# การคัดเลือกตัวแบบทางสถิติด้วยแนวคิดแบบเบส์

## Statistical Model Selection by Bayesian Approach

**พจนา แววสวัสดิ์***

## บทคัดย่อ

การคัดเลือกตัวแบบเป็นวิธีการหนึ่งที่สำคัญในการวิเคราะห์ทางด้านสถิติ ในปัจจุบันการคัดเลือกตัวแบบมีหลายวิธีการด้วยกัน วิธีการหนึ่งที่ถูกนำมาประยุกต์ใช้คือ การคัดเลือกตัวแบบด้วยวิธีการแบบเบส์ เป็นวิธีการคัดเลือกตัวแบบด้วยการใช้การแจกแจงก่อนปรับ รวมกับฟังก์ชันภาวะน่าจะเป็นในการกำหนดการแจกแจงหลังปรับ บทความนี้ ได้กล่าวถึงการแจกแจงก่อนปรับ ในกรณีที่นักวิจัยไม่ทราบทราบข้อสนเทศก่อนปรับ สำหรับใช้ในการอธิบายสภาวะการณ์ที่มีความไม่แน่นอน รวมทั้งใช้ในการพิจารณาตัวแบบสำหรับการพยากรณ์ ด้วยหลักการที่ว่าตัวแบบใดให้ค่าความน่าจะเป็นหลังปรับมากที่สุด ตัวแบบนั้นจะเป็นตัวแบบที่ให้ค่าพยากรณ์แม่นยำมากที่สุด

## ABSTRACT

Model Selection is one of significant method for statistical analysis. Currently, there are several methods of model selection. Bayesian Model Selection (BMS),one of among those methods, is applied to select a model by using prior distribution combines with likelihood function in order to design posterior distribution. Prior distribution is proposed in case of researchers who do not know the noninformative prior, to explain unstable situation, beside, to consider the forecasting model. The principle of this method is that the highest posterior probability model will be the best forecasting model.

* อาจารย์ประจำสำนักวิจัย มหาวิทยาลัยศรีปทุม

## บทนำ

ในงานวิจัยหลายสาขาวิชานักวิจัยต้องอาศัยวิธีการทางสถิติมาใช้เป็นเครื่องมือในการหาคำตอบหรือหาตัวแบบที่สนใจ ซึ่งบางครั้งนักวิจัยพบกับปัญหาการคัดเลือกตัวแบบที่เหมาะสมสำหรับนำไปใช้ในการพยากรณ์ ดังนั้นนักวิจัยจึงมีความจำเป็นในการหาวิธีการในการคัดเลือกตัวแบบที่ดีที่สุด วิธีการคัดเลือกตัวแบบที่ถูกนำไปใช้ในการคัดเลือกตัวแบบ ด้วยวิธีการแบบคลาสสิกนั้นตัวแบบที่นำมาเปรียบเทียบต้องเป็นตัวแบบที่ติดกัน (nested model) นั่นคือ ในกรณีที่เราพิจารณาตัวแบบ A ติดกลุ่มกับตัวแบบ B ถ้าตัวแบบ A เป็นส่วนหนึ่งของตัวแบบ B ตัวแบบสองตัวแบบนี้ ถือว่าติดกลุ่มกัน เมื่อพิจารณาในตัวอย่าง (Steiger, Shapiro & Browne, (1985)) ได้ทำการพิจารณาความแตกต่างระหว่าง สถิติทดสอบไคแสควร์ (chi-square test statistics) ที่มีการแจกแจงอย่างเป็นอิสระในระยะอนันต์ (asymptotically independent) เป็นสถิติที่ใช้ในการทดสอบ ถ้าสถิติทดสอบมีการแจกแจงไคแสควร์ถูกนำมาทำการเปรียบเทียบความแตกต่าง ดังนั้นระดับความเป็นอิสระ (degrees of freedom) ของความแตกต่างทั้งสองตัวแบบจะเท่ากับระดับความเป็นอิสระของสองตัวแบบมาลบกัน ทำให้ค่าระดับความเป็นอิสระเป็นบวก ส่วนการคัดเลือกตัวแบบสำหรับการเปรียบเทียบ 2 ตัวแบบที่ไม่ติดกลุ่มกัน (non-nested model) ถ้านำระดับความเป็นอิสระของ 2 ตัวแบบมาลบกัน จะทำให้ระดับความเป็นอิสระไม่สามารถคำนวณได้อย่างถูกต้อง นอกจากนี้ตัวแบบที่นำมาทำการเปรียบเทียบต้องมีการแจกแจงในรูปแบบเดียวกันอีกด้วย ด้วยข้อจำกัดของวิธีการแบบคลาสสิกปัจจุบันมีนักสถิติหลายท่านได้นำเสนอการคัดเลือกตัวแบบด้วยหลักการของเบส์มาใช้ในการคัดเลือกตัวแบบ เนื่องจากหลักการของเบส์จะนำมาตัวทุกๆ ตัวแบบมาเปรียบเทียบกันด้วยค่าความน่าจะเป็น โดยความน่าจะเป็นที่นำมาทำการเปรียบเทียบคือ ความน่าจะเป็นหลังปรับ (posterior probability) แล้วถ่วงน้ำหนักด้วยความน่าจะเป็นก่อนปรับ (prior probability)

การนำสถิติประยุกต์ไปใช้ในการวิเคราะห์ข้อมูลสิ่งหนึ่งที่ถูกนำมาใช้ก็คือ การหาค่าความน่าจะเป็นของตัวแบบ สำหรับข้อมูลหรือตัวแบบที่นำมาใช้ในการวิเคราะห์ การหาความน่าจะเป็นของตัวแบบที่นำมาทำการวิเคราะห์ จะเริ่มจากการแทนกลุ่มของข้อมูลเป็นตัวแปรสุ่ม แล้วหาตัวแบบที่เหมาะสม (fit a model) นั่นก็หมายความว่า นักวิจัยต้องทำการประมาณการแจกแจง ความน่าจะเป็นของตัวแปรสุ่ม สำหรับตัวแบบที่นำมาทำการพิจารณา หลักการหนึ่งที่ถูกนำมาประยุกต์ใช้ คือ การคัดเลือกตัวแบบ (Model Selection) การคัดเลือกตัวแบบ เป็นวิธีการที่นำไปใช้ในการคัดเลือกตัวแบบบนพื้นฐานของข้อมูลที่มีอยู่จากตัวแบบที่เป็นไปได้ทั้งหมด ซึ่งการคัดเลือกตัวแบบด้วยแนวคิดของเบส์เป็นหลักการที่ใช้ทฤษฎีของเบส์ (Baye's Theory) มาเป็นหลักในการคำนวณ เพื่อคำนวณหาค่าความน่าจะเป็นภายหลังปรับ ที่ได้จากการถ่วงน้ำหนักด้วยค่าความน่าจะเป็นก่อนปรับของแต่ละตัวแบบ แล้วนำความน่าจะเป็นภายหลังปรับมาใช้ในการคัดเลือกตัวแบบ โดยการคัดเลือกตัวแบบด้วยวิธีการของเบส์นี้ จะคัดเลือกจากค่าความน่าจะเป็นภายหลังปรับมากที่สุด เป็นตัวแบบที่ดีที่สุด

การคัดเลือกตัวแบบ (Model Selection) เป็นวิธีการที่นำมาใช้ในการคัดเลือกตัวแบบที่ดีที่สุดเพียงตัวแบบเดียว ซึ่งบทความนี้ได้อธิบายถึงการคำนวณการแจกแจงภายหลังปรับของตัวแบบที่นำมาทำการคัดเลือกตัวแบบที่มีการแจกแจงในรูปแบบต่างๆ สำหรับการแจกแจงก่อนปรับ ในบทความนี้ ผู้เขียนพิจารณาเฉพาะกรณีที่นักวิจัยไม่ทราบทราบข้อสนเทศก่อนปรับ (Non-informative prior) เพื่อใช้ในการอธิบายสภาวการณ์

ที่มีความไม่แน่นอน รวมทั้งใช้ในการพิจารณาตัวแบบสำหรับการพยากรณ์

## ปัจจัยเบส์ (Baye's Factor)

การนำหลักการของเบส์ไปใช้ในการคัดเลือกตัวแบบจะเป็นการพิจารณาโดยใช้การแจกแจงความน่าจะเป็นก่อนปรับ อธิบายความไม่แน่นอนของปัญหาที่ทำการพิจารณา รวมทั้งใช้ในการพิจารณาหาตัวแบบเพื่อใช้ในการพยากรณ์ โดยมีความน่าจะเป็นหลังปรับเป็นค่าที่ใช้ในการกำหนดในการเลือกตัวแบบ ตัวแบบที่ให้ค่าความน่าจะเป็นหลังปรับมากที่สุด จะเป็นตัวแบบที่ดีที่สุด ในการพิจารณาตัวแบบในกรณีที่มีตัวแบบ K ตัวแบบ นั่นคือ $M_1, M_2, \ldots, M_k$ และพิจารณาภายใต้ข้อมูล x ดังนั้นการพิจารณาตัวแบบ $M_k$ ภายใต้ข้อมูล x จะมีความหนาแน่นน่าจะเป็น $p(x|\Theta_k, M_k)$ เมื่อ $\Theta_k$ แทนเวกเตอร์พารามิเตอร์ที่ไม่ทราบค่า จากที่กล่าวมาข้างต้นหลักการของเบส์จะมีการกำหนดการแจกแจงความน่าจะเป็นก่อนปรับ $p(\theta_k|M_k)$ ให้กับพารามิเตอร์ในแต่ละตัวแบบ และค่าความน่าจะเป็นก่อนปรับ $p(M_k)$ ให้กับแต่ละตัวแบบ การพิจารณาการกำหนดค่าความน่าจะเป็นก่อนปรับ สำหรับนำไปหาค่าความน่าจะเป็นหลังปรับของตัวแบบ $M_k$ ได้มาจากทฤษฎีของเบส์ โดยมีสูตรการคำนวณดังนี้

$$p(M_k|x) = \frac{p(x|M_k)p(M_k)}{\sum_{j=1}^{k} p(x|M_j)p(M_j)} \qquad (1)$$

เมื่อ $p(x|M_k) = p(x|\Theta_k, M_k)p(\theta_k, M_k)d\theta_k$ (2)

สมการที่ (2) คือ การแจกแจงชายขอบ (Marginal Distribution) หรือ ค่าคาดหมายความหนาแน่นของ x ภายใต้ตัวแบบ $M_k$ โดย $p(x|\Theta_k, M_k)$ พิจารณาจากฟังก์ชันของ $\Theta_k$ ซึ่งเป็นฟังก์ชันของ $\Theta_k$ ภายใต้ตัวแบบ $M_k$ $p(x|M_k)$ พิจารณาจากค่าชายขอบ (Marginal) หรือการรวมฟังก์ชันภาวะน่าจะเป็น การแจกแจงภายหลังปรับ $p(M_1|x), \ldots, p(M_k|x)$ ต้องมีการกำหนดให้ชัดเจนในแต่ละตัวแบบ หลังจากที่ได้เก็บรวบรวมข้อมูล x การแจกแจงความน่าจะเป็นภายหลังปรับ $p(M_k|x)$ จะเป็นตัววัดความเหมาะสมของตัวแบบ $M_k$ ดังนั้นการคัดเลือกตัวแบบ $M_k$ ที่มีค่า $p(M_k|x)$ มากที่สุด

การเปรียบเทียบความน่าจะเป็นของตัวแบบด้วยอัตราส่วนแปลกหลังปรับ (posterior odds ratio) จากสมการที่ 1 สามารถคำนวณได้ว่า

$$\frac{p(M_k|x)}{p(M_j|x)} = \frac{p(x|M_k)}{p(x|M_j)} \times \frac{p(M_k)}{p(M_j)} \qquad (3)$$

ดังนั้น เมื่อทำการรวมอัตราส่วนภาวะน่าจะเป็นในสมการที่ 3 ได้ว่า

$$B_{kj}(x) = \frac{p(x|M_k)}{p(x|M_j)} \qquad (4)$$

ผลจากการคำนวณในสมการที่ 4 เรียกว่าปัจจัยเบส์สำหรับตัวแบบ $M_k$ เทียบกับตัวแบบ $M_j$ ที่ถูกปรับด้วยอัตราส่วนแปลกก่อนปรับ $\frac{p(M_k)}{p(M_j)}$ ที่มีผลต่ออัตราส่วนแปลกหลังปรับ เมื่อ $B_{kj} > 1$ เมื่อข้อมูล x เพิ่มขึ้นความน่า จะเป็น $M_k$ มากกว่า $M_j$ เนื่องจากตัวแบบนั้นไม่ได้ขึ้น อยู่กับส่วนแปลกก่อนปรับ ปัจจุบันปัจจัย เบส์ ถูกนำไปใช้ในการคัดเลือกตัวแบบ $M_k$ และตัวแบบ $M_j$ ภายใต้ข้อมูลชุดเดียวกัน แต่การคำนวณปัจจัยเบส์ขึ้นอยู่กับการแจกแจงก่อนปรับ $p(\Theta_k|M_k)$ และ $p(\Theta_j|M_j)$ สำหรับการประมาณค่าพารามิเตอร์ของแต่ละตัวแบบ ดังนั้นสามารถคำนวณหาความน่าจะเป็นภายหลังปรับสามารถแสดงในเทอมของ

ปัจจัยเบส์ นั่นคือ

$$p(M_k \mid k) = \left[\sum_{j=1}^{k} \frac{p(M_j)}{p(M_k)} B_{jk}(x)\right]^{-1}$$

## การเปรียบเทียบตัวแบบด้วยปัจจัยเบส์

การเปรียบเทียบตัวแบบด้วยแนวคิดของเบส์ จะทำการเปรียบเทียบครั้งละ 2 ตัวแบบ ปัจจัยเบส์ที่นำมาทำการเปรียบเทียบจะทำการคำนวณจากอัตราส่วนแปลกภายหลังปรับของสองสมมติฐาน เมื่อความน่าจะเป็นก่อนของสองสมมติฐาน คือ $\frac{1}{2}$ ซึ่งจะเป็นตัวที่ใช้วัดความชัดเจนในการยอมรับสมมติฐาน

## นิยาม

กำหนดข้อมูล X ภายใต้การทดสอบสองสมมติฐาน $H_0$ และ $H_1$ มีความหนาแน่นน่าจะเป็น $p(X \mid H_0)$ หรือ $p(X \mid H_1)$ และกำหนดการแจกแจงก่อน $p(H_0)$ และ $p(H_1) = 1\text{-}p(H_0)$ จะได้ความน่าจะเป็นภายหลัง คือ $p(X \mid H_0)$ และ $p(X \mid H_1) = 1 - p(X \mid H_0)$ ดังนั้น แนวคิดของการแจกแจงก่อนที่จะแปลงไปเป็นความน่าจะเป็นภายหลังจากการพิจารณาข้อมูลที่มีอยู่ เราจะทำการแปลงให้อยู่ในรูปแบบที่ง่าย โดยอาศัยอัตราส่วนออดส์ (odds= probability/(1- probability) นั่นคือ

$$p(H_k \mid X) = \frac{p(X \mid H_k)p(H_k)}{p(X \mid H_0)p(H_0) + p(X \mid H_1)p(H_1)} \quad (k = 0,1)$$

ดังนั้นจะได้ว่า

$$\underset{(1)}{\frac{p(H_0 \mid X)}{p(H_1 \mid X)}} = \underset{(2)}{\frac{p(X \mid H_0)}{p(X \mid H_1)}} \times \underset{(3)}{\frac{p(H_0)}{p(H_1)}}$$

โดย (1) แทน Posterior odds (2) แทน Bayes factor และ (3) แทน prior odds

จากสมการข้างต้น เราจะทำการแปลงให้อยู่ในรูปผลคูณอย่างง่าย นั่นคือ

$$B_{01} = \frac{p(H_0 \mid X) / p(H_1 \mid X)}{p(H_0) / p(H_1)}$$

$$= \frac{\dfrac{p(X \mid H_0)p(H_0)}{p(X)} \Big/ \dfrac{p(X \mid H_1)p(H_1)}{p(X)}}{p(H_0) / p(H_1)}$$

$$= \frac{p(X \mid H_0)}{p(X \mid H_1)}$$

โดย $B_{01}$ แทน อัตราส่วนออดส์ภายหลังของ $H_0$ และ $H_1$

## การคำนวณการแจกแจงความน่าจะเป็นภายหลังปรับ

การคำนวณสำหรับการคัดเลือกตัวแบบ ด้วยวิธีการของเบส์ โดยทั่วไปนั้นสูตรการคำนวณจะขึ้นอยู่กับลักษณะการแจกแจงของข้อมูล การพิจารณาปัญหาการคัดเลือกตัวแบบระหว่างกลุ่มของพารามิเตอร์สองกลุ่ม ภายใต้ข้อมูล $x = \{x_1, \ldots, x_n\}$ ซึ่งบทความนี้ผู้เขียนได้นำเสนอหลักการคำนวณความน่าจะเป็นหลักปรับ 3 กรณีด้วยกัน นั่นคือ

**กรณีที่ 1 ตัวแปรสุ่ม x มีการแจกแจงแบบจีออร์เมตริกส์ (geometric distribution)** ดังนั้น

$$p(x \mid \theta_1, M_1) = \pi^n (1-\pi)^s, \; S = \sum_{i=1}^{n} X_i$$

ส่วนตัวแบบที่สอง $M_2$ ที่นำมาเปรียบเทียบตัวแปรสุ่ม x มีการแจกแจงแบบพังซอง (Poisson distribution)

$$p(x \mid \theta_2, M_2) = \frac{e^{-n\lambda}\lambda^{s}}{{}_{i}x_i!}$$

ถ้าเรากำหนดความไม่แน่นอนของ $\theta_1 = \pi$ ภายใต้ตัวแบบ $M_1$ ใช้การแจกแจงก่อนปรับเป็นแบบยูนิฟอร์ม ดังนั้นจะได้ว่า

$$p(\pi \mid M_1) = 1 \text{ เมื่อ } 0 \le \pi \le 1$$

ส่วนความไม่แน่นอนของ $\theta_2 = \lambda$ ภายใต้ตัวแบบ $M_2$ ใช้การแจกแจงก่อนปรับเป็นแบบเอ็กโปแนนเชียล

$$p(\lambda \mid M_2) = e^{-\lambda} \text{ เมื่อ } 0 \le \lambda \le \infty$$

ดังนั้นการกำหนดตัวแบบ สำหรับการแจกแจงชายขอบ

$$p(\pi \mid M_1) = \int_0^1 \pi^{n}(1-\pi)^{s}\, d\pi = \frac{n!\, s!}{(n+s+1)!}$$

และ

$$p(\lambda \mid M_2) = {}_{0}\frac{e^{-(n+1)\lambda}\lambda^{s}}{{}_{i}x_i!}\, d\lambda = \frac{s!}{(n+1)^{s+1}\ {}_{i}x_i!}$$

สำหรับการคัดเลือกความน่าจะเป็นก่อนปรับ $p(M_1)$ และ $p(M_2)$ การแจกแจงหลังปรับสำหรับตัวแบบ $M_1$ สามารถคำนวณจากสูตร

$$p(M_1 \mid X) = \frac{\dfrac{n!}{(n+s+1)!}p(M_1)}{\dfrac{n!}{(n+s+1)!}p(M_1) + \dfrac{n!}{(n+1)^{s+1}\ {}_{i}X_i!}p(M_2)}$$

หลักการของเบส์ สามารถนำไปใช้ในการเปรียบเทียบตัวแบบ ซึ่งสังเกตได้จากการเลือกพารามิเตอร์ก่อนการปรับ (parameter prior) ของ $\pi$ และ $\lambda$ ที่มีผลต่อการคัดเลือกตัวแบบ

สำหรับการพิจารณาปัญหาการคัดเลือกตัวแบบระหว่างการทดสอบสมมติฐานแบบติดกลุ่ม ปัญหาดังกล่าวนี้สามารถพิจารณาด้วยหลัก การคัดเลือกตัวแบบโดยใช้พารามิเตอร์ก่อนปรับของการทดสอบสมติฐาน

**กรณีที่ 2 ตัวแปรสุ่ม x มีการแจกแจงแบบปกติ (normal distribution)**

กำหนดข้อมูล $X = \{x_1, x_2, \ldots, x_n\}$ ประกอบด้วยข้อมูล n ชุด ซึ่งตัวแปรสุ่มสำหรับข้อมูลชุดนี้มีการแจกแจงแบบปกติและทราบค่าความแปรปรวน $\sigma^2$ โดยนำข้อมูลชุดดังกล่าวมาทำการทดสอบสมมติฐานอย่างง่าย $\mu = 0$ และมีสมมติฐานแย้ง $\mu \neq 0$ กับการอธิบายความไม่แน่นอนกับความแปรปรวนขนาดใหญ่ จากข้อมูลข้างต้นเราสามารถกำหนดสูตรการคัดเลือกตัวแบบด้วยแนวคิดของเบส์โดยใช้พารามิเตอร์ที่เหมือนกันจากตัวแบบ $M_1$ และ $M_2$ กล่าวคือ

$$p(x \mid \theta_1, M_1) = p(x \mid \theta_2, M_2) = (2\pi\sigma^2)^{-n/2}\left(\exp - \frac{1}{2\sigma^2}{}_{i=1}^{n}(x_i - \mu)^2\right)$$

และกำหนดความแตกต่างของความน่าจะเป็นก่อนปรับ $\theta_1 = \theta_2 = \mu$ นั่นคือ

$$p(\mu = 0 \mid M_1) = 1 \text{ และ } p(\mu \mid M_2) = (2\pi\tau)^{-n/2}\left(\exp - \frac{1}{2\tau^2}\mu^2\right)$$

เมื่อกำหนดให้ $\tau^2$ มีขนาดใหญ่ เมื่อกำหนดให้ $\bar{x} = \frac{{}_{i=1}^{n}X}{n}$ ความน่าจะเป็นก่อนปรับสำหรับการแจกแจงชายขอบจะอยู่ในรูปของ

$$p(x \mid M_1) = h(x)\exp - \frac{1}{2\frac{\sigma^2}{n}}\bar{X}^2$$

และ

$$p(x \mid M_2) = h(x)\left(\frac{\sigma^2}{n} + \tau^2\right)\exp - \frac{1}{2\left(\frac{\sigma^2}{n} + \tau^2\right)}\bar{X}^2$$

อัตราส่วนแปลกภายหลังปรับระหว่างตัวแบบ $M_1$ และ $M_2$ ได้มาจากอัตราส่วนแปลกก่อนปรับ $\frac{p(M_1)}{p(M_2)}$

$$\frac{p(M_1 \mid x)}{p(M_2 \mid x)} = B_{12}(x) X \frac{p(M_1)}{p(M_2)}$$

ดังนั้น

$$B_{12}(x) = \left(\frac{\sigma^2}{n} + \tau^2\right)^{\frac{n}{2}} \exp - \frac{\tau^2}{2\left(\frac{\sigma^2}{n} + \tau^2\right)\frac{\sigma^2 n}{}}$$

คือ ปัจจัยเบส์สำหรับตัวแบบ $M_1$ และ $M_2$

ค่าของปัจจัยเบส์สำหรับตัวแบบ $M_1$ และ $M_2$ ขึ้นอยู่กับค่า ซึ่งอาจจะทำให้การคำนวณนั้นยุ่งยาก เพราะ $B_{12}(x)$ เมื่อ $\tau^2$ ดังนั้นในการคำนวณนักวิจัยอาจจะกำหนดค่าที่ไม่เกินขอบเขตของเนื้อหาที่ทำการวิเคราะห์ อัตราส่วนแปลกภายหลังปรับสำหรับ $M_1$ สามารถทำให้มีค่ามากได้เมื่อเพิ่มค่า $\tau^2$

**กรณีที่ 3 ตัวแบบความถดถอยเชิงพหุ และตัวแปรสุ่ม X มีการแจกแจงแบบปกติ**

เมื่อเราพิจารณาการประยุกต์ใช้การคัดเลือกตัวแบบด้วยวิธีการของเบส์กับการคัดเลือกตัวแปรในการวิเคราะห์ความถดถอยเชิงพหุ กำหนดให้ y แทนเวกเตอร์ของค่าสังเกตขนาด n X 1 สำหรับการคัดเลือกตัวแปรของตัวแบบความถดถอยเชิงพหุจะทำการคัดเลือกจาก $K = 2^p$ ตัวแบบ ดังรูปสมการ

$$p(y \mid \Theta_k, M_k) = (2\tau\sigma^2)^{-n/2} \exp - \frac{1}{2\sigma^2}(y - z_k\beta_k)(y - z_k\beta_k)$$

เมื่อ $Z_k$ แทน เมตริกส์ขนาด n x $q_k$ และ $\beta_k$ แทนเวกเตอร์ขนาด $q_k$ x 1 ของสัมประสิทธิความถดถอยส่วนความเหมาะสมของการแจกแจงก่อนปรับสำหรับพารามิเตอร์ สำหรับแต่ละตัวแบบคือ มีรูปแบบ standard normal-gamma

$$p(\beta_k, \sigma^2 \mid M_k) = p(\beta_k, \sigma^2 \mid M_k)\, p(\sigma^2 \mid M_k)$$

ดังนั้น

$$p(\beta_k \mid \sigma^2, M_k) = (2\pi\sigma^2)\left| \quad {}_k \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}\beta_k' \,{}_k^{-1}\beta_k' \right\}$$

$$p(\sigma^2 \mid M_k) = \frac{(\nu\lambda)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\sigma^{-(\nu+2)} \exp\left\{ -\frac{1}{2\sigma^2}\nu\lambda \right\}$$

จะได้ว่า

$$p(y \mid M_k) = \frac{\Gamma\frac{\nu+n}{2}(\nu\lambda)^{\nu/2}}{\pi^{\frac{n}{2}}\Gamma\frac{\nu}{2}} \left| I + Z_k' \,{}_k Z_k \right|^{-1/2} \left[\lambda\nu + y'(I + Z_k' \,{}_k Z_k)y\right]^{-(\nu+n)/2}$$

ถ้าทุกตัวแบบถูกพิจารณาด้วยความน่าจะเป็นที่เท่ากัน นั่นคือ $p(M_1) = \ldots\ldots = p(M_2) = 1/2^p$ ความน่าจะเป็น ภายหลังปรับของตัวแบบ $M_k$ สามารถคำนวณได้จากสมการ

$$p(M_k \mid y) = \frac{\left| I + Z_k' \,{}_k Z_k \right|^{-1/2} \left[\lambda\nu + y(I + Z_k' \,{}_k Z_k)^{-1} y\right]^{-(\nu+n)/2}}{\sum_{j=1}^{2^p} \left| I + Z_k' \,{}_k Z_k \right|^{-1/2} \left[\lambda\nu + y(I + Z_k' \,{}_k Z_k)^{-1} y\right]^{-(\nu+n)/2}}$$

จากตัวอย่างก่อนหน้านี้ เราสามารถนำไปประยุกต์ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์

จากสูตรเราสามารถแสดงการนำไปใช้ในรูปของ

$$p(\beta_k \mid \sigma^2, M_k) = (2\pi\sigma^2)^{-q_k/2} \left| \quad _k \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(y - Z\beta)(y - Z\beta) \right\}$$

เมื่อ $Z = [z_1, \ldots, z_p]$ และ $\Theta_k = \beta = (\beta_1, \ldots, \beta_p)$ ดังนั้นการจัดกลุ่มตัวแบบจะใช้ความน่าจะเป็นก่อนปรับ $p(\beta_k, \sigma^2 \mid M_k)$ ร่วมกันในปริภูมิพารามิเตอร์

## การเลือกการแจกแจงก่อนปรับ

การคัดเลือกตัวแบบด้วยวิธีการแบบเบส์จะมีผลกระทบจากการแจกแจงก่อนปรับ ซึ่งการคัดเลือกตัวแบบจะมีการกำหนดการแจกแจงก่อนปรับ $p(M_1) = \ldots = p(M_k)$ เพื่อกำหนดน้ำหนักให้กับแต่ละตัวแบบบนพื้นฐานข้อมูลที่มีอยู่ ส่วนการแจกแจงภายหลังปรับได้มาจากการปรับด้วยการแจกแจงก่อนปรับกับฟังก์ชันภาวะน่าจะเป็น ส่วนค่าของปัจจัยเบส์นั้นขึ้นอยู่กับการเลือกพารามิเตอร์การแจกแจงก่อนปรับ $p(M_1 \mid x), \ldots, p(M_k \mid x)$ หลักการดังกล่าว นำเสนอโดย Lindley การเลือกการแจกแจงก่อนปรับ สำหรับแต่ละตัวแบบมีหลักการง่ายๆ คือ การกำหนดความน่าจะเป็นก่อนปรับสำหรับทุกๆ ตัวแบบให้มีการแจกแจงแบบยูนิฟอร์ม นั่นคือ $p(M_1 \mid x), \ldots, p(M_k \mid x) = 1/k$ อย่างไรก็ตามการเลือกการแจกแจงก่อนปรับจะต้องพยายามเลือกอย่างระมัดระวัง ซึ่งการเลือกการแจกแจงนั้นไม่จำเป็นต้องเป็นการแจกแจงแบบยูนิฟอร์มเสมอไป ขึ้นอยู่กับลักษณะที่เราสนใจ ตัวอย่างเช่น การคัดเลือกตัวแปรเมื่อเรากำหนดการแจกแจงก่อนปรับ ในการกำหนดน้ำหนักในแต่ละตัวแบบ เหตุผลในการเลือกการแจกแจงก่อนปรับจะขึ้นอยู่กับฟังก์ชันของพารามิเตอร์ และจำนวนตัวแปรของแต่ละตัวแบบ แนวคิดในการเลือกการแจกแจงก่อนปรับจะอยู่บนพื้นฐานของการกำหนดข้อสนเทศก่อนปรับในงานวิจัยที่ทำการศึกษากับตัวแปรจำนวนมากและตัวแบบมีความไม่แน่นอน

ดังนั้นบทความนี้จึงได้นำเสนอการแจกแจงก่อนปรับในกรณีที่ไม่ทราบข้อสนเทศ หลักการกำหนดการแจกแจงก่อนปรับในกรณีที่ไม่ทราบข้อสนเทศเป็นหลักการที่นิยมใช้ในปัจจุบัน เนื่องจากการแจกแจงก่อนการปรับแบบอัตวิสัยสามารถกำหนดได้ยาก เนื่องจากข้อจำกัดด้านเวลา และค่าใช้จ่าย และในกรณีที่มีตัวแปรในการวิเคราะห์จำนวนมากทำให้การประมาณค่าพารามิเตอร์มีความซับซ้อนมากยิ่งขึ้น จากเหตุผลข้างต้นทำให้นักวิจัยหลายท่านเลือกใช้การแจกแจงก่อนปรับในกรณีที่ไม่ทราบข้อสนเทศมาใช้ในการวิเคราะห์ บทความนี้ได้นำเสนอแนวคิดในการกำหนดการแจกแจงก่อนปรับในกรณีที่ไม่ทราบข้อสนเทศ 4 แนวคิดด้วยกัน คือ

1. Uniform Prior เป็นการกำหนดการแจกแจงก่อนปรับด้วยการกำหนดให้เป็นค่าคงที่ โดยทั่วไปกำหนดให้มีค่าเท่ากับ 1 แต่ไม่จำเป็นเสมอไป ขึ้นอยู่กับความเหมาะสมของการแจกแจง

2. Jeffreys Prior เป็นการกำหนดการแจกแจงก่อนปรับด้วยการกำหนด เมื่อ $\pi(\theta) = \sqrt{\det(I(\theta))}$ คือ Fisher Information matrix แนวคิดนี้เสนอโดย Jeffreys (1961)

3. Reference Prior เป็นการกำหนดการแจกแจงก่อนปรับที่ถูกพัฒนาโดย Bernado (1979) สำหรับใช้กับปัญหาที่มีจำนวนตัวแปรมากๆ แนวคิดนี้ไม่สามารถอธิบายด้วยวิธีการง่ายๆ แต่วิธีการนี้เป็นวิธีการที่ปรับจาก Jeffreys Prior โดยการลดจำนวนพารามิเตอร์ลง

4. Maximal Data Information Prior เป็นการกำหนดการแจกแจงก่อนปรับที่ถูกพัฒนาโดย Zellner (1971) บนพื้นฐานของข้อสนเทศ โดยการกำหนด เมื่อ $\pi(\theta) = \exp\{p(x \mid \theta) \log p(x \mid \theta) dx\}$ แทน $p(x \mid \theta)$ ฟังก์ชันความหนาแน่นของข้อมูล

การกำหนดการแจกแจงก่อนปรับด้วยวิธี Uniform Prior Jeffreys Prior Reference Prior และ Maximal Data Information Prior(MDIP) ซึ่งบทความนี้ผู้เขียนได้นำเฉพาะการแจกแจงที่ถูกนำไปใช้ในงานวิจัย ซึ่งสามารถดูรายละเอียดดังตาราง 1

## การประมาณค่าวิธีการของเบส์

ถึงแม้ว่าการกำหนดการแจกแจงก่อนปรับกรณีที่ไม่ทราบข้อสนเทศ ถูกนำมาใช้ในการคัดเลือกตัวแบบด้วยหลักการของเบส์ และก็ยังมีหลายหลักการที่ถูกนำมาใช้และเป็นแบบ improper ซึ่งจะมีผลต่อ

**ตาราง 1 การกำหนดการแจกแจงก่อนการปรับ สำหรับการแจกแจงต่างๆ**

| การแจกแจง | การแจกแจงก่อนปรับ | พารามิเตอร์ | การแจกแจงหลังปรับ |
|---|---|---|---|
| 1. Binomial | Uniform Prior | 1 | $Be(p \mid x+1, n-x+1)$ |
| | Jeffreys Prior | $\frac{1}{\pi}p^{-1/2}(1-p)^{-1/2}$ | $Be(p \mid x+\frac{1}{2}, n-x+\frac{1}{2})$ |
| | Reference Prior | | |
| | MDIP | $1.6186p^{p}(1-p)^{(1-p)}$ | proper |
| 2. Bivariate Binomial | Uniform Prior | 1 | proper |
| | Jeffreys Prior | $\frac{1}{2\pi}(1-p)^{-1/2}q^{-1/2}(1-q)^{-1/2}$ | |
| | Reference Prior1 | $\frac{1}{\pi^{2}}p^{-1/2}(1-p)^{-1/2}q^{-1/2}(1-q)^{-1/2}$ | |
| | Reference Prior2 | $\frac{1}{\pi^{2}}(1-p)^{-1/2}p^{-1/2}(1-p)^{-1/2}q^{-1/2}(1-q)^{-1/2}$ | |
| 3. Poisson | Uniform Prior | 1 | $G(\sum_{i=1}^{n}x_i+1, 1/n)$ |
| | Jeffreys Prior | $\lambda^{-1/2}$ | $G(\sum_{i=1}^{n}x_i+1/2, 1/n)$ |
| | Reference Prior | | |
| | MDIP | - | - |
| 4. Gamma | Uniform Prior | 1 | |
| | Jeffreys Prior | $\sqrt{\alpha PG(1,\alpha)-1/\beta}$ | Proper |
| | Reference Prior | $\sqrt{\alpha PG(1,\alpha)-1/\alpha/\beta}$ | |
| | MDIP | $\sqrt{\alpha PG(1,\alpha)/\beta}$ | |
| 5. Multinomial | Uniform Prior | 1 | Proper |
| | Jeffreys Prior | $C_k^{-1}(\sum_{i=1}^{k}p^{-1/2})(1-\delta_k)^{-1/2}$ | |
| | Reference Prior | $(\pi^{-k})\sum_{i=1}^{k}[(p_i^{-1/2})(1-\delta_k)^{-1/2}]$ | |
| | MDIP | $p_1^{p_1}, p_2^{p_2}, \ldots, p_k^{p_k}(1-\sum_{i=1}^{k}p_i)^{1-\sum_{i=1}^{k}p_i}$ | |

**ตาราง 1 การกำหนดการแจกแจงก่อนการปรับ สำหรับการแจกแจงต่างๆ (ต่อ)**

| การแจกแจง | การแจกแจงก่อนปรับ | พารามิเตอร์ | การแจกแจงหลังปรับ |
|---|---|---|---|
| 6. Negative Binomial | Uniform Prior | 1 | $Be(p \mid \alpha, x - \alpha + 1)$ |
| | Jeffreys Prior | $1/[p\sqrt{1-p}]$ | $Be(p \mid \alpha, x - \alpha + 1/2)$ |
| | Reference Prior | $1/[p\sqrt{1-p}]$ | $Be(p \mid \alpha, x - \alpha + 1/2)$ |
| | MDIP | - | - |
| 7.Normal (ทราบ $\sigma^2$) | Uniform Prior | 1 | $\pi(\mu \mid D) = N(\bar{x}, \sigma^2/n)$ |
| | Jeffreys Prior | | |
| | Reference Prior | | |
| | MDIP | | |
| 8. Normal(ทราบ $\mu$) | Uniform Prior | 1 | $\pi(\sigma^2 \mid D) \sim IG(n-2)/2,2/S^2)$ |
| | Jeffreys Prior | $1/\sigma^2$ | $\pi(\sigma^2 \mid D) \sim IG(n/2,2/S^2)$ |
| | Reference Prior | | |
| | MDIP | | |
| 9. Normal (ไม่ทราบ $\sigma^2$และ $\mu$) | Uniform Prior | 1 | $\pi(\mu \mid D) \sim T(n-3, \bar{x}S^2/n(n-3))$<br>$\pi(\sigma^2 \mid D) - IG(n-3)/2,2/S^2)$ |
| | Jeffreys Prior | $1/\sigma^4$ | $\pi(\mu \mid D) \sim T(n+1, \bar{x}, S^2/n(n+1))$<br>$\pi(\sigma^2 \mid D) \sim IG((n+1)/2,2/S^2)$ |
| | Reference Prior | $\pi(\phi,\sigma) \quad (2+\phi)^{-1/2}\sigma^{-1}$ | |
| | MDIP | $1/\sigma^2$ | $\pi(\mu \mid D) - T(n-1, \bar{x}, S^2/n(n+1))$<br>$\pi(\sigma^2 \mid D) \sim IG((n-1)/2,2/S^2)$ |

***หมายเหตุ*** *MDIP แทน Manimal Data Information Prior* *PG แทน การแจกแจงพัวซอง-แกมม่า*
*Be แทน การแจกแจงเบต้า* *IG แทน การแจกแจง INVERSE-GAMMA*
*G แทน การแจกแจงแกมม่า* *PROPER แทน มีการแจกแจงเหมือนกับการแจกแจงก่อนปรับ*

การแปลผลของความน่าจะเป็นภายหลังปรับ เนื่องจากเป็นหลักการหนึ่งของการกำหนดในกรณีที่เราไม่ทราบความไม่แน่นอนเพราะ improper prior ถูกกำหนดโดยการคูณด้วยค่าคงที่และนำไปสู่การหาค่าปัจจัยเบส์ที่ถูกคูณด้วยค่าคงที่ โดยเฉพาะอย่างยิ่งการคำนวณนั้นจะยุ่งยากขึ้นเมื่อทำการเปรียบเทียบตัวแบบที่มีการแจกแจงต่างกัน

การประมาณค่าด้วยวิธีการของเบส์ปัจจุบันได้มีนักวิจัยหาวิธีการประมาณ มาพัฒนาการคำนวณให้มีความง่ายมากขึ้น ตัวอย่างเช่นการคัดเลือกตัวแบบของข้อมูลที่มีการแจกแจงแบบปกติ โดยที่มีการแจกแจงก่อนปรับเป็นแบบ improper Spiegelhalter และ Smith (1982) ได้เสนอหลักการคูณโดยตรงสำหรับหาค่าปัจจัยเบส์บนพื้นฐานของข้อมูลตัวอย่าง Perrichi (1984)

ได้เสนอการคูณโดยตรงบนพื้นฐานของการปรับด้วยข้อสนเทศ แนวทางในการหลีกเหลี่ยงข้อบกพร่องของการกำหนดการแจกแจงก่อนปรับแบบ improper สำหรับใช้ในการคำนวณเพื่อหาความหนาแน่นชายขอบ ถ้ากำหนดให้ $S_1$ และ $S_2$ แทนกลุ่มข้อมูล x 2 กลุ่ม ถ้ากำหนดความหนาแน่นคาดหมายของ $S_1$ และ $S_2$ ภายใต้ตัวแบบ $M_k$ นั่นคือ

$$p(S_1|S_2,M_k) = p(S_1|S_2,\theta_k,M_k)p(\theta_k|S_2,M_k)d\theta_k$$

จะสังเกตได้ว่าจะเปลี่ยน $p(\theta_k|M_k)$ ไปเป็น $p(\theta_k|S_2,M_k)$ เมื่อทำการคำนวณแล้วจะได้การประมาณค่าปัจจัยเบส์จะใช้ $p(S_1|S_2,M_k)$ แทนที่ $p(x|M_k)$ หลักการนี้เสนอโดย Geisser และ Eddy (1979) รวมทั้ง Berger และ Perrichi (1995) ด้วย

## การนำไปใช้ในการตัดสินใจ

ในการคัดเลือกตัวแบบสมมติว่าเรามีตัวแบบทั้งหมด k ตัวแบบ นั่นคือ ตัวแบบ $M_1, M_2, \ldots\ldots, M_k$ โดยแต่ละตัวแบบประกอบด้วยความหนาแน่นน่าจะเป็นสำหรับตัวแปรสุ่ม y

$$M_j = \{P_{\theta_j}(y); \theta_j\ \Omega_j\}$$

เมื่อ $\theta_j$ แทน พารามิเตอร์ที่ไม่ทราบค่าในตัวแบบที่ j ฟังก์ชันภาวะน่าจะเป็นสำหรับตัวแบบ $M_j$ คือ

$$L_j(\theta_j) = {}_i P_{\theta_j}(y_i)$$

เป็นผลคูณของข้อมูล n ชุด $y^n = (y_1,\ldots\ldots,y_n)$ ส่วน log-likelihood แทนด้วย $l_j(\theta_j) = \log(L_j\,\theta_j)$ และกำหนดให้ $\theta_j$ แทนภาวะน่าจะเป็นสูงสุดของตัวประมาณ $\theta_j$ สำหรับการแจกแจงก่อนปรับสำหรับตัวแบบ $M_j$ แทนด้วย $p(M_j)$ ในบทความนี้จะกำหนดให้ $p(M_j) = \frac{1}{k}$ เมื่อ j=1,......,k สำหรับแต่ละตัวแบบเราจะกำหนดการแจกแจงก่อนปรับ $p_j(\theta_j)$ สำหรับพารามิเตอร์ $\theta_j$ ความน่าจะเป็นหลังปรับสำหรับตัวแบบ $M_j$ สามารถคำนวณจากทฤษฎีของเบส์ ซึ่งตัวแบบที่มีความสามารถในการพยากรณ์ได้ดีที่สุดจะเป็นตัวแบบที่ให้ค่าความน่าจะเป็นหลังปรับมากที่สุด

## บทสรุป

การคัดเลือกตัวแบบด้วยแนวคิดของเบส์อาจจะมีความซับซ้อนในเรื่องของการคำนวณและการกำหนดลักษณะการแจกแจงอยู่บ้าง แต่ปัจจุบันเราสามารถใช้เครื่องคอมพิวเตอร์เป็นเครื่องมือที่ช่วยในการคำนวณรวมทั้งยังมีงานวิจัย และรายงานต่างๆ เกิดขึ้นมามากขึ้นทำให้วิธีการของเบส์ได้รับความสนใจมาก และที่น่าสนใจมากที่สุดคือวิธีการคัดเลือกของเบส์เป็นวิธีการที่แปลผลได้ง่ายที่สุด นั่นคือ ตัวแบบใดที่ให้ค่าความน่าจะเป็นมากที่สุด ตัวแบบนั้นจะเป็นตัวแบบที่ให้ค่าพยากรณ์ที่แม่นยำมากที่สุด



1. Joseph B.
De
Av
2. Edward
Sci
3. Hugn Ch
Bay
and
Av
[20
4. Robert E.
254
As
Av
5. Ruoyong
Pu
Av
6. Steiger, J
Seq
7. Scientific
(20
Av
8. Larry Wa
of S
Av

# เอกสารอ้างอิง


1. Joseph B., Kadane & Nicole, Alazar. **Methods and Criteria for Model Selection**. Technical Report759#. Department of Statistics Carnegie Mellon University Pittsburge,2003. [Online]. (2004). Available :http://www.stat.cmu.edu [2004, June 18].
2. Edward I., George. **Bayesian Model Selection**.Preliminary Draft for the Encyclopedia of Statistical Science. University of Texas at Austin. [1995, June].
3. Hugn Chipman, Edward I. George & Robert E., McCulloch. **The Practical Implementation of Bayesian Model Selection** : The University of Waterloo, The University of Texas and Austin, and The University of Chicago, 2001.[Online]. (2005).

   Available : gsbwww:uchicago.edu/fac/hebibert.lopes/teaching/41903/Modelselection.pdf [2005, June 2].
4. Robert E., Kass & Adrian E., Raftery. **Bayes factors and model uncertainty**. Technical Report no. 254, March 1993. A revised version appeared in the Journal of the American Statistical Association, 90(1995):773-795. [Online].

   Available : http://www.stat.washington.edu/raftery/Research/Bayes/bayes_papers.html
5. Ruoyong Yang,James O Berger.**A Catalog of Noninformative Priors**. Department of Statistics, Purdue University.August,1996. [Online]. (2005).

   Available : http://www.isds.duke.edu [2005, January 20].
6. Steiger, J.H., A. Shapiro, & M.W. Browne. (1985). **On the Multivariate Asymptotic Distribution of Sequential Chi-Square Statistics**. Psychometrika 50(3): 253-264.
7. Scientific Software International, Inc., **Model Selection, Estimation and Starting Values**. [Online]. (2004).

   Available :http://www.ssicentral.com/lisrel/models.htm. [2004, June 18].
8. Larry Wasserman. **Bayesian Model Selection and Model Avergeing**. Technical Report 666# Department of Statistics Carnegie Mellon University Pittsburge,2003. [Online]. (2004).

   Available : http://www.stat.cmu.edu [2004, June 18].