

PAPER DETAILS

TITLE: Non-Uniform Random Number Generation from Arbitrary Bivariate Distribution in Polygonal Area

AUTHORS: Bugra Kaan TIRYAKI,Orhan KESEMEN

PAGES: 443-457

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/528915>

Non-Uniform Random Number Generation from Arbitrary Bivariate Distribution in Polygonal Area

Orhan KESEMEN¹, Buğra Kaan TİRYAKİ^{*1}

¹Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, 61080, Trabzon

(Alınış / Received: 10.07.2017, Kabul / Accepted: 10.04.2018, Online Yayınlanma / Published Online: 15.05.2018)

Keywords

Bivariate distribution
function,
Piecewise probability density
function,
Non-uniform random
number generation,
Rejection method,
Inversion method

Abstract: Bivariate non-uniform random numbers are usually generated in a rectangular area. However, this is generally not useful in practice because the arbitrary area in real-life is not always a rectangular area. Therefore, the arbitrary area in real-life can be defined as a polygonal approach. Non-uniform random numbers are generated from an arbitrary bivariate distribution within a polygonal area by using the rejection and the inversion methods. Three examples are given for non-uniform random number generation from an arbitrary bivariate distribution function in polygonal areas. In these examples, the non-uniform random number generation is discussed in the triangular area, the Korea mainland and the Australia mainland. The non-uniform random numbers are generated in these areas from the arbitrary probability density function. The observed frequency values are calculated with using both methods in the simulation study and the generated random numbers are tested with the chi-square goodness of fit test to determine whether or not they come from the given distribution. Also, both methods are compared each other with a simulation study.

Çokgensel Bölgede Keyfi İki Değişkenli Dağılımdan Düzgün Olmayan Rastgele Sayı Üretimi

Anahtar Kelimeler

İki değişkenli dağılım
fonksiyonu,
Parçalı olasılık yoğunluk
fonksiyonu,
Düzgün olmayan rastgele
sayı üretimi,
Ret-kabul yöntemi,
Ters dönüşüm yöntemi

Özet: İki değişkenli bir uzayda rastgele sayılar genellikle dikdörtgensel bir alan içerisinde üretilmektedir. Ancak uygulamalarda her zaman dikdörtgensel bir alan olamayacağı için keyfi bir alan, çokgensel bir yaklaşım kullanılarak tanımlanmaya çalışılmıştır. Çokgensel bir alan içindeki keyfi iki değişkenli dağılımdan düzgün olmayan rastgele sayılar ret ve ters yöntemleri kullanarak üretilmiştir. Çokgensel alanlardaki keyfi iki değişkenli dağılım fonksiyonundan düzgün olmayan rastgele sayı üretimi için üç ayrı örnek verilmiştir. Bu örneklerde, düzgün olmayan rastgele sayılar üçgen alanında, Kore anakarasında ve Avustralya anakarasında üretilmiştir. Düzgün olmayan rastgele sayılar bu alanlarda keyfi olasılık yoğunluk fonksiyonundan üretilmiştir. Gözlenen frekans değerleri, simülasyon çalışmasında her iki yöntem kullanılarak hesaplanmıştır ve üretilen rastgele sayıların verilen dağılımdan gelip gelmediklerini belirlemek için ki-kare uyum iyiliği testi kullanılmıştır. Ayrıca, her iki yöntem bir simülasyon çalışması ile birbirleriyle karşılaştırılmıştır.

1. Introduction

Random number simulations have long been a popular topic in the world of science. Random number simulations have been studied particularly closely within the areas of mathematics and statistics. Random numbers are used in fields such as communication [1], information and control systems [2], signal processing [3], machine learning [4], biostatistics [5], econometrics [6], mathematical

finance [7], biology [8], insurance [9], cryptography [10], simulation [11], computer games [12] and statistical sampling and applications [13]. Therefore, random number generation is quite important.

In the literature, the rejection method or the inversion method from univariate and bivariate probability density functions are the most commonly used non-uniform random number generation algorithms [11, 14, 15, 16]. Particularly, while

* Corresponding author: bugrakaantiryaki@gmail.com

generating non-uniform random numbers from bivariate probability density function, the limit of the probability density functions is defined as rectangular [17, 18, 19, 20]. If the non-uniform random numbers are generated from a bivariate probability density function in an arbitrary area (polygon), then the distribution bounded within the area will be uniform. In this case, the non-uniform random numbers are generated by looking into whether it falls into the area [14, 15, 21, 22, 23]. Furthermore, random numbers are generated over a unit hyperball, hypersphere and hyperellipsoid. But similar to the above, they have a uniform distribution [15]. Contrary to the literature, the non-uniform random numbers are generated from an arbitrary bivariate distribution function in an arbitrary polygonal area.

The process of generating random numbers from the given limits of the bivariate probability density function, which is defined as a rectangular range, can be easily adopted to work with both of the non-uniform random number generation methods. However, in many simulations, it is necessary to use a bivariate probability density function to model the problem within two-dimensional spaces. The definition limits of this probability density function can be an arbitrary non-uniform area rather than a rectangular area [24, 25]. In this case, it is not possible, through conventional methods, to perform conversions required for the non-uniform random number generation. The motivation of this study comes from the fact that the generation of the non-uniform random numbers from an arbitrary bivariate distribution within a polygonal area by using the rejection and the inversion methods is not improved in the literature. Because real-life problems are not always limited to uniform rectangular areas.

The bivariate probability density function, which is bounded by arbitrary non-uniform limits, has many fields of application in real-life problems. These problems include the amount pollution or the crime rate in a city, the distribution of earthquake frequency within a country, the dispersion density of insects in a field, the traffic congestion in a specific region, the documented locations and the frequency of an epidemic disease in a country. Also, a region of interest may have many partial density functions in the rectangular area. Many physically or politically divided cities may be given as examples for this situation (e.g. Belfast, Beirut, Jerusalem, Mostar, and Nicosia). However, the data in the politically divided cities start to change over time. In this case, it may not be possible to evaluate the entire city as a region. In each of these examples, an arbitrary polygonal area may be needed in order to generate non-uniform random numbers from the probability density function.

In this study, the generation of non-uniform random numbers is performed from the bivariate probability

density function with the rejection and the inversion methods. The inversion and the rejection methods for univariate distribution functions are reminded in Section 2.1. The non-uniform bivariate random number generation in a rectangular area and its algorithms are given in Section 2.2.1. In Section 2.2.2, analytical and numerical approaches are presented for the non-uniform random number generation in a triangular area which describes 3-dimensional structures well. The algorithm using the triangular approach mentioned in Section 2.2.2 for the non-uniform random number generation in a polygonal area is given in Section 2.2.3. The performance of these algorithms is tested by simulating for real and artificial data sets in Section 3.

2. Material and Method

2.1. Generating univariate non-uniform random numbers

In all programming languages, random number generator functions have a deterministic algorithm which generates random numbers from uniform distribution in the range $[0, 1]$. Random numbers are generated with a deterministic algorithm since such random numbers are called pseudo random numbers [11, 26]. The rejection and the inversion methods are the most common two approaches which are used to generate non-uniform random numbers.

2.1.1. Rejection method

The rejection method is useful in generating a non-uniform random number in the range $[a, b]$ from an arbitrary distribution. In order to perform this method, the f_{max} value must first be determined as the largest value of the distribution of probability density function $f(x)$ in the range $[a, b]$. The algorithm of the rejection method is given by Rubinstein [15].

Algorithm 1. The univariate non-uniform random number generation algorithm according to the rejection method.

- | | |
|---------|---|
| Step 1. | Determine the probability density function $f(x)$, to produce random number. |
| Step 2. | Generate two random numbers from the uniform distribution $\xi_1 \sim U(0,1)$ and $\xi_2 \sim U(0,1)$. |
| Step 3. | Calculate value $X = a + \xi_1(b - a)$. |
| Step 4. | Calculate condition value $y = \xi_2 f_{max}$. |
| Step 5. | If $y > f(X)$, reject (X) and go to step 2, continue running algorithm steadily, else accept the value (X) . |
-

2.1.2. Inversion method

The inversion method is the most frequently used method to generate a non-uniform random number as it is an analytically approach [26]. The application

of this method is only valid if the inverse of distribution function is available. In cases when the inverse of the distribution function cannot be found, numerical methods [27, 28] are used. $F(x)$ is the desired distribution function used in generating random numbers and Equation (1) is used to generate random number by employing the inversion method.

$$\xi = F(X) \quad (1)$$

$\xi \sim U(0,1)$ indicates a random number is generated from the uniform distribution on the interval $[0,1]$. If the inverse of the distribution function is solved in Equation (1).

$$X = F^{-1}(\xi) \quad (2)$$

Equation (2) is obtained [15]. The variable X shows a random number from the distribution $F(x)$. In this way, all random numbers which are generated from the uniform distribution can be converted to the desired distribution.

2.2. Generating bivariate non-uniform random numbers

Bivariate random numbers are shown in the form of (X, Y) . Generally, these are referred to as two-dimensional random points rather than bivariate random numbers. When generating univariate random numbers within a limited range, only two limit values are used. However, generating random numbers becomes more complicated in a limited area due to an infinite number of limitations of different areas. A general definition is required to overcome this complexity. This definition is the main subject of this study. First, generating a bivariate random number in simple geometric areas will be discussed. In the subsequent sections, the method of generating bivariate random numbers in complex geometric areas will be described.

2.2.1. Generating non-uniform random numbers in a rectangular area

A rectangular area is the most ideal area to generate bivariate random numbers. The joint probability density function in a rectangular area is defined in the shape of $f(x, y): [a, b] \times [c, d] \rightarrow \mathbb{R}^2$. Each random number is generated with two different approaches in this plane. These are listed as

- In the case of being independent variables of the joint distribution function, bivariate random numbers are generated separately random number, according to the marginal distribution,
- In the case of being dependent variables of the joint distribution function, random number generation using the marginal and conditional distribution functions, respectively.

Bivariate non-uniform random numbers are generated according to their marginal distributions. On the other hand, two different methods are generally used for dependent variables. These are the rejection and the inversion methods.

The rejection method is the first method which is referenced for generating a non-uniform random number from an arbitrary joint distribution function in a rectangular area. The non-uniform random number generation algorithm in a rectangular area is given below in Algorithm 2 [15].

Algorithm 2. The random number generation algorithm according to the rejection method in a rectangular area.

- | | |
|---------|--|
| Step 1. | Determine the bivariate probability density function $f_{XY}(x, y)$ to generate a random number. |
| Step 2. | Generate two random numbers from the uniform distribution in the shape of $X \sim U(a, b)$ and $Y \sim U(c, d)$. |
| Step 3. | Generate a random number from the uniform distribution in the shape of $z \sim U(0, f_{max})$. f_{max} value is the largest value of the $f_{XY}(x, y)$ function in the region $[a, b] \times [c, d]$. |
| Step 4. | If $z > f_{XY}(X, Y)$, reject the generated random number (X, Y) and continue processing and go to Step 2, otherwise accept the generated random number (X, Y) . |
-

Algorithm 3 is used for generating a random number from the bivariate distribution function by using the inversion method in a rectangular area [15].

Algorithm 3. The random number generation algorithm from the bivariate distribution function according to the inversion method in a rectangular area.

- | | |
|---------|---|
| Step 1. | Determine the bivariate probability density function $f_{XY}(x, y)$ to generate a random number. |
| Step 2. | Generate two random numbers from the uniform distribution in the shape of $\xi_1 \sim U(0,1)$ and $\xi_2 \sim U(0,1)$. |
| Step 3. | Substitute ξ_1 value in the following inverse of the marginal distribution function and generate the random number component of X . |
| | $X = F_X^{-1}(\xi_1)$ |
| Step 4. | Substitute X and ξ_2 values in the following inverse of the conditional distribution function and generate the random number component of Y . |
| | $Y = F_{Y X}^{-1}(\xi_2 X = x)$ |
-

Thus, the bivariate random number is generated in the form of (X, Y) .

2.2.2. Generating non-uniform random numbers in a triangular area

Generating a random number in the triangular area by using the rejection method will be explained comprehensively in Section 2.2.3. This section will be focused on generating a random number with the inversion method in the triangular area.

Although the calculation of the integration calculus from the bivariate probability density function according a fixed limit on the rectangular area is a relatively simple process, the integration calculus in the triangular area is more complex. To eliminate this complexity, the median of the components $\{x_a, x_b, x_c\}$ of x is found from the corner coordinates of the triangle area of Ω in Figure 1(a). Drawing a parallel line from the y -axis to the median value divides the triangle into two parts. This division makes the calculation of the integral in the triangle area of Ω in (Figure 1(b)) less complicated. The Ω region is divided into two triangles (Ω_L, Ω_R) and the sum of the total integral value can be found separately in Equation (3).

$$\begin{aligned} \iint_{\Omega} f_{XY}(x, y) dx dy \\ = \iint_{\Omega_L} f_{XY}(x, y) dx dy \\ + \iint_{\Omega_R} f_{XY}(x, y) dx dy \end{aligned} \quad (3)$$

The marginal probability density function of the triangle Ω_L is defined in Equation (4).

$$f_{XL}(x) = \int_{\overline{AC}}^{\overline{AB}} f_{XY}(x, y) dy, \quad x \in [x_a, x_b] \quad (4)$$

\overline{AB} shows that the equation of the line segment between A and B points. \overline{AC} shows the equation of the line segment between A and C points. The distribution function of the obtained marginal probability density function is calculated in Equation (5).

$$F_{XL}(x) = \int_{x_a}^x f_{XL}(x) dx \quad (5)$$

The volume of left area (probability value) is obtained in Equation (6).

$$I_L = F_{XL}(x_b) = \int_{x_a}^{x_b} f_{XL}(x) dx \quad (6)$$

Likewise, the marginal probability density function of the triangle Ω_R is defined in Equation (7).

$$f_{XR}(x) = \int_{\overline{AC}}^{\overline{BC}} f_{XY}(x, y) dy, \quad x \in [x_b, x_c] \quad (7)$$

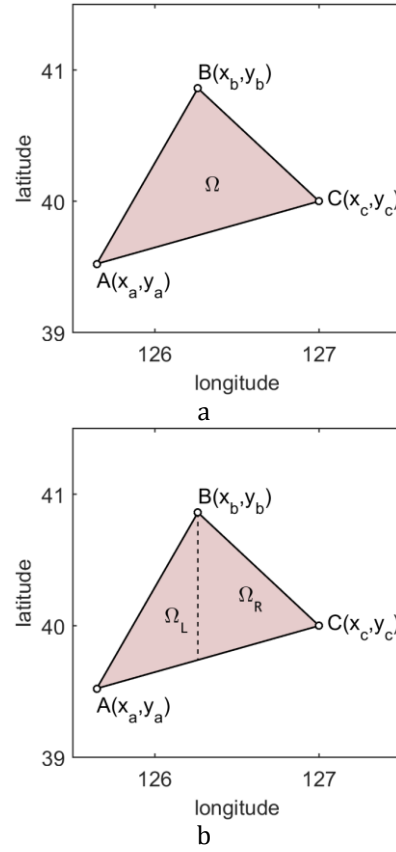


Figure 1. The representation of the triangular region; (a) Ω region; (b) The determination of the triangular region of Ω_L and Ω_R .

\overline{BC} shows the equation of the line segment between B and C points. \overline{AC} shows the equation of the line segment between A and C points. The distribution function from marginal distribution function of the second triangle is obtained in Equation (8).

$$F_{XR}(x) = I_L + \int_{x_b}^x f_{XR}(x) dx \quad (8)$$

The conditional probability density function of Y (given $X = x$ in the Ω_L area) is defined in Equation (9).

$$f_{YL|X}(y|x) = \frac{f_{XY}(x, y)}{f_{XL}(x)} \quad (9)$$

If $X < x_b$, Equation (9) is valid, otherwise the conditional probability density function of Y (given $X = x$ in the Ω_R area) is defined in Equation (10).

$$f_{YR|X}(y|x) = \frac{f_{XY}(x, y)}{f_{XR}(x)} \quad (10)$$

The conditional distribution of Y (given $X = x$ in the Ω_L area) is defined Equation (11).

$$F_{YL|X}(y; y_X|x) = \frac{1}{f_{XL}(x)} \int_{y_X}^y f_{XY}(x, y) dy \quad (11)$$

The conditional distribution function of Y (given $X = x$ in the Ω_R area) is defined in Equation (12).

$$F_{YR|X}(y; y_x | x) = \frac{1}{f_{XR}(x)} \int_{y_x}^y f_{XY}(x, y) dy \quad (12)$$

The analytical equations found here are calculated according the joint distribution function in Algorithm (4a).

Algorithm 4a. The non-uniform random number generation algorithm in the triangular area according to the inversion method (Analytical Part).

- Step 1. Determine the bivariate probability density function $f(x, y)$ to generate a random number and the boundaries of the triangular region (Ω).
- Step 2. Divide the region (Ω) into two triangles according the median apse (x_b).
- Step 3. Compute the marginal probability density functions of each triangle like in Equation (4) and Equation (7).
- Step 4. Calculate the (I_L) value which is the integral of the marginal probability density function from the left triangle in the range $[x_a, x_b]$ like in Equation (6).
- Step 5. Calculate the distribution functions of the marginal probability density functions in Step 3 like in Equation (5) and Equation (8).
- Step 6. Calculate the distribution functions of the conditional probability function of region Ω_L and Ω_R (when $X = x$ given), by utilizing Equation (11) and Equation (12). Determine the inverse functions of these distribution functions.

In Algorithm 4b, the random number generation in the triangular area is performed by utilizing the calculations made analytically in Algorithm 4a.

Algorithm 4b. The non-uniform random number generation algorithm in the triangular area according to the inversion method (Simulation Part).

- Step 1. Generate two random numbers from the uniform distribution $\xi_X \sim U(0,1)$ and $\xi_Y \sim U(0,1)$.
- Step 2. Substitute ξ_X value in the following inverse of the marginal distribution function and generate the component of X of the bivariate random number (Figure 2(a)).

$$X = \begin{cases} F_{XL}^{-1}(\xi_X), & \xi_X \leq I_L \\ F_{XR}^{-1}(\xi_X), & \xi_X > I_L \end{cases} \quad (13)$$

- Step 3. Calculate the component y_x of point $D(X, y_x)$ which is the lower boundary of the region Ω intersected by X , according following equations, (14)

$$y_x = \begin{cases} m_{AB}X + c_{AB}, & m_{AC}x_b + c_{AC} > y_b \text{ AND } \xi_X \leq I_L \\ m_{BC}X + c_{BC}, & m_{AC}x_b + c_{AC} > y_b \text{ AND } \xi_X > I_L \\ m_{AC}X + c_{AC}, & m_{AC}x_b + c_{AC} < y_b \end{cases}$$

where m_* is slope of line (*) and c_* is the constant of line (*).

- Step 4. Substitute ξ_y, y_x and X values in the following inverse of the conditional distribution function and generate the component of Y of the bivariate random number (Figure 2(b)).

$$Y = \begin{cases} F_{YL|X}^{-1}(\xi_y; y_x | x), & \xi_X \leq I_L \\ F_{YR|X}^{-1}(\xi_y; y_x | x), & \xi_X > I_L \end{cases} \quad (15)$$

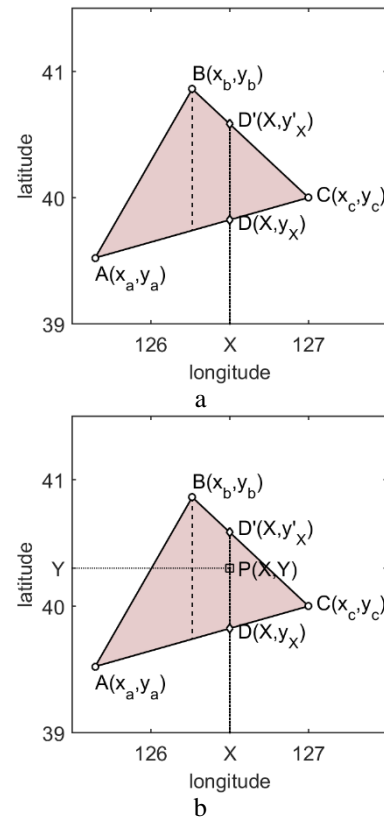


Figure 2. The non-uniform random number generation in the triangular area; (a) the generation of the component X of bivariate random number; (b) the generation of the component Y of the bivariate random number.

The inversion method is performed by calculating the distribution function from a given probability density function and the inverse distribution function from the distribution function. Therefore, the cumulative distribution function of the marginal probability function and the conditional probability density function and their inverse distribution function must be calculated in Section 2.2.2. The distribution function of the selected any probability density function cannot always be calculated. Even when the distribution function is calculated, it is often impossible to have an analytic function of the inverse distribution function. In this case, the rejection

method or a numerical approach is required for solving the problem [28]. A numerical approach is chosen in this section. The bivariate probability density function in the selected triangle must be transformed into a linear function for the numerical approach. First of all, the probability density value at the corner coordinates of the triangle is determined by the following equation.

$$z_i = f_{XY}(x_i, y_i), \quad (i = a, b, c) \quad (16)$$

Thus, the plane equation passing through the points $P_1(x_a, y_a, z_a)$, $P_2(x_b, y_b, z_b)$ and $P_3(x_c, y_c, z_c)$ can be found by the following determinant.

$$\begin{vmatrix} x_b - x_a & y_b - y_a & z_b - z_a \\ x_c - x_a & y_c - y_a & z_c - z_a \\ x - x_a & y - y_a & z - z_a \end{vmatrix} = 0 \quad (17)$$

When the determinant is calculated, the following plane equation is obtained.

$$\hat{f}_{XY}(x, y) = C_1x + C_2y + C_3 \quad (18)$$

The marginal probability density function, the conditional probability function, their distribution and the inverse distribution function always can be found from the obtained plane joint probability density function. Thus, the analytical approach given in Section 2.2.2. can be used easily. A certain amount of error is made with this approach by converting an arbitrary joint probability density function into a planar function. This error can be reduced by dividing the selected triangle into the smaller sub-triangles.

2.2.3. Generating non-uniform random numbers in a polygonal area

Generally, bivariate statistical calculations are performed in a rectangular area. In reality, however, the probability of statistical data being found within a rectangular area is quite low. Thus, statistical calculations must have the potential to be conducted in areas not of a rectangular shape. In this section, the boundaries of Korea are given Figure 3(a). In order to calculate the area of Korea, said area must be defined as a polygonal area. Dominant points of the boundaries are determined in order to change arbitrary area into a polygonal area. These dominant points can be determined manually or automatically by employing using polygonal approximation algorithms [29] (Figure 3(b)).

Whether a point is within the polygon or not must be investigated while generating random numbers with both the rejection and inversion methods. This problem is known as the point problem in the polygons [30, 31]. A perpendicular straight line is drawn by connecting from a randomly selected

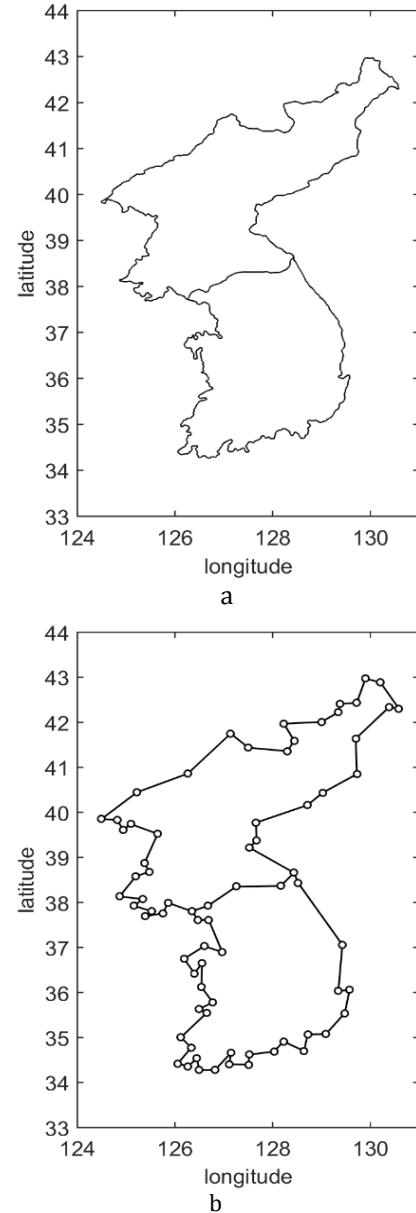


Figure 3. Definition of the polygonal area; (a) An arbitrary piecewise area; (b) The polygonal area obtained by determining the dominant points.

(X, Y) point to line $y = x_{min}$ according the proposed odd-even method for solving this problem. Also, how many points of constituent all lines of the polygon are intercepted by the perpendicular straight line must be determined. If the number of the crossing points is odd, the selected point is determined to be inside of polygon. If the number of crossing points is even, the selected point is determined to be outside of the polygon.

The rejection method is the first method which explains the generation of a non-uniform random number from the arbitrary joint density function in a polygonal area. Algorithm 5 is given to generate a random number from a bivariate distribution function by employing the rejection method within a polygonal area.

Algorithm 5. The non-uniform random number generation algorithm in the polygonal area according to the rejection method.

- Step 1. Determine bivariate probability density function $f_{XY}(x,y)$ to generate a non-uniform random number and the coordinates of the nodal points of the polygonal area (x_k, y_k) .
- Step 2. Generate two random numbers from the uniform distribution $X \sim U(x_{min}, x_{max})$ and $Y \sim U(y_{min}, y_{max})$. These upper and lower boundaries are determined by choosing points which create the polygon nodal points.
- $$\begin{aligned} x_{min} &= \min(x_k) \\ x_{max} &= \max(x_k) \\ y_{min} &= \min(y_k) \\ y_{max} &= \max(y_k) \end{aligned}$$
- Step 3. Examine the generated random number (X,Y) to determine whether is in polygon or not. If the generated number is outside of the polygon, generate a new random number (X,Y) by returning to Step 2.
- Step 4. Generate a random number from the uniform distribution in the shape of $z \sim U(0, f_{max})$. Also, f_{max} value is the largest value of $f_{XY}(x,y)$ function in the region $f_{XY}(x,y) \in \Omega$.
- Step 5. If $z > f_{XY}(X,Y)$, reject the generated random number (X,Y) and continue processing and go to Step 2, otherwise accept the generated random number (X,Y) .

Firstly, the polygonal area needs to be divided into triangles to generate non-uniform random numbers using the inversion method. The Delaunay algorithm is used in this task [32, 33]. After polygonal area is divided into several triangles, the probability value (the probability value in the triangle, P_j) of each triangle (Ω_j) is defined. This is demonstrated in Equation (19) [34].

$$P_j = \iint_{\Omega_j} f(x,y) d_x d_y \quad (19)$$

In order to calculate the integral in a triangular region, as shown in Figure 1(b), the triangle is divided into two parts. The sum of the integral value is calculated separately in Equation (20).

$$P_j = \left| \int_{x_a}^{x_b} \int_{\overline{AC}}^{\overline{AB}} f(x,y) dy dx \right| + \left| \int_{x_b}^{x_c} \int_{\overline{AC}}^{\overline{BC}} f(x,y) dy dx \right| \quad (20)$$

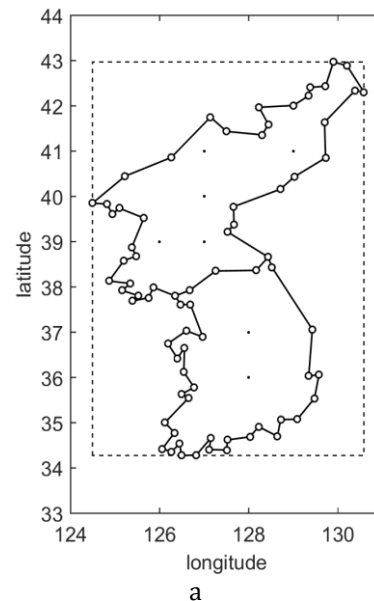
These integrals' boundaries may be replaced. Therefore, the absolute value of these integrals must be calculated. The sum of the probability values of all the triangles must be equal to 1 because of the required probability axioms, such as the one seen in Equation (21).

$$\sum P_j = 1 \quad (21)$$

A random triangle is selected according these probability values. The Algorithm 4a-b is used to generate random numbers in the selected triangle. Algorithm 6 is given in order to generate a random number using the inversion method in a polygonal area.

Algorithm 6. Non-uniform random number generation algorithm in the polygonal area by using the inversion method.

- Step 1. Divide the polygonal area into triangles with dominant points and selected grid points (Figure 4(a)) by using the Delaunay algorithm. If the polygonal area is not a convex area, the Delaunay algorithm determines triangles outside of the polygonal area. The center of gravity of each triangle is found in order to solve this problem and these centers which are outside of the polygonal area are deleted (Figure 4(b)).
- Step 2. Divide all obtained triangle, into two triangles, as can be seen in Figure 1(b) and calculate the selection probability (P_j) of all triangles by implementing Equation (20).
- Step 3. Select a random triangle according P_j probabilities.
- Step 4. Generate a random number (X,Y) in the selected triangle by using the Algorithm 4b. In Step 1 of Algorithm 4b, the value ξ_x must be changed to $\xi_x = U(0, P_j)$.



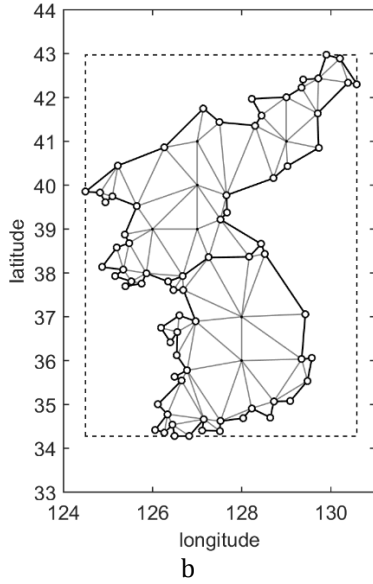


Figure 4. The implementation of the Delaunay triangulation method in the polygonal area; (a) The selection of grid points (b) The division of the polygonal area into triangles.

3. Results

Three separate samples are determined for generating non-uniform random numbers in the various geometric areas with the help of algorithms developed in Section 2.2.2. and Section 2.2.3. In one of these examples, the non-uniform random number generation in a triangular area is discussed. The another example, the non-uniform random number generation in the shape of the mainland of Korea (surrounding islands being excluded) is discussed. The last example, the non-uniform random number generation is performed with the help of the probability density values which are obtained by using the frequencies of the Australian rainforest data. Also, latitude-longitude is used in all samples. The latitudes and longitudes used in calculations are assumed to be in cartesian coordinates.

In three example, the non-uniform random number generation is performed by a simulation. The aims of these simulations are to find which random numbers generation methods run faster than the other methods and whether or not the generated random numbers, those generated by using the rejection and the inversion methods, come from the given joint probability density function. The simulation is made on a computer that has Intel® Core (TM) i5-4440 CPU, 8 GB of RAM. Also, Matlab® R15b software is used to make the calculations.

3.1. Non-uniform random number generation in a triangular area

In the first example, a probability density functions which is bounded by a triangular area, is given in Equation (22). The values of the probability density function are arbitrarily determined to provide that

the total value of the probability density function calculated in the triangular area equals to 1.

$$f(x, y) = \begin{cases} \frac{2}{3} e^{-(x-125)+(y-39)}, & (x, y) \in \Omega \\ 0 & (x, y) \notin \Omega \end{cases} \quad (22)$$

The triangle's corner points are arbitrarily determined. The probability density function is bound by a region consisting of 3 corner points $\Omega = \{p_k : (x_k, y_k) = \{(125.65, 39.52), (126.26, 40.86), (127, 40)\}, k = 1, 2, 3\}$. This region is shown in Figure 5(a) and the probability density function in the same region can be seen in Figure 5(b).

The inverse marginal distribution function and the inverse conditional distribution function of the joint probability density function given in Equation (22) is required for generating the non-uniform random numbers. However, since this process cannot be found analytically, the numerical method given in Section 2.2.2 is preferred. According to this, firstly the triangular area is taken as a single triangle (Figure 6(a)) and then the non-uniform random numbers are generated. After the triangular area is divided into four triangles (Figure 6(b)), the non-uniform random number is generated for each triangle separately. The non-uniform random numbers are compared with a simulation in terms of number of the triangle. The non-uniform random numbers which are generated according number of triangle are compared with a simulation.

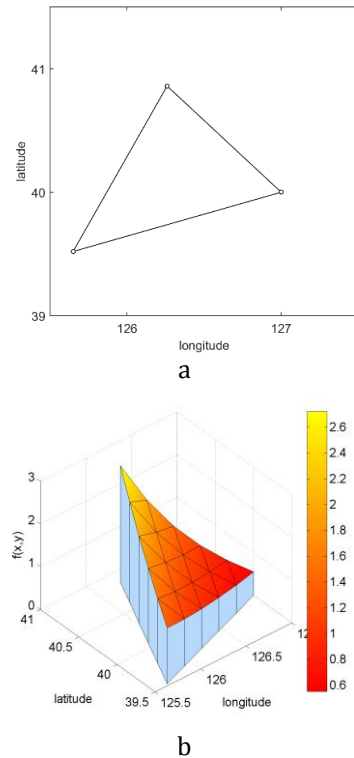


Figure 5. Definition of the first example, (a) The triangular area; (b) The probability density function in the triangular area.

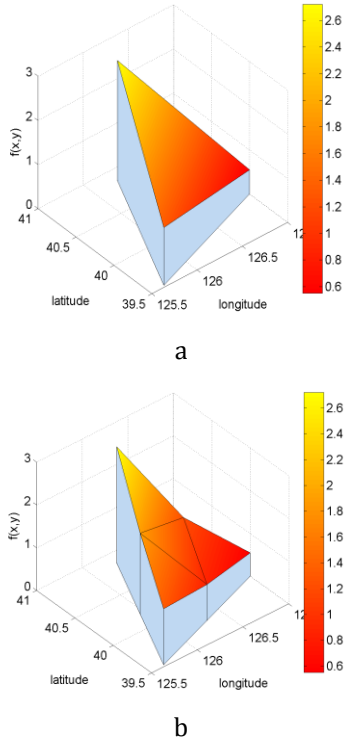


Figure 6. Numerical approximation to JPDF; (a) Definition of JPDF with using a linear plane; (b) Definition of JPDF with using four linear planes.

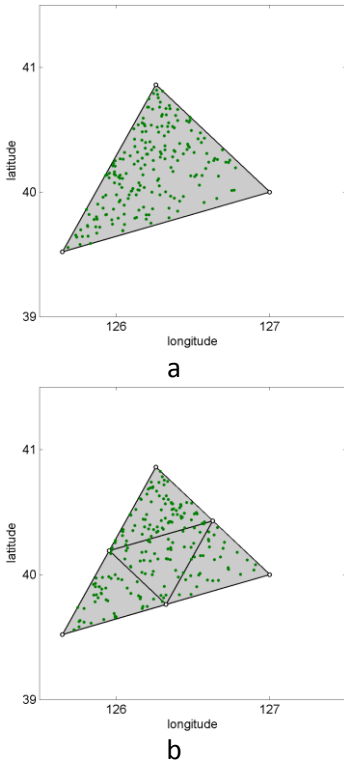


Figure 7. Random number generation; (a) Random numbers generated in a triangle area by using the inversion method; (b) Random numbers generated in four triangle areas by using the inversion method.

Random numbers are generated from the given probability density function according the two different generated non-uniform random number

methods. These random numbers are generated in two different sizes: $n = 200$ and $n = 1000$. This process is repeated 10,000 times and then the simulation is performed. The generated random numbers by the inversion method are illustrated in Figure 7.

3.1.1. χ^2 goodness of fit test for the triangle example

The χ^2 goodness of fit test is used to determine whether or not the generated random numbers come from the given joint probability density function in a triangular area [25]. Firstly, the triangular area is divided into classes by employing the χ^2 goodness of fit test in a triangular area. Generally, a rectangular area is used for the classification for bivariate examples. However, due to the difficulty of rectangular classification in a triangular region, the area is divided into 25 triangular classes (Figure 8(a)). Also, the expected frequency in the same small triangles can be ascertained by calculating the integral over the piece triangle mentioned in the previous section (Figure 8(b)).

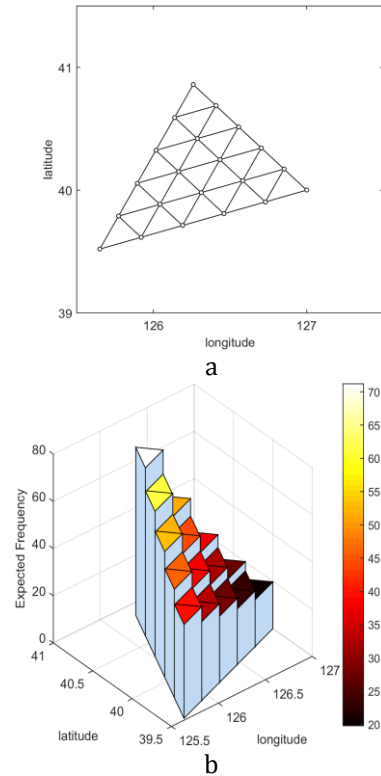


Figure 8. Calculation of the expected frequency in the triangular region for 1000 samples; (a) Dividing the triangular region into small triangles; (b) The expected frequencies of each small triangle.

The observed frequencies are found by dividing the triangular area into small classes in order to test the random numbers which are generated from each method (Figure 9).

The Chi-square test is performed 10,000 times for each number of triangles by using observed and

expected frequencies. A summary of the obtained results from the experiments are shown in Table 1.

Table 1. The results obtained from 10,000 trials for the inversion method

Number of Triangles	n=200		n=1000	
	POA (χ^2)	ACT (sec)	POA (χ^2)	ACT (sec)
1	94.26	0.0039	86.56	0.0063
4	97.52	0.0137	95.83	0.0167

*(POA: The percentage of accepted hypothesis; ACT: Average computational time)

H_0 : The generated non-uniform random numbers come from the given joint probability density function in the Equation (22).

H_1 : The generated non-uniform random numbers do not come from the given joint probability density function in the Equation (22).

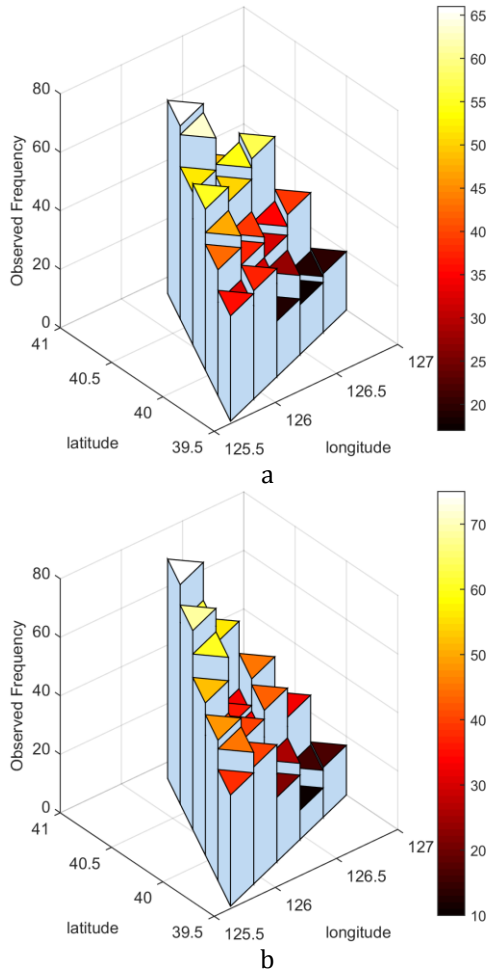


Figure 9. The observed frequencies according each method for 1000 samples; (a) The observed frequencies of the random number generated from the rejection method; (b) The observed frequencies of the random number generated from the inversion method.

The random numbers which are generated by utilizing the inversion method is tested at the 95% confidence level for $n = 200$. The H_0 hypothesis is accepted in approximately 94.26% of the 10,000 trials and the average computational time (ACT) is

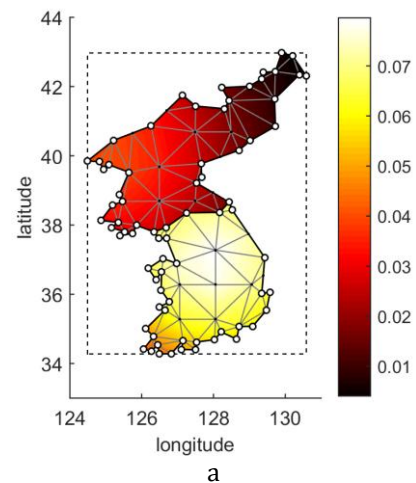
0.0039 second when one triangle is used. Furthermore, the H_0 hypothesis is accepted in approximately 97.52% of the 10,000 trials and the average computational time (ACT) is 0.0137 second when four triangles are used. On the other hand, the random numbers which are generated by utilizing the inversion method is tested at the 95% confidence level for $n = 1000$. The H_0 hypothesis is accepted in approximately 86.56% of the 10,000 trials and the average computational time (ACT) is 0.0063 second when one triangle is used. Furthermore, the H_0 hypothesis is accepted in approximately 95.83% of the 10,000 trials and the average computational time (ACT) is 0.0167 second when four triangles are used.

3.2. Non-uniform random number generation in the polygonal area

3.2.1. Using pdf from theoretical function

The Korea mainland is selected as the polygonal area in the second example. All areas of the Korea mainland are assumed to be surface areas for generating random numbers. This region is shown in Figure 10(a) and the probability density function in the same region can be seen in the Figure 10(b). Also, the piecewise probability density function in this region is given in Equation (23). Equation (23) is modeled based on the night light data which are obtained in the Korean mainland [35]. In the night light data, the amount of light in North Korea is less, and in South Korea is more. The values of the probability density function are approximately estimated from the night light data to ensure that the total value of the probability density function calculated in the bounded Korea mainland equals to 1.

$$f(x, y) = \begin{cases} \frac{1}{25} e^{-\frac{(x-125)^2 + (y-40)^2}{16}} & (x, y) \in \text{in North Korea} \\ \frac{2}{25} e^{-\frac{(x-128)^2 + (y-37)^2}{16}} & (x, y) \in \text{in South Korea} \\ 0 & \text{otherwise} \end{cases} \quad (23)$$



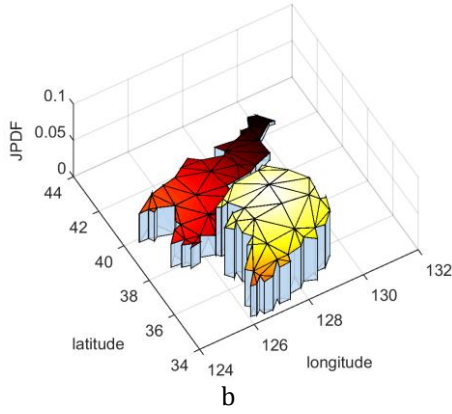


Figure 10. Definition of Example 2; (a) The probability density function in the polygonal area with colored surface on the Korea mainland; (b) The probability density function in the polygonal area with mesh surface on the Korea mainland.

It is aimed to generate the non-uniform random numbers according to the rejection and the inversion methods using JPDF given in Equation (23). However, the numerical approach is preferred to generate random numbers by the inversion method since the inverse marginal distribution function of the given piecewise JPDF functions and the inverse conditional distribution function cannot be found analytically.

Before generating a random number, the polygonal area must be divided into triangles (Figure 10). The expected frequencies of the desired triangles are determined by calculating the probability value (prism volume) of each triangle. The non-uniform random number is generated in the triangle by determining the amount of random numbers according to these frequencies.

Non-uniform random numbers are generated from the given probability density function according to the two different methods. The generated random numbers are illustrated in Figure 11. These random numbers are generated in two different sizes: $n = 1000$ and $n = 2000$. This process is repeated 10,000 times and then the simulation is performed.

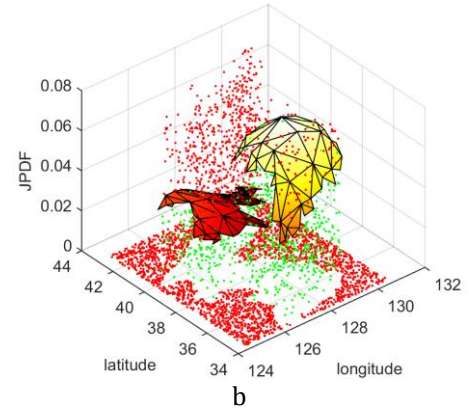
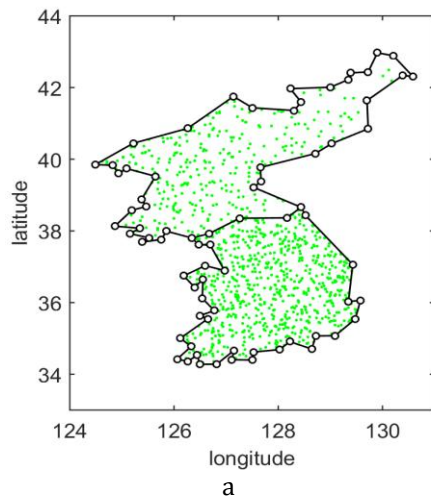


Figure 11. The non-uniform random number generation; (a) Random numbers generated by using the inversion method; (b) Random numbers generated by using the rejection method.

3.2.2. χ^2 goodness of fit test for the Korea example

Firstly, the polygonal area is divided into classes by employing the χ^2 goodness of fit test in polygonal area (Figure 12(a)). Also, the expected frequency in the small triangles can be ascertained by calculating the integral over the piece triangle mentioned in the previous sections (Figure 12(b)).

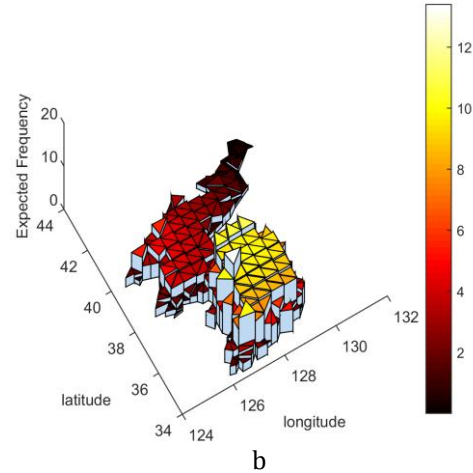
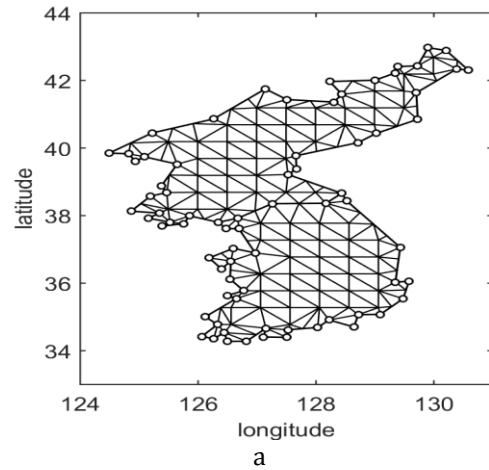


Figure 12. Calculation of the expected frequency in the polygonal region; (a) Dividing the polygonal area into small triangles; (b) The expected frequencies of each small triangle.

The chi-square test is performed 10,000 times for each method by using observed and expected frequencies. A summary of the obtained results from the experiments are shown in Table 2.

Table 2. The results obtained from N=10,000 trials

Methods	n=1000		n=2000	
	POA (%)	ACT (sec)	POA (%)	ACT (sec)
Rejection	94.38	0.3475	95.07	0.7348
Inversion	95.06	0.2841	95.11	0.3047

*(POA: The percentage of accepted hypothesis; ACT: Average computational time)

H_0 : The generated non-uniform random numbers come from the given joint probability density function in the Equation (23).

H_1 : The generated non-uniform random numbers do not come from the given joint probability density function in the Equation (23).

The random numbers which are generated by utilizing the rejection and the inversion methods are tested at the 95% confidence level. In these two methods, the H_0 hypothesis is accepted in approximately 95% of the 10,000 trials for both sample sizes. The average computational time (ACT) is 0.3475 second for $n = 1000$ and the ACT is 0.7348 for $n = 2000$ when the rejection method is utilized. Also, the ACT is 0.2841 second for $n = 1000$ and the ACT is 0.3047 for $n = 2000$ when the inversion method is utilized.

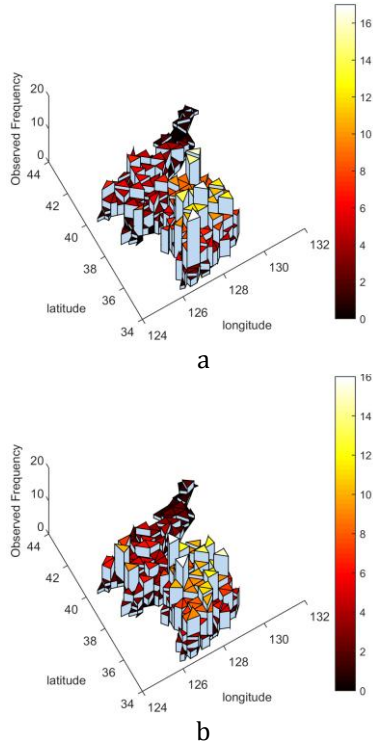


Figure 13. The observed frequencies according each method; (a) The observed frequencies of the non-uniform random numbers generated by utilizing the rejection method; (b) The observed frequencies of the non-uniform random numbers generated by utilizing the inversion method.

3.3.3. Using pdf from rainfall data

It may not always be possible to obtain a probability density function from real-life data. In this case, a probability density function approach can be obtained by taking the frequencies of the data. However, since the JPFD is not determined analytically, the non-uniform random number generation can be performed by using the numeric approach in Section 2.2.2. A model is obtained by using the average annual rainfall data of Australia for this example. The modeling is performed by collecting the real data in ranging (0.05) from latitude-longitude coordinates [36]. The modeled probability density function is shown in Figure 14.

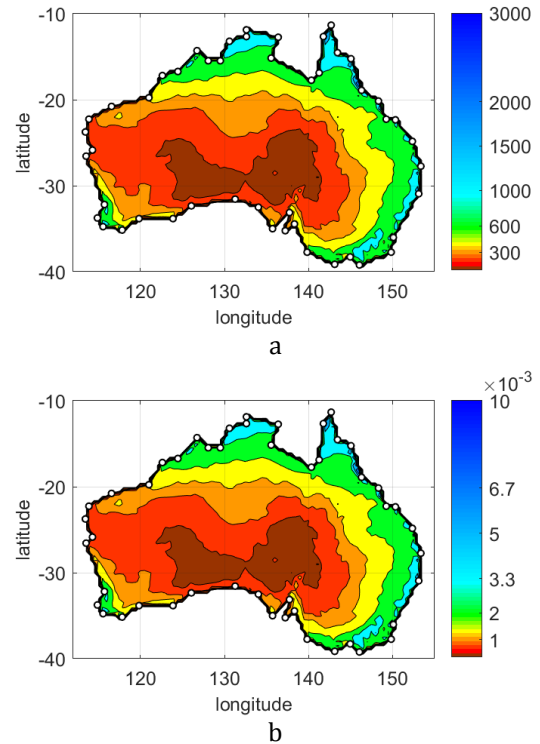


Figure 14. Definition of Example 3 (Australia mainland); (a) The colored contour representation of the observed frequencies; (b) The colored contour representation of the probability density function.

3.3.4. χ^2 goodness of fit test for the Australia example

The polygonal area is divided into classes by employing the χ^2 goodness of fit test in polygonal area. Also, the probability density function bounded in the Australia mainland is determined from rainfall data (Figure 15(a)). The observed frequencies are found by dividing the polygonal area into small classes in order to test the random numbers which are generated from each method (Figure 15 (b)).

The chi-square test is performed 10,000 times for each method by using observed and expected frequencies. A summary of the obtained results from the experiments are shown in Table 3.

Table 3. The results obtained from N=10,000 trials

Methods	n=1000		n=2000	
	POA (%)	ACT (sec)	POA (%)	ACT (sec)
Rejection	94.64	2.9106	95.17	5.6783
Inversion	94.62	1.2611	95.21	1.3725

*(POA: The percentage of accepted hypothesis; ACT: Average computational time)

H_0 : The generated non-uniform random numbers come from the given joint probability density function in the Figure (14).

H_1 : The generated non-uniform random numbers do not come from the given joint probability density function in the Figure (14).

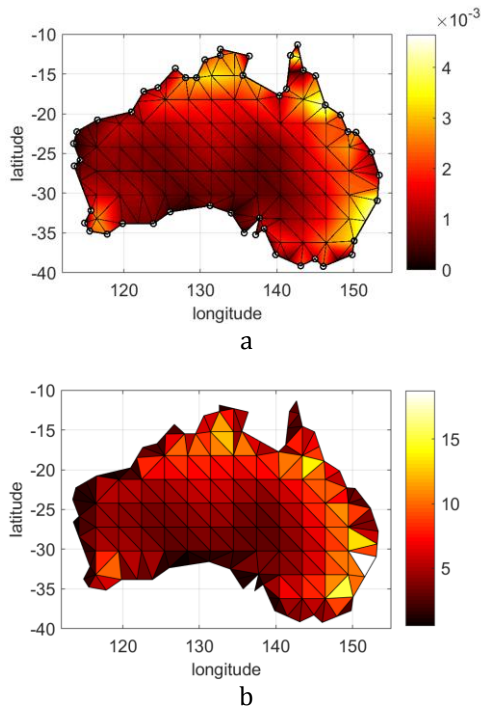


Figure 15. The polygonal area (Australia mainland); (a) The probability density function bounded in the Australia mainland; (b) The observed frequencies on the Australia mainland.

The expected frequency in the small triangles can be ascertained by calculating the integral over the piece triangle mentioned in the previous sections. The expected frequencies are calculated by using the inversion and the rejection methods (Figure 16).

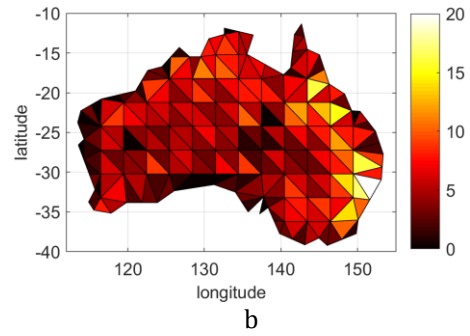
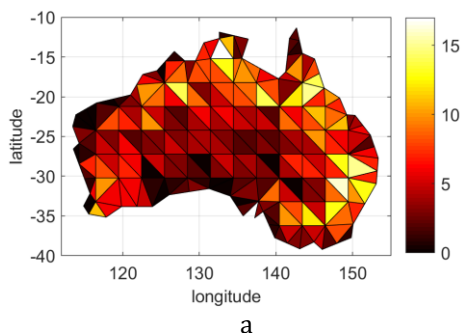


Figure 16. The representation of the expected frequencies; (a) The expected frequencies calculated by the inversion method; (b) The expected frequencies calculated by the rejection method.

The random numbers which are generated by utilizing the rejection and the inversion methods are tested at the 95% confidence level. In these two methods, the H_0 hypothesis is accepted in approximately 95% of the 10,000 trials for both sample sizes. The average computational time (ACT) is 2.9106 second for $n = 1000$ and the ACT is 5.6783 for $n = 2000$ when the rejection method is utilized. Also, the ACT is 1.2611 second for $n = 1000$ and the ACT is 1.3725 for $n = 2000$ when the inversion method is utilized.

4. Discussion and Conclusion

The non-uniform random numbers are generated from an arbitrary bivariate distribution within a polygonal area by using the rejection and the inversion methods. Three separate samples are determined to generate non-uniform random numbers in the various polygonal areas with the help of developed algorithms from bivariate distribution functions.

In the first example, the generated random numbers come from the given joint probability density function in approximately 95% of the 10,000 trials when one triangle and four triangles are used for both sample sizes, $n = 200$ and $n = 1000$. The high percentage of acceptance demonstrates that the random number generation algorithms run correctly for each approximation in the triangular area. That said, the first approximation is faster than the other approximation (for four triangle) in terms of the average computational time. In the second example, the random numbers which are generated with the rejection and the inversion methods are tested at 95% confidence level. In these two methods, the generated random numbers come from the given joint probability density function in approximately 95% of the 10,000 trials for both sample sizes, $n = 1,000$ and $n = 2,000$. The high percentage of acceptance shows that the non-uniform random number generation algorithms run correctly for each method in the polygonal area. Also, the inversion method is faster than the rejection method in terms

of the average computational time. In last example, the random numbers which are generated by utilizing the rejection and the inversion methods are tested at the 95% confidence level. In these two methods, the generated random numbers come from the given joint probability density function in approximately 95% of the 10,000 trials for both sample sizes, $n = 1,000$ and $n = 2,000$. Since the level of significance is taken as $\alpha = 0.05$, this high percentage of acceptance demonstrates that the non-uniform random number generation algorithms run correctly for each method in the polygonal area. That said, the inversion method is faster than the rejection method in terms of the average computational time.

In conclusion, the non-uniform pseudo random numbers are generated simply, consistently and accurately from a bivariate probability density function whose definition range is bounded by an arbitrary polygonal area by employing the rejection and the inversion methods.

References

- [1] Tausworthe, R. 1965. Random numbers generated by linear recurrence modulo two. *Mathematics of Computation*, 19(90), 201-209.
- [2] Hayashi, S., Tanaka, Y., Kodama, E. 2001. A new manufacturing control system using Mahalanobis distance for maximising productivity. *Semiconductor Manufacturing Symposium*, 2001 IEEE International, 59-62.
- [3] Dougherty, E. 1999. Random processes for image and signal processing. *SPIE Optical Engineering Press*, 592s.
- [4] Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5-32.
- [5] Gottselig, J., Adam, M., Retey, J., Khatami, R., Achermann, P., Landolt, H. 2006. Random number generation during sleep deprivation: effects of caffeine on response maintenance and stereotypy. *Journal of sleep research*, 15(1), 31-40.
- [6] Gonzalo, J. 1994. Five alternative methods of estimating long-run equilibrium relationships. *Journal of econometrics*, 60(1), 203-233.
- [7] Kobayashi, H., Mark, B., Turin, W. 2011. Probability, Random Processes, and Statistical Analysis. Cambridge University Press Textbooks, New York, 812s.
- [8] Berg, H. 1993. Random walks in biology. Princeton University Press, 152s.
- [9] Beard, R. 2013. Risk theory: the stochastic basis of insurance. Springer Science & Business Media, 195s.
- [10] Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E. 2001. A statistical test suite for random and pseudorandom number generators for cryptographic applications. Booz-Allen and Hamilton Inc McLean Va.
- [11] Ross, S. 2006. Simulation. Elsevier, 285s.
- [12] Yannakakis, G., Togelius, J. 2011. Experience-driven procedural content generation. *Affective Computing, IEEE Transactions on*, 2(3), 147-161.
- [13] Fuller, W. 2011. Sampling statistics (Vol. 560). John Wiley & Sons, 472s.
- [14] Devroye, L. 1986. Non-uniform Random Variate Generation. Springer, New York, 843s.
- [15] Rubinstein, R., Kroese, D. P. 2008. Simulation and the Monte Carlo method. John Wiley & Sons, 304s.
- [16] Walck, C. 2007. Handbook on statistical distributions for experimentalists, 190s.
- [17] Boes, D., Graybill, F., Mood, A. 1974. Introduction to the Theory of Statistics. McGraw-Hill, 564s.
- [18] Chu, D., Fotouhi, A. 2009. Distance between bivariate beta random points in two rectangular cities. *Communications in Statistics—Simulation and Computation*, 38(2), 257-268.
- [19] Haight, F. 1964. Some probability distributions associated with commuter travel in a homogeneous circular city. *Operations Research*, 12(6), 964-975.
- [20] Wang, M., Kennedy, W. 1990. Comparison of algorithms for bivariate normal probability over a rectangle based on self-validated results from interval analysis. *Journal of Statistical Computation and Simulation*, 37(1-2), 13-25.
- [21] Robert, C., Casella, G. 2013. Monte Carlo Statistical Methods. Springer Science & Business Media, New York, 649s.
- [22] Gentle, J. E. 2006. Random number generation and Monte Carlo methods. Springer Science & Business Media, New York, 247s.
- [23] Johnson, M. E. 2013. Multivariate statistical simulation: A guide to selecting and generating continuous multivariate distributions. John Wiley & Sons, Canada, 240s.
- [24] Kesemen, O., Doğru, F. 2011. Cumulative Distribution Functions of Two Variable in Polygonal Areas. 7. International Statistics Congress, 28 April – 1 May, Antalya, 150-151.
- [25] Kesemen, O., Uluyurt, T. 2013. Bivariate Chi-Square Goodness of Fit Test in Polygonal Areas. 8. International Statistics Congress, 28-30 October, Antalya, 280-281.
- [26] L'Ecuyer, P. 2012. Random Number Generation. ss 35-71. Gentle J.E., Härdle W. K., Mori Y., ed. 2012. Handbooks of Computational Statistics Springer Berlin Heidelberg, 1192s.

- [27] Burden, R., Faires, J. 2011. Numerical Analysis. Brooks/Cole 7, 912s.
- [28] Thomas, D., Luk, W. 2007. Non-uniform random number generation through piecewise linear approximations. IET Computers A Digital Techniques, 1(4), 312-321.
- [29] Douglas, D., Peucker, T. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: The International Journal for Geographic Information and Geovisualization, 10(2), 112-122.
- [30] Haines, E. 1994. Point in polygon strategies. pp 24-26. Heckbert P. S., ed. 1994. Graphics gems IV, Academic Press Professional, Cambridge, 575s.
- [31] Hormann, K., Agathos, A. 2001. The point in polygon problem for arbitrary polygons. Computational Geometry, 20(3), 131-144.
- [32] Gudmundsson, J., Haverkort, H., Van Kreveld, M. 2005. Constrained higher order Delaunay triangulations. Computational Geometry, 30(3), 271-277.
- [33] Shewchuk, J. 1996. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. ss 203-222. Lin M.C., Manocha D., eds. 1996. Applied Computational Geometry Towards Geometric Engineering. Lecture Notes in Computer Science, Springer Berlin Heidelberg, 222s.
- [34] Forbes, C., Evans, M., Hastings, N., Peacock, B. 2011. Statistical distributions. New Jersey: John Wiley & Sons, 212s.
- [35] Earth Observation Group (EOG). <https://www.ngdc.noaa.gov/eog/>. (Date of Access: 27.02.2018).
- [36] Climate Change in Australia. Retrieved from CSIRO and Bureau of Meteorology. <http://www.climatechangeinaustralia.gov.au/> (Date of Access: 12.06.2016).