วารสารวิทยาศาสตร์ลาดกระบัง ปีที่ 20 ฉบับที่ 2 เดือนกรกฎาคม-ธันวาคม 2554

การหาค่าตัวประมาณเบส์ด้วยโปรแกรมวินบั๊ก

The Estimation of Bayes Estimator with WinBUGS Program

อัชฌา อระวีพร

Autcha Araveeporn

สาขาวิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพมหานคร

บทคัดย่อ

โปรแกรมวินบั๊กเป็นโปรแกรมทางสถิติสำหรับการประมาณค่าประมาณเบส์โดยใช้วิธีของมาร์คอฟ เชน มอนติคาร์โล (MCMC) ในการประมาณค่าพารามิเตอร์ ตัวประมาณเบส์เป็นวิธีหนึ่งที่นิยมใช้เนื่องจากมีการใช้ฟังก์ชันการแจกแจงโดยหลักเกณฑ์มาช่วยในการประมาณค่าพารามิเตอร์ ซึ่งวิธีการนี้ค่อนข้างยุ่งยากในการพิสูจน์ในรูปแบบของการแจกแจงแต่โปรแกรมวินบั๊กสามารถช่วยคำนวณค่าของตัวประมาณจากการแจกแจงภายหลังจากตัวประมาณเบส์ โดยผู้ใช้โปรแกรมไม่จำเป็นต้องพิสูจน์ให้ได้รูปแบบของการแจกแจงก็สามารถประมาณค่าได้

คำสำคัญ : การแจกแจงภายหลัง, การแจกแจงโดยหลักเกณฑ์, ตัวประมาณเบส์, มาร์คอฟ เชน มอนติคาร์โล

Abstract

WinBUGS is statistical software to estimate Bayes estimator using Markov Chain Monte Carlo (MCMC) method. For parameter estimation, the Bayesian estimator is one method to use over a wide range because there is a prior distribution to evaluate parameter. However this method is rather complicated to be proved in form of distribution function but WinBUGS program can help to calculate Bayes estimator from posterior distribution. Therefore the user can estimate parameter without proving in order to know the distribution function.

Keywords : Bayes Estimator, Markov Chain Monte Carlo, Posterior Distribution, Prior Distribution

1. บทนำ

ตั้งแต่เริ่มต้นศตวรรษที่ 21 ตัวสถิติเบส์ได้มีการริเริ่มใช้กันอย่างแพร่หลายในด้านวิทยาศาสตร์และสามารถประยุกต์ใช้กับงานด้านต่างๆ วิธีการของตัวสถิติเบส์เป็นการประมาณค่าของพารามิเตอร์จากตัวประมาณเบส์ซึ่งบางครั้งเราอาจทราบข้อมูลหรือลักษณะของพารามิเตอร์โดย

เรียกว่าการแจกแจงโดยหลักเกณฑ์ (Prior Distribution) เมื่อทำการทดลองหรือสุ่มตัวอย่าง จะได้ฟังก์ชันความหนาแน่นของค่าสังเกตที่มีเงื่อนไขเมื่อกำหนดค่าพารามิเตอร์เรียกว่า ฟังก์ชันภาวะความน่าจะเป็น (Likelihood function) และสุดท้ายเมื่อต้องการประมาณค่าพารามิเตอร์เมื่อกำหนดค่าสังเกตจะเรียกว่าการแจกแจงภายหลัง (Posterior Distribution)

จากทฤษฎีในการประมาณค่าพารามิเตอร์ด้วยตัวสถิติเบส์ต้องใช้ความรู้และความเข้าใจในส่วนของทฤษฎีความน่าจะเป็นและสถิติเชิงอนุมานเข้ามาช่วย ซึ่งเป็นปัญหาสำหรับผู้ใช้บางกลุ่มที่ไม่ได้มีความเข้าใจในทฤษฎีสถิติมากนัก รู้สึกถึงความยุ่งยากและอาจเกิดความผิดพลาดในการประมาณค่าได้

ดังนั้นในปี 1990 มีกลุ่มนักสถิติได้คิดค้นวิธีการของ มาร์คอฟ เชน มอนติคาร์โล (Markov Chain Monte Carlo) หรือตัวย่อคือ MCMC [1-2] ซึ่งวิธีการนี้ทำให้การประมาณด้วยวิธีของเบส์ที่ซับซ้อนสามารถแก้ปัญหาได้ง่ายและสะดวกมากขึ้นจนกระทั่งเป็นที่นิยมใช้ในปัจจุบันอย่างแพร่หลาย ทำให้เกิดการพัฒนาวิธีการของ MCMC ในรูปของโปรแกรมเพื่อแก้ปัญหาเมื่อตัวแปร
มีความสัมพันธ์กับรูปแบบของการแจกแจงหลายขั้นตอน (Hierarchical Model)

ปี 1995 มีการตีพิมพ์วิธีงานวิจัยของกรีน [3] ที่เกี่ยวข้องกับขั้นต้อนการคัดเลือกตัวแบบในวิธีการของ MCMC ระหว่างนั้นโปรแกรมบั๊ก (BUGS) ซึ่งย่อมาจาก Bayesian Inference Using Gibbs Sampling ได้พัฒนาเป็นภาษาคอมพิวเตอร์โดยผู้ใช้ต้องกำหนดโครงสร้างของตัวแบบแล้วจึงใช้วิธี MCMC สร้างตัวอย่างจากการแจกแจงภายหลัง ในปัจจุบันระบบวินโดว์ได้เข้ามาช่วยให้การทำงานของโปรแกรมบั๊กง่ายขึ้นโดยมีการคิดค้นเอาระบบวินโดว์มารวมกับโปรแกรมบั๊กเรียกว่า โปรแกรมวินบั๊ก (WinBUGS) [4] โดยผู้ใช้สามารถดาวน์โหลดโปรแกรมได้โดยไม่เสียค่าใช้จ่ายได้ที่ที่เว็บไซต์ http://www.mrc-bsu.cam.ac.uk/bugs/

ในบทความนี้ได้นำเสนอวิธีการใช้โปรแกรมวินบั๊กหาการประมาณค่าตัวประมาณเบส์ พร้อมตัวอย่างในรูปแบบการแจกแจงก่อนแบบคู่สังยุค (Conjugate prior distribution) เมื่อการแจกโดยหลักเกณฑ์เป็นการแจกแจงแบบแกมมา และ ฟังก์ชันภาวะความน่าจะเป็นเป็นการแจกแจงแบบพัวซอง ซึ่งจากตัวอย่างสามารถนำไปปรับเพื่อใช้กับข้อมูลในรูปแบบอื่นๆของการแจกแจงก่อนแบบคู่สังยุคได้

2. ตัวประมาณเบส์

ให้ $\underset{\sim}{x} = x_1, \ldots, x_n$ เป็นแปรสุ่มจากประชากรที่มีอิสระและมีรูปแบบการแจกแจงเหมือนกัน โดยมีฟังก์ชันความหนาแน่น $f(\underset{\sim}{x}|\theta)$ โดยที่ θ เป็นค่าพารามิเตอร์ กำหนดให้ฟังก์ชันการแจกแจงโดยหลักเกณฑ์ (Prior Distribution Function) คือ $g(\theta)$ และ $f(\underset{\sim}{x}|\theta)$ เป็นฟังก์ชันความหนาแน่นแบบมีเงื่อนไข (Condition Density Function) ของตัวแปรสุ่ม $\underset{\sim}{x}$ เมื่อกำหนดค่าพารามิเตอร์คือ θ และ $h(\theta|\underset{\sim}{x})$ เรียกว่าฟังก์ชันความหนาแน่นของ θ เมื่อกำหนดค่าตัวแปรสุ่ม หรือเรียกว่าฟังก์ชันการแจกแจงภายหลัง (Posterior Distribution Function) ของ θ ในการประมาณค่าจากตัวแบบเบส์กำหนดวิธีการประมาณค่าฟังก์ชันต่างๆได้ดังนี้

1. ฟังก์ชันความหนาแน่นร่วมของ $\underset{\sim}{x}$ เมื่อกำหนด θ คือ

$$f(\underset{\sim}{x}|\theta) = f(x_1|\theta)\ldots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

2. ฟังก์ชันความหนาแน่นร่วมของ $\underset{\sim}{x}$ เมื่อกำหนด θ ให้ และฟังก์ชันการแจกแจงโดยหลักเกณฑ์ θ คือ

$$f(\underset{\sim}{x}|\theta)\,g(\theta) = f(x_1|\theta)\ldots f(x_n|\theta)\,g(\theta) = \prod_{i=1}^{n} f(x_i|\theta)\,g(\theta)$$

3. ฟังก์ชันการแจกแจงภายหลังของ θ คือ

$$h(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\,g(\theta)}{\int f(\underset{\sim}{x}|\theta)\,g(\theta)\,d\theta} = \frac{\prod_{i=1}^{n} f(x_i|\theta)\,g(\theta)}{\int \prod_{i=1}^{n} f(x_i|\theta)\,g(\theta)\,d\theta}$$

ฟังก์ชันการแจกแจงภายหลัง $h(\theta|\underset{\sim}{x})$ เป็นฟังก์ชันซึ่งขึ้นอยู่กับค่าพารามิเตอร์ θ ดังนั้นเมื่อพิจารณาฟังก์ชัน $f(x_i|\theta)$ สามารถแยกเป็นค่าคงที่ซึ่งไม่ขึ้นอยู่กับค่าพารามิเตอร์ θ จึงไม่จำเป็นต้องนำมาพิจารณาโดยสามารถเขียนได้ในรูปแบบดังนี้

$$h(\theta|\underset{\sim}{x}) \propto f(\underset{\sim}{x}|\theta)g(\theta)$$

4. ตัวประมาณภายหลังของเบย์ (Posterior Bayes Estimator) หรือเรียกว่าตัวประมาณเบส์ (Bayes Estimator) ของฟังก์ชัน $\tau(\theta)$ เทียบกับฟังก์ชันการแจกแจงโดยหลักเกณฑ์ g(θ) คือ

$$\tau(\theta) = E(\tau(\theta)|\underset{\sim}{x}).$$
$$= \int \tau(\theta)\, h(\theta|\,\underset{\sim}{x})d\theta$$
$$= \frac{\int \tau(\theta)\, f(\underset{\sim}{x}|\theta)\, g(\theta)d\theta}{\int f(\underset{\sim}{x}|\theta) g(\theta)\, d(\theta)}$$

3. ตัวอย่างเพื่อใช้หาค่าตัวประมาณเบส์ [5]

สมมติตัวแปร $\underset{\sim}{x}$ มีการแจกแจงแบบพัวซองด้วยค่าเฉลี่ย λ ซึ่งเป็นค่าพารามิเตอร์ ที่ต้องการประมาณสามารถเขียนได้ในรูปแบบ

$$x_i \,|\, \lambda \sim Poisson(\lambda) \quad ,i = 1,\ldots,n$$

โดยฟังก์ชันภาวะความน่าจะเป็นสามารถเขียนได้เป็น

$$f(\underset{\sim}{x}\,|\,\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i\,!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i\,!}$$

เมื่อพิจารณาฟังก์ชันการแจกแจงโดยหลักเกณฑ์เป็นการแจกแจงแกมมา ที่ค่าพารามิเตอร์ a และ b

$$g(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}$$

จะได้ฟังก์ชันการแจกแจงภายหลังดังนี้

$$h(\theta|\,\underset{\sim}{x}) = \frac{\dfrac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i\,!}\dfrac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}}{\int_0^\infty \dfrac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i\,!}\dfrac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}d\lambda}$$

ค่าคงที่ที่ไม่ขึ้นกับค่าพารามิเตอร์ λ ถูกหารทิ้งจะเหลือแค่

$$= \frac{e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}}{\int_0^\infty e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}d\lambda}$$

พิจารณาเทอมส่วน $\int_0^\infty e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}d\lambda$ ในรูปแบบฟังก์ชันแกมมา

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx = (\alpha-1)\Gamma(\alpha-1)$$

จะได้ว่า

$$\int_0^\infty e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}d\lambda = \frac{1}{(n+b)^{n\bar{x}+a-2}}\int_0^\infty e^{-(n+b)\lambda}[(n+b)\lambda]^{(n\bar{x}+a-1)}d(n+b)\lambda$$

$$= \frac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}$$

ดังนั้น ฟังก์ชันการแจกแจงภายหลังคือ

$$h(\theta|\underset{\sim}{x}) = \frac{e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}}$$

ค่าเฉลี่ยของ λ จากการแจกแจงภายหลังคือ

$$E(\lambda|\underset{\sim}{x}) = \frac{\int_0^\infty \lambda e^{-(n+b)\lambda}\lambda^{n\bar{x}+a-1}d\lambda}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}} = \frac{\int_0^\infty e^{-(n+b)\lambda}\lambda^{n\bar{x}+a}d\lambda}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}}$$

$$= \frac{\dfrac{\Gamma(n\bar{x}+a+1)}{(n+b)^{n\bar{x}+a-1}}}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}} = \frac{\Gamma(n\bar{x}+a+1)(n+b)^{n\bar{x}+a-2}}{\Gamma(n\bar{x}+a)(n+b)^{n\bar{x}+a-1}}$$

$$= \frac{n\bar{x}+a}{n+b}$$

ส่วนค่าความแปรปรวนของ λ จากการแจกแจงภายหลังหาได้จาก

$$V(\lambda|\underset{\sim}{x}) = E(\lambda^2|\underset{\sim}{x}) - \left(E(\lambda|\underset{\sim}{x})\right)^2$$

ซึ่งค่า $E(\lambda^2|\underset{\sim}{x})$ หาจาก

$$E(\lambda^2 \mid \underset{\sim}{x}) = \frac{\int_0^\infty \lambda^2 e^{-(n+b)\lambda} \lambda^{n\bar{x}+a-1} d\lambda}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}} = \frac{\int_0^\infty e^{-(n+b)\lambda} \lambda^{n\bar{x}+a+1} d\lambda}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}}$$

$$= \frac{\dfrac{\Gamma(n\bar{x}+a+2)}{(n+b)^{n\bar{x}+a}}}{\dfrac{\Gamma(n\bar{x}+a)}{(n+b)^{n\bar{x}+a-2}}} = \frac{\Gamma(n\bar{x}+a+2)(n+b)^{n\bar{x}+a-2}}{\Gamma(n\bar{x}+a)(n+b)^{n\bar{x}+a}}$$

$$= \frac{(n\bar{x}+a+1)(n\bar{x}+a)}{(n+b)^2}$$

และหาค่าความแปรปรวนหาได้จาก

$$V(\lambda \mid \underset{\sim}{x}) = \frac{(n\bar{x}+a+1)(n\bar{x}+a)}{(n+b)^2} - \left(\frac{n\bar{x}+a}{n+b}\right)^2$$

$$= \frac{n\bar{x}+a}{(n+b)^2}$$

วิธีการข้างบนค่อนข้างยุ่งยากแต่ถ้าพิจารณาจากรูปแบบการแจกแจงภายหลังดังนี้

$$h(\lambda \mid \underset{\sim}{x}) \propto f(\underset{\sim}{x} \mid \lambda) g(\lambda)$$

$$\propto \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

ค่าคงที่ $\left(b^a, \prod_{i=1}^{n} x_i!, \Gamma(a)\right)$ ไม่ขึ้นกับค่าพารามิเตอร์ λ ถูกหารทิ้งดังวิธีการข้างบนจะเหลือเทอมที่พิจารณาแค่

$$\propto e^{-(n+b)\lambda} \lambda^{(n\bar{x}+a)-1}$$

จากฟังก์ชันการแจกแจงภายหลังที่ได้มาพบว่าอยู่ในรูปการแจกแจงแกมมาและค่าพารามิเตอร์คือ

$$\lambda \mid \underset{\sim}{x} \sim Gamma(n\bar{x}+a, n+b)$$

ค่าเฉลี่ยของ λ จากการแจกแจงภายหลังคือ

$$E(\lambda \mid \underset{\sim}{x}) = \frac{n\bar{x}+a}{n+b}$$

และค่าความแปรปรวนของ λ จากการแจกแจงภายหลังคือ

$$V(\lambda \mid \underset{\sim}{x}) = \frac{n\overline{x}+a}{(n+b)^2}$$

3.1 ตัวอย่างเพื่อทดสอบโปรแกรมวินบั๊ก

จากรูปแบบการแจกแจงจากตัวอย่างข้างบน นำรูปแบบการแจกแจงมาเพื่อจำลองตัวแปรสุ่ม $\underset{\sim}{x}$ จากการแจงแจงพัวซองที่ค่าเฉลี่ย (λ) เท่ากับ 3 มา 50 ค่าสังเกตดังนี้

5	4	3	3	3	5	1	4	0	5
2	5	3	3	5	6	2	4	1	2
5	5	5	5	2	0	5	6	5	3
5	5	3	5	2	5	3	6	0	5
2	4	2	6	4	3	4	3	4	2

พบว่าเมื่อนำค่าสังเกตทั้ง 50 ค่ามาหาค่าเฉลี่ย ($\overline{x}$) ได้เท่ากับ 3.6 และกำหนดรูปแบบการแจกแจงของค่าสังเกตดังนี้

$$x_i \mid \lambda \sim Poisson(\lambda) \quad ,i=1,\ldots,50$$
$$\lambda \sim Gamma(a,b)$$

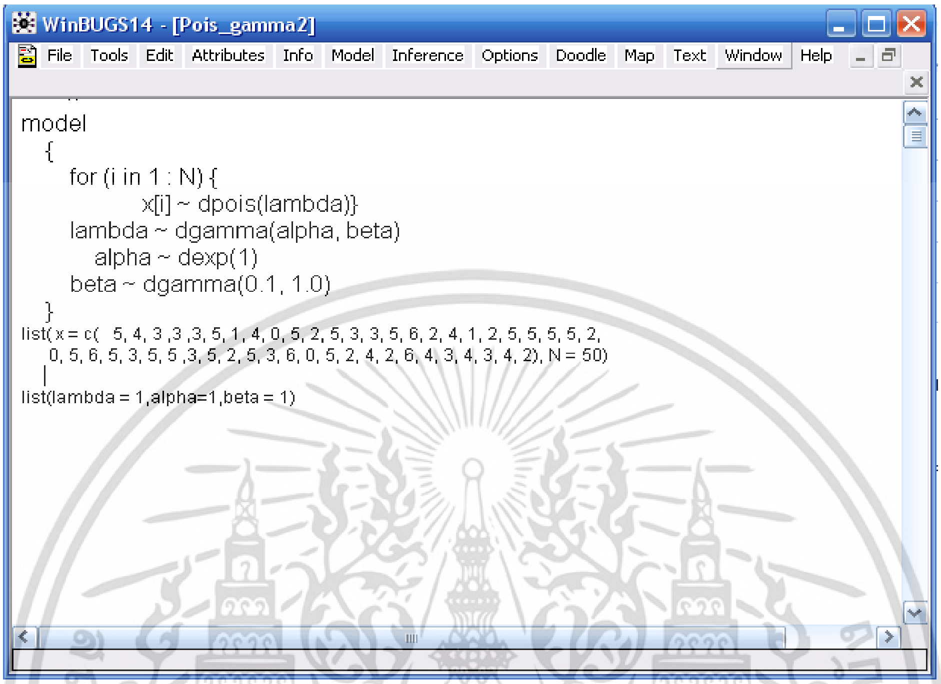
ถ้าทราบค่าพารามิเตอร์ของการแจกแจงแกมมาสามารถกำหนดได้ แต่ถ้าไม่ทราบค่าพารามิเตอร์ a และ b สามารถให้โปรแกรมประมาณได้โดยอาจกำหนดให้อยู่ในรูปแบบการแจกแจง ซึ่งค่าพารามิเตอร์ที่กำหนดขึ้นตามลักษณะข้อมูลดังตัวอย่าง

$$a \sim Exponential(1)$$
$$b \sim Gamma(0.1,1)$$

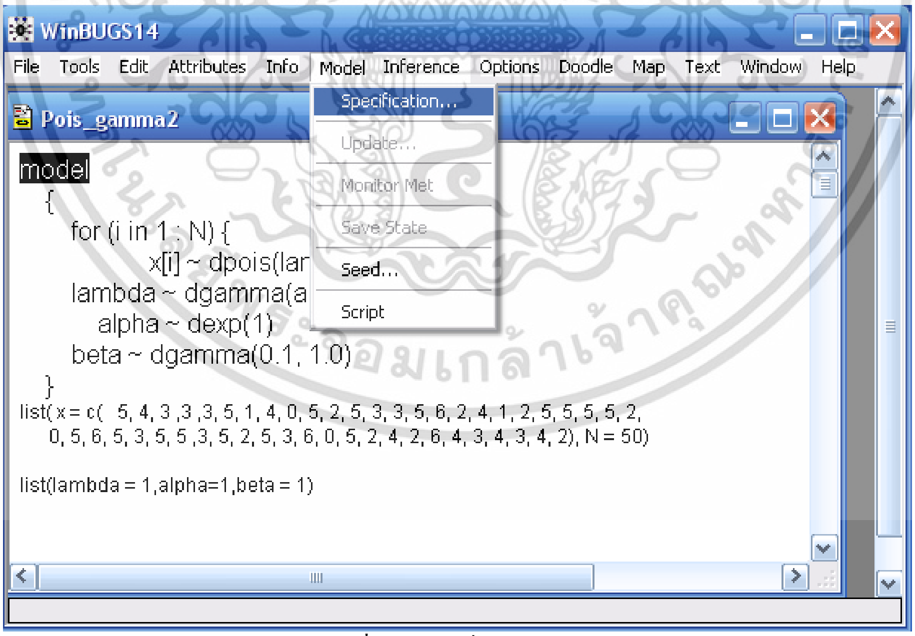
เมื่อกำหนดค่าพารามิเตอร์แล้วจึงเขียนให้อยู่ในรูปแบบของโปรแกรมวินบั๊กหลังคำว่า model และอยู่ภายในเครื่องหมาย {.....} ตามด้วยกำหนดข้อมูลหลังคำว่า list และกำหนดค่าเริ่มต้นในการจำลองค่าพารามิเตอร์หลังคำว่า list ดังรูปที่ 1

ขั้นตอนการประมวลผลโปรแกรมวินบั๊กมีดังต่อไปนี้

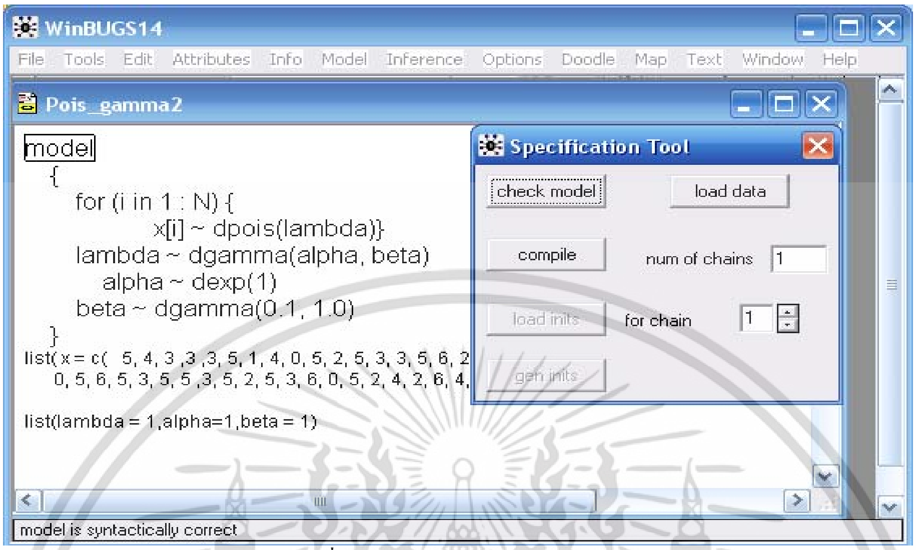
1. ตรวจสอบตัวแบบ (check model) ให้ผู้ใช้เลือกเน้นคำว่า model และไปที่แถบเมนูเลือกคำสั่ง Model ตามด้วย Specification ดังรูปที่ 2 หลังจากนั้นจะได้หน้าจอจะแสดงหน้าต่าง Specification Tool และเลือกปุ่ม check model เพื่อตรวจสอบตัวแบบที่กำหนดว่าถูกต้องหรือไม่ ถ้าถูกต้องไม่มีข้อผิดพลาด ด้านล่างของหน้าจอจะเขียนคำว่า “model is syntactically correct” ดังรูปที่ 3



รูปที่ 1. หน้าจอรูปแบบการแจกแจงของค่าสังเกต

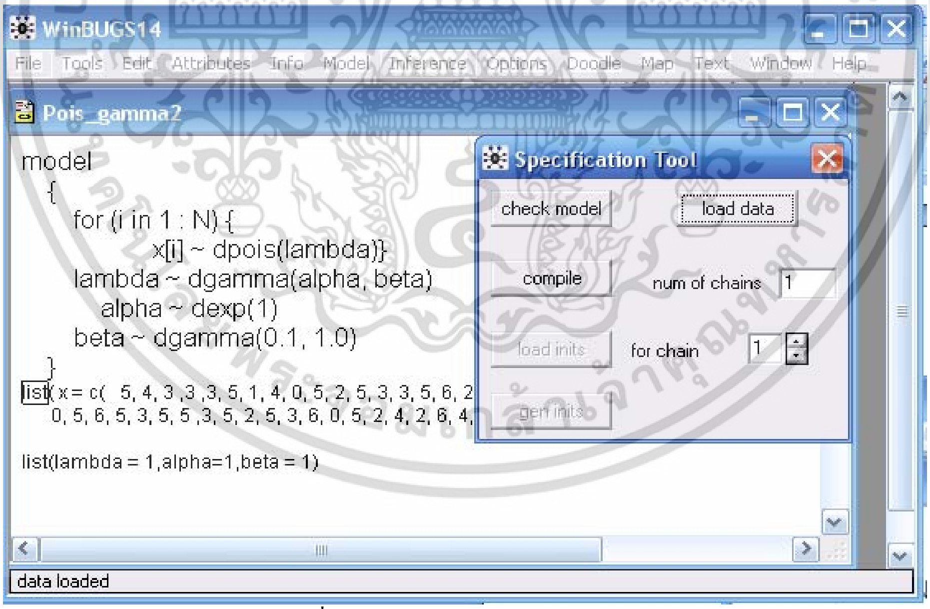


รูปที่ 2. หน้าจอเพื่อตรวจสอบตัวแบบ



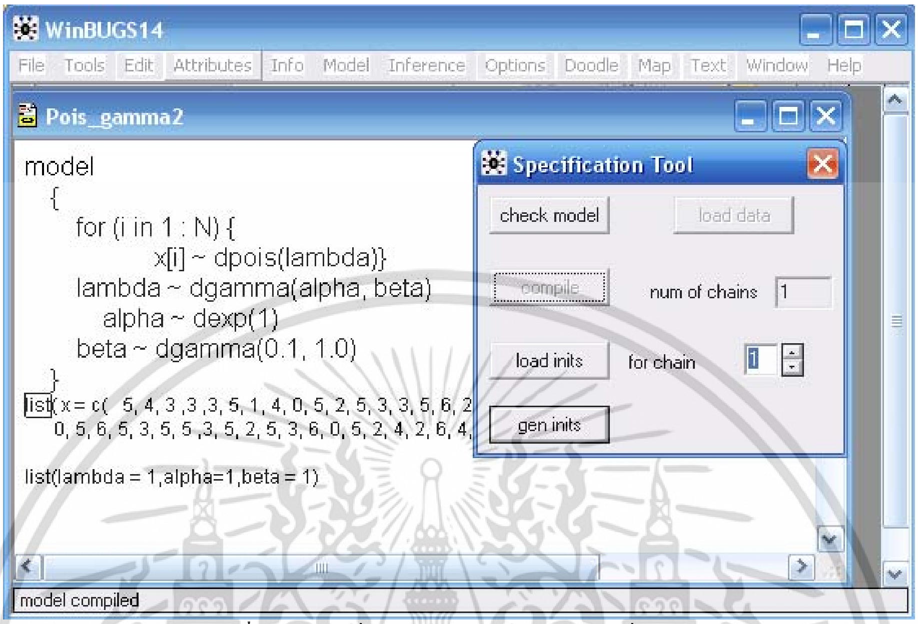
รูปที่ 3. หน้าจอภายหลังการตรวจสอบตัวแบบ

2. เรียกข้อมูล(load data) เมื่อโปรแกรมถูกต้องให้เรียกข้อมูลที่มีอยู่โดยเลือกเน้นคำว่า list ก่อนที่จะเลือกปุ่ม load data หลังจากที่โปรแกรมเรียกข้อมูลครบจะปรากฏคำว่า “data loaded” ด้านล่างของหน้าจอ ดังรูปที่ 4



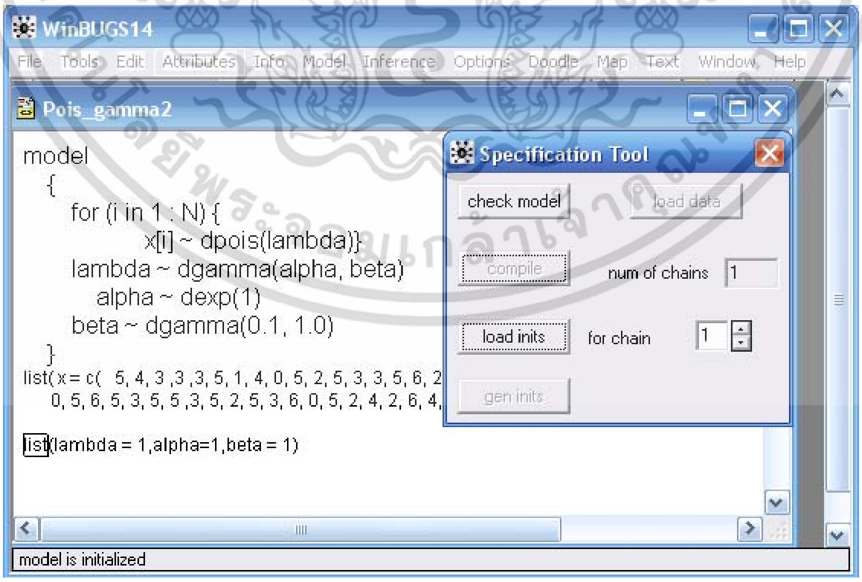
รูปที่ 4. หน้าจอการเรียกข้อมูลเข้าโปรแกรม

3. ประมวลผล (compile) เลือกปุ่ม compile เพื่อยืนยันการประมวลผลของโปรแกรม จะปรากฏคำว่า “model compiled” ดังรูปที่ 5



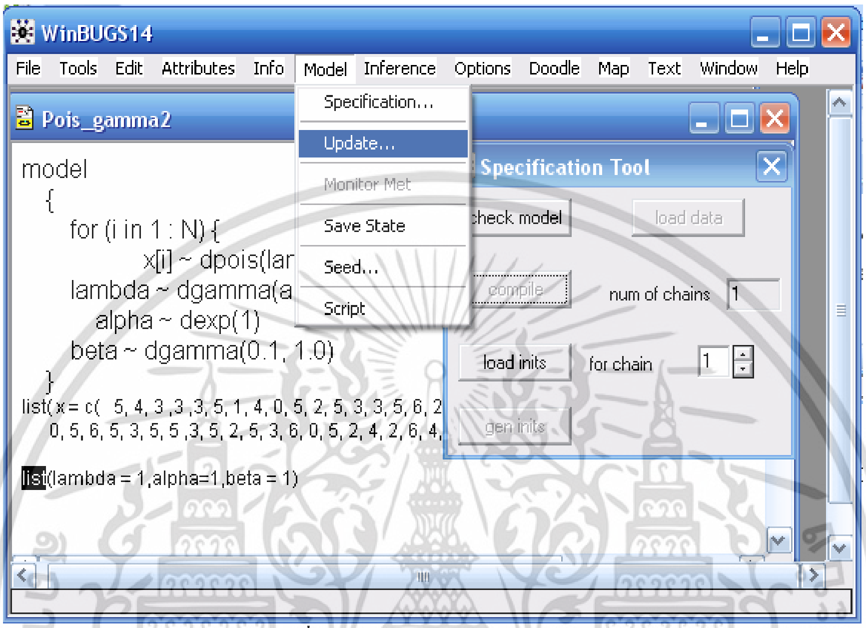
รูปที่ 5. หน้าจอเพื่อประมวลผลตัวแบบและข้อมูลที่เรียกเข้า

4. เรียกค่าเริ่มต้น (load initial values) ผู้ใช้สามารถกำหนดค่าเริ่มต้นซึ่งเป็นค่าเริ่มต้นของค่าพารามิเตอร์ไว้หลังคำว่า list หลังจากนั้นเลือกเน้นคำว่า list และเลือกปุ่ม load inits ใต้หน้าต่างจะขึ้นคำว่า “model is initialized” ดังรูปที่ 6 แต่ถ้าไม่มีการกำหนดค่าเริ่มต้นให้เลือกปุ่ม gen inits เพื่อให้โปรแกรมสร้างค่าเริ่มต้นให้



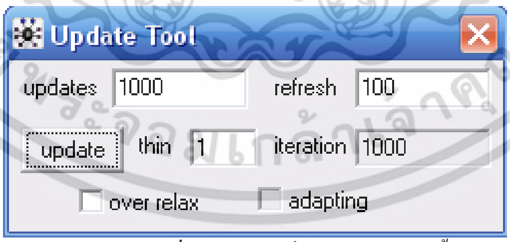
รูปที่ 6. หน้าจอการเรียกค่าเริ่มต้น

5. สร้างค่าทดลอง (Generate Burn-in values) โดยเลือกแถบเมนู Model และ เลือกที่ Update ดังรูปที่ 7



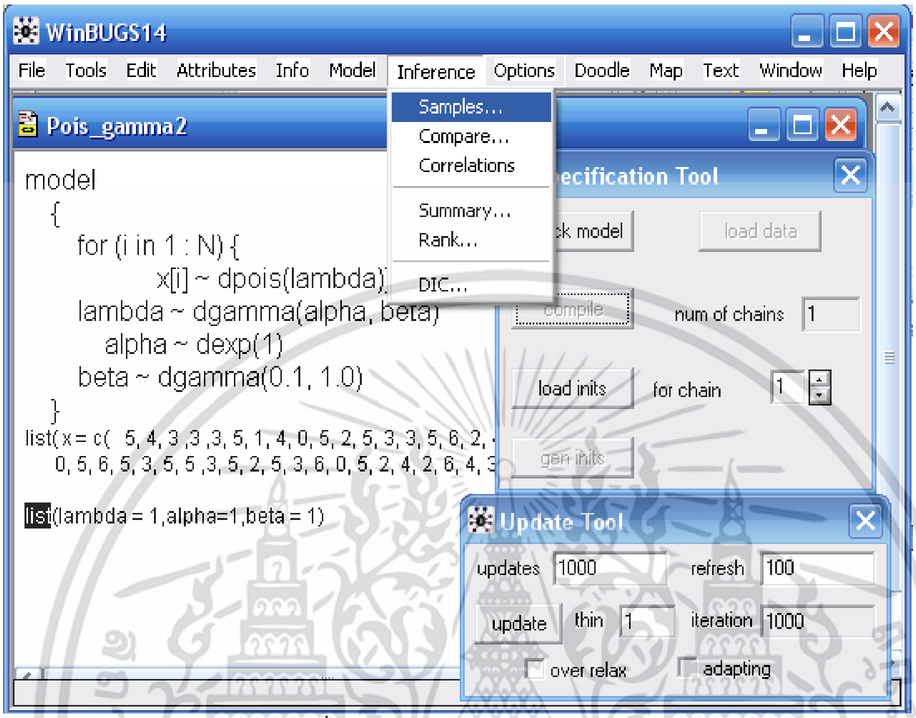
รูปที่ 7. หน้าจอการสร้างค่าทดลอง

หลังจากนั้นจะขึ้นหน้าต่างเพื่อกำหนดจำนวนซ้ำของค่าทดลอง โดยอาจจะกำหนดให้ 1000 หรือค่าเท่าไร ขึ้นอยู่กับผู้ใช้ต้องการ และเลือกปุ่ม update เพื่อสุ่มข้อมูลเริ่มต้น ดังรูปที่ 8



รูปที่ 8. หน้าต่างเพื่อกำหนดจำนวนซ้ำ

6. กำหนดค่าพารามิเตอร์ (Monitor parameters) ไปที่แถบเมนูที่คำว่า Inference และเลือก คำสั่ง Samples ดังรูปที่ 9

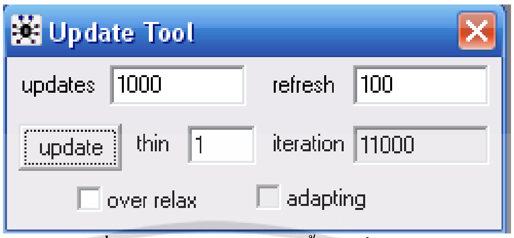


รูปที่ 9. หน้าจอการเลือกค่าพารามิเตอร์

จะได้หน้าต่างใหม่เพื่อใส่ค่าพารามิเตอร์ที่ต้องการประมาณและเลือกปุ่ม set เพื่อกำหนดค่าพารามิเตอร์ที่ต้องการ ดังรูปที่ 10

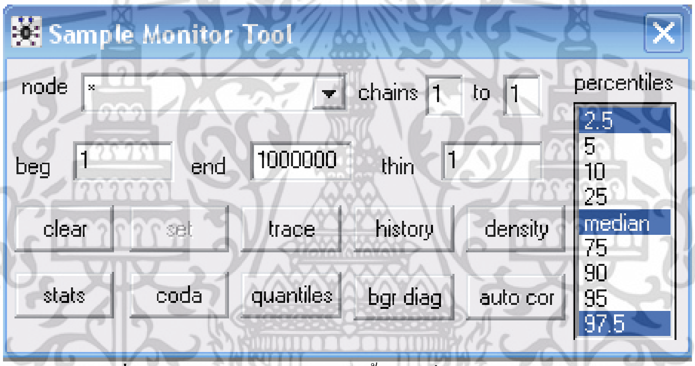
รูปที่ 10. การกำหนดค่าพารามิเตอร์จากตัวแบบ

7. สร้างค่าจากการแจกแจงโดยหลักเกณฑ์ (Generate posterior values) โดยการเลือกปุ่ม update ที่หน้าต่าง Update Tool ตามจำนวนซ้ำที่ต้องการ ดังรูปที่ 11



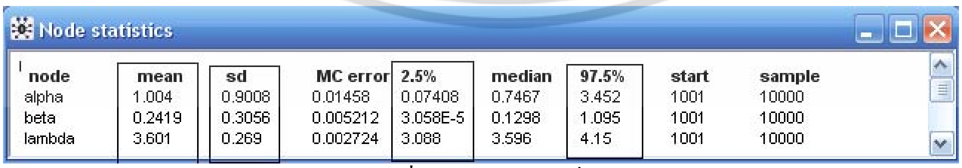
รูปที่ 11. การกำหนดจำนวนซ้ำตามที่ต้องการ

8. ถ้าต้องการค่าประมาณจากการแจกแจงโดยหลักเกณฑ์ทุกค่าที่กำหนดขึ้น ให้ใส่เครื่องหมาย * ในช่อง node และเลือกปุ่ม stats ดังรูปที่ 12



รูปที่ 12. การเลือกค่าพารามิเตอร์ทั้งหมดที่ต้องการประมาณ

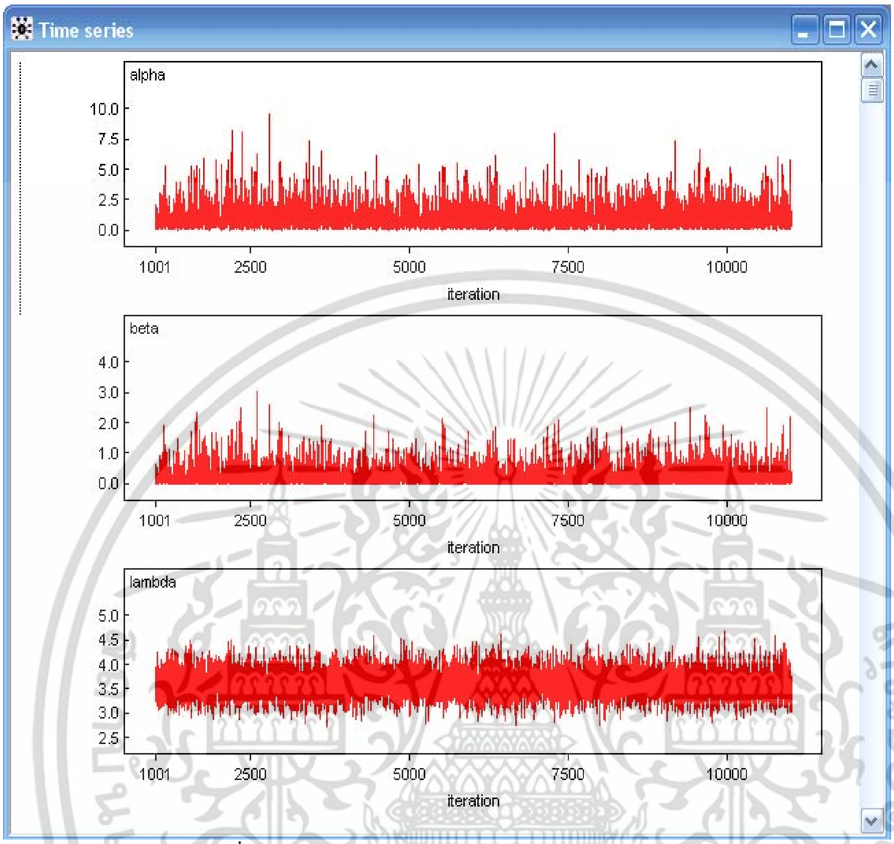
หลังจากนั้นเลือกปุ่ม stats จะได้ค่าพารามิเตอร์ที่ต้องการ โดยพิจารณาจากค่าเฉลี่ย (mean) นอกจากนี้ผลลัพธ์ที่ได้ยังให้ค่าส่วนเบี่ยงเบนมาตรฐาน (sd) ค่าความคลาดเคลื่อนจากการสุ่มแบบ MC (Markov Chain) และช่วงความเชื่อมั่น 95% (2.5% และ 97.5%) ค่าเริ่มต้นและค่าทั้งหมดจากการสุ่มตัวอย่าง ดังรูปที่ 13


Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	1.004	0.9008	0.01458	0.07408	0.7467	3.452	1001	10000
beta	0.2419	0.3056	0.005212	3.058E-5	0.1298	1.095	1001	10000
lambda	3.601	0.269	0.002724	3.088	3.596	4.15	1001	10000

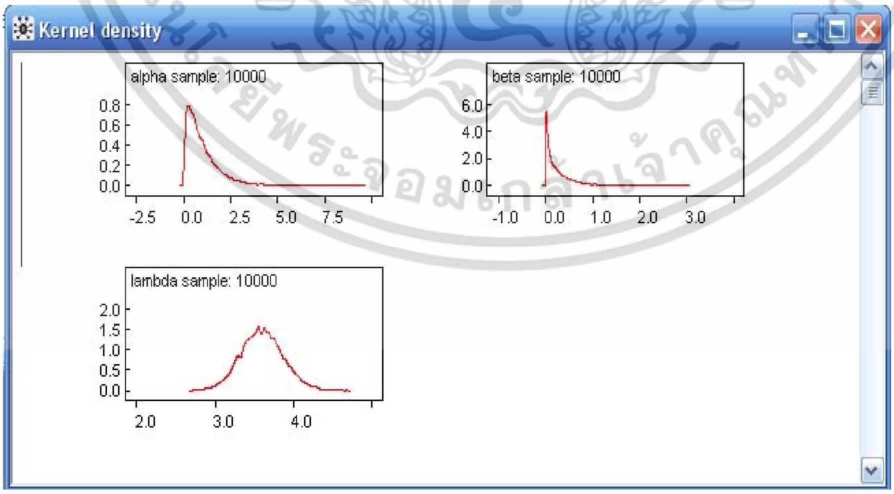
รูปที่ 13. ค่าพารามิเตอร์ที่ต้องการประมาณ

ถ้าเลือกปุ่ม history จากรูปที่ 12 จะแสดงกราฟย้อนหลังการสุ่มตัวอย่างของค่าพารามิเตอร์ดังรูปที่ 14



รูปที่ 14. กราฟย้อนหลังการสุ่มตัวอย่างของค่าพารามิเตอร์

และถ้าเลือกปุ่ม density จากรูปที่ 12 จะแสดงลักษณะการแจกแจงของค่าพารามิเตอร์ ดังรูปที่ 15



รูปที่ 15. กราฟการแจกแจงของค่าพารามิเตอร์

จากรูปที่ 13 พบว่าตัวประมาณการแจกแจงโดยหลักเกณฑ์ λ มีค่าเท่ากับ 3.601 และมีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.269 ส่วนค่า alpha หรือค่า a คือค่าที่ได้จากการแจกแจงแกมมามีค่าเท่ากับ 1.004 และ ค่า beta หรือ ค่า b จากการแจกแจงแกมมามีค่าเท่ากับ 0.2419

จากตัวอย่างข้างบนตัวประมาณที่ได้จากการแจกแจงโดยหลักเกณฑ์ λ มีรูปแบบเป็นการแจกแจงแกมมา

$$\lambda \mid \underset{\sim}{x} \sim Gamma(n\bar{x}+a, n+b)$$

ค่าเฉลี่ยของ λ จากการแจกแจงภายหลังคือ

$$\begin{aligned} E(\lambda \mid \underset{\sim}{x}) &= \frac{n\bar{x}+a}{n+b} \\ &= \frac{(50\times 3.6)+1.004}{50+0.2419} \\ &= 3.602 \end{aligned}$$

ซึ่งค่าที่ได้ใกล้เคียงกับค่าพารามิเตอร์ที่ประมาณได้จากโปรแกรมวินบั๊ก และค่าความแปรปรวนของ λ จากการแจกแจงภายหลังคือ

$$\begin{aligned} V(\lambda \mid \underset{\sim}{x}) &= \frac{n\,\bar{x}+a}{(n+b)^2} \\ &= \frac{(50\times 3.6)+1.004}{(50+0.2419)^2} \\ &= 0.0717 \end{aligned}$$

เมื่อถอดรากที่สองจะได้ค่าส่วนเบี่ยงเบนมาตรฐานคือ 0.267

จากข้อมูลในตัวอย่างข้างต้นจะเป็นว่าถ้าผู้ที่ไม่สามารถพิสูจน์ฟังก์ชันให้อยู่ในรูปของการแจกแจงดังตัวอย่างที่แสดง ก็สามารถใช้โปรแกรมวินบั๊กช่วยในการประมาณค่าพารามิเตอร์ได้ ซึ่งจากผลข้างบนช่วยทำให้ยืนยันได้ว่าโปรแกรมวินบั๊กที่ให้ค่า 3.601 ให้ค่าใกล้เคียงกับกรณีที่เราทราบการแจกแจงคือ 3.602

4. บทสรุป

ปัจจุบันการใช้โปรแกรมสำเร็จรูปสำหรับวิเคราะห์ข้อมูลมีปัญหาทางลิขสิทธิ์ทำให้การวิเคราะห์ข้อมูลและนำเสนอผลที่ได้จากการวิเคราะห์ทางสถิติมีปัญหา แต่โปรแกรมวินบั๊กสามารถดาวน์โหลดโปรแกรมได้โดยไม่เสียค่าใช้จ่ายทำให้ในต่างประเทศโปรแกรมนี้เป็นที่นิยมใช้ในการประมาณค่าตัวประมาณเบส์ แต่ในประเทศไทยยังไม่ค่อยเป็นที่นิยมใช้มากนักเนื่องจากการใช้โปรแกรมในการประมาณค่าของตัวประมาณเบส์ต้องมีการกำหนดรูปแบบของการแจกแจงของตัวอย่าง หลายขั้นตอนและ

ในการประมวลผลต้องมีการเลือกเมนูตามขั้นตอนหลายอย่าง ทำให้ผู้ใช้เกิดความสับสนในการทำงานและสรุปผลการวิจัย

ในบทความนี้จึงได้อธิบายขั้นตอนการทำงานที่จำเป็นในการประมาณค่าตัวประมาณเบส์ในแต่ละขั้นตอนอย่างละเอียดเหมาะสำหรับผู้สนใจที่จะวิเคราะห์ข้อมูลสำหรับตัวประมาณเบส์ นอกจากนี้โปรแกรมวินบั๊กยังสามารถช่วยแก้ปัญหาสำหรับผู้ที่ไม่เข้าใจการประมาณค่าในตัวแบบเบส์ซึ่งต้องมีการใช้สถิติเชิงอนุมานและการพิสูจน์รูปแบบการแจกแจงทางสถิติ ก็สามารถประมาณค่าพารามิเตอร์จากฟังก์ชันการแจกแจงภายหลังซึ่งผลลัพธ์ที่ได้จะให้ค่าที่ใกล้เคียงกัน

เอกสารอ้างอิง

[1] Gelfand, A. and Smith, A., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

[2] Gelfand, A., Hills, S., Racine-Poon, A. and Smith, A., Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association,* 85, 972-985.

[3] Green, P., 1995, Reversible jump Markov chain Monte Carlo computation and Bayesian model Determination. *Biometrika*, 69, 309-314.

[4] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D., 2003. *WinBUGS User Manual, Version 1.4* , MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK.

[5] Ntzoufras, I., 2009. *Bayesian Modeling Using WinBUGS*, John Wiley & Sons, Inc., Hoboken, New Jersey.