บทปริทัศน์เกี่ยวกับการแจกแจงทางสถิติที่น่าสนใจในคณิตศาสตร์ประกันภัย
Review of Interesting Statistical Distributions for Actuarial Science

วุธินันท์ จิรสิริพิพัฒน์
สาขาวิชาคณิตศาสตร์และภูมิสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี
Email: a.rajchavit42h1@gmail.com
เจษฎา ตัณฑนุช
สาขาวิชาคณิตศาสตร์และภูมิสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี
Email: jessada@g.sut.ac.th
ธิดารัตน์ อารีรักษ์
สาขาวิชาคณิตศาสตร์และภูมิสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี
Email: tidarut@g.sut.ac.th

บทคัดย่อ

บทความวิชาการนี้ต้องการนำเสนอบทปริทัศน์ที่เกี่ยวข้องกับการแจกแจงทางสถิติที่น่าสนใจที่นำมาใช้ในการอธิบายปรากฏการณ์หรือการสร้างตัวแบบทางคณิตศาสตร์ที่เกี่ยวข้องกับคณิตศาสตร์ประกันภัย ทั้งนี้ต้องการนำเสนอตั้งแต่การแจกแจงทางสถิติที่มีความซับซ้อนทางคณิตศาสตร์น้อยไปจนถึงการแจกแจงทางสถิติที่มีความซับซ้อนทางคณิตศาสตร์มากและพารามิเตอร์มากขึ้น ได้แก่ การแจกแจงปัวซง การแจกแจงทวินาม การแจกแจงทวินามเชิงลบ การแจกแจงเกาส์ การแจกแจงล็อกนอร์มัล การแจกแจงแกมมา การแจกแจงแบบเลขชี้กำลัง การแจกแจงเกาส์ผกผัน การแจกแจงไวบูล รวมไปถึงการแจกแจงห่อหุ้ม ทั้งนี้นอกจากการนำเสนอฟังก์ชันมวลความน่าจะเป็นและฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงเหล่านั้นแล้ว ยังแสดงให้เห็นถึงความสัมพันธ์ของการแจกแจงดังกล่าวกับสถานการณ์ที่เกี่ยวข้องกับคณิตศาสตร์ประกันภัยอีกด้วย

คำสำคัญ: คณิตศาสตร์ประกันภัย, การแจกแจงทางสถิติ, ตัวแบบทางคณิตศาสตร์

Abstract

This academic article aims to present a review of interesting statistical distributions used to explain phenomena or mathematical models related to actuarial science. It covers a range from statistically simple distributions to those with advanced mathematical complexity and more parameters. These include the Poisson distribution, binomial distribution, negative binomial distribution, Gaussian distribution, log-normal distribution, gamma distribution, exponential distribution, inverse-Gaussian distribution, Weibull distribution, and wrapped distributions. In addition, the probability mass functions and probability density functions of these distributions are presented and the article also demonstrates their relevance to situations in actuarial science.

Keywords: actuarial science, statistical distribution, mathematical model

บทนำ

การประกันภัยเป็นส่วนสำคัญของการวางแผนการเงินในชีวิต เนื่องจากช่วยลดความเสี่ยงทางการเงินในกรณีที่เกิดเหตุการณ์ไม่คาดคิด การประกันภัยตามประมวลกฎหมายแพ่งและพาณิชย์ จะแบ่งเป็น 2 ประเภท ได้แก่ การประกันชีวิต (life insurance) และการประกันวินาศภัย (non-life insurance) นอกจากนี้การประกันภัยบางประเภทสามารถทำหน้าที่เป็นกลไกการออม และค่าเบี้ยประกันบางส่วนยังสามารถหักลดหย่อนภาษีได้อีกด้วย การวางแผนค่าใช้จ่ายในการประกันภัยอย่างเป็นระบบช่วยเพิ่มสภาพคล่องทางการเงิน อย่างไรก็ตามผู้เอาประกันต้องชำระเบี้ยประกันล่วงหน้าเพื่อครอบคลุมความเสี่ยง ซึ่งมีผลต่อเงินสำรองในบางสถานการณ์ และผลตอบแทนอาจต่ำกว่าเมื่อเทียบกับการลงทุนบางประเภท ดังนั้นจึงจำเป็นต้องศึกษาวิจัยและเปรียบเทียบตัวเลือกความคุ้มครองและเข้าใจความคุ้มครองที่เหมาะสม ในทางกลับกันหากบริษัทประกันไม่ได้มีการวางแผนในการสำรองค่าสินไหมของการประกันภัยไว้เป็นอย่างดี ก็ส่งผลต่อธุรกิจในการให้ประกันภัยได้

ในประวัติศาสตร์คณิตศาสตร์ประกันภัย (actuarial science) ได้ถูกนำมาใช้เพื่อพัฒนาวิธีการที่แม่นยำยิ่งขึ้นในการกำหนดราคากรมธรรม์บำนาญและการประกันชีวิต แนวคิดนี้เริ่มต้นจาก Christiaan Huygens ที่ตีพิมพ์งานวิจัยที่เกี่ยวข้องกับทฤษฎีความน่าจะเป็น และ John Graunt ที่สร้างตารางมรณะในยุคแรกจากบันทึกการตายในลอนดอน ต่อมา Edmond Halley ได้นำหลักการทางคณิตศาสตร์เหล่านี้มาใช้ในการกำหนดราคากรมธรรม์บำนาญ เพื่อให้เกิดการทำประกันชีวิตที่มีหลักการทางตัวเลขอ้างอิง (Lewin, 2001) ในปีคริสตศักราช 1936 มหาวิทยาลัย University of Rome ได้ก่อตั้งคณะ The Faculty of Statistical, Demographic, and Actuarial Sciences ขึ้นมาด้วยวัตถุประสงค์หลัก 2 ข้อคือ 1) เพื่อให้เกิดพัฒนาองค์ความรู้และงานวิจัยทางด้านสถิติ ประชากรศาสตร์ และคณิตศาสตร์ประกันภัย และ 2) เตรียมความพร้อมนักศึกษาสำหรับตำแหน่งทางสถิติและคณิตศาสตร์ประกันภัย (Federici & Pedroni, 1951) เห็นได้ว่าคณิตศาสตร์ประกันภัยและการวิเคราะห์ด้านสถิติมีความสัมพันธ์กัน และได้ถูกพัฒนาโดยต่อเนื่องด้วยการสร้างแบบจำลองทางคณิตศาสตร์ที่ซับซ้อน และการประยุกต์ใช้การแจกแจงทางสถิติเพื่อให้สามารถกำหนดราคากรมธรรม์สำหรับการประกันชีวิตและการประกันวินาศภัยที่แม่นยำ การดำเนินการดังกล่าวใช้ความรู้ทางคณิตศาสตร์และสถิติเป็นเครื่องมือในการจำลองอัตราการเสียชีวิตและอายุขัย อัตราการเกิดอุบัติเหตุ หรือความเสื่อมของเครื่องมือและเครื่องจักร ซึ่งมีความสำคัญต่อการกำหนดเบี้ยประกัน การวิเคราะห์ความถี่ของการเรียกร้องค่าสินไหมทดแทน และการประเมินปริมาณของการเรียกร้องค่าสินไหมทดแทน (Bahnemann, 2015) ดังนั้นทฤษฎีความน่าจะเป็นและสถิติจะเข้ามามีบทบาทในการสร้างตัวแบบเป็นอย่างมาก

บทความนี้ต้องการนำเสนอบทปริทัศน์ที่เกี่ยวข้องกับการแจกแจงทางสถิติที่น่าสนใจที่นำมาใช้ในการอธิบายปรากฏการณ์หรือการสร้างตัวแบบทางคณิตศาสตร์ที่เกี่ยวข้องกับคณิตศาสตร์ประกันภัย

วัตถุประสงค์ของการวิจัย

งานวิจัยนี้ต้องการวิเคราะห์สมบัติของการแจกแจงทางสถิติที่สามารถนำไปประยุกต์ใช้ในงานทางด้านการสร้างตัวแบบทางคณิตศาสตร์ที่เกี่ยวข้องกับงานด้านการประกันภัย

การแจกแจงทางสถิติ

หลักการสำคัญทางด้านสถิติคือ การอธิบายปรากฏการณ์ต่าง ๆ ในธรรมชาติด้วยตัวเลข ซึ่งในทางสถิติเราสามารถเชื่อมโยงความสัมพันธ์ระหว่างปรากฏการณ์ต่าง ๆ นั้นกับตัวเลขด้วย **ตัวแปรสุ่ม** (random variable) และเพื่อที่จะวิเคราะห์ว่าตัวแปรสุ่มนั้นมีโอกาสที่จะเกิดค่าต่าง ๆ มากน้อยเพียงใด เราสามารถแสดงให้เห็นได้จาก**การแจกแจงทางสถิติ** (statistical distribution) การแจกแจงทางสถิติเป็นพื้นฐานสำคัญในการวิเคราะห์ข้อมูลและการสรุปผลทางสถิติ เพราะช่วยให้สามารถแสดงให้เห็นถึงค่า**พารามิเตอร์** (parameter) ต่าง ๆ ของข้อมูล เช่น **ค่าเฉลี่ย** (mean) เป็นค่าที่ใช้ในการบ่งบอกตำแหน่งศูนย์กลางของการแจกแจง **ค่ามัธยฐาน** (median) เป็นค่ากลางของข้อมูลเมื่อเรียงลำดับข้อมูลจากจากน้อยไปมาก **ค่าเบี่ยงเบนมาตรฐาน** (standard deviation) เป็นค่าที่ใช้วัดการกระจายของข้อมูลรอบ ๆ ค่าเฉลี่ย **ค่าความเบ้** (skewness) เป็นการวัดความไม่สมมาตร (asymmetry) ของการแจกแจงทางสถิติ (Krishnamoorthy, 2006)

การแจกแจงทางสถิติที่มีความสัมพันธ์กับงานทางด้านคณิตศาสตร์ประกันภัยที่น่าสนใจมีดังต่อไปนี้

การแจกแจงทวินาม

สำหรับการทดลองสุ่ม (trial) ที่เป็นอิสระต่อกัน n ครั้ง โดยแต่ละครั้งของการทดลองสุ่มจะเกิดผลลัพธ์ได้เพียง 2 อย่าง คือ สำเร็จ และล้มเหลว กำหนดให้ $0 \leq p \leq 1$ แทนความน่าจะเป็นที่จะเกิดผลสำเร็จในแต่ละครั้ง ดังนั้น ตัวแปรสุ่ม X ที่แทนจำนวนครั้งของการเกิดผลสำเร็จของการทดลองสุ่มนี้จะมีการแจกแจงทวินาม (binomial distribution) และความน่าจะเป็นของการเกิดผลสำเร็จ k ครั้ง จากการทดลองสุ่มทั้งหมด n ครั้ง สามารถอธิบายได้ด้วยฟังก์ชันมวลความน่าจะเป็น (mass probability function)

$$f(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,\dots,n$$

ในการวิเคราะห์ความเสี่ยง (risk management) การแจกแจงทวินามถูกนำมาใช้ในการประมาณความน่าจะเป็นของความสำเร็จหรือความล้มเหลวในด้านการลงทุนและการเรียกร้องค่าสินไหมทดแทนที่เกี่ยวข้องกับประกันภัย การแจกแจงนี้ช่วยให้บริษัทประกันภัยประเมินความเป็นไปได้ที่ผู้ถือกรมธรรม์จะยื่นเรียกร้องค่าสินไหมทดแทน ซึ่งเป็นสิ่งสำคัญสำหรับการตั้งราคาผลิตภัณฑ์ประกันภัยและการจัดการความเสี่ยง (Dionne, 2000)

การแจกแจงปัวซง

สำหรับตัวแปรสุ่ม X ที่แทนจำนวนครั้งของการเกิดเหตุการณ์ภายในหนึ่งช่วงเวลาที่กำหนดจะมีการแจกแจงปัวซง (Poisson distribution) ซึ่งมีฟังก์ชันมวลความน่าจะเป็นดังนี้

$$f(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0,1,2,\dots$$

เมื่อ λ เป็นค่าเฉลี่ยจำนวนครั้งของการเกิดเหตุการณ์ในหนึ่งช่วงเวลาที่สนใจ สำหรับการแจกแจงทวินามที่ n มีค่ามากและ p มีค่าน้อย สามารถประมาณด้วยการแจกแจงแบบปัวซองที่มีค่าพารามิเตอร์ $\lambda = np$

งานด้านประกันภัยสามารถใช้การแจกแจงแบบปัวซงเพื่อประมาณจำนวนอุบัติเหตุทางรถยนต์ต่อวันบนถนน จำนวนผู้เสียชีวิตในโรงพยาบาลต่อสัปดาห์ รวมไปถึงจำนวนการเรียกร้องค่าสินไหมทดแทนที่คาดหวังในช่วงระยะเวลาของแต่ละกรมธรรม์ การแจกแจงปัวซงสามารถประยุกต์ใช้ในการคำนวณความสูญเสียที่คาดการณ์ไว้ในช่วงเวลาหนึ่ง ซึ่งนำไปสู่ข้อมูลในการกำหนดเบี้ยประกันภัย (Krishnamoorthy, 2006)

การแจกแจงทวินามเชิงลบ

พิจารณาการทดลองสุ่มที่แต่ละครั้งของการทดลองเป็นอิสระต่อกัน และเกิดผลลัพธ์ได้เพียง 2 อย่าง คือ สำเร็จ และล้มเหลว ดังนั้น ตัวแปรสุ่ม X ที่แทนจำนวนครั้งของการทดลองสุ่มที่เกิดความล้มเหลวจนกระทั่งประสบความสำเร็จในครั้งที่ r^{th} จะมีการแจกแจงทวินามเชิงลบ (negative binomial distribution) ซึ่งมีฟังก์ชันมวลความน่าจะเป็นดังนี้

$$f(k|r,p) = \binom{r+k-1}{k} p^r (1-p)^k, k = 0,1,2,\dots$$

เมื่อ $0 < p < 1$ แทนความน่าจะเป็นที่จะเกิดผลสำเร็จในการทดลองสุ่มแต่ละครั้ง

การแจกแจงทวินามเชิงลบเป็นการแจกแจงทางสถิติที่ใช้ในการหาจำนวนครั้งที่ต้องทำการทดลอง (ที่ไม่สำเร็จ) จนกว่าจะได้ผลลัพธ์ที่ต้องการ (สำเร็จ) ครบตามจำนวนที่กำหนด การแจกแจงทางสถิตินี้จึงได้ถูกนำไปจำลองการเรียกร้องค่าสินไหมทดแทนในประกันภัยจนกว่าจะถึงจำนวนครั้งของการเรียกร้องค่าสินไหม (claim count) ที่กำหนด ตัวอย่างงานวิจัยที่มีการประยุกต์ใช้การแจกแจงนี้ เช่น Boucher et al. (2008) ได้นำเสนอแบบจำลองของจำนวนการเรียกร้องค่าสินไหมทดแทนในประกันภัยที่พิจารณาตามช่วงเวลาโดยใช้การแจกแจงปัวซงร่วมกับการแจกแจงทวินามเชิงลบ

การแจกแจงเกาส์

จากทฤษฎีบทขีดจำกัดส่วนกลาง (Central Limit Theorem) ที่กล่าวไว้ว่าเมื่อขนาดตัวอย่าง n มีจำนวนมากพอ ($n \geq 30$) การแจกแจงของค่าเฉลี่ยตัวอย่าง (sample mean) จะเข้าใกล้การแจกแจงเกาส์ (Gaussian distribution) โดยไม่คำนึงถึงการแจกแจงของประชากร ดังนั้นการแจกแจงเกาส์หรืออาจจะเรียกในอีกชื่อหนึ่งคือ การแจกแจงปรกติ (normal distribution) มีความสำคัญทางสถิติเป็นอย่างมาก ตัวแปรสุ่ม X ที่มีการแจกแจงแบบเกาส์มีฟังก์ชันความหนาแน่นความน่าจะเป็น (probability density function) คือ

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty$$

เมื่อ μ เป็นค่าเฉลี่ย ($-\infty < \mu < \infty$) σ เป็นค่าเบี่ยงเบนมาตรฐาน ($\sigma > 0$) และ $\exp(\cdot)$ เป็นฟังก์ชันเลขชี้กำลัง (exponential function)

การแจกแจงเกาส์มักถูกนำมาใช้ในการจำลองการแจกแจงของจำนวนเงินที่เรียกร้องค่าสินไหมทดแทนในประกันภัย โดยเฉพาะในบางประเภทของธุรกิจ เช่น ประกันภัยรถยนต์หรือประกันภัยที่อยู่อาศัย นอกจากนี้ยังถูกใช้ในการกำหนดราคาผลิตภัณฑ์ประกันภัยและการประมาณการสำรองสำหรับการเรียกร้องค่าสินไหมทดแทนในอนาคต การแจกแจงเกาส์เป็นองค์ประกอบสำคัญในแบบจำลองทางคณิตศาสตร์ประกันภัยสำหรับ

การวัดและจัดการความเสี่ยงของประกันภัย เช่น การคำนวณมูลค่าที่เสี่ยง (Value-at-Risk: VaR) และการคาดการณ์ความสูญเสียที่คาดว่าจะเกิดขึ้น (Dionne, 2000)

การแจกแจงล็อกนอร์มัล

ตัวแปรสุ่ม X มีการแจกแจงล็อกนอร์มัล (log-normal distribution) ถ้าตัวแปรสุ่ม $Y = \ln X$ มีการแจกแจงปรกติ โดยที่ฟังก์ชันความหนาแน่นความน่าจะเป็นของ X คือ

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), x > 0$$

เมื่อ μ เป็นค่าเฉลี่ยของ Y $(-\infty < \mu < \infty)$ และ σ เป็นค่าเบี่ยงเบนมาตรฐานของ Y $(\sigma > 0)$

การแจกแจงล็อกนอร์มักถูกนำไปใช้ในการจำลองข้อมูลในลักษณะที่มีการเบ้ขวาและมีความน่าจะเป็นที่ค่าจะออกมาเป็นค่าที่ห่างจากค่าเฉลี่ยสูงกว่าการแจกแจงปรกติ ซึ่งข้อมูลการเคลมประกันมักมีลักษณะหางหนา (heavy-tailed) เนื่องจากมีการเคลมที่มีมูลค่าสูงมากเกิดขึ้นเป็นครั้งคราว เช่น García et al. (2014) ได้วิเคราะห์และทดสอบการเข้ากันได้ของข้อมูลประกันภัยทันตกรรมกับการแจกแจงล็อกนอร์มัลนัยทั่วไป (generalized lognormal distribution) เพื่อแสดงให้เห็นว่าการแจกแจงล็อกนอร์มัลนัยทั่วไป สามารถสะท้อนพฤติกรรมของข้อมูลได้อย่างแม่นยำโดยเฉพาะในส่วนที่มีลักษณะหางหนา

การแจกแจงแกมมา

ตัวแปรสุ่ม X ที่มีการแจกแจงแกมมา (gamma distribution) มีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp(-x/\beta), x > 0$$

เมื่อ α เป็นพารามิเตอร์บ่งรูปร่าง $(\alpha > 0)$ β เป็นพารามิเตอร์บ่งขนาด $(\beta > 0)$ และ $\Gamma(\cdot)$ เป็นฟังก์ชันแกมมา การแจกแจงแกมมามีลักษณะกราฟเป็นโค้งที่ไม่สมมาตร (skewed) โดยมีค่าอยู่ทางด้านบวกเท่านั้น

การแจกแจงแกมมามักนิยมนำมาใช้กับข้อมูลจำนวนเงินเคลมประกันซึ่งเป็นข้อมูลที่มีค่าเป็นบวกและไม่สมมาตร ตัวอย่างงานวิจัยที่มีการประยุกต์ใช้การแจกแจงนี้ เช่น Adam et al. (2021) ได้ศึกษาตัวแบบเชิงเส้นนัยทั่วไป (generalized linear model) และตัวแบบเชิงเส้นนัยทั่วไปผสม (generalized linear mixed model) ซึ่งเป็นเครื่องมือที่สำคัญในการสร้างตัวแบบที่ใช้พยากรณ์จำนวนเงินเคลมประกัน โดยนำการแจกแจงแกมมาเป็นตัวแปรตอบสนองสำหรับตัวแบบดังกล่าว ตัวแบบนี้ช่วยในการทำนายพฤติกรรมของจำนวนเงินเคลมประกันโดยพิจารณาตัวแปรต่าง ๆ ที่มีผลต่อการเกิดเหตุ และผลกระทบต่อความน่าจะเป็นที่จะเกิดเงินเคลมที่มีมูลค่าต่าง ๆ โดยใช้ข้อมูลจริงจากบริษัทประกันของประเทศอินโดนีเซียในปี 2014 เพื่อทำความเข้าใจและวิเคราะห์พฤติกรรมการเคลมเงินประกันในสถานการณ์ที่แตกต่างกันในพื้นที่และช่วงเวลาต่าง ๆ ได้อย่างแม่นยำมากยิ่งขึ้น

การแจกแจงแบบเลขชี้กำลัง

การแจกแจงแบบเลขชี้กำลัง (exponential distribution) เป็นการแจกแจงที่เป็นกรณีเฉพาะของการแจกแจงแกมมา โดยตัวแปรสุ่ม X ที่มีการแจกแจงแบบเลขชี้กำลังมีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(x|\lambda) = \lambda \exp(-\lambda x), x > 0$$

เมื่อ λ เป็นพารามิเตอร์บ่งขนาด ($\lambda > 0$) ลักษณะกราฟของการแจกแจงแบบเลขชี้กำลังมีรูปร่างคล้ายเส้นโค้งลาดเอียงจากซ้ายไปขวา โดยจะมีค่าสูงสุดที่จุดเริ่มต้น ($x = 0$) และลดลงอย่างรวดเร็วเมื่อ x เพิ่มขึ้น กราฟดังกล่าวเป็นกราฟเบ้ขวา (right-skewed) บ่งบอกถึงความน่าจะเป็นที่จะเกิดเหตุการณ์น้อยลงเมื่อเวลาผ่านไป

Lee et al. (2012) ได้พัฒนาการแจกแจงแบบเลขชี้กำลังผสม (mixture exponential distribution) สำหรับการวิเคราะห์ค่าสุดขีด (extreme value analysis) การแจกแจงแบบเลขชี้กำลังผสมถูกเลือกใช้ในงานนี้เนื่องจากสามารถสะท้อนการเคลมประกันที่มักมีลักษณะเป็นกลุ่มย่อย (sub-populations) ที่แตกต่างกัน เช่น การเคลมประกันอาจแบ่งออกเป็นการเคลมเล็กน้อยที่เกิดขึ้นบ่อยครั้ง และการเคลมขนาดใหญ่ที่เกิดขึ้นน้อยกว่า แต่มีผลกระทบสูง การใช้การแจกแจงผสมช่วยให้เราสามารถจับลักษณะเหล่านี้ได้ดีขึ้นเมื่อเทียบกับการใช้การแจกแจงแบบธรรมดาที่อาจไม่สามารถสะท้อนความซับซ้อนของข้อมูลได้และยังช่วยให้การจัดการการเคลมและการบริหารความเสี่ยงในบริษัทประกันภัยเป็นไปอย่างมีประสิทธิภาพมากยิ่งขึ้น

การแจกแจงเกาส์ผกผัน

ตัวแปรสุ่ม X ที่มีการแจกแจงเกาส์ผกผัน (inverse Gaussian distribution) มีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(x|\mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), x > 0$$

เมื่อ μ เป็นค่าเฉลี่ย ($\mu > 0$) และ λ เป็นพารามิเตอร์บ่งรูปร่าง ($\lambda > 0$) การแจกแจงเกาส์เซียนผกผันมีลักษณะเป็นกราฟเบ้ขวา

การแจกแจงเกาส์ผกผันมีคุณสมบัติที่เหมาะสมกับข้อมูลที่มีค่าบวกและมีการกระจายตัวไม่สมมาตร โดยมีลักษณะเบ้ขวา จึงมักนิยมนำการแจกแจงเกาส์ผกผันมาประยุกต์ใช้กับข้อมูลที่เกี่ยวข้องกับการประกันภัย การจัดการความเสี่ยง และข้อมูลทางเศษฐกิจ เช่น การวิเคราะห์เกี่ยวกับข้อมูลการเรียกร้องค่าสินไหมทดแทนจากการบาดเจ็บทางร่างกาย และการประมาณการกระจายรายได้ของครัวเรือน (Punzo, 2019)

การแจกแจงไวบูล

ตัวแปรสุ่ม X ที่มีการแจกแจงไวบูล (weibull distribution) มีฟังก์ชันความหนาแน่นความน่าจะเป็น

$$f(x|\mu, \lambda, k) = \frac{k}{\lambda}\left(\frac{x-\mu}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x-\mu}{\lambda}\right)^{k}\right), x > \mu$$

เมื่อ $-\infty < \mu < \infty,\ k > 0$ และ $\lambda > 0$ ในที่นี้ k คือพารามิเตอร์บ่งรูปร่าง (shape parameter) และ λ คือพารามิเตอร์มาตรา (scale parameter) การแจกแจงไวบูลมีความสัมพันธ์กับการแจกแจงทางสถิติแบบอื่น ๆ ได้แก่ ถ้า $k = 1$ การแจกแจงไวบูลจะกลายเป็นการแจกแจงแบบเลขชี้กำลัง ถ้า $k = 2$ และ $\lambda = 2\sigma$ จะกลายเป็นการแจกแจงเรย์ลี (Rayleigh distribution) และเมื่อ k มีค่ามากขึ้นการแจกแจงไวบูลจะมีแนวโน้มกลายเป็นการแจกแจงเกาส์ (Rinne, 2008)

การแจกแจงไวบูลได้ถูกนำมาใช้หลากหลายในงานทางด้านความน่าเชื่อถือทางวิศวกรรม (reliability engineering) การวิเคราะห์ความล้มเหลว (failure analysis) การประเมินคุณภาพ รวมไปถึงงานทางด้าน

การแพทย์ เช่น การประเมินประสิทธิผลของการรักษา การวิเคราะห์ระยะเวลาและเงื่อนไขของการเกิดโรค และการประเมินช่วงเวลาที่ผู้ป่วยซึ่งเป็นโรคบางชนิดจะเสียชีวิต (McCool, 2012) นอกจากนี้มีการประยุกต์ใช้การแจกแจงไวบูลในงานทางด้านคณิตศาสตร์ประกันภัยโดยใช้ในปัญหาการจัดการความเสี่ยงทางการเงินและการประกันภัย การสร้างแบบจำลองพฤติกรรมของข้อมูลทางการเงินหรือข้อมูลเหตุการณ์ตลอดชีวิต เพื่อคาดการณ์การเคลื่อนไหวของราคาหุ้นหรือการพยากรณ์ความไม่แน่นอน และการสร้างแบบจำลองข้อมูลจำนวนครั้งการเรียกร้องค่าสินไหมทดแทนในการทำประกันภัย การประยุกต์ใช้การแจกแจงไวบูลช่วยให้การสร้างตัวแบบดังกล่าวมีประสิทธิภาพและความแม่นยำสูงขึ้น (Abubakar & Muhammad Sabri, 2023)

การแจกแจงห่อหุ้ม

ในทฤษฎีความน่าจะเป็น การแจกแจงห่อหุ้ม (wrapped distribution) เป็นการนำการแจกแจงทางสถิติแบบฉบับ (classical statistical distribution) หรือบางครั้งเราเรียกว่าการแจกแจงทางสถิติเชิงเส้น (linear statistical distribution) ซึ่งพิจารณาค่าของตัวแปรสุ่มบนจำนวนจริงใด ๆ $(-\infty, \infty)$ มาทำการห่อหุ้ม (wrapped) บนช่วง $[0,2\pi)$ หรือ $[-\pi, \pi)$ การแจกแจงนี้ถูกนำมาใช้ในการอธิบายปรากฏการณ์ที่ข้อมูลมีลักษณะเป็นคาบ (period) หรือมีการวนซ้ำกลับมาที่เดิม (circular) การแจกแจงห่อหุ้มสามารถนำมาประยุกต์ใช้ได้กับแจกแจงทางสถิติแบบฉบับทั้งแบบไม่ต่อเนื่อง (discrete) และต่อเนื่อง (continuous) เช่น การแจกแจงเกาส์ห่อหุ้ม (wrapped Gaussian distribution) ได้มาจากการประยุกต์กับการแจกแจงเกาส์ การแจกแจงแกมมาห่อหุ้ม (wrapped gamma distribution) และการแจกแจงแบบเลขชี้กำลัง (wrapped exponential distribution) ได้มาจากการประยุกต์กับการแจกแจงแกมมาและการแจกแจงแบบเลขชี้กำลังตามลำดับ การนิยามการแจกแจงแบบห่อหุ้มทำได้ดังนี้ สำหรับตัวแปรสุ่ม X ที่นิยามบนจำนวนจริงใด ๆ ซึ่งถือว่าเป็นตัวแปรสุ่มเชิงเส้น (linear random variable) เราสามารถแปลงตัวแปรดังกล่าวไปสู่ตัวแปรสุ่มแบบห่อหุ้มได้โดยนิยาม $\Theta = X(\bmod 2\pi)$ (เศษที่เหลือจากการหารด้วย 2π) ดังนั้นค่าความน่าจะเป็นของตัวแปรสุ่ม Θ จะกลายเป็นค่าความน่าจะเป็นที่สะสมบนจุดที่ซ้อนทับกัน $x = \theta, \theta \pm 2\pi, \theta \pm 4\pi, \dots$ ดังนั้นฟังก์ชันมวลความน่าจะเป็นหรือฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่มห่อหุ้ม คือ

$$f_{\Theta}(\theta) = \sum_{m=-\infty}^{\infty} f_X(\theta + 2\pi m)\, , 0 \leq \theta < 2\pi,$$

เมื่อ $f_X(x)$ แทนฟังก์ชันความหนาแน่นความน่าจะเป็น (หรือฟังก์ชันมวลความน่าจะเป็น) ของตัวแปรสุ่มที่นิยามบนจำนวนจริง และ $m = \cdots, -1, 0, 1, \dots$ (Damnet et al., 2023)

การแจกแจงห่อหุ้มได้ถูกนำมาใช้ในงานทางด้านคณิตศาสตร์ประกันภัยเพราะว่ารูปแบบของค่าใช้จ่ายทั่วไปในชีวิตมีลักษณะเป็นคาบ เช่น การจ่ายค่าประกันภัยไม่ว่าจะเป็นการจ่ายค่าประกันชีวิตหรือประกันภัยรถยนต์มักจ่ายเป็นรายเดือนหรือรายปี ค่าใช้ไฟฟ้ามีค่าลดต่ำในฤดูหนาวและสูงขึ้นมากในฤดูร้อน และพฤติกรรมที่มีโอกาสที่ต้องนำไปสู่การร้องค่าสินไหมทดแทนก็มีรูปแบบที่เป็นคาบหรือเกิดซ้ำ เช่น การเกิดอุบัติเหตุทางรถยนต์มักอยู่ในช่วงที่มีการจราจรหนาแน่น ได้แก่ช่วงเช้าและเย็นของวันทำงานราชการ การเรียกร้องค่าสินไหมทดแทนของอุบัติภัยทางน้ำเกิดในฤดูฝนเป็นส่วนใหญ่ การวิเคราะห์ด้านดอกเบี้ยและการลงทุนในตลาดหลักทรัพย์มีรูปแบบที่เป็นคาบ ดังนั้นในการวิเคราะห์ความเสี่ยงการประยุกต์ใช้การแจกแจงห่อหุ้มทำให้แบบจำลองมีความแม่นยำและประสิทธิภาพสูงขึ้น

****

ผลการวิจัยและอภิปรายผลการวิจัย

จากบทปริทัศน์ที่นำเสนอพบว่ามีงานวิจัยที่เกี่ยวข้องกับการประยุกต์ใช้การแจกแจงทางสถิติเพื่อใช้ในงานทางด้านประกันภัยมากมายโดยแต่ละการแจกแจงทางสถิติก็มีข้อจำกัดบางอย่างในการอธิบายหรือแสดงความสัมพันธ์ของพารามิเตอร์ของการแจกแจงทางสถิติกับเงื่อนไขของงานคณิตศาสตร์ประกันภัย เช่น บางการแจกแจงทางสถิติไม่สามารถอธิบายถึงข้อมูลที่มีการเปลี่ยนแปลงไปตามกาลเวลา หรือมีค่าของข้อมูลที่แตกต่างจากค่าข้อมูลอื่น ๆ อย่างมาก (significant outliers) เช่น ผลตอบแทนทางการเงิน การเรียกร้องประกันภัยและภัยธรรมชาติ รวมไปถึงการพิจารณาใช้เพียงการแจกแจงทางสถิติเพียงการแจกแจงอย่างใดอย่างหนึ่งอาจจะไม่ครอบคลุมเท่ากับการใช้การแจกแจงแบบผสม (mixture distribution) หรือการแจกแจงแบบประกอบ (compound distribution) อย่างไรก็ตามพบว่ายังมีงานวิจัยที่ต้องการพัฒนาและนำเสนอการแจกแจงทางสถิติใหม่ ๆ ที่เหมาะสมต่อการอธิบายงานทางด้านคณิตศาสตร์ประกันภัย แต่ด้วยความซับซ้อนทางด้านองค์ความรู้และเนื้อหาของการแจกแจงทางสถิติเหล่านั้นทำให้ไม่มีการนำมาใช้ในงานด้านคณิตศาสตร์ประกันภัยอย่างแพร่หลายนัก

ข้อเสนอแนะ

เนื่องด้วยยังมีงานวิจัยที่เกี่ยวข้องกับคณิตศาสตร์ประกันภัยอีกหลากหลายที่มีการใช้การแจกแจงทางสถิติอื่นที่ยังไม่ได้ทำบทปริทัศน์ เช่น งานวิจัยในปัจจุบันที่ใช้การแจกแจงทวิดี (Tweedie distribution) การแจกแจง SGT (skewed generalized t distribution) หรือ การแจกแจง log-EIG (log-exponential inverse Gaussian distribution) ในการสร้างแบบจำลองการเรียกร้องค่าสินไหมทดแทน ทั้งนี้การนำเสนอบทปริทัศน์การแจกแจงสถิติที่ร่วมสมัยจะช่วยให้เห็นถึงพัฒนาการทางด้านคณิตศาสตร์ประกันภัยที่ทันสมัยได้เป็นอย่างดี

เอกสารอ้างอิง

Abubakar, H. & Muhammad Sabri, S. R. (2023). A Bayesian approach to Weibull distribution with application to insurance claims data. Journal of Reliability and Statistical Studies, 16(1), 1-24.

Adam, F. A., Kurnia, A., Purnaba, I. G. P., Mangku, I. W. & Soleh, A. M. (2021). Modeling the amount of insurance claim using Gamma linear mixed model with AR(1) random effect. Journal of Physics: Conference Series, 1863, 012027.

Bahnemann, D. (2015). Distributions for actuaries (CAS Monograph Series No. 2). Casualty Actuarial Society.

Boucher, J. P., Denuit, M. & Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. Casualty Actuarial Society, 2(1), 0135-0162.

Damnet, M., Suriyawichitseranee, A. & Tanthanuch, J. (2023). Utilizing Bayesian Analysis of Wrapped Distributions in Computer Technology., Proceedings of the 20th International and National Conference on Applied Computer Technology and Information System (ACTIS). 25 August 2023, 97-101.

Dionne, G. (2000). Handbook of Insurance. Springer Science+Business Media, Llc.

Federici, N., & Pedroni, F. (1951). The Faculty of Statistical, Demographic and Actuarial Sciences at the University of Rome. The American Statistician, 5(4), 22-25.

García, V. J., Gómez-Déniz, E. & Vázquez-Polo, F. J. (2014). On modelling insurance data by using a generalized lognormal distribution, Revista de Métodos Cuantitativos para la Economía y la Empresa, Universidad Pablo de Olavide, Sevilla., (18),146-162.

Krishnamoorthy, K. (2006). Handbook of Statistical Distributions with Applications. Chapman & Hall/CRC.

Lee, D., Li, W. K., & Wong, T. S. T. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. Insurance: Mathematics and Economics, 51(3), 538-550.

Lewin, C. (2001). The creation of actuarial science. ZDM, 33(2), 61-66.

McCool, J. I. (2012). Using the Weibull Distribution: Reliability, Modeling, and Inference. John Wiley & Sons, Inc.

Punzo, A. (2019). A new look at the inverse Gaussian distribution with applications to insurance and economic data. Journal of Modern Applied Statistical Methods, 46(7), 1260-1287.

Rinne, H. (2008). The Weibull Distribution: A Handbook. CRC Press.