GÜFBED/*GUSTIJ* (2019) 9 (1): 88-98
DOI: 10.17714/gumusfenbil.413709 Araştırma Makalesi / *Research Article*

A Novel Approximation for Computation Bivariate Distribution Functions in Polygonal Area

Çokgensel Alanda İki Değişkenli Dağılım Fonksiyonunun Hesaplanmasında Yeni Bir Yaklaşım

Orhan KESEMEN*[1,a], Buğra Kaan TİRYAKİ[1,b], Tuncay ULUYURT[2,c]
[1]*Karadeniz Technical University, Department of Statistics and Computer Sciences, 61080, Trabzon, Turkey*
[2]*Artvin Coruh University, Arhavi Vocational School, 08200, Artvin, Turkey*

• Geliş tarihi / *Received*: 09.04.2018 • Düzeltilerek geliş tarihi / *Received in revised form*: 13.06.2018 • Kabul tarihi / *Accepted*: 22.06.2018

Abstract
Generally bivariate probability density function defined in a rectangular area is used to calculate the cumulative distribution function from the bivariate probability density function. However, definition limits of the probability density functions being non-rectangular are in existence in practice. In this paper, primarily arbitrary non-rectangular areas are defined by applying a polygonal approach. The polygonal area obtained as a result of this approach constitutes boundaries of the probability density function. Thus, the bivariate piecewise probability density function can be defined in an arbitrary area. Then the cumulative distribution function is calculated in the obtained area. Two types of approaches are used for these calculations. The first approach is applied to take integral analytically of bivariate continuous probability density function in the polygonal area. The second approach is developed a numerical method since the explicit integral of the selected probability density function cannot be found.

Keywords: Cumulative distribution function, Probability density function based on polygon, Bivariate distribution functions, Bivariate piecewise distribution functions

Öz
İki değişkenli olasılık yoğunluk fonksiyonundan, birikimli dağılım fonksiyonunu hesaplamak için genellikle dikdörtgensel bir alanda tanımlanmış iki değişkenli olasılık yoğunluk fonksiyonu kullanılır. Ancak uygulamada, tanım bölgesi dikdörtgensel bir alan olmayan birçok olasılık yoğunluk fonksiyonu mevcuttur. Bu çalışmada öncelikle dikdörtgen olmayan keyfi alanlar, çokgensel bir yaklaşım uygulanarak tanımlanmıştır. Bu yaklaşım sonucunda elde edilen çokgensel bölge, olasılık yoğunluk fonksiyonunun tanımlandığı sınırlarını oluşturmuştur. Böylece, iki değişkenli parçalı olasılık yoğunluk fonksiyonu, keyfi bir alanda tanımlanabilir. Elde edilen tanım bölgesinde birikimli dağılım fonksiyonu hesaplamaları yapılmıştır. Bu hesaplamalarda iki tür yaklaşım kullanılmıştır. İlk yaklaşım çokgensel alan üzerinden iki değişkenli sürekli olasılık yoğunluk fonksiyonunun analitik integrali alınarak yapılmıştır. İkinci yaklaşım ise seçilen olasılık yoğunluk fonksiyonun integralinin açık bir şekilde hesaplanamaması durumunda uygulanması için geliştirilen sayısal yöntemdir.

***Anahtar kelimeler**: Birikimli dağılım fonksiyonu, Çokgen tabanlı olasılık yoğunluk fonksiyonu, İki değişkenli dağılım fonksiyonları, İki değişkenli parçalı dağılım fonksiyonları*

*[a] Orhan KESEMEN; okesemen@gmail.com; Tel: (0462) 377 26 55; orcid.org/0000-0002-5160-1178
[b] orcid.org/0000-0003-0995-7389 [c] orcid.org/0000-0002-4331-1592

1. Introduction

Usage of bivariate distribution functions are existence as well as usage of a univariate probability functions generally (Martinez and Martinez, 2002). Usually, the definition range of bivariate probability density function is shown with either a semi-infinite or infinite range such as $f(x,y):(-\infty,\infty)\times(-\infty,\infty)\to\mathbb{R}^2$ or finite range such as $f(x,y):[a,b]\times[c,d]\to\mathbb{R}^2$ (Roussas, 2003; Miller and Childers, 2012). In both cases, it is possible to calculate the cumulative distribution function by conventional methods. However, the definition range of bivariate probability density function is a rectangular or infinite width area is not always possible in practice. Therefore, probability density function in an arbitrary area (Ω) can be defined as $f_{XY}(x,y):(X,Y)\in\Omega\to\mathbb{R}^2$. Consequently, different approaches must be used for calculations in an arbitrary field.

In this paper, an arbitrary area is defined by converting into to the polygonal area with the determination of its dominant points because of the difficulty of mathematical definition of an arbitrary area. As a result of this conversion, the polygonal area constitutes the boundaries of probability density function. Geometric approaches can be used for calculation of distribution function in the obtained polygonal area (confined area) based on the defined boundaries. A similar approach was applied to this probability density function in the polygonal area which has a uniform distribution by (Kesemen and Doğru, 2011). When it has a non-uniform distribution, the cumulative distribution function is calculated approximately by using column blocks based on small rectangles. This method does not provide desired perfection in terms of both computational accuracy and computational time. Unlike their work, the cumulative distribution function is calculated by using column blocks based on small triangles instead of rectangle blocks here. Two types of approaches are used for calculating cumulative distribution function. The first approach is performed by taking integral over the continuous functions. The second approach is a numerical method which is used on the above-mentioned triangle. In this case integral of the selected probability density function is not calculated explicitly.

Bivariate probability density function, which is bounded by arbitrary non-uniform limits, fields of application can be given examples such as rate of pollution or crime rate in a city, distribution of earthquake frequency in a country, dispersion density of insects in the field, traffic congestion in a specific region, seen locations and frequency of epidemic disease in a country. Thus, the region we mentioned may have many partial density functions in rectangular area. Many cities divided physically or politically may be given as examples for this situation (e.g. Belfast, Beirut, Jerusalem, Mostar, and Nicosia). Especially, the data in the cities which are divided politically starts to change gradually. In this case, it may not be possible to evaluate the entire city as a region. Probability density function whose boundaries are entirely arbitrary polygonal area may be needed in each of these examples.

2. Bivariate Distribution Functions

Some definitions have to be explained as the bivariate cumulative distribution function before calculating the distribution value of it. These definitions can be given as joint and marginal probability density functions.

2.1. Joint Cumulative Distribution Function (JCDF)

We assume that X and Y are continuous random variables to calculate the cumulative distribution function of a bivariate probability density function. In this case, the joint cumulative distribution function of (X,Y) is calculated as in Equation (1) (Walck, 2007; Kobayashi et al., 2011).

$$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,\mathrm{v})\,dv\,du \tag{1}$$

Also, the joint cumulative distribution function of (X, Y) is defined as volume in 3-dimensional under the condition below.

$$0\le F_{XY}(x,y)\le 1 \tag{2}$$

If the probability density function is defined in a limited rectangular area ($[a,b]\times[c,d]$), the joint cumulative distribution function is calculated as follows (Kay, 2006; Montgomery and Runger, 2010).

$$F_{XY}(x,y)=\int_{a}^{x}\int_{c}^{y}f(u,\mathrm{v})\,dv\,du \tag{3}$$

2.2. Marginal Probability Density Function

The marginal probability density function is used to obtain one-way variation of bivariate cumulative distribution function. The marginal probability density function of a random variable X is shown with $f_X(x)$ and function curve is called the probability density function curve of X. Integrating over all y's in the range of $(-\infty, \infty)$ is adequate for obtaining the marginal probability density function of X by using the joint probability density function as in the following equation.

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad (4)$$

Cumulative distribution function (CDF) of the probability density function $\left(f_X(x)\right)$ is calculated as follows (Kay, 2006; Montgomery and Runger, 2010).

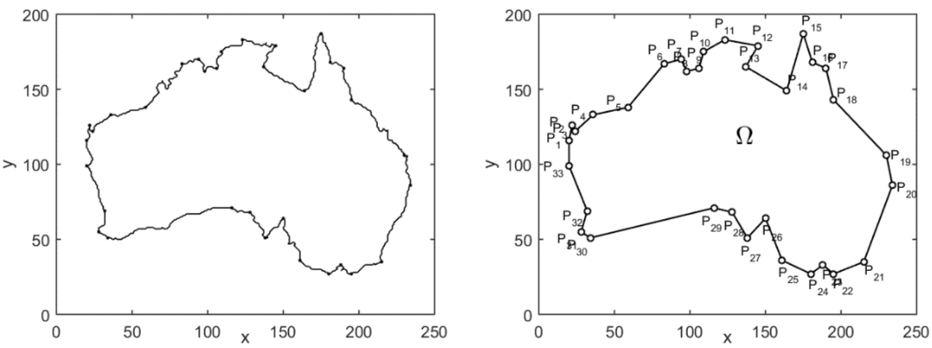
$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx \quad (5)$$

Also, the same procedure can be performed for the variable Y.

3. Computation JCDF in Polygonal Area

The most appropriate method is polygonal definition to determine an arbitrary area as a geometric shape in two dimensions (Kesemen and Doğru, 2011). The calculations can be performed by defining suitable polygon in an arbitrary given area. In this situation, dominant points of the boundaries are determined to define a polygonal area for arbitrary area. These dominant points can be determined manually or automatically by using polygonal approximation algorithms (Douglas and Peucker, 1973). An arbitrary region being bounded by the shape of the Australian mainland is selected as an example for this situation (Figure 1(a)). The shape of the Australian mainland is converted into a polygonal area with the help of the dominant points so as to calculate the cumulative distribution function from the probability density function in a given region in this way (Figure 1(b)). Australian mainland is defined approximately in a representative manner with a polygon $\Omega = \{p_i = (x_i, y_i), i = 1,2, \dots, N\}$ consists of the thirty-three corner points which are selected manually (Figure 1).

Firstly, the polygonal area is divided into triangles to calculate the cumulative distribution function with the help of Delaunay triangulation algorithm (Shewchuk, 1996; Gudmundsson et al., 2005) by using the corner and grid points (Figure 2(a)). In a polygonal structure, these grid points and the triangular grid width are chosen by the researcher. If this width is too small, the computation time will increase while the calculated error reduces. Conversely, if the grid width is too large, the calculation time will decrease while the calculation error increases. Consequently, the optimal grid width is determined when the calculation difference between two different grid widths is smaller than a certain tolerance value.



(a) (b)

Figure 1. The representation of two-dimensional area (Australia) which consists of thirty-three corner points; **(a)** Arbitrary area; **(b)** The polygonal area

If the polygonal area is not a convex polygon as in the Figure 2(a), the triangulation algorithm finds triangles outside of the polygon. Each triangles which are fallen out the polygon are eliminated (Figure 2(b)). In some cases, the edge lines of the triangle intersect the edges of the polygon. In this case, the Delaunay algorithm develops an algorithm for being edges of the triangle instead of polygon edges. Whether a point is in the polygon need to be investigated while deleting the centroids which are fallen outside the polygon. This problem is known as the point problem in the polygons (Haines, 1994; Hormann and Agathos, 2001). A perpendicular straight line is drawn from randomly selected point (X, Y), according the proposed odd-even method to line $y = x_{min}$ for solving this problem. Also, how many points of constituting all lines of the polygon are intersected by the perpendicular straight line must be determined. If the number of point of intersection is odd, the selected point is inside of the polygon, or else the selected point is outside of the polygon.

Two-dimensional colored view of the probability density function defined in Equation (19) which is bounded by Australian continent and the triangulation in this area are shown in Figure 3. Furthermore, the change of color shows the probability density function in the area.

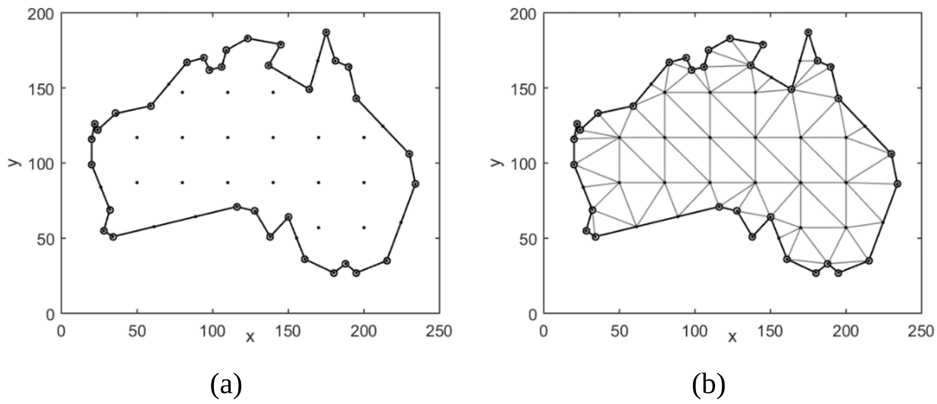


Figure 2. The division of the polygon into triangles; **(a)** Definition of the grid points in polygon; **(b)** Triangulation of the polygon (Ω).

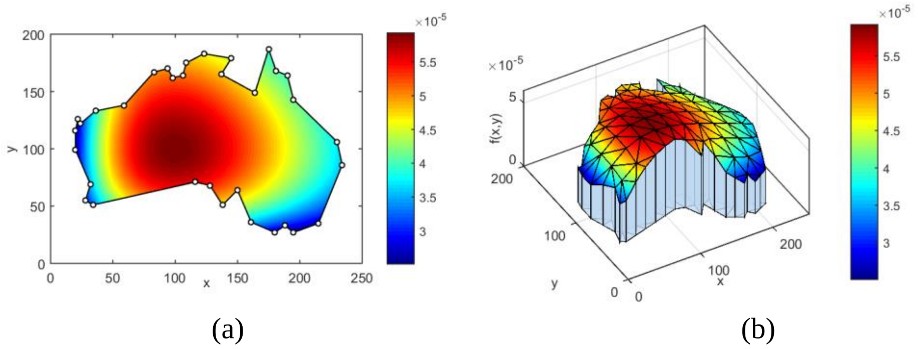


Figure 3. The probability density function bounded by the Australian continent; **(a)** Two-dimensional colored view of the probability density function; **(b)** The mesh view of the probability density function.

After the triangulation process, the probability value of the triangle can be calculated by taking integral separately from each triangle as in the following equation.

$$Pr\left((X, Y) \in \Omega_j\right) = \iint_{\Omega_j} f(x, y)dxdy \qquad (6)$$

Here, intersections of Ω_j's are empty and Ω_j shows the calculated triangular area. Also, j indicates subscript of the selected triangle. The sum of the probability values of each triangle allows to calculate probability value of the whole polygon because of the union of Ω_j is equal to the area of Ω. The sum of all the triangles probabilities must be equal to 1 below.

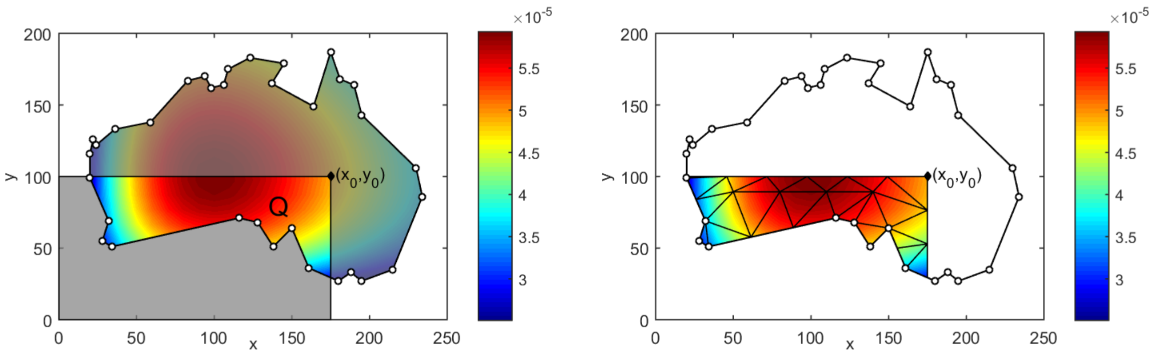
$$Pr\big((X, Y) \in \Omega\big) = \sum_j Pr\left((X, Y) \in \Omega_j\right) = 1 \qquad (7)$$

3.1. Definition of Intersection Region

The intersection of Ω area and the area of $(-\infty, x_0) \times (-\infty, y_0)$ is to be determined for calculation of the cumulative distribution function from probability density function bounded by a polygonal area (Figure 4). The probability density function in Figure 4 is defined in Equation (19) and is the same with Figure 3.

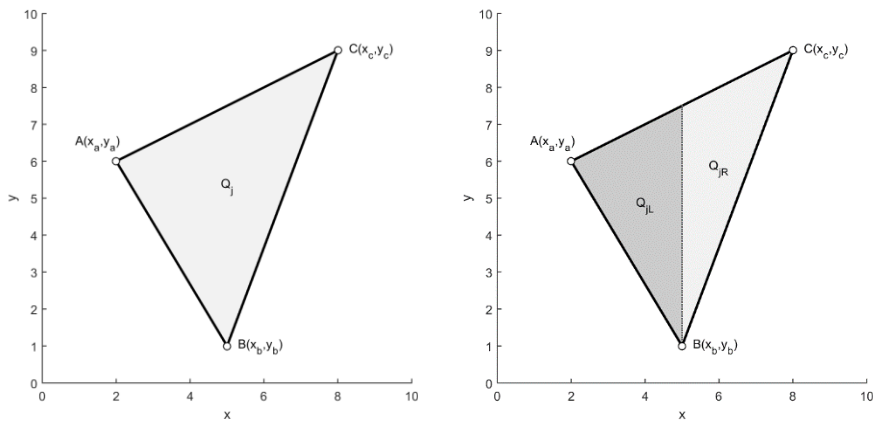
The intersection area is denoted by $Q = \{q_i = (x_i, y_i), i = 0,1,2, \ldots, M\}$. For this calculation, perpendicular lines are drawn from point (x_0, y_0) which is supposed to be calculated the cumulative distribution function value to both x and y axes. Left bottom area of the point (x_0, y_0) is denoted by the intersection area Q. In other words, remaining area between the boundaries of the polygon and these perpendicular lines are determined as the intersection area of Q (Figure 4(a)). As the intersection area can be only a confined area, it can be in the form of multi confined area as well (Margalit and Knott, 1989). Primarily, the intersection area (Q) is divided into triangles for determining the cumulative distribution function with both analytical and numerical methods. Intersection area triangles are obtained by eliminating triangles fallen out the intersection area (Q) (Figure 4(b)). The value of the cumulative distribution function can be calculated by summing up probability values of these triangles.

Although the integration is quite easy to calculate from the bivariate probability density function according to a fixed limit on the rectangular area, the integration in the triangular area is a bit complicated. To eliminate this complexity, the median of the components $\{x_a, x_b, x_c\}$ of x is found from the corners coordinates of triangle area Q_j in Figure 5(a). Drawing a parallel line from y-axis to the median value divides the triangle into two parts. Hence, the division process makes easy to calculate integral in the triangle area Q_j in Figure 5(b).



(a) (b)

Figure 4. Intersection area; **(a)** Determination of the intersection area (Q); **(b)** Triangulation of the intersection area.



(a) (b)

Figure 5. Representation of the triangular area; **(a)** The area Q_j; **(b)** Determination of the triangular areas Q_{j_L} and Q_{j_R}

The region Q_j is divided into two triangles (Q_{jL}, Q_{j_R}) and integral calculus is performed separately in each triangles. Also, their total probability value can be found in the total integration as follows.

$$Pr\left((X,Y) \in Q_j\right) = \iint_{Q_j} f(x,y)dx\,dy = \iint_{Q_{j_L}} f(x,y)dx\,dy + \iint_{Q_{j_R}} f(x,y)dx\,dy \quad (8)$$

Marginal probability density function of the triangle Q_{j_L} is given as Equation (9). $\overline{AB}$ shows the equation of the line through points A and B, also $\overline{AC}$ shows the equation of the line through points A and C.

$$f_{j_L}(x) = \left|\int_{\overline{AB}}^{\overline{AC}} f(x,y)\,dy\right| \quad (9)$$

Volume (probability value) of the left side region of the triangle Q_j is obtained as follows:

$$Pr\left((X,Y) \in Q_{j_L}\right) = \int_{x_a}^{x_b} f_{j_L}(x)dx \quad (10)$$

Likewise, the marginal probability density function of the triangle Q_{j_R} is defined as in Equation (11). $\overline{BC}$ shows the equation of the line through points B and C, also $\overline{AC}$ shows the equation of the line through points A and C.

$$f_{j_R}(x) = \left|\int_{\overline{BC}}^{\overline{AC}} f(x,y)\,dy\right| \quad (11)$$

Volume (probability value) of the right-side region of the triangle Q_j is obtained below.

$$Pr\left((X,Y) \in Q_{j_R}\right) = \int_{x_b}^{x_c} f_{j_R}(x)dx \quad (12)$$

The total probability of the triangle Q_j is calculated as follows.

$$Pr\left((X,Y) \in Q_j\right) = Pr\left((X,Y) \in Q_{j_L}\right) + Pr\left((X,Y) \in Q_{j_R}\right) \quad (13)$$

Total probability value of all triangles Q_j in the intersection region Q gives the distribution value of point (x_0, y_0) as in the following equation.

$$F_{XY}(x_0,y_0) = Pr\big((X,Y) \in Q\big) = \sum_j Pr\left((X,Y) \in Q_j\right) \quad (14)$$

3.3. Numerical Computation of CDF in a Triangle Area

Analytical methods which are mentioned in the previous sections can be preferred for the calculation of probability values in a polygonal area. However, the analytical calculation may not be possible in the case of failure of integration of marginal function from the given probability density function. In this case, numerical methods are applied to calculate the probability value. Whereas, the area under the curve and the volume under the surface is calculated with the univariate function and bivariate function respectively.

The trapezoidal rule is the most common method which used to calculate numerical integration of the univariate functions (Howard and Musto, 2008). When trapezoidal rule is applied to calculate the integral of bivariate function in three-dimensional situation, the truncated right rectangle prism can be used instead of trapezoid. However, the voids around edges of the rectangular pieces in polygon increase the error value in the calculation. Furthermore, triangles are the best shapes to define a surface (Boissonnat and Teillaud, 2007). The volume remaining below surface in the polygonal area which is divided into triangles can be easily calculated with the help of truncated right triangular prisms (Figure 6). The calculation of integral in the area we mentioned can be easily performed by summing volume of all triangular prisms. Also, K shows that the number of parts in an edge of the triangle.

An example is given in Figure 7 for finding the probability value by using the trapezoidal rule in triangle area in concern with sub-triangle division. The triangle area is divided into a triangle in Figure 7(a), when divided into four sub-triangles in Figure 7(b), when divided into nine sub-triangles in Figure 7(c), when divided into sixteen sub-triangles in Figure 7(d), when divided into twenty-five sub-triangles in Figure 7(e) finally, when divided into thirty-six sub-triangles in Figure 7(f) are shown in Figure 7.

Edge height of the sub-triangle prism is found by calculating the probability density function value of each sub-triangle's corner coordinates. The area of polygon which consists of the known coordinates of N corners in two-dimensional plane is given as in Equation (15) (Preparata and Shamos, 2012). This equation is found by $i+1 \rightarrow 1$ when $i = N$. Also, Equation (15) is used when the polygonal area is a triangle (Figure 8).

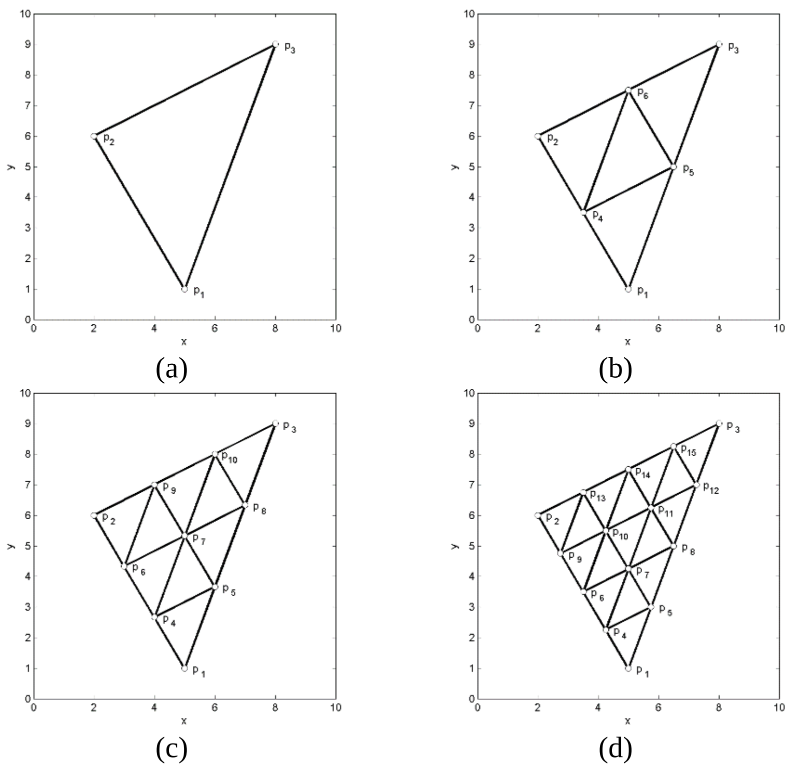


Figure 6. The division of a triangle into the sub-triangles; **(a)** Not to occur a triangle division for $K = 1$ **(b)** Four sub-triangles division of the triangle area for $K = 2$; **(c)** Nine sub-triangles division of the triangle area for $K = 3$; **(d)** Sixteen sub-triangles division of the triangle area for $K = 4$

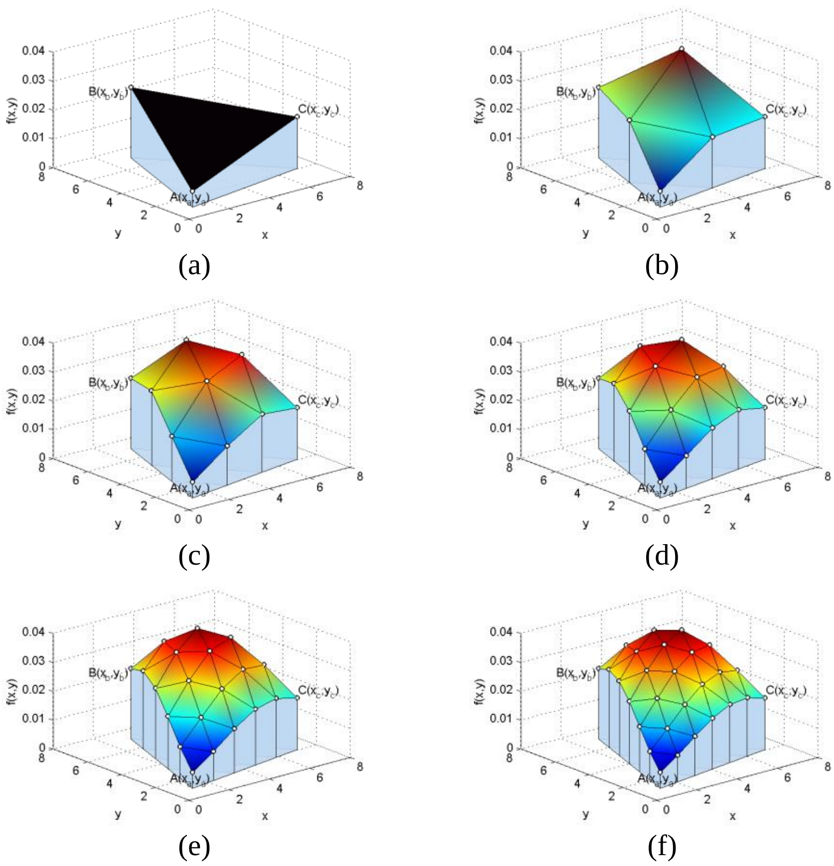


Figure 7. (a) The division to one sub-triangles in the triangle area.; **(b)** The division to four sub-triangles in the triangle area; **(c)** The division to nine sub-triangles in the triangle area; **(d)** The division to sixteen sub-triangles in the triangle area; **(e)** The division to twenty-five sub-triangles in the triangle area; **(f)** The division to thirty-six sub-triangles in the triangle area.

$$A = \frac{1}{2}\left|\sum_{i=1}^{N}(x_{i+1}y_i - x_i y_{i+1})\right| \quad (15)$$

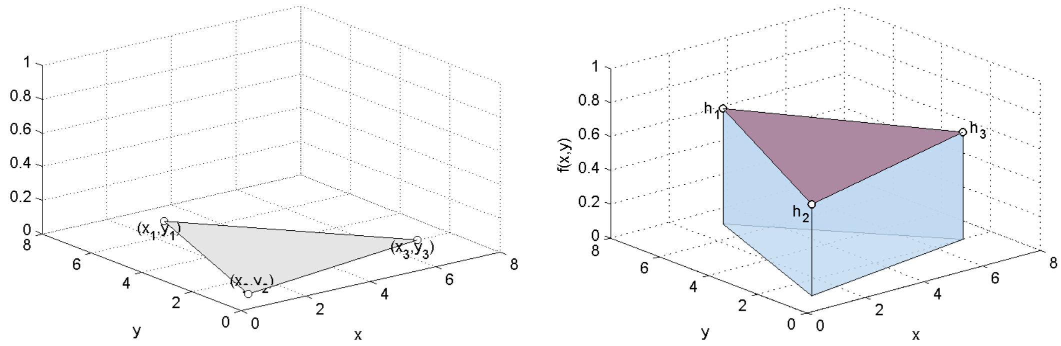
The volume of prism, whose base area is known is calculated below,

$$V = A.h \quad (16)$$

where h is the height of the prism.

If the perpendicular triangle prism has a truncated surface, the volume of k^{th} prism is calculated as in (17) (Eshbach et al., 1990; Badiru and Omitaomu, 2010; Straszewicz, 2014). $\{h_{\mathrm{k1}}, h_{\mathrm{k2}}, h_{\mathrm{k3}}\}$ are the heights of corner coordinates of the triangular prism (Figure 8(b)).

$$V_k = A_k \frac{h_{\mathrm{k1}}+h_{\mathrm{k2}}+h_{\mathrm{k3}}}{3} \quad (17)$$



(a) (b)

Figure 8. The calculation of volume of the truncated triangular prism; **(a)** The base area; **(b)** The truncated triangular prism.

After the calculation of volume of all sub-triangles, the total volume of all sub-triangles is equal to the total probability value for the triangle as follows. Also, K shows that the number of parts in an edge of the triangle.

$$P\left((X,Y) \in Q_j\right) = \sum_{k=1}^{K^2} V_k \quad (18)$$

4. Experimental Results

In this section, the cumulative distribution function is calculated from a probability density function whose definition area (Ω) is bounded by the shape of the continent of Australia for different points (x, y) by using both analytical and numerical methods.

Example 1. The arbitrary probability density function which is defined in region Ω is given as in the following equation.

$$f(x,y) = \begin{cases} \frac{1}{2.2909\times10^7} xy\, e^{-\frac{(x+y)}{100}}, & (x,y) \in \Omega \\ 0, & (x,y) \notin \Omega \end{cases} \quad (19)$$

The values of the probability density function are arbitrarily determined to provide that the total value of the probability density function calculated in the polygonal area (Ω) equals to 1. The polygon's (Ω) corner points also are arbitrarily determined. The mesh and contour view of the probability density function explained in Equation (19) is shown in Figure 3. Firstly, the cumulative distribution values $F(x, y)$ of each grid points (x, y) in area Ω are calculated analytically in Table 1. The cumulative distribution value of 99 grid points is calculated in the Table 1. The cumulative distribution value of the points in $(x < x_{min}\ AND\ y < y_{min})$ area is zero. The reason of this is that the region whose left and below points do not intersect the polygon. On the other points, the cumulative distribution value is bigger than zero, because the region whose left and below points intersects the polygon. On the other hand, the cumulative distribution value of the points in $(x > x_{max}\ AND\ y > y_{max})$ area is 1, for the region whose left and below points intersect the whole polygon.

Secondly, the cumulative distribution values $F(x, y)$ of each grid points (x, y) in area Ω are calculated by using numerical method. The cumulative distribution values of 99 grid points calculated by using numerical method is given in Table 2. This calculation is processed by division of each intersection region with thirty-pixel interval of grid points into sub-triangles. Similar to the analytical method, the cumulative distribution value of the points in $(x < x_{min}\ AND\ y < y_{min})$ area is zero. On the other points, the cumulative distribution value is bigger than zero. Also, the cumulative distribution value of the (250, 200) point in ($x > x_{max}$ AND $y > y_{max}$) area is 0.988. Thus, the cumulative distribution value of the (250, 200) point must be equal to 1 as the analytical method. The difference stems from the numerical method. The difference will be decreased if the number of division of each intersection region into sub-triangles increases in Ω area.

Table 1. The cumulative distribution values calculated by using analytical method

y \ x	0	25	50	75	100	125	150	175	200	225	250
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.028	0.037	0.037
75	0.000	0.000	0.017	0.038	0.053	0.061	0.081	0.119	0.162	0.194	0.195
100	0.000	0.001	0.043	0.097	0.148	0.192	0.246	0.315	0.387	0.444	0.452
125	0.000	0.004	0.071	0.158	0.245	0.326	0.414	0.516	0.616	0.695	0.705
150	0.000	0.004	0.078	0.188	0.310	0.425	0.546	0.678	0.803	0.885	0.895
175	0.000	0.004	0.078	0.189	0.331	0.473	0.618	0.760	0.898	0.980	0.990
200	0.000	0.004	0.078	0.189	0.331	0.477	0.627	0.769	0.908	0.990	1.000

Table 2. The cumulative distribution function values calculated by numerical method

y \ x	0	25	50	75	100	125	150	175	200	225	250
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.027	0.037	0.037
75	0.000	0.000	0.017	0.037	0.052	0.060	0.080	0.117	0.160	0.191	0.193
100	0.000	0.001	0.042	0.095	0.145	0.189	0.243	0.311	0.382	0.439	0.447
125	0.000	0.004	0.069	0.155	0.241	0.320	0.408	0.507	0.607	0.685	0.695
150	0.000	0.004	0.076	0.185	0.305	0.418	0.538	0.669	0.793	0.874	0.884
175	0.000	0.004	0.076	0.186	0.326	0.466	0.610	0.750	0.887	0.968	0.978
200	0.000	0.004	0.076	0.186	0.326	0.470	0.619	0.759	0.897	0.978	0.988

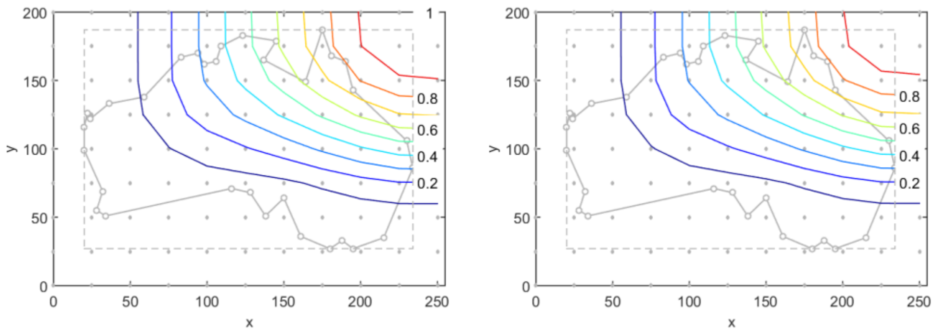
After calculating the cumulative distribution function values both numerically and analytically, the cumulative distribution values calculated numerically are subtracted from the cumulative distribution values calculated analytically to make comparison between each other. The mean relative absolute performance can be calculated as in the following equation.

$$S = \frac{1}{99}\sum_x \sum_y \left(1 - \frac{|F_t(x,y) - F_{n30}(x,y)|}{F_t(x,y)}\right) \tag{20}$$

$F_t(x, y)$ shows the calculated analytically values of the cumulative distribution function, and $F_{n30}(x, y)$ shows the calculated numerically values of triangles which are obtained from the division of each intersection region with thirty-pixel interval of grid points into sub-triangles. As a result, the performance of the numerical method is found as 99.10%. When the intersection region is defined in ten-pixel interval of grid points, the performance of numerical method is found as 99.85%.

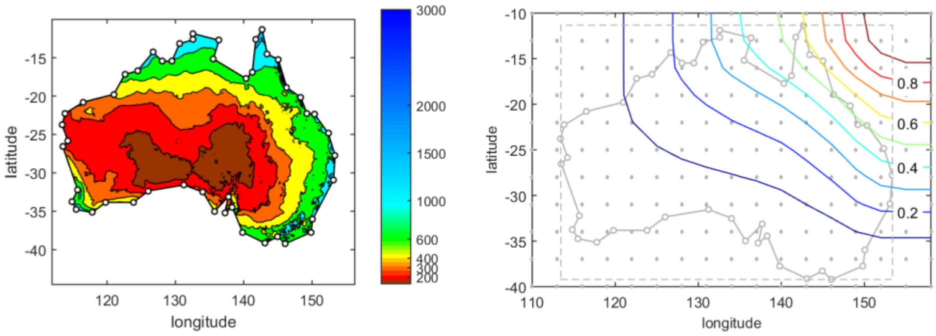
The contour graphic of the cumulative distribution function values obtained analytically in the area Ω is given in Figure 9(a) and the contour graphic of the cumulative distribution function values obtained numerically in the area Ω is given Figure 9(b).

Example 2. Another example is a model based on the annual average amount of rainfall in the Australian mainland (Figure 10(a)). The determination of the probability density function is quite hard because this model is taken from real life. Here, the annual average amount of rainfall data is obtained from [-40, -10] × [110,155] latitude-longitude and 0.05 grid interval. The joint probability density function (JPDF) is determined as a discrete distribution function from the rainfall data (Climate Change in Australia, 2016). The region is divided into small triangles. The constant coefficient of the discrete JPDF is calculated from the total volume of which is summed up volume of all the triangular prism. An analytical method is difficult for calculation of the JCDF. Therefore, the numerical method is calculated and results of the numerical method are given in Figure 10(b).



(a) (b)

Figure 9. The contour graphic of the cumulative distribution values calculated in each grid points in region Ω; **(a)** The contour display of the cumulative distribution function calculated analytically; **(b)** The contour display of the cumulative distribution function calculated numerically.



(a) (b)

Figure 10. The contour graphic of the cumulative distribution values calculated in each grid points in region Ω; **(a)** The contour display of the annual average amount of rainfall as the probability density function **(b)** The contour display of the cumulative distribution function calculated numerically.

5. Conclusion

In this paper, two useful methods are proposed to find the cumulative distribution function when the probability density function is bounded by an arbitrary polygon. For bivariate probability density function which is bounded by an arbitrary polygon has a uniform distribution, the cumulative distribution function can be calculated as a continuous function. However, the calculation of the cumulative distribution function is quite hard when probability density function whose definition range is bounded by a polygon has not a uniform distribution. Analytical and numerical methods are proposed to calculate the cumulative distribution function. These methods can find the cumulative distribution function of a point which is supposed to calculate the cumulative distribution function. Therefore, the cumulative distribution values which cannot be calculated in practice can be calculated quite easily. The cumulative distribution function values are calculated for points which are chosen from the probability density function bounded by the shape of the Australian continent to express these two methods. As the analytical method calculates accurately, the numerical method calculates with regard to the number of triangles. When the number of triangles is increased in the numerical method, the values close to the analytical results are obtained.

On the other hand, another issue is which method will be used firstly. When it is thought that the analytical method gives faultless results in comparison with the numerical method, primarily preferred method must be analytical method. However, the analytical calculation may not be possible in most cases. In this case, numerical methods must be used. Thus, the bivariate piecewise cumulative distribution function can be calculated in an arbitrary area.

References

Badiru, A. and Omitaomu, O., 2010. Handbook of Industrial Engineering Equations, Formulas, and Calculations: CRC Press, 456p.

Boissonnat, J.D. and Teillaud, M., 2007. Effective computational geometry for curves and surfaces: Springer, 344p.

Climate Change in Australia. (2016, 06 December). Retrieved from CSIRO and Bureau of Meteorology, http://www.climatechangeinaustralia.gov.au/.

Douglas, D. and 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.

Eshbach, O., Tapley, B. and Poston, T., 1990. Eshbach's handbook of engineering fundamentals: John Wiley & Sons, 2176p.

Gudmundsson, J., Haverkort, H. and Van Kreveld, M., 2005. Constrained higher order Delaunay triangulations, Computational Geometry, 30 (3), 271-277.

Haines, E., 1994. Point in polygon strategies: In Graphics gems IV: Academic Press, p. 24-26.

Hormann, K. and Agathos, A., 2001. The point in polygon problem for arbitrary polygons, Computational Geometry, 20 (3), 131-144.

Howard, W. and Musto, J., 2008. Engineering Computation: An Introduction Using MATLAB and Excel: McGraw Hill Higher Education, 330p.

Kay, S. M., 2006. Intuitive Probability and Random Processes Using Matlab®, NY: Springer Science & Business Media, 834p.

Kesemen, O. and Doğru, F. Z., 2011. Cumulative Distribution Functions of Two Variable in Polygonal Areas, 7. International Statistics Congress, Antalya, Turkey, p. 150-151.

Kobayashi, H., Mark, B. and Turin, W., 2011. Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance: Cambridge University Press, 812p.

Margalit, A. and Knott, G., 1989. An algorithm for computing the union, intersection or difference of two polygons, Computers & Graphics, 13 (2), 167-183.

Martinez, W. L. and Martinez, A. R., 2002. Computational Statistics Handbook with MATLAB, New York: Crc Press, 731p.

Miller, S. and Childers, D., 2012. Probability and random processes: With applications to signal processing and communications: Academic Press, 611p.

Montgomery, D. and Runger, G., 2010. Applied statistics and probability for engineers: John Wiley & Sons, 784p.

Preparata, F. and Shamos, M., 2012. Computational geometry: an introduction: Springer Science & Business Media, 398p.

Roussas, G., 2003. An introduction to probability and statistical inference: Elsevier, 523p.

Shewchuk, J., 1996. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator: In Applied computational geometry towards geometric engineering: Springer, p. 203-222.

Straszewicz, S., 2014. Mathematical Problems and Puzzles: from the Polish Mathematical Olympiads: Elsevier, 376p.

Walck, C., 2007. Handbook on statistical distributions for experimentalists: University of Stockholm, 202p.