Parameter Estimation of Additive Genetic and Unique Environment (AE) Model on Diabetes Mellitus Type 2 using Bayesian Method

Dewi Retno Sari Saputro¹, Andi Darmawan², and Purnami Widyaningsih³,
¹Department of Mathematics, Universitas Sebelas Maret, Surakarta, 57126, Indonesia
²Department of Mathematics, Universitas Sebelas Maret, Surakarta, 57126, Indonesia
³Department of Mathematics, Universitas Sebelas Maret, Surakarta, 57126, Indonesia
dewiretnoss@staff.uns.ac.id

Abstract: Diabetes mellitus (DM) is a chronic disease occurred in human when the pancreas cannot produce enough insulin hormones or the body uses it ineffectively. The DM is divided into two types, namely DM type 1 and DM type 2. Diabetes type 2 is found more common in patients. The disease with genetic (A) and lifestyle (E) factors can be constructed with additive genetic and unique environment (AE) model. The aims of this research are to estimate the parameter of AE model using bayesian method and to do simulation of DM type 2 on parent-offspring. The simulation showed that the value of genetic variance is 0.3600 and the value of lifestyle variance is 0.0899. Therefore, the variance of genetic factor in DM type 2 is greater than lifestyle.

1. Introduction
Global Status Report by World Health Organization (WHO) on Non Communicable Disease (NCD) in 2010 reported that 60% the cause of dead in this world was NCD. One of the NCD is Diabetes Mellitus (DM). There are two types of DM [1], DM type 1 and type 2. Type 1 diabetes is primarily due to the autoimmune-mediated destruction of pancreatic beta cells leading to insulin deficiency or due to a lack of insulin secretion by beta cells in the pancreas[2][3]. Meanwhile, in type 2 diabetes, increased glucotoxicity, lipotoxicity, endoplasmic reticulum-induced stress, and apoptosis lead to the progressive loss of beta-cells so it caused decreasing sensitivity of target tissue to produce insulin[2][3]. The most one commonly found in patient is DM type 2. Ripsin et al[4]. stated that DM type 2 is due primarily to genetic and lifestyle factors. Environmental factors modulating genetic risk are proving to be important in the development of DM type 2 as shown by Murea et al [5]. According to Hu et al[6], a number of lifestyle factors are known to be important to the development of DM type 2. Obesity is one of the major causes of type 2 diabetes[7][8]. Zahtamal et al [9] showed that the age group of 45 years or older and have a family history of DM disease is the age group at risk of developing DM. According to the factors, a genetic caused DM can be defined as kinship coefficient. Neale et al [11], showed that parent-offspring have kinship coefficient 1/4. The kinship are affected by additive genetic component (A), dominance genetic component (D), common environment component (C), and unique environmet component (E) as shown by Rabe-Hesketh et al [12][13]. According to Betteng et al [13], genetic and environmental are important components that affect the disease. According to these two essential components, DM type 2 can be constructed by additive genetic and unique environment (AE) model.
Referring to Setiawan's research[10] and Betteng et al [13] in this article, the AE model and its simulation to the inheritance of DM type 2 from parent to the child using bayesian method were discussed.

2. Experimental Details

Bayesian method is a theory of probability condition that takes into account of the probability of a case depending on others as shown by Berger[14]. Congdon[15] stated that the method of bayesian inference is based on the posterior distribution \( f(\theta|x) \). According to Bayes theorem, the bayesian inference can be stated as \( f(\theta|x_1, x_2, ..., x_n) = \frac{f(x_1, x_2, ..., x_n | \theta) f(\theta)}{\int_{\Theta} f(x_1, x_2, ..., x_n | \theta) f(\theta) d(\theta)} \). Bayes estimator for \( \theta \) is obtained through the expected value \( \hat{\theta} = E(\theta|x_1, x_2, ..., x_n) \).

Based on Bain and Engelhardt[16], the joint probability density function for \( n \) random variables \( X_1, X_2, ..., X_n \) which is expressed in the form \( f(x_1, x_2, ..., x_n; \theta) \) is called likelihood function. The likelihood function is denoted as \( L(\theta) \).

In Backlawski research[17] it was stated that the bayesian method started with the determination of the prior distribution. Saputro et al[18] also stated that Bayesian method or known as Bayesian, the used parameters are random variables that have a particular distribution (prior distribution). Sample information is extracted from a likelihood function as shown by Bolstad[19]. Based on likelihood function in bayesian, if the \( \theta \) parameter was required as variable then \( \theta \) had a density function \( f(\theta) \).

According to Guure and Noor[20], the posterior distribution is conditional density function of \( \theta \) if known and observed values \( x \) can be defined as \( f(\theta|x) = \frac{f(\theta,x)}{f(x)} \). So that the posterior density function for a continuous random variable can be expressed as \( f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{\Theta} f(x|\theta)f(\theta)d(\theta)} \).

3. Results and Discussion

AE Model

AE model can be applied in DM type 2 disease. The model can be expressed as \( X = A + E \), with \( X, A, \) and \( E \) respectively are the blood glucose random variable, genetic factor, and lifestyle. An \( X \) random variabel is paired data \( (X_1, X_2) \) that can be expressed as \( X_1 = A_1 + E_1 \) and \( X_2 = A_2 + E_2 \) with \( X_1 \) is blood glucose random variable of parent and \( X_2 \) is for the child. The \( A \) and \( E \) factors are independent and the distribution is identical so the variance \( (X_1, X_2) \) is \( V(X_1, X_2) = (\sigma^2_{A_1} + \sigma^2_{E_1}, \sigma^2_{A_2} + \sigma^2_{E_2}) \) with \( \sigma^2_{A_1} \) and \( \sigma^2_{A_2} \) are the variance of each genetic \( A \) factor, \( \sigma^2_{E_1} \) and \( \sigma^2_{E_2} \) are the variance of each lifestyle \( E \).

Suppose \( (Y_{1i}, Y_{2i}) \) are variable that states the size of the trait (diseased or not) in two individuals, parents and children. Random variables \( (Y_{1i}, Y_{2i}) \) depend on \( (X_{1i}, X_{2i}) \) and the threshold \( (b) \) level which can be expressed as \( (Y_{1i}, Y_{2i}) = (0,0) \) if \( (X_{1i} \leq b, X_{2i} \leq b) \), \( (Y_{1i}, Y_{2i}) = (0,1) \) if \( (X_{1i} \leq b, X_{2i} > b) \), \( (Y_{1i}, Y_{2i}) = (1,0) \) if \( (X_{1i} > b, X_{2i} \leq b) \), and \( (Y_{1i}, Y_{2i}) = (1,1) \) if \( (X_{1i} > b, X_{2i} > b) \) for \( i = 1, ..., n \).

The value of \( b \) is the limit of blood sugar level to categorize an individual with diabetes mellitus type 2 or not. According to WHO the limit \( b \) for fasting blood glucose is 0.0126 g/dl. It was assumed that \((X_1, X_2)\) is normal distribution. The probability for \( Y_1 = 0 \) on condition \( A_1 \) and \( A_2 \) is \( P(Y_1 = 0|A_1, A_2) = \Phi \left( \frac{b - \mu_1}{\sigma^2_{E}} \right) \) and the probability for \( Y_1 = 1 \) on condition \( A_1 \) and \( A_2 \) is
\[ P(Y_1 = 1 | A_1, A_2) = 1 - \Phi \left( \frac{b-a}{\sqrt{\sigma^2}} \right) \]

The conditional probability of \( Y_2 \) can be determined in the same manner. The likelihood function for a sample of \( n \) observations pairs (parents and children) selected random is
\[
L(\theta) = \prod_{i=1}^{n} p(y_{1i}, y_{2i}, a_i) = \prod_{i=1}^{n} q(y_{1i}, a_i) q(y_{2i}, a_i) f_A(a_i)
\]

(3.1)

with \( a_i \) is sample random of genetic component and \( f_A(a_i) \) is a density function \( A \) which can be constructed as
\[
f_A(a_i) = \frac{1}{\sqrt{2\pi}\sigma_A^2} \exp \left[ -\frac{a_i^2}{2\sigma_A^2} \right]
\]

(3.2)

and
\[
q(y, a) = \Phi((b-a)/\sqrt{\theta_1}) [1 - \Phi((b-a)/\sqrt{\theta_2})] \text{ with } \theta_1 = \frac{1}{\sigma_A^2}
\]

Furthermore, with equations (3.2) and (3.1) is obtained \( L(\theta) = \prod_{i=1}^{n} q(y_{1i}, a_i) q(y_{2i}, a_i) f_A(a_i) \)
\[
= \left( \frac{1}{2\pi} \sigma_2^2 \right)^n \exp \left[ -\frac{1}{2} \sum_{i=1}^{n} a_i^2 \right] \prod_{i=1}^{n} q(y_{1i}, a_i) q(y_{2i}, a_i)
\]

with \( \theta_2 = \frac{1}{\sigma_A^2} \). The value of \( \sigma_A^2 \) and \( \sigma_B^2 \) were unknown so required to estimate the parameter.

In calculating the posterior density function can be selected prior conjugate for parameter \( \theta_1, \theta_2, \) and \( b \). Parameter \( \theta_1 \) and \( \theta_2 \) are selected from gamma distribution in order to obtain prior density function
\[
f_1(\theta_1) = \frac{1}{\theta_1 \beta_1} \theta_1^{a_1-1} \exp \left[ -\frac{\theta_1}{\beta_1} \right] \text{ and } f_2(\theta_2) = \frac{1}{\theta_2 \beta_2} \theta_2^{a_2-1} \exp \left[ -\frac{\theta_2}{\beta_2} \right] \text{.}
\]

The conjugate prior distribution for \( b \) is selected from the normal distribution family to obtain the prior density function
\[
f_3(b) = \frac{1}{\sqrt{2\pi}\beta_3^2} \exp \left[ -\frac{(b-a_3)^2}{2\beta_3^2} \right].
\]

In this case \( a_1, \beta_1, a_2, \beta_2, a_3, \beta_3 \) are parameters. Therefore the joint posterior density is
\[
f(\theta_1, \theta_2, b) \propto f_1(a_1-1) f_2(a_2+1) \exp \left[ -w_1 \right] f_3 (b)
\]

(3.3)

For \( w_1 = \frac{1}{\beta_1} \theta_1 + \left( \frac{b-a_3}{2\beta_3} \right) \theta_2 + \frac{(b-a_3)^2}{2\beta_3} \) and \( w_2 = \prod_{i=1}^{n} q(y_{1i}, a_i) q(y_{2i}, a_i) \).

Furthermore, parameter \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \) were estimated with posterior distribution (3.3) using bayesian method. Therefore, parameters parameter \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \) were estimated based the joint posterior distribution \( f(\theta_1, \theta_2, b) \) with \( \theta_1 = \frac{1}{\sigma_A^2} \) and \( \theta_2 = \frac{1}{\sigma_A^2} \).

4. Simulation

| \( n \) | (Parents, Children) | \( \sigma_A^2 \) | \( \sigma_B^2 \) |
| --- | --- | --- | --- |
| (0,0) | 13 | 9 | 17 | 11 | 0.3442 | 0.0508 |
| (0,1) | 26 | 39 | 18 | 17 | 0.3582 | 0.0824 |
| (1,0) | 67 | 47 | 42 | 44 | 0.3589 | 0.0825 |
| (1,1) | 90 | 75 | 80 | 55 | 0.3594 | 0.0830 |
| 50 | 140 | 86 | 92 | 82 | 0.3598 | 0.0877 |
| 100 | 187 | 124 | 101 | 88 | 0.3598 | 0.0895 |
| 200 | 213 | 154 | 131 | 102 | 0.3598 | 0.0897 |
| 300 | 269 | 181 | 146 | 104 | 0.3600 | 0.0898 |
| 500 | 351 | 185 | 154 | 110 | 0.3600 | 0.0899 |
| 700 | 382 | 241 | 158 | 119 | 0.3600 | 0.0899 |
| 900 | 427 | 279 | 160 | 134 | 0.3600 | 0.0899 |
| 1000 | 0.3600 | 0.0899 |
The data of DM type 2 from parents to children were generated with different n sample. The data were assumed to

\[
\begin{align*}
X_{1i} &= A_{1i} \sim N(n, \mu, \sigma_{A_{1i}}^2) + E_{1i} \sim N(n, \mu, \sigma_{E_{1i}}^2), \\
X_{2i} &= A_{2i} \sim N(n, \mu, \sigma_{A_{2i}}^2) + E_{2i} \sim N(n, \mu, \sigma_{E_{2i}}^2),
\end{align*}
\]

\(i = 1, 2, \ldots, n\) for \(\mu = 0.110, \sigma_{A_{1i}}^2 = 0.8, \sigma_{A_{2i}}^2 = 0.7, \sigma_{E_{1i}}^2 = 0.4,\) and \(\sigma_{E_{2i}}^2 = 0.3.\) Based on the generated data, then required to estimate the parameter \(\sigma_{A_{2i}}^2\) and \(\sigma_{E_{2i}}^2\) to obtain the value of \(\theta_1\) and \(\theta_2\) for some \(n\). The simulation results are shown in Table 1 that the greater \(n\) was obtained \(\sigma_{A_{2i}}^2\) and \(\sigma_{E_{2i}}^2\) converge to 0.3600 and 0.0899 so the minimum sample which has been taken is 800. Therefore, the variance of genetic factor A on DM type 2 is greater than variance of lifestyle E.

5. Conclusion

Based on the results and discussion, parameters \(\widehat{\theta}_1\) and \(\widehat{\theta}_2\) were estimated based the joint posterior distribution \(f(\theta_1, \theta_2, b)\). The estimation results on simulation that defined in variances are 0.3600 for genetic factor A and 0.0899 for lifestyle E.

References

[1] Ozougwu, J C, O K Clarence, B C Donatus, and C B Unakalamba, 2013 J Phys and Path 4, 46-57
[2] Mishra, M and J F Ndisang, 2013 Current Pharmaceutical Design
[3] Tiwari, S and J F Ndisang, 2013 Current Pharmaceutical Design
[4] Ripsin, C M, Kang H, and Urban R J, 2009 Am Fam Physician, 29-36
[5] Murea, M, L Ma, and B I Freedman, 2012 The Review of Diabetic Studies, 9, 6-22
[6] Hu, F B, Manson J E, Stampfer M J, Colditz G, Liu S, and Solomon C G, 2001 Engl J Med
[7] Hossain, P, B Kawar, and M El Nahas, 2007 The New Engl J Medicine 356, 3, 213–215
[8] WHO, 2006 Obesity and overweight Fact Sheet no 311 in: World Health Organization
[9] Zahtamal, F Chandra, Suyanto, and T Restuastuti, 2007 Berita Kedokteran Masy 23
[10] Setiawan, A, 2008 Prosiding Seminar Basic Science V
[11] Neale, B, M Ferreira, S Medland, and D Posthuma, 2008 Taylor and Francis Group
[12] Rabe-Hesketh, S, A Skrondal, and H K Gjessing, 2008 Biometrics 64, 280-288
[13] Betteng, R, D Pangemanan, and N Mayulu, 2014 Jurnal e-Biomedik (eBM) 2, 404-412
[14] Berger, J O, 1985 Stat decision theory and bayesian analysis, Springer-Verlag, 2
[15] Congdon, P, 2003 Bayesian statistical modeling
[16] Bain, L J and M Engelhardt, 1992 Intro to probability and mathematical statistics 2
[17] Backlawski, K, 2004 Bayesian network development, new trends in software methodologies, tools and techniques (Boston) 18-48
[18] Saputro, D R S, Purnami W, Feri H, and Nughthoh A K, 2017 AIP Conf Proceedings
[19] Bolstad, W M, 2007 Introduction to bayesian statistics, A John Wiley and Sons Inc
[20] Guure C B, and Noor A I, 2012 Mathematical Problems in Engineering