



GENELLEŞTİRİLMİŞ ÜSTEL DAĞILIMIN İLERLEYEN TÜR TİP II SAĞDAN SANSÜRLÜ ÖRNEKLEME DAYALI PARAMETRELERİNİN EM TAHMİN EDİCİSİ

Özlem GÜRÜNLÜ ALMA^{1,*} , Rabia ÖZAR² 

¹ İstatistik Bölümü, Fen Fakültesi, Muğla Sıtkı Koçman Üniversitesi, Muğla, Türkiye

² İstatistik Bölümü, Fen Bilimleri Enstitüsü, Muğla Sıtkı Koçman Üniversitesi, Muğla, Türkiye

ÖZET

Bu çalışmada, genelleştirilmiş üstel dağılıma ait bilinmeyen parametrelerin istatistiksel tahmini ilerleyen tür tip II sağdan sansürlü örneklem durumunda incelenmiştir. Bilinmeyen parametrelerin en çok olabilirlik tahmin edicileri Newton-Raphson 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 bir gerçek yaşam verisi örneği de verilmiştir.

Anahtar Kelimeler: Beklenti Maksimizasyonu Algoritması, En Çok Olabilirlik Tahmincisi, Genelleştirilmiş Üstel Dağılım, İlerleyen Tür Tip II Sağdan Sansürlü Örneklem, Newton-Raphson yöntemi, Sansürleme

ABSTRACT

In this study, the statistical estimation of the unknown parameters of the generalized exponential 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: Expectation Maximization Algorithm, Maximum Likelihood Estimator, Generalized Exponential Distribution, Progressively Type II Right Censored Samples, Newton-Raphson method, Censoring

1. GİRİŞ

Yaşam analizi, başarısızlık olarak adlandırılan, genel olarak da ölüm veya bozulma olarak belirlenen herhangi bir olayın ortaya çıkmasına kadar geçen süre içindeki verilerin analizidir [1]. Burada süre; araştırmanın başından ilgilenilen olay gerçekleşinceye kadar geçen yılları, ayları, haftaları, günleri ya da olay gerçekleştiği anda birimin yaşını ifade eder. Olay ise ölüm, hastalığa yakalanma, hastalığın kötüye gitmesi, iyileşme, işe dönme ya da birimin başına gelebilecek ilgilenilen herhangi bir olayı ifade etmektedir [2]. Yaşam analizi verilerinin kaynağı genel olarak sağlık bilimleri olmakla birlikte nüfus bilimi, mühendislik, ekonomi, işletme ve sosyal bilimler gibi diğer alanlar da olabilir. Sosyolojide olay geçmişi analizi, mühendislikte başarısızlık zamanı analizi, ekonomide süre analizi ya da geçiş analizi ve sağlık alanında ise yaşam analizi olarak adlandırılır [3]. Buradan da anlaşılacağı gibi yaşam analizi sadece hastalık ve yaşamı etkileyen faktörlerin analizinde yararlanılan bir yöntem değildir. Ölüm veya hayatta kalma süreleri ile sınırlanamaz. Örneğin; evli olan çiftlerin evli kalma süreleri, ekonomik alanda şirketlerin aldıkları iş makinelerinin bozulma süreleri veya işlevini göremez duruma gelme süreleri, makinelerin ardışık iki kez bozulma süreleri arasında geçen süre, elektronik parçaların veya aletlerin yaşam sürelerinin analiz edilmesi gibi birçok alanda kullanılır [4]. Bu nedenle söz konusu olay, çalışmanın amacına göre farklılık gösterir. Bu olay canlılar için genellikle ölüm, cansız nesnelere içinde bozulma olarak ele alınır [5].

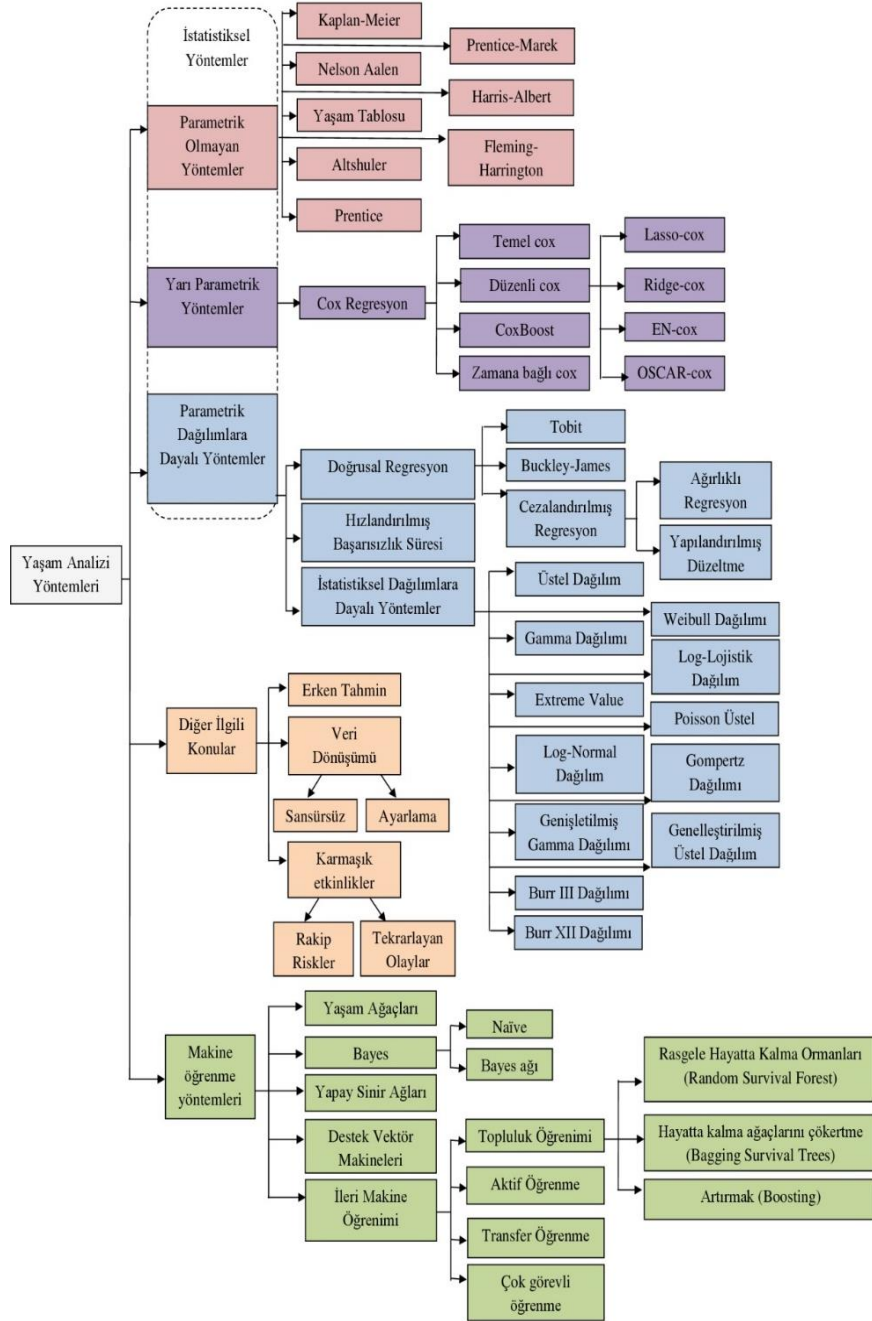
Güvenilirlik çalışmalarında ve yaşam analizlerinde sonuca ulaşmak için sistemi oluşturan tüm bileşenlerin bozulma zamanlarını gözlemlemek her zaman mümkün olmayabilir. Örneğin; bir klinikte tedavi gören hastalara ilişkin veriler, eksiksiz gözlenemeyebilir veya pahalı bir elektronik parçanın yaşam zamanı hakkında bilgi edinmek için yapılan yaşam testinde, parçaların hepsinin bozulmalarının

*Sorumlu Yazar: ozlem.gurunlu@gmail.com

Geliş: 13.10.2019

Yayın: 31.08.2020

gözlenmesi maliyeti ve test zamanını artıracığından istenmeyebilir. Bu tip durumlarda, deney ya da gözlem sonrası sansürlenmiş veri elde edilir. Tıp, biyoloji, mühendislik, kalite kontrol ve birçok alanda sansürlenmiş verilerle karşılaşılır [6]. Bu nedenle yaşam verileri genellikle sansürlüdür. Yaşam verilerinin bir diğer özelliği ise simetrik dağılmaz. Benzer bireylerden oluşan bir gruptan elde edilen yaşam verileri sağa çarpık bir dağılım gösterdiğinden bu tür verilerin normal dağılımlı olduğu kabul edilemez. Bu durumda veri dağılımını simetrik bir dağılıma dönüştürmek için verilere logaritmik dönüşüm yapılır. Fakat bu dönüşümü yaparken hesaplamalarda kayıplar olur. Veri yapısındaki bu farklılık, yaşam analizini diğer istatistiksel analiz yöntemlerinden ayıran en önemli özelliktir. Yaşam analizinde kullanılan yöntemler Şekil 1’de verilmiştir.



Şekil 1. Yaşam Analizi Yöntemleri¹

¹ Wang, P., Li, Y. ve K. Reddy (2017). Machine Learning for Survival Analysis: A Survey, *ACM Computing Surveys*, 1, 1, Article: 1-38.

Sansürlü verilerin analizinde kimi zaman en çok olabilirlik tahminleri kapalı formda elde edilemediğinden, gözlenen ve kayıp verilere göre oluşturulan olabilirlik fonksiyonuna dayalı EM algoritması kullanılarak parametrelerinin en çok olabilirlik tahminleri elde edilir. Bu algoritma [7] tarafından önerilmiştir. Sansürlü veri analizlerine ilişkin daha önce yapılan çalışmalardan bazıları aşağıda verilmiştir:

Gupta ve Kundu (1999, 2000), çalışmalarında genelleştirilmiş üstel dağılımı, dağılımın tahminlerinin farklı yöntemlerle elde edilmesini ve bazı özelliklerinden bahsetmişlerdir. Üstel dağılım ailesinden olan Gamma ve Weibull dağılımları ile karşılaştırmasını yapmışlardır.

Yazıcı (2005), tez çalışmasında istatistiksel analizlerde ve yaşam analizlerinde kullanılan veri türlerinden bahsetmiştir. Genellikle yaşam analizlerinde sık karşılaşılan sansürlü veri durumunda parametre tahmini hesaplamalarında kullanılan EM algoritmasını ve uzantılarını tanıtmayı amaçlamıştır.

Kuş ve Kaya (2006), ilerleyen tür tip II sağdan sansürlemeye dayalı log-lojistik dağılımının parametreleri için en çok olabilirlik tahmin edicilerini EM algoritması kullanarak elde etmiş ve bir örnek üzerinde uygulamışlardır.

Balakrishnan (2007), ilerleyen tür sansürlü sıra istatistiklerinin özelliklerini ve ilerleyen tür sansürlenmiş örnekleme dayalı istatistiksel sonuç çıkarımı üzerinde çalışmıştır. İlerleyen tür sansürleme alanındaki gelişmelerden bahsedip daha sonra yapılacak çalışmalar için olası problemlere değinmiştir.

Pradhan ve Kundu (2008), Genelleştirilmiş üstel dağılımın ilerleyen tür sansürlü örneklem durumunda EM algoritmasını kullanarak bilinmeyen parametrelerinin en çok olabilirlik tahminlerini ve istatistiksel çıkarımını ele almışlardır.

Saraçoğlu (2012), çalışmasında, Pareto dağılımının parametrelerinin ilerleyen tür tip-II sağdan sansürlü örneklere dayalı en çok olabilirlik ve en küçük kareler tahmin edicilerinin yanlılık, hata kareler ortalaması ve varyanslarına ilişkin bir Monte Carlo simülasyon çalışması yapmış ve elde edilen sonuçlar doğrultusunda n sabit tutulup m arttırıldığında, λ ve β parametresinin en çok olabilirlik ve en küçük kareler tahmin edicilerinin yanlılıkları ve hata kareler ortalamalarının azaldığını görmüştür. Bütün sansür şemaları için, β 'nın en küçük kareler tahmin edicisinin yanı ve hata kareler ortalaması, en çok olabilirlik tahmin edicisinin yanına ve hata kareler ortalamasına göre daha büyük elde etmiştir. Değişik sansür şemaları için β için en çok olabilirlik tahmin edicisinin kullanılması gerektiğini önermiştir. Bütün sansür şemaları için, λ 'nın en küçük kareler tahmin edicisinin yanı ve hata kareler ortalaması, en çok olabilirlik tahmin edicisinin yanına ve hata kareler ortalamasına göre daha küçük elde etmiştir. Dolayısıyla değişik sansür şemaları için λ için en küçük kareler tahmin edicisinin kullanılması gerektiğini önermiştir.

Zhu, S. (2013) çalışmasında genelleştirilmiş üstel dağılımın yeni hibrid tahmin yöntemlerinden bahsetmektedir. Monte Carlo simülasyon çalışması ile sonlu örneklem performanslarını karşılaştırmıştır.

Eteman (2014), tez çalışmasında sistemlerin ve ürünlerin güvenilirlik seviyelerinin hesaplanması ve analiz edilmesini sağlayan, güvenilirlik, ortalama yaşam ve diğer genel güvenilirlik sonuçlarını veren istatistiksel analiz seçeneklerini sunan çeşitli hesaplama çizim ve rapor üreten istatistiksel araçları tek bir yazılım bünyesinde toplamayı amaçlamıştır. Bu amaç doğrultusunda geliştirilmiş olan bu yazılım ile çeşitli dağılımlardan tam ya da sansürlü veriler üretilebilmekte, gözlenmiş tam ya da sansürlü verilere dayalı parametre tahmini yapılabildiğini ortaya koymuştur.

Demir (2015), tez çalışmasında, Balakrishnan ve Sandu' nun (1995) ilerleyen tür tip-II sağdan sansürlü örneklem için sayı üretme algoritmasını kullanarak Weibull, Gompertz, Üstelleştirilmiş Üstel, Chen ve Üstelleştirilmiş Pareto dağılımlarından veriler üreterek en çok olabilirlik ve Lindley yaklaşımı kullanarak bayes tahmin edicilerini elde etmiştir. Genelleştirilmiş Gompertz dağılımı için en çok olabilirlik tahmin edicilerini incelemiş ve tüm durumlar için en çok olabilirlik tahmin edicisinin yaklaşık bayes tahmin edicisine göre daha iyi sonuçlar verdiğini gözlemlemiştir. İlerleyen tür tip-II sağdan sansürlü örneklem durumunda incelenen dağılımlar için yaklaşık bayes tahmin edicileri ve en çok olabilirlik tahmin edicilerinin her ikisinin de kullanılabilir olduğu sonucuna ulaşmıştır.

Arabi Belaghi, Valizadeh Gamchi, Bevrani, Gürünlü Alma (2016), çalışmalarında ilerleyen tür sansürleme altında iki parametrelili Burr tipi III dağılımının parametreleri kapalı formda elde edilemediğinden EM ve SEM algoritmalarını kullanarak tahmin edicileri bulmuşlardır. Beyaz kan hücresi kanserli hastalarının analizini yaparak SEM, EM ve MLE sonuçlarını karşılaştırmışlardır.

Bu çalışmada, [8] tarafından önerilen sağlık, ekonomik ve çevresel verilere uygulanabildiği gibi özellikle güvenilirlik ve yaşam analizi için oldukça kapsamlı bir dağılım olan genelleştirilmiş üstel dağılım ele alınmış ve ilerleyen tür tip-II sağdan sansürlü örneklemi üzerinde çalışılmıştır. Genelleştirilmiş üstel dağılıma ait bilinmeyen parametrelerinin en çok olabilirlik tahmin edicileri Newton-Raphson 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 bir gerçek yaşam verisi örneği de verilmiştir.

Çalışmanın ilerleyen bölümleri şu şekildedir: Çalışmanın ilk bölümü olan giriş bölümünde, yaşam çözümlemesinin insan hayatındaki yerinden bahsedilerek kısa bir giriş yapılmıştır. İkinci bölümde ise genelleştirilmiş üstel dağılımın olasılık yoğunluk ve dağılım fonksiyonu, genelleştirilmiş üstel dağılımdan gelen sıralı örneklemin özellikleri ve genelleştirilmiş üstel dağılımın parametre tahmin yöntemlerine yer verilmiş, Genelleştirilmiş üstel dağılımın ilerleyen tür tip-II sağdan sansürlü örneklemi için Newton-Raphson yöntemi ve EM algoritması kullanılarak tahmin edicileri hesaplanmıştır. Üçüncü bölümde simülasyon yoluyla, bu tahminlerin davranışı farklı sansür şemaları ve parametre değerleri altında incelenmiş ve karşılaştırılmış ve daha açıklayıcı olması amacıyla bir gerçek yaşam verisi örneği de verilmiştir. Son olarak dördüncü bölümde elde edilen sonuçlar yorumlanmış ve öneriler sunulmuştur.

2. MATERYAL VE METOT

Yaşam verilerini analiz etmek için Gamma ve Weibull dağılımları yirminci yüzyılın ortasından itibaren popüler olarak kullanılmaktadır. Her iki dağılımda herhangi bir pozitif gerçek veri setini analiz etmek için esnek olmasına rağmen her ikisinin de bazı dezavantajları mevcuttur. Örneğin, Gamma dağılımının parametresi bir tam sayı değilse dağılım fonksiyonu ya da hayatta kalma fonksiyonu kapalı bir biçimde ifade edilememektedir. [19], Weibull dağılımının en çok olabilirlik tahmin edicilerinin tüm parametre değerleri için düzgün bir şekilde çalışmadığını belirtmişlerdir. Bu nedenlerden dolayı negatif olmayan bir rasgele değişken için çarpık dağılım gereken durumlarda Gamma ve Weibull dağılımlarına alternatif olarak kullanılabilen genelleştirilmiş üstel dağılım, [8, 9, 20, 21], tarafından yeni bir dağılım olarak önerilmiştir [15].

2.1 Genelleştirilmiş Üstel Dağılım

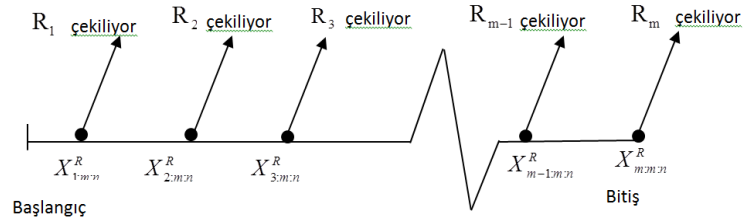
[8] tarafından önerilen iki parametrelili genelleştirilmiş üstel dağılım λ ölçek parametresi ve α şekil parametresi ile tanımlanmaktadır. Dağılım fonksiyonu ve buna karşılık gelen olasılık yoğunluk fonksiyonu Eşitlik 1 ve 2'de verilmiştir.

$$F_{GE}(x; \alpha, \lambda) = (1 - e^{-\lambda x})^\alpha, \quad \alpha, \lambda > 0, x \geq 0 \quad (1)$$

$$f_{GE}(x; \alpha, \lambda) = \alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}, \quad \alpha, \lambda > 0, x \geq 0. \quad (2)$$

İlerleyen tür tip II sansürleme, tip II sansürlemenin genelleştirilmiş halidir. İlerleyen tür tip-II sağdan sansürlü örnekleme, yaşam zamanı analizlerinde veri elde etmede önemli bir yöntemdir ve aşağıda belirtildiği şekilde tanımlanabilir.

Aynı dağılımlı n bağımsız birim için, dağılım fonksiyonu $F(x)$ ve olasılık yoğunluk fonksiyonu $f(x)$ olan X_1, X_2, \dots, X_n bozulma zamanları ile bir yaşam testi düşünölsün. Sistemde meydana gelen 1. bozulma ile rasgele R_1 sayıda bileşenin sistemden çekildiğini, daha sonra hayatta 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 örnekleme ilerleyen tür tip-II sağdan sansürlü örnekleme denir. Burada $n = m + \sum_{i=1}^m R_i$ biçimindedir. Önceden belirlenmiş bozulma sayısı m ve $R = (R_1, R_2, \dots, R_m)$ ilerleyen tür tip-II sağdan sansürleme şeması olarak adlandırılır [22]. Bu sansürleme şeması Şekil 2’de gösterilmiştir:



Şekil 2. İlerleyen tür tip-II sağdan sansürleme şeması

Bu durumda tamamen gözlenmiş bozulma zamanları $X_{i:m:n}^{(R_1, \dots, R_m)}$, $i = 1, 2, \dots, m$ ile gösterilir. Notasyonu basitleştirmek için sansürleme şeması açık olduğunda $X_{i:m:n}^{(R_1, \dots, R_m)}$ yerine $X_{i:m:n}$, $i = 1, 2, \dots, m$ kullanılır. $X_{1:m:n}^R < X_{2:m:n}^R < \dots < X_{m:m:n}^R$, olasılık yoğunluk fonksiyonu $f(x_i; m:n)$ ve dağılım fonksiyonu $F(x_i; m:n)$ 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:

$$f_{X_{1:m:n}^R, X_{2:m:n}^R, \dots, X_{m:m:n}^R}(x_1, x_2, \dots, x_m) = a \prod_{i=1}^m f(x_i; m:n) [1 - F(x_i; m:n)]^{R_i}, \quad -\infty < x_1 < x_2 < \dots < x_m < \infty \quad (3)$$

şeklinde ifade edilir. Burada,

$$a = n(n - R_1 - 1) \times \dots \times (n - R_1 - R_2 - \dots - R_{m-1} - m + 1) \quad (4)$$

şeklinindedir. Eşitlik 3’de $R = (0, \dots, 0)$ olduğunda bilinen sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu, $R = (0, \dots, n - m)$ olduğunda ise tip-II sağdan sansürlü sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu elde edilir [22].

2.2. Newton Raphson Yöntemi İle Tahmin

Newton–Raphson yöntemi, doğrusal olmayan denklemleri çözmek için kullanılabilen bir sayısal yineleme prosedürüdür. Yaklaşımın optimumluğu, gerçek fonksiyon için optimum bir tahmin vermedir. Bu tahmin optimum seviyeye yakın değilse, yeni bir yaklaşım hesaplanır ve işlem tekrarlanır. Bu yöntem Newton Raphson yöntemi olarak belirtilir. Newton-Raphson yönteminin avantajı, hızlı yakınsamasıdır. Özellikle, θ minimuma yakın olduğunda çok hızlı yakınsar. Yöntemin dezavantajları ise, parametrelerin sayısı çok fazla olduğunda Hessian matrisinin tersinin bulunmasının zorlaşması ve θ_i minimuma yakın değilse, yakınsamanın olmaması durumunda H matrisinin negatif tanımlı olmasıdır [23].

Genelleştirilmiş üstel dağılımın ilerleyen tür tip II sağdan sansürlü örnekleme için Newton-Raphson yöntemiyle MLE tahminlerinin elde edilmesi aşağıda verilen eşitliklerle sağlanır.

$$\ln[L(\alpha, \lambda)] = \sum_{j=1}^m \ln \left[\alpha \lambda e^{-\lambda x_j} (1 - e^{-\lambda x_j})^{\alpha-1} \right] + \sum_{j=1}^m R_j \ln(1 - (1 - e^{-\lambda x_j})^\alpha) \quad (5)$$

α parametresi için log-olabilirlik fonksiyonunun ikinci dereceden türevi aşağıdaki gibi elde edilir.

$$\begin{aligned} \frac{\partial^2 \ln[L(\alpha, \lambda)]}{\partial \alpha^2} = & \sum_{j=1}^m \left[\frac{-2\lambda e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1} \ln(e^{-\lambda x_j})}{\alpha \lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1})} \right. \\ & \left. - \frac{\alpha \lambda e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1} \ln(e^{-\lambda x_j})^2}{\alpha \lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1})} \right] \\ & - \left[\frac{\lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1}) - \alpha \lambda e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1} \ln e^{-\lambda x_j}}{\alpha^2 \lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1})} \right] \\ & + \left[\frac{(\lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1}) - \alpha \lambda e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1}) (e^{-\lambda x_j})^{\alpha-1} \ln(e^{-\lambda x_j}))}{\alpha \lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1})^2} \right] \\ & + \sum_{j=1}^m \left[-\frac{R_j (1 - e^{-\lambda x_j})^\alpha \ln(1 - e^{-\lambda x_j})^2}{1 - (1 - e^{-\lambda x_j})^\alpha} - \frac{R_j [(1 - e^{-\lambda x_j})^\alpha]^2 \ln(1 - e^{-\lambda x_j})^2}{(1 - (1 - e^{-\lambda x_j})^\alpha)^2} \right] \end{aligned} \quad (6)$$

λ parametresi için log-olabilirlik fonksiyonunun ikinci dereceden türevi aşağıdaki gibi elde edilir.

$$\begin{aligned} \frac{\partial^2 \ln[L(\alpha, \lambda)]}{\partial \lambda^2} = & \sum_{j=1}^m \frac{-2\alpha x_j e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1}) + 2\alpha e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1} (\alpha - 1) + \alpha \lambda x_j 2e^{-\lambda x_j}}{(1 - (e^{-\lambda x_j})^{\alpha-1}) - 2\alpha \lambda x_j 2e^{-\lambda x_j} (e^{-\lambda x_j})^{\alpha-1} \cdot (\alpha - 1)} \\ & \cdot \frac{-\alpha \lambda e^{-\lambda x_j} \cdot (e^{-\lambda x_j})^{\alpha-1} (\alpha - 1)^2 x_j^2}{\alpha \lambda e^{-\lambda x_j} (1 - (e^{-\lambda x_j})^{\alpha-1})} \end{aligned} \quad (7)$$

$$\begin{aligned}
 & \frac{\alpha e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda x_j e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) + \alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} (\alpha-1) x_j}{\alpha \lambda^2 e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)} \\
 & + \frac{\left(\alpha e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda x_j e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)\right)}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)} \\
 & + \frac{\alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} (\alpha-1) x_j}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)} \\
 & + \frac{\left[\alpha e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda x_j e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)\right]}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)^2} \\
 & + \frac{\alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} (\alpha-1) x_j e^{-\lambda x_j}}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)^2} \\
 & + \sum_{j=1}^m \frac{R_j \left(1 - e^{-\lambda x_j}\right)^\alpha \alpha^2 x_j^2 \left(e^{-\lambda x_j}\right)^2}{\left(1 - e^{-\lambda x_j}\right)^2 \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)} + \frac{R_j \left(1 - e^{-\lambda x_j}\right)^\alpha \alpha x_j 2 e^{-\lambda x_j}}{\left(1 - e^{-\lambda x_j}\right) \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)} \\
 & + \frac{R_j \left(1 - e^{-\lambda x_j}\right)^\alpha \alpha x_j^2 \left(e^{-\lambda x_j}\right)^2}{\left(1 - e^{-\lambda x_j}\right)^2 \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)} - \frac{R_j \left(\left(1 - e^{-\lambda x_j}\right)^\alpha\right)^2 \alpha^2 x_j^2 \left(e^{-\lambda x_j}\right)^2}{\left(1 - e^{-\lambda x_j}\right)^2 \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)^2}
 \end{aligned}$$

α ve λ parametreleri için log-olabilirlik fonksiyonunun türevleri aşağıdaki gibi elde edilir.

$$\begin{aligned}
 \frac{\partial^2 \ln[L(\alpha, \lambda)]}{\partial \alpha \partial \lambda} &= \sum_{j=1}^m \frac{\lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} \ln(e^{-\lambda x_j})}{\alpha \lambda^2 e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)} \\
 & + \frac{\left[\lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} \ln(e^{-\lambda x_j})\right] x_j}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)} \\
 & - \frac{\left(\lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right) - \alpha \lambda e^{-\lambda x_j} \left(e^{-\lambda x_j}\right)^{\alpha-1} \ln(e^{-\lambda x_j})\right) \left(e^{-\lambda x_j}\right)^{\alpha-1} (\alpha-1) x_j}{\alpha \lambda e^{-\lambda x_j} \left(1 - (e^{-\lambda x_j})^{\alpha-1}\right)^2} \\
 & + \sum_{j=1}^m \frac{R_j \left(1 - e^{-\lambda x_j}\right)^\alpha \alpha x_j e^{-\lambda x_j} \ln(1 - e^{-\lambda x_j})}{\left(1 - e^{-\lambda x_j}\right) \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)} - \frac{R_j \left(1 - e^{-\lambda x_j}\right)^\alpha x_j e^{-\lambda x_j}}{\left(1 - e^{-\lambda x_j}\right) \left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)} \\
 & - \frac{R_j \left(\left(1 - e^{-\lambda x_j}\right)^\alpha\right)^2 \ln(1 - e^{-\lambda x_j}) \alpha x_j e^{-\lambda x_j}}{\left(1 - \left(1 - e^{-\lambda x_j}\right)^\alpha\right)^2 \left(1 - e^{-\lambda x_j}\right)}
 \end{aligned} \tag{8}$$

2.3. EM Algoritması ile Tahmin

EM algoritması, sansürlü veri problemlerinde parametre tahminini hesaplamak için en çok olabilirlik tahminlerini içeren tekrarlamalı bir yöntemdir. EM algoritması parametre tahminini gerçekleştirirken, gözlenen ve sansürlü verilere göre oluşturulan olabilirlik fonksiyonundan yararlanır [7].

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 örneklerde kayıp veri ya da eksik veri sorunlarını yaşayan örnekler olduğundan, burada bilinmeyen α ve λ parametrelerinin MLE' sini elde edebilmek için EM yönteminden yararlanılacaktır. $\mathbf{X}=(X_{(1)},\dots,X_{(m)})$ ve $\mathbf{Z}=(Z_1,\dots,Z_m)$ sırasıyla gözlenen ve sansürlü veriler olmak üzere, burada her bir Z_j , $j=1,2,\dots,m$ için $Z_j=(Z_{j1},Z_{j2},\dots,Z_{jR_j})$ ile $1 \times R_j$ vektördür ve gözlemlenemez. Sansürlü veri \mathbf{Z} , kayıp veri olarak düşünülebilir. $\mathbf{W}=(\mathbf{X},\mathbf{Z})$ kombinasyonu ise tam veri setini oluşturmaktadır. Verilerin tamamına dayanarak oluşturulan olasılık fonksiyonunun parametre tahmini Log-olabilirlik fonksiyonunun beklenen değerinin parametrelere göre türevleri alınır ve maksimize edilir. Katkı sabitini göz ardı edecek olursak, $L_c(\mathbf{W};\alpha,\lambda)$ sansürlü verilerin log-olabilirlik fonksiyonu Eşitlik 9'da verildiği gibi elde edilir.

$$l_c(\mathbf{W};\alpha,\lambda) = n \ln \alpha + n \ln \lambda - \lambda \sum_{i=1}^m x_{(i)} + (\alpha - 1) \sum_{i=1}^m \ln(1 - e^{-\lambda x_{(i)}}) - \lambda \sum_{j=1}^m \sum_{k=1}^{R_j} (z_{jk}) + (\alpha - 1) \sum_{j=1}^m \sum_{k=1}^{R_j} \ln(1 - e^{-\lambda z_{jk}}) \quad (9)$$

Genelleştirilmiş üstel dağılımdan gelen ilerleyen tür tip II sağdan sansürlü örnek için EM algoritmasının beklenen ve maksimizasyon adımları şu şekildedir:

E-adımı: Bu adımda öncelikle log-olabilirlik $l_c(\mathbf{W};\alpha,\lambda)$ fonksiyonu hesaplanmalıdır. $l_c(\mathbf{W};\alpha,\lambda)$ 'dan $g(z_{jk})$ gibi z_{jk} 'nın herhangi bir fonksiyonu $E\{g(Z_{jk}) | Z_{jk} > x_{(j)}\}$ ile değiştirilerek elde edilir. Bu nedenle log-olabilirlik fonksiyonunun beklenen değeri Eşitlik 10'da verildiği gibi ifade edilmektedir.

$$n \ln \alpha + n \ln \lambda - \lambda \sum_{i=1}^m x_{(i)} + (\alpha - 1) \sum_{i=1}^m \ln(1 - e^{-\lambda x_{(i)}}) - \lambda \sum_{j=1}^m \sum_{k=1}^{R_j} E[Z_{jk} | z_{jk} > x_{(j)}] + (\alpha - 1) \sum_{j=1}^m \sum_{k=1}^{R_j} E[\ln(1 - e^{-\lambda z_{jk}}) | z_{jk} > x_{(j)}] \quad (10)$$

Burada $E(Z_{jk} | z_{jk} > x_{(j)})$ 'i bulmak için Teorem 1'e ihtiyaç vardır.

Teorem 1:

$f_{z|x}(z_j | X_{(1)} = x_{(1)}, \dots, X_{(j)} = x_{(j)}) = f_{z|x}(z_j | X_{(j)} = x_{(j)}) = \frac{f(z_j | \alpha, \lambda)}{1 - [F(x_{(j)} | \alpha, \lambda)]}$ $X_{(1)} = x_{(1)}, \dots, X_{(j)} = x_{(j)}$ olmak üzere $k=1, \dots, R_j$, $z_j > x_{(j)}$ ve diğer durumlarda 0 için Z_{jk} 'nin şartlı dağılımıdır.

Teoremin ispatı ile ilgili olarak detaylı bilgi için Ng ve diğerlerinin 2002 yılında yapmış olduğu çalışmadan yararlanılabilir [24].

Teorem 1'i kullanarak Eşitlik 11'de verildiği gibi ifade edilir.

$$\underbrace{E(z_{jk} | z_{jk} > c)}_{A(c, \alpha, \lambda)} = \frac{\alpha \lambda}{1 - F(c | \alpha, \lambda)} \int_c^{\infty} x e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha-1} dx$$

Burada $\lambda_x = y$, λ_x 'in yerine y yazılır.

$$= \frac{\alpha}{\lambda(1-F(c|\alpha,\lambda))} x \int_{c/\lambda}^{\infty} y.e^{-y} (1-e^{-y})^{\alpha-1} dy$$

Burada e^{-y} yerine z yazılır.

$$= -\frac{\alpha}{\lambda(1-F(c|\alpha,\lambda))} x \int_0^{e^{-c/\lambda}} \ln z(1-z)^{\alpha-1} dz \tag{11}$$

ve

$$E\left[\underbrace{\ln(1-e^{-\lambda z_{jk}})}_{B(c,\alpha,\lambda)} | z_{jk} > c\right] = \frac{\alpha\lambda}{1-F(c|\alpha,\lambda)} x \int_c^{\infty} \ln(1-e^{-\lambda x}) e^{-\lambda x} (1-e^{-\lambda x})^{\alpha-1} dx$$

$$B(c,\alpha,\lambda) = \frac{1}{\alpha(1-F(c|\alpha,\lambda))} x \int_{(1-e^{-c/\lambda})^\alpha}^1 \ln y dy$$

y 'nin yerine $y=(1-e^{-\lambda x})^\alpha$ yazıldığında Eşitlik 12’de verildiği gibi elde edilir.

$$= \frac{1}{\alpha(1-F(c|\alpha,\lambda))} x \left[(1-e^{-c/\lambda})^\alpha (1-\alpha \ln(1-e^{-c/\lambda})) - 1 \right] \tag{12}$$

M adımı: Eşitlik 11 ve eşitlik 12’deki karşılık gelen değerlerinin Eşitlik 10’da değiştirilmesini ve Eşitlik 10’da verilen log-olabilirlik fonksiyonunun ise maksimize edilmesini içerir. Bu nedenle, (α,λ) 'nın tahmin edicileri $(\alpha^{(k)},\lambda^{(k)})$ k . aşamasında ise eğer α ve λ 'ya göre maksimize edilerek $(\alpha^{(k+1)},\lambda^{(k+1)})$ elde edilir.

$$L_c(\mathbf{W};\alpha,\lambda) = n \ln \alpha + n \ln \lambda - \lambda \sum_{i=1}^m x_{(i)} + (\alpha-1) \sum_{i=1}^m \ln(1-e^{-\lambda x_{(i)}}) - \lambda \sum_{j=1}^m R_j A(x_{(j)},\alpha^k,\lambda^k) + (\alpha-1) \sum_{j=1}^m R_j B(x_{(j)},\alpha^{(k)},\lambda^{(k)}) \tag{13}$$

Eşitlik 13’ün maksimizasyonu, Gupta ve Kundu (2001) tarafından önerilen benzer yöntemle oldukça etkili bir şekilde elde edilir. İlk olarak, fixed-point denklemi çözümlenerek $\lambda^{(k+1)}$ bulunur.

$$h(\lambda) = \lambda, \tag{14}$$

Burada $h(\lambda)$ fonksiyonu Eşitlik 15’de verildiği gibi tanımlanmaktadır.

$$h(\lambda) = \left[\frac{1}{n} \sum_{j=1}^m x_{(j)} + \frac{1}{n} \tilde{A} - \frac{1}{n} (\hat{\alpha}(\lambda) - 1) \sum_{j=1}^m \frac{x_{(j)} e^{-\lambda x_{(j)}}}{1 - e^{-\lambda x_{(j)}}} \right]^{-1} \tag{15}$$

Burada \tilde{A} , \tilde{B} ve $\hat{\alpha}(\lambda)$ sırasıyla aşağıda verildiği gibi ifade edilmektedir.

$$\tilde{A} = \sum_{j=1}^m R_j A(x_{(j)},\alpha^{(k)},\lambda^{(k)}),$$

$$\tilde{B} = \sum_{j=1}^m R_j B(x_{(j)},\alpha^{(k)},\lambda^{(k)}),$$

$$\hat{\alpha}(\lambda) = -\frac{n}{\sum_{j=1}^m \ln(1 - e^{-\lambda x_{(j)}}) + \tilde{B}} .$$

$\lambda^{(k+1)}$ elde edildiğinde $\alpha^{(k+1)}$, $\alpha^{(k+1)} = \hat{\alpha}(\lambda^{(k+1)})$ olarak elde edilir. Bu nedenle, k . tekrardan $(k+1)$. tekrara ilerlemek için aşağıda verilen algoritma kullanılmaktadır.

Adım 1: Eşitlik 14’ü kullanarak eşitlik 13 maksimize edilir. Böylece yakınsayana kadar süreç devam eder. $(k+1)$. aşamada, eşitlik 13’ü maksimize eden λ ’nın değeri $\lambda^{(k+1)}$ ’dir.

Adım 2: $\alpha^{(k+1)} = -\frac{n}{\sum_{j=1}^m \ln(1 - e^{-\lambda^{(k+1)} x_{(j)}}) + \tilde{B}}$ atanır.

Adım 3: $(\alpha^{(k+1)}, \lambda^{(k+1)})$ ’nin yakınsaması kontrol edilir. Eğer yakınsama gerçekleşirse tekrarlama durdurulur. Aksi takdirde 1. Adıma geri dönülür [25].

2.4. Fisher Bilgi Matrisi

Gözlenen ve beklenen Fisher Bilgi matrislerinin hesaplanması için Louis'in (1982) eksik bilgi ilkesi kullanılmıştır. Gözlenen Fisher bilgi matrisi, asimptotik güven aralıkları oluşturmak için kullanılırken, en uygun sansür planlarının yapılması için beklenen Fisher bilgi matrisi kullanılır [25]. Louis'in (1982) eksik bilgi ilkesi fikri şu şekilde ifade edilir:

$$\text{Gözlenen Bilgi} = \text{Tam Bilgi} - \text{Kayıp Bilgi}$$

Burada verinin tamamından elde edilen bilgiyi tam bilgi olarak ifade edersek ve $\theta = (\alpha, \lambda)$ ile gösterirsek X : gözlenen veri, W : tam veri, $I_W(\theta)$: tam bilgi, $I_X(\theta)$: gözlenen bilgi ve $I_{W|X}(\theta)$: kayıp bilgidir. Eşitlik 16’da verildiği gibi ifade edilir.

$$I_X(\theta) = I_W(\theta) - I_{W|X}(\theta) \quad (16)$$

Tam bilgi $I_W(\theta)$ ise Eşitlik 17’de verildiği gibi ifade edilir.

$$I_W(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} L_c(\mathbf{W}; \theta) \right] \quad (17)$$

j . başarısızlık zamanında $x_{(j)}$ gözleminden elde edilen Fisher bilgi matrisi eşitlik 18 ile hesaplanır.

$$I_{W|X}^{(j)}(\theta) = -E_{z_j | x_{(j)}} \left[\frac{\partial^2}{\partial \theta^2} \ln f_{z_j}(z_j | x_{(j)}, \theta) \right] \quad (18)$$

Bu nedenle, X kayıp bilgi koşulunda W ’nin dağılımı için beklenen bilgi matrisi eşitlik 19’da verildiği gibi ifade edilir.

$$I_{W|X}(\theta) = \sum_{j=1}^m R_j I_{W|X}^{(j)}(\theta) . \quad (19)$$

Böylece gözlenen bilgiler eşitlik 16 'dan elde edilir. $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})$ ’nın asimptotik varyans kovaryans matrisi, $I_X(\hat{\theta})$ tersine çevrilerek elde edilir.

$$\Sigma_{[Var-Cov]} = I_X^{-1} = [I_W(\theta) - I_{W|X}(\theta)]^{-1}$$

Olabilirlik fonksiyonu ve soldan budanmış olasılık fonksiyonu kullanılarak kayıp bilgi matrisi aşağıdaki gibi elde edilir.

$$f(z_{jk} | z_{jk} > x_j; \theta) = \prod_{j=1}^m \prod_{k=1}^{R_j} \frac{f(z_{jk})}{1 - F(x_j)}$$

$$\frac{\partial^2}{\partial \theta^2} \ln E(f(z_{jk} | z_{jk} > x_j; \theta)) = E \left[\frac{\partial^2}{\partial \theta^2} \ln(f(z_{jk} | z_{jk} > x_j)) \right]$$

$$\begin{bmatrix} Var(\hat{\alpha}) & Cov(\hat{\alpha}, \hat{\lambda}) \\ Cov(\hat{\alpha}, \hat{\lambda}) & Var(\hat{\lambda}) \end{bmatrix} = I^{-1}(\theta) = [I_W(\theta) - I_{W|X}(\theta)]^{-1}$$

Genelleştirilmiş üstel dağılımdan gelen ilerleyen tür tip II sağdan sansürlü örneklem verilerinin tamamı için elde edilen log-olabilirlik fonksiyonu eşitlik 20’de verildiği gibi elde edilir.

$$f(X | \alpha, \lambda) = \alpha \lambda e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha-1} \tag{20}$$

$$L_W(X | \alpha, \lambda) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \alpha \lambda e^{-\lambda x_i} (1 - e^{-\lambda x_i})^{\alpha-1}$$

$$\ln L_W(X | \alpha, \lambda) = \sum_{i=1}^n [\ln \alpha + \ln \lambda - \lambda x_i + (\alpha - 1) \ln(1 - e^{-\lambda x_i})]$$

$$= \sum_{i=1}^n \ln \alpha + \sum_{i=1}^n \ln \lambda - \lambda \sum_{i=1}^n x_i + (\alpha - 1) \sum_{i=1}^n \ln(1 - e^{-\lambda x_i})$$

Verilerin tamamından elde edilen, tam bilgi matrisinin elemanlarını hesaplayan tam veri kayıp-olabilirlik fonksiyonunun ikinci kısmının beklenen değerleri aşağıdaki gibi verilir.

$$Var(\hat{\alpha}) = E \left[\frac{\partial^2 \ln L_W(X) \alpha, \lambda}{\partial \alpha^2} \right],$$

$$Cov(\hat{\alpha}, \hat{\lambda}) = E \left[\frac{\partial^2 \ln L_W(X | \alpha, \lambda)}{\partial \alpha \partial \lambda} \right],$$

$$Var(\hat{\lambda}) = E \left[\frac{\partial^2 \ln L_W(X | \alpha, \lambda)}{\partial \lambda^2} \right].$$

Böylece $I_W(\alpha)$ ’nin $(i, j)^{th}$ elemanını $a_{ij}(\alpha, \lambda)$ ile belirtirsek, aşağıdaki eşitliklerde verildiği gibi elde edilir.

$$E[a_{mm}] = -\frac{n}{\alpha^2}$$

$$\frac{\partial \ln L_W(X | \alpha, \lambda)}{\partial \lambda} = \frac{n}{\lambda} + \left[\sum_{i=1}^n \left[\frac{X_i e^{(-\lambda x_i)} \alpha}{1 - e^{(-\lambda x_i)}} - \frac{X_i e^{(-\lambda x_i)}}{1 - e^{(-\lambda x_i)}} - X_i \right] \right]$$

$$\frac{\partial^2 \ln L_W(X | \alpha, \lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2} + \left[\sum_{i=1}^n \left[\frac{-x_i^2 e^{(-\lambda x_i)} \alpha}{1 - e^{(-\lambda x_i)}} - \frac{x_i^2 (e^{-\lambda x_i})^2 \alpha}{(1 - e^{-\lambda x_i})^2} + \frac{x_i^2 e^{-\lambda x_i}}{1 - e^{-\lambda x_i}} + \frac{x_i^2 (e^{-\lambda x_i})^2}{(1 - e^{-\lambda x_i})^2} \right] \right]$$

$$\frac{\partial^2 \ln L_W(X | \alpha, \lambda)}{\partial \alpha \partial \lambda} = \sum_{i=1}^n \frac{x_i e^{(-\lambda x_i)}}{1 - e^{-\lambda x_i}}$$

$$a_{11} = \frac{n}{\alpha^2}$$

$$a_{12} = a_{21} = -\frac{n}{\lambda} \left[\frac{\alpha}{\alpha-1} (\psi(\alpha) - \psi(1)) - (\psi(\alpha+1) - \psi(1)) \right], \alpha > 2$$

$$= -\frac{n\alpha}{\lambda} \int_0^{\infty} x e^{-2x} (1 - e^{-x})^{\alpha-2} dx, 0 < \alpha \leq 2,$$

$$a_{22} = \frac{n}{\lambda^2} \left[1 + \frac{\alpha(\alpha-1)}{\alpha-2} (\psi'(1) - \psi'(\alpha-1) + (\psi(\alpha-1) - \psi(1))^2) \right]$$

$$+ \frac{n\alpha}{\lambda^2} [(\psi'(1) - \psi(\alpha)) + (\psi(\alpha) - \psi(1))^2], \alpha > 2,$$

$$= \frac{n}{\lambda^2} + \frac{n\alpha(\alpha-1)}{\lambda^2} \int_0^{\infty} x^2 e^{-2x} (1 - e^{-x})^{\alpha-3} dx, 0 < \alpha \leq 2$$

Burada ψ ve ψ' Digamma ve Trigamma fonksiyonlarıdır [27]. Kayıp bilgi matrisi $I_{w|X}^{(j)}(X|\alpha, \lambda)$ aşağıdaki gibi elde edilir.

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

Burada b_{11} , b_{22} ve b_{12} aşağıdaki eşitliklerde verildiği gibi elde edilmektedir.

$$b_{11}(x_{(j)}; \alpha, \lambda) = \frac{1}{\alpha^2} \left[\ln(1 - e^{-\lambda x_{(j)}}) \right]^2 \frac{(1 - e^{-\lambda x_{(j)}})^{\alpha}}{(1 - (1 - e^{-\lambda x_{(j)}})^{\alpha})^2}$$

$$b_{22}(x_{(j)}; \alpha, \lambda) = \frac{1}{\lambda^2} + (\alpha - 1) h_1(x_{(j)}; \alpha, \lambda) - \frac{\alpha x_{(j)}^2 e^{-\lambda x_{(j)}} (1 - e^{-\lambda x_{(j)}})^{\alpha-2}}{(1 - (1 - e^{-\lambda x_{(j)}})^{\alpha})^2} x \left[\alpha e^{-\lambda x_{(j)}} - 1 + (1 - e^{-\lambda x_{(j)}})^{\alpha} \right]$$

$$b_{12}(x_{(j)}; \alpha, \lambda) = -h_2(x_{(j)}; \alpha, \lambda) + \frac{x_{(j)} e^{-\lambda x_{(j)}} (1 - e^{-\lambda x_{(j)}})^{\alpha-1}}{(1 - (1 - e^{-\lambda x_{(j)}})^{\alpha})^2} x \left[1 + \alpha \ln(1 - e^{-\lambda x_{(j)}}) - (1 - e^{-\lambda x_{(j)}})^{\alpha} \right]$$

Burada,

$$h_1(x_{(j)}; \alpha, \lambda) = \frac{1}{\lambda^2 (1 - (1 - e^{-\lambda x_{(j)}})^{\alpha})} \int_{(1 - e^{-\lambda x_{(j)}})^{\alpha}}^1 (\ln(1 - u^{1/\alpha}))^2 (1 - u^{1/\alpha}) u^{-2/\alpha} du$$

$$h_2(x_{(j)}; \alpha, \lambda) = \frac{1}{\lambda (1 - (1 - e^{-\lambda x_{(j)}})^{\alpha})} \int_{(1 - e^{-\lambda x_{(j)}})^{\alpha}}^1 (-\ln(1 - u^{1/\alpha})) (1 - u^{1/\alpha}) u^{-1/\alpha} du$$

olarak verilmektedir. Daha sonra olasılık fonksiyonu kullanarak kayıp bilgi matrisi aşağıdaki gibi verilir. Soldan budanmış veriler için olasılık fonksiyonu aşağıdaki gibi elde edilir.

$$f_{z_{jk}}(z_{jk} | z_{jk} > x_{j:m:n}, \alpha, \lambda) = \prod_{j=1}^m \prod_{k=1}^{R_j} \frac{f(z_{jk})}{1 - F(x_{j:m:n})}$$

Verilen $x_{j:m:n}$, z_{jk} nin log-olabilirlik fonksiyonu olarak ifade edilebilir.

$$\begin{aligned} \ln \left[f_{z_{jk}} \left(z_{jk} \mid z_{jk} > x_{j:m:n}, \alpha, \lambda \right) \right] &= \sum_{j=1}^m \sum_{k=1}^{R_j} \ln \left(f \left(z_{jk} \right) - \ln \left(1 - F \left(x_{j:m} \right) \right) \right) \\ &= \sum_{j=1}^m \sum_{k=1}^{R_j} \ln \left[\alpha \lambda e^{-\lambda z_{jk}} \left(1 - e^{-\lambda z_{jk}} \right)^{\alpha-1} \right] - \ln \left[1 - \left(1 - e^{-\lambda x_j} \right)^\alpha \right] \end{aligned}$$

Verilen $x_{j:m:n}$, z_{jk} nin log-olabilirlik fonksiyonunun ikinci kısmi beklenen değerleri hesaplanır. Aşamalı olarak sansürlenmiş veriler için beklenen Fisher bilgi matrisi ele alındığında, $j=1, \dots, m$ için $X_{(j)}$ 'nin olasılık yoğunluk fonksiyonu eşitlik 21'deki gibi verilmektedir.

$$f_{X_{(j)}}(x) = c_{j-1} \sum_{i=1}^j a_{i,j} \left(1 - \left(1 - e^{-\lambda x} \right)^\alpha \right)^{r_i-1} \alpha \lambda e^{-\lambda x} \left(1 - e^{-\lambda x} \right)^{\alpha-1} \quad (21)$$

$x > 0$ ve aksi takdirde 0 için.

Burada $j=1, \dots, m$ için $r_j = m - j + 1 + \sum_{i=j}^m R_i$ ve $c_{j-1} = \prod_{i=1}^j r_i$ olarak ifade edilmektedir.

$1 \leq i \leq j \leq m$ için a_{i1} ve $a_{i,j}$ eşitlikleri aşağıdaki gibidir [22].

$$a_{i1} = 1 \text{ ve } a_{i,j} = \prod_{k=1, k \neq i}^j \frac{1}{r_k - r_i}$$

Eşitlik 21'e dayanarak 2×2 'lik beklenen Fisher bilgi matrisini E ile ifade edersek, eşitlik 22'de verildiği gibi elde edilir.

$$E = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} = -E \begin{bmatrix} \frac{\partial^2 \ln l(\alpha, \lambda)}{\partial \alpha^2} & \frac{\partial^2 \ln l(\alpha, \lambda)}{\partial \alpha \partial \lambda} \\ \frac{\partial^2 \ln l(\alpha, \lambda)}{\partial \lambda \partial \alpha} & \frac{\partial^2 \ln l(\alpha, \lambda)}{\partial \lambda^2} \end{bmatrix} \quad (22)$$

Burada E_{11} , E_{22} , E_{12} ve E_{21} eşitlikleri aşağıdaki gibi verilmektedir.

$$\begin{aligned} E_{11} &= \frac{1}{\alpha^2} \left[m + \sum_{j=1}^m R_j h_{1j}(\alpha, \lambda) \right], \\ E_{22} &= \frac{1}{\lambda^2} \left[m + \alpha(\alpha-1) \sum_{j=1}^m h_{2j}(\alpha, \lambda) + \alpha^2 \sum_{j=1}^m h_{3j}(\alpha, \lambda) \right], \\ E_{12} = E_{21} &= -\frac{\alpha}{\lambda} \left[\sum_{j=1}^m h_{4j}(\alpha, \lambda) - \sum_{j=1}^m R_j h_{5j}(\alpha, \lambda) - \alpha \sum_{j=1}^m R_j h_{6j}(\alpha, \lambda) \right], \end{aligned}$$

Burada,

$$\begin{aligned} h_1(\alpha, \lambda) &= c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u (\ln u)^2 (1-u)^{r_i-3} du, \\ h_2(\alpha, \lambda) &= c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u^{\alpha-3} (1-u)^{r_i-1} (1-u) (\ln(1-u))^2 du, \\ h_3(\alpha, \lambda) &= c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u^{2\alpha-3} (1-u)^{r_i-3} (1-u) (\ln(1-u))^2 x (\alpha u - u^\alpha - \alpha + 1) du \\ h_4(\alpha, \lambda) &= c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u^{\alpha-2} (1-u)^{r_i-3} (1-u) (-\ln(1-u)) du, \end{aligned}$$

$$h_{5j}(\alpha, \lambda) = c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u^{2\alpha-2} (1-u^\alpha)^{i-2} (1-u)(-\ln(1-u)) du,$$

$$h_{6j}(\alpha, \lambda) = c_{j-1} \sum_{i=1}^j a_{i,j} \int_0^1 u^{2\alpha-2} (1-u^\alpha)^{i-3} (1-u)(-\ln(1-u)) \ln u du,$$

olarak elde edilir. (α, λ) 'nın MLE'lerinin asimptotik varyans kovaryans matrisi eşitlik 23'deki gibi elde edilecektir.

$$V = E^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \quad (23)$$

3. UYGULAMA

3.1 Benzetim Çalışması ve Sonuçları

Bu alt bölümde, önerilen parametre tahmin edicilerinin performanslarını karşılaştırmak amacıyla Monte Carlo benzetim çalışması R 3.1.3 programı kullanılarak yapılmıştır. İlerleyen tür tip II sağdan sansürlü örneklem Balakrishnan ve Aggarwala'ın (2000) çalışması dikkate alınarak Genelleştirilmiş Üstel Dağılımı için farklı parametre büyüklüklerinde, farklı örneklem genişliklerinde ve çeşitli sansürleme şemaları altında oluşturulmuştur. Her aşamada α ve λ 'nın en çok olabilirlik tahmin edicileri 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 λ parametrelerinin sadece tahmin değerleri değil aynı zamanda her bir tahmine ilişkin hata kareler ortalaması değerleri ve %95 güven düzeyinde Fisher bilgi matrisi kullanılarak parametrelere ilişkin güven aralıkları (CI) oluşturulmuştur. Sonuçların yer aldığı tablolarda verilen CI'lar, 1000 tekrarlı yapılan simülasyon çalışmasında elde edilen güven aralıklarının uç noktalarının ortalaması olarak hesaplanmıştır. Benzetim çalışması için seçilen α ve λ parametre değerlerinden elde edilen farklı kombinasyonlar aşağıda verilen Tablo 1'de görüldüğü gibidir.

Tablo 1. Benzetim Çalışması için (α, λ) Parametre Değerleri Tablosu

		λ		
		1	1.5	2.5
α	0.5	(0.5,1)	(0.5,1.5)	(0.5,2.5)
	1	(1,1)	(1,1.5)	(1,2.5)
	1.5	(1.5,1)	(1.5,1.5)	(1.5,2.5)

Benzetim çalışmasında, her bir (α, λ) parametre çifti için, rasgele olarak türetilen ilerleyen tür tip II sağdan sansürlü örneklem büyüklükleri değerleri ise $n = 20, 30, 50$ olarak seçilmiştir. Benzetim çalışmasında genelleştirilmiş üstel dağılımın ilerleyen tür tip II sağdan sansür örneklem şemaları ise Tablo 2'de verilmiştir.

Tablo 2. Örneklem Şemaları

n	m	R (Sansürleme Şeması)
20	10	$(10, 0^{*9})$
30	20	$(10, 0^{*19})$
50	30	$(20, 0^{*29})$

Tablo 2’de verilen sansürleme şemasında:

- $R = (10, 0^{*9})$ gösterimi 1.başarısızlık gerçekleştiğinde örneklemden rasgele olarak $R_1 = 10$ gözlemin çekildiğini, x_2, \dots, x_{10} başarısızlıkları gerçekleştiğinde ise $R_2 = 0, \dots, R_{10} = 0$ gösterimi ile de örneklemden rasgele gözlemlerin seçilmediğini,
- $R = (10, 0^{*19})$ gösterimi 1.başarısızlık gerçekleştiğinde örneklemden rasgele olarak $R_1 = 10$ gözlemin çekildiğini, x_2, \dots, x_{20} başarısızlıkları gerçekleştiğinde ise $R_2 = 0, \dots, R_{20} = 0$ gösterimi ile de örneklemden rasgele gözlemlerin seçilmediğini ve,
- $R = (20, 0^{*29})$ gösterimi 1.başarısızlık gerçekleştiğinde örneklemden rasgele olarak $R_1 = 20$ gözlemin çekildiğini, x_2, \dots, x_{30} başarısızlıkları gerçekleştiğinde ise $R_2 = 0, \dots, R_{30} = 0$ gösterimi ile de örneklemden rasgele gözlemlerin seçilmediğini ifade etmektedir.

Benzetim çalışması sonucunda parametre tahminleri için kullanılan Newton Raphson yöntemi ve EM algoritmasından elde edilen sonuçlar Tablo 3’den Tablo 11’e kadar verilmiştir.

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

$\alpha = 0.5, \lambda = 1$			NR		EM	
n	m	R	$\hat{\alpha}(\text{MSE})$	$\hat{\lambda}(\text{MSE})$	$\hat{\alpha}(\text{MSE})$	$\hat{\lambda}(\text{MSE})$
20	10	$(10, 0^{*9})$	0.5671(0.5421)	1.3627(0.4891)	0.4563(0.2110)	0.8989(0.1706)
		CI	(0.1770-0.9571)	(0.5163-2.2090)	(0.0662-0.8463)	(0.0525-1.7452)
30	20	$(10, 0^{*19})$	0.5193(0.4118)	1.1388(0.3610)	0.4786(0.1321)	0.9357(0.0975)
		CI	(0.2694-0.7691)	(0.3528-1.9247)	(0.2287-0.7284)	(0.1497-1.7216)
50	30	$(20, 0^{*29})$	0.5284(0.2712)	1.0993(0.1422)	0.5021(0.08325)	1.0641(0.0602)
		CI	(0.3285-0.7282)	(0.5008-1.6977)	(0.3022-0.7019)	(0.4656-1.6625)

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

$\alpha = 0.5, \lambda = 1.5$			NR		EM	
n	m	R	$\hat{\alpha}(\text{MSE})$	$\hat{\lambda}(\text{MSE})$	$\hat{\alpha}(\text{MSE})$	$\hat{\lambda}(\text{MSE})$
20	10	$(10, 0^{*9})$	0.6793(0.5104)	1.4923(0.5470)	0.4507(0.2064)	0.9216(0.2215)
		CI	(0.3942-0.9643)	(0.1686-2.8159)	(0.1656-0.7357)	(-0.4020-2.2452)
30	20	$(10, 0^{*19})$	0.6418(0.3926)	1.2898(0.4775)	0.4978(0.1810)	0.9637(0.1721)
		CI	(0.4080-0.8755)	(0.3871-2.1924)	(0.2640-0.7315)	(0.0610-1.8663)
50	30	$(20, 0^{*29})$	0.4151(0.4216)	1.1208(0.4951)	0.5138(0.1255)	1.0926(0.1139)
		CI	(0.2623-0.5678)	(0.4360-1.8055)	(0.3610-0.6665)	(0.4078-1.7773)

Tablo 5. $\alpha = 0.5$, $\lambda = 2.5$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 0.5, \lambda = 2.5$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	0.6218(0.4918)	3.2042(0.4617)	0.4542(0.2018)	2.6121(0.2357)
		CI	(0.3866-0.8569)	(2.7730-3.6353)	(0.2190-0.6893)	(2.1809-3.0432)
30	20	$(10, 0^{*19})$	0.6101(0.4217)	2.8560(0.3967)	0.5725(0.2025)	2.5632(0.1895)
		CI	(0.3101-0.9100)	(1.1945-4.5174)	(0.2287-0.7284)	(0.9017-4.2246)
50	30	$(20, 0^{*29})$	0.5647(0.2817)	2.7106(0.1609)	0.5086(0.0105)	2.5102(0.0917)
		CI	(0.3873-0.7420)	(1.6130-3.8081)	(0.3312-0.6859)	(1.4126-3.6077)

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

$\alpha = 1, \lambda = 1$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	0.8045(0.6219)	1.7160(0.4926)	1.1947(0.2815)	1.3974(0.1960)
		CI	(1.0407-2.5682)	(1.1741-2.2578)	(0.4309-1.9584)	(0.8555-1.9392)
30	20	$(10, 0^{*19})$	1.4472(0.5009)	1.4913(0.3795)	1.1228(0.2310)	1.2394(0.0768)
		CI	(0.7702-2.1241)	(0.8996-2.0829)	(0.4458-1.7997)	(0.6477-1.8310)
50	30	$(20, 0^{*29})$	1.1187(0.3113)	1.4317(0.2455)	1.0466(0.0807)	1.1107(0.0652)
		CI	(0.6493-1.5880)	(0.9283-1.9350)	(0.5772-1.5159)	(0.6073-1.6140)

Tablo 7. $\alpha = 1$, $\lambda = 1.5$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 1, \lambda = 1.5$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	1.5429(0.4718)	1.9496(0.4221)	1.0186(0.1928)	1.8321(0.1755)
		CI	(0.8791-2.2066)	(1.0343-2.8648)	(0.3548-1.6823)	(0.9168-2.7473)
30	20	$(10, 0^{*19})$	1.2715(0.3267)	1.7461(0.2918)	1.0661(0.07423)	1.4554(0.08190)
		CI	(0.8029-1.7400)	(0.8254-2.6667)	(0.5975-1.5346)	(0.5347-2.3760)
50	30	$(20, 0^{*29})$	0.8917(0.2715)	1.4152(0.1750)	0.9858(0.0663)	1.5348(0.0715)
		CI	(0.4174-1.3659)	(0.8407-1.9896)	(0.5115-1.4600)	(0.9603-2.1092)

Tablo 8. $\alpha = 1, \lambda = 2.5$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 1, \lambda = 2.5$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	1.28090(0.6219)	2.9405(0.4856)	0.9386(0.2316)	2.4872(0.2215)
		CI	(0.4581-2.1036)	(0.9931-4.8878)	(0.1158-1.7613)	(0.5398-4.4345)
30	20	$(10, 0^{*19})$	0.9459(0.4310)	2.0320(0.3765)	0.9781(0.1651)	2.2490(0.1340)
		CI	(0.4404-1.4513)	(0.7671-3.2968)	(0.4726-1.4835)	(0.9841-3.5138)
50	30	$(20, 0^{*29})$	0.8733(0.3229)	2.7842(0.1955)	1.09617(0.1248)	2.5203(0.0917)
		CI	(0.4876-1.2589)	(1.5590-4.0093)	(0.7104-1.4817)	(1.2951-3.7454)

Tablo 9. $\alpha = 1.5, \lambda = 1$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 1.5, \lambda = 1$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	2.1574(0.4923)	1.4953(0.5416)	1.5149(0.2710)	0.8508(0.3315)
		CI	(1.0794-3.2353)	(0.9073-2.0832)	(0.4369-2.5928)	(0.2700-1.4459)
30	20	$(10, 0^{*19})$	1.9826(0.3721)	1.4494(0.3945)	1.4262(0.1968)	1.0808(0.2837)
		CI	(1.2770-2.6881)	(0.7158-2.3546)	(0.7206-2.1317)	(0.3948-1.7667)
50	30	$(20, 0^{*29})$	1.4304(0.2945)	1.0576(0.2658)	1.5074(0.1279)	1.0125(0.1163)
		CI	(0.7248-2.1359)	(1.0568-1.9583)	(0.8018-2.2129)	(0.5617-1.4632)

Tablo 10. $\alpha = 1.5, \lambda = 1.5$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 1.5, \lambda = 1.5$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	1.1904(0.4211)	1.0910(0.3827)	1.6440(0.3684)	1.5728(0.2769)
		CI	(0.3648-2.0159)	(0.2583-1.9236)	(0.8184-2.4695)	(0.7401-2.4054)
30	20	$(10, 0^{*19})$	1.6609(0.3806)	1.3156(0.3023)	1.6129(0.2114)	1.5194(0.1859)
		CI	(1.0237-2.2980)	(0.7190-1.9121)	(0.9757-2.2500)	(0.9228-2.1159)
50	30	$(20, 0^{*29})$	1.2624(0.4752)	1.0479(0.2319)	1.5601(0.1225)	1.4936(0.1270)
		CI	(0.6585-1.8662)	(0.42100-1.6747)	(0.9562-2.1639)	(0.8667-2.1204)

Tablo 11. $\alpha = 1.5$, $\lambda = 2.5$ için EM Algoritması ve Newton-Raphson Sonuçları

$\alpha = 1.5, \lambda = 2.5$			NR		EM	
n	m	R	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
20	10	$(10, 0^{*9})$	1.4891(0.3958)	1.9115(0.4219)	1.4408(0.3450)	2.3116(0.2875)
		CI	(0.6087-2.3694)	(0.5618-3.2611)	(0.5604-2.3211)	(0.9619-3.6612)
30	20	$(10, 0^{*19})$	1.6118(0.4120)	2.9318(0.3786)	1.5729(0.2936)	2.4878(0.1927)
		CI	(0.8571-2.3664)	(1.8226-4.0409)	(0.8182-2.3275)	(1.3786-3.5969)
50	30	$(20, 0^{*29})$	1.7220(0.2817)	3.2540(0.5124)	1.4890(0.2104)	2.4640(0.1137)
		CI	(0.9536-2.4903)	(2.1744-4.3335)	(0.7206-2.2573)	(1.3844-3.5435)

Benzetim çalışmasının sonuçlarının verildiği Tablo 3 – Tablo 11 incelendiğinde EM algoritması ile elde edilen parametre sonuçlarının ve MSE değerlerinin Newton-Raphson yönteminden çok daha iyi olduğu görülmektedir. %95 güven düzeyinde EM algoritması ile elde edilen asimptotik güven aralıkları karşılaştırıldığında ise EM aralıklarının Newton-Raphson yöntemine göre daha dar olduğu gözlemlenmiştir. Her iki güven aralığı da bilinmeyen parametre değerlerini içermiş olsa da EM algoritması Newton-Raphson yöntemine göre tercih edilir. Aynı zamanda tablolar incelendiğinde farklı örneklem büyüklükleri ve farklı sansür şemalarına göre oluşturulan benzetim çalışması senaryolarında tüm durumlar için EM yönteminde örneklem genişliği arttıkça daha tutarlı sonuçlar elde edilmiştir. Newton-Raphson yönteminde ise genellikle tutarlı sonuçlar elde edildiği ancak Tablo 4, Tablo 10 ve Tablo 11’de $\hat{\alpha}$ ’ın MSE değerleri için örneklem hacmi n arttığında düşme eğiliminde olmadığı gözlemlenmiştir.

Genelleştirilmiş üstel dağılımının ilerleyen tür tip II sağdan sansürlü örnekleme parametre tahmini için yapılan bu benzetim çalışmasında EM algoritması yönteminden elde edilen sonuçlardan da görüleceği üzere, EM algoritmasının parametre tahmininde Newton-Raphson yöntemine göre daha üstün olduğu görülmektedir.

3.2 Gerçek Yaşam Verisi Uygulaması ve Sonuçları

Bu bölümde genelleştirilmiş üstel dağılımın ilerleyen tür tip II sağdan sansürlü örnekleme için Newton-Raphson yöntemi ve EM algoritmasıyla elde edilen parametre tahminlerini karşılaştırabilmek amacıyla gerçek yaşam verisi üzerinde uygulama yapılmıştır. Bader ve Priest 1982 yılında bir araştırmada, basit bir fiber lifinden, 1000 karbonluk fiber lifi içeren örneklere kadar veri elde etmişlerdir. Bu veri 1, 10, 20, 50 mm’lik fiber lifi çap uzunluklarında incelemiş olup, aynı zamanda veri 20, 50, 150 ve 300 mm uzunluklarında dayanıklılık testine tabi tutmuşlardır. Araştırmacı tarafından bu veri seti Weibull modeli kullanılarak incelenmiş ve modelin iyi çalışmadığı gözlemlenmiştir. [29] ve [30] bu veri setinde genelleştirilmiş Rayleigh yönteminin iyi çalıştığını belirtmişlerdir. Çalışmamızda kullanmış olduğumuz gerçek veri seti [30] tarafından 2005 yılında dayanıklılık ölçümleri 20 mm olacak şekilde dönüştürülmüş veri setinden oluşmaktadır. Veri seti aşağıda görüldüğü gibidir.

0.312, 0.314, 0.479, 0.552, 0.700, 0.803, 0.861, 0.865, 0.944, 0.958, 0.966, 0.997, 1.006, 1.021, 1.027, 1.055, 1.063, 1.098, 1.140, 1.179, 1.224, 1.240, 1.253, 1.270, 1.272, 1.274, 1.301, 1.301, 1.359, 1.382, 1.382, 1.426, 1.434, 1.435, 1.478, 1.490, 1.511, 1.514, 1.535, 1.554, 1.566, 1.570, 1.586, 1.629, 1.633, 1.642, 1.648, 1.684, 1.697, 1.726, 1.770, 1.773, 1.800, 1.809, 1.818, 1.821, 1.848, 1.880, 1.954, 2.012, 2.067, 2.084, 2.090, 2.096, 2.128, 2.233, 2.433, 2.585, 2.585

Üzerinde çalışmış olduğumuz gerçek yaşam verisinin Genelleştirilmiş üstel dağılımdan gelip gelmediğinin araştırılması için R 3.1.3 programında uyum iyiliği test kodları yazılmış olup Gamma, Weibull, Lognormal ve Genelleştirilmiş üstel dağılım karşılaştırılmıştır. Karşılaştırmak amacıyla kullanılan dağılımların olasılık yoğunluk fonksiyonları Tablo 12’de verilmiştir.

Tablo 12. Dağılımların Olasılık Yoğunluk Fonksiyonları

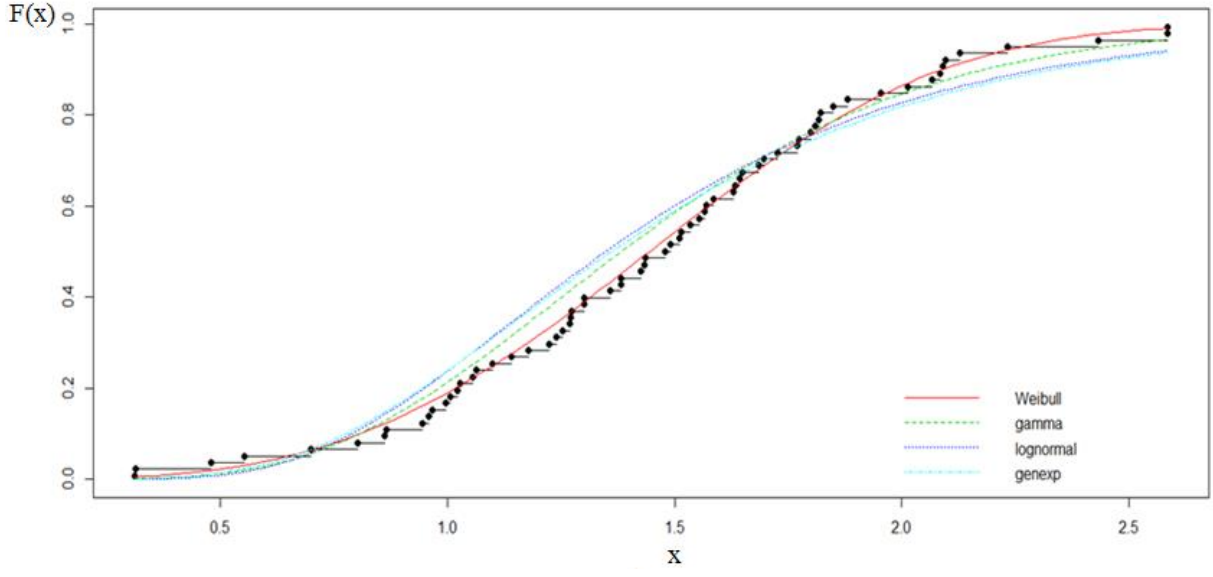
Dağılımlar	Olasılık Yoğunluk Fonksiyonu
$G(\alpha, \beta)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$
$W(\alpha, \beta)$	$\frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$
$\text{Log-N}(\mu, \sigma)$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$
$GED(\alpha, \lambda)$	$f(x; \alpha, \lambda)$

Gamma, Weibull, Log-normal ve Genelleştirilmiş üstel dağılımlarını karşılaştırmak için yapılan Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), Cramer-von Mises (C-M) uyum iyiliği testleri, Akaike (AIC), Bayes (BIC) bilgi kriterleri değerleri ve parametre tahmin değerlerine ait sonuçlar Tablo 13’de verilmiştir.

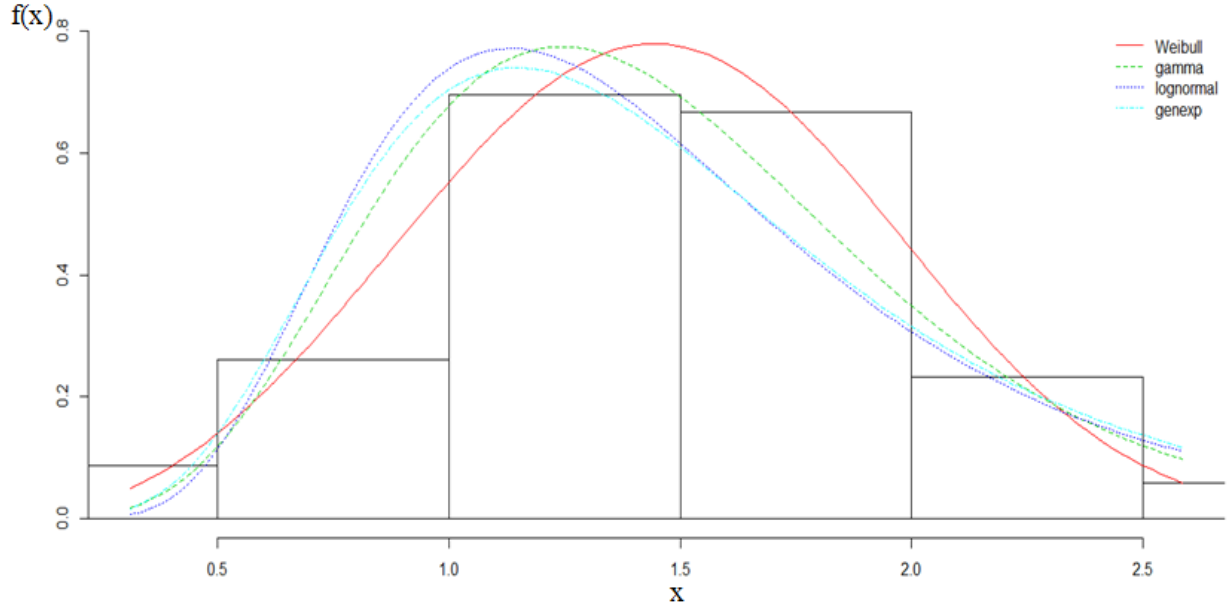
Tablo 13. Dağılımların Uyum İyiliği Test İstatistiği ve Kriter Sonuçları

Model	$\hat{\alpha}$	$\hat{\lambda}$	K-S	A-D	C-M	AIC	BIC
$G(\alpha, \lambda)$	6.9970	4.8209	0.0883	0.8534	0.1233	110.165	114.633
$W(\alpha, \lambda)$	3.2488	1.6171	0.0437	0.2094	0.0219	102.001	106.469
$\text{log-N}(\alpha, \lambda)$	0.2993	0.4180	0.1182	1.6781	0.2501	120.758	125.226
$GED(\alpha, \lambda)$	8.8273	1.8963	0.1118	1.5327	0.2292	117.337	121.805

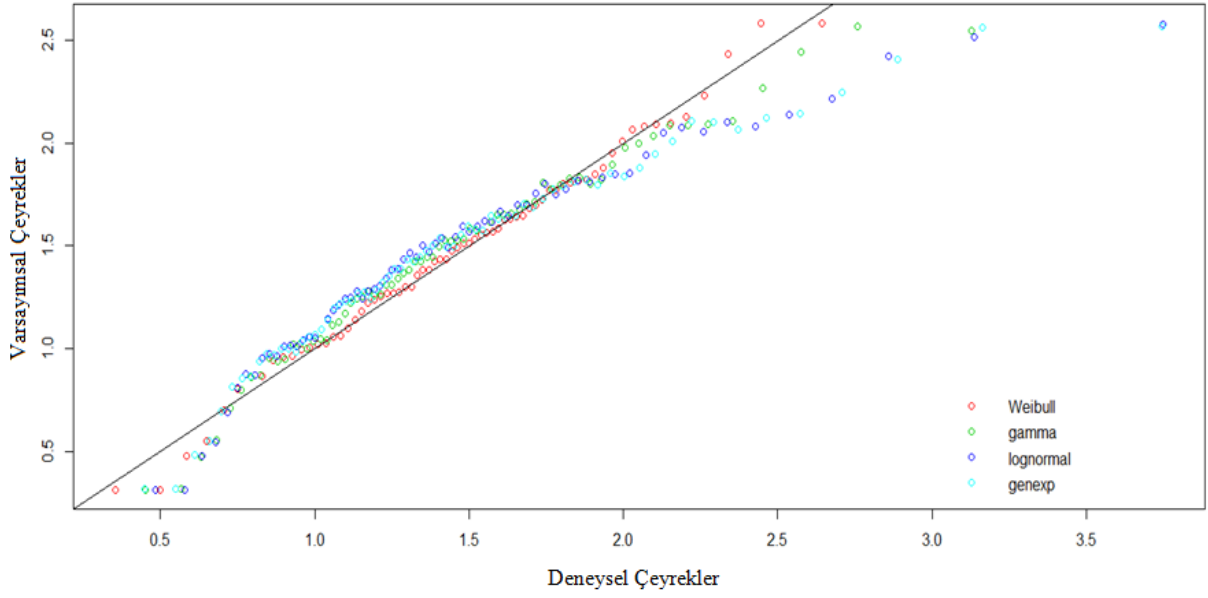
Sonuçlar incelendiğinde sırasıyla Weibull, Gamma ve Genelleştirilmiş üstel dağılımının en küçük uyum iyiliği testlerine ve aynı zamanda en küçük uyum iyiliği kriterleri değerlerine sahip olduğu görülmektedir. Bu çalışmada ise negatif olmayan bir rasgele değişken için çarpık dağılım gereken durumlarda [8, 20, 21] tarafından yeni bir dağılım olarak önerilen, Gamma ve Weibull dağılımlarına alternatif olarak kullanılabilen Genelleştirilmiş üstel dağılım kullanılmıştır [15]. Dağılımları karşılaştırmak amacıyla yapılan birikimli dağılım fonksiyonu, histogram ve yoğunluk fonksiyonları, Q-Q grafikleri ve P-P grafikleri Şekil 3 - Şekil 6’da verilmiştir.



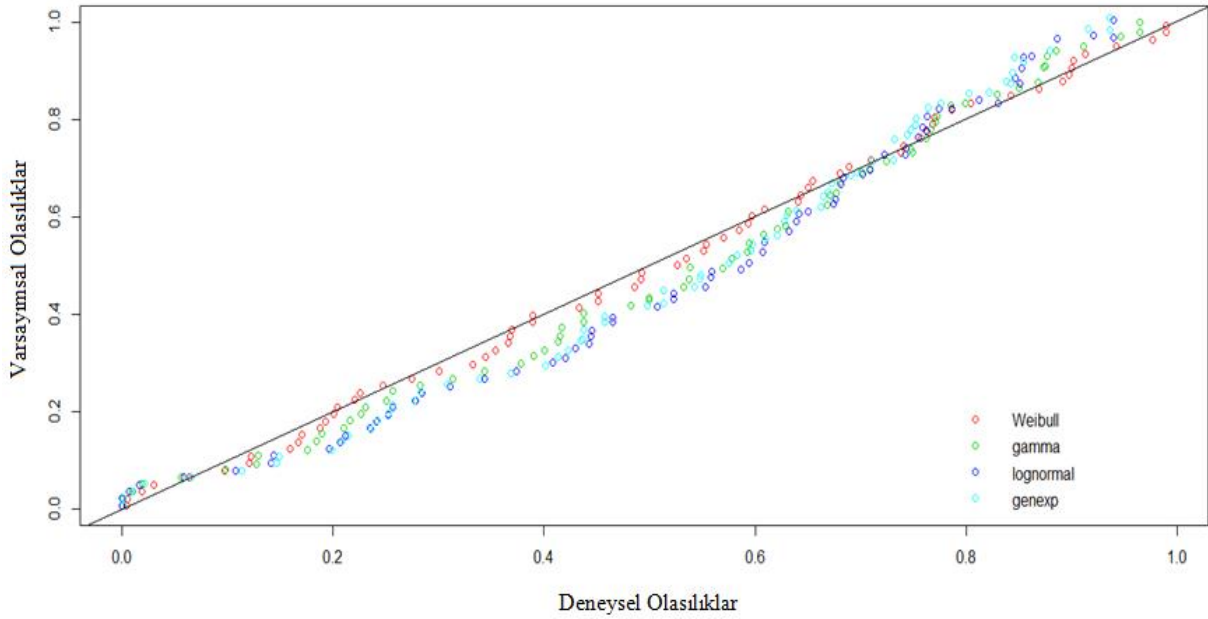
Şekil 3. Weibull, Gamma, Lognormal ve Genelleştirilmiş Üstel Dağılıma İlişkin Deneysel ve Varsayımsal Birikimli Dağılım Fonksiyonu Grafiği



Şekil 4. Weibull, Gamma, Lognormal ve Genelleştirilmiş Üstel Dağılıma İlişkin Histogram ve Olasılık Yoğunluk Fonksiyonu Grafiği



Şekil 5. Weibull, Gamma, Lognormal ve Genelleştirilmiş Üstel Dağılıma İlişkin Q-Q Grafığı



Şekil 6. Weibull, Gamma, Lognormal ve Genelleştirilmiş Üstel Dağılıma İlişkin P-P Grafığı

Şekil 3 – Şekil 6’deki grafikler incelendiğinde veriye en uygun dağılımın sırasıyla Weibull, Gamma ve Genelleştirilmiş üstel dağılım olduğu görülmektedir. Bu çalışmada ise Gamma ve Weibull dağılımlarına alternatif olarak kullanılabilen Genelleştirilmiş üstel dağılım kullanılmıştır. Gerçek yaşam verisinde, Genelleştirilmiş üstel dağılımın veriye uygulanmasına karar verildikten sonra Newton-Raphson yöntemi ve EM algoritmalarının sansürleme şeması altında parametre tahminlerinin karşılaştırmalarını yapabilmek için Tablo 14’ de verilen sansürleme planı kullanılmıştır.

Tablo 14. Sansürleme Şeması

Sansürleme Planı	n	m	R
1	69	40	(29,0 ^{*39})

Burada verilen sansürleme planında, n: örneklem sayısını, m: başarısızlık sayısını, R: sansürleme şemasını göstermektedir. Tablo 14’de verilen sansürleme şeması altında $R=(29,0^{*39})$ gösterimi 1.başarısızlık gerçekleştiğinde örneklemden rasgele olarak $R_1=29$ gözlemin çekildiğini, x_2, \dots, x_{40} başarısızlıkları gerçekleştiğinde ise $R_2=0, \dots, R_{40}=0$ gösterimi ile de örneklemden rasgele gözlemlerin seçilmediğini ifade etmektedir. Tablo 14’de verilen sansürleme planı altında x_2, \dots, x_{40} veriye uygulamasını yapmış olduğumuz Genelleştirilmiş üstel dağılımın ilerleyen tür tip II sağdan sansürlü örneklemelerin Tablo 13’den faydalanarak $\alpha=9, \lambda=2$ olduğu varsayımı altında, Newton-Raphson ve EM algoritmasına ilişkin sonuçlar Tablo 15’ de verilmiştir.

Tablo 15. $\alpha=9$ ve $\lambda=2$ için Newton Raphson ve EM Algoritması Sonuçları

n	m	R	NR		EM	
			$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)	$\hat{\alpha}$ (MSE)	$\hat{\lambda}$ (MSE)
69	40	(29,0 ^{*39})	17.4163 (0.2432)	5.6912 (0.1873)	10.3729 (0.1721)	4.1387 (0.1349)
CI			(16.9397-17.8929)	(5.3241-6.0583)	(10.0356-10.7102)	(3.8743-4.4031)

Tablo 15’ de görüldü gibi $\alpha=9$ ve $\lambda=2$ olduğu durumda, verilen sansürleme şeması altında elde edilen Newton Raphson yöntemi ve EM algoritması tahminlerine bakıldığında EM algoritmasının Newton-Raphson yöntemine göre, parametre değerlerinin daha yakın sonuçlar verdiği söylenebilir.

4. SONUÇ VE ÖNERİLER

Bu çalışmada, Genelleştirilmiş üstel dağılımın ilerleyen tür tip II sağdan sansürlü örnekleme için Newton-Raphson yöntemi ve EM algoritmasıyla elde edilen parametre tahminlerini karşılaştırabilmek amacıyla gerçek yaşam verisi ve benzetim çalışması üzerinde uygulama yapılmıştır. Gerçek yaşam verisinde, Newton-Raphson yöntemi ve EM algoritmalarının $R=(29,0^{*39})$ sansürleme şeması altında parametre tahminleri karşılaştırılmıştır. Sansürleme şeması altında elde edilen Newton-Raphson yöntemi ve EM algoritması tahminlerine bakıldığında EM algoritmasının Newton-Raphson yöntemine göre, parametre değerlerinin daha yakın sonuçlar verdiği gözlemlenmiştir. Benzetim çalışmasında ise: İlerleyen tür tip II sağdan sansürlü örneklemin Genelleştirilmiş Üstel Dağılımı için farklı parametre büyüklüklerinde farklı örneklem genişliklerinde ve çeşitli sansürleme şemaları altında benzetim çalışması yapılmıştır. Analiz sonucu incelendiğinde EM algoritması ile elde edilen parametre sonuçlarının Newton-Raphson yönteminden çok daha iyi olduğu görülmüştür. %95 güven düzeyinde EM algoritması ile elde edilen asimptotik güven aralıkları karşılaştırıldığında ise EM aralıklarının Newton-Raphson yöntemine 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. Aynı zamanda analiz sonucu elde edilen tablolar incelendiğinde farklı örneklem büyüklükleri ve farklı sansür şemalarına göre oluşturulan benzetim çalışması senaryolarında her iki yöntemde örneklem genişliği arttıkça daha tutarlı sonuçlar verdiği gözlemlenmiştir.

Sonuç olarak Genelleştirilmiş üstel dağılımının ilerleyen tür tip II sağdan sansürlü örnekleme parametre tahmini için yapılan bu benzetim çalışması ve gerçek yaşam verisinden EM algoritması

yöntemi ile elde edilen sonuçların, Newton-Raphson yöntemine göre daha iyi olduğu sonucuna varılmıştır.

KAYNAKLAR

- [1] Topçu Ç. Greenwood ve Kaplan-Meier Metodu Yardımı ile Varyans Tahmini, Yüksek Lisans, Muğla Sıtkı Koçman Üniversitesi, 2007, Muğla, 48s.
- [2] Kleinbaum DG ve Klein M. Survival Analysis, A Self Learning Text, 3. Baskı, Springer New York Dordrecht Heidelberg London, 2012, USA, 700s.
- [3] Allison PD. Survival analysis using SAS: A practical guide. 1995, Cary, NC: SAS Institute.
- [4] Kanık A & Kul S.. Sağdan Sansürlü Gözlemlerin Yerleşiminin Hazard Oranı Tahminine Etkisi. Türkiye Klinikleri Biyoistatistik Dergisi, 2012; s. 20-26.
- [5] Nelson, W. Applied Life Data Analysis. Kanada: John Wiley & Sons., 1982.
- [6] Kuş C. İlerleyen Tür Tip II Sağdan Sansürlü Örnekleme Dayalı Düzgün Dağılımın Parametrelerinin Jackknife Tahmin Edicisi. İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi, 2005; 2, 11-29.
- [7] Dempster AP, Laird NM; Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, 1977; Series B 39 : 1–38.
- [8] Gupta RD and Kundu D Generalized Exponential Distribution, Austral. N. Z.Statist. 1999; 41(2), 173-188.
- [9] Gupta R ve Kundu D. Generalized Exponential Distribution: Different Method of Estimations, J. Statist. Comput. Simul., 2000; Vol. 00: 1-22.
- [10] Yazıcı F. EM Algoritması ve Uzantıları, 2005; Ankara, 85s.
- [11] Kuş C, Kaya MF. Estimation of Parameters of Loglogistic Distribution based on Progressive Censoring using Em Algorithm. Hacettepe Journal of Mathematics and Statistics, 2006; 35(2), 203-211.
- [12] Balakrishnan N. Progressive Censoring Methodology: An Appraisal (with discussions), Test 16, 2007; 211–296.
- [13] Pradhan B ve Kundu D. Estimating the Parameters of the Generalized Exponential Distribution in Presence of Hybrid Censoring, Sociedad de Estadística e Investigación Operativa, 2008; 18: 497–515.
- [14] Saraçoğlu B. Pareto Dağılımının Parametrelerinin İlerleyen Tür Tip-II Sağdan Sansürlü Örneklemlere Dayalı En Küçük Kareler Tahmini, İstatistikçiler Dergisi, 2012; 5: 43-50.
- [15] Zhu S. A New Hybrid Estimation Method for the Generalized Exponential Distribution, M.Sc. Thesis, University of Calgary, Alberta, 2013; 45s.
- [16] Eteman V. *Güvenilirlik analizi üzerine bir yazılım*, Yüksek Lisans, Selçuk Üniversitesi, Konya, 2014; 30s.

- [17] Demir E. *İlerleyen tür tip II sağdan sansürleme altında bazı yaşam zamanı dağılımları için parametre tahmini*, Yüksek Lisan, Selçuk Üniversitesi, Konya, 2015; 58s.
- [18] Arabi Belaghi R., Valizadeh Gamchi F, Bevrani H, Gürünlü Alma Ö. 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.
- [19] Bain LJ and Engelhardt M.. *Statistical Analysis of Reliability and Life-testing Models*. New York: Marcel Dekker Inc. 1978.
- [20] Gupta RD and Kundu D. Generalized exponential distributions: Different Method of Esimations. *Journal of Statistics Computation and Simulation*, 2001; 69, 315-338.
- [21] Raqab MZ and Ahsanullah M. Estimation of the location and parameters of the generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, 2001; 69, 109-124.
- [22] Balakrishnan N and Aggarwala R. *Progressive Censoring: Theory, Methods and Applications*, Birkhäuser, Boston, 2000.
- [23] Everitt, BS. *Introduction to optimization methods and their application in statistics*. Chapman and Hall, London, 1987.
- [24] Ng T, Chan CS., Balakrishnan N. Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics and Data Analysis*, 2002; 39, 371-386.
- [25] Pradhan B ve Kundu D. On progressively censored generalized exponential distribution, *Sociedad de Estadística e Investigación Operativa*, 2008; 18: 497–515.
- [26] Louis TA. Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society*, 1982; 44 (2): 226-233.
- [27] Abramowitz M, Stegun IA *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*,1964; Dover, New York.
- [28] Bader MG and Priest AM *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, 1982; 1129-1136.
- [29] Surlles JG and Padgett WJ. Inference for $P(Y < X)$ in the Burr-Type X Model, *Journal of Applied Statistical Sciences* 2001; 7, 225-238.
- [30] Raqab MZ and Kundu D. (2005). Comparison of different estimators of $P(Y < X)$. Rupert Miller and Jerry Halpern. 1982. Regression with censored data. *Biometrika* 69, 3 (1982), 521–531.