IAAOJ, Scientific Science, 2013,1(1),27-33

DOĞRUSAL REGRESYONDA MARKOV ZİNCİRİ MONTE CARLO YAKINSAMA KRİTERLERİNİN KARŞILAŞTIRILMASI

Doç.Dr.M. Ali CENGİZ[1], Yrd.Doç.Dr.Talat SENEL[1], Yrd.Doç.Dr.Erol TERZI[1], Doç.Dr.Yüksel TERZI[1*]

[1] Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Samsun, Türkiye
* Sorumlu yazar e-posta: yukselt@omu.edu.tr

ÖZET

Bayesci çıkarım bir olasılıksal çıkarım metodudur. Son 20 yılda çok boyutlu integrallerin yaklaşık olarak hesaplanmasında Markov zinciri Monte Carlo (MCMC) yöntemlerindeki ve bilgisayar hesaplamalarındaki gelişim nedeni ile bu yaklaşım oldukça popular hale gelmiştir. MCMC yönteminin en önemli tarafı Markov zincirinin yakınsamasının belirlenmesidir. Yakınsama olmayan Markov zincirine dayalı istatistiksel çıkarımlar doğru olmayabilir ve yanlış yönlendirebilir. Bu çalışmada doğrusal regresyon modelinde bazı yakınsama kriterlerinin kullanımı ele alınmıştır.

Anahtar kelimeler: Markov zinciri, Monte Carlo (MCMC), yakınsama, Bayes, doğrusal regresyon

COMPARISIONS OF CONVERGENCE CRITERIAS FOR MARKOV CHAIN MONTE CARLO METHODS

ABSTRACT

Bayesian inference is a probabilistic inferential method. In the last two decades, it has become more popular than ever due to affordable computing power and recent advances in Markov chain Monte Carlo (MCMC) methods for approximating high dimensional integrals. An important aspect of any MCMC is assessing the convergence of the Markov chains. Inferences based on nonconverged. Markov chains can be both inaccurate and misleading. This study invokes the use of convergence criteria for linear regression model.

Keywords: Markov Chain, MCMC, Convergence, Bayesian, Linear regression

1. Giriş

Son yıllarda kompleks istatistiksel modellerin analizinde Markov Zinciri Monte Carlo (MCMC) metotlarının kullanımı çok yaygın hale gelmiştir. Bu metotların uygulamada çok yoğun kullanımı bir takım problemleri de beraberinde getirmiştir (Gastings, 1970; Geman and Geman, 1984). Simülasyona dayalı Bayesci çıkarım, sonsal dağılımı özetlemek yada ilgilenilen herhangi bir parametre değerini tahmin etmek için simülasyonla çekilen örneklemin kullanımını gerektirir. Simülasyonla örnek çekimleri dikkatli yapılmalıdır. Burada iki önemli konu vardır. Birincisi Markov zincirinin durağanlığa yada arzu edilen sonsal dağılıma ulaşıp ulaşmadığına karar vermektir. İkincisi ise Markov zincirini durağanlığa ulaştıracak olan iterasyon sayısının belirlenmesidir. Yakınsama tanısı, bu konuları çözmek için yardımcı olur. Birçok tanı araçları, yakınsama için gerekli fakat yeterli olmayan şartları tanımlamak için tasarlanmıştır. Markov zincirinin durağan dağılıma yakınsadığını söyleyebilen kesin testler yoktur. Dikkatli olmak gerekir. Ayrıca herhangi bir çıkarım yapmadan önce tüm parametrelerin yakınsama durumu kontrol edilmelidir. Bazı modellerde belirli parametreler çok iyi bir yakınsamaya sahip gibi görünebilir. Ancak bu durum, diğer parametrelerin yavaş yakınsaması nedeniyle yanıltıcı olabilir. Bazı parametreler kötü karışıma sahipse, iyi karışıma sahip görünen parametreler için doğru sonsal çıkarım elde edilmeyecektir. Yakınsama tanıları hakkındaki tartışmalar için Cowles and Carlin (1996) ve Brooks (1998) çalışmalarına bakılabilir.

Bu çalışmada yakınsama kriterlerinden bazıları üzerinde durulmuş ve çoklu regresyon modelleri kullanılarak bu yakınsama kriterlerinin karşılaştırılması SAS 9.3 makro kullanımıyla yapılmıştır.

2. Materyal ve Metot

Bayesci yöntemler, Markov zinciri yakınsama değerlendirmesinde yardımcı olabilecek birçok istatistiksel tanı testi içerir. Gelman ve Rubin tanısı; Brooks ve Gelman, zincirler arasındaki varyans ve her bir zincirdeki varyansları karşılaştırarak çoklu simüle edilmiş MCMC zincirlerinin analizine dayalıdır (Gelman and Rubin, 1992; Brooks and Gelman, 1997). Bu iki varyans arasındaki büyük sapma yakınsama olmadığına işaret eder. Heidelberger ve Welch testi durağanlık testi ve yarı-genişlik testi olmak üzere iki bölümden oluşmaktadır (Heidelberger and Welch, 1981). Durağanlık testi, zincirin kovaryans durağan süreçten geldiğini belirten hipotezi test ederek Markov zincirinin durağanlığını değerlendirir. Yarı-genişlik testi, Markov zinciri örnek boyutunun ortalama değerleri tam olarak test etmek için yeterli olup olmadığını kontrol eder. Raftery ve Lewis (1992, 1996) tarafından geliştirilen Raftery-Lewis testi, sonsal yüzdeliklerle ilgilenildiğinde tahmin edilen yüzdeliklerin doğruluğunu değerlendiren bir tanı testidir (Raftery and Lewis, 1992; Raftery and Lewis, 1996). Geweke testi, yakınsamadaki başarısızlıkları belirlemek için Markov zincirinin ilk ve sonraki kısımlarında bulunan değerleri karşılaştıran bir yaklaşımdır. Kass ve ark. (1998) bir Markov zincirinin karışımını incelemek için otokorelasyon ve iz grafiklerini kullanan etkin örneklem büyüklüğü (ESS) yaklaşımını geliştirmiştir (Kass, Carlin, Gelman and Neal, 1998). Diğer yakınsama yaklaşımları otokorelasyon fonksiyonu ve Markov zinciri hata (Thumb kuralı) yaklaşımlarıdır. Yukarıda bahsedilen yakınsama tanı testlerinin tanımı ve yorumlanması Tablo 1' de verilmiştir.

Tablo 1. Bayesci yöntemlerde uygun yakınsama tanı testleri

İsim	Tanım	Testin yorumlanması
Geweke	Markov zincirinin önceki ve sonraki kısımlarının ortalamalarını karşılaştırarak ortalama tahminlerinin yakınsamaya sahip olup olmadığını test eder.	z test istatistiğine dayalı iki yönlü testtir. Büyük mutlak değerli z değeri yakınsamanın sağlanmadığını göstermektedir.
Heidelberger-Welch (durağanlık testi)	Markov zincirinin bir durağan süreç olup olmadığını test eder. Başarısızlık, daha uzun bir Markov zincirine ihtiyaç olduğunu gösterir.	Cramer-von Mises istatistiğine dayalı tek yönlü testtir. Küçük p değerleri yakınsamanın sağlanmadığını göstermektedir.
Heidelberger-Welch (yarı-genişlik testi)	Örneklem büyüklüğünün ortalama tahmini için gerekli hassasiyeti karşılamada yeterli olup olmadığını gösterir. Başarısızlık, daha uzun bir Markov zincirine ihtiyaç olduğunu gösterebilir.	Eğer yarı-genişlik istatistiği, önceden belirlenmiş doğruluk ölçüsünden daha büyükse yakınsamanın sağlanmadığını göstermektedir.
Raftery-Lewis	Yüzdeliklerin istenen doğruluğuna ulaşmak için ihtiyaç duyulan örnek sayısını rapor ederek tahmini yüzdeliklerin doğruluğunu değerlendirir. Başarısızlık, daha uzun bir Markov zincirine ihtiyaç olduğunu gösterebilir.	Eğer gerekli olan toplam örneklem, Markov zinciri örneğinden daha az ise bu yakınsamanın sağlanmadığını gösterir.
Otokorelasyon	Markov zinciri örnekleri arasındaki bağımlılığı ölçer.	Uzun gecikmeler arasındaki yüksek korelasyon, zayıf karışımı dolayısıyla yakınsamanın sağlanmadığını gösterir.
Markov Zinciri hata yaklaşımı	Markov zinciri (MC) standart hatası hesaplanır.	MC hata standart sapmanın %5'inden küçükse yakınsama vardır.
Etkili örneklem büyüklüğü	Otokorelasyonla ilgilidir. Markov zincirinin karışımını ölçer.	Etkili bir örnek büyüklüğü ile simüle edilen örnek büyüklüğü arasındaki büyük fark, zayıf karışımı dolayısıyla yakınsamanın sağlanmadığını gösterir.

3. Uygulama

Tükenmişlik kavramı birçok bilim adamı tarafından incelenmiş olmasına rağmen bu konuda en çok tanınan ve bu alandaki çalışmalara öncülük eden Maslach'tır. Onun çalışmaları bu alanda çalışanlar için en önemli kaynaklar olmuştur (Izgar, 2001). Maslach ve arkadaşları tarafından günümüzde de en çok kabul gören tükenmişlik, "yaygın olarak insanlarla yüz yüze çalışılan mesleklerdeki bireylerin; duygusal olarak kendilerini tükenmiş hissetmeleri, iş gereği karşılaştıkları insanlara karşı duyarsızlaşmaları ve kişisel başarı ya da yeterlilik duygularında azalma olarak görülen bir sendrom" şeklinde tanımlanmıştır (Maslach and Zimbarda, 1982).

Bu çalışmada araştırma görevlilerinin yaş, zeka seviyesi, eğitim durumu ve gelir durumlarının tükenmişlik düzeyleri üzerine etkisini incelemek üzere kurulan çoklu lineer regresyon modeli Bayesci bir yaklaşımla incelenmiştir. Çoklu lineer regresyon modeli (1)'deki gibi olsun.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{1}$$

(1) modelindeki değişkenler şu şekilde tanımlanmıştır.

Y: Tükenmişlik skoru
X_1: Yaş
X_2: Zeka seviyesi
X_3: Eğitim durumu
X_4: Gelir
n=150

Bu amaca yönelik olarak kurduğumuz klasik çoklu regresyon modeli sonuçları Tablo 2' de sunulmuştur.

Tablo 2. Klasik çoklu regresyon modeli

Değişken	β	Std. Hata	t	p
Sabit	162.488	15.618	10.404	.000
Yaş	-.078	.092	-.848	.398
Zeka	-1.056	.181	-5.837	.000
Eğitim	-.894	.301	-2.968	.004
Gelir	-.004	.001	-3.248	.002

Gözlem hataları ε sıfır ortalamalı ve σ^2 sabit varyanslı normal dağılıma sahip olduğu kabul edilir.

$$Y \sim Normal(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma^2)$$

Y için olabilirlik fonksiyonu (2)'de verilmiştir.

$$P(Y \mid \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma^2) = \phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \sigma^2) \tag{2}$$

Burada $P(.\mid .)$ bir koşullu olasılık yoğunluk fonksiyonu ve ϕ bir normal yoğunluk fonksiyonudur. Dolayısıyla olabilirlik fonksiyonunda $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ ve σ^2 gibi 6 parametre vardır. Bu çalışmaya özgün olarak her bir parametre için önsel dağılımlar (3)'de gibi tanımlanmıştır.

$$\begin{gathered} f(\beta_0) = f(\beta_1) = f(\beta_2) = f(\beta_3) = f(\beta_4) = \phi(0, Var = 10000) \\ f(\sigma^2) = \phi(0, Var = 1000) \end{gathered} \tag{3}$$

Bayes teoremi kullanılarak $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma^2$ parametreleri için sonsal dağılım ise (4)'deki gibi olur.

$$P(\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\sigma^2 | D) \propto f(\beta_0)f(\beta_1)f(\beta_2)f(\beta_3)f(\beta_4)f(\sigma^2)\ P(Y | \beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\sigma^2) \tag{4}$$

Metropolis örnekleme algoritması kullanılarak Tablo 3' de verilen sonsal özetler elde edilmiştir.

Tablo 3. Sonsal özetler

Parametre	Ortalama	Standart Sapma	%95 İnanç Aralığı (HPD Interval)	
β_0	132.6	149.250	103.7	160.4
β_1	-0.0672	0.0854	-0.2352	0.0995
β_2	-0.7299	0.1726	-10.548	-0.3955
β_3	-0.8669	0.2804	-14.217	-0.3243
β_4	-0.00391	0.00107	-0.00601	-0.00182
σ^2	110.6	119.329	879.166	134.2

Gerek Tablo 2, gerekse Tablo 3' e bakılarak hem klasik regresyon çözümüyle hem de Bayesci çözümde Zeka, Eğitim ve Gelir'in tükenmişlik skoru üzerinde anlamlı bir etkiye sahip olduğu söylenebilir. Ancak Tablo 3' de verilen sonsal özetlerin geçerli olması için yakınsama kriterlerini sağlaması gerekir. Bu amaca yönelik olarak, materyal metot bölümünde verilen yakınsama kriterleri bu veri seti için uygulanmıştır. İlk olarak Tablo 4' de Monte Carlo standart hataları verilmiştir.

Tablo 4. Monte Carlo standart hatalar

Parametre	MCSE	Standart Sapma	MCSE/SD
β_0	1.1555	14.9250	0.0774
β_1	0.000317	0.0854	0.00372
β_2	0.0126	0.1726	0.0732
β_3	0.00103	0.2804	0.00369
β_4	3.979E-06	0.00107	0.00374
σ^2	0.2052	11.9329	0.0172

β_0, β_2 hariç diğer parametreler için MC hatanın standart sapmasının %5' inden küçük olduğu görülmektedir. Dolayısıyla yakınsama sağlanmıştır. İkinci olarak sonsal otokorelasyonlar Tablo 5' de sunulmuştur.

Tablo 5. Sonsal otokorelasyonlar

Parametre	Lag 1	Lag 5	Lag 10	Lag 50
β_0	0.9966	0.9832	0.9668	0.8489
β_1	0.0528	0.0086	0.0037	0.0055
β_2	0.8981	0.8813	0.8670	0.7591
β_3	0.0434	0.0002	0.0024	-0.0022
β_4	0.0461	0.0044	0.0047	0.0091
σ^2	0.0728	0.0685	0.0589	0.0533

Sonsal otokorelasyonlara bakıldığında β_0 ve β_2 hariç diğer otokorelasyonların hızla düştüğü görülmektedir. Tablo 6'da Geweke test sonuçları verilmiştir.

Tablo 6. Geweke testi

Parametre	z	P
β_0	-0.5058	0.6130
β_1	0.6613	0.5085
β_2	0.5099	0.6101
β_3	-0.0059	0.9953
β_4	-0.5205	0.6027
σ^2	0.8681	0.3854

Mutlak değerce büyük z değerleri dolayısıyla $P > \alpha = 0.05$ değerleri bütün parametreler için yakınsamanın sağlandığını göstermektedir. Tablo 7'de Raftery-Lewis testi, Tablo 8'de Heidelberger-Welch test değerleri verilmiştir.

Tablo 7. Raftery-Lewis testi

Parametre	Yakınsama Periyodu	Toplam	Minimum	Bağlı Faktör
β_0	496	505307	3746	134.8924
β_1	2	3812	3746	1.0176
β_2	447	471529	3746	125.8753
β_3	2	3893	3746	1.0392
β_4	2	3913	3746	1.0446
σ^2	2	3783	3746	1.0099

Tablo 8. Heidelberger-Welch testi

Parametre	Durağanlık testi			Yarı genişlik testi			
	Cramer-von-Mises istatistiği	P	Test sonucu	Yarı genişlik	Ortalama	Relatif yarı genişlik	Test sonucu
β_0	0.0574	0.8301	Başarılı	2.8173	132.6	0.0212	Başarılı
β_1	0.3227	0.1168	Başarılı	0.000993	-0.0672	-0.0148	Başarılı
β_2	0.0565	0.8361	Başarılı	0.0308	-0.7299	-0.0421	Başarılı
β_3	0.0311	0.9725	Başarılı	0.00305	-0.8669	-0.00352	Başarılı
β_4	0.0511	0.8693	Başarılı	0.000018	-0.00391	-0.00461	Başarılı
σ^2	0.0671	0.7696	Başarılı	0.4949	110.6	0.00447	Başarılı

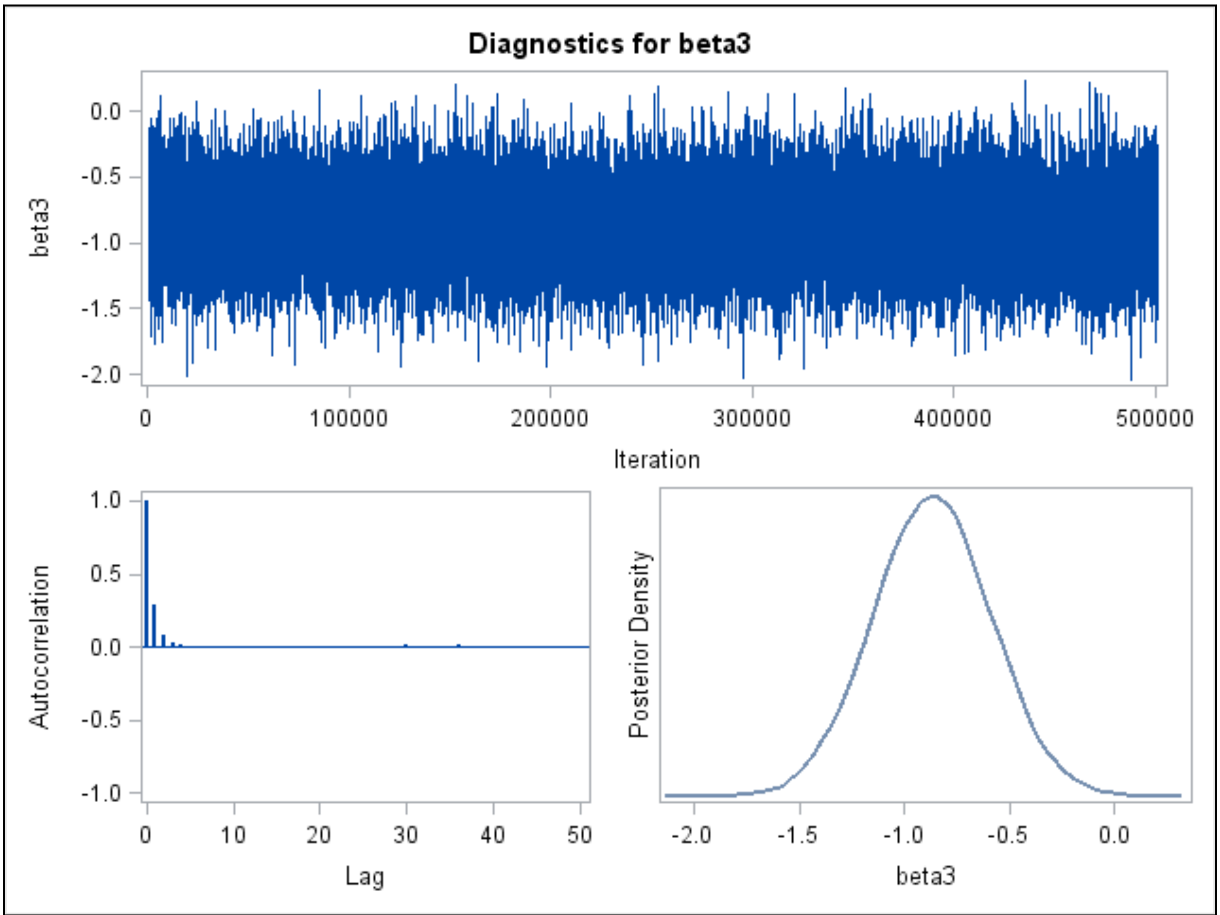
Tablo 7' de yakınsama için β_0 ve β_2'nin dışında problem olmadığı gözükmektedir. Tablo 8' de bütün parametrelerin yakınsadığı açıktır. Tablo 9' da etkin örnek büyüklükleri verilmiştir.

Tablo 9. Simüle edilmiş etkin örnek büyüklüğü

Parametre	ESS	Otokorelasyon zamanı	Etkinlik
β_0	166.8	479.5	0.0021
β_1	72352.6	1.1057	0.9044
β_2	186.8	428.2	0.0023
β_3	73613.7	1.0868	0.9202
β_4	71648.9	1.1166	0.8956
σ^2	3383.4	23.6450	0.0423

Tablo 9' da β_0 ve β_2 dışında bütün parametrelerin yakınsama problemi olmadığı gözükmektedir. β_0 ve β_2 için bazı kriterler dışında çoğu kriter bütün parametreler için yakınsamanın sağlandığını göstermektedir. Dolayısıyla Tablo 3'de verilen sonsal özetler güvenilirdir.

Görsel olarak da yakınsama hakkında bir şeyler söyleyebiliriz. Örnek olarak β_3 için Şekil 1' de iz grafiği, otokorelasyon grafiği ve sonsal yoğunluk fonksiyonu verilmiştir.



Şekil 1. β_3 için diagnostik grafikler

Şekil 1'de her üç grafikte yakınsama kriterinin sağlandığını açık bir şekilde göstermektedir. Bütün yakınsama kriter değerleri 500.000 örneklem büyüklüğü için ifade edilmiştir. Yakınsama kriterlerini karşılaştırmak için hangi örneklem büyüklüğünde tüm parametrelerin yakınsadığını gösteren sonuçlar Tablo 10' da verilmiştir.

Tablo 10. Yakınsama kriterlerinin karşılaştırması

Kriter	Örneklem büyüklüğü
MC Hata	19.300
Geweke testi	20.500
Heildelberger-Welch (Durağanlık)	37.100
Heildelberger-Welch (Yarı genişlik)	37.100
Raftery Lewis	63.000
Sonsal otokorelasyon	65.000
Etkin örnek büyüklüğü	73.614

En küçük örneklem büyüklüğü MC hata kriteriyle sağlanırken, en büyük örneklem büyüklüğünü etkin örnek büyüklüğü vermektedir. Bu kadar büyük örneklemler gerekirken, Bayesci yaklaşımın kullanılmasına gerek olup-olmadığına Tablo 11'de verilen klasik ve Bayesci yaklaşımların karşılaştırması cevap verebilir. Bu amaca yönelik olarak hem klasik hem de Bayesci yaklaşım için Akaike Bilgi Kriteri (AIC) ve Bayesci Bilgi Kriteri (BIC) değerleri hesaplanmış ve sonuçlar Tablo 11' de sunulmuştur.

Tablo 11. Model karşılaştırma kriterleri

Yaklaşım	AIC	BIC
Klasik çoklu regresyon	906.144	922.717
Bayesci çoklu regresyon	693.121	717.421

4. Sonuç ve Tartışma

Bayesci yaklaşım model belirsizliğini göz önüne almayan klasik yaklaşımlara göre bir takım zorluklar içermektedir. Bunların başında sonsal çıkarımları elde etmedeki hesaplama zorlukları gelmektedir. Son yıllarda MCMC yöntemlerinin kullanımı çok yaygın hale gelmiştir. Ancak bu yöntemler kullanılırken kullanıcı oldukça dikkatli olmalıdır. Çünkü tüm tahmin edilecek parametrelere için yakınsamanın gerçekleşmediği durumlarda elde edilen istatistiksel sonsal çıkarımlar yanıltıcı olabilir. Dolayısıyla tüm parametreler için yakınsama incelenmelidir.

Bu çalışmada ilk olarak klasik doğrusal regresyon modeli ile Bayesci doğrusal doğrusal regresyon model sonuçları karşılaştırılmış, parametre tahminleri anlamında yaklaşık aynı sonuçlar elde edilmiştir. Ancak model uyumunun karşılaştırılması durumunda Bayesci yaklaşımının çok daha iyi sonuçlar verdiği Tablo 11' de açıktır. Bu sonuç Bayesci yaklaşımın kullanımındaki zorlukların göz önüne alınmasına değer olduğunu gösterir. Bu zorlukların başında MCMC yaklaşımı kullanıldığında kesin bir yakınsama kriterinin olmamasıdır. İncelenen yakınsama kriterlerinin belli bir örneklem büyüklüğünden sonra sağlandığı Tablo 4-9' larda açıktır. Ancak bu kriterlerden hangisinin daha küçük örneklemelerde yakınsama sağladığı Tablo 10' da görülmektedir. En azından bu çalışma için en küçük örneklem büyüklüğünü *MC hata* kriterinin, en büyük örneklem büyüklüğünü ise *etkin örnek büyüklüğünün* verdiği söylenebilir.

Kaynaklar

Brooks S.P., (1998). Roberts, G.O., Assessing Convergence of Markov Chain Monte Carlo Algorithms, *Statistics and Computing*, 8, 319–335.

Brooks S.P., and Gelman A., (1997). General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 7, 434–455.

Cowles M.K., and Carlin B.P., (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of the American Statistical Association*, 883–904.

Geman S., and Geman D., (1984). Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gelman A., and Rubin D.B., (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7, 457–472.

Hastings W.K., (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 57, 97–109.

Heidelberger P., and Welch P.D., (1981). A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations, *Communication of the ACM*, 24, 233–245.

Izgar H., (2001). Okul Yöneticilerinde Tükenmişlik. Ankara: Nobel Yayın Dağıtım.

Kass R.E., Carlin B.P., Gelman A., and Neal R., (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion, *The American Statistician*, 52, 93–100.

Maslach C., and Zimbardo P.G., (1982). Burnout The Cost of Caring. New Jersey: Prentice-Hall Inc, Englewood Cliffs.

Raftery A.E., and Lewis S.M., (1992). One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo, *Statistical Science*, 7, 493–497.

Raftery A.E., and Lewis S.M., (1996). The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms, in W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, eds., *Markov Chain Monte Carlo in Practice*, London, UK: Chapman & Hall.