Öneri, C.3, S.13.
Ocak 2000, ss.105-109.

ÖRNEK ORTALAMA VE ORANLARININ DENEYSEL BAYES ANALİZİ

Dr.Filiz ÇAKIR

M.Ü. S.B.M.Y.O., Öğretim Görevlisi

***Abstract:** The Empirical Bayes Analysis aims at the prediction of the last possibility values of the sample while the first possibility values are already available (known). In this article, the application of the Empirical Bayes Analysis in predicting the last possibility values of the sample's averages and proportions has been mentioned.*

I-GİRİŞ

İstatistiksel araştırmaların çoğu, benzer nitelikteki olaylar ile ilgili tahmin işlemlerine yöneliktir. Farklı bölgelerde faaliyet gösteren firmaların yıllık gelir tahminleri, sigorta şirketlerinin gruplar bazında risk oranlarını tahmin etmesi ya da aynı işyerinde değişik alanlarda kullanılan makinelerin bozulma oranlarının tahmin edilmesi gibi olaylarda, istatistiksel tahmin yöntemlerini kullanarak bir karar problemi oluşturulmaktadır. Araştırmacının amacı, sahip olduğu örneği kullanmak suretiyle, bilinmeyen parametre hakkında bilgi edinmektir. Parametre değerinin ilk değerlerinin bilinmesi, örneğe ait son olasılık değerlerini ortaya koymaktadır. Deneysel Bayes analizinin konusu bu tip olaylardır. Klasik Bayes analizinden farklı olarak, deneysel bayes analizinde, bir dağılıma sahip olan x tesadüfi değişkenine ait θ parametresinin de dağılımı dikkate alınmakta, bu parametrenin sahip olduğu olasılık dağılımının özelliğine göre, hiperparametreler olarak tanımlanan parametrenin parametreleri tahmin edilmekte ve son olasılık değerinin elde edilmesi amaçlanmaktadır.

Bilindiği gibi, X tesadüfi değişkeninin sahip olduğu $f(x/\theta)$ olasılık fonksiyonu ile θ parametresinin sahip olduğu $\pi(\theta)$ olasılık fonksiyonunun bileşik olasılık dağılımı,

$$P(x,\theta) = f(x/\theta)\,\pi(\theta)$$

şeklinde ifade edilmektedir. Bu durumda X tesadüfi değişkeninin marjinal olasılığı, sürekli durumlarda,

$$P(x) = \int f(x/\theta)\,\pi(\theta)\,d\theta$$

kesikli durumlarda ise,

$$\sum f(x/\theta)\,\pi(\theta)$$

olarak tanımlanmaktadır. X tesadüfi değişkeni ve θ parametresine ait olasılık dağılımlarının normal dağılım olduğu varsayımı altında ise,

$$X \sim N(\theta, \sigma^2_f)$$

$$\pi(\theta) \sim N(\mu_\pi, \sigma^2_\pi)$$

X tesadüfi değişkeninin marjinal son olasılığı ,

$$P(x) \sim N(\mu_\pi, \sigma^2_\pi + \sigma^2_f)$$

şeklinde ifade edilmektedir [1]. Klasik Bayes analizine göre, ilk olasılıklar olarak tanımlanan marjinal olasılıklar, Bayes analizinin bir özelliği olarak, ya verinin kendisi kullanılarak ,başka bir deyişle objektif bilgiler ile, ya da subjektif bilgilere başvurularak belirlenmektedir. Subjektif bilgiler, araştırma yapan kişilerin kendi deney ve tecrübelerine dayanarak oluşturulan, eski dönemlere ait bilgileri kullanarak elde edilen bilgilerdir. Objektif bilgiler ise, verinin kendisinden elde edilmektedir. p sayıdaki X tesadüfi değişkeni söz konusu olduğunda, yine p sayıda θ parametresi de olacağından, her bir x_i tesadüfi değişkeninin marjinal olasılığı, i = 1,2,...., p olmak üzere, $P_0(xi) = \int f(x_i/\theta_i)\,\pi_0(\theta_i)$ dağılımı ile açıklanacaktır. Buna göre, x tesadüfi değişkeninin marjinal olasılığı ise, p sayıda tesadüfi değişken olduğu düşünüldüğünde,

$$\begin{aligned} P(x) &= \int f(x/\theta)\,\pi(\theta)\,d\theta \\ &= \int \left[\prod_{i=1}^{p} f(x_i/\theta_i)\right]\left[\prod_{i=1}^{p} \pi_0(\theta_i)\right] d\theta \\ &= \prod_{i=1}^{p} \int f(x_i/\theta_i)\,\pi_0(\theta_i)\,d\theta \\ &= \prod_{i=1}^{p} P_0(x_i) \end{aligned}$$

şeklinde ifade edilmektedir. Deneysel Bayes analizinde, x tesadüfi değişkenine ait θ_1, θ_2, ...,θ_p parametreleri arasındaki ilişkiler bilinmekte ve bu parametre vektörü, ilk olasılıkların belirlenmesinde kullanılmaktadır.

Deneysel Bayes analizinde, θ_i, genel olarak ortalama ve oran gibi ölçülerdir. Örneğin, θ_i, aynı endüstri dalında çalışan firmalarda meydana gelen işçi kazaları ortalaması ya da bir fabrikanın farklı alanlarında kullanılan bir ürüne ait bozulma oranları olarak tanımlanabilmektedir. Uygulamada daha ziyade, bir ürün için bozulma oranlarının incelenmesinde deneysel bayes yöntemleri özellikle kullanılmaktadır.

Bu makalenin amacı, deneysel bayes analizinin genel teorisini anlatarak, uygulamadaki yerinin belirlenmesi ve bu analizin örnek bir olaya uygulanışını göstermektir. Bu amaçla, ilk bölümde, genel teori verilmiştir. Daha sonra ise, analizin normal süreçte ve üssel dağılımlarda , en yaygın olarak ele alınan Poisson sürecindeki işleyişi anlatılmış ve son olarak bir uygulama yapılmıştır.

II. GENEL TEORİ

Deneysel Bayes analizi iki stokastik süreç ile ilgilenmektedir. Bunların birincisi, verinin kendisi, yani X tesadüfi değişkeni ve bu değişkene ait olasılık fonksiyonu, $f(x/\theta)$, diğeri ise, bu değişken kümesine ait parametre vektörü ve olasılık fonksiyonu, $\pi(\theta)$ dur [2]. Bu iki stokastik sürece ait olasılık dağılımları bilindiğinde, θ parametresinin tahmin edilmesi, bu parametreye ait olasılık dağılımının parametrelerinin tahmin edilmesi ile yapılmaktadır. Hiperparametre olarak tanımlanan bu parametreler bilinmemekte, doğrudan verinin kendisinden elde edilmektedir [3]. Bu tahmin işleminde ise, ya momentler yöntemi ya da maksimum benzerlik yöntemi kullanılmaktadır. Hiperparametrelerin tahmin edilmesinden sonra, θ parametresinin son olasılık dağılımını elde etmek mümkün olmaktadır.

Deneysel Bayes analizi, parametrik ve parametrik olmayan deneysel bayes analizi olmak üzere iki grupta incelenebilmektedir. Parametrik deneysel bayes analizinde, x tesadüfi değişken kümesine ait bilgiler, parametrenin ilk olasılık değerlerini tahmin etmekte kullanılmakta, bilinmeyen hiperparametrelerin tahmin edilmesi ile de, θ parametresinin son dağılımı elde edilmektedir. Parametrik olmayan deneysel bayes analizinde ise, parametreler ile ilgilenilmemekte, verinin kendisi ilk olasılık dağılımının tahmininde kullanılmaktadır. Bu nedenle, parametrik olmayan deneysel bayes analizinin, Bayesyen teori ile benzerlik gösterdiği söylenebilmektedir [2]. Klasik Bayes teorisinde de, genellikle objektif olasılıklar olmak üzere, bilinmeyen parametrelere bir ilk değer tayin edilmekte, bu parametreler için, ikinci bir olasılık modeli uygulanmamaktadır.

II.1. Normal Süreçte Deneysel Bayes Analizi

X tesadüfi değişkeni ile bu değişken kümesine ait θ parametre setinin normal dağıldığı varsayımı altında, deneysel bayes analizi ile ilgili ilk işlem, θ_i parametrelerinin hiperparametrelerini tahmin etmektir. $i = 1,2,...,p$ olmak üzere, $X_i \sim N(\theta_i, \sigma^2_f)$ ve $\theta i \sim N(\mu_\pi, \sigma^2_\pi)$ şeklinde tanımlandığında, tesadüfi değişkenin varyansının bilindiği, ancak θ_i parametresinin ortalama ve varyansının bilinmediği varsayılmaktadır [3]. Bu hiperparametreler, deneysel bayes analizinin bir özelliği olarak, ya momentler yöntemi ile ya da maksimum benzerlik yöntemine göre tahmin edilmektedir [5]. Buna göre, hiperparametrelerin tahminleri, momentler yöntemine göre, $\hat{\mu}_\pi = \bar{x}$ ve $\hat{\sigma}^2_\pi = \dfrac{s^2}{p-1} - \sigma^2_f$ şeklinde açıklanırken, maksimum benzerlik yöntemine göre, $\hat{\mu}_\pi = \bar{x}$ ve $\hat{\sigma}^2_\pi = \dfrac{s^2}{p} - \sigma^2_f$ şeklinde ifade edilmektedir.

Bilindiği gibi, $s^2 = \sum_{i=1}^{p}(x_i - \bar{x}_i)^2$ eşitliği ile bulunmaktadır [8]. Bu iki hiperparametrenin yukarıda açıklanan tahmin yöntemleri ile tahmin edilmesinden sonra, θ_i parametresinin son olasılık dağılımı açıklanabilmektedir. Buna göre, θ_i parametresinin son olasılık dağılımı da normal olmaktadır ve bu dağılım $N(\mu_i(x_i), V)$ şeklinde ifade edilmektedir. Son olasılık dağılımının parametre değerleri ise,

$$\mu_i(x_i) = x_i - (\sigma^2_f / \sigma^2_f + \sigma^2_\pi)\ (x_i - \mu_\pi)$$

$$V = \sigma^2_\pi \sigma^2_f / \sigma^2_\pi + \sigma^2_f = \sigma^2_f\ (1 - (\sigma^2_f / \sigma^2_\pi + \sigma^2_f)$$

şeklinde tanımlanmaktadır [4]. Deneysel Bayes analizi ile ilgili olarak ilk çalışmalardan birini yapan Carl N.MORRİS tarafından 1983 yılında yayınlanan makalede, değişken sayısı olarak ifade edilen p sayısının, $p \geq 4$ olması durumunda, hiperparametreler,

$$\mu_i(x_i) = x_i - \hat{B}\ (x_i - \bar{x})$$

$$V_i(x_i) = \sigma^2_f\ (1 - (p-1/p)\ B) + (2/p-3)\ \hat{B}^2\ (x_i - \bar{x})^2$$

şeklinde tanımlanmıştır [2]. Formüllerde yer alan B 'nin eşiti,

$$\hat{B} = (p-3/p-1)\ (\sigma^2_f / \sigma^2_f + \sigma^2_\pi)$$

olarak ifade edilmektedir.

II.2. Poisson Sürecinde Örnek Oranlarının Deneysel Bayes Analizi

Deneysel Bayes analizi, örnek ortalamaları ile ilgili uygulamalarda değişken değerlerinin normal dağıldığı varsayımı altında incelenebildiği gibi, herhangi bir örneğe ait başarısızlık ya da bozulma oranlarını gösteren veri kümesi ile parametrenin tahmin edilmesinde de kullanılabilmektedir. Genel olarak bozulma oranlarının nadir olarak ortaya çıktığı durumlarda dağılım Poisson sürecine daha uygun oolmaktadır. Dolayısıyla burada üssel bir süreçte deneysel bayes analizinin uygulanışı ele alınmıştır. Poisson süreci, ortalama başarı oranını ifade eden λ parametresi ile tek parametreli bir dağılım olduğundan öncelikle bu parametrenin tahmin edilmesi gerekmektedir [5] .

Genel olarak, oranlar ile ilgili uygulamalarda, x tesadüfi değişkenine ait değerler, bozulma sayıları ve bozulma dönemleri (yıl, ay, gün v.b.) olarak verilmektedir. Buna göre, si değerleri, bozulma sayılarını, ti değerleri ise, bozulma zamanlarını ifade etmek üzere, ri=si/ti değerleri söz konusu olacaktır. Başka bir deyişle, ri değerleri, bozulma oranlarını göstermektedir ve bu değerler λ parametresinin ilk değerleridir. Poisson sürecinde, xi tesadüfi değişkenine ait bilinmeyen parametre olan λi parametresinin Gamma dağılımına sahip olduğu varsayılmaktadır. ri değerlerine ait beklenen değer ve varyans,

$$E[ri] = E[\lambda]$$

$$Var[ri] = Var[\lambda] + E[\lambda]\ (1/ti)$$

şeklinde ifade edilmektedir. λ parametresinin beklenen değer ve varyansı ise, moment yöntemine göre,

$$E[\lambda] = \overline{r_i}$$

$$Var[\lambda] = s^2{}_r - \bar{r}\ \left(1/p \sum_{i=1}^{p} 1/ti \right)$$

denklemleri ile elde edilmektedir [6] . λ parametresinin dağılımı Gamma dağılımı ve hiperparametreler ise Gamma dağılımının parametreleri olan α ve β olduğu için, Gamma dağılımı için beklenen değer ve varyans,

$$E[\lambda] = \beta / \alpha$$

$$Var[\lambda] = \beta / \alpha^2$$

şeklinde açıklanmaktadır. λ parametresinin hiperparametreleri α ve β'nın ilk tahmini değerleri momentler yöntemine göre tahmin edildiğinde, ilk olasılıklarının belirlenmesi ve hiperparametrelerinin elde edilmesinden sonraki adım, λi parametresine ait son olasılık dağılımı, dolayısıyla bu dağılımın parametrelerini tahmin etmektir. Hiperparametrelerin elde edilen son değerleri ise, λ parametresinin son olasılık dağılımını belirlemede kullanılmaktadır. Gamma dağılımında, λi parametresine ait son olasılık dağılımının beklenen değeri ve varyansı,

$$E[\lambda i / si] = (si + \beta) / (ti + \alpha)$$

$$Var[\lambda i / si] = (si + \beta) / (ti + \alpha)^2$$

şeklinde ifade edilmektedir.

Deneysel Bayes analizinin uygulanışını genel olarak tekrar ifade etmek gerekirse; Deneysel Bayes analizinin normal ve üssel süreç içindeki uygulamaları öncelikle yukarıda da izah edildiği gibi, tesadüfi değişken kümesinin dağılımına ve dolayısıyla parametre değerlerinin dağılımına bağlı olarak yapılmaktadır. Bilinmeyen hiperparametre değerlerinin tahmin edilmesi ise, momentler ya da maksimum benzerlik yöntemine göre belirlenmekte, bu hiperparametrelerin tahmin edilmesinden sonra, parametrenin son olasılık değerleri elde edilmektedir.

III.UYGULAMA

Deneysel Bayes analizinin örnek oranlarına uygulanmasını göstermek amacı ile, 10 farklı modeldeki fotokopi makinelerinde kullanılan aynı tip parçanın bozulma oranları veri olarak alınmış ve test edilmiştir. Her bir makinenin ilk alındığı günden itibaren kullanıldığı süre içindeki çekim sayıları, o makineye ait zamanı yani ti değerlerini, bu süre içindeki bozulma sayıları ise si değerlerini göstermektedir. Aşağıdaki tabloda bu veriler görülmektedir.

Tablo.1

Makine No	Bozulma Sayıları	Çekim Sayısı
1	16	665.573
2	4	209.855
3	7	515.285
4	3	135.440
5	12	604.115
6	9	617.095
7	21	743.140
8	5	743.165
9	12	294.850
10	14	184.651

Çekim sayısı, yapılan hesaplamalarda ti/10000 olarak alınmıştır. si ve ti değerlerini kullanarak bulunan ri=si/ti değerleri λ parametresinin ilk değerlerini ifade etmektedir. Bu verileri temel alarak , yapılan işlemler sonucunda elde edilen değerler aşağıdaki tabloda toplu olarak görülmektedir.

Tablo.2

1	2	3	4	5	6	7
Makine No	*B.Sayıları Si*	*Ç.Sayıları ti*	*İlk Değerler ri=si/ti*	*St.Sapma*	*Son Değerler*	*St.Sapma*
1	16	66,5573	0,240	0,25	0,235	0,248
2	4	20,9855	0,190	0,50	0,180	0,484
3	7	51,5285	0,135	0,38	0,134	0,371
4	3	13,544	0,221	0,58	0,201	0,553
5	12	60,4115	0,198	0,29	0,194	0,286
6	9	61,7095	0,145	0,33	0,143	0,32
7	21	74,314	0,282	0,22	0,276	0,217
8	5	74,3165	0,067	0,45	0,068	0,436
9	12	29,485	0,406	0,29	0,381	0,286
10	14	18,4651	0,758	0,27	0,675	0,265

α ve β hiperparametrelerinin ilk tahmini değerleri, momentler yöntemine göre hesaplanmıştır. Buna göre, $\alpha = 2{,}638$ ve $\beta = 0{,}264$ olarak hesaplanmıştır. Tabloda yer alan tahmin değerlerinin açıklamaları sırasıyla aşağıda verilmiştir.

İlk Değerler : λ parametresine ait ilk değerler, veri değerleri kullanılarak bulunmuştur. İlk değerleri ifade eden ri değerleri, si/ti değerlerine eşittir.

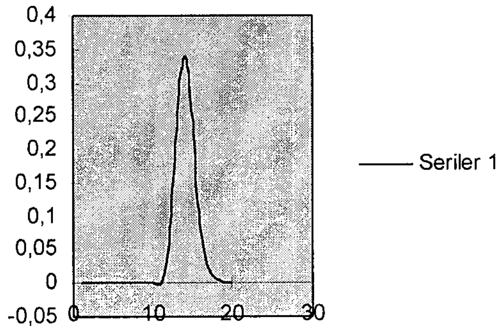
Standart Sapma: λ parametresine ait ilk standart sapmayı ifade etmektedir.

Son Değerler : λ parametresine ait, Gamma olasılık dağılımının son parametre değerlerini kullanarak hesaplanan son beklenen değeri ifade etmektedir.

Standart Sapma : λ parametresinin son standart sapmasını ifade etmektedir.

Aşağıdaki grafik, λ_1 parametresinin hiperparametrelerinin son değerlerini elde ettikten sonra , bu parametrenin son olasılık değerlerini göstermektedir.

Grafik.1



IV. SONUÇ

Örnek ortalamaları ve oranlarına ait bir veri kümesinde deneysel bayes analizi kullanılarak, bilinmeyen parametrelerin son olasılık dağılımının tahmin edilmesi, tesadüfi değişken sayısının fazla olduğu durumlarda, klasik bayes yöntemine göre daha uygun olacaktır[7,s.840]. Bir karşılaştırma yapmak gerekirse, bayes yönteminde ilk olasılıkların belirlenmesinde objektif ve subjektif olasılıklar kullanılırken, deneysel bayes analizinde marjinal olasılıklar dikkate alınarak hiperparametrelerin tahmini verinin kendisinden elde edilmektedir. Deneysel Bayes analizi, θi parametresi için bir olasılık modeli önermekte ve bu olasılık modelini kullanarak hiperparametreleri tahmin etmektedir. Bu tahminler, son olasılık dağılımının parametrelerini belirlemede kullanılmaktadır.

Değişken sayısının az olduğu durumlarda klasik bayes yöntemlerinin kullanılması daha uygun görünse de, bu sayının fazla olduğu uygulamalarda, bileşik olasılık dağılımının elde edilmesinde görülen zorluklar nedeniyle, deneysel bayes yöntemlerinin kullanılması, parametrelerin son olasılık dağılımını elde etmede daha kolay olabilecektir.

YARARLANILAN KAYNAKLAR

[1] Berger. James O.. **Statistical Decision Theory and Bayesian Analysis.** Second Edition. Springer-Verlag. 1985.

[2] Morris. Carl N.. "Parametric Empirical Bayes İnference: Theory and Applications". **Journal of the American Statistical Association**, Volume 78, Number 381. March 1983.

[3] Ghosh. Malay; Meeden. Glen. "Empirical Bayes Estimation in Finite Population Sampling". **Journal of the American Statistical Association**, Volume 81. Number 396. December 1986.

[4] Arora, Vipin; Lahiri, P.; Mukherjee. Kanchan. "Empirical Bayes Estimation of Finite Population Means from Complex Surveys". **Journal of the American Statistical Association**, Volume 92, Number 440. . December 1997.

[5] E. Kass. Robert; Steffey. Duane. "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Bayes Models)". **Journal of the American Statistical Association**, Volume 84, Number 407. September 1989.

[6] P. Gaver. Donald; I.G.O'Muircheartaigh. "Robust Empirical Bayes Analyses of Event Rates". **Tecnometrics**. Vol.29; No:1, February 1987.

[7] Deely. J.J.; Lindley. D.V.. "Bayes Empirical Bayes". **Journal of the American Statistical Association**. Volume 76. Number 376, December 1981.

[8] Efron, Bradley. "Empirical Bayes Methods for Combining Likelihoods", **Journal of the American Statistical Association**. Volume 91. Number 434. June 1996.