Harmonic Probability Density Function, Modeling, and Regression

Reza Ahangar
Texas A & M University- Kingsville

Correspondence concerning this article should be addressed to:
Reza Ahangar,
700 University BLVD, MSC 172, Mathematics Department,
Texas A & M University- Kingsville
Kingsville, TX 78363, USA.
E-mail: reza.ahangar@tamuk.edu
361-593-2235.

Abstract
A continuous case of new density function called harmonic probability density function (HPDF) is structured using mathematical method. The mean and variance of this probability function is presented using both the definition and moment generating function. A discussion has started on many other connections that may exist between this new model and its applications in geometry, probability, and statistics. For a certain set of data we may be able to model by Harmonic Probability Density Function and a Harmonic PDF regression method is also presented in this paper.

Keywords: harmonic probability density function, harmonic distribution function, harmonic regression.

1- Introduction:

The Geometry of statistical notions or the connections between Geometry and Statistics has existed from the beginning of both disciplines (Adler C. F. 1958, and Fisher, J. B, 1978).
The connection between arithmetic, geometric, and harmonic mean can be studied by the same approach of Geometry of Statistics. You may observe that some other investigations are based on the statistical properties of geometric notions (Ahangar R. R. 2010, Hilbert, D. 1902, Pearson, R. 2011, Saville, J. D., and Wood R. G., (1977) .
Arithmetic mean as a great mode of investigation in statistics is used in many natural phenomena. But many researches prefer to use either geometric or harmonic mean in their investigations, (MacCluer, C. R. 2000) .
The arithmetic mean and the median may be the most popular and convinient measures that financial analysts use for their valuations. It is believed by many financial organizations and bankers that the Harmonic Mean provides better information for reasonable measure in investment strategy. (Mathews and Gilbert- 2006, Meyer D., 1970).
In slowly-decaying distributions, the harmonic mean often turns out to be a much better characterization than the arithmetic mean, which is a reciprocal transformation generally not even well-defined theoretically for these distributions (Pearson, R. 2011).
Much research has been done and computational tools designed these days that are equipped to change the mode of computation in either arithmetic, geometric, or harmonic sense. To the best of the author's

knowledge, no direct investigation exists to introduce harmonic probability density functions, its applications, modeling, and regression.

The traditional definitions of the arithmetic, geometric, and harmonic means for a set of two elements or for a set of n-elements data was the main motivation of the author to develop further to the probability density function.

It may be helpful to keep these definitions in mind. Let us denote the arithmetic mean of a set of data E by M(E), geometric mean by G(E), and harmonic mean by H(E). where E= $\{x_1, x_2, x_3, ..., x_n\}$

$$AM = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad GM = \bar{g} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}, \quad HM = \bar{h} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+...+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

(Kapur, D., and Mundy, J. L., 1989).

2- Creating a new Harmonic Probability Density Function:

Domain of a Harmonic Density Function: We should be able to find a real number L such that the function

$$f(x) = \begin{cases} \frac{1}{x} & if \quad 1 \le x \le L \\ 0 & Otherwise \end{cases}$$

such that $\int_1^L f(x)dx = 1$. This simple calculation concludes that L=e is a necessary condition for the area under the curve to be equal to one unit area. Thus, the function f(x) with these conditions will be a probability density function.

Theorem (1): i) For e, as a base of natural logarithm, the following function from the domain [1,e] into a range [0,1],

(2.1)

$$f:[1,e] \to [0,1] \quad and \quad f(x) = \begin{cases} \frac{1}{x} & if \quad 1 \le x \le e \\ 0 & Otherwise \end{cases}$$

is a probability density function.

ii) where the mean and variance are : $E(X) = \mu = e-1, \quad and \quad Var(X) = \frac{1}{2}(3-e)(e-1)$.

iii) The **Probability Distribution Function F(x)= ln(x).**

(2.2)

$$F(x) = \begin{cases} 0 & for \quad x < 1 \\ \ln(x) & 1 \le x \le e \\ 1 & x > e \end{cases}$$

Proof: By a simple integration one can verify (i) $\int_1^e \frac{1}{x}dx = \ln(x)]_1^e = 1$

and for part (ii)

$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_1^e dx = e-1$ **and** $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{x^2}{2}]_1^e = \frac{1}{2}(e^2-1)$

$$Var(X) = \sigma^2 = E(X^2) - (E(X))^2 = \frac{1}{2}(e^2-1) - (e-1)^2 = \frac{1}{2}(3-e)(e-1)$$

As a result the standard deviation can be estimated by $\sigma = .49197$.

iii) The **Probability Distribution Function is F(x) = ln(x) for x in the interval [1,e].**

This can be verified by showing that: $f(x)=\frac{\partial F(x)}{\partial x}$,

$F(-\infty)=0 \quad and \quad F(\infty)=1.$

The Moment Generating Function: To find the moment generating function we will find the expectation:

$$\mu_k^{'}=E(X^k)=\int_{-\infty}^{\infty}x^k f(x)dx=\int_{-\infty}^{\infty}x^k\frac{dx}{x}=\int_1^e x^{k-1}dx=\frac{x^k}{k}]_1^e=\frac{e^k-1}{k}$$

We may apply this result to verify the mean and standard deviation for k=1 and k=2.

$$\mu_1^{'}=\mu=e-1, \mu_2^{'}=\frac{1}{2}(e^2-1)\ and\ Variance=\sigma^2=\mu_2^{'}-\mu^2=\frac{1}{2}(e-1)(3-e)$$

In the following, we are demonstrating the graph of the simple harmonic density and distribution functions.

3-Transformation of the Harmonic Probability Density Function:

3.1-Horizontal Shift: The simple harmonic function can be shifted from z=0 to an arbitrary point z=a. We will investigate conditions for the following function

(3.1)
$$f(z)=\begin{cases}\frac{1}{z-a} & if \quad a+1\le z\le a+e\\ 0 & otherwise\end{cases}$$

It can be verified that the function f(z) is a density function.

$$i)\quad f(z)\ge 0,\quad ii)\quad \int_{-\infty}^{\infty}f(z)dz=1$$

Probability Distribution Function: The probability Distribution function of the random variable X will be in the following form.

$$F(z)=\int_{-\infty}^{z}\frac{du}{u-a}=\ln|z-a|\cdot$$

(3.2)
$$F(z)=\begin{cases}0 \quad if \quad z\le a+1\\ \ln|z-a| \quad if \quad a+1<z<a+e\\ 1 \quad z\ge a+e\end{cases}$$

This is true because $F(-\infty)=0, \quad and \quad F(\infty)=1$.

<u>Expectations:</u> The mean and standard deviation of the random variable Z can be calculated.

$$E(Z)=\int_{-\infty}^{\infty}zf(z)dz=\int_{a+1}^{a+e}\frac{z}{z-a}dz=[z+a\ln|z-a|]_{a+1}^{a+e}=a+e-1$$

To calculate the variance, we will use

$$E(z^2)=\int_{a+1}^{a+e}\frac{z^2}{z-a}dz=[\frac{1}{2}(z-a)^2+2z+a(2-a)\ln|z-a|]_{a+1}^{a+e}$$

$$=a^2+2a((e-1)+\frac{e^2-1}{2}$$

Variance of Harmonic density function also can be calculated:

$$\sigma^2=E(z^2)-(E(z))^2=a^2+2a((e-1)+\frac{e^2-1}{2}-(a+e-1)^2=$$

$$=\frac{(3-e)(e-1)}{2}$$

Theorem (2): The Geometric translation of a simple harmonic probability function will translate the original mean to $\mu=a+e-1$ and the standard deviation will remain the same

Moment Generating Function for shifted harmonic density function:

(3.3)

$$f(z) = \begin{cases} \dfrac{1}{z-a} & if \quad a+1 \le z \le a+e \\ 0 & otherwise \end{cases}$$

By definition of moment generating function

$$\mu_k^{'} = E(Z^k) = \int_{-\infty}^{\infty} z^k f(z)dz = \int_{-\infty}^{\infty} z^k \frac{dz}{z-a} = \int_{1+a}^{e+a} \frac{z^k}{z-a}dz$$

Using a substitution z-a=u the integral can be simplified.

$$\mu_k^{'} = \int_{1+a}^{e+a} \frac{(u+a)^k}{u}du = \int_{1+a}^{e+a} [\frac{\binom{k}{0}(a)^k + \sum_{i=1}^{k}\binom{k}{i}u^i a^{k-i}}{u}du = \int_{1+a}^{e+a} \frac{(a)^k}{u}du + \int_{1+a}^{e+a} \sum_{i=1}^{k}\binom{k}{i}u^{i-1}a^{k-i}du$$

$$= [a^k \ln|u| + \sum_{i=1}^{k}\binom{k}{i}a^{k-i}\frac{u^i}{i}]_{1+a}^{e+a}$$

$$= [a^k + \sum_{i=1}^{k}\binom{k}{i}a^{k-i}\frac{(e)^i}{i}] - [\sum_{i=1}^{k}\binom{k}{i}a^{k-i}\frac{1}{i}]$$

$$\mu_k^{'} = a^k + \sum_{i=1}^{k}\{\binom{k}{i}a^{k-i}(\frac{(e)^i}{i} - \frac{1}{i})\}$$

The moment generating function for harmonic probability function:

$$\mu_k^{'} = a^k + \sum_{i=1}^{k}\binom{k}{i}a^{k-i}\frac{(e)^i - 1}{i}, \quad for \quad k = 1,2,3...$$

We can examine this result for k=1 and k=2 to re-evaluate the mean and standard deviation. The result of this investigation can be expressed in the following theorem:

Theorem 3: The mean of a harmonic probability distribution is $\mu = a + e - 1$ and the variance $\sigma^2 = \frac{1}{2}(3-e)(e-1)$ is independent from the translation point z=a.

For k=1, $\mu_1^{'} = a + \sum_{i=1}^{1}\binom{k}{i}a^{k-i}\frac{(e)^i - 1}{i} = a + e - 1$

As a result for any constant real number a: $\mu = a + e - 1$

For k=2 we will substitute i=1,2.

$$\mu_2^{'} = a^2 + \sum_{i=1}^{2}\binom{2}{i}a^{2-i}\frac{(e)^i - 1}{i} = a^2 + \binom{2}{1}a^{2-1}\frac{(e)^1 - 1}{1} + \binom{2}{2}a^{2-2}\frac{(e)^2 - 1}{2}$$

$$\mu_2^{'} = a^2 + 2a(e-1) + \frac{(e)^2 - 1}{2} = (a+e-1)^2 + \frac{(e-1)(3-e)}{2} = \mu^2 + Variance \quad (a = 0)$$

Variance= $\sigma^2 = \mu_2^{'} - \mu^2 = (a+e-1)^2 + \dfrac{(e-1)(3-e)}{2} - (a+e-1)^2 = \dfrac{1}{2}(3-e)(e-1)$

Standard Deviation: $\sigma = \sqrt{(3-e)(e-1)/2} = .49198115$

Note: Vertical shift of harmonic distribution will not provide an easy and applicable solution. For every real number h, the conditions for probability density function

$$f : R \to [0,1], \quad such \ that \quad f(x) = \begin{cases} \dfrac{1}{x} + b & if \quad 1 \le x \le e \\ 0 & Otherwise \end{cases}$$

changes the area under the curve and the x-axis and will pose challenges to meet the conditions of probability density function.

Harmonic Density Function with Stretch or Contraction:

The original harmonic distribution can be introduced by

(3.5) $$g(x) = \begin{cases} \frac{K}{x} & 1 \le x \le L \\ 0 & Otherwise \end{cases}$$

Set the integral of the function g(x) equal to one in the given domain in order to satisfy the conditions of the probability density function.

(3.6) $$\int_1^L g(x)dx = 1 \Leftrightarrow L = e^{1/K}$$

If K>1, then L=$e^{1/K}$ <e. The transformation is a contraction. If K<1 then (1/K)>1 or $e^{1/K}$ >e, then the transformation is the stretched harmonic density function.

Theorem 4: The following transformed harmonic probability density function

(3.7) $$g(x) = \begin{cases} \frac{K}{x} & 1 \le x \le e^{\frac{1}{K}} \\ 0 & Otherwise \end{cases}$$

is a stretch in y-direction of the basic harmonic transformation if K<1 and it is a contraction in the y-direction for K>1.

General Harmonic Distribution:

The general harmonic probability density function can be defined by considering the assumptions in (3.1) and (3.7). As a result, one can verify that

(3.8) $$f(x) = \begin{cases} \dfrac{k}{x-a} & if \quad a+1 \le x \le a+e^{1/k} \\ 0 & otherwise \end{cases}$$

is a probability density function.

4- Algorithm for Harmonic Modeling and Harmonic Regression:

Suppose a set of given data satisfies a Harmonic Distribution (3.8). By Harmonic Regression we mean finding the parameters k and a such that the relation (3.8) is the best fit distribution for the set of data.

To simulate the regression process we use the Computer Algebra MAPLE to generate a set of data and the best Harmonic Curve fit for this set of data.

Assuming y=f(z) and evaluating z in the domain given in (3.8) will provide:

$$y = \frac{k}{x-a} \Rightarrow \quad \frac{1}{y} = \frac{x-a}{k} = \frac{x}{k} - \frac{a}{k}$$

This relation shows that the reciprocal of y has a linear relationship with x such that

$$z = \frac{1}{y} = mx + b \quad where \quad m = \frac{1}{k} \quad and \quad b = \frac{-a}{k}$$

By this introduction, we can find a linear regression between 1/y and x to evaluate k and a.

The following is the simulation and computation for Harmonic Regression using Excel.

Procedure: To produce and simulate the harmonic regression of (3.8), we will follow the following steps

- Given a set of data {z_i: i=1,..,n}, assign random points x_i for every equal increment in the domain [a+1,a+e^(1/k)].
- Use a linear regression for the new set of data (x_i, z_i) such that z=mx+b.

- find k and a such that

(4.1) $$k = \frac{1}{m}, \quad a = -k*b$$

- Replace k and a in the density function (3.8) and sketch the graph.

The goal in Harmonic regression is to determine values of k and a that satisfy the conditions of (3.8).

```
> restart;
> with(stats[statplots]):
> xv:=[3,34,4,5,77,67,44,23,11,99,47]:
> yv:=[92,52,82,75,4,14,7,8,96,2,16]:
> scatterplot(xv, yv, color = red);
```

$> zv := map\left(yv \to \left(\frac{1}{yv}\right), yv\right):$

```
> with(fit):
> leastsquare[[x,y],y=m*x+b,{m,b}]([xv,zv]); evalf(%);
```

$m := \frac{30746622727}{8033744409600} : \ b := -\frac{11087361}{332660224} :$

$> y := evalf(m \cdot x + b);$

$y := 0.003827184580\, x - 0.03332938596$

$> k := \frac{1}{m} : \ evalf(\%) : a := -k \cdot b : evalf(\%);$

8.708591201

$> ll := evalf(a + 1) : ul := evalf\left(a + \exp\left(\frac{1}{k}\right)\right):$

$g := piecewise\left(x < ll, 0, ll \le x \le ul, \frac{8033744409600}{30746622727\, x - 267759768150}, x > ul, 0\right);$

$> f := x \to \frac{k}{x - a} : y := simplify(f(x));$

#Harmonic Probability Fit

$y := \frac{8033744409600}{30746622727\, x - 267759768150}$

$> plot(f(x), x = 9..10);$

Randomly generated y values are demonstrated in an arbitrary domain for x. We will determine a linear regression for $z = \frac{1}{y} = mx + b.$ where m=.03242 and b=--0.00195

The Excel regression for this particular set of data shows that k=30.84479 and a=-0.0601. Thus the harmonic function will be

$$y = \frac{k}{x-a} = \frac{30.84479}{x+.0601}$$

This is a harmonic distribution in a domain: $[a+1, a+e^{\frac{1}{k}}]$.

5- Future Development: We presented the **mathematical view** of the continuous Harmonic Probability Density Functions and calculated the mean and variance of this new model in this paper. The geometric view, computation, and simulation version of this research can be presented in a separate study.

Assume that two random variables X and Y are related to two points A and B harmonically. What will be the correlation between X and Y? This question opens another stage to study the covariance of (X,Y) and discuss harmonically correlated random variables. The author is planning to apply the HPDF to Michaelis Mentin formulation which has many applications in chemical reactions (see Tesi di, L. D. 2006-2007). Further applications of harmonic distribution in modeling and industrial mathematics like z-transform are promising (Saville, J. D., and Wood, R. G., 1997).

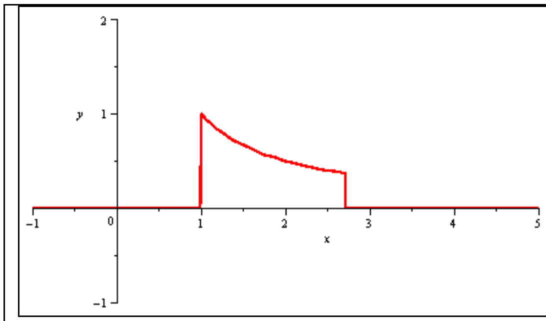

Fig. (1): The graph of the harmonic probability density function

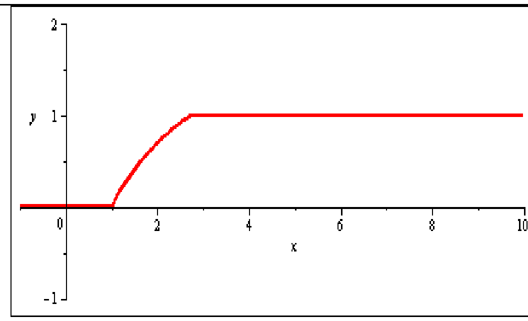

Fig. (2): The graph of the harmonic cumulative distribution function

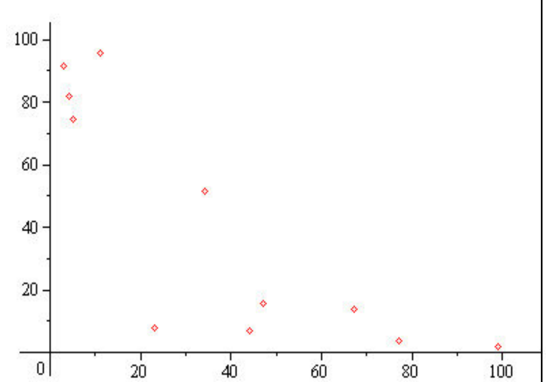

Fig. (3): The scatter plot of the given x-value and y-value data is presented. The goal is to find the Harmonic Regression or model for this set of data.

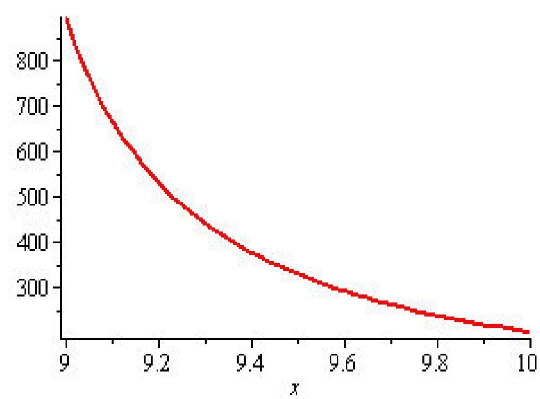

Fig. (4): Using HPDF algorithm for regression, the parameters k and a can be estimated. Maple code for this computation and the resulting graph is presented.

x	y	z=1/y
1	3	0.33333333
1.5	4.5	0.22222222
2	82	0.01219512
2.5	26.5	0.03773585
3	70	0.01428571
3.5	49.5	0.02020202
4	86	0.01162791
4.5	93.5	0.01069519
5	48	0.02083333
5.5	85.5	0.01169591
6	49	0.02040816
6.5	43.5	0.02298851
7	17	0.05882353
7.5	23.5	0.04255319
8	22	0.04545455

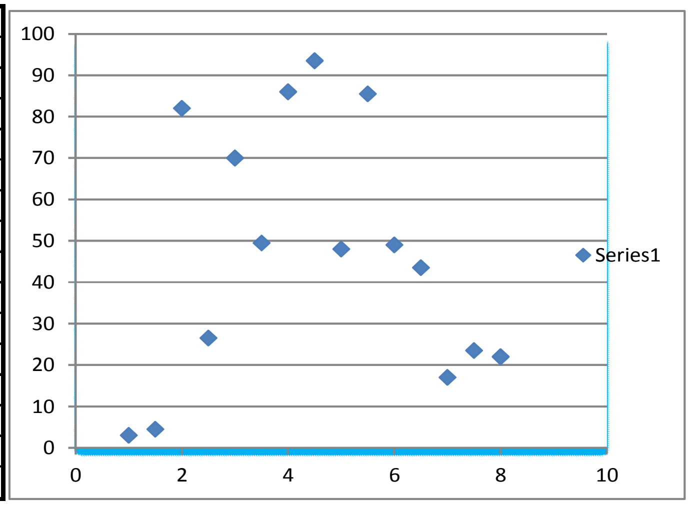

Table (2): In this table, values of y are randomly generated and z=1/y is evaluated.

Fig. (5): The Scatter Plot of the table (2) is demonstrated.

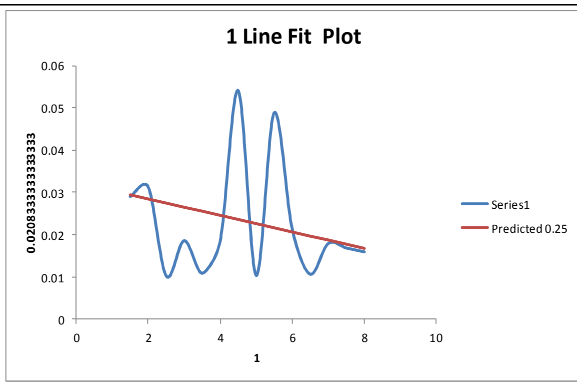

Fig. (6)- The graph of the randomly generated set of data (x,y) of Table (2) and the linear regression of z=mx+b is presented.

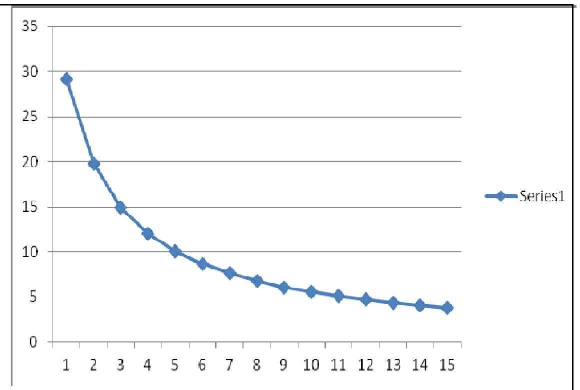

Fig. (7): The graph of the HPDF model y=k/(x-a).

References:

Adler, C. F., (Claire Fisher Adler) 1958, "Modern Geometry", An Integrated First Course, Second Edition, C.W. Post College of Long Island University, McGraw-Hill Book Company.

Ahangar, R. R., 2010, "Introduction to Statistical Geometry", Proceedings of the International Conference on Scientific Computing, Editors: Hamid R. Arabnia, George A. Gravvanis, CSREA Press, www.world-academy-of-Science.org,SCS.

Fisher, B. J., (Box, Joan Fisher), "R. A. Fisher, The Life of Scientist", New York: Wiley, 1978.

Hilbert, D., (1902), "The Foundations of Geometry", Chicago, IL: The Open Court Publishing Company.

Kapur, D., and Mundy, J. L., (1989), Geometric Reasoning", A Bradford Book. A collection of papers presented at international workshop on geometric reasoning held at Oxford University on 1986.

MacCluer, C. R. (2000), "Industrial Mathematics, Modeling in Industry, Science, and Government", Prentice Hall.

Mathews and Gilbert, (2006) "When Averaging Multiples, Apply the Harmonic Mean", BVUpdate: A Business Valuation Library Publication, www.BVLibrary.com.

Meyer D., (1970), "Probability and Statistics an Undergraduate Course", W.A. Benjamin INC. New York.

Pearson, R., (November 11, 2011), "Harmonic means, reciprocals, and ratios of random Variables", this article was first published on ExploringDataBlogs, and later contributed to R-bloggers: www.r-bloggers.com.

Saville J. D., and Wood, R. G., Third print (1997), "Statistical Methods: The Geometric Approach, Springer.

Tesi di, L. D., (2006-2007), "Simulation and Analysis of Chemical Reactions using Stochastic Differential Equations", Thesis, Universita Degli Studi Di Torino, Anno Academico.