Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2020;12(3):242-51

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2020-78851

Random Number Generation Based on Characteristics Functions

Karakteristik Fonksiyonlara Dayalı Rastgele Sayı Üretimi

Xiaohan MEI[a], Xinhe WANG[a], Mehmet Onur KAYA[b], Hakan DEMİRTAŞ[a]

[a]Division of Epidemiology and Biostatistics, University of Illinois at Chicago, USA
[b]Department of Biostatistics and Medical Informatics, Fırat University Faculty of Medicine, Elazığ, TURKEY

ABSTRACT Objective: In probability theory and statistics, some distributions can be better presented by the characteristic function (CF) than the conventional probability density function (PDF). In random number generation (RNG) domains, algorithmic applications based on a known CF is far less common compared to the ones that are driven by a density. An acceptance-rejection algorithm that employs CF appeared in the literature in the context of RNG; however, validity, plausibility, and utility of this algorithm have never been examined by a simulation study, which we attempt to address. **Material and Methods:** We devised a simulation study based on three commonly encountered univariate distributions (Normal, Laplace, and Gamma), and compared the performance of the CF algorithm with the default random number generation tools in R software. **Results:** All three simulation studies yielded similar outcomes across the two methods with indiscernible differences. **Conclusion:** The simulation results for the three distributions we considered suggest that the performance of the CF algorithm aligns well with that of the default algorithm. It is better to harness a direct algorithm when the PDF of a random variable is given. However, the CF-based method can offer a solid alternative for situations where the PDF is hard to sample from. Our findings provide a promising basis for conducting further research on generating data in difficult PDF and straightforward CF scenarios. In brief, the CF method is an excellent alternative to the default for generating data when CF is given or can be derived in closed form.

Keywords: Simulation; random number generation; characteristics function; probability density function; algorithm

ÖZET Amaç: Olasılık teorisi ve istatistikte kullanılan bazı dağılımlar bilinen geleneksel olasılık yoğunluk fonksiyonlarından (PDF) çok karakteristik fonksiyonları (CF) ile daha kolay incelenebilir. Rasgele sayı üretimi (RNG) yöntem ve algoritmalarında çoğunlukla dağılımın PDF'si tercih edilmekte, bilinen bir CF'ye dayalı yöntemler ise daha az kullanılmaktadır. RNG literatüründe CF'yi kullanan kabul-ret algoritmaları kısıtlı sayıda olmasına ragmen bu algoritmaların geçerliliği, akla yatkınlığı ve sağladığı yararlar bizim bu çalışmada ortaya koyduğumuz gibi bir similasyon çalışmasıyla hiç bir zaman incelenmemiştir. **Gereç ve Yöntemler:** Yaygın olarak karşılaşılan tek değişkenli üç dağılımı (Normal, Laplace ve Gamma) temel alan bir simülasyon çalışması tasarladık ve CF algoritmasının performansını R yazılımındaki varsayılan rastgele sayı oluşturma araçlarıyla karşılaştırdık. **Bulgular:** Üç simülasyon çalışmasında da çok küçük farklılıklarla iki yöntem arasında benzer sonuçlar gözlendi. **Sonuç:** Göz önünde bulundurduğumuz üç dağılım için simülasyon sonuçları, CF algoritmasının performansının varsayılan algoritmanın performansıyla iyi hizalandığını göstermektedir. Rastgele bir değişkenin PDF'si verildiğinde doğrudan bir algoritmadan yararlanmak daha iyidir. Bununla birlikte, CF tabanlı yöntem, PDF'den örnek almanın zor olduğu durumlar için sağlam bir alternatif sunabilir. Bulgularımız, zor PDF ve kolay CF senaryolarında veri oluşturma konusunda daha fazla araştırma yapmak için umut verici bir temel sağlamaktadır. Kısacası, CF verildiğinde veya kapalı formda türetilebildiğinde, CF yöntemi veri üretmek için varsayılan yöntemlere mükemmel bir alternatiftir.

Anahtar kelimeler: Simülasyon; rastgele sayı üretimi; karakteristik fonksiyon; olasılık yoğunluk fonksiyonu; algoritma

In statistical analysis, the simulation paradigm is often utilized to evaluate the methods of interest. A key starting step in nearly all simulation studies is RNG. Stochastic simulation is an indispensable part and major focus of scientific inquiry. Model building, estimation, and testing typically require verification through simulation to assess the validity, reliability, and plausibility of inferential techniques,

Correspondence: Mehmet Onur KAYA
Department of Biostatistics and Medical Informatics, Fırat University Faculty of Medicine, Elazığ, TURKEY/TÜRKİYE
E-mail: mokaya@firat.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 08 Sep 2020 ***Received in revised form:*** 09 Nov 2020 ***Accepted:*** 12 Nov 2020 ***Available online:*** 21 Dec 2020

2146-8877 / Copyright © 2020 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

to evaluate how well the implemented models capture the specified true population values, and how reasonably these models respond to departures from underlying assumptions, among other things. Describing a real notion by creating mirror images and imperfect proxies of the perceived underlying truth in a repeated manner allows researchers to study the performance of their methods via simulated data replicates that mimic the real data characteristics of interest in any given setting. Accuracy and precision measures regarding the parameters under consideration signal if the procedure works properly; and may suggest remedial action to minimize the discrepancies between expectation and reality. Simulations have been commonly employed in a wide range of research fields including the physical, medical, social, and managerial sciences. A central aspect of every simulation study is the quantification of the model components and parameters that jointly define a scientific process. When this quantification cannot be performed through deterministic tools, researchers resort to random number generation (RNG) as a starting point in finding simulation-based answers to their questions. One of the most typical methods to conduct RNG is called inversion sampling, which is the default implementation tool in many software packages: Given a continuous distribution function F for random variable X, where U is a uniform $(0,1)$ random variable, then $F^{-1}(U)$ has the same distribution as X. However, the drawback of this method is that the method needs the exact form of the distribution function $F(x)$, which may be unknown or hard to derive. An easier way to perform RNG is through PDF $f(x)$ via approaches such as rejection sampling and importance sampling.

The CF $\phi(t)$ is another way to define the probability distribution. For a univariate random variable X, the relationship between the $f(x)$ and $\phi(t)$ can be expressed by expected value of $\exp(itX)$:

$$\phi(t) = E\left(e^{itX}\right) = \int_x e^{itx} f(x)\, dx \tag{1}$$

where i is the imaginary unit, $i = \sqrt{-1}$, $t \in \mathbb{R}$.

The CF has plenty of decent properties: First, the CF is defined for all real-valued random variables. For example, for the Cauchy distribution, the first order moment does not exist and can be defined in terms of CF, but not using the moment generating function (MGF). In addition, a characteristic function is uniformly continuous on the entire space and is bounded: $|\phi(t)| \leq 1$. Occasionally, there is some convenience to use the CF rather than the PDF. For example, if we have some values of common statistics such as mean, variance, skewness, and kurtosis (first to fourth moments), we can use the CF if the exact PDF is difficult to sample from. Furthermore, if a random variable X has moments up to k-th order, then $\phi(t)$ is k times continuously differentiable on the entire real line. In this case:

$$E(X^k) = i^{-k}\phi^{(k)}(0)$$

In many statistical applications, researchers are given the CF but not the PDF or the distribution function. Devroye[1-3] provided an RNG method where X is a univariate continuous random variable whose CF and its second derivative, $\phi(t)$ and $\phi''(t)$ exist, and $\int|\phi(t)|$ and $\int\left|\phi''(t)\right|$ are finite. In this paper, we first review the Devroye[1-3] RNG method, then provide the simulation examples of RNG of normal, gamma, and Laplace distributions to see whether this CF-based RNG method can perform same as the built-in RNG methods in R[4].

MATERIAL AND METHODS

MATHEMATICAL BACKGROUND

Regarding Devroye's algorithm[1-3], denote X is a univariate continuous random variable with probability function F, density f and CF $\phi(t)$. If $\phi(t)$ is **(a)** twice differentiable with t and **(b)** $|\phi(t)|$, $|\phi'(t)|$ and $|\phi''(t)|$ have finite integral values, then the PDF $f(x)$ satisfies:

$$f(x) \leq \min\left(\frac{1}{2\pi}\int|\phi(t)|, \frac{1}{2\pi x^2}\int|\phi''(t)|\right) \quad (2)$$

We can prove the inequality in (2) from Devroye[3]: Since ϕ is absolutely integrable, f can be computed by the Fourier inversed transformation of (1):

$$f(x) = \frac{1}{2\pi}\int \phi(t)e^{-itx}dt$$

Furthermore, ϕ' and ϕ'' are also absolutely integrable, then similarly:

$$f(x) = \frac{1}{2\pi ix}\int \phi'(t)e^{-itx}dt = -\frac{1}{2\pi x^2}\int \phi''(t)e^{-itx}dt$$

By applying an appropriate theorem in Loeve (1963, p.199)[5], (2) can be proved. The inequality in (2) implies why f can globally be tucked under a bounded integrable curve[3]. Then, the general acceptance-rejection algorithm for characteristic functions satisfying the above conditions is as follows:

Algorithm

1. Set $a = \frac{1}{2\pi}\int|\phi|$, $b = \frac{1}{2\pi}\int|\phi''|$
2. Generate two independently identically distributed (iid) uniform $(-1,1)$ random variates U, V.
3. If $U < 0$, then $X = \sqrt{\frac{b}{a}}V$, $T = a|U|$; else $X = \sqrt{\frac{b}{a}}\frac{1}{V}$, $T = \frac{|U|b}{X^2}$
4. If $T \leq f(X)$, then return X. Otherwise, go back to step 2.

The expected number of iterations is $\frac{2}{\pi}\sqrt{\int|\phi|\int|\phi''|}$, and this algorithm requires the exact form of PDF f. When f is unknown, we can apply the Taylor expansion to the unknown PDF to summation of small parts of integral of CF[3]. The details will not be covered in this work; we only discuss the scenarios where f is accessible to us.

SOME KNOWN DENSITIES

Suppose X is a univariate random variable with known density function $f(x)$ and CF $\phi(t)$, where $\phi(t)$ satisfies the conditions **(a)** and **(b)** mentioned above. Then,

$$\phi(t) = E\left(e^{itX}\right) \quad (1)$$

$$\phi''(t) = E\left(-X^2e^{itX}\right) \quad (3)$$

Where Equation (1) is the definition of CF and Equation (3) provides an optional way to calculate the second derivative of the characteristics function. Suppose a and c are two scalar quantities where $a \neq 0$, then for any scaled random variable aX, we have

$$\phi(at) = E\left(e^{itaX}\right) \tag{4}$$

$$a^2\phi''(at) = E\left(e^{itaX}\right) \tag{5}$$

Equations (4) and (5) show the integral of $\int|\phi|$ and $\int|\phi''|$ will have closed forms when multiplied by a and a^2. For shifted random variable $X + c$, we have

$$e^{itc}\phi(t) = E\left(e^{it(X+c)}\right) \tag{6}$$

Here, Equation (6) shows $\int|\phi(t)|$ is also translation-invariant. However,

$$\int|\phi''| = \int\left|E\left(-(X-c)^2e^{itX}\right)\right|$$

is not invariant here and need further calculation. Therefore, if we try to find a general way to apply the algorithm, it seems that a scaled random variable is easier to deal with than a shifted random variable, and a symmetric density is easier to deal with than an asymmetric one, which will be discussed later.

In the following section, we introduce algorithms for specific densities that conform with the conditions **(a)** and **(b)** above. For simple illustration, we use two symmetric distributions, Normal and Laplace, and one asymmetric distribution, Gamma. All three distributions have relatively simple forms of CF.

Normal Distribution

For the Normal distribution, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(x-\mu)^2}$ and $\phi(t) = e^{it\mu-\frac{1}{2}\sigma^2t^2}$, where μ is the mean parameter and σ^2 is the variance parameter. Then, the second derivative of this CF is $\phi''(t) = (\sigma^2t^2 - 2i\mu\sigma^2t - \sigma^2 - \mu^2)e^{i\mu t-\frac{\sigma^2t^2}{2}}$. It is easy to calculate that the absolute integral $\int|\phi| = \frac{\sqrt{2\pi}}{\sigma}$, but it is hard to find $\int|\phi''|$ given the properties of variant translation above. However, the Normal distribution has nice properties: if $X \sim N(\mu,\sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$. Denote random variable $Y = \frac{X-\mu}{\sigma}$. Therefore, $\int|\phi_Y| = \sqrt{2\pi}$, $\int|\phi_Y''| = \frac{4}{\sqrt{e}}$. Then, we can apply Devroye's algorithm[1-3] to generate Y, then do a random variable transformation on X. The detailed algorithm is shown below:

Algorithm for Normal distribution $N(\mu,\sigma^2)$

1. Set $a = \frac{1}{\sqrt{2\pi}}$, $b = \frac{2}{\pi\sqrt{e}}$

2. Generate two independently identically distributed (iid) uniform $(-1,1)$ random variates U, V.

3. If $U < 0$, then $Y = \sqrt{\frac{b}{a}}V$, $T = a|U|$; else $Y = \sqrt{\frac{b}{a}}\frac{1}{V}$, $T = \frac{|U|b}{Y^2}$

4. If $T \leq \frac{1}{2\pi}e^{-\frac{1}{2}Y^2}$, then return $X = \sigma Y + \mu$. Otherwise, go back to step 2.

Laplace Distribution

The Laplace distribution, also called double-exponential distribution, is another symmetric distribution that satisfies the conditions **(a)** and **(b)** with

$f(x) = \frac{1}{2\lambda}\exp\left(-\frac{|x-\mu|}{\lambda}\right)$ and $\phi(t) = \frac{\exp(\mu it)}{1+\lambda^2t^2}$.

It is difficult to compute $\int|\phi''|$ when the random variable X is scaled and shifted. We can conduct a similar operation as in the Normal distribution, using the "standard type" Laplace distribution by setting $Y = X - \mu$, then $Y \sim Laplace(0, \lambda)$. Comparing it with the Normal distribution, this transformation does not require the standard scale. It is easy to find that $\int|\phi| = \frac{\pi}{\lambda}$, $\int|\phi''| = \frac{\sqrt{27}}{2}\lambda$. Then, the algorithm becomes:

Algorithm for Laplace distribution $Laplace(\mu, \lambda)$

1. Set $a = \frac{1}{2\lambda}$, $b = \frac{\sqrt{27}\lambda}{4\pi}$;
2. Generate two independently identically distributed (iid) uniform $(-1,1)$ random variates U, V.
3. If $U < 0$, then $Y = \sqrt{\frac{b}{a}}V$, $T = a|U|$; else $Y = \sqrt{\frac{b}{a}}\frac{1}{V}$, $T = \frac{|U|b}{Y^2}$
4. If $T \leq \frac{1}{2\lambda}\exp\left(-\frac{|y|}{\lambda}\right)$, then return $X = Y + \mu$. Otherwise, go back to step 2.

Gamma Distribution

Unlike the Normal and Laplace distributions, the Gamma distribution is an asymmetric distribution that does not support the properties of "Standardized distribution". Furthermore, it should be noted that not all Gamma distributions meet the conditions **(a)** and **(b)**. For example, when the shape parameter $\alpha \leq 1$. The absolute integrals of ϕ and ϕ'' are divergent. However, when we fix the shape parameter at beginning, then the rate parameter can be vary. For example, we set $\alpha = 2$ for Gamma distribution, and the rate parameter β can be any value where $\beta > 0$. Then we can have $\int|\phi| = \pi\beta$ and $\int|\phi''| = \frac{3\pi}{\beta}$. Then, the RNG algorithm is as follows:

Algorithm for Gamma distribution $Gamma(\alpha = 2, \beta)$

1. Set $a = \frac{\beta}{2}$, $b = \frac{3}{2\beta}$
2. Generate two independently identically distributed (iid) uniform $(-1,1)$ random variates U, V.
3. If $U < 0$, then $X = \sqrt{\frac{b}{a}}V$, $T = a|U|$; else $X = \sqrt{\frac{b}{a}}\frac{1}{V}$, $T = \frac{|U|b}{X^2}$
4. If $T \leq f(X)$, then return X. Otherwise, go back to step 2.

Note this algorithm is valid for shape parameter $\alpha = 2$. We can also implement this algorithm with any shape parameter $\alpha > 1$ by changing the values of a, b in step 1.

SIMULATION

Random Number Generation Design

In the following part, we introduce our simulation setting for three specific distributions: Normal, Laplace, and Gamma. We set three levels of sample sizes: 30, 100, 1000, and 5000. We specified the default parameters as $N(\mu = 2, \sigma^2 = 9)$, $Laplace(\mu = 5, \lambda = 3)$ and $Gamma(\alpha = 2, \beta = 2)$. We employed the method of moments to calculate the parameter estimates. We repeated the algorithm 1000 times. In summary, we had $3 \times 4 = 12$ scenarios (three types of distributions and four different sample sizes) each of which had 1000 replications.

Evaluation Criteria

The evaluation of the Devroye's algorithm[1-3] was based on the following quantities that are frequently regarded as benchmark accuracy and precision measures:

Standardized bias (SB). The relative magnitude of the raw bias to the overall uncertainty in the system. If the parameter of interest is θ, the standardized bias is $100 \times \frac{(E(\hat{\theta})-\theta)}{SE(\hat{\theta})}$, where SE stands for standard error. If the standardized bias exceeds $40-50\%$ in positive or negative direction, then the bias begins to have a noticeable adverse impact on efficiency, coverage, and error rates.[6]

Percentage bias (PB). The relative magnitude of the raw bias to the true value of the parameter, $100 \times \left[\left(E\left(\hat{\theta}\right)-\theta\right)/\theta\right]$. In our subject opinion, the reasonable upper limit for the PB is 5%.[7]

Root-mean-square error (RMSE). An integrated measure of bias and variance. It is arguably the best criterion for evaluating $\hat{\theta}$ in terms of combined accuracy and precision. RMSE($\hat{\theta}$) is defined as $\sqrt{E_\theta\left[\left(\hat{\theta}-\theta\right)^2\right]}$.

In a nutshell, SB and PB are two accuracy measures. RMSE are two integrated measures of some combination of accuracy and precision. We used R (version 4.0.2)[4] to perform the simulation, and we compared the results with the internal RNG functions $rnorm()$ and $rgamma()$. For Laplace distribution, we compared our method with $rlaplace()$ from R package **rmutil**[8]. We used these functions as a control mechanism to find out whether the CF-based algorithm deliver a comparable performance to the default ones. Furthermore, to make speed comparisons, we stored the run times.

FINANCIAL DATA MODELED BY VARIANCE-GAMMA DISTRIBUTION

In the following example, we performed Devroye's algorithm based on a real-world financial data case to show the scenario when PDF is very difficult to sample from, but CF is relatively straightforward. The data of Standard & Poor's 500 Index, January 1977-December 1981 with sample size $n = 1262$ include an univariate variable X_t denotes the log returns at unit time interval[9], which follows variance-gamma distribution with PDF:

$$f_X(x) = \frac{2\exp\left(\frac{\theta(x-c)}{\sigma^2}\right)}{\sigma\sqrt{2\pi}\nu^{\frac{1}{\nu}}\Gamma\left(\frac{1}{\nu}\right)}\left(\frac{|x-c|}{\sqrt{\frac{2\theta^2}{\nu}+\sigma^2}}\right)^{\frac{1}{\nu}-\frac{1}{2}} K_{\frac{1}{\nu}-\frac{1}{2}}\left(\frac{|x-c|\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}{\sigma^2}\right) \quad (7)$$

Where there are four parameters in this model: c (location parameter), θ (asymmetry parameter), σ (variance parameter) and $\nu > 0$. $K_{\frac{1}{\nu}-\frac{1}{2}}$ denotes a modified Bessel function of the third kind[10].

The characteristic function of X_t is

$$\phi_X(u) = \exp(\mathrm{i}cu)\left(1-\mathrm{i}\theta\nu u+\frac{\sigma^2\nu u^2}{2}\right)^{-\frac{1}{\nu}} \quad (8)$$

In the following example, we generate variance-gamma distribution data of size 1262 based on 1000 replications. The true parameter specification of this variance-gamma distribution is based on the Method of Moments results that are as follows:

$$\sigma^2 = 6.447 \times 10^{-5}$$
$$\hat{\nu} = 0.4220$$
$$\hat{\theta} = -1.510 \times 10^{-4}$$
$$\hat{c} = 2.585 \times 10^{-4}$$

Similarly, our performance is based on the comparisons of the first four moments specified with the empirical results from Devroye's algorithm. The density of variance-gamma distribution is calculated by $dvg()$ function from the R package **VarianceGamma**.[10]

RESULTS

SIMULATION RESULTS

The results are shown in Table 1. In the method column, CF stands for Devroye's characteristics function algorithm, and DE stands for the default internal RNG method for respective distributions in R software. In Table 1, PI means parameters of interest, TV stands for true value of parameters, and AE is the average estimate among 1000 replications. AE in the small sample scenario seems more biased than the larger sample for both CF and DE, which is expected; as the sample size increases, both methods yield more accurate results. Both CF and DE result in similar outputs when the specified sample size is same.

For both methods, we can detect decreasing RMSE when the sample size gets larger. The SB and PB vary without a trend as the sample size changes, but they are always stable in a sensible range. Furthermore, the two symmetric distribution did not show any large value of PB and SB, but for the Gamma distribution, which is asymmetric, the SB is larger than the one in two symmetric distributions when sample size is small ($N = 30, 100$), which might be due to the specific properties of Gamma distribution itself.

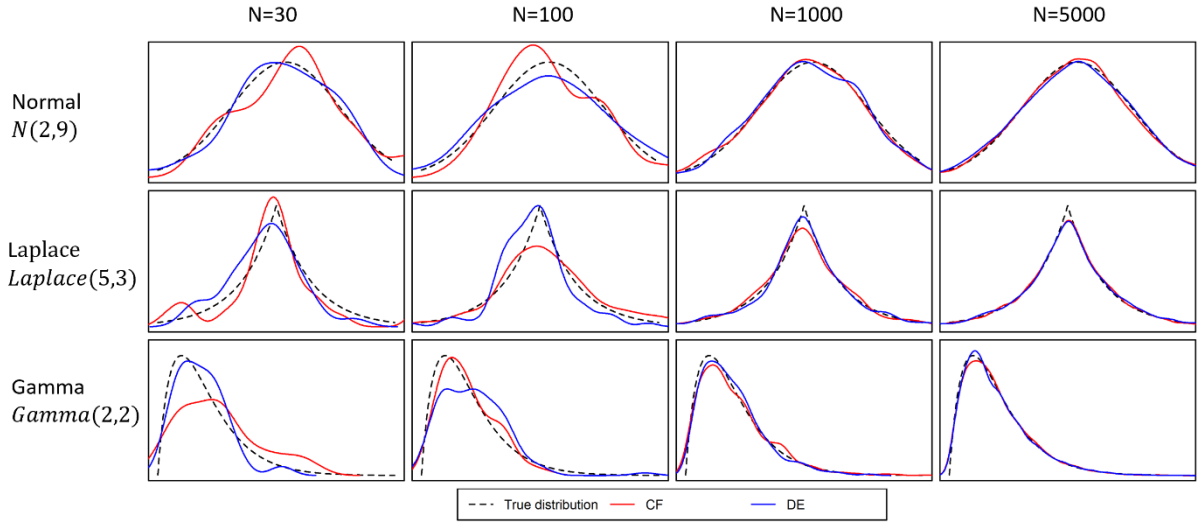

FIGURE 1: Density graphs of Normal, Laplace and Gamma distributions that comes out of two RNG methods where the true density is superimposed using three different sample sizes

In Figure 1, we present the densities for one set of randomly drawn samples by two methods under consideration in addition to the true theoretical density. The black, red, and blue lines line represent the true density and densities that ensue after the implementation of the CF and DE methods, respectively. As can be seen from Figure 1, the CF and DE methods lead to very similar densities, differences are negligibly small across all three sample sizes. The departures from the truth become indiscernible with increasing sample sizes for both methods, as one would expect. It is important to note that had we reported the average trajectories across all replications, three curves would have been almost identical.

TABLE 1: Simulation results for Normal, Laplace and Gamma distributions.

Distribution	Method	PI	n	TV	AE	PB	SB	RMSE
Normal	CF	μ	30	2	2.021	1.08	3.90	0.5550
	DE	μ	30	2	2.001	0.05	0.17	0.5306
	CF	μ	100	2	2.004	0.24	1.61	0.2960
	DE	μ	100	2	1.990	0.52	3.42	0.3044
	CF	μ	1000	2	1.998	0.08	1.62	0.0959
	DE	μ	1000	2	2.000	0.01	0.28	0.0915
	CF	μ	5000	2	1.998	0.09	4.90	0.0407
	DE	μ	5000	2	1.999	0.06	2.71	0.0424
	CF	σ	30	3	2.994	0.19	1.43	0.4054
	DE	σ	30	3	2.972	0.95	7.15	0.3975
	CF	σ	100	3	2.996	0.12	1.66	0.2210
	DE	σ	100	3	2.994	0.21	2.95	0.2121
	CF	σ	1000	3	2.998	0.06	2.69	0.0645
	DE	σ	1000	3	3.001	0.02	0.75	0.0689
	CF	σ	5000	3	3.000	0.01	0.65	0.0310
	DE	σ	5000	3	2.998	0.05	5.02	0.0311
Distribution	**Method**	**PI**	**n**	**TV**	**AE**	**PB**	**SB**	**RMSE**
Laplace	CF	μ	30	5	4.983	0.35	2.82	0.6154
	DE	μ	30	5	4.993	0.14	1.11	0.6140
	CF	μ	100	5	4.988	0.24	3.92	0.3034
	DE	μ	100	5	5.004	0.07	1.09	0.3224
	CF	μ	1000	5	5.000	0.00	0.04	0.0965
	DE	μ	1000	5	5.000	0.00	0.18	0.0938
	CF	μ	5000	5	5.001	0.03	3.16	0.0431
	DE	μ	5000	5	5.002	0.03	3.53	0.0432
	CF	λ	30	3	2.989	0.36	2.01	0.5437
	DE	λ	30	3	2.958	1.40	7.84	0.5357
	CF	λ	100	3	2.989	0.38	3.87	0.2943
	DE	λ	100	3	2.982	0.60	5.98	0.2993
	CF	λ	1000	3	2.996	0.12	3.78	0.0960
	DE	λ	1000	3	2.998	0.08	2.47	0.0930
	CF	λ	5000	3	3.000	0.01	0.53	0.0419
	DE	λ	5000	3	3.000	0.02	1.16	0.0426
Distribution	**Method**	**PI**	**n**	**TV**	**AE**	**PB**	**SB**	**RMSE**
Gamma	CF	α	30	2	2.226	11.32	34.35	0.6969
	DE	α	30	2	2.224	11.19	32.55	0.7227
	CF	α	100	2	2.073	3.63	19.91	0.3715
	DE	α	100	2	2.065	3.26	18.43	0.3599
	CF	α	1000	2	2.004	0.21	3.75	0.1131
	DE	α	1000	2	2.003	0.13	2.35	0.1071
	CF	α	5000	2	2.001	0.06	2.55	0.0499
	DE	α	5000	2	2.002	0.11	4.64	0.0490
	CF	β	30	2	2.243	12.17	34.18	0.7522
	DE	β	30	2	2.260	13.02	33.85	0.8121
	CF	β	100	2	2.075	3.77	18.88	0.4064
	DE	β	100	2	2.077	3.86	19.88	0.3954
	CF	β	1000	2	2.006	0.32	5.14	0.1231
	DE	β	1000	2	2.006	0.31	5.42	0.1136
	CF	β	5000	2	2.002	0.09	3.39	0.0540
	DE	β	5000	2	2.003	0.15	5.65	0.0531

We also compare the calculation time of those two algorithms. When we generated only one sample, the CF method had nearly the same speed as the default method. However, with 1000 replications where $N = 5000$, the CF method needed nearly one minute, but the DE method used two seconds. The difference of the algorithmic speeds makes sense since the CF algorithm includes rejection steps; it takes longer compared to the direct calculation in the default algorithm. In summary, the price to pay for extra generality of being able to employ the CF algorithm is some reduction in speed. On a more positive note, similar accuracy signals the efficacy of the CF algorithm. The reported results are reproducible, and the R code[4] used in this manuscript is publicly accessible at http://demirtas.people.uic.edu/cf_paper_code.R

FINANCIAL DATA RESULTS

The results are shown in Table 2. The skewness and kurtosis are calculated by R function $skewness()$ and $kurtosis()$ from the package **moments**[11]. The formulas for the PB, SB, and RMSE are described in Section 2.3.2. In this table, all four moments of the generated data are close to the theoretical ones, generally indicating the Devroye's algorithm can properly capture the properties of the original sample. Although the PB/SB for the estimates of skewness and kurtosis are not ideal, the mean and variance estimates seem to be very close to the specified quantities.

TABLE 2. Financial data example: Specified, empirical moments and the percent difference between specified and empirical moments.

Moments	Specified	Empirical	PB	SB	RMSE
Mean(μ)	1.052×10^{-4}	1.061×10^{-4}	0.91	0.41	2.3348×10^{-4}
Variance(σ^2)	6.4474×10^{-5}	6.4489×10^{-5}	0.02	0.45	3.1506×10^{-6}
Skewness(β)	-0.023813	-0.021965	6.79	1.21	0.1339
Kurtosis(κ)	4.2661	4.2437	0.53	5.14	0.4362

4. DISCUSSION

The simulation results for the three distributions we considered suggest that the performance of the CF algorithm aligns well with that of the default algorithm. It is better to harness a direct algorithm when the PDF of a random variable is given. However, the CF-based method can offer a solid alternative for situations where the PDF is hard to derive. Our findings provide a promising basis for conducting further research on generating data in complex PDF (which is hard to sample from) and relatively simple CF scenarios, as in the Standard & Poor's data example shown before.

It should also be mentioned that the CF algorithm hinges upon existence of the second derivative ϕ'' and two convergent absolute integrals $\int |\phi|$ and $\int |\phi''|$. Therefore, this method may not be applied to some distributions. For example, we found that the Gamma distribution when the shape parameter $\alpha \leq 1$ did not have convergent integral of second derivative $\int |\phi''|$, just like some types of Beta distributions. Another problem of this CF algorithm is that the absolute integral may be hard to solve.

5. CONCLUSION

In summary, we demonstrated that the CF-based algorithm can perform as well as the internal (built-in) methods in R. Although the algorithm has some problems such as slower speed or difficulties in integration, it is still an excellent alternative to the default methods.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Hakan Demirtaş, Xiaohan Mei, Xinhe Wang;* ***Design:*** *Hakan Demirtaş, Mehmet Onur Kaya, Xiaohan Mei, Xinhe Wang;* ***Control/Supervision:*** *Hakan Demirtaş, Mehmet Onur Kaya, Xiaohan Mei, Xinhe Wang;* ***Data Collection and/or Processing:*** *Xiaohan Mei, Xinhe Wang;* ***Analysis and/or Interpretation:*** *Xiaohan Mei, Xinhe Wang, Hakan Demirtaş;* ***Literature Review:*** *Xiaohan Mei, Xinhe Wang;* ***Writing the Article:*** *Hakan Demirtaş, Xiaohan Mei, Xinhe Wang,* ***Critical Review:*** *Hakan Demirtaş, Mehmet Onur Kaya, Xiaohan Mei, Xinhe Wang;* ***References and Fundings:*** *Xiaohan Mei, Xinhe Wang, Hakan Demirtaş.*

REFERENCES

1. Devroye L. On the computer generation of random variables with a given characteristic function. Computers & Mathematics with Applications. 1981;7(6):547-52.[Crossref]
2. Devroye L. An automatic method for generating random variates with a given characteristic function. SIAM J Appl Math. 1986;46(4):698-719.[Crossref]
3. Devroye L. Nonuniform random variate generation. Handbooks in Operations Research and Management Science. 2006;13:83-121.[Crossref]
4. R Core Team. R: A language and environment for statistical computing. Foundation for Statistical Computing Vienna, Austria; 2020.[Link]
5. Loève M. Probability Theory. Vol. 1963. Princeton: Van Nostrand; 1963.
6. Demirtas H. Simulation driven inferences for multiply imputed longitudinal datasets. Statistica Neerlandica. 2004;58(4):466-82.[Crossref]
7. Demirtas H, Freels SA, Yucel RM. Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. J Stat Comput Simul. 2008;78(1):69-84.[Crossref]
8. Swihart B, Lindsey J. rmutil: Utilities for Nonlinear Regression and Repeated Measurements Models. R package version 1.1.4. 2020.[Link]
9. Seneta E. Fitting the variance-gamma model to financial data. Journal of Applied Probability. 2004;41:177-87.[Crossref]
10. Scott D, Dong CY. VarianceGamma: the variance gamma distribution. R package version 0.4-0. 2018.[Link]
11. Komsta L, Novomestky F. Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. 2015.[Link]