Handbook on probability distributions

R-forge distributions Core Team

University Year 2009-2010

Contents

Introduction **4**

I Discrete distributions **6**

1 Classic discrete distribution **7**

2 Not so-common discrete distribution **27**

II Continuous distributions **34**

3 Finite support distribution **35**

4 The Gaussian family **47**

5 Exponential distribution and its extensions **56**

6 Chi-squared's ditribution and related extensions **75**

7 Student and related distributions **84**

8 Pareto family **88**

9 Logistic distribution and related extensions **108**

10 Extrem Value Theory distributions **111**

III Multivariate and generalized distributions **116**

11 Generalization of common distributions **117**

12 Multivariate distributions **133**

13 Misc **135**

Conclusion **137**

Bibliography **137**

A Mathematical tools **141**

Introduction

This guide is intended to provide a quite exhaustive (at least as I can) view on probability distributions. It is constructed in chapters of distribution family with a section for each distribution. Each section focuses on the tryptic: definition - estimation - application.

Ultimate bibles for probability distributions are Wimmer & Altmann (1999) which lists 750 univariate discrete distributions and Johnson et al. (1994) which details continuous distributions.

In the appendix, we recall the basics of probability distributions as well as "common" mathematical functions, cf. section A.2. And for all distribution, we use the following notations

- X a random variable following a given distribution,
- x a realization of this random variable,
- f the density function (if it exists),
- F the (cumulative) distribution function,
- $P(X = k)$ the mass probability function in k,
- M the moment generating function (if it exists),
- G the probability generating function (if it exists),
- ϕ the characteristic function (if it exists),

Finally all graphics are done the open source statistical software R and its numerous packages available on the Comprehensive R Archive Network (CRAN*). See the CRAN task view† on probability distributions to know the package to use for a given "non standard" distribution, which is not in base R.

*http://cran.r-project.org

†http://cran.r-project.org/web/views/Distributions.html

Part I

Discrete distributions

Chapter 1

Classic discrete distribution

1.1 Discrete uniform distribution

1.1.1 Characterization

The discrete uniform distribution can be defined in terms of its elementary distribution (sometimes called mass probability function):

$$P(X = k) = \frac{1}{n},$$

where $k \in S = \{k_1, \ldots, k_n\}$ (a finite set of ordered values). Typically, the k_i's are consecutive positive integers.

Equivalenty, we have the following cumulative distribution function:

$$F(k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(k_i \leq k)},$$

where $\mathbb{1}$ is the indicator function.

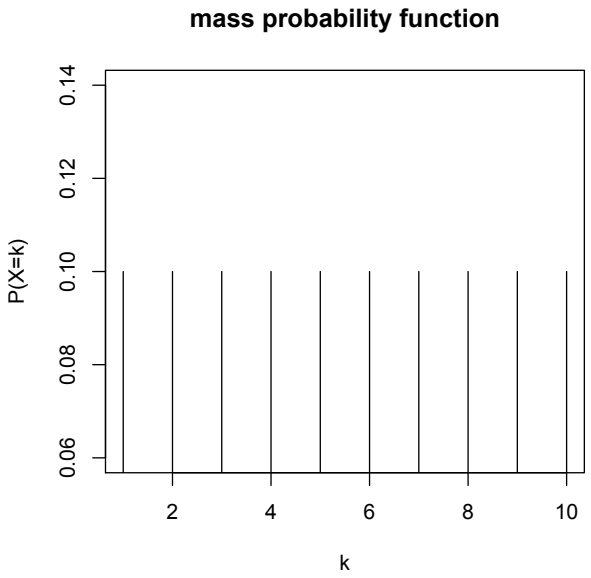


Figure 1.1: Mass probability function for discrete uniform distribution

Furthermore, the probability generating function is given by

$$G(t) \stackrel{\triangle}{=} E(t^X) = \frac{1}{n} \sum_{i=1}^{n} t^{k_i},$$

with the special cases where the k_i's are $\{1, \ldots, n\}$, we get

$$G(t) = z \frac{1 - z^n}{1 - z},$$

when $z \neq 1$.

Finally, the *moment* generating function is expressed as follows

$$M(t) \stackrel{\triangle}{=} E(t^X) = \frac{1}{n} \sum_{i=1}^{n} e^{tk_i},$$

with the special case $e^t \frac{1-e^{tn}}{1-e^t}$ when $S = \{1, \ldots, n\}$.

1.1.2 Properties

The expectation is $\overline{X}_n$, the empirical mean: $E(X) = \frac{1}{n} \sum_{i=1}^{n} k_i$. When $S = \{1, \ldots, n\}$, this is just $\frac{n+1}{2}$. The variance is given by $Var(X) = \frac{1}{n} \sum_{i=1}^{n} (k_i - E(X)^2$ which is $\frac{n^2-1}{12}$ for $S = \{1, \ldots, n\}$.

1.1.3 Estimation

Since there is no parameter to estimate, calibration is pretty easy. But we need to check that sample values are equiprobable.

1.1.4 Random generation

The algorithm is simply

- generate U from a uniform distribution,
- compute the generated index as $I = \lceil n \times U \rceil$,
- finally X is k_I.

where $\lceil . \rceil$ denotes the upper integer part of a number.

1.1.5 Applications

A typical application of the uniform discrete distribution is the statistic procedure called bootstrap or others resampling methods, where the previous algorithm is used.

1.2 Bernoulli/Binomial distribution

1.2.1 Characterization

Since the Bernoulli distribution is a special case of the binomial distribution, we start by explaining the binomial distribution. The mass probability distribution is

$$P(X = k) = C_n^k p^k (1-p)^{n-k},$$

where C_n^k is the combinatorial number $\frac{n!}{k!(n-k)!}$, $k \in \mathbb{N}$ and $0 < p < 1$ the 'success' probability. Let us notice that the cumulative distribution function has no particular expression. In the following, the binomial dsitribuion is denoted by $\mathcal{B}(n,p)$. A special case of the binomial dsitribution is the Bernoulli when $n = 1$. This formula explains the name of this distribution since elementary probabilities $P(X = k)$ are terms of the development of $(p + (1-p))^n$ according the Newton's binom formula.

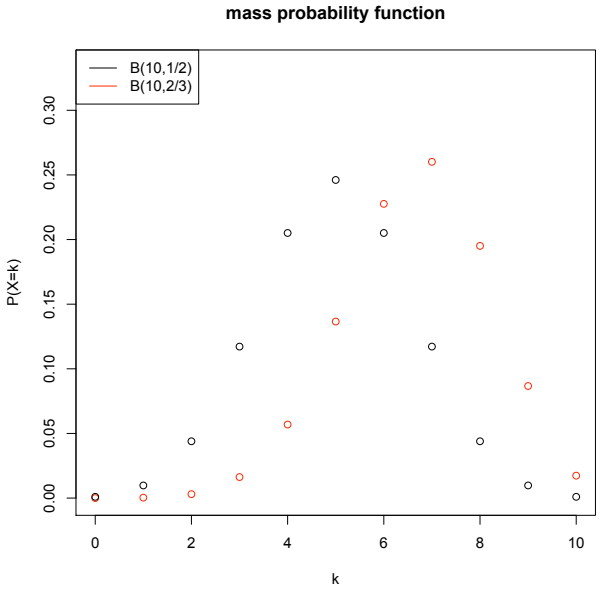


Figure 1.2: Mass probability function for binomial distributions

Another way to define the binomial distribution is to say that's the sum of n identically and independently Bernoulli distribution $\mathcal{B}(p)$. Demonstration can easily be done with probability generating function. The probability generating function is

$$G(t) = (1 - p + pz)^n,$$

while the moment generating function is

$$M(t) = (1 - p + pe^t)^n.$$

The binomial distribution assumes that the events are binary, mutually exclusive, independent and randomly selected.

1.2.2 Properties

The expectation of the binomial distribution is then $E(X) = np$ and its variance $Var(X) = np(1-p)$. A useful property is that a sum of binomial distributions is still binomial if success probabilities are the same, i.e. $\mathcal{B}(n_1,p) + \mathcal{B}(n_2,p) \overset{\mathcal{L}}{=} \mathcal{B}(n_1+n_2,p)$.

We have an asymptotic distribution for the binomial distribution. If $n \to +\infty$ and $p \to 0$ such that np tends to a constant, then $\mathcal{B}(n,p) \to \mathcal{P}(np)$.

1.2.3 Estimation

Bernoulli distribution

Let $(X_i)_{1\leq i\leq m}$ be an i.i.d. sample of binomial distributions $\mathcal{B}(n,p)$. If $n=1$ (i.e. Bernoulli distribution, we have

$$\hat{p}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$$

is the unbiased and efficient estimator of p with minimum variance. It is also the moment-based estimator.

There exists a confidence interval for the Bernoulli distribution using the Fischer-Snedecor distribution. We have

$$I_\alpha(p) = \left[\left(1+\frac{m-T+1}{T}f_{2(m-T+1),2T,\frac{\alpha}{2}}\right)^{-1}, \left(1+\frac{m-T}{T+1}f_{2(m-T),2(T+1),\frac{\alpha}{2}}\right)^{-1}\right],$$

where $T=\sum_{i=1}^{m} X_i$ and $f_{\nu_1,\nu_2,\alpha}$ the $1-\alpha$ quantile of the Fischer-Snedecor distribution with ν_1 and ν_2 degrees of freedom.

We can also use the central limit theorem to find an asymptotic confidence interval for p

$$I_\alpha(p) = \left[\hat{p}_m - \frac{u_\alpha}{\sqrt{n}}\sqrt{\hat{p}_m(1-\hat{p}_m)}, \hat{p}_m + \frac{u_\alpha}{\sqrt{n}}\sqrt{\hat{p}_m(1-\hat{p}_m)}\right],$$

where u_α is the $1-\alpha$ quantile of the standard normal distribution.

Binomial distribution

When n is not 1, there are two cases: either n is known with certainty or n is unknown. In the first case, the estimator of p is the same as the Bernoulli distribution. In the latter case, there are no closed form for the maximum likelihood estimator of n.

One way to solve this problem is to set $\hat{n}$ to the maximum number of 'success' at first. Then we compute the log likelihood for wide range of integers around the maximum and finally choose the likeliest value for n.

Method of moments for n and p is easily computable. Equalling the 2 first sample moments, we have the following solution

$$\begin{cases} \tilde{p} = 1 - \frac{S_m^2}{\bar{X}_m} \\ \tilde{n} = \frac{\bar{X}_m}{\tilde{p}} \end{cases},$$

with the constraint that $\tilde{n}\in\mathbb{N}$.

Exact confidence intervals cannot be found since estimators do not have analytical form. But we can use the normal approximation for $\hat{p}$ and $\hat{n}$.

1.2.4 Random generation

It is easy to simulate Bernoulli distribution with the following heuristic:

- generate U from a uniform distribution,
- compute X as 1 if $U \leq p$ and 0 otherwise.

The binomial distribution is obtained by summing n i.i.d. Bernoulli random variates.

1.2.5 Applications

The direct application of the binomial distribution is to know the probability of obtaining exactly n heads if a fair coin is flipped $m > n$ times. Hundreds of books deal with this application.

In medecine, the article Haddow et al. (1994) presents an application of the binomial distribution to test for a particular syndrome.

In life actuarial science, the binomial distribution is useful to model the death of an insured or the entry in invalidity/incapability of an insured.

1.3 Zero-truncated or zero-modified binomial distribution

1.3.1 Characterization

The zero-truncated version of the binomial distribution is defined as follows

$$P(X = k) = \frac{C_n^k p^k (1-p)^{n-k}}{1-(1-p)^n},$$

where $k \in \{1, \ldots, n\}$, n, p usual parameters. The distribution function does not have particular form. But the probability generating function and the moment generating function exist

$$G(t) = \frac{(1+p(z-1))^n - (1-p)^n}{1-(1-p)^n},$$

and

$$M(t) = \frac{(1+p(e^t-1))^n - (1-p)^n}{1-(1-p)^n}.$$

In the following distribution, we denote the zero-truncated version by $\mathcal{B}_0(n, p)$.

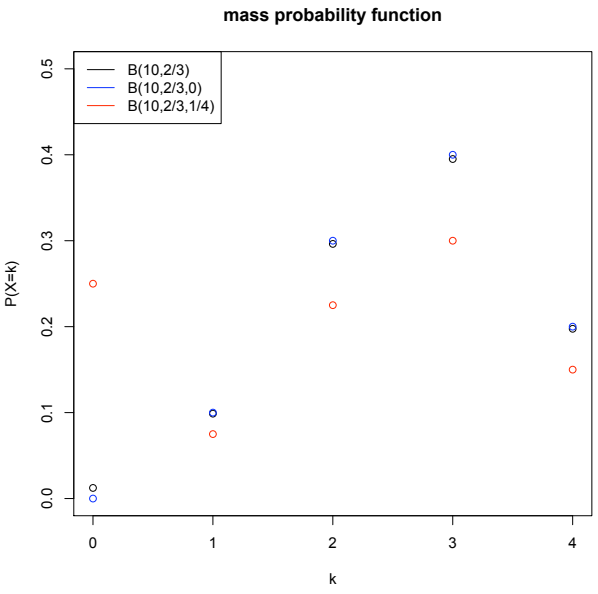


Figure 1.3: Mass probability function for zero-modified binomial distributions

For the zero-modified binomial distribution,
which of course generalizes the zero-truncated
version, we have the following elementary probabilities

$$P(X=k) = \begin{cases} \tilde{p} & \text{if } k=0 \\ KC_n^k p^k (1-p)^{n-k} & \text{otherwise} \end{cases},$$

where K is the constant $\frac{1-\tilde{p}}{1-(1-p)^n}$, $n, p, \tilde{p}$ are the parameters. In terms of probability/moment generating functions we have:

$$G(t) = \tilde{p} + K((1-p+pz)^n - (1-p)^n) \text{ and } M(t) = \tilde{p} + K((1-p+pe^t)^n - (1-p)^n).$$

The zero-modified binomial distribution is denoted by $\mathcal{B}(n, p, \tilde{p})$.

1.3.2 Properties

The expectation and the variance for the zero-truncated version is $E(X) = \frac{np}{1-(1-p)^n}$ and $Var(X) = \frac{np(1-p-(1-p+np)(1-p)^n)}{(1-(1-p)^n)^2}$. For the zero-modified version, we have $E(X) = Knp$ and $Var(X) = Knp(1-p)$.

1.3.3 Estimation

From Cacoullos & Charalambides (1975), we know there is no minimum variance unbiased estimator for p. NEED HELP for the MLE... NEED Thomas & Gart (1971)

Moment based estimators are numerically computable whatever we suppose n is known or unknown.

Confidence intervals can be obtained with bootstrap methods.

1.3.4 Random generation

The basic algorithm for the zero-truncated version $\mathcal{B}_0(n, p)$ is simply

- **do**; generate X binomially distributed $\mathcal{B}(n, p)$; **while** $X = 0$
- return X

In output, we have a random variate in $\{1, \ldots, n\}$.

The zero-modified version $\mathcal{B}(n, p, \tilde{p})$ is a little bit tricky. We need to use the following heuristic:

- generate U from an uniform distribution

- if $U < \tilde{p}$, then $X = 0$
- otherwise
  - **do**; generate X binomially distributed $\mathcal{B}(n, p)$; **while** $X = 0$
- return X

1.3.5 Applications

Human genetics???

1.4 Quasi-binomial distribution

1.4.1 Characterization

The quasi-binomial distribution is a "small" pertubation of the binomial distribution. The mass probability function is defined by

$$P(X = k) = C_n^k p(p + k\phi)^{k-1}(1 - p - k\phi)^{n-k},$$

where $k \in \{0, \ldots, n\}$, n, p usual parameters and $\phi \in]-\frac{p}{n}, \frac{1-p}{n}[$. Of course, we retrieve the binomial distribution with ϕ set to 0.

1.4.2 Properties

NEED REFERENCE

1.4.3 Estimation

NEED REFERENCE

1.4.4 Random generation

NEED REFERENCE

1.4.5 Applications

NEED REFERENCE

1.5 Poisson distribution

1.5.1 Characterization

The Poisson distribution is characterized by the following elementary probabilities

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

where $\lambda > 0$ is the shape parameter and $k \in \mathbb{N}$.

The cumulative distribution function has no particular form, but the probability generating function is given by

$$G(t) = e^{\lambda(t-1)},$$

and the moment generating function is

$$M(t) = e^{\lambda(e^t-1)}.$$

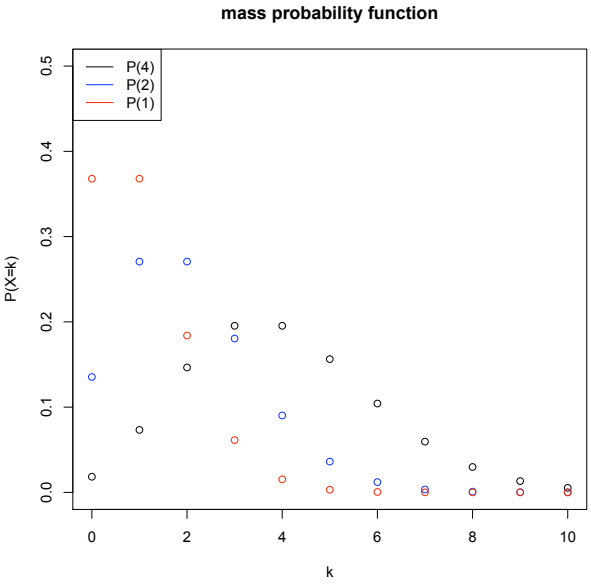


Figure 1.4: Mass probability function for Poisson distributions

Another way to characterize the Poisson distribution is to present the Poisson process (cf. Saporta (1990)). We consider independent and identically events occuring on a given period of time t. We assume that those events can not occur simultaneously and their probability to occur *only* depends on the observation period t. Let c be the average number of events per unit of time (c for cadency). We can prove that the number of events N occuring during the period $[0, t[$ is

$$P(N = n) = \frac{(ct)^k}{k!} e^{-ct},$$

since the interoccurence are i.i.d. positive random variables with the property of 'lack of memory'*.

1.5.2 Properties

The Poisson distribution has the 'interesting' but sometimes annoying property to have the same mean and variance. We have $E(X) = \lambda = Var(X)$.

The sum of two independent Poisson distributions $\mathcal{P}(\lambda)$ and $\mathcal{P}(\mu)$ (still) follows a Poisson distribution $\mathcal{P}(\lambda + \mu)$.

Let N follows a Poisson distribution $\mathcal{P}(\lambda)$. Knowing the value of $N = n$, let $(X_i)_{1 \leq i \leq n}$ be a sequence of i.i.d. Bernoulli variable $\mathcal{B}(q)$, then $\sum_{i=1}^{n} X_i$ follows a Poisson distribution $\mathcal{P}(\lambda q)$.

*i.e. interoccurence are exponentially distributed, cf. the exponential distribution.

1.5.3 Estimation

The estimator maximum likelihood estimator of λ is $\hat{\lambda} = \overline{X}_n$ for a sample $(X_i)_i$. It is also the moment based estimator, an unbiased estimator λ and an efficient estimator.

From the central limit theorem, we have asymptotic confidence intervals

$$I_\alpha(\lambda) = \left[\hat{\lambda}_n - \frac{u_\alpha}{\sqrt{n}}\sqrt{\hat{\lambda}_m}, \hat{\lambda}_n + \frac{u_\alpha}{\sqrt{n}}\sqrt{\hat{\lambda}_m}\right],$$

where u_α is the $1-\alpha$ quantile of the standard normal distribution.

1.5.4 Random generation

A basic way to generate Poisson random variate is the following:

- initialize variable n to 0, l to $e^{-\lambda}$ and P to 1,
- **do**
  - generate U from a uniform distribution,
  - $P = P \times U$,
  - $n = n + 1$,

  while $P \geq l$,
- return $n - 1$.

See Knuth (2002) for details.

TOIMPROVE

Ahrens, J. H. and Dieter, U. (1982). Computer generation of Poisson deviates from modified normal distributions. ACM Transactions on Mathematical Software, 8, 163?179.

1.5.5 Applications

TODO

1.6 Zero-truncated or zero-modified Poisson distribution

1.6.1 Characterization

The zero-truncated version of the Poisson distribution is defined the zero-truncated binomial distribution for the Poisson distribution. The elementary probabilities is defined as

$$P(X=k)=\frac{\lambda^k}{k!}\frac{1}{(e^\lambda-1)},$$

where $k \in \mathbb{N}^*$. We can define probability/moment generating functions for the zero-truncated Poisson distribution $\mathcal{P}_0(\lambda)$:

$$G(t)=\frac{e^{\lambda t}-1}{e^\lambda-1} \quad \text{and} \quad M(t)=\frac{e^{\lambda e^t}-1}{e^\lambda-1}.$$

The zero-modified version of the Poisson distribution (obviously) generalized the zero-truncated version. We have the following mass probability function

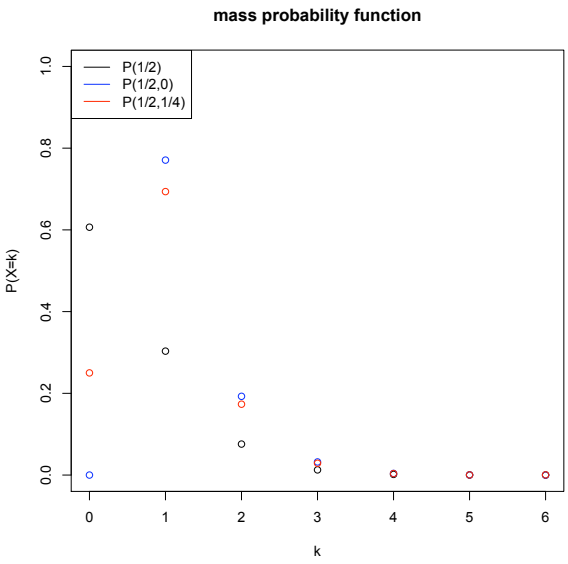


Figure 1.5: Mass probability function for zero-modified Poisson distributions

$$P(X=k)=\begin{cases} p & \text{if } k=0 \\ K\frac{\lambda^k}{k!}e^{-\lambda} & \text{otherwise} \end{cases},$$

where K is the constant $\frac{1-p}{1-e^{-\lambda}}$. The "generating functions" for the zero-modified Poisson distribution $\mathcal{P}(\lambda,p)$ are

$$G(t)=p+K(e^{\lambda t}-1) \quad \text{and} \quad M(t)=p+K(e^{\lambda e^t}-1).$$

1.6.2 Properties

The expectation of the zero-truncated Poisson distribution is $E(X)=\frac{\lambda}{1-e^{-\lambda}}$ and $K\lambda$ for the zero-modified version. While the variance are respectively $Var(X)=\frac{\lambda}{(1-e^{-\lambda})^2}$ and $K\lambda+(K-K^2)\lambda^2$.

1.6.3 Estimation

Zero-truncated Poisson distribution

Let $(X_i)_i$ be i.i.d. sample of truncated Poisson random variables. Estimators of λ for the zero-truncated Poisson distribution are studied in Tate & Goen (1958). Here is the list of possible estimators for λ:

- $\tilde{\lambda}=\frac{T}{n}(1-\frac{{}_2S_{n-1}^{t-1}}{{}_2S_n^t})$ is the minimum variance unbiased estimator,
- $\lambda^*=\frac{T}{n}(1-\frac{N_1}{T})$ is the Plackett's estimator,
- $\hat{\lambda}$, the solution of equation $\frac{T}{n}=\frac{\lambda}{1-e^{-\lambda}}$, is the maximum likelihood estimator,

where $T = \sum_{i=1}^{n} X_i$, ${}_2S_n^k$ denotes the Stirling number of the second kind and N_1 the number of observations equal to 1. Stirling numbers are costly do compute, see Tate & Goen (1958) for approximate of theses numbers.

Zero-modified Poisson distribution

NEED REFERENCE

1.6.4 Random generation

The basic algorithm for the zero-truncated version $\mathcal{P}_0(\lambda)$ is simply

- **do**; generate X Poisson distributed $\mathcal{P}(\lambda)$; **while** $X = 0$
- return X

In output, we have a random variate in $\mathbb{N}^*$.

The zero-modified version $\mathcal{P}(\lambda, p)$ is a little bit tricky. We need to use the following heuristic:

- generate U from an uniform distribution
- if $U < p$, then $X = 0$
- otherwise
  - **do**; generate X Poisson distributed $\mathcal{P}(\lambda)$; **while** $X = 0$
- return X

1.6.5 Applications

NEED REFERENCE

1.7 Quasi-Poisson distribution

NEED FOLLOWING REFERENCE

Biom J. 2005 Apr;47(2):219-29. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. Joe H, Zhu R.

Ecology. 2007 Nov;88(11):2766-72. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ver Hoef JM, Boveng PL.

1.7.1 Characterization

TODO

1.7.2 Properties

TODO

1.7.3 Estimation

TODO

1.7.4 Random generation

TODO

1.7.5 Applications

1.8 Geometric distribution

1.8.1 Characterization

The geometric distribution represents the first outcome of a particular event (with the probability q to raise) in a serie of i.i.d. events. The mass probability function is

$$P(X = k) = q(1 - q)^k,$$

where $k \in \mathbb{N}$ and $0 < q \leq 1$. In terms of cumulative distribution function, it is the same as

$$F(k) = 1 - (1 - q)^{k+1}.$$

The whole question is wether this outcome could be null or at least one (event). If we consider the distribution to be valued in $\mathbb{N}^*$, please see the truncated geometric distribution.

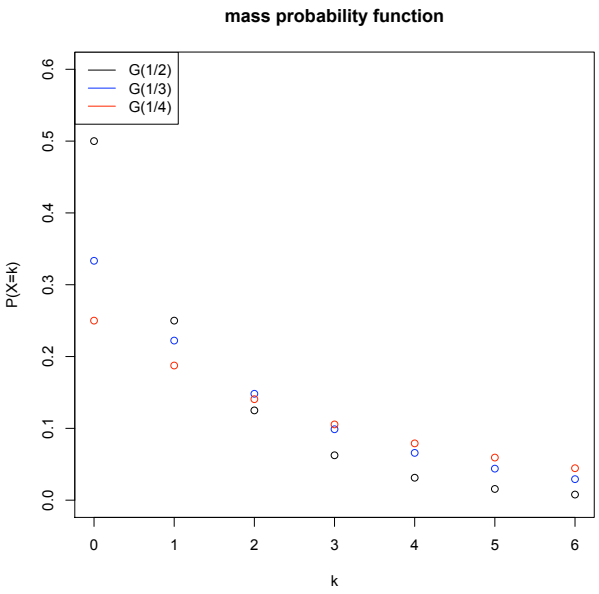


Figure 1.6: Mass probability function for Geometric distributions

The probability generating function of the geometric $\mathcal{G}(q)$ is

$$G(t) = \frac{q}{1-(1-q)t},$$

and its moment generating function

$$M(t) = \frac{q}{1-(1-q)e^t}.$$

1.8.2 Properties

The expecation of a geometric distribution is simply $E(X) = \frac{1-q}{q}$ and its variance $Var(X) = \frac{1-q}{q^2}$.

The sum of n i.i.d. geometric $\mathcal{G}(q)$ random variables follows a negative binomial distribution $\mathcal{NB}(n, q)$.

The minimum of n independent geometric $\mathcal{G}(q_i)$ random variables follows a geometric distribution $\mathcal{G}(q_.)$ with $q_. = 1 - \prod_{i=1}^{n}(1-q_i)$.

The geometric distribution is the discrete analogue of the exponential distribution thus it is memoryless.

1.8.3 Estimation

The maximum likelihood estimator of q is $\hat{q} = \frac{1}{1+\bar{X}_n}$, which is also the moment based estimator.

NEED REFERENCE

1.8.4 Random generation

A basic algorithm is to use i.i.d. Bernoulli variables as follows

- initialize X to 0 and generate U from an uniform distribution,
- **while** $U > p$ **do** ; generate U from an uniform distribution; $X = X + 1$;
- return X.

TOIMPROVE WITH Devroye, L. (1986) Non-Uniform Random Variate Generation. Springer-Verlag, New York. Page 480.

1.8.5 Applications

NEED MORE REFERENCE THAN Mačutek (2008)

1.9 Zero-truncated or zero-modified geometric distribution

1.9.1 Characterization

The zero-truncated version of the geometric distribution is defined as

$$P(X = k) = p(1-p)^{k-1},$$

where $n \in \mathbb{N}^+$. Obviously, the distribution takes values in $\{1, \dots, n, \dots\}$. Its distribution function is

$$F(k) = 1 - (1-p)^k.$$

Finally the probability/moment generating functions are

$$G(t) = \frac{pt}{1-(1-p)t}, \quad \text{and} \quad M(t) = \frac{pe^t}{1-(1-p)e^t}.$$

In the following, it is denoted by $\mathcal{G}_0(p)$.

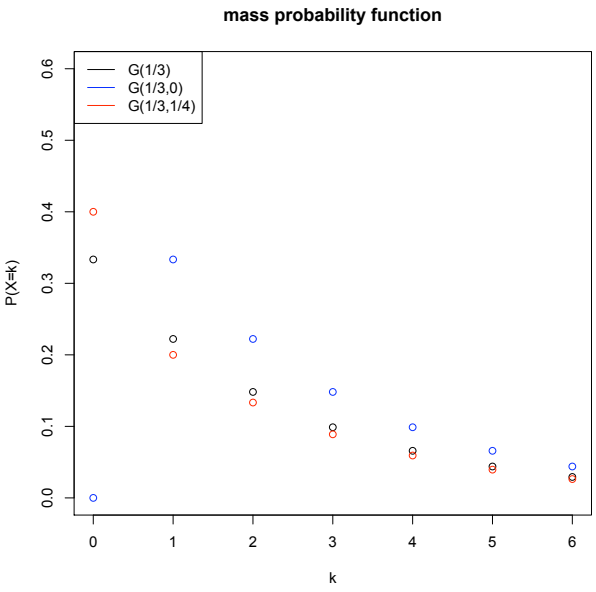


Figure 1.7: Mass probability function for zero-modified geometric distributions

The zero-modified version of the geometric distribution is characterized as follows

$$P(X = k) = \begin{cases} p & \text{if } k = 0 \\ Kq(1-q)^k & \text{otherwise} \end{cases},$$

where the constant K is $\frac{1-p}{1-q}$ and $k \in \mathbb{N}$. Of course special cases of the zero modified version of the geometric $\mathcal{G}(q,p)$ are the zero-truncated version with $p = 0$ and $q = p$ and the classic geometric distribution with $p = q$. The distribution function is expressed as follows

$$F(x) = p + K(1 - (1-p)^k),$$

where $k \geq 0$. The probability/moment generating functions are

$$G(t) = p + K\left(\frac{q}{1-(1-q)t} - q\right) \quad \text{and} \quad M(t) = p + K\left(\frac{q}{1-(1-q)e^t} - q\right).$$

1.9.2 Properties

The expectation of the geometric $\mathcal{G}_0(p)$ distribution is $E(X) = \frac{1}{p}$ and its variance $Var(X) = \frac{1-p}{p^2}$.

For the zero-modified geometric distribution $\mathcal{G}(q,p)$, we have $E(X) = K\frac{1-q}{q}$ and $Var(X) = K\frac{1-q}{q^2}$.

1.9.3 Estimation

Zero-truncated geometric distribution

According to Cacoullos & Charalambides (1975), the (unique) minimim variance unbiased estimator of q for the zero-truncated geometric distribution is

$$\tilde{q} = t\frac{\tilde{S}_n^{t-1}}{\tilde{S}_n^t},$$

where t denotes the sum $\sum_{i=1}^n X_i$, $\tilde{S}_n^t$ is defined by $\frac{1}{n!}\sum_{k=1}^n (-1)^{n-k} C_n^k (k+t-1)_t$*. The maximum likelihood estimator of q is given by

$$\hat{q} = \frac{1}{\bar{X}_n},$$

which is also the moment based estimator. By the uniqueness of the unbiased estimator, $\hat{q}$ is a biased estimator.

Zero-modified geometric distribution

Moment based estimators for the zero-modified geometric distribution $\mathcal{G}(p,q)$ are given by $\hat{q} = \frac{\bar{X}_n}{S_n^2}$ and $\hat{p} = 1 - \frac{(\bar{X}_n)^2}{S_n^2}$.

NEED REFERENCE

1.9.4 Random generation

For the zero-truncated geometric distribution, a basic algorithm is to use i.i.d. Bernoulli variables as follows

- initialize X to 1 and generate U from an uniform distribution,
- **while** $U > q$ **do** ; generate U from an uniform distribution; $X = X + 1$;
- return X.

While for the zero-modified geometric distribution, it is a little bit tricky

- generate U from an uniform distribution
- if $U < p$, then $X = 0$
- otherwise

*where C_n^k's are the binomial coefficient and $(n)_m$ is the falling factorial.

- initialize X to 1 and generate U from an uniform distribution
- **while** $U > q$ **do** ; generate U from an uniform distribution; $X = X + 1$;

- return X

1.9.5 Applications

NEED REFERENCE

1.10 Negative binomial distribution

1.10.1 Characterization

1.10.2 Characterization

The negative binomial distribution can be characterized by the following mass probability function

$$P(X = k) = C^k_{m+k-1} p^m (1-p)^k,$$

where $k \in \mathbb{N}$, C^k_{m+k-1}'s are combinatorial numbers and parameters m, p are constraint by $0 < p < 1$ and $m \in \mathbb{N}^*$. However a second parametrization of the negative binomial distribution is

$$P(X = k) = \frac{\Gamma(r+k)}{\Gamma(r)k!} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k,$$

where $k \in \mathbb{N}$ and $r, \beta > 0$. We can retrieve the first parametrization $\mathcal{NB}(m, p)$ from the second parametrization $\mathcal{NB}(r, \beta)$ with

$$\begin{cases} \frac{1}{1+\beta} = p \\ r = m \end{cases}$$

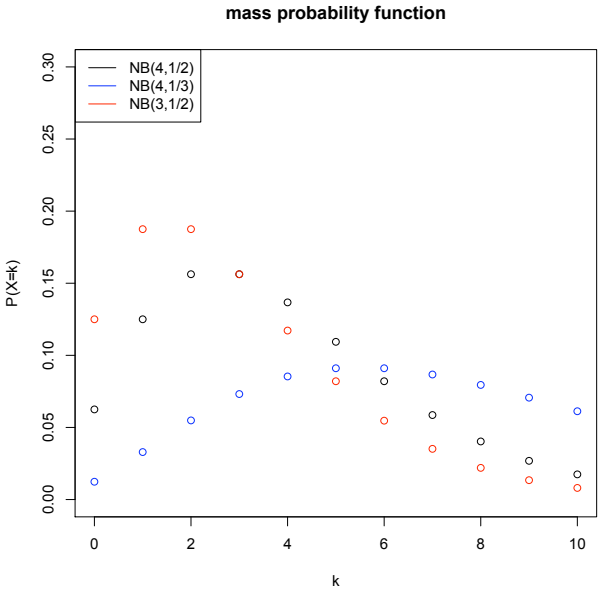


Figure 1.8: Mass probability function for negative binomial distributions

The probability generating functions for these two parametrizations are

$$G(t) = \left(\frac{p}{1-(1-p)t}\right)^m \quad \text{and} \quad G(t) = \left(\frac{1}{1-\beta(t-1)}\right)^r,$$

and their moment generating functions are

$$M(t) = \left(\frac{p}{1-(1-p)e^t}\right)^m \quad \text{and} \quad M(t) = \left(\frac{1}{1-\beta(e^t-1)}\right)^r.$$

One may wonder why there are two parametrization for one distribution. Actually, the first parametrization $\mathcal{NB}(m,p)$ has a meaningful construction: it is the sum of m i.i.d. geometric $\mathcal{G}(p)$ random variables. So it is also a way to characterize a negative binomial distribution. The name comes from the fact that the mass probability function can be rewritten as

$$P(X=k) = C_{m+k-1}^{k} \left(\frac{1-p}{p}\right)^{k} \left(\frac{1}{p}\right)^{-m-k},$$

which yields to

$$P(X=k) = C_{m+k-1}^{k} P^{k} Q^{-m-k}.$$

This is the general term of the development of $(P-Q)^{-m}$.

1.10.3 Properties

The expectation of negative binomial $\mathcal{NB}(m,p)$ (or $\mathcal{NB}(m,p)$) is $E(X) = \frac{m(1-p)}{p}$ or $(r\beta)$, while its variance is $Var(X) = \frac{m(1-p)}{p^2}$ or $(r\beta(1+\beta))$.

Let N be Poisson distributed $\mathcal{P}(\lambda\Theta)$ knowing that $\Theta = \theta$ where Θ is gamma distributed $\mathcal{G}(a,a)$. Then we have N is negative binomial distributed $\mathcal{BN}(a, \frac{\lambda}{a})$.

1.10.4 Estimation

Moment based estimators are given by $\hat{\beta} = \frac{S_n^2}{\bar{X}_n} - 1$ and $\hat{r} = \frac{\bar{X}_n}{\hat{\beta}}$.

NEED REFERENCE

1.10.5 Random generation

The algorithm to simulate a negative binomial distribution $\mathcal{NB}(m,p)$ is simply to generate m random variables geometrically distributed and to sum them.

NEED REFERENCE

1.10.6 Applications

From Simon (1962), here are some applications of the negative binomial distribution

- number of bacterial colonies per microscopic field,
- quality control problem,
- claim frequency in non life insurance.

1.11 Zero-truncated or zero-modified negative binomial distribution

1.11.1 Characterization

The zero-truncated negative binomial distribution is characterized by

$$P(X=k)=\frac{\Gamma(r+k)}{\Gamma(r)k!((r+\beta)^r-1)}(\frac{\beta}{1+\beta})^k,$$

where $k\in\mathbb{N}^*$, r,β usual parameters. In terms of probability generating function, we have

$$G(t)=\frac{(1-\beta(t-1))^r-(1+\beta)^{-r}}{1-(r+\beta)^r}.$$

The zero-modified version is defined as follows

$$P(X=k)=\begin{cases} p & \text{if } k=0 \\ K\frac{\Gamma(r+k)}{\Gamma(r)k!}(\frac{1}{1+\beta})^r(\frac{\beta}{1+\beta})^k & \text{otherwise} \end{cases},$$

where K is defined as $\frac{1-p}{1-(\frac{1}{1+\beta})^r}$, r,β usual parameters and p the new parameter. The probability generating function is given by

$$G(t)=\left((\frac{1}{1-\beta(t-1)})^r-(\frac{1}{1+\beta})^r\right),$$

and

$$M(t)=\left((\frac{1}{1-\beta(e^t-1)})^r-(\frac{1}{1+\beta})^r\right)$$

for the moment generating function.

1.11.2 Properties

Expectations for these two distribution are $E(X)=\frac{r\beta}{1-(r+\beta)^r}$ and $Kr\beta$ respectively for the zero-truncated and the zero-modified versions. Variances are $Var(X)=\frac{r\beta(1+\beta-(1+\beta+r\beta)(1+\beta)^{-r})}{(1-(r+\beta)^r)^2}$ and $Kr\beta(1+\beta)+(K-K^2)E^2[X]$.

1.11.3 Estimation

According to Cacoullos & Charalambides (1975), the (unique) minimim variance unbiased estimator of p for the zero-truncated geometric distribution is

$$\tilde{p}=t\frac{\tilde{S}^{t-1}_{r,n}}{\tilde{S}^t_n},$$

where t denotes the sum $\sum_{i=1}^{n} X_i$, $\tilde{S}_n^t$ is defined by $\frac{1}{n!}\sum_{k=1}^{n}(-1)^{n-k}C_n^k(k+t-1)_t$*. The maximum likelihood estimator of q is given by

$$\hat{q} = \frac{1}{\bar{X}_n},$$

which is also the moment based estimator. By the uniqueness of the unbiased estimator, $\hat{q}$ is a biased estimator.

1.11.4 Random generation

1.11.5 Applications

1.12 Pascal distribution

1.12.1 Characterization

The negative binomial distribution can be constructed by summing m geometric distributed variables $\mathcal{G}(p)$. The Pascal distribution is got from summing n geometrically distributed $\mathcal{G}_0(p)$ variables. Thus possible values of the Pascal distribution are in $\{n, n+1, \dots\}$. The mass probability function is defined as

$$P(X = k) = C_{k-1}^{n-1} p^n (1-p)^{k-n},$$

where $k \in \{n, n+1, \dots\}$, $n \in \mathbb{N}^*$ and $0 < p < 1$. The probability/moment generating functions are

$$G(t) = \left(\frac{pt}{1-(1-p)t}\right)^n \quad \text{and} \quad M(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^n.$$

1.12.2 Properties

For the Pascal distribution $\mathcal{P}a(n,p)$, we have $E(X) = \frac{n}{p}$ and $Var(X) = \frac{n(1-p)}{p^2}$. The link between Pascal distribution $\mathcal{P}a(n,p)$ and the negative binomial distribution $\mathcal{BN}(n,p)$ is to substract the constant n, i.e. if $X \sim \mathcal{P}a(n,p)$ then $X - n \sim \mathcal{BN}(n,p)$.

*where C_n^k's are the binomial coefficient and $(n)_m$ is the increasing factorial.

1.12.3 Estimation

1.12.4 Random generation

1.12.5 Applications

1.13 Hypergeometric distribution

1.13.1 Characterization

The hypergeometric distribution is characterized by the following elementary probabilities

$$P(X = k) = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n},$$

where $N \in \mathbb{N}^+$, $(m, n) \in \{1, \dots, N\}^2$ and $k \in \{0, \dots, \min(m, n)\}$.

It can also be defined though its probability generating function or moment generating function:

$$G(t) = \frac{C_{N-m}^n \, {}_2F_1(-n, -m; N - m - n + 1; t)}{C_N^n} \quad \text{and} \quad M(t) = \frac{C_{N-m}^n \, {}_2F_1(-n, -m; N - m - n + 1; e^t)}{C_N^n},$$

where ${}_2F_1$ is the hypergeometric function of second kind.

1.13.2 Properties

The expectation of an hypergeometric distribution is $E(X) = \frac{nm}{N}$ and $Var(X) = \frac{nm(N-n)(N-m)}{N^2(N-1)}$.

We have the following asymptotic result: $\mathcal{H}(N, n, m) \mapsto \mathcal{B}(n, \frac{m}{N})$ when N and m are large such that $\frac{m}{N} \underset{N \to +\infty}{\longrightarrow} 0 < p < 1$.

1.13.3 Estimation

1.13.4 Random generation

1.13.5 Applications

Let N be the number of individuals in a given population. In this population, m has a particular property, hence a proportion of $\frac{m}{N}$. If we draw n individuals among this population, the random variable associated with the number of people having the desired property follows a hypergeometric distribution $\mathcal{H}(N, n, m)$. The ratio $\frac{n}{N}$ is called the survey rate.

Chapter 2

Not so-common discrete distribution

2.1 Conway-Maxwell-Poisson distribution

2.1.1 Characterization

TODO

2.1.2 Properties

TODO

2.1.3 Estimation

TODO

2.1.4 Random generation

TODO

2.1.5 Applications

2.2 Delaporte distribution

2.2.1 Characterization

TODO

2.2.2 Properties

TODO

2.2.3 Estimation

TODO

2.2.4 Random generation

TODO

2.2.5 Applications

2.3 Engen distribution

2.3.1 Characterization

TODO

2.3.2 Properties

TODO

2.3.3 Estimation

TODO

2.3.4 Random generation

TODO

2.3.5 Applications

2.4 Logaritmic distribution

2.4.1 Characterization

TODO

2.4.2 Properties

TODO

2.4.3 Estimation

TODO

2.4.4 Random generation

TODO

2.4.5 Applications

2.5 Sichel distribution

2.5.1 Characterization

TODO

2.5.2 Properties

TODO

2.5.3 Estimation

TODO

2.5.4 Random generation

TODO

2.5.5 Applications

2.6 Zipf distribution

The name "Zipf distribution" comes from George Zipf's work on the discretized version of the Pareto distribution, cf. Arnold (1983).

2.6.1 Characterization

See Arnold(83) for relationship with Pareto's distribution.

2.6.2 Properties

TODO

2.6.3 Estimation

TODO

2.6.4 Random generation

TODO

2.6.5 Applications

2.7 The generalized Zipf distribution

2.7.1 Characterization

TODO

2.7.2 Properties

TODO

2.7.3 Estimation

TODO

2.7.4 Random generation

TODO

2.7.5 Applications

2.8 Rademacher distribution

2.8.1 Characterization

TODO

2.8.2 Properties

TODO

2.8.3 Estimation

TODO

2.8.4 Random generation

TODO

2.8.5 Applications

2.9 Skellam distribution

2.9.1 Characterization

TODO

2.9.2 Properties

TODO

2.9.3 Estimation

TODO

2.9.4 Random generation

TODO

2.9.5 Applications

2.10 Yule distribution

2.10.1 Characterization

TODO

2.10.2 Properties

TODO

2.10.3 Estimation

TODO

2.10.4 Random generation

TODO

2.10.5 Applications

2.11 Zeta distribution

2.11.1 Characterization

TODO

2.11.2 Properties

TODO

2.11.3 Estimation

TODO

2.11.4 Random generation

TODO

2.11.5 Applications

Part II

Continuous distributions

Chapter 3

Finite support distribution

3.1 Uniform distribution

3.1.1 Characterization

The uniform distribution is the most intuitive distribution, its density function is

$$f(x) = \frac{1}{b-a},$$

where $x \in [a, b]$ and $a < b \in \mathbb{R}$. So the uniform $\mathcal{U}(a, b)$ is only valued in $[a, b]$. From this, we can derive the following distribution function

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

Another way to define the uniform distribution is to use the moment generating function

$$M(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

whereas its characteristic function is

$$\phi(t) = \frac{e^{ibt} - e^{iat}}{i(b-a)t}.$$

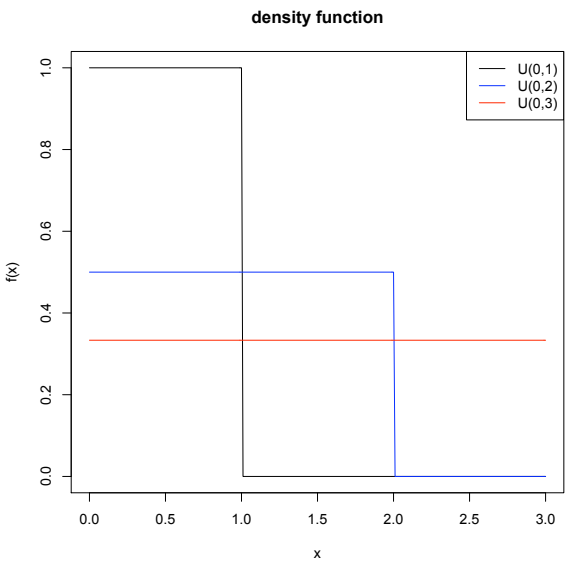


Figure 3.1: Density function for uniform distribution

3.1.2 Properties

The expectation of a uniform distribution is $E(X) = \frac{a+b}{2}$ and its variance $Var(X) = \frac{(b-a)^2}{12}$.

If U is uniformally distributed $\mathcal{U}(0,1)$, then $(b-a)\times U+a$ follows a uniform distribution $\mathcal{U}(a,b)$.

The sum of two uniform distribution does not follow a uniform distribution but a triangle distribution.

The order statistic $X_{k:n}$ of a sample of n i.i.d. uniform $\mathcal{U}(0,1)$ random variable is beta distributed $\mathcal{B}eta(k, n-k+1)$.

Last but not least property is that for all random variables Y having a distribution function F_Y, the random variable $F_Y(Y)$ follows a uniform distribution $\mathcal{U}(0,1)$. Equivalently, we get that the random variable $F_Y^{-1}(U)$ has the same distribution as Y where $U \sim \mathcal{U}(0,1)$ and F_Y^{-1} is the generalized inverse distribution function. Thus, we can generate any random variables having a distribution from the a uniform variate. This methods is called the inverse function method.

3.1.3 Estimation

For a sample $(X_i)_i$ of i.i.d. uniform variate, maximum likelihood estimators for a and b are respectively $X_{1:n}$ and $X_{n:n}$, where $X_{i:n}$ denotes the order statistics. But they are biased so we can use the following unbiased estimators

$$\hat{a} = \frac{n}{n^2-1}X_{1:n} + \frac{1}{1-n^2}X_{n:n} \quad \text{and} \quad \hat{b} = \frac{1}{1-n^2}X_{1:n} + \frac{n}{n^2-1}X_{n:n}.$$

Finally the method of moments gives the following estimators

$$\tilde{a} = \bar{X}_n - \sqrt{3S_n^2} \quad \text{and} \quad \tilde{b} = \bar{X}_n + \sqrt{3S_n^2}.$$

3.1.4 Random number generation

Since this is the core distribution, the distribution can not be generated from another distribution. In our modern computers, we use deterministic algorithms to generate uniform variate initialized with the machine time. Generally, Mersenne-Twister algorithm (or its extensions) from Matsumoto & Nishimura (1998) is implemented, cf. Dutang (2008) for an overview of random number generation.

3.1.5 Applications

The main application is sampling from an uniform distribution by the inverse function method.

3.2 Triangular distribution

3.2.1 Characterization

The triangular distribution has the following density

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{if } c \leq x \leq b \end{cases},$$

where $x \in [a, b]$, $a \in \mathbb{R}$, $a < b$ and $a \leq c \leq b$. The associated distribution function is

$$F(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & \text{if } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{if } c \leq x \leq b \end{cases}.$$

As many finite support distribution, we have a characteristic function and a moment generating function. They have the following expresion:

$$\phi(t) = \frac{(b-c)e^{iat} - (b-a)e^{ict}}{-2(b-a)(c-a)(b-c)t^2} + \frac{-2(c-a)e^{ibt}}{(b-a)(c-a)(b-c)t^2}$$

and

$$M(t) = \frac{(b-c)e^{at} - (b-a)e^{ct}}{2(b-a)(c-a)(b-c)t^2} + \frac{2(c-a)e^{bt}}{(b-a)(c-a)(b-c)t^2}.$$

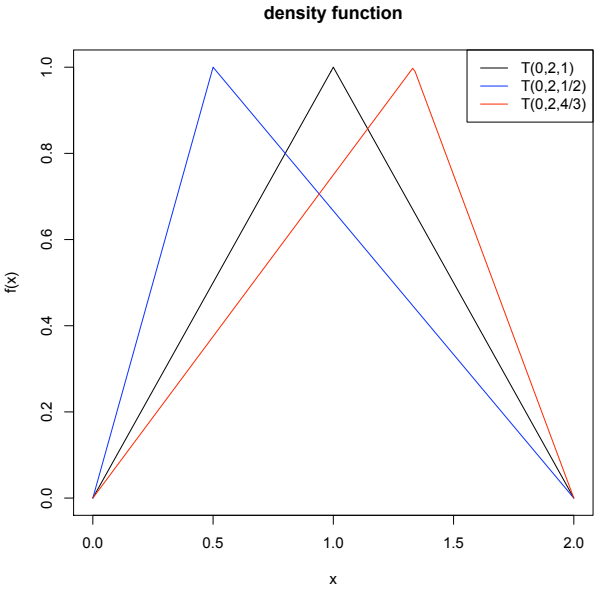


Figure 3.2: Density function for triangular distributions

3.2.2 Properties

The expectation of the triangle distribution is $E(X) = \frac{a+b+c}{3}$ whereas its variance is $Var(X) = \frac{a^2+b^2+c^2}{18} - \frac{ab+ac+bc}{18}$.

3.2.3 Estimation

Maximum likelihood estimators for a, b, c do not have closed form. But we can maximise the log-likelihood numerically. Furthermore, moment based estimators have to be computed numerically solving the system of sample moments and theoretical ones. One intuitive way to estimate the parameters of the triangle distribution is to use sample minimum, maximum and mode: $\hat{a} = X_{1:n}$, $\hat{b} = X_{n:n}$ and $\hat{c} = \text{mode}(X_1, \ldots, X_n)$, where $\text{mode}(X_1, \ldots, X_n)$ is the middle of the interval whose bounds are the most likely order statistics.

3.2.4 Random generation

The inverse function method can be used since the quantile function has a closed form:

$$F^{-1}(u) = \begin{cases} a + \sqrt{u(b-a)(c-a)} & \text{if } 0 \leq u \leq \frac{c-a}{b-a} \\ b - \sqrt{(1-u)(b-a)(b-c)} & \text{if } \frac{c-a}{b-a} \leq u \leq 1 \end{cases}.$$

Thus $F^{-1}(U)$ with U a uniform variable is triangular distributed.

Stein & Keblis (2008) provides new kind of methods to simulate triangular variable. An algorithm for the triangular $\mathcal{T}(0,1,c)$ distribution is provided. It can be adapted for a, b, c in general. Let $\tilde{c}$ be $\frac{c-a}{b-a}$ which is in $]0,1[$. The "minmax" algorithm is

- generate U, V (idependently) from a uniform distribution,
- $X = a + (b-a) \times [(1-\tilde{c}) \min(U,V) + \tilde{c} \max(U,V)]$.

This article also provides another method using a square root of uniform variate, which is called "one line method", but it is not necessary more fast if we use vector operation.

3.2.5 Applications

A typical of the triangle distribution is when we know the minimum and the maximum of outputs of an interest variable plus the most likely outcome, which represent the parameter a, b and c. For example we may use it in business decision making based on simulation of the outcome, in project management to model events during an interval and in audio dithering.

3.3 Beta type I distribution

3.3.1 Characterization

The beta distribution of first kind is a distribution valued in the interval $[0,1]$. Its density is defined as

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)},$$

where $x \in [0,1]$, $a, b > 0$ and $\beta(.,.)$ is the beta function defined in terms of the gamma function.

Since a, b can take a wide range of values, this allows many different shapes for the beta density:

- $a = b = 1$ corresponds to the uniform distribution
- when $a, b < 1$, density is U-shapped

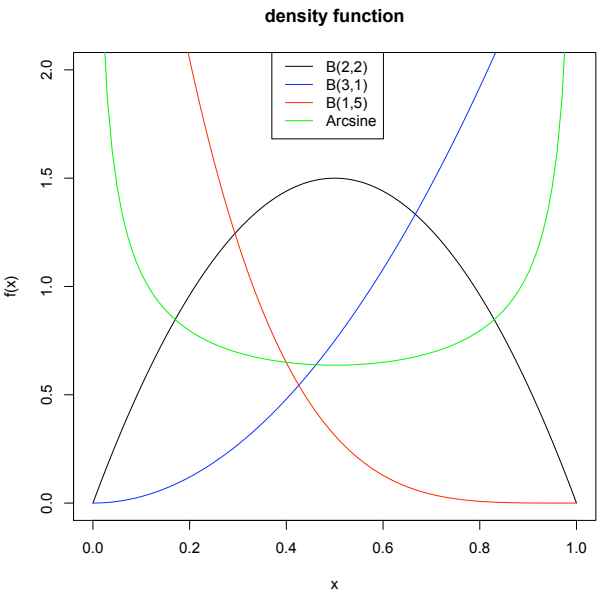


Figure 3.3: Density function for beta distributions

- when $a < 1$, $b \geq 1$ or $a = 1$, $b > 1$, density is strictly decreasing
  - for $a = 1, b > 2$, density is strictly convex
  - for $a = 1, b = 2$, density is a straight line
  - for $a = 1, 1 < b < 2$, density is strictly concave
- when $a = 1, b < 1$ or $a > 1, b \leq 1$, density is strictly increasing
  - for $a > 2, b = 1$, density is strictly convex
  - for $a = 2, b = 1$, density is a straight line
  - for $1 < a < 2, b = 1$, density is strictly concave
- when $a, b > 1$, density is unimodal.

Let us note that $a = b$ implies a symmetric density.

From the density, we can derive its distribution function

$$F(x) = \frac{\beta(a, b, x)}{\beta(a, b)},$$

where $x \in [0, 1]$ and $\beta(.,.,.)$ denotes the incomplete beta function. There is no analytical formula for the incomplete beta function but can be approximated numerically.

There exists a scaled version of the beta I distribution. Let θ be a positive scale parameter. The density of the scaled beta I distribution is given by

$$f(x) = \frac{x^{a-1}(\theta - x)^{b-1}}{\theta^{a+b-1}\beta(a, b)},$$

where $x \in [0, \theta]$. We have the following distribution function

$$F(x) = \frac{\beta(a, b, \frac{x}{\theta})}{\beta(a, b)}.$$

Beta I distributions have moment generating function and characteristic function expressed in terms of series:

$$M(t) = 1 + \sum_{k=1}^{+\infty} \left(\prod_{r=0}^{k-1} \frac{a + r}{a + b + r} \right) \frac{t^k}{k!}$$

and

$$\phi(t) = {}_1F_1(a; a + b; i\, t),$$

where ${}_1F_1$ denotes the hypergeometric function.

3.3.2 Special cases

A special case of the beta I distribution is the arcsine distribution, when $a = b = \frac{1}{2}$. In this special case, we have

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}},$$

from which we derive the following distribution function

$$F(x) = \frac{2}{\pi}\arcsin(\sqrt{x}).$$

Another special case is the power distribution when $b = 1$, with the following density

$$f(x) = ax^{a-1} \quad \text{and} \quad F(x) = x^a,$$

for $0 < x < 1$.

3.3.3 Properties

The moments of the beta I distribution are $E(X) = \frac{a}{a+b}$ and $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$ (and $\frac{\theta a}{a+b}$, $\frac{\theta^2 ab}{(a+b)^2(a+b+1)}$ for the scaled version respectively).

Raw moments for the beta I distribution are given by

$$E(X^r) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)},$$

while central moments have the following expression

$$E\left((X - E(X))^r\right) = \left(-\frac{\alpha}{\alpha+\beta}\right)^r {}_2F_1\left(\alpha, -r, \alpha+\beta, \frac{\alpha+\beta}{\alpha}\right).$$

For the arcsine distribution, we have $\frac{1}{2}$ and $\frac{1}{8}$ respectively. Let us note that the expectation of a arcsine distribution is the least probable value!

Let n be an integer. If we consider n i.i.d. uniform $\mathcal{U}(0,1)$ variables U_i, then the distribution of the maximum $\max_{1\leq i\leq n} U_i$ of these random variables follows a beta I distribution $\mathcal{B}(n,1)$.

3.3.4 Estimation

Maximum likelihood estimators for a and b do not have closed form, we must solve the system

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n}\log(X_i) = \beta(a,b)(\psi(a+b) - \psi(a)) \\ \frac{1}{n}\sum_{i=1}^{n}\log(1-X_i) = \beta(a,b)(\psi(a+b) - \psi(b)) \end{cases}$$

numerically, where $\psi(.)$ denotes the digamma function.

Method of moments gives the following estimators

$$\tilde{a} = \bar{X}_n \left(\frac{\bar{X}_n(1-\bar{X}_n)}{S_n^2} - 1 \right) \quad \text{and} \quad \tilde{b} = \tilde{a}\frac{1-\bar{X}_n}{\bar{X}_n}.$$

3.3.5 Random generation

NEED REFERENCE

3.3.6 Applications

The arcsine distribution (a special case of the beta I) can be used in game theory. If we have two players playint at head/tail coin game and denote by $(S_i)_{i\geq 1}$ the serie of gains of the first player for the different game events, then the distribution of the proportion of gains among all the S_i's that are positive follows asymptotically an arcsine distribution.

3.4 Generalized beta I distribution

3.4.1 Characterization

The generalized beta distribution is the distribution of the variable $\theta X^{\frac{1}{\tau}}$ when X is beta distributed. Thus it has the following density

$$f(x) = \frac{(x/\theta)^{a-1}(1-(x/\theta))^{b-1}}{\beta(a,b)}\frac{\tau}{x}$$

for $0 < x < \theta$ and $a, b, \tau, \theta > 0$. θ is a scale parameter while a, b, τ are shape parameters.

As for the beta distribution, the distribution function is expressed in terms of the incomplete beta function

$$F(x) = \frac{\beta(a,b,(\frac{x}{\theta})^{\tau})}{\beta(a,b)},$$

for $0 < x < \theta$.

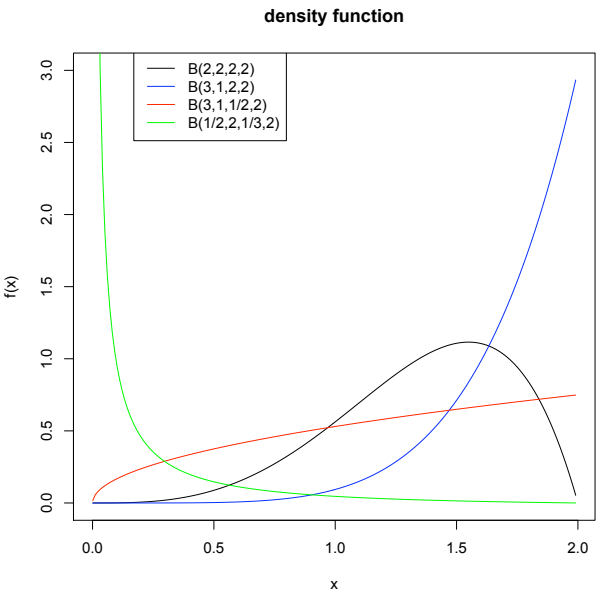


Figure 3.4: Density function for generalized beta distributions

3.4.2 Properties

Moments of the generalized beta distribution are given by the formula

$$E(X^r) = \theta^r \frac{\beta(a + \frac{r}{\tau})}{\beta(a, b)}.$$

For $\tau = \theta = 1$, we retrieve the beta I distribution.

3.4.3 Estimation

Maximum likelihood estimators as well as moment based estimators have no chance to have explicit form, but we can compute it numerically. NEED REFERENCE

3.4.4 Random generation

NEED REFERENCE

3.4.5 Applications

NEED REFERENCE

3.5 Generalization of the generalized beta I distribution

3.5.1 Characterization

A generalization of the generalized beta distribution has been studied in Nadarajah & Kotz (2003). Its density is given by

$$f(x) = \frac{b\beta(a, b)}{\beta(a, b + \gamma)} x^{a+b-1} {}_2F_1(1 - \gamma, a, a + b, x),$$

where $0 < x < 1$ and ${}_2F_1$ denotes the hypergeometric function. Its distribution function is also expressed in terms of the hypergeometric function:

$$F(x) = \frac{b\beta(a, b)}{(a + b)\beta(a, b + \gamma)} x^{a+b} {}_2F_1(1 - \gamma, a, a + b + 1, x),$$

3.5.2 Special cases

Nadarajah & Kotz (2003) list specials cases of this distribution: If $a+b+\gamma=1$ then we get

$$f(x)=\frac{b\Gamma(b)x^{a+b-1}(1-x)^{-a}}{\Gamma(1-a)\Gamma(a+b)}.$$

If $a+b+\gamma=2$ then we get

$$f(x)=\frac{b(a+b-1)\beta(a,b)}{\beta(a,2-a)}\beta(a+b-1,1-a,x)$$

If in addition

- $a+b-1\in\mathbb{N}$, we have

$$f(x)=\frac{b(a+b-1)\beta(a,b)\beta(a+b-1,1-a)}{\beta(a,2-a)}\left(1-\sum_{i=1}^{a+b-1}\frac{\Gamma(i-a)}{\Gamma(1-a)\Gamma(i)}x^{i-1}(1-x)^{1-a}\right)$$

- $a=1/2$ and $b=1$, we have

$$f(x)=\frac{4}{\pi}\arctan\sqrt{\frac{x}{1-x}}$$

- $a=1/2$ and $b=k\in\mathbb{N}$, we have

$$f(x)=\frac{k(2k-1)\beta(1/2,k)\beta(1/2,k-1/2)}{\pi}\left(\frac{2}{\pi}\arctan\sqrt{\frac{x}{1-x}}-\sqrt{x(1-x)}\sum_{i=1}^{k-1}d_i(x,k)\right)$$

If $\gamma=0$ then, we get

$$f(x)=b(a+b-1)(1-x)^{b-1}\beta(a+b-1,1-b,x)$$

If in addition

- $a+b-1\in\mathbb{N}$, we have

$$f(x)=b(a+b-1)\beta(a,b)\beta(a+b-1,1-a)\left(1-\sum_{i=1}^{a+b-1}\frac{\Gamma(i-b)}{\Gamma(1-b)\Gamma(i)}x^{i-1}(1-x)^{1-b}\right)$$

- $a=1$ and $b=1/2$, we have

$$f(x)=\frac{1}{2\sqrt{1-x}}\arctan\sqrt{\frac{x}{1-x}}$$

- $a=k\in\mathbb{N}$, we have

$$f(x)=\frac{(2k-1)\beta(1/2,k-1/2)}{4\sqrt{1-x}}\left(\frac{2}{\pi}\arctan\sqrt{\frac{x}{1-x}}-\sqrt{x(1-x)}\sum_{i=1}^{k-1}d_i(x,k)\right)$$

If $\gamma = 1$, then we get a power function

$$f(x) = (a+b)x^{a+b-1}$$

If $a = 0$, then we get a power function

$$f(x) = bx^{b-1}$$

If $b = 0$, then we get

$$f(x) = \frac{\beta(a,\gamma,x)}{\beta(a,\gamma+1)}$$

If in addition

- $a \in \mathbb{N}$, we have
$$f(x) = \left(\frac{a}{\gamma}+1\right)\left(1 - \sum_{i=1}^{a} \frac{\Gamma(\gamma+i-1)}{\Gamma(\gamma)\Gamma(i)} x^{i-1}(1-x)^{\gamma}\right)$$
- $\gamma \in \mathbb{N}$, we have
$$f(x) = \left(\frac{a}{\gamma}+1\right)\left(1 - \sum_{i=1}^{a} \frac{\Gamma(a+i-1)}{\Gamma(a)\Gamma(i)} x^{a}(1-x)^{i-1}\right)$$
- $a = \gamma = 1/2$, we have
$$f(x) = \frac{4}{\pi} \arctan\sqrt{\frac{x}{1-x}}$$
- $a = k - 1/2$ and $\gamma = j - 1/2$ with $k, j \in \mathbb{N}$, we have
$$f(x) = \left(\frac{a}{\gamma}+1\right)\left(\frac{2}{\pi}\arctan\sqrt{\frac{x}{1-x}} - \sqrt{x(1-x)}\sum_{i=1}^{k-1} d_i(x,k) + \sum_{i=1}^{j-1} c_i(x,k)\right)$$

Where c_i, d_i functions are defined by

$$c_i(x,k) = \frac{\Gamma(k+i-1)x^{k-1/2}(1-x)^{i-1/2}}{\Gamma(k-1/2)\Gamma(i+1/2)}$$

and

$$d_i(x,k) = \frac{\Gamma(i)x^{i-1}}{\Gamma(i+1/2)\Gamma(1/2)}.$$

3.5.3 Properties

Moments for this distribution are given by

$$E(X^n) = \frac{b\beta(a,b)}{(n+a+b)\beta(a,b+\gamma)} x^{a+b}{}_3F_1(1-\gamma, a, n+a+b+1, a+b, n+a+b+1, 1),$$

where ${}_3F_1$ is a hypergeometric function.

3.5.4 Estimation

NEED REFERENCE

3.5.5 Random generation

NEED REFERENCE

3.5.6 Applications

NEED REFERENCE

3.6 Kumaraswamy distribution

3.6.1 Characterization

The Kumaraswamy distribution has the following density function

$$f(x) = abx^{a-1}(1-x^a)^{b-1},$$

where $x \in [0,1]$, $a, b > 0$. Its distribution function is

$$F(x) = 1 - (1-x^a)^b.$$

A construction of the Kumaraswamy distribution use minimum/maximum of uniform samples. Let n be the number of samples (each with m i.i.d. uniform variate), then the distribution of the minimumm of all maxima (by sample) is a Kumaraswamy $\mathcal{K}u(m,n)$, which is also the distribution of one minus the maximum of all minima.

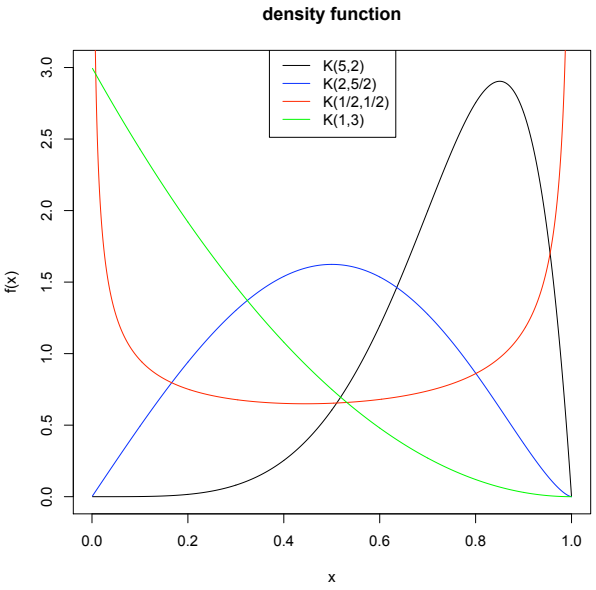


Figure 3.5: Density function for Kumaraswamy distributions

From Jones (2009), the shapes of the density behaves as follows

- $a, b > 1$ implies unimodal density,
- $a > 1, b \leq 1$ implies increasing density,
- $a = b = 1$ implies constant density,
- $a \leq 1, b > 1$ implies decreasing density,

- $a, b < 1$ implies uniantimodal,

which is examplified in the figure on the right.

3.6.2 Properties

Moments for a Kumaraswamy distribution are available and computable with

$$E(X^\tau) = b\beta(1 + \frac{\tau}{a}, b)$$

when $\tau > -a$ with $\beta(.,.)$ denotes the beta function. Thus the expectation of a Kumaraswamy distribution is $E(X) = \frac{b\Gamma(1+1/a)\Gamma(b)}{\Gamma(1+1/a+b)}$ and its variance $Var(X) = b\beta(1 + \frac{2}{a}, b) - b^2\beta^2(1 + \frac{1}{a}, b)$.

3.6.3 Estimation

From Jones (2009), the maximum likelihood estimators are computable by the following procedure

1. solve the equation $\frac{n}{a}\left(1 + \frac{1}{n}\sum_{i=1}^n \frac{\log Y_i}{1-Y_i} + \frac{\sum_{i=1}^n \frac{Y_i \log Y_i}{1-Y_i}}{\sum_{i=1}^n \log(1-Y_i)}\right)$ with $Y_i = X_i^a$ to find $\hat{a}$*,
2. compute $\hat{b} = -n\left(\sum_{i=1}^n \log(1 - X_i^{\hat{a}})\right)^{-1}$.

3.6.4 Random generation

Since the quantile function is explicit

$$F^{-1}(u) = \left(1 - (1-u)^{\frac{1}{b}}\right)^{\frac{1}{a}},$$

an inversion function method $F^{-1}(U)$ with U uniformly distributed is easily computable.

3.6.5 Applications

From wikipedia, we know a good example of the use of the Kumaraswamy distribution: the storage volume of a reservoir of capacity z_{max} whose upper bound is z_{max} and lower bound is 0.

*the solution for this equation exists and is unique.

Chapter 4

The Gaussian family

4.1 The Gaussian (or normal) distribution

The normal distribution comes from the study of astronomical data by the German mathematician Gauss. That's why it is widely called the Gaussian distribution. But there are some hints to think that Laplace has also used this distribution. Thus sometimes we called it the Laplace Gauss distribution, a name introduced by K. Pearson who wants to avoid a querelle about its name.

4.1.1 Characterization

The density of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where $x \in \mathbb{R}$ and $\mu(\in \mathbb{R})$ denotes the mean of the distribution (a location parameter) and $\sigma^2(> 0)$ its variance (a scale parameter).

Its distribution function is then

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, du,$$

which has no explicit expressions. Many softwares have this distribution function implemented, since it is The basic distribution. Generally, we denote by Φ the distribution function a $\mathcal{N}(0, 1)$ normal distribution, called the standard normal distribution. F can be rewritten as

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

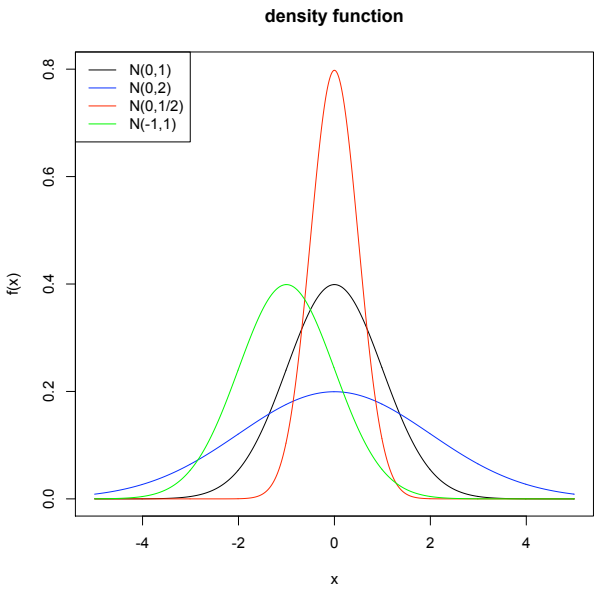


Figure 4.1: The density of Gaussian distributions

Finally, the normal distribution can also be characterized through its moment generating function

$$M(t) = e^{mt+\frac{\sigma^2 t^2}{2}},$$

as well as its characteristic function

$$\phi(t) = e^{imt-\frac{\sigma^2 t^2}{2}}.$$

4.1.2 Properties

It is obvious, but let us recall that the expectation (and the median) of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ is μ and its variance σ^2. Furthermore if $X \sim \mathcal{N}(0,1)$ we have that $E(X^n) = 0$ if x is odd and $\frac{(2n)!}{2^n n!}$ if x is even.

The biggest property of the normal distribution is the fact that the Gaussian belongs to the family of stable distribution (i.e. stable by linear combinations). Thus we have

- if $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\nu, \rho^2)$, then $aX + bY \sim \mathcal{N}(a\mu + b\nu, a^2\sigma^2 + b^2\rho^2 + 2abCov(X,Y))$, with the special case where X, Y are independent cancelling the covariance term.
- if $X \sim \mathcal{N}(\mu, \sigma^2)$, a, b two reals, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

If we consider an i.i.d. sample of n normal random variables $(X_i)_{1 \leq i \leq n}$, then the sample mean $\overline{X}_n$ follows a $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ independently from the sample variance S_n^2 such that $\frac{S_n^2 n}{\sigma^2}$ follows a chi-square distribution with $n-1$ degrees of freedom.

A widely used theorem using a normal distribution is the central limit theorem:
If $(X_i)_{1 \leq i \leq n}$ are i.i.d. with mean m and finite variance s^2, then $\frac{\sum_{i=1}^n X_i - nm}{s\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. If we drop the hypothesis of identical distribution, there is still an asymptotic convergence (cf. theorem of Lindeberg-Feller).

4.1.3 Estimation

The maximum likelihood estimators are

- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ is the unbiased estimator with minimum variance of μ,
- $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$ is the unbiased estimator with minimum variance of σ^{2*},
- $\hat{\sigma}_n = \sqrt{\frac{n-1}{2}}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\sqrt{S_n^2}$ is the unbiased estimator with minimum variance of σ but we generally use $\sqrt{S_n^2}$.

Confidence intervals for these estimators are also well known quantities

*This estimator is not the maximum likelihood estimator since we unbias it.

- $I(\mu) = \left[\overline{X}_n - \sqrt{\frac{S_n^2}{n}} t_{n-1,\alpha/2}; \overline{X}_n + \sqrt{\frac{S_n^2}{n}} t_{n-1,\alpha/2}\right]$,
- $I(\sigma^2) = \left[\frac{S_n^2 n}{z_{n-1,\alpha/2}}; \frac{S_n^2 n}{z_{n-1,1-\alpha/2}}\right]$,

where $t_{n-1,\alpha/2}$ and $z_{n-1,\alpha/2}$ are quantiles of the Student and the Chi-square distribution.

4.1.4 Random generation

The Box-Muller algorithm produces normal random variates:

- generate U, V from a uniform $\mathcal{U}(0,1)$ distribution,
- compute $X = \sqrt{-2\log U}\cos(2\pi V)$ and $Y = \sqrt{-2\log U}\sin(2\pi V)$.

In outputs, X and Y follow a standard normal distribution (independently).

But there appears that this algorithm under estimates the tail of the distribution (called the Neave effect, cf. Patard (2007)), most softwares use the inversion function method, consist in computing the quantile function Φ^{-1} of a uniform variate.

4.1.5 Applications

From wikipedia, here is a list of situations where approximate normality is sometimes assumed

- In counting problems (so the central limit theorem includes a discrete-to-continuum approximation) where reproductive random variables are involved, such as Binomial random variables, associated to yes/no questions or Poisson random variables, associated to rare events;
- In physiological measurements of biological specimens: logarithm of measures of size of living tissue (length, height, skin area, weight) or length of inert appendages (hair, claws, nails, teeth) of biological specimens, in the direction of growth; presumably the thickness of tree bark also falls under this category or other physiological measures may be normally distributed, but there is no reason to expect that a priori;
- Measurement errors are often assumed to be normally distributed, and any deviation from normality is considered something which should be explained;
- Financial variables: changes in the logarithm of exchange rates, price indices, and stock market indices; these variables behave like compound interest, not like simple interest, and so are multiplicative; or other financial variables may be normally distributed, but there is no reason to expect that a priori;
- Light intensity: intensity of laser light is normally distributed or thermal light has a Bose-Einstein distribution on very short time scales, and a normal distribution on longer timescales due to the central limit theorem.

4.2 Log normal distribution

4.2.1 Characterization

One way to characterize a random variable follows a log-normal distribution is to say that its logarithm is normally distributed. Thus the distribution function of a log-normal distribution ($\mathcal{LG}(\mu, \sigma^2)$) is

$$F(x) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right),$$

where Φ denotes the distribution function of the standard normal distribution and $x > 0$.

From this we can derive an explicit expression for the density $\mathcal{LG}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}},$$

for $x > 0$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

A log-normal distribution does not have finite characteristic function or moment generating function.

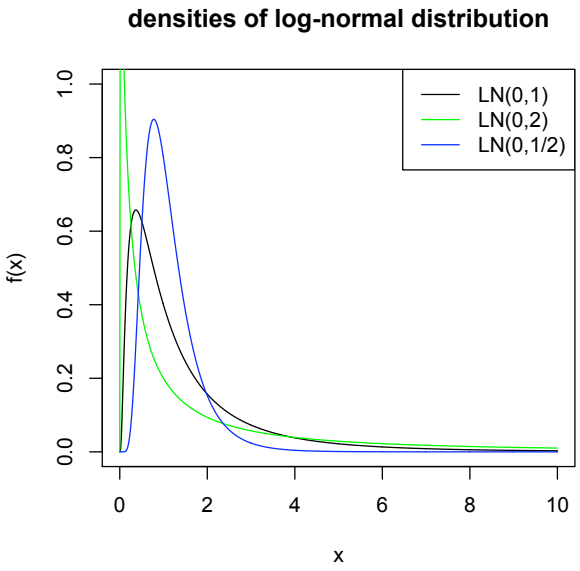


Figure 4.2: The density of log-normal distributions

4.2.2 Properties

The expectation and the variance of a log-normal distribution are $E(X) = e^{\mu + \frac{\sigma^2}{2}}$ and $Var(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$. And raw moments are given by $E(X^n) = e^{n\mu + \frac{n^2\sigma^2}{2}}$. The median of a log-normal distribution is e^{μ}.

From Klugman et al. (2004), we also have a formula for limited expected values

$$E\left((X \wedge L)^k\right) = e^{k(\mu + \frac{k\sigma^2}{2})}\Phi(u - k\sigma) + L^k(1 - \Phi(u)),$$

where $u = \frac{\log(L)-\mu}{\sigma}$.

Since the Gaussian distribution is stable by linear combination, log-normal distribution is stable by product combination. That is to say if we consider X and Y two independent log-normal variables ($\mathcal{LG}(\mu, \sigma^2)$ and $\mathcal{LG}(\nu, \rho^2)$), we have XY follows a log-normal distribution $\mathcal{LG}(\mu+\nu, \sigma^2+\rho^2)$. Let us note that $\frac{X}{Y}$ also follows a log-normal distribution $\mathcal{LG}(\mu - \nu, \sigma^2 + \rho^2)$.

An equivalence of the Limit Central Theorem for the log-normal distribution is the product of i.i.d. random variables $(X_i)_{1\leq i\leq n}$ asymptotically follows a log-normal distribution with paramter $nE(\log(X))$ and $nVar(\log(X))$.

4.2.3 Estimation

Maximum likelihood estimators for μ and σ^2 are simply

- $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log(x_i)$ is an unbiased estimator of μ,
- $\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(\log(x_i)-\hat{\mu})^2$ is an unbiased estimator of σ^{2*}.

One amazing fact about parameter estimations of log-normal distribution is that those estimators are very stable.

4.2.4 Random generation

Once we have generated a normal variate, it is easy to generate a log-normal variate just by taking the exponential of normal variates.

4.2.5 Applications

There are many applications of the log-normal distribution. Limpert et al. (2001) focuses on application of the log-normal distribution. For instance, in finance the Black & Scholes assumes that assets are log-normally distributed (cf. Black & Scholes (1973) and the extraordinary number of articles citing this article). Singh et al. (1997) deals with environmental applications of the log-normal distribution.

*As for the σ^2 estimator of normal distribution, this estimator is not the maximum likelihood estimator since we unbias it.

4.3 Shifted log normal distribution

4.3.1 Characterization

An extension to the log-normal distribution is the translated log-normal distribution. It is the distribution of $X + \nu$ where X follows a log-normal distribution. It is characterized by the following distribution function

$$F(x) = \Phi\left(\frac{\log(x-\nu)-\mu}{\sigma}\right),$$

where Φ denotes the distribution function of the standard normal distribution and $x > 0$. Then we have this expression for the density $\mathcal{TLG}(\nu, \mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma(x-\nu)\sqrt{2\pi}} e^{-\frac{(\log(x-\nu)-\mu)^2}{2\sigma^2}},$$

for $x > 0$, $\mu, \nu \in \mathbb{R}$ and $\sigma^2 > 0$.

As for the log-normal distribution, there is no moment generating function nor characteristic function.

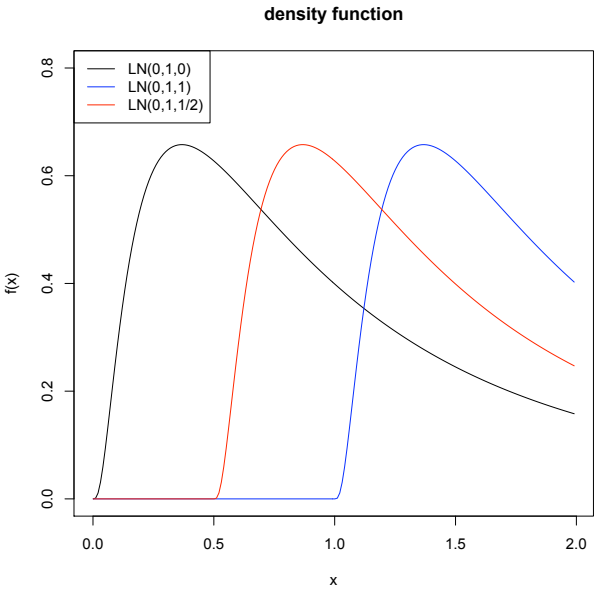


Figure 4.3: The density of shifted log-normal distributions

4.3.2 Properties

The expectation and the variance of a log-normal distribution are $E(X) = \nu + e^{\mu+\frac{\sigma^2}{2}}$ and $Var(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$. And raw moments are given by $E(X^n) = e^{n\mu+\frac{n^2\sigma^2}{2}}$.

4.3.3 Estimation

An intuitive approach is to estimate ν with $X_{1:n}$, then estimate parameters on shifted samples $(X_i - \nu)_i$.

4.3.4 Random generation

Once we have generated a normal variate, it is easy to generate a log-normal variate just by taking the exponential of normal variates and adding the shifted parameter ν.

4.3.5 Applications

An application of the shifted log-normal distribution to finance can be found in Haahtela (2005) or Brigo et al. (2002).

4.4 Inverse Gaussian distribution

4.4.1 Characterization

The density of an inverse Gaussian distribution $\mathcal{IG}(\nu, \lambda)$ is given by

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\lambda \frac{(x-\nu)^2}{2\nu^2 x}\right],$$

while its distribution function is

$$F(x) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\nu} - 1\right)\right] + e^{2\lambda/\nu}\Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\nu} + 1\right)\right]$$

for $x > 0$, $\nu \in \mathbb{R}$, $\lambda > 0$ and Φ denotes the usual standard normal distribution.

Its characteristic function is

$$\phi(t) = e^{\left(\frac{\lambda}{\nu}\right)\left[1-\sqrt{1-\frac{2\nu^2 it}{\lambda}}\right]}.$$

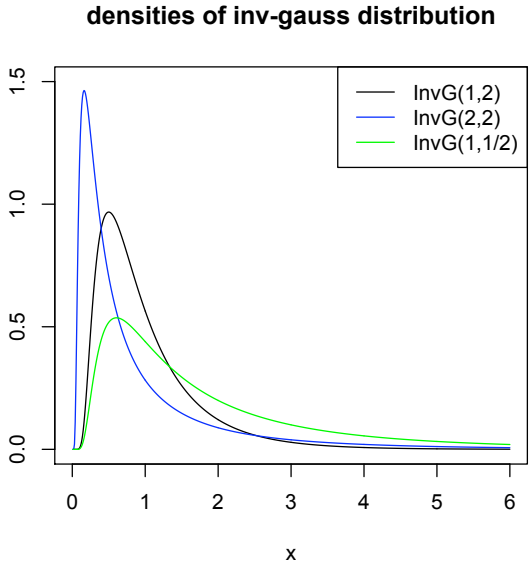


Figure 4.4: The density of inverse Gaussian distributions

The moment generating function is expressed as

$$M(t) = e^{\left(\frac{\lambda}{\nu}\right)\left[1-\sqrt{1-\frac{2\nu^2 t}{\lambda}}\right]}.$$

4.4.2 Properties

The expectation of an inverse Gaussian distribution $\mathcal{IG}(\nu, \lambda)$ is ν and its variance $\frac{\nu^3}{\lambda}$.

Moments for the inverse Gaussian distribution are given $E(X^n) = \nu^n \sum_{i=0}^{n-1} \frac{\Gamma(n+i)}{\Gamma(i+1)\Gamma(n-i)} (\frac{2\lambda}{\nu})^i$ for n integer.

From Yu (2009), we have the following properties

- if X is inverse Gaussian distributed $\mathcal{IG}(\nu, \lambda)$, then aX follows an inverse Gaussian distribution $\mathcal{IG}(a\nu, a\lambda)$ for $a > 0$
- if $(X_i)_i$ are i.i.d. inverse Gaussian variables, then the sum $\sum_{i=1}^{n} X_i$ still follows an inverse Gaussian distribution $\mathcal{IG}(n\nu, n^2\lambda)$

4.4.3 Estimation

Maximum likelihood estimators of ν and λ are

$$\hat{\mu} = \bar{X}_n \quad \text{and} \quad \hat{\lambda} = n \left(\sum_{i=1}^{n} \left(\frac{1}{X_i} - \frac{1}{\hat{\mu}} \right) \right)^{-1}.$$

From previous properties, $\hat{\mu}$ follows an inverse gaussian distribution $\mathcal{IG}(\mu, n\lambda)$ and $\frac{n\lambda}{\hat{\lambda}}$ follows a chi-squared distribution χ^2_{n-1}.

4.4.4 Random generation

NEED

Mitchael,J.R., Schucany, W.R. and Haas, R.W. (1976). Generating random roots from variates using transformations with multiple roots. American Statistician. 30-2. 88-91.

4.4.5 Applications

NEED REFERENCE

4.5 The generalized inverse Gaussian distribution

This section is taken from Breymann & Lüthi (2008).

4.5.1 Characterization

A generalization of the inverse Gaussian distribution exists but there is no closed form for its distribution function and its density used Bessel functions. The latter is as follows

$$f(x) = \left(\frac{\psi}{\chi} \right)^{\frac{\lambda}{2}} \frac{x^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp\left\{ -\frac{1}{2} \left(\frac{\chi}{x} + \psi x \right) \right\},$$

where $x > 0$ and K_λ denotes the modified Bessel function. Parameters must satisfy

- $\chi > 0, \psi \geq 0$, when $\lambda < 0$,
- $\chi > 0, \psi > 0$, when $\lambda = 0$,

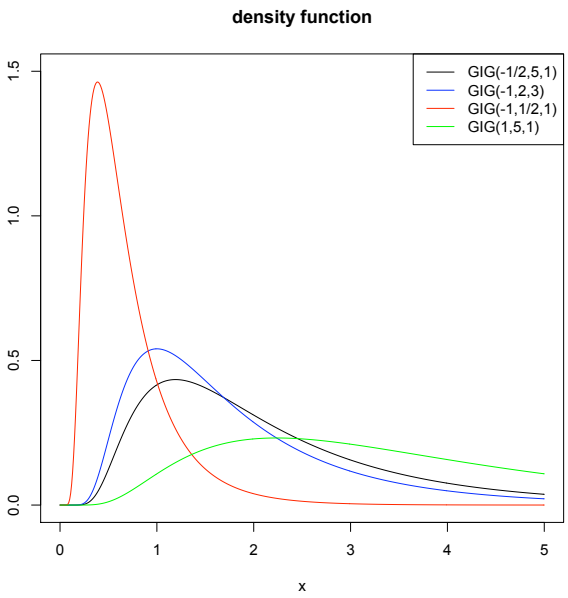


Figure 4.5: The density of generalized inverse Gaussian distributions

- $\chi \geq 0, \psi > 0$, when $\lambda > 0$.

The generalized inverse Gaussian is noted as $\mathcal{GIG}(\lambda, \psi, \chi)$.

Closed form for distribution function??

Plot

The moment generating function is given by

$$M(t) = \left(\frac{\psi}{\psi - 2t}\right)^{\lambda/2} \frac{K_\lambda(\sqrt{\chi(\psi - 2t)})}{K_\lambda(\sqrt{\chi\psi})}. \tag{4.1}$$

4.5.2 Properties

The expectation is given by

$$\sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})},$$

and more generally the n-th moment is as follows

$$E(X^n) = \left(\frac{\chi}{\psi}\right)^{\frac{n}{2}} \frac{K_{\lambda+n}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}.$$

Thus we have the following variance

$$Var(X) = \frac{\chi}{\psi} \frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} - \frac{\chi}{\psi} \left(\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\right)^2.$$

Furthermore,

$$E(\log X) = \left.\frac{\partial dE(X^\alpha)}{\partial d\alpha}\right|_{\alpha=0}. \tag{4.2}$$

Note that numerical calculations of $E(\log X)$ may be performed with the integral representation as well.

4.5.3 Estimation

NEED REFERENCE

4.5.4 Random generation

NEED REFERENCE

Chapter 5

Exponential distribution and its extensions

5.1 Exponential distribution

5.1.1 Characterization

The exponential is a widely used and widely known distribution. It is characterized by the following density

$$f(x) = \lambda e^{-\lambda x},$$

for $x > 0$ and $\lambda > 0$. Its distribution function is

$$F(x) = 1 - e^{-\lambda x}.$$

Since it is a light-tailed distribution, the moment generating function of an exponential distribution $\mathcal{E}(\lambda)$ exists which is

$$M(t) = \frac{\lambda}{\lambda - t},$$

while its characteristic function is

$$\phi(t) = \frac{\lambda}{\lambda - it}.$$

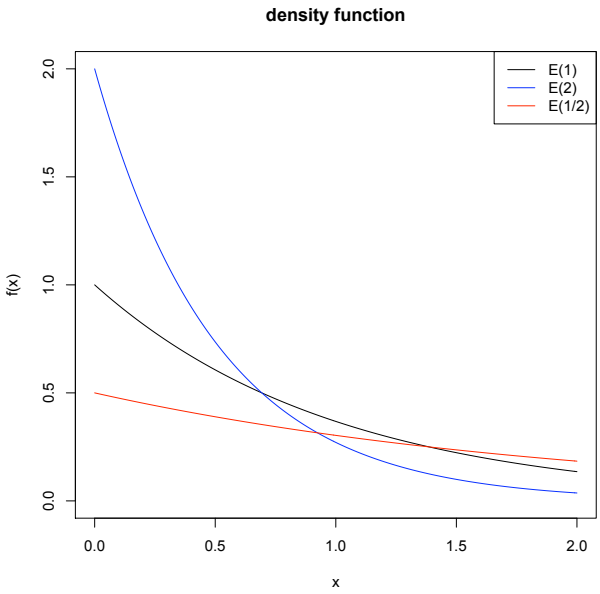


Figure 5.1: Density function for exponential distributions

5.1.2 Properties

The expectation and the variance of an exponential distribution $\mathcal{E}(\lambda)$ are $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$. Furthermore the n-th moment is given by

$$E(X^n) = \frac{\Gamma(n+1)}{\lambda^n}.$$

The exponential distribution is the only one continuous distribution to verify the lack of memory property. That is to say if X is exponentially distributed, we have

$$\frac{P(X > t+s)}{P(X > s)} = P(X > t),$$

where $t, s > 0$.

If we sum n i.i.d. exponentially distributed random variables, we get a gamma distribution $\mathcal{G}(n, \lambda)$.

5.1.3 Estimation

The maximum likelihood estimator and the moment based estimator are the same

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}_n},$$

for a sample $(X_i)_{1 \leq i \leq n}$. But the unbiased estimator with mininum variance is

$$\tilde{\lambda} = \frac{n-1}{\sum_{i=1}^{n} X_i}.$$

Exact confidence interval for parameter λ is given by

$$I_\alpha(\lambda) = \left[\frac{z_{2n,1-\frac{\alpha}{2}}}{2\sum_{i=1}^{n} X_i}, \frac{z_{2n,\frac{\alpha}{2}}}{2\sum_{i=1}^{n} X_i} \right],$$

where $z_{n,\alpha}$ denotes the α quantile of the chi-squared distribution.

5.1.4 Random generation

Despite the quantile function is $F^{-1}(u) = -\frac{1}{\lambda}\log(1-u)$, generally the exponential distribution $\mathcal{E}(\lambda)$ is generated by applying $-\frac{1}{\lambda}\log(U)$ on a uniform variate U.

5.1.5 Applications

From wikipedia, the exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process.

The exponential distribution may be viewed as a continuous counterpart of the geometric distribution, which describes the number of Bernoulli trials necessary for a "discrete" process to change state. In contrast, the exponential distribution describes the time for a continuous process to change state.

In real-world scenarios, the assumption of a constant rate (or probability per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But

if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

- the time until a radioactive particle decays, or the time between beeps of a geiger counter;
- the time it takes before your next telephone call
- the time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit "distance":

- the distance between mutations on a DNA strand;
- the distance between roadkill on a given road;

In queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. (The inter-arrival of customers for instance in a system is typically modeled by the Poisson distribution in most management science textbooks.) The length of a process that can be thought of as a sequence of several independent tasks is better modeled by a variable following the Erlang distribution (which is the distribution of the sum of several independent exponentially distributed variables).

Reliability theory and reliability engineering also make extensive use of the exponential distribution. Because of the "memoryless" property of this distribution, it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory. It is also very convenient because it is so easy to add failure rates in a reliability model. The exponential distribution is however not appropriate to model the overall lifetime of organisms or technical devices, because the "failure rates" here are not constant: more failures occur for very young and for very old systems.

In physics, if you observe a gas at a fixed temperature and pressure in a uniform gravitational field, the heights of the various molecules also follow an approximate exponential distribution. This is a consequence of the entropy property mentioned below.

5.2 Shifted exponential

5.2.1 Characterization

The distribution of the shifted exponential distribution is simply the distribution of $X - \tau$ when X is exponentially distributed. Therefore the density is given by

$$f(x) = \lambda e^{-\lambda(x-\tau)}$$

for $x > \tau$. The distribution function is given by

$$F(x) = 1 - e^{-\lambda(x-\tau)}$$

for $x > \tau$.

As for the exponential distribution, there exists a moment generating function

$$M(t) = e^{-t\tau}\frac{\lambda}{\lambda - t}$$

and also a characteristic function

$$\phi(t) = e^{-it\tau}\frac{\lambda}{\lambda - it}.$$

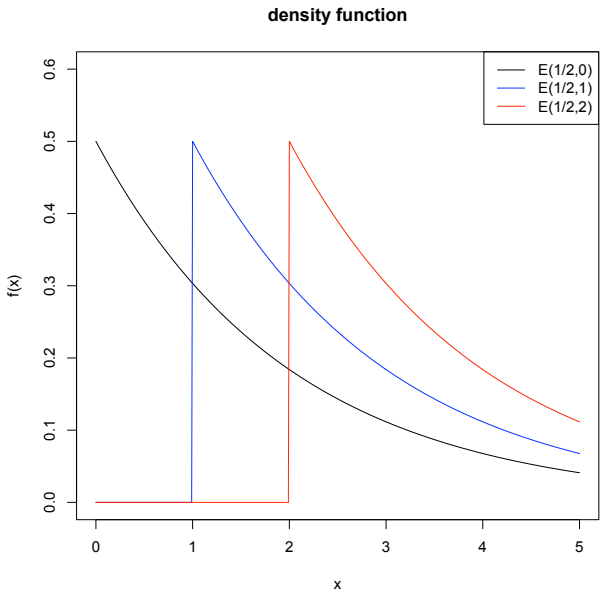


Figure 5.2: Density function for shifted exponential distributions

5.2.2 Properties

The expectation and the variance of an exponential distribution $\mathcal{E}(\lambda, \tau)$ are $\tau + \frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$.

Furthermore the n-th moment (for n integer) is computable with the binomial formula by

$$E(X^n) = \sum_{i=0}^{n} \frac{n!}{(n-i)!}\frac{(-\tau)^n}{(-\lambda\tau)^i}.$$

5.2.3 Estimation

Maximum likelihood estimator for τ and λ are given by

$$\hat{\tau} = X_{1:n} \quad \text{and} \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}(X_i - \hat{\tau})}$$

where $X_{i:n}$ denotes the ith order statistic. Since the minimum $X_{1:n}$ follows a shifted exponential distribution $\mathcal{E}(n\lambda, \tau)$, we have $\hat{\tau}$ is biased but asympotically unbiased.

NEED REFERENCE for unbiased estimators

5.2.4 Random generation

The random generation is simple: just add τ to the algorithm of exponential distribution.

5.2.5 Applications

NEED REFERENCE

5.3 Inverse exponential

5.3.1 Characterization

This is the distribution of the random variable $\frac{1}{X}$ when X is exponentially distributed. The density defined as

$$f(x) = \frac{\lambda}{x^2} e^{-\frac{\lambda}{x}},$$

where $x > 0$ and $\lambda > 0$. The distribution function can then be derived as

$$F(x) = e^{-\frac{\lambda}{x}}.$$

We can define inverse exponential distributions with characteristic or moment generating functions

$$\phi(t) = 2\sqrt{-it\lambda} K_1\left(2\sqrt{-i\lambda t}\right)$$

and

$$M(t) = 2\sqrt{-it\lambda} K_1\left(2\sqrt{-\lambda t}\right).$$

where $K_\cdot(.)$ denotes the modified Bessel function.

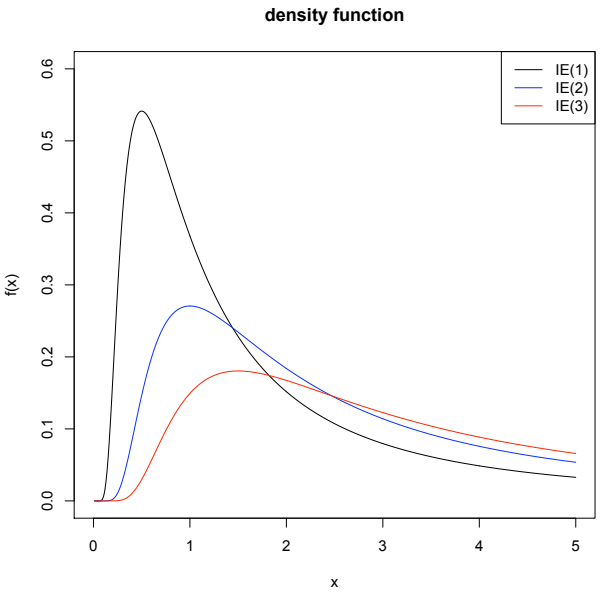


Figure 5.3: Density function for inverse exponential distributions

5.3.2 Properties

Moments of the inverse exponential distribution are given by

$$E(X^r) = \lambda^r * \Gamma(1 - r)$$

for $r < 1$. Thus the expectation and the variance of the inverse exponential distribution do not exist.

5.3.3 Estimation

Maximum likelihood estimator of λ is

$$\hat{\lambda} = n \left(\sum_{i=1}^{n} \frac{1}{X_i} \right)^{-1},$$

which is also the moment based estimator with $E(X^{-1}) = \lambda^{-1}$.

5.3.4 Random generation

The algorithm is simply to inverse an exponential variate of parameter $\frac{1}{\lambda}$, i.e. $(-\lambda \log(U))^{-1}$ for an uniform variable U.

5.3.5 Applications

NEED REFERENCE

5.4 Gamma distribution

5.4.1 Characterization

The gamma distribution is a generalization of the exponential distribution. Its density is defined as

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1},$$

where $x \geq 0$, $\alpha, \lambda > 0$ and Γ denotes the gamma function. We retrieve the exponential distribution by setting α to 1. When α is an integer, the gamma distribution is sometimes called the Erlang distribution.

The distribution function can be expressed in terms of the incomplete gamma distribution. We get

$$F(x) = \frac{\gamma(\alpha, \lambda x)}{\Gamma(\alpha)},$$

where $\gamma(.,.)$ is the incomplete gamma function.

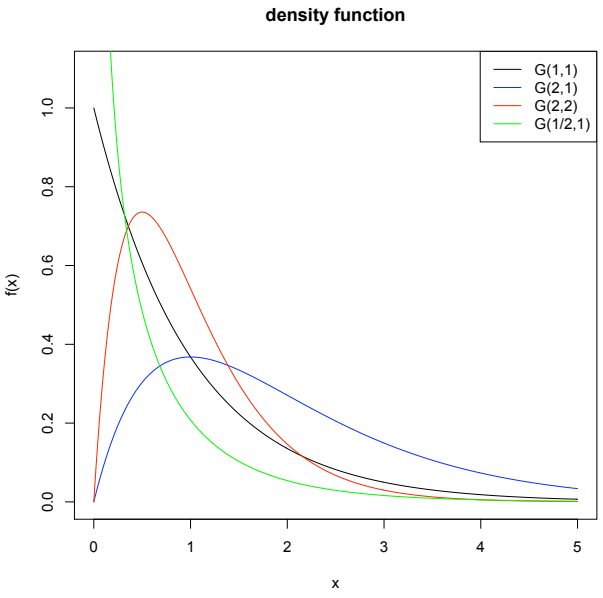


Figure 5.4: Density function for gamma distributions

There is no analytical formula except when we deal with Erlang distribution (i.e. $\alpha \in \mathbb{N}$). In this case, we have

$$F(x) = 1 - \sum_{i=0}^{\alpha-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x}.$$

For the gamma distribution, the moment generating and characteristic functions exist.

$$\phi(t) = \left(\frac{\lambda}{\lambda - it}\right)^{-\alpha},$$

and

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{-\alpha}.$$

5.4.2 Properties

The expectation of a gamma distribution $\mathcal{G}(\alpha, \lambda)$ is $E(X) = \frac{\alpha}{\lambda}$, while its variance is $Var(X) = \frac{\alpha}{\lambda^2}$.

For a gamma distribution $\mathcal{G}(\alpha, \lambda)$, the τth moment is given by

$$E(X^r) = \lambda^r \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)},$$

provided that $\alpha + r > 0$.

As for the exponential, we have a property on the convolution of gamma distributions. Let X and Y be gamma distributed $\mathcal{G}(\alpha, \lambda)$ and $\mathcal{G}(\beta, \lambda)$, we can prove that $X + Y$ follows a gamma distribution $\mathcal{G}(\alpha + \beta, \lambda)$.

For X and Y gamma distributed ($\mathcal{G}(\alpha, \lambda)$ and $\mathcal{G}(\beta, \lambda)$ resp.), we also have that $\frac{X}{X+Y}$ follows a beta distribution of the first kind with parameter α and β.

5.4.3 Estimation

Method of moments give the following estimators

$$\tilde{\alpha} = \frac{(\bar{X}_n)^2}{S_n^2} \quad \text{and} \quad \tilde{\lambda} = \frac{\bar{X}_n}{S_n^2}.$$

with $\bar{X}_n$ and S_n^2 the sample mean and variance.

Maximum likelihood estimators of α, λ verify the system

$$\begin{cases} \log \alpha - \psi(\alpha) = \log(\frac{1}{n}\sum_{i=1}^n X_i) - \frac{1}{n}\sum_{i=1}^n \log X_i \\ \lambda = \frac{n\alpha}{\sum_{i=1}^n X_i} \end{cases},$$

where $\psi(.)$ denotes the digamma function. The first equation can be solved numerically* to get $\hat{\alpha}$ and then $\hat{\lambda} = \frac{\hat{\alpha}}{\bar{X}_n}$. But $\hat{\lambda}$ is biased, so the unbiased estimator with minimum variance of λ is

$$\bar{\lambda} = \frac{\hat{\alpha}n}{\hat{\alpha}n - 1}\frac{\hat{\alpha}}{\bar{X}_n}$$

NEED REFERENCE for confidence interval

5.4.4 Random generation

Simulate a gamma $\mathcal{G}(\alpha, \lambda)$ is quite tricky for non integer shape parameter. Indeed, if the shape parameter α is integer, then we simply sum α exponential random variables $\mathcal{E}(\lambda)$. Otherwise we need to add a gamma variable $\mathcal{G}(\alpha - \lfloor\alpha\rfloor, \lambda)$. This is carried out by an acceptance/rejection method.

NEED REFERENCE

5.4.5 Applications

NEED REFERENCE

5.5 Generalized Erlang distribution

5.5.1 Characterization

As the gamma distribution is the distribution of the sum of i.i.d. exponential distributions, the generalized Erlang distribution is the distribution of the sum independent exponential distributions. Sometimes it is called the hypo-exponential distribution. The density is defined as

$$f(x) = \sum_{i=1}^{d}\left(\prod_{j=1, j\neq i}^{d}\frac{\lambda_j}{\lambda_j - \lambda_i}\right)\lambda_i e^{-\lambda_i x},$$

where $x \geq 0$ and $\lambda_j > 0$'s† are the paremeters (for each exponential distribution building the generalized Erlang distribution). There is an

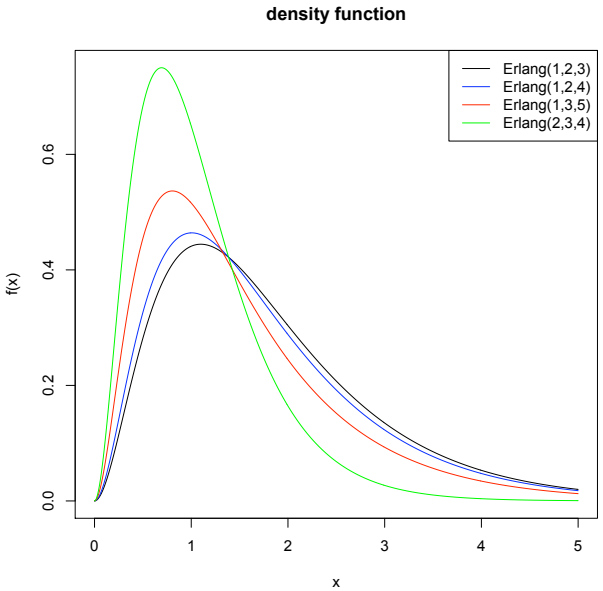


Figure 5.5: Density function for generalized Erlang distributions

*algorithm can be initialized with $\tilde{\alpha}$.

†with the constraint that all λ_j's are strictly different.

explicit form for the distribution function:

$$F(x) = \sum_{i=1}^{d} \left(\prod_{j=1, j\neq i}^{d} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) (1 - e^{-\lambda_i x}).$$

This distribution is noted $\mathcal{E}rlang(\lambda_1, \ldots, \lambda_d)$.
Of course, we retrieve the Erlang distribution when $\forall i, \lambda_i = \lambda$.

Finally, the characteristic and moment generating functions of generalized Erlang distribution are

$$\phi(t) = \prod_{j=1}^{d} \frac{\lambda_j}{\lambda_j - it} \quad \text{and} \quad M(t) = \prod_{j=1}^{d} \frac{\lambda_j}{\lambda_j - t}.$$

5.5.2 Properties

The expectation of the generalized Erlang distribution is simply $E(X) = \sum_{i=1}^{d} \frac{1}{\lambda_i}$ and its variance $Var(X) = \sum_{i=1}^{d} \frac{1}{\lambda_i^2}$.

5.5.3 Estimation

NEED REFERENCE

5.5.4 Random generation

The algorithm is very easy simulate independently d random variables exponentially $\mathcal{E}(\lambda_j)$ distributed and sum them.

5.5.5 Applications

NEED REFERENCE

5.6 Chi-squared distribution

A special case of the gamma distribution is the chi-squared distribution. See section 6.1.

5.7 Inverse Gamma

5.7.1 Characterization

The inverse gamma distribution is the distribution of a random variable $\frac{1}{X}$ when X is gamma distributed. Hence the density is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)x^{\alpha+1}}e^{-\frac{\lambda}{x}},$$

where $x > 0$ and $\beta, \alpha > 0$. From this, we can derive the distribution function

$$F(x) = \frac{\gamma(\alpha, \frac{\lambda}{x})}{\Gamma(\alpha)}.$$

We can define inverse gamma distributions with characteristic or moment generating functions

$$\phi(t) = \frac{2\sqrt{-it\lambda}^\alpha}{\Gamma(\alpha)}K_\alpha(2\sqrt{-i\lambda t})$$

and

$$M(t) = \frac{2\sqrt{-it\lambda}^\alpha}{\Gamma(\alpha)}K_\alpha(2\sqrt{-\lambda t}).$$

where $K_\cdot(.)$ denotes the modified Bessel function.

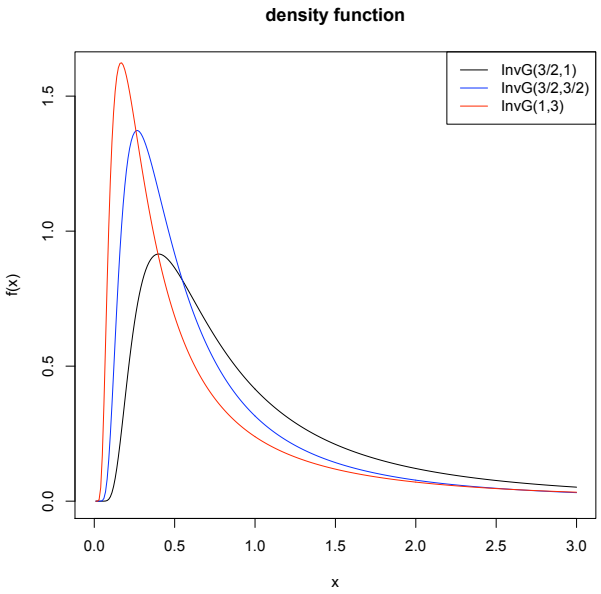


Figure 5.6: Density function for inverse gamma distributions

5.7.2 Properties

The expectation exists only when $\alpha > 1$ and in this case $E(X) = \frac{\lambda}{\alpha-1}$, whereas the variance is only finite if $\alpha > 2$ and $Var(X) = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}$.

5.7.3 Estimation

Method of moments give the following estimators

$$\tilde{\alpha} = 2 + \frac{(\bar{X}_n)^2}{S_n^2} \quad \text{and} \quad \tilde{\lambda} = \bar{X}_n(\tilde{\alpha} - 1)$$

with $\bar{X}_n$ and S_n^2 the sample mean and variance. If the variance does not exist, then α will be 2, it means we must use the maximum likelihood estimator (which works also for $\alpha \leq 2$).

Maximum likelihood estimators of α, λ verify the system

$$\begin{cases} \log\alpha - \psi(\alpha) = \log(\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}) - \frac{1}{n}\sum_{i=1}^n \log\frac{1}{X_i} \\ \lambda = \alpha\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}\right)^{-1} \end{cases},$$

where $\psi(.)$ denotes the digamma function. The first equation can be solved numerically* to get $\hat{\alpha}$ and then $\hat{\lambda}$ with the second equation.

5.7.4 Random generation

Simply generate a gamma variable $\mathcal{G}(\alpha, 1/\lambda)$ and inverse it.

5.7.5 Applications

NEED REFERENCE

5.8 Transformed or generalized gamma

5.8.1 Characterization

The transformed gamma distribution is defined by the following density function

$$f(x) = \frac{\tau(\frac{x}{\lambda})^{\alpha\tau-1}e^{-(\frac{x}{\lambda})^\tau}}{\lambda\Gamma(\alpha)},$$

where $x > 0$ and $\alpha, \lambda, \tau > 0$. Thus, the distribution function is

$$F(x) = \frac{\gamma(\alpha, (\frac{x}{\lambda})^\tau)}{\Gamma(\alpha)}.$$

This is the distribution of the variable $\lambda X^{\frac{1}{\tau}}$ when X is gamma distributed $\mathcal{G}(\alpha, 1)$.

Obviously, a special case of the transformed gamma is the gamma distribution with $\tau = 1$. But we get the Weibull distribution with $\alpha = 1$.

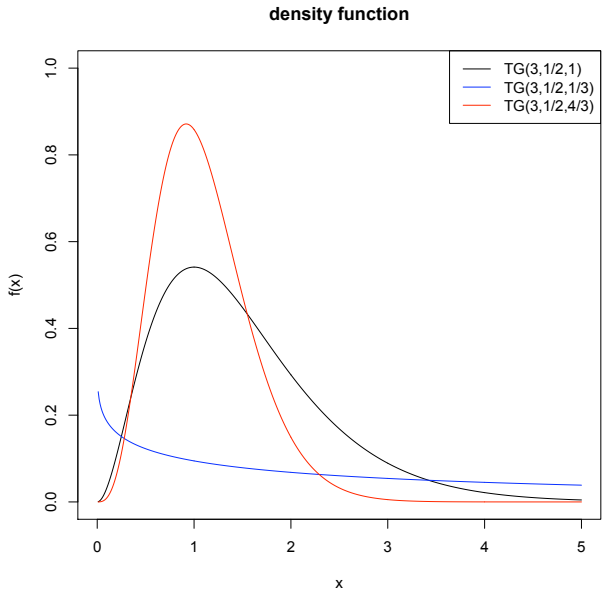


Figure 5.7: Density function for transformed gamma distributions

*algorithm can be initialized with $\tilde{\alpha}$.

5.8.2 Properties

The expectation of the transformed gamma distribution is $E(X) = \frac{\lambda\Gamma(\alpha+\frac{1}{\tau})}{\Gamma(\alpha)}$ and its variance $Var(X) = \frac{\lambda^2\Gamma(\alpha+\frac{2}{\tau})}{\Gamma(\alpha)} - E^2[X]$.

From Venter (1983) moments are given by

$$E(X^r) = \lambda^r \frac{\Gamma(\alpha + \frac{r}{\tau})}{\Gamma(\alpha)},$$

with $\alpha + \frac{r}{\tau} > 0$.

5.8.3 Estimation

Maximum likelihood estimators verify the following system

$$\begin{cases} \psi(\alpha) - \log\alpha = \tau\frac{1}{n}\sum_{i=1}^{n}\log X_i - \log(\frac{1}{n}\sum_{i=1}^{n} X_i^\tau) \\ \alpha = \frac{1}{n}\sum_{i=1}^{n} X_i^\tau \left(\frac{1}{n}\sum_{i=1}^{n} X_i^\tau \log X_i - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^\tau\right)\left(\frac{1}{n}\sum_{i=1}^{n}\log X_i\right)\right)^{-1} \\ \lambda = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^\tau\right)^\tau \alpha^{-\tau} \end{cases},$$

where ψ denotes the digamma function. This system can be solved numerically.

TODO : use Gomes et al. (2008)

5.8.4 Random generation

Generate a gamma distributed variable ($\mathcal{G}(\alpha, 1)$), raise it to power $\frac{1}{\tau}$ and multiply it by λ.

5.8.5 Applications

In an actuarial context, the transformed gamma may be useful in loss severity, for example, in workers' compensation, see Venter (1983).

5.9 Inverse transformed Gamma

5.9.1 Characterization

The transformed gamma distribution is defined by the following density function

$$f(x) = \frac{\tau(\frac{\lambda}{x})^{\alpha\tau}e^{-(\frac{\lambda}{x})^{\tau}}}{x\Gamma(\alpha)},$$

where $x > 0$ and $\alpha, \lambda, \tau > 0$. Thus, the distribution function is

$$F(x) = 1 - \frac{\gamma(\alpha, (\frac{\lambda}{x})^{\tau})}{\Gamma(\alpha)}.$$

This is the distribution of $\left(\frac{\lambda}{X}\right)^{\frac{1}{\tau}}$ when X is gamma distributed $\mathcal{G}(\alpha, 1)$.

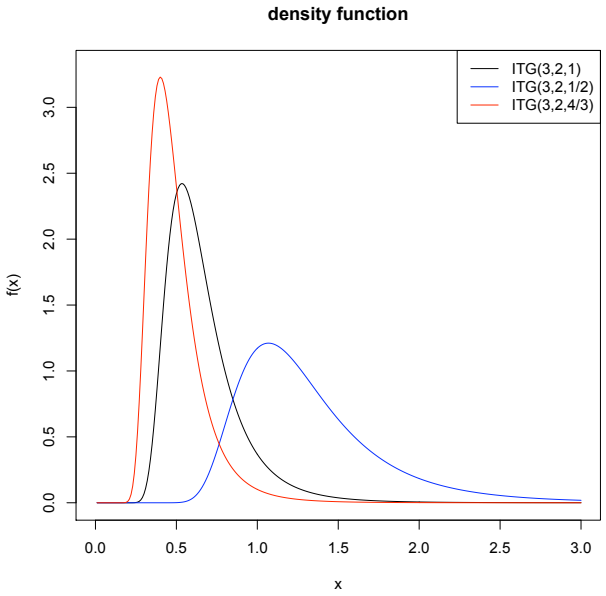


Figure 5.8: Density function for inverse transformed gamma distributions

5.9.2 Properties

The expectation of the transformed gamma distribution is $E(X) = \frac{\lambda\Gamma(\alpha-\frac{1}{\tau})}{\Gamma(\alpha)}$ and its variance $Var(X) = \frac{\lambda^2\Gamma(\alpha-\frac{2}{\tau})}{\Gamma(\alpha)} - E^2[X]$.

From Klugman et al. (2004), we have the following formula for the moments

$$E(X^r) = \frac{\lambda^r\Gamma(\alpha - \frac{r}{\tau})}{\Gamma(\alpha)}.$$

5.9.3 Estimation

NEED REFERENCE

5.9.4 Random generation

Simply simulate a gamma $\mathcal{G}(\alpha, 1)$ distributed variable, inverse it, raise it to power $\frac{1}{\alpha}$ and multiply it by λ.

5.9.5 Applications

NEED REFERENCE

5.10 Log Gamma

5.10.1 Characterization

Density function for log-gamma distribution is expressed as

$$f(x) = \frac{e^{k\frac{x-a}{b} - e^{\frac{x-a}{b}}}}{\Gamma(k)}$$

for $x > 0$, where a is the location parameter, $b > 0$ the scale parameter and $k > 0$ the shape parameter. The distribution function is

$$F(x) = \frac{\gamma(k, e^{\frac{x-a}{b}})}{\Gamma(k)},$$

for $x > 0$. This is the distribution of $a + b\log(X)$ when X is gamma $\mathcal{G}(k, 1)$.

5.10.2 Properties

The expectation is $E(X) = a + b\psi(k)$ and the variance $Var(X) = b^2\psi_1(k)$ where ψ is the digamma function and ψ_1 the trigamma function.

5.10.3 Estimation

NEED REFERENCE

5.10.4 Random generation

Simply simulate a gamma $\mathcal{G}(k, 1)$ distributed variable and returns $a + b\log(X)$.

5.10.5 Applications

NEED REFERENCE

5.11 Weibull distribution

5.11.1 Characterization

Despite the fact the Weibull distribution is not particularly related to the chi distribution, its density tends exponentially fast to zero, as chi's related distribution. The density of a Weibull distribution is given by

$$f(x) = \frac{\beta}{\eta^\beta} x^{\beta-1} e^{-(\frac{x}{\eta})^\beta},$$

where $x > 0$ and $\eta, \beta > 0$. In terms of distribution function, the Weibull can be defined as

$$F(x) = 1 - e^{-(\frac{x}{\beta})^\eta}.$$

There exists a second parametrization of the Weibull distribution. We have

$$f(x) = \tau\lambda x^{\lambda-1} e^{-\tau x^\lambda},$$

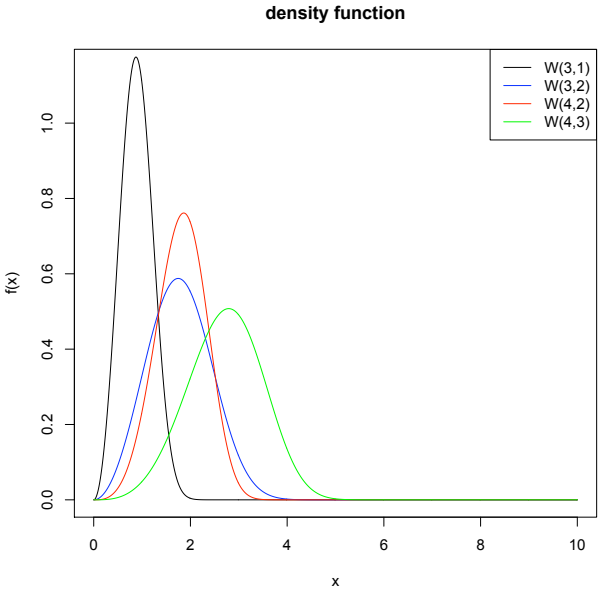


Figure 5.9: Density function for Weibull distributions

with the same constraint on the parameters $\tau, \lambda > 0$. In this context, the distribution function is

$$F(x) = 1 - e^{-\tau x^\lambda}.$$

We can pass from the first parametrization to the second one with

$$\begin{cases} \lambda = \beta \\ \tau = \frac{1}{\eta^\beta} \end{cases}.$$

5.11.2 Properties

The expectation of a Weibull distribution $\mathcal{W}(\eta, \beta)$ is $E(X) = \eta\Gamma(1+\frac{1}{\beta})$ and the variance $Var(X) = \eta^2[\Gamma(\frac{\beta+2}{\beta}) - \Gamma(\frac{\beta+1}{\beta})^2]$. In the second parametrization, we have $E(X) = \frac{\tau(1+\frac{1}{\tau})}{\lambda^{\frac{1}{\tau}}}$ and $Var(X) = \frac{1}{\lambda^{\frac{2}{\tau}}}(\tau(1+\frac{2}{\tau}) - \tau(1+\frac{1}{\tau})^2)$.

The rth raw moment $E(X^r)$ of the Weibull distribution $\mathcal{W}(\eta, \beta)$ is given by $\eta\Gamma(1+\frac{r}{\beta})$ for $r > 0$.

The Weibull distribution is the distribution of the variable $\frac{X^\beta}{\eta}$ where X follows an exponential distribution $\mathcal{E}(1)$.

5.11.3 Estimation

We work in this sub-section with the first parametrization. From the cumulative distribution, we have

$$\log(-\log|1 - F(x)|) = \beta \log x - \beta \log \eta.$$

Thus we can an estimation of β and η by regressing $\log(-\log|\frac{i}{n}|)$ on $\log X_{i:n}$. Then we get the following estimators

$$\tilde{\beta} = \hat{a} \quad \text{and} \quad \tilde{\eta} = e^{-\frac{\hat{b}}{\hat{a}}},$$

where $\hat{a}$ and $\hat{b}$ are respectively the slope and the intercept of the regression line.

The maximum likelihood estimators verify the following system

$$\begin{cases} -\frac{n\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^{n} (x_i)^{\beta} = 0 \\ \frac{n}{\beta} - n\ln(\eta) + \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \ln(x_i)(\frac{x_i}{\eta})^{\beta} = 0 \end{cases},$$

which can be solved numerically (with algorithm initialized by the previous estimators).

5.11.4 Random generation

Using the inversion function method, we simply need to compute $\beta(-\log(1-U))^{\frac{1}{\eta}}$ for the first parametrization or $\left(\frac{-\log(1-U)}{\tau}\right)^{\frac{1}{\lambda}}$ for the second one where U is an uniform variate.

5.11.5 Applications

The Weibull was created by Weibull when he studied machine reliability.

NEED REFERENCE

5.12 Inverse Weibull distribution

5.12.1 Characterization

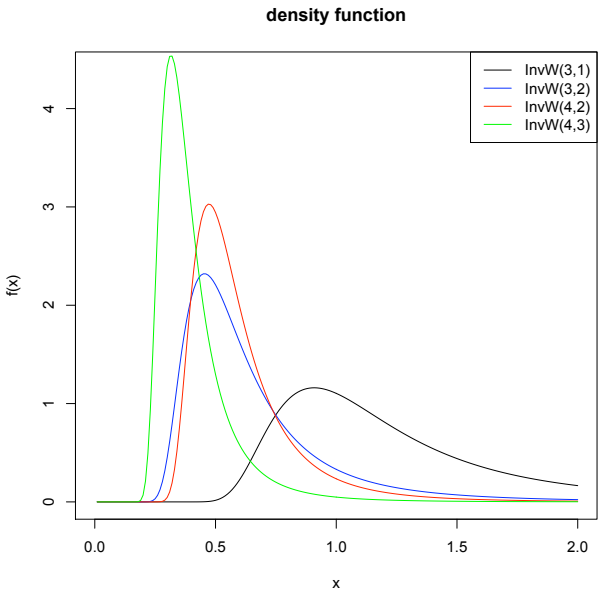
The inverse Weibull distribution is defined as

$$f(x) = \frac{\eta\beta^{\eta} e^{-(\frac{\beta}{x})^{\eta}}}{x^{\eta+1}},$$

where $x > 0$ and $\eta, \beta > 0$. Its distribution function is

$$F(x) = e^{-(\frac{\beta}{x})^{\eta}}.$$

This is the distribution of $1/X$ when X is Weibull distributed $\mathcal{W}(\beta^{-1}, \eta)$.



5.12.2 Properties

The expectation is given by $\eta\Gamma(1-\frac{1}{\beta})$ and the variance $\eta^2[\Gamma(\frac{\beta-2}{\beta}) - \Gamma(\frac{\beta-1}{\beta})^2]$.

The rth moment of the Weibull distribution $\mathcal{IW}(\eta,\beta)$ is given by $\eta^r\Gamma(1-\frac{r}{\beta})$ for $r>0$.

5.12.3 Estimation

Maximum likelihood estimators for β and η verify the following system

$$\begin{cases} \frac{1}{\beta} = \frac{1}{n}\sum_{i=1}^n \left(\frac{\beta}{X_i}\right)^{\eta-1} \\ \frac{1}{\eta} + \log(\beta) = \frac{1}{n}\sum_{i=1}^n \left(\frac{\beta}{X_i}\right)^{\eta} \log\left(\frac{\beta}{X_i}\right) + \frac{1}{n}\sum_{i=1}^n \log(X_i) \end{cases},$$

while the method of moment has the following system

$$\begin{cases} (S_n^2 + (\bar{X}_n)^2)\Gamma^2(1-\frac{1}{\beta}) = (\bar{X}_n)^2\Gamma(1-\frac{2}{\beta}) \\ \eta = \frac{\bar{X}_n}{\Gamma(1-\frac{1}{\beta})} \end{cases}.$$

Both are to solve numerically.

5.12.4 Random generation

Simply generate a Weibull variable $\mathcal{W}(\beta^{-1},\eta)$ and inverse it.

5.12.5 Applications

NEED REFERENCE

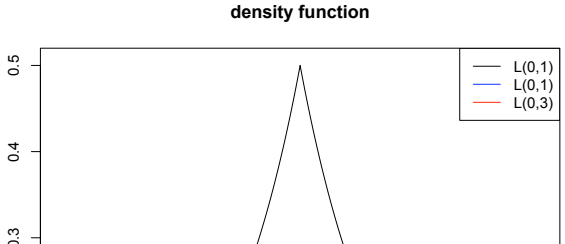
TODO Carrasco et al. (2008)

5.13 Laplace or double exponential distribution

5.13.1 Characterization

Density for the Laplace distribution is given by

$$f(x) = \frac{1}{2\sigma^2} e^{-\frac{|x-m|}{\sigma}},$$



for $x \in \mathbb{R}$, m the location parameter and $\sigma > 0$ the scale parameter. We have the following distribution function

$$F(x) = \begin{cases} \frac{1}{2}e^{-\frac{m-x}{\sigma}} & \text{if } x < m \\ 1 - \frac{1}{2}e^{-\frac{x-m}{\sigma}} & \text{otherwise} \end{cases}.$$

There exists a moment generating function for this distribution, which is

$$M(t) = \frac{e^{mt}}{1 - \sigma^2 t^2},$$

for $|t| < \frac{1}{\sigma}$. The characteristic function is expressed as

$$\phi(t) = \frac{e^{imt}}{1 + \sigma^2 t^2},$$

for $t \in \mathbb{R}$.

5.13.2 Properties

The expectation for the Laplace distribution is given by $E(X) = m$ while the variance is $Var(X) = 2\sigma^2$.

5.13.3 Estimation

Maximum likelihood estimators for m and σ are

$$\hat{m} = \begin{cases} \frac{X_{\frac{n}{2}:n} + X_{\frac{n+2}{2}:n}}{2} & \text{if } n \text{ is even} \\ X_{\lfloor \frac{n}{2} \rfloor :n} & \text{otherwise} \end{cases},$$

where $X_{k:n}$ denotes the kth order statistics and

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} |X_i - \hat{m}|.$$

5.13.4 Random generation

Let U be a uniform variate. Then the algorithm is

- $V = U - 1/2$
- $X = m + \sigma \text{sign}(V) \log(1 - 2|V|)$
- return X

5.13.5 Applications

NEED

The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator Robert M. Norton The American Statistician, Vol. 38, No. 2 (May, 1984), pp. 135-136

Chapter 6

Chi-squared's ditribution and related extensions

6.1 Chi-squared distribution

6.1.1 Characterization

There are many ways to define the chi-squared distribution. First, we can say a chi-squared distribution is the distribution of the sum

$$\sum_{i=1}^{k} X_i^2,$$

where $(X_i)_i$ are i.i.d. normally distributed $\mathcal{N}(0,1)$ and a given k. In this context, k is assumed to be an integer.

We can also define the chi-squared distribution by its density, which is

$$f(x) = \frac{x^{\frac{k}{2}-1}}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} e^{-\frac{x}{2}},$$

where k is the so-called degrees of freedom and $x \geq 0$. One can notice that is the density of a gamma distribution $\mathcal{G}(\frac{k}{2}, \frac{1}{2})$, so k is not necessarily an integer. Thus the distribution function can be expressed with the incomplete gamma function

$$F(x) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})}.$$

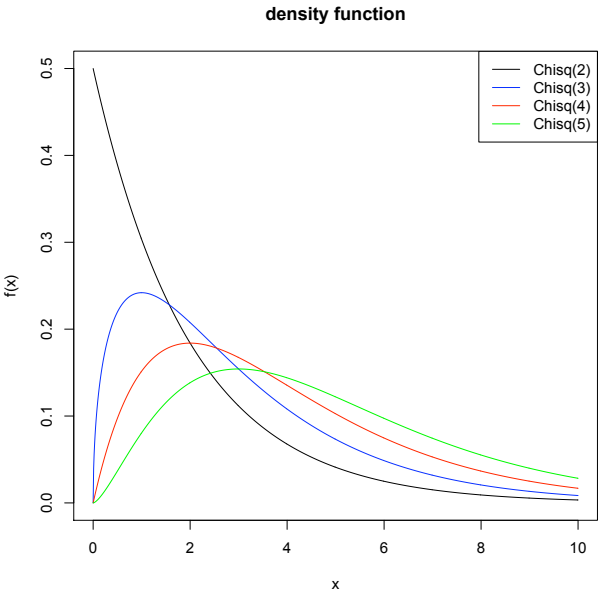


Figure 6.1: Density function for chi-squared distributions

Thirdly, the chi-squared distribution can be defined in terms of its moment generating function

$$M(t) = (1 - 2t)^{-\frac{k}{2}},$$

or its characteristic function

$$\phi(t) = (1 - 2it)^{-\frac{k}{2}}.$$

6.1.2 Properties

The expectation and the variance of the chi-squared distribution are simply $E(X) = k$ and $Var(X) = 2k$. Raw moments are given by

$$E(X^r) = \left(\frac{1}{2}\right)^r \frac{\Gamma(\frac{k}{2} + r)}{\Gamma(\frac{k}{2})}.$$

6.1.3 Estimation

Same as gamma distribution ??

6.1.4 Random generation

For an integer k, just sum the square of k normal variable. Otherwise use the algorithm for the gamma distribution.

6.1.5 Applications

The chi-squared distribution is widely used for inference, typically as pivotal function.

6.2 Chi distribution

6.2.1 Characterization

This is the distribution of the sum

$$\sqrt{\sum_{i=1}^{k} X_i^2},$$

where $(X_i)_i$ are i.i.d. normally distributed $\mathcal{N}(0,1)$ and a given k. This is equivalent as the distribution of a square root of a chi-squared distribution (hence the name).

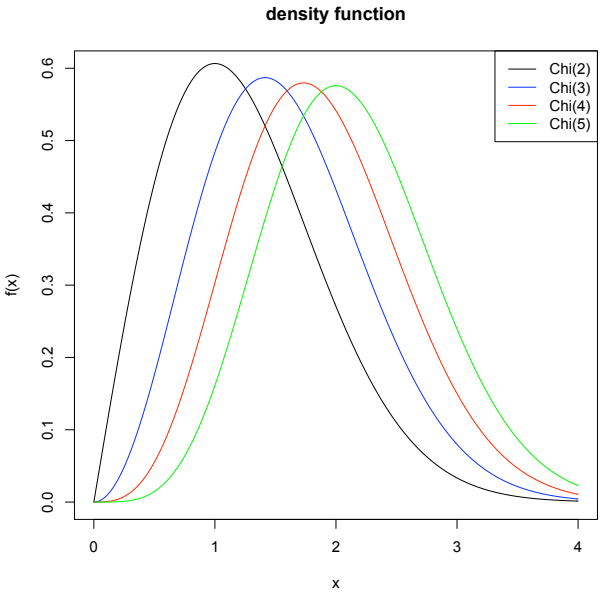


Figure 6.2: Density function for chi distributions

The density function has a closed form

$$f(x) = \frac{x^{k-1}}{2^{\frac{k}{2}-1}\Gamma\left(\frac{k}{2}\right)} e^{-\frac{x^2}{2}},$$

where $x > 0$. The distribution function can be expressed in terms of the gamma incomplete function

$$F(x) = \frac{\gamma(\frac{k}{2}, \frac{x^2}{2})}{\Gamma\left(\frac{k}{2}\right)},$$

for $x > 0$.

Characteristic function and moment generating function exist and are expressed by

$$\phi(t) = {}_1F_1\left(\frac{k}{2}, \frac{1}{2}, \frac{-t^2}{2}\right) + it\sqrt{2}\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$

and

$$M(t) = {}_1F_1\left(\frac{k}{2}, \frac{1}{2}, \frac{t^2}{2}\right) + t\sqrt{2}\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}.$$

6.2.2 Properties

The expectation and the variance of a chi distribution are given by $E(X) = \frac{\sqrt{2}\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}$ and $Var(X) = k - E^2(X)$. Other moments are given by

$$E(X^r) = 2^{\frac{r}{2}}\frac{\Gamma(\frac{k+r}{2})}{\Gamma(\frac{k}{2})},$$

for $k + r > 0$.

6.2.3 Estimation

The maximum likelihood estimator of k satisfies the following equation

$$\frac{1}{2}\psi\left(\frac{k}{2}\right) + \frac{\log(2)}{2} = \frac{1}{n}\sum_{i=1}^{n}\log(X_i),$$

where ψ denotes the digamma function. This equation can be solved on the positive real line or just the set of positive integers.

6.2.4 Random generation

Take the square root of a chi-squared random variable.

6.2.5 Applications

NEED REFERENCE

6.3 Non central chi-squared distribution

6.3.1 Characterization

The non central chi-squared distribution is the distribution of the sum

$$\sum_{i=1}^{k} X_i^2,$$

where $(X_i)_i$ are independent normally distributed $\mathcal{N}(\mu_i, 1)$, i.e. non centered normal random variable. We generally define the non central chi-squared distribution by the density

$$f(x) = \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{k-2}{4}} e^{-\frac{x+\lambda}{2}} I_{\frac{k}{2}-1}\left(\sqrt{\lambda x}\right),$$

for $x > 0$, $k \geq 2$ the degree of freedom, λ the non central parameter and I_λ the Bessel's modified function. λ is related to the previous sum by

$$\lambda = \sum_{i=1}^{k} \mu_i^2.$$

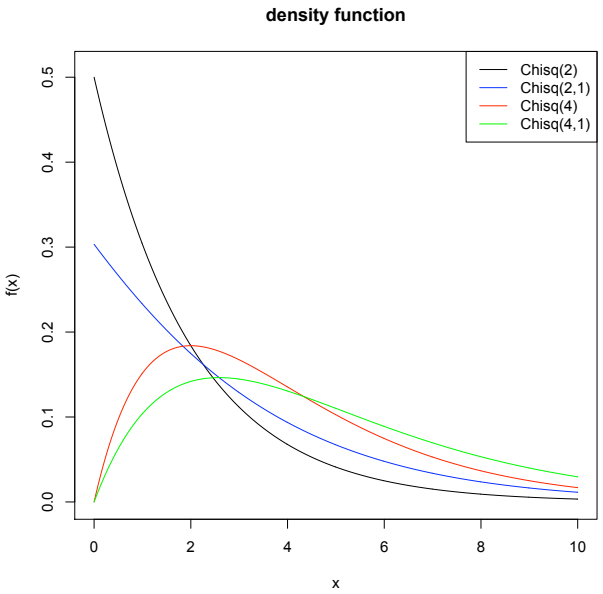


Figure 6.3: Density function for non central chi-squared distributions

The distribution function can be expressed
in terms of a serie

$$F(x) = \sum_{j=0}^{+\infty} e^{\frac{-\lambda}{2}} \frac{(\frac{\lambda}{2})^j \gamma(j + \frac{k}{2}, \frac{x}{2})}{j!\Gamma(j + \frac{k}{2})},$$

for $x > 0$ where $\gamma(.,.)$ denotes the incomplete gamma function.

Moment generating function for the non central chi-squared distribution exists

$$M(t) = \frac{e^{\frac{\lambda t}{1-2t}}}{(1-2t)^{\frac{k}{2}}}$$

and the characteristic function

$$\phi(t) = \frac{e^{\frac{\lambda it}{1-2it}}}{(1-2it)^{\frac{k}{2}}},$$

from which we see it is a convolution of a gamma distribution and a compound Poisson distribution.

6.3.2 Properties

Moments for the non central chi-squared distribution are given by

$$E(X^n) = 2^{n-1}(n-1)!(k+n\lambda) + \sum_{j=1}^{n-1} \frac{(n-1)!2^{j-1}}{(n-j)!}(k+j\lambda)E(X^{n-j}),$$

where the first raw moment is

$$E(X) = k + \lambda.$$

The variance is $Var(X) = 2(k + 2\lambda)$.

6.3.3 Estimation

Li & Yu (2008) and Saxena & Alam (1982)

6.3.4 Random generation

For integer k degrees of freedom, we can use the definition of the sum, i.e. sum k idependent normal random variables $\mathcal{N}(\sqrt{\frac{\lambda}{k}}, 1)$.

6.3.5 Applications

NEED REFERENCE

6.4 Non central chi distribution

6.4.1 Characterization

This is the distribution of the sum

$$\sqrt{\sum_{i=1}^{k} X_i^2},$$

where $(X_i)_i$ are i.i.d. normally distributed $\mathcal{N}(\mu_i, 1)$ and a given k. This is equivalent as the distribution of a square root of a non central chi-squared distribution (hence the name).

We generally define the non central chi distribution by

$$f(x) = \frac{\lambda x^k}{(\lambda x)^{\frac{k}{2}}} e^{-\frac{x^2+\lambda^2}{2}} I_{\frac{k}{2}-1}(\lambda x),$$

where $x > 0$ and $I_{\cdot}(.)$ denotes the modified Bessel's function. The distribution function can be expressed in terms of the gamma incomplete function

$$F(x) = ??,$$

for $x > 0$.

6.4.2 Properties

The expectation and the variance of a chi distribution are given by

$$E(X) = \sqrt{\frac{\pi}{2}} L_{1/2}^{(k/2-1)} \left(\frac{-\lambda^2}{2} \right)$$

and

$$Var(X) = k + \lambda^2 - E^2(X),$$

where $L_{\cdot}^{(\cdot)}$ denotes the generalized Laguerre polynomials. Other moments are given by

$$E(X^r) = ??,$$

for $k + r > 0$.

6.4.3 Estimation

NEED REFERENCE

6.4.4 Random generation

NEED REFERENCE

6.4.5 Applications

NEED REFERENCE

6.5 Inverse chi-squared distribution

The inverse chi-squared distribution is simply the distribution of $\frac{1}{X}$ when X is chi-squared distributed. We can also define the chi-squared distribution by its density, which is

$$f(x) = \frac{2^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})} x^{-\frac{k-2}{2}} e^{-\frac{1}{2x}},$$

where k is the so-called degrees of freedom and $x \geq 0$. Thus the distribution function can be expressed with the incomplete gamma function

$$F(x) = \frac{\Gamma(\frac{k}{2}, \frac{1}{2x})}{\Gamma(\frac{k}{2})},$$

where $\Gamma(.,.)$ the upper incomplete gamma function.

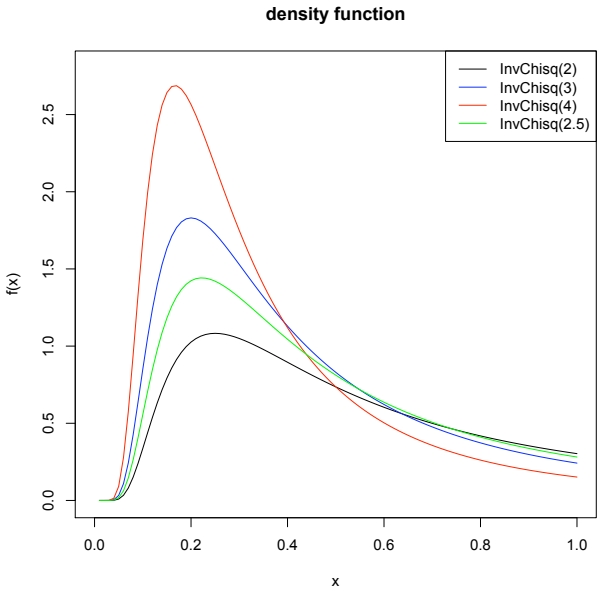


Figure 6.4: Density function for inverse chi-squared distributions

Thirdly, the chi-squared distribution can be defined in terms of its moment generating function

$$M(t) = \frac{2}{\Gamma(\frac{k}{2})} \left(\frac{-t}{2}\right)^{\frac{k}{4}} K_{\frac{k}{2}}\left(\sqrt{-2t}\right),$$

or its characteristic function

$$\phi(t) = \frac{2}{\Gamma(\frac{k}{2})} \left(\frac{-it}{2}\right)^{\frac{k}{4}} K_{\frac{k}{2}}\left(\sqrt{-2it}\right).$$

6.5.1 Properties

The expectation and the variance of the chi-squared distribution are simply $E(X) = \frac{1}{k-2}$ if $k > 2$ and $Var(X) = \frac{2}{(k-2)^2(k-4)}$. Raw moments are given by

$$E(X^r) = ??$$

6.5.2 Estimation

Maximum likelihood estimator for k verifies the equation

$$\psi\left(\frac{k}{2}\right) = -\log(2) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i),$$

where ψ denotes the digamma function.

6.5.3 Random generation

Simply inverse a chi-squared random variable

6.5.4 Applications

NEED REFERENCE

6.6 Scaled inverse chi-squared distribution

6.6.1 Characterization

TODO

6.6.2 Properties

TODO

6.6.3 Estimation

TODO

6.6.4 Random generation

TODO

6.6.5 Applications

TODO

Chapter 7

Student and related distributions

7.1 Student t distribution

Intro?

7.1.1 Characterization

There are many ways to define the student distribution. One can say that it is the distribution of

$$\frac{\sqrt{d}N}{C},$$

where N is a standard normal variable independent of C a chi-squared variable with d degrees of freedom. We can derive the following density function

$$f(x) = \frac{\Gamma(\frac{d+1}{2})}{\sqrt{\pi d}\Gamma(\frac{d}{2})}\left(1+\frac{x^2}{d}\right)^{-\frac{d+1}{2}},$$

for $x \in \mathbb{R}$. d is not necessarily an integer, it could be a real but greater than 1.

The distribution function of the student t distribution is given by

$$F(x) = \frac{1}{2} + x\Gamma\left(\frac{d+1}{2}\right)\frac{{}_2F_1\left(\frac{1}{2},\frac{d+1}{2};\frac{3}{2};-\frac{x^2}{d}\right)}{\sqrt{\pi\nu}\,\Gamma(\frac{d}{2})},$$

where ${}_2F_1$ denotes the hypergeometric function.

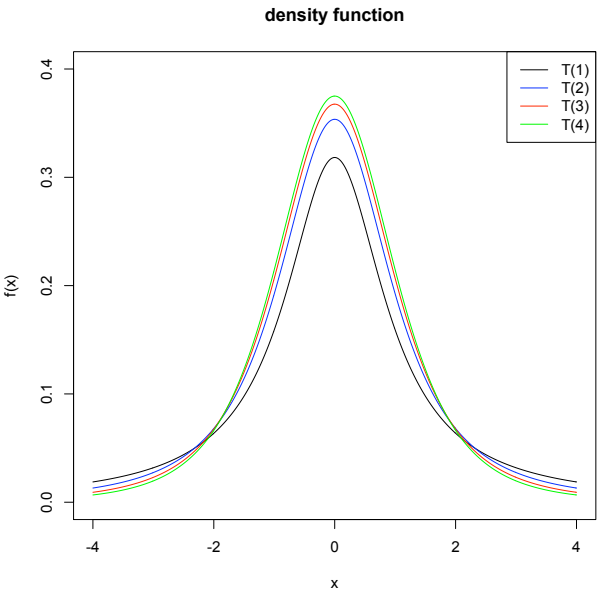


Figure 7.1: Density function for student distributions

7.1.2 Properties

The expectation of a student distribution is $E(X) = 0$ if $d > 1$, infinite otherwise. And the variance is given by $Var(X) = \frac{d}{d-2}$ if $d > 2$.

Moments are given by

$$E(X^r) = \prod_{i=1}^{r/2} \frac{2i-1}{\nu - 2i} \nu^{r/2},$$

where r is an even integer.

7.1.3 Estimation

Maximum likelihood estimator for d can be found by solving numerically this equation

$$\psi\left(\frac{d+1}{2}\right) - \psi\left(\frac{d}{2}\right) = \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + \frac{X_i^2}{d}\right) - \frac{d+1}{n}\sum_{i=1}^{n} \frac{(X_i/d)^2}{1 + X_i^2/d},$$

where ψ denotes the digamma function.

7.1.4 Random generation

The algorithm is simply

- generate a standard normal distribution N
- generate a chi-squared distribution C
- return $\frac{\sqrt{d}N}{\sqrt{C}}$.

7.1.5 Applications

The main application of the student is when dealing with a normally distributed sample, the derivation of the confidence interval for the standard deviation use the student distribution. Indeed for a normally distributed $\mathcal{N}(m, \sigma^2)$ sample of size n we have that

$$\frac{\bar{X}_n - m}{\sqrt{S_n^2}}\sqrt{n}$$

follows a student n distribution.

7.2 Cauchy distribution

7.2.1 Characterization

7.2.2 Characterization

The Cauchy distribution is a special case of the Student distribution when a degree of freedom of 1. Therefore the density function is

$$f(x) = \frac{1}{\pi(1+x^2)},$$

where $x \in \mathbb{R}$. Its distribution function is

$$F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}.$$

There exists a scaled and shifted version of the Cauchy distribution coming from the scaled and shifted version of the student distribution. The density is

$$f(x) = \frac{\gamma^2}{\pi\left[\gamma^2 + (x-\delta)^2\right]},$$

while its distribution function is

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x-\delta}{\gamma}\right) + \frac{1}{2}.$$

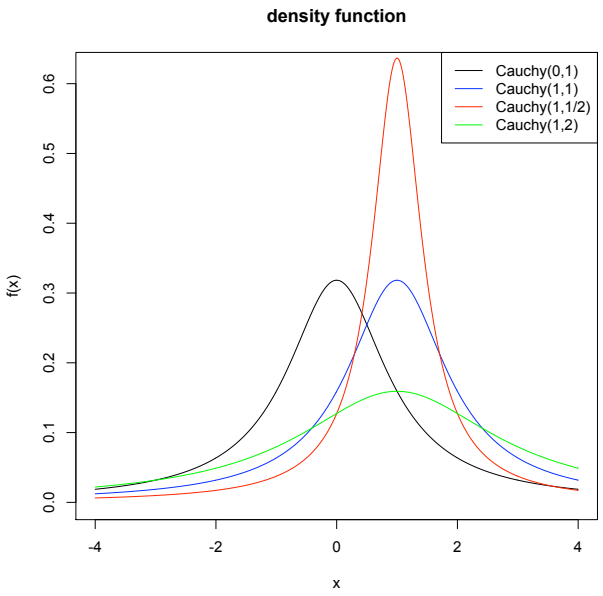


Figure 7.2: Density function for Cauchy distributions

Even if there is no moment generating function, the Cauchy distribution has a characteristic function

$$\phi(t) = \exp(\delta\, i\, t - \gamma\,|t|).$$

7.2.3 Properties

The Cauchy distribution $\mathcal{C}(\delta, \gamma)$ has the horrible feature not to have any finite moments. However, the Cauchy distribution belongs to the family of stable distribution, thus a sum of Cauchy distribution is still a Cauchy distribution.

7.2.4 Estimation

Maximum likelihood estimators verify the following system

$$\begin{cases} \frac{1}{\gamma} = \frac{1}{n}\sum_{i=1}^{n} \frac{\gamma}{\gamma^2+(X_i-\delta)^2} \\ \sum_{i=1}^{n} \frac{X_i}{\gamma^2+(X_i-\delta)^2} = \sum_{i=1}^{n} \frac{\delta}{\gamma^2+(X_i-\delta)^2} \end{cases}.$$

There is no moment based estimators.

7.2.5 Random generation

Since the quantile function is $F^{-1}(u) = \delta + \gamma \tan((u - 1/2)\pi)$, we can use the inversion function method.

7.2.6 Applications

NEED REFERENCE

7.3 Fisher-Snedecor distribution

7.3.1 Characterization

TODO

7.3.2 Properties

TODO

7.3.3 Estimation

TODO

7.3.4 Random generation

TODO

7.3.5 Applications

TODO

Chapter 8

Pareto family

8.1 Pareto distribution

name??

8.1.1 Characterization

The Pareto is widely used by statistician across the world, but many parametrizations of the Pareto distribution are used. Typically two different generalized Pareto distribution are used in extrem value theory with the work of Pickands et al. and in loss models by Klugman et al. To have a clear view on Pareto distributions, we use the work of Arnold (1983). Most of the time, Pareto distributions are defined in terms of their survival function $\bar{F}$, thus we omit the distribution function. In the following, we will define Pareto type I, II, III and IV plus the Pareto-Feller distributions.

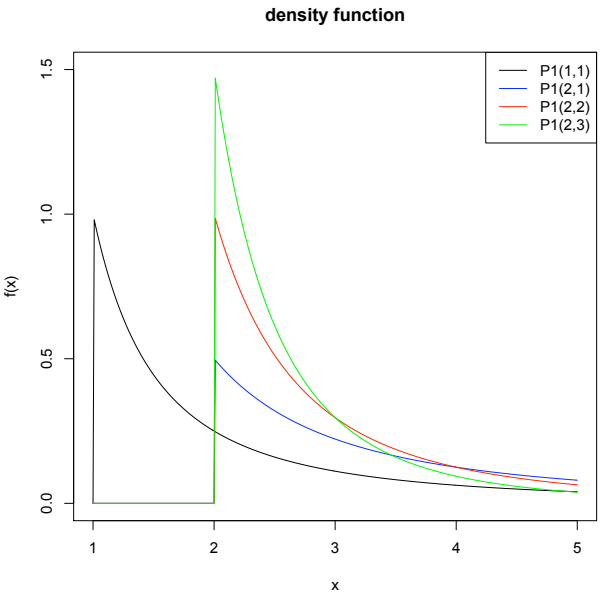


Figure 8.1: Density function for Pareto I distributions

Pareto I

The Pareto type I distribution $\mathcal{P}a_I(\sigma, \alpha)$ is defined by the following survival function

$$\bar{F}(x) = \left(\frac{x}{\sigma}\right)^{-\alpha},$$

where $x > \sigma$ and $\alpha > 0$. Therefore, its density is

$$f(x) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{-\alpha-1},$$

still for $x > \sigma$. α is the positive slope parameter* (sometimes called the Pareto's index) and σ is the scale parameter. Pareto type I distribution is sometimes called the classical Pareto distribution or the European Pareto distribution.

Pareto II

The Pareto type II distribution $\mathcal{P}a_{II}(\mu, \sigma, \alpha)$ is characterized by this survival function

$$\bar{F}(x) = \left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha},$$

where $x > \mu$ and $\sigma, \alpha > 0$. Again α is the shape parameter, while μ is the location parameter. We can derive the density from this definition:

$$f(x) = \frac{\alpha}{\sigma}\left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha-1},$$

for $x > \mu$. We retrieve the Pareto I distribution with $\mu = \sigma$, i.e. if X follows a Pareto I distribution then $\mu - \sigma + X$ follows a Pareto II distribution. The Pareto II is sometimes called the American Pareto distribution.

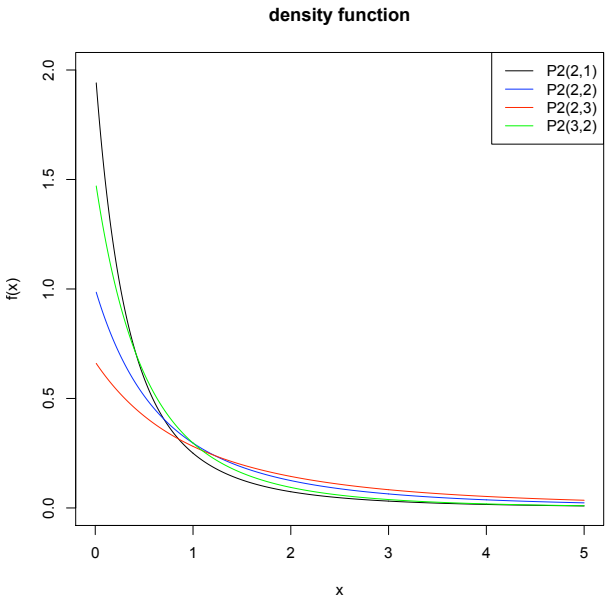


Figure 8.2: Density function for Pareto II distributions

Pareto III

A similar distribution to the type II distribution is the Pareto type III $\mathcal{P}a_{III}(\mu, \sigma, \gamma)$ distribution defined as

$$\bar{F}(x) = \left(1 + \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right)^{-1},$$

where $x > \mu$, $\gamma, \sigma > 0$. The γ parameter is called the index of inequality, and in the special case of $\mu = 0$, it is the Gini index of inequality. The density function is given by

$$f(x) = \frac{1}{\gamma\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}-1}\left(1 + \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right)^{-2},$$

where $x > \mu$. The Pareto III is not a generalisation of the Pareto II distribution, but from these two distribution we can derive more general models. It can be seen as the following transformation $\mu + \sigma Z^{\gamma}$, where Z is a Pareto II $\mathcal{P}a_{II}(0, 1, 1)$.

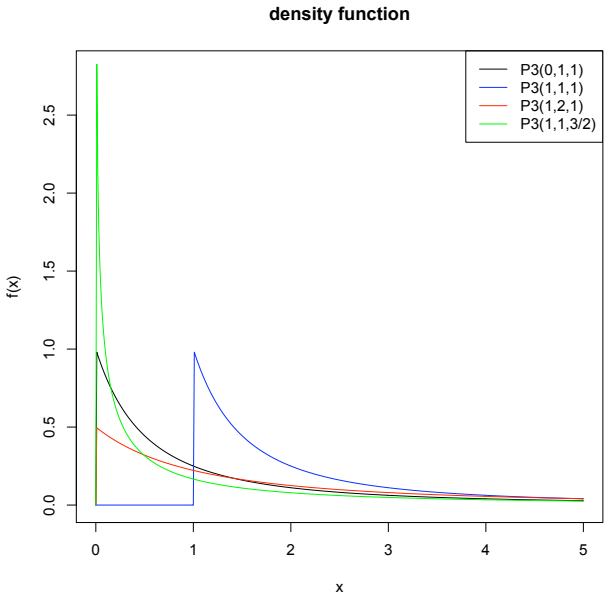


Figure 8.3: Density function for Pareto III distributions

*the slope of the Pareto chart $\log \bar{F}(x)$ vs. $\log x$, controlling the shape of the distribution.

Pareto IV

The Pareto type IV $\mathcal{P}a_{IV}(\mu, \sigma, \gamma, \alpha)$ distribution is defined by

$$\bar{F}(x) = \left(1 + \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right)^{-\alpha},$$

where $x > \mu$ and $\alpha, \sigma, \gamma > 0$. The associated density function is expressed as follows

$$f(x) = \frac{\alpha}{\gamma\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}-1}\left(1 + \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right)^{-\alpha-1}$$

for $x > \mu$.

Quantile functions for Pareto distributions are listed in sub-section random generation.

The generalized Pareto used in extreme value theory due to Pickands (1975) has a limiting distribution with Pareto II $\mathcal{P}a_{II}(0, \sigma, \alpha)$, see chapter on EVT for details. Finally, the Feller-Pareto is a generalisation of the Pareto IV distribution, cf. next section.

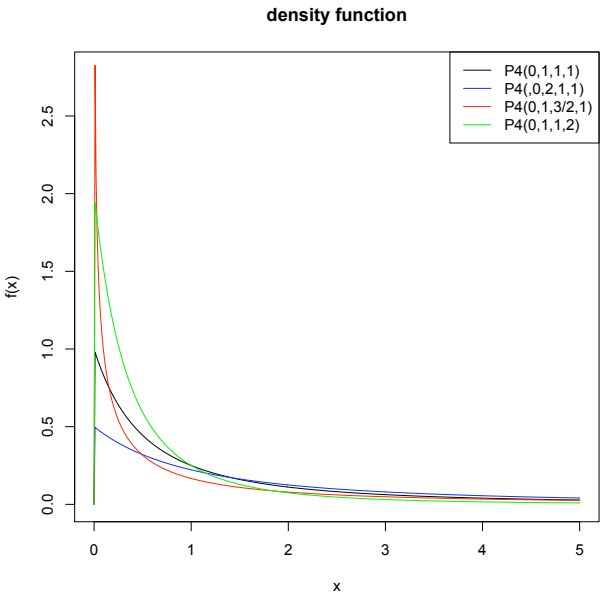


Figure 8.4: Density function for Pareto IV distributions

8.1.2 Properties

Equivalence

It is easy to verify that if X follows a Pareto I distribution $\mathcal{P}a_I(\sigma, \alpha)$, then $\log X$ follows a translated exponential distribution $\mathcal{TE}(\sigma, \alpha?)$.

The Pareto type III distribution is sometimes called the log-logistic distribution, since if X has a logistic distribution then e^X has a Pareto type III distribution with $\mu = 0$.

Moments

Moments for the Pareto I distribution are given by $E(X) = \frac{\alpha\sigma}{\alpha-1}$ if $\alpha > 1$, $Var(X) = \frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}$ and $E(X^\tau) = \tau\frac{\alpha}{\alpha-\tau}$ for $\alpha > \tau$ and $\sigma = 1$.

Moments for the Pareto II, III can be derived from those of Pareto IV distribution, which are

$$E(X^\tau) = \sigma^\tau \frac{\Gamma(1+\tau\gamma)\Gamma(\alpha-\tau\gamma)}{\Gamma(\alpha)},$$

with $-1 < \tau\gamma < \alpha$ and $\mu = 0$.

Convolution and sum

The convolution (i.e. sum) of Pareto I distributions does not have any particular form but the product of Pareto I distributions does have a analytical form.

If we consider of n i.i.d. Pareto I $\mathcal{P}a_I(\sigma, \alpha)$ random variables, then the product Π has the following density

$$f_\Pi(x) = \frac{\alpha \left(\sigma \log(\frac{x-\sigma}{\sigma})\right)^{n-1} \left(\frac{x}{\sigma}\right)^{-\alpha}}{x\Gamma(n)},$$

where $x > \sigma$.

If we consider only independent Pareto I distribution $\mathcal{P}a_I(\sigma_i, \alpha_i)$, then we have for the density of the product

$$f_\Pi(x) = \sum_{i=1}^{n} \frac{\alpha_i}{\sigma} \left(\frac{x}{\sigma}\right)^{-\alpha_i - 1} \prod_{k \neq i} \frac{\alpha_k}{\alpha_i - \alpha_k},$$

where $x > \prod_{i=1}^{n} \sigma_i$.

Other Pareto distributions??

Order statistics

Let $(X_i)_i$ be a sample of Pareto distributions. We denote by $(X_{i:n})_i$ the associated order statistics, i.e. $X_{1:n}$ is the minimum and $X_{n:n}$ the maximum.

For Pareto I distribution, the ith order statistic has the following survival function

$$\bar{F}_{X_{i:n}}(x) = \sum_{j=1}^{i} \left(1 + \frac{x}{\sigma}\right)^{-\alpha(n-j+1)} \prod_{\substack{l=1 \\ l \neq i}}^{i} \frac{n-l+1}{l-j},$$

where $x > 0$. Furthermore moments are given by

$$E(X_{i:n}^{\tau}) = \sigma^{\tau} \frac{n!}{(n-i)!} \frac{\Gamma(n-i+1-\tau\alpha^{-1})}{\Gamma(n+1-\tau\alpha^{-1})},$$

for $\tau \in \mathbb{R}$.

For Pareto II distribution, we get

$$\bar{F}_{X_{i:n}}(x) = \sum_{j=1}^{i} \left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha(n-j+1)} \prod_{\substack{l=1 \\ l \neq i}}^{i} \frac{n-l+1}{l-j},$$

where $x > \mu$. Moments can be derived from those in the case of the Pareto I distribution using the fact $X_{i:n} = \mu - \sigma + Y_{i:n}$ with $Y_{i:n}$ order statistic for the Pareto I case.

For Pareto III distribution, the ith order statistic follows a Feller-Pareto $\mathcal{FP}a(\mu, \sigma, \gamma, i, n-i+1)$. Moments of order statistics can be obtained by using the transformation of Pareto II random

variable: we have $X_{i:n} = \mu + \sigma Z_{i:n}^{\gamma}$ follows a Pareto III distribution, where Z is a Pareto II $\mathcal{P}a_{II}(0,1,1)$. Furthermore, we know the moments of the random variable Z:

$$E(Z_{i:n}^{\tau}) = \frac{\Gamma(i+\tau)\Gamma(n-i+\tau+1)}{\Gamma(i)\Gamma(n-i+1)}$$

The minimum of Pareto IV distributions still follows a Pareto IV distribution. Indeed if we consider n independent random variables Pareto IV $\mathcal{P}a_{IV}(\mu,\sigma,\gamma,\alpha_i)$ distributed, we have

$$\min(X_1,\ldots,X_n) \sim \mathcal{P}a_{IV}\left(\mu,\sigma,\gamma,\sum_{i=1}^{n}\alpha_i\right).$$

But the ith order statistic does not have a particular distribution. The intermediate order statistic can be approximated by the normal distibution with

$$X_{i:n} \underset{n\to+\infty}{\longrightarrow} \mathcal{N}\left(F^{-1}\left(i/n\right), i/n\left(1-i/n\right) f^{-2}\left(F^{-1}\left(i/n\right)\right) n^{-1}\right)$$

where f and F denotes respectively the density and the distribution function of the Pareto IV distribution. Moments for the order statistics are computable from the moments of the minima since we have

$$E(X_{i:n}^{\tau}) = \sum_{r=n-i+1}^{n} (-1)^{r-n+i-1} C_n^r C_{r-1}^{n-i} E(X_{1:r}^{\tau}).$$

Since $X_{1:r}$ still follows a Pareto IV distribution $\mathcal{P}_{IV}(\mu,\sigma,\gamma,r\alpha)$, we have

$$E(X_{1:r}^{\tau}) = E((\mu+\sigma Z_{1:r})^{\tau}),$$

where $Z_{1:r} \sim \mathcal{P}a_{IV}(0,1,\gamma,r\alpha)$ and $E(Z_{1:r}^{\tau}) = \frac{\Gamma(1+\tau\gamma)\Gamma(r\alpha-\tau\gamma)}{\Gamma(r\alpha)}$.

Truncation

Let us denote by $X|X>x_0$ the random variable X knowing that $X>x_0$. We have the following properties (with $x_0>\mu$):

- if $X \sim \mathcal{P}a_I(\sigma,\alpha)$ then $X|X>x_0 \sim \mathcal{P}a_I(x_0,\alpha)$*
- if $X \sim \mathcal{P}a_{II}(\mu,\sigma,\alpha)$ then $X|X>x_0 \sim \mathcal{P}a_I(x_0,\sigma+x_0-\mu,\alpha)$

More general distributions do not have any particular form.

*In this case, the truncation is a rescaling. It comes from the lack of memory property of the log variable since the log variable follows an exponential distribution.

Record values

Geometric minimization

8.1.3 Estimation

Estimation of the Pareto distribution in the context of actuarial science can be found in Rytgaard (1990).

Pareto I

Arnold (1983) notices that from a log transformation, the parameter estimation reduces to a problem for a translated exponentiallly distributed data. From this, we have the following maximum likelihood estimator for the Pareto I distribution

- $\hat{\alpha}_n = X_{1:n}$,
- $\hat{\sigma}_n = \left[\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{X_i}{X_{1:n}}\right)\right]^{-1}$,

where $(X_i)_{1\leq i\leq n}$ denotes a sample of i.i.d. Pareto variables. Those estimators are strongly consistent estimator of α and σ. Let us note that for these estimator we have better than the asymptotic normality (due to the maximum likelihoodness). The distributions for these two estimators are respectively Pareto I and Gamma distribution:

- $\hat{\alpha}_n \sim \mathcal{P}_I(\sigma, n\alpha)$,
- $\hat{\sigma}_n^{-1} \sim \mathcal{G}(n-1, (\alpha n)^{-1})$.

From this, we can see these estimators are biased, but we can derive unbiased estimators with minimum variance:

- $\tilde{\alpha}_n = \frac{n-2}{n}\hat{\alpha}_n$,
- $\tilde{\sigma}_n = \left[1 - \frac{1}{\hat{\alpha}_n}\right]\hat{\sigma}_n$.

Since those statistics $\tilde{\alpha}_n$ and $\tilde{\sigma}_n$ are sufficient, it is easy to find unbiased estimators of functions of these parameters $h(\alpha, \sigma)$ by plugging in $\tilde{\alpha}_n$ and $\tilde{\sigma}_n$ (i.e. $h(\tilde{\alpha}_n, \tilde{\sigma}_n)$).

However other estimations are possible, for instance we may use a least square regression on the Pareto chart (plot of $\log \bar{F}(x)$ against $\log x$). We can also estimate parameters by the method of moments by equalling the sample mean and minimum to corresponding theoretical moments. We get

- $\hat{\alpha}_n^M = \frac{n\bar{X}_n - X_{1:n}}{n(\bar{X}_n - X_{1:n})}$,
- $\hat{\sigma}_n^M = \frac{n\hat{\alpha}_n^M - 1}{n\hat{\alpha}_n^M} X_{1:n}$,

where we assume a finite expectation (i.e. $\alpha > 1$).

Finally, we may also calibrate a Pareto I distribution with a quantile method. We numerically solve the system

$$\begin{cases} p_1 = 1 - \left(\frac{X_{\lfloor np_1 \rfloor:n}}{\sigma}\right)^\alpha \\ p_2 = 1 - \left(\frac{X_{\lfloor np_2 \rfloor:n}}{\sigma}\right)^\alpha \end{cases},$$

for two given probabilities p_1, p_2.

Pareto II-III-IV

Estimation of parameters for Pareto II, III and IV are more difficult. If we write the log-likelihood for a sample $(X_i)_{1 \leq i \leq n}$ Pareto IV distributed, we have

$$\log \mathcal{L}(\mu, \sigma, \gamma, \alpha) = \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^n \log\left(\frac{x_i - \mu}{\sigma}\right) - (\alpha+1) \sum_{i=1}^n \log\left(1 + \left(\frac{x_i - \mu}{\sigma}\right)^{\frac{1}{\gamma}}\right) - n\log\gamma - n\log\sigma + n\log\alpha,$$

with the constraint that $\forall 1 \leq i \leq n, x_i > \mu$. Since the log-likelihood is null when $x_{1:n} \leq \mu$ and a decreasing function of μ otherwise the maximum likelihood estimator of μ is the minimum $\hat{\mu} = X_{1:n}$.

Then if we substract $\hat{\mu}$ to all observations, we get the following the log-likelihood

$$\log \mathcal{L}(\sigma, \gamma, \alpha) = \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^n \log\left(\frac{x_i}{\sigma}\right) - (\alpha+1) \sum_{i=1}^n \log\left(1 + \left(\frac{x_i}{\sigma}\right)^{\frac{1}{\gamma}}\right) - n\log\gamma - n\log\sigma + n\log\alpha,$$

which can be maximised numerically. Since there are no close form for estimators of σ, γ, α, we do not know their distributions, but they are asymptotically normal.

We may also use the method of moments, where again $\hat{\mu}$ is $X_{1:n}$. Substracting this value to all observations, we use the expression of moments above to have three equations. Finally solve the system numerically. A similar scheme can be used to estimate parameters with quantiles.

8.1.4 Random generation

It is very easy to generate Pareto random variate using the inverse function method. Quantiles function can be easily calculated

- for $\mathcal{P}_I(\sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma(1-u)^{\frac{-1}{\alpha}}$,
- for $\mathcal{P}_{II}(\mu, \sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma\left[(1-u)^{\frac{-1}{\alpha}} - 1\right] + \mu$,

- for $\mathcal{P}_{III}(\mu, \sigma, \gamma)$ distribution, $F^{-1}(u) = \sigma \left[(1-u)^{-1} - 1\right]^{\gamma} + \mu$,
- for $\mathcal{P}_{IV}(\mu, \sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma \left[(1-u)^{\frac{-1}{\alpha}} - 1\right]^{\gamma} + \mu$.

Therefore algorithms for random generation are simply

- for $\mathcal{P}_{I}(\sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma U^{\frac{-1}{\alpha}}$,
- for $\mathcal{P}_{II}(\mu, \sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma \left[U^{\frac{-1}{\alpha}} - 1\right] + \mu$,
- for $\mathcal{P}_{III}(\mu, \sigma, \gamma)$ distribution, $F^{-1}(u) = \sigma \left[U^{-1} - 1\right]^{\gamma} + \mu$,
- for $\mathcal{P}_{IV}(\mu, \sigma, \alpha)$ distribution, $F^{-1}(u) = \sigma \left[U^{\frac{-1}{\alpha}} - 1\right]^{\gamma} + \mu$,

where U is an uniform random variate.

8.1.5 Applications

From wikipedia, we get the following possible applications of the Pareto distributions:

- the sizes of human settlements (few cities, many hamlets/villages),
- file size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones),
- clusters of Bose-Einstein condensate near absolute zero,
- the values of oil reserves in oil fields (a few large fields, many small fields),
- the length distribution in jobs assigned supercomputers (a few large ones, many small ones),
- the standardized price returns on individual stocks,
- sizes of sand particles,
- sizes of meteorites,
- numbers of species per genus (There is subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it),
- areas burnt in forest fires,
- severity of large casualty losses for certain lines of business such as general liability, commercial auto, and workers compensation.

In the litterature, Arnold (1983) uses the Pareto distribution to model the income of an individual and Froot & O'Connell (2008) apply the Pareto distribution as the severity distribution in a context of catastrophe reinsurance. Here are just a few applications, many other applications can be listed.

8.2 Feller-Pareto distribution

8.2.1 Characterization

As described in Arnold (1983), the Feller-Pareto distribution is the distribution of

$$X = \mu + \sigma \left(\frac{U}{V}\right)^{\gamma},$$

where U and V are independent gamma variables ($\mathcal{G}(\delta_1, 1)$ and $\mathcal{G}(\delta_2, 1)$ respectively). Let us note that the ratio of these two variables follows a beta distribution of the second kind. In term of distribution function, using the transformation of the beta variable, we get

$$F(x) = \frac{\beta\left(\delta_1, \delta_2, \frac{y}{1+y}\right)}{\beta(\delta_1, \delta_2)} \quad \text{with} \quad y = \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}},$$

with $x \geq \mu$, $\beta(.,.)$ denotes the beta function and $\beta(.,.,.)$ the incomplete beta function.

We have the following density for the Feller-Pareto distribution $\mathcal{FP}(\mu, \sigma, \gamma, \delta_1, \delta_2)$:

$$f(x) = \frac{(\frac{x-\mu}{\sigma})^{\frac{\delta_2}{\gamma}-1}}{\gamma\beta(\delta_1, \delta_2)x(1+(\frac{x-\mu}{\sigma})^{\frac{1}{\gamma}})^{\delta_1+\delta_2}},$$

where $x \geq \mu$. Let y be $\frac{x-\mu}{\sigma}$, the previous expression can be rewritten as

$$f(x) = \frac{1}{\gamma\beta(\delta_1, \delta_2)} \left(\frac{y^{\frac{1}{\gamma}}}{1+y^{\frac{1}{\gamma}}}\right)^{\delta_2} \left(1 - \frac{y^{\frac{1}{\gamma}}}{1+y^{\frac{1}{\gamma}}}\right)^{\delta_1} \frac{1}{xy},$$

for $x \geq \mu$. In this expression, we see more clearly the link with the beta distribution as well as the transformation of the variable $\frac{U}{V}$.

There is a lot of special cases to the Feller-Pareto distribution $\mathcal{FP}(\mu, \sigma, \gamma, \delta_1, \delta_2)$. When $\mu = 0$, we retrieve the transformed beta distribution* of Klugman et al. (2004) and if in addition $\gamma = 1$, we get the "generalized" Pareto distribution† (as defined by Klugman et al. (2004)).

Finally the Pareto IV distribution is obtained with $\delta_1 = 1$. Therefore we have the following equivalences

- $\mathcal{P}_I(\sigma, \alpha) = \mathcal{FP}(\sigma, \sigma, 1, 1, \alpha)$,
- $\mathcal{P}_{II}(\mu, \sigma, \alpha) = \mathcal{FP}(\mu, \sigma, 1, 1, \alpha)$,
- $\mathcal{P}_{III}(\mu, \sigma, \gamma) = \mathcal{FP}(\mu, \sigma, \gamma, 1, 1)$,
- $\mathcal{P}_{IV}(\mu, \sigma, \gamma, \alpha) = \mathcal{FP}(\mu, \sigma, \gamma, 1, \alpha)$.

*sometimes called the generalized beta distribution of the second kind.

†which has nothing to do with the generalized Pareto distribution of the extreme value theory.

8.2.2 Properties

When $\mu = 0$, raw moments are given by

$$E(X^r) = \sigma^r \frac{\Gamma(\delta_1 + r\gamma)\Gamma(\delta_2 - r\gamma)}{\Gamma(\delta_1)\Gamma(\delta_2)},$$

for $-\frac{\delta_1}{\gamma} \leq r \leq \frac{\delta_2}{\gamma}$.

8.2.3 Estimation

NEED REFERENCE

8.2.4 Random generation

Once we have simulated a beta I distribution B, we get a beta II distribution* with $\tilde{B} = \frac{B}{1-B}$. Finally we shift, scale and take the power $X = \mu + \sigma \left(\tilde{B}\right)^{\gamma}$ to get a Feller-Pareto random variable.

8.2.5 Applications

NEED REFERENCE

*We can also use two gamma variables to get the beta II variable.

8.3 Inverse Pareto

8.3.1 Characterization

From the Feller-Pareto distribution, we get the inverse Pareto distribution with $\mu = 0$, $\delta_1 = 1$ and $\gamma = 1$. Thus the density is

$$f(x) = \frac{1}{\beta(1,\delta_2)} \left(\frac{\frac{x}{\sigma}}{1+\frac{x}{\sigma}}\right)^{\delta_2} \frac{1}{1+\frac{x}{\sigma}} \frac{1}{\frac{x}{\sigma}},$$

It can be rewritten as the density

$$f(x) = \frac{\tau\lambda x^{\tau-1}}{(x+\lambda)^{\tau+1}}$$

which implies the following distribution function

$$F(x) = \left(\frac{x}{x+\lambda}\right)^{\tau},$$

for $x \geq 0$. Let us note this is the distribution of $\frac{1}{X}$ when X is Pareto II.

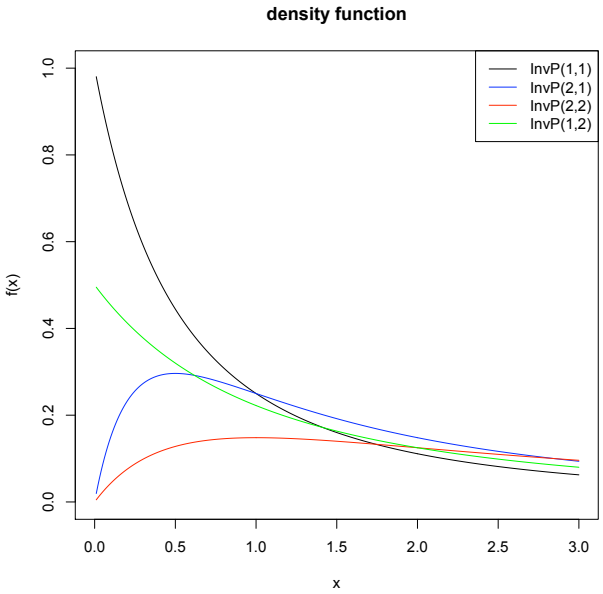


Figure 8.5: Density function for inverse Pareto distributions

8.3.2 Properties

The expectation of the inverse Pareto distribution is $E(X) = \frac{\lambda\Gamma(\tau+1)}{\Gamma(\tau)}$, but the variance does not exist.

8.3.3 Estimation

NEED REFERENCE

8.3.4 Random generation

Simply inverse a Pareto II variable.

8.3.5 Applications

NEED REFERENCE

8.4 Generalized Pareto distribution

8.4.1 Characterization

The generalized Pareto distribution was introduced in Embrechts et al. (1997) in the context of extreme value theory.

We first define the standard generalized Pareto distribution by the following distribution function

$$F(x) = \begin{cases} 1-(1+\xi x)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1-e^{-x} & \text{if } \xi = 0 \end{cases},$$

where $x \in \mathbb{R}_+$ if $\xi \geq 0$ and $x \in \left[0, -\frac{1}{\xi}\right]$ otherwise. This distribution function is generally denoted by G_ξ.

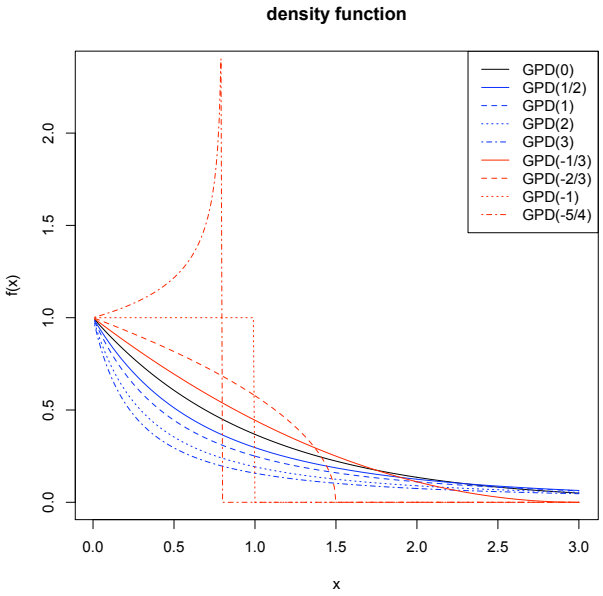


Figure 8.6: Density function for standard generalized Pareto distributions

We can see the impact of the shape parameter ξ on the figure on the right. The case where $\xi = 0$ can be seen as a limiting case of G_ξ when $\xi \to 0$.

To get the "full" generalized Pareto distribution, we introduce a scale β and a location parameter μ. We get

$$F(x) = \begin{cases} 1-\left(1+\xi\frac{x-\nu}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi > 0 \\ 1-e^{-\frac{x-\nu}{\beta}} & \text{if } \xi = 0 \\ 1-\left(1+\xi\frac{x-\nu}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi < 0 \end{cases},$$

where x lies in $[\nu, +\infty[$, $[\nu, +\infty[$ and $\left[\nu, \nu - \frac{\beta}{\xi}\right]$ respectively. We denote it by $G_{\xi,\nu,\beta}(x)$ (which is simply $G_\xi(\frac{x-\nu}{\beta})$). Let us note when $\xi > 0$, we have a Pareto II distribution, when $\xi = 0$ a shifted exponential distribution and when $\xi < 0$ a generalized beta I distribution.

From these expression, we can derive a density function for the generalized Pareto distribution

$$f(x) = \begin{cases} \frac{1}{\beta}\left(1+\xi\frac{x-\nu}{\beta}\right)^{-\frac{1}{\xi}-1} & \text{if } \xi > 0 \\ \frac{1}{\beta}e^{-\frac{x-\nu}{\beta}} & \text{if } \xi = 0 \\ \frac{1}{\beta}\left(1-(-\xi)\frac{x-\nu}{\beta}\right)^{\frac{1}{-\xi}-1} & \text{if } \xi < 0 \end{cases},$$

for x in the same supports as above.

8.4.2 Properties

For a generalized Pareto distribution $G_{\xi,0,\beta}$, we have results on raw moments (for simplicity $\nu = 0$). The expectation $E(X)$ is finite if and only if $\xi < 1$. In this case we have

$$E\left(\left(1+\frac{\xi}{\beta}X\right)^{-r}\right) = \frac{1}{1+\xi r}, \quad \text{for} \quad r > -\frac{1}{\xi}$$

$$E\left(\left(\log\left(1+\frac{\xi}{\beta}X\right)\right)^{k}\right) = \xi^k k!, \quad \text{for} \quad k \in \mathbb{N}$$

$$E\left(X\bar{F}(X)^r\right) = \frac{\beta}{(r+1-\xi)(r+1)}, \quad \text{for} \quad \frac{r+1}{|\xi|} > 0$$

$$E\left(X^k\right) = \frac{\beta^k}{\xi^{k+1}}\frac{\Gamma(\xi^{-1}-k)}{\Gamma(1+\xi^{-1})}k!, \quad \text{for} \quad \xi < \frac{1}{k},$$

see Embrechts et al. (1997) for details.

If X follows a generalized Pareto distribution $GPD(\xi, 0, \beta)$, then the treshold excess random variable $X - u|X > u$ still follows a generalized Pareto distribution $GPD(\xi, 0, \beta + \xi u)$. Let F_u be the distribution function of $X - u|X > u$. We have F is in the maximum domain of attraction H_ξ if and only if

$$\lim_{u \to x_f} \sup_{0<x<x_f-u} \left|F_u(x) - G_{\xi,0,\beta(u)}(x)\right| = 0,$$

where β is a positive function. This makes the link between the generalized Pareto distribution and the generalized extreme value distribution.

8.4.3 Estimation

In this sub-section, we assume $\nu = 0$.

Peak Over a Treshold

We briefly present the Peak Over a Treshold (POT) method to fit the generalized Pareto distribution. Let $(X_i)_{1\leq i\leq n}$ an i.i.d. sample whose distribution function belongs to a maximum domain of attraction H_ξ. For a deterministic treshold $u > 0$, we define the number of exceedances by

$$N_u = \text{Card}(1 \leq i \leq n, X_i > u),$$

with the corresponding excesses $(Y_i)_{1\leq i\leq N_u}$. We want to fit the excess distribution function F_u with the GPD distribution function $G_{\xi,0,\beta(u)}$.

First we can use the linearity of the mean excess function

$$E(X - u|X > u) = \frac{\beta + \xi u}{1 - \xi},$$

for a given u. This can be estimated by the empirical mean of the sample $(Y_i)_{1\leq i\leq N_u}$. Embrechts et al. (1997) warn us about the difficulty of chosing u, since they are many u for wich the plot of $(u, \bar{Y}_{N_u})$.

Once we find the treshold u, we can use conditional likelihood estimation on sample $(Y_i)_{1\leq i\leq N_u}$. Let τ be $-\xi/\beta$. However we can also use a linear regression to fit the shape and the scale parameter.

Maximum likelihood estimation

Maximum likelihood estimators of ξ and β are solutions of the system

$$\left\{ \begin{array}{l} \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\frac{\xi X_i}{\beta^2+\beta\xi X_i} = \frac{n}{\beta} \\ \frac{1}{\xi^2}\sum_{i=1}^{n}\log\left(1+\frac{\xi}{\beta}X_i\right) = (\frac{1}{\xi}+1)\sum_{i=1}^{n}\frac{X_i}{\beta+\xi X_i} \end{array} \right. ,$$

but the system may be instable for $\xi \leq -1/2$. When $\xi > 1/2$, we have some asymptotical properties of maximum likelihood estimators $\hat{\xi}$ and $\hat{\beta}$:

$$\sqrt{n}\left(\hat{\xi}-\xi, \frac{\hat{\beta}}{\beta}-1\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, M^{-1}),$$

where the variance/covariance matrix for the bivariate normal distribution is

$$M^{-1} = (1+\xi)\begin{pmatrix} 1+\xi & 1 \\ 1 & 2 \end{pmatrix}.$$

Let us note that if we estimate a ξ as zero, then we can try to fit a shifted exponential distribution.

Method of moments

From the properties, we know the theoretical expression of $E(X)$ and $E\left(X\bar{F}(X)\right)$. From wich we get the relation

$$\beta = \frac{2E(X)E\left(X\bar{F}(X)\right)}{E(X)-2E\left(X\bar{F}(X)\right)} \quad \text{and} \quad \xi = 2 - \frac{E(X)}{E(X)-2E\left(X\bar{F}(X)\right)}.$$

We simply replace $E(X)$ and $E\left(X\bar{F}(X)\right)$ by the empirical estimators.

8.4.4 Random generation

We have an explicit expression for the quantile function

$$F^{-1}(u) = \left\{ \begin{array}{ll} \nu + \frac{\sigma}{\xi}((1-u)^{-\xi}-1) & \text{if } \xi \neq 0 \\ \nu - \sigma\log(1-u) & \text{if } \xi = 0 \end{array} \right. ,$$

thus we can use the inversion function method to generate GPD variables.

8.4.5 Applications

The main application of the generalized Pareto distribution is the extreme value theory, since there exists a link between the generalized Pareto distribution and the generalized extreme value distribution. Typical applications are modeling flood in hydrology, natural disaster in insurance and asset returns in finance.

8.5 Burr distribution

8.5.1 Characterization

The Burr distribution is defined by the following density

$$f(x) = \frac{\alpha\tau}{\lambda}\frac{(x/\lambda)^{\tau-1}}{(1+(x/\lambda)^{\tau})^{\alpha+1}}$$

where $x \geq 0$, λ the scale parameter and $\alpha, \tau > 0$ the shape parameters. Its distribution function is given by

$$F(x) = 1 - \left(\frac{\lambda^{\tau}}{\lambda^{\tau}+x^{\tau}}\right)^{\alpha},$$

for $x \geq 0$. In a slightly different rewritten form, we recognise the Pareto IV distribution

$$\bar{F}(x) = \left(1+\left(\frac{x}{\lambda}\right)^{\tau}\right)^{-\alpha},$$

with a zero location parameter.

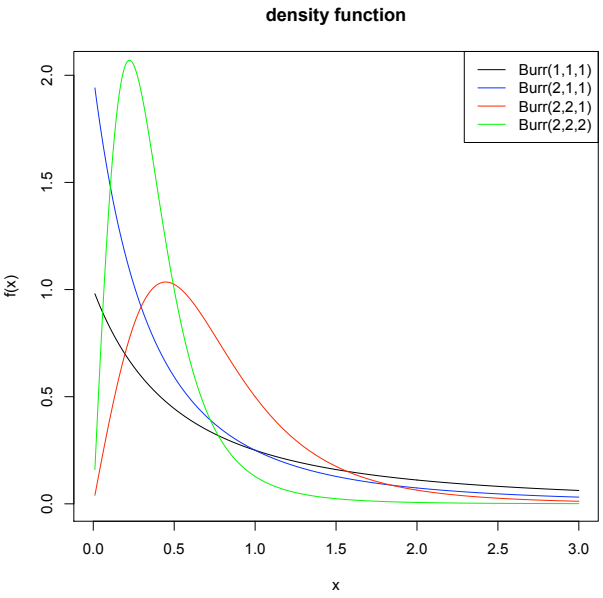


Figure 8.7: Density function for Burr distributions

8.5.2 Properties

The raw moment of the Burr distribution is given by

$$E(X^r) = \lambda^r \frac{\Gamma(1+\frac{r}{\tau})\Gamma(\alpha-\frac{r}{\tau})}{\Gamma(\alpha)},$$

hence the expectation and the variance are

$$E(X) = \lambda\frac{\Gamma(1+\frac{1}{\tau})\Gamma(\alpha-\frac{1}{\tau})}{\Gamma(\alpha)} \quad \text{and} \quad Var(X) = \lambda^2\frac{\Gamma(1+\frac{2}{\tau})\Gamma(\alpha-\frac{2}{\tau})}{\Gamma(\alpha)} - \left(\lambda\frac{\Gamma(1+\frac{1}{\tau})\Gamma(\alpha-\frac{1}{\tau})}{\Gamma(\alpha)}\right)^2.$$

8.5.3 Estimation

Maximum likelihood estimators are solution of the system

$$\begin{cases} \frac{n}{\alpha} = \sum_{i=1}^{n} \log\left(1 + \left(\frac{X_i}{\lambda}\right)^{\tau}\right) \\ \frac{n}{\tau} = -\sum_{i=1}^{n} \log\left(\frac{X_i}{\lambda}\right) + \tau\frac{\alpha+1}{\lambda} \sum_{i=1}^{n} \log\left(\frac{X_i}{\lambda}\right) \frac{X_i^{\tau}}{\lambda^{\tau}+X_i^{\tau}} \\ \frac{n}{\lambda} = -\frac{\tau-1}{\lambda} \sum_{i=1}^{n} \frac{1}{X_i} + \tau\frac{\alpha+1}{\lambda} \sum_{i=1}^{n} \frac{1}{\lambda^{\tau}+X_i^{\tau}} \end{cases},$$

which can be solved numerically.

8.5.4 Random generation

From the quantile function $F^{-1}(u) = \lambda((1-u)^{\frac{1}{\alpha}} - 1)^{\frac{1}{\tau}}$, it is easy to generate Burr random variate with $\lambda(U^{\frac{1}{\alpha}} - 1)^{\frac{1}{\tau}}$ where U is a uniform variable.

8.5.5 Applications

NEED REFERENCE

8.6 Inverse Burr distribution

8.6.1 Characterization

The inverse Burr distribution (also called the Dagum distribution) is a special case of the Feller Pareto distribution $\mathcal{FP}$ with $\delta_2 = 1$. That is to say the density is given by

$$f(x) = \frac{\alpha\gamma}{\sigma} \frac{\left(\frac{x-\mu}{\sigma}\right)^{\alpha\gamma-1}}{\left(1+\left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right)^{\gamma+1}},$$

where $x \leq \mu$, μ the location parameter, σ the scale parameter and α, γ the shape parameters. Klugman et al. (2004) defines the inverse Burr distribution with $\mu = 0$, since this book deals with insurance loss distributions. In this expression, it is not so obvious that this is the inverse Burr distribution and not the Burr distribution. But the density can be rewritten as

$$f(x) = \frac{\alpha\gamma}{\sigma} \frac{\left(\frac{\sigma}{x-\mu}\right)^{\alpha+1}}{\left(\left(\frac{\sigma}{x-\mu}\right)^{\alpha}+1\right)^{\gamma+1}},$$

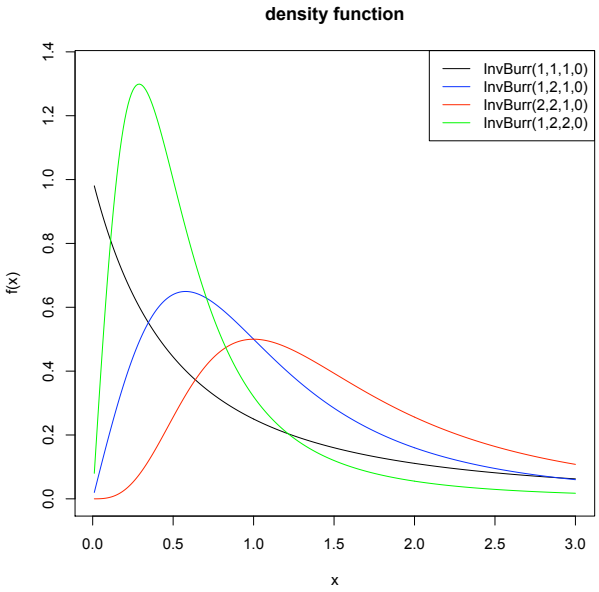


Figure 8.8: Density function for inverse Burr distributions

From this, the distribution function can be derived to

$$F(x) = \left(\frac{1}{\left(\frac{\sigma}{x-\mu}\right)^{\alpha}+1}\right)^{\gamma},$$

for $x \geq \mu$. Here it is also clearer that this is the inverse Burr distribution since we notice the survival function of the Burr distribution taken in $\frac{1}{x}$. We denotes the inverse Burr distribution by $\mathcal{IB}(\gamma, \alpha, \beta, \mu)$.

8.6.2 Properties

The raw moments of the inverse Burr distribution are given by

$$E(X^r) = \sigma^r \frac{\Gamma(\gamma+\frac{r}{\alpha})\Gamma(1-\frac{r}{\alpha})}{\Gamma(\gamma)},$$

when $\mu = 0$ and $\alpha > r$. Thus the expectation and the variance are

$$E(X) = \mu + \sigma \frac{\Gamma(\gamma+\frac{1}{\alpha})\Gamma(1-\frac{1}{\alpha})}{\Gamma(\gamma)}$$

and

$$Var(X) = \sigma^2 \frac{\Gamma(\gamma + \frac{2}{\alpha})\Gamma(1 - \frac{2}{\alpha})}{\Gamma(\gamma)} - \sigma^2 \frac{\Gamma^2(\gamma + \frac{1}{\alpha})\Gamma^2(1 - \frac{1}{\alpha})}{\Gamma^2(\gamma)}$$

Furthermore, we have the following special cases

- with $\gamma = \alpha$, we get the inverse paralogistic distribution,
- with $\gamma = 1$, we have the log logistic distribution,
- with $\alpha = 1$, this is the inverse Pareto distribution.

8.6.3 Estimation

The maximum likelihood estimator of μ is simply $\hat{\mu} = X_{1:n}$ for a sample $(X_i)_i$, then working on the transformed sample $Y_i = X_i - \hat{\mu}$, other maximum likelihood estimators are solutions of the system

$$\begin{cases} \frac{n}{\gamma} = \sum_{i=1}^{n} \log\left(1 + \left(\frac{\lambda}{Y_i}\right)^{\alpha}\right) \\ \frac{n}{\alpha} = -\sum_{i=1}^{n} \log\left(\frac{\sigma}{Y_i}\right) + (\gamma + 1) \sum_{i=1}^{n} \log\left(\frac{\sigma}{Y_i}\right) \frac{\sigma^{\alpha}}{Y_i^{\alpha} + \sigma^{\alpha}} \\ \frac{n}{\sigma} = (\alpha + 1) \sum_{i=1}^{n} \frac{1}{Y_i + \sigma} - \alpha \frac{\gamma + 1}{\sigma} \sum_{i=1}^{n} \frac{\sigma^{\alpha}}{Y_i^{\alpha} + \sigma^{\alpha}} \end{cases},$$

8.6.4 Random generation

Since the quantile function is $F^{-1}(u) = \mu + \sigma^{-1}(u^{-\frac{1}{\gamma}} - 1)^{-\frac{1}{\alpha}}$, we can use the inverse function method.

8.6.5 Applications

NEED REFERENCE

8.7 Beta type II distribution

8.7.1 Characterization

There are many ways to characterize the beta type II distribution. First we can say it is the distribution of $\frac{X}{1-X}$ when X is beta I distributed. But this is also the distribution of the ratio $\frac{U}{V}$

when U and V are gamma distributed ($\mathcal{G}(a,1)$ and $\mathcal{G}(b,1)$ resp.). The distribution function of the beta of the second distribution is given by

$$F(x) = \frac{\beta(a,b,\frac{x}{1+x})}{\beta(a,b)},$$

for $x \leq 0$. The main difference with the beta I distribution is that the beta II distribution takes values in $\mathbb{R}_+$ and not $[0,1]$.

The density can be expressed as

$$f(x) = \frac{x^{a-1}}{\beta(a,b)(1+x)^{a+b}},$$

for $x \leq 0$. It is easier to see the transformation $\frac{x}{1-x}$ if we rewrite the density as

$$f(x) = \left(\frac{x}{1+x}\right)^{a-1}\left(1-\frac{x}{1+x}\right)^{b-1}\frac{1}{\beta(a,b)(1+x)^2}.$$

As already mentioned above, this is a special case of the Feller-Pareto distribution.

8.7.2 Properties

The expectation and the variance of the beta II are given by $E(X) = \frac{a}{b-1}$ and $Var(X) = \frac{a(a+b-1)}{(b-1)^2(b-2)}$ when $b > 1$ and $b > 2$. Raw moments are expressed as follows

$$E(X^r) = \frac{\Gamma(a+r)\Gamma(b-r)}{\Gamma(a)\Gamma(b)},$$

for $b > r$.

8.7.3 Estimation

Maximum likelihood estimators for a and b verify the system

$$\begin{cases} \psi(a) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n}(\log(1+X_i) - \log(X_i)) \\ \psi(b) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n}\log(1+X_i) \end{cases},$$

where ψ denotes the digamma function. We may also use the moment based estimators given by

$$\tilde{b} = 2 + \frac{\bar{X}_n(\bar{X}_n+1)}{S_n^2} \quad \text{and} \quad \tilde{a} = (\tilde{b}-1)\bar{X}_n,$$

which have the drawback that $\tilde{b}$ is always greater than 2.

8.7.4 Random generation

We can simply use the construction of the beta II, i.e. the ratio of $\frac{X}{1-X}$ when X is beta I distributed. However we may also use the ratio of two gamma variables.

8.7.5 Applications

NEED REFERENCE

Chapter 9

Logistic distribution and related extensions

9.1 Logistic distribution

9.1.1 Characterization

The logistic distribution is defined by the following distribution function

$$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{s}}},$$

where $x \in \mathbb{R}$, μ the location parameter and s the scale parameter. TODO

9.1.2 Properties

TODO

9.1.3 Estimation

TODO

9.1.4 Random generation

TODO

9.1.5 Applications

9.2 Half logistic distribution

9.2.1 Characterization

9.2.2 Properties

9.2.3 Estimation

9.2.4 Random generation

9.2.5 Applications

9.3 Log logistic distribution

9.3.1 Characterization

9.3.2 Properties

9.3.3 Estimation

9.3.4 Random generation

9.3.5 Applications

9.4 Generalized log logistic distribution

9.4.1 Characterization

9.4.2 Properties

9.4.3 Estimation

9.4.4 Random generation

9.4.5 Applications

9.5 Paralogisitic distribution

Chapter 10

Extrem Value Theory distributions

10.1 Gumbel distribution

10.1.1 Characterization

The standard Gumbel distribution is defined by the following density function

$$f(x) = e^{-x-e^{-x}},$$

where $x \in \mathbb{R}$. Its distribution function is expressed as follows

$$F(x) = e^{-e^{-x}}.$$

A scaled and shifted version of the Gumbel distribution exists. The density is defined as

$$f(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma} - e^{-\frac{x-\mu}{\sigma}}},$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$. We get back to the standard Gumbel distribution with $\mu = 0$ and $\sigma = 1$. The distribution function of the Gumbel I distribution is simply

$$F(x) = e^{-e^{-\frac{x-\mu}{\sigma}}},$$

for $x \in \mathbb{R}$.

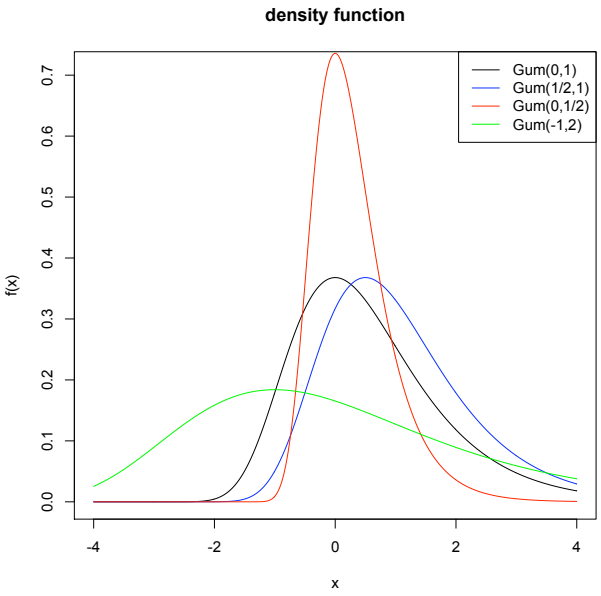


Figure 10.1: Density function for Gumbel distributions

There exists a Gumbel distribution of the second kind defined by the following distribution function

$$F(x) = 1 - e^{-e^{\frac{x-\mu}{\sigma}}},$$

for $x \in \mathbb{R}$. Hence we have the density

$$f(x) = \frac{1}{\sigma} e^{\frac{x-\mu}{\sigma} - e^{\frac{x-\mu}{\sigma}}}.$$

This is the distribution of $-X$ when X is Gumbel I distributed.

The characteristic function of the Gumbel distribution of the first kind exists

$$\phi(t) = \Gamma(1 - i\sigma t)e^{i\mu t},$$

while its moment generating function are

$$M(t) = \Gamma(1 - \sigma t)e^{\mu t}.$$

10.1.2 Properties

The expectation of a Gumbel type I distribution is $E(X) = \gamma$, the Euler constant, roughly 0.57721. Its variance is $Var(X) = \frac{\pi^2}{6}$. Thus for the Fisher-Tippett distribution, we have $E(X) = \mu + \sigma\gamma$ and $Var(X) = \frac{\pi^2\sigma^2}{6}$.

For the Gumbel type II, expectation exists if $a > 1$ and variance if $a > 2$.

10.1.3 Estimation

Maximum likelihood estimators are solutions of the following system

$$\begin{cases} 1 = \frac{1}{n}\sum_{i=1}^{n} e^{-\frac{X_i-\mu}{\sigma}} \\ \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} X_i e^{-\frac{X_i-\mu}{\sigma}} \end{cases},$$

which can solved numerically initialized by the moment based estimators

$$\tilde{\mu} = \bar{X}_n - \tilde{\sigma}\gamma \quad \text{and} \quad \tilde{\sigma} = \sqrt{\frac{6S_n^2}{\pi^2}},$$

where γ is the Euler constant.

10.1.4 Random generation

The quantile function of the Gumbel I distribution is simply $F^{-1}(u) = \mu - \sigma\log(-\log(u))$, thus we can use the inverse function method.

10.1.5 Applications

The Gumbel distribution is widely used in natural catastrophe modelling, especially for maximum flood. NEED REFERENCE

10.2 Fréchet distribution

A Fréchet type distribution is a distribution whose distribution function is

$$F(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}},$$

for $x \geq \mu$. One can notice this is the inverse Weibull distribution, see section 5.12 for details.

10.3 Weibull distribution

A Weibull type distribution is characterized by the following distribution function

$$F(x) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^{\beta}},$$

for $x \geq \mu$. See section 5.11 for details.

10.4 Generalized extreme value distribution

10.4.1 Characterization

The generalized extreme value distribution is defined by the following distribution function

$$F(x) = e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}},$$

for $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$, ξ the shape parameter, μ the location parameter and $\sigma > 0$ the scale parameter. We can derive a density function

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1} e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}}.$$

This distribution is sometimes called the Fisher-Tippett distribution.

Let us note that the values can be taken in $\mathbb{R}$, $\mathbb{R}_-$ or $\mathbb{R}_+$ according to the sign of ξ. The distribution function is generally noted by $H_{\xi,\mu,\sigma}$, wich can expressed with the "standard" generalized extreme value distribution $H_{\xi,0,1}$ with a shift and a scaling. When ξ tends to zero, we get the Gumbel I distribution

$$H_{\xi,\mu,\sigma}(x) \underset{\xi\to 0}{\longrightarrow} e^{-e^{-\frac{x-\mu}{\sigma}}}.$$

10.4.2 Properties

The expectation and the variance are

$$E(X) = \mu - \frac{\sigma}{\xi}\Gamma(1-\xi) \quad \text{and} \quad Var(X) = \frac{\sigma^2}{\xi^2}(\Gamma(1-2\xi) - \Gamma^2(1-\xi))$$

if they exist.

From the extreme value theory, we have the following theorem. Let $(X_i)_{1\leq i\leq n}$ be an i.i.d. sample and $X_{i:n}$ the order statistics. If there exits two sequences $(a_n)_n$ and $(b_n)_n$ valued in $\mathbb{R}_+$ and $\mathbb{R}$ respectively, such that

$$P\left(\frac{X_{n:n}-b_n}{a_n}\right)$$

have a limit in probability distribution. Then the limiting distribution H for the maximum belongs to the type of one the following three distribution functions

$$H(x)=\begin{cases} e^{-x^{-\xi}}, & x\geq 0,\xi>0, \quad \text{MDA of Fréchet} \\ e^{-(-x)^{\xi}}, & x\leq 0,\xi<0, \quad \text{MDA of Weibull} \\ e^{-e^{-x}}, & x\in\mathbb{R},\xi=0, \quad \text{MDA of Gumbel} \end{cases},$$

where MDA stands for maximum domains of attraction*. For all distribution, there is a unique MDA. We quickly see that the limiting distribution for the maximum is nothing else than the generalized extreme value distribution $H_{\xi,0,1}$. This theorem is the Fisher-Tippett-Gnedenko theorem.

For the minimum, assuming that $P\left(\frac{X_{1:n}-b_n}{a_n}\right)$ has a limit, the limiting distribution belongs to

$$\tilde{H}(x)=\begin{cases} 1-e^{-x^{\beta}}, & x\geq 0,\beta>0 \\ 1-e^{-(-x)^{\beta}}, & x\leq 0,\beta<0 \\ 1-e^{-e^{x}}, & x\in\mathbb{R},\beta=0 \end{cases}.$$

Again there are three types for the limiting distribution†.

In the MDA of Fréchet, we have the Cauchy, the Pareto, the Burr, the log-gamma and the stable distributions, while in the Weibull MDA we retrieve the uniform, the beta and bounded support power law distribution. Finally, the MDA of Gumbel contains the exponential, the Weibull, the gamma, the normal, the lognormal, the Benktander distributions.

From the Embrechts et al. (1997), we also have some equivalence given a MDA:

- a distribution function F belongs to the MDA of Fréchet if and only if $1-F(x)=x^{-\alpha}L(x)$ for some slowly varying function L,
- a distribution function F belongs to the MDA of Weibull if and only if $1-F(x_F-1/x)=x^{-\alpha}L(x)$ for some slowly varying function L and $x_F<+\infty$,
- a distribution function F belongs to the MDA of Gumbel if and only if there exists $z<x_F$ such that $1-F(x)=c(x)e^{-\int_z^x \frac{g(t)}{a(t)}dt}$ for some measurable function c, g and a continuous function a.

*Sometimes the distribution characterized by the distribution function $e^{-e^{-x}}$ is called the extreme maximal-value distribution.

†Sometimes the distribution characterized by the distribution function $1-e^{-e^{x}}$ is called the extreme minimal-value distribution.

10.4.3 Estimation

According to Embrechts et al. (1997) maximum likelihood estimation is not very reliable in the case of the generalized extreme value fitting. But that's not surprising since the generalized extreme value distribution is a limiting distribution to very heterogeneous distribution, such as heavy tailed, light tailed or bounded distributions.

We can use weighted moment method, where we estimate moments

$$\omega_r(\xi, \mu, \sigma) = E(XH^r_{\xi,\mu,\sigma}(X))$$

by its empirical equivalent

$$\hat{\omega}_r = \frac{1}{n}\sum_{i=1}^{n} X_{j:n}U^r_{j:n},$$

where $U^r_{j:n}$ are the order statistics of an uniform sample (which can be replaced by its expectation $\frac{(n-r-1)!}{(n-1)!}\frac{(n-j)!}{(n-j-r)!}$). Equalling the theoretical and the empirical moments, we get that ξ is a solution of

$$\frac{3\hat{\omega}_2 - \hat{\omega}_0}{2\hat{\omega}_1 - \hat{\omega}_0} = \frac{3^\xi - 1}{2^\xi - 1}.$$

Then we estimate the other two parameters with

$$\hat{\sigma} = \frac{(2\hat{\omega}_1 - \hat{\omega}_0)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}} - 1)} \quad \text{and} \quad \hat{\mu} = \hat{\omega}_0 + \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma(1-\hat{\xi})).$$

10.4.4 Random generation

The quantile function of the generalized extreme value distribution is $F^{-1}(u) = \mu + \frac{\sigma}{\xi}((-\log u)^{-\xi}) - 1$ for $\xi \neq 0$. So we can use the inverse function method.

10.4.5 Applications

The application of the generalized extreme value distribution is obviously the extremex value theory which can be applied in many fields : natural disaster modelling, insurance/finance extreme risk management,...

10.5 Generalized Pareto distribution

See section 8.4 for details.

Part III

Multivariate and generalized distributions

Chapter 11

Generalization of common distributions

11.1 Generalized hyperbolic distribution

This part entirely comes from Breymann & Lüthi (2008).

11.1.1 Characterization

The first way to characterize generalized hyperbolic distributions is to say that the random vector X follows a multivariate $\mathcal{GH}$ distribution if

$$X \overset{\mathcal{L}}{=} \mu + W\gamma + \sqrt{W}AZ \tag{11.1}$$

where

1. $Z \sim \mathcal{N}_k(\mathbf{0}, I_k)$
2. $A \in \mathbb{R}^{d\times k}$
3. $\mu, \gamma \in \mathbb{R}^d$
4. $W \geq 0$ is a scalar-valued random variable which is independent of Z and has a Generalized Inverse Gaussian distribution, written $GIG(\lambda, \chi, \psi)$.

Note that there are at least five alternative definitions leading to different parametrizations.

Nevertheless, the parameters of a $\mathcal{GH}$ distribution given by the above definition admit the following interpretation:

- λ, χ, ψ determine the shape of the distribution, that is, how much weight is assigned to the tails and to the center. In general, the larger those parameters the closer is the distribution to the normal distribution.

- μ is the location parameter.
- $\Sigma = AA'$ is the dispersion-matrix.
- γ is the skewness parameter. If $\gamma = 0$, then the distribution is symmetric around μ.

Observe that the conditional distribution of $X|W = w$ is normal,

$$X|W = w \sim \mathcal{N}_d(\mu + w\,\gamma, w\Sigma), \tag{11.2}$$

Another way to define a generalized hyperbolic distribution is to use the density. Since the conditional distribution of X given W is Gaussian with mean $\mu + W\gamma$ and variance $W\Sigma$ the $\mathcal{GH}$ density can be found by mixing $X|W$ with respect to W.

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|W}(x|w)\, f_W(w)\, dw \qquad (11.3)\\ &= \int_0^\infty \frac{e^{(x-\mu)'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}\,|\Sigma|^{\frac{1}{2}}\,w^{\frac{d}{2}}} \exp\left\{-\frac{Q(x)}{2w} - \frac{\gamma\Sigma\gamma}{2/w}\right\} f_W(w)dw \\ &= \frac{(\sqrt{\psi/\chi})^\lambda(\psi+\gamma\Sigma\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}\,|\Sigma|^{\frac{1}{2}}\,K_\lambda(\sqrt{\chi\psi})} \times \frac{K_{\lambda-\frac{d}{2}}(\sqrt{(\chi+Q(x))(\psi+\gamma\Sigma\gamma)})\; e^{(x-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi+Q(x))(\psi+\gamma\Sigma\gamma)})^{\frac{d}{2}-\lambda}}, \end{aligned}$$

where $K_\lambda(\cdot)$ denotes the modified Bessel function of the third kind and $Q(x)$ denotes the mahalanobis distance $Q(x) = (x-\mu)'\Sigma^{-1}(x-\mu)$ (i.e. the distance with Σ^{-1} as norm). The domain of variation of the parameters λ, χ and ψ is given in section 11.1.2.

A last way to characterize generalized hyperbolic distributions is the usage of moment generating functions. An appealing property of normal mixtures is that the moment generating function is easily calculated once the moment generating function of the mixture is known. Based on equation (11.4) we obtain the moment generating function of a $\mathcal{GH}$ distributed random variable X as

$$\begin{aligned} M(t) &= E(E(\exp\left\{t'X\right\}|W)) = e^{t'\mu}E(\exp\left\{W\left(t'\gamma + 1/2\; t'\Sigma t\right)\right\}) \\ &= e^{t'\mu}\left(\frac{\psi}{\psi - 2t'\gamma - t'\Sigma t}\right)^{\lambda/2} \frac{K_\lambda(\sqrt{\psi(\chi - 2t'\gamma - t'\Sigma t)})}{K_\lambda(\sqrt{\chi\psi})}, \quad \chi \geq 2\; t'\gamma + t'\Sigma t. \end{aligned}$$

For moment generating functions of the special cases of the $\mathcal{GH}$ distribution we refer to Prause (1999) and Paolella (2007).

11.1.2 Parametrization

There are several alternative parametrizations for the GH distribution. In the R package ghyp the user can choose between three of them. There exist further parametrizations which are not implemented and not mentioned here. For these parametrizations we refer to Prause (1999) and Paolella (2007).

Table 11.1 describes the parameter ranges for each parametrization and each special case. Clearly, the dispersion matrices Σ and Δ have to fulfill the usual conditions for covariance matrices, i.e., symmetry and positive definiteness as well as full rank.

	$(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-Parametrization					
	λ	χ	ψ	μ	Σ	γ
ghyp	$\lambda \in \mathbb{R}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
hyp	$\lambda = \frac{d+1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
NIG	$\lambda = -\frac{1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
t	$\lambda < 0$	$\chi > 0$	$\psi = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
VG	$\lambda > 0$	$\chi = 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$

	$(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-Parametrization				
	λ	$\bar{\alpha}$	μ	Σ	γ
ghyp	$\lambda \in \mathbb{R}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
hyp	$\lambda = \frac{d+1}{2}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
NIG	$\lambda = \frac{1}{2}$	$\bar{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
t	$\lambda = -\frac{\nu}{2} < -1$	$\bar{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$
VG	$\lambda > 0$	$\bar{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^\Sigma$	$\gamma \in \mathbb{R}^d$

	$(\lambda, \alpha, \mu, \Sigma, \delta, \beta)$-Parametrization					
	λ	α	δ	μ	Δ	β
ghyp	$\lambda \in \mathbb{R}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x'\Delta x > 0\}$
hyp	$\lambda = \frac{d+1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x'\Delta x > 0\}$
NIG	$\lambda = -\frac{1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x'\Delta x > 0\}$
t	$\lambda < 0$	$\alpha = \sqrt{\beta'\Delta\beta}$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \mathbb{R}^d$
VG	$\lambda > 0$	$\alpha > 0$	$\delta = 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^\Delta$	$\beta \in \{x \in \mathbb{R}^d : \alpha^2 - x'\Delta x > 0\}$

Table 11.1: The domain of variation for the parameters of the GH distribution and some of its special cases for different parametrizations. We denote the set of all feasible covariance matrices in $\mathbb{R}^{d\times d}$ with $\mathbb{R}^\Sigma$. Furthermore, let $\mathbb{R}^\Delta = \{A \in \mathbb{R}^\Sigma : |A| = 1\}$.

Internally, he package ghyp uses the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization. However, fitting is done in the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-parametrization since this parametrization does not necessitate additional constraints to eliminate the redundant degree of freedom. Consequently, what cannot be represented by the $(\lambda, \alpha, \mu, \Sigma, \delta, \beta)$-parametrization cannot be fitted (cf. section 11.1.2).

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-Parametrization

The $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization is obtained as the normal mean-variance mixture distribution when $W \sim GIG(\lambda, \chi, \psi)$. This parametrization has a drawback of an identification problem. Indeed, the distributions $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $GH_d(\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ are identical for any $k > 0$. Therefore, an identifying problem occurs when we start to fit the parameters of a GH distribution to data. This problem could be solved by introducing a suitable contraint. One possibility is to require the determinant of the dispersion matrix Σ to be 1.

$(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-Parametrization

There is a more elegant way to eliminate the degree of freedom. We simply constrain the expected value of the generalized inverse Gaussian distributed mixing variable W to be 1 (cf. 4.5). This makes the interpretation of the skewness parameters γ easier and in addition, the fitting procedure becomes faster (cf. 11.1.5).
We define

$$E(W) = \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} = 1. \tag{11.4}$$

and set

$$\bar{\alpha} = \sqrt{\chi\psi}. \tag{11.5}$$

It follows that

$$\psi = \bar{\alpha}\, \frac{K_{\lambda+1}(\bar{\alpha})}{K_\lambda(\bar{\alpha})} \text{ and } \chi = \frac{\bar{\alpha}^2}{\psi} = \bar{\alpha}\, \frac{K_\lambda(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})}. \tag{11.6}$$

The drawback of the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-parametrization is that it does not exist in the case $\bar{\alpha} = 0$ and $\lambda \in [-1, 0]$, which corresponds to a Student-t distribution with non-existing variance. Note that the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-parametrization yields to a slightly different parametrization for the special case of a Student-t distribution.

$(\lambda, \alpha, \mu, \Sigma, \delta, \beta)$-Parametrization

When the GH distribution was introduced in Barndorff-Nielsen (1977), the following parametrization for the multivariate case was used.

$$f_X(x) = \frac{(\alpha^2 - \beta'\Delta\beta)^{\lambda/2}}{(2\pi)^{\frac{d}{2}}\sqrt{|\Delta|}\,\delta^\lambda\, K_\lambda(\delta\sqrt{\alpha^2 - \beta'\Delta\beta})} \times \frac{K_{\lambda-\frac{d}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)'\Delta^{-1}(x-\mu)})\; e^{\beta'(x-\mu)}}{(\alpha\sqrt{\delta^2 + (x-\mu)'\Delta^{-1}(x-\mu)})^{\frac{d}{2}-\lambda}}, \tag{11.7}$$

where the determinant of Δ is constrained to be 1. In the univariate case the above expression reduces to

$$f_X(x) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\,\alpha^{\lambda-\frac{1}{2}}\,\delta^\lambda\, K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} \times K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2})\; e^{\beta(x-\mu)}, \tag{11.8}$$

which is the most widely used parametrization of the GH distribution in literature.

Switching between different parametrizations

The following formulas can be used to switch between the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$, $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, and the $(\lambda, \alpha, \mu, \Sigma, \delta, \beta)$-parametrization. The parameters λ and μ remain the same, regardless of the parametrization.
The way to obtain the $(\lambda, \alpha, \mu, \Sigma, \delta, \beta)$-parametrization from the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$-parametrization yields over the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization:

$$(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma) \quad \leftrightarrows \quad (\lambda, \chi, \psi, \mu, \Sigma, \gamma) \quad \leftrightarrows \quad (\lambda, \alpha, \mu, \Sigma, \delta, \beta)$$

$(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma) \rightarrow (\lambda, \chi, \psi, \mu, \Sigma, \gamma)$**:** Use the relations in (11.6) to obtain χ and ψ. The parameters Σ and γ remain the same.

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \rightarrow (\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$**:** Set $k = \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}$.

$$\bar{\alpha} = \sqrt{\chi\psi}, \quad \Sigma \equiv k\,\Sigma, \quad \gamma \equiv k\,\gamma \tag{11.9}$$

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \rightarrow (\lambda, \alpha, \mu, \Sigma, \delta, \beta)$**:**

$$\begin{aligned} \Delta = |\Sigma|^{-\frac{1}{d}}\Sigma \quad &, \quad \beta = \Sigma^{-1}\gamma \\ \delta = \sqrt{\chi|\Sigma|^{\frac{1}{d}}} \quad &, \quad \alpha = \sqrt{|\Sigma|^{-\frac{1}{d}}(\psi + \gamma'\Sigma^{-1}\gamma)} \end{aligned} \tag{11.10}$$

$(\lambda, \alpha, \mu, \Sigma, \delta, \beta) \rightarrow (\lambda, \chi, \psi, \mu, \Sigma, \gamma)$**:**

$$\Sigma = \Delta, \quad \gamma = \Delta\beta, \quad \chi = \delta^2, \quad \psi = \alpha^2 - \beta'\Delta\beta. \tag{11.11}$$

11.1.3 Properties

Moments

The expected value and the variance are given by

$$\begin{aligned} E(X) &= \mu + E(W)\gamma && (11.12) \\ Var(X) &= E(Cov(X|W)) + Cov(E(X|X)) && (11.13) \\ &= Var(W)\gamma\gamma' + E(W)\Sigma. \end{aligned}$$

Linear transformation

The GH class is closed under linear transformations: If $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $Y = BX + \mathbf{b}$, where $B \in \mathbb{R}^{k\times d}$ and $\mathbf{b} \in \mathbb{R}^k$, then $Y \sim GH_k(\lambda, \chi, \psi, B\mu + \mathbf{b}, B\Sigma B', B\gamma)$. Observe that by introducing a new skewness parameter $\bar{\gamma} = \Sigma\gamma$, all the shape and skewness parameters $(\lambda, \chi, \psi, \bar{\gamma})$ become location and scale-invariant, provided the transformation does not affect the dimensionality, that is $B \in \mathbb{R}^{d\times d}$ and $\mathbf{b} \in \mathbb{R}^d$.

11.1.4 Special cases

The GH distribution contains several special cases known under special names.

- If $\lambda = \frac{d+1}{2}$ the name generalized is dropped and we have a multivariate *hyperbolic (hyp)* distribution. The univariate margins are still GH distributed. Inversely, when $\lambda = 1$ we get a multivariate GH distribution with hyperbolic margins.
- If $\lambda = -\frac{1}{2}$ the distribution is called *Normal Inverse Gaussian (NIG).*

- If $\chi = 0$ and $\lambda > 0$ one gets a limiting case which is known amongst others as *Variance Gamma (VG)* distribution.
- If $\psi = 0$ and $\lambda < -1$ one gets a limiting case which is known as a *generalized hyperbolic Student-t* distribution (called simply *Student-t* in what follows).

11.1.5 Estimation

Numerical optimizers can be used to fit univariate GH distributions to data by means of maximum likelihood estimation. Multivariate GH distributions can be fitted with expectation-maximazion (EM) type algorithms (see Dempster et al. (1977) and Meng & Rubin (1993)).

EM-Scheme

Assume we have iid data $x_1, \ldots, x_n$ and parameters represented by $\Theta = (\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$. The problem is to maximize

$$\ln L(\Theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f_X(x_i; \Theta). \tag{11.14}$$

This problem is not easy to solve due to the number of parameters and necessity of maximizing over covariance matrices. We can proceed by introducing an augmented likelihood function

$$\ln \tilde{L}(\Theta; x_1, \ldots, x_n, w_1, \ldots, w_n) = \sum_{i=1}^{n} \ln f_{X|W}(x_i|w_i; \mu, \Sigma, \gamma) + \sum_{i=1}^{n} \ln f_W(w_i; \lambda, \bar{\alpha}) \tag{11.15}$$

and spend the effort on the estimation of the latent mixing variables w_i coming from the mixture representation (11.2). This is where the EM algorithm comes into play.

E-step: Calculate the conditional expectation of the likelihood function (11.15) given the data $x_1, \ldots, x_n$ and the current estimates of parameters $\Theta^{[k]}$. This results in the objective function

$$Q(\Theta; \Theta^{[k]}) = E\left(\ln \tilde{L}(\Theta; x_1, \ldots, x_n, w_1, \ldots, w_n)|x_1, \ldots, x_n; \Theta^{[k]}\right). \tag{11.16}$$

M-step: Maximize the objective function with respect to Θ to obtain the next set of estimates $\Theta^{[k+1]}$.

Alternating between these steps yields to the maximum likelihood estimation of the parameter set Θ. In practice, performing the E-Step means maximizing the second summand of (11.15) numerically. The log density of the GIG distribution (cf. 4.5.1) is

$$\ln f_W(w) = \frac{\lambda}{2}\ln(\psi/\chi) - \ln(2K_\lambda(\sqrt{\chi\psi})) + (\lambda - 1)\ln w - \frac{\chi}{2}\frac{1}{w} - \frac{\psi}{2}w. \tag{11.17}$$

When using the $(\lambda, \bar{\alpha})$-parametrization this problem is of dimension two instead of three as it is in the (λ, χ, ψ)-parametrization. As a consequence the performance increases.

Since the w_i's are latent one has to replace w, $1/w$ and $\ln w$ with the respective expected values in order to maximize the log likelihood function. Let

$$\eta_i^{[k]} := E\left(w_i \,|\, x_i; \Theta^{[k]}\right),\ \delta_i^{[k]} := E\left(w_i^{-1} \,|\, x_i; \Theta^{[k]}\right),\ xi_i^{[k]} := E\left(\ln w_i \,|\, x_i; \Theta^{[k]}\right). \tag{11.18}$$

We have to find the conditional density of w_i given x_i to calculate these quantities.

MCECM estimation

In the `R` implementation a modified EM scheme is used, which is called multi-cycle, expectation, conditional estimation (MCECM) algorithm (Meng & Rubin 1993, McNeil, Frey & Embrechts 2005*a*). The different steps of the MCECM algorithm are sketched as follows:

(1) Select reasonable starting values for $\Theta^{[k]}$. For example $\lambda = 1$, $\bar{\alpha} = 1$, μ is set to the sample mean, Σ to the sample covariance matrix and γ to a zero skewness vector.

(2) Calculate $\chi^{[k]}$ and $\psi^{[k]}$ as a function of $\bar{\alpha}^{[k]}$ using (11.6).

(3) Use (11.18), (11.12) to calculate the weights $\eta_i^{[k]}$ and $\delta_i^{[k]}$. Average the weights to get

$$\bar{\eta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n} \eta_i^{[k]} \text{ and } \bar{\delta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n} \delta_i^{[k]}. \tag{11.19}$$

(4) If an asymmetric model is to be fitted set γ to $\mathbf{0}$, else set

$$\gamma^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n} \delta_i^{[k]}(\bar{x} - x_i)}{\bar{\eta}^{[k]}\bar{\delta}^{[k]} - 1}. \tag{11.20}$$

(5) Update μ and Σ:

$$\mu^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n} \delta_i^{[k]}(x_i - \gamma^{[k+1]})}{\bar{\delta}^{[k]}} \tag{11.21}$$

$$\Sigma^{[k+1]} = \frac{1}{n}\sum_{i=1}^{n} \delta_i^{[k]}(x_i - \mu^{[k+1]})(x_i - \mu^{[k+1]})' - \bar{\eta}^{[k]}\gamma^{[k+1]}\gamma^{[k+1]\prime}. \tag{11.22}$$

(6) Set $\Theta^{[k,2]} = (\lambda^{[k]}, \bar{\alpha}^{[k]}, \mu^{[k+1]}, \Sigma^{[k+1]}, \gamma^{[k+1]})$ and calculate weights $\eta_i^{[k,2]}$, $\delta_i^{[k,2]}$ and $xi_i^{[k,2]}$ using (11.18), (4.2) and (11.12).

(7) Maximize the second summand of (11.15) with density (11.17) with respect to λ, χ and ψ to complete the calculation of $\Theta^{[k,2]}$ and go back to step (2). Note that the objective function must calculate χ and ψ in dependence of λ and $\bar{\alpha}$ using relation (11.6).

11.1.6 Random generation

We can simply use the first characterization by adding

$$\mu + W\gamma + \sqrt{W}AZ$$

where Z is a multivariate gaussian vector $\mathcal{N}_k(\mathbf{0}, I_k)$ and W follows a Generalized Inverse Gaussian $GIG(\lambda, \chi, \psi)$.

11.1.7 Applications

Even though the GH distribution was initially ivented to study the distribution of the logarithm of particle sizes, we will focus on applications of the GH distribution family in finance and risk measurement.

We have seen above that the GH distribution is very flexible in the sense that it nests several other distributions such as the Student-t (cf. 7.1).

To give some references and applications of the GH distribution let us first summarize some of its important properties. Beside of the above mentioned flexibility, three major facts led to the popularity of GH distribution family in finance:

(1) The GH distribution features both *fat tails* and *skewness*. These properties account for the stylized facts of financial returns.

(2) The GH family is naturally extended to multivariate distributions*. A multivariate GH distribution does exhibit some kind of *non-linear dependence*, for example *tail-dependence*. This reflects the fact that extremes mostly occure for a couple of risk-drivers simultaneously in financial markets. This property is of fundamental importance for risk-management, and can influence for instance the asset allocation in portfolio theory.

(3) The GH distribution is infinitely divisible (cf. Barndorff-Nielsen & Halgreen (1977)). This is a necessary and sufficient condition to build *Lévy processes*. Lévy processes are widespread in finance because of their time-continuity and their ability to model jumps.

Based on these properties one can classify the applications of the GH distributions into the fields *empirical modelling*, *risk and dependence modelling*, *derivative pricing*, and *portfolio selection*.

In the following, we try to assign papers to each of the classes of applications mentioned above. Rather than giving abstracts for each paper, we simply cite them and refer the interested reader to the bibliography and to the articles. Note that some articles deal with special cases of the GH distribution only.

Empirical modelling Eberlein & Keller (1995), Barndorff-Nielsen & Prause (2001), Fergusson & Platen (2006)

Risk and dependence modelling Eberlein et al. (1998), Breymann et al. (2003), McNeil et al. (2005*b*), Chen et al. (2005), Kassberger & Kiesel (2006)

Lévy processes Barndorff-Nielsen (1997*a*,*b*), Bibby & Sorensen (1997), Dilip B. Madan et al. (1998), Raible (2000), Cont & Tankov (2003)

*The extension to multivariate distributions is natural because of the mixing structure (see eq. (11.2)).

Portfolio selection Kassberger (2007)

11.2 Stable distribution

A detailed and complete review of stable distributions can be found in Nolan (2009).

11.2.1 Characterization

Stable distributions are characterized by the following equation

$$a\tilde{X} + b\underset{\sim}{X} \overset{\mathcal{L}}{=} cX + d,$$

where $\tilde{X}$ and $\underset{\sim}{X}$ are independent copies of a random variable X and some positive constants a, b, c and d. This equation means stable distributions are distributions closed for linear combinations. For the terminology, we say X is strictly stable if $d = 0$ and symmetric stable if in addition we have $X \overset{\mathcal{L}}{=} -X$. From Nolan (2009), we learn we use the word stable since the shape of the distribution is preserved under linear combinations.

Another way to define stable distribution is to use characteristic functions. X has a stable distribution if and only if its characteristic function is

$$\phi(t) = e^{it\delta} \times \begin{cases} e^{-|t\gamma|^\alpha(1-i\beta\tan(\frac{\pi\alpha}{2})\text{sign}(t))} & \text{if } \alpha \neq 1 \\ e^{-|t\gamma|(1+i\beta\frac{2}{\pi}\log|t|\text{sign}(t))} & \text{if } \alpha = 1 \end{cases},$$

where $\alpha \in]0, 2]$, $\beta \in]-1, 1[$, $\gamma > 0$ and $b \in \mathbb{R}$ are the parameters. In the following, we denote $\mathcal{S}(\alpha, \beta, \gamma, \delta)$, where δ is a location parameter, γ a scale parameter, α an index of stability and β a skewness parameter. This corresponds to the parametrization 1 of Nolan (2009).

We know that stable distributions $\mathcal{S}(\alpha, \beta, \gamma, \delta)$ are continuous distributions whose support is

$$\begin{cases} [\delta, +\infty[& \text{if } \alpha < 1 \text{ and } \beta = 1 \\ ]-\infty, \delta] & \text{if } \alpha < 1 \text{ and } \beta = -1 \\ ]-\infty, +\infty[& \text{otherwise} \end{cases}.$$

11.2.2 Properties

If we work with standard stable distributions $\mathcal{S}(\alpha, \beta, 0, 1)$, we have the reflection property. That is to say if $X \sim \mathcal{S}(\alpha, \beta, 0, 1)$, then $-X \sim \mathcal{S}(\alpha, -\beta, 0, 1)$. This implies the following constraint on the density and the distribution function:

$$f_X(x) = f_{-X}(-x) \text{ and } F_X(x) = 1 - F_{-X}(x).$$

From the definition, we have the obvious property on the sum. If X follows a stable distribution $\mathcal{S}(\alpha, \beta, \gamma, \delta)$, then $aX + b$ follows a stable distribution of parameters

$$\begin{cases} \mathcal{S}(\alpha, \text{sign}(a)\beta, |a|\gamma, a\delta + b) & \text{if } \alpha \neq 1 \\ \mathcal{S}(1, \text{sign}(a)\beta, |a|\gamma, a\delta + b - \frac{2}{\pi}\beta\gamma a \log|a|) & \text{if } \alpha = 1 \end{cases}.$$

Furthermore if X_1 and X_2 follow a stable distribution $\mathcal{S}(\alpha, \beta_i, \gamma_i, \delta_i)$ for $i = 1, 2$, then the sum $X_1 + X_2$ follows a stable distribution $\mathcal{S}(\alpha, \beta, \gamma, \delta)$ with $\beta = \frac{\beta_1\gamma_1^\alpha + \beta_2\gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}$, $\gamma = (\gamma_1^\alpha + \gamma_2^\alpha)^{\frac{1}{\alpha}}$ and $\delta = \delta_1 + \delta_2$.

11.2.3 Special cases

The following distributions are special cases of stable distributions:

- $\mathcal{S}(2, 0, \sigma/\sqrt{2}, \mu)$ is a Normal distribution defined by the density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,
- $\mathcal{S}(1, 0, \gamma, \delta)$ is a Cauchy distribution defined by the density $f(x) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (x-\gamma)^2}$,
- $\mathcal{S}(1/2, 1, \gamma, \delta)$ is a Lévy distribution defined by the density $f(x) = \sqrt{\frac{\gamma}{2\pi}}\frac{1}{(x-\delta)^{\frac{3}{2}}} e^{-\frac{\gamma}{2(x-\delta)}}$.

11.2.4 Estimation

NEED REFERENCE

11.2.5 Random generation

Simulation of stable distributions are carried out by the following algorithm from Chambers et al. (1976). Let Θ be an independent random uniform variable $\mathcal{U}(-\pi/2, \pi/2)$ and W be an exponential variable with mean 1 independent from Θ. For $0 < \alpha \leq 2$, we have

- in the symmetric case,
$$Z = \frac{\sin(\alpha\Theta)}{\cos(\Theta)^{\frac{1}{\alpha}}} \left(\frac{\cos((\alpha-1)\Theta)}{W}\right)^{\frac{1-\alpha}{\alpha}}$$
follows a stable distribution $\mathcal{S}(\alpha, 0, 1, 0)$ with the limiting case $\tan(\Theta)$ when $\alpha \to 1$.
- in the nonsymetric case,
$$Z = \begin{cases} \frac{\sin(\alpha(\Theta+\theta))}{(\cos(\alpha\theta)\cos(\Theta))^{\frac{1}{\alpha}}} \left(\frac{\cos(\alpha\theta + (\alpha-1)\Theta)}{W}\right)^{\frac{1-\alpha}{\alpha}} \\ \frac{2}{\pi}\left(\left(\frac{\pi}{2} + \beta\Theta\right)\tan(\Theta) - \beta\log\left(\frac{\frac{\pi}{2}W\cos(\Theta)}{\frac{\pi}{2} + \beta\Theta}\right)\right) \end{cases}$$
follows a stable distribution $\mathcal{S}(\alpha, \beta, 1, 0)$ where $\theta = \arctan(\beta\tan(\pi\alpha/2))/\alpha$.

Then we get a "full" stable distribution with $\gamma Z + \delta$.

11.2.6 Applications

NEED REFERENCE

11.3 Phase-type distribution

11.3.1 Characterization

A phase-type distribution $PH(\pi, T, m)$ (π a row vector of $\mathbb{R}^m$, T a $m \times m$ matrix) is defined as the distribution of the time to absorption in the state 0 of a Markov jump process, on the set $\{0, 1, \ldots, m\}$, with initial probability $(0, \pi)$ and intensity matrix*

$$\Lambda = (\lambda_{ij})_{ij} = \left(\begin{array}{c|c} 0 & 0 \\ \hline t_0 & T \end{array}\right),$$

where the vector t_0 is $-T\mathbf{1}_m$ and $\mathbf{1}_m$ stands for the column vector of 1 in $\mathbb{R}^m$. This means that if we note $(M_t)_t$ the associated Markov process of a phase-type distribution, then we have

$$P(M_{t+h} = j / M_t = i) = \begin{cases} \lambda_{ij} h + o(h) & \text{if } i \neq j \\ 1 + \lambda_{ii} h + o(h) & \text{if } i = j \end{cases}.$$

The matrix T is called the sub-intensity matrix and t_0 the exit rate vector.

The cumulative distribution function of a phase-type distribution is given by

$$F(x) = 1 - \pi e^{Tx} \mathbf{1}_m,$$

and its density by

$$f(x) = \pi e^{Tx} t_0,$$

where e^{Tx} denote the matrix exponential defined as the matrix serie $\sum_{n=0}^{+\infty} \frac{T^n x^n}{n!}$.

The computation of matrix exponential is studied in details in appendix A.3, but let us notice that when T is a diagonal matrix, the matrix exponential is the exponential of its diagonal terms. Let us note that there also exists discrete phase-type distribution, cf. Bobbio et al. (2003).

11.3.2 Properties

The moments of a phase-type distribution are given by $(-1)^n n! \pi T^{-n} \mathbf{1}$. Since phase-type distributions are platikurtic or light-tailed distributions, the Laplace transform exists

$$\widehat{f}(s) = \pi(-sI_m - T)^{-1} t_0,$$

*matrix such that its row sums are equal to 0 and have positive elements except on its diagonal.

where I_m stands for the $m \times m$ identity matrix.

One property among many is the set of phase-type distributions is dense with the set of positive random variable distributions. Hence, the distribution of any positive random variable can be written as a limit of phase-type distributions. However, a distribution can be represented (exactly) as a phase-type distribution if and only if the three following conditions are verified

- the distribution has a rational Laplace transform;
- the pole of the Laplace transform with maximal real part is unique;
- it has a density which is positive on $\mathbb{R}_+^\star$.

11.3.3 Special cases

Here are some examples of distributions, which can be represented by a phase-type distribution

- exponential distribution $\mathcal{E}(\lambda)$: $\pi = 1$, $T = -\lambda$ and $m = 1$.
- generalized Erlang distribution $\mathcal{G}\left(n, (\lambda_i)_{1 \leq i \leq n}\right)$:

$$\pi = (1, 0, \ldots, 0),$$

$$T = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \ddots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \lambda_{n-1} \\ 0 & 0 & 0 & 0 & -\lambda_n \end{pmatrix},$$

  and $m = n$.

- a mixture of exponential distribution of parameter $(p_i, \lambda_i)_{1 \leq i \leq n}$:

$$\pi = (p_1, \ldots, p_n),$$

$$T = \begin{pmatrix} -\lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & 0 & \ddots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & -\lambda_n \end{pmatrix},$$

  and $m = n$.

- a mixture of 2 (or k) Erlang distribution $\mathcal{G}(n_i, \lambda_i)_{i=1,2}$ with parameter p_i :

$$\pi = (\underbrace{p_1, 0, \ldots, 0}_{n_1}, \underbrace{p_2, 0, \ldots, 0}_{n_2}),$$

$$T = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \ddots & \lambda_1 & 0 & \ddots & 0 & 0 \\ 0 & 0 & -\lambda_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & -\lambda_2 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & \lambda_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2 \end{pmatrix},$$

and $m = n_1 + n_2$.

11.3.4 Estimation

The estimation based on moments can be a starting point for parameters, but according to Feldmann & Whitt (1996) the fit is very poor. Feldmann & Whitt (1996) proposes a recursive algorithm matching theoretical quantiles and empirical quantiles. They illustrates their method with the Weibull and the Pareto distribution by a mixture of exponential distributions.

First Asmussen et al. (1996) and then Lee & Lin (2008) fit phase-type distribution with the EM algorithm. Lee & Lin (2008) also investigates goodness of fit and graphical comparison of the fit. Lee & Lin (2008) focuses on mixture of Erlang distributions while Asmussen et al. (1996) provides an algorithm for general phase-type distributions. Lee & Lin (2008) illustrates their algorithm with an uniform, a Weibull, a Pareto and log-normal distributions.

11.3.5 Random generation

From Neuts (1981), we have the following algorithm to generate phase-type distributed random variate. Let s be the state of the underlying Markov chain.

- S initialized from the discrete distribution characterized by π
- X initialized to 0
- **while** $S \neq 0$ **do**
  - generate U from an uniform distribution,
  - $X = X - \frac{1}{\hat{\lambda}_{ij}} \log(U)$,
  - generate S from a discrete distribution characterized by the row $\hat{\Lambda}$

where $\hat{\Lambda}$ is the transition matrix defined by

$$\hat{\lambda}_{ij} = \begin{cases} 1 & \text{if } i = j = 1 \\ 0 & \text{if } i = 1 \text{ and } j \neq 1 \\ 0 & \text{if } i > 1 \text{ and } j = i \\ \frac{\lambda_{ij}}{-\lambda_{ii}} & \text{if } i > 1 \text{ and } j \neq i \end{cases}.$$

11.3.6 Applications

NEED REFERENCE

11.4 Exponential family

11.4.1 Characterization

Clark & Thayer (2004) defines the exponential family by the following density or mass probability function

$$f(x) = e^{d(\theta)e(x)+g(\theta)+h(x)},$$

where d, e, g and h are known functions and θ the vector of paremeters. Let us note that the support of the distribution can be $\mathbb{R}$ or $\mathbb{R}_+$ or $\mathbb{N}$. This form for the exponential family is called the natural form.

When we deal with generalized linear models, we use the natural form of the exponential family, which is

$$f(x) = e^{\frac{\theta x - b(\theta)}{a(\phi)}+c(x,\phi)},$$

where a, b, c are known functions and θ, ϕ^* denote the parameters. This form is derived from the previous by setting $d(\theta) = \theta$, $e(x) = x$ and adding a dispersion parameter ϕ.

Let μ be the mean of the variable of an exponential family distribution. We have $\mu = \tau(\theta)$ since ϕ is only a dispersion parameter. The mean value form of the exponential family is

$$f(x) = e^{\frac{\tau^{-1}(\mu)x - b(\tau^{-1}(\mu))}{a(\phi)}+c(x,\phi)}.$$

11.4.2 Properties

For the exponential family, we have $E(X) = \mu = b'(\theta)$ and $Var(X) = a(\phi)V(\mu) = a(\phi)b''(\theta)$ where V is the unit variance function. The skewness is given by $\gamma_3(X) = \frac{dV}{d\mu}(\mu)\sqrt{\frac{a(\phi)}{V(\mu)}} = \frac{b^{(3)}(\theta)a(\phi)^2}{Var(Y)^{3/2}}$, while the kurtosis is $\gamma_4(X) = 3 + \left[\frac{d^2V}{d\mu^2}(\mu)V(\mu) + \left(\frac{dV}{d\mu}(\mu)\right)^2\right]\frac{a(\phi)}{V(\mu)} = 3 + \frac{b^{(4)}(\theta)a(\phi)^3}{Var(Y)^2}$.

The property of uniqueness is the fact that the variance function V uniquely identifies the distribution.

11.4.3 Special cases

The exponential family of distributions in fact contains the most frequently used distributions. Here are the corresponding parameters, listed in a table:

*the canonic and the dispersion parameters.

Law	Distribution	θ	ϕ	Expectation	Variance
Normal $\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$\mu = \theta$	1
Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}$	$-\frac{\beta}{\alpha} = \frac{1}{\mu}$	$\frac{1}{\alpha}$	$\mu = -\frac{1}{\theta}$	μ^2
Inverse Normal $\mathcal{I}(\mu, \lambda)$	$\sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}$	$-\frac{1}{2\mu^2}$	$\frac{1}{\lambda}$	$\mu = (-2\theta)^{-\frac{1}{2}}$	μ^3
Bernoulli $\mathcal{B}(\mu)$	$\mu^x (1-\mu)^{1-x}$	$\log(\frac{\mu}{1-\mu})$	1	$\mu = \frac{e^\theta}{1+e^\theta}$	$\mu(1-\mu)$
Poisson $\mathcal{P}(\mu)$	$\frac{\mu^x}{x!} e^{-\mu}$	$\log(\mu)$	1	$\mu = e^\theta$	μ
Overdispersed Poisson $\mathcal{P}(\phi, \mu)$	$\frac{\mu^{\frac{x}{\phi}}}{\frac{x}{\phi}!} e^{-\mu}$	$\log(\mu)$	ϕ	ϕe^θ	$\phi\mu$

11.4.4 Estimation

The log likelihood equations are

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} \frac{X_i}{a(\phi)} = \frac{b'(\theta)}{a(\phi)} \\ \frac{1}{n}\sum_{i=1}^{n} \frac{\theta X_i a'(\phi)}{a^2(\phi)} - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial c}{\partial \phi}(X_i, \phi) = b(\theta)\frac{a'(\phi)}{a^2(\phi)} \end{cases},$$

for a sample $(X_i)_i$.

11.4.5 Random generation

NEED REFERENCE

11.4.6 Applications

GLM, credibility theory, lehman scheffe theorem

11.5 Elliptical distribution

11.5.1 Characterization

TODO

11.5.2 Properties

TODO

11.5.3 Special cases

11.5.4 Estimation

TODO

11.5.5 Random generation

TODO

11.5.6 Applications

Chapter 12

Multivariate distributions

12.1 Multinomial

12.2 Multivariate normal

12.3 Multivariate elliptical

12.4 Multivariate uniform

12.5 Multivariate student

12.6 Kent distribution

12.7 Dirichlet distribution

12.7.1 Characterization

TODO

12.7.2 Properties

TODO

12.7.3 Estimation

TODO

12.7.4 Random generation

TODO

12.7.5 Applications

TODO

12.8 Von Mises Fisher

12.9 Evens

Chapter 13

Misc

13.1 MBBEFD distribution

The MBBEFD distribution comes from the actuarial science due to Bernegger (1997). MBBEFD stands for Maxwell-Boltzmann, Bore-Einstein and Fermi-Dirac distribution.

13.1.1 Characterization

The MBBEFD distribution is characterized by the following distribution function.

$$F(x) = \begin{cases} a\left(\frac{a+1}{a+b^x} - 1\right) & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases},$$

for $x \in \mathbb{R}_+$. The MBBEFD distribution is a mixed distribution of a continuous distribution on $x \in]0,1[$ and a discrete distribution in $x = 1$. We have a mass probability for $x = 1$

$$p = 1 - F(1) = \frac{(a+1)b}{a+b}.$$

The parameters (a, b) are defined on a wide set of intervals, which are not trivial: $]-1, 0[\times]1, +\infty[$ and $]-\infty, -1[\cup]0, +\infty[\times]0, 1[$. The shape of the distribution function F has the following properties

- for $(a, b) \in I_1 =]-1, 0[\times]1, +\infty[$, F is concave,
- for $(a, b) \in I_2 =]-\infty, -1[\times]0, 1[$, F is concave,
- for $(a, b) \in I_3 =]0, b[\times]0, 1[$, F is concave,
- for $(a, b) \in I_4 = [b, 1[\times]0, 1[$, F is convex then concave,
- for $(a, b) \in I_4 = [1, +\infty[\times]0, 1[$, F is convex.

There is no usual density but if we use the Dirac function δ, we can define a function f such that

$$f(x) = \frac{-a(a+1)b^x \ln(b)}{(a+b^x)^2} \mathbb{1}_{]0,1[}(x) + \delta_1.$$

which is a mix between a mass probability and a density functions.

13.1.2 Special cases

TODO

13.1.3 Properties

TODO

13.1.4 Estimation

TODO

13.1.5 Random generation

TODO

13.1.6 Applications

13.2 Cantor distribution

TODO

13.3 Tweedie distribution

TODO

Bibliography

Arnold, B. C. (1983), 'Pareto distributions', *International Co-operative Publishing House* **5**. 30, 88, 93, 95, 96

Asmussen, S., Nerman, O. & Olsson, M. (1996), 'Fitting phase-type distributions via the em algorithm', *Scandinavian Journal of Statistics* **23**(4), 419–441. 129

Barndorff-Nielsen, O. E. (1977), 'Exponentially decreasing distributions for the logarithm of particle size', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **353**(1674), 401–419. 120

Barndorff-Nielsen, O. E. (1997*a*), 'Normal inverse Gaussian distributions and stochastic volatility modelling', *Scandinavian Journal of Statistics* **24**(1), 1–13. 124

Barndorff-Nielsen, O. E. (1997*b*), 'Processes of normal inverse gaussian type', *Finance and Stochastics* **2**(1), 41–68. 124

Barndorff-Nielsen, O. E. & Halgreen, O. (1977), 'Infinite divisibility of the hyperbolic and generalized inverse gaussian distribution', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **38**(4), 309–311. 124

Barndorff-Nielsen, O. E. & Prause, K. (2001), 'Apparent scaling', *Finance and Stochastics* **5**(1), 103–113. 124

Bernegger, S. (1997), 'The swiss re exposure curves and the mbbefd distribution class', *Astin Bull.* **27**(1), 99–111. 135

Bibby, B. M. & Sorensen, M. (1997), 'A hyperbolic diffusion model for stock prices', *Finance & Stochastics 2* pp. 25–41. 124

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3). 51

Bobbio, A., Horvath, A., Scarpa, M. & Telek, M. (2003), 'Acyclic discrete phase type distributions: properties and a parameter estimation algorithm', *performance evaluation* **54**, 1–32. 127

Breymann, W., Dias, A. & Embrechts, P. (2003), 'Dependence Structures for Multivariate High–Frequency Data in Finance', *Quantitative Finance* **3**(1), 1–14. 124

Breymann, W. & Lüthi, D. (2008), *ghyp: A package on generalized hyperbolic distributions*, Institute of Data Analysis and Process Design. 54, 117

Brigo, D., Mercurio, F., Rapisarda, F. & Scotti, R. (2002), 'Approximated moment-matching dynamics for basket-options simulation', *Product and Business Development Group,Banca IMI, SanPaolo IMI Group* . 53

Cacoullos, T. & Charalambides, C. (1975), 'On minimum variance unbiased estimation for truncated binomial and negative binomial distributions', *Annals of the Institute of Statistical Mathematics* **27**(1). 12, 21, 24

Carrasco, J. M. F., Ortega, E. M. M. & Cordeiro, G. M. (2008), 'A generalized modified weibull distribution for lifetime modeling', *Computational Statistics and Data Analysis* **53**, 450–462. 72

Chambers, J. M., Mallows, C. L. & Stuck, B. W. (1976), 'A method for simulating stable random variables', *Journal of the American Statistical Association,* . 126

Chen, Y., Härdle, W. & Jeong, S.-O. (2005), *Nonparametric Risk Management with Generalized Hyperbolic Distributions*, Vol. 1063 of *Preprint / Weierstraß–Institut für Angewandte Analysis und Stochastik*, WIAS, Berlin. 124

Clark, D. R. & Thayer, C. A. (2004), 'A primer on the exponential family of distributions', *2004 call paper program on generalized linear models* . 130

Cont, R. & Tankov, P. (2003), *Financial Modelling with Jump Processes*, Chapman & Hall CRC Financial Mathematics Series. 124

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the Em algorithm', *Journal of the Royal Statistical Society* **39**(1), 1–38. 122

Dilip B. Madan, Peter Carr & Eric C. Chang (1998), 'The variance gamma process and option pricing', *European Finance Review* **2**, 79–105. 124

Dutang, C. (2008), *randtoolbox: Generating and Testing Random Numbers*. 36

Eberlein, E. & Keller, U. (1995), 'Hyperbolic distributions in finance.', *Bernoulli 1* pp. 281–299. 124

Eberlein, E., Keller, U. & Prause, K. (1998), 'New insights into smile, mispricing and value at risk measures.', *Journal of Business* **71**, 371–405. 124

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events*, Springer. 99, 100, 101, 114, 115

Feldmann, A. & Whitt, W. (1996), 'Fitting mixtures of exponentials to long tail distributions to analyze network performance models', *AT&T Laboratory Research* . 129

Fergusson, K. & Platen, E. (2006), 'On the distributional characterization of log-returns of a world stock index', *Applied Mathematical Finance* **13**(1), 19–38. 124

Froot, K. A. & O'Connell, P. G. J. (2008), 'On the pricing of intermediated risks: Theory and application to catastrophe reinsurance', *Journal of banking & finance* **32**, 69–85. 95

Gomes, O., Combes, C. & Dussauchoy, A. (2008), 'Parameter estimation of the generalized gamma distribution', *Mathematics and Computers in Simulation* **79**, 955–963. 67

Haahtela, T. (2005), Extended binomial tree valuation when the underlying asset distribution is shifted lognormal with higher moments. Helsinki University of Technology. 53

Haddow, J. E., Palomaki, G. E., Knight, G. J., Cunningham, G. C., Lustig, L. S. & Boyd, P. A. (1994), 'Reducing the need for amniocentesis in women 35 years of age or older with serum markers for screening', *New England Journal of Medicine* **330**(16), 1114–1118. 11

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous univariate distributions*, John Wiley. 5

Jones, M. C. (2009), 'Kumaraswamy's distribution: A beta-type distribution with some tractability advantages', *Statistical Methodology* **6**, 70. 45, 46

Kassberger, S. (2007), 'Efficient Portfolio Optimization in the Multivariate Generalized Hyperbolic Framework', *SSRN eLibrary* . 125

Kassberger, S. & Kiesel, R. (2006), 'A fully parametric approach to return modelling and risk management of hedge funds', *Financial markets and portfolio management* **20**(4), 472–491. 124

Klugman, S. A., Panjer, H. H. & Willmot, G. (2004), *Loss Models: From Data to Decisions*, 2 edn, Wiley, New York. 50, 68, 96, 104

Knuth, D. E. (2002), *The Art of Computer Programming: seminumerical algorithms*, Vol. 2, 3rd edition edn, Massachusetts: Addison-Wesley. 15

Lee, S. C. & Lin, X. S. (2008), 'Modeling and evaluating insurance losses via mixtures of erlang', *North American Actuarial Journal* . 129

Li, Q. & Yu, K. (2008), 'Inference of non-centrality parameter of a truncated non-central chi-squared distribution', *Journal of Statistical Planning and Inference* . 79

Limpert, E., Stahel, W. A. & Abbt, M. (2001), 'Log-normal distributions across the sciences: Keys and clues', *Bioscience* **51**(5). 51

Matsumoto, M. & Nishimura, T. (1998), 'Mersenne twister: A 623-dimensionnally equidistributed uniform pseudorandom number generator', *ACM Trans. on Modelling and Computer Simulation* **8**(1), 3–30. 36

Mačutek, J. (2008), 'A generalization of the geometric distribution and its application in quantitative linguistics', *Romanian Reports in Physics* **60**(3), 501–509. 20

McNeil, A. J., Frey, R. & Embrechts, P. (2005*a*), *Quantitative risk management: Concepts, techniques and tools*, Princeton University Press, Princeton. 123

McNeil, A. J., Frey, R. & Embrechts, P. (2005*b*), *Quantitative risk management: Concepts, techniques and tools*, Princeton University Press, Princeton. 124

Meng, X.-L. & Rubin, D.-B. (1993), 'Maximum likelihood estimation via the ECM algorithm: A general framework', *Biometrika* **80**(2), 267–278. 122, 123

Moler, C. & Van Loan, C. (2003), 'Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later', *SIAM review* **45**(1), 300. 144

Nadarajah, S. & Kotz, S. (2003), 'A generalized beta distribution ii', *Statistics on the internet* . 42, 43

Neuts, M. F. (1981), Generating random variates from a distribution of phase-type, *in* 'Winter Simulation Conference'. 129

Nolan, J. P. (2009), *Stable Distributions - Models for Heavy Tailed Data*, Birkhäuser, Boston. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan. 125

Paolella, M. (2007), *Intermediate probability: A computational approach*, Wiley, Chichester. 118

Patard, P.-A. (2007), 'Outils numeriques pour la simulation monte carlo des produits derives complexes', *Bulletin français d'actuariat* **7**(14), 74–117. 49

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *Annals of Statistics* **3**, 119–131. 90

Prause, K. (1999), The Generalized Hyperbolic Model: Estimation Financial Derivatives and Risk Measures, PhD thesis, Universität Freiburg i. Br., Freiburg i. Br. 118

Raible, S. (2000), Lévy processes in finance: Theory, numerics, and empirical facts, PhD thesis, Universität Freiburg i. Br. 124

Rytgaard, M. (1990), 'Estimation in the pareto distribution', *Astin Bull.* **20**(2), 201–216. 93

Saporta, G. (1990), *Probabilités analyse des données et statistique*, Technip. 14

Saxena, K. M. L. & Alam, K. (1982), 'Estimation of the non-centrality parameter of a chi-squared distribution', *The Annals of Statistics* **10**(3), 1012–1016. 79

Simon, L. J. (1962), An introduction to the negative binomial distribution and its applications, *in* 'Casualty Actuarial Society', Vol. XLIX. 23

Singh, A. K., Singh, A. & Engelhardt, M. (1997), 'The lognormal distribution in environmental applications', *EPA Technology Support Center Issue* . 51

Stein, W. E. & Keblis, M. F. (2008), 'A new method to simulate the triangular distribution', *Mathematical and computer modelling* . 38

Tate, R. F. & Goen, R. L. (1958), 'Minimum variance unbiased estimation for the truncated poisson distribution', *The Annals of Mathematical Statistics* **29**(3), 755–765. 16, 17

Thomas, D. G. & Gart, J. J. (1971), 'Small sample performance of some estimators of the truncated binomial distribution', *Journal of the American Statistical Association* **66**(333). 12

Venter, G. (1983), Transformed beta and gamma distributions and aggregate losses, *in* 'Casualty Actuarial Society'. 67

Wimmer, G. & Altmann, G. (1999), *Thesaurus of univariate discrete probability distributions*, STAMM Verlag GmbH Essen. 5

Yu, Y. (2009), 'Complete monotonicity of the entropy in the central limit theorem for gamma and inverse gaussian distributions', *Statistics and probability letters* **79**, 270–274. 53

Appendix A

Mathematical tools

A.1 Basics of probability theory

TODO

A.1.1 Characterising functions

For a discrete distribution, one may use the probability generating function to characterize the distribution, if it exists or equivalently the moment generating function. For a continuous distribution, we generally use only the moment generating function. The moment generating function is linked to the Laplace transform of a distribution. When dealing with continuous distribution, we also use the characteristic function, which is related to the Fourrier transform of a distribution, see table below for details.

Probability generating function $G_X(z)$	Moment generating function $M_X(t)$	<=>	Laplace Transform $L_X(s)$	Characteristic function $\phi_X(t)$	<=>	Fourrier transform
$\mathbb{E}\left[z^X\right]$	$\mathbb{E}\left[e^{tX}\right]$	<=>	$\mathbb{E}\left[e^{-sX}\right]$	$\mathbb{E}\left[e^{itX}\right]$	<=>	$\mathbb{E}\left[e^{-itX}\right]$

We have the following results

- $\forall k \in \mathbb{N}, X$ discrete random variable $, P(X = k) = \frac{1}{k!}\frac{d^k G_X(t)}{dt^k}|_{t=0}; E(X \dots (X-k)) = \frac{d^k G_X(t)}{dt^k}|_{t=1}$
- $\forall X$ continuous random variable $E(X^k) = \frac{d^k M_X(t)}{dt^k}|_{t=0}$

A.2 Common mathematical functions

In this section, we recall the common mathematical quantities used in all this guide. By definition, we have

A.2.1 Integral functions

- gamma function: $\forall a > 0,\ \Gamma(a) = \int_0^{+\infty} x^{a-1}e^{-x}dx$
- incomplete gamma function: lower $\forall a, x > 0,\ \gamma(a,x) = \int_0^x y^{a-1}e^{-y}dy$ and upper $\Gamma(a,x) = \int_x^{+\infty} y^{a-1}e^{-y}dy$;
- results for gamma function $\forall n \in \mathbb{N}^\star,\ \Gamma(n) = (n-1)!,\ \Gamma(0) = 1,\ \Gamma(\frac{1}{2}) = \sqrt{\pi},\ \forall a > 1,\ \Gamma(a) = (a-1)\Gamma(a-1)$
- beta function: $\forall a, b > 0,\ \beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$,
- incomplete beta function $\forall 1 \geq u \geq 0, \beta(a,b,u) = \int_0^u x^{a-1}(1-x)^{b-1}dx$;
- results for beta function $\forall a, b > 0, \beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
- digamma function: $\forall x > 0, \psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$
- trigamma function: $\forall x > 0, \psi_1(x) = \frac{\Gamma''(x)}{\Gamma(x)}$
- error function : $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}dt$

A.2.2 Factorial functions

- factorial : $\forall n \in \mathbb{N}, n! = n \times (n-1) \ldots 2 \times 1$
- rising factorial : $\forall n, m \in \mathbb{N}^2, m^{(n)} = m \times (m+1) \ldots (m+n-2) \times (m+n-1) = \frac{\Gamma(n+m)}{\Gamma(n)}$
- falling factorial: $\forall n, m \in \mathbb{N}^2, (m)_n = m \times (m-1) \ldots (m-n+2) \times (m-n+1) = \frac{\Gamma(m)}{\Gamma(m-n)}$
- combination number : $\forall n, p \in \mathbb{N}^2,\ C_n^p = \frac{n!}{p!(n-p)!}$
- arrangement number $A_n^p = \frac{n!}{(n-p)!}$
- Stirling number of the first kind : coefficients ${}_1S_n^k$ of the expansion of $(x)_n = \sum_{k=0}^n {}_1S_n^k x^k$ or defined by the recurrence ${}_1S_n^k = (n-1) \times {}_1S_{n-1}^k + {}_1S_{n-1}^{k-1}$ with ${}_1S_n^0 = \delta_{n0}$ and ${}_1S_0^1 = 0$.
- Stirling number of the second kind : coefficients ${}_2S_n^k$ of the expansion $\sum_{k=0}^n {}_2S_n^k (x)_k = x^n$ or defined by the recurrence ${}_2S_n^k = {}_2S_{n-1}^{k-1} + k \times {}_2S_{n-1}^k$ with ${}_2S_n^1 = {}_2S_n^n = 1$.

A.2.3 Serie functions

- Riemann's zeta function : $\forall s > 1, \zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}$
- Jonquière's function : $\forall s > 1, \forall z > 0, Li_s(z) = \sum_{n=1}^{+\infty} \frac{z^n}{n^s}$
- hypergeometric function : $\forall a, b, c \in \mathbb{N}, \forall z \in \mathbb{R}, {}_1F_1(a, b, z) = \sum_{n=0}^{+\infty} \frac{a^{(n)} z^n}{b^{(n)} n!}$, ${}_2F_1(a, b, c, z) = \sum_{n=0}^{+\infty} \frac{a^{(n)} b^{(n)}}{c^{(n)}} \frac{z^n}{n!}$ and ${}_3F_1(a, b, c, d, e, z) = \sum_{n=0}^{+\infty} \frac{a^{(n)} b^{(n)} c^{(n)}}{d^{(n)} e^{(n)}} \frac{z^n}{n!}$.
- Bessel's functions verify the following ODE: $x^2 y'' + xy' + (x^2 - \alpha^2)y = 0$. We define the Bessel function of the 1st kind by $J_\alpha(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+\alpha+1)} \left(\frac{x}{2}\right)^{2n+\alpha}$ and of the 2nd kind $Y_\alpha(x) = \frac{J_\alpha(x) cos(\alpha\pi) - J_{-\alpha}(x)}{sin(\alpha\pi)}$.
- Hankel's function: $H_\alpha^{(1)}(x) = J_\alpha(x) + iY_\alpha(x)$
- Bessel's modified function $I_\alpha(x) = i^{-\alpha} J_\alpha(ix) = \sum_{k=0}^{+\infty} \frac{(x/2)^{2k+\alpha}}{k!\Gamma(\alpha+k+1)}$ and $K_\alpha(x) = \frac{\pi}{2} i^{\alpha+1} H_\alpha^{(1)}(x) = \frac{1}{2} \int_0^\infty y^{\alpha-1} e^{-\frac{x}{2}(y+y^{-1})} dy$
- Laguerre's polynomials: $L_n(x) = \frac{e^x}{n!} \frac{d^n(e^x x^n)}{dx^n} = \sum_{i=0}^n (-1)^i C_n^{n-i} \frac{x^i}{i!}$
- generalized Laguerre's polynomials: $L_n^{(\alpha)}(x) = \frac{e^x}{n! x^\alpha} \frac{d^n(e^x x^{n+\alpha})}{dx^n} = \sum_{i=0}^n (-1)^i C_{n+\alpha}^{n-i} \frac{x^i}{i!}$

A.2.4 Miscellanous

- Dirac function: $\forall x > 0, \delta_{x_0}(x) = \begin{cases} +\infty & \text{si } x = x_0 \\ 0 & \text{sinon} \end{cases}$ et
- heavyside function : $H_{x_0}(x) = \begin{cases} 0 & \text{si } x < x_0 \\ \frac{1}{2} & \text{si } x = x_0 \\ 1 & \text{sinon} \end{cases}$
- Cantor function : $\forall x \in [0, 1], F_n(x) = \begin{cases} x & \text{si } n = 0 \\ \frac{1}{2} F_{n-1}(3x) & \text{si } n \neq 0 \text{ et } 0 \geq x \geq \frac{1}{3} \\ \frac{1}{2} & \text{si } n \neq 0 \text{ et } \frac{1}{3} \geq x \geq \frac{2}{3} \\ \frac{1}{2} + \frac{1}{2} F_{n-1}(3(x - \frac{2}{3})) & \text{si } n \neq 0 \text{ et } \frac{2}{3} \geq x \geq 1 \end{cases}$

A.3 Matrix exponential

Now let us consider the problem of computing e^{Qu}. We recall that

$$e^{Qu} = \sum_{n=0}^{+\infty} \frac{Q^n u^n}{n!}.$$

There are various methods to compute the matrix exponential, Moler & Van Loan (2003) makes a deep analysis of the efficiency of different methods. In our case, we choose a decomposition method. We diagonalize the $n \times n$ matrix Q and use the identity

$$e^{Qu} = Pe^{Du}P^{-1},$$

where D is a diagonal matrix with eigenvalues on its diagonal and P the eigenvectors. We compute

$$e^{Qu} = \sum_{l=1}^{m} e^{\lambda_l u} \underbrace{PM_lP^{-1}}_{C_l},$$

where λ_i stands for the eigenvalues of Q, P the eigenvectors and $M_l = (\delta_{il}\delta_{lj})_{ij}$ (δ_{ij} is the symbol Kronecker, i.e. equals to zero except when $i = j$). As the matrix M_l is a sparse matrix with just a 1 on the l^{th} term of its diagonal. The constant C_i can be simplified. Indeed, if we denote by X_l the l^{th} column of the matrix P (i.e. the eigenvector associated to the eigenvalue λ_l) and Y_l the l^{th} row of the matrix P^{-1}, then we have

$$C_l \stackrel{\triangle}{=} PM_lP^{-1} = X_l \otimes Y_l.$$

Despite Q is not obligatorily diagonalizable, this procedure will often work, since Q may have a complex eigenvalue (say λ_i). In this case, C_i is complex but as e^{Qu} is real, we are ensured there is $j \in [\![1, \ldots, m]\!]$, such that λ_j is the conjugate of λ_l. Thus, we get

$$e^{\lambda_i u}C_i + e^{\lambda_j u}C_j = 2cos(\Im(\lambda_i)u)e^{\Re\lambda_i u}\Re(X_i \otimes Y_i) - 2sin(\Im(\lambda_i)u)e^{\Re\lambda_i u}\Im(X_i \otimes Y_i) \in \mathbb{R},$$

where $\Re$ and $\Im$ stands resp. for the real and the imaginary part.

A.4 Kronecker product and sum

The Kronecker product $A \otimes B$ is defined as the $mn \times mn$ matrix

$$A \otimes B = (A_{i_1,j_1}B_{i_2,j_2})_{i_1i_2,j_1j_2},$$

when A is a $m \times m$ matrix of general term $(A_{i_1,j_1})_{i_1,j_1}$ and B a $n \times n$ matrix of general term $(B_{i_2,j_2})_{i_2,j_2}$. Note that the Kronecker can also be defined for non-square matrixes.

The Kronecker sum $A \oplus B$ is given by the $mn \times mn$ matrix

$$A \otimes B = A \otimes I_m + B \otimes I_n,$$

where I_m and I_n are the identity matrixes of size m and n. This definition is right only for square matrixes A and B.

Contents

Introduction **4**

I Discrete distributions **6**

1 Classic discrete distribution **7**

1.1 Discrete uniform distribution 7

1.1.1 Characterization 7

1.1.2 Properties 8

1.1.3 Estimation 8

1.1.4 Random generation 8

1.1.5 Applications 8

1.2 Bernoulli/Binomial distribution 8

1.2.1 Characterization 8

1.2.2 Properties 9

1.2.3 Estimation 10

1.2.4 Random generation 11

1.2.5 Applications 11

1.3 Zero-truncated or zero-modified binomial distribution 11

1.3.1 Characterization 11

1.3.2 Properties . . . 12
1.3.3 Estimation . . . 12
1.3.4 Random generation . . . 12
1.3.5 Applications . . . 13
1.4 Quasi-binomial distribution . . . 13
1.4.1 Characterization . . . 13
1.4.2 Properties . . . 13
1.4.3 Estimation . . . 13
1.4.4 Random generation . . . 13
1.4.5 Applications . . . 13
1.5 Poisson distribution . . . 14
1.5.1 Characterization . . . 14
1.5.2 Properties . . . 14
1.5.3 Estimation . . . 15
1.5.4 Random generation . . . 15
1.5.5 Applications . . . 15
1.6 Zero-truncated or zero-modified Poisson distribution . . . 15
1.6.1 Characterization . . . 15
1.6.2 Properties . . . 16
1.6.3 Estimation . . . 16
1.6.4 Random generation . . . 17
1.6.5 Applications . . . 17
1.7 Quasi-Poisson distribution . . . 17
1.7.1 Characterization . . . 18

1.7.2 Properties 18
1.7.3 Estimation 18
1.7.4 Random generation 18
1.7.5 Applications 18
1.8 Geometric distribution 18
1.8.1 Characterization 18
1.8.2 Properties 19
1.8.3 Estimation 19
1.8.4 Random generation 19
1.8.5 Applications 20
1.9 Zero-truncated or zero-modified geometric distribution 20
1.9.1 Characterization 20
1.9.2 Properties 20
1.9.3 Estimation 21
1.9.4 Random generation 21
1.9.5 Applications 22
1.10 Negative binomial distribution 22
1.10.1 Characterization 22
1.10.2 Characterization 22
1.10.3 Properties 23
1.10.4 Estimation 23
1.10.5 Random generation 23
1.10.6 Applications 23
1.11 Zero-truncated or zero-modified negative binomial distribution 24

1.11.1 Characterization 24
1.11.2 Properties 24
1.11.3 Estimation 24
1.11.4 Random generation 25
1.11.5 Applications 25
1.12 Pascal distribution 25
1.12.1 Characterization 25
1.12.2 Properties 25
1.12.3 Estimation 26
1.12.4 Random generation 26
1.12.5 Applications 26
1.13 Hypergeometric distribution 26
1.13.1 Characterization 26
1.13.2 Properties 26
1.13.3 Estimation 26
1.13.4 Random generation 26
1.13.5 Applications 26

2 Not so-common discrete distribution **27**

2.1 Conway-Maxwell-Poisson distribution 27
2.1.1 Characterization 27
2.1.2 Properties 27
2.1.3 Estimation 27
2.1.4 Random generation 27
2.1.5 Applications 28

2.2 Delaporte distribution 28
2.2.1 Characterization 28
2.2.2 Properties 28
2.2.3 Estimation 28
2.2.4 Random generation 28
2.2.5 Applications 28
2.3 Engen distribution 28
2.3.1 Characterization 28
2.3.2 Properties 28
2.3.3 Estimation 28
2.3.4 Random generation 29
2.3.5 Applications 29
2.4 Logaritmic distribution 29
2.4.1 Characterization 29
2.4.2 Properties 29
2.4.3 Estimation 29
2.4.4 Random generation 29
2.4.5 Applications 29
2.5 Sichel distribution 29
2.5.1 Characterization 29
2.5.2 Properties 29
2.5.3 Estimation 30
2.5.4 Random generation 30
2.5.5 Applications 30

2.6 Zipf distribution . . . 30
2.6.1 Characterization . . . 30
2.6.2 Properties . . . 30
2.6.3 Estimation . . . 30
2.6.4 Random generation . . . 30
2.6.5 Applications . . . 31
2.7 The generalized Zipf distribution . . . 31
2.7.1 Characterization . . . 31
2.7.2 Properties . . . 31
2.7.3 Estimation . . . 31
2.7.4 Random generation . . . 31
2.7.5 Applications . . . 31
2.8 Rademacher distribution . . . 31
2.8.1 Characterization . . . 31
2.8.2 Properties . . . 31
2.8.3 Estimation . . . 31
2.8.4 Random generation . . . 32
2.8.5 Applications . . . 32
2.9 Skellam distribution . . . 32
2.9.1 Characterization . . . 32
2.9.2 Properties . . . 32
2.9.3 Estimation . . . 32
2.9.4 Random generation . . . 32
2.9.5 Applications . . . 32

2.10 Yule distribution 32
2.10.1 Characterization 32
2.10.2 Properties 32
2.10.3 Estimation 33
2.10.4 Random generation 33
2.10.5 Applications 33
2.11 Zeta distribution 33
2.11.1 Characterization 33
2.11.2 Properties 33
2.11.3 Estimation 33
2.11.4 Random generation 33
2.11.5 Applications 33

II Continuous distributions **34**

3 Finite support distribution **35**

3.1 Uniform distribution 35
3.1.1 Characterization 35
3.1.2 Properties 35
3.1.3 Estimation 36
3.1.4 Random number generation 36
3.1.5 Applications 36
3.2 Triangular distribution 36
3.2.1 Characterization 36
3.2.2 Properties 37

3.2.3 Estimation . . . 37
3.2.4 Random generation . . . 37
3.2.5 Applications . . . 38
3.3 Beta type I distribution . . . 38
3.3.1 Characterization . . . 38
3.3.2 Special cases . . . 40
3.3.3 Properties . . . 40
3.3.4 Estimation . . . 40
3.3.5 Random generation . . . 41
3.3.6 Applications . . . 41
3.4 Generalized beta I distribution . . . 41
3.4.1 Characterization . . . 41
3.4.2 Properties . . . 42
3.4.3 Estimation . . . 42
3.4.4 Random generation . . . 42
3.4.5 Applications . . . 42
3.5 Generalization of the generalized beta I distribution . . . 42
3.5.1 Characterization . . . 42
3.5.2 Special cases . . . 43
3.5.3 Properties . . . 44
3.5.4 Estimation . . . 45
3.5.5 Random generation . . . 45
3.5.6 Applications . . . 45
3.6 Kumaraswamy distribution . . . 45

3.6.1 Characterization 45
3.6.2 Properties 46
3.6.3 Estimation 46
3.6.4 Random generation 46
3.6.5 Applications 46

4 The Gaussian family **47**

4.1 The Gaussian (or normal) distribution 47
4.1.1 Characterization 47
4.1.2 Properties 48
4.1.3 Estimation 48
4.1.4 Random generation 49
4.1.5 Applications 49
4.2 Log normal distribution 50
4.2.1 Characterization 50
4.2.2 Properties 50
4.2.3 Estimation 51
4.2.4 Random generation 51
4.2.5 Applications 51
4.3 Shifted log normal distribution 52
4.3.1 Characterization 52
4.3.2 Properties 52
4.3.3 Estimation 52
4.3.4 Random generation 52
4.3.5 Applications 53

4.4 Inverse Gaussian distribution 53
4.4.1 Characterization 53
4.4.2 Properties 53
4.4.3 Estimation 54
4.4.4 Random generation 54
4.4.5 Applications 54
4.5 The generalized inverse Gaussian distribution 54
4.5.1 Characterization 54
4.5.2 Properties 55
4.5.3 Estimation 55
4.5.4 Random generation 55

5 Exponential distribution and its extensions 56

5.1 Exponential distribution 56
5.1.1 Characterization 56
5.1.2 Properties 56
5.1.3 Estimation 57
5.1.4 Random generation 57
5.1.5 Applications 57
5.2 Shifted exponential 59
5.2.1 Characterization 59
5.2.2 Properties 59
5.2.3 Estimation 59
5.2.4 Random generation 60
5.2.5 Applications 60

5.3 Inverse exponential 60
5.3.1 Characterization 60
5.3.2 Properties 60
5.3.3 Estimation 61
5.3.4 Random generation 61
5.3.5 Applications 61
5.4 Gamma distribution 61
5.4.1 Characterization 61
5.4.2 Properties 62
5.4.3 Estimation 62
5.4.4 Random generation 63
5.4.5 Applications 63
5.5 Generalized Erlang distribution 63
5.5.1 Characterization 63
5.5.2 Properties 64
5.5.3 Estimation 64
5.5.4 Random generation 64
5.5.5 Applications 64
5.6 Chi-squared distribution 64
5.7 Inverse Gamma 65
5.7.1 Characterization 65
5.7.2 Properties 65
5.7.3 Estimation 65
5.7.4 Random generation 66

5.7.5 Applications 66
5.8 Transformed or generalized gamma 66
5.8.1 Characterization 66
5.8.2 Properties 67
5.8.3 Estimation 67
5.8.4 Random generation 67
5.8.5 Applications 67
5.9 Inverse transformed Gamma 68
5.9.1 Characterization 68
5.9.2 Properties 68
5.9.3 Estimation 68
5.9.4 Random generation 68
5.9.5 Applications 68
5.10 Log Gamma 69
5.10.1 Characterization 69
5.10.2 Properties 69
5.10.3 Estimation 69
5.10.4 Random generation 69
5.10.5 Applications 69
5.11 Weibull distribution 69
5.11.1 Characterization 69
5.11.2 Properties 70
5.11.3 Estimation 70
5.11.4 Random generation 71

5.11.5 Applications 71

5.12 Inverse Weibull distribution 71

5.12.1 Characterization 71

5.12.2 Properties 72

5.12.3 Estimation 72

5.12.4 Random generation 72

5.12.5 Applications 72

5.13 Laplace or double exponential distribution 72

5.13.1 Characterization 72

5.13.2 Properties 73

5.13.3 Estimation 73

5.13.4 Random generation 73

5.13.5 Applications 74

6 Chi-squared's ditribution and related extensions 75

6.1 Chi-squared distribution 75

6.1.1 Characterization 75

6.1.2 Properties 76

6.1.3 Estimation 76

6.1.4 Random generation 76

6.1.5 Applications 76

6.2 Chi distribution 77

6.2.1 Characterization 77

6.2.2 Properties 77

6.2.3 Estimation 78

6.2.4 Random generation 78
6.2.5 Applications 78
6.3 Non central chi-squared distribution 78
6.3.1 Characterization 78
6.3.2 Properties 79
6.3.3 Estimation 79
6.3.4 Random generation 79
6.3.5 Applications 79
6.4 Non central chi distribution 80
6.4.1 Characterization 80
6.4.2 Properties 80
6.4.3 Estimation 80
6.4.4 Random generation 80
6.4.5 Applications 81
6.5 Inverse chi-squared distribution 81
6.5.1 Properties 81
6.5.2 Estimation 82
6.5.3 Random generation 82
6.5.4 Applications 82
6.6 Scaled inverse chi-squared distribution 82
6.6.1 Characterization 82
6.6.2 Properties 82
6.6.3 Estimation 82
6.6.4 Random generation 82

6.6.5 Applications 83

7 Student and related distributions **84**

7.1 Student t distribution 84

7.1.1 Characterization 84

7.1.2 Properties 85

7.1.3 Estimation 85

7.1.4 Random generation 85

7.1.5 Applications 85

7.2 Cauchy distribution 86

7.2.1 Characterization 86

7.2.2 Characterization 86

7.2.3 Properties 86

7.2.4 Estimation 86

7.2.5 Random generation 87

7.2.6 Applications 87

7.3 Fisher-Snedecor distribution 87

7.3.1 Characterization 87

7.3.2 Properties 87

7.3.3 Estimation 87

7.3.4 Random generation 87

7.3.5 Applications 87

8 Pareto family **88**

8.1 Pareto distribution 88

8.1.1 Characterization 88

8.1.2 Properties 90
8.1.3 Estimation 93
8.1.4 Random generation 94
8.1.5 Applications 95
8.2 Feller-Pareto distribution 96
8.2.1 Characterization 96
8.2.2 Properties 97
8.2.3 Estimation 97
8.2.4 Random generation 97
8.2.5 Applications 97
8.3 Inverse Pareto 98
8.3.1 Characterization 98
8.3.2 Properties 98
8.3.3 Estimation 98
8.3.4 Random generation 98
8.3.5 Applications 98
8.4 Generalized Pareto distribution 99
8.4.1 Characterization 99
8.4.2 Properties 100
8.4.3 Estimation 100
8.4.4 Random generation 101
8.4.5 Applications 102
8.5 Burr distribution 102
8.5.1 Characterization 102

8.5.2 Properties . . . 102
8.5.3 Estimation . . . 103
8.5.4 Random generation . . . 103
8.5.5 Applications . . . 103
8.6 Inverse Burr distribution . . . 104
8.6.1 Characterization . . . 104
8.6.2 Properties . . . 104
8.6.3 Estimation . . . 105
8.6.4 Random generation . . . 105
8.6.5 Applications . . . 105
8.7 Beta type II distribution . . . 105
8.7.1 Characterization . . . 105
8.7.2 Properties . . . 106
8.7.3 Estimation . . . 106
8.7.4 Random generation . . . 107
8.7.5 Applications . . . 107

9 Logistic distribution and related extensions 108

9.1 Logistic distribution . . . 108
9.1.1 Characterization . . . 108
9.1.2 Properties . . . 108
9.1.3 Estimation . . . 108
9.1.4 Random generation . . . 108
9.1.5 Applications . . . 110
9.2 Half logistic distribution . . . 110

9.2.1 Characterization . . . 110
9.2.2 Properties . . . 110
9.2.3 Estimation . . . 110
9.2.4 Random generation . . . 110
9.2.5 Applications . . . 110
9.3 Log logistic distribution . . . 110
9.3.1 Characterization . . . 110
9.3.2 Properties . . . 110
9.3.3 Estimation . . . 110
9.3.4 Random generation . . . 110
9.3.5 Applications . . . 110
9.4 Generalized log logistic distribution . . . 110
9.4.1 Characterization . . . 110
9.4.2 Properties . . . 110
9.4.3 Estimation . . . 110
9.4.4 Random generation . . . 110
9.4.5 Applications . . . 110
9.5 Paralogisitic distribution . . . 110
9.5.1 Characterization . . . 110
9.5.2 Properties . . . 110
9.5.3 Estimation . . . 110
9.5.4 Random generation . . . 110
9.5.5 Applications . . . 110
9.6 Inverse paralogistic distribution . . . 110

9.6.1 Characterization . . . 110
9.6.2 Properties . . . 110
9.6.3 Estimation . . . 110
9.6.4 Random generation . . . 110
9.6.5 Applications . . . 110

10 Extrem Value Theory distributions 111

10.1 Gumbel distribution . . . 111
10.1.1 Characterization . . . 111
10.1.2 Properties . . . 112
10.1.3 Estimation . . . 112
10.1.4 Random generation . . . 112
10.1.5 Applications . . . 112
10.2 Fréchet distribution . . . 113
10.3 Weibull distribution . . . 113
10.4 Generalized extreme value distribution . . . 113
10.4.1 Characterization . . . 113
10.4.2 Properties . . . 113
10.4.3 Estimation . . . 115
10.4.4 Random generation . . . 115
10.4.5 Applications . . . 115
10.5 Generalized Pareto distribution . . . 115

III Multivariate and generalized distributions 116

11 Generalization of common distributions 117

11.1 Generalized hyperbolic distribution . . . 117
11.1.1 Characterization . . . 117
11.1.2 Parametrization . . . 118
11.1.3 Properties . . . 121
11.1.4 Special cases . . . 121
11.1.5 Estimation . . . 122
11.1.6 Random generation . . . 123
11.1.7 Applications . . . 124
11.2 Stable distribution . . . 125
11.2.1 Characterization . . . 125
11.2.2 Properties . . . 125
11.2.3 Special cases . . . 126
11.2.4 Estimation . . . 126
11.2.5 Random generation . . . 126
11.2.6 Applications . . . 127
11.3 Phase-type distribution . . . 127
11.3.1 Characterization . . . 127
11.3.2 Properties . . . 127
11.3.3 Special cases . . . 128
11.3.4 Estimation . . . 129
11.3.5 Random generation . . . 129
11.3.6 Applications . . . 130
11.4 Exponential family . . . 130
11.4.1 Characterization . . . 130

11.4.2 Properties . . . 130
11.4.3 Special cases . . . 130
11.4.4 Estimation . . . 131
11.4.5 Random generation . . . 131
11.4.6 Applications . . . 131
11.5 Elliptical distribution . . . 131
11.5.1 Characterization . . . 131
11.5.2 Properties . . . 132
11.5.3 Special cases . . . 132
11.5.4 Estimation . . . 132
11.5.5 Random generation . . . 132
11.5.6 Applications . . . 132

12 Multivariate distributions 133

12.1 Multinomial . . . 133
12.2 Multivariate normal . . . 133
12.3 Multivariate elliptical . . . 133
12.4 Multivariate uniform . . . 133
12.5 Multivariate student . . . 133
12.6 Kent distribution . . . 133
12.7 Dirichlet distribution . . . 133
12.7.1 Characterization . . . 133
12.7.2 Properties . . . 133
12.7.3 Estimation . . . 134
12.7.4 Random generation . . . 134

12.7.5 Applications . . . 134
12.8 Von Mises Fisher . . . 134
12.9 Evens . . . 134

13 Misc **135**

13.1 MBBEFD distribution . . . 135
13.1.1 Characterization . . . 135
13.1.2 Special cases . . . 136
13.1.3 Properties . . . 136
13.1.4 Estimation . . . 136
13.1.5 Random generation . . . 136
13.1.6 Applications . . . 136
13.2 Cantor distribution . . . 136
13.3 Tweedie distribution . . . 136

Conclusion **137**

Bibliography **137**

A Mathematical tools **141**

A.1 Basics of probability theory . . . 141
A.1.1 Characterising functions . . . 141
A.2 Common mathematical functions . . . 142
A.2.1 Integral functions . . . 142
A.2.2 Factorial functions . . . 142
A.2.3 Serie functions . . . 143
A.2.4 Miscellanous . . . 143

A.3 Matrix exponential 143

A.4 Kronecker product and sum 144