

Burr Tipi III Dağılımının İlerleyen Tür Tip II Sağdan Sansürlü Örneklemeye Dayalı Parametrelerinin EM Tahmin Edicisi

Özlem GÜRÜNLÜ ALMA¹, Sinem ARAY²

¹Muğla Sıtkı Koçman Üniv., Fen Fakültesi, İstatistik Bölümü, Muğla, Türkiye
² Muğla Sıtkı Koçman Üniv., Fen Bilimleri Enstitüsü, İstatistik ABD, Muğla, Türkiye

Geliş / Received: 27/09/2019, Kabul / Accepted: 01/06/2020

Öz

Bu çalışmada, Burr Tipi III Dağılım'a ait bilinmeyen parametrelerinin istatistiksel tahmini, ilerleyen tür tip II sağdan sansürlü örneklem durumunda incelenmiştir. Bilinmeyen parametrelerin en çok olabirlik tahmin edicileri Newton-Raphson yöntemi ve Beklenti Maksimizasyonu (EM) algoritması kullanılarak elde edilmiş, parametrelerin asimptotik güven aralıkları kayıp değer prensibine dayalı Fisher bilgi matrisi aracılığıyla bulunmuştur. Tahmin edicilerin performansları farklı sansür şemaları ve parametre değerleri için benzetim çalışması yoluyla karşılaştırılmıştır. Ayrıca çalışmanın daha açıklayıcı olması amacıyla gerçek bir yaşam verisi örneği verilmiştir.

Anahtar Kelimeler: İlerleyen Tür Tip II Sağdan Sansürlü Örneklem, En Çok Olabirlik Tahmin Edicileri (MLE), Olabirlik Fonksiyonu, Burr Tipi III Dağılım, Sansürlü Veri, Beklenti Maksimizasyonu (EM) Algoritması, Newton-Raphson yöntemi.

Parameter Estimation of Burr Type III Distribution Under Progressive Type II Censored Sample Using EM Algorithm

Abstract

In this study, the statistical estimation of the unknown parameters of the Burr Type III Distribution is examined in the case of the progressive type II right censored sample. The maximum likelihood estimators of the unknown parameters were obtained using the Newton Raphson and the Expectation Maximization (EM) algorithm and the asymptotic confidence intervals of the parameters were found through the Fisher information matrix based on the missing value principle. The performances of the estimators are compared by means of a simulation study for different censoring schemes and parameter values. Also, a real-life example is given in order to make the study more descriptive.

Keywords: Progressively Type II Censored Samples, Maximum Likelihood Estimator, Likelihood Function, Burr Type III Distribution, Censored Data, Expectation Maximization (EM) Algorithm

1. Giriş

Yaşam analizi, ilgilenilen herhangi bir olayın ortaya çıkmasına kadar geçen süreyi modelleyen, sağlık alanında daha çok başvurulan, mühendislik alanında güvenilirlik veya güvenilirlik analizi, iktisat ve sosyoloji alanlarında ise süre analizi veya süre modellemesi olarak adlandırılan istatistiksel bir analiz yöntemidir. Burada

olay: bozulma, ölüm ve başarısızlık olarak ifade edilebilir. Bu durumları inceleyebilmek, zaman kaybı ve maliyet gibi nedenlerden dolayı her zaman mümkün olmayabilir. Örneğin; hane halkı araştırmasında aile, gelirlerinin tamamını beyan etmekten kaçınabilir (Yazıcı, 2005). Verinin tamamına erişemediğimiz bu gibi durumlarda araştırmacı ya kayıp verileri göz ardı edecek ya da farklı teknikler

uygulayarak bu sansürlü verileri tamamlama yoluna gitmeyi tercih edecektir (Yazıcı, 2005).

Yaşam ve güvenilirlik analizlerinde yaşam süresi tam olarak bilinmeyen olaylar sansürlü gözlem olarak adlandırılır. Literatürde çeşitli sansürleme türleri mevcuttur, bunlar; sağdan sansürleme, soldan sansürleme, ikili sansürleme ve aralık sansürlemesidir. Sağdan sansürleme, başarısızlık olarak adlandırılan olay, çalışma için belirlenen bir durma zamanına kadar gerçekleşmezse, bireyin yaşam süresinin uzunluğu çalışmanın durma zamanının sağ tarafına geçer. Böyle bir durumda bireyin yaşam süresi kesin olarak bilinmeyecek ve gözlemlenemeyecektir. Soldan sansürlemede, gözlenmek istenen olay bazı bireyler için, verilen zamana gelmeden önce gerçekleşmiştir. Bazı çalışmalarda soldan sansürlemenin meydana geldiği durumlarda, sağdan sansürleme de aynı zamanda ortaya çıkabilir, bu gibi durumlar ikili sansürlemedir. Aralık sansürlemesi ise, çalışmaya konu olan olayın meydana gelme süresi, bir aralıkta ifade edilir. Beklenen olay iki zaman noktası arasında gerçekleşmiştir ve gerçekleşme anı hakkında bilgi yoktur bu da aralık sansürlemesi olarak adlandırılmaktadır.

Sağdan sansürleme, kendi içinde bazı alt gruplara ayrılmaktadır, bunlar; Tip I sansürleme, Tip II sansürleme, İlerleyen tür tip II sansürleme ve Bağımsız rasgele sansürleme' dir (Kale, 2003). Tip I sansürleme olarak adlandırılan sansürleme modeli, t gibi önceden belirlenmiş bir zamandan önce, sistemdeki bozulan birimlerin bozulma zamanının gözlenmesi durumudur. Tip II sansürleme olarak adlandırılan sansürleme modeli, n birimden oluşan bir sistemin bozulan $m \leq n$ biriminin bozulma zamanının gözlenmesi durumudur (Kale, 2003).

İkinci tip sansürlemenin en çok kullanılan türü, ilerleyen tür tip-II sağdan sansürlemedir. Bu sansürleme modelinde, n sayıda özdeş bileşen yaşam analizine tabi

tutulur. Sistemde meydana gelen 1. bozulma ile rasgele R_1 sayıda bileşenin sistemden çekildiğini, daha sonra geriye kalan $n - R_1 - 1$ bileşenden, 2. Bozulma gerçekleşmesi ile rasgele R_2 sayıda bileşenin sistemden çekildiğini ve böylece m. bozulmanın gerçekleşmesinden sonra rasgele R_m sayıda bileşenin sistemden çekilmesiyle m bileşenin bozulma zamanı gözlenir. Bu şekilde elde edilen m hacimli örneklem, ilerleyen tür tip-II sağdan sansürlü örneklemdir. Burada $\sum_{j=1}^m (R_j + m) = n$ 'dir. Bu durumda, $(R_1, R_2, R_3, \dots, R_m)$ ve sansürleme şeması $(X_1, X_2, X_3, \dots, X_m)$ 'e ise gözlenen veriler denilir (Balakrishnan ve Aggarwala 2000).

Yaşam analizinde, elde edilen veriye hangi dağılımın uygulanacağı oldukça önemlidir. Bu nedenle, model seçimi yapılırken ilk olarak modelin veriye uygunluğu için gerekli testin yapılması gerekir. Model veriye uygunsa parametrik test varsayımlarının sağlanıp sağlanmadığına bakılarak modelin geçerliliği ile ilgili gerekli analizler yapılmaktadır. Yaşam analizlerinde, yaşam süreleri verilerini modelleyebilmek için kullanılan dağılımlardan Burr Dağılımı, on iki birikimli dağılım fonksiyonunu içeren bir dağılım ailesinden oluşmaktadır. Bu aileden, Burr tipi III dağılımı, Log-lojistik, Weibull ve Burr XII gibi dağılımlar içinden çarpıklık ve basıklık açısından daha geniş alana izin verir. Bu çalışmada, Burr Tipi III Dağılımının parametrelerinin tahmin edilmesinde kapalı bir form olmaması nedeniyle, parametrelerin tahmin edilmesinde; Newton Raphson yöntemi ve kayıp gözleme dayalı Beklenti Maksimizasyonu (EM) Algoritması kullanılmıştır.

Günümüze kadar ilerleyen tür tip II sağdan sansürlü örneklemeye dayalı ve EM algoritması ile ilgili birçok çalışma gerçekleştirilmiştir. Bu çalışmalardan bazıları aşağıdaki gibi özetlenebilir:

Balakrishnan ve Sandu (1995), ilerleyen tür tip-II sansürlü örnekleme oluşturma yöntemini vermiştir

Balakrishnan ve Aggarwala (2000), çalışmalarında ilerleyen tür sansür teorisi ve uygulamalarını geniş bir şekilde anlatmışlardır.

Dempster, Laird ve Rubin (1977), EM algoritmasının genel formülasyonunu sunmuşlardır.

Zhu, Lee, Wei ve Zhou (2001), mevcut yöntemlerin çözmekte zorlandığı problemlere uygulanabilecek bir yöntem sunmuşlar, EM algoritmasındaki tam olmayan en çok log-olabilirlik fonksiyonunun koşullu beklenen yaklaşımını genelleştirmişlerdir.

Terzi (2003), doktora tezinde, sağdan sansürlü veriler için sağkalım analizi yöntemlerini kullanmış ve gerçek veriler üzerinde uygulamasını gerçekleştirmiştir. Sağkalım fonksiyonları arasındaki farkları parametrik olmayan testlerle bulmuş, iki sağkalım fonksiyonunu karşılaştıran testleri birbirleriyle karşılaştırmıştır. Sağkalım sürelerine etkileri araştırılan prognostik faktörler, yarı parametrik bir yöntem olan Cox regresyon yöntemi ile incelenmiştir.

Kale (2003), tarafından hazırlanan tez çalışmasında ilerleyen tür tip-II sansürlenmiş sıra istatistikleri ve dağılım özellikleri konusu incelenmiştir.

Kuş (2004), tez çalışmasında, tam ve sansürlü verilere dayalı Burr XII, Gompertz ve Pareto I dağılımlarının parametrelerinin güven aralıkları ve güven bölgeleri, ilerleyen tür tip-II sağdan sansürlü örneklemlere dayalı olarak elde edilmiştir.

Kuş (2005), yapmış olduğu çalışmasında, düzgün dağılımın parametrelerinin ilerleyen tür tip-II sağdan sansürlü örnekleme dayalı Jacknife tahmin edicilerini elde etmiştir.

Yazıcı (2005), tez çalışmasında, kayıp veri içeren problemler için parametre tahmini

hesaplamalarında, son yıllarda yoğun olarak kullanılan EM algoritması ve uzantılarının tanıtılması amaçlanmıştır. EM algoritmasının sözel ve kuramsal olarak genel işleyişi ve özellikleri verilmiştir. Normal dağılıma sahip iki örneklem üzerinde, farklı kayıp veri problemleri tasarlanarak benzetim çalışması yapılmıştır.

Bairamov ve Eryılmaz (2006), Bu çalışmada çok değişkenli gözlemler için ilerleyen tür tip-II sansürlü sıra istatistikleri konusu ele alınmıştır.

Saraçoğlu, Kınacı ve Kuş (2006), Burr XII Dağılımının Parametresinin İlerleyen Tür Tip II Sağdan Sansürlü Örneklemlere Dayalı En Küçük Kareler Tahmini üzerinde çalışılmıştır.

İnceoğlu (2013), çalışmasında Turgut Özal Tıp Merkezi Karaciğer Nakil Enstitüsünde nakil yapılan hastalardan alınan veriler yardımıyla sağkalım analizi yöntemlerinin (Yaşam Tablosu Analizi, Kaplan–Meier Analizi, Cox Regresyon Analizi) sonuçlarını birbirleriyle karşılaştırmak amacı ile gerçekleştirmiştir.

Demir (2015), tezinde, ilerleyen tür tip II sansürleme altında, bazı yaşam zamanı dağılımlarının parametreleri için en çok olabilirlik ve lindley yaklaşımı kullanılarak yaklaşık bayes tahmin edicileri elde etmiş ve bu tahmin edicilerin yanları ve hata kareler ortalamaları Monte Carlo Benzetim metodu kullanılarak karşılaştırılmıştır.

Arabi Belaghi, Valizadeh Gamchi, Bevrani ve Gürünlü Alma (2016), İlerleyen Tür Sansür Kullanılarak EM ve SEM Algoritmalarına Dayalı Burr Type III Tahmini isimli makalelerinde, ilerleyen tür sansürleme kullanılarak iki parametrelili Burr tip III dağılımının parametreleri kapalı formda elde edilemeyeceğinden EM ve SEM algoritma tahmincileri elde edilmiştir. Gerçek veri kullanılarak yapılan benzetim çalışmasıyla beyaz kan hücreleri kanserli hastaların analizi yapılarak SEM, EM, MLE sonuçları karşılaştırılmıştır.

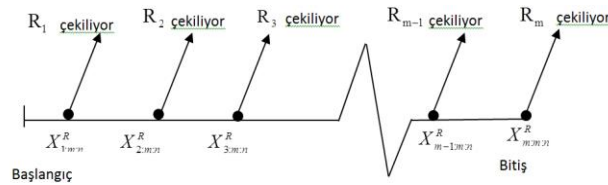
Aray (2018) yapmış olduğu tez çalışmasında, ilerleyen tür tip II sansürleme altında, Burr tipi III dağılımının parametreleri için Newton Raphson ve EM yaklaşımı kullanılarak tahmin edicileri elde etmiş ve bu tahmin edicilerin güven aralıkları ve hata kareler ortalamalarını Monte Carlo Benzetim metodu kullanılarak karşılaştırmıştır.

Çalışmanın ilerleyen bölümleri şu şekildedir: ikinci bölümde, Burr tipi III dağılımının ilerleyen tür Tip II sansürlü örneklem için parametre tahminleri anlatılmıştır. Üçüncü bölümde, benzetim çalışması ve sonuçlarının değerlendirilmesiyle birlikte uygulama bölümünde, sansürlü veri içeren karbon fiberlere ilişkin gerçek veriler için en uygun dağılım belirlenip, bu dağılıma ait parametre tahminleri bulunmuştur. Dördüncü bölümde ise çalışmanın sonuçları değerlendirilmiştir.

2. Materyal ve Metot

Yaşam analizi sadece sağlık alanında kullanılan, ölüm veya hayatta kalma süreleri ile ilgilenen hastalık ve yaşamı etkileyen faktörleri inceleyen bir yöntem değildir. Bozulma, çürüme, ölme, başarısızlık riski taşıyan ve kullanıma bağlı olarak dayanma süreleri farklılaşan materyallerin bozulma zamanlarının analizlerinde de yararlanılmaktadır. Bu nedenle sadece tıp değil, ekonomi, işletme, mühendislik, sosyoloji, zooloji, botanik veya fabrikanın

devamlılığı, bir makinenin bozulma zamanına kadar geçen sürenin analizleri gibi birçok farklı alanda kullanılabilen bir analiz yöntemidir. Yaşam analizinde kullanılan *başarısızlık* terimi, incelenen konunun denekte görülmesi durumudur. Canlılar için genelde ölüm veya hastalık, mekanik aletler için ise bozulma anlamına gelmektedir (Nelson, 1982). Yaşam analizlerinde sağkalım süresi tam olarak bilinmeyen vakalar sansürlü gözlem olarak adlandırılmaktadır. İlerleyen tür tip II sansürleme, sansürlenmiş veri çalışmalarında en popüler sansürleme şemalarından biridir. Bu sansürleme modelinde, n sayıda özdeş bileşenin sağkalım analizine tabi tutulduğu düşünülün. Sistemde meydana gelen 1. bozulma ile rasgele R_1 sayıda bileşenin sistemden çekildiğini, daha sonra geriye kalan $n - R_1 - 1$ bileşenden, 2. bozulma ile rasgele R_2 sayıda bileşenin sistemden çekildiğini ve böylece m . bozulma ile rasgele R_m sayıda bileşenin sistemden çekilmesiyle m bileşenin bozulma zamanı gözlenir. Bu şekilde elde edilen m hacimli örneklem ilerleyen tür tip-II sağdan sansürlü örneklemdir. Burada $n = m + \sum_{i=1}^m R_i$ 'dir. Bu durumda, $(R_1, R_2, R_3, \dots, R_m)$ sansürleme şeması ve $(X_1, X_2, X_3, \dots, X_m)$ 'e gözlenen veriler denilir (Balakrishnan ve Aggarwala 2000). Bu sansürleme şeması Şekil 1'de gösterilmiştir:



Şekil 1. İlerleyen tür tip-II sağdan sansürleme şeması

$X_{1:m:n}^R < X_{2:m:n}^R < \dots < X_{m:m:n}^R$, olasılık yoğunluk fonksiyonu $f(x)$ ve dağılım fonksiyonu $F(x)$ olan dağılımdan alınan ilerleyen tür tip-II

sağdan sansürlü örneklem olmak üzere $X_{1:m:n}^R < X_{2:m:n}^R < \dots < X_{m:m:n}^R$ 'nin ortak olasılık yoğunluk fonksiyonu;

$$f_{X_{1:m:n}^R, \dots, X_{2:m:n}^R, \dots, X_{m:m:n}^R}(x_1, x_2, \dots, x_m) = a \prod_{i=1}^m f(x_i) [1 - F(x_i)]^{R_i}, \quad -\infty < x_1 < x_2 < \dots < x_m < \infty \quad (1)$$

şeklinde ifade edilir. Burada a sabiti,

$$a = n(n - R_1 - 1)x \dots x(n - R_1 - R_2 - \dots - R_{m-1} - m + 1) \text{ 'dir.}$$

İlerleyen tür tip-II sağdan sansürlü örnekleme, yaşam zamanı analizlerinde veri elde etmede önemli bir yöntemdir, deneyin maliyet ve deney süresinin azaltılmasını sağlar (Balakrishnan ve Aggarwala, 2000).

Burr dağılım ailesi, Burr (1942) tarafından yaşam süreleri verilerini modelleyebilmek için kullanılan on iki birikimli dağılım fonksiyonunu içeren bir dağılım ailesi olarak önerilmiştir. Bu ailenin en önemli iki üyesi: Burr tipi III dağılımı ve Burr tipi XII dağılımıdır. Bu iki önemli dağılım, basit dönüşüm ile birbirleriyle ilişkilidir. Weibull,

üstel lojistik, genelleştirilmiş lojistik, Gompertz, normal, aşırı değer (extreme value) dağılımı ve tekdüze dağılımlar da dâhil olmak üzere birçok standart teorik dağılımlar, Burr ailesinin özel durumları veya sınırlayıcı durumlarıdır. Burr tipi III dağılımı, Log-lojistik, Weibull, Burr tipi XII gibi kapalı formdaki dağılımlardan daha çok esnek olup, çarpıklık ve basıklık açısından daha geniş alana izin verir. Burr tipi III dağılımının α ve β parametrelerinin dağılım fonksiyonu ile olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$F(x | \alpha, \beta) = (1 + x^{-\beta})^{-\alpha}, \quad x > 0, \alpha > 0, \beta > 0 \quad (2)$$

$$f(x | \alpha, \beta) = \beta \alpha x^{-\beta-1} (1 + x^{-\beta})^{-\alpha-1}, \quad x > 0, \alpha > 0, \beta > 0. \quad (3)$$

Genel olarak, Burr tipi III dağılımına benzer şekilde parametre kestirimi yapan birçok dağılım mevcuttur. Bunlar; Gamma, Log-normal, Weibull ve Exponential dağılımlardır. Bu dağılımlar güvenilirlik ve kalite kontrolünde önemli bir role sahiptir. Tadikmalla (1980), Burr dağılımının farklı türleri üzerinde çalışmış ve aynı zamanda bu dağılımlar arasındaki ilişkileri incelemiştir. Son yıllarda, Rastogi ve Tripathi (2013),

Gupta ve Kundu (2001) gibi birçok bilim adamı Burr dağılımı ve üstel dağılım ailesi üzerinde çalışarak geliştirmişlerdir. $X_{1:m:n}, X_{2:m:n}, \dots, X_{m:m:n}$ gözlem değerleri (2) ve (3) eşitliğinde verilen dağılımdan çekilmiş ilerleyen tür tip II sansürlü gözlem değerlerini ve R_1, R_2, \dots, R_m de ilerleyen tür tip II sansürleme şemasını gösterebilir. Bu durumda α ve β parametrelerinin olabilirlik fonksiyonu aşağıdaki gibi elde edilir.

$$l(X; \alpha, \beta) = c \prod_{i=1}^m f(x_{i:m:n}; \alpha, \beta) [1 - F(x_{i:m:n}; \alpha, \beta)]^{R_i} \quad (4)$$

Burada, $c = n(n-1)(n-R_1-1)(n-R_1-R_2-2) \dots (n-R_1-R_2-R_3-\dots-R_{m-1}-m+1)$. Burada,

$$R_i > 0 \text{ ve } \sum_{i=1}^m R_i + m = n \text{ 'dir.}$$

En çok olabilirlik tahmincisini elde etmek için log-olabilirlik fonksiyonu ise,

$$\text{Log}(L(X; \alpha, \beta)) = \ell(X; \alpha, \beta)$$

$$\begin{aligned} \ell(X; \alpha, \beta) &= m \log \alpha + m \log \beta + (-\beta - 1) \sum_{i=1}^m \log(x_{i:m:n}) \\ &+ (-\alpha - 1) \sum_{i=1}^m \log(1 + x_{i:m:n}^{-\beta}) + R_i \sum_{i=1}^m \log \left[1 - (1 + x_{i:m:n}^{-\beta})^{-\alpha} \right] \end{aligned} \quad (5)$$

olarak elde edilir. Elde edilen log-olabilirlik türevleri ise aşağıdaki gibidir: fonksiyonunun α ve β parametrelerine göre

$$\frac{\partial \ell(x; \alpha, \beta)}{\partial \alpha} = \frac{m}{\alpha} - \sum_{i=1}^m \log(1 + x_{i:m:n}^{-\beta}) \quad (6)$$

$$-R_i \sum_{i=1}^m \frac{\log(1 - (1 + x_{i:m:n}^{-\beta})) (1 - (1 + x_{i:m:n}^{-\beta})^{-\alpha})}{(1 - (1 + x_{i:m:n}^{-\beta})^{-\alpha})}$$

$$\begin{aligned} \frac{\partial \ell(x; \alpha, \beta)}{\partial \beta} &= \frac{m}{\beta} - \sum_{i=1}^m \log(x_{i:m:n}) - (\alpha + 1) \left[\sum_{i=1}^m \frac{-x_{i:m:n}^{-\beta} \ln x_{i:m:n}}{1 + x_{i:m:n}^{-\beta}} \right] \\ &+ \alpha R_i \sum_{i=1}^m \frac{(1 + x_{i:m:n}^{-\beta})^{-(\alpha+1)} (-x_{i:m:n}^{-\beta} \ln x_{i:m:n})}{(1 - (1 + x_{i:m:n}^{-\beta})^{-\alpha})} \end{aligned} \quad (7)$$

Burada, $\frac{\partial \ell(x; \alpha, \beta)}{\partial \alpha} = 0$ ve $\frac{\partial \ell(x; \alpha, \beta)}{\partial \beta} = 0$

eşitliklerinin analitik olarak çözümlerini elde etmek mümkün olmayacaktır. Bu nedenle MLE tahmincilerini elde edebilmek için uygun numerik çözümlere ihtiyaç vardır. Newton-Raphson ve EM algoritması bu gibi durumlarda kullanılabilir alternatif yöntemlerdir.

2.1 Newton-Raphson Yöntemi

Doğrusal olmayan denklemleri çözmek için kullanılan yinelemeli bir yöntem olan Newton-Raphson yöntemi, fonksiyonun maksimum olduğu noktayı belirler. İlerleyen tür tip II sansürleme planı altında Burr tipi III dağılımının en çok olabilirlik tahmincisinin elde edilmesi için ilgili log-olabilirlik ve α ve β parametrelerine göre türevleri ise aşağıdaki gibidir:

$$\begin{aligned} \ln[L(\theta)] &= m \ln \alpha + m \ln \beta - (\beta + 1) \sum_{j=1}^m \ln(x_{j:m:n}) \\ &- (\alpha + 1) \sum_{j=1}^m \ln(1 + x_{j:m:n}^{-\beta}) - \alpha \sum_{j=1}^m R_j (\ln(1 + x_{j:m:n}^{-\beta})) \end{aligned} \quad (8)$$

Buradan,

$$\frac{\partial^2 \ln[L(\alpha, \beta)]}{\partial \alpha^2} = m \left(\frac{-1}{\alpha^2} \right), \quad (9)$$

$$\frac{\partial^2 \ln[L(\alpha, \beta)]}{\partial \beta^2} = m \left(\frac{-1}{\beta^2} \right) - (\alpha + 1) \sum_{j=1}^m \frac{x_{j:m:n}^{-\beta} \ln x_{j:m:n} \ln(x_{j:m:n}^{-\beta})}{(1 + x_{j:m:n}^{-\beta})} \quad (10)$$

$$\begin{aligned}
 & -(\alpha + 1) \sum_{j=1}^m \frac{x^{-\beta} \ln x_{j:m:n}}{1 + x^{-\beta}} + (\alpha + 1) \sum_{j=1}^m \frac{x^{-\beta^2} \ln(x^{-\beta}) \ln(x)}{(1 + x^{-\beta})^2} \\
 & -\alpha \sum_{j=1}^m R_j \frac{x^{-\beta} \ln x_{j:m:n} \ln(x_{j:m:n}^{-\beta})}{1 + x_{j:m:n}^{-\beta}} - \alpha \sum_{j=1}^m R_j \frac{x_{j:m:n}^{-\beta} \ln x_{j:m:n}}{1 + x_{j:m:n}^{-\beta}} \\
 & + \alpha \sum_{j=1}^m R_j \frac{x_{j:m:n}^{-\beta^2} \ln(x_{j:m:n}^{-\beta}) \ln(x_{j:m:n})}{(1 + x_{j:m:n}^{-\beta})^2},
 \end{aligned}$$

ve

$$\begin{aligned}
 \frac{\partial^2 \ln [L(\alpha, \beta)]}{\partial \alpha \partial \beta} &= - \left(\sum_{j=1}^m \left(- \frac{x_{j:m:n}^{-\beta} \ln(x_{j:m:n})}{1 + x_{j:m:n}^{-\beta}} \right) \right) \\
 & - R_j \left[\sum_{j=1}^m \left(- \frac{x_{j:m:n}^{-\beta} \ln(x_{j:m:n})}{1 + x_{j:m:n}^{-\beta}} \right) \right]
 \end{aligned} \tag{11}$$

olacaktır.

2.2 EM Algoritması

EM algoritması, sansürlü veri problemlerinde parametre tahminini hesaplamak için en çok olabilirlik tahminlerini içeren yinelemeli bir yöntemdir. EM algoritması, gözlenen ve sansürlü verilere göre oluşturulan olabilirlik fonksiyonundan yola çıkmaktadır (Dempster, Laird ve Rubin, 1977). EM türündeki algoritmalarla ilk çalışmalar 1886 yılına dayanmaktadır. Ardından, McKendrick (1926) tarafından, algoritmalar medikal alanda uygulanmaya başlanmıştır. Fesler ve Hero (1994), iki genel EM (General EM, GEM) tanımlamışlardır. EM algoritması tanındıktan sonra çok fazla alanda uygulanmıştır. EM algoritması iki aşamalı olup, E: beklenti adımı ve M: maksimizasyon adımı içerir. **E-Adımı:** Sansürlü veya gözlemlenemeyen veriler için beklenen değerleri yardımıyla bir bilgi elde ederken, **M-Adımı:** Gözlemlenen veri ve beklenen değer bilgileri altında olabilirlik fonksiyonunu maksimize eder. İlerleyen tür tip II sansürlü örneklemlerde kayıp veri ya da

eksik veri sorunlarını yaşayan örneklemler olduğundan, burada bilinmeyen α ve β parametrelerinin MLE' sini elde edebilmek için EM yönteminden yararlanılacaktır.

E adımı:

$X_{(i)}$, i. zamandaki başarısızlık veya gözlemlenen veri, R_i ise, i. başarısızlıktan sonra deneyden çıkarılan gözlemlenemeyen sansürlü verilerin sayısını gösterebilir. Böylelikle \mathbf{X} gözlem vektörü, $\mathbf{X} = (X_{1:m:n}, X_{2:m:n}, \dots, X_{m:m:n})$ ve \mathbf{Z} sansürlü gözlemler vektörü, $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, \dots, Z_{jm})$, $j = 1, 2, \dots, m$ olarak ifade edilir. Burada Z_j , $1 \times R_j$ ' lik bir vektörü ifade etmektedir. \mathbf{X} gözlemlenen, \mathbf{Z} ' de sansürlü verileri göstermek üzere tüm verileri $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$ biçiminde belirtebiliriz. Tüm veriler \mathbf{W} için log-olabilirlik fonksiyonu aşağıdaki gibi ifade edilir.

$$\begin{aligned}
 \lambda(\mathbf{W}; \alpha, \beta) &= n \ln \alpha + n \ln \beta - (\beta + 1) \sum_{j=1}^m \ln x_{j:m:n} - (\alpha + 1) \sum_{j=1}^m \ln(1 + x_{j:m:n}^{-\beta}) \\
 & - (\beta + 1) \sum_{j=1}^m \sum_{k=1}^{R_j} \ln z_{jk} - (\alpha + 1) \sum_{j=1}^m \sum_{k=1}^{R_j} \ln(1 + z_{jk}^{-\beta})
 \end{aligned} \tag{12}$$

$E(Z_{jk} | Z_{jk} > x_{(j)})$ beklenen değerini bulabilmek için, Teorem 1'den yararlanılır.

Teorem 1: Verilen $X_{(1)} = x_1, \dots, X_{(j)} = x_{(j)}$ $k = 1, \dots, R_j$ için aşağıdaki gibidir. değerleri için Z_{jk} 'nın koşullu olasılığı her

$$\begin{aligned} f_{z_{1x}}(z_j | X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}, \dots, X_{(j)} = x_{(j)}) \\ = f_{z_{1x}}(Z_j | X_{(j)} = x_{(j)}) \\ = \frac{f(Z_j | \alpha, \beta)}{1 - F(X_{(j)} | \alpha, \beta)} \end{aligned}$$

$Z_{(j)} > x_{(j)}$ olduğu durumlarda. Diğer durumlarda ise, eşitliğin değeri 0' dır. Teoremin ispatı ile ilgili olarak detaylı bilgi için Ng ve diğerlerinin 2002 yılında yapmış

olduğu çalışmadan yararlanılabilir. Teorem 1' i kullanarak $E[(\ln Z_{jk} | Z_{jk} > x_j)]$ 'i aşağıdaki şekilde yazabiliriz.

$$\begin{aligned} E[\ln Z_{jk} | Z_{jk} > x_j] &= \frac{\alpha\beta}{1 - F_x(x_j; \alpha, \beta)} \int_{x_j}^{\infty} \ln uu^{-(\beta+1)}(1+u^{-\beta})^{-(\alpha+1)} du \\ &= \frac{\alpha}{\beta[1 - F_x(x_j; \alpha, \beta)]} \int_1^{1+x_j^{-\beta}} v^{-(\alpha+1)} \ln(v-1) dv \\ &= A(x_j; \alpha_{(j)}, \beta_{(j)}) \end{aligned}$$

ve,

$$\begin{aligned} E(\ln(1 + Z_{jk}^{-\beta}) | Z_{jk} > x_j) &= \frac{\alpha\beta}{1 - F_x(x_j; \alpha, \beta)} \int_{x_j}^{\infty} \ln(1 + u^{-\beta}) u^{-(\beta+1)} (1 + u^{-\beta})^{-(\alpha+1)} du \\ &= \frac{\alpha}{1 - F_x(x_j; \alpha, \beta)} \int_1^{1+x_j^{-\beta}} v^{-(\alpha+1)} \ln(v) dv \\ &= B(x_j; \alpha_{(j)}, \beta_{(j)}) \end{aligned}$$

M adımı:

Bu adımda, $E[\log(L(x; \alpha, \beta))]$ beklenen değeri bilinmeyen parametreler α ve β ' ya

göre türevi alınarak maksimize edilir. α ve β parametrelerinin tahmincileri aşağıdaki gibi elde edilir:

$$\hat{\alpha} = \frac{n}{\sum_{j=1}^m \log(n + x_j^{-\beta}) + \sum_{j=1}^m R_j B(x_j; \alpha, \beta)} \tag{13}$$

$$\hat{\beta} = \frac{n}{\sum_{j=1}^m \log x_j - (\alpha + 1) \sum_{j=1}^m \frac{x_j^{-\beta}}{(1 + x_j^{-\beta})} + \sum_{j=1}^m R_j A(x_j; \alpha, \beta)} \tag{14}$$

Burada,

$$A(x_j; \alpha, \beta) = E[\log Z_{jk} | Z_{jk} > x_j]$$

$$= \frac{\alpha\beta}{1 - (1 + x_j^{-\beta})^{-\alpha}} \int_{x_j}^{\infty} x^{-\beta-1} (1 + x^{-\beta})^{-(\alpha+1)} \log x dx$$

ve

$$B(x_j; \alpha, \beta) = E[\log(1 + Z_{jk}^{-\beta}) | Z_{jk} > x_j] \\ = \frac{\alpha\beta}{1 - (1 + x_j^{-\beta})^{-\alpha}} \int_{x_j}^{\infty} x^{-(\beta+1)} (1 + x^{-\beta})^{-(\alpha+1)} \log(1 + x^{-\beta}) dx \text{ 'dir.}$$

Yukarıda verilen eşitlikler fixed point yöntemiyle elde edilir. EM algoritmasının uygulama adımları ise aşağıda verilmiştir.

Adım 1: α_0 ve β_0 başlangıç değerlerini seç.

Adım 2: Örneklemi kullanarak α_i ve β_i 'nin beklenen değerlerini hesapla.

Adım 3: Beklenen değerleri α ve β 'ya göre maksimize et.

Adım 4: Adım 3'te ki değerleri α_{i+1} ve β_{i+1} 'e ata.

Adım 5: Algoritmayı durdurma kriteri değerini elde edinceye kadar tekrar et.

2.3 Burr Tipi III Dağılım İlerleyen Tür Tip II Sağdan Sansürlü Örnekleme İçin Fisher Bilgi Matrisi

Bölüm 2.1 ve Bölüm 2.2'de bilinmeyen parametrelerin en çok olabilirlik tahmin edicileri Newton-Raphson ve EM algoritması kullanılarak elde edilmişti. Bu bölümde, 1982 yılında Louis'in kayıp bilgi yönteminden yararlanılarak gözlenen Fisher bilgi matrisi elde edilecek ve parametrelerin asimptotik güven aralıkları kayıp değer prensibine dayalı Fisher bilgi matrisi aracılığıyla bulunacaktır. (Louis, 1982) aşağıda verilen ifadeye dayanmaktadır.

$$I_X(\theta) = I_W(\theta) - I_{W \setminus X}(\theta) \quad (15)$$

X: gözlemlenen veriyi, **W:** tüm verileri, $\theta = (\alpha, \beta)$: parametre vektörünü, $I_X(\theta)$: gözlemlenen bilgiyi, $I_W(\theta)$: bütün verilerden

elde edilen bilgiyi ve $I_{W \setminus X}(\theta)$ ise, kayıp/gözlenemeyen bilgiyi gösterebilir. Burada, $I_W(\theta)$ aşağıda verilen eşitlikle elde edilir.

$$I_W(\theta) = -E \left[\frac{\partial^2 L(W; \theta)}{\partial \theta^2} \right] \quad (16)$$

j. başarısızlıkta tek bir sansürlü gözlem için $I_{W \setminus X}^{(j)}(\theta)$ Fisher bilgisi aşağıda verilmiştir.

$$I_{W \setminus X}^{(j)}(\theta) = -E_{Z_j \setminus x_{(j)}} \left[\frac{\partial^2 \ln f_{Z_j}(Z_j | x_{(j)}, \theta)}{\partial \theta^2} \right] \quad (17)$$

Tüm veriler içerisindeki kayıp/gözlenemeyen veriler için bilgi aşağıdaki eşitlikle elde edilir.

$$I_{W \setminus X}(\theta) = \sum_{j=1}^m R_j I_{W \setminus X}^{(j)}(\theta) \quad (18)$$

Bütün verilerden elde edilen $I_w(\theta)$ bilgi matrisinin elemanları aşağıdaki biçimde gösterilebilir.

$$I_w(\theta) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Burada,

$$a_{11} = \frac{n}{\alpha^2}, \quad (19)$$

$$a_{12} = a_{21} = \frac{n\alpha}{\beta} \int_1^\infty (u-1) \ln(u-1) u^{-(\alpha+2)} du \quad (20)$$

ve

$$a_{22} = \frac{n}{\beta^2} + \frac{n\alpha(\alpha+1)}{\beta^2} \int_1^\infty (u-1) \{\ln(u-1)\}^2 u^{-(\alpha+3)} du, \text{ olacaktır.} \quad (21)$$

(17) eşitliğinde verilen tüm beklenen bilgi matrisi $I_{w|x}^{(j)}(\theta)$, aşağıda belirtildiği gibi değerlerin hesaplanmasının ardından kayıp elde edilir.

$$I_{w|x}^{(j)}(\theta) = \begin{bmatrix} b_{11}(x_{(j)}; \alpha, \beta) & b_{12}(x_{(j)}; \alpha, \beta) \\ b_{21}(x_{(j)}; \alpha, \beta) & b_{22}(x_{(j)}; \alpha, \beta) \end{bmatrix}$$

Burada,

$$b_{11}(x_{(j)}; \alpha, \beta) = \frac{1}{\alpha^2} - \frac{(1+x^{-\beta})^{-\alpha} [\ln(1+x^{-\beta})]^2}{[1-(1+x^{-\beta})^{-\alpha}]^2} \quad (22)$$

$$b_{12}(x_{(j)}; \alpha, \beta) = b_{21}(x_{(j)}; \alpha, \beta) \quad (23)$$

$$= \frac{\alpha}{\beta [1-F(x_j; \alpha, \beta)]} \int_1^{1+x_j^{-\beta}} (u-1) \ln(u-1) u^{-(\alpha+2)} du - \frac{\alpha x_j^{-\beta} (1+x_j^{-\beta})^{-(\alpha+1)} \ln(x_j)}{\beta [1-F(x_j; \alpha, \beta)]^2} [1-(1+x_j^{-\beta})^{-\alpha} - \alpha \ln(1+x_j^{-\beta})]$$

ve

$$b_{22} = \frac{1}{\beta^2} - k_1(x_{(j)}; \alpha, \beta) - k_2(x_{(j)}; \alpha, \beta), \quad (24)$$

$$k_1(x_{(j)}; \alpha, \beta) = \frac{\alpha}{\beta^2 [1-F(x_j; \alpha, \beta)]} \int_1^{1+x_j^{-\beta}} (u-1) \{\ln(u-1)\}^2 u^{-(\alpha+3)} du \quad (25)$$

$$k_2(x_{(j)}; \alpha, \beta) = \frac{\alpha x_j^{-\beta} (1+x_j^{-\beta})^{-(\alpha+2)} \ln(x_{(j)})}{\beta [1-F(x_j; \alpha, \beta)]^2} [(1+x_j^{-\beta})^{-\alpha} - 1 + \alpha x_j^{-\beta}], \quad (26)$$

olarak elde edilir.

Böylece, gözlemlenen bilgi matrisi $I_x(\theta)$ yukarıda belirtilen iki matrisin kullanılmasıyla hesaplanır. α ve β parametrelerinin MLE'nin asimptotik varyans-kovaryans matrisi ise $I_x^{-1}(\hat{\alpha}, \hat{\beta})$ olarak elde edilecektir. MLE'nin asimptotik özelliğinin kullanılmasıyla $\frac{\hat{\alpha} - \alpha}{\sqrt{Var(\hat{\alpha})}}$ ve

$\frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}}$ nın yaklaşık olarak standart normal dağıldığını söyleyebiliriz. Bu özellik ile, α ve β 'nin $100(1-\tau)\%$ güven aralığı ise $\hat{\alpha} \pm Z_{\tau/2} \sqrt{Var(\hat{\alpha})}$ ve $\hat{\beta} \pm Z_{\tau/2} \sqrt{Var(\hat{\beta})}$, ifadelerinden elde edilir.

3. Bulgular

3.1 Benzetim Çalışması ve Sonuçları

Bu bölümde, önerilen parametre tahmincilerinin performanslarını karşılaştırmak amacıyla, Monte Carlo benzetim çalışması R 3.1.3 programı kullanılarak gerçekleştirilmiştir. İlerleyen tür tip II sağdan sansürlü örneklemin Burr tipi III dağılımı için, farklı parametre büyüklüklerinde, farklı örneklem genişliklerinde ve çeşitli sansürleme şemaları altında benzetim çalışması oluşturulmuştur. Her aşamada α ve β 'nin en çok olabilirlik tahmincileri Newton Raphson ve EM

algoritması ile elde edilmiş olup bu parametre tahminleri her bir adımın 1000 kez tekrarından elde edilen değerlerle bulunmuştur. Benzetim çalışmasında α ve β parametrelerinin sadece tahmin değerleri değil aynı zamanda her bir tahmine ilişkin en küçük kareler ortalaması değerleri ve %95 güven aralığında Fisher bilgi matrisi kullanılarak, parametrelere ilişkin güven aralıkları oluşturulmuştur. Benzetim çalışması için seçilen farklı büyüklüklerdeki α ve β parametre çiftlerinin değerleri Tablo 1'de verilmiştir.

Tablo 1. (α, β) değerleri

(α, β)	(0.5, 1)	(0.5, 1.5)	(1, 2)	(1, 2.5)	(2, 2.5)	(2, 3)
-------------------	----------	------------	--------	----------	----------	--------

Benzetim çalışmasında, her bir (α, β) parametre çifti için, rasgele olarak türetilen ilerleyen tür tip II sağdan sansürlü örneklem büyüklükleri değerleri ise $n=20, 30, 50$ ve 60

olarak seçilmiş olup, Burr tipi III dağılımın ilerleyen tür tip II sağdan sansürlü örneklem şemaları ise Tablo 2'de verilmiştir.

Tablo 2. Burr Tipi III Dağılımın İlerleyen Tür Tip II Sağdan Sansürlü Örnekleme Şemaları

n	m	R (Sansürleme Şeması)
20	15	(5, 0 ^{*14})
30	20	(10, 0 ^{*19})

50	35	(15,0 ^{*34})
60	40	(20,0 ^{*39})

Tablo 2’de verilen sansürleme şemasında seçilmediğini ifade etmektedir. Benzetim $R = (5,0^{*14})$ gösterimi, 1. başarısızlık çalışması sonucunda, parametre tahminleri gerçekleştiğinde örneklemden rasgele olarak için kullanılan Newton Raphson yöntemi ve $R_1 = 5$ gözlemin çekildiğini, x_2, \dots, x_{15} EM algoritmasından elde edilen sonuçlar, başarısızlıklar gerçekleştiğinde ise, Tablo 3 - Tablo 8 verilmektedir. $R_2 = 0, \dots, R_{15} = 0$ gösterimi ile de örneklemden rasgele gözlemlerin

Tablo 3. $\alpha = 0.5$ ve $\beta = 1$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	0.3984(0.2135)	0.4798(0.2204)	0.5278(0.0112)	1.0236(0.031)
			CI (0.2714-0.5628)	(0.3149-0.6971)	(0.4916-0.6317)	(0.9305-1.2484)
30	20	(10,0 ^{*19})	0.3035(0.7246)	0.5805(0.3172)	0.5172(0.0193)	1.1006(0.0126)
			CI (0.2973-0.3096)	(0.5743-0.5866)	(0.5110-0.5233)	(1.094-1.1067)
50	35	(15,0 ^{*34})	0.3769(0.4426)	0.6386(0.3678)	0.5117(0.0015)	1.1609(0.0208)
			CI (0.3737-0.3800)	(0.5977-0.6794)	(0.5085-0.5148)	(1.1200-1.2017)
60	40	(20,0 ^{*39})	0.4227(0.2201)	0.7665(0.2722)	0.5057(0.0167)	1.1288(0.0167)
			CI (0.4141-0.4312)	(0.7179-0.8150)	(0.4971-0.5142)	(1.0802-1.1773)

Tablo 4. $\alpha = 0.5$ ve $\beta = 1.5$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	0.5760(0.2703)	0.4193(0.0806)	0.5101(0.0422)	1.5693(0.0498)
			CI (0.6960-0.8959)	(0.3981-0.4405)	(0.4610-0.7103)	(1.5989-1.7988)
30	20	(10,0 ^{*19})	0.5893(0.2910)	0.3552(0.1447)	0.5249(0.0213)	1.5527(0.0170)
			CI (0.5448-0.6338)	(0.3430-0.3673)	(0.5216-0.675)	(1.5365-1.6254)
50	35	(15,0 ^{*34})	0.6571(0.8428)	0.3393(0.1606)	0.5279(0.027)	1.5651(0.025)
			CI (0.6270-0.6873)	(0.3333-0.3534)	(0.5192-0.6338)	(1.5429-1.6471)
60	40	(20,0 ^{*39})	0.6161(0.8838)	0.3625(0.1374)	0.5211(0.0617)	1.5544(0.01463)
			CI (0.5927-0.6395)	(0.3566-0.3685)	(0.5182-0.5994)	(1.5389-1.6258)

Tablo 5. $\alpha = 1$ ve $\beta = 2$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	0.5895(0.4104)	1.9217(0.0782)	0.9855(0.0120)	1.9721(0.0527)
			CI (0.5770-0.6519)	(1.8756-1.9677)	(0.9230-1.0479)	(1.9260-2.0181)
30	20	(10,0 ^{*19})	0.5917(0.4082)	1.9177(0.4496)	0.9884(0.0390)	2.1052(0.0211)
			CI (0.5511-0.6322)	(1.7739-2.0614)	(0.9478-1.0289)	(1.9614-2.2489)
50	35	(15,0 ^{*34})	0.6442(0.3855)	1.9303(0.3696)	1.0370(0.0324)	1.9874(0.0239)
			CI (0.6231-0.6652)	(1.8500-2.0105)	(1.0159-1.0580)	(1.9071-2.0676)
60	40	(20,0 ^{*39})	0.6981(0.2746)	1.9811(0.2118)	1.0176(0.01237)	2.0391(0.0185)
			CI (0.6883-0.7078)	(1.7528-2.2093)	(1.0078-1.0273)	(1.8026-2.2591)

Tablo 6 $\alpha = 1$ ve $\beta = 2.5$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	0.5192(0.4807)	2.0609(0.2391)	0.9942(0.0403)	2.2541(0.2382)
			CI (0.4817-0.5566)	(1.7157-2.4060)	(0.9567-1.0316)	(1.9089-2.5992)
30	20	(10,0 ^{*19})	0.6792(0.046)	2.1944(0.5543)	1.0466(0.0223)	2.3765(0.2056)
			CI (0.6380-0.7203)	(2.0319-2.3568)	(1.0054-1.0877)	(2.2140-2.5389)
50	35	(15,0 ^{*34})	0.7080(0.4197)	2.2541(0.3801)	1.0537(0.0315)	2.5728(0.1885)
			CI (0.6918-0.7241)	(2.0066-2.5015)	(1.0375-1.0698)	(2.3253-2.8202)
60	40	(20,0 ^{*39})	0.9386(0.3054)	2.7855(0.2248)	1.0195(0.0214)	2.5376(0.1238)
			CI (0.9270-0.9501)	(2.5764-2.9945)	(1.0079-1.0310)	(2.3285-2.7466)

Tablo 7. $\alpha = 2$ ve $\beta = 2.5$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	1.2746(0.4575)	1.8247(0.5612)	1.8247(0.0888)	2.4734(0.0443)
			CI (1.1497-1.3994)	(1.4936-2.1557)	(1.6998-1.9495)	(2.1423-2.8044)
30	20	(10,0 ^{*19})	1.1539(0.3781)	1.9204(0.4834)	1.9119(0.0150)	2.7446(0.0198)
			CI (1.0761-1.2316)	(1.6410-2.1997)	(1.8341-1.9896)	(2.4652-3.0239)
50	35	(15,0 ^{*34})	1.5869(0.3611)	1.9520(0.4174)	2.0644(0.0123)	2.4591(0.0532)
			CI (1.5422-1.6315)	(1.7903-2.1136)	(2.0197-2.1090)	(2.2974-2.6207)
60	40	(20,0 ^{*39})	1.7595(0.3117)	1.9848(0.2813)	2.0410(0.0648)	2.5378(0.2695)
			CI (1.7245-1.7944)	(1.8749-2.0946)	(2.0060-2.0759)	(2.4279-2.6476)

Tablo 8. $\alpha = 2$ ve $\beta = 3$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)
20	15	(5,0 ^{*14})	1.4936(0.4012)	2.2815(0.4735)	1.7905(0.0963)	2.5766(0.0388)

			CI (1.4235-1.5636)	(1.9302-2.6327)	(1.7204-1.8605)	(2.2253-2.9278)
30	20	(10,0 ^{*19})	1.6099(0.3874)	2.5450(0.4356)	1.9572(0.0174)	2.8472(0.0448)
			CI (1.5000-1.7197)	(2.1568-2.9331)	(1.8473-2.0670)	(2.4590-3.2353)
50	35	(15,0 ^{*34})	1.7576(0.3216)	2.5636(0.4227)	2.1247(0.0138)	3.0281(0.0321)
			CI (1.7186-1.7965)	(2.2807-2.8464)	(2.0857-2.1636)	(2.7452-3.3109)
60	40	(20,0 ^{*39})	1.8913(0.3517)	2.6112(0.2390)	2.0297(0.0182)	3.1753(0.0131)
			CI (1.8630-1.9195)	(2.4705-2.7518)	(2.0014-2.0579)	(3.0346-3.3159)

Benzetim çalışmasının sonuçlarının verildiği Tablo 3 - Tablo 8 incelendiğinde, EM algoritması ile elde edilen parametre sonuçlarının Newton Raphson yönteminden çok daha iyi olduğu görülmektedir. Aynı zamanda, EM algoritması ile elde edilen parametre tahminlerine ait MSE değerleri, Newton Raphson yöntemine göre daha küçüktür. %95 güven düzeyinde, EM algoritması ile elde edilen asimptotik güven aralıkları karşılaştırıldığında ise, EM aralıklarının Newton Raphson' a göre daha dar olduğu gözlemlenmiştir. Her iki güven aralığında bilinmeyen parametre değerlerini içermiş olsa da EM algoritması Newton Raphson yöntemine göre tercih edilir. Aynı zamanda tablolar incelendiğinde, farklı örneklem büyüklükleri farklı sansür şemalarına göre oluşturulan benzetim çalışması senaryolarında her iki yöntemin de

örneklem genişliği arttıkça, daha tutarlı sonuçlar verdiği görülmektedir.

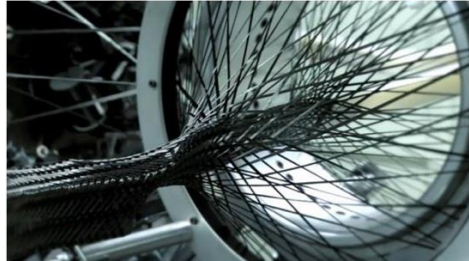
3.2. Gerçek Yaşam Verisi Uygulama Örneği

Bu bölümde, gerçek veri seti analiz edilmektedir. Literatürde, bu veri seti Burr tipi III dağılımı için tartışılmıştır. Bader ve Priest tarafından 1982 yılında yapılan araştırmada basit bir fiber lifinden, 1000 karbonluk fiber lifi içeren örneklerin elde edildiği veri seti kullanılmıştır. Bader ve Priest (1982), elde edilen fiber liflerinin dayanıklılığını 1,10,20 ve 50 mm'lik fiber lifi çap uzunluklarında incelemiş ve veriyi bu uzunlukları göz önüne alarak elde etmiştir. Ardından verileri 20, 50, 150 ve 300 mm uzunluklarında dayanıklılık testine tabi tutmuşlardır. Ölçüm uzunlukları 10 mm'lik olan tek karbon fiber lifleri için GPA'da ölçülen dayanıklılık değerlerine ilişkin veri seti aşağıdaki gibidir.

6.693, 8.432, 9.052, 9.281, 9.554, 10.486, 10.602, 10.979, 10.990, 11.531, 11.635, 11.870, 12.404, 12.453, 12.491, 12.579, 13.131, 13.654, 13.681, 13.708, 13.791, 14.282, 14.512, 15.456, 15.487, 17.392, 18.486, 18.690, 18.859, 18.859, 19.629, 20.005, 20.697, 22.760, 23.081, 23.220, 25.028, 25.103, 25.406, 25.610, 26.154, 26.364, 26.950, 27.994, 28.389,

29.283, 30.205, 31.031, 32.884, 33.149, 34.364, 34.953, 35.234, 37.637, 47.087, 47.990, 48.716, 53.038, 55.924, 56.092, 68.374, 81.045, 151.41

Şekil 2’de bir karbon lifinin gösterimi verilmiştir.



Şekil 2. Karbon Fiber Kablo Lifleri

Üzerinde çalışmış olduğumuz gerçek yaşam amacıyla; Gamma dağılım, Weibull dağılım, verisinin Burr tipi III dağılımdan gelip Log-Normal dağılım ve Üstel dağılım gelmediğinin araştırılması için, R 3.1.3 incelenmiştir. Karşılaştırmak amacıyla programında uyum iyiliği test kodları kullanılan dağılımların olasılık yoğunluk yazılmış olup, ilgili dağılım karşılaştırmak fonksiyonları Tablo 9’da verilmiştir.

Tablo 9. Dağılımların Olasılık Yoğunluk Fonksiyonları

Dağılım	Olasılık Yoğunluk Fonksiyonu
$G(\alpha, \beta)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$
$W(\alpha, \beta)$	$\frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$
$\text{Log-N}(\alpha, \beta)$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$
$\text{Exp}(\alpha)$	$\alpha e^{-\alpha x}$
$\text{Burr III}(\alpha, \beta)$	$\beta \alpha x^{-\beta-1} (1 + x^{-\beta})^{-\alpha-1}$

Gamma, Weibull, Log-Normal, Üstel ve Burr Darling (AD) uyum iyiliği testleri ile Akaike tipi III dağılımlarını karşılaştırmak için (AIC) ve Bayes (BIC) uyum iyiliği uygulanan Kolmogorov-Smirnov (KS), kriterlerine ait sonuçlar Tablo 10’ da Cramer-von Mises (CVM), Anderson- verilmiştir.

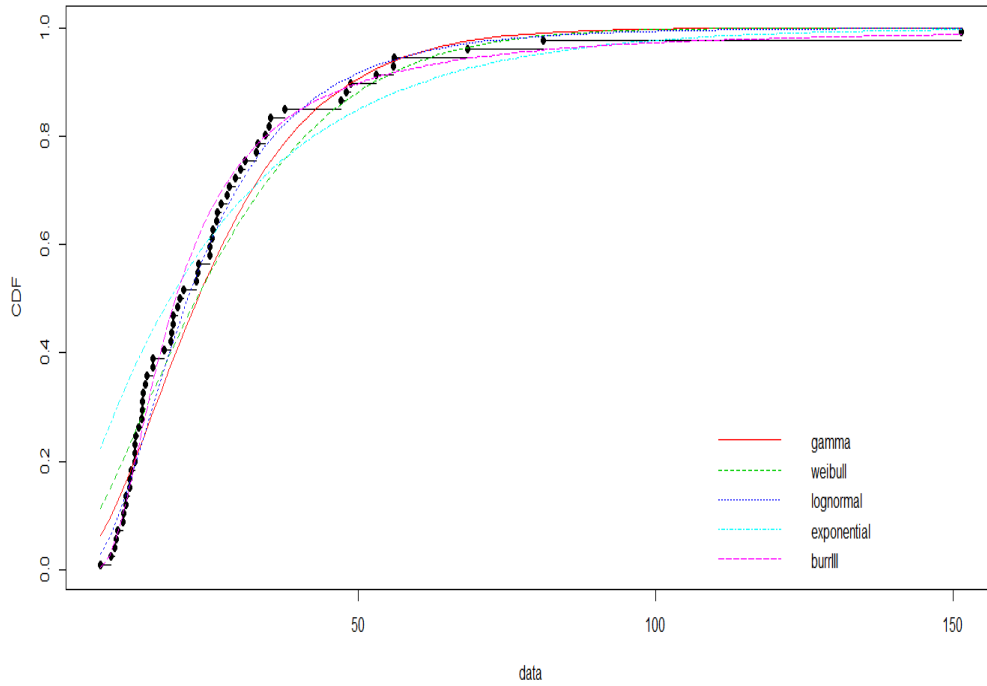
Tablo 10 Dağılımların Uyum İyiliği Test Sonuçları

Dağılım	$\hat{\alpha}$	$\hat{\beta}$	KS	CVM	AD	AIC	BIC
$G(\alpha, \beta)$	2.4731	0.0935	0.1059	0.2101	1.3532	519.6326	523.9188

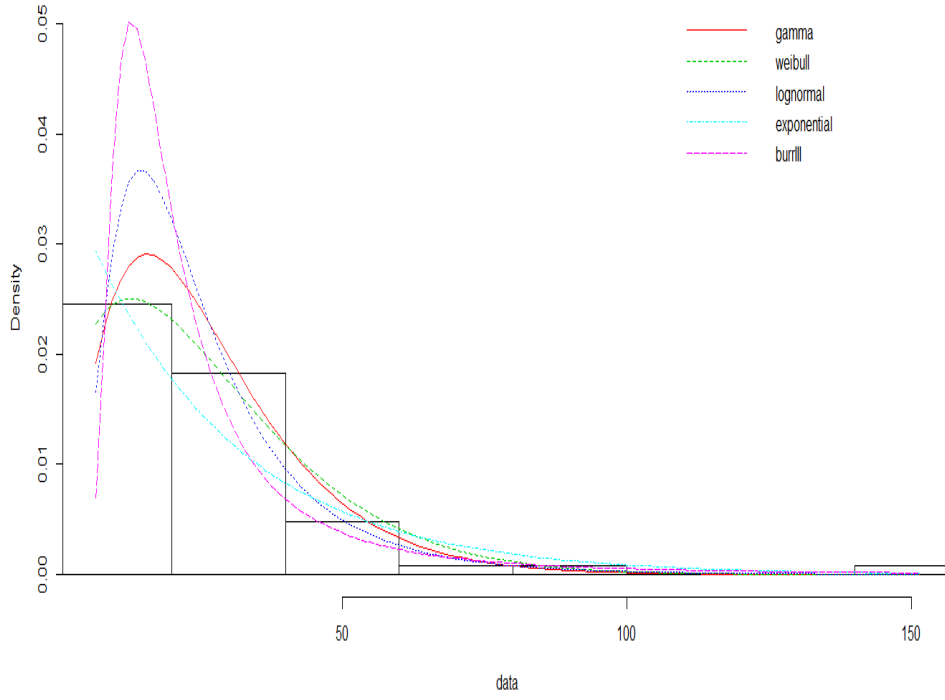
$W(\alpha, \beta)$	1.4298	29.4758	0.1379	0.2967	2.0354	528.8282	533.1145
$\text{Log-N}(\alpha, \beta)$	3.0593	0.6159	0.0987	0.0837	0.5698	507.2042	511.4905
$\text{Exp}(\alpha)$	0.0378	0.0378	0.2581	0.8697	5.0336	540.6426	542.7858
$\text{Burr III}(\alpha, \beta)$	225.6142	1.9554	0.0886	0.0684	0.3554	502.4946	506.7809

Tablo 10’ da sonuçlar incelendiğinde, açıkça görülmektedir ki Burr tipi III dağılımının en küçük test istatistiklerine ve aynı zamanda en küçük AIC ve BIC değerlerine sahip olduğu görülmektedir. Böylelikle, Burr tipi III dağılımımızın veriye en uygun dağılım

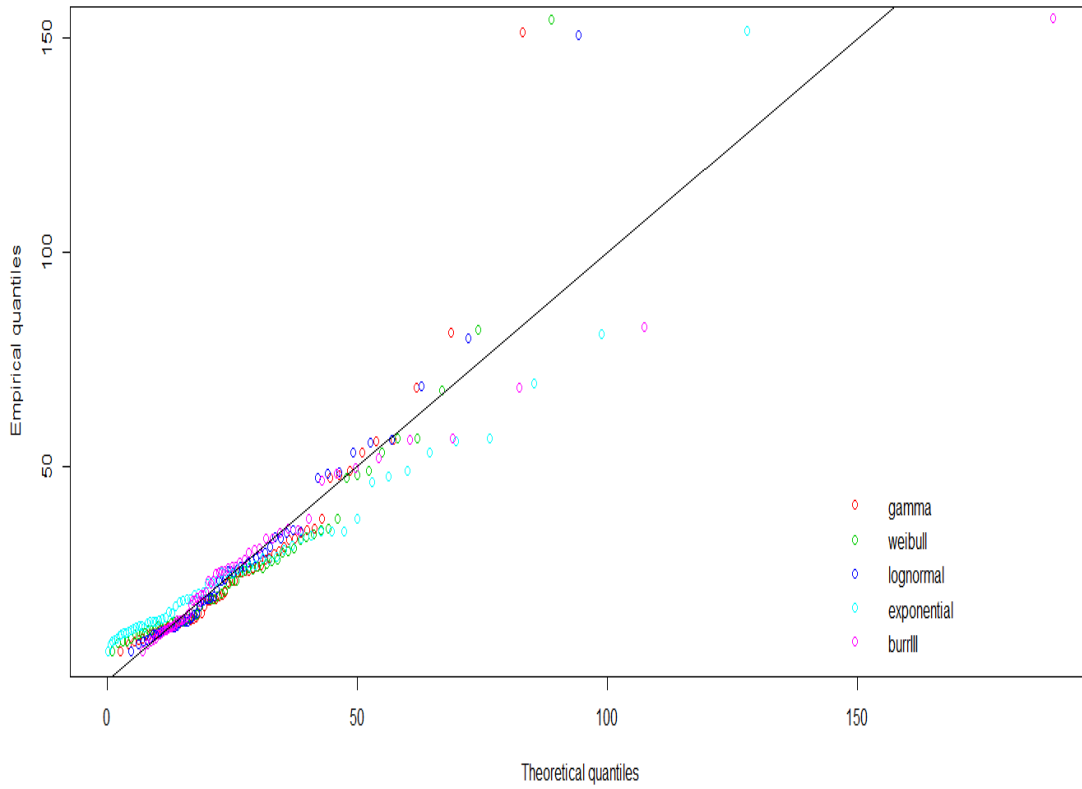
olduğunu söyleyebiliriz. Dağılımları karşılaştırmak amacıyla yapılan birikimli dağılım fonksiyonu, histogram ve yoğunluk fonksiyonları, Q-Q grafikleri ve P-P grafikleri Şekil 3-Şekil 6’da verilmiştir.



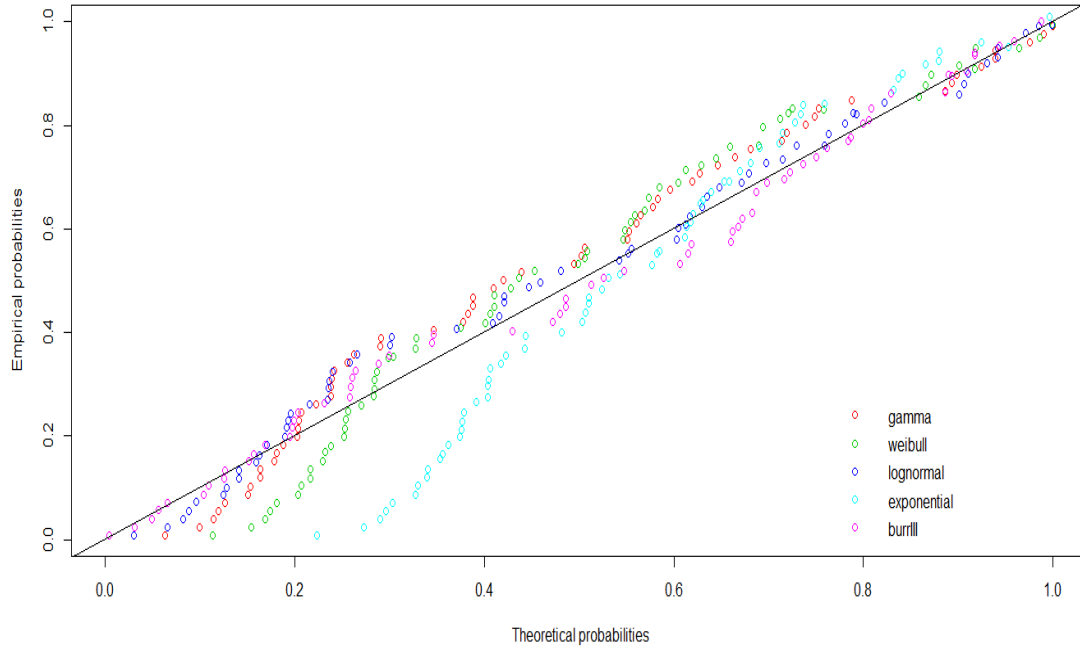
Şekil 3 Dağılımlara İlişkin Birikimli Dağılım Fonksiyonu Grafikleri



Şekil 4 Dağılımlara İlişkin Histogram ve Yoğunluk Fonksiyonu Grafikleri



Şekil 5 Dağılımlara İlişkin Q-Q Grafikleri



Şekil 6 Dağılımlara İlişkin P-P Grafikleri

Şekil 3- Şekil 6’ da ki grafikler incelendiğinde görüldüğü gibi Burr tipi III dağılım veriye en uygun dağılımdır. Gerçek yaşam verisinde, Burr tipi III dağılımı veriye uygunluğu karar verildikten sonra Newton

Raphson yöntemi ve EM algoritmalarının farklı sansürleme şemaları altında parametre tahminlerinin karşılaştırmalarını yapabilmek için Tablo 11’ de verilen sansürleme planları kullanılmıştır.

Tablo 11 Sansürleme Şeması

Sansürleme Planları	n	m	R
1	63	20	(43,0 ^{*19})
2	63	30	(33,0 ^{*29})
3	63	40	(23,0 ^{*39})
4	63	50	(13,0 ^{*49})

Burada verilen sansürleme planında,

n: örneklem sayısı, **m:** başarısızlık sayısı, **R:** sansürleme şemasını, göstermektedir.

- Sansürleme planı 1’ de görülen, n=63 ve m=20 olduğunda, 1.başarısızlık gerçekleştiğinde rasgele olarak veriden 43 gözlem çıkarılıyor. 2. ve 3.

başarısızlıklar gerçekleştiğinde örneklemden gözlemler çekilmiyor.

- Sansürleme planı 2’ de görülen, n=63 ve m=30 olduğunda, 1.başarısızlık gerçekleştiğinde rasgele olarak veriden 33 gözlem çıkarılıyor. 2. ve 3. başarısızlıklar gerçekleştiğinde

- örneklemeden gözlemler çekilmiyor. başarısızlıklar gerçekleştiğinde
- Sansürleme planı 3' de görülen, $n=63$ ve $m=40$ olduğunda, 1.başarısızlık gerçekleştiğinde rasgele olarak veriden 23 gözlem çıkarılıyor. 2. ve 3. başarısızlıklar gerçekleştiğinde örneklemeden gözlemler çekilmiyor. Tablo 11'de verilen farklı sansürleme planları altında, gerçek veriye uygulamasını yapmış olduğumuz Burr tipi III dağılımı ilerleyen tür tip II sağdan sansürlü örneklemlerin $\alpha = 2, \beta = 1$ olduğu varsayılan durumda, Newton Raphson ve EM algoritmasına ilişkin tablo sonuçları Tablo 12' de verilmiştir.
 - Sansürleme planı 4' de görülen, $n=63$ ve $m=50$ olduğunda, 1.başarısızlık gerçekleştiğinde rasgele olarak veriden 13 gözlem çıkarılıyor. 2. ve 3.

Tablo 12 . $\alpha = 2$ ve $\beta = 1$ için, Newton Raphson-EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
63	20	(43,0 ^{*19})	1.4866	2.3440	2.5241	0.9358
63	30	(33,0 ^{*29})	5.2797	1.0416	2.7469	1.0443
63	40	(23,0 ^{*39})	3.7198	0.7858	2.9107	1.1707
63	50	(13,0 ^{*49})	4.7460	0.7007	3.1776	1.3322

Tablo 12' de görüldü gibi, $\alpha = 2$ ve $\beta = 1$ olduğu durumda, farklı sansürleme planları altında elde edilen Newton Raphson yöntemi ve EM algoritması tahminlerine bakıldığında en iyi sansürleme planının 1. ve 2. sansürleme planının olduğu görülmektedir. Burada, EM algoritmasının Newton Raphson yöntemine göre, parametre değerlerinin daha yakın sonuçlar verdiği 1. ve 2. sansürleme planı altında gerçekleştiğini söyleyebiliriz.

4. Sonuç

Bu çalışmada, Burr tipi III dağılımına ilişkin ilerleyen tür tip II sağdan sansürlü örnekleme dayalı parametrelerinin EM algoritmasıyla tahmini gerçekleştirilmiştir. Önerilen

parametre tahmin edicilerinin performanslarını karşılaştırmak amacıyla R 3.1.3 programı kullanılarak Monte Carlo benzetim uygulanmıştır. İlerleyen tür tip II sağdan sansürlü örneklemin Burr tipi III dağılımı için, farklı parametre büyüklüklerinde, farklı örneklem genişliklerinde ve çeşitli sansürleme şemaları altında en çok olabilirlik tahmincileri Newton Raphson ve EM algoritması ile elde edilmiş olup bu parametre tahminleri her bir adımın 1000 kez tekrarından elde edilen değerlerle bulunmuştur. EM algoritmasının performansı, Newton Raphson yöntemiyle karşılaştırılmıştır ve araştırmacılara performansı daha iyi sonuç veren EM

algoritması önerilmiştir. Gerçek yaşam verisi, Badar ve Priest tarafından 1982 yılında basit bir fiber lifinden, 1000 karbonluk fiber lifi içeren örneklere kadar elde edilmiştir. 1, 10, 20, 50 mm'lik fiber lifi çap uzunluklarını incelemişlerdir. Veriyi bu uzunlukları göz önüne alarak elde etmişler ve veriyi 20, 50, 150 ve 300 mm uzunluklarında GPA dayanıklılık testine tabi tutmuşlardır. Veri setinin R 3.1.3 programında gerçekleştirilen uyum iyiliği testlerinden ve uyum iyiliği kriterlerinden en küçük sonucu vererek en iyi Burr tipi III dağılımına uyduğu gözlenmiştir. Farklı sansürleme planları altında elde edilen Newton Raphson ve EM algoritması tahminlerine bakıldığında EM algoritması parametre değerlerinin daha yakın sonuç verdiği görülmüş ve araştırmacılara EM algoritması önerilmiştir. %95 güven düzeyinde, EM algoritması ile elde edilen asimptotik güven aralıkları karşılaştırıldığında ise EM aralıklarının Newton-Raphson' a göre daha dar olduğu gözlemlenmiştir. Her iki güven aralığı da bilinmeyen parametre değerlerini içermiş olsa da EM algoritmasının Newton-Raphson yöntemine göre tercih edilebileceği ortaya konmuştur.

Sonuç olarak, gerçek yaşam verisi ve benzetim çalışmasında, EM algoritması ile elde edilen parametre tahminlerine ait MSE değerleri, Newton Raphson yöntemine göre daha küçüktür. %95 güven düzeyinde, EM algoritması ile elde edilen asimptotik güven

aralıkları karşılaştırıldığında ise, EM aralıklarının Newton Raphson'a göre daha dar olduğu gözlemlenmiştir. Her iki güven aralığı da bilinmeyen parametre değerlerini içermiş olsa da EM algoritması Newton Raphson yöntemine göre tercih edilir.

5. Kaynaklar

Aray S., (2018). "Burr Tipi III Dağılımının İlerleyen Tür Tip II Sağdan Sansürlü Örnekleme Dayalı Parametrelerinin EM Tahmin Edicisi".Yüksek Lisans. *Muğla Sıtkı Koçman Üniv., Fen Bilimleri Enstitüsü.* Muğla, 1-84.

Arabi Belaghi, R., Valizadeh Gamchi, F., Bevrani, H., Gürünlü Alma, Ö. 2016. "Estimation on Burr type III by progressive censoring using the EM and SEM algorithms". *13th Iranian Statistics Conference Shadid Bahonar University of Kerman* 23-25 August 2016.

Bader, M.G. and Priest, A.M. (1982) "Statistical Aspects of Fibre and Bundle Strength in Hybrid Composites". In: *Hayashi, T., Kawata, K. and Umekawa, S., Eds., Progress in Science and Engineering of Composites, ICCM-IV, Tokyo, 1129-1136.*

Bairamov I, Eryılmaz S .,2006. "Spacings, exceedances and concomitants in progressive type II censoring scheme". *Journal of Statistical Planning and Inference* 136(3), 527-536.

Balakrishnan and Sandhu 1995. "A simple simulation algorithm for generating progressive Type-II censored sample". *American Statisticians.* 49:229-230.

Balakrishnan, N. and Aggarwala, R. 2000. "Progressive Censoring: Theory, Methods and Applications". *Birkhäuser, Boston.*

- Burr, I. W., 1942. "Cumulative frequency functions". *The Annals of mathematical statistics* 13(2): 215-232.
- Demir, E. 2015. "Eğitim bilimlerinde olay geçmişi çözümlemesi", *Uluslararası Sosyal Araştırmalar Dergisi*, Cilt 8:570-582
- Dempster, A.P., Laird, N.M, and Rubin, D., 1977. "Maximum likelihood estimation from incomplete data via the EM algorithm", *Computational Statistics and Data Analysis, Journal Royal Statistical Soc.* Vol: 39, 1–38.
- Fesler, J.A. and Hero, A.O., 1994." Space-alternating generalized expectation maximization algorithm". *IEEE Transactions on Signal Processing*, 42, 2664-2677.
- Gupta, R. D., & Kundu, D., 2001) "Exponentiated Exponential Family:an Alternative to Gamma and Weibull Distribution". *Biometrika* 43, 117-130.
- İnceoğlu, F. (2013), "Sağkalım analizi yöntemleri ve karaciğer nakli verileri ile bir uygulama". *Yüksek Lisans, İnönü Üniversitesi*, Malatya, 142s.
- Kale, B. 2003. "İlerleyen Tür Sansürlenmiş Sıra İstatistikleri:Dağılım Özellikleri Ve Uygulamalar". Yüksek Lisans Tezi, *Ankara Üniversitesi Fen Bilimleri Enstitüsü*.
- Kus C., 2004. "Bazı yaşam zamanı dağılımlarının parametrelerinin tam ve sansürlü verilere dayalı tahmini", *S.Ü. Fen Bilimleri Enstitüsü*, Doktora tezi.
- Kuş, C. 2005. "İlerleyen Tür Tip II Sağdan Sansürlü Örnekleme Dayalı Düzgün Dağılımın Parametrelerinin Jackknife Tahmin Edicisi". *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 2, 11-29.
- Louis, T.1982. A., "Finding the observed information matrix when using the EM algorithm". *J. Roy. Statist. Soc. Ser. B*, 44, 226-233.
- McKendrick, A.G., 1926. "Applications of mathematics to medical problems". *Proc. Edinb. Math. Soc.* 13, 98–130.
- Nelson, W. 1982, "Applied Life Data Analysis". *Wiley, New York*, 584.
- Ng T., Chan CS., Balakrishnan N.,2002. "Estimation of parameters from progressively censored data using EM algorithm". *Computational Statistics and Data Analysis*, 39, 371-386.
- Rastogi M., Tripathi Y., 2013. "Inference on unknown parameters of a Burr distribution under hybrid censoring". *Statistical Papers*, Springer, 54(3), 619-643.
- Saraçoğlu, B., Kınacı, İ., Kuş, C. 2006. "Power Hazard Fonksiyonuna Sahip Dağılım Fonksiyonlarının Parametrelerinin İlerleyen Tür Tip II Sağdan Sansürlü Örneklemlere Dayalı En Küçük Kareler Tahmini". *5. İstatistik Günleri Sempozyumu*(82).
- Tadikamalla, P. R. (1980). "A look at the Burr and related distributions". *International Statistical Review/Revue Internationale de Statistique*: 337-344.
- Terzi, Y. 2003. "Sansürlü Veriler İçin Sağkalım Analizi ve Gerçek Verilere Uygulaması", Doktora Tezi, *Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü*.
- Yazıcı, F. 2005. "EM Algoritması ve Uzantıları", Ankara, 85s.
- Zhu HT, Lee SY, Wei BC, Zhou J. 2001. "Case-Deletion Measures for Models With Incomplete Data".*Biometrika*. 88:727–737.