Volume: 09 Number: 01 Category: 02 Page: 65-71 ISSN No: 1303-6319 Cilt: 09 Sayı: 01 Kategori: 02 Sayfa: 65-71

ÇOKGEN ALANLARDA İKİ DEĞİŞKENLİ BİRİKİMLİ DAĞILIM FONKSİYONUNUN BULUNMASI

Orhan KESEMEN * **Fatma Zehra DOĞRU** **

ÖZET

İki değişkenli olasılık yoğunluk fonksiyonundan birikimli dağılım fonksiyonunu hesaplamak için genellikle dikdörtgen alan kullanılır. Ancak uygulamada dikdörtgen olmayan birçok alan mevcuttur. Bu çalışmada, bu alanlar çokgenlerle yaklaşım yapılarak hesaplandı. Hesaplamada iki tür yöntem kullanıldı. İlk yöntem sürekli fonksiyonlar için geliştirildi, ancak bu yöntem yalnızca düzgün dağılım için uygulandı. İkinci yöntem ise ayrık fonksiyonlar için geliştirildi ve herhangi bir olasılık yoğunluk fonksiyonu için kullanılabilir bir yöntemdir.

Anahtar Kelimeler: Birikimli dağılım fonksiyonu, Çokgen tabanlı olasılık yoğunluk fonksiyonu, İki değişkenli dağılımlar, Kapalı bölgede iki değişkenli dağılım fonksiyonları.

1. GİRİŞ

Olasılık fonksiyonları uygulamada genellikle tek değişkenli bir fonksiyon olarak kullanılmasına rağmen, iki değişkenli kullanımları da bulunmaktadır (Martinez vd., 2002). İki değişkenli olasılık fonksiyonlarının uygulama alanları olarak bir göletteki balık popülasyonu, bir şehirdeki kirlilik oranı, bir ormandaki bir ağaç türünün yoğunluğu, bir bölgedeki trafik akışı, bir hava sahasındaki uçuş yoğunluğu, bir bölgedeki yaban hayatın çeşitliliği, bir şehirdeki suç işleme oranı vb. verilebilir. İki değişkenli olasılık yoğunluk fonksiyonları genelde dikdörtgen şekilli veya fonksiyonu bilinen bir alan içerisinde incelenmektedir (Whitt, 1976; Kay, 2006). Ancak gerçek yaşamda olasılık yoğunluk fonksiyonuna temel teşkil eden alanların istenen şekilde olması olasılığı oldukça düşüktür. Bu alanlar değişik geometrik şekillerde olabilmektedir. Bu çalışmada her türlü şekle sahip alanlar, hesaplamada kolaylık olması açısından çokgen yaklaşımıyla gösterilir. Bu gösterimden yola çıkarak elde edilen çokgen alanların (kapalı alan) dağılım fonksiyonlarının hesaplanmasında geometrik yaklaşımlar kullanılmıştır.

2. YÖNTEM

Sınırlı bir bölgede iki değişkenli olasılık yoğunluk fonksiyonu, genelde sınırları belli dörtgen alanlar içerisinde tanımlanmaktadır. Bu iki değişken bağımsız olması durumunda her birinin olasılık yoğunluk fonksiyonlarının çarpımı biçiminde yazıldığı bilinmektedir (Yates vd., 2005). İki değişkenli birikimli dağılım fonksiyonu, iki değişkenin karma olması durumunda çift katlı integralle hesaplanabilmektedir.

*Yrd. Doç. Dr., Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, e-posta: okesemen@gmail.com

**Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, e-posta: fatmazehradogru@gmail.com

Bu integral,

$$F(u,v) = \int_{-\infty}^{v}\int_{-\infty}^{u} f(x,y)\,dx\,dy \quad (1)$$

biçiminde gösterilmektedir. Burada $f(x,y)$ iki değişkenli olasılık yoğunluk fonksiyonudur. Öte yandan bu fonksiyon farklı koordinat sistemlerine çevrilerek farklı geometrik sınırlamalar için hesaplanabilir (Nelsen, 1993).

2.1. Düzgün Dağılımlı Çokgen Alanlarda Dağılım Fonksiyonu

Bu fonksiyonun düzgün dağılıma sahip olduğu varsayıldığında $x \in [a,b]$ ve y $\in [e,f]$ biçiminde sınırlandırılmış bir bölge için birikimli dağılım fonksiyonu,

$$F(u,v) = c\int_{e}^{v}\int_{a}^{u} dx\,dy \quad (2)$$

eşitliğiyle verilir. Ancak alan dikdörtgen şekilli değilse integralin değerini bulmak sorun olacaktır. Olasılık yoğunluk fonksiyonun tanımlı ($f(x,y) > 0$) olduğu bölge N köşeden $\Omega = \{p_i = (x_i, y_i), i = 1,2,\ldots,N\}$ oluşan bir çokgen alan Şekil 1(a) ise bu alan üzerinden integralin değeri,

$$F(u,v) = c\iint_{\Omega} dx\,dy = 1 \quad (3)$$

biçiminde verilir. Burada integralin değerinin ikinci olasılık aksiyomuna ($F(\Omega) = 1$) göre 1'e eşit olması için c ölçek değerinin bulunması gerekir. Kapalı çokgen alanın hesaplanması için çokgen alan yöntemine göre Ω kapalı çokgen alanın değeri,

$$A_{\Omega} = \frac{1}{2}\left|\sum_{i=1}^{N}(x_{i+1}\,y_i - x_i y_{i+1})\right| \quad (4)$$

eşitliğiyle bulunur (Wikipedia, 2011). Burada $i = \{1,$ $i + 1 > N$ biçiminde kullanılır. Bu alan değeri eşitlik (2)'de verilen kapalı integral yerine konur, c değişkeni yalnız bırakılırsa,

$$c = \frac{1}{A_{\Omega}} = \frac{2}{\left|\sum_{i=1}^{N}(x_{i+1}\,y_i - x_i y_{i+1})\right|} \quad (5)$$

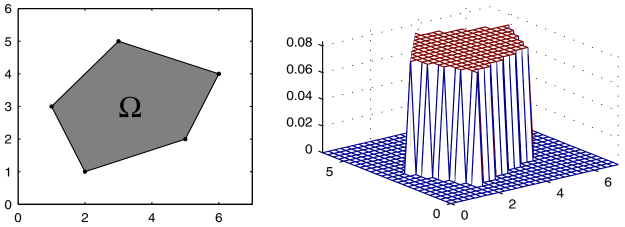
eşitliği elde edilir. Burada,

$$c = f(x,y) = \begin{cases} \frac{1}{A_{\Omega}}, & x,y \in \Omega \\ 0, & x,y \notin \Omega \end{cases} \quad (6)$$

parçalı fonksiyon ile tanımlanmaktadır. Ω kapalı çokgen alanının olasılık yoğunluk fonksiyonun üç boyutlu grafiği Şekil 1(b)'de verilmektedir. (5) eşitliği (2) eşitliğinde yerine konur sınırlar düzenlenirse,

$$F(u,v) = \frac{2}{\left|\sum_{i=1}^{N}(x_{i+1}\, y_i - x_i y_{i+1})\right|} \int_{-\infty}^{v}\int_{-\infty}^{u} dx\, dy \qquad (7)$$

eşitliği elde edilir. Burada, olasılık dağılım fonksiyonu, $y \le v$ ile $x \le u$ ifadeleri ile tanımlanan bölge ile çokgenin kesişiminden oluşan ara kesit bölgesinin (Q) alanı yardımıyla bulunabilir. Ara kesit bölgesi $Q_{u,v} = \{q_j = (\hat{x}_j, \hat{y}_j), j = 1,2,\ldots,M\}$ çokgeni q_i köşe noktaları yardımıyla yeniden tanımlanır.



(a) (b)

Şekil 1. Olasılık yoğunluk fonksiyonları; (a) olasılık yoğunluk fonksiyonun sınırlı olduğu alan; (b) üç boyutlu olasılık yoğunluk fonksiyonu.

Yeni ara kesit bölgesinin dağılım fonksiyonu aşağıdaki eşitlikte gösterildiği gibi Q bölgesinin alanının Ω bölgesinin alanına oranıyla hesaplanır,

$$F_Q(u,v) = \frac{A_Q}{A_\Omega} = \frac{\left|\sum_{j=1}^{M}(\hat{x}_{j+1}\, \hat{y}_j - \hat{x}_j \hat{y}_{j+1})\right|}{\left|\sum_{i=1}^{N}(x_{i+1}\, y_i - x_i y_{i+1})\right|}. \qquad (8)$$

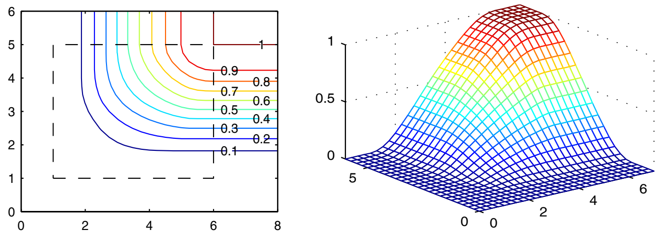
Bu dağılım fonksiyonun değişik bölgelerdeki değerleri,

$$F(u,v) = \begin{cases} 0, & u < x_{min} \bigvee v < y_{min} \\ F_Q(u,v), & x_{min} \le u \le x_{max} \bigwedge y_{min} \le v \le y_{max} \\ 1, & u > x_{max} \bigwedge v > y_{max} \end{cases} \qquad (9)$$

biçiminde verilmektedir. Burada $\wedge$ simgesi mantıksal VE, $\vee$ simgesi ise mantıksal VEYA işlecine karşılık gelmektedir. (9) eşitliğine göre değişim bölgesinin olasılık dağılım fonksiyonu hesaplamak için (7) eşitliğindeki sınırlar aşağıdaki gibi yeniden düzenlenirse,

$$F(u,v) = \frac{2}{\left|\sum_{i=1}^{N}(x_{i+1}\, y_i - x_i y_{i+1})\right|} \int_{y_{min}}^{v}\int_{x_{min}}^{u} dx\, dy, \qquad (10)$$

eşitliği elde edilir. Bu eşitlikte $v \in [y_{min}, y_{max}]$ ve $u \in [x_{min}, x_{max}]$ verilen aralıklarında hesaplanır. Diğer bölgelerin değişimi ise Şekil 2 ve (9) eşitliğinde verildiği gibidir.



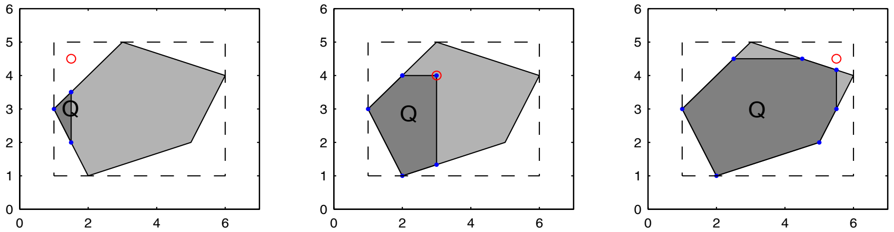
(a) (b)

Şekil 2. Birikimli dağılım fonksiyonları; (a) eş yükselti eğrileri ile birikimli dağılım fonksiyonu; (b) üç boyutlu birikimli dağılım fonksiyonu.

Şekil 2(a)'da verilen eş yükselti grafiği birikimli dağılım fonksiyonun grafiğini göstermektedir. Burada kesikli çizgiyle gösterilen alan hesaplama bölgesi olarak tanımlanmaktadır. Bu bölgenin solunda, altında ve sol-altındaki dağılım fonksiyonu değerleri 0'dır. Üst ve sağ tarafındaki değerler için, hesaplama bölgesine en yakın değer alınarak o noktanın dağılım fonksiyonunun değeri hesaplanabilir. Bu hesaplama bir çeşit değerlerin yinelenmesi tekniği olarak kullanılabilir. Hesaplama bölgesinin sağ-üst köşesinde tüm değerler 1'e eşittir.

2.2. Q Bölgesinin Tanımlanması

Dağılım fonksiyonun istenen aralıktaki her (u,v) noktası için Q bölgesinin yeniden tanımlanması gerekir. Q bölgesi, Ω bölgesinin bir alt bölgesi olduğundan, $x \leq u$ ve $y \leq v$ bölgeleri ile Ω bölgesinin kesişim bölgesi olarak tanımlanır. Q bölgesinin değişik durumlardaki görünümü Şekil 3'te verilmektedir. Buradaki yuvarlak nokta birikimli dağılım fonksiyonun hesaplandığı (u,v) noktasını göstermektedir. Diğer küçük noktalar ise Q bölgesinin köşe konumlarını vermektedir. Yuvarlak noktalar çokgen alan dışındayken bu noktalar Q bölgesinin köşe konumlarına dahil edilmez.



(a) (b) (c)

Şekil 3. Q bölgesinin hesaplanması, (a) kesilen doğrular yalnızca bir noktadan kesilmesi ve (u,v) noktasının çokgenin dışında olduğu durum; (b) (u,v) noktası çokgenin içinde olduğunda; (c) (u,v) noktasının çokgenin dışında olması ve bazı kenarları iki noktadan kesmesi.

Çokgen alan hesaplama yöntemi, çokgenin noktalarının sırasıyla yazılması koşuluyla hesaplanır. Bu sıralama ister saat yönünde, isterse saatin tersi yönünde hesaplanabilir (Beyer, 1987). Bu çalışmada hesaplamalar saatin tersi yönünde gerçekleştirilmiştir. Q bölgesinin köşe noktalarının belirlenmesi için geliştirilen yöntem Algoritma 1'de verilmiştir.

Algoritma 1. Q bölgesinin köşe noktalarının belirlenmesi

Adım 1. Ω bölgesinin i inci noktası Q bölgesinin köşe noktalarına eklenir,

Adım 2. $x = u$ ve $y = v$ doğrularının (p_i, p_{i+1}) doğru parçasını kestiği noktalar bulunur. Eğer,

- Kesilen nokta sayısı bir tane ise, kesişim noktası Q bölgesine köşe diye eklenir.
- Kesilen nokta sayısı iki tane ise, kesişim noktalarından hangisi p_i noktasına yakınsa öncelikle o nokta Q bölgesi köşesi diye eklenir, sonra diğer nokta eklenir.

Adım 3. Son doğru parçasına gelindiyse 5. adıma geçilir. Eğer değilse, Ω bölgesini bir sonraki doğru parçasına geçilerek, 1. adıma gidilir ve işlemlere devam edilir.

Adım 4. Ω bölgesinin tüm köşe noktaları ve bu köşe noktalarının oluşturduğu tüm doğru parçaları ile $x = u$ ve $y = v$ doğrularının tüm kesişim noktaları sırasıyla Q bölgesinin köşe noktaları diye eklenir. Sonra, (u, v) noktası Ω bölgesinin içindeyse bu noktada son olarak Q bölgesine köşe diye eklenir.

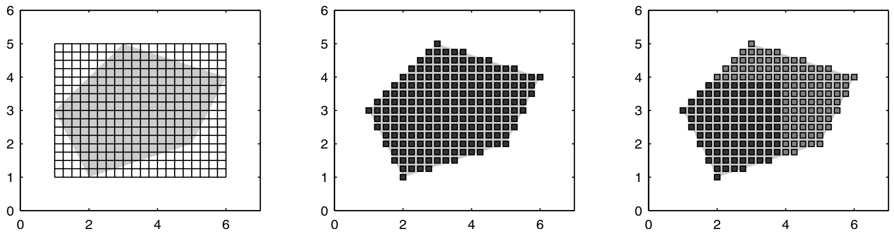
Adım 5. Q bölgesini tanımlayan tüm noktalardan $\hat{x}_j > u$ veya $\hat{y}_j > v$ koşulunu sağlayanlar köşe noktaları dizisinden silinir. Geriye kalan tüm noktalar Q bölgesinin kesin köşe noktalarıdır.

2.3. Çokgen Alanlarda Düzgün Olmayan Dağılım Fonksiyonu

Çokgen içindeki olasılık yoğunluk fonksiyonu düzgün olmayan bir dağılıma sahipse hesaplama için sayısal yöntemler (sonlu elemanlar) kullanılır. Sonlu elemanlar yöntemi için dikdörtgen ızgara düzeneği kullanılarak dağılım fonksiyonu,

$$F(u,v) = (\Delta x \cdot \Delta y) \sum_{i=1}^{x_i \le u} \sum_{j=1}^{y_j \le v} f(x_i, y_j), \qquad \{f(x_i, y_j) \ge 0, \qquad x_i, y_j \in \Omega\} \qquad (11)$$

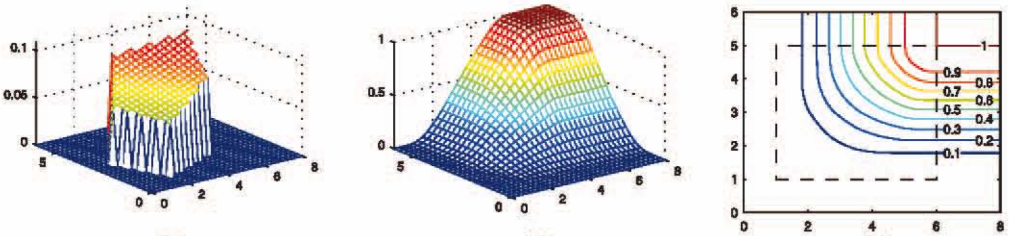
biçiminde hesaplanır. Burada $x_i = x_{min} + i\Delta x$ ve $y_i = y_{min} + j\Delta y$ olarak verilmektedir. Δx ve Δy Şekil 4'teki ızgara çizgileriyle gösterilen alanın x ve y yönündeki birim alan ızgara genişliğidir.



(a) (b) (c)

Şekil 4. Hesaplama bölgesinin ızgaraya bölümlemesi, (a) Ω bölgesinin sonlu elemanlara bölümlemesi; (b) Ω bölgesinin kare biçimli sonlu elemanlara bölümlemesi; (c) Q bölgesi $x < 4$ ve $y < 4$ için hesaplanacak sonlu elemanlar koyu kareler olarak gösterilmiştir.

Şekil 4(c)'de verilen her bir sonlu eleman bir dikdörtgen prizma yaklaşımıyla ele alınır. Prizmada her elemanın orta noktasının olasılık yoğunluk fonksiyonu Şekil 5(a) tüm prizma yüzeyinin olasılık yoğunluk fonksiyonu olarak alınıp prizmanın yüksekliği bulunur. Prizmanın yüksekliği ile prizma hacmi hesaplanıp birikimli dağılım fonksiyonu hesaplanır Şekil 5(b), (c).



(a) (b) (c)

Şekil 5. Çokgen alanlarda düzgün olmayan dağılım fonksiyonu, (a) olasılık yoğunluk fonksiyonu; (b) birikimli dağılım fonksiyonu; (c) birikimli dağılım fonksiyonunun eş yükselti grafiği.

3. SONUÇLAR

Bu çalışmada değişkenlerin sınırlandığı alanın keyfi bir çokgen olduğu zaman dağılım fonksiyonunun bulunması için kullanışlı bir yöntem önerildi. Çokgen içindeki iki değişkenli olasılık yoğunluk fonksiyonun düzgün dağılama sahip durumunda birikimli dağılım fonksiyonu sürekli bir fonksiyon olarak hesaplanabilmektedir. Ancak çokgen alan ile sınırlandırılmış bölgedeki olasılık yoğunluk fonksiyonu düzgün olmayan bir dağılıma sahip olduğunda birikimli dağılım fonksiyonunun hesaplanması için sonlu elemanlar yöntemi önerilmiştir. Bu yöntem herhangi iki rastgele değişkenin belirli bir noktadan itibaren istenen dağılım fonksiyonunu bulmaktadır. Dolayısıyla uygulamada hesaplanamayan dağılımlar bu yöntem ile kolayca elde edilebilmektedir.

4. KAYNAKLAR

Martinez, W. L., Martinez, A. R., 2002. Computational Statistics Handbook with MATLAB, Chapman & Hall/Crc, NY.

Whitt, W., 1976. Bivariate Distributions with Given Marginals, The Annals of Statistics, 6 (4) 1280-1289.

Kay, S. M., 2006. Intuitive Probability and Random Processes Using Matlab®, Springer, NY.

Yates, R. D. and Goodman, D. J., 2005. Probability and Stochastic Processes, John Wiley & Sons, Inc., USA.

Nelsen, R. B., 1993. Some Concepts of Bivariate Symmetry, Journal of Nonparametric Statistics, 3(1), 95-101.

Wikipedia, 2011. Polygon, http://en.wikipedia.org/wiki/Polygon, 3 April 2011.

Beyer, W. H., 1987. CRC Standard Mathematical Tables, 28th ed. Boca Raton, FL: CRC Press, p. 123.

FINDING CUMULATIVE DISTRIBUTION FUNCTIONS OF TWO VARIABLES IN POLYGONAL AREAS

ABSTRACT

In general, rectangular area is used to calculate cumulative distribution function from bivariate probability density function. However, in practice a many areas are available that aren't rectangular. In this study, these areas were calculated by polygons approach. Two types of methods were used in the calculation. First method was developed for continuous functions; however this method was applied only for uniform distribution. The second method was developed for discrete functions and can be used for any probability density function.

Keywords: Cumulative distribution function, Polygonal based probability density function, Bivariate distributions, Bivariate distributions functions over bounded domain.