MENGHITUNG NILAI PROBABILITAS PADA DISTRIBUSI NORMAL MULTIVARIATE

Jaka Nugraha
Jurusan Statistika
FMIPA Universitas Islam Indonesia
e-mail: jnugraha@fmipa.uii.ac.id

Abstrak

Menghitung nilai probabilitas variabel random yang mempunyai distribusi multivariat normal, merupakan salah satu kendala dalam model probit. Persamaan probabilitas dalam model probit tidak berbentuk persamaan tertutup dan harus diselesaikan secara numerik maupun simulasi. Metode Gezn merupakan metode yang paling efektif untuk menghitung nilai probabilitas normal multivariat. Metode ini banyak diimplementasikan dalam beberapa "package" pada program R. Metode ini juga telah diimplementasikan dalam model multinomial probit yang populer disebut dengan simulator Geweke-Hajivassiliou-Keane (GHK).

Kata kunci : model multinomial probit, faktor Cholesky, random utiliti, deret Taylor

PENDAHULUAN

Menghitung nilai probabilitas variabel random yang mempunyai distribusi multivariat normal, merupakan salah satu kendala dalam model probit. Persamaan probabilitas dalam model probit tidak berbentuk persamaan tertutup dan harus diselesaikan secara numerik. Beberapa prosedur non simulasi telah banyak digunakan seperti metode kuadrat untuk mendekati hitungan integral menggunakan fungsi pembobot. Metode ini hanya efektif untuk hitungan integral dimensi rendah. Metode ini dapat digunakan untuk integral dengan dimensi tidak lebih dari empat atau lima (Geweke, 1996). Metode non simulasi yang lain adalah Clark algoritma, yang telah disampaikan oleh Daganzo dkk (1977). Metode ini sangat tidak akurat sebab akurasi dipengaruhi oleh modelnya (Horowitz dkk, 1982).

Untuk menghitung probabilitas dalam model probit dapat juga digunakan metode simulasi. Banyak cara simulasi yang dapat digunakan dalam model probit, secara ringkas dijelaskan oleh Hajivassiliou dkk. (1996) dan menyimpulkan bahwa Geweke-Hajivassiliou-Keane (GHK) adalah metode yang paling baik paling baik. GHK simulator menghasilkan nilai antara nul s/d satu, dan bersifat unbiased dan konsisten. (Paul C. dkk, 2001)

Permasalahan yang sering muncul dalam statistika adalah menghitung probabilitas normal multivariate. Dalam makalah ini akan dibahas metode menghitung nilai probabilitas variabel random yang mempunyai distribusi multivariat normal menggunakan program R. Pembahasan dimulai dari distribusi normal univariat, distribusi normal bivariate, Distribusi normal multivariat dan terakhir adalah pembahasan mengenai pengggunakan Metode GHK dalam model Multinomial Probit.

DISTRIBUSI NORMAL UNIVARIAT

Distribusi Normal (Gaussian) yang dinotasikan dengan N(μ ,σ^2) mempunyai mean μ dan variansi σ^2 mempunyai densitas

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{dengan } \mu,\ \sigma\in R \quad (1)$$

Selanjutnya, jika variabel random Z berdistribusi normal standar (mean μ = 0 dan variansi σ^2 = 1). Fungsi distribusi $F(z_0)$ yang dinotasikan dengan $\Phi(z_0)$ adalah

$$\Phi(z_0)=P(Z\le z_0)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z_0}\exp\left[-t^2/2\right]dt \quad (2)$$

Untuk mendapatkan nilai $\Phi(z_0)$ untuk suatu nilai z_0, memerlukan tranformasi deret Taylor

$$\exp(-z^2/2)=\sum_{i=1}^{\infty}(-1)^i\frac{z^{2i}}{2^i i!}$$

sehingga

$$\Phi(z)=F_Z(z)=F_Z(0)+\frac{1}{\sqrt{2\pi}}\int_0^z\exp\left[-t^2/2\right]dt$$

$$=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\sum_{i=0}^{\infty}(-1)^i\frac{z^{2i+1}}{(2i+i)2^i i!} \quad (3)$$

(Paolella, 2006). Invers fungsi $\Phi(z_0)$, yaitu $\Phi^{-1}(p)$, $0 < p < 1$, juga dapat dihitung.

DISTRIBUSI NORMAL BIVARIATE

Variabel random (X,Y) yang berdistribusi normal bivariate, dengan

Var(X) = Var(Y) = 1 dan Korr(X,Y) = ρ,

fungsi distribusinya dapat dituliskan dalam bentuk

$$\Phi(b,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{b_1}\int_{-\infty}^{b_2}\exp\left[-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}\right]dydx \qquad (4)$$

Tersa dan Wellan, (1990) menghitung probabilitas normal bivariate dengan mendefinisikan probabilitas dalam bentuk

$$L(h,k,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{h}^{\infty}\int_{k}^{\infty}\exp\left[-\frac{(x^2-2\rho xy+y^2)}{2(1-\rho^2)}\right]dydx \qquad (5)$$

dimana $\Phi(\mathbf{b},\rho) = L(-b_1,-b_2,\rho)$.

PROBABILITAS NORMAL TRIVARIATE

Distribusi normal trivariat dapat dituliskan sebagai

$$\Phi(b,R)=\frac{1}{(2\pi)^{3/2}\,|R|^{1/2}}\int_{-\infty}^{b_1}\int_{-\infty}^{b_2}\int_{-\infty}^{b_3}\exp\left[-\frac{x^T R^{-1} x}{2}\right]dx_1dx_2dx_3 \qquad (6)$$

dimana $b=(b_1,b_2,b_3)$ dan $R = (\rho_{ij})$ adalah matrik korelasi. Genz (1992) menyajikan distribusi normal Trivariat berdasarkan distribusi normal bivariat dalam bentuk

$$\Phi(b,R)=\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{b_1}\exp\left[-\frac{x^2}{2}\right]F(x)dx \qquad (7)$$

dimana

$$F(x)=\Phi\left(\left(\frac{b_2-\rho_{21}x}{(1-\rho_{21}^2)^{1/2}},\frac{b_3-\rho_{31}x}{(1-\rho_{31}^2)^{1/2}}\right),\frac{\rho_{32}-\rho_{31}\rho_{21}}{(1-\rho_{21}^2)^{1/2}(1-\rho_{31}^2)^{1/2}}\right)$$

Terdapat dua metode yang digunakan untuk menghitung F(.) pada persamaan (7), yaitu pertama adalah menggunakan transformasi $x=\Phi^{-1}(t)$, sehingga

$$\Phi(b,R)=\int_{0}^{\Phi(b_1)}F(\Phi^{-1}(t)dt \qquad (8)$$

dan selanjutnya itegrasi numerik menggunakan aturan Gauss-Legendre pada interval $[0,\Phi(b_1)]$. Yang kedua adalah menggunakan metode Drezner (1992), yang menggunakan modifikasi aturan Gauss-Hermite. Aturan tersebut sangat cocok untuk integral dalam bentuk

$$\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\exp(-\frac{x^2}{2})\,f(x)dx$$

sehingga persamaan (7) ditranformasi kedalam interval [0,∞]. Misal $y=x-b_1$ dan selanjutnya

$$\Phi(b,R) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{0} \exp\left[-\frac{(y+b_1)^2}{2}\right] F(y+b_1)dy$$

$$= \frac{\exp(-b_1^2/2)}{(2\pi)^{1/2}} \int_{-\infty}^{0} \exp\left[-\frac{y^2/2 - yb_1}{2}\right] F(y+b_1)dy \qquad (9)$$

jika $z=-y$ maka

$$\Phi(b,R) = \frac{\exp(-b_1^2/2)}{(2\pi)^{1/2}} \int_{0}^{\infty} \exp\left[-z^2/2 + zb_1\right] F(b_1 - z)dz \qquad (10)$$

Kedua metode sangat akurat untuk menghitung integral pada daerah terbatas (Genzs,1993).

Rumus untuk menghitung nilai probabilitas normal trivariat yang lain didasarkan pada rumus Plackett, yang didasarkan atas derivarif parsial

$$\frac{\partial \Phi(b,R)}{\partial \rho_{21}} = \frac{\exp(-f_3(\rho_{21})/2)}{2\pi\sqrt{1-\rho_{21}^2}} \Phi(u_3(\rho_{21}))$$

dimana

$$f_3(r) = \frac{b_1^2 + b_2^2 - 2rb_1b_2}{(1-r^2)} \text{ dan } u_3(r) = \frac{b_3(1-r^2) - b_1(\rho_{31} - r\rho_{32}) - b_2(\rho_{32} - r\rho_{31})}{((1-r^2)(1-r^2-\rho_{31}^2-\rho_{32}^2+2r\rho_{31}\rho_{32})^{1/2}}$$

Terdapat dua metode Plackett, yaitu yang pertama menggunakan

$$\rho_{21}^* = \rho_{31}\rho_{32} \pm \sqrt{(1-\rho_{31}^2)(1-\rho_{32}^2)} \text{ dan } R^* = \begin{bmatrix} 1 & \rho_{21}^* & \rho_{31} \\ \rho_{21}^* & 1 & \rho_{32} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$

Metode Plackett yang ke dua adalah

$$\Phi(b,R) = \Phi(b,R') + \frac{1}{2\pi}\int_0^1 \left(\rho_{21} \frac{\exp(-f_3(\rho_{21}t)/2}{\sqrt{1-\rho_{21}^2 t^2}} \Phi(\hat{u}_3(t)) + \rho_{31} \frac{\exp(-f_2(\rho_{31}t)/2}{\sqrt{1-\rho_{31}^2 t^2}} \Phi(\hat{u}_2(t)) \right) dt \quad (11)$$

dimana $f_2(r) = \frac{b_1^2 + b_3^2 - 2rb_1b_3}{(1-r^2)}$ dan $R' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho_{32} \\ 0 & \rho_{32} & 1 \end{bmatrix}$

$$\hat{u}_2(t) = \frac{b_3(1-\rho_{31}^2 t^2) - b_1 t(\rho_{21} - \rho_{31}\rho_{32}) - b_3(\rho_{32} - t^2\rho_{21}\rho_{31})}{((1-\rho_{31}^2 t^2)(1-\rho_{31}^2 t^2 - \rho_{21}^2 t^2 - \rho_{32}^2 + 2t^2\rho_{31}\rho_{21}\rho_{32})^{1/2}}$$

$$\hat{u}_3(t) = \frac{b_3(1-\rho_{21}^2 t^2) - b_1 t(\rho_{31} - \rho_{21}\rho_{32}) - b_2(\rho_{32} - t^2\rho_{31}\rho_{21})}{((1-\rho_{21}^2 t^2)(1-\rho_{21}^2 t^2 - \rho_{31}^2 t^2 - \rho_{32}^2 + 2t^2\rho_{31}\rho_{21}\rho_{32})^{1/2}}$$

Dalam kasus ini, $\Phi(b,R^1)$ dihitung menggunakan $\Phi(b_1)\Phi((b_2,b_3),\rho_{32})$. Metode ke dua ini telah diimplementasikan oleh Drezner (1994).

DISTRIBUSI MULTIVARIATE NORMAL

Fungsi densitas multivariate normal dapat disajikan dalam bentuk

$$F(a,b) = \frac{1}{\sqrt{|\Sigma|(2\pi)^m}} \int_{a_1}^{b_1}\int_{a_2}^{b_2}....\int_{a_m}^{b_m} \exp\left(-\frac{1}{2}\theta^t\Sigma^{-1}\theta\right) d\theta \qquad (12)$$

dimana $\theta = (\theta_1, \theta_2, ..., \theta_m)^t$ dan Σ adalah matrik covarian mxm yang simetrik positif definet.

Pada korelasi konstan, $\sigma_{ij} = \rho$, untuk $i \neq j$ dan $\sigma_{ii} = 1$, nilai F(.) dapat dihitung dengan akurat menggunakan normal independen. (Tong, 90).

$$F(b) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left[-\frac{t^2}{2}\right] \prod_{i=1}^{m} \Phi\left(\frac{b_i + \sqrt{\rho}t)}{\sqrt{1-\rho}}\right) dt \qquad (13)$$

Secara umum, untuk mengitung integral rangkap, fungsi densitasnya ditranformasi menggunakan faktor cholesky, yaitu $\theta = Cy$ dimana $CC^t = \Sigma$.

Sehingga $\theta^t\Sigma^{-1}\theta = y^tC^tC^{-t}C^{-1}Cy = y^ty$ dan $d\theta = |C|dy = |\Sigma|^{1/2}dy$. Karena $a \leq \theta = Cy \leq b$ maka

$$\frac{(a_i - \sum_{j=1}^{i-1} c_{ij}y_j)}{c_{ii}} \leq y_i \leq \frac{(b_i - \sum_{j=1}^{i-1} c_{ij}y_j)}{c_{ii}} \quad \text{untuk } i=1,2,..,m$$

Oleh karena itu

$$F(a,b) = \frac{1}{\sqrt{|\Sigma|(2\pi)^m}} \int_{a_1}^{b'_1} \exp(-\frac{y_1^2}{2}) \int_{a'_2(y_1)}^{b'_2(y_1)} \exp(-\frac{y_2^2}{2}) .... \int_{a'_m(y_1,..,y_{m-1})}^{b'_m(y_1,..,y_{m-1})} \exp(-\frac{y_1^2}{2}) dy \qquad (14)$$

dengan $a_i^t(y_1,...,y_{i-1}) = (a_i - \sum_{j=1}^{i-1} c_{ij}y_j)/c_{ii}$ dan $b_i^t(y_1,...,y_{i-1}) = (b_i - \sum_{j=1}^{i-1} c_{ij}y_j)/c_{ii}$

Untuk masing-masing y_i dapat ditransformasi menggunakan $y_i = \Phi^{-1}(z_i)$, dengan

$$\Phi(y) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{y} \exp\left[-\frac{\theta^2}{2}\right] d\theta$$

yang merupakan fungsi distribusi normal standar univariat. setelah transformasi itu, F(a,b) menjadi

$$F(a,b) = \int_{d_1}^{e_1} \int_{d_2(z_1)}^{e'_2(z_1)} \dots \int_{d_m(z_1,..,z_{m-1})}^{e_m(z_1,..,z_{m-1})} dz \tag{15}$$

dengan

$$d_i(z_1,...,z_{i-1}) = \Phi\left([a_i - \sum_{i=1}^{i-1} c_{ij}\Phi^{-1}(z_j)] / c_{ii} \right) \text{dan} \quad e_i(z_1,...,z_{i-1}) = \Phi\left([b_i - \sum_{i=1}^{i-1} c_{ij}\Phi^{-1}(z_j)] / c_{ii} \right)$$

Bentuk integran ini lebih sederhana dibanding integran aslinya. Integrasi wilayah (region) adalah lebih complek, sehingga tidak bisa ditangani secara langsung menggunakan algoritma numerik standar untuk integral rangkap. Penyelesaian masalah ini menggunakan tranformasi $z_i = d_i + w_i (e_i - d_i)$. Setelah tranformasi akhir ini

$$F(a,b) = (e_1 - d_1)\int_0^1 (e_2 - d_2) \dots \dots \int_0^1 (e_m - d_m) \int_0^1 dw \tag{16}$$

dengan $d_i = \Phi\left([a_i - \sum_{i=1}^{i-1} c_{ij}\Phi^{-1}(d_j + w_j(e_j - d_j))] / c_{ii} \right)$ dan

$$e_i = \Phi\left((b_i - \sum_{i=1}^{i-1} c_{ij}\Phi^{-1}(d_j + w_j(e_j - d_j))] / c_{ii} \right)$$

Integrasi pertama terhadap variabel w_1 , w_2 ...dst.

$$F(a,b) = (e_1 - d_1)\int_0^1 (e_m - d_m) \dots \dots \int_0^1 (e_2 - d_2) \int_0^1 dw_1 \dots . dw_m \tag{17}$$

Integral ini dapat dinyatakan dalam algoritma sebagai berikut

1. Input Σ, α, ε,a,b dan N_{max}
2. menghitung faktor Cholesky bagian bawah, C untuk Σ.
3. nilai awal Itsum=0, N=0, Varsum=0, $d_1 = \Phi(a_1/c_{1,1})$, $e_1 = \Phi(b_1/c_{1,1})$ dan $f_1=e_1-d_1$.
4. Repeat
   a. Membangkitkan variabel random uniform $w_1,w_2,...,w_{m-1} \in [0,1]$
   b. untuk i=2,3,....,m

      $y_{i-1} = \Phi^{-1}(d_{i-1} + w_{i-1}(e_{i-1}-d_{i-1}))$; $d_i = \Phi\left((a_i - \sum_{i=1}^{i-1} c_{i,j} y_j) / c_{ii} \right)$

      $e_i = \Phi\left((b_i - \sum_{i=1}^{i-1} c_{i,j} y_j) / c_{ii} \right)$ dan $f_i = (e_i-d_i)f_{i-1}$

c. N=N+1; δ=(f_m – Itsum)/N; Itsum=Itsum+δ; Varsum=(N-2)Varsum/N + δ^2
   dan Error = $\alpha\sqrt{Varsum}$

Until Error < ε atau N = N_{max}

5. Output : Itsum, Error,dan N

Parameter input α adalah faktor confidensi montecarlo untuk error standar. Misalnya untuk α = 2,5, kita mengharapkan error aktual dalah F menjadi lebih kecil dari Error 99%. N_{max} adalah parameter input yang merupakan total jumlah perhitungan.

Untuk a_i = - ∞ atau b_i = ∞ maka d_i = 0 atau e_i =1.

Dalah beberapa aplikasi, kadang kita menghitung integral

$$G9g) = \frac{1}{\sqrt{|\Sigma|(2\pi)^m}}\int_{a_1}^{b_1}\int_{a_2}^{b_2}....\int_{a_m}^{b_m}\exp\left(-\frac{1}{2}\theta^t\Sigma^{-1}\theta\right)g(\theta)d\theta \tag{18}$$

Untuk menghitung integral ini memerlukan modeifikasi dari algoritma di atas. Karena g(θ) yang tergantung pada θ_m dan w_m. Sehingga perlu menghitung θ = Cy dan menghitung f_m(g(θ)) dalam langkah 4c.

MENGHITUNG NILAI PROBABILITAS DALAM PROGRAM R

Terdapat beberapa software yang cukup reliabel dan efisien untuk kasus m=1 dan m=2 [Drezner dkk, 1990] dan [Cox dkk, 1991]. Algoritma Schervish [Schervish , 1984] telah diimplementasikan untuk m kurang dari 8. Algoritma ini menggunakan integrasi secara *locally adaptive numerical* yang didasarkan pada *Simpson's rule*. Algoritma ini membutuhkan waktu yang lama untuk m lebih dari 4 (Genz, 1993). Pendekatan lain adalah dengan metode monte Carlo atau metode subregion adaptive. Metode ini cukup reliabel dan efisien untuk menghitung F(.). [Genz, A., 1992]. Genz (1993) telah melakukan perbandingan beberapa metode untuk menghitung nilai probabilitas normal multivariat. Metode yang dibandingkan adalah *acceptance-rejection sampling*, metode Deák, metode Genz dan metode Schervish. Disimpulkan bahwa metode Genz adalah yang paling efisien.

Pada distribusi normal, untuk menghitung fungsi distribusi maupun fungsi densitas dalam program R, dapat digunakan perintah

>*pnorm(x, μ, σ)* : untuk menghitung nilai fungsi densitas normal

>dnorm(x, μ, σ) : untuk menghitung nilai fungsi distribusi normal

Dalam program R juga terdapat beberapa paket yang dapat digunakan untuk menghitung nilai probabilitas multivariat. Perhitungan probabilitas yang didasarkan pada algoritma Genz (1992, 1993) dapat di akses dalam "*fmultivar Package*", "*mnormt Package*", "*mvtnorm Package*", "*adapt package* ".Paket lain yang didasarkan pada metode Schervish's dan metode "Joe" dapat dperoleh pada "*mprobit Package*". (R Core Team, 2007)

SIMULASI PROBABILITAS DENGAN METODE GHK DALAM MODEL PROBIT

Dalam pemodelan yang didasarkan pada Random Utility, seorang pembuat keputusan dinotasikan dengan i, yang berhadapan dengan pilihan sebanyak J anternatif. Pembuat keputusan mempunyai tingkat utiliti (keuntungan) untuk setiap alternatif. Misalkan U_{ij} untuk j=1,…,J adalah utiliti pembuat keputusan i jika memilih alternatif j. Nilai U_{ij} yang sesungguhya tidak diketahui oleh pengamat (peneliti). Pembuat keputusan memilih alternatif yang mempunyai utiliti terbesar, sehingga memilih alternatif k jika dan hanya jika $U_{ik} > U_{ij}\ \forall j \neq k$.

Peneliti tidak mengetahui nilai utiliti untuk pembuat keputusan pada masing-masing alternatif. Peneliti hanya mengamati atribut yang ada untuk masing-masing alternatifnya, dan atribut pembuat keputusan yang dinotasikan dengan x_{ij}. Secara fungsi dapat dinotasikan sebagai $V_{ij}= V(x_{ij})$, $\forall j$ yang biasa dinamakan *representative utility*.

Utiliti dapat dipisahkan menjadi komponen terobservasi (V_{ij}) dan komponen tidak terobservasi (ε_{ij}),

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad \forall j,i$$

Probabilitas pembuat keputusan i memilih alternatif k adalah

$$\begin{aligned} p_{ik} &= P(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) \quad \forall j\neq k \\ &= \int I(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij})\phi(\varepsilon_i)d\varepsilon_i \quad \forall j\neq k \\ &= \int I((V_{ij} - V_{ik}) - (\varepsilon_{ij} - \varepsilon_{ik}) < 0)\phi(\varepsilon_i)d\varepsilon_i \quad \forall j\neq k \end{aligned} \tag{19}$$

dengan I(.) merupakan fungsi indikator dan integral terhadap semua nilai ε_i.

Model multinomial probit disusun dari asumsi bahwa vektor $\varepsilon_i^t = (\varepsilon_{i1}, \ldots, \varepsilon_{1J})$ berdistribusi multivariat normal dengan mean nol dan matrik kovariansi Σ. Densitas untuk ε_i adalah

$$\phi(\varepsilon_i) = \frac{1}{(2\pi)^{J/2} \, |\Sigma|^{1/2}} \exp[-\frac{1}{2}\varepsilon_i^t \Sigma^{-1} \varepsilon_i \,] \tag{20}$$

Selanjutnya dapat dilakukan transformasi utiliti ke selisih setiap utiliti terhadap alternatif k,

$$\tilde{U}_{ijk} = U_{ij} - U_{ik}, \tilde{V}_{ijk} = V_{ij} - V_{ik} \text{ dan } \tilde{\varepsilon}_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}.$$

Sehingga persamaan utilitinya menjadi

$$\tilde{U}_{ijk} = \tilde{V}_{ijk} + \tilde{\varepsilon}_{ijk} \text{ untuk j}\neq\text{k ; } \tilde{\varepsilon}_{ik}^t = (\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{iJ})$$

dimana tanda " . . ." adalah simbol semua kecuali k, sehingga $\tilde{\varepsilon}_{ik}$ adalah matrik (J-1)x1 dan $\tilde{\varepsilon}_{ik}$~N(0, $\tilde{\Sigma}_k$). Matrik kovariansi $\tilde{\Sigma}_k$ dapat diturunkan dari Σ dan dapat dicari faktor Choleski, L_k sedemikian hingga $L_k L_k^t = \tilde{\Sigma}_k$.

$$L_k = \begin{pmatrix} c_{11} & 0 & 0 & 0 & 0 & \ldots\ldots & 0 \\ c_{21} & c_{22} & 0 & 0 & 0 & \ldots\ldots. & 0 \\ c_{31} & c_{32} & c_{33} & 0 & 0 & \ldots\ldots. & 0 \\ \ldots. & \ldots & \ldots & \ldots & \ldots & \ldots\ldots. & \\ c_{J1} & c_{J2} & c_{J3} & \ldots & \ldots & c_{(J-1)1} & 0 \end{pmatrix} \text{ dimana}$$

Persamaan utiliti dapat disusun sebagai

$$\tilde{U}_{ik} = \tilde{V}_{ik} + L_k \eta_i$$

Dimana $\eta^t_i = (\eta_{1i}, \ldots, \eta_{(J-1)i})$ adalah vektor yang berdistribusi normal standard independen, $\eta_{ji} \sim N(0, 1)$ $\forall j$., $\tilde{U}_{ik}^t = (\tilde{U}_{i1k}, \ldots, \tilde{U}_{iJk})$ dan $\tilde{V}_{ik}^t = (\tilde{V}_{i1k}, \ldots, \tilde{V}_{iJk})$

Model dapat dituliskan secara eksplisit

$$\tilde{U}_{i1k} = \tilde{V}_{i1k} + c_{11}\eta_{1i},$$

$$\tilde{U}_{i2k} = \tilde{V}_{i2k} + c_{21}\eta_{1i} + c_{21}\eta_{2i}$$

$$\tilde{U}_{i3k} = \tilde{V}_{i3k} + c_{31}\eta_{1i} + c_{32}\eta_{2i} + c_{33}\eta_{3i}$$

....

$$\tilde{U}_{iJk} = \tilde{V}_{iJk} + c_{J1}\eta_{1i} + c_{J2}\eta_{2i} + c_{J3}\eta_{3i} + \ldots + c_{JJ}\eta_{Ji}$$

Probabilitas individu i memilih alternatif k adalah

$$
\begin{aligned}
p_{ik} &= \mathrm{P}(\tilde{U}_{ijk} < 0) \quad \forall \mathrm{j} \neq k \\
&= \mathrm{P}(\tilde{U}_{i1k} < 0, \tilde{U}_{i2k} < 0, \ldots, \tilde{U}_{iJk} < 0) \\
&= \mathrm{P}\left(\eta_{1\mathrm{i}} < \frac{-\tilde{V}_{i1k}}{c_{11}}\right) . \mathrm{P}\left(\eta_{2\mathrm{i}} < \frac{-(\tilde{V}_{i2k} + c_{21}\eta_{1i})}{c_{22}} \middle| \eta_{1i} < \frac{-\tilde{V}_{i1k}}{c_{11}}\right) . \\
&\quad \mathrm{P}\left(\eta_{3\mathrm{i}} < \frac{-(\tilde{V}_{i3k} + c_{31}\eta_{1i} + c_{32}\eta_{2i})}{c_{33}} \middle| \eta_{1i} < \frac{-\tilde{V}_{i1k}}{c_{11}} \text{ dan } \eta_{2\mathrm{i}} < \frac{-(\tilde{V}_{i2k} + c_{21}\eta_{1i})}{c_{22}}\right) \ldots \\
&\quad \mathrm{P}\left(\eta_{\mathrm{Ji}} < \frac{-(\tilde{V}_{iJk} + c_{J1}\eta_{1i} + \ldots + c_{J(J-1)}\eta_{(J-1)i})}{c_{33}} \middle| \eta_{1i} < \frac{-\tilde{V}_{i1k}}{c_{11}}, \ldots, \right. \\
&\quad \left. \ldots, \eta_{(\mathrm{J}-1)\mathrm{i}} < \frac{-(\tilde{V}_{i(J-1)k} + c_{(J-1)1}\eta_{1i} + \ldots + c_{(J-1)(J-2)}\eta_{(J-2)i})}{c_{(J-1)(J-1)}}\right) \qquad (21)
\end{aligned}
$$

GHK simulator dapat dihitung menggunakan algoritma berikut ini, (Train, 2003)

1. menghitung

$$\mathrm{P}\left(\eta_1 < \frac{-\tilde{V}_{i1k}}{c_{11}}\right) = \Phi\left(\frac{-\tilde{V}_{i1k}}{c_{11}}\right)$$

2. mengambil sebuah nilai η_1, yang diberi label $\eta^r{}_1$ dari distribusi normal terpotong pada $\Phi\left(-\frac{\tilde{V}_{1ki}}{c_{11}}\right)$. Pengambilan dapat dilakukan sebagai berikut :

   (a) mengambil dari distribusi standard uniform $\mu^r{}_1$.

   (b) menghitung $\eta^r{}_1 = \Phi^{-1}\left(\mu_1^r \Phi\left(-\frac{\tilde{V}_{1ki}}{c_{11}}\right)\right)$

3. menghitung :

$$\mathrm{P}\left(\eta_2 < \frac{-(\tilde{V}_{i2k} + c_{21}\eta_1)}{c_{22}} \middle| \eta_1 = \eta_1^r\right) = \Phi\left(\frac{-(\tilde{V}_{i2k} + c_{21}\eta_1^r)}{c_{22}}\right)$$

   Mengambil sebuah nilai η_2, diberi label $\eta^r{}_2$ dari normal standart terpotong $-\frac{(\tilde{V}_{i2k} + c_{21}\eta_1^r)}{c_{22}}$. Pengambilan ini dapat dilakukan sebagai berikut :

   (a) mengambil dari standard uniform $\mu^r{}_2$.

(b) menghitung $\eta^r 2 = \Phi^{-1}\left(\mu_2^r \Phi\left(-\frac{(\tilde{V}_{i2k} + c_{21}\eta_1^r)}{c_{22}}\right)\right)$

5. menghitung

$$\mathrm{P}\left(\eta_3 < \frac{-(\tilde{V}_{i3k} + c_{31}\eta_1 + c_{32}\eta_2)}{c_{33}} \middle| \eta_1 = \eta_1^r, \eta_2 = \eta_2^r\right) = \Phi\left(\frac{-(\tilde{V}_{i3k} + c_{31}\eta_1^r + c_{32}\eta_2^r)}{c_{33}}\right)$$

6. dan seterusnya untuk semua alternatif kecuali k..

7. probabilitas simulasi untuk pengambilan ke-r dari η_1, η_2, . . . adalah dihitung sebagai

$$\breve{\mathrm{p}}_{\mathrm{ik}}^{\mathrm{r}} = \Phi\left(\frac{-\tilde{V}_{i1k}}{c_{11}}\right).\Phi\left(\frac{-(\tilde{V}_{i2k} + c_{21}\eta_1^r)}{c_{22}}\right)\Phi\left(\frac{-(\tilde{V}_{i3k} + c_{31}\eta_1^r + c_{32}\eta_2^r)}{c_{33}}\right)....$$

$$....\Phi\left(\frac{-(\tilde{V}_{iJk} + c_{J1}\eta_1^r + c_{J2}\eta_2^r + ... + c_{J(J-1)}\eta_{(J-1)}^r)}{c_{JJ}}\right)$$

8. mengulangi langkah 1-7 beberapa kali, untuk $r = 1, \ldots, R$.

9. probabilitas simulasinya adalah

$$\breve{p}_{ik} = \frac{1}{R}\sum_{r=1}^{R} \breve{p}_{ik}^r$$

Algoritma ini sama dengan Algoritma Genz, yaitu dengan batas bawahnya adalah – Inf.

$$p_{ik} = \mathrm{P}(\tilde{U}_{ijk} < 0) = \mathrm{P}(-\infty < \tilde{U}_{ijk} < 0) \quad \forall \mathrm{j} \neq k$$

$$= \mathrm{P}(-\infty < \tilde{U}_{i1k} < 0, -\infty < \tilde{U}_{i2k} < 0, ..., -\infty < \tilde{U}_{iJk} < 0)$$

$$= \mathrm{P}(-\infty < \tilde{\varepsilon}_{i1k} < -\tilde{V}_{i1k}, -\infty < \tilde{\varepsilon}_{i2k} < -\tilde{V}_{i2k}, ..., -\infty < \tilde{\varepsilon}_{iJk} < -\tilde{V}_{iJk})$$

sehingga dalam algoritma Genz, diambil nilai $d_i = 0$ dan $f_i = e_i f_{i-1}$. Sehingga nilai Itsum sama dengan nilai $\breve{p}_{ik}$.

Misalkan Pengamatan dilakukan dengan mengambil model untuk tiga alternatif dengan memasukkan variabel atribut pembuat keputusan (X_i) dan variabel atribut masing masing alternatif (Z_{ij}). Variabel X_i biasa disebut variabel sosio ekonomik/geografi, misalnya penghasilan, jenis kelamin, asal daerah, jumlah anak. Sedangkan variabel Z_{ij} misalkan untuk pilihan penggunaan alat tranportasi (Bus, mobil pribadi, sepeda motor) maka Z_{ij} dapat berupa waktu tempuh, biaya. Model utilitinya adalah

$U_{ij} = X_i\beta_j + Z_{ij}\gamma + \varepsilon_{ij}$ untuk i=1,2,...,n dan j=1,2,3.

$U_{i1} = \beta_{01} + X_i\beta_1 + Z_{i1}\gamma + \varepsilon_{i1}$, $U_{i2} = \beta_{02} + X_i\beta_2 + Z_{i2}\gamma + \varepsilon_{i2}$

$U_{i3} = \beta_{03} + X_i\beta_3 + Z_{i3}\gamma + \varepsilon_{i3}$

Dengan mengambil alternatif pertama sebagai *base line*, maka model terestimasinya yang merupakan tranformasi selisih U_{ij}-U_{i1} adalah

$U^*_{i21} = (\beta_{02} - \beta_{01}) + X_i(\beta_2-\beta_1) + (Z_{i2} - Z_{i1})\gamma + (\varepsilon_{i2}-\varepsilon_{i1})$

$U^*_{i31} = (\beta_{03} - \beta_{01}) + X_i(\beta_3-\beta_1) + (Z_{i1} - Z_{i1})\gamma + (\varepsilon_{i3}-\varepsilon_{i1})$

$U^*_{i11} = 0$

Dengan matrik kovariansi (cov($\varepsilon_{i1},\varepsilon_{i2}, \varepsilon_{i3}$)) sama dengan :

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

Algoritma GHK tersebut dapat dituliskan dalam program R sebagai berikut :

```
#Program mengitung nilai probailitas dengan metode GHK

GHK<-function(b)
{
b01=b[1]; b02=b[2]; b03=b[3]; b1=b[4]; b2=b[5]; b3=b[6]
g=b[7]; s11=b[8];s12=b[9];s13=b[10];s22=b[11]
s23=b[12];s33=b[13]
V21=(b02-b01) + ( b2-b1)*X+g*(Z[,2]-Z[,1])
V31=(b03-b01) + (b3-b1)*X+g*(Z[,3]-Z[,1])
V32=V31-V21
S=matrix(c(s11,s12,s13,s12,s22,s23,s13,s23,s33),3,3)
M1= matrix(c(-1,-1,1,0,0,1),2,3)
M2=matrix(c(1,0,-1,-1,0,1),2,3)
S1= M2%*%S%*%t(M2)
S2=M2%*%S%*%t(M2)
F1=matrix(,N,1); F2= matrix(,N,1);F3= matrix(,N,1)
Prob=matrix(,N,3)
      for (i  in 1:N)
      {
      F1= sadmvn(-Inf, c(-V21[i],-V31[i]), mean=rep(0,2),S1)
      F2= sadmvn(-Inf, c(V21[i],-V32[i]), mean=rep(0,2),S2)
      F3= 1 – F1-F2
      Prob[i,]=c(F1,F2,F3)
      }
Prob
}
```

Misalkan diambil data dengan N= 500, X_i ~ NID(0,1) , Z_{ij} ~ NID(0,1) dan ε_i ~ N(**0,Σ**)

$$\Sigma = \begin{pmatrix} 1 & 0.2 & 0.7 \\ 0.2 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{pmatrix}$$

parameter β_{01} =2 , β_{02} =1, β_{03} =0.5, β_1=-2, β_2=-1, β_3 =-1 dan γ=0.8, maka nilai probabilitas masing-masing alternatif(pilihan) pada masing-masing individu dapat dihitung sebagai berikut:

```
> library(mnormt)
>X=rnorm(N,3); prop.pilihan= matrix(,N,3); Z=matrix(,N,3); V=matrix(,N,3)
>for (i in 1:3){Z[,i]=rnorm(N)}
>prop.pilihan=GHK(c(2,1,0.5,-2,-1,-1,0.8,1,0.2,0.7,1,0.7,1))
```

KESIMPULAN

Seiring dengan kemajuan teknologi komputasi, saat ini memungkinkan menghitung nilai probabilitas dari variabel random yang berdistribusi normal multivariate. Metode yang banyak diaplikasikan dalam program R adalah Metode Genz yang disusun berdasarkan simulasi Monte Carlo. Aplikasi metode ini dalam bidang ekonometrika, adalah untuk simulasi nilai probabilitas dikenal dengan nama metode GHK.

Dengan adanya simulasi probabilitas (GHK) ini, maka dengan program R adalah sangat memungkinkan untuk menyusun model multinomial probit.

DAFTAR PUSTAKA

Contoyannis P, 2001, An Introduction to Simulation-Based Estimation, *Working Paper* Department of Economics and Related Studies, University of York.

Cox, D.R. dan Wermuth, N. (1991), A Simple Approximationfor Bivariateand Normal Integrals, International Statistics Review 59, pp. 263-269

Daganzo, C., F. Bouthelier, and Y. Sheffi (1977), ‘Multinomial probit and qualitative choice: A computationally efficient algorithm’, *Transportation Science* 11, 338–358.

Drezner, Z. (1992), Computation of Multivariate Normal Integral, ACM Transactions on Mathematics Software 18, pp. 450-460

Drezner, Z. (1994), Computation of the Trivariate Normal Integral, Mathematics of Computation 62, pp. 289-294

Drezner, Z. dan Wesolowsky G.O., (1989), On the Computation of Bivariate Normal Integral, Journal of Statist. Comput. Simul. 35. pp. 101-107

Genz, A. (1992). Numerical Computation of Multivariate Normal Probabilities. *J. Computational and Graphical Statist.*, 1, 141-149.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, 25, 400-405.

Geweke, J. (1996), 'Monte Carlo simulation and numerical integration', in D. Kendrick and J. Rust, eds., *Handbook of Computational Economics*, Elsevier Science, Amsterdam, pp. 731–800.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996), 'Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results',*Journal of Econometrics* 72, 85–134.

Horowitz, J., J. Sparmann, and C. Daganzo (1982), 'An investigation of the accuracy of the Clark approximation for the multinomial probit model', *Transportation Science* 16, 382–401

Paolella, Marc S. (2006), Fundamental Probability (A Computational Approach), John Wiley & Sons Ltd,

Schervish, M.J. (1984). Multivariate normal probabilities with error bound. Appl. Statist., 33, 81-94.

Tong, Y.L. (190), The Multivariate Normal Distribution, Springer-Verlag, New York

Train, Kenneth (2003), *Discrete Choice Methods with Simulation*, UK Press, Cambridge.