

3.6: Distribution and Quantile Functions

As usual, our starting point is a random experiment modeled by a with probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So to review, Ω is the set of outcomes, \mathcal{F} is the collection of events, and \mathbb{P} is the probability measure on the sample space (Ω, \mathcal{F}) . In this section, we will study two types of functions that can be used to specify the distribution of a real-valued random variable.

Distribution Functions

Definition

Suppose that X is a random variable with values in \mathbb{R} . The (cumulative) *distribution function* of X is the function $F : \mathbb{R} \rightarrow [0, 1]$ defined by

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R} \quad (3.6.1)$$

The distribution function is important because it makes sense for any type of random variable, regardless of whether the distribution is discrete, continuous, or even mixed, and because it completely determines the distribution of X . In the picture below, the light shading is intended to represent a continuous distribution of probability, while the darker dots represents points of positive probability; $F(x)$ is the total probability mass to the left of (and including) x .



Figure 3.6.1: $F(x)$ is the total probability to the left of (and including) x

Basic Properties

A few basic properties completely characterize distribution functions. Notationally, it will be helpful to abbreviate the limits of F from the left and right at $x \in \mathbb{R}$, and at ∞ and $-\infty$ as follows:

$$F(x^+) = \lim_{t \downarrow x} F(t), \quad F(x^-) = \lim_{t \uparrow x} F(t), \quad F(\infty) = \lim_{t \rightarrow \infty} F(t), \quad F(-\infty) = \lim_{t \rightarrow -\infty} F(t) \quad (3.6.2)$$

Suppose that F is the distribution function of a real-valued random variable X .

1. F is increasing: if $x \leq y$ then $F(x) \leq F(y)$.
2. $F(x^+) = F(x)$ for $x \in \mathbb{R}$. Thus, F is *continuous from the right*.
3. $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$. Thus, F has *limits from the left*.
4. $F(-\infty) = 0$.
5. $F(\infty) = 1$.

Proof

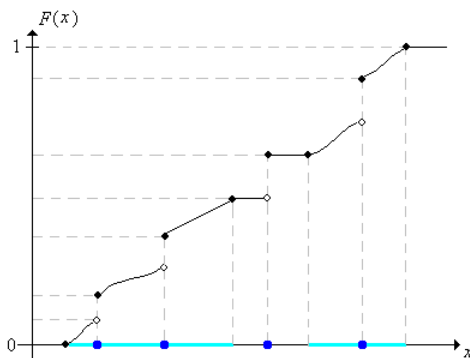


Figure 3.6.2: The graph of a distribution function

The following result shows how the distribution function can be used to compute the probability that X is in an interval. Recall that a probability distribution on \mathbb{R} is completely determined by the probabilities of *intervals*; thus, the *distribution function* determines

the *distribution* of X .

Suppose again that F is the distribution function of a real-valued random variable X . If $a, b \in \mathbb{R}$ with $a < b$ then

1. $\mathbb{P}(X = a) = F(a) - F(a^-)$
2. $\mathbb{P}(a < X \leq b) = F(b) - F(a)$
3. $\mathbb{P}(a < X < b) = F(b^-) - F(a)$
4. $\mathbb{P}(a \leq X \leq b) = F(b) - F(a^-)$
5. $\mathbb{P}(a \leq X < b) = F(b^-) - F(a^-)$

Proof

These results follow from the [definition](#), the [basic properties](#), and the difference rule: $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ if A, B are events and $A \subseteq B$.

1. $\{X = a\} = \{X \leq a\} \setminus \{X < a\}$, so $\mathbb{P}(X = a) = \mathbb{P}(X \leq a) - \mathbb{P}(X < a) = F(a) - F(a^-)$.
2. $\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$, so $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a)$.
3. $\{a < X < b\} = \{X < b\} \setminus \{X \leq a\}$, so $\mathbb{P}(a < X < b) = \mathbb{P}(X < b) - \mathbb{P}(X \leq a) = F(b^-) - F(a)$.
4. $\{a \leq X \leq b\} = \{X \leq b\} \setminus \{X < a\}$, so $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = F(b) - F(a^-)$.
5. $\{a \leq X < b\} = \{X < b\} \setminus \{X < a\}$, so $\mathbb{P}(a \leq X < b) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = F(b^-) - F(a^-)$.

Conversely, if a Function $F : \mathbb{R} \rightarrow [0, 1]$ satisfies the [basic properties](#), then the [formulas above](#) define a probability distribution on \mathbb{R} , with F as the distribution function. For more on this point, read the section on Existence and Uniqueness.

If X has a continuous distribution, then the distribution function F is continuous.

Proof

If X has a continuous distribution, then by definition, $\mathbb{P}(X = x) = 0$ so $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$. Hence from part (a) of the [previous theorem](#), $F(x^-) = F(x^+) = F(x)$.

Thus, the two meanings of *continuous* come together: continuous distribution and continuous function in the calculus sense. Next recall that the distribution of a real-valued random variable X is *symmetric* about a point $a \in \mathbb{R}$ if the distribution of $X - a$ is the same as the distribution of $a - X$.

Suppose that X has a continuous distribution on \mathbb{R} that is symmetric about a point a . Then the distribution function F satisfies $F(a - t) = 1 - F(a + t)$ for $t \in \mathbb{R}$.

Proof

Since $X - a$ and $a - X$ have the same distribution,

$$F(a - t) = \mathbb{P}(X \leq a - t) = \mathbb{P}(X - a \leq -t) = \mathbb{P}(a - X \leq -t) = \mathbb{P}(X \geq a + t) = 1 - F(a + t) \quad (3.6.3)$$

Relation to Density Functions

There are simple relationships between the distribution function and the probability density function. Recall that if X takes value in $S \subseteq \mathbb{R}$ and has probability density function f , we can extend f to all of \mathbb{R} by the convention that $f(x) = 0$ for $x \in S^c$. As in Definition (1), it's customary to define the distribution function F on all of \mathbb{R} , even if the random variable takes values in a subset.

Suppose that X has discrete distribution on a countable subset $S \subseteq \mathbb{R}$. Let f denote the probability density function and F the distribution function.

1. $F(x) = \sum_{t \in S, t \leq x} f(t)$ for $x \in \mathbb{R}$
2. $f(x) = F(x) - F(x^-)$ for $x \in S$

Proof

1. This follows from the definition of the PDF of X , $f(t) = \mathbb{P}(X = t)$ for $t \in S$, and the additivity of probability.
2. This is a restatement of part (a) of the [theorem above](#).

Thus, F is a step function with jumps at the points in S ; the size of the jump at x is $f(x)$.

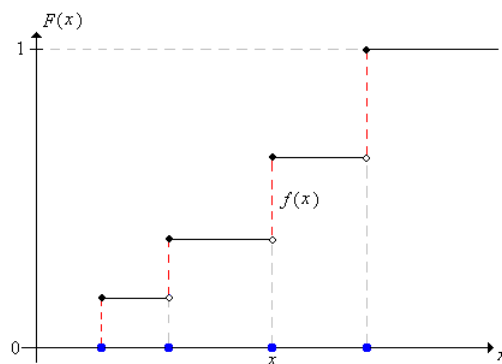


Figure 3.6.3: The distribution function of a discrete distribution

There is an analogous result for a continuous distribution with a probability density function.

Suppose that X has a continuous distribution on \mathbb{R} with probability density function f and distribution function F .

1. $F(x) = \int_{-\infty}^x f(t)dt$ for $x \in \mathbb{R}$.
2. $f(x) = F'(x)$ if f is continuous at x .

Proof

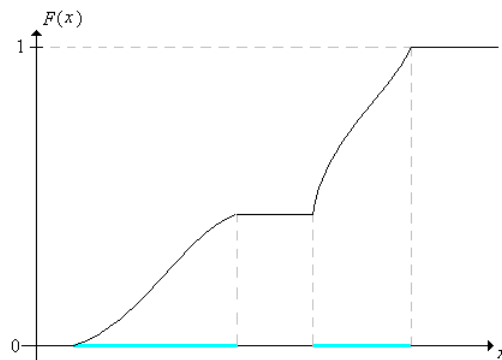


Figure 3.6.4: The distribution function of a continuous distribution

The last result is the basic probabilistic version of the fundamental theorem of calculus. For mixed distributions, we have a combination of the results in the last two theorems.

Suppose that X has a mixed distribution, with discrete part on a countable subset $D \subseteq \mathbb{R}$, and continuous part on $\mathbb{R} \setminus D$. Let g denote the partial probability density function of the discrete part and assume that the continuous part has partial probability density function h . Let F denote the distribution function.

1. $F(x) = \sum_{t \in D, t \leq x} g(t) + \int_{-\infty}^x h(t)dt$ for $x \in \mathbb{R}$
2. $g(x) = F(x) - F(x^-)$ for $x \in D$
3. $h(x) = F'(x)$ if $x \notin D$ and h is continuous at x

Go back to the [graph](#) of a general distribution function. At a point of positive probability, the probability is the size of the jump. At a smooth point of the graph, the continuous probability density is the slope.

Recall that the existence of a probability density function is not guaranteed for a continuous distribution, but of course the distribution function always makes perfect sense. The advanced section on absolute continuity and density functions has an example of a continuous distribution on the interval $(0, 1)$ that has no probability density function. The distribution function is continuous and strictly increases from 0 to 1 on the interval, but has derivative 0 at almost every point!

Naturally, the distribution function can be defined relative to any of the conditional distributions we have discussed. No new concepts are involved, and all of the results above hold.

Reliability

Suppose again that X is a real-valued random variable with distribution function F . The function in the following definition clearly gives the same information as F .

The function F^c defined by

$$F^c(x) = 1 - F(x) = \mathbb{P}(X > x), \quad x \in \mathbb{R} \quad (3.6.4)$$

is the *right-tail distribution function* of X . Give the mathematical properties of F^c analogous to the properties of F in (2).

Answer

1. F^c is decreasing.
2. $F^c(t) \rightarrow F^c(x)$ as $t \downarrow x$ for $x \in \mathbb{R}$, so F^c is continuous from the right.
3. $F^c(t) \rightarrow \mathbb{P}(X \geq x)$ as $t \uparrow x$ for $x \in \mathbb{R}$, so F^c has left limits.
4. $F^c(x) \rightarrow 0$ as $x \rightarrow \infty$.
5. $F^c(x) \rightarrow 1$ as $x \rightarrow -\infty$.

So F might be called the *left-tail distribution function*. But why have two distribution functions that give essentially the same information? The right-tail distribution function, and related functions, arise naturally in the context of *reliability theory*. For the remainder of this subsection, suppose that T is a random variable with values in $[0, \infty)$ and that T has a continuous distribution with probability density function f . Here are the important definitions:

Suppose that T represents the lifetime of a device.

1. The right tail distribution function F^c is the *reliability function* of T .
2. The function h defined by $h(t) = f(t)/F^c(t)$ for $t \geq 0$ is the *failure rate function* of T .

To interpret the reliability function, note that $F^c(t) = \mathbb{P}(T > t)$ is the probability that the device lasts at least t time units. To interpret the failure rate function, note that if dt is “small” then

$$\mathbb{P}(t < T < t + dt \mid T > t) = \frac{\mathbb{P}(t < T < t + dt)}{\mathbb{P}(T > t)} \approx \frac{f(t) dt}{F^c(t)} = h(t) dt \quad (3.6.5)$$

So $h(t) dt$ is the approximate probability that the device will fail in the interval $(t, t + dt)$, given survival up to time t . Moreover, like the distribution function and the reliability function, the failure rate function also completely determines the distribution of T .

The reliability function can be expressed in terms of the failure rate function by

$$F^c(t) = \exp\left(-\int_0^t h(s) ds\right), \quad t \geq 0 \quad (3.6.6)$$

Proof

At the points of continuity of f we have $\frac{d}{dt} F^c(t) = -f(t)$. Hence

$$\int_0^t h(s) ds = \int_0^t \frac{f(s)}{F^c(s)} ds = \int_0^t -\frac{\frac{d}{ds} F^c(s)}{F^c(s)} ds = -\ln[F^c(t)] \quad (3.6.7)$$

The failure rate function h satisfies the following properties:

1. $h(t) \geq 0$ for $t \geq 0$
2. $\int_0^\infty h(t) dt = \infty$

Proof

1. This follows from the definition.
2. This follows from the [previous result](#) and the fact that $F^c(t) \rightarrow 0$ as $t \rightarrow \infty$.

Conversely, a function that satisfies these properties is the failure rate function for a continuous distribution on $[0, \infty)$:

Suppose that $h : [0, \infty) \rightarrow [0, \infty)$ is piecewise continuous and $\int_0^\infty h(t) dt = \infty$. Then the function G defined by

$$F^c(t) = \exp\left(-\int_0^t h(s) ds\right), \quad t \geq 0 \quad (3.6.8)$$

is a reliability function for a continuous distribution on $[0, \infty)$

Proof

The function F^c is continuous, decreasing, and satisfies $F^c(0) = 1$ and $F^c(t) \rightarrow 0$ as $t \rightarrow \infty$. Hence $F = 1 - F^c$ is the distribution function for a continuous distribution on $[0, \infty)$.

Multivariate Distribution Functions

Suppose now that X and Y are real-valued random variables for an experiment (that is, defined on the same probability space), so that (X, Y) is random vector taking values in a subset of \mathbb{R}^2 .

The *distribution function* of (X, Y) is the function F defined by

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2 \quad (3.6.9)$$

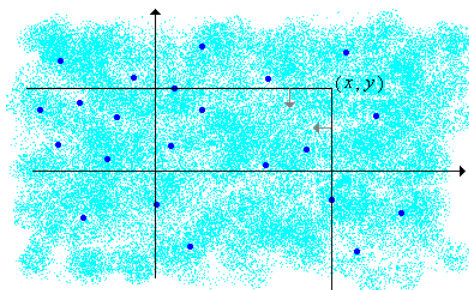


Figure 3.6.5: $F(x, y)$ is the total probability below and to the left of (x, y) .

In the graph above, the light shading is intended to suggest a continuous distribution of probability, while the darker dots represent points of positive probability. Thus, $F(x, y)$ is the total probability mass below and to the left (that is, southwest) of the point (x, y) . As in the single variable case, the distribution function of (X, Y) completely determines the distribution of (X, Y) .

Suppose that $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$. Then

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c) \quad (3.6.10)$$

Proof

Note that $\{X \leq a, Y \leq d\} \cup \{X \leq b, Y \leq c\} \cup \{a < X \leq b, c < Y \leq d\} = \{X \leq b, Y \leq d\}$. The intersection of the first two events is $\{X \leq a, Y \leq c\}$ while the first and third events and the second and third events are disjoint. Thus, from the inclusion-exclusion rule we have

$$F(a, d) + F(b, c) + \mathbb{P}(a < X \leq b, c < Y \leq d) - F(a, c) = F(b, d) \quad (3.6.11)$$

A probability distribution on \mathbb{R}^2 is completely determined by its values on rectangles of the form $(a, b] \times (c, d]$, so just as in the single variable case, it follows that the distribution function of (X, Y) completely determines the distribution of (X, Y) . See the advanced section on existence and uniqueness of positive measures in the chapter on Probability Measures for more details.

In the setting of the previous result, give the appropriate formula on the right for all possible combinations of weak and strong inequalities on the left.

The joint distribution function determines the individual (marginal) distribution functions.

Let F denote the distribution function of (X, Y) , and let G and H denote the distribution functions of X and Y , respectively. Then

1. $G(x) = F(x, \infty)$ for $x \in \mathbb{R}$

2. $H(y) = F(\infty, y)$ for $y \in \mathbb{R}$

Proof

These results follow from the continuity theorem for increasing events. For example, in (a)

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y < \infty) = \lim_{y \rightarrow \infty} \mathbb{P}(X \leq x, Y \leq y) = \lim_{y \rightarrow \infty} F(x, y) \quad (3.6.12)$$

On the other hand, we cannot recover the distribution function of (X, Y) from the individual distribution functions, except when the variables are independent.

Random variables X and Y are independent if and only if

$$F(x, y) = G(x)H(y), \quad (x, y) \in \mathbb{R}^2 \quad (3.6.13)$$

Proof

If X and Y are independent then $F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) = G(x)H(y)$ for $(x, y) \in \mathbb{R}^2$. Conversely, suppose $F(x, y) = G(x)H(y)$ for $(x, y) \in \mathbb{R}^2$. If $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$ then from (15),

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = G(b)H(d) - G(a)H(d) - G(b)H(c) + G(a)H(c) \quad (3.6.14)$$

$$= [G(b) - G(a)][H(d) - H(c)] = \mathbb{P}(a < X \leq b)\mathbb{P}(c < Y \leq d) \quad (3.6.15)$$

so it follows that X and Y are independent. (Recall again that a probability distribution on \mathbb{R}^2 is completely determined by its values on rectangles.)

All of the results of this subsection generalize in a straightforward way to n -dimensional random vectors. Only the notation is more complicated.

The Empirical Distribution Function

Suppose now that X is a real-valued random variable for a basic random experiment and that we repeat the experiment n times independently. This generates (for the new compound experiment) a sequence of independent variables (X_1, X_2, \dots, X_n) each with the same distribution as X . In statistical terms, this sequence is a *random sample* of size n from the distribution of X . In statistical inference, the observed values (x_1, x_2, \dots, x_n) of the random sample form our *data*.

The *empirical distribution function*, based on the data (x_1, x_2, \dots, x_n) , is defined by

$$F_n(x) = \frac{1}{n} \# \{i \in \{1, 2, \dots, n\} : x_i \leq x\} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i \leq x), \quad x \in \mathbb{R} \quad (3.6.16)$$

Thus, $F_n(x)$ gives the proportion of values in the data set that are less than or equal to x . The function F_n is a statistical estimator of F , based on the given data set. This concept is explored in more detail in the section on the sample mean in the chapter on Random Samples. In addition, the empirical distribution function is related to the Brownian bridge stochastic process which is studied in the chapter on Brownian motion.

Quantile Functions

Definitions

Suppose again that X is a real-valued random variable with distribution function F .

For $p \in (0, 1)$, a value of x such that $F(x^-) = \mathbb{P}(X < x) \leq p$ and $F(x) = \mathbb{P}(X \leq x) \geq p$ is called a *quantile* of order p for the distribution.

Roughly speaking, a quantile of order p is a value where the graph of the distribution function crosses (or jumps over) p . For example, in the picture below, a is the unique quantile of order p and b is the unique quantile of order q . On the other hand, the quantiles of order r form the interval $[c, d]$, and moreover, d is a quantile for all orders in the interval $[r, s]$. Note also that if X has a continuous distribution (so that F is continuous) and x is a quantile of order $p \in (0, 1)$, then $F(x) = p$.

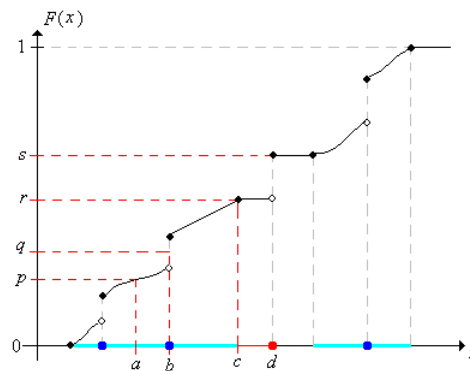


Figure 3.6.6: Quantiles of various orders

Note that there is an inverse relation of sorts between the quantiles and the cumulative distribution values, but the relation is more complicated than that of a function and its ordinary inverse function, because the distribution function is not one-to-one in general. For many purposes, it is helpful to select a specific quantile for each order; to do this requires defining a *generalized inverse* of the distribution function F .

The *quantile function* F^{-1} of X is defined by

$$F^{-1}(p) = \min\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in (0, 1) \quad (3.6.17)$$

F^{-1} is well defined

Since F is right continuous and increasing, $\{x \in \mathbb{R} : F(x) \geq p\}$ is an interval of the form $[a, \infty)$. Thus, the minimum of the set is a .

Note that if F strictly increases from 0 to 1 on an interval S (so that the underlying distribution is continuous and is supported on S), then F^{-1} is the ordinary inverse of F . We do not usually define the quantile function at the endpoints 0 and 1. If we did, note that $F^{-1}(0)$ would always be $-\infty$.

Properties

The following exercise justifies the name: $F^{-1}(p)$ is the minimum of the quantiles of order p .

Let $p \in (0, 1)$.

1. $F^{-1}(p)$ is a quantile of order p .
2. If x is a quantile of order p then $F^{-1}(p) \leq x$.

Proof

Let $y = F^{-1}(p)$.

1. Note that $F(y) \geq p$ by definition, and if $x < y$ then $F(x) < p$. Hence $F(y^-) \leq p$. Therefore y is a quantile of order p .
2. Suppose that x is a quantile of order p . Then $F(x) \geq p$ so by definition, $y \leq x$.

Other basic properties of the quantile function are given in the following theorem.

F^{-1} satisfies the following properties:

1. F^{-1} is increasing on $(0, 1)$.
2. $F^{-1}[F(x)] \leq x$ for any $x \in \mathbb{R}$ with $F(x) < 1$.
3. $F[F^{-1}(p)] \geq p$ for any $p \in (0, 1)$.
4. $F^{-1}(p^-) = F^{-1}(p)$ for $p \in (0, 1)$. Thus F^{-1} is continuous from the left.
5. $F^{-1}(p^+) = \inf\{x \in \mathbb{R} : F(x) > p\}$ for $p \in (0, 1)$. Thus F^{-1} has limits from the right.

Proof

1. Note that if $p, q \in (0, 1)$ with $p \leq q$, then $\{x \in \mathbb{R} : F(x) \geq q\} \subseteq \{x \in \mathbb{R} : F(x) \geq p\}$.
2. This follows from the definition: $F^{-1}[F(x)]$ is the smallest $y \in \mathbb{R}$ with $F(y) \geq F(x)$.

3. This also follows from the definition: $F^{-1}(p)$ is a value $y \in \mathbb{R}$ satisfying $F(y) \geq p$.
4. This follows from the fact that F is continuous from the right
5. This follows from the fact that F has limits from the left.

As always, the inverse of a function is obtained essentially by reversing the roles of independent and dependent variables. In the graphs below, note that jumps of F become flat portions of F^{-1} while flat portions of F become jumps of F^{-1} . For $p \in (0, 1)$, the set of quantiles of order p is the closed, bounded interval $[F^{-1}(p), F^{-1}(p^+)]$. Thus, $F^{-1}(p)$ is the smallest quantile of order p , as we noted earlier, while $F^{-1}(p^+)$ is the largest quantile of order p .

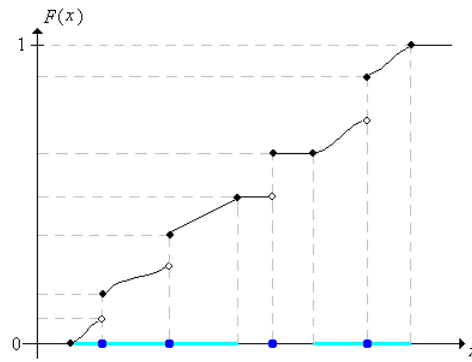


Figure 3.6.7: Graph of the distribution function

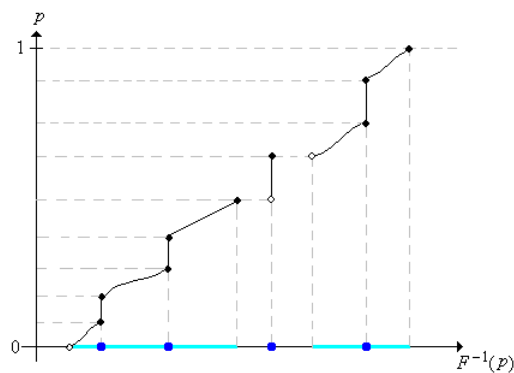


Figure 3.6.8: Graph of the quantile function

The following basic property will be useful in simulating random variables, a topic explored in the section on transformations of random variables.

For $x \in \mathbb{R}$ and $p \in (0, 1)$, $F^{-1}(p) \leq x$ if and only if $p \leq F(x)$.

Proof

Suppose that $F^{-1}(p) \leq x$. Then, since F is increasing, $F[F^{-1}(p)] \leq F(x)$. But $p \leq F[F^{-1}(p)]$ by part (c) of the previous result, so $p \leq F(x)$. Conversely, suppose that $p \leq F(x)$. Then, since F^{-1} is increasing, $F^{-1}(p) \leq F^{-1}[F(x)]$. But $F^{-1}[F(x)] \leq x$ by part (b) of the previous result, so $F^{-1}(p) \leq x$.

Special Quantiles

Certain quantiles are important enough to deserve special names.

Suppose that X is a real-valued random variable.

1. A quantile of order $\frac{1}{4}$ is a *first quartile* of the distribution.
2. A quantile of order $\frac{1}{2}$ is a *median* or *second quartile* of the distribution.
3. A quantile of order $\frac{3}{4}$ is a *third quartile* of the distribution.

When there is only one median, it is frequently used as a measure of the *center* of the distribution, since it divides the set of values of X in half, by probability. More generally, the quartiles can be used to divide the set of values into fourths, by probability.

Assuming uniqueness, let q_1 , q_2 , and q_3 denote the first, second, and third quartiles of X , respectively, and let $a = F^{-1}(0^+)$ and $b = F^{-1}(1)$.

1. The *interquartile range* is defined to be $q_3 - q_1$.
2. The five parameters (a, q_1, q_2, q_3, b) are referred to as the *five number summary* of the distribution.

Note that the interval $[q_1, q_3]$ roughly gives the middle half of the distribution, so the interquartile range, the length of the interval, is a natural measure of the dispersion of the distribution about the median. Note also that a and b are essentially the minimum and maximum values of X , respectively, although of course, it's possible that $a = -\infty$ or $b = \infty$ (or both). Collectively, the five parameters give a great deal of information about the distribution in terms of the center, spread, and skewness. Graphically, the five numbers are often displayed as a *boxplot* or *box and whisker plot*, which consists of a line extending from the minimum value a to the maximum value b , with a rectangular box from q_1 to q_3 , and “whiskers” at a , the median q_2 , and b . Roughly speaking, the five numbers separate the set of values of X into 4 intervals of approximate probability $\frac{1}{4}$ each.

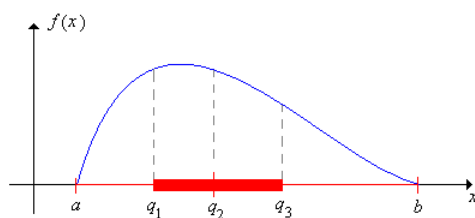


Figure 3.6.9: The probability density function and boxplot for a continuous distribution

Suppose that X has a continuous distribution that is symmetric about a point $a \in \mathbb{R}$. If $a + t$ is a quantile of order $p \in (0, 1)$ then $a - t$ is a quantile of order $1 - p$.

Proof

Note that this is the quantile function version of [symmetry result](#) for the distribution function. If $a + t$ is a quantile of order p then (since X has a continuous distribution) $F(a + t) = p$. But then $F(a - t) = 1 - F(a + t) = 1 - p$ so $a - t$ is a quantile of order $1 - p$.

Examples and Applications

Distributions of Different Types

Let F be the function defined by

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{10}, & 1 \leq x < \frac{3}{2} \\ \frac{3}{10}, & \frac{3}{2} \leq x < 2 \\ \frac{6}{10}, & 2 \leq x < \frac{5}{2} \\ \frac{9}{10}, & \frac{5}{2} \leq x < 3 \\ 1, & x \geq 3; \end{cases} \quad (3.6.18)$$

1. Sketch the graph of F and show that F is the distribution function for a discrete distribution.
2. Find the corresponding probability density function f and sketch the graph.
3. Find $\mathbb{P}(2 \leq X < 3)$ where X has this distribution.
4. Find the quantile function and sketch the graph.
5. Find the five number summary and sketch the boxplot.

Answer

1. Note that F increases from 0 to 1, is a step function, and is right continuous.

2. $f(x) = \begin{cases} \frac{1}{10}, & x = 1 \\ \frac{1}{5}, & x = \frac{3}{2} \\ \frac{3}{10}, & x = 2 \\ \frac{3}{10}, & x = \frac{5}{2} \\ \frac{1}{10}, & x = 3 \end{cases}$
3. $\mathbb{P}(2 \leq X < 3) = \frac{3}{5}$
4. $F^{-1}(p) = \begin{cases} 1, & 0 < p \leq \frac{1}{10} \\ \frac{3}{2}, & \frac{1}{10} < p \leq \frac{3}{10} \\ 2, & \frac{3}{10} < p \leq \frac{6}{10} \\ \frac{5}{2}, & \frac{6}{10} < p \leq \frac{9}{10} \\ 3, & \frac{9}{10} < p \leq 1 \end{cases}$
5. $(1, \frac{3}{2}, 2, \frac{5}{2}, 3)$

Let F be the function defined by

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{x+1}, & x \geq 0 \end{cases} \quad (3.6.19)$$

1. Sketch the graph of F and show that F is the distribution function for a continuous distribution.
2. Find the corresponding probability density function f and sketch the graph.
3. Find $\mathbb{P}(2 \leq X < 3)$ where X has this distribution.
4. Find the quantile function and sketch the graph.
5. Find the five number summary and sketch the boxplot.

Answer

1. Note that F is continuous and increases from 0 to 1.
2. $f(x) = \frac{1}{(x+1)^2}, \quad x > 0$
3. $\mathbb{P}(2 \leq X < 3) = \frac{1}{12}$
4. $F^{-1}(p) = \frac{p}{1-p}, \quad 0 < p < 1$
5. $(0, \frac{1}{3}, 1, 3, \infty)$

The expression $\frac{p}{1-p}$ that occurs in the quantile function in the last exercise is known as the *odds ratio* associated with p , particularly in the context of gambling.

Let F be the function defined by

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x, & 0 \leq x < 1 \\ \frac{1}{3} + \frac{1}{4}(x-1)^2, & 1 \leq x < 2 \\ \frac{2}{3} + \frac{1}{4}(x-2)^3, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases} \quad (3.6.20)$$

1. Sketch the graph of F and show that F is the distribution function of a mixed distribution.
2. Find the partial probability density function of the discrete part and sketch the graph.
3. Find the partial probability density function of the continuous part and sketch the graph.
4. Find $\mathbb{P}(2 \leq X < 3)$ where X has this distribution.
5. Find the quantile function and sketch the graph.
6. Find the five number summary and sketch the boxplot.

Answer

1. Note that F is piece-wise continuous, increases from 0 to 1, and is right continuous.

2. $g(1) = g(2) = g(3) = \frac{1}{12}$
3. $h(x) = \begin{cases} \frac{1}{4}, & 0 < x < 1 \\ \frac{1}{2}(x-1), & 1 < x < 2 \\ \frac{3}{4}(x-2)^2, & 2 < x < 3 \end{cases}$
4. $\mathbb{P}(2 \leq X < 3) = \frac{1}{3}$
5. $F^{-1}(p) = \begin{cases} 4p, & 0 < p \leq \frac{1}{4} \\ 1, & \frac{1}{4} < p \leq \frac{1}{3} \\ 1 + \sqrt{4(p - \frac{1}{3})}, & \frac{1}{3} < p \leq \frac{7}{12} \\ 2, & \frac{7}{12} < p \leq \frac{2}{3} \\ 2 + \sqrt[3]{4(p - \frac{2}{3})}, & \frac{2}{3} < p \leq \frac{11}{12} \\ 3, & \frac{11}{12} < p \leq 1 \end{cases}$
6. $(0, 1, 1 + \sqrt{\frac{2}{3}}, 2 + \sqrt[3]{\frac{1}{3}}, 3)$

The Uniform Distribution

Suppose that X has probability density function $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$, where $a, b \in \mathbb{R}$ and $a < b$.

1. Find the distribution function and sketch the graph.
2. Find the quantile function and sketch the graph.
3. Compute the five-number summary.
4. Sketch the graph of the probability density function with the boxplot on the horizontal axis.

Answer

1. $F(x) = \frac{x-a}{b-a}, \quad a \leq x < b$
2. $F^{-1}(p) = a + (b-a)p, \quad 0 \leq p \leq 1$
3. $(a, \frac{3a+b}{4}, \frac{a+b}{2}, \frac{a+3b}{4}, b)$

The distribution in the last exercise is the *uniform distribution* on the interval $[a, b]$. The left endpoint a is the *location parameter* and the length of the interval $w = b - a$ is the *scale parameter*. The uniform distribution models a point chose “at random” from the interval, and is studied in more detail in the chapter on Special Distributions.

In the special distribution calculator, select the continuous uniform distribution. Vary the location and scale parameters and note the shape of the probability density function and the distribution function.

The Exponential Distribution

Suppose that T has probability density function $f(t) = re^{-rt}$ for $0 \leq t < \infty$, where $r > 0$ is a parameter.

1. Find the distribution function and sketch the graph.
2. Find the reliability function and sketch the graph.
3. Find the failure rate function and sketch the graph.
4. Find the quantile function and sketch the graph.
5. Compute the five-number summary.
6. Sketch the graph of the probability density function with the boxplot on the horizontal axis.

Answer

1. $F(t) = 1 - e^{-rt}, \quad 0 \leq t < \infty$
2. $F^c(t) = e^{-rt}, \quad 0 \leq t < \infty$
3. $h(t) = r, \quad 0 \leq t < \infty$
4. $F^{-1}(p) = -\frac{1}{r} \ln(1-p), \quad 0 \leq p < 1$
5. $(0, \frac{1}{r}[\ln 4 - \ln 3], \frac{1}{r} \ln 2, \frac{1}{r} \ln 4, \infty)$

The distribution in the last exercise is the *exponential distribution* with rate parameter r . Note that this distribution is characterized by the fact that it has *constant failure rate* (and this is the reason for referring to r as the rate parameter). The reciprocal of the rate parameter is the *scale parameter*. The exponential distribution is used to model failure times and other random times under certain conditions, and is studied in detail in the chapter on The Poisson Process.

In the special distribution calculator, select the exponential distribution. Vary the scale parameter b and note the shape of the probability density function and the distribution function.

The Pareto Distribution

Suppose that X has probability density function $f(x) = \frac{a}{x^{a+1}}$ for $1 \leq x < \infty$ where $a > 0$ is a parameter.

1. Find the distribution function.
2. Find the reliability function.
3. Find the failure rate function.
4. Find the quantile function.
5. Compute the five-number summary.
6. In the case $a = 2$, sketch the graph of the probability density function with the boxplot on the horizontal axis.

Answer

1. $F(x) = 1 - \frac{1}{x^a}, \quad 1 \leq x < \infty$
2. $F^c(x) = \frac{1}{x^a}, \quad 1 \leq x < \infty$
3. $h(x) = \frac{a}{x}, \quad 1 \leq x < \infty$
4. $F^{-1}(p) = (1 - p)^{-1/a}, \quad 0 \leq p < 1$
5. $\left(1, \left(\frac{3}{4}\right)^{-1/a}, \left(\frac{1}{2}\right)^{-1/a}, \left(\frac{1}{4}\right)^{-1/a}, \infty\right)$

The distribution in the last exercise is the *Pareto distribution* with shape parameter a , named after Vilfredo Pareto. The Pareto distribution is a heavy-tailed distribution that is sometimes used to model income and certain other economic variables. It is studied in detail in the chapter on Special Distributions.

In the special distribution calculator, select the Pareto distribution. Keep the default value for the scale parameter, but vary the shape parameter and note the shape of the density function and the distribution function.

The Cauchy Distribution

Suppose that X has probability density function $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$.

1. Find the distribution function and sketch the graph.
2. Find the quantile function and sketch the graph.
3. Compute the five-number summary and the interquartile range.
4. Sketch the graph of the probability density function with the boxplot on the horizontal axis.

Answer

1. $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x, \quad x \in \mathbb{R}$
2. $F^{-1}(p) = \tan\left[\pi\left(p - \frac{1}{2}\right)\right], \quad 0 < p < 1$
3. $(-\infty, -1, 0, 1, \infty), \text{ IQR} = 2$

The distribution in the last exercise is the *Cauchy distribution*, named after Augustin Cauchy. The Cauchy distribution is studied in more generality in the chapter on Special Distributions.

In the special distribution calculator, select the Cauchy distribution and keep the default parameter values. Note the shape of the density function and the distribution function.

The Weibull Distribution

Let $h(t) = kt^{k-1}$ for $0 < t < \infty$ where $k > 0$ is a parameter.

1. Sketch the graph of h in the cases $0 < k < 1$, $k = 1$, $1 < k < 2$, $k = 2$, and $k > 2$.
2. Show that h is a failure rate function.
3. Find the reliability function and sketch the graph.
4. Find the distribution function and sketch the graph.
5. Find the probability density function and sketch the graph.
6. Find the quantile function and sketch the graph.
7. Compute the five-number summary.

Answer

1. h is decreasing and concave upward if $0 < k < 1$; $h = 1$ (constant) if $k = 1$; h is increasing and concave downward if $1 < k < 2$; $h(t) = t$ (linear) if $k = 2$; h is increasing and concave upward if $k > 2$;
2. $h(t) > 0$ for $0 < t < \infty$ and $\int_0^\infty h(t) dt = \infty$
3. $F^c(t) = \exp(-t^k)$, $0 \leq t < \infty$
4. $F(t) = 1 - \exp(-t^k)$, $0 \leq t < \infty$
5. $f(t) = kt^{k-1} \exp(-t^k)$, $0 \leq t < \infty$
6. $F^{-1}(p) = [-\ln(1-p)]^{1/k}$, $0 \leq p < 1$
7. $(0, [\ln 4 - \ln 3]^{1/k}, [\ln 2]^{1/k}, [\ln 4]^{1/k}, \infty)$

The distribution in the previous exercise is the *Weibull distributions* with shape parameter k , named after Walodi Weibull. The Weibull distribution is studied in detail in the chapter on Special Distributions. Since this family includes increasing, decreasing, and constant failure rates, it is widely used to model the lifetimes of various types of devices.

In the special distribution calculator, select the Weibull distribution. Keep the default scale parameter, but vary the shape parameter and note the shape of the density function and the distribution function.

Beta Distributions

Suppose that X has probability density function $f(x) = 12x^2(1-x)$ for $0 \leq x \leq 1$.

1. Find the distribution function of X and sketch the graph.
2. Find $\mathbb{P}(\frac{1}{4} \leq X \leq \frac{1}{2})$.
3. Compute the five number summary and the interquartile range. You will have to approximate the quantiles.
4. Sketch the graph of the density function with the boxplot on the horizontal axis.

Answer

1. $F(x) = 4x^3 - 3x^4$, $0 \leq x \leq 1$
2. $\mathbb{P}(\frac{1}{4} \leq X \leq \frac{1}{2}) = \frac{67}{256}$
3. $(0, 0.4563, 0.6413, 0.7570, 1)$, $IQR = 0.3007$

Suppose that X has probability density function $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ for $0 < x < 1$.

1. Find the distribution function of X and sketch the graph.
2. Compute $\mathbb{P}(\frac{1}{3} \leq X \leq \frac{2}{3})$.
3. Find the quantile function and sketch the graph.
4. Compute the five number summary and the interquartile range.
5. Sketch the graph of the probability density function with the boxplot on the horizontal axis.

Answer

1. $F(x) = \frac{2}{\pi} \arcsin(\sqrt{x})$, $0 \leq x \leq 1$
2. $\mathbb{P}(\frac{1}{3} \leq X \leq \frac{2}{3}) = 0.2163$
3. $F^{-1}(p) = \sin^2(\frac{\pi p}{2})$, $0 < p < 1$

$$4. \left(0, \frac{1}{2} - \frac{\sqrt{2}}{4}, \frac{1}{2}, \frac{1}{2} + \frac{\sqrt{2}}{4}, 1\right), \text{IQR} = \frac{\sqrt{2}}{2}$$

The distributions in the last two exercises are examples of *beta distributions*. The particular beta distribution in the last exercise is also known as the *arcsine distribution*; the distribution function explains the name. Beta distributions are used to model random proportions and probabilities, and certain other types of random variables, and are studied in detail in the chapter on Special Distributions.

In the special distribution calculator, select the beta distribution. For each of the following parameter values, note the location and shape of the density function and the distribution function.

1. $a = 3, b = 2$. This gives the [first beta distribution](#) above.
2. $a = b = \frac{1}{2}$. This gives the [arcsine distribution](#) above

Logistic Distribution

Let $F(x) = \frac{e^x}{1+e^x}$ for $x \in \mathbb{R}$.

1. Show that F is a distribution function for a continuous distribution, and sketch the graph.
2. Compute $\mathbb{P}(-1 \leq X \leq 1)$ where X is a random variable with distribution function F .
3. Find the quantile function and sketch the graph.
4. Compute the five-number summary and the interquartile range.
5. Find the probability density function and sketch the graph with the boxplot on the horizontal axis.

Answer

1. Note that F is continuous, and increases from 0 to 1.
2. $\mathbb{P}(-1 \leq X \leq 1) = 0.4621$
3. $F^{-1}(p) = \ln\left(\frac{p}{1-p}\right), \quad 0 < p < 1$
4. $(-\infty, -\ln 3, 0, \ln 3, \infty)$
5. $f(x) = \frac{e^x}{(1+e^x)^2}, \quad x \in \mathbb{R}$

The distribution in the last exercise is an *logistic distribution* and the quantile function is known as the *logit function*. The logistic distribution is studied in detail in the chapter on Special Distributions.

In the special distribution calculator, select the logistic distribution and keep the default parameter values. Note the shape of the probability density function and the distribution function.

Extreme Value Distribution

Let $F(x) = e^{-e^{-x}}$ for $x \in \mathbb{R}$.

1. Show that F is a distribution function for a continuous distribution, and sketch the graph.
2. Compute $\mathbb{P}(-1 \leq X \leq 1)$ where X is a random variable with distribution function F .
3. Find the quantile function and sketch the graph.
4. Compute the five-number summary.
5. Find the probability density function and sketch the graph with the boxplot on the horizontal axis.

Answer

1. Note that F is continuous, and increases from 0 to 1.
2. $\mathbb{P}(-1 \leq X \leq 1) = 0.6262$
3. $F^{-1}(p) = -\ln(-\ln p), \quad 0 < p < 1$
4. $(-\infty, -\ln(\ln 4), -\ln(\ln 2), -\ln(\ln 4 - \ln 3), \infty)$
5. $f(x) = e^{-e^{-x}} e^{-x}, \quad x \in \mathbb{R}$

The distribution in the last exercise is the *type 1 extreme value distribution*, also known as the *Gumbel distribution* in honor of Emil Gumbel. Extreme value distributions are studied in detail in the chapter on Special Distributions.

In the special distribution calculator, select the extreme value distribution and keep the default parameter values. Note the shape and location of the probability density function and the distribution function.

The Standard Normal Distribution

Recall that the *standard normal distribution* has probability density function ϕ given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \quad (3.6.21)$$

This distribution models physical measurements of all sorts subject to small, random errors, and is one of the most important distributions in probability. The normal distribution is studied in more detail in the chapter on Special Distributions. The distribution function Φ , of course, can be expressed as

$$\Phi(z) = \int_{-\infty}^z \phi(x) dx, \quad z \in \mathbb{R} \quad (3.6.22)$$

but Φ and the quantile function Φ^{-1} cannot be expressed, in closed form, in terms of elementary functions. Because of the importance of the normal distribution Φ and Φ^{-1} are themselves considered *special functions*, like \sin , \ln , and many others. Approximate values of these functions can be computed using most mathematical and statistical software packages. Because the distribution is symmetric about 0, $\Phi(-z) = 1 - \Phi(z)$ for $z \in \mathbb{R}$, and equivalently, $\Phi^{-1}(1-p) = -\Phi^{-1}(p)$. In particular, the median is 0.

Open the special distribution calculator and choose the normal distribution. Keep the default parameter values and select CDF view. Note the shape and location of the distribution/quantile function. Compute each of the following:

1. The first and third quartiles
2. The quantiles of order 0.9 and 0.1
3. The quantiles of order 0.95 and 0.05

Miscellaneous Exercises

Suppose that X has probability density function $f(x) = -\ln x$ for $0 < x \leq 1$.

1. Sketch the graph of f .
2. Find the distribution function F and sketch the graph.
3. Find $\mathbb{P}(\frac{1}{3} \leq X \leq \frac{1}{2})$.

Answer

2. $F(x) = x - x \ln x, \quad 0 < x < 1$
3. $\mathbb{P}(\frac{1}{3} \leq X \leq \frac{1}{2}) = \frac{1}{6} + \frac{1}{2} \ln 2 - \frac{1}{3} \ln 3$

Suppose that a pair of fair dice are rolled and the sequence of scores (X_1, X_2) is recorded.

1. Find the distribution function of $Y = X_1 + X_2$, the sum of the scores.
2. Find the distribution function of $V = \max\{X_1, X_2\}$, the maximum score.
3. Find the conditional distribution function of Y given $V = 5$.

Answer

The random variables are discrete, so the CDFs are step functions, with jumps at the values of the variables. The following tables give the values of the CDFs at the values of the random variables.

1.	y	2	3	4	5	6	7	8	9	10	11	12
	$\mathbb{P}(Y \leq y)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

2.	v	1	2	3	4	5	6
	$\mathbb{P}(V \leq v)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	1

3.	y	6	7	8	9	10
	$\mathbb{P}(Y \leq y \mid V = 5)$	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{6}{9}$	$\frac{8}{9}$	1

Suppose that (X, Y) has probability density function $f(x, y) = x + y$ for $0 \leq x \leq 1, 0 \leq y \leq 1$.

- Find the distribution function of X, Y .
- Compute $\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{1}{2}, \frac{1}{3} \leq Y \leq \frac{2}{3}\right)$.
- Find the distribution function of X .
- Find the distribution function of Y .
- Find the conditional distribution function of X given $Y = y$ for $0 < y < 1$.
- Find the conditional distribution function of Y given $X = x$ for $0 < x < 1$.
- Are X and Y independent?

Answer

- $F(x, y) = \frac{1}{2}(xy^2 + yx^2); \quad 0 < x < 1, 0 < y < 1$
- $\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{1}{2}, \frac{1}{3} \leq Y \leq \frac{2}{3}\right) = \frac{7}{96}$
- $G(x) = \frac{1}{2}(x + x^2), \quad 0 < x < 1$
- $H(y) = \frac{1}{2}(y + y^2), \quad 0 < y < 1$
- $G(x \mid y) = \frac{x^2/2 + xy}{y + 1/2}; \quad 0 < x < 1, 0 < y < 1$
- $H(y \mid x) = \frac{y^2/2 + xy}{x + 1/2}; \quad 0 < x < 1, 0 < y < 1$

Statistical Exercises

For the M&M data, compute the empirical distribution function of the total number of candies.

Answer

Let N denote the total number of candies. The empirical distribution function of N is a step function; the following table gives the values of the function at the jump points.

n	50	53	54	55	56	57	58	59	60	61
$\mathbb{P}(N \leq n)$	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{7}{30}$	$\frac{11}{30}$	$\frac{14}{30}$	$\frac{23}{30}$	$\frac{26}{30}$	$\frac{28}{30}$	1

For the cicada data, let BL denotes body length and let G denote gender. Compute the empirical distribution function of the following variables:

- BL
- BL given $G = 1$ (male)
- BL given $G = 0$ (female).
- Do you believe that BL and G are independent?

For statistical versions of some of the topics in this section, see the chapter on Random Samples, and in particular, the sections on empirical distributions and order statistics.

This page titled [3.6: Distribution and Quantile Functions](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.