

**ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

**YARIŞAN BAĞIMLI RİSKLERLE SAĞKALIM ANALİZİNDE
ARCHIMEDEAN KAPULA YAKLAŞIMI**

Çiğdem TOPÇU

İSTATİSTİK ANABİLİM DALI

**ANKARA
2013**

Her Hakkı Saklıdır

ÖZET

Doktora Tezi

YARIŞAN BAĞIMLI RİSKLERLE SAĞKALIM ANALİZİNDE ARCHIMEDEAN KAPULA YAKLAŞIMI

Çiğdem TOPÇU

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Fahrettin ARSLAN

Sağkalım analizi, başlama anı ile çalışmanın başında belirlenen ve başarısızlık olarak adlandırılan bir sonucun ortaya çıkmasına kadar geçen süre olarak elde edilen verilerin analizidir. Sağkalım analizinde birey için başarısızlık birden fazla nedenden meydana gelebilmektedir. Bu tür çalışmalarda, bireye ait sağkalım süresi T ve meydana gelebilecek başarısızlığın olası nedeni ise kesikli C rasgele değişkeni ile gösterilir. Burada, toplam k tane farklı neden vardır ve bunlar 1. neden, 2. neden, ... , k . neden biçiminde ifade edilir. Başarısızlığın bu olası nedenleri yarışan risk olarak adlandırılmaktadır. Yarışan risklerin var olduğu durumlarda temel varsayım her bireye ait potansiyel $(T_1, T_2, \dots, T_k)$ sağkalım süresi vektörünün olduğudur. Bu sağkalım süreleri aynı anda gözlenemeyeceğinden örtük (latent) bir yapıya sahiptir. Tüm risklerin varlığında gözlenen gerçek sağkalım süresi $T = \min(T_1, T_2, \dots, T_k)$ olmaktadır. Risklerin bağımlılığı ile bu örtük sağkalım sürelerinin bağımlılığı anlatılmaktadır. Çalışmada iki bağımlı yarışan riskin varlığında, bu riskler arasındaki bağımlılığı en iyi şekilde temsil eden Archimedean kapula fonksiyonunun tahmin edilmesi amaçlanmıştır. Bu amaç için Genest ve Rivest (1993) 'in önerdiği parametrik olmayan yöntemin duruma uyarlaması yapılmıştır. Son olarak yöntem gerçek bir veri seti üzerinde uygulanmış, elde edilen sonuçlar ve öneriler verilmiştir.

Ocak 2013, 111 sayfa

Anahtar Kelimeler: Sağkalım analizi, yarışan riskler, bağımlılık, kapula, Archimedean kapula

ABSTRACT

Ph. D. Thesis

ARCHIMEDEAN COPULA APPROACH TO DEPENDENT COMPETING RISKS IN SURVIVAL ANALYSIS

Çiğdem TOPÇU

Ankara University
Graduate School of Natural and Applied Science
Department of Statistics

Advisor: Prof. Dr. Fahrettin ARSLAN

Survival analysis is related to analyze the data which are obtained as a time until a prespecified event is called failure event. In survival analysis, failure may occur from several causes for any individual. In such studies, survival time that is assigned to an individual and potential cause of failure are represented by T and C , respectively. There are k different causes and these are represented by 1st cause, 2nd cause, ..., k^{th} cause. These potential causes are called as competing risks. In the presence of competing risks, the basic assumption is there is a potential $(T_1, T_2, \dots, T_k)$ survival time vector that is assigned to every individual. Since these survival times are not observable random variables they have a latent structure. When all the risks are present the actual observed survival time is $T = \min(T_1, T_2, \dots, T_k)$. Dependent risks means dependent these latent survival times. The aim of this study is to estimate the Archimedean copula function that provide best fit for the dependency between two dependent competing risks. For the purpose, the method that is suggested by Genest and Rivest (1993) is adapted for the case. At the end of the study, the method is applied to a real-life data set, the obtained results and suggestions are presented.

January 2013, 111 pages

Key Words: Survival analysis, competing risks, dependency, copula, Archimedean copulas

TEŞEKKÜR

İhtiyaç duyduğum her anda değerli zamanını bana ayırarak gerek akademik gerekse hayat deneyimini benimle cömertçe paylaşan, çalışmalarımın her aşamasında önerileriyle beni yönlendiren değerli danışman hocam Prof. Dr. Fahrettin ARSLAN'a (Ankara Ünv. İstatistik Bölümü) sonsuz teşekkürlerimi sunarım. Çalışmalarımı titizlikle takip ederek, yapmış olduğu katkılar ve örnek aldığım anlayışlı ve hoşgörülü kişiliği için değerli hocam Prof. Dr. Hamza GAMGAM'a (Gazi Ünv. İstatistik Bölümü) teşekkürlerimi sunarım. Çalışmalarımı büyük bir özenle takip ederek son ana kadar yaptığı değerli katkıları, önerileri ve yardımları için değerli hocam Doç. Dr. Halil AYDOĞDU'ya (Ankara Ünv. İstatistik Bölümü) teşekkürlerimi sunarım. Dikkatle incelediği çalışmama yaptığı önemli katkılarından, önerilerinden ve eksik etmediği moral desteğinden dolayı değerli hocam Yrd. Doç. Dr. İhsan KARABULUT 'a (Ankara Ünv. İstatistik Bölümü) teşekkürlerimi sunarım. Tez çalışmama gösterdiği özen, yaptığı değerli katkı ve önerilerinin yanı sıra benden esirgemediği moral destekten dolayı değerli hocam Doç. Dr. Çağdaş Hakan ALADAĞ'a (Hacettepe Ünv. İstatistik Bölümü) teşekkürlerimi sunarım. Her alanda olduğu gibi doktora süresince de manevi desteğini benden esirgememiş olan sevgili dostum Yrd. Doç. Dr. Pelin KASAP'a (Ondokuz Mayıs Ünv. İstatistik Bölümü) teşekkür ederim. Çalışmalarım süresince hep yanımda olan, kader ortaklığı yaptığımız ve moral desteğini her an hissettiğim sevgili arkadaşım Ayhan TOPCU'ya teşekkür ederim. Manevi desteklerini, yardımlarını ve güler yüzlerini benden esirgemeyen Öğr. Gör. Dr. M. Bahar BAŞKIR (Ankara Ünv. İstatistik Bölümü), Arş. Gör. Yeşim DÖNE (Ankara Ünv. İstatistik Bölümü) ve Arş. Gör. Abdullah YALÇINKAYA (Ankara Ünv. İstatistik Bölümü) başta olmak üzere tüm arkadaşlarıma teşekkür ederim. Yaşamımın her anında varlıklarından güç aldığım, maddi ve manevi desteklerini esirgmeden yanımda olan sevgili aileme en içten teşekkürlerimi sunarım.

Çiğdem TOPÇU

Ankara, Ocak 2013

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	vi
ÇİZELGELER DİZİNİ	vii
1. GİRİŞ	1
2. SAĞKALIM ANALİZİNDE BAZI TEMEL KAVRAMLAR	4
2.1 Sağkalım Analizinde Sansürleme Kavramı ve Parametrik Olmayan Tahminler	7
2.2 Kaplan-Meier Tahmin Edicisi	9
2.3 Kaplan-Meier Tahmin Edicisi İçin Ortalama ve Varyansın Bulunması	13
2.4 Parametrik Olmayan En Çok Olabilirlik Tahmin Edicisi Olarak Kaplan-Meier Tahmin Edicisi	15
2.5 Nelson-Aalen Tahmin Edicisi	20
2.6 Nelson-Aalen Tahmin Edicisi İçin Ortalama ve Varyansın Bulunması	22
3. YARIŞAN RİSKLERİN VARLIĞINDA SAĞKALIM ANALİZİ	27
3.1 Temel Kavramlar ve Gösterimler	27
3.2 Örtük Başarısızlık Zamanları Gösterimi	31
3.3 Kesikli Başarısızlık Zamanları	35
4. YARIŞAN RİSKLER TEORİSİNDE TAHMİNLER	39
4.1 Parametrik Tahminler	39
4.2 Sansürlemenin Varlığında Tahminler	43
5. BİRİKİMLİ SIKLIK FONKSİYONU	48
5.1 Yarışan Risklerin Varlığında Birikimli Sıklık Fonksiyonunun Tahmini	48
5.1.1 K-M yöntemi	49
5.1.2 Yarışan riskler yöntemi	52
6. YARIŞAN RİSKLER TEORİSİNDE BELİRLENEBİLME PROBLEMİ	54

7. KAPULA FOKİYONLARI VE ARCHIMEDEAN KAPULA AİLESİ	60
7.1 Kapula Kavramı ve Özellikleri	61
7.2 Bağımlılık Kavramı ve Kuyruk Bağımlılığı	66
7.3 Archimedean Kapula Ailesi	68
8. İKİ BOYUTLU ARCHİMEDEAN KAPULA FONKSİYONUNUN TAHMİNİ	73
8.1 Birliktelik Parametresinin Tahmini	75
8.1.1 Kendall’ın tau ilişki katsayısına dayanan tahmin yöntemi	75
8.1.2 En çok olabilirlik yöntemi ile tahmin	77
8.1.3 Simülasyon sonuçları	80
8.2 Archimedean Kapula Fonksiyonunun Tahmini	84
8.3 Yöntemin Uygulanması ve Elde Edilen Sonuçlar	85
8.4 Yarışan Risklerin Varlığında Bağımlılığın Modellenmesinde Kapula Yaklaşımı	90
9. UYGULAMA	92
10. TARTIŞMA VE SONUÇ	96
KAYNAKLAR	101
EKLER	107
EK1 Tamoxifen çalışması sonucunda elde edilmiş verilerin bir kısmı	108
EK2 Tamoxifen çalışmasındaki değişkenlerin tanımları	109
ÖZGEÇMİŞ	110

ŞEKİLLER DİZİNİ

Şekil 8.1 Ampirik ve teorik λ -fonksiyonu grafikleri	89
Şekil 9.1 Kombine tedavi altında iki riske ait ampirik ve teorik λ fonksiyonu	95

ÇİZELGELER DİZİNİ

Çizelge 2.1 Kaplan-Meier tahminleri	12
Çizelge 5.1 K-M yöntemiyle hesaplanan sağkalım olasılıkları	51
Çizelge 5.2 Yarışan riskler yöntemiyle göre hesaplanan sağkalım olasılıkları	53
Çizelge 7.1 Çalışmada kullanılan iki boyutlu Archimedean kapulalar.....	71
Çizelge 8.1 Çalışmada kullanılan Archimedean kapulaların dağılımı	75
Çizelge 8.2 Çalışmada kullanılan Archimedean kapulalar ile Kendall'ın tau katsayısı arasındaki ilişki	77
Çizelge 8.3 Clayton($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları	81
Çizelge 8.4 Clayton($\theta = 3$) kapulası için elde edilen simülasyon sonuçları	81
Çizelge 8.5 Clayton($\theta = 5$) kapulası için elde edilen simülasyon sonuçları	81
Çizelge 8.6 Gumbel ($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları	82
Çizelge 8.7 Gumbel ($\theta = 3$) kapulası için elde edilen simülasyon sonuçları	82
Çizelge 8.8 Gumbel ($\theta = 5$) kapulası için elde edilen simülasyon sonuçları	82
Çizelge 8.9 Frank($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları	83
Çizelge 8.10 Frank($\theta = 5$) kapulası için elde edilen simülasyon sonuçları	83
Çizelge 8.11 Frank($\theta = 15$) kapulası için elde edilen simülasyon sonuçları	83
Çizelge 8.12 Birliktelik parametresinin TAU ve MLE tahminleri	86
Çizelge 8.13 $\hat{\theta}_{TAU}$ kullanılarak elde edilen sonuçlar	87
Çizelge 8.14 $\hat{\theta}_{MLE}$ kullanılarak elde edilen sonuçlar	88
Çizelge 9.1 Tamoxifen tedavisi altındaki hasta grubu için, uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks riski arasındaki bağımlılık modelini seçimi	93
Çizelge 9.2 Tamoxifen ve radyoterapi tedavisi altındaki hasta grubu için koltukaltı lenf nodu bölgesinde nüks riski uzak organ yayılım riski arasındaki bağımlılık modelini seçimi	94

1. GİRİŞ

Sağkalım analizi, başarısızlık olarak belirlenen bir sonucun ortaya çıkmasına kadar geçen süre ile ilgili olarak elde edilen verilerin analizidir. Özellikle, tıp alanındaki çalışmaların birçoğunda çalışmanın başında belirlenen ve ortaya çıkması gözlenen bu olay, hastanın ölümü, hastalığın nüks etmesi, hastalığın başka bir hastalığı tetiklemesi gibi olaylar olabilmektedir. Sağkalım analizinde meydana gelmesi gözlenen olaylar genellikle olumsuz olduğundan bu olaylar başarısızlık olarak adlandırmaktadır. Yarışan risklerin olmadığı, klasik sağkalım analizinde başarısızlığın tek bir nedeni ile ilgilenilmekte ve başarısızlığa neden olan diğer olaylar göz ardı edilmektedir. Birçok çalışmada, başarısızlığın birden çok nedeni olmaktadır (Moeschberger ve David 1971). Birden fazla başarısızlık nedeninin olduğu durumlarda yapılan sağkalım analizinde ilgilenilen başarısızlık nedeninin yanında diğer nedenlerin de göz önüne alınması daha gerçekçi bir yaklaşımdır.

Örneğin, belli bir tür kanser tanısı konmuş bir hasta grubuna uygulanan bir ya da birkaç farklı tedavi türünün etkinliğinin araştırılmak istendiği bir çalışma düzeni göz önüne alınsın. Başarısızlık olarak adlandırılan sonuç hastanın ölümüdür. Hastanın ölümünü ortaya çıkaran risklerin, cerrahi operasyon sonucu gelişen bir komplikasyon, hastalığın nüks etmesi, enfeksiyon ve kalp krizinin olabileceği düşünölsün. Ölümü ortaya çıkartacak bu ‘*risk*’ler, ölüm meydana geldikten sonra hastanın ölümü için ‘*neden*’ olarak adlandırılmaktadır. Riskler neden olmak için yarışmıştır. Çalışmanın amacı doğrultusunda, bu risklerden hastalığın nüks etmesine ait veriler toplanıp analizler yapılırken diğer riskler de göz önüne alınmalıdır. Başarısızlığın tek bir nedeni ile ilgilenilen klasik sağkalım analizinde, nüks sebebiyle ölüm dışında meydana gelen ölümler sansürlü gözlem olarak düşünölmektedir. Kalp krizi nedeniyle ölümlerin sansürlü gözlem olarak düşünölməsi, uygulanan tedavinin kalp krizi riskini arttırabileceği bilgisini gözden kaçırmaya sebep olabilecektir. Bu nedenle, tedavi etkinliğinin araştırıldığı bu çalışmada, diğer riskler nedeniyle ölümleri de göz önüne almak daha gerçekçi bir sonuçlar verecektir.

Yarışan risklerin var olduğu standart düzenlerin analiz ve modellemesi için iki yapı söz konusudur. Bu yapılar;

i) Nedene özel alt dağılım fonksiyonu ya da birikimli sıklık fonksiyonu (cumulative incidence function, CI);

$$F_j(t) = P(T \leq t, C = j), \quad j = 1, 2, \dots, k$$

ii) Nedene özel hazard fonksiyonu;

$$\lambda_j(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, C = j | T \geq t)}{\Delta t}, \quad j = 1, 2, \dots, k$$

şeklindedir.

Çalışmanın ikinci bölümünde sağkalım analizinde yer alan bazı temel kavramlar üzerinde durulmuştur. Yine, başta sağkalım analizinde önemli bir yere sahip olan Kaplan-Meier (K-M) tahmin edicisi tanıtılmış ve parametrik olmayan en çok olabilirlik tahmin edicisi olarak nasıl elde edildiği gösterilmiştir. Birikimli hazard fonksiyonunun parametrik olmayan tahmin edicisi olan Nelson-Aalen tahmin edicisi üzerinde durulmuş, K-M tahmin edicisi ile aralarındaki ilişkiye değinilmiştir.

Üçüncü bölümünde ise, çalışmanın konusunu oluşturan yarışan riskler teorisine ait kavramlara ve gösterimlere yer verilmiştir. Matematiksel olarak yarışan riskler modeli incelenmiş ve yarışan riskler modelinin örtük (latent) başarısızlık süreleri olarak ifade edilmesi üzerinde durulmuştur.

Dördüncü ve beşinci bölümde sırasıyla, yarışan riskler teorisinde nedene özel hazard fonksiyonuna ve birikimli sıklık fonksiyonuna değinilmiş bu fonksiyonlarının tahminlerinin nasıl elde edileceği açıklanmıştır.

Altıncı bölümde, örtük başarısızlık süreleri gösterimi ile göz önüne alınan yarışan riskler modelinde, değişkenlerin bağımlılığı söz konusu iken bu değişkenlere ait ortak dağılımın belirlenmesinde karşılaşılan belirlenebilme problemine değinilmiştir.

Yedinci bölümde, yarışan risklerin bağımlılığı durumunda ortak dağılımın belirlenmesinde kapula fonksiyonlarının kullanımı üzerinde durulmuştur. Matematiksel özellikleri, uygulamada sağladığı kolaylık ve sağkalım tipi verilerine uygunluğu nedeniyle çalışmada geniş yer verilecek olan Archimedean kapula ailesi tanıtılmıştır. Bu aileye ait olan, pozitif bağımlılığın modellenmesi ve kuyruk bağımlılığın yorumlanmasına olanak tanıyan Gumbel ve Clayton kapulalarının yanı sıra, hem pozitif hem de negatif bağımlılığın modellenmesinde kullanılan Frank kapula incelenmiştir.

Sekizinci bölümde, değişkenlerin sahip olduğu bağımlılığı en iyi modelleyecek Archimedean kapula fonksiyonunun seçimi diğer bir ifadeyle, iki boyutlu Archimedean kapula fonksiyonun tahmin edilmesi üzerinde durulmuştur. Veriyi modelleyecek uygun kapulanın seçim yöntemi için, Genest ve Rivest (1993)'in önerdiği parametrik olmayan yöntem esas alınmıştır. Genest ve Rivest (1993) çalışmasında, kapula parametresinin (birliktelik parametresinin) tahmini için Kendall'ın tau katsayısına dayanarak elde edilen tahmin ediciyi kullanırken, bu çalışmada aynı yöntem için kapula parametresinin en çok olabilirlik tahmin edicisi kullanılmıştır. Bu iki tahmin edicinin karşılaştırılması amacıyla bir simülasyon çalışması yapılmıştır.

Çalışmanın devamında uygun kapula fonksiyonunun seçim kriteri için Cramer Von Misses istatistiği kullanılabileceğinden bahsedilmiş ve Gumbel kapuladan simülasyon yoluyla edilen bir veri seti üzerinde, yöntem uygulanmış elde edilen sonuçlar tartışılmıştır.

Çalışmanın uygulama bölümünde ise Fyles vd. (2004) çalışmasının sonuçlarından oluşan gerçek bir veri seti kullanılarak, bu çalışmada önerilen yöntem uygulanmış ve bağımlılığı modelleyen uygun kapula fonksiyonu seçilmiş ve yorumlanmıştır.

2. SAĞKALIM ANALİZİNDE BAZI TEMEL KAVRAMLAR

Genel olarak sağkalım analizi; çalışmada başarısızlık kabul edilen bir sonucun ortaya çıkmasına kadar geçen süre ile ilgili verilerin analizidir.

Negatif olmayan sürekli bir T rasgele değişkeni, bir kitledeki bireylerin sağkalım süresini gösterebilir ve T rasgele değişkeninin olasılık yoğunluk fonksiyonu, f olsun.

Herhangi bir bireyin t zamanından sonra yaşama olasılığı;

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx \quad (2.1)$$

şeklinde tanımlanır ve ‘Sağkalım Fonksiyonu’ olarak adlandırılır. Bu fonksiyon aşağıdaki özellikleri taşıyan, artmayan bir fonksiyondur.

$$\lim_{t \rightarrow \infty} S(t) = 0 \quad \text{ve} \quad \lim_{t \rightarrow 0} S(t) = 1.$$

Ayrıca,

$$F(t) = P(T \leq t) = \int_0^t f(x)dx, \quad t > 0 \quad (2.2)$$

şeklinde tanımlanan fonksiyona, ‘Birikimli Başarısızlık (Failure) Fonksiyonu’ denir ve bireyin, t zamanından önce ölmesi olasılığını verir.

Sağkalım fonksiyonu ile birikimli başarısızlık fonksiyonu arasında;

$$S(t) = 1 - F(t)$$

ile ifade edilen bir ilişki vardır.

t zamanına kadar yaşadığı bilinen bir bireyin t zamanındaki ani başarısızlık (failure) ya da ölüm oranı,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.3)$$

şeklinde tanımlanır ve $h(t)$ fonksiyonuna ‘hazard fonksiyonu’ denir.

Daha önce tanımları verilen $S(t)$ ve $f(t)$ fonksiyonları kullanılarak $h(t)$ ve $S(t)$ arasındaki ilişkiler aşağıdaki gibi elde edilir.

$$h(t) = -\frac{d}{dt} \ln S(t)$$

$$\ln S(x) \Big|_0^t = \ln S(t) - \ln S(0) = -\int_0^t h(x) dx$$

$S(0) = 1$ olduğundan $\ln S(0) = 0$ ve böylece

$$\ln S(t) = -\int_0^t h(x) dx$$

$$S(t) = \exp\left(-\int_0^t h(x) dx\right) \quad (2.4)$$

yazılır.

Böylelikle birikimli hazard fonksiyonunun tanımını yapmak mümkün olacaktır ve (2.5) ile gösterildiği şekilde tanımlanır.

$$H(t) = \int_0^t h(x) dx = \int_0^t \frac{f(x)}{S(x)} dx \quad (2.5)$$

(2.4) ile gösterilen sağkalım fonksiyonu, birikimli hazard fonksiyonu yardımıyla

$$S(t) = \exp[-H(t)] \quad (2.6)$$

olarak yazılabilir. Ayrıca, f olasılık fonksiyonu

$$f(t) = h(t) \exp\left(-\int_0^t h(x) dx\right) \quad (2.7)$$

olur.

T rasgele değişkeni kesikli bir değişken olarak düşünüldüğünde $\{t_1, t_2, \dots\}$, $(0 < t_1 < t_2 < \dots)$ T 'nin aldığı değerlerin kümesi olmak üzere, T 'nin olasılık fonksiyonu;

$$f(t_j) = P(T = t_j), \quad j = 1, 2, \dots \quad (2.8)$$

ve sağkalım fonksiyonu;

$$S(t) = P(T \geq t) = \sum_{j: t \leq t_j} P(T = t_j) = \sum_{t \leq t_j} f(t_j) \quad (2.9)$$

dır.

Sağkalım fonksiyonu, her $t \geq 0$ için $S(0) = 1$ ve $S(\infty) = 0$ özelliklerini taşıyan soldan sürekli artmayan bir basamak fonksiyonudur.

Kesikli zaman hazard fonksiyonu;

$$h(t_j) = P(T = t_j | T \geq t_j) \quad (2.10)$$

olmak üzere

$$h(t_j) = \frac{f(t_j)}{S(t_j)}, \quad j = 1, 2, \dots \quad (2.11)$$

ile tanımlanır.

Ayrıca;

$$f(t_j) = S(t_j) - S(t_{j+1}) \quad (2.12)$$

olduğundan (2.11) ile gösterilen ifadeden yola çıkılarak,

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)}, \quad j = 1, 2, \dots \quad (2.13)$$

eşitliği ve bu eşitliğin yardımıyla

$$S(t) = \prod_{j:t_j < t} [1 - h(t_j)] \quad (2.14)$$

ifadesine ulaşılır. (2.14) ile gösterilen ifade kesikli durum için yazılabilmektedir ve (2.12) eşitliği ardışık olarak çalıştırılarak elde edilişi gösterilebilmektedir.

2.1 Sağkalım Analizinde Sansürleme Kavramı ve Parametrik Olmayan Tahmin Ediciler

Sağkalım analizi, başlama anı ile belirlenen bir olayın ortaya çıkmasına kadar geçen süre boyunca birimlerden elde edilecek gözlemlerin, bir takım nedenlerden dolayı tam olarak gözlenememesi sansürlü verilerin ortaya çıkmasına neden olur. Sansürlü verilerin analizinin hedeflendiği sağkalım analizinde, ilgilenilen temel konu $P(T > t) = S(t)$ sağkalım fonksiyonunun tahmin edilmesidir. Sağkalım analizinde T ve C , sırasıyla sağkalım süresi ve sansürleme süresini gösteren negatif olmayan rasgele değişkenler olmak üzere, gözlemler $X = \min(T, C)$ formunda elde edilmektedir. T ve C 'nin

bağımsızlığının varsayımını içeren bilgilendirici olmayan (noninformative) sansürleme modeli çerçevesinde, $S(t)$ sağkalım fonksiyonu için en uygun tahmin edici, Kaplan-Meier (K-M) tahmin edicisi olmaktadır (Kaplan-Meier 1958). Sansürlü verilerin analizi için önerilen neredeyse tüm istatistiksel yöntemlerde T ve C 'nin bağımsızlığı varsayımı kullanılmaktadır.

Bu varsayımda, sağkalım süresinin sansürlemenin meydana gelme sebeplerinden etkilenmediği düşünülmektedir. Örneğin; klinik çalışmalarda hastalar tedaviden, tedavinin yan etkilerinden veya tedavinin etkisiz kalmasından veya beklenmeyen bir durumun ortaya çıkmasından dolayı ayrılabilir. Bu durumların sonucunda, sansürlü gözlemler elde edilmektedir. Böyle durumlarda, sağkalım süresinin tahmini aşamasında, her bireyin sansürlenme nedeninin sağkalım süresine etkisi göz ardı edilmektedir. Yani; sansürlemenin bireyin sağkalım süresi hakkında bilgi vermediği düşünülmektedir (Lagakos 1979).

Bu varsayımın geçerliliğinin aşağıda belirtilen üç durum için sorgulanabilir olduğunu öne sürülmüştür (Lagakos 1979).

- Hastaların, klinik denemelerde tedavinin yan etkilerinden dolayı çalışmadan ayrılmaları
- Klinik çalışmalarda, hastaların metazas gibi kritik sebeplerden dolayı çalışmaya devam edememesi
- Asıl ortaya çıkması beklenen olay dışında, ikinci bir olayın ortaya çıkması nedeniyle çalışmadan ayrılması

Bu üç durumda da bireylerin sağkalım süreleri sansürlenmiştir. Sansürleme sonucunda; hastalığın seyri hakkında hüküm vermek için kullanılan gelecek sağkalım tahminleri, sansürlenmiş bu bireylerin olası sağkalımlarından etkilenmektedir. Fakat, K-M tahmin edicisi bu durumu gözardı etmektedir. Sansürlemenin gerçekçi olmayan prognoz ortaya çıkardığı durumlarda, uygun tahmin ediciler K-M tahmin edicisi ile üstten, gözlenen X rasgele değişkeninin ampirik sağkalım fonksiyonu ile alttan sınırlı olmalıdır (Link 1989).

2.2 Kaplan-Meier Tahmin Edicisi

Sansürlenmiş gözlemlerin olduğu örneklemelerde, sağkalım süresi t 'den büyük gözlemlerin sayısı kesin olarak bilinemeyecektir. Bu durum, sansürlemenin olmadığı durumda elde edilen sağkalım fonksiyonunun tahmininin sansürlemenin olduğu duruma uyarlanması gerektirir. Bu uyarlama sonucu elde edilen sağkalım fonksiyonunun tahmini "Çarpım Limit (Ç-L) tahmin edicisi" olarak bilinir.

Sağkalım fonksiyonu için, K-M tahmin edicisi aşağıdaki belirtilen aşamalara dayanır (Kaplan ve Meier 1958).

a) Zaman eksenini $[0, \infty)$, $0 = a_0 < a_1 < \dots < a_k < a_{k+1} = \infty$ olmak üzere uygun olarak seçilen $k+1$ tane $I_j = [a_{j-1}, a_j)$ olacak şekilde alt aralıklara bölünür.

b) Her bir, $[a_{j-1}, a_j)$ aralığı için, a_{j-1} 'den sonra yaşadığı bilinen bir bireyin, a_j 'den sonrada sağkalma olasılığı olarak tanımlanan koşullu olasılığı,

$$p_j = P(T > a_j | T > a_{j-1}) = \frac{P(T > a_j)}{P(T > a_{j-1})} = \frac{P_j}{P_{j-1}} = \frac{S(a_j)}{S(a_{j-1})} \text{ olarak ifade edilir.}$$

c) t noktası için; $P(t) = S(t) = \Pr(T > t)$ olasılığı t 'den önceki tüm aralıklar için tahmin edilen p_j 'lerin çarpımı olarak bulunur.

Hiç bir zaman aralığının her bir birey için sansürleme ve başarısızlık durumunu aynı anda içermediği kabul edilmektedir.

$$P_j = P(T > a_j) = S(a_j), \quad j = 1, \dots, k+1 \quad (2.15)$$

$$j = 0 \text{ için } S(a_0) = S(0) = P_0 = 1$$

olur.

$$p_j = \frac{P_j}{P_{j-1}} \text{ olduğundan;}$$

$$j = 1 \text{ için } p_1 = \frac{P_1}{P_0} = P_1 \rightarrow P_1 = p_1$$

$$j = 2 \text{ için } p_2 = \frac{P_2}{P_1} \rightarrow P_2 = p_2 p_1$$

$$j = 3 \text{ için } p_3 = \frac{P_3}{P_2} \rightarrow P_3 = p_3 p_2 p_1$$

.

.

.

$$j = k \text{ için } p_k = \frac{P_k}{P_{k-1}} \rightarrow P_k = p_k p_{k-1} \dots p_1$$

$$P_j = p_1 p_2 \dots p_k = \prod_{j=1}^k p_j$$

olarak bulunur.

Buradan, bir t noktası için; $P(t) = S(t) = P(T > t)$ olasılığı t 'yi içeren ve öncesinde bulunan tüm aralıklar için tahmin edilen $\hat{p}_j$ 'lerin çarpımı olarak tahmin edilir.

$$\hat{P}(t) = \hat{S}(t) = \prod_{j=1}^k \hat{p}_j \quad (2.16)$$

n_j ; $[a_{j-1}, a_j)$ zaman aralığında risk altında olan, yani a_{j-1} 'de yaşayan ve öncesinde gözlem altında olan (sansürlenmemiş) toplam birey sayısı

d_j ; $[a_{j-1}, a_j)$ zaman aralığında gözlenen toplam başarısızlık sayısı

olarak tanımlanırsa;

$$\hat{P}(t) = \hat{P}_j = \frac{n_j - d_j}{n_j} = \frac{n'_j}{n_j} \quad (2.17)$$

ile tahmin edilir. Burada, $n'_j = n_j - d_j$ ifadesi riske maruz kalan birey sayısını vermektedir. Bu durumda, K-M tahmin edicisi;

$$\hat{S}(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j} = \prod_{j=1}^k \frac{n'_j}{n_j} \quad (2.18)$$

ile tanımlanır (Kaplan and Meier 1958).

Bununla birlikte, herhangi bir alt zaman aralığında başarısızlık yok ise o aralık için $\hat{p}_j = 1$ olacaktır.

Eğer, elde edilen tüm gözlemler, $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ olarak sıralandığında, $t_{(n)} = t^*$, en büyük gözlem zamanı, bir sansürleme zamanı ise, (2.18) formülü $t > t^*$ için kullanılamayacaktır. Bu durumda, $\hat{S}(t)$ tahmini $t \in (0, t^*)$ için tanımlı olacaktır ve $\hat{S}(t^*)$ ise, net bir biçimde tanımlı olmayacaktır (Kaplan ve Meier 1958).

(2.17) ifadesinde, $n_1 = n$ ve $n_j = n_{j-1} - d_{j-1}$, $j = 1, 2, \dots, k$ yazılırsa, K-M tahmin edicisi sansürlemenin olmadığı durumda elde edilen, ampirik sağkalım fonksiyonu tahminine indirgenir.

Örnek 2.1 Kaplan-Meier (1958) çalışmasında, K-M tahminlerinin elde edilmesinin anlatıldığı örnek aşağıdaki biçimde özetlenebilir. Çalışmada, 8 birey gözlenmiş ve sağkalım süreleri ay olarak kaydedilmiştir. Bu bireylerden 4 tanesi için başarısızlık gerçekleşmiş ve geri kalan bireylerin sağkalım süreleri sansürlenmiştir. Bilgiler aşağıdaki gibi özetlenmiştir.

Başarısızlık zamanları: 0.8 3.1 5.4 9.2 ay

Sansürleme zamanları: 1.0 2.7 7.0 12.1 ay

n_j ; $[a_{j-1}, a_j)$ zaman aralığında risk altında olan toplam birey sayısı

d_j ; $[a_{j-1}, a_j)$ zaman aralığında gözlenen toplam başarısızlık sayısı olmak üzere,

$$n'_j = n_j - d_j \text{ ve } \lambda_j = n'_j - n_{j+1}$$

olarak tanımlanır. Burada, $\lambda_j [a_{j-1}, a_j)$ zaman aralığındaki ölüm ve kayıpların sayısını göstermektedir.

K-M tahminlerinin elde edilme yönteminde, başarısızlık zamanları büyükten küçüğe doğru sıralanır ve gözlem aralıkları oluşturulur. Buna göre, $[0.8, 3.1)$ aralığının başında risk altında bulunan birey sayısı 8 ve bu aralıkta gözlenen başarısızlık sayısı ise 1'dir. Çünkü veriye bakıldığında, sağkalım süresi bu aralığa düşen 1 tane başarısızlık vardır. Diğer aralıklarda, aynı şekilde yorumlanarak Çizelge 2.1 oluşturulmuştur.

Çizelge 2.1 Kaplan-Meier tahminleri

j	a_{j-1}	n_j	d_j	n'_j	λ_j	$\hat{S}(a_j)$
1	0	8	0	8	0	1
2	0.8	8	1	7	2	7/8
3	3.1	5	1	4	0	7/10
4	5.4	4	1	3	1	21/40
5	9.2	2	1	1	0	21/80
6	(12.1)	1	0	1	1	21/80

Son satırda, gösterilen $(t^* = 12.1)$ değeri bir sansürleme zamanı olduğu için, $\hat{P}(T > t^*)$ tahmini tanımsız olacaktır. Çünkü $(t^* = 12.1)$ zamanından sonra sağ kalan birey sayısı, gözlenen başarısızlık sayısı bilinemeyecektir.

2.3 Kaplan Meier Tahmin Edicisi İçin Ortalama ve Varyansın Bulunması

Sağkalım fonksiyonu $S(t)$ 'nin K-M tahmin edicisi (2.16) ifadesinden

$$\hat{S}(t) = \hat{p}_1 \hat{p}_2 \dots \hat{p}_k \quad (2.19)$$

olarak yazılır.

Bu kısımda, K-M tahmin edicisinin ortalama ve varyansının elde edilişi üzerinde durulacaktır. Bu amaçla, aşağıdaki iki önemli varsayım göz önüne alınacaktır.

$n'_j = n_j - d_j$ riske maruz kalan birey sayısı olmak üzere,

i) n'_j verildiğinde $\hat{p}_j = \frac{n'_j}{n_j}$, $i = 1, 2, \dots, k$ tahmin edicisinin binom oranı varsayımı

ile p_j ortalama ve $\frac{p_j q_j}{n'_j}$ varyansına sahip olduğu varsayılır. Burada,

$q_j = 1 - p_j$ dır.

ii) $n'_1, n'_2, \dots, n'_k$ istatistikleri verildiğinde, $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k$ tahmin edicileri birbirinden bağımsızdır.

Birinci ve ikinci varsayımlar göz önüne alındığında,

$$E[\hat{p}_j | n'_1, n'_2, \dots, n'_k] = p_j \quad (2.20)$$

ve

$$\begin{aligned} E[\hat{S}(t) | n'_1, n'_2, \dots, n'_k] &= E[\hat{p}_1 \dots \hat{p}_k | n'_1, n'_2, \dots, n'_k] \\ &= \prod_{j=1}^k E[\hat{p}_j | n'_1, n'_2, \dots, n'_k] = p_1 \dots p_k = S(t) \end{aligned} \quad (2.21)$$

olur. Böylece, kabul edilen varsayımlar altında, $\hat{S}(t)$, $S(t)$ için yansız bir tahmin edicidir. K-M tahmin edicisi anı zamanda tutarlı bir tahmin edicidir (Kaplan ve Meier 1958, Fleming ve Harrington 2005).

K-M tahmin edicisine ilişkin varyans ise aşağıdaki gibi elde edilir.

$$\begin{aligned}
 Var[\hat{S}(t) | n'_1, n'_2, \dots, n'_k] &= E[\hat{S}^2(t) | n'_1, n'_2, \dots, n'_k] - \left(E[\hat{S}(t) | n'_1, n'_2, \dots, n'_k]\right)^2 \\
 &= E[\hat{p}_1^2 \dots \hat{p}_k^2 | n'_1, n'_2, \dots, n'_k] - \left(E[\hat{p}_1 \dots \hat{p}_k | n'_1, n'_2, \dots, n'_k]\right)^2 \\
 &= \prod_{j=1}^k E(\hat{p}_j^2 | n'_1, n'_2, \dots, n'_k) - \left(\prod_{j=1}^k E(\hat{p}_j | n'_1, n'_2, \dots, n'_k)\right)^2 \quad (2.22)
 \end{aligned}$$

$E(\hat{p}_j^2 | n'_1, n'_2, \dots, n'_k) = Var(\hat{p}_j | n'_1, n'_2, \dots, n'_k) + E(\hat{p}_j | n'_1, n'_2, \dots, n'_k)^2$ olduğunun göz önüne alınmasıyla,

$$\begin{aligned}
 Var[\hat{S}(t) | n'_1, n'_2, \dots, n'_k] &= \prod_{j=1}^k \left(\frac{p_j q_j}{n_j} + p_j^2\right) - \left(\prod_{j=1}^k p_j\right)^2 \\
 &= \prod_{j=1}^k p_j^2 \left(1 + \frac{q_j}{p_j n_j}\right) - \left(\prod_{j=1}^k p_j\right)^2 \\
 &= \prod_{j=1}^k p_j^2 \prod_{j=1}^k \left(1 + \frac{q_j}{p_j n_j}\right) - \left(\prod_{j=1}^k p_j\right)^2 \\
 &= \left(\prod_{j=1}^k p_j\right)^2 \prod_{j=1}^k \left(1 + \frac{q_j}{p_j n_j}\right) - \left(\prod_{j=1}^k p_j\right)^2 \\
 &= \left(\prod_{j=1}^k p_j\right)^2 \left[\prod_{j=1}^k \left(\frac{q_j}{p_j n_j} + 1\right) - 1\right] \quad (2.23) \\
 &= [S(t)]^2 A
 \end{aligned}$$

olur. Burada,

$$A = \left(1 + \frac{q_1}{p_1 n_1}\right) \left(1 + \frac{q_2}{p_2 n_2}\right) \dots \left(1 + \frac{q_k}{p_k n_k}\right) - 1$$

$$= \frac{q_1}{p_1 n_1} + \frac{q_2}{p_2 n_2} + \dots + \frac{q_k}{p_k n_k}$$

Yüksek dereceli terimler ihmal edilmesiyle

$$\text{Var}[\hat{S}(t)] \approx [S(t)]^2 \sum_{j=1}^k \frac{q_j}{p_j n_j} \quad (2.24)$$

ifadesine ulaşılır.

(2.24) ifadesi Greenwood formülü olarak bilinir (Greenwood 1926).

$\hat{q}_j = \frac{d_j}{n_j}$ ve $\hat{p}_j = \left(1 - \frac{d_j}{n_j}\right)$ olduğunun göz önüne alınmasıyla, $\text{Var}[\hat{S}(t)]$ için bir tahmin edici

$$\hat{\text{Var}}[\hat{S}(t)] \approx [\hat{S}(t)]^2 \sum_{j=1}^k \frac{d_j}{n_j (n_j - d_j)} \quad (2.25)$$

olur.

2.4 Parametrik Olmayan En Çok Olabilirlik Tahmin Edicisi Olarak K-M Tahmin Edicisi

K-M tahmin edicisi; $S(t)$ sağkalım fonksiyonunun parametrik olmayan en çok olabilirlik tahmin edicisi olarak elde edilebilir. Bu amaçla, n tane bireyin $t=0$ anından başarısızlık olarak belirlenen olay gerçekleşene kadar takip edildiği düşünölsün. Bu bireylere ait sağkalım süreleri $T_1, T_2, \dots, T_n$ olarak gösterilsin ve bu sürelerin birbirinden bağımsız olduğu varsayölsin. $t = 0, 1, 2, \dots$ zaman noktaları için, i . bireye ait sağkalım

fonksiyonu $S_i(t)$ ve hazard fonksiyonu $h_i(t)$ olduğu göz önüne alınarak aşağıdaki tanımlar yapılır (Lawless 2003).

$$Y_i(t) = I(T_i \geq t, \text{ bireyin sağ kalım süresi } t \text{ zamanından önce sansürlenmemiştir}) \quad (2.26)$$

olarak tanımlanır ve i . bireyin t zamanında yaşadığını ve öncesinde gözlem altında olduğunu gösteren risk göstergesidir. Eğer, birey risk altında ise yani t zamanından önce bireyin sağkalım süresi sansürlenmemişse bu değişken 1 değerini alır, aksi durumda 0 değerini alacaktır.

$$Y(t) = \sum_{i=1}^n Y_i(t); \quad t \text{ zamanında risk altında bulunan birey sayısını ifade eder.} \quad (2.27)$$

$$dN_i(t) = Y_i(t)I(T_i = t) \quad (2.28)$$

ifadesi ise, i . birey için $[t, t + \Delta t)$ zaman aralığında başarısızlığın gerçekleşip gerçekleşmediğini ve t 'nin öncesinde gözlem altında olup olmadığını gösterir. Eğer, birey için başarısızlık gerçekleşmiş ise, $dN_i(t) = 1$ olacak aksi durumda 0 değerini alacaktır.

$$dN(t) = \sum_{i=1}^n dN_i(t) \quad (2.29)$$

ifadesi $[t, t + \Delta t)$ zaman aralığında gözlenen başarısızlık sayısını ifade eder.

$$dC_i(t) = Y_i(t)I(i. \text{ birey } t \text{ zamanında sansürlüdür}) \quad (2.30)$$

ifadesi ise, i . birey sağkalım süresinin $[t, t + \Delta t)$ zaman aralığında sansürlü olup olmadığını gösterir. Eğer, bireyin sağkalım süresi sansürlenmiş ise, $dC_i(t) = 1$ olacak aksi durumda 0 değerini alacaktır.

$$dC(t) = \sum_{i=1}^n dC_i(t) \quad (2.31)$$

ifadesi $[t, t + \Delta t)$ zaman aralığındaki sansürlü gözlemlerin sayısını ifade eder.

Herhangi bir birey için, $\{dN_i(t), dC_i(t), t \geq 0\}$ değerlerinden sadece bir tanesi sıfırdan farklıdır. Yukarıdaki tanımlardan yola çıkarak,

$$dN(t) = (dN_1(t), \dots, dN_n(t)) \quad (2.32)$$

$$dC(t) = (dC_1(t), \dots, dC_n(t)) \quad (2.33)$$

vektörleri tanımlanabilir. t zamanında meydana gelen başarısızlık ve sansürleme geçmişi;

$$H(t) = \{(dN(s), dC(s)), s = 0, 1, \dots, t-1\} \quad (2.34)$$

ile özetlenmektedir. Burada; $(t-1)$ zamanına kadar olan başarısızlık ve sansürleme bilgisi yer almaktadır.

Gözlenen veriye bağlı olarak L , olabilirlik fonksiyonu

$$L = \prod_{t=0}^{\infty} P(dN(t) | dH(t)) P(dC(t) | dN(t), dH(t)) \quad (2.35)$$

olarak yazılır (Lawless 2003).

i . bireye ait hazard fonksiyonu $h_i(t)$; bireyin t zamanına kadar yaşadığı bilindiğinde $[t, t + \Delta t)$ zaman aralığında başarısızlığa uğrama olasılığını göstermektedir.

Bu durumda,

$$P(dN_i(t) = 0 | H(t)) = \begin{cases} 1, & Y_i(t) = 0 \\ 1 - h_i(t), & Y_i(t) = 1 \end{cases} \quad (2.36)$$

ve

$$P(dN_i(t) = 1 | H(t)) = h_i(t) \quad (2.37)$$

olacaktır. Buradan;

$$P(dN(t) | H(t)) = \prod_{i=1}^n h_i(t)^{dN_i(t)} [1 - h_i(t)]^{Y_i(t)(1-dN_i(t))} \quad (2.38)$$

elde edilir.

Eğer, $P(dC(t) | dN(t), dH(t))$ ifadesi, $h_i(t)$ 'yi belirleyen herhangi bir parametre içermiyorsa, sansürleme bilgilendirici olmayan ya da hüküm vermeyen adını taşımaktadır. Bu durumda, bu terimler olabilirlik fonksiyonundan düşürülerek ve (2.38) ifadesi, (2.35)'in içerisine yerleştirilerek;

$$L = \prod_{i=1}^n \prod_{t=0}^{\infty} h_i(t)^{dN_i(t)} [1 - h_i(t)]^{Y_i(t)(1-dN_i(t))} \quad (2.39)$$

ifadesi elde edilir.

$T_1, T_2, \dots, T_n$ olarak gösterilen sağkalım sürelerinin bağımsız olduğu yukarıda belirtilmişti, bu değişkenlerin aynı dağılımlı olduğu varsayalım. Yani $S_i(t) = S(t)$ ve $h_i(t) = h(t)$ olsun. T 'nin dağılımı parametre olarak düşünülen $h(t)$ hazard fonksiyonu aracılığı ile incelenmektedir. Bu durumda, (2.39) ile gösterilen olabilirlik fonksiyonu

$$L = \prod_{i=1}^n \prod_{t=0}^{\infty} h(t)^{dN_i(t)} [1 - h(t)]^{Y_i(t)(1-dN_i(t))} \quad (2.40)$$

şeklinde olacaktır.

(2.29) ve (2.31) ifadeleri,

$$d_t = dN(t) = \sum_{i=1}^n dN_i(t) \quad (2.41)$$

$$n_t = \sum_{i=1}^n Y_i(t) \quad (2.42)$$

olarak düzenlenerek, (2.40) ile gösterilen olabilirlik fonksiyonu,

$$L = \prod_{t=0}^{\infty} h(t)^{d_t} [1-h(t)]^{n_t-d_t} \quad (2.43)$$

ifadesine dönüşecektir (Lawless 2003).

(2.43) ile gösterilen olabilirlik fonksiyonunda $h(t)$ parametresi için en çok olabilirlik tahmin edicisi yani fonksiyonu maksimum değeri;

$$\hat{h}(t) = \frac{d_t}{n_t} \quad (t = 0, 1, \dots, \tau), \tau = \max(t : n_t > 0) \quad (2.44)$$

olarak elde edilir.

Kesikli modellerde, sağkalım fonksiyonu;

$$S(t) = \prod_{j:t_j < t} [1-h(t_j)] \quad (2.45)$$

olarak ifade edilmekteydi. Bu durum göz önüne alınarak, $S(t)$ için en çok olabilirlik tahmin edicisi, $t = 0, 1, \dots, \tau$ için,

$$\begin{aligned}\hat{S}(t) &= \prod_{s=0}^{t-1} [1 - \hat{h}(s)] \\ &= \prod_{s=0}^{t-1} \left(1 - \frac{d_s}{n_s}\right)\end{aligned}\tag{2.46}$$

olarak elde edilir. Bu tahminin (2.17) ifadesi ile eş değer olduğu görülür. Yani, K-M tahmin edicisi $S(t)$ sağkalım fonksiyonunun parametrik olmayan en çok olabilirlik tahmin edicisidir (Lawless 2003). n_t 'nin sıfıra eşit olması durumunda, $h(t), h(t+1), \dots$ içeren terim olmadığından, $s > \tau$ için, $h(s)$ hakkında bilgi olmayacaktır. Eğer, $d_\tau < n_\tau$ ise, $\hat{S}(\tau+) > 0$ olacak ve $\tau+$ 'den sonra tahmin tanımlı olmayacaktır. Böyle bir durumla, en büyük gözlemin değerinin sansürleme zamanı olduğunda karşılaşılır. $d_\tau = n_\tau$ ise, $\hat{S}(\tau+) = 0$ olacak, $S(t)$ artmayan bir fonksiyon olduğundan, $S(t)$ 'nin tahmini $t > \tau$ için sıfır olacaktır.

2.5 Nelson-Aalen Tahmin Edicisi

$[0, \infty)$ zaman eksenini, $0 = a_0 < a_1 < \dots < a_k < a_{k+1} = \infty$ olmak üzere $I_j = [a_{j-1}, a_j)$ biçiminde $k+1$ tane alt aralıklara bölünsün ve bölüm 2.4 'te yapılan tanımlamalar kullanılarak, K-M tahmin edicisi

$$\hat{S}(t) = \prod_{a+\Delta a \leq t} \left\{1 - \frac{dN(a)}{Y(a)}\right\}, \Delta a \rightarrow 0\tag{2.47}$$

olarak yazılabilir. Burada, a noktaları için $a + \Delta a \leq t$ olduğu düşünülmelidir. Örneklem genişliği yeterince büyük ve Δa yeterince küçük ise, $\frac{dN(a)}{Y(a)}$ çok küçük bir sayı hatta

0'a yakın olacaktır. $\frac{dN(a)}{Y(a)} \rightarrow 0$ iken $e^{-\frac{dN(a)}{Y(a)}}$ fonksiyonunu MacLaurent açılımı yapılır

ve ilk iki terime yakınsadığı düşünülürse;

$$\exp\left\{-\frac{dN(a)}{Y(a)}\right\} \approx \left\{1 - \frac{dN(a)}{Y(a)}\right\} \quad (2.48)$$

yazılabilir (Zhang 2005).

Böylece, K-M tahmin edicisi;

$$\hat{S}(t) \approx \prod_{a+\Delta a \leq t} \exp\left\{-\frac{dN(a)}{Y(a)}\right\} = \exp\left\{-\sum_{a < t} \frac{dN(a)}{Y(a)}\right\} \quad (2.49)$$

ifadesine yakınsamaktadır.

$S(t) = e^{-H(t)}$ ilişki göz önüne alındığında,

$$\hat{H}(t) = \sum_{a < t} \frac{dN(a)}{Y(a)} = \sum_{j: t_j \leq t} \frac{d_j}{n_j} \quad (2.50)$$

tahmin edicisi, $H(t) = \sum_{j: t_j \leq t} h(t_j)$ olarak ifade edilen birikimli hazard fonksiyonunun

Nelson-Aalen (N-A) tahmin edicisi olarak bilinir (Nelson 1969).

Temel olarak bu tahmin edici, t zamanından önceki belirli zamanlarda gözlenen ölümlerin (başarısızlık) sayısının o zamanlarda risk altında bulunan birey sayısına bölünüp toplanmasıyla elde edilir.

Δa yeterince küçük ($\Delta a \rightarrow 0$) alınır ve her aralıkta bir kere başarısızlık gözlemlendiği düşünüldüğünde, N-A tahmin edicisi

$$\hat{H}(t) = \int_0^t \frac{dN(a)}{Y(a)} \quad (2.51)$$

olarak yazılabilir (Nelson 1969).

2.6 Nelson Tahmin Edicisi İçin Ortalama ve Varyansın Bulunması

Bölüm 2.4'te verilen tanımlar dikkate alındığında ve $dN(t) | H(t) : \text{Binom}(Y(t), h(t))$ olduğundan,

$$E[dN(t) | H(t)] = Y(t)h(t) \quad (2.52)$$

$$\text{Var}[dN(t) | H(t)] = Y(t)h(t)(1 - h(t)) \quad (2.53)$$

$$E\left[\frac{dN(t)}{Y(t)} | H(t)\right] = h(t) \quad (2.54)$$

yazılabilir.

N-A tahmin edicisinin beklenen değeri ve varyansı aşağıdaki gibi elde edilmiştir. Bu amaçla, koşullandırma yaparak çift beklenen değer alma işlemi kullanılmıştır.

X ve Y herhangi iki rasgele değişken ya da rasgele vektör olmak üzere,

$$E[X] = E[E(X | Y)] \quad (2.55)$$

$$\text{Var}(X) = \text{Var}[E(X | Y)] + E[\text{Var}(X | Y)]$$

yazılabilir.

Buna göre tahmin edicinin beklenen değeri aşağıdaki gibi bulunur.

$$\begin{aligned} E[\hat{H}(t)] &= E\left[\sum_{a < t} \frac{dN(a)}{Y(a)}\right] = \sum_{a < t} E\left[\frac{dN(a)}{Y(a)}\right] \\ &= \sum_{a < t} E\left[E\left[\frac{dN(a)}{Y(a)} | H(a)\right]\right] = \sum_{a < t} h(a) \\ &\approx \sum_{a < t} h(a)\Delta a \approx \int_0^t h(a)da = H(a) \quad (\Delta a \rightarrow 0) \end{aligned} \quad (2.56)$$

Böylece,

$$E[\hat{H}(t)] = \sum_{a < t} h(a) \quad (2.57)$$

olmaktadır.

Eğer, Δa yeterince küçük alınırsa $\sum_{a < t} h(a)$ ifadesinin limiti $H(t)$ ye eşit olacaktır. Bu durumda, $\hat{H}(t)$, $H(t)$ için yaklaşık yansız bir tahmin edici olacaktır.

N-A tahmin edicisinin varyansı ise aşağıdaki gibi elde edilir. Varyans tanımı kullanılarak,

$$\begin{aligned} Var(\hat{H}(t)) &= E[\hat{H}(t) - E(\hat{H}(t))]^2 \\ &= E\left[\sum_{a < t} \frac{dN(a)}{Y(a)} - \sum_{a < t} h(a)\right]^2 \\ &= E\left[\sum_{a < t} \left\{ \frac{dN(a)}{Y(a)} - h(a) \right\}\right]^2 \end{aligned} \quad (2.58)$$

yazılabilir.

Yukarıdaki beklenen değer,

$$\begin{aligned} &E\left[\sum_{a < t} \left\{ \frac{dN(a)}{Y(a)} - h(a) \right\}\right]^2 + \sum_{a \neq a' < t} \left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} \left\{ \frac{dN(a')}{Y(a')} - h(a') \right\} \\ &= \sum_{a < t} E\left[\frac{dN(a)}{Y(a)} - h(a)\right]^2 + \sum_{a \neq a' < t} E\left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} \left\{ \frac{dN(a')}{Y(a')} - h(a') \right\}\right] \end{aligned} \quad (2.59)$$

ifadesine eşit olur.

Çarpım terimlerinin beklenen değeri sıfır olmaktadır. Bir tanesi için bu durum aşağıdaki gibi gösterilir.

$a < a'$ için

$$\begin{aligned} & E \left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} \left\{ \frac{dN(a')}{Y(a')} - h(a') \right\} \right] \\ &= E \left[E \left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} \left\{ \frac{dN(a')}{Y(a')} - h(a') \right\} \middle| H(a') \right] \right] \end{aligned} \quad (2.60)$$

$a < a'$ için olduğu için $H(a')$ ifadesine koşullu olarak yazılan $dN(a), Y(a)$ ve $h(a)$ ifadeleri sabit birer sayıdır. Buradan

$$E \left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} E \left[\left\{ \frac{dN(a')}{Y(a')} - h(a') \right\} \middle| H(a') \right] \right] \quad (2.61)$$

yazılır.

$$E \left\{ \frac{dN(a')}{Y(a')} \middle| H(a') \right\} = h(a') \quad (2.62)$$

olduğundan,

$$E \left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\} \left\{ \frac{dN(a')}{Y(a')} - h(a') \right\} \right] = 0 \quad (2.63)$$

dır.

Buradan,

$$Var(\hat{H}(t)) = \sum_{a < t} E \left[\left(\frac{dN(a)}{Y(a)} - h(a) \right)^2 \right] \quad (2.64)$$

olarak yazılır. Benze koşullandırma işlemi kullanılarak,

$$\begin{aligned}
E \left[\frac{dN(a)}{Y(a)} - h(a) \right]^2 &= E \left[E \left[\left\{ \frac{dN(a)}{Y(a)} - h(a) \right\}^2 \mid H(a) \right] \right] \\
&= E \left[Var \left[\frac{dN(a)}{Y(a)} \mid H(a) \right] \right] = E \left[\frac{h(a)[1-h(a)]}{Y(a)} \right]
\end{aligned} \tag{2.65}$$

ifadesine ulaşılır.

Böylece $\hat{H}(t)$ tahmin edicisinin varyansı,

$$Var(\hat{H}(t)) = \sum_{a < t} E \left[\frac{h(a)[1-h(a)]}{Y(a)} \right] \tag{2.66}$$

olarak elde edilir ve $\frac{h(a)[1-h(a)]}{Y(a)}$ ifadesini tahmin etmek için

$$\frac{\frac{dN(a)}{Y(a)} \left[1 - \frac{dN(a)}{Y(a)} \right]}{Y(a) - 1} \tag{2.67}$$

tahmin edicisi kullanılırsa aşağıdaki ifadeye ulaşılır.

$$\hat{Var}(\hat{H}(t)) = \sum_{a < t} E \left[\frac{\frac{dN(a)}{Y(a)} \left[1 - \frac{dN(a)}{Y(a)} \right]}{Y(a) - 1} \right] \tag{2.68}$$

Yukarıdaki varyans tahmin edicisinin $Var(\hat{\Lambda}(t))$ için yansız bir tahmin edici olduğu aşağıdaki gibi gösterilir.

$$\begin{aligned}
E \left[\sum_{a < t} \frac{\frac{dN(a)}{Y(a)} \left[1 - \frac{dN(a)}{Y(a)} \right]}{Y(a) - 1} \right] &= \sum_{a < t} E \left[\frac{\frac{dN(a)}{Y(a)} \left[1 - \frac{dN(a)}{Y(a)} \right]}{Y(a) - 1} \right] \\
&= \sum_{a < t} E \left[E \left[\frac{\frac{dN(a)}{Y(a)} \left[1 - \frac{dN(a)}{Y(a)} \right]}{Y(a) - 1} \middle| H(a) \right] \right] \\
&= \sum_{a < t} E \left[\frac{h(a) [1 - h(a)]}{Y(a)} \right] = Var(\hat{H}(t))
\end{aligned} \tag{2.69}$$

$\hat{H}(t)$ asimptotik olarak $H(t)$ ortalama ve $Var(\hat{H}(t))$ varyans ile normal dağılıma sahiptir. $\hat{H}(t)$ tahmin edicisinin yansızlık ve asimptotik normallik özellikleri kullanılarak $H(t)$ için $(1 - \alpha)$ güven düzeyinde güven aralığı aşağıdaki gibi yazılır.

$$\hat{H}(t) \pm z_{\alpha/2} \left[\hat{Var}(\hat{H}(t)) \right]^{1/2} \tag{2.70}$$

Bu güven aralığı $S(t) = e^{-H(t)}$ ilişkisi kullanılarak sağkalım fonksiyonu için aşağıdaki gibi düzenlenebilir.

$$\left[e^{-\hat{H}(t) - z_{\alpha/2} [\hat{Var}(\hat{H}(t))]^{1/2}}, e^{-\hat{H}(t) + z_{\alpha/2} [\hat{Var}(\hat{H}(t))]^{1/2}} \right] \tag{2.71}$$

$\hat{S}(t)$ ve $\hat{H}(t)$, tahmin edicilerinin ikisi de parametrik olmayan en çok olabilirlik tahmin edicisidir. Dikkat edilmelidir ki, kesikli modellerde $\hat{S}(t)$ ve $\hat{H}(t)$ için, sürekli dağılımlar için geçerli olan $H(t) = -\log S(t)$ ilişkisi geçerli olmamaktadır. Sürekli modellerde, bazen sağkalım fonksiyonun tahmini için alternatif olarak, $\tilde{S}(t) = \exp[-\hat{H}(t)]$ önerilmektedir. $H(t)$ için önerilen tahmin ise; $-\log \hat{S}(t)$ olacaktır.

3. YARIŞAN RİSKLERİN VARLIĞINDA SAĞKALIM ANALİZİ

Yarışan riskler teorisi ve çeşitli alanlardaki uygulamasıyla ilgili literatürde birçok makale ve kitap bulunmaktadır. Bu kayaklardan bazıları çalışmada başvuru kaynağı olarak kullanılmıştır. Moeschberger ve David (1971) yarışan riskler teorisine ilişkin yayınlanan ilk çalışmalardan biridir. Gail (1975), yarışan riskler analizinde kullanılan bazı modeller üzerine yaptığı inceleme çalışması literatürün önemli çalışmalarından biridir. Bu alanda diğer bir önemli çalışma ise Moeschberger ve David (1971) çalışmasıdır. Prentice vd. (1978), yarışan risklerin varlığında sağkalın analizini göz önüne alan çalışmalardan biridir. Crowder (2001) yarışan riskler teorisinin ele alındığı ve uygulamalarına yer verildiği geniş kapsamlı bir kaynak olarak literatürde önemli bir yere sahiptir. Yarışan riskler teorisinin medikal alandaki uygulamalarına yer veren Pintilie (2006) farklı uygulama örneklerini bilgisayar programları ile incelemesi açısından önemli kaynaklardan biridir. Kalbfleisch ve Prentice (1980) ve Lawless (2003) sağkalım analizinin temel kavramlarını kapsamlı olarak inceledikleri kitaplarının bir bölümünü yarışan riskler teorisine ayırmışlardır.

3.1 Temel Kavramlar ve Gösterimler

Yarışan risklerin var olduğu sağkalım analizi çalışmalarında, toplanan veriler için önerilen bir istatistiksel model, gözlenebilen (T, C, x) büyüklüklerinin dağılımını belirlemeyi amaçlar. Genellikle, verilen bir x açıklayıcı değişken vektörü için (T, C) ikilisine ilişkin bir model belirlemek yeterli olmaktadır (Prentice ve Kalbfleisch 1978). Çalışmada, x açıklayıcı değişkeni göz ardı edilerek (T, C) diğer bir gösterimle, (T_1, T_2) ikilisinin dağılımı göz önüne alınmıştır.

Öncelikle, her birey için (T, C) ikilisi tanımlanır. Burada, T ; sağkalım süresi (başarısızlık zamanı) ve C ; başarısızlık nedeni yani gözlemin sonlanmasına neden olan olaydır ve $1, 2, \dots, k$ değerlerini alan rasgele değişkendir. Böyle bir durumda, T sürekli bir değişken iken, C kesikli bir değişken olmaktadır. Her birey için gözlenen bu (T, C) ikilisinin ortak dağılımının belirlenmesi, yarışan riskler teorisinin temel olasılık

çerçevesini oluşturmaktadır. Her birey için gözlenen, (T, C) ikilisine ait ortak dağılım aşağıdaki gibi ifade edilir.

$$F_j(t) = P(T \leq t, C = j), \quad j = 1, 2, \dots, k \quad (3.1)$$

Bu ortak dağılıma, j . neden için alt dağılım fonksiyonu ya da j . nedene özel dağılım fonksiyonu denir.

Ayrıca, j . nedene özel dağılım fonksiyonu, $S_j(t)$ ise j . nedene özel sağkalım fonksiyonu ise aşağıdaki gibi ifade edilir.

$$S_j(t) = P(T > t, C = j), \quad j = 1, 2, \dots, k \quad (3.2)$$

j . nedene özel yoğunluk fonksiyonu ise,

$$f_j(t) = \frac{d}{dt} F_j(t) \quad (3.3)$$

ya da

$$f_j(t) = -\frac{d}{dt} S_j(t) \quad (3.4)$$

olarak ifade edilmektedir.

(3.1), (3.2) ve (3.3) ile gösterilen bu nedene özel fonksiyonlar bazı kaynaklarda sırasıyla, $F(t, j)$, $S(t, j)$ ve $f(t, j)$ olarak gösterilmektedir. Bu çalışmada, güncel olarak kullanılan gösterimler tercih edilmiştir.

Bu tanımlardan sonra, j . nedenin meydana gelme olasılığı, yani C değişkeninin j değerini alma olasılığı C değişkenin marjinal dağılımıdır ve aşağıdaki gibi ifade edilir.

$$P(C = j) = F(\infty, j) = S(0, j) = \pi_j, \quad j = 1, 2, \dots, k, \quad (3.5)$$

Burada, $\pi_j > 0$ ve $\sum_{j=1}^k \pi_j = 1$ olduğu varsayılır.

Aynı şekilde C rasgele değişkeni üzerinden toplam olasılık formülünün çalıştırılmasıyla, T 'nin marjinal sağkalım fonksiyonu (tüm risklerin varlığında) ve marjinal dağılımı sırasıyla aşağıda gösterildiği gibi ifade edilir.

$$P(T > t) = S_T(t) = \sum_{j=1}^k S_j(t) \quad (3.6)$$

$$P(T \leq t) = F_T(t) = \sum_{j=1}^k F_j(t) \quad (3.7)$$

Buradan, T 'ye ait marjinal yoğunluk fonksiyonu

$$\begin{aligned} f_T(t) &= -\frac{d}{dt} S_T(t) = -\frac{d}{dt} \sum_{j=1}^k S_j(t) \\ &= -\sum_{j=1}^k \frac{d}{dt} S_j(t) = \sum_{j=1}^k f_j(t) \end{aligned} \quad (3.8)$$

ya da

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt} \sum_{j=1}^k F_j(t) = \sum_{j=1}^k f_j(t) \quad (3.9)$$

olarak elde edilir. Burada, $f_j(t)$ 'nin j . nedene özel yoğunluk fonksiyonudur.

Dikkat edilmelidir ki, $S(t, j)$ j . nedene ait sağkalım fonksiyonu değildir. (T, C) 'nin ortak dağılımıdır. j . nedene ait sağkalım fonksiyonu, yani başarısızlığın j . nedenden

kaynaklandığı bilindiğinde elde edilecek sağkalım ve başarısızlık fonksiyonu sırasıyla aşağıdaki gibidir.

$$S_j^*(t) = P(T > t | C = j) = \frac{P(T > t, C = j)}{P(C = j)} = \frac{S_j(t)}{\pi_j} \quad (3.10)$$

ve

$$F_j^*(t) = P(T \leq t | C = j) = \frac{P(T \leq t, C = j)}{P(C = j)} = \frac{F_j(t)}{\pi_j} \quad (3.11)$$

(T, C) ikilisinin dağılımının belirlenmesinde, hazard fonksiyonlarının kullanılması sıkça tercih edilen bir yoldur. Sonra ki bölümlerde nedene özel hazard fonksiyonlarına geniş olarak yer verilecektir. Birey ölüm oranıyla (mortality) ilgilenilen çalışmalarda, t yaşına kadar sağ olduğu bilinen bireyin, t yaşında belirli bir nedenden ölüm oranını temsil eden nedene özel hazard fonksiyonu diğer bir ifadeyle, j . nedenden kaynaklanan başarısızlık için alt hazard fonksiyonu,

$$\begin{aligned} \lambda_j(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t, C = j | T \geq t)}{\Delta t} \\ &= \frac{f_j(t)}{S_T(t)} \quad j = 1, 2, \dots, k \end{aligned} \quad (3.12)$$

olarak tanımlanır ve marjinal hazard fonksiyonu ise,

$$\lambda_T(t) = \sum_{j=1}^k \lambda_j(t) = \sum_{j=1}^k \frac{f_j(t)}{S(t)} = \frac{1}{S(t)} \sum_{j=1}^k f_j(t) \quad (3.13)$$

biçimindedir.

Nedene özel dağılım, yoğunluk ve hazard fonksiyonları arasında aşağıda gösterileceği şekilde bir ilişki vardır. (3.12) ile gösterilen ifade göz önüne alındığında,

$$f_j(t) = \lambda_j(t) S_T(t) \quad (3.14)$$

yazılır.

(3.3) ifadesi göz önüne alındığında,

$$F_j(t) = P(T \leq t, C = j) = \int_0^t \lambda_j(u) S(u) du \quad (3.15)$$

ifadesine ulaşılır.

T 'ye ait marjinal hazard fonksiyonu ise,

$$\lambda_T(t) = \sum_{j=1}^k \lambda_j(t) \quad (3.16)$$

olarak ifade edilir. Yine T 'ye birikimli hazard fonksiyonu ise,

$$\Lambda(t) = \sum_{j=1}^k \Lambda_j(t) \quad (3.17)$$

olarak elde edilir.

Bölüm 2'de yer verilen tanımlamalar doğrultusunda, (3.6) ile verilen T 'nin marjinal sağkalım fonksiyonu ile (3.17) ile gösterilen marjinal birikimli hazard fonksiyonu arasında aşağıdaki ilişki olduğu açıktır.

$$S_T(t) = e^{-\sum_{j=1}^k \Lambda_j(t)} = e^{-\Lambda(t)} \quad (3.18)$$

3.2 Örtük Başarısızlık Zamanları Gösterimi

Yarışan risklerin varlığındaki durumu tanımlamadaki genel yaklaşımın, örtük başarısızlık zamanı modelinin kullanılması yönündedir. Bu modelde, her bir riskle ilişkilendirilen, $(j = 1, \dots, k)$, k tane potansiyel başarısızlık zamanının (T_j) , olduğu

varsayılmaktadır. Bu k tane başarısızlık zamanından her biri, diğer risklerin var olmadığı durumdaki, varsayımsal (hipotetik) başarısızlık zamanı olarak düşünülür ve bu başarısızlık zamanlarına örtük (latent) başarısızlık zamanı adı verilir. Tüm risklerin var olduğu durumda ise, gözlenen başarısızlık zamanı, $T = \min(T_1, \dots, T_k)$ olarak ifade edilir. Eğer, $T = T_j$ ise $C = j$ olmaktadır. Burada, T_j , j . risk dışındaki tüm risklerin olmadığı durumdaki gözlenen başarısızlık zamanı olarak tanımlanır. Diğer bir ifadeyle, j . riskin tek başarısızlık nedeni olduğu durumdaki başarısızlık zamanıdır.

$T_1, T_2, \dots, T_k$ sağkalım sürelerinin negatif olmayan rasgele değişkenler olduğu göz önüne alınsın. $\mathbf{t} = (t_1, t_2, \dots, t_k)$ olmak üzere, $\mathbf{T} = (T_1, T_2, \dots, T_k)$ 'ya ait ortak dağılım fonksiyonu ve sağkalım fonksiyonu sırasıyla,

$$H(\mathbf{t}) = P(\mathbf{T} \leq \mathbf{t}) = P(T_1 \leq t_1, T_2 \leq t_2, \dots, T_k \leq t_k) \quad (3.19)$$

$$G(\mathbf{t}) = P(\mathbf{T} > \mathbf{t}) = P(T_1 > t_1, T_2 > t_2, \dots, T_k > t_k) \quad (3.20)$$

olarak tanımlanır.

T_j 'lerin bağımsızlığı durumunda, ortak sağkalım fonksiyonu;

$$G(\mathbf{t}) = \prod_{j=1}^k G_j(t_j) \quad (3.21)$$

olarak tanımlanır ve T_j 'nin marjinal sağkalım fonksiyonu,

$$G_j(t_j) = P(T_j > t_j) \quad (3.22)$$

olarak ifade edilir. Bu fonksiyon diğer risklerin olmadığı sadece j . riskin olduğu durumdaki T_j 'ye ait marjinal fonksiyondur.

Eski literatürde, bu fonksiyonlar net sağkalım fonksiyonları ve daha önceki bölümde tanımlanan $S_j(t)$ nedene özel sağkalım fonksiyonları ise kaba (crude) sağkalım fonksiyonları olarak adlandırılmaktadır.

$T_1, T_2, \dots, T_k$ sağkalım sürelerinin hepsinin aynı anda gözlenebilen değişkenler olmadığına yukarıda değinilmişti. Bu nedenle bu değişkenler arasındaki yapı örtük (latent) adını almaktadır.

Nedene özel yoğunluk fonksiyonları ve nedene özel hazard fonksiyonları ise, ortak sağkalım fonksiyonu kullanılarak doğrudan elde edilebilir. $(T_1, T_2, \dots, T_k)$ 'ya ait ortak sağkalım fonksiyonu (3.20) ile ifade edildiği gibi

$$G(t_1, t_2, \dots, t_k) = P(T_1 > t_1, \dots, T_k > t_k) \quad (3.23)$$

şeklindedir.

(3.20) ile verilen ortak dağılım bilinen bir forma sahipse, (3.6) ile verilen (T, C) ikilisindeki değişkenlerden biri olan T 'nin marjinal sağkalım fonksiyonu,

$$S_T(t) = G(t, t, \dots, t) \quad (3.24)$$

olarak bulunabilir (Tsiatis 1975).

Nedene özel yoğunluk fonksiyonları, örtük sağkalım sürelerinin ortak dağılımından doğrudan elde edilir.

$$f_j(t) = - \left(\frac{\partial G(t_1, \dots, t_k)}{\partial t_j} \right)_{t_1 = \dots = t_k = t} \quad (3.25)$$

Yukarıdaki teoreme ek olarak, $g(\mathbf{t})$, $(T_1, T_2, \dots, T_k)$ 'nin ortak olasılık yoğunluk fonksiyonu olmak üzere,

$$G(\mathbf{t}) = \int_{t_1}^{\infty} ds_1 \dots \int_{t_p}^{\infty} ds_p \{g(\mathbf{s})\}$$

$$f_j(t) = \int_t^{\infty} ds_1 \dots \int_t^{\infty} ds_{j-1} \int_t^{\infty} ds$$

$$p=2 \text{ için } f_j(t) = \int_t^{\infty} g(\mathbf{r}_j) ds, \mathbf{r}_1 = (t, s) \text{ ve } \mathbf{r}_2 = (s, t)$$

olur.

Buradan (3.3) ve (3.4) ile verilen nedene özel yoğunluk fonksiyonuna ise,

$$f_j(t) = - \left(\frac{\partial G(t_1, \dots, t_k)}{\partial t_j} \right)_{t_1 = \dots = t_k = t} \quad (3.26)$$

olarak geçilebilir. Ayrıca, (3.12) ile verilen nedene özel hazard fonksiyonu ise,

$$\lambda_j(t) = \frac{f_j(t)}{S(t)} = - \left(\frac{\partial \log G(t_1, \dots, t_k)}{\partial t_j} \right)_{t_1 = \dots = t_k = t} \quad (3.27)$$

olur.

(3.18) göz önüne alındığında;

$$S(t) = \exp \left\{ - \sum_{j=1}^k \Lambda_j(t) \right\} = \prod_{j=1}^k G_j^*(t) \quad (3.28)$$

$$G_j^*(t) = \exp \{ - \Lambda_j(t) \}, j = 1, \dots, k \quad (3.29)$$

yazılabilir.

Dikkat edilmelidir ki burada, $G_j^*(t)$ fonksiyonları, matematiksel olarak sürekli sağkalım fonksiyonları özelliklerini taşır fakat herhangi gözlenebilir bir rasgele değişkenin sağkalım fonksiyonu olmamaktadır (Lawless 2003, Lindqvist 2006). $G_j^*(t)$ ve $G_j(t)$ fonksiyonları ancak risklerin bağımsızlığı durumunda birbirine eşit olmaktadır. Bölüm 6’da bu konuya değinilecektir. Benzer biçimde,

$$h_j(t) = -\frac{G_j'(t)}{G_j(t)} = -\frac{g_j(t)}{G_j(t)} \quad (3.30)$$

olarak ifade edilen hazard fonksiyonu $h_j(t)$ ile daha önceki bölümlerde tanımlanan nedene özel hazard fonksiyonu $\lambda_j(t)$ birbirinden farklı anlamlar taşır. Burada, $h_j(t)$, j . risk dışındaki riskler ayrı tutularak yani izole edilerek elde edilen hazard fonksiyonu iken, $\lambda_j(t)$ diğer risklerin varlığında j . nedene ait hazard fonksiyonu olmaktadır.

Tüm $t > 0$ için, $h_j(t) = \lambda_j(t)$ eşitliği, T_j ’lerin bağımsızlığı durumunda geçerli olmaktadır.

3.3 Kesikli Başarısızlık Zamanları

Bazı durumlarda sağkalım süresi ya da başarısızlık süresi kesikli olarak düşünülebilir. Örneğin, bir sistem için bozulma, sistemin beklenen performansı yerine getirememesi anlamına gelmektedir. Kırılma, patlama gibi durumların meydana gelmesinden sonra, sistem beklendiği gibi çalışmayacaktır. Sürekli bir işleyişe sahip bir sistem göz önüne alınsın ve sistemin belirli aralıklarla kontrol edildiği düşünölsün. Sistemin parçalarından biri ya da birkaç kullanım ömrünün sonuna geldiğinde sistemin kullanım dışı olduđu varsayölsün. Gözlemler aynı türdeki bir işleyişin döngü sayısı olarak elde edilmektedir. Burada, sistemin ömrü, kullanım dışı olduđu tespit edilene kadar yapılan denetimlerin sayısıdır.

(T, C) ikilisindeki iki değişkenin de kesikli olduğu göz önüne alınırsa, bölüm 3.1’ de yapılan tanımlamalar yapılabilir.

T ’nin alacağı olası değerler, $0 = \tau_0 < \tau_1 < \dots < \tau_m$ (m sonlu ya da sonsuz olabilir) ve C ’nin alacağı değerler ise $\{1, 2, \dots, k\}$ olarak düşünölsün. Nedene özel kesikli dağılım ve nedene özel kesikli sağkalım fonksiyonları,

$$F_c(t) = P(T \leq t, C = c) \quad (3.31)$$

$$S_c(t) = P(T > t, C = c) \quad (3.32)$$

şeklinde ifade edilir.

Ayrıca,

$$S(0, C = c) = P(C = c) = \pi_c \quad (3.33)$$

yazılabilir.

T ’ye ait marjinal kesikli dağılım ve sağkalım fonksiyonu sırasıyla,

$$F(t) = P(T \leq t) = \sum_{c=1}^k F_c(t) \quad (3.34)$$

$$S(t) = P(T > t) = \sum_{c=1}^k S_c(t) \quad (3.35)$$

olarak ifade edilir.

Kesiki nedene özel olasılık fonksiyonları ve marjinal olasılık fonksiyonu ise aşğıdaki gibi tanımlanır. $l = 1, \dots, m$ için;

$$f_c(\tau_l) = \Pr(T = \tau_l, C = c) = S_c(\tau_l^-) - S_c(\tau_l) \quad (3.36)$$

$$f(\tau_l) = \sum_c f_c(\tau_l) \quad (3.37)$$

Eğer, $t \neq \tau_l$ ise $f_c(\tau_l) = 0$ 'dır. Aynı zamanda $f_c(0) = 0$ ve $l > 1$ için

$$S_c(\tau_l) = \sum_{s=l+1}^m f_c(\tau_s) = \pi_c - \sum_{s=1}^l f_c(\tau_s) \quad (3.38)$$

şeklinde tanımlanır.

Nedene özel kesikli hazard fonksiyonları ve marjinal hazard fonksiyonu $l = 1, \dots, m$ için aşağıdaki gibi tanımlanır

$$\begin{aligned} h_c(\tau_l) &= P(T = \tau_l, C = c | T \geq \tau_l) = \frac{P(T = \tau_l, C = c)}{P(T \geq \tau_l)} \\ &= \frac{f_c(\tau_l)}{P(T > \tau_l) + P(T = \tau_l)} = \frac{f_c(\tau_l)}{S(\tau_l) + S(\tau_l^-) - S(\tau_l)} \\ &= \frac{f_c(\tau_l)}{S(\tau_l^-)} \end{aligned} \quad (3.39)$$

$$h(\tau_l) = f(\tau_l) / S(\tau_l^-) = \sum_{c=1}^k h_c(\tau_l) \quad (3.40)$$

$l = 0$ için $h_c(\tau_0) = 0$ 'dır. Olası en son başarısızlık zamanı τ_m 'de $S(\tau_m) = 0$, $S(\tau_{m-1}) = f(\tau_m)$ ve $h(\tau_m) = 1$ değerlerini alır.

(2.47) göz önüne alınarak $l = 1, \dots, m$ için

$$S(\tau_l) = \prod_{s=0}^l \{1 - h(\tau_s)\} \quad (3.41)$$

$$f(\tau_l) = h(\tau_l) \prod_{s=0}^{l-1} \{1 - h(\tau_s)\} \quad (3.42)$$

olarak yazılabilir.

4. YARIŞAN RİSKLER TEORİSİNDE TAHMİNLER

4.1 Parametrik Tahminler

Bu bölümde, $T_1, \dots, T_k$ başarısızlık sürelerinin parametrik bir aileden geldiği varsayılarak, en çok olabilirlik fonksiyonunun elde edilmesi incelenecektir. Genel olarak, yarışan riskler teorisinde k tane başarısızlık riski ve bu risklere karşılık gelen potansiyel başarısızlık/sağkalım sürelerinin olduğu düşünülmektedir. Bu potansiyel sağkalım sürelerine birçok kaynakta net sağkalım süreleri adı verilir ve aşağıdaki gibi k -boyutlu bir vektör ile gösterilir.

$$\mathbf{T} = (T_1, T_2, \dots, T_k) \quad (4.1)$$

Buna göre bu değişkenlerin belirli bir parametrik forma sahip ortak olasılık yoğunluk fonksiyonu;

$$f(\mathbf{t}; \boldsymbol{\theta}) = f(t_1, \dots, t_k; \theta_1, \dots, \theta_m) \quad (4.2)$$

olarak ifade edilir.

Gözlenen sağkalım/başarısızlık süresi ise,

$$T = \min(T_1, \dots, T_k) \quad (4.3)$$

olur.

Burada, $P(T_i = T_j) = 0$, $i \neq j$ olduğu varsayılmaktadır.

(T, C) ikilisinin ortak dağılımı ise aşağıdaki gibidir.

$$\begin{aligned}
F_j(t) &= P(T \leq t, C = j) \\
&= P(T_j \leq t, T_j < T_i), \quad i \in \{1, \dots, j-1, j+1, \dots, k\}, i \in \{1, \dots, k\}, i \neq j \\
&= \int_0^t \left[\int_{t_j}^{\infty} \dots \int_{t_j}^{\infty} f(t_1, \dots, t_k) dt_1 dt_2 \dots dt_{j-1} dt_{j+1} \dots dt_k \right] dt_j
\end{aligned} \tag{4.4}$$

olarak elde edilir.

(T, C) ikilisinin ortak sağkalım fonksiyonu ise (3.2) ile ifade edildiği gibi bulunur.

T 'nin marjinal dağılımı ve marjinal sağkalım fonksiyonu sırasıyla aşağıdaki gibidir.

$$F_T(t) = \sum_{j=1}^k F_j(t) \tag{4.5}$$

$$S_T(t) = 1 - \sum_{j=1}^k F_j(t) \tag{4.6}$$

olarak ifade edilir.

Belirlenmiş bir j . için,

$$\pi_j = P(T = T_j), \quad j = 1, \dots, k \tag{4.7}$$

ifadesi başarısızlık nedeninin j . neden olması olasılığıdır. Bu olasılık

$$\pi_j = \int_{t_j=0}^{\infty} \int_{t_1=t_j}^{\infty} \dots \int_{t_{j-1}=t_j}^{\infty} \int_{t_{j+1}=t_j}^{\infty} \dots \int_{t_k=t_j}^{\infty} f(t; \theta) \prod_{i \neq j} dt_i dt_j \tag{4.8}$$

ile elde edilir (Moeschberger David 1971, Birnbaum 1979).

Y_j , başarısızlığının j . nedenden kaynaklandığı bilindiğindeki sağkalım süresi, $Y_j = T_j | C = j$ olsun. üzere, (3.9) göz önüne alındığında, bu sağkalım süresinin dağılım ve olasılık fonksiyonları aşağıdaki gibi bulunur.

$$F_Y(y_j) = P(Y_j \leq y_j) = P(T \leq y_j | C = j) = \frac{P(T \leq y_j, C = j)}{P(C = j)} = \frac{F_j(y_j)}{\pi_j}. \quad (4.9)$$

Bu ifadeden (4.4) ile türev alınarak;

$$f_{Y_j}(y_j; \theta) = \frac{1}{\pi_j} \int_{y_j}^{\infty} \dots \int_{y_j}^{\infty} f(t_1, \dots, t_{j-1}, y_j, t_{j+1}, \dots, t_k; \theta_1, \dots, \theta_m) dt_1 dt_2 \dots dt_{j-1} dt_{j+1} \dots dt_k \quad (4.10)$$

elde edilir.

N tane bireyin gözleendiği bir çalışmada, N_j tanesi j . nedenden dolayı başarısızlığa uğradığı göz önüne alınsın. N_j 'ler bilindiğinde gözlenen Y_{jr} , $r = 1, 2, \dots, N_j$ sağkalım sürelerinin ortak olasılık yoğunluk fonksiyonu;

$$\prod_{j=1}^k 1/\pi_j^{N_j} \prod_{r=1}^{N_j} \int_{y_{jr}}^{\infty} \dots \int_{y_{jr}}^{\infty} f(t_1, \dots, t_{j-1}, y_{jr}, t_{j+1}, \dots, t_k; \theta_1, \dots, \theta_N) dt_1 dt_2 \dots dt_j \quad (4.11)$$

olacaktır (Moeschberger ve David 1971).

(4.11) ile verilen ortak olasılık yoğunluk fonksiyonu N_j 'ye göre koşullandırılmıştır. Bir rasgele değişken olan N_j çokterimli dağılıma sahiptir. Buna göre

$$h(N_1, \dots, N_k) = \left(N! / \prod_{j=1}^{N_j} N_j! \right) \left(\prod_{j=1}^{N_j} \pi_j^{N_j} \right) \quad (4.12)$$

yazılır.

Sağkalım sürelerinin olabilirlik fonksiyonu;

$$L = L(..., y_{jr}, ...; \theta_1, ..., \theta_m) \quad (4.13)$$

$$L = \left(m! / \prod_{j=1}^k m_j! \right) \prod_{j=1}^k \prod_{r=1}^{m_j} \int_{y_{jr}}^{\infty} \dots \int_{y_{jr}}^{\infty} f(t_1, ..., t_{j-1}, y_{jr}, t_{j+1}, ..., t_k; \theta_1, ..., \theta_m) dt_1 dt_2 \dots dt_j \quad (4.14)$$

biçiminde yazılır (Moeschberger ve David 1971).

Dikkat edilmelidir ki, bu ifade, net sağkalım sürelerinin bağımsızlığı varsayımını gerektirmemektedir. Burada, sağlanması gereken tek varsayım, bu sürelerin ortak dağılımının belirli bir parametrik forma sahip olmasıdır. Parametrik aile bilindiğinde ve gözlem değerleri elde edildiğinde, olabilirlik fonksiyonu, $\theta_1, ..., \theta_m$ parametreleri için en çok olabilirlik tahmin edicilerinin elde edilmesi için kullanılabilir. Aşağıdaki şekilde ifade edilen kısmi türevlerin sıfıra eşitlenmesi ve bunun sonucunda elde edilecek m tane eşitliğin, $\theta_1, ..., \theta_m$ için çözülmesiyle istenen tahmin edicilere ulaşılır.

$$\frac{\partial L}{\partial \theta_s} \text{ (veya } \frac{\partial \ln L}{\partial \theta_s} \text{)} \quad (4.15)$$

Eğer, gerçek sağkalım süreleri bağımsız, olasılık yoğunluk fonksiyonu $f_j(t_j; \theta)$ ve sağkalım fonksiyonu $S_j(t_j; \theta)$ ise ortak olasılık yoğunluk fonksiyonu;

$$f_{Y_j}(y_j; \theta) = \frac{f_j(t_j; \theta)}{\pi_j S_j(t_j; \theta)} \prod_{l=1}^k S_l(t_l; \theta) \quad (4.16)$$

biçimini alır ve olabilirlik fonksiyonu;

$$L = \left(N! / \prod_{j=1}^k N_j! \right) \prod_{j=1}^k \prod_{r=1}^{N_j} \left[\frac{f_j(y_{jr}; \theta)}{S_j(y_{jr}; \theta)} \right] \prod_{l=1}^k [S_l(y_{jr}; \theta)] \quad (4.17)$$

olarak elde edilir ve bilinen yollarla en çok olabilirlik tahmin edicileri bulunur.

4.2 Sansürlemenin Varlığında Tahminler

Sağdan sansürlemenin var olduğu, n bireyden oluşan ve (T_i, C_i) , $i = 1, 2, \dots, n$ ikilisinden elde edilen bir rasgele örneklemden gözlemler alındığı düşünölsün. Bağımsız sansürleme varsayımı altında, olabilirlik fonksiyonu;

$$L = \prod_{i=1}^n f_{C_i}(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \quad (4.18)$$

şeklinde olacaktır.

Burada,

$$\delta_i = \begin{cases} 1, & t_i \text{ başarısızlık / sağkalım zamanı} \\ 0, & t_i \text{ sansürleme zamanı} \end{cases}$$

olarak tanımlanır.

T 'ye ait marjinal sağkalım fonksiyonu (3.24) ve (3.25) göz önüne alınarak aşağıdaki gibi tanımlanır.

$$S(t) = \exp \left\{ - \sum_{j=1}^k \Lambda_j(t) \right\} = \prod_{j=1}^k G_j(t) \quad (4.19)$$

Olabilirlik fonksiyonu, $\delta_{ij} = I(C_i = j)$ notasyonu kullanılarak yeniden yazılırsa;

$$\begin{aligned}
L &= \prod_{i=1}^n \prod_{j=1}^k f_j(t_i)^{\delta_{ij}} S(t_i)^{1-\delta_{ij}} \\
&= \prod_{i=1}^n \prod_{j=1}^k [\lambda_j(t_i) S(t_i)]^{\delta_{ij}} S(t_i)^{1-\delta_{ij}} \\
&= \prod_{i=1}^n \prod_{j=1}^k g_j(t_i)^{\delta_{ij}} G_j(t_i)^{1-\delta_{ij}}
\end{aligned} \tag{4.20}$$

olarak elde edilir.

Son olarak;

$$L = \prod_{i=1}^n \prod_{j=1}^k g_j(t_i)^{\delta_{ij}} G_j(t_i)^{1-\delta_{ij}} \tag{4.21}$$

biçimindeki olabilirlik fonksiyonunun çarpım sırası ters çevrilirse; k tane olabilirlik fonksiyonunun çarpımı olduğu görülür. Bunlar, j . başarısızlık nedenine karşılık gelen sansürlenmiş örnekleme temsil eden olabilirlik fonksiyonlarıdır. Çarpımın j . kısmı;

$$L_j = \prod_{i=1}^n g_j(t_i)^{\delta_{ij}} G_j(t_i)^{1-\delta_{ij}} \tag{4.22}$$

olarak yazılır.

(4.22) ile gösterilen olabilirlik fonksiyonu $\lambda_j(t)$ veya $\Lambda_j(t)$ fonksiyonlarının (T, C) ikilisinden elde edilen gözlemlerden tahmin edilebileceğini göstermektedir. Eğer, $\lambda_j(t)$ fonksiyonları farklı θ_j , $j=1, \dots, k$ parametrelerine sahip ise $L(\theta_1, \dots, \theta_k)$, $L_j(\theta)$ olarak parçalanabilir. Verilen j için, θ_j 'ye ilişkin çıkarım (4.22) ile gösterilen olabilirlik fonksiyonuna dayanmaktadır. Burada, t_i . zamanda j . sebepten meydana gelen başarısızlık, ilgilenilen başarısızlık; diğer sebeplerden meydana gelen başarısızlık ise sansürlenmiş gözlem olarak ele alınır.

Bu tipteki modeller için parametrik çıkarımlar standart sağkalım analizi yöntemleri kullanılarak yapılabilir. Benzer olarak, parametrik olmayan ve yarı parametrik yöntemler bu modeller için $\lambda_j(t)$ 'ye dayanarak uygulanabilir.

Daha önce sözü edilen (T_i, C_i) ikilisinden elde edilen sansürlenmiş örnekleme göz önüne alalım. Genel olarak tanımlanan olabilirlik fonksiyonu, her bir nedene özgü olarak (4.22) ile gösterildiği biçimde parçalanabilmektedir. (4.22); sağkalım fonksiyonu $G_j(t)$, yoğunluk fonksiyonu $g_j(t)$ ve hazard fonksiyonu $\lambda_j(t)$ olan rasgele örneklemin olabilirlik fonksiyonudur. Buna göre; (t_i, δ_{ij}) , $j=1,2,\dots,n$ 'den elde edilen veriye dayanarak $G_j(t)$ Kaplan-Meier yöntemi kullanılarak tahmin edilebilir. $G_j(t)$ fonksiyonu yerine sebebe özel birikimli hazard fonksiyonunu, $\Lambda_j(t)$, göz önüne almak daha uygundur. $\Lambda_j(t) = -\log G_j(t)$ şeklindedir ve K-M yöntemi ile tahmin edilebilmektedir ancak bu fonksiyonun tahmini için, bölüm 2.5'te anlatılan ve (2.50) ile gösterilen Nelson-Aalen tahminini kullanmak daha yaygındır.

$$\hat{\Lambda}_j(t) = \sum_{i:t_i \leq t} \frac{\delta_{ij}}{n_i}, \quad j=1,2,\dots,k \quad (4.23)$$

$Y_l(t) = I(t_l \geq t)$, $\delta_{ij} = I(C_i = j)$ olarak tanımlanır ve $n_i = \sum_{l=1}^n Y_l(t_i)$; t_i 'den önce yaşayan ve sansürlenmemiş birey sayısını ifade eder, varyans tahmini ise,

$$\hat{Var}[\hat{\Lambda}_j(t)] = \sum_{i:t_i \leq t} \frac{\delta_{ij}}{n_i^2} \quad (4.24)$$

olarak verilir.

T 'nin marjinal sağkalım fonksiyonu $S(t)$, K-M tahmini kullanılarak (ilgili sebebin göz ardı edilmesiyle), (t_i, δ_i) , $i=1,\dots,n$ verisine dayanarak elde edilir.

$t_{(1)} < \dots < t_{(k)}$ başarısızlık zamanlarını, d_i ; $t_{(i)}$ zamanındaki başarısızlık sayısını ve n_i , $t_{(i)}$ zamanındaki risk altında bulunan birey sayısını göstermek üzere marjinal sağkalım fonksiyonunun K-M tahmini aşağıdaki gibidir.

$$\hat{S}(t) = \prod_{i: t_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i} \right) \quad (4.25)$$

$S(t)$ 'nin alternatif tahmini aşağıdaki gibi verilir.

$$\hat{S}(t) = \exp[-\hat{\Lambda}(t)] = \exp \left[- \sum_{j=1}^k \hat{\Lambda}_j(t) \right] \quad (4.26)$$

$\hat{\Lambda}_j(t)$ için yukarıda (4.18) ile verilen Nelson-Aalen tahmini kullanılır.

(3.5) ile gösterilen nedene özel dağılım fonksiyonu ifadesi göz önüne alındığında bu fonksiyonun tahmin edicisi

$$\hat{F}_j(t) = \int_0^t \hat{S}(u) d\hat{\Lambda}_j(u) \quad (4.27)$$

olur ve buradan,

$$\hat{F}_j(t) = \sum_{i: t_i \leq t} \hat{S}(t_i) \frac{\delta_{ij}}{n_i}, \quad j = 1, 2, \dots, k \quad (4.28)$$

elde edilir.

Eğer, sadece bir tane başarısızlık nedeni var ise, (4.28) ile gösterilen ifade, $\hat{S}(t)$ 'nin (4.25) ile gösterildiği haliyle $1 - \hat{S}(t)$ ifadesine indirgenir. Eğer, $k \geq 2$ tane başarısızlık nedeni var ve sansürleme yok ise $\hat{F}_j(t)$, $t_i \leq t$ ve $C_i = j$ olan bireylerin dağılımlarının

arpımıdır. Yani, j . başarısızlık nedeni iin ampirik (deneysel) nedene zel daėılım fonksiyonlarının arpımıdır.

$F_j(t)$, nedene zel daėılım fonksiyonu bazı alıřma alanlarında birikimli sıklık fonksiyonu olarakta bilinir. Uygulamalarda nemli bir yere sahip olan bu fonksiyona sonraki blmde yer verilmiřtir.

5. BİRİKİMLİ SIKLIK FONKSİYONU

Bölüm 3.1’de (3.1) ve (3.15) ile gösterilen, $F_j(t)$, nedene özel dağılım fonksiyonları demografi, epidemiyoloji biyoistatistik literatüründe birikimli sıklık (cumulative incidence-CI) fonksiyonları olarak adlandırılmaktadır. Farklı tedavi uygulanan hasta grupları için birikimli sıklık fonksiyonlarının karşılaştırılmasıyla elde edilen sonuçlar hasta için en uygun tedavinin seçiminde çok yararlı bilgiler vermektedir (Gray 1988). Birikimli sıklık fonksiyonları (T, C) ’den elde edilen veriden tahmin edilebilmektedir (Prentice vd. 1978). $\lambda_j(t)$, nedene özel hazard fonksiyonları da (T, C) ’den elde edilen veriden tahmin edilebilir fakat bu fonksiyonlar her bir başarısızlık riskine maruz kalan hastaların oranını doğrudan göstermemektedir (Pepe ve Mori 1993). Bu nedenle birikimli sıklık fonksiyonu ile çalışmak sonuçların yorumlanması açısından daha yararlı olmaktadır.

Bir olayın olayın ortaya çıkış sıklığı medikal araştırmalarda ilgilenilen konulardan biri olduğundan, birikimli sıklık fonksiyonunun tahmin edilmesi ve yorumlanması da önemli olmaktadır. K-M yöntemi, sağkalım fonksiyonu ve onun tümleyeni $(1 - S(t))$ olarak birikimli sıklık fonksiyonunun tahmininde yaygın olarak kullanılan bir yöntemdir. Birden fazla başarısızlık riskinin var olduğu ve bu risklerin bağımsız olmadığı durumlarda, birikimli sıklık fonksiyonunun tahmin edilmesinde K-M yönteminin kullanılması yanlış tahminler vermektedir (Kim 2007). Çünkü K-M yöntemi sadece ilgilenilen başarısızlık riskini dikkate almakta ve diğer başarısızlık risklerini sansürlü gözlem olarak değerlendirmektedir. Bu nedenle K-M tahminleri birikimli sıklık fonksiyonunun gerçek değerinin üzerinde tahminler olmaktadır (Satagopan vd. 2004, Kim 2007).

5.1 Yarışan Risklerin Varlığında Birikimli Sıklık Fonksiyonunun Tahmini

Bu bölümde, birikimli sıklık fonksiyonun yarışan risklerin varlığında parametrik olmayan tahmininin K-M ve yarışan riskler yöntemleri kullanılarak elde edilmesi

üzerinde durulacaktır. Bu iki yöntem Kim (2007) çalışmasında yer alan varsayımsal gözlem değerleri ve konuyla ilgili oluşturduğu senaryo kullanılarak ele alınacaktır.

Örnek 5.1. Daha önce yapılmış olan bir kemik iliği nakli (HSCT; allogeneic hematopoietic stem cell transplantation) çalışmasında, ortaya çıkması gözlenen başarısızlığın ilgili kan hastalığının nüks etmesi sonucu ölüm olarak belirlenmiştir. Ayrıca; nakil sonrasında oluşabilecek komplikasyonların da (TRM; transplant-related mortality) hastalığın nüks etmesi kadar ölümcül olduğu göz önüne alınmaktadır. Buna senaryoda, konuyla ilgili kan hastalığına sahip 10 hastanın farklı zamanlarda kliniğe yattığı düşünülmüş ve bu hastaların kliniğe yatış zamanından belirlenmiş bir zamana kadar takip edildiği varsayılmıştır. Takip süresi sonunda farklı başarısızlık zamanlarının elde edildiği göz önüne alınmıştır. Ortaya çıkması gözlenen olay, hastalığın nüks etmesi sonucu ölüm olarak belirlenen bu çalışmada nüks dışındaki herhangi bir nedenden meydana gelen ölümler ya da bireyin sağ kalması durumları sansürlü gözlem olarak kaydedilmiştir. Ayrıca, takip süresi boyunca herhangi bir nedenden dolayı takip edilememiş bireyler sağ olarak düşünülmüş ve nüks meydana gelmemiş hastalar takip sonunda sağ olduğundan sansürlü gözlem olarak kabul edilmiştir.

5.1.1 K-M yöntemi

K-M yöntemi (Kaplan ve Meier 1958) bölüm 2.2’de kapsamlı olarak anlatılmış ve örnek 2.1 ile örneklendirilmiştir. Buna göre,

$t_0 = 0$, çalışmanın başlangıç zamanı, t_1 en küçük gözlem ve t_{10} en büyük gözlem olarak kaydedilsin. t zamanına kadar nüks nedeniyle ölümü gerçekleşmemiş bir hastanın t zamanından sonra da yaşaması olasılığının K-M tahmini $S_{KM}^{NüksYok}(t)$ ile gösterilmiştir.

Birikimli sıklık fonksiyonunun K-M yöntemiyle tahmini, yani en fazla t zamanına kadar nüks nedeniyle ölüm gerçekleşme olasılığı ise $CI_{KM}^{Nüks}(t)$ ile gösterilmiş ve aşağıdaki gibi hesaplanmıştır.

$CI_{KM}^{Nüks}(t_0) = 1 - S_{KM}^{NüksYok}(t_0) = 0 \Rightarrow$ çünkü, çalışmanın başında tüm bireyler sağ ve nüks gerçekleşmemiş

$$CI_{KM}^{Nüks}(t_1 = 10) = 1 - S_{KM}^{NüksYok}(t_1 = 10) = 1 - 0.9 = 0.1$$

$$CI_{KM}^{Nüks}(t_1 = 10) = CI_{KM}^{Nüks}(t_0 = 0) + S_{KM}^{NüksYok}(t_0 = 0) \left(\frac{t_1 \text{ zamanındaki nüks sayısı}}{t_1 \text{ zamanında önce risk altındaki birey sayısı}} \right)$$

$$CI_{KM}^{Nüks}(t_1 = 10) = 0 + 1 \left(\frac{1}{10} \right) = 0.1$$

$$\begin{aligned} CI_{KM}^{Nüks}(t_2 = 20) &= 1 - S_{KM}^{NüksYok}(t_2 = 20) = 1 - 0.90 = 0.1 \\ &= CI_{KM}^{Nüks}(t_1 = 10) + S_{KM}^{NüksYok}(t_1 = 10) \left(\frac{t_2 \text{ zamanındaki nüks sayısı}}{t_2 \text{ zamanında önce risk altındaki birey sayısı}} \right) \end{aligned}$$

$$CI_{KM}^{Nüks}(t_2 = 20) = 0 + 1 \left(\frac{1}{10} \right) = 0.1$$

N

$$\begin{aligned} CI_{KM}^{Nüks}(t_{10} = 90) &= 1 - S_{KM}^{NüksYok}(t_{10} = 90) = 1 - 0.315 = 0.685 \\ &= CI_{KM}^{Nüks}(t_9 = 80) + S_{KM}^{NüksYok}(t_9 = 80) \left(\frac{t_{10} \text{ zamanındaki nüks sayısı}}{t_{10} \text{ zamanında önce risk altındaki birey sayısı}} \right) \end{aligned}$$

$$CI_{KM}^{Nüks}(t_{10} = 90) = 0.685 + 0.315 \left(\frac{0}{1} \right) = 0.685$$

Tüm hesaplamalar çizelge 5.1 ile özetlenmiştir.

Böyle bir çalışma düzeninde, hastalığın nüksü ve komplikasyonların oluşması birbirinden bağımsız olaylar değildir. Çünkü bu iki olay da nakilden sonraki bağışıklık mekanizmasıyla ilgilidir. Komplikasyonun azaltılması yönündeki çabalar, hastalığın

nüks etmesine neden olduğu gibi nüks etmemesi için uygulanan tedaviler farklı komplikasyonları meydana getirebilmektedir.

Birikimli sıklık fonksiyonunu K-M yöntemi kullanarak tahmin etmek tek bir başarısızlık riski varken (hastalığın nüks etmesi) anlamlı ve kullanışlı olmaktadır. Fakat, nüks dışındaki başka risklerin (TRM, komplikasyon) meydana gelmesi de ilgilenilen konu olduğunda, K-M yöntemi bu riskler arasındaki bağımlılığı yansıtmayacaktır. Çünkü, diğer riskler sonucu meydana gelen ölümler sansürlü gözlem olarak düşünülecektir (Kaplan ve Meier 1958).

Çizelge 5.1 K-M yöntemiyle hesaplanan sağkalım olasılıkları

Hasta No	İzlem süresi	Olay	Risk Altındaki Birey Sayısı	Nüks Sayısı	Sansürlü Gözlem Sayısı	K-M Yöntemi	
						$S_{KM}^{NüksYok}$	$CI_{KM}^{Nüks}$
1	10	Nüks	10	1	0	$1 \times (9/10) = 0.9$	0.1
2	20 ⁺	Sağ	9	0	1	$0.9 \times (9/9) = 0.9$	0.1
3	35	Nüks	8	1	0	$0.9 \times (7/8) = 0.7875$	0.2125
4	40	TRM	7	0	1	$0.7875 \times (7/7) = 0.7875$	0.2125
5	50 ⁺	Sağ	6	0	1	$0.7875 \times (6/6) = 0.7875$	0.2125
6	55	Nüks	5	1	0	$0.7875 \times (4/5) = 0.63$	0.37
7	70	TRM	4	0	1	$0.63 \times (4/4) = 0.63$	0.37
8	71	TRM	3	0	1	$0.63 \times (3/3) = 0.63$	0.37
9	80	Nüks	2	1	0	$0.63 \times (1/2) = 0.315$	0.685
10	90	Sağ	1	0	1	$0.315 \times (1/1) = 0.315$	0.685

5.1.2 Yarışan riskler yöntemi

Hastalığın nüks etmesi nedeniyle meydana gelen ölümler ilgilenilen asıl sonuç iken ölüme neden olan diğer bir risk olan komplikasyonun (TRM) yarışan risk olarak göz önüne alındığı durumlarda birikimli sıklık fonksiyonunun tahmini 5.1.1 ile verilen K-M yöntemine benzer şekilde elde edilir. İki yöntem arasındaki farklılık, t zamanına kadar nüks görülmemiş bir hastada t zamanından sonra da nüks görülmemesi olasılığının $(S_{KM}^{NüksYok}(t))$ yerine, $(S_{KM}^{NüksYok,Komp.Yok}(t))$ nüksün ve komplikasyonun meydana gelmeme olasılığının kullanılmasıdır (Prentice 1978, Kim 20007). Komplikasyonun varlığında, t zamanına kadar nüks görülmesi olasılığı $CI_{CR}^{Nüks}(t)$ olarak gösterilir ve aşağıdaki gibi hesaplanır.

$$CI_{CR}^{Nüks}(t_0) = 0$$

$$CI_{CR}^{Nüks}(t_1) = CI_{CR}^{Nüks}(t_0) + S_{KM}^{NüksYok,Komp.Yok}(t_0) \left(\frac{t_1 \text{ zamanındaki nüks sayısı}}{t_1 \text{ zamanında önce risk altındaki birey sayısı}} \right)$$

$$CI_{CR}^{Nüks}(t_1 = 10) = 0 + 1 \left(\frac{1}{10} \right) = 0.1$$

$$CI_{CR}^{Nüks}(t_2) = CI_{CR}^{Nüks}(t_1) + S_{KM}^{Nüks,Komp.Yok}(t_1) \left(\frac{t_2 \text{ zamanındaki nüks sayısı}}{t_2 \text{ zamanında önce risk altındaki birey sayısı}} \right)$$

$$CI_{CR}^{Nüks}(t_1 = 20) = 0.1 + 0.9 \left(\frac{0}{10} \right) = 0.1$$

N

$$CI_{CR}^{Nüks}(t_{10} = 90) = 0.4825 + 0.135 \left(\frac{0}{1} \right) = 0.4825$$

Hesaplamalar çizelge 5.2 ile özetlenmiştir. Çizelgede, hastalığın nüks etmesi yarışan risk olarak düşünüldüğünde birikimli sıklık fonksiyonunun tahminlerine de yer verilmiştir.

Çizelge 5.2 Yarışan riskler yöntemiyle göre hesaplanan sağkalım olasılıkları

Nüks ya da TRM sayısı	Sansürlü gözlem sayısı	Nüksün Meydana Gelmeme Olasılığı	Yarışan Risk Yöntemi	
			Nüks Yarışan Risk iken CI Tahmini	TRM Yarışan Risk iken CI Tahmini
1	0	$1 \times (9/10) = 0.9$	$0 + 1 \times (1/10) = 0.1$	0
0	1	$0.9 \times (9/9) = 0.9$	$0.1 + 0.9 \times (0/9) = 0.1$	0
1	0	$0.9 \times (7/8) = 0.7875$	$0.1 + 0.9 \times (1/8) = 0.2125$	0
1	0	$0.7875 \times (6/7) = 0.675$	$0.2125 + 0.7875 \times (0/7) = 0.2125$	0.1125
0	1	$0.675 \times (6/6) = 0.675$	$0.2125 + 0.675 \times (0/6) = 0.2125$	0.1125
1	0	$0.675 \times (4/5) = 0.54$	$0.2125 + 0.675 \times (1/5) = 0.3475$	0.1125
1	0	$0.54 \times (3/4) = 0.405$	$0.3475 + 0.54 \times (0/4) = 0.3475$	0.2475
1	0	$0.405 \times (2/3) = 0.27$	$0.3475 + 0.405 \times (0/3) = 0.3475$	0.3825
1	0	$0.27 \times (1/2) = 0.135$	$0.3475 + 0.27 \times (1/2) = 0.4825$	0.3825
0	1	$0.135 \times (0/1) = 0.135$	$0.4825 + 0.135 \times (0/1) = 0.4825$	0.3825

Yarışan riskler yönteminde, K-M yönteminin aksine, kompliasyon sonucu ölen hastalarda başarısızlık olarak sayılmaktadır. Sadece, sağ kalan bireyler nüksün meydana gelmesi açısından risk altında olarak düşünülmektedir. Bu nedenle, K-M yönteminde hesaplanan $S_{KM}^{NüksYok}(t)$ olasılıkları her zaman, yarışan riskler yönteminde hesaplanan $S_{KM}^{NüksYok,Komp.Yok}$ olasılıklarından daha büyük olmaktadır (Gooley vd. 1999).

6. YARIŞAN RİSKLER TEORİSİNDE BELİRLENEBİLME PROBLEMİ

Yarışan risklerin var olduğu durumları ifade etmek için yaygın olarak örtük zaman gösterimi kullanılmaktadır. $(T_1, T_2, \dots, T_k)$ 'nin k tane riske maruz kalan bir bireyin (ya da herhangi bir gözlem biriminin) her bir riske karşılık gelen potansiyel sağkalım zamanlarının rasgele bir vektörü ve gözlenebilen sağkalım süresinin ise $T = \min(T_1, T_2, \dots, T_k)$ olduğuna bölüm 3.2'de değinilmiştir.

Yarışan riskler analizinde genellikle amaç, $(T_1, T_2, \dots, T_k)$ örtük başarısızlık zamanlarına ait marjinal dağılımların ve ortak dağılımın elde edilmesi olabilmektedir. Bunun için, geleneksel yaklaşım, sağkalım süresini etkilediği düşünülen risklerin varlığında gözlem alıp, bu risklerin bir kısmı izole edilerek çıkarım yapma yönündedir. Bu girişim, $S_j(t)$, nedene özel sağkalım fonksiyonlarının yardımıyla $G_j(t)$, marjinal sağkalım fonksiyonlarını belirlemek olarak ifade edilir. Risklerin bağımsızlığı durumunda, marjinal dağılımların ve ortak dağılımların belirlenmesi aşamasında bir sorun yoktur. Bağımsızlık durumunda, marjinal dağılımların çarpımıyla ortak dağılım elde edilmektedir. Yarışan riskler teorisindeki belirlenebilme problemi göz önüne alındığında, $G_j(t)$ için parametrik bir model varsayımı gibi bir takım kısıtlamalar yapmadan, gözlemler yardımıyla elde edilen nedene özel fonksiyonları kullanılarak marjinal dağılımların belirlenmesi mümkün olmamaktadır.

Belirlenebilme problemi, gözlemlerden elde edilen (T, C) ikilisinin ortak dağılımının $(S_j(t)$ veya $F_j(t))$, $(T_1, T_2, \dots, T_k)$ örtük sağkalım sürelerinin ortak dağılımını $(G(t))$ belirlememesi olarak ifade edilir (Cox 1959, Tsiatis 1975, Crowder 1991). Yani, aynı nedene özel dağılım fonksiyonuna $(S_j(t)$ veya $F_j(t))$ sahip $(T_1, T_2, \dots, T_k)$ 'ye ait birden fazla ortak dağılım fonksiyonu vardır. $k=2$ için Cox (1959)'ın ifade ettiği belirlenebilme problemi, Tsiatis (1975) tarafından genelleştirilmiştir.

Eğer bağımlı bir model için nedene özel sağkalım fonksiyonları $S_j(t)$ ise, aynı nedene özel sağkalım fonksiyonlarına sahip (6.2) ile gösterildiği gibi ifade edilen bir tek bağımsız model vardır.

$$G^*(t) = \prod_{j=1}^k G_j^*(t) \quad (6.1)$$

$G_j^*(t)$, (3.29) ile verildiği gibi

$$G_j^*(t) = \exp \left\{ - \int_0^t \lambda_j(s) ds \right\} \quad (6.2)$$

biçimindedir.

Bu durum (T, C) gözlemlerine dayanarak hangi modelin doğru olduğuna karar vermeyi olanaksız kılar. Aşağıdaki örnekler bu duruma örnek olarak verilebilir.

Örnek 6.1 Gumbel (1960), iki değişkenli üstel dağılım için önerdiği üç farklı formdan ilki, kolaylığı açısından bu örnekte kullanılmıştır. Bu dağılım iki değişkenli üstel dağılım R_1 ve R_2 olarak adlandırılan, $k=2$ tane başarısızlık nedeninin varlığında, (T_1, T_2) potansiyel başarısızlık zamanlarının gerçek ortak dağılımını gösterebilir.

$$G(t_1, t_2) = \exp \{ -\alpha_1 t_1 - \alpha_2 t_2 - \nu t_1 t_2 \} \quad (6.3)$$

Burada $\alpha_1 > 0$, $\alpha_2 > 0$ ve bağımlılık parametresi olan ν ise $0 \leq \nu \leq \lambda_1 \lambda_2$ koşulunu sağlar. $\nu = 0$ olması, T_1 ve T_2 arasındaki bağımsızlığı belirtir. R_1 ve R_2 'nin başarısızlık nedenlerinden, herhangi birinin ortaya çıkması gözlenen başarısızlığın tek nedeni olduğu durumdaki (3.22) ile gösterilen gerçek sağkalım fonksiyonları aşağıdaki gibidir.

$$G_1(t) = \exp\{-\alpha_1 t\}, \quad \alpha_1 > 0 \quad (6.4)$$

$$G_2(t) = \exp\{-\alpha_2 t\}, \quad \alpha_2 > 0 \quad (6.5)$$

(3.6) ile gösterilen, (T, C) ikilisinin dağılımından elde edilen T 'nin marjinal sağkalım fonksiyonu ise

$$S_T(t) = \exp\{-(\alpha_1 + \alpha_2)t - \nu t^2\} \quad (6.6)$$

olarak ifade edilir.

$j = 1, 2$ için (3.3) ya da (3.4) ile gösterilen nedene özel yoğunluk fonksiyonları

$$f_j(t) = \left[(\alpha_j + \nu t_{3-j}) G(t) \right]_{(t_1=t_2=t)} = (\alpha_j + \nu t) S_T(t) \quad (6.7)$$

olarak bulunur. (3.2) ile tanımı verilen nedene özel sağkalım fonksiyonları ise

$$\begin{aligned} S_j(t) &= \int_t^\infty (\alpha_j + \nu s) \exp\{-(\alpha_1 + \alpha_2)s - \nu s^2\} ds \\ &= \frac{1}{2} \left[-\exp\{-(\alpha_1 + \alpha_2)s - \nu s^2\} \right]_t^\infty + \frac{1}{2} (\alpha_j - \alpha_{3-j}) \int_t^\infty \exp\{-(\alpha_1 + \alpha_2)s - \nu s^2\} ds \\ &= \frac{1}{2} \exp\{-(\alpha_1 + \alpha_2)t - \nu t^2\} + \frac{1}{2} (\alpha_j - \alpha_{3-j}) (\pi/\nu)^{1/2} \Phi\{-(2\nu)^{1/2}(t + (\alpha_1 + \alpha_2)/2\nu)\} \end{aligned} \quad (6.8)$$

şeklinde elde edilir. İfade de, Φ , standart normal dağılım fonksiyonudur.

$\alpha_1 = \alpha_2$ ise,

$$S_j(t) = \frac{1}{2} \exp\{-(\alpha_1 + \alpha_2)t - \nu t^2\} \quad (6.9)$$

şeklinde nedene özel sağkalım fonksiyonunun daha basit bir formuna ulaşılır. (3.12) ile tanımlanan nedene özel hazard fonksiyonu ise

$$\lambda_j(t) = \alpha_j + \nu t \quad (6.10)$$

olarak bulunur ve buradan (3.16) ile gösterilen T 'ye ilişkin marjinal hazard fonksiyonu

$$\lambda_T(t) = (\alpha_1 + \alpha_2) + 2\nu t \quad (6.11)$$

şeklinde elde edilir.

Bağımlı modele karşılık gelen $S_j(t) = \frac{1}{2} \exp(-(\alpha_1 + \alpha_2)t - \nu t^2)$, $j = 1, 2$ olarak bulundu.

Bu fonksiyonun gözlemlerden elde edilebildiğine daha önce değinilmişti. Tsiatis (1975) 'in çalışmasında belirttiği üzere, bu fonksiyona karşılık gelen bağımsız model,

$$G_j^*(t) = \exp \left\{ - \int_0^t (\alpha_j + \nu s) ds \right\} = \exp \left\{ - (\alpha_j t - \nu t^2 / 2) \right\} \quad (6.12)$$

$$G^*(t) = \prod_{j=1}^k G_j^*(t) = \exp \left\{ - (\alpha_1 t - \nu t^2 / 2) - (\alpha_2 t - \nu t^2 / 2) \right\} \quad (6.13)$$

olacaktır.

Görülmektedir ki, risklerin bağımlı olması durumunda, bağımlı (6.3) ve bağımsız model (6.13) birbirinin aynısı olmaktadır. Dolayısıyla, $S_j(t)$ fonksiyonundan yani gözlemlerden yola çıkarak hangi modelin doğru olduğunu söylemek olanaksızdır. Gerçek marjinal dağılım $G_j(t) = \exp(-\alpha_j t)$, $j = 1, 2$ iken gözlemlerden elde edilen dağılım $S_j(t) = \frac{1}{2} \exp(-(\alpha_1 + \alpha_2)t - \nu t^2)$, $j = 1, 2$ olmaktadır. Dolayısıyla $S_j(t)$ ne marjinal dağılımı ($G_j(t)$) ne de ortak dağılımı ($G(t_1, t_2, \dots, t_k)$) belirlememektedir.

Örnek 6.2 İki boyutlu bir sağkalım fonksiyonu (6.14) ile verilmiş olsun.

$$G(t_1, t_2) = \exp\{1 - \alpha_1 t_1 - \alpha_2 t_2 - e^{\beta(\alpha_1 t_1 + \alpha_2 t_2)}\} \quad (6.14)$$

Burada, $\alpha_1, \alpha_2 > 0$ ve β , T_1 ve T_2 arasındaki bağımlılığı ölçmektedir. (3.27) kullanılarak $\lambda_j(t)$

$$\lambda_j(t) = \alpha_j(1 + \beta e^{\beta(\alpha_1 + \alpha_2)t}) \quad (6.15)$$

olarak elde edilir. Nedene özel hazard fonksiyonu da, nedene özel dağılım ya da sağkalım fonksiyonu gibi gözlem sonu elde edilen fonksiyonlardır.

İkinci bir bağımlı model aşağıdaki gibi verilmiş olsun.

$$G(t_1, t_2) = \exp\left\{1 - \alpha_1 t_1 - \alpha_2 t_2 - \frac{\alpha_{12} e^{\alpha_{12}(\alpha_1 + \alpha_2)t_1} + \alpha_{12} e^{\alpha_{12}(\alpha_1 + \alpha_2)t_2}}{\alpha_1 + \alpha_2}\right\} \quad (6.16)$$

Yine bu iki boyutlu sağkalım fonksiyonu için de, $\alpha_1, \alpha_2 > 0$ dır ve β , T_1 ve T_2 arasındaki bağımlılığı ifade eder. (3.27) kullanılarak bu sağkalım fonksiyonu için de aşağıdaki nedene özel hazard fonksiyonu elde edilir.

$$\lambda_j(t) = \alpha_j(1 + \beta e^{\beta(\alpha_1 + \alpha_2)t}) \quad (6.17)$$

Görülmektedir ki, farklı iki bağımlı model aynı nedene özel hazard fonksiyonuna sahiptir. Yani, gözlemlere dayanarak T_1 ve T_2 'nin ortak dağılımı belirlenmemektedir.

Her iki örnekten görüldüğü üzere, (T, C) ikilisinin dağılımı, $(T_1, T_2, \dots, T_k)$ örtük sağkalım sürelerinin ortak dağılımını belirlememektedir. Belirlenebilme problemi olarak adlandırılan bu problem yarışan risklerin varlığında bağımlılığı modelleme konusunda sıkıntılar yaratmaktadır.

Yarışan riskler teorisinde klasik bir varsayım olarak risklerin bağımsız oldukları düşünülmektedir. Bu varsayımdan hareketle, T_j 'nin marjinal sağkalım fonksiyonu ($G_j(t)$), nedene özel sağkalım fonksiyonları ($S_j(t)$) kullanılarak hesaplanabilmektedir. Bağımsızlık durumunda (3.21) ile gösterildiği gibi

$$G(t_1, t_2, \dots, t_k) = \prod_{j=1}^k G_j(t) \quad (6.18)$$

yazılabilmektedir. Bu eşitlikten yola çıkarak,

$$G_j(t) = G_j^*(t) \quad (6.19)$$

ve

$$\lambda_j(t) = h_j(t) \quad (6.20)$$

olmaktadır.

Literatürde belirlenebilme sorununa ilişkin çalışmalar mevcuttur. Nadar (1971) iki değişkenli normal dağılımda, Arnold ve Brockett (1983), iki değişkenli Makeham ve pareto dağılımının iki versiyonu için belirlenebilme probleminin olmadığını göstermiştir. Ayrıca, çalışmalarında dağılımların parametrelerinin açık çözümlerini de vermişlerdir. Zheng ve Klein (1995), marjinal dağılımlar arasındaki bağımlılığı ifade eden kapula fonksiyonunun bilinmesi halinde, yarışan risklerin varlığında elde edilen verinin marjinal dağılımları belirlemede yeterli olacağını göstermiştir.

Çalışmanın ilerleyen bölümlerinde, başarısızlığı ortaya çıkaran riskler arasındaki bağımlılık örtük sağkalım süreleri arasındaki bağımlılık olarak ifade edilmiş ve bu bağımlılık yapısı Archimedean kapula fonksiyonları yardımıyla modellenmiştir.

7. KAPULA FONKSİYONLARI VE ARCHİMEDEAN KAPULA AİLESİ

Sağkalım zamanlarının bağımsız rasgele değişkenler olarak düşünüldüğü yarışan riskler modeli üzerine, bioistatistik, ekonometri, medikal, demografik ve aktüerya literatüründe birçok araştırmacı tarafından çalışmalar ortaya konmuştur.

Sağkalım literatüründeki bazı çalışmalarda değinildiği gibi $T_1, T_2, \dots, T_k$ ile gösterilen sağkalım süreleri (başarısızlık zamanları) bağımlı olmaya daha eğilimlidir (Carriere 1994, Valdez 2001). Örneğin, Weir (2005)'in kronik böbrek rahatsızlığı (chronic kidney disease-CKD) olan hastalarda, CKD ile kardiyovasküler rahatsızlıklar (cardiovascular disease CVD) arasındaki ilişkiyi araştırdığı çalışmasında, CKD'nin varlığının CVD riskini arttırdığını gözlemlemiştir. Yani, hasta için CVD riskinden ölümün meydana gelmesi, CKD riskinden ölümün meydana gelmesinden bağımsız değildir. Aynı şekilde, Lobo (2008) çalışmasında, Alzheimer hastalığı, diyabet, kardiyovasküler hastalıklar arasındaki ilişkileri ortaya çıkarmak amacıyla, genler arasındaki etkileşimleri (bağımlılık yapısı) araştırmıştır. Tıp literatürdeki bu tipteki gibi çalışmaların amacı, belli bir hastalıktan meydana gelebilecek ölümleri azaltmanın ya da tamamen ortadan kaldırmanın, diğer ilişkili hastalıklardan meydana gelecek ölüm oranlarını anlamlı bir biçimde iyileştirip iyileştirmedini araştırmak olabilmektedir. Bu durum, sağkalım süreleri kullanılarak ifade edilirse, birbirini tetikleyen hastalıklara ilişkin sağkalım sürelerinin bağımlılığının araştırılması ve modellenmesi anlamına gelmektedir

David and Moeschberger (1978), Elandth-Johnson and Johnson (1980), Bowers vd. (1997), Salinas-Terris vd. (2002), Bryant ve Dignam (2004), Zheng ve Klein (1994, 1995) gibi çalışmalarda ise birbiriyle ilişkili sağkalım fonksiyonlarının tahmini için istatistiksel yöntemler üzerine çalışmıştır.

Yarışan riskler teorisinde genellikle başarısızlık sürelerinin istatistiksel olarak birbirinden bağımsız olduğu varsayılır. Bağımsızlık varsayımı altında, model kolaylıkla oluşturulmakta ve istatistiksel çıkarımda karşılaşılabilecek belirlenebilme sorunu ortadan kalkmaktadır. Bununla birlikte, Carriere (1949), Makeham (1874), Seal (1977) gibi bazı

uygulamacı, akademisyen ve araştırmacılar bu varsayımın gerçekçi olmadığını vurgulamışlardır.

Yarışan risk modellerinde, bağımlılık durumu göz önüne alındığında, genel yaklaşım kapulaların kullanılmasıdır. Literatürde, kapulaların kullanıldığı çalışmalar mevcuttur. Oakes (1989) çalışmasında, 109 kanser hastasına uygulanan iki farklı kemoterapi rejimi sonunda elde edilen sonuçları Archimedean kapula kullanarak değerlendirmiştir. Shih ve Louis (1975), HIV virüsü ile enfekte olan hastalardan elde edilen verilerin analizini Clayton kapula, pozitif durağan zayıflık (positive stable frailties), Frank kapula kullanarak yapmıştır. Zheng ve Klein (1995), non-Hodkings lenfoma hastalarının yer aldığı klinik denemelerden elde edilen verilerin modellenmesi için kapula yaklaşımını kullanmıştır. Medikal olmayan alanda ise, Hougaard (1987) makine bozulmalarının dağılımını incelemek için pozitif durağan kapulanın yarışan risklerin modellenmesinde nasıl kullanıldığını araştırmıştır.

Yarışan risk modeli, bazı çalışmalarda özellikle aktüeryal bilimlerde, çoklu azalan modeli (multiple decrement model) olarak adlandırılmaktadır. Elandth-Johnson (1976), başarısızlık zamanlarının bağımlılığı varsayımı altında yarışan riskler modelinin göz önüne alındığı ilk çalışmalardandır. Daha sonra, Yashin vd. (1986) ile konuyla ilgili çalışmalar devam etmiştir. Daha güncel olarak, Carriere (1994), Escarela ve Carriere (2003) iki başarısızlık zamanı arasındaki bağımlılığı iki boyutlu kapula ile modellemiştir.

7.1 Kapula Kavramı ve Özellikleri

Rasgele değişkenler arasındaki bağımlılık yapısını modellemek ve ortak dağılım fonksiyonunu oluşturmak için kapula fonksiyonları kullanışlı araçlardır. Kapula, marjinal dağılımları ortak dağılımı oluşturmak amacıyla ilişkilendiren bir fonksiyondur. Bağımlılık yapısını yani ortak dağılımı kapula fonksiyonları kullanılarak belirlemek marjinal dağılımların seçiminden bağımsız olarak yapılabilir. Bu nedenle kapula, bağımlılığı modellemek için gerçekçi ve daha az kısıtlayıcı bir araçtır.

Yarışan riskler teorisinin gerçek yaşamdaki uygulamalarında birbirine bağımlı ikiden fazla riskin olduğu durumlar yer almaktadır. Buna karşın, iki boyutlu kapula teorisinin doğrudan çok boyutlu duruma genişletilememesi çok boyutlu kapula teorisinin uygulanabilirliği açısından bir sorundur oluşturmaktadır. Sklar (1959)'ın çalışmasında yer alan temel sonuçlara rağmen çok boyutlu kapulaların oluşturulmasında hala açık problemler yer almaktadır. Örneğin, kapula parametresinin tahmin edilmesi için kullanılan Kendall'ın tau ve Spearman'ın rho ilişki katsayısı gibi iki boyutlu ölçülere karşılık gelen çok boyutlu bir bağımlılık ölçüsünün olmaması çok boyutlu kapula teorisinde parametre tahmini açık bir problemdir. Bu durumun varlığı çok boyutlu kapulaları daha az cazip hale getirmektedir. İki den fazla riskin bağımlılığın kapula ile modellendiği çalışmalarda kapula parametrelerine ilişkin varsayımlar yaparak modellemeler yapılmıştır (Kaishev 2007, Dimitrova 2009). Çalışmada, kapula parametresinin tahmininin elde edilmesi için Kendall'ın tau katsayısına dayanan tahmin yönteminin yanı sıra en çok olabilirlik yöntemine de yer verilmiştir. Ancak, çok boyutlu durumda parametre tahmin problemi çalışma kapsamına alınmadığından iki boyutlu kapulalar ile çalışılmıştır. Ayrıca, çalışmada kapula parametresinin Kendall'ın tau katsayısına dayanan ve aslında momentler tahmin yöntemiyle elde edilen tahminleri ile en çok olabilirlik yöntemine göre elde edilen tahminlerinin karşılaştırılması amaçlandığından iki boyutlu kapulalar çalışmanın konusunu oluşturmuştur.

Tanım 7.1 İki boyutlu kapula, aşağıdaki özellikleri olan ve $C : [0,1]^2 \rightarrow [0,1]$ şeklinde bir fonksiyondur.

1. Her $u \in [0,1]$ için $C(0,u) = C(u,0) = 0$
2. Her $u \in [0,1]$ için $C(u,1) = u$ ve $C(1,u) = u$
3. $u_1 \leq v_1$ ve $u_2 \leq v_2$ olan her $(u_1, u_2), (v_1, v_2) \in [0,1] \times [0,1]$ için

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$$

Sklar'ın teoremi ile 'Copula' kelimesinin kullanımı açıklık kazanmaktadır. Sklar (1959) bu terimi 'çok boyutlu dağılımı kendi tek boyutlu marjinali ile ilişkilendiren (bağlayan)' fonksiyonu adlandırmak için kullanmıştır ve Copula terimi matematik literatüründe ilk kez Sklar (1959)'da görülmüştür.

Sklar (1959), C bir kapula fonksiyonunu $U \in (0,1)$ olmak üzere, U_1, U_2 için bir ortak dağılım fonksiyonu olarak ifade etmiştir.

$$C(u_1, u_2) = P\{U_1 \leq u_1, U_2 \leq u_2\} \quad (7.1)$$

Buna göre Sklar teoremi aşağıdaki gibi ifade edilebilir.

T_1, T_2 , sırasıyla, $F_1(t_1), F_2(t_2)$ sürekli dağılım fonksiyonuna sahip rasgele değişkenler olsun. Sklar (1959) teoremine göre, ortak dağılım fonksiyonu $F(t_1, t_2)$ 'yi tanımlayan tek bir C kapula fonksiyonu vardır ve aşağıdaki gibi ifade edilir (Sklar 1959).

$$F(t_1, t_2) = C(F_1(t_1), F_2(t_2)) \quad (7.2)$$

(7.1) göz önüne alındığında, U_1, U_2 marjinal dağılımlar seçilirse,

$$U_1 \rightarrow F_1(t_1), U_2 \rightarrow F_2(t_2) \quad (7.3)$$

$$\begin{aligned} F(t_1, t_2) &= P\{T_1 \leq t_1, T_2 \leq t_2\} \\ &= P\{F_1(T_1) \leq F_1(t_1), F_2(T_2) \leq F_2(t_2)\} \\ &= C(F_1(t_1), F_2(t_2)) \end{aligned} \quad (7.4)$$

olur.

Sklar (1959) bu ifadenin tersinin de doğruluğunu göstermiştir. C iki boyutlu bir kapula fonksiyonu ve $F_1(t_1), F_2(t_2)$ sürekli marjinal dağılımlar olmak üzere,

(T_1, T_2) rasgele vektörünün ortak dağılımı, $(F_1(t_1), F_2(t_2))$ rasgele vektörünün ortak dağılımı ile ifade edilebilir.

$$\begin{aligned}
 C\{F_1(t_1), F_2(t_2)\} &= P\{F_1(T_1) \leq F_1(t_1), F_2(T_2) \leq F_2(t_2)\} \\
 &= P\{T_1 \leq F_1^{-1}(t_1), T_2 \leq F_2^{-1}(t_2)\} \\
 &= P\{T_1 \leq t_1, T_2 \leq t_2\} \\
 &= F(t_1, t_2)
 \end{aligned} \tag{7.5}$$

olur.

Yaygın olarak bilinen kapula fonksiyonu örneği çarpım kapulasıdır. Çarpım kapulası aşağıdaki gibi ifade edilir.

$$C(F_1(t_1), F_2(t_2)) = F(t_1, t_2) = \prod_{j=1}^2 F_j(t_j) \tag{7.6}$$

Ayrıca, $C(F_1(t_1), F_2(t_2)) = \min(F_1(t_1), F_2(t_2))$ olarak tanımlanan kapula da bir diğer örnek olarak göz önüne alınabilir. Daha birçok kapula örneğine çeşitli referanslardan (Nelsen 2006) ulaşmak mümkündür.

Marjinal dağılım fonksiyonları ile ortak dağılım fonksiyonu arasında kapula fonksiyonu ile kurulan ilişkinin bir benzeri, marjinal sağkalım fonksiyonları ile ortak sağkalım fonksiyonu arasında da kurulabilir (Sklar 1959).

Buna göre, (T_1, T_2) 'ye ait ortak sağkalım fonksiyonu aşağıdaki gibi tanımlanır.

$$\begin{aligned}
 S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) \\
 &= 1 - F_1(t_1) + F_2(t_2) + F(t_1, t_2) \\
 &= S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2))
 \end{aligned} \tag{7.7}$$

$\mathcal{C}^0[0,1]^2 \rightarrow [0,1]$ olan bir $\mathcal{C}^0$ fonksiyonu aşağıdaki gibi tanımlansın.

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v), \quad u, v \in (0, 1) \quad (7.8)$$

Sklar (1959) teoremine göre $\tilde{C}$ de bir kapuladır ve ortak sağkalım fonksiyonu bu kapula fonksiyonu kullanılarak

$$S(t_1, t_2) = \tilde{C}(S_1(t_1), S_2(t_2)) \quad (7.9)$$

olarak yazılır. Burada, $\tilde{C}$ sağkalım kapulası olarak adlandırılır.

$\tilde{C} = C$ eşitliği sadece eliptik kapulalar için geçerlidir. Bu tür kapulalara örnek olarak, Frank, Student t kapulası, Normal kapula verilebilir. Bu tür kapula fonksiyonları kullanıldığı takdirde, kapula ya da sağkalım kapulası ile çalışmak aynıdır.

Kapula yaklaşımını kullanarak, farklı aralıklardaki farklı ilişkiler modellenenebilir. Bağımlılık yapısının belirlenmesinde kapula fonksiyonunun kullanılmasının bir takım getirileri vardır. Bu getiriler aşağıdaki gibi özetlenebilir.

1. Kapula marjinal dağılımların ve ortak dağılımın yani bağımlılık yapısının ayrı ayrı modellenmesine olanak tanır.
2. Kapula fonksiyonu, bağımlılığın derecesini ve aynı zamanda bağımlılığın yapısının elde edilmesini sağlar. Doğrusal korelasyon, kuyruk bağımlılığı için bilgi vermezken, kapula fonksiyonu asimetrik bağımlılık hakkında bilgiye ulaşmaya olanak tanır.
3. Kapulalar, artan ve sürekli dönüşümler altında değişmezdir. Örneğin, kapula ile ifade edilen bir bağımlılık yapısı, logaritmik dönüşüm altında değişmez. Buna karşın, korelasyonlar sadece doğrusal dönüşümler altında değişmezdir.

Çok boyutlu dağılımların modellenmesinde çok kullanışlı bir araç olan kapula ailesinin belirlenmesi önemlidir. Kapula ailelerinden, matematiksel özelliklerinin avantajları ve

sağkalım tipi verilerin modellenmesine uygunluğu göz önüne alınarak Archimedean kapula aileleri incelenmiş ve bağımlılığın modellenmesinde bu aileler tercih edilmiştir (Oakes 1989).

7.2 Bağımlılık Kavramı ve Kuyruk Bağımlılığı

Rasgele değişkenlerin arasındaki bağımlılık, karşılık gelen çok boyutlu dağılımın kapulası tarafından tümüyle karakterize edilir (Deheuvels 1978, Schweizer ve Wolf 1981). Buna karşın, ‘ilişkili sağkalım süreleri’ ifadesi için korelasyon ölçüsünü kullanmak yaygındır. İki değişken arasındaki bağımlılığı anlatmak için ilişki katsayısını (Pearson ‘ın ilişki katsayısı) kullanmak geleneksel bir yoldur (Hougaard, 2000). Pearson’ın ilişki katsayısı, rasgele değişkenlerin ortak dağılımının normal olduğu durumlarda bağımlılığın uygun bir ölçüsüdür. Sağkalım sürelerinin dağılımının normal dağılmadığı göz önüne alındığında, Pearson’ın ilişki katsayısı sağkalım modellerinde bağımlılığın açıklanması için uygun bir ölçü olmamaktadır. Spearman’ın rho ve Kendall’ın tau ilişki katsayısı bağımlılığın açıklanması için daha uygun ölçülerdir.

Kapula fonksiyonu T_1 ve T_2 rasgele değişkenleri arasındaki tüm bağımlılığı göz önüne almaktadır. g_1 ve g_2 , T_1 ve T_2 ’nin tanım aralığı üzerinde kesinlikle artan iki fonksiyon olsun. $g_1(T_1)$ ve $g_2(T_2)$ dönüşümleri T_1 ve T_2 ile aynı kapulayı vermektedir (Schweizer ve Wolff 1981). Başka bir deyişle T_1 ve T_2 ’nin birlikte hareketi kapula tarafından ortaya konmaktadır. Ayrıca, iki standart parametrik olmayan korelasyon ölçüsü Spearman’ın rho ve Kendall’ın tau ilişki katsayısı kapula fonksiyonu kullanılarak ifade edilebilmektedir (Schweizer ve Wolff 1981).

$$\rho_s(T_1, T_2) = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3 \quad (7.10)$$

$$\tau(T_1, T_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 1 \quad (7.11)$$

Kendall'nın τ ve Spearman'nın ρ ilişki katsayıları, birlikteliğin geniş çaplı ölçüleri olmasının yanında, bazı durumlarda daha lokal bağımlılıkları da ölçmek daha yararlı olmaktadır. Diğer bir ifadeyle kuyruk bağımlılığı hakkında bilgi edinmek amaçlanabilir. Kuyruk bağımlılığı sağkalım tipi verilerin modellenmesinde önemli bir kavramdır. Çünkü (T_1, T_2) sağkalım süresi vektörünün $(0, 0)$ ve (∞, ∞) limit durumlarındaki davranışını göstermektedir. Aynı zamanda, bu limit durumları, ‘ani ortak başarısızlık (ölüm)’ ve ‘uzun dönem ortak sağkalım’ durumlarını ifade etmektedir.

C iki boyutlu bir kapula ise ve

$$\lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lambda_U \quad (7.12)$$

olarak tanımlanan limit var ise, $\lambda_U \in (0, 1]$ ise, C üst kuyruk bağımlılığına sahiptir. $\lambda_U = 0$ ise, kuyruk bağımlılığına sahip değildir.

Benzer olarak,

$\lim_{u \rightarrow 1} \frac{C(u, u)}{u} = \lambda_L$ limiti var ise ve, $\lambda_L \in (0, 1]$ ise, C alt kuyruk bağımlılığına sahiptir. $\lambda_L = 0$ ise, kuyruk bağımlılığına sahip değildir.

Yukarıdaki tanımlamalar aşağıdaki gibi açıklanabilir. T_1 ve T_2 iki sağkalım süresi olsun. λ_U , aşağıdaki fonksiyonun limit değeri olsun.

$$\begin{aligned} \lambda_U(u) &= \Pr\{T_2 > F_2^{-1}(u) \mid T_1 > F_1^{-1}(u)\} \\ &= \frac{1 - 2u + C(u, u)}{1 - u} \end{aligned} \quad (7.13)$$

Aynı şekilde $\lambda_L(u)$ için,

$$\begin{aligned}\lambda_L(u) &= \Pr\{T_2 < F_2^{-1}(u) \mid T_1 < F_1^{-1}(u)\} \\ &= \frac{C(u,u)}{u}\end{aligned}\tag{7.14}$$

olsun.

λ_U , T_1 rasgele değişkeninin $t_1 = F_1^{-1}(u)$ ‘değerinden büyük olduğu bilindiğinde, T_2 rasgele değişkeninin $t_2 = F_2^{-1}(u)$ değerinden büyük olma olasılığını vermektedir. Benzer açıklama, λ_L için de yapılabilir. Dikkat edilmelidir ki, λ_U ve λ_L değerleri sadece kapula fonksiyona bağlıdır, marjinal dağılımlardan bağımsızdır.

7.3 Archimedean Kapula Ailesi

Archimedean kapula ailesi literatürde, çok boyutlu dağılımların tek boyutlu olarak ifade edilmesine olanak tanıırken kolaylık sağlayan bazı matematiksel özelliklerinden dolayı oldukça önemli bir yere sahiptir (Genest ve Mackay 1986b). Örneğin, Gaussian kapula kuyruklardaki bağımlılığın modellenmesinde, Student t kapula asimetric kuyruk bağımlılığının modellenmesinde yeterli olmazken, Archimedean kapula sınıfı, üst ve alt kuyruk bağımlılığın modellenmesinde oldukça kullanışlı modeller içermektedir.

Kapulanın üretici fonksiyonu olarak adlandırılan $\phi: [0,1] \rightarrow [0,\infty]$ fonksiyonu aşağıdaki özelliklere sahip bir fonksiyondur.

1. $\phi(1) = 0$
2. $\phi(0) = \infty$
3. Tüm $t \in (0,1)$ için, $\phi'(t) < 0$ 'dır.
4. Tüm $t \in (0,1)$ için, $\phi''(t) \geq 0$ 'dır.

ϕ fonksiyonunun tersi vardır ve $\phi^{-1}: [0, \infty] \rightarrow [0, 1]$ biçiminde tanımlanır. ϕ^{-1} fonksiyonu $\phi^{-1}(0)=1$ ve $\phi^{-1}(\infty)=0$ özellikleri hariç, ϕ fonksiyonu ile aynı özelliklere sahiptir. Buna göre, bir kapula eğer aşağıdaki biçimde yazılabiliyorsa, Archimedean kapula olarak adlandırılır. Üretici fonksiyon, Archimedean kapulayı tek olarak belirler.

$$C_{\phi}(u, v) = \phi^{-1}[\phi(u) + \phi(v)], \quad u, v \in (0, 1) \quad (7.17)$$

Teorem 7.1 $X \sim U(0,1)$ ve $Y \sim U(0,1)$ olsun. X ve Y 'nin bağımlılık fonksiyonu (dependency function) $C(x, y)$ aşağıdaki gibi tanımlansın (Genest ve Rivest 1993).

$$C(x, y) = \phi^{-1}\{\phi(x) + \phi(y)\} \quad (7.18)$$

ϕ , yukarıdaki özelliklere sahip bir fonksiyon olmak üzere aşağıdaki tanımlamalar yapılsın.

$$U \sim \frac{\phi(X)}{\phi(X) + \phi(Y)} \quad (7.19)$$

$$Z = C(X, Y) \quad (7.20)$$

$$\lambda(z) = \frac{\phi(z)}{\phi'(z)}, \quad 0 < z \leq 1 \quad (7.21)$$

Buna göre,

- a) $U \sim (0,1)$
- b) $Z \sim K(z)$
- c) $K(z) = z - \lambda(z)$ olmak üzere $(0,1)$ aralığında tanımlıdır.
- d) U ve V bağımsız rasgele değişkenlerdir.

Gerçekte, (a), (b) ve (c) özelliklerini sağlayan ϕ fonksiyonunun varlığı tüm tanım aralığı boyunca $C(x, y) = \phi^{-1}\{\phi(x) + \phi(y)\}$ fonksiyonun varlığını gösterir (Genest ve Rivest 1993).

İlerleyen bölümlerde, $Z = C(X, Y)$ 'nin dağılımı olan $K(z)$ 'nin, X ve Y ikilisinden elde edilen rasgele örneklemden parametrik olmayan tahminin nasıl elde edileceği gösterilecektir. $C(X, Y)$ 'nin Archimedean kapula olduğu göz önüne alınırsa $K(v)$ 'nin tahmini, iki boyutlu dağılım fonksiyonun tahmini anlamına gelmektedir. Çünkü ϕ fonksiyonuna

$$\frac{\phi(z)}{\phi'(z)} = z - K(z) \quad (7.22)$$

diferansiyel denklemi çözülerek ulaşılabilir. Bu denklemin çözümü ise,

$$\phi(z) = \exp \left\{ \int_{z_0}^z \frac{1}{\lambda(t)} dt \right\}, \quad 0 < z_0 < 1 \quad (7.23)$$

olmaktadır.

Burada (7.22) ile gösterilen fonksiyon, λ fonksiyonu olarak adlandırılır ve aşağıdaki gibi gösterilir.

$$\lambda(z; \theta) = \frac{\phi(z)}{\phi'(z)} = z - K(z; \theta) \quad (7.24)$$

λ fonksiyonu, her iki boyutlu kapula ailesini karakteristiğidir. Bu çalışmada, bu fonksiyona ait grafikler çizdirilerek yorumlar yapılmıştır.

Archimedean kapulalar, Archimedean aksiyom özelliğini sağlamasından dolayı bu isimle adlandırılmaktadır ve ilk olarak Ling (1965) çalışmasında bu kapula fonksiyonları için Archimedean kapula terimi kullanılmıştır. Archimedean kapula sınıfı,

Gumbel (Gumbel 1960), Ali-Mikhail-Had-Thélot (Ali, Mikhail ve Hog 1978), Clayton (Clayton 1978, Cook-Johnson 1981, Oakes 1982), Frank (Frank 1979, Nelsen 1986, Genest 1987), Hougaard (1987) gibi kapula fonksiyonlarını içermektedir. Oakes (1982) çalışmasında, bağımlılık söz konusu olunca, Archimedean kapulaların iki boyutlu sağkalım tipi verilerinin modellenmesinde önemli bir araç olduğunu ortaya koymuştur. Archimedean kapulaların temel özelliklerini, yakınsama sonuçları Genest ve MacKay (1986b)'ın çalışmasında incelenmiş ve ortaya konmuştur.

Archimedean sınıfına ait bazı aileler, farklı yapıdaki bağımlılıkları ortaya koymada çok kullanışlıdır. Aynı zamanda Spearman'ın rho ve Kendall'ın tau ilişki katsayıları ile ilişkilendirilebildiğinden uygulamada Archimedean kapulalar tercih edilmektedir. Bu ailelere örnek olarak, Gumbel, Clayton ve Frank kapulaları örnek olarak verilebilir. Çalışmada, yukarıda bahsedilen özelliklerinden dolayı bu kapulalar ile çalışılmıştır.

Çizelge 7.1 ile çalışmada kullanılan Gumbel, Clayton ve Frank kapulaları, bu kapulalara ait üretici fonksiyonlar gösterilmiştir.

Çizelge 7.1 Çalışmada kullanılan iki boyutlu Archimedean kapulalar

Aile	Kapula parametre uzayı	İki boyutlu kapula fonksiyonu
Gumbel	$\theta \geq 1$	$C(u_1, u_2) = \exp[-(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}$
Clayton	$\theta > 1$	$C(u_1, u_2) = ((u_1)^{-\theta} + (u_2)^{-\theta} - 1)^{-1/\theta}$
Frank	$-\infty < \theta < \infty$	$C(u_1, u_2) = (-\frac{1}{\theta}) \ln \left\{ \frac{[(1 - e^{-\theta}) - (1 - e^{-\theta u_1})(1 - e^{-\theta u_2})]}{(1 - e^{-\theta})} \right\}$

Gumbel kapulada, $\theta = 1$ durumu bağımsızlığı gösterir ve $\theta \rightarrow \infty$ durumu ise güçlü bağımlılığa işaret etmektedir. Gumbel kapulanın asimetric bir yapıya sahip olmasından dolayı kuyruk bağımlılığını yorumlamada kullanışlıdır. Gumbel kapula ile modellenen

gözlemler için sağ kuyruk bağımlılığı yorumlanmalıdır. Yani, Gumbel kapula ile modellenen iki boyutlu rasgele değişkenler, birlikte artış göstermeye birlikte azalış göstermekten daha yatkındır. Gözlemler küçük değerler için düşük korelasyona ve yüksek değerler için yüksek korelasyona sahip ise bu tür gözlemleri Gumbel kapula ile modellemek doğru bir tercih olmaktadır.

Archimedean kapula ailesine ait bir diğer asimetrik kapula ise Clayton kapulasıdır. Gumbel kapulasının aksine, Clayton kapula için sol kuyruk bağımlılığı dikkate alınmalı ve yorumlanmalıdır. Birlikte azalış göstermeye, birlikte artış göstermekten daha eğimli olan gözlemler Clayton kapula ile modellenebilmektedir.

Asimetrik olan Gumbel ve Clayton kapulalarından sonra Archimedean kapula ailesine ait ve simetrik olan bir kapula Frank kapulasıdır. Diğer iki kapuladan farklı olarak Frank kapula ile negatif bağımlılığın da modellenebilir olması ve geniş bir parametre uzayına sahip olması Frank kapulayı uygulamada tercih edilir yapmaktadır. Bağımlılık Frank kapula için iki kuyrukta da simetriktir. Kuyruklarda bağımlılık zayıftır ve buralarda yer alan olaylar sıfır olasılıkla meydana gelmektedir. Frank kapulada güçlü bağımlılık dağılımının ortasında yer almaktadır. Frank kapula ile modellenebilen gözlemler çok güçlü pozitif ya da çok güçlü negatif bağımlılığa sahiptir.

8. İKİ BOYUTLU ARCHIMEDEAN KAPULA FONKSİYONUNUN TAHMİNİ

Bir rasgele vektöre ait ortak dağılım fonksiyonunun, marjinal dağılım fonksiyonları ve kapula fonksiyonu yardımıyla aşağıdaki gibi ifade edildiğine önceki bölümde değinilmişti.

$$F(t_1, t_2, \dots, t_k) = C(F_1(t_1), F_2(t_2), \dots, F_k(t_k)) \quad (8.1)$$

(8.1) ifadesini yazabilmek için kapula fonksiyonunun ve marjinal dağılımların ayrı ayrı bilinmesi gerekmektedir. Çoğu kez marjinallerin önceden biliniyor olduğu varsayılmaktadır. Buna karşın kapula fonksiyonun belirlenmesi marjinal dağılımlardan bağımsız olduğu göz önüne alındığında marjinallerin önceden biliniyor olmasının bir getirisi olmamaktadır (Genest ve Rivest 1993).

Bu çalışmada, yarışan riskler arasındaki bağımlılığın Archimedean kapula ile modellenebileceği varsayılmış ve bağımlılığı modelleyecek olan Archimedean kapula fonksiyonun tahmini üzerinde çalışılmıştır. Bu amaçla, iki boyutlu Archimedean kapula fonksiyonun parametrik olmayan tahmininin elde edilmesi için Genest ve Rivest'in (1993) önerdiği parametrik olmayan yöntem esas alınmıştır. Genest ve Rivest (1993), bu yöntemde iki boyutlu rasgele değişkenin ortak dağılımının iki boyutlu Archimedean kapula fonksiyonlarından seçtiği aday kapulalardan biriyle modellenebileceği varsaymış ve örneklemden elde ettiği ampirik dağılımın aday kapulalardan hangisine yakın olduğunu araştırmıştır. Veriden elde ettiği ampirik dağılıma en yakın olan aday kapula ile ortak dağılımı tahmin etmiştir. Bu tahmin sürecinde, öncelikle marjinal dağılımların daha sonra da her aday kapula için kapula parametresinin tahminine ihtiyaç vardır. θ kapula parametresi (birliktelik parametresi-association parameter) Genest ve Rivest (1993)'ın çalışmasında Kendall'in ilişki katsayısı kullanılarak momentler tahmin yöntemine göre tahmin edilmiştir. Bu çalışmada, aday kapulaların birliktelik parametreleri ise en çok olabilirlik yöntemi kullanılarak tahmin edilmiştir. Ayrıca, çalışmada Kendall'in ilişki katsayısına dayanan yöntem ve en çok olabilirlik yöntemi ile elde edilen tahminlerin karşılaştırıldığı bir simülasyon çalışması da yapılmıştır. Çalışmada, verinin Archimedean kapula ailesine

ait olan Gumbel, Clayton ve Frank kapulalarından biri ile modellenenebildiği varsayılmış ve bu aday kapulalar arasından verideki bağımlılık yapısını en iyi modelleyen Archimedean kapula fonksiyonu tahmin edilmiştir.

Archimedean kapula sınıfına ait bir kapula fonksiyonu daha önceki bölümlerde (8.1) ile gösterildiği gibi

$$C_{\phi}(u, v) = \phi^{-1}[\phi(u) + \phi(v)] \quad (8.2)$$

biçiminde ifade edilmekteydi. Bu fonksiyon, $[0,1]$ aralığında tanımlı olan

$$K(t) = t - \frac{\phi(t)}{\phi'(t)} \quad (8.3)$$

fonksiyonu ile tek boyutlu olarak belirlenebilmektedir. Bu özelliği nedeniyle, Archimedean kapula ailesi matematiksel olarak birçok kolaylık sağlamaktadır. $\phi(t)$ 'nin doğası gereği, $K(t)$ fonksiyonu $(0,1)$ aralığında tanımlı olan bir dağılım fonksiyonudur.

$$F(X, Y) = C(U, V) = Z \quad (8.4)$$

olarak ifade edildiğinde,

$$K(z) = P[F(X, Y) \leq z] = P[C\{F_X(X), F_Y(Y)\} \leq z] \quad (8.5)$$

olarak tanımlanır.

Çizelge 8.1'de çalışmada kullanılan Archimedean kapula aileleri karşılık gelen üretici fonksiyonları ve dağılım fonksiyonları gösterilmiştir.

Çizelge 8.1 Çalışmada kullanılan Archimedean kapulaların dağılımı

Aile	$\phi(z)$	$\phi'(z)$	$K(z) = z - \frac{\phi(z)}{\phi'(z)}$
Gumbel	$(-\ln z)^\theta$	$-\theta(-\ln z)^{\theta-1} \frac{1}{z}$	$z - \frac{(z \ln z)}{\theta}$
Clayton	$z^{-\theta} - 1$	$-\theta z^{-\theta-1}$	$z - \frac{(z^{\theta+1} - z)}{\theta}$
Frank	$-\ln \frac{e^{-\theta z} - 1}{e^{-\theta} - 1}$	$\frac{\theta}{1 - e^{\theta z}}$	$z - \frac{\ln \frac{e^{-\theta z} - 1}{e^{-\theta} - 1}}{\theta} (e^{\theta z} - 1)$

8.1 Birliktelik Parametresinin Tahmini

Rasgele değişkenler arasındaki bağımlılık yapısı parametrik kapulalar tarafından modellenenmektedir. Bu kapulalar tek bir parametre ile belirlenebildiği gibi bir parametre vektörü ile de belirlenebilmektedir. Bağımlılık ya da birliktelik parametresi olarak adlandırılan bu parametre(ler) bağımlılığın derecesini göstermektedir. Bu çalışmada kullanılan Archimedean kapula ailesinden olan, Clayton, Gumbel, Frank kapulaları tek parametre ile belirlenen kapulalardır.

Bağımlılık parametresinin tahmini, gözlemlerin sıralaması (rank) temel alınarak elde edilebilir. Bu yöntem Kendall'nın tau ilişki katsayısını kullanmaktadır.

8.1.1 Kendall'ın tau ilişki katsayısı ile tahmin

Gözlemlerin sıralanması temeline dayanan önemli bir bağımlılık ölçüsü Kendall'ın tau ilişki katsayısıdır. Bu katsayı örneklemden aşağıdaki gibi tahmin edilir.

$$\hat{\tau}_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)} P_n - 1 \quad (8.6)$$

P_n ve Q_n , sırasıyla birbiriyle uyumlu (concordant) ve uyumsuz (discordant) çiftlerin sayısıdır. Burada, $(X_i - X_j)(Y_i - Y_j) > 0$ ise (X_i, Y_i) ve (X_j, Y_j) gözlem çifti birbiriyle uyumlu, tersine $(X_i - X_j)(Y_i - Y_j) < 0$ ise birbiriyle uyumsuz olara düşünülür. X ve Y nin sürekli olduğunun varsayımı altında, $(X_i - X_j)(Y_i - Y_j) = 0$ değerini sıfır olasılıkla almaktadır.

Kendall'ın tau değerine dayalı parametre tahminindeki fikir momentler tahmin yöntemine dayanmaktadır. Archimedean kapula ailesi ile Kendall'ın tau ilişki katsayısı arasındaki ilişki aşağıdaki gibi ifade edilir (Genest ve MacKay 1986a,1986b).

$$\tau(X, Y) = 4E(Z) - 1 = 1 + 4 \int_0^1 \lambda(z) dz \quad (8.7)$$

Kendall'ın tau katsayısının örneklemden tahmini ise,

$$\hat{\tau}_n(X, Y) = 4E(\bar{Z}) - 1 = 1 + 4 \int_0^1 \lambda_n(z) dz \quad (8.8)$$

olarak elde edilir. Burada, Z , (8.4) ile ve $\lambda(z)$ ise (7.24) ile gösterildiği gibidir. $\bar{Z}$ ve $\lambda_n(z)$ ise bu ifadelerin örneklemden elde edilen tahmin değeridir. Bu ifadelerden yola çıkarak momenler tahmin yöntemine göre Kendall'ın tau ilişki katsayısı kullanılarak birliktelik parametresi θ 'nın tahmini elde edilir.

Kendall'ın tau değeri kullanılarak elde edilen tahmin edici verilerin zayıf bağımlılık gösterdiği durumda iyi sonuçlar vermekte, yüksek bağımlılık ya da zayıf bağımsızlık durumlarında iyi sonuçlar vermemektedir (Phelps ve Weissfeld 1997). Bu nedenle, uygulamacı yüksek bağımlılık ya da zayıf bağımlılık olan durumlarda tahmin yöntemini seçerken dikkatli olmalıdır. Buna karşın hesaplama kolaylığı açısından ve marjinal dağılımların bilinmesine gerek duymayan bu parametrik olmayan yaklaşım bir çok çalışmada tercih edilmektedir. Çizelge (8.2) ile çalışmada kullanılan Archimedean kapula ailesine ait kapulaların parametresi ile Kendall'ın tau ilişki katsayısı arasındaki ilişki örneklendirilmiştir.

Çizelge 8.2 Çalışmada kullanılan Archimedean kapulalar ile Kendall’ın tau katsayısı arasındaki ilişki

Aile	θ değer kümesi	τ
Gumbel	$\theta \in [1, \infty)$	$\frac{\theta - 1}{\theta}$
Clayton	$\theta \in [0, \infty)$	$\frac{\theta}{\theta + 2}$
Frank	$\theta \in (-\infty, +\infty)$	$1 - \frac{4}{\theta} [1 - D_1(\theta)]$

$D_1(\theta)$, Debye fonksiyonudur ve $D_n(x) = \frac{n}{x^n} \int_0^x \frac{t^n}{e^t - 1} dt$, $n > 0$ olarak tanımlanmaktadır.

8.1.2 En çok olabilirlik yöntemi ile tahmin

Kapula parametresinin tahmin problemi birçok araştırmacının ilgisini çekmiştir (Genest ve Rivest 1993, Genest vd. 1995, Joe 1997). Özellikle parametrenin çok boyutlu olduğu durumlarda momentler tahmin yöntemine iyi bilinen bir alternatif en çok olabilirlik yöntemidir ve bu yöntem daha etkin tahmin ediciler sunmaktadır.

İki boyutlu ortak dağılım fonksiyonu kapula fonksiyonu ve marjinalleri yardımıyla aşağıdaki gibi yazılabilmektedir (Sklar 1959, Sklar 1973).

$$F(x_1, x_2) = C(F_1(x; \alpha_1), F_2(x; \alpha_2); \theta) \quad (8.9)$$

Buna göre, (8.9) ifadesi için tahmin edilecek parametreler marjinal dağılımların parametreleri ve kapulaya ait parametredir (ya da parametre vektörleridir). Tam en çok olabilirlik (full maximum likelihood, FML) yaklaşımı kullanılarak, tüm parametrelerin eşanlı tahmini elde edilebilir (Joe 1997). Genellikle elde edilen skor fonksiyonlarının doğrusal olmaması çözümü zahmetli kılmaktadır. Buna karşın bilgisayar programları yardımıyla bu fonksiyonlar çözülüp tahminler elde edilebilmektedir.

Parametrelerin tahmin edilmesine ilişkin alternatif bir yöntem ise aşağıdaki gibi açıklanabilir. Bu yöntem, kapula fonksiyonunun marjinallerden bağımsız olma özelliğini kullanmaktadır.

Sklar'ın teoremindeki gösterimler kullanılarak, iki boyutlu ortak olasılık yoğunluk fonksiyonu aşağıdaki gibidir.

$$f(x_1, x_2) = c(F_1(x_1; \alpha_1), F_2(x_2; \alpha_2); \theta) f_1(x_1) f_2(x_2) \quad (8.10)$$

Burada, $c(u, v)$ fonksiyonu, iki boyutlu $C(u, v)$ kapula fonksiyonun yoğunluk fonksiyonudur.

$$c(u, v) = \frac{\partial C(u, v)}{\partial u \partial v} \quad (8.11)$$

(8.9) kullanılarak aşağıdaki olabilirlik fonksiyonu elde edilir.

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \ln(f_i(x_1, x_2)) = L_{Cop} + L_{Mar} \\ &= \sum_{i=1}^n \ln c(F_1(x_1), F_2(x_2)); \theta) + \sum_{i=1}^n [\ln(f_1(x_i)) + \ln(f_2(x_i))] \end{aligned} \quad (8.12)$$

olarak yazılır.

L_{Cop} , verinin kapula ile temsil edilen bağımlı kısmının log-olabilirlik fonksiyonudur.

L_{Mar} ise her bir marjinalden elde edilen log-olabilirlik fonksiyonunu ifade etmektedir.

C kapula fonksiyonunun, θ parametresi (ya da parametre vektörü) ile gösterilen bir kapula ailesine ait olduğu ve marjinal dağılımların da α_j , $j=1,2$ parametresi (ya da parametre vektörü) ile gösterilen ailelere ait olduğu göz önüne alınsın. Bu durumda, ilk

olarak marjinal dağılımlara ait parametreler genellikle en çok olabilirlik tahmin yöntemi kullanılarak tahmin edilir.

$$\hat{\alpha}_j^{MLE} = \frac{\partial L_j^{Mar}}{\partial \alpha_j} = 0, \quad j = 1, 2 \quad (8.13)$$

Daha sonra ve bu tahmin ediciler kullanılarak, kapula parametresine ait tahmin edici elde edilir.

$$\frac{\partial}{\partial \theta} [L_{Cop}(\theta; \mathbf{x}) + L_{Mar}(\hat{\alpha}_{MLE}; \mathbf{x})] = 0 \quad (8.14)$$

Tahmin sürecinin bu şekilde iki basamaktan oluşması nedeniyle bu yöntem 'iki basamaklı parametrik yöntem' adı verilir ve literatürde marjinallerden çıkarım tekniği (Inference from margins-IFM) olarak adlandırılır.

Tam en çok olabilirlik yöntemi ve iki basamaklı en çok olabilirlik yöntemi bazı durumlarda eş değerler sonuçlar verebilmektedir. Örneğin, marjinal dağılımları normal olan, Gaussian kapula için bu eş değerlik söz konusudur (Joe 1997). İki yöntem sonucu elde edilen tahmin ediciler, tutarlı ve asimptotik olarak normal dağılım göstermektedir (Serfling 1980). Çok boyutlu parametreler söz konusu olduğunda iki aşamalı tahmin yöntemi, eş anlı doğrudan tahmin yöntemine iyi bir alternatif olmaktadır (Joe 1997).

Bu çalışmada birliktelik parametresinin en çok olabilirlik tahmini iki basamaklı yöntem kullanılarak elde edilmiştir. İlk basamakta marjinal dağılımların ampirik tahminlerini elde edilmiş ve bu tahminler ikinci aşamada kullanılarak (8.15) ile gösterildiği gibi yarı parametrik bir yöntem izlenmiştir. Bu iki basamaklı yarı parametrik yöntem ile elde edilen tahmin ediciler, θ için tutarlı tahmin ediciler olmakta ve düzgünlük koşulları altında asimptotik normal dağılım göstermektedir (Genes vd. 1995).

$$\hat{\theta}_{MLE} = \arg \max_{\theta} L_{Cop}(\theta; \mathbf{x}) = \arg \max_{\theta} \sum_{i=1}^n \ln \square (\hat{F}_1(x_i), \hat{F}_2(x_i); \theta) \quad (8.15)$$

8.1.3 Simülasyon Sonuçları

Simülasyon çalışması ile kapula parametresinin tahmin edilmesi için kullanılan, (8.1.1) ve (8.1.2) bölümlerinde anlatılan iki yöntem ile elde edilen tahminler karşılaştırılmıştır. Kapula parametresi θ 'nın Kendall'ın tau katsayısını temel alan tahmini $\hat{\theta}_{TAU}$ ve en çok olabilirlik yöntemi ile elde edilen tahminler $\hat{\theta}_{MLE}$ olarak gösterilmiştir.

Çizelge 8.3-Çizelge 8.11 ile $\hat{\theta}_{TAU}$ ve $\hat{\theta}_{MLE}$ tahminlerini karşılaştıran simülasyon sonuçları özetlenmiştir. Çizelge 8.3, Çizelge 8.4, Çizelge 8.5 ile birliktelik parametresi sırasıyla $\theta=1.5$, $\theta=3$ ve $\theta=5$ Clayton kapulasından $n=30$, $n=50$, $n=100$, $n=300$ genişliğinde örneklem simüle edilmiş ve her iki yöntem kullanılarak parametrenin tahmin edicisinin değerleri hesaplanmıştır. Sonuçlar 10000 tekrar yapılarak elde edilmiştir. Tahmin edicilerin ortalama, varyans ve hata kareler ortalaması (HKO) değerleri verilmiştir. Aynı işlemler, Gumbel ve Frank kapulası için yapılmıştır. Sonuçlar Çizelge 8.6, Çizelge 8.7 ve Çizelge 8.8 ile Gumbel kapula için, Çizelge 8.9, Çizelge 8.10 ve Çizelge 8.11 ile de Frank kapula için özetlenmiştir. Parametreler değişkenler arasındaki korelasyona göre belirlenmiştir. Yani, aynı aile için artan kapula parametresi artan ilişki miktarını göstermektedir.

Örneklem genişliğinin artması ile tahminler hem parametreye hem de birbirine oldukça yaklaşmıştır. Simülasyon sonuçlarına incelendiğinde HKO ölçütüne göre, en çok olabilirlik tahmin edicisi tau tahmin edicisine göre daha iyi bir tahmin edicidir. Ancak, sonuçlar tekrar değerlendirildiğinde, Tau tahmin edicisinin de kötü bir tahmin edici olmadığı görülebilir. Özellikle artan örneklem genişliğinde en çok olabilirlik tahmin edicisine yakın sonuçlar vermektedir. Özellikle, bazı kapula aileleri için tau tahmin edicinin kapalı formunun bulunması uygulamada büyük kolaylık sağlamaktadır.

Çizelge 8.3 Clayton($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları

Clayton ($\theta = 1.5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	1.5622	1.5375	1.5193	1.5030
	$\hat{\theta}_{TAU}$	1.6070	1.5809	1.5481	1.5129
Yan	$\hat{\theta}_{MLE}$	0.0670	0.0375	0.0193	0.0030
	$\hat{\theta}_{TAU}$	0.1069	0.0809	0.0481	0.0129
HKO	$\hat{\theta}_{MLE}$	0.1807	0.1250	0.0590	0.0177
	$\hat{\theta}_{TAU}$	0.1883	0.1301	0.0609	0.0179

Çizelge 8.4 Clayton($\theta = 3$) kapulası için elde edilen simülasyon sonuçları

Clayton ($\theta = 3$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	3.1279	3.0498	3.0413	3.0014
	$\hat{\theta}_{TAU}$	3.2812	3.1209	3.0695	3.0176
Yan	$\hat{\theta}_{MLE}$	0.1279	0.0498	0.0413	0.0014
	$\hat{\theta}_{TAU}$	0.2812	0.1209	0.0695	0.0176
HKO	$\hat{\theta}_{MLE}$	0.5150	0.2727	0.1325	0.0414
	$\hat{\theta}_{TAU}$	0.5777	0.2848	0.1356	0.0417

Çizelge 8.5 Clayton($\theta = 5$) kapulası için elde edilen simülasyon sonuçları

Clayton ($\theta = 5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	5.1695	5.1033	5.0361	5.0213
	$\hat{\theta}_{TAU}$	5.4395	5.3086	5.1147	5.0339
Yan	$\hat{\theta}_{MLE}$	0.1695	0.1033	0.0361	0.0213
	$\hat{\theta}_{TAU}$	0.4395	0.3086	0.1147	0.0339
HKO	$\hat{\theta}_{MLE}$	1.0238	0.6111	0.2920	0.0984
	$\hat{\theta}_{TAU}$	1.1883	0.6957	0.3038	0.0990

Çizelge 8.6 Gumbel ($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları

Gumbel ($\theta = 1.5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	1.5178	1.5163	1.5093	1.4997
	$\hat{\theta}_{TAU}$	1.5307	1.5260	1.5101	1.5004
Yan	$\hat{\theta}_{MLE}$	0.0178	0.0163	0.0093	0.0003
	$\hat{\theta}_{TAU}$	0.0307	0.0260	0.0101	0.0004
HKO	$\hat{\theta}_{MLE}$	0.0330	0.0311	0.0151	0.0040
	$\hat{\theta}_{TAU}$	0.0337	0.0315	0.0151	0.0040

Çizelge 8.7 Gumbel ($\theta = 3$) kapulası için elde edilen simülasyon sonuçları

Gumbel ($\theta = 3$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	3.0937	3.0401	3.0284	3.0103
	$\hat{\theta}_{TAU}$	3.1961	0.0913	3.0581	3.0208
Yan	$\hat{\theta}_{MLE}$	0.0937	0.0401	0.0284	0.0103
	$\hat{\theta}_{TAU}$	0.1961	0.0913	0.0581	0.0208
HKO	$\hat{\theta}_{MLE}$	0.2481	0.1382	0.0656	0.0222
	$\hat{\theta}_{TAU}$	0.2778	0.1449	0.0682	0.0225

Çizelge 8.8 Gumbel ($\theta = 5$) kapulası için elde edilen simülasyon sonuçları

Gumbel ($\theta = 5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	5.1138	5.0704	5.0270	5.0189
	$\hat{\theta}_{TAU}$	5.3619	5.1844	0.0680	5.0346
Yan	$\hat{\theta}_{MLE}$	0.1138	0.0704	0.0270	0.0186
	$\hat{\theta}_{TAU}$	0.3919	0.1844	0.0680	0.0343
HKO	$\hat{\theta}_{MLE}$	0.6587	0.3774	0.1748	0.0584
	$\hat{\theta}_{TAU}$	0.7767	0.4064	0.1786	0.0592

Çizelge 8.9 Frank ($\theta = 1.5$) kapulası için elde edilen simülasyon sonuçları

Frank ($\theta = 1.5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	1.5623	1.5393	1.5480	1.5033
	$\hat{\theta}_{TAU}$	1.5771	1.5531	1.5493	1.5069
Yan	$\hat{\theta}_{MLE}$	0.0623	0.0393	0.0480	0.0033
	$\hat{\theta}_{TAU}$	0.0771	0.0531	0.0493	0.0069
HKO	$\hat{\theta}_{MLE}$	1.4504	0.7828	0.3735	0.1215
	$\hat{\theta}_{TAU}$	1.4525	0.7841	0.3736	0.1215

Çizelge 8.10 Frank ($\theta = 5$) kapulası için elde edilen simülasyon sonuçları

Frank ($\theta = 5$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	5.1895	5.1379	5.0597	5.0110
	$\hat{\theta}_{TAU}$	5.2451	5.1667	5.0441	5.0207
Yan	$\hat{\theta}_{MLE}$	0.1895	0.1379	0.0597	0.0110
	$\hat{\theta}_{TAU}$	0.2451	0.1667	0.0441	0.0207
HKO	$\hat{\theta}_{MLE}$	2.0158	1.1851	0.5852	0.2026
	$\hat{\theta}_{TAU}$	2.0399	1.1939	0.5836	0.2029

Çizelge 8.11 Frank ($\theta = 15$) kapulası için elde edilen simülasyon sonuçları

Frank ($\theta = 15$)	Tahmin Edici	$n = 30$	$n = 50$	$n = 100$	$n = 300$
Ortalama	$\hat{\theta}_{MLE}$	15.4626	15.3116	15.1448	15.0023
	$\hat{\theta}_{TAU}$	15.7945	15.4595	15.1634	15.0332
Yan	$\hat{\theta}_{MLE}$	0.4626	0.3116	0.1448	0.0023
	$\hat{\theta}_{TAU}$	0.7945	0.4595	0.1634	0.0332
HKO)	$\hat{\theta}_{MLE}$	8.0727	4.4796	2.2828	0.7620
	$\hat{\theta}_{TAU}$	8.4899	4.5936	2.2885	0.7631

8.2 Archimedean Kapula Fonksiyonunun Parametrik Olmayan Tahmin Yöntemi

Bu bölümde, bağımlılık yapısını modelleyebileceği varsayılan Archimedean kapula ailesine ait kapulalardan hangisinin bağımlılığı en iyi modellediği araştırılacaktır. Bunun için, Genest ve Rivest (1993) önerdiği yöntem esas alınmıştır fakat parametre tahmini ve karşılaştırma kriterlerinde farklı alternatifler kullanılmıştır. Çalışmada uygulanacak yöntem aşağıdaki gibi kısaca özetlenebilir. Gösterimlerde kolaylık açısından sadece yöntemi anlatmak için T_1 yerine X ve T_2 yerine Y kullanılmıştır.

1. Kapula parametresinin tahmini elde edilir.
2. $K(z)$ 'nin yani, $F(X,Y) = C(U,V) = z$ 'nin dağılım fonksiyonun parametrik olmayan tahmini, $K_n(z)$ elde edilir. Bunun için ilk olarak sözde (pseudo) gözlemler elde edilmelidir.

$$\text{i. } Z_i = F_n(X_i, Y_i) = \sum_{j=1}^n \left[I\{X_j \leq X_i \& Y_j \leq Y_i\} \right] / (n+1), \quad i = 1, \dots, n$$

$$\text{ii. } K_n(z) = \frac{\sum_{i=1}^n I(Z_i \leq z)}{n+1}$$

$$\text{iii. } K(z) \text{ 'nin parametrik tahmini } (K_\phi(z)), \quad K(z) = z - \frac{\phi(z)}{\phi'(z)} \text{ ilişkisi}$$

kullanarak elde edilir.

3. Örneklemden elde edilen $K_n(z)$ değeri ile teorik $K_{\phi_n}(z)$ değerlerinin birbirinden uzaklığı karşılaştırılır.

$K_n(z)$ aslında $K(z)$ 'nin yani, $F(X,Y) = C(U,V) = Z$ 'nin dağılım fonksiyonun ampirik dağılımıdır. Verilen veri setinden, F için $F_n(z)$ 'nin elde edilmesi ile aynı yol izlenerek $K_n(z)$ bulunmaktadır. Burada temel farklılık, Z_i 'lerin elimizde olmamasıdır. O nedenle ilk adım sözde gözlemleri elde etmektir. Yukarıdaki eşitlikten (2i)

anlaşılacağı üzere, Z_i değerleri iki boyutlu dağılımın ampirik karşılığıdır. Yani, (X_i, Y_i) 'den küçük ya da eşit kaç tane (X_j, Y_j) varsa elde edilen sayı $(n+1)$ 'e bölünür. $K_n(z)$ 'de aynı mantıkla elde edilmektedir. z 'den küçük veya eşit kaç tane Z_i olduğunu bulunarak $K_n(z)$ yani ampirik dağılım elde edilir. Burada z , $(0,1)$ aralığında herhangi bir değer olarak düşünülmektedir. Çünkü z_i değerleri dağılım fonksiyonunun değeridir.

$$K_n(z) = \sum_{j=1}^n I[\{Z_j \leq z\}] / (n+1) \quad (8.13)$$

Bu çalışmada temel varsayım verinin Archimedean kapula ailesinden geldiğidir. Bu nedenle, veriyi modellemeye aday kapulalar olarak, Clayton, Gumbel ve Frank kapulaları seçilmiştir. Uygun kapula ailesinin seçimi için yukarıda anlatılan yöntem kullanılacaktır. Yöntemde, kapula parametresinin tahmini için (8.1.1) ve (8.1.2) bölümlerinde anlatılan her iki yöntem de kullanılacak ve sonuçlar karşılaştırılacaktır.

Karşılaştırma kriteri olarak, (8.14) ile gösterilen Cramer-Von Mises istatistiği yani kare uzaklık (square distance) kullanılmıştır.

$$\sum [K_{\phi_n}(z) - K_n(z)]^2 \quad (8.14)$$

(8.14) ifadesini minimize eden değere göre verideki bağımlılığı aday kapulalardan hangisinin daha iyi modellediği tespit edilir.

8.3 Yöntemin Uygulanması ve Elde Edilen Sonuçlar

Bu bölümde, bilinen bir Archimedean kapula sınıfından simüle edilen iki boyutlu bir veri seti kullanılarak, kapula fonksiyonunun parametrik olmayan tahmininin elde edilmesi için bölüm 8.2'de anlatılan yöntem kullanılmıştır. Sonuçlar Çizelge 8.13 ve Çizelge 8.14 ile özetlenmiştir. $\theta = 3$ birliktelik parametresine sahip Gumbel kapuladan

$n = 100$ birimlik örnek veri seti üretilmiştir. Bu örnek veri setinin Archimedean kapula ailesinden bir kapula ile modellenebileceği varsayımı altında aday kapulalar olarak yine Archimedean kapula sınıfına ait olan Gumbel, Clayton ve Frank seçilmiştir.

İlk olarak aday kapulaların parametreleri bölüm 8.1.1 ve bölüm 8.1.2 ‘de yöntemlerle tahmin edilmiştir. Buna göre sonuçlar Çizelge 8.12 ile verilmiştir.

Çizelge 8.12 Birliktelik parametresinin TAU ve MLE tahminleri

Clayton Kapula	$\hat{\theta}_{TAU}=3.76$	$\hat{\theta}_{MLE}=2.06$
Gumbel Kapula	$\hat{\theta}_{TAU}=2.88$	$\hat{\theta}_{MLE}=2.91$
Frank Kapula	$\hat{\theta}_{TAU}=9.52$	$\hat{\theta}_{MLE}=9.03$

Çizelgede verilen tahminlerin elde edilmesinde R programı kullanılmıştır. Daha sonra bölüm 8.2’de verilen yöntem adım adım uygulanmıştır.

Çizelge 8.13 ve Çizelge 8.14 verideki bağımlılık yapısını aday kapulalar arasından en iyi modelleyen kapula fonksiyonun seçim yönteminden elde edilen sonuçları özetlemektedir. Her iki çizelge için, z ile sözde gözlemler gösterilmektedir. $K_n(z)$ ile kapula fonksiyonunun parametrik olmayan tahmini, $K_C(z)$, $K_G(z)$ ve $K_F(z)$ ile de sırasıyla Clayton, Gumbel ve Frank aday kapulaların dağılımının parametrik tahminleri elde edilmiştir. Son olarak, $K_n(z)-K_C(z)$, $K_n(z)-K_G(z)$ ve $K_n(z)-K_F(z)$ ile dağılım fonksiyonunun örneklemden elde edilen değeri ile teorik olarak elde edilen değeri arasındaki fark gösterilmiştir. Çizelgelerde gösterilen sonuçlar için Matlab ile yazılan program kullanılmıştır.

Çizelge 8.13 $\hat{\theta}_{TAU}$ kullanılarak elde edilen sonuçlar

z	cum	$K_n(z)$	$K_C(z)$	$K_G(z)$	$K_F(z)$	$K_n(z)-K_C(z)$	$K_n(z)-K_G(z)$	$K_n(z)-K_F(z)$
0.00001	2	0.019802	0.00001	0.00005	0.00010	0.00039	0.00039	0.00039
0.05001	12	0.118812	0.06331	0.10203	0.11218	0.00308	0.00028	0.00004
0.10001	17	0.168317	0.12660	0.17997	0.18150	0.00174	0.00014	0.00017
0.15001	25	0.247525	0.18987	0.24882	0.24129	0.00332	0.00000	0.00004
0.20001	34	0.336634	0.25308	0.31178	0.29676	0.00698	0.00062	0.00159
0.25001	41	0.405941	0.31614	0.37035	0.34996	0.00806	0.00127	0.00313
0.30001	44	0.435644	0.37894	0.42542	0.40185	0.00322	0.00010	0.00114
0.35001	50	0.49505	0.44130	0.47759	0.45295	0.00289	0.00030	0.00177
0.40001	52	0.514851	0.50300	0.52727	0.50354	0.00014	0.00015	0.00013
0.45001	59	0.584158	0.56375	0.57478	0.55377	0.00042	0.00009	0.00092
0.50001	62	0.613861	0.62318	0.62035	0.60371	0.00009	0.00004	0.00010
0.55001	68	0.673267	0.68084	0.66418	0.65333	0.00006	0.00008	0.00040
0.60001	73	0.722772	0.73621	0.70643	0.70255	0.00018	0.00027	0.00041
0.65001	75	0.742574	0.78866	0.74723	0.75120	0.00212	0.00002	0.00007
0.70001	77	0.762376	0.83749	0.78670	0.79895	0.00564	0.00059	0.00134
0.75001	80	0.792079	0.88185	0.82492	0.84530	0.00806	0.00108	0.00283
0.80001	86	0.851485	0.92083	0.86199	0.88938	0.00481	0.00011	0.00144
0.85001	90	0.891089	0.95337	0.89797	0.92985	0.00388	0.00005	0.00150
0.90001	95	0.940594	0.97830	0.93293	0.96450	0.00142	0.00006	0.00057
0.95001	97	0.960396	0.99432	0.96693	0.98979	0.00115	0.00004	0.00086
1.00000	100	1.00000	1.00000	1.00000	1.00000	0.05765	0.00569	0.01886

Çizelge 8.13 verilen sonuçlarda aday kapulaların parametre tahminleri için Çizelge 8.12’ de verilen $\hat{\theta}_{TAU}$ değerleri kullanılmıştır. Karşılaştırma kriteri olarak, (8.14) ile verilen Cramer- Von Mises ölçüsü kullanılmıştır. Bu ölçü ile ifade edilen uzaklığı en küçük değerini $\hat{\theta}_{TAU} = 2.88$ değeriyle Gumbel kapula için almaktadır. Buna göre, verideki bağımlılığı modelleyebilecek aday kapulalar içinde en iyi modelleyen kapula

fonksiyonun $\hat{\theta}_{TAU} = 2.88$ parametresiyle Gumbel kapula olduğu söylenir. Verinin $\theta = 3$ ile Gumbel kapuladan simüle edildiği göz önüne alındığında yöntemin beklenen doğrultuda sonuç verdiği görülür.

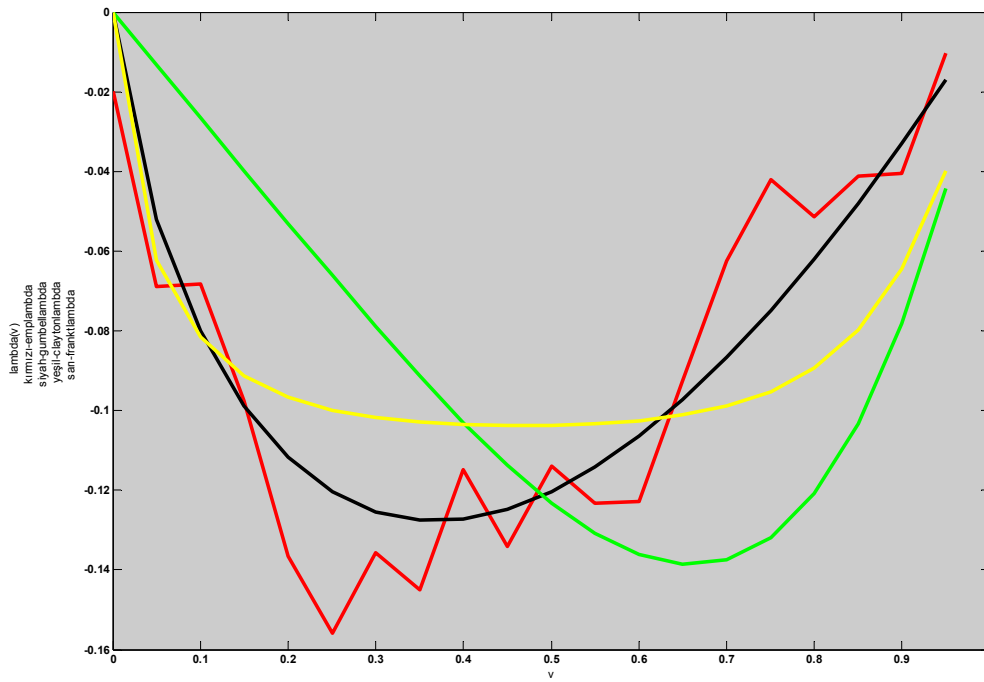
Çizelge 8.14 verilen sonuçlarda aday kapulaların parametre tahminleri için Çizelge 8.12’ de verilen $\hat{\theta}_{MLE}$ değerleri kullanılmıştır. Çizelge 8.13 için yorumlar benzer olarak Çizelge 8.14 içinde yapılır. Buna göre, verideki bağımlılığı modelleyebilecek aday kapulalar içinde en iyi modelleyen kapula fonksiyonun $\hat{\theta}_{MLE} = 2.91$ parametresiyle Gumbel kapuladır.

Çizelge 8.14 $\hat{\theta}_{MLE}$ kullanılarak elde edilen sonuçlar

z	cum	$K_n(z)$	$K_C(z)$	$K_G(z)$	$K_F(z)$	$K_n(z)-K_C(z)$	$K_n(z)-K_G(z)$	$K_n(z)-K_F(z)$
0.00001	2	0.019802	0.00001	0.00005	0.00010	0.00039	0.00039	0.00039
0.05001	12	0.118812	0.07424	0.10149	0.11399	0.00199	0.00030	0.00002
0.10001	17	0.168317	0.14814	0.17914	0.18444	0.00041	0.00012	0.00026
0.15001	25	0.247525	0.22137	0.24780	0.24501	0.00068	0.00000	0.00001
0.20001	34	0.336634	0.29358	0.31063	0.30104	0.00185	0.00068	0.00127
0.25001	41	0.405941	0.36439	0.36911	0.35463	0.00173	0.00136	0.00263
0.30001	44	0.435644	0.43345	0.42413	0.40679	0.00000	0.00013	0.00083
0.35001	50	0.49505	0.50037	0.47628	0.45807	0.00003	0.00035	0.00137
0.40001	52	0.514851	0.56478	0.52596	0.50877	0.00249	0.00012	0.00004
0.45001	59	0.584158	0.62629	0.57349	0.55904	0.00178	0.00011	0.00063
0.50001	62	0.613861	0.68452	0.61911	0.60895	0.00499	0.00003	0.00002
0.55001	68	0.673267	0.73908	0.66300	0.65847	0.00433	0.00011	0.00022
0.60001	73	0.722772	0.78958	0.70533	0.70753	0.00446	0.00030	0.00023
0.65001	75	0.742574	0.83563	0.74623	0.75591	0.00866	0.00001	0.00018
0.70001	77	0.762376	0.87683	0.78581	0.80329	0.01310	0.00055	0.00167
0.75001	80	0.792079	0.91280	0.82415	0.84912	0.01457	0.00103	0.00325
0.80001	86	0.851485	0.94312	0.86135	0.89253	0.00840	0.00010	0.00168
0.85001	90	0.891089	0.96740	0.89748	0.93216	0.00582	0.00004	0.00169
0.90001	95	0.940594	0.98524	0.93259	0.96585	0.00199	0.00006	0.00064
0.95001	97	0.960396	0.99624	0.96675	0.99024	0.00129	0.00004	0.00089
1.00000	100	1.00000	1.00000	1.00000	1.00000	0.07897	0.00583	0.01792

Sonuç olarak, için kapula parametresinin tahmini için her iki yöntemin kullanılması kapulanın tahmin edilmesinde sonucu deęiřtirmemiřtir.

Seřim kriteri olarak (8.2.3) ile gsterilen AIC kriteri kullanıldığında rnek veri seti iin Clayton, Gumbel ve Frank kapulaları arasından en iyi uyum saęlayan kapulanın Gumbel kapula ailesi sylemektedir.



Şekil 8.1 Ampirik ve teorik λ -fonksiyonu grafikleri

Son olarak, Şekil 8.1 ile λ - fonksiyonun veriden elde edilen ve aday kapulaların teorik daęılımlarından elde edilen grafikleri çizdirilmiştir. Ampirik daęılım fonksiyonuna ait grafięinin aday kapulalardan hangisinin daęılım fonksiyonuna ait garifikle temas ettięi nokta fazlaysa, o aday kapulanın (8.14) ile gsterilen lye gre baęımlılıęı modelleyen en iyi kapula olduęu sylenir.

Kırmızı renkte olan fonksiyon, veriden elde edilerek izilen ampirik fonksiyondur. Siyah renkte olan Gumbel kapulaya ait teorik λ -fonksiyonun grafięi, sırasıyla yeşil ve sarı renkte olanlar ise Clayton ve Frank kapulalara ait teorik λ -fonksiyonun grafięidir.

Grafik incelendiğinde, kırmızı renkte olan ampirik dağılım grafiğine en çok noktada temas eden $\hat{\theta}_{MLE} = 2.91$ olan Gumbel kapulanın grafiğidir.

8.4 Yarışan Risklerin Varlığında Bağımlılığın Modellenmesinde Kapula Yaklaşımı

Yarışan risklerin varlığındaki sağkalım analizinde genel yaklaşım, risklerin birbirinden bağımsız davrandığının varsayılmasıdır. Bu varsayım, örtük başarısızlık sürelerinin $(T_j, j = 1, \dots, k)$ bağımsızlığı anlamına gelmektedir. Bağımsızlık varsayımı ile birlikte bölüm 6'da verilen belirlenebilme problemi ortadan kalkmakta ve $(T_j, j = 1, \dots, k)$ 'lere ait marjinal dağılımlar ve $(T_1, T_2, \dots, T_k)$ 'ye ait ortak dağılım veriden elde edilen nedene özel fonksiyonlar tarafından belirlenebilmektedir.

Bu çalışmada, risklerin bağımlı olması durumunda örtük başarısızlık sürelerine ait ortak dağılımın elde edilmesi yani riskler arasındaki bağımlılık yapısını modellemek amaçlanmıştır. Bu amaçla, riskler arasındaki bağımlılığı modellemek için Archimedean fonksiyonlarının kullanılması önerilmiştir.

Bağımlılığı modelleyecek Archimedean kapula fonksiyonun parametrik olmayan tahmini için marjinal dağılımlara ihtiyaç vardır. Yarışan risklerin varlığındaki çalışmalarda elde edilen dağılım fonksiyonları nedene dağılım özel fonksiyonlardır. Nedene özel dağılım fonksiyonları, önceki bölümlerde de ifade edildiği gibi tüm risklerin varlığında belirli bir riske ait olasılıkları veren fonksiyonlardır. Yarışan risklerin bağımlılığı göz önüne alındığında bu fonksiyonlar marjinal fonksiyonları dolayısıyla da ortak dağılımı belirleyememektedir.

Yarışan risklerin olduğu düzenlerde veriden elde edilebilecek tüm bilgi nedene özel fonksiyonlardır. Bu nedenle veriden elde edilen tüm bilgi kullanılarak marjinal dağılımların tahminleri elde edilmelidir. Bu amaçla, Chiang (1970) 'ın önerdiği (8.16) ile gösterilen tahmin edici kullanılarak, nedene özel dağılımlar kullanılarak marjinal dağılımların tahminlerine ulaşılabilir.

$$1 - \hat{G}_j(t) = \hat{F}_T(t)^{\frac{\hat{F}_j(t)}{\hat{F}_1(t) + \hat{F}_2(t) + \dots + \hat{F}_k(t)}}, \quad j = 1, 2, \dots, k \quad (8.16)$$

(8.16) aslında bir koşullu olasılıktır. t noktasında veya belirlenen aralıkta diğer risklerin olmadığı bilindiğinde elde edilen olasılıktır. Bu haliyle bile marjinal dağılıma nedene özel dağılım fonksiyonlarından daha yakın olmaktadır.

Çalışmada, risklere ait marjinal dağılımların tahmini için (8.16) ile verilen tahmin edici ve risklerin arasındaki bağımlılığın modellenmesi için ise Archimedean kapula fonksiyonunun tahmininin elde edildiği bölüm 8.2 ile verilen yöntem önerilmiştir.

Çalışmanın dokuzuncu bölümünde, önerilen yöntem gerçek bir veri seti üzerinde uygulanmıştır.

9. UYGULAMA

Aralık 1992 tarihinde, meme kanseri tanısı konmuş hastalara uygulanan tedaviler süresince ve sonrasında tedavi etkinliğinin araştırılması amacıyla hastaların durumunun izlendiği çok merkezli rasgele bir klinik deneme başlatılmıştır. Bu amaçla, 320 hasta tamoxifen adlı ilaçla tedavi edilmeye çalışılırken, diğer 320 hastaya da tamoxifen ve radyoterapi tedavisi uygulanmıştır. Kombine tedavinin tek ilaçla tedaviye üstünlüğünün belirlendiği bu çalışmada son kayıt 2002 yılının yazında tutulmuştur. Çalışmanın detayları (Fyles vd. 2004) çalışmasında yer almaktadır. Söz konusu çalışmadan elde edilen verilerin bir kısmı EK1 ile ve değişkenler ile ilgili açıklamalar EK2 ile verilmiştir. Verinin çok boyutu ve 640 hastaya ait olması nedeniyle veriye ait fikir vermek amacıyla sadece bir kısmına verilmiştir. Verinin tamamına ilgili referanstan ulaşılabilir. (Fyles vd. 2004)'nin yürüttüğü çalışmada meydana geliş izlenmiş beş olay vardır. Bunlar, bölgesel nüks (local relaps), koltuk altı lenf nodu bölgesinde nüks (axillary relaps), uzak organ yayılımı (distant relaps), ikinci bir kötü huylu tümörün meydana gelmesi ve hastanın ölümüdür. Sözü edilen bu olayların her biri hasta için birer ölüm riskidir. Bu beş olayın her birinden ilk olarak gerçekleşen zaman sağkalım süresi olarak kaydedilmiştir. Örneğin, bir hastada ilk yılda bölgesel nüks, ikinci yılda başka bir bölgesel nüks ve üçüncü yılda ise koltuk altı lenf nodu bölgesinde nüks gerçekleşmiş olsun. Bu durumda, ikinci bölgesel nüks kaydedilmemiş, sadece ilk nüks ve koltuk altı lenf nodu bölgesindeki nüks gerçekleşene kadar geçen süre kaydedilmiş ve veriler oluşturulmuş.

Çalışmada, klinik olarak aralarındaki ilişkinin varlığının anlamlı olması nedeniyle uzak organ yayılım riski ve koltuk altı lenf nodu bölgesinde nüks ortaya çıkması riski göz önüne alınmıştır. Çalışmada, uzak organ yayılım riski ve koltuk altı lenf nodu bölgesinde nüks olayları arasındaki bağımlılık her iki tedavi türüne göre kapula ile modellenmiştir. Uzak organ yayılım riski ve koltuk altı lenf nodu bölgesinde nüks riskinin arasındaki bağımlılığın modellenmesi için elde edilen sonuçlar sadece tamoxifen ile tedavi edilen grup için Tablo 9.1 ve kombine tedavi uygulanan hasta grubu için ise Tablo 9.2 ile özetlenmiştir.

Çizelge 9.1 Tamoxifen tedavisi altındaki hasta grubu için uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks riski arasındaki bağımlılık modelinin seçimi

z	cum	$K_n(z)$	$K_C(z)$	$K_G(z)$	$K_F(z)$	$K_n(z)-K_C(z)$	$K_n(z)-K_G(z)$	$K_n(z)-K_F(z)$
0.00001	2	0.026316	0.00002	0.00003	0.00009	0.00069	0.00069	0.00069
0.05001	6	0.078947	0.08456	0.07797	0.07845	0.00003	0.00000	0.00000
0.10001	18	0.236842	0.16746	0.14299	0.13136	0.00481	0.00881	0.01113
0.15001	23	0.302632	0.24811	0.20313	0.18193	0.00297	0.00990	0.01457
0.20001	24	0.315789	0.32606	0.26009	0.23205	0.00011	0.00310	0.00701
0.25001	24	0.315789	0.40096	0.31470	0.28208	0.00725	0.00000	0.00114
0.30001	26	0.342105	0.47251	0.36743	0.33208	0.01701	0.00064	0.00010
0.35001	28	0.368421	0.54044	0.41859	0.38208	0.02959	0.00252	0.00019
0.40001	33	0.434211	0.60449	0.46842	0.43208	0.02899	0.00117	0.00000
0.45001	36	0.473684	0.66444	0.51708	0.48208	0.03639	0.00188	0.00007
0.50001	39	0.513158	0.72007	0.56470	0.53208	0.04281	0.00266	0.00036
0.55001	42	0.552632	0.77120	0.61138	0.58208	0.04777	0.00345	0.00087
0.60001	46	0.605263	0.81762	0.65722	0.63208	0.04510	0.00270	0.00072
0.65001	50	0.657895	0.85917	0.70227	0.68208	0.04051	0.00197	0.00059
0.70001	55	0.723684	0.89567	0.74661	0.73208	0.02958	0.00053	0.00007
0.75001	63	0.828947	0.92696	0.79028	0.78207	0.00961	0.00150	0.00220
0.80001	65	0.855263	0.95289	0.83333	0.83202	0.00953	0.00048	0.00054
0.85001	67	0.881579	0.97329	0.87579	0.88178	0.00841	0.00003	0.00000
0.90001	69	0.907895	0.98804	0.91771	0.93066	0.00642	0.00010	0.00052
0.95001	73	0.960526	0.99699	0.95910	0.97535	0.00133	0.00000	0.00022
1.00000	75	1.00000	1.00000	1.00000	1.00000	0.36892	0.04213	0.04097

Tamoxifen tedavisi altındaki hasta grubu için, uzak bölge tutulum riski ve ikinci bir tümör meydana gelme riski arasındaki bağımlılık modelinin seçimi aday kapulaların birliktelik parametresinin tahmini en çok olabilirlik yöntemi kullanılarak elde edilmiştir. Buna göre, $\hat{\theta}_{MLE}$ değeri, sırasıyla Gumbel, Clayton ve Frank kapula için 5.357594, 1.427231 ve 31.17925 olarak bulunmuştur. Bölüm (8.2) ile anlatılan yollar izlenerek elde edilen sonuçlar Çizelge 9.1 ile özetlenmiştir. Buna göre, uzak bölge tutulum riski ve ikinci bir tümör meydana gelme riski arasındaki bağımlılık yapısını en iyi modelleyen kapula Frank ($\hat{\theta}_{MLE} = 31.17325$) kapuladır. Buna göre, sadece tamoxifen ile tedavi edilen hastalar için uzak organ yayılım riski ve koltuk altı lenf nodu bölgesinde nüks meydana gelme riski arasındaki bağımlılık kuyruklarda zayıf, dağılım ortasında

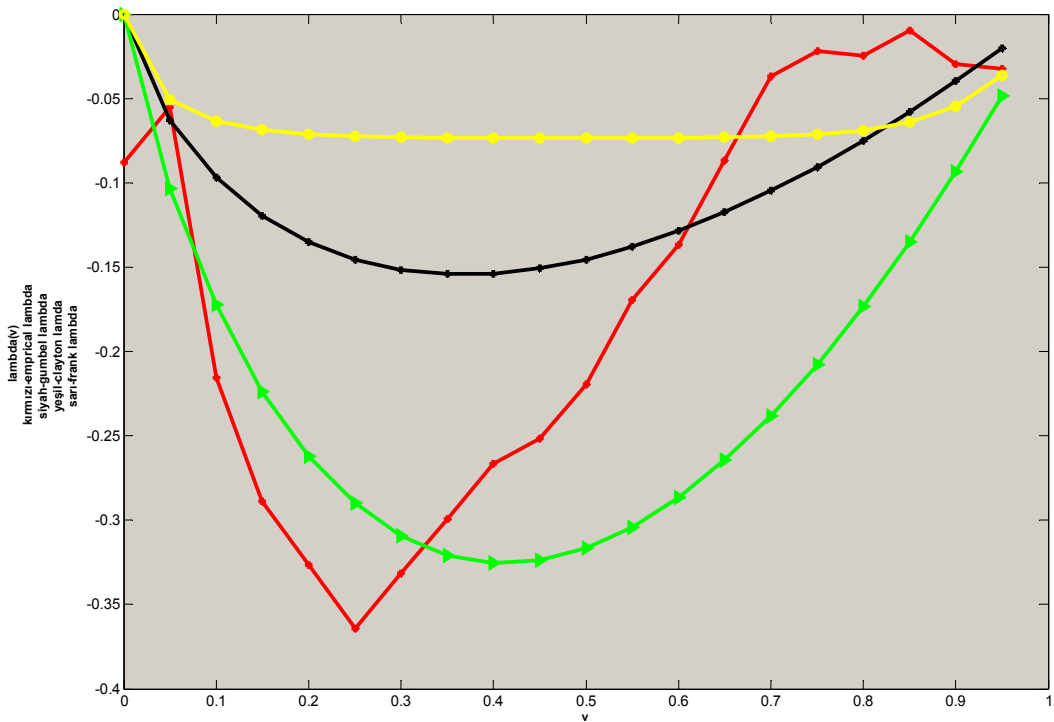
güçlüdür. Ayrıca, söz konusu iki risk için için Kendall'ın ilişki katsayısı 0.8515315 olarak bulunması da bu iki riskin güçlü pozitif bağımlılığa sahip olduğunu göstermektedir.

Çizelge 9.2 Tamoxifen ve radyoterapi tedavisi altındaki hasta grubu için koltukaltı lenf nodu bölgesinde nüks riski uzak organ yayılım riski arasındaki bağımlılık modelinin seçimi

z	cum	$K_n(z)$	$K_C(z)$	$K_G(z)$	$K_F(z)$	$K_n(z)-K_C(z)$	$K_n(z)-K_G(z)$	$K_n(z)-K_F(z)$
0.00001	5	0.08772	0.00005	0.00006	0.00010	0.00769	0.00768	0.00768
0.05001	6	0.10526	0.15335	0.11287	0.10071	0.00231	0.00006	0.00002
0.10001	18	0.31579	0.27231	0.19664	0.16334	0.00189	0.01420	0.02324
0.15001	25	0.43860	0.37360	0.26942	0.21874	0.00423	0.02862	0.04834
0.20001	30	0.52632	0.46198	0.33508	0.27129	0.00414	0.03657	0.06504
0.25001	35	0.61404	0.53998	0.39544	0.32253	0.00548	0.04778	0.08497
0.30001	36	0.63158	0.60918	0.45157	0.37316	0.00050	0.03240	0.06678
0.35001	37	0.64912	0.67070	0.50419	0.42346	0.00047	0.02100	0.05092
0.40001	38	0.66667	0.72536	0.55381	0.47361	0.00345	0.01274	0.03727
0.45001	40	0.70175	0.77381	0.60079	0.52367	0.00519	0.01019	0.03171
0.50001	41	0.71930	0.81657	0.64544	0.57367	0.00946	0.00546	0.02121
0.55001	41	0.71930	0.85406	0.68798	0.62361	0.01816	0.00098	0.00916
0.60001	42	0.73684	0.88663	0.72862	0.67346	0.02244	0.00007	0.00402
0.65001	42	0.73684	0.91460	0.76750	0.72315	0.03160	0.00094	0.00019
0.70001	42	0.73684	0.93822	0.80477	0.77253	0.04055	0.00461	0.00127
0.75001	44	0.77193	0.95773	0.84054	0.82130	0.03452	0.00471	0.00244
0.80001	47	0.82456	0.97333	0.87491	0.86889	0.02213	0.00254	0.00196
0.85001	49	0.85965	0.98521	0.90797	0.91413	0.01576	0.00234	0.00297
0.90001	53	0.92982	0.99351	0.93980	0.95477	0.00406	0.00010	0.00062
0.95001	56	0.98246	0.99840	0.97045	0.98632	0.00025	0.00014	0.00001
1.00000	56	1.00000	1.00000	1.00000	1.00000	0.23429	0.23313	0.45983

Tamoxifen ve radyoterapi kombine tedavisi uygulanan hasta grubu için, uzak organ yayılım riski ve ikinci bir tümör meydana gelme riski arasındaki bağımlılık modelinin seçimi aday kapulaların birliktelik parametresinin tahmini de en çok olabilirlik yöntemi kullanılarak elde edilmiştir. Buna göre, $\hat{\theta}_{MLE}$ değeri, sırasıyla Gumbel, Clayton ve Frank kapula için 2.383121, 0.2654083 ve 13.55248 olarak bulunmuştur. Aynı şekilde, bölüm (8.2) ile anlatılan yollar izlenerek elde edilen sonuçlar Çizelge 9.2 ile

özetlenmiştir. Buna göre, uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks meydana gelme riski arasındaki bağımlılık yapısını en iyi modelleyen kapula yine Gumbel ($\hat{\theta}_{MLE} = 2.383121$) kapula olarak bulunmuştur. Gumbel kapula ile modellenen bu veri için sağ kuyruk bağımlılığı daha güçlü iken sol kuyruk bağımlılığı ise daha zayıftır. Yani, Tamoxifen ve radyoterapi kombine tedavisi uygulanan hasta grubu için, uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks meydana gelme riski birlikte artış göstermeye birlikte azalış göstermekten daha yatkın olduğunu söylenebilir.



Şekil 9.1 Kombine tedavi altında iki riske ait ampirik ve teorik λ fonksiyonu grafikleri

Şekil 9.1 incelendiğinde, kombine tedavi uygulanan hasta grubuna için koltukaltı lenf nodu bölgesinde nüks riski uzak organ yayılım riski için ampirik dağılım kırmızı ile gösterilmiştir. Bu dağılım, siyah ile gösterilen Gumbel kapula ile en fazla temas noktasına sahiptir. Bu durumda, koltukaltı lenf nodu bölgesinde nüks riski ile uzak organ yayılım riski arasındaki bağımlılığı en iyi modelleyen kapula fonksiyonunun gumbel kapula olduğu söylenir.

10. TARTIŞMA VE SONUÇ

Sağkalım analizi, çalışmanın başında belirlenen ve başarısızlık olarak adlandırılan bir sonucun ortaya çıkmasına kadar geçen süre ile ilgilenir. Bu şekilde elde edilen verilen sağkalım tipi veriler olarak adlandırılmaktadır. Bu şekilde elde edilen verilerin analizi için sağkalım fonksiyonunun ve hazard fonksiyonunun yorumlanması önemli olmaktadır. Bu nedenle çalışmada öncelik bu fonksiyonların tanıtılmasına ve tahmin edicilerinin elde edilmesi üzerinde yoğunlaşmıştır. İlk olarak, $S(t) = P(T > t)$ olarak tanımlanan sağkalım fonksiyonu için en iyi tahmin edici olan Kaplan-Meier tahmin edicisi incelenmiş ve parametrik olmayan en çok olabilirlik tahmin edicisi olarak nasıl elde edileceği gösterilmiştir. Diğer önemli bir fonksiyon olan hazard fonksiyonuna ait Nelson-Aalen tahmin edicisi incelenerek iki fonksiyon arasındaki ilişkiye dikkat çekilmiştir. Bu iki tahmin ediciye ait, ortalama ve varyans hesaplanarak istatistiksel özellikleri incelenmiş, yansızlık ve tutarlılık özellikleri ortaya konmuştur.

Çalışmada, bioistatistik, güvenilirlik analizi, finans gibi birçok alanda uygulamaları bulunan yarışan riskler teorisi incelenmiştir. Yarışan riskler teorisi, canlı bir organizmanın ya da bir sistemin başarısızlığına neden olan iki veya daha fazla riskin var olduğu durumlarla ilgilenir. Bu riskler başarısızlık ortaya çıkmadan önce birer risk iken başarısızlık ortaya çıktıktan sonra başarısızlığın nedeni olarak adlandırılmaktadır. Bir anlamda bu riskler neden olmak için yarışmaktadır. İlk bölümlerde tanıtılan sağkalım ve hazard fonksiyonları yarışan riskler analizi içinde önemli fonksiyonlardır. Farklı olarak, birden fazla riskin olduğu durumlar için tüm risklerin varlığında başarısızlığın ortaya çıkmasına neden olan riske özel elde edildiği için, nedene özel hazard, nedene özel sağkalım fonksiyonu gibi isimler almaktadır. Yarışan riskler analizinde başarısızlığın nedeni olabilecek tüm risklerin varlığı göz önüne alınmaktadır. Genel yaklaşım bu risklerin birbirlerinden bağımsız olduğunu varsayarak ilgili fonksiyonların elde edilmesidir. Fakat risklerin birbirlerini etkiledikleri durumlarda bu yaklaşım gerçekçi olmamaktadır. Bu nedenle bu risklerin arasındaki ilişki yapısının ortaya konması önemli olmaktadır. Çalışmada söz konusu risklerin birbirinden bağımsız olmadığı durumlarda aralarındaki bağımlılık yapısının nasıl araştırılabileceği ve yorumlanabileceği ortaya konmuştur.

Yarışan risklerin varlığındaki çoğu çalışmalarda, özellikle başarısızlık ölüm ise, tek bir riskten dolayı ortaya çıkan başarısızlık gözlenebilmektedir. Ortaya çıkması gözlenen başarısızlığa neden olan riskin meydana geliş süresi $T = \min(T_1, T_2, \dots, T_k)$ olarak ifade edilir. Böylece, ortaya çıkması süresi en kısa olan risk gözlenir. Bu durumda rasgele değişkenin çok boyutlu bir rasgele değişken vektörü olabileceği göz önüne alındığında, tüm risklere ait sağkalım ya da başarısızlık sürelerinin gözlenebilir rasgele değişkenler olmadığı ortaya çıkmaktadır. Değişkenler arasındaki bu yapıya örtük (latent) yapı adı verilmektedir. Çalışmada, bu yapıya sahip değişkenlere ait ortak dağılımın yani riskler arasındaki bağımlılık yapısının nasıl belirlenebileceği araştırılmıştır. Bağımlılık durumunda gözlem sonucu elde edilen nedene özel dağılım fonksiyonlar yardımıyla bu ortak dağılım yani riskler arasındaki bağımlılık yapısının belirlenememesi literatürde belirlenebilme sorunu olarak adlandırılmaktadır. Çalışmada, bu sorun Gumbel'in önerdiği için iki boyutlu üstel dağılım fonksiyonları kullanılarak örneklendirilmiştir.

Çalışmanın amacına uygun olarak bu riskler arasındaki bağımlılık yapısını daha detaylı bir biçimde ortaya koyabilecek bir araç olarak kapula fonksiyonları seçilmiştir. Özellikle finans alanında çok yaygın olarak kullanılan kapula fonksiyonları günümüzde yaygın olarak başka alanlarda da kullanılmaya başlanmıştır. Çok kullanışlı bir araç olan kapula teorisine değinilmiştir. Bioistatistik, güvenilirlik gibi alanlarında daha çok karşılaşılan sağkalım tipi verilerin modellenmesi ve matematiksel özellikleri nedeniyle önemli bir yere sahip olan Archimedean kapula sınıfıyla çalışmak tercih edilmiştir. Özellikle, bu aileye ait olan Gumbel, Clayton ve Frank kapulaları ile çalışılmıştır.

İki boyutlu kapula teorisini çok boyutlu kapula teorisine genişletilmesinde çözülmemiş bir takım problemler olduğundan çalışmada sadece iki boyutlu Archimedean kapula fonksiyonları ile çalışma yürütülmüştür. Bu aileye ait kapulalardan, Gumbel, Clayton ve Frank kapulalar kullanılarak Genest ve River (1993)'in önerdiği parametrik olmayan yöntem temel alınmış ve gözlemlerin bağımlılığını en yakın modelleyen kapula fonksiyonu seçilmiştir. Öncelikle söz konusu yöntem detaylı olarak incelenmiş ve sonuçlar yorumlanmıştır. Çalışmada bağımlılığın Archimedean kapula ailesinden Clayton, Gumbel ve Frank kapulalarından biriyle modellendiği varsayılmıştır. Özellikle birliktelik parametresinin tahmini üzerinde durulmuştur. Momentler yöntemi ile elde

edilen ve Kendall'ın ilişki katsayısı kullanılarak elde edilen TAU tahmin edicisi ile en çok olabilirlik yöntemi ile elde edilen MLE tahmin edicisi karşılaştırılmıştır ve simülasyon sonuçları sunulmuştur. Sonuçlara göre, MLE tahmin edicinin daha iyi bir tahmin edici olduğu görülmüştür fakat TAU tahmin edicisinin de MLE'ye göre çok kötü bir tahmin edici olmadığına dikkat çekilmiş ve bazı aileler için kapalı formunu olması ve hesaplama kolaylığı nedeniyle tercih edilebilir olduğu vurgulanmıştır. Çalışmada marjinal dağılımların tahmini için ampirik dağılımlar kullanılmış, bilinen parametrik dağılımların kullanılması daha sonraki çalışmalara bırakılmıştır. Yöntemin uygulanabilirliğinin denetlenmesi Gumbel kapuladan simüle edilen bir veri seti üzerinde yapılmıştır. Kapula parametresini tahmini için Kendall'ın tau ilişki katsayısını kullanan momentler tahmin yöntemi ile en çok olabilirlik yöntemi kullanılmıştır. Uygun modelin seçim kriteri olarak Cramer-Von Misses uzaklığı kullanılarak elde edilen sonuçlar sunulmuştur. Her iki tahmin edici ve seçim kriteri birbirini destekler nitelikte sonuçlar vermiştir.

Çalışmada, birliktelik parametresini tahmini ve simülasyon sonuçları için R programı ile ve model seçimi için MATLAB ile yazılan programlar kullanılmıştır. Ayrıca uygulama kısmında kullanılan veri seti için nedene özel fonksiyonların tahminlerinin elde edilmesine ilişkin hesaplamalar SAS programı kullanılarak yapılmıştır.

Yöntemin uygulanma aşamalarını ortaya koymak adına gumbel kapuladan üretilen veri seti üzerinde çalışılmış, marjinal dağılımların tahmini için ampirik dağılımlar olarak kullanılmıştır. Yüksek bağımlılığın olması ve örneklem genişliğinin az olması nedeniyle birliktelik parametreleri tahminleri için MLE tahmin edicileri ve Kendall'ın ilişki katsayısına dayanan tahmin edici kullanılmış ve sonuçların değişmediği görülmüştür.

Yarışan risklerin varlığında ve bu risklerin bağımlı olduğu varsayımı altında bu bağımlılığı modelleyecek kapula fonksiyonun seçimi Genest ve Rivest (1993) 'in önerdiği ve çalışmada incelenen yöntem kullanılarak yapılmıştır. Bu amaçla, Fyles vd. (2004) yürüttüğü bir çalışmadan elde edilen veriler kullanılarak tanıtılan yöntemin yarışan risklerin varlığında, riskler arasındaki bağımlılığın modellenmesine uyarlanabilirliği tartışılmıştır.

Yarışan risklerin varlığındaki çalışma düzenlerinde, ortaya çıkma süresi en kısa olan risk gözlenmektedir. Bu nedenle de her riske ait elde edilen dağılım fonksiyonunu nedene özel dağılım fonksiyonu olmaktadır. Belirlenebilme problemini varlığı göz önüne alındığında, bir takım varsayımlar yapmadan veriden elde edilen nedene özel dağılım fonksiyonları kullanılarak ortak dağılımın belirlenmesi mümkün olmamaktadır. Bu nedenle, her bir riske ait marjinal dağılım fonksiyonların elde edilmesi gerekmektedir. Marjinal dağılım fonksiyonları, sadece tek riskin var olduğu varsayıldığı durumdaki marjinal dağılımları ifade etmektedir. Bu marjinal dağılım fonksiyonlarının tahminin elde edilmesi için (8.16) ile gösterilen Chiang (1970) 'ın önerdiği tahmin edici kullanmıştır. Birliktelik parametresini tahmin edicisi için ise bu çalışmada en çok olabilirlik tahmin edicisi kullanılmıştır.

Meme kanseri tanısı konmuş hastalar için meydana gelişi gözlenen ve ölüm riski olarak belirlenen beş olaydan uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks riski seçilerek bu iki risk arasındaki bağımlılığı modelleyen Archimedean kapula fonksiyonu seçilmiş ve yorumlanmıştır. Aralarındaki bağımlılığın klinik olarak anlamlı olması nedeniyle bu iki risk seçilmiştir. Sadece tamoxifen adlı ilaçla tedavi edilen hasta grubu için uzak organ yayılım riski ve koltukaltı lenf nodu bölgesinde nüks riski arasındaki bağımlılığın Frank kapula ile modellenemediği görülmüştür. Aynı şekilde, tamoxifen ve radyoterapi uygulanan hasta grubu içinde aynı risklerin bağımlılık yapısı araştırılmış ve Gumbel kapula ile modellenebileceği görülmüştür.

Bağımlı arışan risklerin varlığında örtük başarısızlık sürelerine ait ortak dağılımın belirlenmesi diğer bir ifadeyle riskler arasındaki bağımlılık yapısının belirlenmesi bir takım varsayımlar yapmadan mümkün olmamaktadır. Çalışmada yarışan risklerin bağımlılığı göz önüne alındığında varsayımları en az seviyede tutarak gözlemlerden elde edilen bilgiler doğrultusunda bağımlılık yapısının belirlenmesine çalışılmıştır.

Kapula teorisinin henüz gelişmekte olan bir alan olması çalışmanın devamı için bir takım düşünceler ortaya çıkmasına olanak tanımıştır. Çalışmada, gelişmekte olan bu

alanın yarışan riskler teorisinde bağımlılığın modellenmesi için, uygulanabilirliği ve geliştirilebilirliği için temel oluşturulmaya çalışılmıştır.

Çalışmanın devamı için çok boyutlu durumlar incelebilir. Ayrıca, belirlenebilme problemi varlığında net dağılımların tahmin edilmesi için daha uygun tahmin edicilerin araştırılması düşünülebilir. Yarışan riskler teorisinde önemli bir kısıt olan belirlenebilme problemi için, bu problemin olmadığı farklı parametrik ailelerin araştırılması araştırmanın devamı için düşünülen bir diğer konu olabilir.

KAYNAKLAR

- Ali, M.M. Mikhail, N.N. and Haq, M.S. 1978. A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, Vol. 8, pp.405-412.
- Arnold, B.C. and Brockett, P.L. 1983. Identifiability for dependent multiple decrement/competing risk models. *Scand. Actuar. J.* 117–127.
- Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. 1997. *Actuarial Mathematics*. Itasca, III: Society of Actuaries.
- Birnbaum, Z.W. 1979. On the mathematics of competing risks. *Data evaluation and methods research*, (2), 77.
- Bryant, J. and Dignam, J.J. 2004. Semiparametric models for cumulative incidence Functions. *Biometrics*, 60, 182-190.
- Carriere, J. 1994. Dependent decrement theory. *Transactions, society of Actuaries XLVI*, 45-65.
- Chiang, C.L. 1970. Competing risks and conditional probabilities. *Biometrics* 26, 767-76.
- Clayton, D.G. 1978. A Model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141–151.
- Cook, R.D. and Johnson, M.E. 1981. A family of distributions for modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society B* 43:210–218.
- Cox, D. R. 1959. The analysis of exponentially distributed lifetimes with two types of failure. *Journal of the Royal Statistical Society Seires. B* 21:411-421.
- Crowder, M. 2001. *Classical Competing Risks*, Chapman & Hall.
- Crowder, M. 1997. A test for independence of competing risks with discrete failure times. *Lifetime Data Analysis*, (3), 215-223.
- David, H.A. and Moeschberger, M. L. 1978. *The theory of competing risks*. Griffin, London.
- Deshpande, J. V. and Purohit, S. G. 2001. Survival, hazard and competing risks. *Current Science*, Vol:80, No:9.

- Deheuvels, P. 1978. Caractérisation complète des Lois extrêmes multivariées et de la convergence des types extrêmes, Publ. l'Institut de Statist. L'université de Paris, 23, 1-36.
- Dimitrova, D.S., Haberman, S., and Kaishev, V.K. 2009. Dependent competing risks: Cause elimination. Faculty of Actuarial Science and Insurance, Cass Business School, City University, London
- Elandt-Johnson R.C. 1976. Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. *Scandinavian Actuarial Journal*, 37-51.
- Escarela, G. and Carriere, J. 2003. Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, 12 (4), 333-349.
- Fleming, T.R. and Harrington D.P. 2005. *Counting Processes and Survival Analysis*. John Wiley and Sons, 429, Hoboken, New Jersey.
- Frees W.E. and Valdez A.E. 1997. Understanding relationships using copulas. 32nd. Actuarial Research Conference, 6-8 August at University of Calgary, Albert Canada.
- Frank, M.J. 1979. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. *Aequationes Mathematicae*, 19, 194-226.
- Fyles, A. W., McCready, D. R., Manchul, L. A., Trudeau, M. E., Merante, P., Pintilie, M., Weir, L. M. and Olivotto, I. A. 2004. Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer. *New England Journal of Medicine*, 351, 963-970.
- Gail, M. 1975. A review and critique of some models used in competing risk analysis. *Biometrics*, 31, 209-222.
- Gaynor J., Feuer, E., Tan, C., Wu, D., Little, C., Straus D., Clarkson B. and Brennan M. 1993. On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *American Statistical Association* (88), 400-409.
- Genest, C. and Mackay, J. 1986b. The Joy of Copulas: Bivariate Distributions with Uniform Marginals. *The American Statistician*, 40, 280-283.
- Genest, C. 1987. Frank's family of bivariate distributions. *Biometrika*, 74, 549-555.
- Genest, C. and Rivest L.P. 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of The American Statistical Association Theory and Methods* (88), No: 423

- Genest, C., Ghoudi, K. and Rivest L.P. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543-552.
- Gray, R. J. 1988. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*, 19, 1141-1154.
- Greenwood, M. 1926. The natural of cancer. Reports on Public Health and Medical Subjects 33:1-26, London: Her Majesty's Stationery Office.
- Gumbel, E.J. 1960. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698-707.
- Gooley, T. A., Leisenring, W., Crowley, J. and Stroer, B. E. 1999. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, 18, 695-706.
- Heckman, J. and Honore, B.E. 1989. The identifiability of the competing risk model. *Biometrika*, 76 (2), 325-330.
- Hougaard, P. 1987. Modeling multivariate survival. *Scandinavian Journal of Statistics*, 14, 291-304.
- Hougaard, P. 2000. Analysis for Multivariate Survival Data. *Statistics for Biology and Health*. Springer-Verlag, New York.
- Joe, H. and Xu, J. J. 1996. The estimation method of inference functions for margins formultivariate models; Technical Report no. 166, Department of Statistics, University of British Columbia
- Joe, H. 1997. *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York.
- Kaarik E. 2006. Imputation algortihm using copulas. *Metodološki zvezki* ,(3), No:1, 109-120.
- Kaarik M. and Kaarik E. 2010. Imputation by gaussian copula model with an application to incomplete Customer Satisfaction Data. *COMPSTAT*, Paris.
- Kaishev, K.V., Dimitrova, D. S. and Haberman S. 2007. Modelling the joint distribution of competing risks survival times using copula funcitons. *Insurance: Mathematics and Economics*, 41, 339-361.
- Kalbfleisch, J. D., Prentice, R. L., 1980. *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kaplan, E.J. and Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal American Statistics Association*, (53), 457-481.

- Kim, H.T. 2007. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research*, (13), 2.
- Kimball, A. W. 1969. Models for the estimation of competing risks from grouped data. *Biometrics*, 25, 329-37.
- Kleinbaum, D.G. and Klein, M. 2005. *Survival Analysis A Self- Learning Text*, Springer, 332.
- Lagakos, S. W. 1979. General right censoring and its impact on the analysis of survival data. *Biometrics*, (35); 139-156.
- Lawless, J. F. 2003. *Statistical Models and Methods for Lifetime Data*, 2nd edn, Wiley.
- Lindqvist, B.H. 2006. Eqr 067 Competing Risks. Lecture Notes.
- Ling, C. H. 1965. Representation of associative functions. *Publicationes Mathematicae*, 12, 189-212.
- Link, A. W. 1989. A model for informative censoring. *Journal of the American Statistical Association*, (84), No:407.
- Lobo, I. 2008. Epistasis: Gene interaction and the phenotypic expression of complex diseases like Alzheimer's. *Nature Education* 1, 1.
- Makeham, W.M. 1874. On an application of the theory of decremental forces. *Journal of the Institute of Actuaries*, 18, 317–322.
- Melchiori R. M. 2003. Which Archimedean copula is the right one? *YieldCurve.com* (e-Journal)
- Moeschberger, M. L. and David, H. A. 1971. Life tests under competing causes of failure and the theory of competing risks. *Biometrics*, (4), 909-933.
- Nadas, A. 1971. The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics* 13, 201-202.
- Naifar N. 2011. Modelling dependence structure with Archimedean copulas and applications iTraxx CDS index. *Journal of Computational and Applied Mathematics*, (235), 2459-2466.
- Nelsen, R.B. 1986. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics—Theory and Methods*, 15, 3277–85.
- Nelsen, R. 1999. *An Introduction to Copulas*. Springer, New York.

- Nelson, W. B. 1969. Hazard plotting for incomplete failure data. *J. Qual. Technol.* 1, 27-52.
- Oakes, D. 1982. A Model for association in bivariate survival data. *Journal of The Royal Statistical Society. Ser B.* (44), 414-422.
- Oakes, D. 1989. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84:487–493.
- Pepe, M. S. and Mori, M. 1993. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risk failure time data. *Statistics in Medicine*, 16,901-910.
- Phelps, A. L. and Weissfeld, L. A. 1997. A comparison of dependence estimators in bivariate copula models. *Communications in Statistics-Simulation and Computation.* 26(4):1583-1597.
- Pintile, M. 2006. *Competing Risks, A Practical Perspective.* Wiley, England.
- Prentice, R. L., Kalbfleisch, J. D., Peterson, A.V., Flournoy N., Farewell, V. T. and Breslow, N. E. 1978. The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541-554.
- Salinas-Torres, V. H., de Brança Pereira, C. A., and Tiwardi, R.C. 2002. Bayesian nonparametric estimation in a series or a competing risk model. *Journal of Nonparametric Statistics*, 14, 449-458.
- Sankaran, P.G., Lawlees, J.F., B.Abraham and Antony A.A. 2006. Estimation of distribution function in bivariate competing risk models. *Biometrical Journal*, (48), 399-410.
- Satagopan, J.M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D. and Auerbach, A. D. 2004. A note competing risks in survival data analysis. *British Journal of Cancer*, 91, 1229-1235.
- Schweizer, B. and Wolff, E.F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9:879–885.
- Seal, H.L. 1977. Multiple decrements or competing risks. *Biometrika*, 64:429–439.
- Serfling, R.J. 1980. *Approximation theorems of mathematical statistics*, Wiley, New York
- Shih, J.H. and Louis, T.A. 1995. Inferences on the Association Parameter in Copula Models for Bivariate Survival Data, *Biometrics* 51:1384–1399.
- Sklar, A. 1959. Functions de repartition a n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* 8, 229-231.

- Sklar, A. 1973. Random Variables, Joint Distribution Functions and Copulas. *Kybernetika*, 9:449–460.
- Southern, D.A., Faris, P.D., Brant, R., Galbraith, P.D., Norris, C. M., Knudtson, M.L. and Ghali, W. A. 2006. Kaplan-Meier methods yielded misleading results in competing risks scenarios. *Journal of Clinical Epidemiology*, (59), 1110-1114.
- Slud, E. and Rubinstein, L. 1983. Dependent competing risks and summary survival curves. *Biometrika*(70), 643-649.
- Tsiatis, A. 1975. A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Academic Science*, (72), USA.
- Weir, M.R. 2005. Cardiovascular disease in patients with chronic kidney disease: The CVD/CKD interaction. *Medscape Cardiology*, 9, 2.
- Valdez, E. 2001. Bivariate analysis of survivorship and persistency. *Insurance: Mathematics and Economics* 29, 357-373.
- Yashin, A., Manton, K., and Stallard, E. 1986. Dependent competing risks: A stochastic process model. *Journal of Mathematical Biology*, 24, 119-140.
- Zhang, D. 2005. Analysis of survival data. *Lectures Notes* (Modified from Dr. A. Tsiatis 's lecture notes.)
- Zheng, M. and Klein 1994. A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics - Theory and Methods*, 23 (8), 2299-2311.
- Zheng, M. and Klein 1995. Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika*, 82 (1), 127-138.

EKLER

EK1 Tamoxifen çalışması sonucunda elde edilmiş verilerin bir kısmı

EK2 Tamoxifen çalışmasındaki değişkenlerin tanımları

EK1 Tamoxifen çalışması sonucunda elde edilmiş verilerin bir kısmı

stnum	tx	pathsize	hist	hrlevel	hgb	nodediss	age	survtime	stat	loctime	lcens	axltime	acens	distime	dcens	maltime	mcens
1	B	1.0	DUC	POS	140	Y	51	8.2683093771	0	8.2683093771	0	8.2683093771	0	8.2683093771	0	8.2683093771	0
2	B	0.5	DUC	POS	138	Y	74	6.173853525	0	6.173853525	0	6.173853525	0	1.3524982888	1	6.173853525	0
3	B	1.1	DUC	POS	157	Y	71	7.1759069131	0	7.1759069131	0	7.1759069131	0	7.1759069131	0	7.1759069131	0
4	B	0.8	DUC	POS	136	Y	52	9.5058179329	0	9.5058179329	0	9.5058179329	0	9.5058179329	0	9.5058179329	0
5	B	1.5	DUC	POS	123	Y	62	9.0951403149	0	9.0951403149	0	9.0951403149	0	9.0951403149	0	9.0951403149	0
6	B	0.7	DUC	POS	122	Y	75	8.0958247775	1	8.0958247775	0	8.0958247775	0	8.0958247775	0	8.0958247775	0
7	B	2.4	DUC	POS	139	Y	77	0.7255304586	1	0.7255304586	0	0.7255304586	0	0.6023271732	1	0.7255304586	0
8	B	2.0	DUC	POS	142	Y	78	4.9637234771	0	4.9637234771	0	4.9637234771	0	4.9637234771	0	4.9637234771	0
9	B	2.0	DUC	POS	121	Y	65	8.848733744	0	8.848733744	0	8.848733744	0	8.848733744	0	8.848733744	0
10	B	1.2	DUC	POS	132	Y	67	5.1635865845	0	5.1635865845	0	5.1635865845	0	5.1635865845	0	5.1635865845	0
513	T	1.0	MIX	POS	133	Y	54	6.2313483915	0	6.2313483915	0	6.2313483915	0	6.2313483915	0	6.2313483915	0
514	T	2.5	DUC	POS	159	N	66	6.0424366872	0	6.0424366872	0	2.9596167009	1	6.0424366872	0	6.0424366872	0
515	T	1.0	MIX	POS	127	Y	52	5.0266940452	0	5.0266940452	0	5.0266940452	0	5.0266940452	0	5.0266940452	0
516	T	1.4	DUC	POS	126	N	65	5.5660506502	0	5.5660506502	0	5.5660506502	0	5.5660506502	0	5.5660506502	0
517	T	2.0	MIX	POS	130	Y	71	6.0616016427	0	6.0616016427	0	6.0616016427	0	6.0616016427	0	6.0616016427	0
518	T	2.0	LOB	POS	147	Y	75	5.4099931554	0	5.4099931554	0	5.4099931554	0	5.4099931554	0	5.4099931554	0
519	T	1.7	DUC	POS	144	Y	56	6.1957563313	0	6.1957563313	0	6.1957563313	0	6.1957563313	0	6.1957563313	0
520	T	1.0	DUC	POS	135	N	71	6.0616016427	0	6.0616016427	0	6.0616016427	0	6.0616016427	0	6.0616016427	0

EK 2 Tamoxifen çalışmasındaki değişkenlerin tanımları

Değişken İsmi	Tanımları
stnum	Hasta numarası
tx	Tedavi türü: T:tamoxifen, B:tomoxifen+radyoterapi
pathsize	Tümör boyutu
hist	Histoloji, DUC:Ductal,MIX:Mixed,OTH:diğer
hrlevel	Hormon düzeyi, NEG:Negatif, POS:Pozitif
hgb	Hemoglobin
nodeiss	Lenf nodu incelemesi yapılp yapılmadığı
age	yaş
survtime	Sağkalım süresi
stat	Durum değişkeni, 1: ölü, 0:sağ
loctime	Bölgesel nüks gerçekleşene kadar geçen süre
lcens	Bölgesel nüks için durum değişkeni, 1: var, 0:yok
axltime	Koltuk altı lenf nodu bölgesinde nüks gerçekleşene kadar geçen süre
acens	Koltuk altı lenf nodu bölgesinde nüks için durum değişkeni, 1: var, 0:yok
distime	Uzak organ yayılımı gerçekleşene kadar geçen süre
dcens	Uzak organ yayılımı için durum değişkeni, 1: var, 0:yok
maltime	Kötü huylu ikinci bir tümör ortaya çıkana kadar geçen süre
mcens	Kötü huylu ikinci bir tümör değişkeni için durum değişkeni, 1: var, 0:yok

ÖZGEÇMİŞ

Adı Soyadı : Çiğdem TOPÇU

Doğum Yeri : Burdur

Doğum Tarihi : 02.01.1983

Yabancı Dili : İngilizce

Eğitim Durumu:

Lise : Antalya Aldemir Atilla Konuk Yabancı Dil Ağırlıklı Lise, 2000.

Lisans : Hacettepe Üniversitesi, İstatistik Bölümü 2004.

Yüksek Lisans: Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik A.B.D
(2004-2007)

Doktora : Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı
(Şubat 2007- Şubat 2013)

Çalıştığı Kurum:

Ankara Üniversitesi Fen Fakültesi: Araştırma Görevlisi, 2005-

Seminerler:

‘Sağdan Sansürlemenin Sağkalım Analizindeki Etkisi’, 14.11.2008, Ankara Üniversitesi
İstatistik Bölümü

‘Kaplan-Meier and Greenwood Yöntemi ile Varyans Tahmini’, 02.06.2006, Ankara
Üniversitesi İstatistik Bölümü

Bildiriler:

Topçu, Ç., Arslan, F. 2006. Sağdan Sansürlenmiş Veriler İçin Kaplan-Meier Yöntemi Yardımıyla Yaşam Fonksiyonlarının Parametrik Olmayan Tahmini, 5. İstatistik Günleri Sempozyumu 24-27 Mayıs, ANTALYA

Topçu Ç., Genç Y., ‘Faktöriyel Tasarımlarda Uzunlamasına Veriler İçin Parametrik Olmayan Analiz’, 9. Bioistatistik Kongresi, 27 -30 Mayıs 2008, Malatya,Türkiye

Topçu Ç., Arslan F., Sağdan Sansürlemenin Sağkalım Analizindeki Etkisi ve Simulasyon Çalışması’, 6.İstatisitk Kongresi 29 Nisan-03 Mayıs 2009,Antalya,Türkiye

Yayınlar:

Topçu Ç., Arslan F. ‘Copula Approach for modelling dependence structure.’ NWSA Uluslararası Hakemli e-Dergi. Ekim 2012.

Topçu Ç., Arslan F. ‘Selecting the Suitable Copula Function When Only Values of Distribution Functions are Avaliable.’ Selçuk Journal of Applied Mathematics (Accepted/In press)