

## GIBBS SAMPLING APPROACH TO VARIABLE SELECTION IN LINEAR REGRESSION WITH OUTLIER VALUES

Atilla YARDIMCI\*

Türkiye Odalar ve Borsalar Birliği, Bilgi Hizmetleri Dairesi, 06640,  
Ankara, TÜRKİYE, e-mail: atilla@tobb.org.tr

Aydın ERAR

Mimar Sinan Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü,  
34349, İstanbul, TÜRKİYE

### ABSTRACT

In this study, Gibbs sampling has been applied to the variable selection in the linear regression model with outlier values. Gibbs sampling has been compared with classical variable selection criteria by using dummy data with different  $\beta$  and priors.

**Key Words:** Bayesian variable selection, prior distribution, Gibbs sampling, Markov Chain Monte Carlo, outlier values, entropy

## AYKIRI DEĞER VARLIĞINDA DOĞRUSAL REGRESYONDA DEĞİŞKEN SEÇİMİNE GIBBS ÖRNEKLEMESİ YAKLAŞIMI

### ÖZET

Bu çalışmada, aykırı değerlerin olduğu doğrusal regresyon modelinde değişken seçimine, Gibbs örnekleme yaklaşımı uygulanmıştır. Klasik değişken seçimi ölçütleri ile Gibbs örnekleme yaklaşımı, değişik  $\beta$  ve önsel yapıları için yapay veriler üzerinden karşılaştırılmıştır.

**Anahtar Kelimeler:** Bayesci değişken seçimi, önsel dağılım, Gibbs örnekleme, Markov Chain Monte Carlo, aykırı değer, entropi

### 1. GİRİŞ

Gözlem sayısı  $n$ , bağımlı değişken sayısı  $k$  iken çoklu doğrusal regresyon denklemi,

$$Y = X\beta + \varepsilon \quad [1]$$

biçiminde tanımlanır.  $Y_{n \times 1}$ , bir raslantı değişkeni olan yanıt vektörü;  $X_{n \times k'}$ ,  $k' = k+1$  iken  $k'$  ranklı stokastik olmayan girdi matrisi;  $\beta_{k' \times 1}$ , bilinmeyen katsayılar vektörü;  $\varepsilon_{n \times 1}$ ,  $E(\varepsilon) = 0$ ,  $E(\varepsilon\varepsilon') = \sigma^2 I_n$  koşullarını sağlayan yanılgi vektörüdür. Kestirim ve önkestim (prediction) denklemleri ise, artıklar vektörü  $e$  iken,  $Y = X\hat{\beta} + e$  ve  $\hat{Y} = X\hat{\beta}$  olarak tanımlanır.  $\hat{\beta}$  kestiricisi, En Küçük Kareler (EKK) ya da  $\varepsilon$ 'nın Normal dağılımlı olduğu varsayımı altında En Çok Olabilirlik (EÇO) yöntemleri kullanılarak elde edilir.

Olabilirlik fonksiyonunun enbüyüklenmesi ile elde edilen  $Y$ 'nin birleşik dağılımı,

$$L(\beta, \sigma / y) = f(y / \beta, \sigma, X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)\right) \quad [2]$$

### 1. INTRODUCTION

A multiple linear regression equation with  $n$  observations and  $k$  independent variables can be defined as

$$Y = X\beta + \varepsilon \quad [1]$$

where  $Y_{n \times 1}$ , is a response vector;  $X_{n \times k'}$ , is non-stochastic input matrix with rank  $k'$ , where  $k' = k+1$ ;  $\beta_{k' \times 1}$ , is as unknown vector of coefficients;  $\varepsilon_{n \times 1}$  is an error vector with  $E(\varepsilon) = 0$ ,  $E(\varepsilon\varepsilon') = \sigma^2 I_n$ . Where  $e$  is the residual vector, estimation and prediction equations can be defined as,  $Y = X\hat{\beta} + e$  and  $\hat{Y} = X\hat{\beta}$ .  $\hat{\beta}$  estimator can be obtained using the Least Square Error (LSE) method or the Maximum Likelihood (MC) method under the assumption that  $\varepsilon$  is normally distributed.

The joint distribution of  $Y$  is obtained by maximizing the likelihood function as follows

$$L(\beta, \sigma / y) = f(y / \beta, \sigma, X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)\right) \quad [2]$$

biçiminde yazılır. Buradan,  $\hat{\beta}$  ve  $\hat{\sigma}^2$  kestiricileri,

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad [3]$$

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) / (n - k') = \text{AKT}(\hat{\beta}) / (n - k')$$

olarak elde edilir. Burada  $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$  ve  $E(\hat{\sigma}^2) = \sigma^2$  dır (3).

Regresyon analizlerinde değişken seçimi, şu amaçlarla yapılır: Daha düşük maliyetle kestirim ya da önkestim yapmak, bilgi verici olmayan değişkenlerin çıkarılması ile daha küçük yanılğı kareler ortalamalı kestirimler elde etmek, yüksek derecede ilişkili değişkenlerin varlığında regresyon katsayılarının küçük standart yanılğı ile kestirimini elde etmek.

## 2. KLASİK DEĞİŞKEN SEÇİMİ ÖLÇÜTLERİ

Doğrusal regresyonda değişken seçimi işlemi için sık sık, adimsal ve tüm olası altküme tekniklerine başvurulur. Tüm olası altkümeler tekniği, artık kareler ortalaması, çoklu belirtme katsayısı,  $F_p$  istatistiği,  $C_p$  ve  $\text{PRESS}_p$  gibi seçim ölçütleri ile birlikte kullanılır (9). Son dönemlerde bilgi miktarına bağlı olarak değişken seçimi yapılmasına olanak tanıyan AIC, AICc ve BIC ölçütleri de kullanılmıştır. Altkümedeki parametre sayısı  $p$  olmak üzere AIC ve AICc,

$$\text{AIC} = n \log(\text{AKO} + 1) + 2p \quad [4]$$

$$\text{AICc} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p-2} \quad [5]$$

eşitlikleri ile tanımlanır.  $p$  alt kümesi için olabilirlik fonksiyonu  $f(\mathbf{Y}/\hat{\beta}_p, \sigma_p, \mathbf{X})$  biçiminde tanımlanmak üzere BIC kriteri,

$$\text{BIC} = \log f(\mathbf{Y}/\hat{\beta}_p, \sigma_p, \mathbf{X}) - \frac{p}{2} \log n \quad [6]$$

ile tanımlanır ve (6) eşitliğini en büyük yapan altküme seçmeyi amaçlar (2).

Ayrıca, sonsal model olasılıkları ve risk ölçüleri de zaman zaman değişken seçiminde kullanılmıştır. Sonsal model olasılıkları, Wassermann (12) tarafından verilen logaritmik yaklaşım kullanılarak, eşit önsel altküme olasılıkları olduğu varsayımı altında tüm altkümeler için hesaplanmıştır. Risk kriteri ise Toutenburg (10) tarafından;  $\mathbf{V}(\hat{\beta})$ ,  $\hat{\beta}$  için varyans-kovaryans matrisi iken,

$$\text{Risk} = \sigma^2 \text{tr}(\mathbf{V}(\hat{\beta})) \quad [7]$$

eşitliği yardımıyla değişken seçimine uyarlanmıştır. Tüm olası altkümeler için bulunan (7) değeri, yapılacak kestirimlerin riskine karşılık gelmektedir. Amaç, riski en küçük modeli elde etmektir. Ancak bu noktada risk ile sonsal olasılık değerlerinin, yorumlama aşamasında birlikte değerlendirilmesi gerektiği Yardımcı (13) tarafından belirtilmiştir. İdeal durumda sonsal olasılığı yüksek ve riski düşük altkümeye ulaşılmak istenir. Buna karşın her ikisinin de sağlandığı duruma ulaşmak güç

Where,  $\hat{\beta}$  and  $\hat{\sigma}^2$  estimators are defined by

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad [3]$$

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) / (n - k') = \text{SSE}(\hat{\beta}) / (n - k')$$

Where  $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$  and  $E(\hat{\sigma}^2) = \sigma^2$  (3).

The aims of variable selection in regression analysis are; to make estimation and prediction in a lower cost, to obtain estimation with mean square error by subtractions of the noninformative variables, to obtain regression coefficient with estimation of low standard error.

## 2. CLASSICAL VARIABLE SELECTION CRITERIA

For variable selection procedure in linear regression, stepwise and all possible subset methods are frequently used. The all possible subsets method include mean square error, coefficient of determination,  $F_p$  statistics,  $C_p$  and  $\text{PRESS}_p$  (9). In recent years AIC, AICc and BIC criteria which make variable selection possible to do information amount are used. Where  $p$  is the parameter number in subset, AIC and AICc are,

$$\text{AIC} = n \log(\text{MSE} + 1) + 2p \quad [4]$$

$$\text{AICc} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p-2} \quad [5]$$

If  $f(\mathbf{Y}/\hat{\beta}_p, \sigma_p, \mathbf{X})$  is the likelihood function for the subset  $p$ , the criterion BIC aims to select the subset which maximizes the equation (6) given below (2)

$$\text{BIC} = \log f(\mathbf{Y}/\hat{\beta}_p, \sigma_p, \mathbf{X}) - \frac{p}{2} \log n \quad [6]$$

Posterior model probabilities and risk criteria have also been used occasionally in variable selection. Posterior model probabilities have been assessed for all the subsets using a logarithmic approach given by Wassermann (12) under the assumption of equal prior subset probabilities. The risk criterion has been adapted to variable selection with the help of the equation

$$\text{Risk} = \sigma^2 \text{tr}(\mathbf{V}(\hat{\beta})) \quad [7]$$

given by Toutenburg (10), where  $\mathbf{V}(\hat{\beta})$  is the variance-covariance matrix of  $\hat{\beta}$ . The rate (7) found for all the probable subsets corresponds to the risk of the estimations. The aim is to get the model with the smallest risk.

On the other hand, Yardımcı (13) claims that the rates of risk and posterior probability should be evaluated together during the interpretation. Ideally, the aim is to find a subset that has higher posterior probability and lower risk. Since it is hard to satisfy this condition, preferences are made within acceptable limits (13). Furthermore, the Risk Inflation Criterion (RIC) has been found for all probable subsets with the help of the

olduğundan kabul edilebilir sınırlar içerisinde tercihler yapılabilir (13).

Ayrıca Risk şişme ölçütü, Foster ve George (4) tarafından verilen,

$$RIC = AKT_p + 2p AKO_p \log(k) \quad [8]$$

eşitliği yardımıyla tüm olası altkümeler için bulunmuştur. Burada  $AKT_p$  ve  $AKO_p$ , sırasıyla, Artık Kareler Toplamına ve Artık Kareler Ortalamasına karşılık gelmektedir. Doğru açıklayıcıların seçilmesi ile RIC değerinin enküçükleneceği ve böylece doğru değişkenlerin seçilebileceği belirtilmiştir (13).

### 3. GIBBS ÖRNEKLEMESİ YAKLAŞIMI

Bayesci değişken seçimi yaklaşımlarında sonsal olasılık ya da dağılımlarının belirlenmesi üzerine yoğunlaşmıştır. Ancak kimi durumlarda özellikle sonsal momentlerin hesaplanması için gerekli olan integrallerin analitik olarak çözümleri mümkün olmamakta ya da güç olmaktadır. Bu durumlarda, Markov zinciri türetme ve yakınsaklık özellikleri ile sonsal dağılıma erişme biçiminde yaklaşımlar verilmektedir (15). Böylece Monte Carlo ve stokastik benzetim tekniklerinin bir arada kullanılması ile zincirleme veri çoğaltma adı verilen yaklaşımlar geliştirilmiştir. Bu yaklaşımlar genel olarak Markov Zinciri Monte Carlo (MCMC) başlığı altında toplanmaktadır.

MCMC yaklaşımları aracılığıyla sonlu sayıda gözlem değeri kullanılarak, sonsuz sayıda veri elde etmek mümkündür. Böylece çözümü analitik olarak zor olan bazı problemlerin, benzetim teknikleri ve bilgisayar yazılımları sayesinde hızlı biçimde çözülmesi mümkün olmaktadır.

MCMC yaklaşımının amacı,  $\theta$  parametre uzayında rasgele yürüyüş oluşturup hedef olan sonsal dağılıma yakınsamaktır (1). MCMC yaklaşımlarında sonsal dağılıma yakınsamak amacıyla türetilen  $\theta^{j+1}$ , önceki  $\theta^j$  değerine bağlı olmaktadır. Bu amaçla örnekleme,  $g(\theta/\theta^j)$  Markov geçiş dağılımından yapılmaktadır (6).

Hedeflenen sonsal dağılım için geçiş dağılımından örneklem almak ve bunu işlemek için birçok yöntem bulunmaktadır. Sonsal dağılımdan alınan bu örneklemelerin Markov zinciri özelliğini göstermesi amacıyla Metropolis algoritmaları ile Gibbs örnekleme yaklaşımları kullanılmaktadır.

Gibbs örneklemede,  $x^{(1)}, x^{(2)}, \dots, x^{(t)}$  ile gösterilen Markov zincirinin,  $p(x)$  dağılımına yakınsaması için  $p(x_i / x_{-i})$  koşullu dağılımından türetilen örneklemeler kullanılmaktadır. Burada  $x_{-i}$ ,  $x_i$  dışındakileri değişkenleri simgelemektedir. Böylece Gibbs örneklemesinin türetilme süreci,  $j$  iterasyon sayısı olmak üzere aşağıda verilmiştir (5).

1.  $j = 1$ , başlangıç değeri  $(x_1^0, x_2^0, \dots, x_k^0)$  olarak alınır.

equation defined by Foster and George (4) given below

$$RIC = SSE_p + 2p MSE_p \log(k) \quad [8]$$

where  $SSE_p$  and  $MSE_p$  are the sum of squares error and mean square error for the subset  $\pi_i$  respectively. It is claimed that RIC will produce the lowest values and thus give more consistent deductions when trying to select the correct variables (13).

### 3. THE GIBBS SAMPLING APPROACH

In Bayesian variable selection approaches, it is concentrated on the determination of posterior probabilities and distributions. However in some cases, the analytic analysis of integrals needed for the calculation of posterior moments is difficult or even impossible. In such cases, by the creation of Markov's chain and convergence characteristics approaches to posterior distribution are obtained (15). Thus, by using Markov chain and stochastic stimulation techniques together, Markov Chain Monte Carlo (MCMC) approaches have been developed. These approaches are generally gathered together under Markov Chain Monte Carlo method.

Using limited observation values in MCMC approachment, it is possible to obtain unlimited data. Thus by using simulation techniques and computer software, it is possible to easily solve the problems hard to be calculated in analytic analysis.

The aim of the MCMC approach is to create a random walk which converges to the target posterior distributions on the  $\theta$  parameters space (1). In the MCMC approach,  $\theta^{j+1}$  created for the convergence to the posterior distribution, depends on the previous  $\theta^j$  value. For this purpose sampling is done by using the Markov transition distributions  $g(\theta/\theta^j)$  (6).

For a posterior distribution, there are various methods that take samples and use from transition distribution Metropolis algorithms and Gibbs sampling approaches are used to show how these samples taken from a posterior distribution achieve the characteristic of a Markov chain.

In Gibbs sampling, the Markov Chain is defined as  $x^{(1)}, x^{(2)}, \dots, x^{(t)}$ . For the convergence  $p(x)$  distribution, samples created from  $p(x_i / x_{-i})$  conditional distribution are used. Here  $x_{-i}$ , represents variables different from  $x_i$ . Being  $j$  the iteration number, the creation process of Gibbs sampling is below; (5)

1.  $j = 1$  the initial value is  $(x_1^0, x_2^0, \dots, x_k^0)$

$$2. \quad x_1^{(j)} \sim p(x_1 / x_2^{(j-1)}, x_3^{(j-1)}, \dots, x_k^{(j-1)})'$$

$$x_2^{(j)} \sim p(x_2 / x_1^{(j)}, x_3^{(j-1)}, \dots, x_k^{(j-1)})'$$

...

$$x_k^{(j)} \sim p(x_k / x_1^{(j)}, x_2^{(j)}, \dots, x_{k-1}^{(j)})'$$

3.  $j=j+1$  alınır ve 2. adıma gidilir.

Yakınsaklık sağlandığında  $x^{(j)}$  değerleri  $p(x_1, x_2, x_3, \dots, x_k)$ ' dağılımdan alınmış değerlere karşılık gelmektedir.

Doğrusal regresyonda Gibbs örneklemesine dayanan birçok Bayesci değişken seçimi yaklaşımları bulunmaktadır. Burada, çalışmada kullanılan, Kuo ve Mallick (8) tarafından verilen yaklaşım açıklanacaktır.

Gibbs örneklemesinde  $2^k-1$  sayıda açıklayıcı değişken altkümesi için, bunların modeldeki durumlarını tanımlayacak  $\gamma=(\gamma_1, \gamma_2, \dots, \gamma_k)$ ' göstermelik değişkeni kullanılmaktadır. İlgilenilen değişken modelde ise  $\gamma=1$ , modelde değilse  $\gamma=0$  değerini almaktadır.  $\gamma$  'nın değeri bilinmediğinden Bayes sürecine eklenmesi gerekmektedir. Bu amaçla değişken seçimi işleminde kullanılacak bileşik önsel,

$$p(\beta, \sigma, \gamma) = p(\beta / \sigma, \gamma) p(\sigma / \gamma) p(\gamma)$$

biçiminde tanımlanır. Burada,  $\beta$  katsayıları için tanımlanan önselin,

$$\beta \sim N(\beta_0, D_0)$$

olduğu varsayılmıştır (8). Bununla beraber  $\sigma$  için,

$$p(\sigma) \propto \frac{1}{\sigma} \quad [9]$$

ile ifade edilen bilgi vermeyen önsel tanımlanır.  $\gamma$  için önsel de,

$$\gamma_j \sim \text{Bernoulli}(1, p) \quad [10]$$

biçiminde verilir. Yapılan bu tanımlamalar sonucunda  $p(\gamma/y)$  marjinal sonsal dağılımı, değişken seçimi için istenen tüm bilgileri içermiş olduğundan yapılan çıkarsamalarda bu dağılım üzerine yoğunlaşmaktadır. Kuo ve Mallick (8) tarafından, Gibbs örneklemesi ile  $p(\gamma/y)$  'nin kestirilebileceği belirtilmiştir. Ayrıca bu işlem için gerekli olan  $\beta^0$  ve  $\sigma^0$  başlangıç değerleri için  $\gamma_j=1$  ( $j=1,2,\dots,k$ ) olduğu durumda, tam model için bulunacak EKK kestirimlerinin kullanılabilirliği de ifade edilmiştir.  $\gamma^0$  için ise başlangıçta  $(1,1,\dots,1)$  alınabilir. Böylece  $\gamma$  'nın marjinal sonsal yoğunluğu,

$$p(\gamma_j / \gamma_{-j}, \beta, \sigma, Y) = B(1, \tilde{p}_j) \quad , \quad j=1,\dots,k$$

eşitliği ile verilebilir. Burada  $\gamma_{-j}=(\gamma_1, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_k)$  olmak üzere,

$$2. \quad x_1^{(j)} \sim p(x_1 / x_2^{(j-1)}, x_3^{(j-1)}, \dots, x_k^{(j-1)})'$$

$$x_2^{(j)} \sim p(x_2 / x_1^{(j)}, x_3^{(j-1)}, \dots, x_k^{(j-1)})'$$

...

$$x_k^{(j)} \sim p(x_k / x_1^{(j)}, x_2^{(j)}, \dots, x_{k-1}^{(j)})'$$

3.  $j=j+1$  then return to the 2nd step.

When convergence is obtained,  $x^{(j)}$  values are taken from the  $p(x_1, x_2, x_3, \dots, x_k)$ ' distribution.

In linear regression, there are many Bayesian selection methods that rely on Gibbs sampling. Here the approach of Kuo and Mallick (8) that is used in this study will be explained.

In Gibbs sampling, a dummy variable given by  $\gamma=(\gamma_1, \gamma_2, \dots, \gamma_k)$ ' is used to define the position of the  $2^k-1$  independent variable subsets in the model. If the variable considered is in the model, then  $\gamma=1$ , if not  $\gamma=0$ . Since the  $\gamma$  rate is not known, it has to be added to the Bayes process. Thus, the joint prior that will be used in the variable selection process is defined as

$$p(\beta, \sigma, \gamma) = p(\beta / \sigma, \gamma) p(\sigma / \gamma) p(\gamma)$$

Here, the prior defined by Kuo and Mallick (8) for the  $\beta$  coefficients is assumed to be,

$$\beta \sim N(\beta_0, D_0)$$

In addition, a noninformative prior for  $\sigma$  is defined by

$$p(\sigma) \propto \frac{1}{\sigma} \quad [9]$$

The noninformative prior for  $\gamma$  is given as

$$\gamma_j \sim \text{Bernoulli}(1, p) \quad [10]$$

The deductions concentrate on these distributions since the marginal posterior distribution  $p(\gamma/y)$  contains all the information that is needed for the variable selection. In Kuo and Mallick (8), it has been shown that  $p(\gamma/y)$  can be estimated using Gibbs sampling. Moreover, the least square estimation for the full model can be used when the starting values of  $\beta^0$  and  $\sigma^0$ , that are needed for this process, are  $\gamma_j=1$  ( $j=1,2,\dots,k$ ). Initially we may take  $(1,1,\dots,1)$  for the  $\gamma^0$ . In this way, the marginal posterior density of  $\gamma$  can be given as,

$$p(\gamma_j / \gamma_{-j}, \beta, \sigma, Y) = B(1, \tilde{p}_j) \quad , \quad j=1,\dots,k$$

where, if  $\gamma_{-j}=(\gamma_1, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_k)$ , then

$$\tilde{p}_j = \frac{c_j}{(c_j + d_j)} \quad [11]$$

$$c_j = p_j \exp\left(-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{Xv}_j^*)' (\mathbf{Y} - \mathbf{Xv}_j^*)\right)$$

$$\tilde{p}_j = \frac{c_j}{(c_j + d_j)} \quad [11]$$

$$c_j = p_j \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{Xv}_j^*)(\mathbf{Y} - \mathbf{Xv}_j^*)\right)$$

$$d_j = (1 - p_j) \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{Xv}_j^{**})(\mathbf{Y} - \mathbf{Xv}_j^{**})\right)$$

olarak tanımlanır.  $\mathbf{v}_j^*$  vektörü, j inci elemanı  $\beta_j$  ile değiştirilen  $\mathbf{v}$  kolon vektörü,  $\mathbf{v}_j^{**}$  ise j inci elemanı 0 ile değiştirilen  $\mathbf{v}$  vektördür. Böylece  $c_j$ , j inci açıklayıcının kabul edilmesine,  $d_j$  reddedilmesine,  $p_j$  modelde bulunma olasılıklarına karşılık gelmektedir (8).

Bir sisteme özgü belirsizlik ve bilinmezliğin, bir düzendeki karmaşa miktarının matematiksel ölçüsü entropi olarak tanımlanmaktadır (11). Entropi özellikleri Gray (7) tarafından verilmiştir. Deneyden önce bulunan entropi, deney sonuçları ile ilgili belirsizliğin, deney sonrasında bulunan entropi ise deneyden elde edilen bilgi miktarını vermektedir. Böylece Gibbs yaklaşımı uygulandığında, entropi değerleri bulunarak modeldeki bağımsız değişkenlerin beklenen bilgi miktarlarının da elde edilmesi olanaklıdır.

Bu amaçla tam modelde bulunan bağımsız değişkenlerin sonsal ortalama ve varyans,

$$\beta_G = \frac{1}{T} \sum_{i=1}^T \beta_{(i)}, \quad v(\beta_G) = \frac{1}{T-1} \left( \sum_{i=1}^T \beta_{(i)}^2 - \left( \sum_{i=1}^T \beta_{(i)} \right)^2 \right)$$

eşitlikleri yardımıyla elde edilir. Eşitlikte T, Gibbs örnekleme tekrar sayısını,  $\beta_{(i)}$ , i. adımda Gibbs algoritmasına uygun türetilen parametre değeridir. Entropi değerleri, değişkenlerin modelde bulunma olasılıklarını veren (8) eşitliği kullanılarak, bu çalışmada,

$$H(\beta_j) = - \sum_{i=1}^T \tilde{p}_{j(i)} \log(\tilde{p}_{j(i)})$$

ile hesaplanacaktır.

#### 4. SEÇİM ÖLÇÜTLERİNİN AYKIRI DEĞER İÇEREN YAPAY VERİLER ÜZERİNDEN KARŞILAŞTIRILMASI

Seçim ölçütlerinin karşılaştırılması amacıyla, Kuo ve Mallick (8) ile George ve McCulloch (6) tarafından yapılan çalışmalara paralel olarak, n=20 ve k=5 olacak şekilde, yapay veriler türetilmiştir. Böylece verilen tepkilerin gözlenmesi mümkün olmuştur.  $X_j$  değişkenlerinin dağılımlarının, N(0,1) ile normal dağılımlı oldukları yönünde varsayım yapılmıştır. Modelde bulunan değişkenlerin anlamlılık düzeylerinin yapılacak seçim işlemindeki etkilerini görmek amacıyla ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ ) için 4 farklı  $\beta$  yapısı belirlenmiştir. Veri türetilmesi sırasında kullanılan  $\beta$  yapısı:

$$d_j = (1 - p_j) \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{Xv}_j^{**})(\mathbf{Y} - \mathbf{Xv}_j^{**})\right)$$

Here, the vector  $\mathbf{v}_j^*$  is the column vector of  $\mathbf{v}$  with the  $j^{\text{th}}$  entry replaced by  $\beta_j$ , similarly,  $\mathbf{v}_j^{**}$  is the column vector of  $\mathbf{v}$  with the  $j^{\text{th}}$  entry replaced by 0. Thus  $c_j$ , is a probability for the acceptance of the  $j^{\text{th}}$  variable and  $d_j$  a probability for its rejection (8).

The mathematical degree of uncertainty, uninformation and the amount of complexity in a process is described as entropy (11). Gray (7) has described the properties of entropy. Entropy found before experiment represents the uncertainty about the experiment results, the entropy calculated after the experiment represents the amount of information obtained from the experiment. Therefore by applying the Gibbs approach, using the entropy values, it is possible to obtain expected information amount from the independent variables in the model.

For this purpose, posterior mean and variance of independent variables in the full model B are calculated as follows;

$$\beta_G = \frac{1}{T} \sum_{i=1}^T \beta_{(i)}, \quad v(\beta_G) = \frac{1}{T-1} \left( \sum_{i=1}^T \beta_{(i)}^2 - \left( \sum_{i=1}^T \beta_{(i)} \right)^2 \right)$$

In the equation T, while Gibbs sampling represents the repeat number,  $\beta_{(i)}$  is the parameter value in the step of i simulated by using the Gibbs algorithm. In this study, the entropy values will be calculated by using the equation (8) that gives the inclusion probability of variables in the model described below.

$$H(\beta_j) = - \sum_{i=1}^T \tilde{p}_{j(i)} \log(\tilde{p}_{j(i)})$$

#### 4. THE CORRESPONDENCE OF SELECTION CRITERIA BY USING DUMMY DATA INCLUDING OUTLIER VALUES

For the correspondence of selection criteria, we have simulated dummy data as n=20 and k=5 in parallel to Kuo and Mallick's (8) and George and McCulloch's (6) studies. By this way it has been possible to obtain the reactions. The distribution of the variables  $X_j$  has been assumed to be normal with N(0,1). Four types of  $\beta$  structure have been determined to observe the effects of the significance level of variables ( $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ ) in the model. The  $\beta$  structure that has been used is as follows;

$$\text{Beta1} : (3,2,1,0,0)$$

$$\text{Beta2} : (1,1,1,0,0)$$

$$\text{Beta3} : (3,-2,1,0,0)$$

$$\text{Beta4} : (1,-1,1,0,0)$$

Beta1 için (3,2,1,0,0)      Beta2 için (1,1,1,0,0)  
 Beta3 için (3,-2,1,0,0)      Beta4 için (1,-1,1,0,0)

biçiminde belirlenmiştir. Her bir  $\beta$  yapısında geçerli olmak üzere aykırı değer oluşturmak amacıyla bağımlı değişkenin 10. gözlemi gerçek değerinin 8 katı olacak biçimde yapay veriler üretilmiştir. Ayrıca aykırı değerlerin sonuç üzerindeki etkilerini görmek amacıyla, aykırı değer olmadan aynı koşullar altında yapay veriler de üretilmiştir.

Bunun yanında katsayıların üretilmesi sırasında kullanılan  $\sigma^2$ , 1.0 olarak seçilmiştir. Çalışmada önsel bilgi düzeylerinin sonuç üzerindeki etkilerini gözlemlemek amacıyla, Gibbs örnekleme için  $D_0=1$ , 16 ve eşit değişken önsel olasılıkları tanımlanmıştır. Ayrıca hatalar ile ilgili  $\varepsilon \sim N(0, \sigma^2)$  varsayımı yapılmıştır.  $k=5$  açıklayıcı değişken için tüm olası altkümelerin listesi Ek-1 de verilmiştir.

Karşılaştırmalar sırasında Yardımcı (13) tarafından, Bayesci değişken seçimi işlemleri için özel olarak geliştirilen ve BARVAS adı verilen özel bilgisayar yazılımı kullanılmıştır. BARVAS yazılımı, doğrusal regresyonda Bayesci regresyon ve değişken seçimi işlemleri için kullanılabilen, ayrıca değişken seçimi için klasik değişken seçimi ölçütlerinin de bir arada uygulanabildiği bir paket yazılımdır (14).

#### 4.1. Klasik Değişken Seçimi

Klasik değişken seçimi karşılaştırmaları yapılırken I. Grup altında, en bilinen ölçütler olan  $C_p$ , AICc, BIC ve Press yer alırken, istatistiksel olarak kullanılan ancak uygulamada henüz tercih edilmeyen ölçütler olan Risk, RIC II. Grup altında toplanmışlardır. Aykırı değer varlığında ve bozulmanın olmadığı durumlarda yapay veriler için klasik değişken seçimi sonuçları Çizelge 4.1'de özetlenmiştir.

**Table 4.1.** The results of classical variable selection in the presence / absence of outlier value  
**Çizelge 4.1.** Aykırı değer olduğu ve olmadığı durumlarda klasik değişken seçim sonuçları

Beta	Aykırı değer olmadığı durum/ Outlier value absence						Aykırı değer olduğu durum/ Outlier value presence					
	I. Grup/ Group I				II. Grup/ Group II		I. Grup / Group I				II. Grup/ Group II	
	$C_p$	AICc	BIC	Press	Risk	RIC	$C_p$	AICc	BIC	Press	Risk	RIC
1	5	5	5	5	5	5	5	1	5	5	1	5
	10	2	10	26	10	10	10	2	10	10	2	10
	26	10	26	10	26	26	26	5	26	26	14	26
2	5	5	5	5	5	5	1	1	5	1	1	5
	10	1	10	26	1	10	2	2	1	2	6	2
	26	6	26	10	6	26	6	6	2	29	2	1
3	5	5	5	5	5	5	26	2	26	5	31	26
	10	10	10	10	10	26	5	31	21	10	3	21
	26	26	26	26	26	10	21	1	5	26	28	5
4	5	5	5	5	7	10	6	6	6	6	7	5
	26	6	26	26	6	5	5	7	5	25	6	6

I. Grup seçim ölçütlerinin aykırı değer varlığında Beta 1 durumu için etkilenmedikleri söylenebilir. Beta 1 durumu için II. Grupta yer alan RIC, I. Grup ile uyumlu

In all the beta structures above, the 10th observation of dependent variable has been obtained as the 8 times of real values. Thus outlier values have been created. Moreover for the same structures dummy data has been obtained without the existence of outlier value.

During the creation of coefficients, it has been chosen as  $\sigma^2=1.0$  In the study for the observation of prior information levels on the results, for Gibbs sampling  $D_0=1$ , 16 and equal prior probability for variables have been defined. Furthermore, the assumption  $\varepsilon \sim N(0, \sigma^2)$  is made for the errors. The list of all possible subsets for 5 independent variables is given in appendix 1.

During the comparison of these approaches, the BARVAS (Bayesian Regression-Variable Selection) computer program, which is developed especially for the Bayesian variable selection process by Yardımcı(13) is used. The BARVAS program can be used for Bayesian regression and variable selection in linear regression, and it is a packet program where the classical variable selection criteria for variable selection could also be used together (14).

#### 4.1. Classical Variable Selection

In the comparison of classical variable selection, the well-known criteria ( $C_p$ , AICc, BIC, Press) will be referred to as group 1, and posterior model probability, Risk RIC as Group 2. In the situations when outlier value absent or present, the results of classical variable selection for dummy data have been shown on Table 4.1.

The selection criteria for Group I in the outlier value presence have not been affected for the Beta 1 situation. For Beta 1 structure, RIC that has been in Group II has

sonuç vermiştir. Buna karşın Risk ölçütü değişken sayısı az olan altkümelere daha fazla ağırlık vermiştir. Ancak Beta 2 yapısı için en iyi alt küme olan 5. kümenin bulunmasında farklılıklar olmuştur. Beta 2 için I. Grupta yer alanlar arasında sadece BIC ölçütü doğru altküme bulmuştur. Grupta yer alan diğer ölçütler ise 1. altküme bulmuştur. II. Grupta yer alan ölçütlerden RIC doğru altküme olan 5. altküme bulmuşlardır. Ancak seçen altkümelere bulunmasında, I. Grup ile uyumludurlar. Beta 3 yapısında ise I. Grupta PRESS hariç diğer ölçütler ile II. Gruptaki ölçütler seçen altkümelere ağırlık vermişlerdir. Beta 4 yapısında ise I. Grupta yer alan ölçütlerin hiçbirini eniyi altküme olarak 5. altküme bulamamışlardır. II. Grupta ise RIC eniyi altküme bulmuştur.

#### 4.2. Gibbs Örnekleme Yaklaşımı

Aykırı değer varlığında doğrusal regresyonda Bayesci değişken seçimi için Kuo ve Mallick (7) tarafından önerilen yaklaşım kullanılarak sonuçlar, BARVAS yazılımından elde edilmiş ve Çizelge 4.2'de özetlenmiştir. Gibbs örnekleme sonucunda değişkenlerin sahip oldukları entropi değerleri ise Çizelge 4.3'de verilmiştir.

Gibbs yaklaşımında tekrar sayısı 10.000 olarak alınmıştır. Değişken seçimi işlemi için önsel parametre kestirimleri  $\beta_0=0$  alınmıştır.  $D_0$  önsel varyans tüm değişkenler için eşit alınmakla birlikte önsel beklentilerdeki değişikliklerin sonuçlar üzerindeki etkisini gözlemek amacıyla  $D_0=1$  ve  $D_0=16$  alınmıştır.

**Table 4.2.** The results of variable selection with Gibbs sampling approach in the presence / absence of outlier value  
**Çizelge 4.2.** Aykırı değer olduğu ve olmadığı durumlarda Gibbs örnekleme yaklaşımı ile değişken seçimi sonuçları

Beta	Aykırı değer olmadığı durum/ Outlier value absence				Aykırı değer olduğu durum/ Outlier value presence			
	$D_0=1$		$D_0=16$		$D_0=1$		$D_0=16$	
	Altküme No/ Subset No	Sonsal Olasılık/ Posterior probability	Altküme No/ Subset No	Sonsal Olasılık/ Posterior probability	Altküme No/ Subset No	Sonsal Olasılık/ Posterior probability	Altküme No/ Subset No	Sonsal Olasılık/ Posterior probability
1	5	0.37	2	0.63	21	0.41	1	0.30
	2	0.31	5	0.27	5	0.29	2	0.29
	21	0.24	1	0.05	2	0.13	5	0.21
2	5	0.72	5	0.66	1	0.38	1	0.73
	21	0.17	3	0.22	21	0.26	5	0.12
	3	0.05	6	0.07	5	0.18	30	0.09
3	5	0.70	5	0.75	21	0.43	29	0.29
	21	0.18	2	0.19	26	0.31	26	0.28
	2	0.04	18	0.02	5	0.11	5	0.18
4	5	0.64	5	0.59	6	0.41	6	0.70
	21	0.13	2	0.15	21	0.26	5	0.13
	26	0.11	3	0.12	26	0.21	7	0.08

Elde edilen sonuçlara bakıldığında, öncelikle Beta yapılarının en iyi altküme ulaşmada etkili olmadıkları görülmüştür. Ancak önsel bilgi düzeyinin göstergesi olan  $D_0$  değerinin sonuçlar üzerinde etkili olduğu söylenebilir. Seçilen altkümelere sonsal olasılıklarının aykırı değer varlığında düştüğü görülmektedir. Beta 1 yapısında aykırı değer olmadığı durumda 5 nolu altkümenin sonsal olasılığının düşüklüğü, belirlenen önsel olasılık düzeyinin düşüklüğünden ve seçen altkümelere yüksek sonsal

been in harmony with Group 1. Wherever, the risk criteria have given more importance to subsets that have few variables. But there have been differences in finding the 5th subset that has been the best subset for Beta 2 structure. The other criteria in the Group 1 have found the subset 1. RIC that has been in the Group 2 has found the subset 5 as the correct subset. But also in finding the alternative subsets RIC is in harmony with the Group 1. In the Beta 3 structure, the criteria except PRESS in the Group 1 and the criteria in the Group 2 have given importance to alternative subsets. In the Beta 4 structure none of the criteria in the Group 1 have been able to find the subset 5 which is the best subset. In the Group 2 RIC has found the best subset.

#### 4.2. The Gibbs Sampling Approach

For the Bayesian variable selection in linear regression with outlier value using the approach proposed by Kuo and Mallick (7), the results are obtained from Barvas software and summarized in Table 4.2. The entropy values which variables have in the result of Gibbs sampling are shown in Table 4.3.

In Gibbs approach repeat number is taken as 10.000. For variable selection prior parameter estimation is  $\beta_0=0$ . For all variables  $D_0$  prior variance is taken as equal. Furthermore to evaluate the effects of the differences of prior expectations on the results;  $D_0=1$  and  $D_0=16$ .

Examining the results it is observed that Beta structures are not effective to reach to the best subset. But  $D_0$  value which is the indicator of prior information level is effective on the results. In the presence of outlier value posterior probability of selected subsets decreases. In the absence of outlier value in beta 1 structure the posterior probability of subset 5 is low. The reason for this is low level of prior probability and high posterior probability of

olasılığın kaynaklanmaktadır.

Bağımsız değişkenlerin Beta yapılarına göre entropi değerleri Çizelge 4.3 yardımıyla incelendiğinde, aykırı değerlerin yanında önsel bilgi düzeyinin de etkili olduğu görülmektedir. Ancak Beta yapılarına uygun entropi değerleri elde edilmiştir. Aykırı değer varlığında bağımsız değişkenlerin entropi değerleri önemlilik derecelerini daha belirginleştirecek biçimde artmıştır.

alternative subsets.

The entropy values of independent variables for beta structures are shown in Table 4.3. On entropy values besides outlier value prior information level is also effective. Moreover entropy values suitable for beta structures are obtained. In the presence of outlier value the entropy values increase by making the importance degree of independent variables more clear.

**Table 4.3.** The entropy values of independent variables with Gibbs sampling approach in the presence / absence of outlier value

**Çizelge 4.3.** Aykırı değer olduğu ve olmadığı durumlarda Gibbs örnekleme yaklaşımı uygulandığında bağımsız değişkenlerin entropi değerleri

		Bağımsız değişkenler/Independent variables									
		X1		X2		X3		X4		X5	
Beta	Durum/ Case	D <sub>0</sub> =1	D <sub>0</sub> =16	D <sub>0</sub> =1	D <sub>0</sub> =16	D <sub>0</sub> =1	D <sub>0</sub> =16	D <sub>0</sub> =16	D <sub>0</sub> =16	D <sub>0</sub> =16	D <sub>0</sub> =16
1	A	3.97	4.00	3.96	3.86	3.97	3.66	3.92	3.46	3.86	3.51
	B	4.00	4.00	4.00	3.76	4.00	3.86	3.65	3.11	3.73	3.12
2	A	3.98	3.98	3.89	3.48	3.89	3.48	3.91	3.45	3.87	3.48
	B	3.93	3.85	3.88	3.72	3.94	3.72	3.70	3.15	3.74	3.28
3	A	3.98	3.95	3.89	3.96	3.89	3.72	3.91	3.56	3.87	3.88
	B	3.94	3.91	3.77	3.53	3.94	3.71	3.91	3.87	3.76	3.26
4	A	3.96	3.99	3.84	3.50	3.96	3.99	3.85	3.40	3.87	3.40
	B	3.65	3.19	3.69	3.22	3.67	3.94	3.81	3.45	3.78	3.34

A: outlier value present, B: outlier value absent /A: Aykırı değer var, B: Aykırı değer yok

## 5. SONUÇ

Bu çalışma sonucunda doğrusal regresyonda değişken seçimine Gibbs örnekleme yaklaşımının D<sub>0</sub> önsel bilgi düzeyindeki değişimlerden etkilendiği görülmüştür. Bunun yanında aykırı değer varlığında Gibbs örnekleme yaklaşımı, seçenek altkümeleri Beta yapısına bağlı bulmasına karşın altküme sonsal olasılıklarında düşme ve birbirlerine yakınlık gözlemlenmiştir.

Beta 1 ve Beta 4 yapısında aykırı değer varlığında, Gibbs yaklaşımı C<sub>p</sub> ile benzer sonuçlar vermiştir. Ancak seçenek altkümelerin sonsal olasılıklarında aykırı değer varlığında yükseliş görülmüştür. Gibbs yaklaşımında aykırı değer varlığında altküme sonsal olasılıklarının düşük ve birbirine yakın oldukları gözlemlenmiştir.

Böylece Gibbs yaklaşımlarının aykırı değer varlığından daha fazla etkilendiği, altküme sonsal olasılıkların birbirine yakın olması sonucunda seçenek altkümelerin tercih edilebileceği söylenebilir.

Gibbs örnekleme yaklaşımı, tüm olası alt kümeler yerine sadece sonsal olasılıkları yüksek alt kümeler ile değişken seçimi işlemi yapmaktadır. Bunun sonucunda değişken seçimi işlemi daha hızlı uygulamaktadır.

Aykırı değer varlığında Gibbs yaklaşımı sonucunda önemli bağımsız değişkenlerin entropi değerlerinde yükselme olmuştur. Önemli değişkenlerdeki entropi değerlerinde ise değişim daha az olmuştur. Bunun sonucunda önemli bağımsız değişkenler, aykırı değer varlığında daha fazla bilgi miktarına sahip olurlarken, önemli değişkenlerdeki bilgi miktarı azalmıştır.

## 5. RESULTS

As a result of this study it is shown that Gibbs sampling approach in variable selection of linear regression is affected from the changes in D<sub>0</sub> prior information level. Although Gibbs sampling approach in the presence of outlier value finds alternative subsets dependent to beta structure, subset posterior probabilities decrease and get closer.

In the presence of outlier value in Beta 1 and Beta 4 structures Gibbs approach give similar results with C<sub>p</sub>. But in the presence of outlier value posterior probabilities of alternative subsets increase. In Gibbs approach in the presence of outlier value subset posterior probabilities are lower and closer.

Thus Gibbs approaches are more affected from the presence of outlier value, alternative subsets can be selected because of subset posterior probabilities being closer.

Gibbs sampling approach makes variable selection by using only subsets with high posterior probability instead of all possible subsets. By this way applies variable selection faster.

In Gibbs sampling approach result with outlier value, there is increase in entropy values of important independent variables. The change in entropy values of unimportant variables is lesser. As a result while important independent variables in the presence of outlier value have increase in information amount, unimportant variables have decrease.



**KAYNAKLAR/ REFERENCES**

1. Carlin, B.P. and Chib, S., "Bayesian model choice via markov chain monte carlo methods", *J.R. Statist. Soc.*, B, 57, No. 3: 473-484 (1995).
2. Cavanaugh, J.E., "Unifying the derivations for the Akaike and corrected Akaike information criteria", *Statistics and Probability Letters*, Vol 33, 201-208 (1997).
3. Draper, N. and Smith, H., *Applied Regression Analysis*, **John Wiley**, (1981).
4. Foster, D.P., and George, E.I., "The risk inflation criterion for multiple regression", *The Annals of Statistics*, Vol. 22, No. 4: 1947-1975 (1995).
5. Gamerman, D., *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, **Chapman and Hall**, London, (1997).
6. George, E.I. and McCulloch, R.E., "Approaches for Bayesian variable selection", *Statistica Sinica*, 7, 339-373 (1997).
7. Gray, R.M., *Entropy and Information Theory*, **Springer Verlag**, New York, (1990).
8. Kuo, L. and Mallick, B., "Variable selection for regression models", *Sankhya*, B, 60, 65-81 (1998).
9. Miller, A.J., "Selection of subsets of regression variables", *J.R. Statist. Soc.*, A 147, part 3: 389-425 (1984).
10. Toutenburg, H., *Prior Information in Linear Models*, **John Wiley Inc.**, (1982).
11. Shannon, C.E., "A mathematical theory of communication", *Bell System Tech. J.*, 27, 379-423, 623-656, (1948).
12. Wasserman, L., "Bayesian model selection", *Symposium on Methods for Model Selection*, Indiana Uni., Bloomington (1997).
13. Yardımcı, A., "Doğrusal regresyonda değişken seçimine Bayes yaklaşımlarının karşılaştırılması", Doktora Tezi, *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, (2000).
14. Yardımcı, A. ve Erar, A., "Doğrusal regresyonda Bayesci değişken seçimi ve BARVAS yazılımı", *2. İstatistik Kongresi*, Antalya, (2001).
15. Yardımcı, A and Erar, A., "Bayesian variable selection in linear regression an a comparison", *Hacettepe Journal of Mathematics and Statistics*, 31, 63-76 (2002).

**EK-1:** k=5 için olası altkümeler listesi

**Appendix 1:** List of subsets for k=5

Altküme No Subset No	Değişkenler/ Variables	Altküme No /Subset No	Değişkenler/ Variables
1	$x_1$	17	$x_1, x_4, x_5$
2	$x_1, x_2$	18	$x_1, x_2, x_4, x_5$
3	$x_2$	19	$x_2, x_4, x_5$
4	$x_2, x_3$	20	$x_2, x_3, x_4, x_5$
5	$x_1, x_2, x_3$	21	$x_1, x_2, x_3, x_4, x_5$
6	$x_1, x_3$	22	$x_1, x_3, x_4, x_5$
7	$x_3$	23	$x_3, x_4, x_5$
8	$x_3, x_4$	24	$x_3, x_5$
9	$x_1, x_3, x_4$	25	$x_1, x_3, x_5$
10	$x_1, x_2, x_3, x_4$	26	$x_1, x_2, x_3, x_5$
11	$x_2, x_3, x_4$	27	$x_2, x_3, x_5$
12	$x_2, x_4$	28	$x_2, x_5$
13	$x_1, x_2, x_4$	29	$x_1, x_2, x_5$
14	$x_1, x_4$	30	$x_1, x_5$
15	$x_4$	31	$x_5$
16	$x_4, x_5$		