Azalan Bozulma Oranına Sahip (DFR) Yeni bir Yaşam Dağılımı ve Parametrelerinin EM Algoritması Kullanılarak Tahmin Edilmesi

Mehmet Fedai KAYA* Coşkun KUŞ Buğra SARAÇOĞLU

ÖZET

Bu çalışmada azalan bozulma oranına sahip iki parametreli yeni bir yaşam dağılımı önerildi. Önerilen dağılımın çeşitli özellikleri incelendi ve parametrelerin en çok olabilirlik tahmin edicileri EM algoritması yardımıyla elde edildi. Daha sonra bir simulasyon çalışması yapılarak sonuçlar değerlendirildi.

Anahtar Kelimeler: *EM algoritması, üstel dağılım, poisson dağılımı, yaşam zamanı dağılımı, bozulma(hazard) oranı, en çok olabilirlik tahmin edicisi, azalan bozulma oranı(DFR).*

1. GİRİŞ

Bozulma oranı fonksiyonunun zamanla azaldığı durumlar, bir çok istatistikçi tarafından belirtilmiştir. Bunun örnekleri olarak, işyerlerinin iflası(Lomax,1954), Boing 720 filosundaki uçakların havalandırma sistemindeki bozulmalar(Proschan, 1963) ve entegre devre modüllerinin ömrü(Saunders and Mhyre, 1983) verilebilir. Genel olarak bir populasyonun zamana bağlı davranışı biyolojide bağışıklık "immune" ile karakterize edildiğinde o populasyonun azalan bozulma oranına sahip olması beklenir.

2. YAŞAM DAĞILIMI VE KARAKTERİSTİKLERİ

$\{Y_i\}_{i=1}^{Z+1}$, bağımsız ve aynı $f(y;\beta)=\beta e^{-\beta y}, \beta>0, y>0$ yoğunluğuna sahip rasgele değişkenlerin bir dizisi olsun. Burada Z, $f(z;\lambda)=\dfrac{e^{-\lambda}\lambda^z}{z!}$, $z\in\{0,1,2,\ldots\}$, $\lambda>0$ olasılık fonksiyonuna sahip bir rasgele değişkendir. $X=\min\left(\{Y_i\}_{i=1}^{Z+1}\right)$ olsun. O zaman,

$$f(x/z;\beta)=\beta(z+1)e^{-\beta(z+1)x}$$

ve $\theta=(\beta,\lambda)$ olmak üzere

$$f(x;\theta)=\beta\left[1+\lambda e^{-\beta x}\right]e^{-\lambda-\beta x+\lambda e^{-\beta x}},\ \underline{x}>0,\ \lambda>0, \beta>0$$

* Selçuk Üni. Fen-Edeb. Fak. İstatistik Bölümü, Tel:(332) 241 00 11 /1350, e-mail: fkaya@selcuk.edu.tr

olacaktır. Kolaylık olması bakımından $f(x;\theta)$'ya EP(Exponential-Poisson) dağılımı diyeceğiz. X rasgele değişkenin dağılım fonksiyonu,

$$F(x;\theta)=1-e^{-\lambda-\beta x+\lambda e^{-\beta x}},\ x>0$$

medyanı,

$$F(x;\theta)=1-e^{-\lambda-\beta x+\lambda e^{-\beta x}}=\frac{1}{2}$$

Burada denklemin kökü(medyan), Newton-Raphson yöntemini,

$$x^{(t+1)}=x^{(t)}-\frac{-\log 2+\lambda+\beta x^{(t)}-\lambda e^{-\beta x^{(t)}}}{\beta+\lambda\beta e^{-\beta x^{(t)}}}$$

kullanarak elde edilir.

Beklenen değeri,

$$E(X)=\frac{1-e^{-\lambda}}{\lambda\beta}$$

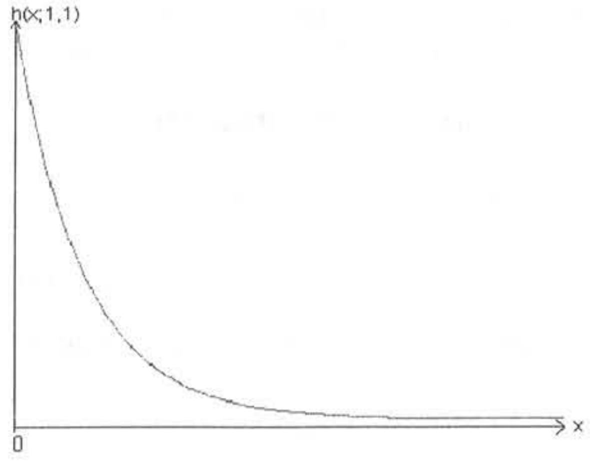
yaşam fonksiyonu,

$$s(x;\theta)=1-F(x;\theta)=e^{-\lambda-\beta x+\lambda e^{-\beta x}}$$

bozulma(hazard) oranı fonksiyonu da

$$h(x;\theta)=f(x;\theta)\{s(x;\theta)\}^{-1}=\beta\left(1+\lambda e^{-\beta x}\right)$$

olup grafiği



Şekil 2. $\beta=1$ ve $\lambda=1$ için bozulma oranının grafiği

biçimindedir. Buradan da görüldüğü gibi EP dağılımı azalan bozulma oranı (DFR) özelliğine sahiptir.

Ortalama geriye kalan ömür,

$$m(x_0,\theta)=E(X-x_0/X\geq x_0)=\frac{e^{\beta x_0}}{\beta\lambda}\left(1-e^{-\lambda e^{-\beta x_0}}\right)$$

şeklindedir.

3. EM ALGORİTMASI

EM algoritması, rasgele değişkenlerin bazılarının gözlenmediği yani kayıp veya tamamlanmamış olduğu durumlarda, parametrelerin en çok olabilirlik tahminlerini elde etmek için kullanılan çok genel bir iteratif algoritmadır. EM (Expectation-Maximization) terimi ilk kez (Dempster, Laird ve Rubin, 1977)'de ortaya atılmış ve algoritmanın yapısı ve davranışı ile genel sonuçların ispatı ve çok sayıda uygulama bu çalışmada sunulmuştur.

Varsayalım ki $\underline{Y}=(Y_1,...,Y_n)$, $f(\underline{y};\theta),\theta\in\Re^p$ ortak o.(y.)f.'na sahip bir örneklem olsun. $Y_1,...,Y_n$'lerin hepsi gözlenmişse θ'nın en çok olabilirlik tahmin edicisi, tam log-olabilirlik fonksiyonunun "$\ell(\theta,\underline{y})$" maksimize edilmesiyle elde edilir. Burada,

$$\ell(\theta,\underline{y})=\log L(\theta,\underline{y})=\log f(\underline{y},\theta)$$

şeklindedir.

EM algoritması, $\ell(\theta,\underline{y})$'nin θ'ya göre maksimize edilmesinin kolay fakat gözlenen log-olabilirlik fonksiyonunun "$\ell_g(\theta,\underline{y}_g)$" maksimize edilmesinin zor olduğu durumlarda kullanışlıdır. $\underline{Y}$'nin bir kısmı gözlenemediğinde $\ell(\theta,\underline{y})$ maksimize edilemez. EM algoritması $\underline{Y}_g$ verildiğinde $\ell(\theta,\underline{y})$ nın koşullu beklenen değerini iteratif olarak maksimize eder.

Y tam veriyi(complete data); $\underline{Y}_g$, gözlenen veya tamamlanmayan(incomplete) verileri; $\underline{Y}_k$, kayıp(missing data) veya sansürlü verileri göstermek üzere Y'yi $\underline{Y}=(\underline{Y}_g,\underline{Y}_k)$şeklinde ifade edebiliriz(Rubin, D.B., 1976).

$$f(\underline{y};\theta)=f(\underline{y}_g,\underline{y}_k;\theta)$$
$$=f_1(\underline{y}_g;\theta)f_2(\underline{y}_k/\underline{y}_g;\theta)$$

Böylece

$$\ell_g(\theta;\underline{y}_g)=\ell(\theta;\underline{y})-\log f_2(\underline{y}_k/\underline{y}_g;\theta)$$

Burada f_1, $\underline{Y}_g$'nin ortak o.(y.)f., f_2, $\underline{Y}_g=\underline{y}_g$ verilmişken $\underline{Y}_k$'nın ortak o.(y.)f., $\ell_g(\theta;\underline{y}_g)$ gözlenmiş log-olabilirlik fonksiyonudur(observed log-likelihood)

E- adımı.

$$Q(\theta;\theta^{(t)})=E\left[\ell(\theta,\underline{Y})/\underline{Y}_g=\underline{y}_g,\theta'\right]$$

M-adımı.

$E\left[\ell(\theta,\underline{Y})/\underline{Y}_g=\underline{y}_g,\theta'\right]$'yi maksimize eden θ değeri bir sonraki adımda $\theta^{(t+1)}$'in kullanılarak yakınsama gerçekleşinceye kadar E ve M adımları tekrar edilir. Algortimanın yakınsadığı değer θ'nın en çok olabilirlik tahminidir(Little, R.J.A., Rubin, D.B.,1987).

4. PARAMETRELERİN EN ÇOK OLABİLİRLİK TAHMİN EDİCİLERİ

En çok olabilirlik tahmin edicisini bulmak için $y_g=(x_i,i=1,...,n)$ olmak üzere gözlenen verilerin olabilirlik fonksiyonu,

$$L(\theta;y_g)=\beta^n\prod_{j=1}^{n}\left(1+\lambda e^{-\beta x_j}\right)e^{-n\lambda-\beta\sum_{i=1}^{n}x_i+\lambda\sum_{i=1}^{n}e^{-\beta x_i}}$$

ve gözlenen verilerin log-olabilirlik fonksiyonu,

$$\log\left[L(\theta;y_g)\right]=\ell(\theta;y_g)=n\log(\beta)+\sum_{i=1}^{n}\log\left(1+\lambda e^{-\beta x_i}\right)-n\lambda-\beta\sum_{i=1}^{n}x_i+\lambda\sum_{i=1}^{n}e^{-\beta x_i}$$

şeklindedir. En çok olabilirlik tahmin edicilerini Newton-Raphson yöntemiyle bulmak için $\ell(\theta;y_{obs})$ fonksiyonunun gradyent vektörü ve hessian matrisinin bulunması gerekmektedir.
Buradaki türev alma zorluğundan kurtulmak için EM algoritmasını önereceğiz.

$\{Y_i\}_{i=1}^{Z+1}$, bağımsız ve aynı $f(y;\beta)=\beta e^{-\beta y},\beta,y\in\Re_+$ yoğunluğuna sahip rasgele değişkenlerin bir dizisi olsun. Burada Z, $P(z;p)=\dfrac{e^{-\lambda}\lambda^z}{z!},z\in N,\lambda\in\Re_+$ olasılık fonksiyonuna sahip bir rasgele değişkendir. $X=\min\left(\{Y_i\}_{i=1}^{Z+1}\right)$ olsun. O zaman, $\underline{X}=(X_1,X_2,...,X_n)$ gözlenmiş, $\underline{Z}=(Z_1,Z_2,...,Z_n)$'ler de gözlenemeyen veri' ler olsun. Burada Z kayıp veri olarak düşünülebilir.

Modelimiz tamamlanmamış veri problemi gibi görülebileceğinden parametrelerinin en çok olabilirlik tahmin edicileri elde etmede EM algoritması uygulanabilir.

EM algoritmasını uygulayabilmek için varsayılan tam veri 'nin dağılımı

$$f(\underline{x},\underline{z})=\exp(-n\lambda)\lambda^{\sum_{i=1}^{n}z_i}\prod_{i=1}^{n}(z_i!)^{-1}\prod_{i=1}^{n}(z_i+1)\beta^n e^{-\beta\sum_{i=1}^{n}(z_i+1)x_i}$$

$$x_i>0\ ,\ z_i=0,1,2,...\ ,\ \lambda>0\ ,\ \beta>0\ ,\ i=1,2,...,n$$

şeklindedir. Buradan da

$$\log\left[f(\underline{x},\underline{z})\right]=-n\lambda+\sum_{i=1}^{n}z_i\log(\lambda)+\log\prod_{i=1}^{n}(z_i!)^{-1}+\log\prod_{i=1}^{n}(z_i+1)n\log(\beta)-\beta\sum_{i=1}^{n}(z_i+1)x_i$$

ve

$$E\{\log[f(\underline{X},\underline{Z})]\}=$$

$$=-n\lambda+\log(\lambda)\sum_{i=1}^{n}E\left(Z_i/\underline{X}=\underline{x},\beta^{(t)},\lambda^{(t)}\right)+E\left(\log\prod_{i=1}^{n}(Z_i!)^{-1}\Big/\underline{X}=\underline{x},\beta^{(t)},\lambda^{(t)}\right)+$$

$$+E\left(\log\prod_{i=1}^{n}(Z_i+1)\Big/\underline{X}=\underline{x},\beta^{(t)},\lambda^{(t)}\right)+n\log(\beta)-\beta\sum_{i=1}^{n}E\left[(Z_i+1)X_i/\underline{X}=\underline{x},\beta^{(t)},\lambda^{(t)}\right]$$

olur. EM algoritmasının E-adımını gerçekleştirebilmek için $E(Z/X\,;\theta)$ koşullu beklenen değerin hesaplanması yeterlidir.

$$f(z/x)=\frac{f(x,z)}{f(x)}=\frac{e^{-\lambda}\lambda^{z}(z!)^{-1}\beta(z+1)e^{-\beta(z+1)x}}{\beta\left[1+\lambda e^{-\beta x}\right]e^{-\lambda-\beta x+\lambda e^{-\beta x}}}$$

$$=\frac{e^{\left(-\beta xz-\lambda e^{-\beta x}\right)}\lambda^{z}(z+1)}{z!\left(1+\lambda e^{-\beta x}\right)}$$

Buradan

$$E(Z/X)=\sum_{z=0}^{\infty}z\frac{e^{\left(-\beta xz-\lambda e^{-\beta x}\right)}\lambda^{z}(z+1)}{z!\left(1+\lambda e^{-\beta x}\right)}=\frac{\lambda\left(2+\lambda e^{-\beta x}\right)e^{-\beta x}}{1+\lambda e^{-\beta x}}$$

elde edilir. EM algoritmasının M adımında ise tam veriyi kullanarak elde edilen en çok olabilirlik tahmin edicisinde kayıp veri olan $Z_i, i=1,\dots,n$'lerin yerine $E(Z/X\,;\theta)$ yazılarak EM algoritması elde edilir. Tam veriyi kullanarak elde edilen en çok olabilirlik tahmin edicileri,

$$\hat{\beta}_{MLE}=n\left\{\sum_{i=1}^{n}(Z_i+1)X_i\right\}^{-1}\quad,\quad\hat{\lambda}_{MLE}=n^{-1}\sum_{i=1}^{n}Z_i$$

şeklindedir. Böylece

$$\hat{\beta}^{(t+1)}=n\left\{\sum_{i=1}^{n}\left(\frac{\lambda^{(t)}\left(2+\lambda^{(t)}e^{-\beta^{(t)}X_i}\right)e^{-\beta^{(t)}X_i}}{1+\lambda^{(t)}e^{-\beta^{(t)}X_i}}+1\right)X_i\right\}^{-1}$$

$$\hat{\lambda}^{(t+1)}=\sum_{i=1}^{n}\frac{\lambda^{(t)}\left(2+\lambda^{(t)}e^{-\beta^{(t)}X_i}\right)e^{-\beta^{(t)}X_i}}{1+\lambda^{(t)}e^{-\beta^{(t)}X_i}}n^{-1}$$

olacaktır.

5. SİMÜLASYON ALGORİTMASI

EP dağılımından sayı üretmek için aşağıdaki iki alternatif yöntem ileri önereceğiz.

I. Yöntem.

$Z\sim Poisson(\lambda)$ bir sayı üretilir. Daha sonra $Y_1,Y_2,\dots,Y_{z+1}\sim \ddot{U}stel(\beta)$'dan bir örneklem alınır. $X=Min(Y_1,Y_2,\dots,Y_{z+1})$, EP dağılımına sahip bir rasgele değişkendir.

II. Yöntem. (Newton-Raphson Yöntemi kullanılarak)

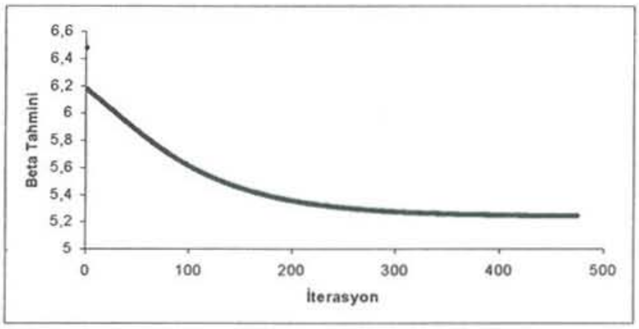
i. $U \sim Düzgün(0,1)$'den bir sayı üretilir.

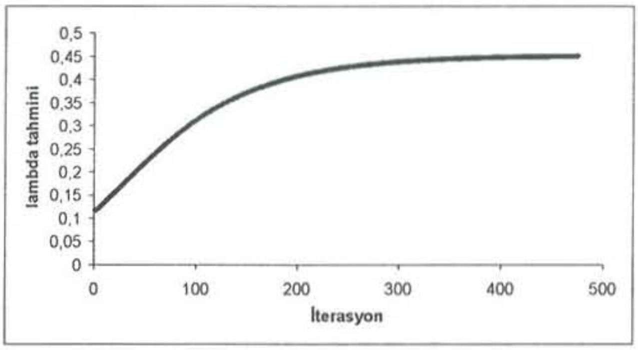
ii. $X^{(t+1)} = X^{(t)} - \dfrac{\log(1-U) + \lambda + \beta X^{(t)} - \lambda e^{-\beta X^{(t)}}}{\beta + \lambda\beta e^{-\beta X^{(t)}}}$

$X^{(0)} = \dfrac{1-e^{-\lambda}}{\lambda\beta}$ alınması uygundur.

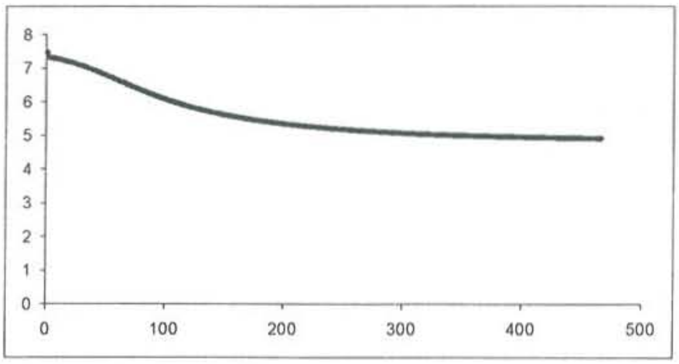
iteresyonunun yakınsadığı değer EP dağılımına sahiptir.

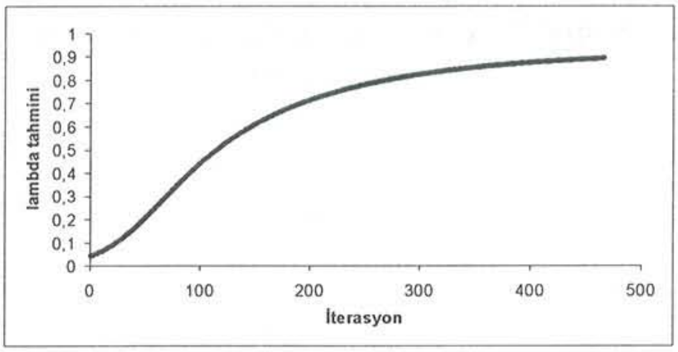
$X_1, X_2, ..., X_{100} \sim EP(\beta = 5, \lambda = .3)$ bir örneklem ve $\beta^{(0)} = 100$, $\lambda^{(0)} = 1$ durumunda (β, λ)'nın en çok olabilirlik tahmin edicilerinin EM iterasyonu sonucunda aldığı değerler





$X_1, X_2, ..., X_{1000} \sim EP(\beta = 5, \lambda = .9)$ bir örneklem ve $\beta^{(0)} = 100$, $\lambda^{(0)} = .3$ durumunda (β, λ)'nın en çok olabilirlik tahmin edicilerinin EM iterasyonu sonucunda aldığı değerler





Burada durdurma kuralı $\beta^{(t+1)} - \beta^{(t)} < \varepsilon = 0.00001$ biçimindedir.

6. SONUÇ

Elde edilen EP dağılımı DFR özelliğine sahip olayların modellenmesinde kullanılabilir. Newton-Raphson yöntemiyle elde edilen parametrelerin en çok olabilirlik tahminleri, EM algoritmasından elde edilen sonuçlarla karşılaştırılır ve modele uygun bir örnek bulunabilirse bu çalışma güçlendirilebilir.

KAYNAKLAR

ADAMIDIS, K., LOUKAS, S. (1998), *A Life Time Distribution with Decreasing Failure Rate*, Statistics&Probability Letters, 39,35-42

DEMPSTER,A.P., LAIRD, N.M., RUBIN, D.B.(1977), *Maximum Likelihood From Incomplete Data via The EM Algorithm (with discussion)*, J. Roy. Statist. Soc. Ser. B 39, 1-38

LITTLE, R.J.A., RUBIN, D.B.(1987) ,*Statistical Analysis with Missing Data*, John Willey&Sons

LOMAX, K.S.(1954), Business Failures: *Another Example of The Analysis of The Failure Data*, J. Roy. Statist. Soc. Ser. B 44, 226-233.

LOUIS, T.A.(1982) *Finding The Observed Information When Using The EM Algorithm*, J. Roy. Statist. Soc. Ser. B 44, 226-233

PROSCHAN, F.(1963),*Teoratical Explanation of Observed Decreasing Failure Rate*, Technometrics, 5, 375-383.

RUBIN, D.B.(1976), *Inference with Missing Data, Biometrika*,63,581-592

SAUNDERS, S. C. and MYHRE, J. M. (1983), *Maximum likelihood estimation for two-parameter decreasing hazard rate distributions using censored data*. Journal of the American Statistical Association ,78, 664-673.

A New Life Time Distribution with Decreasing Failure and Estimation of Parameters using EM Algorithm

ABSTRACT

In this study a two parameter distribution with decreasing failure rate is introduced. Various properties are discussed and the estimators of parameters are obtanied by the method of maximum likelihood using EM algorithm and results are evaluated with simulation study.

Key Words: *EM algorithm, exponential distribution, decreasing failure rate (DFR), failure(hazard) rate, lifetime distribution, maximum likelihood estimator,poisson distribution*