NU Science Journal 2012; 9(1): 30- 44

การวิเคราะห์เบส์จากโปรแกรมวินบั๊กสู่โปรแกรมอาร์

อัชฌา อระวีพร

Bayesian Analysis from WinBUGS Program to R Program

Autcha Araveeporn

สาขาวิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
กรุงเทพมหานคร

*Corresponding author. E-mail: kaautcha@hotmail.com

บทคัดย่อ

โปรแกรมวินบั๊กเป็นโปรแกรมทางสถิติสำหรับการประมาณค่าประมาณเบส์โดยใช้วิธีของมาร์คอฟ เชน มอนติคาร์โล (MCMC) ซึ่งชุดคำสั่ง R2WinBUGS สร้างขึ้นเพื่ออำนวยความสะดวกให้ผู้ใช้สามารถเรียกโปรแกรมวินบั๊กได้จากโปรแกรมอาร์ซึ่งสามารถเขียนคำสั่ง ข้อมูล และประมวลผลโดยใช้โปรแกรมวินบั๊กพร้อมๆกับโปรแกรมอาร์ โดยผลลัพธ์ที่ได้สามารถเรียกดูได้จากโปรแกรมอาร์ ซึ่งบทความนี้นำเสนอตัวอย่างจากตัวแบบเบส์เพื่อวิเคราะห์ด้วยโปรแกรมวินบั๊ก และเขียนคำสั่งจากโปรแกรมอาร์เพื่อเรียกโปรแกรมวินบั๊กมาประมวลผล ซึ่งผลลัพธ์จากตัวประมาณที่ได้ให้ค่าที่ใกล้เคียงกัน

คำสำคัญ: โปรแกรมวินบั๊ก โปรแกรมอาร์ มาร์คอฟ เชน มอนติคาร์โล

Abstract

WinBUGS is statistical software to estimate Bayes estimator using Markov Chain Monte Carlo (MCMC) method. The R2WinBUGS package provides convenient functions to call WinBUGS from R program. It automatically writes the scripts and data by WinBUGS processing with R program. After the WinBUGS process has finished, the results can seen from R program. This article shows the example from Bayesian analysis with WinBUGS program and the code from R program that linked to Win BUGS for processing. However, the output from 2 program is similar for parameter estimation.

Keywords: WinBUGS program, R program, Markov Chain Monte Carlo

บทนำ

กฎของเบส์เป็นส่วนหนึ่งของวิชาความน่าจะเป็นโดยเป็นการประมาณค่าความน่าจะเป็นจากเหตุการณ์หรือค่าสังเกตของตัวแปรสุ่มโดยใช้ทฤษฎีของเบส์ (Carlin and Louis, 2009) ซึ่งเริ่มต้นจากความน่าจะเป็นแบบมีเงื่อนไข กฎการคูณ และกฎความน่าจะเป็นรวม นอกจากนี้ในวิชาการอนุมานเชิงสถิติก็ยังสามารถนำการวิเคราะห์ตัวแบบเบส์มาใช้ในการประมาณค่าแบบจุด โดยพิจาณาจากฟังก์ชันการแจกแจงของค่าสังเกตของตัวแปรสุ่มเมื่อกำหนดค่าพารามิเตอร์โดยเรียกว่า ฟังก์ชันภาวะความน่าจะเป็น (Likelihood Function) ซึ่งทราบลักษณะการแจกแจงของค่าพารามิเตอร์ซึ่งเรียกว่า ฟังก์ชันการแจกแจงโดยหลักเกณฑ์ (Prior Density Function) เพื่อใช้ประมาณค่าพารามิเตอร์เมื่อกำหนดค่าสังเกตจากฟังก์ชันการแจกแจงโดยประสบการณ์ (Posterior Density Function)

วิธีการวิเคราะห์ด้วยตัวแบบเบส์ค่อนข้างยุ่งยากสำหรับผู้ที่ไม่ถนัดเนื้อหาวิชาทางสถิติ จึงได้มีนักสถิติพัฒนาวิธีของ มาร์คอฟ เชน มอนติคาร์โล (Markov Chain Monte Carlo) หรือเรียกว่าวิธี MCMC โดยวิธีนี้เป็นที่นิยมใช้ในการหาตัวประมาณจากฟังก์ชันการแจกแจงโดยประสบการณ์จึงมีการพัฒนาเป็นโปรแกรมสำเร็จรูปเรียกว่า โปรแกรมวินบั๊ก (WinBUGS) ซึ่งย่อมาจากคำ Bayesian Inference Using Gibbs Sampling โดย Spiegelhalter และคณะ (2003)

โปรแกรมวินบั๊กเป็นโปรแกรมที่ใช้วิธีของ MCMC (Gilks *et al*., 1996)ในการวิเคราะห์ตัวแบบเบส์ที่ซับซ้อน ซึ่งเป็นการสุ่มตัวอย่างด้วยวิธีกิบส์ (Gibbs Sampling) (Geman and Geman, 1984; Gelfand and Smith, 1990; Casella and George, 1992) เพื่อสร้างค่าจากวิธี มาร์คอฟ เชน (Markov Chain) โดยการสุ่มตัวอย่างจากการแจกแจง แล้วจึงใช้วิธีการประมาณค่าของมาร์คอฟ (Markov) ในการประมาณค่าพารามิเตอร์ ซึ่งโปรแกรมวินบั๊กเป็นโปรแกรมที่สามารถดาวน์โหลดมาใช้ได้โดยไม่เสียค่าใช้จ่ายที่เว็บไซต์ http://www.mrc-bsu.cam.ac.uk/bugs/

ในการใช้โปรแกรมวินบั๊กผู้ใช้ต้องสร้างตัวแบบที่ใช้ในการประมวลผล เรียกข้อมูล และกำหนดค่าเริ่มต้นสำหรับจำนวนรอบของมาร์คอฟ เชน ซึ่งปัญหาที่เกิดสำหรับผู้ที่ใช้โปรแกรมวินบั๊กคือผลลัพธ์ที่ได้จากการประมวลผลแต่ละครั้งเก็บอยู่ในโปรแกรมวินบั๊กเท่านั้น แต่กระนั้นในบางครั้งผู้ใช้ต้องการจำลองข้อมูลหลายๆ รอบเพื่อหาผลลัพธ์จากการวิเคราะห์เบส์และต้องการเก็บผลลัพธ์ในลักษณะของโปรแกรมอาร์ (R Program) (R Development Core Team, 2004) จึงได้มีการสร้างชุดคำสั่ง R2WinBUGS เพื่อใช้ในการเรียกโปรแกรม WinBUGS เพื่อประมวลผลและกลับมาเก็บผลลัพธ์ในโปรแกรมอาร์

โปรแกรมอาร์ เป็นโปรแกรมที่เขียนเป็นภาษาเพื่อใช้ในการวิเคราะห์ข้อมูลทางสถิติ ผู้ที่ต้องการใช้สามารถดาวน์โหลดโดยไม่เสียค่าใช้จ่ายที่เว็บไซต์ http://www.R-project.org ซึ่งโปรแกรมนี้ได้พัฒนามาจากภาษาเอส (S Language) โดย Becker และ Chambers (1984) Becker และคณะ, (1988) Chambers และ Hastie, (1992) Chambers, (1998) ในโปรแกรมอาร์มีชุดคำสั่งสำหรับวิเคราะห์ข้อมูลในแต่ละประเภทอยู่มากมายแล้วแต่ผู้ใช้ต้องการใช้เรื่องใดก็สามารถดาวน์โหลดได้โดยไม่เสียค่าใช้จ่าย ส่วนชุดคำสั่ง R2WinBUGS เป็นชุดคำสั่งหนึ่งสำหรับผู้สนใจการวิเคราะห์ข้อมูลในรูปของการแจกแจงที่เกี่ยวข้องกันหลายๆขั้นหรือการวิเคราะห์เบส์ ซึ่งสามารถดาวน์โหลดเพื่อมาติดตั้งได้ทางอินเตอร์เน็ตขณะที่เรียกใช้อยู่ในโปรแกรมอาร์และผู้ใช้ควรจะดาวน์โหลดชุดคำสั่ง coda (Plunmer *et al*., 2004) ซึ่งใช้สำหรับการวิเคราะห์ผลลัพธ์จากวินบั๊ก เพื่อให้การทำงานสมบูรณ์

ในบทความนี้ได้นำเสนอวิธีการของมาร์คอฟ เชน มอนติคาร์โล (MCMC) ด้วยวิธีการสุ่มตัวอย่างจากวิธีคาร์คอฟ เชน วิธีการสุ่มตัวอย่างแบบกิบส์ และนำเสนอตัวอย่างเพื่อเปรียบเทียบการทำงานของโปรแกรมวินบั๊กและการเขียนชุดคำสั่ง R2WinBUGS ในโปรแกรมอาร์

วิธี มาร์คอฟ เชน มอนติคาร์โล (MCMC)

วิธี MCMC เป็นวิธีที่นิยมใช้เมื่อไม่ทราบฟังก์ชันการแจกแจงบางส่วน (Marginal Distribution) ของตัวแปรสุ่ม โดยประกอบด้วยการสุ่มตัวอย่างตัวแปรจากวิธี มาร์คอฟ เชน (Markov Chain) จากการแจกแจงโดยหลักเกณฑ์ของค่าพารามิเตอร์ที่สนใน และวิธีการสุ่มตัวอย่างแบบกิบส์ (Gibbs sampling) ได้ถูกนำมาใช้ในวิธีของ MCMC สำหรับประมาณค่าพารามิเตอร์จากฟังก์ชันการแจกแจงโดยประสบการณ์

ขั้นตอนการสร้างสุ่มตัวอย่างจากจากวิธี มาร์คอฟ เชน ประกอบด้วย

1. กำหนดค่าเริ่มต้น $\theta^{(0)}$ จากฟังก์ชันการแจกแจงโดยหลักเกณฑ์
2. สร้างค่าจากข้อ 1. มา T ค่าจนกระทั่งได้ลักษณะการแจกแจงที่ต้องการ
3. ตรวจสอบการแจกแจงในข้อ 2. ถ้าไม่เป็นไปตามลักษณะที่ต้องการให้สร้างค่ามากขึ้น
4. เมื่อค่าที่ได้เป็นไปตามที่ต้องการเลือกค่าที่สังเกตที่ค่า B เป็นต้นไป
5. พิจารณาค่า $\{\theta^{(B+1)}, \theta^{(B+2)}, ..., \theta^{(T)}\}$ ในรูปของตัวอย่างสำหรับวิเคราะห์ฟังก์ชันจากการแจงโดยประสบการณ์
6. สร้างกราฟเพื่อดูลักษณะการแจกแจงของฟังก์ชันการแจกแจงโดยประสบการณ์
7. คำนวณค่าเฉลี่ย ค่ากลาง ค่าส่วนเบี่ยงเบนมาตรฐาน จากฟังก์ชันการแจกแจงโดยประสบการณ์

วิธีการสุ่มตัวอย่างอีกวิธีหนึ่งที่นิยมใช้กรณีที่ทราบการแจกแจงโดยประสบการณ์เมื่อกำหนดค่าของตัวแปรสุ่ม คือวิธีกิบส์ ศึกษาโดย (Geman and Geman, 1984) มีขั้นตอนดังนี้คือ

1. กำหนดค่าเริ่มต้น $\theta^{(0)}$ จากฟังก์ชันการแจกแจงโดยหลักเกณฑ์
2. เมื่อ t = 1,2,…,T ทำซ้ำตามขั้นตอน

2.1 กำหนด $\theta = \theta^{(t-1)}$

2.2 สร้างค่า θ_j จาก $\theta_j \sim f(\theta_j \mid \theta_1, ..., \theta_{j-1}, \theta_{j+1}, ...\theta_d, \underset{\sim}{x})$ เมื่อ j=1,…,d

2.3 กำหนดให้ค่า $\theta^{(t)} = \theta$ โดยเก็บค่านี้ไว้สร้างพารามิเตอร์ใหม่ที่รอบ t+1 ตามกระบวนการนี้

$\theta_1^{(t)}$ จาก $f(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}, \underset{\sim}{x})$

$\theta_2^{(t)}$ จาก $f(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}, \underset{\sim}{x})$

$\theta_3^{(t)}$ จาก $f(\theta_3 \mid \theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)}, ..., \theta_p^{(t-1)}, \underset{\sim}{x})$

$\theta_j^{(t)}$ จาก $f(\theta_j \mid \theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{j-1}^{(t-1)}, \theta_{j+1}^{(t-1)}, ..., \theta_p^{(t-1)}, \underset{\sim}{x})$

$\theta_p^{(t)}$ จาก $f(\theta_p \mid \theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{p-1}^{(t)}, \underset{\sim}{x})$

ค่าพารามิเตอร์ที่ได้จากการสุ่มตัวอย่างแบบกิบส์สามารถสรุปได้ว่า

$$f(\theta_j \mid \theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{j-1}^{(t-1)}, \theta_{j+1}^{(t-1)}, ..., \theta_p^{(t-1)}, \underset{\sim}{x}) \propto f(\theta \mid \underset{\sim}{x})$$

เมื่อได้ค่าพารามิเตอร์ที่สร้างจากวิธีของ มาร์คอฟ เชน และ การสุ่มตัวอย่างแบบกิบส์ เพื่อประมาณค่าพารามิเตอร์ในสถิติแบบเบส์โดยใช้วิธีการประมาณค่าด้วยมอนติคาร์โล (Monte Carlo) เพื่อประมาณค่าจากฟังก์ชันการแจกแจงภายหลัง จาก

$$E(g(\theta) \mid x_1, ..., x_n) = \int f(\theta \mid x_1, ..., x_n) g(\theta)\, d\theta$$

ซึ่งสามารถประมาณค่า $\overline{\theta}$ ได้โดย

$$\overline{\theta} = \frac{1}{T} \sum_{t=1}^{T} \theta^{(t)}$$

ตัวอย่างกรณีตัวแปรสุ่มแบบต่อเนื่อง (Ntzoufras, 2009)

สมมติตัวแปร X มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 ซึ่งเป็นค่าพารามิเตอร์ θ ที่ต้องการประมาณสามารถเขียนได้ในรูปแบบ

$$x_i \sim Normal(\mu, \sigma^2) \quad , i = 1, ..., n$$

โดยฟังก์ชันภาวะความน่าจะเป็น (Likelihood Function) สามารถเขียนได้เป็น

$$f(\underset{\sim}{x} \mid \mu, \sigma^2) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right) \right\}$$

$$= (\sqrt{2\pi}\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 \right)$$

$$= (\sqrt{2\pi}\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (\sum_{i=1}^{n} x_i^2 + n\mu^2 - 2n\mu\overline{x}) \right)$$

เมื่อพิจารณาฟังก์ชันการแจกแจงโดยหลักเกณฑ์เป็นการแจกแจงแบบปกติที่ค่าเฉลี่ย μ_0 และค่าความแปรปรวน σ_0^2 $(\mu \sim Normal(\mu_0, \sigma_0^2))$ ซึ่งเรียกว่ารูปแบบการแจกแจงก่อนแบบคู่สังยุค (Conjugate prior distribution) เมื่อฟังก์ชันภาวะความน่าจะเป็นเป็นการแจกแจงปกติ และฟังก์ชันการแจกแจงโดยหลักเกณฑ์เป็นการแจกแจงแบบปกติ ซึ่งจะสามารถเขียนฟังก์ชันการแจกแจงโดยประสบการณ์คือ

$$f(\mu \mid \sigma^2, \underset{\sim}{x}) = f(\underset{\sim}{x} \mid \mu, \sigma^2) g(\mu \mid \sigma^2)$$

$$\propto (\sqrt{2\pi}\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (\sum_{i=1}^{n} x_i^2 + n\mu^2 - 2n\mu\overline{x}) \right)$$

$$\times (\sqrt{2\pi}\sigma_0^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_0^2} (\mu-\mu_0)^2 \right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(n\mu^2 - 2n\mu\overline{x}) - \frac{1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0) \right)$$

$$\propto \exp\left(-\frac{1}{2} \left\{ \left[\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right] \mu^2 - 2\mu \left[\frac{n\,\overline{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right] \right\} \right)$$

$$\propto \exp\left(-\frac{1}{2} \left[\frac{n\sigma_0^2 + \sigma^2}{\sigma^2\sigma_0^2} \right] \left\{ \mu^2 - 2\mu \left[\frac{n\,\overline{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \right] \right\} \right)$$

หรือค่าพารามิเตอร์ μ สามารถเขียนในรูปการแจกแจงแบบปกติได้

$$\mu \mid \sigma^2, \underset{\sim}{x} \sim Normal\left(\frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right)$$

ค่าเฉลี่ยของ μ จากการแจกแจงโดยประสบการณ์คือ

$$E(\mu \mid \underset{\sim}{x}) = \frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

ตัวอย่างจากการจำลองข้อมูลเพื่อทดสอบโปรแกรมวินบั๊กและชุดคำสั่ง R2WinBUGS

จำลองตัวแปร X จากการแจงแจงแบบปกติงที่ค่าเฉลี่ย (μ) เท่ากับ 1 ความแปรปรวน (σ^2) เท่ากับ 0.5 มา 50 ค่าสังเกตดังนี้

1.567	1.470	0.709	-0.253	1.372	2.143	0.747	1.073	0.840	0.434
1.659	1.027	0.744	0.634	1.598	1.266	1.548	0.596	1.263	2.087
0.178	1.114	2.138	0.105	0.531	1.927	0.156	0.303	0.928	0.411
1.588	0.862	1.415	0.867	-0.066	1.577	0.002	0.851	0.812	1.667
-0.107	0.418	0.705	1.412	0.166	1.018	1.012	1.779	1.183	1.814

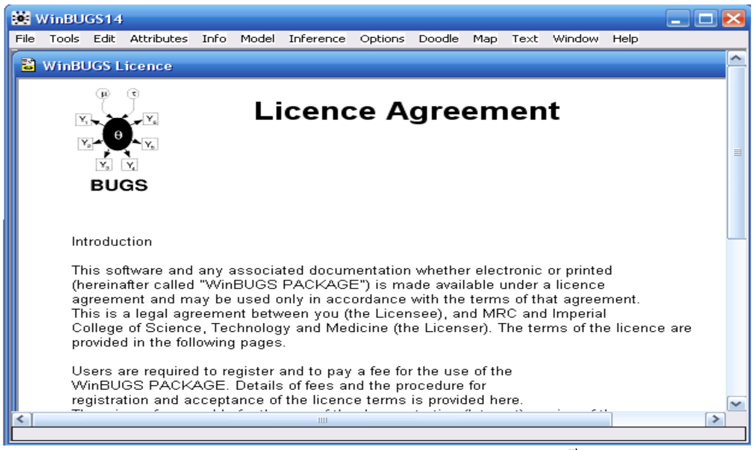
พบว่าเมื่อนำค่าสังเกตทั้ง 50 ค่ามาหาค่าเฉลี่ย ($\bar{x}$) ได้เท่ากับ 0.985 และกำหนดรูปแบบการแจกแจงของค่าสังเกตดังนี้

$$x_i \sim Normal(\mu, \sigma^2) \quad , i = 1, \ldots, n$$
$$\mu \sim Normal(\mu_0, \sigma_0^2), \quad \mu_0 = 0, \quad \sigma_0^2 = 100$$

แต่ในโปรแกรมวินบั๊กกำหนดว่า $\mu \sim Normal(0, 0.01) \sim Normal(mu, tau)$ โดยที่ค่า

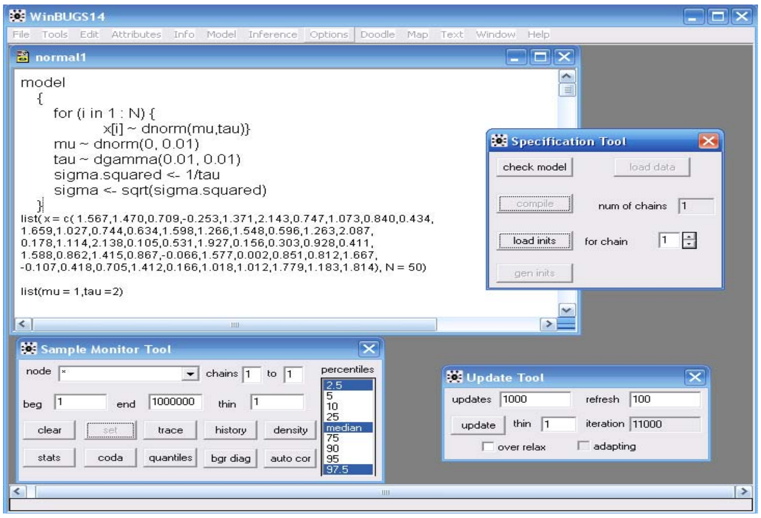
$$tau = \frac{1}{\sigma^2} = \frac{1}{100} = 0.01$$

เมื่อเลือกไอคอนที่ดาวน์โหลดจากโปรแกรมวินบั๊กจะปรากฏหน้าแรกของโปรแกรมดังรูป 1



รูป 1 หน้าแรกในการแสดงโปรแกรมวินบั๊ก

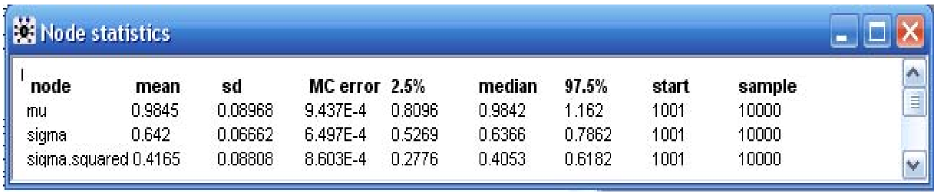
ผู้ใช้สามารถเขียนในโปรแกรมวินบั๊กโดยตัวแบบจะอยู่ในเครื่องหมาย { } ได้ดังรูป 2



รูป 2 หน้าต่างสำหรับประมวลผลการวิเคราะห์เบส์จากโปรแกรมวินบั๊ก

วิธีการประมวลผลสามารถอ่านได้จากหนังสือ Bayesian Modeling Using WinBUGS หรือที่เว็บไซต์ที่ดาวน์โหลดโปรแกรมวินบั๊ก นอกจากนี้ที่มีรายละเอียดและขั้นตอนที่เป็นภาษาไทยสามารถศึกษาได้จากบทความเรื่อง การหาตัวประมาณเบส์ด้วยโปรแกรมวินบั๊ก (อัชฌา อระวีพร, 2554)

นำตัวแบบที่สร้างมาจากฟังก์ชันภาวะความน่าจะเป็นและฟังก์ชันการแจกแจงโดยหลักเกณฑ์ เมื่อตัวแบบถูกต้อง ทำการเรียกข้อมูลมาประมวลผลที่ 10000 รอบจะได้ผลลัพธ์ดังรูป 3

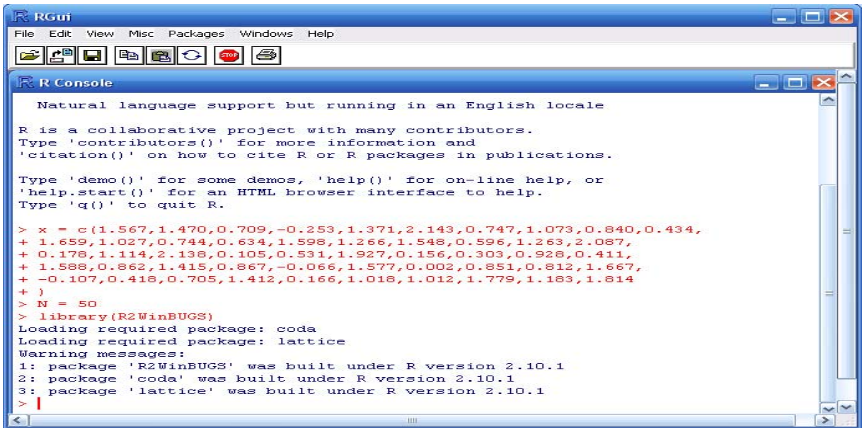


node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mu	0.9845	0.08968	9.437E-4	0.8096	0.9842	1.162	1001	10000
sigma	0.642	0.06662	6.497E-4	0.5269	0.6366	0.7862	1001	10000
sigma.squared	0.4165	0.08808	8.603E-4	0.2776	0.4053	0.6182	1001	10000

รูป 3 หน้าต่างแสดงผลลัพธ์ที่ได้จากประมวลผลการวิเคราะห์เบส์ด้วยโปรแกรมวินบั๊ก

บางครั้งในการทำงานที่ต้องจำลองข้อมูลหลายๆรอบและประมวลผลจากโปรแกรมอาร์แต่มีความจำเป็นต้องวิเคราะห์ด้วยโปรแกรมวินบั๊ก ซึ่งสามารถเชื่อมระหว่างสองโปรแกรมด้วยชุดคำสั่ง R2WinBUGS ในโปรแกรมอาร์ จะเริ่มต้นการทำงานด้วยเครื่องหมาย > เสมอก่อนที่จะเขียนคำสั่งลงไป ตามขั้นตอนดังนี้

1. เรียกข้อมูลจากไฟล์หรือสร้างข้อมูลจากการแจกแจงที่ต้องการและชุดคำสั่ง R2WinBUGS ด้วย library(R2WinBUGS) ดังรูป 4



รูป 4 หน้าจอของโปรแกรมอาร์และการเรียกชุดคำสั่ง R2WinBUGS

2. สร้างฟังก์ชันสำหรับสร้างตัวแบบเพื่อวิเคราะห์ตัวแบบเบส์ซึ่งสามารถนำคำสั่งมาจากโปรแกรมวินบั๊กโดยตั้งชื่อได้ตามที่ต้องการ

```
> bayesian <- function(){
+ for (i in 1 : N) {
+ x[i] ~ dnorm(mu,tau)}
+ mu ~ dnorm(0, 0.01)
+ tau ~ dgamma(0.01, 0.01)
+ sigma.squared <- 1/tau
+ sigma <- sqrt(sigma.squared)
+ }
```

3. กำหนดชื่อของฟังก์ชันในข้อ 2 โดยการสร้างที่สำหรับจัดเก็บไว้จากคำสั่ง

```
> filename <- file.path(tempdir(), "bayesian.bug")
> write.model(bayesian, filename)
```

4. กำหนดตัวแปรที่จะนำเข้ามาประมวลผลในโปรแกรมวินบั๊กจากคำสั่ง จากตัวอย่างนำตัวแปร x และ ค่า N ด้วยคำสั่ง

```
> data <- list("x", "N")
```

5. สร้างค่าเริ่มต้นในรูปของฟังก์ชันด้วยคำสั่ง

```
> inits <- function(){list(mu = 1, tau = 2)}
```

6. กำหนดค่าพารามิเตอร์ที่ต้องการให้ประมวลผล ซึ่งจากตัวอย่างต้องการค่า μ, σ และ σ^2 จากคำสั่ง

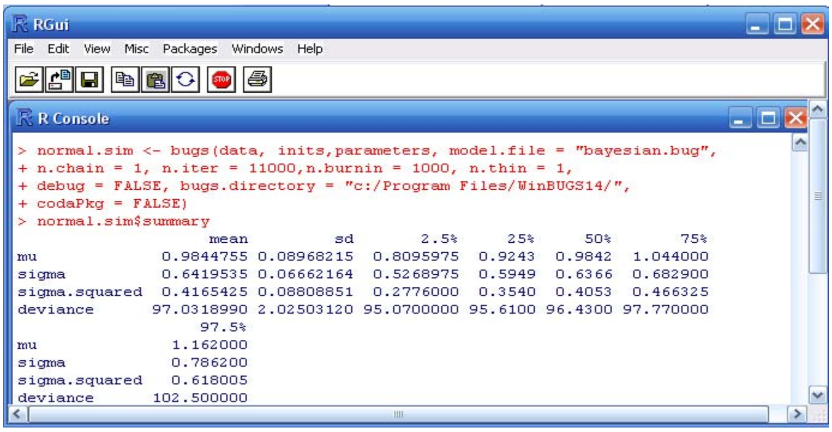
```
> parameters <- list("mu", "sigma", "sigma.squared")
```

7. ให้โปรแกรมอาร์เริ่มการประมวลผลจากโปรแกรมวินบั๊กโดยใส่ค่าที่กำหนดไว้ในข้อ 4 5 และ 6 โดยใช้คำสั่ง

```
> normal.sim <- bugs(data, inits,parameters, model.file = "bayesian.bug",
+ n.chain = 1, n.iter = 11000,n.burnin = 1000, n.thin = 1,
+ debug = TRUE, bugs.directory = "c:/Program Files/WinBUGS14/",
+ codaPkg = FALSE)
```

8. หลังจากโปรแกรมอาร์เรียกโปรแกรมวินบั๊กเพื่อประมวลผลถ้าไม่มีข้อผิดพลาดใดๆหน้าจอจะแสดงกลับที่โปรแกรมอาร์ ให้ใช้คำสั่ง

> normal.sim$summary เพื่อแสดงค่าประมาณพารามิเตอร์ที่ได้จากโปรแกรมวินบั๊กดังรูป 5



รูป 5 ผลลัพธ์ที่ได้จากโปรแกรมอาร์โดยชุดคำสั่ง R2WinBUGS

จากรูป 5 พบว่าตัวประมาณที่ได้มีค่าเท่ากับค่าที่ได้จากโปรแกรมวินบั๊ก และโปรแกรมอาร์โดยใช้ชุดคำสั่งR2WinBUGS ได้ค่าเท่ากัน โดยตัวประมาณการแจกแจงโดยหลักเกณฑ์ μ มีค่าเท่ากับ 0.984 ส่วนค่า σ มีค่าเท่ากับ 0.641 และ ค่า σ^2 มีค่าเท่ากับ 0.416

จากตัวอย่างข้างบนตัวประมาณที่ได้จากการแจกแจงโดยหลักเกณฑ์ μ มีรูปแบบเป็นการแจกแจงแบบปกติที่

$$\mu \mid \sigma^2, \underset{\sim}{x} \sim Normal\left(\frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right)$$

ค่าเฉลี่ยของ μ จากการแจกแจงโดยประสบการณ์คือ

$$E(\mu \mid \underset{\sim}{x}) = \frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{(50 \times 0.985 \times 100) + (0 \times 0.416)}{(50 \times 100) + 0.416}$$

$$= 0.984$$

ค่าที่ได้จากตัวประมาณเบส์จากฟังก์ชันการแจกแจงโดยหลักเกณฑ์ให้ค่าใกล้เคียงกับค่าพารามิเตอร์ที่ประมาณได้จากโปรแกรมวินบั๊กและโปรแกรมอาร์โดยใช้ชุดคำสั่ง R2WinBUGS

ตัวอย่างกรณีตัวแปรสุ่มแบบไม่ต่อเนื่อง (Ntzoufras, 2009)

สมมติตัวแปร X มีการแจกแจงแบบทวินาม เป็นจำนวนความสำเร็จที่เกิดขึ้นจากการกระทำ N_i ครั้ง ด้วยค่าความน่าจะเป็นที่จะเกิดความสำเร็จ p เป็นค่าพารามิเตอร์ θ ที่ต้องการประมาณ สามารถเขียนได้ในรูปแบบ

$$x_i \sim Binomial(p) \quad ,i = 1,...,n$$

โดยฟังก์ชันภาวะความน่าจะเป็น (Likelihood Function) สามารถเขียนได้เป็น

$$\begin{aligned} f(\underset{\sim}{x} \mid p) &= \prod_{i=1}^{n} \left\{ \binom{N_i}{x_i} p^{x_i} (1-p)^{N_i - x_i} \right\} \\ &= \prod_{i=1}^{n} \binom{N_i}{x_i} p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i} \\ &= \prod_{i=1}^{n} \binom{N_i}{x_i} p^{n\bar{x}} (1-p)^{N-n\bar{x}} \end{aligned}$$

เมื่อ $N = \sum_{i=1}^{n} N_i$ คือจำนวนของการทดลองเบอร์นูลลี่ในแต่ละครั้ง พิจารณาฟังก์ชันการแจกแจงโดยหลักเกณฑ์เป็นการแจกแจงแบบเบต้ามีค่าพารามิเตอรคือ a และ b $(p \sim Beta(a,b))$ มีฟังก์ชันการแจกแจงความน่าจะเป็นดังนี้

$$g(p) \quad = \quad \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} p^{a-1}(1-p)^{b-1} \quad , \quad 0 < p < 1$$

ซึ่งเรียกว่ารูปแบบการแจกแจงก่อนแบบคู่สังยุค (Conjugate prior distribution) เมื่อฟังก์ชันภาวะความน่าจะเป็นเป็นการแจกแจงทวินาม และฟังก์ชันการแจกแจงโดยหลักเกณฑ์เป็นการแจกแจงแบบเบต้า ซึ่งจะสามารถเขียนฟังก์ชันการแจกแจงโดยประสบการณ์คือ

$$\begin{aligned} f(p \mid \underset{\sim}{x}) &= f(\underset{\sim}{x} \mid p) g(p) \\ &\propto \prod_{i=1}^{n} \binom{N_i}{x_i} p^{n\bar{x}} (1-p)^{N-n\bar{x}} \times \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} p^{a-1}(1-p)^{b-1} \\ &\propto p^{n\bar{x}+a-1}(1-p)^{N-n\bar{x}+b-1} \end{aligned}$$

หรือค่าพารามิเตอร์ p สามารถเขียนในรูปการแจกแจงเบต้าได้

$$p \mid \underset{\sim}{x} \sim Beta(n\bar{x}+a, N-n\bar{x}+b)$$

ค่าเฉลี่ยของ p จากการแจกแจงโดยประสบการณ์คือ

$$E(p\,|\,\underset{\sim}{x}) = \frac{n\bar{x}+a}{N+a+b}$$

ตัวอย่างจากการจำลองข้อมูลเพื่อทดสอบโปรแกรมวินบั๊กและชุดคำสั่ง R2WinBUGS

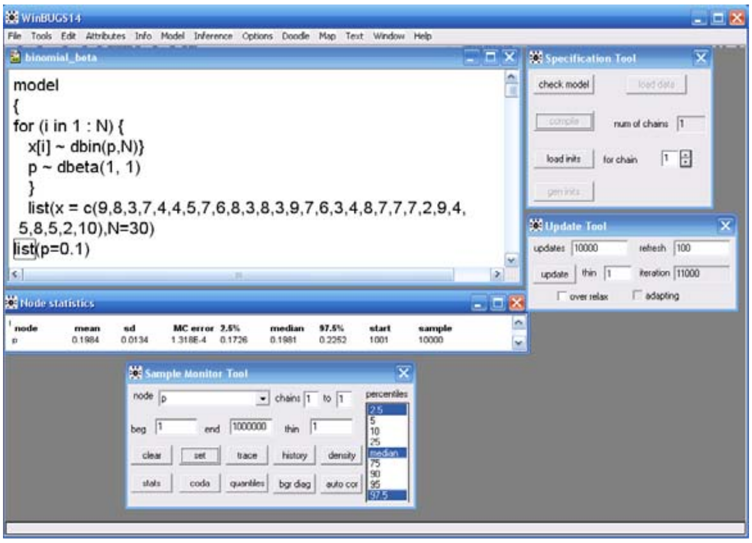
จำลองตัวแปร X จากการแจงแจงทวินามที่ค่าความน่าจะเป็นที่จะเกิดความสำเร็จ 0.2 มา 30 ค่าสังเกต โดยในแต่ละค่าสังเกตเป็นการทดลองแบบเบอร์นูลลี่ 30 ครั้ง ดังนี้

9	8	3	7	4	4	5	7	6	8
3	8	3	9	7	6	3	4	8	7
7	7	2	9	4	5	8	5	2	10

พบว่าเมื่อนำค่าสังเกตทั้ง 30 ค่ามาหาค่าเฉลี่ย ($\bar{x}$) ได้เท่ากับ 5.933 และกำหนดรูปแบบการแจกแจงของค่าสังเกตดังนี้

$$x_i \sim Binomial(p) \quad ,i=1,\ldots,n$$
$$p \sim Beta(a,b), \quad a=1, \quad b=1$$

นำข้อมูลที่จำลองมาวิเคราะห์ด้วยโปรแกรมวินบั๊กโดยดาวน์โหลดจากเว็บไซต์ที่กล่าวมาแล้ว และเขียนรูปแบบฟังก์ชันพร้อมประมวลผลได้ผลลัพธ์ดังรูป 6



รูป 6 หน้าต่างแสดงผลลัพธ์ที่ได้จากประมวลผลการวิเคราะห์เบส์ด้วยโปรแกรมวินบั๊ก

จากรูป 6 พบว่าตัวประมาณที่ได้จากโปรแกรมวินบั๊ก ให้ค่า p เท่ากับ 0.1984 และเมื่อใช้โปรแกรมอาร์โดยเรียกชุดคำสั่ง R2WinBUGS ความหมายได้กล่าวมาแล้วในหัวข้อ 3.2 จะได้ผลลัพธ์ดังรูป 7 ซึ่งได้ค่า p เท่ากับ 0.2 ซึ่งมีค่าใกล้เคียงกันเนื่องจากอาจเกิดจากการปัดตำแหน่งของทศนิยม

จากตัวอย่างที่ 4 ตัวประมาณที่ได้จากการแจกแจงโดยหลักเกณฑ์ p มีรูปแบบเป็นการแจกแจงแบบเบต้าที่

$$p \mid \underset{\sim}{x} \sim Beta(n\bar{x}+a, N-n\bar{x}+b)$$

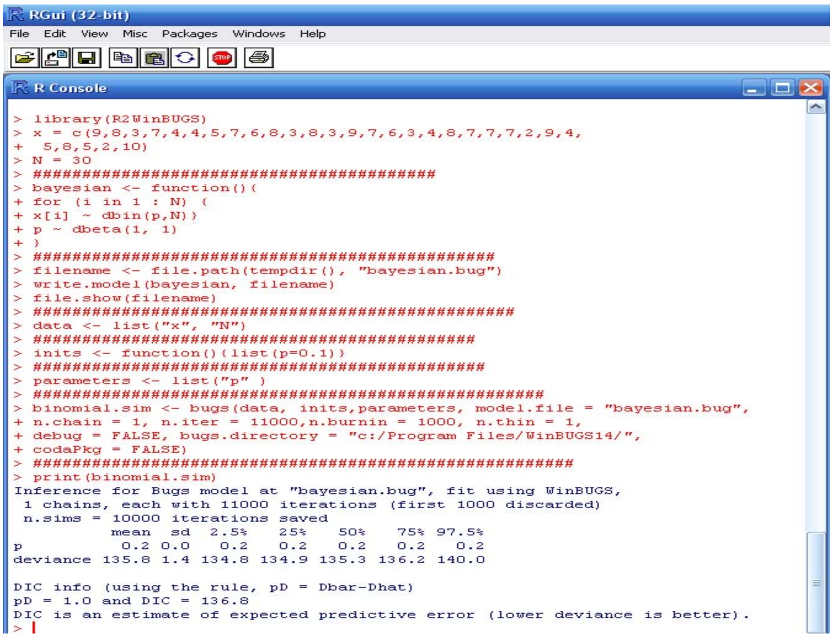
ค่าเฉลี่ยของ p จากการแจกแจงโดยประสบการณ์คือ

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

$$= \frac{(30\times 5.933)+1}{(30\times 30)+1+1}$$

$$= 0.1984$$

ค่าที่ได้จากตัวประมาณเบส์จากฟังก์ชันการแจกแจงโดยหลักเกณฑ์ให้ค่าใกล้เคียงกับค่าพารามิเตอร์ที่ประมาณ ได้จากโปรแกรมวินบั๊กและโปรแกรมอาร์โดยใช้ชุดคำสั่ง R2WinBUGS


```
> library(R2WinBUGS)
> x = c(9,8,3,7,4,4,5,7,6,8,3,8,3,9,7,6,3,4,8,7,7,7,2,9,4,
+  5,8,5,2,10)
> N = 30
> ############################################
> bayesian <- function(){
+ for (i in 1 : N) {
+ x[i] ~ dbin(p,N)}
+ p ~ dbeta(1, 1)
+ }
> ##################################################
> filename <- file.path(tempdir(), "bayesian.bug")
> write.model(bayesian, filename)
> file.show(filename)
> ####################################################
> data <- list("x", "N")
> ################################################
> inits <- function(){list(p=0.1)}
> #################################################
> parameters <- list("p" )
> ######################################################
> binomial.sim <- bugs(data, inits,parameters, model.file = "bayesian.bug",
+ n.chain = 1, n.iter = 11000,n.burnin = 1000, n.thin = 1,
+ debug = FALSE, bugs.directory = "c:/Program Files/WinBUGS14/",
+ codaPkg = FALSE)
> #######################################################
> print(binomial.sim)
Inference for Bugs model at "bayesian.bug", fit using WinBUGS,
 1 chains, each with 11000 iterations (first 1000 discarded)
 n.sims = 10000 iterations saved
          mean  sd  2.5%   25%   50%   75% 97.5%
p          0.2 0.0   0.2   0.2   0.2   0.2   0.2
deviance 135.8 1.4 134.8 134.9 135.3 136.2 140.0

DIC info (using the rule, pD = Dbar-Dhat)
pD = 1.0 and DIC = 136.8
DIC is an estimate of expected predictive error (lower deviance is better).
>
```

รูป 7 ผลลัพธ์ที่ได้จากโปรแกรมอาร์โดยชุดคำสั่ง R2WinBUGS

สรุป

การประมาณค่าเบส์ด้วยวิธีการทางสถิติมีความยุ่งยากในการคำนวณและการพิสูจน์ เนื่องจากแต่ละตัวแปรอาจมีมีความเกี่ยวข้องกับรูปแบบการแจกแจงหลายรูปแบบ การใช้โปรแกรมสำเร็จรูปมาช่วยในการวิเคราะห์ทำให้มีความสะดวกมากยิ่งขึ้น โดยเฉพาะ โปรแกรมอาร์เป็นโปรแกรมที่นิยมใช้ในต่างประเทศเนื่องจากผู้ใช้สามารถดาวน์โหลดโปรแกรมมาใช้ได้โดยไม่เสียค่าใช้จ่ายและไม่มีปัญหาเรื่องลิขสิทธิ์ ซึ่งได้มีการร่วมมือจากองค์กรต่างๆทั่วโลกที่สร้างโปรแกรมวิเคราะห์ข้อมูลทางสถิติ และแจกจ่ายแก่ผู้สนใจฟรี แต่โปรแกรมนี้ค่อนข้างใช้ยาก ผู้ใช้ไม่คุ้นเคยจึงไม่ได้รับความนิยมในประเทศไทยเนื่องจากทุกคำสั่งเป็นภาษาอังกฤษ แต่โปรแกรมอาร์มีประสิทธิภาพและมีความหลากหลายในการใช้งานทางสถิติที่ตัวแบบที่ยุ่งยากซับซ้อน ส่วนโปรแกรมวินบั๊กเป็นแก้ปัญหาการประมาณค่าจากตัวประมาณเบส์เท่านั้นแต่โปรแกรมอาร์สามารถใช้งานได้หลากหลายมากกว่าดังนั้นการที่สามารถเรียกโปรแกรมวินบั๊กเพื่อมาประมวลผลในโปรแกรมอาร์ช่วยให้การทำงานที่ต้องมีการจำลองปัญหาซ้ำหลายๆรอบเพื่อเก็บข้อมูลในโปรแกรมอาร์สะดวกมากขึ้นและได้ผลลัพธ์ใกล้เคียงกัน การใช้ชุดคำสั่ง R2WinBUGS ในโปรแกรมอาร์ ผู้ใช้ควรจะรู้ความหมายในคำสั่งต่างๆที่เขียนขึ้นในการเรียกโปรแกรมวินบั๊กมาประมวลผล ซึ่งจะประโยชน์ในการนำโปรแกรมทั้งสองไปใช้ประยุกต์ตัวสถิติเบส์กับงานด้านอื่นๆได้มากยิ่งขึ้น

เอกสารอ้างอิง

อัชฌา อระวีพร. (2554). การหาตัวประมาณเบส์ด้วยโปรแกรมวินบั๊ก. *วารสารวิทยาศาสตร์ลาดกระบัง. 20(*2), 45-60.

Becker, R.A. and Chambers, J.M. (1984). *S. An Interactive Environment for Data Analysis and Graphics*. Wadsworth and Brooks/Cole, Monterey.

Becker, R. A., Chambers, J. M., and Wilk, A. R. (1988). *The NEW S Language-A Programming Environment for Data Analysis and Graphics.* Chapman & Hall, New York.

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data analysis*. Chapman & Hall / CRC, Florida.

Casella, G. and George , E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46, 167-174.

Chambers, J. M. (1998*). Programming with Data. A Guide to the S Language*. Springer-Verlag, New York.

Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S.* Chapman & Hall, NewYork.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*, 398-409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transactions on Pattern analysis and Machine Intelligence*, *6*, 721-741.

Gilk, W., Richardson, S. and Spiegelhalter. D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.

Plummer, M., Best, N.G., Cowles, K., and Vines, K. (2004). *coda : Output Analysis and Diagnostics for MCMC*. R package version 0.9-1, URL http://www-fis.iarc.fr/coda/.

R Development Core Team. (2004). *R: A Language and Environment for Statistical Computing*. R Foundation forStatistical Computing, Vienna, Austria.

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.