WP/23/xx

GTAP Working Paper

A Summary Guide to the Latin Hypercube Sampling (LHS) Utility

by Dominique van der Mensbrugghe

Center for Global Trade Analysis
Purdue University

First edition: May 2023

Please cite using the following reference:
van der Mensbrugghe, Dominique, 2023.
A Summary Guide to the Latin Hypercube Sampling (LHS) Utility
GTAP Working Paper, TP/23/xx
Center for Global Trade Analysis, Purdue University
West Lafayette, IN
https://www...
DOI://www...

Center for Global Trade Analysis
Department of Agricultural Economics
Purdue University
West Lafayette, IN 47907
https://www.gtap.agecon.purdue.edu

A Summary Guide to the Latin Hypercube Sampling (LHS) Utility

Dominique van der Mensbrugghe[†]

May 8, 2023

Abstract: Latin Hypercube Sampling (LHS) is one method of Monte Carlo-type sampling, which is useful for limiting sample size yet maximizing the range of sampling of the underlying distributions. The LHS utility, for which this document describes the usage, also allows for user-specified correlations between two or more of the sampled distributions. The LHS utility described herein is a full re-coding using C/C++ of the original LHS utility—developed at Sandia National Labs (Swiler and Wyss (2004)), written in FORTRAN and freely available. The re-coding hones close to the original FORTRAN code, but allows for significantly more flexibility. For example, dynamic memory allocation is used for all internal variables and hence there are no pre-determined dimensions. The new utility has additional features compared to the original FORTRAN code: (1) it includes 10 new statistical distributions; (2) it has four additional output formats; and (3) it has an alternative random number generator. This guide provides a summary of the full features of the LHS utility. For a complete reference, with the exception of the new features, as well as a description of the intuition behind the LHS algorithm users are referred to Swiler and Wyss (2004).

JEL Codes: C15, C53, C63
Keywords: Monte Carlo simulations, Latin hypercube sampling, statistical sampling

[†] The Center for Global Trade Analysis, Purdue University, 403 Mitch Daniels Boulevard, West Lafayette, IN 47906. Corresponding e-mail: vandermd@purdue.edu.

Contents

Chapter 1

Introduction

Latin Hypercube Sampling (LHS) is one method of Monte Carlo-type sampling, which is useful for limiting sample size yet maximizing the range of sampling of the underlying distributions. The LHS utility, for which this document describes the usage, also allows for user-specified correlations between two or more of the sampled distributions. The LHS utility described herein is a full re-coding using C/C++ of the original LHS utility—developed at Sandia National Labs (Swiler and Wyss (2004)), written in FORTRAN and freely available. The re-coding hones close to the original FORTRAN code, but allows for significantly more flexibility. For example, dynamic memory allocation is used for all internal variables and hence there are no pre-determined dimensions. The new utility has additional features compared to the original FORTRAN code: (1) it includes 10 new statistical distributions; (2) it has four additional output formats; and (3) it has an alternative random number generator. This guide provides a summary of the full features of the LHS utility. For a complete reference, with the exception of the new features, as well as a description of the intuition behind the LHS algorithm users are referred to Swiler and Wyss (2004).

An introduction to the LHS algorithm and its application to an Integrated Assessment Model (IAM) is provided in van der Mensbrugghe (forthcoming). The GAMS implementation of the IAM, based on the META 21 model (Dietz et al., 2021), coupled with the LHS utility is described in van der Mensbrugghe (2023a). With the increasing sophistication of desktop computers, parallel processing can vastly decrease the time for running large sets of Monte Carlo simulations. An example of parallelization using Python as the scripting language is provided in van der Mensbrugghe (2023b). The time to simulate META 21 10,000 times decreased from around 3 hours with sequential processing to around 15 minutes with parallel processing on a desktop computer.

Chapter 2

User guide

2.1 Input file

The input file is a simple text file that is composed of two sections. The first section is essentially the preamble where the user provides key information for the overall LHS run, such as the initial seed, the number of draws, etc. The second section provides the specific information on the shape of the draw, i.e., the list of statistical distributions with their respective parameters, and any correlation across the sampled variables.

2.1.1 The run's user options

The user options can be divided into two parts: (1) file information; and (2) overall parameters for a specific run. Each option is composed of a recognized keyword and a possible option. Not all options are required. In the absence of a specific user input, the program will assume a default option. Additional notes: (1) the original FORTRAN code assumes all options start with the letters `LHS`. The new code will recognize these letters, but strips them before parsing the option; and (2) the description below will indicate which options are new to the C/C++ version of the LHS utility.

File options

Table 2.1 lists the keywords that relate to the output files of the LHS tool. These are described further below:

1. **OUT** represents the name of the output file containing the values from the draw for this run. It is a required option. The file extension will be parsed to determine the file format, of which there are five. An extension of `.csv`, will create a comma separated text file with the draw's values. This format is commonly used in integrated workflows as most software packages have the capability of reading these flat files. For example, a CSV file can readily be read into an Excel Pivot table. An extension of `.xml` will create an XML-formatted file. These files are Excel ready. If the user requests more than one replication, each replication will be saved in individual worksheets. An extension of `.gms` will create a GAMS-ready data text file along with the set definitions to read the sampled data. An extension of `.gdx` will create a GAMS-ready GDX file, along with a GAMS code file with the set definitions and instructions to read the GDX file. If any other extension (including no extension) is specified, the program will default to the standard output format, which is the only format available with the original Sandia code. This format lends itself for easy access in FORTRAN-style input statements.
2. **ALL** is an optional keyword [NEW]. It forces the output to include variables that are defined as `CONSTANT` and/or as an alias using the `SAME AS` keyword. Both of these types of variables are further described below.
3. **MSG** Is the name of the file containing the run's main characteristics and optional summary output detailed further below. It is a required keyword. It has been traditional to use the `.lst` extension, i.e., the listing file.
4. **TITL** Is an optional keyword that allows the user to provide a title for this run.
5. **CODE** This option allows the sample output to prepend a specific alphanumeric label to the sample observation indicator [NEW]. The default is to output the indicator in numerical sequence, i.e., $1, 2, 3, \ldots$. If the user supplies a code, for example 'S', each observation index will be output with a prefix using the user-supplied code, e.g., $S1, S2, S3, \ldots$. The code must start and end with a letter. If the user requests the GMS or GDX

Table 2.1: **Keywords related to output files**

Keyword	Specification	Required	Value
`OUT` *filename*	Name of LHS sampled data output file	YES	Valid filename
`ALL`	Forces output to include 'SAME AS' and 'CONSTANT' variables	NO	Default: False
`MSG` *filename*	Name of output file that contains the LHS user output and messages from LHS execution	YES	Valid filename
`TITL` *title*	Title of LHS run	NO	User-defined name
`CODE` *title*	Code of LHS run	NO	User-defined code
`RPTS` *options*	Reporting options	NO	Default = None. `HIST` = prints histogram of each sampled random variable, `CORR` = prints achieved correlation matrices, `DATA` = prints raw and rank sample values. Any combination of `HIST`, `CORR`, and `DATA` may be used.

format for the output, the appropriate set definition will include the code, e.g., `set obs / S1*S2000 / ;`. In CSV files, the observation indicator will be output with double quotes if the user requests a code, otherwise, the indicator will be an integer. This option has no impact on XML files, or the default file format.

6. **RPTS** This is an optional keyword with three possible options and relates to the summary output that will be provided in the listing file. The option *DATA* will output the draw and the rankings for each individual variable in the listing file. N.B. This can create a large file if the user requests a large sample size. The option *HIST* will create a histogram for each of the sampled variables. Given the 'width' of the listing file, draws with a large sample size may see the histogram truncated as the largest size of any bar is limited to 90 characters.[1] The third option is *CORR* that will provide the correlation matrix for each draw across all variables, as well as the rank correlation matrix. The three options can be specified in any sequence and combination. If the `REPTS` keyword is absent, the default is to leave out this optional output from the listing file.

Sampling options

Table 2.2 lists the keywords that relate to the overall sample parameters of the run. These are enumerated further below:

1. **SEED** is a user-provided initial seed for the random number generator. This is a required option that must be in the range $[1, 2^{31} - 1]$.
2. **RAND** The default random number generator is quite portable across operating systems [NEW]. The new code also permits the use of the Mersenne Twister algorithm.[2] This code is not necessarily portable but has been tested under Windows with a 64-bit operating system.
3. **REPS** The user can request more than 1 sample per run. The option is not required and the program defaults to 1 replication. This is somewhat different than doing multiple single sample runs as the starting seed is not reset for each sample.

[1] This version of LHS will save the histogram data to a CSV file that can be read into a visualization program. An example Python program that creates the histograms from this dataset is provided in the Appendix.

[2] See Matsumoto and Nishimura (1998) and Nishimura (2000)

Table 2.2: **Keywords related to run options**

Keyword	Specification	Required	Value
`SEED` *iseed*	Random number seed	YES	Integer n $(1 \leq n \leq 2^{31} - 1)$
`RAND`	If keyword is present, the Mersenne Twister algorithm is used to generate random numbers	NO	Default: False
`REPS` *nreps*	Number of multiple LHS sample sets generated	NO	Default: $nreps = 1$
`OBS` *num*	Number of observations per entire LHS sample	YES	Integer number of observations
`OPTS` *options*	How the sampling is done	NO	`RANDOM SAMPLE` is pure Monte Carlo; `RANDOM PAIRING` refers to random pairing for correlation. Omitting the `OPTS` keyword results in restricted pairing.

4. **OBS** is a required parameter that determines the number of draws for each replication. There is no upper bound, though the number of draws will be limited by memory.
5. **OPTS** There are three pairing methods available with the LHS tool. The default is so-called restricted pairing, which uses the LHS methodology for the drawing random deviates. In this case, the ordering of the sample prioritizes the desired correlation across variables. In the absence of user-specified correlation, the target correlation matrix is the identity matrix. If the user specifies a correlation for a subset of variables, all non-specified correlations default to 0. The second sampling method requires `OPTS` to take the value `RANDOM SAMPLE`. In this case, a fully random sample is taken independently for each variable and no attempt is made to pair across variables. The third option is a hybrid and requires setting `OPTS` to take the value `RANDOM PAIRING`. This option causes the tool to draw random deviates according to the LHS methodology, but generates a sample with purely random pairing.

Several keywords from the original package are still active mainly for backward compatibility. These are `WCOL`, `SCOL`, `LHNONAM`, `PVAL`, `PRETRIN` and `DATA:`. Interested users can consult Swiler and Wyss (2004). The first three relate to the output of the data file using the default file format, which is the only option using the original FORTRAN-based package. The option `WCOL` uses a so-called wide format for outputting the data, i.e., the sample for each observation is output on a single line (for all sampled variables). The option `SCOL` outputs the data in a single column. The option `LHNONAM` omits variable names and the point values from the output file. This option is strictly for backward compatibility. `PVAL` refers to options for outputting the so-called point values, a seemingly minor feature of the package. `PRETRIN` allows for splitting the main input file into two—with the run options in one file and the distribution information in a second. `DATA:` allows for interspersing keywords with the other options. The use of `DATASET:` obviates the need for the `DATA:` keyword.

2.1.2 The distribution section

The distribution section starts with the keyword `DATATSET:`. It is followed by a list of the desired distributions, which may also include the keywords `CONSTANT` and `SAME AS`, described further below, and the keyword `CORRELATE`, which can be used to specify a target correlation between any two variables.

1. Each desired distribution is composed of three parts—the variable name, the name of the distribution, for example `NORMAL` and the parameters that define the distribution, for example the mean and the standard deviation.[3]

[3] The original LHS package also allows for the user to specify the so-called point value of the distribution. If the user desires to specify the point value, it should be provided as the second element of the distribution, i.e.,

Table 2.3: **Normal Distribution**—$\mathcal{N}(\mu, \sigma^2)$

PDF	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$
CDF	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	
Mean	μ	$-\infty < \mu < \infty$
Variance	σ^2	$\sigma > 0$
Quantile	$F^{-1}(U) = \mu + \sigma\sqrt{2}\text{erf}^{-1}(2U - 1)$	$0 \leq U \leq 1$
Usage	`NORMAL` *mean standard-deviation*	
Example	`ArmElas NORMAL 3.2 1.43`	

2. The `CONSTANT` keyword allows for inserting a constant in the sample. Note that it will actually only be in the output sample if the keyword `ALL` is specified—see above.
3. The `SAME AS` keyword allows for specifying an alias name for one of the user-specified distribution. As with the `CONSTANT` keyword, the alias will only be in the output fall of the `ALL` keyword is specified. Aliases can only be declared after a variable with its distribution has been defined.
4. The `CORRELATE` keyword has three fields. The first two fields specify the variables with the targeted correlation and the third field is the target correlation. The correlation must be in the range $-1 < corr < 1$.

2.2 The available distributions

This section provides the key information related to the distributions currently available with the LHS tool. There are 41 individual distributions—some of which are derivative of a generic distribution—most of which are continuous, but there are a handful of discrete distributions. Each section will provide the underlying probability distribution function (PDF) and the cumulative distribution function (CDF), the mean and variance, the quantile function (where defined explicitly), the usage and an example. The LHS utility also accepts user-defined continuous and discrete distributions. These options are described in Swiler and Wyss (2004).

2.2.1 The Normal distribution

The LHS tool includes the Normal distribution with a number of its variants. Table 2.3 provides the key specification for the Normal distribution. Figure 1 provides an illustration of a handful of examples of the Normal distribution PDF and the corresponding CDF.

Random samples from the Normal distribution can be generated using inverse transform sampling, i.e., sampling from the Uniform distribution and using the quantile function to derive the relevant deviate. The code includes an algorithm that calculates numerically the inverse of the standard Normal CDF, which itself relies on a numerical approximation of the so-called error function, designated by the symbol erf, which is defined as:

$$\text{erf} = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

The error function is also closely related to the complementary error function, erfc, defined as erfc $= 1 -$ erf. The CDF of the standard normal, designated by Φ, is related to erf and erfc via the following relations:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}}\right) \iff \text{erf}(z) = 2\Phi(\sqrt{2}z) - 1$$
$$\Phi(x) = 1 - \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right) \iff \text{erfc}(z) = 2\left[1 - \Phi(\sqrt{2}z)\right]$$

between the name of the variable and the name of the distribution. If the `PVAL` option is set to 0, the program will output the user specified point value. The default option, when `PVAL` is set to 1, is to output the sample mean as the point value. If the option is set to 2, the program will output the sample median.

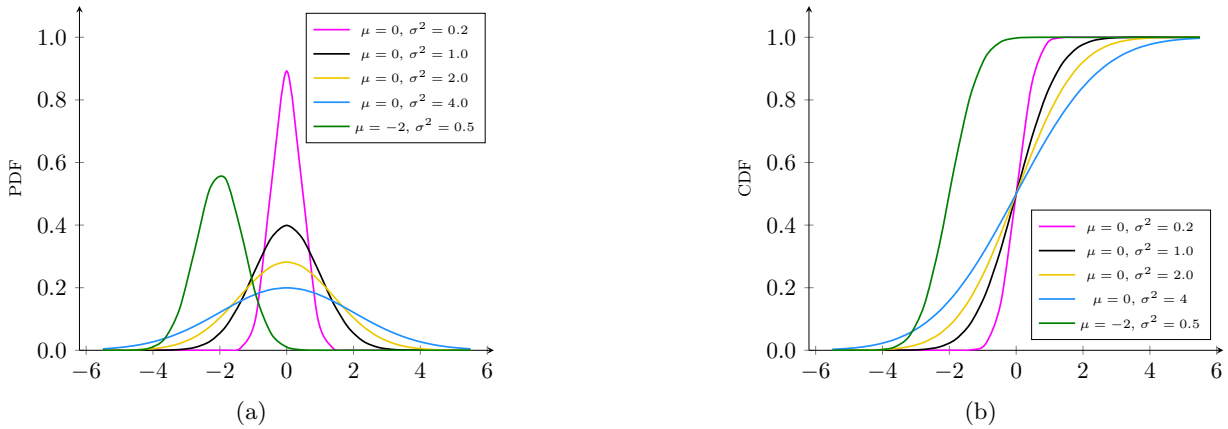


Figure 1: **Examples of the Normal distribution (a) PDF and (b) CDF**

Table 2.4: **Truncated Normal Distribution**

Usage	`TRUNCATED NORMAL` *mean standard-deviation lower upper*
Restrictions	$\sigma > 0$, $0 \leq$ *lower* $<$ *upper* ≤ 1
Example	`ArmElas TRUNCATED NORMAL 3.2 1.43 0.1 0.8`

The Truncated Normal distribution

The Truncated Normal distribution derives from the Normal distribution, but only samples between two quantile limits—lower and upper—provided by the user. Table 2.4 describes its usage. The example is the same as the previous, but limits sampling between the 10 and 80 percent quantiles.

The Bounded Normal distribution

The Bounded Normal distribution derives from the Normal distribution, but only samples between two value limits—lower and upper—provided by the user. Table 2.5 describes its usage. The example is the same as the previous, but limits sampling between the the values of 1 and 6.[4]

2.2.2 The Lognormal distribution

The LHS tool has the Lognormal distribution with a number of its variants. Table 2.6 provides the key specification for the Lognormal distribution. The core functionality of the Lognormal distribution uses the methods of the Normal distribution with the appropriate adjustments. The parameters μ^x and σ^x represent respectively the mean and

[4] The LHS utility also has a Normal distribution variant called `NORMAL-B`. This is included for backward compatibility. Interested users should refer to the Swiler and Wyss (2004) document to see the specification requirements.

Table 2.5: **Bounded Normal Distribution**

Usage	`BOUNDED NORMAL` *mean standard-deviation lower_bound upper_bound*
Restrictions	$\sigma > 0$, *lower_bound* $<$ *upper_bound*
Example	`ArmElas BOUNDED NORMAL 3.2 1.43 1.0 6.0`

Table 2.6: **Lognormal Distribution**—Lognormal(μ, σ^2)

PDF	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$	$0 < x < \infty$
CDF	$F(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right)\right]$	
Mean	$\mu^x = e^{\mu+\frac{\sigma^2}{2}}$	$\mu^x > 0$
Variance	$(\sigma^x)^2 = \left[e^{\sigma^2} - 1\right] e^{2\mu+\sigma^2}$	$\sigma^x > 0$
Quantile	$F^{-1}(U) = \exp\left(\mu + \sigma\sqrt{2}\text{erf}^{-1}(2U-1)\right)$	$0 \leq U \leq 1$
Usage	`LOGNORMAL` *mean error_factor*	
Example	`Inc LOGNORMAL 10 3.66`	

standard deviation of the random variable X, whereas μ and σ are the means of $\ln(X)$. The formulas in Table 2.6 can also be inverted to derive the mean and variable of $\ln(X)$:[5]

$$\mu = \ln\left(\frac{(\mu^x)^2}{\sqrt{(\sigma^x)^2 + (\mu^x)^2}}\right) \qquad \sigma^2 = \ln\left(1 + \frac{(\sigma^x)^2}{(\mu^x)^2}\right)$$

Figure 2 provides an illustration of a handful of examples of the Lognormal distribution PDF and the corresponding CDF.

There are three ways to specify the parameters of the standard Lognormal distribution. In the first, the user provides the mean of the variable X and the so-called *error factor*, described further below. The second, new to this version of the LHS utility, lets the user provide the mean and standard deviation of the variable X. The third method allows the user to specify the mean and standard deviation of the variable $\ln(X)$.

The error factor measures dispersion around the median and is defined as the square root of the ratio of the 95$^{\text{th}}$ percentile with respect to the 5$^{\text{th}}$ percentile. It is linked to the standard deviation of $\ln(X)$, i.e, σ with the following formula:

$$\sigma = \frac{\ln(EF)}{1.645} \iff EF = e^{1.645\sigma}$$

where 1.645 represents the value of the standard normal at the 95$^{\text{th}}$ percentile and the variable EF represents the error factor. In summary, if one knows the mean and standard deviation of X, one can convert the standard deviation to the relevant error factor using the formula:

$$EF = e^{1.645\sigma} = \exp\left(1.645\sqrt{\left[\ln\left(1 + \frac{(\sigma^x)^2}{(\mu^x)^2}\right)\right]}\right)$$

The example below is taken from a possible income distribution, with a mean income of 10 (e.g, \$10,000 per capita) and a Gini coefficient of 0.423. The latter is directly linked to the standard deviation of $\ln(X)$, i.e., σ, via the following relation:

$$\sigma = \sqrt{(2)}F^{-1}\left(\frac{G+1}{2}\right)$$

where F is the CDF of the standard normal distribution and G is the Gini coefficient. In the example $\sigma = 0.62222$. From the equation above, one can derive the error factor: $EF = 3.66$.

A first alternative form for the Lognormal

In addition to the `LOGNORMAL` form, this version of the LHS utility includes `LOGNORMALALT`, which allows the user to input the mean and standard deviation of the distribution, instead of the mean and the error factor. Table 2.7 describes the usage and the example will provide the same deviates as above if starting from the same seed.

A second alternative form for the Lognormal

[5] We can also show that $\mu = ln(\mu^x) - \sigma^2/2$.

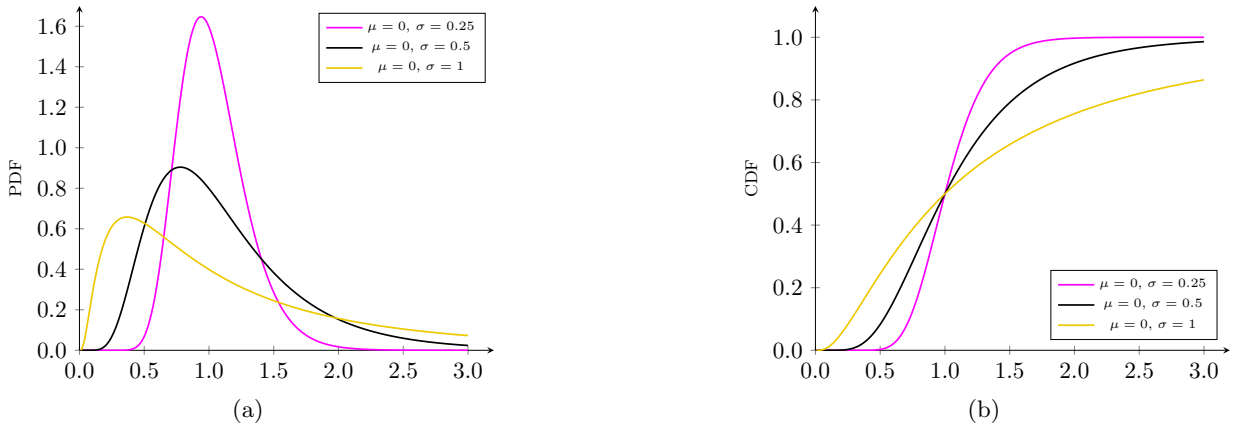


Figure 2: **Examples of the Lognormal distribution (a) PDF and (b) CDF**

Table 2.7: **First alternative Lognormal Distribution**—Lognormal(μ, σ^2)

Usage	`LOGNORMALALT` *mean standard_deviation*
Restrictions	*mean* > 0 and *standard_deviation* > 0
Example	`Inc LOGNORMALALT 10 9.29`

The second alternative allows the user to specify the mean and standard deviation of $\ln(X)$, i.e., the parameters of the underlying normal distribution. The example in Table 2.8 is the same as above and will re-produce the same deviates if starting from the same seed.

Similar to the Normal distribution, the Lognormal distribution includes truncated and bounded variants. For each, the user has two options—specifying the parameters with the mean of the distribution and the error factor, or else with the mean and standard deviation of the underlying Normal distribution.[6]

First version of the truncated Lognormal distribution

The specification for the first version of the truncated Lognormal is provided in Table 2.9. The user inputs the mean and the error factor of the distribution, as well as the quantile limits to be sampled. In the example, the sampling will occur between the 5^{th} and 95^{th} percentiles.

Second version of the truncated Lognormal distribution

The specification for the second version of the truncated Lognormal is provided in Table 2.10. The user inputs the mean and the standard deviation of the underlying Normal distribution, as well as the quantile limits to be sampled. In the example, the sampling will occur between the 5^{th} and 95^{th} percentiles.

[6] We have not added the third option of specifying the mean and standard deviation of the actual distribution.

Table 2.8: **Second alternative Lognormal Distribution**—Lognormal(μ, σ^2)

Usage	`LOGNORMAL-N` *mean standard_deviation*
Restrictions	*standard_deviation* > 0
Example	`Inc LOGNORMAL-N 1.9915 0.7888`

Table 2.9: **Truncated Lognormal distribution—First version**

Usage	`TRUNCATED LOGNORMAL` *mean error_factor lower upper*
Restrictions	*mean* > 0, *error_factor* > 1, $0 \leq$ *lower* $<$ *upper* ≤ 1
Example	`Inc TRUNCATED LOGNORMAL 10 3.66 0.05 0.95`

Table 2.10: **Truncated Lognormal distribution—Second version**

Usage	`TRUNCATED LOGNORMAL-N` *mean standard_deviation lower upper*
Restrictions	*standard_deviation* > 0, $0 \leq$ *lower* $<$ *upper* ≤ 1
Example	`Inc LOGNORMAL-N 1.9915 0.7888 0.05 0.95`

First version of the bounded Lognormal distribution

The specification for the first version of the bounded Lognormal is provided in Table 2.11. The user inputs the mean and the error factor of the distribution, as well as the range to be sampled. In the example, the sampling will occur between 0.5 and 12, i.e., between \$500 and \$12,000.

Second version of the bounded Lognormal distribution

The specification for the second version of the bounded Lognormal is provided in Table 2.12. The user inputs the mean and the standard deviation of the underlying Normal distribution, as well as the limits of the range to be sampled. In the example, the sampling will occur between 0.5 and 12, i.e., between \$500 and \$12,000.

2.2.3 Uniform distribution

The Uniform distribution has the simplest PDF, just the value 1 in the case of the unit support domain. The main characteristics of the Uniform distribution are provided in Table 2.13, for a generic distribution with a support range of [A,B]. Various depictions of the PDF and CDF for the Uniform distribution are provided in Figure 3.

2.2.4 The Log-uniform distribution

The Log-uniform distribution assumes that $\ln(X)$ is distributed uniformly between $\ln(A)$ and $\ln(B)$. The main characteristics of the Log-uniform distribution are provided in Table 2.14, for a distribution with a support range of [A,B]. Various depictions of the PDF and CDF for the Log-uniform distribution are provided in Figure 4.

2.2.5 Beta distribution

The Beta distribution is supported over a range from A to B, and is based on two shape factors, α and β. The key features of the distribution are provided in Table 2.15. Note that there is no explicit form for the quantile function. The source code provides two methods for producing the deviates. The first uses an approximation of the quantile function for the Beta distribution. The second uses the acceptance-rejection sampling technique, which is used for cases when the quantile function is not available. Figure 5 provides an illustration of the Beta distribution for various

Table 2.11: **Bounded Lognormal distribution—First version**

Usage	`BOUNDED LOGNORMAL` *mean error_factor lower_bound upper_bound*
Restrictions	*mean* > 0, *error_factor* > 1, $0 <$ *lower_bound* $<$ *upper_bound*
Example	`Inc TRUNCATED LOGNORMAL 10 3.66 0.5 12`

Table 2.12: **Bounded Lognormal distribution—Second version**

Usage	`BOUNDED LOGNORMAL-N` *mean standard_deviation lower_bound upper_bound*
Restrictions	$standard\_deviation > 0$, $0 \leq lower\_bound < upper\_bound$
Example	`Inc LOGNORMAL-N 1.9915 0.7888 0.5 12`

Table 2.13: **Uniform Distribution—$\mathcal{U}(A, B)$**

PDF	$f(x) = \dfrac{1}{B-A}$	$A \leq x \leq B$
CDF	$F(x) = \dfrac{x-A}{B-A}$	$A \leq x \leq B$
Mean	$\dfrac{A+B}{2}$	
Variance	$\dfrac{(B-A)^2}{12}$	
Quantile	$F^{-1}(U) = A + (B-A)U$	$0 \leq U \leq 1$
Usage	`UNIFORM` A B	$A < B$
Example	`ArmElas UNIFORM 3 6`	

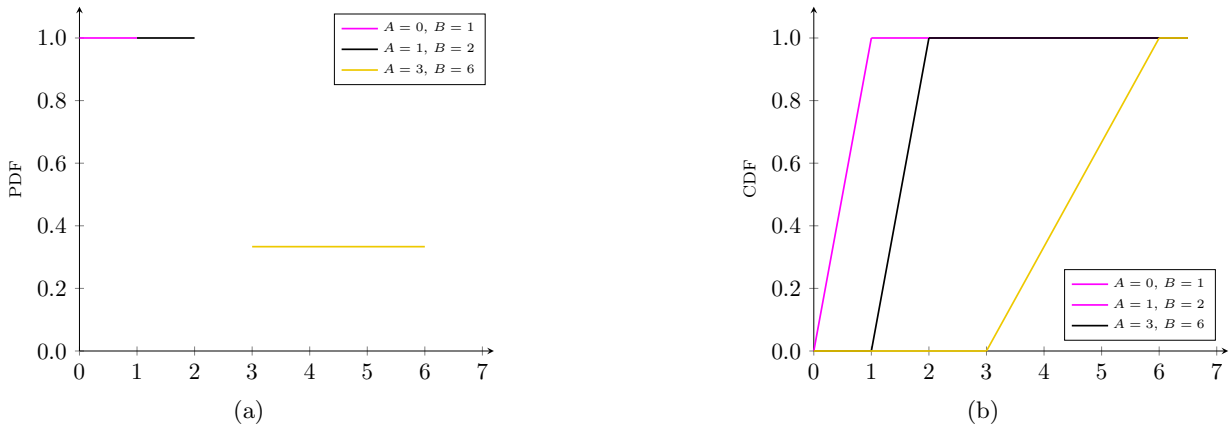


Figure 3: **Examples of the Uniform distribution (a) PDF and (b) CDF**

Table 2.14: **Log-uniform Distribution**—$\mathcal{LU}(A, B)$

PDF	$f(x) = \frac{1}{x\left(\ln(B) - \ln(A)\right)} = \frac{1}{x \ln(\frac{B}{A})}$	$0 < A \leq x \leq B$
CDF	$F(x) = \frac{\ln(x) - \ln(A)}{\ln(B) - \ln(A)}$	
Mean	$\frac{B - A}{\ln(B) - \ln(A)}$	
Variance	$\frac{1}{2}\left(\frac{B^2 - A^2}{\ln(B) - \ln(A)}\right) - \left(\frac{B - A}{\ln(B) - \ln(A)}\right)^2$	
Quantile	$F^{-1}(U) = Ae^{U(\ln(B) - \ln(A))}$	$0 \leq U \leq 1$
Usage	`LOGUNIFORM` A B	$0 < A < B$
Example	`ArmElas LOGUNIFORM 3 6`	

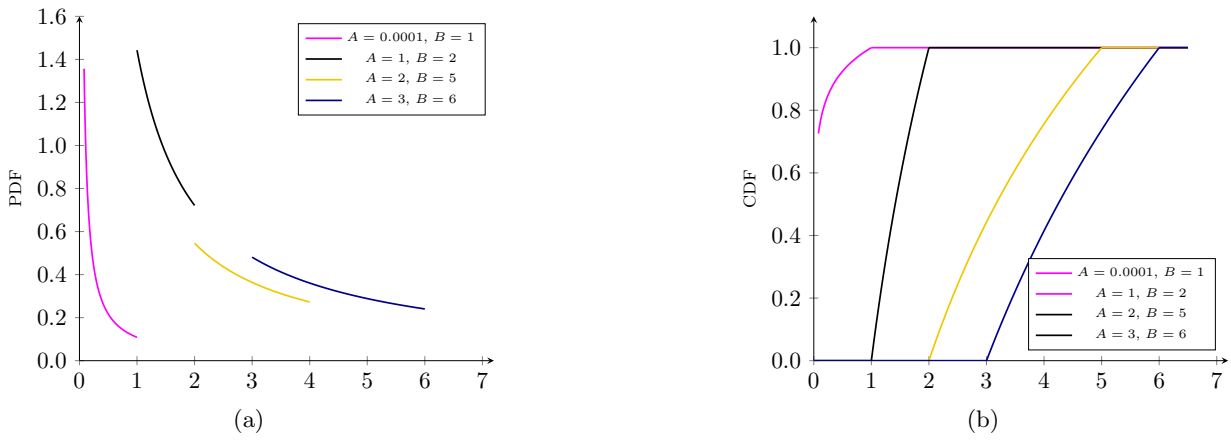


Figure 4: **Examples of the Log-Uniform distribution (a) PDF and (b) CDF**

Table 2.15: **Beta Distribution**—Beta(A, B, α, β)

PDF	$f(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_A^B u^{\alpha-1}(1-u)^{\beta-1}du}$	$A < x < B$
CDF	$F(x) = \dfrac{\int_A^x u^{\alpha-1}(1-u)^{\beta-1}du}{\int_A^B u^{\alpha-1}(1-u)^{\beta-1}du}$	
Mean	$\dfrac{\alpha}{\alpha+\beta}$	
Variance	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
Usage	`BETA` A B α β	
Restrictions	$0 \leq A < B$, $0.001 < \alpha, \beta < 1e7$	
Example	`inc Beta 0.1 50 2 5`	

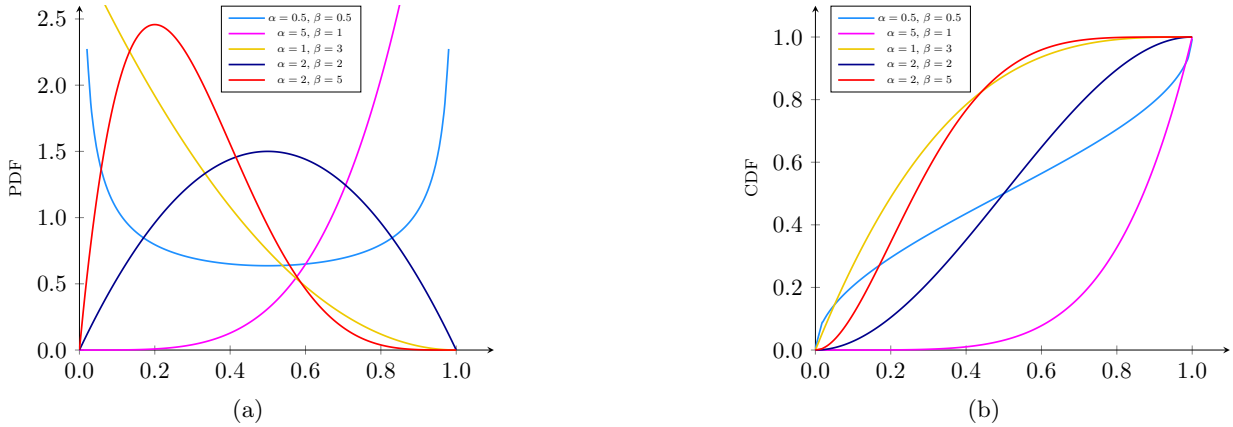


Figure 5: **Examples of the Beta distribution (a) PDF and (b) CDF**

values of the shape parameters. The example samples the Beta distribution over the range 0.1 through 50 with shape parameters 2 and 5.

2.2.6 The Cauchy distribution

The Cauchy distribution has two parameters: μ, which is the location parameter and γ, which is the scale parameter. The key characteristics of the distribution are provided in Table 2.16. Figure 6 provides indicative profiles of the Cauchy PDF and CDF for various values of the parameters.

2.2.7 The Dagum distribution

The Dagum distribution has three parameters: a, b and p. The distribution has often been used in the analysis of income distribution. The key characteristics of the distribution are provided in Table 2.17. Figure 7 provides indicative profiles of the Dagum PDF and CDF for various values of the parameters.

2.2.8 The Exponential distribution

The exponential distribution has a single parameter, λ. The key characteristics of the distribution are provided in Table 2.18. Figure 8 provides indicative profiles of the exponential PDF for various values of the λ parameter.

Table 2.16: **Cauchy Distribution**—Cauchy(μ, γ)

PDF	$f(x) = \frac{1}{\pi}\left[\frac{\gamma}{\gamma^2 + (x-\mu)^2}\right]$	$x \in \mathbb{R}$
CDF	$F(x) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\gamma}\right) + \frac{1}{2}$	
Mean	Not defined	
Variance	Not defined	
Quantile	$F^{-1}(U) = \mu + \gamma\tan\left(\pi\left(p - \frac{1}{2}\right)\right)$	$[0, 1)$
Usage	`CAUCHY mu gamma`	$\gamma > 0$
Example	`x1 CAUCHY 0 1`	

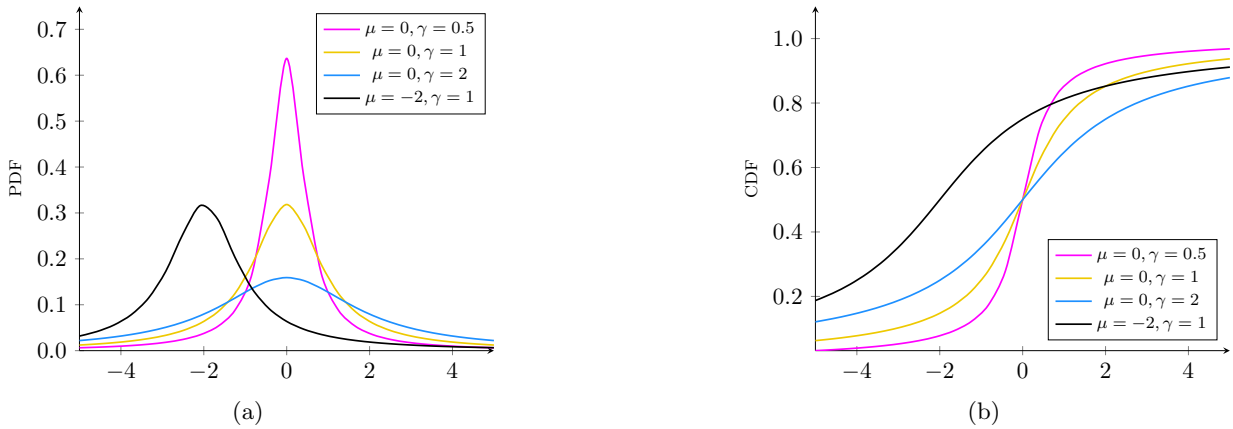


Figure 6: **Examples of the Cauchy distribution (a) PDF and (b) CDF**

Table 2.17: **Dagum Distribution**—Dagum(a, b, p)

PDF	$f(x) = \frac{ap}{x}\left(\frac{\left(\frac{x}{b}\right)^{ap}}{\left(\left(\frac{x}{b}\right)^{a} + 1\right)^{p+1}}\right)$	$x > 0$
CDF	$F(x) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p}$	
Mean	$b\frac{\Gamma\left(1 - \frac{1}{a}\right)\Gamma\left(p + \frac{1}{a}\right)}{\Gamma(p)}$	if $a > 1$
Variance	$b^2\left[\frac{\Gamma\left(1-\frac{2}{a}\right)\Gamma\left(p+\frac{2}{a}\right)}{\Gamma(p)} - \left(\frac{\Gamma\left(1-\frac{1}{a}\right)\Gamma\left(p+\frac{1}{a}\right)}{\Gamma(p)}\right)^2\right]$	
Quantile	$F^{-1}(U) = b\left[U^{-1/p} - 1\right]^{-1/a}$	$[0, 1)$
Usage	`DAGUM a b p`	$a > 0$, $b > 0$, $p > 0$
Example	`x5 DAGUM 2 1 2`	

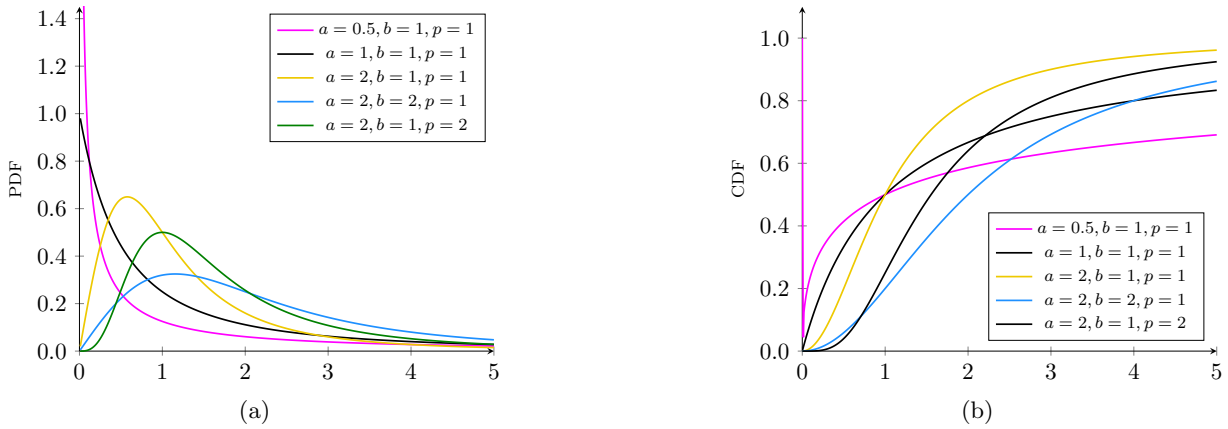


Figure 7: **Examples of the Dagum distribution (a) PDF and (b) CDF**

Table 2.18: **Exponential Distribution**—Exponential(λ)

PDF	$f(x) = \lambda e^{-\lambda x}$	$x \geq 0$
CDF	$F(x) = 1 - e^{-\lambda x}$	
Mean	$\frac{1}{\lambda}$	
Variance	$\frac{1}{\lambda^2}$	
Quantile	$F^{-1}(U) = -\frac{\ln(1-U)}{\lambda}$	$[0, 1)$
Usage	`EXPONENTIAL` *lambda*	$\lambda > 0$
Example	`decay EXPONENTIAL 2`	

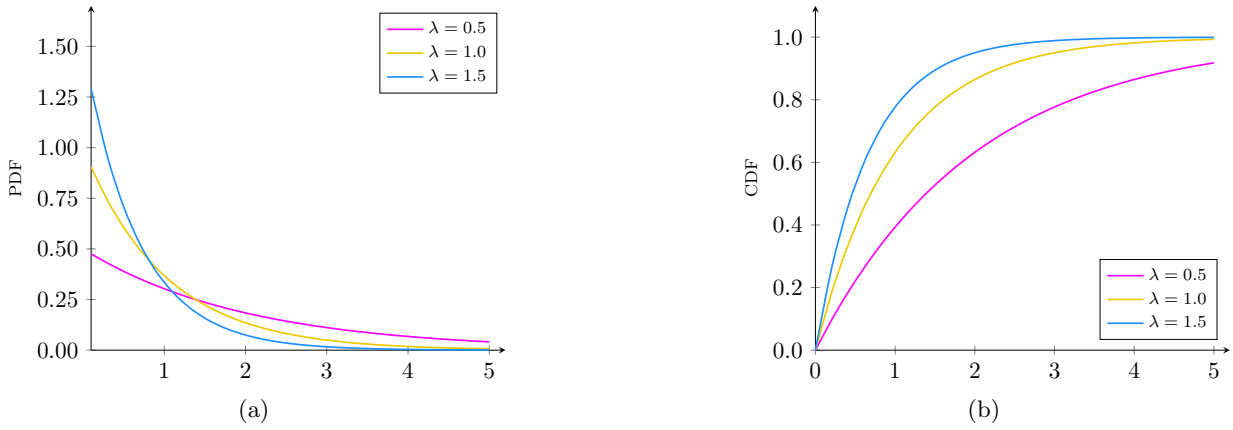


Figure 8: **Examples of the Exponential distribution (a) PDF and (b) CDF**

Table 2.19: **Fréchet Distribution**—Frechet(α, s, m)

PDF	$f(x) = \frac{\alpha}{s}\left(\frac{x-m}{s}\right)^{-1-\alpha} e^{-\left(\frac{x-m}{s}\right)^{-\alpha}}$	$x > m$
PDF	$f(x) = \alpha x^{-1-\alpha} e^{-x^{-\alpha}}$	$x > 0$, $m = 0$, $s = 1$
CDF	$F(x) = e^{-\left(\frac{x-m}{s}\right)^{-\alpha}}$	$x > m$
CDF	$F(x) = e^{-x^{-\alpha}}$	$x > 0$, $m = 0$, $s = 1$
Mean	$\mu = \begin{cases} m + s\Gamma\left(1 - \frac{1}{\alpha}\right) & \text{if } \alpha > 1 \\ \infty & \text{if } \alpha \leq 1 \end{cases}$	
Variance	$\sigma^2 = \begin{cases} s^2\left[\Gamma\left(1 - \frac{2}{\alpha}\right) - \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2\right] & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \leq 2 \end{cases}$	
Quantile	$F^{-1}(U) = m + s\left(-\ln(U)\right)^{-1/\alpha}$	$(0, 1]$
Usage	`FRECHET` *alpha s m*	$\alpha > 0$, $s > 0$
Example	`productivity FRECHET 4.5 1 0`	

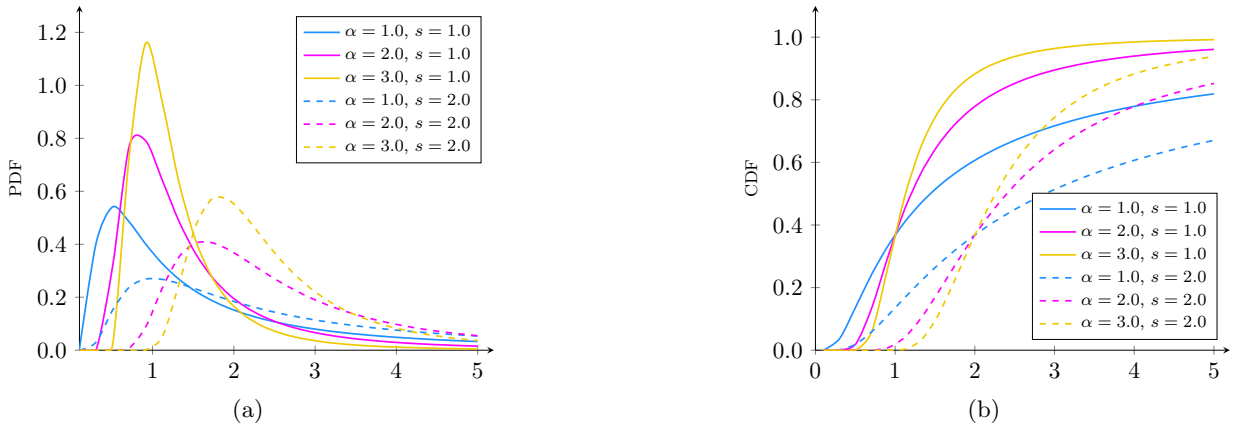


Figure 9: **Examples of the Fréchet distribution (a) PDF and (b) CDF**

2.2.9 Fréchet distribution

The Fréchet distribution is a special case of the generalized extreme value distribution. It is used in economics, for example to sample the distribution of productivity across heterogeneous firms. The main characteristics of the distribution are provided in Table 2.19. The parameter $\alpha > 0$ is called the shape parameter. In the generalized form, the parameter $s > 0$ is a scale parameter and the parameter m, the minimum, is the so-called location parameter. The user must specify all three parameters in the order α, s and m, where α and s must be greater than zero. Figure 9 provides a graphical depiction of the Fréchet distribution for various values of the α and s parameters.

2.2.10 Gamma distribution

The Gamma distribution is a two-parameter distribution the key characteristics of which are provided in Table 2.20. The user must specify α, the parameter of the Gamma function, and β, a scaling factor, both of which are real numbers. Examples of the PDF and CDF are provided in figure 10. The $\gamma(\alpha, \beta x)$ function is known as the lower incomplete gamma function. There is no explicit quantile function for this distribution and sampling uses the acceptance/rejection technique. The mean provided by this distribution is α/β. Note that there are two different forms for the Gamma distribution, which differ in the definition of β. In this form, β is often referred to as a *rate*. In

Table 2.20: **Gamma Distribution**—Gamma(α, β)

PDF	$f(x) = \frac{\beta^{\alpha} x^{(\alpha-1)} e^{-\beta x}}{\Gamma(\alpha)}$	$x > 0$
Note	$\Gamma(\alpha) = \int_0^{\infty} y^{(\alpha-1)} e^{-y} dy$	
CDF	$F(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$	
Note	$\gamma(\alpha, \beta x) = \int_0^{\beta x} y^{(\alpha-1)} e^{-y} dy$	
Mean	α/β	
Variance	α/β^2	
Usage	`GAMMA` *alpha beta*	$\alpha > 0$, $\beta > 0$
Example	`delay GAMMA 2 3`	

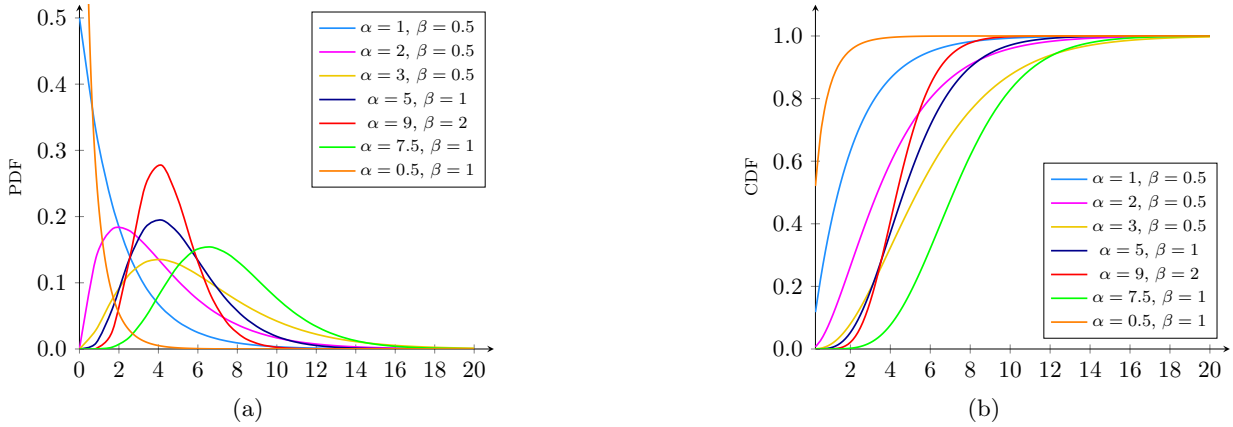


Figure 10: **Examples of the Gamma distribution (a) PDF and (b) CDF**

the other form, that uses the reciprocal of β, it is often referred to as a *scale* parameter. The mean in the other form of the Gamma distribution is $\alpha\beta$. Users should make sure that they are using the correct value for β, i.e., consistent with its use herein.

2.2.11 Gompertz distribution

The Gompertz distribution has two parameters: η, which is the shape parameter and γ, which is the scale parameter. The key characteristics of the distribution are provided in Table 2.21. Figure 11 provides indicative profiles of the Gompertz PDF and CDF for various values of the parameters.

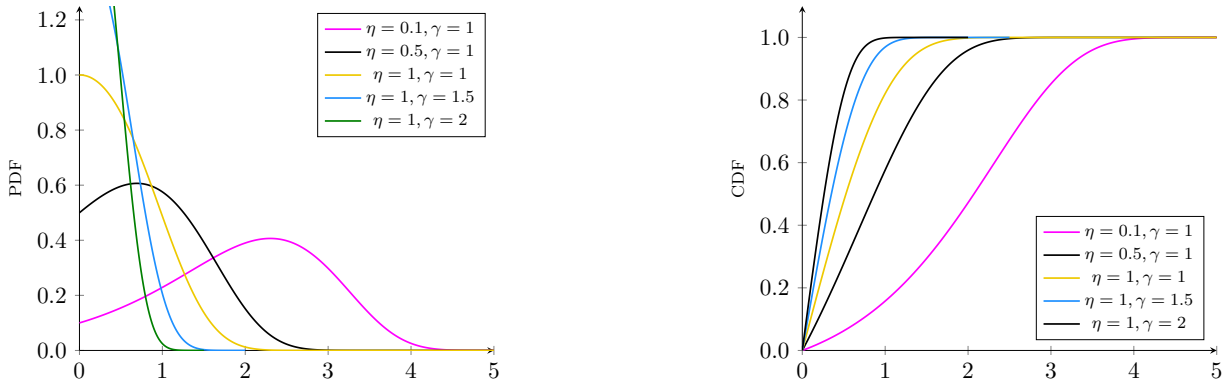
2.2.12 Gumbel distribution

The Gumbel distribution is another special case of the generalized extreme value distribution. The user must specify two parameters in the order μ and β, where β must be greater than zero:[7] Figure 12 provides a graphical depiction of the Gumbel distribution for various values of the μ and β parameters.

[7] The latest LHS code from Sandia includes both the Fréchet and Gumbel distributions. The input of the parameters of the FORTRAN-based Gumbel is reversed—the scale parameter comes first and the locational parameter comes second. In addition, the input scale parameter (*beta*) is the inverse.

Table 2.21: **Gompertz Distribution**—Gompertz(η, γ)

PDF	$f(x) = \eta\gamma \exp\left(\eta + \gamma x - \eta e^{\gamma x}\right)$	$x \in [0, \infty)$
CDF	$F(x) = 1 - \exp\left(\eta - \eta e^{\gamma x}\right)$	
Mean	$\frac{1}{\gamma} e^{\eta} \mathrm{Ei}(-\eta) \qquad$ where $\mathrm{Ei}(x) = \int_{-x}^{\infty} \left(e^{-v}/v\right) dv$	
Quantile	$F^{-1}(U) = \frac{1}{\gamma} \ln\left(1 - \frac{\ln(1-U)}{\eta}\right)$	$[0, 1)$
Usage	`GOMPERTZ eta gamma`	$\eta > 0$, $\gamma > 0$
Example	`x5 GOMPERTZ 2 1`	

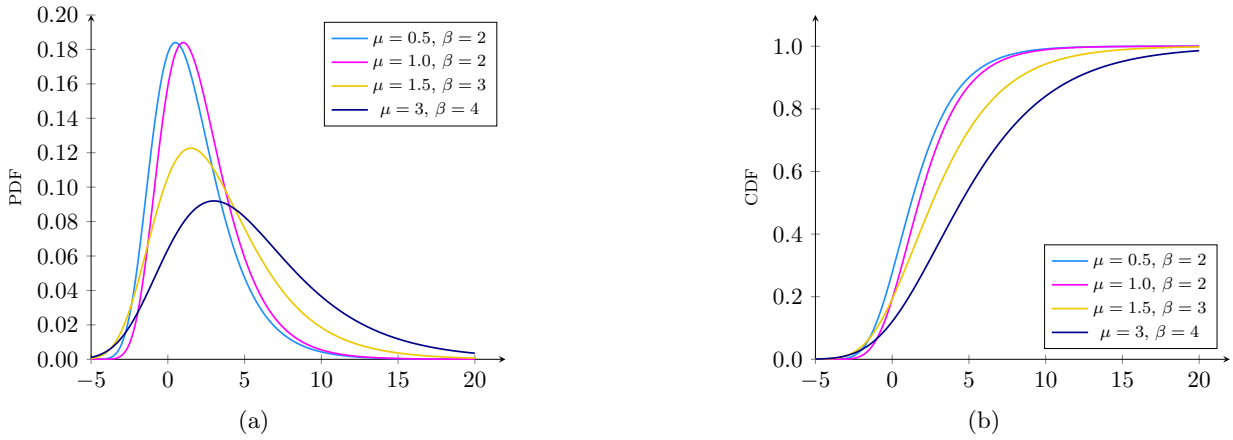


(a) (b)

Figure 11: **Examples of the Gompertz distribution (a) PDF and (b) CDF**

Table 2.22: **Gumbel Distribution**—Gumbel(μ, β)

PDF	$f(x) = \frac{1}{\beta} e^{-\left(z + e^{-z}\right)}$	$z = \frac{x-\mu}{\beta}$
PDF	$f(x) = e^{-\left(x + e^{-x}\right)}$	$\mu = 0, \beta = 1$
CDF	$F(x) = e^{-e^{-(x-\mu)/\beta}}$	
CDF	$F(x) = e^{-e^{-x}}$	$\mu = 0, \beta = 1$
Mean	$\mu + \beta \cdot \gamma$	
Note	$\gamma \approx 0.5772$ is the Euler-Mascheroni constant	
Variance	$\frac{\pi^2}{6} \beta^2$	
Quantile	$F^{-1}(U) = \mu - \beta \ln\left(-\ln(U)\right)$	
Usage	`GUMBEL` *mu beta*	$\beta > 0$
Example	`productivity GUMBEL 0 1`	



(a) (b)

Figure 12: **Examples of the Gumbel distribution (a) PDF and (b) CDF**

Table 2.23: **Inverse Gaussian Distribution**—IG(μ, λ)

PDF	$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\left(\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)\right]$	$x > 0, \mu > 0, \lambda > 0$
CDF	$F(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)$	
Note	Φ is the standard normal distribution CDF	
Mean	μ	
Variance	μ^3/λ	
Usage	`INVERSE GAUSSIAN` *mu lambda*	$\mu > 0, \lambda > 0$
Example	`income INVERSE GAUSSIAN 10 100`	

2.2.13 Inverse Gaussian distribution

An inverse Gaussian distribution is a two-parameter distribution where the user must specify the distribution parameters μ and λ, both greater than zero. The key characteristics are provided in Table 2.23. Examples of the distribution are depicted in figure 13. There is no explicit quantile function of the Inverse Gaussian Distribution and another technique is used to generate random deviates.

2.2.14 Kumaraswamy distribution

The Kumaraswamy distribution is a double-bounded distribution defined on the support domain $[A, B]$. The normalized Kumaraswamy distribution is defined on the unit support domain, i.e., $[0, 1]$, and we can use the transformation $x = (z - A)/(B - A)$ to convert from the generic support domain to the unit support domain. The parameters α and β are non-negative shape parameters. Representative forms of the PDF and CDF are provided in Figure 14. The user must specify four parameters in the order α, β, A and B, where $\alpha > 0$, $\beta > 0$, and $A < B$.

The moments of the Kumaraswamy distribution can be derived from the following expression:

$$M(n) = \int_0^1 x^n f(x) dx = \int_0^1 x^n \alpha\beta x^{\alpha-1} \left(1 - x^\alpha\right)^{\beta-1} dx$$

Let $t = x^\alpha$, which implies $dt = \alpha x^{\alpha-1} dx$ or $dx = t^{1/\alpha - 1} dt/\alpha$, thus

$$M(n) = \alpha\beta \int_0^1 (t^{1/\alpha})^{n+\alpha-1} (1-t)^{\beta-1} t^{1/\alpha-1} dt/\alpha = \beta \int_0^1 t^{n/\alpha} (1-t)^{\beta-1} dt = \beta B(1 + \frac{n}{\alpha}, \beta)$$

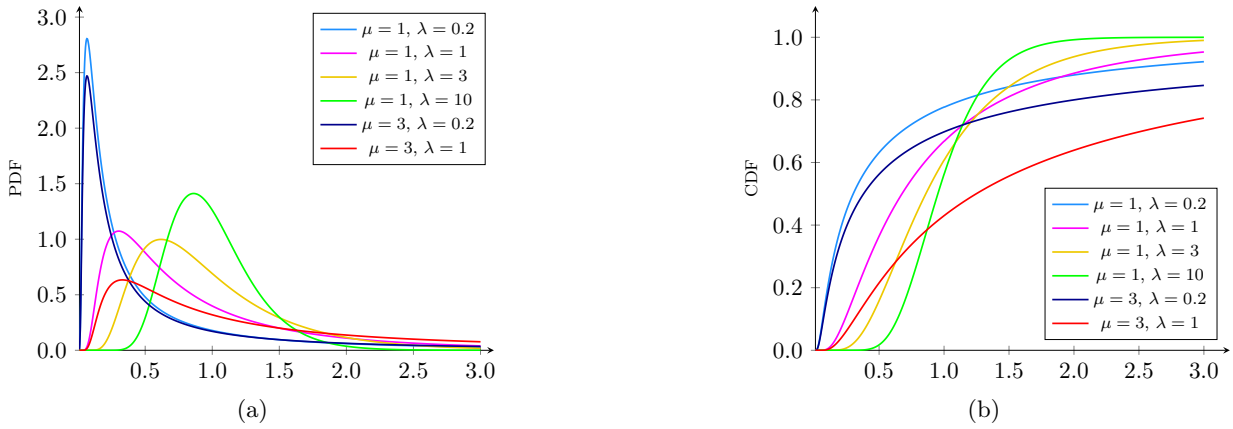


Figure 13: **Examples of the Inverse Gaussian distribution (a) PDF and (b) CDF**

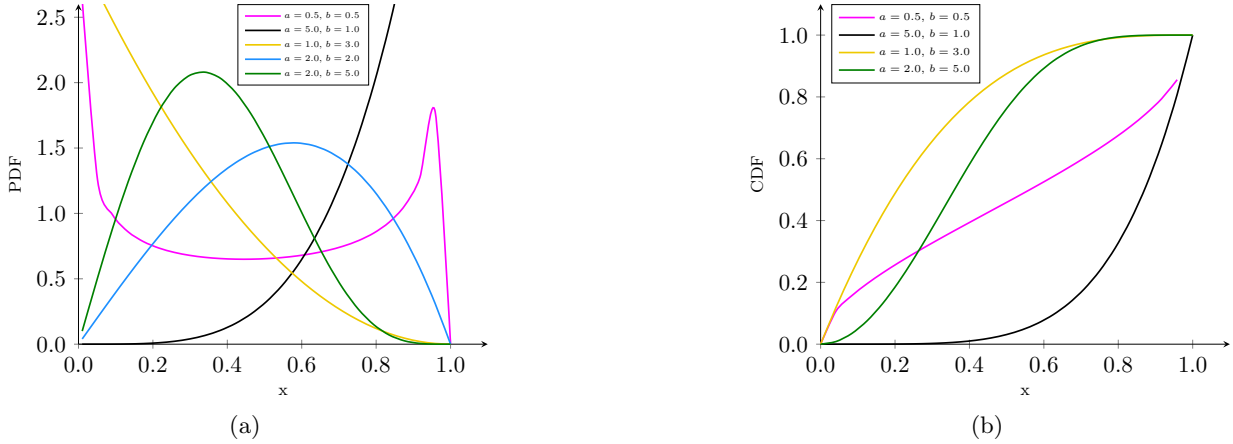


Figure 14: **Examples of the Kumaraswamy distribution (a) PDF and (b) CDF**

where B is the Beta function.[8]

2.2.15 Laplace distribution

The Laplace distribution has two parameters: μ, which is the location parameter and γ, which is the scale parameter. The key characteristics of the distribution are provided in Table 2.25. Figure 15 provides indicative profiles of the Laplace PDF and CDF for various values of the parameters.

2.2.16 Lévy distribution

The Lévy distribution is a two-parameter distribution defined over the domain $x \geq \mu$, where μ is the *location* parameter and c is the *scale* parameter. The main features of the distribution are provided in Table 2.26. The user must specify two parameters in the order μ and c, where $c > 0$. Examples of the PDF and CDF are provided in Figure 16.

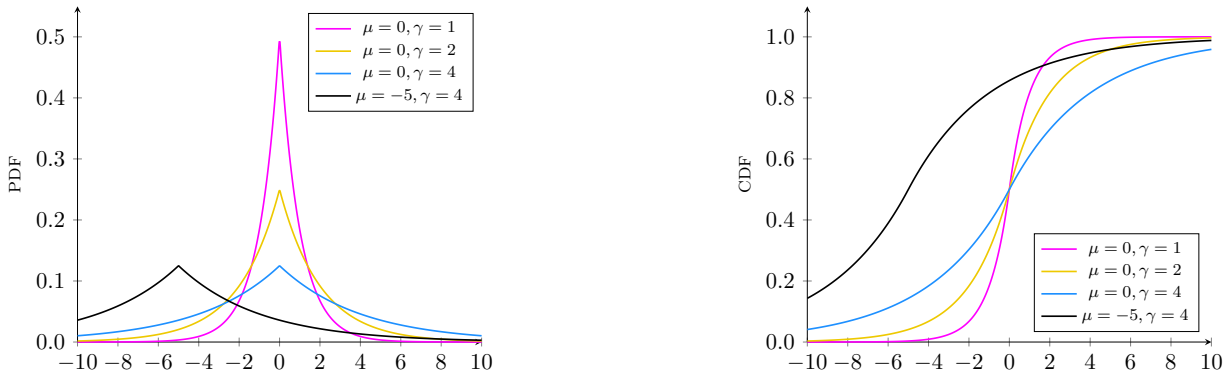
[8] Excel does not include the Beta function but it can be derived from the log of the Gamma function, which is in Excel. `B(a,b) = EXP(GAMMALN(a)+GAMMALN(b)-GAMMALN(a+b))`.

Table 2.24: **Kumaraswamy Distribution**—Kumaraswamy(a, b)

PDF	$f(z) = \dfrac{\alpha\,\beta}{B-A} z^{\alpha-1}\left(1-z^{\alpha}\right)^{\beta-1}$	$A < z < B$
PDF	$f(x) = \alpha\beta x^{\alpha-1}\left(1-x^{\alpha}\right)^{\beta-1}$	$x \in (0,1)$
CDF	$F(z) = 1 - \left[1 - \left(\dfrac{z-A}{B-A}\right)^{\alpha}\right]^{\beta}$	
CDF	$F(x) = 1 - (1-x^{\alpha})^{\beta}$	
Mean	$A + (B-A)\dfrac{\beta\,\Gamma\left(1+\dfrac{1}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\dfrac{1}{\alpha}+\beta\right)}$ $= A + (B-A)\beta\,\mathrm{B}\left(1+\dfrac{1}{\alpha},\beta\right)$	
Mean	$\dfrac{\beta\,\Gamma\left(1+\dfrac{1}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\dfrac{1}{\alpha}+\beta\right)} = \beta\,\mathrm{B}\left(1+\dfrac{1}{\alpha},\beta\right)$	
Note	B is the Beta function, in addition $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$	
Variance	$(B-A)^2\left[\beta\,\mathrm{B}\left(1+\dfrac{2}{\alpha},\beta\right) - \left(\beta\,\mathrm{B}\left(1+\dfrac{1}{\alpha},\beta\right)\right)^2\right]$	
Variance	$\beta\,\mathrm{B}\left(1+\dfrac{2}{\alpha},\beta\right) - \left(\beta\,\mathrm{B}\left(1+\dfrac{1}{\alpha},\beta\right)\right)^2$	
Quantile	$F^{-1}(U) = A + (B-A)\left(1-(1-U)^{1/\beta}\right)^{1/\alpha}$	
Usage	`KUMARASWAMY` *alpha beta A B*	$\alpha > 0$, $\beta > 0$, $A < B$
Example	`rho2 KUMARASWAMY 1.013 1.691 0.0167 0.048`	

Table 2.25: **Laplace Distribution**—Laplace(μ, γ)

PDF	$f(x) = \dfrac{1}{2\gamma}\exp\left(-\dfrac{\lvert x-\mu\rvert}{\gamma}\right)$	$x \in \mathbb{R}$
CDF	$F(x) = \begin{cases} \frac{1}{2}exp\left(\frac{x-\mu}{\gamma}\right) & \text{if } x < \mu \\ 1 - \frac{1}{2}exp\left(-\frac{x-\mu}{\gamma}\right) & \text{if } x \geq \mu \end{cases}$	
Mean	μ	
Variance	$2\gamma^2$	
Quantile	$F^{-1}(U) = \begin{cases} \mu + \gamma\ln(2U) & \text{if } U \leq \frac{1}{2} \\ \mu - \gamma\ln(2-2U) & \text{if } U > \frac{1}{2} \end{cases}$	$[0,1)$
Usage	`LAPLACE mu gamma`	$\gamma > 0$
Example	`x2 LAPLACE 0 1`	

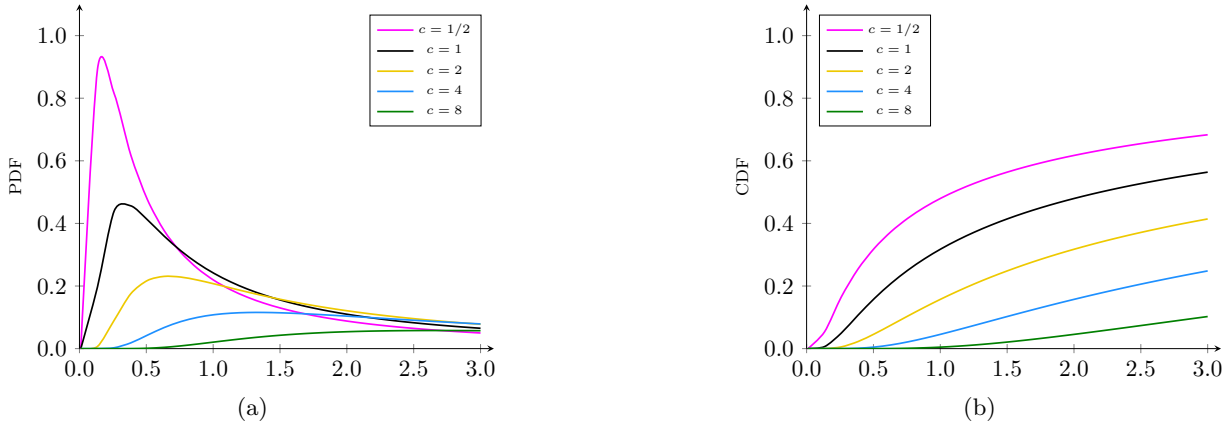


(a) (b)

Figure 15: **Examples of the Laplace distribution (a) PDF and (b) CDF**

Table 2.26: **Lévy Distribution**—Lévy(μ, c)

PDF	$f(x) = \sqrt{\frac{c}{2\pi}} \frac{1}{(x-\mu)^{3/2}} e^{-c/(2(x-\mu))}$	$x \geq \mu$
CDF	$F(x) = \operatorname{erfc}\left(\sqrt{\frac{c}{2(x-\mu)}}\right)$	
Note	$\operatorname{erfc}(z)$ is the complementary error function: $1 - \operatorname{erf}(z)$	
Mean	∞	
Variance	∞	
Quantile	$F^{-1}(U) = \frac{c}{(\Phi^{-1}(1-U/2))^2} + \mu$	
Note	Φ is the CDF of the standard normal distribution	
Usage	`LEVY` *mu c*	$c > 0$
Example	`alpha3 LEVY 0.276 0.00000039629`	



(a) (b)

Figure 16: **Examples of the Lévy distribution (a) PDF and (b) CDF**

Table 2.27: **Logistic Distribution**—Logistic(μ, s)

PDF	$f(x) = \dfrac{\mathrm{e}^{-(x-\mu)/s}}{s\left(1+\mathrm{e}^{-(x-\mu)/s}\right)^2} = \dfrac{1}{4s}\mathrm{sech}^2\left(\dfrac{x-\mu}{2s}\right)$	$s > 0$
CDF	$F(x) = \dfrac{1}{1+\mathrm{e}^{-(x-\mu)/s}} = \dfrac{1}{2} + \dfrac{1}{2}\tanh\left(\dfrac{x-\mu}{2s}\right)$	
Mean	μ	
Variance	$\dfrac{s^2\pi^2}{3}$	
Quantile	$F^{-1}(U) = 2s\tanh^{-1}(2U-1) + \mu = s\ln\left(\dfrac{U}{1-U}\right) + \mu$	
Note	$\tanh^{-1}(x) = \dfrac{1}{2}\ln\left(\dfrac{1+x}{1-x}\right)$	
Usage	`LOGISTIC` *mu s*	$s > 0$
Example	`alpha1 LOGISTIC 0.224 0.0000339623`	

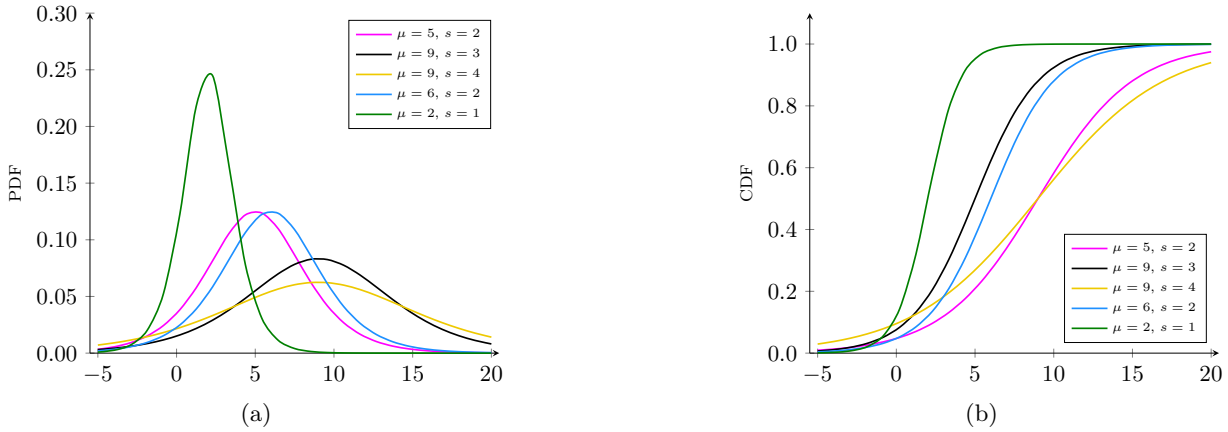


Figure 17: **Examples of the Logistic distribution (a) PDF and (b) CDF**

2.2.17 Logistic distribution

The Logistic distribution is a two-parameter distribution where μ is the *location* parameter and s is the *scale* parameter. It is similar to the Normal distribution, but with heavier tails. The key characteristics are provided in Table 2.27. Examples of the Logistic distribution PDF and CDF are provided in Fig 17.

2.2.18 Maximum Entropy distribution

The Maximum Entropy distribution is a truncated Exponential distribution, where $[A, B]$ represents the support range of the distribution, and μ is the mean. The main characteristics of the Maximum Entropy distribution are provided in Table 2.28. The user must specify $0 \leq A < \mu < B$. The LHS utility will calculate the parameter λ using a numerical technique based on the formula for the mean in Table 2.28, which has no explicit expression for λ as a function of the mean. The parameter β can be derived from the CDF at its value 1, and is equal to $\left(e^{-\lambda A} - e^{-\lambda B}\right)^{-1}$. With A equal to 0 and B converging towards infinity, the distribution converges to the Exponential distribution, with the mean converging towards $1/\lambda$ and β converging towards 1. Figure 18 shows several PDFs and CDFs for the Maximum Entropy distribution with $A = 0.0$, $B = 3.0$, and varying μ.

Table 2.28: **Maximum Entropy Distribution**—Maximum Entropy(μ)

PDF	$f(x) = \beta\lambda e^{-\lambda x}$	$0 \leq A \leq x \leq B$
CDF	$F(x) = \beta\left(e^{-\lambda A} - e^{-\lambda x}\right)$	
Mean	$\frac{\beta}{\lambda}\left[(1+\lambda A)\, e^{-\lambda A} - (1+\lambda B)\, e^{-\lambda B}\right]$	
MGF	$MGF(t) = \frac{\beta\lambda}{t-\lambda}\left[e^{B(t-\lambda)} - e^{A(t-\lambda)}\right]$	
Quantile	$F^{-1}(U) = -\frac{\ln\left(e^{-\lambda A} - U/\beta\right)}{\lambda}$	
Usage	`MAXIMUM ENTROPY` A mu B	$0 \leq A < \mu < B$
Example	`alpha1 MAXIMUM ENTROPY 0 1 3`	

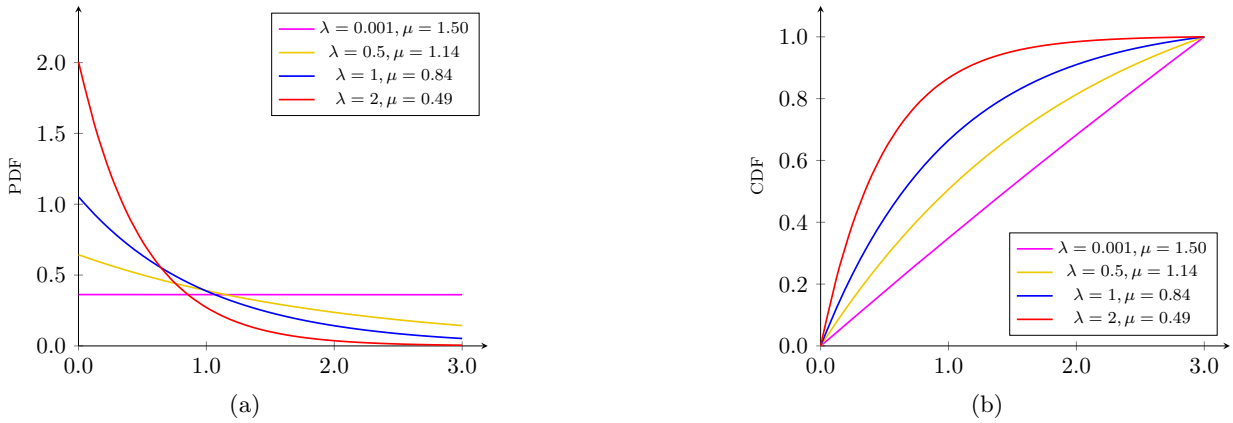


Figure 18: **Examples of the Maximum Entropy distribution (a) PDF and (b) CDF ; for $A = 0$, $B = 3$**

Table 2.29: **Pareto Distribution**—Pareto(α, β)

PDF	$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} = \frac{\alpha}{x}\left(\frac{\beta}{x}\right)^\alpha$	$0 < \beta \leq x < \infty$
CDF	$F(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha$	
Mean	$\mu = \begin{cases} \frac{\alpha\beta}{\alpha - 1} & \text{if } \alpha > 1 \\ \infty & \text{if } \alpha \leq 1 \end{cases}$	
Variance	$\sigma^2 = \begin{cases} \frac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2} & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \leq 2 \end{cases}$	
Quantile	$F^{-1}(U) = \beta\,(1-U)^{(-1/\alpha)}$	$U \in [0, 1]$
Usage	`PARETO` *alpha beta*	$\alpha > 2$, $\beta > 0$
Example	`alpha1 PARETO 2.4 0.5`	

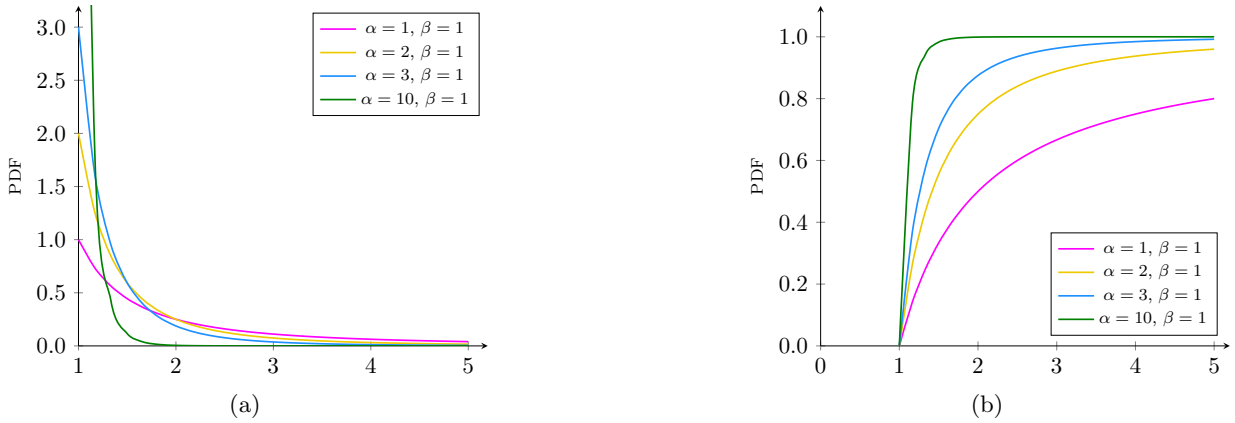


Figure 19: **Examples of the Pareto distribution (a) PDF and (b) CDF**

2.2.19 Pareto distribution

The Pareto distribution is a two-parameter distribution with shape parameter α and minimum value given by β. It has often been used to describe income distribution. Table 2.29 provides the key characteristics of the distribution. Sample distributions are provided in Figure 19.

The original LHS program restricted the parameters so that variance was finite, i.e., $\alpha > 2$. The new LHS utility adds the distribution `PARETOALT`, which only restricts α to be positive, see Table 2.30.

2.2.20 Rayleigh distribution

The Rayleigh distribution has a single scalar parameter, σ. The key characteristics of the distribution are provided in Table 2.31. Figure 20 provides indicative profiles of the Rayleigh PDF and CDF for various values of the parameters.

2.2.21 Triangular distribution

The Triangular distribution is a continuous distribution in the shape of a triangle with a lower limit of A, an upper limit of B and a maximum, or mode, at m. The main characteristics of the Triangular distribution are provided in Table 2.32. Figure 21 shows different shapes for the distribution. Three of the distributions have the same end points, but different modes. There are two special cases. The first is when the mode equals the left end point ($m = A$), and the second is when the mode equals the right end point ($m = B$). The special cases are detailed in Table 2.32.

Table 2.30: **Alternative Pareto Distribution**

Usage	`PARETOALT` *alpha beta*
Restrictions	$\alpha > 0$, $\beta > 0$
Example	`income PARETOALT 1.7 53`

Table 2.31: **Rayleigh Distribution**—Rayleigh(σ)

PDF	$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$	$x \in [0, \infty)$
CDF	$F(x) = 1 - e^{-x^2/(2\sigma^2)}$	
Mean	$\sigma\sqrt{\frac{\pi}{2}}$	
Variance	$\frac{4-\pi}{2}\sigma^2$	
Quantile	$F^{-1}(U) = \sigma\sqrt{-2\ln(1-U)}$	$[0, 1)$
Usage	`RAYLEIGH sigma`	$\sigma > 0$
Example	`x3 RAYLEIGH 2`	

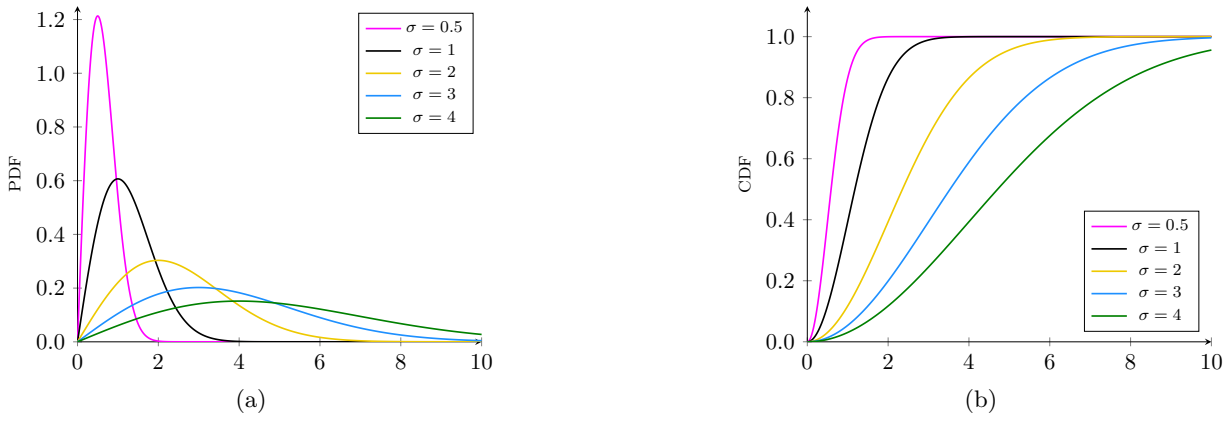


Figure 20: **Examples of the Rayleigh distribution (a) PDF and (b) CDF**

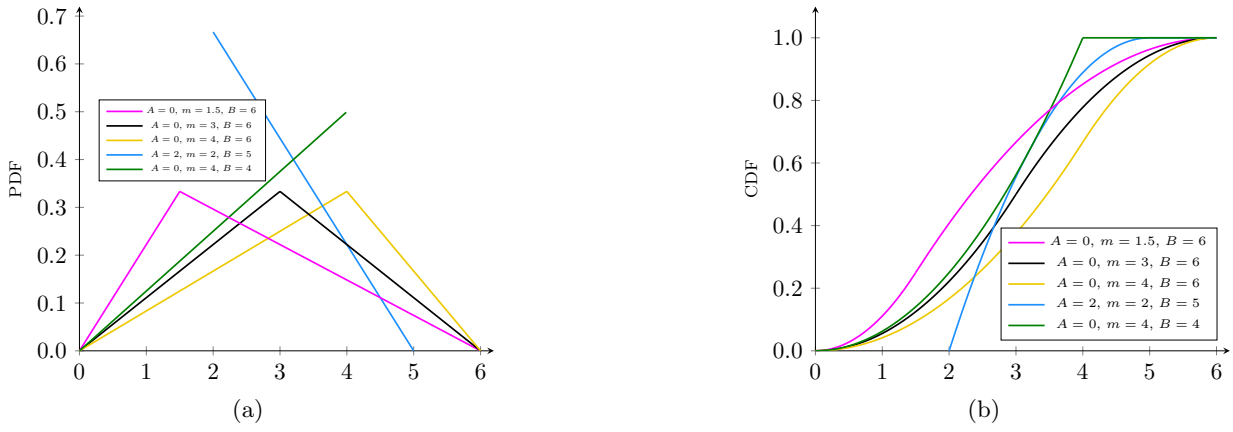


Figure 21: **Examples of the Triangular distribution (a) PDF and (b) CDF**

Table 2.32: **Triangular Distribution**—$Tr(A, B)$

PDF	$f(x) = \begin{cases} \frac{2(x-A)}{(B-A)(m-A)} & \text{if } A \le x \le m \\ \frac{2(B-x)}{(B-A)(B-m)} & \text{if } m \le x \le B \end{cases}$	$A < B$ $A \le m \le B$
PDF	$f(x) = \frac{2(B-x)}{(B-A)^2}$ if $A \le x \le B$	$A < B$ $m = A$
PDF	$f(x) = \frac{2(x-A)}{(B-A)^2}$ if $A \le x \le B$	$A < B$ $m = B$
CDF	$F(x) = \begin{cases} \frac{(x-A)^2}{(B-A)(m-A)} & \text{if } A \le x \le m \\ 1 - \frac{(B-x)^2}{(B-A)(B-m)} & \text{if } m \le x \le B \end{cases}$	
CDF	$F(x) = 1 - \left(\frac{B-x}{B-A}\right)^2$ if $A \le x \le B$	$m = A$
CDF	$F(x) = \left(\frac{x-A}{B-A}\right)^2$ if $A \le x \le B$	$m = B$
Mean	$\frac{A+B+m}{3}$	
Mean	$\frac{2A+B}{3}$	$m = A$
Mean	$\frac{A+2B}{3}$	$m = B$
Variance	$\frac{A^2+B^2+m^2-AB-m(A+B)}{18}$	
Variance	$\frac{(A-B)^2}{18}$	$m = A$ or $m = B$
Quantile	$A + \sqrt{U(B-A)(m-A)}$ if $0 \le U \le \frac{m-A}{B-A}$ $B - \sqrt{(1-U)(B-A)(B-m)}$ if $\frac{m-A}{B-A} \le U \le 1$	
Quantile	$B - (B-A)\sqrt{(1-U)}$	$m = A$
Quantile	$A + (B-A)\sqrt{U}$	$m = B$
Usage	`TRIANGULAR` *A mode B*	$A < B$, $A \le mode \le B$
Example	`ArmElas TRIANGULAR 1 4 6`	

Table 2.33: **Weibull Distribution**—Weibull(λ, k)

PDF	$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$	$x \geq 0$
CDF	$F(x) = 1 - e^{-(x/\lambda)^k}$	
Mean	$\lambda\Gamma\left(1 + \frac{1}{k}\right)$	
Variance	$\lambda^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right]$	
Quantile	$F^{-1}(U) = \lambda\left(-\ln(1-U)\right)^{1/k}$	$U \in [0, 1)$
Usage	`WEIBULL` k *lambda*	$k > 0$, $\lambda > 0$
Example	`productivity WEIBULL 0.2 0.4`	

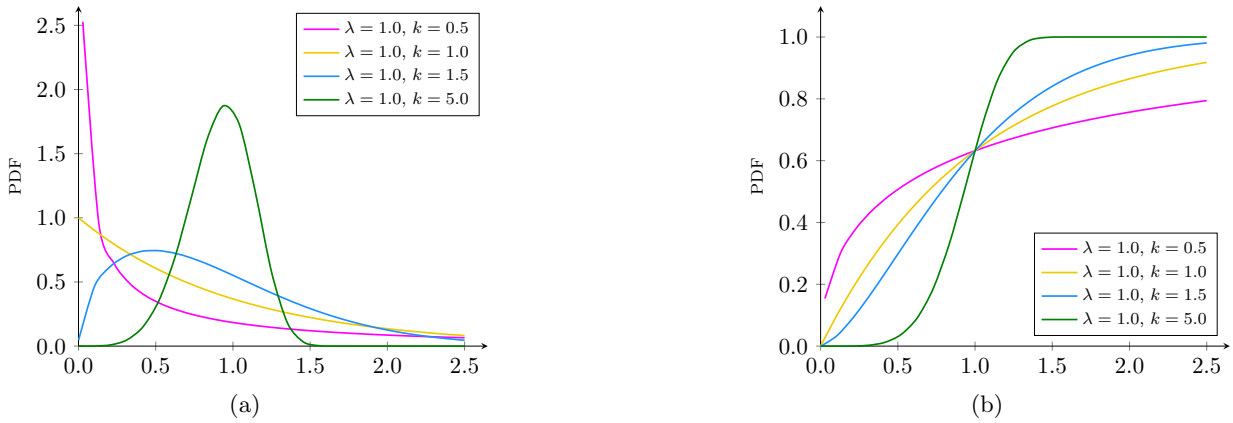


Figure 22: **Examples of the Weibull distribution (a) PDF and (b) CDF**

2.2.22 Weibull distribution

The Weibull distribution is a two-parameter distribution, where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. Among other uses, it has been used to characterize productivity, Sharif and Islam (1980). The key characteristics of the distribution are provided in Table 2.33. Figure 22 provides indicative profiles of the Weibull PDF and CDF for various values of the k and λ parameters.

2.2.23 Other continuous distributions

The original LHS package includes user-defined continuous distributions. These are largely a combination of uniform and log-uniform distributions where the user provides various ranges (e.g., quantiles) over which to sample. These are included in the LHS tool. Interested users should refer to the original LHS documentation (Swiler and Wyss, 2004) for full details and the appropriate syntax.

2.2.24 Poisson distribution

The Poisson distribution is a discrete distribution, typically the probability of having n discrete events occurring over a fixed time period, e.g., the Poisson distribution could characterize the probability of floods over a 100-year period. Table 2.34 describes the main characteristics of the Poisson distribution. Figure 23 depicts the PDF for two values of λ, for example, it shows the impact of doubling the frequency of floods occurring within a 100-year time period.

Table 2.34: **Poisson Distribution**—Poisson(λ)

PDF	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$x = 0, 1, 2, 3, \ldots$
Mean	λ	
Variance	λ	
Usage	`POISSON` *frequency*	*frequency* > 0
Example	`frqFlood POISSON 3`	

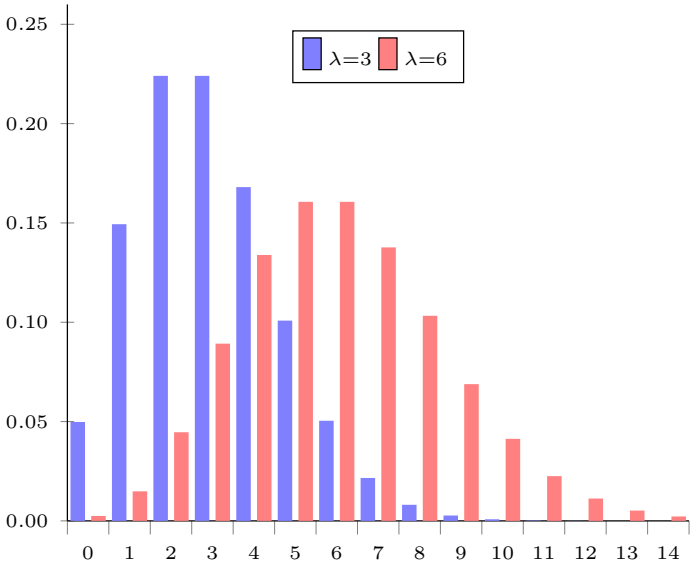


Figure 23: **Poisson with Varying λ**

Table 2.35: **Binomial Distribution**—B(n, p)

PDF	$f(x) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$	$k = 0, 1, 2, \ldots, n$
Mean	np	
Variance	$np(1-p)$	
Usage	`BINOMIAL` *probability* n	$0 < p < 1$, $n > 1$
Example	`survey BINOMIAL 0.52 50`	

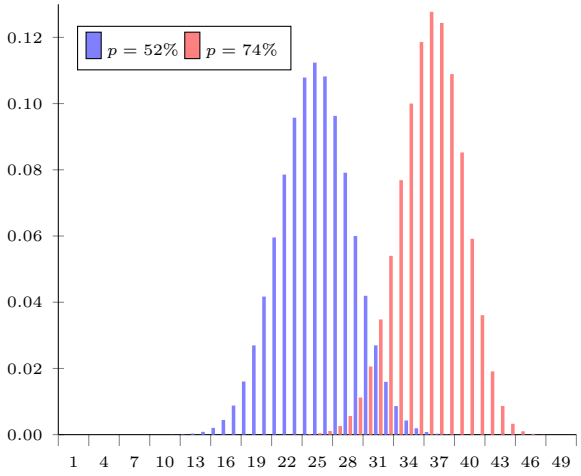


Figure 24: **Binomial PDF** ($n = 50$)

2.2.25 Binomial distribution

The Binomial distribution is a discrete distribution, which can be used to describe the probability of an event occurring over an experiment repeated n times. For example, one could use the Binomial distribution to calculate probabilities from a survey. If the average positive response to a binary survey is 52%, what is the probability out of a sample of 50 persons, that more than 30 respond would respond positively? The answer is:

$$\sum_{k=30}^{50} P[X = k] = \sum_{k=30}^{50} \binom{50}{k} p^k (1-p)^{50-k} = 16.1\%$$

Table 2.35 describes the main characteristics of the Binomial distribution. Figure 24 depicts the PDF for two values of the probability, for example, it shows the impact of doubling the belief in climate change from 52 to 74%.

2.2.26 Negative Binomial distribution

The Negative Binomial distribution is used to describe the number of tests needed to have n successes, where each success is assumed to occur with probability p. The main characteristics of the distribution are provided in Table 2.36.

Take as an example that the probability of civic disorder in any given year is 1.5%. In a 50-year time span, the probability of civic disorder is 53%:

$$\sum_{n=0}^{50} P[X = 1] = \sum_{n=0}^{50} \binom{n+1-1}{1} 0.015^1 (0.985)^n \sim 53\%$$

The probability rises to 78% for a doubling of the annual probability.

Table 2.36: **Negative Binomial Distribution**—NB(r, p)

PDF	$f(x) = \binom{n+r-1}{r} p^n (1-p)^r$	$n = 1, 2, \ldots$
Mean	$r(1-p)/p$	
Variance	$r(1-p)/^2$	
Usage	`NEGATIVE BINOMIAL` *probability* n	$0 < p < 1$, $n > 0$
Example	`survey NEGATIVE BINOMIAL 0.5 50`	

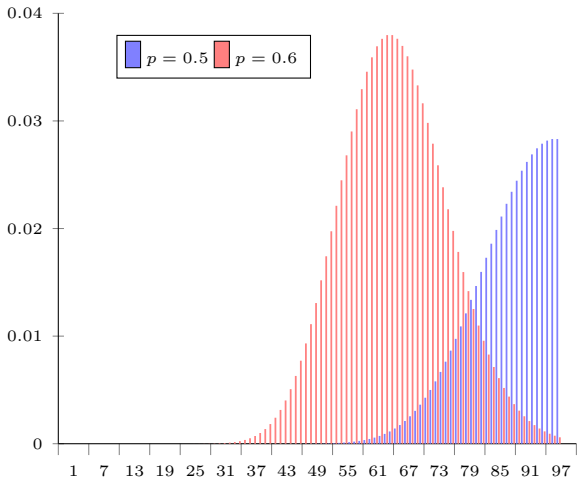


Figure 25: **Negative Binomial PDF** ($n = 100$)

Table 2.37: **Geometric Distribution**—Geometric(p)

PDF	$f(x) = (1-p)^x\, p$	$x = 0, 1, 2, 3, \ldots$
Mean	$\frac{1-p}{p}$	
Variance	$\frac{1-p}{p^2}$	
Usage	`GEOMETRIC` p	$0 < p \leq 1$
Example	`heat GEOMETRIC 0.4`	

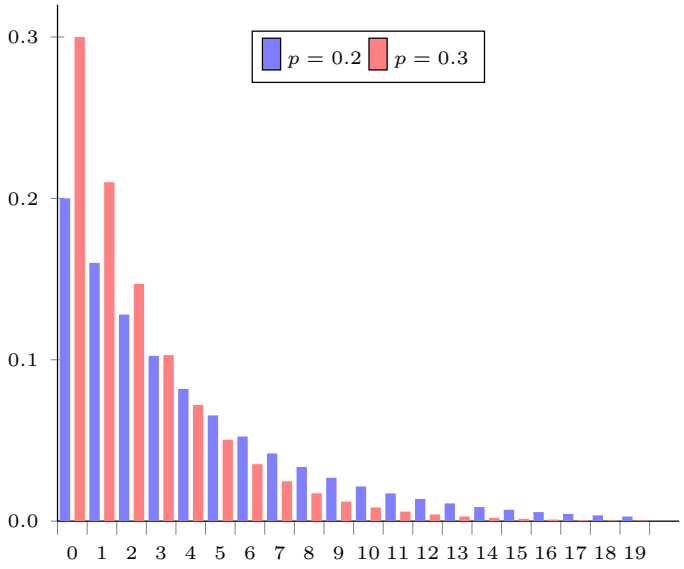


Figure 26: **Geometric PDF**

2.2.27 The Geometric distribution

A Geometric distribution is used to measure the number of successes before a failure occurs. This has been used, for example, to study the lengths of hot spells where 'success' is measured as having a day with extreme heat and 'failure' represents a break in the heat spell, Furrer et al. (2010).[9] The key characteristics are provided in Table 2.37. Figure 26 shows the PDF of a geometric distribution. Say the annual probability of a hurricane in a given location is 20%. The probability of going 5 years without a hurricane is only 6%. If the annual probability increases to 30%, the probability of going five years without a hurricane drops to 5%.

Figure 26 shows the PDF of a geometric distribution.

2.2.28 The Hypergeometric Distribution

A Hypergeometric distribution can characterize selection of n objects from a total population of N, with a binary characteristic, for example red or white, where the full population has N_1 of the first type (and thus $N_2 = N - N_1$ represents the number of the other type from the full population). The Hypergeometric distribution then represents the probability that the sample of n objects has exactly k objects of type 1. The main characteristics of the distribution are provided in Table 2.38.

Take an example of a population of 1500, of which 53%, or 795, are from the Blue party and the remainder are from the Red party. 100 are selected at random for a committee. What is the probability that the Red party has a majority? The answer is:

[9] N.B. The Furrer et al. (2010) uses a variant of the Geometric distribution, not the one used by the LHS utility.

Table 2.38: **Hypergeometric Distribution**—Hypergeometric(N, N_1, n)

PDF	$f(x) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}$	
Mean	$n\frac{N_1}{N}$	
Variance	$n\frac{N_1}{N}\frac{N-N_1}{N}\frac{N-n}{N-1}$	
Usage	`HYPERGEOMETRIC` N n N_1	$n < N_1 < N$
Example	`committee HYPERGEOMETRIC 100 10 40`	

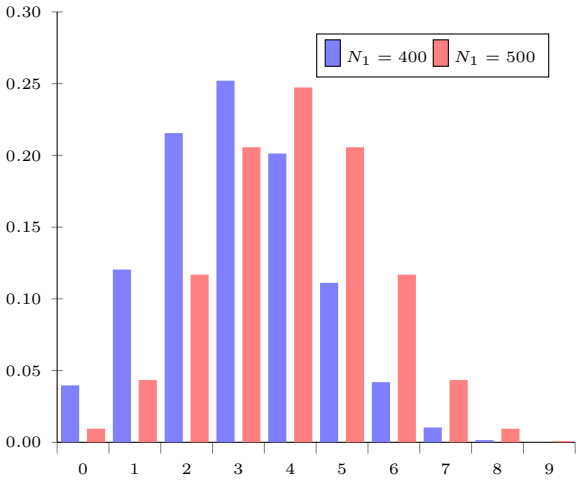


Figure 27: **Hypergeometric PDF** $(N = 100, n = 10)$

$$\sum_{k=51}^{100} \frac{\binom{705}{k}\binom{795}{100-k}}{\binom{1500}{100}} = 23.4\%$$

Figure 27 provides two distributions, both for a total population of 1000, of which 10 are selected. In one case, the 'Red' party represents 40% of the population, and in the alternate case, the 'Red' party represents half of the population. The chances of having a majority in the first case is around 5%. In the second case, the chance is around 17%.

2.2.29 Other discrete distributions

Similar to continuous distributions, the LHS utility allows for the possibility of user-defined discrete distributions. These are detailed in Swiler and Wyss (2004).

Bibliography

Dietz, S., J. Rising, T. Stoerk, and G. Wagner. 2021. "Economic impacts of tipping points in the climate system." *Proceedings of the National Academy of Sciences*, 118(34). doi:10.1073/pnas.2103081118.

Furrer, E.M., R.W. Katz, M.D. Walter, and R. Furrer. 2010. "Statistical modeling of hot spells and heat waves." *Climate Research*, 43(3): 191–205. doi:10.3354/cr00924.

Matsumoto, M., and T. Nishimura. 1998. "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator." *ACM Transactions on Modeling and Computer Simulation*, 8(1): 3–30. doi:10.1145/272991.272995.

Nishimura, T. 2000. "Tables of 64-Bit Mersenne Twisters." *ACM Transactions on Modeling and Computer Simulation*, 10(4): 348–357. doi:10.1145/369534.369540.

Sharif, M., and M. Islam. 1980. "The Weibull distribution as a general model for forecasting technological change." *Technological Forecasting and Social Change*, 18(3): 247–256. doi:10.1016/0040-1625(80)90026-8.

Swiler, L.P., and G.D. Wyss. 2004. "A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library Standalone Version." Sandia National Laboratories, Sandia Technical Report No. SAND2004-2439. doi:10.2172/919175.

van der Mensbrugghe, D. forthcoming. "A Latin Hypercube Sampling Utility: with an application to an Integrated Assessment Model." *The Journal of Global Economic Analysis*, pp. .

van der Mensbrugghe, D. 2023a. "The META 21 Integrated Assessment Model in GAMS and LHS Sampling." Global Trade Analysis Project (GTAP), Department of Agricultural Economics, Purdue University, West Lafayette, IN, GTAP Working Paper No. forthcoming.

van der Mensbrugghe, D. 2023b. "Using Python for Parallelization." Global Trade Analysis Project (GTAP), Department of Agricultural Economics, Purdue University, West Lafayette, IN, GTAP Working Paper No. forthcoming.

Appendix A

Generating the sample histograms with Python

If the user requests the sample histograms in the listing file (i.e., uses the option `RPTS HIST`), in the case of large sample sizes, the histograms will be meaningless given the limits on line length in the listing file. The new LHS utility will save the underlying data for the histogram(s) in a CSV cube with the name of the run, for example `META21LHS`, appended with the letters `Hist` and a file extension of `csv`. The following Python code shows how to read the cube and create a separate histogram figure for each of the sampled variables.

Listing A.1: **Python code to generate sample histograms**

```
import os
import pandas as pd
import matplotlib.pyplot as plt
import sys

#     User options

folder = './'                                   # Folder containing the input CSV cube
infile = 'xxxHist.csv'                        # File name of the CSV cube---replace XXX
outfolder = folder + 'Figures/'               # Code assumes the outfolder exists

os.chdir(folder)

#   Get the CSV cube
data = pd.read_csv(infile, index_col=None,dtype={'CellNo':'string'})

os.chdir(outfolder)

#   Get the number of replications and variables in the cube
nreps = data.Rep.unique()
nvars = data.VName.unique()

nFig = 0

#   Loop over all replications and variables
for r in nreps:
      for v in nvars:
            nFig = nFig + 1
            # Get the x- and y-values
            # !!!! For some reason these need to be converted to a list
            x = ((data[(data.Rep == r) & (data.VName == v)]).loc[:,['CellNo']]).astype(str)
            x = x['CellNo'].tolist()
            y = ((data[(data.Rep == r) & (data.VName == v)]).loc[:,['Value']]).astype(int)
            y = y['Value'].tolist()

            # Setup the plot
            fig, ax = plt.subplots(figsize=(6, 6), dpi=300)
            ax.spines['top'].set_visible(False)
```

```
            ax.spines['right'].set_visible(False)
            ax.spines['left'].set_visible(False)
            ax.spines['bottom'].set_color('#DDDDDD')
            ax.tick_params(bottom=False, left=False)
            ax.set_axisbelow(True)
            ax.yaxis.grid(True, color='#EEEEEE')
            ax.xaxis.grid(False)

            plt.xticks(fontsize=8, rotation=90)
            plt.yticks(fontsize=9, rotation=0)
            plt.title('Histogram for variable '+v, fontsize=10)

            #   Create the bar chart
            bars = ax.bar(x, y,width=0.6, color='dodgerblue')

            # Add the values to the top of the bars
            # Grab the color of the bars so we can make the
            # text the same color.
            bar_color = bars[0].get_facecolor()

            # Note, you might have to adjust this slightly (the 0.3)
            # with different data.
            for bar in bars:
                  ax.text(
                        bar.get_x() + bar.get_width() / 2,
                        bar.get_height() + 0.3,
                        round(bar.get_height(), 1),
                        horizontalalignment='center',
                        color=bar_color,
                        size=8,
                        weight='light'
                  )

            #   Display or save the figure
            fig.tight_layout()
            filename = v.partition('[')
            filename = filename[0] + '_Rep' + str(r) + '.png'
            fig.savefig(filename, dpi=300)
            plt.close(fig)
```