İstatistik Araştırma Dergisi Cilt 01 No 02 Özel Sayı Sayfa 225-235 Ağustos 2002

Gibbs Örneklemesi İle Karışık Doğrusal Bir Modeldeki Varyans Unsurları Hakkında Bayesci Yorumlama

Mehmet Ziya FIRAT*

ÖZET

Markov zinciri Monte Carlo yöntemleri sayısal genetik modellerde parametrelerin marginal sonsal dağılımları hakkında yorumlamalar yapmada giderek artan bir biçimde uygulanmaktadır. Bu makale, böyle metotlardan biri olan Gibbs örneklemesinin karışık bir dogrusal modelde parametreler hakkında Bayesci yorumlamaya uygulanmasını ele almaktadır. Gibbs örneklemesi, bileşik veya marjinal yoğunluklar doğrudan doğruya elde edilmeseler dahi, yorumlamalar yapılmasına izin veren sayısal bir integral yöntemidir. Tam şartlı yoğunluk fonksiyonlarının tamamından sırayla değişkenlerin üretilmesi esasına dayanmaktadır. Tam şartlı yoğunluk, modelde bütün diğer parametreler verildiğinde bir değişkenin yoğunluğudur. Varyans unsurlarını tahmin probleminde, ilgi duyulan bileşik yoğunluk fonksiyonu, gözlemler verildiğinde sabit etkiler, bileşik etkiler, rassal etkiler ve varyans unsurlarının dağılımıdır ve marjinal yoğunluklar, gözlemler verildiğinde sabit etkiler, bileşik etkiler, rassal etkiler veya varyans unsurlarının dağılımlarıdır. Bu araştırmada, kısıtlanmış maksimum olabilirlik (REML) ve Gibbs örneklemesi yöntemleri kullanılarak 20438 Türk Holstein-Friesian süt sığırı için süt verimine ait genetik ve fenotipik parametrelerin sonsal dağılımlarının tahminleri elde edilmiştir.

Anahtar Kelimeler *: Gibbs örneklemesi, Bayesci yorumlama, parametre tahmini*

1. GİRİŞ

Bayesci yaklaşımın kullanıldığı birçok problemde, uygun yorumlamalar yapabilmek için parametrelerin marjinal dağılımlarına ihtiyaç duyulmaktadır. Bununla birlikte, bileşik sonsal dağılımların karmaşıklığı nedeniyle, parametreler veya bunların fonksiyonlarının marjinal dağılımlarını analitik yöntemlerle elde etmek oldukça zor veya imkansızdır. Bu durum varyans unsurları hakkındaki yorumlamalar dahil olmak üzere birçok uygulamalı problemler için geçerlidir. Dolayısı ile, bileşik yoğunluk fonksiyonundan herbir parametreye ait marjinal fonksiyonu elde etmede analitik integral yerine sayısal integral yöntemlerine gereksinim bulunmaktadır.

*Akdeniz Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, ANTALYA, e-mail : mzf@agric.akdeniz.edu.tr

Son yıllarda, sadece şartlı dağılımlar mevcut olduğunda çoklu bileşik ve marjinal sonsal dağılımların örneklemesini yapan sayısal integral yöntemleri birçok bilim adamının dikkatini çekmiştir. Bu yöntemlerin ortak yönü, örneklemenin stokastik süreçle yapılmasıdır. Bu fikir esasını Metropolis algoritmasının uyarlanmasından almaktadır. Bununla birlikte, Metropolis algoritmasının genelleştirilmiş şekli olan Markov-Zinciri Monte Carlo örnekleme yöntemlerini tavsiye eden Hastings (1970), Gibbs örneklemesinin yaygın olarak kullanılmasını sağlamıştır. Geman ve Geman (1984)'ün Bayesci görüntü analizi konusunda yaptığı çalışma Bayesci istatistiksel yorumlamanın uygulanabilirliğini daha da arttırmıştır.

Yakın zamanlarda, Gelfand ve Smith (1990) Gibbs örneklemesi ve diğer Monte Carlo yöntemlerini kapsamlı bir şekilde incelemişlerdir. Gibbs örneklemesi algoritması, tam şartlı yoğunluk fonksiyonlarının tamamından örnekleme yaparak modeldeki bütün parametrelerin bileşik yoğunluk fonksiyonuna yaklaşımda bulunur. Tam şartlı yoğunluk fonksiyonu, modelde bütün diğer parametreler verildiğinde ilgi duyulan değişkenin yoğunluğudur. Varyans unsurlarının tahmini durumunda, ilgi duyulan bileşik yoğunluk fonksiyonu, gözlemler verildiğinde sabit etkiler, rassal etkiler ve varyans unsurlarının tamamının dağılımıdır. Bu durumda ilgi duyulan marjinal yoğunluklar, gözlemler verildiğinde sabit etkiler, rassal etkiler veya varyans unsurlarının dağılımlarıdır.

Bu makalenin esas amacı, Gibbs örneklemesi kullanılarak sürekli değişkenler için karışık doğrusal modellerden varyans unsurları ve fonksiyonları hakkında Bayesci yorumlamaların nasıl yapıldığını göstermektir. Verilerin tek değişkenli normal bir dağılıma sahip oldukları varsayılmaktadır. Analizde kullanılan karışık model, rassal etki, sabit etkiler ve bileşik etkilerden oluşmaktadır. Makalede önce teorik varsayımlar ve Gibbs örneklemesi hakkında ayrıntılar verildikten sonra 20438 Türk Holstein-Friesian sığırının süt verimine ait sayısal örnek kullanılarak genetik ve fenotipik parametrelerin REML tahminleri ve Gibbs örneklemesinden elde edilen sonsal beklenen değerleri sunulmuştur.

2. TEORİK VARSAYIMLAR

Bu kısımda, kullanılan model ve varsayımlar, genel Bayesci tahmin yöntemi, kullanılan algoritma ve hesaplamayla ilgili Gibbs adımları hakkında bazı bilgiler verilmeye çalışılacaktır.

2.1 Model ve Varsayımlar

Matris gösteriminde N boyutlu gözlem vektörü $\mathbf{y}$ için genel karışık bir model şöyledir

$$\mathbf{y} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{C}\boldsymbol{\beta} + \mathbf{D}\mathbf{s} + \mathbf{e} \tag{1}$$

burada $\boldsymbol{\alpha}$, g boyutlu sabit etkiler vektörü, $\mathbf{H}$, Nxg boyutlu sabit etkilere ait desen matrisi, $\mathbf{C}$, bileşik etkilerin merkezi değerini içeren Nxc boyutlu bir matris, $\boldsymbol{\beta}$, bileşik etkilere ait c boyutlu regresyon katsayıları vektörü, $\mathbf{D}$, Nxs boyutlu rassal etkilere ait desen matrisi, $\mathbf{s}$, s boyutlu rassal etkiler vektörü ve $\mathbf{e}$, N boyutlu rassal hatalar vektörüdür. Yukarıdaki (1) nolu modelle ilgili varsayımlar şöyledir:

$$E(\mathbf{y} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{s}, \sigma_s^2, \sigma_e^2) = \mathbf{H}\boldsymbol{\alpha} + \mathbf{C}\boldsymbol{\beta} + \mathbf{D}\mathbf{s}, \quad E(\mathbf{s}) = \mathbf{0}, \quad E(\mathbf{e}) = \mathbf{0}, \mathrm{Var}(\mathbf{s} \mid \sigma_s^2) = \mathbf{G} = \mathbf{I}_s \sigma_s^2,$$

$$\mathrm{Var}(\mathbf{e} \mid \sigma_e^2) = \mathbf{R} = \mathbf{I}_N \sigma_e^2, \quad \mathrm{Cov}(\mathbf{s}, \mathbf{e} \mid \sigma_s^2, \sigma_e^2) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{y} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{s}, \sigma_s^2, \sigma_e^2) = \mathbf{I}_N \sigma_e^2,$$

$$E\begin{bmatrix} \mathbf{y} \\ \mathbf{s} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{H}\boldsymbol{\alpha} + \mathbf{C}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \text{ve} \quad \mathrm{Var}\begin{bmatrix} \mathbf{y} \\ \mathbf{s} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{D}\mathbf{G}\mathbf{D}' + \mathbf{R} & \mathbf{G}\mathbf{D}' & \mathbf{R} \\ \mathbf{G}\mathbf{D}' & \mathbf{G} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{bmatrix}.$$

2.1.1 Bazı kısıtlamalar

Bazı hayvan ıslahı uygulamalarında, varyans unsurları σ_s^2 ve σ_e^2 yerine bunların oranları veya fonksiyonları hakkında yorumlamaların yapılması gerekebilir. γ, boğa ve hata varyans unsurlarının oranını, σ_s^2 / σ_e^2, temsil etsin. Bu durumda, bir karakterin kalıtım derecesi, h^2, γ'nin artan bir fonksiyonu olup $h^2 = 4/(1+\gamma^{-1})$ şeklinde verilir. γ varyansların bir oranı olduğundan, pozitiftir. Ayrıca, (1)'deki boğa modelinin kullanımı γ üzerine bir üst sınır konulmasını zorunlu kılar, bu u ile gösterilsin. Böylece $0 \le \gamma \le u$ veya $0 \le \sigma_s^2 \le u\sigma_e^2$ sınırları elde edilir. Baba bir üvey kardeş aile yapısı için kalıtım derecesinin 0 ile 1 arasında olması u'nun 1/3'e eşit olmasını gerektirir ve σ_s^2, 0 ve $1/3\,\sigma_e^2$ arasında değerler alır. Bu kısıtlamayı dikkate almayan bir uygulama, kalıtım derecesi tahminlerinin kendi parametre sınırları dışında değerler almasına yol açabilir.

2.2. Bayesci Formülasyon

2.2.1. Önsel dağılımlar

Varyans unsurlarının tahmini amacıyla modeli tam olarak tanımlamak için ilave varsayımlar yapılması gerekmektedir. Önce, modeldeki $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{s}$, σ_s^2 ve σ_e^2 parametreleri için önsel dağılım tayin edilmesi zorunludur. Bu çalışmada sabit ve bileşik etkilere ait parametreler, $\boldsymbol{\alpha}$ ve $\boldsymbol{\beta}$, için düzgün önsel dağılım kullanılacaktır

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \text{sabite}. \qquad (2)$$

Bu, başlangıçta bu parametrelerin değerleri hakkında çok az bilgiye sahip olduğumuz anlamındadır. Sonra, rassal etkilerin, $\mathbf{s}$, normal dağılım gösterdiği varsayılmaktadır

$$\mathbf{s} \mid \sigma_s^2 \sim N_s(\mathbf{0}, \mathbf{I}_s \sigma_s^2). \qquad (3)$$

Varyans unsurları σ_s^2 ve σ_e^2'nin önsel dağılımlarının bileşik yoğunluk fonksiyonunun aşağıdaki gibi olduğu varsayılmaktadır

$$f\left(\sigma_s^2,\sigma_e^2 | v_s,s_s^2,v_e,s_e^2\right) \propto (\sigma_s^2)^{-\frac{1}{2}(v_s+2)} (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp\left[-\frac{1}{2}\left(\frac{v_s s_s^2}{\sigma_s^2}+\frac{v_e s_e^2}{\sigma_e^2}\right)\right]. \quad (4)$$

Bu fonksiyondan, σ_s^2 ve σ_e^2'nin birbirinden bağımsız ve ters χ^2 dağılımları, $\chi^2(v_s,v_s s_s^2)$ ve $\chi^2(v_e,v_e s_e^2)$, izledikleri anlaşılmaktadır. Böylece s_s^{-2} ve s_e^{-2}, sırasıyla σ_s^{-2} ve σ_e^{-2}'nin önsel beklenen değerleri olarak yorumlanabilirler, buna karşılık v_s ve v_e serbestlik derecesine eşdeğer doğruluk (precision) parametreleridir.

Varyans unsurları için diğer dağılımlar kullanılması durumunda Gibbs örneklemesi algoritması daha karmaşık olacaktır. Daha önce yapılan Gibbs örneklemesi çalışmalarında bu parametreler için düzgün dağılımlar kullanılmıştır (Wang vd, 1994; Jensen vd, 1994). Hobert (1994), varyans unsurları için düzgün dağılımlar kullanıldığı zaman bileşik dağılımın her zaman mevcut olmayacağını ve Gibbs örneklemesinden elde edilen sonuçların doğru olmayacağını belirtmiştir. Varyans unsurları için düzgün dağılımların güvenli olarak ne zaman kullanılabileceğini belirlemek için daha fazla çalışmaya gerek vardır.

2.2.2. Olabilirlik fonksiyonu

Son olarak, rassal hata etkilerinin aşağıdaki normal dağılıma sahip olduğu varsayılmaktadır

$$f(\mathbf{y} | \boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{s},\sigma_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}N} \exp\left\{-\frac{1}{2\sigma_e^2}\left[(\mathbf{y}-\mathbf{H}\boldsymbol{\alpha}-\mathbf{C}\boldsymbol{\beta}-\mathbf{D}\mathbf{s})'(\mathbf{y}-\mathbf{H}\boldsymbol{\alpha}-\mathbf{C}\boldsymbol{\beta}-\mathbf{D}\mathbf{s})\right]\right\}. \quad (5)$$

Yukarıda (2)-(5) eşitliklerinde verilen varsayımlar birçok olabilirliğe dayalı yöntemlerde kullanılanlarla aynıdır. Ayrıca, varyans unsurları bilindiği zaman, bu varsayımlar sabit ve rassal etkiler için BLUE ve BLUP çözümleri ile sonuçlanmaktadırlar (Gianola ve Fernando, 1986; Gianola vd, 1990).

2.2.3. Bileşik sonsal yoğunluk fonksiyonu

Bileşik sonsal olasılık yoğunluk fonksiyonu yukarıda (2)-(5)'de tanımlanan önsel ve şartlı dağılımların çarpımı olarak yazılabilir. Gözlemler ve önsel bilgi verildiğinde $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{s}$, σ_s^2 ve σ_e^2 parametrelerinin bileşik yoğunluk fonksiyonu aşağıdaki gibidir

$$f(\alpha,\beta,s,\sigma_s^2\sigma_e^2 \mid y,v_s,s_s^2,v_e,s_e^2) \propto f(y \mid \alpha,\beta,s,\sigma_e^2)\, f(\alpha,\beta)\, f(s \mid \sigma_s^2) f(\sigma_s^2,\sigma_e^2 \mid v_s,s_s^2,v_e,s_e^2)$$

$$\propto (\sigma_e^2)^{-\frac{1}{2}(N+v_e+2)} (\sigma_s^2)^{-\frac{1}{2}(s+v_s+2)} \exp\left[-\frac{1}{2\sigma_s^2}(s's+v_s s_s^2)\right] \qquad (6)$$

$$\times \exp\left\{-\frac{1}{2\sigma_e^2}\left[(y - H\alpha - C\beta - Ds)'(y - H\alpha - C\beta - Ds) + v_e s_e^2\right]\right\}$$

2.2.4. Tam şartlı yoğunluk fonksiyonu

Gibbs örneklemesi yöntemini kullanabilmek için α, β, s, σ_s^2 ve σ_e^2 parametrelerinin herbirinin tam şartlı dağılımına, yani modeldeki diğer parametreler verildiğinde herbir parametrenin sonsal dağılımına ihtiyacımız bulunmaktadır. Bu dağılımların ilk üçünü elde etmek için aşağıdaki yaklaşım kullanılır. Boyutu d olan bir θ vektörünün olasılık yoğunluk fonksiyonu aşağıdaki ifadeye oransal ise

$$\exp\left[-\frac{1}{2}(\theta' A\theta - 2\theta' a)\right],$$

bu durumda θ'nın dağılımı $N_d(A^{-1}a, A^{-1})$'dir, burada **A** pozitif tanımlı bir matristir. α, β, s, σ_s^2 ve σ_e^2'nin şartlı sonsal dağılımları (6) nolu eşitlikteki bileşik sonsal olasılık fonksiyonundan elde edilir ve sırasıyla şöyledir:

$$[\alpha \mid \beta,s,\sigma_s^2,\sigma_e^2,y] = N_g\left((H'H)^{-1}H'(y - C\beta - Ds),\ \sigma_e^2(H'H)^{-1}\right), \qquad (7)$$

$$[\beta \mid \alpha,s,\sigma_s^2,\sigma_e^2,y] = N_c\left((C'C)^{-1}C'(y - H\alpha - Ds),\ \sigma_e^2(C'C)^{-1}\right), \qquad (8)$$

$$[s \mid \alpha,\beta,\sigma_s^2,\sigma_e^2,y] = N_s\left((D'D+\frac{\sigma_s^2}{\sigma_e^2}I_s)^{-1}D'(y - H\alpha - C\beta), \sigma_e^2(D'D+\frac{\sigma_s^2}{\sigma_e^2}I_s)^{-1}\right), \qquad (9)$$

$$[\sigma_s^2 \mid \alpha,\beta,s,\sigma_e^2,y] = \chi^{-2}\left(s+v_s,\ s's+v_s s_s^2\right), \qquad (10)$$

$$[\sigma_e^2 \mid \alpha,\beta,s,\sigma_s^2,y] = \chi^{-2}\left(N+v_e,\ (y - H\alpha - C\beta - Ds)'(y - H\alpha - C\beta - Ds)+v_e s_e^2\right). \qquad (11)$$

burada **H'H** g tane sabit etkiye ait frekansları veren köşegen matris, **C'C** bileşik etkilerin düzeltilmiş kareler ve çarpımlar matrisi ve **D'D** boğaların kızlarının sayısını, n_i, veren köşegen matrisi temsil etmektedir.

2.2 Gibbs Örneklemesi

Bayesci modellerde sonsal dağılımları incelemede güçlü bir iteratif algoritma olan Gibbs örneklemesi, sıra ile tam şartlı dağılımlardan örnekleme yaparak modeldeki parametrelerin hepsinin bileşik sonsal yoğunluk dağılımına örnekler üretir. Bu çalışmada, algoritmanın kullanımı Wang vd (1993) ve Fırat vd (1997)'nin tanımlandığı şekli ile tek değişkenli analizler için aşağıdaki gibi yerine getirilmiştir:

1) Sırasıyla bütün değişkenlere (α, β, s, σ_s^2 ve σ_e^2) rastgele bir başlangıç değeri tayin edilir, α_0, β_0, s_0, σ_{s0}^2 ve σ_{e0}^2.
2) $[\alpha \mid \beta_0, s_0, \sigma_{s0}^2, \sigma_{e0}^2, y]$'den sabit etkiler üretilir (α_1) ve α güncelleştirilir,
3) $[\beta \mid \alpha_1, s_0, \sigma_{s0}^2, \sigma_{e0}^2, y]$'den bileşik etkiler üretilir (β_1) ve β güncelleştirilir,
4) $[s \mid \alpha_1, \beta_1, \sigma_{s0}^2, \sigma_{e0}^2, y]$'den rassal etkiler üretilir (s_1) ve s güncelleştirilir,
5) $[\sigma_s^2 \mid \alpha_1, \beta_1, s_1, \sigma_{e0}^2, y]$'den σ_{s1}^2 üretilir ve σ_s^2 güncelleştirilir,
6) $[\sigma_e^2 \mid \alpha_1, \beta_1, s_1, \sigma_{s1}^2, y]$'den σ_{e1}^2 üretilir ve σ_e^2 güncelleştirilir,
7) $3\sigma_s^2 \geq \sigma_e^2$ ise 5. ve 6. adımlar $3\sigma_s^2 \leq \sigma_e^2$ koşulu sağlanana tekrar edilir.

Böylece her parametre (=değişken) rastgele bir sıra ile ziyaret edilmiş olur ve bu döngü (α_0, β_0, s_0, σ_{s0}^2 ve σ_{e0}^2)'dan $(\alpha_1, \beta_1, s_1, \sigma_{s1}^2, \sigma_{e1}^2)$'e bir geçişi tamamlar. Gibbs örneklemesinin geçerliliği, algoritmanın her döngüsünün Markov zincirinin bir adımına karşılık gelmesinden kaynaklanmaktadır (Geman ve Geman, 1984). Sonsal dağılımdan üretilen değerler aşağıdaki şekilde elde edilebilirler:

8) Adım 2-7 güncelleştirilmiş değerler kullanılarak ve tüm değerler kaydedilmek suretiyle *m* defa tekrarlanır.

Örneğin eğer σ_s^2'ye ait değerlerin σ_{s1}^2, σ_{s2}^2, ..., σ_{sm}^2 olduğu varsayılırsa, bunlar σ_s^2'nin marjinal sonsal dağılımından simüle edilmiş değerleri oluştururlar. Böylece Gibbs örneklemesi algoritması *m* tane Gibbs örneği (α_l, β_l, s_l, σ_{sl}^2 ve σ_{el}^2), (l=1,...,m) üretir. Herhangi bir parametre için, *m* örnek değerlerinin kaydedilmesi marjinal sonsal yoğunluk fonksiyonundan simüle edilmiş bir örnek olarak algılanabilir.

Gibbs örneklemesinin bazı özel problemler için nasıl uygulanacağı konusunda farklı görüşler vardır. Bazı araştırıcılar Gibbs döngüsü sürecinin başlangıcında kaydedilen değerlerin bir kısmının atılmasını tavsiye etmektedirler. Bunun nedeni ilk değerlerin populasyon parametresine yaklaşmaktan çok uzak olmalarıdır. Bazı araştırıcılar ise sadece örneğin her onuncu değerinin kaydedilmesini tavsiye etmektedirler, çünkü birbirini takip eden örnekler arasında yüksek korelasyon bulunmaktadır. Bazıları da sürecin farklı başlangıç değerleri ile yeniden başlatılmasından yanadırlar. Simüle edilmiş verilerle olan deneyimler (Fırat, 1995) ve Raftery ve Lewis (1992)'nin bu konudaki teorik iddiaları, mevcut durum için sonsal dağılıma yakın başlangıç değerleri kullanılması halinde böyle olumsuzluklarla karşılaşılmayacağı inancına neden olmuştur.

3. SAYISAL ÖRNEK

Bu çalışmada kullanılan veri seti, 20438 Türk siyah-alaca sığırlarının ilk üç laktasyon süt verimlerinden ibarettir. Tablo 1 veri setinin yapısını göstermektedir. Boğa başına düşen kız sayısı 3 ile 5195 arasında değişirken sürü-yıl-mevsim başına düşen gözlem sayısı 3 ile 247 arasındadır. Analize dahil edilen kovaryetler şunlardır: (i) süt verimlerinin toplandığı iller, (ii) laktasyon sayısı ve (iii) doğrusal ve karesel olarak buzağılamadaki yaş (ay). Laktasyon süt verimi kg olarak ifade edilmiştir.

Tablo 1 Veri setinin yapısı

	Veri seti
Süt verim kayıtlarının sayısı	20438
İl sayısı	17
Sürü-yıl-mevsim kategorisi sayısı	376
Boğa sayısı	714
Boğa başına ortalama kız sayısı	28.6
Laktasyon başına kayıt sayısı	
İlk laktasyon	10219
İkinci laktasyon	6614
Üçüncü laktasyon	3605
Ortalama	5542.38
Standart sapma	1428.43

Denemede *s* boğa olduğunu ve bunların *N* kızından süt verimleri elde edildiğini varsayalım. Model (1)'de, **y** süt verimleri olsun. **H** herbir kızın ait olduğu sürü-yıl-mevsim grubunu tanımlar, **α** *g* sürü-yıl-mevsim etkilerini temsil eder, **C** kovaryetlerin merkezi değerlerini içerir, **β** *c* kovaryetlere ait regresyon katsayılar vektörüdür, s boğa etkilerini temsil eder ve **D** herbir kızın babasını belirler. **y** ve **e** vektörleri eşit uzunluğa, *N*, sahiptirler. s'nin uzunluğu boğaların toplam sayısı *s*'dir. **H**, **C** ve **D** matrislerinin boyutları sırasıyla *N*x*g*, *N*x*c* ve *N*x*s*'dir. Boğaların birbiriyle ilişkili olmadığı ve ineklerin ise sadece kendi boğaları aracılığı ile ilişkili oldukları varsayılmıştır.

4. SONUÇLAR ve TARTIŞMA

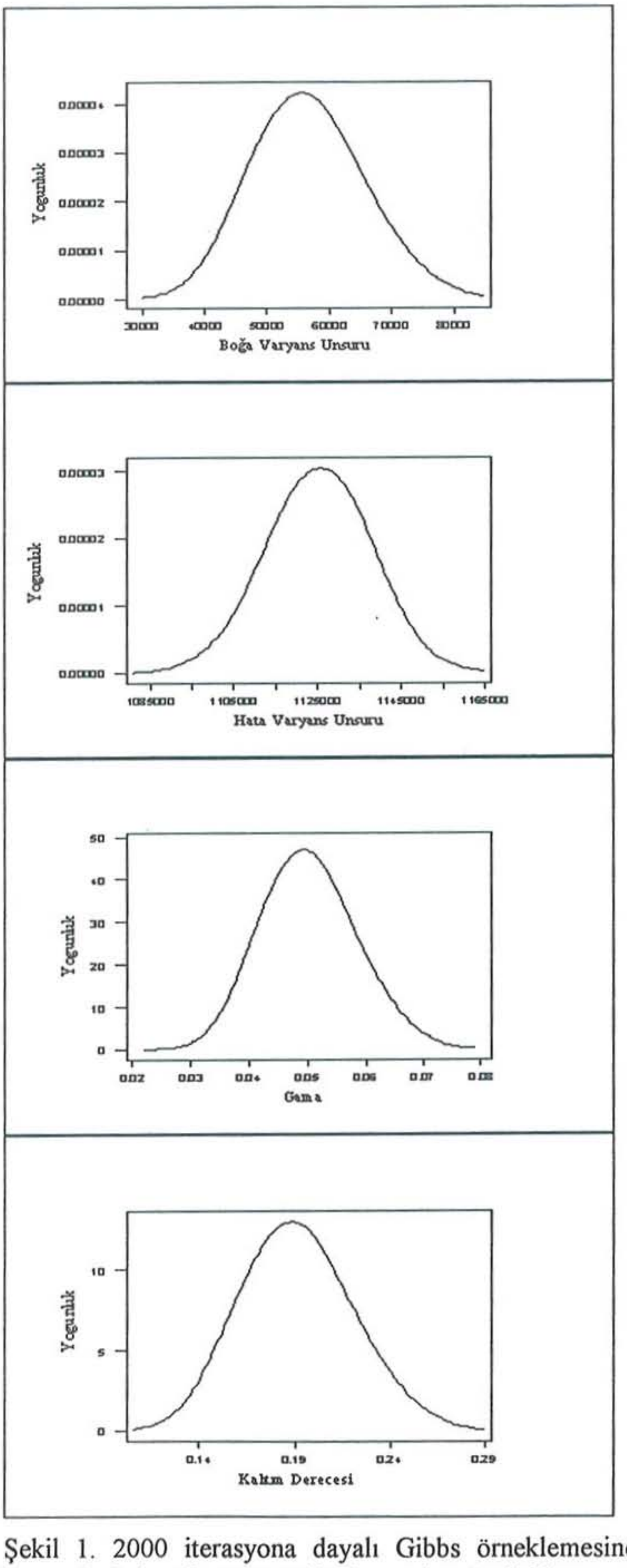
Model (1) için tek değişkenli Gibbs örneklemesi işlemi iki kez tekrarlanmıştır. İlkinde, ilgi duyulan parametrelere rastgele başlangıç değerleri verilmiş ve 2500 iterasyondan elde edilen değerler kaydedilmiştir. 2000 iterasyonun yakınsama için yeterli olduğu anlaşıldığından, ilk 500 değer atılarak 2000 iterasyonun ortalaması alınmıştır. İkincide bu ortalamalar başlangıç değerleri olarak kullanılarak 2000 iterasyondan elde edilen değerler kaydedilmiştir. Böylece model (1)'deki parametreler hakkındaki marjinal sonsal yorumlamalar ikinci koşumdan elde edilen 2000 iterasyona dayanarak yapılmıştır. Serbestlik dereceleri ν_s ve ν_e 1 olarak alınmıştır. Süt verimine ait varyans unsurları ve fonksiyonlarının ve regresyon katsayılarının REML tahminleri ve 2000 iterasyona dayalı Gibbs örneklemesinden elde edilen sonsal beklenen değerleri Tablo 2'de verilmiştir. Bu çizelge incelendiğinde REML tahminlerinin sonsal beklenen değerlere oldukça yakın oldukları görülmektedir. Normal kernellere sahip kernel yoğunluk tahmin edicisi yöntemi kullanılarak Gibbs algoritmasından elde edilen model parametreleri σ_s^2 ve σ_e^2, γ ve h^2'nin herbirinin marjinal sonsal yoğunluklarının tahminleri Şekil 1'de verilmiştir.

Tablo 2: Süt verimine ait varyans unsurları ve fonksiyonlarının ve regresyon katsayılarının REML tahminleri ve 2000 iterasyona dayalı Gibbs örneklemesinden elde edilen sonsal beklenen değerleri (standart sapmalar parantez içinde verilmiştir).

Yöntem	σ_s^2	σ_e^2	γ^*	h^2
REML	56827.276	1124962.259	0.0505	0.1923
GİBBS	56584.100	1125185.000	0.0503	0.1914
	(8378.988)	(11590.880)	(0.0075)	(0.0273)
	İl	Lakt. No.	Yaş (Lin.)	Yaş (Kuad.)
REML	-2.0137	122.1576	40.8786	-0.2637
GİBBS	-2.0187	122.5836	40.6324	-0.2607
	(0.5402)	(25.1737)	(5.6234)	(0.0684)

*Varyansların oranı, $\gamma = \sigma_s^2 / \sigma_e^2$.

Bayesci yaklaşımın avantajı, parametreler hakkındaki yorumlamaların, sadece notka tahminleri ve yaklaşık standart sapmalar yerine, bunların sonsal dağılımlarının dikkate alınarak yapılmasıdır. Bu makalede, dengesiz tek değişkenli karışık bir modelde kovaryetlere ait bilgiyi kullanarak nisbeten büyük bir veri setinin Gibbs örneklemesi ile analizinin mümkün olabileceği irdelenmiştir. Ayrıca Bayesci marjinal sonsal beklenen değerlerin Gibbs örneklemesi kullanılarak kolayca elde edilebileceği gösterilmiştir. Fakat, Gibbs örneklemesinin mahsuru hesaplamalar bakımından oldukça uzun zaman almasıdır. Örneğin, REML programı 5 dakikada analizi yaparken, Gibbs örneklemesi 2000 iterasyonu tamamlamak için 17 dakika harcamaktadır. Bu ilk bakışta büyük bir dejavantaj olarak görülsede, sonsal dağılım REML nokta tahmininden daha fazla bilgi vermektedir. Burada sunulan metedolojinin önemli bir uzantısı, tek değişkenli bir karışık modeli çok değişkenliye genelleştirmektir.



Şekil 1. 2000 iterasyona dayalı Gibbs örneklemesinden elde edilen varyans unsurları ve fonksiyonlarının marjinal sonsal yoğunlukları.

KAYNAKLAR

FIRAT, M.Z. (1995), *Bayesian Methods in the Selection of Farm Animals for Breeding*. PhD Thesis, Edinburgh University.

FIRAT, M.Z., THEOBALD, C.M., THOMPSON, R. (1997), "*Univariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling*", Acta Agric. Scand., Sect. A, Anim. Sci., 47, 213-220.

GELFAND, A.E., SMITH, A.F.M. (1990), "*Sampling-based approaches to calculating marginal densities*", Journal of Amer. Statist. Assoc., 85, 398-409.

GEMAN, S., GEMAN, D. (1984), "*Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*". IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

GIANOLA, D., FERNANDO, R.L. (1986), "*Bayesian methods in animal breeding theory*". Journal of Animal Science, 63, 217-244.

GIANOLA, D., IM, S., FERNANDO, R.L., FOULLEY, J.L. (1990), "*Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach*". In: Advances in Statistical Methods for the Genetic Improvement of Livestock (Gianola, D. and Hammond, K. eds) Springer-Verlag, Berlin, 210-238.

HASTINGS, W.K. (1970), "*Monte Carlo sampling methods using Markov chains and their applications*". Biometrika, 57, 97-109.

HOBERT, J.P. (1994) *Occurences and Consequences of Nonpositive Markov Chains in Gibbs Sampling*. PhD Thesis, Cornell University.

JENSEN, J., WANG, C.S., SORENSEN, D.A., GIANOLA, D. (1994), "*Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects using the Gibbs sampler*". Acta Agric. Scand., 44, 193-201.

RAFTERY, A.E., LEWIS, S.M. (1992), "*How many iterations in the Gibbs sampler?*" In Bayesian Statistics 4, Bernardo, J.M. Berger, J.O., David, A.P. and Smith, A.F.M. (eds). Oxford: Clarendon Press, 763-773.

WANG, C.S., RUTLEDGE, J.J., GIANOLA, D. (1993), "*Marginal inferences about variance components in a mixed linear model using Gibbs sampling*". Genet. Sel. Evol., 25, 41-62.

WANG, C.S., RUTLEDGE, J.J., GIANOLA, D. (1994), "*Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs*". Genet. Sel. Evol., 26, 91-115.

Bayesian Inferences About Variance Components In A Mixed Linear Model Via Gibbs Sampling

ABSTRACT

Markov chain Monte Carlo methods are increasingly being applied to make inferences about the marginal posterior distributions of parameters in quantitative genetic models. This paper considers the application of one such method, Gibbs sampling, to Bayesian inferences about parameters in a mixed linear model. Gibbs sampling is a method of numerical integration that allows inferences to be made about joint or marginal density, even those densities cannot be evaluated directly. It is based on generation, in sequence, of variables from all of the full conditional densities. Full conditional density is the density of a variable given all other parameters in the model. In the problem of estimation of variance components, the joint density of interest is the distribution of fixed effects, covariate effects, random effects and variance components, all given the data and the marginal densities are the distributions of fixed effects, covariate effects, random effects, or variance components, given the data. In this research, estimates of posterior distributions of genetic and phenotypic parameters for milk yield are obtained for 20438 Turkish Holstein-Friesian cows using restricted maximum likelihood (REML) and Gibbs sampling methods.

Key Words : *Gibbs sampling, Bayesian inference, parameter estimation*