สถิติและความน่าจะเป็น
สำหรับการวิเคราะห์ข้อมูลเศรษฐศาสตร์และการเงิน
Statistics and Probability
For Economic and Financial Data Analysis

วีระชาติ กิเลนทอง

มหาวิทยาลัยหอการค้าไทย

สารบัญ

บทนำ

หนังสือเล่มนี้ต้องการนำเสนอแนวคิดทางสถิติที่สำคัญและจำเป็นต่อการวิเคราะห์ข้อมูลด้านเศรษฐศาสตร์และการเงิน

บทที่ 1

ความน่าจะเป็นเบื้องต้น (Basic Probability)

ถึงแม้ว่าอาจจะมียังการโต้เถียงกันในทางทฤษฎีอยู่ว่า ความน่าจะเป็นมีอยู่จริงหรือไม่ De Finetti (1974) แต่ในปัจจุบันความน่าจะเป็นได้เข้ามามีบทบาทในชีวิตประจำวันของทุกคนอย่างหลีกเลี่ยงไม่ได้

1.1 ความน่าจะเป็น

ความน่าจะเป็น (probability) เป็นพื้นฐานสำคัญของทุกสิ่งทุกอย่างที่เราจะกล่าวถึงในหนังสือเล่มนี้ โดยเรามักจะใช้ความน่าจะเป็นเพื่อบอกโอกาสที่แต่ละสถานการณ์หรือเหตุการณ์จะเกิดขึ้นจากการทดลองอันใดอันหนึ่ง ไม่ว่าจะเป็นกระบวนการทดลองจริง (real experiment) หรือการทดลองทางความคิด (thought experiment) ยกตัวอย่างเช่น ความน่าจะเป็นที่ผลลัพธ์ของการโยนเหรียญเป็นหัว (หรือก้อย) เท่ากับ $\frac{1}{2}$ หมายความว่าโอกาส (likelihood) ที่ผลลัพธ์ของการโยนเหรียญจะเป็นหัวหรือก้อยน่าจะพอๆ กัน หรือความน่าจะเป็นที่เด็กไทยจะเป็นโรคออทิสติก (autistic disorder) มีค่าประมาณ 0.006 เพราะเราพบในข้อมูลว่าร้อยละ 0.6 ของเด็กไทยป่วยเป็นโรคดังกล่าว เป็นต้น

ความน่าจะเป็นดังกล่าวเป็นผลมาจากการทดลองหรือการสังเกตที่เราทราบได้ถึงผลลัพธ์ (outcome) ที่เป็นไปได้ก่อนที่จะทำการทดลอง ยกตัวอย่างเช่น เรารู้ว่าผลลัพธ์ที่ได้จากการโยนเหรียญคือ หัวหรือก้อย ส่วนผลลัพธ์ที่เกี่ยวกับการป่วยเป็นโรคออทิสติก (autistic disorder) ก็คือ ป่วยหรือไม่ป่วย เป็นต้น นักคณิตศาสตร์มักเรียกกระบวนการที่เราสามารถระบุผลลัพธ์ได้เหล่านี้ว่า การทดลอง (experiment) ซึ่งมีนิยามดังต่อไปนี้

บทนิยามที่ 1.1 (การทดลอง)**.** *การทดลอง (experiment) คือกระบวนการ (process) ที่สามารถระบุได้ว่าผลลัพธ์ที่เป็นไปได้ (possible outcomes) ประกอบด้วยอะไรบ้าง เราเรียกการทดลองที่ไม่สามารถกำหนดผลลัพธ์ได้อย่างแน่นอนล่วงหน้าเพียงอันเดียวว่า การทดลองสุ่ม (random experiment)*

ผลลัพธ์บางกลุ่มอาจจะมีคุณสมบัติร่วมบางประการ เช่น ผลลัพธ์ของการโยนลูกเต๋าที่มีค่าน้อยกว่าสี่ ประกอบไปด้วยผลลัพธ์สามแบบคือ หนึ่ง สอง และสาม ในกรณีนี้เราอาจเรียกกลุ่มของผลลัพธ์ชุดนี้ว่า เหตุการณ์ที่ผลการโยนลูกเต๋ามีค่าน้อยกว่าสี่ ซึ่งในทางคณิตศาสตร์ เราเรียกกลุ่มของผลลัพธ์นี้ซึ่งอยู่ในรูปของเซ็ตว่า เหตุการณ์ (event)

บทนิยามที่ 1.2 (เหตุการณ์). *เหตุการณ์ (event) หมายถึงเซ็ตซึ่งระบุไว้ชัดเจน (well-defined set) ของผลลัพธ์ที่เป็นไปได้ (possible outcomes)*

ยิ่งไปกว่านั้น เราเรียกเซ็ตของผลลัพธ์ที่เป็นไปได้ (possible outcomes) ทั้งหมดว่า ปริภูมิตัวอย่าง (sample space) ซึ่งมีนิยามดังต่อไปนี้

บทนิยามที่ 1.3 (ปริภูมิตัวอย่าง). *ปริภูมิตัวอย่าง (sample space) คือเซ็ตของผลลัพธ์ที่เป็นไปได้ (possible outcomes) ทั้งหมดของการทดลองอันใดอันหนึ่ง*

ตัวอย่างที่ 1.1. พิจารณาการโยนลูกเต๋าที่มีทั้งหมด 6 หน้า ดังแสดงในรูป XXX ปริภูมิตัวอย่างสำหรับกรณีนี้คือเซ็ต

$$S = \{1, 2, 3, 4, 5, 6\}$$

ซึ่งมีสมาชิกทั้งหมด 6 อัน (element) ในกรณีที่ปริภูมิตัวอย่างเป็นแบบไม่ต่อเนื่อง ทุกๆ สับเซ็ตของปริภูมิตัวอย่างเป็นเซ็ตซึ่งระบุไว้ชัดเจน (well-defined set) ดังนั้น ทุกๆ สับเซ็ตจึงถือเป็นเหตุการณ์ ยกตัวอย่างเช่น $E = \{1, 2, 3\} \subset S$ เป็นเหตุการณ์ที่ผลการโยนลูกเต๋ามีค่าน้อยกว่าสี่ ▫

แน่นอนว่า ไม่ใช่ทุกสับเซ็ตของปริภูมิตัวอย่าง (sample space) จะมีคุณสมบัติที่จะเป็นเหตุการณ์ (event) ซึ่งจะเกิดขึ้นได้ในกรณีที่ปริภูมิตัวอย่างเป็นเซ็ตแบบต่อเนื่อง เครื่องมือที่ใช้ในหนังสือเล่มนี้จะไม่สามารถใช้ได้กับกรณีพิเศษเหล่านี้ ผู้อ่านที่สนใจสามารถศึกษาเพิ่มเติมได้จาก Billingsley (2008)

1.1.1 ทฤษฎีเซ็ตที่จำเป็นต่อทฤษฎีความน่าจะเป็นพื้นฐาน

ดังเช่นคณิตศาสตร์แขนงอื่นๆ เราจำเป็นต้องมีจุดเริ่มต้นของเรื่องเสมอ โดยในที่นี้ เราจำเป็นต้องเงื่อนไขเบื้องต้นทั้งหมด 4 เงื่อนไขเป็นพื้นฐานที่นำไปสู่ผลลัพธ์ต่างๆ อย่างไรก็ตาม เนื่องจากผลลัพธ์บางอย่างต้องการเพียงสองเงื่อนไข ดังนั้น เราจึงขอนำเสนอเงื่อนไขตามลำดับความจำเป็นของแต่ละเงื่อนไข สำหรับผู้อ่านที่มีความรู้พื้นฐานของทฤษฎีเซ็ตเป็นอย่างดีน่าจะสามารถทำความเข้าใจหัวข้อนี้ได้อย่างรวดเร็วหรืออาจจะข้ามหัวข้อนี้ได้โดยไม่น่าจะมีผลเสียใดๆ ส่วนผู้ที่ยังไม่เข้าใจดีนักควรพยายามที่จะศึกษาและฝึกฝนขั้นตอนการคำนวณและการพิสูจน์ในส่วนนี้ให้ดี เพราะเป็นพื้นที่สำคัญในการทำความเข้าใจทฤษฎีความน่าจะเป็นอย่างมาก

เพื่อความสะดวก ขอเริ่มจากทฤษฎีบทที่มาจากทฤษฎีเซ็ตโดยตรง ซึ่งไม่จำเป็นต้องเชื่อมโยงกับความน่าจะเป็นมากนัก แต่ก็มีประโยชน์ต่อการทำความเข้าใจทฤษฎีความน่าจะเป็นไม่น้อย โดยเริ่มจากเงื่อนไขแรกซึ่งกำหนดให้ปริภูมิตัวอย่างเป็นเหตุการณ์ๆ หนึ่ง

เงื่อนไขที่ 1.1. *ปริภูมิตัวอย่าง (sample space) ต้องเป็นหนึ่งในเหตุการณ์*

ทฤษฎีบทที่ 1.1. *กำหนดให้ A, B, และ C เป็นเหตุการณ์ ถ้า $A \subset B$ และ $B \subset A$ แล้ว $A = B$ นอกจากนี้ ถ้า $A \subset B$ และ $B \subset C$ แล้ว $A \subset C$*

ยกตัวอย่างเช่น A เป็นเหตุการณ์การที่ผลการโยนลูกเต๋ามีค่าเป็นเลขคู่ ส่วน B เป็นเหตุการณ์การที่ผลการโยนลูกเต๋ามีค่ามากกว่าหรือเท่ากับสอง เนื่องจากเลขคู่มีค่ามากกว่าหนึ่ง ดังนั้น $A \subset B$ ส่วน C เป็นเหตุการณ์การที่ผลการโยนลูกเต๋ามีค่ามากกว่าหรือเท่ากับหนึ่ง ดังนั้น $B \subset C$ จากทฤษฎีบทที่ 1.1 เราจึงสรุปได้ว่า $A \subset C$

ในทางคณิตศาสตร์ เรามีเซ็ตอันหนึ่งที่มีความพิเศษ และไม่มีสมาชิกเลย กล่าวคือมีจำนวนสมาชิกเท่ากับศูนย์ เราเรียกเซ็ตนี้ว่า เซ็ตว่าง (empty set) ซึ่งมีความคล้ายคลึงกับเลขศูนย์ในระบบตัวเลขปกติที่เราคุ้นเคย

บทนิยามที่ 1.4 (เซ็ตว่าง)**.** *เซ็ตว่าง (empty set) $\varnothing$ คือสับเซ็ตของปริภูมิตัวอย่างที่ไม่มีสมาชิกอยู่เลย*

ทฤษฎีบทที่ 1.2. *กำหนดให้ A เป็นเหตุการณ์ $\varnothing \subset A$*

บทนิยามที่ 1.5 (ส่วนเติมเต็มของเซ็ต)**.** *ส่วนเติมเต็ม (complement) ของเซ็ต A หมายถึงเซ็ตซึ่งประกอบไปด้วยสมาชิกทุกตัวของปริภูมิตัวอย่าง S ที่ไม่เป็นสมาชิกของ A โดยเรามักแทนส่วนเติมเต็มของเซ็ต A ด้วย A^c*

ยกตัวอย่างเช่น A เป็นเหตุการณ์การที่ผลการโยนลูกเต๋ามีค่าเป็นเลขคู่ ดังนั้น $A = \{2, 4, 6\}$ ในขณะที่ $A^c = \{1, 3, 5\}$ คือส่วนเติมเต็มของเซ็ต A

เงื่อนไขที่สำคัญอันที่สองเกี่ยวข้องกับส่วนเติมเต็มของเซ็ต ดังต่อไปนี้

เงื่อนไขที่ 1.2. *ถ้า A เป็นเหตุการณ์ แล้วส่วนเติมเต็มของมัน A^c ต้องเป็นเหตุการณ์ด้วย*

ในทางปฏิบัติ นักคณิตศาสตร์มักแทนเซ็ตและการดำเนินการทางคณิตศาสตร์ที่เกี่ยวข้องโดยใช้รูปภาพที่เรียกว่า แผนภาพเวนน์ (Venn diagram) ดังแสดงตัวอย่างในรูปที่ XXX ซึ่งแสดงถึงส่วนเติมเต็มของเซ็ต A โดยพื้นที่ทั้งหมดในกรอบสี่เหลี่ยมนั้นแทนปริภูมิตัวอย่าง S

ADD FIGURE OF VENN DIAGRAM WITH COMPLEMENT OF A

ทฤษฎีบทที่ 1.3. *กำหนดให้ A เป็นเหตุการณ์ ดังนั้น $(A^c)^c = A$, $\varnothing^c = S$, $S^c = \varnothing$*

สังเกตว่าเนื่องจาก ปริภูมิตัวอย่าง เป็นเหตุการณ์ (ตามเงื่อนไขที่) และ $S^c = \varnothing$ ดังนั้น เหตุการณ์ว่าง (empty event) $\varnothing$ ต้องเป็นเหตุการณ์ด้วย

บทนิยามที่ 1.6 (ยูเนียน (union))**.** *ยูเนียน (union) ของเซ็ต A_1 และ A_2 ซึ่งแทนด้วย $A_1 \cup A_2$ ประกอบไปด้วยสมาชิกทุกตัวที่อยู่ใน A_1 อย่างเดียว ที่อยู่ใน A_2 อย่างเดียว และที่อยู่ในทั้งสองเซ็ต*

เราสามารถแสดงยูเนียนของเซ็ต A_1 และ A_2 ในรูปแบบของแผนภาพเวนน์ (Venn diagram) ได้ดังแสดงในรูปที่ XXX

ทฤษฎีบทที่ 1.4. *กำหนดให้ A และ B เป็นเซ็ตใดๆ ดังนั้น*

$$A \cup B = B \cup A, A \cup A = A, AUA^c = S, A \cup \varnothing = A, A \cup S = S$$

ยิ่งไปกว่านั้น ถ้า $A \subset B$ แล้ว $A \cup B = B$

อันที่จริงแล้ว หลักการยูเนียนสามารถใช้ได้กับกรณีที่มีเซ็ตมากกว่าสองเซ็ต ยกตัวอย่างเช่น กรณีที่มีเซ็ตทั้งหมด n เซ็ต ประกอบไปด้วย $A_1, A_2, \ldots, A_n$ เราสามารถเขียนยูเนียนของ n เหล่านี้ได้เป็น

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

ส่วนที่ยุ่งยากมากกว่าคือ การยูเนียนของอนุกรมอนันต์ของเซ็ต ยกตัวอย่างเช่น อนุกรมของเซ็ต $A_1, A_2, \ldots$ หรืออนุกรม $(A_\alpha)_{\alpha \in \mathbb{R}_{+]}}$ เป็นต้นส่วนที่แตกต่างจากกรณีก่อนหน้านี้คือ จำนวนของเซ็ตทั้งหมดในที่นี้คืออนันต์ เพื่อความสะดวก เรามักแยกความเป็นอนันต์ออกเป็นสองประเภทคือ แบบที่นับได้ (countable) และแบบที่นับไม่ได้ (uncountable) ซึ่งเป็นประเด็นทางเทคนิคที่ค่อนข้างเป็นนามธรรม แต่ก็สำคัญกับทฤษฎีความน่าจะเป็นอย่างมาก ดังนั้น จึงจำเป็นจะต้องเข้าใจนิยามและความหมายพื้นฐานของเรื่องนี้ให้ดี ในมุมของทฤษฎีเซ็ต เรามักสามารถแยกเซ็ตออกเป็นสองกลุ่มตามจำนวนสมาชิกของเซ็ตได้ดังนี้ คือ เซ็ตอนันต์ (infinite set) ซึ่งมีจำนวนสมาชิกไม่จำกัดหรืออนันต์ ส่วนที่เหลือเรียกว่า เซ็ตจำกัด (finite set) นอกจากนี้ เซ็ตอนันต์ยังออกแบ่งได้เป็นสองประเภท คือ เซ็ตอนันต์ A เป็นเซ็ตแบบนับได้ (countable set) ถ้ามีความสมนัยแบบหนึ่งต่อหนึ่ง (one-to-one correspondence) ระหว่างสมาชิกของเซ็ตนั้นกับจำนวนธรรมชาติ (natural numbers) $\{1, 2, 3, \ldots\}$ ส่วนเซ็ตที่ไม่ใช่เซ็ตจำกัดและไม่ใช่แบบนับได้จะถูกเรียกว่า เซ็ตแบบนับไม่ได้ (uncountable set) ตัวอย่างของเซ็ตแบบนับไม่ได้คือ เซ็ตจำนวนจริง (set of real numbers) ผู้อ่านสามารถศึกษาการพิสูจน์ว่าเซ็ตจำนวนจริงเป็นเซ็ตแบบนับไม่ได้จากส่วนท้ายของหัวข้อที่ 1.4 ในหนังสือ DeGroot and Schervish (2012) ในขณะเดียวกันหากเราประยุกต์ใช้หลักการนับได้ (countability) ดังที่กล่าวมาแล้ว เราก็จะสามารถเข้าใจได้ว่า อนุกรมของเซ็ต $A_1, A_2, \ldots$ เป็นอนุกรมอนันต์แบบนับได้ ความถ้าทายก็คือ เราจะรู้ได้อย่างไรว่า ยูเนียนของอนุกรมอนันต์ของเหตุการณ์เป็นเหตุการณ์หรือไม่ คำถามก็คือเราไม่สามารถรู้ได้ ดังนั้น เราจึงต้องกำหนดไปเงื่อนไขที่ต้องเป็นจริงเสมอ

เงื่อนไขที่ 1.3. *ถ้า $A_1, A_2, \ldots$ เป็นอนุกรมอนันต์ที่นับได้ (countable) ของเหตุการณ์ แล้ว $\bigcup_{i=1}^{\infty} A_i$ จะต้องเป็นเหตุการณ์*

ประเด็นที่สำคัญทางเทคนิคของเงื่อนไขคือ เรากำหนดให้ $\bigcup_{i\in I} A_i$ ต้องเป็นเหตุการณ์ ก็ต่อเมื่อ I เป็นเซ็ตที่นับได้ (countable set) กล่าวคือ เราไม่ได้บังคับให้ $\bigcup_{i\in I} A_i$ ต้องเป็นเหตุการณ์สำหรับทุกๆ เซ็ต I ยกเว้นว่า I เป็นเซ็ตที่นับได้

ผลที่สำคัญของเงื่อนไขที่ 1.3 คือการที่เราสามารถบอกได้ว่าการยูเนียนของจำนวนจำกัดของเหตุการณ์เป็นเหตุการณ์

ทฤษฎีบทที่ 1.5. *ผลการยูเนียนของจำนวนจำกัดของเหตุการณ์ $A_1, \ldots, A_n$ เป็นเหตุการณ์*

การพิสูจน์. กำหนดให้ $A_m = \varnothing$ สำหรับ $m = n+1, n+2, \ldots$ เนื่องจาก $\varnothing$ เองก็เป็นเหตุการณ์ ดังนั้น เราสามารถสร้างอนุกรมอนันต์ที่นับได้ (countable) ของเหตุการณ์ $A_1, A_2, \ldots$ ได้จาก $A_1, \ldots, A_n$ และ $A_{n+1}, A_{n+2}, \ldots$ ในขณะเดียวกัน เราสามารถใช้คุณสมบัติของยูเนียนของเซ็ตแสดงได้ว่า

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i + \bigcup_{m=n+1}^{\infty} A_m = \bigcup_{i=1}^{n} A_i$$

โดยที่ขั้นตอนสุดท้ายประยุกต์ใช้ผลลัพธ์ที่ว่ายูเนียนของเซ็ตว่างต้องเป็นเซ็ตว่าง หลังจากนั้น เราสามารถประยุกต์ใช้เงื่อนไขที่ 1.3 ที่กำหนดให้ $\bigcup_{i=1}^{\infty} A_i$ จะต้องเป็นเหตุการณ์ เพื่อสรุปได้ว่า ผลการยูเนียนของจำนวนจำกัดของเหตุการณ์ $\bigcup_{i=1}^{n} A_i$ เป็นเหตุการณ์ ■

ทฤษฎีบทที่ 1.6 (คุณสมบัติการเชื่อมโยงของยูเนียน (associative property of union))**.** *ความสัมพันธ์ระหว่าง 3 เหตุการณ์ A, B และ C ใดๆ จะต้องสอดคล้องกับคุณสมบัติการเชื่อมโยงของยูเนียน (associative property of union) ต่อไปนี้*

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$$

การดำเนินการ (operation) ที่สำคัญในทฤษฎีเซ็ตคือ การหาส่วนร่วม (intersection) ของเซ็ตตั้งแต่สองเซ็ตเป็นต้นไป

บทนิยามที่ 1.7 (ส่วนร่วม (intersection))**.** *ส่วนร่วม (intersection) ของเซ็ต A_1 และ A_2 ซึ่งแทนด้วย $A_1 \cap A_2$ ประกอบไปด้วยสมาชิกทุกตัวที่อยู่ในทั้ง A_1 และ A_2*

เราสามารถแสดงส่วนร่วมของเซ็ต A_1 และ A_2 ในรูปแบบของแผนภาพเวนน์ (Venn diagram) ได้ดังแสดงในรูปที่ XXX สังเกตได้ว่าความแตกต่างระหว่างยูเนียนและส่วนรวมคือ ส่วนร่วมนั้นเป็นข้อจำกัดที่เข้มงวดกว่าเพราะจะเลือกเอาเฉพาะผลลัพธ์ที่อยู่ในทั้งสองเหตุการณ์เท่านั้น ซึ่งกำหนดโดยการใช้คำว่า "และ'" ในขณะที่การยูเนียนใช้คำว่า "หรือ"

ทฤษฎีบทที่ 1.7. *ถ้า A และ B เป็นเหตุการณ์ แล้ว $A \cap B$ เป็นเหตุการณ์ด้วย นอกจากนี้ สำหรับเหตุการณ์ A และ B ใดๆ เราสามารถแสดงได้ว่า*

$$A \cap B = B \cap A, A \cap A = A, A \cap A^c = \varnothing, A \cap \varnothing = \varnothing, A \cap S = A.$$

ยิ่งไปกว่านั้น ถ้า $A \subset B$ แล้ว $A \cap B = A$

เราสามารถขยายนิยามของส่วนร่วมของเซ็ตให้ครอบคลุมกลุ่มของเซ็ตที่มากกว่าสองอันได้ในทำนองเดียวกับกรณีของยูเนียน ยกตัวอย่างเช่น ส่วนร่วมของ n เหตุการณ์ $A_1, \ldots, A_n$ ประกอบไปด้วยสมาชิกทุกตัวที่อยู่ในทุกเหตุการณ์ ซึ่งแทนในรูปสัญลักษณ์ได้เป็น $\bigcap_{i=1}^{n} A_i$ ในทำนองเดียวกัน เราสามารถนิยามส่วนร่วมของเซ็ตจากกลุ่มของเซ็ต I ใดๆ เป็น $\bigcap_{i \in I} A_i$ เป็นที่น่าสังเกตว่า เราไม่จำเป็นสร้างข้อกำหนดเพิ่มเติมในกรณีของส่วนร่วมของเซ็ต ซึ่งแตกต่างจากการยูเนียนที่จำเป็นต้องกำหนดเงื่อนไขที่ 1.3 เพิ่มเติม

ทฤษฎีบทที่ 1.8 (คุณสมบัติการเชื่อมโยงของส่วนร่วม (associative property of intersection))**.** *ความสัมพันธ์ระหว่าง 3 เหตุการณ์ A, B และ C ใดๆ จะต้องสอดคล้องกับคุณสมบัติการเชื่อมโยงของส่วนร่วม (associative property of intersection) ต่อไปนี้*

$$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$$

นิยามที่เกี่ยวข้องกับส่วนร่วมของเซ็ตที่มีความสำคัญต่อการดำเนินการ (operation) ในทฤษฎีความน่าจะเป็น คือ การไม่มีส่วนร่วมต่อกัน (disjoint) หรือการไม่เกิดร่วมกัน (mutually exclusive) ซึ่งเป็นคุณสมบัติพิเศษของกลุ่มของเซ็ตที่นำไปสู่คุณสมบัติต่างๆ ที่มีส่วนสำคัญในทฤษฎีความน่าจะเป็นและสถิติ ดังนั้น ผู้อ่านควรให้ความสำคัญกับนิยามนี้เป็นพิเศษ รวมทั้งผู้ที่อาจจะมีความรู้ด้านทฤษฎีเซ็ตที่ดีพอสมควรแล้วก็ตาม นอกจากนี้ สังเกตว่านิยามการไม่มีส่วนร่วมต่อกัน (disjoint) หรือการไม่เกิดร่วมกัน (mutually exclusive) เป็นนิยามที่เกี่ยวกับเซ็ตที่ไม่จำเป็นต้องใช้หลักการของความน่าจะเป็น (probability) แต่อย่างใด ซึ่งจะแตกต่างจากนิยามความเป็นอิสระ (independence) ซึ่งต้องอาศัยหลักการของความน่าจะเป็น ดังที่แสดงในหัวข้อ XXX

บทนิยามที่ 1.8. *เซ็ต A และ B ไม่มีส่วนร่วมต่อกัน (disjoint) หรือไม่เกิดร่วมกัน (mutually exclusive) ถ้า $A \cap B = \varnothing$*

ตัวอย่างที่ 1.2. พิจารณาการโยนลูกเต๋า 6 หน้า ปริภูมิตัวอย่างในกรณีนี้คือ $\{1, 2, 3, 4, 5, 6\}$ เหตุการณ์ที่ผลลัพธ์เป็นเลขคี่คือ $A_1 = \{1, 3, 5\}$ ส่วนเหตุการณ์ที่ผลลัพธ์เป็นเลขคู่คือ $A_2 = \{2, 4, 6\}$ ชัดเจนว่า $A_1 \cap A_2 = \varnothing$ ดังนั้น เราสามารถสรุปได้ว่า เหตุการณ์ A_1 และ A_2 ไม่มีส่วนร่วมต่อกัน (disjoint) ในทางตรงกันข้าม เหตุการณ์ที่ผลลัพธ์น้อยกว่า 3 คือ $A_3 = \{1, 2\}$ ดังนั้น เหตุการณ์ A_1 และ A_3 มีส่วนร่วมต่อกัน เพราะ $A_1 \cap A_3 = \{1\} \neq \varnothing$ □

ส่วนกรณีที่มีมากกว่าสองเซ็ตสามารถนิยามได้ดังต่อไปนี้

บทนิยามที่ 1.9 (การไม่มีส่วนร่วมต่อกัน (disjoint) หรือการไม่เกิดร่วมกัน (mutually exclusive)). *อนุกรมเหตุการณ์ $A_1, A_2, \ldots$ ไม่มีส่วนร่วมต่อกัน (disjoint) หรือไม่เกิดร่วมกัน (mutually exclusive) ถ้า $A_i \cap A_j = \varnothing$ สำหรับทุกๆ $i \neq j$ กล่าวคือ ทุกๆ คู่เซ็ตไม่มีส่วนร่วมต่อกัน ในทำนองเดียวกัน กลุ่มของเซ็ต I ใดๆ ไม่มีส่วนร่วมต่อกันหรือไม่เกิดร่วมกันก็ต่อเมื่อๆ ทุกๆ คู่เซ็ตไม่มีส่วนร่วมต่อกัน ไม่มีส่วนร่วมต่อกัน (disjoint)*

ตัวอย่างที่ 1.3. พิจารณาการโยนลูกเต๋า 6 หน้า จำนวนสองครั้ง กำหนดให้ A_1 คือเหตุการณ์ที่ผลบวกของผลลัพธ์ในการโยนทั้งสองครั้งเป็นเลขคี่ A_2 คือเหตุการณ์ที่ผลลัพธ์ในการโยนทั้งสองครั้งเป็นเลขคี่ทั้งสองครั้ง ส่วน A_3 คือเหตุการณ์ที่ผลลัพธ์ในการโยนทั้งสองครั้งเป็นเลขคู่ทั้งสองครั้ง ชัดเจนว่า $A_2 \cap A_3 = \varnothing$ ในขณะเดียวกัน ผลบวกของการโยนทั้งสองครั้งจะเป็นเลขคี่ได้ก็ต่อเมื่อผลการโยนครั้งหนึ่งจะต้องเป็นเลขคี่และอีกครั้งหนึ่งจะต้องเป็นเลขคู่ เราจึงสรุปได้ว่า $A_1 \cap A_3 = \varnothing$ และ $A_1 \cap A_2 = \varnothing$ โดยสรุป A_1, A_2 และ A_3 ในทางตรงกันข้าม หากกำหนดให้ A_4 คือเหตุการณ์ที่ผลบวกของผลลัพธ์ในการโยนทั้งสองครั้งเป็นเลขคู่ จะพบว่า $A_2 \cap A_4 \neq \varnothing$ และ $A_3 \cap A_4 \neq \varnothing$ ถึงแม้ว่า $A_1 \cap A_4 = \varnothing$ ดังนั้น ข้อสรุปในกรณีนี้คือ A_1, A_2, A_3 และ A_4 มีส่วนร่วมต่อกัน ▫

คุณสมบัติที่มีประโยชน์และจำเป็นที่เกี่ยวข้องกับส่วนร่วมของเซ็ตและยูเนียนประกอบไปด้วย

ทฤษฎีบทที่ 1.9 (De Morgan's Laws). *ถ้า A และ B เป็นเหตุการณ์ แล้ว*

$$(A \cup B)^c = A^c \cap B^c, \text{ และ } (A \cap B)^c = A^c \cup B^c$$

ดูรูปที่ XXX ในรูปแบบของแผนภาพเวนน์ (Venn diagram) ประกอบ

ทฤษฎีบทที่ 1.10 (คุณสมบัติการกระจาย (distributive properties)). *สำหรับเหตุการณ์ A, B และ C ใดๆ*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \text{ และ } A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

ดูรูปที่ XXX ในรูปแบบของแผนภาพเวนน์ (Venn diagram) ประกอบ

ทฤษฎีบทต่อไปนี้เกี่ยวข้องกับการแบ่งส่วนเซ็ตหรือเหตุการณ์ (partitioning) ซึ่งมีประโยชน์ในการคำนวณความน่าจะเป็นของเหตุการณ์ที่สามารถแบ่งส่วนเป็นเหตุการณ์ย่อยๆ ได้

ทฤษฎีบทที่ 1.11. *สำหรับเหตุการณ์ Aและ B ใดๆ คุณสมบัติต่อไปนี้เป็นจริงเสมอ*

1. *$A \cap B$ และ $A \cap B^c$ ไม่มีส่วนร่วมต่อกัน (disjoint)*

2. *B และ $A \cap B^c$ ไม่มีส่วนร่วมต่อกัน (disjoint)*

3. $A = (A \cap B) \cup (A \cap B^c)$

4. $A \cup B = B \cup (A \cap B^c)$

คุณสมบัติการแยกส่วนเซ็ตในทฤษฎีบทที่ 1.11 เป็นส่วนเริ่มต้นสำคัญของหลักการทางคณิตศาสตร์สำหรับสถิติขั้นสูงที่เรามักจะเรียกว่า การแบ่งส่วน (partition) ซึ่งบอกถึงการแบ่งส่วนของปริภูมิตัวอย่างออกเป็นส่วนๆ ที่ไม่มีส่วนร่วมกัน (คล้ายกับการแบ่งส่วนหรือแบ่ง partition ในอาคารบ้านเรือนของเรานั่นเอง)

บทนิยามที่ 1.10 (การแบ่งส่วน (partition)). *กลุ่มของเหตุการณ์ $\{A_i : i \in I\}$ เป็นการแบ่งส่วน (partition) หรือการแยก (decomposition) ของปริภูมิตัวอย่าง (sample space) ถ้าทุกๆ คู่เซ็ตไม่มีส่วนร่วมกัน นั่นคือ $A_i \cap A_j = \varnothing$ สำหรับทุกๆ $i \neq j \in I$ และ $\bigcup_{i \in I} = S$*

สังเกตว่า คุณสมบัติเบื้องต้นของการแบ่งส่วน (partition) คือจะต้องไม่เกิดร่วมกัน (mutually exclusive)

1.1.2 นิยามของความน่าจะเป็น (The Definition of Probability)

หัวข้อนี้นำเสนอนิยามของความน่าจะเป็น (probability) ของเหตุการณ์ $A \subset S$ ซึ่งแทนด้วย $Pr(A)$ โดยอาศัยสัจพจน์ (axiom) 3 อันซึ่งประกอบไปด้วย

สัจพจน์ที่ 1.1 (คุณสมบัติการเป็นบวกของความน่าจะเป็น). *ความน่าจะเป็นของเหตุการณ์ A ใดๆ จะต้องไม่น้อยกว่าศูนย์ นั่นคือ*

$$Pr(A) \geqslant 0 \tag{1.1}$$

สัจพจน์ที่ 1.2 (คุณสมบัติความแน่นอนของปริภูมิตัวอย่าง). *ความน่าจะเป็นของปริภูมิตัวอย่าง S ซึ่งเกิดขึ้นอย่างแน่นอนจะต้องมีค่าเท่ากับหนึ่ง นั่นคือ*

$$Pr(S) = 1 \tag{1.2}$$

ส่วนสัจพจน์ที่สามนั้นเกี่ยวข้องกับชุดของเหตุการณ์ (collection of events) ที่ไม่มีส่วนร่วมต่อกัน (disjoint)

สัจพจน์ที่ 1.3 (คุณสมบัติการบวกกันของความน่าจะเป็น (additive property of probability)). *ความน่าจะเป็นของยูเนียนของอนุกรมอนันต์ของเหตุการณ์ที่ไม่มีส่วนร่วมต่อกัน (disjoint events) มีค่าเท่ากับผลบวกของความน่าจะเป็นของแต่ละเหตุการณ์ นั่นคือ*

$$Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i) \tag{1.3}$$

บทนิยามที่ 1.11 (ความน่าจะเป็น (probability))**.** *ความน่าจะเป็น (probability) หรือมาตรวัดความน่าจะเป็น (probability measure) หมายถึงฟังก์ชัน (function)* $Pr\left(A\right)$ *ที่ทำหน้าที่แปลงเหตุการณ์* $A \subset S$ *ใดๆ ให้อยู่ในรูปของจำนวนจริงที่สอดคล้องกับสัจพจน์ 1.1-1.3*

ประเด็นที่สำคัญก็คือ เราสามารถนิยามหรือสร้างความน่าจะเป็น (probability) ที่สอดคล้องกับสัจพจน์ที่ 1.1-1.3 สำหรับการทดลองอันใดอันหนึ่งได้มากกว่าหนึ่งแบบ กล่าวคือ ฟังก์ชันความน่าจะเป็นไม่มีความเป็นหนึ่งเดียว (not unique) ยกตัวอย่างเช่น การทดลองโยนลูกเต๋า 6 หน้า ซึ่งมีปริภูมิตัวอย่างเป็น $S = \{1, 2, 3, 4, 5, 6\}$ ความน่าจะเป็น (probability) ที่คนส่วนใหญ่อาจจะคุ้นเคยก็คือ ฟังก์ชันที่กำหนดให้ความน่าจะเป็นของแต่ละผลลัพธ์ที่เป็นไปได้มีค่าเท่ากัน นั่นคือ ความน่าจะเป็นของผลการโยนที่ได้ค่า k มีค่าเท่ากับ $p_k = \frac{1}{6}$ สำหรับ $k = 1, 2, 3, 4, 5, 6$ แน่นอนว่าฟังก์ชันอันนี้สอดคล้องกับสัจพจน์ที่ 1.1-1.3 ทุกข้อ ดังนั้น มันจึงเป็นความน่าจะเป็น (probability) หรือมาตรวัดความน่าจะเป็น (probability measure) ในขณะเดียวกัน บางคนอาจจะมองว่าลูกเต๋าที่ใช้ไม่สมดุล จึงกำหนดให้ความน่าจะเป็นของผลการโยนที่ได้ค่า k มีค่าเท่ากับ $\tilde{p}_k = \frac{2}{9}$ สำหรับ $k = 1, 3, 5$ และ ความน่าจะเป็นของผลการโยนที่ได้ค่า k มีค่าเท่ากับ $\tilde{p}_k = \frac{1}{9}$ สำหรับ $k = 2, 4, 6$ ฟังก์ชันอันนี้ก็สอดคล้องกับสัจพจน์ที่ 1.1-1.3 ทุกข้อเช่นเดียวกัน ดังนั้น มันจึงเป็นความน่าจะเป็น (probability) หรือมาตรวัดความน่าจะเป็น (probability measure) ของการทดลองอันเดียวกันนี้ โดยสรุป ฟังก์ชัน p_k และ $\tilde{p}_k$ เป็น ความน่าจะเป็น (probability) ที่ถูกต้องตามหลักการสำหรับการโยนลูกเต๋าทั้งคู่ อันที่จริงแล้วยังมีฟังก์ชันอีกนับไม่ถ้วนที่เป็นความน่าจะเป็น (probability) ที่ถูกต้องตามหลักการ ในทางปฏิบัติ เราจึงจำเป็นต้องเลือกว่าจะใช้ความน่าจะเป็น (probability) หรือมาตรวัดความน่าจะเป็น (probability measure) ใดในการวิเคราะห์แต่ละปัญหา ซึ่งก็ขึ้นอยู่กับสถานการณ์และวิจารณญานของผู้วิเคราะห์เอง

ทฤษฎีบทที่ 1.12. $Pr\left(\varnothing\right) = 0$

การพิสูจน์. พิจารณาอนุกรมเหตุการณ์ $A_1, A_2, \ldots$ โดยที่ $A_i = \varnothing$ สำหรับทุก $i = 1, 2, \ldots$ ซึ่งไม่มีส่วนร่วมต่อกัน (disjoint) เพราะ $A_i \cap A_j = \varnothing$ สำหรับทุก $i \neq j$ และ $\bigcup_{i=1}^{\infty} A_i = \varnothing$ ดังนั้น เราจึงสามารถประยุกต์ใช้สัจพจน์ที่ 1.3 เพื่อแสดงว่า

$$Pr\left(\varnothing\right) = Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr\left(A_i\right)$$

มีทางเดียวที่สมการนี้จะเป็นจริงได้คือ $Pr\left(\varnothing\right) = 0$ ■

ทฤษฎีบทที่ 1.13. *สำหรับอนุกรมจำกัดของเหตุการณ์* $A_1, A_2, \ldots, A_n$ *ที่ไม่มีส่วนร่วมต่อกัน (disjoint) ใดๆ*

$$Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} Pr\left(A_i\right) \tag{1.4}$$

การพิสูจน์. พิจารณาอนุกรมเหตุการณ์ $A_1, A_2, \ldots$ โดยที่ $A_1, A_2, \ldots, A_n$ ไม่มีส่วนร่วมต่อกัน (disjoint) และ $A_i = \varnothing$ เมื่อ $i > n$ ดังนั้น อนุกรมเหตุการณ์ $A_1, A_2, \ldots$ ไม่มีส่วนร่วมต่อกัน (disjoint) และ $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$ ดังนั้น เราจึงสามารถประยุกต์ใช้สัจพจน์ที่ 1.3 เพื่อแสดงว่า

$$\begin{aligned} Pr\left(\bigcup_{i=1}^{n} A_i\right) = Pr\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} Pr\left(A_i\right) \\ &= \sum_{i=1}^{n} Pr\left(A_i\right) + \sum_{i=n+1}^{\infty} Pr\left(A_i\right) = \sum_{i=1}^{n} Pr\left(A_i\right) + 0 = \sum_{i=1}^{n} Pr\left(A_i\right) \end{aligned}$$

โดยที่สมการที่สี่เป็นผลมาจากทฤษฎีบทที่ 1.12 ■

ทฤษฎีบทที่ 1.14. *สำหรับเหตุการณ์ A ใดๆ*

$$Pr\left(A^c\right) = 1 - Pr\left(A\right) \tag{1.5}$$

การพิสูจน์. เนื่องจาก A และ A^c ไม่มีส่วนร่วมต่อกัน (disjoint) และ $A \cup A^c = S$ ดังนั้น

$$Pr\left(A\right) + Pr\left(A^c\right) = Pr\left(S\right) = 1$$

ดังนั้น เราสามารถสรุปได้ว่า $Pr\left(A^c\right) = 1 - Pr\left(A\right)$ ■

ทฤษฎีบทที่ 1.15. *ถ้า $A \subset B$ แล้ว $Pr\left(A\right) \leqslant Pr\left(B\right)$*

การพิสูจน์. เนื่องจาก A และ $\left(B \cap A^c\right)$ ไม่มีส่วนร่วมต่อกัน (disjoint) และ $B = A \cup \left(B \cap A^c\right)$ ดังนั้น

$$Pr\left(B\right) = Pr\left(A\right) + Pr\left(B \cap A^c\right)$$

เนื่องจาก $Pr\left(B \cap A^c\right) \geqslant 0$ เราจึงสามารถสรุปได้ว่า $Pr\left(A\right) \leqslant Pr\left(B\right)$ ■

ทฤษฎีบทที่ 1.16. *สำหรับเหตุการณ์ A ใดๆ $0 \leqslant Pr\left(A\right) \leqslant 1$*

การพิสูจน์. ส่วนแรกเป็นผลมาจากสัจพจน์ที่ 1.1 ที่กำหนดให้ $0 \leqslant Pr\left(A\right)$ ส่วนที่สองเป็นผลมาจากทฤษฎีบทที่ 1.15 เนื่องจาก $A \subset S$ ดังนั้น $Pr\left(A\right) \leqslant Pr\left(S\right) = 1$ โดยที่สมการสุดท้ายเป็นผลมาจากสัจพจน์ที่ 1.2 ■

ทฤษฎีบทที่ 1.17. *สำหรับเหตุการณ์ A และ B ใดๆ*

$$Pr\left(A \cap B^c\right) = Pr\left(A\right) - Pr\left(A \cap B\right)$$

การพิสูจน์. จากทฤษฎีบทที่ 1.11 ซึ่งระบุว่า $A = (A \cap B) \cup (A \cap B^c)$ และ $(A \cap B)$ และ $(A \cap B^c)$ ไม่มีส่วนร่วมต่อกัน (disjoint) เราสามารถประยุกต์ใช้ทฤษฎีบทที่ 1.13 เพื่อแสดงว่า

$$Pr(A) = Pr(A \cap B) + Pr(A \cap B^c)$$

■

ทฤษฎีบทที่ 1.18. *สำหรับเหตุการณ์ A และ B ใดๆ*

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) \tag{1.6}$$

การพิสูจน์. จากทฤษฎีบทที่ 1.11 ซึ่งระบุว่า $A \cup B = B \cup (A \cap B^c)$ และ B และ $(A \cap B^c)$ ไม่มีส่วนร่วมต่อกัน (disjoint) เราสามารถประยุกต์ใช้ทฤษฎีบทที่ 1.13 เพื่อแสดงว่า

$$Pr(A \cup B) = Pr(B) + Pr(A \cap B^c)$$

เมื่อแทนค่า $Pr(A \cap B^c) = Pr(A) - Pr(A \cap B)$ จากทฤษฎีบทที่ 1.17 ลงไปจะได้ว่า

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

■

ทฤษฎีบทที่ 1.19. *สำหรับอนุกรมจำกัดของเหตุการณ์ $A_1, \ldots, A_n$*

$$Pr\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} Pr(A_i) \tag{1.7}$$

และ

$$Pr\left(\bigcap_{i=1}^{n} A_i\right) \geqslant 1 - \sum_{i=1}^{n} Pr(A_i^c) \tag{1.8}$$

การพิสูจน์. ดูจากแบบฝึกหัด ■

ทฤษฎีบทที่ 1.20. *สำหรับอนุกรมจำกัดของเหตุการณ์ $A_1, \ldots, A_n$*

$$\begin{aligned} Pr\left(\bigcup_{i=1}^{n} A_i\right) = &\sum_{i=1}^{n} Pr(A_i) - \sum_{i<j} Pr(A_i \cap A_j) + \sum_{i<j<k} Pr(A_i \cap A_j \cap A_k) \\ &- \sum_{i<j<k<l} Pr(A_i \cap A_j \cap A_k \cap A_l) + \cdots + (-1)^{n+1} Pr\left(\bigcap_{i=1}^{n} A_i\right) \end{aligned} \tag{1.9}$$

การพิสูจน์. ทฤษฎีบทนี้ใช้วิธีการพิสูจน์แบบอุปนัย (induction) โดยมีทฤษฎีบทที่ 1.4 ซึ่งระบุว่าทฤษฎีบทที่ต้องการพิสูจน์นี้เป็นจริงในกรณีที่ $n = 2$ ส่วนกรณีที่ $n = 1$ นั้นสามารถเข้าใจได้ไม่ยาก ดังนั้น ส่วนที่เหลือคือการพิสูจน์ว่าถ้าทฤษฎีบทนี้เป็นจริงสำหรับ n ใดๆ แล้วจะต้องเป็นจริงสำหรับ $m = n + 1$ เสมอ สำหรับกรณีที่ $n = 1$ ทฤษฎีบทที่ 1.18 ยืนยันแล้วว่า ทฤษฎีบทนี้เป็นจริงสำหรับ $m = 2$ ดังนั้น เราจำเป็นเพียงต้องพิสูจน์ว่าถ้าทฤษฎีบทนี้เป็นจริงสำหรับ $n \geqslant 2$ แล้วจะต้องเป็นจริงสำหรับ $m = n + 1$

สมมุติให้ทฤษฎบทนี้เป็นจริงสำหรับ $n \geqslant 2$ และกำหนดให้ $A_1, \ldots, A_{n+1}$ เป็นอนุกรมของเหตุการณ์ และ $A = \bigcup_{i=1}^{n} A_i$ และ $B = A_{n+1}$ เราสามารถประยุกต์ใช้ทฤษฎีบทที่ 1.18 ได้ว่า

$$\begin{aligned} Pr\left(\bigcup_{i=1}^{n+1} A_i\right) &= Pr\left(\left[\bigcup_{i=1}^{n} A_i\right] \cup A_{n+1}\right) \\ &= Pr\left(\bigcup_{i=1}^{n} A_i\right) + Pr\left(A_{n+1}\right) - Pr\left(\left[\bigcup_{i=1}^{n} A_i\right] \cap A_{n+1}\right) \end{aligned} \tag{1.10}$$

หลังจากนั้น เราสามารถประยุกต์ใช้คุณสมบัติการกระจายเพื่อแสดงว่า

$$\left[\bigcup_{i=1}^{n} A_i\right] \cap A_{n+1} = \bigcup_{i=1}^{n} (A_i \cap A_{n+1})$$

ซึ่งมีทั้งหมด n พจน์ ดังนั้น เราจึงสามารถประยุกต์ใช้สมการ 1.9 สำหรับ n พจน์ และแปลงพจน์ที่หนึ่งและพจน์ที่สามในสมการตามลำดับได้เป็น

$$\begin{aligned} Pr\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} Pr\left(A_i\right) - \sum_{i<j} Pr\left(A_i \cap A_j\right) + \sum_{i<j<k} Pr\left(A_i \cap A_j \cap A_k\right) \\ &\quad - \sum_{i<j<k<l} Pr\left(A_i \cap A_j \cap A_k \cap A_l\right) + \cdots + (-1)^{n+1} Pr\left(\bigcap_{i=1}^{n} A_i\right) \\ -Pr\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) &= -\sum_{i=1}^{n} Pr\left(A_i \cap A_{n+1}\right) + \sum_{i<j} Pr\left(A_i \cap A_j \cap A_{n+1}\right) \\ &\quad - \sum_{i<j<k} Pr\left(A_i \cap A_j \cap A_k \cap A_{n+1}\right) \\ &\quad + \sum_{i<j<k<l} Pr\left(A_i \cap A_j \cap A_k \cap A_l \cap A_{n+1}\right) + \cdots + (-1)^{n+2} Pr\left(\bigcap_{i=1}^{n+1} A_i\right) \end{aligned}$$

เมื่อแทนผลที่ได้เข้าไปในสมการที่ 1.10 เราจะได้ว่า

$$\begin{aligned} Pr\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^{n+1} Pr\left(A_i\right) - \sum_{i<j} Pr\left(A_i \cap A_j\right) + \sum_{i<j<k} Pr\left(A_i \cap A_j \cap A_k\right) \\ &\quad - \sum_{i<j<k<l} Pr\left(A_i \cap A_j \cap A_k \cap A_l\right) + \cdots + (-1)^{n+2} Pr\left(\bigcap_{i=1}^{n+1} A_i\right) \end{aligned}$$

ซึ่งเป็นการพิสูจน์ว่า หากทฤษฎีบทนี้เป็นจริงสำหรับ $n \geqslant 2$ แล้วจะต้องเป็นจริงสำหรับ $m = n + 1$ ■

ข้อสังเกต: เหตุการณ์ที่มีความน่าจะเป็นเท่ากับศูนย์ไม่ได้หมายความว่าเป็นไปไม่ได้ ในชีวิตประจำวัน เรามักจะบอกว่าเหตุการณ์ที่เป็นไปไม่ได้ (impossible events) มีความน่าจะเป็นเท่ากับศูนย์ แต่ในทางกลับกัน เหตุการณ์ที่มีความน่าจะเป็นเท่ากับศูนย์นั้นอาจจะเกิดขึ้นได้ ยกตัวอย่างเช่น เหตุการณ์ที่ปริมาณน้ำฝนที่ตกในวันพรุ่งนี้มีค่าระหว่าง $x - \epsilon$ และ $x + \epsilon$ มิลลิเมตร โดยที่ x และ ϵ เป็นค่าจำนวนจริงที่มากกว่าศูนย์ ความน่าจะเป็นของเหตุการณ์นี้เมื่อเราพิจารณาลิมิตที่ ϵ ลู่เข้าสู่ศูนย์จะเท่ากับศูนย์โดยอัตโนมัติ ซึ่งในทางคณิตศาสตร์ เราสามารถสรุปได้ว่า ความน่าจะเป็นของเหตุการณ์ที่ปริมาณน้ำฝนจะเท่ากับ x มิลลิเมตรมีค่าเท่ากับศูนย์ แต่ไม่ได้หมายความว่า เหตุการณ์นี้เกิดขึ้นไม่ได้ เพราะถ้าเป็นเช่นนั้นย่อมหมายความว่า จะไม่สามารถมีฝนตกในปริมาณใดๆ ได้เลย (เพราะเหตุผลที่ใช้เป็นจริงสำหรับค่าจำนวนจริง x ที่มากกว่าศูนย์ใดๆ) ในเชิงเทคนิค ตัวอย่างนี้เป็นคุณสมบัติพื้นฐานของการอินทิเกรต (integration)

บทที่ 2

ความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability)

เรามีคำถามมากมายในชีวิตประจำวันที่อยู่ในรูปแบบของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ยกตัวอย่างเช่น เราอาจจะอยากทราบว่า ผู้ชายมีโอกาสป่วยเป็นโรคหัวใจมากน้อยแค่ไหน? ซึ่งแตกต่างกับคำถามที่ว่าคนทั่วไปมีโอกาสป่วยเป็นโรคหัวใจมากน้อยแค่ไหนตรงที่ว่า ในกรณีแรกเรามีเงื่อนไขว่ากลุ่มประชากรที่เราสนใจต้องเป็นผู้ชายเท่านั้น ซึ่งเป็นเพียงส่วนหนึ่งของประชากรทั้งหมด นอกจากนี้ ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ยังเป็นรากฐานที่สำคัญของการพยากรณ์ (prediction) หรือการประมาณการณ์ (estimation) ซึ่งล้วนแล้วแต่เป็นการหาความคาดหวังแบบมีเงื่อนไข (conditional expectation) ซึ่งมีรากฐานมาจากความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ดังนั้น เราจำเป็นจะต้องเข้าใจเกี่ยวกับความน่าจะเป็นแบบมีเงื่อนไขให้ถ่องแท้ เพื่อให้สามารถพยากรณ์หรือประมาณการณ์ได้ถูกต้องและแม่นยำ นอกจากนี้ บทนี้ยังเป็นจุดเริ่มต้นของการนิยามความสัมพันธ์ระหว่างเหตุการณ์ (events) ว่าเป็นอิสระต่อกัน (independence) หรือไม่ รวมถึงทฤษฎีบทของเบส์ ซึ่งเป็นทฤษฎีบทที่สำคัญต่อสถิติแบบเบส์ (Bayesian statistics) ที่มีส่วนสำคัญอย่างมากต่อการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data analysis)

2.1 นิยามของความน่าจะเป็นแบบมีเงื่อนไข (Definition of Conditional Probability)

เรามักใช้ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) เมื่อเราทราบว่าเหตุการณ์อันใดอันหนึ่งได้เกิดขึ้นแล้ว เช่น เราได้ทราบแล้วว่าคนหนึ่งเป็นผู้ชาย ซึ่งแทนในรูปสัญลักษณ์ด้วยเหตุการณ์ B และต้องการหาความ

น่าจะเป็นของเหตุการณ์อีกอันหนึ่ง เช่น การป่วยเป็นโรคหัวใจซึ่งแทนในรูปสัญลักษณ์ด้วยเหตุการณ์ A โดยคำนึงถึงการที่เราทราบแล้วว่าเขาเป็นผู้ชาย กล่าวคือ เราต้องการหาความน่าจะเป็นที่ผู้ชายมีโอกาสป่วยเป็นโรคหัวใจ ซึ่งในที่นี้สามารถเขียนในรูปสัญลักษณ์ได้เป็น $Pr\left(A|B\right)$

บทนิยามที่ 2.1. *กำหนดให้ A และ B เป็นเหตุการณ์ใดๆ ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ของเหตุการณ์ A เมื่อกำหนดให้ว่าเหตุการณ์ B เกิดขึ้น คือ*

$$Pr\left(A|B\right) = \frac{Pr\left(A \cap B\right)}{Pr\left(B\right)} \tag{2.1}$$

โดยที่ $Pr\left(B\right) > 0$ จะต้องมีค่ามากกว่าศูนย์เท่านั้น หาก $Pr\left(B\right)$ มีค่าเท่ากับศูนย์แล้ว $Pr\left(A|B\right)$ จะไม่มีความหมาย

ข้อควรระวังอันหนึ่งคือ เหตุการณ์ B ไม่ได้จำเป็นต้องเกิดขึ้นก่อน A เลยแม้แต่น้อย ซึ่งสะท้อนให้เห็นว่า ความหมายของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ไม่มีส่วนเกี่ยวกับช่วงเวลาแต่อย่างใด แต่ในขณะเดียวกัน เหตุการณ์ทั้งสองเหตุการณ์นี้อาจจะเกิดขึ้นในคนละช่วงเวลาก็ได้ แต่จะต้องเป็นส่วนหนึ่งของปริภูมิตัวอย่าง (sample space) เดียวกัน

ตัวอย่างที่ 2.1. สมมุติว่าเราทราบว่าผลรวมของการโยนลูกเต๋าสองครั้ง Y มีค่าเป็นเลขคู่ แล้วความน่าจะเป็นที่ Y จะมีค่าน้อยกว่า 8 เป็นเท่าใด กำหนดให้ A คือเหตุการณ์ที่ Y จะมีค่าน้อยกว่า 8 และ B คือเหตุการณ์ที่ Y มีค่าเป็นเลขคู่ ดังนั้นสิ่งที่เราต้องการหาก็คือ $Pr\left(A|B\right)$ ซึ่งสามารถคำนวณได้ดังต่อไปนี้ เริ่มจากการหาความน่าจะเป็นของเหตุการณ์ B หรือความน่าจะเป็นของการที่ผลรวมของการโยนลูกเต๋าสองครั้ง Y มีค่าเป็นเลขคู่ ผลบวกจะเท่ากับ 2 ได้เพียงรูปแบบเดียวคือผลการโยนเป็น 1 ทั้งสองครั้ง เช่นเดียวกับกรณีที่ผลบวกเป็น 12 ในทำนองเดียวกันผลบวกจะเท่ากับ 4 ได้เพียง 3 รูปแบบ (โดยได้ 1 และ 3 หรือ 3 และ 1 หรือ 2 และ 2) เมื่อพิจรณาครบทุกกรณีของเลขคู่ที่เป็นไปได้จะพบว่า

$$Pr\left(B\right) = \frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \frac{5}{36} + \frac{3}{36} + \frac{1}{36} = \frac{18}{36} = \frac{1}{2}$$

ทั้งนี้ผลรวมของการโยนลูกเต๋าสองครั้ง Y มีทั้งหมด 36 รูปแบบที่แตกต่างกัน (มาจาก 6 คูณ 6) ในทำนองเดียวกันเราสามารถคำนวณหา ความน่าจะเป็นที่ ผลรวมของการโยนลูกเต๋าสองครั้ง Y มีค่าเป็นเลขคู่และมีค่าน้อยกว่า 8 ได้เป็น

$$Pr\left(A \cap B\right) = \frac{1}{36} + \frac{3}{36} + \frac{5}{36} = \frac{9}{36} = \frac{1}{4}$$

ดังนั้น ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ของเหตุการณ์ A เมื่อกำหนดให้ว่าเหตุการณ์ B เกิดขึ้น มีค่าเท่ากับ

$$Pr\left(A|B\right) = \frac{Pr\left(A \cap B\right)}{Pr\left(B\right)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

□

'ตัวอย่างต่อไปนี้แสดงให้เห็นถึงประโยชน์ในการคำนวณความน่าจะเป็นโดยประยุกต์ใช้หลักการของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ซึ่งช่วยให้สามารถโฟกัสกับผลลัพธ์ของการทดลองที่จำเป็นต่อการคำนวณเท่านั้น

ตัวอย่างที่ 2.2. พิจารณาการโยนลูกเต๋าสองลูกพร้อมกันซ้ำแล้วซ้ำอีก โดยให้ความสนใจกับผลรวมของการโยนลูกเต๋าทั้งสองลูกในแต่ละครั้ง Y เป็นหลัก สมมุติว่าเราต้องการทราบถึง ความน่าจะเป็นที่ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ ก่อนที่จะได้ค่า $Y = 8$

แน่นอนว่า เราสามารถใช้วิธีที่ตรงไปตรงมาที่สุดในการคำนวณคือ การหาจำนวนรูปแบบของผลการโยนที่จะทำให้ได้ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ ก่อนที่จะได้ค่า $Y = 8$ ทั้งหมด แล้วนำมาหารด้วยรูปแแบบของผลรวมของการโยนลูกเต๋าทั้งสองลูกที่เป็นได้ทั้งหมด ซึ่งเป็นงานที่ไม่ง่ายเลย

แต่หากเราประยุกต์ใช้หลักการของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ให้ดีแล้วเราจะสามารถหาผลลัพธ์ได้ง่ายกว่ามาก โดยเริ่มจากความเข้าใจที่ว่าหากเราเปลี่ยนการทดลองไปเล็กน้อยด้วยการโยนลูกเต๋าสองลูกพร้อมกันซ้ำแล้วซ้ำอีกเช่นเดิม แต่จะทำไปจนกระทั่งผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ หรือ $Y = 8$ แล้วจึงหยุด เมื่อมองผ่านการทดลองอันใหม่นี้ เราจะสามารถคำนวณหาความน่าจะเป็นที่ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ ก่อนที่จะได้ค่า $Y = 8$ (คำถามเริ่มต้น) ได้โดยพิจารณาเพียงว่าหากเราทราบว่าเหตุการณ์ B คือเหตุการณ์ที่ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ หรือ $Y = 8$ แล้ว ความน่าจะเป็นของเหตุการณ์ A ที่ผลรวมจะเป็น $Y = 7$ มีค่าเท่าใด เนื่องจากเหตุการณ์ A เป็นสับเซ็ตของเหตุการณ์ B ดังนั้น $Pr(A \cap B) = Pr(A)$ ซึ่งมีผลทำให้เราสามารถเขียนได้ว่า

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A)}{Pr(B)}$$

ซึ่งโดยหลักการแล้วการประยุกต์ใช้หลักการของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ในที่นี้ช่วยให้เราโฟกัสกับผลลัพธ์ของการทดลองที่จำเป็นต่อการคำนวณเท่านั้น ซึ่งในที่นี้ก็คือ ความน่าจะเป็นที่ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ ซึ่งมีค่าเท่ากับ $Pr(A) = \frac{6}{36}$ และผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ หรือ $Y = 8$ ซึ่งมีค่าเท่ากับ $Pr(B) = \frac{11}{36}$ ดังนั้น เราจึงสามารถสรุปได้ว่า ความน่าจะเป็นที่ผลรวมของการโยนลูกเต๋าทั้งสองลูก $Y = 7$ ก่อนที่จะได้ค่า $Y = 8$ มีค่าเท่ากับ $Pr(A|B) = \frac{6}{11}$ □

ทฤษฎีบทที่ 2.1 (กฎการคูณของความน่าจะเป็นแบบมีเงื่อนไข (multiplication rule for conditional probabilities))**.** *กำหนดให้ A และ B เป็นเหตุการณ์ใดๆ ถ้า $Pr(B) > 0$ แล้ว*

$$Pr(A \cap B) = Pr(B)\, Pr(A|B) \tag{2.2}$$

และในทำนองเดียวกัน ถ้า $Pr(A) > 0$ *แล้ว*

$$Pr(A \cap B) = Pr(A)\, Pr(B|A) \tag{2.3}$$

กฎการคูณของความน่าจะเป็นแบบมีเงื่อนไข (multiplication rule for conditional probabilities) นี้สามารถใช้ได้กับกรณีที่มีเหตุการณ์จำนวนจำกัดใดๆ ก็ได้ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 2.2. *กำหนดให้* $A_1, \ldots, A_n$ *เป็นเหตุการณ์ใดๆ โดยที่* $Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$ *แล้ว*

$$Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = Pr(A_1)\, Pr(A_2|A_1)\, Pr(A_3|A_1 \cap A_2) \cdots Pr(A_n|A_1 \cap \cdots \cap A_{n-1}) \tag{2.4}$$

การพิสูจน์. ประเด็นที่น่าสังเกตสำหรับการพิสูจน์ทฤษฎีบทนี้ก็คือ เราจะต้องแน่ใจก่อนว่าสามารถนิยามความน่าจะเป็นแบบมีเงื่อนไขได้หรือไม่ ซึ่งก็จะขึ้นอยู่กับว่าตัวหารของแต่ละพจน์มีค่ามากกว่าศูนย์หรือไม่ ส่วนที่เหลือก็เป็นเรื่องที่ไม่มีอะไรยุ่งยากนัก ดังนั้น เราจึงเริ่มด้วยการยืนยันว่าสามารถนิยามความน่าจะเป็นแบบมีเงื่อนไขได้ ดังต่อไปนี้

เนื่องจาก $Pr(A_1 \cap \cdots \cap A_{n-1}) > 0$ แต่ละพจน์ต่อไปนี้ $Pr(A_1 \cap \cdots \cap A_k) > 0$ สำหรับ $1 \leqslant k \leqslant n-2$ จะต้องมีค่ามากกว่าศูนย์ด้วยเช่นกัน ดังนั้น ความน่าจะเป็นแบบมีเงื่อนไข $Pr(A_{k+1}|A_1 \cap A_2 \cap \cdots \cap A_k)$ สำหรับ $1 \leqslant k \leqslant n-2$ มีความหมายและสามารถแทนได้ด้วย

$$Pr(A_{k+1}|A_1 \cap A_2 \cap \cdots \cap A_k) = \frac{Pr(A_1 \cap \cdots \cap A_{k+1})}{Pr(A_1 \cap \cdots \cap A_k)}$$

ดังนั้น เราสามารถเขียนพจน์ด้านขวาของสมการ 2.4 ใหม่ได้เป็น

$$Pr(A_1)\, \frac{Pr(A_1 \cap A_2)}{Pr(A_1)}\, \frac{Pr(A_1 \cap A_2 \cap A_3)}{Pr(A_1 \cap A_2)} \cdots \frac{Pr(A_1 \cap \cdots \cap A_n)}{Pr(A_1 \cap \cdots \cap A_{n-1})}$$

ซึ่งจะเห็นได้ว่าแต่ละพจน์ที่เป็นตัวตั้งมีตัวหารที่มีค่าเท่ากันซึ่งตัดกันไปได้ยกเว้นพจน์เดียวซึ่งก็คือผลลัพธ์สุดท้าย $Pr(A_1 \cap \cdots \cap A_n)$ ■

ตัวอย่างที่ 2.3. พิจารณาการทดลองที่ต้องจับลูกบอลครั้งละลูกทั้งหมด 4 ครั้ง โดยไม่ใส่ลูกบอลกลับเข้าไป จากกล่องที่มีลูกบอลสีแดง ρ ลูก และ ลูกบอลสีฟ้า β ลูก สมมุติว่าเราต้องการหาความน่าจะเป็นที่ผลของการจับลูกบอลเป็น สีแดง สีฟ้า สีแดง สีฟ้า แน่นอนว่า เราสามารถที่จะคำนวณหาความน่าจะเป็นนี้ได้โดยหาว่ามีรูปแบบกี่แบบที่จะได้ผลลัพธ์แบบนี้และนำไปหารด้วยจำนวนรูปแบบทั้งหมด แต่ในที่นี้ เราสามารถใช้เครื่องมือของกฎการคูณของความน่าจะเป็นแบบมีเงื่อนไขในการคำนวณ ซึ่งง่ายกว่ามาก

ก่อนอื่นเราจะต้องนิยามเหตุการณ์ R_j เพื่อแทนเหตุการณ์ที่จับได้ลูกบอลสีแดงในการจับครั้งที่ j^{th} และ B_j แทนเหตุการณ์ที่จับได้ลูกบอลสีฟ้าในการจับครั้งที่ j^{th} ดังนั้น เหตุการณ์ที่เป็นตัวแทนของการจับลูกบอลเป็น สีแดง สีฟ้า สีแดง สีฟ้า คือ $R_1 \cap B_2 \cap R_3 \cap B_4$ ดังนั้น สิ่งที่เราต้องการหาคือ

$$Pr\left(R_1 \cap B_2 \cap R_3 \cap B_4\right) = Pr\left(R_1\right) Pr\left(B_2|R_1\right) Pr\left(R_3|R_1 \cap B_2\right) Pr\left(B_4|R_1 \cap B_2 \cap R_3\right)$$

เราเหลือเพียงแค่ต้องหาค่าของความน่าจะเป็นแบบมีเงื่อนไขแต่ละพจน์ซึ่งมีค่าเท่ากับ

$$\begin{aligned} Pr\left(R_1\right) = \frac{\rho}{\rho+\beta}, Pr\left(B_2|R_1\right) = \frac{\beta}{\rho+\beta-1}, \\ Pr\left(R_3|R_1 \cap B_2\right) = \frac{\rho-1}{\rho+\beta-2}, Pr\left(B_4|R_1 \cap B_2 \cap R_3\right) = \frac{\beta-1}{\rho+\beta-3} \end{aligned}$$

ดังนั้น ความน่าจะเป็นที่ผลของการจับลูกบอลเป็น สีแดง สีฟ้า สีแดง สีฟ้า มีค่าเท่ากับ

$$Pr\left(R_1 \cap B_2 \cap R_3 \cap B_4\right) = \frac{\rho}{\rho+\beta} \times \frac{\beta}{\rho+\beta-1} \times \frac{\rho-1}{\rho+\beta-2} \times \frac{\beta-1}{\rho+\beta-3}$$

□

ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) เป็นมาตรวัดความน่าจะเป็น (probability measure) อันหนึ่ง

ข้อสังเกตที่น่าสนใจอันหนึ่งคือ ความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) เป็นความนาจะเป็นหรือมาตรวัดความน่าจะเป็น (probability measure) ที่สอดคล้องกับสัจพจน์ 1.1-1.3 โดยจะเห็นได้ว่า

1. $Pr\left(A|B\right) \geqslant 0$ ซึ่งเป็นผลมาจากการที่ $Pr\left(A \cap B\right) \geqslant 0$ และ $Pr\left(B\right) > 0$

2. $Pr\left(S|B\right) \geqslant 0$ ซึ่งเป็นผลมาจากการที่

$$Pr\left(S|B\right) = \frac{Pr\left(S \cap B\right)}{Pr\left(B\right)} = \frac{Pr\left(B\right)}{Pr\left(B\right)} = 1$$

   ซึ่งเป็นผลมาจากการที่ $Pr\left(S \cap B\right) = Pr\left(B\right)$

3. ความน่าจะเป็นแบบมีเงื่อนไขของยูเนียนของอนุกรมอนันต์ของเหตุการณ์ที่ไม่มีส่วนร่วมต่อกัน (disjoint events) $Pr\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} Pr\left(A_i|B\right)$ ซึ่งเป็นผลมาจากการที่

$$Pr\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{Pr\left(\left[\bigcup_{i=1}^{\infty} A_i\right] \cap B\right)}{Pr\left(B\right)} = \frac{Pr\left(\bigcup_{i=1}^{\infty} \left[A_i \cap B\right]\right)}{Pr\left(B\right)}$$

เนื่องจาก $A_1, A_2, \ldots$ เป็นอนุกรมอนันต์ของเหตุการณ์ที่ไม่มีส่วนร่วมต่อกัน (disjoint events) ดังนั้น $A_1 \cap B, A_2 \cap B, \ldots$ เป็นอนุกรมอนันต์ของเหตุการณ์ที่ไม่มีส่วนร่วมต่อกันเช่นเดียวกัน ดังนั้นเราสามารถประยุกต์ใช้สัจพจน์ที่ 1.3 เพื่อแสดงว่า

$$Pr\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{Pr\left(\bigcup_{i=1}^{\infty}\left[A_i \cap B\right]\right)}{Pr\left(B\right)} = \sum_{i=1}^{\infty} \frac{Pr\left(A \cap B\right)}{Pr\left(B\right)} = \sum_{i=1}^{\infty} Pr\left(A_i|B\right)$$

ผลตามมาก็คือว่า ทฤษฎีบทและคุณสมบัติต่างๆ ที่นำเสนอและพิสูจน์มาแล้วสำหรับความน่าจะเป็นในบทที่ 1 สามารถใช้ได้กับความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) เช่น กฎของการคูณของความน่าจะเป็นแบบมีเงื่อนไข ดังแสดงในทฤษฎีบทข้างล่าง

ทฤษฎีบทที่ 2.3. *กำหนดให้ $A_1, \ldots, A_n, B$ เป็นเหตุการณ์ใดๆ โดยที่ $Pr\left(A_1 \cap A_2 \cap \cdots \cap A_{n-1}|B\right) > 0$ และ $Pr\left(B\right) > 0$ แล้ว*

$$Pr\left(A_1 \cap A_2 \cap \cdots \cap A_n|B\right) = Pr\left(A_1|B\right) Pr\left(A_2|A_1 \cap B\right) Pr\left(A_3|A_1 \cap A_2 \cap B\right) \cdots$$
$$\cdots Pr\left(A_n|A_1 \cap \cdots \cap A_{n-1} \cap B\right) \tag{2.5}$$

สมการนี้คล้ายกับสมการ 2.4 มาก ส่วนที่แตกต่างมีเพียงการมีเงื่อนไขที่ขึ้นอยู่กับเหตุการณ์ B ในทุกๆ พจน์นั่นเอง

ส่วนต่อไปเป็นการเชื่อมโยงระหว่างความน่าจะเป็นแบบมีเงื่อนไขและการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง (sample space) เพื่อใช้ในการคำนวณความน่าจะเป็นของเหตุการณ์

ถึงแม้ว่าเราจะนิยามการแบ่งส่วน (partition) ไว้แล้วในบทนิยามที่ 1.10 แต่เพื่อความสะดวกเราจึงนิยามอีกครั้งดังต่อไปนี้

บทนิยามที่ 2.2. *กำหนดให้ S คือปริภูมิตัวอย่าง (sample space) ของการทดลองอันหนึ่ง และ $B_1, B_2, \ldots, B_k$ คือเหตุการณ์ใดๆ ในปริภูมิตัวอย่าง S ที่ไม่มีส่วนร่วมต่อกัน (disjoint) และ $\bigcup_{i=1}^{k} B_i = S$ ในกรณีนี้ เราเรียกเหตุการณ์เหล่านี้ว่าเป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง S*

ทฤษฎีบทที่ 2.4 (กฎของความน่าจะเป็นรวม (Law of total probability))**.** *กำหนดให้ $B_1, B_2, \ldots, B_k$ เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง S และ $Pr\left(B_i\right) > 0$ สำหรับ $i = 1, \ldots, k$ ดังนั้น สำหรับเหตุการณ์ $A \subset S$ ใดๆ*

$$Pr\left(A\right) = \sum_{i=1}^{k} Pr\left(B_i\right) Pr\left(A|B_i\right) \tag{2.6}$$

การพิสูจน์. เนื่องจาก $B_1, B_2, \ldots, B_k$ เป็นการแบ่งส่วนของปริภูมิตัวอย่าง S ดังนั้น $B_1 \cap A, B_2 \cap A, \ldots, B_k \cap A$ เป็นการแบ่งส่วนของ A ซึ่งก็หมายความว่า

$$A = \bigcup_{i=1}^{k} B_i \cap A$$

ดังนั้น

$$Pr\left(A\right) = Pr\left(\bigcup_{i=1}^{k} B_i \cap A\right) = \sum_{i=1}^{k} Pr\left(B_i\right) Pr\left(A|B_i\right)$$

โดยที่สมการสุดท้ายเป็นผลจากการที่ $B_1 \cap A, B_2 \cap A, \ldots, B_k \cap A$ ไม่เกิดร่วมกัน (disjoint) ■

ตัวอย่างที่ 2.4. สมมุติว่ามีกล่องบรรจุลูกบอลทั้งหมด 2 กล่อง โดยกล่องแรกมีลูกบอลสีฟ้า 70 ลูก และลูกบอลสีแดง 30 ลูก ส่วนกล่องที่สองมีลูกบอลสีฟ้า 10 ลูก และลูกบอลสีแดง 30 ลูก กำหนดให้ B_1 และ B_2 คือเหตุการณ์ที่กล่องที่หนึ่งและกล่องที่สองถูกเลือกตามลำดับ คำถามก็คือ หากการเลือกกล่องเป็นแบบสุ่ม (random) ความน่าจะเป็นของการจับได้ลูกบอลสีฟ้าเป็นเท่าใด?

เพื่อตอบคำถาม กำหนดให้ A คือเหตุการณ์ที่จับได้ลูกบอลสีฟ้า ดังนั้น ความน่าจะเป็นของการจับได้ลูกบอลสีฟ้าสามารถเขียนในรูปของกฎของความน่าจะเป็นรวมได้เป็น

$$Pr\left(A\right) = Pr\left(B_1\right) Pr\left(A|B_1\right) + Pr\left(B_2\right) Pr\left(A|B_2\right)$$

เนื่องเราสุ่มเลือกกล่อง ดังนั้น $Pr\left(B_1\right) = Pr\left(B_2\right) = \frac{1}{2}$ และเราสามารถคำนวณหาพจน์ที่เหลือในด้านขวามือได้ดังต่อไปนี้

$$Pr\left(A|B_1\right) = \frac{7}{10}, Pr\left(A|B_2\right) = \frac{1}{4}$$

ดังนั้น ความน่าจะเป็นของการจับได้ลูกบอลสีฟ้าเท่ากับ

$$Pr\left(A\right) = \frac{1}{2} \times \frac{7}{10} + \frac{1}{2} \times \frac{1}{4} = \frac{19}{40}$$

ซึ่งมีค่าต่างจากการที่หากเราคิดโดยรวมเอาทั้งสองกล่องเข้าด้วยกันทำให้มีลูกบอลสีฟ้า 80 ลูก และลูกบอลสีแดง 60 ลูก ซึ่งในกรณีนี้จะได้ความน่าจะเป็นของการจับได้ลูกบอลสีฟ้าเท่ากับ $\frac{4}{7}$ แทน บทเรียนที่ได้ในที่นี้คือ ลำดับของการทดลองหรือการเลือกมีส่วนสำคัญต่อความน่าจะเป็นของเหตุการณ์ □

ตัวอย่างที่ 2.5. พิจารณาเกมส์จับฉลากที่ผู้เล่นจะจับฉลากที่มีตัวเลขตั้งแต่ 1 ถึง 100 และทุกครั้งที่จับเสร็จจะนำฉลากใส่กลับ กำหนดให้ X เป็นคะแนนที่จับได้ครั้งแรก แต่ละคนจะจับต่อไปจนกว่าจะได้ตัวเลข $Y \geqslant X$ สิ่งที่เราสนใจคือ ความน่าจะเป็นที่จะได้คะแนน $Y = 100$

กำหนดให้ B_j คือเหตุการณ์ที่ $X = j$ โดยที่ $j = 1, 2, \ldots, 100$ และ A คือเหตุการณ์ที่ $Y = 100$ เมื่อพิจารณาในรูปแบบของความน่าจะเป็นแบบมีเงื่อนไขจะพบว่า หากเริ่มจากเหตุการณ์ B_j คะแนนที่เป็นไปได้สำหรับ $Y = j, j+1, \ldots, 100$ (เพราะ $Y \geqslant X$) ซึ่งมีทั้งหมด $101 - j$ และเนื่องจากเราใส่ฉลากกลับไป ความน่าจะเป็นของแต่ละคะแนนในการจับแต่ละครั้งมีค่าเท่ากัน ดังนั้น

$$Pr\left(A|B_j\right) = \frac{1}{101 - j}$$

นอกจากนี้ เรายังทราบอีกว่า $Pr\left(B_j\right) = \frac{1}{100}$ สำหรับทุกค่า $j = 1, 2, \ldots, 100$ ดังนั้น เราสามารถประยุกต์ใช้กฎของความน่าจะเป็นรวมเพื่อคำนวณหาค่าความน่าจะเป็นที่จะได้คะแนน $Y = 100$ ได้ดังต่อไปนี้

$$\begin{aligned} Pr\left(A\right) &= \sum_{j=1}^{100} Pr\left(B_j\right) Pr\left(A|B_j\right) \\ &= \frac{1}{100}\left[\frac{1}{100} + \frac{1}{99} + \cdots + 1\right] = 0.0519 \end{aligned}$$

□

ทฤษฎีบทต่อไปนี้เป็นส่วนขยายของกฎของความน่าจะเป็นรวมไปสู่ความน่าจะเป็นแบบมีเงื่อนไข โดยใช้หลักการที่ว่า ความน่าจะเป็นแบบมีเงื่อนไขก็เป็นความน่าจะเป็นหรือมาตรวัดความน่าจะเป็นอันหนึ่ง

ทฤษฎีบทที่ 2.5. *กำหนดให้ $B_1, B_2, \ldots, B_k$ เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง S โดยที่ $Pr\left(B_i\right) > 0$ สำหรับ $i = 1, \ldots, k$ และ C เป็นเหตุการณ์ใดๆ ที่ $Pr\left(C\right) > 0$ ดังนั้น สำหรับเหตุการณ์ $A \subset S$ ใดๆ*

$$Pr\left(A|C\right) = \sum_{i=1}^{k} Pr\left(B_i|C\right) Pr\left(A|B_i \cap C\right) \tag{2.7}$$

ตัวอย่างต่อไปนี้แสดงให้เห็นถึงการประยุกต์ใช้หลักการทดลองเสริม (augmented experiment) ซึ่งช่วยให้เราสามารถแก้ปัญหาได้ด้วยการเปลี่ยนมุมมองในการมองการทดลองใดๆ ให้อยู่ในรูปของการแบ่งส่วน (partition)

ตัวอย่างที่ 2.6. พิจารณาการทดลองอันหนึ่งซึ่งคล้ายคลึงกับการทดลองในตัวอย่างที่ 2.4 แต่ไม่เหมือนกันซะทีเดียว โดยในที่นี้มีกล่องอยู่แค่หนึ่งอันที่มีลูกบอลสีฟ้าและสีแดงอยู่ข้างใน แต่เนื่องจากฝาปิดสนิทอยู่และหากุญแจยังไม่เจอ จึงไม่ทราบว่ามีสีฟ้าและสีแดงอย่างละเท่าใด แต่ต้องการทราบว่า ความน่าจะเป็นในการหยิบได้ลูกบอลสีฟ้าเป็นเท่าใด นาย ก ซึ่งต้องการทราบค่าความน่าจะเป็นจึงไปถาม นาย ข ซึ่งเคยเปิดกล่องมาแล้ว และได้คำตอบว่ามีลูกบอลสีฟ้า 70 ลูก และลูกบอลสีแดง 30 ลูก แต่พอไปถามนาย ค ซึ่งก็เคยเปิดกล่องมาแล้วเช่นกัน กลับได้คำตอบว่ามีลูกบอลสีฟ้า 10 ลูก และลูกบอลสีแดง 30 ลูก สมมุติว่านาย ก มีความเชื่อมั่นในคำตอบของนาย ข และนาย ค พอๆ กัน วิธีการแก้ปัญหาที่เหมาะสมในกรณีนี้คือ การกำหนดให้คำตอบของนาย ข เป็นกล่องใบแรก

ซึ่งแทนด้วยเหตุการณ์ B_1 และคำตอบของนาย ค เป็นกล่องใบที่สองซึ่งแทนด้วยเหตุการณ์ B_2 ซึ่งจะทำให้ปัญหานี้กลายเป็นปัญหาเดียวกับในตัวอย่างที่ 2.4 จะเห็นได้ว่า การสร้างการทดลองเสริม (augmented experiment) ขึ้นมาให้มีกล่องสองใบ ทั้งที่ในความเป็นจริงมีกล่องเพียงแค่หนึ่งใบ ช่วยให้เราหาคำตอบได้สะดวกมากยิ่งขึ้น □

2.2 ความเป็นอิสระต่อกัน (Independent Events)

เหตุการณ์สองเหตุการณ์จะเป็นอิสระต่อกันก็ต่อเมื่อ การทราบว่าเหตุการณ์หนึ่งเกิดขึ้นแล้วไม่ได้มีผลกระทบต่อความน่าจะเป็นของอีกเหตุการณ์หนึ่ง กล่าวคือ ถ้า A และ B คือเหตุการณ์ใดๆ ที่ $Pr(A) > 0$ และ $Pr(B) > 0$ สองเหตุการณ์นี้จะเป็นอิสระต่อกัน ก็ต่อเมื่อ $Pr(A|B) = Pr(A)$ และ $Pr(B|A) = Pr(B)$ อย่างไรก็ตาม เนื่องจากทั้งสองเงื่อนไขนำไปสู่ข้อสรุปเดียวกันคือ $Pr(A \cap B) = Pr(A) Pr(B)$ เราจึงมักนิยามความเป็นอิสระต่อกันในรูปของผลคูณของความน่าจะเป็นดังนี้

บทนิยามที่ 2.3. *เหตุการณ์ A และ B เป็นอิสระต่อกัน (independent) ถ้า*

$$Pr(A \cap B) = Pr(A) Pr(B) \tag{2.8}$$

สังเกตว่านิยามความเป็นอิสระต่อกันนี้ไม่จำเป็นต้องสมมุติว่า $Pr(A) > 0$ และ $Pr(B) > 0$ แน่นอนว่าหากเราสมมุติเงื่อนไขทั้งสองนี้เราจะสามารถนิยามความเป็นอิสระต่อกันได้ว่า เหตุการณ์ A และ B เป็นอิสระต่อกัน (independent) *ก็ต่อเมื่อ* สมการ (2.8) เป็นจริง

ตัวอย่างต่อไปนี้แสดงให้เห็นว่าความเป็นอิสระต่อกันเป็นคุณสมบัติของเหตุการณ์ไม่ใช่คุณสมบัติของการทดลองหรือมาตรวัดความน่าจะเป็น ด้วยตัวอย่างของสองเหตุการณ์ที่เป็นผลจากการโยนลูกเต๋าครั้งเดียวกัน แต่มีอิสระต่อกัน ในขณะที่ หากเราเปลี่ยนเป็นเหตุการณ์คู่ใหม่กลับพบว่าไม่ได้เป็นอิสระต่อกัน

ตัวอย่างที่ 2.7. พิจารณาการโยนลูกเต๋าหนึ่งครั้ง กำหนดให้ A เป็นเหตุการณ์ที่ผลการโยนเป็นเลขคี่ และ B เป็นเหตุการณ์ที่ผลการโยนเป็นตัวเลขใดตัวเลขหนึ่งในเซ็ต $\{1, 2, 3, 4\}$ ความน่าจะเป็นของทั้งสองเหตุการณ์เท่ากับ $Pr(A) = \frac{1}{2}$ และ $Pr(B) = \frac{2}{3}$ ส่วนที่สำคัญที่จะช่วยตรวจสอบว่าสองเหตุการณ์นี้เป็นอิสระต่อกันหรือไม่ คือ $Pr(A \cap B)$ ซึ่งหาได้จากการที่ $A \cap B = \{1, 3\}$ ดังนั้น $Pr(A \cap B) = \frac{1}{3}$ ซึ่งมีค่าเท่ากับ $Pr(A) \times Pr(B) = \frac{1}{2} \times \frac{2}{3}$ ดังนั้น เราจึงสามารถสรุปได้ว่า A และ B เป็นอิสระต่อกัน (independent)

อย่างไรก็ตาม หากเราพิจารณาเหตุการณ์ใหม่ C ซึ่งแทนเหตุการณ์ที่ผลการโยนเป็นตัวเลขใดตัวเลขหนึ่งในเซ็ต $\{1, 3, 4\}$ และ D ซึ่งแทนเหตุการณ์ที่ผลการโยนเป็นตัวเลขใดตัวเลขหนึ่งในเซ็ต $\{1, 2, 5\}$ ดังนั้น $C \cap D = \{1\}$ ดังนั้น ในกรณีนี้ $Pr(C) = \frac{1}{2}$ และ $Pr(D) = \frac{1}{2}$ ซึ่งหมายความว่า $Pr(C) Pr(D) = \frac{1}{4}$ แต่ $Pr(C \cap D) = \frac{1}{6}$ นั่นคือ เหตุการณ์ C และ D ไม่เป็นอิสระต่อกัน

□

ทฤษฎีบทที่ 2.6. *ถ้าเหตุการณ์ A และ B เป็นอิสระต่อกัน (independent) แล้วเหตุการณ์ A และ B^c เป็นอิสระต่อกัน*

การพิสูจน์. จากทฤษฎีบทที่ว่า

$$Pr\left(A \cap B^c\right) = Pr\left(A\right) - Pr\left(A \cap B\right)$$

และจากความเป็นอิสระต่อกันของ A และ B ซึ่งหมายความว่า

$$Pr\left(A \cap B\right) = Pr\left(A\right) Pr\left(B\right)$$

เราสามารถสรุปได้ว่า

$$Pr\left(A \cap B^c\right) = Pr\left(A\right) - Pr\left(A\right) Pr\left(B\right) = Pr\left(A\right)\left[1 - Pr\left(B\right)\right] = Pr\left(A\right) Pr\left(B^c\right)$$

ซึ่งหมายความว่า A และ B^c เป็นอิสระต่อกัน ■

ความเป็นอิสระต่อกันสามารถขยายไปถึงกลุ่มเหตุการณ์ที่มากกว่าสองเหตุการณ์ได้ดังแสดงในนิยามข้างล่าง

บทนิยามที่ 2.4. *เหตุการณ์ $A_1, \ldots, A_n$ เป็นอิสระต่อกัน (independent) ถ้าทุกๆ กลุ่มเหตุการณ์ k เหตุการณ์ใดๆ $A_{i_1}, \ldots, A_{i_k}$ สำหรับ $k = 2, \ldots, n$*

$$Pr\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) = Pr\left(A_{i_1}\right) \cap \cdots \cap Pr\left(A_{i_k}\right) \tag{2.9}$$

โดยที่ดัชนี $i_j \in \{1, 2, \ldots, n\}$ คือดัชนีที่บอกว่าเป็นเหตุการณ์ใดเช่น $i_1 = 4$ หมายความว่า $A_{i_1} = A_4$

เราสามารถสสรุปความหมายของความเป็นอิสระต่อกันในรูปที่ไม่เป็นทางการได้ว่า ความเป็นอิสระต่อกันของกลุ่มใหญ่ย่อมหมายถึงความเป็นอิสระต่อกันของกลุ่มย่อยเสมอ แต่ในทางกลับกันไม่จำเป็นต้องเป็นจริง ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 2.8. พิจารณาการโยนเหรียญสองครั้งติดต่อกัน โดยมีปริภูมิตัวอย่าง $S = \{HH, HT, TH, TT\}$ โดยที่ H แทนผลการโยนที่เป็นหัว และ T แทนผลการโยนที่เป็นก้อย กำหนดให้

$$A = \{HH, HT\},$$
$$B = \{HH, TH\},$$
$$C = \{HH, TT\},$$

ดังนั้น $Pr(A) = Pr(B) = Pr(C) = \frac{1}{2}$ ในขณะเดียวกัน เราสามารถแสดงได้ว่า $A \cap B = A \cap C = B \cap C = \{HH\}$ ซึ่งทำให้เราสามารถคำนวณได้ว่า

$$Pr(A \cap B) = Pr(A \cap C) = Pr(B \cap C) = \frac{1}{4}$$

โดยสรุปเหตุการณ์ A, B และ C เป็นอิสระต่อกันเป็นคู่ (pairwise independent) แต่เหตุการณ์เหล่านี้ไม่ได้เป็นอิสระต่อกันเพราะ

$$Pr(A \cap B \cap C) = \frac{1}{4}$$

ซึ่งมีค่าแตกต่างจาก $Pr(A)\,Pr(B)\,Pr(C)$ □

ตัวอย่างที่ 2.9. พิจารณาการโยนเหรียญจนกระทั่งได้ผลเป็นหัว ความน่าจะเป็นที่จะได้ผลลัพธ์เป็นหัวหลังจากโยนไปแล้ว n ครั้งคำนวณได้จากการที่การโยนแต่ละครั้งเป็นอิสระต่อกันและความน่าจะเป็นที่จะได้ผลการโยนเป็นหัวเท่ากับ $\frac{1}{2}$ เหตุการณ์ที่เราสนใจคือ ผลการโยนตลอด $n-1$ ครั้งแรกได้ผลเป็นก้อยและผลการโยนครั้งสุดท้ายเป็นหัว กำหนดให้ A_i คือเหตุการณ์ที่ผลการโยนครั้งที่ i^{th} ออกเป็นหัว ดังนั้น สิ่งที่เราต้องการคือ $Pr\left(A_1^c \cap \cdots A_{n-1}^c \cap A_n\right)$ เนื่องจากเหตุการณ์เหล่านี้เป็นอิสระต่อกัน เราจึงสามารถคำนวณได้ว่า

$$Pr\left(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n\right) = Pr\left(A_1^c\right) \cdots Pr\left(A_{n-1}^c\right) Pr\left(A_n\right) = \left(\frac{1}{2}\right)^n$$

ในขณะเดียวกัน เราสามารถนิยามเหตุการณ์ของการที่ผลการโยนจะเป็นหัวครั้งแรกในการโยนครั้งที่ n เท่ากับ $B_n = A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n$ ดังนั้น เหตุการณ์ที่ผลการโยนจะต้องออกเป็นหัวไม่ครั้งใดก็ครั้งหนึ่งหากเราโยนไปเรื่อยๆ สามารถนิยามได้เป็น

$$C = \bigcup_{i=1}^{\infty} B_i$$

ในขณะเดียวกัน $B_i \cap B_j = \varnothing$ สำหรับ $i \neq j$ ใดๆ ดังนั้น ความน่าจะเป็นของเหตุการณ์ที่ผลการโยนจะต้องออกเป็นหัวไม่ครั้งใดก็ครั้งหนึ่งเท่ากับ

$$Pr(C) = Pr\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} Pr(B_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1$$

ซึ่งก็เป็นไปตามสามัญสำนึกที่ว่าเราจะต้องได้ผลการโยนเป็นหัวแน่ๆ ขึ้นอยู่เพียงว่าจะเป็นผลของการโยนครั้งที่เท่าไหร่เท่านั้นเอง □

ทฤษฎีบทต่อไปนี้ นำเสนอการเป็นอิสระต่อกันในรูปแบบของความน่าจะเป็นแบบมีเงื่อนไขสำหรับกลุ่มเหตุการณ์ใดๆ

ทฤษฎีบทที่ 2.7. *กำหนดให้ $A_1, \ldots, A_n$ เป็นเหตุการณ์ใดๆ โดยที่ $Pr\left(A_1 \cap \cdots \cap A_{n-1} \cap A_n\right) > 0$ ดังนั้น $A_1, \ldots, A_n$ เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ สำหรับเซ็ตที่ไม่มีส่วนร่วมกันสองเซ็ต $\{i_1, \ldots, i_m\}$ และ $\{j_1, \ldots, j_l\}$ ซึ่งเป็นสับเซ็ตของ $\{1, \ldots, n\}$*

$$Pr\left(A_{i_1} \cap \cdots \cap A_{i_m} | A_{j_1} \cap \cdots \cap A_{j_l}\right) = Pr\left(A_{i_1} \cap \cdots \cap A_{i_m}\right) \tag{2.10}$$

แน่นอนว่าในแง่ของการตีความ นิยามการเป็นอิสระต่อกันในรูปแบบของความน่าจะเป็นแบบความน่าจะเป็นแบบมีเงื่อนไขให้ความหมายที่ตรงไปตรงมาที่สุด กล่าวคือ ถ้าการเรียนรู้เกี่ยวกับเหตุการณ์ B ไม่ได้มีผลต่อความน่าจะเป็นของเหตุการณ์ A ย่อมแสดงว่า เหตุการณ์ทั้งสองเป็นอิสระต่อกัน (independent)

เช่นเดียวกับก่อนหน้านี้ เนื่องจากความน่าจะเป็นแบบมีเงื่อนไขก็คือความน่าจะเป็นอันหนึ่ง ดังนั้น เราจึงสามารถนิยามการเป็นอิสระต่อกันแบบมีเงื่อนไขและพิสูจน์ทฤษฎีบทที่เกี่ยวข้องได้ในทำนองเดียวกับ การเป็นอิสระต่อกันที่นำเสนอมาก่อนหน้านี้

บทนิยามที่ 2.5. *กลุ่มเหตุการณ์ $A_1, \ldots, A_n$ เป็นอิสระต่อกันแบบมีเงื่อนไข (conditionally independent) ภายใต้เงื่อนไขของเหตุการณ์ B ถ้า*

$$Pr\left(A_{i_1} \cap \cdots \cap A_{i_m} | B\right) = Pr\left(A_{i_1} | B\right) \cdots Pr\left(A_{i_m} | B\right) \tag{2.11}$$

สำหรับทุกๆ กลุ่มย่อย $A_{i_1}, \ldots, A_{i_m}$ ของ $2 \leqslant m \leqslant n$ เหตุการณ์

ทฤษฎีบทที่ 2.8. *กำหนดให้ A_1, A_2, B เป็นเหตุการณ์ใดๆ โดยที่ $Pr\left(A \cap B\right) > 0$ แล้ว A_1 และ A_2 จะเป็นอิสระต่อกันแบบมีเงื่อนไข (conditionally independent) ภายใต้เงื่อนไขของ B ก็ต่อเมื่อ*

$$Pr\left(A_2 | A_1 \cap B\right) = Pr\left(A_2 | B\right) \tag{2.12}$$

ข้อสังเกต: ความสัมพันธ์ระหว่างการเป็นอิสระต่อกัน (independent) และการไม่มีส่วนร่วมต่อกัน (mutually exclusive) ถึงแม้ว่าชื่อของทั้งสองหลักการจะดูมีส่วนคล้ายกันไม่น้อย แต่ในความเป็นจริงแล้ว กลุ่มของเหตุการณ์โดยทั่วไปจะไม่สามารถสอดคล้องกับทั้งสองหลักการพร้อมกันได้ ทั้งนี้ เนื่องจากการเป็นอิสระต่อกัน (independent) เป็นเงื่อนไขที่กำหนดว่า การเรียนรู้เกี่ยวกับเหตุการณ์หนึ่งจะต้องไม่มีผลต่อความน่าจะเป็นของเหตุการณ์หนึ่ง ในขณะที่ การไม่มีส่วนร่วม (mutually exclusive or disjoint) กลับบอกว่าเมื่อทราบว่าเหตุการณ์หนึ่งเกิดแล้วย่อมทราบทันทีว่าอีกเหตุการณ์หนึ่งเกิดขึ้นไม่ได้ กล่าวคือ ภายใต้หลักการไม่มีส่วนร่วมต่อกัน การเรียนรู้เกี่ยวกับเหตุการณ์หนึ่งมีผลต่อความน่าจะเป็นของเหตุการณ์หนึ่งเสมอ ยกเว้นในกรณีที่ เหตุการณ์อันหลังนั้นมีความน่าจะเป็นเท่ากับศูนย์อยู่แล้ว ดังนั้น เราสรุปได้ว่า หลักการสองอันนี้จะเป็นจริงพร้อมกันไม่ได้ ยกเว้นในกรณีที เหตุการณ์ทุกเหตุการณ์ยกเว้นเพียงหนึ่งอันมีความน่าจะเป็นเท่ากับศูนย์

ทฤษฎีบทที่ 2.9. *กำหนดให้กลุ่มเหตุการณ์ $A_1, \ldots, A_n$ ไม่มีส่วนร่วมต่อกัน (mutually exclusive or disjoint) เหตุการณ์กลุ่มนี้จะเป็นอิสระต่อกัน (mutually independent) ก็ต่อเมื่อ ทุกเหตุการณ์ยกเว้นเพียงหนึ่งเหตุการณ์มีความน่าจะเป็นเท่ากับศูนย์*

2.3 ทฤษฎบทของเบส์ (Bayes' Theorem)

หัวข้อนี้นำเสนอทฤษฎีบทของเบส์ซึ่งเป็นรากฐานสำคัญของทฤษฎีการเรียนรู้ (learning theory) ในสถิติ โดยในปัจจุบัน ได้รับความสนใจมากขึ้นในรูปแบบของการปรับปรุงแบบเบส์ (Baysean updating) ซึ่งเป็นเครื่องมือหนึ่งที่ได้รับความนิยมในการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data)

ทฤษฎีบทที่ 2.10 (ทฤษฎีบทของเบส์ (Bayes' theorem))**.** *กำหนดให้ $B_1, B_2, \ldots, B_k$ เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง S โดยที่ $Pr(B_i) > 0$ สำหรับ $i = 1, \ldots, k$ และ A เป็นเหตุการณ์ใดๆ ที่ $Pr(A) > 0$ ดังนั้น สำหรับ $i = 1, \ldots, k$*

$$Pr(B_i|A) = \frac{Pr(B_i)\, Pr(A|B_i)}{\sum_{j=1}^{k} Pr(B_j)\, Pr(A|B_j)} \tag{2.13}$$

ในมุมมองของสถิติแบบเบส์ (Bayesian Statistics) $Pr(B_i)$ มักจะถูกเรียกว่าความน่าจะเป็นก่อน (prior probability) ซึ่งสะท้อนถึงความรู้หรือความเชื่อที่ผู้วิเคราะห์มีต่อเหตุการณ์ B_i ซึ่งจะถูกปรับปรุงเมื่อมีข้อมูลใหม่เพิ่มเข้ามาซึ่งในที่นี้หมายถึงเหตุการณ์ A และผลลัพธ์ที่ได้จากการปรับปรุง (updating) $Pr(B_i|A)$ มักจะถูกเรียกว่าความน่าจะเป็นหลัง (posterior probability)

ตัวอย่างต่อไปนี้เป็นการประยุกต์ใช้ทฤษฎีบทของเบส์ (Bayes' theorem) กับคำถามทางการแพทย์ที่มีการผสมผสานข้อมูลภาพรวมและผลการทดสอบที่ไม่สมบูรณ์เข้าด้วยกัน ซึ่งเป็นรูปแบบของคำถามที่นำมาซึ่งการออกแบบหรือการค้นพบทฤษฎีบทของเบส์ (Bayes' theorem)

ตัวอย่างที่ 2.10. พิจารณาเทคโนโลยีการตรวจการติดเชื้ออันหนึ่งที่มีระดับความเชื่อมั่น (reliability) ที่ 90 เปอร์เซนต์ ซึ่งหมายความว่า หากคนใดติดเชื้อจริง ความน่าจะเป็นที่ผลการตรวจจะออกมาเป็นบวกหรือบอกว่าติดเชื้อมีค่าเท่ากับ 0.90 ในขณะที่ ความน่าจะเป็นที่ผลการตรวจของคนที่ไม่ติดเชื้อจะออกมาเป็นบวกมีค่าเท่ากับ 0.10 นอกจากนี้ ข้อมูลในภาพรวมยังบอกอีกว่า โอกาสในการติดเชื้อในประชากรอยู่ที่ประมาณ 1 ใน 10,000 สมมุติว่า ผู้ชายคนหนึ่งไปตรวจและพบว่าได้ผลการตรวจเป็นบวก คำถามคือ ความน่าจะเป็นที่จริงๆ แล้วเขาติดเชื้อมีค่าเท่าใด?

กำหนดให้ A แทนเหตุการณ์ที่ผลของการตรวจเชื้อออกมาเป็นบวก ส่วน B_1 แทนเหตุการณ์ที่ผู้ชายคนนั้นติดเชื้อ และ B_2 แทนเหตุการณ์ที่เขาไม่ได้ติดเชื้อ แน่นอนว่า B_1 และ B_2 เป็นการแบ่งส่วน (partition) ดังนั้น สิ่งที่

เราต้องการทราบคือ $Pr\left(B_1|A\right)$ ซึ่งสามารถคำนวณหาได้โดยใช้ทฤษฎีบทของเบส์ (Bayes' theorem) ดังต่อไปนี้

$$Pr\left(B_1|A\right) = \frac{Pr\left(B_1\right) Pr\left(A|B_1\right)}{Pr\left(B_1\right) Pr\left(A|B_1\right) + Pr\left(B_2\right) Pr\left(A|B_2\right)}$$

ซึ่งจะเห็นได้ว่า เราจำเป็นต้องใช้ความน่าจะเป็นที่มาจากข้อมูลในภาพรวม $Pr\left(B_1\right)$ และ $Pr\left(B_2\right)$ ประกอบในการคำนวณด้วย ซึ่งในที่นี้แต่ละอันมีค่าเท่ากับ $Pr\left(B_1\right) = 0.0001$ (จาก 1 ใน 10,000) และ $Pr\left(B_2\right) = 0.9999$ ดังนั้น ความน่าจะเป็นที่จริงๆ แล้วเขาติดเชื้อมีค่าเท่ากับ

$$Pr\left(B_1|A\right) = \frac{0.0001 \times 0.90}{0.0001 \times 0.90 + 0.9999 \times 0.10} = 0.0009$$

ซึ่งน้อยมากเมื่อเปรียบเทียบกับระดับความน่าเชื่อถือ (reliability) ของการตรวจสอบซึ่งอยู่ที่ระดับ 90 เปอร์เซนต์ อันที่จริงแล้วผลลัพธ์ที่ได้นี้เป็นผลมาจากการที่เราทราบดีอยู่แล้วว่า จะมีเพียงหนึ่งคนในทุกๆ หนึ่งหมื่นคนติดเชื้อ แต่ในขณะเดียวกันการตรวจสอบจะให้ผลเป็นบวกหนึ่งคนในทุกๆ สิบคนที่ทำการตรวจ ดังนั้น หากมีคนรับการตรวจทั้งหมด 10,000 คน จะพบว่าประมาณ 1,000 คนจะได้ผลตรวจที่เป็นบวก ทั้งที่มีเพียงคนเดียวที่ติดเชื้อ นั่นคือ ผลการตรวจของ 999 คนเป็นผลที่ผิด หรือกล่าวอีกนัยหนึ่งได้ว่า โอกาสที่ผลการตรวจที่เป็นบวกจะหมายถึงการติดเชื้อจริงมีแค่ 1 ใน 1,000 ดังนั้น เมื่อรวมกับระดับความน่าเชื่อถือ 0.9 ทำให้เราพบว่า ความน่าจะเป็นที่จริงๆ แล้วเขาติดเชื้อมีค่าเท่ากับ $\frac{0.9}{1000} = 0.0009$ (ซึ่งเป็นตัวเลขโดยประมาณ)

เพื่อให้เห็นผลของข้อมูลภาพรวม เราลองสมมุติว่าโอกาสในการติดเชื้อในประชากรอยู่ที่ประมาณ 1 ใน 100 ในกรณีนี้ ความน่าจะเป็นที่จริงๆ แล้วเขาติดเชื้อมีค่าเท่ากับ

$$Pr\left(B_1|A\right) = \frac{0.01 \times 0.90}{0.01 \times 0.90 + 0.99 \times 0.10} = 0.0833$$

ซึ่งชี้ให้เห็นว่า หากโอกาสในการติดเชื้อในประชากรไม่สูงมากนัก การตรวจการติดเชื้อที่มีระดับความเชื่อมั่นไม่ถึงร้อยเปอร์เซนต์อาจจะไม่คุ้มค่าก็เป็นได้

โดยสรุป บทเรียนจากตัวอย่างนี้สอนให้เราระมัดระวังในการสรุปผลการทดสอบใดๆ ซึ่งมักจะมีโอกาสที่จะผิดพลาดเสมอ และที่สำคัญการสรุปที่ถูกต้องจำเป็นต้องใช้ข้อมูลทั้งหมดที่มีอยู่ □

ทฤษฎีบทต่อไปนี้เป็นทฤษฎีบทของเบส์ (Bayes' theorem) แบบมีเงื่อนไข

ทฤษฎีบทที่ 2.11. *กำหนดให้ $B_1, B_2, \ldots, B_k$ เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง S โดยที่ $Pr\left(B_i\right) > 0$ สำหรับ $i = 1, \ldots, k$ และ A และ C เป็นเหตุการณ์ใดๆ ที่ $Pr\left(A\right) > 0$ และ $Pr\left(C\right) > 0$ ดังนั้น สำหรับ $i = 1, \ldots, k$*

$$Pr\left(B_i|A \cap C\right) = \frac{Pr\left(B_i|C\right) Pr\left(A|B_i \cap C\right)}{\sum_{j=1}^{k} Pr\left(B_j|C\right) Pr\left(A|B_j \cap C\right)} \tag{2.14}$$

ตัวอย่างต่อไปนี้แสดงวิธีการคำนวณความน่าจะเป็นหลัง (posterior probability) ของการทดลองที่มี 2 ขั้นตอน ซึ่งส่งผลให้เราสามารถคำนวณได้หลายวิธี

ตัวอย่างที่ 2.11. กล่องหนึ่งมีเหรียญอยู่สองเหรียญ โดยอันหนึ่งเป็นเหรียญที่เที่ยงตรง (fair coin) ส่วนอีกอันหนึ่งมีด้านทั้งสองด้านเป็นหัวเพียงอย่างเดียว พิจารณาการทดลองที่เมื่อเราจะสุ่มเลือกเหรียญขึ้นมาและโยนเหรียญนั้นและได้ผลลัพธ์เป็นหัว และเมื่อโยนอีกครั้งก็ได้ผลลัพธ์เป็นหัวเช่นเดิม คำถามก็คือ ความน่าจะเป็นที่เหรียญที่หยิบขึ้นมาจะเป็นเหรียญที่เที่ยงตรงมีค่าเท่าใด

พิจารณาการโยนครั้งที่หนึ่งดังต่อไปนี้ กำหนดให้ B_1 เป็นเหตุการณ์ที่เหรียญเป็นแบบเที่ยงตรง ในขณะที่ B_2 เป็นเหตุการณ์ที่เหรียญเป็นแบบมีหัวทั้งสองด้าน ส่วน H_1 คือเหตุการณ์ที่ผลการโยนครั้งที่หนึ่งออกเป็นหัว ดังนั้น โดยใช้ทฤษฎีบทของเบส์ เราสามารถคำนวณได้ว่า

$$\begin{aligned} Pr\left(B_1|H_1\right) &= \frac{Pr\left(B_1\right)Pr\left(H_1|B_1\right)}{Pr\left(B_1\right)Pr\left(H_1|B_1\right)+Pr\left(B_2\right)Pr\left(H_1|B_2\right)} \\ &= \frac{\frac{1}{2}\times\frac{1}{2}}{\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times 1} = \frac{1}{3} \end{aligned}$$

สังเกตว่าเนื่องจากการหยิบเหรียญเป็นแบบสุ่ม ดังนั้น จึงเป็นเรื่องปกติที่จะสมมุติให้ $Pr\left(B_1\right) = Pr\left(B_2\right) = \frac{1}{2}$

ต่อไปเราจะพิจารณาการโยนครั้งที่สอง โดยในส่วนนี้เราสามารถคำนวณได้สองวิธี ดังต่อไปนี้ กำหนดให้ H_2 คือเหตุการณ์ที่ผลการโยนครั้งที่สองออกเป็นหัว และสมมุติว่าการโยนเหรียญทั้งสองครั้งนี้เป็นอิสระต่อกันอย่างมีเงื่อนไขภายใต้เหตุการณ์ B_1 และ B_2

1. **การประยุกต์ใช้ทฤษฎีบทของเบส์ครั้งเดียว:** สิ่งเราต้องการคำนวณหาสามารถเขียนในรูปสัญลักษณ์ได้เป็น $Pr\left(B_1|H_1\cap H_2\right)$ นั่นคือ เราแทนเหตุการณ์ที่ผลการโยนทั้งสองครั้งออกมาเป็นหัวทั้งคู่ด้วย $H_1\cap H_2$ นอกจากนี้ เนื่องจากการโยนเหรียญทั้งสองครั้งนี้เป็นอิสระต่อกันอย่างมีเงื่อนไขภายใต้เหตุการณ์ B_1 เราจึงสามารถคำนวณได้ว่า $Pr\left(H_1\cap H_2|B_1\right) = Pr\left(H_1|B_1\right)Pr\left(H_2|B_1\right) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$ ในทำนองเดียวกัน $Pr\left(H_1\cap H_2|B_2\right) = Pr\left(H_1|B_2\right)Pr\left(H_2|B_2\right) = 1\times 1 = 1$ ดังนั้น โดยใช้ทฤษฎีบทของเบส์ เราสามารถคำนวณได้ว่า

$$\begin{aligned} Pr\left(B_1|H_1\cap H_2\right) &= \frac{Pr\left(B_1\right)Pr\left(H_1\cap H_2|B_1\right)}{Pr\left(B_1\right)Pr\left(H_1\cap H_2|B_1\right)+Pr\left(B_2\right)Pr\left(H_1\cap H_2|B_2\right)} \\ &= \frac{\frac{1}{2}\times\frac{1}{4}}{\frac{1}{2}\times\frac{1}{4}+\frac{1}{2}\times 1} = \frac{1}{5} \end{aligned}$$

2. **การประยุกต์ใช้ทฤษฎีบทของเบส์ต่อเนื่องหลายขั้น:** ส่วนนี้จะประยุกต์ใช้ทฤษฎีบทของเบส์แบบมีเงื่อนไขโดยเริ่มจากผลการโยนครังที่หนึ่ง ซึ่งเราพบว่า $Pr\left(B_1|H_1\right) = \frac{1}{3}$ ซึ่งหมายความว่า $Pr\left(B_2|H_1\right) = \frac{2}{3}$

หลังจากนั้น เราสามารถใช้ความน่าจะเป็นทั้งสองนี้เป็นความน่าจะเป็นก่อน (prior probability) ในขั้นต่อไป โดยใช้ทฤษฎีบทของเบส์แบบมีเงื่อนไข ดังนี้

$$Pr\left(B_1|H_1 \cap H_2\right) = \frac{Pr\left(B_1|H_1\right) Pr\left(H_2|B_1 \cap H_1\right)}{Pr\left(B_1|H_1\right) Pr\left(H_2|B_1 \cap H_1\right) + Pr\left(B_2|H_1\right) Pr\left(H_2|B_2 \cap H_1\right)}$$

โดยในที่นี้เรากำหนดให้ $A = H_2$ และ $C = H_1$ สิ่งที่เราต้องการเพิ่มเติมก็คือ $Pr\left(H_2|B_1 \cap H_1\right) = \frac{1}{2}$ และ $Pr\left(H_2|B_2 \cap H_1\right) = 1$ ดังนั้น

$$Pr\left(B_1|H_1 \cap H_2\right) = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times 1} = \frac{1}{5}$$

ซึ่งมีค่าเท่ากับผลที่คำนวณได้โดยใช้ทฤษฎีบทของเบส์ครั้งเดียว

□

บทที่ 3

ตัวแปรสุ่มและการแจกแจง (Random Variables and Ditributions)

ถึงแม้ว่าในทางทฤษฎีแล้ว สถิติมีรากฐานมาจากความน่าจะเป็น (probability) ที่อภิปรายมาแล้วในบทที่ 1 และ 2 แต่ในทางปฏิบัติการนิยามความน่าจะเป็นในรูปแบบของเหตุการณ์นั้นอาจจะไม่สะดวกมากนัก เนื่องจากบางครั้งผลลัพธ์ของการทดลองอาจจะไม่ได้อยู่ในรูปของตัวเลขที่สามารถนำมาเปรียบเทียบกันโดยตรง ซึ่งทำให้ยากต่อการเปรียบเทียบ และไม่สามารถประยุกต์ใช้เครื่องมือทางคณิตศาสตร์ที่สามารถใช้ได้กับจำนวนจริง เช่น แคลคูลัส เป็นต้น

บทนี้จึงนำเสนอเครื่องมือที่สำคัญที่แปลงค่าผลลัพธ์ของการทดลองให้อยู่ในรูปของจำนวนจริงหรือตัวเลข ซึ่งช่วยให้เราสามารถประยุกต์ใช้ทฤษฎีสถิติในการวิเคราะห์ข้อมูลต่างๆ ได้สะดวก นั่นคือ ตัวแปรสุ่ม (random variable) สิ่งที่ผู้อ่านต้องให้ความสำคัญที่สุดไม่ใช่นิยามของตัวแปรสุ่มว่าเขียนว่าอย่างไร แต่ต้องศึกษาให้ดีเพื่อให้เกิดความเข้าใจและเกิดความรู้สึกเชื่อมโยงกับความหมายของตัวแปรสุ่ม ซึ่งหากเข้าใจดีแล้วย่อมจะสามารถบอกได้ว่า เราจะสามารถบอกคุณสมบัติของตัวแปรสุ่มแต่ละตัวได้อย่างไร ซึ่งหากสรุปแบบคร่าวๆ จะได้ว่า เรามักจะอธิบายตัวแปรสุ่มแต่ละตัวด้วย ฟังก์ชันความน่าจะเป็น (distribution function) ของมัน กล่าวคือ ทุกครั้งที่เรานึกถึงตัวแปรสุ่ม เราควรต้องถามตัวเองทันทีว่า มันมีการแจกแจงตัวอย่างไร เพราะคุณสมบัติทุกอย่างของมันได้ถูกบรรจุไว้ในฟังก์ชันความน่าจะเป็นของมัน หวังเป็นอย่างยิ่งว่า หลังจากศึกษาบทนี้อย่างละเอียดแล้ว ผู้อ่านจะถามตัวเองทุกครั้งที่เห็นตัวแปรสุ่มว่า การแจกแจงของตัวแปรนั้นเป็นอย่างไร?

3.1 ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variables)

ตัวอย่างของตัวแปรสุ่มง่ายๆ ที่เราคุ้นเคยอันหนึ่งซึ่งเป็นผลจากการทดลองสุ่มเลือกคนมาหนึ่งคนจากประชากร และดูว่าคนที่สุ่มมาได้เป็นผู้หญิงหรือผู้ชาย ตัวแปรสุ่มนี้คือ เพศ (gender) โดยอาจจะกำหนดให้เพศชายมีค่าเท่ากับ 1 และเพศหญิงมีค่าเท่ากับ 3 นั่นคือ ตัวแปรสุ่มจะแปลงผลลัพธ์ของการทดลองให้เป็นตัวเลข ดังนิยามต่อไปนี้

บทนิยามที่ 3.1 (ตัวแปรสุ่ม (random variable))**.** *ตัวแปรสุ่ม (random variable)* X *คือฟังก์ชันค่าจริง (real-valued function) ซึ่งกำหนดค่าจำนวนจริงสำหรับแต่ละผลลัพธ์ (outcomes) หรือเหตุการณ์ (events) ในปริภูมิตัวอย่าง (sample space) ซึ่งสามารถเขียนในรูปแบบทางคณิตศาสตร์ได้เป็น* $X : S \to \mathbb{R}$

โดยทั่วไป เรามักจะแทนค่าที่เกิดขึ้นจริง (realization) ของตัวแปรสุ่ม X สำหรับผลลัพธ์ $s \in S$ ในรูป $X(s) = x$ ซึ่งเป็นค่าจำนวนจริง ทำให้เราสามารถวิเคราะห์บนค่าจำนวนจริงโดยไม่ต้องสนใจว่าเหตุการณ์ที่อยู่เบื้องหลังคืออะไร ซึ่งช่วยให้การวิเคราะห์มีความสะดวกอย่างมาก

ตัวอย่างที่ 3.1. พิจารณาการทดลองโยนเหรียญ 5 ครั้ง ดังนั้น ปริภูมิตัวอย่าง (sample space) ในกรณีนี้คือ

$$S = \Big\{ s : s \text{ ผลของการโยนเหรียญ 5 ครั้ง} \Big\}$$

สมมุติว่าสิ่งที่เราสนใจคือจำนวนครั้งที่ผลการโยนเป็นก้อย ซึ่งสามารถแทนได้ด้วยตัวแปรสุ่ม X ซึ่งเป็นฟังก์ชันที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย เช่น ค่าที่เกิดขึ้นจริง (realization) สำหรับผลลัพธ์ $s = HTTHT$ (H แทนหัว และ T แทนก้อย) คือ $X(s) = 3$ ในขณะเดียวกันหากเราสนใจจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นหัว เราสามารถแทนด้วยตัวแปรสุ่ม Y ซึ่งเป็นฟังก์ชันที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นหัว เช่น ค่าที่เกิดขึ้นจริง (realization) สำหรับผลลัพธ์ $s = HTTHT$ (H แทนหัว และ T แทนก้อย) คือ $Y(s) = 2$ บทเรียนจากตัวอย่างนี้คือ เราสามารถสร้างตัวแปรสุ่มสำหรับการทดลองอันหนึ่งได้ไม่จำกัด ในขณะเดียวกัน เราจะเลือกใช้ตัวแปรสุ่มที่ช่วยให้เราวิเคราะห์ปัญหาของเราได้สะดวกที่สุด นั่นคือ นักวิเคราะห์มีทางเลือกในการนิยามตัวแปรสุ่มมากมาย และควรเลือกให้ดี

นอกจากนี้ เรายังสามารถสรุปได้ว่า $Y(s) = 5 - X(s)$ ซึ่งเป็นการแทนทางพีชคณิต (algebraic representation) ของผลลัพธ์ที่เกิดจากการทดลอง การที่เราสามารถใช้การดำเนินการทางพีชคณิต (algebraic operation) กับตัวแปรสุ่มได้ ในขณะที่เราอาจจะไม่สามารถทำได้กับผลลัพธ์จากการทดลองโดยตรง เป็นประโยชน์ที่สำคัญของการวิเคราะห์ความน่าจะเป็นโดยใช้ตัวแปรสุ่ม □

ตัวอย่างต่อไปนี้แสดงให้เห็นว่า เราสามารถนิยามตัวแปรสุ่มแบบไม่ต่อเนื่อง (discrete random variable) จากปริภูมิตัวอย่างที่ค่าของผลลัพธ์เป็นแบบต่อเนื่องได้ ซึ่งเป็นการตอกย้ำถึงการที่นักวิเคราะห์มีทางเลือกในการออกแบบตัวแปรสุ่มได้มากมาย

ตัวอย่างที่ 3.2. พิจารณาการทดลองที่ผลลัพธ์คือความต้องการใช้ไฟฟ้าและน้ำประปา โดยในที่นี้สมมุติให้มีค่าตั้งแต่ 0 ถึง 1 ทั้งสองอย่าง กำหนดให้ผลลัพธ์ของการทดลองนี้อยู่ในรูป (x, y) โดยที่ $x \in [0, 1]$ คือความต้องการใช้ไฟฟ้า และ $y \in [0, 1]$ คือความต้องการใช้น้ำประปา สมมุติว่าเราสนใจว่าความต้องการใช้ไฟฟ้าและน้ำประปาสูงเกินไปหรือไม่ โดยกำหนดว่า ความต้องการที่สูงเกินไปคือระดับความต้องการที่มากกว่า 0.5 สำหรับการใช้ไฟฟ้าและ 0.75 สำหรับการใช้น้ำประปา ดังนั้น เราสามารถสร้างตัวแปรสุ่มเพื่อแทนสิ่งที่เราต้องการได้เป็น

$$Z(x, y) = \begin{cases} 1, \text{ ถ้า } x > 0.5 \text{ และ } y > 0.75 \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

โดยที่ $s = (x, y)$ คือผลลัพธ์ของการทดลองใดๆ □

คำถามที่สำคัญต่อไปคือ ตัวแปรสุ่มที่นิยามมานี้มีความสัมพันธ์กับความน่าจะเป็นที่นิยามมาก่อนหน้านี้อย่างไร?

3.1.1 การแจกแจงของตัวแปรสุ่มไม่ต่อเนื่อง (The Distribution of a Discrete Random Variable)

ดังที่อภิปรายมาแล้วในบทที่ 1 เราสามารถนิยามความน่าจะเป็นที่ถูกต้องตามหลักการ (well-defined probability) สำหรับปริภูมิตัวอย่าง S ของการทดลองอันหนึ่งได้มากมาย ดังนั้น เพื่อให้การอภิปรายมีความชัดเจน เราจำเป็นต้องเริ่มด้วยการกำหนดก่อนว่าเรากำลังใช้ความน่าจะเป็นหรือมาตรวัดความน่าจะเป็นอันใด ดังที่นำเสนอในบทที่ 1 เราแทนความน่าจะเป็นของเหตุการณ์ A ซึ่งเป็นสับเซ็ตของปริภูมิตัวอย่าง S ด้วย $Pr(A)$ ดังนั้น สิ่งที่เราต้องทำก็คือ สร้างเหตุการณ์จากตัวแปรสุ่มที่เราสนใจ โดยกำหนดให้ C คือสับเซ็ตของจำนวนจริงที่ $\{s : X(s) \in C\}$ หรือเขียนแบบย่อได้เป็น $\{X \in C\}$ เป็นเหตุการณ์ ซึ่งเป็นสับเซ็ตของปริภูมิตัวอย่าง S ยกตัวอย่างเช่น กรณีของการทดลองโยนเหรียญห้าครั้ง สมมุติว่า $C = \{1\}$ และ X คือตัวแปรสุ่มที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย เหตุการณ์ที่ได้จาก C คือ $A = \{THHHH, HTHHH, \ldots, HHHHT\}$ โดยที่ H แทนหัว และ T แทนก้อย นั่นคือ $X(s) = 1$ สำหรับทุกๆ $s \in A$ จากการนิยามแบบนี้ทำให้เราสามารถกำหนดความน่าจะเป็นจากตัวแปรสุ่ม X ได้เป็น $Pr(X \in C) = Pr(X \in A) = \frac{5}{32}$ สำหรับ $C = \{1\}$ ผลของการรวบรวมความน่าจะเป็นในรูปแบบนี้ทั้งหมดคือ การแจกแจงของตัวแปรสุ่ม X

บทนิยามที่ 3.2 (การแจกแจงของตัวแปรสุ่ม (distribution of random variable))**.** *สำหรับตัวแปรสุ่ม X และความน่าจะเป็น $Pr(\cdot)$ ที่นิยามบนปริภูมิตัวอย่าง S การแจกแจงของตัวแปรสุ่ม (distribution of random variable) X คือเซ็ตของความน่าจะเป็นทุกอันที่อยู่ในรูปแบบ $Pr(X \in C)$ สำหรับเซ็ตของจำนวนจริง C ทุกเซ็ต โดยที่ $\{X \in C\}$ เป็นเหตุการณ์*

ข้อสังเกตอันหนึ่งคือ การแจกแจงของ X ก็คือมาตรวัดความน่าจะเป็น (probability measure) ที่นิยามบนเซ็ตของจำนวนจริงนั่นเอง ซึ่งสะท้อนข้อเท็จจริงที่ว่า ตัวแปรสุ่มคือฟังก์ชันจำนวนจริงที่แปลงผลลัพธ์ของการทดลองเป็นค่าจำนวนจริง

ตัวอย่างที่ 3.3. พิจารณาการทดลองโยนเหรียญที่เที่ยงตรงจำนวน 5 ครั้ง เช่นเดียวกับตัวอย่างที่ 3.1 เราสนใจในตัวแปรสุ่ม X ที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย ค่าที่เป็นไปได้ทั้งหมดของ X คือ $0, 1, 2, 3, 4, 5$ ในการหาความน่าจะเป็นเราสามารถใช้ค่าสัมประสิทธิ์ทวินาม (binomial coefficient) ช่วยในการคำนวณดังต่อไปนี้ จำนวนผลลัพธ์ที่จะมีก้อย 0 ครั้งจากการโยนทั้งหมด 5 ครั้งเท่ากับ $\binom{5}{0} = \frac{5!}{0!5!} = 1$ ในทำนองเดียวกัน จำนวนผลลัพธ์ที่จะมีก้อย 3 ครั้งจากการโยนทั้งหมด 5 ครั้งเท่ากับ $\binom{5}{3} = \frac{5!}{3!2!} = 10$ โดยสรุป การแจกแจงของตัวแปรสุ่ม X สามารถรวบรวมได้เป็น $Pr(X \in \{0\}) = \frac{1}{32}$, $Pr(X \in \{1\}) = \frac{5}{32}$, $Pr(X \in \{2\}) = \frac{10}{32}$, $Pr(X \in \{3\}) = \frac{10}{32}$, $Pr(X \in \{4\}) = \frac{5}{32}$ และ $Pr(X \in \{5\}) = \frac{1}{32}$ ดูรูปที่ XXX ประกอบ □

บทนิยามที่ 3.3. *ตัวแปรสุ่ม X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) หรือเป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) ถ้าค่าจำนวนจริงที่เป็นไปได้ของ X มีจำนวนจำกัด $x_1, \ldots, x_n$ หรืออย่างมากต้องเป็นอนุกรมอนันต์ที่นับได้ $x_1, x_2, \ldots$*

บทนิยามที่ 3.4 (ฟังก์ชันความน่าจะเป็น (probability function))**.** *ถ้าตัวแปรสุ่ม X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) ฟังก์ชันความน่าจะเป็น (probability function หรือ p.f.) ของ X คือฟังก์ชัน f โดยที่สำหรับทุกค่าจำนวนจริง x*

$$f(x) = Pr(X = x) \tag{3.1}$$

นอกจากนี้ เราเรียกส่วนปิดคลุม (closure) ของเซ็ต $\{x : f(x) > 0\}$ ว่า ส่วนค้ำจุน (support) ของ X

ตัวอย่างที่ 3.4. ฟังก์ชันความน่าจะเป็น (probability function) ของตัวแปรสุ่ม X ในตัวอย่างที่ 3.3 สามารถเขียนได้เป็น

$$f(0) = \frac{1}{32}, f(1) = \frac{5}{32}, f(2) = \frac{10}{32}, f(3) = \frac{10}{32}, f(4) = \frac{5}{32}, f(5) = \frac{1}{32}$$

ในขณะที่ส่วนค้ำจุน (support) ของ X คือ $\{0, 1, 2, 3, 4, 5\}$ □

คุณสมบัติสำคัญอันหนึ่งของฟังก์ชันความน่าจะเป็น (probability function) ของตัวแปรสุ่มคือ ผลรวมจากทุกค่าในส่วนค้ำจุนจะต้องเท่ากับหนึ่ง ทั้งนี้เพราะค่าแต่ละค่าในส่วนค้ำจุนเป็นผลมาจากเหตุการณ์ที่ไม่มีส่วนร่วมต่อกัน (disjoint) และเมื่อมองในรูปของเหตุการณ์ ทุกค่าในส่วนค้ำจุนรวมกันเป็นปริภูมิตัวอย่าง ดังนั้น ผลรวมของความน่าจะเป็นของทุกค่าจากส่วนค้ำจุนจึงเท่ากับความน่าจะเป็นของปริภูมิตัวอย่างซึ่งมีค่าเท่ากับหนึ่ง

ทฤษฎีบทที่ 3.1. *กำหนดให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) ที่มี f แทนฟังก์ชันความน่าจะเป็น (probability function) ถ้า x ไม่ใช่ค่าที่เป็นไปได้ $f(x) = 0$ และ ถ้าอนุกรม $x_1, x_2, \ldots$ ประกอบด้วยค่าที่เป็นไปได้ทั้งหมด แล้ว*

$$\sum_{i=1}^{\infty} f(x_i) = 1 \tag{3.2}$$

ทฤษฎีบทต่อไปนี้บอกความสัมพันธ์ระหว่างความน่าจะเป็นที่นิยามบนเหตุการณ์และฟังก์ชันความน่าจะเป็น (probability function) ที่นิยามบนจำนวนจริง

ทฤษฎีบทที่ 3.2. *ถ้า X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) แล้ว เราสามารถคำนวณหาค่าความน่าจะเป็นของสับเซ็ตของจำนวนจริง $C \subset \mathsf{R}$ ได้จาก*

$$Pr(X \in C) = \sum_{x_i \in C} f(x_i) \tag{3.3}$$

ตัวแปรสุ่มไม่ต่อเนื่องบางตัวเป็นที่นิยมมากทำให้มีชื่อเฉพาะ เช่น ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ซึ่งมีฟังก์ชันความน่าจะเป็นดังต่อไปนี้

บทนิยามที่ 3.5. *ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีค่าพารามิเตอร์ p คือ ตัวแปรสุ่ม X ที่มีค่าได้แค่สองค่าคือ 0 และ 1 โดยที่ $Pr(X = 1) = p$ และมีการแจกแจงแบบเบอร์นูลี (Bernoulli distribution) ที่มีค่าพารามิเตอร์ p ซึ่งสามารถเขียนแทนได้ด้วย*

$$f(x) = p^x (1-p)^{1-x}, \text{ สำหรับ } x = 0, 1 \tag{3.4}$$

สังเกตได้ว่า การระบุเพียงชื่อของตัวแปรสุ่มหรือการแจกแจงไม่เพียงพอที่จะบอกถึงคุณสมบัติของตัวแปรสุ่มนั้น ทั้งนี้เนื่องจากตัวแปรสุ่มประเภทเดียวกันแต่ค่าพารามิเตอร์ที่ต่างกันย่อมเป็นตัวแปรสุ่มคนละตัวกัน ดังนั้นจำเป็นจะต้องระบุทั้งชื่อและค่าพารามิเตอร์ทุกครั้งที่อ้างอิงถึงตัวแปรสุ่ม ในทางกลับกัน สิ่งที่ต้องการทราบเพื่อบอกถึงคุณสมบัติของตัวแปรสุ่มใดๆ ก็คือการแจกแจงของมัน ซึ่งในที่นี้สามารถทราบได้จากชื่อของตัวแปรสุ่มและค่าพารามิเตอร์ที่เกี่ยวข้อง

ตัวแปรสุ่มไม่ต่อเนื่องที่มีชื่อเสียงอีกอันหนึ่งคือ ตัวแปรสุ่มทวินาม (binomial random variable) ที่มีพารามิเตอร์ n และ p ซึ่งเกิดจากการสุ่มเลือกตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p ต่อเนื่องกันจำนวน n ครั้งดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 3.5. พิจารณาการตรวจสอบคุณภาพของบะหมี่กึ่งสำเร็จรูป โดยที่ความน่าจะเป็นที่จะเกิดผลิตภัณฑ์ด้อยคุณภาพเท่ากับ p ดังนั้นความน่าจะเป็นที่จะได้ผลิตภัณฑ์ที่ได้คุณภาพเท่ากับ $1 - p$ สมมุติว่าเจ้าหน้าที่ทำการ

ตรวจสอบผลิตภัณฑ์ทั้งหมด n ชิ้น และเหตุการณ์ที่บอกถึงผลการตรวจสอบเป็นอิสระต่อกันหรือการตรวจสอบแต่ละครั้งเป็นอิสระต่อกัน ความน่าจะเป็นที่ผลการตรวจสอบแต่ละครั้งจะออกมาว่ามีผลิตภัณฑ์ด้อยคุณภาพทั้งหมด $0 \leqslant x \leqslant n$ ชิ้นเท่ากับ $\binom{n}{x} p^x (1-p)^{n-x}$ ทั้งนี้เพราะมีทั้งหมด $\binom{n}{x}$ รูปแบบที่ผลการทดสอบจะพบว่ามีผลิตภัณฑ์ด้อยคุณภาพทั้งหมด x ชิ้น และความน่าจะเป็นของแต่ละรูปแบบเท่ากับผลคูณของความน่าจะเป็นของผลลัพธ์ที่เกิดขึ้นในแต่ละครั้งซึ่งจะต้องมี x ครั้งจากทั้งหมด n ครั้ง ที่พบผลิตภัณฑ์ด้อยคุณภาพ ในขณะที่ อีก $n-x$ ครั้งจะพบผลิตภัณฑ์ที่ได้คุณภาพ ทั้งนี้เพราะการตรวจสอบแต่ละครั้งเป็นอิสระต่อกัน ดังนั้นความน่าจะเป็นของแต่ละรูปแบบเท่ากับ $p^x (1-p)^{n-x}$

นอกจากนี้ เราสามารถมองได้ว่า การตรวจสอบแต่ละครั้งคือ ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p หนึ่งตัว ดังนั้น การตรวจสอบทั้งหมด n ชิ้นจึงเสมือนกับ ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p และเป็นอิสระต่อกันทั้งหมด n ตัว ดังนั้น ความน่าจะเป็นของผลการตรวจสอบ n ชิ้นมีค่าเท่ากับ

$$\begin{aligned} Pr(Y_1 = y_1, \ldots, Y_n = y_n) &= Pr(Y_1 = y_1) \cdots Pr(Y_n = y_n) \\ &= p^{y_1} (1-p)^{1-y_1} \cdots p^{y_n} (1-p)^{1-y_n} = p^{\sum_{i=1}^n y_i} (1-p)^{1-\sum_{i=1}^n y_i} \end{aligned}$$

สำหรับกรณีที่มีผลิตภัณฑ์ด้อยคุณภาพทั้งหมด x ชิ้น จะได้ว่า $\sum_{i=1}^n y_i = x$ ความน่าจะเป็นของรูปแบบนี้มีค่าเท่ากับ $p^x (1-p)^{n-x}$ ซึ่งมีค่าเท่ากับการคำนวณในรูปแบบแรก นั่นคือ ตัวแปรสุ่มแบบทวินามที่มีพารามิเตอร์ n และ p เกิดจากตัวแปรสุ่มเบอร์นูลีที่มีพารามิเตอร์ p และเป็นอิสระต่อกันทั้งหมด n ตัว □

3.2 ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variables)

ตัวแปรที่ได้รับความสนใจในทางเศรษฐศาสตร์และการเงินมักอยู่ในรูปของตัวแปรสุ่ม เช่น รายได้ การบริโภค ราคาหลักทรัพย์ อัตราผลตอบแทน เป็นต้น จุดเด่นของตัวแปรสุ่มเหล่านี้คือ ค่าที่เกิดขึ้นจริง (realization) สามารถเป็นได้ทุกค่า ทำให้เซ็ตของจำนวนจริงที่เป็นไปได้มีลักษณะที่ต่อเนื่อง ทำให้เราเรียกตัวแปรสุ่มกลุ่มนี้ว่า ตัวแปรสุ่มต่อเนื่อง

บทนิยามที่ 3.6. *ตัวแปรสุ่ม X เป็นตัวแปรสุ่มต่อเนื่อง (continuous random variable) และมีการแจกแจงต่อเนื่อง (continuous distribution) ถ้ามีฟังก์ชันที่ไม่เป็นลบ (non-negative function) f นิยามบนเส้นจำนวนจริง โดยที่ ความน่าจะเป็นที่ X จะมีค่าที่เกิดขึ้นจริงในช่วงจำนวนจริง (interval of real numbres) ใดๆ จะมีค่าเท่ากับผลการอินทิเกรตของ f บนช่วงของจำนวนจริงนั้น นั่นคือ ความน่าจะเป็นที่ X จะมีค่าอยู่ในช่วง $[a, b]$*

เท่ากับ

$$Pr\left(a \leqslant X \leqslant b\right) = \int_a^b f\left(x\right) dx \tag{3.5}$$

ส่วนกรณีที่เป็นช่วงที่ไม่มีขอบเขตบน (unbounded above) จะได้ว่า

$$Pr\left(X \geqslant a\right) = \int_a^\infty f\left(x\right) dx \tag{3.6}$$

และกรณีที่เป็นช่วงที่ไม่มีขอบเขตล่าง (unbounded below) จะได้ว่า

$$Pr\left(X \leqslant b\right) = \int_{-\infty}^b f\left(x\right) dx \tag{3.7}$$

ประเด็นสำคัญมากอันหนึ่งที่ซ่อนอยู่ในนิยามข้างบนคือ ฟังก์ชัน f นั้นมีหน้าที่สำคัญในการบรรยายคุณสมบัติของตัวแปรสุ่มต่อเนื่อง ซึ่งก็คือการแจกแจงของตัวแปรสุ่มนั่นเอง นั่นคือ ถ้าเรารู้จักหรือเข้าใจฟังก์ชัน f ก็เหมือนกับว่าเราเข้าใจตัวแปรสุ่มนั้น ฟังก์ชัน f นี้มีชื่อเรียกกันทั่วไปว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function)

บทนิยามที่ 3.7. *ถ้าตัวแปรสุ่ม X มีการแจกแจงต่อเนื่อง (continuous distribution) เราจะเรียกฟังก์ชัน f ในนิยามที่ 3.6 ว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) หรือเขียนย่อสั้นๆ เป็น p.d.f. นอกจากนี้ เราเรียกส่วนปิดคลุม (closure) ของเซ็ต $\{x : f\left(x\right) > 0\}$ ว่า ส่วนค้ำจุน (support) ของ X*

รูปภาพที่ XXX แสดงตัวอย่างของฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ของตัวแปรสุ่ม พร้อมทั้งแสดงความสัมพันธ์ระหว่างความน่าจะเป็นและพื้นที่ใต้เส้นฟังก์ชันความหนาแน่นของความน่าจะเป็น โดยจะเห็นได้ว่า ความน่าจะเป็นที่ X จะมีค่าอยู่ในช่วง $[a, b]$ มีค่าเท่ากับพื้นที่ที่แรเงาในรูป ซึ่งสามารถแทนได้ด้วยผลการอินทิเกรตฟังก์ชันความหนาแน่นของความน่าจะเป็นระหว่างช่วง $[a, b]$ ดังนั้น จากหลักการของการอินทิเกรต เราสามารถสรุปได้ว่า ความน่าจะเป็นของค่าจำนวนจริงค่าใดค่าหนึ่งมีค่าเท่ากับศูนย์หากฟังก์ชันเป็นแบบต่อเนื่อง เพราะพื้นที่ใต้กราฟของฟังก์ชันต่อเนื่อง ณ จุดใดจุดหนึ่งมีค่าเท่ากับศูนย์เสมอ นั่นคือ $Pr\left(a \leqslant X \leqslant a\right) = Pr\left(X = a\right) = \int_a^a f\left(x\right) dx = 0$ ซึ่งเป็นไปตามหลักการอินทิเกรตแบบรีมันน์ แต่ไม่ได้หมายความว่า $X = a$ เป็นไปไม่ได้ ดังที่ได้อภิปรายมาแล้วในหน้าที่ 13 อย่างไรก็ตาม ความน่าจะเป็นระหว่างช่วง $[a - \epsilon, a + \epsilon]$ สำหรับ $\epsilon > 0$ ใดๆ ย่อมมีค่ามากกว่าศูนย์เสมอ นั่นคือ $Pr\left(a - \epsilon \leqslant X \leqslant a + \epsilon\right) = \int_{a-\epsilon}^{a+\epsilon} f\left(x\right) dx \approx 2\epsilon f\left(a\right) > 0$ ยิ่งไปกว่านั้น หลักการอินทิเกรตแบบรีมันน์ยังบอกอีกว่า การเปลี่ยนแปลงของค่าฟังก์ชัน f หรือปริพัทธ์ (integrand) ณ จำนวนจุดที่จำกัด (finite points) จะไม่มีผลต่อการอินทิเกรต ดังนั้น เราจึงสามารถสรุปได้ว่า การเปลี่ยนแปลงค่าฟังก์ชันความหนาแน่นของความ

น่าจะเป็น (probability density function) ของตัวแปรสุ่มเพียงบางจุดในระหว่างช่วง $[a, b]$ จะไม่ส่งผลต่อการคำนวณความน่าจะเป็นระหว่างช่วงนั้น นั่นหมายความว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ที่เกิดจากการเปลี่ยนค่าดังกล่าวก็เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวแปรสุ่มนั้นเช่นเดียวกัน เพราะสามารถบรรยายคุณสมบัติของตัวแปรสุ่มนั้นได้เช่นเดียวกัน โดยสรุป ตัวแปรสุ่มอันหนึ่งสามารถมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ได้มากมาย โดยทั่วไปเรามักเรียกคุณสมบัตินี้ว่า การไม่มีความเป็นหนึ่งเดียวของฟังก์ชันความหนาแน่นของความน่าจะเป็น (non-uniqueness of p.d.f.) อย่างไรก็ตาม ถึงแม้ว่าจะมีฟังก์ชันจำนวนมากมายที่แทนตัวแปรสุ่มตัวหนึ่งได้ แต่เรามักจะเลือกใช้อันที่เป็นฟังก์ชันต่อเนื่อง (continuous function) เพื่อให้สะดวกต่อการวิเคราะห์ ดังแสดงในรูปภาพที่ XXX

ADD FIGURE OF DENSITY WITH PROB AS SHADED REGION

ตัวอย่างที่ 3.6. พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นดังต่อไปนี้

$$f(x) = \begin{cases} \frac{x}{2}, \text{ ถ้า } x \in [0, 2] \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

ความน่าจะเป็นที่ค่าที่เกิดขึ้นจริงจะอยู่ในช่วง $[1, 3]$ เท่ากับ

$$Pr(1 \leqslant X \leqslant 3) = \int_1^2 \frac{x}{2} dx + \int_2^3 0 dx = \left. \frac{x^2}{4} \right|_1^2 = \frac{3}{4}$$

□

ในทางปฏิบัติ เรามักจะอ้างถึงคุณสมบัติที่สำคัญสองประการของฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ดังต่อไปนี้

ทฤษฎีบทที่ 3.3. *ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) f ใดๆ จะต้องสอดคล้องกับเงื่อนไขสองข้อต่อไปนี้*

$$f(x) \geqslant 0, \text{ สำหรับทุกค่าจำนวนจริง } x, \tag{3.8}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{3.9}$$

ข้อสังเกตหนึ่งจากทฤษบทที่ 3.3 คือ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ไม่มีขอบเขตบน ซึ่งแตกต่างจากความน่าจะเป็นที่มีทั้งขอบเขตล่างที่ต้องมีค่ามากกว่าศูนย์และขอบเขตบนที่ต้องมีค่าไม่เกินหนึ่ง ความแตกต่างส่วนนี้ตอกย้ำว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ไม่ใช่ความน่าจะเป็นทำให้ไม่จำเป็นต้องมีค่าน้อยกว่าหนึ่ง

ประโยชน์อย่างหนึ่งของทฤษฎีบทที่ 3.3 คือการหาค่าคงที่มาตรฐาน (normalizing constant) สำหรับฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) กล่าวคือ เราอาจจะทราบความสัมพันธ์ระหว่างฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) และค่าที่เกิดขึ้นจริงแต่ละค่าของตัวแปรสุ่ม X ซึ่งแทนด้วย x แต่ยังไม่สามารถกำหนดรูปแบบฟังก์ชันได้ทั้งหมด โดยมากมักจะอยู่ในรูปที่มีค่าคงที่ที่ยังไม่ทราบค่า ซึ่งมักแทนด้วย c เช่น $f(x) = cx$ สำหรับ $1 \leqslant x \leqslant 10$ และ $f(x) = 0$ สำหรับกรณีอื่น เป็นต้น ในกรณีแบบนี้ เราสามารถประยุกต์ใช้สมการที่ 3.8 และ 3.9 เพื่อกำหนดค่าคงที่ c ที่ทำให้ฟังก์ชัน f เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)

ตัวอย่างที่ 3.7 (การแจกแจงเอกรูป (uniform distribution))**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นดังต่อไปนี้

$$f(x) = \begin{cases} c, \text{ ถ้า } x \in [a,b] \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

คำถามก็คือ ค่าคงที่มาตรฐาน (normalizing constant) c ต้องมีค่าเท่าใดฟังก์ชัน f จึงจะเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)? ก่อนอื่นสมการที่ 3.8 ส่งผลให้ค่าคงที่ c จะต้องไม่น้อยกว่าศูนย์ ดังนั้น $c \geqslant 0$ ส่วนสมการที่ 3.9 มีผลทำให้

$$\int_a^b c dx = 1 \Rightarrow c(b-a) = 1 \Rightarrow c = \frac{1}{b-a}$$

โดยสรุป ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่มที่มีการแจกแจงเอกรูป (uniform distribution) คือ

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ ถ้า } x \in [a,b] \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

□

ตัวอย่างที่ 3.8. พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นแบบเลขชี้กำลัง (exponential) ดังต่อไปนี้

$$f(x) = \begin{cases} ce^{-2x}, \text{ ถ้า } x > 0 \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

ค่าคงที่มาตรฐาน (normalizing constant) สำหรับกรณีหาได้จาก

$$\int_0^\infty ce^{-2x} dx = 1 \Rightarrow -\frac{c}{2} e^{-2x} \Big|_0^\infty = \frac{c}{2} = 1 \Rightarrow c = 2$$

โดยสรุป ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่มที่มีการแจกแจงแบบเลขชี้กำลัง (exponential) คือ

$$f(x) = \begin{cases} ce^{-2x}, \text{ ถ้า } x > 0 \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases} \tag{3.10}$$

□

ตัวอย่างที่ 3.9 (การแจกแจงแบบปกติ (normal distribution))**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นแบบปกติ (normal) ดังต่อไปนี้

$$f(x) = ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

โดยที่ μ คือค่าจำนวนจริงที่แทนค่าความคาดหวัง (expectation) และ σ คือค่าจำนวนจริงที่เป็นบวกที่แทนค่าความเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งจะอภิปรายอย่างละเอียดในบทที่ XXX โดยหลักการ เราสามารถหาค่าคงที่มาตรฐาน (normalizing constant) สำหรับการแจกแจงแบบปกติ (normal distribution) ได้โดยใช้สมการที่ 3.9 ดังต่อไปนี้

$$1 = \int_{-\infty}^{\infty} ce^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = c\sigma \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2} d\left(\frac{x-\mu}{\sigma}\right) = c\sigma\sqrt{2\pi} \Rightarrow c = \frac{1}{\sqrt{2\pi\sigma^2}}$$

โดยในที่นี้ เราได้ประยุกต์ใช้เทคนิคการใช้พิกัดเชิงขั้ว (polar coordinates) เพื่อหาแสดงว่า $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$ ซึ่งมีขั้นตอนดังต่อไปนี้ เริ่มด้วยการกำหนดให้

$$I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

ดังนั้น

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{y^2+z^2}{2}} dzdy$$

ขั้นตอนต่อไปคือการแปลง (y, z) ให้อยู่ในรูปของพิกัดเชิงขั้ว (polar coordinates) โดยใช้

$$y = r\cos\theta, z = r\sin\theta$$

ซึ่งมีผลทำให้

$$y^2 + z^2 = r^2, dzdy = rdrd\theta$$

ADD FIGURE OF POLAR TRANSFORMATION ดังนั้น

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{y^2+z^2}{2}} dzdy = \int_0^{2\theta}\int_0^{\infty} e^{-\frac{r^2}{2}} rdrd\theta \\ &= \int_0^{2\theta}\int_0^{\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) d\theta = \int_0^{2\theta} -e^{-\frac{r^2}{2}}\Big|_0^{\infty} d\theta \\ &= \int_0^{2\theta} d\theta = 2\pi \end{aligned}$$

ซึ่งหมายความว่า

$$I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$$

□

3.3 ฟังก์ชันความน่าจะเป็นสะสม (The Cumulative Distribution Function)

ที่ผ่านมา เราอธิบายการแจกแจงของตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่องด้วยฟังก์ชันที่แตกต่างกัน ซึ่งในทางคณิตศาสตร์แล้วไม่ค่อยสะดวกนัก ดังนั้น จึงมีความจำเป็นที่จะต้องหาฟังก์ชันที่สามารถอธิบายการแจกแจงของตัวแปรสุ่มทั้งสองประเภทให้ได้ โดยเรียกฟังก์ชันอันใหม่นี้ว่า ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ซึ่งมีคุณสมบัติที่ตัวแปรสุ่มตัวใดตัวหนึ่งจะมีฟังก์ชันความน่าจะเป็นสะสมเพียงอันเดียว (uniqueness) นอกจากนี้ ในทางเทคนิคฟังก์ชันความน่าจะเป็นสะสมเป็นจุดเชื่อมระหว่างทฤษฎีความน่าจะเป็นพื้นฐานและทฤษฎีมาตรวัดความน่าจะเป็น (probability measure theory) ซึ่งใช้ฟังก์ชันความน่าจะเป็นสะสมเป็นมาตรวัด (measure) สำหรับการอินทิเกรต ผู้อ่านสามารถศึกษาเพิ่มเติมได้ใน Billingsley (2008)

บทนิยามที่ 3.8 (ฟังก์ชันความน่าจะเป็นสะสม (Cumulative Distribution Function: C.D.F.))**.** *ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) หรือฟังก์ชันความน่าจะเป็น (distribution function) ของตัวแปรสุ่ม X ซึ่งเรียกสั้นๆ ว่า C.D.F. ของ X คือฟังก์ชัน*

$$F(x) = Pr(X \leqslant x), \text{ สำหรับ } -\infty < x < \infty \tag{3.11}$$

สิ่งที่สำคัญก็คือ ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ที่นิยามนี้ใช้ได้กับทั้งตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง

ตัวอย่างที่ 3.10. ตัวอย่างนี้แสดงให้เห็นถึงวิธีการแปลงฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ให้เป็นฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) โดยใช้ตัวแปรสุ่ม X ที่มีการแจกแจงดังต่อไปนี้

$$f(x) = \begin{cases} 2e^{-2x}, \text{ ถ้า } x > 0 \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) หรือ C.D.F. ของตัวแปรสุ่ม X คือฟังก์ชัน

$$F(x) = Pr(X \leqslant x) = \begin{cases} \int_0^x 2e^{-2y}dy = 1 - e^{-2x}, \text{ สำหรับ } x > 0 \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

สังเกตว่า เพื่อความสะดวกจึงเปลี่ยนตัวแปรสำหรับการอินทิเกรตจาก x เป็น y เพราะฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. นิยามโดยใช้ x ไปแล้ว

DRAW GRAPHS FOR PDF AND CDF HERE □

ตัวอย่างต่อไปนี้แสดงฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) หรือ C.D.F. ของตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable)

ตัวอย่างที่ 3.11. พิจารณาตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) X ที่มีพารามิเตอร์ p ดังนั้น $Pr(X = 0) = 1 - p$ และ $Pr(X = 1) = p$ ในกรณีนี้ เราสามารถเขียนฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. ได้เป็น

$$F(x) = \begin{cases} 0, \text{ สำหรับ } x < 0 \\ 1 - p, \text{ สำหรับ } 0 \leqslant x < 1 \\ 1, \text{ สำหรับ } x \geqslant 1 \end{cases}$$

DRAW GRAPH FOR CDF HERE □

คุณสมบัติที่ 3.1. *ฟังก์ชัน $F(x)$ เป็นฟังก์ชันไม่ลดลง (decreasing function) เมื่อค่า x เพิ่มขึ้น ซึ่งหมายความว่า ถ้า $x_1 < x_2$ แล้ว $F(x_1) \leqslant F(x_2)$*

การพิสูจน์. ถ้า $x_1 < x_2$ แล้ว เหตุการณ์ $\{X \leqslant x_1\}$ จะเป็นสับเซ็ตของ $\{X \leqslant x_2\}$ ดังนั้น $Pr(X \leqslant x_1) \leqslant Pr(X \leqslant x_2)$ ซึ่งหมายความว่า $F(x_1) \leqslant F(x_2)$ ■

คุณสมบัติที่ 3.2. $\lim_{x\to-\infty} F(x) = 0$ *และ* $\lim_{x\to\infty} F(x) = 1$

นอกจากนี้ ฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. ไม่จำเป็นต้องต่อเนื่อง (continuous) โดยสามารถมีการกระโดด (jump) ได้ตราบเท่าที่จำนวนครั้งในการกระโดดสามารถนับได้ (countable) ดังแสดงตัวอย่างในรูปที่ XXX ซึ่งเป็นตัวอย่างของฟังก์ชันความน่าจะเป็นสะสมตัวแปรสุ่มที่มีการแจกแจงผสม (mixed distribution) อย่างไรก็ตาม ฟังก์ชันความน่าจะเป็นสะสมจำเป็นจะต้องมีคุณสมบัติความต่อเนื่องจากด้านขวา (continuous from the right) ซึ่งนิยามโดยใช้การลู่เข้าจากด้านขวา (limit from the right)

$$F\left(x^{+}\right)=\lim_{\substack{y\to x\\ y>x}}F\left(y\right) \tag{3.12}$$

ซึ่งหมายถึงค่าลิมิตที่ลู่มาจากค่าที่มากกว่าค่าที่พิจารณา ในทำนองเดียวกัน การลู่เข้าจากด้านซ้าย (limit from the left) คือ

$$F\left(x^{-}\right)=\lim_{\substack{y\to x\\ y<x}}F\left(y\right) \tag{3.13}$$

นอกจากนี้ เราสามารถนิยามความต่อเนื่อง (continuity) ของฟังก์ชัน ณ จุด x ใดๆ โดยใช้การลู่เข้าจากด้านขวาและการลู่เข้าจากด้านซ้ายได้ว่า ถ้า ฟังก์ชัน $F\left(x\right)$ ต่อเนื่อง ณ จุด x แล้ว $F\left(x^{+}\right)=F\left(x^{-}\right)=F\left(x\right)$

คุณสมบัติที่ 3.3. *ฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. มีความต่อเนื่องจากด้านขวา (continuous from the right) เสมอ นั่นคือ $F\left(x^{+}\right)=F\left(x\right)$ สำหรับทุกจุด x นอกจากนี้ ณ จุดที่มีการกระโดด (jump) จะพบว่า $F\left(x\right)>F\left(x^{-}\right)$*

การพิสูจน์. กำหนดให้ $y_1>y_2>\ldots$ คืออนุกรมของจำนวนจริงที่มีค่าลดลงโดยที่ $\lim_{n\to\infty}y_n=x$ ผลที่ตามมาก็คือ เหตุการณ์ $\{X\leqslant x\}=\bigcap_{i=1}^{\infty}\{X\leqslant y_i\}$ และ $\{X\leqslant y_i\}\subset\{X\leqslant y_j\}$ สำหรับ $j\leqslant i$ ดังนั้น

$$F\left(x\right)=Pr\left(X\leqslant x\right)=\lim_{i\to\infty}Pr\left(X\leqslant y_i\right)=F\left(x^{+}\right)$$

■

ตัวอย่างต่อไปนี้แสดงวิธีการคำนวณความน่าจะเป็นจากฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง

ตัวอย่างที่ 3.12. พิจารณาตัวแปรสุ่มต่อเนื่อง X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ดังนี้

$$f\left(x\right)=\begin{cases}0,\ \text{สำหรับ } x\leqslant 0\\ \frac{1}{(1+x)^2},\ \text{สำหรับ } x>0\end{cases}$$

ซึ่งมีฟังก์ชันแจกแจงสะสมเป็น

$$F\left(x\right)=\begin{cases}0,\ \text{สำหรับ } x\leqslant 0\\ \int_0^x\frac{1}{(1+y)^2}=1-\frac{1}{1+x},\ \text{สำหรับ } x>0\end{cases}$$

ในขณะเดียวกัน เราสามารถคำนวณความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าอยู่ระหว่าง $(1, 4]$ โดยใช้ฟังก์ชันความน่าจะเป็นสะสมดังต่อไปนี้

$$Pr\left(1 < X \leqslant 4\right) = Pr\left(X \leqslant 4\right) - Pr\left(X \leqslant 1\right) = F\left(4\right) - F\left(1\right) = \frac{4}{5} - \frac{1}{2} = \frac{3}{10}$$

ดูรูปที่ XXX ประกอบ

ADD FIGURE WITH AREA SHADED BETWEEN 1 TO 4 □

ทฤษฎีบทที่ 3.4. *สำหรับค่าจำนวนจริง x ใดๆ*

$$Pr\left(X > x\right) = 1 - F\left(x\right) \tag{3.14}$$

การพิสูจน์. เริ่มจากการที่เหตุการณ์ $\{X > x\}$ และ $\{X \leqslant x\}$ ไม่มีส่วนร่วมกัน (disjoint) และยูเนียนของทั้งสองเหตุการณ์ครอบคลุมปริภูมิตัวอย่าง (sample space) นั่นคือทั้งสองเหตุการณ์นี้เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง ดังนั้น

$$Pr\left(X \leqslant x\right) + Pr\left(X > x\right) = 1 \Rightarrow Pr\left(X > x\right) = 1 - F\left(x\right)$$

■

ทฤษฎีบทต่อไปนี้เป็นบทสรุปอย่างเป็นทางการของเทคนิคการคำนวณความน่าจะเป็นจากฟังก์ชันความน่าจะเป็นสะสมที่ใช้ในตัวอย่างที่ 3.12

ทฤษฎีบทที่ 3.5. *สำหรับค่าจำนวนจริง x_1 และ x_2 ที่ $x_1 < x_2$*

$$Pr\left(x_1 < X \leqslant x_2\right) = F\left(x_2\right) - F\left(x_1\right) \tag{3.15}$$

การพิสูจน์. กำหนดให้ $A = \{x_1 < X \leqslant x_2\}$, $B = \{x_1 \leqslant X\}$ และ $C = \{X \leqslant x_2\}$ เมื่อพิจารณาอย่างรอบคอบจะเห็นได้ว่า เหตุการณ์ A และ B เป็นการแบ่งส่วน (partition) ของ C นั่นคือ $A \cap B = \varnothing$ และ $A \cup B = C$ ดังนั้น

$$Pr\left(x_1 < X \leqslant x_2\right) + Pr\left(X \leqslant x_1\right) = Pr\left(X \leqslant x_2\right) \Rightarrow Pr\left(x_1 < X \leqslant x_2\right) = F\left(x_2\right) - F\left(x_1\right)$$

ซึ่งเป็นผลมาจากการแทนค่า $Pr\left(X \leqslant x_n\right) = F\left(x_n\right)$ ■

ทฤษฎีบทต่อไปนี้แสดงความสัมพันธ์ระหว่างความน่าจะเป็นและฟังก์ชันความน่าจะเป็นสะสมในกรณีที่อาจจะมีการกระโดด (jump)

ทฤษฎีบทที่ 3.6. *สำหรับค่าจำนวนจริง x ใดๆ*

$$Pr\left(X < x\right) = F\left(x^{-}\right) \tag{3.16}$$

การพิสูจน์. ดังที่อภิปรายมาก่อนหน้านี้ สำหรับกรณีที่ไม่มีการกระโดด ความน่าจะเป็นของค่าจำนวนจริงค่าใดค่าหนึ่งมีค่าเท่ากับศูนย์ $PrPr\left(X = x\right) = 0$ ดังนั้น $Pr\left(X < x\right) = Pr\left(X \leqslant x\right) = F\left(x\right)$ ในขณะเดียวกัน ในกรณีนี้ $F\left(x\right) = F\left(x^{-}\right) = F\left(x^{+}\right)$

ส่วนกรณีที่มีการกระโดด (jump) เราสามารถพิสูจน์ความสัมพันธ์ได้โดยพิจารณาอนุกรมที่มีค่าเพิ่มขึ้น $y_1 < y_2 < \ldots$ โดยที่ $\lim_{i\to\infty} y_i = x$ ซึ่งช่วยให้เราสามารถแทนเหตุการณ์ $\{X < x\}$ ในรูปของการยูเนียนของเหตุการณ์ทีเป็นผลของค่า y_i ได้เป็น

$$\{X < x\} = \bigcup_{i=1}^{\infty} \{X < y_i\}$$

นอกจากนี้ เรายังทราบอีกว่า $\{X < y_i\} \subset \{X < y_j\}$ สำหรับ $y_i < y_j$ ดังนั้น เราสามารถแสดงได้ว่า

$$\begin{aligned} Pr\left(X \leqslant x\right) = Pr\left(\bigcup_{i=1}^{\infty} \{X < y_i\}\right) &= \lim_{i\to\infty} Pr\left(X \leqslant y_i\right) \\ = \lim_{i\to\infty} F\left(y_i\right) = \lim_{\substack{y\to x \\ y<x}} Pr\left(X \leqslant y\right) &\equiv F\left(x^{-}\right) \end{aligned}$$

■

ถึงแม้ว่า เราจะสรุปไว้ก่อนหน้านี้ว่า ความน่าจะเป็นของค่าจำนวนจริงค่าใดค่าหนึ่งมีค่าเท่ากับศูนย์ $Pr\left(X = x\right) = 0$ แต่ข้อความนี้เป็นจริงในกรณีที่ไม่มีการกระโดด ณ จุด x เท่านั้น หากมีการกระโดด (jump) ความน่าจะเป็นของค่าจำนวนจริง x จะไม่เท่ากับศูนย์ทันที ดังนั้น เพื่อความสะดวก จึงสรุปผลที่ครอบคลุมทั้งสองกรณีไว้ในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.7. *สำหรับค่าจำนวนจริง x ใดๆ*

$$Pr\left(X = x\right) = F\left(x\right) - F\left(x^{-}\right) \tag{3.17}$$

การพิสูจน์. เมื่อแทนค่า $Pr\left(X < x\right) = F\left(x^{-}\right)$ จากทฤษฎีบทที่ 3.6 เข้าไปในความสัมพันธ์ที่ว่า $Pr\left(X = x\right) = Pr\left(X \leqslant x\right) - Pr\left(X < x\right)$ จะได้ว่า

$$Pr\left(X = x\right) = F\left(x\right) - F\left(x^{-}\right)$$

■

ข้อสังเกตอันหนึ่งจากทฤษฎีบทที่ 3.7 คือ ถ้าตัวแปรสุ่มไม่มีการกระโดด ณ จุด x เราจะได้ว่า $F(x) = F(x^-)$ ซึ่งหมายความว่า $Pr(X = x)$ ในทางกลับกัน ถ้ามีการกระโดด ณ จุด x เราจะพบว่า $F(x) > F(x^-)$ เพราะความน่าจะเป็น ณ จุดนั้นจะต้องมีค่าเป็นบวก $Pr(X = x) > 0$

เมื่อพิจารณาจากฟังก์ชันความน่าจะเป็นสะสมหรือ C.D.F. จะเห็นได้ว่า ตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) คือตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็นสะสมที่มีการกระโดดเป็นจำนวนที่จำกัด (finite jumps) หรือนับได้ (countably many jumps) โดยในช่วง (a, b) ที่ไม่มีการกระโดด ($Pr(a < Xb) = 0$) ฟังก์ชันความน่าจะเป็นสะสมหรือ C.D.F. จะมีค่าคงที่และเป็นเส้นขนานตลอดช่วง ดังแสดงในรูปที่ XXX ข้างล่าง นอกจากนี้ขนาดของการกระโดดแต่ละครั้งจะเท่ากับฟังก์ชันความน่าจะเป็น ณ จุดนั้น เช่น ถ้าฟังก์ชันความน่าจะเป็น ณ จุด x มีค่าเท่ากับ $f(x_i)$ ขนาดของการกระโดดของฟังก์ชันความน่าจะเป็นสะสมหรือ C.D.F. ณ จุด x มีค่าเท่ากับ $f(x_i)$

ADD CDF FOR A DISCRETE WITH FLAT LINES

ในส่วนของตัวแปรสุ่มต่อเนื่อง (continuous random variable) เราสามารถประยุกต์ใช้หลักการพื้นฐานของแคลคูลัส (fundamental theorem of calculus) เพื่อกำหนดความสัมพันธ์ระหว่างฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) และฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ได้ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.8. *กำหนดให้ตัวแปรสุ่ม X มีการกระจายต่อเนื่อง โดยมี $f(x)$ และ $F(x)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) และฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ดังนั้น*

1. *ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function)*

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \tag{3.18}$$

และมีความต่อเนื่อง (continuous) ทุกๆ ค่าจำนวนจริง x

2. *ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function)*

$$f(x) = \frac{dF(x)}{dx} \tag{3.19}$$

สำหรับทุกๆ ค่าจำนวนจริง x ที่ฟังก์ชัน $f(x)$ มีความต่อเนื่อง (continuous)

ข้อสังเกตที่น่าสนใจอันหนึ่งจากทฤษฎีบทนี้คือ ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ของตัวแปรสุ่มต่อเนื่องนั้นมีความต่อเนื่องที่ทุกค่าจำนวนจริง ในขณะที่ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) อาจจะไม่มีความต่อเนื่องที่บางค่าได้ และผลที่ตามมาคือ สมการ

ที่ 3.19 จะไม่สามารถใช้ได้ที่จุดนั้น ดังจะเห็นได้จากตัวอย่างที่ 3.10 ซึ่งมีจุดหักงอ (kink) ณ จุด $x = 0$ ทำให้ ณ จุดนี้เราไม่สามารถหาค่าอนุพันธ์ (derivative) ของฟังก์ชันความน่าจะเป็นสะสม $F(x)$ ได้ (ดูรูปที่ XXX ประกอบ) นอกจากนี้ ในตัวอย่างนี้ ฟังก์ชันความหนาแน่นของความน่าจะเป็น $f(x)$ มีการกระโดด (jump) ณ จุดเดียวกันนี้

ตัวอย่างต่อไปนี้แสดงวิธีการคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็นจากฟังก์ชันความน่าจะเป็นสะสมโดยใช้สมการที่ 3.19

ตัวอย่างที่ 3.13. พิจารณาตัวแปรสุ่ม X ที่มีการแจกแจงแบบเลขชี้กำลัง (exponential distribution) ด้วยค่าพารามิเตอร์ $\lambda > 0$ ดังนี้

$$F(x) = \begin{cases} 0, \text{ สำหรับ } x < 0 \\ 1 - e^{-\lambda x}, \text{ สำหรับ } x \geqslant 0 \end{cases}$$

ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่ม X ที่มีการแจกแจงแบบเลขชี้กำลัง คือ

$$f(x) = \begin{cases} 0, \text{ สำหรับ } x < 0 \\ \frac{d\left[1-e^{-\lambda x}\right]}{dx} = \lambda e^{-\lambda x}, \text{ สำหรับ } x \geqslant 0 \end{cases}$$

□

ค่าสถิติแบบหนึ่งที่ได้รับความนิยมในการนำเสนอการแจกแจง (distribution) ของตัวแปรสุ่มคือ ควอนไทล์ (quantile) ซึ่งเป็นการบอกค่าของตัวแปรสุ่มโดยอ้างอิงกับอับดับของค่าตัวแปรสุ่มที่เรียงจากน้อยไปหามาก โดยบางครั้งอาจจะอ้างอิงกับเปอร์เซนต์ของลำดับและเรียกว่า เปอร์เซ็นไทล์ (percentile)

บทนิยามที่ 3.9 (ควอนไทล์ (quantile) และ เปอร์เซ็นไทล์ (percentile))**.** *กำหนดให้ $F(x)$ แทนฟังก์ชันความน่าจะเป็นสะสมของตัวแปรสุ่ม X และ $F^{-1}(p)$ แทนค่าที่เล็กที่สุด (smallest value) ที่ $F(x) \geqslant p$ สำหรับค่า $p \in (0,1)$ โดยเรียก $F^{-1}(p)$ ว่า ควอนไทล์ที่ p (p quantile) หรือ $100p$ เปอร์เซ็นไทล์ ($100p$ percentile) ของ X และเรียกฟังก์ชันส่วนกลับ (inverse function) F^{-1} ว่า ฟังก์ชันควอนไทล์ (quantile function)*

ตัวอย่างที่มักจะเห็นในชีวิตประจำวันคือ การหาคะแนนมาตรฐาน (standardized score) ซึ่งเป็นการแปลงคะแนนผลการทดสอบให้อยู่ในรูปของเปอร์เซ็นไทล์ (percentile) โดยการเปรียบเทียบกับผลการทดสอบของผู้เข้ารับการทดสอบจำนวนมาก โดยเรียงคะแนนจากน้อยไปหามาก ยกตัวอย่างเช่น เด็กชายสมชายสอบได้คะแนน 75 จากคะแนนเต็ม 100 แต่เมื่อนำมาจัดเรียงตามลำดับคะแนนแล้วปรากฏว่าคะแนนของเขาอยู่ที่เปอร์เซ็นต์ไทล์ที่ 50 หรือควอนไทล์ที่ 0.50 เพราะครึ่งหนึ่งของผู้เข้ารับการทดสอบมีคะแนนมากกว่า 75 คะแนน นั่นคือ $F(X = 75) = 0.50$ เป็นต้น ในขณะเดียวกัน หากเราสนใจว่าผู้เข้ารับการทดสอบจะต้องได้คะแนนเท่าใดจึงจะถูกจัดว่ามีคะแนนที่เปอร์เซ็นต์ไทล์ที่ 75 หรือควอนไทล์ที่ 0.75 ยกตัวอย่างเช่น มีผู้เข้าทดสอบร้อยละ 25 ที่

มีคะแนนมากกว่า 85 คะแนน ซึ่งหมายความว่าระดับคะแนนที่เปอร์เซ็นต์ไทล์ที่ 75 หรือควอนไทล์ที่ 0.75 มีค่าเท่ากับ 85 คะแนน นั่นคือ $F^{-1}(0.75) = 85$ เป็นต้น

ในทางคณิตศาสตร์ ฟังก์ชันควอนไทล์ (quantile function) มีสัญลักษณ์คล้ายกับฟังก์ชันส่วนกลับ (inverse function) ของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ซึ่งเป็นจริงในกรณีของตัวแปรสุ่มที่มีการแจกแจงสะสมต่อเนื่อง (continuous distribution) และเป็นแบบหนึ่งต่อหนึ่ง (one-to-one) โดยในกรณีนี้ฟังก์ชันส่วนกลับ (inverse function) ของ F จะมีอยู่ (exist) และมีค่าเท่ากับฟังก์ชันควอนไทล์ (quantile function) ส่วนกรณีที่ฟังก์ชันความน่าจะเป็นสะสมไม่ได้เป็นแบบหนึ่งต่อหนึ่ง เราอาจจะไม่สามารถหาฟังก์ชันส่วนกลับ (inverse function) ของ F ได้แต่เรายังสามารถหาค่าควอนไทล์ได้ตามที่กำหนดในนิยามที่ 3.9

ตัวอย่างที่ 3.14 (มูลค่าที่เสี่ยง (Value at Risk))**.** หลักการหามูลค่าที่เสี่ยง (Value at Risk) หรือ VaR ถูกสร้างขึ้นเพื่อวัดบอกถึงระดับความเสียหายที่อาจจะขึ้นจากการลงทุนในกรณีที่แย่ที่สุด (worst case) แน่นอนว่า กรณีที่แย่ที่สุดจริงๆ คือ การสูญเสียมูลค่าการลงทุนทั้งหมด ซึ่งหมายถึงการกำหนดให้ $p = 1$ (ครอบคลุมทุกๆ กรณี) แต่อาจจะไม่ใช่จุดที่น่าสนใจที่สุด ดังนั้น ในทางปฏิบัติเรามักจะสนใจกรณีที่แย่สุดๆ นั่นคือ ค่าความน่าจะเป็นมีค่าใกล้เคียงกับหนึ่งมาก แต่ก็ไม่ใช่หนึ่ง ยกตัวอย่างเช่น เราอาจจะกำหนดให้ระดับความเชื่อมั่น $p = 0.99$ ซึ่งหมายความว่า เราอยากทราบว่ามูลค่าของหลักทรัพย์จะลดลงมากที่สุดเท่าไร หากเราพิจารณาครอบคลุมร้อยละ 99 ของเหตุการณ์ที่เป็นไปได้ทั้งหมด เป็นต้น

พิจารณาตัวแปรสุ่ม X ที่บ่งบอกถึงการเปลี่ยนแปลงของมูลค่ากลุ่มหลักทรัพย์ (portfolio value) ในช่วงเวลาหนึ่งเดือนโดยมีหน่วยเป็นล้านบาท ซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ดังแสดงในรูปที่ XXX คำถามที่สนใจในที่นี้คือ มูลค่ากลุ่มหลักทรัพย์จะลดลงมากที่สุดเท่าใด ภายใต้ระดับความเชื่อมั่นหรือความน่าจะเป็น p ในทางเทคนิค คำถามนี้สามารถเขียนในรูปของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ได้ว่า x จะต้องมีค่าเท่ากับเท่าใดจึงจะทำให้ $p = Pr(X \geqslant x) = F(x)$ ยกตัวอย่างเช่น มูลค่าที่เสี่ยง (Value at Risk) ที่ $p = 0.99$ มีค่าเท่ากับ $F^{-1}(0.99) = -5$ (ค่าควอนไทล์ที่ 0.99) กล่าวคือ มูลค่าของหลักทรัพย์จะลดลงมากที่สุด 5 ล้านบาท หากเราพิจารณาครอบคลุมร้อยละ 99 ของเหตุการณ์ที่เป็นไปได้ทั้งหมด

โดยสรุป มูลค่าที่เสี่ยง (Value at Risk) คือฟังก์ชันควอนไทล์ของตัวแปรสุ่มนั่นเอง

ADD FIGURE OF PDF FROM FIG 3.7 BUT DOES NOT NEED TO BE EXACTLY THE SAME □

ข้อสังเกตที่สำคัญอันหนึ่งคือ p เป็นค่าจำนวนจริงระหว่างศูนย์กับหนึ่ง ซึ่งหมายความว่าเราสามารถกำหนดค่าควอนไทล์ (quantile) ที่ค่าเท่าใดก็ได้ แต่ในทางปฏิบัติ เรามักให้ความสนใจกับบางค่าเป็นพิเศษ พร้อมทั้งตั้งชื่อให้เป็นพิเศษ เช่น $p = 0.250.500.75$ ซึ่งถูกเรียกว่าควอร์ไทล์ (quartile) ซึ่งหมายถึงค่าควอนไทล์ที่เฉพาะเจาะจง

บทนิยามที่ 3.10. *ค่าควอนไทล์ที่ 0.50 หรือเปอร์เซ็นต์ไทล์ที่ 50 คือค่ามัธยฐาน (median) ค่าควอนไทล์ที่ 0.25 หรือเปอร์เซ็นต์ไทล์ที่ 25 คือค่าควอร์ไทล์ล่าง (lower quartile) และค่าควอนไทล์ที่ 0.75 หรือเปอร์เซ็นต์ไทล์ที่*

75 คือค่าควอร์ไทล์บน (upper quartile)

ตัวอย่างต่อไปนี้แสดงการหาค่ามัธยฐาน (median) ค่าควอร์ไทล์ล่าง (lower quartile) และค่าควอร์ไทล์บน (upper quartile) ซึ่งในทางปฏิบัติสามารถทำได้หลายรูปแบบที่นำไปสู่คำตอบที่แตกต่างกัน

ตัวอย่างที่ 3.15. พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ดังต่อไปนี้

$$F(x) = \begin{cases} 0, \text{ สำหรับ } x < 0, \\ \frac{1}{4}, \text{ สำหรับ } 0 \leqslant x < 4, \\ \frac{1}{2}, \text{ สำหรับ } 4 \leqslant x < 6, \\ \frac{3}{4}, \text{ สำหรับ } 6 \leqslant x < 10, \\ 1, \text{ สำหรับ } x \geqslant 10 \end{cases}$$

ถ้าหากเราใช้นิยามของควอนไทล์ซึ่งใช้ค่าที่เล็กที่สุดในการกำหนดค่า จะได้ว่า ค่ามัธยฐาน (median) มีค่าเท่ากับ 4 ค่าควอร์ไทล์ล่าง (lower quartile) มีค่าเท่ากับ 0 และค่าควอร์ไทล์บน (upper quartile) มีค่าเท่ากับ 6

อย่างไรก็ตาม จะสังเกตได้ว่า $F(x) = \frac{1}{2}$ สำหรับทุกค่า x ที่อยู่ระหว่าง $4 \leqslant x < 6$ ดังนั้น จึงเป็นเรื่องปกติที่จะกำหนดให้ค่า x ใดๆ ที่อยู่ระหว่าง $4 \leqslant x < 6$ เป็นค่ามัธยฐาน ในขณะเดียวกัน ยังมีรูปแบบการคำนวณที่ได้รับความนิยมอย่างมากอีกรูปแบบหนึ่งที่ทำให้ได้ค่าที่ชัดเจนเพียงค่าเดียวคือ การเลือกค่ากลางของช่วงที่มีค่า $F(x) = \frac{1}{2}$ เช่นในกรณีนี้ค่ามัธยฐานจะมีค่าเท่ากับ $\frac{4+6}{2} = 5$ โดยสรุป ในทางปฏิบัติ การกำหนดค่ามัธยฐาน (median) ค่าควอร์ไทล์ล่าง (lower quartile) และค่าควอร์ไทล์บน (upper quartile) มีอยู่หลายวิธีและยังไม่มีข้อสรุปร่วมกันว่าวิธีใดดีที่สุด □

ทำไมจึงต้องสนใจค่าสถิติ เช่น ควอนไทล์ (quantile) หรือค่าคาดหวัง (expectation)?

ผู้อ่านจะต้องตระหนักเสมอว่า สิ่งที่เราต้องการเข้าใจจริงๆ คือคุณสมบัติของตัวแปรสุ่ม ซึ่งบรรจุอยู่ในการแจกแจงของตัวแปรสุ่มนั้น กล่าวคือเราสามารถเข้าใจตัวแปรสุ่มได้ด้วยการทำความเข้าใจการแจกแจงของมัน ดังนั้น หากเราสามารถนำเสนอการแจกแจงของตัวแปรสุ่มโดยตรงได้ย่อมเป็นการดีที่สุด แต่ในทางปฏิบัติ เราอาจจะไม่มีข้อมูลที่เพียงพอ (แต่อาจจะไม่ใช่ปัญหาอีกต่อไปเมื่อเรามี Big Data) ที่จะหาการแจกแจงของตัวแปรสุ่ม ทำให้เราต้องเลือกใช้วิธีอื่น ไม่ว่าจะเป็นการนำเสนอโดยใช้ควอนไทล์ซึ่งอภิปรายในส่วนนี้ หรือการนำเสนอโดยใช้ค่าคาดหวัง (expectation) ดังที่แสดงในบทที่ XXX ซึ่งเป็นการนำเสนอที่สะดวกมากกว่าแบบควอนไทล์ แต่ในขณะเดียวกัน ก็จะสูญเสียสารสนเทศ (information) ที่เกี่ยวข้องกับตัวแปรสุ่มมากกว่าเช่นกัน ดังนั้น นักวิเคราะห์ข้อมูลจะต้องเข้าใจว่า การเลือกใช้ค่าสถิติแต่ละแบบมีข้อดีข้อเสียอย่างไร และตระหนักเสมอว่า สิ่งที่เราต้องการเข้าใจคือคุณสมบัติของตัวแปรสุ่ม ซึ่งสามารถอธิบายได้อย่างสมบูรณ์ด้วยการแจกแจง (distribution) ของตัวแปรสุ่ม เพราะถ้าเราทราบการแจกแจงของตัวแปรสุ่มก็เหมือนกับว่าเราเข้าใจคุณสมบัติของตัวแปรสุ่มนั้น

3.4 การแจกแจงร่วมของสองตัวแปร (Bivariate Distributions)

บทนิยามที่ 3.11. *การแจกแจงร่วม (joint distribution) ของตัวแปรสุ่ม X และ Y คือคือเซ็ตของความน่าจะเป็นทุกอันที่อยู่ในรูปแบบ $Pr\left((X,Y)\in C\right)$ สำหรับเซ็ตของจำนวนจริงสองจำนวน C ทุกเซ็ต โดยที่ $\{(X,Y)\in C\}$ เป็นเหตุการณ์*

บทนิยามที่ 3.12. *ตัวแปรสุ่ม X และ Y มีการแจกแจงร่วมไม่ต่อเนื่อง (discrete joint distribution) ถ้าค่าจำนวนจริงทั้งสองจำนวน (x,y) ที่เป็นไปได้ของ (X,Y) มีจำนวนจำกัด (finite) หรืออย่างมากต้องเป็นอนุกรมอนันต์ที่นับได้ (countable)*

ทฤษฎีบทที่ 3.9. *ถ้า X และ Y ต่างเป็นตัวแปรสุ่มที่มีการแจกแจงไม่ต่อเนื่อง แล้ว (X,Y) มีการแจกแจงร่วมไม่ต่อเนื่อง (discrete joint distribution)*

การพิสูจน์. ถ้าค่าที่เป็นไปได้ของตัวแปรสุ่มแต่ละตัวมีจำนวนจำกัด (finite) แล้วค่าจำนวนจริงทั้งสองจำนวน (x,y) ที่เป็นไปได้ของ (X,Y) จะต้องมีจำนวนจำกัดตามไปด้วย ในทำนองเดียวกัน ถ้าแต่ละตัวมีค่าที่เป็นไปได้เป็นอนันต์แต่นับได้ (countably infinite) แล้วค่าจำนวนจริงทั้งสองจำนวน (x,y) ที่เป็นไปได้ของ (X,Y) จะต้องเป็นอนันต์แต่นับได้ (countably infinite) ■

เช่นเดียวกับกรณีของตัวแปรสุ่มไม่ต่อเนื่อง ฟังก์ชันความน่าจะเป็นร่วม (joint probability function) สามารถนิยามได้จากความน่าจะเป็นของเหตุการณ์ร่วม

บทนิยามที่ 3.13. *ฟังก์ชันความน่าจะเป็นร่วม (joint probability function) ของตัวแปรสุ่ม X และ Y คือฟังก์ชัน f ซึ่งนิยามสำหรับทุกค่า (x,y) ในระนาบ xy โดยที่*

$$f\left(x,y\right)=Pr\left(X=x\text{ และ }Y=y\right) \tag{3.20}$$

คุณสมบัติที่ 3.4. *ฟังก์ชันความน่าจะเป็นร่วม (joint probability function) ของตัวแปรสุ่ม X และ Y มีคุณสมบัติต่อไปนี้*

1. *ถ้า (x,y) เป็นค่าที่เป็นไปไม่ได้สำหรับ (X,Y) แล้ว $f\left(x,y\right)=0$*

2. *สำหรับเซ็ตของจำนวนจริงสองจำนวน C ใดๆ*

$$Pr\left((X,Y)\in C\right)=\sum_{(x,y)\in C}f\left(x,y\right) \tag{3.21}$$

3. ค่าผลรวมของฟังก์ชันความน่าจะเป็นร่วมจากทุกค่า (x, y) ที่เป็นไปได้จะต้องเท่ากับหนึ่ง นั่นคือ

$$\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = 1 \tag{3.22}$$

ตัวอย่างการประยุกต์ใช้ฟังก์ชันความน่าจะเป็นร่วม (joint probability function) ในชีวิตประจำวันคือการนำเสนอข้อมูลในรูปแบบตารางไขว้ (cross tabulation) ซึ่งช่วยให้สามารถสังเกตความสัมพันธ์ระหว่างตัวแปรได้โดยสะดวก ยกตัวอย่างเช่น ตารางที่ XXX ซึ่งแสดงสัดส่วนของเด็กโดยแบ่งตามเพศและระยะเวลาที่ได้รับนมแม่ซึ่งแบ่งเป็นสองกลุ่มคือกลุ่มที่ได้รับนมแม่ไม่เพียงพอ (น้อยกว่า 6 เดือน) และกลุ่มที่ได้รับนมแม่เพียงพอ (อย่างน้อย 6 เดือน) ตัวเลขในแต่ละช่องนั้นแทนฟังก์ชันความน่าจะเป็นร่วมของตัวแปรสุ่มสองตัวคือ ตัวแปรเพศและตัวแปรแทนการได้รับนมแม่อย่างเพียงพอ

ตารางที่ 3.1: สัดส่วนของเด็กอายุระหว่าง ถึง 3 ปี แยกตามเพศและการได้รับนมแม่อย่างเพียงพอ โดยใช้ข้อมูล Multiple Indicator Cluster Survey (MICS4) ปี 2012

	เพศหญิง	เพศชาย
ได้รับนมแม่เพียงพอ	0.244	0.253
ได้รับนมแม่ไม่เพียงพอ	0.266	0.237

เช่นเดียวกันกับการแจกแจงต่อเนื่องของตัวแปรสุ่มตัวเดียว เราสามารถนิยามการแจกแจงต่อเนื่องของตัวแปรสุ่มสองตัวได้โดยใช้ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.)

บทนิยามที่ 3.14. *ตัวแปรสุ่ม X และ Y มีการแจกแจงต่อเนื่องร่วม (continuous joint distribution) ถ้ามีฟังก์ชันที่ไม่เป็นลบ (nonnegative function) f ที่*

$$Pr((X,Y)\in C) = \iint_C f(x,y)\,dxdy \tag{3.23}$$

สำหรับทุกสับเซ็ตของจำนวนจริงสองจำนวน C ใดๆ ที่ผลของการอินทิเกรตสามารถหาค่าได้ โดยเรียกฟังก์ชัน f ว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint probability density function) นอกจากนี้ เราเรียกส่วนปิดคลุม (closure) ของเซ็ต $\{(x,y) : f(x,y) > 0\}$ ว่า ส่วนค้ำจุน (support) ของ (X,Y)

คุณสมบัติที่ 3.5. *ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint probability density function) สอดคล้องกับเงื่อนไขต่อไปนี้*

$$f(x,y) \geqslant 0, \text{ สำหรับ } (x,y)\in\mathbb{R}^2 \tag{3.24}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1 \tag{3.25}$$

ในทางกลับกัน ฟังก์ชัน f ที่สอดคล้องกับสมการทั้งสองเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมของการแจกแจงร่วมอันใดอันหนึ่งอย่างแน่นอน

ทฤษฎีบทที่ 3.10. *การแจกแจงต่อเนื่องร่วม (continuous joint distribution) นิยามบนระนาบ xy ใดๆ สอดคล้องกับเงื่อนไขต่อไปนี้*

1. *ความน่าจะเป็นของจุดใดๆ หรือของอนุกรมอนันต์ของจุดใดๆ ในระนาบ xy มีค่าเท่ากับศูนย์เสมอ*

2. *ถ้า g เป็นฟังก์ชันต่อเนื่อง (continuous function) ของตัวแปรหนึ่งตัวที่นิยามบนช่วง $[a,b]$ แล้ว เซ็ตหรือเหตุการณ์ $\{(x,y) : y = g(x), a < x < b\}$ และ $\{(x,y) : x = g(y), a < y < b\}$ ต่างมีความน่าจะเป็นเท่ากับศูนย์*

ข้อสังเกตที่น่าสนใจอันหนึ่งคือ เงื่อนไขที่สองของทฤษฎีบทที่ 3.10 บอกเป็นนัยว่า ความน่าจะเป็นที่นิยามบนเส้นทั้งแบบตรงและโค้งบนระนาบ xy ต่างมีความน่าจะเป็นเท่ากับศูนย์ ซึ่งเป็นผลมาจากทฤษฎีการอินทิเกรตที่บอกว่า การอินทิเกรตบนช่วงของเส้น (line) ใดๆ มีค่าเท่ากับศูนย์ ผลที่ตามมาที่น่าสนใจคือ การแจกแจงร่วม (joint distribution) ของตัวแปรสุ่มต่อเนื่องสองตัวไม่จำเป็นต้องเป็นแบบต่อเนื่องซึ่งแตกต่างจากกรณีของตัวแปรสุ่มไม่ต่อเนื่อง ยกตัวอย่างเช่น ตัวแปรสุ่มต่อเนื่อง X และ Y โดยที่ $X = Y$ ถึงแม้ว่าทั้งคู่จะมีการแจกแจงต่อเนื่อง แต่การแจกแจงร่วมของทั้งสองตัวแปรลลับเป็นแบบไม่ต่อเนื่อง เพราะความน่าจะเป็นที่ค่าของ (X,Y) จะอยู่บนเซ็ต $\{(x,y) : x = y\}$ มีค่าเท่ากับหนึ่งซึ่งไม่เป็นไปตามทฤษฎีบทที่ 3.10 ดังนั้น เราจึงสรุปได้ว่า (X,Y) มีการแจกแจงร่วมไม่ต่อเนื่อง

ตัวอย่างต่อไปนี้แสดงการคำนวณหาค่าคงที่มาตรฐาน (normalizing constant) สำหรับฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.)

ตัวอย่างที่ 3.16. กำหนดให้ตัวแปรสุ่ม X และ Y มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ดังต่อไปนี้

$$f(x,y) = \begin{cases} cx^2y, & \text{สำหรับ } x^2 \leqslant y \leqslant 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

คำถามก็คือ ค่าคงที่มาตรฐาน (normalizing constant) c ต้องมีค่าเท่าใดเพื่อให้ฟังก์ชัน $f(x,y)$ เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.)? คำตอบหาได้โดยใช้สมการที่ 3.25 ดังต่อไปนี้

กำหนดให้ S คือส่วนค้ำจุน (support) ของ (X,Y) ดังแสดงในรูปที่ XXX ดังนั้น เราสามารถเขียนการอินทิ

เกรตสองชั้นของฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ได้เป็น

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = \iint_S cx^2ydxdy = \int_{-1}^{1} cx^2\left[\int_{x^2}^{1} ydy\right]dx$$
$$= \int_{-1}^{1} cx^2\left[\frac{y^2}{2}\Big|_{x^2}^{1}\right]dx = c\int_{-1}^{1} x^2\left[\frac{1}{2}-\frac{x^4}{2}\right]dx = \frac{4}{21}c$$

เนื่องจากผลการอินทิเกรตสองชั้นของฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) จะต้องมีค่าเท่ากับหนึ่ง (สมการที่ 3.25) ดังนั้น ค่าคงที่มาตรฐาน (normalizing constant) $c = \frac{21}{4}$ นั่นคือ

$$f(x,y) = \begin{cases} \frac{21}{4}x^2y, \text{ สำหรับ } x^2 \leqslant y \leqslant 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ข้อสังเกตในทางคณิตศาสตร์อันหนึ่งคือ ลำดับในการอินทิเกรตไม่มีผลต่อการหาค่าคงที่มาตรฐานในข้อนี้ ทั้งนี้เป็นผลมาจากหลักการพื้นฐานทางแคลคูลัสที่ว่า ลำดับในการอินทิเกรตสองชั้นนั้นไม่มีผลต่อผลลัพธ์ ผู้อ่านสามารถทดสอบได้ด้วยการเริ่มอินทิเกรตจากตัวแปร x แล้วจึงอินทิเกรตตัวแปร y แน่นอนว่าขอบเขตของการอินทิเกรตอาจจะเปลี่ยนไปด้วย แต่ผลลัพธ์ที่ได้จะต้องเหมือนเดิม □

ตัวอย่างต่อไปนี้แสดงวิธีการคำนวณความน่าจะเป็นจากฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) โดยใช้สมการที่ 3.23

ตัวอย่างที่ 3.17. พิจารณาตัวแปรสุ่ม X และ Y ซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) เช่นเดียวกับในตัวอย่างที่ 3.16 คำถามก็คือ ความน่าจะเป็นที่ $X \geqslant Y$ มีค่าเท่ากับเท่าใด เราสามารถหาคำตอบนี้ได้โดยใช้สมการที่ 3.23 ดังต่อไปนี้

$$Pr(X \geqslant Y) = \int_0^1\int_{x^2}^{x} \frac{21}{4}x^2ydydx = \frac{3}{20}$$

โดยในที่นี้ เราจะเริ่มอินทิเกรตจากตัวแปร y ก่อนและมีขอบเขตของการอินทิเกรตดังแสดงในรูปที่ XXX □

ในทางปฏิบัติ เราอาจจะต้องวิเคราะห์ปัญหาที่ประกอบไปด้วยตัวแปรสุ่มทั้งที่เป็นแบบต่อเนื่องและไม่ต่อเนื่อง ซึ่งมีการแจกแจงผสม (mixed distribution)

บทนิยามที่ 3.15. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่มีการแจกแจงไม่ต่อเนื่อง (discrete) และต่อเนื่อง (continuous) ตามลำดับ และถ้ามีฟังก์ชัน $f(x,y)$ ที่นิยามบนระนาบ xy โดยที่เซ็ตจำนวนจริง A และ B ใดๆ*

$$Pr(X \in AY \in B) = \int_B \sum_{x\in A} f(x,y)\,dy \tag{3.26}$$

ในกรณีที่ผลการอินทิเกรตมีค่าจริง (integral exists) แล้ว เราจะเรียกฟังก์ชัน f ว่า ฟังก์ชันความน่าจะเป็นร่วม (joint probability function: joint p.f.) ของ X และ Y

เช่นเดียวกับกรณีของตัวแปรสุ่มไม่ต่อเนื่องและตัวแปรสุ่มต่อเนื่อง ฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) ของตัวแปรสุ่มที่มีการแจกแจงผสมสอดคล้องกับคุณสมบัติหลักสองประการ คือ ความไม่เป็นลบ (non-negativity) และผลรวมเท่ากับหนึ่ง (summing to one) ดังต่อไปนี้

คุณสมบัติที่ 3.6. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่มีการแจกแจงไม่ต่อเนื่อง (discrete) และต่อเนื่อง (continuous) ตามลำดับ ฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) ของ X และ Y สอดคล้องกับเงื่อนไขต่อไปนี้*

$$f(x,y) \geqslant 0, \text{ สำหรับ } (x,y) \text{ ใดๆ}, \tag{3.27}$$

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} f(x_i, y)\, dy = 1 \tag{3.28}$$

ตัวอย่างที่ 3.18 (ประยุกต์จาก DeGroot and Schervish (2012))**.** พิจารณาการทดลองทางคลินิก (clinical trial) ที่ต้องการประเมินผลของการรักษาเบาหวานด้วยยาชนิดหนึ่ง โดยกำหนดให้ X แทนตัวแปรสุ่มที่บ่งบอกว่าผลการรักษาได้ผล ($X = 1$) หรือไม่ได้ผล ($X = 0$) ซึ่งหากเราทราบว่าความน่าจะเป็นที่ผลการรักษาได้ผลหรือสัดส่วนของผู้ที่หายป่วยหลังจากได้รับการรักษา p อย่างแน่นอน เราก็จะสามารถแทนตัวแปรสุ่ม X ได้ด้วยตัวแปรสุ่มเบอร์นูลีที่มีพารามิเตอร์ p แต่เพื่อให้ปัญหานี้น่าสนใจมากยิ่งขึ้น สมมุติว่าเราไม่สามารถระบุค่าสัดส่วนหรือความน่าจะเป็นนั่นได้อย่างแน่นอน ดังนั้น เราจึงกำหนดให้สัดส่วนหรือความน่าจะเป็นนั้นเป็นตัวแปรสุ่ม P ซึ่งเป็นตัวแปรสุ่มต่อเนื่องที่มีค่าอยู่ระหว่างศูนย์ถึงหนึ่ง ดังนั้น ฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) ของ X และ P คือ

$$f(x,p) = cp^x (1-p)^{1-x}, \text{ สำหรับ } x = 0, 1 \text{ และ } 0 < p < 1$$

โดยที่ค่าคงที่มาตรฐาน c สามารถคำนวณได้จาก

$$1 = c \int_0^1 \sum_{x=0}^{1} p^x (1-p)^{1-x}\, dp = c \int_0^1 [(1-p) + p]\, dp = c$$

นั่นคือ $c = 1$ หลังจากนั้น เราสามารถคำนวณหาความน่าจะเป็นของเหตุการณ์ที่ผลการรักษาจะได้ผล $X = 1$ และสัดส่วนของผู้ป่วยที่ได้รับการรักษาได้ผลมากกว่าหรือเท่ากับหนึ่งส่วนสาม $P \geqslant \frac{1}{3}$ ได้ดังต่อไปนี้

$$Pr\left(X = 1, P \geqslant \frac{1}{3}\right) = \int_{\frac{1}{3}}^1 p dp = \frac{1}{2} - \frac{\left(\frac{1}{3}\right)^2}{2} = \frac{4}{9}$$

ในทำนองเดียวกัน ความน่าจะเป็นที่ผลการรักษาจะไม่ได้ผลมีค่าเท่ากับ

$$Pr(X = 0) = \int_0^1 (1-p)\, dp = -\frac{1}{2}\left[(1-1)^2 - (1-0)^2\right] = \frac{1}{2}$$

□

เช่นเดียวกับกรณีของตัวแปรสุ่มเดี่ยว เราสามารถนิยามฟังก์ชันความน่าจะเป็นสะสมร่วม (joint cumulative distribution function) ดังนิยามต่อไปนี้

บทนิยามที่ 3.16. *ฟังก์ชันความน่าจะเป็นสะสมร่วม (joint cumulative distribution function หรือ joint C.D.F.) ของตัวแปรสุ่ม X และ Y คือ*

$$F(x,y) = Pr(X \leqslant x, Y \leqslant y) \tag{3.29}$$

ซึ่งเป็นฟังก์ชันที่นิยามสำหรับค่า $(x,y) \in \mathbb{R}^2$

หากตัวแปรสุ่ม X และ Y เป็นแบบต่อเนื่อง เราสามารถเขียนฟังก์ชันความน่าจะเป็นสะสมร่วม (joint C.D.F.) ของตัวแปรสุ่ม X และ Y ได้เป็น

$$F(x,y) = Pr(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(r,s)\,drds \tag{3.30}$$

ในทางกลับกัน เราสามารถประยุกต์ใช้ทฤษฎีบทพื้นฐานของแคลคูลัส (fundamental theorem of calculus) เพื่อพิสูจน์ความสัมพันธ์ที่ใช้ในการคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) จากระหว่างฟังก์ชันความน่าจะเป็นสะสมร่วม (joint C.D.F.)

คุณสมบัติที่ 3.7. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มต่อเนื่อง (continuous random variable) ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) สามารถคำนวณจากฟังก์ชันความน่าจะเป็นสะสมร่วม (joint C.D.F.) ได้ดังต่อไปนี้*

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 F(x,y)}{\partial y \partial x} \tag{3.31}$$

สำหรับค่าจำนวนจริง x และ y ใดๆ ที่ค่าอนุพันธ์กำลังสอง (second-order derivative) หาค่าได้ (exist)

ตัวอย่างที่ 3.19. พิจารณาตัวแปรสุ่ม X และ Y ที่มีฟังก์ชันความน่าจะเป็นร่วม (joint C.D.F.) เป็น

$$F(x,y) = \frac{1}{156}xy\left(x^2 + y\right)$$

สำหรับค่า (x,y) ที่ $0 \leqslant x \leqslant 3$ และ $0 \leqslant y \leqslant 4$

เราสามารถหาฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ได้จากอนุพันธ์กำลังสอง (second-order derivative) ของฟังก์ชันความน่าจะเป็นร่วม ดังต่อไปนี้

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{156}\frac{\partial}{\partial x}\left[x^3 + 2xy\right] = \frac{1}{156}\left[3x^2 + 2y\right]$$

□

3.5 การแจกแจงตามขอบ (Marginal Distributions)

หัวข้อนี้นำเสนอหลักการหาการแจกแจงของตัวแปรสุ่มตัวใดตัวหนึ่งจากฟังก์ชันความน่าจะเป็นร่วม โดยเริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของตัวแปรสุ่มตัวใดตัวหนึ่งจากฟังก์ชันความน่าจะเป็นสะสมร่วม (joint C.D.F.) ซึ่งเรียกผลลัพธ์ที่ได้ว่า ฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.)

บทนิยามที่ 3.17. *กำหนดให้ X และ Y เป็นตัวแปรสุ่ม ฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของตัวแปรสุ่ม X และ Y สามารถนิยามได้เป็น*

$$F_1(x) = Pr(X \leqslant x), \tag{3.32}$$

$$F_2(y) = Pr(Y \leqslant y) \tag{3.33}$$

ตามลำดับ

เราสามารถประยุกต์ใช้หลักการของทฤษฎีเซ็ตที่ว่า

$$\{X \leqslant x\} = \{X \leqslant x\} \cap \{Y < \infty\}, \tag{3.34}$$

$$\{Y \leqslant y\} = \{Y \leqslant y\} \cap \{X < \infty\} \tag{3.35}$$

เพื่อเขียนฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของตัวแปรสุ่ม X และ Y ใหม่ได้เป็น

$$F_1(x) = Pr(X \leqslant x, Y < \infty) = \lim_{y \to \infty} F(x, y), \tag{3.36}$$

$$F_2(y) = Pr(X < \infty, Y \leqslant y) = \lim_{x \to \infty} F(x, y) \tag{3.37}$$

ในทำนองเดียวกับกรณีตัวแปรสุ่มเดี่ยว เราสามารถคำนวณค่าฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ได้โดยอาศัยความสัมพันธ์ระหว่างฟังก์ชันความน่าจะเป็นและฟังก์ชันความหนาแน่นของความน่าจะเป็น ทฤษฎีบทต่อไปนี้แสดงผลในกรณีของตัวแปรสุ่มต่อเนื่อง

ทฤษฎีบทที่ 3.11. *ถ้า X และ Y เป็นตัวแปรสุ่มต่อเนื่องที่มี $f(x, y)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) แล้ว ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม X และ Y คือ*

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \text{ สำหรับ } -\infty < x < \infty \tag{3.38}$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx, \text{ สำหรับ } -\infty < y < \infty \tag{3.39}$$

การพิสูจน์. เริ่มจากนิยามของฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของ X

$$F_1(x) = Pr(X \leqslant x, Y < \infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(r,y)\,dydr = \int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f(r,y)\,dy\right]dr \qquad (3.40)$$

ในขณะเดียวกัน เราสามารถนิยามฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของ X ในรูปฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ได้เป็น

$$F_1(x) = Pr(X \leqslant x) = \int_{-\infty}^{x} f_1(r)\,dr$$

ดังนั้น เราสามารถสรุปได้ว่า

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

ในทำนองเดียวกัน เราสามารถแสดงได้ว่า

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

■

ผลที่ตามมาจากทฤษฎีบทที่ 3.11 คือ เราสามารถคำนวณหาค่าฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ได้จากอนุพันธ์อันดับที่หนึ่งของฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ซึ่งเป็นผลมาจากการประยุกต์ทฤษฎีบทพื้นฐานของแคลคูลัสกับสมการที่ 3.40 เช่นเดียวกับกรณีก่อนหน้านี้ นั่นคือ

$$f_1(x) = \frac{dF_1(x)}{dx}, \qquad (3.41)$$

$$f_2(y) = \frac{dF_2(y)}{dx} \qquad (3.42)$$

นอกจากนี้ เราสามารถประยุกต์ใช้หลักการที่ว่า สามารถแทนการอินทิเกรตได้ด้วยการหาผลรวม (summation) หากตัวแปรสุ่มเป็นแบบไม่ต่อเนื่อง (discrete) ดังแสดงผลในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.12. *ถ้า X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete) ที่มี $f(x,y)$ แทนฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) แล้ว ฟังก์ชันความน่าจะเป็นตามขอบ (marginal p.f.) ของตัวแปรสุ่ม X และ Y คือ*

$$f_1(x) = \sum_y f(x,y), \qquad (3.43)$$

$$f_2(y) = \int_x f(x,y), \qquad (3.44)$$

โดยที่ส่วนห้อยของการหาผลรวม (summation) ในสมการที่ 3.43 และ 3.44 หมายถึงให้หาผลรวมจากทุกๆ ค่าของ y หรือ x ตามลำดับ

ส่วนกรณีที่มีการแจกแจงผสม (mixed distribution) ซึ่งประกอบไปด้วยตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง เราก็สามารถหาการแจกแจงขอบ (marginal distribution) ได้ในทำนองเดียวกัน ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.13. *ถ้า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete) และ Y เป็นตัวแปรสุ่มต่อเนื่อง (continuous) ที่มี $f(x,y)$ แทนฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) แล้ว ฟังก์ชันความน่าจะเป็นตามขอบ (marginal p.f.) ของตัวแปรสุ่ม X คือ*

$$f_1(x) = Pr(X = x) = \int_{-\infty}^{\infty} f(x,y)\, dy, \tag{3.45}$$

ส่วนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม Y คือ

$$f_2(y) = \sum_x f(x,y), \tag{3.46}$$

(3.47)

NEED AN EXAMPLE WITH DISCRETE DISTRIBUTIONS FROM FINANCIAL DATA AND PRESENT THE RESULT IN A TABLE

ตัวอย่างต่อไปนี้แสดงวิธีการคำนวณหาค่าการแจกแจงตามขอบ (marginal distribution) สำหรับกรณีที่การแจกแจงร่วมเป็นแบบผสม

ตัวอย่างที่ 3.20 (จาก DeGroot and Schervish (2012))**.** พิจารณา (X,Y) โดยที่ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete) ที่เป็นไปได้สามค่าคือ $1,2,3$ และ Y เป็นตัวแปรสุ่มต่อเนื่อง (continuous) ที่มีค่าอยู่ระหว่าง $(0,1)$ และมีฟังก์ชันความน่าจะเป็นดังต่อไปนี้

$$f(x,y) = \frac{xy^{x-1}}{3}, \text{ สำหรับ } x = 1,2,3 \text{ และ } 0 < y < 1$$

ดังนั้น ฟังก์ชันความน่าจะเป็นตามขอบ (marginal p.f.) ของตัวแปรสุ่ม X เท่ากับ

$$f_1(x) = \int_0^1 \frac{xy^{x-1}}{3} dy = \frac{1}{3}\left[y^x\Big|_0^1\right] = \frac{1}{3}, \text{ สำหรับ } x = 1,2,3$$

ส่วนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม Y มีค่าเท่ากับ

$$f_2(y) = \sum_{x=1}^{3} \frac{xy^{x-1}}{3} = \frac{1}{3} + \frac{2y}{3} + y^2, \text{ สำหรับ } 0 < y < 1$$

ผู้อ่านสามารถตรวจสอบดูได้ว่าฟังก์ชันความน่าจะเป็นตามขอบ (marginal p.f.) แต่ละอันมีคุณสมบัติที่เหมาะสมหรือไม่ ได้โดยตรวจสอบจากคุณสมบัติการที่ผลรวมต้องเท่ากับหนึ่ง □

ตัวอย่างต่อไปนี้แสดงการคำนวณหาค่าฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่มจากฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.)

ตัวอย่างที่ 3.21 (จาก Hogg et al. (2005))**.** พิจารณาตัวแปรสุ่ม (X, Y) ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) เท่ากับ

$$f(x,y) = \begin{cases} 4xye^{-x^2-y^2}, \text{ สำหรับ } 0 < x < \infty, 0 < y < \infty, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม X เท่ากับ

$$\begin{aligned} f_1(x) &= \int_0^\infty 4xye^{-x^2-y^2}dy = 2xe^{-x^2}\int_0^\infty 2ye^{-y^2}dy \\ &= 2xe^{-x^2}\left[-e^{-y^2}\Big|_0^\infty\right] = 2xe^{-x^2} \end{aligned}$$

ในทำนองเดียวกัน ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม Y เท่ากับ

$$f_2(y) = \int_0^\infty 4xye^{-x^2-y^2}dx = 2ye^{-y^2}$$

ข้อสังเกตที่น่าสนใจคือ ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม X ไม่ขึ้นอยู่กับค่าของตัวแปรสุ่ม Y เลย เช่นเดียวกับกรณีของฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม Y ที่ขึ้นอยู่กับค่า y เพียงอย่างเดียว ผลลัพธ์นี้เป็นคุณสมบัติที่สำคัญในทางสถิติที่เราเรียกว่า ความเป็นอิสระต่อกัน (independence) ซึ่งจะอภิปรายในรายละเอียดในหัวข้อถัดไป □

บทนิยามที่ 3.18. *ตัวแปรสุ่มสองตัว (X, Y) เป็นอิสระต่อกัน (independent) ถ้า*

$$Pr(X \in A \text{ และ } Y \in B) = Pr(X \in A)\, Pr(Y \in B) \tag{3.48}$$

สำหรับทุกเซ็ตของจำนวนจริง A และ B ที่ $\{X \in A\}$ และ $\{Y \in B\}$ เป็นเหตุการณ์

นอกจากนี้ เราสามารถสรุปได้ว่า ถ้า (X, Y) เป็นอิสระต่อกัน (independent) แล้ว

$$Pr(X \leqslant x \text{ และ } Y \leqslant y) = Pr(X \leqslant x)\, Pr(Y \leqslant y) \tag{3.49}$$

สำหรับทุกๆ จำนวนจริง x และ y ข้อสรุปนี้ช่วยให้เราสามารถนิยาม ความเป็นอิสระต่อกัน (independence) ในรูปของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ได้ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.14. *กำหนดให้ $F(x,y)$ แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ (X,Y) ส่วน $F_1(x)$ แทนฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของ X และ $F_2(y)$ แทนฟังก์ชันความน่าจะเป็นสะสมตามขอบ (marginal C.D.F.) ของ Y ดังนั้น X และ Y เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ (if and only if)*

$$F(x,y) = F_1(x) F_2(y) \tag{3.50}$$

สำหรับทุกๆ จำนวนจริง x และ y

สำหรับกรณีของตัวแปรสุ่มต่อเนื่อง เราสามารถแปลงเงื่อนไขที่ (3.50) ให้อยู่ในรูปของฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) และฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) โดยใช้ความสัมพันธ์เชิงอนุพันธ์ของการแจกแจงสะสมและฟังก์ชันความหนาแน่นของความน่าจะเป็น ได้ดังต่อไปนี้

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial F_1(x)}{\partial x}\frac{\partial F_2(y)}{\partial y} = f_1(x) f_2(y) \tag{3.51}$$

โดยสมการแรกเป็นผลมาจากสมการที่ (3.31) ส่วนสมการสุดท้ายเป็นผลมาจากสมการที่ (3.41) และ (3.42)

ทฤษฎีบทที่ 3.15. *ตัวแปรสุ่มสองตัว (X,Y) เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ (if and only if) เงื่อนไขการแยกตัวประกอบ (factorization)*

$$f(x,y) = f_1(x) f_2(y) \tag{3.52}$$

เป็นจริงสำหรับทุกๆ จำนวนจริง x และ y

อย่างไรก็ตาม ผู้อ่านควรระมัดระวังในการประยุกต์ใช้ทฤษฎีบทที่ 3.15 โดยเฉพาะเงื่อนไขที่ว่าสมการที่ 3.52 จะต้องเป็นจริงสำหรับทุกๆ จำนวนจริง x และ y เพราะตัวแปรสุ่มสองตัวที่สอดคล้องกับเงื่อนไขดังกล่าวภายใต้ข้อจำกัดพิเศษบางประเภทเท่านั้นอาจจะไม่เป็นอิสระต่อกัน ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 3.22 (จาก DeGroot and Schervish (2012))**.** พิจารณาตัวแปรสุ่ม X และ Y ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) เป็น

$$f(x,y) = \begin{cases} kx^2y^2, \text{ สำหรับ } x^2+y^2 \leqslant 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

โดยที่ $k > 0$ คือค่าคงที่มาตรฐาน สังเกตได้ว่า เราสามารถเขียนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ให้อยู่ในรูปของการแยกตัวประกอบตามสมการที่ 3.52 สำหรับ (x,y) ที่สอดคล้องกับเงื่อนไข

$x^2+y^2 \leqslant 1$ อย่างไรก็ตาม เราจำเป็นจะต้องตรวจสอบดูว่า เงื่อนไขดังกล่าวเป็นจริงในส่วนอื่นๆ ด้วยหรือไม่ โดยเริ่มจากการหาฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของแต่ละตัวแปรสุ่ม ดังนี้

$$f_1(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} kx^2y^2dy = \frac{2k}{3}x^2\left(1-x^2\right)^{\frac{3}{2}},$$
$$f_2(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} kx^2y^2dx = \frac{2k}{3}y^2\left(1-y^2\right)^{\frac{3}{2}}$$

จากฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ที่กำหนดให้ เราสามารถคำนวณหาค่า $f\left(\sqrt{0.75},\sqrt{0.75}\right) = 0$ เนื่องจาก $\sqrt{0.75}^2+\sqrt{0.75}^2 = 1.5 < 1$ ในขณะเดียวกัน ค่าฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ที่จุดดังกล่าวมีค่าเท่ากับ $f_1\left(\sqrt{0.75}\right) = f_2\left(\sqrt{0.75}\right) = 0.0625k \neq 0$ นั้นคือ เงื่อนไขการแยกตัวประกอบ (factorization) ไม่เป็นจริงสำหรับค่า (x,y) ที่อยู่นอกส่วนค้ำจุน ดังนั้น เราสามารถสรุปได้ว่า ตัวแปรสุ่ม X และ Y ไม่เป็นอิสระต่อกัน □

ข้อสังเกตจากตัวอย่างที่ 3.22 คือ ขอบเขต (boundaries) ของส่วนค้ำจุน (support) ของตัวแปรสุ่ม (X,Y) ซึ่งนิยามโดย $\{(x,y) : f(x,y) > 0\}$ มีลักษณะเป็นเส้นโค้ง (curved) หรือไม่ขนานกับแกนของปริภูมิของยุคลิด (Euclidean space) ทำให้ตัวแปรทั้งสองขาดความเป็นอิสระต่อกัน ในขณะเดียวกัน หากส่วนค้ำจุนของเป็นรูปสี่เหลี่ยมมุมฉาก (rectangular) ซึ่งมีขอบเขตขนานกับแกนตั้งและแกนนอน ตัวแปรสุ่มทั้งสองจะเป็นอิสระต่อกัน ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 3.23. พิจารณาตัวแปรสุ่ม X และ Y ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) เป็น

$$f(x,y) = \begin{cases} 9x^2y^2, \text{ สำหรับ } 0 \leqslant x \leqslant 1 \text{ และ } 0 \leqslant y \leqslant 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ซึ่งมีรูปแบบคล้ายกับตัวอย่างที่ 3.22 ต่างกันเพียงส่วนค้ำจุน (support) ซึ่งในที่นี้มีรูปแบบเป็นสี่เหลี่ยมมุมฉาก แต่เนื่องจากมีการเปลี่ยนขอบเขต จึงจำเป็นต้องคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของแต่ละตัวแปรสุ่มใหม่ ดังนี้

$$f_1(x) = \int_0^1 9x^2y^2dy = 3x^2, \text{ สำหรับ } 0 \leqslant x \leqslant 1,$$
$$f_2(y) = \int_0^1 9x^2y^2dx = 3y^2, \text{ สำหรับ } 0 \leqslant y \leqslant 1$$

ในกรณีจะเห็นได้ว่า เงื่อนไขการแยกตัวประกอบตามสมการที่ 3.52 เป็นจริงสำหรับค่า $0 \leqslant x \leqslant 1$ และ $0 \leqslant y \leqslant 1$ ส่วนกรณีอื่นๆ นั้น $f_1(x) = f_2(y) = f(x,y) = 0$ ซึ่งสอดคล้องสมการที่ 3.52 เช่นกัน ดังนั้น เราสามารถสรุปได้ว่า ตัวแปรสุ่ม X และ Y เป็นอิสระต่อกัน □

ทฤษฎีบทต่อไปนี้ขยายผลจากตัวอย่างที่ 3.23 ซึ่งช่วยให้เรามั่นใจได้ว่าเมื่อใดเราสามารถประยุกต์ใช้ทฤษฎีบทที่ 3.15 ได้โดยไม่ต้องกังวลว่าส่วนค้ำจุนจะสร้างปัญหา

ทฤษฎีบทที่ 3.16. *ตัวแปรสุ่มสองตัว (X,Y) มีการแจกแจงร่วมต่อเนื่อง (continuous joint distribution) สมมุติว่า ส่วนค้ำจุน (support) ของตัวแปรสุ่ม (X,Y) ซึ่งนิยามโดย $\{(x,y) : f(x,y) > 0\}$ เป็นสี่เหลี่ยมมุมฉาก (อาจจะไม่มีขอบเขตก็เป็นได้) โดยแต่ละด้านขนานกับแกนของปริภูมิของยุคลิด (Euclidean space) แล้ว (X,Y) เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ (if and only if) เงื่อนไขการแยกตัวประกอบ (factorization)*

$$f(x,y) = f_1(x) f_2(y) \tag{3.53}$$

เป็นจริงสำหรับทุกๆ จำนวนจริง x และ y

ทฤษฎีบทต่อไปนี้ขยายผลทฤษฎีบทที่ 3.15 จากความเป็นอิสระต่อกันของตัวแปรสุ่มไปสู่ความเป็นอิสระต่อกันของฟังก์ชันของตัวแปรสุ่มที่เป็นอิสระต่อกัน

ทฤษฎีบทที่ 3.17. *ถ้าตัวแปรสุ่ม X และ Y เป็นอิสระต่อกัน (independent) แล้ว ตัวแปรสุ่ม $h(X)$ และ $g(Y)$ เป็นอิสระต่อกัน (independent) ไม่ว่าฟังก์ชัน h และ g จะมีรูปแบบอย่างไร*

การพิสูจน์. พิจารณาเหตุการณ์ $H = \{h(X) \leqslant r\}$ ซึ่งสามารถแปลงให้อยู่ในรูปของเซ็ตของ x ได้เป็น $A \equiv \{x : h(x) \leqslant r\}$ ในทำนองเดียวกัน สามารถแปลงเหตุการณ์ $G = \{g(Y) \leqslant s\}$ ได้เป็น $B \equiv \{y : g(y) \leqslant s\}$ เนื่องจาก X และ Y เป็นอิสระต่อกัน (independent) ดังนั้น

$$Pr(X \in A \text{ และ } Y \in B) = Pr(X \in A)\, Pr(Y \in B)$$

ซึ่งสามารถเขียนในรูปของเหตุการณ์ H และ G ได้เป็น

$$\begin{aligned} Pr(h(X) \in H \text{ และ } g(Y) \in G) &= Pr(X \in A \text{ และ } Y \in B) \\ &= Pr(X \in A)\, Pr(Y \in B) \\ &= Pr(h(X) \in H)\, Pr(g(Y) \in G) \end{aligned}$$

นั่นคือ ตัวแปรสุ่ม $h(X)$ และ $g(Y)$ สอดคล้องกับเงื่อนไขการแยกตัวประกอบของความน่าจะเป็นตามสมการที่ 3.48 ดังนั้น เราจึงสามารถสรุปได้ว่า $h(X)$ และ $g(Y)$ เป็นอิสระต่อกัน (independent) ■

3.6 การแจกแจงแบบมีเงื่อนไข (Conditional Distributions)

การวิเคราะห์ทางการเงินและเศรษฐศาสตร์มักให้ความสนใจในการพยากรณ์โดยมีเงื่อนไขมาจากสิ่งที่เราทราบแล้ว เช่น ต้องการทราบว่าผลตอบแทนของหลักทรัพย์มีค่าเท่าไหร่ หากเราทราบว่าธนาคารแห่งประเทศไทยได้ประกาศ

ขึ้นอัตราดอกเบี้ยนโยบาย หรือต้องการทราบว่า อัตราการออมของครัวเรือนในชนบทที่ระดับรายได้ต่างๆ มีค่าเป็นเท่าใด เป็นต้น สิ่งที่ต้องการทราบเหล่านี้ล้วนอยู่ในรูปแบบของการแจกแจงแบบมีเงื่อนไข (conditional distribution) แน่นอนว่า ผู้อ่านอาจจะคุ้นเคยกับการรายงานผลในรูปของค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ซึ่งเป็นรูปแบบหนึ่งของการรายงานการแจกแจงแบบมีเงื่อนไข แต่ควรตระหนักอยู่เสมอว่า ถ้ามีข้อมูลขนาดใหญ่เพียงพอ เราควรสนใจการแจกแจงมากกว่าค่าความคาดหวัง เพราะการแจกแจงนั้นบอกคุณสมบัติของตัวแปรสุ่มได้ครบถ้วนสมบูรณ์มากกว่า

การแจกแจงแบบมีเงื่อนไขของตัวแปรสุ่มไม่ต่อเนื่อง (discrete) เป็นผลโดยตรงจากหลักการความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) ซึ่งอภิปรายในหน้าที่ 16 โดยในที่นี้ จะเขียนในรูปของฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X ภายใต้เงื่อนไขที่ $Y = y$ ได้เป็น

$$g_1(x|y) \equiv Pr(X = x|Y = y) = \frac{Pr(X = x \text{ และ } Y = y)}{Pr(Y = y)} = \frac{f(x,y)}{f_2(y)} \tag{3.54}$$

บทนิยามที่ 3.19. *กำหนดให้ X และ Y คือตัวแปรสุ่มไม่ต่อเนื่อง (discrete) ที่มี $f(x,y)$ แทนฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) และ $f_2(y)$ แทนฟังก์ชันความน่าจะเป็นตามขอบ (marginal p.f.) ของตัวแปรสุ่ม Y ดังนั้น สำหรับค่าจำนวนจริง y ที่ $f_2(y) > 0$*

$$g_1(x|y) \equiv \frac{f(x,y)}{f_2(y)} \tag{3.55}$$

เรามักเรียก g_1 ว่าฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไข (conditional p.f.) ของ X ภายใต้เงื่อนไขของ Y ส่วนฟังก์ชัน $g_1(\cdot|y)$ คือการแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ X เมื่อค่าของ $Y = y$ ในทำนองเดียวกัน การแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ Y เมื่อค่าของ $X = x$ คือ

$$g_2(y|x) \equiv \frac{f(x,y)}{f_1(x)} \tag{3.56}$$

สำหรับค่าจำนวนจริง x ที่ $f_1(x) > 0$

ADD TABLE PRESENTING CONDITIONAL DISTRIBUTION FROM FINANCIAL DATA

ส่วนการแจกแจงแบบมีเงื่อนไขของตัวแปรสุ่มต่อเนื่อง (continuous) ก็สามารถนิยามได้โดยใช้สูตรเดียวกัน จะแตกต่างกันก็เพียงว่า เรามีอิสระในการกำหนดค่าของการแจกแจงแบบมีเงื่อนไข ณ จุดที่ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบเท่ากับศูนย์ ตราบเท่าที่ผลลัพธ์ที่ได้ยังมีคุณสมบัติเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)

บทนิยามที่ 3.20. *กำหนดให้ X และ Y คือตัวแปรสุ่มต่อเนื่อง (continuous) ที่มี $f(x,y)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) และ $f_1(x)$ และ $f_2(y)$ แทนฟังก์ชันความหนาแน่นของความ*

น่าจะเป็นตามขอบ (marginal p.d.f.) ของตัวแปรสุ่ม X และ Y ตามลำดับ ดังนั้น สำหรับค่าจำนวนจริง y ที่ $f_2(y) > 0$ การแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ X เมื่อค่าของ $Y = y$

$$g_1(x|y) \equiv \frac{f(x,y)}{f_2(y)}, \text{ สำหรับ } -\infty < x < \infty \tag{3.57}$$

ในทำนองเดียวกัน สำหรับค่าจำนวนจริง x ที่ $f_1(x) > 0$ การแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ Y เมื่อค่าของ $X = x$ คือ

$$g_2(y|x) \equiv \frac{f(x,y)}{f_1(x)}, \text{ สำหรับ } -\infty < y < \infty \tag{3.58}$$

ส่วนกรณีที่ $f_2(y) = 0$ หรือ $f_1(x) = 0$ เราสามารถกำหนดค่าของการแจกแจงแบบมีเงื่อนไข $g_1(x|y)$ หรือ $g_2(y|x)$ ตามลำดับ ได้โดยอิสระ ตราบเท่าที่การแจกแจงแบบมีเงื่อนไขนั้นเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น

เช่นเดียวกับความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) การแจกแจงแบบมีเงื่อนไข (conditional distribution) เป็นการแจกแจงแบบหนึ่ง กล่าวคือ สำหรับตัวแปรสุ่มต่อเนื่อง การแจกแจงแบบมีเงื่อนไขมีคุณสมบัติเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.18. *กำหนดให้ X และ Y คือตัวแปรสุ่มต่อเนื่อง (continuous) สำหรับค่า y ใดๆ การแจกแจงแบบมีเงื่อนไข (conditional distribution) $g_1(x|y)$ มีคุณสมบัติเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ที่เป็นฟังก์ชันของตัวแปร x ในทำนองเดียวกัน สำหรับค่า x ใดๆ การแจกแจงแบบมีเงื่อนไข (conditional distribution) $g_2(y|x)$ มีคุณสมบัติเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ที่เป็นฟังก์ชันของตัวแปร y*

การพิสูจน์. เพื่อประหยัดพื้นที่ เราจะแสดงการพิสูจน์สำหรับ $g_2(y|x)$ เท่านั้น โดยเริ่มจากกรณีที่ $f_1(x) > 0$ ดังนั้น การแจกแจงแบบมีเงื่อนไข (conditional distribution) $g_2(y|x)$ จะต้องสอดคล้องตามสมการที่ 3.58 ซึ่งมีผลทำให้ $g_2(y|x) \geqslant 0$ เนื่องจากทั้งตัวตั้งมีค่าเป็นบวกหรือศูนย์ส่วนตัวหารมีค่ามากกว่าศูนย์ ส่วนที่เหลือคือการพิสูจน์ว่า $g_2(y|x)$ สอดคล้องกับคุณสมบัติผลรวมเท่ากับหนึ่ง นั่นคือ

$$\int_{-\infty}^{\infty} g_2(y|x)\,dy = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_1(x)}dy = \frac{1}{f_1(x)}\int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{f_1(x)}f_1(x) = 1$$

โดยที่สมการรองสุดท้ายเป็นผลมาจากนิยามของฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) $f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$ ส่วนกรณีที่ $f_1(x) = 0$ นั้น นิยามที่ 3.20 ได้กำหนดให้ $g_2(y|x)$ มีคุณสมบัติเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) อยู่แล้ว ■

ส่วนกรณีของตัวแปรสุ่มไม่ต่อเนื่อง (discrete) การแจกแจงแบบมีเงื่อนไขมีคุณสมบัติเป็นฟังก์ชันความน่าจะเป็น (p.f.) เช่นเดียวกัน และวิธีการพิสูจน์ก็คล้ายคลึงกันมาก จึงขอไม่นำเสนอในรายละเอียด

เช่นเดียวกับความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) สอดคล้องกับกฎการคูณดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.19. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่มีการแจกแจง $f_1(x)$ และ $f_2(y)$ ตามลำดับ ในทำนองเดียวกัน กำหนดให้ $g_1(x|y)$ และ $g_2(y|x)$ แทนการแจกแจงแบบมีเงื่อนไขของ X เมื่อค่าของ $Y = y$ และ การแจกแจงแบบมีเงื่อนไขของ Y เมื่อค่าของ $X = x$ ตามลำดับ ดังนั้น การแจกแจงร่วมของ X และ Y สามารถหาค่าได้จาก*

$$f(x,y) = g_1(x|y)\, f_2(y), \tag{3.59}$$

$$f(x,y) = g_2(y|x)\, f_1(x) \tag{3.60}$$

โดยที่สมการแรกเป็นจริงสำหรับทุกๆ ค่า x และค่า y ที่ $f_2(y) > 0$ และในทางกลับกันสมการที่สองเป็นจริงสำหรับทุกๆ ค่า y และค่า x ที่ $f_1(x) > 0$

ทฤษฎีบทต่อไปนี้เป็นผลโดยตรงจากทฤษฎบทที่ 3.19 และนิยามของฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) โดยเรามักจะเรียกทฤษฎีบทนี้ว่า ทฤษฎีบทของเบส์ (Bayes' Theorem) เพื่อให้เกียรติกับโทมัส เบส์ (Thomas Bayes) ทฤษฎีบทของเบส์ (Bayes' Theorem) เป็นส่วนสำคัญในการวิเคราะห์ข้อมูลแบบของเบส์ (Bayesian analysis) ที่ได้รับความนิยมสำหรับการศึกษาแบบจำลองการเรียนรู้ (learning model) ซึ่งมีบทบาทสำคัญในการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) ถึงแม้ว่า ทฤษฎีบทต่อไปนี้สามารถใช้ได้กับทั้งตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง แต่เพื่อความสะดวก ขอนำเสนอในรูปแบบของตัวแปรสุ่มต่อเนื่องเป็นหลัก หากต้องการประยุกต์ใช้กับตัวแปรสุ่มไม่ต่อเนื่องก็เพียงเปลี่ยนจากฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นฟังก์ชันความน่าจะเป็น (p.f.)

ทฤษฎีบทที่ 3.20 (ทฤษฎีบทของเบส์ (Bayes' Theorem))**.** *กำหนดให้ X และ Y เป็นตัวแปรสุ่ม โดยที่ $f_1(x)$ และ $f_2(y)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของ X และ Y ตามลำดับ ความสัมพันธ์ระหว่างฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) สามารถเขียนได้ในรูปสมการดังต่อไปนี้*

$$g_1(x|y) = \frac{g_2(y|x)\, f_1(x)}{f_2(y)} \tag{3.61}$$

$$g_2(y|x) = \frac{g_1(x|y)\, f_2(y)}{f_1(x)} \tag{3.62}$$

ในทางปฏิบัติทฤษฎีของเบส์ (Bayes' Theorem) ช่วยให้เราสามารถคำนวณหาฟังก์ชันความน่าจะเป็น (distribution function) ที่เกี่ยวข้องในทุกรูปแบบ ไม่ว่าจะเป็นการแจกแจงร่วม การแจกแจงตามขอบของทุกตัวแปร ในกรณีที่ทราบการแจกแจงย่อยเพียงสองอย่าง คือ การแจกแจงตามขอบและการแจกแจงแบบมีเงื่อนไข โดยมักจะใช้ร่วมกับกฎของความน่าจะเป็นรวม (Law of total probability)

ทฤษฎีบทที่ 3.21. *ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) และฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) สำหรับตัวแปรสุ่ม X และ Y สอดคล้องกับกฎของความน่าจะเป็นรวม (Law of total probability) ดังต่อไปนี้*

$$f_1(x) = \int_{-\infty}^{\infty} g_1(x|y)\, f_2(y)\, dy, \tag{3.63}$$

$$f_2(y) = \int_{-\infty}^{\infty} g_2(y|x)\, f_1(x)\, dx \tag{3.64}$$

โดยที่ $f_1(x)$ และ $f_2(y)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของ X และ Y ตามลำดับ และ $g_1(x|y)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไขของ X เมื่อค่าของ $Y = y$ และ $g_2(y|x)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไขของ Y เมื่อค่าของ $X = x$

ตัวอย่างที่ 3.24 (จาก DeGroot and Schervish (2012))**.** สมมุติให้ค่า X คือค่าที่สุ่มจากการแจกแจงเอกรูป (uniform distribution) ในช่วง $[0, 1]$ และ Y คือค่าที่สุ่มหลังจากที่ทราบว่าค่า $X = x$ จากการแจกแจงเอกรูป (uniform distribution) ในช่วง $[x, 1]$ คำถามก็คือ ฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของ Y จะมีลักษณะอย่างไร?

เนื่องจาก X มีการแจกแจงแบบเอกรูปในช่วง $[0, 1]$ ดังนั้น เราสามารถเขียนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของ X ได้เป็น

$$f_1(x) = \begin{cases} 1, & \text{สำหรับ } 0 \leqslant x \leqslant 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

ในทำนองเดียวกัน ฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) ของ Y เมื่อค่าของ $X = x$ เท่ากับ

$$g_2(y|x) = \begin{cases} \frac{1}{1-x}, & \text{สำหรับ } x \leqslant y \leqslant 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

เมื่อประยุกต์ใช้กฎการคูณในทฤษฎีบทที่ 3.19 เราจะพบว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของ (X, Y) เท่ากับ

$$f(x, y) = \begin{cases} \frac{1}{1-x}, & \text{สำหรับ } 0 < x \leqslant y \leqslant 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

นอกจากนี้ เราสามารถหาฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ของ Y โดยประยุกต์ใช้กฎของความน่าจะเป็นรวมในทฤษฎีบทที่ 3.21 ดังต่อไปนี้

$$f_2(y) = \begin{cases} \int_0^y \frac{1}{1-x}dx = -\ln(1-y) & \text{สำหรับ } 0 < y < 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

สุดท้ายเราสามารถหาฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) ของ X เมื่อค่าของ $Y = y$ โดยประยุกต์ใช้ทฤษฎีบทของเบส์ (Bayes' theorem) ดังต่อไปนี้

$$g_1(x|y) = \begin{cases} \frac{\frac{1}{1-x}}{-\ln(1-y)} = -\frac{1}{(1-x)\ln(1-y)} & \text{สำหรับ } 0 < x \leqslant y \leqslant 1, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

แน่นอนว่า ผลลัพธ์เดียวกันนี้สามารถหาได้จากการใช้นิยามของฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional p.d.f.) เพราะเราทราบฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) □

ทฤษฎีบทต่อไปนี้เชื่อมโยงหลักการของการแจกแจงแบบมีเงื่อนไข (conditional distribution) กับความเป็นอิสระต่อกัน (independence) ซึ่งช่วยให้สามารถทำความเข้าใจเรื่องความเป็นอิสระต่อกันได้ดียิ่งขึ้น กล่าวคือ ตัวแปรสุ่ม X และ Y เป็นอิสระต่อกันก็ต่อเมื่อการรู้ค่าของ Y ไม่มีผลกระทบต่อการแจกแจงตามขอบของ X และในทางกลับกัน การรู้ค่าของ X ไม่มีผลกระทบต่อการแจกแจงตามขอบของ Y

ทฤษฎีบทที่ 3.22. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่มีการแจกแจง $f_1(x)$ และ $f_2(y)$ ตามลำดับ ในทำนองเดียวกัน กำหนดให้ $g_1(x|y)$ และ $g_2(y|x)$ แทนการแจกแจงแบบมีเงื่อนไขของ X เมื่อค่าของ $Y = y$ และ การแจกแจงแบบมีเงื่อนไขของ Y เมื่อค่าของ $X = x$ ตามลำดับ ดังนั้น X และ Y เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ สำหรับทุกๆ ค่าของ y ที่ $f_2(y) > 0$ และทุกๆ ค่าของ x*

$$g_1(x|y) = f_1(x) \tag{3.65}$$

และสำหรับทุกๆ ค่าของ x ที่ $f_1(x) > 0$ และทุกๆ ค่าของ y

$$g_2(y|x) = f_2(y) \tag{3.66}$$

3.7 ฟังก์ชันของตัวแปรสุ่ม (Functions of Random Variables)

ในทางปฏิบัติ คำตอบของปัญหาที่เราต้องการแก้ไขมักจะอยู่ในรูปของฟังก์ชันของตัวแปรพื้นฐาน ทั้งนี้อาจจะเป็นผลมาจากการสร้างแบบจำลองทางเศรษฐศาสตร์หรือการเงินที่ซับซ้อน ทำให้ได้สมการซึ่งอยู่ในรูปของฟังก์ชันของ

ตัวแปรพื้นฐาน และที่สำคัญตัวแปรพื้นฐานส่วนใหญ่มักจะอยู่ในรูปของตัวแปรสุ่ม ทำให้คำตอบที่ได้เป็นตัวแปรสุ่มเช่นกัน ดังนั้น เพื่อให้เข้าใจคำตอบที่ได้ จึงจำเป็นจะต้องหาคุณสมบัติของคำตอบที่ได้ในฐานะที่เป็นตัวแปรสุ่ม ซึ่งก็หมายความว่า เราจะต้องหาการแจกแจงของฟังก์ชันของตัวแปรสุ่มให้ได้ เพื่อจะได้เข้าใจคำตอบของแบบจำลองที่สนใจ ดังนั้น เราจึงควรที่จะต้องเข้าใจหลักการหาการแจกแจงของฟังก์ชันของตัวแปรสุ่ม

ทฤษฎีบทที่ 3.23. *กำหนดให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องที่มีฟังก์ชันความน่าจะเป็น (p.f.) f และ $Y = r(X)$ เป็นฟังก์ชันที่นิยามบนเซ็ตของค่าที่เป็นไปได้ทั้งหมดของ X ดังนั้น ฟังก์ชันความน่าจะเป็น (p.f.) ของ Y ซึ่งแทนด้วย g มีค่าเท่ากับ*

$$g(y) = Pr(Y = y) = Pr(r(X) = y) = \sum_{x:r(X)=y} f(x) \tag{3.67}$$

สำหรับแต่ละค่า y ที่เป็นไปได้สำหรับตัวแปรสุ่ม Y

ในทำนองเดียวกัน การแจกแจงของตัวแปรสุ่มต่อเนื่องสามารถคำนวณได้โดยใช้หลักการสองขั้นตอนโดยเริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ก่อน แล้วจึงหาหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จากอนุพันธ์ของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.24. *กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) f และ $Y = r(X)$ เป็นฟังก์ชันที่นิยามบนเซ็ตของค่าที่เป็นไปได้ทั้งหมดของ X ดังนั้น ฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y ซึ่งแทนด้วย G มีค่าเท่ากับ*

$$G(y) = Pr(Y \leqslant y) = Pr(r(X) \leqslant y) = \int_{\{x:r(X)\leqslant y\}} f(x)\, dx \tag{3.68}$$

สำหรับแต่ละค่า y ที่เป็นไปได้สำหรับตัวแปรสุ่ม Y และถ้าฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y สามารถหาค่าอนุพันธ์ได้ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จะมีค่าเท่ากับ

$$g(y) = \frac{dG(y)}{dy} \tag{3.69}$$

ตัวอย่างที่ 3.25. สมมุติว่าตัวแปรสุ่ม X มีการแจกแจงเอกรูป (uniform distribution) ในช่วง $[0, 1]$ คำถามก็คือตัวแปรสุ่ม $Y = X^2$ มีการแจกแจงอย่างไร?

เนื่องจาก X มีการแจกแจงเอกรูป (uniform distribution) ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ X มีค่าเท่ากับ

$$f(x) = \begin{cases} 1 \text{ สำหรับ } 0 \leqslant x \leqslant 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

นอกจากนี้ เนื่องจาก $0 \leqslant x \leqslant 1$ ดังนั้น $0 \leqslant y \leqslant 1$ เช่นเดียวกัน

ในขั้นแรก เราจะต้องคำนวณหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y สำหรับค่า $0 \leqslant y \leqslant 1$ ดังต่อไปนี้

$$\begin{aligned} G(y) &= Pr(Y \leqslant y) = Pr\left(X^2 \leqslant y\right) = Pr\left(-\sqrt{y} \leqslant X \leqslant \sqrt{y}\right) = Pr\left(0 \leqslant X \leqslant \sqrt{y}\right) \\ &= \int_0^{\sqrt{y}} f(x)\, dx = \sqrt{y} \end{aligned}$$

โดยในที่นี้ สมการที่สี่เป็นผลมาจากการที่ $f(x) = 0$ สำหรับค่า x ที่ติดลบ ส่วนในขั้นตอนที่สองคือการหาอนุพันธ์ของ $G(y)$ เพื่อให้ได้ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)

$$g(y) = \frac{dG(y)}{dy} = \frac{1}{2\sqrt{y}}$$

สำหรับ $0 \leqslant y \leqslant 1$ ส่วนในช่วงอื่น $g(y) = 0$ □

ตัวอย่างต่อไปนี้แสดงให้เห็นว่าผลการแปลงแบบเชิงเส้นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (normal distribution) จะได้เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (normal distribution) เช่นเดียวกัน แต่อาจจะมีค่าพารามิเตอร์ที่ต่างออกไป

ตัวอย่างที่ 3.26. สมมุติว่า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (normal distribution) ที่มีพารามิเตอร์ μ และ σ ซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

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

คำถามก็คือ ตัวแปรสุ่ม $Y = aX + b (a \neq 0)$ โดยที่ ซึ่งเป็นฟังก์ชันเชิงเส้นของ X มีการแจกแจงอย่างไร?

เช่นเดียวกับตัวอย่างก่อนหน้านี้ เราจะใช้หลักการสองขึ้นคือเริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y และหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จากการหาอนุพันธ์ ในทางเทคนิคเราจำเป็นต้องแยกออกเป็นสองกรณีคือ

1. กรณีที่ $a > 0$:

$$\begin{aligned} G(y) &= Pr(Y \leqslant y) = Pr(aX + b \leqslant y) = Pr\left(X \leqslant \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

ขั้นตอนต่อไปคือการหาอนุพันธ์

$$g(y) = \frac{dG(y)}{dy} = \frac{d\int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{dy} = \frac{1}{a\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}$$
$$= \frac{1}{\sqrt{2\pi\left(a\sigma\right)^2}} e^{-\frac{(y-(a\mu+b))^2}{2(a\sigma)^2}}$$

โดยในที่นี้ สมการที่สามเป็นผลมาจากกฎการอินทิเกรตของไลบิซ[1] (Leibniz integral rule)

2. กรณีที่ $a < 0$:

$$G(y) = Pr(Y \leqslant y) = Pr(aX + b \leqslant y) = Pr\left(X \geqslant \frac{y-b}{a}\right)$$
$$= 1 - \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

ขั้นตอนต่อไปคือการหาอนุพันธ์

$$g(y) = \frac{dG(y)}{dy} = -\frac{d\int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{dy} = \frac{1}{-a\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}$$
$$= \frac{1}{\sqrt{2\pi\left(a\sigma\right)^2}} e^{-\frac{(y-(a\mu+b))^2}{2(a\sigma)^2}}$$

โดยสรุป ตัวแปรสุ่ม $Y = aX + b$ โดยที่ ซึ่งเป็นฟังก์ชันเชิงเส้นของ X มีการแจกแจงแบบปกติ (normal distribution) ที่มีพารามิเตอร์ $a\mu + b$ และ $a\sigma$

□

บทเรียนหนึ่งจากตัวอย่างที่ 3.26 คือ ผลลัพธ์ที่ได้จากกรณีที่ $a > 0$ และ $a < 0$ นั้นเหมือนกัน ซึ่งชี้ให้เห็นว่าสิ่งที่มีผลต่อการแปลงการแจกแจงในกรณีที่เป็นฟังก์ชันเชิงเส้นคือค่าสมบูรณ์ (absolute value) ซึ่งสามารถขยายผลไปสู่ตัวแปรสุ่มต่อเนื่องใดๆ ได้ดังแสดงในทฤษฎีบทต่อไปนี้

[1] กฎการอินทิเกรตของไลบิซบอกว่า

$$\frac{d\int_{a(y)}^{b(y)} f(x)\,dx}{dy} = f(b(y))\frac{db(y)}{dy} - f(a(y))\frac{da(y)}{dy}$$

ทฤษฎีบทที่ 3.25. *สมมุติให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มี f เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) และตัวแปรสุ่ม $Y = aX + b$ โดยที่ $a \neq 0$ แล้ว ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y เท่ากับ*

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right), \text{ สำหรับ } -\infty < y < \infty \tag{3.70}$$

อันที่จริง ทฤษฎีบทนี้เป็นกรณีพิเศษของทฤษฎีบทที่กล่าวถึงการแปลงการแจกแจงด้วยฟังก์ชันหนึ่งต่อหนึ่งที่หาอนุพันธ์ได้ (one-to-one and differentiable function) ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 3.26. *กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มี f เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) โดยที่ $Pr(a < X < b) = 1$ และตัวแปรสุ่ม $Y = r(X)$ โดยที่ฟังก์ชัน $r(x)$ เป็นฟังก์ชันหนึ่งต่อหนึ่งที่หาอนุพันธ์ได้ (one-to-one and differentiable function) สำหรับ $a < x < b$ ซึ่งมีผลทำให้เกิดภาพฉาย[2] (image) ของช่วง (a, b) เท่ากับ (α, β)*

กำหนดให้ $s(y)$ คือฟังก์ชันผกผัน (inverse function) ของ $r(x)$ นั่นคือ $x = s(y)$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y มีค่าเท่ากับ

$$g(y) = \begin{cases} f(r(x)) \left|\frac{ds(y)}{dy}\right| \text{ สำหรับ } \alpha \leqslant y \leqslant \beta, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases} \tag{3.71}$$

การพิสูจน์. การพิสูจน์แบ่งออกเป็นสองส่วน ดังต่อไปนี้

1. กรณีที่ r เป็นฟังก์ชั่นที่เพิ่มขึ้น (increasing function): ผลที่ตามมาก็คือ s ก็เป็นฟังก์ชั่นที่เพิ่มขึ้น ซึ่งช่วยให้เราเขียนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y สำหรับ $y \in (\alpha, \beta)$

   $$G(y) = Pr(Y \leqslant y) = Pr(r(X) \leqslant y) = Pr(X \leqslant s(y)) = F(s(y))$$

   เราสามารถประยุกต์ใช้กฎลูกโซ่ (chain rule) เพื่อหาอนุพันธ์ของ G ได้ดังต่อไปนี้

   $$g(y) = \frac{dG(y)}{dy} = \frac{dF(s(y))}{dy} = \frac{dF(s(y))}{dx}\frac{ds(y)}{dy} = f(s(y))\frac{ds(y)}{dy} = f(s(y))\left|\frac{ds(y)}{dy}\right|$$

2. กรณีที่ r เป็นฟังก์ชั่นที่ลดลง (decreasing function): ผลที่ตามมาก็คือ s ก็เป็นฟังก์ชั่นที่ลดลง ซึ่งช่วยให้เราเขียนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y สำหรับ $y \in (\alpha, \beta)$

   $$G(y) = Pr(Y \leqslant y) = Pr(r(X) \leqslant y) = Pr(X \geqslant s(y)) = 1 - F(s(y))$$

[2]หากพิจารณาอย่างไม่เป็นทางการ ภาพฉาย (image) หมายถึงเซ็ตของจำนวนจริงที่เป็นผลลัพธ์ของฟังก์ชัน r ที่เป็นไปได้ทั้งหมด หรือในทางเทคนิค เราอาจจะเขียนได้เป็น $r : (a, b) \to (\alpha, \beta)$

เราสามารถประยุกต์ใช้กฎลูกโซ่ (chain rule) เพื่อหาอนุพันธ์ของ G ได้ดังต่อไปนี้

$$g(y) = \frac{dG(y)}{dy} = -\frac{dF(s(y))}{dy} = -\frac{dF(s(y))}{dx}\frac{ds(y)}{dy} = -f(s(y))\frac{ds(y)}{dy} = f(s(y))\left|\frac{ds(y)}{dy}\right|$$

■

ตัวอย่างที่ 3.27. สมมุติให้ X เป็นตัวแปรสุ่มที่บอกถึงอัตราการให้บริการลูกค้าที่เข้าแถวรอของธนาคารแห่งหนึ่ง โดยกำหนดให้ f แทนฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ X คำถามที่สนใจก็คือ การแจกแจงของเวลาที่ใช้ในการรอ (waiting time) เป็นอย่างไร?

เวลาที่ใช้ในการรอ (waiting time) Y เป็นส่วนกลับของอัตราการให้บริการ นั่นคือ $Y = \frac{1}{X}$ ดังนั้น เราสามารถประยุกต์ใช้ทฤษฎีบทที่ 3.26 โดยนิยามฟังก์ชัน $r(x) = \frac{1}{x}$ ดังนั้น $s(y) = \frac{1}{y}$ และอนุพันธ์ของ s เท่ากับ

$$\frac{ds(y)}{dy} = -\frac{1}{y^2}$$

ซึ่งมีค่าน้อยกว่าศูนย์ ดังนั้น $\left|\frac{ds(y)}{dy}\right| = \frac{1}{y^2}$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y เท่ากับ

$$g(y) = f\left(\frac{1}{y}\right)\frac{1}{y^2}$$

□

ในตัวอย่างที่ 3.27 เราหาอนุพันธ์ของ s โดยการหา $s(y)$ จากฟังก์ชันผกผันของ $r(x)$ ก่อนแล้วจึงหาค่าอนุพันธ์ ซึ่งในบางครั้งการหาอนุพันธ์ของฟังก์ชันผกผันโดยตรงอาจจะมีความยุ่งยาก อย่างไรก็ดี เราสามารถประยุกต์ใช้ทฤษฎีบทจากแคลคูลัสสำหรับการหาอนุพันธ์ของฟังก์ชันผกผัน ดังนี้

$$\frac{ds(y)}{dy} = \left.\frac{1}{\frac{dr(x)}{dx}}\right|_{x=s(y)} \tag{3.72}$$

สังเกตว่า พจน์ด้านขวาหาได้ด้วยการหาอนุพันธ์ของฟังก์ชัน $r(x)$ เทียบกับตัวแปร x ก่อน แล้วจึงแทนค่า $x = s(y)$ เข้าไปในผลลัพธ์ที่ได้ ซึ่งจะทำให้ได้ผลลัพธ์สุดท้ายในรูปของ y

ตัวอย่างที่ 3.28 (จาก DeGroot and Schervish (2012))**.** พิจารณาแบบจำลองการขยายตัวของจุลินทรีย์แบบหนึ่ง ซึ่งกำหนดว่า ขนาดประชากรของจุลินทรีย์ ณ เวลา t จะมีค่าเท่ากับ αe^{Xt} โดยที่ $\alpha > 0$ คือปริมาณจุลินทรีย์ ณ จุดเริ่มต้น $t = 0$ และ X คืออัตราการขยายตัวของจุลินทรีย์ ซึ่งในที่นี้จะสมมุติว่านักวิเคราะห์ไม่สามารถระบุได้แน่ชัดว่ามีค่าเท่าใด แต่จากข้อมูลที่มีอยู่ทราบว่ามีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$f(x) = \begin{cases} 3(1-x^2) \text{ สำหรับ } 0 \leqslant x \leqslant 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

เราสามารถประยุกต์ใช้ทฤษฎีบทที่ 3.26 เพื่อหาการแจกแจงของจำนวนประชากรของจุลินทรีย์ได้ดังต่อไปนี้ โดยเริ่มจากการหาฟังก์ชันผกผัน

$$s(y) = x \Rightarrow y = \alpha e^{s(y)t} \Rightarrow s(y) = \frac{1}{t} \ln \frac{y}{\alpha}$$

ในขณะเดียวกัน เราสามารถหาอนุพันธ์ของฟังก์ชันผกผัน s ได้โดยเริ่มจากการหาอนุพันธ์ของ r

$$\frac{dr(x)}{dx} = t\alpha e^{xt}$$

ดังนั้น อนุพันธ์ของ s หาได้จากการแทนค่า $x = s(y) = \frac{1}{t} \ln \frac{y}{\alpha}$ ลงในส่วนกลับของสมการนี้ ดังนี้

$$\frac{ds(y)}{dy} = \frac{1}{t\alpha e^{\left(\frac{1}{t} \ln \frac{y}{\alpha}\right)t}} = \frac{1}{ty}$$

ซึ่งมีค่ามากกว่าศูนย์สำหรับกรณีที่ค่า y อยู่ระหว่าง $\alpha < y < \alpha e^t$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y เท่ากับ

$$g(y) = \begin{cases} \frac{3\left[1-\left(\frac{1}{t} \ln \frac{y}{\alpha}\right)^2\right]}{ty} \text{ สำหรับ } \alpha < y < \alpha e^t, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

□

3.7.1 ฟังก์ชันของตัวแปรสุ่มสองตัวขึ้นไป (Functions of Two or More Random Variables)

หลักการที่ใช้ในการแปลงการแจกแจงในกรณีของฟังก์ชันของตัวแปรสุ่มสองตัวขึ้นไปก็ยังเป็นเช่นเดิมคือ เริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ก่อน แล้วจึงหาหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จากอนุพันธ์ของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) อาจจะมีความแตกต่างกันบ้างในทางเทคนิค โดยเฉพาะในกรณีของสูตรที่ใช้ในการแปลงโดยตรง (ในทำนองเดียวกับทฤษฎีบทที่ 3.26) ซึ่งจำเป็นต้องใช้เครื่องมือที่เรียกว่าจาโคเบียนเมตริกซ์ (Jacobian matrix) ซึ่งเป็นส่วนขยายของอนุพันธ์ของฟังก์ชันผกผันที่ใช้ในกรณีที่มีตัวแปรเดียว

ทฤษฎีบทที่ 3.27. *กำหนดให้ $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มต่อเนื่องที่มี f แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) สมมุติให้เซ็ต $S \subset \mathbb{R}^n$ โดยที่ $Pr((X_1, \ldots, X_n) \in S) = 1$*

พิจารณาตัวแปรสุ่ม $Y_1, \ldots, Y_n$ ซึ่งเป็นฟังก์ชันของ $X_1, \ldots, X_n$ ดังนี้

$$\begin{aligned} Y_1 &= r_1(X_1, \ldots, X_n) \\ &\vdots \\ Y_n &= r_n(X_1, \ldots, X_n) \end{aligned} \tag{3.73}$$

โดยที่ r_i สำหรับ $i = 1, \ldots, n$ เป็นฟังก์ชันหนึ่งต่อหนึ่งที่หาอนุพันธ์ได้ (one-to-one and differentiable function) ซึ่งแปลงค่าจากเซ็ต S ไปสู่เซ็ต $T \subset \mathbb{R}^n$ และกำหนดให้ฟังก์ชันผกผัน s_i ที่สอดคล้องกับฟังก์ชัน r_i สำหรับ $i = 1, \ldots, n$ เท่ากับ

$$\begin{aligned} x_1 &= s_1(y_1, \ldots, y_n) \\ &\vdots \\ x_n &= s_n(y_1, \ldots, y_n) \end{aligned} \tag{3.74}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของ $Y_1, \ldots, Y_n$ หาได้จาก

$$g(y_1, \ldots, y_n) = \begin{cases} f(s_1, \ldots, s_n)\,|J| \text{ สำหรับ } (y_1, \ldots, y_n) \in T, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases} \tag{3.75}$$

โดยที่ J แทนดีเทอร์มิแนนต์ (determinant) ของจาโคเบียนเมตริกซ์ (Jacobian matrix) ซึ่งมีค่าเท่ากับ

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{bmatrix} \tag{3.76}$$

ในที่นี้ สัญลักษณ์ det *หมายถึงดีเทอร์มิแนนต์ (determinant) ของเมตริกซ์จตุรัส*

สำหรับผู้อ่านที่ต้องการทำความเข้าใจเกี่ยวกับการพิสูจน์ทฤษฎีบทนี้ สามารถศึกษาเพิ่มเติมได้จาก Fitzpatrick (2005) ในหัวข้อเรื่องการเปลี่ยนตัวแปรของการอินทิเกรต (change of variables)

ตัวอย่างที่ 3.29 (จาก DeGroot and Schervish (2012))**.** พิจารณาตัวแปรสุ่มต่อเนื่อง X และ Y ซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) เท่ากับ

$$f(x_1, x_2) = \begin{cases} 4xy \text{ สำหรับ } 0 < x_1 < 1 \text{ และ } 0 < y < 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

สิ่งที่ต้องการทราบในตัวอย่างนี้คือ ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ตัวแปรสุ่ม $Z_1 = \frac{X}{Y}$ และ $Z_2 = XY$

ในที่นี้ เราสามารถเขียนฟังก์ชันที่เราสนใจได้เป็น

$$r_1(x, y) = \frac{x}{y} \text{ และ } r_2(x, y) = xy$$

เราสามารถหาฟังก์ชันผกผันของฟังก์ชันทั้งสองได้จากการแก้ระบบสมการ

$$z_1 = \frac{x}{y}$$

$$z_2 = xy$$

ซึ่งได้คำตอบเป็นฟังก์ชันผกผัน

$$x_1 = s_1(z_1, z_2) = \sqrt{\frac{z_1}{z_2}} \text{ และ } x_2 = s_2(z_1, z_2) = \sqrt{z_1 z_2}$$

นอกจากนี้ S ในตัวอย่างนี้เท่ากับ $S = \{(x, y) \in \mathbb{R}^2 : 0 < x_1 < 1 \text{ และ } 0 < y < 1\}$ ดังนั้น เราหาเซ็ต T ได้ดังต่อไปนี้ ก่อนอื่น เนื่องจาก $x > 0$, $y > 0$, $z_1 = \frac{x}{y}$ และ $z_2 = xy$ ดังนั้น $z_1, z_2 > 0$ หลังจากนั้น เราสามารถประยุกต์ใช้ฟังก์ชันผกผันร่วมกับเงื่อนไขของ x และ y ที่กำหนดให้ $0 < x_1 < 1$ และ $0 < y < 1$ ในการคำนวณหาเซ็ต $T = \left\{(z_1, z_2) \in \mathbb{R}^2 : 0 < z_2 < \frac{1}{z_1} \text{ และ } 0 < z_2 < z_1\right\}$ INSERT FIGURE OF S AND T BELOW

ขั้นตอนต่อไปคือ การหาดีเทอร์มิแนนต์ (determinant) ของจาโคเบียนเมตริกซ์ (Jacobian matrix)

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial z_1} & \frac{\partial s_1}{\partial z_2} \\ \frac{\partial s_2}{\partial z_1} & \frac{\partial s_2}{\partial z_2} \end{bmatrix} = \det \begin{bmatrix} \frac{1}{2}\sqrt{\frac{z_2}{z_1}} & \frac{1}{2}\sqrt{\frac{z_1}{z_2}} \\ -\frac{1}{2}\sqrt{\frac{z_2}{z_1^3}} & \frac{1}{2}\sqrt{\frac{1}{z_1 z_2}} = \frac{1}{2z_1} \end{bmatrix}$$

เนื่องจาก $z_1 > 0$ ดังนั้น $|J| = \frac{1}{2z_1}$ และฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ตัวแปรสุ่ม $Z_1 = \frac{X}{Y}$ และ $Z_2 = XY$ เท่ากับ

$$g(z_1, z_2) = \begin{cases} 2\frac{z_2}{z_1} \text{ สำหรับ } (z_1, z_2) \in T, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

□

เช่นเดียวกับกรณีฟังก์ชันหนึ่งตัวแปร การแปลงการแจกแจงในกรณีของฟังก์ชันเชิงเส้นนั้นเป็นกรณีพิเศษ ซึ่งผลที่ได้ขึ้นอยู่กับค่าสัมประสิทธิ์ของสมการเชิงเส้นเป็นหลัก ในขณะเดียวกัน ฟังก์ชันเชิงเส้นเป็นรูปแบบการวิเคราะห์ข้อมูลที่ได้รับความนิยม ทั้งนี้อาจเป็นผลมาจากความสะดวกในการวิเคราะห์และการคำนวณ

ทฤษฎีบทที่ 3.28. *กำหนดให้ $\boldsymbol{X} = (X_1, \ldots, X_n)$ เป็นตัวแปรสุ่มต่อเนื่องที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) f และนิยามฟังก์ชันเชิงเส้น*

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \tag{3.77}$$

โดยที่ $\boldsymbol{A}$ คือเมทริกซ์ไม่เอกฐาน (nonsingular matrix) ที่มีขนาด $n \times n$ (บางครั้งเราเรียกค่าจำนวนจริงในเมทริกซ์นี้ว่าค่าสัมประสิทธิ์) ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของ $\mathbf{Y}$ หาได้จาก

$$g(\boldsymbol{y}) = \frac{1}{|det\boldsymbol{A}|} f\left(\boldsymbol{A}^{-1}\boldsymbol{y}\right) \tag{3.78}$$

โดยที่ $\boldsymbol{A}^{-1}$ แทนส่วนผกผันของ $\boldsymbol{A}$

ทฤษฎีบทสองอันหลังนั้นมีประโยชน์ต่อการหาการแจกแจงของฟังก์ชันของตัวแปรสุ่มอย่างมาก แต่ก็มีข้อจำกัดในบางประการ เช่น จำนวนฟังก์ชันที่พิจารณาต้องเท่ากับจำนวนตัวแปรสุ่มพื้นฐาน เพื่อให้จาโคเบียนเมตริกซ์ (Jacobian matrix) เป็นจัตรัสจึงจะหาค่าได้ ในขณะเดียวกัน เราอาจจะสนใจปัญหาที่มีจำนวนฟังก์ชันแตกต่างจากจำนวนตัวแปรสุ่มพื้นฐาน ซึ่งโดยมากมักจะมีจำนวนฟังก์ชันน้อยกว่าจำนวนตัวแปร ยกตัวอย่างเช่น การประมาณการณ์สมการเชิงเส้นอย่างง่ายมักจะมีฟังก์ชันเดียวแต่มีตัวแปรสุ่มพื้นฐานจำนวนมาก กรณีเช่นนี้เราสามารถกลับไปใช้หลักพื้นฐานที่เริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ก่อน แล้วจึงหาหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จากอนุพันธ์ของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ดังแสดงในสามตัวอย่างต่อไปนี้

ตัวอย่างที่ 3.30. พิจารณาตัวอย่างสุ่ม (random sample) ขนาด n ซึ่งสุ่มเลือกมาจากตัวแปรสุ่ม $X_1, X_2, \ldots, X_n$ โดยที่แต่ละตัวมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) f และฟังก์ชันความน่าจะเป็นสะสมร่วม (C.D.F.) F ที่เหมือนกัน พิจารณาตัวแปรสุ่มที่แทนค่าสูงสุด Y_m และค่าต่ำสุด Y_0 ของตัวแปรสุ่มเหล่านั้นดังต่อไปนี้

$$Y_m = \max\{X_1, X_2, \ldots, X_n\},$$
$$Y_0 = \min\{X_1, X_2, \ldots, X_n\}$$

สิ่งที่ต้องการทราบคือ การแจกแจงของ Y_m และ Y_0 ซึ่งสามารถคำนวณได้ตามขั้นตอนต่อไปนี้ เริ่มจากฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของค่าสูงสุด Y_m

$$G_m(y) = Pr(Y_m \leqslant y) = Pr(X_1 \leqslant y, X_2 \leqslant y, \ldots, X_n \leqslant y)$$

นอกจากนี้ การที่ข้อมูลเป็นแบบตัวอย่างสุ่ม (random sample) ทำให้ทราบว่า $X_1, X_2, \ldots, X_n$ เป็นอิสระต่อกัน (independent) ดังนั้น

$$\begin{aligned} G_m(y) &= Pr(X_1 \leqslant y, X_2 \leqslant y, \ldots, X_n \leqslant y) = Pr(X_1 \leqslant y)\, Pr(X_2 \leqslant y) \cdots Pr(X_n \leqslant y) \\ &= [F(y)]^n \end{aligned}$$

ส่วนฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y_n หาได้จากอนุพันธ์ของ $G_n(y)$ ดังนี้

$$g_m(y) = \frac{dG_m(y)}{dy} = nf(y)[F(y)]^{n-1}$$

ในทำนองเดียวกัน การแจกแจงของฟังก์ชันค่าต่ำสุด Y_0 เป็นดังนี้

$$G_0(y) = Pr(Y_0 \leqslant y) = 1 - Pr(Y_0 > y) = 1 - Pr(X_1 > y, X_2 > y, \ldots, X_n > y) = 1 - Pr(X_1 > y)\,Pr(X_2$$
$$= 1 - [1 - F(y)]^n$$

และ

$$g(y) = \frac{dG(y)}{dy} = nf(y)[1 - F(y)]^{n-1}$$

นอกจากนี เรายังสามารถหาการแจกแจงร่วม (joint distribution) ของ (Y_0, Y_m) ได้ดังต่อไปนี้

$$G(y_0, y_m) = Pr(Y_0 \leqslant y_0, Y_m \leqslant y_n)$$

กำหนดให้ A แทนเหตุการณ์ที่ $Y_m \leqslant y_m$ นั่นคือเซ็ต $\{(x_1, \ldots, x_n) : Y_m \leqslant y_n\}$ และ B แทนเหตุการณ์ที่ $Y_0 \leqslant y_0$ นั่นคือเซ็ต $\{(x_1, \ldots, x_n) : Y_0 \leqslant y_0\}$ จากทฤษฎีบทที่ 1.17 ซึ่งกล่าวว่า $Pr(A \cap B) = Pr(A) - Pr(A \cap B^c)$ เราสามารถเขียน G ได้เป็น

$$\begin{aligned} G(y_0, y_m) &= Pr(Y_0 \leqslant y_0, Y_m \leqslant y_m) = Pr(Y_m \leqslant y_m) - Pr(Y_m \leqslant y_m, Y_0 > y_0) \\ &= G_m(y_m) - Pr(y_0 < X_1 \leqslant y_m, \ldots, y_0 < X_n \leqslant y_m) \\ &= G_m(y_m) - \Pi_{i=1}^n Pr(y_0 < X_i \leqslant y_m) = [F(y_m)]^n - [F(y_m) - F(y_0)]^n \end{aligned}$$

ส่วนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ในกรณีที่ $-\infty < y_0 < y_m < \infty$ สามารถหาได้จากอนุพันธ์ของ $G(y_0, y_m)$ ดังนี้

$$g(y_0, y_m) = \frac{\partial^2 G(y_0, y_m)}{\partial y_0 \partial y_m} = n(n-1) f(y_0) f(y_m) [F(y_m) - F(y_0)]^{n-2} \tag{3.79}$$

ส่วนในกรณีอื่นๆ จะกำหนดให้ $g(y_0, y_m) = 0$ □

ตัวอย่างต่อไปนี้แสดงการหาการแจกแจงของฟังก์ชันเชิงเส้นอย่างง่าย ซึ่งเป็นรูปแบบที่ได้รับความนิยม

ตัวอย่างที่ 3.31. กำหนดให้ X_1 และ X_2 เป็นตัวแปรสุ่มที่มี $f(x_1, x_2)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) คำถามก็คือ การแจกแจงของ $Y = a_1X_1 + a_2X_2 + b$ โดยที่ $a_1 \neq 0$ เป็นอย่างไร?

เช่นเดียวกับกรณีก่อนหน้านี้ เราเริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y

$$G\left(y\right)=Pr\left(Y\leqslant y\right)=Pr\left(a_1X_1+a_2X_2+b\leqslant y\right)$$

ซึ่งหมายถึงความน่าจะเป็นของเหตุการณ์ $A_y=\{(x_1,x_2):a_1x_1+a_2x_2+b\leqslant y\}$ ในกรณีที่ $a_1>0$ เราสามารถเขียนเหตุการณ์นี้ใหม่ได้เป็น $A_y=\left\{(x_1,x_2):x_1\leqslant\frac{y-a_2x_2-b}{a_1}\right\}$ ดังนั้น เขียนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ใหม่ในรูปการอินทิเกรตได้เป็น

$$G\left(y\right)=\iint\limits_{A_y}f\left(x_1,x_2\right)dx_1dx_2=\int_{-\infty}^{\infty}\int_{-\infty}^{\frac{y-a_2x_2-b}{a_1}}f\left(x_1,x_2\right)dx_1dx_2$$

หลังจากนั้น จึงประยุกต์ใช้กฎการอินทิเกรตของไลบิซ (Leibniz integral rule) เพื่อคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y ดังนี้

$$g\left(y\right)=\frac{dG\left(y\right)}{dy}=\int_{-\infty}^{\infty}f\left(\frac{y-a_2x_2-b}{a_1},x_2\right)\frac{1}{a_1}dx_2$$

ส่วนในกรณีที่ $a_1<0$ จะได้ว่า

$$g\left(y\right)=\frac{dG\left(y\right)}{dy}=-\int_{-\infty}^{\infty}f\left(\frac{y-a_2x_2-b}{a_1},x_2\right)\frac{1}{a_1}dx_2$$

ดังนั้น เมื่อนำมารวมกันแล้วสามารถสรุปได้ว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ $Y=a_1X_1+a_2X_2+b$ $(a_1\neq0)$ เท่ากับ

$$g\left(y\right)=\frac{dG\left(y\right)}{dy}=\int_{-\infty}^{\infty}f\left(\frac{y-a_2x_2-b}{a_1},x_2\right)\frac{1}{|a_1|}dx_2 \tag{3.80}$$

□

ตัวอย่างต่อไปนี้ผสมผสานผลลัพธ์ของตัวอย่างที่ 3.30 และ 3.31 เพื่อคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของค่าพิสัย (range) ของข้อมูล ซึ่งนิยามได้เป็น $Z=Y_m-Y_0$ โดยที่ Y_m คือค่าสูงสุด และ Y_0 คือค่าต่ำสุด ค่าพิสัย (range) เป็นเครื่องมือหนึ่งที่สามารถบอกถึงลักษณะการกระจายของข้อมูล

ตัวอย่างที่ 3.32. พิจารณาฟังก์ชัน $Z=Y_m-Y_0$ ซึ่งเรียกว่าค่าพิสัย (range) คำถามก็คือ การแจกแจงของค่าพิสัยเป็นอย่างไร?

จากตัวอย่างที่ 3.30 เราทราบแล้วว่าฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของตัวแปรสุ่ม (Y_0,Y_m) เท่ากับ

$$g\left(y_0,y_m\right)=\frac{\partial^2G\left(y_0,y_m\right)}{\partial y_0\partial y_m}=n\left(n-1\right)f\left(y_0\right)f\left(y_m\right)\left[F\left(y_m\right)-F\left(y_0\right)\right]^{n-2}$$

ในขณะเดียวกัน เราสามารถเทียบเคียงฟังก์ชันค่าพิสัย Z กับฟังก์ชันเชิงเส้นในตัวอย่างที่ 3.31 ได้โดยกำหนดให้ $a_1 = -1$, $a_2 = 1$, และ $b = 0$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Z เท่ากับ

$$h(z) = \int_{-\infty}^{\infty} g(y_n - z, y_n)\, dy_n$$

นอกจากนี้ เรายังสามารถประยุกต์ใช้หลักการเปลี่ยนตัวแปรการอินทิเกรต (change of variable) โดยกำหนดให้ $w = y_n - z$ ดังนั้น $dy_n = dw$ และ $y_n = w + z$ เพื่อแปลงฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Z เป็น

$$h(z) = \int_{-\infty}^{\infty} g(w, w + z)\, dw$$

□

บทที่ 4

ค่าคาดหมายและโมเมนต์ของตัวแปรสุ่ม (Expectation and Moments of Radom Variables)

บทที่ 3 ได้นำเสนอหลักการสำคัญที่ระบุว่า การแจกแจงของตัวแปรสุ่มเป็นเครื่องมือที่เก็บหรือบรรจุคุณสมบัติเชิงความน่าจะเป็น (probabilistic properties) ทั้งหมดของตัวแปรสุ่มนั้น กล่าวคือ หากเราทราบการแจกแจงก็หมายความว่าเราทราบคุณสมบัติเชิงความน่าจะเป็นของตัวแปรสุ่มนั้น อย่างไรก็ตาม เนื่องจากบรรจุคุณสมบัติทั้งหมดไว้ทำให้โดยทั่วไปการแจกแจงมีรายละเอียดมากจนทำให้เข้าใจและนำไปสื่อสารได้ยาก ทำให้จำเป็นจะต้องหาตัวแทนหรือจำนวนที่สามารถบอกถึงคุณสมบัติเชิงความน่าจะเป็นที่สำคัญของตัวแปรสุ่ม ซึ่งอาจจะไม่ครบถ้วนแต่ก็สามารถสะท้อนถึงส่วนที่สำคัญของตัวแปรสุ่มนั้น ในบทนี้ จะกล่าวถึง ค่าคาดหมาย (expectation or expected value) หรือบางครั้งอาจจะเรียกว่าค่ามัชฌิม[1] (mean) ซึ่งเป็นตัวแทนที่ได้รับความนิยมที่ใช้บอกถึงคุณสมบัติเชิงความน่าจะเป็นที่สำคัญของตัวแปรสุ่ม

[1]ผู้คนจำนวนมากมักแปลว่าค่าเฉลี่ย แต่เพื่อแยกความแตกต่างระหว่างเครื่องมือทางทฤษฎีและตัวประมาณการณ์ (estimator) ที่พยายามประมาณค่าจากข้อมูลจริง โดยค่าเฉลี่ย (average) ในหนังสือเล่มนี้จะหมายถึงตัวประมาณการณ์ซึ่งมีค่าเท่ากับการนำเอาข้อมูลทั้งหมดมารวมกันแล้วการด้วยจำนวนตัวอย่าง

4.1 ค่าคาดหมาย (Expectation)

บทนิยามที่ 4.1. *ค่าคาดหมาย (expectation or expected value or mean) ของตัวแปรสุ่ม X สำหรับตัวแปรสุ่มไม่ต่อเนื่อง (discrete) เท่ากับ*

$$E[X] = \sum_{x} x f(x)\, dx \tag{4.1}$$

โดยที่ $f(x)$ แทนฟังก์ชันความน่าจะเป็น (p.f.) ของ X และของตัวแปรสุ่มต่อเนื่อง (continuous) เท่ากับ

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx \tag{4.2}$$

โดยที่ $f(x)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ X

ข้อสังเกตที่สำคัญอันหนึ่งคือ ตัวแปรสุ่มสองตัวที่มีการแจกแจง (distribution) เหมือนกันย่อมมีค่าคาดหมาย (expectation) เหมือนกัน ดังนั้น จึงสามารถสรุปได้ว่า อันที่จริงแล้วค่าคาดหมาย (expectation) เป็นคุณสมบัติของการแจกแจง ทำให้อาจจะใช้ค่าคาดหมายของตัวแปรสุ่มกับค่าคาดหมายของการแจกแจงสลับกันไปมาโดยถือว่ามีความหมายเหมือนกัน

ตัวอย่างที่ 4.1 (การแจกแจงทวินาม (Binomial Distribution) สำหรับพารามิเตอร์ n และ p)**.** พิจารณาตัวแปรสุ่มแบบทวินาม (binomial random variable) ซึ่งมีแทนฟังก์ชันความน่าจะเป็น (p.f.) เท่ากับ

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{สำหรับ } x = 0, 1, \ldots, n, \\ 0, & \text{สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$\begin{aligned} E[X] &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{1} p (1-p)^{n-1} + 2\binom{n}{2} p^2 (1-p)^{n-2} + \cdots + n\binom{n}{n} p^n \\ &= np\binom{n-1}{0} (1-p)^{n-1-0} + np\binom{n-1}{1} p (1-p)^{n-1-1} + \cdots + np\binom{n-1}{n-1} p^{n-1} \\ &= np\left[\binom{n-1}{0} (1-p)^{n-1-0} + \binom{n-1}{1} p (1-p)^{n-1-1} + \cdots + \binom{n-1}{n-1} p^{n-1}\right] \\ &= np\left[(p + (1-p))^{n-1}\right] = np \end{aligned}$$

□

ตัวอย่างไปนี้เป็นการหาค่าคาดหมาย (expectation) ของตัวตัวแปรสุ่มปัวซองส์ (Poisson random variable) สำหรับพารามิเตอร์ λ ซึ่งเป็นตัวแปรสุ่มไม่ต่อเนื่องที่ได้รับนิยมใช้ในการวิเคราะห์สถานการณ์ที่เกี่ยวข้องกับการมาถึง (arrival) ของสิ่งใดสิ่งหนึ่งในช่วงเวลาใดเวลาหนึ่ง ยกตัวอย่างเช่น จำนวนการเปลี่ยนปแปลงของสารพันธุกรรม (mutation) ในแต่ละช่วงความยาวของดีเอ็นเอ (DNA) จำนวนครั้งที่จะมีลูกค้าโทรเข้ามาที่ศูนย์บริการ จำนวนการเรียกร้องค่าสินไหมทดแทน (claims) ในแต่ละช่วงเวลา เป็นต้น

ตัวอย่างที่ 4.2 (การแจกแจงปัวซองส์ (Poisson distribution) สำหรับพารามิเตอร์ λ)**.** พิจารณาตัวแปรสุ่มแบบปัวซองส์ (Poisson random variable) ซึ่งมีแทนฟังก์ชันความน่าจะเป็น (p.f.) เท่ากับ

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} \text{ สำหรับ } x = 0, 1, \ldots, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$E[X] = \sum_{x=0}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} = \lambda\left[\sum_{x=1}^{\infty}\frac{\lambda^{x-1}e^{-\lambda}}{(x-1)!}\right] = \lambda\left[\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!}\right] = \lambda$$

โดยที่สมการสุดท้ายเป็นมาจาก $\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!} = 1$[2] □

ตัวอย่างไปนี้แสดงการหาค่าคาดหมาย (expectation) ของตัวแปรสุ่มต่อเนื่องที่น่าจะพูดได้ว่าเป็นที่นิยมมากที่สุด นั่นคือ ตัวแปรสุ่มปกติ (normal random variable) สำหรับพารามิเตอร์ μ และ σ

ตัวอย่างที่ 4.3 (การแจกแจงแบบปกติ (normal distribution) สำหรับพารามิเตอร์ μ และ σ)**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

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

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$E[X] = \int_{-\infty}^{\infty}\frac{x}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

[2]พิจารณา

$$e^{\lambda} = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots = \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}$$

ซึ่งสามารถเขียนใหม่ได้เป็น

$$\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!} = 1$$

การเปลี่ยนตัวแปรของการอินทิเกรต (change of variables) $d\left[\frac{(x-\mu)^2}{2\sigma^2}\right] = \left[\frac{x-\mu}{\sigma^2}\right]dx$ ช่วยให้หาค่าการอินทิเกรตได้เป็น

$$E[X] = \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma^2}\right]^2}\left[\frac{x-\mu}{\sigma^2}\right]dx + \int_{-\infty}^{\infty}\mu f(x)\,dx = \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma^2}\right]^2} d\left[\frac{(x-\mu)^2}{2\sigma^2}\right] + \mu$$
$$= -\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} + \mu = \mu$$

□

ตัวอย่างไปนี้แสดงการหาค่าคาดหมาย (expectation) ของตัวแปรสุ่มต่อเนื่องที่ได้รับความนิยมโดยเฉพาะในด้านการเงิน และมีความสัมพันธ์กับตัวแปรสุ่มปกติ นั่นคือ ตัวแปรสุ่มปกติด้วยล็อก (lognormal random variable) ซึ่งล็อการิธึมของตัวแปรสุ่มนี้จะมีการแจกแจงปกติ (normal distribution) นั่นคือ ถ้าตัวแปรสุ่ม X มีการแจกแจงปกติด้วยล็อก (lognormal distribution) แล้ว ตัวแปรสุ่ม $\ln X$ จะมีการแจกแจงปกติ (normal distribution)

ตัวอย่างที่ 4.4 (การแจกแจงปกติด้วยล็อก (lognormal distribution))**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)[3] เท่ากับ

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \text{ สำหรับ } x > 0, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases} \tag{4.3}$$

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$E[X] = \int_0^{\infty} x\left[\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\right]dx$$

[3]ทั้งนี้ เป็นผลมาจากการที่ X มีการแจกแจงปกติด้วยล็อก (lognormal distribution) แล้ว ตัวแปรสุ่ม $\ln X$ จะมีการแจกแจงปกติ (normal distribution)

$$h(\ln x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \text{ สำหรับ } -\infty < \ln x < \infty, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$E[X] = \int_{-\infty}^{\infty} x\left[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\right]d\ln x = \int_0^{\infty} x\left[\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\right]dx$$

กำหนดให้ $y = \ln x$ และประยุกต์ใช้เทคนิคการเปลี่ยนตัวแปรของการอินทิเกรต (change of variables) จะได้ว่า

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{y} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= e^{\mu+\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left[y-\left(\mu+\sigma^2\right)\right]^2}{2\sigma^2}} dy = e^{\mu+\frac{\sigma^2}{2}} \end{aligned}$$

โดยที่สมการสุดท้ายเป็นผลมาจากคุณสมบัติของการแจกแจงที่การอินทิเกรตของฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ตลอดส่วนค้ำจุนมีค่าเท่ากับหนึ่ง สังเกตว่าค่าคาดหมาย (expectation) ของตัวแปรสุ่มปกติด้วยล็อก (lognormal random variable) มีส่วนสัมพันธ์กับค่าคาดหมาย (expectation) ของตัวแปรสุ่มปกติ (normal random variable) แต่มีส่วนเพิ่มเติมคือ $\frac{\sigma^2}{2}$ □

ตัวอย่างต่อไปนี้แสดงการหาค่าคาดหมาย (expectation) ของตัวแปรสุ่มที่มีการแจกแจงโลจิสติกส์ (logistic distribution) ซึ่งเป็นพื้นฐานสำคัญสำหรับการประมาณค่าแบบโลจิสติกส์ (logistic estimation) ซึ่งจะกล่าวถึงอย่างละเอียดในบทที่ XXX

ตัวอย่างที่ 4.5 (การแจกแจงโลจิสติกส์ (logistic distribution) สำหรับพารามิเตอร์ μ และ s)**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$f(x) = \frac{e^{\frac{x-\mu}{s}}}{s\left(1+e^{\frac{x-\mu}{s}}\right)^2} \tag{4.4}$$

โดยที่พารามิเตอร์ $s > 0$ และฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เท่ากับ

$$F(x) = \frac{1}{1+e^{-\frac{x-\mu}{s}}} \tag{4.5}$$

ดังนั้น ค่าคาดหมาย (expectation) ของ X มีค่าเท่ากับ

$$E[X] = \int_{-\infty}^{\infty} \frac{xe^{\frac{x-\mu}{s}}}{s\left(1+e^{\frac{x-\mu}{s}}\right)^2} dx$$

กำหนดให้ $y = e^{\frac{x-\mu}{s}}$ และประยุกต์ใช้เทคนิคการเปลี่ยนตัวแปรของการอินทิเกรต (change of variables) จะได้ว่า

$$E[X] = \int_{0}^{\infty} \frac{\mu + s\ln y}{(1+y)^2} dy = -\mu \int d\left[\frac{1}{1+y}\right] - s\int \ln y d\left[\frac{1}{1+y}\right]$$

สังเกตว่าพจน์สุดท้ายไม่ได้กำหนดขอบเขตของอินทิเกรต ทั้งนี้เพื่อให้ง่ายในการคำนวณ โดยเก็บไว้เป็นนัยว่าค่าของ y อยู่ในช่วง $[0,\infty)$ ในขั้นตอนต่อไป เราสามารถประยุกต์ใช้การอินทิเกรตโดยการแยกส่วน (integration by part) ทำให้ได้ว่า

$$E[X] = -\mu\left[\frac{1}{1+y}\right]_0^{\infty} - s\left[\frac{(1+y)\ln(1+y) - y\ln y}{1+y}\right]_0^{\infty} = \mu - s\left[\frac{(1+y)\ln(1+y) - y\ln y}{1+y}\right]_0^{\infty}$$
$$= \mu$$

โดยที่สมการสุดท้ายเป็นผลมาจากการที่ $\left[\frac{(1+y)\ln(1+y)-y\ln y}{1+y}\right]_0^{\infty} = 0$[4] □

ตัวอย่างที่ผ่านมาล้วนแล้วแต่เป็นตัวแปรสุ่มที่สามารถหาค่าคาดหมาย (expectation) ได้ อย่างไรก็ตาม ตัวแปรสุ่มบางตัวอาจจะมีการแจกแจงที่ไม่สามารถหาค่าคาดหมาย (expectation) ได้ ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 4.6 (การแจกแจงที่หาค่าคาดหมายไม่ได้ (non-existence of expectation))**.** พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$f(x) = \frac{1}{\pi(1+x^2)} \tag{4.6}$$

ขั้นแรกเราควรตรวจสอบว่าการแจกแจงนี้ถูกต้องตามหลักการ (well-defined distribution) ด้วยการตรวจสอบว่า

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

หรือไม่ นั่นคือ

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)}dx = \frac{1}{\pi}\tan^{-1}x\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1$$

[4]พิจารณาส่วนลิมิตล่างของพจน์นี้

$$\lim_{y\to 0}\frac{(1+y)\ln(1+y) - y\ln y}{1+y} = -\lim_{y\to 0}\frac{y\ln y}{1+y}$$

การประยุกต์ใช้กฎของโลปิตาล (L'Hôpital's rule) ทำให้สามารถสรุปได้ว่า

$$\lim_{y\to 0}\frac{y\ln y}{1+y} = \lim_{y\to 0}\frac{\ln y}{\frac{1+y}{y}} = \lim_{y\to 0}\frac{\frac{1}{y}}{-\frac{1}{y^2}} = 0$$

ในทำนองเดียวกัน

$$\lim_{y\to\infty}\frac{(1+y)\ln(1+y) - y\ln y}{1+y} = \lim_{y\to\infty}\frac{(1+y)+1-\ln y - 1}{1} = \lim_{y\to\infty}\ln\left[1+\frac{1}{y}\right] = 0$$

ซึ่งหมายความว่าการแจกแจงนี้ถูกต้องตามหลักการ (well-defined distribution) อันที่จริง การแจกแจงนี้มีชื่อเฉพาะว่า การแจกแจงโคชี (Cauchy distribution) อย่างไรก็ตาม

$$\int_0^{\infty} \frac{x}{\pi\left(1+x^2\right)} dx = \frac{1}{2}\int_0^{\infty} \frac{1}{\pi\left(1+x^2\right)} d\left[1+x^2\right] = \frac{1}{2\pi} \ln\left(1+x^2\right)\Big|_0^{\infty}$$

ซึ่งหาค่าไม่ได้เพราะมีค่าไม่จำกัด ในทำนองเดียวกัน

$$\int_{-\infty}^{0} \frac{x}{\pi\left(1+x^2\right)} dx = \frac{1}{2\pi} \ln\left(1+x^2\right)\Big|_{-\infty}^{0}$$

ซึ่งหาค่าไม่ได้เพราะมีค่าไม่จำกัด ดังนั้น เราจึงสรุปได้ว่าตัวแปรสุ่มที่มีการแจกแจงโคชี (Cauchy distribution) หาค่าคาดหมายไม่ได้ (non-existence) □

4.1.1 ค่าคาดหมายของฟังก์ชันของตัวแปรสุ่ม (Expectation of a Function of Random Variables)

4.2 subsec-exp-function

โดยปกติ แบบจำลองหรือทฤษฎีที่นักวิจัยหรือนักวิเคราะห์สร้างขึ้นมักจะอยู่ในรูปของฟังก์ชันของตัวแปรสุ่มพื้นฐาน (underlying random variables) ยกตัวอย่างเช่น อุปสงค์ (demand) ของสินค้าเป็นฟังก์ชันของราคาของสินค้านั้น ราคาของสินค้าอื่น รายได้ และปัจจัยอื่นๆ โดยในที่นี้ อาจจะมองได้ว่า ราคาและรายได้เป็นตัวแปรสุ่มพื้นฐาน ดังนั้น หาต้องการหาค่าคาดหมาย (expectation) ของอุปสงค์ของสินค้า ก็จำเป็นที่จะต้องสามารถคำนวณหาค่าคาดหมาย (expectation) ของฟังก์ชันอุปสงค์ของตัวแปรสุ่มพื้นฐานเหล่านั้น

พิจารณาตัวแปรสุ่ม $Y = r(X)$ ซึ่งเป็นฟังก์ชันของตัวแปรสุ่ม X แน่นอนว่า เราสามารถคำนวณหาค่าคาดหมาย (expectation) ของ Y ได้โดยเริ่มจากการหาการแจกแจงของ Y จากการแจกแจงของ X แล้วคำนวณหาค่าคาดหมาย (expectation) ของ Y โดยตรงจาก $\int yg(y)\,dy$ โดยที่ $g(y)$ คือฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y อย่างไรก็ตาม ดังจะเห็นได้จากการอภิปรายในหัวข้อที่ 3.7 การหาการแจกแจงของฟังก์ชันของตัวแปรสุ่มนั้นค่อนข้างยุ่งยาก หากต้องทำอย่างนั้นทุกครั้งก็คงทำให้การวิเคราะห์แต่ละครั้งต้องเสียเวลาอย่างมาก แต่ก็โชคดีที่เราไม่จำเป็นต้องทำอย่างนั้นเลยเพราะสามารถคำนวณหาค่าคาดหวังดังกล่าวได้โดยไม่ต้องเสียเวลาคำนวณหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y ดังสรุปในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 4.1 (กฎของนักสถิติที่ไม่รู้ตัว (Law of the Unconscious Statistician))**.** *กำหนดให้ X เป็นตัวแปรสุ่ม และ r เป็นฟังก์ชันค่าจริง ถ้า X มีการแจกแจงต่อเนื่อง (continuous distribution) และค่าคาดหมายมีค่า*

จำกัด แล้ว

$$E\left[r\left(X\right)\right]=\int_{-\infty}^{\infty}r\left(x\right)f\left(x\right)dx \tag{4.7}$$

ในทำนองเดียวกัน ถ้า X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) และค่าคาดหมายมีค่าจำกัด แล้ว

$$E\left[r\left(X\right)\right]=\sum_{x}r\left(x\right)f\left(x\right) \tag{4.8}$$

การพิสูจน์. เพื่อประหยัดพื้นที่ จึงขอแสดงเฉพาะการพิสูจน์ในกรณีของการแจกแจงต่อเนื่อง (continuous distribution) ดังนี้

$$E\left[r\left(X\right)\right]=\int_{-\infty}^{\infty}r\left(x\right)f\left(x\right)dx$$

กำหนดให้ $x=s\left(y\right)$ และ ประยุกต์ใช้เทคนิคการเปลี่ยนตัวแปรของการอินทิเกรต (change of variables) จะได้ว่า

$$E\left[r\left(X\right)\right]=\int_{-\infty}^{\infty}yf\left(s\left(y\right)\right)\left|\frac{ds\left(y\right)}{dy}\right|dy$$

ในขณะเดียวกัน ทฤษฎีบทที่ 3.26 ระบุว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ Y เท่ากับ

$$g\left(y\right)=f\left(s\left(y\right)\right)\left|\frac{ds\left(y\right)}{dy}\right|$$

ดังนั้น เราจึงสามารถสรุปได้ว่า

$$\int_{-\infty}^{\infty}r\left(x\right)f\left(x\right)dx=E\left[r\left(X\right)\right]=\int_{-\infty}^{\infty}yg\left(y\right)dy=E\left[Y\right]$$

■

ตัวอย่างที่ 4.7. สมมุติว่าเครื่องใช้ไฟฟ้ายี่ห้อหนึ่งมีสัดส่วนหรืออัตราการพัง X ต่อปี แต่นักวิเคราะห์ยังไม่ทราบว่า X มีค่าเท่าใด ดังนั้นจึงควรพิจารณาเป็นตัวแปรสุ่ม ในขณะเดียวกันสิ่งที่ต้องการทราบคือ เครื่องใช้ไฟฟ้าแต่ละชิ้นจะใช้ได้นานแค่ไหนจึงจะพัง?

คำถามแรกคือ ระยะเวลาที่ใช้ได้ก่อนที่เครื่องใช้ไฟฟ้าจะพัง Y สามารถเขียนในรูปฟังก์ชันของ X ได้อย่างไร? คำตอบก็คือ $Y=\frac{1}{X}$ แน่นอนว่า ไม่มีทางที่จะทราบค่าที่แน่นอนของ Y เนื่องจากไม่ทราบค่าที่แน่นอนของ X แต่อย่างน้อยก็สามารถหาค่าคาดหมายของ Y หากทราบการแจกแจงของ X

สมมุติให้ X มีการแจกแจงต่อเนื่องดังนี้

$$f\left(x\right)=\begin{cases}3x^{2}\text{ สำหรับ }0<x<1,\\0,\text{ สำหรับกรณีอื่น}\end{cases}$$

ดังนั้น ค่าคาดหมายของระยะเวลาที่ใช้ได้ก่อนที่เครื่องใช้ไฟฟ้าจะพัง Y เท่ากับ

$$E[Y] = \int_0^1 \frac{1}{x} 3x^2 dx = \int_0^1 3x dx = \frac{3}{2}$$

กล่าวคือ ค่าคาดหมาย (expectation) ของอายุการใช้งานของเครื่องใช้ไฟฟ้ามีค่าเท่ากับหนึ่งปีครึ่ง □

ทฤษฎีบทนี้สามารถขยายผลไปสู่กรณีที่มีตัวแปรสุ่มพื้นฐานมากกว่าหนึ่งตัวได้เช่นกัน แต่เพื่อประหยัดพื้นที่ขอไม่นำเสนอการพิสูจน์

ทฤษฎีบทที่ 4.2 (กฎของนักสถิติที่ไม่รู้ตัว (Law of the Unconscious Statistician))**.** *กำหนดให้ $\boldsymbol{X} = (X_1, \ldots, X_n)$ เป็นตัวแปรสุ่มที่มีการแจกแจงร่วม $f(\boldsymbol{x})$ และ r เป็นฟังก์ชันค่าจริง ถ้า $\boldsymbol{X}$ มีการแจกแจงต่อเนื่อง (continuous distribution) และค่าคาดหมายมีค่าจำกัด แล้ว*

$$E[r(\boldsymbol{X})] = \int \cdots \int_{\mathbb{R}^n} r(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n \tag{4.9}$$

ในทำนองเดียวกัน ถ้า $\boldsymbol{X}$ มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) และค่าคาดหมายมีค่าจำกัด แล้ว

$$E[r(\boldsymbol{X})] = \sum_{\boldsymbol{x}} (x_1, \ldots, x_n) f(x_1, \ldots, x_n) \tag{4.10}$$

ตัวอย่างที่ 4.8. พิจารณาการสุ่มเลือกจุด (X, Y) จากสี่เหลี่ยมจัตุรัส S ซึ่งครอบคลุมพื้นที่ระหว่าง $0 \leqslant x \leqslant 1$ และ $0 \leqslant y \leqslant 1$ ดังแสดงในรูปภาพที่ XXX คำถามก็คือ ค่าคาดหวังของ $X^2 + Y^2$ เท่ากับเท่าใด?

เนื่องจากเป็นการสุ่มเลือก (random) ดังนั้นแต่ละจุดย่อมมีโอกาสที่จะถูกเลือกเท่ากัน ซึ่งมีผลทำให้สรุปได้ว่า การแจกแจงของ X และ Y เป็นแบบเอกรูป (uniform distribution) และเนื่องจากพื้นที่ของ S เท่ากับหนึ่งหน่วย ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของ (X, Y) เท่ากับ

$$f(x, y) = \begin{cases} 1 \text{ สำหรับ } (x, y) \in S, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ค่าคาดหวังของ $X^2 + Y^2$ เท่ากับ

$$E\left[X^2 + Y^2\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x^2 + y^2\right) f(x, y) \, dxdy = \int_0^1 \int_0^1 \left(x^2 + y^2\right) dxdy = \frac{2}{3}$$

□

4.3 คุณสมบัติของค่าคาดหมาย (Properties of Expectation)

ทฤษฎีบทที่ 4.3. *ถ้า* $Y = aX + b$ *โดยที่* a *และ* b *เป็นค่าคงที่ แล้ว*

$$E[Y] = aE[X] + b \tag{4.11}$$

การพิสูจน์.

$$E[Y] = \int_{-\infty}^{\infty} (ax+b) f(x)\, dx = a\int_{-\infty}^{\infty} xf(x)\, dx + b\int_{-\infty}^{\infty} f(x)\, dx = aE[X] + b$$

■

ทฤษฎีบทที่ 4.4. *ถ้ามีค่าคงที่* a *ที่ทำให้* $Pr(X \geqslant a) = 1$ *แล้ว* $E[X] \geqslant a$ *ในทางกลับกัน ถ้ามีค่าคงที่* b *ที่ทำให้* $Pr(X \leqslant b) = 1$ *แล้ว* $E[X] \leqslant b$

การพิสูจน์. ก่อนอื่น เนื่องจาก $Pr(X \geqslant a) = 1$ ทำให้สามารถสรุปได้ว่า $Pr(X < a) = 0$ และสามารถสมมุติได้โดยไม่มีผลเสีย (without loss of generality) ว่า $f(x) = 0$ สำหรับ $x < a$

$$E[X] = \int_{-\infty}^{\infty} xf(x)\, dx = \int_{a}^{\infty} xf(x)\, dx \geqslant \int_{a}^{\infty} af(x)\, dx = aPr(X \geqslant a) = a$$

ส่วนการพิสูจน์ของอีกกรณีหนึ่งทำได้ในทำนองเดียวกัน ■

ทฤษฎีบทที่ 4.5. *ถ้า* $X = c$ *ด้วยความน่าจะเป็นเท่ากับหนึ่ง แล้ว* $E[X] = c$

ทฤษฎีบทที่ 4.6. *สมมุติให้* $E[X] = a$ *และกรณีใดกรณีหนึ่งระหว่าง* $Pr(X \geqslant a) = 1$ *หรือ* $Pr(X \leqslant a) = 1$ *เป็นจริง แล้ว* $Pr(X = a) = 1$

ทฤษฎีบทที่ 4.7. *ถ้า* $X_1, \ldots, X_n$ *คือตัวแปรสุ่ม* n *ตัว ที่ค่าคาดหมาย (expectation) ของแต่ละตัวมีค่าจำกัด และ* $a_1, \ldots, a_n$ *และ* b *คือค่าคงที่ แล้ว*

$$E\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i E[X_i] + b \tag{4.12}$$

การพิสูจน์.

$$\begin{aligned}
E\left[\sum_{i=1}^{n} a_i X_i + b\right] &= \int \cdots \int_{\mathbb{R}^n} \left[\sum_{i=1}^{n} a_i x_i + b\right] f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\
&= \sum_{i=1}^{n} a_i \int \cdots \int_{\mathbb{R}^n} x_i f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\
&\quad + b\int \cdots \int_{\mathbb{R}^n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\
&= \sum_{i=1}^{n} a_i \int_{-\infty}^{\infty} x_i f_i(x_i)\, dx_i + b = \sum_{i=1}^{n} a_i E[X_i] + b
\end{aligned}$$

■

ตัวอย่างที่ 4.9. กำหนดให้ R_i แทนอัตราผลตอบแทนรวม (gross return) ต่อปีของกองทุนรวม (mutual fund) i สำหรับ $i = 1, \ldots, n$ กล่าวคือ ลงทุน 1 บาทในกองทุนรวม i จะได้กลับมา R_i บาทหลังจากผ่านไปหนึ่งปี แต่ไม่มีใครทราบว่า R_i จริงๆ แล้วเท่าไหร่ เราจึงต้องพิจารณาเป็นตัวแปรสุ่ม นอกจากนี้ กำหนดให้ $\alpha_i \in [0, 1]$ โดยที่ $\sum_i \alpha_i = 1$ แทนสัดส่วนการลงทุนในกองทุน i กล่าวคือ หากมีเงินลงทุนทั้งหมด $10,000$ บาท จะลงทุนในกองทุน i ทั้งหมด $10,000\alpha_i$ คำถามก็คือ นักลงทุนควรจะคาดหมายที่จะได้รับอัตราผลตอบแทนรวมจากการลงทุนเป็นเท่าไหร่หลังจากลงทุนเป็นเวลาหนึ่งปี?

เริ่มจากการสร้างตัวแปรสุ่มที่เรียกว่าอัตราผลตอบแทนรวมจากการลงทุน ซึ่งหาได้จากจำนวนเงินที่ได้คืนทั้งหมดหารด้วยเงินลงทุน นั่นคือ

$$R = \sum_{i=1}^{n} \frac{10,000\alpha_i R_i}{10,000} = \sum_{i=1}^{n} \alpha_i R_i$$

บทเรียนหนึ่งที่ได้จากส่วนนี้คือ อัตราผลตอบแทนรวมจากการลงทุนไม่ขึ้นอยู่กับจำนวนเงินลงทุน นั่นคือ อัตราผลตอบแทนรวมจากการลงทุนมีความสัมพันธ์เชิงเส้นกับอัตราผลตอบแทนรวมของแต่ละกองทุนรวม ดังนั้น จึงสามารถประยุกต์ใช้ทฤษฎีบทที่ 4.7 เพื่อหาค่าคาดหมาย (expectation) ของอัตราผลตอบแทนรวมจากการลงทุนได้เป็น

$$E[R] = E\left[\sum_{i=1}^{n} \alpha_i R_i\right] = \sum_{i=1}^{n} \alpha_i E[R_i]$$

ส่วนค่าคาดหมาย (expectation) ของอัตราผลตอบแทนสุทธิจากการลงทุน (return on investment: ROI) เท่ากับ

$$E[r] = 1 - E[R] = \sum_{i=1}^{n} \alpha_i E[1 - R_i]$$

เพื่อให้เห็นความเชื่อมโยงกับการปฏิบัติมากขึ้น ขอยกตัวอย่างกองทุนรวม 5 กองทุน ซึ่งประกอบไปด้วย XXX รูปภาพที่ XXX แสดงมูลค่าหน่วยลงทุนสุทธิ (Net Asset Value: NAV) รายวันของกองทุนรวมแต่ละกองทุน ในที่นี้จะประมาณค่าคาดหมาย (expectation) ของอัตราผลตอบแทนรวมของแต่ละกองทุนรวมด้วยค่าเฉลี่ยของอัตราผลตอบแทนรวมของกองทุนนั้นโดยใช้ข้อมูลในอดีต ส่วนเหตุผลที่ว่าทำไมจึงควรใช้ค่าเฉลี่ยในการประมาณการณ์ค่าคาดหมายนั้นจะอภิปรายในหัวข้อ XXX ต่อไป แต่ขอให้เข้าใจไว้ก่อนว่า ค่าเฉลี่ยคือวิธีการประมาณค่าคาดหมายอย่างหนึ่ง แต่ไม่ใช่สิ่งเดียวกัน เพราะค่าคาดหมายควรจะเป็นค่าที่แท้จริง (true value) ที่เป็นคุณสมบัติของการแจกแจงหรือตัวแปรสุ่ม ส่วนค่าเฉลี่ยเป็นเพียงตัวประมาณค่า (estimator) ของค่าคาดหมายตัวหนึ่งเท่านั้น

ADD FIGURE FROM FIN4FUTURE HERE □

ตัวอย่างที่ 4.10 (จาก DeGroot and Schervish (2012)). สมมุติว่า ผู้หญิงคนหนึ่งพิมพ์จดหมาย n ฉบับ และที่อยู่หน้าซองจดหมาย n ซอง แล้วนำจดหมายใส่ซองด้วยวิธีการสุ่ม (random) คำถามก็คือ ค่าคาดหมาย (expectation) ของจำนวนจดหมายและซองที่ตรงกันพอดีมีค่าเท่ากับเท่าไหร่?

กำหนดให้ X_i แทนตัวแปรสุ่มที่ระบุว่าซองที่ i ได้รับจดหมายที่ตรงกันหรือไม่ นั่นคือ $X_i = 1$ ถ้าซองจดหมายที่ i ตรงกันกับจดหมาย และ $X_i = 0$ ถ้าซองจดหมายที่ i ไม่ตรงกันกับจดหมาย ดังนั้น จำนวนจดหมายและซองที่ตรงกันพอดีเท่ากับ $Y = \sum_{i=1}^{n} X_i$ ซึ่งสามารถหาค่าคาดหมาย (expectation) ได้โดยประยุกต์ใช้ทฤษฎีบทที่ 4.7 ดังนี้

$$E\left[Y\right] = \sum_{i=1}^{n} E\left[X_i\right]$$

สิ่งที่ต้องการในขั้นต่อไปคือ ค่าคาดหมาย (expectation) ของ X_i ซึ่งมีค่าเท่ากับ

$$E\left[X_i\right] = (1) \times Pr\left(X_i = 1\right) + (0) \times Pr\left(X_i = 0\right) = Pr\left(X_i = 1\right)$$

ซึ่งมีค่าเท่ากับ $\frac{1}{n}$ เพราะการจัดซองจดหมายกับจดหมายเป็นแบบสุ่ม (random) ดังนั้น คำตอบเท่ากับ

$$E\left[Y\right] = \sum_{i=1}^{n} E\left[X_i\right] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

นั่นคือ ไม่ว่าจะมีจดหมายและซองจดหมายกี่อัน ค่าคาดหมาย (expectation) ของจำนวนจดหมายและซองที่ตรงกันพอดีจะมีค่าเท่ากับหนึ่งเสมอ □

ที่ผ่านมา จะเห็นได้ว่าหากฟังก์ชันเป็นแบบเชิงเส้น (linear) แล้วค่าคาดหมายของฟังก์ชันจะเท่ากับฟังก์ชันของค่าคาดหมาย อย่างไรก็ตาม ข้อสรุปนี้ไม่สามารถใช้ได้กับฟังก์ชันที่ไม่ใช่เชิงเส้น แน่นอนว่า อาจจะไม่สามารถมีข้อสรุปใดๆ ได้เลยในกรณีที่เป็นฟังก์ชันทั่วไป แต่อย่างน้อย ก็ยังสามารถสร้างข้อสรุปบางอย่างที่มีประโยชน์สำหรับฟังก์ชันบางกลุ่มที่มีลักษณะเฉพาะ ยกตัวอย่างเช่น ฟังก์ชันเว้า (concave function) หรือฟังก์ชันนูน (convex function) ซึ่งนำไปสู่ทฤษฎีบทที่สำคัญชื่อ อสมการเจนเซ็น (Jensen's inequality) ซึ่งมีประโยชน์ต่อวงการเศรษฐศาสตร์ การเงิน และการประกันภัยอย่างมาก

บทนิยามที่ 4.2. *ฟังก์ชัน g เป็นฟังก์ชันเว้า (concave function) ถ้าสำหรับทุกๆ ค่า $\alpha \in (0,1)$ และทุกๆ เวคเตอร์ $\boldsymbol{x}$ และ $\boldsymbol{y}$*

$$g\left(\alpha\boldsymbol{x} + \left[1-\alpha\right]\boldsymbol{y}\right) \leqslant \alpha g\left(\boldsymbol{x}\right) + \left[1-\alpha\right] g\left(\boldsymbol{y}\right) \tag{4.13}$$

ในทางกลับกัน ฟังก์ชัน g เป็นฟังก์ชันนูน (convex function) ถ้าสำหรับทุกๆ ค่า $\alpha \in (0,1)$ และทุกๆ เวคเตอร์ $\boldsymbol{x}$ และ $\boldsymbol{y}$

$$g\left(\alpha\boldsymbol{x} + \left[1-\alpha\right]\boldsymbol{y}\right) \geqslant \alpha g\left(\boldsymbol{x}\right) + \left[1-\alpha\right] g\left(\boldsymbol{y}\right) \tag{4.14}$$

รูปที่ XXX และ XXX แสดงตัวอย่างของฟังก์ชันเว้า (concave) และนูน (convex) ตามลำดับ

ทฤษฎีบทที่ 4.8 (อสมการเจนเซ็น (Jensen's Inequality))**.** *กำหนดให้ X แทนเวคเตอร์ของตัวแปรสุ่มที่ค่าคาดหมายมีค่าจำกัด ถ้า g เป็นฟังก์ชันเว้า (concave function) แล้ว*

$$E\left[g\left(X\right)\right] \leqslant g\left(E\left[X\right]\right) \tag{4.15}$$

ในทางกลับกัน ถ้า g เป็นฟังก์ชันนูน (convex function) แล้ว

$$E\left[g\left(X\right)\right] \geqslant g\left(E\left[X\right]\right) \tag{4.16}$$

ตัวอย่างต่อไปนี้แสดงการประยุกต์ใช้ ในทางเศรษฐศาสตร์ซึ่งเกี่ยวข้องกับการประกันภัยด้วย

ตัวอย่างที่ 4.11. กำหนดให้ $U\left(x\right)$ แทนฟังก์ชันอรรถประโยชน์ (utility function) ซึ่งแสดงถึงคุณค่าหรืออรรถประโยชน์ที่ได้รับจาก x และที่สำคัญ ฟังก์ชันอรรถประโยชน์ (utility function) มักถูกใช้เป็นเครื่องมือในการตัดสินใจ โดยสมมุติว่า ผู้มีอำนาจในการตัดสินใจ (decision maker) จะเลือกทางเลือกที่นำไปสู่ค่าอรรถประโยชน์สูงที่สุด เพื่อให้สามารถเข้าใจได้ง่ายขึ้น ขอพิจารณากรณีของการตัดสินใจซื้อประกันรถยนต์ (car insurance) เป็นกรณีตัวอย่างในที่นี้

ประเด็นสำคัญของการประกันภัย (insurance) คือผลลัพธ์ที่ไม่แน่นอน ขึ้นอยู่กับเหตุการณ์ที่เกิดขึ้นจากการขับขี่รถยนต์ □

ถึงแม้ว่าโดยทั่วไป ค่าคาดหมายของฟังก์ชันจะไม่เท่ากับฟังก์ชันของค่าคาดหมาย แต่ก็มีบางกรณีที่เป็นไปได้ ซึ่งรวมถึงกรณีที่ฟังก์ชันเป็นแบบเชิงเส้นซึ่งกล่าวถึงแล้วก่อนหน้านี้ และกรณีของฟังก์ชันที่เป็นผลคูณของตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) ซึ่งจะกล่าวในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 4.9. *ถ้า $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) และมีค่าคาดหมาย (expectation) ที่จำกัด แล้ว*

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E\left[X_i\right] \tag{4.17}$$

การพิสูจน์. เพื่อประหยัดพื้นที่จะขอนำเสนอการพิสูจน์ในกรณีของตัวแปรสุ่มต่อเนื่องเท่านั้น ส่วนกรณีของตัวแปรสุ่มไม่ต่อเนื่องสามารถพิสูจน์ได้โดยใช้ขั้นตอนที่คล้ายคลึงกับที่ใช้ในที่นี้

กำหนดให้ $f\left(x_1, \ldots, x_n\right)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) ของ $\left(X_1, \ldots, X_n\right)$ และ $f_i\left(x_i\right)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ X_i เนื่องจาก$X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) ดังนั้น

$$f\left(x_1, \ldots, x_n\right) = \prod_{i=1}^{n} f_i\left(x_i\right)$$

ทำให้สามารถหาค่าคาดหมายได้เป็น

$$\begin{aligned} E\left[\prod_{i=1}^{n} X_i\right] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i\right] f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i\right] \left[\prod_{i=1}^{n} f_i(x_i)\right] dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i f_i(x_i)\right] dx_1 \cdots dx_n \\ &= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i f_i(x_i)\, dx_i = \prod_{i=1}^{n} E[X_i] \end{aligned}$$

■

ตัวอย่างที่ 4.12. พิจารณากระบวนการกรองน้ำวิธีหนึ่งที่สามารถกรองสิ่งเจือปนออกไปได้ แต่ก็ไม่สามารถบอกได้อย่างแน่นอนร้อยเปอร์เซ็นต์ว่ากรองออกไปได้เท่าใด จึงต้องพิจารณาว่าเป็นตัวแปรสุ่ม สมมุติให้ตัวแปรสุ่ม X_1 คือสัดส่วนของสิ่งเจือปนที่ถูกกรองออกไปได้ในการกรองครั้งที่หนึ่ง และ X_2 คือสัดส่วนของสิ่งเจือปนที่ถูกกรองออกไปได้ในการกรองครั้งที่สอง ยิ่งไปกว่านั้น สมมุติว่ากระบวนการกรองทั้งสองครั้งเป็นอิสระต่อกัน (independent) ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เหมือนกันคือ

$$f(x) = \begin{cases} 4x^3 \text{ สำหรับ } 0 < x < 1, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

กำหนดให้ Y แทนสัดส่วนของสิ่งเจือปนที่เหลืออยู่หลังจากการกรองแล้วสองรอบ ดังนั้น $Y = (1 - X_1)(1 - X_2)$ เนื่องจาก X_1 และ X_2 เป็นอิสระต่อกัน (independent) จึงสรุปได้ว่า $1 - X_1$ และ $1 - X_2$ เป็นอิสระต่อกัน (independent) ดังนั้น ค่าคาดหมายของสัดส่วนของสิ่งเจือปนที่เหลืออยู่หลังจากการกรองแล้วสองรอบเท่ากับ

$$E[Y] = E[(1 - X_1)(1 - X_2)] = E[(1 - X_1)]\, E[(1 - X_2)] = (1 - E[X_1])(1 - E[X_2])$$

ในขณะกัน $E[X_i] = \int_0^1 x 4x^3 dx = \frac{4}{5}$ ดังนั้น

$$E[Y] = (1 - E[X_1])(1 - E[X_2]) = \left(1 - \frac{4}{5}\right)^2 = 0.04$$

□

ทฤษฎีบทที่ 4.10. *กำหนดให้ X เป็นตัวแปรสุ่มที่มีค่าเป็นจำนวนธรรมชาติ (natural numbers) $0, 1, \ldots$ แล้ว*

$$E[X] = \sum_{n=1}^{\infty} Pr(X \geqslant n) \tag{4.18}$$

ตัวอย่างที่ 4.13 (จาก DeGroot and Schervish (2012)). สมมุติว่าผู้ชายคนหนึ่งพยายามที่จะโยนเหรียญให้ลงในขวดไปเรื่อยจนกว่าจะสำเร็จ สมมุติอีกด้วยว่าความน่าจะเป็นที่จะสำเร็จในแต่ละครั้งเท่ากับ p และการโยนแต่ละครั้งเป็นอิสระต่อกัน (independent) ถ้า X แทนจำนวนครั้งที่โยนจนสำเร็จเป็นครั้งแรก คำถามก็คือ ค่าคาดหมายของจำนวนครั้งการโยนดังกล่าวเท่ากับเท่าใด?

เนื่องจากเพื่อให้สำเร็จเขาจำเป็นต้องโยนอย่างน้อยหนึ่งครั้ง ดังนั้น $Pr(X \geqslant 1) = 1$ ส่วนการที่ประสบความสำเร็จครั้งแรกในการโยนครั้งที่ n ย่อมหมายความว่า การโยน $n-1$ ครั้งก่อนหน้านั้นไม่ประสบความสำเร็จ ดังนั้น

$$Pr(X \geqslant n) = (1-p)^{n-1}$$

เมื่อประยุกต์ใช้ทฤษฎีบทที่ 4.10 จะได้ว่า ค่าคาดหมายของจำนวนครั้งการโยนจนสำเร็จเป็นครั้งแรกเท่ากับ

$$E[X] = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{1-(1-p)} = \frac{1}{p} \tag{4.19}$$

□

ทฤษฎีบทที่ 4.11. *กำหนดให้ X แทนตัวแปรสุ่มที่มีค่าไม่เป็นลบ (nonnegative random variable) ที่มีฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เท่ากับ F แล้ว*

$$E[X] = \int_0^{\infty} [1 - F(x)]\, dx \tag{4.20}$$

ตัวอย่างที่ 4.14. สมมุติว่า X คือเวลาที่ลูกค้ารอในการรับบริการ ซึ่งมีฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เท่ากับ

$$F(x) = \begin{cases} 1 - e^{-2x} \text{ สำหรับ } 0 < x, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ดังนั้น ประยุกต์ใช้ทฤษฎีบทที่ 4.11 จะได้ว่า ค่าคาดหมายของเวลาที่ต้องรอรับบริการเท่ากับ

$$E[X] = \int_0^{\infty} e^{-2x} dx = \frac{1}{2}$$

□

4.4 ความแปรปรวน (Variance)

ความแปรปรวน (variance) เป็นเครื่องมือที่นิยมใช้เพื่อบอกถึงลักษณะการกระจายตัวของการแจกแจง ซึ่งเป็นคุณสมบัติที่สำคัญอีกอย่างหนึ่งของการแจกแจงที่จำเป็นต้องทราบ ทั้งนี้เนื่องจากการแจกแจงที่แตกต่างกันอาจจะ

มีค่าคาดหมายที่เท่ากันก็เป็นได้ และที่สำคัญความแปรปรวนยังสะท้อนถึงระดับความไม่มีแบบแผน (randomness) ของตัวแปรสุ่มนั้นอีกด้วย

ในทางเทคนิค ความแปรปรวน (variance) ก็คือค่าคาดหมายของฟังก์ชันของตัวแปรสุ่มอย่างหนึ่ง ซึ่งนิยามได้ดังต่อไปนี้

บทนิยามที่ 4.3. *กำหนดให้ X เป็นตัวแปรสุ่มที่ค่าคาดหมายมีค่าจำกัดเท่ากับ $\mu = E[X]$ ความแปรปรวน (variance) ของ X นิยามได้เป็น*

$$Var[X] = E\left[(X-\mu)^2\right] \tag{4.21}$$

ในกรณีที่ความแปรปรวนหาค่าได้ เราจะเรียกรากที่สอง (square root) ของความแปรปรวน $\sqrt{Var[X]}$ ว่า ค่าเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งมักแทนด้วยสัญลักษณ์ σ_X

สังเกตว่า หากค่าคาดหมายหาค่าไม่ได้ (does not exist) แล้วค่าความแปรปรวนจะหาค่าไม่ได้ตามไปด้วย เพราะไม่สามารถนำ X ไปลบกับจำนวนใด นอกจากนี้ ความแปรปรวน (variance) ขึ้นอยู่กับการแจกแจงเท่านั้น กล่าวคือ ตัวแปรสุ่มต่างกันแต่มีการแจกแจงแบบเดียวกันย่อมมีค่าความแปรปรวน (variance) เท่ากัน เช่นเดียวกับกรณีของค่าคาดหมาย

ทฤษฎีบทที่ 4.12. *สำหรับตัวแปรสุ่ม X ใดๆ*

$$Var[X] = E\left[X^2\right] - (E[X])^2 \tag{4.22}$$

การพิสูจน์. กำหนดให้ $\mu = E[X]$ ดังนั้น

$$\begin{aligned} Var[X] &= E\left[(X-\mu)^2\right] = E\left[X^2 - 2X\mu + \mu^2\right] \\ &= E\left[X^2\right] - 2E[X]\mu + \mu^2 = E\left[X^2\right] - \mu^2 = E\left[X^2\right] - (E[X])^2 \end{aligned}$$

■

ทฤษฎีบทที่ 4.13. *กำหนดให้ X เป็นตัวแปรสุ่ม ดังนั้น $Var[X] \geqslant 0$ และถ้า X เป็นตัวแปรสุ่มที่มีขอบเขต (bounded) แล้ว $Var[X]$ มีค่าจำกัด*

ทฤษฎีบทที่ 4.14. *กำหนดให้ X เป็นตัวแปรสุ่ม ดังนั้น $Var[X] = 0$ ก็ต่อเมื่อ มีค่าคงที่ c ที่ทำให้ $Pr(X = c) = 1$*

การพิสูจน์. กรณีแรก เริ่มจากการสมมุติว่า $Pr(X = c) = 1$ แล้ว $E[X] = c$ และ $Var[X] = 0$

กรณีที่สอง เริ่มจากการสมมุติว่า $Var[X] = 0$ ซึ่งหมายความว่า

$$E\left[(X-\mu)^2\right] = 0$$

ในขณะเดียวกัน เราก็ทราบดีว่า $Pr\left([X-\mu]^2 \geqslant 0\right) = 1$ ดังนั้น ประยุกต์ใช้ทฤษฎีบทที่ 4.6 ทำให้สรุปได้ว่า $Pr\left([X-\mu]^2 = 0\right) = 1$ ซึ่งสมมูล (equivalent) กับข้อสรุปที่ว่า $Pr(X = \mu) = 1$ โดยค่าคงที่ในที่นี้คือค่า $c = \mu$ นั่นเอง ■

ทฤษฎีบทที่ 4.15. *สำหรับค่าคงที่ a และ b ใดๆ กำหนดให้ตัวแปรสุ่ม $Y = aX + b$ แล้ว*

$$Var[Y] = a^2 Var[X] \tag{4.23}$$

และ $\sigma_Y = |a|\sigma_X$

ทฤษฎีบทที่ 4.16. *ถ้า $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) และมีค่าคาดหมาย (expectation) ที่จำกัด และ $a_1, \ldots, a_n$ และ b คือค่าคงที่ แล้ว*

$$Var\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 Var[X_i] \tag{4.24}$$

การพิสูจน์. กำหนดให้ $\mu_i = E[X_i]$ ดังนั้น

$$E\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i \mu_i + b$$

และ

$$\begin{aligned}
Var\left[\sum_{i=1}^{n} a_i X_i + b\right] &= E\left[\left(\left(\sum_{i=1}^{n} a_i X_i + b\right) - \left(\sum_{i=1}^{n} a_i \mu_i + b\right)\right)^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right] \\
&= \sum_{i=1}^{n} E\left[(a_i (X_i - \mu_i))^2\right] + \sum_{i \neq j} a_i a_j E\left[(X_i - \mu_i)(X_j - \mu_j)\right]
\end{aligned}$$

เพื่อให้ได้คำตอบที่ต้องการจะต้องพิสูจน์ว่า $E[(X_i - \mu_i)(X_j - \mu_j)] = 0$ สำหรับทุกๆ $i \neq j$ ดังนี้

$$E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i - \mu_i]\, E[X_j - \mu_j] = 0$$

โดยที่สมการแรกเป็นผลมาจากการที่ $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) ดังนั้น จึงสามารถสรุปได้ว่า

$$Var\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 Var\left[X_i\right]$$

■

ในกรณีที่ค่าความแปรปรวน (variance) หาค่าไม่ได้ ก็ยังพอมีเครื่องมือที่สามารถบอกถึงการกระจายตัวของการแจกแจง แต่เป็นเครื่องมือที่ไม่เกี่ยวข้องกับค่าคาดหมาย แต่คำนวณจากค่าควอนไทล์ (quantile) ซึ่งหาค่าได้เสมอ ตัวอย่างหนึ่งที่สำคัญคือช่วงระหว่างควอไทล์ (interquartile range) ซึ่งเท่ากับความแตกต่างระหว่างค่าควอไทล์ที่ 0.75 กับควอไทล์ที่ 0.25 ดังแสดงในนิยามต่อไปนี้

บทนิยามที่ 4.4. *กำหนดให้ X เป็นตัวแปรสุ่มที่มีฟังก์ชันควอไทล์ (quartile function) เท่ากับ $F^{-1}(p)$ สำหรับ $0 < p < 1$ ค่าช่วงระหว่างควอไทล์ (interquartile range) มีค่าเท่ากับ $F^{-1}(0.75) - F^{-1}(0.25)$*

ตัวอย่างที่ 4.15. พิจารณาตัวแปรสุ่มที่มีการแจกแจงโคชี (Cauchy distribution)

$$f(x) = \frac{1}{\pi\left(1 + x^2\right)}$$

จากตัวอย่างที่ 4.6 จะเห็นได้ว่าค่าคาดหมาย (mean) ของตัวแปรสุ่ม X หาค่าไม่ได้ ทำให้ค่าความแปรปรวน (variance) ก็หาค่าไม่ได้เช่นกัน แต่สามารถหาค่าช่วงระหว่างควอไทล์ (interquartile range) ได้ดังต่อไปนี้

เริ่มจากการหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.)

$$F(x) = \int_{-\infty}^{x} \frac{1}{\pi\left(1 + y^2\right)} dy = \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi}$$

หลังจากนั้นจึงใช้ความสัมพันธ์ที่ว่า $p = F(x)$ เพื่อคำนวณหาฟังก์ชันควอไทล์ (quartile function)

$$F^{-1}(p) = \tan\left(\pi\left(p - \frac{1}{2}\right)\right)$$

ดังนั้น ค่าช่วงระหว่างควอไทล์ (interquartile range) มีค่าเท่ากับ □

$$IQR(X) = \tan\left(\pi\left(0.75 - \frac{1}{2}\right)\right) - \tan\left(\pi\left(0.25 - \frac{1}{2}\right)\right) = \tan\left(\frac{\pi}{4}\right) - \tan\left(-\frac{\pi}{4}\right) = 2$$

ซึ่งมีค่าจำกัด (finite)

4.5 ความแปรปรวนร่วมและสหสัมพันธ์ (Covariance and Correlation)

ความสัมพันธ์ระหว่างตัวแปรเป็นสิ่งที่นักวิเคราะห์ต้องการทราบ หัวข้อนี้เป็นความพยายามขั้นแรกๆ ที่จะแสดงความสัมพันธ์ของตัวแปรสุ่ม โดยนำเสนอในรูปแบบของ ความแปรปรวนร่วม (covariance) และสหสัมพันธ์ (correlation) ถึงแม้อาจจะไม่ใช่วิธีที่ดีที่สุด แต่ก็สะดวกต่อการคำนวณและสามารถเข้าใจและสื่อสารได้ง่าย

บทนิยามที่ 4.5. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่ค่าคาดหมายมีค่าจำกัด และ $E[X] = \mu_X$ และ $E[Y] = \mu_Y$ ความแปรปรวนร่วม (covariance) ของ X และ Y นิยามได้เป็น*

$$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] \tag{4.25}$$

ถ้าค่าคาดหมายมีค่าจำกัด

สังเกตได้ว่า ความแปรปรวนร่วม (covariance) ก็คือค่าคาดหมายของฟังก์ชันของตัวแปรสุ่มอย่างหนึ่งนั่นเอง

ทฤษฎีบทที่ 4.17. *สำหรับตัวแปรสุ่ม X และ Y ที่ความแปรปรวนร่วม (covariance) มีค่าจำกัดใดๆ*

$$Cov[X,Y] = E[XY] - E[X]E[Y] \tag{4.26}$$

การพิสูจน์. กำหนดให้ $E[X] = \mu_X$ และ $E[Y] = \mu_Y$

$$\begin{aligned} Cov[X,Y] &= E[(X-\mu_X)(Y-\mu_Y)] = E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y = E[XY] - \mu_X\mu_Y = E[XY] - E[X]E[Y] \end{aligned}$$

■

ข้อจำกัดอย่างหนึ่งคือการที่ขนาดของความแปรปรวนร่วม (covariance) ขึ้นอยู่กับขนาดของแต่ละตัวแปร ซึ่งทำให้ยากที่จะบอกได้ว่าตัวแปรคู่ใดมีความสัมพันธ์ระหว่างกัน (association) มากกว่ากัน โดยการเปรียบเทียบค่าความแปรปรวนร่วม (covariance) ดังนั้น จึงจำเป็นต้องนิยามค่าสถิติที่บอกถึงความสัมพันธ์แต่ไม่ขึ้นอยู่กับขนาดของแต่ละตัวแปรโดยตรง หรือมีการปรับขนาด (re-scaling) นั่นเอง

บทนิยามที่ 4.6. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่ความแปรปรวน (variance) มีค่าจำกัด นั่นคือ $\sigma_X^2 < \infty$ และ $\sigma_Y^2 < \infty$ สหสัมพันธ์ (correlation) ของ X และ Y นิยามได้เป็น*

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sigma_X\sigma_Y} \tag{4.27}$$

ทฤษฎีบทต่อไปนี้มีความสำคัญต่อคุณสมบัติที่สำคัญอันหนึ่งของสหสัมพันธ์ (correlation) นั่นคือ ค่าสมบูรณ์ของสหสัมพันธ์มีค่าน้อยกว่าหรือเท่ากับหนึ่งเสมอ

ทฤษฎีบทที่ 4.18. *สำหรับตัวแปรสุ่ม X และ Y ใดๆ ที่ $E[XY]$ หาค่าได้*

$$(E[XY])^2 \leqslant E\left[X^2\right] E\left[Y^2\right] \tag{4.28}$$

ความสัมพันธ์นี้เป็นจริงในรูปแบบสมการก็ต่อเมื่อมีจำนวนจริงที่ไม่เท่ากับศูนย์ $a \neq 0$ และ $b \neq 0$ ที่ทำให้ $aX + bY = 0$ ด้วยความน่าจะเป็นเท่ากับหนึ่ง

การพิสูจน์. สำหรับกรณีที่ $E[X^2] = \infty$ หรือ $E[Y^2] = \infty$ ไม่จำเป็นต้องพิสูจน์อะไรมากเพราะพจน์ด้านขวาเท่ากับอนันต์ดังนั้นอสมการที่ 4.28 เป็นจริง ส่วนในกรณีที่ $E[X^2] = 0$ นั้นสามารถพิสูจน์ได้โดยเริ่มจาก $Pr(X = 0) = 1$ ดังนั้น $E[XY] = 0$ ซึ่งมีผลทำให้อสมการที่ 4.28 เป็นจริงเช่นกัน ส่วนกรณีที่ $E[Y^2] = 0$ ก็สามารถพิสูจน์ได้ในทำนองเดียวกัน

กรณีสุดท้ายที่ต้องพิสูจน์คือ กรณีที่ $0 < E[X^2] < \infty$ และ $0 < E[Y^2] < \infty$ ซึ่งสามารถพิสูจน์ได้โดยเริ่มจากการพิจารณาความสัมพันธ์ต่อไปนี้ สำหรับค่าจำนวนจริงที่ไม่เท่ากับศูนย์ a, b ใดๆ

$$E\left[(aX + bY)^2\right] = a^2 E\left[X^2\right] + b^2 E\left[Y^2\right] + 2abE[XY] \geqslant 0 \tag{4.29}$$

$$E\left[(aX - bY)^2\right] = a^2 E\left[X^2\right] + b^2 E\left[Y^2\right] - 2abE[XY] \geqslant 0 \tag{4.30}$$

หากเลือก $a = \sqrt{E[Y^2]}$ และ $b = \sqrt{E[X^2]}$ จะสามารถเขียนสมการที่ 4.29 และ 4.30 ตามลำดับ ได้ใหม่เป็น

$$E[XY] \geqslant -\sqrt{E[X^2] E[Y^2]} \tag{4.31}$$

$$E[XY] \leqslant \sqrt{E[X^2] E[Y^2]} \tag{4.32}$$

ซึ่งทำให้สามารถสรุปได้ว่า

$$(E[XY])^2 \leqslant E\left[X^2\right] E\left[Y^2\right] \tag{4.33}$$

ในส่วนของสมการนั้นจะเห็นได้ว่า อสมการที่ 4.28 เป็นจริงในรูปของสมการก็ต่อเมื่อสมการที่ 4.29 และ 4.30 เป็นจริงในรูปของสมการ ซึ่งหมายความว่า

$$E\left[(aX + bY)^2\right] = 0 \Rightarrow Pr(aX + bY = 0) = 1$$

และ

$$E\left[(aX - bY)^2\right] = 0 \Rightarrow Pr(aX - bY = 0) = 1$$

■

ทฤษฎีบทที่ 4.19 (อสมการโคชีและชวาร์ซ (Cauchy-Schwarz Inequality))**.** *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่ความแปรปรวน (variance) มีค่าจำกัด นั่นคือ $\sigma_X^2 < \infty$ และ $\sigma_Y^2 < \infty$ สหสัมพันธ์ (correlation) ของ X และ Y มีค่าอยู่ระหว่าง $[-1, 1]$ นั่นคือ*

$$-1 \leqslant \rho[X, Y] \leqslant 1 \tag{4.34}$$

และความแปรปรวนร่วม (covariance) ของ X และ Y มีความสัมพันธ์กับความแปรปรวน (variance) ของแต่ละตัวแปรสุ่มดังนี้

$$(Cov[X, Y])^2 \leqslant \sigma_X^2 \sigma_Y^2 \tag{4.35}$$

โดยที่ อสมการที่ จะเป็นจริงในรูปแบบสมการก็ต่อเมื่อ มีจำนวนจริงที่ไม่เท่ากับศูนย์ $a, b \neq 0$ และจำนวนจริง c ที่ทำให้ $aX + bY = c$ ด้วยความน่าจะเป็นเท่ากับ 1

การพิสูจน์. ประยุกต์ใช้ทฤษฎีบทที่ 4.18 เพื่อแสดงว่า

$$(Cov[X, Y])^2 = (E[(X - \mu_X)(Y - \mu_Y)])^2 \leqslant E\left[(X - \mu_X)^2\right] E\left[(Y - \mu_Y)^2\right] = \sigma_X^2 \sigma_Y^2$$

ซึ่งจะเป็นจริงในรูปแบบสมการก็ต่อเมื่อมีจำนวนจริงที่ไม่เท่ากับศูนย์ $a \neq 0$ และ $b \neq 0$ ที่ทำให้ $aX + bY = 0$ ด้วยความน่าจะเป็นเท่ากับหนึ่ง เช่นเดียวกับกรณีของทฤษฎีบทที่ 4.18 ยิ่งไปกว่านั้นความสัมพันธ์ที่ได้นำไปสู่ข้อสรุปที่ว่า

$$-\sigma_X \sigma_Y \leqslant Cov[X, Y] \leqslant \sigma_X \sigma_Y \Rightarrow -1 \leqslant \frac{Cov[X, Y]}{\sigma_X \sigma_Y} \leqslant 1 \Rightarrow -1 \leqslant \rho[X, Y] \leqslant 1$$

■

บทนิยามที่ 4.7. *ตัวแปรสุ่ม X และ Y มีสหสัมพันธ์เชิงบวก (positively correlated) ถ้า $\rho[X, Y] > 0$ และมีสหสัมพันธ์เชิงลบ (positively correlated) ถ้า $\rho[X, Y] < 0$ ส่วนกรณีที่ $\rho[X, Y] = 0$ หมายความว่า X และ Y ไม่มีสหสัมพันธ์ (uncorrelated)*

ทฤษฎีบทต่อไปนี้บอกถึงความสัมพันธ์ระหว่างความเป็นอิสระต่อกัน (independent) และการไม่มีสหสัมพันธ์ (uncorrelated) ซึ่งกล่าวว่า ตัวแปรสุ่มที่เป็นอิสระต่อกันต้องไม่มีสหสัมพันธ์

ทฤษฎีบทที่ 4.20. *ถ้าตัวแปรสุ่ม X และ Y เป็นอิสระต่อกัน (independent) และ $\sigma_X^2 < \infty$ และ $\sigma_Y^2 < \infty$ แล้ว*

$$Cov[X, Y] = \rho[X, Y] = 0 \tag{4.36}$$

การพิสูจน์. เนื่องจาก X และ Y เป็นอิสระต่อกัน (independent) ดังนั้น $E[XY] = E[X]E[Y]$ ซึ่งหมายความว่า $E[XY] - E[X]E[Y] = Cov[X,Y] = 0$ ผลที่ตามมาอีกอย่างหนึ่งก็คือ $\rho[X,Y] = 0$ ■

ในทางกลับกัน ตัวแปรสุ่มที่ไม่มีสหสัมพันธ์อาจจะไม่เป็นอิสระต่อกันก็ได้ ซึ่งสะท้อนให้เห็นว่า การเป็นอิสระต่อกัน (independent) นั้นเป็นข้อจำกัดที่เข้มข้นมากกว่าการไม่มีสหสัมพันธ์ (uncorrelated)

ตัวอย่างที่ 4.16. กำหนดให้ X เป็นตัวแปรสุ่มที่มีค่าที่เป็นไปได้สามค่าคือ $\{-1, 0, 1\}$ และมีความน่าจะเป็นเท่ากับ

$$Pr(X = -1) = Pr(X = 0) = Pr(X = 1) = \frac{1}{3}$$

สมมุติให้ $Y = X^2$ ดังนั้น X และ Y ไม่เป็นอิสระต่อกันอย่างชัดเจน ส่วนที่เหลืออยู่คือการตรวจสอบว่า X และ Y มีสหสัมพันธ์ต่อกันหรือไม่? ก่อนอื่น $E[X] = 0$ ดังนั้น

$$\begin{aligned} Cov[X,Y] &= E[XY] - E[X]E[Y] = E[XY] \\ &= Pr(X = -1, Y = 1)(-1) + Pr(X = 0, Y = 0)(0) + Pr(X = 1, Y = 1)(1) \\ &= -\frac{1}{3} + 0 + \frac{1}{3} = 0 \end{aligned}$$

ซึ่งหมายความว่า X และ Y ไม่มีสหสัมพันธ์ (uncorrelated) ทั้งที่ไม่เป็นอิสระต่อกัน □

ข้อสังเกตอันหนึ่งจากตัวอย่างนี้คือความสัมพันธ์ระหว่าง X และ Y เป็นแบบไม่เป็นเชิงเส้น (nonlinear) ซึ่งมีส่วนสำคัญที่ทำให้เกิดปรากฏการณ์ที่ตัวแปรสุ่มสองตัวที่ไม่เป็นอิสระต่อกัน แต่กลับไม่มีสหสัมพันธ์ (uncorrelated) ในขณะเดียวกัน หากความสัมพันธ์ของตัวแปรเป็นแบบเชิงเส้น (linear) จะสามารถสรุปได้ว่า ตัวแปรสุ่มสองตัวที่ไม่เป็นอิสระต่อกันจะมีสหสัมพันธ์ต่อกัน ดังสรุปในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 4.21. *กำหนดให้ X เป็นตัวแปรสุ่มที่ $0 < \sigma_X^2 < \infty$ และ $Y = aX + b$ สำหรับค่าคงที่ $a \neq 0$ และ b แล้วสามารถสรุปได้ว่า*

1. *ถ้า $a > 0$ แล้ว $\rho[X,Y] = 1$*
2. *ถ้า $a < 0$ แล้ว $\rho[X,Y] = -1$*

ยิ่งไปกว่านั้น ทฤษฎีบทต่อไปนี้แสดงให้เห็นอย่างชัดเจนมากยิ่งขึ้นไปอีกว่า สหสัมพันธ์ (correlation) บอกถึงความสัมพันธ์เชิงเส้นเท่านั้น นั่นหมายความว่า การที่พบว่าค่าสหสัมพันธ์ (correlation) ของตัวแปรสุ่มสองตัวมีค่าเท่ากับศูนย์ไม่ได้หมายความว่าตัวแปรสุ่มทั้งสองไม่มีความสัมพันธ์กันหรือเป็นอิสระต่อกัน แต่เป็นเพียงว่าไม่มีความสัมพันธ์เชิงเส้นระหว่างกัน ในทางกลับกัน หากพบว่า ค่าสัมบูรณ์ของค่าสหสัมพันธ์ (correlation) เท่ากับหนึ่ง จะสามารถสรุปได้ว่าตัวแปรสุ่มทั้งสองมีความสัมพันธ์กันเชิงเส้นอย่างแน่นอน ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 4.22. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่ความแปรปรวน (variance) มีค่าจำกัด นั่นคือ $\sigma_X^2 < \infty$ และ $\sigma_Y^2 < \infty$ ถ้า $|\rho[X,Y]| = 1$ แล้ว จะต้องมีค่าคงที่ $a \neq 0$, $b \neq 0$, และ c ที่ทำให้ $aX + bY = c$ ด้วยความน่าจะเป็นเท่ากับหนึ่ง*

การพิสูจน์. เริ่มจากอสมการโคชีและชวาร์ซ (Cauchy-Schwarz Inequality)

$$(Cov[X,Y])^2 \leqslant \sigma_X^2 \sigma_Y^2$$

ซึ่งจะเป็นจริงในรูปแบบสมการก็ต่อเมื่อมีจำนวนจริงที่ไม่เท่ากับศูนย์ $a, b \neq 0$ และจำนวนจริง c ที่ทำให้ $aX + bY = c$ ด้วยความน่าจะเป็นเท่ากับ 1 ซึ่งหมายความว่า ถ้า

$$(Cov[X,Y])^2 = \sigma_X^2 \sigma_Y^2$$

ก็ต่อเมื่อมีจำนวนจริงที่ไม่เท่ากับศูนย์ $a, b \neq 0$ และจำนวนจริง c ที่ทำให้ $aX + bY = c$ ด้วยความน่าจะเป็นเท่ากับ 1 ■

ทฤษฎีบทต่อไปนี้แสดงวิธีการคำนวณค่าความแปรปรวนของผลรวมของตัวแปรสุ่มสองตัว ซึ่งรวมทั้งกรณีที่ตัวแปรสุ่มทั้งสองมีสหสัมพันธ์ต่อกันด้วย

ทฤษฎีบทที่ 4.23. *ถ้า X และ Y เป็นตัวแปรสุ่มที่ความแปรปรวน (variance) มีค่าจำกัด นั่นคือ $\sigma_X^2 < \infty$ และ $\sigma_Y^2 < \infty$ แล้ว*

$$Var[aX + bY + c] = a^2 Var[X] + b^2 Var[Y] + 2abCov[X,Y] \tag{4.37}$$

ทฤษฎีบทที่ 4.24. *ถ้า $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่ความแปรปรวน (variance) มีค่าจำกัด นั่นคือ $Var[X_i] < \infty$ แล้ว*

$$Var\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 Var[X_i] + 2a_i a_j \sum_{i<j} Cov[X_i, X_j] \tag{4.38}$$

การพิสูจน์.

$$\begin{aligned} Var\left[\sum_{i=1}^{n} a_i X_i\right] &= Cov\left[\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov[X_i, X_j] \\ &= \sum_{i=1}^{n} a_i^2 Cov[X_i, X_i] + \sum_{i \neq j} a_i a_j Cov[X_i, X_j] \\ &= \sum_{i=1}^{n} a_i^2 Var[X_i] + 2a_i a_j \sum_{i<j} Cov[X_i, X_j] \end{aligned}$$

■

NEXT EXAMPLE COMES AFTER WE DISCUSS ABOUT COVARIANCE

ตัวอย่างต่อไปนี้แสดงค่าคาดหมาย (mean) และค่าเบี่ยงเบนมาตรฐาน (standard deviation) ของกลุ่มหลักทรัพย์ (portfolio) ที่สร้างขึ้นจากหลักทรัพย์พื้นฐาน (underlying assets) จำนวนหนึ่ง โดยแบ่งเป็นสองกรณี คือ กรณีที่อัตราผลตอบแทนของหลักทรัพย์ไม่มีสหสัมพันธ์ (uncorrelated) ส่วนอีกกรณีหนึ่งนั้นมีสหสัมพันธ์ (correlated) โดยมีเป้าหมายเพื่อให้ผู้อ่านได้เห็นถึงความสำคัญของความแปรปรวนร่วม (covariance) ในการคำนวณหาความแปรปรวนของกลุ่มหลักทรัพย์ ซึ่งมีความสำคัญในด้านเศรษฐศาสตร์และการเงินอย่างมาก

ตัวอย่างที่ 4.17 (ขอบเขตของค่าคาดหมายและความแปรปรวนของกลุ่มหลักทรัพย์ (mean-variance frontier))**.** กำหนดให้ R_i แทนอัตราผลตอบแทนรวม (gross return) ต่อปีของกองทุนรวม (mutual fund) i สำหรับ $i = 1, \ldots, n$ และกำหนดให้ $\alpha_i \in [0, 1]$ โดยที่ $\sum_i \alpha_i = 1$ แทนสัดส่วนการลงทุนในกองทุน i ของกลุ่มหลักทรัพย์ (portfolio) ที่สนใจ

สิ่งที่ต้องการทราบคือกลุ่มหลักทรัพย์ (portfolio) คืออัตราผลตอบแทนรวมของกลุ่มหลักทรัพย์ (portfolio) หลังจากลงทุนเป็นเวลาหนึ่งปี

$$R = \sum_{i=1}^{n} \alpha_i R_i$$

ซึ่งเป็นตัวแปรสุ่มเพราะเป็นฟังก์ชันของตัวแปรสุ่ม □

4.6 ค่าคาดหมายแบบมีเงื่อนไข (Conditional Expectation)

การคาดการณ์ (prediction) หรือการประมาณค่า (estimation) ที่ได้รับความสนใจมักจะอยู่ในรูปของค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ทั้งนี้เป็นเพราะว่า การคาดการณ์ (prediction) หรือการประมาณค่า (estimation) มักจะต้องเริ่มจากการที่รู้อะไรบางอย่างแล้วจึงพยายามที่จะบอกว่าจะเกิดอะไรขึ้นกับตัวแปรหรือสิ่งที่สนใจ ยกตัวอย่างเช่น รายได้ของครัวเรือนที่หัวหน้าครัวเรือนเรียนจบปริญญาตรีแตกต่างจากครัวเรือนที่หัวหน้าครัวเรือนจบอาชีวศึกษานั้นมีค่าเท่าใด หรืออัตราผลตอบแทนของกองทุนจะเป็นเท่าใดหากอัตราผลตอบแทนของตลาดเท่ากับ 10 เปอร์เซ็นต์ต่อปี เป็นต้น

บทนิยามที่ 4.8. *กำหนดให้ X และ Y เป็นตัวแปรสุ่ม โดยที่ค่าคาดหมายของ Y มีค่าจำกัด นั่นคือ $\mu_Y < \infty$ สำหรับกรณีของตัวแปรสุ่มต่อเนื่อง ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ Y เมื่อค่าของ $X = x$ เท่ากับ*

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_2(y|x)\, dy \tag{4.39}$$

และค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ X เมื่อค่าของ $Y = y$ เท่ากับ

$$E\left[X|Y=y\right] = \int_{-\infty}^{\infty} x f_1\left(x|y\right) dx \tag{4.40}$$

สำหรับกรณีของตัวแปรสุ่มไม่ต่อเนื่อง ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ Y เมื่อค่าของ $X = x$ เท่ากับ

$$E\left[Y|X=x\right] = \sum_{y} y f_2\left(y|x\right) dy \tag{4.41}$$

และค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ X เมื่อค่าของ $Y = y$ เท่ากับ

$$E\left[X|Y=y\right] = \sum_{x} x f_1\left(x|y\right) dx \tag{4.42}$$

ภายใต้นิยามข้างต้นค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ Y เมื่อค่าของ $X = x$ ซึ่งแทนด้วย $E\left[Y|X=x\right]$ นั่นจะถูกพิจารณาเป็นค่าคงที่ค่าหนึ่ง แต่หากนำเอาค่าคาดหมายแบบมีเงื่อนไขสำหรับค่าต่างๆ ของ x มารวบรวมเข้าในรูปแบบของการแจกแจง (distribution) แล้ว ก็จะเห็นได้ว่า ค่าคาดหมายแบบมีเงื่อนไขเป็นตัวแปรสุ่มแบบหนึ่ง ซึ่งในกรณีนี้มักเขียนแทนด้วย $E\left[Y|X\right]$ โดยไม่ระบุค่าของตัวแปรที่เป็นตัวกำหนดเงื่อนไขซึ่งในที่นี้ก็คือ X ผู้อ่านควรตระหนักถึงความแตกต่างระหว่าง $E\left[Y|X=x\right]$ และ $E\left[Y|X\right]$ โดยเฉพาะในประเด็นที่ต้องพิจารณาให้ $E\left[Y|X\right]$ เป็นตัวแปรสุ่ม เพราะจะมีส่วนความสำคัญอย่างมากหลังจากนี้

บทนิยามที่ 4.9. *กำหนดให้ $h\left(x\right) = E\left[Y|X=x\right]$ เป็นฟังก์ชันของ x นิยามให้ $E\left[Y|X\right]$ แทนค่าคาดหมาย (mean) ของ $h\left(X\right)$ และเรียก $E\left[Y|X\right] \equiv E\left[h\left(X\right)\right]$ ว่า ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ Y เมื่อทราบ X*

ตัวอย่างต่อไปนี้แสดงตัวอย่างของค่าคาดหมายแบบมีเงื่อนไขของผลตอบแทนการลงทุนของกองทุนรวมในปีปัจจุบันภายใต้เงื่อนไขที่ว่าเราทราบผลตอบแทนการลงทุน ในปีที่ผ่านมา เพื่อความสะดวกในการนำเสนอ จึงแบ่งอัตราผลตอบแทนในปีที่ผ่านมาออกเป็น 10 ช่วงที่ไม่ต่อเนื่องกัน (discrete) ประเด็นที่ต้องการให้สังเกตในตัวอย่างนี้คือ การที่ค่าคาดหมายของอัตราผลตอบแทนในปัจจุบันสำหรับแต่ละช่วงอัตราผลตอบแทนในปีที่ผ่านมานั้นเป็นค่าคงที่ค่าหนึ่ง แต่เมื่อพิจารณาค่าคาดหมายของทั้ง 10 ช่วงโดยรวม ก็จะได้ผลออกมาเป็นการแจกแจงของค่าคาดหมายแบบมีเงื่อนไข ซึ่งหมายความว่า ค่าคาดหมายแบบมีเงื่อนไขที่ไม่ได้ระบุค่าของตัวแปรเงื่อนไขอย่างเจาะจงนั้นเป็นตัวแปรสุ่ม

ตัวอย่างที่ 4.18. ตัวอย่างนี้นำเสนอค่าคาดหมายแบบมีเงื่อนไขของผลตอบแทนการลงทุนของกองทุนรวมในปี 2560 ภายใต้เงื่อนไขที่ว่าเราทราบผลตอบแทนการลงทุน ในปี 2559 □

ตัวอย่างที่ 4.19. พิจารณาการทดลองทางคลินิค (clinical trial) โดยกำหนดให้ X_i แทนตัวแปรสุ่มที่บ่งบอกถึงผลลัพธ์ของการรักษาสำหรับผู้ป่วย i นั่นคือ $X_i = 1$ ถ้ารักษาได้สำเร็จ แต่ $X_i = 0$ ถ้าไม่สำเร็จ ในขณะเดียวกัน นักวิจัยยังไม่ทราบว่า ความน่าจะเป็นที่จะรักษาสำเร็จมีค่าเท่าใด จึงต้องพิจารณาให้ความน่าจะเป็นดังกล่าวเป็นตัวแปรสุ่ม ซึ่งแทนด้วย P ในขณะเดียวกัน สมมุติให้ $X_1, \ldots, X_n$ มีการแจกแจงแบบมีเงื่อนไขเมื่อทราบว่า $P = p$ เหมือนกัน นั่นคือ $Pr\left(X_i = 1 | P = p\right) = p$ สำหรับทุก $i = 1, \ldots, n$

สิ่งที่เราสนใจในตัวอย่างนี้คือ ค่าคาดหมายของจำนวนผู้ป่วยที่จะหายจากอาการป่วยหรือรักษาสำเร็จ ซึ่งแทนด้วย $X = X_1 + \cdots + X_n$ คุณสมบัติเชิงเส้นของค่าคาดหมายแบบมีเงื่อนไขทำให้สามารถสรุปได้ว่า

$$E\left[X | P = p\right] = \sum_{i=1}^{n} E\left[X_i | P = p\right] = np$$

ดังนั้น

$$E\left[X|P\right] = nP$$

เนื่องจาก P เป็นตัวแปรสุ่ม ดังนั้น $E\left[X|P\right]$ เป็นตัวแปรสุ่มด้วย □

ทฤษฎีบทต่อไปนี้แสดงวิธีการคำนวณหาค่าคาดหมายของตัวแปรสุ่มจากค่าคาดหมายแบบมีเงื่อนไข โดยอาศัยหลักการที่ว่าค่าคาดหมายแบบมีเงื่อนไข $E\left[Y|X\right]$ เป็นตัวแปรสุ่มที่ขึ้นอยู่กับตัวแปรสุ่มที่เป็นเงื่อนไข ซึ่งในที่นี้หมายถึงตัวแปรสุ่ม X ดังนั้น หากเราดำเนินการหาค่าคาดหมายโดยใช้การแจกแจงของตัวแปรสุ่ม X อีกครั้งหนึ่งก็จำให้ได้ค่าคาดหมายของ X (ไม่ใช่ตัวแปรสุ่มอีกต่อไป) หลักการนี้มีประโยชน์อย่างมากในการวิเคราะห์เชิงเศรษฐศาสตร์และการเงิน ซึ่งมักเรียกคุณสมบัตินี้ว่า กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) ในขณะที่ นักสถิติเรียกทฤษฎีบทนี้ว่ากฎของความน่าจะเป็นรวมสำหรับค่าคาดหมาย (Law of Total Probability for Expectations)

ทฤษฎีบทที่ 4.25. *กำหนดให้ X และ Y เป็นตัวแปรสุ่ม โดยที่ค่าคาดหมายของ Y มีค่าจำกัด นั่นคือ $\mu_Y < \infty$ ดังนั้น*

$$E\left[E\left[Y|X\right]\right] = E\left[Y\right] \tag{4.43}$$

การพิสูจน์. เพื่อประหยัดพื้นที่ขอนำเสนอเฉพาะการพิสูจน์สำหรับตัวแปรสุ่มต่อเนื่องดังต่อไปนี้

$$E\left[E\left[Y|X\right]\right] = \int_{-\infty}^{\infty} E\left[Y|x\right] f_1\left(x\right) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y g_2\left(y|x\right) f_1\left(x\right) dx$$

เนื่องจาก $g_2\left(y|x\right) f_1\left(x\right) = f\left(x, y\right)$ ทำให้สรุปได้ว่า

$$E\left[E\left[Y|X\right]\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f\left(x, y\right) dx = E\left[Y\right]$$

■

ตัวอย่างต่อไปนี้ใช้ข้อมูลจากตัวอย่างที่ 4.18 เพื่อคำนวณหาค่าคาดหมายของอัตราผลตอบแทนในปีปัจจุบัน บทเรียนที่สำคัญอันหนึ่งก็คือ กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) มีลักษณะคล้ายกับการหาค่าเฉลี่ยถ่วงน้ำหนัก (weighted average) นั่นเอง

ตัวอย่างที่ 4.20. จากตัวอย่างที่ 4.18 เราสามารถคำนวณหาค่าคาดหมายของอัตราผลตอบแทนในปีปัจจุบันได้เป็น □

ตัวอย่างที่ 4.21 (จาก DeGroot and Schervish (2012))**.** สมมุติว่า X และ Y เป็นตัวแปรสุ่มที่มีการแจกแจงเอกรูป (uniform distribution) พิจารณาการสุ่มที่เริ่มจากสุ่มเลือกค่า x จากการแจกแจงเอกรูปที่มีค่าอยู่ในช่วง $[0,1]$ หลังจากนั้นจึงสุ่มเลือกค่า y จากจากการแจกแจงเอกรูปที่มีค่าอยู่ในช่วง $[x,1]$ นั่นคือการแจกแจงแบบมีเงื่อนไขของ Y เมื่อทราบค่า $X=x$ เป็นการแจกแจงเอกรูปที่มีค่าอยู่ในช่วง $[x,1]$ คำถามก็คือ ค่าคาดหมาย (expectation) ของ Y มีค่าเท่าใด?

ก่อนอื่น ค่าคาดหมายแบบมีเงื่อนไขของ Y เมื่อทราบค่า $X=x$ มีค่าเท่ากับ

$$E\left[Y|X=x\right]=\frac{1+x}{2}$$

ดังนั้น $E\left[Y|X\right]=\frac{1+X}{2}$ ส่วนต่อไปคือการประยุกต์ใช้กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) เพื่อหาค่าคาดหมาย (expectation) ของ Y ซึ่งมีค่าเท่ากับ

$$E\left[Y\right]=E\left[E\left[Y|X\right]\right]=E\left[\frac{1+X}{2}\right]=\frac{1+\frac{1}{2}}{2}=\frac{3}{4}$$

ทั้งนี้สมการที่สามใช้ผลลัพธ์ที่ว่าค่าคาดหมายของตัวแปรสุ่มที่มีการแจกแจงเอกรูปในช่วง $[0,1]$ เท่ากับ $E\left[X\right]=\frac{1}{2}$ □

ตัวอย่างต่อไปนี้แสดงวิธีการประยุกต์ใช้กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) เพื่อหาค่าคาดหมายต่างๆ สำหรับกรณีที่ค่าคาดหมายแบบมีเงื่อนไขเป็นแบบเชิงเส้น

ตัวอย่างที่ 4.22 (จาก DeGroot and Schervish (2012))**.** สมมุติว่า $E\left[Y|X\right]=aX+b$ สำหรับค่าคงที่ a และ b ใดๆ คำถามคือ ค่าคาดหมาย $E\left[XY\right]$ ในรูปของ $E\left[X\right]$ และ $E\left[X^2\right]$ มีค่าเท่าใด?

เริ่มจากหลักการที่ว่า เราสามารถพิจารณาให้ตัวแปรสุ่ม X เป็นค่าคงที่ได้ในกรณีของค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ของ Y เมื่อทราบ X ทั้งนี้เพราะเราทราบค่า X แล้ว ดังนั้น

$$E\left[XY|X\right]=XE\left[Y|X\right]=aX^2+bX$$

หลักจากนั้นจึงประยุกต์ใช้กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) เพื่อคำนวณหาค่า $E\left[XY\right]$

$$E\left[XY\right]=E\left[E\left[XY|X\right]\right]=E\left[aX^2+bX\right]=aE\left[X^2\right]+bE\left[X\right]$$

□

อันที่จริงแล้ว หากทราบว่าค่าคาดหมายแบบมีเงื่อนไขมีความสัมพันธ์แบบเชิงเส้น (linear relationship) และทราบค่าคาดหมาย (mean) μ_X, μ_Y ค่าความแปรปรวน (variance) σ_X^2, σ_Y^2 และค่าสหสัมพันธ์ (correlation) ρ ของตัวแปรสุ่มทั้งสองตัว X และ Y ก็จะสามารถระบุค่าคงที่ a และ b ในรูปค่าสถิติดังกล่าวและเขียนค่าคาดหมายแบบมีเงื่อนไขได้เป็น

$$E[Y|X] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - \mu_X) \tag{4.44}$$

ที่สำคัญรูปแบบความสัมพันธ์นี้ที่ได้มีความลักษณะคล้ายคลึงกับความสัมพันธ์ที่ได้จากการประมาณค่าแบบกำลังสองน้อยที่สุด (Ordinary Least Square หรือที่เรียกสั้นๆ ว่า OLS) ซึ่งจะกล่าวถึงในบทที่ XXX โดยจะเป็นได้ว่าค่าสัมประสิทธิ์ของ X นั้นขึ้นอยู่กับค่าสหสัมพันธ์เป็นหลัก โดยเฉพาะอย่างยิ่งในการกำหนดว่าค่าจะเป็นบวกหรือลบ ผู้อ่านควรจะพยายามพิสูจน์ความสัมพันธ์นี้เป็นแบบฝึกหัด

ส่วนต่อไปนำเสนอความแปรปรวนแบบมีเงื่อนไข (conditional variance) ซึ่งใช้หลักการเดียวกับค่าคาดหมายแบบมีเงื่อนไข หรือในมุมหนึ่งก็คือค่าคาดหมายแบบมีเงื่อนไขของฟังก์ชันของตัวแปรสุ่มนั่นเอง

บทนิยามที่ 4.10. *กำหนดให้ X และ Y เป็นตัวแปรสุ่ม ค่าความแปรปรวนแบบมีเงื่อนไข (conditional variance) ของ Y เมื่อค่าของ $X = x$ เท่ากับ*

$$Var[Y|x] = E\left[(Y - E[Y|x])^2 |x\right] \tag{4.45}$$

ในทำนองเดียวกันกับค่าคาดหมายแบบมีเงื่อนไข $Var[Y|x]$ เป็นค่าคงที่ แต่ $Var[Y|X]$ เป็นตัวแปรสุ่ม

ทฤษฎีบทต่อไปนี้แสดงถึงวิธีการแยกส่วนค่าความแปรปรวน (variance decomposition) ซึ่งมีประโยชน์ในการทำความเข้าใจปัจจัยที่มีผลต่อค่าความแปรปรวน ยกตัวอย่างเช่น XXX เป็นต้น FIND PAPERS TO CITE HERE ทางสถิติมักเรียกทฤษฎีบทนี้ว่า กฎของความน่าจะเป็นรวมสำหรับค่าความแปรปรวน (Law of Total Probability for Variances)

ทฤษฎีบทที่ 4.26. *สมมุติให้ X และ Y เป็นตัวแปรสุ่มที่มีค่าคาดหมายและค่าความแปรปรวนที่จำกัด แล้ว*

$$Var[Y] = E[Var[Y|X]] + Var[E[Y|X]] \tag{4.46}$$

การพิสูจน์. เริ่มจากความแปรปรวนแบบมีเงื่อนไข (conditional variance)

$$Var[Y|X] = E\left[Y^2|X\right] - E[Y|X]^2$$

ซึ่งเป็นตัวแปรสุ่ม ดังนั้น

$$\begin{aligned} E[Var[Y|X]] &= E\left[E\left[Y^2|X\right]\right] - E\left[E[Y|X]^2\right] \\ &= E\left[Y^2\right] - E\left[E[Y|X]^2\right] \end{aligned} \tag{4.47}$$

ในขณะเดียวกัน

$$\begin{aligned} Var\left[E\left[Y|X\right]\right] &= E\left[E\left[Y|X\right]^2\right] - \left(E\left[E\left[Y|X\right]\right]\right)^2 \\ &= E\left[E\left[Y|X\right]^2\right] - E\left[Y^2\right] \end{aligned} \tag{4.48}$$

เมื่อนำสมการที่ 4.47 และ 4.48 มารวมกัน ทำให้สรุปได้ว่า

$$E\left[Var\left[Y|X\right]\right] + Var\left[E\left[Y|X\right]\right] = E\left[Y^2\right] - E\left[Y^2\right] = Var\left[Y\right]$$

■

4.7 ค่าคาดหมายและค่ามัธยฐานในฐานะตัวทำนาย (Mean and Median as Predictors)

หัวข้อนี้นำเสนอแนวคิดพื้นฐานของการทำนายในทางสถิติ โดยมีจุดประสงค์หลักเพื่อชี้ให้เห็นว่า ค่าสถิติที่แตกต่างกัน เช่น ค่าคาดหมาย (mean) ค่ามัธยฐาน (median) เป็นต้น ต่างสามารถเป็นตัวทำนาย (predictor) ที่เหมาะสมได้ขึ้นอยู่กับเงื่อนไขที่ใช้ในการกำหนดความเหมาะสมหรือจุดประสงค์ที่ต้องการตอบ เช่น ค่าคาดหมาย (mean) เป็นตัวทำนายที่เหมาะสมหากจุดประสงค์คือการทำให้ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (mean square error: m.s.e.) มีค่าต่ำที่สุด ในขณะที่ ค่ามัธยฐาน (median) เป็นตัวทำนายที่เหมาะสมหากจุดประสงค์คือการทำให้ค่าคาดหมายของค่าสัมบูรณ์ของค่าคลาดเคลื่อน (mean absolute error: m.a.e.) มีค่าต่ำที่สุด เป็นต้น นอกจากนี้ ยังต้องการให้ผู้อ่านได้เริ่มเห็นภาพที่ชัดเจนมากขึ้นว่า ทำไมนักสถิติจึงให้ความสนใจกับค่าสถิติบางค่าเป็นพิเศษ เช่น ค่าคาดหมาย (mean) ค่ามัธยฐาน (median) เป็นต้น

ปัญหาการทำนายที่สนใจในที่นี้ คือความต้องการทำนายว่าค่าตัวเลขที่สุ่มได้จากการแจกแจงที่มีค่าคาดหมายเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 จะมีค่าเท่าใด? นั่นคือ ต้องการหาค่าตัวทำนายที่เหมาะสมนั่นเอง? คำตอบที่ได้ขึ้นอยู่กับฟังก์ชันจุดประสงค์ (objective function) ที่ใช้ในการกำหนดตัวทำนาย โดยในที่นี้จะยกตัวอย่างฟังก์ชันจุดประสงค์ 2 แบบคือ การหาตัวทำนายที่ทำให้ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) มีค่าต่ำที่สุด และการหาตัวทำนายที่ทำให้ค่าคาดหมายของค่าสัมบูรณ์ของค่าคลาดเคลื่อน (mean absolute error: m.a.e.) มีค่าต่ำที่สุด

บทนิยามที่ 4.11. *กำหนดให้ Y เป็นตัวแปรสุ่ม และ d คือตัวทำนาย (predictor) ดังนั้น ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) ของตัวทำนาย d นิยามได้เป็น $E\left[(Y-d)^2\right]$*

ทฤษฎีบทต่อไปนี้แสดงให้เห็นว่า ค่าคาดหมาย (mean) คือตัวทำนายที่ทำให้ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) มีค่าต่ำที่สุด

ทฤษฎีบทที่ 4.27. *กำหนดให้ X เป็นตัวแปรสุ่มที่ค่าความแปรปรวนมีค่าจำกัด และกำหนดให้ $\mu = E[Y]$ ดังนั้น สำหรับค่าจำนวนจริง d ใดๆ*

$$E\left[(Y-d)^2\right] \geqslant E\left[(Y-\mu)^2\right] \tag{4.49}$$

ซึ่งจะเป็นจริงในรูปสมการก็ต่อเมื่อ $d = \mu$

การพิสูจน์. ก่อนอื่นเราสามารถเขียนค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) ในรูปฟังก์ชันของตัวทำนาย d ได้เป็น

$$E\left[(Y-d)^2\right] = E\left[Y^2 - 2dY + d^2\right] = E\left[Y^2\right] - 2d\mu + d^2$$

ซึ่งเป็นฟังก์ชันกำลังสอง (quadratic function) ดังนั้น สามารถประยุกต์ใช้หลักการแคลคูลัสพื้นฐานในการหาค่าต่ำสุดได้ด้วยการกำหนดให้ตัวทำนายที่ทำให้ได้ค่าฟังก์ชันนี้มีค่าต่ำสุด d^* คือค่าที่ทำให้ค่าอนุพันธ์อันดับที่หนึ่งมีค่าเท่ากับศูนย์ นั่นคือ

$$\frac{\partial}{\partial d}\left(E\left[Y^2\right] - 2d^*\mu + (d^*)^2\right) = 0 \Rightarrow -2\mu + 2d^* = 0 \Rightarrow d^* = \mu$$

เนื่องจากฟังก์ชันจุดประสงค์ (objective function) เป็นแบบกำลังสอง ค่าทำให้ได้ค่าต่ำสุดจึงมีค่าเดียว (unique) ผลที่ตามมาก็คือ สำหรับค่าจำนวนจริง d ใดๆ

$$E\left[(Y-d)^2\right] \geqslant E\left[(Y-\mu)^2\right]$$

■

ทฤษฎีบทต่อไปนี้แสดงให้เห็นว่า ข้อสรุปแบบเดียวกันนี้เป็นจริงสำหรับกรณีที่ใช้ค่าคาดหมายแบบมีเงื่อนไข (conditional mean) แทนค่าคาดหมายธรรมดา โดยสมมุติว่ามีตัวแปรสุ่มอีกตัวหนึ่ง X ซึ่งมีความสัมพันธ์กับตัวแปรสุ่มที่ต้องการทราบหรือทำนายค่า Y และที่สำคัญเนื่องจากทราบค่าของ X ตัวทำนายจึงอยู่ในรูปของฟังก์ชันของ X นั่นคือ $d(X)$ ทฤษฎีบทต่อไปนี้แสดงให้เห็นว่า ตัวทำนายที่ทำให้ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) มีค่าต่ำที่สุดก็คือ ค่าคาดหมายแบบมีเงื่อนไข (conditional mean) ของ Y เมื่อทราบ X

ทฤษฎีบทที่ 4.28. *กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่ค่าความแปรปรวนมีค่าจำกัด ตัวทำนาย (predictor) ของตัวแปรสุ่ม Y เมื่อทราบค่า X ที่ทำให้ค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) $E\left[(Y-d(X))^2\right]$ มีค่าต่ำที่สุด คือ $d^*(X) = E[Y|X]$*

การพิสูจน์. กำหนดให้ $d(X)$ คือตัวทำนายใดๆ ซึ่งเป็นฟังก์ชันของ X และ $d^*(X) = E[Y|X]$ สิ่งที่ต้องการพิสูจน์คือ

$$E\left[(Y - d^*(X))^2\right] \leqslant E\left[(Y - d(X))^2\right]$$

ซึ่งทำได้ด้วยการประยุกต์ใช้กฎการหาค่าคาดหมายซ้ำ (Law of Iterative Expectation) ซึ่งทำให้สามารถเขียนค่าคาดหมายของค่าคลาดเคลื่อนกำลังสอง (m.s.e.) สำหรับตัวทำนาย $d(X)$ ใดๆ ในรูปของค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) ได้เป็น

$$E\left[(Y - d(X))^2\right] = E\left[E\left[(Y - d(X))^2 |X\right]\right]$$

ซึ่งบอกเป็นนัยว่า สิ่งที่ต้องการพิสูจน์คือ

$$E\left[E\left[(Y - d^*(X))^2 |X\right]\right] \leqslant E\left[E\left[(Y - d(X))^2 |X\right]\right]$$

ความสัมพันธ์นี้จะเป็นจริงถ้า

$$E\left[(Y - d^*(X))^2 |X = x\right] \leqslant E\left[(Y - d(X))^2 |X = x\right] \tag{4.50}$$

สำหรับค่า x ใดๆ เมื่อนำเอาอสมการที่ 4.50 ไปเปรียบเทียบกับอสมการที่ 4.49 จะเห็นได้ว่า อสมการทั้งสองนั้นสมมูล (equivalent) กัน หากพิจารณาว่าตัวทำนายใน $E\left[(Y - d(X))^2 |X = x\right]$ นั้นเป็นค่าจำนวนจริงเช่นเดียวกับ d ในอสมการที่ 4.49 ดังนั้น จึงสามารถประยุกต์ใช้ทฤษฎีบทที่ 4.27 สรุปได้ว่า

$$E\left[(Y - E[Y|X])^2 |X = x\right] \leqslant E\left[(Y - d(X))^2 |X = x\right] \tag{4.51}$$

สำหรับฟังก์ชัน $d(X)$ และค่าจำนวนจริง x ใดๆ ดังนั้น จึงสามารถสรุปได้ว่า

$$E\left[E\left[(Y - d^*(X))^2 |X\right]\right] \leqslant E\left[E\left[(Y - d(X))^2 |X\right]\right]$$

ซึ่งนำไปสู่บทสรุปของทฤษฎีบทที่ต้องการ ■

ทฤษฎีบทต่อไปนี้แสดงให้เห็นว่าหากเปลี่ยนฟังก์ชันจุดประสงค์ (objective function) เป็นอย่างอื่น ตัวทำนายที่เหมาะสมก็อาจจะเปลี่ยนไปได้ โดยในที่นี้จะพิจารณากรณีที่ค่ามัธยฐาน (median) เป็นตัวทำนายที่เหมาะสมหากจุดประสงค์คือการทำให้ค่าคาดหมายของค่าสัมบูรณ์ของค่าคลาดเคลื่อน (m.a.e.) มีค่าต่ำสุด

ก่อนอื่นจำเป็นต้องกำหนดนิยามของค่ามัธยฐาน (median) ซึ่งครอบคลุมทั้งกรณีของตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง ดังนี้

บทนิยามที่ 4.12. *ค่ามัธยฐาน (median) m ของการแจกแจงของตัวแปรสุ่ม X ต้องสอดคล้องกับเงื่อนไขต่อไปนี้*

$$Pr\left(X \leqslant m\right) \geqslant \frac{1}{2} \text{ และ } Pr\left(X \geqslant m\right) \geqslant \frac{1}{2} \tag{4.52}$$

ตัวอย่างต่อไปนี้ช่วยให้เข้าใจนิยามของค่ามัธยฐาน (median) ได้ดียิ่งขึ้น

ตัวอย่างที่ 4.23. พิจารณาตัวแปรสุ่มไม่ต่อเนื่อง X ที่มีการแจกแจงดังนี้

$$Pr\left(X=1\right)=0.2, Pr\left(X=3\right)=0.4, Pr\left(X=5\right)=0.1, Pr\left(X=7\right)=0.3$$

ในกรณีนี้ จะเห็นได้ว่าค่ามัธยฐาน $m=3$ เพราะ

$$Pr\left(X \leqslant 3\right)=0.6 \geqslant 0.5 \text{ และ } Pr\left(X \geqslant 3\right)=0.8 \geqslant 0.5$$

ยิ่งไปกว่านั้น ค่าที่เป็นไปได้ของตัวแปรสุ่ม X ค่าอื่นไม่สามารถเป็นค่ามัธยฐานได้ ดังนั้น $m=3$ จึงเป็นค่ามัธยฐานเพียงค่าเดียว (unique) □

ตัวอย่างที่ 4.24. พิจารณาตัวแปรสุ่มไม่ต่อเนื่อง X ที่มีการแจกแจงต่างออกไปจากตัวอย่างก่อนหน้านี้เล็กน้อย

$$Pr\left(X=1\right)=0.1, Pr\left(X=3\right)=0.4, Pr\left(X=5\right)=0.2, Pr\left(X=7\right)=0.3$$

ซึ่งจะเห็นได้ว่า

$$Pr\left(X \leqslant 3\right)=0.5 \text{ และ } Pr\left(X \geqslant 5\right)=0.5$$

ดังนั้น ค่าจำนวนจริง m ที่อยู่ระหว่างช่วง $3 \leqslant m \leqslant 5$ ล้วนเป็นค่ามัธยฐานทั้งสิ้นเพราะสอดคล้องกับสมการที่ 4.52 กล่าวคือ ค่ามัธยฐานในกรณีนี้ไม่มากกว่าหนึ่งค่า (not unique) ในทางปฏิบัติ จึงมักนิยมที่จะใช้ค่ากึ่งกลางของช่วงดังกล่าวแทนค่ามัธยฐาน ซึ่งในกรณีมีค่าเท่ากับ 4 □

กรณีที่มีค่ามัธยฐานมากกว่าหนึ่งค่าสามารถเกิดขึ้นกับการแจกแจงแบบต่อเนื่องได้เช่นกัน ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 4.25. พิจารณาตัวแปรสุ่ม X ที่มีการแจกแจงต่อเนื่องและมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$f\left(x\right)=\begin{cases} \frac{1}{4} \text{ สำหรับ } -1 \leqslant x \leqslant 1 \\ \frac{1}{2}, \text{ สำหรับ } 3 \leqslant x \leqslant 4 \\ 0, \text{ สำหรับกรณีอื่น} \end{cases}$$

ซึ่งสามารถใช้คำนวณหาฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ได้เป็น

$$F(x)=\begin{cases}0,\ \text{สำหรับ}\ -\infty\leqslant x\leqslant -1\\ \frac{1}{4}+\frac{1}{4}x,\ \text{สำหรับ}\ -1\leqslant x\leqslant 1\\ \frac{1}{2},\ \text{สำหรับ}\ 1\leqslant x\leqslant 3\\ -1+\frac{1}{2}x,\ \text{สำหรับ}\ 3\leqslant x\leqslant 4\\ 1,\ \text{สำหรับ}\ 4\leqslant x\leqslant \infty\end{cases}$$

ดูรูปที่ XXX ประกอบ จะเห็นได้ว่า ค่าจำนวนจริง m ที่อยู่ระหว่าง $1\leqslant m\leqslant 3$ ล้วนแต่มีผลทำให้ $Pr(X\leqslant m)\geqslant \frac{1}{2}$ และ $Pr(X\geqslant m)\geqslant \frac{1}{2}$ นั่นหมายความว่า ค่าจำนวนจริงเหล่านี้ล้วนเป็นค่ามัธยฐาน ในทำนองเดียวกับตัวอย่างก่อนหน้านี้ ในทางปฏิบัติมักนิยมที่จะใช้ค่ากึ่งกลางของช่วงดังกล่าวแทนค่ามัธยฐาน ซึ่งในกรณีมีค่าเท่ากับ 2 □

ทฤษฎีบทต่อไปนี้แสดงว่า ค่ามัธยฐาน (median) เป็นตัวทำนายที่เหมาะสมหากจุดประสงค์คือการทำให้ค่าคาดหมายของค่าสัมบูรณ์ของค่าคลาดเคลื่อน (m.a.e.) มีค่าต่ำสุด

ทฤษฎีบทที่ 4.29. *กำหนดให้ X เป็นตัวแปรสุ่มที่ค่าความคาดหมายมีค่าจำกัด และ m แทนค่ามัธยฐาน (median) ของการแจกแจงของตัวแปรสุ่ม X ดังนั้น สำหรับค่าจำนวนจริง d ใดๆ*

$$E[|X-m|]\leqslant E[|X-d|] \tag{4.53}$$

การพิสูจน์. ในที่นี้จะพิจารณาเฉพาะกรณีที่ $d>m$ ส่วนกรณีตรงกันข้ามสามารถพิสูจน์ได้ด้วยขั้นตอนที่คล้ายคลึงกัน

$$\begin{aligned}E[|X-d|]-E[|X-m|] &= \int_{-\infty}^{\infty}(|x-d|-|x-m|)\,f(x)\,dx\\ &= \int_{-\infty}^{m}(d-m)\,f(x)\,dx+\int_{m}^{d}(d+m-2x)\,f(x)\,dx\\ &\quad +\int_{d}^{\infty}(m-d)\,f(x)\,dx\end{aligned}$$

เนื่องจาก $d+m-2x\leqslant d+m-2d=m-d$ สำหรับจำนวนจริง x ใดๆ ที่ $m\leqslant x\leqslant d$ จึงสามารถเขียน

สมการข้างบนใหม่ได้เป็น

$$\begin{aligned} E\left[|X-d|\right]-E\left[|X-m|\right] &\geqslant \int_{-\infty}^{m}(d-m)\,f\left(x\right)dx+\int_{m}^{d}(m-d)\,f\left(x\right)dx \\ &\quad+\int_{d}^{\infty}(m-d)\,f\left(x\right)dx \\ &=(d-m)\left[\int_{-\infty}^{m}f\left(x\right)dx-\int_{m}^{\infty}f\left(x\right)dx\right] \\ &=(d-m)\left[2\int_{-\infty}^{m}f\left(x\right)dx-\left[\int_{-\infty}^{m}f\left(x\right)dx+\int_{m}^{\infty}f\left(x\right)dx\right]\right] \\ &=(d-m)\left[2\int_{-\infty}^{m}f\left(x\right)dx-\int_{-\infty}^{\infty}f\left(x\right)dx\right] \\ &=(d-m)\left[2Pr\left(X\leqslant m\right)-1\right] \end{aligned}$$

ในขณะเดียวกัน ค่ามัธยฐาน m จะต้องสอดคล้องกับเงื่อนไขที่ว่า $Pr\left(X\leqslant m\right)\geqslant\frac{1}{2}$ ซึ่งส่งผลให้ $2Pr\left(X\leqslant m\right)-1\geqslant 0$ ดังนั้น จึงสามารถสรุปได้ว่า

$$E\left[|X-d|\right]-E\left[|X-m|\right]\geqslant 0\Rightarrow E\left[|X-d|\right]\geqslant E\left[|X-m|\right]$$

สำหรับค่าจำนวนจริง d ใดๆ ■

4.8 โมเมนต์และฟังก์ชันก่อกำเนิดโมเมนต์ (Moments and Moment Generating Function)

ที่ผ่านมา เรารวบรวมคุณสมบัติทั้งหมดของตัวแปรสุ่มในรูปของฟังก์ชันการแจกแจง (distribution function) ไม่ว่าจะเป็นในรูปของฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) หรือฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ส่วนบทนี้ได้นำเสนอค่าคาดหมาย (expectation) ซึ่งเป็นเครื่องมือที่ช่วยสรุปคุณสมบัติของการแจกแจงให้เข้าใจได้ง่ายและสะดวกมากยิ่งขึ้น ในขณะเดียวกัน ค่าคาดหมาย (expectation) อันเดียวไม่สามารถบ่งบอกถึงคุณสมบัติของการแจกแจงได้ครบถ้วน แต่หากรวบรวมเอาค่าคาดหมายของฟังก์ชันยกกำลังของตัวแปรสุ่มสำหรับทุกๆ ค่ายกกำลังที่เป็นจำนวนเต็มบวกแล้วก็จะได้คุณสมบัติที่ครบถ้วนตามที่ต้องการ นักสถิติมักเรียกค่าคาดหมายของฟังก์ชันยกกำลังของตัวแปรสุ่มว่า ค่าโมเมนต์ (moment) และที่สำคัญเราสามารถรวบรวมเอาค่าโมเมนต์ทั้งหมดนั้นมาไว้ด้วยกันในรูปของฟังก์ชัน โดยเรียกฟังก์ชันนี้ว่า ฟังก์ชันก่อกำเนิดโมเมนต์ (moment generating function หรือ m.g.f.) ซึ่งเป็นฟังก์ชันที่ช่วยให้สามารถหาค่าโมเมนต์ต่างๆ ได้โดยสะดวก ยิ่งไปกว่านั้น การแจกแจงอันใดอันหนึ่ง จะมีฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ได้เพียงอันเดียว (unique) ซึ่งแตกต่างจากกรณีของฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ซึ่งมีได้มากมาย

โมเมนต์ (moment) ที่ k ของ X คือค่าคาดหมายของ X^k ซึ่งแทนด้วย $E\left[X^k\right]$ ส่วน $E\left[(X-\mu)^k\right]$ นั้นจะถูกเรียกว่า โมเมนต์ศูนย์กลาง (central moment) ที่ k ของ X โดยที่ $\mu = E[X]$ แทนค่าคาดหมายของ X หากพิจารณาให้ดีแล้วจะเห็นได้ว่า โมเมนต์ศูนย์กลางที่สองก็คือค่าความแปรปรวนนั่นเอง ส่วนโมเมนต์ศูนย์กลางที่สามนั้นใช้เพื่อบ่งบอกถึงความสมมาตร (symmetry) ของการแจกแจงซึ่งวัดด้วยค่าสถิติที่เรียกว่าความเบ้ (skewness) ซึ่งนิยามได้เป็น $\frac{E\left[(X-\mu)^3\right]}{\sigma^3}$

ทฤษฎีบทต่อไปนี้แสดงว่าหากโมเมนต์ที่ k มีค่าจำกัด ค่าโมเมนต์ที่ต่ำกว่านั้น (สำหรับ $j<k$) จะมีค่าจำกัดเสมอ ซึ่งหมายความว่า หากเราสามารถหาค่าโมเมนต์ในอันดับที่สูงของการแจกแจงได้ เราย่อมสามารถหาค่าโมเมนต์ในอันดับที่ต่ำลงมาได้เสมอ ยกตัวอย่างเช่น หากสามารถหาค่าความแปรปรวนได้ ย่อมหมายความว่าค่าความคาดหมายย่อมหาค่าได้เช่นกัน

ทฤษฎีบทที่ 4.30. *ถ้า $E\left[|X|^k\right]<\infty$ สำหรับค่าจำนวนเต็มบวก k แล้ว $E\left[|X|^j\right]<\infty$ สำหรับค่าจำนวนเต็มบวก j ใดๆ ที่ $j<k$*

การพิสูจน์. กำหนดให้ j และ k โดยที่ $j<k$

$$\begin{aligned}
E\left[|X|^j\right] &= \int_{-\infty}^{\infty}|x|^j f(x)\,dx = \int_{|x|\leqslant 1}|x|^j f(x)\,dx + \int_{|x|>1}|x|^j f(x)\,dx \\
&\leqslant \int_{|x|\leqslant 1} f(x)\,dx + \int_{|x|>1}|x|^j f(x)\,dx \leqslant \int_{|x|\leqslant 1} f(x)\,dx + \int_{|x|>1}|x|^k f(x)\,dx \\
&\leqslant \int_{|x|\leqslant 1} f(x)\,dx + \int_{|x|>1}|x|^k f(x)\,dx + \int_{|x|\leqslant 1}|x|^k f(x)\,dx \\
&= Pr(|X|\leqslant 1) + E\left[|X|^k\right] < \infty
\end{aligned}$$

■

บทนิยามที่ 4.13. *กำหนดให้ X เป็นตัวแปรสุ่ม สำหรับค่าจำนวนจริง t ใดๆ นิยามฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ X เป็น*

$$\psi(t) = E\left[e^{tX}\right] \tag{4.54}$$

ทฤษฎีบทต่อไปนี้นำเสนอวิธีการคำนวณหาค่าโมเมนต์ต่างๆ จากฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.)

ทฤษฎีบทที่ 4.31. *กำหนดให้ X เป็นตัวแปรสุ่มที่ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) มีค่าจำกัดทุกๆ ค่า t ที่อยู่รอบๆ ศูนย์ (t in some open interval around zero) ดังนั้น สำหรับค่าจำนวนเต็มบวก n ใดๆ โมเมนต์ที่ n ของ X มีค่าเท่ากับอนุพันธ์ลำดับที่ n ของฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) $\psi(t)$ ที่ $t=0$ นั่นคือ*

$$E\left[X^n\right] = \psi^n(0) \tag{4.55}$$

โดยที่ $\psi^n(0)$ หมายถึงอนุพันธ์ลำดับที่ n ของฟังก์ชัน $\psi(t)$ ที่ $t=0$

การพิสูจน์. เริ่มจากการประยุกต์ใช้หลักการที่ว่าการหาอนุพันธ์และการอินทิเกรตสามารถสลับที่ได้ นั่นคือ

$$\frac{d}{dt}E\left[r\left(X,t\right)\right]=\frac{d}{dt}\int_{-\infty}^{\infty}r\left(x,t\right)f\left(x\right)dx=\int_{-\infty}^{\infty}\left[\frac{d}{dt}r\left(x,t\right)\right]f\left(x\right)dx=E\left[\frac{d}{dt}r\left(X,t\right)\right]$$

ดังนั้น ในกรณีที่ $r\left(x,t\right)=e^{tx}$ จะได้ว่า สำหรับกรณีที่ $n=1$

$$\frac{d}{dt}E\left[e^{tX}\right]\Big|_{t=0}=\int_{-\infty}^{\infty}\left[\frac{d}{dt}e^{tx}\Big|_{t=0}\right]f\left(x\right)dx=\int_{-\infty}^{\infty}xf\left(x\right)dx=E\left[X\right]$$

ในทำนองเดียวกัน สามารถหาอนุพันธ์ไปเรื่อยๆ จำนวน n ครั้งจะได้ว่า

$$\frac{d^{n}}{dt^{n}}E\left[e^{tX}\right]\Big|_{t=0}=E\left[\frac{d^{n}}{dt^{n}}e^{tX}\Big|_{t=0}\right]=\int_{-\infty}^{\infty}x^{n}f\left(x\right)dx=E\left[X^{n}\right]$$

■

ทฤษฎีบทที่ 4.32. *กำหนดให้ X เป็นตัวแปรสุ่มที่ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) เท่ากับ $\psi_{x}\left(t\right)$ และ $Y=aX+b$ สำหรับจำนวนจริง a และ b ใดๆ แล้ว ตัวแปรสุ่มที่ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ Y เท่ากับ*

$$\psi_{y}\left(t\right)=e^{bt}\psi_{x}\left(at\right) \tag{4.56}$$

การพิสูจน์.

$$\psi_{y}\left(t\right)=E\left[e^{tY}\right]=E\left[e^{t(aX+b)}\right]=e^{bt}E\left[e^{(at)X}\right]=e^{bt}\psi_{x}\left(at\right)$$

■

ทฤษฎีบทต่อไปนี้แสดงถึงวิธีการหาฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่มที่เกิดจากการบวกกันของตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent)

ทฤษฎีบทที่ 4.33. *สมมุติให้ $X_{1},\ldots,X_{n}$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) ที่มีฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของแต่ละตัวแปรเท่ากับ $\psi_{i}\left(t\right)$ สำหรับ $i=1,\ldots,n$ และกำหนดให้ $Y=\sum_{i=1}^{n}X_{i}$ ดังนั้น ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ Y เท่ากับ*

$$\psi_{Y}\left(t\right)=\prod_{i=1}^{n}\psi_{i}\left(t\right) \tag{4.57}$$

บทที่ 5

การแจกแจงที่ได้รับความนิยม (Popular Distributions)

บทนี้นำเสนอตัวอย่างตัวแปรสุ่มและการแจกแจงที่ได้รับความนิยม พร้อมทั้งตัวอย่างการประยุกต์ใช้ในปัญหาต่างๆ โดยพิจารณาระดับความนิยมจากการประยุกต์ใช้ในการวิเคราะห์ข้อมูลด้านเศรษศาสตร์และการเงินเป็นหลัก ดังนั้น จึงเริ่มจากการแจกแจงปกติ (normal distribution) ซึ่งได้รับความนิยมสูงสุด

5.1 การแจกแจงปกติ (Normal Distribution)

การแจกแจงปกติ (normal distribution) เป็นหนึ่งในรูปแบบของการแจกแจงที่ได้รับความนิยมสูงสุด ซึ่งมีเหตุผลหลักสามประการดังนี้ (1) คุณสมบัติทางสถิติที่ทำให้สะดวกต่อการวิเคราะห์ เช่น ผลรวมของตัวแปรสุ่มที่แต่ละตัวมีการแจกแจงปกติจะมีการแจกแจงปกติ การแจกแจงปกติสามารถอธิบายได้ด้วยพารามิเตอร์เพียงสองตัวคือ ค่าความคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 เป็นต้น (2) ทฤษฎีบทลิมิตของค่ากลาง (Central Limit Theorem) ซึ่งจะนำเสนอในบทที่ XXX ได้ระบุว่าค่าเฉลี่ยของตัวอย่างจะลู่เข้าสู่ (converge) การแจกแจงปกติ (normal distribution) เมื่อจำนวนตัวอย่างมากขึ้นเรื่อยๆ ทำให้การแจกแจงของตัวประมาณค่า (estimator) หรือตัวทำนาย (predictor) ส่วนใหญ่จะอยู่ในรูปแบบที่เกี่ยวข้องกับการแจกแจงปกติ ส่วนสุดท้ายเป็นผลมาจากการสังเกตในธรรมชาติว่า ปรากฏการณ์จำนวนไม่น้อยที่นำไปสู่การแจกแจงที่มีลักษณะคล้ายกับการแจกแจงปกติ (normal distribution) ยกตัวอย่างเช่น XXX KEI CAN WE GET NORMAL DISTRIBUTIONF ROM THE RETURNS DATA?

การแจกแจงปกติที่มีความพิเศษคือการแจกแจงปกติมาตรฐาน (standard normal distribution) ซึ่งมีฟังก์ชัน

ความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{5.1}$$

ในขณะเดียวกัน ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่ม Z ที่มีการแจกแจงปกติมาตรฐาน (standard normal distribution) เท่ากับ

$$\begin{aligned}
\psi_z(t) = E\left[e^{tZ}\right] &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \int_{-\infty}^{\infty} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(z^2-2tz+t^2\right)}{2}} dz = \int_{-\infty}^{\infty} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz \\
&= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} d\tilde{z} = e^{\frac{t^2}{2}}
\end{aligned}$$

โดยที่ในขั้นตอนก่อนสุดท้ายใช้เทคนิคการเปลี่ยนตัวแปร $\tilde{z} = z - t$

เมื่อประยุกต์ใช้ทฤษฎีบทที่ จะสามารถหาค่าความคาดหมาย (mean) และค่าความแปรปรวน (variance) ของการแจกแจงปกติมาตรฐาน (standard normal distribution) ได้เป็น

$$E[Z] = \frac{d}{dt}\psi_z(t)\bigg|_{t=0} = \frac{d}{dt} e^{\frac{t^2}{2}}\bigg|_{t=0} = te^{\frac{t^2}{2}}\bigg|_{t=0} = 0 \tag{5.2}$$

$$Var[Z] = E\left[Z^2\right] = \frac{d^2}{dt^2} e^{\frac{t^2}{2}}\bigg|_{t=0} = \left[e^{\frac{t^2}{2}} + t^2 e^{\frac{t^2}{2}}\right]\bigg|_{t=0} = 1 \tag{5.3}$$

นั่นคือ การแจกแจงปกติมาตรฐาน (standard normal distribution) คือการแจกแจงแบบปกติที่มีค่าความคาดหมายเท่ากับศูนย์และค่าความแปรปรวนเท่ากับหนึ่ง ซึ่งมักเขียนแทนการแจกแจงแบบนี้ด้วย $N(0,1)$

ยิ่งไปกว่านั้น เราสามารถสร้างตัวแปรสุ่ม X ที่มีการการแจกแจงปกติ (normal distribution) ที่มีค่าความคาดหมาย (mean) เท่ากับ μ และค่าความแปรปรวน (variance) เท่ากับ σ^2 ได้จากตัวแปรสุ่ม Z โดยใช้ความสัมพันธ์เชิงเส้นต่อไปนี้

$$X = \sigma Z + \mu \tag{5.4}$$

เมื่อประยุกต์ใช้สมการที่ 4.56 จะพบว่า ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่ม X เท่ากับ

$$\psi_x(t) = e^{\mu t} e^{\frac{(\sigma t)^2}{2}} = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \tag{5.5}$$

ซึ่งสามารถใช้ในการคำนวณหาค่าความคาดหมาย (mean) และค่าความแปรปรวน (variance) ได้ดังต่อไปนี้

$$\begin{aligned}
E[X] &= \frac{d}{dt}\psi_x(t)\bigg|_{t=0} = \frac{d}{dt} e^{\mu t + \frac{1}{2}\sigma^2 t^2}\bigg|_{t=0} = \left(\mu + t\sigma^2\right) e^{\mu t + \frac{1}{2}\sigma^2 t^2}\bigg|_{t=0} = \mu \\
Var[X] &= E\left[X^2\right] - \mu^2 = \frac{d^2}{dt^2} e^{\mu t + \frac{1}{2}\sigma^2 t^2}\bigg|_{t=0} - \mu^2 \\
&= \left[\sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \left(\mu + t\sigma^2\right)^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]\bigg|_{t=0} = \sigma^2 + \mu^2 - \mu^2 = \sigma^2
\end{aligned}$$

นั่นคือ ตัวแปรสุ่ม $X = \sigma Z + \mu$ มีการการแจกแจงปกติ (normal distribution) ที่ค่าความคาดหมาย (mean) เท่ากับ μ และค่าความแปรปรวน (variance) เท่ากับ σ^2 ซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เท่ากับ

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.6}$$

โดยมักเขียนแทนการแจกแจงแบบนี้ด้วย $N(\mu, \sigma^2)$

ในทางกลับกัน หากทราบว่าตัวแปรสุ่ม X มีการการแจกแจงปกติ (normal distribution) ที่ค่าความคาดหมาย (mean) เท่ากับ μ และค่าความแปรปรวน (variance) เท่ากับ $\sigma^2 \neq 0$ แล้วก็จะทราบทันทีว่าตัวแปรสุ่ม $Z = \frac{X-\mu}{\sigma}$ จะมีการแจกแจงปกติมาตรฐาน (standard normal distribution) การแปลงค่ารูปแบบนี้คือการปรับคะแนนมาตรฐาน (standardized score) ซึ่งนิยมใช้ในการปรับค่าผลลัพธ์หรือผลการทดสอบให้อยู่ในรูปของตัวแปรที่มีการแจกแจงปกติมาตรฐาน ยกตัวอย่างเช่น XXX NEED TO ASK NICK FOR REFERENCES

ตัวอย่างที่ 5.1. แบบจำลองทางการเงินมักสมมุติให้มูลค่าของหลักทรัพย์ ณ เวลา t เป็นไปตามความสัมพันธ์ต่อไปนี้

$$V(t) = V_0 e^{tr}$$

โดยที่ V_0 คือมูลค่าของหลักทรัพย์ ณ เวลาที่ $t = 0$ และ r คืออัตราผลตอบแทนสุทธิ (net return) ของหลักทรัพย์ดังกล่าว[1] ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงปกติ (normal distribution) โดยที่ค่าคาดหมายมีค่าเท่ากับ μ และค่าความเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งในทางการเงินมักเรียกว่า ความผันผวน (volatility) มีค่าเท่ากับ σ

ค่าคาดหมาย (mean) ของมูลค่าของหลักทรัพย์ ณ เวลา t มีค่าเท่ากับ

$$E[V(t)] = E\left[V_0 e^{tr}\right] = V_0 E\left[e^{tr}\right]$$

สังเกตได้ว่า $E\left[e^{tr}\right]$ คือฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ $N(\mu, \sigma^2)$ ซึ่งมีค่าเท่ากับ $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ (ดูสมการที่ 5.5 ประกอบ) ดังนั้น ค่าคาดหมาย (mean) ของมูลค่าของหลักทรัพย์ ณ เวลา t มีค่าเท่ากับ

$$E[V(t)] = V_0 e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

[1]โดยปกติอัตราผลตอบแทนของหลักทรัพย์ต่อหนึ่งหน่วยเวลาสามารถเขียนในรูปของล็อกการิทึมได้เป็น

$$r = \frac{1}{t} \ln \frac{V(t)}{V_0} = \frac{1}{t} \ln \left(1 + \frac{V(t) - V_0}{V_0}\right) \approx \frac{1}{t}\left[\frac{V(t) - V_0}{V_0}\right]$$

□

ในทำนองเดียวกันกับฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) นักสถิติได้กำหนดสัญลักษณ์พิเศษ $\Phi(x)$ เพื่อแทนฟังก์ชันความนาจะเป็นสะสม (C.D.F.) ของการแจกแจงปกติมาตรฐาน (standard normal distribution) ซึ่งมีนิยามดังต่อไปนี้

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)\, dx, \text{ สำหรับ } -\infty < z < \infty \tag{5.7}$$

โดยที่ $\phi(x)$ คือฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของการแจกแจงปกติมาตรฐานดังแสดงในสมการที่ 5.1 ข้อสังเกตที่สำคัญอันหนึ่งคือ การอินทิเกรตในสมการที่ 5.7 นั้นไม่มีคำตอบที่เป็นรูปแบบปิด (closed form) ทำให้ต้องใช้การประมาณ (approximation) ซึ่งในอดีตมักแสดงในรูปของตาราง แต่ในปัจจุบันสามารถหาได้จากโปรแกรมคอมพิวเตอร์ทางสถิติทั่วไป เหตุผลเดียวกันนี้ทำให้ เรามักเขียนแทนผลลัพธ์ที่เกี่ยวข้องกับการแจกแจงปกติโดยใช้สัญลักษณ์ $\phi(x)$ หรือ $\Phi(x)$ โดยตรง ทั้งนี้เพราะไม่มีรูปแบบฟังก์ชันอย่างง่ายที่ใช้แทนได้

ทฤษฎีบทต่อไปนี้กล่าวถึงคุณสมบัติความสมมาตรของการแจกแจงปกติในรูปแบบของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.)

ทฤษฎีบทที่ 5.1 (ความสมมาตรของการแจกแจงปกติ)**.** *สำหรับค่าจำนวนจริง z ใดๆ*

$$\Phi(-z) = 1 - \Phi(z) \tag{5.8}$$

และสำหรับ $0 < p < 1$

$$\Phi^{-1}(p) = -\Phi^{-1}(1-p) \tag{5.9}$$

การพิสูจน์. เริ่มจากความสัมพันธ์ที่ว่า

$$Pr(Z \leqslant -z) + Pr(Z > -z) = 1$$

เนื่องจากการแจกแจงปกติเป็นการแจกแจงที่สมมาตร (symmetric) ทำให้ $Pr(Z > -z) = Pr(Z \leqslant z)$ (ดูรูปที่ XXX ประกอบ) ดังนั้น

$$Pr(Z \leqslant -z) + Pr(Z \leqslant z) = 1 \Rightarrow \Phi(-z) = 1 - \Phi(z)$$

ส่วนสมการที่ 5.9 พิสูจน์ได้ด้วยการแทนค่า $z = \Phi^{-1}(p)$ และส่วนกลับของมัน $\Phi(z) = p$ ลงในด้านซ้ายและด้านขวาสมการที่ 5.8 ตามลำดับ ดังต่อไปนี้

$$\Phi\left(-\Phi^{-1}(p)\right) = 1 - p \Rightarrow -\Phi^{-1}(p) = \Phi^{-1}(1-p)$$

โดยที่ในขั้นตอนสุดท้ายเป็นผลมาจากการดำเนินการด้วยฟังก์ชันส่วนกลับ Φ^{-1} ทั้งสองข้างของสมการก่อนหน้านั้น ■

สมการที่ 5.8 และ 5.9 สามารถทำความเข้าใจได้ไม่ยากโดยอาศัยรูปภาพดังต่อไปนี้ จากรูปที่ XXX จะเห็นได้ว่า พื้นที่แรเงาด้านซ้ายซึ่งมีค่าเท่ากับ $\Phi(-z)$ มีขนาดเท่ากันกับพื้นที่แรเงาด้านขวาซึ่งมีค่าเท่ากับ $1-\Phi(z)$ ทั้งนี้เป็นผลมาจากความสมมาตรของการแจกแจงปกติ ในทำนองเดียวกัน สมมุติว่าพื้นที่แรเงาในรูปที่ XXX มีค่าเท่ากับ p ดังนั้น จุดด้านขวาของพื้นที่แรเงามีค่าเท่ากับ $\Phi^{-1}(p)$ ส่วนพื้นที่ที่เหลือมีขนาดเท่ากับ $1-p$ ในขณะเดียวกัน เราสามารถประยุกต์ใช้คุณสมบัติความสมมาตรของการแจกแจงปกติเพื่อสร้างพื้นที่ที่เท่ากันทั้งด้านซ้ายและด้านขวา ดังแสดงในรูปที่ XXX ซึ่งจะเห็นได้ว่า จุดขอบของพื้นที่แรเงาของทั้งสองด้านจะเป็นจำนวนตรงข้ามกัน นั่นคือ หากจุดด้านขวามีค่าเท่ากับ $z=\Phi^{-1}(p)$ จุดด้านซ้ายก็จะมีค่าเท่ากับ $-\Phi^{-1}(p)$ ในขณะเดียวกัน เมื่อพิจารณาโดยใช้นิยามของ จุดด้านซ้ายจะมีค่าเท่ากับ $\Phi^{1}(1-p)$ ดังนั้นจึงสรุปได้ว่า $\Phi^{-1}(p)=-\Phi^{-1}(1-p)$

ADD FIGURES HERE

ทฤษฎีบทต่อไปนี้แสดงวิธีการแปลงตัวแปรสุ่มที่มีการแจกแจงปกติเป็นตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐาน ซึ่งช่วยให้การประยุกต์ใช้การแจกแจงปกติมาตรฐานในการทดสอบสมมุติฐาน (อภิปรายอย่างละเอียดต่อไปในบทที่ XXX) มีความสะดวกมากยิ่งขึ้น

ทฤษฎีบทที่ 5.2. *กำหนดให้ X เป็นตัวแปรสุ่มที่มีแจกแจงปกติ (normal distribution) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 และกำหนดให้ F แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ X ดังนั้น ตัวแปรสุ่ม $Z=\frac{X-\mu}{\sigma}$ มีการแจกแจงปกติมาตรฐาน (standard normal distribution) โดยที่สำหรับค่าจำนวนจริง x ใดๆ*

$$F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right) \tag{5.10}$$

และสำหรับค่าจำนวนจริง $0<p<1$ ใดๆ

$$F^{-1}(p)=\mu+\sigma\Phi^{-1}(p) \tag{5.11}$$

การพิสูจน์. เริ่มจากนิยามของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ X

$$F(x)=Pr(X\leqslant x)=Pr\left(\frac{X-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right)=Pr\left(Z\leqslant\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

ส่วนที่สองพิสูจน์ได้โดยเริ่มจากการกำหนดให้ $p=F(x)$ แล้วประยุกต์ใช้ฟังก์ชันส่วนกลับ (inverse function) ดังต่อไปนี้

$$F(x)=p=\Phi\left(\frac{x-\mu}{\sigma}\right)\Rightarrow\Phi^{-1}(p)=\frac{x-\mu}{\sigma}\Rightarrow x=\mu+\sigma\Phi^{-1}(p)$$

เนื่องจาก $p=F(x)$ ดังนั้น $x=F^{-1}(p)$ ซึ่งทำให้สรุปได้ว่า

$$F^{-1}(p)=\mu+\sigma\Phi^{-1}(p)$$

■

ตัวอย่างที่ 5.2. สมมุติให้ตัวแปรสุ่ม X มีการแจกแจงแบบปกติ ด้วยด้วยค่าคาดหมาย (mean) $\mu = 10$ และค่าความแปรปรวน (variance) $\sigma^2 = 5$ คำถามก็คือ ความน่าจะเป็นที่ X จะมีค่าอยู่ในช่วง $5 < X < 15$ มีค่าเท่าใด?

เริ่มจากการแปลงตัวแปรสุ่มให้เป็นแบบปกติมาตรฐานด้วยฟังก์ชัน $Z = \frac{X-\mu}{\sigma}$

$$Pr\left(5 < X < 15\right) = Pr\left(\frac{5-10}{5} < \frac{X-10}{5} < \frac{15-10}{5}\right) = Pr\left(-1 < Z < 1\right)$$

ซึ่งสามารถเขียนในรูปของ Φ ได้เป็น

$$Pr\left(-1 < Z < 1\right) = Pr\left(|Z| < 1\right) = \Phi\left(1\right) - \Phi\left(-1\right) \approx 0.6826$$

ดูรูปที่ XXX ประกอบ □

ทฤษฎีบทต่อไปนี้กล่าวถึงคุณสมบัติที่สำคัญอันหนึ่งของการแจกแจงปกติ (normal distribution) นั่นคือ ฟังก์ชันเชิงเส้น (linear) ของตัวแปรสุ่มปกติจะยังมีการแจกแจงปกติ

ทฤษฎีบทที่ 5.3. *ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 และ $Y = aX + b$ สำหรับค่าคงที่ a และ b โดยที่ $a \neq 0$ แล้ว Y จะมีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) $a\mu + b$ และค่าความแปรปรวน (variance) $a^2\sigma^2$*

การพิสูจน์. ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่ม X เท่ากับ

$$\psi_x\left(t\right) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

ดังนั้น ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรสุ่ม $Y = aX + b$ เท่ากับ

$$\psi_y\left(t\right) = e^{bt}\psi_x\left(at\right) = e^{bt}e^{\mu at + \frac{1}{2}\sigma^2 a^2 t^2} = e^{(a\mu+b)t + \frac{1}{2}\left(a^2\sigma^2\right)t^2}$$

ซึ่งเป็นฟังก์ชันที่ตรงกับรูปแบบของฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของการแจกแจงปกติ ที่มีค่าคาดหมาย (mean) $a\mu + b$ และค่าความแปรปรวน (variance) $a^2\sigma^2$ ■

ทฤษฎีบทที่ 5.4. *ถ้าตัวแปรสุ่ม $X_1, \ldots, X_n$ เป็นอิสระต่อกัน และ X_i มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) μ_i และค่าความแปรปรวน (variance) σ_i^2 สำหรับ $i = 1, \ldots, n$ แล้ว $Y = \sum_{i=1}^n a_i X_i + b$ มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) $\sum_{i=1}^n a_i\mu_i + b$ และค่าความแปรปรวน (variance) $\sum_{i=1}^n a_i^2\sigma_i^2$*

การพิสูจน์. ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ X_i เท่ากับ

$$\psi_i(t) = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}$$

เนื่องจากตัวแปรสุ่มเหล่านี้เป็นอิสระต่อกัน ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ Y เท่ากับ

$$\psi_Y(t) = \prod_{i=1}^{n} e^{bt}\psi_i(a_i t) = \prod_{i=1}^{n} e^{bt} e^{\mu_i a_i t + \frac{1}{2}\sigma_i^2 a_i^2 t^2} = e^{\left(\sum_{i=1}^{n} a_i\mu_i + b\right)t + \frac{1}{2}\left(\sum_{i=1}^{n} a_i^2\sigma_i^2\right)t^2}$$

ซึ่งบ่งบอกว่า Y มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) $\sum_{i=1}^{n} a_i\mu_i + b$ และค่าความแปรปรวน (variance) $\sum_{i=1}^{n} a_i^2\sigma_i^2$ ■

ผลที่ตามมาจากทฤษฎีบทนี้ก็คือ ค่าเฉลี่ยของตัวอย่างสุ่ม (random sample) ที่สุ่มมาจากการแจกแจงปกติ (normal distribution) จะมีการแจกแจงปกติ

บทนิยามที่ 5.1. *กำหนดให้ $X_1, \ldots, X_n$ เป็นตัวแปรสุ่ม ค่าคาดหมายของตัวอย่าง (sample mean) นิยามได้เป็น*

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \tag{5.12}$$

หากพิจารณาจากข้อมูล (observed data) n ตัวอย่าง ค่าเฉลี่ย (average) นั้นหมายถึง $\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n}$ โดยที่ x_i คือค่าที่เกิดขึ้นจริง (realized value) ของตัวแปรสุ่ม X_i หรือค่าตัวแปรของตัวอย่าง $1, \ldots, n$ กล่าวคือ ค่าเฉลี่ย (average) $\bar{x}_n$ คือสำเนาจากตัวอย่าง (sample counterpart) ของค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ สิ่งที่แตกต่างกันก็คือ ค่าเฉลี่ย (average) $\bar{x}_n$ เป็นจำนวนจริงค่าหนึ่ง ส่วนค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ เป็นตัวแปรสุ่ม

ทฤษฎีบทที่ 5.5. *สมมุติให้ $X_1, \ldots, X_n$ เป็นตัวแปรสุ่มที่นำไปสู่ตัวอย่างสุ่ม (random sample) ขนาด n และแต่ละตัวเป็นตัวแปรสุ่มที่มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 ดังนั้น ค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) $\frac{\sigma^2}{n}$ นั่นคือ*

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{5.13}$$

การพิสูจน์. ทฤษฎีบทนี้เป็นผลมาจากการประยุกต์ใช้ทฤษฎีบทที่ 5.4 โดยกำหนดให้ $a_i = \frac{1}{n}$ และ $b = 0$ ซึ่งช่วยให้สรุปได้ว่า $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) $\sum_{i=1}^{n} \frac{1}{n}\mu = \mu$ และค่าความแปรปรวน (variance) $\sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}$ ■

ตัวอย่างต่อไปนี้ แสดงวิธีการคำนวณหาขนาดกลุ่มตัวอย่างเพื่อทำให้เกิดความมั่นใจได้ว่าค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ จะแตกต่างจากค่าคาดหมาย (mean) μ ไม่มากเกินขอบเขตที่กำหนด บทเรียนที่สำคัญจากตัวอย่างนี้คือ ขนาดตัวอย่างนั้นขึ้นอยู่กับระดับความน่าจะเป็นที่ต้องการและขนาดความคาดเคลื่อนที่สนใจ จึงไม่มีสูตรตายตัวที่จะบอกได้ทุกกรณี

ตัวอย่างที่ 5.3. สมมุติว่า สุ่มเลือกกลุ่มตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง มาจากการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 คำถามก็คือ n ควรมีขนาดเท่าใดจึงจะทำให้

$$Pr\left(|\overline{X}_n - \mu| \leqslant c\right) \geqslant 0.95$$

โดยที่ c คือค่าจำนวนจริงบวกใดๆ กล่าวคือ เราต้องการหาขนาดของกลุ่มตัวอย่างที่จะทำให้มั่นใจได้ว่า ค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ จะแตกต่างจากค่าคาดหมาย (mean) μ ไม่เกินค่า c โดยกำหนดระดับความมั่นใจด้วยความน่าจะเป็นที่ไม่น้อยกว่า 0.95

ก่อนอื่นต้องทำการแปลงตัวแปรสุ่ม $\overline{X}_n$ ให้อยู่ในรูปของตัวแปรสุ่มปกติมาตรฐาน (standard normal) โดย

$$Z = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right)$$

ดังนั้น

$$Pr\left(|\overline{X}_n - \mu| \leqslant c\right) = Pr\left(|\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right)| \leqslant \frac{\sqrt{n}}{\sigma}c\right) = Pr\left(|Z| \leqslant \frac{\sqrt{n}}{\sigma}c\right)$$

เมื่อพิจารณาจากฟังก์ชันควอนไทล์ของการแจกแจงปกติมาตรฐาน (standard normal) พบว่า $Pr\left(|Z| \leqslant 1.96\right) \geqslant 0.95$ แต่เพื่อความสะดวกในการจดจำ อาจจะเลือกใช้ 2 แทนค่า 1.96 ก็ได้ ดังนั้น $Pr\left(|\overline{X}_n - \mu| \leqslant c\right) \geqslant 0.95$ ก็ต่อเมื่อ

$$\frac{\sqrt{n}}{\sigma}c \geqslant 2 \Rightarrow n \geqslant 4\frac{\sigma^2}{c^2}$$

หากเลือก $c = 0.1\sigma$ (ร้อยละ 10 ของค่าเบี่ยงเบนมาตรฐาน) ก็จะได้ว่าขนาดตัวอย่างขั้นต่ำในกรณีนี้เท่ากับ $n = 400$ ตัวอย่าง ในขณะที่หากเลือก $c = 0.05\sigma$ (ต้องการให้แม่นยำมากขึ้น) ก็จะได้ว่า ขนาดตัวอย่างขั้นต่ำในกรณีนี้เท่ากับ $n = 1,600$ ตัวอย่าง

□

นอกจากนี้ ตัวอย่างด้านบนนี้ยังได้ชี้ให้เห็นว่า เมื่อพิจารณาเกี่ยวกับค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ ก็ต้องใช้ค่าความแปรปรวนของค่าคาดหมายของตัวอย่างซึ่งมีค่าเท่ากับ $\frac{\sigma^2}{n}$ ไม่ใช่ค่าความแปรปรวนของกลุ่มตัวอย่าง

คุณสมบัติอีกอันหนึ่งที่สำคัญของการแจกแจงปกติที่เชื่อมโยงกับการแปลงค่าเป็นการแจกแจงปกติมาตรฐาน (standard normal) ก็คือ การที่สามารถใช้ค่าเบี่ยงเบนมาตรฐาน σ เป็นหน่วยในการวัดเพื่อบอกค่าความน่าจะเป็น ยกตัวอย่างเช่น ความน่าจะเป็นที่ค่าคาดหมายของตัวอย่าง (sample mean) $\overline{X}_n$ จะแตกต่างจากค่าคาดหมาย (mean) หนึ่งหน่วยของค่าเบี่ยงเบนมาตรฐาน σ เท่ากับ $Pr\left(|X-\mu| \leqslant 2\sigma\right) = Pr\left(|Z| \leqslant 2\right) \approx 0.95$ ไม่ว่าการแจกแจงนั้นจะมีค่าค่าเบี่ยงเบนมาตรฐานเท่ากับเท่าใด ในทำนองเดียวกัน $Pr\left(|X-\mu| \leqslant 3\sigma\right) = Pr\left(|Z| \leqslant 2\right) \approx 0.997$ คุณสมบัตินี้ช่วยให้การประยุกต์ใช้การแจกแจงปกติ (และการแจกแจงที หรือ t-distribution) ในการทดสอบสมมุติฐานทำได้โดยง่าย ซึ่งโดยทั่วไป มักจะใช้งานในลักษณะของเกณฑ์มาตรฐาน (benchmark) คือ ที่สองเท่าของค่าเบี่ยงเบนมาตรฐาน 2σ จะหมายถึงประมาณร้อยละ 95 ส่วนที่สามเท่าของค่าเบี่ยงเบนมาตรฐาน 3σ จะหมายถึงมากกว่าร้อยละ 99 เกณฑ์มาตรฐานแบบนี้ช่วยให้นักวิเคราะห์สามารถอ่านผลการประมาณค่าได้อย่างรวดเร็วโดยไม่ต้องพึ่งตารางสถิติ ถึงแม้ว่าอาจจะเป็นเพียงค่าประมาณก็ตาม

การแจกแจงอันหนึ่งที่เชื่อมโยงกับการแจกแจงปกติคือ การแจกแจงปกติด้วยล็อก (lognormal distribution) ซึ่งหมายถึงการแจกแจงของตัวแปรสุ่มที่ล็อกการิธึมของตัวแปรนั้นมีการแจกแจงปกติ (normal distribution)

บทนิยามที่ 5.2. *ถ้า* $\ln X$ *มีการแจกแจงปกติ (normal distribution) ด้วยค่าคาดหมาย (mean)* μ *และค่าความแปรปรวน (variance)* σ^2 *แล้ว ตัวแปรสุ่ม* X *จะมีการแจกแจงปกติด้วยล็อก (lognormal distribution) ที่มีพารามิเตอร์* μ *และ* σ^2

ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่ม X ที่มีการแจกแจงปกติด้วยล็อก (lognormal distribution) สำหรับพารามิเตอร์ μ และ σ^2 เท่ากับ

$$f\left(x\right) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\ln x-\mu\right)^2}{2\sigma^2}} \text{ สำหรับ } x > 0, \\ 0, \text{ สำหรับกรณีอื่น} \end{cases} \tag{5.14}$$

ส่วนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เท่ากับ

$$F\left(x\right) = Pr\left(X \leqslant x\right) = Pr\left(\frac{\ln X-\mu}{\sigma} \leqslant \frac{\ln x-\mu}{\sigma}\right) = \Phi\left(\frac{\ln x-\mu}{\sigma}\right) \tag{5.15}$$

ส่วนการหาค่าคาดหมาย (mean) และค่าความแปรปรวน (variance) นั้นมีเทคนิคหนึ่งที่ช่วยให้คำนวณหาค่าได้ง่ายคือ การใช้ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของตัวแปรที่มีการแจกแจงปกติ ซึ่งในที่นี้หมายถึง $Y = \ln X$

$$\psi_Y\left(t\right) = E\left[e^{tY}\right] = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$$

โดยที่สมการสุดท้ายเป็นผลมาจากการที่ $Y = \ln X$ มีการแจกแจงปกติ (normal distribution) ด้วยค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) σ^2 ในขณะเดียวกัน เราสามารถเขียนฟังก์ชันก่อกำเนิด

โมเมนต์ (m.g.f.) ในรูปของ X ได้เป็น

$$\psi_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\ln X}\right] = E\left[X^t\right]$$

ดังนั้น ค่าคาดหมายของ X เท่ากับ

$$E[X] = \psi_Y(1) = e^{\mu+\frac{1}{2}\sigma^2}$$

ส่วนค่าความแปรปรวนของ X เท่ากับ

$$\begin{aligned} Var[X] &= E\left[X^2\right] - E[X]^2 = \psi_Y(2) - \psi_Y(1)^2 = e^{2\mu+2\sigma^2} - \left(e^{\mu+\frac{1}{2}\sigma^2}\right)^2 \\ &= e^{2\mu+\sigma^2}\left[e^{\sigma^2} - 1\right] \end{aligned}$$

ตัวอย่างต่อไปนี้แสดงการประยุกต์ใช้การแจกแจงปกติด้วยล็อกในการกำหนดราคาอนุพันธ์ (option pricing) หรือที่รู้จักกันภายใต้ชื่อ กำหนดราคาอนุพันธ์ของ Black and Scholes (1973) ซึ่งสมมุติให้การแจกแจงของราคาหลักทรัพย์พื้นฐานในแต่ละช่วงเวลาเป็นการแจกแจงปกติด้วยล็อก (lognormal distribution) ผู้อ่านสามารถศึกษาเพิ่มเติมเกี่ยวกับการกำหนดราคาอนุพันธ์ได้จาก Hull (2010)

ตัวอย่างที่ 5.4. สมมุติให้ราคาหลักทรัพย์ ณ เวลา t, S_t, มีการแจกแจงปกติด้วยล็อก (lognormal distribution) นั่นคือ $S_t = S_0 e^{Z_t}$ โดยที่ Z_t มีการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมาย (mean) μt และค่าความแปรปรวน (variance) $\sigma^2 t$

เพื่อความสะดวกในการคำนวณ สามารถเขียน Z_t ในรูปของตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐาน (standard normal distribution) Z ได้เป็น

$$Z_t = \mu t + \sigma\sqrt{t}Z$$

ดังนั้น สามารถเขียนราคาหลักทรัพย์ในรูปการแจกแจงปกติมาตรฐาน (standard normal distribution) Z ได้เป็น

$$S_t = S_0 e^{\mu t + \sigma\sqrt{t}Z}$$

พิจารณาอนุพันธ์เพื่อซื้อ (call options) ซึ่งหมายถึงอนุพันธ์ที่ให้สิทธิแก่ผู้ซื้อในการซื้อหลักทรัพย์ที่ราคาอ้างอิง (strike price) K ณ เวลาที่กำหนด T แต่ไม่จำเป็นต้องใช้สิทธิ์นั้นถ้าไม่ต้องการ ดังนั้นมูลค่า (value) ของอนุพันธ์เพื่อซื้อ (call option) ณ เวลาที่กำหนด T มีค่าเท่ากับ

$$V(S) = \max\{S-K, 0\} = \begin{cases} S-K, \text{ ถ้า } S > K \\ 0, \text{ ถ้าเป็นอย่างอื่น} \end{cases}$$

การกำหนดราคาอนุพันธ์ (option pricing) ในที่นี้จะประยุกต์ใช้หลักการเป็นกลางต่อความเสี่ยง (risk neutrality) ซึ่งกำหนดว่า ค่าคาดหมายของมูลค่าของหลักทรัพย์ควรจะมีค่าเท่ากับผลลัพธ์ที่ได้จากการฝากเงินแบบไม่มีความเสี่ยงที่อัตราดอกเบี้ย r นั่นคือ $e^{rt}S_0 = E[S_t]$ ซึ่งมีผลทำให้สามารถสรุปได้ว่า

$$e^{rt}S_0 = E\left[S_0 e^{\mu t+\sigma\sqrt{t}Z}\right] \Rightarrow e^{rt} = e^{\mu t+\frac{\sigma^2}{2}t} \Rightarrow \mu = r - \frac{\sigma^2}{2}$$

ยิ่งไปกว่านั้น หลักการเป็นกลางต่อความเสี่ยง (risk neutrality) ยังระบุว่าราคาอนุพันธ์เพื่อซื้อ, $C(K,T)$, ควรจะมีค่าเท่ากับค่าคาดหมายของมูลค่าของอนุพันธ์ที่ปรับส่วนลด (discount) ด้วยอัตราดอกเบี้ย r นั่นคือ

$$C(K,T) = e^{-rT}E[V(S_T)] = e^{-rT}E[\max\{S_T - K, 0\}]$$

เพื่อจะคำนวณค่าคาดหมายนี้ จำเป็นต้องแยกออกเป็นสองกรณีคือ

1. กรณีที่ $S_T > K$ ซึ่งจะเป็นจริงก็ต่อเมื่อ

$$S_0 e^{\mu T+\sigma\sqrt{T}Z} > K \Rightarrow Z > \frac{\ln\frac{K}{S_0} - \mu T}{\sigma\sqrt{T}} \equiv \bar{z}_T$$

2. กรณีที่ $S_T > K$ ซึ่งจะเป็นจริงก็ต่อเมื่อ

$$Z < \frac{\ln\frac{K}{S_0} - \mu T}{\sigma\sqrt{T}} \equiv \bar{z}_T$$

ดังนั้น ราคาอนุพันธ์เพื่อซื้อ (call options) ที่สามารถใช้สิทธิ์ได้ ณ เวลา T ที่ราคาอ้างอิง (strike price) K มีค่าเท่ากับ

$$\begin{aligned} C(K,T) = e^{-rT}E[\max\{S_T - K, 0\}] &= e^{-rT}\int_{\bar{z}_T}^{\infty}\left[S_0 e^{\mu T+\sigma\sqrt{T}z} - K\right]\phi(z)\,dz \qquad (5.16)\\ &= e^{-rT+\mu T}S_0\int_{\bar{z}_T}^{\infty} e^{\sigma\sqrt{T}z}\phi(z)\,dz - Ke^{-rT}\int_{\bar{z}_T}^{\infty}\phi(z)\,dz \end{aligned}$$

พิจารณาการอินทิเกรตในพจน์แรกด้านขวา ดังนี้

$$\begin{aligned} \int_{\bar{z}_T}^{\infty} e^{\sigma\sqrt{T}z}\phi(z)\,dz &= \int_{\bar{z}_T}^{\infty} e^{\sigma\sqrt{T}z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz = e^{\frac{\sigma^2}{2}T}\int_{\bar{z}_T}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-\sigma\sqrt{T})^2}{2}}dz\\ &= e^{\frac{\sigma^2}{2}T}\int_{\bar{z}_T-\sigma\sqrt{T}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz = e^{\frac{\sigma^2}{2}T}\left[1-\Phi\left(\bar{z}_T-\sigma\sqrt{T}\right)\right] = e^{\frac{\sigma^2}{2}T}\Phi\left(\sigma\sqrt{T}-\bar{z}_T\right) \end{aligned}$$

โดยที่สมการที่สองเป็นผลมาจากการเปลี่ยนรูปให้เป็นกำลังสองสัมบูรณ์ ส่วนสมการที่สามเป็นผลมาจากการเปลี่ยนตัวแปรอินทิเกรตจาก z เป็น $z - \sigma\sqrt{T}$ และสมการสุดท้ายเป็นผลมาจากคุณสมบัติความสมมาตรของ

การแจกแจงปกติมาตรฐาน เมื่อแทนค่ากลับเข้าไปในสมการราคาอนุพันธ์เพื่อซื้อ (call options) จะได้ว่า

$$\begin{aligned} C(K,T) &= e^{-rT+\mu T} S_0 e^{\frac{\sigma^2}{2}T} \Phi\left(\sigma\sqrt{T} - \bar{z}_T\right) - Ke^{-rT}\Phi\left(-\bar{z}_T\right) \\ &= e^{-rT+\mu T+\frac{\sigma^2}{2}T} \Phi\left(\sigma\sqrt{T} - \bar{z}_T\right) - Ke^{-rT}\Phi\left(-\bar{z}_T\right) \end{aligned}$$

ดังนั้น เมื่อแทนค่า $\mu = r - \frac{\sigma^2}{2}$ จะได้ว่า

$$C(K,T) = S_0\Phi\left(\sigma\sqrt{T} - \bar{z}_T\right) - Ke^{-rT}\Phi\left(-\bar{z}_T\right) \tag{5.17}$$

โดยที่ $\bar{z}_T = \frac{\ln\frac{K}{S_0} - \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$ สมการที่ คือสมการที่มีชื่อเสียงของ Black-Scholes สำหรับการกำหนดราคาอนุพันธ์ (option pricing)

ที่จริงแล้วสมการนี้สามารถพิสูจน์ได้โดยใช้เครื่องมือแคลคูลัสแบบสุ่ม (stochastic calculus) ซึ่งเป็นผลงานชิ้นสำคัญของ Merton (1973) นอกจากนี้ Cox et al. (1979) ยังได้นำเสนอวิธีการกำหนดราคาอนุพันธ์ (option pricing) ได้สมมุติว่าการแจกแจงของราคาหลักทรัพย์พื้นฐานเป็นการแจกแจงทวินาม (binomial distribution)

□

5.2 การแจกแจงปกติหลายตัวแปร (Multivariate Normal Distribution)

ก่อนหน้านี้เราโฟกัสกับการแจกแจงปกติของตัวแปรสุ่มตัวเดียว สิ่งที่ขาดหายไปจากฟังก์ชันความหนาแน่นความน่าจะเป็น (p.d.f.) ในกรณีที่มีตัวแปรเดียว คือ ความแปรปรวน (covariance) หรือสหสัมพันธ์ (correlation) ซึ่งจะปรากฏในกรณีที่มีตัวแปรหลายตัว

เริ่มจากตัวแปรสุ่มปกติมาตราฐานที่เป็นอิสระต่อกัน ซึ่งเขียนแทนด้วยเวคเตอร์สุ่ม (random vector) $\boldsymbol{Z} = [Z_1, \ldots, Z_n]'$ โดยที่ $\boldsymbol{X}'$ หมายถึง เมทริกซ์สลับเปลี่ยน (transpose matrix) ของ $\boldsymbol{X}$ และ $Z_i \sim N(0,1)$ มีการแจกแจงปกติมาตรฐาน (standard normal) และเป็นอิสระต่อกัน (independent) นั่นคือตัวแปรสุ่มเหล่านี้มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (identically and independently distributed หรือ i.i.d.) ซึ่งในกรณีที่การแจกแจงเป็นแบบปกติมาตรฐาน สามารถเขียนแทนด้วย $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$

ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ $\boldsymbol{Z}$ คือ

$$\begin{aligned} f(\boldsymbol{z}) &= \prod_{i=1}^{n} \phi(z_i) = \prod_{i=1}^{n} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\} \end{aligned} \tag{5.18}$$

โดยที่สมการแรกเป็นผลมาจากการที่ตัวแปรทั้งหมดมีการแจกแจงปกติมาตรฐานเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) นอกจากนี้ ยังเห็นได้อย่างชัดเจนว่า ค่าคาดหมาย (mean) และเมทริกซ์ของค่าความแปรปรวนร่วม (covariance matrix) ของ $\boldsymbol{Z}$ มีค่าเท่ากับ $\mathbf{0}$ และ $\boldsymbol{I}_n$ ตามลำดับ ส่วนฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{Z}$ สามารถเขียนได้เป็น

$$\begin{aligned}\psi_{\boldsymbol{z}}(t) &= E\left[\exp\left\{\boldsymbol{Z}'\boldsymbol{t}\right\}\right] = E\left[\exp\left\{\sum_{i=1}^{n} Z_i t_i\right\}\right] = E\left[\prod_{i=1}^{n}\exp\left\{Z_i t_i\right\}\right] = \prod_{i=1}^{n} E\left[\exp\left\{Z_i t_i\right\}\right] \\ &= \prod_{i=1}^{n}\psi_z(t_i) = \prod_{i=1}^{n} e^{\frac{1}{2}t_i^2} = \exp\left\{\frac{1}{2}\sum_{i=1}^{n} t_i^2\right\} = \exp\left\{\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right\} \qquad (5.19)\end{aligned}$$

เราสามารถสร้างเวคเตอร์ของตัวแปรสุ่มปกติ (normal random vector) $\boldsymbol{X}$ ที่มีค่าคาดหมาย (mean) เท่ากับ $\boldsymbol{\mu}$ และเมทริกซ์ของค่าความแปรปรวนร่วม (covariance matrix) เท่ากับ $\boldsymbol{\Sigma}$ โดยใช้เทคนิคเดียวกันกับที่ใช้สร้างตัวแปรสุ่มตัวเดียวที่มีการแจกแจงปกติแบบทั่วไปจากการแจกแจงปกติมาตรฐาน (ดูรายละเอียดในหน้าที่ 118) เพียงแต่มีข้อยุ่งยากทางเทคนิคนิดหน่อยในการนิยามค่าความเบี่ยงเบนในรูปของเมทริกซ์ ซึ่งต้องใช้เทคนิคที่เรียกว่า การแยกส่วนสเปกตรัล (spectral decomposition) ดังต่อไปนี้

คุณสมบัติที่สำคัญอันหนึ่งของเมทริกซ์ของค่าความแปรปรวนร่วม (variance-covariance matrix) $\boldsymbol{\Sigma}$ คือเป็นเมทริกซ์จัตุรัสและสมมาตร (symmetric square matrix) คุณสมบัติเป็นบวกเกือบแน่นอน (positive semidefinite หรือ p.s.d.) ซึ่งทำให้สามารถใช้หลักการทางพีชคณิตเชิงเส้น (linear algebra) แยกส่วน $\boldsymbol{\Sigma}$ ออกได้เป็น

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma} \qquad (5.20)$$

โดยที่ $\boldsymbol{\Lambda}$ เป็นเมทริกซ์แนวทแยง (diagonal matrix) นั่นคือ

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix} \qquad (5.21)$$

และ $\boldsymbol{\Gamma}$ เป็นเมทริกซ์ตั้งฉาก (orthogonal matrix) นั่นคือ $\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Gamma}'$ ซึ่งทำให้สรุปได้ว่า $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}$ โดยทั่วไปเพื่อความสะดวก เรามักจะเรียงลำดับ λ_i จากมากไปหาน้อย นั่นคือ $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ และเรียก λ_i แต่ละตัวว่า ค่าลักษณะเฉพาะ (eigenvalue) ส่วนคอลัมน์ที่ i แต่ละอัน (สอดคล้องกับ λ_i) ซึ่งแทนด้วย $\boldsymbol{\gamma}_i$ มักเรียกว่าเวคเตอร์ลักษณะเฉพาะ (eigenvector) ของ $\boldsymbol{\Sigma}$ ยิ่งไปกว่านั้น เราสามารถเขียนเมทริกซ์ของค่าความแปรปรวนร่วม (variance-covariance matrix) $\boldsymbol{\Sigma}$ ในรูปของค่าลักษณะเฉพาะ (eigenvalue) และเวคเตอร์ลักษณะเฉพาะ

(eigenvector) ได้ดังนี้

$$\mathbf{\Sigma} = \mathbf{\Gamma}^{'}\mathbf{\Lambda}\mathbf{\Gamma} = \sum_{i=1}^{n} \lambda_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^{'} \tag{5.22}$$

ประโยชน์ที่สำคัญของการแยกส่วนนี้ก็คือ การได้มาซึ่งเมทริกซ์แนวทแยง (diagonal matrix) $\mathbf{\Lambda}$ ซึ่งค่าลักษณะเฉพาะ (eigenvalue) λ_i แต่ละค่ามีค่าไม่ติดลบ (nonnegative) ซึ่งช่วยให้สามารถนิยามรากที่สอง (square root) ของ $\mathbf{\Lambda}$ ได้เป็น

$$\mathbf{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_{n-1}} & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} \tag{5.23}$$

ซึ่งมีคุณสมบัติคล้ายกับรากกำลังสองของจำนวนจริง นั่นคือ

$$\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}} = \mathbf{\Lambda} \tag{5.24}$$

ดังนั้น จึงสามารถเขียนสมการการแยกส่วนสเปกตรัล (spectral decomposition) ใหม่ได้เป็น

$$\mathbf{\Sigma} = \mathbf{\Gamma}^{'}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Gamma} = \left(\mathbf{\Gamma}^{'}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Gamma}\right)\left(\mathbf{\Gamma}^{'}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Gamma}\right) \tag{5.25}$$

โดยที่สมการที่สองประยุกต์ใช้คุณสมบัติการตั้งฉาก (orthogonal) ของเมทริกซ์ $\mathbf{\Gamma}$ นั่นคือ $\mathbf{\Gamma}\mathbf{\Gamma}^{'} = I$ ผลที่ตามมาจากสมการที่ 5.25 คือ ทำให้สามารถนิยามรากที่สอง (square root) ของ $\mathbf{\Sigma}$ ได้เป็น

$$\mathbf{\Sigma}^{\frac{1}{2}} = \mathbf{\Gamma}^{'}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Gamma} \tag{5.26}$$

นอกจากนี้ หากเมทริกซ์ของค่าความแปรปรวนร่วม (variance-covariance matrix) $\mathbf{\Sigma}$ เป็นบวกแน่นอน (positive definite หรือ p.d.) แล้วค่าลักษณะเฉพาะ (eigenvalue) มีค่าเป็นบวกทุกค่า นั่นคือ $\lambda_i > 0$ สำหรับทุกๆ $i = 1, \ldots, n$ ซึ่งมีผลทำให้ค่าดีเทอร์มิแนนต์ (determinant) ของ $\mathbf{\Lambda}^{\frac{1}{2}}$ มีค่ามากกว่าศูนย์อย่างแน่นอน ทำให้สามารถหาเมทริกซ์ส่วนกลับ (inverse matrix) ของ $\mathbf{\Lambda}^{\frac{1}{2}}$ ได้ ดังนั้น จึงสามารถนิยามเมทริกซ์ส่วนกลับ (inverse matrix) ของ $\mathbf{\Sigma}^{\frac{1}{2}}$ ได้เป็น

$$\mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{\Gamma}^{'}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Gamma} \tag{5.27}$$

โดยที่ $\mathbf{\Lambda}^{-\frac{1}{2}}$ แทนเมทริกซ์ส่วนกลับ (inverse matrix) ของ $\mathbf{\Lambda}^{\frac{1}{2}}$

ณ จุดนี้ เราพร้อมที่จะสร้างเวคเตอร์ของตัวแปรสุ่มปกติ (normal random vector) $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ที่มีค่าคาดหมาย (mean) เท่ากับ $\boldsymbol{\mu}$ และเมทริกซ์ของค่าความแปรปรวนร่วม (covariance matrix) เท่ากับ $\boldsymbol{\Sigma}$ โดยกำหนดให้

$$\boldsymbol{X} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu} \tag{5.28}$$

ดังนั้น ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{X}$ เท่ากับ

$$\begin{aligned}\psi_{\boldsymbol{x}}(t) &= E\left[\exp\left\{\boldsymbol{X}'\boldsymbol{t}\right\}\right] = E\left[\exp\left\{\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}\right)'\boldsymbol{t}\right\}\right] = \exp\left\{\boldsymbol{\mu}'\boldsymbol{t}\right\} E\left[\exp\left\{\boldsymbol{Z}'\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right)\right\}\right] \\ &= \exp\left\{\boldsymbol{\mu}'\boldsymbol{t}\right\}\exp\left\{\frac{1}{2}\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right)'\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}\right)\right\} = \exp\left\{\boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right\}\end{aligned} \tag{5.29}$$

โดยที่สมการที่สี่ประยุกต์ใช้ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{Z}$ ในสมการที่ 5.19 โดยใช้ $\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{t}$ แทน $\boldsymbol{t}$ สังเกตด้วยว่า ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ในที่นี้สามารถหาค่าได้ตราบเท่าที่เมทริกซ์ของค่าความแปรปรวนร่วม (variance-covariance matrix) $\boldsymbol{\Sigma}$ เป็นบวกเกือบแน่นอน (p.s.d.) ในขณะที่ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ $\boldsymbol{X}$ จะหาค่าได้ก็ต่อเมื่อ $\boldsymbol{\Sigma}$ เป็นบวกแน่นอน (p.d.) ทั้งนี้เพราะการแปลงฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) จำเป็นต้องใช้ค่าดีเทอร์มิแนนต์ (determinant) ของ $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ (ดูทฤษฎีบทที่ 3.28 ประกอบ)

หาก$\boldsymbol{\Sigma}$ เป็นบวกแน่นอน (p.d.) แล้ว จะสามารถเขียนสมการที่ 5.28 ใหม่ได้เป็น

$$\boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{5.30}$$

ดังนั้น สามารถประยุกต์ใช้ทฤษฎีบทที่ 3.28 เพื่อหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของ $\boldsymbol{X}$ ได้เป็น

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\det\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\} \tag{5.31}$$

ทฤษฎีบทต่อไปนี้แสดงให้เห็นว่า ฟังก์ชันเชิงเส้นของตัวแปรสุ่มปกติจะมีการแจกแจงปกติ (normal distribution)

ทฤษฎีบทที่ 5.6. *กำหนดให้ $\boldsymbol{X}$ เป็นเวคเตอร์ของตัวแปรสุ่ม n ตัว ซึ่งมีการแจกแจงร่วมแบบปกติ นั่นคือ $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ และ $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ โดยที่ $\boldsymbol{A}$ เป็นเมทริกซ์ของค่าคงที่ขนาด $m \times n$ และ $\boldsymbol{b}$ เป็นเวคเตอร์ของค่าคงที่ m ตัว แล้ว $\boldsymbol{Y} \sim N\left(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'\right)$*

การพิสูจน์. ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{Y}$ เท่ากับ

$$\begin{aligned}\psi_{\boldsymbol{y}}(t) &= E\left[\exp\left\{\boldsymbol{Y}'\boldsymbol{t}\right\}\right] = E\left[\exp\left\{(\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b})'\,\boldsymbol{t}\right\}\right] \\ &= \exp\left\{\boldsymbol{b}'\boldsymbol{t}\right\} E\left[\exp\left\{\boldsymbol{X}'\left(\boldsymbol{A}'\boldsymbol{t}\right)\right\}\right] = \exp\left\{\boldsymbol{b}'\boldsymbol{t}\right\}\exp\left\{\boldsymbol{\mu}'\left(\boldsymbol{A}'\boldsymbol{t}\right)+\frac{1}{2}\left(\boldsymbol{A}'\boldsymbol{t}\right)'\boldsymbol{\Sigma}\left(\boldsymbol{A}'\boldsymbol{t}\right)\right\} \\ &= \exp\left\{(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})'\,\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}'\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'\right)\boldsymbol{t}\right\}\end{aligned}$$

นั่นคือ $\boldsymbol{Y} \sim N\left(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'\right)$

■

นอกจากจะมีประโยชน์ในการแปลงตัวแปรสุ่มในรูปแบบฟังก์ชันเชิงเส้น (linear transformation) แล้ว ทฤษฎีบทด้านบนยังช่วยให้สามารถหาการแจกแจงของตัวแปรสุ่มเพียงบางส่วนในกรณีที่การแจกแจงร่วมเป็นแบบปกติ โดยเริ่มจากการที่สามารถแบ่งเวคเตอร์ $\boldsymbol{X}$ ออกได้เป็นสองส่วน คือ

$$\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_m \\ \boldsymbol{X}_k\end{bmatrix} \tag{5.32}$$

โดยที่ $\boldsymbol{X}_m$ คือเวคเตอร์ของตัวแปรสุ่มที่สนใจ โดยในที่นี้กำหนดให้มีขนาด m ในขณะที่ตัวแปรสุ่ม $\boldsymbol{X}_k$ มีขนาด $k = n - m$ ดังนั้น หากกำหนดให้ $\boldsymbol{b} = \boldsymbol{0}$ และ

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{I}_{mm} & \boldsymbol{0}_{mk} \\ \boldsymbol{0}_{km} & \boldsymbol{0}_{kk}\end{bmatrix} \tag{5.33}$$

จะทำให้ฟังก์ชันที่เกิดจากการแปลงค่าโดยใช้ $\boldsymbol{A}$ และ $\boldsymbol{b}$ เท่ากับ

$$\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{I}_{mm} & \boldsymbol{0}_{mk}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_m \\ \boldsymbol{X}_k\end{bmatrix} = \boldsymbol{X}_m$$

ในทำนองเดียวกัน ค่าคาดหมาย (mean) และเมทริกซ์ความแปรปรวนร่วม (variance-covariance matrix) ของ $\boldsymbol{X}$ สามารถแบ่งส่วนได้เป็น

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_m \\ \boldsymbol{\mu}_k\end{bmatrix} \tag{5.34}$$

และ

$$\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{mm} & \boldsymbol{\Sigma}_{mk} \\ \boldsymbol{\Sigma}_{km} & \boldsymbol{\Sigma}_{kk}\end{bmatrix} \tag{5.35}$$

โดยที่ $\boldsymbol{\Sigma}_{mk}$ คือเมทริกซ์ความแปรปรวนร่วม (variance-covariance matrix) ระหว่าง $\boldsymbol{X}_m$ และ $\boldsymbol{X}_k$

ดังนั้น ค่าคาดหมาย (mean) และเมทริกซ์ความแปรปรวนร่วม (variance-covariance matrix) ของ $\boldsymbol{Y}$ สามารถแบ่งส่วนได้เป็น

$$\begin{bmatrix} \boldsymbol{I}_{mm} & \boldsymbol{0}_{mk} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_m \\ \boldsymbol{\mu}_k \end{bmatrix} = \boldsymbol{\mu}_m \tag{5.36}$$

และ

$$\begin{bmatrix} \boldsymbol{I}_{mm} & \boldsymbol{0}_{mk} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{mm} & \boldsymbol{\Sigma}_{mk} \\ \boldsymbol{\Sigma}_{km} & \boldsymbol{\Sigma}_{kk} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{mm} \\ \boldsymbol{0}_{mk} \end{bmatrix} = \boldsymbol{\Sigma}_{mm} \tag{5.37}$$

ดังนั้น สามารถสรุปได้ว่า $\boldsymbol{Y} = \boldsymbol{X}_m \sim N(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_{mm})$ กล่าวคือ การแจกแจงร่วมเป็นแบบปกติ แล้วการแจกแจงของตัวแปรสุ่มบางส่วนก็จะมีการแจกแจงปกติเช่นกัน โดยที่ค่าคาดหมาย (mean) และค่าความแปรปรวน (variance) ของชุดตัวแปรบางส่วนสามารถหาได้จากค่าสถิติของตัวแปรสุ่มทั้งหมดได้ไม่ยากนัก

ตัวอย่างที่ 5.5. พิจารณาตัวแปรสุ่มสองตัว X_1 และ X_2 ที่มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (μ_1, μ_2) และเมทริกซ์ความแปรปรวนร่วม (variance-covariance matrix)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

จะเห็นได้ว่า $\det \boldsymbol{\Sigma} = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$ โดยที่ ρ แทนค่าสหสัมพันธ์ระหว่าง X_1 และ X_2 ดังนั้นส่วนกลับ (inverse) ของ $\boldsymbol{\Sigma}$ เท่ากับ

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix}$$

ซึ่งช่วยให้สามารถคำนวณได้ว่า

$$\begin{aligned} (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) &= [x_1-\mu_1, x_2-\mu_2] \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \\ &= \frac{1}{1-\rho^2} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \end{aligned}$$

ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (p.d.f.) ของ X_1 และ X_2 คือ

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{Q}{2}} \tag{5.38}$$

โดยที่

$$Q=\frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] \tag{5.39}$$

ยิ่งไปกว่านี้ ถ้า $\rho=0$ แล้ว ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (p.d.f.) ของ X_1 และ X_2 คือ

$$\begin{aligned} f(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} \\ &= \left[\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2}\right]\left[\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}\right]=f_1(x_1)\,f_2(x_2) \end{aligned}$$

ซึ่งหมายความว่า X_1 และ X_2 เป็นอิสระต่อกัน กล่าวคือ ในกรณีของการแจกแจงปกติ (normal distribution) การไม่มีสหสัมพันธ์ (uncorrelated) หมายถึงการเป็นอิสระต่อกัน (independent) ด้วย ดังสรุปในทฤษฎีบทต่อไปนี้ □

ทฤษฎีบทที่ 5.7. *กำหนดให้ $\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ และสามารถแยกส่วนได้เป็นสองส่วนคือ $\boldsymbol{X}=(\boldsymbol{X}_m,\boldsymbol{X}_k)$ ซึ่งสอดคล้องกับสมการที่ 5.32, 5.34, และ 5.35 ดังนั้น $\boldsymbol{X}_m$ และ $\boldsymbol{X}_k$ เป็นอิสระต่อกัน (independent) ก็ต่อเมื่อ (if and only if) $\boldsymbol{\Sigma}_{mk}=\boldsymbol{0}$*

การพิสูจน์. พิจารณาฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{X}$ ในรูปของ $\boldsymbol{t}_m$ และ $\boldsymbol{t}_k$

$$\begin{aligned} \psi(\boldsymbol{t}_m,\boldsymbol{t}_k) &= \exp\left\{\boldsymbol{\mu_m}'\boldsymbol{t}_m+\boldsymbol{\mu_k}'\boldsymbol{t}_k+\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_m+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_k+\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mk}\boldsymbol{t}_k+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{km}\boldsymbol{t}_m\right]\right\} \\ &= \exp\left\{\boldsymbol{\mu_m}'\boldsymbol{t}_m+\boldsymbol{\mu_k}'\boldsymbol{t}_k+\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_m+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_k\right]\right\}\exp\left\{\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mk}\boldsymbol{t}_k+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{km}\boldsymbol{t}_m\right]\right\} \end{aligned} \tag{5.40}$$

ในขณะเดียวกัน หาก $\boldsymbol{X}_m$ และ $\boldsymbol{X}_k$ เป็นอิสระต่อกัน (independent) แล้ว ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ $\boldsymbol{X}$ เท่ากับ

$$\begin{aligned} \psi(\boldsymbol{t}_m,\boldsymbol{t}_k) &= \exp\left\{\boldsymbol{\mu_m}'\boldsymbol{t}_m+\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_m\right]\right\}\exp\left\{\boldsymbol{\mu_k}'\boldsymbol{t}_k+\frac{1}{2}\left[\boldsymbol{t}_k'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_k\right]\right\} \\ &= \exp\left\{\boldsymbol{\mu_m}'\boldsymbol{t}_m+\boldsymbol{\mu_k}'\boldsymbol{t}_k+\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_m+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{mm}\boldsymbol{t}_k\right]\right\} \end{aligned} \tag{5.41}$$

ดังนั้น สมการที่ 5.40 และสมการที่ 5.41 จะเท่ากันก็ต่อเมื่อ

$$\exp\left\{\frac{1}{2}\left[\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mk}\boldsymbol{t}_k+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{km}\boldsymbol{t}_m\right]\right\}=1\Rightarrow\boldsymbol{t}_m'\boldsymbol{\Sigma}_{mk}\boldsymbol{t}_k+\boldsymbol{t}_k'\boldsymbol{\Sigma}_{km}\boldsymbol{t}_m=0$$

ซึ่งจะเป็นจริงก็ต่อเมื่อ $\boldsymbol{\Sigma}_{mk}=\boldsymbol{\Sigma}_{km}'=\boldsymbol{0}$ ■

ทฤษฎีบทที่ 5.8. *กำหนดให้* $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *และสามารถแยกส่วนได้เป็นสองส่วนคือ* $\boldsymbol{X} = (\boldsymbol{X}_m, \boldsymbol{X}_k)$ *ซึ่งสอดคล้องกับสมการที่ 5.32, 5.34, และ 5.35 และสมมุติว่า* $\boldsymbol{\Sigma}$ *เป็นบวกแน่นอน (positive definite) ดังนั้น การแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ* $\boldsymbol{X}_m$ *เมื่อทราบ* $\boldsymbol{X}_k$ *เป็นแบบปกติ (normal) นั่นคือ*

$$\boldsymbol{X}_m \big| \boldsymbol{X}_k \sim N\left(\boldsymbol{\mu}_m + \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\left(\boldsymbol{X}_k - \boldsymbol{\mu}_k\right), \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{km}\right) \tag{5.42}$$

การพิสูจน์. กำหนดให้ $\boldsymbol{W} = \boldsymbol{X}_m - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{X}_k$ และพิจารณาระบบสมการ

$$\begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_m & -\boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$$

และประยุกต์ใช้ทฤษฎีบทที่ เพื่อคำนวณหาเมทริกซ์ของค่าความแปรปรวน (variance-covariance matrix) ได้เป็น

$$\begin{bmatrix} \boldsymbol{I}_m & -\boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{mm} & \boldsymbol{\Sigma}_{mk} \\ \boldsymbol{\Sigma}_{km} & \boldsymbol{\Sigma}_{kk} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1} & \boldsymbol{I}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{km} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{kk} \end{bmatrix}$$

ซึ่งชี้ให้เห็นว่า $\boldsymbol{W}$ และ $\boldsymbol{X}_k$ เป็นอิสระต่อกัน (independent) ดังนั้น การแจกแจงแบบมีเงื่อนไข $\boldsymbol{W}\big|\boldsymbol{X}_k$ จะเหมือนกับการแจกแจงของ $\boldsymbol{W}$ ซึ่งเป็นการแจกแจงปกติ (normal distribution) ด้วยค่าคาดหมาย (mean) เท่ากับ $\boldsymbol{\mu}_m - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\mu}_k$ ดังนั้น จึงสามารถสรุปได้ว่า

$$\boldsymbol{W}\big|\boldsymbol{X}_k \sim N\left(\boldsymbol{\mu}_m - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{km}\right)$$

ในขณะเดียวกัน เนื่องจากสิ่งที่เราสนใจคือการแจกแจงแบบมีเงื่อนไขเมื่อทราบ $\boldsymbol{X}_k$ ดังนั้น การบวกเพิ่มด้วยฟังก์ชันของ $\boldsymbol{X}_k$ เข้าไปย่อมมีผลต่อเช่นเดียวกับการบวกเพิ่มด้วยค่าคงที่ ซึ่งมีผลต่อค่าคาดหมาย แต่ไม่มีผลต่อค่าความแปรปรวน และที่สำคัญ เราสามารถคำนวณหา $\boldsymbol{X}_m$ ได้จาก $\boldsymbol{W} + \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{X}_k = \boldsymbol{X}_m$ ดังนั้น การแจกแจงแบบมีเงื่อนไข (conditional distribution) ของ $\boldsymbol{X}_m$ เมื่อทราบ $\boldsymbol{X}_k$ เท่ากับ

$$\boldsymbol{X}_m\big|\boldsymbol{X}_k \sim N\left(\boldsymbol{\mu}_m - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\mu}_k + \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{X}_k, \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mk}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{km}\right)$$

ซึ่งตรงกับผลลัพธ์ที่ต้องการ ■

ตัวอย่างต่อไปนี้แสดงให้เห็นว่า หากตัวแปรสุ่มสองตัวมีการแจกแจงร่วมแบบปกติ (normally distributed) แล้วค่าคาดหมายแบบมีเงื่อนไข (conditional mean) ของตัวแปรหนึ่งเมื่อทราบอีกอันหนึ่งจะอยู่ในรูปฟังก์ชันเชิงเส้น (linear function) เสมอ ซึ่งเป็นข้อจำกัดของการวิเคราะห์ข้อมูลภายใต้การแจกแจงปกติ ทำให้จำเป็นต้องสมมุติการแจกแจงที่แตกต่างจากการแจกแจงปกติหากต้องการความสัมพันธ์ที่ไม่เป็นเชิงเส้น (nonlinear relationship) ยกตัวอย่างเช่น Attanasio et al. (2015) จำเป็นต้องเลือกใช้การแจกแจงผสม (mixture distribution)

เพื่อให้สามารถประมาณการณ์ฟังก์ชันการผลิตทุนมนุษย์ (human capital production function) ที่เป็นแบบไม่เชิงเส้น (nonlinear)

ตัวอย่างที่ 5.6. เช่นเดียวกับตัวอย่างที่ 5.5 สมมุติให้ตัวแปรสุ่ม X_1 และ X_2 มีการแจกแจงปกติ (normally distributed) ด้วยค่าคาดหมาย (μ_1, μ_2) และเมทริกซ์ความแปรปรวนร่วม (variance-covariance matrix)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

ดังนั้น สามารถใช้สมการที่ 5.42 เพื่อคำนวณหาค่าคาดหมายแบบมีเงื่อนไข (conditional mean) ของ X_1 เมื่อทราบ $X_2 = x_2$ ได้เป็น

$$\begin{aligned} E\left[X_1 | X_2 = x_2\right] &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{X}_2 - \boldsymbol{\mu}_2\right) = \mu_1 + \left(\rho\sigma_1\sigma_2\right)\left(\sigma_2^2\right)^{-1}\left(x_2 - \mu_2\right) \\ &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(x_2 - \mu_2\right) \end{aligned}$$

ซึ่งเป็นฟังก์ชันเชิงเส้น (linear function) ของ x_2 สังเกตว่าเราไม่ได้กำหนดหรือบังคับให้ค่าคาดหมายแบบมีเงื่อนไข (conditional mean) นี้มีความสัมพันธ์เชิงเส้นแต่อย่างใด แต่คุณสมบัติของการแจกแจงปกติ นอกจากนี้ยังสามารถเขียนในรูปของตัวแปรสุ่มได้เป็น

$$E\left[X_1 | X_2\right] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(X_2 - \mu_2\right) \tag{5.43}$$

□

5.3 การแจกแจงไคกำลังสอง (Chi-square distribution)

การแจกแจงไคกำลังสอง (chi-square distribution) เป็นการแจกแจงอีกแบบหนึ่งที่ได้รับความนิยม และเป็นหนึ่งในการแจกแจงแบบแกมมา (gamma distribution) นอกจากนี้ คุณสมบัติหนึ่งที่มีความสำคัญในทางสถิติและจะมีประโยชน์สำหรับการหาการแจกแจงของค่าสถิติที่ใช้ในการทดสอบสมมุติฐานในอนาคตคือ เราสามารถสร้างตัวแปรสุ่มที่มีการแจกแจงไคกำลังสอง (chi-square distribution) ได้ด้วยการนำเอาตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐาน (standard normal) มายกกำลังสอง กล่าวคือ ตัวแปรไคกำลังสอง (chi-square) คือกำลังสองของตัวแปรปกติ (normal)

กำหนดให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงไคกำลังสอง (chi-square distribution) ด้วยระดับความอิสระ (degree of freedom) เท่ากับ r นั่นคือ $X \sim \chi^2(r)$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม

(p.d.f.) ของ X คือ

$$f(x)=\begin{cases}\frac{1}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}x^{\frac{r}{2}-1}e^{-\frac{x}{2}},\ \text{สำหรับ } 0<x<\infty,\\ 0,\ \text{สำหรับกรณีอื่น}\end{cases} \tag{5.44}$$

โดยที่ $\Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx$ แทนฟังก์ชันแกมมา (gamma function) ค่าแกมมาฟังก์ชันอันหนึ่งที่มีความสัมพันธ์กับการแจกแจงปกติมาตราคือค่าที่มาจากกรณีที่ระดับความอิสระ (degree of freedom) $r=1$ ซึ่งมีค่าเท่ากับ $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$ ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (p.d.f.) ของ $\chi^2(1)$ คือ

$$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}},\ \text{สำหรับ } 0<x<\infty,\\ 0,\ \text{สำหรับกรณีอื่น}\end{cases} \tag{5.45}$$

ส่วนฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ X เท่ากับ

$$\psi(t)=(1-2t)^{-\frac{r}{2}},\ \text{สำหรับ } t<\frac{1}{2} \tag{5.46}$$

นอกจากนี้ ค่าคาดหมาย (mean) ของ $\chi^2(r)$ เท่ากับ $\psi^1(0)=r$ และค่าความแปรปรวน (variance) ของ $\chi^2(r)$ เท่ากับ $\psi^2(0)-r^2=r(r+2)-r^2=2r$

ทฤษฎีบทต่อไปนี้แสดงว่าฟังก์ชันกำลังสองของตัวแปรสุ่มปกติมาตรฐาน (standard normal) มีการแจกแจงไคกำลังสอง (chi-square) ด้วยระดับความอิสระ (degree of freedom) $r=1$

ทฤษฎีบทที่ 5.9. *สมมุติให้ $X\sim N(0,1)$ แล้ว ตัวแปรสุ่ม $Y=X^2\sim\chi^2(r)$*

การพิสูจน์. เริ่มจากฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y เท่ากับ

$$G(y)=Pr(Y\leqslant y)=Pr\left(X^2\leqslant y\right)=Pr\left(-\sqrt{y}\leqslant X\leqslant\sqrt{y}\right),\ \text{สำหรับ } y>0$$

เนื่องจาก $X\sim N(0,1)$ ดังนั้น สำหรับ $y>0$

$$G(y)=Pr\left(-\sqrt{y}\leqslant X\leqslant\sqrt{y}\right)=\int -\sqrt{y}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=2\int 0^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz$$

โดยที่สมการสุดท้ายเป็นผลมาจากความสมมาตรของการแจกแจงปกติ หลังจากนั้นคือการเปลี่ยนตัวแปรการอินทิเกรต โดยกำหนดให้ $w=z^2$ ซึ่งส่งผลให้ $dz=\frac{dw}{2\sqrt{w}}$ ดังนั้น

$$G(y)=\int 0^{y}\frac{1}{\sqrt{2\pi w}}e^{-\frac{w}{2}}dw$$

ซึ่งมีค่าเท่ากับฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ $\chi^2(1)$ ตามที่ต้องการ ■

คุณสมบัติอีกอันหนึ่งที่การแจกแจงไคกำลังสอง (chi-square) ที่คล้ายกับคุณสมบัติของการแจกแจงปกติ (normal) คือคุณสมบัติการรักษารูปแบบการกระจาย กล่าวคือ ผลบวกของตัวแปรสุ่มที่เป็นอิสระต่อกัน (independent) และแต่ละตัวมีการแจกแจงไคกำลังสอง (chi-square) จะมีการแจกแจงไคกำลังสอง (chi-square) เช่นกัน

ทฤษฎีบทที่ 5.10. *สมมุติให้* $X_i \sim \chi^2(r_i)$ *สำหรับ* $i = 1, \ldots, n$ *และ* $X_1, \ldots, X_n$ *เป็นอิสระต่อกัน (independent) แล้ว ตัวแปรสุ่ม* $Y = \sum_{i=1}^{n} X_i \sim \chi^2(R)$ *โดยที่* $R = \sum_{i=1}^{n} r_i$

การพิสูจน์. เนื่องจาก $X_1, \ldots, X_n$ เป็นอิสระต่อกัน (independent) ดังนั้น ฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของ Y เท่ากับ

$$\psi_y(t) = \prod_{i=1}^{n} (1-2t)^{-\frac{r_i}{2}} = (1-2t)^{-\frac{\sum_{i=1}^{n} r_i}{2}}$$

ซึ่งมีค่าเท่ากับฟังก์ชันก่อกำเนิดโมเมนต์ (m.g.f.) ของการแจกแจงไคกำลังสอง (chi-square) ที่มีระดับความอิสระ (degree of freedom) เท่ากับ $\sum_{i=1}^{n} r_i$ ■

ทฤษฎีบทนี้สามารถขยายผลไปสู่กรณีที่มีตัวแปรมากกว่าหนึ่งตัวได้ดังต่อไปนี้

ทฤษฎีบทที่ 5.11. *สมมุติให้* $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *โดยที่* $\boldsymbol{\Sigma}$ *เป็นบวกแน่นอน (positive definite) แล้ว ตัวแปรสุ่ม* $\boldsymbol{Y} = (\boldsymbol{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(n)$ *โดยที่* n *คือจำนวนตัวแปรสุ่มทั้งหมดใน* $\boldsymbol{X}$

การพิสูจน์. กำหนดให้ $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{X} - \boldsymbol{\mu})$ และจากที่อภิปรายมาก่อนหน้านี้ ทำให้ทราบดีว่า $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ และกำหนดให้ $\boldsymbol{W} = \boldsymbol{Z}'\boldsymbol{Z} = \sum_{i=1}^{n} Z_i^2$ เนื่องจาก $Z_i \sim N(0,1)$ ส่งผลให้ $Z_i^2 \sim \chi^2(1)$ ดังนั้น จึงสามารถประยุกต์ใช้ทฤษฎบทที่ 5.10 สรุปได้ว่า $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$ นั่นคือ $\boldsymbol{W} \sim \chi^2(n)$ ซึ่งคือผลลัพธ์ที่ต้องการ ■

ตัวอย่างต่อไปนี้แสดงการประยุกต์ใช้ทฤษฎีบทที่ 5.11 ในการคำนวณค่าการแจกแจงของตัวแปรสุ่มที่เป็นผลมาจากการประมาณค่า (estimation) ด้วยแบบจำลองเชิงเส้น (linear model) ซึ่งจะกล่าวถึงอย่างละเอียดในบทที่ XXX แต่ในที่นี้ ขอยกเอาบางส่วนที่ประยุกต์ใช้ความรู้ในส่วนนี้ให้ดูเพื่อให้ตระหนักถึงประโยชน์ในเบื้องต้นก่อน

ตัวอย่างที่ 5.7. พิจารณาแบบจำลองเชิงเส้น

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

โดยที่ $\boldsymbol{\beta}$ คือเวคเตอร์ของพารามิเตอร์ที่ต้องการประมาณค่า (vector of estimated parameters) และสมมุติให้ค่าคลาดเคลื่อน (error terms) $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ และสมมุติว่ามีข้อมูลแบบตัวอย่างสุ่มขนาด n ตัวอย่าง ดังนั้นในการทดสอบสมมุติฐานที่เกี่ยวข้องกับแบบจำลองเชิงเส้นนี้ จำเป็นจะต้องหาการแจกแจงของ $\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$ ซึ่งเมื่อประยุกต์ใช้ทฤษฎีบทที่ 5.11 จะสามารถสรุปได้ว่า $\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \sim \chi^2(n)$ □

นอกจากนี้ การแจกแจงไคกำลังสอง (chi-square) ยังมีความสำคัญเพราะเป็นส่วนสำคัญในการสร้างการแจกแจงที (t distribution) และการแจกแจงเอฟ (F distribution) ดังจะกล่าวในรายละเอียดในหัวข้อถัดไป

5.4 การแจกแจงที (t Distribution)

กำหนดให้ $X \sim N(0,1)$ และ $Y \sim \chi^2(r)$ และเป็นอิสระต่อกัน (independent) ดังนั้น

$$T = \frac{X}{\sqrt{\frac{Y}{r}}} \tag{5.47}$$

มีการแจกแจงที (t distribution) ด้วยระดับความอิสระ (degree of freedom) r ซึ่งมักแทนด้วย $t(r)$ นั่นคือ การแจกแจงที (t distribution) เกิดจากการผสมผสานกันของการแจกแจงปกติ (normal) และการแจกแจงไคกำลังสอง (chi-square)

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (p.d.f.) ของ T ด้วยระดับความอิสระ (degree of freedom) r คือ

$$f(t) = \frac{\Gamma\left(\frac{1+r}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\sqrt{\pi r}}\left(1+\frac{t^2}{r}\right)^{-\frac{1+r}{2}} \tag{5.48}$$

ในทางปฏิบัติค่าสถิติจากการแจกแจงทีมีค่าใกล้เคียงค่าสถิติจากการแจกแจงปกติอย่างมาก (ดูรูปที่ XXX ประกอบ) โดยเฉพาะอย่างยิ่งเมื่อกลุ่มตัวอย่างหรือระดับความอิสระ (degree of freedom) มีค่ามากพอ นั่นคือ $\lim_{r\to\infty} t(r) \sim N(0,1)$

ADD FIGURE DENSITY OF t WITH DIFFERENT DOF

5.5 การแจกแจงเอฟ (F Distribution)

กำหนดให้ $X \sim \chi^2(r_1)$ และ $Y \sim \chi^2(r_2)$ และเป็นอิสระต่อกัน (independent) ดังนั้น

$$F = \frac{\frac{X}{r_1}}{\frac{Y}{r_2}} \tag{5.49}$$

มีการแจกแจงเอฟ (F distribution) ด้วยระดับความอิสระ (degree of freedom) r_1 และ r_2 ซึ่งมักแทนด้วย $F(r_1, r_2)$ นั่นคือ การแจกแจงที (F distribution) เป็นผลลัพธ์จากการหารกันของการแจกแจงไคกำลังสอง (chi-square) สองอัน

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (p.d.f.) ของ T ด้วยระดับความอิสระ (degree of freedom) r คือ

$$f(x)=\begin{cases}\frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}x^{\frac{r_1}{2}-1}\left(1+\frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}}, & \text{สำหรับ } 0<x<\infty,\\ 0, & \text{สำหรับกรณีอื่น}\end{cases} \tag{5.50}$$

นอกจากนี้ ยังสามารถแสดงได้ว่า ลิมิตของการแจกแจงเอฟ คือการแจกแจงไคกำลังสอง นั่นคือ

$$\lim_{r_2\to\infty} F(r_1,r_2)\sim\chi^2(r_1) \tag{5.51}$$

$$\lim_{r_1\to\infty}\lim_{r_2\to\infty} F(r_1,r_2)\sim N(0,1) \tag{5.52}$$

บทที่ 6

ทฤษฎีที่อ้างอิงกับกลุ่มตัวอย่างขนาดใหญ่ (Large-Sample Theories)

บทนี้นำเสนอทฤษฎีบทที่เป็นรากฐานที่สำคัญในทางสถิติสองอัน คือ กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) และทฤษฎีบทลิมิตของค่ากลาง (Central Limit Theorem)

6.1 กฎว่าด้วยจำนวนมาก (Law of Large Numbers)

กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เป็นทฤษฎีบทที่สำคัญทางสถิติที่ทำให้สามารถประมาณค่า (estimate) ค่าคาดหมาย (mean) ของตัวแปรที่สนใจ (ซึ่งเป็นค่าทางทฤษฎี) ด้วยค่าเฉลี่ย (average) ซึ่งหมายถึงการนำเอาข้อมูลตัวแปรที่สนใจนั้นมารวมกันแล้วหารด้วยจำนวนตัวอย่าง หลักการนี้เป็นหลักการพื้นฐานที่ช่วยให้สามารถคำนวณหาค่าสถิติต่างๆ ได้สะดวก

ในทางเทคนิค การพิสูจน์กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) จำเป็นต้องใช้ความสัมพันธ์ทางคณิตศาสตร์ที่เรียกว่า อสมการเชบิเชฟ (Chebychev inequality) ทำให้หนังสือส่วนใหญ่เริ่มอภิปรายจากอสมการเชบิเชฟก่อน แต่เพื่อให้ผู้อ่านได้เห็นและทำความเข้าใจเกี่ยวกับทฤษฎีก่อน จึงขอนำเสนอกฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) โดยใช้เทคนิคทางคณิตศาสตร์ที่จะนำเสนอในภายหลัง

ก่อนอื่นเนื่องจากกฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เกี่ยวข้องกับการลู่เข้า แต่เป็นการลู่เข้าเชิงความน่าจะเป็น (converge in probability) ซึ่งมีนิยามดังต่อไปนี้

บทนิยามที่ 6.1. *อนุกรมของตัวแปรสุ่ม* $X_1, X_2, \ldots$ *ลู่เข้าสู่ตัวแปรสุ่ม* X *ในเชิงความน่าจะเป็น (converge to*

X in probability) ถ้า สำหรับทุกๆ ค่า $\varepsilon > 0$

$$\lim_{n\to\infty} Pr\left(|X_n - X| < \varepsilon\right) = 1 \Leftrightarrow \lim_{n\to\infty} Pr\left(|X_n - X| \geqslant \varepsilon\right) = 0 \tag{6.1}$$

ซึ่งอาจจะเขียนแทนด้วย $X_n \xrightarrow{p} X$ หรือ $plim\ X_n = c$

ความหมายของการลู่เข้าเชิงความน่าจะเป็น (converge in probability) คือ ความน่าจะเป็นที่ X_n จะมีค่าใกล้เคียงกับ c นั้นมีค่าเข้าใกล้หนึ่งเมื่อ n ลู่เข้าสู่อนันต์ หรือในทางตรงกันข้าม อาจจะมองได้ว่า เมื่อ n มีค่าสูงมากๆ แล้วโอกาสที่ค่าของตัวแปรสุ่ม X_n จะแตกต่างจาก c นั้นจะมีค่าเข้าใกล้ศูนย์มากๆ

เหตุผลที่สำคัญที่ทำให้ต้องพิจารณาในรูปแบบของการลู่เข้าเชิงความน่าจะเป็น (converge in probability) ก็เพราะว่าในโลกของความน่าจะเป็นหรือความไม่แน่นอน ไม่สามารถระบุค่าอะไรได้อย่างแน่นอน สิ่งที่พอทำได้คือบอกว่าเหตุการณ์อย่างอื่นมีความน่าจะเป็นลู่เข้าสู่ศูนย์ แต่ก็ไม่ได้หมายความว่าความน่าจะเป็นเท่ากับศูนย์

ทฤษฎีบทต่อไปนี้คือกฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers)

ทฤษฎีบทที่ 6.1. *สมมุติให้ $X_1, \ldots, X_n$ คือกลุ่มตัวอย่างขนาด n ตัวอย่างที่สุ่มมาจากการแจกแจงที่มีค่าคาดหมาย (mean) μ และค่าความแปรปรวน (variance) มีค่าจำกัด และกำหนดให้ $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ คือค่าคาดหมายจากตัวอย่าง (sample mean) แล้ว*

$$\overline{X}_n \xrightarrow{p} \mu \tag{6.2}$$

การพิสูจน์. ดังที่กล่าวมาแล้วในทฤษฎีบทที่ 5.5 $E\left[\overline{X}_n\right] = \mu$ และ $Var\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$ ดังนั้น สามารถประยุกต์ใช้อสมการเชบิเชฟ (Chebyshev Inequality) เพื่อแสดงว่า สำหรับทุกๆ ค่าจำนวนจริง $\varepsilon > 0$

$$Pr\left(\left|\overline{X}_n - \mu\right| \leqslant \varepsilon\right) = 1 - Pr\left(\left|\overline{X}_n - \mu\right| \geqslant \varepsilon\right) \geqslant 1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2}$$

ซึ่งลู่เข้าสู่หนึ่งเมื่อ n ลู่เข้าสู่อนันต์

$$\lim_{n\to\infty} Pr\left(\left|\overline{X}_n - \mu\right| \leqslant \varepsilon\right) \geqslant \lim_{n\to\infty} 1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = 1$$

เนื่องจากความน่าจะเป็นมีค่าไม่เกินหนึ่ง ดังนั้น จึงสามารถสรุปได้ว่า

$$\lim_{n\to\infty} Pr\left(\left|\overline{X}_n - \mu\right| \leqslant \varepsilon\right) = 1$$

สำหรับทุกๆ ค่าจำนวนจริง $\varepsilon > 0$ ■

กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ระบุว่า ค่าคาดหมายจากตัวอย่าง (sample mean) $\overline{X}_n$ จะมีค่าใกล้เคียงกับค่าคาดหมาย (mean) μ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ นั่นหมายความว่า หากมีข้อมูลขนาดใหญ่มากพอแล้ว ค่าเฉลี่ยเลขคณิต (arithmetic average) $\overline{x}_n$ ที่คำนวณจากตัวอย่างสุ่ม (random sample) จะมีค่าใกล้เคียงกับค่าคาดหมาย (mean) μ ดังนั้น จึงนิยมที่จะประมาณค่าคาดหมาย (mean) ของตัวแปรต่างๆ ด้วยค่าเฉลี่ยเลขคณิต (arithmetic average) ของตัวแปรนั้น

การประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ที่น่าสนใจอันหนึ่งคือ การสร้างกราฟแสดงความถี่ (histogram)

บทนิยามที่ 6.2. *กำหนดให้ $x_1, \ldots, x_n$ คือค่าจำนวนจริงที่ $a < x_i < b$ สำหรับทุก i และ a, b คือค่าคงที่ เราสามารถสร้างกราฟแสดงความถี่ (histogram) จากเซ็ตของจำนวนจริงดังกล่าวได้โดยเลือกจำนวนเต็ม $k > 1$ และแบ่งช่วง $[a, b]$ เป็น k ช่วงที่มีความกว้างเท่ากันที่ $\frac{b-a}{k}$ หลังจากนั้น ให้นับดูว่ามีกี่จำนวนจากเซ็ตดังกล่าวที่มีค่าตกอยู่ในช่วงย่อย (subinterval) แต่ละช่วง $j = 1, \ldots, k$ และแทนจำนวนที่นับได้ด้วย c_j ขั้นตอนต่อไปคือการเลือกรูปแบบของกราฟแสดงความถี่ (histogram) ซึ่งทำได้ด้วยการเลือกจำนวนจริง r เพื่อกำหนดระดับความสูงของแท่ง j เท่ากับ $\frac{c_j}{r}$*

โดยทั่วไปมักจะเลือกค่า r จากค่าใดค่าหนึ่งต่อไปนี้ $\left\{1, n, \frac{n(b-a)}{k}\right\}$ ค่า $r = 1$ แทนกราฟที่ความสูงแทนจำนวนที่นับได้ในแต่ละช่วง ส่วน $r = n$ แทนกราฟที่ความสูงเท่ากับสัดส่วนของจำนวนที่ตกอยู่ในแต่ละช่วง ซึ่งกรณีที่ได้รับความนิยมมากที่สุดเพราะมีความสัมพันธ์กับฟังก์ชันความหนาแน่นความน่าจะเป็น (p.d.f.) ส่วนกรณีที่ $r = \frac{n(b-a)}{k}$ แทนกราฟที่พื้นที่ของแต่ละแท่งมีค่าเท่ากับสัดส่วนของจำนวนที่ตกอยู่ในช่วงนั้น ทฤษฎีบทต่อไปนี้ ประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เพื่อบอกว่า ความสูงของแท่งในกรณีที่ $r = n$ นั้นลู่เข้าสู่ค่าความน่าจะเป็นที่จะสุ่มได้จำนวนที่มีค่าอยู่ในช่วงนั้น

ทฤษฎีบทที่ 6.2. *กำหนดให้ $X_1, X_2, \ldots$ เป็นอนุกรมของตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) และค่าคงที่ $c_1 < c_2$ พร้อมทั้งกำหนดให้ $Y_i = 1$ ถ้า $c_1 \leqslant X_i < c_2$ และ $Y_i = 0$ ถ้าไม่ใช่ แล้ว $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ คือสัดส่วนของ $X_1, X_2, \ldots$ ที่มีค่าอยู่ในช่วง $[c_1, c_2)$ และ*

$$\overline{Y}_n \xrightarrow{p} Pr\left(c_1 \leqslant X_i < c_2\right) \tag{6.3}$$

การพิสูจน์. จากนิยามที่กำหนดให้ จะเห็นได้ว่า $Y_1, Y_2, \ldots$ เป็นตัวแปรสุ่มแบบเบอร์นูลี (Bernoulli) ด้วยค่าพารามิเตอร์ $p = Pr\left(c_1 \leqslant X_i < c_2\right)$ ที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) ดังนั้น สามารถประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เพื่อสรุปว่า $\overline{Y}_n \xrightarrow{p} p$ เพราะ p คือค่าคาดหมายของ Y_i ในขณะเดียวกัน $p = Pr\left(c_1 \leqslant X_i < c_2\right)$ ■

ตัวอย่างต่อไปนี้แสดงการสร้างกราฟแสดงความถี่ (histogram) ของอัตราผลตอบแทน (returns) ของตลาดหลักทรัพย์แห่งประเทศไทย ประเด็นสำคัญของตัวอย่างนี้คือ บทบาทของความกว้างแถบความถี่ (bandwidth) $[c_1, c_2)$ ต่อความราบเรียบ (smoothness) ของกราฟแสดงความถี่ (histogram) โดยจะเห็นได้ว่า กราฟแสดงความถี่จะมีความราบเรียบมากขึ้นเมื่อความกว้างแถบความถี่ (bandwidth) มีค่าสูงขึ้น

ตัวอย่างที่ 6.1. พิจารณาข้อมูลรายวันของอัตราผลตอบแทนต่อปี (annualized returns) ของตลาดหลักทรัพย์แห่งประเทศไทยระหว่าง 1 มกราคม 2015 ถึง 30 กันยายน 2018 โดยในกราฟแรกกำหนดให้ความกว้างแถบความถี่ (bandwidth) แต่ละแถบมีความกว้างเท่าๆ กันที่ 0.025 หรือร้อยละ 2.5 และเพื่อความสะดวก กำหนดให้ค่าต่ำสุดเท่ากับ -0.25 และสูงสุดเท่ากับ 0.25 ส่วนในกราฟที่สองกำหนดให้กว้าง 0.01 หรือร้อยละ 1

ADD FIGURES

บทเรียนที่สำคัญจากตัวอย่างนี้คือ กราฟแรกมีความราบเรียบมากกว่าอันที่สอง ซึ่งสามารถอธิบายได้โดยพิจารณาตัวแปรสุ่ม X ซึ่งมีการแจกแจงทวินาม (binomial distribution) ด้วยค่าพารามิเตอร์ n และ p และตัวแปรสุ่ม Y ซึ่งมีการแจกแจงทวินาม (binomial distribution) ด้วยค่าพารามิเตอร์ n และ $\frac{p}{k}$ โดยที่ $k > 1$ เราสามารถสร้างตัวแปรใหม่ $Z = kY$ ที่มีค่าคาดหมาย (mean) เท่ากับ $\mu_Z = np$ ซึ่งมีค่าเท่ากับค่าคาดหมายของ X แต่มีค่าความแปรปรวน (variance) เท่ากับ $\sigma_Z^2 = k^2 n \frac{p}{k}\left(1 - \frac{p}{k}\right) = knp\left(1 - \frac{p}{k}\right)$ ในขณะเดียวกัน ค่าความแปรปรวน (variance) ของ X มีค่าเท่ากับ $\sigma_X^2 = np\left(1 - p\right)$

หากพิจารณาในกรณีที่ p มีค่าน้อยมาก (กรณีที่กำหนดให้ความกว้างแถบความถี่ (bandwidth) $[c_1, c_2)$ มีค่าน้อยมาก) จนทำให้สามารถกำหนดให้ค่า $p^2 \approx 0$ ซึ่งมีผลทำให้สามารถแสดงได้ว่า $\sigma_X^2 \approx np$ และ $\sigma_Z^2 \approx knp$ นั่นคือ ค่าความแปรปรวน (variance) ของ Z มีค่าประมาณ k เท่าของค่าความแปรปรวน (variance) ของ X

ผลการคำนวณในตัวอย่างนี้แสดงให้เห็นว่า ตัวแปร X และ Z มีค่าคาดหมายที่เท่ากัน แต่มีช่วงความกว้างแถบความถี่ (bandwidth) ที่แตกต่างกัน โดยตัวแปร X นำไปสู่กราฟแสดงความถี่ (histogram) ที่มีช่วงความกว้างแถบความถี่ (bandwidth) ที่กว้างกว่าเพราะ $k > 1$ นอกจากนี้ การที่ค่าความแปรปรวน (variance) ของ Z มีค่ามากกว่าค่าความแปรปรวน (variance) ของ X สะท้อนให้เห็นว่า กราฟแสดงความถี่จะมีความราบเรียบมากขึ้นเมื่อความกว้างแถบความถี่ (bandwidth) มีค่าสูงขึ้น □

ทฤษฎีบทต่อไปนี้เป็นการขยายผลการลู่เข้าเชิงความน่าจะเป็น (converge in probability) สำหรับฟังก์ชันเชิงเส้นและฟังก์ชันต่อเนื่อง (continuous) ที่ลู่เข้าเชิงน่าจะเป็น ซึ่งทำให้สามารถประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) กับค่าสถิติที่เป็นฟังก์ชันของค่าคาดหมายจากตัวอย่าง (sample mean)

ทฤษฎีบทที่ 6.3. *สมมุติว่า* $X_n \xrightarrow{p} X$ *และ* $Y_n \xrightarrow{p} Y$ *แล้ว* $X_n + Y_n \xrightarrow{p} X + Y$

การพิสูจน์. กำหนดให้ $\epsilon > 0$ ที่สอดคล้องกับอสมการต่อไปนี้

$$\left| \left(X_n + Y_n\right) - \left(X + Y\right) \right| \geqslant \epsilon$$

ดังนั้น อสมการสามเหลี่ยม (triangular inequality) ช่วยให้สามารถบอกได้ว่า

$$|X_n - X| + |Y_n - Y| \geqslant |\,(X_n + Y_n) - (X + Y)\,| \geqslant \epsilon$$

ซึ่งสามารถประยุกต์ใช้กับค่าความน่าจะเป็นได้ว่า

$$Pr\left(|\,(X_n + Y_n) - (X + Y)\,| \geqslant \epsilon\right) \leqslant Pr\left(|X_n - X| + |Y_n - Y| \geqslant \epsilon\right)$$

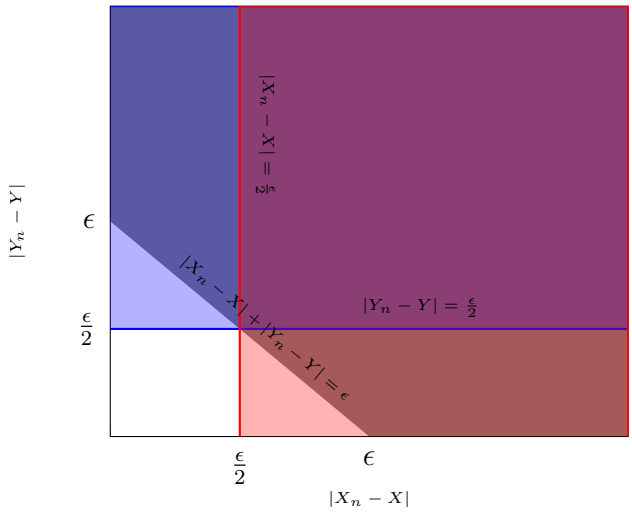
เพื่อให้สามารถเขียนพจน์ด้านขวาในรูปของ $|X_n - X|$ และ $|Y_n - Y|$ จำเป็นต้องพิสูจน์ว่าเหตุการณ์

$$A = \{|X_n - X| + |Y_n - Y| \geqslant \epsilon\}$$

เป็นสับเซ็ตของเหตุการณ์[1]

$$B = \left\{|X_n - X| \geqslant \frac{\epsilon}{2}\right\} \cup \left\{|Y_n - Y| \geqslant \frac{\epsilon}{2}\right\}$$

ซึ่งเห็นได้อย่างชัดเจนจากรูปที่ 6.1



รูปภาพที่ 6.1: พื้นที่ที่อยู่เหนือเส้น $|X_n - X| + |Y_n - Y| = \epsilon$ แทนเหตุการณ์ A ซึ่งเห็นได้อย่างชัดเจนว่าเป็นส่วนหนึ่งของเหตุการณ์ B ส่วนพื้นที่สามเหลี่ยมสองอันด้านล่างเส้น $|X_n - X| + |Y_n - Y| = \epsilon$ แทนส่วนที่อยู่ใน B แต่ไม่อยู่ใน A ดังนั้น จึงสามารถสรุปได้ว่า $A \subset B$

[1]การพิสูจน์ในส่วนนี้ไม่ได้ผูกติดกับค่า $\frac{\epsilon}{2}$ แต่อย่างใด ผู้อ่านสามารถพิสูจน์ทฤษฎีบทเดียวกันนี้โดยเลือกใช้ $\frac{\epsilon}{k}$ สำหรับ X และ $\frac{(k-1)\epsilon}{k}$ สำหรับ Y โดยที่ค่าจำนวนจริง $k > 1$

ผลที่ตามมาก็คือ

$$\begin{aligned} Pr\left(|X_n - X| + |Y_n - Y| \geqslant \epsilon\right) &\leqslant Pr\left(|X_n - X| \geqslant \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geqslant \frac{\epsilon}{2}\right) \\ &\quad + Pr\left(|X_n - X| \geqslant \frac{\epsilon}{2} \text{ และ } |Y_n - Y| \geqslant \frac{\epsilon}{2}\right) \\ &\leqslant Pr\left(|X_n - X| \geqslant \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geqslant \frac{\epsilon}{2}\right) \end{aligned}$$

ดังนั้น

$$Pr\left(|\left(X_n + Y_n\right) - \left(X + Y\right)| \geqslant \epsilon\right) \leqslant Pr\left(|X_n - X| \geqslant \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geqslant \frac{\epsilon}{2}\right)$$

ในขณะเดียวกัน คุณสมบัติการลู่เข้าเชิงความน่าจะเป็นของ X_n และ Y_n ทำให้สรุปได้ว่า

$$\lim_{n\to\infty} Pr\left(|\left(X_n + Y_n\right) - \left(X + Y\right)| \geqslant \epsilon\right) \leqslant \lim_{n\to\infty} Pr\left(|X_n - X| \geqslant \frac{\epsilon}{2}\right) + \lim_{n\to\infty} Pr\left(|Y_n - Y| \geqslant \frac{\epsilon}{2}\right) = 0$$

ซึ่งหมายความว่า $X_n + Y_n \xrightarrow{p} X + Y$ ■

ทฤษฎีบทที่ 6.4. *สมมุติว่า $X_n \xrightarrow{p} X$ และ a เป็นค่าคงที่ แล้ว $aX_n \xrightarrow{p} aX$*

การพิสูจน์. ถ้า $a = 0$ ค่าลิมิตต้องเท่ากับศูนย์อย่างแน่นอน ดังนั้น จำเป็นต้องพิสูจน์เพิ่มเติมเฉพาะในกรณีที่ $a \neq 0$ โดยเริ่มจากการกำหนดให้ $\epsilon > 0$ ดังนั้น

$$\lim_{n\to\infty} Pr\left(|aX_n - aX| \geqslant \epsilon\right) = \lim_{n\to\infty} Pr\left(|a|X_n - X| \geqslant \epsilon\right) = \lim_{n\to\infty} Pr\left(|X_n - X| \geqslant \frac{\epsilon}{|a|}\right) = 0$$

■

ทฤษฎีบทที่ 6.5. *ถ้า $X_n \xrightarrow{p} c$ และ $g\left(x\right)$ เป็นฟังก์ชันที่ต่อเนื่องที่จุด $x = c$ แล้ว $g\left(X_n\right) \xrightarrow{p} g\left(c\right)$*

การพิสูจน์. กำหนดให้ $\epsilon > 0$ เนื่องจาก $g\left(x\right)$ เป็นฟังก์ชันที่ต่อเนื่องที่จุด c ดังนั้น จะต้องมีค่าคงที่ $\delta > 0$ ที่ทำให้ $|g\left(x\right) - g\left(a\right)| < \epsilon$ สำหรับทุกๆ ค่า x ที่สอดคล้องกับเงื่อนไข $|x - c| < \delta$ ดังนั้น

$$|g\left(x\right) - g\left(c\right)| \geqslant \epsilon \Rightarrow |x - c| \geqslant \delta$$

ซึ่งสามารถเขียนในรูปของเหตุการณ์ได้ว่า

$$\{|g\left(x\right) - g\left(c\right)| \geqslant \epsilon\} \subset \{|x - c| \geqslant \delta\}$$

ดังนั้น

$$Pr\left(|g\left(x\right) - g\left(c\right)| \geqslant \epsilon\right) \leqslant Pr\left(|x - c| \geqslant \delta\right)$$

โดยนิยามของการลู่เข้าเชิงความน่าจะเป็น $X_n \xrightarrow{p} c$ ค่าลิมิตของพจน์ด้านขวามือมีค่าเท่ากับ

$$\lim_{n\to\infty} Pr\left(|x-c| \geqslant \delta\right) = 0$$

ซึ่งทำให้สรุปได้ว่า

$$\lim_{n\to\infty} Pr\left(|g\left(x\right) - g\left(c\right)| \geqslant \epsilon\right) = 0 \Rightarrow g\left(X_n\right) \xrightarrow{p} g\left(c\right)$$

■

ทฤษฎีบทนี้สามารถขยายผลสู่กรณีของเวคเตอร์ของตัวแปรสุ่ม $\boldsymbol{Z}_n$ นั่นคือ ถ้า $\boldsymbol{Z}_n \xrightarrow{p} \boldsymbol{z}$ และ $g\left(\boldsymbol{x}\right)$ เป็นฟังก์ชันที่ต่อเนื่องที่จุด $\boldsymbol{x} = \boldsymbol{z}$ แล้ว $g\left(\boldsymbol{Z}_n\right) \xrightarrow{p} g\left(\boldsymbol{z}\right)$ ดังแสดงตัวอย่างในรูปของฟังก์ชันผลคูณในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 6.6. *สมมุติว่า* $X_n \xrightarrow{p} X$ *และ* $Y_n \xrightarrow{p} Y$ *แล้ว* $X_nY_n \xrightarrow{p} XY$

การพิสูจน์. พิจาณาผลคูณ

$$X_nY_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}\left(X_n + Y_n\right)^2 \xrightarrow{p} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}\left(X - Y\right)^2 = XY$$

■

ตัวอย่างต่อไปนี้แสดงการประยุกต์กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) และคุณสมบัติการลู่เข้าเชิงความน่าจะเป็นเพื่อหาค่าลิมิตของค่าความแปรปรวนจากตัวอย่าง (sample variance)

ตัวอย่างที่ 6.2. สมมุติว่า $X_1, \ldots, X_n$ คือตัวอย่างสุ่มที่สุ่มเลือกมาจากการแจกแจงที่ค่าคาดหมายเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 พิจารณาค่าความแปรปรวนจากตัวอย่าง (sample variance)

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n}X_i^2}{n} - \bar{X}_n^2\right]$$

กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ช่วยให้สรุปได้ว่า

$$\bar{X}_n \xrightarrow{p} \mu$$
$$\frac{\sum_{i=1}^{n}X_i^2}{n} \xrightarrow{p} E\left[X_i^2\right] = \sigma^2 + \mu^2$$

ในขณะเดียวกัน ทฤษฎีบทก่อนหน้านี้ช่วยให้สามารถสรุปได้ว่า

$$\bar{X}_n^2 \xrightarrow{p} \mu^2$$

ดังนั้น

$$S_n^2 = \frac{n}{n-1}\left[\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}_n^2\right] \xrightarrow{p} \left(\lim_{n\to\infty}\frac{n}{n-1}\right)\left[\sigma^2 + \mu^2 - \mu^2\right] = \sigma^2$$

□

อันที่จริงตัวอย่างนี้แสดงให้เห็นว่า ค่าความแปรปรวนจากตัวอย่าง (sample variance) นั้นลู่เข้าเชิงความน่าจะเป็นสู่ค่าความแปรปรวนที่แท้จริง ซึ่งเป็นคุณสมบัติที่ต้องการจากตัวประมาณ (estimator) ใดๆ คุณสมบัตินี้เรียกว่า ความคงเส้นคงวา (consistency) ซึ่งจะอภิปรายในรายละเอียดในบทที่ XXX

6.2 ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem)

ทฤษฎีบทลิมิตของค่ากลาง (Central Limit Theorem หรือที่เรียกสั้นๆ ว่า CLT) เป็นทฤษฎีทางสถิติที่มีความสำคัญมากอันหนึ่ง ถึงแม้ว่าทฤษฎีบทนี้จะมีข้อจำกัดในกรณีที่กลุ่มตัวอย่างมีขนาดเล็ก (finite sample) เพราะทฤษฎีบทนี้ตั้งอยู่บนพื้นฐานของการหาค่าลิมิตเมื่อกลุ่มตัวอย่างมีขนาใหญ่มากพอ (large sample) และในปัจจุบัน ความสามารถของคอมพิวเตอร์ที่พัฒนาอย่างก้าวกระโดดได้ช่วยให้เราสามารถคำนวณหาการแจกแจงของค่าสถิติในกรณีที่ตัวอย่างมีขนาดจำกัด (finite sample) ได้สะดวกมากยิ่งขึ้น (ยกตัวอย่างเช่น วิธี bootstrapping) แต่อย่างไรก็ตาม ทฤษฎีบทลิมิตของค่ากลาง (CLT) ก็ยังไปสิ่งที่นักสถิติหรือนักวิเคราะห์ข้อมูลต้องทำความเข้าใจ เพราะเป็นรากฐานที่สำคัญที่นำไปสู่วิธีการทางสถิติและสูตรที่ใช้ในการคำนวณการแจกแจงของค่าสถิติในโปรแกรมทางสถิติที่ใช้อยู่ในปัจจุบัน

เหตุผลที่ทฤษฎีบทนี้ถูกเรียกว่า ทฤษฎีบทลิมิตของค่ากลาง (CLT) ก็เพราะมันเกี่ยวข้องกับค่ากลาง ซึ่งในที่นี้หมายถึงค่าคาดหมายจากตัวอย่าง (sample mean) $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ รูปแบบอย่างง่ายของทฤษฎีบทนี้กล่าวว่า สำหรับตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง ที่ถูกสุ่มมาจากการแจกแจงที่มีค่าคาดหมาย μ และค่าความแปรปรวน σ^2 เราสามารถประมาณค่าการแจกแจงของค่าคาดหมายจากตัวอย่าง (sample mean) ได้โดยใช้การแจกแจงปกติ (normal distribution) ซึ่งมีค่าคาดหมายเท่ากับ μ และค่าความแปรปรวน $\frac{\sigma^2}{n}$ ราวกับว่าเป็นการสุ่มตัวอย่าง n ตัวอย่างมาจากการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมาย μ และค่าความแปรปรวน σ^2 ทั้งๆ ที่เป็นการสุ่มจากการแจกแจงใดๆ ขอเพียงแค่ว่า เป็นการสุ่มอย่างอิสระจากการแจกแจงเดียวกัน

ทฤษฎีบทลิมิตของค่ากลาง (CLT) ต่อไปนี้เป็นรูปแบบที่ J. W. Lindeberg และ P. Levy ได้พิสูจน์สำเร็จในช่วงต้นทศวรรษที่ 1920 โดยเป็นการพิสูจน์ที่ต่างคนต่างทำแต่นำไปสู่ผลลัพธ์เดียวกัน ในทางประวัติศาสตร์ จริงๆ แล้ว A. de Moivre ได้พิสูจน์รูปแบบที่เจาะจงกว่า (เฉพาะกรณีที่การแจกแจงเริ่มต้นเป็นแบบเบอร์นูลลี) ไว้แล้วในช่วงศตวรรษที่ 18

ทฤษฎีบทที่ 6.7 (Lindeberg and Levy). *ถ้าตัวแปรสุ่ม $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง ซึ่งสุ่มมาจากการแจกแจงเดียวกัน ที่มีค่าคาดหมาย μ และค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว สำหรับค่าคงที่ x ใดๆ*

$$\lim_{n\to\infty} Pr\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leqslant x\right) = \Phi(x) \tag{6.4}$$

โดยที่ Φ แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของการแจกแจงปกติมาตรฐาน (standard normal distribution)

ยิ่งไปกว่านั้น ในทางทฤษฎี เราสามารถเขียนทฤษฎีบทลิมิตของค่ากลาง (CLT) โดยใช้หลักการลู่เข้าเชิงการแจกแจง (convergence in distribution) ซึ่งมีนิยามดังต่อไปนี้

บทนิยามที่ 6.3 (Convergence in Distribution หรือ Asymptotic Distribution). *กำหนดให้ $Y_1, Y_2, \ldots$ เป็นอนุกรมของตัวแปรสุ่ม และ F_n แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ Y_n สำหรับ $n = 1, 2, \ldots$ และกำหนดให้ F^* เป็นฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) อันหนึ่ง ดังนั้น อนุกรม $Y_1, Y_2, \ldots$ ลู่เข้าเชิงการแจกแจง (converge in distribution) สู่ F^* ถ้า*

$$\lim_{n\to\infty} F_n(y) = F^*(y) \tag{6.5}$$

สำหรับทุกๆ ค่า y ที่ $F^*(y)$ ต่อเนื่อง (continuous) และเพื่อความกระชับ เราอาจจะกล่าวสั้นๆ แทนได้ว่า Y_n ลู่เข้าเชิงความน่าจะเป็นสู่ F^* ซึ่งอาจจะเขียนแทนได้ด้วย

$$Y_n \xrightarrow{d} F^* \tag{6.6}$$

และเรียก F^* ว่าการแจกแจงที่ลิมิต (asymptotic distribution) ของ Y_n

ทฤษฎีบทลิมิตของค่ากลาง (CLT) สำหรับตัวอย่างสุ่มสามารถเขียนใหม่ในรูปของการลู่เข้าเชิงการแจกแจงได้ดังต่อไปนี้

ทฤษฎีบทที่ 6.8 (Lindeberg and Levy). *ถ้าตัวแปรสุ่ม $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง ซึ่งสุ่มมาจากการแจกแจงเดียวกัน ที่มีค่าคาดหมาย μ และค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว สำหรับค่าคงที่ x ใดๆ*

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0, 1) \tag{6.7}$$

เพื่อให้เข้าใจทฤษฎีบทลิมิตของค่ากลาง (CLT) ได้ดี ขอนำเสนอผลการจำลอง (simulation) จากการแจกแจงรูปแบบต่างๆ ดังตัวอย่างต่อไปนี้

ADD SIMULATION RESULTS FROM BENOULLI, POISSON, EXPONENTIAL, LOG NORMAL

ทฤษฎีบทต่อไปนี้ช่วยให้สามารถประยุกต์ทฤษฎีบทลิมิตของค่ากลาง (CLT) ได้กว้างขวางมากยิ่งขึ้น

ทฤษฎีบทที่ 6.9. *ถ้า* $X_n \xrightarrow{p} X$ *แล้ว* $X_n \xrightarrow{d} X$

ทฤษฎีบทที่ 6.10. *ถ้า* $X_n \xrightarrow{d} c$ *โดยที่* c *เป็นค่าคงที่ แล้ว* $X_n \xrightarrow{p} c$

ทฤษฎีบทที่ 6.11. *ถ้า* $X_n \xrightarrow{d} X$ *และ* $Y_n \xrightarrow{p} 0$ *แล้ว* $X_n + Y_n \xrightarrow{d} X$

ทฤษฎีบทที่ 6.12 (Slutsky's Theorem)**.** *กำหนดให้* $X, X_n A_n, B_n$ *เป็นตัวแปรสุ่ม (random variable) ส่วน* a *และ* b *เป็นค่าคงที่ ดังนั้น ถ้า* $X_n \xrightarrow{d} X$*,* $A_n \xrightarrow{p} a$ *และ* $B_n \xrightarrow{p} b$ *แล้ว*

$$A_n + B_n X_n \xrightarrow{d} a + bX \tag{6.8}$$

ทฤษฎีบทที่ 6.13. *สมมุติว่า* $X_n \xrightarrow{d} X$ *และ* g *เป็นฟังก์ชันที่ต่อเนื่องตลอดช่วงค้ำจุน (support) ของ* X *แล้ว* $g(X_n) \xrightarrow{d} g(X)$

ในทางคณิตศาสตร์ ทฤษฎีบทนี้กล่าวว่า การลู่เข้าเชิงการแจกแจง (convergence in distribution) สามารถส่งผ่านฟังก์ชันต่อเนื่องได้ โดยมีทฤษฎีบทของสลูสกี้ (Slutsky's Theorem) เป็นกรณีพิเศษสำหรับฟังก์ชันเชิงเส้นซึ่งต่อเนื่องตลอดช่วงค้ำจุน

ถึงแม้ว่าทฤษฎีบทนี้จะช่วยให้สามารถหาการแจกแจงที่ลิมิตของฟังก์ชันต่อเนื่องของตัวแปรสุ่มที่ลู่เข้าเชิงการแจกแจงได้ แต่ในทางปฏิบัติก็ยังมีความยุ่งยากพอสมควร เพราะต้องแปลงการแจกแจงของ X ให้เป็นการแจกแจงของ $g(X)$ ซึ่งก็ขึ้นอยู่กับรูปแบบของฟังก์ชัน g ดังนั้น ในทางปฏิบัติ จึงนิยมใช้วิธีการเดลต้า (Delta method) ซึ่งสามารถประยุกต์ใช้กับการแจกแจงปกติได้อย่างสะดวก ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 6.14. *สมมุติว่า* g *เป็นฟังก์ชันที่หาค่าอนุพันธ์อันดับหนึ่ง (first-order derivative) ได้ที่จุด* θ *และค่าอนุพันธ์อันดับหนึ่ง* $g'(\theta) \neq 0$ *และ*

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N\left(0, \sigma^2\right) \tag{6.9}$$

แล้ว

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N\left(0, [g'(\theta)]^2 \sigma^2\right) \tag{6.10}$$

สังเกตว่าวิธีการเดลต้า (Delta method) นี้ใช้ได้กับการแจกแจงปกติ (normal distribution) และฟังก์ชันที่ค่าอนุพันธ์อันดับหนึ่ง (first-order derivative) ไม่เท่ากับศูนย์ ณ จุดที่สนใจเท่านั้น

6.3 อสมการที่จำเป็นสำหรับการพิสูจน์กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers)

ทฤษฎีบทที่ 6.15 (อสมการมาร์คอฟ (Markov Inequality)). *สมมุติให้ X เป็นตัวแปรสุ่มที่ $Pr(X \geqslant 0) = 1$ แล้ว*

$$Pr(X \geqslant t) \leqslant \frac{E[X]}{t} \tag{6.11}$$

สำหรับทุกๆ ค่าจำนวนจริง $t > 0$

การพิสูจน์. ค่าคาดหมายของ X เท่ากับ

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{t} xf(x)\,dx + \int_{t}^{\infty} xf(x)\,dx \\ &\geqslant \int_{t}^{\infty} xf(x)\,dx \geqslant t\int_{t}^{\infty} f(x)\,dx = tPr(X \geqslant t) \end{aligned}$$

ซึ่งช่วยให้สรุปได้ว่า

$$Pr(X \geqslant t) \leqslant \frac{E[X]}{t}$$

■

ทฤษฎีบทที่ 6.16 (อสมการเชบิเชฟ (Chebyshev Inequality)). *สมมุติให้ X เป็นตัวแปรสุ่มที่ค่าความแปรปรวนมีค่าจำกัด แล้ว*

$$Pr\left(\left|X - E[X]\right| \geqslant t\right) \leqslant \frac{Var[X]}{t^2} \tag{6.12}$$

สำหรับทุกๆ ค่าจำนวนจริง $t > 0$

การพิสูจน์. ก่อนอื่นเราสามารถใช้อสมการมาร์คอฟ (Markov Inequality) เพื่อบอกได้ว่า

$$Pr\left(X^2 \geqslant t^2\right) \leqslant \frac{E[X^2]}{t^2}$$

เพราะ X^2 เป็นตัวแปรสุ่มที่ $Pr(X^2 \geqslant 0) = 1$ ในขณะเดียวกัน เราก็ทราบว่า

$$Pr\left((X - E[X])^2 \geqslant t^2\right) = Pr\left(\left|X - E[X]\right| \geqslant t\right)$$

ดังนั้น จึงสามารถสรุปได้ว่า

$$Pr\left(\left|X - E[X]\right| \geqslant t\right) = Pr\left((X - E[X])^2 \geqslant t^2\right) \leqslant \frac{E\left[(X - E[X])^2\right]}{t^2} = \frac{Var[X]}{t^2}$$

■

บทที่ 7

หลักการประมาณค่าแบบจุด (Principle of Estimation)

บทนี้นำเสนอหลักการพื้นฐานของการประมาณค่า (estimation) โดยพิจารณาวิธีการประมาณค่าทั้งหมด 3 วิธี คือ วิธีการประมาณค่าแบบเบส์ (Bayes Estimation) และวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation หรือเรียกสั้นๆ ว่า MLE) และวิธีการประมาณค่าด้วยโมเมนต์ (Method of Moments หรือเรียกสั้นๆ ว่า MM)

7.1 นิยามพื้นฐานเกี่ยวกับการประมาณค่า (Basic Definitions regarding Estimation)

บทนิยามที่ 7.1. *พารามิเตอร์ (parameter)* $\boldsymbol{\theta} \in \Theta$ *หมายถึงลักษณะเฉพาะ (characteristics) ที่กำหนดการแจกแจงร่วม (joint distribution) ของตัวแปรสุ่มที่สนใจ โดยเรียกเซ็ต* Θ *ว่าปริภูมิพารามิเตอร์ (parameter space)*

บทนิยามที่ 7.2. *แบบจำลองทางสถิติ (statistical model) ประกอบไปด้วย*

1. *การระบุอย่างชัดเจนว่าตัวแปรสุ่มที่สนใจมีอะไรบ้าง ทั้งที่สามารถสังเกตได้ (observable) และไม่สามารถสังเกตได้ (unobservable) แต่กำหนดให้มีอยู่ในทางทฤษฎี เช่น*
2. *ข้อกำหนดหรือสมการที่บอกถึงความสัมพันธ์ระหว่างตัวแปรสุ่มที่สนใจ โดยอาจอยู่ในรูปของการแจกแจงร่วม (joint distribution) สำหรับตัวแปรสุ่มที่สามารถสังเกตได้ (observable random variables) เช่น*

3. *การระบุค่าของพารามิเตอร์ (parameter identification) $\boldsymbol{\theta}$ ที่ทำหน้าที่กำหนดการแจกแจงร่วม (joint distribution) ของตัวแปรสุ่มที่สนใจ โดยในส่วนนี้จะมองว่าพารามิเตอร์ที่ไม่ทราบค่า (unknown parameters) เป็นค่าคงที่ นั่นคือ สิ่งที่ต้องการทราบค่าเป็นค่าจำนวนจริงหรือจุด ทำให้เรียกการประมาณค่าแบบนี้ว่า การประมาณค่าแบบจุด (point estimation)*

4. *ในบางกรณีที่สมมุติให้พารามิเตอร์ $\boldsymbol{\theta}$ เป็นตัวแปรสุ่ม สิ่งที่จำเป็นสำหรับแบบจำลองทางสถิติอีกอย่างหนึ่งคือการกำหนดรูปแบบการแจกแจงร่วม (joint distribution) ของพารามิเตอร์ที่ไม่ทราบค่า (unknown parameters) ส่วนการแจกแจงร่วมของตัวแปรสุ่มที่สามารถสังเกตได้ (observable random variables) ซึ่งขึ้นอยู่กับพารามิเตอร์ $\boldsymbol{\theta}$ จะต้องพิจารณาเป็นการแจกแจงร่วมแบบมีเงื่อนไข (joint conditional probability) สำหรับค่าที่เกิดขึ้นจริง (realized value) ของ $\boldsymbol{\theta}$ นั่นคือ $f(\boldsymbol{x}|\boldsymbol{\theta})$*

โดยทั่วไป เราจำเป็นต้องประมาณค่าพารามิเตอร์เพราะ เราต้องการทดสอบสมมุติฐาน (hypothesis testing) ที่ต้องการทราบ ซึ่งมักจะอยู่ในรูปของข้อความเชิงความน่าจะเป็น (probabilistic statement) ที่เรียกอย่างเป็นทางการว่า การอนุมานทางสถิติ (statistical inference) ยกตัวอย่างเช่น เราอาจจะสนใจว่า สามารถบอกด้วยความมั่นใจแค่ไหนว่าค่าพารามิเตอร์ θ มีค่ามากกว่าศูนย์ นั่นคือ ต้องการทดสอบสมมุติฐานที่ว่า $\theta > 0$ เป็นต้น

บทนิยามที่ 7.3. *การอนุมานทางสถิติ (statistical inference) คือกระบวนการ (procedure) ที่สร้างข้อความเชิงความน่าจะเป็น (probabilistic statement) ที่เกี่ยวข้องกับแบบจำลองทางสถิติ (statistical model)*

การอนุมานทางสถิติ (statistical inference) แบ่งได้เป็น 4 รูปแบบดังนี้

1. การคาดการณ์ (prediction) เป็นคาดเดาค่าตัวแปรสุ่มหรือฟังก์ชันของตัวแปรสุ่มที่ยังไม่ทราบค่า ยกตัวอย่างเช่น การคาดการณ์ว่าค่าเฉลี่ยของผลตอบแทนของกองทุนรวมในปีหน้ามีค่าเท่ากับเท่าใด? การคาดการณ์ว่าผลิตภัณฑ์มวลรวม (GDP) ของประเทศในปีหน้าจะขยายตัวร้อยละเท่าใด? การคาดการณ์ว่าในปีหน้าจะมีเด็กที่มีพัฒนาช้ากว่าวัยร้อยละเท่าใด? เป็นต้น เรามักเรียกการคาดการณ์ในกรณีที่สิ่งที่ต้องคาดการณ์หรือพยากรณ์ว่า การประมาณค่า (estimation) ซึ่งคือประเด็นหลักของบทนี้

2. ปัญหาการตัดสินใจทางสถิติ (statistical decision problems) หมายถึงกระบวนการตัดสินใจที่ผลที่ตามมาขึ้นอยู่กับค่าของพารามิเตอร์ที่ไม่ทราบค่า (แต่ต้องใช้ข้อมูลและเครื่องมือทางสถิติประมาณค่า) ดังนั้นเพื่อให้สามารถตัดสินใจได้อย่างมีหลักการ จึงจำเป็นจะต้องพยายามประมาณค่าพารามิเตอร์ดังกล่าว โดยธรรมชาติแล้ว เราไม่มีทางที่จะแน่ใจได้ว่าพารามิเตอร์ดังกล่าวมีค่าเท่าใดกันแน่ แต่เป็นไปได้ที่จะทราบการแจกแจงของพารามิเตอร์ ดังนั้น การตัดสินใจในกรณีจึงต้องทำภายใต้เงื่อนไขเชิงความน่าจะเป็น ซึ่งขึ้นอยู่กับการแจกแจงของตัวประมาณค่า (distribution of estimators) ยกตัวอย่างเช่น เราจะตัดสินใจซื้อกองทุนรวม A ถ้าคาดว่าอัตราผลตอบแทนไม่น้อยกว่าร้อยละ 3 ดังนั้น ปัญหาทางสถิติในกรณีคือ ความน่า

จะเป็นที่อัตราผลตอบแทนของกองทุน A จะมีค่าไม่น้อยกว่าร้อยละ 3 มีค่าเท่าใด? เป็นต้น วิธีการอนุมานทางสถิติ (statistical inference) แบบนี้เกี่ยวข้องกับสิ่งที่เรียกว่า การทดสอบสมมุติฐาน (hypothesis testing) ซึ่งจะกล่าวถึงอย่างละเอียดในบทที่ XXX

3. การออกแบบการทดลอง (experimental design) เป็นการกำหนดว่าควรจะเก็บข้อมูลอย่างไรและมากน้อยเพียงใด รวมทั้งรูปแบบการทดลอง ยกตัวอย่างเช่น การทดลองแบบสุ่ม (randomized controlled trials) ที่ดีควรจะมีการออกแบบจำนวนกลุ่มตัวอย่างที่เหมาะสม เพื่อให้มั่นใจได้ว่าการทดสอบสมมุติฐานที่จะดำเนินการมีพลังทางสถิติ (statistical power) มากเพียงพอ เป็นต้น

4. การอนุมานทางสถิติแบบอื่นๆ ที่ไม่สามารถจัดอยู่ในสามประเภทแรกได้

บทนิยามที่ 7.4. *กำหนดให้ $X_1, \ldots, X_n$ เป็นข้อมูลที่สังเกตค่าได้ (observed data) ที่การแจกแจงร่วมกำหนดได้ด้วยพารามิเตอร์ $\theta \in \Theta$ ตัวประมาณค่า (estimator) ของพารามิเตอร์ θ หมายถึงฟังก์ชันจำนวนจริง $\hat{\theta}(X_1, \ldots, X_n)$ และกำหนดให้ x_i คือค่าที่เกิดขึ้นจริง (realized value) ของ X_i แล้วเราจะเรียก $\hat{\theta}(x_1, \ldots, x_n)$ ว่าค่าประมาณ (estimate) ของพารามิเตอร์ θ*

ในขณะเดียวกัน เรามักเรียกฟังก์ชันของตัวอย่าง $\hat{\theta}(X_1, \ldots, X_n)$ ว่าค่าสถิติ (statistic) ดังนั้น จึงอาจจะสรุปได้ว่า ตัวประมาณค่า (estimator) ก็คือค่าสถิติ (statistic) อย่างหนึ่งนั่นเอง

7.2 ตัวประมาณค่าแบบเบส์ (Bayes Estimator)

บทนิยามที่ 7.5. *กำหนดให้ $\boldsymbol{\theta}$ คือพารามิเตอร์ที่ไม่ทราบค่าซึ่งสมมุติให้เป็นตัวแปรสุ่ม การแจกแจงก่อนการสังเกต (prior distribution) คือการแจกแจง (distribution) ของพารามิเตอร์ θ ที่กำหนดขึ้นก่อนที่จะสังเกตค่าของตัวแปรสุ่มอื่นๆ ซึ่งมักแทนด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนการสังเกต (prior p.d.f.) หรือฟังก์ชันความน่าจะเป็นก่อนการสังเกต (prior p.f.) $h(\boldsymbol{\theta})$*

ตัวอย่างที่ 7.1. กำหนดให้ θ แทนความน่าจะเป็นที่จะได้ผลการโยนเหรียญดังกล่าวเป็นหัว ซึ่งในที่นี้เป็นพารามิเตอร์ของแบบจำลองทางสถิติที่สนใจ สมมุติว่าเหรียญหนึ่งอาจจะเป็นเหรียญที่ด้านหนึ่งเป็นหัวส่วนอีกด้านหนึ่งเป็นก้อย หรืออาจจะเป็นเหรียญที่มีแต่หัวทั้งสองด้าน ดังนั้น θ เป็นไปได้สองค่าคือ $\theta = \frac{1}{2}$ และ $\theta = 1$ การแจกแจงก่อน (prior distribution) สำหรับกรณีคือการแจกแจงของ θ เช่น $h\left(\theta = \frac{1}{2}\right) = 0.6$ และ $h(\theta = 1) = 0.4$ สังเกตว่า h ต้องรวมแล้วเท่ากับหนึ่งเพราะเป็นฟังก์ชันความน่าจะเป็นก่อนการสังเกต (prior p.f.) □

บทนิยามที่ 7.6. *กำหนดให้ $\boldsymbol{\theta}$ คือพารามิเตอร์ที่ไม่ทราบค่าซึ่งสมมุติให้เป็นตัวแปรสุ่มและ $X_1, \ldots, X_n$ คือตัวแปรสุ่มที่สังเกตค่าได้ การแจกแจงหลังการสังเกต (posterior distribution) คือ การแจกแจงแบบมีเงื่อนไขของ $\boldsymbol{\theta}$*

เมื่อทราบค่า $X_1 = x_1, \ldots, X_n = x_n$ *ซึ่งมักแทนด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นหลังการสังเกต (posterior p.d.f.) หรือฟังก์ชันความน่าจะเป็นหลังการสังเกต (posterior p.f.)* $h(\boldsymbol{\theta}|\boldsymbol{x})$

สังเกตว่าความแตกต่างเชิงสัญลักษณ์ระหว่างการแจกแจงก่อนการสังเกต (prior distribution) $h(\boldsymbol{\theta})$ และการแจกแจงหลังการสังเกต (posterior distribution) $h(\boldsymbol{\theta}|\boldsymbol{x})$ คือการที่อันแรกไม่มีเงื่อนไขแต่อันหลังเป็นการแจกแจงแบบมีเงื่อนไข

ทฤษฎีบทต่อไปนี้ประยุกต์ใช้ทฤษฎีบทของเบส์ (3.20) เพื่อแปลงการแจกแจงก่อนการสังเกต (prior distribution) ให้เป็นการแจกแจงหลังการสังเกต (posterior distribution) โดยอาศัยสารสนเทศ (information) ที่ได้จากข้อมูลที่สังเกตได้ $\boldsymbol{x} = (x_1, \ldots, x_n)$ ประเด็นทางเทคนิคที่สำคัญในที่นี้คือหลักการที่ว่า ข้อมูลที่ได้มานั้นมาจากการสุ่มจากการแจกแจงแบบมีเงื่อนไข (conditional distribution) ที่ถูกกำหนดโดยพารามิเตอร์ $\boldsymbol{\theta}$ ซึ่งมักแทนด้วย $f(\boldsymbol{x}|\boldsymbol{\theta})$

ทฤษฎีบทที่ 7.1. *กำหนดให้* $X_1, \ldots, X_n$ *คือตัวอย่างสุ่มที่เกิดจากการสุ่มเลือกจากการแจกแจง* $f(\boldsymbol{x}|\theta)$ *และกำหนดให้* $h_0(\theta)$ *แทนการแจกแจงก่อนการสังเกต (prior distribution) ของ* θ *แล้ว ฟังก์ชันความหนาแน่นของความน่าจะเป็นหลังการสังเกต (posterior p.d.f.) หรือฟังก์ชันความน่าจะเป็นหลังการสังเกต (posterior p.f.) เท่ากับ*

$$h(\theta|\boldsymbol{x}) = \frac{f(x_1|\theta)\cdots f(x_n|\theta)\, h(\theta)}{\int_{\hat{\theta}\in\Theta} f\left(\boldsymbol{x}|\hat{\theta}\right) h\left(\hat{\theta}\right) d\hat{\theta}},\ \text{สำหรับ } \theta \in \Theta \tag{7.1}$$

การพิสูจน์. จากนิยามของการแจกแจงแบบมีเงื่อนไข เราสามารถเขียนได้ว่า

$$h(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)}{g(\boldsymbol{x})},\ \text{สำหรับ } \theta \in \Theta$$

โดยที่ $g(\boldsymbol{x})$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ซึ่งมีค่าเท่ากับ

$$g(\boldsymbol{x}) = \int_{\hat{\theta}\in\Theta} f\left(\boldsymbol{x}|\hat{\theta}\right) h\left(\hat{\theta}\right) d\hat{\theta}$$

ส่วนตัวตั้งสามารถเขียนใหม่ได้เป็น

$$f(\boldsymbol{x},\theta) = f(\boldsymbol{x}|\theta)\, h(\theta)$$

ในขณะเดียวกัน การที่ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่ม ช่วยให้สามารถเขียน $f(\boldsymbol{x}|\theta)$ ได้เป็น

$$f(\boldsymbol{x}|\theta) = f(x_1|\theta)\cdots f(x_n|\theta)$$

ดังนั้น จึงสามารถสรุปได้ว่า

$$h\left(\theta|\boldsymbol{x}\right)=\frac{f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)h\left(\theta\right)}{\int_{\hat{\theta}\in\Theta}f\left(\boldsymbol{x}|\hat{\theta}\right)h\left(\hat{\theta}\right)d\hat{\theta}},\ \text{สำหรับ}\ \theta\in\Theta$$

■

ประเด็นที่น่าสังเกตอันหนึ่งคือ ตัวหารในสมการที่ 7.1 ไม่ขึ้นอยู่กับพารามิเตอร์ θ เพราะได้อินทิเกรต θ ออกไปหมดแล้ว และอันที่จริงพจน์นี้ทำหน้าที่หลักเป็นค่าคงที่ที่ทำให้การแจกแจงหลังการสังเกต (posterior distribution) มีคุณสมบัติที่เหมาะสม ซึ่งในที่นี้หมายถึง การที่ผลรวมหรือผลการอินทิเกรตของการแจกแจงหลังการสังเกต (posterior distribution) มีค่าเท่ากับหนึ่งนั่นเอง ดังนั้น บางครั้งเราอาจจะมองข้ามพจน์นี้ไปได้และเขียน $h\left(\theta|\boldsymbol{x}\right)$ ในรูปของการแปรผันตามส่วน (proportionality) ได้เป็น

$$h\left(\theta|\boldsymbol{x}\right)\propto f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)h\left(\theta\right) \tag{7.2}$$

ตัวอย่างที่ 7.2. กำหนดให้ θ แทนสัดส่วนของสินค้าที่มีตำหนิ (defective items) ที่ยังไม่ทราบค่า และการแจกแจงก่อนการสังเกต (prior distribution) เป็นแบบเอกรูป (uniform distribution) ในช่วง $[0,1]$ สิ่งที่ต้องการทราบคือการแจกแจงหลังการสังเกต (posterior distribution) ของ θ หลังจากสุ่มตรวจสินค้าทั้งหมด n ชิ้น

กำหนดให้ X_i แทนผลการตรวจสินค้าชิ้นที่ $i=1,\ldots,n$ โดยที่ $X_i=1$ ถ้าสินค้าที่ i มีตำหนิ ไม่เช่นนั้นจะมีค่าเท่ากับศูนย์ นั่นคือ X_i มีการแจกแจงแบบเบอร์นูลลี ซึ่งมีฟังก์ชันความน่าจะเป็น (p.f.) เท่ากับ

$$f\left(x_i|\theta\right)=\begin{cases}\theta_i^x\left(1-\theta\right)^{1-x_i},\ \text{สำหรับ}\ x_i=0,1\\ 0,\ \text{ถ้าเป็นอย่างอื่น}\end{cases}$$

การที่ตัวอย่างที่ได้มาเป็นตัวอย่างสุ่ม (random sample) ทำให้สามารถเขียนได้ว่า

$$f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)=\prod_{i=1}^{n}\theta_i^x\left(1-\theta\right)^{1-x_i}=\theta^{\sum_{i=1}^{n}x_i}\left(1-\theta\right)^{1-\sum_{i=1}^{n}x_i}=\theta^y\left(1-\theta\right)^{1-y}$$

โดยที่ $y=\sum_{i=1}^{n}x_i$ ส่วนการที่การแจกแจงก่อนการสังเกต (prior distribution) เป็นแบบเอกรูป (uniform distribution) ในช่วง $[0,1]$ ทำให้เขียนได้ว่า $h\left(\theta\right)=1$ ดังนั้น การแจกแจงหลังการสังเกต (posterior distribution) ของ θ เท่ากับ

$$h\left(\theta|\boldsymbol{x}\right)\propto\theta^y\left(1-\theta\right)^{1-y}$$

เมื่อพิจารณาให้ดีแล้วจะเห็นว่าพจน์ด้านขวานั้นเป็นส่วนหนึ่งของการแจกแจงเบต้า (beta distribution) ของ θ ที่มีค่าพารามิเตอร์ $\alpha=y+1$ และ $\beta=n-y+1$ ดังนั้น จึงสามารถกำหนดค่าคงที่ที่ต้องการได้โดยไม่ต้องอินทิ

เกรต ทำให้ได้การแจกแจงหลังการสังเกต (posterior distribution) ของ θ เป็น

$$h\left(\theta|\boldsymbol{x}\right)=\frac{\Gamma\left(n+2\right)}{\Gamma\left(y+1\right)\Gamma\left(n-y+1\right)}\theta^{y}\left(1-\theta\right)^{1-y}$$

สังเกตได้ว่า ในกรณีนี้ เราสามารถเขียนการแจกแจงร่วมในรูปของค่าสถิติ $y=\sum_{i=1}^{n}x_i$ นั่นคือ เราไม่จำเป็นต้องทราบค่า x_i แต่ละค่า สิ่งเดียวที่ต้องการทราบคือผลรวมหรือจำนวนสินค้าที่มีตำหนิเท่านั้น หลักการนี้ช่วยให้เราวิเคราะห์ปัญหาที่สนใจได้สะดวกขึ้นมากเพราะค่าผลรวมนี้เป็นค่าสถิติที่รวบรวมเอาสารสนเทศ (information) ทั้งหมดจาก $\boldsymbol{x}$ ที่จำเป็นในการวิเคราะห์ปัญหานี้ไว้อย่างครบถ้วน เราจะอภิปรายประเด็นนี้อย่างละเอียดในหัวข้อสถิติที่เพียงพอ (sufficient statistics) □

ในกรณีที่มีตัวอย่างสุ่มมากกว่าหนึ่งตัวอย่าง เราสามารถแปลงการแจกแจงได้ด้วยการปรับเป็นลำดับ (sequential updating) โดยเริ่มการแจกแจงหลังการสังเกต (posterior distribution) ซึ่งเป็นผลจากการสังเกตุ X_1

$$h\left(\theta|x_1\right)\propto f\left(x_1|\theta\right)h\left(\theta\right) \tag{7.3}$$

หลังจากนั้นจึงคำนวณหาการแจกแจงหลังการสังเกต (posterior distribution) จาก X_1,X_2 โดยใช้ $h\left(\theta|x_1\right)$ เป็นการแจกแจงก่อนการสังเกต (prior distribution)

$$h\left(\theta|x_1,x_2\right)\propto f\left(x_2|\theta\right)h\left(\theta|x_1\right)\propto f\left(x_1|\theta\right)f\left(x_2|\theta\right)h\left(\theta\right) \tag{7.4}$$

ดังนั้น การดำเนินการในรูปแบบนี้ต่อไปจนครบทั้งหมด n ครั้งก็จะได้การแจกแจงหลังการสังเกต (posterior distribution)

$$h\left(\theta|\boldsymbol{x}\right)\propto f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)h\left(\theta\right)=f\left(\boldsymbol{x}|\theta\right)h\left(\theta\right) \tag{7.5}$$

ซึ่งตรงกับการแจกแจงหลังการสังเกต (posterior distribution) ที่แสดงในสมการ (7.1) ซึ่งได้จากการปรับเพียงโดยใช้ข้อมูล $X_1,\ldots,X_n$ เพียงครั้งเดียว

อย่างไรก็ตาม การปรับเป็นลำดับ (sequential updating) นี้มีความสำคัญอย่างมากในโลกของข้อมูลขนาดใหญ่ (big data) เพราะพัฒนาการทางเทคโนโลยีทำให้มีการเพิ่มข้อมูลตลอดเวลา ในขณะเดียวกัน ก็มีความต้องการพยากรณ์ที่ทันท่วงที ซึ่งในที่นี้สามารถทำได้โดยการปรับการแจกแจงทุกครั้งที่มีการรับข้อมูลเข้ามาใหม่ด้วยหลักการปรับเป็นลำดับ (sequential updating) ซึ่งเป็นพื้นฐานสำคัญอันหนึ่งของรูปแบบการเรียนรู้ของเครื่องจักร (machine learning) นอกจากนี้ เครื่องมือที่นิยมใช้ในการปรับการแจกแจงหรือปรับความเชื่อ (updating beliefs) ในทางเศรษฐศาสตร์และการเงินคือการกรองแบบคาลแมน (Kalman filtering)

โดยทั่วไป การแจกแจงหลังการสังเกต (posterior distribution) จะมีความแตกต่างจากการแจกแจงก่อนการสังเกต (prior distribution) โดยสิ้นเชิง ดังแสดงในตัวอย่างที่ 7.2 ซึ่งเริ่มจากการแจกแจงเอกรูป (uniform distribution) แต่ได้การแจกแจงหลังการสังเกตที่เป็นการแจกแจงแบบเบต้า (beta distribution) แต่ก็มีการแจกแจง

บางรูปแบบที่ถ้าเริ่มจากการแจกแจงก่อนการสังเกต (prior distribution) ในกลุ่มนี้แล้วจะได้การแจกแจงหลังการสังเกต (posterior distribution) ในรูปแบบเดียวกันนี้ เราเรียกการแจกแจงที่มีคุณสมบัติแบบนี้ว่าการแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution) ตัวอย่างที่สำคัญอันหนึ่งของการแจกแจงก่อนการสังเกตคู่คือการแจกแจงปกติ (normal distribution)

ทฤษฎีบทที่ 7.2. *สมมุติว่า $X_1, \ldots, X_n$ คือตัวอย่างสุ่มที่สุ่มเลือกมาจากการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ ซึ่งไม่ทราบค่า และค่าความแปรปรวนเท่ากับ σ_x^2 ซึ่งทราบค่า และสมมุติว่าการแจกแจงก่อนการสังเกต (prior distribution) ของ μ เป็นการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_0 และค่าความแปรปรวนเท่ากับ σ_0^2 แล้วการแจกแจงหลังการสังเกต (posterior distribution) ของ μ หลังจากทราบค่าของ $X_1, \ldots, X_n$ เป็นการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_1 และค่าความแปรปรวนเท่ากับ σ_1^2 โดยที่*

$$\mu_1 = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{x}_n \tag{7.6}$$

$$\sigma_1^2 = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\sigma_0^2 \tag{7.7}$$

การพิสูจน์. พิจารณาฟังก์ชันความเป็นไปได้ (likelihood function)

$$f(\boldsymbol{x}|\mu) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

เนื่องจากสิ่งที่เราต้องการจริงๆ คือ รูปแบบของการแจกแจงหลังการสังเกต (posterior distribution) ที่เกี่ยวข้องกับ μ ดังนั้นเราจึงสามารถที่จะละเลยพจน์ที่แยกออกไปและที่ไม่เกี่ยวข้องกับ μ ได้โดยไม่ส่งผลเสียต่อสิ่งที่ต้องการพิสูจน์ โดยเริ่มจาก

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \mu)^2 &= \sum_{i=1}^{n}((x_i - \bar{x}_n) + (\bar{x}_n - \mu))^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{n}(\bar{x}_n - \mu)^2 \\ &\quad + (\bar{x}_n - \mu)\sum_{i=1}^{n}(x_i - \bar{x}_n) \\ &= \sum_{i=1}^{n}(\mu - \bar{x}_n)^2 + \sum_{i=1}^{n}(x_i - \bar{x}_n)^2\end{aligned}$$

เนื่องจากพจน์ด้านขวาไม่ขึ้นอยู่กับ μ ดังนั้น เราสามารถสรุปได้ว่าฟังก์ชันความเป็นไปได้ (likelihood function)

$$f(\boldsymbol{x}|\mu) \propto \exp\left[-\frac{1}{2\sigma^2}n(\mu - \bar{x}_n)^2\right]$$

ในขณะเดียวกัน การแจกแจงก่อนการสังเกต (prior distribution) ของ μ เขียนได้เป็น

$$h(\mu) \propto \exp\left[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right]$$

ดังนั้น การแจกแจงหลังการสังเกต (posterior distribution) เท่ากับ

$$\begin{aligned} h(\theta|\boldsymbol{x}) &\propto \exp\left[-\frac{n}{2\sigma^2}(\mu-\bar{x}_n)^2\right]\exp\left[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right] \\ &= \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\mu-\bar{x}_n)^2+\frac{1}{\sigma_0^2}(\mu-\mu_0)^2\right]\right\} \end{aligned}$$

ขั้นตอนต่อไปคือการแยกพจน์ที่ไม่เกี่ยวกับ μ ออกโดยการจัดรูป

$$\begin{aligned} \frac{n}{\sigma^2}(\mu-\bar{x}_n)^2+\frac{1}{\sigma_0^2}(\mu-\mu_0)^2 &= \frac{n}{\sigma^2}\left(\mu^2-2\mu\bar{x}_n+\bar{x}_n^2\right)+\frac{1}{\sigma_0^2}\left(\mu^2-2\mu\mu_0+\mu_0^2\right) \\ &= \frac{1}{\sigma^2\sigma_0^2}\left[\left(\sigma^2+n\sigma_0^2\right)\mu^2-2\left(\sigma^2\mu_0+n\sigma_0^2\bar{x}_n\right)\mu\right] \\ &+ \frac{1}{\sigma^2\sigma_0^2}\left[\sigma^2\mu_0^2+n\sigma_0^2\bar{x}_n^2\right] \\ &= \frac{\left(\sigma^2+n\sigma_0^2\right)}{\sigma^2\sigma_0^2}\left[\mu^2-2\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)\mu+\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)^2\right] \\ &+ \frac{1}{\sigma^2\sigma_0^2}\left[\sigma^2\mu_0^2+n\sigma_0^2\bar{x}_n^2-\left(\sigma^2+n\sigma_0^2\right)\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)^2\right] \\ &= \frac{1}{\sigma_1^2}(\mu-\mu_1)^2+\frac{n}{\sigma^2+n\sigma_0^2}(\bar{x}_n-\mu_0)^2 \end{aligned}$$

โดยที่ $\mu_1 = \frac{\sigma_x^2}{\sigma_x^2+n\sigma_0^2}\mu_0+\frac{n\sigma_0^2}{\sigma_x^2+n\sigma_0^2}\bar{x}_n$ และ $\sigma_1^2 = \frac{\left(\sigma^2+n\sigma_0^2\right)}{\sigma^2\sigma_0^2}$ เนื่องจากพจน์ด้านขวาไม่เกี่ยวข้อง μ ดังนั้นเราสามารถสรุปได้ว่า

$$h(\theta|\boldsymbol{x}) \propto \exp\left[-\frac{1}{\sigma_1^2}(\mu-\mu_1)^2\right]$$

■

ตัวอย่างต่อไปนี้เป็นการประยุกต์ใช้ทฤษฎีบทที่ 7.2 เพื่อพยากรณ์ความเสี่ยงเชิงระบบ (systematic risks) β ในทางการเงิน

ตัวอย่างที่ 7.3. พิจารณา □

ตัวประมาณค่าแบบเบส์ (Bayes Estimator) ของ θ คือตัวประมาณค่า (estimator) $\hat{\theta}(X_1,\ldots,X_n)$ ซึ่งทำให้ค่าคาดหมายของฟังก์ชันสูญเสีย (loss function) ที่คำนวณโดยใช้การแจกแจงหลังการสังเกต (posterior distribution) มีค่าต่ำที่สุด

บทนิยามที่ 7.7. *ฟังก์ชันสูญเสีย (loss function) หมายถึงฟังก์ชันของสองกลุ่มตัวแปร $L(\theta, a)$ ซึ่งตีความว่าเป็นการสูญเสียทางสถิติถ้าพารามิเตอร์มีค่าเท่ากับ θ แต่ตัวประมาณค่ามีค่าเท่ากับ a*

ในทางปฏิบัติ เรามักจะใช้ค่าคาดหมายของการสูญเสีย (expected loss)

$$E\left[L(\theta, a)\,|\boldsymbol{x}\right] = \int_{\Theta} L(\theta, a)\, h(\theta|\mathbf{x})\, d\theta \tag{7.8}$$

เป็นฟังก์ชันเป้าหมาย (objective function) ที่ใช้ในการเลือกตัวประมาณค่าแบบเบส์ (Bayes Estimator)

บทนิยามที่ 7.8. *กำหนดให้ $L(\theta, a)$ แทนฟังก์ชันสูญเสีย (loss function) และ $\hat{\theta}(\boldsymbol{x})$ เป็นคำตอบของปัญหาการหาค่าต่ำสุด (minimization problem) ต่อไปนี้*

$$E\left[L\left(\theta, \hat{\theta}(\boldsymbol{x})\right)\right] = \min_{a} E\left[L(\theta, a)\,|\boldsymbol{x}\right] \tag{7.9}$$

แล้ว ฟังก์ชัน $\hat{\theta}(\boldsymbol{X})$ คือตัวประมาณค่าแบบเบส์ (Bayes Estimator) ส่วนค่า $\hat{\theta}(\boldsymbol{x})$ คือค่าประมาณแบบเบส์ (Bayes estimate) ของ θ เมื่อข้อมูลที่ใช้ในการประมาณค่าคือ $\boldsymbol{x}$

แน่นอนว่า ตัวประมาณค่าแบบเบส์ (Bayes Estimator) ย่อมขึ้นอยู่กับรูปแบบของฟังก์ชันสูญเสียและ เพราะค่าคาดหมายของการสูญเสีย (expected loss) ซึ่งเป็นฟังก์ชันเป้าหมาย (objective function) ขึ้นอยู่กับทั้งสองอย่าง กล่าวอีกนัยหนึ่งได้ว่า การกำหนดรูปแบบฟังก์ชันสูญเสียที่แตกต่างกันย่อมนำไปสู่ตัวประมาณค่าที่แตกต่างกัน ในขณะเดียวกัน ก็ไม่มีทฤษฎีที่บอกได้ว่าควรจะใช้ฟังก์ชันสูญเสียแบบใดดี ดังนั้น จึงเป็นหน้าที่ของนักวิเคราะห์ที่จะต้องเลือกฟังก์ชันสูญเสียให้เหมาะสม ซึ่งอาจจะต้องอาศัยประสบการณ์เป็นสำคัญ

บทนิยามที่ 7.9. *ฟังก์ชันสูญเสียกำลังสอง (square error loss function) นิยามได้เป็น*

$$L(\theta, a) = (\theta - a)^2 \tag{7.10}$$

ทฤษฎีบทต่อไปนี้ระบุว่า ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) คือตัวประมาณค่าแบบเบส์ (Bayes Estimator) ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบกำลังสอง (square error loss function) อันที่จริงแล้วทฤษฎีบทนี้คือทฤษฎีบทที่ 4.28 ในรูปแบบของเบส์นั่นเอง ทำให้สามารถใช้วิธีการพิสูจน์แบบเดียวกันได้ แต่ขอแสดงวิธีการพิสูจน์อีกครั้งเพื่อให้เกิดความกระจ่าง

ทฤษฎีบทที่ 7.3. *สมมุติว่าฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบฟังก์ชันสูญเสียกำลังสอง (square error loss function) ดังแสดงในสมการที่ 7.10 แล้ว ตัวประมาณค่าแบบเบส์ (Bayes Estimator)*

$$\hat{\theta}(\boldsymbol{X}) = E\left[\theta|\boldsymbol{X}\right] = \int_{\Theta} \theta h(\theta|\boldsymbol{X})\, d\theta \tag{7.11}$$

สังเกตว่า ค่าคาดหมายในที่นี้คำนวณจากการแจกแจงหลังการสังเกต (posterior distribution) ของ θ

การพิสูจน์. พิจารณาปัญหาการหาค่าคาดหมายของการสูญเสีย (expected loss) ที่ต่ำที่สุดดังต่อไปนี้

$$\min_a E\left[L\left(\theta,a\right)|\boldsymbol{x}\right]=\min_a\int_\Theta\left(\theta-a\right)^2h\left(\theta|\boldsymbol{x}\right)d\theta$$

เงื่อนไขอันดับที่หนึ่ง (first-order condition) สำหรับปัญหานี้คือ

$$\left.\frac{\partial E\left[L\left(\theta,a\right)|\boldsymbol{x}\right]}{\partial a}\right|_{a=\hat{\theta}(\boldsymbol{x})}=0$$

นั่นคือ

$$\begin{aligned}\int_\Theta\left.\frac{\partial\left(\theta-a\right)^2}{\partial a}h\left(\theta|\boldsymbol{x}\right)d\theta\right|_{a=\hat{\theta}(\boldsymbol{x})}&=-2\int_\Theta\left(\theta-\hat{\theta}\left(\boldsymbol{x}\right)\right)h\left(\theta|\boldsymbol{x}\right)d\theta=0\\&\Rightarrow\hat{\theta}\left(\boldsymbol{x}\right)=\int_\Theta\theta h\left(\theta|\boldsymbol{x}\right)d\theta=E\left[\theta|\boldsymbol{x}\right]\end{aligned}$$

■

ตัวอย่างที่ 7.4. พิจารณาตัวอย่างสุ่ม $X_1,\ldots,X_n$ ซึ่งสุ่มเลือกมาจากการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมาย μ และค่าความแปรปรวน σ^2 สมมตุว่าเราทราบค่าความแปรปรวน แต่ไม่ทราบค่าคาดหมาย ดังนั้นจึงต้องการประมาณค่าคาดหมาย μ จากข้อมูลที่มีอยู่ด้วยการประมาณค่าแบบเบส์ (Bayes estimation) สมมุติอีกว่า การแจกแจงก่อนการสังเกต (prior distribution) ของ μ เป็นการแจกแจงแบบปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_0 และค่าความแปรปรวนเท่ากับ σ_0

ทฤษฎีบทที่ 7.3 ระบุว่าตัวประมาณค่าของเบส์ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบกำลังสองคือ ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) นั่นคือ

$$\hat{\theta}\left(\boldsymbol{X}\right)=E\left[\theta|\boldsymbol{X}\right]=\frac{\sigma_x^2}{\sigma_x^2+n\sigma_0^2}\mu_0+\frac{n\sigma_0^2}{\sigma_x^2+n\sigma_0^2}\bar{X}_n$$

โดยที่สมการสุดท้ายประยุกต์ใช้ผลลัพธ์จากทฤษฎีบทที่ 7.2 □

ทฤษฎีบทต่อไปนี้แสดงถึงตัวประมาณค่าแบบเบส์ (Bayes Estimator) ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) ในทำนองเดียวกับทฤษฎีบทก่อนหน้านี้ ทฤษฎีบทนี้เกี่ยวข้องกับทฤษฎีบทที่ 4.29 ดังนั้น จึงขอละเว้นการพิสูจน์ทฤษฎีบทนี้เพื่อประหยัดพื้นที่

บทนิยามที่ 7.10. *ฟังก์ชันสูญเสียแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) นิยามได้เป็น*

$$L\left(\theta,a\right)=\left|\theta-a\right| \tag{7.12}$$

ทฤษฎีบทที่ 7.4. *สมมุติว่าฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) ดังแสดงในสมการที่ 7.12 แล้ว ตัวประมาณค่าแบบเบส์ (Bayes Estimator) มีค่าเท่ากับค่ามัธยฐาน (median) ของการแจกแจงหลังการสังเกต (posterior distribution) ของ θ*

โดยทั่วไป ฟังก์ชันสูญเสียที่แตกต่างกันมักจะนำไปสู่ตัวประมาณค่าที่ต่างกันดังแสดงในทฤษฎีบทที่ 7.3 และทฤษฎีบทที่ 7.4 แต่อาจจะเป็นไปได้ว่าค่าประมาณ (estimate) ที่ได้จากตัวประมาณที่แตกต่างกันมีค่าเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบการแจกแจงหลังการสังเกต ดังแสดงในตัวอย่างต่อไปนี้สำหรับกรณีของการแจกแจงปกติซึ่งมีค่าคาดหมายและค่ามัธยฐานมีค่าเท่ากัน

ตัวอย่างที่ 7.5. พิจารณาสถานการณ์ที่เหมือนกับตัวอย่างที่ 7.4 แต่คราวนี้กำหนดให้ฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) ดังนั้น จากทฤษฎีบทที่ 7.4 ตัวประมาณค่าแบบเบส์ (Bayes Estimator) มีค่าเท่ากับค่ามัธยฐาน (median) ของการแจกแจงหลังการสังเกต (posterior distribution) ของ μ ซึ่งในที่นี้จะมีค่าเท่ากับค่าคาดหมาย นั่นคือ

$$\hat{\theta}\left(\boldsymbol{X}\right) = E\left[\theta|\boldsymbol{X}\right] = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{X}_n$$

□

KEI AND FAI WE SHOULD HAVE AN EXERCISE FOR THEM TO ESTIMATE RETURNS HERE

7.2.1 คุณสมบัติความคงเส้นคงวาของตัวประมาณค่าแบบเบส์เมื่อตัวอย่างมีขนาดใหญ่

ความคงเส้นคงวา (consistency) เป็นคุณสมบัติการลู่เข้าของตัวประมาณค่าที่เกิดจากการเพิ่มขึ้นของขนาดตัวอย่าง (sample size) จนมีขนาดเข้าใกล้อนันต์ ($n \to \infty$) โดยใช้หลักการลู่เข้าเชิงความน่าจะเป็น (convergence in probability) เป็นเครื่องมือในการวิเคราะห์ ถึงแม้ว่าในโลกความจริง เราจะไม่เคยมีข้อมูลขนาดอนันต์ แต่หลักการความคงเส้นคงวา (consistency) ก็เป็นเครื่องมือทางสถิติที่มีประโยชน์และสะดวกในการใช้งาน โดยที่บางครั้งอาจจะเป็นเครื่องมือเดียวที่สามารถบอกถึงความแม่นยำของตัวประมาณค่า เพราะไม่สามารถพิสูจน์ในกรณีที่มีตัวอย่างจำกัดได้ (finite sample) ทั้งนี้เพราะว่าคุณสมบัติของการลู่เข้าเชิงความน่าจะเป็น (convergence in probability) สามารถส่งผ่านฟังก์ชันที่ต่อเนื่องใดๆ ในขณะที่หลักการหาค่าคาดหมาย (expectation) สามารถส่งผ่าานได้เพียงฟังก์ชันเชิงเส้นเท่านั้น

บทนิยามที่ 7.11. *ตัวประมาณค่า $\hat{\theta}$ ของพารามิเตอร์ θ มีความคงเส้นคงว่า (consistent) ถ้า*

$$\hat{\theta} \xrightarrow{p} \theta \tag{7.13}$$

คำถามที่ตามมาก็คือ ตัวประมาณค่าแบบเบส์ (Bayes estimator) มีความคงเส้นคงวา (consistent) หรือไม่? คำตอบโดยทั่วไปก็คือ ภายใต้เงื่อนไขที่ค่อนข้างมาตรฐาน ตัวประมาณค่าแบบเบส์ (Bayes estimator) มีความคงเส้นคงวา (consistent) แต่การพิสูจน์ความคงเส้นคงวาของตัวประมาณค่าแบบเบส์ (Bayes estimator) อยู่เหนือขอบเขตของหนังสือเล่มนี้ เพราะจำเป็นต้องใช้เทคนิคขั้นสูงของทฤษฎีการวัด (measure theory) ผู้อ่านสามารถศึกษาเพิ่มเติมได้จาก Prakasa Rao (1987) และดูต้นฉบับของการพิสูจน์ใน Doob (1949)

แต่อย่างไรก็ตาม ยังสามารถแสดงให้เห็นถึงความคงเส้นคงวา (consistency) ของตัวประมาณค่าแบบเบส์ (Bayes estimator) ได้อย่างไม่ยากเย็น โดยใช้ตัวอย่างต่อไปนี้

ตัวอย่างที่ 7.6. พิจารณาสถานการณ์ที่เหมือนกับตัวอย่างที่ 7.4 ดังนั้น ตัวประมาณค่าแบบเบส์ (Bayes estimator) เท่ากับ

$$\hat{\theta}\left(\boldsymbol{X}\right)=\frac{\sigma_x^2}{\sigma_x^2+n\sigma_0^2}\mu_0+\frac{n\sigma_0^2}{\sigma_x^2+n\sigma_0^2}\bar{X}_n$$

ขั้นตอนต่อไปคือ การตรวจสอบว่า $\hat{\theta}\left(\boldsymbol{X}\right)$ ลู่เข้าเชิงความน่าจะเป็นสู่ μ หรือไม่?

$$\begin{aligned}plim_{n\to\infty}\hat{\theta}\left(\boldsymbol{X}\right)&=\left[\lim_{n\to\infty}\frac{\sigma_x^2}{\sigma_x^2+n\sigma_0^2}\right]\mu_0+\left[\lim_{n\to\infty}\frac{n\sigma_0^2}{\sigma_x^2+n\sigma_0^2}\right]\left[plim_{n\to\infty}\bar{X}_n\right]\\&=\left[0\right]\mu_0+\left[1\right]\left[plim_{n\to\infty}\bar{X}_n\right]=\mu\end{aligned}$$

โดยที่สมการสุดท้ายเป็นผลมาจากกฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) □

บทเรียนอีกอย่างหนึ่งที่ได้จากตัวอย่างนี้คือ โดยทั่วไป การแจกแจงก่อนการสังเกต (prior distribution) ที่แตกต่างกันมักนำไปสู่การแจกแจงหลังการสังเกต (posterior distribution) ที่แตกต่างกัน ซึ่งส่งผลให้ได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) ที่แตกต่างกันด้วย แต่เมื่อตัวอย่างมีขนาดใหญ่มากพอ ผลของการแจกแจงก่อนการสังเกต (prior distribution) ดังกล่าวจะหมดไป ทำให้ได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) ที่เหมือนกัน ไม่ว่าจะเริ่มด้วยการแจกแจงก่อนการสังเกต (prior distribution) แบบใด หรือหากพิจารณาจากมุมมองของการปรับเป็นลำดับ (sequential updating) อาจจะสรุปได้ว่า เมื่อเราปรับการแจกแจง (updating distribution) ไปเรื่อยๆ ก็จะได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) ที่เหมือนกัน ไม่ว่าจะเริ่มด้วยการแจกแจงก่อนการสังเกต (prior distribution) แบบใดก็ตาม

ตัวอย่างที่ 7.7. ตัวอย่างต่อไปนี้แสดงการประมาณค่าแบบเบส์ (Bayes estimation) จากการแจกแจงก่อนการสังเกต (prior distribution) ที่แตกต่างกันสองอัน อันแรกเป็นการแจกแจงปกติที่มีค่าคาดหมาย $\mu_0=10$ และค่าความแปรปรวน $\sigma_0^2=10$ ส่วนอันที่สองเป็นการแจกแจงปกติที่มีค่าคาดหมาย $\mu_0=100$ และค่าความแปรปรวน $\sigma_0^2=100$ ในขณะที่ กลุ่มตัวอย่างที่ใช้ในตัวอย่างนี้สุ่มเลือกมาจากการแจกแจงปกติที่มีค่าคาดหมาย $\mu=50$ และค่าความแปรปรวน $\sigma^2=50$ (ด้วยการจำลอง (simulation) ในคอมพิวเตอร์)

รูปที่ XXX แสดงค่าประมาณแบบเบส์ (Bayes estimate) สำหรับการแจกแจงก่อนการสังเกต (prior distribution) ทั้งสองอัน

ADD FIGURE WHEN X-AXIS IS THE NUMBER OF SAMPLE AND Y IS THE ESTIMATES

บทเรียนที่สำคัญจากตัวอย่างนี้คือ ในช่วงแรกที่ขนาดของตัวอย่างยังไม่มากนัก ค่าประมาณแบบเบส์ (Bayes estimate) ที่ได้จากการแจกแจงก่อนการสังเกต (prior distribution) ทั้งสองอัน มีความแตกต่างกันอย่างชัดเจน แต่เมื่อตัวอย่างมีจำนวนมากพอ ความแตกต่างดังกล่าวแทบจะไม่เหลืออยู่เลย □

7.3 วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation)

หลักการสำคัญของการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation หรือเรียกสั้นๆ ว่า MLE) คือ ข้อมูลที่ได้จากการสังเกตเกิดจากการสุ่มเลือกมาจากประชากรที่มีการแจกแจงที่ขึ้นอยู่กับค่าพารามิเตอร์เฉพาะค่าหนึ่ง ซึ่งควรจะเป็นค่าที่ทำให้ความเป็นไปได้ (likelihood) ที่จะสุ่มเลือกได้ตัวอย่างดังกล่าวมีค่าสูงสุด

บทนิยามที่ 7.12. *ฟังก์ชันความเป็นไปได้ (likelihood function) คือฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (joint p.d.f.) หรือฟังก์ชันความน่าจะเป็นร่วม (joint p.f.)* $f(\boldsymbol{x}|\theta)$ *โดยที่* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *คือค่าที่เกิดขึ้นจริงที่ได้จากข้อมูล*

ในมุมหนึ่งรูปแบบของฟังก์ชันความเป็นไปได้ (likelihood function) ขึ้นอยู่กับข้อมูลที่ได้จากการสังเกต ดังนั้น หากตัวแปรที่เกี่ยวข้องเป็นแบบไม่ต่อเนื่อง (discrete) ฟังก์ชันความเป็นไปได้ (likelihood function) ก็จะอยู่ในรูปแบบฟังก์ชันความน่าจะเป็นร่วม (joint p.f.) ถึงแม้ว่าตัวแปรสุ่มพื้นฐานที่ระบุในแบบจำลองจะเป็นแบบต่อเนื่องก็ตาม ในอีกด้านหนึ่ง ฟังก์ชันความเป็นไปได้ (likelihood function) ก็เป็นฟังก์ชันของพารามิเตอร์ที่ต้องการทราบค่า θ โดยมองว่าค่าของ $\boldsymbol{x} = (x_1, \ldots, x_n)$ เป็นเพียงค่าคงที่ ซึ่งทราบค่าแล้วจากข้อมูลที่มีอยู่ ดังนั้น ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) จึงหมายถึงการแก้ปัญหาการหาค่าสูงสุดซึ่งมีฟังก์ชันความเป็นไปได้ (likelihood function) เป็นฟังก์ชันจุดประสงค์ (objective function) และพารามิเตอร์ที่ต้องการทราบค่า θ เป็นตัวแปรที่ต้องเลือก

บทนิยามที่ 7.13. *พิจารณาข้อมูลที่สังเกตได้ (observed data)* $\boldsymbol{x}$ *ชุดหนึ่ง และ* $\hat{\theta}_{MLE}(\boldsymbol{x})$ *เป็นคำตอบของปัญหาการหาค่าสูงสุด (maximization problem) ต่อไปนี้*

$$f\left(\boldsymbol{x}|\hat{\theta}_{MLE}(\boldsymbol{x})\right) = \max_{\theta} f(\boldsymbol{x}|\theta) \tag{7.14}$$

หรือเขียนอีกรูปแบบหนึ่งได้เป็น

$$\hat{\theta}_{MLE}(\boldsymbol{x}) = \underset{\theta}{\text{argmax}}\, f(\boldsymbol{x}|\theta) \tag{7.15}$$

แล้ว ฟังก์ชัน $\hat{\theta}_{MLE}(\boldsymbol{X})$ คือตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimator หรือเรียกสั้นๆ ว่า MLE) ส่วนค่า $\hat{\theta}_{MLE}(\boldsymbol{x})$ คือค่าประมาณด้วยความเป็นไปได้สูงสุดของ θ เมื่อข้อมูลที่ใช้ในการประมาณค่าคือ $\boldsymbol{x}$

โดยทั่วไป ฟังก์ชันความเป็นไปได้ (likelihood function) มักอยู่ในรูปของผลคูณของฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) หรือฟังก์ชันความน่าจะเป็น (p.f.) ดังนั้น เพื่อความสะดวกจึงมักจะแปลงฟังก์ชันจุดประสงค์ (objective function) ให้อยู่ในรูปของล็อกการิธึมแทน ซึ่งจะอยู่ในรูปของผลบวกซึ่งจัดการได้ง่ายกว่ามาก และที่สำคัญยังได้คำตอบเท่าเดิม เพราะการแปลงค่าฟังก์ชันจุดประสงค์โดยใช้ฟังก์ชันที่เพิ่มขึ้นทางเดียว (monotonically increasing) จะไม่ทำให้คำตอบของปัญหาการหาค่าสูงสุดเปลี่ยนไป

ตัวอย่างที่ 7.8. พิจารณาตัวอย่าง $X_1, \ldots, X_n$ โดยที่ X_i มีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) สำหรับพารามิเตอร์ θ ซึ่งเป็นสิ่งที่ต้องการประมาณค่า และกำหนดให้ $\boldsymbol{x} = (x_1, \ldots, x_n)$ คือค่าที่เกิดขึ้นจริงที่ได้จากข้อมูล ดังนั้น ฟังก์ชันความเป็นไปได้ (likelihood function) ในกรณีนี้เท่ากับ

$$\mathscr{L}(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} \theta_i^x (1-\theta)^{1-x_i} \tag{7.16}$$

ดังนั้น ฟังก์ชันล็อกการิธึมของความเป็นไปได้ (log-likelihood function) เท่ากับ

$$\begin{aligned}\log \mathscr{L} &= \sum_{i=1}^{n} [x_i \log\theta + (1-x_i)\log(1-\theta)] \\ &= \left(\sum_{i=1}^{n} x_i\right)\log\theta + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-\theta)\end{aligned}$$

ปัญหาการหาค่าสูงสุดในกรณีคือ

$$\max_{\theta} \left(\sum_{i=1}^{n} x_i\right)\log\theta + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

ซึ่งสามารถหาคำตอบได้โดยการกำหนดให้ค่าอนุพันธ์อันดับที่หนึ่งของฟังก์ชันจุดประสงค์ (objective function) หรือฟังก์ชันล็อกการิธึมของความเป็นไปได้ (log-likelihood function) เท่ากับศูนย์ หรือที่เรียกว่า เงื่อนไขอันดับที่หนึ่ง (first-order conditions)

$$\begin{aligned}\left.\frac{\partial \log \mathscr{L}}{\partial\theta}\right|_{\hat{\theta}_{MLE}(\boldsymbol{x})} = 0 &\Rightarrow \left(\sum_{i=1}^{n} x_i\right)\frac{1}{\hat{\theta}_{MLE}(\boldsymbol{x})} - \left(n - \sum_{i=1}^{n} x_i\right)\frac{1}{1-\hat{\theta}_{MLE}(\boldsymbol{x})} = 0 \\ &\Rightarrow \hat{\theta}_{MLE}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n\end{aligned}$$

ดังนั้น ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE)

$$\hat{\theta}_{MLE}\left(\boldsymbol{X}\right)=\frac{\sum_{i=1}^{n}X_{i}}{n}=\bar{x}_{n}$$

□

สังเกตว่า ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) $\hat{\theta}_{MLE}\left(\boldsymbol{X}\right)$ ในตัวอย่างข้างบนนั้นเขียนอย่างชัดเจนว่าเป็นฟังก์ชันของตัวอย่าง $\boldsymbol{X}$ ทั้งนี้เพื่อชี้ให้เห็นอย่างชัดเจนว่าตัวประมาณค่าเป็นฟังก์ชันของตัวอย่าง แต่หลังจากนี้ต่อไป จะขอไม่เขียนส่วนนี้ ทั้งนี้เพื่อความกระชับ โดยจะเขียนสั้นๆ เป็น $\hat{\theta}_{MLE}$ แต่ขอให้เข้าใจว่ามันคือ $\hat{\theta}_{MLE}\left(\boldsymbol{X}\right)$

ตัวอย่างที่ 7.9. สมมุติให้ $X_1,\ldots,X_n$ เป็นกลุ่มตัวอย่างสุ่มที่เลือกมาการแจกแจงแบบปกติ (normal distribution) ที่มีค่าคาดหมาย μ และค่าความแปรปรวน σ^2 โดยทั้งสองเป็นพารามิเตอร์ที่ต้องการทราบค่า และกำหนดให้ $\boldsymbol{x}=(x_1,\ldots,x_n)$ คือค่าที่เกิดขึ้นจริงที่ได้จากข้อมูล ดังนั้น ฟังก์ชันความเป็นไปได้ (likelihood function) ในกรณีนี้เท่ากับ

$$\mathscr{L}=\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\middle|\mu,\sigma^2\right) \tag{7.17}$$

ดังนั้น ฟังก์ชันล็อกการิธึมของความเป็นไปได้ (log-likelihood function) เท่ากับ

$$\log\mathscr{L}=-\frac{n}{2}\log\left(2\pi\right)-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \tag{7.18}$$

ปัญหาการหาค่าสูงสุดในกรณีคือ

$$\max_{\mu,\sigma^2}-\frac{n}{2}\log\left(2\pi\right)-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

เงื่อนไขอันดับที่หนึ่ง (first-order conditions) ของปัญหานี้เท่ากับ

$$\left.\frac{\partial\log\mathscr{L}}{\partial\mu}\right|_{\hat{\mu}_{MLE},\hat{\sigma}^2_{MLE}}=0\Rightarrow\frac{1}{\hat{\sigma}^2_{MLE}}\sum_{i=1}^{n}\left(x_i-\hat{\mu}_{MLE}\right)=0 \tag{7.19}$$

$$\left.\frac{\partial\log\mathscr{L}}{\partial\sigma^2}\right|_{\hat{\mu}_{MLE},\hat{\sigma}^2_{MLE}}=0\Rightarrow-\frac{n}{2\hat{\sigma}^2_{MLE}}+\frac{1}{2\left(\hat{\sigma}^2_{MLE}\right)^2}\sum_{i=1}^{n}\left(x_i-\hat{\mu}_{MLE}\right)^2=0 \tag{7.20}$$

ซึ่งอยู่ในรูปของระบบสมการของสองตัวแปร และสามารถหาคำตอบได้เป็น

$$\hat{\mu}_{MLE}=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}_n \tag{7.21}$$

$$\hat{\sigma}^2_{MLE}=\frac{1}{n}\left(x_i-\bar{x}_n\right)^2 \tag{7.22}$$

เรามักเรียกตัวประมาณค่าสำหรับค่าคาดหมาย $\bar{X}_n$ ว่าค่าคาดหมายจากตัวอย่าง (sample mean) และตัวประมาณค่าสำหรับค่าความแปรปรวน $\frac{1}{n}\left(x_i - \bar{X}_n\right)^2$ ว่าค่าความแปรปรวนจากตัวอย่าง (sample variance) ซึ่งแตกต่างจาก $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$ ตรงที่ส่วนตัวหารที่ $\hat{\sigma}_{MLE}^2$ นั้นหารด้วย n แต่ S_n^2 หารด้วย $n-1$ ตัวประมาณค่าทั้งสองของความแปรปรวนนี้มีคุณสมบัติลู่เข้าเชิงความน่าจะเป็นที่เหมือนกันเพราะ n และ $n-1$ นั้นมีพฤติกรรมที่ลิมิตเหมือนกัน แต่มีคุณสมบัติความไม่เบี่ยงเบน (unbiasness) ต่างกัน ซึ่งจะอภิปรายในรายละเอียดในหัวข้อที่ XXX □

ตัวอย่างที่ 7.10. สมมุติให้ $X_1, \ldots, X_n$ เป็นกลุ่มตัวอย่างสุ่มที่เลือกมาการแจกแจงแบบเอกรูป (uniform distribution) ในช่วง $[0, c]$ โดยในที่นี้ c คือพารามิเตอร์ที่ต้องการทราบค่า และกำหนดให้ $\boldsymbol{x} = (x_1, \ldots, x_n)$ คือค่าที่เกิดขึ้นจริงที่ได้จากข้อมูล ดังนั้น ฟังก์ชันความเป็นไปได้ (likelihood function) ในกรณีนี้เท่ากับ

$$\mathscr{L} = \prod_{i=1}^{n} \frac{1}{c} = \frac{1}{c^n}, \text{ สำหรับ } 0 \leqslant x_i \leqslant c \text{ สำหรับทุก } i = 1, \ldots, n \tag{7.23}$$

ส่วนกรณีอื่นๆ $\mathscr{L} = 0$ ปัญหาการหาค่าสูงสุดในกรณีคือ

$$\max_c \frac{1}{c^n}$$

โดยที่

$$0 \leqslant x_i \leqslant c \text{ สำหรับทุก } i = 1, \ldots, n$$

ปัญหานี้เป็นปัญหาการหาค่าสูงสุดแบบมีข้อจำกัด (constrained optimization problem) หากไม่มีข้อจำกัดคำตอบของ c ที่ทำให้ค่าฟังก์ชันความเป็นไปได้ (likelihood function) มีค่าสูงสุดคือ $c = 0$ แต่ข้อจำกัด $0 \leqslant x_i \leqslant c$ ทำให้ค่า c จะต้องไม่น้อยกว่าค่า x_i แต่ละค่า ซึ่งทำให้ $c = 0$ เป็นไปไม่ได้ (ยกเว้น $X_i = 0$ ทุกค่า ซึ่งเป็นกรณีที่ไม่น่าสนใจ) คำตอบของปัญหานี้จึงถูกกำหนดโดยข้อจำกัดเป็นหลัก นั่นคือ $\hat{c}_{MLE} \geqslant x_i$ สำหรับทุก $i = 1, \ldots, n$ แต่ในขณะเดียวกันฟังก์ชันวัตถุประสงค์ก็ต้องการให้ค่า $\hat{c}_{MLE}$ มีค่าต่ำที่สุดเท่าที่จะเป็นไปได้ ดังนั้นคำตอบที่ได้คือ $\hat{c}_{MLE} = \max(x_1, \ldots, x_n)$

บทเรียนทางคณิตศาสตร์อันหนึ่งในตัวอย่างนี้คือ บางครั้งเราอาจจะไม่สามารถแก้ปัญหาการหาค่าสูงสุดแบบมีข้อจำกัด (constrained optimization problem) ได้โดยใช้เงื่อนไขอันดับที่หนึ่ง (first-order conditions) โดยตรงเหมือนกับตัวอย่างที่ 7.9 ทั้งนี้เพราะบางครั้งคำตอบที่ได้อาจจะอยู่บนขอบ (boundary) ของขอบเขตที่เป็นไปได้ ในขณะที่ เงื่อนไขอันดับที่หนึ่ง (first-order conditions) ใช้ได้กับกรณีที่คำตอบอยู่ภายในขอบเขตที่เป็นไปได้ (interior solution) □

ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) มีคุณสมบัติอย่างหนึ่งที่สำคัญที่ช่วยให้สามารถประยุกต์ใช้ได้อย่างกว้างขวาง โดยเรามักเรียกคุณสมบัตินี้ว่า คุณสมบัติไม่แปรเปลี่ยน (invariance property) ซึ่งมีความหมายดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 7.5. *ถ้า $\hat{\theta}_{MLE}$ คือตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) ของ θ และ g เป็นฟังก์ชันที่กำหนดโดย $\gamma = g(\theta)$ แล้ว $g\left(\hat{\theta}_{MLE}\right)$ เป็นตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) ของ γ*

การพิสูจน์. เพื่อความสะดวก ขอนำเสนอการพิสูจน์ในกรณีที่ g เป็นฟังก์ชันหนึ่งต่อหนึ่ง (one-on-one) เท่านั้น โดยเริ่มจาก การที่ $\hat{\theta}_{MLE}$ เป็นตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) ของ θ หมายความว่า

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f_i(x_i|\theta) \tag{7.24}$$

ในขณะเดียวกัน การที่ g เป็นฟังก์ชันหนึ่งต่อหนึ่ง (one-on-one) หมายความว่า เราสามารถใช้ฟังก์ชันส่วนกลับได้เป็น $\theta = g^{-1}(\gamma)$ ดังนั้น เราสามารถเขียนฟังก์ชันความเป็นไปได้ (likelihood function) ในรูปฟังก์ชันของ γ ได้เป็น

$$\mathscr{L}(\gamma|\boldsymbol{x}) = \underset{\gamma}{\operatorname{argmax}} \prod_{i=1}^{n} f_i\left(x_i|g^{-1}(\gamma)\right) \tag{7.25}$$

ดังนั้น เนื่องจากฟังก์ชันความเป็นไปได้ (likelihood function) ในสมการที่ 7.24 มีค่าสูงสุดเมื่อพารามิเตอร์ θ มีค่าเท่ากับ $\hat{\theta}_{MLE}$ ดังนั้น ฟังก์ชันความเป็นไปได้ (likelihood function) ในสมการที่ 7.25 จะมีค่าสูงสุด ถ้าพารามิเตอร์ γ มีค่าเท่ากับ $\hat{\gamma}_{MLE}$ ที่ทำให้ $\hat{\theta}_{MLE} = g^{-1}(\hat{\gamma}_{MLE})$ นั่นคือ $\hat{\gamma}_{MLE} = g\left(\hat{\theta}_{MLE}\right)$

ส่วนในกรณีที่ฟังก์ชัน g ไม่ใช่ฟังก์ชันหนึ่งต่อหนึ่ง ก็สามารถพิสูจน์ได้ด้วยวิธีการที่คล้ายคลึงกัน แต่ต้องตระหนักว่า สิ่งที่ต้องการคือค่าที่ต้องการนำไปสู่ค่าสูงสุด ถึงแม้ว่าอาจจะมีค่าพารามิเตอร์ที่ทำให้ได้ค่าสูงสุดหลายค่า ■

ตัวอย่างที่ 7.11. สมมุติให้ $X_1, \ldots, X_n$ เป็นกลุ่มตัวอย่างสุ่มที่เลือกมาการแจกแจงแบบเอกรูป (uniform distribution) ในช่วง $[0, c]$ เช่นเดียวกับตัวอย่างที่ 7.10 ซึ่งพิสูจน์แล้วว่าตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) ของ c มีค่าเท่ากับ $\hat{c}_{MLE} = \max(x_1, \ldots, x_n)$

เราสามารถประยุกต์ใช้คุณสมบัติไม่แปรเปลี่ยน (invariance property) เพื่อประมาณค่าความแปรปรวน (variance) ของ X ได้โดยเริ่มจากการที่ทราบว่า $Var[X] = \frac{c^2}{12}$ ในขณะเดียวกันเราก็ทราบว่า $\hat{c}_{MLE} = \max(x_1, \ldots, x_n)$ ดังนั้น

$$\widehat{Var}[X] = \frac{\hat{c}^2_{MLE}}{12} = \frac{\max(x_1, \ldots, x_n)^2}{12}$$

□

คุณสมบัติอันหนึ่งที่สำคัญของตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) คือความคงเส้นคงวา (consistency) นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากพอแล้ว ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) จะมีค่าใกล้เคียงกับค่าที่แท้จริง

ทฤษฎีบทและการพิสูจน์ด้านล่างนี้ดัดแปลงมาจาก Hogg et al. (2005) โดยเริ่มจากการกำหนดเงื่อนไขปกติ (regularity conditions) ต่อไปนี้

RC1 ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) $f(x,\theta)$ แตกต่างกันชัดเจน นั่นคือ ถ้า $\theta \neq \theta'$ แล้ว $f(x,\theta) \neq f(x,\theta')$

RC2 ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) $f(x,\theta)$ มีส่วนค้ำจุน (support) S เหมือนกันสำหรับทุกๆ ค่าพารามิเตอร์ θ

RC3 ค่าที่แท้จริง θ_0 ของพารามิเตอร์ θ เป็นจุดภายใน (interior point) ของเซ็ต Θ

ทฤษฎีบทที่ 7.6. *กำหนดให้ θ_0 เป็นค่าที่แท้จริงของพารามิเตอร์ที่สนใจ ดังนั้น ภายใต้เงื่อนไขปกติ (regularity conditions) RC1-RC2*

$$\lim_{n\to\infty} Pr_{\theta_0}\left[\mathscr{L}(\theta_0|\boldsymbol{X}) > \mathscr{L}(\theta|\boldsymbol{X})\right] = 1, \text{ สำหรับทุกๆ ค่า } \theta \neq \theta_0 \tag{7.26}$$

การพิสูจน์. เริ่มจากนิยามของฟังก์ชันความเป็นไปได้ (likelihood function)

$$\mathscr{L}(\theta_0|\boldsymbol{X}) = \prod_{i=1}^{n} f(X_i|\theta_0) \Rightarrow \log \mathscr{L}(\theta_0|\boldsymbol{X}) = \sum_{i=1}^{n} \log f(X_i|\theta_0)$$
$$\mathscr{L}(\theta|\boldsymbol{X}) = \prod_{i=1}^{n} f(X_i|\theta) \Rightarrow \log \mathscr{L}(\theta_0|\boldsymbol{X}) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

ในขณะเดียวกัน เราสามารถเปลี่ยนเงื่อนไข $\mathscr{L}(\theta_0|\boldsymbol{X}) > \mathscr{L}(\theta|\boldsymbol{X})$ ให้อยู่ในรูปของล็อกการิธึมได้เป็น

$$\log \mathscr{L}(\theta_0|\boldsymbol{X}) > \log \mathscr{L}(\theta|\boldsymbol{X}) \Rightarrow \sum_{i=1}^{n} \log \frac{f(X_i|\theta)}{f(X_i|\theta_0)} < 0$$

เพื่อให้สามารถประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ได้ เราจึงคูณ $\frac{1}{n}$ เพิ่มเข้าไป โดยที่อสมการยังมีทิศทางเช่นเดิม

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{f(X_i|\theta)}{f(X_i|\theta_0)} < 0$$

ในขณะเดียวกัน กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ช่วยให้สามารถสรุปได้ว่า

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \xrightarrow{p} E_{\theta_0}\left[\log \frac{f(X_i|\theta)}{f(X_i|\theta_0)}\right]$$

สังเกตุว่าค่าคาดหมายในที่นี้เป็นการหาค่าคาดหมายภายใต้การแจกแจงด้วยพารามิเตอร์จริง θ_0 ส่วนในขั้นตอนต่อไป จะประยุกต์ใช้อสมการของเจนเซ็น (Jensen's Inequality) เพื่อสรุปว่า

$$E_{\theta_0}\left[\log \frac{f\left(X_i|\theta\right)}{f\left(X_i|\theta_0\right)}\right] < \log E_{\theta_0}\left[\frac{f\left(X_i|\theta\right)}{f\left(X_i|\theta_0\right)}\right]$$

ทั้งนี้เนื่องจากฟังก์ชันล็อกการิธึมเป็นฟังก์ชันนูน (convex function) ส่วนที่เหลือคือต้องแสดงว่า

$$E_{\theta_0}\left[\frac{f\left(X_i|\theta\right)}{f\left(X_i|\theta_0\right)}\right] = \int_S \left[\frac{f\left(x|\theta\right)}{f\left(x|\theta_0\right)}\right] f\left(x|\theta_0\right) dx = \int_S f\left(x|\theta\right) dx = 1$$

สังเกตว่าขอบเขตของการอินทิเกรตในที่นี้คือส่วนค้ำจุนของ $f\left(x|\theta_0\right)$ ในขณะที่สมการที่สมการสุดท้ายเป็นการอินทิเกรตฟังก์ชัน $f\left(x|\theta\right)$ ซึ่งจะมีค่าเท่ากับหนึ่งก็ต่อเมื่อขอบเขตของการอินทิเกรตที่ใช้เท่ากับส่วนส่วนค้ำจุนของ $f\left(x|\theta\right)$ ซึ่งในที่นี้เป็นจริงได้เพราะเงื่อนไขปกติ RC2 ■

อันที่จริงแล้ว ทฤษฎีบทที่ 7.6 เป็นเครื่องยืนยันว่า ค่าพารามิเตอร์ที่ทำให้ฟังก์ชันความเป็นไปได้ (likelihood function) มีค่าสูงสุดที่ลิมิต (asymptotically) คือค่าพารามิเตอร์ที่แท้จริง (true parameter) นั่นเอง ซึ่งเป็นเหตุผลที่ทำให้เชื่อได้ว่า เมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากพอ ตัวประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) น่าจะมีค่าใกล้เคียงกับค่าที่แท้จริงของพารามิเตอร์ (true parameter)

ทฤษฎีบทที่ 7.7. *สมมุติว่าตัวแปรสุ่ม $X_1, \ldots, X_n$ สอดคล้องกับเงื่อนไขปกติ (regularity conditions) RC1-RC3 โดยที่ θ_0 เป็นค่าพารามิเตอร์ที่แท้จริง (true parameter) และ $f\left(x|\theta\right)$ สามารถหาค่าอนุพันธ์เทียบกับ θ ได้ แล้ว คำตอบของสมการความเป็นไปได้ (likelihood equation) $\hat{\theta}$ ซึ่งสอดคล้องกับสมการต่อไปนี้*

$$\frac{\partial \log \mathscr{L}\left(\hat{\theta}|\boldsymbol{X}\right)}{\partial \theta} = 0 \tag{7.27}$$

จะมีความคงเส้นคงวา (consistent) นั่นคือ $\hat{\theta} \xrightarrow{p} \theta_0$

การพิสูจน์. เนื่องจาก θ_0 เป็นจุดภายใน (interior point) ของเซ็ต Θ (RC3) นั่นคือ สามารถหาจำนวนจริง $\delta > 0$ ที่ทำให้ $(\theta_0 - \delta, \theta_0 + \delta) \subset \Theta$ ผลที่ตามมาคือ เราสามารถนิยามเหตุการณ์

$$S_n = \{\boldsymbol{X} : \mathscr{L}\left(\theta_0|\boldsymbol{X}\right) > \mathscr{L}\left(\theta_0 - \delta|\boldsymbol{X}\right)\} \cup \{\boldsymbol{X} : \mathscr{L}\left(\theta_0|\boldsymbol{X}\right) > \mathscr{L}\left(\theta_0 + \delta|\boldsymbol{X}\right)\}$$

โดยที่ ■

7.4 การประมาณค่าด้วยโมเมนต์ (Method of Moments)

การประมาณค่าด้วยโมเมนต์[1] (method of moments หรือเรียกสั้นๆ ว่า MM) เป็นวิธีประมาณค่าอย่างง่ายที่อาศัยหลักการพื้นฐานที่ว่า ค่าคาดหมายจากตัวอย่าง (sample mean) $\bar{X}_n$ คือตัวประมาณค่า (estimator) ที่ดีสำหรับค่าคาดหมาย (population mean) $E[X]$ ซึ่งเป็นหลักการที่ประยุกต์ใช้อย่างกว้างขวาง แม้แต่ในกรณีของการประมาณค่าที่ซับซ้อน ข้อดีอีกอย่างหนึ่งของการประมาณค่าด้วยโมเมนต์ (MM) ก็คือการที่สามารถหาตัวประมาณค่า (estimator) ได้ค่อนข้างแน่นอน ซึ่งต่างจากวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation หรือเรียกสั้นๆ ว่า MLE) ที่อาจจะไม่สามารถหาตัวประมาณค่าได้ (estimator)

กำหนดให้ $X_1, \ldots, X_n$ เป็นตัวอย่างที่สุ่มมาจากประชากรที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) หรือฟังก์ชันความน่าจะเป็น (p.f.) $f(\boldsymbol{x}|\theta_1, \ldots, \theta_k)$ วิธีการประมาณค่าด้วยโมเมนต์ (MM) ของพารามิเตอร์ $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ เริ่มได้ด้วยการกำหนดให้ k โมเมนต์จากตัวอย่าง (sample moments) มีค่าเท่ากับ k โมเมนต์จากประชากร (population moments) ซึ่งเป็นฟังก์ชันของพารามิเตอร์ $\boldsymbol{\theta}$ ผลลัพธ์จากขั้นตอนนี้คือ ระบบสมการของ k สมการที่มีตัวไม่ทราบค่า (unknown) ทั้งหมด k ตัว ขั้นตอนสุดท้ายคือการแก้ระบบสมการนี้ โดยที่ผลลัพธ์ที่ได้จะอยู่ในรูปของฟังก์ชัน $\theta_j(X_1, \ldots, X_n)$

เพื่อให้ชัดเจนมากยิ่งขึ้น กำหนดให้ $M_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$ แทนโมเมนต์จากตัวอย่าง (sample moments) อันดับที่ $j = 1, \ldots, k$ และ $\mu_j(\theta_1, \ldots, \theta_k) = E[X^j]$ แทนโมเมนต์จากประชากร (population moments) อันดับที่ $j = 1, \ldots, k$ ดังนั้น การประมาณค่าด้วยโมเมนต์ (MM) คือการแก้ระบบสมการต่อไปนี้

$$M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \tag{7.28}$$

$$M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \tag{7.29}$$

$$\vdots \tag{7.30}$$

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \mu_k\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \tag{7.31}$$

นี่คือระบบสมการที่มีทั้งหมด k สมการและมีตัวไม่ทราบค่า $\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right)$ ซึ่งมีทั้งหมด k ตัว

ตัวอย่างที่ 7.12. สมมุติว่า เป็นตัวอย่างสุ่มที่มีมีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) แบบปกติ $N(\mu, \sigma^2)$ นั่นคือ พารามิเตอร์ที่ต้องการทราบค่าในตัวอย่างนี้มีทั้งหมด 2 ตัวคือ $\hat{\theta}_1 = \mu$ $\hat{\theta}_2 = \sigma^2$ ดังนั้น เรา

[1]การประมาณค่าด้วยโมเมนต์ (MM) เป็นวิธีประมาณค่าที่เก่าแก่ ซึ่งน่าจะเริ่มมาจาก Karl Pearson ในช่วงทศวรรษที่ 1800 (Casella and Berger, 2002)

จำเป็นต้องใช้โมเมนต์ที่หนึ่งและโมเมนต์ที่สอง

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu}$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E\left[X^2\right] = \hat{\sigma}^2 + \hat{\mu}^2$$

ตัวประมาณค่าด้วยโมเมนต์ (MM estimators) ในกรณีนี้เท่ากับ

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \bar{X}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

ซึ่งก็คือตัวประมาณค่าสำหรับค่าคาดหมายและค่าความแปรปรวนที่ใช้กันอยู่ทั่วไป อาจจะแตกต่างบ้างเล็กน้อยในกรณีของค่าความแปรปรวนที่หารด้วย n ไม่ใช่ $n-1$ ความแตกต่างตรงนี้มีผลต่อความเบี่ยงเบน (bias) ของตัวประมาณค่าในทางทฤษฎี แต่ในทางปฏิบัติหาก n มีค่ามากพอแล้ว ค่าประมาณที่ได้จะมีค่าใกล้เคียงกันมาก □

อย่างไรก็ตาม วิธีการประมาณค่าด้วยโมเมนต์ (MM) ก็มีจุดอ่อน ทั้งนี้ส่วนหนึ่งเป็นเพราะโมเมนต์เพียงไม่กี่อันอาจจะไม่สามารถบอกคุณสมบัติของการแจกแจงได้ดีพอ เพราะหากต้องการให้ครอบคลุมทั้งหมดต้องใช้โมเมนต์ทุกอัน ตัวอย่างต่อไปนี้เป็นกรณีที่ตัวประมาณค่าด้วยโมเมนต์ (MM) ไม่สมเหตุสมผล

ทฤษฎีบทต่อไปนี้สรุปว่า ตัวประมาณค่าด้วยโมเมนต์ (MM) มีความคงเส้นคงวา (consistent) นั่นคือ เมื่อจำนวนตัวอย่างมากพอแล้ว ค่าประมาณด้วยโมเมนต์ (MM) จะมีค่าใกล้เคียงกับค่าพารามิเตอร์ที่แท้จริง (true parameter)

ทฤษฎีบทที่ 7.8. *สมมุติว่าตัวแปรสุ่ม $X_1, \ldots, X_n$ ที่เป็นอิสระต่อกันและมีการแจกแจงเหมือนกัน (i.i.d.) โดยมี $\boldsymbol{\theta}_0$ เป็นค่าพารามิเตอร์ที่แท้จริง (true parameter) ซึ่งมีทั้งหมด k ตัว และสมมุติว่า k โมเมนต์ของการแจกแจงหาค่าได้และมีค่าจำกัด แล้ว ตัวประมาณค่าด้วยโมเมนต์ (MM) จะมีความคงเส้นคงวา (consistent) นั่นคือ $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$*

การพิสูจน์. กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ทำให้สรุปได้ว่า สำหรับค่าจำนวนเต็ม $0 < j \leqslant k$

$$plim\ M_j = plim\ \frac{1}{n}\sum_{i=1}^{n} X_i^j = E\left[X_i^j\right]$$

ซึ่งมีค่าเท่ากับ $\mu_j\left(\boldsymbol{\theta}_0\right)$ ในขณะเดียวกัน เมื่อเราประยุกต์ใช้การลู่เข้าเชิงความน่าจะเป็นกับระบบสมการที่ (7.28)-(7.31) จะสามารถสรุปได้ว่า

$$plim\ M_j = plim\ \mu_j\left(\hat{\boldsymbol{\theta}}\right) = \mu_j\left(plim\ \hat{\boldsymbol{\theta}}\right)$$

ดังนั้นเราจึงสามารถสรุปได้ว่า

$$\mu_j\left(plim\ \hat{\boldsymbol{\theta}}\right) = \mu_j\left(\boldsymbol{\theta}_0\right) \Rightarrow plim\ \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$$

■

บทที่ 8

การวิเคราะห์แบบกำลังสองน้อยสุด (ordinary Least Square)

บทนี้นำเสนอ

บทที่ 9

การวิเคราะห์ความแปรปรวน (Analysis of Variance)

บทนี้นำเสนอหลักการพื้นฐานของการวิเคราะห์ความแปรปรวน (Analysis of Variance) ซึ่งเป็นเครื่องมือสำคัญในการทดสอบความแตกต่างของค่าคาดหวัง (expected value) ของกลุ่มตัวอย่างแต่ละกลุ่ม ไม่ใช่การทดสอบที่เกี่ยวกับความแปรปรวน (variance) แต่ค่าสถิติที่ใช้ในการวิเคราะห์แบบนี้อยู่ในรูปแบบของความแปรปรวน (variance) ทำให้ได้ชื่อว่าการวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA)

สิ่งหนึ่งที่ผู้อ่านควรจะเข้าใจหลังจากศึกษาบทนี้คือ ค่าสถิติที่ใช้ในการทดสอบ ANOVA เป็นเพียงรูปแบบหนึ่งที่เป็นไปได้ แต่เหตุผลที่คนส่วนใหญ่นิยมใช้รูปแบบดังกล่าวก็เป็นเพราะว่า ค่าสถิติดังกล่าวมีการกระจายตัวที่แน่ชัด นั่นคือ มีการกระจายตัวแบบ F (non-central F)

9.1 สมมุติฐานสำหรับการวิเคราะห์ความแปรปรวน (Hypothesis for ANOVA)

Bibliography

Attanasio, O., Meghir, C., and Nix, E. (2015). Human capital development and parental investment in india. Technical report, National Bureau of Economic Research.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637--654.

Casella, G. and Berger, R. L. (2002). *Statistical inference*. Duxbury Pacific Grove, CA.

Cox, J. C., Ross, S. A., Rubinstein, M., et al. (1979). Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229--263.

De Finetti, B. (1974). *Theory of probability: a critical introductory treatment*, volume 1. John Wiley & Sons.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and statistics*. Pearson Education.

Doob, J. L. (1949). Application of the theory of martingales. *Le calcul des probabilites et ses applications*, pages 23--27.

Fitzpatrick, P. M. (2005). *Advanced Calculus*. Cole Publishing Company.

Hogg, R. V., McKean, J., and Craig, A. T. (2005). *Introduction to mathematical statistics*. Pearson Education.

Hull, J. (2010). *Options, Futures, and Other Derivatives, 7/e (With CD)*. Pearson Education India.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141--183.

Prakasa Rao, B. (1987). *Asymptotic theory of statistical inference*. John Wiley & Sons, Inc.

Index