บทความวิจัย

ค่าพารามิเตอร์ที่เหมาะสมของการแจกแจงก่อน ในตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซง

วราฤทธิ์ พานิชกิจโกศลกุล*

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาค่าพารามิเตอร์ที่เหมาะสมของการแจกแจงก่อน (prior distribution)ในตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซง ซึ่งคือค่าพารามิเตอร์ λ กำหนดค่า λ ให้เป็นจำนวนเต็มโดยมีค่าตั้งแต่ 1 ถึง 10 การแจกแจงก่อนของค่า λ มีการแจกแจงแกมมา (gamma distribution) ด้วยพารามิเตอร์ α และ β โดยช่วงของพารามิเตอร์ α และ β ที่เป็นไปได้อยู่ระหว่าง 0.1 ถึง 5.0 เพิ่มขึ้นครั้งละ 0.1 ในการวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล โดยกำหนดขนาดตัวอย่างเท่ากับ 10, 20, 30, 40 และ 50 และทำการทดลองซ้ำๆ กัน 10,000 ครั้งในแต่ละสถานการณ์ ในแต่ละสถานการณ์จะประมาณค่าพารามิเตอร์ เปรียบเทียบค่าประมาณพารามิเตอร์กับค่าพารามิเตอร์จริง และคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (mean squared error) ค่า α และ β ใดให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด ค่า α และ β นั้นจะเป็นค่าที่เหมาะสม ผลการวิจัยสรุปได้ว่า ค่า α และ β ที่เหมาะสม คือ ค่า α และ β ที่ให้ค่า $\alpha\beta$ มีค่าใกล้เคียงกับค่า λ และค่า α มีค่ามากกว่าค่า β เสมอ โดยทั่วไปแล้ว ค่า α จะมีค่าอยู่ระหว่าง [4.0, 5.0] และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแปรผกผันกับขนาดตัวอย่าง

คำสำคัญ: พารามิเตอร์ การแจกแจงก่อน ตัวประมาณเบส์ การแจกแจงปัวซง

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

*ผู้นิพนธ์ประสานงาน, email: wararit@mathstat.sci.tu.ac.th

The Optimal Parameters of Prior Distribution for Bayes' Estimator of Poisson Mean

Wararit Panichkitkosolkul*

ABSTRACT

The objective of this research is to find the optimal parameters of prior distribution for Bayes' estimator of poisson mean. The poisson mean is denoted by parameter λ and its value is given to be an integer from 1 to 10. The prior distribution of λ is assumed to be a gamma distribution with α and β parameters. The possibility of α and β values vary from 0.1 to 5.0 which increase by 0.1. This research used the Monte Carlo simulation method by varying the sample sizes from 10, 20, 30, 40, and 50. In each condition, the simulation was repeated 10,000 times and each time the estimate of parameter was compared to its true value. Then the mean squared error (MSE) of the Bayes estimator were calculated and used to indicate the suitable parameters. The lowest of MSE represents the optimal value. The finding of this study can be concluded as that the optimal values of α and β are their values that provide $\alpha\beta$ closed to the λ parameter. The α value is always higher than the β value. Generally, the α lies between 4.0 and 5.0. Moreover, MSE is inversely proportional to sample size.

Keywords: parameter, prior distribution, Bayes' estimator, Poisson distribution

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University
*Corresponding author, email: wararit@mathstat.sci.tu.ac.th

บทนำ

โดยทั่วไปแล้วการศึกษาข้อมูลเกี่ยวกับจำนวนครั้งของเหตุการณ์ที่สนใจต่อคาบเวลาใดคาบเวลาหนึ่ง หรือต่อพื้นที่ใดพื้นที่หนึ่ง เช่น จำนวนเครื่องบินที่มาลงที่ท่าอากาศยานสุวรรณภูมิภายใน 1 ชั่วโมง จำนวนลูกค้าที่เข้ามาซื้อสินค้าในห้างสรรพสินค้าต่อชั่วโมง หรือจำนวนรอยตำหนิบนถ้วยเซรามิกต่อใบ เป็นต้น โดยข้อมูลดังกล่าวมักสมมุติให้มีการแจกแจงปัวซง (Poisson distribution) ซึ่งมีฟังก์ชันความน่าจะเป็น ดังนี้ [1]

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & , \quad x = 0,1,2,\ldots \\ 0 & , x \text{ มีค่าอื่นๆ} \end{cases}$$

ในการประมาณค่าพารามิเตอร์ λ ซึ่งเป็นค่าเฉลี่ยของการแจกแจงปัวซง สามารถประมาณค่าได้ทั้งการประมาณค่าแบบจุด (point estimation) และการประมาณค่าแบบช่วง (interval estimation) วิธีการประมาณค่าแบบจุดมีหลายวิธี เช่น วิธีกำลังสองน้อยที่สุด (least squares method) วิธีความควรจะเป็นสูงสุด (maximum likelihood method) และวิธีของเบส์ (Bayesian method) ซึ่งวิธีของเบส์จะแตกต่างจากวิธีอื่น โดยทั่วไปจะถือว่าพารามิเตอร์ λ เป็นค่าคงที่ที่ไม่ทราบค่า แต่ตามวิธีของเบส์จะถือว่า λ เป็นค่าของตัวแปรสุ่ม Λ ที่มีการแจกแจงแสดงได้ในรูปการแจกแจงความน่าจะเป็น ซึ่งเรียกว่า การแจกแจงก่อน (prior distribution) เพราะเป็นการแจกแจงที่กำหนดขึ้นก่อนที่จะมีการรวบรวมข้อมูล เมื่อมีการรวบรวมข้อมูลจะใช้ความรู้ที่ได้จากข้อมูลมาปรับปรุงการแจกแจงก่อน การแจกแจงที่ได้จากการปรับปรุงนี้ เรียกว่า การแจกแจงภายหลัง (posterior distribution) [2] ในที่นี้ศึกษาการแจกแจงก่อนของพารามิเตอร์ λ มีการแจกแจงแกมมา (gamma distribution) ด้วยพารามิเตอร์ α และ β ดังนั้นงานวิจัยครั้งนี้จึงสนใจหาค่าพารามิเตอร์ α และ β ที่เหมาะสมของการแจกแจงแกมมาว่าควรมีค่าเป็นเท่าใด จึงทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด

ขอบเขตการวิจัย

ขอบเขตการวิจัย มีดังนี้

1. กำหนดขนาดตัวอย่าง (n) ที่ศึกษา เท่ากับ 10, 20, 30, 40 และ 50
2. กำหนดค่าพารามิเตอร์ λ เท่ากับ 1, 2, 3, ..., 10
3. การแจกแจงก่อนเป็นการแจกแจงแกมมา (α, β) โดยกำหนดช่วงของพารามิเตอร์ α และ β ที่เป็นไปได้อยู่ระหว่าง 0.1 ถึง 5.0 เพิ่มขึ้นครั้งละ 0.1
4. โปรแกรมที่ใช้ในการวิจัยครั้งนี้ทั้งหมดเขียนด้วยโปรแกรม R เวอร์ชัน 2.4.0 ซึ่งทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีวิธีดำเนินการวิจัยดังนี้

1. การจำลองข้อมูลที่ใช้ในการวิจัย

จำลองข้อมูลที่มีการแจกแจงปัวซง ข้อมูลที่ได้กำหนดให้เป็นตัวแปรสุ่ม X ซึ่งการกำหนดค่าพารามิเตอร์ λ ขนาดตัวอย่าง เป็นไปตามขอบเขตของการวิจัย

2. การคำนวณตัวประมาณค่าเฉลี่ยแบบเบส์ โดยกำหนดการแจกแจงก่อนเป็นการแจกแจงแกมมา (α, β) ดังนั้น ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซง คำนวณได้ดังนี้ [3]

$$\hat{\lambda} = \frac{\left(\alpha + \sum_{i=1}^{n} X_i\right)\beta}{n\beta + 1}$$

3. คำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) คำนวณได้ดังนี้

$$\text{MSE} = \frac{\sum_{i=1}^{10,000}(\lambda - \hat{\lambda}_i)^2}{10,000}$$

4. สรุปผล เมื่อจำลองครบทุกกรณีจะทำการสรุปผลการทดลองว่าค่าพารามิเตอร์ของการแจกแจงก่อนมีค่าเท่ากับค่าใด จึงทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด

ผลการวิจัย

จากการหาค่าพารามิเตอร์ที่เหมาะสมของการแจกแจงก่อน ซึ่งมีการแจกแจงแกมมา (α, β) ผลการวิจัยจะแสดงค่าพารามิเตอร์ α และ β ที่ทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีค่า MSE ต่ำที่สุด 2 อันดับ โดยจำแนกตามค่าพารามิเตอร์ λ และขนาดตัวอย่าง แสดงดังตารางที่ 1

ตารางที่ 1 ค่าพารามิเตอร์ α และ β ที่ให้ค่า MSE ต่ำที่สุด 2 อันดับ จำแนกตามค่าพารามิเตอร์ λ และขนาดตัวอย่าง

ค่า λ	ขนาดตัวอย่าง	อันดับที่ 1			อันดับที่ 2		
		α	β	MSE	α	β	MSE
1	10	4.9	0.2	0.04298	4.7	0.2	0.04462
	20	4.9	0.2	0.03149	4.8	0.2	0.03169
	30	5.0	0.2	0.02387	4.6	0.2	0.02426
	40	4.7	0.2	0.01916	5.0	0.2	0.01935
	50	4.8	0.2	0.01614	4.2	0.2	0.01631
2	10	4.6	0.4	0.12617	4.4	0.4	0.12798
	20	4.9	0.4	0.07797	4.7	0.4	0.07837
	30	4.9	0.4	0.05605	5.0	0.3	0.05634
	40	4.8	0.4	0.04306	4.4	0.4	0.04329
	50	4.1	0.4	0.03545	4.6	0.5	0.03576
3	10	5.0	0.5	0.21237	4.6	0.5	0.21839
	20	5.0	0.5	0.12614	4.2	0.6	0.12687
	30	4.8	0.5	0.08840	4.2	0.5	0.08846
	40	4.8	0.6	0.06662	4.4	0.6	0.06754
	50	4.7	0.6	0.05475	4.2	0.5	0.05486
4	10	5.0	0.7	0.30824	4.6	0.8	0.30849
	20	5.0	0.8	0.17252	4.8	0.8	0.17373
	30	4.2	0.8	0.12070	5.0	0.8	0.12077
	40	4.5	0.8	0.09116	4.9	0.7	0.09138
	50	4.2	0.9	0.07398	4.8	0.7	0.07438

ตารางที่ 1 (ต่อ)

ค่า λ	ขนาดตัวอย่าง	อันดับที่ 1			อันดับที่ 2		
		α	β	MSE	α	β	MSE
5	10	4.8	0.8	0.40097	4.7	0.8	0.40311
	20	5.0	0.8	0.21998	4.0	1.0	0.22180
	30	4.6	0.9	0.15034	4.2	1.0	0.15091
	40	5.0	0.8	0.11452	5.0	0.9	0.11518
	50	3.8	1.0	0.09276	4.9	0.8	0.09304
6	10	5.0	0.9	0.49804	5.0	1.0	0.50192
	20	5.0	1.2	0.26672	4.7	1.2	0.26824
	30	5.0	1.1	0.18204	4.0	1.7	0.18281
	40	3.9	1.0	0.13942	4.8	0.9	0.14025
	50	4.5	1.0	0.11269	3.9	1.9	0.11308
7	10	4.3	1.3	0.58850	5.0	1.2	0.58970
	20	4.8	1.1	0.30950	4.4	1.3	0.31843
	30	4.1	1.4	0.21500	4.1	1.7	0.21596
	40	5.0	1.2	0.16163	4.0	1.5	0.16378
	50	4.8	1.1	0.13078	4.5	1.1	0.13175
8	10	5.0	1.6	0.68960	4.7	1.5	0.69380
	20	4.5	1.4	0.36164	3.8	1.7	0.36551
	30	4.1	1.5	0.24798	4.8	1.6	0.24799
	40	4.9	1.3	0.18785	3.9	2.2	0.18818
	50	4.6	1.5	0.15084	3.5	2.5	0.15094

ตารางที่ 1 (ต่อ)

ค่า λ	ขนาดตัวอย่าง	อันดับที่ 1			อันดับที่ 2		
		α	β	MSE	α	β	MSE
9	10	4.9	1.6	0.77340	4.3	1.8	0.78720
	20	4.0	1.7	0.41225	4.8	1.4	0.41241
	30	3.8	1.8	0.28003	4.2	1.9	0.28079
	40	3.9	2.2	0.21214	4.2	2.0	0.21287
	50	4.1	1.9	0.16897	3.7	1.5	0.17045
10	10	5.0	1.5	0.87990	4.9	2.0	0.88010
	20	4.7	1.7	0.46140	5.0	1.8	0.46200
	30	5.0	1.9	0.31055	4.9	1.7	0.31150
	40	4.9	2.0	0.23311	2.5	3.6	0.23491
	50	3.3	2.2	0.18836	4.6	1.4	0.18909

จากตารางที่ 1 ซึ่งแสดงค่าพารามิเตอร์ α และ β ที่ทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีค่า MSE ต่ำที่สุด พบว่า ค่า α และ β ที่เหมาะสม คือ ค่า α และ β ที่ให้ค่า $\alpha\beta$ มีค่าใกล้เคียงกับค่า λ และค่า α มีค่ามากกว่าค่า β เสมอ ซึ่งสามารถสรุปช่วงของค่าพารามิเตอร์ที่ให้ค่า MSE ต่ำที่สุด จำแนกตามค่าพารามิเตอร์ λ ดังตารางที่ 2

ตารางที่ 2 ช่วงของค่าพารามิเตอร์ α และ β ที่ทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีค่า MSE ต่ำที่สุด จำแนกตามค่าพารามิเตอร์ λ

ค่า λ	ช่วงของค่าพารามิเตอร์	
	α	β
1	[4.7, 5.0]	0.2
2	[4.1, 4.9]	0.4
3	[4.7, 5.0]	[0.5, 0.6]
4	[4.2, 5.0]	[0.7, 0.9]

ตารางที่ 2 (ต่อ)

ค่า λ	ช่วงของค่าพารามิเตอร์	
	α	β
5	[3.8, 5.0]	[0.8, 1.0]
6	[3.9, 5.0]	[0.9, 1.2]
7	[4.1, 5.0]	[1.1, 1.4]
8	[3.9, 5.0]	[1.3, 1.6]
9	[3.9, 4.9]	[1.6, 2.2]
10	[3.3, 5.0]	[1.5, 2.2]

สรุปผลการวิจัย

จากการหาค่าพารามิเตอร์ α และ β ที่ทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีค่า MSE ต่ำที่สุด สรุปได้ว่า ค่า α และ β ที่เหมาะสม คือ ค่า α และ β ที่ให้ค่า $\alpha\beta$ มีค่าใกล้เคียงกับค่า λ และค่า α มีค่ามากกว่าค่า β เสมอ โดยทั่วไปแล้ว ค่า α จะมีค่าอยู่ระหว่าง [4.0, 5.0] และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) แปรผกผันกับขนาดตัวอย่าง แสดงดังรูปที่ 1

ข้อเสนอแนะ

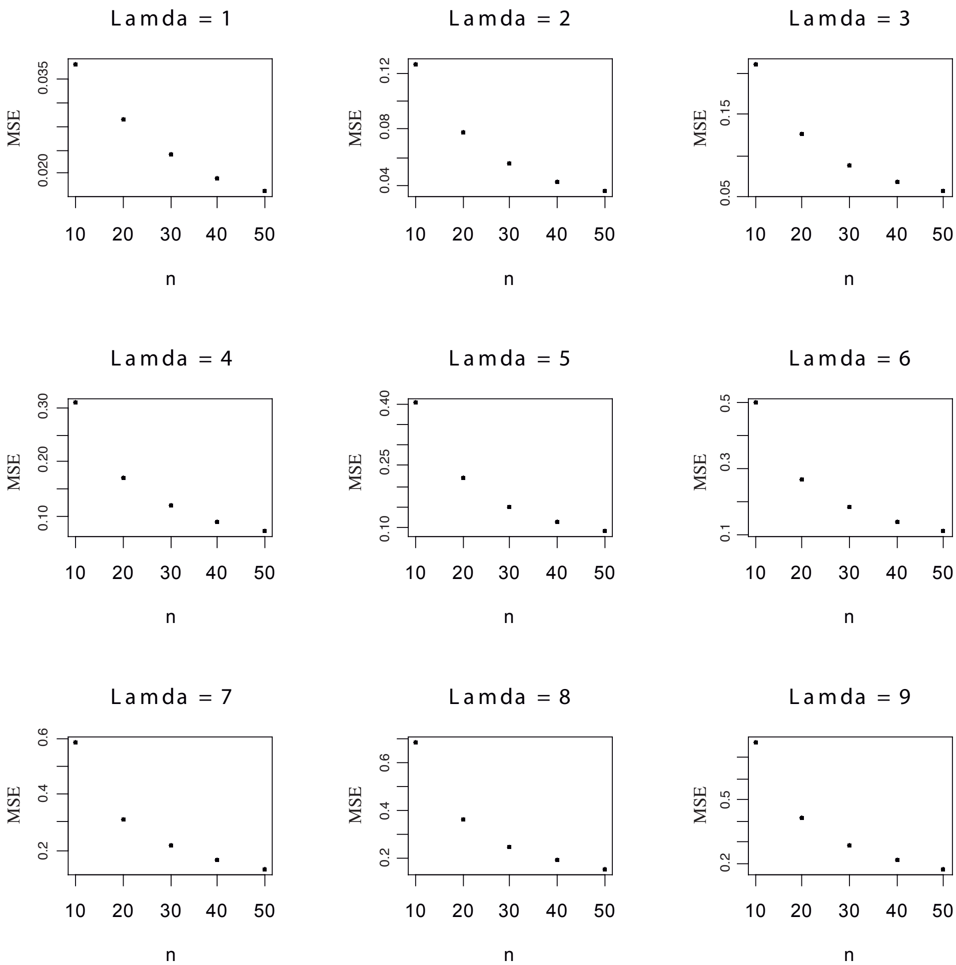
จากผลการวิจัย พบว่า ค่า α และ β ที่ทำให้ตัวประมาณค่าเฉลี่ยแบบเบส์ของปัวซงมีค่า MSE ต่ำที่สุดจะขึ้นอยู่กับค่า λ ซึ่งเป็นค่าเฉลี่ยของการแจกแจงปัวซง แต่ในทางปฏิบัติเราไม่สามารถทราบค่าพารามิเตอร์ก่อนทำการประมาณค่าพารามิเตอร์ ดังนั้นเพื่อให้สามารถเลือกค่า α และ β ที่เหมาะสม ควรทำการประมาณค่า λ เบื้องต้น (preliminary estimation) ซึ่งตัวประมาณค่าพารามิเตอร์ λ เบื้องต้นโดยใช้วิธีกำลังสองน้อยที่สุด (least square method) คือ

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n}$$

เอกสารอ้างอิง

1. Bowerman, B. L. and O'Connell, R. T. Business Statistics in Practice. New York. Mc-Graw-Hill Companies, Inc. p. 182.
2. Bolstad, W. M. 2004. Introduction to Bayesian Statistics. New York. John Wiley and Sons, Inc. p. 6.

3. ชินนะพงษ์ บำรุงทรัพย์. 2548. ทฤษฎีการอนุมานเชิงสถิติ. กรุงเทพมหานคร. หจก.ฟันนี่พลับบลิชชิ่ง. หน้า 121-123.



รูปที่ 1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด จำแนกตามค่าพารามิเตอร์ λ

ได้รับบทความวันที่ 30 พฤศจิกายน 2549
ยอมรับตีพิมพ์วันที่ 8 มกราคม 2550