

Estimasi Bayesian untuk Penentuan Besarnya Pengaruh Genetik terhadap Sifat Fenotip dan Studi Simulasinya

Adi Setiawan (adi_setia_03@yahoo.com)
Program Studi Matematika, Fakultas Sains dan Matematika
Universitas Kristen Satya Wacana
Jl. Diponegoro 52-60 Salatiga 50711, Indonesia

Abstract

Twins that have a particular categorical trait can be used to determine the genetic contribution to the trait. In this paper it is described a simulation study to generate a particular categorical data trait in MZ and DZ twin. The data is then used to find the genetic contribution to the trait by using Bayesian method.

Key-words: twin, genetic contribution, Bayesian method.

A. Pendahuluan

Dalam upaya menentukan besarnya pengaruh genetik terhadap sifat fenotip (trait) dapat digunakan metode momen dan metode maksimum likelihood yang menggunakan data trait pada pasangan kembar hasil simulasi (lihat dalam makalah Setiawan (2008)). Dalam makalah ini akan dijelaskan metode lain yang menggunakan pendekatan Bayesian.

B. Dasar Teori

Misalkan dimiliki suatu trait kuantitatif X dari suatu individu yang dipilih secara random dari suatu populasi. Trait X dapat dianggap mengikuti model

$$X = A + E,$$

dengan A dan E masing-masing adalah faktor genetik dan faktor lingkungan yang saling bebas. Dua individu yang diambil secara random dari suatu populasi masing-masing dengan trait X_1 dan X_2 dapat dimodelkan sebagai

$$X_1 = A_1 + C + E_1,$$

$$X_2 = A_2 + C + E_2,$$

dengan (G_1, G_2) , C , E_1 , E_2 saling bebas dan E_1 , E_2 berdistribusi identik. Dalam hal ini C adalah faktor lingkungan bersama dan E adalah faktor lingkungan bukan bersama. Model ini dinamakan model *ACE*.

Apabila suatu trait yang didekomposisi sebagai $X = A + C + E$ maka heritabilitas (heritability) didefinisikan sebagai

$$\frac{V(A)}{V(X)} = \frac{V(A)}{V(A)+V(C)+V(E)},$$

yang dapat diestimasi berdasarkan dari data pengamatan X . Heritabilitas dapat dipandang sebagai besarnya pengaruh faktor genetik terhadap sifat fenotip. Dalam model ACE, variansi dan kovariansi dari trait X_1 dan X_2 antara dua individu yang bersaudara dapat didekomposisi menjadi

$$V(X_1)=V(X_2)=\sigma^2 = \nu^2 + \eta^2 + \kappa^2, \quad (1)$$

$$Cov(X_1, X_2) = 2\Psi\nu^2 + \eta^2,$$

dengan ν^2 , η^2 dan κ^2 masing-masing adalah variansi faktor genetic A , variansi lingkungan bersama C dan variansi lingkungan bukan bersama E sedangkan Ψ merupakan koefisien kinship yang tergantung pada hubungan antara dua saudara tersebut (Lange, 2002). Untuk pasangan kembar MZ dan DZ masing-masing dapat digunakan $\Psi = 1/4$ dan $\Psi = 1/2$. Dekomposisi variansi pada persamaan (1) tersebut dapat diterapkan untuk trait kuantitatif maupun trait yang merupakan data kategori (*categorical trait*). Trait ini dapat dianggap dipengaruhi oleh trait lain yang tidak teramati (*unobservable trait*) yang dinamakan liabilitas (*liability*). Trait kategori yang teramati (*observable trait*) seperti berpenyakit tertentu atau tidak, akan berkaitan dengan suatu liabilitas yang melampaui batas (*threshold*). Hal tersebut dapat dijelaskan dalam model matematika berikut ini.

Misalkan Y_1 dan Y_2 adalah ukuran dari suatu trait dikotomi pada 2 anggota pasangan kembar. Kita menganggap bahwa vektor $(Y_1, Y_2)^t$ tergantung pada variabel laten $(X_1, X_2)^t$ dan suatu batas b melalui persamaan $Y_i = 0$ jika $X_i \leq b$ dan $Y_i = 1$ jika $X_i > b$ untuk $i = 1, 2$. Diasumsikan bahwa $(X_1, X_2)^t$ mempunyai distribusi normal dan dapat didekomposisi ke dalam komponen genetik A_1, A_2 , komponen lingkungan bersama C dan komponen lingkungan tidak bersama E_1, E_2 sebagai berikut :

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} + \begin{pmatrix} C \\ C \end{pmatrix} + \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \quad (2)$$

dengan

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \nu^2 & \nu^2 \\ \nu^2 & \nu^2 \end{pmatrix} \right)$$

pada pasangan kembar MZ dan

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \nu^2 & 0,5\nu^2 \\ 0,5\nu^2 & \nu^2 \end{pmatrix} \right)$$

pada pasangan kembar DZ, sedangkan

$$\begin{pmatrix} C \\ C \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2 & \eta^2 \\ \eta^2 & \eta^2 \end{pmatrix} \right),$$

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \kappa^2 & 0 \\ 0 & \kappa^2 \end{pmatrix} \right).$$

Dalam hal ini $(A_1, A_2)^t$, $(C, C)^t$ dan $(E_1, E_2)^t$ saling bebas. Probabilitas bersyarat bahwa $Y_1 = 0$ diberikan A_1, A_2 dan C adalah

$$\begin{aligned} P(Y_1=0|A_1, A_2, C) &= P(X_1 \leq b | A_1, A_2, C) \\ &= P(A_1 + C + E_1 \leq b | A_1, A_2, C) \\ &= P(E_1 \leq b - A_1 - C | A_1, A_2, C) \\ &= \Phi \left(\frac{b - a_1 - c}{\sqrt{\kappa^2}} \right) \end{aligned}$$

dan probabilitas bersyarat bahwa $Y_1 = 1$ diberikan A_1, A_2 dan C adalah

$$\begin{aligned} P(Y_1=1|A_1, A_2, C) &= P(X_1 > b | A_1, A_2, C) \\ &= P(A_1 + C + E_1 > b | A_1, A_2, C) \\ &= P(E_1 > b - A_1 - C | A_1, A_2, C) \\ &= 1 - \Phi \left(\frac{b - a_1 - c}{\sqrt{\kappa^2}} \right). \end{aligned}$$

Jika diberikan A_1, A_2 dan C maka Y_1 dan Y_2 variabel saling bebas. Probabilitas bersyarat dari Y_2 jika diberikan A_1, A_2 dan C dapat ditentukan dengan cara yang sama. Batas b dapat distandardisasi menjadi

$$b' = \frac{b}{\sqrt{V(X_1)}} = \frac{b}{\sqrt{\nu^2 + \eta^2 + \kappa^2}}. \quad (3)$$

Lebih jauh, heritabilitas atau komponen genetic $V_g = \sigma^2_A$ dari status berpenyakit atau tidak dapat ditentukan dengan

$$V_g = \sigma^2_A = \frac{V(A_1)}{V(X_1)} = \frac{v^2}{v^2 + \eta^2 + \kappa^2},$$

komponen lingkungan bersama $V_s = \sigma^2_C$ dengan

$$V_s = \sigma^2_C = \frac{V(C)}{V(X_1)} = \frac{\eta^2}{v^2 + \eta^2 + \kappa^2}$$

dan komponen lingkungan tidak bersama $V_u = \sigma^2_E$ dengan

$$V_u = \sigma^2_E = \frac{V(E_1)}{V(X_1)} = \frac{\kappa^2}{v^2 + \eta^2 + \kappa^2}.$$

Misalkan pemetaan $(y_1, y_2, a, c) \rightarrow p(y_1, y_2, a, c)$ adalah fungsi densitas dari vektor (Y_1, Y_2, A, C) dan $y_j \rightarrow p(y_j | a, c)$ adalah densitas bersyarat dari Y_j diberikan (A, C) untuk $j = 1, 2$. Fungsi likelihood untuk pengamatan vektor dalam n pasangan kembar MZ yang dipilih secara acak (dengan $A := A_1, A_2$) adalah

$$\begin{aligned} L_{MZ} &= \prod_{i=1}^n p(y_{i1}, y_{i2}, a_i, c_i) \\ &= \prod_{i=1}^n p(y_{i1}, y_{i2} | a_i, c_i) f_A(a_i) f_C(c_i) \\ &= \prod_{i=1}^n p(y_{i1} | a_i, c_i) p(y_{i2} | a_i, c_i) f_A(a_i) f_C(c_i) \\ &= \prod_{i=1}^n q(y_{i1}, a_i, c_i) q(y_{i2}, a_i, c_i) f_A(a_i) f_C(c_i) \end{aligned}$$

dengan f_A dan f_C masing-masing adalah fungsi densitas dari A dan C yang diberikan oleh

$$\begin{aligned} f_A(a_i) &= \frac{1}{\sqrt{2\pi v^2}} \exp\left[-\frac{a_i}{2v^2}\right], \\ f_C(c_i) &= \frac{1}{\sqrt{2\pi \eta^2}} \exp\left[-\frac{c_i}{2\eta^2}\right], \end{aligned}$$

dan

$$q(y, a, c) = \left[\Phi\left(\frac{b-a-c}{\sqrt{\kappa^2}}\right) \right]^{1-y} \left[1 - \Phi\left(\frac{b-a-c}{\sqrt{\kappa^2}}\right) \right]^y$$

$$= [\Phi((b-a-c)\sqrt{\theta_1})]^{1-y} [1 - \Phi((b-a-c)\sqrt{\theta_1})]^y$$

dengan $\theta_1 = 1/\kappa^2$. Fungsi likelihood ini akan sebanding dengan

$$L_{MZ} \propto [v^2]^{-n/2} [\eta^2]^{-n/2} \exp\left[-\frac{\sum_{i=1}^n a_i^2}{2v^2}\right] \exp\left[-\frac{\sum_{i=1}^n c_i^2}{2\eta^2}\right] \prod_{i=1}^n q(y_{i1}, a_i, c_i) q(y_{i2}, a_i, c_i)$$

Dengan kata lain, fungsi likelihood untuk n pasangan kembar MZ akan sebanding dengan

$$L_{MZ} \propto \theta_3^{-n/2} \theta_4^{-n/2} \exp\left[-\frac{\theta_3 \sum_{i=1}^n a_i^2}{2}\right] \exp\left[-\frac{\theta_4 \sum_{i=1}^n c_i^2}{2}\right] \prod_{i=1}^n q(y_{i1}, a_i, c_i) q(y_{i2}, a_i, c_i)$$

dengan $\theta_3 = 1/v^2$ dan $\theta_4 = 1/\eta^2$.

Fungsi likelihood untuk pengamatan (Y_1, Y_2, A_1, A_2, C) dalam m pasangan kembar DZ yang dipilih secara acak diperoleh dengan cara yang sama dengan hanya fungsi densitas bersama (A_1, A_2) yang berubah yaitu

$$\begin{aligned} L_{DZ} &= \prod_{i=1}^n p(y_{i1}, y_{i2}, a_{i1}, a_{i2}, c_i) \\ &= \prod_{i=1}^n p(y_{i1}, y_{i2} | a_{i1}, a_{i2}, c_i) f(a_{i1}, a_{i2}) f_C(c_i) \\ &= \prod_{i=1}^n p(y_{i1} | a_{i1}, a_{i2}, c_i) p(y_{i2} | a_{i1}, a_{i2}, c_i) f(a_{i1}, a_{i2}) f_C(c_i) \\ &= \prod_{i=1}^n q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i) g(a_{i1} | a_{i2}) f_A(a_{i2}) f_C(c_i) \end{aligned}$$

dengan f adalah fungsi densitas bersama dari (A_1, A_2) dan g adalah fungsi densitas bersyarat dari A_1 jika diberikan A_2 yang dinyatakan dengan rumus

$$f(a_{i1}, a_{i2}) = \frac{1}{2\pi \sqrt{1 - \left(\frac{1}{2}\right)^2}} \exp\left[-\frac{1}{2v^2 \left(1 - \left(\frac{1}{2}\right)^2\right)} \left(a_{i1}^2 - 2(0,5) a_{i1} a_{i2} + a_{i2}^2\right)\right]$$

$$g_A(a_{i1} | a_{i2}) = \frac{1}{\sqrt{2\pi \left(\frac{3}{4}\right) v^2}} \exp \left[-\frac{(a_{i1} - 0,5 a_{i2})^2}{2 \left(\frac{3}{4}\right) v^2} \right].$$

Hal itu berarti bahwa fungsi likelihoodnya sebanding dengan

$$L_{DZ} \propto [v^2]^{-n/2} [\eta^2]^{-n/2} \exp[-x] \prod_{i=1}^n q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i)$$

dengan

$$x = \frac{2 \sum_{i=1}^m (a_{i1} - 0,5 a_{i2})^2}{3v^2} + \frac{\sum_{i=1}^m a_{i2}^2}{2v^2} + \frac{\sum_{i=1}^m c_i^2}{2\eta^2}.$$

Dengan kata lain, fungsi likelihood sebanding dengan

$$L_{DZ} \propto \theta_3^m \theta_4^{m/2} \exp[-z] \prod_{i=1}^n q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i)$$

dengan

$$z = \frac{2 \theta_3 \sum_{i=1}^m (a_{i1} - 0,5 a_{i2})^2}{3} + \frac{\theta_3 \sum_{i=1}^m a_{i2}^2}{2} + \frac{\theta_4 \sum_{i=1}^m c_i^2}{2}.$$

Kita memilih prior konjugat untuk parameter θ_3 , θ_4 , θ_1 dari keluarga distribusi gamma sehingga fungsi densitas priornya adalah

$$\pi_1(\theta_3) = \frac{p_2^{p_1}}{\Gamma(p_1)} \theta_3^{p_1-1} \exp(-p_2 \theta_3),$$

$$\pi_2(\theta_4) = \frac{p_4^{p_3}}{\Gamma(p_3)} \theta_4^{p_3-1} \exp(-p_4 \theta_4),$$

$$\pi_3(\theta_1) = \frac{p_6^{p_5}}{\Gamma(p_5)} \theta_1^{p_5-1} \exp(-p_6 \theta_1),$$

dan distribusi konjugat prior untuk parameter b dari keluarga distribusi normal, yaitu fungsi densitas prior :

$$\pi_4(b) = \frac{1}{\sqrt{2\pi p_8}} \exp\left(-\frac{(b - p_7)^2}{2 p_8}\right).$$

Dalam hal ini p_1, p_2, \dots, p_8 adalah parameter yang dipilih yang sesuai. Berdasarkan pada anggapan bahwa parameter saling bebas maka fungsi densitas bersamanya adalah

$$\pi(\theta_3, \theta_4, \theta_1, b) = \pi_1(\theta_3) \pi_2(\theta_4) \pi_3(\theta_1) \pi_4(b).$$

Akibatnya fungsi densitasnya sebanding dengan

$$(\theta_3, \theta_4, \theta_1, b) \rightarrow \pi(\theta_3, \theta_4, \theta_1, b) L_{MZ} L_{DZ}$$

Fungsi densitas posterior bersama $\pi(\theta_1, \theta_3, \theta_4, b | data)$ memenuhi

$$\pi(\theta_1, \theta_3, \theta_4, b | data) \propto \theta_3^{(p_1 + \frac{n}{2} + m - 1)} \theta_4^{(p_3 + \frac{n}{2} + \frac{m}{2} - 1)} \theta_1^{p_5 - 1} \exp[-w_1] w_2,$$

dengan

$$w_1 = p_6 \theta_1 + u \theta_3 + v \theta_4 + \frac{(b - p_7)^2}{2p_8},$$

$$w_2 = \prod_{i=1}^n q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i) \prod_{i=1}^m q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i)$$

dan

$$u = p_2 + \frac{1}{2} \sum_{i=1}^n a_i^2 + \frac{2 \sum_{i=1}^m (a_{i1} - 0,5a_{i2})^2}{3} + \frac{\sum_{i=1}^m a_{i2}^2}{2}$$

$$v = p_4 + \frac{1}{2} \sum_{i=1}^n c_{i1}^2 + \frac{1}{2} \sum_{i=1}^m c_{i2}^2.$$

Berdasarkan pada fungsi densitas bersama, distribusi bersyarat penuh (*full conditional distribution*) untuk masing-masing parameter dapat dinyatakan sebagai berikut :

$$\pi(\theta_3 | \text{yang lain}) \propto \theta_3^{(p_1 + \frac{n}{2} + m - 1)} \exp p[-u\theta_3] \prod_{i=1}^m q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i),$$

$$\pi(\theta_4 | \text{yang lain}) \propto \theta_4^{(p_3 + \frac{n}{2} + \frac{m}{2} - 1)} \exp p[-v\theta_4] \prod_{i=1}^m q(y_{i1}, a_{i1}, c_i) q(y_{i2}, a_{i2}, c_i),$$

$$\pi(\theta_1 | \text{yang lain}) \propto \theta_1^{p_5 - 1} \exp p[-p_6\theta_1] w_2,$$

$$\pi(b | \text{yang lain}) \propto \exp \left[-\frac{(b - p_7)^2}{2p_8} \right] w_2.$$

Untuk variabel laten, distribusi bersyarat penuh dapat ditentukan dengan

$$\pi(a_i | \text{yang lain}) \propto \exp \left[-\frac{a_i^2 \theta_3}{2} \right] q(y_{i1}, a_i, c_i) q(y_{i2}, a_i, c_i)$$

untuk $i=1, 2, \dots, n$,

$$\pi(c_i | \text{yang lain}) \propto \exp \left[-\frac{c_i^2 \theta_4}{2} \right] q(y_{i1}, a_i, c_i) q(y_{i2}, a_i, c_i)$$

untuk $i=1, 2, \dots, n + m$,

$$\pi(a_{i1} | \text{yang lain}) \propto \exp\left[-\frac{2(a_{i1} - 0,5 a_{i2})^2 \theta_3}{2}\right] q(y_{i1}, a_{i1}, c_i)$$

untuk $i=1,2, \dots, m$,

$$\pi(a_{i2} | \text{yang lain}) \propto \exp\left[-\frac{2(a_{i2} - 0,5 a_{i1})^2 \theta_3}{2}\right] q(y_{i2}, a_{i2}, c_i)$$

untuk $i=1,2, \dots, m$.

Untuk mengkonstruksikan penyampel Markov Chain Monte Carlo (MCMC), kita menggunakan algoritma Gibbs Sampler sebagai berikut :

1. Inisialisasi parameter $[\theta_3]_0, [\theta_4]_0, [\theta_1]_0, b_0, [a_1]_0, [c_1]_0, \dots, [a_n]_0, [c_n]_0, [a_{11}]_0, [a_{12}]_0, [c_1]_0, \dots, [a_{m1}]_0, [a_{m2}]_0, [c_m]_0$, dan diset $j=1$.
2. Dibangkitkan $[\theta_3]_j \sim \theta_3 \rightarrow \pi(\theta_3 | \text{yang lain})$ dengan yang lain berarti parameter yang lain.
3. Dibangkitkan $[\theta_4]_j \sim \theta_4 \rightarrow \pi(\theta_4 | \text{yang lain})$.
4. Dibangkitkan $\theta_1 \sim \theta_1 \rightarrow \pi(\theta_1 | \text{yang lain})$.
5. Dibangkitkan $b_j \sim b \rightarrow \pi(b | \text{yang lain})$.
6. Dibangkitkan $[a_i]_j \sim a_i \rightarrow \pi(a_i | \text{yang lain})$.
7. Dibangkitkan $[c_i]_j \sim c_i \rightarrow \pi(c_i | \text{yang lain})$.
8. Dibangkitkan $[a_{i1}]_j \sim a_{i1} \rightarrow \pi(a_{i1} | \text{yang lain})$.
9. Dibangkitkan $[a_{i2}]_j \sim a_{i2} \rightarrow \pi(a_{i2} | \text{yang lain})$.
10. Langkah 2 sampai 9 untuk $j = 1, 2, \dots$ sampai rantai Markov (*Markov chain*) menjadi konvergen.

Distribusi bersyarat di atas tidak ada yang merupakan anggota keluarga standard. Untuk menyampelnya kita menggunakan algoritma Metropolis-Hasting yang diusulkan sebagai berikut :

1. Distribusi eksponensial dengan rata-rata nilai θ_3 sebelumnya untuk parameter θ_3 yaitu

$$p(y | \theta_3) = \frac{1}{\theta_3} \exp\left(-\frac{y}{\theta_3}\right), y > 0,$$

2. Distribusi eksponensial dengan rata-rata nilai θ_4 sebelumnya untuk parameter θ_4 ,

3. Distribusi eksponensial dengan rata-rata nilai θ_1 sebelumnya untuk parameter θ_1 ,
4. Distribusi normal dengan rata-rata nilai b sebelumnya dan variansi 1 untuk variabel laten b ,
5. Distribusi normal dengan rata-rata nilai a_i sebelumnya dan variansi 1 untuk variabel laten a_i ,
6. Distribusi normal dengan rata-rata nilai c_i sebelumnya dan variansi 1 untuk variabel laten c_i ,
7. Distribusi normal dengan rata-rata nilai a_{i1} sebelumnya dan variansi 1 untuk variabel laten a_{i1} ,
8. Distribusi normal dengan rata-rata nilai a_{i2} sebelumnya dan variansi 1 untuk variabel laten a_{i2} .

C. Studi Simulasi, Hasil dan Pembahasan

Paket program R digunakan untuk membangkitkan data kategorikal pada sampel pasangan kembar MZ dan DZ dengan menggunakan model pada persamaan (2). Dalam simulasi ini, kami memilih menggunakan 400 pasangan kembar dan menggunakan input pengaruh faktor genetik $\sigma_A^2 = 0,3$, pengaruh lingkungan bersama $\sigma_C^2 = 0,4$ dan pengaruh lingkungan tidak bersama $\sigma_E^2 = 0,3$ untuk memberikan gambaran bagaimana metode ini digunakan. Tabel 1 dan Tabel 2 merupakan contoh dari hasil simulasi tersebut. Dari tabel tersebut, terlihat bahwa terdapat 400 pasangan kembar. Tabel 1 berarti bahwa dari 400 pasang tersebut, terdapat 313 pasang yang keduanya tidak berpenyakit tertentu yang menjadi perhatian (*categorical trait*), 30 pasang kembar yang orang kembar 1 tidak berpenyakit sedangkan orang kembar 2 berpenyakit, 25 pasang kembar yang orang kembar 1 berpenyakit sedangkan orang kembar 2 tidak berpenyakit, dan 32 pasang yang keduanya berpenyakit.

Tabel 1. Tabel kontingensi dari status berpenyakit tertentu (kategori 1) atau tidak (kategori 0) pada pasangan kembar MZ.

	Kembar 2	
Kembar 1	Kategori 0	Kategori 1
Kategori 0	313	30
Kategori 1	25	32

Tabel 2. Tabel kontingensi dari status berpenyakit tertentu (kategori 1) atau tidak (kategori 0) pada pasangan kembar DZ.

	Kembar 2	
Kembar 1	Kategori 0	Kategori 1
Kategori 0	300	40
Kategori 1	34	26

Tabel 3. Data hasil simulasi dari status berpenyakit tertentu atau trait kategori (*categorical trait*) pada pasangan kembar MZ dan DZ serta estimasi kembali pengaruh faktor genetik σ^2_A , lingkungan bersama σ^2_C dan lingkungan tidak bersama σ^2_E dengan menggunakan metode Bayesian.

No	MZ				DZ				Metode Bayesian		
	(0,0)	(0,1)	(1,0)	(1,1)	(0,0)	(0,1)	(1,0)	(1,1)	σ^2_A	σ^2_C	σ^2_E
1	313	30	25	32	300	40	34	26	0.26 (0.10, 0.55)	0.36 (0.20, 0.49)	0.38 (0.28, 0.48)
2	298	33	29	40	306	34	38	22	0.34 (0.04, 0.63)	0.36 (0.12, 0.63)	0.30 (0.20, 0.42)
3	305	34	29	32	308	37	28	27	0.20 (0.06, 0.52)	0.47 (0.17, 0.63)	0.33 (0.22, 0.44)
4	292	27	36	45	296	44	37	23	0.56 (0.27, 0.75)	0.16 (0.02, 0.41)	0.27 (0.18, 0.38)
5	318	22	27	33	284	41	41	31	0.55 (0.10, 0.67)	0.19 (0.02, 0.50)	0.26 (0.16, 0.38)

Bila simulasi dilakukan $n=5$ kali maka akan diperoleh hasil lengkap seperti dinyatakan pada Tabel 3. Berdasarkan data kategorikal pasangan kembar MZ dan DZ, untuk mengestimasi besarnya komponen genetik $V_g=\sigma^2_A$, komponen lingkungan bersama $V_s=\sigma^2_C$ dan komponen lingkungan tidak bersama $V_u=\sigma^2_E$ metode yang telah dijelaskan di atas diimplementasikan dalam WinBUGS versi 1.4 (untuk gambaran penggunaan WINBUGS versi 1.4 dalam estimasi Bayesian, lihat makalah Cowles, 2004). Untuk parameter $1/\nu^2$, $1/\eta^2$ dan $1/\kappa^2$, dipilih prior $\Gamma(1,1)$. Berdasarkan pada prior ini, variansi dari variabel laten X_1 yang dinyatakan dengan $V(X_1) = \nu^2 + \eta^2 + \kappa^2$ akan memberikan probabilitas yang tinggi pada interval (2,20). Prior dari b' digunakan distribusi $N(0,1)$ sehingga dalam pandangan persamaan (3), parameter b memberikan probabilitas yang tinggi pada interval

$$(-3\sqrt{(20-2)}, 3\sqrt{(20-2)})$$

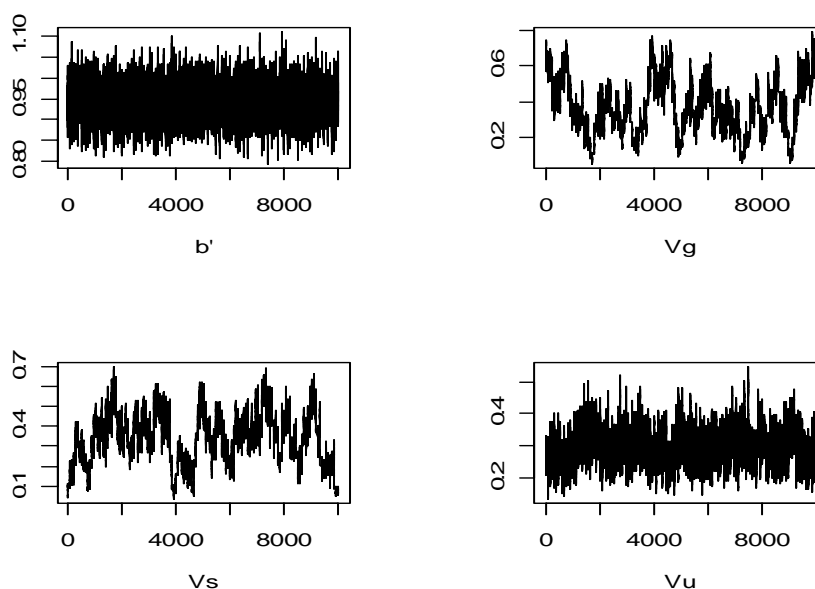
sehingga prior distribusi $N(0, 18)$ akan merupakan pemilihan yang beralasan untuk parameter b . Dalam hal ini digunakan median dari rantai Markov dalam MCMC untuk mengestimasi ketiga komponen tersebut. Nilai dalam tanda kurung memberikan estimasi interval kredibel 95 % (*credible interval*) untuk metode Bayesian. Gambar 1 dan Gambar 2 masing-masing memberikan plot MCMC dan estimasi density (*kernel density estimation*) untuk parameter-parameter yang diperlukan. Apabila digunakan

input pengaruh faktor genetik $\sigma_A^2 = 0,8$, pengaruh lingkungan bersama $\sigma_C^2 = 0,1$ dan pengaruh lingkungan tidak bersama $\sigma_E^2 = 0,1$ akan dihasilkan Tabel 4 sedangkan bila digunakan input pengaruh faktor genetik $\sigma_A^2 = 0,1$, pengaruh lingkungan bersama $\sigma_C^2 = 0,4$ dan pengaruh lingkungan tidak bersama $\sigma_E^2 = 0,3$ akan dihasilkan Tabel 5. Berdasarkan pada Tabel 3, Tabel 4 dan Tabel 5, terlihat bahwa metode Bayesian memberikan estimasi yang relatif memuaskan karena sesuai dengan parameter yang digunakan untuk membangkitkan data pasangan kembar MZ dan DZ.

Tabel 4. Data hasil simulasi dari status berpenyakit tertentu atau trait kategori (*categorical trait*) pada pasangan kembar MZ dan DZ serta estimasi kembali pengaruh faktor genetik σ_A^2 , lingkungan bersama σ_C^2 dan lingkungan tidak bersama σ_E^2 dengan menggunakan metode Bayesian. Dalam hal ini digunakan input $\sigma_A^2 = 0,8$, $\sigma_C^2 = 0,1$ dan $\sigma_E^2 = 0,1$.

No	MZ				DZ				Metode Bayesian		
	(0,0)	(0,1)	(1,0)	(1,1)	(0,0)	(0,1)	(1,0)	(1,1)	σ_A^2	σ_C^2	σ_E^2
1	313	17	21	49	285	46	40	29	0.77 (0.52, 0.89)	0.12 (0.03, 0.34)	0.12 (0.06, 0.19)
2	321	13	20	46	306	32	39	23	0.79 (0.57, 0.92)	0.11 (0.01, 0.29)	0.09 (0.05, 0.15)
3	323	19	21	37	302	42	35	21	0.68 (0.43, 0.85)	0.15 (0.02, 0.38)	0.17 (0.09, 0.26)
4	319	12	20	49	294	34	45	27	0.85 (0.57, 0.94)	0.07 (0.01, 0.32)	0.08 (0.04, 0.15)
5	320	16	13	51	288	36	46	30	0.81 (0.58, 0.91)	0.12 (0.03, 0.38)	0.07 (0.04, 0.17)

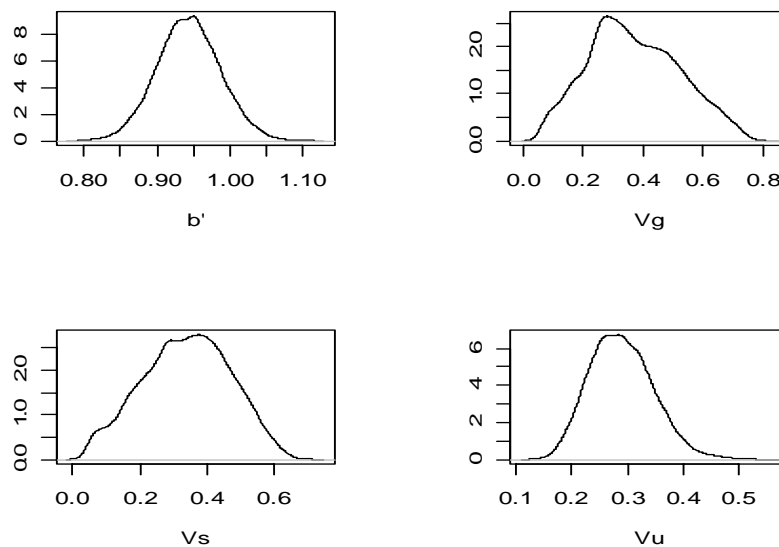
Gambar 1. Plot MCMC dengan ukuran sampel 10000 untuk batas b' pengaruh faktor genetik $V_g = \sigma_A^2$, lingkungan bersama $V_s = \sigma_C^2$ dan lingkungan tidak bersama $V_u = \sigma_E^2$.



Tabel 5. Data hasil simulasi dari status berpenyakit tertentu atau trait kategori (*categorical trait*) pada pasangan kembar MZ dan DZ serta estimasi kembali pengaruh faktor genetik σ^2_A , lingkungan bersama σ^2_C dan lingkungan tidak bersama σ^2_E dengan menggunakan metode Bayesian. Dalam hal ini digunakan input $\sigma^2_A = 0,1$, $\sigma^2_A = 0,8$ dan $\sigma^2_A = 0,1$.

No	MZ				DZ				Metode Bayesian		
	(0,0)	(0,1)	(1,0)	(1,1)	(0,0)	(0,1)	(1,0)	(1,1)	σ^2_A	σ^2_C	σ^2_E
1	321	22	13	44	324	18	18	40	0.10 (0.01, 0.24)	0.80 (0.67, 0.89)	0.10 (0.05, 0.15)
2	318	20	19	43	301	27	31	41	0.19 (0.07, 0.45)	0.65 (0.42, 0.78)	0.14 (0.08, 0.22)
3	307	19	17	57	315	27	24	34	0.25 (0.06, 0.55)	0.65 (0.36, 0.82)	0.10 (0.05, 0.17)
4	320	20	18	42	321	15	21	43	0.05 (0.00, 0.16)	0.84 (0.72, 0.90)	0.11 (0.07, 0.17)
5	310	16	24	50	307	17	21	55	0.06 (0.02, 0.17)	0.84 (0.73, 0.90)	0.10 (0.06, 0.15)

Gambar 2. Estimasi densitas untuk batas b' pengaruh faktor genetik $V_g = \sigma^2_A$, lingkungan bersama $V_s = \sigma^2_C$ dan lingkungan tidak bersama $V_u = \sigma^2_E$.



Pendekatan Bayesian dengan bantuan MCMC dalam kasus ini merupakan masalah yang relatif baru (Eaves dan Erkanli, 2003). Makalah atau hasil penelitian lain yang terkait dengan pendekatan Bayesian dalam studi pasangan kembar (*twin study*) sebagai contohnya adalah ditulis oleh Eaves et al. (2004), Eaves et al. (2005), van den Berg et al. (2006) dan Setiawan (2007).

D. Kesimpulan

Dalam makalah ini telah dijelaskan bagaimana pendekatan Bayesian digunakan untuk menentukan besarnya pengaruh genetik terhadap sifat fenotip tertentu dari data yang diperoleh dengan cara simulasi. Penelitian ini dapat dikembangkan pada studi simulasi yang menggunakan dua trait kategorikal pada pasangan kembar MZ dan DZ.

E. Daftar Pustaka

- [1] Lange, K. [2002], *Mathematics and Statistical Methods for Genetic Analysis*, Springer, New York
- [2] Berg, S. M. van den, Setiawan, A., Bartels, M., Polderman, T.J.C., van der Vaart, A.W., Boomsma, D.I., (2006), Individual Differences in Puberty Onset in Girls : Bayesian Estimation of Heritabilities and Genetic Correlations, *Behavior Genetics*, 36 (2) : 261-270.
- [3] Cowles, M. K., (2004), Review of WinBUGS 1.4, *Am. Stat.* 58:330-336.
- [4] Eaves, L. J., and Erkanli, A. (2003). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and $G \times E$ interaction. *Behav. Genet.* 33:279-299.
- [5] Eaves, L., Silberg, J., Foley, D., Bulik, C., Maes H., Erkanli A., Angold A., Costello E. J., Worthman C. (2004). Genetic and environmental influences on the relative timing of pubertal change, *Twin Res.* 7:471-481.
- [6] Eaves, L., Erkanli, A., Silberg, J., Angold, A., Maes, H. H., Foley, D., (2005), Application of Bayesian Inference using Gibbs Sampling to Item Response Theory Modeling of Multi Symptom genetic Data, *Behavior Genetics* 35 (6) : 765-780.
- [7] Setiawan, A. [2007] , *Statistical Data Analysis of Genetic Data in Twin Studies and Association Studies*, Vrije Universiteit, Amsterdam, Ph.D Thesis, ISBN 978-90-9021728.
- [8] Setiawan, A. [2008] , Penentuan Besarnya Pengaruh Faktor Genetik terhadap Sifat Fenotip dengan Metode Pasangan Kembar, *Prosiding Seminar Basic Science V 2008 di Universitas Brawijaya, Malang*.