Sine Square Distribution

A new statistical model based on the sine function

Rand Q. Al-Faris
Institute for Mathematical Research
University Putra Malaysia
Serdang 43400, Selangor Darul Ehsan, MALAYSIA
Email: randalfaris@hotmail.com
&
Shahjahan Khan*
Department of Mathematics & Computing
University of Southern Queensland
Toowoomba, Queensland, AUSTRALIA
Email: khans@usq.edu.au

Abstract

This paper introduces a new continuous distribution based on the sine function. The proposed Sine Square distribution has one parameter and its domain depends on this parameter. The probability density function $f(x)$ of a Sine Square variable X as well as its cumulative distribution function $F(x)$ are defined. The formulas for the r^{th} raw moment and central moments, moments generating function ($m.g.f.$), characteristic function ($c.f.$) and some other properties of the new distribution are provided. A method to generate random variables from the Sine Square distribution is analyzed and applied.

AMS 2000 Subject Classification : Primary 11K06, secondary 62E25.

Keywords: Sine function, probability and distribution functions, generating functions, and simulation of random variables.

1 Introduction

There are many probability distributions in statistical literature that are used in real-life for modelling a varieties of random phenomenon. These distributions cover both discrete and continuous variables. No one particular distribution is appropriate for modelling every phenomenon. Different variables are modelled by different probability distributions. Some distributions are based on algebraic functions of the underlying random variable (for instance normal and gamma distributions), and some others are based on trigonometric functions (such as von Mises distribution). Johnson et al. (1994, pp.172) covers almost all available statistical distributions along with their properties. Unfortunately, there are

*Corresponding author

only a few distributions that are based on trigonometric functions. In fact there are very limited number of distributions that are based on trigonometric functions. One of them is the von Mises distribution (also called the circular normal distribution) and is found in Kotz and Johnson (1982), for instance. Fisher (1993) provides a number of statistical tools to analyse circular data along with appropriate statistical models and inferencial methods. Nadarajah and Kotz (2006) discussed several beta-type distributions using some trigonometric functions. With the increased interest in the directional data researchers are looking forward to have more and more options in terms of the availability of distributions based on trigonometric functions.

This paper introduces a new statistical distribution using commonly used Sine function having one shape/growth parameter λ. The domain of the distribution depends on the parameter of the model. In addition to defining the probability density function $p.d.f.$ of the Sine square distribution, we derive the cumulative distribution function $c.d.f.$, moment generating function $m.g.f.$, characteristic function $c.f.$, and raw and central moments of the distribution. Furthermore, we discuss some important properties of the distribution.

The next Section introduces the $p.d.f.$ and $c.d.f.$ of the proposed Sine Square distribution. The moment and characteristic generating functions are derived in Section three. Section four provides the raw and central moments of the distribution. Some distributional properties of the Sine square variable are included in Section five. A method to generate Sine Square variables is covered in Section six. The final section contains some concluding remarks.

2 The Sine Square Distribution

Definition: Let X be a continuous random variable and $\lambda > 0$ be a positive real number. Then X is said to have a Sine Square distribution with parameter λ, if the probability density function $p.d.f.$ of X is expressed as:

$$f_X(x;\lambda) = \begin{cases} \frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}, & \text{if } 0 < x < \lambda\pi, \text{ and } \lambda > 0; \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

Note that $f_X(x;\lambda) \geq 0$ for all values of X in the domain $(0, \lambda\pi)$, and

$$\begin{aligned} \int_0^{\lambda\pi} f_X(x;\lambda)dx &= \int_0^{\lambda\pi} \frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}dx = \frac{1}{\lambda\pi}\int_0^{\lambda\pi}\left(1-\cos\frac{x}{\lambda}\right)dx \\ &= \left.\frac{x}{\lambda\pi}\right|_0^{\lambda\pi} - \left.\frac{\lambda}{\lambda\pi}\sin\frac{x}{\lambda}\right|_0^{\lambda\pi} = 1 \end{aligned} \tag{2.2}$$

We denote the above distribution as $X \sim Sin^2(\lambda)$, that is, the random variable X follows a Sine square distribution with parameter λ. Here λ is the shape/growth parameter. Smaller values of λ represent higher growth rate of the $p.d.f.$ curve, and larger values are related to its lower growth rate. The *p.d.f.* of the Sine Square distribution is displayed for some selected values of the parameter λ.

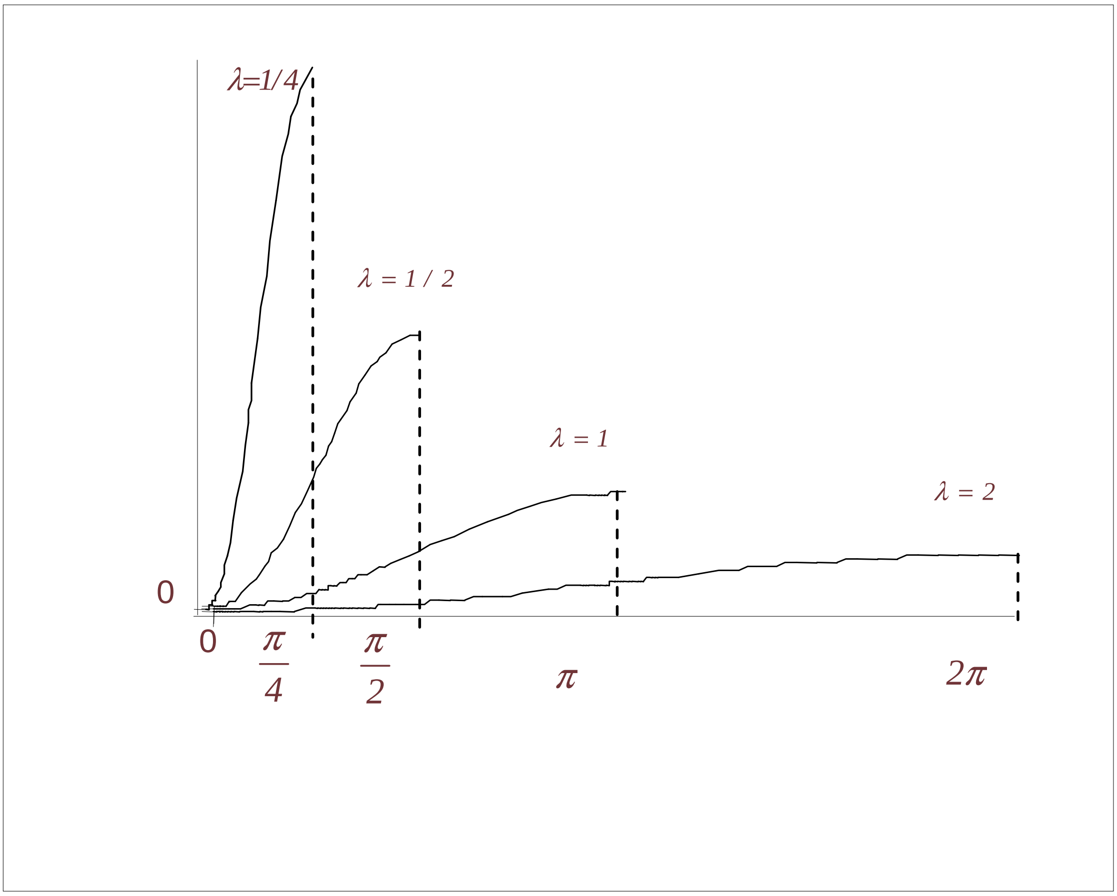

Figure 1: Graph of p.d.f. of Sine Square distribution for selected values of λ .

The cumulative distribution function *c.d.f.* of the Sine Square distribution is given by

$$F_X(t;\lambda) = P(X \leq t) = \int_0^t \frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}dx = \begin{cases} 0, & \text{if } t \leq 0 \\ \frac{t}{\lambda\pi} - \frac{1}{\pi}\sin\frac{t}{\lambda}, & \text{if } 0 < t < \lambda\pi \\ 1, & \text{if } t \geq \lambda\pi \end{cases} \tag{2.3}$$

For the proof, note that

$$P(X \leq t) = \int_0^t \frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}dx = \frac{2}{\lambda\pi}\left[\frac{x}{2} - \frac{\lambda}{2}\sin\frac{x}{\lambda}\right]_0^t = \frac{t}{\lambda\pi} - \frac{1}{\pi}\sin\frac{t}{\lambda} \tag{2.4}$$

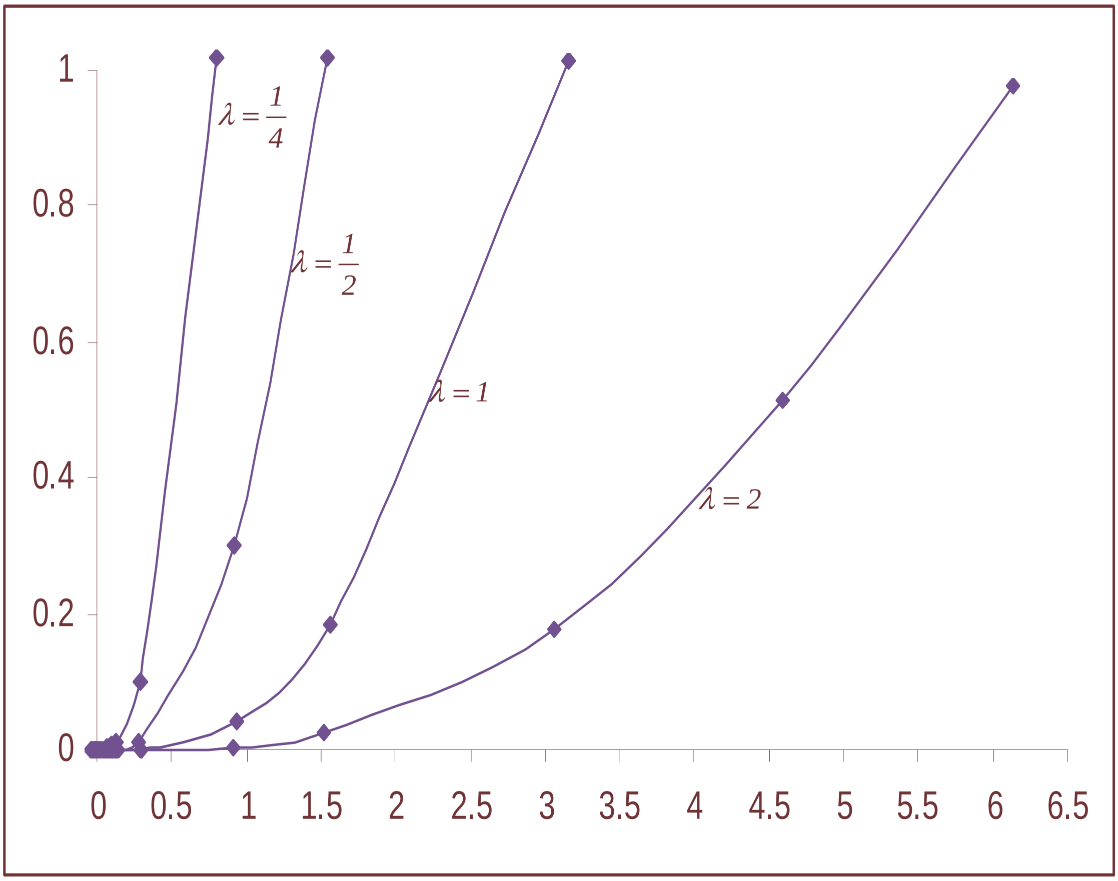

Figure 1: Graph of c.d.f. of Sine Square distribution for selected values of λ .

To work out the probability that X is less than any arbitrary value (say) x, we provide Table 1 containing the cumulative probabilities for selected choices of the parameter λ. Probability of any event defined within the domain of X can be obtained from this table. The table can be expanded for other choices of x, or λ, or both x and λ.

3 The Generating Functions

The moment generating function $m.g.f.$ and characteristic function $c.f.$ of the Sine Square distribution are provided in this Section.

Theorem 3.1: *If the random variable X follows a Sine Square distribution with parameter λ then the $m.g.f.$ and $c.f.$ of X are given by*

$$M_X(t) = \frac{4\left(e^{\pi\lambda}-1\right)}{\pi\left(4\lambda t+\lambda^3t^3\right)} \text{ and} \tag{3.1}$$

$$C_X(t) = \frac{\left(e^{\pi\lambda}-1\right)}{\pi}\frac{\left(16\lambda ti-4\lambda^3t^3(2-i)-\lambda^5t^5(1+i)\right)}{4\lambda t+\lambda^3t^3} \tag{3.2}$$

respectively.

Table 1: The cumulative probability of Sine Square Distribution for selected values of X and λ

x/λ	0.01	0.015	0.02	0.025	0.03	0.035	0.04	0.045	0.05	0.1	0.3	0.5	1	1.5	2
0.01π	1	0.391002	0.181690	0.097269	0.057669	0.036850	0.024921	0.017617	0.012902	0.001637	0.000061	0.000013	0.000002	0.000000	0.000000
0.015π		1	0.524921	0.297269	0.181690	0.118242	0.080920	0.057669	0.042482	0.005490	0.000205	0.000044	0.000006	0.000002	0.000001
0.02π			1	0.612902	0.391002	0.261099	0.181690	0.130970	0.097269	0.012902	0.000486	0.000105	0.000013	0.000004	0.000002
0.025π				1	0.674178	0.465421	0.330920	0.242082	0.181690	0.024921	0.000949	0.000205	0.000026	0.000008	0.000003
0.03π					1	0.719033	0.524921	0.391002	0.297269	0.042482	0.001637	0.000355	0.000044	0.000013	0.000006
0.035π						1	0.753188	0.573172	0.442482	0.066384	0.002595	0.000563	0.000070	0.000021	0.000009
0.04π							1	0.780020	0.612902	0.097269	0.003865	0.000840	0.000105	0.000031	0.000013
0.045π								1	0.801637	0.135609	0.005490	0.001194	0.000150	0.000044	0.000019
0.05π									1	0.181690	0.007512	0.001637	0.000205	0.000061	0.000026
0.1π										1	0.057669	0.012902	0.001637	0.000486	0.000205
0.3π											1	0.297269	0.042482	0.012902	0.005490
0.5π												1	0.1816	0.0576	0.0249
π													1	0.3910	0.1816
1.5π														1	0.5249
2π															1

Proof:
The moment generating function of X is given by

$$\begin{aligned} \mathrm{M}_X(t) &= \mathrm{E}(e^{tX}) = \int_0^{\lambda\pi} e^{tx}\frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}dx \\ &= \frac{1}{\lambda\pi}\int_0^{\lambda\pi} e^{tx}\left(1-\cos\frac{x}{2\lambda}\right)dx \\ &= \frac{e^{\lambda\pi t}-1}{\lambda\pi t}+\frac{\lambda^2 t(e^{\lambda\pi t}-1)}{\lambda\pi(4+\lambda^2t^2)} \end{aligned}$$

Simplification of the above expression establishes the result. Similarly, the characteristic function of X is given by:

$$\begin{aligned} \mathrm{C}_X(it) &= \mathrm{E}(e^{itX}) = \int_0^{\lambda\pi} e^{itx}\frac{2}{\lambda\pi}\sin^2\frac{x}{2\lambda}dx \\ &= \frac{-ie^{it\lambda\pi}-1}{\lambda\pi t}-\frac{\lambda^2 it(e^{it\lambda\pi}-1)}{\lambda\pi(4+it^2\lambda^2)} \end{aligned}$$

Proceeding in the same way as for the proof of $\mathrm{M}_X(t)$ above, the evaluation of the integral completes the proof.

For the verification of $M_X(t=0)=1$ consider the following:

$$M_X(t=0)=\left[\frac{e^{\lambda\pi t}-1}{\lambda\pi t}\right]_{t=0}+\left[\frac{\lambda^2t(e^{\lambda\pi t}-1)}{\lambda\pi(4+\lambda^2t^2)}\right]_{t=0}=\left[\frac{e^{\lambda\pi t}-1}{\lambda\pi t}\right]_{t=0} \tag{3.3}$$

Straightforward substitution of $t=0$ in the above expression would lead to an undefined value. However, direct expansion of $\frac{e^{\lambda\pi t}-1}{\lambda\pi t}$ by Maclaurin's series lead to

$$M_X(t=0)=\left[\frac{1+\lambda\pi t+\frac{(\lambda\pi t)^2}{2!}+\frac{(\lambda\pi t)^3}{3!}+\cdots-1}{\lambda\pi t}\right]_{t=0}=1 \tag{3.4}$$

4 The Moments

Finding of moments using $M_X(t)$ involves more complicated computations than finding them by using the direct definition of moments. In this Section we derive the general formulas for the r^{th} raw and central moments. The mean and variance are also provided here.

Theorem 4.1: *If the random variable X follows a Sine Square distribution with parameter λ then the mean, variance, and r^{th} raw and central moments of X are given by*

$$\mu = \mathrm{E}(X)=\frac{\lambda(\pi^2+4)}{2\pi} \tag{4.1}$$

$$\sigma^2 = \mathrm{Var}(X)=\frac{\lambda^2\pi^4-48}{12\pi^2} \tag{4.2}$$

$$\mu_r' = \frac{\lambda\pi}{r+1}-\frac{\lambda^r}{\pi}\begin{cases}\sum_{k=1}^{\frac{r}{2}}(-1)^{\frac{r}{2}-k+1}\pi^{2k-1}\frac{r!}{(2k-1)!}, & \text{if } r \text{ is even;}\\ 2(-1)^{\frac{r+1}{2}}r!+\sum_{k=2}^{\frac{r+1}{2}}(-1)^{\frac{r+1}{2}-k+1}\pi^{2k-2}\frac{r!}{(2k-2)!}, & \text{if } r \text{ is odd}\end{cases} \tag{4.3}$$

$$\mu_r = \frac{\lambda^r\left[(\pi^2+4)^{r+1}+(-1)^r(\pi^2-4)^{r+1}\right]}{(r+1)\pi}-\frac{\lambda^r}{\pi}\begin{cases}\zeta_r, & \text{if } r \text{ is even;}\\ \eta_r, & \text{if } r \text{ is odd}\end{cases} \tag{4.4}$$

where

$$\zeta_r = \sum_{k=1}^{\frac{r}{2}} (-1)^{\frac{r}{2}-k+1} \frac{r!}{(2k-1)!} \left[(\pi^2+4)^{2k-1} - (\pi^2-4)^{2k-1} \right] , \quad (4.5)$$

$$\eta_r = (-1)^{\frac{r+1}{2}} r! + \sum_{k=2}^{\frac{r+1}{2}} (-1)^{\frac{r+1}{2}-k+1} \frac{r!}{(2k-2)!} \left[(\pi^2+4)^{2k-2} - (\pi^2-4)^{2k-2} \right] \quad (4.6)$$

respectively.

Proof:

The proof of the mean and variance follows from the r^{th} raw and central moments for $r = 1$ and $r = 2$ respectively. First we proof equation (4.3). The r^{th} raw moment is given by

$$\mathrm{E}(X^r) = \int_0^{\lambda\pi} x^r f_X(x;\lambda) dx \text{ for } r = 1, 2, 3, ... \quad (4.7)$$

Then for $r = 1$ we get

$$\mathrm{E}(X) = \int_0^{\lambda\pi} x f_X(x;\lambda) dx = \frac{2}{\lambda\pi} \int_0^{\lambda\pi} x \sin^2 \frac{x}{2\lambda} dx = \frac{\lambda\pi}{2} - \frac{1}{\lambda\pi} \int_0^{\lambda\pi} x \cos \frac{x}{\lambda} dx \quad (4.8)$$

Let $I_1 = \int_0^{\lambda\pi} x \cos \frac{x}{\lambda} dx$, then

$$\mathrm{E}(X) = \frac{\lambda\pi}{2} - \frac{I_1}{\lambda\pi} \quad (4.9)$$

Similarly, for $r = 2$ we get

$$\begin{aligned} \mathrm{E}(X^2) &= \int_0^{\lambda\pi} x^2 f_X(x;\lambda) dx = \frac{2}{\lambda\pi} \int_0^{\lambda\pi} x^2 \sin^2 \frac{x}{2\lambda} dx \\ &= \frac{(\lambda\pi)^2}{3} - \frac{1}{\lambda\pi} \int_0^{\lambda\pi} x^2 \cos \frac{x}{\lambda} dx \end{aligned} \quad (4.10)$$

Let $I_2 = \int_0^{\lambda\pi} x^2 \cos \frac{x}{\lambda} dx$, then

$$\mathrm{E}(X^2) = \frac{(\lambda\pi)^2}{3} - \frac{I_2}{\lambda\pi} \quad (4.11)$$

Continuing the process we get

$$\mathrm{E}(X^r) = \frac{(\lambda\pi)^r}{r+1} - \frac{I_r}{\lambda\pi} \quad (4.12)$$

where

$$I_r = \int_0^{\lambda\pi} x^r \cos \frac{x}{\lambda} dx \text{ for } r = 1, 2, 3, \dots \quad (4.13)$$

The evaluation of I_r leads to

$$\begin{aligned} I_1 &= -2\lambda^2 \\ I_2 &= -2\lambda^3\pi \\ I_3 &= -3\lambda^4(\pi^2-4) \\ &\vdots \\ I_r &= \begin{cases} \lambda^{r+1} \sum_{k=1}^{\frac{r}{2}} (-1)^{\frac{r}{2}-k+1} \pi^{2k-1} \frac{r!}{(2k-1)!}, & \text{if } r \text{ is even;} \\ \lambda^{r+1} \left[2(-1)^{\frac{r+1}{2}} r! + \sum_{k=2}^{\frac{r+1}{2}} (-1)^{\frac{r+1}{2}-k+1} \pi^{2k-2} \frac{r!}{(2k-2)!} \right], & \text{if } r \text{ is odd} \end{cases} \end{aligned} \quad (4.14)$$

Substituting the value of I_r from (4.14) in (4.12), we get the required equation (4.3). To find the mean of X, we substitute the value of I_1 in (4.9).

To prove (4.4), by definition, the r^{th} central moment is given by

$$\mu_r = \mathrm{E}[(X-\mu)^r] = \int_0^{\lambda\pi} (x-\mu)^r f_X(x;\lambda)dx \text{ for } r = 1,2,3,\ldots \tag{4.15}$$

Following similar steps, as above, the expressions for μ_r is obtained. Finally, to find the variance of X, we substitute $r = 2$ in (4.15)

With this,the proof is accomplished.

5 Some Distributional Properties

If the random variable X follows a Sine Square distribution with parameter λ then the following properties hold.

1. In general, if

$$f_X(x) = \begin{cases} \frac{2}{n\lambda\pi}\sin^2\frac{x}{2n\lambda}, & 0 < x < n\lambda\pi, \text{ for any real } n; \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

   then $X \sim Sin^2(n\lambda)$.

2. If $X \sim Sin^2(\lambda)$ and $Y = mX$, then $Y \sim Sin^2(m\lambda)$ where m is a non-zero real valued constant.

3. If $X \sim Sin^2(\lambda)$ and $\lambda = \frac{n}{\pi}$ where n is an integer number then the *p.d.f.* of X becomes

$$f_X(x) = \begin{cases} \frac{2}{n}\sin^2\frac{x\pi}{2n}, & 0 < x < n; \quad n = 1,2,3,\ldots \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

   In this case X is said to have a Sine Square distribution over the interval $(0, n)$. Here the domain of X does not depend on λ when $\lambda = \frac{n}{\pi}$.

4. If $X \sim Sin^2(\lambda)$ then the *coefficient of skewness* of the distribution is given by

$$\mathrm{Sk} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = -\frac{12\sqrt{12}(\lambda\pi)^3\left[\frac{4\pi}{3} + \frac{64}{3\pi^3} + \frac{96}{\pi}\right]}{[\lambda^2\pi^4 - 48]^{\frac{3}{2}}} \tag{5.3}$$

5. If $X \sim Sin^2(\lambda)$ then the *coefficient of kurtosis* of the distribution is given by

$$\mathrm{Ku} = \frac{\mu_4}{\mu_2^2} - 3 = \frac{144(\lambda\pi)^4\left[\frac{\pi^4}{80} + \frac{24}{\pi^2} + \frac{128}{5\pi} - 1\right]}{[\lambda^2\pi^4 - 48]^2} - 3 \tag{5.4}$$

6. If $X \sim Sin^2(\lambda)$ then the *median* of the distribution is given by the solution $\tilde{x}$ of the following equation:

$$P(X \leq \tilde{x}) = 0.50 \tag{5.5}$$

or equivalently the solution of

$$\tilde{x} - \lambda \sin\left(\frac{\tilde{x}}{\lambda}\right) = \frac{\lambda\pi}{2} \tag{5.6}$$

Solving the above non-linear equation by Newton-Raphson (see for example, Epperson, 2002) method we get the *median* as

$$\tilde{x} = 1.659019676\lambda \tag{5.7}$$

Therefore, both the mean and median of the Sine Square distribution depends on the shape/growth parameter, λ, of the distribution.

6 Generating Sine Square Random Variables

In this Section the *Acceptance-Rejection method* is used to generate Sine Square random variables (Rubinstein, 1981, p.45-50). This method generates random variables from any specific distribution and accept it after checking if it is from the distribution, otherwise reject it. The method is described as follows.

Let X be a random variable generated from a distribution with *p.d.f.* $f(x)$. The density function of X can be expressed as

$$f(x) = c \times h(x) \times g(x) \tag{6.1}$$

where $c \geq 1$ represents the mean number of trails, $h(x)$ is a *p.d.f.* and $0 < g(x) < 1$. Generate one random variable U from $U(0, 1)$ distribution and another one Y from the density $g(y)$. If the inequality $U \leq g(Y)$ holds, accept $X = Y$ as a variable generated from $f(x)$, otherwise reject it and repeat the trail again.

For generating random variables from the Sine Square distribution, the *p.d.f.* of the distribution of X can be written as

$$f(x) = 2 \times \frac{1}{\lambda\pi} \times \sin^2 \frac{x}{2\lambda}, \text{ when } 0 < x < \lambda\pi \tag{6.2}$$

so that $c = 2$, $h(x) = \frac{1}{\lambda\pi}$, an uniform *p.d.f.* over the interval $(0, \lambda\pi)$, and $g(x) = \sin^2 \frac{x}{2\lambda}$. Note that $0 < \sin^2 \frac{x}{2\lambda} < 1$.

First generate a random variate U_1 from a $U(0, 1)$ distribution and then independently generate another random variate Y from $h(y)$. Set $U_2 = H(Y)$ where $H(Y)$ is the *c.d.f.* associated with $h(y)$. Then from $H(y) = \frac{y}{\lambda\pi}$, set $\frac{Y}{\lambda\pi} = U_2$ so that $Y = \lambda\pi U_2$. If $U_1 \leq g(Y)$, with $g(Y) = \sin^2 \frac{\pi U_2}{2}$, accept U_1 as a Sine Square random variate, otherwise repeat the process again.

The Algorithm

1. Select an appropriate value of λ.
2. Generate U_1 and U_2 from $U(0,1)$ and $U(0,\lambda\pi)$ distributions respectively.
3. Set $Y = \lambda\pi U_2$.
4. If $U_1 \leq \sin^2 \frac{\pi U_2}{2}$, accept $X = Y$ as a random variate from the Sine Square distribution, otherwise go to step 2.

In Mont Carlo simulation, we compute the simulated mean of the distribution, the simulated mean number of trials and the simulated efficiency for the parameter values vary $0.01(0.005)0.05, 0.1(0.2)0.5, 1(0.5)2$ with run size 1000 and the results represented below in Table 2 together with the theoretical mean, mean number of trials and the efficiency.

Table 2: Theoretical and simulated means and efficiency by *Monte Carlo* simulation with theoretical mean No. of trials ($c = 2$) and theoretical efficiency ($1/c = 0.5$).

λ	Theoretical Mean	Simulation Mean	Simulation Mean No. of Trails	Simulation Efficiency
0.01	0.0221	0.0219	1.984	0.504
0.015	0.0331	0.0333	2.0084	0.4979
0.02	0.0441	0.0441	2.0024	0.4994
0.025	0.0552	0.0548	1.9842	0.504
0.03	0.0662	0.0662	2.0088	0.4978
0.035	0.0773	0.0779	1.9998	0.5001
0.04	0.0883	0.0879	1.9936	0.5016
0.045	0.0993	0.0988	1.973	0.5068
0.05	0.1104	0.1097	2.0092	0.4977
0.1	0.2207	0.2213	1.9826	0.5044
0.3	0.6622	0.6602	1.9748	0.5064
0.5	1.1037	1.1021	1.9812	0.5047
1	2.2074	2.2182	1.984	0.504
1.5	3.3111	3.2984	2.0028	0.4993
2	4.4148	4.3771	2.0028	0.4993

7 Concluding Remarks

The proposed distribution can be used to model truncated data. The truncation may be due to Government policies, natural causes or termination of a study at some point. Unlike the maximum likelihood estimator of the parameter λ, the method of moment estimator is easily obtainable. For an independent sample of size n from the Sine Square distribution the method of moment estimator for the shape/growth parameter is $\hat{\lambda} = \frac{2\pi}{\pi^2+4}\bar{X}$ where $\bar{X}$ is the sample mean.

References

Epperson, J. F. (2002). An Introduction to Numerical Methods and Analysis. Wiley, New York.

Fisher, N. I. (1993). Statistical Analysis of Directional Data, Cambridge University Press, UK.

Johnson, N. L., Kotz, S. and Balakrsihnan, N. (1994). Continuous Univariate Distributions. **1**, 2nd ed., Wiley, New York.

Kotz, S. and Johanson, N. L. (1982). Circular normal distribution. Encyclopedia of Statistical Sciences, **1**, 479.

Nadarajah, S. and Kotz, S. (2006). Beta Trogonometric Distribution. Portuguese Economic Journal, Vol 5, No 3, pp 207-224.

Rubinstein, R. Y. (1981). Simulation and Mont Carlo Methods, Wiley, New York.