

Sağlık Sistemlerinde Karma Dağılım Modellerinin Uygulanması

Melik KOYUNCU*¹

¹Çukurova Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Adana

Geliş tarihi: 26.07.2019

Kabul tarihi: 30.09.2019

Öz

Sağlık sistemleri belirsizliğin genellikle fazla olduğu sistemlerden bir tanesidir. Örneğin bir hastanenin polikliniklerine gelen hasta sayısı, günün saatlerine göre değişkenlik gösterir. Hastane servislerinde bir hastanın kalış süresi, hastanın yaşı, cinsiyeti, başka bir hastalığı olup olmadığına bağlı olarak farklılık gösterir. Sağlık sistemlerinde, kapasite, personel vb. gibi planlamalar yapabilmek için, hasta geliş sayıları, hastanelerde kalış süreleri, ameliyat süreleri gibi parametrelerin tahmin edilmesi gerekir. Bu parametreleri tahmin etmek için kullanılan yöntemlere örnek olarak, talep tahmin yöntemleri, istatistiksel dağılım uydurma, makine öğrenmesi gibi yöntemler verilebilir. Bu yöntemlerden istatistiksel dağılım uydurma, belirli bir zaman diliminde toplanan verinin, teorik bir dağılıma uyup uymadığını test eder. Ancak belirsizliğin fazla olduğu durumlarda, yukarıda bahsedilen parametreleri saf bir istatistiksel dağılımla tahmin etmek her zaman mümkün olmayabilir. Bu çalışmada, heterojen yapıya sahip verilerin modellenmesinde kullanılan ve birden fazla dağılımın birleşmesiyle veya iç içe geçmesiyle oluşan karma dağılımların, sağlık sistemlerinde kullanılmasının, veri setine bağlı olarak, belirsizlikleri daha iyi modellediği gösterilmiştir.

Anahtar Kelimeler: Sağlık sistemleri, Stokastik modeller, Karma dağılımlar, Veri analizi

Applications of Mixed Distribution Models in Healthcare Systems

Abstract

Healthcare systems are one of the systems which usually operate under uncertainty. For example, incoming number of patients of a hospital's outpatient clinics varies according to the time of the day. Patient's length of stay at the hospital services differs depending upon the age of patient, gender of patient, having another disease or not. In healthcare systems; in order to be able to do plannings such as capacity or personnel planning; parameters, like number of incoming patients, length of hospital stay, operation time of surgery should be estimated. Some methods to estimate these parameters are forecasting, distribution fitting and machine learning. From these methods, distribution fitting tests whether the data which is collected in a specific time period fits to a theoretical distribution or not. Nevertheless, in the circumstances which has much uncertainty, estimating the aforementioned parameters may always not be possible by a pure statistical distribution. In this study, it has been shown that the use of mixed distributions in health systems, which are used in modeling of heterogeneous data and formed by merging or overlapping multiple distributions, better model uncertainties depending on the data set.

Keywords: Healthcare systems, Stochastic models, Mixed distribution models, Data analysis

*Sorumlu yazar (Corresponding author): Melik KOYUNCU, mkoyuncu@cu.edu.tr

1. GİRİŞ

Sağlık harcamaları ülkelerin gayri safi milli hasılasının (GSYH) önemli bir yüzdesini oluşturmaktadır. Örneğin, 2017 yılında ülkemizdeki sağlık harcamalarının toplamı 140 milyar 647 milyon TL'ye ulaşarak GSYH'nin %4,5'ine denk gelmiştir [1]. Harcamaların bu kadar büyük olduğu bir sistemin iyi yönetilmesi ekonomik bir zorunluluktur. Birçok sektörde, kişilerin taleplerinin zamanında karşılanması müşteri memnuniyeti, ekonomik getiri gibi faktörler için önemli olsa da, sağlık sistemlerinde, ihtiyaç duyan kişilerin taleplerinin vakit kaybedilmeden karşılanması, bazı durumlarda hayati açıdan önemlidir. Hastanede tedaviye ihtiyaç duyan bir hastaya ilgili serviste yer bulunması, acil servise gelen bir hastaya fazla bekletilmeden müdahale edilmesi, poliklinikte muayene olmak isteyen bir kişiye makul bir sürede randevu verilmesi için yeterli miktarda kaynağın olması kadar, bu kaynakların etkili bir biçimde yönetilmesi de gereklidir. Sağlık sistemlerini etkin yönetebilmek için stratejik ve operasyonel kararlar alınırken, planlamacılar genellikle sayısal analiz teknikleri kullanır. Matematiksel modelleme, simülasyon modelleme ve kuyruk teorisi bu tekniklere örnek olarak verilebilir. Bu teknikler kullanılarak sağlık sistemlerindeki problemlerin bir çoğuna çözüm önerisi getirilmiştir. Özellikle simülasyon modelleme, acil servislerde ve yoğun bakım ünitelerinde optimal yatak sayısının belirlenmesi, optimal sağlık personeli sayısının belirlenmesi vb. problemlerinin çözümü için çok değişik durum ve senaryolar için kullanılmıştır. Simülasyon modellemeyi kullanabilmek için sisteme ait bazı parametrelerin tahmin edilmesi gerekir. Örneğin bir hastanenin optimal yatak sayısını belirlemek isteniyorsa, hastaneye gelen hastaların geliş sıklığı ve hastaların hastanede kalış sürelerine ait zamana bağlı bir tahmin yapılmalıdır. Modellerin iyi sonuç vermesi, modellerde kullanılan parametrelerin gerçekçi tahmin edilmesine bağlıdır. Ancak sağlık sistemleri doğası gereği belirsizliğin fazla olduğu sistemlerden bir tanesidir. Bir polikliniğe gelen hasta sayısı günün saatine bağlı olarak değişkenlik gösterebilir. Hastane servislerinde bir hastanın yatış süresi hastanın yaşına, cinsiyetine, hastalığının türüne,

eşlik eden başka bir hastalık olup olmadığı gibi faktörlere bağlı olarak değişkenlik gösterir. Ameliyat süreleri, muayene süreleri, tedavi süreleri hastadan hastaya farklılık gösteren parametrelerdir. Bu parametreler genellikle, istatistiksel dağılım uydurma, talep tahmin yöntemleri ve son yıllarda makina öğrenmesi teknikleri ile tahmin edilmektedir.

2. ÖNCEKİ ÇALIŞMALAR

Sağlık sistemlerinde simülasyon modelleme çok geniş bir literatüre sahiptir. Günel ve Pidd [2] yaptıkları çalışmada, sağlık sistemlerinde simülasyon modellerinin kullanımına ait 80'e yakın çalışmanın değerlendirmesini yapmışlardır. Benzer şekilde Gül ve Güneri [3] tarafından yapılan acil servislerde simülasyon modelinin kullanılmasına yönelik literatür değerlendirmesinde, 100'e yakın çalışma bulunmaktadır. Bai ve arkadaşları [4], yoğun bakım ünitelerinin yönetiminde yöneylem araştırması adlı literatür değerlendirmesinin, önemli bir kısmını simülasyon çalışmaları oluşturmaktadır. Bu alanda yapılan çalışmalar incelendiğinde, birçok araştırmacının simülasyon modeline girdi parametresi olarak çoğunlukla teorik dağılımlar kullandığı görülmektedir. Bir veri setinin normal, lognormal, gamma dağılımı gibi teorik bir dağılıma uyup uymadığının belirlenmesi için Ki-kare testi ve Kolmogorov Smirnov testleri gibi uyum iyiliği testleri yapılmaktadır [5]. Ancak veri setinin herhangi bir teorik dağılıma uymadığı durumda çok az araştırmacı sağlık sistemlerinin analizi için karma dağılım kullanmıştır. Bu tür durumlarda, araştırmacılar genellikle ampirik dağılımı tercih etmişlerdir. Ampirik dağılımlar, teorik dağılımlara göre daha az bilgi içermektedir ve toplanan veri aralığı dışında veri üretilmesine imkan vermemektedir. Karma dağılım modelleri ilk olarak Pearson tarafından önerilmiştir. Günümüzde karma dağılım modelleri biyoloji, genetik, sağlık ve mühendislik gibi birçok alanda uygulama alanı bulmuştur [6]. Ancak sağlık sistemlerinin operasyonel planlamasında, karma dağılımların kullanımına çok sık rastlanılmamaktadır. Ickowicz ve Sparks [7] hastanede toplam kalış süresini karma dağılımla modellemiştir. Sağlık sistemlerinde kalış süresi bir hastanın sağlık sistemine girdiği

zaman ile çıktığı zaman arasındaki fark olarak tanımlanabilir. Kalış süresinin tahmini birçok modelde, girdi parametresi olarak kullanıldığı için son derece önemlidir. Sorensen [8] tarafından geliştirilen modelde kalış süresinin normal dağılıma uyduğu varsayılmıştır. Ancak kalış süresi ile ilgili bu varsayımı, birçok veri setinde, uyum iyiliği testlerini kullanarak, istatistiksel olarak kabul etmemiz mümkün görünmemektedir. Bu çalışmada yoğun bakım ünitelerinden, beyin cerrahi yoğun bakım ünitesindeki kalış süresinin karma dağılımla modellenmesine yönelik bir uygulama yapılmıştır. Karma dağılım modelinin seçilmesi bu üniteye hasta kalış süresinin hiçbir teorik istatistiksel dağılıma uymamasıdır.

3. METOD

İki veya daha fazla bileşenden meydana gelen dağılımlar karma dağılımlar olarak tanımlanır. Karma dağılım modellerini aşağıdaki gibi tanımlayabiliriz.

X: S örneklem uzayındaki değerleri alan bir rasgele değişken

$f(x)$: X rasgele değişkeninin olasılık yoğunluk fonksiyonu olsun.

Eğer X rasgele değişkeni bir karma dağılım modeline sahipse, X rasgele değişkeninin olasılık yoğunluk fonksiyonu; (Eşitlik 1)

$$f(x) = \pi_1 f_1(x) + \dots + \pi_k f_k(x) \quad x \in S \quad (1)$$

(1) no'lu denklemden $\pi_1, \pi_2, \dots, \pi_k$ karma dağılım modelindeki her bir dağılımın karma oranını $f_1(\cdot), f_2(\cdot), \dots, f_k(\cdot)$, ise karma dağılım modelinde yer alan dağılımların olasılık yoğunluk fonksiyonunu göstermektedir.

Bu durumda, Eşitlik 2-4 yazılabilir.

$$0 \leq \pi_i \leq 1 \quad i=1 \dots k \quad (2)$$

$$\sum_{i=1}^k \pi_i = 1 \quad (3)$$

$$\int_S f_i(\cdot) dx = 1 \quad i=1 \dots k \quad (4)$$

(1) no'lu denklemden yer alan $f_1(x), f_2(x), \dots, f_n(x)$ karma dağılım modelinde yer alan dağılımların olasılık yoğunluk fonksiyonlarının parametrelerini de gösterebilmek için (1) no'lu denklem;

θ_i : i. dağılımın parametresi $i=1, \dots, n$ olmak üzere; (Eşitlik 5)

$$f(x) = \pi_1 f_1(x; \theta_1) + \dots + \pi_n f_n(x; \theta_n) \quad (5)$$

Eşitlik 5'teki gibi yazılabilir.

Karma dağılımları, içerdiği dağılımların parametreleri ile göstermek için genellikle Ψ sembolü kullanılır ve karma dağılımlar Eşitlik 6'daki gibi de gösterilebilir.

$$f(x, \Psi) = \sum_{i=1}^n \pi_i f(x; \theta_i) \quad (6)$$

Eğer μ ortalamalı, σ standart sapmalı bir normal dağılımı;

$N \sim (\mu, \sigma)$ biçiminde gösterecek olursak;

Üç bileşen ve tek değişkene sahip normal dağılımlardan oluşan bir karma dağılım modeli; (Eşitlik 7 ve 8)

μ_i : i. dağılımın populasyon ortalaması

σ_i : i. dağılımın populasyon standart sapması

π : 3,14159...

e: 2,71... olmak üzere;

$$f(x; \Psi) = \pi_1 f(x; \theta_1) + \dots + \pi_3 f(x; \theta_3)$$

$$f(x; \Psi) = \pi_1 f(x; \mu_1, \sigma_1) + \dots \quad (7)$$

$$+ \pi_3 f(x; \mu_3, \sigma_3) \quad (8)$$

biçiminde yazılabilir.

$$\pi = (\pi_1, \pi_2, \pi_3), \theta_1 = (\mu_1, \sigma_1), \theta_2 = (\mu_2, \sigma_2), \theta_3 = (\mu_3, \sigma_3)$$

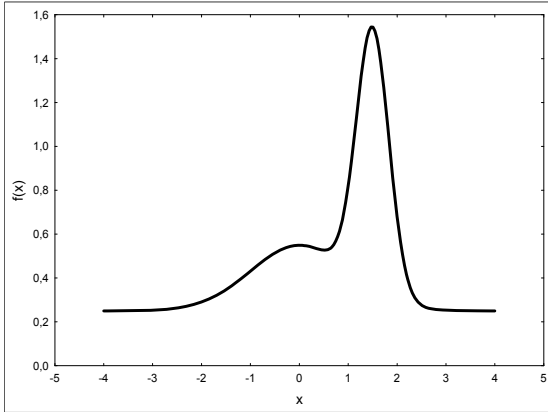
Tek değişkenli normal dağılımın olasılık yoğunluk fonksiyonu Eşitlik 9'da gösterilmiştir.

$$f(x; \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (9)$$

$$-\infty < x < \infty \quad \text{ve} \quad -\infty < \sigma_i < \infty$$

Tek değişkenli, iki tane normal dağılımdan oluşan bir karma dağılım modeline örnek olarak;

$f(x;\Psi)=\frac{3}{4}f(x;0,1)+\frac{1}{4}f(x;\frac{3}{2},\frac{1}{9})$ yazarsak, bu karma dağılım modelinin olasılık yoğunluk fonksiyonunun $-4 \leq x \leq 4$ aralığındaki grafiği Şekil 1'deki gibi olacaktır.



Şekil 1. Örnek karma dağılım modeli grafiği

Analiz edilen bir veri setinin karma dağılımla modellenebilmesi için;

- Bileşen sayısının tahmin edilmesi
- Karma oranlarının tahmin edilmesi
- Dağılım parametrelerinin tahmin edilmesi gerekmektedir.

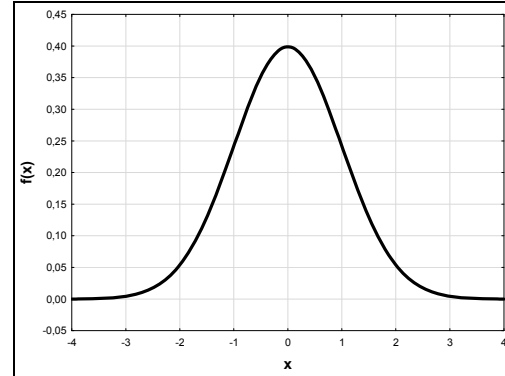
3.1. Karma Dağılım Modellerinde Bileşen Sayısının Tahmininde Kullanılan Grafikselle Yöntemler

Grafikselle yöntemlerin karma dağılımların analizinde kullanılmasının iki temel amacı vardır. Bunlar;

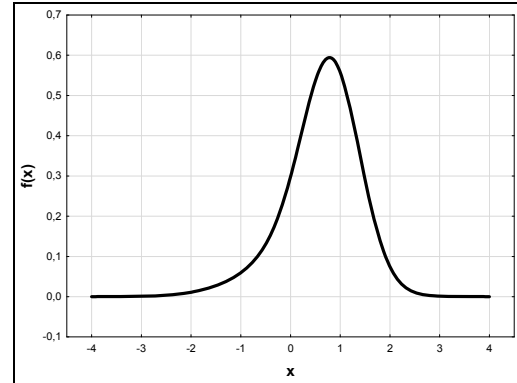
- İncelenen veri setinde karma dağılım olup olmadığının belirlenmesi, eğer karma dağılım varsa türü hakkında ön fikir elde edilmesi,
- En azından karma dağılım parametrelerinin ham bir tahminin yapılmasıdır [9].

Karma dağılımların analizinde kullanılan birden fazla grafikselle yöntem vardır. Grafikselle yöntemler incelenen veri setinin kaç tane bileşenden oluştuğunu göstermesi açısından başlangıçta araştırmacılara iyi bir fikir verebilmektedir [10]. İncelenen veri setinin grafiği çizilerek aşağıda örnekleri verilen karma dağılım modellerine benzerlikleri araştırmacılara bileşen sayısı hakkında

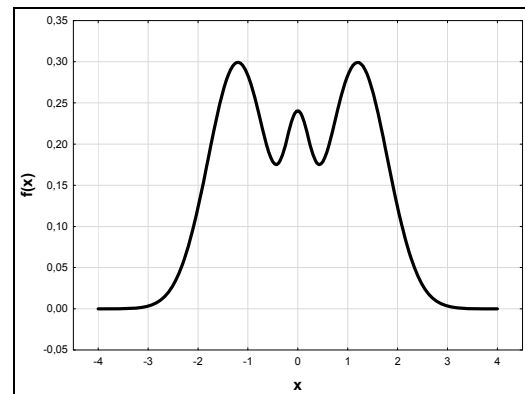
bir başlangıç noktası sağlayacaktır. Şekil 2'de saf bir normal dağılıma ait, Şekil 3-5 arasında ise bazı karma dağılım modellerine ait olasılık yoğunluk fonksiyonları gösterilmiştir. Çizelge 1'de ise bu modellerin olasılık yoğunluk fonksiyonlarının formülü ile karma oranları verilmiştir.



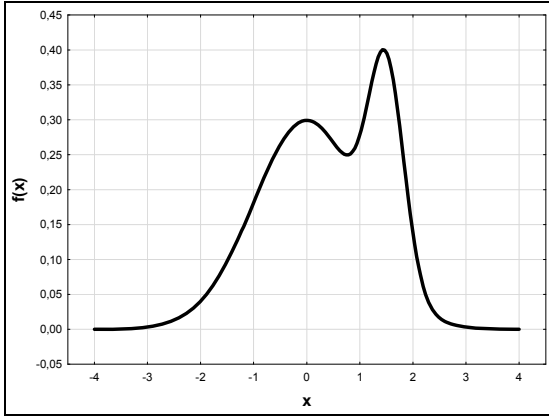
Şekil 2. Karma dağılım yok



Şekil 3. Sola çarpık karma dağılım modeli



Şekil 4. Üç tepeli karma dağılım modeli



Şekil 5. Asimetrik çift tepeli karma dağılım modeli

Çizelge 1. Şekil 2-5'de çizilen dağılımların olasılık yoğunluk fonksiyonları

Şekil No	Olasılık Yoğunluk Fonksiyonu f(x)
2	$N(0,1)$
3	$\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2}, \frac{2}{3}\right) + \frac{3}{5}N\left(\frac{13}{15}, \frac{5}{9}\right)$
4	$\frac{9}{20}N\left(-\frac{6}{5}, \frac{3}{5}\right) + \frac{9}{20}N\left(\frac{6}{5}, \frac{3}{5}\right) + \frac{1}{10}N\left(0, \frac{1}{4}\right)$
5	$\frac{3}{4}N(0,1) + \frac{1}{4}N\left(\frac{3}{2}, \frac{1}{3}\right)$

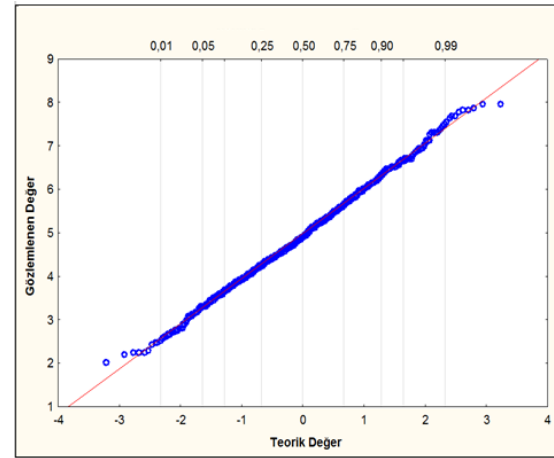
3.2. Bileşen Sayısının Belirlenmesinde Histogram ve Olasılık Yoğunluk Fonksiyonu

Karma dağılım modelinde bileşen sayısını belirlemek için kullanılan grafiksel yöntemlerden bir tanesi incelediğimiz verinin histogramının çizilmesidir. Histogram çizilirken aralık sayısı ve genişliğinin belirlenmesi son derece önemlidir. Sezgisel olarak n tane gözlemin olduğu bir veride aralık sayısı \sqrt{n} olarak belirlenebilir [11]. Ancak araştırmacı aralık sayısını verinin yapısına göre değiştirebilir. Histogram çizilmesinin amacı olasılık yoğunluk fonksiyonunu tahmin etmek ve buna bağlı olarak bileşen sayısını hakkında ön fikir sahibi olmaktır. Ancak verinin histogramı incelenerek bileşen sayısının kesin olarak tahmin edilmesi garanti bir yöntem değildir [12].

3.3. Birikimli Dağılım Fonksiyonuna Dayalı Yöntemler

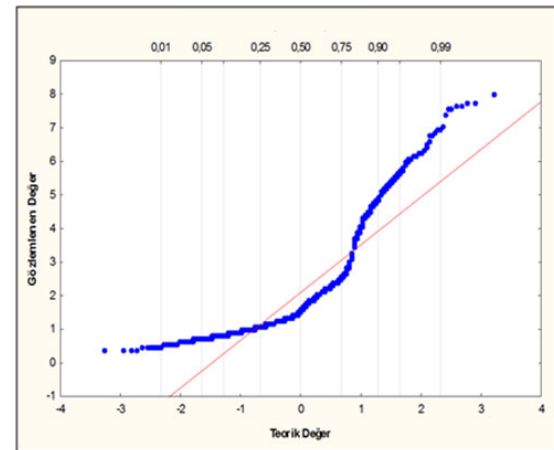
Karma dağılımın varlığını araştırmak için yoğunluk fonksiyonunun tahminine alternatif olarak Q-Q ve

P-P grafiklerinin çizimi kullanılabilir. Q-Q çizimi, $0 < p < 1$ aralığında $F^{-1}(p)$ nin $\Phi^{-1}(p)$ 'e karşı çizilmesidir. Burada, $F(\cdot)$ karma dağılımın, $\Phi(\cdot)$ ise standart normal dağılımın birikimli dağılım fonksiyonudur. Eğer incelenen veri setinde karma dağılım mevcut değilse, Q-Q çizimi doğrusal olacaktır ancak normal dağılımların karmasından meydana gelmişse Q-Q çizimi doğrusal olmayacaktır.



Şekil 6. Normal dağılım Q-Q grafiği

Şekil 6 ve Şekil 7'de gösterilen Q-Q çizimleri, verilerin dağılımı hakkında ön fikir vermektedir.



Şekil 7. Örnek bir karma dağılım Q-Q grafiği

Karma dağılımlarda bileşen sayısının belirlenmesinde, Q-Q grafiğine alternatif olarak

P-P (yüzde-yüzde) grafiği Fowlkes (1979) tarafından önerilmiştir [13].

n: örneklem büyüklüğü

x_i: i. örneklemin değeri $i=1, \dots, n$

$$p_i = (i - \frac{1}{2}) / n \quad i=1, \dots, n$$

\bar{x} : örneklem ortalaması

s: örneklem standart sapması olmak üzere;

$x_1 \leq x_2 \leq \dots \leq x_n$ sıralı örnekleminde

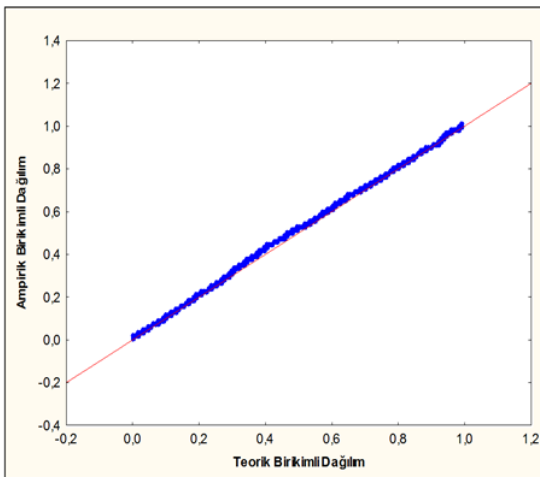
$(x_i - \bar{x})/s$ değerine karşı $\Phi\left(\frac{x_i - \bar{x}}{s}\right) - p_i$ 'nin grafiği; Φ -P'ye karşılık Q'nun grafiğine bir örnek teşkil edecektir. Bu çizim populasyon parametreleri üzerinden yazılacak olursa;

μ : populasyon ortalaması

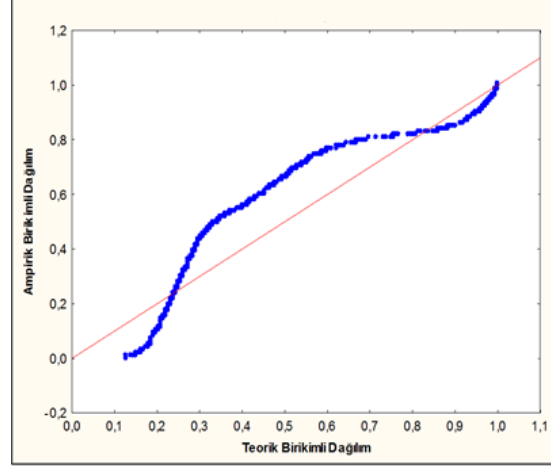
σ : populasyon standart sapması olmak üzere;

$\Phi\left(\frac{x - \mu}{\sigma}\right) - F(x)$ 'in $(x - \mu)/\sigma$ ya karşı çizimi olacaktır.

İncelediğimiz veri seti normal dağılıma sahipse Fowlkes çizimi düz bir çizgi olacaktır. Ancak normal dağılımlardan oluşan karma bir dağılım ise bu çizim düz bir çizgiye sahip olmayacaktır. Şekil 8 ve Şekil 9'da Fowlkes çizimleri gösterilmiştir. Bu çizimler parametre tahmini yapmamasına rağmen ilk analiz için önemlidir [9].



Şekil 8. Normal dağılım P-P grafiği



Şekil 9. Örnek bir karma dağılım P-P grafiği

3.4. Karma Dağılım Modellerinde Bileşen Sayısının Belirlenmesinde Modele Dayalı Yöntemler

i. Olabilirlik Kestirim Oranı Test İstatistiği

$X_i = x_i, i=1, \dots, n$ gözlemlerinden oluşan bir örneklem uzayında X_i 'nin olasılık yoğunluk fonksiyonu Eşitlik 6'da gösterilmiştir.

Bu karma dağılım modelinin olabilirlik kestirim (likelihood) fonksiyonu; (Eşitlik 10)

$$L(\Psi) = \prod_{i=1}^n f(x_i, \Psi) = \prod_{i=1}^n \left[\sum_{j=1}^k \pi_j f(x_i, \theta_j) \right] \quad (10)$$

logaritması alınmış olabilirlik kestirim fonksiyonu ise; (Eşitlik 11)

$$\log L(\Psi) = \sum_{i=1}^n \log(f(x_i; \Psi)) = \sum_{i=1}^n \log\left(\sum_{j=1}^k \pi_j f_i(x_i, \theta_j) \right) \quad (11)$$

$i = 1, \dots, n, j = 1, \dots, k$ biçimindedir. Olabilirlik kestirim oran testi, farklı karma dağılım modelleri arasındaki uyum iyiliğini gösteren istatistiksel bir testtir [14].

Karma dağılımdaki bileşen sayısını belirlemek için başlangıçta k_0 ve k_1 ($k_1 > k_0$) sayıları belirlenerek iterasyona başlanır. Her bir iterasyonda ; $k_1 = k_0 + 1$ alınarak logaritmik olabilirlik kestirim fonksiyonunun değerindeki artış miktarının, azalmaya başladığı noktaya kadar iterasyona devam edilir. Bu noktada aşağıdaki hipotez test edilir.

$$H_0 : k = k_0$$

$$H_1 : k = k_1$$

Karma dağılım modelindeki bileşen sayısı k 'yı bulabilmek için olabilirlik kestirim oranı test istatistiği, $\chi^2_{v_1-v_0}$ (Ki-kare) dağılımına yakınsayan $-2\log\Lambda$ dir [15].

$\hat{\Psi}_0$: k_0 bileşen sayısına karşı gelen olabilirlik kestirim fonksiyonunun maksimum değeri

$\hat{\Psi}_1$: k_1 bileşen sayısına karşı gelen olabilirlik kestirim fonksiyonunun maksimum değeri olmak üzere; (Eşitlik 12 ve 13)

$$-2\log\Lambda = -2\log \left[\frac{\max_{\Psi \in \Psi_0} L(\Psi)}{\max_{\Psi \in \Psi_1} L(\Psi)} \right] \quad (12)$$

Buradan;

$$-2\log\Lambda = -2\{\log\hat{\Psi}_0 - \log\hat{\Psi}_1\} \quad (13)$$

elde edilir. Olabilirlik kestirim oran testinde Λ yeterince küçük olduğunda, $-2\log\Lambda$ değeri büyük olacaktır. Bu durumda, H_0 hipotezini reddetmek için yeterli kanıt olmayacaktır. Bu testteki amaç iki karma dağılım modelinin olabilirlik kestirim fonksiyonu değerleri arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemektir. χ^2 dağılım tablosu kullanılarak test istatistiğinin kritik değeri belirlenir. Test istatistiğinin serbestlik derecesi v_1 ve v_0 , sırasıyla Ψ_1 ve Ψ_0 'ın boyutuna eşittir.

ii. Akaike Bilgi Kriteri (AIC)

Tahmin edilen bileşen sayısı k olan bir karma dağılım modeli için $L(\Psi)$ olabilirlik fonksiyonunun maksimum değeri olsun. Bu durumda Akaike Bilgi Kriteri (AIC);

$AIC = 2k - 2\ln L(\Psi)$ ile hesaplanır. Bu yöntem 1974 yılında Akaike tarafından ileri sürülmüştür. Farklı sayıdaki k bileşenleri için AIC değeri en küçük olan değer, k bileşen sayısı için seçilir [12].

iii. Bayesci Bilgi Kriteri (BIC)

Bileşen sayısının tahmininde kullanılan diğer bir bilgi kriteri ise Schwarz tarafından önerilen Bayesci Bilgi Kriteri (BIC) dir [12].

k bileşen sayısı , n gözlem sayısı olmak üzere;

$BIC = -\ln(n)k - 2\ln L(\Psi)$ formülü ile hesaplanır. Farklı sayıdaki k değerleri için BIC değeri en küçük olan k değeri seçilir.

Yukarda açıklanan her üç yöntemde, karma dağılım modellerinde bileşen sayısını tahmin etmek için kullanılır. Koehler ve Murhee, çalışmalarında AIC'nin bileşen sayısını belirlemede tutarsız davrandığını belirtmişlerdir [16]. Roeder ve Wasserman, BIC'nin normal karma dağılım modelinin bileşen sayısını tutarlı bir biçimde tahmin ettiğini iddia etmişlerdir [17]. Karma dağılım modellerinde bileşen sayısının tahmininde her üç yöntem kullanıldığında, olabilirlik fonksiyonu değerinin en büyük, AIC ve BIC değerlerinin en küçük olduğu k değeri bileşen sayısı olarak seçilmelidir.

3.5. Karma Dağılım Modellerinde Parametre Tahmini

Karma dağılım modellerinde parametre tahmini için araştırmacılar tarafından çok sayıda yöntem önerilmiştir. Tittirengton, bu durumunun nedenini, parametre tahmini için kesin formüllerin olmamasına bağlamıştır [18]. Karma dağılımlarda parametre tahmini için kullanılan yöntemler genel olarak, grafiksel yöntemler, momentler yöntemi, en küçük mesafe yöntemi, en çok olabilirlik yöntemi ve Bayesci yaklaşımlar olarak gruplara ayrılabilir.

3.5.1. En Çok Olabilirlik Tahmin Yöntemi

X_1, \dots, X_n örnekleminin, k tane bileşenden oluşan $f(\cdot, \Psi)$, $\Psi \in \Omega$ karma dağılım modelinden seçildiğini varsayalım. Gözlemlenen x_1, \dots, x_n değerleri için, Ω

parametre kümesi için $L(\Psi)$ 'yi en büyük yapan $\theta(x_1, \dots, x_n)$ değerine, θ 'nın en çok olabilirlik tahmini denir. Ancak matematiksel olarak logaritması alınan olabilirlik fonksiyonunun en büyük değerini bulmak kolay olduğundan, Ω parametre kümesi için, $\max = \ln(L(\Psi))$ modeli çözülür. Denklemin her parametreye göre I. türevi alınır ve sıfıra eşitlenir. Elde edilen denklem sistemi, ele alınan dağılımlara ve parametrelerine göre çok karmaşık olabilir. Bu nedenle, bu denklem sistemini çözmek için doğrusal olmayan optimizasyon yöntemleri kullanılır [14].

3.5.2. Beklenti Maksimizasyonu Algoritması

Beklenti maksimizasyonu algoritması Dempster ve arkadaşları tarafından 1977 yılında önerilmiştir. Elimizde $x = \{x_1, \dots, x_n\}$ veri kümesinin olduğunu varsayalım. Başlangıçta, x_i gözleminin hangi dağılıma ait olduğu bilinmediği için x tamamlanmamış veri kümesi olarak adlandırılır. Beklenti maksimizasyonu algoritmasının amacı $L(\Psi) = f(x, \Psi)$ fonksiyonunu maksimum yapan $\Psi = \hat{\Psi}$ tahmininin bulunmasıdır [12]. Algoritma tamamlandığında x tamamlanmamış veri kümesindeki her bir gözlemin hangi kümeye ait olduğu belirlenir.

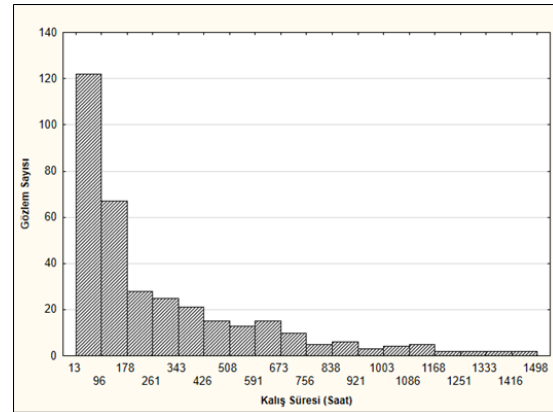
4. UYGULAMA

Çalışmanın uygulama kısmında, üçüncü basamak yoğun bakım ünitelerinden, beyin cerrahi ünitesindeki hasta kalış sürelerinin tahmini, karma dağılım modelleri kullanılarak yapılmıştır. Uygulama olarak yoğun bakım ünitelerinin seçim nedeni, bu ünitelerin hastanelerin en kritik bölümlerinden birisi olmasıdır. Tanım olarak yoğun bakımlar bir yada daha fazla organ sistemlerinde ciddi işlev bozukluğu nedeniyle yoğun bakım gereksinimi olan hastaların iyileştirilmesini amaçlayan ve hasta bakımı açısından ayrıcalık taşıyan, ileri teknolojiye sahip cihazlarla donatılmış, 24 saat yaşamsal göstergelerin gözlemi ve hasta tedavisinin yapıldığı kliniklerdir. Hastanelerde bir çok ünite olmasına rağmen, Yoğun Bakım Üniteleri, hastanelerin, sağlık açısından önemli risk taşıyan hastalara hizmet veren ünitelerinin başında gelmektedir. Yoğun bakım

ünitelerinde yeterli sayıda yatak olmaması durumunda bu üniteye ihtiyaç duyan hastalar için telafisi mümkün olmayan sonuçlara, bu ünitelerde fazla kapasite olması durumunda ise yüksek işletme maliyetleri nedeniyle finansal sıkıntıların ortaya çıkmasına neden olmaktadır [19]. Ortalama olarak hastane yatak sayısının %10'undan daha az bir orana sahip olan yoğun bakım üniteleri, işletme maliyeti açısından hastane bütçelerinin yaklaşık %20-25'ini oluşturmaktadır [20,21].

4.1. Veri Toplanması ve Analizi

Çalışmada kullanılan veri, Çukurova Üniversitesi Tıp Fakültesi Balcalı Hastanesi otomasyon sisteminden alınmıştır. 2014-2015 yılı için yoğun bakım ünitelerinde kalan hastaların, bu üniteye kabul zamanı ile çıkış zamanı arasındaki farklar analiz edilmiştir. Şekil 10'da çizilen Histogram Beyin Cerrahi yoğun bakımında yatan hasta kalış sürelerine ait histogramdır. Kalış süresinin herhangi bir teorik dağılıma uyup uymadığını belirlemek için χ^2 (ki-kare) uyum iyiliği testi yapılmış ve kalış sürelerinin herhangi bir istatistiksel dağılıma uymadığı gözlemlenmiştir. Çizelge 2'de sonuçlar gösterilmiştir.



Şekil 10. Beyin cerrahi yoğun bakım ünitesi hasta kalış süreleri

Bu nedenle beyin cerrahi yoğun bakım ünitesinde hasta kalışlarını tahmini için karma dağılım modelleri kullanılmıştır. Karma normal dağılım ve karma gamma dağılımlara ait bileşen sayıları ve dağılım parametreleri beklenti maksimizasyonu

algoritması ile tahmin edilmiştir. Çizelge 3’de AIC, BIC ve loglikelihood değerleri gösterilmiştir. Yazılım olarak R programı kullanılmıştır.

Çizelge 2. χ^2 (ki-kare) uyum iyiliği testi sonuçları

Dağılım	p değeri
Normal	0,00
Gamma	0,05
Lognormal	0,00
Üstel	0,00
Üçgen	0,00

Çizelge 3. Kalış Sürelerine ait BIC, AIC ve loglikelihood değerleri

Bileşen sayısı	BIC	AIC	Loglikelihood Normal	Loglikelihood Gamma
k=2	4707,89	4970	-2290,24	-2339,30
k=3	4676,78	4972,79	-2296,34	-2314,88
k=4	4649,01	4974,79	-2287,43	-2292,22
k=5	4616,80	4976,79	-2299,44	-2287,02
k=6	4624,60	4978,79	-2291,25	-2262,02
k=7	4640,49	4980,79	-2286,62	-2254,16

Çizelge 3’de gösterilen, loglikelihood, AIC ve BIC değerlerine göre, bileşen sayılarının tahmini için $k=2$ ve $k=3$ değerleri seçilebilir. Beklenti maksimizasyonu algoritmasının sonuçlarına göre, iki ve üç bileşenli karma normal dağılım denklemleri; $f(x;\Psi)=\pi_1f(x;\theta_1)+\dots+\pi_n f(x;\theta_3)$ gösterimini kullanarak $k=2$ için normal dağılımlardan oluşan olasılık yoğunluk fonksiyonu; (Eşitlik 14)

$$f(x;\Psi)=0,51f(x;\theta_1)+0,49f(x;\theta_2) \quad (14)$$

$$\theta_1=(81,33, 50,83) \theta_2=(501,82, 322,48)$$

Benzer şekilde, üç bileşenli normal dağılım karması için; (Eşitlik 15)

$$f(x;\Psi)=0,28f(x;\theta_1)+0,45f(x;\theta_2)+0,27f(x;\theta_3) \quad (15)$$

$$\theta_1=(261,97, 117,68) \theta_2=(70,55, 42,48) \\ \theta_3=(698,62, 319,67) \text{ biçiminde yazılabilir.}$$

Gamma dağılımdan oluşan karma dağılım modellerini de; $k=2$ bileşen için; (Eşitlik 16)

$$f(x;\Psi)=0,45f(x;\theta_1)+0,55f(x;\theta_2) \quad (16) \\ \theta_1=(35,02, 2,15) \theta_2=(245,13, 1,85)$$

Üç bileşenli gamma dağılımından oluşan karma dağılım modeli ise; (Eşitlik 17)

$$f(x;\Psi)=0,02f(x;\theta_1)+0,54f(x;\theta_2)+0,46f(x;\theta_3) \quad (17)$$

$$\theta_1=(0,04, 303,20) \theta_2=(31,60, 2,65)$$

$$\theta_3=(197,99, 3,30) \text{ biçiminde yazılabilir.}$$

4.2. Ampirik Dağılımla Karma Dağılımların Karşılaştırılması

Karma dağılım modeli ile ampirik dağılımı karşılaştırmak için gerçek veri ile ampirik dağılımdan üretilmiş verinin histogramı Şekil 11’de gösterilmiştir. Benzer şekilde parametreleri Eşitlik 17’de verilen üç tane gamma dağılımından oluşan karma dağılım modelinden üretilen sayılarla, gerçek verinin histogramları Şekil 12’de gösterilmiştir. Görsel olarak karma dağılım modelinden üretilen sayıların gerçek veri’yi daha iyi temsil ettiği gözlemlenmektedir. Karma dağılım modelinin bu veri seti için daha iyi sonuç verdiğini göstermek için parametrik olmayan istatistiksel testlerden Mann Whitney U testi yapılmıştır. Bu test için;

H_0 : Ampirik dağılımdan üretilen değerler ile gerçek değerler aynı dağılımdan gelmektedir.

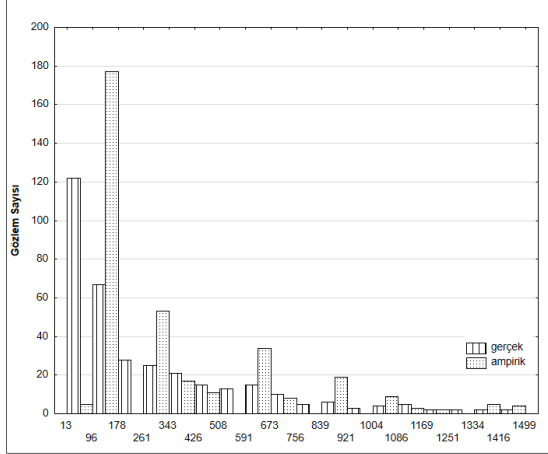
H_1 : Ampirik dağılımdan üretilen değerler ile gerçek değerler aynı dağılımdan gelmemektedir.

p değeri çok küçük bir değer olduğu için H_0 hipotezi red edilir. Aynı hipotezler karma dağılımdan üretilmiş için sayılar için yapılırsa;

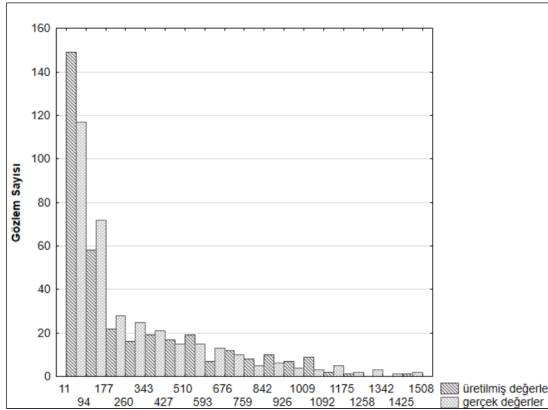
H_0 : Karma dağılımdan üretilen değerler ile gerçek değerler aynı dağılımdan gelmektedir.

H_1 : Karma dağılımdan üretilen değerler ile gerçek değerler aynı dağılımdan gelmemektedir.

p değeri 0,31 olduğu için, H_0 hipotezini reddetmek için seçilen önem düzeyine bağlı olarak, güçlü bir kanıt bulunmamaktadır. Bu sonuç ele alınan veri setini, karma dağılım modelinin, $\approx \sqrt{n}$ kategori aralığına göre oluşturulmuş ampirik dağılıma göre, daha iyi temsil ettiğini göstermektedir.



Şekil 11. Gerçek değerler ile amprik dağılımdan üretilmiş değerlerin karşılaştırılması



Şekil 12. Gerçek değerler ile karma dağılımdan üretilmiş değerlerin karşılaştırılması

5. SONUÇLAR VE TARTIŞMA

Sağlık sistemlerinin modellenmesinde simülasyon modelleri çok yoğun biçimde kullanılmaktadır. Çalışmalarda kullanılan veri setinin teorik bir dağılıma uymadığı durumlarda bir çok araştırmacı simülasyon modellerinde ampirik dağılım

kullanılmaktadır. Gözlem sayısının küçük olduğu veri setlerinde ampirik dağılımdan üretilen sayılar ile gerçek sayılar birbirine yakın değerler üretirken, gözlem sayısının büyük olduğu veri setlerinde gerçek değerler ile üretilen değerler arasında büyük farklar oluşmaktadır veya veri setini çok fazla kategoriye bölmek gerekmektedir. Bu tür durumlarda, karma dağılım modellerinin kullanılması araştırmacılara veri setini daha iyi modelleme imkanı sağlayacaktır. Bu çalışmada beyin cerrahi yoğun bakım ünitesinde yatan hastaların kalış sürelerinin karma dağılım modelleri ile tahmin edilmesinin ampirik dağılımla tahmin edilmesine göre daha iyi sonuç verdiği gösterilmiştir. Bir veri setinin temsili için ampirik dağılımlar kullanılmadan önce, karma dağılım modellerine uyup uymadığına bakılması, veri setinin daha iyi modellenmesi için son derece önemlidir.

6. KAYNAKLAR

1. Türkiye İstatistik Kurumu. Haber Bülteni. Sağlık Harcamaları İstatistikleri, 2017. Sayı: 27621.
2. Günel, M.M., Pidd, M., 2010. Discrete Event Simulation for Performance Modelling in Healthcare: A Review of the Literature. Journal of Simulation, 4, 42-51.
3. Gül, M., Güneri, A.F., 2015. A Comprehensive Review of Emergency Department Simulation Applications for Normal and Disaster Conditions. Computers & Industrial Engineering, 83, 327-344
4. Bai, J., Fügenger, A., Schoenfelder, J., Brunner, J., 2018. Operations Research in Intensive Care Unit Management: A Literature Review, 21, 1-24.
5. Law, A., Kelton, D., 2000. Simulation Modeling and Analysis. 3rd Edition. McGraw-Hill Higher Education.
6. McLachlan, G., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons, Inc.
7. Ickowicz, A., Sparks, R., 2016. Modelling Hospital Length of Stay Using Convolutional Mixture Distributions. Statistics in Medicine. DOI :10.1002/sim.7135

8. Sorensen, J. 1996. Multi-phased Bed Modelling. Health Services Management Research, 9, 61-67.
9. Titterington, D.M., Smith, A.F., Makov, U.E., 1985. Statistical Analysis of Finite Mixture Distributions. John Wiley & Sons, Inc.
10. Oliveira, A., Oliveira, T., 2011. Method for Detection of Mixtures of Normal Distributions with Application to Vine Varieties. Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment. (HAICTA 2011). 771-781.
11. Montgomery, D., Runger, G., 2010. Applied Statistics and Probability for Engineers. 5th Edition. Wiley.
12. Çalış, N., 2005. Karma Dağılım Modellerinde Bileşen Sayısını Tahmin Etmek için Yöntemler, Ç.Ü., Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 10-30, Adana.
13. Fowlkes, E.B., 1979. Some Methods for Studying the Mixture of two normal (lognormal) distributions. JASA. 74, 561-575.
14. Açıkgöz, İ., 2007. Sonlu Karma Dağılımlarda Parametre Tahmini, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 68-72, Ankara.
15. Johnson, R.A., Wichern, D.W., 1982. Applied Multivariate Statistical Analysis. Prentice-Hall, Inc. Englewood Cliffs.
16. Koehler, D., Murphee, E., 1988. A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order. Applied Statistics, 187-195.
17. Roeder, K., Wasserman, L., 1997. Practical Density Estimation Using Mixtures of Normal. Journal of the American Statistical Association. 894-902.
18. Titterington, D.M., 1990. Some Recent Research in the Analysis of Mixture Distributions. Statistics 21, 619-641.
19. Antmen, Z.F., 2012. Üçüncü Basamak Yoğun Bakım Üniteleri Kapasite Planlama Problem İçin Benzetim Modelleri ve Uygulamaları. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 3-5, Adana.
20. Edbrooke, D., Hibbert, C., Ridley, S., Long, T., Dickie, H., 1999. The Development of a Method for Comparative Costing of Individual Intensive Care Units. The Intensive Care Working Group on Costing, Anaesthesia 54(2), 110-120.
21. Gyldmark, M., 1995. A Review of Cost Studies of Intensive Care Units: Problems with the Cost Concept. Critical Care Medicine 23, 964-972.

