AN IMPROVEMENT OF MINH'S ALGORITHM FOR GENERATING GAMMA VARIATES WITH ANY VALUE OF SHAPE PARAMETER

Nguyen Van Hung
Danang Vocational Training College, 99 To HienThanh
Danang City, Vietnam
nguyenvanhungdn1@gmail.com

NgoThiThanhTrang
Danang Vocational Training College, 99 To HienThanh
Danang City, Vietnam
trangngothanh@gmail.com

Tran QuocChien
Danang University, 41 Le Duan,
Danang City, Vietnam
tqchien@dce.ud.vn

Abstract

The algorithm of Minh as in [Minh (1988)] was used to generate variates having a gamma distribution with shape parameter a>1 only. In this paper, a method, which is the improvement of the algorithm of Minh is introduced for the generation of independent random variables from a gamma distribution with all values of shape parameter and is compared with the method of Marsaglia and Tsang. By means of computer simulation, for each method and each value of shape parameter, a series of 10.000 gamma variables was generated, and then the speed, the randomness and the preservation of the numerical characteristics, namely expected value, variance and skewness coefficient are considered. It is found that the speed and the randomness of the two methods are the same, however, the preservation of the numerical characteristics of the gamma distribution by the proposed method is much better than the method of Marsaglia and Tsang.

Keywords**:** Algorithm of Minh; Algorithm of Marsaglia and Tsang; Gamma random variable; Improvement of Minh's algorithm.

1. Introduction

Generating gamma random variates is a very important problem in the statistical literature. It is well-known that the available algorithms can be divided into two distinct cases. Case 1: Shape parameter $a\leq 1$; Case 2: Shape parameter a>1. For case 1, the most popular and very simple method proposed by [Ahrens and Dieter (1974)]. For case 2, by the acceptance-rejection principle [Minh (1988)] proposed the very good algorithm to generate the gamma variates, and [Hung and Chien (2013)] used these algorithms to generate the gamma random numbers and obtained the very good results in computer simulation of reservoir storage. For both cases, [Marsaglia and Tsang (2000)] used the Monty Python method and proposed the algorithm for generating gamma variates for all values of shape parameter, and recently this algorithm was introduced by [Hong LiangJie(2012)] as a one of the best algorithms which was used in GSL Library and Matlab "gamrnd".

By theoretical and statistical analysis, so many algorithms have been proposed for generating the random numbers with the specific type of distribution, in which the speed, the simplicity and the ease to implement on the computers were considered, several works considered the randomness of the random numbers generators such as in the work of [Boiroju and Reddy (2012)] but there is no any work evaluates the efficiency of the different random numbers algorithms based on the preservation of the numerical characteristics of the distribution. To do this, the computer simulation experiments can be applied. The autocorrelation coefficient of the generated random number series is investigated to test the randomness and to evaluate the preservation of the numerical characteristics of the distribution based on the mean, variance and the skewness of the series of generated data. In this paper, the proposed method which is an improvement of Minh's algorithm to generate random gamma variables for all values of shape parameter is compared with the algorithm of Marsaglia and Tsang. By computer simulation experiments, the series of gamma random numbers are generated by using these algorithms, and the speed, the randomness and the preservation of the numerical characteristics of the gamma distribution are considered, and, that is the subject of this paper.

2. Literature Review

2.1. *The statistical descriptors*

The statistical descriptors of a series of random numbers$\{X_1, X_2, \ldots, X_N\}$, namely *m, s*², *g, r* are the mean, variance, skewness coefficient and the lag-one autocorrelation coefficient, respectively. These statistical descriptors are expressed as follow:

$$m = \frac{1}{N}\sum_{i=1}^{N} X_i \quad (1)$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - m)^2 \quad (2)$$

$$g = \frac{N}{(N-1)(N-2)s^3}\sum_{i=1}^{N}(X_i - m)^3 \quad (3)$$

$$r = \frac{1}{(N-1)s^2}\sum_{i=1}^{N-1}(X_i - m)(X_{i+1} - m) \quad (4)$$

2.2. *The gamma distribution*

A continuous random variable *X* is said to have a three-parameter gamma distribution if its density can be expressed as

$$f(x) = \frac{(x-c)^{a-1}e^{-(x-c)/b}}{b^a\Gamma(a)}$$

Where *a>0, b>0, c>0, x≥c and a, b, c* are respectively the shape, scale, and location parameters. The gamma function is defined by

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt, \quad a > 0$$

This function satisfies the following recursive formula

$$\Gamma(a+1) = a\Gamma(a)$$

and for *a*= *k* (a positive integer), we have :

$$\Gamma(k) = (k-1)! = 1 * 2 * \ldots (k-1)$$

The numerical characteristics of the three-parameter gamma distribution are given by the following formulas:

Expected value: $E(X) = ab + c$

Variance: $Var(X) = ab^2$

Skewness coefficient: $g = 2/\sqrt{a}$

When $c = 0$ we have the two-parameter gamma distribution, and, when $c = 0$ and $b = 1$ we have the one-parameter gamma distribution. By transformation method, the gamma variables with two parameters or three parameters can be converted into the gamma variables with one parameter. For three-parameter variables, the transformed variables can be obtained by setting $y = (x-c)/b$ or $x = c + by$. For two-parameter variables, the transformation used is $y = x/b$ or $x = by$.

2.3. *Generation of gamma variables*

2.3.1. *Minh's algorithm*

The algorithm works for *X* is the gamma variable with one parameter for shape $a>1$:

Initialization:

(1) Set $m = a-1$, $D = \sqrt{m}$

(2) If $1 < a \le 2$, set $D_1 = m/2$, $x_1 = 0$, $x_2 = D_1$, $x_s = -1$, $f_1 = 0$, goto (4)

(3) Set $D_1 = D - 0.5$, $x_2 = m - D_1$, $x_1 = x_2 - D_1$, $x_s = 1 - m/x_1$, $f_1 = \sum_{i=1}^{m} -\ln U_i$

(4) Set $f_2 = e^{m\ln(x_2/m)+D_1}$, $x_4 = m+D$, $x_5 = x_4+D$, $x_r = 1 - m/x_5$, $f_4 = e^{m\ln(x_4/m)-D}$, $f_5 = e^{m\ln(x_5/m)-2D}$

$p_1 = 2Df_4$, $p_2 = 2D_1f_2+p_1$, $p_3 = f_5/x_r+p_2$, $p_4 = -f_1/x_s+p_3$

Generation:

(5) Generate a uniform random number $U \sim U(0,1)$

Set $U = Up_4$

If $U > p_1$, goto (7)

Set $w = U/D - f_4$

If $w \leq 0$, deliver $X = m+U/f_4$

If $w \leq f_5$, deliver $X = x_4+wD/f_5$

(6) Generate a uniform random number $V \sim U(0,1)$

Set $X = x_4+VD$, $x' = 2x_4-X$

If $w \geq f_4+(f_4-1)(X-x_4)/(x_4-m)$, deliver $X=x'$

If $w \leq f_4 + (m/x_4-1)f_4(X-x_4)$, deliver $X= x_4+VD$

If $w < 2f_4-1$, goto (11)

If $w < 2f_4 - e^{mln(X'/m)+m-X'}$, goto (11); otherwise, deliver $X = x'$

(7) If $U > P_2$, goto (9)

Set $w = (U-p_1)/D_1 - f_2$

If $w \leq 0$, deliver $X = m-(U-p_1)/f_2$

If $w \leq f_1$, deliver $X = x_1 + wD_1/f_1$

(8) Generate a uniform random number $V \sim U(0,1)$

Set $X = x_1+VD_1$, $x' = 2x_2 - X$

If $w \geq f_2+(f_2-1)(X-x_2)/(x_2-m)$, deliver $X=x'$

If $w \leq f_2 + (X-x_1)/D_1$, deliver $X= x_1+VD_1$

If $w < 2f_2-1$, goto (11)

If $w < 2f_2 - e^{mln(X'/m)+m-X'}$, goto (11); otherwise, deliver $X = x'$

(9) Generate a uniform random number $w \sim U(0,1)$

If $U \geq p_3$, goto (10)

Set $U = (p_3-U)/(p_3-p_2)$, $X = x_5 - ln(U)/x_r$

If $w \leq [x_r(x_5-X)+1]/U$, deliver $X = x_5 - ln(U)/x_r$

Set $w = wf_5U$, goto (11)

(10) Set $U = (p_4-U)/(p_4-p_3)$, $X = x_1-ln(U)/x_s$

If $X \leq 0$, goto (5)

If $w \leq [x_s(x_1-X)+1]/U$, deliver X

Set $w = wf_1U$

(11) If $ln(w) > mln(X/m) + m - X$, goto (5); otherwise, deliver $X = x_1-ln(U)/x_s$.

Basically, this algorithm requires two uniform random variables to generate a single gamma random number with one-parameter gamma distribution, and just one time of initialization to generate a series of gamma random numbers.

2.3.2. *Marsaglia and Tsang's algorithm*

The algorithm works for X is the gamma variable with one parameter for $a \geq 1$:

(1) Set $d = a - 1/3$ and $c = 1/\sqrt{9d}$.

(2) Generate a standard normal random number $Z \sim N(0,1)$ and a uniform random number $U \sim U(0,1)$ independently.

(3) If $Z > -1/c$ and $\ln U < 1/2Z^2 + d - dV + d \times lnV$, where $V = (1+cZ)^3$, deliver $X = d \times V$; otherwise, go back to Step 2.

The algorithm can be easily extended to the cases where $1>a>0$. Generating X' is the gamma variable with one parameter for shape $a+1$, then deliver $X = X' \times U^{1/a}$ where $U \sim U(0,1)$. Thus, X is the gamma variable with one parameter for shape a. See details in the work of [Marsaglia and Tsang (2000)].

This algorithm requires one uniform random variable and one standard normal random variable to generate a single gamma random number for shape $a \geq 1$ and one more of uniform random variable for shape $1 > a > 0$.

To generate a single standard normal random number, the algorithm proposed by [Rao *et al.*(2011)] is used:

(1) Generate a uniform random number $U \sim U(0,1)$.

(2) Deliver $Z = \frac{-\ln\left(\frac{1}{U}-1\right)}{1.702}$

3. The Proposed Method

Based on the work of [Marsaglia and Tsang (2000)] the proposed method from an improvement of Minh's algorithm is as follows:

(1) If $a>1$ generate X *by* Minh's algorithm with shape a; Deliver X .

(2) If $1 \geq a > 0$ generate X' *by* Minh's algorithm with shape $a+1$; Deliver $X = X' \times U^{1/a}$ where $U \sim U(0,1)$.

This method is applied for any value of shape parameter.

4. Computer simulation

To generate the series of the gamma random variables, the proposed method and the method of Marsaglia and Tsang were used. All the programs were coded in the C language for computer simulation experiments and tested on the computer with Intel(R) Atom CPU N570- 32 bit.

For each value of the skewness coefficient of the gamma distribution, a moderate sample of 10,000 gamma random numbers was generated on the computer using the algorithms of the two methods. The speed, the randomness and the statistical descriptors, namely mean value, variance, skewness coefficient and lag-one autocorrelation coefficient were investigated. In these experiments, the value of skewness coefficient of the gamma distribution was in the range [0.1, 500]. The time consuming was recorded, the statistical descriptors of the series of generated data were computed by using Eqs. (1) – (3) and used to consider the preservation of the numerical characteristics of the gamma distribution, and the lag-one autocorrelation coefficient is computed by using Eq. (4) and used to test the randomness. Table 1 indicates the speed and the randomness, and Tables 2-4 indicate the preservation of the numerical characteristics of the distribution. The results are as follows:

Table 1. The Autocorrelation coefficients of 10,000 generated gamma variables and average generation times (in Milliseconds).

Shape *a*	**a) Average Marginal Time**		**b) Autocorrelation Coefficient**	
	Proposed method	**Marsaglia and Tsang's method**	**Proposed method**	**Marsaglia and Tsang's method**
0.1	0.55	0.57	0.015	0.004
0.3	0.50	0.61	0.008	0.008
0.5	0.55	0.55	0.001	-0.004
0.7	0.49	0.52	0.000	-0.013
0.9	0.53	0.55	-0.012	-0.013
1.0	0.55	0.55	0.013	0.003
1.5	0.25	0.28	-0.003	-0.011
2.0	0.27	0.34	-0.003	-0.028
2.5	0.29	0.33	0.006	-0.045
3.0	0.27	0.29	0.006	-0.034
5.0	0.27	0.28	-0.005	-0.003
10.0	0.29	0.30	-0.019	-0.015
15.0	0.33	0.30	-0.004	-0.018
30.0	0.33	0.33	-0.004	-0.002
50.0	0,32	0.31	-0.001	0.020
100.0	0.32	0.29	-0.014	-0.003
300.0	0.33	0.31	0.008	0.011
500.0	0.33	0.33	0.013	0.019

Table 2. Mean values of 10,000 generated gamma variables.

Shape *a*	a) Proposed method		b) Marsaglia and Tsang's method	
	Generated data	**Relative error in percentage**	**Generated data**	**Relative error in percentage**
0.1	0.099	-0.78	0.114	14.32
0.3	0.296	-1.27	0.343	14.38
0.5	0.498	-0.41	0.564	12.79
0.7	0.693	-1.04	0.778	11.14
0.9	0.914	1.55	0.980	8.94
1.0	0.984	-1.60	1.350	35.03
1.5	1.508	0.55	1.838	22.53
2.0	2.004	0.19	2.324	16.20
2.5	2,506	0.25	2.820	12.81
3.0	3.034	1.12	3.325	10.82
5.0	4.967	-0.68	5.333	6.67
10.0	10.018	0.18	10.391	3.91
15.0	14.911	-0.59	15.314	2.09
30.0	30.026	0.09	30.307	1.02
50.0	50.114	0.23	50.138	0.28
100.0	100.130	0.13	100.287	0.29
300.0	300.020	0.01	300.183	0.06
500.0	499.874	-0.03	500.223	0.05

Table 3. Variances of 10,000 generated gamma variables.

Shape *a*	a) Proposed method		b) Marsaglia and Tsang's method	
	Generated data	**Relative error in percentage**	**Generated data**	**Relative error in percentage**
0.1	0.098	-1.79	0.094	-6.44
0.3	0.273	-8.03	0.270	-10.08
0.5	0.483	-3.42	0.416	-16.71
0.7	0.668	-4.53	0.562	-19.74
0.9	0.937	4.12	0.684	-23.99
1.0	0.961	-3.86	1.351	35.06
1.5	1.531	2.04	1.874	25.02
2.0	1.983	-0.84	2.322	16.09
2.5	2.511	0.44	2.932	17.29
3.0	3.132	4.40	3.295	9.83
5.0	4.922	-1.56	5.213	4.25
10.0	10.245	2.45	10.491	4.91
15.0	14.934	-0.44	15.399	2.66
30.0	29.272	-2.43	30.630	2.10
50.0	49.548	-0.90	49.924	-0.15
100.0	98.376	-1.62	98.790	-1.21
300.0	299.550	-0.15	289.798	-3.40
500.0	499.916	-0.02	507.709	1.54

Table 4. Skewness coefficients of 10,000 generated gamma variables.

Shape *a*	Skewness coefficients of gamma distribution	a) Proposed method		b) Marsaglia and Tsang's method	
		Generated data	Relative error in percentage	Generated data	Relative error in percentage
0.1	6.325	6.752	6.75	4.524	-28.47
0.3	3.651	3.530	-3.34	2.429	-33.47
0.5	2.828	2.898	2.45	1.912	-32.40
0.7	2.390	2.422	1.30	1.653	-30.87
0.9	2.108	2.048	-2.86	1.393	-33.93
1.0	2.000	2.046	2.28	1.698	-15.08
1.5	1.633	1.704	4.33	1.538	-5.81
2.0	1.414	1.445	2.20	1.292	-8.63
2.5	1.265	1.306	3.25	1.315	3.94
3.0	1.155	1.260	9.14	1.028	-10.96
5.0	0.894	0.838	-6.37	0.780	-12.79
10.0	0.632	0.637	0.68	0.606	-4.131
15.0	0.516	0.515	-0.31	0.490	-5.143
30.0	0.365	0.335	-8.16	0.332	-9.013
50.0	0.283	0.255	-9.85	0.248	-12.25
100.0	0.200	0.193	-3.43	0.163	-18.41
300.0	0.115	0.133	14.88	0.107	-7.27
500.0	0.089	0.117	30.48	0.071	-20.40

5. Conclusion

The following conclusions are drawn from this study:

- The algorithm proposed by Minh is used for the case of the value of shape parameter a>1 only, whereas, the proposed method from the improvement of Minh's algorithm will be applied for any value of shape parameter of the gamma distribution.
- To evaluate the efficiency of an algorithm to generate the random numbers, the randomness and the preservation of the numerical characteristics, namely the expectation, variance and skewness coefficient of the distribution are considered also.
- As the results indicate in Table 1(a), the time consuming on the computer to generate the series of gamma random numbers by the proposed method and the method of Marsaglia and Tsang are the same and the results in Table 1(b) showed that the absolute values of the autocorrelations of the series of gamma random numbers generated are significantly in the range of zero to 0.05. This proves that the generated samples are the random samples.
- From the Tables 2 - 4, it is observed that the values of the relative errors in percentage of the mean, variance and skewness coefficient of the series data obtained by using the proposed method are very small for almost the values of shape parameter. Especially, for the cases of shape a<3, it is found that the numerical characteristics of the gamma distribution can be preserved very well by the proposed method, much better than those by the method of Marsaglia and Tsang.

Acknowledgments

The authors wish to express deep gratitude and sincere appreciation to Professor Huynh Ngoc Phien, former Professor of Asian Institute of Technology for his kind guidance, helpful suggestions and keen interest throughout the study period of the work reported in this paper.

References

[1] Ahrens, J. H; Dieter, U. (1974): Computer methods for sampling from gamma, beta, Poisson and Binomial distribution. Computing, vol. 12. pp. 223-246.

[2] Boiroju; Naveen Kumar; M. Krishna Reddy.(2012):Generation of standard normal random numbers.InterStat.

[3] HongLiangJie. (2012): How to Generate Gamma Random Variables. Research Scientist at Yahoo Labs http://www.Hongliangjie.com/2012/12/19/how-to-generate-gamma-random-variables/.

[4] Marsaglia, G.;Tsang, W. (2000): A simple method for generating gamma variables. ACM Transactions on Mathematical Software, vol. 26, No. 3, pp. 363-372.

[5] Minh, Do Le. (1988): Generating gamma variates. ACM Transactions on Mathematical Software (TOMS), 14.3 (1988): 261-266.

[6] Hung, Nguyen Van; Chien, Tran Quoc. (2013): Computer simulation and approximate expression for the mean range of reservoir storage with GAR (1) inflows. In Proceedings of the Fourth Symposium on Information and Communication Technology, ACM New York, NY, USA, pp. 11-17.

[7] Rao;Kollu Ranga; Naveen Kumar Boiroju; M. Krishna Reddy.(2011): Generation of standard normal random variables. Indian Journal of Scientific Research 2, no. 4, pp. 83-85.