EKSAKTA

Jurnal Ilmu-Ilmu MIPA

EKSAKTA / VOL. 11 NO. 1 / FEBRUARI 2010 / Hal 1-47 / ISSN 1411-1047

Segmentasi Bayesian Hirarki Untuk Model Ma Konstan Sepotong Demi Sepotong Berbasis Algoritma *Reversible Jump MCMC*

Suparman

Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UTY
Email : suparman@netcourrier.com

ABSTRACT

This paper addresses the problem of the signal segmentation within a hierarchical Bayesian framework by using reversible jump MCMC sampling. The signal is modelled by piecewise constant MA processes where the numbers of segments, the position of abrupt, the order and the coefficients of the MA processes for each segment are unknown.

The reversible jump MCMC algorithm is then used to generate samples distributed according to the joint posterior distribution of the unknown parameters. These samples allow to compute some interesting features of the a posterior distribution. Main advantage of the algorithm reversible jump MCMC algorithm is produce the joint estimators for the parameter and hyper parameter in hierarchical Bayesian

The performance of the this methodology is illustrated via several simulation results.

Keywords : *Hierarchical Bayesian model, Reversible Jump MCMC methods, Signal Segmentation,* piecewise constant Moving-Average (MA) processes.

ABSTRAK

Makalah ini membahas masalah segmentasi sinyal dalam kerangka Bayesian hirarki dengan menggunakan sampling Reversible Jump MCMC. Sinyal dimodelkan oleh MA konstan sepotong demi sepotong dimana jumlah segmen, posisi, orde dan koefisien MA proses untuk setiap segmen tidak diketahui.

Algoritma Reversible Jump MCMC kemudian digunakan untuk menghasilkan sampel yang didistribusikan sesuai dengan distribusi posterior gabungan dari parameter yang tidak diketahui. Sampel ini memungkinkan untuk menghitung beberapa fitur menarik dari distribusi posterior. Keuntungan utama dari algoritma Reversibel Jump MCMC adalah menghasilkan penduga bersama untuk parameter dan hyperparameter dalam Bayesian hirarki

Kinerja metodologi ini digambarkan melalui beberapa hasil simulasi.

Kata-Kata Kunci : Model Bayesian Hirarki, Metode Reversible Jump MCMC, Segmentasi Sinyal, Prosess Moving-Average (MA) konstan sepotong demi sepotong.

Pendahuluan

Model MA (*Moving Average*) konstan sepotong demi sepotong merupakan model yang sering digunakan untuk memodelkan signal. Sinyal **EEG** (*Electro Enchepalo Gram* (**EEG**) dan sinyal *Electro Cardio Gram* (**ECG**) merupakan beberapa contoh data riil yang dapat dimodelkan oleh model MA konstan sepotong demi sepotong. Apabila model MA konstan sepotong demi sepotong dicocokkan terhadap sinyal, umumnya parameter model tidak diketahui. Parameter model di sini meliputi : banyaknya segmen, lokasi perubahan model MA dan parameter model MA untuk tiap-tiap segmen. Parameter model MA meliputi orde, koefisien dan variansi galat.

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan bayesian. Parameter model dipertimbangkan sebagai variabel random yang mempunyai distribusi tertentu. Distribusi ini dikenal sebagai distribusi prior. Distribusi prior dari parameter model dan fungsi kemungkinan dari sinyal dikombinasikan untuk mendapatkan distribusi posterior dari parameter model. Estimasi Bayesian didasarkan pada distribusi posterior.

Distribusi posterior mempunyai bentuk yang sangat rumit menyebabkan penentuan estimator tidak dapat dilakukan secara analitis. Untuk mengatasi masalah ini, digunakan algoritma *reversibel jump MCMC*.

Metode

Misalkan $X_1, X_2, \ldots, X_n$ adalah suatu sinyal. Sinyal ini dikatakan mempunyai model MA konstan sepotong demi sepotong dengan banyaknya segmen k (k = 0, 1, …, k_{max}) apabila (untuk t = 1, 2, …, n) sinyal tersebut memenuhi persamaan stokastik berikut (Suparman and Doisy, 2002) :

$$X_t = Z_t + \sum_{j=1}^{q_{i,k}} \theta_{i,k,j}^{(q_{i,k})} Z_{t-j}, \tag{1}$$

di mana $\tau_{i,k} < t \leq \tau_{i+1,k}$, $i=0,1,\Lambda,k$ dan di bawah asumsi k segmen : $\tau_{i,k}$ adalah waktu terjadinya perubahan model MA ke-i, dengan konvensi $\tau_{0,k} = 0$ dan $\tau_{k+1,k} = n$ dan untuk tiap-tiap segmen ke-i :

1. $q_{i,k}$ dan $\theta_{i,k}^{(q_{i,k})} = \left(\theta_{i,k,1}^{(q_{i,k})}, L, \theta_{i,k,q_{i,k}}^{(q_{i,k})}\right)$ adalah orde dan koefisien model MA yang bersesuaian dengan segmen ke-i.
2. Z_t adalah nilai galat pada saat t yang bersesuaian dengan segmen ke-i. Z_t dimodelkan sebagai distribusi normal dengan mean 0 dan variansi $\sigma_{i,k}^2$

Selanjutnya model MA ke-i (i=0,1, …, k) disebut inversibel jika dan hanya jika persamaan suku banyak

$$\phi(b) = 1 + \sum_{j=1}^{q_{i,k}} \theta_{i,k,j}^{(q_{i,k})} b$$

bernilai nol untuk nilai b di luar lingkaran dengan jari-jari sama dengan satu (Brockwell and Davis, 1991).

Apabila banyaknya segmen k diasumsikan diketahui, lokasi perubahan model MA diasumsikan diketahui dan orde yang diasumsikan diketahui, maka permasalahan segmentasi model MA konstan per segmen menjadi permasalahan identifikasi orde dan estimasi parameter model MA untuk tiap-tiap segmen.

Apabila orde model MA diasumsikan diketahui, maka permasalahan identifikasi orde model MA dan estimasi parameter model MA menjadi permasalahan estimasi parameter model MA. Estimasi parameter model MA dapat dilakukan dengan menggunakan berbagai metode. Metode-metode tersebut diantaranya diusulkan oleh Shaarawy and Broemeling (1984), Brockwell and Davis. (1991), Box *et al.* (1994) dan Suparman and Soejoeti (1999). Shaarawy and Broemeling. (1984) menggunakan Metode Bayesian untuk mengestimasi parameter MA. Sedangkan ketiga peneliti lainnya, Brockwell and Davis. (1991), Box *et al.* (1994) dan Suparman dan Soejoeti (1999), menggunakan Metode Kemungkinan Maksimum untuk mengestimasi parameter model MA. Selanjutnya metode identifikasi orde dan estimasi parameter model MA diusulkan oleh Suparman (2006).

Dalam penelitian ini, banyaknya segmen dan orde model MA untuk masing-masing segmen diasumsikan tidak diketahui. Algoritma *reversible jump MCMC* (Green 1995) digunakan untuk mendeteksi banyaknya segmen, lokasi perubahan model MA, mengidentifikasi orde model MA dan mengestimasi parameter model MA secara bersamaan dalam satu tahap. Untuk mengatasi masalah hiperparameter yang muncul, diadopsi Bayesian hirarki (Robert 1999). Kinerja algoritma yang diusulkan akan diuji dengan menggunakan sinyal sintesis.

Metode Bayesian Hirarki

Dalam penelitian ini digunakan pendekatan bayesian hirarki, yang akan diuraikan dalam bagian berikut. Andaikan

$$s = \left(x_{q_{maks}+1}, x_{q_{maks}+2}, L, x_n\right) \text{ adalah}$$

suatu realisasi dari model MA konstan sepotong demi sepotong. Jika nilai

$$s_0 = \left(x_1, x_2, L, x_{q_{maks}}\right) \text{ diketahui dan}$$

$$\theta = \left(k, \tau^{(k)}, \left\{\theta_{i,k}^{(q_{i,k})}\right\}_{i=0}^{k}, \sigma^{(k)}\right), \text{ maka fungsi}$$

kemungkinan dari s dapat ditulis kurang lebih sebagai berikut :

$$l(s|\theta) = \prod_{i=0}^{k} (2\pi\sigma_{i,k}^2)^{-\frac{1}{2}(\tau_{i+1,k}-\tau_{i,k})} \exp -\frac{1}{2\sigma_{i,k}^2} \sum_{t=\tau_{i,k}+1}^{\tau_{i+1,k}} \left(y_t - \sum_{j=1}^{q_{i,k}} G(\rho_{i,k}^{(q_{i,k})}) \hat{z}_{t-j}\right)^2 \qquad (2)$$

di mana

$$\hat{z}_1 = L = \hat{z}_{q_{maks}} = 0$$

Dan di mana untuk semua

$$t = q_{maks} + 1, L, n$$

$$\hat{z}_t = x_t - \sum_{j=1}^{q_{i,k}} G(\rho_{i,k}^{(q_{i,k})}) \hat{z}_{t-j},$$

$$n \in \left[\tau_{i,k}+1, \tau_{i+1,k}\right]$$

Misalkan $I_{q_{i,k}}$ adalah daerah inversilibite.

Dengan menggunakan transformasi

$$F: \theta_{i,k}^{(q_{i,k})} \in I_q \; a \; \rho_{i,k}^{(q_{i,k})} \in (-1,1)^{q_{i,k}} \qquad (3)$$

maka model MA $(X_t)_{t\in Z}$ inversibel jika dan hanya jika $\rho_{i,k}^{(q_{i,k})} \in (-1,1)^{q_{i,k}}$ (Bhansali, 1983). Apabila $\rho = \left(k, \tau^{(k)}, \left\{\rho_{i,k}^{(q_{i,k})}\right\}_{i=0}^{k}, \sigma^{(k)}\right)$, maka fungsi kemungkinan dapat ditulis kembali sebagai :

$$l(s|\rho)=$$

$$\prod_{i=0}^{k}(2\pi\sigma_{i,k}^{2})^{-\frac{1}{2}(\tau_{i+1,k}-\tau_{i,k})}\exp-\frac{1}{2\sigma_{i,k}^{2}} \quad (4)$$

$$\sum_{t=\tau_{i,k}+1}^{\tau_{i+1,k}}(x_t-\sum_{j=1}^{q_{i,k}}G^{-1}(\theta_{i,k}^{(q_{i,k})})\hat{z}_{t-j})^2$$

Penentukan distribusi prior untuk parameter -parameter tersebut di atas adalah sebagai berikut :

a) Orde $q_{i,k}$ berdistribusikan Binomial

dengan parameter λ :

$$\pi(q_{i,k}|\lambda)=C_{q_{maks}}^{q_{i,k}}\lambda^{q_{i,k}}(1-\lambda)^{q_{maks}-q_{i,k}}$$

b) Untuk orde $q_{i,k}$ ditentukan terlebih dahulu, vektor koefisien $\rho_{i,k}^{q_{i,k}}$ berdistribusikan seragam pada interval $(-1,1)^{q_{i,k}}$.

c) Variansi $\sigma_{i,k}^2$ berdistribusikan invers gamma dengan parameter $\alpha/2$ dan $\beta/2$:

$$\pi(\sigma_{i,k}^2|\alpha,\beta)=$$

$$\frac{(\beta/2)^{\alpha/2}}{\Gamma(\alpha/2)}(\sigma_{i,k}^2)^{-(1+\alpha/2)}\exp-\beta/(2\sigma_{i,k}^2)$$

Di sini parameter λ dia sumsikan berdistribusi seragam pada interval (0,1), nilai α diambil sama dengan 2 dan parameter β diasumsikan berdistribusi Jeffrey. Sehingga distribusi prior untuk parameter $H_1=(q_{i,k},\rho_{i,k}^{q_{i,k}},\sigma_{i,k}^2)$ dan $H_2=(\lambda,\beta)$ dapat dinyatakan sebagai :

$$\pi(H_1,H_2)=$$

$$\pi(q_{i,k}|\lambda)\pi(\rho_{i,k}^{q_{i,k}}|q_{i,k})\pi(\sigma_{i,k}^2|\alpha,\beta) \quad (5)$$

$$\pi(\lambda)\pi(\beta)$$

Menurut Teorema Bayes, maka distribusi a posteriori untuk parameter H_1 dan H_2 dapat dinyatakan sebagai :

$$\pi(H_1,H_2|s)\propto\lambda(s|H_1)\pi(H_1,H_2) \quad (6)$$

Distribusi a posteriori merupakan gabungan dari fungsi kemungkinan dan distribusi prior yang kita asumsikan sebelum sampel diambil. Fungsi kemungkinan bersifat obyektif sementara distribusi prior ini bersifat subyektif. Dalam kasus ini, distribusi a posteriori $\pi(H_1,H_2|s)$ mempunyai bentuk yang sangat rumit sehingga tidak dapat diselesaikan secara analitis. Untuk mengatasi masalah tersebut, diusulkan metode MCMC.

Metode MCMC

Misalkan M = (H_1, H_2). Secara umum, metode Markov Chain Monte Carlo (MCMC) merupakan suatu metode sampling, yaitu dengan cara membuat rantai Markov homogen $M_1,M_2,L\ ,M_m$ dapat dipertimbangkan sebagai variabel random yang mengikuti distribusi $\pi(H_1,H_2|s)$ Dengan demikian $M_1,M_2,L\ ,M_m$ dapat digunakan sebagai sarana untuk menaksir parameter M. Untuk merealisasikan itu diadopsi algoritma Gibbs Hibrida (Robert, 1996) yang terdiri dari dua tahap :

1. Simulasi distribusi $\pi(H_2|s)$
2. Simulasi distribusi $\pi(H_1,H_2|s)$

Algoritma Gibbs digunakan untuk mensimulasikan distribusi $(q_{i,k},\rho_{i,k}^{(q_{i,k})})$ dengan algoritma Gibbs untuk mensimulasikan parameter $\sigma_{i,k}^2$ digunakan untuk mensimulasikan distribusi $\pi(H_1,H_2|s)$ Algoritma RJMCMC merupakan rampatan dari algoritma Metropolis-Hastings (Metropolis *et al.*, 1953; Hastings, 1970).

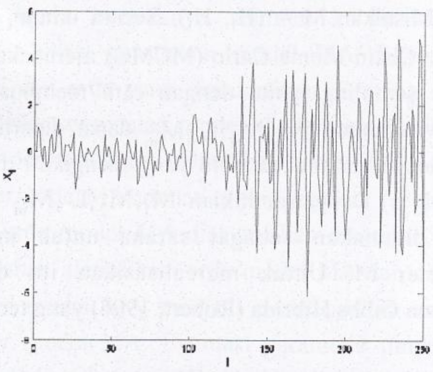
Estimator yang dihasilkan oleh metode MCMC dalam dua tahap. Tahap pertama adalah estimator dari orde $q_{i,k}$ Tahap kedua adalah estimator dari parameter model MA dan variansi $\sigma_{i,k}^2$ yang bersesuaian dengan orde $q_{i,k}$ yang diperoleh pada tahap pertama.

Pembahasan

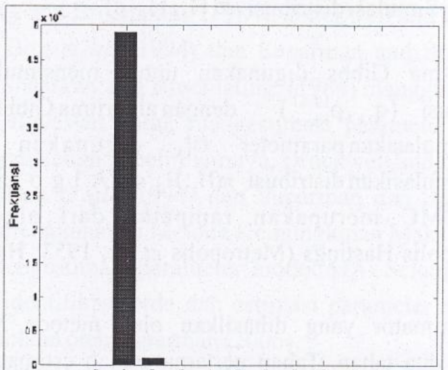
Sebagai ilustrasi, kita akan menerapkan metode ini untuk mengidentifikasi orde dan menaksir parameter sinyal sintesis. Studi simulasi ditempuh untuk mengkonfirmasi kinerja dari algoritma *reversible jump* MCMC apakah dapat berkerja dengan baik.

Algoritma *reversible jump* MCMC digunakan untuk mengestimasi banyaknya segmen, lokasi perubahan model MA, orde model MA untuk masing-masing segmen, dan koefisien model MA untuk masing-masing model MA serta variansi galat yang bersesuaiaan. Untuk keperluan itu, algoritma *reversible jump* MCMC dimplementasikan sebanyak 70000 iterasi dengan periode pemanasan sebanyak 10000 iterasi. Nilai orde q_{maks} dibatasi maksimum 10 sehingga $q_{maks} = 10$

Gambar 1 merupakan sinyal sintesis dengan model MA konstan sepotong demi sepotong yang dibuat menurut persamaan (1) di atas.



Gambar 1. Sinyal model MA sepotong demi sepotong.



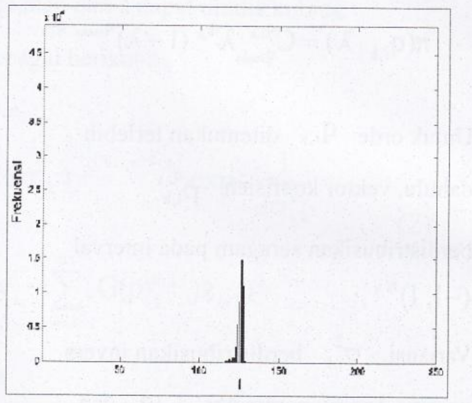
Gambar 2. Histogram dari banyaknya segmen

Pembuatan sinyal sintesis dilakukan dengan menggunakan bahasa pemograman MATLAB (Hanselman and Littlefield, 1997), dengan jumlah data n = 250, k = 1 dan waktu terjadinya perubahan model MA adalah $\tau = (125)$ sedangkan orde, koefisien, dan galat model MA untuk masing-masing segmen dinyatakan dalam Tabel 1 berikut.

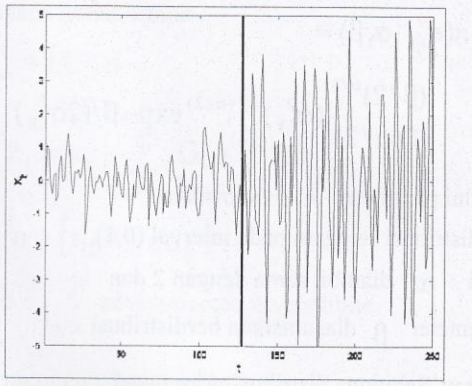
Tabel 1. Nilai parameter sinyal sintesis

Segmen ke - i	$\sigma_{i,1}$	$q_{i,1}$	$\phi_{i,1}^{(q_i,1)}$
0	0.5	1	(0.78)
1	1,5	3	(0.52, -0.09, -0.96)

Berdasarkan data dalam Gambar 1, selanjutnya parameter model diestimasi dengan menggunakan *reversible jump* MCMC. Histogram dari k disajikan pada Gambar 2. Hasilnya adalah

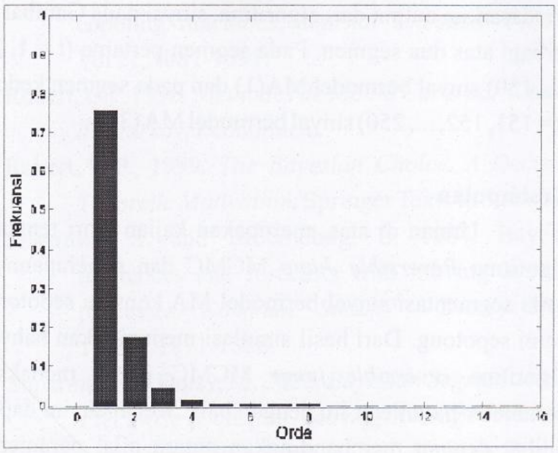


Gambar 3. Histogram lokasi perubahan model

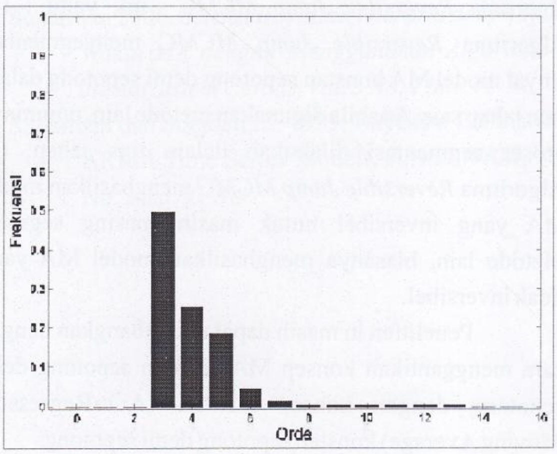


Gambar 4. Segmentasi sinyal

Histogram untuk τ yang bersesuaian dengan nilai $\hat{k} = 1$ diberikan pada Gambar 3. Hasilnya adalah $\hat{\tau} = (127)$ Hasil segmentasi disajikan dalam Gambar 4. Histogram untuk orde yang bersesuaian dengan nilai $\hat{k} = 1$ dan $\hat{\tau} = (127)$ diberikan pada Gambar 5 dan Gambar 6.

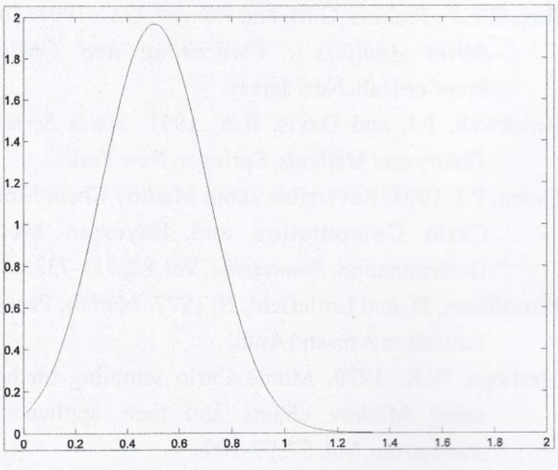


Gambar 5. Histogram orde segmen ke-0

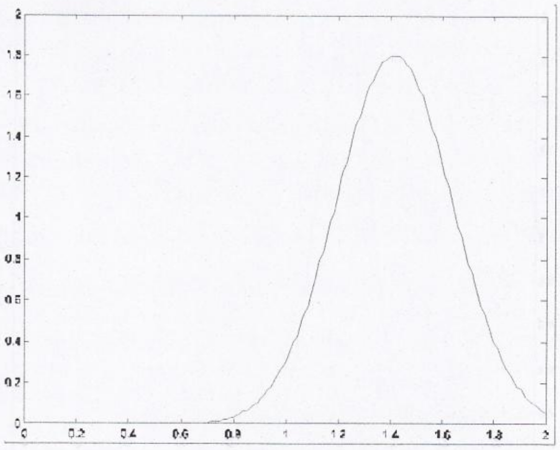


Gambar 6. Histogram orde segmen ke-1

Histogram untuk $\sigma_{i,1}$ yang bersesuaian dengan nilai $\hat{k} = 1$ dan $\hat{\tau} = (127)$ diberikan pada Gambar 7 dan Gambar 8.

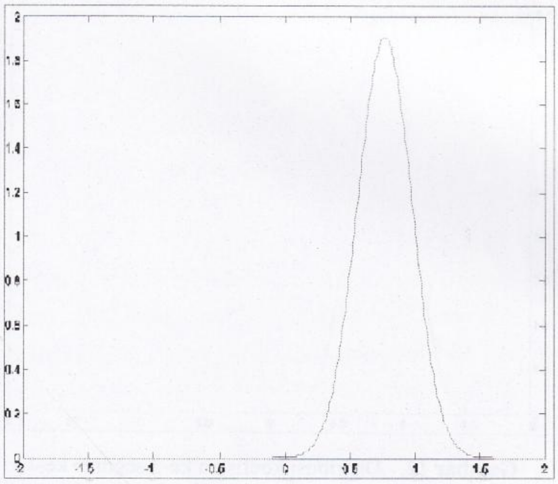


Gambar 7. Distribusi variansi galat segmen ke-0

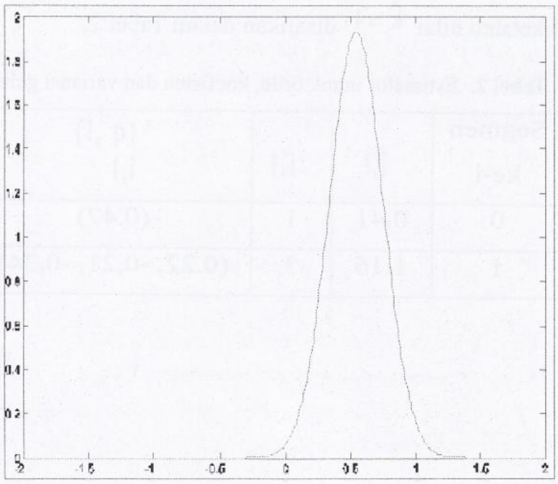


Gambar 8. Distribusi variansi galat segmen ke-1

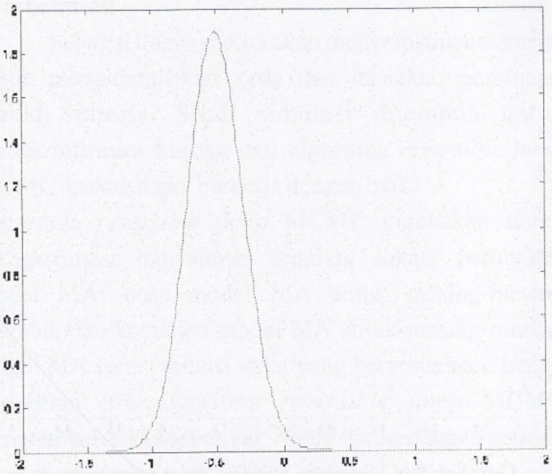
Sedangkan histogram untuk koefisien model MA yang bersesuaian diberikan pada Gambar 9-12.



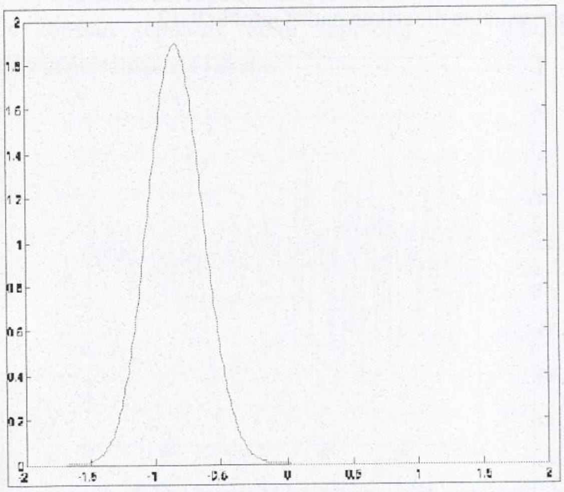
Gambar 9. Distribusi koefisien ke-1 segmen ke-0



Gambar 10. Distribusi koefisien ke-1 segmen ke-1



Gambar 11. Distribusi koefisien ke-2 segmen ke-1



Gambar 12. Distribusi koefisien ke-3 segmen ke-1

Hasil estimasi untuk orde, koefisien dan variansi galat diketahui nilai $\hat{k} = 1$ disajikan dalam Tabel 2.

Tabel 2. Estimator untuk orde, koefisien dan variansi galat

Segmen ke-i	i,1	i,1	(q ,1) i,1
0	0.41	1	(0.47)
1	1.16	3	(0.22, -0,21, -0,34)

Berdasarkan output dari algoritma, sinyal pada Gambar 1 terbagi atas dua segmen. Pada segmen pertama (t = 1, 2, ..., 150) sinyal bermodel MA(1) dan pada segmen kedua (t = 151, 152, ..., 250) sinyal bermodel MA (3).

Kesimpulan

Uraian di atas, merupakan kajian teori tentang algoritma *Reversible Jump* MCMC dan penerapannya pada segmentasi sinyal bermodel MA konstan sepotong demi sepotong. Dari hasil simulasi menunjukkan bahwa algoritma *reversible jump* MCMC dapat menaksir parameter-parameter itu dengan baik. Kebaikan ini dapat dilihat dengan membandingkan antara nilai parameter dari sinyal pada Tabel 1 dan nilai estimasi pada Tabel 2.

Beberapa kelebihan yang diperoleh dari algoritma *Reversible Jump MCMC* ini, yaitu : (1) Algoritma *Reversible Jump MCMC* menyegmentasi sinyal model MA konstan sepotong demi sepotong dalam satu tahap saja. Apabila digunakan metode lain, umumnya proses segmentasi dilakukan dalam tiga tahap. (2) Algoritma *Reversible Jump MCMC* menghasilkan model MA yang inversibel untuk masing-masing segmen. Metode lain, biasanya menghasilkan model MA yang tidak inversibel.

Penelitian in masih dapat dikembangkan dengan cara menggantikan konsep MA konstan sepotong demi sepotong dengan konsep **ARMA** (**A**uto**R**egressive **M**oving **A**verage) konstan sepotong demi sepotong.

Pustaka

Bhansali, R.J. 1983. The inverse partial correlation function of a time series and its applications, *J. Multivar. Anal.*, Vol. 13, 310-327.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis : Forecasting and Control*, Prentice Hall, New Jersey.

Brockwell, P.J. and Davis, R.A. 1991. *Times Series : Theory and Methods*, Springer, New York.

Green, P.J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, Vol. 82, 711-732.

Hanselman, D. and Littlefield, B. 1977. *Matlab*, Pearson Education Asia and Andi.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, Vol. 57, 97-109.

Metropolis, N., Rosenbluth, A.W., Teller, A.H. and Teller, E. 1953. Equations of state calculations by fast coputing machines, *Journal Chemical Physics*, Vol 21, 1087-1091.

Robert, C.P. 1996. *Méthodes de Monte Carlo par Chaînes de Markov*, Economica.

Robert, C.P., 1999. *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Texts in Statistics..

Shaarawy, S. and Broemeling, L. 1984. Bayesian inferences and forecasts with moving averages processes. *Commun. Statist. – Theory Meth.*, 13(15), 1871-1888.

Suparman and Doisy, M. 2002. *Bayesian Segmentation of Piecewise Constant Moving-Average Processes using Reversible Jump MCMC Methods Proc. Of the 7th Indonesian Student's Scientific Meeting, pp. 481-485, Berlin Germany.*

Suparman 2006. *Seleksi Bayesian Hierarki dalam Runtun Waktu MA dengan Menggunakan Algoritma SA, Jurnal Sistem Cerdas, Vol. 5 No. 1 hal. 43-54.*

Suparman dan Soejoeti, Z. 1999. Bayesian Estimation of ARMA Time Series Models, *Jurnal WKSI*, Vol. 2 No. 3 hal. 91-98.