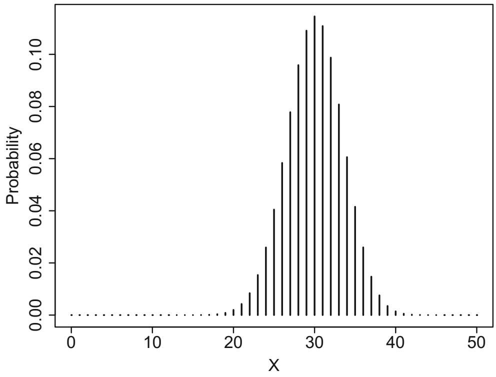


Fig. 8.2 PMF of a $B(50, 0.6)$ distribution

```
n <- 50
p <- 0.6
k <- 0:n
pmf <- dbinom(k,n,p)
plot(k,pmf, type=h)
```

A plot of the PMF of a binomial distribution with $n = 50$ and $p = 0.6$ (i.e. $B(50, 0.6)$) is given in Fig. 8.2.

Note that we can also calculate the CDF with R. We can use the `pbinom(x,n,p)` command, where the prefix `p` stands for probability, to calculate the CDF at any point. For example, suppose we are interested in $P(X \geq 30) = 1 - F(29)$, that is the probability of observing thirty or more tails; then we write

```
1-pbinom(29,50,0.6)
[1] 0.5610349
```

Similarly, we can determine quantiles. For instance, the 80 % quantile q which describes that $P(X \leq q) \geq 0.8$ can be obtained by the `qbinom(q,n,p)` command as follows:

```
qbinom(0.8,50,0.6)
[1] 33
```

If we want to generate 100 random realizations from a $B(50, 0.6)$ distribution we can use the `rbinom` command.

```
rbinom(100,50,0.6)
```

The binomial distribution has some nice properties. One of them is described in the following theorem:

Theorem 8.1.1 *Let $X \sim B(n; p)$ and $Y \sim B(m; p)$ and assume that X and Y are (stochastically) independent. Then*

$$X + Y \sim B(n + m; p). \tag{8.9}$$

This is intuitively clear since we can interpret this theorem as describing the additive combination of two independent binomial experiments with n and m trials, with equal probability p, respectively. Since every binomial experiment is a series of independent Bernoulli experiments, this is equivalent to a series of $n + m$ independent Bernoulli trials with constant success probability p which in turn is equivalent to a binomial distribution with $n + m$ trials.

8.1.5 Poisson Distribution

Consider a situation in which the number of events is very large and the probability of success is very small: for example, the number of alpha particles emitted by a radioactive substance entering a particular region in a given short time interval. Note that the number of emitted alpha particles is very high but only a few particles are transmitted through the region in a given short time interval. Some other examples where Poisson distributions are useful are the number of flu cases in a country within one year, the number of tropical storms within a given area in one year, or the number of bacteria found in a biological investigation.

Definition 8.1.5 A discrete random variable X is said to follow a Poisson distribution with parameter $\lambda > 0$ if its PMF is given by

$$P(X = x) = \frac{\lambda^x}{x!} \exp(-\lambda) \quad (x = 0, 1, 2, \ldots). \tag{8.10}$$

We also write $X \sim Po(\lambda)$. The mean and variance of a Poisson random variable are identical:

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

Example 8.1.6 Suppose a country experiences $X = 4$ tropical storms on average per year. Then the probability of suffering from only two tropical storms is obtained by using the Poisson distribution as

$$P(X = 2) = \frac{\lambda^x}{x!} \exp(-\lambda) = \frac{4^2}{2!} \exp(-4) = 0.146525.$$

If we are interested in the probability that not more than 2 storms are experienced, then we can apply rules (7.7)–(7.13) from Chap. 7: $P(X \le 2) = P(X = 2) + P(X = 1) + P(X = 0) = F(2) = 0.2381033$. We can calculate $P(X = 1)$ and $P(X = 0)$ from (8.10) or using R. Similar to Example 8.1.5, we use the prefix `d` to obtain the PMF and the prefix `p` to work with the CDF, i.e. we can use `dpois(x,`λ`)` and `ppois(x,`λ`)` to determine $P(X = x)$ and $P(X \le x)$, respectively.

```
dpois(2,4) + dpois(1,4) + dpois(0,4)
[1] 0.2381033
ppois(2,4)
[1] 0.2381033
```

8.1.6 Multinomial Distribution

We now consider random experiments where k distinct or disjoint events $A_1, A_2, \ldots, A_k$ can occur with probabilities $p_1, p_2, \ldots, p_k$, respectively, with the restriction $\sum_{j=1}^{k} p_j = 1$. For example, if eight parties compete in a political election, we may be interested in the probability that a person votes for party A_j, $j = 1, 2, \ldots, 8$. Similarly one might be interested in the probability whether tuberculosis is detected in the lungs (A_1), in other organs (A_2), or both (A_3). Practically, we often use the multinomial distribution to model the distribution of categorical variables. This can be interpreted as a generalization of the binomial distribution (where only two distinct events can occur) to the situation where more than two events or outcomes can occur. If the experiment is repeated n times independently, we are interested in the probability that

$$A_1 \quad \text{occurs} \quad n_1\text{-times}, \; A_2 \quad \text{occurs} \quad n_2\text{-times}, \ldots, A_k \quad \text{occurs} \quad n_k\text{-times}$$

with $\sum_{j=1}^{k} n_j = n$. Since several events can occur, the outcome of one (of the n) experiments is conveniently described by binary indicator variables. Let V_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, k$, denote the event "A_j is observed in experiment i", i.e.

$$V_{ij} = \begin{cases} 1 & \text{if} \quad A_j \quad \text{occurs in experiment } i \\ 0 & \text{if} \quad A_j \quad \text{does not occur in experiment } i \end{cases}$$

with probabilities $P(V_{ij} = 1) = p_j$, $j = 1, 2, \ldots, k$; then, the outcome of one experiment is a vector of length k,

$$V_i = (V_{i1}, \ldots, V_{ij}, \ldots, V_{ik}) = (0, \ldots, 1, \ldots, 0),$$

with "1" being present in only one position, i.e. in position j, if A_j occurs in experiment i. Now, define (for each $j = 1, \ldots, k$) $X_j = \sum_{i=1}^{n} V_{ij}$. Then, X_j is counting how often event A_j was observed in the n independent experiments (i.e. how often V_{ij} was 1 in the n experiments).

Definition 8.1.6 The random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ is said to follow a **multinomial distribution** if its PMF is given as

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \tag{8.11}$$

with the restrictions $\sum_{j=1}^{k} n_j = n$ and $\sum_{j=1}^{k} p_j = 1$. We also write $\mathbf{X} \sim M(n; p_1, \ldots, p_k)$. The mean of $\mathbf{X}$ is the (component-wise) vector

$$\begin{aligned} \mathrm{E}(\mathbf{X}) &= (\mathrm{E}(X_1), \mathrm{E}(X_2), \ldots, \mathrm{E}(X_k)) \\ &= (np_1, np_2, \ldots, np_k). \end{aligned}$$

The (i, j)th element of the covariance matrix $V(\mathbf{X})$ is

$$\mathrm{Cov}(X_i, X_j) = \begin{cases} np_i(1-p_i) & \text{if } i = j \\ -np_i p_j & \text{if } i \neq j. \end{cases}$$

Remark 8.1.2 Due to the restriction that $\sum_{j=1}^{k} n_j = \sum_{j=1}^{k} X_j = n$, $X_1, \ldots, X_k$ are not stochastically independent which is reflected by the negative covariance. This is also intuitively clear: if one X_j gets higher, another $X_{j'}$, $j \neq j'$, has to become lower to satisfy the restrictions.

We use the multinomial distribution to describe the randomness of categorical variables. Suppose we are interested in the variable "political party"; there might be eight political parties, and we could thus summarize this variable by eight binary variables, each of them describing the event of party A_j, $j = 1, 2, \ldots, 8$, being voted for. In this sense, $\mathbf{X} = (X_1, X_2, \ldots, X_8)$ follows a multinomial distribution.

Example 8.1.7 Consider a simple example of the urn model. The urn contains 50 balls of three colours: 25 red balls, 15 white balls, and 10 black balls. The balls are drawn from the urn with replacement. The balls are placed back into the urn after every draw, which means the draws are independent. Therefore, the probability of drawing a red ball in every draw is $p_1 = \frac{25}{50} = 0.5$. Analogously, $p_2 = 0.3$ (for white balls) and $p_3 = 0.2$ (for black balls). Consider $n = 4$ draws. The probability of the random event of drawing "2 red balls, 1 white ball, and 1 black ball" is:

$$P(X_1 = 2, X_2 = 1, X_3 = 1) = \frac{4!}{2!1!1!}(0.5)^2(0.3)^1(0.2)^1 = 0.18. \quad (8.12)$$

We would have obtained the same result in *R* using the `dmultinom` function:

```
dmultinom(c(2,1,1),prob=c(0.5,0.3,0.2))
```

This example demonstrates that the multinomial distribution relates to an experiment with replacement and without considering the order of the draws. Instead of the urn model, consider another example where we may interpret these three probabilities as probabilities of voting for candidate A_j, $j = 1, 2, 3$, in an election. Now, suppose we ask four voters about their choice, then the probability of candidate A_1 receiving 2 votes, candidate A_2 receiving 1 vote, and candidate A_3 receiving 1 vote is 18 % as calculated in (8.12).

Remark 8.1.3 In contrast to most of the distributions, the CDF of the multinomial distribution, i.e. the function calculating $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k)$, is not contained in the base *R*-distribution. Please note that for $k = 2$, the multinomial distribution reduces to the binomial distribution.

8.1.7 Geometric Distribution

Consider a situation in which we are interested in determining how many independent Bernoulli trials are needed until the event of interest occurs for the first time. For instance, we may be interested in how many tickets to buy in a raffle until we win for the first time, or how many different drugs to try to successfully tackle a severe migraine, etc. The geometric distribution can be used to determine the probability that the event of interest happens at the kth trial for the first time.

Definition 8.1.7 A discrete random variable X is said to follow a geometric distribution with parameter p if its PMF is given by

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, 3, \ldots \tag{8.13}$$

The mean (expectation) and variance are given by $\mathrm{E}(X) = 1/p$ and $\mathrm{Var}(X) = 1/p(1/p - 1)$, respectively.

Example 8.1.8 Let us consider an experiment where a coin is tossed until "head" is obtained for the first time. The probability of getting a head is $p = 0.5$ for each toss. Using (8.13), we can determine the following probabilities:

$$\begin{aligned}
P(X = 1) &= 0.5 \\
P(X = 2) &= 0.5(1 - 0.5) = 0.25 \\
P(X = 3) &= 0.5(1 - 0.5)^2 = 0.125 \\
P(X = 4) &= 0.5(1 - 0.5)^3 = 0.0625 \\
\ldots \quad &\ldots
\end{aligned}$$

Using the command structure for obtaining PMF's in *R* (Appendix A as well as Examples 8.1.5 and 8.1.6), we can determine the latter probability of $P(X = 4)$ as follows:

```
dgeom(3,0.5)
```

Note that the definition of X in R slightly differs from our definition. In R, k is the number of failures before the first success. This means we need to specify $k - 1$ in the `dgeom` function rather than k. The mean and variance for this setting are

$$\mathrm{E}(X) = \frac{1}{0.5} = 2; \quad \mathrm{Var}(X) = \frac{1}{0.5}\left(\frac{1}{0.5} - 1\right) = 2.$$

8.1.8 Hypergeometric Distribution

We can again use the urn model to motivate another distribution, the hypergeometric distribution. Consider an urn with We randomly draw n balls without replacement,

M	white balls
$N-M$	black balls
N	total balls

i.e. we do not place a ball back into the urn once it is drawn. The order in which the balls are drawn is assumed to be of no interest; only the number of drawn white balls is of relevance. We define the following random variable

$$X: \text{“number of white balls } (x) \text{ among the } n \text{ drawn balls”}.$$

To be more precise, among the n drawn balls, x are white and $n-x$ are black. There are $\binom{M}{x}$ possibilities to choose x white balls from the total of M white balls, and analogously, there are $\binom{N-M}{n-x}$ possibilities to choose $(n-x)$ black balls from the total of $N-M$ black balls. In total, we draw n out of N balls. Recall the probability definition of Laplace as the number of simple favourable events divided by all possible events. The number of combinations for all possible events is $\binom{N}{n}$; the number of favourable events is $\binom{M}{x}\binom{N-M}{n-x}$ because we draw, independent of each other, x out of M balls and $n-x$ out of $N-M$ balls. Hence, the PMF of the hypergeometric distribution is

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{8.14}$$

for $x \in \{\max(0, n-(N-M)), \ldots, \min(n, M)\}$.

Definition 8.1.8 A random variable X is said to follow a **hypergeometric** distribution with parameters n, M, N, i.e. $X \sim H(n, M, N)$, if its PMF is given by (8.14).

Example 8.1.9 The German national lottery draws 6 out of 49 balls from a rotating bowl. Each ball is associated with a number between 1 and 49. A simple bet is to choose 6 numbers between 1 and 49. If 3 or more chosen numbers correspond to the numbers drawn in the lottery, then one wins a certain amount of money. What is the probability of choosing 4 correct numbers? We can utilize the hypergeometric distribution with $x=4$, $M=6$, $N=49$, and $n=6$ to calculate such probabilities. The interpretation is that we “draw” (i.e. bet on) 4 out of the 6 winning balls and “draw” (i.e. bet on) another 2 out of the remaining 43 (49 − 6) losing balls. In total, we draw 6 out of 49 balls. Calculating the number of the favourable combinations and all possible combinations leads to the application of the hypergeometric distribution as follows:

$$P(X=4) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} \approx 0.001 \text{ (or } \ 0.1\,\%).$$

We would have obtained the same results using the `dhyper` command. Its arguments are `x, M, N, n`, and thus, we specify

```r
dhyper(4,6,43,6)
```

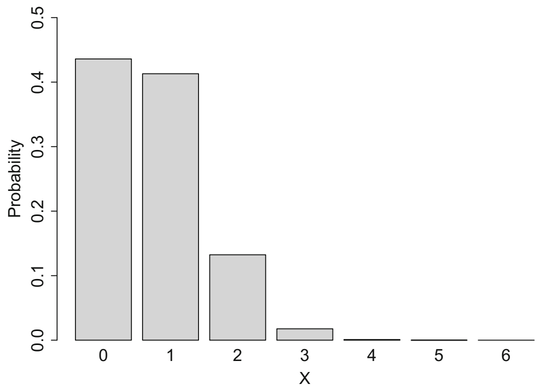


Fig. 8.3 The $H(6, 43, 6)$ distribution

The $H(6, 43, 6)$ distribution is also visualized in Fig. 8.3. It is evident that the cumulative probability of choosing 2 or fewer correct numbers is greater than 0.9 (or 90 %), but it is very unlikely to have 3 or more numbers right. This may explain why the national lottery pays out money only for 3 or more correct numbers.

8.2 Standard Continuous Distributions

Now, we discuss some standard probability distributions of (absolute) continuous random variables. Characteristics of continuous random variables are that the number of possible outcomes is uncountably infinite and that they have a continuous distribution function $F(x)$. It follows that the point probabilities are zero, i.e. $P(X = x) = 0$. Further, we assume a unique density function f exists, such that $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$.

8.2.1 Continuous Uniform Distribution

A continuous analogue to the discrete uniform distribution is the continuous uniform distribution on a closed interval in $\mathbb{R}$.

Definition 8.2.1 A continuous random variable X is said to follow a (continuous) **uniform distribution** in the interval $[a, b]$, i.e. $X \sim U(a, b)$, if its probability density function (PDF) is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \quad (a < b) \\ 0 & \text{otherwise.} \end{cases}$$

The mean and variance of $X \sim U(a, b)$ are

$$\mathrm{E}(X) = \frac{a+b}{2} \quad \text{and} \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12},$$

respectively.

Example 8.2.1 Suppose a train arrives at a subway station regularly every 10 min. If a passenger arrives at the station without knowing the timetable, then the waiting time to catch the train is uniformly distributed with density

$$f(x) = \begin{cases} \frac{1}{10} & \text{if } 0 \leq x \leq 10 \\ 0 & \text{otherwise.} \end{cases}$$

The "average" waiting time is $\mathrm{E}(X) = (10 + 0)/2 = 5$ min. The probability of waiting for the train for less than 3 min is obviously 0.3 (or 30 %) which can be calculated in *R* using the `punif(x,a,b)` command (see also Appendix A.3):

```r
punif(3,0,10)
```

8.2.2 Normal Distribution

The normal distribution is one of the most important distributions used in statistics. The name was given by Carl Friedrich Gauss (1777–1855), a German mathematician, astronomer, geodesist, and physicist who observed that measurements in geodesy and astronomy randomly deviate in a symmetric way from their true values. The normal distribution is therefore also often called a Gaussian distribution.

Definition 8.2.2 A random variable X is said to follow a **normal distribution** with parameters μ and σ^2 if its PDF is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right); \qquad -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0. \tag{8.15}$$

We write $X \sim N(\mu, \sigma^2)$. The mean and variance of X are

$$\mathrm{E}(X) = \mu\,; \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2,$$

respectively. If $\mu = 0$ and $\sigma^2 = 1$, then X is said to follow a ***standard*** **normal distribution**, $X \sim N(0, 1)$. The PDF of a standard normal distribution is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}); \quad -\infty < x < \infty.$$

The density of a normal distribution has its maximum (see Fig. 8.4) at $x = \mu$. The density is also symmetric around μ. The inflexion points of the density are at $(\mu - \sigma)$ and $(\mu + \sigma)$ (Fig. 8.4). A lower σ indicates a higher concentration around the mean μ. A higher σ indicates a flatter density (Fig. 8.5).

The cumulative distribution function of $X \sim N(\mu, \sigma^2)$ is

$$F(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t \tag{8.16}$$

which is often denoted as $\Phi(x)$. The value of $\Phi(x)$ for various values of x can be obtained in *R* following the rules introduced in Appendix A.3. For example,

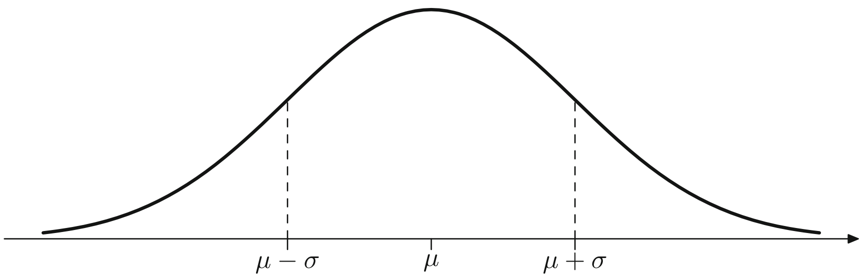


Fig. 8.4 PDF of a normal distribution*

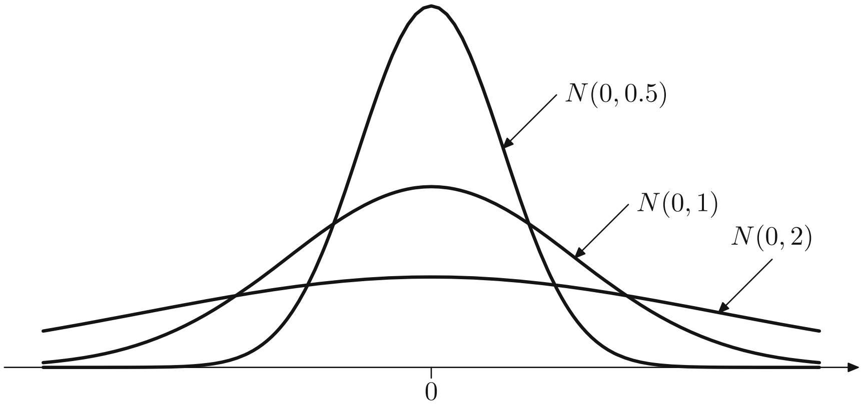


Fig. 8.5 PDF of $N(0, 2)$, $N(0, 1)$ and $N(0, 0.5)$ distributions*

```
pnorm(1.96, mean = 0, sd = 1)
```

calculates $\Phi(1.96)$ as approximately 0.975. This means, for a standard normal distribution the probability $P(X \leq 1.96) \approx 0.975$.

Remark 8.2.1 There is no explicit formula to solve the integral in Eq. (8.16). It has to be solved by numerical (or computational) methods. This is the reason why CDF tables are presented in almost all statistical textbooks, see Table C.1 in Appendix C.

Example 8.2.2 An orange farmer sells his oranges in wooden boxes. The weights of the boxes vary and are assumed to be normally distributed with $\mu = 15$ kg and $\sigma^2 = \frac{9}{4}$ kg^2. The farmer wants to avoid customers being unsatisfied because the boxes are too low in weight. He therefore asks the following question: What is the probability that a box with a weight of less than 13 kg is sold? Using the `pnorm(x,` μ, σ`)` command in *R*, we get

```
pnorm(13,15,sqrt(9/4))
[1] 0.09121122
```

To calculate the probability in Example 8.2.2 manually, we first have to introduce some theoretical results.

Calculation rules for normal random variables.

Let $X \sim N(\mu, \sigma^2)$. Using the transformation

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1), \tag{8.17}$$

every normally distributed random variable can be transformed into a *standard* normal random variable. We call this transformation the Z-transformation. We can use this transformation to derive convenient calculation rules. The probability for $X \leq b$ is

$$P(X \leq b) = P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P\left(Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right). \tag{8.18}$$

Consequently, the probability for $X > a$ is

$$P(X > a) = 1 - P(X \leq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right). \tag{8.19}$$

The probability that X realizes a value in the interval $[a, b]$ is

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \tag{8.20}$$

Because of the symmetry of the probability density function $\phi(x)$ around its mean 0, the following equation holds for the distribution function $\Phi(x)$ of a standard normal random variable for any value a:

$$\Phi(-a) = 1 - \Phi(a). \tag{8.21}$$

It follows that $P(-a < Z < a) = 2 \cdot \Phi(a) - 1$, see also Fig. 8.6.

Example 8.2.3 Recall Example 8.2.2 where a farmer sold his oranges. He was interested in $P(X \leq 13)$ for $X \sim N(15, 9/4)$. Using (8.17), we get

$$\begin{aligned} P(X \leq 13) &= \Phi\left(\frac{13 - 15}{\frac{3}{2}}\right) \\ &= \Phi\left(-\frac{4}{3}\right) = 1 - \Phi\left(\frac{4}{3}\right) \approx 0.091 \text{ (or } 9.1\,\%). \end{aligned}$$

To obtain $\Phi(4/3) \approx 90.9\,\%$, we could either use *R* (`pnorm(4/3)`) or use Table C.1.

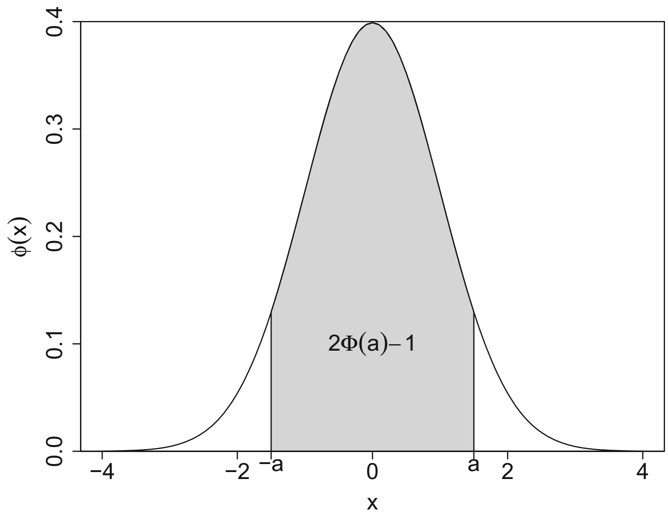


Fig. 8.6 Distribution function of the standard normal distribution

Distribution of the Arithmetic Mean.

Assume that $X \sim N(\mu, \sigma^2)$. Consider a random sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of independent and identically distributed random variables X_i with $X_i \sim N(\mu, \sigma^2)$. Then, the arithmetic mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ follows a normal distribution with mean

$$\mathrm{E}(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \mu$$

and variance

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n} \tag{8.22}$$

where $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$. In summary, we get

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

Remark 8.2.2 In fact, in Eq. (8.22), we have used the fact that the sum of normal random variables also follows a normal distribution, i.e.

$$(X_1 + X_2) \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right).$$

This result can be generalized to n (not necessarily identically distributed but independent) normal random variables. In fact, it holds that if $X_1, X_2, \ldots, X_n$ are independent normal variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, then for any real numbers $a_1, a_2, \ldots, a_n$, it holds that

$$(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \sim N\left(a_1\mu_1 + a_2\mu_2 + \cdots a_n\mu_n, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots a_n^2\sigma_n^2\right).$$

In general, it cannot be taken for granted that the sum of two random variables follows the same distribution as the two variables themselves. As an example, consider the sum of two independent uniform distributions with $X_1 \sim U[0, 10]$ and $X_2 \sim U[20, 30]$. It holds that $\mathrm{E}(X_1 + X_2) = \mathrm{E}(X_1) + \mathrm{E}(X_2)$ and $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$, but $X_1 + X_2$ is obviously not uniformly distributed.

8.2.3 Exponential Distribution

The exponential distribution is useful in many situations, for example when one is interested in the waiting time, or lifetime, until an event of interest occurs. If we assume that the future lifetime is independent of the lifetime that has already taken place (i.e. no "ageing" process is working), the waiting times can be considered to be exponentially distributed.

Definition 8.2.3 A random variable X is said to follow an exponential distribution with parameter $\lambda > 0$ if its PDF is given by

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } \; x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8.23}$$

We write $X \sim Exp(\lambda)$. The mean and variance of an exponentially distributed random variable X are

$$\mathrm{E}(X) = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2},$$

respectively. The CDF of the exponential distribution is given as

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{if } \; x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8.24}$$

Note, that $P(X > x) = 1 - F(x) = \exp(-\lambda x)$ $(x \geq 0)$. An interesting property of the exponential distribution is its **memorylessness**: if time t has already been reached, the probability of reaching a time greater than $t + \Delta$ does not depend on t. This can be written as

$$P(X > t + \Delta | X > t) = P(X > \Delta) \quad t, \Delta > 0.$$

The result can be derived using basic probability rules as follows:

$$\begin{aligned} P(X > t + \Delta | X > t) &= \frac{P(X > t + \Delta \text{ and } X > t)}{P(X > t)} = \frac{P(X > t + \Delta)}{P(X > t)} \\ &= \frac{\exp[-\lambda(t + \Delta)]}{\exp[-\lambda t]} = \exp[-\lambda \Delta] \\ &= 1 - F(\Delta) = P(X > \Delta). \end{aligned}$$

For example, suppose someone stands in a supermarket queue for t minutes. Say the person forgot to buy milk, so she leaves the queue, gets the milk, and stands in the queue again. If we use the exponential distribution to model the waiting time, we say that it does not matter what time it is: the random variable "waiting time from standing in the queue until paying the bill" is not influenced by how much time has elapsed already; it does not matter if we queued before or not. Please note that the memorylessness property is shared by the geometric and the exponential distributions.

There is also a relationship between the Poisson and the exponential distribution:

Theorem 8.2.1 *The number of events Y occurring within a continuum of time is Poisson distributed with parameter λ if and only if the time between two events is exponentially distributed with parameter λ.*

The continuum of time depends on the problem at hand. It may be a second, a minute, 3 months, a year, or any other time period.

Example 8.2.4 Let Y be the random variable which counts the "number of accesses per second for a search engine". Assume that Y is Poisson distributed with parameter $\lambda = 10$ ($\mathrm{E}(Y) = 10$, $\mathrm{Var}(Y) = 10$). The random variable X, "waiting time until the next access", is then exponentially distributed with parameter $\lambda = 10$. We therefore get

$$\mathrm{E}(X) = \frac{1}{10}, \quad \mathrm{Var}(X) = \frac{1}{10^2}.$$

In this example, the continuum is 1 s. The expected number of accesses per second is therefore $\mathrm{E}(Y) = 10$, and the expected waiting time between two accesses is $\mathrm{E}(X) = 1/10$ s. The probability of experiencing a waiting time of less than 0.1 s is

$$F(0.1) = 1 - \exp(-\lambda x) = 1 - \exp(-10 \cdot 0.1) \approx 0.63.$$

In R, we can obtain the same result as

```
pexp(0.1,10)
[1] 0.6321206
```

8.3 Sampling Distributions

All the distributions introduced in this chapter up to now are motivated by practical applications. However, there are theoretical distributions which play an important role in the construction and development of various statistical tools such as those introduced in Chaps. 9–11. We call these distributions "sampling distributions". Now, we discuss the χ^2-, t-, and F-distributions.

8.3.1 χ^2-Distribution

Definition 8.3.1 Let $Z_1, Z_2 \ldots, Z_n$ be n independent and identically $N(0, 1)$-distributed random variables. The sum of their squares, $\sum_{i=1}^n Z_i^2$, is then χ^2-**distributed** with n degrees of freedom and is denoted as χ_n^2. The PDF of the χ^2-distribution is given in Eq. (C.7) in Appendix C.3.

The χ^2-distribution is not symmetric. A χ^2-distributed random variable can only realize values greater than or equal to zero. Figure 8.7a shows the χ^2_1-, χ^2_2-, and χ^2_5-distributions. It can be seen that the "degrees of freedom" specify the shape of the distribution. Their interpretation and meaning will nevertheless become clearer in the following chapters. The quantiles of the CDF of different χ^2-distributions can be obtained in *R* using the `qchisq(p,df)` command. They are also listed in Table C.3 for different values of n.

Theorem 8.3.1 *Consider two independent random variables which are χ^2_m- and χ^2_n-distributed, respectively. The sum of these two random variables is χ^2_{n+m}-distributed.*

An important example of a χ^2-distributed random variable is the sample variance (S^2_X) of an i.i.d. sample of size n from a normally distributed population, i.e.

$$\frac{(n-1)S^2_X}{\sigma^2} \sim \chi^2_{n-1}. \tag{8.25}$$

8.3.2 *t*-Distribution

Definition 8.3.2 Let X and Y be two independent random variables where $X \sim N(0,1)$ and $Y \sim \chi^2_n$. The ratio

$$\frac{X}{\sqrt{Y/n}} \sim t_n$$

follows a ***t*-distribution** (Student's t-distribution) with n degrees of freedom. The PDF of the t-distribution is given in Eq. (C.8) in Appendix C.3.

Figure 8.7b visualizes the t_1-, t_5-, and t_{30}-distributions. The quantiles of different t-distributions can be obtained in *R* using the `qt(p,df)` command. They are also listed in Table C.2 for different values of n.

An application of the t-distribution is the following: if we draw a sample of size n from a normal population $N(\mu, \sigma^2)$ and calculate the arithmetic mean $\bar{X}$ and the sample variance S^2_X, then the following theorem holds:

Theorem 8.3.2 (Student's theorem) *Let* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *with* $X_i \overset{iid.}{\sim} N(\mu, \sigma^2)$*. The ratio*

$$\frac{(\bar{X}-\mu)\sqrt{n}}{S_X} = \frac{(\bar{X}-\mu)\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum_{i+1}^{n}(X_i-\bar{X})^2}} \sim t_{n-1} \tag{8.26}$$

is then t-distributed with $n-1$ degrees of freedom.

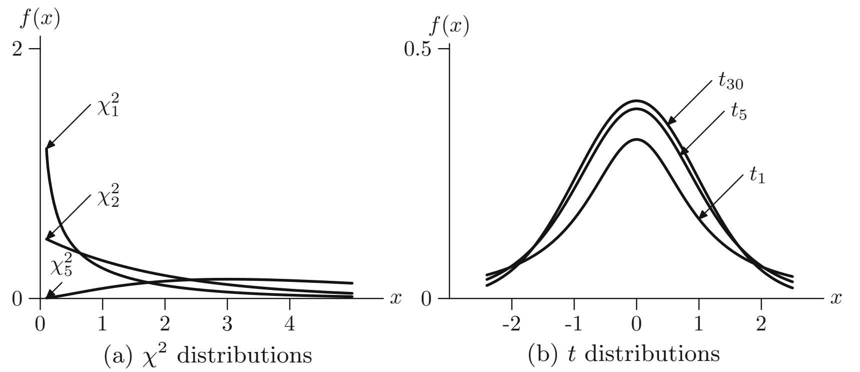


Fig. 8.7 Probability density functions of χ^2 and t distributions*

8.3.3 *F*-Distribution

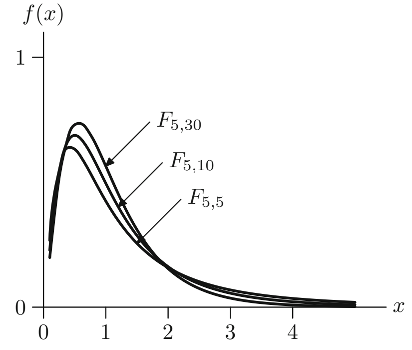
Definition 8.3.3 Let X and Y be independent χ^2_m and χ^2_n-distributed random variables, then the distribution of the ratio

$$\frac{X/m}{Y/n} \sim F_{m,n} \tag{8.27}$$

follows the **Fisher F-distribution** with (m, n) degrees of freedom. The PDF of the F-distribution is given in Eq. (C.9) in Appendix C.3.

If X is a χ^2_1-distributed random variable, then the ratio (8.27) is $F_{1,n}$-distributed. The square root of this ratio is t_n-distributed since the square root of a χ^2_1-distributed random variable is $N(0, 1)$-distributed. If W is F-distributed, $F_{m,n}$, then $1/W$ is $F_{n,m}$-distributed. Figure 8.8 visualizes the $F_{5,5}$, $F_{5,10}$ and $F_{5,30}$ distributions. The

Fig. 8.8 Probability density functions for different F-distributions*



quantiles of different F-distributions can be obtained in R using the `qf(p,df1,df2)` command.

One application of the F-distribution relates to the ratio of two sample variances of two independent samples of size m and n, where each sample is an i.i.d. sample from a normal population, i.e. $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$. For the sample variances $S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$ and $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ from the two populations, the ratio

$$\frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

is F-distributed with $(m-1)$ degrees of freedom in the numerator and $(n-1)$ degrees of freedom in the denominator.

8.4 Key Points and Further Issues

Note:

✓ Examples of different distributions are:

Distribution	Example
Uniform	Rolling a die (discrete)
	Waiting for a train (continuous)
Bernoulli	Any binary variable such as gender
Binomial	Number of "heads" when tossing a coin n times
Poisson	Number of particles emitted by a radioactive source
	entering a small area in a given time interval
Multinomial	Categorical variables such as "party voted for"
Geometric	Number of raffle tickets until first ticket wins
Hypergeometric	National lotteries; Fisher's test, see p. 428
Normal	Height or weight of women (men)
Exponential	Survival time of a PC
χ^2	Sample variance; χ^2 tests, see p. 235 ff
t	Confidence interval for the mean, see p. 197
F	Tests in the linear model, see p. 272

Note:

✓ One can use R to determine values of densities (PDF/PMF), cumulative probability distribution functions (CDF), quantiles of the CDF, and random numbers:

First letter	Function	Further letters	Example
`d`	Density	`distribution name`	`dnorm`
`p`	Probability	`distribution name`	`pnorm`
`q`	Quantiles	`distribution name`	`qnorm`
`r`	Random number	`distribution name`	`rnorm`

We encourage the use of R to obtain quantiles of sampling distributions, but Tables C.1–C.3 also list some of them.

✓ In this chapter, we assumed the parameters such as μ, σ, λ, and others to be known. In Chap. 9, we will propose how to estimate these parameters from the data. In Chap. 10, we test statistical hypotheses about these parameters.

✓ For n i.i.d. random variables $X_1, X_2, \ldots, X_n$, the arithmetic mean $\bar{X}$ converges to a $N(\mu, \sigma^2/n)$ distribution as n tends to infinity. See Appendix C.3 as well as Exercise 8.11 for the Theorem of Large Numbers and the Central Limit Theorem, respectively.

8.5 Exercises

Exercise 8.1 A company producing cereals offers a toy in every sixth cereal package in celebration of their 50th anniversary. A father immediately buys 20 packages.

(a) What is the probability of finding 4 toys in the 20 packages?
(b) What is the probability of finding no toy at all?
(c) The packages contain three toys. What is the probability that among the 5 packages that are given to the family's youngest daughter, she finds two toys?

Exercise 8.2 A study on breeding birds collects information such as the length of their eggs (in mm). Assume that the length is normally distributed with $\mu = 42.1$ mm and $\sigma^2 = 20.8^2$. What is the probability of

(a) finding an egg with a length greater than 50 mm?
(b) finding an egg between 30 and 40 mm in length?

Calculate the results both manually and by using R.

Exercise 8.3 A dodecahedron is a die with 12 sides. Suppose the numbers on the die are 1–12. Consider the random variable X which describes which number is shown after rolling the die once. What is the distribution of X? Determine $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Exercise 8.4 Felix states that he is able to distinguish a freshly ground coffee blend from an ordinary supermarket coffee. One of his friends asks him to taste 10 cups of coffee and tell him which coffee he has tasted. Suppose that Felix has actually no clue about coffee and simply guesses the brand. What is the probability of at least 8 correct guesses?

Exercise 8.5 An advertising board is illuminated by several hundred bulbs. Some of the bulbs are fused or smashed regularly. If there are more than 5 fused bulbs on a day, the owner of the board replaces them, otherwise not. Consider the following data collected over a month which captures the number of days (n_i) on which i bulbs were broken:

Fused bulbs	0	1	2	3	4	5
n_i	6	8	8	5	2	1

(a) Suggest an appropriate distribution for X: "number of broken bulbs per day".
(b) What is the average number of broken bulbs per day? What is the variance?
(c) Determine the probabilities $P(X = x)$ using the distribution you chose in (a) and using the average number of broken bulbs you calculated in (b). Compare the probabilities with the proportions obtained from the data.
(d) Calculate the probability that at least 6 bulbs are fused, which means they need to be replaced.
(e) Consider the random variable Y: "time until next bulb breaks". What is the distribution of Y?
(f) Calculate and interpret $\mathrm{E}(Y)$.

Exercise 8.6 Marco's company organizes a raffle at an end-of-year function. There are 4000 raffle tickets to be sold, of which 500 win a prize. The price of each ticket is €1.50. The value of the prizes, which are mostly electrical appliances produced by the company, varies between €80 and €250, with an average value of €142.

(a) Marco wants to have a 99 % guarantee of receiving three prizes. How much money does he need to spend? Use *R* to solve the question.
(b) Use *R* to plot the function which describes the relationship between the number of tickets bought and the probability of winning at least three prizes.
(c) Given the value of the prizes and the costs of the tickets, is it worth taking part in the raffle?

Exercise 8.7 A country has a ratio between male and female births of 1.05 which means that 51.22 % of babies born are male.

(a) What is the probability for a mother that the first girl is born during the first three births?
(b) What is the probability of getting 2 girls among 4 babies?

Exercise 8.8 A fishermen catches, on average, three fish in an hour. Let Y be a random variable denoting the number of fish caught in one hour and let X be the time interval between catching two fishes. We assume that X follows an exponential distribution.

(a) What is the distribution of Y?
(b) Determine $\mathrm{E}(Y)$ and $\mathrm{E}(X)$.
(c) Calculate $\mathrm{P}(Y = 5)$ and $\mathrm{P}(Y < 1)$.

Exercise 8.9 A restaurant sells three different types of dessert: chocolate, brownies, yogurt with seasonal fruits, and lemon tart. Years of experience have shown that the probabilities with which the desserts are chosen are 0.2, 0.3, and 0.5, respectively.

(a) What is the probability that out of 5 guests, 2 guests choose brownies, 1 guest chooses yogurt, and the remaining 2 guests choose lemon tart?
(b) Suppose two out of the five guests are known to always choose lemon tart. What is the probability of the others choosing lemon tart as well?
(c) Determine the expectation and variance assuming a group of 20 guests.

Exercise 8.10 A reinsurance company works on a premium policy for natural disasters. Based on experience, it is known that W = "number of natural disasters from October to March" (winter) is Poisson distributed with $\lambda_W = 4$. Similarly, the random variable S = "number of natural disasters from April to September" (summer) is Poisson distributed with $\lambda_S = 3$. Determine the probability that there is at least 1 disaster during both summer and winter based on the assumption that the two random variables are independent.

Exercise 8.11 Read Appendix C.3 to learn about the Theorem of Large Numbers and the Central Limit Theorem.

(a) Draw 1000 realizations from a standard normal distribution using *R* and calculate the arithmetic mean. Repeat this process 1000 times. Evaluate the distribution of the arithmetic mean by drawing a kernel density plot and by calculating the mean and variance of it.

(b) Repeat the procedure in (a) with an exponential distribution with $\lambda = 1$. Interpret your findings in the light of the Central Limit Theorem.
(c) Repeat the procedure in (b) using 10,000 rather than 1000 realizations. How do the results change and why?

→ Solutions to all exercises in this chapter can be found on p. 375

**Source* Toutenburg, H., Heumann, C., *Induktive Statistik*, 4th edition, 2007, Springer, Heidelberg