



การประมาณขนาดประชากรภายใต้การแจกแจง

คอนเวย์แมกซ์เวลล์ปัวซอง

โดย

นางสาวจันทิรา แยมสรवल

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ภาควิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2559
ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

การประมาณขนาดประชากรภายใต้การแจกแจง
คอนเวย์แมกซ์เวลล์ปัวซอง

โดย

นางสาวจันทิรา แยมสรवल



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ภาควิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2559
ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

POPULATION SIZE ESTIMATION BASED UPON
CONWAY-MAXWELL-POISSON DISTRIBUTION

BY

MISS JUNTIRA YAMSUAN



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (APPLIED STATISTICS)

DEPARTMENT OF MATHEMATICS AND STATISTICS

FACULTY OF SCIENCE AND TECHNOLOGY

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2016

COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นางสาวจันทิรา แยมสรวล

เรื่อง

การประมาณขนาดประชากรภายใต้การแจกแจงคอนเวย์แมกซ์เวลล์ปัวซอง
ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

เมื่อ วันที่ 31 กรกฎาคม พ.ศ. 2560

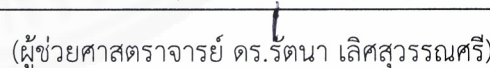
ประธานกรรมการสอบวิทยานิพนธ์



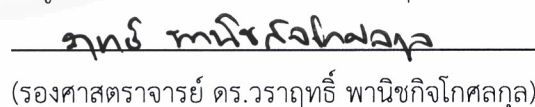
กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์


(ผู้ช่วยศาสตราจารย์ ดร.รมิดา ศรีเหรา)

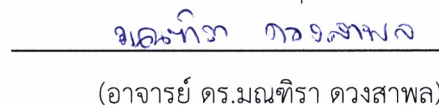
กรรมการสอบวิทยานิพนธ์


(ผู้ช่วยศาสตราจารย์ ดร.รัตนา เลิศสุวรรณศรี)

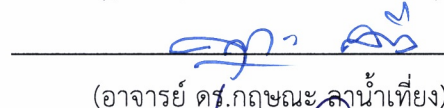
กรรมการสอบวิทยานิพนธ์


(รองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล)

กรรมการสอบวิทยานิพนธ์


(อาจารย์ ดร.มณฑิรา ดวงสภาพ)

คณบดี


(อาจารย์ ดร.กฤษณะ สาน้ำเที่ยง)


(รองศาสตราจารย์ปกรณ์ เสริมสุข)

| | |
|-----------------------------|--|
| หัวข้อวิทยานิพนธ์ | การประมาณขนาดประชากรภายใต้การแจกแจง คอนเวย์แมกซ์เวลล์ปัวซอง |
| ชื่อผู้เขียน | นางสาวจันทิรา แย้มสรवल |
| ชื่อปริญญา | วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์) |
| สาขาวิชา/คณะ/มหาวิทยาลัย | ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ | ผู้ช่วยศาสตราจารย์ ดร.รัตนา เลิศสุวรรณศรี |
| ปีการศึกษา | 2559 |

บทคัดย่อ

Capture-Recapture เป็นวิธีที่ใช้ประมาณขนาดประชากรที่พบได้ยากต่อการสำรวจ พบ ข้อมูลได้จากการนับซ้ำๆกันหลายครั้ง การแจกแจงสำหรับข้อมูลแจกนับโดยทั่วไปมักใช้การแจกแจงปัวซองที่มีพารามิเตอร์ λ เพื่อที่จะขยายจากการแจกแจงปัวซอง เราพิจารณาการแจกแจง Conway-Maxwell-Poisson (CMP) ที่มีพารามิเตอร์ λ และ ν โดยประโยชน์ของการแจกแจง CMP คือสามารถใช้ได้กับข้อมูลที่มีการกระจายมากเกินไป (Over-dispersion) หรือมีการกระจายน้อยเกินไป (Under-dispersion) ได้ นอกจากนี้ CMP ยังเป็นการแจกแจงนับทั่วไปของการแจกแจงอื่นๆ เช่น การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงแบร์นูลลี เป็นต้น งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวประมาณขนาดประชากร 2 ตัวภายใต้การแจกแจง Conway-Maxwell-Poisson นั่นคือ ตัวประมาณขนาดประชากรที่พัฒนาภายใต้ Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ และภายใต้ Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3 โดยใช้กระบวนการหาค่าคาดหวังและการหาค่าสูงสุดในการประมาณค่าพารามิเตอร์ในตัวแบบ

เมื่อเปรียบเทียบตัวประมาณขนาดประชากรที่พัฒนาทั้ง 2 ตัวกับตัวประมาณขนาดประชากรอื่นๆได้แก่ ตัวประมาณขนาดประชากรของ Chao, ตัวประมาณภาวะน่าจะเป็นสูงสุด (MLE), ตัวประมาณของ Lanumteang-Böhning และตัวประมาณที่พัฒนาโดยสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) โดยการจำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson $CMP(\lambda, \nu)$ เมื่อ $\lambda = 0.5, 1.0, 2.0$ และ $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$ จำลองสถานการณ์ภายใต้การแจกแจงปัวซอง $Poisson(\lambda)$ เมื่อ $\lambda = 0.5, 1.0, 2.0$ จำลองสถานการณ์ภายใต้การแจกแจงเรขาคณิต $Geometric(p)$ เมื่อ $p = 0.3, 0.5, 0.7$ และ

จำลองภายใต้การแจกแจงทวินามเชิงลบ *Negative Binomial*(k, θ) เมื่อ $(k, \theta) = (1, 0.3)$, $(3, 0.5)$, $(5, 0.7)$ เพื่อประมาณประชากรขนาด $N = 500, 1,000$ และ $10,000$ โดยแต่ละสถานการณ์ทำซ้ำๆ 2,000 รอบ

ส่วนผลลัพธ์จากการจำลองภายใต้การแจกแจง Conway-Maxwell-Poisson เห็นได้ว่าเมื่อประชากรขนาด N เท่ากับ 500 และ 1,000 ที่พารามิเตอร์ λ และ ν มีค่าสูงจะทำให้ตัวประมาณขนาดประชากรที่พัฒนาขึ้นมาทั้ง 2 ตัวให้ค่าความเอนเอียงสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ต่ำที่สุด ส่วนประชากรขนาด $N = 10,000$ ตัวประมาณขนาดประชากรของ Chao และตัวประมาณขนาดประชากรของ Lanumteang-Böhning มีประสิทธิภาพสูงสุด เมื่อพิจารณาความแปรปรวนสัมพัทธ์ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson ให้ค่าต่ำที่สุด ในการจำลองภายใต้การแจกแจงปัวซอง ตัวประมาณภาวะน่าจะเป็นสูงสุด ให้ค่าความเอนเอียงสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจะให้ค่าที่ต่ำสุด ในขณะที่การจำลองภายใต้การแจกแจงเรขาคณิตและภายใต้การแจกแจงทวินามเชิงลบเห็นได้ว่าตัวประมาณขนาดประชากรของ Lanumteang-Böhning ให้ค่าความเอนเอียงสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ต่ำที่สุด นอกจากนี้ตัวประมาณขนาดประชากรที่พัฒนาขึ้นมาให้ค่าความเอนเอียงสัมพัทธ์ที่ต่ำใกล้เคียงกับตัวประมาณขนาดประชากรของ Chao ทั้งนี้ตัวประมาณขนาดประชากรที่พัฒนาทั้ง 2 ตัวยังสามารถนำไปประยุกต์ใช้กับข้อมูลเสถียรในเมืองแวนคูเวอร์ ประเทศแคนาดาอีกด้วย

คำสำคัญ: Capture-Recapture, Conway-Maxwell-Poisson, กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด, ตัดปลายที่ศูนย์, ตัดปลายที่มากกว่า 3

| | |
|--------------------------------|---|
| Thesis Title | Population Size Estimation Based Upon Conway-Maxwell-Poisson Distribution |
| Author | Miss Juntira Yamsuan |
| Degree | Master of Science (Applied Statistics) |
| Major Field/Faculty/University | Mathematics and Statistics Faculty of Science and Technology Thammasat University |
| Thesis Advisor | Associate Professor Dr. Rattana Lerdsuwansri |
| Academic Year | 2016 |

ABSTRACT

Capture-Recapture techniques have been used to estimate the unknown population size. The data are collected by the repeated identifications. The widely used distribution for count data is the Poisson distribution with parameter λ . To extend Poisson distribution, we consider the Conway-Maxwell-Poisson (CMP) distribution having mean parameter λ and dispersion parameter ν . The utility of CMP distribution allows for over-dispersion or under-dispersion. Moreover, it can be some well-known distribution for count data. The objective of this study is to develop the two new estimators based upon the Conway-Maxwell-Poisson distribution. The EM algorithm is discussed to estimate of parameters of the Conway-Maxwell-Poisson distribution. The two new estimators of the population size are introduced. One is developed based on the zero-truncated CMP, \hat{N}_{ZCMP} and the other is developed based on truncation larger than three, \hat{N}_{YCMP} .

The performance of the two new estimators is compared to the conventional estimators including the Chao's estimator, the maximum likelihood estimator under the zero-truncated Poisson distribution, the Lanumteang-Böhning's estimator, and the linear regression estimator based on the Conway-Maxwell-Poisson distribution. Data are generated from distributions. Firstly, populations are simulated based upon Conway-Maxwell-Poisson distribution; $CMP(\lambda, \nu)$, where $\lambda = 0.5, 1.0$,

2.0 and $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$. Secondly, populations are simulated based upon Poisson distribution; $Poisson(\lambda)$, where $\lambda = 0.5, 1.0, 2.0$. Next, populations are simulated based upon Geometric distribution; $Geometric(p)$ where $p = 0.3, 0.5, 0.7$. Finally, populations are simulated based upon Negative Binomial distribution; $Negative\ Binomial(k, \theta)$, where $(k, \theta) = (1, 0.3), (3, 0.5), (5, 0.7)$. The population to be estimated is $N = 500, 1,000$ and $10,000$. Each situation is repeated 2,000 times.

Results from the simulation study based upon Conway-Maxwell-Poisson distribution show that two new estimators yield the best with smallest relative bias and relative mean square error when λ and ν are biggest for population size $N = 500$ and $1,000$. Chao's estimator and Lanumteang-Böhning's estimator are the best performance for $N = 10,000$. Considering the variance, the population size estimators are developed based upon the Conway-Maxwell-Poisson distribution perform the smallest relative variance. The results, based upon Poisson distribution, show that the maximum likelihood estimator under the zero-truncated Poisson distribution provides the smallest relative bias and relative mean square error. The Lanumteang-Böhning's estimator performs the smallest relative mean square error based upon geometric distribution and Negative Binomial distribution. Moreover, the two new estimators give small relative bias similar to Chao's estimator. The two proposed estimators are illustrated to the real data on street prostitutes in Vancouver.

Keywords: Capture-Recapture, Conway-Maxwell-Poisson, EM algorithm, zero-truncated, truncated larger than three

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้จะไม่สำเร็จลงได้หากปราศจากความอนุเคราะห์จากอาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.รัตนา เลิศสุวรรณศรี ที่ให้คำปรึกษาและให้คำแนะนำที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์ รวมถึงการตรวจทานแก้ไขข้อผิดพลาดต่างๆ เพื่อให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ จึงขอใช้โอกาสนี้ในการกราบขอบพระคุณอย่างสูงไว้ ณ ที่นี้ด้วย

กราบขอบพระคุณประธานกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.รมิตา ศรีเหรา รวมทั้งกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล และอาจารย์ ดร.มณฑิรา ดวงสาพล รวมทั้ง อาจารย์ ดร.กฤษณะ ลาน้ำเที่ยง ผู้ทรงคุณวุฒิภายนอก ที่กรุณาเสียสละเวลาและให้คำแนะนำที่เป็นประโยชน์ต่อการพัฒนาเนื้อหาวิทยานิพนธ์ฉบับนี้ให้เสร็จสมบูรณ์ยิ่งขึ้น นอกจากนี้ขอขอบพระคุณคณาจารย์ภาควิชาคณิตศาสตร์ และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี ที่ได้มอบวิชาความรู้ให้ และขอขอบพระคุณเจ้าหน้าที่ประจำโครงการปริญญาโท ที่คอยช่วยประกาศข่าวและประสานงานให้ อีกทั้งขอขอบคุณคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ที่ได้มอบทุนบัณฑิตเรียนดีเพื่อศึกษาต่อระดับบัณฑิตศึกษา คณะวิทยาศาสตร์และเทคโนโลยี ตลอดระยะเวลา 2 ปีที่ผ่านมา

สุดท้ายนี้ขอขอบคุณครอบครัวที่เป็นแรงผลักดันสู่ความสำเร็จ คอยให้กำลังใจ ให้คำปรึกษาและให้คำแนะนำเพื่ก้าวข้ามอุปสรรคต่างๆ รวมทั้งขอบคุณเพื่อนทุกคนที่คอยให้กำลังใจมาโดยตลอด

นางสาวจันทิรา แยมสรवल

สารบัญ

| | หน้า |
|--|------|
| บทคัดย่อภาษาไทย | (1) |
| บทคัดย่อภาษาอังกฤษ | (3) |
| กิตติกรรมประกาศ | (5) |
| สารบัญตาราง | (9) |
| สารบัญภาพ | (13) |
| รายการสัญลักษณ์และคำย่อ | (14) |
| บทที่ 1 บทนำ | 1 |
| 1.1 ความเป็นมาและความสำคัญ | 1 |
| 1.2 วัตถุประสงค์ของการวิจัย | 4 |
| 1.3 ขอบเขตการศึกษา | 4 |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ | 5 |
| 1.5 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพตัวประมาณขนาดประชากร | 5 |
| บทที่ 2 ทฤษฎี และงานวิจัยที่เกี่ยวข้อง | 7 |
| 2.1 การแจกแจง Conway-Maxwell-Poisson | 7 |
| 2.2 การแจกแจงเรขาคณิต | 8 |
| 2.3 การแจกแจงทวินามเชิงลบ | 9 |

สารบัญ(ต่อ)

| | | |
|----------------------|---|----|
| 2.4 | กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) | 9 |
| 2.5 | ตัวประมาณขนาดประชากร | 11 |
| 2.5.1 | ตัวประมาณขนาดประชากรภาวะน่าจะเป็นสูงสุด | 11 |
| 2.5.2 | ตัวประมาณขนาดประชากรของ Chao | 14 |
| 2.5.3 | ตัวประมาณขนาดประชากรของ Lanumteang-Böhning | 16 |
| 2.5.4 | ตัวประมาณขนาดประชากร LCMP | 18 |
| 2.6 | งานวิจัยที่เกี่ยวข้อง | 22 |
| บทที่ 3 วิธีการวิจัย | | 24 |
| 3.1 | การนำเสนอตัวประมาณขนาดประชากร | 24 |
| 3.1.1 | ตัวประมาณขนาดประชากรตัดปลายที่ค่าศูนย์ | 25 |
| 3.1.2 | ตัวประมาณขนาดประชากรตัดปลายที่มากกว่า 3 | 28 |
| 3.2 | การจำลองสถานการณ์ | 33 |
| 3.2.1 | กำหนดสถานการณ์จำลอง | 33 |
| 3.2.2 | ขั้นตอนการดำเนินแผนการจำลอง | 34 |
| บทที่ 4 ผลการวิจัย | | 38 |
| 4.1 | ผลลัพธ์จากการจำลองสถานการณ์ | 38 |
| 4.1.1 | การจำลองภายใต้การแจกแจง Conway-Maxwell-Poisson | 38 |
| 4.1.2 | การจำลองภายใต้การแจกแจงปัวซอง | 39 |
| 4.1.3 | การจำลองภายใต้การแจกแจงเรขาคณิต | 40 |
| 4.1.4 | การจำลองภายใต้การแจกแจงทวินามเชิงลบ | 41 |
| 4.2 | การประยุกต์ใช้กับสถานการณ์จริง | 67 |
| 4.2.1 | ข้อมูลโสเภณีในเมืองแวนคูเวอร์ (Street prostitutes in Vancouver) | 67 |

สารบัญ(ต่อ)

| | |
|---|----|
| บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ | 70 |
| 5.1 สรุปผลการวิจัย | 70 |
| 5.1.1 สรุปผลการพัฒนาตัวประมาณขนาดประชากร | 70 |
| 5.1.2 สรุปผลการเปรียบเทียบประสิทธิภาพตัวประมาณขนาดประชากร | 71 |
| 5.2 ข้อเสนอแนะเพิ่มเติม | 78 |
| รายการอ้างอิง | 79 |
| ภาคผนวก | |
| ภาคผนวก ก คำสั่งการทำงานของโปรแกรม Rstudio สำหรับเปรียบเทียบ ประสิทธิภาพตัวประมาณขนาดประชากร | 82 |
| ประวัติผู้เขียน | 96 |

สารบัญตาราง

| ตารางที่ | หน้า |
|---|------|
| 1.1 จำนวนหมึกกรีซลีเพคเมียในอุทยานแห่งชาติเยลโลว์สโตนถูกสำรวจพบ y ครั้ง โดยที่ $y = 1, 2, 3, \dots, 7$ | 2 |
| 4.1 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 0.5$, $\nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 43 |
| 4.2 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 1$, $\nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 44 |
| 4.3 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 2$, $\nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 45 |
| 4.4 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 0.5$, $\nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 46 |
| 4.5 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 1$, $\nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 47 |
| 4.6 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 2$, $\nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 48 |
| 4.7 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 0.5$, $\nu = 0.8$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 49 |
| 4.8 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อ กำหนด $\lambda = 1$, $\nu = 0.8$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 50 |

สารบัญตาราง(ต่อ)

| | |
|--|----|
| 4.18 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Poisson</i> (λ) เมื่อ กำหนด $\lambda = 2$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 60 |
| 4.19 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Geometric</i> (p) เมื่อ กำหนด $p = 0.3$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 61 |
| 4.20 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Geometric</i> (p) เมื่อ กำหนด $p = 0.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 62 |
| 4.21 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Geometric</i> (p) เมื่อ กำหนด $p = 0.7$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 63 |
| 4.22 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Negative Binomial</i> (k, θ) เมื่อกำหนด $(k, \theta) = (1, 0.3)$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 64 |
| 4.23 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Negative Binomial</i> (k, θ) เมื่อกำหนด $(k, \theta) = (3, 0.5)$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 65 |
| 4.24 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง <i>Negative Binomial</i> (k, θ) เมื่อกำหนด $(k, \theta) = (5, 0.7)$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000 | 66 |
| 4.25 จำนวนโสเภณีที่ถูกจับในเมืองแวนคูเวอร์ ประเทศแคนาดา | 67 |
| 4.26 ค่าอนุमानโสเภณีในเมืองแวนคูเวอร์ ประเทศแคนาดา | 69 |

สารบัญตาราง(ต่อ)

| | | |
|-----|--|----|
| 5.1 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความเอนเอียงสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson | 72 |
| 5.2 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความแปรปรวนสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson | 73 |
| 5.3 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson | 73 |
| 5.4 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความเอนเอียงสัมพัทธ์ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ | 75 |
| 5.5 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความแปรปรวนสัมพัทธ์ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ | 76 |
| 5.6 | ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจาก ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ | 76 |

สารบัญภาพ

| ภาพที่ | หน้า |
|--------|------|
| 3.1 | 37 |
| 4.1 | 68 |

3.1 แผนผังแสดงการทำงานของโปรแกรมจำลองตัวแบบ

4.1 แสดงการอนุมานข้อมูลโสมถณีในเมืองแวนคูเวอร์ ประเทศแคนาดา ด้วยการแจกแจง
ปัวซอง การแจกแจงทวินามเชิงลบ การแจกแจง Conway-Maxwell-Poisson ตัดปลาย
ที่ศูนย์ และการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3



รายการสัญลักษณ์และคำย่อ

| สัญลักษณ์/คำย่อ | คำเต็ม/คำจำกัดความ |
|-----------------|---|
| N | ขนาดประชากร |
| \hat{N} | ตัวประมาณขนาดประชากร |
| f_y | จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ y ครั้ง โดยที่ $y = 0,1,2,\dots$ |
| f_0 | จำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบหรือถูกสำรวจพบศูนย์ครั้ง |
| m | จำนวนครั้งสูงสุดที่หน่วยตัวอย่างใดๆถูกสำรวจพบซ้ำ |
| n | ขนาดตัวอย่าง เมื่อ $n = f_1 + f_2 + \dots + f_m$ |
| \hat{f}_0 | ค่าประมาณจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบหรือถูกสำรวจพบศูนย์ครั้ง |
| p_y | โอกาสที่หน่วยตัวอย่างใดๆถูกสำรวจพบ y ครั้ง โดยที่ $y = 0,1,2,\dots$ |
| p_0 | โอกาสที่หน่วยตัวอย่างไม่ถูกสำรวจพบ |
| \hat{p}_0 | ค่าประมาณโอกาสที่หน่วยตัวอย่างไม่ถูกสำรวจพบ |
| $1 - p_0$ | โอกาสที่หน่วยตัวอย่างถูกสำรวจพบ |
| $1 - \hat{p}_0$ | ค่าประมาณโอกาสที่หน่วยตัวอย่างถูกสำรวจพบ |

รายการสัญลักษณ์และคำย่อ(ต่อ)

| สัญลักษณ์/คำย่อ | คำเต็ม/คำจำกัดความ |
|------------------|--|
| \hat{N}_{HT} | ตัวประมาณขนาดประชากรของ Horvitz-Thomson |
| \hat{N}_{MLE} | ตัวประมาณขนาดประชากรภาวะน่าจะเป็นสูงสุด |
| \hat{N}_{Chao} | ตัวประมาณขนาดประชากรของ Chao |
| \hat{N}_{LB} | ตัวประมาณขนาดประชากรของ Lanumteang-Böhning |
| \hat{N}_{LCMP} | ตัวประมาณขนาดประชากรจากสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) |
| \hat{N}_{ZCMP} | ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ |
| \hat{N}_{YCMP} | ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3 |
| $RBias$ | ความเอนเอียงสัมพัทธ์ |
| $RVar$ | ความแปรปรวนสัมพัทธ์ |
| $RMSE$ | ความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ |

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

Capture-Recapture เป็นวิธีที่ใช้ประมาณขนาดประชากรที่พบได้ยากในงานด้านต่างๆ ได้แก่ ด้านชีววิทยา จะประมาณขนาดประชากรของสัตว์ที่มีการหลบซ่อนตัว เคลื่อนย้ายถิ่นอาศัยอยู่ตลอดเวลา ซึ่งยากต่อการสำรวจขนาดประชากรสัตว์ทั้งหมด เช่น นักชีววิทยาสนใจจำนวนของหมึกกรีซลีเพศเมียในอุทยานแห่งชาติเยลโลว์สโตน ปี ค.ศ. 1996-1998 (Chao and Huggins, 2013) ซึ่งในการสำรวจประชากรสัตว์จะทำการสุ่มตัวอย่างครั้งแรกจากประชากร โดยทำเครื่องหมายแล้วปล่อยกลับคืน หลังจากนั้นรอจนระยะเวลาหนึ่งจึงทำการสุ่มตัวอย่างจากประชากรเดิมอีกครั้งและทำเครื่องหมาย ทำซ้ำๆ กันหลายครั้ง (Repeated Identification) ซึ่งในหน่วยตัวอย่างที่ได้ทำการสุ่มในแต่ละครั้งอาจมีหน่วยตัวอย่างซ้ำจากการสุ่มครั้งก่อนๆ ที่ถูกทำเครื่องหมายไว้ ด้านระบาดวิทยา ด้านการแพทย์และสาธารณสุข จะสนใจจำนวนการติดเชื้อ การเจ็บป่วย และการเสียชีวิต เพื่อตรวจสอบสุขภาพของประชากร หาสาเหตุของการเกิดโรค การวางแผนและหาวิธีป้องกัน นอกจากนี้ยังประยุกต์ใช้ในด้านสังคมและด้านวิศวกรรมอีกด้วย ซึ่งเครื่องมือเพื่อให้ได้หน่วยตัวอย่างอาจจะได้มาจาก การวางกับดัก การบันทึก และการลงทะเบียน ยกตัวอย่างเช่น จำนวนเด็กป่วยด้วยโรคเบาหวานในเมืองมาดริดประเทศสเปนที่ได้จากบันทึกของโรงพยาบาล (McCarty et al., 1992) จำนวนโสเภณีในเมืองลาสโกว์ ประเทศอังกฤษจากบันทึกของตำรวจ (Mckeganey et al., 1992) เป็นต้น

ลักษณะของประชากรที่เราสนใจในงานด้าน Capture-Recapture จะเป็นประชากรปิด (Closed Population) ซึ่งไม่มีการเกิด การตาย หรือการย้ายถิ่นที่อยู่อาศัยในช่วงระยะเวลาของการศึกษา ในที่นี้กำหนดให้ N แทนขนาดประชากร ซึ่งเป็นค่าคงที่แต่ไม่ทราบค่า จากการสุ่มตัวอย่างโดยทำซ้ำๆ กันหลายครั้งภายใต้ระยะเวลาการศึกษา ทำให้ทราบจำนวนครั้งที่หน่วยตัวอย่างแต่ละหน่วยถูกสำรวจพบ กำหนดให้ Y แทนจำนวนครั้งที่หน่วยตัวอย่างถูกสำรวจพบ โดยที่ $y = 0, 1, 2, \dots, m$ ซึ่ง m เป็นจำนวนครั้งที่สูงสุดที่หน่วยตัวอย่างถูกสำรวจพบ และให้ $f_0, f_1, f_2, \dots, f_m$ แทนจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ $0, 1, 2, \dots, m$ ครั้งตามลำดับ จะเห็นว่า f_0 คือจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบหรือถูกสำรวจพบศูนย์ครั้ง ซึ่งเราไม่ทราบค่า ตัวอย่างเช่น จำนวนหมึกกรีซลีเพศเมียในอุทยานแห่งชาติเยลโลว์สโตนที่ถูกสำรวจพบในปี ค.ศ. 1996-1998 (Chao and Huggins, (2013)) โดยให้ f_y แทนจำนวนหมึกกรีซลีเพศเมียที่ถูกสำรวจพบ y ครั้ง โดยที่ $y = 1, 2, \dots, 7$ ดังตารางที่ 1.1

ตารางที่ 1.1 จำนวนหมีกรีซลีเพศเมียในอุทยานแห่งชาติเยลโลว์สโตนที่ถูกสำรวจพบ y ครั้ง

| ปี | จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ y ครั้ง (f_y) | | | | | | | | |
|------|---|-------|-------|-------|-------|-------|-------|-------|-----|
| | f_0 | f_1 | f_2 | f_3 | f_4 | f_5 | f_6 | f_7 | n |
| 1996 | ? | 15 | 10 | 2 | 1 | 0 | 0 | 0 | 28 |
| 1997 | ? | 13 | 7 | 4 | 1 | 3 | 0 | 1 | 29 |
| 1998 | ? | 11 | 13 | 5 | 1 | 1 | 0 | 2 | 33 |

จะเห็นว่าจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ $n = f_1 + f_2 + \dots + f_m$ ดังนั้นขนาดประชากร $N = f_0 + n$ ซึ่ง f_0 เราไม่ทราบค่า ด้วยเหตุนี้จึงมีฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มที่เป็นข้อมูลแจงนับเข้ามาเกี่ยวข้อง นั่นคือ ให้ p_y แทนโอกาสที่หน่วยตัวอย่างแต่ละตัวถูกสำรวจพบ y ครั้ง สำหรับ $y=0,1,2,\dots,m$ อย่างไรก็ตาม โดยธรรมชาติของข้อมูลที่เก็บโดยวิธี Capture-Recapture เราไม่ทราบจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ (f_0) และไม่ทราบ โอกาสที่หน่วยตัวอย่างไม่ถูกสำรวจพบ (p_0) เราจึงจะทำการประมาณค่าทั้ง f_0 และ p_0 และเนื่องจากจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบคือ $f_0 = Np_0$ ขณะเดียวกันจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ $n = N(1 - p_0)$ ดังนั้นขนาดประชากร $N = N(1 - p_0) + Np_0$ ทำให้ได้ตัวประมาณขนาดประชากรของ Horvitz-Thompson (Horvitz and Thomson, 1952) คือ $\hat{N}_{HT} = n / (1 - \hat{p}_0)$

จากการที่ข้อมูลใน Capture-Recapture เป็นข้อมูลแจงนับ การแจกแจงของข้อมูลแจงนับที่รู้จักโดยทั่วไปได้แก่ การแจกแจงปัวซองที่มีพารามิเตอร์ λ ซึ่งเป็นตัวแบบพื้นฐานในการวิเคราะห์ข้อมูลโดยมีคุณสมบัติการกระจายเท่ากัน (Equi-dispersion) นั่นคือ ค่าเฉลี่ยมีค่าเท่ากับ ความแปรปรวน อย่างไรก็ตาม ในทางปฏิบัติข้อมูลอาจมีค่าเฉลี่ยที่ไม่เท่ากับความแปรปรวน นั่นคือ ถ้าความแปรปรวนมากกว่าค่าเฉลี่ย เรียกกรณีนี้ว่า Over-dispersion และกรณีที่ความแปรปรวนน้อยกว่าค่าเฉลี่ย ซึ่งเรียกกรณีนี้ว่า Under-dispersion ด้วยเหตุนี้เราจึงสนใจการแจกแจง Conway-Maxwell-Poisson (CMP) (Shmueli et al., 2005) ซึ่งเป็นการแจกแจงนับทั่วไปและครอบคลุมข้อมูลที่ เป็น Over-dispersion และ Under-dispersion โดยฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม Y ภายใต้การแจกแจง Conway-Maxwell-Poisson (CMP) ที่มีพารามิเตอร์ λ และ ν เป็นดังนี้

$$f_{CMP}(y; \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)} \quad y = 0, 1, 2, \dots$$

โดยที่

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

$\lambda > 0$, $\nu \geq 0$ และ $Z(\lambda, \nu)$ เป็นค่าคงที่ซึ่งได้จากการประมาณ ในที่นี้การแจกแจง Conway-Maxwell-Poisson สามารถทำให้อยู่ในรูปการแจกแจงไม่ต่อเนื่องอื่นๆ ได้ เมื่อพารามิเตอร์ ν มีการแปรผัน เช่น เมื่อ $\nu = 1$ ทำให้ $Z(\lambda, \nu) = e^\lambda$ ซึ่งสอดคล้องกับการแจกแจงปัวซองที่มีพารามิเตอร์ λ ในกรณีที่ ν ลู่เข้าใกล้อนันต์ ($\nu \rightarrow \infty$) จะทำให้ $Z(\lambda, \nu)$ ลู่เข้าใกล้ $1 + \lambda$ ซึ่งสอดคล้องกับการแจกแจงแบร์นูลลีที่มีพารามิเตอร์ $\lambda/(1 + \lambda)$ และในกรณีที่ $\nu = 0$ และ $0 < \lambda < 1$ จะทำให้ $Z(\lambda, \nu) = 1/(1 - \lambda)$ จะสอดคล้องกับการแจกแจงเรขาคณิตที่มีพารามิเตอร์ $1 - \lambda$

การแจกแจง Conway-Maxwell-Poisson มีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}$$

$$V(Y) \approx \frac{\lambda^{1/\nu}}{\nu}$$

อัตราส่วนของความแปรปรวนและค่าเฉลี่ย $\frac{V(Y)}{E(Y)} \approx \frac{1}{\nu}$ จะได้ว่า $\nu < 1$ กรณีที่เป็น Over-dispersion และ $\nu > 1$ กรณีที่เป็น Under-dispersion

เนื่องจากข้อมูลแจงนับใน Capture-Recapture ไม่มีศูนย์ครั้งและข้อมูลส่วนใหญ่ถูกสำรวจพบ 1, 2 และ 3 ครั้ง ซึ่งจำนวนครั้งที่ถูกสำรวจพบที่มากกว่า 3 ครั้งเป็นไปได้ยาก ดังนั้นเราจึงพัฒนาตัวประมาณขนาดประชากร 2 ตัว ภายใต้การแจกแจง Conway-Maxwell-Poisson โดยตัวประมาณขนาดประชากรตัวแรกจะอยู่ภายใต้การตัดปลายที่ศูนย์ (Zero-truncated) ซึ่งเป็นรูปแบบการตัดปลายโดยทั่วไปของวิธี Capture-Recapture ส่วนตัวประมาณขนาดประชากรอีกตัวจะอยู่ภายใต้การตัดปลายที่มากกว่า 3 (truncated larger than three)

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson

1.3 ขอบเขตการศึกษา

1.3.1 ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson โดยตัดปลายที่ศูนย์

1.3.2 ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson โดยอาศัยจำนวนหน่วยตัวอย่างที่ถูกสุ่มสำรวจพบ 1, 2 และ 3 ครั้ง

1.3.3 ตัวประมาณพารามิเตอร์ λ และ ν ที่ได้จากกระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) ในสถานการณ์จำลองทำวนซ้ำๆจนกระทั่งค่าประมาณพารามิเตอร์ในรอบสุดท้ายแตกต่างจากรอบก่อนหน้าทศนิยม 4 ตำแหน่ง โดยใช้ฟังก์ชัน nlminb ในโปรแกรม Rstudio

1.3.4 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาทั้ง 2 ตัวกับตัวประมาณขนาดประชากรของ Chao, Lanumteang-Böhning, สมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) และ ภาวะน่าจะเป็นสูงสุด (MLE) โดยศึกษาจากการจำลองสถานการณ์ด้วยโปรแกรม Rstudio ในแต่ละสถานการณ์จำลองทำการวนซ้ำๆจนครบ 2,000 รอบ ซึ่งขอบเขตการศึกษาเป็นดังต่อไปนี้

1.3.4.1 จำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson กำหนดพารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ และ $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$ และกำหนดขนาดประชากร $N = 500, 1,000$ และ $10,000$ ตามลำดับ

1.3.4.2 จำลองสถานการณ์ภายใต้การแจกแจงอื่นๆได้แก่ การแจกแจงปัวซอง กำหนดพารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ การแจกแจงเรขาคณิต กำหนดพารามิเตอร์ $p = 0.3, 0.5, 0.7$ และการแจกแจงทวินามเชิงลบ $(k, \theta) = (1, 0.3), (3, 0.5), (5, 0.7)$ และกำหนดขนาดประชากร $N = 500, 1,000$ และ $10,000$ ตามลำดับ

1.3.4.3 ประยุกต์ใช้ตัวประมาณขนาดประชากรที่นำเสนอประมาณขนาดประชากร
ในสถานการณ์จริง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-
Poisson

1.4.2 ได้ตัวประมาณขนาดประชากรที่มีประสิทธิภาพ

1.4.3 สามารถประยุกต์ใช้ตัวประมาณขนาดประชากรที่พัฒนากับสถานการณ์จริงได้

1.5 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพตัวประมาณขนาดประชากร

1.5.1 พิจารณาความแม่นยำ (Accuracy) ในการประมาณขนาดประชากรจากค่าความ
เอนเอียงสัมพัทธ์ (Relative Bias)

$$RBias = \frac{\bar{\hat{N}} - N}{N}$$

โดยที่ $\bar{\hat{N}} = \frac{1}{B} \sum_{l=1}^B \hat{N}_l$ เมื่อ B แทนจำนวนรอบของการวนซ้ำๆ ในการจำลอง และค่าความ
คลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative Mean Square Error)

$$RMSE = \frac{1}{N^2} \sum_{l=1}^B \frac{(\hat{N}_l - N)^2}{B}$$

และพิจารณาความเที่ยง (Precision) ในการประมาณขนาดประชากรจากค่าความแปรปรวนสัมพัทธ์
(Relative Variance)

$$RVar = \frac{1}{N^2} \sum_{l=1}^B \frac{(\hat{N}_l - \bar{\hat{N}})^2}{B}$$

1.5.2 การเปรียบเทียบตัวประมาณขนาดประชากรที่มีประสิทธิภาพในการประมาณขนาดประชากรที่เหมาะสมที่สุด พิจารณาจากตัวประมาณขนาดประชากรที่ให้ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์มีค่าต่ำสุด เป็นตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงที่สุด



บทที่ 2

ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในบทที่ 2 จะกล่าวถึงทฤษฎี และงานวิจัยที่เกี่ยวข้องในการพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson โดยได้รวบรวมวรรณกรรมและงานวิจัยที่อยู่ภายใต้ขอบเขตที่ได้กล่าวไว้ข้างต้น

2.1 การแจกแจง Conway-Maxwell-Poisson

การแจกแจงปัวซอง เป็นหนึ่งในการแจกแจงไม่ต่อเนื่องที่ถูกนำไปใช้มากที่สุด โดยมีคุณสมบัติ ค่าเฉลี่ยเท่ากับความแปรปรวน (Equi-dispersion) แต่ข้อมูลอาจมีค่าเฉลี่ยไม่เท่ากับความแปรปรวน นั่นคือ การกระจายมากเกินไป (Over-dispersion) และการกระจายน้อยเกินไป (Under-dispersion) ดังนั้น Conway และ Maxwell (1962) จึงได้แนะนำการแจกแจงที่มีความหลากหลาย นั่นคือ การแจกแจง Conway-Maxwell-Poisson (CMP) (Shmueli et al., 2005) ซึ่งครอบคลุมทั้งข้อมูลที่เป็น Over-dispersion และ Under-dispersion และเป็นส่วนขยายของการแจกแจงปัวซองที่ประกอบด้วยพารามิเตอร์นอกเหนือจาก λ นั่นคือ ν และมีความพิเศษตรงที่การแจกแจง Conway-Maxwell-Poisson เป็นการแจกแจงย่อยทั่วไปของการแจกแจงอื่นๆ เมื่อพารามิเตอร์ ν แปรผัน เช่น การแจกแจงปัวซอง การแจกแจงแบร์นูลลี และการแจกแจงเรขาคณิต เป็นต้น สำหรับฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson เป็นดังต่อไปนี้

$$f_{CMP}(y; \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)} \quad y = 0, 1, 2, \dots$$

โดยที่

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

$\lambda > 0$, $\nu \geq 0$ และ $Z(\lambda, \nu)$ เป็นค่าคงที่ที่ได้จากการประมาณ การแจกแจง Conway-Maxwell-Poisson จะเป็นรูปแบบโดยทั่วไปของการแจกแจงไม่ต่อเนื่อง เนื่องจากพารามิเตอร์ ν มีค่าแปรผันดังต่อไปนี้

กรณี $\nu = 1$ จะทำให้ $Z(\lambda, \nu) = e^\lambda$ สอดคล้องกับการแจกแจงปัวซอง

กรณี $\nu \rightarrow \infty$ จะทำให้ $Z(\lambda, \nu) \rightarrow 1 + \lambda$ สอดคล้องกับการแจกแจงแบร์นูลลีที่มีพารามิเตอร์ของการแจกแจง $\lambda/(1 + \lambda)$

กรณี $\nu = 0$ และ $0 < \lambda < 1$ จะทำให้ $Z(\lambda, \nu) = 1/(1 - \lambda)$ สอดคล้องกับการแจกแจงเรขาคณิตที่มีพารามิเตอร์ $1 - \lambda$

การแจกแจง Conway-Maxwell-Poisson มีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}$$

$$V(Y) \approx \frac{\lambda^{1/\nu}}{\nu}$$

อัตราส่วนของความแปรปรวนและค่าเฉลี่ย $\frac{V(Y)}{E(Y)} \approx \frac{1}{\nu}$ จะได้ว่า $\nu < 1$ กรณีที่เป็น Over-dispersion และ $\nu > 1$ กรณีที่เป็น Under-dispersion

2.2 การแจกแจงเรขาคณิต (Geometric distribution)

ให้ตัวแปรสุ่ม Y แทนจำนวนครั้งของการทดลองจนกว่าจะล้มเหลวเป็นครั้งแรก ซึ่ง จะทำการทดลองแบบแบร์นูลลีซ้ำๆ กัน โดยให้ p แทนโอกาสที่การทดลองแต่ละครั้งล้มเหลว สำหรับ ฟังก์ชันความน่าจะเป็นของการแจกแจงเรขาคณิต เป็นดังนี้

$$f(y; p) = p(1 - p)^y, \quad y = 0, 1, 2, \dots$$

มีค่าคาดหวังและความแปรปรวน ดังนี้

$$E(Y) = \frac{1-p}{p}$$

$$V(Y) = \frac{1-p}{p^2}$$

2.3 การแจกแจงทวินามเชิงลบ (Negative Binomial distribution)

ให้ตัวแปรสุ่ม Y มีการแจกแจงทวินามเชิงลบ ซึ่งขยายจากการแจกแจงปัวซองผสม การแจกแจงแกมมา สำหรับฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามเชิงลบ เป็นดังนี้

$$f(y; k, \theta) = \frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)} \theta^k (1-\theta)^y, \quad y = 0, 1, 2, \dots$$

โดยที่ $0 < \theta < 1$ และ $k > 0$ และมีค่าคาดหวังและความแปรปรวน ดังนี้

$$E(Y) = \frac{k(1-\theta)}{\theta}$$

$$V(Y) = \frac{k(1-\theta)}{\theta^2}$$

2.4 กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (Expectation-Maximization Algorithm)

กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (Expectation-Maximization Algorithm) หรือ EM-algorithm เป็นวิธีการเชิงตัวเลขที่อาศัยหลักการวนซ้ำๆ (iterative procedure) เพื่อหาค่าประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ของพารามิเตอร์ θ ของการแจกแจงในกรณีที่มีข้อมูลสูญหาย โดย EM-algorithm แบ่งออกเป็น 2 ขั้นตอนคือ กระบวนการหาค่าคาดหวัง (Expectation: E-step) และกระบวนการหาค่าสูงสุด (Maximization: M-step) โดยกำหนดให้ Y_{obs} แทนเซตของข้อมูลที่สังเกตได้ของการแจกแจงหนึ่ง และ Y_{mis} แทน

ค่าที่สูญหาย ดังนั้น $Y = (Y_{obs}, Y_{mis})$ ซึ่งเป็นข้อมูลที่สมบูรณ์ (Complete data) ให้ $p(y | \theta)$ แทนฟังก์ชันความน่าจะเป็นของการแจกแจงที่มีพารามิเตอร์ θ จากฟังก์ชันความน่าจะเป็นของการแจกแจง จะได้

$$L(\theta) = \prod_Y p(y | \theta)$$

และล็อกฟังก์ชันความน่าจะเป็น

$$\log L(\theta) = \sum_Y \log(p(y | \theta))$$

เมื่อมีข้อมูลสูญหาย จะเรียกล็อกฟังก์ชันความน่าจะเป็นที่ข้อมูลไม่สมบูรณ์ (Observed Incomplete Log-likelihood) ซึ่งในกระบวนการหาค่าคาดหวัง (Expectation: E-step) จะประมาณค่าคาดหวังของล็อกฟังก์ชันความน่าจะเป็นภายใต้ข้อมูลที่สมบูรณ์ จะได้ว่า

$$Q(\theta) = E(\log L(\theta) | Y_{obs}; \theta) = \hat{Y}_{mis}$$

โดยมีค่าคาดหวังของข้อมูลที่สูญหายและล็อกฟังก์ชันความน่าจะเป็นดังนี้

$$Q(\theta) = E(Y_{mis} | Y_{obs}; \theta) \log(p(y_{mis}; \theta)) + \sum_{Y_{obs}} \log(p(y_{obs}; \theta))$$

สำหรับกระบวนการหาค่าสูงสุด (Maximization: M-step) จะแทนค่าคาดหวังของข้อมูลที่สูญหายจากขั้นตอน E-step และประมาณค่าคาดหวังจากล็อกฟังก์ชันความน่าจะเป็นภายใต้ข้อมูลที่สมบูรณ์ (Unobserved Complete Log-likelihood) ซึ่งมีขั้นตอนดังนี้

ขั้นตอนที่ 0: กำหนดค่า $\hat{\theta}^{(r)}$ รอบที่ 0 และกำหนดค่าเริ่มต้น $r=0$ โดยที่ r แทนจำนวนรอบ

ขั้นตอนที่ 1: แทนค่า $\hat{\theta}^{(r)}$ ลงใน $\hat{Y}_{mis}^{(r+1)}$

ขั้นตอนที่ 2: แทนค่า $\hat{Y}_{mis}^{(r+1)}$ ที่ได้จากขั้นตอนที่ 1 ลงใน $\hat{\theta}^{(r+1)}$ และให้ $r = r + 1$

จากขั้นตอนที่ 2 จะได้ค่า $\hat{\theta}$ ที่ปรับค่าแล้ว หลังจากนั้นนำค่า $\hat{\theta}$ ดังกล่าวโดยในรอบที่ $r = r + 1$ กลับไปแทนค่าในขั้นตอนที่ 1 เพื่อหา \hat{Y}_{mis} อีกครั้ง ทำการวนซ้ำๆจนกระทั่งค่าประมาณ $\hat{\theta}$ มีการลู่เข้า จึงจะได้ตัวประมาณพารามิเตอร์ของ θ

2.5 ตัวประมาณขนาดประชากร

ในปี ค.ศ. 1952 Horvitz และ Thompson ได้เสนอตัวประมาณขนาดประชากร ภายใต้การแจกแจงทวินาม (Binomial Distribution) โดยกำหนดให้ N แทนขนาดประชากร ทำการสุ่มตัวอย่างขนาด n ซึ่งเป็นหน่วยตัวอย่างที่ถูกสำรวจพบ โดยโอกาสที่หน่วยตัวอย่างแต่ละหน่วย ถูกสำรวจพบเท่ากับ $1 - p_0$ และกำหนดให้ Z_i แทนตัวชี้วัดที่หน่วยตัวอย่างตัวที่ i ถูกสำรวจพบ โดยที่ $i = 1, 2, \dots, N$ โดยที่ $Z_i = 1$ เมื่อหน่วยตัวอย่างตัวที่ i ถูกสำรวจพบด้วยความน่าจะเป็น $1 - p_0$ และ $Z_i = 0$ เมื่อหน่วยตัวอย่างที่ i ไม่ถูกสำรวจพบด้วยความน่าจะเป็น p_0 และกำหนดให้ $n = \sum_{i=1}^N Z_i$ แทนจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ ซึ่งค่าที่เป็นไปได้ของ n คือ $n = 1, 2, \dots, n$ ดังนั้นโอกาสที่หน่วยตัวอย่างขนาด n ถูกสำรวจพบภายใต้การแจกแจงทวินามมีค่าเท่ากับ $\binom{N}{n} (1 - p_0)^n (p_0)^{N-n}$ และตัวประมาณภาวะน่าจะเป็นสูงสุดของ N ซึ่งเรียกตัวประมาณนี้ว่าตัวประมาณขนาดประชากรของ Horvitz-Thompson

$$\hat{N}_{HT} = \frac{n}{1 - \hat{p}_0}$$

2.5.1 ตัวประมาณขนาดประชากรภาวะน่าจะเป็นสูงสุด (Maximum Likelihood)

การแจกแจงปัวซองที่มีพารามิเตอร์ λ เป็นการแจกแจงของข้อมูลแจกแจงนับที่นิยมใช้ โดยทั่วไป โดยมีฟังก์ชันความน่าจะเป็นของการแจกแจง ดังนี้

$$f(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

ในการหาตัวประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด เป็นวิธีที่ได้รับความนิยมในการประมาณค่าพารามิเตอร์ ในการศึกษา Capture-Recapture กำหนดให้ Y แทนจำนวนครั้งที่หน่วยตัวอย่างถูกสำรวจพบ และให้ f_y แทนจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ y ครั้ง โดยที่ $y=1,2,\dots,m$ โดยที่ m แทนจำนวนครั้งสูงสุดในงานด้าน Capture-Recapture ข้อมูลจะไม่มีคุณสมบัติ เนื่องจากหน่วยตัวอย่างไม่มีค่าศูนย์ นั่นคือเราไม่ทราบ f_0 ดังนั้นฟังก์ชันภาวะน่าจะเป็นที่นำมาใช้จะมาจากการแจกแจงปัวซองตัดปลายที่ศูนย์ โดยมีฟังก์ชันภาวะน่าจะเป็นดังต่อไปนี้

$$f_{ZTP}(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!(1 - e^{-\lambda})}, \quad y = 1, 2, \dots$$

ฟังก์ชันภาวะน่าจะเป็นของการแจกแจงปัวซองตัดปลายที่ศูนย์ ได้ดังนี้

$$L(\lambda) = \prod_{y=1}^m \left(\frac{e^{-\lambda} \lambda^y}{y!(1 - e^{-\lambda})} \right)^{f_y}$$

ลือกฟังก์ชันภาวะน่าจะเป็นที่ไม่สมบูรณ์ของการแจกแจงปัวซองตัดปลายที่ศูนย์ที่ f_0 ไม่ทราบค่า (Observed Incomplete Log-likelihood) ได้ดังนี้

$$\log L(\lambda) = -\lambda n + \log \lambda \sum_{y=1}^m y f_y - \sum_{y=1}^m f_y \log(y!) - n \log(1 - e^{-\lambda}) \quad (2.1)$$

พิจารณาหาค่าพารามิเตอร์ λ ที่ทำให้ค่าของลือกฟังก์ชันภาวะน่าจะเป็นของการแจกแจงปัวซองตัดปลายที่ศูนย์มีค่าสูงสุด ด้วยการหาอนุพันธ์อันดับ 1 ของฟังก์ชันเทียบ λ

$$\begin{aligned}\hat{\lambda} &= \frac{1}{n} \sum_{y=1}^m y f_y (1 - e^{-\hat{\lambda}}) \\ &= \frac{S}{n} (1 - e^{-\hat{\lambda}}) \\ &= \frac{S}{\hat{N}}\end{aligned}$$

โดยที่ $S = \sum_{y=1}^m y f_y$ และเนื่องจากตัวประมาณค่าพารามิเตอร์ λ ไม่ได้อยู่ในรูปของสมการปิด (No Closed Form Solution) จึงใช้วิธีการเชิงตัวเลขเข้ามาช่วยในการประมาณค่าพารามิเตอร์ λ โดยเลือกใช้กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) จะเห็นว่าลูปฟังก์ชันภาวะน่าจะเป็นข้างต้น ข้อมูลสามารถสังเกตได้แต่ไม่สมบูรณ์เพราะ f_0 ไม่ทราบค่า แต่เราต้องการลูปฟังก์ชันภาวะน่าจะเป็นที่ข้อมูลสมบูรณ์ ดังนั้นกระบวนการหาค่าคาดหวัง (Expectation: E-step) หาได้จาก

$$\begin{aligned}\hat{f}_0 &= E(f_0 | f_1, f_2, \dots, f_m; \lambda) \\ &= N p_0 \\ &= (n + f_0) e^{-\lambda}\end{aligned}$$

ดังนั้น

$$\hat{f}_0 = \frac{ne^{-\lambda}}{1 - e^{-\lambda}}$$

ในกระบวนการหาค่าสูงสุด (Maximization: M-step) แทน f_0 ด้วย \hat{f}_0 จะได้ลูปฟังก์ชันภาวะน่าจะเป็นที่ข้อมูลสมบูรณ์ (Unobserved Complete Log-likelihood) ของการแจกแจงปัวซองตัดปลายที่ศูนย์ จะได้ว่า

$$\log L(\lambda) = -\hat{f}_0 \lambda + \sum_{y=1}^m f_y \log \left(\frac{e^{-\lambda} \lambda^y}{y!} \right) \quad (2.2)$$

หาอนุพันธ์อันดับ 1 ของฟังก์ชันภาวะน่าจะเป็นเทียบกับ λ จะได้ตัวประมาณของ λ ดังนี้

$$\hat{\lambda} = \frac{S}{\hat{N}} = \frac{S}{\hat{f}_0 + n}$$

จากนั้นใช้กระบวนการเชิงตัวเลขในการหาค่า $\hat{\lambda}$ และ \hat{f}_0 ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 0: กำหนดค่า $\hat{\lambda}^{(0)}$ รอบที่ 0 และกำหนดค่าเริ่มต้น $r = 0$

$$\text{ขั้นตอนที่ 1: ให้ } \hat{f}_0^{(r+1)} = \frac{ne^{(-\hat{\lambda}^{(r)})}}{1 - e^{(-\hat{\lambda}^{(r)})}}$$

ขั้นตอนที่ 2: แทนค่า $\hat{f}_0^{(r+1)}$ ที่ได้จากขั้นตอนที่ 1 ลงใน $\hat{\lambda}^{(r+1)} = \frac{S}{\hat{f}_0^{(r+1)} + n}$ และให้ $r = r + 1$

จากขั้นตอนที่ 2 จะได้ค่า $\hat{\lambda}$ ที่ปรับค่าแล้ว หลังจากนั้นนำค่า $\hat{\lambda}$ ดังกล่าวโดยให้ $r = r + 1$ กลับไปขั้นตอนที่ 1 ทำการวนซ้ำๆจนกระทั่งค่าประมาณ $\hat{\lambda}$ มีการลู่เข้า นั่นคือ $\hat{\lambda}_{MLE}$ และทำให้ทราบว่า $\hat{p}_0 = e^{-\hat{\lambda}_{MLE}}$ จากตัวประมาณขนาดประชากรของ Horvitz-Thompson จะได้ค่าประมาณขนาดประชากรภาวะน่าจะเป็นสูงสุด ดังนี้

$$\hat{N}_{MLE} = \frac{n}{1 - e^{-\hat{\lambda}_{MLE}}}$$

2.5.2 ตัวประมาณขนาดประชากรของ Chao

พิจารณาตัวประมาณขนาดประชากรภายใต้การแจกแจงปัวซองที่มีพารามิเตอร์ λ คงที่ (Homogeneity Poisson Model) ซึ่งหมายความว่า ในการสุ่มหน่วยตัวอย่างแต่ละหน่วย ตัวอย่างนั้นโอกาสที่หน่วยตัวอย่างแต่ละหน่วยถูกสำรวจพบ y ครั้ง มีค่าเท่ากัน แต่ในทางปฏิบัติพบว่าประชากรอาจจะประกอบไปด้วยกลุ่มย่อยๆที่หลากหลาย ในแต่ละกลุ่มย่อยมีการแจกแจงปัวซองซึ่งมีพารามิเตอร์ λ ที่แตกต่างกัน (Heterogeneity Poisson Model) ดังนั้นโอกาสที่หน่วยตัวอย่างแต่ละหน่วยถูกสำรวจพบ y ครั้ง อาจมีค่าไม่เท่ากัน Chao จึงพัฒนาตัวแบบของตัวประมาณขนาดประชากรภายใต้การแจกแจงปัวซองซึ่งมีพารามิเตอร์ λ ที่แตกต่างกัน (Chao, 1987) โดยหน่วยตัวอย่างมาจากการแจกแจงปัวซองที่มีพารามิเตอร์ λ ที่แตกต่างกันและให้ $f(\lambda)$ แทนฟังก์ชันหนาแน่นน่าจะเป็นของ λ ดังนี้

$$p_y = P(Y = y) = \int_0^{\infty} \frac{e^{-\lambda} \lambda^y}{y!} f(\lambda) d\lambda \quad (2.3)$$

โดยที่ $y=0,1,2, \dots$ และ $\lambda > 0$ และกำหนดให้ X และ Y เป็นตัวแปรสุ่ม โดยที่ $X = u(\lambda)$ และ $Y = v(\lambda)$ ซึ่งจะใช้สมการของ Cauchy-Schwarz ดังนี้

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

$$\left(\int_0^\infty u(\lambda)v(\lambda)f(\lambda)d\lambda \right)^2 \leq \left(\int_0^\infty u(\lambda)^2 f(\lambda)d\lambda \right) \left(\int_0^\infty v(\lambda)^2 f(\lambda)d\lambda \right)$$

ให้ $u(\lambda) = \sqrt{e^{-\lambda} \lambda^{y-1}}$ และ $v(\lambda) = \sqrt{e^{-\lambda} \lambda^{y+1}}$ จะได้อสมการของ Cauchy-Schwarz ดังต่อไปนี้

$$\begin{aligned} \left(\int_0^\infty e^{-\lambda} \lambda^y f(\lambda)d\lambda \right)^2 &\leq \left(\int_0^\infty e^{-\lambda} \lambda^{y-1} f(\lambda)d\lambda \right) \left(\int_0^\infty e^{-\lambda} \lambda^{y+1} f(\lambda)d\lambda \right) \\ \left(y! \int_0^\infty \frac{e^{-\lambda} \lambda^y}{y!} f(\lambda)d\lambda \right)^2 &\leq \left((y-1)! \int_0^\infty \frac{e^{-\lambda} \lambda^{y-1}}{(y-1)!} f(\lambda)d\lambda \right) \left((y+1)! \int_0^\infty \frac{e^{-\lambda} \lambda^{y+1}}{(y+1)!} f(\lambda)d\lambda \right) \\ (y! p_y)^2 &\leq (y-1)! p_{y-1} (y+1)! p_{y+1} \\ \frac{y p_y}{p_{y-1}} &\leq \frac{(y+1) p_{y+1}}{p_y} \end{aligned}$$

อัตราส่วนของความน่าจะเป็นจำนวนนับที่ใกล้เคียง (Chao, 1987) ได้ดังนี้

$$\frac{p_1}{p_0} \leq \frac{2p_2}{p_1} \leq \frac{3p_3}{p_2} \leq \dots$$

กำหนดให้ $y=1$ จะได้ตัวประมาณเขตล่างของการประมาณจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ ดังนี้

$$\begin{aligned} \frac{p_1}{p_0} &\leq \frac{2p_2}{p_1} \\ \frac{f_1/N}{f_0/N} &\leq \frac{2f_2/N}{f_1/N} \end{aligned}$$

$$\frac{f_1}{f_0} \leq \frac{2f_2}{f_1}$$

จะได้ตัวประมาณของ f_0 คือ

$$\hat{f}_{0Chao} \geq \frac{f_1^2}{2f_2}$$

แทน \hat{f}_{0Chao} ลงใน $\hat{N} = n + \hat{f}_0$ จะได้ตัวประมาณขนาดประชากรของ Chao ดังนี้

$$\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}$$

เห็นได้ว่า ตัวประมาณขนาดประชากรของ Chao ใช้จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบเพียง 1 และ 2 ครั้งเท่านั้น จึงสะดวกและง่ายต่อการนำไปใช้

2.5.3 ตัวประมาณขนาดประชากรของ Lanumteang-Böhning

ในปี ค.ศ. 2011 Lanumteang และ Böhning (Lanumteang-Böhning, 2011) ได้ขยายตัวประมาณขนาดประชากรของ Chao โดยใช้จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 3 ครั้ง (f_3) เข้ามาช่วยในการประมาณขนาดประชากร จากสมการ (2.3) Lanumteang และ Böhning กำหนดให้ $f(\lambda) = \lambda^{k-1} \exp(-\lambda/\theta') / (\theta'^k \Gamma(k))$ แทนฟังก์ชันหนาแน่นน่าจะเป็นของการแจกแจงแกมมา (Gamma Distribution) ที่มีพารามิเตอร์ θ' และ $k > 0$ ทำให้สมการ (2.3) มีฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามเชิงลบ (Negative Binomial Distribution) ดังนี้

$$\begin{aligned} p_y &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^y}{y!} f(\lambda) d\lambda \\ &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^y}{y!} \frac{\lambda^{k-1} e^{-\lambda/\theta'}}{\Gamma(k) (\theta')^k} d\lambda \end{aligned}$$

กำหนดให้ $\theta' = (1-\theta)/\theta$ จะได้ว่า

$$p_y = P(Y = y) = \frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)} \theta^k (1-\theta)^y, \quad y=0,1,2,\dots \quad (2.4)$$

จากสมการ (2.4) เห็นได้ว่า Y มีการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ k และ θ และกำหนดให้ $r_y = yp_y / p_{y-1}$ ดังนั้นจะได้ $r_y = (k+y-1)(1-\theta)$ หรือเขียนให้อยู่ในรูปสมการเส้นตรง $r_y = (k-1)(1-\theta) + (1-\theta)y$ โดยสร้างกราฟของความสัมพันธ์ระหว่าง r_y และ y ซึ่งจะบ่งชี้ว่าเป็นการแจกแจงทวินามเชิงลบ เพื่อประมาณขนาดประชากรจะพิจารณาอนุกรมเทเลอร์ของ $\log(r_y)$ รอบจุด $k-1$ จะได้ว่า

$$\log(r_y) = \log(k+y-1) + \log(1-\theta) \approx \underbrace{\log(1-\theta) + \log(k-1)}_{\alpha} + \frac{1}{\underbrace{k-1}_{\beta}} y \quad (2.5)$$

เมื่อทำการกำหนดให้ $y=2$ และ $y=3$ จะทำให้ได้ $\log r_2 = \log(2f_2/f_1) = \alpha + 2\beta$ และ $\log r_3 = \log(3f_3/f_2) = \alpha + 3\beta$ แก้สมการจะได้ $\hat{\alpha} = 3\log(2f_2/f_1) - 2\log(3f_3/f_2)$ และ $\hat{\beta} = \log(3f_3/f_2) - \log(2f_2/f_1)$ แทนค่า $\hat{\alpha}$ และ $\hat{\beta}$ ที่ได้ลงในสมการ (2.5) โดยให้ $y=1$ จะได้ $\log r_1 = \log(f_1/f_0) = \alpha + \beta$ หรือจะได้อีกต่อไปนี้

$$\log\left(\frac{f_1}{f_0}\right) = 3\log\left(\frac{2f_2}{f_1}\right) - 2\log\left(\frac{3f_3}{f_2}\right) + \log\left(\frac{3f_3}{f_2}\right) - \log\left(\frac{2f_2}{f_1}\right) = 2\log\left(\frac{2f_2}{f_1}\right) - \log\left(\frac{3f_3}{f_2}\right)$$

จัดรูปใหม่ได้ดังนี้

$$\log(f_0) = \log(f_1) - \log\left(\frac{4f_2^2}{f_1^2}\right) + \log\left(\frac{3f_3}{f_2}\right) = \log\left(\frac{3f_1^3 f_3}{4f_2^3}\right)$$

ดังนั้นค่าประมาณของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ คือ

$$\hat{f}_{0LB} = \frac{3f_1^3 f_3}{4f_2^3}$$

แทน \hat{f}_{0LB} ลงใน $\hat{N} = n + \hat{f}_0$ จะได้ตัวประมาณขนาดประชากรของ Lanumteang-Böhning ดังนี้

$$\hat{N}_{LB} = n + \frac{3f_1^3 f_3}{4f_2^3}$$

เห็นได้ว่า ตัวประมาณขนาดประชากรของ Lanumteang-Böhning ได้ขยายตัวประมาณขนาดประชากรของ Chao โดยใช้จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 1, 2 และ 3 ครั้งเท่านั้น นั่นคือ f_1, f_2 และ f_3

2.5.4 ตัวประมาณขนาดประชากรของสมการถดถอยเชิงเส้นภายใต้ CMP

จากที่กล่าวข้างต้น การแจกแจง Conway-Maxwell-Poisson (CMP) เป็นการแจกแจงที่มีความพิเศษตรงที่สามารถเป็นตัวแทนของการแจกแจงไม่ต่อเนื่องที่สำคัญ ซึ่งครอบคลุมทั้งข้อมูลที่มีการกระจายมากเกินไป และการกระจายน้อยเกินไป ในปี ค.ศ. 2016 Anan, Böhning และ Maruotti ได้ใช้ประโยชน์จากการสร้างกราฟ โดยใช้กราฟสัดส่วน (ratio-plot) ซึ่งเป็นวิธีการเขียนกราฟสำหรับระบุรูปแบบการแจกแจงที่เป็น heterogeneity ในข้อมูล Capture-Recapture (Böhning et al., 2013) โดยที่ ratio-plot เป็นเครื่องมือในการประมาณความน่าจะเป็นของหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ p_0 และใช้สัดส่วนของความน่าจะเป็นของ 2 ค่าที่ติดกันในการประมาณพารามิเตอร์ของการแจกแจง เพื่อนำไปประมาณจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบและขนาดประชากรต่อไป โดยตัวประมาณถดถอยอยู่ภายใต้ ratio-plot ที่ทำการแปลงให้อยู่ในรูปของล็อกกาลิทึมหาค่าประมาณของพารามิเตอร์ เพื่อพัฒนาตัวประมาณขนาดประชากรภายใต้สมการถดถอยเชิงเส้นของการแจกแจง Conway-Maxwell-Poisson (Anan, 2016) ให้ X แทนจำนวนครั้งที่หน่วยตัวอย่างถูกสำรวจพบ โดยที่ $x = 0, 1, 2, \dots, m$ โดยที่ m คือจำนวนครั้งสูงสุดที่หน่วยตัวอย่างถูกสำรวจพบ ซึ่งฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson คือ

$$p_x = \frac{\lambda^x}{(x!)^\nu} \frac{1}{Z(\lambda, \nu)}, \quad x=0,1,2,\dots$$

โดยที่

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

โดยที่ $\lambda > 0$ และ $\nu \geq 0$ การแจกแจง Conway-Maxwell-Poisson มีสัดส่วนของความน่าจะเป็นของตัวแปรสุ่ม 2 ค่าที่ติดกัน ดังนี้

$$r_x = (x+1) \frac{p_{x+1}}{p_x}$$

เนื่องจากงานด้าน Capture-Recapture จะไม่มีศูนย์ ดังนั้นการแจกแจงที่ตัดปลายที่ศูนย์ (Zero-truncated Distribution) คือ $p_x / (1 - p_0)$ ซึ่งจะได้อัตราส่วนความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson ดังต่อไปนี้

$$r_x = (x+1) \frac{p_{x+1} / (1 - p_0)}{p_x / (1 - p_0)}$$

ดังนั้น

$$r_x = (x+1) \frac{\frac{\lambda^{x+1}}{((x+1)!)^\nu}}{\frac{\lambda^x}{(x!)^\nu}} = \lambda(x+1)^{1-\nu} \quad (2.6)$$

เนื่องจากสมการ (2.6) ไม่เป็นสมการเชิงเส้น จึงแปลงอัตราส่วนความน่าจะเป็นให้อยู่ในรูปของล็อกกาลีทิม จะได้สมการเชิงเส้นดังนี้

$$\log(r_x) = \log(\lambda) + (1 - \nu)\log(x + 1)$$

เมื่อพิจารณาในรูปสมการเชิงเส้น จะได้

$$\log(r_x) = \beta_0 + \beta_1 \log(x + 1) \quad (2.7)$$

โดยที่ $\lambda = \exp(\beta_0)$, $\nu = 1 - \beta_1$ จะได้ $\beta_0 = \log(\lambda)$ และ $\beta_1 = 1 - \nu$ เนื่องจาก $\lambda > 0$ และ $\nu \geq 0$ ดังนั้น $\beta_1 \leq 1$ ซึ่งเป็นข้อจำกัดของงานนี้ (Anan, 2016) ในการประมาณค่าความน่าจะเป็น โดยใช้ความสัมพันธ์ของจำนวนหน่วยตัวอย่าง f_x จากอัตราส่วนในสมการ (2.6)

$$r_x^* = (x + 1) \frac{\hat{p}_{x+1}}{\hat{p}_x} = (x + 1) \frac{f_{x+1}/N}{f_x/N} = (x + 1) \frac{f_{x+1}}{f_x}$$

และถือก้ออัตราส่วนจากสมการ (2.7)

$$\log(r_x^*) = \log\left((x + 1) \frac{f_{x+1}}{f_x}\right) = \beta_0 + \beta_1 \log(x + 1) + \varepsilon_x$$

ประมาณพารามิเตอร์ของสมการถดถอยเชิงเส้น β_0 และ β_1 ใช้ตัวแบบถดถอยเชิงเส้นแบบถ่วงน้ำหนักในการประมาณค่าพารามิเตอร์ในตัวแบบ จะได้

$$\sum_{x=1}^{m-1} W_x [\log(r_x^*) - \beta_0 - \beta_1 \log(x + 1)]^2$$

โดยที่ W_x แทนเมทริกซ์ถ่วงน้ำหนัก และ m คือจำนวนครั้งสูงสุดที่หน่วยตัวอย่างถูกสำรวจพบ

$$Y = \begin{pmatrix} \log \frac{2f_2}{f_1} \\ \log \frac{3f_3}{f_2} \\ \vdots \\ \log \frac{mf_m}{f_{m-1}} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & \log(2) \\ 1 & \log(3) \\ \vdots & \vdots \\ 1 & \log(m) \end{pmatrix}$$

ดังนั้น

$$W = \begin{bmatrix} \frac{1}{f_1} + \frac{1}{f_2} & 0 & \cdots & 0 \\ 0 & \frac{1}{f_2} + \frac{1}{f_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{f_{m-1}} + \frac{1}{f_m} \end{bmatrix}^{-1}$$

จะได้ว่า

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X'WX)^{-1} X'WY \quad (2.8)$$

จะได้ตัวประมาณของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ ดังต่อไปนี้

$$\log(r_0^*) = \log\left(\frac{f_1}{f_0}\right) = \hat{\beta}_0$$

จะได้ว่า

$$\begin{aligned} \frac{f_1}{f_0} &= \exp(\hat{\beta}_0) \\ \hat{f}_{0LCMP} &= f_1 \exp(-\hat{\beta}_0) \end{aligned}$$

แทน f_{0LCMP} ลงใน $\hat{N} = n + f_0$ จะได้ขนาดประชากรที่พัฒนาจากสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson ดังนี้

$$\hat{N}_{LCMP} = n + f_1 \exp(-\hat{\beta}_0)$$

จะเห็นว่าตัวประมาณขนาดประชากร LCMP (\hat{N}_{LCMP}) ถึงแม้ว่าจะไม่จำเป็นต้องใช้ $\hat{\beta}_1$ แต่ $\hat{\beta}_1 \leq 1$

2.6 งานวิจัยที่เกี่ยวข้อง

ในการหาตัวประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุดเป็นวิธีที่นิยมใช้ในงานด้าน Capture-Recapture เนื่องจากตัวประมาณค่าพารามิเตอร์ภายใต้การแจกแจงปัวซองนั้นไม่ได้อยู่ในรูปของสมการปิด (No Closed Form Solution) จึงใช้กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) ที่เป็นวิธีการเชิงตัวเลขเข้ามาช่วยในการประมาณค่าพารามิเตอร์

ในปี ค.ศ. 1987 Chao ได้พัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจงปัวซองซึ่งมีพารามิเตอร์ λ ที่แตกต่างกัน โดยเลือกใช้จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 1 และ 2 ครั้ง เนื่องจากจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบส่วนใหญ่ถูกสำรวจพบน้อยครั้ง ดังนั้นตัวประมาณขนาดประชากรของ Chao จึงมีความน่าเชื่อถือและถูกเลือกใช้บ่อยครั้ง โดยเงื่อนไขของจำนวนครั้งที่ถูกสำรวจพบ t ครั้งควรมีค่าสูงๆ และโอกาสที่หน่วยตัวอย่างที่ i ใดๆ ถูกสำรวจพบควรมีค่าน้อยๆ ซึ่งทำให้หน่วยตัวอย่างส่วนใหญ่ นั้นถูกสำรวจพบจำนวนน้อยครั้ง ตัวประมาณขนาดประชากรของ Chao ให้เพียงขอบเขตล่างของขนาดประชากร และเมื่อจำนวนครั้งที่ถูกสำรวจพบ t ครั้งสูงขึ้น ทำให้ค่าความคลาดเคลื่อนมาตรฐานต่ำลงด้วย นอกจากนี้โอกาสของการตกอยู่ในช่วงความเชื่อมั่นแบบ Log-transformation เข้าใกล้ 0.95 มากกว่าการตกอยู่ในช่วงเชื่อมั่นแบบปกติ ในบางสถานการณ์ตัวประมาณขนาดประชากรของ Chao จะให้ค่าประมาณขนาดประชากรและโอกาสของการตกอยู่ในช่วงความเชื่อมั่นจะต่ำกว่าค่าที่แท้จริง แต่ตัวประมาณขนาดประชากรของ Chao ยังคงมีความน่าเชื่อถือและใช้ได้จริงในทางปฏิบัติ

ในปี ค.ศ. 2011 Lanumteang และ Böhning ได้พัฒนาตัวประมาณขนาดประชากรที่มีแนวคิดต่อจาก Chao โดยพัฒนาภายใต้การแจกแจงปัวซอง ซึ่งพารามิเตอร์ λ ที่มีการแจกแจงแกมมา (Gamma Distribution) ซึ่งจะทำให้ได้จำนวนครั้งที่พบหน่วยตัวอย่างเป็นตัวแปรสุ่มที่มีการแจกแจงทวินามเชิงลบ นอกจากนี้ยังใช้จำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 1, 2 และ 3 ครั้ง และใช้ Log-Ratio ในการพัฒนาตัวประมาณขนาด พบว่า ตัวประมาณขนาดประชากรของ Lanumteang-

Böhning เป็นตัวประมาณไม่เอนเอียง (Unbiased estimator) ที่ให้ค่าความเอนเอียงต่ำ แต่ให้ค่าความแปรปรวนสูง

ในปี ค.ศ. 2016 Anan, Böhning และ Maruotti ได้พัฒนาตัวประมาณขนาดประชากรจากสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) โดยเลือกใช้ Ratio-plot ที่ทำการแปลงให้อยู่ในรูปของล็อกกาลีทิม ทำให้ได้สมการถดถอยเชิงเส้นและประมาณจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบศูนย์ครั้งได้จากค่าประมาณสัมประสิทธิ์การถดถอย ตัวประมาณขนาดประชากรที่พัฒนาขึ้นมานั้นมีความแม่นยำสูงสุด และมีความแปรปรวนสูง แต่เมื่อประชากรมีขนาดใหญ่ขึ้นจะทำให้ความแปรปรวนมีค่าลดต่ำลง ซึ่งตัวประมาณขนาดประชากร LCMP ยังเป็นตัวเลือกที่ดีในการประมาณขนาดประชากรเมื่อประชากรมีขนาดใหญ่



บทที่ 3 วิธีการวิจัย

ในบทที่ 3 นี้จะอธิบายเกี่ยวกับการหาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์และตัดปลายที่มากกว่า 3 นอกจากนี้ยังอธิบายเกี่ยวกับขั้นตอนการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาขึ้นมาจากการจำลองสถานการณ์จากการใช้โปรแกรม Rstudio โดยเปรียบเทียบกับตัวประมาณขนาดประชากรของ Chao, Lanumteang-Böhning, ภาวะน่าจะเป็นสูงสุด (MLE) และสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) ดังต่อไปนี้

3.1 ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson

จากที่กล่าวมาแล้วข้างต้นว่าข้อมูลแจกแจงนับโดยทั่วไปนิยมใช้การแจกแจงปัวซองที่มีค่าเฉลี่ยเท่ากับความแปรปรวน (Equi-dispersion) อย่างไรก็ตาม ข้อมูลอาจมีค่าเฉลี่ยที่ไม่เท่ากับ ความแปรปรวนเสมอไปนั่นคือ ความแปรปรวนมากกว่าค่าเฉลี่ยหรือการกระจายมากเกินไป (Over-dispersion) และความแปรปรวนน้อยกว่าค่าเฉลี่ยหรือการกระจายน้อยเกินไป (Under-dispersion) ด้วยเหตุนี้ เราจึงสนใจการแจกแจง Conway-Maxwell-Poisson (CMP) (Shmueli et al., 2005) ซึ่งเป็นการแจกแจงนับทั่วไปและครอบคลุมข้อมูลที่เป็น Over-dispersion และ Under-dispersion สำหรับฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson เป็นดังนี้

$$f_{CMP}(y; \lambda, \nu) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)}, \quad y = 0, 1, 2, \dots \quad (3.1)$$

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \quad (3.2)$$

โดยที่ $\lambda > 0$, $\nu \geq 0$ และ $Z(\lambda, \nu)$ เป็นค่าคงที่ การแจกแจง Conway-Maxwell-Poisson สามารถทำให้อยู่ในรูปการแจกแจงไม่ต่อเนื่องอื่น ๆ ได้เมื่อพารามิเตอร์ ν มีการแปรผันดังนี้ กรณี $\nu = 1$ ทำให้ $Z(\lambda, \nu) = e^\lambda$ ซึ่งสอดคล้องกับการแจกแจงปัวซองที่มีพารามิเตอร์ λ ส่วนที่ $\nu \rightarrow \infty$ ซึ่ง $Z(\lambda, \nu) \rightarrow 1 + \lambda$ จะลู่เข้าสู่การแจกแจงแบร์นูลลีที่มีพารามิเตอร์ของการแจกแจง $\lambda / (1 + \lambda)$

เช่นเดียวกันกับกรณีที่ $\nu=0$ และ $0 < \lambda < 1$ ทำให้ $Z(\lambda, \nu) = 1/(1-\lambda)$ สอดคล้องกับการแจกแจงเรขาคณิตที่มีพารามิเตอร์ $1-\lambda$

สำหรับการพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson จะนำเสนอตัวประมาณขนาดประชากร 2 ตัวคือ ตัวประมาณขนาดประชากรตัดปลายที่ศูนย์ และตัวประมาณขนาดประชากรตัดปลายที่มากกว่า 3 ดังต่อไปนี้

3.1.1 ตัวประมาณขนาดประชากรตัดปลายที่ศูนย์

ข้อมูลที่เก็บโดยวิธี Capture-Recapture โดยปกติแล้วเป็นข้อมูลที่ได้จากการนับซ้ำ ซึ่งข้อมูลไม่มีศูนย์ กำหนดให้ Y แทนจำนวนครั้งที่หน่วยตัวอย่างถูกสำรวจพบ และให้ f_y แทนจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ y ครั้ง โดยที่ $y = 1, 2, \dots, m$ โดยที่ m เป็นจำนวนครั้งสูงสุดที่หน่วยตัวอย่างถูกสำรวจพบ เนื่องจากไม่ทราบ f_0 ดังนั้น ฟังก์ชันความน่าจะเป็นที่นำมาใช้จะมาจาก การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ ดังนี้

$$\begin{aligned} f(y; \lambda, \nu) &= P(Y = y \mid y = 1, 2, 3, \dots, m) \\ &= \frac{P(Y = y \mid y = 1, 2, 3, \dots, m)}{1 - P(Y = 0)} \\ &= \frac{\frac{\lambda^y}{(y!)^\nu}}{Z(\lambda, \nu) - 1} \end{aligned} \quad (3.3)$$

โดยที่ $y = 1, 2, \dots, m$ และ $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ ใช้วิธีภาวะน่าจะเป็นสูงสุดในการประมาณพารามิเตอร์ของการแจกแจง ซึ่งฟังก์ชันภาวะน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ เป็นดังนี้

$$L(\lambda, \nu) = \prod_{y=1}^m \left[\frac{\frac{\lambda^y}{(y!)^\nu}}{Z(\lambda, \nu) - 1} \right]^{f_y} \quad (3.4)$$

ล็อกฟังก์ชันภาวะน่าจะเป็นที่ไม่ทราบ f_0 (Observed Incomplete Log-likelihood) ของการแจกแจง Conway-Maxwell-Poisson ได้ดังนี้

$$\log L(\lambda, \nu) = \sum_{y=1}^m \log \left(\frac{\lambda^y}{(y!)^\nu}{Z(\lambda, \nu) - 1} \right)^{f_y} = \sum_{y=1}^m f_y \log \left(\frac{\lambda^y}{(y!)^\nu}{Z(\lambda, \nu) - 1} \right) \quad (3.5)$$

พิจารณาค่าประมาณของพารามิเตอร์ λ และ ν ที่ทำให้ค่าของล็อกฟังก์ชันภาวะน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่าศูนย์มีค่าสูงสุดด้วยการหาอนุพันธ์อันดับ 1 ของฟังก์ชันเทียบ λ และ ν จะได้ ตัวประมาณภาวะน่าจะเป็นสูงสุด

$$\hat{\lambda} = \frac{\sum_{y=1}^m f_y}{\sum_{j=1}^{\infty} \frac{j \hat{\lambda}^{j-1}}{(j!)^\nu}} \left(\sum_{j=1}^{\infty} \frac{\hat{\lambda}^j}{(j!)^\nu} - 1 \right)$$

เนื่องจากตัวประมาณค่าพารามิเตอร์ λ และ ν ไม่ได้อยู่ในรูปของสมการปิด (No Closed Form Solution) จึงใช้วิธีการเชิงตัวเลขเพื่อเข้ามาช่วยในการประมาณค่าพารามิเตอร์ λ และ ν โดยเลือกใช้กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) ซึ่งมักถูกนำมาใช้ในงานด้าน Capture-Recapture เนื่องจากว่าล็อกฟังก์ชันภาวะน่าจะเป็นข้างต้น ข้อมูลสามารถสังเกตได้แต่ไม่สมบูรณ์เพราะ f_0 ไม่ทราบค่า เพื่อให้ได้ล็อกฟังก์ชันภาวะน่าจะเป็นที่สมบูรณ์ โดยเริ่มจากกระบวนการหาค่าคาดหวัง (Expectation: E-step)

$$Q(\lambda, \nu) = E(\log L(\lambda, \nu) | f_1, f_2, \dots, f_m; \lambda, \nu)$$

โดยมีค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบและล็อกฟังก์ชันภาวะน่าจะเป็นดังนี้

$$Q(\lambda, \nu) = \{E(f_0 | f_1, f_2, \dots, f_m; \lambda, \nu)\} \log \left(\frac{1}{Z(\lambda, \nu)} \right) + \sum_{y=1}^m f_y \log \left(\frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)} \right)$$

นั่นคือ

$$\log L(\lambda, \nu) = f_0 \log\left(\frac{1}{Z(\lambda, \nu)}\right) + \sum_{y=1}^m f_y \log\left(\frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)}\right) \quad (3.6)$$

พิจารณาค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ f_0 จะได้ว่า

$$\begin{aligned} \hat{f}_0 &= E(f_0 | f_1, f_2, \dots, f_m; \lambda, \nu) \\ &= Np_0 \\ &= (n + f_0) \frac{1}{Z(\lambda, \nu)} \\ &= \frac{n}{Z(\lambda, \nu) - 1} \end{aligned} \quad (3.7)$$

ในกระบวนการหาค่าสูงสุด (Maximization: M-step) จากล็อกฟังก์ชันสถานะน่าจะเป็นในสมการ (3.6) แทน f_0 ด้วย \hat{f}_0 จะได้ล็อกฟังก์ชันสถานะน่าจะเป็นที่สมบูรณ์ (Unobserved Complete Log-likelihood) ของการแจกแจง Conway-Maxwell-Poisson ดังนี้

$$\log L(\lambda, \nu) = \hat{f}_0 \log\left(\frac{1}{Z(\lambda, \nu)}\right) + \sum_{y=1}^m f_y \log\left(\frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)}\right) \quad (3.8)$$

ซึ่ง $Z(\lambda, \nu)$ เป็นค่าคงที่และสามารถประมาณโดยการตัดปลายที่มากกว่า k (Minka, 2003) ได้ดังนี้

$$Z(\lambda, \nu) = \sum_{j=1}^k \frac{\lambda^j}{(j!)^\nu} + R_k$$

โดยที่ $R_k = \sum_{j=k+1}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ เป็นความคลาดเคลื่อนจากการตัดปลาย โดยเลือกตัดปลายที่มากกว่า 3 ใช้

โปรแกรม Rstudio ในการคำนวณหาค่า $Z(\lambda, \nu)$ และจากล็อกฟังก์ชันความควรจะเป็นที่สมบูรณ์ในสมการ (3.8) ในการประมาณค่าพารามิเตอร์ λ และ ν เลือกใช้ฟังก์ชัน nlminb (Sellers, 2011) ในโปรแกรม Rstudio

หลังจากนั้นใช้กระบวนการเชิงตัวเลขในการหาค่า $\hat{\lambda}$, $\hat{\nu}$ และ \hat{f}_0 ที่ทำให้ฟังก์ชันความควรจะเป็นมีค่าสูงสุด ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 0: กำหนดค่า $\hat{\lambda}^{(0)}$ และ $\hat{\nu}^{(0)}$ รอบที่ 0 และกำหนดค่าเริ่มต้น $r=0$

$$\text{ขั้นตอนที่ 1: ให้ } \hat{f}_0^{(r+1)} = \frac{n}{Z(\hat{\lambda}^{(r)}, \hat{\nu}^{(r)}) - 1}$$

ขั้นตอนที่ 2: แทนค่า $\hat{f}_0^{(r+1)}$ ที่ได้จากขั้นตอนที่ 1 ลงใน $\hat{\lambda}^{(r+1)}$ และ $\hat{\nu}^{(r+1)}$ และให้ $r = r + 1$ แล้วกลับไปขั้นตอนที่ 1 ทำการวนซ้ำๆจนกระทั่งค่าประมาณ $\hat{\lambda}$, $\hat{\nu}$ และ \hat{f}_0 มีการลู่เข้า ดังนั้นตัวประมาณค่า f_0 จะได้ว่า

$$\hat{f}_0 = \frac{n}{Z(\hat{\lambda}, \hat{\nu}) - 1} \quad (3.9)$$

ดังนั้นได้ตัวประมาณขนาดประชากร $\hat{N} = n + \hat{f}_0$ ดังต่อไปนี้

$$\hat{N}_{ZCMP} = n + \frac{n}{Z(\hat{\lambda}, \hat{\nu}) - 1}$$

เรียกตัวประมาณขนาดประชากรที่พัฒนานี้ว่า ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์

3.1.2 ตัวประมาณขนาดประชากรตัดปลายที่มากกว่า 3

เมื่อพิจารณาตัวประมาณขนาดประชากรของ Chao ซึ่งเป็นตัวประมาณที่ถูกเลือกใช้บ่อยและพบว่าหน่วยตัวอย่างถูกใช้เพียงกรณีที่สำรวจพบ 1 และ 2 ครั้งเท่านั้นในการประมาณส่วนที่ยังไม่ทราบค่า ซึ่งสะดวกสำหรับผู้นำไปใช้ต่อไป ในงานวิจัยด้าน Capture-Recapture ที่ผ่านๆมา จึงเลือกตัดปลายที่ค่าศูนย์และมากกว่า 2 ซึ่งข้อมูลแฉงนับใน Capture-Recapture ปกติแล้วจะไม่มีศูนย์ครั้งและข้อมูลส่วนใหญ่ถูกพบสำรวจพบ 1, 2 และ 3 ครั้ง แต่จำนวนครั้งที่ข้อมูลถูกสำรวจพบ 4, 5 และ 6 ครั้งนั้นเป็นไปได้ยาก Lanumteang-Böhning ได้เสนอว่าถ้าทำการเพิ่มจำนวนครั้งบนความเป็นไปได้แล้วตัวประมาณขนาดประชากรที่ได้สามารถประมาณขนาดประชากรได้ดีขึ้นหรือไม่ จึงได้พัฒนาตัวประมาณประชากรโดยเพิ่มหน่วยตัวอย่างที่ถูกสำรวจพบเป็นจำนวน 3 ครั้ง ในที่นี้เราสมมติว่าการประมาณจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ f_0 สามารถประมาณได้ด้วยจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 1, 2 และ 3 ครั้ง เมื่อพิจารณาตัวประมาณขนาดประชากรโดยประมาณจาก

หน่วยตัวอย่างที่ถูกสำรวจพบหรือถูกนับซ้ำจำนวน 1, 2 และ 3 ครั้ง จากฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3 ดังต่อไปนี้

$$\begin{aligned}
 f_{Y=1,2,3}(y; \lambda, \nu) &= P(Y = y | y = 1, 2, 3) \\
 &= \frac{P(Y = y | y = 1, 2, 3)}{P(Y = 1) + P(Y = 2) + P(Y = 3)} \\
 &= \frac{\frac{\lambda^y}{(y!)^\nu}}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}} \quad (3.10)
 \end{aligned}$$

โดยที่ $y = 1, 2, 3$ ดังนั้นฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson เมื่อ $y = 1, 2, 3$ ได้ดังต่อไปนี้

$$L(\lambda, \nu) = \prod_{y=1}^3 \left[\frac{\frac{\lambda^y}{(y!)^\nu}}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}} \right]^{f_y} \quad (3.11)$$

ล็อกฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson เมื่อ $y = 1, 2, 3$ ได้ดังนี้

$$\log L(\lambda, \nu) = \sum_{y=1}^3 \log \left(\frac{\frac{\lambda^y}{(y!)^\nu}}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}} \right) = \sum_{y=1}^3 f_y \log \left(\frac{\frac{\lambda^y}{(y!)^\nu}}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}} \right) \quad (3.12)$$

พิจารณาค่าพารามิเตอร์ λ และ ν ที่ทำให้ค่าของล็อกฟังก์ชันความน่าจะเป็นของการแจกแจง Conway-Maxwell-Poisson เมื่อทราบว่า $y = 1, 2, 3$ มีค่าสูงสุด ด้วยการหาอนุพันธ์อันดับ 1 ของฟังก์ชันเทียบ λ และ ν จะได้ ตัวประมาณความน่าจะเป็นสูงสุด

$$\hat{\lambda} = \frac{\sum_{y=1}^3 y f_y}{\left(1 + \frac{2\hat{\lambda}}{(2!)^\nu} + \frac{3\hat{\lambda}^2}{(3!)^\nu}\right)} \left(\hat{\lambda} + \frac{\hat{\lambda}^2}{(2!)^\nu} + \frac{\hat{\lambda}^3}{(3!)^\nu} \right)$$

แต่เนื่องจากตัวประมาณค่าพารามิเตอร์ไม่ได้อยู่ในรูปสมการปิด (No Closed Form Solution) จึงใช้
 เลือกใช้กระบวนการเชิงตัวเลขช่วยในการแก้ปัญหาการประมาณค่าพารามิเตอร์ λ และ ν จะเห็น
 ว่าล็อกฟังก์ชันภาวะน่าจะเป็นข้างต้นข้อมูลสามารถสังเกตหรือถูกสำรวจพบได้แต่ไม่สมบูรณ์ เพราะไม่
 ทราบค่า f_0 แต่เราต้องการล็อกฟังก์ชันภาวะน่าจะเป็นที่สมบูรณ์ จึงพิจารณาฟังก์ชันภาวะน่าจะเป็น
 ที่สมบูรณ์ โดยมีค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบและล็อกฟังก์ชันภาวะน่าจะเป็น
 เป็นที่ไม่ทราบ f_0 (Observed Incomplete Log-likelihood) ดังต่อไปนี้

$$\log L(\lambda, \nu) = f_0 \log\left(\frac{1}{Z(\lambda, \nu)}\right) + \sum_{y=1}^3 f_y \log\left(\frac{\frac{\lambda^y}{(y!)^\nu}}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}}\right) \quad (3.13)$$

เมื่อพิจารณาค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ f_0 ดังนี้

$$\hat{f}_0 = E(f_0 | f_1, f_2, \dots, f_m; \lambda, \nu) = Np_0$$

กำหนดให้ e_y แทนค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ y ครั้ง โดยที่
 $y = 1, 2, 3, \dots, m$ จะได้ว่า

$$e_y = Np_y = (e_0 + f_1 + f_2 + f_3 + e_4^+) p_y \quad (3.14)$$

เมื่อ $e_4^+ = \sum_{y=4}^{\infty} e_y$ จากสมการ (3.14) จะได้ e_0 และ e_4^+ ดังนี้

$$e_0 = (e_0 + f_1 + f_2 + f_3 + e_4^+) p_0 \quad (3.15)$$

$$e_4^+ = (e_0 + f_1 + f_2 + f_3 + e_4^+)p_4^+ \quad (3.16)$$

จากสมการ (3.15) และ (3.16) จะได้ $e_0 + e_4^+$ ดังนี้

$$\begin{aligned} e_0 + e_4^+ &= (e_0 + f_1 + f_2 + f_3 + e_4^+)(p_0 + p_4^+) \\ &= (e_0 + e_4^+ + f_1 + f_2 + f_3)(1 - p_1 - p_2 - p_3) \\ &= (e_0 + e_4^+)(1 - p_1 - p_2 - p_3) + (f_1 + f_2 + f_3)(1 - p_1 - p_2 - p_3) \\ &= \frac{(f_1 + f_2 + f_3)(1 - p_1 - p_2 - p_3)}{p_1 + p_2 + p_3} \end{aligned} \quad (3.17)$$

แทนค่า $e_0 + e_4^+$ จากสมการ (3.17) ลงในสมการ (3.15) จะได้ดังต่อไปนี้

$$\begin{aligned} e_0 &= (e_0 + e_4^+ + f_1 + f_2 + f_3)p_0 \\ &= \frac{(f_1 + f_2 + f_3)(1 - p_1 - p_2 - p_3)}{p_1 + p_2 + p_3} p_0 + (f_1 + f_2 + f_3)p_0 \\ &= \frac{(f_1 + f_2 + f_3)(1 - p_1 - p_2 - p_3)}{p_1 + p_2 + p_3} p_0 + \frac{(f_1 + f_2 + f_3)(p_1 + p_2 + p_3)}{p_1 + p_2 + p_3} p_0 \\ &= \frac{(f_1 + f_2 + f_3)p_0}{p_1 + p_2 + p_3} \end{aligned} \quad (3.18)$$

แทนค่า $p_0 = \frac{1}{Z(\lambda, \nu)}$, $p_1 = \frac{\lambda}{Z(\lambda, \nu)}$, $p_2 = \frac{\lambda^2}{(2!)^\nu} \frac{1}{Z(\lambda, \nu)}$ และ $p_3 = \frac{\lambda^3}{(3!)^\nu} \frac{1}{Z(\lambda, \nu)}$ ลงในสมการ (3.18) จะได้ว่า

$$\begin{aligned} e_0 &= \frac{(f_1 + f_2 + f_3) \frac{1}{Z(\lambda, \nu)}}{\frac{1}{Z(\lambda, \nu)} \left(\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu} \right)} \\ &= \frac{f_1 + f_2 + f_3}{\lambda + \frac{\lambda^2}{(2!)^\nu} + \frac{\lambda^3}{(3!)^\nu}} \end{aligned} \quad (3.19)$$

ดังนั้น ค่าคาดหวังของจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ คือ

$$\hat{f}_0 = \frac{f_1 + f_2 + f_3}{\lambda + \frac{\lambda^2}{(2!)^v} + \frac{\lambda^3}{(3!)^v}} \quad (3.20)$$

ในกระบวนการหาค่าสูงสุด (Maximization: M-step) จากล็อกฟังก์ชันภาวะน่าจะเป็นในสมการ (3.13) แทน f_0 ด้วย \hat{f}_0 จะได้ล็อกฟังก์ชันภาวะน่าจะเป็นที่สมบูรณ์ (Unobserved Complete Log-likelihood) ของการแจกแจง Conway-Maxwell-Poisson ดังนี้

$$\log L(\lambda, \nu) = \hat{f}_0 \log\left(\frac{1}{Z(\lambda, \nu)}\right) + \sum_{y=1}^3 f_y \log\left(\frac{\frac{\lambda^y}{(y!)^v}}{\lambda + \frac{\lambda^2}{(2!)^v} + \frac{\lambda^3}{(3!)^v}}\right) \quad (3.21)$$

ใช้โปรแกรม Rstudio ในการประมาณค่า $Z(\lambda, \nu)$ โดยการตัดปลายที่มากกว่า 3 และประมาณค่าพารามิเตอร์ λ และ ν จากล็อกฟังก์ชันภาวะน่าจะเป็นที่สมบูรณ์ในสมการ (3.21) โดยในการประมาณพารามิเตอร์ใช้ฟังก์ชัน nlminb (Sellers, 2011)

จากนั้นใช้กระบวนการเชิงตัวเลขในการหาค่า $\hat{\lambda}, \hat{\nu}$ และ \hat{f}_0 ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 0: กำหนดค่า $\hat{\lambda}^{(0)}$ และ $\hat{\nu}^{(0)}$ รอบที่ 0 และกำหนดค่าเริ่มต้น $r = 0$

ขั้นตอนที่ 1: ให้ $\hat{f}_0^{(r+1)} = \frac{f_1 + f_2 + f_3}{\hat{\lambda}^{(r)} + \frac{\hat{\lambda}^{2(r)}}{(2!)^{\hat{\nu}^{(r)}}} + \frac{\hat{\lambda}^{3(r)}}{(3!)^{\hat{\nu}^{(r)}}}}$

ขั้นตอนที่ 2: แทนค่า $\hat{f}_0^{(r+1)}$ ที่ได้จากขั้นตอนที่ 1 ลงใน $\hat{\lambda}^{(r+1)}$ และ $\hat{\nu}^{(r+1)}$ และให้ $r = r + 1$ แล้วกลับไปขั้นตอนที่ 1 ทำการวนซ้ำๆ จนกระทั่งค่าประมาณ $\hat{\lambda}, \hat{\nu}$ และ \hat{f}_0 มีการลู่เข้า นำตัวประมาณพารามิเตอร์มาใช้ในสมการ (3.20) จะได้ตัวประมาณค่า f_0 ดังนั้น ตัวประมาณขนาดประชากร $\hat{N} = n + \hat{f}_0$ ดังต่อไปนี้

$$\hat{N}_{YCMP} = n + \left(\frac{f_1 + f_2 + f_3}{\hat{\lambda} + \frac{\hat{\lambda}^2}{(2!)^\nu} + \frac{\hat{\lambda}^3}{(3!)^\nu}} \right)$$

เรียกตัวประมาณขนาดประชากรที่พัฒนานี้ว่า ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3

3.2 การจำลองสถานการณ์

เพื่อศึกษาประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson จึงทำการจำลองสถานการณ์โดยเปรียบเทียบกับตัวประมาณขนาดประชากรของ Chao, ตัวประมาณของ Lanumteang-Böhning, ตัวประมาณภาวะน่าจะเป็นสูงสุด (MLE) และตัวประมาณที่พัฒนาโดยสมการถดถอยเชิงเส้นภายใต้การแจกแจง Conway-Maxwell-Poisson (LCMP) มีขั้นตอนดังต่อไปนี้

3.2.1 กำหนดสถานการณ์จำลอง

3.2.1.1 กำหนดขนาดประชากร N ให้มีขนาดเท่ากับ 500, 1,000 และ 10,000 ตามลำดับ

3.2.1.2 กำหนดพารามิเตอร์

(1) จำลองประชากรจากการแจกแจง Conway-Maxwell-Poisson ด้วยฟังก์ชัน `rcom(N,lambda,nu)` ในโปรแกรม Rstudio กำหนดพารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ และ $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$ สำหรับการจำลองประชากรขนาด 500, 1,000 และ 10,000 ตามลำดับ รวมทั้งสิ้น 45 สถานการณ์

(2) จำลองประชากรจากการแจกแจงปัวซอง ด้วยฟังก์ชัน `rpois(N,lambda)` ในโปรแกรม Rstudio กำหนดพารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ สำหรับการจำลองประชากรขนาด 500, 1,000 และ 10,000 ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์

(3) จำลองประชากรจากการแจกแจงเรขาคณิต ด้วยฟังก์ชัน $\text{rgeom}(N,p)$ ในโปรแกรม Rstudio กำหนดพารามิเตอร์ $p = 0.3, 0.5, 0.7$ สำหรับการจำลองประชากรขนาด 500, 1,000 และ 10,000 ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์

(4) จำลองประชากรจากการแจกแจงทวินามเชิงลบ ด้วยฟังก์ชัน $\text{rbinom}(N,k,\theta)$ ในโปรแกรม Rstudio กำหนดพารามิเตอร์ $(k, \theta) = (1,0.3), (3,0.5), (5,0.7)$ สำหรับการจำลองประชากรขนาด 500, 1,000 และ 10,000 ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์

3.2.2 ขั้นตอนการดำเนินแผนการจำลอง

3.2.2.1 จำลองประชากรตามขนาดและการแจกแจงต่างๆ ตามสถานการณ์ที่กำหนด
ในข้อ 3.2.1

3.2.2.2 คำนวณค่า $f_1, f_2, \dots, f_m, S = \sum_{y=1}^m yf_y$ และ $n = \sum_{y=1}^m f_y$ จากประชากร
จำลองในข้อ 3.2.1.1 เมื่อ m คือจำนวนครั้งสูงสุด

3.2.2.3 คำนวณค่าประมาณขนาดประชากรด้วยตัวประมาณ ต่อไปนี้

(1) ตัวประมาณขนาดประชากรภาวะน่าจะเป็นสูงสุด

$$\hat{N}_{MLE} = \frac{n}{1 - e^{-\hat{\lambda}_{MLE}}}$$

(2) ตัวประมาณขนาดประชากรของ Chao

$$\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}$$

(3) ตัวประมาณขนาดประชากรของ Lanumteang-Böhning

$$\hat{N}_{LB} = n + \frac{3f_1^3 f_3}{4f_2^3}$$

(4) ตัวประมาณขนาดประชากร LCMP

$$\hat{N}_{LCMP} = n + f_1 e^{-\hat{\beta}_0}$$

(5) ตัวประมาณขนาดประชากร ZCMP

$$\hat{N}_{ZCMP} = n + \frac{n}{Z(\hat{\lambda}, \hat{\nu}) - 1}$$

(6) ตัวประมาณขนาดประชากร YCMP

$$\hat{N}_{YCMP} = n + \left(\frac{f_1 + f_2 + f_3}{\hat{\lambda} + \frac{\hat{\lambda}^2}{(2!)^{\hat{\nu}}} + \frac{\hat{\lambda}^3}{(3!)^{\hat{\nu}}} } \right)$$

3.2.2.4 สำหรับตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson และตัวประมาณขนาดประชากรน่าจะจะเป็นสูงสุด ประมาณพารามิเตอร์โดยใช้กระบวนการหาค่าคาดหวังและการหาค่าสูงสุด (EM-algorithm) จนกระทั่งค่าประมาณพารามิเตอร์ในรอบสุดท้ายแตกต่างจากรอบก่อนหน้าทศนิยม 4 ตำแหน่ง

3.2.2.5 ทำซ้ำจากข้อ 3.2.2.1 ถึง 3.2.2.3 จำนวน $B = 2,000$ รอบ

3.2.2.6 คำนวณค่าค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

(1) ค่าความเอนเอียงสัมพัทธ์

$$RBias = \frac{\bar{N} - N}{N}$$

โดยที่ $\bar{\hat{N}} = \frac{1}{B} \sum_{l=1}^B \hat{N}_l$ เมื่อ B แทนจำนวนรอบของการวนซ้ำๆ ในการจำลอง

(2) ค่าความแปรปรวนสัมพัทธ์

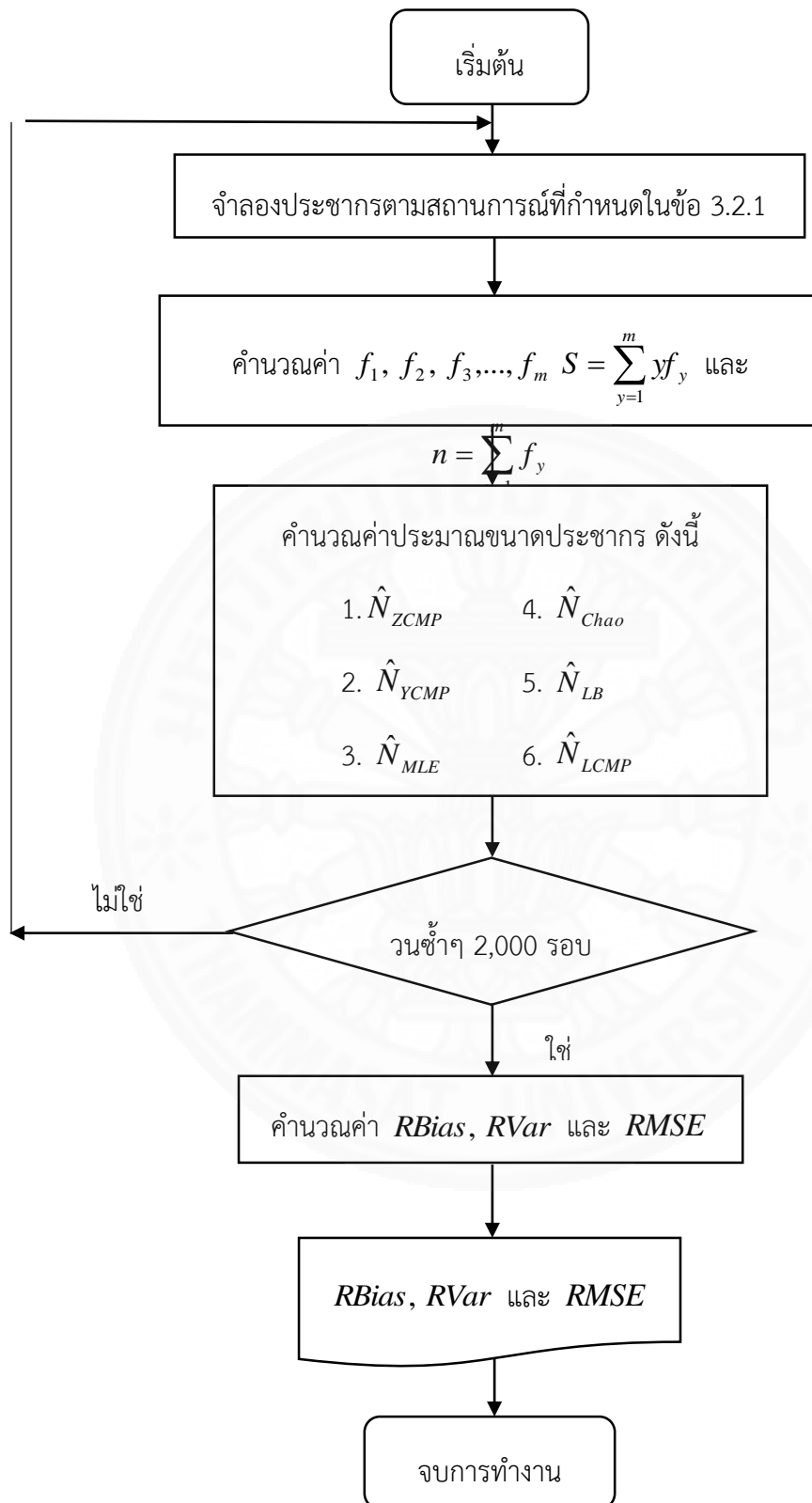
$$RVar = \frac{1}{N^2} \sum_{l=1}^B \frac{(\hat{N}_l - \bar{\hat{N}})^2}{B}$$

(3) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

$$RMSE = \frac{1}{N^2} \sum_{l=1}^B \frac{(\hat{N}_l - N)^2}{B}$$

3.2.2.7 จำลองซ้ำจากข้อ 3.2.2.1 ถึง 3.2.2.6 จนกระทั่งครบ 72 สถานการณ์

ใช้โปรแกรม Rstudio ในการจำลองสถานการณ์และการคำนวณ โดยทำวนซ้ำๆ จำนวน $B = 2,000$ รอบ สามารถสรุปการทำงานของโปรแกรมได้ตามแผนผังในภาพที่ 3.1



ภาพที่ 3.1 แผนผังแสดงการทำงานของโปรแกรมจำลองตัวแบบ

บทที่ 4

ผลการวิจัย

จากบทที่ 3 ได้อธิบายเกี่ยวกับการหาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์และตัดปลายที่มากกว่า 3 และขั้นตอนการจำลองสถานการณ์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาขึ้นมา กับตัวประมาณขนาดประชากรของเดิมที่มีอยู่แล้ว สำหรับบทที่ 4 จะแสดงผลที่ได้จากการจำลองสถานการณ์ พร้อมทั้งอธิบายผลที่ได้ นอกจากนี้ยังนำเสนอผลของตัวประมาณขนาดประชากรจากการประยุกต์ใช้กับข้อมูลจริงอีกด้วย

4.1 ผลลัพธ์จากการจำลองสถานการณ์

4.1.1 การจำลองภายใต้การแจกแจง Conway-Maxwell-Poisson

ตัวประมาณขนาดประชากรทั้ง 2 ตัว พัฒนาขึ้นมาภายใต้การแจกแจง Conway-Maxwell-Poisson เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร จึงจำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson ซึ่งเหมาะสมต่อการใช้ตัวประมาณขนาดประชากรที่พัฒนาขึ้นมาทั้ง 2 ตัว โดยกำหนดสถานการณ์ที่ให้พารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ และ $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$ สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ตามลำดับ รวมทั้งสิ้น 45 สถานการณ์ แสดงผลลัพธ์ไว้ดังตารางที่ 4.1-4.15

4.1.1.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.1-4.15 จะเห็นว่า สำหรับขนาดประชากร $N = 500$ นั้น \hat{N}_{Chao} จะมีความแม่นยำที่สุดหรือค่า $RBias$ เข้าใกล้ศูนย์มากที่สุด และสำหรับ $N = 1,000$ และ $N = 10,000$ จะเห็นว่า \hat{N}_{LB} จะมีความแม่นยำที่สุด ในกรณีที่ประชากรมีขนาดใหญ่ $\lambda = 2$ และ ν มีค่าเพิ่มขึ้น จะทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีความแม่นยำที่สุด ส่วนกรณีที่ขนาดประชากร $N = 500$ ทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะมีความแม่นยำสูงกว่า \hat{N}_{LB} , \hat{N}_{MLE} และ \hat{N}_{LCMP} นอกจากนี้กรณีขนาดประชากร $N = 500, 1,000$ และ $10,000$ เมื่อ ν มีค่าเพิ่มขึ้นทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีความแม่นยำที่สุดและยังทำให้ \hat{N}_{Chao} , \hat{N}_{LB} , \hat{N}_{MLE} , \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ได้ค่าประมาณขนาดประชากรสูงกว่าค่าที่แท้จริง อย่างไรก็ตาม \hat{N}_{LCMP} จะให้ค่าประมาณขนาดประชากรที่ต่ำกว่าค่าจริง สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ส่วนในกรณีที่ λ มีค่าเพิ่มขึ้น จะเห็นว่า

ตัวประมาณขนาดประชากรทั้ง 6 ตัว จะมีความแม่นยำเพิ่มขึ้นในขนาดประชากร $N = 500, 1,000$ และ $10,000$

4.1.1.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตารางที่ 4.1-4.15 จะเห็นว่า สำหรับขนาดประชากร $N = 500$ และ λ มีค่าน้อยๆ \hat{N}_{LB} มีค่า $RVar$ ค่อนข้างสูงมากเมื่อเปรียบเทียบกับตัวประมาณขนาดประชากรอีก 5 ตัว เมื่อพิจารณาขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{MLE} จะให้ค่า $RVar$ ต่ำที่สุด ในกรณีที่ขนาดประชากร $N = 500$ จะทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีค่า $RVar$ ต่ำที่สุด แต่จะเห็นได้ว่าเมื่อเปรียบเทียบ $RVar$ ของ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} เมื่อ λ มีค่าเพิ่มขึ้นจะทำให้ $RVar$ ของ \hat{N}_{YCMP} มีค่าที่ต่ำกว่า $RVar$ ของ \hat{N}_{ZCMP}

4.1.1.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตารางที่ 4.1-4.15 จะเห็นว่า สำหรับ $N = 1,000$ และ $10,000$ \hat{N}_{Chao} ให้ค่า $RMSE$ ต่ำที่สุด ในกรณีที่ประชากรมีขนาดใหญ่และ ν มีค่าน้อยๆ จะทำให้ \hat{N}_{LB} ให้ค่า $RMSE$ ต่ำกว่า $\hat{N}_{Chao}, \hat{N}_{MLE}, \hat{N}_{LCMP}, \hat{N}_{YCMP}$ และ \hat{N}_{ZCMP} ตามลำดับ และจะเห็นว่า ตัวประมาณขนาดประชากรทั้ง 6 ตัว ให้ค่าประมาณขนาดประชากรที่มีความแม่นยำมากขึ้นเมื่อ λ มีค่าเพิ่มขึ้น

4.1.2 การจำลองภายใต้การแจกแจงปัวซอง

เมื่อมีการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรโดยจำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson แล้ว เพื่อให้สถานการณ์จำลองนั้น ใกล้เคียงตัวประมาณประชากรทั้ง 2 ตัวที่พัฒนาขึ้นมา จึงทำการจำลองสถานการณ์ภายใต้การแจกแจง อื่นๆ ในที่นี้ได้แก่ การแจกแจงปัวซอง โดยกำหนดสถานการณ์ที่ให้พารามิเตอร์ $\lambda = 0.5, 1.0, 2.0$ สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์ แสดงผลลัพธ์ไว้ดังตารางที่ 4.16-4.18

4.1.2.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.16-4.18 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{MLE} มีความแม่นยำสูงสุดที่สุดหรือค่า $RBias$ เข้าใกล้ศูนย์มากที่สุด โดยจะต่ำกว่า $\hat{N}_{Chao}, \hat{N}_{LB}, \hat{N}_{YCMP}, \hat{N}_{ZCMP}$ และ \hat{N}_{LCMP} ตามลำดับ และเมื่อ λ มีค่าเพิ่มขึ้น \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่าประมาณขนาดประชากรสูงกว่าค่าที่แท้จริง ในขณะที่ \hat{N}_{LCMP} ให้ค่าประมาณขนาดประชากรต่ำกว่าค่าที่แท้จริง สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ นอกจากนี้ เมื่อ λ มีค่าเพิ่มขึ้นจะเห็นว่า ตัวประมาณขนาดประชากรทั้ง 6 ตัวจะมีความแม่นยำเพิ่มขึ้นเมื่อ λ มีค่าสูงขึ้น

4.1.2.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตารางที่ 4.16-4.18 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{MLE} จะให้ค่า $RVar$ ต่ำที่สุด ในขณะที่ \hat{N}_{LB} ให้ค่า $RVar$ ค่อนข้างสูงมาก เมื่อเปรียบเทียบกับตัวประมาณขนาดประชากรตัวอื่น และเมื่อ λ มีค่าเพิ่มขึ้นจะทำให้ \hat{N}_{MLE} และ \hat{N}_{LCMP} ให้ค่า $RVar$ ที่ใกล้เคียงกันมาก ส่วน \hat{N}_{YCMP} จะให้ค่า $RVar$ ที่ต่ำกว่า $RVar$ ของ \hat{N}_{ZCMP}

4.1.2.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตารางที่ 4.16-4.18 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{MLE} ให้ค่า $RMSE$ ต่ำที่สุด นอกจากนี้ $RMSE$ จะลดต่ำลงเมื่อประชากรมีขนาดใหญ่ขึ้น และ λ มีค่าเพิ่มขึ้น

4.1.3 การจำลองภายใต้การแจกแจงเรขาคณิต

นอกจากการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรโดยจำลองสถานการณ์ภายใต้การแจกแจงปัวซองแล้วยังจำลองสถานการณ์ภายใต้การแจกแจงเรขาคณิตโดยกำหนดสถานการณ์ที่ให้พารามิเตอร์ $p = 0.3, 0.5, 0.7$ สำหรับการจำลองประชากรขนาด $500, 1,000$ และ $10,000$ ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์ แสดงผลลัพธ์ไว้ดังตารางที่ 4.19-4.21

4.1.3.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.19-4.21 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{LB} มีความแม่นยำสูงสุดที่สุดหรือค่า $RBias$ เข้าใกล้ศูนย์มากที่สุด ในส่วนของ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะมีความแม่นยำสูงสุดที่สุด นอกจากนี้กรณีที่ p มีค่าเพิ่มขึ้นจะทำให้ \hat{N}_{Chao} , \hat{N}_{MLE} , \hat{N}_{LCMP} , \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ยกเว้น \hat{N}_{LB} ให้ค่าประมาณขนาดประชากรลดลงและตัวประมาณขนาดประชากรทั้ง 6 ตัวให้ค่าประมาณขนาดประชากรที่ต่ำกว่าค่าที่แท้จริง

4.1.3.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตารางที่ 4.19-4.21 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{MLE} มีค่า $RVar$ ต่ำที่สุด ในส่วนของ \hat{N}_{LB} จะให้ค่า $RVar$ สูงที่สุด สำหรับขนาดประชากร $N = 500, 1,000$ จะเห็นว่า \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่า $RVar$ ต่ำกว่าของ \hat{N}_{Chao} , \hat{N}_{LCMP} และ \hat{N}_{LB} และเมื่อเปรียบเทียบระหว่าง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะเห็นว่า \hat{N}_{YCMP} ให้ค่า $RVar$ ต่ำกว่า $RVar$ ของ \hat{N}_{ZCMP}

4.1.3.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตารางที่ 4.19-4.21 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ ทำให้ \hat{N}_{LB} มีค่า $RMSE$ ต่ำที่สุด ในขณะที่ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่า $RMSE$ สูงที่สุด เมื่อเปรียบเทียบระหว่าง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะเห็นว่า \hat{N}_{ZCMP} จะให้ $RMSE$ ต่ำกว่า $RMSE$ ของ \hat{N}_{YCMP} นอกจากนี้ $RMSE$ ของตัวประมาณขนาดประชากรทั้ง 6 ตัวจะลดต่ำลงเมื่อ p มีค่าลดลง

4.1.4 การจำลองภายใต้การแจกแจงทวินามเชิงลบ

นอกจากการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรโดยจำลองสถานการณ์ภายใต้การแจกแจงปัวซองและการแจกเรขาคณิตแล้ว ยังจำลองสถานการณ์ภายใต้การแจกแจงทวินามเชิงลบ โดยกำหนดสถานการณ์ที่ให้พารามิเตอร์ $k = 1, 3, 5$ และ $\theta = 0.3, 0.5, 0.7$ สำหรับการจำลองประชากรขนาด $500, 1,000$ และ $10,000$ ตามลำดับ รวมทั้งสิ้น 9 สถานการณ์ แสดงผลลัพธ์ไว้ดังตารางที่ 4.22-4.24

4.1.4.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.22-4.24 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ นั้น \hat{N}_{LB} มีความแม่นยำสูงสุดที่สุดหรือค่า $RBias$ เข้าใกล้ศูนย์มากที่สุด ในกรณีที่ k และ θ มีค่าเพิ่มขึ้น จะเห็นว่าตัวประมาณขนาดประชากรทั้ง 6 ตัว จะมีความแม่นยำเพิ่มขึ้น นอกจากนี้ \hat{N}_{LB} ให้ค่าประมาณขนาดประชากรสูงกว่าค่าที่แท้จริงเมื่อ k และ θ มีค่าเพิ่มขึ้น ส่วนตัวประมาณอีก 5 ตัวจะให้ค่าประมาณขนาดประชากรต่ำกว่าค่าที่แท้จริง

4.1.4.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตารางที่ 4.22-4.24 จะเห็นว่า สำหรับขนาดประชากร $N = 500, 1,000$ และ $10,000$ จะทำให้ \hat{N}_{MLE} มีค่า $RVar$ ต่ำที่สุด ส่วน \hat{N}_{LB} มีค่า $RVar$ สูงที่สุด เมื่อประชากรขนาดเล็ก และ k และ θ มีค่าลดลงจะทำให้ \hat{N}_{LCMP} ให้ค่า $RVar$ สูงที่สุด เมื่อเปรียบเทียบระหว่าง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะเห็นว่า \hat{N}_{YCMP} ให้ค่า $RVar$ ที่ต่ำกว่า $RVar$ ของ นอกจากนี้ $RVar$ ของตัวประมาณขนาดประชากรทั้ง 6 ตัวจะลดต่ำลงเมื่อ k และ θ มีค่าเพิ่มขึ้น

4.1.4.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตารางที่ 4.22-4.24 จะเห็นว่า สำหรับประชากรที่มีขนาดใหญ่จะทำให้ \hat{N}_{LB} มีค่า $RMSE$ ต่ำที่สุด ส่วนขนาดประชากร $N = 500, 1,000$ และ k และ θ มีค่าเพิ่มขึ้นจะทำให้ \hat{N}_{Chao} ให้ค่า $RMSE$ ต่ำที่สุดในขณะที่ประชากรขนาดเล็ก \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่า $RMSE$ ที่ต่ำกว่า $RMSE$ ของ \hat{N}_{LB} , \hat{N}_{LCMP} และ \hat{N}_{MLE} ตามลำดับ นอกจากนี้ $RMSE$ ของตัวประมาณขนาดประชากรทั้ง 6 ตัวจะลดต่ำลงเมื่อ k และ θ มีค่าเพิ่มขึ้น

ตารางที่ 4.1 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 0.5$, $\nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 410 | -0.1798 | 0.0079 | 0.0402 | 811 | -0.1888 | 0.0040 | 0.0396 | 8085 | -0.1915 | 0.0004 | 0.0371 |
| LB | 493 | -0.0149 | 0.1089 | 0.1091 | 940 | -0.0598 | 0.0426 | 0.0462 | 9126 | -0.0874 | 0.0034 | 0.0111 |
| MLE | 377 | -0.2457 | 0.0034 | 0.0638 | 749 | -0.2513 | 0.0018 | 0.0649 | 7474 | -0.2526 | 0.0002 | 0.0640 |
| LCMP | 307 | -0.3854 | 0.0184 | 0.1670 | 619 | -0.3809 | 0.0049 | 0.1499 | 5987 | -0.4013 | 0.0001 | 0.1612 |
| ZCMP | 278 | -0.4442 | 0.0050 | 0.2024 | 554 | -0.4463 | 0.0045 | 0.2037 | 5519 | -0.4481 | 0.0044 | 0.2051 |
| YCMP | 278 | -0.4441 | 0.0044 | 0.2015 | 556 | -0.4444 | 0.0039 | 0.2014 | 5522 | -0.4478 | 0.0037 | 0.2042 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.2 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 1, \nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 458 | -0.0849 | 0.0018 | 0.0090 | 912 | -0.0877 | 0.0009 | 0.0086 | 9106 | -0.0894 | 8.44E-05 | 0.0081 |
| LB | 491 | -0.0189 | 0.0122 | 0.0125 | 968 | -0.0318 | 0.0053 | 0.0063 | 9591 | -0.0409 | 0.0005 | 0.0021 |
| MLE | 421 | -0.1578 | 0.0007 | 0.0256 | 841 | -0.1591 | 0.0003 | 0.0257 | 8404 | -0.1596 | 3.16E-05 | 0.0255 |
| LCMP | 418 | -0.1635 | 0.0055 | 0.0322 | 831 | -0.1693 | 0.0007 | 0.0294 | 8175 | -0.1825 | 4.47E-05 | 0.0333 |
| ZCMP | 440 | -0.1192 | 0.0026 | 0.0168 | 879 | -0.1212 | 0.0023 | 0.0170 | 8788 | -0.1212 | 0.0021 | 0.0168 |
| YCMP | 429 | -0.1425 | 0.0019 | 0.0222 | 856 | -0.1444 | 0.0015 | 0.0224 | 8552 | -0.1448 | 0.0013 | 0.0222 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.3 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 2$, $\nu = 0.4$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|-----------------|---------------|-------------|----------------|-----------------|-----------------|--------------|----------------|-----------------|-----------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 497 | -0.0069 | 9.87E-05 | 0.0001 | 993 | -0.0074 | 5.06E-05 | 0.0001 | 9922 | -0.0078 | 4.65E-05 | 6.56E-05 |
| LB | 501 | 0.0027 | 0.0006 | 0.0006 | 999 | -0.0008 | 0.0002 | 0.0002 | 9966 | -0.0034 | 1.57E-05 | 2.72E-05 |
| MLE | 489 | -0.0218 | 4.64E-05 | 0.0005 | 978 | -0.0217 | 2.37E-05 | 0.0005 | 9782 | -0.0218 | 2.25E-05 | 0.0005 |
| LCMP | 490 | -0.0199 | 0.0010 | 0.0014 | 980 | -0.0197 | 2.54E-05 | 0.0004 | 9793 | -0.0207 | 2.27E-05 | 0.0004 |
| ZCMP | 503 | 0.0556 | 6.91E-05 | 0.0001 | 1006 | 0.0057 | 4.17E-05 | 7.46E-05 | 10055 | 0.0055 | 1.76E-05 | 4.81E-05 |
| YCMP | 492 | -0.0155 | 4.93E-05 | 0.0003 | 985 | -0.0153 | 2.56E-05 | 0.0003 | 9846 | -0.0154 | 3.46E-05 | 0.0002 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.4 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์
การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 0.5$, $\nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 434 | -0.1320 | 0.0130 | 0.0304 | 865 | -0.1348 | 0.0056 | 0.0238 | 8596 | -0.1404 | 0.0005 | 0.0202 |
| LB | 510 | 0.0200 | 0.1662 | 0.1666 | 971 | -0.0287 | 0.0557 | 0.0565 | 9415 | -0.0585 | 0.0045 | 0.0079 |
| MLE | 412 | -0.1750 | 0.0065 | 0.0371 | 825 | -0.1753 | 0.0029 | 0.0337 | 8198 | -0.1802 | 0.0003 | 0.0327 |
| LCMP | 391 | -0.4181 | 0.0298 | 0.2047 | 606 | -0.3937 | 0.0068 | 0.1618 | 5890 | -0.4110 | 0.0002 | 0.1691 |
| ZCMP | 267 | -0.4670 | 0.0054 | 0.2234 | 532 | -0.4680 | 0.0048 | 0.2238 | 5313 | -0.4687 | 0.0048 | 0.2245 |
| YCMP | 271 | -0.4585 | 0.0053 | 0.2155 | 541 | -0.4590 | 0.0048 | 0.2154 | 5394 | -0.4606 | 0.0046 | 0.2167 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.5 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์
การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 1, \nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 464 | -0.0710 | 0.0027 | 0.0077 | 927 | -0.0731 | 0.0014 | 0.0067 | 9251 | -0.0749 | 0.0001 | 0.0057 |
| LB | 495 | -0.0100 | 0.0166 | 0.0167 | 977 | -0.0226 | 0.0078 | 0.0083 | 9668 | -0.0332 | 0.0007 | 0.0018 |
| MLE | 439 | -0.1216 | 0.0011 | 0.0159 | 878 | -0.1220 | 0.0006 | 0.0154 | 8771 | -0.1229 | 5.63E-05 | 0.0152 |
| LCMP | 405 | -0.1901 | 0.0045 | 0.0407 | 803 | -0.1968 | 0.0009 | 0.0397 | 7910 | -0.2090 | 6.39E-05 | 0.0437 |
| ZCMP | 427 | -0.1452 | 0.0047 | 0.0258 | 853 | -0.1474 | 0.0044 | 0.0262 | 8500 | -0.1500 | 0.0044 | 0.0269 |
| YCMP | 421 | -0.1585 | 0.0035 | 0.0286 | 842 | -0.1581 | 0.0032 | 0.0282 | 8386 | -0.1614 | 0.0032 | 0.0292 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.6 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 2$, $\nu = 0.6$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|-----------------|---------------|-------------|----------------|-----------------|-----------------|--------------|----------------|-----------------|-----------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 497 | -0.0069 | 9.87E-05 | 0.0001 | 993 | -0.0074 | 5.06E-05 | 0.0001 | 9922 | -0.0078 | 4.65E-05 | 6.56E-05 |
| LB | 501 | 0.0027 | 0.0006 | 0.0006 | 999 | -0.0008 | 0.0002 | 0.0002 | 9966 | -0.0034 | 1.57E-05 | 2.72E-05 |
| MLE | 489 | -0.0218 | 4.64E-05 | 0.0005 | 978 | -0.0217 | 2.37E-05 | 0.0005 | 9782 | -0.0218 | 2.25E-05 | 0.0005 |
| LCMP | 490 | -0.0199 | 0.0010 | 0.0014 | 980 | -0.0197 | 2.54E-05 | 0.0004 | 9793 | -0.0207 | 2.27E-05 | 0.0004 |
| ZCMP | 503 | 0.0056 | 6.91E-05 | 0.0001 | 1006 | 0.0057 | 4.17E-05 | 7.46E-05 | 10055 | 0.0055 | 1.76E-05 | 4.81E-05 |
| YCMP | 492 | -0.0155 | 4.93E-05 | 0.0003 | 985 | -0.0153 | 2.56E-05 | 0.0003 | 9846 | -0.0154 | 3.46E-05 | 0.0002 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.7 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 0.5$, $\nu = 0.8$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 470 | -0.0591 | 0.0168 | 0.0203 | 932 | -0.0682 | 0.0074 | 0.0120 | 9229 | -0.0771 | 0.0007 | 0.0067 |
| LB | 537 | -0.0731 | 0.2347 | 0.2400 | 1010 | 0.0099 | 0.0754 | 0.0755 | 9677 | -0.0323 | 0.0063 | 0.0073 |
| MLE | 459 | -0.0820 | 0.0096 | 0.0163 | 912 | -0.0882 | 0.0045 | 0.0123 | 9055 | 0.0945 | 0.0004 | 0.0094 |
| LCMP | 247 | -0.5056 | 0.0680 | 0.3237 | 581 | -0.4195 | 0.0181 | 0.1940 | 5818 | -0.4182 | 0.0002 | 0.1750 |
| ZCMP | 259 | -0.4817 | 0.0056 | 0.2377 | 515 | -0.4852 | 0.0053 | 0.2406 | 5158 | -0.4842 | 0.0054 | 0.2399 |
| YCMP | 266 | -0.4680 | 0.0059 | 0.2249 | 536 | -0.4639 | 0.0056 | 0.2208 | 5295 | -0.4705 | 0.0054 | 0.2268 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.8 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 1$, $\nu = 0.8$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 480 | -0.0407 | 0.0038 | 0.0055 | 959 | -0.0415 | 0.0019 | 0.0036 | 9560 | -0.0440 | 0.0002 | 0.0021 |
| LB | 502 | 0.0034 | 0.0215 | 0.0216 | 997 | -0.0032 | 0.0105 | 0.0105 | 9819 | -0.0181 | 0.0009 | 0.0013 |
| MLE | 467 | -0.0661 | 0.0018 | 0.0062 | 933 | -0.0674 | 0.0009 | 0.0055 | 9321 | -0.0679 | 8.68E-05 | 0.0047 |
| LCMP | 398 | -0.2041 | 0.0042 | 0.0458 | 788 | -0.2120 | 0.0014 | 0.0464 | 7765 | -0.2235 | 7.94E-05 | 0.0500 |
| ZCMP | 411 | -0.1770 | 0.0063 | 0.0377 | 823 | -0.1770 | 0.0061 | 0.0374 | 8220 | -0.1780 | 0.0062 | 0.0379 |
| YCMP | 410 | -0.1798 | 0.0052 | 0.0375 | 819 | -0.1818 | 0.0049 | 0.0380 | 8107 | -0.1893 | 0.0049 | 0.0407 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.9 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 2$, $\nu = 0.8$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 494 | -0.0124 | 0.0058 | 0.0007 | 987 | -0.0129 | 0.0003 | 0.0005 | 9868 | -0.0132 | 2.84E-05 | 0.0002 |
| LB | 501 | 0.0011 | 0.0022 | 0.0022 | 998 | -0.0022 | 0.0011 | 0.0011 | 9947 | -0.0053 | 9.10E-05 | 0.0001 |
| MLE | 485 | -0.0309 | 0.0026 | 0.0012 | 969 | -0.0308 | 0.0001 | 0.0011 | 9691 | -0.0309 | 1.33E-05 | 0.0010 |
| LCMP | 462 | -0.0755 | 0.0003 | 0.0060 | 923 | -0.0774 | 0.0001 | 0.0061 | 9202 | -0.0798 | 1.21E-05 | 0.0064 |
| ZCMP | 516 | 0.0315 | 0.0018 | 0.0028 | 1032 | 0.0319 | 0.0018 | 0.0028 | 10339 | 0.0339 | 0.0016 | 0.0027 |
| YCMP | 500 | 0.0002 | 0.0010 | 0.0010 | 999 | -0.0012 | 0.0009 | 0.0009 | 9994 | -0.0006 | 0.0007 | 0.0007 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.10 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 0.5$, $\nu = 1.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 649 | 0.2972 | 0.0612 | 0.1496 | 1271 | 0.2713 | 0.0271 | 0.1007 | 12606 | 0.2606 | 0.0026 | 0.0705 |
| LB | 661 | 0.3213 | 1.2106 | 1.3138 | 1163 | 0.1629 | 0.2819 | 0.3085 | 11019 | 0.1019 | 0.0205 | 0.0308 |
| MLE | 664 | 0.3287 | 0.0445 | 0.1526 | 1311 | 0.3110 | 0.0205 | 0.1172 | 13007 | 0.3007 | 0.0020 | 0.0924 |
| LCMP | 41 | -0.9184 | 0.0506 | 0.8942 | 174 | -0.8259 | 0.0813 | 0.7633 | 5589 | -0.4411 | 0.0132 | 0.2078 |
| ZCMP | 243 | -0.5142 | 0.0067 | 0.2712 | 483 | -0.5170 | 0.0063 | 0.2736 | 4769 | -0.5231 | 0.0060 | 0.2797 |
| YCMP | 256 | -0.4881 | 0.0081 | 0.2463 | 512 | -0.4881 | 0.0081 | 0.2463 | 5095 | -0.4905 | 0.0085 | 0.2490 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.11 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 1, \nu = 1.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 591 | 0.1822 | 0.0127 | 0.0459 | 1179 | 0.1791 | 0.0064 | 0.0385 | 11720 | 0.1720 | 0.0006 | 0.0302 |
| LB | 555 | 0.1102 | 0.0763 | 0.0885 | 1092 | 0.0918 | 0.0334 | 0.0419 | 10676 | 0.0676 | 0.0028 | 0.0074 |
| MLE | 618 | 0.2352 | 0.0089 | 0.0642 | 1233 | 0.2327 | 0.0044 | 0.0585 | 12276 | 0.2276 | 0.0004 | 0.0522 |
| LCMP | 339 | -0.3214 | 0.0850 | 0.1883 | 766 | -0.2337 | 0.0169 | 0.0715 | 7710 | -0.2290 | 0.0001 | 0.0526 |
| ZCMP | 377 | -0.2460 | 0.0105 | 0.0710 | 752 | -0.2484 | 0.0102 | 0.0719 | 7506 | -0.2494 | 0.0101 | 0.0723 |
| YCMP | 384 | -0.2311 | 0.0104 | 0.0638 | 772 | -0.2276 | 0.0102 | 0.0620 | 7663 | -0.2337 | 0.0104 | 0.0650 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.12 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 2$, $\nu = 1.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|---------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 541 | 0.0816 | 0.0026 | 0.0093 | 1083 | 0.0830 | 0.0012 | 0.0081 | 10813 | 0.0813 | 0.0001 | 0.0067 |
| LB | 520 | 0.0397 | 0.0083 | 0.0099 | 1037 | 0.0372 | 0.0037 | 0.0051 | 10312 | 0.0312 | 0.0004 | 0.0013 |
| MLE | 570 | 0.1399 | 0.0015 | 0.0211 | 1141 | 0.1414 | 0.0007 | 0.0207 | 11403 | 0.1403 | 7.63E-05 | 0.0198 |
| LCMP | 445 | -0.1103 | 0.0012 | 0.0134 | 888 | -0.1123 | 0.0005 | 0.0131 | 8831 | -0.1169 | 4.74E-05 | 0.0137 |
| ZCMP | 503 | 0.0067 | 0.0091 | 0.0092 | 1008 | 0.0082 | 0.0088 | 0.0089 | 10073 | 0.0073 | 0.0093 | 0.0093 |
| YCMP | 504 | 0.0071 | 0.0073 | 0.0073 | 1008 | 0.0078 | 0.0074 | 0.0074 | 10036 | 0.0036 | 0.0074 | 0.0074 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.13 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 0.5$, $\nu = 2$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 851 | 0.7026 | 0.1474 | 0.6410 | 1679 | 0.6787 | 0.0705 | 0.5311 | 16417 | 0.6417 | 0.0063 | 0.4180 |
| LB | 805 | 0.6094 | 4.3242 | 4.6955 | 1381 | 0.3811 | 0.9973 | 1.1425 | 12259 | 0.2259 | 0.0548 | 0.1059 |
| MLE | 880 | 0.7593 | 0.1166 | 0.6931 | 1743 | 0.7426 | 0.0563 | 0.6078 | 17103 | 0.7103 | 0.0053 | 0.5097 |
| LCMP | 5 | -0.9901 | 0.0079 | 0.9881 | 27 | -0.9726 | 0.0195 | 0.6078 | 3004 | -0.6996 | 0.0889 | 0.5783 |
| ZCMP | 238 | -0.5248 | 0.0077 | 0.2831 | 471 | -0.5289 | 0.0072 | 0.2870 | 4653 | -0.5347 | 0.0067 | 0.2926 |
| YCMP | 253 | -0.4947 | 0.0095 | 0.2543 | 502 | -0.4984 | 0.0092 | 0.2576 | 4974 | -0.5026 | 0.0096 | 0.2622 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.14 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 1, \nu = 2$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 729 | 0.4589 | 0.0307 | 0.2413 | 1451 | 0.4509 | 0.0165 | 0.2198 | 14397 | 0.4397 | 0.0015 | 0.1948 |
| LB | 616 | 0.2324 | 0.2233 | 0.2773 | 1194 | 0.1944 | 0.0915 | 0.1293 | 11491 | 0.1491 | 0.0067 | 0.0289 |
| MLE | 779 | 0.5584 | 0.0228 | 0.3346 | 1552 | 0.5517 | 0.0125 | 0.3169 | 15440 | 0.5440 | 0.0011 | 0.2971 |
| LCMP | 132 | -0.7359 | 0.1554 | 0.6970 | 426 | -0.5744 | 0.1635 | 0.4934 | 7889 | -0.2111 | 0.0008 | 0.0454 |
| ZCMP | 358 | -0.2806 | 0.0119 | 0.0906 | 719 | -0.2809 | 0.0110 | 0.0899 | 7120 | -0.2880 | 0.0115 | 0.0944 |
| YCMP | 380 | -0.2402 | 0.0141 | 0.0718 | 761 | -0.2390 | 0.0139 | 0.0711 | 7525 | -0.2475 | 0.0147 | 0.0759 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.15 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $CMP(\lambda, \nu)$ เมื่อกำหนด $\lambda = 2, \nu = 2$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 620 | 0.2403 | 0.0060 | 0.0637 | 1240 | 0.2397 | 0.0029 | 0.0604 | 12358 | 0.2358 | 0.0003 | 0.0559 |
| LB | 545 | 0.0970 | 0.0188 | 0.0283 | 1091 | 0.0913 | 0.0085 | 0.0168 | 10802 | 0.0802 | 0.0008 | 0.0073 |
| MLE | 679 | 0.3583 | 0.0041 | 0.1325 | 1358 | 0.3579 | 0.0022 | 0.1303 | 13552 | 0.3552 | 0.0002 | 0.1264 |
| LCMP | 443 | -0.1139 | 0.0216 | 0.0346 | 901 | -0.0994 | 0.0018 | 0.0117 | 8954 | -0.1046 | 9.29E-05 | 0.0110 |
| ZCMP | 486 | -0.0283 | 0.0134 | 0.0142 | 971 | -0.0295 | 0.0132 | 0.0140 | 9665 | -0.0335 | 0.0139 | 0.0150 |
| YCMP | 492 | -0.0155 | 0.0127 | 0.0130 | 982 | -0.0182 | 0.0121 | 0.0124 | 9817 | -0.1826 | 0.0130 | 0.0133 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.16 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $Poisson(\lambda)$ เมื่อกำหนด $\lambda = 0.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 509 | 0.0184 | 0.0241 | 0.0244 | 1006 | 0.0056 | 0.0108 | 0.0108 | 10007 | 0.0007 | 0.0011 | 0.0011 |
| LB | 560 | 0.1192 | 0.3240 | 0.3382 | 1046 | 0.0465 | 0.1148 | 0.1170 | 10058 | 0.0058 | 0.0088 | 0.0088 |
| MLE | 506 | 0.0122 | 0.0153 | 0.0154 | 1004 | 0.0045 | 0.0067 | 0.0067 | 9996 | -0.0004 | 0.0007 | 0.0007 |
| LCMP | 178 | -0.6450 | 0.1022 | 0.5182 | 497 | -0.5033 | 0.0570 | 0.3102 | 5816 | -0.4184 | 0.0002 | 0.1752 |
| ZCMP | 251 | -0.4980 | 0.0056 | 0.2536 | 503 | -0.4968 | 0.0055 | 0.2523 | 5025 | -0.4975 | 0.0053 | 0.2528 |
| YCMP | 263 | -0.4744 | 0.0066 | 0.2317 | 526 | -0.4739 | 0.0064 | 0.2310 | 5228 | -0.4772 | 0.0065 | 0.2342 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.17 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $Poisson(\lambda)$ เมื่อกำหนด $\lambda = 1$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|-----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 505 | 0.0091 | 0.0052 | 0.0531 | 1002 | 0.0025 | 0.0026 | 0.0026 | 10003 | 0.0003 | 0.0003 | 0.0003 |
| LB | 519 | 0.0383 | 0.0312 | 0.0327 | 1013 | 0.0133 | 0.0139 | 0.0140 | 10019 | 0.0019 | 0.0013 | 0.0013 |
| MLE | 502 | 0.0042 | 0.0028 | 0.0028 | 1002 | 0.0018 | 0.0014 | 0.0014 | 10000 | 1.43E-05 | 0.0001 | 0.0001 |
| LCMP | 396 | -0.2081 | 0.0039 | 0.0472 | 781 | -0.2188 | 0.0015 | 0.0493 | 7697 | -0.2303 | 0.0001 | 0.0532 |
| ZCMP | 399 | -0.2014 | 0.0076 | 0.0482 | 796 | -0.2039 | 0.0075 | 0.0491 | 7947 | -0.2053 | 0.0078 | 0.0499 |
| YCMP | 401 | -0.1971 | 0.0070 | 0.0458 | 797 | -0.2031 | 0.0066 | 0.0478 | 7941 | -0.2059 | 0.0065 | 0.0489 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.18 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์
การแจกแจง $Poisson(\lambda)$ เมื่อกำหนด $\lambda = 2$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|-----------------|-----------------|-----------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 500 | 0.0098 | 0.0010 | 0.0010 | 1002 | 0.0018 | 0.0005 | 0.0005 | 9998 | -0.0002 | 4.85E-05 | 4.85E-05 |
| LB | 504 | 0.0071 | 0.0034 | 0.0035 | 1006 | 0.0058 | 0.0016 | 0.0016 | 9998 | -0.0002 | 0.0001 | 0.0001 |
| MLE | 500 | 0.0003 | 0.0005 | 0.0005 | 1001 | 0.0006 | 0.0002 | 0.0002 | 10000 | 3.75E-05 | 2.34E-05 | 2.34E-05 |
| LCMP | 454 | -0.0924 | 0.0005 | 0.0091 | 906 | -0.0942 | 0.0002 | 0.0091 | 9018 | -0.0982 | 1.99E-05 | 0.0097 |
| ZCMP | 520 | 0.0409 | 0.0035 | 0.0051 | 1037 | 0.0374 | 0.0036 | 0.0050 | 10378 | 0.0378 | 0.0034 | 0.0048 |
| YCMP | 508 | 0.0161 | 0.0024 | 0.0023 | 1017 | 0.0167 | 0.0022 | 0.0025 | 10147 | 0.0147 | 0.0021 | 0.0023 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.19 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $Geometric(p)$ เมื่อกำหนด $p = 0.3$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 427 | -0.1470 | 0.0018 | 0.0234 | 851 | -0.1487 | 0.0009 | 0.0230 | 8504 | -0.1496 | 8.98E-05 | 0.0225 |
| LB | 475 | -0.0494 | 0.0179 | 0.0203 | 938 | -0.0620 | 0.0075 | 0.0113 | 9263 | -0.0737 | 0.0006 | 0.0060 |
| MLE | 365 | -0.2702 | 0.0005 | 0.0735 | 730 | -0.2701 | 0.0002 | 0.0732 | 7300 | -0.2700 | 2.44E-05 | 0.0729 |
| LCMP | 299 | -0.4015 | 0.1537 | 0.3149 | 754 | -0.2459 | 0.0750 | 0.1355 | 8205 | -0.1795 | 0.0001 | 0.0324 |
| ZCMP | 390 | -0.2200 | 0.0011 | 0.0495 | 780 | -0.2196 | 0.0008 | 0.0490 | 7797 | -0.2203 | 0.0005 | 0.0490 |
| YCMP | 379 | -0.2429 | 0.0007 | 0.0597 | 757 | -0.2427 | 0.0005 | 0.0594 | 7579 | -0.2421 | 0.0002 | 0.0588 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.20 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $Geometric(p)$ เมื่อกำหนด $p = 0.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 378 | -0.2431 | 0.0044 | 0.0635 | 754 | -0.2458 | 0.0022 | 0.0626 | 7501 | -0.2499 | 0.0002 | 0.0626 |
| LB | 466 | -0.0689 | 0.0614 | 0.0661 | 903 | -0.0969 | 0.0227 | 0.0321 | 8761 | -0.1239 | 0.0018 | 0.0172 |
| MLE | 314 | -0.3710 | 0.0012 | 0.1388 | 628 | -0.3717 | 0.0006 | 0.1388 | 6277 | -0.3723 | 5.50E-05 | 0.1387 |
| LCMP | 319 | -0.3627 | 0.0509 | 0.1824 | 680 | -0.3196 | 0.0165 | 0.1187 | 6653 | -0.3347 | 0.0002 | 0.1122 |
| ZCMP | 310 | -0.3810 | 0.0027 | 0.1478 | 619 | -0.3813 | 0.0024 | 0.1478 | 6169 | -0.3831 | 0.0021 | 0.1490 |
| YCMP | 303 | -0.3942 | 0.0020 | 0.1574 | 605 | -0.3946 | 0.0018 | 0.1576 | 6056 | -0.3944 | 0.0015 | 0.1571 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.21 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง $Geometric(p)$ เมื่อกำหนด $p = 0.7$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|---------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 331 | -0.3378 | 0.0109 | 0.1251 | 657 | -0.3426 | 0.0053 | 0.1226 | 6508 | -0.3492 | 0.0005 | 0.1224 |
| LB | 470 | -0.0590 | 0.2896 | 0.2931 | 889 | -0.1112 | 0.1103 | 0.1227 | 8309 | -0.1691 | 0.0068 | 0.0354 |
| MLE | 285 | -0.4305 | 0.0042 | 0.1895 | 568 | -0.4321 | 0.0020 | 0.1887 | 5635 | -0.4365 | 0.0002 | 0.1191 |
| LCMP | 223 | -0.5544 | 0.0326 | 0.3399 | 478 | -0.5217 | 0.0130 | 0.2852 | 4672 | -0.5328 | 0.0003 | 0.2841 |
| ZCMP | 191 | -0.6170 | 0.0029 | 0.3836 | 383 | -0.6172 | 0.0026 | 0.3835 | 3817 | -0.6183 | 0.0023 | 0.3846 |
| YCMP | 194 | -0.6124 | 0.0027 | 0.3778 | 388 | -0.6123 | 0.0025 | 0.3774 | 3842 | -0.6158 | 0.0022 | 0.3814 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.22 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง *Negative Binomial*(k, θ) เมื่อกำหนด $k = 1, \theta = 0.3$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|----------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 427 | -0.1457 | 0.0019 | 0.0231 | 851 | -0.1490 | 0.0008 | 0.0230 | 8500 | -0.1500 | 8.76E-05 | 0.0226 |
| LB | 478 | -0.0438 | 0.0179 | 0.0198 | 935 | -0.0651 | 0.0068 | 0.0110 | 9264 | -0.0736 | 0.0006 | 0.0060 |
| MLE | 365 | -0.2700 | 0.0005 | 0.0734 | 730 | -0.2701 | 0.0002 | 0.0732 | 7297 | -0.2703 | 2.57E-05 | 0.0731 |
| LCMP | 292 | -0.4154 | 0.1587 | 0.3316 | 764 | -0.2360 | 0.0692 | 0.1249 | 8198 | -0.1802 | 0.0001 | 0.0326 |
| ZCMP | 390 | -0.2191 | 0.0011 | 0.0491 | 781 | -0.2195 | 0.0007 | 0.0489 | 7796 | -0.2204 | 0.0005 | 0.0491 |
| YCMP | 379 | -0.2421 | 0.0007 | 0.0593 | 757 | -0.2428 | 0.0004 | 0.0594 | 7567 | -0.2433 | 0.0002 | 0.0594 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.23 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง *Negative Binomial*(k, θ) เมื่อกำหนด $k = 3, \theta = 0.5$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|----------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 485 | -0.0295 | 0.0008 | 0.0017 | 970 | -0.0299 | 0.0004 | 0.0013 | 9688 | -0.0312 | 3.71E-05 | 0.0010 |
| LB | 502 | 0.0033 | 0.0044 | 0.0045 | 999 | -0.0014 | 0.0018 | 0.0018 | 9927 | -0.0073 | 0.0002 | 0.0002 |
| MLE | 454 | -0.0911 | 0.0003 | 0.0086 | 909 | -0.0914 | 0.0001 | 0.0085 | 9084 | -0.0916 | 1.25E-05 | 0.0084 |
| LCMP | 463 | -0.0733 | 0.0112 | 0.0166 | 931 | -0.0686 | 0.0003 | 0.0050 | 9197 | -0.0803 | 2.12E-05 | 0.0065 |
| ZCMP | 486 | -0.0288 | 0.0009 | 0.0018 | 970 | -0.0295 | 0.0008 | 0.0017 | 9707 | -0.0293 | 0.0006 | 0.0015 |
| YCMP | 469 | -0.0616 | 0.0005 | 0.0043 | 939 | -0.0610 | 0.0003 | 0.0041 | 9390 | -0.0610 | 0.0002 | 0.0039 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

ตารางที่ 4.24 ค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ การแจกแจง *Negative Binomial*(k, θ) เมื่อกำหนด $k = 5$, $\theta = 0.7$ และขนาดประชากร N เท่ากับ 500, 1,000, 10,000

| <i>Estimator</i> | $N = 500$ | | | | $N = 1,000$ | | | | $N = 10,000$ | | | |
|------------------|------------|---------------|---------------|---------------|-------------|---------------|---------------|---------------|--------------|----------------|-----------------|---------------|
| | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> | \bar{N} | <i>RBias</i> | <i>RVar</i> | <i>RMSE</i> |
| Chao | 488 | -0.0250 | 0.0012 | 0.0018 | 974 | -0.0264 | 0.0006 | 0.0013 | 9721 | -0.0279 | 5.45E-05 | 0.0008 |
| LB | 505 | 0.0093 | 0.0062 | 0.0063 | 1002 | 0.0017 | 0.0026 | 0.0026 | 9960 | -0.0040 | 0.0002 | 0.0002 |
| MLE | 462 | -0.0769 | 0.0004 | 0.0063 | 923 | -0.0768 | 0.0002 | 0.0061 | 9223 | -0.0777 | 2.10E-05 | 0.0061 |
| LCMP | 453 | -0.0931 | 0.0018 | 0.0105 | 900 | -0.1004 | 0.0004 | 0.0105 | 8903 | -0.1097 | 2.82E-05 | 0.0121 |
| ZCMP | 488 | -0.0233 | 0.0024 | 0.0029 | 975 | -0.0253 | 0.0022 | 0.0028 | 9748 | -0.0252 | 0.0020 | 0.0027 |
| YCMP | 474 | -0.0517 | 0.0014 | 0.0041 | 948 | -0.0523 | 0.0013 | 0.0040 | 9477 | -0.0523 | 0.0011 | 0.0038 |

หมายเหตุ \bar{N} แทนค่าประมาณขนาดประชากรเฉลี่ยจากการคำนวณ 2,000 รอบ (แสดงค่าเป็นจำนวนเต็ม)

4.2 การประยุกต์ใช้กับสถานการณ์จริง

ถึงแม้ว่าการจำลองสถานการณ์จะสามารถยืนยันได้ว่าตัวประมาณขนาดประชากรที่นำเสนอ นั้นมีประสิทธิภาพ เพื่อที่จะนำมาประมาณขนาดประชากรที่แท้จริงได้ แต่จากการจำลองสถานการณ์นั้นเอื้อต่อการใช้ตัวประมาณขนาดประชากรที่นำเสนอ จึงทำการประยุกต์ใช้ตัวประมาณขนาดประชากรกับสถานการณ์จริง เพื่อเป็นการยืนยันประสิทธิภาพของตัวประมาณขนาดประชากรดังกล่าวต่อไป

4.2.1 ข้อมูลโสเภณีในเมืองแวนคูเวอร์ (Street prostitutes in Vancouver)

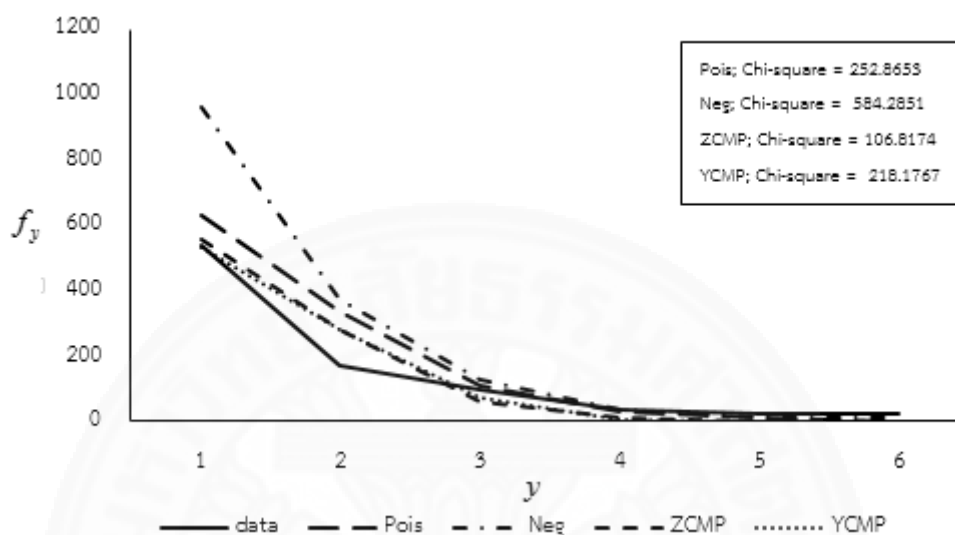
จากการสำรวจจำนวนโสเภณีในเมืองแวนคูเวอร์ ประเทศแคนาดาของ Rossmo และคณะ ในปี ค.ศ. 1990 (Rossmo et al., 1990) ซึ่งข้อมูลของโสเภณีในเมืองแวนคูเวอร์ที่ถูกจับ นั้นได้มาจากบันทึกของตำรวจ โดยแสดงจำนวนครั้งที่โสเภณีถูกจับในช่วงระยะเวลา 23 เดือน ดังตารางที่ 4.25

ตารางที่ 4.25 จำนวนโสเภณีที่ถูกจับในเมืองแวนคูเวอร์ ประเทศแคนาดา

| y | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|-----|-----|----|----|----|----|
| f_y | 541 | 169 | 95 | 37 | 21 | 23 |

จากตารางที่ 4.25 เห็นได้ว่า โสเภณีที่ถูกจับในเมืองแวนคูเวอร์มีจำนวนทั้งหมด 886 คน โดยมีโสเภณีจำนวน 541 คนที่ถูกจับเป็นครั้งแรก มีโสเภณีจำนวน 169 คนที่ถูกจับมาแล้ว 2 ครั้ง และมีโสเภณีที่ถูกจับหลายๆ ครั้ง โดยจำนวนครั้งที่ถูกจับสูงสุด คือ 6 ครั้ง แต่ในความเป็นจริงอาจมีโสเภณีที่ยังไม่เคยถูกจับเลยสักครั้ง ซึ่งเราไม่ทราบจำนวนและต้องการประมาณค่า ดังนั้นเราจะทำการประยุกต์ใช้ \hat{N}_{Chao} , \hat{N}_{LB} , \hat{N}_{MLE} , \hat{N}_{LCMP} , \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ประมาณจำนวนโสเภณีในเมืองแวนคูเวอร์ ประเทศแคนาดา เมื่อพิจารณาการแจกแจงปัวซองที่มีพารามิเตอร์ λ โดยที่ λ เท่ากับค่าเฉลี่ยของจำนวนครั้งที่โสเภณีถูกจับ ในที่นี้ได้ว่า $\lambda = 1.76$ เมื่อพิจารณาการแจกแจง Conway-Maxwell-Poisson ที่มีพารามิเตอร์ λ และ ν โดยให้ λ เท่ากับค่าเฉลี่ยของจำนวนครั้งของข้อมูลและ ν ได้จากค่าผกผันของอัตราส่วนของความแปรปรวนและค่าเฉลี่ย จะได้ค่า λ และ ν เท่ากับ 1.76 และ 1.21 ตามลำดับ นอกจากนี้ ยังได้ตรวจสอบความเหมาะสมของการแจกแจงของข้อมูลด้วยการ

ทดสอบภาวะสารูปดี (Goodness of Fit) โดยทดสอบด้วยไคกำลังสอง (Chi-squared test) พร้อมกับการใช้กราฟในการร่วมอธิบาย ดังแสดงในภาพที่ 4.1



ภาพที่ 4.1 แสดงการอนุมานข้อมูลโสมถณีในเมืองแวนคูเวอร์ ประเทศแคนาดา ด้วยการแจกแจงปัวซอง การแจกแจงทวินามเชิงลบ การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ และการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3

จากภาพที่ 4.1 เห็นได้จากกราฟว่า การอธิบายข้อมูลจำนวนโสมถณีในเมืองแวนคูเวอร์ ประเทศแคนาดาด้วยการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์และตัดปลายที่มากกว่า 3 นั้นเหมาะสมมากกว่าการอธิบายด้วยการแจกแจงปัวซองและการแจกแจงทวินามเชิงลบ โดยมีการทดสอบภาวะสารูปดีช่วยในการยืนยันว่า ข้อมูลจำนวนโสมถณีในเมืองแวนคูเวอร์ ประเทศแคนาดาเหมาะสมที่จะอธิบายด้วยการแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์และตัดปลายที่มากกว่า 3 มากกว่าการแจกแจงปัวซองและการแจกแจงทวินามเชิงลบ แต่การอธิบายด้วยการแจกแจงปัวซองและการแจกแจงทวินามเชิงลบก็สามารถเหมาะสมกับข้อมูลจำนวนโสมถณีในเมืองแวนคูเวอร์ได้เช่นกัน ส่วนผลลัพธ์ของตัวประมาณขนาดประชากรทั้ง 6 ตัวที่ได้จากการประยุกต์ใช้กับข้อมูลจำนวนโสมถณีในเมืองแวนคูเวอร์ ประเทศแคนาดา ซึ่งประมาณจำนวนโสมถณีในเมืองแวนคูเวอร์ได้ดังแสดงในตารางที่ 4.26

ตารางที่ 4.26 ค่าประมาณจำนวนโศภณีในเมืองแวนคูเวอร์ ประเทศแคนาดา

| \hat{N}_{Chao} | \hat{N}_{LB} | \hat{N}_{MLE} | \hat{N}_{LCMP} | \hat{N}_{ZCMP} | \hat{N}_{YCMP} |
|------------------|----------------|-----------------|------------------|------------------|------------------|
| 1752 | 3223 | 1240 | 3841 | 1213 | 1187 |

จากตารางที่ 4.26 แสดงค่าประมาณของจำนวนโศภณีในเมืองแวนคูเวอร์ ประเทศแคนาดา และจากการใช้กราฟเข้ามาช่วยในการอธิบาย จะเห็นว่า \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ให้ค่าประมาณจำนวนโศภณีที่เหมาะสมที่สุด เนื่องจากข้อมูลจำนวนโศภณีในเมืองแวนคูเวอร์ ประเทศแคนาดานั้น เป็น Under-dispersion และจากการจำลองสถานการณ์ พบว่า \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีประสิทธิภาพกรณีที่จำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson ที่เป็น Under-dispersion ดังนั้น \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จึงเป็นตัวประมาณที่น่าไว้วางใจในการนำมาใช้ประมาณขนาดประชากร

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวประมาณขนาดประชากรสำหรับประมาณขนาดประชากรที่พบได้ยากหรือโอกาสที่หน่วยตัวอย่างถูกสำรวจพบมีค่าน้อย โดยพัฒนาตัวประมาณขนาดประชากร 2 ตัว ซึ่งพัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ และตัดปลายที่มากกว่า 3 ได้นำเสนอไว้ในบทที่ 3 นอกจากนี้ยังทำการจำลองสถานการณ์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson กับตัวประมาณขนาดประชากรเดิมที่มีอยู่แล้ว พร้อมกับประยุกต์ใช้ตัวประมาณขนาดประชากรกับข้อมูลจริงถูกแสดงไว้ในบทที่ 4 ในส่วนของการสรุปผลการวิจัยและข้อเสนอแนะ ซึ่งเป็นการสรุปผลโดยรวม โดยจะทำการสรุปทั้งผลของตัวประมาณขนาดประชากร 2 ตัวที่พัฒนาขึ้นมาและผลการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรอื่นๆที่ได้นำเสนอ พร้อมกับทำให้ข้อเสนอแนะเพิ่มเติม ซึ่งจะแสดงไว้ในบทที่ 5 นี้ด้วย

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์และตัดปลายที่มากกว่า 3 และเปรียบเทียบประสิทธิภาพกับตัวประมาณขนาดประชากรอื่นๆที่นำเสนอ ดังนั้นในการสรุปผลการวิจัยนี้จึงแบ่งออกเป็น 2 ส่วน ดังต่อไปนี้

5.1.1 สรุปผลการพัฒนาตัวประมาณขนาดประชากร

การพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ ซึ่งได้อธิบายไว้ในบทที่ 3 ช่วยในการประมาณค่าจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ (f_0) ทำให้ได้ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่ศูนย์ ดังนี้

$$\hat{N}_{ZCMP} = n + \frac{n}{Z(\hat{\lambda}, \hat{\nu}) - 1}$$

โดยที่

$$Z(\hat{\lambda}, \hat{\nu}) = \sum_{j=0}^{\infty} \frac{\hat{\lambda}^j}{(j!)^{\hat{\nu}}}$$

นอกจากนี้ยังพัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3 หรืออาศัยจำนวนหน่วยตัวอย่างที่ถูกสำรวจพบ 1, 2 และ 3 ครั้ง ช่วยในการประมาณค่าจำนวนหน่วยตัวอย่างที่ไม่ถูกสำรวจพบ (f_0) ทำให้ได้ตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ตัดปลายที่มากกว่า 3 ดังนี้

$$\hat{N}_{YCMP} = n + \left(\frac{f_1 + f_2 + f_3}{\hat{\lambda} + \frac{\hat{\lambda}^2}{(2!)^{\hat{\nu}}} + \frac{\hat{\lambda}^3}{(3!)^{\hat{\nu}}}} \right)$$

5.1.2 สรุปผลการเปรียบเทียบประสิทธิภาพตัวประมาณขนาดประชากร

การจำลองสถานการณ์ด้วยโปรแกรม Rstudio เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร ซึ่งกำหนดสถานการณ์จำลองภายใต้การแจกแจงไว้ 4 การแจกแจง ได้แก่ จำลองภายใต้การแจกแจง Conway-Maxwell-Poisson เมื่อ $\lambda = 0.5, 1.0, 2.0$ และ $\nu = 0.4, 0.6, 0.8, 1.5, 2.0$ จำลองภายใต้การแจกแจงปัวซอง เมื่อ $\lambda = 0.5, 1.0, 2.0$ จำลองภายใต้การแจกแจงเรขาคณิต เมื่อ $p = 0.3, 0.5, 0.7$ และจำลองภายใต้การแจกแจงทวินามเชิงลบ เมื่อ $(k, \theta) = (1, 0.3), (3, 0.5), (5, 0.7)$ สำหรับประมาณขนาดประชากร N เท่ากับ 500, 1,000 และ 10,000 โดยแต่ละสถานการณ์จำลองจะทำซ้ำๆ จนครบ 2,000 รอบ รวมทั้งสิ้น 72 สถานการณ์ และเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร โดยพิจารณาจากค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ดังนี้

5.1.1.1 ภายใต้การแจกแจง Conway-Maxwell-Poisson

ผลที่ได้จากการจำลองจะเห็นว่า เมื่อพารามิเตอร์ λ และ ν มีค่าสูงๆ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ให้ค่าความเอนเอียงสัมพัทธ์สัมพัทธ์ต่ำที่สุด นั่นแสดงว่าทั้ง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่าประมาณขนาดประชากรได้แม่นยำกว่า \hat{N}_{Chao} , \hat{N}_{LB} , \hat{N}_{MLE} และ \hat{N}_{LCMP} เมื่อพิจารณาค่าความแปรปรวนสัมพัทธ์ \hat{N}_{MLE} ให้ต่ำที่สุด แต่เมื่อขนาดประชากร $N = 500$ และ λ และ ν มีค่าสูง \hat{N}_{YCMP} ให้ค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ต่ำ ดังนั้นสามารถสรุปได้ว่า \hat{N}_{YCMP} มีประสิทธิภาพในการประมาณขนาดประชากรสูงที่สุดเมื่อขนาดประชากร $N = 500$ ในขณะที่ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ให้ค่าประมาณขนาดประชากรได้แม่นยำกว่า \hat{N}_{LB} , \hat{N}_{MLE} และ \hat{N}_{LCMP} ดังนั้นสามารถสรุปได้ว่า \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะมีประสิทธิภาพการประมาณขนาดประชากรสูงใกล้เคียงกัน นอกจากนี้จะเห็นว่าเมื่อ λ มีค่าสูง ตัวประมาณขนาดประชากรทั้ง 6 ตัวจะมีประสิทธิภาพในการประมาณขนาดประชากรเพิ่มขึ้น

ตารางที่ 5.1 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความเอนเอียงสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson

| (λ, ν) | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|------------------|------------------|------------------|
| (0.5,0.4) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (1,0.4) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,0.4) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (0.5,0.6) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (1,0.6) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,0.6) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (0.5,0.8) | \hat{N}_{Chao} | \hat{N}_{LB} | \hat{N}_{LB} |
| (1,0.8) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,0.8) | \hat{N}_{YCMP} | \hat{N}_{YCMP} | \hat{N}_{YCMP} |
| (0.5,1.5) | \hat{N}_{Chao} | \hat{N}_{LB} | \hat{N}_{LB} |
| (1,1.5) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,1.5) | \hat{N}_{ZCMP} | \hat{N}_{YCMP} | \hat{N}_{YCMP} |
| (0.5,2) | \hat{N}_{YCMP} | \hat{N}_{LB} | \hat{N}_{LB} |
| (1,2) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,2) | \hat{N}_{YCMP} | \hat{N}_{YCMP} | \hat{N}_{LB} |

ตารางที่ 5.2 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความแปรปรวนสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson

| (λ, ν) | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|------------------|---------------------------------|------------------|
| (0.5,0.4) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{LCMP} |
| (1,0.4) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| (2,0.4) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{LB} |
| (0.5,0.6) | \hat{N}_{YCMP} | \hat{N}_{MLE} | \hat{N}_{LCMP} |
| (1,0.6) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| (2,0.6) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{LB} |
| (0.5,0.8) | \hat{N}_{ZCMP} | \hat{N}_{MLE} | \hat{N}_{LCMP} |
| (1,0.8) | \hat{N}_{MLE} | \hat{N}_{LCMP} | \hat{N}_{LCMP} |
| (2,0.8) | \hat{N}_{LCMP} | $\hat{N}_{MLE}, \hat{N}_{LCMP}$ | \hat{N}_{LCMP} |
| (0.5,1.5) | \hat{N}_{ZCMP} | \hat{N}_{ZCMP} | \hat{N}_{MLE} |
| (1,1.5) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{LCMP} |
| (2,1.5) | \hat{N}_{LCMP} | \hat{N}_{LCMP} | \hat{N}_{LCMP} |
| (0.5,2) | \hat{N}_{ZCMP} | \hat{N}_{ZCMP} | \hat{N}_{MLE} |
| (1,2) | \hat{N}_{ZCMP} | \hat{N}_{ZCMP} | \hat{N}_{LCMP} |
| (2,2) | \hat{N}_{MLE} | \hat{N}_{LCMP} | \hat{N}_{LCMP} |

ตารางที่ 5.3 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ภายใต้การแจกแจง Conway-Maxwell-Poisson

| (λ, ν) | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|------------------|------------------|------------------|
| (0.5,0.4) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{Chao} |
| (1,0.4) | \hat{N}_{Chao} | \hat{N}_{LB} | \hat{N}_{LB} |
| (2,0.4) | \hat{N}_{ZCMP} | \hat{N}_{ZCMP} | \hat{N}_{LB} |
| (0.5,0.6) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (1,0.6) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (2,0.6) | \hat{N}_{Chao} | \hat{N}_{ZCMP} | \hat{N}_{LB} |

ตารางที่ 5.3 (ต่อ)

| (λ, ν) | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|------------------|------------------|------------------|
| (0.5,0.8) | \hat{N}_{MLE} | \hat{N}_{Chao} | \hat{N}_{Chao} |
| (1,0.8) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (2,0.8) | \hat{N}_{YCMP} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (0.5,1.5) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (1,1.5) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (2,1.5) | \hat{N}_{YCMP} | \hat{N}_{LB} | \hat{N}_{LB} |
| (0.5,2) | \hat{N}_{YCMP} | \hat{N}_{YCMP} | \hat{N}_{LB} |
| (1,2) | \hat{N}_{YCMP} | \hat{N}_{YCMP} | \hat{N}_{LB} |
| (2,2) | \hat{N}_{YCMP} | \hat{N}_{LCMP} | \hat{N}_{LB} |

5.1.1.2 ภายใต้การแจกแจงปัวซอง

ผลที่ได้จากการจำลองจะเห็นว่าทั้ง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะมีประสิทธิภาพในการประมาณขนาดประชากรที่ต่ำกว่าการจำลองภายใต้การแจกแจง Conway-Maxwell-Poisson เนื่องจาก \hat{N}_{MLE} พัฒนภายใต้การแจกแจงปัวซองตัดปลายที่ค่าศูนย์ ซึ่งทำให้ \hat{N}_{MLE} ให้ค่าความเอนเอียงสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ต่ำที่สุด แสดงว่า \hat{N}_{MLE} มีประสิทธิภาพการประมาณขนาดประชากรสูงที่สุดในส่วนของ \hat{N}_{YCMP} จะมีประสิทธิภาพค่อนข้างสูงกว่า \hat{N}_{ZCMP} นอกจากนี้จะเห็นว่าเมื่อ λ มีค่าสูง ตัวประมาณขนาดประชากรทั้ง 6 ตัวจะมีประสิทธิภาพในการประมาณขนาดประชากรเพิ่มขึ้น

5.1.1.2 ภายใต้การแจกแจงเรขาคณิต

ผลที่ได้จากการจำลองจะเห็นว่าทั้ง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะมีประสิทธิภาพการประมาณขนาดประชากรที่ต่ำกว่าการจำลองภายใต้การแจกแจง Conway-Maxwell-Poisson เช่นเดียวกันกับการแจกแจงปัวซอง ซึ่ง \hat{N}_{LB} ให้ค่าความเอนเอียงสัมพัทธ์ต่ำที่สุด และเมื่อประชากรมีขนาดใหญ่ \hat{N}_{LB} ให้ค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ต่ำที่สุด แสดงว่า \hat{N}_{LB} จะมีประสิทธิภาพการประมาณขนาดประชากรสูงสุดเมื่อประชากรมีขนาดใหญ่ เมื่อพิจารณาค่าความแปรปรวนสัมพัทธ์จะเห็นว่า \hat{N}_{MLE} ให้ค่าต่ำที่สุดในกรณีขนาดประชากร $N = 500$ ทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} จะให้ค่าความแปรปรวนสัมพัทธ์ต่ำกว่า \hat{N}_{Chao} , \hat{N}_{LCMP} และ \hat{N}_{LB} นอกจากนี้ \hat{N}_{YCMP} มีประสิทธิภาพในการประมาณขนาดประชากรสูงกว่า \hat{N}_{ZCMP}

5.1.1.3 ภายใต้การแจกแจงทวินามเชิงลบ

เช่นเดียวกับผลลัพธ์ที่ได้จากการแจกแจงเรขาคณิต เนื่องจาก \hat{N}_{LB} พัฒนาภายใต้การแจกแจงทวินามเชิงลบ ทำให้ \hat{N}_{LB} ให้ค่าความเอนเอียงสัมพัทธ์ต่ำที่สุด และให้ค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ต่ำที่สุดเมื่อประชากรมีขนาดใหญ่ แสดงว่า \hat{N}_{LB} มีประสิทธิภาพในการประมาณขนาดประชากรสูงสุดเมื่อประชากรมีขนาดใหญ่ ในขณะที่ขนาดประชากร $N = 500$ เมื่อ k และ θ เพิ่มขึ้น \hat{N}_{ZCMP} และ \hat{N}_{Chao} ให้ค่าความเอนเอียงสัมพัทธ์ต่ำใกล้เคียงกันมาก ดังนั้น \hat{N}_{ZCMP} มีประสิทธิภาพในการประมาณขนาดประชากรสูงเมื่อขนาดประชากร $N = 500$ นอกจากนี้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ให้ค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ต่ำกว่า \hat{N}_{LB} , \hat{N}_{LCMP} และ \hat{N}_{MLE} เมื่อขนาดประชากร $N = 500$ ในการพิจารณาค่าความแปรปรวนสัมพัทธ์จะเห็นว่า \hat{N}_{MLE} ให้ค่าต่ำที่สุด อย่างไรก็ตามเมื่อ k และ θ สูงขึ้นขึ้นจะทำให้ตัวประมาณขนาดประชากรทั้ง 6 ตัว มีประสิทธิภาพในการประมาณขนาดประชากรเพิ่มขึ้น

ตารางที่ 5.4 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความเอนเอียงสัมพัทธ์ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ

| <i>Parameter</i> | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|--|-----------------|-----------------|
| λ | <i>Poisson</i> (λ) | | |
| $\lambda = 0.5$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $\lambda = 1$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $\lambda = 2$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| p | <i>Geometric</i> (p) | | |
| $p = 0.3$ | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| $p = 0.5$ | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| $p = 0.7$ | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (k, θ) | <i>Negative Binomial</i> (k, θ) | | |
| (1,0.3) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (3,0.5) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| (5,0.7) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |

ตารางที่ 5.5 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความแปรปรวนสัมพัทธ์ ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ

| <i>Parameter</i> | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|-----------|-------------|--------------|
|------------------|-----------|-------------|--------------|

| λ | <i>Poisson</i> (λ) | | |
|-----------------|--|---------------------------------|---------------------------------|
| $\lambda = 0.5$ | \hat{N}_{ZCMP} | \hat{N}_{ZCMP} | \hat{N}_{LCMP} |
| $\lambda = 1$ | \hat{N}_{MLE} | \hat{N}_{MLE} | $\hat{N}_{MLE}, \hat{N}_{LCMP}$ |
| $\lambda = 2$ | $\hat{N}_{MLE}, \hat{N}_{LCMP}$ | $\hat{N}_{MLE}, \hat{N}_{LCMP}$ | \hat{N}_{LCMP} |
| p | <i>Geometric</i> (p) | | |
| $p = 0.3$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $p = 0.5$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $p = 0.7$ | \hat{N}_{ZCMP} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| (k, θ) | <i>Negative Binomial</i> (k, θ) | | |
| (1,0.3) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| (3,0.5) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| (5,0.7) | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |

ตารางที่ 5.6 ตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ภายใต้การแจกแจงปัวซอง การแจกแจงเรขาคณิต และการแจกแจงทวินามเชิงลบ

| <i>Parameter</i> | $N = 500$ | $N = 1,000$ | $N = 10,000$ |
|------------------|------------------------------|------------------|-----------------|
| λ | <i>Poisson</i> (λ) | | |
| $\lambda = 0.5$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $\lambda = 1$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| $\lambda = 2$ | \hat{N}_{MLE} | \hat{N}_{MLE} | \hat{N}_{MLE} |
| p | <i>Geometric</i> (p) | | |
| $p = 0.3$ | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |
| $p = 0.5$ | \hat{N}_{Chao} | \hat{N}_{LB} | \hat{N}_{LB} |
| $p = 0.7$ | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |

ตารางที่ 5.6 (ต่อ)

| (k, θ) | <i>Negative Binomial</i> (k, θ) | | |
|---------------|--|----------------|----------------|
| (1,0.3) | \hat{N}_{LB} | \hat{N}_{LB} | \hat{N}_{LB} |

| | | | |
|---------|------------------|------------------|----------------|
| (3,0.5) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |
| (5,0.7) | \hat{N}_{Chao} | \hat{N}_{Chao} | \hat{N}_{LB} |

จากการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร จะเห็นว่า ในการจำลองสถานการณ์ภายใต้การแจกแจง Conway-Maxwell-Poisson กรณีที่ข้อมูลเป็น Under-dispersion หรือ $\nu > 1$ ที่ขนาดประชากร $N = 500, 1,000$ และ λ มีค่าสูงๆ จะทำให้ \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีประสิทธิภาพสูงสุด ถึงแม้ว่า \hat{N}_{LB} จะมีความแปรปรวนสูงมาก แต่เมื่อประชากรมีขนาดใหญ่ขึ้นความแปรปรวนจะลดลง ส่วนในขนาดประชากร $N = 10,000$ จะทำให้ \hat{N}_{LB} มีประสิทธิภาพสูงสุด นอกจากนี้ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจง Conway-Maxwell-Poisson จะมีความเที่ยงสูง กรณีที่ข้อมูลเป็น Over-dispersion หรือ $\nu < 1$ ในทุกขนาดประชากร \hat{N}_{Chao} และ \hat{N}_{LB} จะมีประสิทธิภาพสูงสุด ในการจำลองสถานการณ์ภายใต้การแจกแจงปัวซอง \hat{N}_{MLE} มีประสิทธิภาพสูงสุด เนื่องจาก \hat{N}_{MLE} นั้นพัฒนาภายใต้การแจกแจงปัวซอง ส่วนในการจำลองสถานการณ์ภายใต้การแจกแจงเรขาคณิตและการแจกแจงทวินามเชิงลบ ในทุกขนาดประชากร \hat{N}_{Chao} และ \hat{N}_{LB} จะมีประสิทธิภาพสูงสุด ซึ่งเราสามารถนำตัวประมาณขนาดประชากรทั้ง 6 ตัวที่นำเสนอมาใช้ในการประมาณขนาดประชากรได้

นอกจากนี้การประยุกต์ใช้ตัวประมาณขนาดประชากรกับข้อมูลจำนวนโสภณีในเมืองแวนคูเวอร์ ประเทศแคนาดายังเป็นตัวยืนยันประสิทธิภาพของตัวประมาณขนาดประชากร ซึ่งข้อมูลจำนวนโสภณีในเมืองแวนคูเวอร์นั้นเป็น Under-dispersion และจากการการจำลองสถานการณ์ พบว่า \hat{N}_{ZCMP} และ \hat{N}_{YCMP} มีประสิทธิภาพกรณีที่จำลองภายใต้การแจกแจง Conway-Maxwell-Poisson ที่ข้อมูลเป็น Under-dispersion ดังนั้น \hat{N}_{ZCMP} และ \hat{N}_{YCMP} เป็นอีกตัวเลือกที่ใช้ในการประมาณขนาดประชากรของข้อมูลจำนวนโสภณีในเมืองแวนคูเวอร์ ประเทศแคนาดา

5.2 ข้อเสนอแนะเพิ่มเติม

5.2.1 สำหรับตัวประมาณขนาดประชากรภายใต้การแจกแจง Conway-Maxwell-Poisson ทั้ง \hat{N}_{ZCMP} และ \hat{N}_{YCMP} ให้ผลลัพธ์ที่มีประสิทธิภาพที่ใกล้เคียงกันมาก สามารถเลือกใช้ \hat{N}_{YCMP} ในการประมาณขนาดประชากรแทน \hat{N}_{ZCMP} ได้ เนื่องจาก \hat{N}_{YCMP} ใช้เวลาในการคำนวณน้อยกว่า

5.2.2 งานวิจัยนี้พัฒนาเพียงตัวประมาณขนาดประชากรแบบจุดเท่านั้น และเพื่อประโยชน์ในการประมาณขนาดประชากรควรมีการศึกษาการประมาณค่าแบบช่วงร่วมด้วย ซึ่งในงานวิจัยนี้ยังไม่มีการพัฒนาตัวประมาณแบบช่วงภายใต้การแจกแจง Conway-Maxwell-Poisson ดังนั้นควรมีการศึกษาและพัฒนาตัวประมาณขนาดประชากรแบบช่วง



รายการอ้างอิง

หนังสือและบทความในหนังสือ

Geoffrey, J. M., Thriyambakam, K. (1996). The EM algorithm and Extensions. Wiley Series in Probability and Statistics.

บทความวารสาร

Anan, O., Böhning, D., and Maruotti A. (2016). Population size estimation and heterogeneity in capture-recapture data: a linear regression estimator based on the Conway-Maxwell-Poisson distribution. *Statistical Methods and Applications*, DOI:10. 1007

Baker, S.G. (2016). A simple EM algorithm for capture-recapture data with categorical covariates. *International Biometric Society*, 46, 1193-1200.

Böhning, D., Dietz, E., Kuhnert, R. and Schön, D. (2005). Mixture models for capture-recapture count data. *Statistical Methods & Applications*, DOI:10. 1007

Böhning, D., Baksh, M.F., Lerdsuwansri, R., and Gallagher, J. (2013). The use of the ratio-plot in capture-recapture estimation. *Journal of Computational and Graphical Statistics*, 22, 133-155.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catechability. *International Biometric Society*, 43, 783-791.

Gupta, M.R., Chen, Y. (2010). Theory and use of the EM Algorithm. *Foundations and Trends in Signal Processing*, 4(3), 223-296.

Horvitz, D.G., Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

รายการอ้างอิง(ต่อ)

บทความวารสาร

- Köse, T., Orman, M., Ikiz, F., Baksh, M.F., Gallagher, J., and Böhning, D. (2014). Extending the Lincoln-Petersen estimator for multiple indentifications in one source. *Statistics in Medicine*, 33, 4237-4249.
- Lanumteang, K. and Böhning, D. (2011). An extension of Chao's estimator of population size based on the first three capture frequency counts. *Computational Statistics and Data Analysis*, 55, 2302-2311.
- McCarty, D.J., Tull, E.S., Moy, C.S., Kwoh, C.K., LaPorte R.E. (1992). Ascertainment corrected rates: Applications of capture-recapture methods. *Int. J. Epidemiol*, 22, 559-562.
- Mckeganey, N., Barnard, M., Leyland, A., Coote, I., and Follet, E. (1992). Female streetworking prostitution and HIV infection in Glasgow. *British Medical Journal*, 305(6857), 801-804.
- Rossmo, D.K. and Routledge, R. (1990). Estimating the size of criminal populations. *J. Quant. Crim.* 6, 293-314.
- Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., and Boatwright P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, 54(1), 127-142.
- Sellers, K.F., Borle, S., and Shmueli, G. (2011). The COM-Poisson model for count data: a survey of the methods and applications. *Applied Stochastic Models in Business and Industry*, 28, 104-116.

วิทยานิพนธ์

- Thepmong, N. (2014), Population Size Estimation Based Upon Zero-Modified Poisson Models. M.S. (Applied Statistics) Thesis, Thammasat University



ภาคผนวก

ภาคผนวก ก

คำสั่งการทำงานของโปรแกรม Rstudio สำหรับเปรียบเทียบประสิทธิภาพตัวประมาณ
ขนาดประชากร

```
#####

##For Zero-truncated Based Upon Conway-Maxwell-Poisson Distribution##
##Nhat.ZCMP##

#####

library(compoisson)

library(stats)

library(VGAM)

mle.zcmp <- function(par,Y){
  lambda <- par[1]
  nu <- par[2]
  Z <- 1+lambda+(lambda^2)/(factorial(2)^nu)+(lambda^3)/(factorial(3)^nu)
  error <- runif(1000)
  error.m <- mean(error)
  Re.k <- lambda^4/(((factorial(4))^nu)*(1-error.m))
  Z.k <- Z+Re.k
  #find m
  L <- length(Y)
  m <- 0
  for(j in 1:L){
```

```

    if(Y[j]>m){m=Y[j]}
  }

f <- numeric(m)
k <- numeric(m)
s <- numeric(m)
for(i in 1:m){
  f[i] <- length(Y[Y==i])
  k[i] <- log(((lambda^i)/(factorial(i)^nu))/Z.k)
  s[i] <- f[i]*k[i]
}
n <- length(Y[Y!=0])
f0.hat <- n/(Z.k-1)
log.L <- f0.hat*log(1/Z.k)+sum(s)
return(-log.L)
}

```

```

MLE.Z <- function(lambda,nu,Y){
  fit <- nlm(b(c(lambda,nu), mle.zcmp,Y=Y,lower=2)
  return(fit)
}

```

```

Z.k <- function(lambda,nu){
  Z <- 1+lambda+(lambda^2)/(factorial(2)^nu)+(lambda^3)/(factorial(3)^nu)
  error <- runif(1000)
}

```

```

error.m <- mean(error)

Re.k <- lambda^4/(((factorial(4))^nu)*(1-error.m))

Z.k <- Z+Re.k

return(Z.k)

}

```

```

Nhat.Z <- function(N,lambda,nu,Y){

  n<- length(Y[Y!=0])

  S <- sum(Y)

  f0.hat <- n/(Z.k(lambda,nu)-1)

  nu.hat <- MLE.Z(lambda,nu,Y)$par[2]

  ## EM algorithm

  oldf0 <- 0

  oldlambda <- S/n

  oldnu <- nu.hat

  f0.new <- n/(Z.k(oldlambda,oldnu)-1)

  diff.f0 <- abs(oldf0-f0.new)

  Y <- rcom(N,oldlambda,oldnu)

  Mle <- MLE.Z(oldlambda,oldnu,Y)

  lambda.new <- mle$par[1]

  nu.new <- mle$par[2]

  crit.lambda <- abs(oldlambda-lambda.new)

  crit.nu <- abs(oldnu-nu.new)

```

```

while(diff.f0>=0.0001&&crit.lambda>=0.0001 && crit.nu>=0.0001){

oldf0 <- f0.new

oldlambda <- lambda.new

oldnu <- nu.new

f0.new <- n/(Z.k(lambda.new,nu.new)-1)

Y <- rcom(N,oldlambda,oldnu)

Mle <- MLE.Z(oldlambda,oldnu,Y)

lambda.new <- mle$par[1]

nu.new <- mle$par[2]

diff.f0 <- abs(oldf0-f0.new)

crit.lambda <- abs(oldlambda-lambda.new)

crit.nu <- abs(oldnu-nu.new)

}

return(n+f0.new)

}

#####

##For Truncated Larger Than Three Based Upon Conway-Maxwell-Poisson ##

##Nhat.YCMP##

#####

mle.ycmp <- function(par,Y){

lambda <- par[1]

nu <- par[2]

Z <- 1+lambda+(lambda^2)/(factorial(2)^nu)+(lambda^3)/(factorial(3)^nu)

error <- runif(1000)

```



```

error.m <- mean(error)

Re.k <- lambda^4/(((factorial(4))^nu)*(1-error.m))

Z.k <- Z+Re.k

f1 <- length(Y[Y==1])
f2 <- length(Y[Y==2])
f3 <- length(Y[Y==3])

f0.hat <- (f1+f2+f3)/(lambda*(1+lambda/((factorial(2))^nu)
          +lambda^2/((factorial(3))^nu)))

k1 <- (log((lambda)/(factorial(1)^nu)) log((lambda+(lambda^2/factorial(2)^nu)
          +(lambda^3/factorial(3)^nu))))

k2 <- (log((lambda^2)/(factorial(2)^nu))
       -log((lambda+(lambda^2/factorial(2)^nu)+(lambda^3/factorial(3)^nu))))

k3 <- (log((lambda^3)/(factorial(3)^nu))
       -log((lambda+(lambda^2/factorial(2)^nu)+(lambda^3/factorial(3)^nu))))

log.L <- f0.hat*log(1/Z.k)+(k1*f1+k2*f2+k3*f3)

return(-log.L)

}

MLE <- function(lambda,nu,Y){

  fit <- nlm(b,c(lambda,nu),mle.ycmp,Y=Y,lower = 2)

  return(fit)

}

Nhat<- function(N,lambda,nu,Y){

  f1 <- length(Y[Y==1])

```

```

f2 <- length(Y[Y==2])
f3 <- length(Y[Y==3])
n <- length(Y[Y!=0])
S <- sum(Y)
f0.hat <- (f1+f2+f3)/(lambda*(1+lambda/((factorial(2))^nu)
          +lambda^2/((factorial(3))^nu)))
nu.hat <- MLE(lambda,nu,Y)$par[2]

#EM algorithm
oldf0 <- 0
oldlambda <- S/n
oldnu <- nu.hat
f0.new <- (f1+f2+f3)/(oldlambda*(1+oldlambda/((factorial(2))^oldnu)
          +oldlambda^2/((factorial(3))^oldnu)))
diff.f0 <- abs(oldf0-f0.new)
Y <- rcom(N,oldlambda,oldnu)
mle <- MLE(oldlambda,oldnu,Y)
lambda.new <- mle$par[1]
nu.new <- mle$par[2]
crit.lambda <- abs(oldlambda-lambda.new)
crit.nu <- abs(oldnu-nu.new)
while(diff.f0>=0.0001&&crit.lambda>=0.0001 && crit.nu>=0.0001){
oldf0 <- f0.new
oldlambda <- lambda.new

```

```

oldnu <- nu.new

f0.new <- (f1+f2+f3)/(lambda.new*(1+lambda.new/((factorial(2))^nu.new)
          +lambda.new^2/((factorial(3))^nu.new)))

Y <- rcom(N,oldlambda,oldnu)

mle <- MLE(oldlambda,oldnu,Y)

lambda.new <- mle$par[1]

nu.new <- mle$par[2]

diff.f0 <- abs(oldf0-f0.new)

crit.lambda <- abs(oldlambda-lambda.new)

crit.nu <- abs(oldnu-nu.new)

}

return(n+f0.new)

}

#####

##For Linear Regression Estimator Based On Coway-Maxwell-Poisson##

##Nhat.LCMP##

#####

getdata <- function(getY,getfY){ ## adjust for freq = 0

  Ylist <- c(0:max(getY))

  if(length(getY)==length(Ylist)) {

    outY <- matrix(cbind(getY,getfY),nrow=length(getY),ncol=2,byrow=FALSE)}

  else {

    "%w/o%" <- function(u,v) u[!u %in% v] #-- u without v

    Ymis <- Ylist %w/o% getY

```

```

fmis <- rep(NA,length(Ymis))

mis_fig <- cbind(Ymis,fmis)

obs_fig <- cbind(getY,getfY)

com_list <- rbind(obs_fig,mis_fig)

aa <- com_list[,1]

bb <- com_list[,2]

o <- order(aa)

outY <- matrix(cbind(aa[o],bb[o]),nrow=length(Ylist),ncol=2,byrow=FALSE)}

return(outY)

}

#####

##For Zero-truncated Poisson Distribution##

## Nhat.MLE##

#####

EM.MLE <- function(S,n){

  f0.old <- 0

  lamb.old <- S/n

  f0.new <- n*exp(-lamb.old)/(1-exp(-lamb.old))

  lamb.new <- S/(n+f0.new)

  diff_f0 <- f0.old-f0.new

  diff_lam <- lamb.old-lamb.new

  while(abs(diff_f0)>=0.0001 && abs(diff_lam)>=0.0001){

    f0.old <- f0.new

    lamb.old <- lamb.new

```

```

f0.new <- n*exp(-lamb.old)/(1-exp(-lamb.old))

lamb.new <- S/(n+f0.new)

diff_f0 <- f0.old-f0.new

diff_lam <- lamb.old-lamb.new

}

lambdahatMLE <- lamb.new

return(lambdahatMLE)

}

#####

##For performance comparison Relative Bias, Relative Variance and Relative MSE##

    ##Based upon Conway-Maxwell-Poisson distribution##

#####

perform_cmp <- function(N,lambda,nu,B){

  res <- matrix(numeric(6*B),ncol = 6,nrow = B)

  Nhat.Chao <- rep(0,B)

  Nhat.LB <- rep(0,B)

  Nhat.LCMP <- rep(0,B)

  Nhat.MLE <- rep(0,B)

  Nhat.ZCMP <- rep(0,B)

  Nhat.YCMP <- rep(0,B)

  for(l in 1:B){

    Y <- rcom(N,lambda,nu)

    S <- sum(Y)

    f1 <- length(Y[Y==1])

```

```
f2 <- length(Y[Y==2])
f3 <- length(Y[Y==3])
n <- length(Y[Y!=0])
lambdahatMLE <- EM.MLE(S,n)

# ratio plot Nhat.LCMP
tab<-table(Y)
dtab <- as.data.frame(tab)
out <- attributes(tab)
outY <- out$dimnames$Y
getY <- as.numeric(outY)
getfY <- as.numeric(dtab$Freq) ## get fy
size <- max(getY)
Ydat <- getdata(getY,getfY)
obsdat <- Ydat[-1,] ### observed frequency
n <- sum(obsdat[,2],na.rm=TRUE)
newsiz <- size
track.na <- 0
for(j in 1:size){
  if(is.na(obsdat[j,2])){
    track.na <- cbind(track.na,j)}
}
track.na <- track.na[-1]
if(length(track.na > 0)){
```

```

        obsdat <- obsdat[-track.na,]
        size <- size-length(track.na)
    }
r_Y <- rep(0,size)
w_Y <- rep(0,size)
for (j in 1:size){
    if(j==size){r_Y[j]<-NA; w_Y[j]<-NA }
    else{r_Y[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
        w_Y[j]<-(((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
    } ## end loop j
ex_na <- na.omit(r_Y)
pnt_no <- length(ex_na)
if(pnt_no > 2){
    mod <- lm(r_Y~obsdat[,1],weights=w_Y)
    sum<-summary.lm(mod)
    beta0_hat<-coefficients(summary(mod))[1]
    beta1_hat<-coefficients(summary(mod))[2]

    if(beta1_hat<=1){
        Nhat.LCMP[l] = n+f1*exp(-beta0_hat)
    }
}
}

```

```

Nhat.Chao[l] <- n+((f1^2)/(2*f2))
Nhat.LB[l] <- n+(f1^2/(2*f2))*(3*f1*f3/(2*f2^2))
Nhat.MLE[l] <- n/(1-exp(-lambdahatMLE))
Nhat.ZCMP[l] <- Nhat.Z(N,lambda,nu,Y)
Nhat.YCMP[l] <- Nhat(N,lambda,nu,Y)
}#end loop l

res[,1] <- Nhat.Chao
res[,2] <- Nhat.LB
res[,3] <- Nhat.LCMP
res[,4] <- Nhat.MLE
res[,5] <- Nhat.ZCMP
res[,6] <- Nhat.YCMP
return(res)
}

Nes <- perform_cmp(N,lambda,nu,B)
mean(Nes[,1]) #Nhat.Chao
mean(Nes[,2]) #Nhat.LB
mean(Nes[,3]) #Nhat.LCMP
mean(Nes[,4]) #Nhat.MLE
mean(Nes[,5]) #Nhat.ZCMP
mean(Nes[,6]) #Nhat.YCMP

```



```
Chao <- mean(Nes[,1]) #Nhat.Chao
```

```
LB <- mean(Nes[,2]) #Nhat.LB
```

```
LCMP <- mean(Nes[,3]) #Nhat.LCMP
```

```
MLE <- mean(Nes[,4]) #Nhat.MLE
```

```
ZCMP <- mean(Nes[,5]) #Nhat.ZCMP
```

```
YCMP <- mean(Nes[,6]) #Nhat.YCMP
```

```
Rbias.Chao <- (Chao-N)/N
```

```
Rbias.LB <- (LB-N)/N
```

```
Rbias.LCMP <- (LCMP-N)/N
```

```
Rbias.MLE <- (MLE-N)/N
```

```
Rbias.ZCMP <- (ZCMP-N)/N
```

```
Rbias.YCMP <- (YCMP-N)/N
```

```
RVar.Chao <- mean((Nes[,1]-Chao)^2,na.rm = TRUE)/N^2
```

```
RVar.LB <- mean((Nes[,2]-LB)^2,na.rm = TRUE)/N^2
```

```
RVar.LCMP <- mean((Nes[,3]-LCMP)^2,na.rm = TRUE)/N^2
```

```
RVar.MLE <- mean((Nes[,4]-MLE)^2,na.rm = TRUE)/N^2
```

```
RVar.ZCMP <- mean((Nes[,5]-ZCMP)^2,na.rm = TRUE)/N^2
```

```
RVar.YCMP <- mean((Nes[,6]-YCMP)^2,na.rm = TRUE)/N^2
```

```
RMSE.Chao <- mean((Nes[,1]-N)^2,na.rm = TRUE)/N^2
```

```
RMSE.LB <- mean((Nes[,2]-N)^2,na.rm = TRUE)/N^2
```

```
RMSE.LCMP <- mean((Nes[,3]-N)^2,na.rm = TRUE)/N^2
```

```
RMSE.MLE <- mean((Nes[,4]-N)^2,na.rm = TRUE)/N^2
```

```
RMSE.ZCMP <- mean((Nes[,5]-N)^2,na.rm = TRUE)/N^2
```

```
RMSE.YCMP <- mean((Nes[,6]-N)^2,na.rm = TRUE)/N^2
```

```
*****
```

```
#####
```

```
##For performance comparison Relative Bias, Relative Variance and Relative MSE##
```

```
#####
```

```
perform_cmp <- function(N,lambda,nu,B){
```

```
  for(l in 1:B){
```

```
    Y <- rpois(N,lambda)
```

```
  }#end loop l
```

```
}#compute Nhat Rbias RVar RMSE
```

```
perform_cmp <- function(N,lambda,nu,p,B){
```

```
  for(l in 1:B){
```

```
    Y <- rgeom(N,p)
```

```
  }#end loop l
```

```
}#compute Nhat Rbias RVar RMSE
```

```
perform_cmp <- function(N,lambda,nu,k,theta,B){
```

```
  for(l in 1:B){
```

```
    Y <- rnegbin(N,k,theta)
```

```
  }#end loop l
```

```
}#compute Nhat Rbias RVar RMSE
```

```
*****
```

ประวัติผู้เขียน

| | |
|-----------------|---|
| ชื่อ | นางสาวจันทิรา แยมสรवल |
| วันเดือนปีเกิด | 12 กุมภาพันธ์ พ.ศ. 2534 |
| ประวัติการศึกษา | สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต ภาควิชา คณิตศาสตร์และสถิติ สาขาคณิตศาสตร์ประยุกต์ คณะ วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2557 |

