รายงานผลการวิจัย

เรื่อง การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ สำหรับข้อมูลเชิงนับและประยุกต์ใช้ในผู้ป่วยโรคเบาหวานที่ต้องเข้ารับการรักษาตัวในโรงพยาบาล

A Negative Binomial – New Weighted Lindley Distribution for Count Data and its Application to Hospitalized Patients with Diabetes

ได้รับการจัดสรรงบประมาณวิจัย ประจำปี 2561

จำนวน 50,000 บาท

หัวหน้าโครงการ นางสาวศิริพร สมุทรวชิรวงษ์

งานวิจัยเสร็จสิ้นสมบูรณ์

........../............/..........

คำนำ

งานวิจัยเรื่องนี้ได้ทำการศึกษาเรื่อง การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่สำหรับข้อมูลเชิงนับและประยุกต์ใช้ในผู้ป่วยโรคเบาหวานที่ต้องเข้ารับการรักษาตัวในโรงพยาบาล ได้ทำการศึกษาตัวแบบของข้อมูลเชิงนับ ซึ่งเป็นการแจกแจงใหม่เพื่อแนวทางหนึ่งที่มีความสำคัญในการวิเคราะห์ข้อมูล ที่ไม่เป็นไปตามข้อสมมติของการแจกแจงพื้นฐาน ข้อมูลเชิงนับที่สนใจในการศึกษาครั้งนี้ คือจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาลและนำมาสร้างตัวแบบที่เหมาะสมกับข้อมูลที่มีความแปรปรวนมากกว่าค่าเฉลี่ย อันจะส่งผลต่อการวิเคราะห์ข้อมูลเชิงนับที่มีประสิทธิภาพมากขึ้น และนำมาประยุกต์ใช้ในข้อมูลเชิงนับเพื่อการอธิบายจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาล เพื่อกำหนดแนวทางในการรักษาให้มีประสิทธิภาพขึ้นได้ต่อไป ทั้งนี้หากผู้อ่านมีข้อเสนอแนะหรือข้อคิดเห็นเพิ่มเติมเกี่ยวกับเนื้อหา สามารถติดต่อได้ที่ห้องทำงาน สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้ เพื่อปรับปรุงแก้ไขให้มีความสมบูรณ์เป็นประโยชน์ให้แก่ผู้ที่ต้องการศึกษาต่อไป

อาจารย์ ดร.ศิริพร สมุทรวชิรวงษ์

สารบัญ

สารบัญตาราง

สารบัญภาพ

บทคัดย่อ

การใช้แจกแจงปัวซงในการอธิบายลักษณะข้อมูลนั้น แต่การแจกแจงดังกล่าวมีข้อสมมติของการแจกแจงคือ ค่าเฉลี่ยต้องเท่ากับค่าความแปรปรวน ซึ่งสภาพของข้อมูลเชิงนับบางครั้งเกิดสภาพที่ขัดแย้งกับข้อสมมติข้างต้น เช่น กรณีที่ค่าสังเกตมีความถี่ใด ๆ สูงผิดปกติ งานวิจัยเรื่องนี้มีวัตถุประสงค์เพื่อพัฒนาการแจกแจงใหม่ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (NB-NWL) สำหรับข้อมูลเชิงนับ เพื่อศึกษาคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL เพื่อศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด และเพื่อประยุกต์ใช้ในการวิเคราะห์ข้อมูลเชิงนับ จำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาล ผลการศึกษาพบว่าการแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ เป็นการผสมระหว่างการแจกแจงทวินามลบกับการแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่ ซึ่งจะได้เป็นการแจกแจงแบบไม่ต่อเนื่องและนำเสนอฟังก์ชันการแจกแจงน่าจะเป็น ค่าเฉลี่ย และคามแปรปรวนของการแจกแจงใหม่ เมื่อนำไปประยุกต์ใช้กับข้อมูลผู้ป่วยเบาหวานที่เข้ารับการรักษาตัวในโรงพยาบาล พบว่า การแจกแจงใหม่เหมาะสมกับข้อมูลมากกว่าการแจกแจงแบบดั้งเดิมซึ่งก็คือการแจกแจงปัวซงและการแจกแจงทวินามลบ

คำสำคัญ: การแจกแจงทวินามลบ การแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่ ข้อมูลแจงนับ

ABSTRACT

The Poisson distribution plays a central role in count data analysis. The most important characteristic of the Poisson distribution is its mean and variance must be equal. In practice, the count data often exhibit the overdispersion, which is the variance is greater than the mean. The negative binomial distribution offers a remedy to this problem. Moreover, the mixed negative binomial distribution is an alternative to count data with overdispersion. The objectives of research are to proposed new distributions for count data, namely, the negative binomial-new weighted Lindley distribution, to derive parameter estimation of the proposed distribution by using the maximum likelihood estimation and to compare efficiencies of the proposed distribution with other distributions. This research found that the negative binomial-new weighted Lindley distribution, obtained by mixing the negative binomial distribution with the new weighted Lindley distribution is another mixed negative binomial distribution that provided an appropriate fit for count data with overdispersion. Some characteristics of the proposed distribution, such as mean, variance, the coefficient of skewness and kurtosis are derived. Finally, application of the proposed distribution for the number of hospitalized patients with diabetes at Ratchaburi hospital, Thailand is presented.

Key words: negative binomial distribution, NB-NWL distribution, count data

บทนำ

ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

การศึกษาข้อมูลเชิงนับ (Count Data) เช่น จำนวนผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล จำนวนการแพร่ระบาดของไข้หวัดนกสายพันธุ์ H5N1 จำนวนอุบัติเหตุบนท้องถนน เป็นต้น โดยทั่วไปมักใช้การแจกแจงปัวซง (Poisson) ในการอธิบายลักษณะข้อมูลนั้น แต่การแจกแจงดังกล่าวมีข้อสมมติ (Assumption) ของการแจกแจงคือ ค่าเฉลี่ย (Mean) ต้องเท่ากับค่าความแปรปรวน (Variance) ซึ่งสภาพของข้อมูลเชิงนับบางครั้งเกิดสภาพที่ขัดแย้งกับข้อสมมติข้างต้น เช่น กรณีที่ค่าสังเกตมีความถี่ใด ๆ สูงผิดปกติ เป็นต้น ต่อมาในปี ค.ศ. 1920 Greenwood และ Yule ได้นำเสนอการแจกแจงทวินามลบ (Negative Binomial: NB Distribution) ซึ่งสามารถใช้อธิบายชุดข้อมูลที่มีค่าความแปรปรวนมากกว่าค่าเฉลี่ยได้ดีกว่าการแจกแจงปัวซง โดยพัฒนาการจากการผสมระหว่างการแจกแจงปัวซงกับการแจกแจงแกมม่า (Gamma Distribution) (Greenwood and Yule, 1920)

การสร้างการแจกแจงใหม่ (A New Distribution) จึงเป็นแนวทางหนึ่งที่มีความสำคัญในการวิเคราะห์ข้อมูล ที่ไม่เป็นไปตามข้อสมมติของการแจกแจงพื้นฐานได้ซึ่งการแจกแจงผสม (A Mixed Distribution) เป็นวิธีการหนึ่งที่นำมาซึ่งการแจกแจงใหม่ ดังที่มีผู้วิจัยหลายท่าน (Panjer and Willmot 1981, G´omez *et al.* 2008, Zamani and Ismail 2010, Saengthong and Bodhisuwan 2013) ได้เสนอการแจกแจงทวินามลบผสม (A Mixed Negative Binomial) เพื่อแก้ไขปัญหาข้อมูลที่มีความแปรปรวนมากกว่าค่าเฉลี่ยข้อมูลเชิงนับที่สนใจในการศึกษาครั้งนี้ คือจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาล โรคเบาหวาน คือโรคที่เซลร่างกายมีความผิดปรกติในขบวนการเปลี่ยนน้ำตาลในเลือดให้เป็นพลังงาน เมื่อน้ำตาลไม่ได้ถูกใช้จึงทำให้ระดับน้ำตาลในเลือดสูงขึ้นกว่าระดับผิดปรกติ ในปัจจุบันหลายประเทศใช้เกณฑ์ระดับน้ำตาลที่มากกว่า 126 มก./ดล. โดยมีข้อแม้ว่าเป็นค่าของน้ำตาลในน้ำเลือดหลังจากอดอาหารมาอย่างน้อย 8 ชม. จากข้อมูลรายงานขององค์การอนามัยโลก (The World Health Organization; WHO) ปี 2560 พบว่าโรคเบาหวานเป็นสาเหตุการเสียชีวิตอันดับที่ 9 ของผู้หญิงทั่วโลก คิดเป็นอัตรา 2.1 ล้านคนต่อปี โดยสภาวะเศรษฐกิจและสังคม เฉพาะอย่างยิ่งในประเทศที่กำลังพัฒนา ทำให้เด็กหญิงและมารดาที่เป็นเบาหวานมีอุปสรรคในการเข้าถึงวิธีป้องกันโรคเบาหวาน ความไม่เท่าเทียมกันทางเศรษฐกิจและสังคมทำให้ผู้หญิงมีความเสี่ยงต่อโรคเบาหวาน เนื่องจากอาหารและโภชนาการที่ไม่ดี การไม่มีกิจกรรมทางกาย การสูบบุหรี่และการดื่ม เครื่องดื่มแอลกอฮอล์จนก่อให้เกิดอันตราย (สมาคม

โรคเบาหวานแห่งประเทศไทยฯ, 2560) การรักษาในโรงพยาบาลจากโรคผู้ป่วยเบาหวาน หมายถึง การที่ผู้ป่วยเบาหวานต้องรับไว้รักษาในโรงพยาบาลเนื่องจากภาวะฉุกเฉินที่มีระดับน้ำตาลในเลือดสูง (Hyperglycemia) และภาวะฉุกเฉินที่มีระดับน้ำตาลในเลือดต่ำ (Hypoglycemia) ยกเว้นผู้ป่วยเบาหวานที่รับส่งต่อมาจากโรงพยาบาล

ทั้งนี้จะทำการเก็บรวบรวมข้อมูลผู้ป่วยโรคเบาหวานในโรงพยาบาลสันทราย อำเภอสันทราย จังหวัดเชียงใหม่ อำเภอสันทราย เป็นอำเภอหนึ่งในเขตปริมณฑลของนครเชียงใหม่ ที่มีการขยายตัวอย่างรวดเร็ว เป็นอำเภอที่มีความเจริญรองจากนครเชียงใหม่ จนปัจจุบันอำเภอสันทราย มีสถานะอำเภอขนาดใหญ่มีจำนวน ประชากรเป็นอันดับ 2 ของจังหวัดเชียงใหม่ รองจากอำเภอเมืองเชียงใหม่ มีสถานประกอบการจำนวนมากและการขยายตัวของชุมชนเมืองแม่โจ้ ทำให้อำเภอสันทราย มีความพร้อมทุก ๆ ด้าน ทั้งสถานศึกษา มหาวิทยาลัย สถานประกอบการ โรงงาน บ้านจัดสรร ฯลฯ และมีทรัพยากรทางด้านการดูแลสุขภาพ ที่เป็นโรงพยาบาลของรัฐ จำนวนเพียง 1 แห่ง จึงทำให้มีผู้ป่วยเป็นจำนวนมาก รวมถึงผู้ป่วยโรคเบาหวานรายใหม่ ค่อนข้างสูงกว่าเป้าหมายที่กำหนดไว้เนื่องจาก ในบริบทพื้นที่อำเภอสันทรายทำงานค่อนข้างยาก เป็นหมู่บ้านจัดสรร โรงงานอุตสาหกรรม ชุมชนเมืองค่อนข้างมาก และพื้นที่อำเภอสันทราย รวมถึงมหาวิทยาลัยแม่โจ้ มีข้าวกล้อง พืชผักปลอดสารพิษค่อนข้างมาก แต่ยังพบผู้ป่วยรายใหม่ค่อนข้างสูง จึงทำให้เกิดมาตรการเรื่อง การกินผักลดเค็ม ขึ้น (สรุปผลการนิเทศโรงพยาบาลสันทรายประจำปี, 2560) และข้อมูลผู้ป่วยโรคเบาหวานในโรงพยาบาลราชบุรีซึ่งเป็นปัญหาสาธารณสุขที่สำคัญของจังหวัดราชบุรี โดยทางโรงพยาบาลราชบุรีมีนโยบายแผนการทำงานด้านการดูแลสุขภาพของทีมสุขภาพซึ่งพบว่าผู้ป่วยเบาหวานที่เข้ารับการรักษาแผนกผู้ป่วยนอกของโรงพยาบาลราชบุรีมีจำนวนมากขึ้น ตั้งแต่ปี พ.ศ. 2550 ถึง ปีพ.ศ. 2551 จาก 5,656 ราย เป็น 6,270 ราย และพบว่า มีจำนวนผู้ป่วยเบาหวานที่เข้ารับการลงทะเบียนมากขึ้นในปีพ.ศ. 2550 ถึง 2552 จาก 3,557 ราย เป็ น 4,176 ราย และ 4,490 ราย ตามลำดับ และมีความชุกจำนวนผู้ป่วยโรคเบาหวานสูงมีแนวโน้มเพิ่มขึ้นร้อยละ 40 ของผู้ป่วยโรคเบาหวานที่ควบคุมได้ (การตรวจราชการและนิเทศงานกระทรวงสาธารณสุขโรงพยาบาลราชบุรี ประจำปีงบประมาณ 2561)

ดังนั้นการพัฒนาการแจกแจงใหม่ที่จะช่วยในการแก้ปัญหาข้อมูลเชิงนับที่มีความแปรปรวนมากกว่าค่าเฉลี่ยของข้อมูลดังกล่าว จึงควรมีการศึกษาวิจัย เพื่อสร้างตัวแบบที่เหมาะสมกับข้อมูลที่มีความแปรปรวนมากกว่าค่าเฉลี่ย อันจะส่งผลต่อการวิเคราะห์ข้อมูลเชิงนับที่มีประสิทธิภาพมากขึ้น และนำมาประยุกต์ใช้ในข้อมูลเชิงนับเพื่อการอธิบายจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาล เพื่อกำหนดแนวทางในการรักษาให้มีประสิทธิภาพขึ้นได้ต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาการแจกแจงใหม่ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (A Negative Binomial-New Weighted Lindley Distribution; NB-NWL) และ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ที่มีศูนย์มาก (Hurdel Negative Binomial-New Weighted Lindley Distribution; HNB-NWL) สำหรับข้อมูลเชิงนับ

2. เพื่อศึกษาคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL และ HNB-NWL

3. เพื่อศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL และ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimate: MLE)

4. เพื่อประยุกต์ใช้ในการวิเคราะห์ข้อมูลเชิงนับ จำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาในโรงพยาบาล

ขอบเขตของการวิจัย

1. พัฒนาการแจกแจงใหม่ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (A Negative Binomial-new weighted Lindley Distribution) และ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ที่มีศูนย์มาก (Hurdel Negative Binomial-New Weighted Lindley Distribution; HNB-NWL) สำหรับข้อมูลเชิงนับ

2. ศึกษาคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL และ HNB-NWL

3. ศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL และ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate: MLE)

4. ประยุกต์ใช้ในการวิเคราะห์ข้อมูลเชิงนับ

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้การแจกแจงใหม่ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (A Negative Binomial-new weighted Lindley Distribution) และ การแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ที่มีศูนย์มาก (Hurdel Negative Binomial-New Weighted Lindley Distribution; HNB-NWL) สำหรับข้อมูลเชิงนับ

2. ทราบคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL และ HNB-NWL เพื่อหาคุณลักษณะบางประการของข้อมูลที่ศึกษา

3. ประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWLและ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุดได้

4. นำการแจกแจง NB-NWL และ HNB-NWL ประยุกต์ใช้ในการวิเคราะห์ข้อมูลเชิงนับ

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ตัวแบบที่ใช้อธิบายการแจกแจงสำหรับข้อมูลเชิงนับ โดยทั่วไปมักใช้การแจกแจงปัวซง ทวินาม หรือทวินามลบ แต่บางครั้งพบว่าการแจกแจงดังกล่าวมีประสิทธิภาพต่ำ อีกทั้งยังไม่เป็นไปตามข้อสมมติของการแจกแจง เช่น การแจกแจงปัวซงจะมีข้อสมมติว่าค่าเฉลี่ยต้องเท่ากับค่าความแปรปรวน สภาพของข้อมูลเชิงนับบางครั้งเกิดสภาพที่ขัดแย้งกับข้อสมมติข้างต้น เช่น กรณีที่ค่าสังเกตมีความถี่สูงผิดปกติ ดังนั้นการแจกแจงทวินามลบจึงถูกนำมาใช้กับข้อมูลในลักษณะนี้ แต่อย่างไรก็ตามการแจกแจงทวินามลบอาจไม่เหมาะกับข้อมูล Heavy Tailed ดังนั้นการแจกแจงทวินามลบผสม (Mixed NB) จึงถูกนำเสนอสำหรับข้อมูลที่เกิดปัญหาดังกล่าว และเมื่อเกิดปัญหาศูนย์มากจากการแจกแจงทวินามลบผสม สามารถพัฒนาเป็นการแจกแจงเป็นรูปแบบจำลองที่ถูกดัดแปลง โดยกำหนดกระบวนการประมาณให้ค่าศูนย์และจำนวนบวกมีความแตกต่างกัน เหมาะสมสำหรับการวิเคราะห์ตัวแปรตามที่มีค่าศูนย์และจำนวนบวกที่มีการแจกแจงแบบทวินาม (Binomial Probability) ซึ่งมีความคล้ายคลึงกับรูปด้วยแบบจำลองโทบิท (Tobit Model)

1. ตัวแบบสำหรับข้อมูลเชิงนับ

ข้อมูลเชิงนับ (Count Data) เป็นข้อมูลที่ได้จากการใช้วิธีการนับ ลักษณะของข้อมูลที่ได้จากการนับนี้ โดยทั่วไปจะเป็นตัวเลขจำนวนเต็ม เช่น จำนวนผู้เข้ามาใช้บริการ ณ โรงพยาบาลในแต่ละวัน จำนวนลูกค้าที่มาซื้อสินค้าในร้านค้าสะดวกซื้อ สาขาแม่โจ้ เป็นต้น

ถ้าตัวแปรสุ่ม X เป็นข้อมูลเชิงนับ การแจกแจงความน่าจะเป็นของแต่ละค่าของตัวแปรสุ่ม X บางครั้งสามารถเขียนแทนด้วย $f(x)$ ซึ่งจะมีลักษณะแตกต่างกันออกไป ทั้งนี้ขึ้นอยู่กับแต่ละตัวแปรสุ่ม X ตัวอย่างการแจกแจงความน่าจะเป็นสำหรับข้อมูลเชิงนับ มีดังนี้

1.1 การแจกแจงทวินาม

การแจกแจงทวินาม (Binomial Distribution) จะได้จากการทดลองที่เป็นแบบเบอร์นูลลีต่างกันที่การแจกจงแบบเบอร์นูลลีทำการทดลองเพียงครั้งเดียว แต่การแจกแจงทวินามการทดลองจะซ้ำๆ กัน n ครั้ง ทั้งนี้การทดลองแต่ละครั้งเป็นอิสระต่อกัน

บทนิยาม 1 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงทวินาม เขียนแทนด้วย $X \sim \mathrm{B}(n, p)$ จะได้ฟังก์ชันความน่าจะเป็นของการแจกแจงทวินาม ดังนี้

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad , x = 0,1,2,\ldots,n \text{ สำหรับ } 0 < p < 1$$

ทฤษฎีบท 1 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงทวินามแล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$\mathrm{E}(X) = np \text{ และ } \mathrm{Var}(X) = np(1-p)$$

พิสูจน์

ค่าเฉลี่ย

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{\infty} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{\infty} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= n \sum_{x=1}^{\infty} \binom{n-1}{x-1} p^x (1-p)^{n-x}
\end{aligned}$$

แทน $y = x - 1$

ดังนั้น
$$\begin{aligned}
\mathrm{E}(X) &= n \sum_{y=0}^{\infty} \binom{n}{y} p^{y+1} (1-p)^{n-y-1} \\
&= np \sum_{y=0}^{\infty} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \\
&= np
\end{aligned}$$

ความแปรปรวน

พิจารณา $\mu_2'(X) = \mathrm{E}(X(X-1)) = \mathrm{E}(X^2 - X)$

$$\mathrm{E}(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)\binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=2}^{\infty} x(x-1)\binom{n}{x} p^x (1-p)^{n-x}$$

$$= n(n-1)\sum_{x=2}^{\infty} \binom{n-2}{x-2} p^x (1-p)^{n-x}$$

แทน $y = x - 2$

$$\mathrm{E}(X(X-1)) = n(n-1)\sum_{y=0}^{\infty} \binom{n-2}{y} p^{y+2} (1-p)^{n-y-2}$$

$$= n(n-1)p^2\sum_{y=0}^{\infty} \binom{n-2}{y} p^{y} (1-p)^{(n-2)-y}$$

$$= n(n-1)p^2$$

จาก $\quad \mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$

$$= n(n-1)p^2 + np$$

$$= n^2p^2 + np(1-p).$$

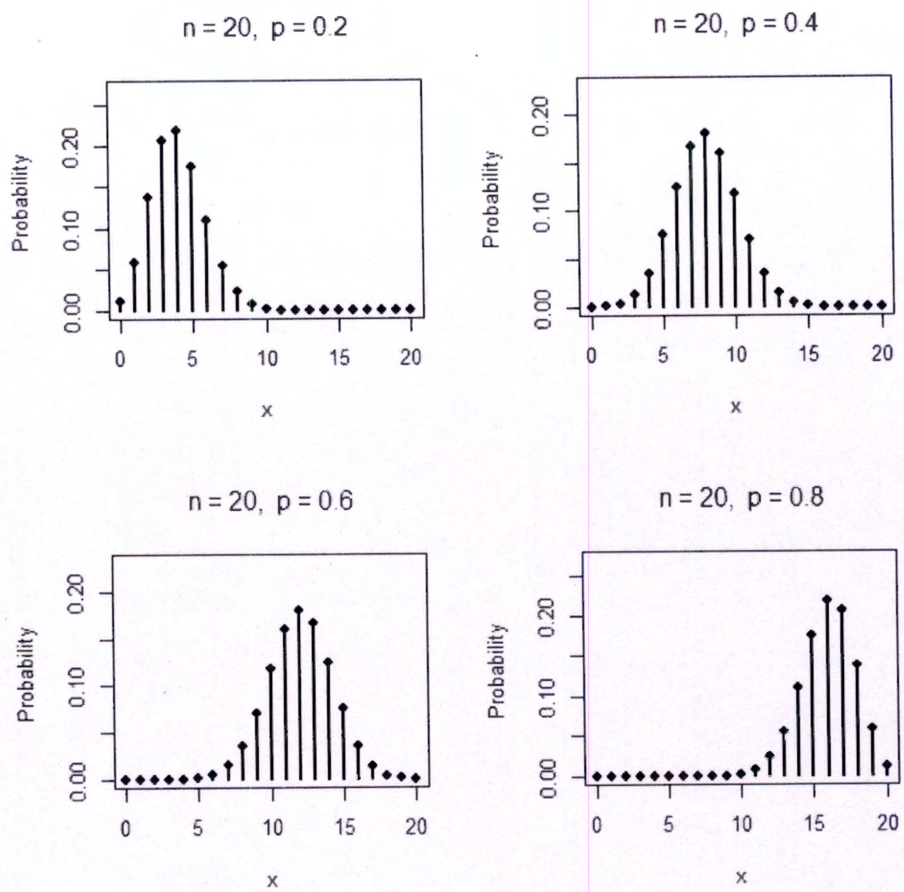
จาก $\quad \mathrm{Var}(X) = np(1-p) + n^2p^2 - (np)^2$

$$= np(1-p)$$

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจงทวินามในโปรแกรม R

```
library("stats")
x<-seq (0,100,1)
r<- k1
c<- k2
k<- k3
dnbel<-function(x,r,k,c)
{
 f=function(lam)
 {
  f<-(exp(-lam*r))*((1-exp(-lam))^x)*(lam^(k-1))*(exp(-c*lam))
 }
 fun=integrate(f,0,Inf)
 fun_lam=fun$value
 p=(((factorial(r+x-1)/(factorial(r-1)*factorial(x)))*(c^k))/((factorial(k-1))))*fun_lam
}
out.dist=c()
for(ii in 1 : length(x)){
 out.dist[ii]<-dnbel(x[ii],r,k,c)
}
sum(out.dist)
out.dist
```

รูปภาพที่ 1 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจงทวินาม สำหรับค่าพารามิเตอร์ n และ p ที่กำหนดไว้



รูปภาพที่ 1 ฟังก์ชันความน่าจะเป็นของการแจกแจงทวินาม

1.2 การแจกแจงปัวซง

ในการทดลองหนึ่ง ๆ ถ้าให้ X เป็นจำนวนของเหตุการณ์ต่างๆ ที่เกิดขึ้นในช่วงเวลาใดเวลาหนึ่งหรือในอาณาบริเวณหนึ่งๆ ถ้า λ เป็นค่าเฉลี่ยของจำนวนเหตุการณ์ที่เกิดขึ้นในช่วงเวลาที่จำกัด หรือในอาณาบริเวณที่จำกัดดังกล่ าวนั้น จะได้ว่าตัวแปรสุ่ม X เป็นตัวแปรสุ่มโดยที่มีการแจกแจง ความน่าจะเป็นปัวซง (Poisson Distribution)

บทนิยาม 2 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซง เขียนแทนด้วย $X \sim \text{Poisson}(\lambda)$ จะได้ฟังก์ชันความน่าจะเป็นของการแจกแจง ดังนี้

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad , \ x = 0, 1, 2, \ldots \ \text{สำหรับ} \ \lambda > 0$$

ทฤษฎีบท 2 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซงแล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$\text{E}(X) = \lambda \ \text{และ} \ \text{Var}(X) = \lambda$$

พิสูจน์

ค่าเฉลี่ย

$$\begin{aligned}\text{E}(X) &= \sum_{x=0}^{\infty} x\frac{\exp(-\lambda)\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x\frac{\exp(-\lambda)\lambda^x}{x!} \\ &= \lambda\sum_{x=1}^{\infty} \frac{\exp(-\lambda)\lambda^{x-1}}{(x-1)!}.\end{aligned}$$

แทน $y = x - 1$

ดังนั้น $\text{E}(X) = \lambda\sum_{y=0}^{\infty} \frac{\exp(-\lambda)\lambda^y}{y!}$

$$= \lambda$$

ความแปรปรวน

พิจารณา $\mu_2'(X) = \text{E}(X(X-1)) = \text{E}(X^2 - X)$

$$\text{E}(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)\frac{\exp(-\lambda)\lambda^x}{x!}$$

$$= \sum_{x=2}^{\infty} x(x-1)\frac{\exp(-\lambda)\lambda^{x}}{x!}$$

$$= \lambda^2 \sum_{x=2}^{\infty} \frac{\exp(-\lambda)\lambda^{x-2}}{(x-2)!}$$

แทน $y = x - 2$

$$\mathrm{E}(X(X-1)) = \lambda^2 \sum_{y=0}^{\infty} \frac{\exp(-\lambda)\lambda^{y}}{y!}$$

$$= \lambda^2$$

จาก $\quad \mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$

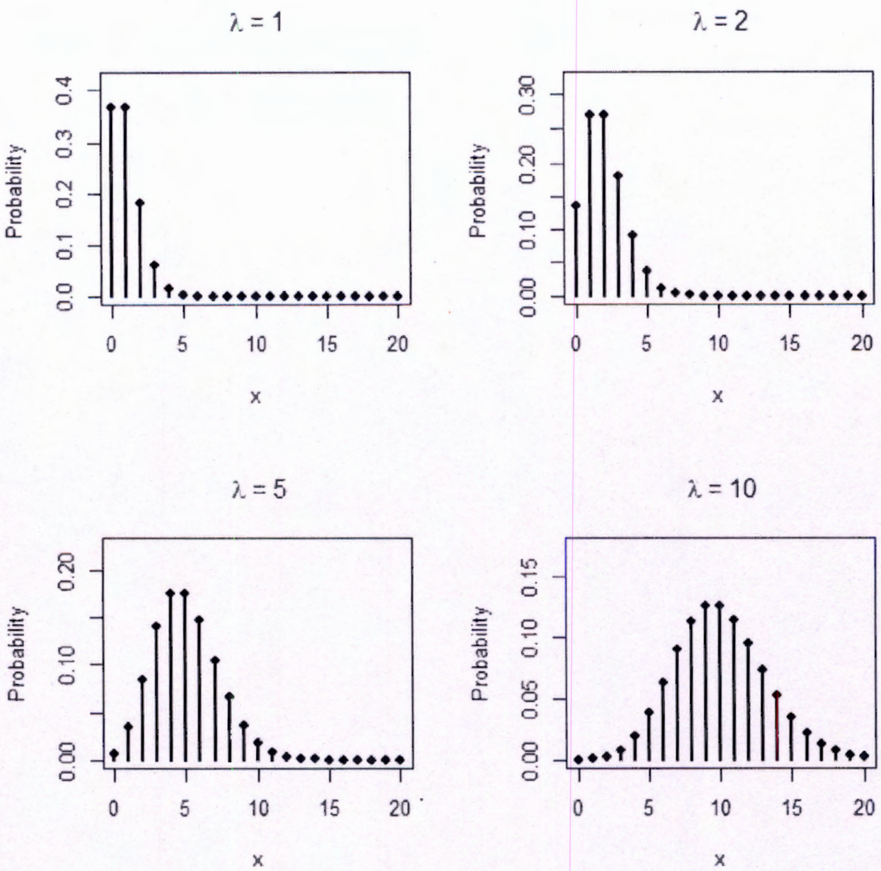
$$= \lambda^2 + \lambda.$$

ดังนั้น $\quad \mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจงปัวซงในโปรแกรม R

```
library("stats")
x<-seq(0,30,1)
l<- k
poi <-function(x,l)
{
  p<-((exp(-l))*(l^x))/factorial(x)
}
out.dist=c()
for(ii in 1 : length(x)){
  out.dist[ii]<-poi(x[ii],l)
}
sum(out.dist)
out.dist
```

รูปภาพที่ 2 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจงปัวซง สำหรับค่าพารามิเตอร์ λ ที่กำหนดไว้



รูปภาพที่ 2 ฟังก์ชันความน่าจะเป็นของการแจกแจงปัวซง

1.3 การแจกแจงทวินามลบ

ในการทดลองซ้ำ ๆ กัน โดยที่การทดลองแต่ละครั้งมีสิ่งที่เรียกว่าความไม่สำเร็จ (Failure) กับความสำเร็จ (Success) นั้น ถ้าให้ X เป็นจำนวนครั้งที่ทดลองแล้วได้ความสำเร็จเป็นครั้งที่ r ทั้งนี้ p คือความน่าจะเป็นของความสำเร็จในแต่ละครั้ง จะได้ว่า X เป็นตัวแปรสุ่มโดยที่มีการแจกแจงความน่าจะเป็นทวินามลบ (Negative Binomial Distribution) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

บทนิยาม 3 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงทวินามลบ เขียนแทนด้วย $X \sim \mathrm{NB}(r,p)$ จะได้ฟังก์ชันความน่าจะเป็นของการแจกแจง ดังนี้

$$f(x)=\binom{r+x-1}{x}p^r(1-p)^x \quad , x=0,1,2,\ldots \text{ สำหรับ } r>0 \text{ และ } 0<p<1$$

ทฤษฎีบท 3 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงทวินามลบ แล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$\mathrm{E}(X)=r\left(\frac{1-p}{p}\right) \text{และ } \mathrm{Var}(X)=r\left(\frac{1-p}{p^2}\right).$$

พิสูจน์

ค่าเฉลี่ย

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{\infty} x\binom{r+x-1}{x}p^r(1-p)^x \\
&= \sum_{x=1}^{\infty} x\binom{r+x-1}{x}p^r(1-p)^x \\
&= r\sum_{x=1}^{\infty} \binom{r+x-1}{x-1}p^r(1-p)^x
\end{aligned}$$

แทน $y=x-1$

ดังนั้น
$$\begin{aligned}
\mathrm{E}(X) &= r\sum_{y=0}^{\infty}\binom{r+y}{y}p^r(1-p)^{y+1} \\
&= \frac{r(1-p)}{p}\sum_{y=0}^{\infty}\binom{(r+1)+y-1}{y}p^{r+1}(1-p)^y \\
&= \frac{r(1-p)}{p}
\end{aligned}$$

ความแปรปรวน

พิจารณา $\mu_2'(X)=\mathrm{E}(X(X-1))=\mathrm{E}(X^2-X)$

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1)\binom{r+x-1}{x}p^r(1-p)^x \\
&= \sum_{x=2}^{\infty} x(x-1)\binom{r+x-1}{x}p^r(1-p)^x \\
&= r(r+1)\sum_{x=2}^{\infty}\binom{r+x-1}{x-2}p^r(1-p)^x
\end{aligned}$$

แทน $y = x-2$

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= r(r+1)\sum_{y=0}^{\infty}\binom{r+y+1}{y}p^r(1-p)^{y+2} \\
&= \frac{r(r+1)(1-p)^2}{p^2}\sum_{y=0}^{\infty}\binom{(r+2)+y-1}{y}p^{r+2}(1-p)^y \\
&= \frac{r(r+1)(1-p)^2}{p^2}
\end{aligned}$$

จาก $\quad \mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$

$$= r\left[(r+1)\frac{(1-p)^2}{p^2} + \frac{(1-p)}{p}\right].$$

ดังนั้น $\mathrm{Var}(X) = \dfrac{r(r+1)(1-p)^2}{p^2} + \dfrac{r(1-p)}{p} - \left(\dfrac{r(1-p)}{p}\right)^2$

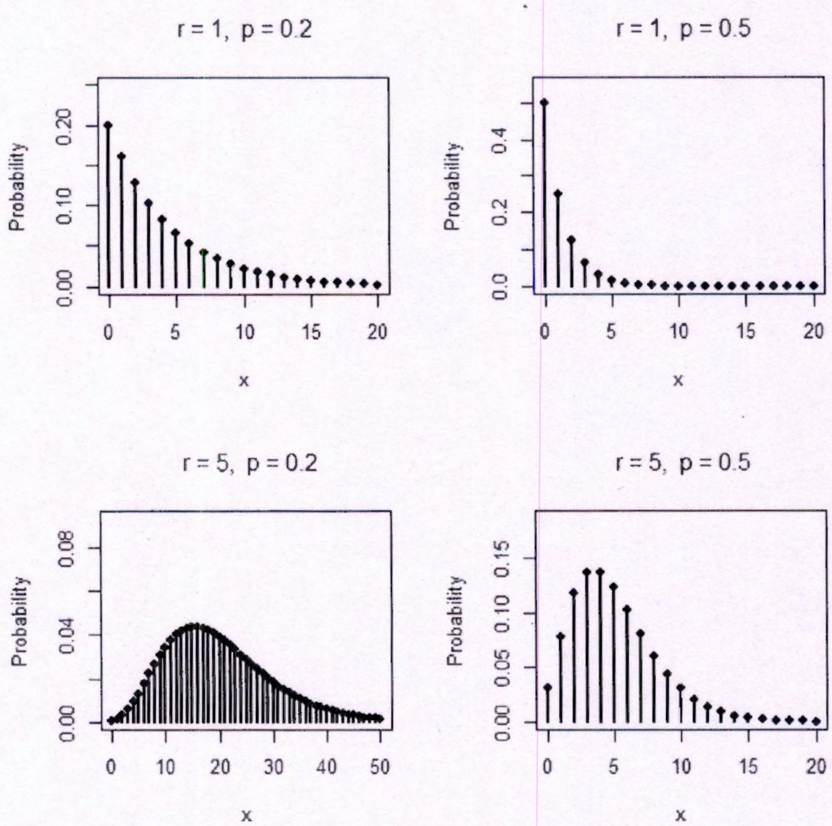
$$= \frac{r(1-p)}{p^2}$$

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจงปัวซงในโปรแกรม R

```
library("stats")
x<-seq (0,100,1)
r<- k1
c<- k2
k<- k3
dnbel<-function(x,r,k,c)
{
 f=function(lam)
```

```
 {f<-(exp(-lam*r))*((1-exp(-lam))^x)*(lam^(k-1))*(exp(-c*lam))}
 fun=integrate(f,0,Inf)
 fun_lam=fun$value
 p=(((factorial(r+x-1)/(factorial(r-1)*factorial(x)))*(c^k))/((factorial(k-1))))*fun_lam
}
out.dist=c()
for(ii in 1 : length(x)){out.dist[ii]<-dnbel(x[ii],r,k,c)}
sum(out.dist)
out.dist
```

รูปภาพที่ 3 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามลบ สำหรับค่าพารามิเตอร์ r และ p บางค่า จะเห็นว่าเมื่อเรากำหนดให้ p คงที่ เมื่อ r เปลี่ยนไป การแจกแจงจะมีรูปร่างเปลี่ยนไป ดังนั้น r เป็นพารามิเตอร์ shape



รูปภาพที่ 3 ฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามลบ

1.4 การแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่

การแจกแจงลินด์เลย์ (Lindley Distribution) เป็นที่รู้จักกันและใช้งานกันอย่างกว้างขวางในการวิเคราะห์ความน่าเชื่อถือ (Reliability Analysis) นำเสนอโดย Ghitany และคณะในปี ค.ศ. 2008 ต่อมามีการพัฒนา การแจกแจงลินด์เลย์ถ่วงน้ำหนัก (weighted Lindley distribution) สำหรับการจำลองข้อมูลการอยู่รอด (Modelling Survival Data) และล่าสุด Asgharzadeha และคณะได้นำเสนอ การแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่ (New Weighted Lindley Distribution) ในปี ค.ศ. 2006 สำหรับประยุกต์ใช้ในข้อมูลอายุการใช้งาน (Lifetime Data)

บทนิยาม 4 กำหนดให้ตัวแปรสุ่ม X มีการแจกแจงแบบลินด์เลย์ถ่วงน้ำหนักใหม่ (NWL) ด้วยพารามิเตอร์ 2 ตัวได้แก่ $\alpha > 0$ และ $\beta > 0$ มีฟังก์ชันความน่าจะเป็นของการแจกแจงคือ

$$f(x) = \frac{\beta^2(1+\alpha)^2}{\alpha\beta(1+\alpha)+\alpha(2+\alpha)}(1+x)(1-\exp(-\alpha\beta x))\exp(-\beta x), \text{ สำหรับ } x > 0.$$

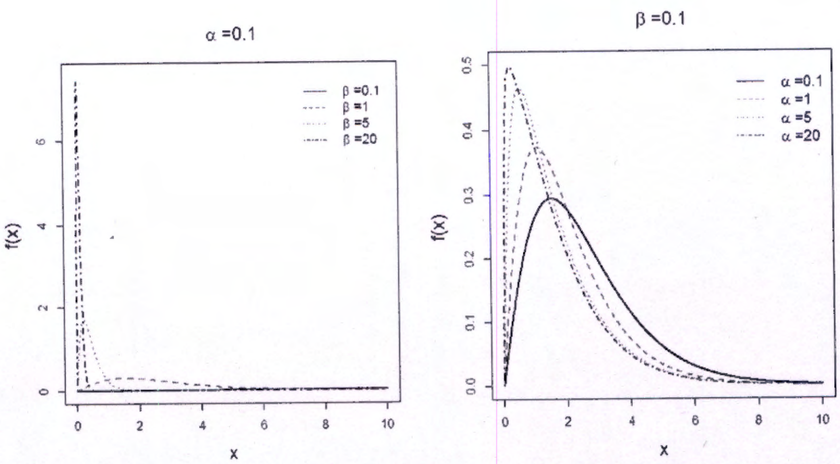
บทนิยาม 5 กำหนดให้ตัวแปรสุ่ม X มีการแจกแจงแบบลินด์เลย์ถ่วงน้ำหนักใหม่ (NWL) มีฟังก์ชันก่อกำเนิดโมเมนต์คือ

$$M_X(t) = \frac{\beta^2(1+\alpha)^2}{\alpha\beta(1+\alpha)+\alpha(2+\alpha)}\left\{\frac{\beta-t+1}{(\beta-t)^2} - \frac{\beta(1+\alpha)-t+1}{[\beta(1+\alpha)-t]^2}\right\}.$$

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่ในโปรแกรม R

```
x <- seq(0,20,0.001)
a <- k1
b <- k2
f <- (((b^2)*(1+a)^2)/(a*b*(1+a)+a*(2+a)))*((1+x)*(1-exp(-a*b*x))*exp(-b*x))
```

รูปภาพที่ 4 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่ สำหรับค่าพารามิเตอร์ α และ β ที่กำหนดไว้



รูปภาพที่ 4 ฟังก์ชันความน่าจะเป็นของการแจกแจงลินด์เลย์ถ่วงน้ำหนักใหม่

2. ตัวแบบสำหรับข้อมูลเชิงนับที่มีศูนย์มาก

ข้อมูลเชิงนับที่มีศูนย์มาก (Count Data with Extra Zeros) เป็นข้อมูลที่ได้จากการใช้วิธีการนับจำนวนครั้งของการเกิดเหตุการณ์ที่เราสนใจมีค่าเท่ากับ 0 ปริมาณมากหรือความถี่ที่ไม่เกิดเหตุการณ์ที่เราสนใจนั้นมีค่าสูง เช่น จำนวนครั้งของการเรียกสินไหม (Claim) ของแต่ละกรมธรรม์ที่สนใจ จำนวนครั้งของการเกิดแผนดินไหวในหน่วยพื้นที่หนึ่ง ๆ เป็นต้น ตัวอย่างการแจกแจงความน่าจะเป็นสำหรับข้อมูลเชิงนับที่มีศูนย์มาก มีดังนี้

2.1 การแจกแจง Zero Inflated Poisson Distribution

การแจกแจง ZIP (zero inflated Poisson Distribution) นำเสนอโดย Lambert (1992) โดยการผสมระหว่างการแจกแจงเบอร์นูลลีและการแจกแจงปัวซง เพื่อเป็นทางเลือกสำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีศูนย์มาก โดยมีฟังก์ชันความน่าจะเป็นดังนี้

บทนิยาม 6 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจง ZIP (zero inflated Poisson) เขียนแทนด้วย $X \sim \text{ZIP}(\lambda, \phi)$ จะได้ฟังก์ชันความน่าจะเป็นของการแจกแจง (Lambert, 1992) ดังนี้

$$f(x) = \begin{cases} \phi + (1-\phi)e^{-\lambda} & ; x = 0 \\ (1-\phi)\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x = 1, 2, \ldots \end{cases} \qquad \text{สำหรับ } \lambda > 0 \text{ และ } 0 < \phi < 1$$

ทฤษฎีบท 4 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจง ZIP แล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$\mathrm{E}(X) = (1-\phi)\lambda \text{ และ } \mathrm{Var}(X) = (1-\phi)\left(\phi\lambda^2 + \lambda\right)$$

พิสูจน์

ค่าเฉลี่ย

$$\begin{aligned}\mathrm{E}(X) &= (0)\left[\phi + (1-\phi)\exp(\lambda)\right] + \sum_{x=1}^{\infty} x(1-\phi)\frac{\exp(-\lambda)\lambda^x}{x!} \\ &= (1-\phi)\sum_{x=1}^{\infty}\frac{\exp(-\lambda)\lambda^x}{(x-1)!} \\ &= (1-\phi)\lambda\sum_{x=1}^{\infty}\frac{\exp(-\lambda)\lambda^{x-1}}{(x-1)!}\end{aligned}$$

แทน $y = x - 1$

ดังนั้น $$\begin{aligned}\mathrm{E}(X) &= (1-\phi)\lambda\sum_{y=0}^{\infty}\frac{\exp(-\lambda)\lambda^y}{y!} \\ &= (1-\phi)\lambda\end{aligned}$$

ความแปรปรวน

พิจารณา $\mu_2'(X) = \mathrm{E}(X(X-1)) = \mathrm{E}(X^2 - X)$

$$\begin{aligned}\mathrm{E}(X(X-1)) &= \sum_{x=1}^{\infty} x(x-1)(1-\phi)\frac{\exp(-\lambda)\lambda^x}{x!} \\ &= (1-\phi)\sum_{x=2}^{\infty} x(x-1)\frac{\exp(-\lambda)\lambda^x}{x!} \\ &= (1-\phi)\lambda^2\sum_{x=2}^{\infty}\frac{\exp(-\lambda)\lambda^{x-2}}{(x-2)!}.\end{aligned}$$

แทน $y = x - 2$

$$\mathrm{E}(X(X-1)) = (1-\phi)\lambda^2\sum_{y=0}^{\infty}\frac{\exp(-\lambda)\lambda^y}{y!}$$

$$= (1-\phi)\lambda^2$$

จาก $$\mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$$

$$= (1-\phi)\lambda^2 + (1-\phi)\lambda.$$

$$= (1-\phi)\left(\lambda^2+\lambda\right).$$

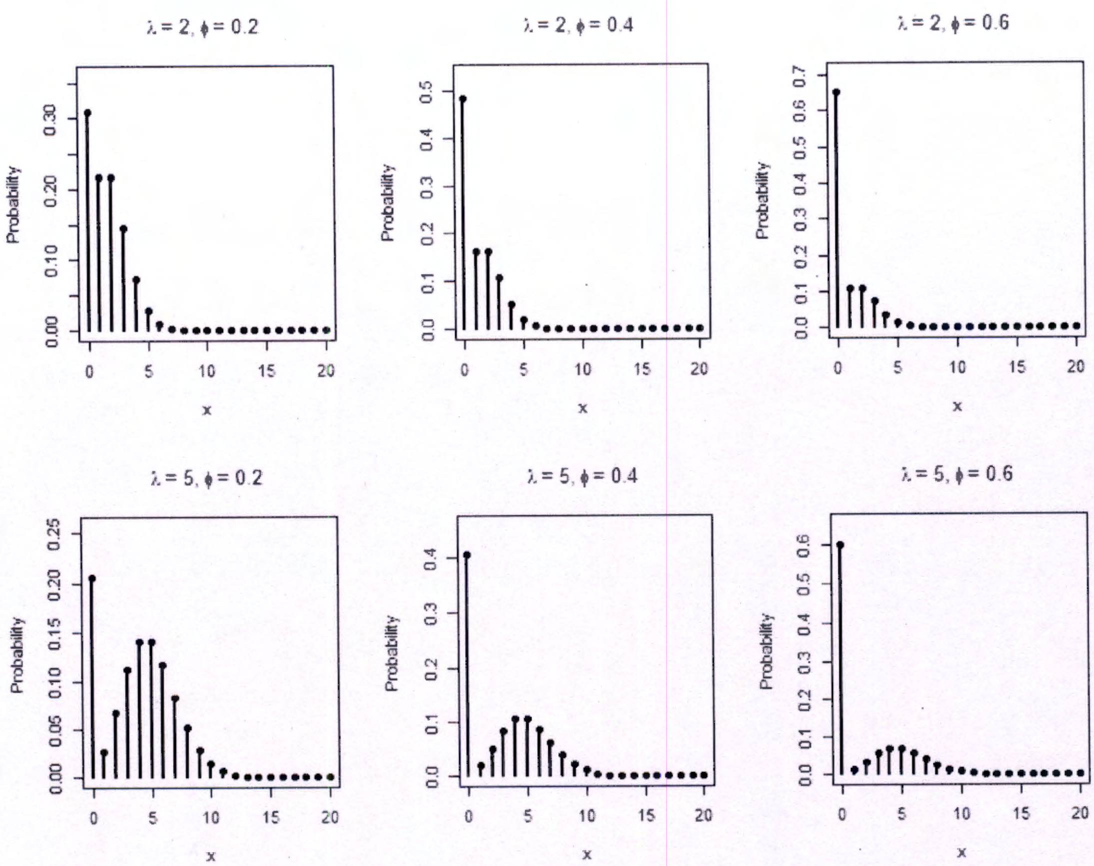
ดังนั้น $$\mathrm{Var}(X) = (1-\phi)\left(\lambda^2+\lambda\right) - \left[(1-\phi)\lambda\right]^2$$

$$= (1-\phi)\left(\phi\lambda^2+\lambda\right)$$

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจง ZIP ในโปรแกรม R

```
x<-seq(0,25,1)
lambda<-k1
pstr0<- k2
dzipois(x, lambda, pstr0, log = FALSE)
out.dist=c()
for(ii in 1 : length(x)){
  out.dist[ii]<-dzipois(x[ii],lambda, pstr0 )
}
sum(out.dist)
out.dist
```

รูปภาพที่ 5 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง ZIP สำหรับค่าพารามิเตอร์ λ และ ϕ ที่กำหนดไว้



รูปภาพที่ 5 ฟังก์ชันความน่าจะเป็นของการแจกแจง ZIP

2.2 การแจกแจง ZINB (zero inflated Negative Binomial Distribution)

การแจกแจง ZINB นำเสนอโดย Greene (1994) โดยการผสมระหว่างการแจกแจงเบอร์นูลลีและการแจกแจงทวินามลบ เพื่อเป็นทางเลือกสำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีศูนย์มาก โดยมีฟังก์ชันความน่าจะเป็น ดังนี้

บทนิยาม 7 ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจง ZINB (zero inflated Negative Binomial) เขียนแทนด้วย $X \sim \text{ZINB}(r, p, \phi)$ จะได้ฟังก์ชันความน่าจะเป็นของการแจกแจง ดังนี้

$$f(x) = \begin{cases} \phi + (1-\phi)\, p^r & ; x = 0 \\ (1-\phi)\dbinom{r+x-1}{x} p^r (1-p)^x & ; x = 1, 2, \ldots \end{cases}$$ สำหรับ $r > 0, 0 < p < 1$ และ

$$0 < \phi < 1$$

ทฤษฎีบท 5 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจง ZINB แล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$\mathrm{E}(X) = r(1-\phi)\left(\frac{1-p}{p}\right) \text{ และ } \mathrm{Var}(X) = r(1-\phi)\left(\frac{(1+r\phi)(1-p)^2}{p^2} + \frac{1-p}{p}\right)$$

พิสูจน์

ค่าเฉลี่ย

$$\begin{aligned}\mathrm{E}(X) &= \sum_{x=1}^{\infty} x(1-\phi)\binom{r+x-1}{x}p^r(1-p)^x \\ &= (1-\phi)\sum_{x=1}^{\infty} x\binom{r+x-1}{x}p^r(1-p)^x \\ &= (1-\phi)r\sum_{x=1}^{\infty}\binom{r+x-1}{x-1}p^r(1-p)^x.\end{aligned}$$

แทน $y = x-1$

ดังนั้น
$$\begin{aligned}\mathrm{E}(X) &= r(1-\phi)\sum_{y=0}^{\infty}\binom{r+y}{y}p^r(1-p)^{y+1} \\ &= \frac{(1-\phi)r(1-p)}{p}\sum_{y=0}^{\infty}\binom{(r+1)+y-1}{y}p^{r+1}(1-p)^y \\ &= (1-\phi)\frac{r(1-p)}{p}.\end{aligned}$$

ความแปรปรวน

พิจารณา $\mu_2'(X) = \mathrm{E}(X(X-1)) = \mathrm{E}(X^2 - X)$

$$\begin{aligned}\mathrm{E}(X(X-1)) &= \sum_{x=1}^{\infty} x(x-1)(1-\phi)\binom{r+x-1}{x}p^r(1-p)^x \\ &= (1-\phi)r(r+1)\sum_{x=2}^{\infty}\binom{r+x-1}{x-2}p^r(1-p)^x.\end{aligned}$$

แทน $y = x - 2$

$$\begin{aligned}
&\mathrm{E}(X(X-1)) \\
&= (1-\phi)r(r+1)\sum_{y=0}^{\infty}\binom{r+y+1}{y}p^r(1-p)^{y+2} \\
&= \frac{(1-\phi)r(r+1)(1-p)^2}{p^2}\sum_{y=0}^{\infty}\binom{(r+2)+y-1}{y}p^{r+2}(1-p)^y \\
&= \frac{(1-\phi)r(r+1)(1-p)^2}{p^2}.
\end{aligned}$$

จาก $\quad \mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$

$$\begin{aligned}
&= r(1-\phi)(r+1)\frac{(1-p)^2}{p^2} + r(1-\phi)\frac{(1-p)}{p} \\
&= r(1-\phi)\left[(r+1)\frac{(1-p)^2}{p^2} + \frac{(1-p)}{p}\right]
\end{aligned}$$

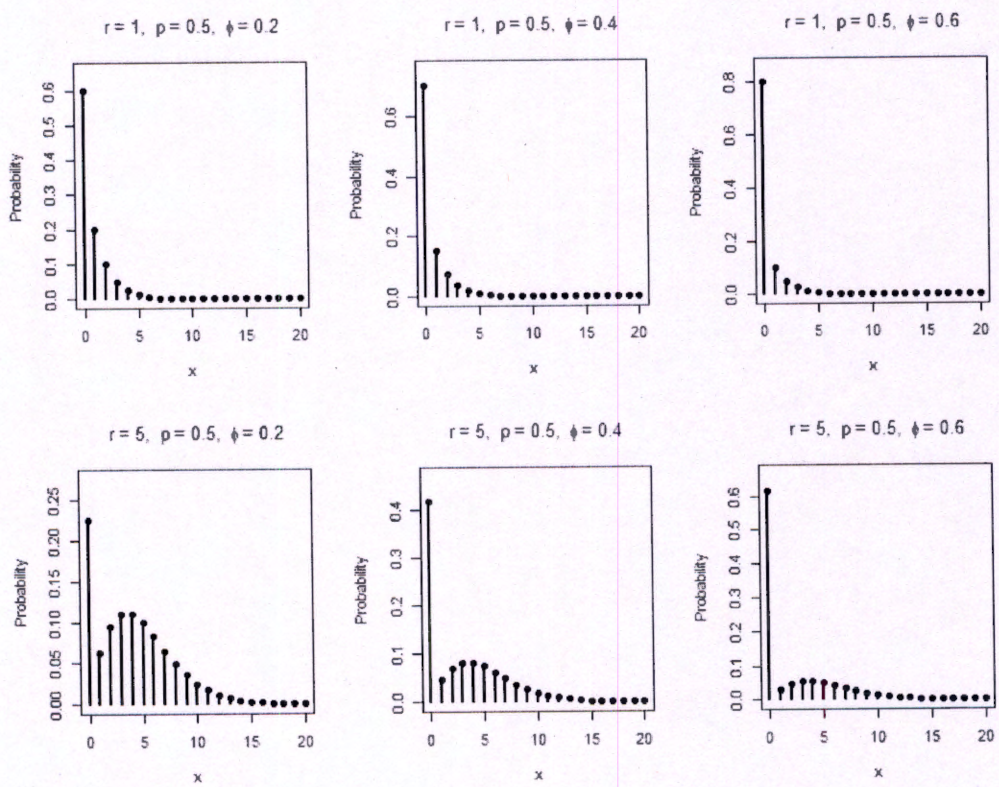
ดังนั้น $\quad \mathrm{Var}(X)$

$$\begin{aligned}
&= (1-\phi)r(1-\phi)(r+1)\frac{(1-p)^2}{p^2} + r(1-\phi)\frac{(1-p)}{p} \\
&\quad - \left[r(1-\phi)\frac{(1-p)}{p}\right]^2 \\
&= r(1-\phi)\left[\frac{(1+r\phi)(1-p)^2}{p^2} + \frac{(1-p)}{p}\right].
\end{aligned}$$

รูปภาพที่ 6 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง ZINB สำหรับค่าพารามิเตอร์ r, p และ ϕ ที่กำหนดไว้

การสร้างการแจกแจงความน่าจะเป็นของการแจกแจง ZINB ในโปรแกรม R

```
x<-seq(0,25,1)
size<-k1
munb<-k2
pstr0<-k3
dzinegbin(x, size, prob = NULL, munb, pstr0, log = FALSE)
out.dist=c()
for(ii in 1 : length(x))
{ out.dist[ii]<-dzinegbin(x[ii],size,prob = NULL, munb, pstr0) }
sum(out.dist)
out.dist
```



รูปภาพที่ 6 ฟังก์ชันความน่าจะเป็นของการแจกแจง ZINB

3. การพัฒนาตัวแบบสำหรับข้อมูลเชิงนับที่มีศูนย์มาก

การแจกแจง Hurdle สามารถนำมาใช้ในการวิเคราะห์ข้อมูลเชิงนับที่มีศูนย์มาก โดยข้อมูลจะได้จากสภาวะ (State) คือ สภาวะแรกเกิดขึ้นด้วยความน่าจะเป็น ϕ และส่วนที่เหลือเกิดขึ้นด้วยความน่าจะเป็น $1-\phi$ และนำไปสู่การแจกแจงสำหรับข้อมูลเชิงนับ ด้วยฟังก์ชันความน่าจะเป็น $f(x;\theta)$ ด้วยพารามิเตอร์ θ โดยที่ศูนย์ที่มาจากสภาวะที่เรียกว่า "Structural Zeros" (Lambert, 1992; Jansakul and Hinde, 2008) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$g(x) = \begin{cases} \omega & ,x=0 \\ \dfrac{(1-\omega)h(x)}{1-h(0)} & ,x=1,2,3,\ldots, \end{cases}$$

โดยที่ $1-\omega$ คือ ความน่าจะเป็นของการแจกแจงเมื่อมีค่าสังเกตที่ไม่ใช่ศูนย์

$h(x)$ คือ การแจกแจงฟังก์ชันความน่าจะเป็น

$h(0)$ คือ การแจกแจงพื้นฐานเมื่อค่าสังเกตเท่ากับ 0

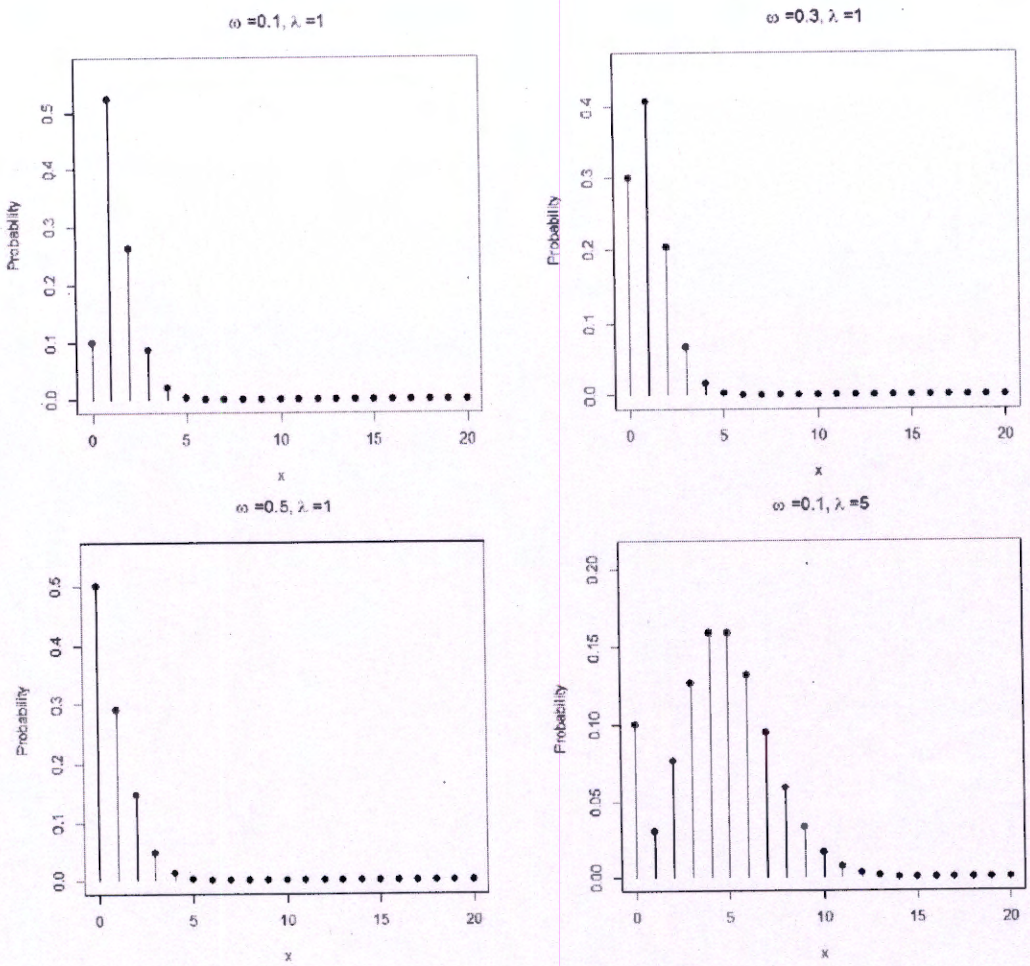
3.1 การแจกแจง Hurdle Poisson (HP) distribution

การแจกแจงปัวซงเป็นตัวเลือกหนึ่งที่ใช้กันทั่วไปสำหรับการแจงแจงแบบ Hurdle โดยมีฟังก์ชันความน่าจะเป็น ดังนี้

$$g(x) = \begin{cases} \omega & ,x=0 \\ (1-\omega)\dfrac{\exp(-\lambda)\lambda^x}{(1-\exp(-\lambda))x!} & ,x=1,2,3,\ldots, \end{cases}$$

โดยที่ $\lambda > 0$ และ $0 \le \omega \le 1$

รูปภาพที่ 7 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง HP สำหรับค่าพารามิเตอร์ λ และ ω ที่กำหนดไว้



รูปภาพที่ 7 ฟังก์ชันความน่าจะเป็นของการแจกแจง HP

ทฤษฎีบท 6 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจง HP แล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$E(X) = \frac{(1-\omega)\lambda}{1-\exp(-\lambda)}$$

และ

$$Var(X) = \frac{(1-\omega)\lambda^2+(1-\omega)\lambda}{1-\exp(-\lambda)} - \left[\frac{(1-\omega)\lambda}{1-\exp(-\lambda)}\right]^2.$$

พิสูจน์

ค่าเฉลี่ย

ค่าคาดหวังของตัวแปรสุ่ม x

$$
\begin{aligned}
E(X) &= (0)[\omega] + \sum_{x=1}^{\infty} x(1-\omega)\frac{\exp(-\lambda)\lambda^x}{(1-\exp(-\lambda))x!} \\
&= (1-\omega)\sum_{x=1}^{\infty} \frac{\exp(-\lambda)\lambda^x}{(1-\exp(-\lambda))(x-1)!} \\
&= \frac{(1-\omega)\lambda}{1-\exp(-\lambda)}\sum_{x=1}^{\infty} \frac{\exp(-\lambda)\lambda^{x-1}}{(x-1)!}.
\end{aligned}
$$

แทนค่า $y = x-1$ จะได้

$$
\begin{aligned}
E(X) &= \frac{(1-\omega)\lambda}{1-\exp(-\lambda)}\sum_{y=0}^{\infty} \frac{\exp(-\lambda)\lambda^{y}}{y!} \\
&= \frac{(1-\omega)\lambda}{1-\exp(-\lambda)}.
\end{aligned}
$$

ความแปรปรวน

$$
\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
&= E(X(X-1)) + E(X) - (E(X))^2 \\
E(X(X-1)) &= \sum_{x=1}^{\infty} x(x-1)(1-\omega)\frac{\exp(-\lambda)\lambda^x}{(1-\exp(-\lambda))x!} \\
&= (1-\omega)\sum_{x=2}^{\infty} x(x-1)\frac{\exp(-\lambda)\lambda^x}{(1-\exp(-\lambda))x!} \\
&= \frac{(1-\omega)\lambda^2}{1-\exp(-\lambda)}\sum_{x=2}^{\infty} \frac{\exp(-\lambda)\lambda^{x-2}}{(x-2)!}.
\end{aligned}
$$

แทนค่า $y = x-2$ จะได้

$$
\begin{aligned}
E(X(X-1)) &= \frac{(1-\omega)\lambda^2}{1-\exp(-\lambda)}\sum_{y=0}^{\infty} \frac{\exp(-\lambda)\lambda^{y}}{y!} \\
&= \frac{(1-\omega)\lambda^2}{1-\exp(-\lambda)} \\
E(X^2) &= E(X(X-1)) + E(X) \\
&= \frac{(1-\omega)\lambda^2 + (1-\omega)\lambda}{1-\exp(-\lambda)}.
\end{aligned}
$$

ดังนั้น

$$
\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
&= \frac{(1-\omega)\lambda^2 + (1-\omega)\lambda}{1-\exp(-\lambda)} - \left[\frac{(1-\omega)\lambda}{1-\exp(-\lambda)}\right]^2. \quad \square
\end{aligned}
$$

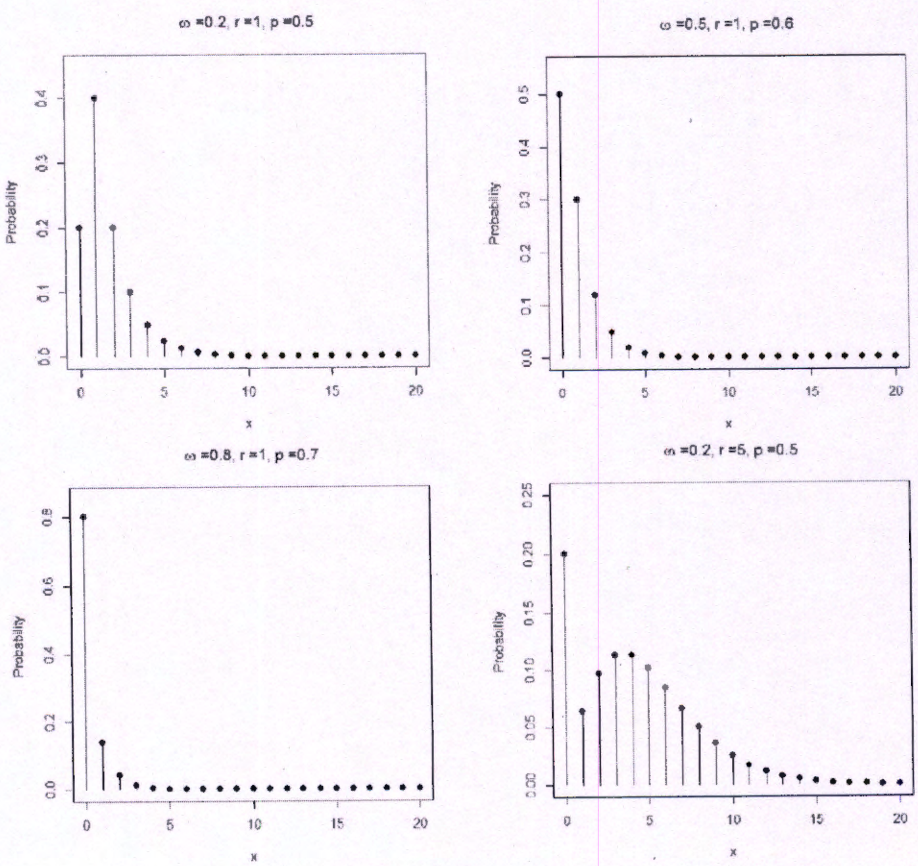
3.2 การแจกแจง Hurdle Negative Binomial (HNB) distribution

การแจกแจงทวินามลบเป็นอีกตัวเลือกหนึ่งที่ใช้กันทั่วไปสำหรับการแจงแจงแบบ Hurdle โดยมีฟังก์ชันความน่าจะเป็น ดังนี้

$$
g(x) = \begin{cases} \omega & ,x=0 \\ (1-\omega)\dfrac{\binom{r+x-1}{x} p^r (1-p)^x}{1-p^r} & ,x=1,2,3,\ldots, \end{cases}
$$

โดยที่ $r, p > 0$ และ $0 \le \omega \le 1$

รูปภาพที่ 8 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง HP สำหรับค่าพารามิเตอร์ r, p และ ω ที่กำหนดไว้



รูปภาพที่ 8 ฟังก์ชันความน่าจะเป็นของการแจกแจง HNB

ทฤษฎีบท 7 ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจง HNB แล้วค่าเฉลี่ยและความแปรปรวนมีค่าเท่ากับ

$$E(X) = (1-\omega)\frac{r(1-p)}{p(1-p^r)}$$

และ

$$Var(X) = \frac{(1-\omega)r(1-p)(r(1-p)+1)}{p^2(1-p^r)} - \left[\frac{(1-\omega)r(1-p)}{p(1-p^r)}\right]^2.$$

พิสูจน์

ค่าเฉลี่ย

ค่าคาดหวังของตัวแปรสุ่ม x

$$\begin{aligned}
E(X) &= (0)[\omega] + \sum_{x=1}^{\infty} x(1-\omega)\frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r} \\
&= \sum_{x=1}^{\infty} x(1-\omega)\frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r} \\
&= (1-\omega)\sum_{x=1}^{\infty} x\frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r} \\
&= \frac{(1-\omega)r}{1-p^r}\sum_{x=1}^{\infty}\binom{r+x-1}{x-1}p^r(1-p)^x.
\end{aligned}$$

แทนค่า $y = x-1$ จะได้

$$\begin{aligned}
E(X) &= \frac{r(1-\omega)}{1-p^r}\sum_{y=0}^{\infty}\binom{r+y}{y}p^r(1-p)^{y+1} \\
&= \frac{(1-\omega)r(1-p)}{p(1-p^r)}\sum_{y=0}^{\infty}\binom{(r+1)+y-1}{y}p^{r+1}(1-p)^y \\
&= (1-\omega)\frac{r(1-p)}{p(1-p^r)}.
\end{aligned}$$

ความแปรปรวน

$$\begin{aligned}
Var(X) &= E(X^2)-(E(X))^2\\
E(X^2) &= E(X(X-1))+E(X)\\
E(X(X-1)) &= \sum_{x=1}^{\infty} x(x-1)(1-\omega)\frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r}\\
&= \frac{(1-\omega)}{1-p^r}\sum_{x=1}^{\infty} x(x-1)\binom{r+x-1}{x}p^r(1-p)^x\\
&= \frac{(1-\omega)r(r+1)}{1-p^r}\sum_{x=2}^{\infty}\binom{r+x-1}{x-2}p^r(1-p)^x.
\end{aligned}$$

แทนค่า $y=x-2$ จะได้

$$\begin{aligned}
E(X(X-1)) &= \frac{(1-\omega)r(r+1)}{1-p^r}\sum_{y=0}^{\infty}\binom{r+y+1}{y}p^r(1-p)^{y+2}\\
&= \frac{(1-\omega)r(r+1)(1-p)^2}{p^2(1-p^r)}\sum_{y=0}^{\infty}\binom{(r+2)+y-1}{y}p^{r+2}(1-p)^y\\
&= \frac{(1-\omega)r(r+1)(1-p)^2}{p^2(1-p^r)},\\
E(X^2) &= E(X(X-1))+E(X)\\
&= (1-\omega)r(r+1)\frac{(1-p)^2}{p^2(1-p^r)}+(1-\omega)r\frac{(1-p)}{p(1-p^r)}.
\end{aligned}$$

ดังนั้น

$$\begin{aligned}
Var(X) &= E(X^2)-(E(X))^2\\
&= (1-\omega)r(r+1)\frac{(1-p)^2}{p^2(1-p^r)}+(1-\omega)r\frac{(1-p)}{p(1-p^r)}-\left[(1-\omega)r\frac{(1-p)}{p(1-p^r)}\right]^2\\
&= \frac{(1-\omega)r(1-p)(r(1-p)+1)}{p^2(1-p^r)}-\left[\frac{(1-\omega)r(1-p)}{p(1-p^r)}\right]^2. \quad \square
\end{aligned}$$

งานวิจัยที่เกี่ยวข้อง

Lambert (1992) พัฒนาตัวแบบ Zero inflated Poisson (ZIP) เพื่อใช้ในการอธิบายลักษณะการแจกแจงข้อมูลเชิงนับและประยุกต์การนับจำนวนผลิตภัณฑ์ที่ชำรุดในกระบวนการผลิตซึ่งพบว่าการแจกแจงที่เขาพัฒนาขึ้นมีประสิทธิภาพสูงกว่าการแจกแจงปัวซง

Greene (1994) พัฒนาตัวแบบ Zero inflated Negative Binomial (ZINB) เพื่อใช้ในการอธิบายลักษณะการแจกแจงข้อมูลเชิงนับและประยุกต์การตรวจนับการทุจริตในการใช้บัตรเครดิต ซึ่งพบว่าการแจกแจงที่เขาพัฒนาขึ้นมีประสิทธิภาพสูงกว่าการแจกแจงปัวซง

Ridout *et al.* (1998) ศึกษาปัญหาของการสร้างแบบจำลองข้อมูลเชิงนับด้วยตัวแบบ ZIP และตัวแบบ HP (Hurdle Poisson) โดยประยุกต์ใช้กับตัวอย่างเกี่ยวกับการวิจัยพืชสวนเพื่อประกอบการอธิบาย และศึกษาแนวทางในการพัฒนาการทดสอบคะแนน (score test) สำหรับการเปรียบเทียบตัวแบบ ZIP กับตัวแบบ ZINB

Hall (2000) ได้พัฒนาตัวแบบ Zero inflated Binomial (ZIB) โดยปรับปรุงมาจากตัวแบบ ZIP ที่เสนอโดย Lambert (1992) ซึ่งได้ประยุกต์ใช้ตัวแบบทั้งสองกับข้อมูลตัวอย่างทางด้านการเกษตร ซึ่งผลการศึกษาพบว่าตัวแบบ ZIB และ ZIP ให้ประสิทธิภาพที่ใกล้เคียงกัน

Ridout *et al.* (2001) พัฒนาการทดสอบคะแนน (score test) สำหรับการทดสอบระหว่างตัวแบบ ZIP และ ZINB เพื่อเป็นตัวบ่งบอกว่าควรใช้ตัวแบบใดจึงจะเหมาะสมกับข้อมูลแต่ละชุดระหว่างตัวแบบ ZIP และ ZINB

Famoye and Singh (2006) พัฒนาตัวแบบ Zero inflated generalized Poisson regression (ZIGP) เพื่อใช้ในการอธิบายลักษณะการแจกแจงข้อมูลเชิงนับและประยุกต์กับจำนวนความรุนแรงในครอบครัวซึ่งพบว่ามีศูนย์เป็นจำนวนมาก

Yip and Yau (2005) ศึกษาตัวแบบข้อมูลความถี่ของการ Claim ใน General Insurance with Extra Zeros โดยทำการเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับข้อมูลเชิงนับที่มีศูนย์จำนวนมาก ได้แก่ ตัวแบบ ZIP, ZINB, ZIGP และตัวแบบ ZIDP (Zero inflated double Poisson) ผลการศึกษาพบว่าตัวแบบ ZIDP มีประสิทธิภาพมากที่สุดในการวิเคราะห์ข้อมูลเชิงนับที่มีศูนย์จำนวนมาก

Jansakul and Hinde (2008) ได้ศึกษาคะแนนทดสอบสำหรับตัวแบบที่มีศูนย์มากใน Zero-Inflated โดยได้เสนอสถิติคะแนนทดสอบทั่วไปสำหรับเปรียบเทียบประสิทธิภาพตัวแบบ ZINB (zero-inflated Negative Binomial) กับตัวแบบ NB (Negative Binomial) เมื่อเกิดปัญหาความแปรปรวนมีค่ามากกว่าค่าเฉลี่ย ผลจากการศึกษาจำลองสถานการณ์ (simulation results) ชี้ให้เห็นถึงประสิทธิภาพของการทดสอบว่าการทดสอบที่มีประสิทธิภาพดีพอสมควรและมีการกระจายอย่างสุ่มภายใต้สมมติฐาน (ตัวแบบNB) ประมาณด้วยการแจกแจงปกติ (usual Distribution) นอกจากนี้ก็ทำการใช้การทดสอบด้วยข้อมูล apple shoot propagation ด้วยคะแนนทดสอบพบว่าบ่งบอกถึงรูปแบบดังกล่าวเหมาะสม

Flynn (2009) ได้ประยุกต์ใช้ตัวแบบ GLMs สำหรับข้อมูลด้านประกันภัย (insurance) ที่มีจำนวนการเรียกสินไหม (claim) เป็นศูนย์จำนวนมาก โดยใช้ตัวแบบ ZIP และ ZINB โดยผสมผสานความยืดหยุ่นมากขึ้นในรูปแบบนับ GLM ซึ่งในทางปฏิบัติอาจจะสามารถปรับปรุงรูปแบบที่เหมาะสมและเพิ่มประสิทธิภาพของตัวแบบจำลอง การใช้ตัวแบบ ZIP หรือ ZINB

ปรับปรุงแบบแบบจำลองที่มีฐานข้อมูลการประกัน (insurance database) พบว่าตัวแบบทั้งสองให้การประมาณโดยรวมดีกว่าตัวแบบที่ไม่มีฐานข้อมูลการประกัน (non-insurance database)

Bodhisuwan, W. (2011) พัฒนาการแจกแจง Zero inflated Waring (ZIW) เพื่อใช้ในการอธิบายลักษณะการแจกแจงข้อมูลเชิงนับและประยุกต์การนับกับจำนวนอุบัติเหตุในกระบวนการผลิตซึ่งพบว่าการแจกแจงที่เขาพัฒนาขึ้นมีประสิทธิภาพสูงกว่าการแจกแจง ZIP และ ZINB

Pudprommarat *et al.* (2012) พัฒนาตัวแบบผสมระหว่างการแจกแจงทวินามลบและการแจกแจงบีต้าเอกซ์โพเนนเชียลและนำไปประยุกต์ใช้ในการวิเคราะห์จำนวนอุบัติเหตุพบว่าการแจกแจงที่พัฒนาขึ้นมีประสิทธิภาพสูงกว่าการแจกแจงปัวซง และการแจกแจงทวินามลบ จากผลการศึกษาของ พบว่าจากการใช้ข้อมูลจริงเกี่ยวกับจำนวนอุบัติเหตุ (accident) การแจกแจงดังกล่าวมีประสิทธิภาพดีกว่าการแจกแจงปัวซงและการแจกแจงทวินามลบ

วิธีดำเนินการวิจัย

การวิจัยเป็นการพัฒนาทฤษฎีและศึกษาคุณสมบัติเชิงความน่าจะเป็นของการแจกแจง NB-NWL รวมทั้งการทดลองโดยใช้เทคนิคการจำลองด้วยคอมพิวเตอร์ โดยมีขั้นตอนการดำเนินการวิจัย ดังนี้

1. การพัฒนาการแจกแจง NB-NWL และ HNB-NWL
2. การศึกษาคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL และ HNB-NWL
3. การสร้างตัวแปรสุ่ม NB-NWL และ HNB-NWL
4. การศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL และ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด โดยใช้วิธีการจำลองสถานการณ์
5. การประยุกต์ใช้การแจกแจง NB-NWL และ HNB-NWL ในการวิเคราะห์ข้อมูลเชิงนับ

1. การพัฒนาการแจกแจง NB-NWL

พัฒนาการแจกแจง NB-NWL โดยการผสม 2 การแจกแจง คือ การแจกแจงเบอร์นูลลีด้วยพารามิเตอร์ ϕ และการแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ ด้วยพารามิเตอร์ r,a,b โดยที่ฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL ได้แก่ $f(x;r,a,b)$ ดังนั้นจะได้ฟังก์ชันความน่าจะเป็น (pmf) ของการแจกแจง NB-NWL ดังนี้

$$h(x)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{b^2(1+a)^2}{ab(1+a)+a(2+a)}\left(\frac{b+r+j+1}{(b+r+j)^2}-\frac{b+ab+r+j+1}{(b+ab+r+j)^2}\right),$$

ทฤษฎีบท 8 กำหนดให้ $X \sim$ NB-NWL(r,a,b). ฟังก์ชันความน่าจะเป็นของ X คือ

$$f(x;r,a,b)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{b^2(1+a)^2}{ab(1+a)+a(2+a)}\left(\frac{b+r+j+1}{(b+r+j)^2}-\frac{b+ab+r+j+1}{(b+ab+r+j)^2}\right),\quad x=0,1,2,\ldots$$

โดยที่ a และ $b>0$.

พิสูจน์ ถ้า $X|\lambda \sim \mathrm{NB}(r,p=\exp(-\lambda))$ และ $\lambda \sim \mathrm{NWL}(\alpha,\beta))$, แล้ว pmf ของ X สามารถเขียนอธิบายได้โดย

$$f(x)=\int_0^\infty f_1(x|\lambda)g(\lambda;\alpha,\beta)d\lambda,\ \text{where}\ f_1(x|\lambda)\ \text{is express as}$$

$$f_1(x|\lambda)=\binom{r+x-1}{x}\exp(-\lambda r)(1-\exp(-\lambda))^x,$$

$$=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\exp(-\lambda(r+j)).$$

แทนสมการ $f_1(x|\lambda)$ ใน $f(x)=\int_0^\infty f_1(x|\lambda)g(\lambda;\alpha,\beta)d\lambda$, แล้วจะได้

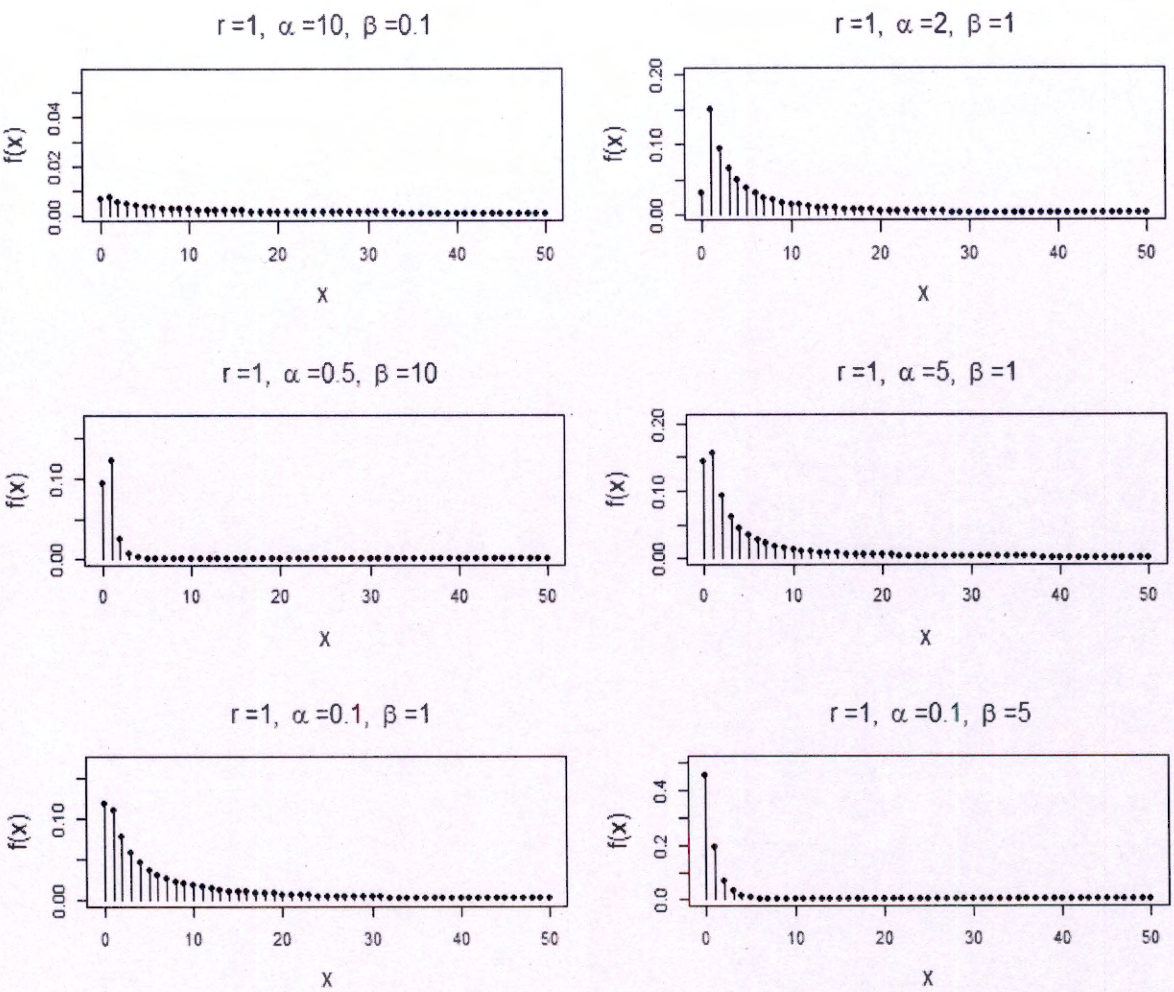
$$f(x)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\int_0^\infty \exp(-\lambda(r+j))g(\lambda;\alpha,\beta)d\lambda\right),$$

$$=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j M_\lambda(-(r+j)).$$

แทนค่าฟังก์ชันก่อกำเนิดโมเมนต์ของการแจกแจง WL ในสมการด้านบนจะได้ pmf ของการแจกแจง NB-NWL(r,α,β) ดังนี้

$$f(x;r,\alpha,\beta)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\beta^2(1+\alpha)^2}{\alpha\beta(1+\alpha)+\alpha(2+\alpha)}\left\{\frac{\beta+r+j+1}{(\beta+r+j)^2}-\frac{\beta(1+\alpha)+r+j+1}{[\beta(1+\alpha)+r+j]^2}\right\}.$$

รูปภาพที่ 9 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL สำหรับค่าพารามิเตอร์ r,a,b ที่กำหนดไว้



รูปภาพที่ 9 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL

2. การศึกษาคุณสมบัติเชิงความน่าจะเป็นสำหรับการแจกแจง NB-NWL

ศึกษาค่าเฉลี่ยและความแปรปรวนของการแจกแจง NB-NWL จากฟังก์ชันความน่าจะเป็นของการแจกแจง $f(x;\phi,r,a,b,c)$ ตลอดจนศึกษารูปแบบการกระจายตัวของการแจกแจง NB-NWL เมื่อกำหนดค่าบางค่าให้พารามิเตอร์ในการแจกแจง

3. การสร้างตัวแปรสุ่ม NB-NWL

สร้างตัวแปรสุ่ม NB-NWL โดยใช้การสร้างตัวแปรสุ่มเบอร์นูลลีร่วมกับการสร้างตัวแปรสุ่ม NWL โดยที่จะกำหนดขั้นตอนการสร้างตัวแปรสุ่ม NB-NWL ดังนี้

1. กำหนดความน่าจะเป็นที่เกิด structural zeros เท่ากับ ϕ และ $(1-\phi)$ เป็นความน่าจะเป็นที่เกิดสภาวะการแจกแจง NB-NWL

2. สร้างตัวแปรสุ่ม Y จากการแจกแจง NB-NWL, $Y \sim NB-NWL(r,a,b,c)$

3. สร้างตัวแปรสุ่ม u จากการแจกแจงยูนิฟอร์ม (0,1), $u \sim U(0,1)$

4. ถ้า $u > \phi$ แล้วให้กำหนด $X = Y$ กรณีอื่นๆ กำหนดให้ $X = 0$

4. การศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด โดยใช้วิธีการจำลองสถานการณ์

ตัวอย่างข้อมูลจะถูกสร้างจากการแจกแจง NB-NWL ด้วยพารามิเตอร์ที่กำหนดไว้ 4 กรณี ได้แก่

1. $r=5,\ a=3,\ b=3,\ c=3$
2. $r=5,\ a=10,\ b=5,\ c=3$
3. $r=5,\ a=10,\ b=3,\ c=3$
4. $r=10,\ a=10,\ b=3,\ c=3$

ค่าประมาณเฉลี่ยของพารามิเตอร์ θ : $\hat{\theta}_{av} = \frac{1}{500}\sum_{t=1}^{500}\hat{\theta}_t$ (เมื่อ $\hat{\theta}_t$ คือ ค่าประมาณของพารามิเตอร์ θ รอบที่ t โดยที่ $t = 1,2,...,500$)

ค่าเอนเอียง (Bias) : $\text{Bias}(\hat{\theta}_{av}) = \hat{\theta}_{av} - \theta$

ความแปรปรวน (Variance) : $\text{Var}(\hat{\theta}_{av}) = \frac{1}{500-1}\sum_{t=1}^{500}(\hat{\theta}_t - \hat{\theta}_{av})^2$

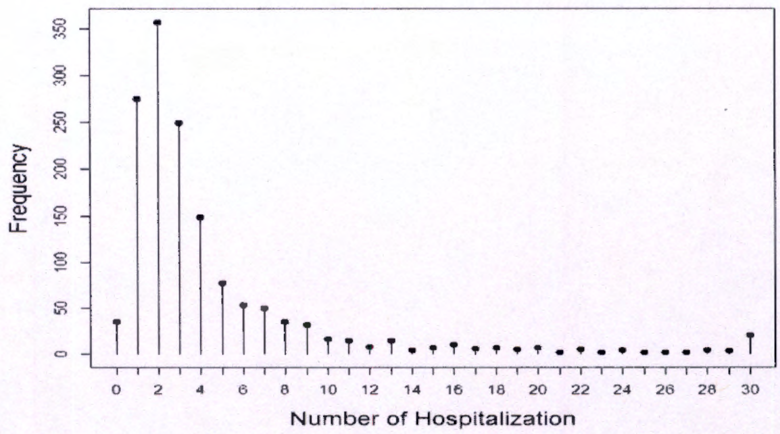
ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : SD) : $\text{SD}(\hat{\theta}_{av}) = \sqrt{\text{Var}(\hat{\theta}_{av})}$

ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error : MSE) : $\text{MSE}(\hat{\theta}_{av}) = \text{Var}(\hat{\theta}_{av}) + \text{Bias}^2(\hat{\theta}_{av})$

5. การประยุกต์ใช้การแจกแจง NB-NWL ในการวิเคราะห์ข้อมูลเชิงนับ

การศึกษาเชิงประยุกต์จะใช้ข้อมูลจริงนั่นคือ จำนวนของผู้ป่วยในของโรงพยาบาลในสหรัฐอเมริกาที่อายุตั้งแต่ 66 ปี (Flynn, 2009) โดยทำการเปรียบเทียบประสิทธิภาพในการหาค่า

คาดหวังระหว่างการแจกแจง ZIP, ZINB และ NB-NWL เกณฑ์ที่ใช้เลือกตัวแบบสำหรับการศึกษา คือ AIC (Akaike information criteria) และ BIC (Bayesian information criteria) และใช้การทดสอบสารูปสนิทดี (goodness of fit test: Anderson-Darling test statistic) เพื่อทำการทดสอบว่ามีการแจกแจงดังที่คาดหวังหรือไม่



รูปภาพที่ 10 จำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาตัวที่โรงพยาบาลราชบุรี

ผลการวิจัย

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาการแจกแจง Negative Binomial - New Weighted Lindley Distribution (NB-NWL) และการแจกแจง Hurdel Negative Binomial - New Weighted Lindley Distribution (HNB-NWL) ตลอดจนศึกษาคุณสมบัติเชิงความน่าจะเป็น เช่น ค่าเฉลี่ย และความแปรปรวน นอกจากนี้ยังศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL และ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด และประยุกต์ใช้การแจกแจง NB-NWL ในการวิเคราะห์ข้อมูลเชิงนับ โดยผลการวิจัยมีดังต่อไปนี้

1. การแจกแจง NB-NWL
2. การสร้างตัวแปรสุ่มของการแจกแจง NB-NWL
3. การประมาณค่าพารามิเตอร์
4. การศึกษาจำลองสถานการณ์
5. การศึกษาเชิงประยุกต์

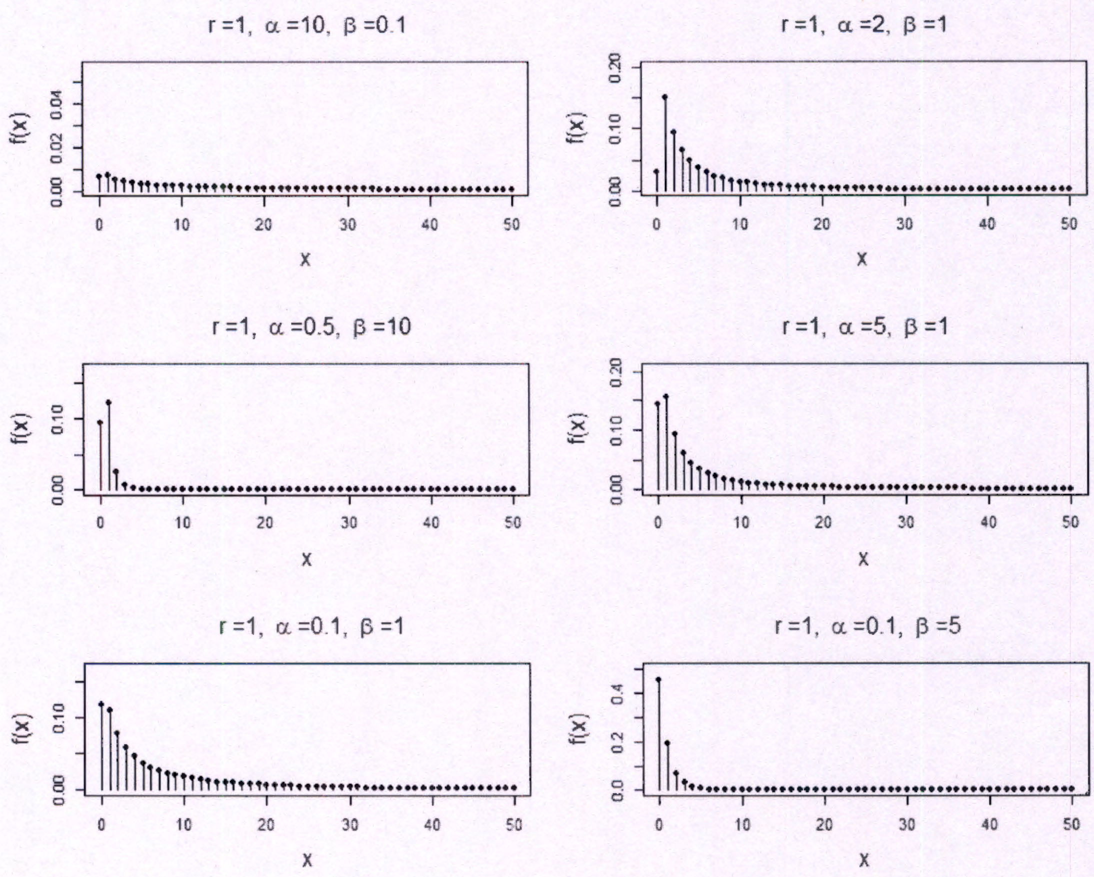
1. การแจกแจง NB-NWL

ผู้วิจัยจะนำเสนอการแจกแจง NB-NWL ซึ่งเป็นการนำสองการแจกแจงมาผสมกัน ได้แก่ การแจกแจงเบอร์นูลลี (Bernoulli) และการแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (NB-BE) จากการศึกษาของ Pudprommarat *et al.* (2012) ฟังก์ชันความน่าจะเป็น (probability mass function: pmf) และมีคุณลักษณะที่สำคัญของการแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่แสดงไว้ดังบทนิยาม 2.4 และทฤษฎีบท 4.2 ตามลำดับ ดังนั้นจะได้การแจกแจง NB-NWL ดังนี้

บทนิยาม 8 ให้ตัวแปรสุ่ม X มีการแจกแจง Negative Binomial-New Weighted Lindley (NB-NWL) ด้วยพารามิเตอร์ ϕ, r, a, b และ c ฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL ได้แก่

$$h(x)=\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}\,\frac{b^2(1+a)^2}{ab(1+a)+a(2+a)}\left(\frac{b+r+j+1}{(b+r+j)^2}-\frac{b+ab+r+j+1}{(b+ab+r+j)^2}\right),$$

เมื่อ $r, a, b > 0$



รูปภาพที่ 11 ฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL

รูปภาพที่ 11 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL สำหรับค่าพารามิเตอร์ r,a,b ที่กำหนดไว้ ซึ่งพบว่าการแจกแจงจะหนาแน่นจะมีรูปร่าง (shape) แตกต่างกัน เมื่อพารามิเตอร์ r และ a เปลี่ยนไป

ทฤษฎีบท 9 ให้ $X \sim \text{NB-NWL}(r,a,b)$ แล้วค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่ม X มีค่าเท่ากับ

1. $\mathrm{E}(X)=\dfrac{(1-\phi)\left[r\,\mathrm{B}\left(b-\frac{1}{c},a\right)-r\mathrm{B}(a,b)\right]}{\mathrm{B}(a,b)}$ เมื่อ $b>\dfrac{1}{c}$

2. $$\mathrm{Var}(X)=\frac{(1-\phi)}{\mathrm{B}(a,b)}\left((r^2+r)\mathrm{B}\left(b-\frac{2}{c},a\right)-(2r^2+r)\mathrm{B}\left(b-\frac{1}{c},a\right)+r^2\mathrm{B}(a,b)\right)$$
$$-\left(\frac{1-\phi}{\mathrm{B}(a,b)}\right)^2\left(r\,\mathrm{B}\left(b-\frac{1}{c},a\right)-r\mathrm{B}(a,b)\right)^2 \quad \text{เมื่อ } b>\frac{1}{c}$$

พิสูจน์ ถ้า $X \sim \text{NB-NWL}(r,a,b)$ แล้วค่าคาดหวัง (Expected) ของตัวแปรสุ่ม X มีค่าเท่ากับ

$$\mathrm{E}(X)=\sum_{x=0}^{\infty}x\,g(x)=\sum_{x=1}^{\infty}x(1-\phi)\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}$$
$$=\frac{(1-\phi)\left[r\,\mathrm{B}\left(b-\frac{1}{c},a\right)-r\mathrm{B}(a,b)\right]}{\mathrm{B}(a,b)}$$

ความแปรปรวนของตัวแปรสุ่ม X มีค่าเท่ากับ

$$\mathrm{Var}(X)=\mathrm{E}(X^2)-\left(\mathrm{E}(X)\right)^2=\sum_{x=0}^{\infty}x^2g(x)-\left[\sum_{x=0}^{\infty}xg(x)\right]^2$$
$$=\sum_{x=1}^{\infty}x^2(1-\phi)\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}$$
$$-\left[\sum_{x=1}^{\infty}x(1-\phi)\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}\right]^2$$

$$= (1-\phi)\left(\frac{(r^2+r)\mathrm{B}\left(b-\frac{2}{c},a\right)-(2r^2+r)\mathrm{B}\left(b-\frac{1}{c},a\right)+r^2\mathrm{B}(a,b)}{\mathrm{B}(a,b)}\right)$$

$$-\left[(1-\phi)\frac{r\,\mathrm{B}\left(b-\frac{1}{c},a\right)-r\mathrm{B}(a,b)}{\mathrm{B}(a,b)}\right]^2$$

ดังนั้น

$$\mathrm{Var}(X)=\frac{(1-\phi)}{\mathrm{B}(a,b)}\left((r^2+r)\mathrm{B}\left(b-\frac{2}{c},a\right)-(2r^2+r)\mathrm{B}\left(b-\frac{1}{c},a\right)+r^2\mathrm{B}(a,b)\right)$$

$$-\left(\frac{1-\phi}{\mathrm{B}(a,b)}\right)^2\left(r\,\mathrm{B}\left(b-\frac{1}{c},a\right)-r\mathrm{B}(a,b)\right)^2$$

1.1 การสร้างตัวแปรสุ่มของการแจกแจง NB-NWL

การสร้างตัวแปรสุ่มของการแจกแจง NB-NWL (Random Generation of NB-NWL Distribution) ในการสร้าง (Generate) ตัวแปรสุ่ม X จาก $X \sim \text{NB-NWL}(r,a,b)$ สามารถสร้างตามขั้นตอนต่อไปนี้

1) สร้าง u จากการแจกแจงยูนิฟอร์ม (Uniform Distribution : U(0,1))

2) สร้าง k จากการแจกแจง NWL

3) กำหนดให้ $\lambda = -\frac{1}{c}\log(-k+1)$ โดยที่ $\lambda \sim \mathrm{BE}(a,b,c)$.

4) สร้าง Y จากการแจกแจง NB($r, p = \exp(-\lambda)$)

5) สร้าง $u*$ จากแจกแจงยูนิฟอร์ม (Uniform Distribution : U(0,1))

6) ถ้า $u^* > \phi$ แล้วให้กำหนด $X = Y$ กรณีอื่นๆ กำหนดให้ $X = 0$

1.2 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ (Parameter Estimation) ของการแจกแจง NB-NWL จะใช้วิธี MLE (Maximum Likelihood Estimation) โดยที่ฟังก์ชัน likelihood ของ NB-NW, $X \sim$ NB-NWL(r,a,b) มีค่าเท่ากับ

$$L = L(\theta,r,a,b,c) = \prod_{i=1}^{n}\left[I_{(x_i=0)}\left(\phi+(1-\phi)\frac{\mathrm{B}\left(b+\frac{r}{c},a\right)}{\mathrm{B}(a,b)}\right)+I_{(x_i>0)}\left((1-\phi)\binom{r+x_i-1}{x_i}\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}\right)\right]$$

$$= \prod_{i=1}^{n}\left[I_{(x_i=0)}\left(\phi+(1-\phi)\frac{\mathrm{B}\left(b+\frac{r}{c},a\right)}{\mathrm{B}(a,b)}\right)+I_{(x_i>0)}\left((1-\phi)\frac{\Gamma(\mathrm{r}+\mathrm{x}_i)}{\Gamma(r)\Gamma(x_i+1)}\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}\right)\right]$$

แล้วเราสามารถเขียนฟังก์ชัน log likelihood ของ NB-NWL (ϕ,r,a,b,c) ได้ดังนี้

$$\log L = \sum_{i=1}^{n}\left[I_{(x_i=0)}\log\left(\phi+(1-\phi)\frac{\mathrm{B}(b+r/c,a)}{\mathrm{B}(a,b)}\right)+I_{(x_i>0)}\left(\log(1-\phi)+\log\Gamma(r+x_i)\right.\right.$$

$$\left.\left.-\log\Gamma(x_i+1)-\log\Gamma(r)+\log\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}\left(b+\frac{r+j}{c},a\right)}{\mathrm{B}(a,b)}\right)\right]$$

สำหรับอนุพันธ์ (Differentiating) ของฟังก์ชัน Log Likelihood ของ NB-NWL เราจะได้ Partial Derivatives ของฟังก์ชัน Log Likelihood เมื่อเทียบกับ r,a,b ดังนี้

$$\frac{\partial}{\partial r}\log L = \sum_{i=1}^{n}\left[I_{(x_i=0)}\frac{\partial}{\partial r}\log\left(\phi+(1-\phi)\frac{\mathrm{B}(b+r/c,a)}{\mathrm{B}(a,b)}\right)\right.$$

$$\left.+I_{(x_i>0)}\left(\psi(r+x_i)-\psi(r)+\frac{\partial}{\partial r}\log\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}(b+(r+j)/c,a)}{\mathrm{B}(a,b)}\right)\right] \text{ เมื่อ } \psi(s)=\frac{\Gamma'(s)}{\Gamma(s)}$$

$$\frac{\partial}{\partial a}\log L = \sum_{i=1}^{n}\left[I_{(x_i=0)}\frac{\partial}{\partial a}\log\left(\phi+(1-\phi)\frac{\mathrm{B}(b+r/c,a)}{\mathrm{B}(a,b)}\right)\right.$$

$$\left.+I_{(x_i>0)}\left(\frac{\partial}{\partial a}\log\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\mathrm{B}(b+(r+j)/c,a)}{\mathrm{B}(a,b)}\right)\right],$$

$$\frac{\partial}{\partial b}\log L=\sum_{i=1}^{n}\Bigg[I_{(x_i=0)}\frac{\partial}{\partial b}\log\left(\phi+(1-\phi)\frac{\mathrm{B}\left(b+r/c,a\right)}{\mathrm{B}(a,b)}\right)$$
$$+I_{(x_i>0)}\left(\frac{\partial}{\partial b}\log\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}\frac{\mathrm{B}\left(b+(r+j)/c,a\right)}{\mathrm{B}(a,b)}\right)\Bigg],$$

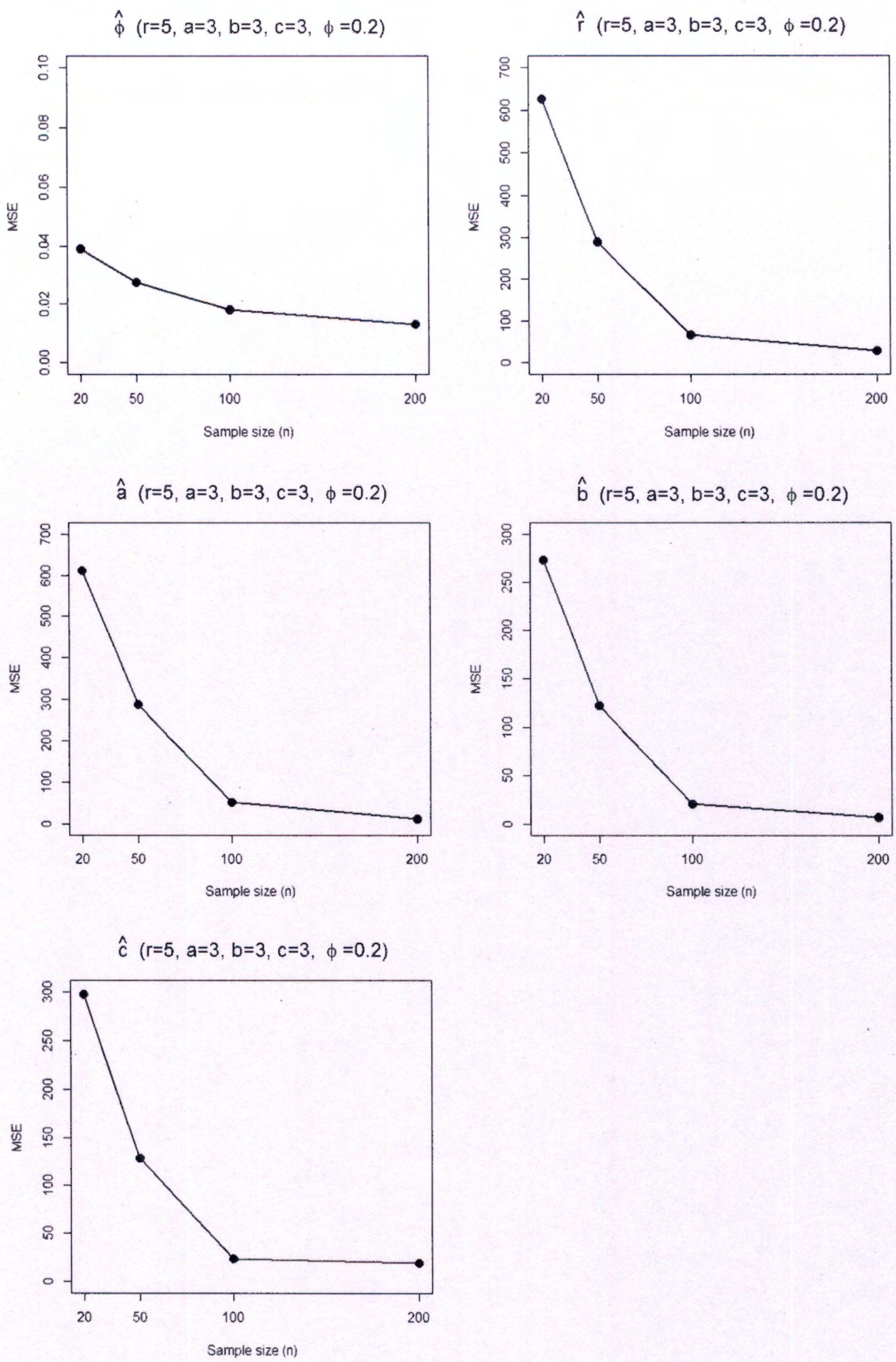
เราจะได้ค่าประมาณของพารามิเตอร์ *r, a, b* โดยวิธี Numerical Optimization ด้วยฟังก์ชัน *optim* ในโปรแกรม R (R Core Team, 2012).

1.3 การศึกษาจำลองสถานการณ์

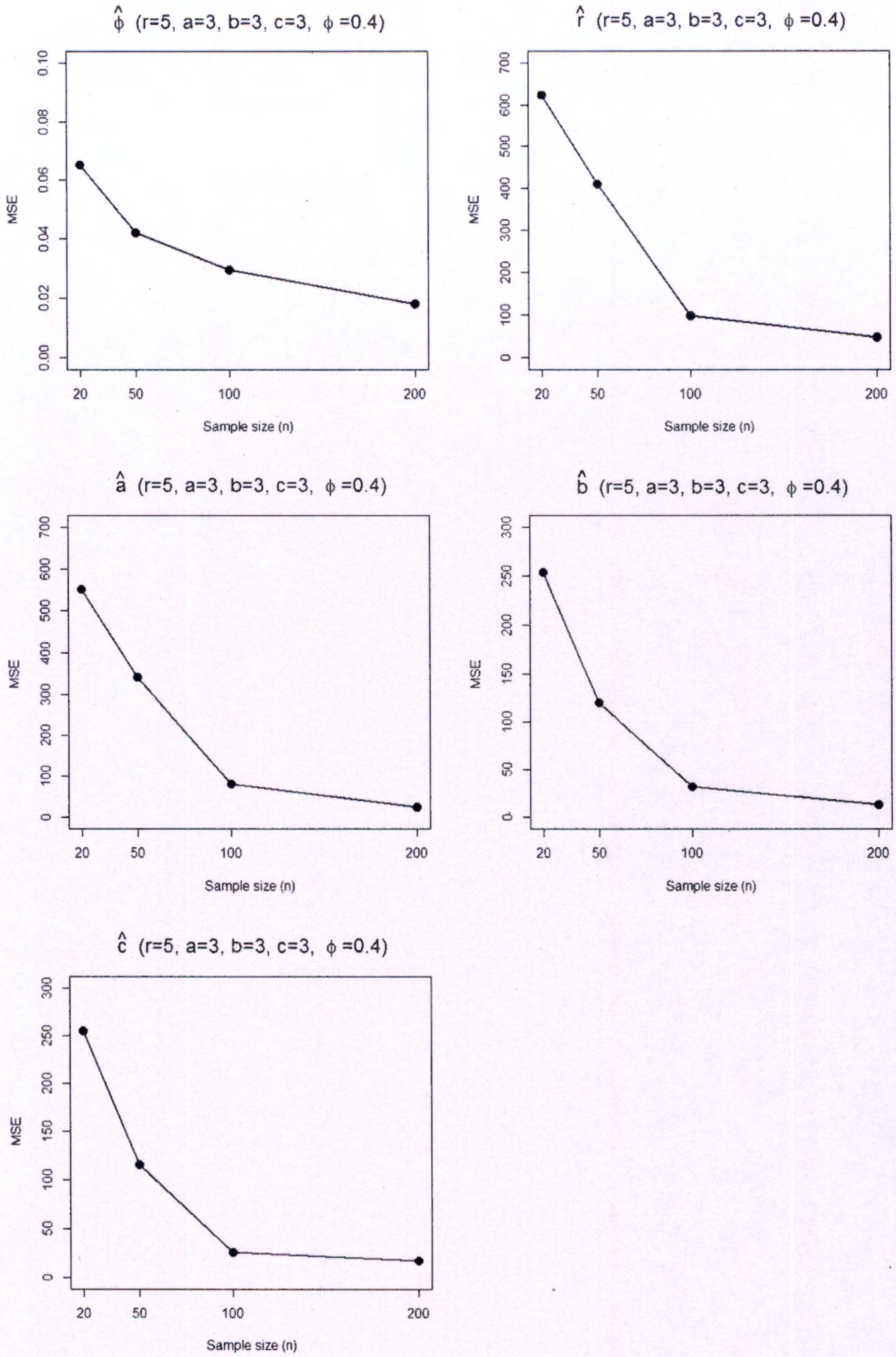
การศึกษาจำลองสถานการณ์ (Simulation Study) จากตัวอย่างข้อมูลจะถูกสร้างจากการแจกแจง NB-NWL ด้วยพารามิเตอร์ที่กำหนดไว้ 4 กรณี ได้แก่

1. $r=5,\ a=3,\ b=3$
2. $r=5,\ a=10,\ b=5$
3. $r=5,\ a=10,\ b=3$
4. $r=10,\ a=10,\ b=3$

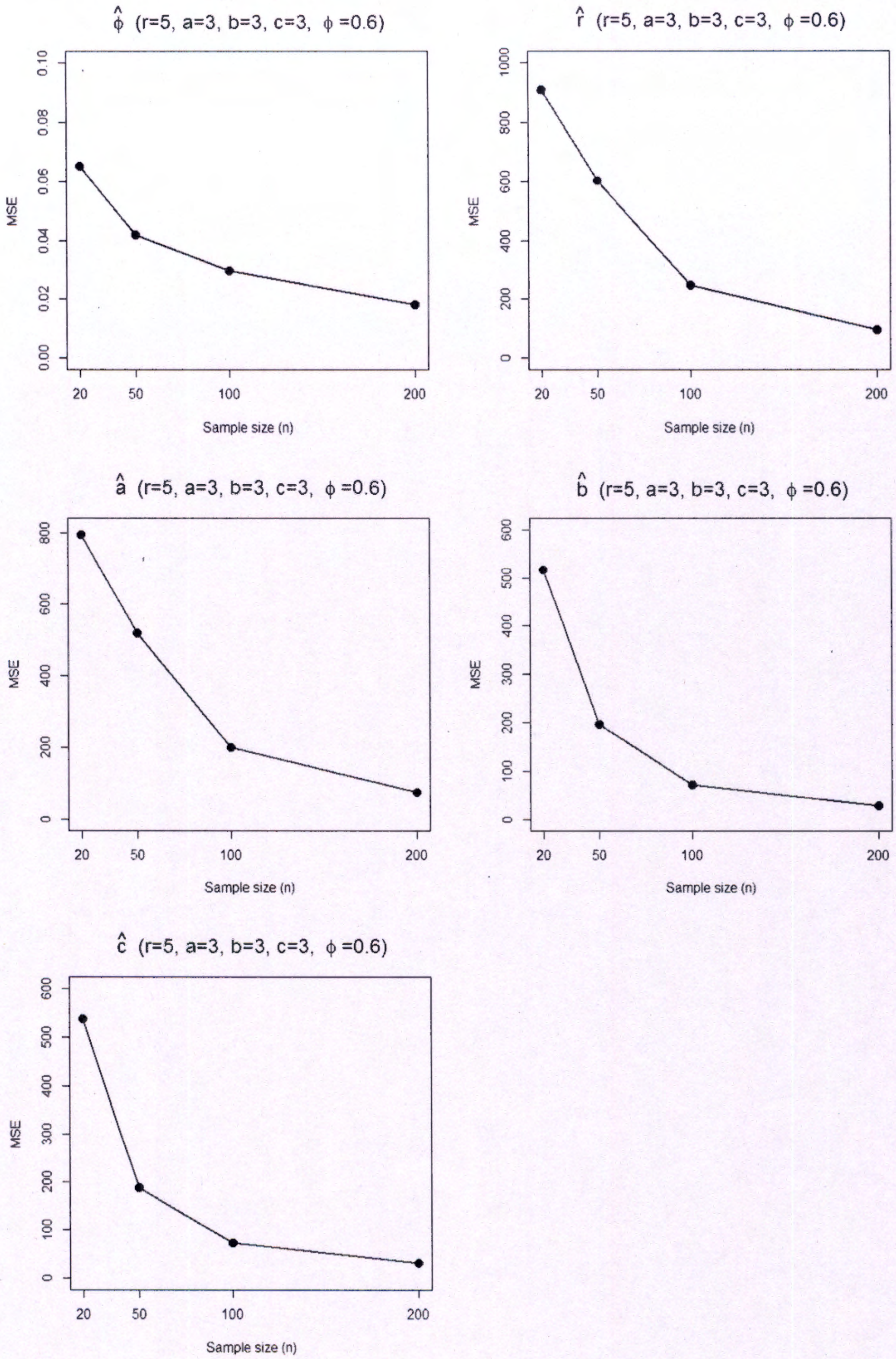
แล้วแต่ละกรณีจะใช้ใช้ขนาดตัวอย่าง (Sample Sizes : n) เท่ากับ 20, 50, 100 และ 200 ตามลำดับ ในแต่ละสถานการณ์จะทำการประมาณค่าซ้ำ 500 รอบ จากผลการศึกษาพบว่าวิธี MLE จะมีประสิทธิภาพสูงเมื่อขนาดตัวอย่างใหญ่ และจะมีประสิทธิภาพต่ำเมื่อขนาดตัวอย่างเล็ก และเมื่อพิจารณาค่า MSE พบว่าจะมีค่าเพิ่มขึ้นเมื่อขนาดสัดส่วนของศูนย์สูงขึ้นจะได้รูปภาพที่ 12 – 15



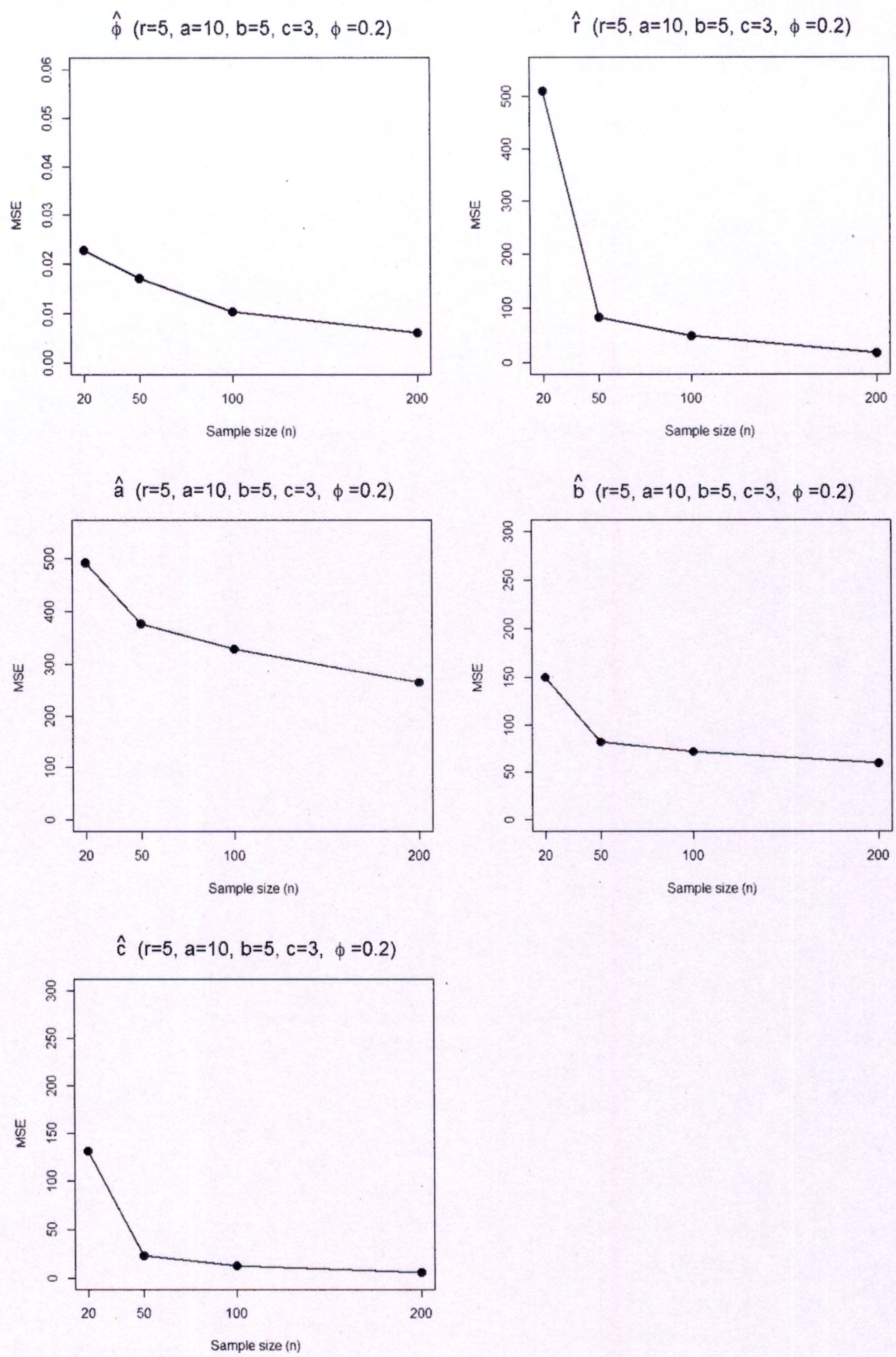
รูปภาพที่ 12 ค่า MSE เมื่อ $X \sim \text{NB-NWL}(r,a,b) = 5,3,3$



รูปภาพที่ 13 ค่า MSE เมื่อ $X \sim \text{NB-NWL}(r,a,b) = 5,10,5$



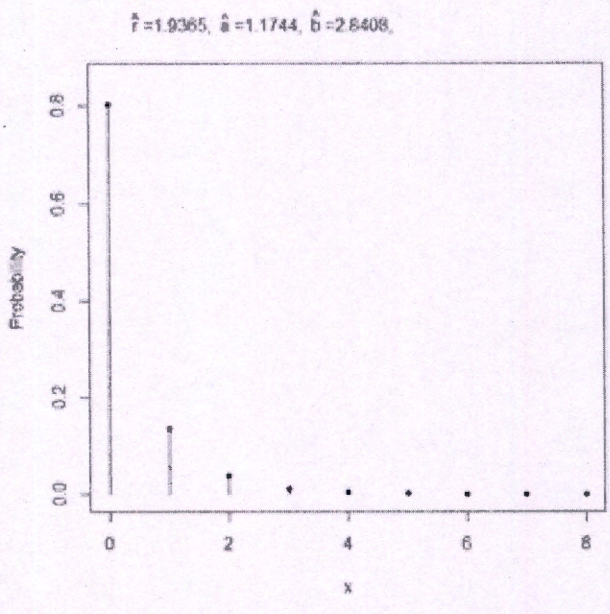
รูปภาพที่ 14 ค่า MSE เมื่อ $X \sim$ NB-NWL$(r,a,b) = 5,10,3$



รูปภาพที่ 15 ค่า MSE เมื่อ $X \sim$ NB-NWL$(r,a,b) = 10,10,3$

1.4 การศึกษาเชิงประยุกต์

การศึกษาเชิงประยุกต์ (Application study) พบว่า จากตารางที่ 4.1 แสดงค่าประมาณพารามิเตอร์และค่าคาดหวังที่หาได้จากการแจกแจง ZIP ZINB และ NB-NWL เกณฑ์ที่ใช้เลือกตัวแบบสำหรับการศึกษา คือ AIC (Akaike information criteria) และ BIC (Bayesian information criteria) และใช้การทดสอบสารูปสนิทดี (Goodness of Fit Test (Chi-Squared Test)) เพื่อเปรียบเทียบค่าสังเกต (Observed Value) และค่าคาดหวัง (Expected Value) ของข้อมูลจากผลการศึกษาพบว่า การแจกแจง NB-NWL มีค่า AIC และ BIC น้อยกว่าการแจกแจง ZIP และ ZINB นั่นแสดงว่าข้อมูลชุดนี้ควรเลือกใช้ตัวแบบ NB-NWL สอดคล้องกับค่าพีของการแจกแจง NB-NWL มีค่ามากกว่าการแจกแจง ZIP และ ZINB นั่นแสดงว่า การแจกแจง NB-NWL ให้ค่าคาดหวังที่ใกล้เคียงกับค่าสังเกตของข้อมูลชุดนี้มากกว่าการแจกแจง ZIP และ ZINB



รูปภาพที่ 16 ฟังก์ชันความน่าจะเป็นของการแจกแจง NB-NWL สำหรับค่าพารามิเตอร์ที่ถูกกำหนดไว้บางค่าของข้อมูลจริงที่ใช้ในการศึกษาเชิงประยุกต์

ตารางที่ 1 ค่าสังเกตและค่าคาดหวังของจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาตัวที่โรงพยาบาลราชบุรี

No. of hospitalization	No. of cases	Expected value by fitting distribution		
		Poisson	NB	NB-NWL
0	35	261.2574	73.5711	34.3315
1	275	449.3630	155.7058	189.0602
2	357	386.4520	205.2514	171.5407
3	249	221.5659	215.9518	147.2508
4	149	95.2733	198.4813	124.3442
5	77	32.7740	166.5815	104.7453
6	53	9.3952	130.9441	88.5186
7	50	2.3085	97.9537	75.2206
8	35	0.4963	70.4826	64.3310
9	32	0.0949	49.1530	55.3816
10	16	0.0163	33.4063	47.9862
11	15	0.0026	22.2195	41.8357
12	8	0.0004	14.5100	36.6865
13	15	0.0000	9.3271	32.3472
14	4	0.0000	5.9136	28.6669
15	7	0.0000	3.7045	25.5265
16	10	0.0000	2.2959	22.8310
17	6	0.0000	1.4095	20.5048
18	7	0.0000	0.8579	18.4868
19	5	0.0000	0.5182	16.7277
20	7	0.0000	0.3108	15.1872
21	2	0.0000	0.1853	13.8323
22	5	0.0000	0.1098	12.6358
23	2	0.0000	0.0647	11.5752
24	4	0.0000	0.0380	10.6315
25	2	0.0000	0.0222	9.7890
26	2	0.0000	0.0129	9.0345
27	2	0.0000	0.0075	8.3566
28	4	0.0000	0.0043	7.7458
29	3	0.0000	0.0025	7.1940
30+	21	0.0000	0.0014	6.6942
Total	1459			
Parameter estimates		$\hat{\lambda} = 1.72$	$\hat{r} = 4.07$ $\hat{p} = 0.48$	$\hat{r} = 4.15$ $\hat{\alpha} = 0.52$ $\hat{\beta} = 2.01$
log-likelihood		-1140.449	-1014.642	-825.985
K-S test		0.319	0.013	0.018
p-value		<0.001	0.086	0.572

2. การแจกแจง HNB-NWL

แนวทางแบบคลาสสิกสำหรับการวิเคราะห์ข้อมูลการนับคือการแจกแจงปัวซง อย่างไรก็ตามข้อมูลการนับมักจะแสดงการกระจายตัวมากและศูนย์ที่แท้จริงเกิดขึ้นที่อาจจะเกิดเหตุการณ์นั้นแต่ไม่พบ หรือศูนย์ที่เกิดจากการไม่เกิดขึ้นของเหตุการณ์นั้นจริง ๆ ในปัจจุบันการนำไปสู่การพัฒนาเทคนิคทางเลือกสำหรับการสร้างแบบจำลองข้อมูลการนับ ได้มีการพัฒนาขึ้นเพื่อแก้ปัญหาและเพื่อให้เกิดประสิทธิภาพในการประมาณค่าได้มากขึ้น สำหรับตัวแบบการแจกแจงที่มีศูนย์มากกรณีที่การเกิดศูนย์นั้นเป็นศูนย์ที่แท้จริง มีตัวแบบการแจกแจงคือ

$$g(x)=\begin{cases}\omega & ,x=0\\ \dfrac{(1-\omega)h(x)}{1-h(0)} & ,x=1,2,...,\end{cases}$$

โดยที่ $1-\omega$ คือโอกาสที่ไม่เกิดศูนย์ขึ้นในเหตุการณ์

$h(x)$ คือการแจกแจงที่สนใจและ

$h(0)$ คือการแจกแจงที่สนใจเมื่อเกิดเหตุการณ์ที่มีศูนย์

บทนิยาม 9 ให้ตัวแปรสุ่ม X มีการแจกแจง Hurdle Negative Binomial - New Weighted Lindley (HNB-NWL) ด้วยพารามิเตอร์ $X \sim HNB\text{-}NWL(r,a,b,\omega)$, มีฟังก์ชันการแจกแจงความน่าจะเป็นของ X คือ

$$g(x)=\begin{cases}\omega & ,x=0\\ \dfrac{(1-\omega)\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^{j}D\left(\dfrac{b+r+j+1}{(b+r+j)^2}-\dfrac{b+ab+r+j+1}{(b+ab+r+j)^2}\right)}{1-D\left(\dfrac{b+ab+r+1}{(b+ab+r)^2}-\dfrac{b+r+1}{(b+r)^2}\right)} & ,x=1,2,...\end{cases}$$

โดยที่ $D=\dfrac{b^2(1+a)^2}{ab(1+a)+a(2+a)}$ และพารามิเตอร์ $r,a,b>0,\ 0<\omega<1$.

คุณสมบัติเบื้องต้นของการแจกแจง

1. $g(x)\geq 0$ ทุกค่าของ x และ $g(x)\geq 0$

2. $\sum_{x=0}^{\infty} g(x)=1$ จาก

$$\sum_{x=0}^{\infty} g(x) = \sum_{x=0}^{\infty}\left(\omega + (1-\omega)\frac{h(x>0)}{1-h(x=0)}\right)$$

$$= \omega + (1-\omega)\frac{\left(\sum_{x=0}^{\infty} h(x \geq 0) - h(x=0)\right)}{1-h(x=0)}$$

$$= \omega + (1-\omega)\frac{\left(\sum_{x=0}^{\infty}\binom{r+x-1}{x}\exp(-\lambda r)(1-\exp(-\lambda)-1)^{x} - h(x=0)\right)}{1-h(x=0)}$$

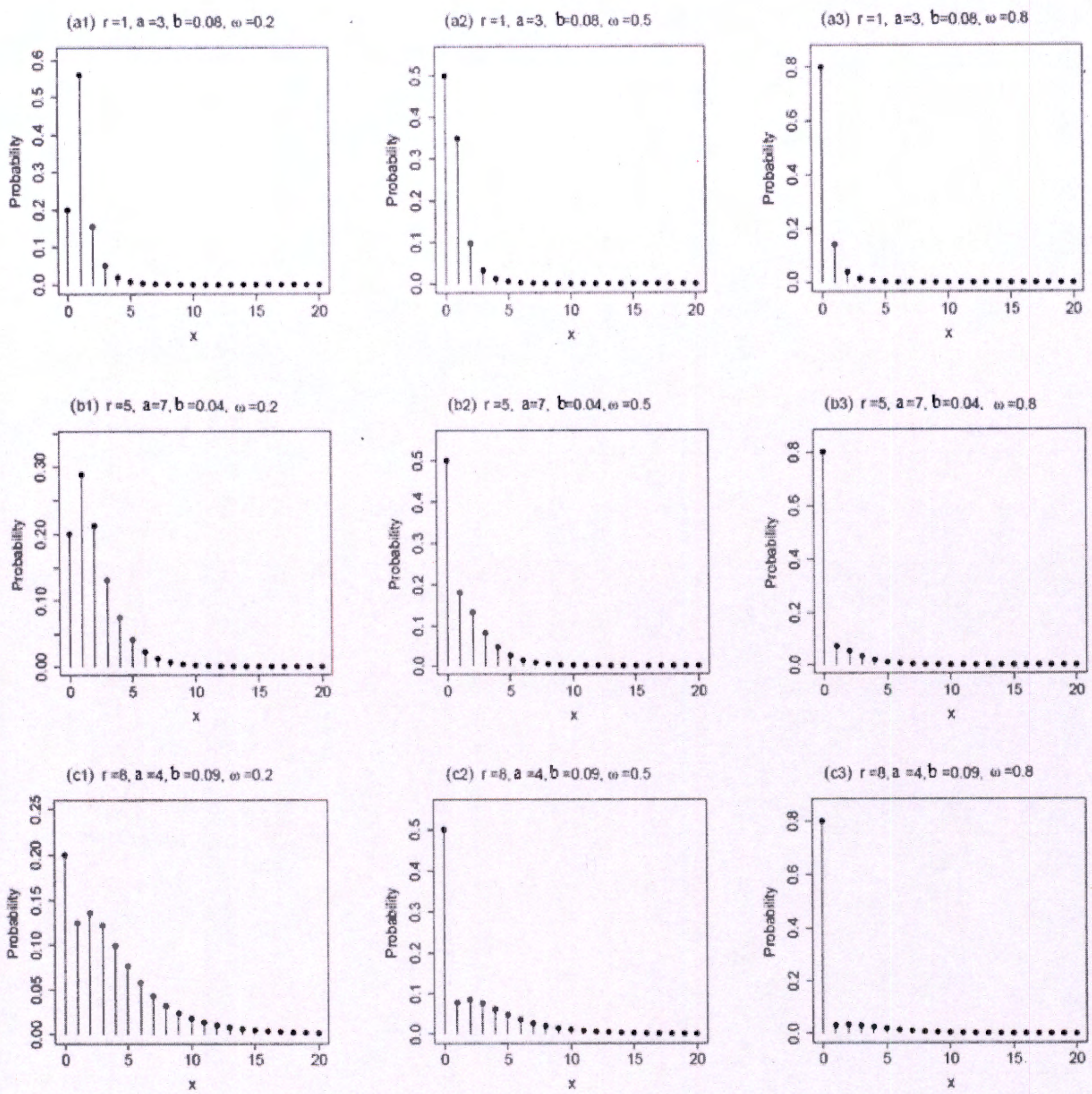
$$= \omega + (1-\omega)\frac{\exp(-\lambda r)(1+(\exp(-\lambda)-1)^{-r} - h(x=0)}{1-h(x=0)}$$

$$= \omega + (1-\omega)\frac{\exp(-\lambda r)\exp(\lambda r) - h(x=0)}{1-h(x=0)}$$

$$= \omega + (1-\omega)$$

$$= 1$$

รูปภาพที่ 17 แสดงกราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง HNB-NWL สำหรับค่าพารามิเตอร์ที่กำหนดไว้ ซึ่งพบว่าการแจกแจงจะหนาแน่นจะมีรูปร่างแตกต่างกัน เมื่อพารามิเตอร์เปลี่ยนไป



รูปภาพที่ 17 กราฟของฟังก์ชันความน่าจะเป็นของการแจกแจง HNB-NWL

บทนิยาม 10 ถ้า $X \sim HNB\text{-}NWL(r,a,b,\omega)$, ฟังก์ชันรอบจุดกำเนิด k^{th} ของตัวแปรสุ่ม X คือ

$$\mu_k'(X)=\frac{1-\omega}{1-D\left(\frac{b+ab+r+1}{(b+ab+r)^2}-\frac{b+r+1}{(b+r)^2}\right)}\sum_{x=1}^{\infty}x^k\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j D\left(\frac{b+r+j+1}{(b+r+j)^2}-\frac{b+ab+r+j+1}{(b+ab+r+j)^2}\right),$$

$$D=\frac{b^2(1+a)^2}{ab(1+a)+a(2+a)} \text{ for } k=1,2,....$$

นอกจากนี้ค่าเฉลี่ยและความแปรปรวนตามลำดับคือ

$$E(X) = \frac{(1-\omega)(r\ (\phi_1 - 1))}{1-D\left(\dfrac{b+ab+r+1}{(b+ab+r)^2}-\dfrac{b+r+1}{(b+r)^2}\right)},$$

$$Var(X) = \frac{(1-\omega)(r(r+1)\phi_2 - r\ (1+r\phi_1)\phi_1)}{1-D\left(\dfrac{b+ab+r+1}{(b+ab+r)^2}-\dfrac{b+r+1}{(b+r)^2}\right)}.$$

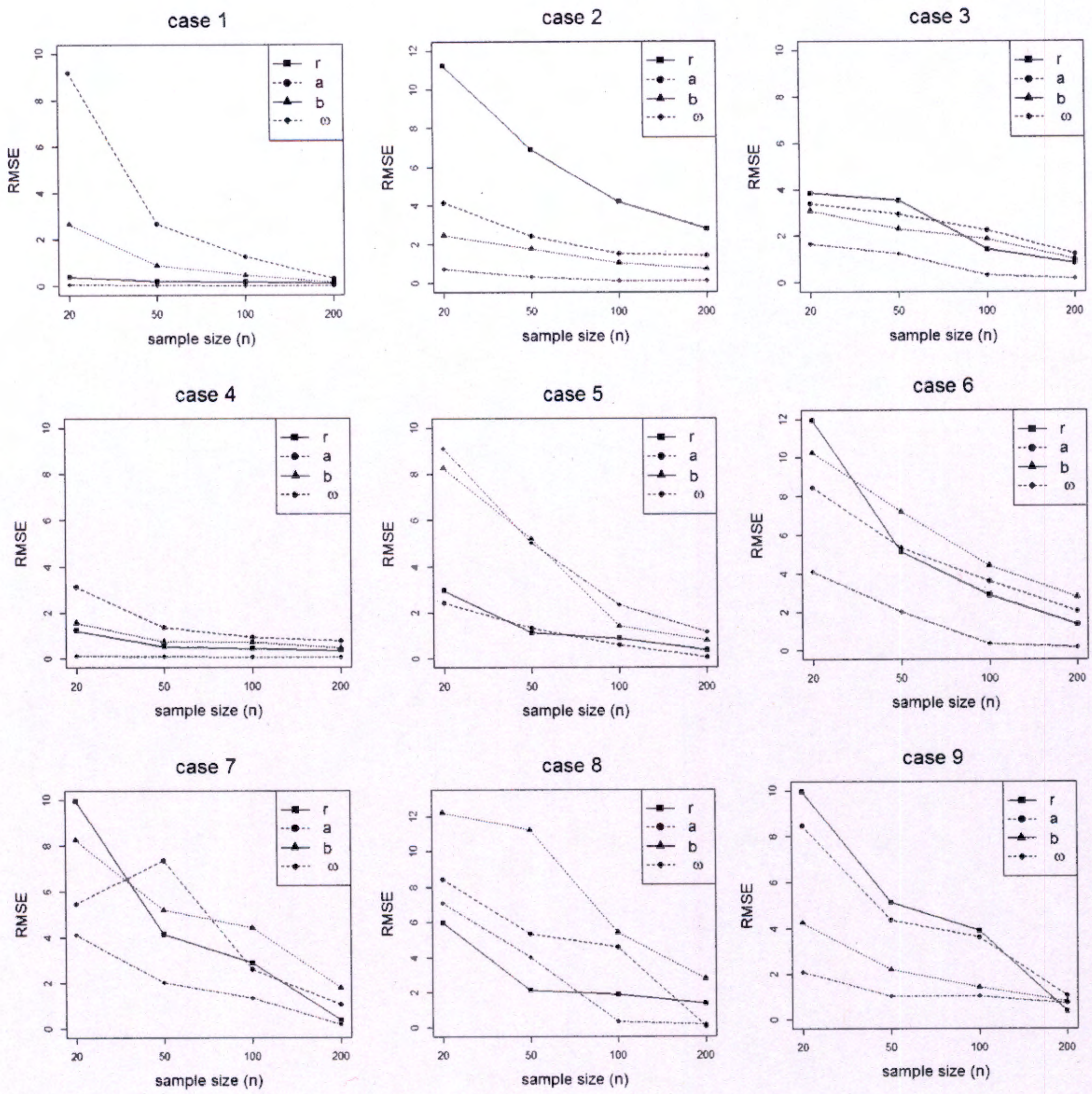
โดยที่ $D=\dfrac{b^2(1+a)^2}{ab(1+a)+a(2+a)}$ และ $\phi_k=\dfrac{b^2(1+a)^2}{ab(1+a)+a(2+a)}\left\{\dfrac{b-k+1}{(b-k)^2}-\dfrac{b(1+a)-k+1}{\left[b(1+a)-k\right]^2}\right\}, k=1,2$

2.1 การศึกษาจำลองสถานการณ์

การศึกษาจำลองสถานการณ์ (Simulation Study) จากตัวอย่างข้อมูลจะถูกสร้างจากการแจกแจง HNB-NWL ด้วยพารามิเตอร์ที่กำหนดไว้ 9 กรณี ได้แก่

ตารางที่ 2 การศึกษาจำลองสถานการณ์จากตัวอย่างข้อมูลจะถูกสร้างจากการแจกแจง HNB-NWL

Cases	r	a	b	ω	E(X)	V(X)
1	1	3	0.8	0.2	16.97	2.86
2	1	3	0.8	0.5	0.88	2.44
3	1	3	0.8	0.8	2.48	1.24
4	5	7	0.4	0.2	68.23	9.29
5	5	7	0.4	0.5	42.65	7.44
6	5	7	0.4	0.8	17.06	3.64
7	8	4	0.3	0.2	32.34	10.32
8	8	4	0.3	0.5	18.96	9.03
9	8	4	0.3	0.8	15.59	4.65



รูปภาพที่ 18 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าจากการจำลองแบบของการแจกแจง $X \sim HNB\text{-}NWL(r,a,b,\omega)$.

แต่ละกรณีจะใช้สัดส่วนของศูนย์ (Proportion of zero : ω) 3 ค่า คือ 0.2, 0.5 และ 0.8 โดยใช้ขนาดตัวอย่าง (Sample Sizes : n) เท่ากับ 20, 50, 100 และ 200 ตามลำดับ ในแต่ละสถานการณ์จะทำการประมาณค่าซ้ำ 500 รอบ จากผลการศึกษาพบว่าวิธี MLE จะมีประสิทธิภาพสูงเมื่อขนาดตัวอย่างใหญ่ และจะมีประสิทธิภาพต่ำเมื่อขนาดตัวอย่างเล็ก และเมื่อพิจารณาค่า MSE พบว่าจะมีค่าเพิ่มขึ้นเมื่อขนาดสัดส่วนของศูนย์สูงขึ้นจะได้รูปภาพที่ 18

2.2 การศึกษาเชิงประยุกต์

การศึกษาเชิงประยุกต์ (Application study) พบว่าจากตารางที่ 3 แสดงค่าประมาณพารามิเตอร์และค่าคาดหวังที่หาได้จากการแจกแจง HP, HNB, NB-NWL และ HNB-NBL โดยมีเกณฑ์ที่ใช้เลือกตัวแบบสำหรับการศึกษา คือ AIC (Akaike information criteria) และ BIC (Bayesian information criteria) และใช้การทดสอบสารูปสนิทดี (goodness of fit test (chi-squared test)) เพื่อเปรียบเทียบค่าสังเกต (observed value) และค่าคาดหวัง (expected value) ของข้อมูลจากผลการศึกษาพบว่า การแจกแจง HNB-NBL มีค่า AIC และ BIC น้อยกว่าการแจกแจง HP, HNB, NB-NWL นั่นแสดงว่าข้อมูลชุดนี้ควรเลือกใช้ตัวแบบ HNB-NBL สอดคล้องกับค่าพีของการแจกแจง HNB-NBL มีค่ามากกว่าการแจกแจง HP, HNB, NB-NWL นั่นแสดงว่า การแจกแจง HNB-NBL ให้ค่าคาดหวังที่ใกล้เคียงกับค่าสังเกตของข้อมูลชุดนี้มากกว่าการแจกแจง HP, HNB และ NB-NWL

ตารางที่ 3 ค่าสังเกตและค่าคาดหวังของจำนวนผู้ป่วยโรคเบาหวานที่เข้ารับการรักษาตัวที่โรงพยาบาลราชบุรี

No. of hospitalization	No. of cases	Expected value by fitting distribution			
		HP	HNB	NB-NWL	HNB-NBL
0	35	227.294	42.9712	34.3315	35.2000
1	275	348.240	108.1082	189.0602	257.3264
2	357	301.240	142.7810	171.5407	374.5123
3	249	110.240	148.4853	147.2508	224.2159
4	149	85.125	134.9065	124.3442	124.9979
5	77	28.513	110.9387	104.7453	74.7692
6	53	8.174	92.2798	88.5186	30.3299
7	50	2.008	66.5267	75.2206	39.3059
8	35	0.432	45.7475	64.3310	34.8999
9	32	0.083	32.3902	55.3816	24.8759
10	16	0.014	20.0525	47.9862	14.3318
11	15	0.002	14.9980	41.8357	15.1872
12	8	0.000	10.4472	36.6865	7.9327
13	15	0.000	6.7155	32.3472	14.2359
14	4	0.000	4.2578	28.6669	3.4155

ตารางที่ 3 (ต่อ)

No. of hospitalization	No. of cases	Expected value by fitting distribution			
		HP	HNB	NB-NWL	HNB-NBL
15	7	0.000	2.6672	25.5265	7.2272
16	10	0.000	1.6530	22.8310	9.7388
17	6	0.000	1.0148	20.5048	3.3424
18	7	0.000	0.6177	18.4868	6.1387
19	5	0.000	0.3731	16.7277	4.8105
20	7	0.000	0.2238	15.1872	5.5158
21	2	0.000	0.1334	13.8323	1.8884
22	5	0.000	0.0791	12.6358	4.3274
23	2	0.000	0.0466	11.5752	2.8561
24	4	0.000	0.0274	10.6315	1.6738
25	2	0.000	0.0160	9.7890	1.5439
26	2	0.000	0.0093	9.0345	0.5880
27	2	0.000	0.0054	8.3566	0.6684
28	4	0.000	0.0031	7.7458	0.6230
29	3	0.000	0.0018	7.1940	2.6344
30+	21	0.000	0.0010	6.6942	17.9383
Total	1459				
Parameter estimates		$\hat{\lambda} = 1.42$ $\hat{\omega} = 0.24$	$\hat{r} = 3.27$ $\hat{p} = 0.78$ $\hat{\omega} = 0.11$	$\hat{r} = 4.15$ $\hat{\alpha} = 0.52$ $\hat{\beta} = 2.01$	$\hat{r} = 2.47$ $\hat{a} = 0.98$ $\hat{b} = 1.68$ $\hat{\omega} = 0.11$
log-likelihood		-1240.449	-978.642	-825.985	-525.985
K-S test		0.519	0.053	0.018	0.007
p-value		<0.001	0.026	0.572	0.789

สรุปผลการวิจัยและข้อเสนอแนะ

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาการแจกแจง Negative Binomial - New Weighted Lindley Distribution (NB-NWL) และ Hurdle Negative Binomial - New Weighted Lindley Distribution (HNB-NWL) ตลอดจนศึกษาคุณสมบัติเชิงความน่าจะเป็น เช่น ค่าเฉลี่ย และความแปรปรวน นอกจากนี้ยังศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL และ HNB-NWL ด้วยวิธีภาวะน่าจะเป็นสูงสุด และประยุกต์ใช้การแจกแจง NB-NWL และ HNB-NWL ในการวิเคราะห์ข้อมูลเชิงนับ ซึ่งผู้วิจัยสรุปผลการวิจัยและข้อเสนอแนะ ดังรายละเอียดต่อไปนี้

สรุปผลการวิจัย

การแจกแจง Negative Binomial - New Weighted Lindley Distribution (NB-NWL) ได้มาจากการนำสองการแจกแจงมาผสมกัน ได้แก่ การแจกแจงเบอร์นูลลี (Bernoulli) และการแจกแจงทวินามลบ-ลินด์เลย์ถ่วงน้ำหนักใหม่ (NB-NWL) และ Hurdle Negative Binomial - New Weighted Lindley Distribution (HNB-NWL) เป็นการนำการแจกแจง NBWL มาศึกษาเพิ่มเติมเมื่อการแจงแจงนั้นมีค่าสังเกตที่เกิดศูนย์ที่แท้จริงมากเกินไป โดยทำการศึกษาคุณลักษณะของการแจกแจง เช่น ค่าเฉลี่ย ความแปรปรวน และทำการประมาณพารามิเตอร์ด้วยวิธี MLE ซึ่งวิธีการประมาณค่าดังกล่าวจะมีประสิทธิภาพสูงเมื่อขนาดตัวอย่างใหญ่ การแจกแจง NB-NWL และ HNB-NWL จะเป็นทางเลือกสำหรับการวิเคราะห์ข้อมูลเชิงนับที่มีศูนย์มากและเกิดปัญหาที่ค่าความแปรปรวนมากกว่าค่าเฉลี่ย

การศึกษาจำลองสถานการณ์ผลการศึกษาพบว่าวิธี MLE จะมีประสิทธิภาพสูงเมื่อขนาดตัวอย่างใหญ่ และจะมีประสิทธิภาพต่ำเมื่อขนาดตัวอย่างเล็กและเมื่อพิจารณาค่า MSE พบว่าจะมีค่าเพิ่มขึ้นเมื่อขนาดสัดส่วนของศูนย์สูงขึ้น

การศึกษาเชิงประยุกต์ จากผลการศึกษาพบว่าการแจกแจง NB-NWLและ HNB-NWL มีค่า AIC และ BIC น้อยกว่าการแจกแจงดั้งเดิม นั่นแสดงว่าข้อมูลชุดนี้ควรเลือกใช้ตัวแบบ NB-NWL สอดคล้องกับค่าพีของการแจกแจง NB-NWL มีค่ามากกว่าการแจกแจง ZIP และ ZINB และการแจกแจง HNB-NBL มีค่า AIC และ BIC น้อยกว่าการแจกแจง HP, HNB, NB-NWL นั่นแสดงว่าข้อมูลชุดนี้ควรเลือกใช้ตัวแบบ HNB-NBL สอดคล้องกับค่าพีของการแจกแจง HNB-NBL มีค่า

มากกว่าการแจกแจง HP, HNB, NB-NWL นั่นแสดงว่า การแจกแจง NB-NWL และ HNB-NBL ให้ค่าคาดหวังที่ใกล้เคียงกับค่าสังเกตของข้อมูลมากกว่าการแจกแจงแบบดั้งเดิม

ข้อเสนอแนะ

ในการวิจัยนี้เป็นการศึกษาการแจกแจง Negative Binomial - New Weighted Lindley Distribution (NB-NWL) และ Hurdle Negative Binomial - New Weighted Lindley Distribution (HNB-NWL) ศึกษาคุณลักษณะของการแจกแจง เช่น ค่าเฉลี่ย ความแปรปรวน และทำการประมาณพารามิเตอร์ด้วยวิธี MLE ซึ่งวิธีการประมาณค่าดังกล่าวจะมีประสิทธิภาพสูงเมื่อขนาดตัวอย่างใหญ่ ในการวิจัยครั้งต่อไปอาจศึกษาตัวแบบถดถอยเมื่อตัวแปรที่ศึกษามีการแจกแจง NB-NWL และ HNB-NBL และอาจศึกษาวิธีประมาณค่าพารามิเตอร์ด้วยวิธีเบส์ (Bayesian approach) และนำการแจกแจงเหล่านี้ไปประยุกต์ใช้กับข้อมูลจริงที่ผู้วิจัยสนใจศึกษา

บรรณานุกรม

รายงานการประชุม. 2561. **การตรวจราชการและนิเทศงานกระทรวงสาธารณสุขโรงพยาบาลราชบุรี ประจำปีงบประมาณ 2561.**

รายงานการประชุม. 2560. **สรุปผลการนิเทศโรงพยาบาลสันทรายประจำปี 2560.**

สมาคมโรคเบาหวานแห่งประเทศไทยฯ. 2560. **แนวทางเวชปฏิบัติสำหรับโรคเบาหวาน พ.ศ. 2560.** จาก http://www.dmthai.org/ [30 กันยายน 2560]

Bodhisuwan, W. 2011. Zero inflated Waring Distribution and its application. *In* **Proceedings of the 37th Congress on Science and Technology of Thailand**.The Science Society of Thailand under the Patronage of His Majesty the King and Faculty of Science, Mahidol University, Thailand.

Famoye, F. and K.P. Singh. 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. **Journal of Data Science**.4(1): 117–130.

Flynn, M. 2009. More flexible GLMs zero-inflated models and hybrid models. **Casualty Actuarial Society E-Forum, Winter**. pp. 148-224.

G´omez-D´eniz E., J.M. Sarabia and E. Calder´ın-Ojeda. 2008. Univariate and multivariate versions of the Negative Binomial-inverse Gaussian Distributions with applications. **Insurance: Mathematics and Economics**. 42: 39–49.

Greene, W.H. 1994. Accounting for excess zeros and sample selection in Poisson and Negative Binomial regression models. **Working Papers EC-94-10**. Department of Economics, New York University.

Hall, D. 2000. Zero-inflated Poisson and Binomial regression with random effects: a case study. **Biometrics**. 56: 1030–1039.

Jansakul, N. and J.P. Hinde. 2008. Score Tests for Extra-Zero Models in Zero-Inflated Negative Binomial Models. **Communications in Statistics - Simulation and Computation**. 38: 92–108.

Lambert, D. 1992. Zero-inflated Poisson regression with application to defects in manufacturing. **Technometrics**. 34: 1–14.

Meng, S., Y. Wei and G. Whitmore. 1999. Accounting for individual overdispersion in a Bonus - Malus system. **Astin Bulletin**. 29(2): 327–337.

Panjer, H. and G. Willmot. 1981. Finite sum evaluation of the Negative Binomial exponential model. **Astin Bulletin**. 12: 133–137.

Pudprommarat, C., W. Bodhisuwan and P. Zeephongsekul. 2012. A new mixed negative Binomial Distribution. **Journal Applied Sciences**. 12(17): 1853–1858.

R Core Team. 2012. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria.

Ridout, M., C.G. Dem´etrio and J. Hinde. 1998. Models for count data with many zeros, pp. 179–192. *In* **Proceedings of the XIXth international biometric conference**, vol. 19.

Ridout, M., J. Hinde and C.G. DeméAtrio. 2001. A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. **Biometrics**. 57(1): 219-223.

Wang, Z. 2011. One mixed Negative Binomial Distribution with application. **Journal of Statistical Planning and Inference.** 141(3): 1153-1160.

Yip, K.C. and K.K. Yau. 2005. On modeling claim frequency data in general insurance with extra zeros. **Insurance: Mathematics and Economics**, 36(2): 153-163.

Zamani, H. and N. Ismail. 2010. Negative Binomial Lindley Distribution and its application. **Jonrual Mathematics and Statistics**. 6(1): 4–9.

ภาคผนวก

ตารางผนวกที่ 1 ค่าประมาณพารามิเตอร์ของการแจกแจง NB-NWL ของพารามิเตอร์ $r = 5$, $a = 3$, $b = 3$

n	**Parameter**	**Estimate**	**SD**	**MSE**
20	r	20.89	19.32	625.85
	a	17.10	20.34	612.55
	b	11.92	13.92	273.40
50	r	15.25	13.61	290.25
	a	11.35	14.77	288.10
	b	7.70	9.49	112.06
100	r	10.23	6.19	65.75
	a	5.34	6.77	51.26
	b	5.20	3.99	20.75
200	r	8.41	3.89	26.75
	a	4.15	3.08	10.81
	b	4.00	2.33	6.42

ตารางผนวกที่ 2 ค่าประมาณพารามิเตอร์ของการแจกแจง NB-NWL ของพารามิเตอร์ $r = 5$, $a = 3$, $b = 3$

n	**Parameter**	**Estimate**	**SD**	**MSE**
20	r	22.80	17.52	623.69
	a	17.96	18.07	550.64
	b	13.08	12.31	253.27
50	r	17.56	15.92	411.15
	a	11.78	16.27	341.90
	b	8.48	9.47	119.71
100	r	11.12	7.78	97.97
	a	6.31	8.32	80.19
	b	5.78	4.85	31.31
200	r	9.81	4.96	47.70
	a	4.38	4.67	23.69
	b	4.57	3.16	12.42

ตารางผนวกที่ 3 ค่าประมาณพารามิเตอร์ของการแจกแจง NB-NWL ของพารามิเตอร์ $r = 5$, $a = 3$, $b = 3$

n	Parameter	Estimate	SD	MSE
20	r	26.29	21.34	908.85
	a	21.76	21.10	797.19
	b	17.68	17.39	517.88
50	r	21.08	18.57	603.34
	a	15.06	19.30	518.09
	b	10.94	11.54	196.09
100	r	15.30	11.83	246.05
	a	9.52	12.52	199.32
	b	7.53	7.22	72.64
200	r	11.29	7.46	95.27
	a	5.84	8.20	75.30
	b	5.44	4.84	29.35

ตารางผนวกที่ 4 ค่าประมาณพารามิเตอร์ของการแจกแจง NB-NWL ของพารามิเตอร์ $r = 5$, $a = 10,\ b = 5$

n	Parameter	Estimate	SD	MSE
20	r	18.02	18.45	509.76
	a	24.06	17.17	492.39
	b	12.53	9.65	149.77
50	r	7.86	8.69	83.79
	a	22.03	15.22	376.31
	b	10.22	7.37	81.61
100	r	6.78	6.79	49.23
	a	21.09	14.33	328.47
	b	9.71	7.03	71.65
200	r	5.68	3.99	16.36
	a	19.36	13.36	266.02
	b	8.66	6.78	59.40

การสร้างตัวแปรสุ่ม NB-NWL โดยใช้โปรแกรม R

****

R code of Random generation for NB-BE Distribution

****

```
rnbbe <- function(n,r,a,b,c)
{
    rbetaexp <- function(n,a,b,c)
        {
            x <- rep(0,n)
            k <- rbeta(n,a,b)
            for(i in 1:n)    {
                x[i] <- (-1/c)*log(-k[i]+1)
                             }
            sample<-x
        }
    lambda <- rbetaexp(n,a,b,c)
    v <- rnbinom(n,size=r,prob=exp(-lambda))
    sample <- v
}
```

****

****

R code of Random generation for NB-NWL Distribution

****

****

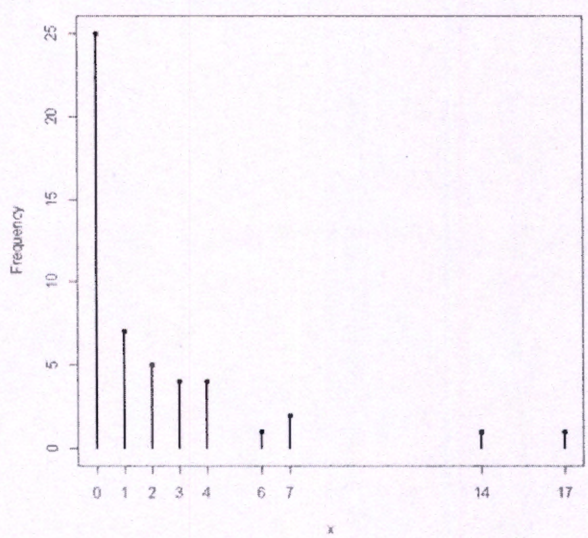
```
rzinbbe <- function(n,r,a,b,c,phi)
{
    x <- rep(0,n)
    for (i in 1:n)
        {
            U<-runif(1,0,1)
            V<-rnbbe(1,r,a,b,c)
            If (U >= phi)    {
                x[i]<-V
                         }
            else if (U< phi)  {
                x[i]<-0
                         }
        }
    v <- x
    sample <- v
}
```

****
****

ตัวอย่างการสร้างตัวแปรสุ่ม NB-NWL โดยใช้โปรแกรม R

****
****

```
> n <- 50; r<- 3; a <- 5; b <- 2; c <- 2; theta <- 0.2
> x <- rzinbbe(n,r,a,b,c,theta)
> x
[1] 0 0 1 0 4 0 1 17 0 0 2 0 0 3 0 0 1 1 0 0 0 0 3 2 0 3 0 4 6 14 0 3 2
[34] 0 2 0 0 4 0 4 7 7 0 1 0 1 0 0 2 1
> Frequency <- table(x)
> Frequency
x
 0  1  2  3  4  6  7 14 17
25  7  5  4  4  1  2  1  1
> plot (Frequency)
> points (Frequency,cex =1.1,pch=20, type="p")
```



****

****

ฟังก์ชันความหนาแน่น NB-NWL โดยใช้โปรแกรม R

****

****

```
zinbbe <- function(x,r,a,b,c,phi)
{
        if (x==0)        {
                p <-     phi+(1-phi) * (factorial(r+x-1)/( factorial(r-1)*factorial(x)))
                         *(beta(b+(r/c),a)/beta(a,b))
                         }
        else if (x>0)    {
                pp1 <- beta(b+(r/c),a)
                for (j in 1:x)      {
        p1 <- ((factorial(x)/( factorial(x-j)*factorial(j)))*((-1)^j)*(beta(b+((r+j)/c),a)))
        pp1 <- pp1+p1
                                    }
        p <- (1-phi)*(factorial(r+x-1)/( factorial(r-1)*factorial(x)*beta(a,b)))*pp1
                         }
        p
}
```

****
****

การประมาณค่าพารามิเตอร์ของการแจกแจง NB-NWL ด้วยวิธี MLE โดยใช้โปรแกรม R

****
****

```
mlogl <- function(theta,x)
{
        fnbbe <- function(theta,x)
        {
                mm <- length (x)
                k <- numeric (mm)
                nbbe <- function (theta,x)
                {
        if (x==0)       {
                p<- -log(theta[5]+(1-theta[5])
                                *(beta(theta[3]+(theta[1]/theta[4]),theta[2])
                                /beta(theta[2],theta[3])))
                        }
                        else if (x>0)   {
                                pp1<-(beta(theta[3]+(theta[1]/theta[4]),theta[2])
                                /beta(theta[2],theta[3]))
                for (j in 1:x)  {
                        p1 <- ((factorial(x)/( factorial(x-j)*factorial(j)))
                                *((-1)^j)*(beta(theta[3]+((theta[1]+j)/theta[4]),
                                theta[2])/beta(theta[2],theta[3])))
                pp1 <- pp1+p1
                                }
                                p <- (-log(1-theta[5]))-log(factorial(theta[1]+x-1))+
                                log(factorial(theta[1]-1))+log(factorial(x))-log(pp1);
```

```
                                    }
                        p
                }
                for(i in 1 : length(x))      {
                        k[i]<-nbbe(theta,x[i])
                                    }
                k
        }
        sum(fnbbe(theta,x))
}
theta.start <- c(1,1,1,1,.1)
out <- nlm(mlogl, theta.start, x = x)
r_MLE <- out$estimate[1]
a_MLE <- out$estimate[2]
b_MLE <- out$estimate[3]
c_MLE <- out$estimate[4]
omega_MLE <- out$estimate[5]
```

ตัวอย่างการสร้างตัวแปรสุ่ม HNB-NWL โดยใช้โปรแกรม R

```
library(VGAM)
rhnbtwcr<-function(n,r,lambda,theta,omega)
{
rposnbtwcr<-function(n,r,lambda,theta)
{
rposnegbin<-function (n,size,prob=NULL,munb=NULL)
{
k<-r
size<-k
munb<-(k*((1-((theta/(1+theta))*(1-sqrt(1-(2*theta)))))
*exp(lambda*(1-sqrt(1-(2*theta))))
/(sqrt(1-(2*theta)))))-k
qnbinom(p=runif(n,min=dnbinom(0,size,mu=munb)),size,mu=munb)
}
X<-rposnegbin(n,size,prob=NULL,munb=NULL)
}
x<-rep(0,n)
{
for(i in 1:n){
u<-runif(1)
v<-rposnbtwcr(1,r,lambda,theta)
if(u<=omega){x[i]<-0}
if(u>omega){x[i]<-v}
}
}
sample<-x
}
```

ฟังก์ชันความหนาแน่น HNB-NWL โดยใช้โปรแกรม R

```
dhnbnwl<-function(x,r,a,b,omega)
281
{
mm<-length(x)
k<-numeric(mm)
hnbtwcr<-function(x,r,a,b,omega)
{
if (x==0){
p<-omega
}else
if (x>0){
pp1<-(exp((-lambda)*sqrt(1+(2*theta*r)))/sqrt(1+(2*theta*r)))
*(1-((theta/(1+theta))*(1-sqrt(1+(2*theta*r)))))
for(j in 1:x){
p1<-((factorial(x)/(factorial(x-j)*factorial(j)))*((-1)^j)
*(exp((-lambda)*sqrt(1+(2*theta*(r+j))))
/sqrt(1+(2*theta*(r+j))))
*(1-((theta/(1+theta))*(1-sqrt(1+(2*theta*(r+j)))))))
pp1<-pp1+p1
}
p<-(1-omega)*(factorial(r+x-1)/( factorial(r-1)
*factorial(x)))*(exp(lambda))*pp1
*(1/(1-((exp(lambda*(1-sqrt(1+(2*theta*r)))))
*(1-((theta/(1+theta))
*(1-sqrt(1+(2*theta*r)))))
/(sqrt(1+(2*theta*r))))))
}
p
```

การประมาณค่าพารามิเตอร์ของการแจกแจง HNB-NWL ด้วยวิธี MLE โดยใช้โปรแกรม R

```
library(stats)
mlogl<-function(theta,x)
{
fhnbtwcr<-function(theta,x)
{
mm<-length(x)
k<-numeric(mm)
hnbtwcr<-function(theta,x)
{
if (x==0){
p<-(-log(theta[4]))
}else
if (x>0){
pp1<-(exp((-theta[2])*sqrt(1+(2*theta[3]*theta[1])))
/sqrt(1+(2*theta[3]*theta[1])))
*(1-((theta[3]/(1+theta[3]))
*(1-sqrt(1+(2*theta[3]*theta[1])))))
for(j in 1:x){
p1<-((factorial(x)/(factorial(x-j)*factorial(j)))*((-1)^j)
*(exp((-theta[2])*sqrt(1+(2*theta[3]*(theta[1]+j))))
/sqrt(1+(2*theta[3]*(theta[1]+j))))
*(1-((theta[3]/(1+theta[3]))
*(1-sqrt(1+(2*theta[3]*(theta[1]+j)))))))
pp1<-pp1+p1
}
```