Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 21, Sayı 3, 999-1010, 2017

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 21, Issue 3, 999-1010, 2017

DOI: 10.19113/sdufbed.44445

Bayesci Genelleştirilmiş Lineer Karma Modellerde Önsel Seçimleri ve Karşılaştırılması

Zeynep ÖZTÜRK*[1], Mehmet Ali CENGİZ[2]

[1]Artvin Çoruh Üniversitesi, Hopa İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, 08600, Artvin
[2]Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55200, Samsun

(Alınış / Received: 18.10.2016, Kabul / Accepted: 26.04.2017, Online Yayınlanma / Published Online: 14.06.2017)

Anahtar Kelimeler
Genelleştirilmiş lineer karma modeller,
Bayesci yaklaşım,
Önsel dağılımlar,
Önsel seçimi

Özet: Parametre tahmininde ve model seçiminde uzman görüşlerin modele katılmasını öngören Bayesci yaklaşımda en önemli nokta parametreler hakkında önsel bilgi seçimi ve kullanımıdır. Bu nedenle, bu çalışmada, önsel dağılımların seçimleri tanıtılmış, genelleştirilmiş lineer karma modeller için farklı önseller ile elde edilen modeller hem parametre tahmini hem de model uyumu anlamında karşılaştırılmıştır. Akciğer kanseri hastalarının yaşam kaliteleri ve aldıkları tedavinin çıktılarını ölçmek amacıyla yapılan bir çalışmaya ait gerçek bir veri kullanılmıştır. SAS 9.3 programında Markov Zinciri Monte Carlo algoritması ile elde edilen parametre tahminleri bulunmuş ve modeller kurulmuştur. Farklı önsel seçimleri ile Sapma Bilgi kriteri (DIC) ne göre en iyi model bilgilendirici önsel dağılım ile elde edilen model olmuştur.

Prior Selection and Comparisons in Bayesian Generalized Linear Mixed Models

Keywords
Generalized linear mixed models,
Bayesian approach,
Prior distribution,
Prior selection

Abstract: In parameter estimation and model selection, selection and use of prior information about the parameters is the most important point in the Bayesian approach that envisage to participate in the expert opinion. Therefore, in this study, the choice of the prior distributions was introduced, generalized linear mixed models with different prior were compared in terms of both models and parameter estimation. It was used a real data that measure outcomes of the treatment and qualities of life of lung cancer patients In SAS 9.3 software, It was found parameter estimates with the Markov Chain Monte Carlo algorithm and models were established. Model that obtained with informative prior distribution is the best model according to Deviance information criterion(DIC).

1. Giriş

Genelleştirilmiş lineer karma modeller (GLKM), tıp biliminde ve birçok disiplinde rasgele etkili normal dağılımlı olmayan verilerin analizini yapabilmek için son yıllarda oldukça kullanışlı ve popülerliğini gün geçtikçe arttıran bir modeldir. [1] tarafından ileri sürülen genelleştirilmiş lineer modeller (GLM), bağımlı değişkenin olasılık dağılımının bir üstel ailesinin herhangi bir üyesi olmasını sağlayan ve lineer olmayan bir link fonksiyonu sayesinde bir lineer tahmin ediciye bağlı bir örneklemin ortalamasını elde eden klasik lineer modelin genişlemesidir. Bir çok kullanışlı istatistiksel modeller uygun link fonksiyonu ve bağımlı değişkenin olasılık dağılımı seçimiyle genelleştirilmiş lineer modeller olarak formüle edilebilir. Genelleştirilmiş lineer karma modeller ise, lineer tahmin ediciye rasgele etkilerin eklenmesiyle genelleştirilmiş lineer modellerin genişlemesi olarak adlandırılır. GLKM de rasgele etkilerin varlığı olabilirlik fonksiyonunun rasgele etkilerin dağılımı üzerine yüksek boyutlu bir integral içermesine neden olduğundan, olabilirliğin hesaplanmasını ve dolayısıyla da maksimum olabilirlik tahminlerini oldukça zorlaştırır. Bilgisayar kullanımının artması ve karmaşık çözümlerin elde edilmelerine getirilen kolaylıklar nedeniyle klasik yaklaşımın yanısıra Bayesci yaklaşımların uygulamada kullanımı da artmıştır. Bu nedenle, GLKM' nin analitik zorluğunu gidermek için, olabilirliğe dayalı yaklaşımlara (klasik yaklaşımlara) alternatif Bayesci yaklaşım önerilebilir. Bayesci yaklaşımlar doğrudan çıkarım elde etmenin ve Monte Carlo yaklaşımlarından Markov Zinciri Monte Carlo yaklaşımını kullanarak sonsal ve tahmin edici dağılımlardan rasgele örneklem çekmeye dayalı

*İlgili yazar: zozturk@artvin.edu.tr

bir süreç yürütmenin kolaylığını sağlar. [2] GLKM ye Gibbs örneklemini uyguladı ve Monte Carlo metodlarını içeren diğer uygulamalar [3], [4], [5] tarafından yapıldı. [1] MCMC sayesinde GLKM analizi ile ilgili bir araştırma örneği ortaya çıkardılar. [2] ise, rasgele etkilerin kovaryans matrisi için yaklaşık Jeffreys önselli bir Bayesci yaklaşım geliştirdi. [3] GLKM analizi için Bayesci ve klasik yöntemleri ile ilgili bilgileri bir arada toparlayarak literatüre katkıda bulundu.

Klasik yaklaşım ile Bayesci yaklaşımın en belirgin farkı, Bayesci yaklaşımda bütün bilinmeyen parametrelere rasgele değişken gibi davranılmasıdır. [4] regresyon parametreleri için büyük örneklem durumunda normal önselin uygun olduğunu gösterdiler. [5] normal dağılımlı rasgele etkili modellere karşılık gelen G kovaryans matrisi için bir eşlenik önsel seçimi olarak Wishart dağılımını belirlemişlerdir. Wishart için hiper parametrelerini seçmenin oldukça zor olduğunu ve Gibbs örnekleminin yakınsama oranını belirlemenin sıkıntılı olduğunu belirtmişlerdir. [6] ve [7] konumsal GLKM de önsel seçimini tanımlamış ve [8] kovaryans matrisinin önsel seçimleri hakkında çalışmışlar yapmışlardır. Rasgele etkili varyans için [9] uniform önselin ters gama önseline göre daha iyi olduğunu belirlemişlerdir. [10] eşlenik ters Wishart dağılımını önermişlerdir. Klasik modeller için [11] regresyon parametrelerini tahmin etmek için yeterli veriye sahip olunduğunda, herhangi bir bilgilendirici olmayan önselin yeterli olduğunu belirttiler. [11] bilgilendirici olmayan önsel dağılımların ters gama ailesi ile ilgili ciddi problemlerin olduğunu göstermişler ve yerine standart sapma parametreleri üzerine uniform önsel ve zayıfça (weakly) bilgilendirici önsel(yarı-t önseli) önermişlerdir. [12] model seçimi için regresyon parametreleri için bilgilendirici önsel önerdiler. Sabit etkilere ait parametre vektörüne ait ön bilgi yoksa regresyon katsayıları için sezgisel bir seçim uniformdur ve G için Jeffreys önseldir [13], [14]. Bilgilendirici olmayan ve subjektif önsellere alternatif olarak şartlı eşlenik önsel [15], referans önsel [8], difüz önsel [9] ve zayıfça bilgilendirici önsel [11] tarafından seçilmiş ve gösterilmiştir.

Bayesci yaklaşımda parametreler için önsel dağılmı belirlemek büyük bir sorundur. İyi bir önsel belirlemenin dışında da Bayesci GLKM de hala birçok problem vardır. Bunlar hesaplama ve yakınsama sorunları ve hesaplama yükünü engellemek için klasik yaklaşımdaki büyük örneklem teorisini uygulama sorunları olarak belirtilebilir. Bayesci genelleştirilmiş lineer karma modellerde sabit etkilere rasgele etkiler eklendiğinden ve parametre tahmini için rasgele etkiler üzerinden işlem yapıldığından rasgele etkilerin boyutu arttığında hesaplama oldukça zor ve karmaşık hale gelir. Böyle durumlarda bu hesaplama zorluğundan kaçınmak için yaklaşık çıkarım ele alınmalıdır. Bu nedenle, hem yüksek boyutlu integral alma hem de Bayesci yaklaşımdaki hesaplama ve yakınsama sorunlarıyla karşılaşılacağından önsel dağılımın önemle seçilmesi ve dikkatli olunması gereken bir noktadır. Bu çalışmada, önceki çalışmalardan ele edilen önsel dağılım seçimleri, gerçek bir tıp verisine uygulanarak parametre ve model uyumu bakımdan karşılaştırmalı olarak değerlendirilmiştir. Günümüz teknolojisinde yüksek boyutlu integraller birçok sayısal integrasyon yöntemiyle elde edilebilir. Bu nedenle, bu çalışmada Bayesci yaklaşım ile sonsal dağılımları ve parametre tahminlerini hesaplamak için Markov Zinciri Monte Carlo(MZMC) teknikleri kullanılarak SAS 9.3 bilgisayar yazılım programı kullanılmıştır.

2. Modeller

2.1. Genelleştirilmiş lineer karma modeller

GLKM' nin üç temel bileşeni vardır:

i. Üstel dağılım ailesinden, rasgele etki üzerinde şartlı bağımlı değişken dağılımının seçimi,
ii. Sabit ve rasgele etkileri içeren lineer tahmin edicinin formüle edilmesi,
iii. Bir link fonksiyonu seçimidir.

Y bağımlı vektör $f(y_{ij} \mid \beta, \gamma_j, G)$ yoğunluklu üstel dağılım ailesinden γ_j rasgele etkiler üzerinde şartlı yoğunluk fonksiyonu bağımsız rasgele olarak dağılır. Bir GLKM aşağıdaki gibi gösterilebilir:

$$\begin{aligned} E[y_{ij} \mid \gamma_j] &= g^{-1}(X_{ij}\beta + Z_{ij}\gamma_j) = g^{-1}(\eta_{ij}) = \mu_{ij} \\ \gamma_j &\sim N(0,G), \quad j = 1,\dots,N;\ i = 1,\dots,n_j \end{aligned} \tag{1}$$

Burada, g fonksiyonu uygun bilinen bir link fonksiyonu, η_{ij} lineer tahmin edici, N grupların toplam sayısı, n_j $j-$inci gruptaki deneklerin toplam sayısıdır. X sabit etkilere ilişkin tasarım matrisi, β sabit etkilere ait parametre vektörü, Z rasgele etkilere ilişkin tasarım matrisi, γ 0 ortalamalı ve G varyans matrisli normal dağılımlı olduğu varsayılan rasgele-etkilere ait parametre vektörüdür. G, σ_γ^2 varyans bileşeni köşegen elemanı olan bir köşegen bir matristir. σ_ε^2 hatadaki değişimi ifade eden parametre olmak üzere, modelde yer alan γ ve ε rasgele vektörleri üzerine kurulan temel varsayımlar ise, aşağıdaki gibidir :

- $E(\gamma) = 0$ ve $Cov(\gamma) = E(\gamma\gamma') = G$ olmak üzere, $\gamma \sim N(0,G)$
- $E(\varepsilon) = 0$ ve $Cov(\varepsilon) = E(\varepsilon\varepsilon') = R = \sigma_\varepsilon^2 I_{n_j}$ olmak üzere, $\varepsilon \sim N(0,R)$
- $Cov(\gamma, \varepsilon') = Cov(\varepsilon, \gamma') = 0$
- $Cov\begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G_{NxN} & 0 \\ 0 & R_{n_j x n_j} \end{bmatrix}$

Genel olarak GLKM' ler boylamsal verilerin (longitudinal data) ve ilişkili tekrarlı ölçümlerin analizinde kullanıldığı söylenebilir. Bu çalışmada iki çıktılı boylamsal veriye sahip olunduğundan, ikili sonuçlardan oluşan veri yapısını analiz etmek için kullanılan rasgele etkili lojistik model Denklem 2'dir.

$$Y_{ij} \mid \gamma_j \sim Bağımsız.Bernoulli[\pi(x_{ij})]$$
$$\gamma_j \sim N(0,\sigma^2_{\gamma_j}) \qquad (2)$$
$$\log it(\pi(x_{ij})) = X_{ij}\beta + Z_{ij}\gamma_j$$

$f(.)$ olasılık yoğunluk fonksiyonu için genel bir terim olarak kullanılmak üzere, y ve γ nın ortak dağılımı

$$f(y,\gamma \mid \beta,G) = \prod_{j=1}^{N}\prod_{i=1}^{n_j} f(y_{ij} \mid \gamma_j,\beta) f(\gamma_j \mid G) \qquad (3)$$

dir. Bir karma modelde verinin marjinali

$$f(y \mid \beta,G) = \prod_{j=1}^{N}\int\prod_{i=1}^{n_j} f(y_{ij} \mid \gamma_j,\beta) f(\gamma_j \mid G) d\gamma_j \qquad (4)$$

olarak ifade edilir. Bu integrasyonu hesaplamak ve maksimize etmek oldukça zor olduğundan sayısal yaklaşımlara ihtiyaç duyulur Özel olarak rasgele etkili lojistik modellerde olabilirlik fonksiyonu aşağıdaki şekilde hesaplanabilir [1]:

$$L = \prod_{j=1}^{N}\int_{-\infty}^{\infty}\prod_{i=1}^{n_j} P(Y_{ij}=1 \mid \gamma_j,\beta)\, f(\gamma_j)\, d\gamma_j \qquad (5)$$

2.2. Rasgele etkili lojistik modeller

Rasgele etkili lojistik modeller ikili sonuçlardan oluşan veri yapısını analiz etmek için sıkça kullanılan popüler bir yöntemdir. Y_{ij}; $j-\text{inci}$ gruptaki $i-\text{inci}$ gözlemin ikili bağımlı değişkeni, N; grupların toplam sayısı, n_j; $j-\text{inci}$ gruptaki deneklerin toplam sayısı, $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$; sabit etkilere ilişkin parametre vektörü ve $g_1, g_2, \ldots, g_N$ rasgele seçilmiş düzey etkileri, yani, $i-\text{inci}$ gözleme karşılık gelen etkiyi modellemede bir parametre olsun. Rasgele-etkili lojistik regresyon model(bağımlı değişkenin şartlı beklenen değeri) aşağıdaki formdadır:

$$\pi(x_{ij}) = E(Y_{ij} \mid \gamma_j) = P(Y_{ij}=1 \mid \gamma_j) = \frac{e^{X_{ij}\beta + Z_{ij}\gamma_j}}{1+e^{X_{ij}\beta + Z_{ij}\gamma_j}} \qquad (6)$$

veya, denk olarak

$$\log it(\pi(x_{ij})) = X_{ij}\beta + Z_{ij}\gamma_j \qquad (7)$$

$$Y_{ij} \sim Bağımsız.Bernoulli[\pi(x_{ij})]$$
$$\gamma_j \sim N(0,\sigma^2_{\gamma_j})$$

dir. Rasgele etkili lojistik modellerin varsayımları

i) γ_j verildiğinde Y_{ij} nin şartlı dağılımı $f(y_{ij} \mid \gamma_j,\beta)$ yoğunluk fonksiyonu bir üstel dağılım ailesinin üyesi olan Bernoulli dağılımına sahip olmalıdır.

ii) γ_j verildiğinde tekrarlı ölçümler $Y_{i1}, Y_{i2}, \ldots, Y_{inj}$ bağımsızdır.

iii) γ_j bağımsız ve özdeş olarak dağılımlı $f(\gamma_j)$ yoğunluk fonksiyonlu normal dağılımlıdır.

Hata ve rasgele etkilere ait varsayımlar ise;

i) $j=1,2,\ldots,N,\quad i=1,2,\ldots,n_j$ için gözlenemeyen rasgele hata terimi $e_{ij} : N(0,s_e^2)$ ve e_{ij} 'ler bağımsız,

ii) $g_1, g_2, \ldots, g_N$ rasgele faktör etkileri bağımsız ve her biri $N(0,s_g^2)$ dağılımlı,

iii) $j=1,2,\ldots,N,\quad i=1,2,\ldots,n_j$ için e_{ij} 'ler ile γ_j 'ler bağımsız varsayılmaktadır.

$Var(Y_{ij}) = s_g^2 + s_e^2$ olmak üzere, s_g^2 ile s_e^2 varyanslarına, varyans bileşenleri (variance components) denir.

3. Bayesci Yaklaşım

Klasik yaklaşımdan farklı olan Bayesci yaklaşımda parametrelere olasılık dağılımı olan bir rasgele değişken gibi davranılır. Ayrıca olasılık kavramını; bir olayın olasılığı o olayın doğru olduğuna inanmanın derecesi olan inanç derecesi olarak tanımlar. Bayesci yaklaşımda önce, parametre hakkında ortalama, yayılım, sivrilik gibi kesin olmayan bilgileri ifade eden bir önsel dağılım belirlenmelidir.

$p(y|\theta)$ yoğunluk fonksiyonu ile tanımlı istatistiksel bir model kullanarak $y = \{y_1, \ldots, y_n\}$ verisinden θ tahmin edilmek istensin. Bayes çıkarımına göre, θ üç adımda tahmin edilebilir:

1. θ için bir olasılık dağılımı, önsel dağılım olarak da bilinen $\pi(\theta)$ formüle edilmeli,

2. y gözlem vektörü olmak üzere, θ verildiğinde y'nin dağılımını belirlemek için $p(y|\theta)$ uygun bir istatistiksel bir model seçilmeli,

3. Sonsal dağılımın, $p(\theta|y)$, hesaplanması için veri ve önsel dağılımdan elde edilen bilgiyi birleştirerek θ hakkındaki inanışlar güncellenmelidir [17].

Bu üç adım, önsel dağılım kullanarak Bayes teoremi kullanımını gerektirir:

$$p(\theta|y) = \frac{p(\theta,y)}{p(y)} = \frac{p(y|\theta)\pi(\theta)}{p(y)} = \frac{p(y|\theta)\pi(\theta)}{\int p(y|\theta)\pi(\theta)\,d\theta} \quad (8)$$

Burada $\int p(y|\theta)\pi(\theta)\,d\theta$, sonsal dağılımın y' nin marjinal dağılımıdır. θ nın olabilirlik fonksiyonu $L(\theta)$, $p(y|\theta)$ fonksiyonu ile orantılı bir fonksiyonudur, yani, $L(\theta) \propto p(y|\theta)$. Bayes teoremini yazmanın diğer bir yolu da aşağıdaki gibidir:

$$p(\theta|y) = \frac{L(\theta)\pi(\theta)}{\int L(\theta)\pi(\theta)d\theta} \quad (9)$$

$p(y)$ marjinal dağılımı bir integraldir. İntegral sonlu ise, bu integralin değeri, sonsal dağılım için herhangi bir katkı sağlamaz. Bu nedenle, sonsal dağılım $p(\theta|y)$ aşağıda verilen orantılı formda rastgele sabit olarak yazılabilir:

$$p(\theta|y) \propto L(\theta)\pi(\theta) \quad (10)$$

4. Önsel Dağılımlar

Bir parametrenin önsel dağılımı, veriyi analiz etmeden önce parametre hakkında kesin olmayan ön bilgileri içeren olasılık dağılımıdır. Önsel dağılım ve olabilirlik fonksiyonunun çarpımı parametrenin sonsal dağılımını verir. Sonsal dağılımı kullanarak tüm çıkarımlar yapılabilir. Önsel dağılım kullanmadan Bayesci çıkarım elde edilemez. Önsel dağılım bilgilendirici (informative)-subjektif veya bilgilendirici olmayan (non-informative)-objektif şeklinde olabilir. Bilgilendirici olmayan önselin kullanılması durumunda, sonsal yoğunluk, olabilirlik fonksiyonuna benzer.

4.1. Önsel dağılımın belirlenmesi sorunu

Önsel dağılımı belirlemede yani önsel dağılımın seçiminde veya oluşturulmasında çok çeşitli sınıflandırmalar yapılabilir. Bayesci yaklaşım içinde önsel dağılım seçiminden kaynaklanan fikir ayrılıkları vardır. Bu bağlamda öncelikle önsel dağılım tercihine göre Bayesciler gruplandırılabilir:

- Klasik Bayesciler, düzgün önsel dağılım gibi bilgilendirici olmayan önsel dağılım (non-informative prior) kullanmayı tercih ederler.
- Modern Parametrik Bayesciler, tasarlanmış özelliklere sahip eşlenik önsel (conjugate prior) tercih ederler.
- Sübjektif Bayesciler, benzer bir alanda daha önce elde edilen izlenimler doğrultusunda, çoğunlukla uzman görüşünden elde edilen bilgiye göre ortaya çıkarılan önsel dağılım seçerler.

4.2. Uygun olmayan(improper) önseller

Eğer $\int \pi(\theta)d\theta = \infty$ ise, $\pi(\theta)$ önselinin uygun olmadığı söylenir. Örneğin, θ $[-\infty,\infty]$ aralığı için $\pi(\theta) \propto 1$ önseli reel uzayda tanımlı düzgün önsel dağılıma sahip ise, uygun olmayan (improper) önseldir. Bir sonsal dağılımın uygun olup-olmadığına karar vermek için, tüm y'ler için $\int p(y|\theta)\pi(\theta)\,d\theta$ normalleştirme sabitinin sonlu olduğundan emin olunmalıdır. Uygun olmayan bir önsel dağılım uygun olmayan bir sonsal dağılıma öncülük ediyorsa, sonsal dağılım üzerine yapılacak tüm çıkarımlar geçersiz olur.

4.3. Bilgilendirici olmayan önseller

Bilgilendirici olmayan önsel dağılımlar, parametrenin muhtemel gerçek değeri hakkında hemen hemen hiçbir bilgi taşımayan önsel dağılımlar olarak tanımlanır. Örneğin, bir lojistik rastgele etkili modelde ölçülen rastgele etkiler için, sıfır ortalama ve 10000 varyans ile bir normal dağılım muhtemel gerçek parametre değeri hakkında hemen hemen hiçbir bilgi taşımaz ve böylece, bilgilendirici olmayan bir önsel dağılımı oluşturur. Varyans parametresi için, bilgilendirici olmayan bir önsel dağılım örneği için hem şekil ve hem de ölçek parametreleri 10^{-3} olan bir ters gama dağılımı olabilir. Bilgilendirmeyen önsel dağılımlar Düzgün önsel dağılım (Flat, Uniform Prior), Jeffreys önsel dağılım (Jeffreys's Prior), Belirsiz önsel dağılım (Diffuse, Vague, Weak, Locally Uniform Prior) lardır.

4.3.1. Düzgün önsel dağılım

Düzgün önsel dağılımda, belirlenen aralıkta parametreye aynı olasılık değerleri atanır; $p(\theta) = c = 1\ ve\ 0 \leq \theta \leq m$ iken, belirtilen aralıktaki her noktada, parametrenin olasılığı c'ye eşittir [20]. Bu, "yetersiz neden ilkesi"ne ("principle of insufficient reason") dayanır. Belirli bir neden olmadıkça, bir olayın gerçekleşme olasılığı başka bir olaya göre daha olası değildir. Bu nedenle eşit olasılıklar tayin edilir.

Düzgün önsel dağılım genel olarak parametrenin belirli bir aralıkta yer aldığı, sınırlandırılabildiği durumlarda kullanılabilir. Örneğin, parametre $[0,m]$ aralığında ise, m sonsuza giderken önsel dağılım daha az bilgi verir hale gelir. Reel eksende $[-\infty,\infty]$ aralığında, θ'nın tüm değerleri için $p(\theta)=c$ iken, düzgün önsel dağılım "uygun olmayan" dır. Yani, olasılık yoğunluk fonksiyonunun integrali (veya toplamı) alındığında sonsuz çıkar ve olasılıklar toplamının 1'e eşit olma aksiyomunu bozar. Uygun olmayan önsel dağlımlar hesaplamada güçlük çıkarır. Şunu da belirtmelidir ki, böyle bir önsel dağılımdan elde edilen sonsal dağılım, yanlış olmak zorunda da değildir.

4.3.2. Jeffreys' in sabit önseli

[3] tarafından önerilen ve oldukça kolay hesaplanabilen Jeffreys'in önseli, düzgün dağılım özelliğine sahip olup parametrelerin tanımlı olduğu aralık dışında büyük değerleri içermez. Ayrıca olabilirliğin anlamlı olduğu bölgede çok fazla değişmeyen ve Fisher bilgi matrisi kullanılarak hesaplanan bir önseldir [20]. Jeffreys'in önseli çok kullanışlı bir önseldir. Aralık dışında büyük değerleri olmayan ve olabilirliğin anlamlı olduğu bölgede çok fazla değişmeyen bir önseldir. Jeffreys önseli, sözü edilen temel motivasyondan hareketle, parametre belirli bir aralıkla sınırlı olsun ya da olmasın ($[-\infty,\infty]$ veya $[0,\infty]$) önsel dağılımı bir sabite eşitler. Jeffreys'in önseli aşağıdaki gibi tanımlanır:

$$\pi(\theta) \propto |I(\theta)|^{1/2} \tag{11}$$

Burada, $|.|$ determinantı ve $I(\theta)$ ise, $p(y|\theta)$ olabilirlik fonksiyonu için Fisher bilgi matrisini gösterir [3]. Jeffreys'in önseli yüksek boyutlularda kullanımı yavaş olur.

4.3.3. Belirsiz (vague) önsel dağılım

Bilindiği gibi parametrenin dağılımının türüne göre, bu parametrenin dağılımının ölçüsünün şeklini veren parametre değişmektedir. Sözgelimi konum parametresi için bir normal dağılım alınabilir. Şekil parametresi için de gama dağılımı alınabilir. Büyük varyansın belirsizlikle özdeş olması fikrine dayanarak, bu dağılımların şekil parametrelerine öyle değerler atanır ki, oldukça geniş bir aralıkta yer alan, neredeyse düzgün önsel dağılım kadar düz, bilgilendirici olmayan bir önsel dağılım oluşturulur; $p(\theta)\sim N(\mu,\sigma^2)$, σ^2'ye büyük değer atanır. $p(\sigma)\sim G(\alpha,\beta)$; α,β'ya (0,001 gibi) çok küçük değerler atanır.

Bu şekilde oluşturulan belirsiz önsel dağılımın, en güçlü yanı uygun dağılım olmasıdır. Düzgün önsel dağılım yerine belirsiz önsel dağılım almakla, sözü edilen özelliğinden ötürü hesaplamada oldukça kolaylık sağlanmış olunur.

4.4. Bilgilendirici önseller

Bilgilendirici bir önsel, sonsal dağılım üzerine etkisi olan olabilirliğe etkisi olmayan bir önseldir. Eğer önselin olabilirliğe etkisi var ise, açıkça bilgilendirici bir önsel olur. Bu tip dağılımlar gerçek uygulamaların çözümünde önemlidir. Diğer bir ifade ile önceki çalışmalardan elde edilmiş bilgi, geçmiş deneyimler veya doğal bir şekilde şimdiki bilgi ile birleştirilebilecek bir uzman görüşü gibi uygun önsel dağılımların kullanımı Bayesci metodun gücünü arttırır.

4.5. Eşlenik önseller

Eğer önsel ve sonsal dağılımlar aynı aileden ise, önselin dağılımlar ailesi için eşlenik önsel olduğu söylenir. Örneğin, olabilirlik binom ise, $y \sim Bin(n,\theta)$ θ için eşlenik önsel beta dağılımıdır ki, θ'nın sonsal dağılımı Tablo 1 de verildiği gibi beta dağılımıdır. Hesaplamalardaki kolaylığı arttırdığı için Eşlenik önseller genelde tercih edilir. Ancak, Bayesci yöntemler sonsal örneklemede eşlenik kullanmaz.

Tablo 1. Olabilirlik fonksiyonuna göre eşlenik önsel dağılım ve sonsal dağılım

Olabilirlik Fonksiyonu	Önsel Dağılım	Sonsal Dağılım
Binom	Beta	Beta
Negatif Binom	Beta	Beta
Normal	Normal	Normal
Poisson	Gama	Gama
Üstel	Gama	Gama
Gama	Gama	Gama

5. Genelleştirilmiş Lineer Karma Modellerde Bayesci Yaklaşım

Maksimum olabilirlik tekniklerinde rasgele etkilerin çok boyutlu olarak integrallenmesi oldukça zor ve karmaşık olduğundan lineer tahmin edici için karmaşık yapılara (örneğin, rasgele etkili, hem çapraz hem de iç içe etkileri birleştiren) imkân tanıyan GLKM' lerin Bayesci yapısı ele alınır. Bayesci yaklaşımın avantajı, rasgele etkiler için kullanılabilen çeşitli dağılımların (normalden başka) olmasıdır. Bayesci yaklaşım GLKM' deki bütün bilinmeyen parametrelere rasgele değişkenler gibi davranır. β parametrelerine ve normal dağılımlı rasgele etkili G varyans-kovaryans matrisine önsel dağılımlar belirlenir. GLKM' deki sonsal dağılımları hesaplamak sayısal ve analitik olarak zor olduğundan, sonsal ve tahmin edici çıkarım MCMC teknikleri ile elde edilen sonsal ve tahmin edici dağılımlardan rasgele örneklem çekmeye dayalıdır [2].

GLKM, genel olarak bir Bayesci yapıda formüle edilebilir ve böylece Bayesci metotlar kullanılarak analiz edilebilir. Bayesci yapı için β ve G parametreleri rasgele değişkenlerdir. β ve G parametreleri için olabilirlik fonksiyonu

$$L(\beta,G\,|\,y)\propto\prod_{j=1}^{N}\int\prod_{i=1}^{n_j}f(y_{ij}\,|\,\gamma_j)|G|^{-1/2}\exp(-\frac{1}{2}\gamma_j^{'}G^{-1}\gamma_j)d\gamma_j \quad (12)$$

dir. Yukarıdaki integralin boyutu rasgele değişkenlerin sayısına eşit olduğundan bir analitik çözüme sahip değildir, bu nedenle sayısal çözümlemelere ihtiyaç duyulur. Rasgele etkiler normal dağılım varsayımına sahip olduğundan,

$$L(\beta,G\,|\,y)\propto\prod_{j=1}^{N}\int\prod_{i=1}^{n_j}f(y_{ij}\,|\,\gamma_j)|G|^{-1/2}\exp(-\frac{1}{2}\gamma_j^{'}G^{-1}\gamma_j)d\gamma_j \quad (13)$$

$$f(\gamma_j\,|\,G)=\frac{1}{(2\pi)^{q/2}|G|^{1/2}}\exp(-\frac{1}{2}\gamma_j^{'}G^{-1}\gamma_j) \quad (14)$$

olur. $\pi(\beta,G)$, β ve G için ortak önsel dağılımı göstersin. Bayesci çıkarımın genel amacı,

$$f(\beta,G\,|\,y)=\frac{\prod_{j=1}^{N}\int\prod_{i=1}^{n_j}f(y_{ij}\,|\,\gamma_j)|G|^{-1/2}\exp(-\frac{1}{2}\gamma_j^{'}G^{-1}\gamma_j)d\gamma_j\pi(\beta,G)}{\prod_{j=1}^{N}\int\prod_{i=1}^{n_j}f(y_{ij}\,|\,\gamma_j)|G|^{-1/2}\exp(-\frac{1}{2}\gamma_j^{'}G^{-1}\gamma_j)d\gamma_j\pi(\beta,G)d\beta dG} \quad (15)$$

sonsal dağılımı elde edebilmektir. Payda sonsal dağılımın normalleştirme sabitidir. Bu yüzden sonsal tahmin ediciler sadece paydan elde edilebilir. Eğer $\pi(\beta,G)$ düzgün bir önsel ise, (15) eşitliği basitçe olabilirlik fonksiyonuna dönüşür. Bazı problemlerde $f(\gamma_j\,|\,y)$ sonsal dağılımları önemli olabilir. Benzer şekilde γ_j için sonsal

$$f(\gamma_j\,|\,y)=\frac{\int f(y_i\,|\,\gamma_j,\beta)f(\gamma_j\,|\,G)\pi(\beta,G)d\beta dG}{\int f(y_i\,|\,\beta,\gamma_j)f(\gamma_j\,|\,G)\pi(\beta,G)d\gamma_j d\beta dG} \quad (16)$$

dir. Eşitlik (15) ve (16)' da verilen sonsalları özellikle γ_j rasgele değişkenin, N boyutu birden daha büyük olduğunda sayısal olarak çözümü zor olacaktır. Bu yüzden, hesaplamaların yapılabilmesi için Monte Carlo metotları önerilir. Örneğin, [15] de sonsalları bulabilmek için Gibbs örneklemi kullanımı uygun bulunmuştur.

6. Rasgele Etkili Lojistik Modellerde Bayesci Yaklaşim

Rasgele etkili lojistik modellerde β ve σ^2 model parametreleri için $\pi(\beta)$ ve $\pi(\sigma^2)$ önseller belirlenmelidir. Rasgele etkili modellerde β ve σ^2 için genellikle bilgilendirici olmayan önseller kullanılır. Ayrıca önsellerin bağımsız olduğu kabul edilir. Sonsal dağılımın hesaplanması için, veri ve önsel dağılımdan elde edilen bilgi birleştirilerek β hakkındaki inanışlar belirlenmelidir. O halde, rasgele etkili lojistik modeller için sonsal dağılım

$$p(\beta,\sigma^2)=$$

$$\frac{\prod_{j=1}^{N}\int_{-\infty}^{\infty}\prod_{i=1}^{n_j}P(Y_{ij}=1\,|\,\gamma_j,\beta)\,f(\gamma_j)\,d\gamma_j\,p(\beta)p(\sigma^2)}{\int\prod_{i=1}^{n_j}\int_{-\infty}^{\infty}\prod_{j=1}^{N}P(Y_{ij}=1\,|\,\gamma_j,\beta)\,f(\gamma_j)\,d\gamma_j\,p(\beta)p(\sigma^2)d\beta d\gamma} \quad (17)$$

$$p(\beta,\sigma^2)\propto\prod_{j=1}^{N}\int_{-\infty}^{\infty}\prod_{i=1}^{n_j}P(Y_{ij}=1\,|\,\gamma_j,\alpha)\,f(\gamma_j)\,d\gamma_j\,p(\beta)p(\sigma^2)$$

olur. Analitik olarak çözümü zor olduğundan, sonsal tahminleri elde etmek için MCMC metodu gibi simülasyon algoritması kullanılabilir.

7. Genelleştirilmiş Lineer Karma Modellerde Önsellerin Belirlenmesi

Tüm istatistiksel modellerin Bayesci analizinde olduğu gibi, GLKM' de de önsel dağılımların iki sınıfı olan bilgilendirici olmayan önsel dağılımlar ve bilgilendirici önsel dağılımlardan bahsedilebilir Bilgilendirici olmayan önsel dağılımlarda az bilgi varken, bilgilendirici önsellerde yeteri kadar bilgi mevcuttur, örneğin elde edilen veriler hakkında var olan önceki bilgiler gibi. Bayesci analizde uygun bir önsel dağılıma ihtiyaç duyulur. Bu nedenle, önsel dağılımın seçimi özellikle Bayesci faktörlerin ve sonsal değerlerin hesaplamasında ve Bayesci hipotez test problemlerinde çok önemli bir yere sahiptir.

Literatürde genelde sabit etkili parametre β için önsel dağılım genelde bir belirsiz normal önsel dağılımdır:

$$\beta\sim N(0,\sigma^2) \quad (18)$$

[21]' de normal modellerde parametreleri tahmin etmek için yeterli veri varsa, herhangi bir bilgilendirici olmayan önselin yeterli olduğundan bahsettiler. İki bağımlı model için [8] β için Uniform dağılımı kullanımını önerdi.

Lojistik modeller için, [4] büyük veya yeteri kadar örneklem durumunda β için normal önselin kullanışlı olduğuna işaret etti. Büyük örneklemin olmadığı durumda büyük kovaryanslı bir normal önsel kullanımında dikkatli olunması gerektiğini bildirdiler. [12]' de GLKM için bilgilendirici önselin bir sınıfını önerdi ve hesaplama özellikleri yanı sıra teorisinden bahsetti.

Son yıllarda Bayesci GLKM' de varyans bileşenleri için önsel belirlemek aktif bir istatistiksel araştırma alanı oldu. Bu parametreler için uygun olmayan önsellerin kullanımı uygun olmayan sonsallara neden olabilmektedir. Varyans bileşenleri için önsel seçimi önemlidir ve zordur. Seçimde dikkatli olunmalıdır. Normal rasgele etkili GLKM de en yaygın olarak kullanılan önsel, tek bir varyans bileşeni için normal dağılımın eşlenik dağılımı olan ters Gamma ve varyans kovaryans matrisi için ters Wishart dağılımıdır [11]. Fakat birçok uygulamalarda varyans bileşenleri hakkında az bilgi vardır ya da hiç bilgi yoktur. Bu belirsizliği gidermek için ya difüz önsel ya da bilgilendirici olmayan önsel kullanılır [22]. Ayrıca, [4] varyans bileşeni için Jeffreys önseli üzerine çalıştı. [8], [11] GLKM' de varyans bileşenleri için difüz önsellere alternatif önseller önerdiler. Bu önsellerin bazıları, yaklaşık Uniform ve tanımlı eşlenik önseller veya rasgele etkilerin varyansları yerine standart sapmalar üzerine Uniform önsellerdir. Ayrıca [8]' de varyans kovaryans matrisi için bilgi matrisine bir yaklaşım kullanarak Jeffreysin genel yapısını önerdi. [5]' de varyans bileşenleri için Cauchy dağılımı, ters Gamma ve Uniform önselli modeller karşılaştırıldı. [23]' de 0 merkezli ve 2.5 ölçekli Cauchy dağılımını önerdi.

8. Markov Zinciri Monte Carlo Algoritması, Yakınsama Teşhisleri ve Uyum Kriteri

Markov Zinciri Monte Carlo (MCMC) metodu, sonsal miktarların hesaplanması ve sonsal dağılımlardan örnekleme elde etmek için yapılan genel bir simülasyon yöntemidir. MCMC, bir hedef dağılımdan örnekleme yapma yöntemidir.

Monte Carlo Standart Hata (MCSE) sonsal ortalama tahminin standart hatası olarak bilinir ve simülasyonun doğruluğunun ölçüsüdür. Markov Zincirinin yakınsamasını değerlendirmek amacıyla Monte Carlo standart hata/ standard sapma oranı %5 den küçük olduğunda Thumb kuralına göre MCMC zinciri yakınsamaya ulaşır.

Yakma, pratikte sonsal dağılımdan çıkarım yapılırken başlangıç değerinin etkisini minimum yapmak için Markov Zincir örnekleminin başlangıç bölümünün atılmasını ifade eder. Böylece sonsal çıkarım için iyi örnekler kullanılır

İnceltme, örneklem otokorelasyonlarını azaltmak için her seriden elde edilen k. yinci simülasyonun tutulması, Markov Zincirinin küçültülmesi anlamına gelir. Markov Zinciri başlangıç değeri uygulamalarda genellikle maksimum olabilirlik tahmin değerleri kullanılır.

Etkin örneklem büyüklüğü(ESS) Markov zincirinin karışımı ile yakından ilgili bir ölçümdür. Etkin bir örnek büyüklüğü ile simülasyon örnek büyüklüğü arasındaki büyük çelişki, zayıf karışımı dolayısıyla yakınsamanın sağlanmadığını göstermektedir.

İz grafikleri zincirin hedef dağılıma ulaşıp ulaşmadığını görsel yöntemler ile gösterir. Küçük dalgalanmalarla durağan olan bir zincir hedef dağılıma yakınsadığını gösterir.

Sapma Bilgi kriteri(DIC) [27] tarafından geliştirilen ve son zamanlarda kullanımı yaygın olan ve özellikle Bayesci yaklaşımda kullanılan bir uyum kriterdir. . y örneklem verisi, θ ise veriyi ürettiği varsayılan modelin parametresi olmak üzere $DIC = Dbar + pD = Dhat + 2pD$ ile hesaplanır. Burada pD veriye en iyi uyumlu modelde yer alması gereken etkin parametre sayısı olarak ölçülen model kompleksliğidir ve $pD = Dbar - Dhat$ eşitliği ile elde edilir. pD, modeldeki parametrelerin sayısı için bir sınırlama sağlar ve yaklaşık olarak modeldeki parametrelerin sayısı ile ilgilenir. Önsel bilgi ve veri arasında önemli bir fikir ayrılığı (çelişki) olduğu zaman ya da bir parametre için sonsal dağılımın çok zayıf özet istatistiği ve çok büyük sapma vermesi durumlarında pD'nin değeri negatif olacaktır. Sapmanın sonsal dağılımının ortalaması Dbar dır ve Gibbs örneklemesinin bir iterasyonunun sonunda hesaplanmış log-olabilirliklerin ortalamasıdır. Dhat ise θ'nın sonsal ortalamasındaki sapmaya dayalıdır. Kurulan modellerde DIC değeri küçük olan model en uyumlu model olarak belirtilir.

9. Bulgular ve Tartışma

Bu çalışmada, farklı doktorların akciğer kanseri hastalarının aldıkları tedaviden sonraki yaşam kalitelerini etkileyen faktörleri ölçmek için yapılan bir çalışmaya ait verinin belli bir kısmı kullanılmıştır. Hastaların aldıkları tedaviden sonra iyileşip iyileşmeme durumları hastaların bazı demografik ve hastalık etkenlerinin etkisi olup olmadığı incelendi. Çalışmada rasgele seçilen 9 doktor tarafından görülen 360 akciğer hastası için iyileşme durumu (1-İyileşti, 2-İyileşmedi) bağımlı değişken; yaş, genetik etken (1-var, 2-yok), pro-enflamatuar stokin değerlerinin düşük düzeyi (İnterlökin-6 değeri), kreaktiv protein değeri, kanser evresi (1- 1. Evre, 2- 2.evre, 3- 3.evre, 4- 4. evre), akciğer kapasitesi ölçümleri sabit etkili değişkenler ve doktorlar rasgele etkili değişken olarak ele alınmıştır. Bu değişkenlere ait demografik özellikler aşağıda verilen Tablo 2 deki gibidir.

Akciğer kanser verisi ikili bağımlı değişken, rasgele seçilen 9 grup doktor ve her grupta 40 hasta bulunmasıyla genelleştirilmiş lineer karma modele uymaktadır. Bu model

$$\begin{aligned} & f(y_{ij} \mid \beta, \gamma_j) = \exp\left\{\left(y_{ij}\theta_{ij} - b(\theta_{ij})\right)\Big/\phi^2 - c(y_{ij}, \phi)\right\} \\ & \eta_{ij} = g(\mu_{ij}) = X_{ij}\beta + Z_{ij}\gamma_j \\ & \gamma_j \sim N(0, G), \qquad j = 1, \ldots, 9; \qquad i = 1, \ldots, 40 \end{aligned} \tag{19}$$

şekilde ifade edilebilir. Bağımlı değişken (iyilesmedrm) Y_{ij} j yinci doktorun i yinci hastası iyileşti ise, 1 ve iyileşmedi ise, 0 olacak şekilde üstel dağılım ailesinden biri olan Bernoulli dağılımına $Y_{ij} \mid \gamma_j \sim Bernoulli(p)$ uygundur. Lineer tahmin edici $\eta_{ij} = X_{ij}\beta + Z_{ij}\gamma$ dır. Burada X matrisi 360x7 boyutlu sabit etkili tasarım matrisi, Z matrisi 360x9 rasgele etkili tasarım matrisidir. β_{7x1} sabit etkiye ait parametre vektörü ve γ_{9x1} rasgele etkiye ait parametre vektörüdür. Rasgele etki $\gamma_j \sim N(0, \sigma^2_{\gamma_j})$ olmak üzere, link fonksiyonu lojit: $\eta = lojit(p) = \log\left(\frac{p}{1-p}\right)$ dir.

Tablo 2. Hastaların demografik ve hastalık etkenlerinin karakteristik özellikleri (n=360)

Değişkenler	Değer(Yüzde)
İyileşme durumu	
İyileşti(0)	91 (%25.3)
İyileşmedi(1)	269 (%74.7)
Genetik etken	
Var(1)	72(%20)
Yok (0)	288(%80)
Kanser evresi	
1.Kanser evresi(1)	115(%31.9)
2.Kanser evresi(2)	141(%39.2)
3.Kanser evresi(3)	66(%18.3)
4.Kanser evresi(4)	38(%10.6)
Akciğer kapasitesi	
Ortalama	0.778
Min-Max	0.134-0.999
Range	0.778
Yaş	
Ortalama	50.87
Min-Max	35-69
Range	34
Pro-enflamatuar stokin değerlerinin düşük düzeyi;	
İnterlökin-6 değeri	
Ortalama	4.100
Min-Max	0.17-18.14
Range	17.96
Kreaktiv protein değeri	
Ortalama	5.178
Min-Max	0.355-20.485
Range	20.130

10. Genelleştirilmiş Lineer Karma Modeller için Bayesci Yaklaşım

Bayesci yaklaşımın en önemli ve en dikkat edilmesi gereken konusu önsel seçimidir. Önsel seçiminde kesin bir yöntem yoktur. Özellikle parametre sayısının fazla olduğu modellerde hesaplama oldukça zordur. Bu çalışmada da parametre sayısı fazla olduğundan önsel seçiminde, hiper parametreler ve yakınsama konusunda sıkıntılar yaşanmıştır. Modelde bulunan sabit etkili parametre vektörüne (Beta), rasgele etkili parametre vektörüne (Gama) ve varyans bileşenine (S2) önceki çalışmalardan faydalanılarak farklı önseller atandı ve Tablo 3 de görülen 6 durum oluşturulmuştur. Burada N(,sd=) normal dağılımı, Cauchy(,) Cauchy dağılımını ve Igamma(,s=) ters gamma dağılımını ve Uniform(,) Uniform dağılımını temsil etmektedir.

Tablo 3. Bayesci GLKM için farklı önselli durumlar

Durumlar	Parametreler		
	Beta	Gama	S2
Durum1	N(0,sd= 0.001)	N(0.sd= 1000)	Cauchy(0, 2.5)
Durum2	Bilgilendirici	N(0,sd= 10000)	Cauchy(0, 2.5)
Durum3	Bilgilendirici	N(0,sd= 10000)	Igamma(1,s= 0.01)
Durum4	N(0,sd= 10000)	N(0,sd= 10000)	Igamma(1,s= 0.01)
Durum5	N(0,sd= 10000)	N(0,sd= 10000)	Uniform(0,100)
Durum6	Bilgilendirici	N(0,sd= 10000)	Uniform(0,100)

Aşağıda Durum 1 için belirlenen önsel dağılımlar, sonsal özetler, yakınsama teşhisleri açıklamalı ve ayrıntılı olarak verilmiştir. Diğer durumlara ait sadece önsel dağlımlar ve gerekli açıklamalar verilecektir.

Durum 1

[26] da 0 merkezli ve 2.5 ölçekli Cauchy dağılımı önerildi ve [12] de bilgilendirici olmayan önsel olarak bir belirsiz dağılım olan normal önselin kullanımını gösterdi. [9] yeteri kadar örneklem durumunda β için normal önselin kullanışlı olduğunu belirtti. Bundan dolayı, Durum 1' de modeldeki parametreler için önsel dağılımlar aşağıdaki gibi seçilmiştir. Uygun önselleri elde etmek için hiperparametreler pozitif olmak üzere,

$$\begin{aligned} &\beta \propto N(0, sd = 0.001) \\ &\gamma \sim N(0,\ sd = 1000) \\ &\sigma^2_{\gamma} \sim Cauchy(0, 2.5) \end{aligned} \tag{20}$$

Durum 1 için SAS 9.3 de MCMC modülü [28] için program girdisi aşağıdaki gibi oluşturulmuştur.

```
title 'Durum 1';
ods graphics on;
      proc mcmc data=akciger outpost=postout
      seed=27513 nmc=1000000 ntu=10000
      nbi=10000 thin=25 dic diag=all plots=all
      stats=all;
      parms beta0 0 beta1 0 beta2 0 beta3 0 beta4 0
      beta5 0 beta6 0 s2 1;
prior s2 ~ cauchy(0,2.5);
prior beta: ~ normal(0,sd=0.001);
random gama ~ normal(0, sd=1000) subject=did
monitor=(gama);
      y=beta0+beta1*akcigerkap+beta2*yas+beta3*
      genetikdrm+beta4*kanserevresi+beta5*IL6+b
      eta6*CRP;
pi = logistic(y);
```

```
model iyilesmedrm ~ binary(p = pi);
run;
ods graphics off;
```

Yakınsamanın sağlanması için regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 1000000 MCMC örneklemi ile ilk 10000 iterasyona bir yakma periyodu uygulanarak elde edildi. Örneklem otokorelasyonlarını azaltmak için inceltme oranı 25 olarak alındı. Sonsal örneklem sayısı 40000 olarak elde edildi. Veri büyük olduğundan yakınsama problemlerini gidermek fazla zaman aldı. Yakınsama teşhisi olarak Monte Carlo standart hata ve etkin örneklem büyüklükleri ve yakınsama iz grafikleri dikkate alındı.

Sonsal özetler her parametrenin temel istatistiklerini içerir. Durum 1 için sonsal ortalamalar, standart sapma tahminleri ve HPD aralığı Tablo 4 de görüldüğü gibidir. HPD aralığı sıfırı kapsarsa, bu o değişkenin istatistiksel olarak anlamlı olmadığını gösterir. Sonuçlara göre, modelde yer alan sabit etkili parametreler yaş, akciğer kapasitesi (akcigerkap), İnterlökin-6 değeri (IL6), kreaktiv protein değeri (CRP), genetik etken (genetikdrm), rasgele etkili parametreler kovaryans değişken ve doktorların (DID) akciğer kanseri hastalarının aldıkları tedaviden sonra iyileşip iyileşmeme durumlarını önemli bir şekilde etkilemediği %95 güvenle elde edilmiştir. Kanser evresi değişkeninin ise, akciğer kanseri hastalarının iyileşmesi üzerinde önemli etkiye sahip olduğu söylenebilir.

Markov Zincirin yakınsamasını değerlendirmek amacıyla Monte Carlo standart hata/ standard sapma oranı %5 den küçük olduğundan thumb kuralına göre yakınsama sağlanmıştır. Etkin örneklem büyüklüğün de ise, simülasyon örnek büyüklüğü (40000) arasındaki büyük çelişki, zayıf karışımı gösterir ki, burada sadece kovaryans parametresinin zayıf karışıma sahip olduğu iz grafiklerine bakılmadan anlaşılabilir. Böylece Durum 1 de yakınsamanın her parametre için sağlandığı görülmektedir. Zincirin karışımını gösteren Markov zinciri iz grafikleri, sonsal yoğunluk grafikleri ve otokorelasyon grafikleridir. Şekil 1 de görüldüğü gibidir.

Diğer durumlarda da önsel dağılımlarda önceki çalışmalar ve deneyimlerden yararlanılarak seçilmiş, benzer yollar izlenmiş ve aşağıda belirtilmiştir.

Durum 2

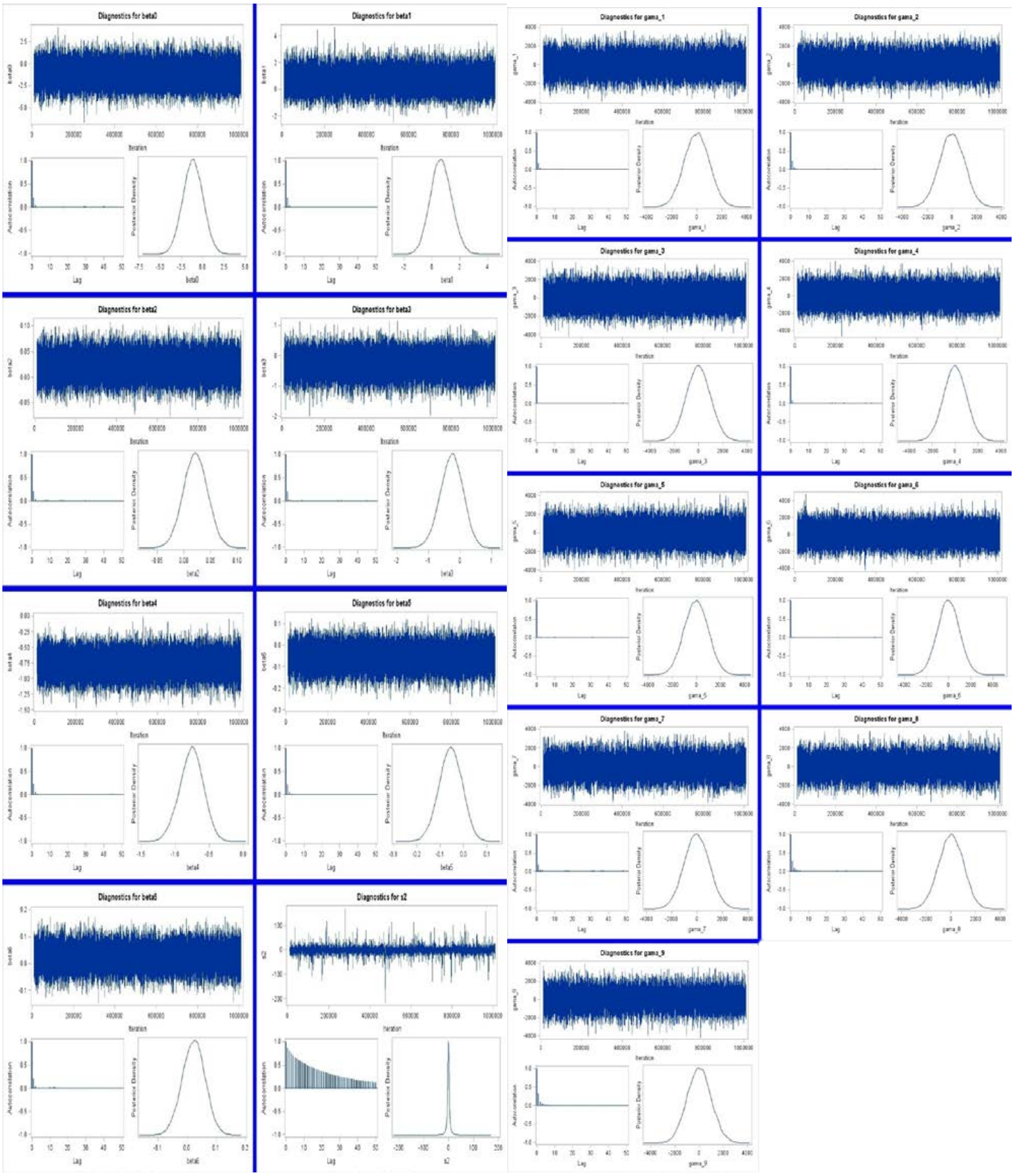
[16] model seçiminde regresyon parametreleri için bilgilendirici önsel önerdiler. Modelde sabit etkili parametreler ile ilgili önceki çalışmalardan elde edilmiş bilgi ile sabit etkili parameterelere bilgilendirici önsel dağılım ve varyans bileşenine Cauchy dağılımı aşağıdaki gibi alınmıştır:

$$\begin{aligned}
\beta_0 &\sim N(-0.4122\ , sd = 1.6524) \\
\beta_1 &\sim N(1.2551,\ sd = 1.0149\ ) \\
\beta_2 &\sim N(0.004391,\ sd = 0.02866\ ) \\
\beta_3 &\sim N(-1.0170,\ sd = 0.5048\ ) \\
\beta_4 &\sim N(-1.0472,\ sd = 0.2306\ ) \\
\beta_5 &\sim N(-0.07727,\ sd = 0.06722) \\
\beta_6 &\sim N(0.02154,\ sd = 0.04798\ ) \\
\gamma &\sim N(0,\ sd = 10000) \\
\sigma_\gamma^2 &\sim Cauchy(0, 2.5)
\end{aligned} \tag{21}$$

Regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 1000000 MCMC örneklem ile sağlandı. İlk 10000 iterasyona bir yakma periyodu uygulandı ve inceltme oranı 25 alındı. Sonsal örneklem sayısı 40000 olarak elde edildi.

Tablo 4. Durum 1 için sonsal özetler ve yakınsama teşhisleri

	Ortalama	Standart sapma	HPD Aralığı		MCSE/SD	Etkin Örn.B.
Sabit	-1.140	12.043	-3.555	1.167	0.006	27098.100
akcigerkap	0.675	0.763	-0.729	2.237	0.006	27382.300
yas	0.022	0.023	-0.023	0.066	0.006	26884.800
genetikdrm	-0.268	0.361	-0.968	0.443	0.006	27267.600
kanserevresi	**-0.761**	**0.175**	**-1.115**	**-0.427**	0.006	25636.300
IL6	-0.055	0.0450	-0.152	0.043	0.006	26467.200
CRP	0.023	0.037	-0.054	0.101	0.006	27071.700
kovaryans	-0.557	16.256	-28.204	26.640	0.033	923.700
DID1	2.647	1001.300	-1895.900	1996.500	0.006	28276.400
DID2	4.138	1003.600	-2025.100	1893.500	0.006	25524.000
DID3	3.483	995.400	-1926.800	1968.200	0.005	40000.000
DID4	-3.708	998.700	-1974.200	1957.000	0.005	34905.100
DID5	-11.010	1000.100	-1941.600	1956.000	0.005	38814.600
DID6	4.114	997.500	-1944.300	1969.100	0.005	39203.500
DID7	-7.103	996.800	-1934.200	1970.100	0.006	27527.000
DID8	12.553	999.700	-2014.900	1875.100	0.007	21325.100
DID9	30.015	998.500	-1980.900	1908.800	0.007	19242.300



Şekil 1. Durum 1 için MCMC yakınsama iz grafikleri

Durum 3

Genelde, literatürde varyans bileşeni için normal dağılımın eşlenik dağılımı olan ters gamma önseli önerilmiştir. Bundan dolayı, Durum 3 de parametreler için önsel dağılımlar Denklem 22'deki gibidir.

Yakınsamanın sağlanması için regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 900000 MCMC örneklem sayısı ile sağlandı. İlk 10000 iterasyona bir yakma periyodu uygulandı. Örneklem otokorelasyonlarını azaltmak için inceltme oranı 25 olarak alındı. Sonsal örneklem sayısı 36000 olarak elde edildi.

$$
\begin{aligned}
\beta_0 &\sim N(-0.4122\ , sd = 1.6524) \\
\beta_1 &\sim N(1.2551,\ sd = 1.0149\ ) \\
\beta_2 &\sim N(0.004391,\ sd = 0.02866\ ) \\
\beta_3 &\sim N(-1.0170,\ sd = 0.5048\ ) \\
\beta_4 &\sim N(-1.0472,\ sd = 0.2306\ ) \\
\beta_5 &\sim N(-0.07727,\ sd = 0.06722) \\
\beta_6 &\sim N(0.02154,\ sd = 0.04798\ ) \\
\gamma \sim N(0,\ sd &= 10000)\ \sigma_\gamma^2 \sim \text{Igamma}(1, 0.01)
\end{aligned}
\qquad (22)
$$

Durum 4

Durum 4 de sabit etkili parametreler için difüz normal önsel ve varyans bileşeni için ters gamma önsel alınmıştır:

$$\begin{aligned} &\gamma \sim N(0,\ sd = 10000) \\ &\beta \propto N(0, sd = 10000) \\ &\sigma_\gamma^2 \sim \text{igamma}(1, 0.01) \end{aligned} \tag{23}$$

Yakınsamanın sağlanması için regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 900000 MCMC örneklem ile sağlandı. İlk 10000 iterasyona bir yakma periyodu uygulandı. Örneklem otokorelasyonlarını azaltmak için inceltme oranı 25 olarak alındı. Sonsal örneklem sayısı 36000 olarak elde edildi.

Durum 5

[13] ve Gelman [15] tarafından varyans bileşenleri için Uniform önsel önerildi. Durum 5 de Uniform dağılım içinde alınan hiperparametreler dikkate alındı. Alınan hiperparemetreler pozitif olmak üzere,

$$\begin{aligned} &\beta \propto N(0, sd = 10000) \\ &\gamma \sim N(0,\ sd = 10000) \\ &\sigma_\gamma^2 \sim \text{uniform}(0, 100) \end{aligned} \tag{24}$$

Yakınsamanın sağlanması için regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 900000 MCMC örneklemi ile sağlandı. İlk 10000 iterasyona bir yakma periyodu uygulandı. Örneklem otokorelasyonlarını azaltmak için inceltme oranı 25 olarak alındı. Sonsal örneklem sayısı 36000 olarak elde edildi.

Durum 6

Durum 6 da varyans bileşenleri için Uniform önsel ve sabit etkili parametreler için önceki çalışmalardan elde edilmiş bilgi ile bilgilendirici önsel alınmıştır.

$$\begin{aligned} &\beta_0 \sim N(-0.4122\ , sd = 1.6524) \\ &\beta_1 \sim N(1.2551,\ sd = 1.0149\ ) \\ &\beta_2 \sim N(0.004391,\ sd = 0.02866 \\ &\beta_3 \sim N(-1.0170,\ sd = 0.5048\ ) \\ &\beta_4 \sim N(-1.0472,\ sd = 0.2306\ ) \\ &\beta_5 \sim N(-0.07727,\ sd = 0.06722 \\ &\beta_6 \sim N(0.02154,\ sd = 0.04798 \\ &\gamma \sim N(0,\ sd = 10000) \\ &\sigma_\gamma^2 \sim \text{uniform}(0, 100) \end{aligned} \tag{25}$$

Yakınsamanın sağlanması için regresyon katsayılarının tümünün başlangıç değerleri 0 ve varyans bileşeninin başlangıç değeri 1 alındı. Yakınsama 900000 MCMC örneklem ile sağlandı. İlk 10000 iterasyona bir yakma periyodu uygulandı. Örneklem otokorelasyonlarını azaltmak için inceltme oranı 25 olarak alındı. Sonsal örneklem sayısı 36000 olarak elde edildi.

Tablo 5. Durumların Karşılaştırılması

	Durumlar	**Parametreler**								**Uyum Kriteri**
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	S2	DIC
Bayesci Yaklaşım	**Durum 1**	-	-	-	-	√	-	-	-	393.641
	Durum 2	-	-	-	-	√	-	-	-	390.043
	Durum 3	-	-	-	-	√	-	-	√	390.034
	Durum 4	-	-	-	-	√	-	-	√	393.667
	Durum 5	-	-	-	-	√	-	-	√	393.663
	Durum 6	-	-	-	-	√	-	-	√	390.038

11. Tartışma ve Sonuç

Yüksek boyutlu verilerde önsel dağılımın hiperparametre seçiminin önemli ve zor olması daha fazla zaman ve çaba gerektirmektedir. Bayesci yaklaşımda Markov zincirinin yakınsama problemini gidermek için hiperparametrelerin yeterli büyüklükte olması gerekir. Bu çalışmada, hiperparametreler Markov zinciri yakınsamasının zamandan tasarruf sağlayacak ve daha az örneklem çekilecek şekilde seçilmeye çalışılmıştır. Durum 1, Durum4 ve Durum 5 de regresyon parametreleri için literatürde kullanılan bilgilendirici olmayan belirsiz normal önsel dağılım verinin modele uyumunu azaltmıştır. Ancak, regresyon parametrelerinin bilgilendirici önsel dağılıma sahip olması verinin modele uyumunu arttırmaktadır. Tablo 5 te, Durum 2, Durum3 ve Durum 6 da uyum kriteri DIC in en küçük değerleri regresyon parametrelerine bilgilendirici önsel dağılım seçildiğinde olduğu görülmektedir Bu yüzden regresyon parametreleri için bilgilendirici önsel dağılım seçilmesi en uygundur. Varyans bileşenleri için seçilen önsel dağılımın modelin uyumuna katkısı olmadığı, yakınsama için önemli olduğu yakınsama teşhislerinden ve iz grafiklerinden de görülür.

Kaynakça

[1] Nelder, J. A and Wedderburn, R. W. M., 1972. Generalized Linear Models, Journal of the Royal Statistical Society. Series A (General) Vol. 135, No. 3 , pp. 370-384.

[2] Zeger, S. L., Karim, M. R., 1991. Generalized Linear Models with Random Effects: A Gibbs Sampling Approach, Journal of the American Statistical Association, 86, 79–86.

[3] Gamerman, D. Statistics and Computing (1997) 7: 57. doi:10.1023/A:1018509429360.

[4] Booth, J. G., Hobert, J. P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society Series B (Statistical Methodology), 61, 265–285.

[5] Natarajan, R., Kass, R. E., 2000. Reference Bayesian Methods for Generalized Linear Mixed Models, Journal of the American Statistical Association, 95, 227–237.

[6] Zhao, Y., Staudenmayer, J., Coull, B.A., Wand, M.P., 2006. General design Bayesian generalized linear mixed models. Statistical Science, 21, 35–51.

[7] Tsai, M. & Hsiao, C.K. Comput Stat (2008) 23: 587. doi:10.1007/s00180-007-0100-x.

[8] Evangelou, E. A. 2009. Bayesian and Frequentist Methods for Approximate Inference in Generalized Linear Mixed Models. University of North Carolina at Chapel Hill. Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

[9] Bedrick, E. J., Christensen, R., Johnson, W., 1997. Bayesian binomial regression: Predicting survival at a trauma center. Amer. Statist. 51 211–218.

[10] Natarajan, R., McCulloch, C.E., 1998. Gibbs sampling with diffuse proper priors: a valid approach to data-driven ınference? Journal of Computational and Graphical Statistics,7, 67/277.

[11] Diggle, P. J., Tawn, J. A., Moyeed, R. A., 1998. Model-based Geostatistics, Journal of the Royal Statistical Society, Series C: Applied Statistics, 47, 299–326.

[12] Berger, J. O., De Oliveira, V., Sans´o, B., 2001. Objective Bayesian Analysis of Spatially Correlated Data, Journal of the American Statistical Association, 96, 1361–1374.

[13] Browne, W.J., Draper, D., 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models, International Society for Bayesian Analysis, 1, Number 3, pp. 473-514.

[14] Kass, R.E. , Natarajan, R., 2006. A Default Conjugate Prior for Variance Components in Generalised Linear Mixed Models (Comment on Article by Browne and Draper). Bayesian Analysis, 1(3), 535-542.

[15] Gelman, A., 2006. Prior Distributions for Variance Parameters in Hierarchical Models, International Society for Bayesian Analysis, 1, Number 3, pp. 515-534.

[16] Chen M-H, Ibrahim JG, Shao Q-M, Weiss RE, 2003. Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models. Journal of Statistical Planning and Inference; 111:57–76. 586.

[17] Tiao, G. C. and Tan, W., 1965. Bayesian analysis of random effect models in the analysis of variance. i. Posterior distribution of variance components. Biometrika, 51: 37-53.

[18] Box, G.,Tiao, G., 1973. Bayesian inference in statistical analysis, John Wiley&Sons.

[19] Demidenko, E., 2004. Mixed models: Theory and Applications. In: Wiley Series in Probability and Statistics. Hoboken, New Jersey.

[20] Congdon, P., 2003. Applied Bayesian Modelling. John Wiley Sons, England Jeffreys, A., 1961, The Theory of Probability, Cambridge University Press, Cambridge 2nd Edition. .

[21] Antoni, K., Beirlant, J., 2006. Actuarial statistics with generalized linear mixed models, Insurance Mathematics ve Economics.

[22] Jeffreys, H., 1961. Theory of probability Oxford univ Press, New York.

[23] Mehmet A. Cengiz, M. A., Terzi, E., Şenel, T. ve Murat, N., 2012. Lojistik Regresyonda Parametre Tahmininde Bayesci Bir Yaklaşım, AKU J. Sci.12 (2012), 011302 ,(15-22).

[24] Gelman, A., 2005. Analysis of variance: why it is more important than ever (with discussion). Annals of Statistics.

[25] Daniels, M.J., 1999.A prior for the variance in hierarchical models. Canadian Journal of Statistics, 27:569/580.

[26] Gelman, A., Jakulin, A., Pittau, M. G., Su¸S. 2008. A weakly informative default prior distribution for logistic and other regression models. Annals of Applied Statistics, 1360-1383.

[27] Spiegelhalter, David J.; Best, Nicola G.; Carlin, Bradley P.; Van der Linde, Angelika (2002). Bayesian measures of model complexity and fit (with discussion)". Journal of the Royal Statistical Society, Series B. 64 (4): 583–639. doi:10.1111/1467-9868.00353. JSTOR 3088806. MR 1979380.

[28] SAS Institute Inc. 2011. SAS/STAT 9.3 User's Guide. Cary, NC, 7745 s.,USA.