Markov Zinciri Monte Carlo Yaklaşımı
ve
Aktüeryal Uygulamaları

ŞİRZAT ÇETİNKAYA
Aktüer
YapıKredi Sigorta |Sistem Araştırma Geliştirme Bölümü

AKTÜERLER DERNEĞİ
21.10.2008 - İSTANBUL

Sunum Planı

1. Giriş
2. Bayesci Metodun Temelleri
3. Markov Zinciri Monte Carlo (MZMC) Yaklaşımı
   3.1 Monte Carlo İntegrallemesi ve Markov Zinciri
   3.2 Gibbs Örneklemesi
4. MZMC Yönteminin Aktüeryal Uygulamaları
   4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli
   4.2 Hasar Büyüklüğü Dağılımının Modellenmesi
   4.3 Gruplandırılmış Hasar Büyüklüğü Dağılımının Modellenmesi

1. Giriş

Markov Zinciri Monte Carlo (MZMC) yaklaşımı markov zinciri kullanarak Monte Carlo integrasyonunun yapıldığı bir yöntemdir. ***Çoğu zaman uygulamacılar ya da bayes yaklaşımı ile problemleri çözmeye çalışan kişiler yüksek boyutlu olasılık dağılımlarının model parameterelerini bulmak ya da tahmin yapmak için integrallemek isterler.*** Monte Carlo integrasyonunda dağılımdan örneklemler çekilir ve daha sonra bu örneklemlerin ortalamaları beklenen değerin yaklaşık değerini bulmak için kullanılır. MZMC yönteminde bu çekilişler markov zinciri sistematiğine uygun olarak çekilir. Zincirleri oluşturmanın değişik yolları vardır Gibbs (Geman and Geman, 1984), Metropolis et al. (1953) ve Hastings (1970).

2. Bayesci Metodun Temelleri

Önsel dağılım, $\pi(\theta)$, belirsiz θ parametrisi -bu parametre yeni veriye ait olasılık dağılımı ile birleştirilerek sonsal dağılımı elde etmek için kullanılır - hakkında bilgileri gösterir.

Model dağılımı, belirlenen parametrelere göre toplanan verilere ilişkin olasılık dağılımıdır. Olasılık yoğunluk fonksiyonu, $f_{X|\Theta}(x|\theta)$ ile gösterilip, bazen olabilirlik fonksiyonu olarakta isimlendirilebilinir.

$x=(x_1,x_2,\ldots,x_n)^T$, birbirlerinden bağımsız ve aynı dağılımlı iseler;

$$f_{X|\Theta}(x|\theta)=f_{X|\Theta}(x_1|\theta)\ldots f_{X|\Theta}(x_n|\theta)$$

2. Bayesci Metodun Temelleri

Birleşik dağılım,

$$f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta)\pi(\theta)$$

olasılık yoğunluk fonksiyonuna sahiptir.

x'in marjinal dağılımı,

$$f_X(x) = \int f_{X|\theta}(x|\theta)\pi(\theta)d\theta$$

olasılık yoğunluk fonksiyonuna sahiptir.

Sonsal dağılım gözlenen değerler verildiğinde gözlenen değerlerin parametreleri üzerine koşullu olasılık dağılımıdır ve

$$\pi_{\Theta|X}(\theta|x)$$

ile gösterilir.

Öngörü dağılımı x verildiğinde yeni gözlem y'nin koşullu olasılık dağılımıdır ve

$$f_{Y|X}(y|x)$$

ile gösterilir.

2. Bayesci Metodun Temelleri

$$\pi_{\Theta|X}(\theta \mid x) = \frac{f_{X|\Theta}(x \mid \theta)\pi(\theta)}{\int f_{X|\Theta}(x \mid \theta)\pi(\theta)d\theta}$$

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

2. Bayesci Metodun Temelleri

D gözlenen veriler, teta model parametreleridir.

$$P(D,\theta) = P(D \mid \theta)P(\theta)$$

D gözlemi elde edildiğinde, Bayes teoremi *D* koşulu altında θ'nın dağılımını bulmada kullanılabilir.

$$P(\theta \mid D) = \frac{P(\theta)P(D \mid \theta)}{\int P(\theta)P(D \mid \theta)d\theta}$$

Sonsal dağılıma ilişkin momentler, yüzdelikler vs. θ'nın fonksiyonunun beklenen değerleri bulunabilir.

$f(\theta)$ fonksiyonunun sonsal beklenen değeri,

$$E[f(\theta) \mid D] = \frac{\int f(\theta)P(\theta)P(D \mid \theta)d\theta}{\int P(\theta)P(D \mid \theta)d\theta}$$

3. Markov Zinciri Monte Carlo Yaklaşımı
3.1 Monte Carlo İntegrallemesi ve Markov Zinciri

Monte Carlo integrallemesinde, $E[\,f(X)\,]$ beklenen değerine $\pi(.)$ dağılımından çekilen rasgele $\{X_t, t=1,\dots,n\}$ değerleri çekilerek,

$$E[\,f(X)\,] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t)$$

ile yaklaşılır.

$$\boxed{\overline{f} \approx \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t)}$$

Markov Zinciri;

$$\begin{aligned}\Pr(X_{n+1} = x \mid X_n = x_n, \dots, X_1 = x_1) \\ = \Pr(X_{n+1} = x \mid X_n = x_n)\end{aligned}$$

$$\begin{aligned}\Pr(X_{n+1} = x \mid X_n = x_n, \dots, X_1 = x_1) \\ = \Pr(X_{n+1} = x \mid X_n = x_n, \dots, X_{n-m} = x_{n-m})\end{aligned}$$

3. Markov Zinciri Monte Carlo Yaklaşımı
3.2 Gibbs Örneklemesi

Algoritmanın adımları aşağıdaki gibidir:

$\theta^{(0)}$'ın verilmesi durumunda,

1. Çekiliş: $\theta_1^{(1)} \sim q(\theta_1 \mid \theta_2^{(0)}, \ldots, \theta_r^{(0)})$
2. Çekiliş: $\theta_2^{(1)} \sim q(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)})$

.

.

.

r. Çekiliş: $\theta_r^{(1)} \sim q(\theta_r \mid \theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{r-1}^{(1)})$

Ardından $q(X \mid \theta)$'dan X değeri çekilir.

(Geman ve Geman (1984))

4. Aktüeryal Uygulamaları

The BUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling)

http://www.mrc-bsu.cam.ac.uk/bugs/

4. Aktüeryal Uygulamaları
4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli

Tablo 1: Poliçe ve hasar sayıları

YIL	Grup 1		Grup 2		Grup 3	
	Poliçe Sayısı	**Hasar Sayısı**	**Poliçe Sayısı**	**Hasar Sayısı**	**Poliçe Sayısı**	**Hasar Sayısı**
1	285	9	265	6	?	?
2	325	7	270	5	140	8
3	270	6	240	3	115	4
4	345	13	265	8	110	5
5	?	?	295	?	125	?

(Scollnik (2000))

4. Aktüeryal Uygulamaları
4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli

- X_{ij} , hasar sayısını, P_{ij} ise kesilen poliçe sayısını göstersin, $i = 1,2,3,\ \ j = 1,...,5$.
- $X_{ij} \sim Pois(\lambda_{ij})$ ve $\lambda_{ij} = P_{ij}\theta_i$ olsun.
- $\theta_i \sim Gamma(\alpha,\beta)$
- $\alpha \sim Gamma(5,5)$
- $\beta \sim Gamma(25,1)$
- $P_{ij} \sim Gamma(a_i,b_i)$
- $a_i \sim Uniform(0,100)$
- $b_i \sim Uniform(0,100)$

4. Aktüeryal Uygulamaları
4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli

X, α, θ, β değişkenlerinin bileşik dağılımı,

$$\prod_{j=1}^{4} f(X_{1j} \mid P_{1j}\theta_1) \prod_{j=1}^{4} f(X_{2j} \mid P_{2j}\theta_2) \prod_{j=2}^{4} f(X_{3j} \mid P_{3j}\theta_3) \prod_{j=1}^{3} f(\theta_j \mid \alpha, \beta) f(\alpha) f(\beta)$$

Veri gözlendikten sonra sonsal dağılım, $f(\theta, \alpha, \beta \mid X)$, hesaplanabilir. Bayes teoremine göre $f(\theta, \alpha, \beta \mid X) \propto f(\theta, \alpha, \beta, X)$ olarak yazılabilir.

Yukarıda verilenlere göre örnek sonsal dağılım,

$$f(\theta_1 \mid \alpha, \beta, X, \theta_2, \theta_3) \propto \prod_{j=1}^{4} f(X_{1j} \mid P_{1j}\theta_1) f(\theta_1 \mid \alpha, \beta)$$

$$\sim Gamma(\alpha + \sum_{j=1}^{4} X_{1j}, \beta + \sum_{j=1}^{4} P_{1j})$$

4. Aktüeryal Uygulamaları
4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli

$$f(\alpha \mid \beta, X, \theta) \propto \prod_{i=1}^{3} f(\alpha) f(\theta_i \mid \alpha, \beta)$$

$$f(\beta \mid \alpha, X, \theta) \propto \prod_{i=1}^{3} f(\beta) f(\theta_i \mid \alpha, \beta)$$

$$f(\theta_i \mid X) = \int_{\alpha}\int_{\beta} f(\theta_i \mid X, \alpha, \beta) f(\alpha, \beta \mid X) d\beta d\alpha$$

$$\approx \frac{1}{n-m} \sum_{t=m+1}^{n} f(\theta_i \mid X, \alpha^{(t)}, \beta^{(t)}), \; i = 1,2,3$$

$$f(X_{i5} \mid X) = \int_{\theta_i} f(X_{i5} \mid P_{i5}\theta_i) f(\theta_i \mid X) d\theta_i$$

$$\approx \frac{1}{n-m} \sum_{t=m+1}^{n} f(X_{i5} \mid P_{i5}\theta_i^{(t)}), \; i = 1,2,3$$

4. Aktüeryal Uygulamaları
4.1 Hasar Sıklığı İçin Hiyerarşik Poisson Modeli

	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
x[1,5]	8.921	3.638	0.03265	3.0	9.0	17.0	1	10000
x[2,5]	6.378	2.856	0.02899	2.0	6.0	13.0	1	10000
x[3,1]	5.676	2.878	0.02843	1.0	5.0	12.0	1	10000
x[3,5]	5.844	2.774	0.02898	1.0	6.0	12.0	1	10000

4. Aktüeryal Uygulamaları
4.2 Hasar Büyüklüğü Dağılımının Modellenmesi

$$X \sim Gamma(k, \theta)$$

$$f(x|\theta) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-\theta x}, \qquad 0 < x < \infty$$

$$\theta \sim Gamma(\alpha, \lambda)$$

$$f(\theta) = \frac{\lambda^k}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}, \qquad 0 < \theta < \infty$$

$$X \sim Pareto(\alpha, \lambda, k)$$

$$f(x) = \frac{\Gamma(\alpha+k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda+x)^{\alpha+k}}, \qquad 0 < x < \infty$$

$$f(\theta|x) \sim Gamma(\alpha+k, \lambda+x), \qquad 0 < \theta < \infty$$

4. Aktüeryal Uygulamaları
4.2 Hasar Büyüklüğü Dağılımının Modellenmesi

1. Rasgele, $X^{(0)}, \theta^{(0)}$ başlangıç değerleri seçilir,
2. i=0; $X^{(i+1)} \sim f\left(x \middle| \theta^{(i)}\right) \sim Gamma(k, \theta^{(i)})$
3. $\theta^{(i+1)} \sim f\left(\theta \middle| X^{(i+1)}\right) \sim Gamma(\alpha + k, \lambda + X^{(i+1)})$

$X^{(0)} = 20, \theta^{(0)} = 10, \alpha = 5, \lambda = 1000, k = 2,$

500 iterasyon sonunda; $X^{(0)}, \theta^{(0)}; X^{(1)}, \theta^{(1)}; \dots; X^{(500)}, \theta^{(500)}$

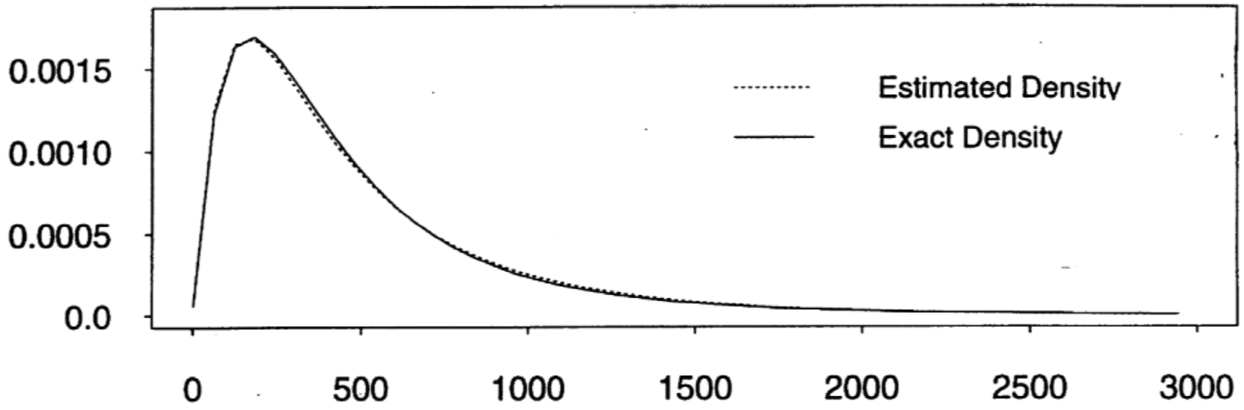
4. Aktüeryal Uygulamaları
4.2 Hasar Büyüklüğü Dağılımının Modellenmesi

$$f(x) = \int f(x|\theta)f(\theta)d\theta$$

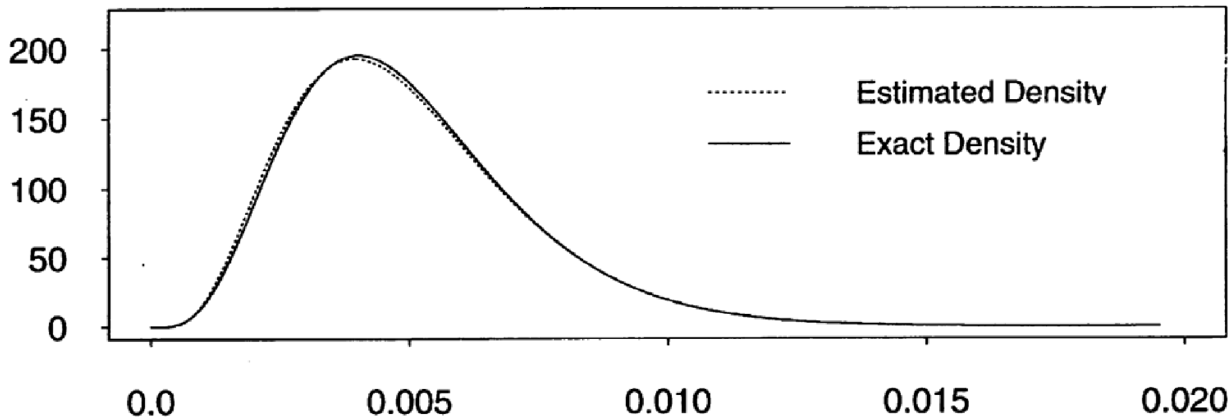
$$\hat{f}(x) = \frac{1}{500}\sum_{i=1}^{500} f(x|\theta^{(i)})$$

$$\hat{f}(\theta) = \frac{1}{500}\sum_{i=1}^{500} f(\theta|X^{(i)})$$

4. Aktüeryal Uygulamaları
4.2 Hasar Büyüklüğü Dağılımının Modellenmesi



Marjinal dağılım-X



Marjinal dağılım-Teta

4. Aktüeryal Uygulamaları
4.3 Gruplandırılmış Hasar Büyüklüğü Dağılımının Modellenmesi

$$X_i, i = 1, 2, \dots, 25$$

- X, hasar büyüklüğü, gruplandırılmış,
- Gruplar: (0,1000], (1000,2000], (2000,3000], (3000,∞],
- Gruplara ilişkin sıklıklar: 12, 8, 3 ve 2'dir.

$$Pareto(\theta, \lambda) \sim f(x|\theta) = \frac{\theta \lambda^{\theta}}{(\lambda + x)^{\theta+1}}, \quad 0 < x < \infty$$, Lambda=5000

$$f(\theta) \sim Gamma(\alpha, \beta)$$

$$L(\theta|X) = \frac{\theta^{25} \lambda^{25\theta}}{\prod_{i=1}^{25} (\lambda + x_i)^{\theta+1}}$$

4. Aktüeryal Uygulamaları
4.3 Gruplandırılmış Hasar Büyüklüğü Dağılımının Modellenmesi

$$f(\theta|X) \propto \theta^{24+\alpha} \exp\left(-\theta\left(\beta - 25\ln\lambda + \sum_{i=1}^{25} \ln\left[\lambda + x_i\right]\right)\right)$$

$$\sim Gamma\left(25+\alpha, \beta - 25\ln\lambda + \sum_{i=1}^{25} \ln\left[\lambda + x_i\right]\right)$$

$$\alpha = \beta = 0{,}001$$

$$x_i \sim Truncated\ Pareto(\theta, \lambda), (0{,}1000], i = 1, 2, \ldots, 12$$

$$x_i \sim Truncated\ Pareto(\theta, \lambda), (1000{,}2000], i = 13, 14, \ldots, 20$$

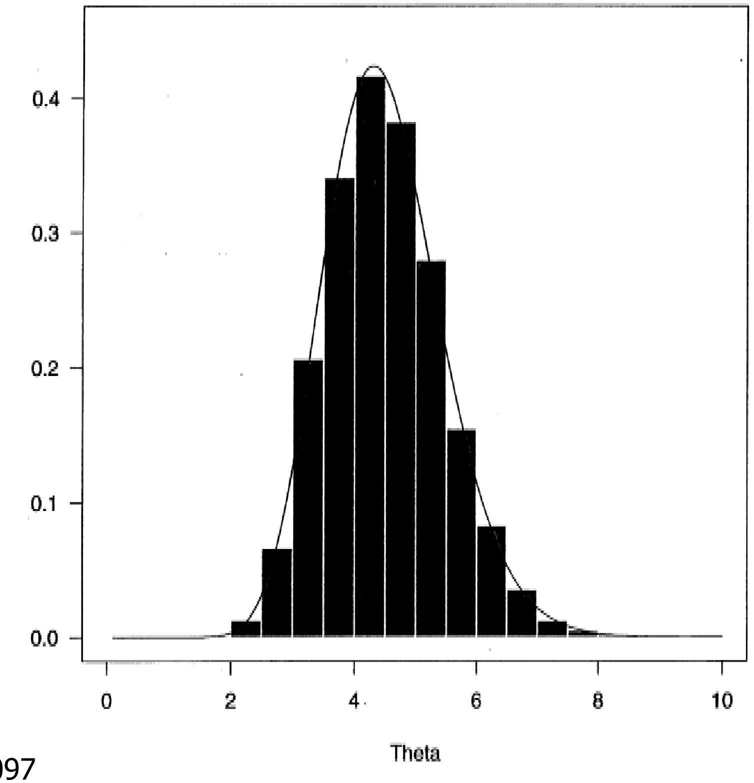
$$x_i \sim Truncated\ Pareto(\theta, \lambda), (2000{,}3000], i = 21{,}22, 23$$

$$x_i \sim Truncated\ Pareto(\theta, \lambda), (3000, \infty], i = 24, 25$$

$$Truncated\ Pareto\ \text{Dağılımı} \sim g(x|\theta) = \frac{f(x|\theta)}{\Pr(l < x \le u|\theta)}, \qquad l < x \le u$$

4. Aktüeryal Uygulamaları
4.3 Gruplandırılmış Hasar Büyüklüğü Dağılımının Modellenmesi

10. iterasyondaki θ değeri alınmak üzere 1000 farklı θ değeri



Örneklem Ortalaması:4,5097

Bazı Kaynaklar

- Gilks, W. R., Richardson, S., Spiegelhalter, D., 1996, Markov Chain Monte Carlo In Practice, Chapman&Hall.
- Klugman, S., Panjer, H. H., Willmot G.E., 2004, Loss Models: From Data to Decisions, John Wiley & Sons, New York.
- Scollnik D.P.M., 2000, Actuarial Modeling with MCMC and BUGS: Additional Worked Examples, Actuarial Research Clearing House, 2000.2, 433-585.
- Scollnik, D.P.M., 2001, Actuarial Modeling with MCMC and BUGS, North American Actuarial Journal, 5 (2), 96-124.