REGRESYON ANALİZİNDE BAYESCİ YAKLAŞIM

Doç. Dr. Emel ŞIKLAR
Anadolu Üniversitesi
Fen Fakültesi
İstatistik Bölümü

ABSTRACT

In this study it has been shown how the Bayesian approach can be used in the process of parameter estimation in regression analysis. For this purpose, simple and multiple linear Bayesian regressions have been theoretically explained after the Bayes estimator and its properties were analysed.

ÖZET

Bu çalışmada regresyon analizinde parametre kestirimlerini yaparken Bayes yaklaşımının nasıl kullanıldığı gösterilmeye çalışılmıştır. Bu amaçla Bayes kestiricisi ve özellikleri incelendikten sonra basit ve çoklu doğrusal bayesci regresyon teorik olarak açıklanmıştır.

GİRİŞ

Son yıllarda Bayes yaklaşımı istatistiğin hemen hemen tüm alanlarında kullanılmaktadır. Bayes yaklaşımının temelini, 18. Yüzyılda Thomas Bayes adlı bir adam tarafından geliştirilen teorem oluşturmaktadır. Bu teoreme göre A olayının gerçekleşmesi olasılığı bilindiğinde B olayının da gerçekleşmesi olasılığını hesaplamak olanaklıdır. Sözü edilen olasılığın hesaplanmasında kullanılan Bayes teoremi modern istatistik karar kuramında yer alan karar verme tekniklerinde uygulanmaya başlanmıştır.

Karar vermede Bayes karar verme yöntemi kullanılırken parametreler hakkında yapılacak kestirimlerde kişisel deneyim ve bilgiler, yani objektif bilgilerin yanında subjektif bilgiler de karar verme sürecine katılmaktadır. Bayes yaklaşımının temel ilkesi elde edilen bu bilgileri kullanarak yanlış karar verme riskini azaltmaktır.

Regresyon analizinde de daha etkili parametre kestirimleri yapmak ve en önemlisi artık kareler ortalamasını minimum kılmak gibi amaçlarla Bayes yaklaşımı kullanılmaya başlanmıştır.

Bu çalışmada amaç, basit ve çoklu doğrusal regresyon modelinin parametre kestirimlerini Bayes yaklaşımını kullanarak nasıl elde edildiğini araştırmaktır. Bu amaçla, temel bazı kuramsal özellikler ile ilgili genel bilgiler verilmiştir.

I. BAYES KESTİRİCİSİ VE ÖZELLİKLERİ

Genel doğrusal modelde $(X'X)$ matrisi birim olarak alındığında Y_i, θ_i ortalama ve σ^2 varyansı ile normal dağılımlı olacaktır. Önsel (prior) varsayım

$$E(\theta_i) = \mu \qquad E(Yi) = \theta_i^*$$

dır. μ, θ_i için tanımlanan önsel dağılımın ortalaması olup, parametrelerin varyansları da sırasıyla τ^2 ve σ^2 ile gösterilmektedir. θ_i'nin sonsal (posterior) dağılımı Bayes teoremi kullanılarak elde edilirse bu sonsal dağılımın ortalaması olan Bayes kestiricisi

$$(1) \qquad E(\theta/Y) = \frac{\overline{Y}/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2} = \theta_i^*$$

şeklinde elde edilmektedir. Bu eşitlikte $\overline{Y} = \sum Y_i/n$ olmaktadır. (1)'de verilen ifade EKK yerine önerilen Bayes kestiricisidir[1].

[1] LINDLEY, D.V. and SMITH, A.F.M., "Bayes estimates for the linear model", **J.R.Statistical Soc.**, 1972, s.10.

θ_i^* kestiricisi, EKK kestiricisi ile μ'nün tartılı ortalamasıdır. Tartılar Y_i ve μ'nün varyansları ile ters orantılıdır. Bazı araştırmacılar θ_i^*'in duyarlı ve EKK'dan daha iyi bir kestirici olduğunu ileri sürmüşlerdir. Ancak θ_i^* kestiricisinin, θ_i'nin yanlı bir kestiricisi olduğu Lindley ve Smith (1972) tarafından gösterilmiştir. Yanlı olmasının nedeni, kullanılan önsel dağılımların kişiden kişiye farklılıklar göstermesindendir.

$E(\theta_i - \theta)^2/(n-1) \geq (2\tau^2 + \sigma^2)$ koşulunun sağlandığı durumlarda Bayes kestiricilerinin EKK kestiricilerinden daha üstün olacağı belirtilmiştir.[2] Örneklem hacmi yeteri kadar büyük ise, bilinmeyen birleşik dağılım için belirsiz önselin geniş aralığı altında Bayes kestiricisinin EKK kestiricisine oldukça yakın olacağı da Goldstein tarafından ifade edilmiştir.[3] Olabilirlik fonksiyonunun baskısı olmadığında önsel dağılımın sonsal dağılım ve Bayes kestiricilerinin hesaplanmasında etkili olduğu da Gunst tarafından belirtilmiştir.[4]

II. BASİT DOĞRUSAL BAYESCİ REGRESYON

(2) $$Y = X\beta + \varepsilon$$

doğrusal regresyon modelinde bağımsız değişken sayısı bire eşit olduğunda regresyon modeli basit doğrusal regresyon modeli olacaktır. Basit doğrusal regresyon denklemi için ε_i hata terimi 0 ortalamalı ve σ^2 varyanslı normal dağılıma sahipken parametrelere bağlı olabilirlik fonksiyonu,

(3) $$p(Y/X,\beta_0,\beta_1,\sigma)\alpha\frac{1}{\sigma^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2\right)$$

şeklinde yazılır.

---

2 LINDLEY, D.V. and SMITH, A.F.M., s.14.

3 GOLDSTEIN,M. "Bayesian analysis of regression problems", **Biometrika**,63,1, s.54.

4 GUNST, R.F. "Regression analysis with multicollinear predictor variables: definition, detection and effects", **Commun. Statist Theory Meth.** 12(19), s.2240.

Parametreler ile ilgili yetersiz önsel bilgi varsa belirsiz önsel dağılım,

$$p(\beta_0,\beta_1,\sigma)\alpha\frac{1}{\sigma} \tag{4}$$

ile verilir.[5] (3) olabilirlik fonksiyonu ile (4) önsel dağılımın çarpımı sonucunda bileşik sonsal dağılım,

$$p(\beta_0,\beta_1,\sigma/Y,X)\alpha\frac{1}{\sigma^{(n+1)}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2\right)$$

olarak yazılır ve açılırsa,

$$p(\beta_0,\beta_1,\sigma/Y,X)\alpha\frac{1}{\sigma^{(n+1)}}\exp(-\frac{1}{2\sigma^2}(vs^2+n(\beta_0-\beta_{0e})^2 + (\beta_1-\beta_{1e})^2\sum_{i=1}^{n}X_i^2+2(\beta_0-\beta_{0e})(\beta_1-\beta_{1e})\sum_{i=1}^{n}X_i)) \tag{5}$$

elde edilir. Burada, v ve s^2,

$$v=n-2 \qquad s^2=\left(\sum_{i=1}^{n}(Y_i-\beta_{0e}-\beta_{1e}X_i)^2\right)\Big/v$$

olarak tanımlanır.[6] β_{0e} , β_{1e} terimleri bilinen EKK kestiricilerini ifade etmektedirler. Bu dağılım, (β_{0e} , β_{1e}) ortalamalı ve $\sigma^2(X'X)^{-1}$ kovaryans matrisi ile iki değişkenli normal dağılım yapısına uymaktadır. Ancak uygulamada σ^2 çok seyrek bilindiğinden parametrelere ilişkin marjinal sonsal dağılımların kullanılması daha kolaydır. (5) ifadesinin σ'ya göre integrali alındığında, β_0 ve β_1 parametrelerinin bileşik marjinal sonsal dağılımları,

$$p(\beta_0,\beta_1/Y,X)\alpha(vs^2+n(\beta_0-\beta_{0e})^2+(\beta_1-\beta_{1e})^2\sum_{i=1}^{n}X_i^2 +2(\beta_0-\beta_{0e})(\beta_1-\beta_{1e})\sum_{i=1}^{n}X_i)^{-n/2} \tag{6}$$

[5] ZELLNER, A., **An introduction to Bayesian inference in econometrics**, John Wiley, Inc.1971, s.221.
[6] ZELLNER, A., s.222.

olur. (6) ifadesi iki değişkenli Student-t dağılımı yapısındadır. Bu dağılımın özelliğinden β_0, β_1 parametreleri için marjinal sonsal dağılımları ise,

$$p(\beta_0 / Y, X) \alpha \left(v + \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{s^2 \sum_{i=1}^{n} X_i^2 / n} (\beta_0 - \beta_{0e})^2 \right)^{-(v+1)/2}$$

ve

(7) $$p(\beta_1 / Y, X) \alpha \left(v + \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{s^2} (\beta_1 - \beta_{1e})^2 \right)^{-(v+1)/2}$$

biçiminde verilir.[7] Böylece,

$$\sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{s^2 \sum_{i=1}^{n} X_i^2 / n}} (\beta_0 - \beta_{0e}) = t_v$$

ve

$$\frac{(\beta_1 - \beta_{1e})}{\sqrt{s^2 / \sum_{i=1}^{n} (X_i - \overline{X})^2}} = t_v$$

dönüşümleri yapılarak β_0 ve β_1 parametreleri ile ilgili işlemlerde t çizgisinden yararlanılması olanaklıdır.[8] Benzer biçimde, (6) ifadesinin β_0 ve β_1 parametrelerine göre integrali ile σ'nın marjinal sonsal dağılımı,

[7] PRESS, S.J., **Bayesian Statistics: principles, models and applications**, John Wiley, Inc.1989, s.110.
[8] PRESS, S.J., s.114.

$$p(\sigma/X,Y)\alpha\frac{1}{\sigma^{v+1}}\exp\left(-\frac{vs^2}{2\sigma^2}\right)$$

olur. Bu dağılım dönüştürülmüş (inverted) Gamma dağılımı yapısındadır. Bu dağılımın ortalaması ve varyansı,

(8) $$E(\sigma)=s\sqrt{\sqrt{v/2}\,\frac{\Gamma((v-1)/2)}{\Gamma(v/2)}} \qquad v>1$$

(9) $$V(\sigma)=\left(vs^2/(v-2)\right)-\left(E(\sigma)\right)^2 \qquad v>2$$

olarak verilir.[9]

Bu kısma kadar basit doğrusal Bayesci regresyonda β_0 ve β_1 parametrelerinin her ikisinin de aynı anda önsel bilgiye sahip olması durumu incelenmiştir. Ancak sadece β_0 ya da β_1 parametrelerinden herhangi biri de önsel bilgiye sahip olabilir. Bu durumda H_0: $\beta_0 = \beta_{\ddot{0}}$ hipotezi test edilerek kestiricilerin bulunabileceği Rahim tarafından belirtilmiştir.[10]

III. ÇOKLU DOĞRUSAL BAYESCİ REGRESYON

Bir çoklu doğrusal regresyon modeli k, parametre sayısını göstermek üzere

(10) $$Y = X\beta + \varepsilon$$

şeklinde tanımlanır. n gözlem sayısını gösterdiğinde Y, $n\times1$ boyutlu vektör (bağımlı değişken) X, $n\times k$ boyutlu bağımsız değişken matrisi, β, $k\times1$ boyutlu bilinmeyen katsayılar vektörü, ε, $n\times1$ boyutlu ve $E(\varepsilon) = 0$, $E(\varepsilon'\varepsilon)=\sigma^2 I_n$ koşullarına sahip hata vektörüdür. Y değişkeni $X\beta$ ortalamalı ve $\sigma^2 I_n$ varyans-kovaryans matrisi ile çok değişkenli normal dağılıma sahiptir.

[9] ZELLNER, A., s.227.

[10] RAHİM, M.A. "Pre-test estimation of simple linear regression model using prior knowledge of one of the parameters" **Techometrics**, Vol.18, No:4, 1976, s.442.

A. Belirsiz Önsel Kullanılarak Çoklu Doğrusal Bayesci Regresyon

β ve σ parametreleri için belirsiz önsel dağılım $p(\beta,\sigma) = p(\beta/\sigma)p(\sigma)\alpha 1/\sigma$ olacaktır. Olabilirlik (maximum likeli-hood) fonksiyonuyla bileşik sonsal dağılım

$$p(\beta,\sigma/Y)\alpha L(\beta,\sigma)p(\beta,\sigma)$$

(11) $$p(\beta,\sigma/Y)\alpha\frac{1}{\sigma^{n+1}}\exp\left[-\frac{1}{2\sigma^2}\left(AKT(\beta_e)+(\beta-\beta_e)'X'X(\beta-\beta_e)\right)\right]$$

şeklinde ifade edilmektedir.[11]

Bulunan bu bileşik sonsal dağılım k boyutlu çok değişkenli sonsal dağılım yapısındadır. Bunun sonucunda parametre kestirimi

(12) $$\beta_b = \beta_e = (X'X)^{-1}X'Y$$

olarak elde edilmektedir. β için marjinal sonsal dağılım (11)'in σ'ya göre integrali alınması ile bulunmakta ve

(13) $$p(\beta/Y)\alpha\left[AKT(\beta_e)+(\beta-\beta_b)'X'X(\beta-\beta_b)\right]^{-n/2}$$

şeklindeki ifade ile verilmektedir.[12] Elde edilen bu marjinal sonsal dağılım β_b ortalama vektörü ve varyans-kovaryans matrisi ile (n-k) serbestlik derecesi ile Student dağılımına sahiptir.[13]

$$E(\beta/Y) = \beta_b$$
$$V(\beta/Y) = AKO(\beta_e)(X'X)^{-1}$$

Aynı şekilde σ parametresi için marjinal sonsal dağılım da (11)'in integralinin alınması ile elde edilmektedir.

---

11 ZELLNER, A., s.230, William H.Greene, **Econometric Analysis**, 1997, s.312.
12 ZELLNER, A., s.231.
13 ZELLNER, A., s.232.

(14) $$p(\sigma/Y)\alpha\frac{1}{\sigma^{n+1}}\exp\left[-\frac{1}{2\sigma^2}AKT(\beta_e)\right]$$

Elde edilen bu marjinal sonsal dağılım dönüştürülmüş Gamma dağılımı yapısındadır. Dağılımın ortalama ve varyansı

$$E(\sigma/Y)=\sqrt{\sqrt{AKO(\beta_e)(n-k)/2}\,\frac{\Gamma(n-k-1)/2}{\Gamma(n-k)/2}}$$

$$V(\sigma/Y)=AKT(\beta_e)/[((n-k)/2)-(E(\sigma/Y))]^2$$

olarak verilmektedir.[14]

B. Bayesci Regresyon Kestiricilerinin Özellikleri

β ve σ için belirsiz önsel dağılım kullanıldığında bulunacak Bayesci regresyon kestiricisi en çok olabilirlik kestiricileri ile aynı olmaktadır. Bunun nedeni belirsiz önsel dağılım kullanımında olabilirlik fonksiyonunun önsel dağılıma karşı daha etkili olmasıdır. Parametre kestirimlerinin güven aralıkları ile klasik güven aralıklarının hesaplanması aynıdır. Ancak bu aralıkların yorumu farklı olmaktadır.

Bayesci regresyon kestiricisi doğrusal olmayan yanlı bir kestiricidir. Kestiricinin yanlılığı

$$E(\beta_b)=\beta-(X'X+A)^{-1}A(\beta-\beta_ö)$$

eşitliğinden

$$E(\beta_b)-\beta=\beta-(X'X+A)^{-1}A(\beta-\beta_ö)-\beta$$

olarak bulunur. Burada $\beta-\beta_ö=\beta^*$, $(X'X+A)^{-1}=\mu$ alınarak

(15) $$Y_{an}(\beta_b)=-\mu A\beta^*$$

şeklinde ifade edilmektedir.[15]

---

[14] ZELLNER, A., "Bayesian statistics in econometrics", **Bayesian Statistics**, 2, s.575.
[15] VINOD, H.D. and ULLAH, A., **Recent advances in Regression Methods**, Murcel Dekker, Inc., New York, 1981, s.200.

Bayesci regresyon kestiricilerinin varyansı ise,

$$V(\beta_{bi}) = E(\beta_{bi} - E\beta_{bi})^2$$

olarak tanımlanmaktadır. Bu nedenle öncelikle $\beta_b - E\beta_b$ bulunmalıdır.

$$\beta_b - E(\beta_b) = (I - \mu A)(\beta - \beta_ö)$$

$$(\beta_b - E\beta_b)(\beta_b - E\beta_b)' = (I - \mu A)(\beta_e - \beta)(\beta_e - \beta)'(I - \mu A)'$$

Bunun sonucunda bulunacak parametre kestirimlerinin varyansı

(16) $$V(\beta_b) = \sigma^2 \mu X'X \mu$$

olmaktadır.[16]

Doğrusal regresyon modelinde kesin olarak doğrusal bir ilişki varsa bu ilişkiyi önsel bilgi olarak almak sezgisel bir yaklaşım olmamaktadır. Buna karşın doğrusal ilişkiyi belirleyen önsel bilgi güçlü olsa bile elde edilen model çok sık olmamakla birlikte doğrusal olmayabilir. Böylece önsel varsayımlar her zaman regresyon modelinin uygunluğu için gerekli olan varsayımları sağlamayacağı belirtilmiştir.[17]

SONUÇ VE ÖNERİLER

Bu çalışmada giriş bölümünde belirtilen amaç doğrultusunda regresyon çözümlemesinde Bayesgil yaklaşımının nasıl kullanıldığı araştırılmıştır.

Bayes kestiricisinin temel özelliklerinin özet olarak açıklanmasından sonra basit ve çoklu doğrusal regresyonda Bayes yaklaşımı teorik olarak irdelenmiştir. Yapılan çalışma sonucunda ön kestirim açısından Bayes kestiricilerinin değişim aralığının diğerlerine göre daha dar olacağı görülmüştür. Ancak belirlenen önsel parametre değerlerine göre katsayıların işaretçe ve/veya değerce etkilendikleri söylenebilir.

Regresyonda Bayes yaklaşımı kullanılmak istendiğinde, önsel bilginin analize başlamadan önce irdelenmesi gerekir. Ayrıca iyi belirlenmiş önsel bilgiler olmasına karşı regresyon modelinde karşılaşılan varsayımlardan sapmaların sonsal analizlerdeki etkilerinin araştırılması kanısındayız.

---

[16] VINOD, H.D. and ULLAH, A., s.202.

[17] GOLDSTEİN, M., s.57.

YARARLANILAN KAYNAKLAR

GOLDSTEIN, M., "Bayesian Analysis of Regression Problems", **Biometrika**, 63(1) 1976, s.51-58.

GUNST, R.F., "Regression Analysis with Multicollinear Predictor Variables: Definition, Detection and Effects", **Commun. Statist Theory Meth.** 12(19), 1983, s.2217-2260.

LINDLEY, D.V. and SMITH, A.F.M., "Bayes Estimates for the Linear Model", **J.R.Statistical Soc.**, 1972, s.1-41.

PRESS, S.J., **Bayesian Statistics: Principles, Models and Applications**, John Wiley, Inc.1989.

RAHİM, M.A., "Pre-test Estimation of Simple Linear Regression Model Using Prior Knowledge of One of the Parameters" **Techometrics**, Vol.18, No:4, 1976, s.439-445.

VINOD, H.D. and ULLAH, A., **Recent Advances in Regression Methods**, Murcel Dekker, Inc., New York, 1981.

ZELLNER, A., **An introduction to Bayesian Inference in Econometrics:** John Wiley, Inc.1971.

ZELLNER,A.,"Bayesian Statistics in Econometrics",**Bayesian Statistics**, 2, 1985, s.571-586.

GREENE,W.A., **Econometric Analysis**, New York University, Third Edition,1997.