บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ผู้วิจัยต้องการเปรียบเทียบผลการใช้สถิติในการทดสอบเกี่ยวกับค่าเฉลี่ยของสองประชากร คือ การทดสอบด้วยสถิติที (t-test) และการทดสอบด้วยสถิติแบบเบย์เซียน (Bayesian Statistics) ขอเสนอเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้

1. ความรู้เกี่ยวกับการทดสอบสมมติฐานด้วยสถิติแบบที
2. ความรู้เกี่ยวกับการทดสอบสมมติฐานด้วยสิถิติแบบเบย์เซียน
3. ความรู้เกี่ยวกับการแจกแจงของประชากรที่ใช้ในการศึกษาครั้งนี้
4. ความรู้เกี่ยวกับระเบียบวิธีมอนติคาร์โล
5. เอกสารและงานวิจัยที่เกี่ยวข้อง

1. ความรู้เกี่ยวกับทดสอบสมมติฐานด้วยสถิติแบบที

สถิติทดสอบแบบที ได้ถูกพัฒนาโดยวิลเลียม กอสเซท (William Gosset) นักสถิติชาวอังกฤษ ซึ่งใช้นามแฝงว่า สติวเดนท์ (Student) และต่อมาได้ถูกพัฒนาโดยฟิชเชอร์ (Fisher) และเรียกการแจกแจงว่า การแจกแจงแบบทีของฟิชเชอร์ (Fisher's t Distribution) และเรียกการทดสอบว่า การทดสอบโดยวิธีฟิชเชอร์ (Fisherion approach) โดยแบบการแจกแจงแบบที จะเป็นรูปแบบการแจกแจงแบบระฆังคว่ำ สมมาตร มีผลรวมความน่าจะเป็นเท่ากับ 1 มีจุดศูนย์กลางที่ t=0 มีค่าเฉลี่ย ค่ามัธยฐาน และฐานนิยมอยู่ที่จุดศูนย์กลาง มีความแปรปรวน เท่ากับ n/n-2 โดยมีระดับของความเป็นอิสระ (Degree of Freedom) เท่ากับ n ซี.เอ. บูโน (C.A. Bouneau) กล่าวถึงสถิติทดสอบแบบทีว่า มีข้อตกลงเบื้องต้น 3 ข้อ

1) กลุ่มตัวอย่างได้มาจากประชากรที่มีการแจกแจงเป็นแบบปกติ (Sampling from Normal Distribution)

2) ค่าความแปรปรวนของประชากรทั้งสองมีค่าเท่ากัน (Homogeneity of Variances)

3) กลุ่มตัวอย่างทั้งสองกลุ่มมีความเป็นอิสระจากกัน (Independent)

บูโน กล่าวว่า ข้อตกลงที่มีผลกระทบต่อการทดสอบแบบทีมากคือ การแจกแจงของประชากรและค่าความแปรปรวน ส่วนความเป็นอิสระของกลุ่มตัวอย่างมีผลกระทบน้อย ซึ่งในทาง

ปฏิบัติจริงของการทดสอบความแตกต่างของค่าเฉลี่ยสองประชากร นักวิจัยส่วนใหญ่จะไม่ทราบการแจกแจงของประชากรและความแปรปรวน ซึ่งส่วนใหญ่จะสมมุติให้มีการแจกแจงของประชากรเป็นแบบปกติ และความแปรปรวนเท่ากัน ซึ่งเป็นการเสี่ยงหรือฝ่าฝืนข้อตกลง ทั้ง ๆ ที่ยังไม่มีการทดสอบหรือการทำวิจัยมาก่อนในเรื่องการแจกแจงของประชากรและทดสอบค่าความแปรปรวนของประชากร สำหรับกรณีของความแปรปรวนไม่เท่ากันนั้น เวลช์ (Welch) ได้ปรับแก้สูตรค่าสถิติทีใหม่ โดยปรับแก้ระดับความเป็นอิสระ ซึ่งจริง ๆ แล้วเป็นเพียงการประมาณการแจกแจงแบบทีเท่านั้น ไม่ใช่การแจกแจงที่แบบทีที่แท้จริง (E.J. Dudewicz and S.N. Mishra, 1987, pp. 495-502, J.A.Steger, 1971 อ้างใน อุษาพร เสวกวิ, 2533, หน้า 2-3)

การทดสอบสมมติฐานเปรียบเทียบค่าเฉลี่ยของสองประชากร

ในการทดสอบสมมติฐานเปรียบเทียบค่าเฉลี่ยของสองประชากร ด้วยสถิติแบบที ข้อมูลต้องเป็นไปตามข้อตกลงเบื้องต้น ดังได้กล่าวแล้วนั้น สามารถแยกได้เป็น 2 กรณีดังนี้

1. กรณีความแปรปรวนของสองประชากรเท่ากัน

ถ้าให้ $x_{1i} = (x_{11}, x_{12}, x_{13}, \ldots, \quad 1_{1})$ เป็นตัวอย่างจากประชากรที่มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย เป็น μ_1

$x_{2i} = (x_{21}, x_{22}, x_{23}, \ldots, \quad 2_{2})$ เป็นตัวอย่างจากประชากรที่มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย เป็น μ_2

ประชากรทั้งสองมีค่าความแปรปรวนเท่ากัน คือ σ^2

นั่นคือ $x_{1i} \sim N(\mu_1,\sigma^2)$ และ $x_{2i} \sim N(\mu_2,\sigma^2)$

ในการทดสอบสมมติฐานเปรียบเทียบค่าเฉลี่ยของสองประชากร $H_0 : \mu_1 - \mu_2 = \delta$

ถ้าให้ n_1 เป็นขนาดตัวอย่างที่ได้จากประชากร $N(\mu_1,\sigma^2)$

n_2 เป็นขนาดตัวอย่างที่ได้จากประชากร $N(\mu_2,\sigma^2)$

$\bar{x}_1 = \sum_{i=1}^{n_1} \frac{x_{1i}}{n_1}$ เป็นค่าเฉลี่ยของกลุ่มตัวอย่างที่ได้จากประชากร $N(\mu_1,\sigma^2)$

$\bar{x}_2 = \sum_{i=1}^{n_2} \frac{x_{2i}}{n_2}$ เป็นค่าเฉลี่ยของกลุ่มตัวอย่างที่ได้จากประชากร $N(\mu_2,\sigma^2)$

$s_1^2 = \sum_{i=1}^{n_1} \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1}$ เป็นค่าความแปรปรวนของกลุ่มตัวอย่างที่ได้จากประชากร $N(\mu_1,\sigma^2)$

$s_2^2 = \sum_{i=1}^{n_2} \frac{(x_{2i} - \bar{x}_2)}{n_2 - 1}$ เป็นค่าความแปรปรวนของกลุ่มตัวอย่างที่ได้จากประชากร $N(\mu_2, \sigma^2)$

จะได้ตัวสถิติทดสอบสมมติฐานคือ

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

ที่ระดับความเป็นอิสระ $n_1 + n_2 - 2$

เมื่อ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

และจะปฏิเสธสมมติฐาน H_0 เมื่อ $|t| > t_{v=n+n-2,\, 1-\alpha/2}$

(E.J. Dudewicz and S.N. Mishra, 1987, pp. 494-496)

2. กรณีความแปรปรวนไม่เท่ากัน

สถิติทดสอบสมมติฐาน $H_0 : \mu_1 - \mu_2 = \delta$ คือ

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

ซึ่งสถิติดังกล่าวนี้ไม่ใช่การแจกแจงทีที่แท้จริง แม้ว่าเวลช์ (Welch) ได้แนะนำให้มีการปรับระดับความเป็นอิสระเป็น

$$\frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left[\frac{s_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{s_2^2}{n_2}\right]^2}{n_2 - 1}}$$

เพื่อให้การแจกแจงของค่าสถิติดังกล่าวเข้าใกล้การแจกแจงแบบที

(E.J. Dudewicz and S.N. Mishra, 1987, pp. 495-502)

2. ความรู้เกี่ยวกับการทดสอบสมมติฐานด้วยสิถิติแบบเบย์เซียน

ทฤษฎีเบย์ (Bayes' Theorem)

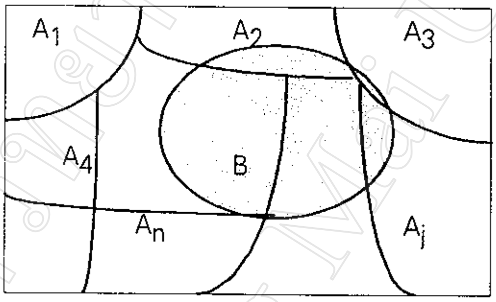
ทฤษฎีเบย์เป็นทฤษฎีสำหรับหาค่าความน่าจะเป็นแบบมีเงื่อนไขของเหตุการณ์ที่สนใจจากเหตุการณ์ที่เป็นเงื่อนไข

เมื่อกำหนดให้ $A_1, A_2, A_3, \ldots, A_n$ เป็นเหตุการณ์ไม่เกิดขึ้นร่วมกันและเป็นอิสระจากกัน (Mutually Exclusive Independent events) จะได้ว่า

$A_i \cap A_j = \phi$ เมื่อ $i \neq j$

$A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = S$ (Sample Space)

และ กำหนดให้ B เป็นเหตุการณ์ใด ๆ ที่ $B \subset A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$



จะได้ความน่าจะเป็นแบบมีเงื่อนไขของ A_j หรือ ทฤษฎีเบย์ ดังนี้

$$P(A_j / B) = \frac{P(B / A_j)P(A_j)}{\sum_{i=1}^{n} P(B / A_i)P(A_i)}$$

หรือเขียนในรูปของตัวแปรสุ่มของเหตุการณ์ของ A_j ดังนี้

$$P(x = A_j / y = B) = \frac{P(y = B / x = A_j)P(x = A_j)}{\sum_{i=1}^{n} P(y = B / x = A_i)P(x = A_i)}$$

(R.L. Winkler, 1972, pp. 40-42)

ซึ่งทฤษฎีความน่าจะเป็นแบบมีเงื่อนไขหรือทฤษฎีเบย์สามารถนำไปประยุกต์ในการหาการแจกแจงของตัวแปรได้

ถ้าให้ $x = (x_1, x_2, x_3, \ldots, x_n)$ เป็นเวคเตอร์ของค่าสังเกตขนาด n ซึ่งมีค่าการแจกแจงความน่าจะเป็นคือ $P(x/\theta)$ ที่ขึ้นอยู่กับค่าพารามิเตอร์ θ จำนวน k ค่า คือ $\theta' = (\theta_1, \theta_2, \theta_3, \ldots, \theta_k)$ และมีค่าการแจกแจงความน่าจะเป็นคือ $P(\theta)$ แล้วจะได้ว่า

$P(x/\theta)\,P(\theta) = P(x,\theta) = P(\theta/x)\,P(x)$

และได้ความน่าจะเป็นแบบมีเงื่อนไขของ θ เป็น

$$P(\theta / x) = \frac{P(x / \theta)P(\theta)}{P(x)}$$

และจาก

$P(x) = E[\,P(x/\theta)] = c^{-1}$

$= \int f(x/\theta)\,f(\theta)\,d\theta$ เมื่อ θ เป็นค่าที่ต่อเนื่อง

$= \sum P(x/\theta)P(\theta)$ เมื่อ θ เป็นค่าที่ไม่ต่อเนื่อง

ดังนั้น

$P(\theta/x) = c\,P(x/\theta)\,P(\theta)$

จากสมการจะได้ P(θ) เรียกว่า การแจกแจงเบื้องต้น **(Prior Distribution)** เป็นการแจกแจงของค่า θ ซึ่งมีรายละเอียดเกี่ยวกับข้อมูลหลายส่วน และ P(θ/x) จะเรียกว่า การแจกแจงโพสทีเรีย **(Posterior Distribution)** เป็นการแจกแจงของ θ ที่ทราบรายละเอียดของข้อมูล x เพิ่มเข้ามาจากการแจกแจงเบื้องต้น (E.P. Box and G.C. Tiao, 1973, p. 10)

จากสมการฟังก์ชัน P(x/θ) ฟิชเชอร์ (1922) เรียกว่า ไลค์ลิฮูดฟังก์ชันของ θ เมื่อกำหนด x ซึ่งเขียนแทนด้วย $\ell(\theta / x)$ และเขียนสมการความน่าจะเป็นแบบมีเงื่อนไขได้ใหม่เป็น

$$P(\theta / x) = c\,\ell(\theta / x)\,P(\theta)$$

หรือ
$$P(\theta / x) = \frac{P(\theta)\ell(\theta / x)}{\int P(\theta)\ell(\theta / x)d\theta}$$

หรือ
$$P(\theta / x) = \frac{P(\theta)\ell(\theta / x)}{\sum P(\theta)\ell(\theta / x)}$$

นั่นคือ
$$\text{Posterior probability} = \frac{(\text{Pr ior probability})(\text{likelihood})}{\sum(\text{Prior probability})(\text{likelihood})}$$

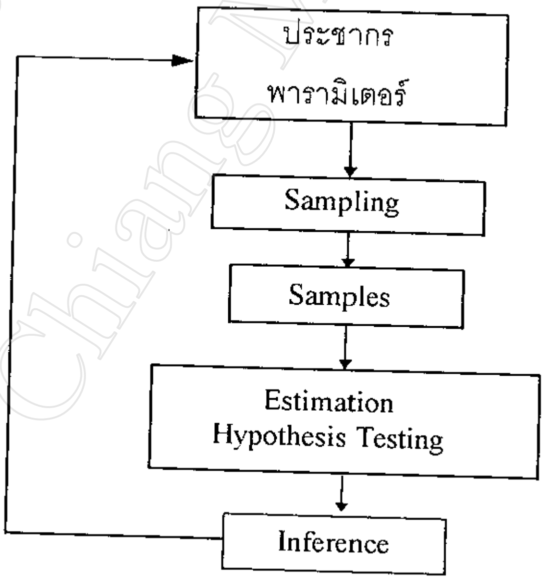
แสดงว่า การแจกแจงโพสทีเรียจะแปรผันตามไลค์ลิฮูดฟังก์ชันและการแจแจงเบื้องต้น หรือเขียนได้เป็น

การแจกแจงโพสทีเรีย $\propto$ ไลค์ลิฮูดฟังก์ชัน x การแจกแจงเบื้องต้น

ซึ่งหมายถึงไลค์ลิฮูดฟังก์ชัน จะมีบทบาทสำคัญในการเปลี่ยนแปลงหรือทำให้ทราบการเปลี่ยนแปลงของ θ มากขึ้น เมื่อเพิ่มเติมรายละเอียดของข้อมูล x (G.E.P. Box and G.C.Tiao, 1973, pp. 10-11; R.L. Winkler, 1972, pp. 143-145)

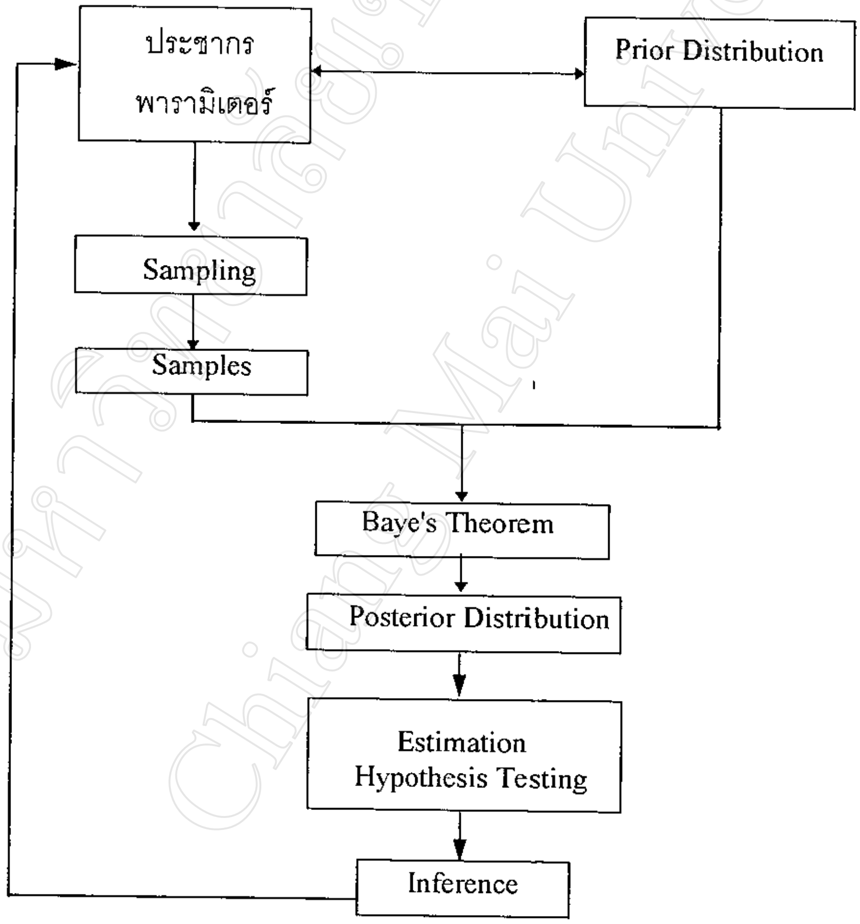
วิธีการและการอ้างอิงทางสถิติของสถิติแบบเบย์เซียน (Bayesian process and inference)

ในกระบวนการของสถิติแบบดั้งเดิมนั้น ในการอ้างอิงเกี่ยวกับพารามิเตอร์ของประชากรนั้นจะอาศัยข้อมูลจากกลุ่มตัวอย่าง ซึ่งจะมีข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลที่จะได้มากำกับอยู่ นำข้อมูลจากการสุ่มตัวอย่าง ซึ่งเรียกว่า ตัวอย่าง (Samples) นั้นมาทำการประมาณค่า (Estimation) ตัวสถิติที่สอดคล้องกับพารามิเตอร์ที่ต้องการอ้างอิงนั้น และทำการทดสอบสมมติฐาน (Hypothesis Testing) ตัวสถิติว่า สามารถนำมาเป็นตัวประมาณ (Estimator) ค่าพารามิเตอร์นั้น ๆ ได้หรือไม่ ถ้าทดสอบสมมติฐานแล้วปรากฏว่า ตัวสถิติที่ประมาณนั้นสามารถนำไปเป็นค่าประมาณพารามิเตอร์ได้ จึงจะสามารถนำไปทำการอ้างอิงประมาณค่าพารามิเตอร์ที่สอดคล้องดังกล่าวได้



ส่วนกระบวนการอ้างอิงของสถิติแบบเบย์เซียนนั้น ต่างไปจากกระบวนการของสถิติแบบดั้งเดิม ในกระบวนการของสถิติแบบเบย์เซียนนั้นจะถือว่าค่าพารามิเตอร์ของประชากรนั้นเป็นตัวแปร (variable) ซึ่งสถิติแบบดั้งเดิมนั้นถือว่าเป็นค่าคงที่ (constant) ดังนั้นในสถิติแบบเบย์เซียนค่าพารามิเตอร์จะมีการแจกแจง (Distribution) และจะอาศัยข้อมูลเบื้องต้นที่มีอยู่ ซึ่งรวบรวมข้อมูลทุก

อย่าง เช่น ความเชื่อ ปรัชญา การวิจัยที่มีอยู่ เป็นต้น เป็นการแจกแจงที่เรียกว่า การแจกแจงเบื้องต้น (Prior Distribution) เมื่อทำการศึกษาจะทำการสุ่มตัวอย่างมาเพื่อศึกษาคุณลักษณะของประชากรที่ต้องการ อาศัยทฤษฎีเบย์รวมข้อมูลเบื้องต้นและข้อมูลที่ได้จากตัวอย่างเข้าด้วยกัน จะได้การแจกแจงที่เรียกว่า การแจกแจงโพสทีเรีย (Posterior Distribution) แล้วทำการประมาณค่า (estimation) ค่าพารามิเตอร์จากการแจกแจงดังกล่าว



ความน่าจะเป็นแบบออฟเจคทีพและซับเจคทีฟ (objective and subjective probability)

ความน่าจะเป็นออฟเจคทีฟ คือ ความน่าจะเป็นที่เกิดจากการทดลองให้เห็นชัดเจนว่าจะเกิดเหตุการณ์ที่น่าสนใจในสัดส่วนหรือความถี่ เช่น การทดลองโยนเหรียญหลาย ๆ ครั้งก็จะเกิดความน่าจะเป็นสำหรับเกิดหัวหรือก้อยเท่าไร ซึ่งทำให้เกิดแนวคิดของสถิติแบบดั้งเดิม ทำให้การอ้างอิงทางสถิติประกอบด้วยการประมาณค่าและการทดสอบสมมติฐาน ซึ่งในการทดสอบสมมติฐานก็จะทำให้การทดสอบว่าเป็นจริงหรือเท็จในระดับนัยสำคัญที่กำหนดขึ้น โดยการนำข้อมูลจากการสุ่มตัวอย่างจากประชากรเดียวกันหลาย ๆ กลุ่มมาสร้างช่วงประมาณแล้วจะมีสัดส่วนให้ความสามารถบรรจุพารามิเตอร์ เช่น ถ้ากำหนดระดับนัยสำคัญ .05 การสร้างช่วงประมาณ ก็จะพบว่า จากการสุ่มตัวอย่าง 100 ครั้ง จะมี 95 ครั้งที่ค่าพารามิเตอร์ตกอยู่ในช่วงประมาณดังกล่าว จะมีเพียง 5 ครั้งที่ค่าพารามิเตอร์ไม่ตกอยู่ในช่วงดังกล่าว (G.R.Inversen, 1984, p.11)

ความน่าจะเป็นซับเจคทีพ คือ ความน่าจะเป็นที่เกิดจากการวัดความไม่แน่นอนของเหตุการณ์หนึ่ง ๆ ซึ่งจะเกิดขึ้นกับข้อมูลหรือหลักฐานที่มีอยู่ขณะนั้น ขึ้นอยู่กับความเชื่อของบุคคลที่มีต่อเหตุการณ์ที่สนใจ เช่น มหาวิทยาลัยจะคาดหวังว่านักศึกษาที่สอบเข้าได้คะแนนสูง จะมีโอกาสสำเร็จการศึกษาได้มากกว่านักศึกษาที่สอบเข้าได้คะแนนต่ำ (R.L.Winkler, 1974, pp.72-74)

การแจกแจงเบื้องต้น (Prior Distribution)

การแจกแจงเบื้องต้นเป็นหน้าที่ของผู้วิจัยว่าจะกำหนดเป็นอย่างไร ซึ่งอาจแตกต่างกันไปสำหรับนักวิจัยแต่ละคน ขึ้นอยู่กับข้อมูลเริ่มแรกที่ผู้วิจัยแต่ละคนมีอยู่ ความเชื่อต่าง ๆ ของผู้วิจัย บางครั้งการกำหนดการแจกแจงเบื้องต้นที่แตกต่างกันนี้ก็อาจส่งผลทำให้ผลการวิเคราะห์เหมือนกันก็ได้ แต่ในทางตรงกันข้ามก็อาจส่งผลต่อการวิเคราะห์และการตัดสินใจที่แตกต่างกันหรือคลาดเคลื่อนกันได้

ข้อมูลเบื้องต้น หรือข้อมูลเริ่มแรก (Prior Information) จะเป็นส่วนสำคัญอย่างยิ่งที่ทำให้ผู้วิจัยได้กำหนดการแจกแจงเบื้องต้นของข้อมูล ส่วนใหญ่แล้วหากนักวิจัยไม่มีข้อมูลเริ่มต้นเลยหรือไม่มีความรู้เกี่ยวกับพารามิเตอร์ที่สนใจเลย ผู้วิจัยมักกำหนดให้การแจกแจงเบื้องต้นเป็นการแจกแจงแบบสม่ำเสมอ (Uniform Distribution) หรือค่าความน่าจะเป็นของตัวอย่างมีโอกาสเกิดขึ้นเท่า ๆ กัน ซึ่งการวิเคราะห์หรือตัดสินใจอ้างอิงก็จะขึ้นกับข้อมูลจากกลุ่มตัวอย่างเพียงอย่างเดียว

ซึ่งในกรณีนี้ถือได้ว่าสถิติแบบเบย์เซียนเหมือนกับสถิติแบบดั้งเดิม เพราะขึ้นอยู่กับข้อมูลจากกลุ่มตัวอย่างเพียงอย่างเดียวเหมือนกัน แต่การตีความหมายจะแตกต่างกันออกไป เพราะการวิเคราะห์แบบเบย์เซียนถือว่าค่าพารามิเตอร์เป็นตัวแปรตัวหนึ่ง ส่วนในสถิติดั้งเดิมค่าพารามิเตอร์เป็นเพียงค่าคงที่ค่าหนึ่งเท่านั้น

ดังนั้นจึงจำเป็นอย่างยิ่งที่ผู้วิจัยจะต้องมีการหาความรู้เกี่ยวกับค่าพารามิเตอร์ที่สนใจศึกษามาก่อน ซึ่งอาจได้มาจากการวิจัยเรื่องก่อน ๆ แล้วผนวกกับความเชื่อของตนเองที่มีต่อค่าพารามิเตอร์นั้น ๆ หรือในกรณีของการทำซ้ำก็อาจนำการแจกแจงโพสทีเรียครั้งแรกมาเป็นการแจกแจงเบื้องต้นของการทำซ้ำได้ แล้วนำมาสร้างเป็นการแจกแจงเบื้องต้นเพื่อให้การตีความของการวิเคราะห์ด้วยสถิติแบบเบย์เซียนได้ผลดียิ่งขึ้น

ข้อมูลเบื้องต้นเกี่ยวกับค่าพารามิเตอร์จะเป็นตัวกำหนดความน่าจะเป็นและการแจกแจงเบื้องต้น ซึ่งขึ้นอยู่กับความเชื่อของบุคคล และข้อมูลที่มีอยู่ในขณะนั้น ซึ่งจะเป็นความน่าจะเป็นแบบซับเจคทีพ ซึ่งจะมีลักษณะที่สำคัญ 2 ลักษณะคือ

1. การกำหนดความน่าจะเป็นทุก ๆ ค่าของพารามิเตอร์ จะมีค่ามากกว่าศูนย์ (Non- negative)

2. ผลรวมทั้งหมดของความน่าจะเป็นทุก ๆ ค่าของพารามิเตอร์จะรวมกันเป็น 1 (W.L.Hays, 1970, pp. 456-457 อ้างในอุษาพร เสวกวิ, 2533, หน้า 28)

การสร้างแบบการแจกแจงเบื้องต้น (Prior distribution) ในกรณีของตัวแปรต่อเนื่องทำได้หลายวิธี คือ

1. การเลือกฟังก์ชันทางคณิตศาสตร์ เป็นฟังก์ชันการแจกแจง โดยความเชื่อของตนเองว่าการแจกแจงเบื้องต้นของค่าพารามิเตอร์น่าจะเป็นตามฟังก์ชันคณิตศาสตร์ที่เลือกมา (อุษาพร เสวกวิ, 2533, หน้า 28)

2. การสร้างโค้งความถี่จากข้อมูลเบื้องต้นที่มีอยู่ โดยการสร้างงฮิสโตรแกรมแล้วปรับโค้งให้เรียบ (Grouping and Smoothing Technique) แล้วพิจารณาถึงค่ากำหนดตำแหน่งต่าง ๆ เช่น Mean, Median, Mode ค่าการกระจาย ช่วงความเชื่อมั่น ค่าเปอร์เซนต์ไทล์ต่าง ๆ แล้วประเมินว่าโค้งดังกล่าวน่าจะมีการแจกแจงเป็นแบบใด (R.L. Winkler, 1972, pp.183-186; อุษาพร เสวกวิ, 2533, หน้า 29-30)

3. การสร้างการแจกแจงโดยกำหนดคอนจุเกต (Conjugate prior distribution) เป็นการการสร้างชุดการกระจาย (Family of distribution) เมื่อนำมาใช้สำหรับสถิติแบบเบย์เซียน จำเป็นต้องพิจารณาถึงความง่ายในการคำนวณ หาไลค์ลิฮูดฟังก์ชันได้ คำนวณการแจกแจงโพสทีเรียง่าย ลักษณะการแจกแจงของประชากร การสุ่มตัวอย่าง หรือข้อตกลงเบื้องต้นของชุดการแจกแจง (เช่น ถ้าเป็นการแจกแจงแบบปกติ แล้วตัวอย่างที่ได้มาจะต้องเป็นกระบวนการอิสระแบบเบอร์นูลี (Bernoulli process)

ข้อมูลของกลุ่มตัวอย่างที่มาจากการแจกแจงแบบเบอร์นูลี (Bernoulli process) จะพบว่า สัดส่วนของผลสำเร็จ (success) เป็นการแจกแจงแบบไบโนเมียล (Binomial distribution) จะได้ว่า ไลค์ลิฮูดฟังก์ชันเป็นไบโนเมียล เมื่อพิจารณาถึงกลุ่มตัวอย่างต่าง ๆ และลักษณะประชากรแล้วจะพบว่า ฟังก์ชันการแจกแจงแบบเบต้า (Beta distribution) เป็นไลค์ลิฮูดฟังก์ชันที่เหมาะสม ดังนั้นคอนจุเกตที่เหมาะสมกับการทดลองแบบเบอร์นูลีก็คือ การแจกแจงแบบเบต้า (Beta distribution)(R.L.Winkler, 1974,pp. 497-498)

ขั้นตอนการสร้างการแจกแจงเบื้องต้นด้วยวิธีนี้ คือ พิจารณาข้อมูลเบื้องต้น (prior information) และข้อมูลจากกลุ่มตัวอย่าง (Sample information) จะสามารถหาฟังก์ชันคอนจุเกตในการสร้างการแจกแจงเบื้องต้นได้

คุณสมบัติของคอนจุเกตฟังก์ชัน ที่สำคัญมีดังนี้ พารามิเตอร์ต่าง ๆ ในฟังก์ชันคอนจุเกตเพื่อให้ฟังก์ชันสมบูรณ์ และสอดคล้องกับเหตุการณ์ที่สนใจอยู่

1. คุณสมบัติทางคณิตศาสตร์ (อุษาพร เสวกวิ, 2533, หน้า 31)

1.1 คอนจุเกตจะทำให้หาการแจกแจงโพสทีเรียง่ายขึ้น

1.2 ถ้าเลือกคอนจุเกตที่มีลักษณะเหมือนกัน แต่ค่าคงที่ต่างกัน ก็ทำให้การหาการแจกแจงโพสทีเรียไม่แตกต่างกันมากนัก

1.3 เมื่อเลือกคอนจุเกตฟังก์ชันสำหรับสร้างการแจกแจงโพสทีเรียได้แล้ว ควรจะสามารถคำนวณหาค่าต่าง ๆ ที่ใช้ในการอ้างอิง และตัดสินใจได้ไม่ยากนัก

2. คุณสมบัติริชเนส (Richness) เป็นคุณสมบัติที่สามารถเลือคอนจุเกตในการสร้างการแจกแจงเบื้องต้นที่เหมือนกันได้แต่แตกต่างในกันบ้างในรูปร่างและตำแหน่ง การกระจาย

3. เมื่อสร้างการแจกแจงเบื้องต้นได้แล้วสามารถตีความโดยไม่ต้องพิจารณาผลจากกลุ่มตัวอย่าง

ตัวอย่างการหาแบบการแจกแจงโดยกำหนดคอนจุเกต

1. ในการทดลองแบบเบอร์นูลี (Bernoulli process) จะได้ฟังก์ชันการแจกแจงแบบเบต้า เป็นชุดคอนจุเกต ดังนี้

$$f(p) = \frac{(n-1)!}{(r-1)!(n-r-1)!}p^{r-1}(1-p)^{n-r-1}$$

เมื่อ p แทนความน่าจะเป็นของการทดลองประสบผลสำเร็จ (success), $0 \le p \le 1$

n เป็นจำนวนครั้งของการทดลอง

r เป็นจำนวนครั้งของการทดลองประสบผลสำเร็จ

จะได้การแจกแจงเบื้องต้นเป็นแบบเบต้าและการแจกแจงโพสทีเรียก็เป็นแบบเบต้า ดังนี้

การแจกแจงเบื้องต้นคือ

$$f'(p) = \frac{(n'-1)!}{(r'-1)!(n'-r'-1)!}p^{r'-1}(1-p)^{n'-r'-1}$$

และการแจกแจงโพสทีเรียคือ

$$f''(p) = \frac{(n''-1)!}{(r''-1)!(n''-r''-1)!}p^{r''-1}(1-p)^{n''-r''-1}$$

$$n'' = n'+1,\ r'' = r'+1$$

(R.L.Winkler, 1974, pp. 498-503)

2. ในกรณีของการทดลองเป็นแบบปกติ (Normal process) จะได้ชุดคอนจุเกตเป็นการแจกแจงเป็นแบบปกติ (Normal distribution)

จะได้การแจกแจงเบื้องต้นเป็นแบบปกติและการแจกแจงโพสทีเรียก็เป็นปกติ

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\mu-m)^2}{2\sigma^2}}$$

การแจกแจงเบื้องต้นคือ

$$f'(\mu) = \frac{1}{\sqrt{2\pi}\sigma'}e^{-\frac{(\mu-m')^2}{2\sigma'^2}}$$

การแจกแจงโพสทีเรียคือ

$$f''(\mu) = \frac{1}{\sqrt{2\pi\sigma''}} e^{-\frac{(\mu - m'')^2}{2\sigma''^2}}$$

เมื่อ $\frac{1}{\sigma''^2} = \frac{1}{\sigma'^2} + \frac{1}{\sigma^2}$

$$m' = \frac{(\frac{1}{\sigma'^2})m' + (\frac{n}{\sigma^2})m}{(\frac{1}{\sigma'^2}) + (\frac{n}{\sigma^2})}$$

(R.L. Winkler, 1974, pp. 507-508)

คุณสมบัติบางประการของการแจกแจงเบื้องต้น

1. ความไวของการทดสอบ

การแจกแจงเบื้องต้นเป็นความน่าจะเป็นแบบซับเจคทีพ ซึ่งขึ้นอยู่กับแต่ละบุคคล ทำให้แตกต่างกันออกไปตามความสามารถของบุคคลและวิธีการกำหนดการแจกแจงเบื้องต้นที่แตกต่างกันอาจทำให้เกิดผลต่อการแจกแจงโพสทีเรียที่แตกต่างกันไปด้วย ซึ่งมีผลต่อการตัดสินใจหรือเกิดความคลาดเคลื่อนในการวิเคราะห์ได้

กรณีการแจกแจงเบื้องต้นไม่มีผลต่อการเปลี่ยนแปลงของการแจกแจงโพสทีเรีย เราจะเรียกว่า การทดสอบไม่มีความไว ซึ่งทำให้การตัดสินใจหรือการวิเคราะห์ไม่มีความแตกต่างกันมากแต่หากการแจกแจงเบื้องต้นมีผลต่อการเปลี่ยนแปลงของการแจกแจงโพสทีเรีย จะเรียกว่า การทดสอบมีความไว ซึ่งทำให้ผลการตัดสินใจหรือผลการวิเคราะห์แตกต่างกันไปด้วย

2. จำนวนข้อมูลเบื้องต้น

ในสถานการณ์การกำหนดการแจกแจงเบื้องต้นหากนักวิจัยมีข้อมูลเบื้องต้นอยู่น้อย การวิเคราะห์หรือการตัดสินใจจะขึ้นอยู่กับข้อมูลจากกลุ่มตัวอย่างเพียงอย่างเดียว ซึ่งมักจะกำหนดให้การแจกแจงเบื้องต้นเป็นแบบแฟลต **(flat)** หรือ การแจกแจงเป็นแบบสม่ำเสมอ **(Uniform)** จะมีผลทำให้การวิเคราะห์ทางสถิติแบบเบย์เซียนจะเหมือนกับการวิเคราะห์แบบสถิติดั้งเดิม แต่แตกต่างกันเพียงค่าพารามิเตอร์เท่านั้น ที่การวิเคราะห์แบบดั้งเดิมถือว่าเป็นค่าคงที่ ส่วนการวิเคราะห์แบบเบย์เซียนถือว่าค่าพารามิเตอร์เป็นตัวแปร ดังนั้นจึงทำให้การตี ความหมายแตกต่างกันออกไป (อุษาพร เสวกวิ, 2533, หน้า 31-35)

การแจกแจงโพสทีเรีย (Posterior distribution)

ในกระบวนการทางสถิติแบบเบย์เซียนค่าพารามิเตอร์เป็นตัวแปรสุ่มที่ทำการศึกษา อาศัยทฤษฎีเบย์ทำการสร้างการแจกแจงโพสทีเรียจากการแจกแจงเบื้องต้นและข้อมูลที่ได้จากกลุ่มตัวอย่างที่ทำการศึกษา เพื่อหาค่าความน่าจะเป็นสำหรับพารามิเตอร์นั้น ๆ ในการอ้างอิงค่าพารามิเตอร์จะอ้างอิงหรือประมาณค่าจะทำการอ้างอิงหรือประมาณค่าจากการแจกแจงโพสทีเรีย

การอ้างอิงค่าความแตกต่างของค่าเฉลี่ยสองประชากร (Inferences concerning the difference between two means) กรณีการแจกแจงเบื้องต้นเป็นการแจกแจงปกติทั้งสองประชากร

ให้ $(x_{11}, x_{12}, x_{13}, \ldots, x_{1i}, \ldots, x_{1n_1})$ เป็นตัวอย่างที่ได้จากประชากรที่มีการแจกแจง $N(\mu_1, \sigma_1^2)$ และ $(x_{21}, x_{22}, x_{23}, \ldots, x_{2j}, \ldots, x_{2n_2})$ เป็นตัวอย่างที่ได้จากประชากรที่มีการแจกแจง $N(\mu_2, \sigma_2^2)$

เมื่อพารามิเตอร์ μ_1, μ_2, σ_1^2, σ_2^2 ไม่ทราบค่า จะพิจารณา การแจกแจงโพสทีเรียของ $\delta = \mu_1 - \mu_2$ ดังนี้ (E.P. Box and G.C. Tiao, pp. 140-141)

1. การแจกแจงของ δ เมื่อ $\sigma_1^2 = \sigma_2^2$ เท่ากัน มีค่าเป็น σ^2

   δ จะมีการแจกแจงโพสทีเรียเป็นการแจกแจงแบบ t ดังนี้

   $$\delta \sim t(\bar{x}_1 - \bar{x}_2, s^2(\frac{1}{n_1} + \frac{1}{n_2}), \nu_1 + \nu_2)$$

   เมื่อ $\nu_1 = n_1 - 1$

   $$\nu_2 = n_2 -$$
   $$s^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2}$$
   $$s_1^2 = \nu_1^{-1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$
   $$s_2^2 = \nu_2^{-1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2$$

2. การแจกแจงของ δ เมื่อ $\sigma_1^2 \neq \sigma_2^2$

   δ จะมีการแจกแจงโพสทีเรียเป็นการแจกแจงแบบ t ดังนี้

   $$\delta \sim t(\bar{x}_1 - \bar{x}_2, a^2(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}), b)$$

   เมื่อ $b = 4 + \frac{f_1^2}{f_2}$, $a^2 = (\frac{b-2}{b}) f_1$

$$f_1 = \left(\frac{\nu_2}{\nu_2 - 2}\right)\cos^2\phi + \left(\frac{\nu_1}{\nu_1 - 2}\right)\sin^2\phi$$

$$f_2 = \frac{\nu_2^2 \cos^4\phi}{(\nu_2 - 2)^2(\nu_2 - 4)} + \frac{\nu_1^2 \sin^4\phi}{(\nu_1 - 2)^2(\nu_1 - 4)}$$

และ $$\cos^2\phi = \frac{\frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad \sin^2\phi = 1 - \cos^2\phi$$

การทดสอบสมมติฐาน (Hypotesis testing)

ในการทดสอบสมมติฐานทางสถิติของสถิติแบบเบย์เชียนจะเป็นการพิจารณาตัดสินใจเลือกทางเลือกที่ดีที่สุด คือ ปฏิเสธหรือยอมรับสมมติฐานที่ตั้งไว้โดยอาศัยข้อเท็จจริงจากตัวอย่างที่ได้ประกอบกับข้อมูลเบื้องต้นและความเชื่อ หรืออื่น ๆ ที่นักวิจัยมีมาก่อนการสุ่มตัวอย่าง โดยจะพิจารณาถึงความเสี่ยงที่จะตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมติฐน นั่นคือ ต้องเลือกทางเลือกที่มีความเสี่ยงน้อยที่สุด หรือถ้าพิจารณาในเรื่องฟังก์ชันความสูญเสียแล้วจะมีความสูญเสียน้อยที่สุด

เมื่อพิจารณาสมมติฐานอย่างง่ายเกี่ยวกับการทดสอบค่าเฉลี่ยของประชากร ดังนี้

$H_0 : \theta = \theta_0$ กับ $H_1 : \theta = \theta_1$

และพิจารณาเหตุการณ์และทางเลือกต่อไปนี้

เหตุการณ์	ทางเลือก	
	a_1 : ยอมรับ H_0	a_2 : ปฏิเสธ H_0
H_0 เป็นจริง ($\theta \in H_0$) (θ_1เป็นพารามิเตอร์ที่แท้จริง)	$L(a_1,\theta_1) = 0$	$L(a_2,\theta_1) > 0$ (Type I Error)
H_1 เป็นจริง ($\theta \in H_1$) (θ_2เป็นพารามิเตอร์ที่แท้จริง)	$L(a_1,\theta_2) > 0$ (Type II Error)	$L(a_2,\theta_2) = 0$

กำหนดให้ $L(a_i,\theta_j)$ คือ ฟังก์ชันความสูญเสียเมื่องเลือกทางเลือกที่ a_i และ θ_j เป็นพารามิเตอร์ที่แท้จริง

และกำหนดให้ $R(\theta)$ คือ ฟังก์ชันความเสี่ยงของค่า θ ซึ่งจะได้ว่า

$R(\theta) = E[L(a_i,\theta j)]$

= ค่าคาดหวังของฟังก์ชันความสูญเสีย

ถ้ากำหนดให้ C เป็นอาณาเขตวิกฤตของแบบทดสอบ

และ $\pi(\theta)$ เป็นความน่าจะเป็นที่ θ จะตกอยู่ในอาณาเขตวิกฤต C จะได้ว่า

ถ้า $\theta \in H_0$ หรือ θ_1 เป็นพารามิเตอร์ที่แท้จริง จะได้

$$R(\theta_1) = L(a_2,\theta_1)\ \pi(\theta_1)$$

ถ้า $\theta \in H_1$ หรือ θ_2 เป็นพารามิเตอร์ที่แท้จริง จะได้

$$R(\theta_2) = L(a_1,\theta_2)\ [1- \pi(\theta_2)]$$

หรือ
$$\begin{aligned} R(\theta) &= R(\theta_1) + R(\theta_2) \\ &= L(a_2,\theta_1)\ \pi(\theta_1) + L(a_1,\theta_2)\ [1 - \pi(\theta_2)] \end{aligned}$$

เมื่อเทียบกับการแจกแจงเบื้องต้น ($P(\theta)$) แล้ว จะได้แบบทดสอบเป็น

$$P(\theta_1)R(\theta_1) + P(\theta_2)R(\theta_2)$$

และจะได้

$$\begin{aligned} P(\theta_1)R(\theta_1) + P(\theta_2)R(\theta_2) &= P(\theta_1)[L(a_2,\theta_1)\ \pi(\theta_1)] + P(\theta_2)[L(a_1,\theta_2)\ (1 - \pi(\theta_2))] \\ &= P(\theta_1)[L(a_2,\theta_1)\iint_c\dots\int\prod_{i=1}^{n} f(x_i,\theta_1)dx_1dx_2\dots dx_n \\ &\quad + P(\theta_2)[L(a_1,\theta_2)\ [1 - \iint_c\dots\int\prod_{i=1}^{n} f(x_i,\theta_2)dx_1dx_2\dots dx_n] \\ &= P(\theta_2)[L(a_1,\theta_2) + P(\theta_1)[L(a_2,\theta_1)\iint_c\dots\int\prod_{i=1}^{n} f(x_i,\theta_1)dx_1dx_2\dots dx_n \\ &\quad - P(\theta_2)[L(a_1,\theta_2)\ \iint_c\dots\int\prod_{i=1}^{n} f(x_i,\theta_2)dx_1dx_2\dots dx_n \\ &= P(\theta_2)[L(a_1,\theta_2) + \iint_c\dots\int[P(\theta_1)L(a_2,\theta_1)\prod_{i=1}^{n} f(x_i,\theta_1) \\ &\quad - P(\theta_2)L(a_1,\theta_2)\prod_{i=1}^{n} f(x_i,\theta_2)]dx_1dx_2\dots dx_n \end{aligned}$$

ถ้า $(x_1, x_2, x_3, \dots, x_n) \in c$ และทำให้ $P(\theta_1)R(\theta_1) + P(\theta_2)R(\theta_2)$ มีค่าน้อยที่สุดจะได้ว่า

$$P(\theta_1)L(a_2,\theta_1)\prod_{i=1}^{n}f(x_i,\theta_1) - P(\theta_2)L(a_1,\theta_2)\prod_{i=1}^{n}f(x_i,\theta_2) < 0$$

$$P(\theta_1)L(a_2,\theta_1)\prod_{i=1}^{n}f(x_i,\theta_1) < P(\theta_2)L(a_1,\theta_2)\prod_{i=1}^{n}f(x_i,\theta_2)$$

$$\frac{\prod_{i=1}^{n}f(x_i,\theta_1)}{\prod_{i=1}^{n}f(x_i,\theta_2)} < \frac{P(\theta_2)L(a_1,\theta_2)}{P(\theta_1)L(a_2,\theta_1)}$$

$$\frac{\prod_{i=1}^{n}f(x_i,\theta_1)}{\prod_{i=1}^{n}f(x_i,\theta_2)} \bullet \frac{P(\theta_1)L(a_2,\theta_1)}{P(\theta_2)L(a_1,\theta_2)} < 1$$

ในทางปฏิบัติสัดส่วน $\frac{L(a_2,\theta_1)}{L(a_1,\theta_2)}$ จะสามารถหาได้จากข้อมูลที่มีอยู่เบื้องต้นแล้ว หรือหากไม่มีข้อมูลอยู่ก็จะกำหนดให้ฟังก์ชันการสูญเสียมีค่าเท่ากันหรือ $\frac{L(a_2,\theta_1)}{L(a_1,\theta_2)} = 1$ ซึ่งจะได้

$$\frac{\prod_{i=1}^{n}f(x_i,\theta_1)}{\prod_{i=1}^{n}f(x_i,\theta_2)} \bullet \frac{P(\theta_1)}{P(\theta_2)} < 1$$

หรือ $$\frac{P(\theta_1)}{P(\theta_2)} \bullet \frac{\prod_{i=1}^{n}f(x_i,\theta_1)}{\prod_{i=1}^{n}f(x_i,\theta_2)} < 1$$

และถ้าให้ Ω' คือ สัดส่วนความน่าจะเป็นเบื้องต้น (Prior odds ratio)

Ω'' คือ สัดส่วนความน่าจะเป็นโพสทีเรีย (Posterior odds ratio)

$L(\theta_1) = \prod_{i=1}^{n}f(x_i,\theta_1)$ คือ ไลค์ลิฮูดฟังก์ชันของ θ_1 (Likelihood function)

$L(\theta_2) = \prod_{i=1}^{n}f(x_i,\theta_2)$ คือ ไลค์ลิฮูดฟังก์ชันของ θ_2 (Likelihood function)

และ L.R. เป็นสัดส่วนความน่าจะเป็นของกลุ่มตัวอย่าง (Likelihood Ratio)

จากอสมการ จะได้

$$\Omega' \bullet \frac{L(\theta_1)}{L(\theta_2)} = \Omega' \bullet \text{L.R.} < 1$$

$$\Omega'' = \Omega' \bullet \text{L.R.}$$

$\therefore$ จะได้แบบทดสอบคือ $\Omega'' < 1$ จึงจะปฏิเสธสมมติฐาน H_0

(R.L. Winkler, 1974, pp. 621-625; ฝันจิต แต้มทอง, 2537, หน้า 36-41)

ในการทดสอบสมมติฐานแบบเบย์เซียนสามารถเขียนได้เป็น

$$\Omega'' = \frac{P''(H_0)}{P''(H_1)}$$

เมื่อ $P''(H_0)$ เป็นการแจกแจงโพสทีเรียของ H_0 และ

$P''(H_1)$ เป็นการแจกแจงโพสทีเรียของ H_1

สำหรับในกรณีของการทดสอบสมมติฐานประกอบหรือสมมติฐานที่มีค่าพารามิเตอร์มากกว่า 1 ค่า เช่น $H_0: \mu \geq \mu_0$ และ $H_1: \mu < \mu_0$ จะหาค่าความน่าจะเป็นของ H_0 และ H_1 ของการแจกแจงโพสทีเรียมากกว่าจะหาค่า Ω'' จาก L.R. x Ω' นั่นคือ

$P''(H_0$ เป็นจริง$) = P''(\mu \geq \mu_0)$

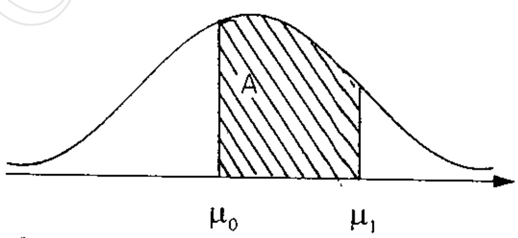
$P''(H_1$ เป็นจริง$) = P''(\mu < \mu_0)$

ถ้า Ω'' น้อยกว่า 1 แล้วจะปฏิเสธสมมติฐาน H_0

สำหรับกรณีของ $H_0: \mu = \mu_0$ และ $H_1: \mu \neq \mu_0$

สำหรับในกรณีของตัวแปรต่อเนื่องแล้ว $P''(H_0)$ นั้นไม่เกิดขึ้นอย่างแน่นอน ดังนั้นจึงไม่ควรตั้งสมมติฐานแบบนี้ ดังนั้นควรจะตั้งสมมติฐานในลักษณะค่าพารามิเตอร์เข้าใกล้ค่าคงที่ค่าหนึ่ง เช่น $H_0: \mu_0 \leq \mu \leq \mu_1$ และ $H_1: \mu < \mu_0$ หรือ $\mu > \mu_1$

นั่นคือ



$P''(H_0) = A$

$P''(H_1) = A' = 1 - P''(H_0)$

การทดสอบสมมติฐานในวิธีเบย์เซียน จะทำการวิเคราะห์หาค่าความน่าจะเป็นของค่าพารามิเตอร์ที่แท้จริง หรือ การหาค่าความน่าจะเป็นของสมมติฐาน H_0 และ H_1

3. ความรู้เกี่ยวกับการแจกแจงของประชากรที่ใช้ในการศึกษาครั้งนี้

การแจกแจงเบื้องต้นที่ใช้ในการศึกษาครั้งนี้มีดังนี้

1. การแจกแจงแบบสมมาตร (Symetry)

ในการศึกษาครั้งนี้ในลักษณะการแจกแจงแบบสมมาตรจะอาศัยชุดการแจกแจงของเอกซ์โปเนนเชียลพาวเวอร์ (Exponential power function) ซึ่งมีฟังก์ชันดังนี้

$$P(x/\theta,\phi,\beta) = k\phi^{-1}\exp[\frac{1}{2}\left|\frac{x-\theta}{\phi}\right|^{2/(1+\beta)}]$$

เมื่อ $k^{-1} = \Gamma(1+\frac{1+\beta}{2})\cdot 2^{1+(1+\beta)/2}$

β เป็นพารามิเตอร์ที่สัมพันธ์กับความโด่ง (Kurtosis)

θ เป็นพารามิเตอร์บอกตำแหน่ง (Location Parameter)

ϕ เป็นสเกลพารามิเตอร์ (Scale Parameter)

$\phi > 0$, $-\infty < \theta < \infty$, $-1 < \beta \leq 1$

โดยฟังก์ชันการแจกแจงนี้มีค่าเฉลี่ยและความแปรปรวนดังนี้

a) ค่าเฉลี่ย (mean)

$E(x) = \theta$

b) ค่าความแปรปรวน (Variance)

$$V(x) = 2^{(1+\beta)}\cdot\left[\frac{\Gamma(\frac{3}{2}(1+\beta))}{\Gamma(\frac{1}{2}(1+\beta))}\right]\cdot\phi^2$$

และฟังก์ชันการแจกแจงดังกล่าวสามารถเขียนได้อีกรูปแบบหนึ่งดังนี้

$$P(x/\theta,\sigma,\beta) = w(\beta)\sigma^{-1}\exp[-c(\beta)\left|\frac{x-q}{s}\right|^{2/(1+\beta)}, \quad -\infty < x < \infty$$

เมื่อ $$c(\beta) = \left[\frac{\Gamma[\frac{3}{2}(1+\beta)]}{\Gamma[\frac{1}{2}(1+\beta)]}\right]^{1/(1+\beta)}$$

$$w(\beta) = \frac{[\Gamma(\frac{3}{2}(1+\beta))]^{1/2}}{(1+\beta)[\Gamma(\frac{1}{2}(1+\beta))]^{3/2}}, \quad \sigma > 0,\ -\infty < x < \infty,\ -1 < \beta \leq 1$$

θ เป็นค่าเฉลี่ย

σ^2 เป็นค่าความแปรปรวน

การแจกแจงตามฟังก์ชันดังกล่าวสามารถกำหนดได้ดังนี้

1.1 เมื่อ $\beta = 0$ จะได้ฟังก์ชันเป็นการแจกแจงแบบปกติ (Normal Distribution)

$$c(\beta) = \left[\frac{\Gamma[\frac{3}{2}(1+0)]}{\Gamma[\frac{1}{2}(1+0)]}\right]^{1/(1+0)} = \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{1}{2})} = \frac{1}{2}$$

$$w(\beta) = \frac{[\Gamma(\frac{3}{2}(1+0))]^{1/2}}{(1+0)[\Gamma(\frac{1}{2}(1+0))]^{3/2}} = \frac{[\Gamma(\frac{3}{2})]^{1/2}}{[\Gamma(\frac{1}{2})]^{3/2}} = \frac{[\frac{1}{2}\Gamma(-\frac{1}{2})]^{1/2}}{[\Gamma(\frac{1}{2})]^{3/2}}$$

$$w(\beta) = \frac{[\frac{1}{2}]^{1/2}[\pi^{1/2}]^{1/2}}{[\pi^{1/2}][\pi^{1/2}]^{1/2}} = \frac{1}{\sqrt{2\pi}}$$

$$P(x/\theta,\sigma,\beta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{1}{2}\left|\frac{x-\theta}{\sigma}\right|^2] = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2]$$

1.2 เมื่อ $\beta = 1$ จะได้ฟังก์ชันเป็นการแจกแจงแบบดับเบิ้ลเอกซ์โปเนนเชียล (Double Exponential Distribution)

$$c(\beta) = \left[\frac{\Gamma[\frac{3}{2}(1+1)]}{\Gamma[\frac{1}{2}(1+1)]}\right]^{1/(1+1)} = \left[\frac{\Gamma(3)}{\Gamma(1)}\right] = [2]^{1/2} = \sqrt{2}$$

$$w(\beta) = \frac{[\Gamma(\frac{3}{2}(1+1))]^{1/2}}{(1+1)[\Gamma(\frac{1}{2}(1+1))]^{3/2}} = \frac{[\Gamma(3)]^{1/2}}{2\cdot[\Gamma(1)]^{3/2}} = \frac{[2!)]^{1/2}}{2\cdot[\Gamma(1)]^{3/2}}$$

$$= \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$$

$$P(x/\theta,\sigma,\beta) = \frac{1}{\sqrt{2}}\sigma^{-1}\exp[-\sqrt{2}\left|\frac{x-\theta}{\sigma}\right|^{2/2}] = \frac{1}{\sqrt{2\sigma^2}}\exp[-\sqrt{2}\left|\frac{x-\theta}{\sigma}\right|]$$

1.3 เมื่อ $\beta \to -1$ จะได้ฟังก์ชันเป็นการแจกแจงแบบสม่ำเสมอ (Uniform Distribution)

$$\lim_{\beta \to -1} P(x/\theta,\sigma,\beta) = \frac{1}{2\sqrt{3}\sigma} \quad , \ \theta - \sqrt{3}\sigma < x < \theta + \sqrt{3}\sigma$$

2. การแจกแจงแบบไม่สมมาตร (Asymtotic)

ในการศึกษาครั้งนี้จะใช้การแจกแจงที่มีความเบ้บวกเป็นตัวศึกษา โดยทำการแปลงข้อมูลที่ได้จากการแจกแจงปกติ ให้มีความเบ้เป็นบวก (Positive Skewness) โดยใช้สมการที่ใช้แปลงคือ

$$y = -0.3268 + 1.16050961x + 0.2909708x^2 - 0.0886191x^3$$

เมื่อ x เป็นข้อมูลที่มีการแจกแจงเป็นปกติ

y เป็นข้อมูลที่มีการแจกแจงเบ้บวกจากการแจกแจงปกติ

(Fleishman, 1978, pp. 521-532)

โดยเมื่อใช้สมการดังกล่าวแปลงข้อมูลที่มีการแจกแจงปกติ x ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 จะได้ค่า y ที่มีการแจกแจงเบ้บวกที่มีค่าเฉลี่ยเป็น 0 ความแปรปรวนเท่ากับ 1 สัมประสิทธิ์ความเบ้เท่ากับ 1 และสัมประสิทธิ์ความโด่ง เท่ากับ 3

4. ความรู้เกี่ยวกับระเบียบวิธีมอนติคาร์โล (Monte Carlo Method)

มอนติคาร์โล เป็นเทคนิคการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่ม (Random Number) มาสร้างสถานการณ์ให้เหมือนหรือคล้ายคลึงกับสถานการณ์จริง และมีการทดลองซ้ำได้หลาย ๆ ครั้ง เพื่ออธิบายปรากฏการณ์หรือหาคำตอบของปัญหาที่สงสัยในสถานการณ์ที่สร้างขึ้น วิธีการของมอนติคาร์โล ถูกเริ่มนำมาใช้ครั้งแรกเมื่อศตวรรษที่ 17 ในการพัฒนาทฤษฎีความน่าจะเป็น (Probability Theory) และการหาค่าพาย (π) และในปี 1908 กอสเซท ก็ได้นำมาศึกษาการแจกแจงแบบที (t-distribution)

ในช่วงสงครามโลกครั้งที่ 2 อูลามและนิวแมน (Ulam and Van Newman) ได้ใช้คำว่า "มอนติคาร์โล" มาเป็นชื่อวิธีการเป็นครั้งแรก ซึ่งชื่อมอนติคาร์โลนี้เป็นรหัสลับของการทำงานที่ลอสอาลามอส (Los Alamos) โดยนำวิธีการนี้มาหาคำตอบของการสร้างระเบิดปรมณูเพื่อหาผลลัพธ์ที่แท้จริง และต่อมาก็ได้ถูกนำมาใช้กันอย่างแพร่หลายและกว้างขวางทั้งในด้านฟิสิกส์ คณิตศาสตร์ สถิติและการวิจัย

ขั้นตอนของมอนติคาร์โล

1. สร้างตัวเลขสุ่ม (Generate random number)

เป็นการสร้างตัวเลขสุ่มโดยอาศัยเครื่องมือต่าง ๆ ในการหาตัวเลขสุ่ม แต่ในปัจจุบันนิยมใช้คอมพิวเตอร์ในการช่วยคำนวณ โดยสูตรทางคณิตศาสตร์ วิธีที่นิยมในปัจจุบันคือ

1.1 วิธีส่วนกลางกำลังสอง (Mid Square Method)

1.2 วิธีเศษเหลือ (Congrunent Method) นิยมใช้เศษเหลือของผลคูณ

ในปัจจุบันการใช้คอมพิวเตอร์ช่วยในการสร้างก็มีโดยมีฟังก์ชันในการสร้างตัวเลขสุ่มแบบสำเร็จรูปเช่น ในภาษาเบสิค จะมีฟังก์ชัน RND และ RANDOMIZE ซึ่งมีความยาวของการซ้ำของเลขสุ่มยาวเหมาะสำหรับการใช้งาน

คุณสมบัติของตัวเลขสุ่ม ต้องมีคุณสมบัติดังนี้

ก) ตัวเลขสุ่มมีการแจกแจงเป็นแบบสม่ำเสมอ

ข) ตัวเลขสุ่มมีความเป็นอิสระต่อกัน

ค) อนุกรมตัวเลขที่จะกลับมาซ้ำเดิมนั้นห่างพอสมควร มีช่วงยาวพอที่จะใช้

ง) สามารถสร้างซ้ำได้

2. นำตัวเลขสุ่มที่ได้มาประยุกต์ใช้กับปัญหาต่าง ๆ ตามลักษณะตัวแปรนั้น ๆ ตามข้อกำหนดที่ต้องการจะให้เป็น

3. ทำการทดลองซ้ำได้หลาย ๆ ครั้ง และสามารถสรุปปัญหานั้นได้

(ศิริจันทร์ ทองประเสริฐ, 2529, หน้า 87-90; บัญชา พนเจริญสวัสดิ์, 2527, หน้า 101 อ้างใน อุษาพร เสวกวิ, 2533, หน้า 62-63 ; ต่าย เซี่ยงฉี, วารสารวัดผลและวิจัยทางการศึกษา, 2537, หน้า 1-5)

5. เอกสารและงานวิจัยที่เกี่ยวข้อง

ในปี 1960 บูโนได้ศึกษาผลของการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากร 2 กลุ่ม ใช้วิธีมอนติคาร์โล โดยมีวัตถุประสงค์เพื่อศึกษาถึงความคลาดเคลื่อนชนิดที่ 1 (Type I error) ของการแจกแจงแบบทีที่เกิดขึ้นกับค่าทางทฤษฎีใน 3 กรณี คือ

1) กลุ่มตัวอย่างทั้งสองมาจากประชากรที่มีการแจกแจงปกติแต่ความแปรปรวนและขนาดของตัวอย่างแตกต่างกัน

2) กลุ่มตัวอย่างทั้งสองมาจากประชากรที่มีการแจกแจงที่ไม่เป็นแบบปกติ แต่เป็นการแจกแจงแบบเดียวกัน

3) กลุ่มตัวอย่างทั้งสองมาจากประชากรที่มีการแจกแจงต่างกัน

พบว่า เกิดความคลาดเคลื่อนชนิดที่ 1 ที่แตกต่างไปจากค่าทางฤษฏีของการแจกแจงแบบที ดังนี้

1) กรณีประชากรทั้งสองกลุ่มมีความแปรปรวนไม่เท่ากันและกลุ่มตัวอย่างมีขนาดแตกต่างกัน และกลุ่มตัวอย่างขนาดเล็กมีความแปรปรวนมาก และกลุ่มตัวอย่างขนาดใหญ่มีความแปรปรวนต่ำ ทำให้เกิดความคลาดเคลื่อนชนิดที่ 1 สูงกว่าค่าทางทฤษฎีมาก

2) กรณีประชากรทั้งสองกลุ่มมีการแจกแจงเหมือนกัน ขนาดตัวอย่างเล็ก (n=5) และความแปรปรวนต่างกัน จะทำให้เกิดความคลาดเคลื่อนชนิดที่ 1 สูงกว่าค่าทางทฤษฎี

3) กรณีประชากรทั้งสองกลุ่มมีการแจกแจงต่างกัน ขนาดตัวอย่างเท่ากัน แต่มีขนาดเล็ก (n=5) และความแปรปรวนเท่ากัน ทำให้เกิดความคลาดเคลื่อนชนิดที่ 1 สูงกว่าค่าทางทฤษฎี (J.A.Steger, 1971 อ้างใน อุษาพร เสวกวิ, 2533 : หน้า 5-6)

สำหรับกรณีการใช้สถิติเบย์เซียนในการทดสอบสมมติฐานในการเปรียบเทียบค่าเฉลี่ยของประชากรสองกลุ่มนั้น ได้มีผู้ศึกษาไว้ดังนี้

มอสเทลเลอร์ และวาเลซ (Mosteller and Wallace)(J.M.Tanur, 1972, p. 114-115) ได้ศึกษาถึงงานเขียนของผู้เขียนต่าง ๆ ว่างานเขียนนั้น ๆ ควรเป็นผลงานของผู้เขียนคนใด โดยศึกษาอัตราการใช้คำโดยใช้สถิติหลาย ๆ แบบศึกษา พบว่า การใช้สถิติแบบเบย์เซียนให้ผลในการศึกษาคมชัดกว่าการใช้สถิติแบบอื่น ๆ

โนวิคและแจกสัน (M.R.Novick and P.H. Jackson) ได้เปรียบเทียบการทดสอบเปรียบเทียบค่าเฉลี่ยของสองประชากร ของการทดสอบแบบดั้งเดิมกับการทดสอบแบบเบย์เซียนในกรณีของการฝ่าฝืนข้อตกลงของการทดสอบแบบเดิม พบว่า การทดสอบแบบเบย์เซียนให้ผลการทดสอบที่คมชัดกว่าการทดสอบแบบเดิม (M.R. Novick and P.H. Jackson, 1974, p. 243-255 : อ้างในอุษาพร เสวกวิ, 2533, หน้า 7)

บอกซ์ ได้นำข้อมูลจากการทดลองของดาร์วินในเรื่องการเจริญเติบโตด้านความสูงของพืชในการทดลองเกี่ยวกับการผสมพันธุ์พืชสายพันธุ์เดียวกับและข้ามสายพันธุ์จำนวน 15 คู่

เมื่อนำข้อมูลดังกล่าวมาเขียนกราฟพบว่า เป็นรูปสี่เหลี่ยมผืนผ้า บอกซ์ได้แนะนำว่าข้อมูลดังกล่าวไม่สมควรใช้สถิติที (t-test) มาทำการทดสอบ เพราะเป็นการฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบแบบที ในเรื่องการแจกแจงของประชากร ควรนำแนวคิดการทดสอบแบบเบย์เซียนมาทำการทดสอบแทน (E.P. Box and G.C. Tiao, 1973, pp. 152-156)

อุษาพร ได้ศึกษาเปรียบเทียบการใช้สถิติแบบเบย์เซียนกับสถิติแบบดั้งเดิม ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่ม ที่มีค่าการกระจายต่างกัน โดยศึกษาเปรียบเทียบการใช้สถิติแบบเบย์เซียนและการใช้สถิติทีในการทดสอบ ในกรณีของการฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงประชากรและความแปรปรวน ซึ่งศึกษาเฉพาะกรณีการแจกแจงแบบสมมาตร โดยทำการจำลองสถานการณ์โดยใช้ระเบียบวิธีมอนติคาร์โล ผลการศึกษาพบว่า

1) ประชากรมีการแจกแจงเป็นแบบปกติ (Normal Distribution) ทั้งคู่ ควรใช้สถิติทดสอบแบบที ถึงแม้ว่าจะมีการฝ่าฝืนข้อตกลงเบื้องต้นในเรื่องความแปรปรวนของประชากร

2) ประชากรมีการแจกแจงเป็นแบบดับเบิ้ลเอกซ์โปเนนเชียล (Double Exponential) ทั้งคู่ และความแปรปรวนไม่เท่ากัน ควรใช้สถิติแบบเบย์เซียนทดสอบ ยกเว้น กรณีตัวอย่างขนาดเล็กและความแปรปรวนสูง

3) การใช้สถิติแบบเบย์เซียนและสถิติที ไม่เหมาะสมในการทดสอบกับกรณีต่อไปนี้

3.1) ประชากรมีการแจกแจงเป็นแบบสม่ำเสมอ (Uniform distribution) ทั้งคู่ แต่ความแปรปรวนไม่เท่ากัน

3.2) ประชากรมีการแจกแจงเป็นปกติและดับเบิ้ลเอกซ์โปเนนเชียล และความแปรปรวนไม่เท่ากัน

3.3) ประชากรมีการแจกแบบปกติและแบบสม่ำเสมอ และความแปรปรวนไม่เท่ากัน

3.4) ประชากรมีการแจกแจงแบบสม่ำเสมอและแบบดับเบิ้ลเอกซ์โปเนนเชียล และความแปรปรวนไม่เท่ากัน

สำหรับกรณี 3.1 ถึง 3.4 อุษาพรแนะนำว่า ทั้ง 4 กรณีควรใช้สถิติแบบเบย์เซียนทดสอบมากกว่าสถิติที เพราะให้อำนาจการทดสอบสูงกว่า ถ้าหากใช้การคำนวณโพสทีเรีย (Posterior) ที่แท้จริง ซึ่งทำให้เกิดความคลาดเคลื่อนชนิดที่ 1 น้อยลง (อุษาพร เสวกวิ, 2533, หน้า 120-122)

จากผลการศึกษาข้างต้นจะเห็นว่าการใช้สสถิติแบบเบย์เซียนช่วยแก้ปัญหาในเรื่องการฝ่าฝืนข้อตกลงของการใช้สถิติแบบดั้งเดิมได้ และช่วยให้นักวิจัยสามารถแปลผลการวิจัยได้ดีขึ้นด้วย จากผลการการศึกษาที่ผ่านมายังไม่มีผู้ศึกษาในกรณีของการแจกแจงประชากรเป็นแบบเบ้ (Asymtotics) เลย ดังนั้นผู้วิจัยจึงต้องการศึกษาในกระบวนการเดิมในส่วนของการแจกแจงแบบสมมาตร และศึกษาเพิ่มเติมในส่วนของการแจกแจงที่ไม่สมมาตรด้วย โดยอาศัยระเบียบวิธีการจำลองสถานการณ์ของมอนติคาร์โล โดยกำหนดอัตราความแปรปรวนของประชากรสองกลุ่มเป็น 1:3, 1:5 และ 1:10 และขนาดกลุ่มตัวอย่าง 5, 15 และ 25