Universitat Autònoma de Barcelona

FACULTAT DE CIÈNCIES I BIOCIÈNCIES

DEPARTAMENT DE MATEMÀTIQUES

Treball de Final de Grau d'Estadística Aplicada:

INTRODUCING HARMONIC DISTRIBUTION IN WIKIPEDIA

Alumne: Víctor Martínez Ruiz
Tutor: Dr. Pere Puig

Metadata

Títol treball: **Introducing Harmonic Distribution in Wikipedia**
Alumne: **Víctor Martínez Ruiz**
Tutor: **Pere Puig**

Abstract

Català
Volem estudiar la Distribució Harmònica, aleshores el que farem amb aquest treball és implementar aquesta distribució a Wikipedia on descriurem la seva funció de densitat, la funció de distribució, els moments i els estimadors dels paràmetres entre d'altres propietats.

Castellano
Queremos estudiar la Distribución Harmónica, entonces lo que haremos en este trabajo será implementar ésta distribución en Wikipedia donde describiremos su función de densidad , la función de distribución, los momentos y los estimadores de los parámetros entre otras propiedades.

English
We want to study the Harmonic Distribution. To do this we will implement this distribution in Wikipedia where we will describe its probability density function, its cumulative distribution function, the moments and some parameter estimators among other properties.

Paraules claus:

Català
Funció de Densitat de Probabilitat, Funció de Disribució, mediana, estimació de paràmetres.

Castellano
Función de Densidad de Probabilidades, Función de Distribución, mediana, estimación de parámetros.

English
Probability Density Function, Cumulative Distribution Function, median, estimation of parameters.

Contents

1 Introduction

In probability theory and statistics, the Harmonic Law is a continuous probability distribution. It was discovered by Étienne Halphen, who had become interested in the statistical modeling of natural events. His practical experience in data analysis motivated him to pioneer a new system of distributions that provided sufficient flexibility to fit a large variety of data sets. Halphen restricted his search to distributions whose parameters could be estimated using simple statistical approaches. Then, Halphen introduced for the first time what he called the Harmonic distribution pr Harmonic Law.
It is important to remark that the Harmonic Law is a special case of the Generalized Inverse Gaussian (GIG) family of distributions when $\gamma = 0$.

2 History

One of Halphen's tasks, while working as statistician for Electricité de France, was the modeling of the monthly flow of water in hydroelectric stations. Halphen realized that the Pearson system of probability distributions could not be solved, it was inadequate for his purposes despite its remarkable properties. Therefore, Halphen's objective was to obtain a probability distribution with two parameters, subject to an exponential decay both for large and small flows[1].

In 1941, Halphen decided that, in suitably scaled units, the density of the monthly flow X should be the same as 1/X [1]. Taken this consideration, Halphen found the Harmonic density function. Nowadays known as an hyperbolic distribution, it has been studied by Rukhin (1974) and Barndorff-Nielsen (1978) [5].

The Harmonic Law is the only one two-parameter family of distributions that is closed under change of scale and under reciprocals, such that the maximum likelihood estimator of the population mean is the sample mean (Gauss' principle) [4].

In 1946, Halphen realized that introducing an additional parameter, flexibility could be improved. His efforts led him to generalize the Harmonic Law to obtain the GIG density[1].

3 Definition

3.1 Notation

The Harmonic distribution will be denoted by $\theta(m, a)$. As a result, when a random variable X is distributed following a Harmonic Law, the parameter of scale m is just the population median and a can be interpreted as a parameter of shape. We write:

$$X \sim \theta(m, a)$$

3.2 Probability Density Function

The density function of the Harmonic Law, which depends of two parameters, has the form[4],

$$f(x; m, a) = \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})],\ x \geq 0$$

Where:

- $K_0(a)$ denotes the modified Bessel function of the third kind with index 0.
- $m \geq 0$.
- $a \geq 0$.

We can see in Figure 1 that depending on the values of a and m, the probability density function has different profiles.

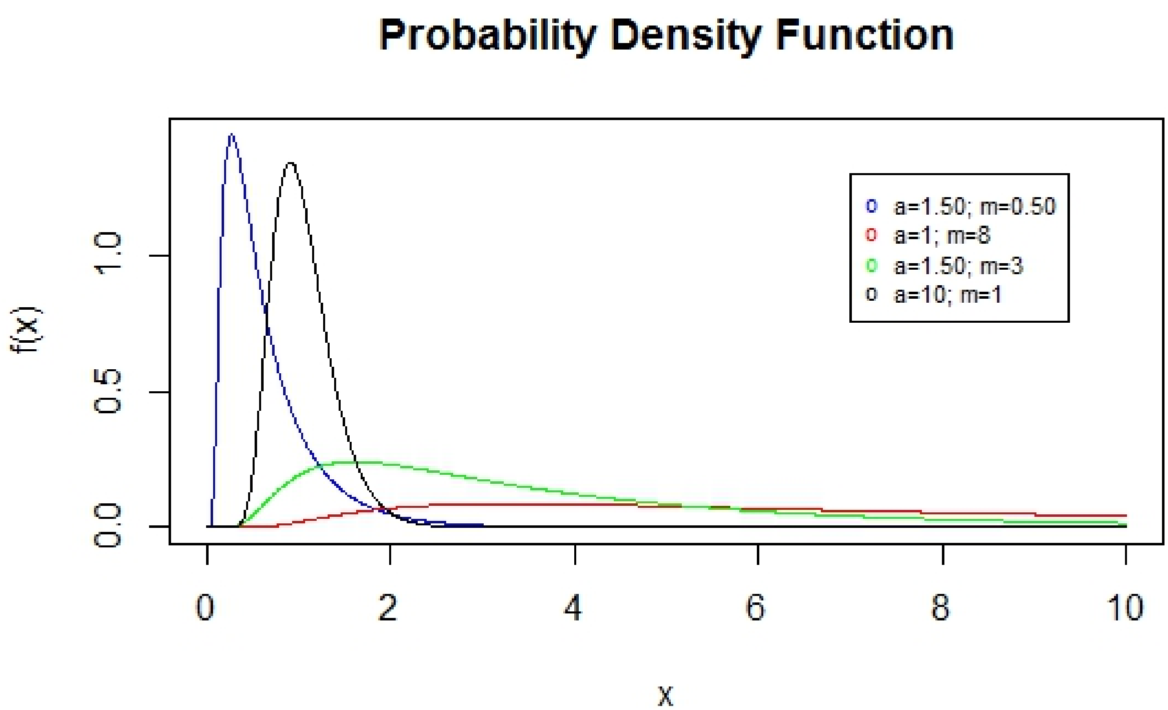

Figure 1: Probability Density Function

The highest value of the probability density function corresponds to its modal value or mode (see 4.1). Once it reaches the maximum, the density decreases tending to 0.

Now, we are going to describe the changes found in the shape of the densities, when we fix one parameter and the other one increases or decreases.

3.2.1 a constant

When a is a constant, for example $a = 1.5$, and m takes different values:

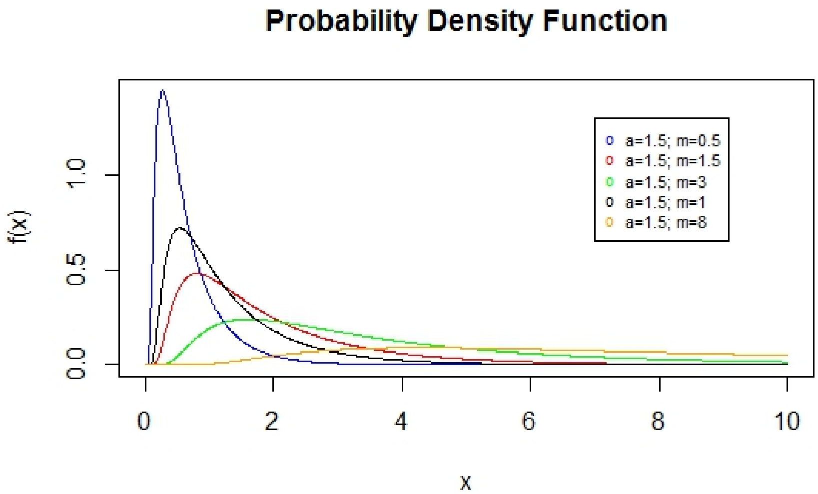

Figure 2: Probability Density Function $a = 1.5$

First when $m = 0.5$, the density function shown in Figure 2 increases quickly but it also tends to zero very fast after its modal value.
On the other hand, when $m = 8$, the density function increases slowly and its maximum value is lower than for $m = 0.5$.
Note that, fixing a, as m increases the density function is more flat.

3.2.2 m constant

When m is a constant, for example $m = 1.5$, and a takes different values:

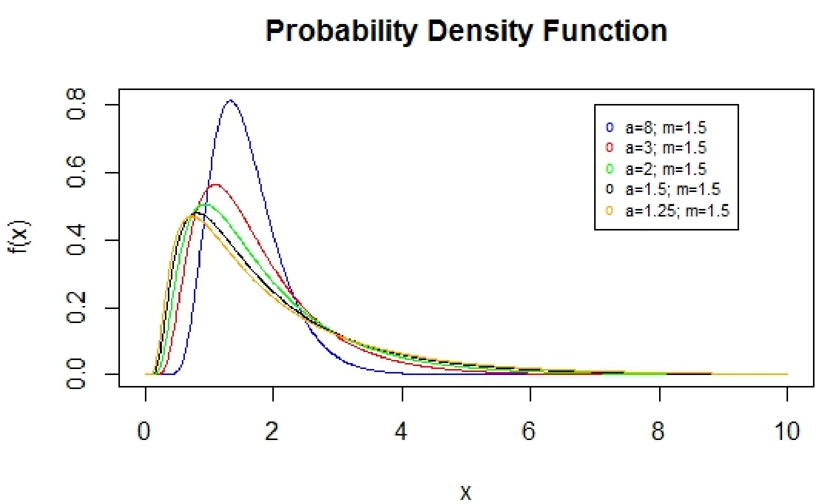

Figure 3: Probability Density Function $m = 1.5$

Taking m as a fixed value, we can see in Figure 3 that the mode increases as a increases. However, after its modal point the density goes faster to zero for low values than for large values of the parameter a.

3.3 Cumulative Distribution Function

The Cumulative distribution function (cdf) of the Harmonic Law does not have a closed form, and consequently it is not possible to derive an explicit expression.
The Cumulative distribution function has to be calculated solving numerically the following integral[3],

$$F(x;m,a) = \int_0^x \frac{1}{2tK_0(a)} \exp[-\frac{a}{2}(\frac{t}{m} + \frac{m}{t})]dt$$

Although we only can get the cdf numerically, we can also explore the different profiles changing the values of the parameters a and m.

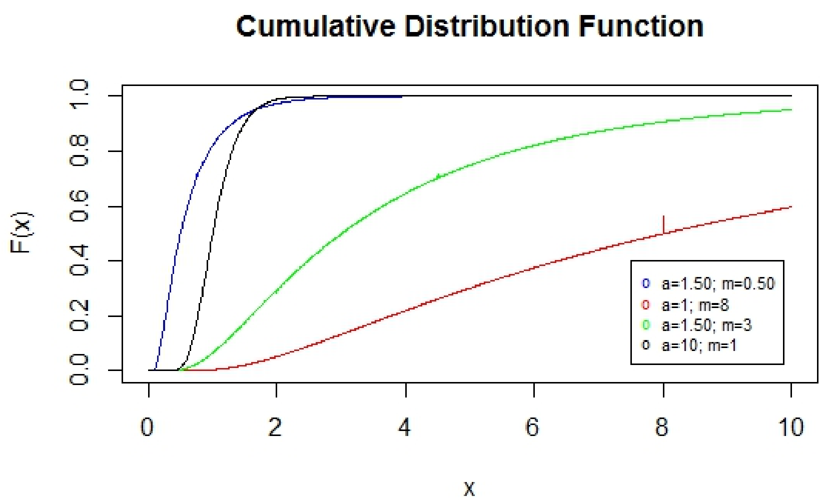

Figure 4: Cumulative Distribution Function

We are going to describe the changes found in the shape of the cumulative distribution function when we fix one parameter and the other one changes.

3.3.1 a constant

When a is a constant, for example $a = 1.5$, Figure 5 shows the different profiles of the cdf changing m.

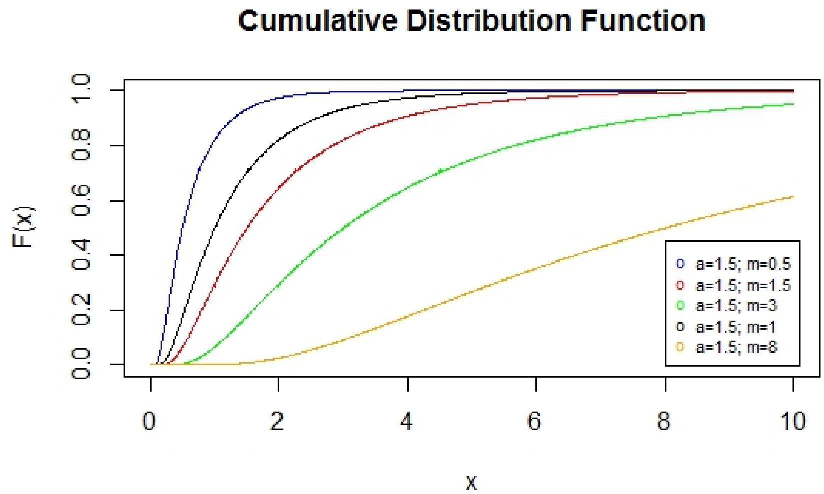

Figure 5: Cumulative Distribution Function $a = 1.5$

With a fixed, the cdf tends to 1 more slowly as m increases. Note that for $m = 0.5$, cdf increases quickly and arrives close to 1 faster than for the other values of m.

3.3.2 m constant

When m is a constant, for example $m = 1.5$, Figure 6 shows the different shapes of the cdf for several values of a.

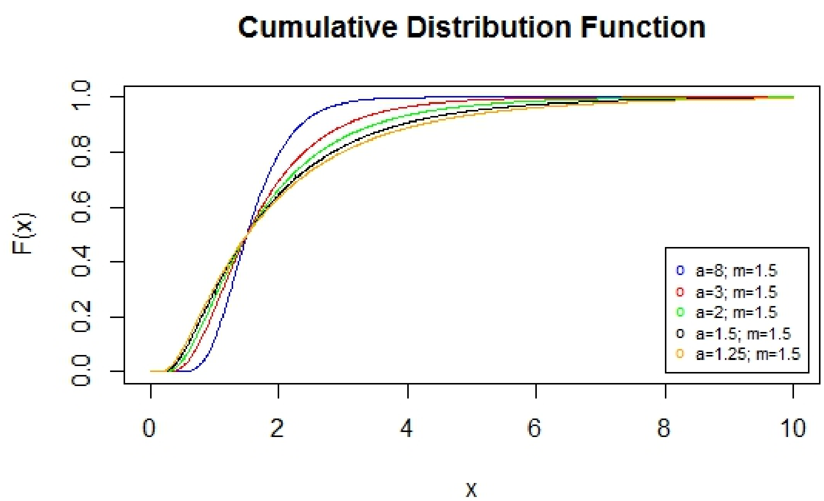

Figure 6: Cumulative Distribution Function $m = 1.5$

With the value of m fixed at 1.5, the first thing we see is that all the curves cut at $x = m = 1.5$. It happens

because m is the median and $F(1.5) = 0.5$.
Note that, for $x > m$, the cdf grows to 1 more fast as a increases.

3.4 Quantiles

The Quantiles of the Harmonic Law are calculated using the Cumulative distribution. In general does not exist a closed form except for the second quantile[3].
In general we can only calculate the quantiles numerically.

- q1
  The first quantile, can be obtained solving the following equation expressed in terms of the integral of the probability density function:

$$\int_0^{q1} \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]dx = \frac{1}{4}$$

- q2
  In this distribution, m is the median, that is, $q2 = m$. Therefore,

$$\int_0^{m} \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]dx = \frac{1}{2}$$

- q3
  Finally, the third quantile, comes from the solution of the equation,

$$\int_0^{q3} \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]dx = \frac{3}{4}$$

4 Properties

4.1 Mode

The modal value or mode can be obtained finding the maximum of the probability density function. It can be done equating to zero the derivative of the pdf with respect to x, and solving the corresponding equation.

We derive first:

$$\frac{\partial}{\partial x} f(x; m, a) = \frac{\partial}{\partial x} \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})] = \frac{1}{2K_0(a)} \frac{\partial}{\partial x} (\frac{\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]}{x}) =$$

$$= \frac{1}{2K_0(a)} \frac{1}{x^2} \frac{\partial}{\partial x} (\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]) =$$

$$= \frac{1}{2K_0(a)} \frac{1}{x^2} (-\frac{a}{2} x (\frac{1}{m} - \frac{m}{x^2}) \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})] - \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]) =$$

$$= -\frac{1}{4xK_0(a)}a(\frac{1}{m} - \frac{m}{x^2})\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})] - \frac{1}{2x^2K_0(a)}\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})])$$

Next step, to equal to 0:

$$-\frac{1}{4xK_0(a)}a(\frac{1}{m} - \frac{m}{x^2})\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})] - \frac{1}{2x^2K_0(a)}\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]) = 0$$

Finally, to solve the equation with respect x:

$$-\frac{1}{2xK_0(a)}\exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]\frac{1}{2}a(\frac{1}{m} - \frac{m}{x^2}) + \frac{1}{x}) = 0$$

$$-\frac{1}{2}a(\frac{1}{m} - \frac{m}{x^2}) - \frac{1}{x}) = 0$$

$$0 = \frac{a}{2m} - \frac{am}{2x^2} + \frac{1}{x}$$

$$0 = \frac{ax^2 - am^2 + 2mx}{2mx^2}$$

$$0 = ax^2 - am^2 + 2mx$$

$$X_{Mode} = \frac{-2m \pm \sqrt{4m^2 + 4a^2m^2}}{2a} = \frac{-2m \pm 2m\sqrt{1 + a^2}}{2a} = \frac{-m \pm m\sqrt{1 + a^2}}{a}$$

$$X_{Mode} = \frac{m(\sqrt{a^2 + 1} - 1)}{a}$$

Note that the negative solution of the quadratic equation has not been considered because it has not sense.

4.2 Moments

To derive an expression for the non-central moment of order r, the integral representation of the Bessel functions are used [2]. It is direct to show that,

$$\mu'_r = \int_0^{+\infty} x^r f(x; m, a)dx = m^r \frac{K_r(a)}{K_0(a)}$$

Where:

- r denotes the order of the moment.

The four principal central moments of the Harmonic Law are:

Order	Moment	Cumulant
1	$\mu_1 = m\frac{K_1(a)}{K_0(a)}$	μ
2	$\mu_2 = m^2(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)})$	σ^2
3	$\mu_3 = m^3(\frac{K_3(a)}{K_0(a)} - 3\frac{K_1(a)K_2(a)}{K_0^2(a)} + 2\frac{K_1^3(a)}{K_0^3(a)})$	k_3
4	$\mu_4 = m^4(\frac{K_4(a)}{K_0(a)} - 4\frac{K_1(a)K_3(a)}{K_0^2(a)} + 6\frac{K_1^2(a)K_2(a)}{K_0^3(a)} - 3\frac{K_1^4(a)}{K_0^4(a)})$	k_4

Table 1: Moments

4.3 Skewness

The coefficient of skewness is the third standardized moment around the mean divided by the 3/2 power of the standard deviation[2]. Then, it follows,

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{m^3(\frac{K_3(a)}{K_0(a)} - 3\frac{K_1(a)K_2(a)}{K_0^2(a)} + 2\frac{K_1^3(a)}{K_0^3(a)})}{(m^2(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)}))^{\frac{3}{2}}} = \frac{\frac{K_3(a)}{K_0(a)} - 3\frac{K_1(a)K_2(a)}{K_0^2(a)} + 2\frac{K_1^3(a)}{K_0^3(a)}}{(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)})^{\frac{3}{2}}} =$$

$$= \frac{\frac{K_0^3(a)K_3(a) - 3K_0(a)K_1(a)K_2(a) + 2K_1^3(a)}{K_0^3(a)}}{(\frac{K_0(a)K_2(a) - K_1^2(a)}{K_0^2(a)})^{\frac{3}{2}}} = \frac{\frac{K_0^3(a)K_3(a) - 3K_0(a)K_1(a)K_2(a) + 2K_1^3(a)}{K_0^3(a)}}{\frac{(K_0(a)K_2(a) - K_1^2(a))^{\frac{3}{2}}}{K_0^3(a)}} =$$

$$= \frac{K_0^2(a)K_3(a) - 3K_0(a)K_1(a)K_2(a) + 2K_1^3(a)}{(K_0(a)K_2(a) - K_1^2(a))^{\frac{3}{2}}}$$

Because $\gamma_1 > 0$, the mass of the distribution is always concentrated on the left.

4.4 Kurtosis

The coefficient of kurtosis is the fourth standardized moment divided by the square of the variance. For the Harmonic distribution it is [2],

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{m^4(\frac{K_4(a)}{K_0(a)} - 4\frac{K_1(a)K_3(a)}{K_0^2(a)} + 6\frac{K_1^2(a)K_2(a)}{K_0^3(a)} - 3\frac{K_1^4(a)}{K_0^4(a)})}{(m^2(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)}))^2} = \frac{\frac{K_4(a)}{K_0(a)} - 4\frac{K_1(a)K_3(a)}{K_0^2(a)} + 6\frac{K_1^2(a)K_2(a)}{K_0^3(a)} - 3\frac{K_1^4(a)}{K_0^4(a)}}{(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)})^2} =$$

$$= \frac{\frac{K_0^3(a)K_4(a) - 4K_0^2(a)K_1(a)K_3(a) + 6K_0(a)K_1^2(a)K_2(a) - 3K_1^4(a)}{K_0^4(a)}}{(\frac{K_0(a)K_2(a) - K_1^2(a)}{K_0^2(a)})^2} = \frac{\frac{K_0^3(a)K_4(a) - 4K_0^2(a)K_1(a)K_3(a) + 6K_0(a)K_1^2(a)K_2(a) - 3K_1^4(a)}{K_0^4(a)}}{\frac{(K_0(a)K_2(a) - K_1^2(a))^2}{K_0^4(a)}} =$$

$$= \frac{K_0^3(a)K_4(a) - 4K_0^2(a)K_1(a)K_3(a) + 6K_0(a)K_1^2(a)K_2(a) - 3K_1^4(a)}{(K_0(a)K_2(a) - K_1^2(a))^2}$$

Because $\gamma_2 > 0$, the density has a high acute peak around the mean and fatter tails than the Normal distribution.

5 Parameter estimation

5.1 Maximum Likelihood Estimator

Given a sample $X = (x_1, ...x_n)$, the likelihood function of the Harmonic Distribution with parameters a and m is,

$$L(X;\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{2xK_0(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]$$

Then, the log-likelihood function is,

$$l = \log(L(X;\theta)) = \sum_{i=1}^{n} \log(f(x_i;\theta)) = \sum_{i=1}^{n} \log(\frac{1}{2xK_0(a)}) - \sum_{i=1}^{n} \frac{ax_i}{2m} - \sum_{i=1}^{n} \frac{am}{2x_i} =$$

$$= \sum_{i=1}^{n} \log(1) - [\sum_{i=1}^{n} \log(2K_0(a)) + \sum_{i=1}^{n} \log(x_i)] - \frac{a}{2m} \sum_{i=1}^{n} x_i - \frac{am}{2} \sum_{i=1}^{n} \frac{1}{x_i} =$$

$$= -n\log(2K_0(a)) - \sum_{i=1}^{n} \log(x_i) - \frac{a}{2m} \sum_{i=1}^{n} x_i - \frac{am}{2} \sum_{i=1}^{n} \frac{1}{x_i} =$$

$$= n\log(2K_0(a)) + \sum_{i=1}^{n} \log(x_i) + \frac{a}{2m} \sum_{i=1}^{n} x_i + \frac{am}{2} \sum_{i=1}^{n} \frac{1}{x_i}$$

From the log-likelihood function, the likelihood equations are,

$$\frac{\partial l}{\partial a} = n(\log(2K_0(a)))' + \frac{1}{2m} \sum_{i=1}^{n} x_i + \frac{m}{2} \sum_{i=1}^{n} \frac{1}{x_i}$$

$$\frac{\partial l}{\partial m} = \frac{a}{2m^2} \sum_{i=1}^{n} x_i + \frac{a}{2} \sum_{i=1}^{n} \frac{1}{x_i}$$

Then, the maximum likelihood estimators of the parameters $\hat{m}$ and $\hat{a}$ are the solutions of the likelihood equations, that is,

$$\bar{H} = \hat{m}\frac{K_1(\hat{a})}{K_0(\hat{a})}$$

,

$$\bar{H}_{-1} = \hat{m}^{-1}\frac{K_1(\hat{a})}{K_0(\hat{a})},$$

where $\bar{H}$ is the sample mean and $\bar{H}_{-1} = (\sum_{i=1}^{n} 1/x_i)/n$ is the sample mean of the reciprocals.

It follows directly that $\hat{m} = \sqrt{\frac{\bar{H}}{\bar{H}_{-1}}}$, and $\hat{a}$ can be found solving numerically the equation,

$$\sqrt{\bar{H}\bar{H}_{-1}} = \frac{K_1(\hat{a})}{K_0(\hat{a})}.$$

Based on the maximum likelihood estimators we can also estimate the variance efficiently. Taking into account the identity $K_2(a) = K_0(a) + \frac{2K_1(a)}{a}$ we can obtain the following estimation of variance[4],

$$\begin{cases} \bar{H} = m\frac{K_1(a)}{K_0(a)} \\ \sigma^2 = m^2(\frac{K_2(a)}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)}) = m^2(\frac{K_0(a)+\frac{2K_1(a)}{a}}{K_0(a)} - \frac{K_1^2(a)}{K_0^2(a)}) = m^2(1 + \frac{2K_1(a)}{K_0(a)a} - \frac{K_1^2(a)}{K_0^2(a)}) \end{cases}$$

$$\sigma^2 = m^2 + \frac{2}{a}m^2\frac{K_1(a)}{K_0(a)} - m^2\frac{K_1^2(a)}{K_0^2(a)}$$

$$\hat{\sigma}^2 = \hat{m}^2 + \frac{2\hat{m}\bar{H}}{\hat{a}} - \bar{H}^2$$

5.2 Method of Moments

The population mean and variance of the Harmonic distribution are,

$$\begin{cases} \mu = m\frac{K_1(a)}{K_0(a)} \\ \sigma^2 = m^2(1 + \frac{2K_1(a)}{K_0(a)a} - \frac{K_1^2(a)}{K_0^2(a)}) \end{cases}$$

Note that,

$$\sigma^2 = m^2 + \frac{2m\mu}{a} - \mu^2 = \mu^2\left(\frac{K_0(a)}{K_1(a)}\right)^2 + \frac{2K_0(a)\mu^2}{K_1(a)a} - \mu^2$$

The method of moments consists in to solve the following equations:

$$\begin{cases} \bar{H} = m\frac{K_1(a)}{K_0(a)} \\ s^2 = \bar{H}^2\left(\frac{K_0(a)}{K_1(a)}\right)^2 + \frac{2K_0(a)\bar{H}^2}{K_1(a)a} - \bar{H}^2 \end{cases}$$

where s^2 is the sample variance and $\bar{H}$ is the sample mean. Solving the second equation we obtain $\hat{a}$, and then we calculate $\hat{m}$ using,

$$\hat{m} = \frac{\bar{H}K_0(\hat{a})}{K_1(\hat{a})}$$

5.3 Example of application

To illustrate the calculation of the different estimators (MLE and Method of Moments) we will work with a real data example.

The feed conversion index is a frequently used ratio in animal production. For a particular animal in the considered period of time, the index is calculated as the weight of the feed eaten divided by the increment of the weight of the animal. The following data set corresponds to the feed conversion indexes of 24 calves coming from a lot:

3.65 4.03 4.58 4.61 4.70 4.85 3.21 3.93 3.15 3.00 2.93 3.56
4.13 3.68 3.88 3.25 3.92 3.99 3.04 3.10 3.20 3.35 3.19 3.10

First, we are going to obtain the MLE estimators $\hat{a}$ and $\hat{m}$. To do this, we calculate $\bar{H}$ and $\bar{H}_{-1}$:

$$\bar{H} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{24}\sum_{i=1}^{n}(3.65 + 4.03 + \cdots + 3.10) = 3.668$$

$$\bar{H}_{-1} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i} = \frac{1}{24}\sum_{i=1}^{n}(\frac{1}{3.65} + \frac{1}{4.03} + \cdots + \frac{1}{3.10}) = 0.279$$

And with these two values, we obtain $\hat{m}$:

$$\hat{m} = \sqrt{\frac{\bar{H}}{\bar{H}_{-1}}} = \sqrt{\frac{3.668}{0.279}} = 3.626$$

In order to calculate $\hat{a}$ we have to solve numerically the following equation (see the Appendix):

$$\sqrt{\bar{H}\bar{H}_{-1}} = \frac{K_1(\hat{a})}{K_0(\hat{a})}$$

$$1.011619 = \frac{K_1(\hat{a})}{K_0(\hat{a})}$$

$$\hat{a} = 42.789$$

Now we are going to calculate the Moment estimators.

First, we have to calculate s^2 from the data set, and then proceed to solve the corresponding equations.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{H})^2 = \frac{1}{24-1}\sum_{i=1}^{24}(x_i - 3.668)^2 = 0.35$$

Now, solving the equation,

we obtain $\hat{a} = 39.39$

And then $\hat{m}$ is 3.62,

If we compare the MLE and the Method of Moments estimators,

$$\begin{cases} \hat{m}_{MLE} = 3.626 \\ \hat{m}_{MM} = 3.62 \end{cases}$$

we can see that the estimators of m are very similar.

$$\begin{cases} \hat{a}_{MLE} = 42.789 \\ \hat{a}_{MM} = 39.39 \end{cases}$$

On the other hand, the estimators of a are more different.

6 Related Distributions

Harmonic Law is a sub-family of the Generalized Inverse Gaussian distribution. The density of the GIG family has the form,

$$f(x; a, m, \gamma) = \frac{x^{\gamma-1}}{2m^{\gamma}K_{\gamma}(a)} \exp[-\frac{a}{2}(\frac{x}{m} + \frac{m}{x})]$$

The density of the GIG family corresponds to the Harmonic Law when $\gamma = 0$[4].

When a tends to infinity, the Harmonic Law can be approximated by a Normal distribution. It can be shown checking that if a tends to infinity, then $U = \sqrt{a}(\frac{X}{m} - 1)$, which is a lineal transformation of X, tends to a standard Normal distribution (N(0,1)).

This explains why the Normal distribution can be used successfully for certain data sets of ratios [2].

Another related distribution is the log-Harmonic Law, which is the probability distribution of a random variable whose logarithm follows an Harmonic Law.
This family has an interesting property, the Pitman estimator of the location parameter does not depend on the choice of the loss function. Only two statistical models satisfy this property: One is the normal family of distributions and the other one is a three-parameter statistical model which contains the log-Harmonic Law[5].

7 Conclusions

The Harmonic distribution is a especial case of the Generalized Inverse Gaussian distribution when $\gamma = 0$. Both were discovered by Halphen who was modelling data related with Hydrology.

Harmonic distribution has two parameters a and m.

The parameter of scale m is the median of the distribution (second quartile q_2) and a is the parameter of shape.

The distribution is unimodal, having a modal value that can be written in terms of a and m.

The maximum likelihood and moment estimators are easy to obtain, solving a numerical equation to find the estimator of the parameter a.

The Wikipedia Link to see the Harmonic Distribution is:
`https://en.wikipedia.org/wiki/Harmonic_Distribution`

8 References

[1] Kotz, Samuel L., *Encyclopedia of statistical sciences*, Second Edition, Volume 5, pgs. 3059–3061 & 3069–3072, 1982–1989.

[2] Perrault,L. , Bobée, B. & Rasmussen, P.F., 1999a., *Halphen distribution system. I: Mathematical and statistical properties.*, J. Hydrol. Eng. 4 (3), 189–199.

[3] Perrault,L. , Bobée, B. & Rasmussen, P.F., 1999b., *Halphen distribution system. II: Parameter and quantile estimation.*, J. Hydrol. Eng. 4 (3), 200–209.

[4] Puig, P., *A note on the harmonic law: A two-parameter family of distributions for ratios*, Statistics & Probability Letters 78, pgs. 320–326, 2008.

[5] Rukhin, A.L., *Strongly symmetrical families and statistical analysis of their parameters*, J. Sov.,Math. 9, pgs. 886–910, 1978.

A Appendix

Graphic 3.2

```
> require(GeneralizedHyperbolic)
> x<- seq(0,10,.01)
> #pdf
> a=1.5;m=0.5
> plot(x,dgig(x,a*m,a/m,0, param=c(a*m,a/m,0)), col="blue",type="l",xlab="x",ylab="f(x)",
+      main="Probability Density Function")
> a=1;m=8
> lines(x,dgig(x,a*m,a/m,0), col="red", type="l")
> a=1.5;m=3
> lines(x,dgig(x,a*m,a/m,0), col="green", type="l")
> a=10;m=1
> lines(x,dgig(x,a*m,a/m,0), col="black", type="l")
> legend(7,1.3, c("a=1.50; m=0.50","a=1; m=8","a=1.50; m=3","a=10; m=1"),
+        pch=c("o"),col=c("blue","red","green","black"),cex=0.7)
```

Graphic 3.2.1

```
> #pdf a=1.5
> a=1.5;m=0.5
> plot(x,dgig(x,a*m,a/m,0, param=c(a*m,a/m,0)), col="blue",type="l",xlab="x",ylab="f(x)",
+      main="Probability Density Function")
> a=1.5;m=1.5
> lines(x,dgig(x,a*m,a/m,0), col="red", type="l")
> a=1.5;m=3
> lines(x,dgig(x,a*m,a/m,0), col="green", type="l")
> a=1.5;m=1
> lines(x,dgig(x,a*m,a/m,0), col="black", type="l")
> a=1.5;m=8
> lines(x,dgig(x,a*m,a/m,0), col="orange", type="l")
> legend(7,1.3, c("a=1.5; m=0.5","a=1.5; m=1.5","a=1.5; m=3","a=1.5; m=1","a=1.5; m=8"),
+        pch=c("o"),col=c("blue","red","green","black","orange"),cex=0.7)
```

Graphic 3.2.2

```
> #pdf m=1.5
> a=8;m=1.5
> plot(x,dgig(x,a*m,a/m,0, param=c(a*m,a/m,0)), col="blue",type="l",xlab="x",ylab="f(x)",
+      main="Probability Density Function")
> a=3;m=1.5
> lines(x,dgig(x,a*m,a/m,0), col="red", type="l")
> a=2;m=1.5
> lines(x,dgig(x,a*m,a/m,0), col="green", type="l")
> a=1.5;m=1.5
> lines(x,dgig(x,a*m,a/m,0), col="black", type="l")
> a=1.25;m=1.5
> lines(x,dgig(x,a*m,a/m,0), col="orange", type="l")
> legend(7,0.8, c("a=8; m=1.5","a=3; m=1.5","a=2; m=1.5","a=1.5; m=1.5","a=1.25; m=1.5"),
+        pch=c("o"),col=c("blue","red","green","black","orange"),cex=0.7)
```

Graphic 3.3

```
> #cdf
> a=1.5;m=0.5
> plot(x,pgig(x,a*m,a/m,0), col="blue",type="l",xlab="x",ylab="F(x)",
+      main="Cumulative Distribution Function")
> a=1;m=8
> lines(x,pgig(x,a*m,a/m,0), col="red", type="l")
> a=1.5;m=3
> lines(x,pgig(x,a*m,a/m,0), col="green", type="l")
> a=10;m=1
> lines(x,pgig(x,a*m,a/m,0), col="black", type="l")
> legend(7.5,0.4, c("a=1.50; m=0.50","a=1; m=8","a=1.50; m=3","a=10; m=1"),
+        pch=c("o"),col=c("blue","red","green","black"), cex=0.7)
```

Graphic 3.3.1

```
> #cdf a=1.5
> a=1.5;m=0.5
> plot(x,pgig(x,a*m,a/m,0, param=c(a*m,a/m,0)), col="blue",type="l",xlab="x",ylab="F(x)",
+      main="Cumulative Distribution Function")
> a=1.5;m=1.5
> lines(x,pgig(x,a*m,a/m,0), col="red", type="l")
> a=1.5;m=3
> lines(x,pgig(x,a*m,a/m,0), col="green", type="l")
> a=1.5;m=1
> lines(x,pgig(x,a*m,a/m,0), col="black", type="l")
> a=1.5;m=8
> lines(x,pgig(x,a*m,a/m,0), col="orange", type="l")
> legend(8,0.45, c("a=1.5; m=0.5","a=1.5; m=1.5","a=1.5; m=3","a=1.5; m=1","a=1.5; m=8"),
+        pch=c("o"),col=c("blue","red","green","black","orange"),cex=0.7)
```

Graphic 3.3.2

```
> #cdf m=1.5
> a=8;m=1.5
> plot(x,pgig(x,a*m,a/m,0, param=c(a*m,a/m,0)), col="blue",type="l",xlab="x",ylab="F(x)",
+      main="Cumulative Distribution Function")
> a=3;m=1.5
> lines(x,pgig(x,a*m,a/m,0), col="red", type="l")
> a=2;m=1.5
> lines(x,pgig(x,a*m,a/m,0), col="green", type="l")
> a=1.5;m=1.5
> lines(x,pgig(x,a*m,a/m,0), col="black", type="l")
> a=1.25;m=1.5
> lines(x,pgig(x,a*m,a/m,0), col="orange", type="l")
> legend(8,0.45, c("a=8; m=1.5","a=3; m=1.5","a=2; m=1.5","a=1.5; m=1.5","a=1.25; m=1.5"),
+        pch=c("o"),col=c("blue","red","green","black","orange"),cex=0.7)
```

Calculate $\hat{a}$ MLE, $\hat{a}$ and $\hat{m}$ Method of Moments by Maple.

MLE:

$$fsolve(\frac{BesselK(1,a)}{BesselK(0,a)} = 1.011619, a)$$

Method of Moments:
$\hat{a}$:

$$fsolve(a = 2*BesselK(0,a)*3.668^2*(1/(BesselK(0,a)*(0.35+3.668^2-3.668^2*(BesselK(0,a)/BesselK(1,a))^2))), a)$$

$\hat{m}$:

$$fsolve(m = 3.668 * BesselK(0, 39.38958214)/BesselK(1, 39.38958214), m)$$