พื้นฐานการจำลองสถานการณ์เชิงสุ่ม

เพื่อการประยุกต์ใช้กับปัญหาจริง

ผศ. ดร. จุฑา พิชิตลำเค็ญ
ภาควิชาวิศวกรรมอุตสาหการ
คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเกษตรศาสตร์

สำนักพิมพ์แห่งมหาวิทยาลัยเกษตรศาสตร์

คำชี้แจง

ผู้เขียนรวบรวมตำราฉบับนี้ เพื่อใช้ประกอบการสอนวิชาการจำลองสถานการณ์ (Simulation modeling) ในระดับปริญญาตรีและโท ตำรานี้กล่าวถึงประเด็นที่ผู้สร้างแบบจำลองควรคำนึงถึง และทฤษฎีที่เกี่ยวข้องในขั้นตอนต่าง ๆ ของการพัฒนาแบบจำลอง มิได้มุ่งสอนวิธีการใช้ซอฟท์แวร์แบบใดแบบหนึ่ง หรือวิธีการพัฒนาแบบจำลองบนโปรแกรมคอมพิวเตอร์โดยเฉพาะ

ตำราเล่มนี้เหมาะสำหรับผู้อ่านที่มีพื้นฐานแคลคูลัส สถิติ และความน่าจะเป็นในระดับที่เป็นวิชาบังคับในหลักสูตรวิศวกรรมศาสตร์และวิทยาศาสตร์ประยุกต์ทั่วไป โดยบทที่ 2 ทบทวนความรู้พื้นฐานด้านความน่าจะเป็นและสถิติที่ใช้บ่อยๆ

ผู้เขียนได้โครงสร้างหลักๆ จากเอกสารคำสอนของ Dr. Barry L. Nelson (Nelson, 2003) ซึ่งเป็นอาจารย์ที่ปรึกษาระดับปริญญาเอกของผู้เขียนที่ Northwestern University เมือง Evanston มลรัฐอิลลินอยส์ ประเทศสหรัฐ อเมริกา นอกจากนี้หนังสือเรียนที่ใช้อ้างอิงหลักๆ คือ Discrete-Event System Simulation (Banks et al., 2005) และ Simulation Modeling and Analysis (Law, 2006) ซึ่งถือว่าเป็นหนังสือประกอบการสอนวิชาจำลองสถานการณ์ที่รู้จักกันดี ส่วนคู่มือการจำลองสถานการณ์ที่ใช้เป็นแหล่งค้นคว้าและอ้างอิงได้เช่น Banks (1998) (ฐานข้อมูลออนไลน์ของห้องสมุดมหาวิทยาลัยเกษตรศาสตร์มี e-book แยกเป็นบท) และ Chung (2004) ซึ่งเป็น e-book ฟรี

การแปลคำศัพท์เทคนิคภาษาอังกฤษให้เป็นภาษาไทย ผู้เขียนอ้างอิงจากเว็บไซต์ศัพท์บัญญัติของราชบัณฑิตยสถาน (The Royal Institute of Thailand, 2555) ใช้โปรแกรม R (R Development Core Team, 2008) และ Microsoft Visio เพื่อสร้างรูปภาพ

ตำราการจำลองสถานการณ์ที่เขียนเป็นภาษาไทยเล่มอื่นๆ ที่ไม่เน้นสอนการใช้ซอฟต์แวร์สำเร็จรูป ที่ผู้เขียนทราบ เรียงตามลำดับปีที่พิมพ์ มีดังนี้: สงวน ตั้งโพธิธรรม (2538), ศิริจันทร์ ทองประเสริฐ (2540), วุฒิชัย วงษ์ทัศนีย์กร (2555), และ เสกสรร เกียรติสุไพบูลย์ (2555)

หากผู้อ่านพบข้อผิดพลาดในหนังสือหรือมีข้อแนะนำ กรุณาแจ้งผู้เขียนที่ `juta.p@ku.ac.th` จักเป็นพระคุณยิ่ง

ผศ. ดร. จุฑา พิชิตลำเค็ญ
มหาวิทยาลัยเกษตรศาสตร์

จากใจ

ผู้เขียนขออุทิศหนังสือเล่มนี้ให้กับมารดาและบิดา ผู้ให้อิสระและความไว้วางใจกับลูก ๆ เสมอมา และเป็นแบบอย่างของการทำงานอย่างตั้งใจและจริงจัง

ขอขอบคุณภาควิชาวิศวกรรมอุตสาหการ, คณะวิศวกรรมศาสตร์, และมหาวิทยาลัยเกษตรศาสตร์ที่อนุญาตให้ผู้เขียนได้ลาเพื่อเพิ่มพูนความรู้ทางวิชาการในปีการศึกษา 2555 ตำราฉบับนี้อาจจะไม่เกิดขึ้นหากผู้เขียนไม่ได้รับโอกาสนี้

ขอขอบคุณน้อง ๆ และบิดาของข้าพเจ้าที่ให้โอกาสทำในสิ่งที่อยากทำ: ได้เดินทางและได้ปฏิบัติธรรม

I am also indebted to Barry L. Nelson, my beloved former PhD advisor, who taught me academically, professionally and athletically.

กราบขอบพระคุณพระอาจารย์ของผู้เขียน พระสุทธิศาสตร์ ปญฺญาปทีโป หรือพระอาจารย์โน้ส แห่งวัดป่าสุคะโต (จังหวัดชัยภูมิ) สำหรับการรับฟังและกำลังใจ ทั้งทางโลกย์และทางธรรม ให้ธรรมะและความสงบเย็นที่ทำให้หนังสือเล่มนี้เสร็จสมบูรณ์ได้

ขอขอบคุณผศ. นุชรินทร์ ทิพยวรรณากร (ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ) ดร. วนัฐฌพงษ์ คงแก้ว (ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์) อ.ดร. สุวิชภรณ์ วิชกูล (ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์) นิสิตปริญญาโท ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ รุ่นรหัส 56xx ผู้จัดการและบรรณาธิการผู้ทรงคุณวุฒิแห่งสำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์ ที่ช่วยอ่านบทร่างของหนังสือเล่มนี้อย่างละเอียด และเสนอแนะจุดบกพร่อง

จุฑา พิชิตลำเค็ญ

สารบัญ

1

บทนำ

คอมพิวเตอร์ส่วนตัว (Personal computers) สมาร์ทโฟน (Smart phones) และไอแพด (iPad) หรือเทเบลดพีซี (Tablet PCs) ได้กลายเป็นส่วนสำคัญในการดำรงชีวิตของผู้คน พร้อมกันนี้ได้นำการจำลองสถานการณ์ (Simulation) เข้าสู่ชีวิตเราด้วย หนึ่งในตัวอย่างที่เด่นชัดที่สุดคือเกมคอมพิวเตอร์ ซึ่งคือการสร้างโลกเสมือนจริง (Virtual world) หรือการจำลองสถานการณ์นั่นเอง โดยมีผู้เล่นเกมเป็นผู้ตอบสนองต่อสถานการณ์ในเกม ทั้งนี้ เพื่อให้เกมสนุกสนานและแปลกใหม่ สถานการณ์ที่นำเสนอจึง "ไม่เหมือนเดิม" ในแต่ละครั้งที่เล่น ในแง่วิชาการ เรียกได้ว่าเกมสร้างผลลัพธ์สุ่ม (Random outcomes)

นอกจากการเล่นเกมเพื่อความเพลิดเพลินแล้ว การเล่นเกมบางอย่างมีจุดประสงค์เพื่อการเรียนรู้หรือการฝึกอบรม เช่น การฝึกนักบินโดยใช้การจำลองห้องควบคุมเครื่องบิน (Flight simulator) ซึ่งประหยัดค่าใช้จ่ายและปลอดภัยกว่าการอบรมบนเครื่องบินจริง กระทรวงกลาโหมของสหรัฐอเมริกา (Department of Defense) ใช้การจำลองสถานการณ์อย่างกว้างขวางในการฝึกอบรม, การทดสอบอาวุธยุทโธปกรณ์, และการประเมินผลได้ผลเสียของทางเลือกต่าง ๆ (อ่านบทความล่าสุดเกี่ยวกับการประยุกต์ใช้ได้จากวารสารออนไลน์ Training and Simulation Journal: `http://www.tsjonline.com/`)

การจำลองสถานการณ์โดดเด่นในแง่เป็นวิธีที่ใช้ศึกษาระบบที่ซับซ้อนซึ่งมีความไม่แน่นอน (Uncertainty) หรือความสุ่ม (Randomness) ในระบบ จึงได้รับความนิยมอย่างแพร่หลายในกลุ่มนักวิจัยดำเนินงาน (Operations research practitioners) ซึ่งจัดให้การจำลองสถานการณ์เป็นหนึ่งในเทคนิคที่ถูกใช้มากที่สุด การจำลองสถานการณ์เป็นสหวิทยาการที่อาศัยหลักสถิติ, ความน่าจะเป็น, ทฤษฎีตัวเลข, และวิทยาศาสตร์ด้านคอมพิวเตอร์ (Haas, 2006) หัวข้อแรกจะกล่าวถึงคำจำกัดความของการจำลองสถานการณ์

1.1 คำจำกัดความของการจำลองสถานการณ์

การจำลองสถานการณ์มีความหมายต่างกันในสาขาวิชาต่าง ๆ ในตำราเล่มนี้ การจำลองสถานการณ์เป็นวิธีวิเคราะห์การปฏิบัติการ (Performance) ของระบบซึ่งมีพฤติกรรมที่ขึ้นอยู่กับปฏิสัมพันธ์ (Interaction) ของกระบวนการสุ่ม (Random process) หลาย ๆ ตัว กระบวนการสุ่มเหล่านี้สามารถระบุได้ด้วยแบบจำลองความน่าจะเป็น (Nelson, 2013) เช่น ดัชนีชี้วัดอาจเป็นเวลาในการจัดส่งสินค้าจากศูนย์กระจายสินค้าไปยังร้านค้า ซึ่งขึ้นอยู่กับอุปสงค์ (Demand) ของลูกค้าและการส่งมอบสินค้าจากโรงงาน อาจสมมติให้อุปสงค์ของลูกค้าแต่ละรายเป็นอิสระต่อกันและมีการแจกแจงแบบปัวซอง และเวลาในการส่งมอบ (Lead time) มีการแจกแจงแบบปกติ

ระบบที่เราสนใจนี้อาจมีอยู่จริงแล้ว เช่น กรณีที่เราต้องการปรับปรุงระบบที่มีอยู่ให้"ดีขึ้น" (ต้องกำหนดเกณฑ์ในการวัด"ความดี"นี้) หรือเป็นสถานการณ์ที่ยังไม่มีอยู่จริงแต่กำลังพิจารณาอยู่ เช่น การเปลี่ยนผังการจัดเก็บสินค้าในคลังโดยจะใช้ชั้นวางแทนการวางบนแพเล็ต (Pallet) อย่างที่ใช้อยู่ มักใช้การจำลองสถานการณ์ในกรณีที่ไม่สามารถคำนวณดัชนีชี้วัด (Performance measure) ตรง ๆ ได้ หรือในกรณีที่การ

ประมาณค่าด้วยวิธีเชิงตัวเลข (Numerical method) ก็ยังไม่สามารถกำหนดขอบเขตของความผิดพลาดในการประมาณค่า (Estimation error) ได้ (Nelson, 2013) หัวข้อต่อไปจะกล่าวถึงประวัติศาสตร์สั้น ๆ ของการจำลองสถานการณ์

1.2 ความเป็นมาของการจำลองสถานการณ์

การจำลองสถานการณ์บนคอมพิวเตอร์ได้รับการพัฒนาควบคู่ไปกับการเติบโตอย่างรวดเร็วของคอมพิวเตอร์ทั้งในด้านฮาร์ดแวร์และซอฟต์แวร์ สืบเนื่องมาจากการประยุกต์ใช้ขนานใหญ่เป็นครั้งแรกสำหรับโครงการแมนฮัตตัน (Manhattan Project) ในช่วงสงครามโลกครั้งที่สอง เพื่อจำลองการระเบิดแบบนิวเคลียร์ด้วยวิธีมอนติคาร์โล (Wikipedia, 2012a) ช่วงปี พ.ศ. 2503-2508 ซอฟต์แวร์ได้รับการพัฒนาเพื่อสร้างแบบจำลองสถานการณ์โดยเฉพาะ เช่น บริษัทไอบีเอ็มมีภาษา GPSS (**G**eneral **P**urpose **S**ystem **S**imulator) นอกจากนี้ยังมี SIMSCRIPT ที่พัฒนาโดย Harry Markowitz จากสถาบัน RAND

เดิมแบบจำลองสถานการณ์ต้องประมวลผลบนเครื่องคอมพิวเตอร์เมนเฟรม (Mainframe computer) ซึ่งจำกัดให้การใช้งานอยู่เฉพาะในองค์กรขนาดใหญ่ที่สามารถมีเครื่องเหล่านี้ได้ แต่บุคคลทั่วไปเข้าไม่ถึง การจำลองสถานการณ์เริ่มมีการใช้อย่างแพร่หลายมากขึ้นในช่วงปี พ.ศ. 2509-2513 เพราะมีซอฟต์แวร์เพื่อจำลองสถานการณ์สำหรับเครื่องคอมพิวเตอร์ส่วนตัว (Nance, 1993) นอกจากนี้คอมพิวเตอร์ยังมีราคาถูกลงและมีกำลังการประมวลผลและความจุของการเก็บข้อมูลเพิ่มขึ้นอย่างรวดเร็ว ปัจจัยเหล่านี้ส่งเสริมให้มีการประยุกต์ใช้การจำลองสถานการณ์มากขึ้นเรื่อย ๆ ดังจะเห็นได้จากความหลากหลายของสาขาวิชาที่มีการประยุกต์ใช้การจำลองสถานการณ์

1.3 สาขาวิชาที่มีการประยุกต์ใช้การจำลองสถานการณ์

สาขาวิชามากมายอาศัยการจำลองสถานการณ์เป็นเครื่องมือวิเคราะห์ ผู้สนใจสามารถสืบค้นงานวิจัยและการประยุกต์ใช้ล่าสุดได้จากเอกสารประกอบการประชุมวิชาการของ The Winter Simulation Conference (`www.wintersim.org`) ซึ่งจัดเป็นประจำในช่วงต้นเดือนธันวาคมของทุกปี

เว็บไซต์ของสมาคมการจำลองสถานการณ์ (Simulation Society) เก็บรวบรวมบทความประกอบการประชุมวิชาการย้อนหลังตั้งแต่ปี พ.ศ. 2511 ที่ `http://informs-sim.org/` สมาคมนี้สังกัดสถาบันการวิจัยการดำเนินงานและวิทยาการจัดการ (Institute for Operations Research and the Management Sciences: INFORMS) ในปี พ.ศ. 2555 หัวข้อด้านการประยุกต์ใช้มีดังนี้

- การผลิต (Manufacturing)
- การดูแลสุขภาพ (Health care)

- โลจิสติกส์และการขนส่ง (Logistics and transportation)
- สังคมศาสตร์และองค์กร (Social science and organization)
- การทหาร (Military applications)
- วิทยาศาสตร์สิ่งแวดล้อม (Environmental sciences)
- การก่อสร้างและการบริหารโครงการ (Construction and project management)

เพื่อช่วยให้ผู้อ่านเข้าใจวิธีการและการลำดับการเกิดขึ้นของเหตุการณ์ในแบบจำลอง ผู้เขียนจึงนำเสนอแบบจำลองอย่างง่ายที่สามารถประมวลผลด้วยมือได้

1.4 การจำลองสถานการณ์ด้วยมือ: ระบบแถวคอยอย่างง่าย

พิจารณาแถวคอยของลูกค้าที่รอชำระเงินในซุปเปอร์มาร์เก็ต โดยที่มีเคาน์เตอร์คิดเงินเพียง 1 จุด (Banks et al., 2005) ตารางที่ 1.1 แสดงข้อมูลนำเข้าที่ได้จากข้อมูลในอดีต ดังนี้

- **เวลาระหว่างการมาถึง** (Interarrival time) ของลูกค้า สมมติให้ลูกค้าคนแรกมาถึงระบบที่เวลา 0 เช่น ลูกค้าคนที่สองมาต่อคิวหลังจากลูกค้าแรก 2 นาที
- **เวลาที่ใช้ในการบริการลูกค้าแต่ละคน** (เวลาคิดเงิน) เช่น พนักงานใช้เวลา 2 นาทีเพื่อคิดเงินลูกค้าคนแรก และ 1 นาทีสำหรับลูกค้าคนที่สอง

ตารางที่ 1.1: เวลาระหว่างการมาถึงและเวลาที่ใช้คิดเงินลูกค้าของระบบแถวคอยอย่างง่าย

ลูกค้าคนที่	เวลาระหว่างการมาถึง (นาที)	เวลาที่ใช้คิดเงินลูกค้า (นาที)
1	-	2
2	2	1
3	4	3
4	1	2
5	2	1
6	6	4

สมมตินาฬิกา (Clock) ในแบบจำลองเริ่มที่เวลาศูนย์ และเมื่อพนักงานคิดเงินพร้อมที่จะทำงานและว่าง ตารางที่ 1.2 เป็นผลการจำลอง ข้อมูลเวลาระหว่างการมาถึงถูกนำไปใช้คำนวณเวลาตามนาฬิกาที่ลูกค้าเข้าสู่ระบบ กล่าวคือ ลูกค้าคนที่สองเข้ามาต่อคิวที่เวลา 0+2=2 ส่วนคนที่สามมาถึงที่เวลา 2+4=6 สังเกตว่าลูกค้าคนแรกได้รับการบริการทันทีที่มาถึง และสิ้นสุดการบริการที่เวลาที่เริ่มคิดเงินบวกด้วยเวลารับบริการ ลูกค้า

คนที่สี่ต้องรอให้ลูกค้าคนที่สามคิดเงินเสร็จก่อน จึงจะเริ่มได้รับการบริการ ดังนั้นเวลาที่เริ่มรับบริการจึงเป็นค่าที่มากกว่าระหว่างเวลาที่ตนเองมาถึง กับเวลาที่คนก่อนหน้าเขาเสร็จสิ้นการรับบริการ หรืออาจใช้ฟังก์ชัน $\max(\cdot,\cdot)$ เวลารอคอยในแถวคิดจากส่วนต่างระหว่างเวลาที่เริ่มได้รับบริการและเวลาที่มาถึงระบบ เวลาในระบบของลูกค้าคือเวลาที่คิดเงินเสร็จหักลบกับเวลาที่มาถึงระบบ

ตารางที่ 1.2: ตารางการจำลองสถานการณ์สำหรับระบบแถวคอยอย่างง่าย (ข้อมูลนำเข้าในตารางที่ 1.1)

ลูกค้าคนที่	**เวลาที่เข้ามาต่อคิว (ตามนาฬิกา)**	**เวลาที่เริ่มคิดเงิน**	**เวลาที่ใช้คิดเงิน (นาที)**	**เวลาที่คิดเงินเสร็จ (ตามนาฬิกา)**	**เวลารอคอยในแถว (นาที)**	**เวลาในระบบ (นาที)**
1	0	0	2	0+2 = 2	0	2-0 = 2
2	0+2 = 2	2	1	2+1 = 3	2-2 = 0	1
3	2+4 = 6	6	3	6+3 = 9	0	3
4	6+1 = 7	max(7, 9) = 9	2	9+2 = 11	9-7 = 2	11-7 = 4
5	7+2 = 9	11	1	11+1 = 12	2	3
6	9+6 = 15	15	4	15+4 = 19	0	4

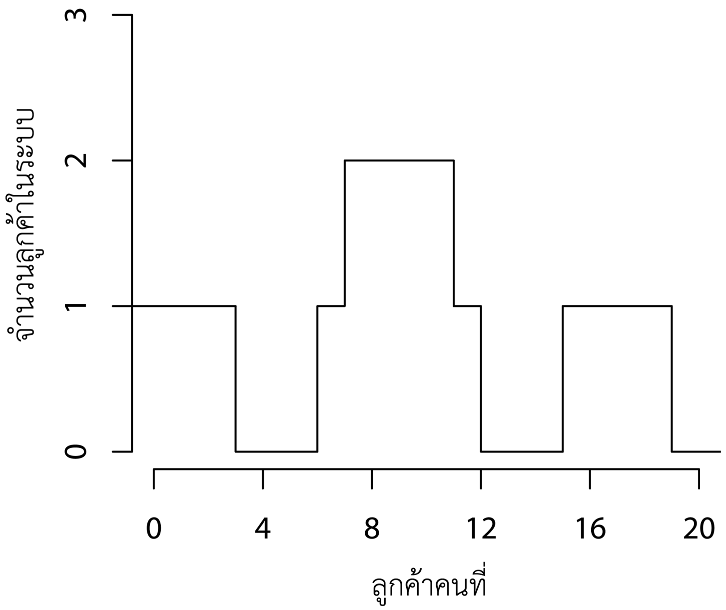
เวลารอคอยในแถวและเวลาในระบบของลูกค้าแต่ละคนที่จำลองได้ ถูกนำมาใช้เป็นค่าประมาณค่าเฉลี่ยของเวลารอคอยในแถว ($\widehat{W}_q$) และเวลาเฉลี่ยในระบบ ($\widehat{W}$) ดังนี้ (สัญลักษณ์ $\widehat{\ }$ เพื่อแสดงว่าเป็นค่าประมาณ)

$$\begin{aligned}\widehat{W}_q &= \frac{(0\times 4)+(2\times 2)}{6} = 2/3 \text{ นาที} \\ \widehat{W} &= \frac{2+1+3+4+3+4}{6} = 17/6 \text{ นาที}\end{aligned}$$

หัวข้อที่ 5.3 จะกล่าวถึงการคำนวณค่าความแปรปรวนและช่วงความเชื่อมั่น นอกจากนี้ ยังสามารถสร้างกราฟจำนวนลูกค้าที่มีในระบบที่เวลาต่าง ๆ (รูปที่ 1.1) และนำกราฟที่ได้มาประมาณจำนวนลูกค้าเฉลี่ยในระบบ ณ ขณะเวลาใดเวลาหนึ่ง

$$\begin{aligned}\widehat{L} &= \frac{\text{พื้นที่ใต้กราฟ}}{\text{เวลาทั้งหมด}} \\ &= \frac{[3\times 1]+[(12-6)\times 1]+[(11-7)\times 1]+[(19-15)\times 1]}{19} = 17/19\end{aligned}$$

การตีความ $\widehat{L}$ คือ หากถ่ายรูประบบนี้หลาย ๆ ครั้ง, นับจำนวนลูกค้าในรูป, แล้วเฉลี่ยจำนวนลูกค้าในรูป จะได้ประมาณ 17/19 คน



รูปที่ 1.1: จำนวนลูกค้าในระบบที่เวลาต่าง ๆ สำหรับระบบแถวคอยอย่างง่าย

การจำลองสถานการณ์เป็นเครื่องมือที่สามารถรองรับความต้องการได้หลากหลายก็จริง แต่ก็เป็นเทคนิคที่มีทั้งข้อเด่นและกรณีพิเศษที่การจำลองสถานการณ์เป็นเครื่องมือที่ไม่เหมาะสม

1.5 ข้อได้เปรียบของเทคนิคการจำลองสถานการณ์

การจำลองสถานการณ์มีจุดเด่นหลายประการที่ส่งผลให้มีการประยุกต์ใช้อย่างแพร่หลายในปัจจุบัน (Banks et al., 2005) อาทิ

- สามารถจำลองระบบที่ซับซ้อน ช่วยให้สามารถศึกษาและพิจารณาความสัมพันธ์ระหว่างระบบย่อยที่อยู่ภายใต้ระบบใหญ่ที่ซับซ้อนได้ เช่น ปฏิสัมพันธ์ระหว่างอุปสงค์และอุปทานของระบบการให้บริการทางสาธารณสุขและสังคมที่ประเทศอังกฤษ (Brailsford et al., 2011)

- สามารถจำลองผลกระทบที่เกิดขึ้นจากการเปลี่ยนแปลงเชิงข้อมูล, เชิงองค์กร, หรือเชิงสิ่งแวดล้อม เช่น การจำลองคลินิกกระดูกของโรงพยาบาลขนาดใหญ่ของรัฐแห่งหนึ่งในประเทศไทย เพื่อใช้ศึกษาแนวทางปรับปรุงการบริการ เพื่อลดเวลาการรอคอยของผู้ป่วย (Weerawat et al., 2013)

- ก่อให้เกิดความรู้ ความเข้าใจในระหว่างการพัฒนาแบบจำลอง ซึ่งอาจนำไปสู่การปรับปรุงระบบ และความเข้าใจที่เพิ่มขึ้น เช่น การจำลองห้องปฏิบัติการตรวจเลือดเพื่อใช้ปรับปรุงเวลาในการส่งผล (Turnaround time) นำไปสู่ข้อเสนอการปรับเปลี่ยนขั้นตอนการทำงานและผัง (Layout) ของห้องปฏิบัติการ (Luangmul et al., 2012)
- เป็นประโยชน์สำหรับการวิเคราะห์แบบตั้งคำถามโดยใช้คำว่า "ถ้า..." (What-if analysis) ซึ่งเป็นการเปลี่ยนค่าพารามิเตอร์นำเข้า และสังเกตผลกระทบที่ตามมาต่อค่าเอาท์พุท (Output) จากแบบจำลอง หรือการวิเคราะห์ความไว (Sensitivity analysis)
- สามารถใช้ทดสอบการออกแบบใหม่ ๆ ก่อนนำไปใช้จริง เพื่อเตรียมรับมือกับเหตุการณ์ที่อาจเกิดขึ้น เช่น การจำลองแผนกตรวจสุขภาพ เพื่อใช้ทดสอบวิธีลดเวลารอคอยของคนไข้ในรูปแบบต่าง ๆ เช่น การจัดระบบนัดให้คนไข้ (Wongsammacheep et al., 2012)
- สามารถนำไปสู่การกำหนดข้อบังคับ (Requirements) ของระบบ หากมีการจำลองระบบนั้น ๆ ที่ขีดความสามารถ (Capability) ต่าง ๆ เช่น การใช้แบบจำลองเพื่อศึกษาผลกระทบจากการเพิ่มหัวจ่ายก๊าซที่คลังจ่าย ในแง่ของเวลารอคอยของรถบรรทุกที่ลดลงและสัดส่วนของเวลาที่หัวจ่ายก๊าซได้ใช้ประโยชน์
- สามารถใช้ตรวจสอบความถูกต้องของคำตอบเชิงวิเคราะห์ (Analytical solutions) ในกรณีที่คำตอบมีรูปแบบสมการที่ซับซ้อน
- สามารถใช้เพื่อการฝึกอบรม เช่น การฝึกนักบินหรือฝึกพนักงานควบคุมโรงงานผลิตกระแสไฟฟ้าพลังงานนิวเคลียร์

ถึงแม้ว่ามีข้อดีหลายประการ การจำลองสถานการณ์เป็นเพียงเครื่องมือสำหรับวิเคราะห์ชนิดหนึ่ง จึงมีกรณีที่เหมาะจะนำไปใช้ และในบางกรณี มีเครื่องมืออื่นเหมาะสมกว่า

1.6 กรณีที่การจำลองสถานการณ์เป็นเครื่องมือที่ไม่เหมาะสม

กรณีเหล่านี้เป็นกรณีที่ผู้ใช้ควรพิจารณาเลือกใช้เครื่องมืออื่นในการวิเคราะห์ (Banks et al., 2005)

- เมื่อปัญหาสามารถแก้ได้ง่าย ๆ โดยใช้สามัญสำนึก เช่น สำหรับระบบให้บริการลูกค้าทางโทรศัพท์ (Call center) ที่ลูกค้าโทรเข้ามาด้วยอัตรา 100 สายต่อชั่วโมง และพนักงานแต่ละคนสามารถให้บริการด้วยอัตรา 12 สายต่อชั่วโมง จำนวนพนักงานที่ต่ำที่สุดที่ต้องมี คือ 100/12 = 8.33 หรือ 9 คน
- เมื่อปัญหาสามารถแก้ด้วยวิธีเชิงวิเคราะห์ (Analytically) เช่น ระบบแถวคอยอย่างง่ายบางอย่างอาจถูกจำลองด้วยแบบจำลองแถวคอยแบบมาร์คอฟ (Markovian queueing models)

- เมื่อการทดลองด้วยสถานการณ์จริงมีค่าใช้จ่ายต่ำกว่าการจำลองสถานการณ์ด้วยคอมพิวเตอร์ เช่น การจำลองการให้บริการของร้านอาหารฟาสต์ฟู้ดด้วยพนักงานจริง
- เมื่อค่าใช้จ่ายสูงกว่าความประหยัด เช่น ค่าว่าจ้างที่ปรึกษาสูงกว่าความประหยัดที่คาดว่าจะเกิดขึ้น
- เมื่อไม่มีเวลาหรือทรัพยากร เช่น ต้องตัดสินใจพรุ่งนี้ แต่แบบจำลองต้องใช้เวลาหลายสัปดาห์ในการพัฒนา
- เมื่อไม่มีข้อมูลหรือกระทั่งค่าประมาณสำหรับค่าพารามิเตอร์นำเข้าต่าง ๆ
- เมื่อผู้ว่าจ้างหรือเจ้าของโครงการมีความคาดหวังที่เกินจริง
- เมื่อพฤติกรรมของระบบมีความซับซ้อนมาก ๆ เช่น พฤติกรรมของมนุษย์

เมื่อเลือกที่จะใช้เทคนิคการจำลองสถานการณ์แล้ว มีบางประเด็นที่ผู้วิเคราะห์ควรพิจารณาก่อนเริ่มงาน

1.7 ประเด็นคำถามหลักที่ควรพิจารณาในการจำลองสถานการณ์

ผู้วิเคราะห์ควรคำนึงถึงประเด็นต่อไปนี้ (Haas, 2006)

- **จะจำลองระบบอย่างไร** เช่น จำลองระบบการผลิตเฉพาะโมเดลที่มียอดสั่งซื้อสูงสุด หรือที่ทุก ๆ สถานการณ์ที่อาจเป็นได้
- **สามารถประมาณดัชนีชี้วัดที่สนใจจากค่าที่ประมวลผลได้จากแบบจำลองหรือไม่** เช่น ถ้าจะประมาณสัดส่วนของการสั่งซื้อที่ส่งตรงเวลา แบบจำลองต้องมีตัวแปรตัวหนึ่งเพื่อเก็บตัวเลขจำนวนการสั่งซื้อทั้งหมด และอีกตัวหนึ่งเพื่อเก็บจำนวนการสั่งซื้อที่ส่งตรงเวลา
- **นิยามดัชนีชี้วัดที่สนใจอย่างชัดเจนแล้วหรือไม่** เช่น ความพึงพอใจของลูกค้าของศูนย์โทรศัพท์ อาจนิยามในแง่เวลาที่ลูกค้ารอโดยเฉลี่ยจนกว่าจะได้พูดกับพนักงาน หรือสัดส่วนของลูกค้าที่รอนานไม่เกินเป้าที่กำหนด (เช่น 20 วินาที)
- **ดัชนีที่จะวัดในแต่ละรอบทำซ้ำ (Replication) คืออะไร และจะคำนวณด้วยสูตรใด** เช่น ค่าใช้จ่ายในการจัดเก็บและส่งสินค้าต่อชิ้น ควรประกอบด้วยค่าใช้จ่ายใดบ้าง
- **นิยามรอบทำซ้ำบนคอมพิวเตอร์อย่างไร** เช่น สภาวะของระบบที่เวลาศูนย์และสภาวะที่เวลาสิ้นสุดรอบ

- **จำนวนรอบทำซ้ำ (Number of replications) ที่ควรใช้และความยาวของแต่ละรอบควรเป็นเท่าใด** สองปัจจัยนี้ส่งผลต่อค่าความผิดพลาดของค่าประมาณ

- **ประมาณความผิดพลาดในการประมาณค่า (Estimation error) ได้อย่างไร**

- **สามารถเพิ่มประสิทธิภาพการวิเคราะห์ได้หรือไม่** เช่น ในกรณีที่ใช้แบบจำลองเพื่อเลือกทางเลือกที่ดีที่สุดจากหลาย ๆ ตัวเลือก การออกแบบการทดลองที่เหมาะสมจะช่วยกำหนดจำนวนรอบทำซ้ำที่จำเป็นต้องใช้

- **จะใช้ค่าประมาณของดัชนีชี้วัดที่ได้จากแบบจำลองเพื่อการตัดสินใจอย่างไร** เช่น การเลือกวิธีการจัดคนไข้เข้าแผนกตรวจร่างกาย ควรพิจารณาเวลารอเฉลี่ยของคนไข้ และสัดส่วนที่ทำงานจริง (Utilization) ของแพทย์หรือทรัพยากรอื่น ๆ ด้วย เพราะถ้าพิจารณาเฉพาะเวลารอคอยของคนไข้เพียงอย่างเดียว จำนวนแพทย์ที่ใช้อาจสูงเกินกว่าสามารถปฏิบัติจริงได้

เมื่อตอบคำถามเบื้องต้นได้แล้ว โดยมากขั้นตอนในการศึกษาจะคล้ายคลึงกัน

1.8 ขั้นตอนในการศึกษาด้วยการจำลองสถานการณ์

ในภาพรวม การวิเคราะห์ด้วยแบบจำลองสถานการณ์จะดำเนินไปตามขั้นตอนที่จะกล่าวถึงต่อไปนี้ (Banks et al., 2005) อาจไม่เรียงกันตามลำดับก่อนหลังอย่างชัดเจน บางขั้นตอนอาจทำไปได้พร้อมกันหรือบางขั้นอาจต้องกลับมาทำใหม่ เช่น การเก็บข้อมูลตรวจวัด (Data), การพัฒนาแบบจำลองคอมพิวเตอร์, และการตรวจสอบเหมาะสมของแบบจำลอง เมื่อพัฒนาแบบจำลองไปแล้วอาจพบว่าแบบจำลองยังไม่สมจริง เช่น ในแบบจำลองสถานีจ่ายก๊าซ LPG สังเกตพบว่าเวลาในคลังก๊าซที่ได้จากแบบจำลองแตกต่างจากค่าจริงที่เก็บได้มาก จึงทบทวนข้อสมมติ (Assumption) พบว่าเวลาในการเติมก๊าซขึ้นอยู่กับขนาดความจุของรถ มิใช่มีค่าเฉลี่ยเดียวกันอย่างที่สมมติมาก่อนหน้า จึงต้องไปเก็บข้อมูลความสัมพันธ์ระหว่างเวลาในการเติมก๊าซและความจุของรถเพิ่มเติม ขั้นตอนการวิเคราะห์มีดังนี้

1. **กำหนดปัญหาที่จะศึกษา** ให้แน่ใจว่าผู้พัฒนาตัวแบบและผู้ใช้ตัวแบบเพื่อการตัดสินใจ (Decision maker) มีความเห็นตรงกัน

2. **กำหนดวัตถุประสงค์และแผนโครงการในภาพรวม** เช่น ตัวเลือกอื่น ๆ นอกจากระบบปัจจุบันที่พิจารณา, เกณฑ์การเลือกหรือดัชนีชี้วัด, และรายละเอียดเชิงปฏิบัติการของการศึกษา เช่น จำนวนคนที่จะทำโครงการ, ค่าใช้จ่าย, ตารางเวลา, และผลที่คาดหวังในแต่ละช่วงเวลา

3. **วางแนวคิดของตัวแบบ** (Model conceptualization) ถึงแม้ว่าการวางแนวคิดอาจขึ้นกับผู้พัฒนาแบบจำลอง แต่ก็สามารถปรับให้ดีขึ้นได้โดยการดึงเอาคุณลักษณะที่สำคัญ ๆ ของระบบจริงออกมาเลือกข้อสมมติหลัก ๆ ที่จำเป็น และปรับเปลี่ยนจนกระทั่งแบบจำลองที่ได้สามารถจำลองสถานการณ์จริงได้ ดังนั้นจึงควรเริ่มจากแบบจำลองที่เรียบง่ายและไม่ใหญ่เกินไป แล้วค่อยปรับเพิ่มรายละเอียด ทั้งนี้ระดับรายละเอียด (Fidelity) ของแบบจำลองขึ้นอยู่กับคำถามที่ต้องการจะตอบด้วย ไม่จำเป็นต้องสร้างแบบจำลองที่เหมือนระบบจริงทุกประการ เช่น ถ้าต้องการออกแบบระบบให้สามารถรองรับช่วงการใช้งานสูง ไม่จำเป็นต้องจำลองระบบในช่วงเวลาอื่น ๆ

4. **เก็บข้อมูลตรวจวัด** รายละเอียดของแบบจำลองและวัตถุประสงค์ของการศึกษาเป็นตัวกำหนดชนิดของข้อมูลตรวจวัดที่ต้องใช้ เนื่องจากการเก็บข้อมูลใช้เวลามาก จึงควรเริ่มให้เร็วที่สุดเท่าที่จะเป็นได้ ทั้งนี้ต้องพิจารณาถึงคุณภาพของข้อมูลด้วยว่าสอดคล้องกันหรือไม่ เช่น หากเวลานาฬิกา (Clock time) ที่ลูกค้าออกจากระบบน้อยกว่าเวลาที่ลูกค้าเข้าระบบแสดงว่าข้อมูลนั้นผิดพลาด หรือมีลูกค้าเข้ามาในเวลาที่ระบบปิด ดังนั้น จึงควรคัดกรองข้อมูลตรวจวัดก่อนนำไปใช้

5. **พัฒนาตัวแบบบนคอมพิวเตอร์** ในปัจจุบันมักใช้ซอฟต์แวร์สำเร็จรูป เช่น Arena (Rockwell Software) Promodel (Promodel Corp.) หรือ Simio (Simio LLC) สำหรับการจำลองสถานการณ์เชิงสุ่มแบบเหตุการณ์ไม่ต่อเนื่อง (Discrete-event stochastic simulation) แต่หากผู้วิเคราะห์มีเงินจำกัดก็สามารถใช้ฟรีแวร์ที่เขียนด้วยภาษาจาวา เช่น SSJ (L'Ecuyer, 2012) หรือภาษาซีพลัสพลัส เช่น SIMLIB (Peringer et al., 2011) ซอฟต์แวร์ทางการค้าสำหรับการจำลองสถานการณ์บนสเปรดชีทที่รู้จักกันแพร่หลาย มีดังนี้: @Risk (Palisade Corp.), CrystalBall (Oracle), หรือฟรีแวร์ เช่น PopTools (Hood, 2010a)

6. **ตรวจสอบความถูกต้องของตัวแบบคอมพิวเตอร์** (Verification) หรือ Debugging เป็นการตรวจสอบว่าพัฒนาโปรแกรมคอมพิวเตอร์ตรงตามที่คิดไว้หรือไม่

7. **ตรวจสอบความเหมาะสมของตัวแบบจำลองในการเลียนแบบระบบจริง** (Validation) เรียกอีกอย่างหนึ่งว่า การตั้งค่า (Calibration)

8. **ออกแบบการทดลอง** (Experimental design) การออกแบบการทดลองช่วยกำหนดจำนวนรอบทำซ้ำที่ต้องประมวลผลจากแบบจำลองของแต่ละตัวเลือก เช่น ระหว่างระบบควบคุมการผลิตแบบผลักและดึง เพื่อให้สามารถเลือกตัวเลือกที่ดีที่สุดโดยใช้จำนวนรอบทำซ้ำที่สมเหตุสมผล และมีการประกันความน่าจะเป็นที่จะเลือกพลาด

9. **ประมวลผลแบบจำลองและวิเคราะห์** เนื่องจากค่าที่ประมวลผลได้จากแบบจำลอง (Simulation observations) เป็นค่าสุ่ม จึงต้องอาศัยสถิติในการวิเคราะห์

10. **ตัดสินใจว่าจำเป็นต้องประมวลผลเพิ่มเติมหรือไม่** ในกรณีที่ค่าความผิดพลาดของดัชนีชี้วัดที่เป็นเอาท์พุทสูง อาจต้องเพิ่มขนาดตัวอย่าง ซึ่งคือการประมวลผลเพิ่ม

11. **ทำคู่มือการใช้และรายงานผลการศึกษา** ทั้งนี้ อาจมีการรายงานผลเป็นระยะ ๆ ระหว่างการศึกษาแทนที่จะรอส่งรายงานขั้นสุดท้ายเพียงฉบับเดียว

12. **นำผลการศึกษาไปใช้งานจริง** (Implementation) การนำไปใช้จริงที่ประสบความสำเร็จขึ้นอยู่กับการมีส่วนร่วมของผู้ใช้งานแบบจำลองหรือผู้มีอำนาจตัดสินใจในระหว่างการศึกษา เพราะถ้าผู้ใช้งานมีส่วนร่วมในการพัฒนา เขาจะมีความเข้าใจและความเชื่อมั่นในแบบจำลองและผลที่ได้ ทำให้มั่นใจในการนำไปใช้จริง

ตำราฉบับนี้จะกล่าวถึงทฤษฎีที่เป็นระเบียบวิธีวิจัย (Methodology) ในการจำลองสถานการณ์ โดยจะเริ่มจากการทบทวนพื้นฐานสถิติและความน่าจะเป็นในบทต่อไป

แบบฝึกหัด

1.1. ให้เลือกบทความจากเอกสารประกอบการประชุมของ The Winter Simulation Conference (`http://informs-sim.org/`) 1 ฉบับ แล้วระบุว่าสิ่งใดที่เป็นตัวแปรนำเข้าที่ไม่แน่นอนและสิ่งใดเป็นดัชนีชี้วัดในบทความนั้น

1.2. ให้พิจารณาตัวอย่างระบบแถวคอยอย่างง่ายในหัวข้อ 1.4 แต่ให้ใช้เวลาที่ลูกค้ามาถึง (เวลาตามนาฬิกา) และเวลาที่ให้บริการดังตารางที่ 1.3 กำหนดให้เวลานาฬิกาเริ่มที่ศูนย์

   (i) สร้างตารางจำลองสถานการณ์ เลียนแบบตารางที่ 1.2

   (ii) ให้คำนวณค่าเฉลี่ยเวลาที่ลูกค้ารอในแถว ($\widehat{W}_q$) และเวลาเฉลี่ยที่ลูกค้ารอในระบบ ($\widehat{W}$)

   (iii) สร้างกราฟจำนวนลูกค้าที่เคาน์เตอร์กับเวลา (เลียนแบบรูปที่ 1.1) และใช้กราฟเพื่อหาจำนวนลูกค้าเฉลี่ยในระบบ ($\widehat{L}$)

1.3. ที่สหกรณ์ออมทรัพย์มีลูกค้า 6 คนมาใช้บริการในชั่วโมงแรกที่เปิดทำการ มีพนักงานที่หน้าเคาน์เตอร์ 1 คน เวลาระหว่างการมาถึง (Interarrival time) ในหน่วยนาที มีดังนี้

   0   1.1   4.4*   3.2*   21.8   2.3

   ลูกค้าแต่ละคนใช้เวลารับบริการ (ไม่รวมเวลารอ) คงที่คือ 10 นาทีเท่ากัน เวลาระหว่างการมาถึงที่มีสัญลักษณ์ * เป็นลูกค้าพิเศษที่มีลำดับการให้บริการก่อนลูกค้าที่ไม่มี * ลูกค้าประเภทเดียวกันได้รับการ

ตารางที่ 1.3: เวลาที่ลูกค้ามาถึงและเวลาที่ใช้คิดเงินลูกค้า

ลูกค้าคนที่	เวลาที่ลูกค้ามาถึงตามนาฬิกา (นาที)	เวลาที่ใช้คิดเงินลูกค้า (นาที)
1	0	4
2	8	1
3	14	4
4	15	3
5	23	2
6	34	5
7	41	4
8	43	5
9	46	3

บริการแบบมาก่อนได้ก่อน ลูกค้าได้รับบริการต่อเนื่องรวดเดียวจนเสร็จ ไม่มีการสับเปลี่ยนลูกค้า ให้สร้างตารางจำลองสถานการณ์แล้วตอบคำถามดังต่อไปนี้

(i) มีลูกค้ากี่คนที่ไม่ต้องรอเพื่อรับการบริการ (จากทั้งหมด 6 คน) คือมาถึงแล้วได้รับบริการทันที

(ii) ลูกค้าคนที่ 6 ออกจากธนาคารที่เวลาเท่าใด

(iii) เวลาในระบบเฉลี่ยของลูกค้า 6 คนนี้เป็นเท่าใด

(iv) จำนวนลูกค้าในธนาคารเฉลี่ยในช่วง 30 นาทีแรก และ 60 นาทีแรกเป็นเท่าใด (คำแนะนำ: ให้สร้างกราฟคล้ายรูปที่ 1.1 แล้วคำนวณพื้นที่ใต้กราฟ)

2

บททบทวนความน่าจะเป็นและสถิติ

เนื่องจากระบบที่สนใจจำลองเป็นระบบที่มีความไม่แน่นอนอยู่ การจำลองสถานการณ์จึงต้องอาศัยพื้นฐานทฤษฎีความน่าจะเป็นเพื่อจำลองความไม่แน่นอนนี้ ในแง่ดังต่อไปนี้

- **เพื่อกำหนดรายละเอียดของตัวแปรนำเข้าที่มีความแปรปรวน** เช่น เวลาในการให้บริการลูกค้าแต่ละคน หรือชนิดของการบริการที่ลูกค้าต้องการ เช่น สอบถามเกี่ยวกับผลิตภัณฑ์ ก.หรือเกี่ยวกับผลิตภัณฑ์ ข.

- **เพื่อวิเคราะห์ค่าเอาท์พุทแบบจำลองประมวลผลได้** เนื่องจากค่าเอาท์พุท (เช่น เวลาที่ลูกค้าคอยในแถว) มีความไม่แน่นอนอยู่ จึงไม่สามารถรายงานเฉพาะค่าเฉลี่ย แต่จะต้องมีค่าบ่งชี้ความแปรปรวนด้วย

บทนี้มีวัตถุประสงค์เพื่อทบทวนความน่าจะเป็นและสถิติที่เกี่ยวข้องกับการจำลองสถานการณ์

2.1 ความน่าจะเป็นและความไม่แน่นอน

ความน่าจะเป็นให้ค่าเป็นตัวเลขหรือมาตราส่วน (Scale) สำหรับโอกาสที่เหตุการณ์ใดเหตุการณ์หนึ่งจะปรากฏ เช่น ฝนจะตกพรุ่งนี้ หรือเครื่องจักรจะติดขัดในอีก 1 ชั่วโมงการใช้งานข้างหน้า ด้วยค่าตั้งแต่ 0 ถึง 1 (หรือ 100%) ความน่าจะเป็นเท่ากับหนึ่งหมายความว่าเหตุการณ์นั้น ๆ จะเกิดขึ้นแน่นอน เช่น ทุกคนจะต้องตายด้วยความน่าจะเป็นหนึ่ง ส่วนความน่าจะเป็นศูนย์แปลว่าเหตุการณ์นั้น ๆ จะไม่เกิดขึ้นแน่นอน เช่น แก้วที่แตกแล้วไม่มีรอยร้าว ส่วนความน่าจะเป็นที่มีค่าอยู่ระหว่างศูนย์ถึงหนึ่ง เช่น 0.1 มีความหมายว่า เมื่อพิจารณาเหตุการณ์นี้หลาย ๆ ครั้งภายใต้สถานการณ์เดียวกัน เหตุการณ์นั้น ๆ จะเกิดขึ้น 10% ของจำนวนครั้งทั้งหมด เช่น มีรายงานว่า 5%-10% ของหญิงที่เป็นมะเร็งเต้านมมีสาเหตุจากกรรมพันธุ์ที่ยีนถ่ายทอดจากแม่สู่ลูก (Breastcancer.org, 2012) นั่นคือหญิงทั่วไปที่มีมารดาเป็นมะเร็งเต้านมจะมีโอกาสเป็นมะเร็งเต้านมเช่นกันประมาณ 5%-10% ซึ่งค่าความน่าจะเป็นนี้ประมาณด้วยการสุ่มตัวอย่าง (Sampling) หรือการวิเคราะห์ข้อมูลในอดีต สมมติให้ A เป็นเหตุการณ์ที่หญิงที่เป็นมะเร็งเต้านมมีมารดาที่เป็นมะเร็งเต้านม ให้ n เป็นจำนวนหญิงที่เป็นมะเร็งเต้านมในกลุ่มตัวอย่างและในจำนวนนี้มี x คนที่มีมารดาเป็นมะเร็งเต้านม จึงสามารถประมาณความน่าจะเป็นที่ลูกสาวของหญิงที่เป็นมะเร็งเต้านมจะเป็นมะเร็งด้วยสัดส่วน

$$\widehat{\Pr\{A\}} = \frac{x}{n}$$

สังเกตว่าความน่าจะเป็นสัมพันธ์กับเหตุการณ์ หัวข้อต่อไปจะกล่าวถึงคุณสมบัติพื้นฐานของความน่าจะเป็น

2.2 คุณสมบัติพื้นฐานของความน่าจะเป็น

สมมติให้ A และ B เป็นเหตุการณ์ที่เราพิจารณา ความน่าจะเป็นของเหตุการณ์ A คือ $\Pr\{A\}$ และความน่าจะเป็นของเหตุการณ์ B คือ $\Pr\{B\}$ เช่น A เป็นเหตุการณ์ที่ประชาชนตรวจสุขภาพทุกปี และ B เป็นเหตุการณ์ที่ประชาชนอายุอย่างน้อย 50 ปี **ปริภูมิตัวอย่าง** (Sample space; Φ) คือเซตของผลของเหตุการณ์ทั้งหมด เช่น ปริภูมิตัวอย่างของ A อาจเป็นประชากรไม่เคยตรวจสุขภาพ, ตรวจทุกปี, ตรวจทุกสองปี, $\ldots$, หรือด้วยความถี่ไม่แน่นอน สามารถสร้างเหตุการณ์ใหม่จาก A และ B ได้ดังนี้

- **เหตุการณ์ที่ไม่ใช่** A: แทนด้วยสัญลักษณ์ A' หรือ A^c (A complement) คือเหตุการณ์ที่ประชาชนไม่ตรวจสุขภาพทุกปี คือ อาจจะไม่เคยตรวจเลย หรือหลาย ๆ ปีตรวจครั้งหนึ่ง สามารถคำนวณ $\Pr\{A^c\}$ จาก $\Pr\{A\}$ ดังนี้
$$\Pr\{A^c\} = 1 - \Pr\{A\} \tag{2.1}$$
- **เหตุการณ์** A ***และ*** B: แสดงด้วยสัญลักษณ์ $A \cap B$ คือเหตุการณ์ที่ประชาชนที่อายุอย่างน้อย 50 ปีตรวจสุขภาพทุกปี
- **เหตุการณ์** A ***หรือ*** B: แสดงด้วยสัญลักษณ์ $A \cup B$ คือเหตุการณ์ที่ประชาชนมีอายุอย่างน้อย 50 ปีหรือเป็นคนที่ตรวจสุขภาพทุกปี มองอีกนัยหนึ่งคือเป็นเหตุการณ์ที่ประชาชนคนใดคนหนึ่งที่สุ่มเลือกมามีอายุอย่างน้อย 50 ปีหรือเป็นคนที่ตรวจสุขภาพทุกปี ซึ่งสามารถคำนวณได้จาก $\Pr\{A\}, \Pr\{B\}$ และ $\Pr\{A \cap B\}$ ดังนี้
$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \tag{2.2}$$
- **เหตุการณ์** A **ที่ไม่อยู่ในเหตุการณ์** B: แสดงด้วยสัญลักษณ์ $A - B$ คือเหตุการณ์ที่ประชาชนที่ตรวจสุขภาพทุกปีที่มีอายุน้อยกว่า 50 ปี ซึ่งสามารถคำนวณได้ดังนี้
$$\Pr\{A - B\} = \Pr\{A\} - \Pr\{A \cap B\}$$
- **กฎของความน่าจะเป็นรวม (Law of total probability)**: หากเหตุการณ์ A ประกอบด้วยเหตุการณ์ย่อยอื่น ๆ ที่ไม่ทับซ้อนกัน เช่น A เป็นเหตุการณ์ที่คนไข้ตรวจสุขภาพประจำปีที่โรงพยาบาลแห่งหนึ่งประกอบด้วยคนไข้ภายใต้ชุดตรวจสุขภาพประจำปีชุดต่าง ๆ กัน สมมติว่ามี n ชุดตรวจ (Package) โดย A_i เป็นเหตุการณ์ที่คนไข้ซื้อชุดตรวจที่ i สำหรับ $i = 1, 2, \ldots, n$ สมมติว่าคนไข้ไม่ตรวจสุขภาพสองชุดในปีเดียวกัน ($A_i \cap A_j = \varnothing$ เมื่อ $i \neq j$ และ $\varnothing$ แทนเซ็ตว่าง) จึงเขียนได้ว่า
$$\begin{aligned} A &= A_1 \cup A_2 \cup \ldots \cup A_n \\ \Pr\{A\} &= \Pr\{A_1\} + \Pr\{A_2\} + \ldots + \Pr\{A_n\} \end{aligned} \tag{2.3}$$

ตัวอย่างที่ 2.1 แสดงวิธีการใช้คุณสมบัติเหล่านี้เพื่อคำนวณความน่าจะเป็นที่สนใจ

ตัวอย่างที่ 2.1. สมมติให้ $\Pr\{A\} = 0.55, \Pr\{B^c\} = 0.35$ และ $\Pr\{A \cup B\} = 0.75$ สามารถคำนวณความน่าจะเป็นดังต่อไปนี้ได้

- $\Pr\{B\} = 1 - \Pr\{B^c\} = 1 - 0.35 = 0.65$ ด้วยสมการที่ (2.1)
- $\Pr\{A \cap B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cup B\} = 0.55 + 0.65 - 0.75 = 0.45$ ด้วยสมการที่ (2.2)

□

ความเป็นอิสระต่อกันเป็นข้อสมมติที่สำคัญมาก ที่เป็นรากฐานในการวิเคราะห์ทางสถิติขั้นพื้นฐาน

2.2.1 ความเป็นอิสระต่อกัน

ความเป็นอิสระต่อกัน (Independence) เป็นการประมาณ (Approximation) เชิงทฤษฎีเพื่อให้การวิเคราะห์และการคำนวณเป็นไปได้สะดวกขึ้น (ปกติการพิสูจน์ว่าสองเหตุการณ์เป็นอิสระต่อกันอย่างแท้จริงทำได้ยากมาก เพราะถึงแม้ว่าความเชื่อมโยงของสองเหตุการณ์จะมีน้อยมาก ความน่าจะเป็นก็อาจไม่เป็นศูนย์) เมื่อกล่าวว่าเหตุการณ์ A เป็นอิสระต่อเหตุการณ์ B หมายความว่า ความรู้หรือข้อมูลที่มีเกี่ยวกับเหตุการณ์ A ไม่เปลี่ยนความรู้หรือข้อมูลที่เรามีต่อเหตุการณ์ B เช่น หากสมมติว่าแผ่นดินไหวแต่ละครั้งเป็นอิสระต่อกัน เมื่อสัปดาห์ที่แล้วเกิดแผ่นดินไหวหรือไม่ก็ตาม ความน่าจะเป็นที่จะเกิดแผ่นดินไหวในสัปดาห์นี้ก็คงเดิม แต่ในความเป็นจริงแล้ว นักปฐพีวิทยาทราบว่าแผ่นดินไหวมักจะเกิดไล่เลี่ยกันเป็นชุด

เมื่อเหตุการณ์ A และ B เป็นอิสระต่อกัน คุณสมบัติต่อไปนี้จะเป็นจริง

$$\Pr\{A \cap B\} = \Pr\{A\} \times \Pr\{B\} \tag{2.4}$$

ตัวอย่างที่ 2.2 แสดงวิธีนำคุณสมบัติความเป็นอิสระต่อกันเพื่อคำนวณ

ตัวอย่างที่ 2.2. การพยากรณ์อากาศของกรมอุตุนิยมวิทยาของประเทศจีนทำนายได้ถูกต้อง (เหตุการณ์ A) 82% ส่วนการพยากรณ์อากาศของกรมอุตุฯ ของไทยทำนายได้ถูกต้อง (เหตุการณ์ B) 65% ความน่าจะเป็นที่การพยากรณ์อากาศที่ประเทศใดประเทศหนึ่งหรือทั้งสองประเทศจะถูกต้อง คือ

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}$$

ด้วยสมการที่ (2.2) แต่โจทย์ไม่ได้ระบุ $\Pr\{A \cap B\}$ จึงสมมติเพิ่มเติมว่า A และ B เป็นอิสระต่อกัน จากคุณสมบัติ (2.4) ทำให้

$$\begin{aligned}\Pr\{A \cup B\} &= \Pr\{A\} + \Pr\{B\} - (\Pr\{A\} \times \Pr\{B\}) \\ &= 0.82 + 0.65 - (0.82 \times 0.65) = 0.937\end{aligned}$$

ความน่าจะเป็นที่การพยากรณ์อากาศที่ประเทศใดประเทศหนึ่งหรือทั้งสองประเทศจะถูกต้อง คือ 93.7% □

ในเชิงปฏิบัติ ความน่าจะเป็นที่ประมาณได้จากข้อมูลจริงเป็นความน่าจะเป็นแบบมีเงื่อนไข ซึ่งนำมาใช้คำนวณความน่าจะเป็นที่ต้องการ

2.2.2 ความน่าจะเป็นแบบมีเงื่อนไข

โดยมากข้อมูลตรวจวัดที่เก็บได้หรือสังเกตได้จริงเป็น**ความน่าจะเป็นแบบมีเงื่อนไข** (Conditional Probability) เช่น ถ้าตรวจพบว่าติดเชื้อโรคเอดส์ โอกาสที่จะเป็นโรคเอดส์จริง ๆ ประมาณ $x\%$, สินค้าที่ส่งมาจากซัพพลายเออร์ ก. มีของเสีย 1%, หรือเที่ยวบินของสายการบินไทยถึงที่หมายตรงเวลา 80% (FlightStats, Inc., 2012) ดังนั้นความน่าจะเป็นแบบมีเงื่อนไขนำไปสู่ค่าความน่าจะเป็นของเหตุการณ์ที่สนใจจริง ๆ ความน่าจะเป็นแบบมีเงื่อนไขมีลักษณะ ”ถ้า B เป็นจริง โอกาสที่ A จะเป็นจริงคือ...” สามารถคำนวณได้ดังนี้

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \tag{2.5}$$

เช่น สำหรับตัวอย่างการตรวจโรคเอดส์ เหตุการณ์เงื่อนไข (B) คือผลตรวจระบุว่ามีเชื้อเอดส์ ส่วนเหตุการณ์ที่สนใจ (A) คือติดเชื้อจริง ๆ ในตัวอย่างของเสียจากซัพพลายเออร์ เหตุการณ์เงื่อนไข (B) คือเป็นสินค้าจากซัพพลายเออร์ ก. และ A คือ เป็นของเสีย ในตัวอย่างเที่ยวบิน เหตุการณ์เงื่อนไข (B) คือเป็นเที่ยวบินของการบินไทย และเหตุการณ์ที่เราสนใจ (A) คือเป็นเที่ยวบินที่ตรงเวลา

กฎของความน่าจะเป็นรวม (สมการที่ 2.3) สามารถเขียนในรูปของความน่าจะเป็นแบบมีเงื่อนไข สมมติให้ปริภูมิตัวอย่าง Φ ประกอบด้วย $A_i, i = 1, 2, \ldots, n$ ที่ไม่ทับซ้อนกัน ($A_i \cap A_j = \varnothing$ สำหรับ $i \neq j$ และ $A_1 \cup A_2 \cup \ldots \cup A_n = \Phi$) สามารถคำนวณความน่าจะเป็นของ B จากส่วนต่าง ๆ ของ B ที่อยู่ใน A_i แต่ละอันได้ เพราะ $(B \cap A_i) \cap (B \cap A_j) = \varnothing$ ดังนี้

$$\Pr\{B\} = \Pr\{B \cap A_1\} + \Pr\{B \cap A_2\} + \ldots + \Pr\{B \cap A_n\} \tag{2.6}$$

$$= \Pr\{B|A_1\}\Pr\{A_1\} + \Pr\{B|A_2\}\Pr\{A_2\} + \ldots + \Pr\{B|A_n\}\Pr\{A_n\} \tag{2.7}$$

สมการที่ (2.6) กลายเป็นสมการที่ (2.7) ด้วยคุณสมบัติ (2.5) ตัวอย่างที่ 2.3 แสดงวิธีใช้คุณสมบัติความน่าจะเป็นแบบมีเงื่อนไข

ตัวอย่างที่ 2.3. ถ้ามีเพลิงไหม้ (เหตุการณ์ B) อยู่ในพื้นที่ ๆ หนึ่งในอาคาร เครื่องตรวจควันไฟตรวจจับได้และส่งสัญญาณเตือน (เหตุการณ์ A) ด้วยความน่าจะเป็น 0.99; ถ้าไม่มีเพลิง เครื่องตรวจนี้ส่งสัญญาณเตือนภัยที่ผิด (False alarm) ด้วยความน่าจะเป็น 0.1 สมมติว่าความน่าจะเป็นที่มีเพลิงไหม้ในพื้นที่นี้คือ 0.05

โจทย์ให้ข้อมูลว่า $\Pr\{A|B\} = 0.99, \Pr\{A|B'\} = 0.1$ และ $\Pr\{B\} = 0.05$

- ถ้ามีเพลิงไหม้ ความน่าจะเป็นที่เครื่องตรวจจับควันจะไม่ส่งสัญญาณ เขียนด้วยสัญลักษณ์ได้ดังนี้

$$\Pr\{A'|B\} = 1 - \Pr\{A|B\} = 1 - 0.99 = 0.01$$

- ความน่าจะเป็นที่ไม่มีเพลิงไหม้และเครื่องตรวจส่งสัญญาณผิด คือ

$$\Pr\{A \cap B'\} = \Pr\{A|B'\}\Pr\{B'\} = \Pr\{A|B'\}(1 - \Pr\{B\}) = 0.1 \times (1 - 0.05) = 0.095$$

- ความน่าจะเป็นที่มีเพลิงไหม้และเครื่องตรวจไม่ส่งสัญญาณ คือ

$$\Pr\{A' \cap B\} = \Pr\{A'|B\}\Pr\{B\} = 0.01 \times 0.05 = 5 \times 10^{-4}$$

- ความน่าจะเป็นที่เครื่องตรวจทำงานผิดพลาด

$$\Pr\{A' \cap B\} + \Pr\{A \cap B'\} = 0.0005 + 0.095 = 0.0955$$

□

ตัวแปรสุ่มเป็นฟังก์ชันที่แทนค่าที่ไม่แน่นอน ค่าความไม่แน่นอนนี้แจกแจงด้วยการแจกแจงความน่าจะเป็น

2.3 ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น

การแจกแจงความน่าจะเป็น (Probability distribution) กระจายค่าความน่าจะเป็นที่เหตุการณ์ใด ๆ ที่พิจารณามีผล (Outcome) ต่าง ๆ ดังนั้นเมื่อต้องการจำลองสถานการณ์ที่ไม่แน่นอนด้วยการแจกแจงความน่าจะเป็น คำถามที่ควรตอบมีดังนี้

- อะไรเป็นผลที่เป็นไปได้
- ผลเป็นแบบต่อเนื่อง (Continuous) หรือแบบไม่ต่อเนื่อง เรียกอีกชื่อหนึ่งว่า **วิยุต** (Discrete)
- ผลมีขอบเขต (Bounded) หรือไม่มีขอบเขต (Unbounded)

ตัวแปรสุ่ม มีค่าเป็นจำนวนจริง (Real number) มักแสดงตัวแปรสุ่มด้วยตัวอักษรโรมันใหญ่ เช่น X หรือ Y และแสดงค่าจริงที่เกิดขึ้นแล้วด้วยตัวอักษรเล็ก เช่น x หรือ y ปริภูมิตัวอย่างของการทดลองหรือเหตุการณ์ที่พิจารณาเป็นเซตของผลที่เป็นไปได้ทั้งหมด เขียนแทนด้วย Φ_X และ Φ_Y ตามลำดับ

มักแบ่งตัวแปรสุ่มเป็นสองประเภท ตามลักษณะของค่าที่เป็นได้ คือ แบบไม่ต่อเนื่อง (หัวข้อ 2.3.1) และแบบต่อเนื่อง (หัวข้อ 2.3.2)

2.3.1 ตัวแปรสุ่มแบบไม่ต่อเนื่อง

ตัวแปรสุ่มแบบไม่ต่อเนื่อง (Discrete random variable) เป็นตัวแปรสุ่มที่มีค่าที่เป็นไปได้แบบไม่ต่อเนื่อง กำหนดให้ตัวแปรสุ่มแบบไม่ต่อเนื่อง X มีความน่าจะเป็นที่แทนด้วยสัญลักษณ์ ดังนี้

$$p(a) = \Pr\{X = a\}$$

เรียกว่า **ฟังก์ชันมวลความน่าจะเป็น** (Probability mass function, PMF) ตัวแปรสุ่มแบบไม่ต่อเนื่องมีผลจำนวนจำกัดหรือไม่จำกัดแบบนับได้ (Countable finite) เช่น ผ่านหรือไม่ผ่าน ($\Phi = \{0, 1\}$); จำนวนชิ้นงานที่เสียจากล็อตขนาด 20 ชิ้น ($\Phi = \{0, 1, 2, \ldots, 20\}$); หรือจำนวนลูกค้าที่รออยู่ในแถวคอย ($\Phi = \{0, 1, 2, \ldots\}$)

สังเกตว่า $p(x)$ หรือ PMF มีคุณสมบัติดังนี้

- $0 \leqslant p(x) \leqslant 1$ สำหรับทุกค่า x กล่าวคือ มวลความน่าจะเป็นไม่ติดลบและไม่เกินหนึ่ง
- $\sum_{x \in \Phi_X} p(x) = 1$ คือ ผลรวมของความน่าจะเป็นของทุก ๆ เหตุการณ์เท่ากับหนึ่ง

การแจกแจงความน่าจะเป็นยังสามารถแสดงในรูป**ฟังก์ชันมวลความน่าจะเป็นสะสม** (Cumulative mass function, CMF) ซึ่งมีนิยาม ดังนี้

$$F_X(a) = \Pr\{X \leqslant a\} \tag{2.8}$$

ตัวอย่างที่ 2.4 แสดงการคำนวณค่ามวลความน่าจะเป็นสะสม

ตัวอย่างที่ 2.4. สมมติให้ปริภูมิตัวอย่างของ X เป็น $\Phi_X = \{0, 1, 2, 3\}$ นั่นคือ $X \in \{0, 1, 2, 3\}$ ดังนี้

$$p(0) = 1/8, p(1) = 3/8, p(2) = 3/8, p(3) = 1/8$$

รูปที่ 2.1ก เป็นกราฟค่า PMF ซึ่งนำมาคำนวณ CMF ได้ ดังตัวอย่างต่อไปนี้

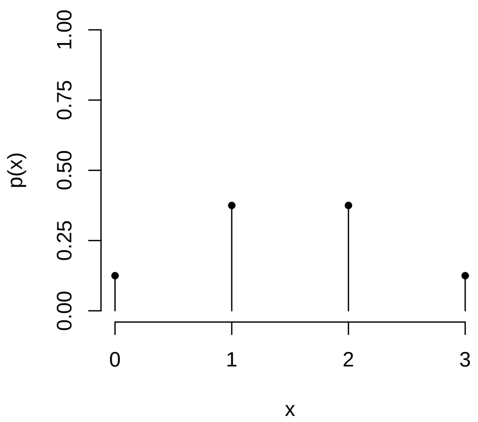
$$F_X(2) = \Pr\{X \leqslant 2\} = \Pr\{X = 0\} + \Pr\{X = 1\} + \Pr\{X = 2\} = 1/8 + 3/8 + 3/8 = 7/8$$

รูปที่ 2.1ข เป็นกราฟค่า CMF เนื่องจากผลบวกของความน่าจะเป็นของทุกเหตุการณ์รวมกันได้หนึ่ง จึงสามารถคำนวณค่าความน่าจะเป็นของเหตุการณ์ที่ต้องการโดยการหักความน่าจะเป็นของเหตุการณ์ที่เรา*ไม่ต้องการ*ออกจาก 1 (ซึ่งก็คือ A complement หรือ A^c) เช่น

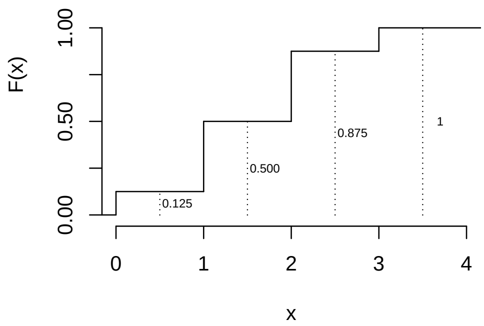
$$\begin{aligned} \Pr\{X > 2\} &= 1 - \Pr\{X \leqslant 2\} = 1 - F(2) = 1 - (7/8) = 1/8 \\ \Pr\{X \neq 0\} &= 1 - \Pr\{X = 0\} = 1 - (1/8) = 7/8 \end{aligned}$$

□

เมื่อค่าที่เป็นไปได้มีค่าต่อเนื่อง ตัวแปรสุ่มนั้นจะเป็นตัวแปรสุ่มแบบต่อเนื่อง ซึ่งจะกล่าวถึงในหัวข้อถัดไป



(ก) ฟังก์ชันมวลความน่าจะเป็น



(ข) ฟังก์ชันมวลความน่าจะเป็นสะสม

รูปที่ 2.1: กราฟมวลความน่าจะเป็นสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง (ตัวอย่างที่ 2.4)

2.3.2 ตัวแปรสุ่มแบบต่อเนื่อง

ตัวแปรสุ่มแบบต่อเนื่อง (Continuous random variable) เป็นตัวแปรสุ่มที่มีค่าต่อเนื่อง คือมีค่าเป็นช่วง (Interval) หรือกลุ่มของช่วง (Banks et al., 2005) ดังนั้นหากต้องการคำนวณค่าความน่าจะเป็นที่ตัวแปรสุ่มแบบต่อเนื่อง Y จะมีค่าอยู่ในช่วง a ถึง b จึงต้องหาปริพันธ์ (Integrate)

$$\Pr\{a \leqslant Y \leqslant b\} = \int_a^b f(y)dy$$

โดยเรียก $f(y)$ ว่า**ฟังก์ชันความหนาแน่นของความน่าจะเป็น** (Probability density function หรือ PDF)

ตัวอย่างที่ 2.5. สมมติให้ Y แสดงอายุการใช้งานของเครื่องมือตรวจสอบชิ้นหนึ่ง ดังนั้น $Y \geqslant 0$ สมมติให้ Y มีการแจกแจงดังนี้

$$f(y) = \begin{cases} \frac{1}{2}e^{-y/2}, & y \geqslant 0 \\ 0, & y \text{ อื่น ๆ} \end{cases}$$

รูปที่ 2.2 แสดงกราฟของการแจกแจง ความน่าจะเป็นที่อายุการใช้งานของเครื่องมือนี้จะไม่เกิน 2 ปีคำนวณได้ดังนี้

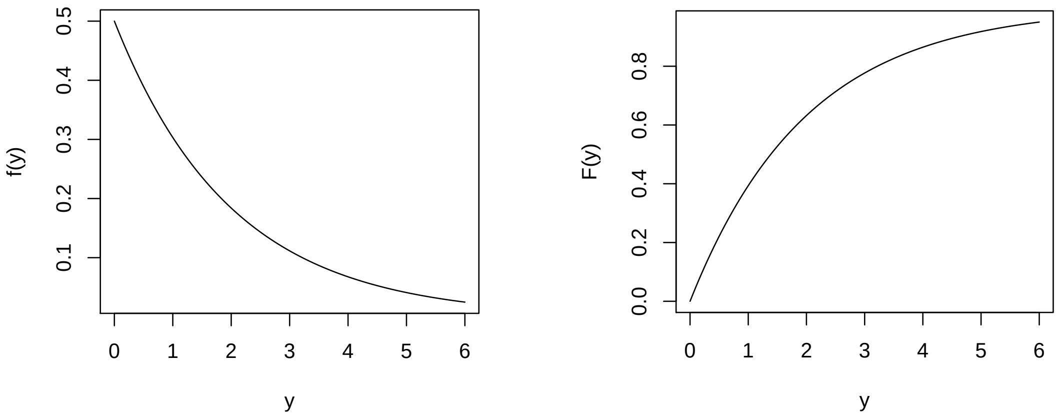
$$\begin{aligned} \Pr\{Y \leqslant 2\} &= \int_0^2 \frac{1}{2}e^{-y/2}dy \\ &= -e^{-2/2} + e^0 = 0.632 \end{aligned}$$

จากนิยามของความน่าจะเป็นสะสมในสมการที่ (2.8) CDF ของตัวแปรสุ่มแบบต่อเนื่องมีนิยามดังนี้

$$F_Y(b) = \int_{-\infty}^b f(y)dy$$

เช่น $F_Y(2) = \int_{-\infty}^{2} f(y)dy = \int_{-\infty}^{0}(0)dy + \int_{0}^{2} \frac{1}{2}e^{-y/2}dy = 0.632$

□



(ก) ฟังก์ชันความหนาแน่นความน่าจะเป็น (ข) ฟังก์ชันความหนาแน่นความน่าจะเป็นสะสม

รูปที่ 2.2: กราฟความหนาแน่นความน่าจะเป็นสำหรับตัวแปรสุ่มแบบต่อเนื่อง (ตัวอย่างที่ 2.5)

ฟังก์ชั่นความหนาแน่นความน่าจะเป็นเป็นฟังก์ชั่นที่มีคุณสมบัติเฉพาะ สมมติให้ค่าที่เป็นไปได้ของ Y อยู่ในช่วง Φ_Y นั่นคือ $Y \in \Phi_Y$ สังเกตว่าฟังก์ชันความหนาแน่นของความน่าจะเป็น $f(y)$ หรือ PDF มีคุณสมบัติดังนี้

- $f(y) \geqslant 0$ สำหรับทุกค่า $y \in \Phi_Y$ และ $f(y) = 0$ สำหรับ $Y \notin \Phi_Y$
- $\int_{-\infty}^{\infty} f(y)dy = \int_{y\in\Phi_Y} f(y)dy = 1$ คือ ผลรวมของความน่าจะเป็นของทุก ๆ เหตุการณ์คือหนึ่ง
- $\Pr\{Y = a\} = 0$ สำหรับทุกค่า a เพราะ $\int_{a}^{a} f(y)dy = 0$ พื้นที่ใต้กราฟของจุดใด ๆ เป็นศูนย์
- ด้วยคุณสมบัติข้อที่แล้ว จึงทำให้

$$\Pr\{a \leqslant Y \leqslant b\} = \Pr\{a < Y \leqslant b\} = \Pr\{a \leqslant Y < b\} = \Pr\{a < Y < b\} = F(b) - F(a)$$

จากฟังก์ชั่นมวลความน่าจะเป็นและฟังก์ชั่นความหนาแน่นของความน่าจะเป็น สามารถคำนวณดัชนีที่เกี่ยวข้องกับตัวแปรสุ่มได้

2.4 ดัชนีที่เกี่ยวข้องกับตัวแปรสุ่ม

หัวข้อนี้กล่าวถึงดัชนีที่บ่งบอกคุณสมบัติของตัวแปรสุ่ม เช่น ค่าคาดหมายซึ่งแสดงค่ากลาง ๆ ของตัวแปรสุ่ม ค่าความแปรปรวนซึ่งแสดงการกระจายตัว (Spread) ของตัวแปรสุ่ม

2.4.1 ค่าคาดหมาย

ค่าคาดหมาย (Expectation) หรือค่าเฉลี่ย (Mean) เป็นดัชนีแสดงแนวโน้มของตำแหน่งกลาง ๆ ของตัวแปรสุ่ม โดยมีคำจำกัดความดังนี้

- **ถ้า X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง**

$$\mu_X \text{ หรือ } \mathrm{E}(X) = \sum_{x \in \Phi_X} x \times p(x)$$

ตัวอย่างที่ 2.6 แสดงตัวอย่างการคำนวณค่าคาดหมายสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง

ตัวอย่างที่ 2.6. ต่อเนื่องจากตัวอย่างที่ 2.4 ซึ่งมีการแจกแจงดังนี้

$$p(0) = 1/8, p(1) = 3/8, p(2) = 3/8, p(3) = 1/8$$

จึงสามารถคำนวณค่าคาดหมายได้เป็น

$$\mathrm{E}(X) = \left(0 \times \frac{1}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(2 \times \frac{3}{8}\right) + \left(3 \times \frac{1}{8}\right) = 1.5$$

□

- **ถ้า Y เป็นตัวแปรสุ่มแบบต่อเนื่อง**

$$\mu_Y \text{ หรือ } \mathrm{E}(Y) = \int_{y \in \Phi_Y} y \times f(y) dy$$

ตัวอย่างที่ 2.7 แสดงตัวอย่างการคำนวณค่าคาดหมายสำหรับตัวแปรสุ่มแบบต่อเนื่อง

ตัวอย่างที่ 2.7. ต่อเนื่องจากตัวอย่างที่ 2.5 ซึ่งมีการแจกแจงดังนี้

$$f(y) = \frac{1}{2} e^{-y/2}, y \geqslant 0$$

จึงสามารถคำนวณค่าคาดหมายได้เป็น

$$\mathrm{E}(Y) = \int_0^\infty y \left(\frac{1}{2} e^{-y/2}\right) dy = -e^{-y/2}(2+y)\Big|_0^\infty = 2$$

□

ค่าเฉลี่ยเป็นค่าสถิติที่เป็นที่รู้จักและใช้กันมาก แต่ค่าความแปรปรวนก็สำคัญ เพราะแสดงการกระจายตัวของข้อมูล

2.4.2 ค่าความแปรปรวน

ค่าความแปรปรวน (Variance) แสดงการกระจายตัวของตัวแปรสุ่มรอบ ๆ ค่าคาดหมาย มีนิยามดังต่อไปนี้

$$\mathrm{Var}(X) \text{ หรือ } \sigma_X^2 = \mathrm{E}\left[(X-\mu_X)^2\right] \tag{2.9}$$

ยกกำลังสองส่วนต่างระหว่าง X และ μ_X เพื่อให้ค่าที่ได้มีเครื่องหมายบวกทั้งหมด ซึ่งชัดเจนว่าค่าความแปรปรวนไม่ติดลบ หากคำนวณแล้วได้ค่าลบแปลว่าคำนวณผิด สังเกตว่าหน่วยของความแปรปรวนเป็นกำลังสองของหน่วยของค่าคาดหมาย เช่น ถ้าค่าคาดหมายมีหน่วยเป็นเมตร ความแปรปรวนมีหน่วยเป็นเมตร^2 ไม่สะดวกต่อการเข้าใจ ดังนั้นจึงมี**ส่วนเบี่ยงเบนมาตรฐาน** (Standard deviation) ซึ่งเป็นรากที่สองของค่าความแปรปรวน

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

สามารถเขียนนิยาม (2.9) ในรูปที่ง่ายต่อการคำนวณ ดังนี้

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu_X^2 \tag{2.10}$$

ถ้าค่าความแปรปรวนต่ำ แสดงว่าตัวแปรสุ่มนั้นกระจุกตัวใกล้กัน ถ้าค่าความแปรปรวนสูง หมายความว่าตัวแปรสุ่มนั้นกระจัดกระจายห่างกัน การคำนวณค่าความแปรปรวนสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่องและแบบต่อเนื่องมีรูปแบบต่างกัน ดังนี้

- **ถ้า X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง**

$$\mathrm{Var}(X) = \sum_{x\in\Phi_X} (x-\mu_X)^2 p(x)$$

หากจะใช้สมการที่ (2.10) คำนวณ $\mathrm{E}(X^2)$ ได้ดังนี้

$$\mathrm{E}(X^2) = \sum_{x\in\Phi_X} x^2 p(x)$$

ตัวอย่างที่ 2.8 แสดงการคำนวณค่าความแปรปรวนสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง

ตัวอย่างที่ 2.8. ต่อเนื่องจากตัวอย่างที่ 2.4 ซึ่งมี PMF ดังนี้

$$p(0) = 1/8, p(1) = 3/8, p(2) = 3/8, p(3) = 1/8$$

จึงสามารถคำนวณ $\mathrm{E}(X^2)$ ได้ดังนี้

$$\mathrm{E}(X^2) = \left(0^2 \times \frac{1}{8}\right) + \left(1^2 \times \frac{3}{8}\right) + \left(2^2 \times \frac{3}{8}\right) + \left(3^2 \times \frac{1}{8}\right) = 3$$

แทนค่าในสมการที่ (2.10) ได้ $\mathrm{Var}(X) = 3 - 1.5^2 = 0.75$ และ $\sigma_X = 0.866$

$\square$

- **ถ้า Y เป็นตัวแปรสุ่มแบบต่อเนื่อง**

$$\mathrm{Var}(Y) = \int_{y \in \Phi_Y} (y - \mu_Y)^2 f(y) dy$$

หากจะใช้สมการที่ (2.10) คำนวณ $\mathrm{E}(Y^2)$ ได้โดย

$$\mathrm{E}(Y^2) = \int_{y \in \Phi_Y} y^2 f(y) dy$$

ตัวอย่างที่ 2.9 แสดงการคำนวณค่าความแปรปรวนสำหรับตัวแปรสุ่มแบบต่อเนื่อง

ตัวอย่างที่ 2.9. ต่อเนื่องจากตัวอย่างที่ 2.5 ซึ่งมีการแจกแจงดังนี้

$$f(y) = \frac{1}{2} e^{-y/2}, y \geqslant 0$$

จึงสามารถคำนวณ $\mathrm{E}(Y^2)$ ได้ดังนี้

$$\mathrm{E}(Y^2) = \int_0^\infty y^2 \left(\frac{1}{2} e^{-y/2}\right) dy = 8$$

แทนค่าในสมการที่ (2.10) ได้ $\mathrm{Var}(Y) = 8 - 2^2 = 4$ และ $\sigma_Y = 2$

$\square$

ถึงแม้ว่าส่วนเบี่ยงเบนมาตรฐานมีหน่วยเดียวกับค่าคาดหมาย แต่ก็ยังไม่สะดวกในการพิจารณาขนาดของการกระจายเมื่อเทียบกับค่าคาดหมาย ดังนั้นจึงมี**สัมประสิทธิ์การแปรผัน** (Coefficient of variation) ซึ่งมีนิยามดังนี้

$$cv_X = \frac{\sigma_X}{\mu_X}$$

ดังนั้น cv_X จึงไม่มีหน่วย (Dimensionless) ถ้าค่า cv_X ต่ำแปลว่าส่วนเบี่ยงเบนมาตรฐานน้อย (หรือการกระจายตัวของการแจกแจงต่ำ) เมื่อเทียบกับค่าคาดหมาย แต่ถ้าค่า cv_X สูงแปลว่าส่วนเบี่ยงเบนมาตรฐานมากเมื่อเทียบกับค่าคาดหมาย

ตัวอย่างที่ 2.10 แสดงการคำนวณสัมประสิทธิ์การแปรผันสำหรับตัวแปรสุ่มแบบต่อเนื่องและไม่ต่อเนื่อง

ตัวอย่างที่ 2.10. ต่อเนื่องจากตัวอย่างที่ 2.6 ตัวแปรสุ่ม X มีค่าคาดหมาย $\mathrm{E}(X) = 1.5$ และจากตัวอย่างที่ 2.8 ค่าส่วนเบี่ยงเบนมาตรฐาน $\sigma_X = 0.866$ ดังนั้น

$$cv_X = \frac{0.866}{1.5} = 0.577.$$

นั่นคือส่วนเบี่ยงเบนมาตรฐานมีค่าประมาณ 60% ของค่าคาดหมาย และสำหรับตัวอย่างที่ 2.9 ตัวแปรสุ่ม Y มี $cv_Y = 1$

□

ค่าสถิติที่กล่าวมาเกี่ยวข้องกับตัวแปรสุ่มตัวเดียว แต่ค่าความแปรปรวนร่วมและค่าสหสัมพันธ์เป็นค่าความสัมพันธ์สำหรับตัวแปรสุ่ม 2 ตัว

2.4.3 ค่าความแปรปรวนร่วมและค่าสหสัมพันธ์

ระบบที่สนใจอาจมีตัวแปรสุ่มสองตัวที่มีความสัมพันธ์กัน: ในเชิงบวกคือมีแนวโน้มที่จะเพิ่มขึ้นพร้อมกันหรือลดลงพร้อมกัน เช่น ราคาน้ำมันและราคาทองคำ, ค่าแรงขั้นต่ำและยอดขายรถจักรยานยนต์ หรือในเชิงลบคือมีแนวโน้มที่เคลื่อนไหวในทางตรงกันข้าม หากตัวหนึ่งเพิ่มอีกตัวหนึ่งมีแนวโน้มว่าจะลด หรือหากตัวหนึ่งลดอีกตัวหนึ่งมีแนวโน้มว่าจะเพิ่ม เช่น เวลาที่ลูกค้าต้องรอในแถวหรือในสายโทรศัพท์และระดับความพึงพอใจของลูกค้าต่อการบริการ, อุปสงค์และราคาของสินค้าหรือบริการโดยทั่วไป สมมติว่าตัวแปรสุ่มสองตัวนี้คือ M และ N และมีค่าคาดหมาย μ_M และ μ_N ตามลำดับ **ความแปรปรวนร่วม** (Covariance) มีนิยามดังนี้

$$\sigma_{M,N} = \mathrm{E}\left[(M-\mu_M)(N-\mu_N)\right] \tag{2.11}$$

ถ้าค่าความแปรปรวนร่วมเป็นลบ ตัวแปรสุ่มนี้มีแนวโน้มที่จะเคลื่อนไหวในทางตรงกันข้าม แต่หากค่าความแปรปรวนร่วมนี้เป็นบวก หมายความว่าตัวแปรสุ่มนี้มีแนวโน้มที่จะเคลื่อนไหวในทางเดียวกัน สมการที่ (2.11) สามารถเขียนในรูปที่สะดวกต่อการคำนวณ ดังนี้

$$\sigma_{M,N} = \mathrm{E}(MN) - \mu_M\mu_N \tag{2.12}$$

โดย $\mathrm{E}(MN)$ สามารถคำนวณจาก**ฟังก์ชันมวลความน่าจะเป็นร่วม** (Joint probability mass function), $p_{M,N}(m,n)$, และ**ฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม** (Joint probability density function), $f_{M,N}(m,n)$, ซึ่งมีนิยามดังนี้

$$\begin{aligned} p_{M,N}(m,n) &= \Pr\{M=m \text{ และ } N=n\} \\ f_{M,N}(m,n) &= \frac{d^2 \Pr\{M\leqslant m \text{ และ } N\leqslant n\}}{dmdn} \end{aligned}$$

สูตรคำนวณ $\mathrm{E}(MN)$ ขึ้นกับชนิดตัวแปรสุ่ม ดังนี้

- ถ้า M และ N เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง

$$\mathrm{E}(MN) = \sum_{m\in\Phi_M}\sum_{n\in\Phi_N} mn \times p_{M,N}(m,n) \tag{2.13}$$

- ถ้า M และ N เป็นตัวแปรสุ่มแบบต่อเนื่อง

$$\mathrm{E}(MN) = \int_{m\in\Phi_M}\int_{n\in\Phi_N} mn \times f_{M,N}(m,n)dndm \tag{2.14}$$

ตัวอย่างที่ 2.11 แสดงการคำนวณค่าความแปรปรวนร่วมและค่าสหสัมพันธ์

ตัวอย่างที่ 2.11. สมมติให้ตัวแปรสุ่ม M และ N มีฟังก์ชันมวลความน่าจะเป็นร่วมและความน่าจะเป็นของแต่ละตัวแปรสุ่ม ดังแสดงในตาราง 2.1

ตารางที่ 2.1: ฟังก์ชันมวลความน่าจะเป็นร่วมของ M และ N

m/n	1	2	3	$p(m)$
1	0.1	0.2		0.3
3	0.2	0.2	0.3	0.7
$p(n)$	0.3	0.4	0.3	1.0

สมการที่ (2.12) อาศัย $\mathrm{E}(MN)$ ซึ่งสามารถคำนวณด้วยสมการที่ (2.13) ดังนี้

$$\mathrm{E}(MN) = (1\times1\times0.1) + (1\times2\times0.2) + (3\times1\times0.2) + (3\times2\times0.2) + (3\times3\times0.3) = 5$$

ส่วน $\mu_M = (1\times0.3) + (3\times0.7) = 2.4$ และ $\mu_N = (1\times0.3) + (2\times0.4) + (3\times0.3) = 2$ เมื่อแทนค่าในสมการที่ (2.12) จะได้ว่า

$$\sigma_{M,N} = \mathrm{E}(MN) - \mu_M \times \mu_N = 5 - (2.4\times2) = 0.2$$

□

กรณีที่ตัวแปรสุ่มตัวหนึ่งเป็นแบบต่อเนื่องและอีกตัวเป็นแบบไม่ต่อเนื่อง ก็ยังสามารถคำนวณความแปรปรวนร่วมได้โดยใช้ความน่าจะเป็นแบบมีเงื่อนไขเพื่อตรึงค่าตัวแปรหนึ่งในขณะที่หาผลบวกหรือปริพันธ์ของอีกตัวแปรหนึ่งในสมการที่ (2.11)

สังเกตว่าหน่วยของความแปรปรวนร่วมเป็นผลคูณของหน่วยของ M และ N ซึ่งยากที่จะวิเคราะห์ว่าความแปรปรวนร่วมนี้เข้มข้นเพียงใด จึงปรับสัดส่วนของความแปรปรวนร่วมด้วยผลคูณของส่วนเบี่ยงเบนมาตรฐานของทั้งสองตัวแปรนี้ สิ่งที่ได้คือ**ค่าสหสัมพันธ์** (Correlation) ซึ่งมีค่าอยู่ระหว่าง -1 ถึง 1 โดยเครื่องหมายของค่าแสดงว่าความสัมพันธ์เป็นแนวโน้มในเชิงทิศเดียวกัน (+) หรือทิศตรงกันข้าม (-) ส่วนขนาดของตัวเลขแสดงขนาดของความสัมพันธ์ ยิ่งสูงยิ่งมีความสัมพันธ์กันมาก

$$\rho_{M,N} = \frac{\sigma_{M,N}}{\sigma_M\sigma_N} \tag{2.15}$$

ตัวอย่างที่ 2.12 แสดงการคำนวณค่าสหสัมพันธ์

ตัวอย่างที่ 2.12. ต่อเนื่องจากตัวอย่างที่ 2.11 เมื่อใช้สมการ (2.15) จะได้ว่า

$$\rho_{M,N} = \frac{\sigma_{M,N}}{\sigma_M \sigma_N} = \frac{0.2}{0.917 \times 0.775} = 0.282$$

ถึงแม้ว่า M และ N มีแนวโน้มที่จะเคลื่อนไปทางเดียวกัน แต่ความสัมพันธ์ไม่เข้มข้นนัก กล่าวคือ 0.282 ยังห่างจาก 1

□

การพัฒนาแบบจำลองสถานการณ์อาศัยข้อมูลเชิงตัวเลขสำหรับตัวแปรนำเข้าเพื่อนำไปจำลองในแบบจำลอง (รายละเอียดในบทที่ 4) เช่น เวลาในการประกอบชิ้นส่วน, จำนวนครั้งที่เครื่องจักรเสียในช่วงเวลาต่าง ๆ, หรือจำนวนคนไข้ที่เข้าสู่ระบบในแต่ละชั่วโมง นอกจากนี้ยังต้องการข้อมูลเชิงตัวเลขสำหรับดัชนีชี้วัดของระบบที่สนใจเพื่อนำไปตรวจสอบความถูกต้องของแบบจำลอง (หัวข้อที่ 9.3) เช่น เวลารอคอยของลูกค้าจนกว่าจะได้รับการบริการ ข้อมูลเหล่านี้เป็นตัวอย่างหรือตัวแทนของค่าจริงที่ไม่สามารถสังเกตได้ทั้งหมด อาจได้มาจากการจับเวลา (Time study) หรือการรวบรวมข้อมูลในอดีต (Historical data) หัวข้อต่อไปจะกล่าวถึงการสรุปข้อมูลตรวจวัดในเบื้องต้น

2.5 การสรุปข้อมูลตรวจวัด

เมื่อได้ข้อมูลตรวจวัดมาแล้วหรือในขณะที่เก็บ ควรพิจารณาไปด้วยว่าสมจริงหรือไม่ (เช่น ข้อมูลเวลาไม่ควรติดลบ, เวลาตามนาฬิกาที่ลูกค้าออกควรมาหลังเวลาที่ลูกค้าเข้า) และควรสรุปข้อมูลด้วยค่าทางสถิติและกราฟ การนำเสนอข้อมูลด้วยตัวเลขมีข้อดีที่ละเอียดและชัดเจน แต่ไม่เห็นภาพรวมของชุดตัวอย่าง ส่วนการใช้กราฟดีตรงที่นำเสนอชุดตัวอย่างในภาพรวม แต่ไม่ละเอียดเท่าตัวเลข ดังนั้นควรสรุปข้อมูลด้วยทั้งสองวิธีแล้วนำมาพิจารณาประกอบกัน

2.5.1 การสรุปข้อมูลตรวจวัดเชิงตัวเลข

สมมติให้ตัวอย่างมีจำนวน n ค่า $X_1, X_2, \ldots X_n$ อาจเป็นเวลาที่ให้บริการลูกค้าคนที่ i และสำหรับลูกค้า n คน ค่าสรุปทางสถิติหลัก ๆ ที่ควรพิจารณา มีดังนี้

1. **ค่าเฉลี่ยตัวอย่าง (Sample mean)** แสดงตำแหน่งกลางของตัวอย่าง คำนวณได้ดังนี้

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.16}$$

ซึ่งสามารถใช้ประมาณค่าคาดคะเนของตัวแปรสุ่มได้ นั่นคือ $\hat{\mu}_X = \bar{X}$ ฟังก์ชันไมโครซอฟท์เอกซ์เซลที่ใช้คำนวณค่าเฉลี่ยคือ `=average(data array)` โดย `data array` คือข้อมูลที่เรียงชิดกันในแนวแถวหรือคอลัมน์ และมี 1 ค่าใน 1 เซลล์ ตัวอย่างที่ 2.13 แสดงการคำนวณค่าเฉลี่ยจากตัวอย่าง

ตัวอย่างที่ 2.13. ข้างล่างนี้เป็นข้อมูลน้ำหนักของชิ้นงานในหน่วยกรัม

$$62\ 64\ 66\ 67\ 65\ 68\ 61\ 65\ 67\ 65\ 64\ 63\ 67\ 68\ 64\ 66\ 68\ 69\ 65\ 67$$

ขนาดตัวอย่าง (n) คือ 20 และ $\bar{X} = (62 + 64 + \ldots + 67)/20 = 65.55$

□

2. **ส่วนเบี่ยงเบนมาตรฐานตัวอย่าง (Sample standard deviation)** แสดงการกระจายตัวของตัวอย่าง คำนวณได้จากความแปรปรวนของตัวอย่างสำหรับตัวแปรสุ่ม X, S_X^2, ดังนี้

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \qquad (2.17)$$

สามารถนำไปใช้ประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มได้ นั่นคือ $\hat{\sigma}_X = S_X$ ฟังก์ชันไมโครซอฟท์เอกซ์เซลที่ใช้คำนวณส่วนเบี่ยงเบนมาตรฐานคือ `=stdev(data array)` ตัวอย่างที่ 2.14 แสดงการคำนวณส่วนเบี่ยงเบนมาตรฐานจากตัวอย่าง

ตัวอย่างที่ 2.14. ข้อมูลจากตัวอย่างที่ 2.13 จะได้ $S_X = 2.14$

□

3. **ควอนไทล์ (Quantile) หรือเปอร์เซ็นไทล์ (Percentile)** เปอร์เซ็นไทล์ที่ $p, 0 \leqslant p \leqslant 100$, คือค่าที่มี p% ของข้อมูลน้อยกว่าค่านี้ ควอร์เตอร์ (Quarter) แปลว่า 1/4 หรือ 25% ดังนั้น**ควอร์ไทล์** (Quartile) ที่หนึ่งคือเปอร์เซ็นไทล์ที่ 25 สังเกตว่าควอร์ไทล์**ไม่เหมือน**กับควอนไทล์ คือค่าที่มีประมาณ 25% ของข้อมูลในตัวอย่างน้อยกว่าค่านี้และประมาณ 75% ของตัวอย่างมากกว่าค่านี้ ควอร์ไทล์ที่สองคือ**ค่ามัธยฐาน** (Median) หรือเปอร์เซ็นไทล์ที่ 50 ซึ่งมีตัวอย่างประมาณ 50% มากกว่าค่านี้และอีก 50% น้อยกว่าค่านี้ ขั้นตอนการคำนวณมีดังนี้

   (i) เรียงข้อมูลตัวอย่างจากน้อยไปหามาก

   ตัวอย่างที่ 2.15. ใช้ข้อมูลในตัวอย่างที่ 2.13 เรียงได้ดังนี้

$$61\ 62\ 63\ 64\ 64\ 64\ 65\ 65\ 65\ 65\ 66\ 66\ 67\ 67\ 67\ 67\ 68\ 68\ 68\ 69$$

(ii) คำนวณลำดับของข้อมูลที่ต้องการ คือ np โดย n คือขนาดของข้อมูลและ p คือเปอร์เซ็นไทล์ที่ต้องการในรูปจุดทศนิยม (0 ถึง 1) เช่น ควอร์ไทล์ที่หนึ่งมี $p = 0.25$

สำหรับตัวอย่างที่ 2.13 และควอร์ไทล์ที่หนึ่ง $n = 20, p = 0.25$ ดังนั้นลำดับที่ต้องการคือ $np = 20 \times 0.25 = 5$

(iii) ค่าในตัวอย่างลำดับที่ np คือเปอร์เซ็นไทล์ที่ต้องการ อาจต้องเทียบบรรญัติไตรยางค์ (Interpolate) หากค่า np ไม่ใช่จำนวนเต็ม

ต่อจากตัวอย่างที่ 2.15 ค่าประมาณของควอร์ไทล์ที่หนึ่งคือข้อมูลลำดับที่ 5 มีค่า 64

ฟังก์ชันไมโครซอฟท์เอกซ์เซลที่ใช้คำนวณควอร์ไทล์คือ `=quartile(data array,` ควอร์ไทล์ที่ต้องการ`)` เช่น ถ้าต้องการมัธยฐานซึ่งก็คือควอร์ไทล์ที่สอง จะกำหนดให้ควอร์ไทล์ที่ต้องการเป็น 2 หรือจะใช้ฟังก์ชัน `=percentile(data array,` เปอร์เซ็นที่ต้องการ`)`

ตัวอย่างที่ 2.16. ข้อมูลจากตัวอย่างที่ 2.13 จะได้ควอร์ไทล์แรกเป็น 64 ควอร์ไทล์ที่สองเป็น 65.5 และควอร์ไทล์ที่สามเป็น 67

□

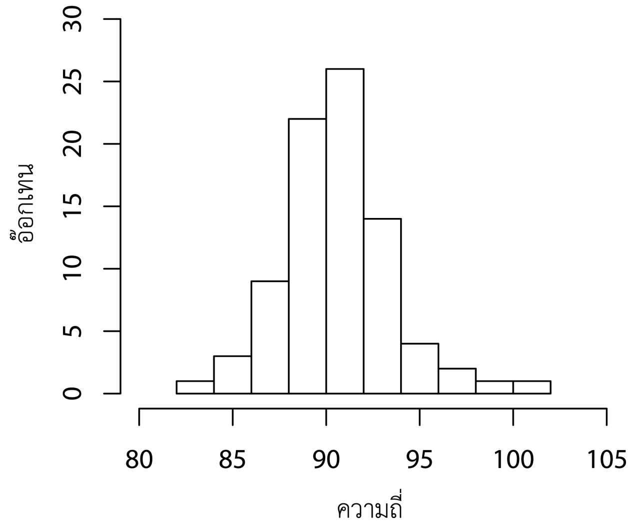
นอกเหนือจากการสรุปข้อมูลเชิงตัวเลขแล้ว ควรสรุปข้อมูลเชิงแผนภาพประกอบด้วย

2.5.2 การสรุปข้อมูลเชิงแผนภาพ

การสรุปข้อมูลเชิงแผนภาพแสดงให้เห็นภาพรวมของข้อมูล เช่น แนวโน้มหรือการกระจายตัวของข้อมูล หัวข้อนี้จะกล่าวถึงเพียงฮีสโตแกรมและกราฟอนุกรมเวลา

ฮีสโตแกรม (Histogram)

ฮีสโตแกรมเป็นตัวแทนของกราฟฟังก์ชันความหนาแน่นความน่าจะเป็นที่สร้างจากข้อมูลตัวอย่าง แกนนอนแบ่งข้อมูลเป็นช่วง ๆ แกนตั้งแสดงความถี่ (Frequency) ของข้อมูลหรือความถี่สัมพัทธ์ของตัวอย่างที่ตกอยู่ในช่วงนั้น ๆ ฮีสโตแกรมมีประโยชน์สำหรับการพิจารณาการกระจายตัวของข้อมูล เช่น พิสัยกว้างหรือไม่, มีจุดยอดของกราฟหลายจุดหรือไม่, สมมาตร (Symmetric) หรือไม่, หรือว่าเบ้ซ้ายหรือเบ้ขวา รูปที่ 2.3 แสดงตัวอย่างของฮีสโตแกรมโดยใช้ข้อมูลจากแบบฝึกหัดข้อ 6.14 ใน (Montgomery and Runger, 2002) จะเห็นว่าตัวอย่างชุดนี้ค่อนข้างสมมาตรรอบ ๆ ค่าเฉลี่ย โดยข้อมูลส่วนใหญ่ตกอยู่ในช่วง 87-92 การพัฒนาแบบจำลองใช้ฮีสโตแกรมเพื่อคัดเลือกการแจกแจงพาราเมตริกที่เหมาะสม (รายละเอียดในบทที่ 4)



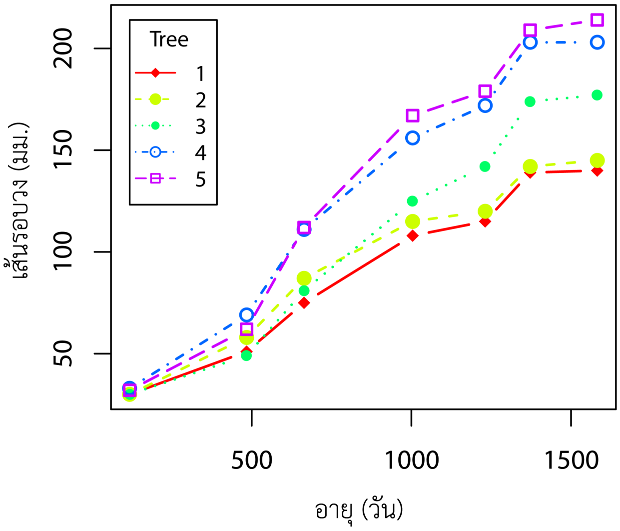
รูปที่ 2.3: ตัวอย่างฮีสโตแกรม

จำนวนช่วง (Bin) ที่แนะนำให้ใช้อยู่ระหว่าง 5 ถึง 20 ขึ้นอยู่กับขนาดข้อมูล เมื่อมีข้อมูลจำนวนมาก จำนวนกล่องที่ควรใช้ยิ่งมาก คำแนะนำคือใช้จำนวนช่วงประมาณรากที่สองของขนาดข้อมูล (Montgomery and Runger, 2002) มีซอฟต์แวร์สำเร็จรูปหลากหลายที่สามารถสร้างฮีสโตแกรมได้ เช่น ไมโครซอฟท์เอกซ์เซล มีฟังก์ชัน `Histogram` ใน `Add-in` ที่ชื่อ `Data analysis` หรือจะใช้โปรแกรมคำนวณทางสถิติ เช่น `Minitab` หรือ `R` ซึ่งเป็นฟรีแวร์ โปรแกรมเลือกจำนวนช่วงให้อัตโนมัติ แต่ผู้ใช้ควรทดลองเปลี่ยนจำนวนช่วงด้วย เพราะรูปลักษณ์ของฮีสโตแกรมเปลี่ยนไปเมื่อจำนวนช่วงเปลี่ยน ซึ่งอาจบิดเบือนการตีความได้

อนุกรมเวลา (Time series)

หากชุดตัวอย่างเป็นข้อมูลอนุกรมเวลา กล่าวคือ เป็นข้อมูลที่สัมพันธ์กับเวลา เช่น ยอดขายสินค้าในแต่ละวัน หรือจำนวนชั่วโมงที่เครื่องจักรทำงานได้ในแต่ละเดือน ควรสร้างกราฟอนุกรมเวลานี้ด้วย ให้แกนนอนเป็นเวลาและแกนตั้งเป็นค่าที่พิจารณา ซึ่งก็คือ**แผนภูมิกระจาย** (Scatter plot) นั่นเอง ที่มีข้อมูลแกนนอนเป็นเวลา ประเด็นที่พึงพิจารณาจากกราฟ มีดังนี้ กราฟมีแนวโน้มหรือไม่ มีรูปแบบ (Pattern) หรือไม่ เช่น สินค้าบางอย่างขึ้นอยู่กับฤดูกาล, เครื่องปรับอากาศและพัดลมขายดีในฤดูร้อน, เครื่องทำน้ำอุ่นและผ้าห่มขายดีในฤดูหนาว กราฟเพิ่มขึ้นอย่างต่อเนื่อง หรือลดลงอย่างต่อเนื่อง หรือกราฟดูเป็นเชิงสุ่ม ไม่มีรูปแบบใด ๆ ที่เด่นชัด รูปที่ 2.4 แสดงขนาดเส้นรอบวงของต้นส้มที่ระยะเวลาในการปลูกต่าง ๆ กันจำนวน 5 ต้น (Kabacoff, 2012) นอกจากจะพิจารณาการเติบโตของต้นใดต้นหนึ่งแล้ว ยังสามารถพิจารณาหลาย ๆ ต้นพร้อมกันได้

ซึ่งช่วยในการเปรียบเทียบและศึกษาข้อเหมือนหรือข้อแตกต่าง การพิจารณาอนุกรมเวลาเช่นนี้ช่วยให้เข้าใจธรรมชาติของข้อมูล และเอื้อประโยชน์ในการจำลอง



รูปที่ 2.4: ตัวอย่างกราฟอนุกรมเวลา

2.6 บทส่งท้าย

บทนี้มีวัตถุประสงค์เพื่อให้ผู้อ่านทบทวนหลักความน่าจะเป็นและสถิติที่ใช้บ่อยในการวิเคราะห์ด้วยแบบจำลองสถานการณ์ มิใช่เป็นการสอนพื้นฐาน หากผู้อ่านต้องการทบทวนเพิ่มเติม ให้ลองสืบค้นเอกสารประกอบการสอนและแบบฝึกหัดออนไลน์ในอินเตอร์เน็ตด้วยคำสำคัญ ”Prob stat tutorial for simulation” นอกจากนี้มีเอกสารภาษาไทย เช่น บทที่ 2 ของตำราการจำลองสถานการณ์โดยผศ.ดร.วุฒิชัย วงษ์ทัศนีย์กร จากมหาวิทยาลัยธรรมศาสตร์ (วุฒิชัย วงษ์ทัศนีย์กร, 2555) หรือตำราสถิติพื้นฐานทั่วไป เช่น (ประไพศรี สุทัศน์ ณ อยุธยา และ พงศ์ชนัน เหลืองไพบูลย์, 2554)

แบบฝึกหัด

2.1. ที่นิคมอุตสาหกรรมแห่งหนึ่ง มีโอกาสประมาณ 35% ที่จะไม่มีอุบัติเหตุเกิดขึ้นในช่วง 1 สัปดาห์ของการทำงาน

(i) ให้หาความน่าจะเป็นที่จะมีอุบัติเหตุอย่างน้อย 1 ครั้งใน 1 สัปดาห์

(ii) ให้คำนวณความน่าจะเป็นที่จะไม่มีอุบัติเหตุเกิดขึ้น 3 สัปดาห์ต่อเนื่องกัน สมมติให้อุบัติเหตุแต่ละครั้งไม่เกี่ยวข้องกัน

2.2. ผู้ผลิตรายหนึ่งรับวัตถุดิบจากซัพพลายเออร์สองแห่ง การสุ่มตรวจในอดีตพบว่าสินค้าจากโรงงาน ก มีของเสีย 5% และสินค้าจากโรงงาน ข มีของเสีย 3% เราสุ่มวัตถุดิบนี้ 1 ชิ้นจากล็อตสินค้าของบริษัท ก และอีกชิ้นจากล็อตสินค้าของบริษัท ข โดยสมมติว่าการสุ่มจากล็อตทั้งสองนี้เป็นอิสระต่อกัน ให้คำนวณความน่าจะเป็นดังต่อไปนี้

(i) ได้วัตถุดิบดี 1 ชิ้นและเสีย 1 ชิ้น

(ii) ได้ชิ้นส่วนดีทั้งสองชิ้น

(iii) สัดส่วนของเสียรวมที่ผู้ผลิตรายนี้รับเข้ามา หากผู้ผลิตรายนี้สั่งของจากโรงงาน ก และ ข ในปริมาณเท่า ๆ กัน

2.3. *ความสัมพันธ์ระหว่างความดันโลหิตของแม่และลูกคนแรก*: ให้ A เป็นเหตุการณ์ที่แม่มีความดันเลือดอย่างน้อย 95 มม.ปรอท และให้ B เป็นเหตุการณ์ที่เด็กมีความดันเลือดอย่างน้อย 80 มม.ปรอท หากรู้ว่า

$$\Pr\{A\} = 0.1, \Pr\{B\} = 0.2 \text{ และ } \Pr\{A \cap B\} = 0.05$$

ความดันโลหิตของแม่และลูกเป็นอิสระต่อกันหรือไม่

2.4. ตัวแปรสุ่มแบบไม่ต่อเนื่อง X มีฟังก์ชันมวลความน่าจะเป็นดังต่อไปนี้

x	0	1	2
$p(x)$	0.5	0.2	0.3

ให้คำนวณค่าต่อไปนี้

(i) $\Pr\{X = 0.5\}$

(ii) ฟังก์ชันมวลความน่าจะเป็นสะสม $F(x)$ ที่ x ต่าง ๆ

(iii) ค่าคาดหมาย

(iv) ค่าความแปรปรวน, ส่วนเบี่ยงเบนมาตรฐาน, และสัมประสิทธิ์การแปรผัน

2.5. กำหนดให้ตัวแปรสุ่มแบบต่อเนื่อง Y แทนเส้นผ่านศูนย์กลางของรูที่เจาะในชิ้นงานโลหะ เป้าหมายคือ 12.5 มิลลิเมตรแต่ขนาดจริงมักจะใหญ่กว่านี้ ข้อมูลที่เก็บระบุว่าการแจกแจงของ Y มีรูปแบบดังนี้

$$f(y) = 20e^{-20(y-12.5)}, y \geqslant 12.5$$

ให้คำนวณค่าต่อไปนี้

(i) ชิ้นงานที่ใช้ได้ต้องมีรูอยู่ระหว่าง 12.5 มม. ถึง 12.6 มม. สัดส่วนของชิ้นงานที่ต้องแก้ไขหรือคัดทิ้งเป็นเท่าใด

(ii) ฟังก์ชันความหนาแน่นความน่าจะเป็นสะสม $F(y)$

(iii) ค่าคาดหมาย

(iv) ค่าความแปรปรวน, ส่วนเบี่ยงเบนมาตรฐาน, และสัมประสิทธิ์การแปรผัน

2.6. พิจารณา PDF ดังต่อไปนี้

$$f(t) = \begin{cases} \frac{1}{b-a} & a \leqslant t \leqslant b \\ 0 & \text{หากอยู่นอกช่วง} \end{cases}$$

(i) กำหนด CDF $F(t)$

(ii) กำหนดฟังก์ชัน CDF ผกผัน $F^{-1}(t)$.

2.7. พิจารณาข้อมูลดังต่อไปนี้

$$5.3 \ \ 10.1 \ \ 5.9 \ \ 12.2 \ \ 11.2 \ \ 12.4 \ \ 9.2 \ \ 5 \ \ 5.8 \ \ 7.2 \ \ 8.5 \ \ 7.3 \ \ 3.9 \ \ 10.5 \ \ 9.5 \ \ 6.2 \ \ 10 \ \ 4.7 \ \ 6.4 \ \ 8.1$$

ให้คำนวณหรือสร้างกราฟ ดังนี้

(i) ค่าเฉลี่ยตัวอย่าง

(ii) ส่วนเบี่ยงเบนมาตรฐานตัวอย่าง

(iii) ค่ามัธยฐาน

(iv) ฮีสโตแกรม และอธิบายลักษณะของฮีสโตแกรมนี้

2.8. จำนวนรถยนต์ที่มาถึงร้านเปลี่ยนน้ำมันเครื่อง K-Quik ในช่วง 200 ชั่วโมงการทำงาน เป็นดังแสดงในตาราง ร้านนี้ใช้ระบบแถวคอยแบบแถวเดียว (Single waiting line)

จำนวนรถที่มาถึงใน 1 ชั่วโมง	ความถี่
3 หรือน้อยกว่า	0
4	10
5	30
6	70
7	50
8	40
9 หรือมากกว่า	0

กำหนดให้ X เป็นจำนวนรถที่มาถึงในช่วงหนึ่งชั่วโมงการทำงาน

(i) ให้ประมาณค่า $\mathrm{Pr}\{X = 7\}$.

(ii) หากจำลอง X ว่าเป็นตัวแปรสุ่มแบบปัวซอง ให้ประมาณค่าอัตราการมาถึงสำหรับตัวแปรสุ่มปัวซองนี้ (ต่อชั่วโมง)

3

การจำลองสถานการณ์ด้วยสเปรดชีต

โปรแกรมสเปรดชีต (Spreadsheet) ที่ใช้กันอย่างแพร่หลายที่สุด คือ ไมโครซอฟท์เอกซ์เซล (Microsoft Excel) นอกจากนี้ยังมีซอฟต์แวร์และฟรีแวร์อื่นๆ เช่น Google Spreadsheet, OpenOffice Calc, และ Lotus 1-2-3 จากที่ผู้เขียนสังเกตการใช้งานจริง หากไม่เก็บข้อมูลในฐานข้อมูล ก็มักจะเก็บในรูปไฟล์เอกซ์เซล เพราะเอกซ์เซลมีรูปลักษณ์เป็นตาราง และมีเครื่องมือพร้อมสำหรับการวิเคราะห์เพิ่มเติม เช่น คำนวณค่าเฉลี่ย สร้างกราฟ หรือวิเคราะห์ด้วยการถดถอยเชิงเส้น การจำลองสถานการณ์ช่วยให้การวิเคราะห์ด้วยเอกซ์เซลสมบูรณ์มากขึ้นในแง่การวิเคราะห์ความเสี่ยงหรือประเมินโอกาสที่เหตุการณ์ที่สนใจจะเกิดขึ้นได้ โดยแสดงด้วยค่าความน่าจะเป็น ฮีสโตแกรมของค่าเอาท์พุท หรือเปอร์เซ็นไทล์

เริ่มแรก โปรแกรมสเปรดชีตได้รับการออกแบบเพื่องานบัญชีและด้านการเงิน จึงมีรูปลักษณ์เป็นตารางที่มีแถวและคอลัมน์ (Wikipedia, 2012b) นอกจากนี้สเปรดชีตยังเหมือนเครื่องคิดเลขขนาดใหญ่ สามารถใช้คำนวณได้หลากหลายประเภท ผู้ใช้สามารถนิยามฟังก์ชันเพิ่มเติมด้วยมาโคร (Macro) หรือใช้แอดอิน (Add-in) จึงมีหนังสือเรียนทางด้านการวิจัยการดำเนินงาน (Operations Research) หลายเล่มที่ใช้เอกซ์เซลเป็นเครื่องมือคำนวณหรือสร้างแบบจำลอง เช่น Albright et al. (2009), Ragsdale (2008), และ Winston (2004) บางส่วนของเนื้อหาและตัวอย่างในบทนี้ได้จากสไลด์สำหรับผู้สอนของหนังสือเรียนโดย Cliff Ragsdale (Ragsdale, 2008)

บทนี้จะใช้เอกซ์เซลแอดอินที่ชื่อพ็อพทูล (Hood, 2010b) เพื่อพัฒนาตัวอย่างแบบจำลอง ผู้อ่านสามารถดาวน์โหลดพ็อพทูลได้ฟรีจาก `http://www.poptools.org/download/` ผู้เขียนเลือกใช้ซอฟต์แวร์นี้เนื่องจากไม่ต้องเสียค่าใช้จ่าย มีคุณภาพดี และยังมีการปรับปรุงดูแลอย่างต่อเนื่อง พ็อพทูลมีฟังก์ชันอื่นๆ ด้วย เช่น การจัดการเมตริกซ์ ผู้อ่านสามารถศึกษาวิธีใช้งานพ็อพทูลได้จากตัวอย่าง (เลือก **Add-Ins** > **PopTools** > **Demos**)

ผู้เขียนจะไม่กล่าวถึงซอฟต์แวร์ทางการค้าสำหรับการจำลองสถานการณ์บนสเปรดชีต เช่น @Risk หรือ CrystalBall ซึ่งมีข้อดีที่ใช้งานง่าย และมีเครื่องมือวิเคราะห์ให้เลือกหลากหลายทั้งในเชิงตัวเลขและเชิงแผนภาพ ถึงแม้จะไม่ใช้ซอฟต์แวร์เหล่านี้ แต่หลักการจำลองความไม่แน่นอนในแบบจำลองที่จะนำเสนอยังคงเป็นหลักการทั่วไปที่ใช้ได้กับซอฟต์แวร์อื่นด้วย เพียงแต่คำสั่งของโปรแกรมแตกต่างออกไป ผู้อ่านควรมีทักษะไมโครซอฟท์เอกซ์เซลพื้นฐาน หากไม่เคยใช้ก็สามารถศึกษาด้วยตนเองจากเว็บไซต์ต่างๆ

หัวข้อแรกของบทนี้จะกล่าวถึงการใช้ตัวแปรสุ่มเพื่อจำลองตัวแปรนำเข้าที่ไม่แน่นอน

3.1 ตัวแปรสุ่มและความไม่แน่นอน

เมื่อวิเคราะห์ด้วยสเปรดชีต ค่าในเซลล์บางเซลล์เป็นตัวแปรอิสระ (Independent variable) ที่ไม่แน่นอน สมมติให้ค่าเอาท์พุทหรือตัวแปรตาม Y (Dependent variable) เกี่ยวข้องกับตัวแปรอิสระ X_i ด้วยฟังก์ชัน

$$Y = f(X_1, X_2, \ldots, X_k) \tag{3.1}$$

หากตัวแปรอิสระบางตัวใน $\{X_1, X_2, \ldots, X_k\}$ เป็นค่าสุ่ม ค่า Y ก็ย่อมเป็นค่าสุ่มเช่นกัน การจำลองสถานการณ์เป็นเครื่องมือวิเคราะห์ที่เหมาะสมสำหรับแบบจำลองเช่นนี้ โดยเฉพาะอย่างยิ่งการวิเคราะห์ทางการเงินหรือการประมาณอย่างหยาบ (Rough-cut approximation)

บทที่ 2 ได้กล่าวถึงการจำลองความไม่แน่นอนด้วยตัวแปรสุ่ม ซึ่งคือตัวแปรที่ไม่สามารถระบุได้อย่างตายตัวแน่นอน ค่าตัวแปรนำเข้าบางตัวในแบบจำลองสเปรดชีตที่สมมติว่าเป็นค่าคงที่ จริง ๆ อาจเป็นค่าที่ไม่แน่นอนเพราะเป็นค่าในอนาคต เช่น ราคาของวัตถุดิบปีหน้า อัตราดอกเบี้ยในไตรมาสหน้า จำนวนพนักงานในบริษัทปีหน้า หรือยอดการสั่งซื้อในรอบถัดไปของลูกค้า ค่าที่คำนวณโดยอาศัยตัวแปรนำเข้าที่ไม่แน่นอน (เช่น สมการที่ 3.1) ก็ย่อมผันผวนตามไปด้วย การตัดสินใจที่กระทำบนพื้นฐานข้อมูลที่มีความผันแปรจึงมีความเสี่ยง ซึ่งมีนัยว่ามีโอกาสเกิดความสูญเสีย

การวิเคราะห์ความเสี่ยงมีความจำเป็นเพราะการกำหนดค่าตัวแปรนำเข้าที่จริง ๆ แล้วไม่แน่นอนด้วยค่าคาดหมาย ซึ่งเป็นค่าคงที่ค่าหนึ่ง ไม่สามารถบอกถึงการกระจายตัว (Variability) ของดัชนีชี้วัดที่เป็นค่าเอาท์พุทได้ ตัวอย่างที่ 3.1 ต่อไปนี้แสดงให้เห็นว่า ค่าเฉลี่ยเพียงอย่างเดียวนั้นไม่เพียงพอสำหรับการตัดสินใจ จำเป็นต้องพิจารณาการกระจายตัวประกอบด้วย

ตัวอย่างที่ 3.1. สมมติว่ามีการลงทุนที่ใช้เงินลงทุน 10,000 บาท และคาดว่าจะได้รับผลตอบแทนพร้อมเงินต้นเป็นเงิน 100,000 บาทใน 2 ปี คุณจะลงทุนหรือไม่ หากผลตอบแทนนี้เป็นไปได้ 2 ทาง คือ

1. ผลตอบแทนมีการแจกแจงอย่างสม่ำเสมอในช่วง 90,000 ถึง 110,000 บาท

2. ผลตอบแทนมีมีการแจกแจงอย่างสม่ำเสมอในช่วง -300,000 ถึง 500,000 บาท

สังเกตว่าสองสถานการณ์นี้มีค่าเฉลี่ยเท่ากัน คนทั่วไปที่มิใช่นักพนันคงจะเลือกแบบแรกมากกว่าแบบที่สองเพราะความเสี่ยงต่ำกว่า อย่างไรก็ได้เงินต้นคืน แต่แบบที่สองมีโอกาสให้ผลตอบแทนสูงมาก ๆ หรือขาดทุนได้ □

ตัวอย่างที่ 3.1 นี้แสดงให้เห็นว่าค่าเฉลี่ยหนึ่งค่าเกิดขึ้นได้หลายรูปแบบซึ่งเกี่ยวข้องกับระดับความเสี่ยงที่ต่างกัน การจำลองสถานการณ์สามารถให้ค่าเชิงตัวเลขกับความเสี่ยง

การจำลองสถานการณ์

ในกรณีที่ไม่ทราบค่าข้อมูลนำเข้าบางตัว เช่น ยอดสั่งซื้อของลูกค้าในปีหน้า ผู้วิเคราะห์อาจใช้ค่าประมาณของยอดซื้อหลาย ๆ ค่า เช่น ขอบเขตบน ขอบเขตล่าง หรือค่าที่น่าจะเป็นไปได้มากที่สุด แล้วประมวลผลแบบจำลองภายใต้สถานการณ์ 3 ชนิดนี้ เรียกอีกนัยหนึ่งคือ**การวิเคราะห์แบบถ้า** (What-if analysis) การจำลองสถานการณ์ก็คล้ายการวิเคราะห์แบบนี้แต่เป็นอัตโนมัติกว่า คือสุ่มเลือกสถานการณ์ของชุดตัวแปรนำเข้าได้

อย่างไม่มีอคติ และสามารถพิจารณาได้หลากหลายสถานการณ์กว่ามาก ค่าเอาท์พุทที่ได้จากการประมวลผลเหล่านี้จะถูกวิเคราะห์รวมกันเพื่อเข้าใจพฤติกรรมของเอาท์พุทที่เป็นดัชนีชี้วัด และประเมินความเสี่ยงที่เกิดจากความไม่แน่นอนของตัวแปรนำเข้า ผู้วิเคราะห์สามารถใช้สเปรดชีตเพื่อจำลองค่าของตัวแปรนำเข้าที่อาจจะเกิดขึ้นในอนาคต และประมวลผลเพื่อประมาณค่าของดัชนีชี้วัด ประเด็นที่ผู้วิเคราะห์ควรพึงตระหนักคือ

- การวิเคราะห์ความเสี่ยงเกิดจากข้อสังเกตว่าค่าเฉลี่ยอย่างเดียวไม่เพียงพอ ถึงแม้ว่าจะเป็นค่าเฉลี่ยระยะยาว และควรพิจารณาค่าสถิติอื่นประกอบด้วย เช่น ส่วนเบี่ยงเบนมาตรฐาน หรืออาจพิจารณากราฟ เช่น ฮีสโตแกรม
- ถึงแม้ว่าต้องการเพียงค่าเฉลี่ยของดัชนีชี้วัด การแทนค่าตัวแปรนำเข้าที่ไม่แน่นอนด้วยค่าเฉลี่ยก็ยังไม่ถูกต้อง เพราะตัวแปรนำเข้าเหล่านี้อาจมีปฏิสัมพันธ์ต่อกัน เช่น ล็อตสั่งซื้อที่มีขนาดใหญ่อาจใช้เวลาในการผลิตนานกว่าล็อตเล็ก ถ้าแบบจำลองสมมติว่าเวลาในการผลิตไม่ขึ้นกับขนาดล็อตสั่งซื้อ ผลการคำนวณระดับการให้บริการ (Service level) ก็ผิดพลาด

บทนี้จะนำเสนอการพัฒนาแบบจำลองสเปรดชีตโดยใช้ตัวอย่างเป็นหลัก

3.2 ตัวอย่างการวางแผนงบประมาณของบริษัทมายา

ตัวอย่างนี้ดัดแปลงจากตัวอย่าง Hungry Dawg Restaurant ใน Ragsdale (2008) บริษัทมายาจำกัดกำลังเติบโต มีรายได้และจำนวนพนักงานเพิ่มขึ้นอย่างต่อเนื่อง เพื่อให้พนักงานมีกำลังใจ บริษัทจึงสนับสนุนทุนการศึกษาให้กับบุตรหลานของพนักงาน โดยให้เบิกจ่ายค่าธรรมเนียมการศึกษาตามจริงในใบเสร็จ ปัจจุบันมีจำนวนพนักงาน 63 คนและเพิ่มด้วยอัตรา 2% ต่อปี และค่าธรรมเนียมฯ เฉลี่ย ที่พนักงานเบิกคือ 16,541 บาทต่อคนต่อปี และเพิ่มด้วยอัตรา 5% ต่อปี บริษัทต้องการวางแผนงบประมาณรายจ่ายล่วงหน้า 5 ปี โดยจะยังไม่คำนึงถึงอัตราเงินเฟ้อ ตารางที่ 3.1 แสดงการวิเคราะห์เบื้องต้นด้วยเอกซ์เซล และตารางที่ 3.2 แสดงสูตรคำนวณ สังเกตว่าเมื่อคัดลอกสูตร เอกซ์เซลเลียนแบบตำแหน่งของเซลล์ เช่น เมื่อคัดลอกเซลล์ `B11` มาวางที่ `B12` สูตรในเซลล์คือ `=B11*(1+$F$5)` การวาง `$` หน้าตำแหน่งแถวหรือตำแหน่งคอลัมน์เพื่อตรึงไม่ให้ขยับเมื่อคัดลอก

สังเกตว่าการวิเคราะห์ในตัวอย่างนี้เกี่ยวข้องกับเหตุการณ์ในอนาคต แบบจำลองได้สมมติว่าอัตราที่เพิ่มขึ้นคงที่และเท่ากันทุกปี นอกจากนี้แล้วค่าใช้จ่ายรวมตลอด 5 ปียังเป็นค่า ๆ เดียว แบบจำลองไม่สามารถประเมินได้ว่าโอกาสที่ค่าใช้จ่ายจะเป็น 6.43 ล้านบาทจริง ๆ นั้นเป็นเท่าใด หรือความน่าจะเป็นที่ค่าใช้จ่ายจะเกินเป้าที่ตั้งไว้ (เช่น 6 ล้านบาท) เป็นเท่าใด การจำลองสถานการณ์สามารถตอบประเด็นเหล่านี้ได้ ก่อนที่จะปรับแบบจำลองในตารางที่ 3.1 ผู้เขียนจะกล่าวถึงการสร้างตัวเลขสุ่มจากซอฟต์แวร์พ็อพทูล

ตารางที่ 3.1: แบบจำลองการวิเคราะห์บนเอกซ์เซลสำหรับบริษัทมายา (หัวข้อที่ 3.2)

	A	B	C	D	E	F	G
1							
2			**บริษัทมายา จำกัด**				
3							
4	**เงื่อนไขเริ่มต้น**				**ข้อสมมุติ**		
5	จำนวนพนักงานในปัจจุบัน			63	อัตราที่เพิ่มขึ้น	2%	ต่อปี
6	ค่าธรรมเนียมการศึกษาเฉลี่ยที่พนักงานเบิก (ต่อคน)			16,541	อัตราที่เพิ่มขึ้น	5%	ต่อปี
7							
8		จำนวน	ค่าธรรมเนียมที่เบิก	ค่าใช้จ่าย			
9	ปีที่	พนักงาน	ต่อพนักงาน 1 คน	รวม			
10	1	64	฿17,368.05	฿1,116,070.89			
11	2	66	฿18,236.45	฿1,195,311.93			
12	3	67	฿19,148.28	฿1,280,179.07			
13	4	68	฿20,105.69	฿1,371,071.79			
14	5	70	฿21,110.97	฿1,468,417.88			
15			**ค่าใช้จ่ายรวมของบริษัท**	**฿6,431,051.56**			

ตารางที่ 3.2: สูตรเอกซ์เซลสำหรับตัวอย่างการวางแผนงบประมาณบริษัทมายาในตารางที่ 3.1

ตำแหน่งเซลล์	สูตรเอกซ์เซล	คัดลอกไปยังเซลล์
B10	=D5*(1+F5)	
C10	=D6*(1+F6)	
D10	=C10*B10	D11:D14
B11	=B10*(1+$F$5)	B12:B14
C11	=C10*(1+$F$6)	C12:C14
D15	=sum(D10:D14)	

3.2.1 การสร้างตัวเลขสุ่มในพ็อพทูล

การสุ่มค่าตัวแปรนำเข้าที่ไม่แน่นอนในแบบจำลองอาศัยฟังก์ชันเพื่อสร้างตัวเลขสุ่ม (Random variate generation, RVG) ผู้ใช้กำหนดการแจกแจงความน่าจะเป็นและค่าพารามิเตอร์ที่เกี่ยวข้อง สมมติว่าเป็นการแจกแจงแบบปกติที่มีค่าเฉลี่ย 20 และส่วนเบี่ยงเบนมาตรฐาน 2 เมื่อพิจารณาค่าที่สุ่มได้หลาย ๆ ค่า ตัวเลขที่สร้างได้นี้ดูเหมือนมาจากการแจกแจงแบบปกติจริง ๆ เช่น เมื่อพิจารณาฮีสโตแกรมหรือค่าทางสถิติอื่น ๆ ฟังก์ชัน RVG ที่มีในเอกซ์เซลคือ `RAND()` ซึ่งให้ตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง [0,1] บทนี้จะกล่าวถึงการสร้างตัวเลขสุ่มอย่างสั้น ๆ ให้พอเข้าใจการกำหนดฟังก์ชันสร้างตัวเลขสุ่ม บทที่ 4 จะอธิบายเรื่องการแจกแจงความน่าจะเป็นโดยละเอียด

ตัวแปรสุ่มแบ่งเป็นสองประเภทหลัก คือ ตัวแปรสุ่มแบบไม่ต่อเนื่อง (หัวข้อที่ 2.3.1) และตัวแปรสุ่มแบบต่อเนื่อง (หัวข้อที่ 2.3.2) ตารางที่ 3.3 และ 3.4 แสดงฟังก์ชันการสร้างตัวเลขสุ่มและการกำหนดค่าพารามิเตอร์สำหรับการแจกแจงแบบไม่ต่อเนื่องและการแจกแจงแบบต่อเนื่องในพ็อพทูล ตามลำดับ

ตารางที่ 3.3: สูตรพ็อพทูลสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง

การแจกแจงความน่าจะเป็น	**พารามิเตอร์**	**สูตรพ็อพทูล**
เบอร์นูลลี (Bernoulli)	`p` ความน่าจะเป็นที่เหตุการณ์ที่สนใจจะเกิดขึ้น	`dBinomialDev(1, p)`
ทวินาม (Binomial)	`n` จำนวนของเหตุการณ์ `p` ความน่าจะเป็นที่เหตุการณ์จะเกิดขึ้น	`dBinomialDev(n, p)`
พหุนาม (Multinomial) ไม่ต่อเนื่อง (Discrete)	p_i ความน่าจะเป็นที่จะได้ค่า v_i $i = 1, \ldots, n$	`DiscreteDev(`$v_1 : v_n$`,` $p_1 : p_n$`)` $v_1 : v_n$ และ $p_1 : p_n$ อยู่ในรูปอาร์เรย์
สม่ำเสมอแบบไม่ต่อเนื่อง (Discrete uniform)	`a` ขอบเขตล่าง `b` ขอบเขตบน โดยที่ `a` < `b`	`dRandInt(a, b)`
เรขาคณิต (Geometric)	`p` ความน่าจะเป็นที่เหตุการณ์จะเกิดขึ้น	`dGeomDev(p)`
ไฮเปอร์จีโอเมตริก (Hypergeometric)	`n` จำนวนของครั้งที่สุ่ม `m` จำนวนของชิ้นที่ "ดี" หรือที่สนใจ `N` จำนวนชิ้นทั้งหมด	`dHyperDev(n, m, N)`
ปัวซอง (Poisson)	`m` ค่าเฉลี่ยจำนวนครั้งที่เหตุการณ์เกิดขึ้นใน 1 ช่วงเวลา	`dPoissonDev(m)`

ตารางที่ 3.4: สูตรพ็อพทูลสำหรับตัวแปรสุ่มแบบต่อเนื่อง

การแจกแจงความน่าจะเป็น	พารามิเตอร์	สูตรพ็อพทูล
ปกติ ล็อกนอร์มอล (Lognormal)	μ ค่าเฉลี่ย σ ส่วนเบี่ยงเบนมาตรฐาน	`dNormalDev(`μ, σ`)` `dLogNormalDev(`μ, σ`)`
สม่ำเสมอ (Uniform)	`a` ขอบเขตล่าง `b` ขอบเขตบน โดยที่ `a` < `b`	`dRandReal(a, b)`
เอ็กซ์โพเนนเชียล (Exponential)	μ ค่าเฉลี่ย	`dExpDev(`μ`)`
เบต้า (Beta)	α, β ค่าพารามิเตอร์ระบุรูปร่างของ PDF	`dBetaDev(`α, β`)`
สามเหลี่ยม (Triangular)	`a` ขอบเขตล่าง `b` ค่าที่เป็นไปได้มากที่สุด `c` ขอบเขตบน	`dTRand(a, b, c)`
ไวบูล (Weibull)	λ ค่าพารามิเตอร์ระบุรูปร่างของ PDF k ค่าพารามิเตอร์ระบุสเกลของ PDF	`dWeibullDev(`λ, k`)`

3.2.2 การจำลองการวางแผนงบประมาณของบริษัทมายา

จากข้อมูลที่บริษัทมายารวบรวมไว้ ผู้วิเคราะห์พบว่า

- จำนวนของพนักงานเพิ่มขึ้นด้วยอัตราที่มีการแจกแจงแบบสม่ำเสมอ มีค่าต่ำที่สุดคือลดลงอย่างมากที่สุด 3% หรือเพิ่มขึ้นสูงสุด 7%
- ค่าธรรมเนียมการศึกษาที่พนักงานเบิกต่อคนมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเพิ่มขึ้นด้วยอัตรา 1% ต่อปี และมีส่วนเบี่ยงเบนมาตรฐาน 124 บาท

แบบจำลองในตารางที่ 3.5 แสดงแบบจำลองที่เพิ่มข้อมูลนี้แล้ว เนื่องจากฟังก์ชัน RVG ของพ็อพทูลปรากฏอยู่ในสูตรที่มีการคำนวณทางคณิตศาสตร์อื่น ๆ ด้วย ผู้ใช้จึงต้องพิมพ์สูตร RVG เข้าไปในเซลล์เอง เมื่อพัฒนาแบบจำลองแล้ว ควรตรวจดูความสมจริงของค่าที่ปรากฏในแบบจำลองเอกซ์เซล เช่น ความน่าจะเป็นต้องไม่ติดลบและตัวเลขต่าง ๆ มีค่าสมจริง ไล่เรียงไปทีละเซลล์ว่ากำหนดสูตรถูกต้องตามที่คิดแล้ว

เมื่อได้แบบจำลองแล้ว ผู้ใช้ต้องระบุให้พ็อพทูลรู้ว่าเซลล์ใดเป็นค่าเอาท์พุทหรือดัชนีชี้วัด พร้อมทั้งกำหนดค่าพารามิเตอร์ในการประมวลผลอื่น ๆ ให้เลือกคำสั่งจากเมนู `Add-ins` > `PopTools` > `Simulation tools` > `Monte Carlo analysis` แล้วไดอะล็อกบ็อกซ์ (Dialog box) ในรูปที่ 3.1 จะปรากฏ มีหลายช่องให้ระบุค่า แต่จะอธิบายเฉพาะช่องที่จำเป็นต้องใช้ ดังนี้

ตารางที่ 3.5: แบบจำลองสถานการณ์บริษัทมายาที่ปรับจากตารางที่ 3.1

	A	B	C	D	E	F	G	H	I
1									
2			**บริษัทมายา จำกัด**						
3									
4	**เงื่อนไขเริ่มต้น**				**ข้อสมมุติ**				
5	จำนวนพนักงานในปัจจุบัน			63	อัตราที่ลดลงสูงสุด	3%	อัตราที่เพิ่มสูงสุด	7%	การแจกแจงแบบสม่ำเสมอ
6	ค่าธรรมเนียมการศึกษาเฉลี่ยที่พนักงานเบิก (ต่อคน)			16,541	อัตราที่เพิ่มขึ้น	1.0%	ส่วนเบี่ยงเบนมาตรฐาน	124	การแจกแจงแบบปกติ
7									
8		จำนวน	ค่าธรรมเนียมที่เบิก	ค่าใช้จ่าย					
9	ปีที่	พนักงาน	ต่อพนักงาน 1 คน	รวม					
10	1	64	16,685.70	1,069,906.02					
11	2	68	16,918.56	1,111,065.95					
12	3	65	16,925.91	1,106,118.12					
13	4	65	17,031.50	1,107,288.19					
14	5	68	17,200.40	1,168,504.70					
15			**ค่าใช้จ่ายรวมของบริษัท**	**5,562,882.98**					

ตารางที่ 3.6: สูตรเอกซ์เซลสำหรับแบบจำลองสถานการณ์บริษัทมายา (ตารางที่ 3.5)

ตำแหน่งเซลล์	**สูตรเอกซ์เซล**	**คัดลอกไปยังเซลล์**
B10	`=D5*(1+dRandReal(-$F$5, $H$5))`	
C10	`=dNormalDev(D6*(1+$F$6), $H$6)`	
D10	`=C10*B10`	D11:D14
B11	`=B10*(1+dRandReal(-$F$5, $H$5))`	B12:B14
C11	`=dNormalDev(C10*(1+$F$6), $H$6)`	C12:C14
C15	`=sum(D10:D14)`	

- `Dependent range` ตำแหน่งเซลล์ของเอาท์พุทหรือดัชนีชี้วัด อาจมีได้มากกว่า 1 เซลล์ แต่หากมากกว่า 1 เซลล์ เอาท์พุทต้องเรียงชิดกันในแนวแถวหรือคอลัมน์ คืออยู่ในรูปอาร์เรย์

- `Number of replicates` จำนวนรอบทำซ้ำ เป็นจำนวนครั้งที่แบบจำลองเอกซ์เซลนี้ได้รับการประมวลผล คล้ายกับทุกครั้งที่มีการกด `Enter` หรือกด `Calculate now` เอกซ์เซลประมวลผลใหม่ทั้งเวิร์คชีต แต่ละครั้งที่ประมวลผล เซลล์ที่มีสูตรคณิตศาสตร์จะถูกคำนวณใหม่ นั่นคือเซลล์ที่มีฟังก์ชัน RVG ก็จะผลิตค่าเลขสุ่มใหม่เช่นกัน พ็อพทูลเก็บค่าเอาท์พุทที่คำนวณใหม่นี้ไว้ทุกครั้งเพื่อคำนวณค่าสรุปทางสถิติ และแสดงผลเมื่อการประมวลผลเสร็จสิ้นตามจำนวนรอบที่ผู้ใช้ระบุ ค่าโดยปริยาย (Default value) คือ 100 รอบ

- `Output` ตำแหน่งของเซลล์มุมบนซ้ายสุดที่ผู้ใช้ต้องการให้วางตารางผลสรุปที่ได้จากการประมวลผล เช่น ตารางที่ 3.7ก

- `Keep results` หากผู้ใช้คลิกเลือกตัวเลือกนี้ พ็อพทูลจะแสดงค่าของเอาท์พุทที่ทุกรอบการประมวลผลบนอีกแผ่นเวิร์คชีตหนึ่ง เช่น ตารางที่ 3.7ข

เมื่อกำหนดค่าพารามิเตอร์เหล่านี้แล้ว คลิก `Go` หากแบบจำลองไม่มีข้อผิดพลาด จะได้ตารางผลสรุปดังเช่นรูปที่ 3.7ก หากคลิกเลือกตัวเลือก `Keep results` จะได้เวิร์คชีตอีกหนึ่งแผ่นเพิ่มด้วย ดังแสดงในรูปที่ 3.7ข

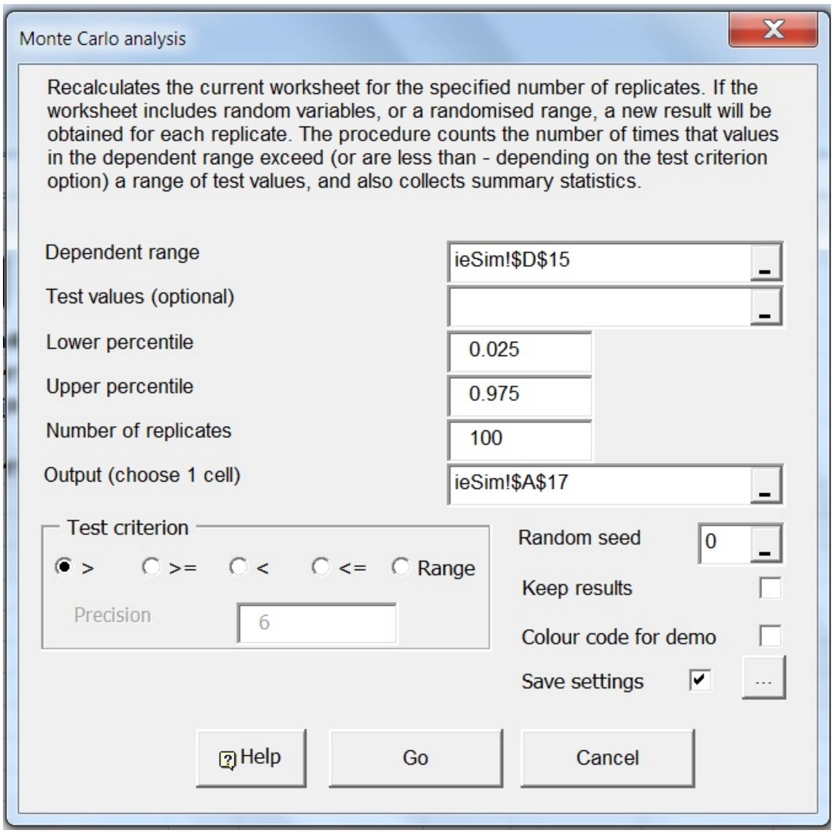
ตารางที่ 3.7: ผลการประมวลผลแบบจำลองการวางแผนงบประมาณบริษัทมายา (ตารางที่ 3.5)

(ก) ค่าสรุปทางสถิติ

Mean	Variance	Lower CL	Upper CL	Valid Iterations	Time taken
5665292	55133160018	5160482.121	614215.196	100	0

(ข) ผลของแต่ละรอบทำซ้ำ

Replicate	Var 1
1	5344870
2	6092414
⋮	⋮
98	5950705
99	5230595
100	5296585



รูปที่ 3.1: ไดอะล็อกบ็อกซ์สำหรับกำหนดพารามิเตอร์เพื่อประมวลผล

อ้างอิงจากแบบจำลองในตารางที่ 3.5 สมมติให้ X แทนค่าใช้จ่ายรวมของบริษัท ค่าของผลในตารางที่ 3.7ก สามารถอธิบายได้ดังนี้

- `Mean` ค่าเฉลี่ยเซลล์ที่นิยามว่าเป็น `Dependent range` ซึ่งมีขนาดข้อมูลเท่ากับจำนวนรอบทำซ้ำที่ผู้ใช้กำหนด ในตารางที่ 3.5 ค่าเอาท์พุทคือค่าใช้จ่ายรวมของบริษัท (เซลล์ `D15`) ตัวเลข 5,665,292 บาทเป็นค่าเฉลี่ยของค่าใช้จ่ายที่ประมวลผลได้ 100 ค่า (แสดงในตารางที่ 3.7ข) ด้วยสูตร (2.16) ดังนี้

$$\bar{X} = 5.66 \times 10^6 = \left(\frac{5.34 + 6.09 + \ldots + 5.23 + 5.30}{100}\right) \times 10^6 \qquad (3.2)$$

- `Variance` ความแปรปรวนที่คำนวณจากเอาท์พุทที่ได้จากการประมวลผล คำนวณโดยใช้สูตร (2.17) ดังนั้นส่วนเบี่ยงเบนมาตรฐานจึงเป็น

$$S_X = 2.34 \times 10^5 = \sqrt{5.51 \times 10^{10}} \qquad (3.3)$$

- `Lower CL` CL ย่อมาจาก Confidence limit ให้ค่าเปอร์เซ็นไทล์ตามนัยสำคัญที่ระบุในช่อง `Lower percentile` (มีค่าอยู่ในช่วง 0 ถึง 1) ในไดอะล็อกบ็อกซ์รูปที่ 3.1 ค่าโดยปริยายคือ 0.025 จากผลในตารางที่ 3.7ก ค่า Lower percentile คือ 5,160,482 บาท หมายความว่า

$$\Pr\{X \leqslant 5{,}160{,}482\} = 0.025$$

นั่นคือ มีโอกาสเพียงแค่ 0.025 หรือ 2.5% ที่ค่าใช้จ่ายจะน้อยกว่า 5.16 ล้านบาท และมีความน่าจะเป็น 1-0.025 = 0.975 หรือ 97.5% ที่ค่าใช้จ่ายจะมากกว่าค่านี้ สามารถเขียนในรูปทั่วไปได้ดังนี้

$$\Pr\{\texttt{Dependent range} \leqslant \texttt{Lower CL}\} = \texttt{Lower percentile} \qquad (3.4)$$

สังเกตว่า `Lower CL` เป็นผลจากการประมวลผล ส่วน `Lower percentile` เป็นพารามิเตอร์นำเข้าที่ผู้ใช้กำหนด

- `Upper CL` เป็นเปอร์เซ็นไทล์เช่นเดียวกับ Lower CL แต่ `Upper percentile` มีค่าโดยปริยายคือ 0.975 ในตารางที่ 3.7ก ค่า Upper percentile คือ 6,141,215 บาท ซึ่งก็หมายความว่า

$$\Pr\{X \leqslant 6{,}141{,}215\} = 0.975$$

นั่นคือ มีโอกาสเพียงแค่ 2.5% ที่ค่าใช้จ่ายจะมากกว่า 6.14 ล้านบาท หรือเขียนในรูปแบบทั่วไปได้ดังนี้

$$\Pr\{\texttt{Dependent range} \leqslant \texttt{Upper CL}\} = \texttt{Upper percentile} \qquad (3.5)$$

ทั้งนี้ ค่าเปอร์เซ็นไทล์ที่ได้จากพ็อพทูลอาจต่างจากที่ได้จากสูตรเอกซ์เซลที่ผู้ใช้คำนวณเองจากเอาท์พุทที่ทุกรอบการประมวลผลเล็กน้อย เพราะความผิดพลาดจากการปัด (Roundoff error)

- `Valid iterations` จำนวนรอบทำซ้ำที่ประมวลผลเสร็จ
- `Time taken` เวลาที่ใช้ในการประมวลผล

พ็อพทูลรายงานค่าสรุปทางสถิติ ซึ่งผู้วิเคราะห์สามารถตีความและวิเคราะห์เพิ่มเติมได้

3.2.3 การตีความผล

เมื่อการประมวลผลเสร็จสิ้น ผู้ใช้สามารถคำนวณค่าทางสถิติเพิ่มเติมจากที่พ็อพทูลรายงาน หรือวิเคราะห์ผลที่ได้ด้วยกราฟ บทที่ 5 จะอธิบายเรื่องการวิเคราะห์ผลลัพธ์อย่างละเอียด

- **ช่วงความเชื่อมั่น** (Confidence interval) มีสูตรการคำนวณจากค่าเฉลี่ย $\bar{X}$ และส่วนเบี่ยงเบนมาตรฐาน S_X ที่ได้จากการประมวลผลดังสมการที่ (3.2) และ (3.3) ดังนี้

$$
\begin{aligned}
\texttt{Lower bound} &= \bar{X} - t_{1-\alpha/2,R-1}\frac{S_X}{\sqrt{R}} && (3.6)\\
\texttt{Upper bound} &= \bar{X} + t_{1-\alpha/2,R-1}\frac{S_X}{\sqrt{R}} && (3.7)\\
(\texttt{Lower bound, Upper bound}) &\leftarrow \bar{X} \pm \texttt{Half width} && (3.8)
\end{aligned}
$$

ความกว้างของครึ่งช่วง (Half width) คือ $t_{1-\alpha/2,R-1}S_X/\sqrt{R}$ ซึ่งเป็นครึ่งหนึ่งของความกว้างของช่วงความเชื่อมั่น ค่าสถิติ $t_{1-\alpha/2,R-1}$ เป็นค่าเปอร์เซ็นไทล์ของการแจกแจง Student t ที่ความน่าจะเป็น $1-\alpha/2$ และองศาอิสระ (Degree of freedom) เท่ากับ $R-1$ โดยที่ R คือขนาดของข้อมูลซึ่งก็คือจำนวนรอบทำซ้ำที่ประมวลผล หากต้องการช่วงความเชื่อมั่นที่นัยสำคัญ $(1-\alpha)\times 100$ =95% โอกาสที่จะผิดพลาดคือช่วงความเชื่อมั่นไม่ครอบค่าเฉลี่ยที่แท้จริงคือ 5% ซึ่งอาจจะผิดด้านที่มากเกินไป (หางขวา) หรือน้อยเกินไป (หางซ้าย) ดังนั้นความผิดพลาดของหางข้างใดข้างหนึ่ง คือ 5%/2 จึงใช้ค่าเปอร์เซ็นไทล์ของ Student t ที่ความน่าจะเป็น $1-\alpha/2$ ซึ่งสามารถค้นได้จากตารางสถิติท้ายเล่มของหนังสือเรียนสถิติทั่วไปหรือจากอินเตอร์เน็ต แต่วิธีที่แนะนำคือใช้สูตรเอกซ์เซล ดังนี้

$$t_{1-\alpha/2,R-1} \leftarrow \texttt{tinv}(\alpha,\texttt{R-1})$$

เพราะเอกซ์เซลต้องการให้ระบุโอกาสที่จะผิดพลาด**ทั้งหมด** จึงใส่ α สูตร (3.8) ให้คำนวณความกว้างของครึ่งช่วงก่อน แล้วจึงนำไปหักออกหรือเติมเข้าไปกับค่าเฉลี่ย ตัวอย่างที่ 3.2 แสดงการคำนวณช่วงความเชื่อมั่นสำหรับผลการประมวลผลในตารางที่ 3.7ก

ตัวอย่างที่ 3.2. กำหนดให้ช่วงความเชื่อมั่นมีนัยสำคัญ 95% ดังนั้น $\alpha = 5\%$ จากตารางที่ 3.7ก $R =$

100 แทนค่าในสมการที่ (3.6)-(3.7) จะได้ (หน่วย: ล้านบาท)

$$\texttt{Half width} = 1.984\left(\frac{0.235}{\sqrt{100}}\right) = 0.0466 \text{ ล้านบาท} \tag{3.9}$$

$$\texttt{Lower bound} = 5.665 - 0.0466 = 5.619 \text{ ล้านบาท} \tag{3.10}$$

$$\texttt{Upper bound} = 5.665 + 0.0466 = 5.712 \text{ ล้านบาท} \tag{3.11}$$

การแปลความหมายของช่วงความเชื่อมั่นอย่างคร่าว ๆ คือ โอกาสที่**ค่าคาดหมาย**ของค่าใช้จ่ายทุนการศึกษารวมจะมีค่ามากกว่า 5.619 ล้านบาท แต่ไม่เกิน 5.712 ล้านบาทคือ 95% หัวข้อที่ 5.3 จะกล่าวถึงความหมายที่ถูกต้องทางทฤษฎีกว่านี้

□

- **เปอร์เซ็นไทล์** จากสมการที่ (3.4)-(3.5) จะได้ว่า

$$\Pr\{\texttt{Lower CL} \leqslant \texttt{Dependent range} \leqslant \texttt{Upper CL}\} = \texttt{Upper percentile} - \texttt{Lower percentile} \tag{3.12}$$

ค่าเปอร์เซ็นไทล์สามารถให้ช่วงที่**ค่าดัชนีชี้วัด**ที่สังเกตได้จริงจะตกอยู่ในช่วงนี้ด้วยความน่าจะเป็นสูง เช่น 95% ซึ่งเป็นค่าโดยปริยายของพ็อพทูล

ตัวอย่างที่ 3.3. สำหรับผลการประมวลผลในตารางที่ 3.7ก ช่วงเปอร์เซ็นไทล์จากสมการที่ (3.12) คือ [5.160, 6.141] ล้านบาท หมายความว่าโอกาสที่ค่าใช้จ่ายทุนการศึกษารวมมีค่ามากกว่า 5.160 ล้านบาท แต่ไม่เกิน 6.141 ล้านบาทคือ 95% สังเกตว่าในตัวอย่างนี้และโดยทั่วไป ช่วงเปอร์เซ็นไทล์ (สมการที่ 3.12) กว้างกว่าช่วงความเชื่อมั่น (สมการที่ 3.8) มาก เพราะช่วงเปอร์เซ็นไทล์กล่าวถึงค่าที่เป็นไปได้ของ**ตัวแปรสุ่ม**ซึ่งเป็นค่าที่ยังไม่เกิดขึ้น แต่ช่วงความเชื่อมั่นกล่าวถึงค่าที่เป็นไปได้ของค่าเฉลี่ยซึ่งเป็นค่าคงที่ที่เราไม่ทราบค่า

□

- **ความน่าจะเป็นที่จะสูงหรือต่ำกว่าเป้าหมาย** เมื่อคำนวณเปอร์เซ็นไทล์ ผู้ใช้กำหนดค่าความน่าจะเป็น ณ ขณะนี้ จะทำตรงกันข้าม กล่าวคือจะกำหนดค่าที่เป็นเป้าหมาย แล้วจึงคำนวณโอกาสที่ค่าดัชนีชี้วัดจะมากหรือน้อยกว่าค่านี้ ผู้ใช้กำหนด c ที่เป็นเป้าหมาย สมมติให้ X เป็นดัชนีชี้วัด ต้องการคำนวณ p_c ที่ทำให้สมการที่ (3.13) ต่อไปนี้เป็นจริง

$$\Pr\{X \leqslant c\} = p_c \tag{3.13}$$

ความน่าจะเป็นนี้จะประมาณจากผลที่พ็อพทูลรายงานสำหรับทุกรอบทำซ้ำ เช่น ในตารางที่ 3.7ข

เอกซ์เซลมีฟังก์ชันสำหรับประมาณค่า p_c คือ

$$\Pr\{X \leqslant c\} \leftarrow \texttt{PERCENTRANK(dataArray, c)} \tag{3.14}$$

ตัวอย่างที่ 3.4. สำหรับผลในตารางที่ 3.7ข ผลจากการประมวลผลอยู่ในอาร์เรย์ B2:B101 สมมติว่าบริษัทมายามีงบประมาณสำหรับทุนนี้ 6 ล้านบาท (รวมทั้ง 5 ปี) ผู้บริหารต้องการทราบโอกาสที่ทุนการศึกษาที่จ่ายจริงจะเกินงบประมาณ ดังนั้น c = 6 ล้านบาทและเมื่อใช้สูตรเอกซ์เซล PERCENTRANK(B2:B101, $6*10^6$) จะได้ 0.912 หมายความว่าโอกาสที่จะเกินงบฯ คือ 1-0.912 หรือ 8.8%

□

- *ฮีสโตแกรม* เพื่อแสดงการกระจายตัวของผลเอาท์พุทจากการประมวลผล อ่านเพิ่มเติมได้ในหัวข้อที่ 2.5.2

ถึงแม้จะเป็นปัญหาเดียวกันและผลการคำนวณจะไม่แตกต่างอย่างมีนัยสำคัญ ผู้พัฒนาแต่ละคนก็อาจจะได้แบบจำลองที่แตกต่างกันในรูปแบบหรือรายละเอียด เพราะการจำลองไม่ใช่วิทยาศาสตร์ทั้งหมดแต่เป็นงานศิลปะด้วย

3.2.4 สิ่งที่ปรับแต่งได้ในแบบจำลองสเปรดชีต

เอกซ์เซลและพ็อพทูลเปิดโอกาสให้ผู้ใช้สามารถปรับรายละเอียดของแบบจำลองได้ เช่น การนิยามชื่อให้ตัวแปรในเซลล์ หรือการใช้พื้นสีสำหรับเซลล์ต่างกัน

การนิยามตัวแปรในเซลล์

เอกซ์เซลให้ผู้ใช้ตั้งชื่อเซลล์หรือบล็อกของเซลล์ การนิยามเซลล์ด้วยชื่อมีประโยชน์ตรงที่ทำให้แบบจำลองเข้าใจง่าย และสามารถหาตำแหน่งของตัวแปรเหล่านี้ได้สะดวก เช่น หากนิยามเซลล์ F5 ว่า `MaxDecrease` และเรียกเซลล์ H5 ว่า `MaxIncrease` ในเซลล์ B11 จะเห็นเป็นสูตร `=B10*(1+dRandReal(-maxDecrease,maxIncrease))` ซึ่งช่วยให้เข้าใจง่าย และหากว่าต้องการหาตำแหน่งตัวแปรที่ชื่อ `maxDecrease` ก็สามารถเลือกได้จากเมนูดึงลง (Pull-down menu) ด้านบนซ้ายมือ

ผู้ใช้สามารถนิยามชื่อตัวแปรในเอกซ์เซลได้ 2 วิธี ดังนี้

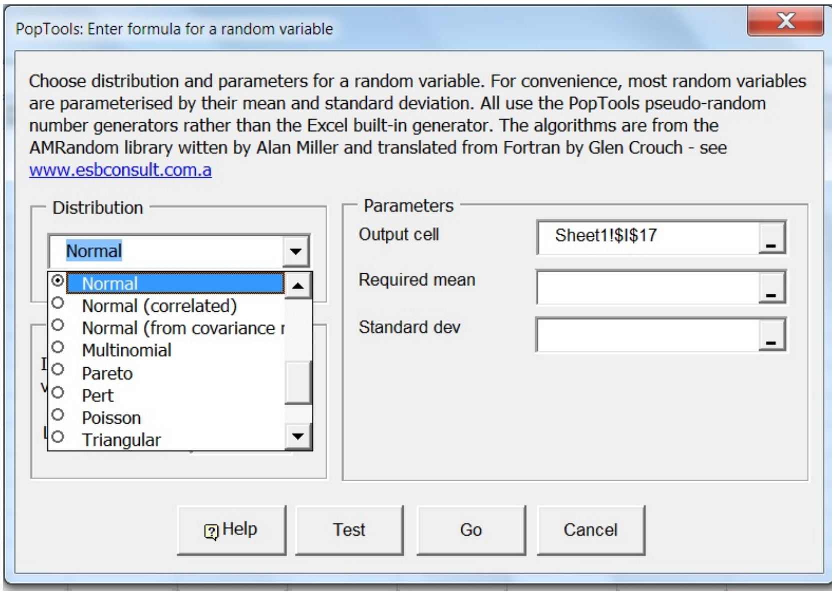
1. คลิกเลือกเซลล์หรืออาร์เรย์ของเซลล์ที่ต้องการนิยาม พิมพ์ชื่อตัวแปรใน `Name box` แล้วกด `Enter`

2. เลือก `Formula` > `Define names` แล้วไดอะล็อกบ็อกซ์จะปรากฏ ให้ใส่ชื่อที่ต้องการในกล่อง `Name:`

หากฟังก์ชัน RVG ปรากฏเดี่ยว ๆ ในเซลล์ ไม่ฝังตัวอยู่ในสูตรคณิตศาสตร์อื่น ก็สามารถใช้ไดอะล็อกบ็อกซ์เพื่อนิยามตัวสร้างตัวแปรสุ่มได้

การนิยามฟังก์ชัน RVG ด้วยไดอะล็อกบ็อกซ์

บางครั้งผู้ใช้อาจจำชื่อคำสั่งหรือวิธีการระบุพารามิเตอร์ไม่ได้ ฟังก์ชัน RVG สามารถนิยามด้วยไดอะล็อกบ็อกซ์ โดยเลือก `Add-Ins > PopTools > Random variable` และจะได้ไดอะล็อกบ็อกซ์ดังแสดงในรูปที่ 3.2 `Output cell` คือเซลล์ที่เราเลือกที่จะวางตัวแปรสุ่ม อาจเป็นเซลล์เดียวหรือเป็นบล็อกของเซลล์ที่ติดกัน เมื่อเลือกการแจกแจงความน่าจะเป็นจากเมนูดึงลงทางซ้าย กล่องใส่พารามิเตอร์ที่เกี่ยวข้องจึงปรากฏด้านขวาซึ่งอาจเป็นค่าตัวเลขหรือตำแหน่งเซลล์ที่บรรจุค่า เมื่อใส่ค่าพารามิเตอร์แล้ว จึงกด `Go`



รูปที่ 3.2: ไดอะล็อกบ็อกซ์ของพ็อพทูลเพื่อกำหนดการแจกแจงสำหรับตัวแปรสุ่ม

หัวข้อที่ 3.2 ได้นำเสนอการปรับปรุงการวิเคราะห์ในเอกซ์เซลโดยเพิ่มข้อมูลความไม่แน่นอนของตัวแปรนำเข้า เพื่อให้ได้กลุ่มตัวอย่างของเอาท์พุทซึ่งสามารถนำไปวิเคราะห์ความเสี่ยงได้ ส่วนที่เหลือของบทนี้จะกล่าวถึงแบบจำลองสเปรดชีตสำหรับโจทย์อื่น ๆ เพื่อให้ผู้อ่านเรียนรู้ผ่านตัวอย่างและได้เห็นความหลากหลายของแบบจำลองที่สามารถสร้างบนสเปรดชีตได้ ตัวอย่างในหัวข้อที่ 3.3-3.5 ต่อไปนี้มีลักษณะร่วมกันคือมีตัวแปรตัดสินใจ (Decision variable) ซึ่งผู้วิเคราะห์ต้องกำหนดก่อนที่จะประมวลผลแบบจำลอง เช่น จุดสั่งซื้อและยอดสั่งซื้อ (หัวข้อที่ 3.4) และมีเอาท์พุทเป็นค่าดัชนีชี้วัดที่ต้องการให้เป็นค่าที่มากที่สุดหรือน้อยที่สุด เช่น ต้องการให้ระดับสินค้าคงคลังเฉลี่ยน้อยที่สุดที่ยังได้ระดับการบริการที่เหมาะสม (หัวข้อที่ 3.4)

3.3 ปัญหาเด็กส่งหนังสือพิมพ์

ปัญหาเด็กส่งหนังสือพิมพ์ (Newsboy problem หรือ Newsvendor problem) เป็นกลุ่มของปัญหาด้านการวิจัยดำเนินงานที่มีลักษณะเฉพาะ (อ่านรายละเอียดเพิ่มเติมใน Kalvelagen 2003) ปัญหาดั้งเดิมคือตอนเช้ามืดหรือก่อนเปิดร้าน เด็กส่งหนังสือพิมพ์จะตัดสินใจว่าจะซื้อหนังสือพิมพ์กี่ฉบับสำหรับขายในวันนั้น หากซื้อมากเกินไปก็จะเหลือซึ่งขายคืนเป็นเศษกระดาษได้ในราคาที่ต่ำกว่าราคาทุนมาก ในทางตรงกันข้ามหากซื้อมาน้อยเกินไปก็จะไม่พอขายและเสียโอกาสในการทำรายได้ หนังสือพิมพ์มีอุปสงค์ที่ไม่แน่นอนจึงถือว่าเป็นค่าสุ่ม การเสียโอกาสทำรายได้และขาดทุนจากการซื้อมากเกินไปนี้ถ่วงดุลกัน วัตถุประสงค์ของปัญหาคือกำหนดยอดสั่งซื้อหนังสือพิมพ์ที่ทำให้ผลกำไร (รายรับ - รายจ่าย) เฉลี่ยสูงที่สุด ปัญหานี้สามารถประยุกต์ใช้กับสินค้าหรือผลิตภัณฑ์ที่มีอายุขัย เช่น ผลิตภัณฑ์ทางการเกษตร อาหาร นิตยสาร ปฏิทินประจำปี เสื้อผ้าแฟชั่น หรือแม้กระทั่งตั๋วเครื่องบิน ดังแสดงด้วยตัวอย่างต่อไปนี้

ตัวอย่างระบบจองตั๋วเครื่องบิน

ตัวอย่างนี้ดัดแปลงจาก Pietmont Commuter Airlines ใน Ragsdale (2008) Thai Commuter Airlines (TCA) เป็นสายการบินขนาดเล็กที่ให้บริการระหว่างชัยภูมิและอุบลราชธานี TCA มีเครื่องบินขนาดเล็กจำนวน 19 ที่นั่ง และมักจะมีที่นั่งว่างหลายที่ ราคาตั๋ว 4,500 บาทต่อที่นั่ง โอกาสที่ผู้โดยสารแต่ละคนที่ซื้อตั๋วไปแล้วแต่ไม่มาขึ้นเครื่องคือ 10% ถ้า TCA ขายตั๋วจำนวนมากกว่าที่นั่งที่มีและมีผู้โดยสารมาขึ้นเครื่องมากกว่าจำนวนที่นั่งบนเครื่องบิน สายการบินจะต้องออกตั๋วชดเชยให้กับลูกค้าหรือหาที่พักให้ ซึ่งมีค่าใช้จ่าย 9,750 บาทต่อใบ อุปสงค์ของลูกค้าในแต่ละเที่ยวบินไม่แน่นอน ข้อมูลที่เก็บในอดีตแสดงว่ามีการแจกแจงดังแสดงในตารางที่ 3.8 ผู้บริหาร TCA ต้องการทราบว่าควรจะเปิดขายตั๋วกี่ใบต่อหนึ่งเที่ยวบิน เพื่อให้ผลกำไรสุทธิเฉลี่ยสูงที่สุด

ตารางที่ 3.8: การแจกแจงของอุปสงค์ของจำนวนที่นั่งสำหรับตัวอย่างการจัดการระบบจองตั๋วเครื่องบิน

อุปสงค์	14	15	16	17	18	19	20	21	22	23	24	25
Probability	0.03	0.05	0.07	0.09	0.11	0.15	0.18	0.14	0.08	0.05	0.03	0.02

ตัวอย่างนี้มีปัจจัยสองด้านถ่วงดุลกันอยู่ เนื่องจากมีผู้โดยสารที่อาจตกเครื่อง TCA จึงควรเปิดขายตั๋วมากกว่าจำนวนที่นั่งที่มีเพื่อเพิ่มผลกำไร แต่หากขายตั๋วมากเกินไป จะต้องชดเชยให้ลูกค้าด้วยค่าใช้จ่ายต่อคนที่มากกว่าราคาตั๋วที่ขายได้ นอกจากนี้ยังมีค่าเสียความรู้สึกของลูกค้าที่วัดเป็นตัวเงินไม่ได้อีก ตารางที่ 3.9 แสดงแบบจำลองสถานการณ์สำหรับปัญหานี้และตารางที่ 3.10 แสดงสูตรพ็อพทูลหรือสูตรเอกซ์เซลที่ใช้ สังเกตว่าในเซลล์ B4 ถึง B7 เป็นพารามิเตอร์นำเข้า แต่เซลล์ B8 เป็นตัวแปรตัดสินใจที่ผู้วิเคราะห์เป็นผู้กำหนด

เนื่องจากจำนวนที่นั่งบนเครื่องบินมี 19 ที่นั่งและมีผู้โดยสารบางคนตกเครื่อง จำนวนตั๋วที่เปิดให้จองจึงควรมากกว่า 19 ใบ เซลล์ B10 สุ่มจากการแจกแจงความน่าจะเป็นที่ระบุในตารางในเซลล์ D5:E16 จำนวนตั๋วที่ได้ขาย (เซลล์ B11) ขึ้นอยู่กับอุปสงค์ของลูกค้าและโควต้าที่เปิดให้ซื้อ

แบบจำลองนี้ใช้สูตรเอกซ์เซล =MIN(a,b) และ =MAX(a,b) ซึ่งให้ผลเป็นค่าที่น้อยกว่าและค่าที่มากกว่าตามลำดับ โดยอาจเปรียบเทียบได้มากกว่า 2 ค่าและใช้จุลภาค (,) เพื่อแยกค่าที่ระบุหรือแยกเซลล์ สามารถระบุเป็นอาร์เรย์ เช่น =MAX(B1:B5, 0) ซึ่งจะให้ค่าที่มากกว่าหรือเท่ากับ 0 ในอาร์เรย์ B1:B5

จำนวนผู้โดยสารที่มาขึ้นเครื่อง (เซลล์ B12) เป็นค่าสุ่มแบบทวินามที่ขึ้นอยู่กับจำนวนตั๋วที่ขายและความน่าจะเป็นที่ลูกค้าแต่ละคนจะตกเครื่อง รายได้จากการขายตั๋ว (เซลล์ B14) แปรผันตรงกับราคาหน้าตั๋วและจำนวนตั๋วที่ได้ขาย; เมื่อหักจากค่าใช้จ่ายของการเปลี่ยนเที่ยวบินให้ลูกค้าที่ไม่ได้ขึ้นเครื่อง (เซลล์ B15, หากมี) แล้วก็จะได้กำไรสุทธิ (เซลล์ B16) ซึ่งเป็นเอาท์พุท อีกดัชนีชี้วัดหนึ่งที่น่าสนใจคือโอกาสที่ไฟลท์จะเต็มจนต้องจัดให้ลูกค้าไปเที่ยวบินอื่น นิยามไว้ในเซลล์ B17 ซึ่งแสดงค่า 1 หากไฟลท์ล้น และ 0 หากไม่ล้น

ตารางที่ 3.9: แบบจำลองตัวอย่างจองตั๋วเครื่องบิน (หัวข้อที่ 3.3)

	A	B	C	D	E
1					
2	Thai Commuter Airlines				
3					
4	จำนวนที่นั่งบนเครื่องบิน	19		Demand	Probability
5	ราคาตั๋วต่อที่นั่ง	฿4,500		14	0.03
6	ความน่าจะเป็นที่ลูกค้าไม่ขึ้นเครื่อง	0.10		15	0.05
7	ค่าใช้จ่ายในการเปลี่ยนไฟลท์ลูกค้า	฿9,750		16	0.07
8	จำนวนตั๋วที่ให้จอง	23		17	0.09
9				18	0.11
10	จำนวนที่นั่งที่ลูกค้าต้องการ	16		19	0.15
11	จำนวนตั๋วที่ได้ขาย	16		20	0.18
12	จำนวนผู้โดยสารที่ขึ้นเครื่อง	15		21	0.14
13				22	0.08
14	รายได้จากการขายตั๋ว	฿72,000		23	0.05
15	ค่าใช้จ่ายจากการเปลี่ยนไฟลท์ลูกค้า	฿ -		24	0.03
16	**กำไรสุทธิ**	**฿72,000**		25	0.02
17	**ถ้าไฟลท์ล้นได้ 1 และ 0 ถ้าเป็นอื่น**	**0.00**			

ตารางที่ 3.10: สูตรเอกซ์เซลในแบบจำลองระบบจองตั๋วเครื่องบิน (ตารางที่ 3.9)

ตำแหน่งเซลล์	**สูตรเอกซ์เซล**
B10	=DiscreteDev(D5:D16,E5:E16)
B11	=MIN(B10,B8)
B12	=dBinomialDev(B11,1-B6)
B14	=B11*B5
B15	=MAX(B12-B4,0)*B7
B16	=B14-B15
B17	=IF(B15>0,1,0)

ตัวอย่างนี้มีเอาท์พุท 2 ตัวซึ่ง**ต้องวางไว้ทับซ้อนกัน** (Stacked) ในเวิร์คชีตเพื่อระบุในพ็อพทูลว่าเป็น `Dependent range` คือ `$B$16:$B$17` ตารางที่ 3.11ก แสดงค่าสรุปผลการประมวลผล จะเห็นว่ามีสองแถว แถวแรกเป็นผลสำหรับเอาท์พุทตัวบนสุดในอาร์เรย์ของ `Dependent range` ซึ่งก็คือค่าเฉลี่ยผลกำไร และแถวที่สองเป็นผลสำหรับเอาท์พุทตัวที่สองซึ่งก็คือความน่าจะเป็นที่จะต้องจัดลูกค้าไปไฟลท์อื่น มีค่าประมาณ 17% สำหรับโควต้าตั๋ว 23 ใบ ส่วนผลอย่างละเอียดของแต่ละรอบการประมวลผลก็จะมี 2 คอลัมน์เช่นกัน ในรูปที่ 3.11ข คอลัมน์แรกเป็นหมายเลขรอบทำซ้ำ คอลัมน์ที่สองที่มีชื่อว่า `Var 1` เป็นเอาท์พุทตัวแรก (ผลกำไร) และคอลัมน์ที่สามแสดงหัวด้วย `Var 2` เป็นเอาท์พุทตัวที่สอง (ผลลัพธ์ของเหตุการณ์ต้องจัดลูกค้าไปไฟลท์อื่น)

ตารางที่ 3.11: ผลการประมวลผลตัวอย่างระบบจองตั๋วเครื่องบิน (ตารางที่ 3.9)

(ก) ค่าสรุปทางสถิติ

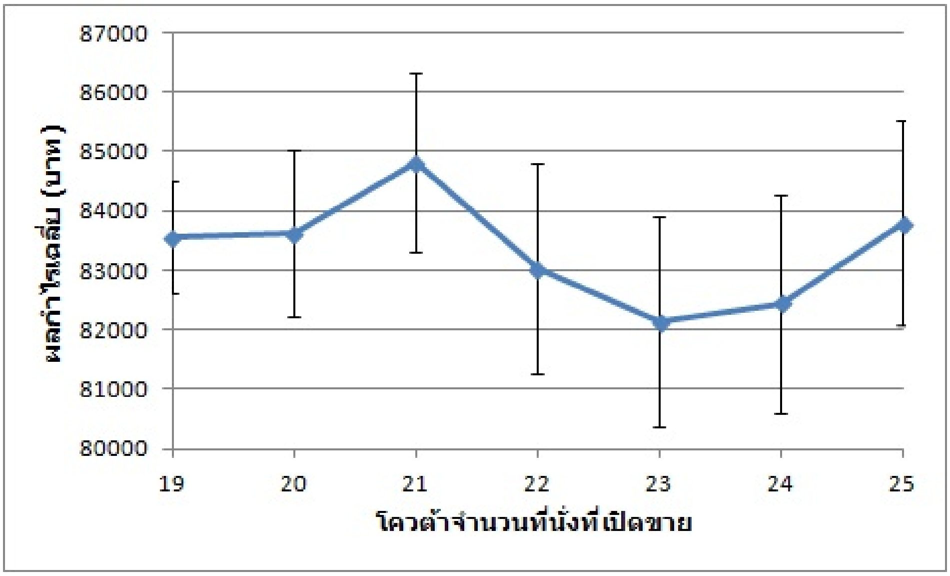
Mean	Variance	Lower CL	Upper CL
84225	82017045.5	63000	99000
0.17	0.14252525	0	1

(ข) ผลของแต่ละรอบทำซ้ำ

Replicate	Var 1	Var 2
1	99000	0
2	94500	0
3	81000	0

ผู้วิเคราะห์สามารถกำหนดจำนวนตั๋วที่ทำให้ผลกำไรเฉลี่ยสูงที่สุดได้ โดยทดลองประมวลผลแบบจำลองที่หลาย ๆ ค่าตัวแปรตัดสินใจ ดังแสดงในรูปที่ 3.3 แถบเส้นตั้งแสดงค่าช่วงความเชื่อมั่นที่นัยสำคัญ 95% ถึงแม้ว่าจำนวนตั๋ว 21 ใบจะทำให้เกิดผลกำไรเฉลี่ยสูงที่สุด แต่ช่วงความเชื่อมั่นของผลกำไรที่ตัวแปรตัดสินใจอื่นก็ทับกับช่วงความเชื่อมั่นที่จำนวนตั๋ว 21 ใบ จึงไม่สามารถสรุปในทันทีว่า 21 ใบดีที่สุด ต้องเพิ่มจำนวนรอบทำซ้ำและประมวลผลใหม่จนกว่าช่วงความเชื่อมั่นจะแคบพอที่จะไม่ทับกัน หัวข้อที่ 5.3.5 จะกล่าวถึงวิธีการ

คำนวณจำนวนรอบทำซ้ำที่เหมาะสมสำหรับการตัดสินใจ



รูปที่ 3.3: ผลกำไรเฉลี่ยที่โควต้าการซื้อตั๋วต่าง ๆ เมื่อประมวลผล 100 รอบ (แบบจำลองในหัวข้อที่ 3.3)

3.4 ตัวอย่างปัญหาสินค้าคงคลัง

ตัวอย่างนี้ดัดแปลงจาก Ragsdale (2008) บริษัทไฮเทคคอมพิวเตอร์ (HTC) เป็นผู้ค้าปลีกอุปกรณ์คอมพิวเตอร์ บางครั้งหน้าร้านขาดสินค้าที่กำลังอยู่ในช่วงโปรโมชั่น เช่น จอมอนิเตอร์รุ่น S นโยบายสินค้าคงคลังปัจจุบัน คือสั่งสินค้ารุ่นนี้ 50 ตัวเมื่อสต็อกที่ร้านอยู่ที่ระดับ 28 ตัวหรือต่ำกว่า (จุดสั่งซื้อ = 28) จากการวิเคราะห์ประวัติการขายจอมอนิเตอร์รุ่นนี้ ภายใน 6 เดือนที่ผ่านมา HTC มีข้อมูลการแจกแจงของอุปสงค์และเวลาในการส่งมอบดังแสดงในตารางที่ 3.12

ตารางที่ 3.12: การแจกแจงความน่าจะเป็นสำหรับตัวอย่างบริษัทไฮเทคคอมพิวเตอร์ (หัวข้อที่ 3.4)

อุปสงค์ต่อวัน	0	1	2	3	4	5	6
ความน่าจะเป็น	0.03	0.05	0.13	0.25	0.22	0.20	0.12

เวลาในการส่งมอบ	1	2	3
ความน่าจะเป็น	0.20	0.50	0.30

สต็อกเริ่มต้นมี 50 ชิ้น ผู้จัดการเช็คสต็อกเมื่อปิดร้านตอนเย็นเพื่อตัดสินใจว่าจะสั่งสินค้าเพิ่มหรือไม่ หาก

สั่งสินค้าตอนเย็นวันที่ t และเวลาในการส่งมอบ d วัน สินค้าจะพร้อมขายในตอนเช้าวันที่ $t+d+1$ ผู้จัดการร้านต้องการกำหนดจุดสั่งซื้อและปริมาณสั่งซื้อที่ทำให้ระดับการบริการไม่ต่ำกว่า 98% และก่อให้เกิดระดับสินค้าคงคลังเฉลี่ยรายวันน้อยที่สุด นิยามระดับการบริการดังนี้

$$\text{ระดับการให้บริการ} = \frac{\text{อุปสงค์ที่ได้ตอบสนองลูกค้า}}{\text{อุปสงค์รวม}} \tag{3.15}$$

อุปสงค์ที่ได้ตอบสนองเกิดขึ้นเมื่อลูกค้ามาซื้อแล้วหน้าร้านแล้วมีสินค้าขายให้ทันที หากมาแล้วไม่มีของ HTC ไม่จองสินค้าให้ลูกค้า (No backorder)

แบบจำลองสถานการณ์

ตารางที่ 3.13ก แสดงข้อมูลและตารางที่ 3.13ข แสดงแบบจำลองบนสเปรดชีต การแยกข้อมูลนำเข้าให้อยู่อีกเวิร์คชีตหนึ่งช่วยเน้นว่าเซลล์ใดเป็นพารามิเตอร์นำเข้า ทำให้ง่ายในการเปลี่ยนค่าพารามิเตอร์ เช่น เมื่อวิเคราะห์ความไว สังเกตว่าการวางตำแหน่งเซลล์ที่ดีจะไล่เรียงจากแถวบนลงแถวล่างและจากซ้ายไปขวา เพราะเข้ากับธรรมชาติการอ่านภาษาไทยและภาษาอังกฤษ ไม่ควรวางสะเปะสะปะลงบนหน้าเวิร์คชีต ตัวอย่างในตารางที่ 3.13ข นี้ใช้หนึ่งแถวสำหรับหนึ่งวัน แต่ละวันเกี่ยวเนื่องกันโดยสต็อกตอนสิ้นวันนี้จะเป็นสต็อกของต้นวันของวันถัดไป อาจมีสินค้าที่ได้สั่งไปแล้วมาถึงตอนเช้าพร้อมขายในวันนั้น วันแรกพิเศษกว่าวันอื่นเพราะต้องนิยามเงื่อนไขเริ่มต้น เช่น ค่าของสต็อกเริ่มต้นและไม่มีสินค้าที่กำลังเดินทางมาถึง จากนั้นสูตรของตัวแปรในระบบ (State variable) ของวันที่สองถูกคำนวณที่เกี่ยวเนื่องกับวันแรก วันที่ t สัมพันธ์กับวันที่ $t-1$ เช่น สต็อกตอนเปิดร้านและจำนวนสินค้าที่เพิ่งมาส่ง อุปสงค์ของลูกค้าในแต่ละวันถูกสุ่มอย่างเป็นอิสระต่อกัน เมื่อนิยามสูตรของวันที่สองเสร็จก็สามารถคัดลอกไปวางยังแถวอื่น ๆ ได้ หลังจากนิยามสูตรสำหรับทุกวันแล้วจึงคำนวณดัชนีชี้วัดของระบบ ระดับการให้บริการคำนวณเป็นสัดส่วนของอุปสงค์ที่ได้ขายรวมทั้งเดือนเมื่อเทียบกับอุปสงค์รวมที่เข้ามาด้วยสมการที่ (3.15) ระดับสินค้าคงคลังเฉลี่ยคำนวณจากสต็อกตอนสิ้นวัน (คอลัมน์ `F`) หรือตอนต้นวัน (คอลัมน์ `B`) ก็ได้

ตารางที่ 3.14 แสดงสูตรเอกซ์เซลที่ใช้ในแบบจำลองในตารางที่ 3.13ข คอลัมน์ `A` แสดงลำดับที่ของวันสมมติว่าประมวลผล 1 เดือน คอลัมน์ `B` แสดงสต็อกตอนเริ่มเปิดร้านซึ่งก็คือสต็อกตอนปิดร้านของวันก่อนหน้า; คอลัมน์ `C` แสดงสินค้าที่เพิ่งมาส่งโดยตรวจสอบจากวันก่อนหน้าว่ามีสินค้าที่สั่งไปแล้วกำลังจะมาถึงหรือไม่ หากลำดับที่ของวันนั้น ๆ ตรงกับวันที่สินค้าจะมาถึงในคอลัมน์ `J` ก็แปลว่ามีของมาส่งในวันนั้น; ฟังก์ชัน `countif`(อาร์เรย์, เงื่อนไข) นับจำนวนเซลล์ในอาร์เรย์ที่ทำให้เงื่อนไข (Condition) เป็นจริง ซึ่งก็คือจำนวนคำสั่งซื้อ (Order) ที่สั่งไปแล้วและกำหนดมาถึงวันนั้น ๆ คูณด้วยขนาดของคำสั่งซื้อ; คอลัมน์ `D` แสดงอุปสงค์รายวันซึ่งสุ่มจากการแจกแจงแบบไม่ต่อเนื่องที่โจทย์ให้มา; คอลัมน์ `E` คำนวณอุปสงค์ของลูกค้าที่ HTC มีสินค้าขายให้ สต็อกที่มีคือสต็อกที่เหลือจากวันก่อนหน้ากับที่มาส่งตอนเช้า; เปรียบเทียบสต็อกที่มีกับอุปสงค์ของ

ตารางที่ 3.13: แบบจำลองสินค้าคงคลัง (หัวข้อที่ 3.4)

(ก) พารามิเตอร์นำเข้าในหน้าเวิร์คชีต `Data`

	A	B	C	D	E	F
1						
2	**ไฮเทคคอมพิวเตอร์**					
3						
4		**เวลาในการส่งมอบ**			**อุปสงค์**	
5		จำนวนวัน	ความน่าจะเป็น		จำนวนหน่วย	ความน่าจะเป็น
6		3	0.20		0	0.01
7		4	0.60		1	0.02
8		5	0.20		2	0.04
9		**รวม**	**1.00**		3	0.06
10					4	0.09
11					5	0.14
12					6	0.18
13					7	0.22
14					8	0.16
15					9	0.06
16					10	0.02
17					**รวม**	**1.00**

(ข) แบบจำลอง

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		**ไฮเทคคอมพิวเตอร์**											
2													
3		สต็อก	จำนวนชิ้น	อุปสงค์	อุปสงค์	สต็อก	ตำแหน่ง	สั่งหรือไม่?	เวลานำ	วันที่สินค้า			
4	วันที่	ตอนต้นวัน	ที่รับใหม่	ที่เข้ามา	ที่ได้ขาย	ตอนสิ้นวัน	สต็อก	(0=ไม่,1=ใช่)		มาถึง		ตัวแปรตัดสินใจ	
5	1	50	0	6	6	44	44	0	0	0		จุดสั่งซื้อ	28
6	2	44	0	10	10	34	34	0	0	0		เพดานสั่งซื้อ	50
7	3	34	0	5	5	29	29	0	0	0			
8	4	29	0	7	7	22	22	0	0	0		ดัชนีชี้วัด	
9	5	22	0	9	9	13	13	1	4	10		ระดับการบริการ	98.84%
10	6	13	0	6	6	7	17	1	4	11		สต๊อกเฉลี่ย	30.03

ลูกค้า จึงใช้ฟังก์ชัน `MIN()`; คอลัมน์ `F` คำนวณสต็อกที่เหลือตอนปิดร้าน ซึ่งก็คือสต็อกที่มีตอนเปิดร้านและของใหม่ที่เพิ่งมาส่ง (ถ้ามี) หักด้วยของที่ขายไปในวันนั้น; สต็อกที่เหลือตอนปิดร้านแตกต่างจาก**ตำแหน่งสต็อก** (Inventory position, IP) ซึ่งเป็นผลบวกของสต็อกที่มีอยู่ที่ร้านและออร์เดอร์ที่สั่งไปแล้วและกำลังจะมาถึง; HTC ใช้ตำแหน่งสต็อก (คอลัมน์ `G`) สำหรับตัดสินใจว่าควรจะสั่งสินค้าเพิ่มหรือไม่โดยเปรียบเทียบกับจุดสั่งซื้อ (Reorder point, ROP) หากตำแหน่งสต็อกมีค่าน้อยกว่าหรือเท่ากับจุดสั่งซื้อ (IP $\leqslant$ ROP) ก็จะสั่ง การตรวจสอบใช้ฟังก์ชัน `if(`เงื่อนไข, ผลหากเงื่อนไขเป็นจริง, ผลหากเงื่อนไขเป็นเท็จ`)`; โดยปริยายเอกซ์เซลและโปรแกรมคอมพิวเตอร์ถือว่าจำนวนบวกเทียบเท่าค่าจริงทางตรรกะ (Logical true) และศูนย์เป็นค่าเท็จทางตรรกะ (Logical false); เวลานำของการส่งมอบไม่แน่นอน (คอลัมน์ `I`); จากวันที่สั่งและเวลานำจะได้วันที่สินค้าไปถึงร้าน (คอลัมน์ `J`); เมื่อคัดลอกสูตรจนครบ 30 แถวสำหรับ 1 เดือนก็จะคำนวณดัชนีชี้วัด ระดับการให้บริการ (เซลล์ `M9`) คำนวณจากผลรวมอุปสงค์ที่ได้ขายหารด้วยผลรวมของอุปสงค์ที่เข้ามา และระดับสินค้าคงคลัง (เซลล์ `M10`) เฉลี่ยทั้งเดือนจากอุปสงค์ตอนปิดร้าน

ตารางที่ 3.14: สูตรเอกซ์เซลในแบบจำลองสินค้าคงคลัง (หัวข้อที่ 3.4)

ตำแหน่งเซลล์	**สูตรเอกซ์เซล**	**คัดลอกไปยัง**
วันแรก		
D6	=DiscreteDev(Data!$E$7:$E$17,Data!$F$7:$F$17)	D7:D35
E6	=MIN(B6+C6,D6)	E7:E35
F6	=B6+C6-E6	F7:F35
G6	=F6	
H6	=IF(G6 < $M$5,1,0)	H7:H35
I6	=IF(H6=0,0,DiscreteDev(Data!$B$7:$B$9,Data!$C$7:$C$9))	I7:I35
J6	=IF(H6=0,0,A6+I6+1)	J7:J35
วันที่สองและวันถัดไป		
B7	=F6	B8:B35
C7	=COUNTIF($J$6:J6,A7)*$M$6	C8:C35
G7	=G6-E7+IF(H6=1,$M$6,0)	G8:G35
ดัชนีชี้วัด		
M9	=SUM(E6:E35)/SUM(D6:D35)	
M10	=AVERAGE(B6:B35)	

ตารางที่ 3.15ก แสดงผลการประมวลผลที่ 100 รอบทำซ้ำ ตารางที่ 3.15ข แสดงผลอย่างละเอียดบาง

ส่วน สังเกตว่านโยบายสินค้าคงคลังที่ใช้อยู่ยังไม่สามารถให้ระดับการบริการที่ต้องการได้ คือทั้งช่วงความเชื่อมั่นยังต่ำกว่า 98% ผู้วิเคราะห์อาจพิจารณาเพิ่มจุดสั่งซื้อหรือเพิ่มปริมาณสั่งซื้อ เช่นที่จุดสั่งซื้อ 34 ชิ้นและปริมาณสั่งซื้อเท่าเดิมจะได้ค่าเฉลี่ยระดับบริการเป็น 99.3% ช่วงความเชื่อมั่นที่ 95% เป็น [94.4%, 100.0%] ซึ่งครอบคลุมค่าเป้าหมายที่ต้องการที่ 98% แต่ราคาที่ต้องจ่ายคือต้องเก็บสต็อกมากขึ้น (ระดับสต็อกเฉลี่ยสูงขึ้น) แบบจำลองสถานการณ์เอื้อให้ผู้ใช้ลองทดสอบตัวแปรตัดสินใจหลาย ๆ ชุดว่ามีผลกระทบต่อดัชนีชี้วัดอย่างไร นอกจากนี้แล้วยังสามารถวิเคราะห์ผลกระทบด้านอื่น ๆ ของตัวแปรนำเข้าได้ เช่น หากลดความแปรปรวนของเวลานำแล้ว จะทำให้ค่าดัชนีชี้วัดเปลี่ยนแปลงอย่างไร ตรรกะช่วยบอกว่าถ้าลดความแปรปรวนของเวลานำในการส่งมอบสินค้า ระบบน่าจะมีระดับการบริการสูงขึ้น แบบจำลองช่วยตอบคำถามได้ว่าดีขึ้น*เท่าใด*

ตารางที่ 3.15: ผลการประมวลผลสำหรับแบบจำลองสถานการณ์สินค้าคงคลังเมื่อประมวลผล 100 รอบ (ตารางที่ 3.13)

(ก) ผลสรุป

ดัชนีชี้วัด	ค่าเฉลี่ย	ช่วงความเชื่อมั่นที่นัยสำคัญ 95%
ระดับการบริการ	95.9%	[95.2%, 96.6%]
สต็อกเฉลี่ย (ชิ้น)	25.72	[25.32, 26.12]

(ข) บางส่วนของเอาท์พุทที่ทุกรอบทำซ้ำ

A	B	C
Replicate	Var 1	Var 2
1	1	28.03
2	1	26.17
⋮	⋮	⋮
99	1	31.40
100	1	29.77

3.5 ตัวอย่างการเลือกโครงการเพื่อลงทุน

ตัวอย่างนี้ดัดแปลงจาก Ragsdale (2008) บริษัทปัญญาอินเวสเม้นท์มีเงินสด 2 ล้านบาทสำหรับการลงทุนในงานวิจัยและพัฒนาใหม่ ๆ แต่ละโครงการมีเงินลงทุน ความน่าจะเป็นที่จะได้ผลงานที่เป็นประโยชน์เชิง

พาณิชย์ และศักยภาพในการทำเงินที่แตกต่างกัน หากโครงการใดไม่มีผลงานที่ขายได้ ผู้ลงทุนก็จะสูญเงินลงทุนทั้งหมด ถึงแม้ว่าจะมีผลงานที่ขายได้ รายได้ที่จะเกิดขึ้นก็ไม่แน่นอน แต่สามารถประมาณขอบเขตล่าง, ค่าที่น่าจะเป็นไปได้มากที่สุด, และขอบเขตบน ตารางที่ 3.16 แสดงข้อมูลนำเข้า ผู้ลงทุนรายนี้ต้องการทราบว่าควรลงทุนในโครงการใดบ้างเพื่อให้ค่าคาดหมายของผลตอบแทนสู งที่สุด ในขณะเดียวกันเงินลงทุนรวมต้องไม่เกินเงินต้นที่วางไว้

ตารางที่ 3.16: ข้อมูลนำเข้าสำหรับตัวอย่างการเลือกโครงการเพื่อลงทุน (หัวข้อที่ 3.5)

	ค่าใช้จ่ายเริ่มต้น	ความน่าจะเป็น	ศักยภาพในการทำเงิน (1,000 บาท)		
โครงการ	(1,000 บาท)	ที่จะสำเร็จ	ขอบเขตล่าง	ค่าที่เป็นไปได้มากที่สุด	ขอบเขตบน
1	250	0.9	600	750	900
2	650	0.7	1,250	1,500	1,600
3	250	0.6	500	600	750
4	500	0.4	1,600	1,800	1,900
5	700	0.8	1,150	1,200	1,400
6	30	0.6	150	180	250
7	350	0.7	750	900	1,000
8	70	0.9	220	250	320

ตารางที่ 3.17 แสดงแบบจำลองและตารางที่ 3.18 แสดงสูตรเอกซ์เซล ใช้หนึ่งแถวสำหรับหนึ่งโครงการ คอลัมน์ `F` แสดงผลลัพธ์ของการลงทุน หากไม่เลือกลงทุนก็ไม่สำเร็จ แต่หากลงทุน ผลที่เกิดมีเพียงสองค่า คือ สำเร็จและไม่สำเร็จ หรือการแจกแจงแบบเบอร์นูลลี ซึ่งก็คือกรณีพิเศษของการแจกแจงแบบทวินามที่สุ่มเพียง 1 ครั้ง; เมื่อการลงทุนสำเร็จ จำลองผลตอบแทนที่ได้ด้วยค่าสุ่มที่มีการแจกแจงแบบสามเหลี่ยม เพราะมีข้อมูลเพียงแค่ขอบเขตล่าง, ขอบเขตบน, และค่าที่เป็นไปได้มากที่สุด; จุดเด่นของตัวอย่างนี้ก็คือการแสดงให้เห็นถึงวิธีการจำลองความสัมพันธ์ที่เกี่ยวเนื่องกันเป็นทอด ๆ เหตุการณ์หนึ่งจะเกิดขึ้นเมื่อเหตุการณ์ก่อนหน้าเป็นจริง; คำนวณเงินลงทุนที่ใช้ด้วยสูตร `=SUMPRODUCT(D6:D13,C6:C13)` ซึ่งคำนวณ $\sum_{i=6}^{13} D_i C_i$ หรือ Dot product ของเวคเตอร์นั่นเอง เนื่องจากตัวอย่างนี้ยังไม่ใช้เครื่องมือหาค่าที่ดีที่สุด (Optimization tool) ผู้อ่านจึงต้องลองเลือกชุดโครงการที่จะลงทุนเอง ให้ต้นทุนรวมไม่เกิน 2 ล้านบาท แล้วจึงประมวลผลแบบจำลองเพื่อประมาณผลตอบแทนเฉลี่ยรวมที่ได้ สำหรับโครงงานที่เลือกในตารางที่ 3.17 ช่วงความเชื่อมั่นของผลตอบแทนเฉลี่ย ที่นัยสำคัญ 95% คือ [1.40, 2.94] ล้านบาท และความน่าจะเป็นที่จะขาดทุน คือ [0%, 3%]

การลงทุนนั้นควรพิจารณาผลตอบแทนเฉลี่ยและความเสี่ยงด้วย เพราะการลงทุนที่ให้ผลตอบแทนสูงมัก

จะมีความเสี่ยงสูง และการลงทุนที่ให้ผลตอบแทนต่ำมักจะมีความเสี่ยงต่ำ เช่น ฝากออมทรัพย์กับธนาคารมีความเสี่ยงต่ำกว่าการลงทุนในตลาดหุ้น แต่ก็ได้รับผลตอบแทนที่ต่ำกว่า ในตัวอย่างท้ายบทผู้อ่านจะได้ลองพิจารณาดัชนีชี้วัดทั้งสองด้านนี้

ตารางที่ 3.17: แบบจำลองการเลือกโครงการในหัวข้อที่ 3.5 (หน่วยของเงิน: พันบาท)

	A	B	C	D	E	F	G	H	I	J	K
1											
2					ปัญญาอินเวสเม้นท์						
3											
4			เลือกหรือไม่?	เงินลงทุน	ความน่าจะเป็น	สำเร็จหรือไม่?	ค่าที่	ค่าที่เป็นไปได้	ค่าที่		
5		โครงการ	(1=ใช่, 0=ไม่ใช่)	เบื้องต้น	ที่จะสำเร็จ	(1=ใช่, 0=ไม่ใช่)	น้อยที่สุด	มากที่สุด	มากที่สุด	ผลตอบแทน	ผลกำไร
6		1	1	฿250.00	90%	1	฿600.00	฿750.00	฿900.00	฿663.71	฿413.71
7		2	1	฿650.00	70%	0	฿1,250.00	฿1,500.00	฿1,600.00	฿-	-฿650.00
8		3	1	฿250.00	60%	0	฿500.00	฿600.00	฿750.00	฿-	-฿250.00
9		4	1	฿500.00	40%	1	฿1,600.00	฿1,800.00	฿1,900.00	฿1,784.75	฿1,284.75
10		5	0	฿700.00	80%	0	฿1,150.00	฿1,200.00	฿1,400.00	฿-	฿-
11		6	0	฿30.00	60%	0	฿150.00	฿180.00	฿250.00	฿-	฿-
12		7	1	฿350.00	70%	0	฿750.00	฿900.00	฿1,000.00	฿-	-฿350.00
13		8	0	฿70.00	90%	0	฿220.00	฿250.00	฿320.00	฿-	฿-
14			**เงินที่ใช้ไป**	**฿2,000.00**						**ผลกำไรสุทธิ**	**฿448**
15			**เงินที่มี**	**฿2,000.00**						**ขาดทุนหรือไม่?**	**0**
16			**คงเหลือ**	**฿-**							

ตารางที่ 3.18: สูตรเอกซ์เซลในแบบจำลองการเลือกโครงการ (ตารางที่ 3.17)

ตำแหน่งเซลล์	สูตรเอกซ์เซล	คัดลอกไปยัง
F6	`=IF(C6=1,dBinomialDev(1,E6),0)`	F7:F13
J6	`=IF(F6=1,dTRand(G6,H6,I6),0)`	J7:J13
K6	`=J6-C6*D6`	K7:K13
D14	`=SUMPRODUCT(D6:D13,C6:C13)`	
D16	`=D15-D14`	
K14	`=SUM(K6:K13)`	
K15	`=IF(K14<0,1,0)`	

3.6 บทส่งท้าย

บทนี้ได้นำเสนอแบบจำลองขนาดเล็กบนสเปรดชีตเพื่อช่วยให้ผู้อ่านเข้าใจแนวคิดของการดัดแปลงแบบจำลองสเปรดชีตที่อาจมีอยู่แล้ว ให้สามารถรองรับความไม่แน่นอนของตัวแปรนำเข้า ความไม่แน่นอนนี้ส่งผลต่อ

ความแปรปรวนของเอาท์พุท การจำลองสถานการณ์ช่วยให้แบบจำลองมีประโยชน์มากกว่าเดิม โดยเฉพาะในแง่การวิเคราะห์ความเสี่ยง การจำลองบนสเปรดชีตช่วยให้ผู้อ่านเข้าใจการจำลองสถานการณ์แบบเหตุการณ์ไม่ต่อเนื่องที่จะนำเสนอต่อไปได้ง่ายขึ้น เพราะการจำลองบนสเปรดชีตแสดงความสัมพันธ์ระหว่างตัวแปรให้ผู้ใช้เห็นชัดกว่า หากผู้อ่านสนใจที่จะเรียนรู้การจำลองสถานการณ์บนสเปรดชีตเพิ่มเติม สามารถอ่านได้จาก Winston (2000) และ Seila et al. (2003) นอกจากนี้ยังมีหนังสือที่กล่าวถึงการประยุกต์ใช้การวิจัยดำเนินงาน โดยเน้นที่การสร้างแบบจำลองบนสเปรดชีต เช่น Albright et al. (2009), Ragsdale (2008), Barlow (2005), และ Powell and Baker (2009) ซึ่งล้วนมีบทเกี่ยวกับการจำลองสถานการณ์

แบบฝึกหัด

3.1. พิจารณาระบบสินค้าคงคลังที่ใช้นโยบาย (s, S) ร้านค้าขายเก้าอี้ในราคาตัวละ 400 บาท นโยบายสินค้าคงคลังคือให้เก็บสต็อคไว้อย่างน้อย 6 ตัวตอนเปิดร้าน ถ้าสต็อกเหลือน้อยกว่า 6 ตัวตอนปิดร้าน ก็จะสั่งสินค้าเพิ่มเพื่อให้มีสต็อก 12 ตัว สินค้ามาถึงตอนเปิดร้านใน*ตอนเช้าวันมะรืน* (วันที่ถัดจากวันถัดไป) เช่น ถ้าสั่งตอนเย็นวันจันทร์ จะได้ของตอนเช้าวันพุธ ทางร้านมีคำสั่งซื้อที่กำลังมาส่งไม่เกิน 1 คำสั่ง คำสั่งซื้อหนึ่งมีต้นทุนการสั่งสินค้า 50 บาท ต้นทุนต่อชิ้น 100 บาท ต้นทุนการถือครองอีก 20 บาทต่อชิ้นสำหรับเก็บสินค้าข้ามคืน และค่าเสียโอกาสอีก 60 บาทต่อชิ้นหากลูกค้ามาซื้อแล้วไม่มีของ ลูกค้าก็จะไปซื้อที่อื่นแทน (No backlogs)

สมมติว่าตอนเช้าวันจันทร์ ร้านเปิดพร้อมเก้าอี้ 12 ตัวและในสัปดาห์นี้ (เปิดแค่วันจันทร์ถึงศุกร์) มีอุปสงค์ของลูกค้า (ทางร้านไม่ทราบอุปสงค์เหล่านี้) ดังนี้

$$D_1 = 10; D_2 = 2; D_3 = 14; D_4 = 6; D_5 = 7$$

ให้พัฒนาแบบจำลองระบบนี้บนกระดาษหรือซอฟต์แวร์ด้วยเอ็กซเซล (ยังไม่ใช้พ็อพทูล) เพื่อตอบคำถามต่อไปนี้

(i) มีเก้าอี้กี่ตัวที่เหลือตอนเย็นวันพฤหัสบดี (วันที่ 4)

(ii) มีกี่วันใน 5 วันที่มีเหตุการณ์สินค้าหมดสต็อก

(iii) ร้านค้ามีผลกำไรเท่าใดใน 1 สัปดาห์ (รวมถึงต้นทุนของคำสั่งซื้อที่อาจจะมีตอนเย็นวันศุกร์)

3.2. ค่าใช้จ่ายในการผลิตนิตยสารรายเดือน*ไทยแลนด์ซิม* คือ 45 บาทต่อเล่ม ราคาขายคือ 60 บาท อุปสงค์ของนิตยสารในแต่ละเดือนเป็นดังนี้

จำนวนนิตยสารที่ขายได้	500	600	700	800	900	1000
ความน่าจะเป็น	0.1	0.15	0.2	0.25	0.2	0.1

นิตยสารที่ขายไม่ได้มีมูลค่า 25 บาท สำนักพิมพ์ควรพิมพ์นิตยสาร*ไทยแลนด์ซิม*กี่เล่ม (ในหน่วย 100 เล่ม) เพื่อทำให้ผลกำไรสุทธิสูงสุด ให้ใช้จำนวนรอบทำซ้ำเพียงพอที่จะเลือกตัวเลือกที่ดีที่สุดอย่างมีนัยสำคัญได้ ให้สร้างกราฟคล้ายรูปที่ 3.3 โดยที่แกนนอนเป็นยอดพิมพ์และแกนตั้งเป็นผลกำไรเฉลี่ย

3.3. บริษัทสุคะโตต้องการให้มีระดับการให้บริการสูงอย่างน้อย 98% สำหรับสินค้า ก. นั่นคือต้องกำหนดนโยบายสินค้าคงคลังให้มีสต็อกที่หน้าร้านเพียงพอที่จะรองรับอุปสงค์และเวลาในการส่งมอบที่ไม่แน่นอน นโยบายปัจจุบันที่ใช้ คือ สั่งสินค้า ก. 30 ชิ้นเมื่อสต็อกที่หน้างานอยู่ที่ระดับ 15 ชิ้นหรือต่ำกว่า (จุดสั่งซื้อ = 15) จากที่ศึกษาประวัติการขายสินค้า ก. ภายใน 6 เดือนที่ผ่านมา บริษัทสุคะโตมีข้อมูลการแจกแจงของอุปสงค์และเวลาในการส่งมอบดังนี้

อุปสงค์ของสินค้า ก. ต่อวัน	0	1	2	3	4	5	6
ความน่าจะเป็น	0.03	0.05	0.13	0.25	0.22	0.20	0.12

เวลาในการส่งมอบ	1	2	3
ความน่าจะเป็น	0.20	0.50	0.30

สมมติให้สต็อกเริ่มต้นที่ 30 ชิ้น ผู้จัดการเช็คสต็อกเมื่อปิดร้านตอนเย็นเพื่อตัดสินใจว่าจะสั่งสินค้าเพิ่มหรือไม่ หากสั่งสินค้าตอนเย็นวันที่ t และเวลาในการส่งมอบ d วัน สินค้าจะพร้อมขายในตอนต้นวันที่ $t + d + 1$ ภายใต้นโยบายสินค้าคงคลังปัจจุบัน ให้รายงานช่วงความเชื่อมั่นของดัชนีชี้วัดต่อไปนี้ ที่นัยสำคัญ 95%

(i) ระดับการให้บริการ (ดูนิยามที่สมการที่ 3.15)

(ii) ระดับสต็อกรายวันเฉลี่ย

ผู้จัดการต้องการทราบว่ามีนโยบายสินค้าคงคลัง (จุดสั่งซื้อและปริมาณสั่งซื้อ) อื่นที่ยังรักษาระดับการให้บริการให้ไม่ต่ำกว่า 98% โดยที่สามารถลดระดับสต็อกเฉลี่ยลงได้หรือไม่ (ให้ใช้จำนวนรอบทำซ้ำ 500 รอบ)

3.4. นราธิปเพิ่งสำเร็จการศึกษา พ่อจึงให้เงินลงทุน 100,000 บาท เขาเลือกลงทุนในหุ้น 4 ตัว: บริษัท A, B, C และ D ตารางข้างล่างนี้แสดงผลตอบแทนของหุ้นและส่วนเบี่ยงเบนมาตรฐานของผลตอบแทน (เพื่อให้การวิเคราะห์นี้ไม่ซับซ้อนจนเกินไป สมมติให้ผลตอบแทนเหล่านี้ไม่มีความสัมพันธ์กันและผลตอบแทนมีการแจกแจงแบบปกติ) นราธิปต้องการทราบว่าเขาควรจะกระจายเงินลงทุนอย่างไรเพื่อให้ผลตอบแทนเฉลี่ยสูงที่สุด แต่ในขณะเดียวกันก็รักษาระดับความเสี่ยงให้อยู่ในขอบเขตที่รับได้ (ไม่เกิน 13%)

หุ้น	สัดส่วนการลงทุน	ค่าเฉลี่ยผลตอบแทน	ส่วนเบี่ยงเบนมาตรฐาน
A	x_1	40%	28%
B	x_2	16%	22%
C	x_3	20%	24%
D	x_4	30%	18%

สังเกตว่า $\sum_{i=1}^{4} x_i = 1$ และ $0 \leqslant x_i \leqslant 1$ ดังนั้นตัวแปรตัดสินใจจริง ๆ คือ x_1, x_2 และ x_3 เอาท์พุทซึ่งเป็นดัชนีชี้วัดคือ

$$\text{ผลตอบแทนรวม (หน่วยเป็น \%)} = 0.4x_1 + 0.16x_2 + 0.2x_3 + 0.3x_4$$

ให้ประมาณระดับความเสี่ยงของพอร์ตโฟลิโอการลงทุนนี้จากส่วนเบี่ยงเบนมาตรฐานของผลตอบแทนรวม ให้ทดลองใช้ x_1, x_2 และ x_3 2-3 ชุด สำหรับแต่ละชุด รายงาน

(i) ค่าเฉลี่ยผลตอบแทนรวมและช่วงความเชื่อมั่นที่นัยสำคัญ 95%

(ii) ระดับความเสี่ยงซึ่งคือส่วนเบี่ยงเบนมาตรฐานของผลตอบแทน

(iii) ความน่าจะเป็นที่ผลตอบแทนรวมจะมีค่าอย่างน้อย 20%

3.5. บริษัทยาแห่งหนึ่งวางแผนที่จะลงทุนในยาตัวใหม่ ค่าใช้จ่ายในการวิจัยมีการแจกแจงแบบสม่ำเสมอระหว่าง 90 ล้านบาท ถึง 150 ล้านบาท มีความน่าจะเป็น 40% ที่ยาจะไม่ได้ผลในการรักษา หากไม่ได้ผลก็จะไม่มีการขายเกิดขึ้น จำนวนคนไข้ในประเทศทั้งหมดประมาณ 400,000 คน ถ้ายานี้ใช้ได้ผลบริษัทนี้คาดว่าส่วนแบ่งทางการตลาดมีการแจกแจงแบบสามเหลี่ยมที่มีขอบเขตล่างคือ 5%, ค่าที่น่าจะเป็นไปได้มากที่สุด 20% และขอบเขตบนคือ 30% สำหรับลูกค้าแต่ละคนที่จะใช้ยาใหม่นี้บริษัทคาดว่าจะเกิดยอดขาย 3,000 บาทในปีแรก บริษัทต้องการทราบการแจกแจงของผลกำไรที่สามารถคาดหวังได้จากการลงทุนนี้

(i) ให้รายงานช่วงความเชื่อมั่นของผลกำไรที่นัยสำคัญ 95%

(ii) ให้แสดงฮีสโตแกรมของผลกำไร

(iii) ให้รายงานความน่าจะเป็นที่ผลกำไรจะเกิน 24 ล้านบาท (ดูคำแนะนำในแบบฝึกหัดข้อ 3.4.)

(iv) ให้กำหนดค่าเปอร์เซ็นไทล์ที่ 80% ของผลกำไร

3.6. บริษัทตีโลปะให้สวัสดิการด้านประกันสุขภาพต่อพนักงาน 3 คน พนักงานแต่ละคนมีโอกาสที่จะใช้สิทธิประมาณ 90% ถ้าพนักงานใช้สิทธิประกันในปีนั้น ค่าใช้จ่ายมีค่าประมาณ 3,000 บาทต่อคนด้วยความน่าจะเป็น 0.9 และ 300,000 บาทต่อคนด้วยความน่าจะเป็น 0.1 บริษัทตีโลปะซื้อประกันฯ แบบที่

บริษัทประกันจะจ่ายค่าใช้จ่ายด้านสุขภาพทั้งหมด แต่ตีโลปะจะต้องจ่าย 150,000 บาทแรก ตีโลปะสนใจในค่าใช้จ่ายรวมที่ต้องจ่ายเองใน 1 ปี ให้ใช้พ็อพทูลเพื่อจำลองสถานการณ์นี้ โดยใช้ 100 รอบทำซ้ำ และรายงานช่วงความเชื่อมั่นของค่าใช้จ่ายที่ตีโลปะต้องจ่ายเองใน 1 ปี ที่นัยสำคัญ 95%

3.7. บริษัทผู้ผลิตเครื่องหนังแห่งหนึ่งมีทีมซ่อมบำรุง ผู้บริหารต้องการใช้การจำลองสถานการณ์เพื่อกำหนดจำนวนคนที่เหมาะสมสำหรับทีมนี้ จำนวนคนที่พิจารณาคือ 2, 3 หรือ 4 คน เวลาที่ใช้ในการซ่อมเครื่องจักรมีการแจกแจงแบบสม่ำเสมอที่มีค่าต่ำสุด 0 ถึงสองเท่าของค่าเฉลี่ย ซึ่งค่าเฉลี่ยนี้ขึ้นอยู่กับจำนวนช่างในทีม ดังนี้

จำนวนช่างในทีม	**เวลาซ่อมเฉลี่ย (ชั่วโมง)**
2	4
3	3
4	2

เวลาระหว่างการเสียของเครื่องจักร (Time to failure) มีการแจกแจงแบบเอ็กซโพเนนเชียลที่มีค่าเฉลี่ย 5 ชั่วโมง เมื่อเครื่องจักรเสียและต้องการการซ่อม (สมมติว่ามีจำนวนเครื่องจักรจำนวนมาก) ผู้บริหารตั้งเป้าให้เวลารอเพื่อซ่อมไม่เกิน 3 ชั่วโมง ผู้บริหารไม่ต้องการให้จำนวนช่างในทีมมีมากเกินไป ให้สมมติว่าช่างทุกคนในทีมทำงานร่วมกันเพื่อซ่อมเครื่องจักรทีละเครื่อง

ให้ใช้พ็อพทูลเพื่อจำลองระบบนี้ให้พิจารณาการเสียของเครื่องจักร 1000 ครั้ง เพื่อเลือกจำนวนช่างในทีมที่เหมาะสมให้ได้ตามเป้าหมายของผู้บริหาร โดยพิจารณาค่าเฉลี่ยเวลารอซ่อมและความน่าจะเป็นที่เวลารอซ่อมมากกว่า 3 ชั่วโมง

4

การจำลองตัวแปรนำเข้า

แบบจำลองตัวแปรนำเข้า (Input model) เป็นตัวแทนความไม่แน่นอนในการจำลองสถานการณ์เชิงสุ่ม ดังนั้นคุณสมบัติพื้นฐานที่แบบจำลองตัวแปรนำเข้าจำเป็นต้องมี มีดังนี้

- สามารถเป็นตัวแทนความจริงทางกายภาพ เช่น การชักตัวอย่างเพื่อตรวจรับสินค้า (Acceptance sampling) หรืออายุขัย (Lifetime) ของอะไหล่
- สามารถปรับให้เข้ากับสถานการณ์จริงได้ง่าย เช่น หลังจากปรับปรุงกระบวนการแล้ว ค่าเฉลี่ยเวลาในการบริการเท่าเดิม แต่ความแปรปรวนของเวลาลดลง หากจำลองเวลาการให้บริการด้วยการแจกแจงความน่าจะเป็นแบบปกติ ผู้วิเคราะห์สามารถปรับค่าส่วนเบี่ยงเบนมาตรฐานได้โดยง่าย
- สามารถสร้างตัวเลขสุ่มด้วยวิธีทางคณิตศาสตร์ได้

ถึงแม้ว่าจุดประสงค์ของบทนี้เพื่อเลือกการแจกแจงความน่าจะเป็นที่สอดคล้องกับข้อมูล แต่ผู้อ่านควรระลึกไว้ว่าการแจกแจงความน่าจะเป็นเป็นเพียงการจำลองความจริง ไม่มี*การแจกแจงที่แท้จริง*ที่รอให้ค้นพบ แบบจำลองที่ดีมีหน้าที่เพียงให้ผลการคำนวณที่เป็นประโยชน์สำหรับการตัดสินใจ

บทนี้อธิบายการคัดสรรการแจกแจงความน่าจะเป็นให้สอดคล้องกับสถานการณ์ที่ต้องการจำลอง และให้เข้ากับข้อมูลจริงที่เก็บมาได้ (ถ้ามี) แบ่งเป็นสองกรณี คือ **มีระบบที่สนใจอยู่แล้วจริง** นั่นคือมีข้อมูลจริง และกรณีที่**ยังไม่มีระบบที่สนใจยังอยู่จริง** เช่น การศึกษาความเป็นไปได้ (Feasibility study) ซึ่งหมายความว่าต้องกำหนดตัวแปรนำเข้าด้วยวิธีอื่น เนื้อหาในบทนี้สมมติว่าข้อมูลในชุดเป็นอิสระต่อกันและเป็นอิสระต่อข้อมูลชุดอื่น ๆ (อ่านหัวข้อ 2.2.1 สำหรับคำอธิบายเพิ่มเติม) หากข้อมูลมีความสัมพันธ์ในตัวเอง (Autocorrelation) ควรใช้แบบจำลองอื่นนอกเหนือจากการแจกแจงความน่าจะเป็น เช่น อนุกรมเวลา (อ่านเพิ่มเติมใน Banks et al. 2005, Law 2006) หากข้อมูลชุดหนึ่งมีความเกี่ยวข้องกับอีกข้อมูลอีกชุดหนึ่ง อาจต้องกำหนดระดับความสัมพันธ์ด้วยค่าสหสัมพันธ์ (สมการที่ 2.15) ซอฟต์แวร์สำเร็จรูปบางชนิดสามารถจำลองข้อมูลลักษณะนี้ได้ เช่น @Risk เป็นต้น

4.1 การเก็บข้อมูล

ในทางปฏิบัติ ขั้นตอนนี้อาจใช้เวลามากกว่าขั้นตอนอื่น ๆ ในการศึกษาด้วยสาเหตุหลายประการ เช่น อาจจะต้องลงสนามไปเก็บข้อมูล (Data) เองด้วยนาฬิกาจับเวลาหรือการสัมภาษณ์ การไปขอความอนุเคราะห์นี้อาศัยความร่วมมือของเจ้าหน้าที่ที่หน้างาน ผู้วิเคราะห์ควรใช้เวลาทำความคุ้นเคยกับพวกเขาและสื่อสารให้ชัดเจนว่าการศึกษาที่กำลังทำนั้นมีวัตถุประสงค์ใด ถึงแม้บางครั้งมีข้อมูลจากเอกสารที่เก็บไว้ ข้อมูลนั้นอาจยังไม่สามารถนำมาใช้ได้ทันที ควรกลั่นกรองพิจารณาความสมจริงและคัดกรองข้อมูลก่อน เช่น หากเป็นระบบคนไข้นอกของโรงพยาบาล ควรตรวจสอบว่าเวลาที่คนไข้เข้ามานั้นเป็นเวลาที่คลินิกเปิดหรือไม่ เวลาที่เริ่มต้นในบันทึกเวลา (Time log) เป็นเวลาที่ลูกค้ามาถึงระบบจริงหรือไม่ หรือมีขั้นตอนที่กระทำก่อนหน้านี้แต่ไม่ปรากฏในบันทึก ถึงแม้ข้อมูลอยู่ในรูปประเบียนอิเลกทรอนิกส์ (Electronic records) ก็อาจใช้เวลาในการค้น

คืน (Retrieve) เพราะต้องขออนุญาตและต้องอาศัยบุคลากรที่รู้วิธีทำ

ก่อนที่จะไปเก็บข้อมูลจริง ควรสังเกตระบบเบื้องต้น และควรรู้พื้นฐานเกี่ยวกับระบบที่ศึกษา ในแง่ต่อไปนี้

- **ขั้นตอนการทำงาน** อาจแสดงด้วยแผนภาพการไหล (Flow diagram) ของระบบ
- **สิ่งที่ไหลในระบบที่ต้องการรับบริการ** ซึ่งจะเรียกว่า **เอ็นทิตี้** (Entity) เช่น คนไข้, ชิ้นงาน, หรือลูกค้า เอ็นทิตี้มีอัตราการมาถึง (จำนวนต่อหนึ่งหน่วยเวลา) เป็นเท่าใด ด้วยค่าเฉลี่ยที่ค่อนข้างคงที่ตลอดเวลาที่ระบบทำการหรือไม่ ถ้าไม่ ควรสมมติว่าอัตราคงที่ในแต่ละชั่วโมง แต่ละครึ่งชั่วโมง หรือแต่ละ 15 นาที
- **ทรัพยากรที่ให้บริการเอ็นทิตี้**มีอะไรบ้าง และมีจำนวนเท่าใด มีตารางการทำงานอย่างไร
- **เวลาที่ระบบทำการ**

วัตถุประสงค์หลักของการเก็บข้อมูลคือให้มีรายละเอียดมากพอที่จะเลียนแบบพฤติกรรมของระบบจริงได้ (Halverson, 2011) เช่น ในระบบแถวคอย (Queueing system) อย่างง่ายที่มีผู้ให้บริการ (Server) 1 คน และลูกค้า 1 ชนิดที่มาถึงระบบด้วยอัตราคงที่ (หัวข้อ 1.4) ข้อมูลที่ต้องการคือ จำนวนลูกค้ามาถึงในแต่ละช่วงเวลาหรืออัตราการมาถึง (Arrival rate), เวลานาฬิกาที่ลูกค้าเริ่มรับบริการและเวลานาฬิกาที่ลูกค้าเสร็จสิ้นการบริการ เพื่อคำนวณเวลาที่ใช้บริการลูกค้า 1 คน (ไม่นับรวมเวลารอ)

เมื่อมีรายละเอียดเบื้องต้นของระบบจริงแล้ว ประเด็นปลีกย่อยอื่น ๆ ที่ควรพิจารณา เช่น

- **วันและเวลาที่จะเข้าไปเก็บข้อมูล** ระบบให้บริการบางชนิด เช่น เคาน์เตอร์ธนาคาร หรือระบบผู้ป่วยนอกของโรงพยาบาล มีลูกค้าในวันจันทร์มากกว่าวันอื่น ๆ ในสัปดาห์ ดังนั้นจึงต้องพิจารณาว่าอัตราการมาถึงของเอ็นทิตี้นั้นขึ้นอยู่กับวันของสัปดาห์ (จันทร์, อังคาร, ...) และเวลาของวัน (8.00, 9.00, ...) หรือไม่ ตัวแปรนำเข้าอื่น ๆ ก็ต้องพิจารณาในแง่นี้เช่นกัน
- **หน่วยของเวลาที่ใช้ในระหว่างเก็บข้อมูล** เช่น ชั่วโมงหรือนาที
- **ตำแหน่งของผู้เก็บข้อมูลสัมพัทธ์กับระบบ** ว่าจะประกบเอ็นทิตี้ เช่น คนไข้ แล้วไหลผ่านระบบพร้อมกัน หรือจะประจำอยู่ที่แต่ละสถานีแล้วจับเวลา ผู้เก็บข้อมูลอาจต้องทำทั้งสองวิธีเพื่อวัตถุประสงค์ที่แตกต่างกัน การพัฒนาแบบจำลองต้องการข้อมูลเพื่อกำหนดตัวแปรนำเข้าและสำหรับการตรวจสอบความถูกต้องของแบบจำลอง

ก่อนไปเก็บข้อมูลที่หน้างาน ควรออกแบบฟอร์มสำหรับเก็บข้อมูล เช่น สำหรับระบบแถวคอยอย่างง่ายในหัวข้อ 1.4 แบบฟอร์มอาจเป็นดังแสดงในตารางที่ 4.1 เมื่อได้ข้อมูลเวลาตามนาฬิกาที่เริ่มรับบริการและสิ้นสุดการบริการแล้ว นำมาหักลบกันเพื่อให้ได้เวลาในการบริการลูกค้า 1 คน แต่หากเอ็นทิตี้มาถึงระบบด้วยความถี่สูง อาจใช้วิธีการแบ่งเวลาเป็นช่วง ช่วงละ 15 นาที, 30 นาที, หรือ 1 ชั่วโมง แล้วนับจำนวนที่เอ็นทิตี้มาถึงในช่วงเวลานั้น ๆ แทนการบันทึกเวลาที่มาถึง เช่น ตารางที่ 4.2

ตารางที่ 4.1: ตัวอย่างแบบฟอร์มสำหรับเก็บข้อมูลระบบแถวคอยอย่างง่ายในหัวข้อ 1.4

ลูกค้าคนที่	**เวลาที่ลูกค้ามาถึงระบบ** (ตามนาฬิกา)	**เวลาที่เริ่มรับบริการ** (ตามนาฬิกา)	**เวลาที่สิ้นสุดการรับบริการ** (ตามนาฬิกา)
1	8.31 น.	8.31 น.	8.36 น.
2	8.32 น.	8.36 น.	8.45 น.
$\vdots$	$\vdots$	$\vdots$	$\vdots$

ตารางที่ 4.2: ตัวอย่างแบบฟอร์มสำหรับเก็บข้อมูลอัตราการมาถึง

	จำนวนที่มาถึง				**อัตราการมาถึงเฉลี่ย**
ช่วงเวลา	**วันที่ 1**	**วันที่ 2**	...	**วันที่ n**	(ต่อ 30 นาที)
8.00-8.30	21	34	$\dots$	25	(21+34 + $\dots$ + 25)/n
8.30-9.00	a_1	a_2	$\dots$	a_n	$(a_1 + a_2 + \dots + a_n)/n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

แนวทางทั่วไปในการเก็บข้อมูล มีดังนี้ (Kelton et al., 2009)

- **วิเคราะห์ข้อมูลขณะที่เก็บ** ตรวจสอบความสอดคล้อง (Consistency) และความสมจริงของข้อมูล เช่น เวลาที่ออกต้องมีค่ามากกว่าเวลาเข้า เวลาที่ใช้ที่แต่ละสถานีมีค่าสมเหตุสมผล ตรวจสอบว่ามีค่าน้อยหรือมากผิดปกติหรือไม่

- **พิจารณาว่าสามารถรวมชุดข้อมูลที่ใกล้เคียงกันหรือไม่** เช่น ช่วงเวลาเดียวกันจากหลาย ๆ วัน ดังเช่นในตารางที่ 4.2 หรือภายในวันเดียวกัน

- **ระวังการเซ็นเซอร์ตัวเองของข้อมูล** (Data censoring) หมายถึง กรณีที่ไม่สามารถเห็นข้อมูลบางส่วนเพราะระยะเวลาที่สังเกตไม่ยาวพอ เช่น หากเก็บข้อมูลเป็นเวลา 5 ปี ก็จะไม่เห็นการเสียของชิ้นงานที่มีอายุมากกว่า 5 ปี

- **ตรวจสอบความสัมพันธ์ระหว่างตัวแปรนำเข้า** เช่น ที่เคาน์เตอร์ธนาคาร เวลาในการทำธุรกรรมแปรผันตรงกับจำนวนธุรกรรม แนวโน้มหรือความสัมพันธ์นี้อาจปรากฏชัดหากสร้างแผนภาพกระจาย (Scatter diagram) เช่น ให้จำนวนธุรกรรมอยู่บนแกนนอน และเวลาอยู่บนเป็นแกนตั้ง

- **ตรวจสอบค่าสหสัมพันธ์ภายในตัวเอง** เช่น คำสั่งซื้อขายหุ้นมักจะเกิดขึ้นตาม ๆ กัน คือ เวลาที่เกิดขึ้นไม่เป็นอิสระต่อกัน หากผู้วิเคราะห์สมมติว่าเป็นอิสระต่อกัน ก็อาจทำให้ผลการวิเคราะห์คลาดเคลื่อน

ไปมาก

- **แยกแยะว่าตัวแปรใดเป็นตัวแปรนำเข้า และตัวแปรใดเป็นดัชนีชี้วัด** เช่น เวลาในการให้บริการที่แต่ละสถานีในระบบเป็นตัวแปรนำเข้า แต่เวลารอคอยในระบบและเวลารวมในระบบเป็นเอาท์พุทและดัชนีชี้วัด บทนี้เน้นพิจารณาตัวแปรนำเข้า แต่การศึกษาด้วยการจำลองสถานการณ์ต้องใช้เอาท์พุทจากหน้างานจริงด้วยสำหรับตรวจสอบความถูกต้องของแบบจำลอง (หัวข้อที่ 9.3)

เมื่อได้ข้อมูลมาแล้ว ควรวิเคราะห์ทางสถิติเบื้องต้น: ด้วยแผนภาพ เช่น ฮีสโตแกรม, กราฟอนุกรมเวลา; หรือสรุปค่าทางสถิติ เช่น คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (รายละเอียดในหัวข้อ 2.5)

4.2 การจำลองตัวแปรนำเข้าเมื่อมีข้อมูล

ขั้นตอนการเลือกการแจกแจงความน่าจะเป็นสำหรับข้อมูลเชิงตัวเลข มีดังนี้

1. **เลือกการแจกแจงความน่าจะเป็นหลาย ๆ ชนิดที่น่าจะสอดคล้องกับข้อมูล** โดยพิจารณาจากลักษณะทางกายภาพของขบวนการสุ่มที่เก็บข้อมูลมา (รายละเอียดในหัวข้อ 4.2.1) และพิจารณาฮีสโตแกรมของข้อมูล (หัวข้อ 2.5.2) ว่ามีลักษณะใกล้เคียงกับกราฟความหนาแน่นความน่าจะเป็น (PDF) ของการแจกแจงใด ฮีสโตแกรมเทียบเคียงได้กับกราฟ PDF จากกลุ่มตัวอย่าง

2. **ปรับการแจกแจงความน่าจะเป็นให้สอดคล้องกับข้อมูลที่มี** (Distribution fitting) คือ การกำหนดค่าพารามิเตอร์ของการแจกแจงจากข้อมูลจริง เช่น การแจกแจงแบบปกติมีพารามิเตอร์ 2 ตัว คือ ค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ ซึ่งประมาณจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจากกลุ่มตัวอย่าง: สมการที่ (2.16) และ (2.17)

3. **ตรวจสอบว่าการแจกแจงที่เลือกสอดคล้องกับข้อมูลจริงหรือไม่** โดยการทดสอบสมมุติฐาน (หัวข้อ 4.3.1) และพิจารณาเปรียบเทียบฮีสโตแกรมกับกราฟ PDF (หัวข้อ 4.3.2)

4. **ถ้าการแจกแจงที่เลือกไม่สอดคล้อง อาจเลือกการแจกแจงอื่นแล้วทำซ้ำข้อ 2-3** หรืออาจพิจารณาใช้ข้อมูลจริงเพื่อสร้างเลขสุ่มโดยตรงด้วย**การแจกแจงเชิงประสบการณ์**หรือเรียกอีกชื่อหนึ่งว่า**การแจกแจงตามตัวอย่าง** (รายละเอียดในหัวข้อ 4.2.16)

ผู้วิเคราะห์ควรพิจารณาการแจกแจงความน่าจะเป็นที่อาจเหมาะสมกับข้อมูลโดยพิจารณาพื้นฐานทางกายภาพของการแจกแจง

4.2.1 พื้นฐานทางกายภาพของการแจกแจงความน่าจะเป็น

การแจกแจงความน่าจะเป็นที่มีพารามิเตอร์ (Parametric distribution) แต่ละชนิดได้รับการคิดค้นเพื่อเป็นตัวแทนของสถานการณ์ทางกายภาพเฉพาะเจาะจงบางอย่าง เช่น การแจกแจงแบบปกติมีชื่อเช่นนั้นเพราะเป็นลักษณะที่พบได้ทั่วไปเมื่อมีข้อมูลจำนวนมาก เช่น น้ำหนัก, ส่วนสูง, หรือระดับไอคิวของประชากร เกรดเฉลี่ยของนักศึกษาทั้งชั้นปีที่มีหลายร้อยคน ดังนั้นถ้ารู้พื้นฐานทางกายภาพของการแจกแจงต่าง ๆ ผู้วิเคราะห์สามารถจับคู่การแจกแจงความน่าจะเป็นให้ตรงกับข้อมูลที่ต้องการจำลองได้ ในขั้นตอนการเลือกนี้ คำถามเบื้องต้นที่ควรถาม คือ

- **ผล** (Outcome) ที่เป็นไปได้มีอะไรบ้าง
- **ผลเหล่านี้มีลักษณะอย่างไร** เป็นค่าต่อเนื่องหรือค่าไม่ต่อเนื่อง ช่วงของผลหรือค่าที่เป็นไปได้จำกัดหรือไม่จำกัด

การตอบคำถามนี้ช่วยลดทอนจำนวนตัวเลือกที่ต้องพิจารณา ดังนี้

- **ค่าไม่ต่อเนื่องและอยู่ในช่วงหรือพิสัย (Range) ที่จำกัด** เช่น จำนวนพนักงานที่ขาดงานจากทั้งหมด 120 คนใน 1 กะ, จำนวนชิ้นงานที่ไม่ผ่านการตรวจคุณภาพใน 1 ล็อตที่มีจำนวน 1,000 ชิ้น, จำนวนของรถบัสที่พร้อมใช้การได้ในวันจันทร์ (ขอบเขตบนคือจำนวนรถบัสที่มี และขอบเขตล่างคือศูนย์)

  การแจกแจงความน่าจะเป็นที่เข้าข่ายนี้ เช่น ทวินาม, สม่ำเสมอแบบไม่ต่อเนื่อง, พหุนามหรือไม่ต่อเนื่อง (Discrete), ไฮเปอร์จีโอเมตริก (Hypergeometric)

- **ค่าไม่ต่อเนื่องและอยู่ในช่วงหรือพิสัยที่ไม่จำกัด** (รวมถึงกรณีที่อาจมีขอบเขตใหญ่มากและไม่ทราบค่า) เช่น อุปสงค์ของโทรศัพท์มือถือรุ่นใหม่, จำนวนอุบัติเหตุทางถนนทั่วประเทศที่เกิดขึ้นใน 1 ปี, จำนวนชิ้นงานที่ผลิตได้จนกว่าเครื่องจักรจะติดขัด

  การแจกแจงความน่าจะเป็นที่เข้าข่ายนี้ เช่น เรขาคณิต, ปัวซอง (Poisson)

- **ค่าต่อเนื่องและอยู่ในช่วงหรือพิสัยที่จำกัด** เช่น เส้นผ่านศูนย์กลางของชิ้นงาน (ขนาดที่ผลิตได้ถูกจำกัดด้วยเงื่อนไขทางกายภาพ), ระยะเวลาทำงานที่เพิ่มมูลค่าใน 1 วัน, การเปลี่ยนแปลงของอัตราดอกเบี้ยเงินฝากในเดือนหน้า

  การแจกแจงความน่าจะเป็นที่เข้าข่ายนี้ เช่น เบต้า (Beta), เพิร์ท (Pert), สามเหลี่ยม (Triangular), สม่ำเสมอแบบต่อเนื่อง

- **ค่าต่อเนื่อง มีขอบเขตล่างที่ตรึงไว้แต่ขอบเขตบนไม่จำกัด** เช่น เวลาในการก่อสร้างอาคาร, เวลาที่เครื่องจักรทำงานได้ก่อนที่จะเสีย, สังเกตว่าข้อมูลช่วงเวลามีขอบเขตล่างที่ชัดเจนคือศูนย์ เพราะระยะเวลาไม่ติดลบ

การแจกแจงความน่าจะเป็นที่เข้าข่ายนี้ เช่น เอ็กซ์โพเนนเชียล (Exponential), แกมมา (Gamma), และล็อกนอร์มอล (Lognormal)

หัวข้อย่อยต่อไปนี้จะอธิบายถึงการแจกแจงความน่าจะเป็นแต่ละแบบ (Banks et al., 2005) ให้สังเกตว่าค่า PDF , $f(x)$, จะเป็นบวกเฉพาะสำหรับค่าในช่วงที่เป็นไปได้เท่านั้น ($x \in \Phi_X$) ส่วนที่ค่าอื่น ๆ ($x \notin \Phi_X$) $f(x) = 0$

4.2.2 การแจกแจงแบบทวินาม

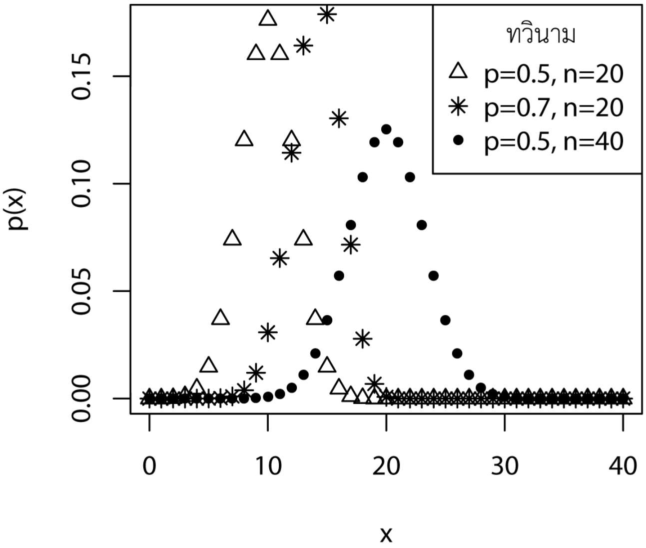
การแจกแจงแบบทวินาม (Binomial distribution) จำลอง**จำนวนความสำเร็จ**จากการสุ่ม n ครั้ง เมื่อการสุ่มแต่ละครั้งเป็นอิสระต่อกันและมีความน่าจะเป็นของความสำเร็จในแต่ละครั้งเท่ากันคือ p *ความสำเร็จ*คือเมื่อเหตุการณ์ที่สนใจเกิดขึ้น ไม่จำเป็นต้องเป็นเรื่องดีเสมอไป เช่น จำนวนชิ้นงานที่เสียใน 1 ล็อตที่มีชิ้นงานทั้งหมด n ชิ้น; จำนวนผู้โดยสารที่มาขึ้นเครื่องไม่ทันจากจำนวนตั๋วที่ขายไป n ใบใน 1 เที่ยวบินโดยสมมติว่าผู้โดยสารแต่ละคนมีความน่าจะเป็นที่จะตกเครื่องเท่ากันคือ p และสมมติว่าเหตุการณ์ที่ผู้โดยสารแต่ละคนจะตกเครื่องเป็นอิสระต่อกัน ซึ่งเห็นได้ชัดว่าเป็นการประมาณความจริง เพราะผู้โดยสารบางส่วนเดินทางเป็นคู่หรือเป็นกลุ่ม (แบบจำลองในตารางที่ 3.9 มี $p = 0.1$ และ n คือ จำนวนผู้โดยสารที่ซื้อตั๋วในเที่ยวบินนั้น); จำนวนโครงการลงทุนที่ไม่ประสบความสำเร็จจากทั้งหมด n โครงการ; รูปที่ 4.1 แสดงฟังก์ชันมวลความน่าจะเป็น (PMF) ที่ชุดค่าพารามิเตอร์ต่าง ๆ ตารางที่ 4.3 แสดงคุณสมบัติของการแจกแจง สัญลักษณ์ $\lfloor x \rfloor$ แสดงค่าที่ปัดลงเป็นจำนวนเต็มที่ใกล้ที่สุด เช่น $\lfloor 3.55 \rfloor = 3$ และสัญลักษณ์ $\lceil x \rceil$ แสดงค่าที่ปัดขึ้นเป็นจำนวนเต็มที่ใกล้ที่สุด เช่น $\lceil 3.55 \rceil = 4$ นิยามจำนวนที่เกิดขึ้นได้ทั้งหมด (Combination)

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

เมื่อสุ่มเพียง 1 ครั้ง ($n = 1$) การแจกแจงนี้มีชื่อเฉพาะว่า **การแจกแจงแบบเบอร์นูลลี** (Bernoulli distribution)

4.2.3 การแจกแจงแบบทวินามลบ

การแจกแจงแบบทวินามลบ (Negative binomial distribution) หรือการแจกแจงแบบ**ปาสคาล** (Pascal distribution) จำลองจำนวนครั้งที่สุ่มจนกว่าจะได้จำนวนครั้งที่สำเร็จครบ r ครั้ง เมื่อการสุ่มแต่ละครั้งเป็นอิสระต่อกันและมีความน่าจะเป็นของความสำเร็จเท่ากันคือ p เช่นเดียวกับการแจกแจงแบบทวินาม *ความสำเร็จ*คือเมื่อเหตุการณ์ที่สนใจเกิดขึ้น เช่น เมื่อมีชิ้นงานเสียเกิดขึ้น และต้องการจำลองจำนวนชิ้นส่วนที่ตรวจสอบจนกว่าจะพบชิ้นเสีย 4 ชิ้น สามารถใช้ตัวแปรสุ่มแบบทวินามลบโดยที่ $r = 4$ และ p คือความน่าจะเป็นที่

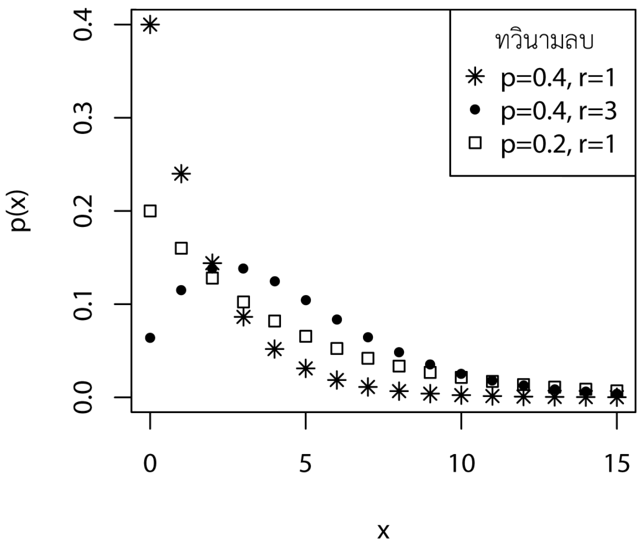


รูปที่ 4.1: PMF ของการแจกแจงแบบทวินาม

ตารางที่ 4.3: คุณสมบัติของการแจกแจงแบบทวินาม

ค่าสถิติ	**ทวินาม**	**ฟังก์ชันเอกซ์เซล**
พารามิเตอร์	n จำนวนเต็มบวก และ $0 \leqslant p \leqslant 1$	
ช่วงของค่าที่เป็นได้ (Support, Φ)	$\{0, 1, 2, \ldots, n\}$	
PMF $p(x)$	$\binom{n}{x} p^x (1-p)^{n-x}$	`BINOMDIST(`x, n, p `FALSE)`
CMF, $F(x)$	$\sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$	`BINOMDIST(`x, n, p `TRUE)`
ค่าเฉลี่ย	np	
ความแปรปรวน	$np(1-p)$	

จะสุ่มชิ้นงานแล้วพบชิ้นเสีย รูปที่ 4.2 เป็นกราฟของ PMF ตารางที่ 4.4 แสดงคุณสมบัติของการแจกแจง เมื่อสนใจจำนวนครั้งที่สำเร็จเพียง 1 ครั้ง ($r = 1$) การแจกแจงที่ได้มีอีกชื่อเฉพาะว่า **การแจกแจงแบบเรขาคณิต** (Geometric distribution)



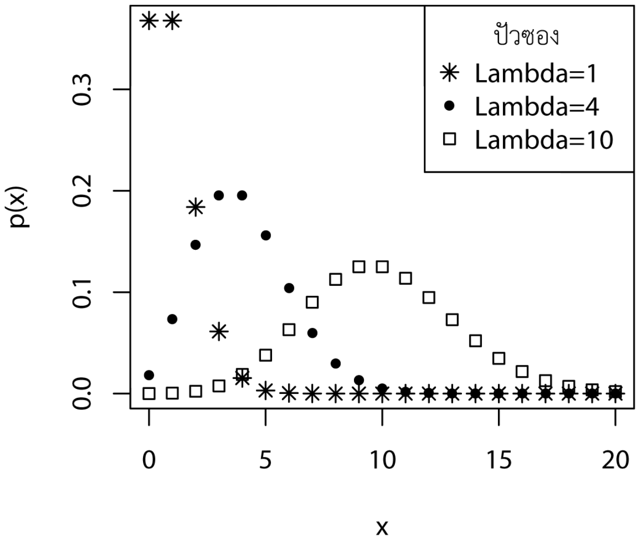
รูปที่ 4.2: PMF ของการแจกแจงแบบทวินามลบ

ตารางที่ 4.4: คุณสมบัติของการแจกแจงแบบทวินามลบ

ค่าสถิติ	**ทวินามลบ**	**ฟังก์ชันเอกซ์เซล**
พารามิเตอร์	r จำนวนเต็มบวก และ $0 \leqslant p \leqslant 1$	
ช่วงของค่าที่เป็นได้ (Φ)	$\{r, r+1, r+2, \ldots\}$	
PMF $p(x)$	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$,	NEGBINOMDIST($x-r, r, p$ FALSE)
CMF, $F(x)$	$\sum_{i=0}^{\lfloor x \rfloor} \binom{x+r-1}{x} p^x (1-p)^r$	NEGBINOMDIST($x-r, r, p$ TRUE)
ค่าเฉลี่ย	r/p	
ความแปรปรวน	$r(1-p)/p^2$	

4.2.4 การแจกแจงแบบปัวซอง

การแจกแจงแบบปัวซอง (Poisson distribution) จำลองจำนวนครั้งที่*เหตุการณ์*ที่สนใจเกิดขึ้นภายในระยะเวลา, ระยะทาง, ขนาดพื้นที่, หรือปริมาตรหนึ่ง ๆ ที่กำหนดค่าเอาไว้แน่นอน เหตุการณ์นี้เป็นอิสระต่อกัน คือไม่มีอิทธิพลต่อกันและเกิดขึ้นทีละครั้ง เช่น ลูกค้าเข้ามาทีละ 1 คนหรือทีละกลุ่ม (ถ้าลูกค้าเข้ามาเป็นกลุ่ม ก็จะจำลองการเข้ามาของกลุ่มลูกค้าด้วยการแจกแจงปัวซอง และจำลองจำนวนลูกค้าในกลุ่มด้วยการแจกแจงแบบพหุนาม) ช่วงเวลาที่เหตุการณ์เกิดขึ้นครั้งล่าสุดไม่ส่งผลต่อเวลาที่จะเกิดครั้งถัดไป เช่น จำนวนลูกค้าที่มาถึงร้านภายใน 1 ชั่วโมง (λ เป็นจำนวนลูกค้าเฉลี่ยที่เข้าร้านใน 1 ชม.) จำนวนตำหนิที่เกิดขึ้นบนสายเคเบิลยาว 30 เมตร (λ เป็นจำนวนตำหนิเฉลี่ยบนสายเคเบิลยาว 30 เมตร) หรือจำนวนตำหนิของงานพ่นสีบนประตูรถยนต์ 1 บาน (λ เป็นจำนวนตำหนิเฉลี่ยบนประตูรถยนต์ 1 บาน) รูปที่ 4.3 เป็นกราฟของ PMF ของการแจกแจงแบบปัวซอง ตารางที่ 4.5 แสดงคุณสมบัติของการแจกแจงแบบปัวซอง



รูปที่ 4.3: PMF ของการแจกแจงแบบปัวซอง

4.2.5 การแจกแจงแบบปัวซองที่อัตราไม่คงที่

ในเชิงปฏิบัติสำหรับระบบจริง เมื่อพิจารณาระบบที่เวลาต่าง ๆ อัตราการมาถึงมักไม่คงที่ในแต่ละช่วงเวลา เช่น ที่ศูนย์โทรศัพท์ของบัตรเครดิต ลูกค้าโทรเข้ามาถี่ในตอนเย็นหรือตอนค่ำแต่ไม่มากนักตอนกลางดึกหรือเช้ามืด ผู้วิเคราะห์สามารถใช้กระบวนการปัวซองแบบอัตราไม่คงที่ (Non-stationary Poisson process) เพื่อจำลองระบบนี้ ดังนั้นอัตราการเกิดเหตุการณ์หรืออัตราการมาถึงจึงขึ้นอยู่กับเวลาหรือแสดงด้วยฟังก์ชัน $\lambda(t)$

ตารางที่ 4.5: คุณสมบัติของการแจกแจงแบบปัวซอง

ค่าสถิติ	ปัวซอง	ฟังก์ชันเอกซ์เซล
พารามิเตอร์	$\lambda > 0$	
ช่วงของค่าที่เป็นได้ (Φ)	$\{0, 1, 2, \ldots\}$	
PMF $p(x)$	$e^{-\lambda}\lambda^x/x!$	POISSON(x, λ FALSE)
CMF, $F(x)$	$\sum_{i=0}^{\lfloor x \rfloor} \frac{e^{-\lambda}\lambda^x}{x!}$	POISSON(x, λ TRUE)
ค่าเฉลี่ย	λ	
ความแปรปรวน	λ	

ข้อสมมติของการแจกแจงนี้เหมือนกับการแจกแจงแบบปัวซองคือเหตุการณ์เกิดขึ้นทีละครั้ง และเหตุการณ์ที่เกิดขึ้นภายในระยะเวลาที่ไม่ทับกันเป็นอิสระต่อกัน กำหนดให้ $N(t)$ เป็นจำนวนครั้งที่เหตุการณ์เกิดขึ้นตั้งแต่เวลา 0 ถึงเวลา t และ

$$b(t, s) = \int_t^{t+s} \lambda(y)dy$$

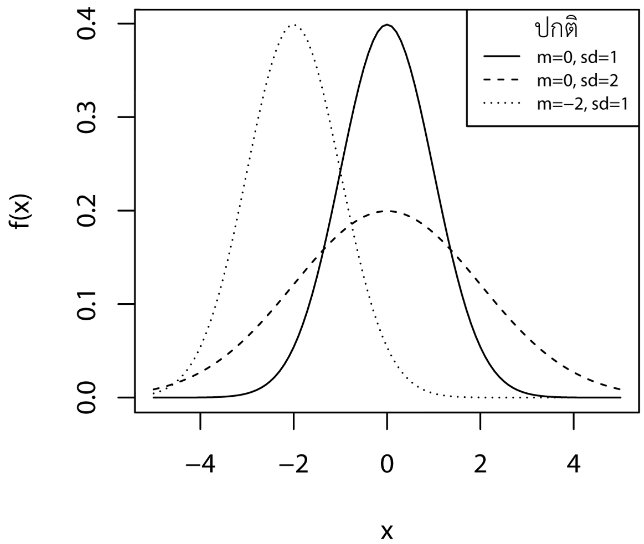
เป็นจำนวนครั้งที่เหตุการณ์เกิดขึ้นโดยเฉลี่ยในช่วงเวลา $(t, t+s]$ ฟังก์ชันความน่าจะเป็น คือ

$$\Pr\{N(t+s) - N(t) = k\} = \frac{e^{-b(t,s)}[b(t, s)]^k}{k!}$$

ซอฟต์แวร์สำเร็จรูปเช่นอารีน่า ให้ผู้ใช้กำหนด $\lambda(t)$ ได้

4.2.6 การแจกแจงแบบปกติ

การแจกแจงแบบปกติ (Normal distribution) จำลองกระบวนการที่สามารถมองว่าเป็น**ผลบวก**ของกระบวนการย่อย เช่น เวลารวมที่ใช้ในการประกอบผลิตภัณฑ์ต่าง ๆ เพราะเป็นผลบวกของเวลาในแต่ละขั้นตอน, ค่าคลาดเคลื่อน (Error) ในค่าที่สังเกตได้เมื่อทำการทดลอง, และค่าผิดพลาดอื่น ๆ เช่น เวลาส่งมอบที่ช้าหรือเร็วกว่ากำหนด ดังนั้นจึงใช้จำลองตัวแปรที่รู้เพียงค่าเฉลี่ยและความแปรปรวนรอบค่าเฉลี่ย เช่น จำลองเวลาในการส่งมอบที่มีค่าเฉลี่ย 4 $\pm$ 1 วัน ด้วยการแจกแจงแบบปกติที่มี $\mu = 4$ และ $\sigma = 1$ วัน นอกจากนี้ยังใช้ประมาณค่าความน่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่องบางชนิด เช่น ทวินามและปัวซอง (เมื่อขนาดตัวอย่าง n และอัตราการเกิดเหตุการณ์ λ มีค่ามาก) รูปที่ 4.4 เป็นกราฟของ PDF จะเห็นว่า μ แสดงตำแหน่งกึ่งกลางของกราฟ PDF และ σ แสดงความกว้างของระฆังคว่ำ ถ้า σ สูง การกระจายตัวของข้อมูลมาก ระฆังจะกว้างและเตี้ย ตารางที่ 4.6 แสดงคุณสมบัติของการแจกแจง



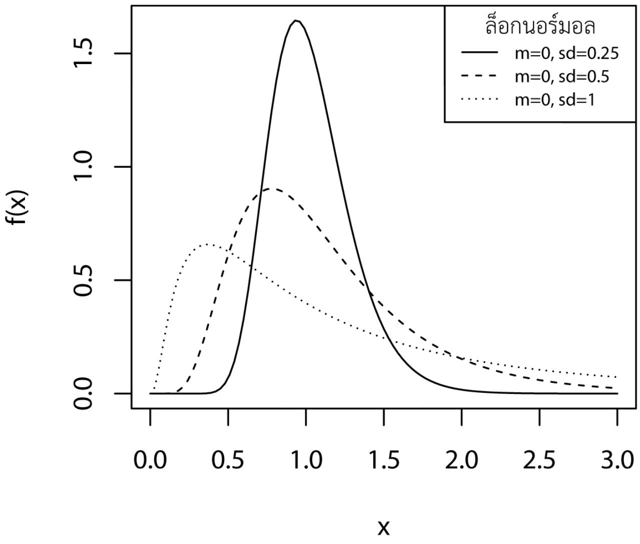
รูปที่ 4.4: กราฟ PDF ของการแจกแจงแบบปกติ

ตารางที่ 4.6: คุณสมบัติของการแจกแจงแบบปกติ

ค่าสถิติ	**ปกติ**	**ฟังก์ชันเอกซ์เซล**
พารามิเตอร์	μ จำนวนจริง และ $\sigma > 0$	
ช่วงของค่าที่เป็นได้ (Φ)	$(-\infty, \infty)$	
PDF $f(x)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	`NORMDIST(`x, μ, σ`, FALSE)`
CDF, $F(x)$	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-t^2}{2}\right) dt$ โดยที่ $z = \frac{x-\mu}{\sigma}$	`NORMDIST(`x, μ, σ`, TRUE)`
ค่าเฉลี่ย	μ	
ความแปรปรวน	σ^2	

4.2.7 การแจกแจงแบบล็อกนอร์มอล

การแจกแจงแบบล็อกนอร์มอล (Lognormal distribution) จำลองค่าที่สามารถมองได้ว่าเป็น**ผลคูณ**ของกระบวนการย่อย เช่น อัตราผลตอบแทนจากการลงทุนเมื่อดอกเบี้ยทบต้น เงินฝากธนาคาร นอกจากนี้ยังใช้จำลองผลตอบแทนจากการลงทุนในหุ้นสามัญ ชื่อของการแจกแจงนี้แสดงว่ามีความเกี่ยวข้องกับการแจกแจงแบบปกติ กล่าวคือ ถ้า Y มีการแจกแจงแบบล็อกนอร์มอล $\ln(Y)$ จะมีการแจกแจงแบบปกติ ในทางกลับกัน ถ้า X มีการแจกแจงแบบปกติ $\exp(X)$ จะมีการแจกแจงแบบล็อกนอร์มอล รูปที่ 4.5 เป็นกราฟของ PDF ตารางที่ 4.7 แสดงคุณสมบัติของการแจกแจง



รูปที่ 4.5: กราฟ PDF ของการแจกแจงแบบล็อกนอร์มอล

4.2.8 การแจกแจงแบบเอ็กซ์โพเนนเชียล

การแจกแจงแบบเอ็กซ์โพเนนเชียล (Exponential distribution) จำลองเวลาระหว่างเหตุการณ์ที่เป็นอิสระต่อกัน หรือเวลาของกระบวนการที่จำอดีตไม่ได้ (Memoryless) หมายถึง กระบวนการที่แม้ว่าจะรู้เวลาที่ผ่านไปแล้ว ก็ไม่สามารถระบุว่าได้ว่าจะเสร็จสิ้นเมื่อใด เช่น หากจำลองอายุการใช้งาน (Time to failure) ของอุปกรณ์ด้วยการแจกแจงแบบเอ็กซ์โพเนนเชียล นัยคืออุปกรณ์นี้ที่ใช้งานไปแล้วและยังไม่เสียมีการแจกแจงของอายุการใช้งานที่เหลือเหมือนอุปกรณ์ใหม่แกะกล่อง เมื่อจำลองอายุการใช้งานด้วยการแจกแจงแบบเอ็กซ์โพเนนเชียล สิ่งที่สมมติคือมีอัตราการเกิดความล้มเหลว (Failure rate) คงที่เมื่อเวลาดำเนินไป ให้สังเกตว่าการ

ตารางที่ 4.7: คุณสมบัติของการแจกแจงแบบล็อกนอร์มอล

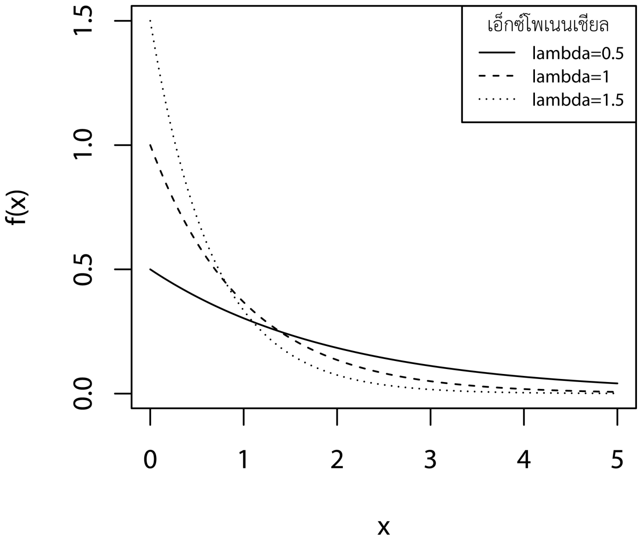
ค่าสถิติ	ล็อกนอร์มอล	ฟังก์ชันเอกซ์เซล
พารามิเตอร์	μ และ $\sigma > 0$	
ช่วงของค่าที่เป็นได้ (Φ)	$[0, \infty)$	
PDF $f(x)$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$	ไม่มี
CDF, $F(x)$	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt$ โดยที่ $z = \frac{\ln(x)-\mu}{\sigma}$	LOGNORMDIST(x, μ, σ)
ค่าเฉลี่ย	$e^{\mu+\sigma^2/2}$	
ความแปรปรวน	$(e^{\sigma^2}+1)e^{2\mu+\sigma^2}$	

แจกแจงแบบเอ็กซ์โพเนนเชียลและปัวซองสัมพันธ์กัน หากเวลาระหว่างที่เหตุการณ์เกิดขึ้นมีการแจกแจงแบบเอ็กซ์โพเนนเชียล จำนวนครั้งที่เหตุการณ์เกิดขึ้นภายในช่วงเวลาที่ตรึงไว้ค่าหนึ่ง (เช่น 1 ชั่วโมง) มีการแจกแจงแบบปัวซอง

มีการนำการแจกแจงแบบเอ็กซ์โพเนนเชียลไปใช้ในแบบจำลองหลายชนิดถึงแม้ว่าข้อสมมติเรื่องอัตราการเกิดเหตุการณ์เป็นค่าคงที่แทบไม่เป็นจริงในทางปฏิบัติ เพราะการแจกแจงแบบเอ็กซ์โพเนนเชียลทำให้ตัวแบบจัดการได้ง่ายทางคณิตศาสตร์ ข้อจำกัดเรื่องอัตราการเกิดคงที่สามารถแก้ไขด้วยการแบ่งระยะเวลาเป็นช่วงสั้นพอที่จะสมมติว่าอัตราคงที่ในช่วงนั้น ๆ ได้ เช่น ช่วง 10-11 น. มีอัตรา 14 คน จึงกำหนดให้ λ_1 = 14 ต่อชั่วโมง; ช่วง 11-12 น. มีอัตรา 18 คน λ_2 = 18 ต่อชั่วโมง และอื่น ๆ; หากเวลาระหว่างการโทรเข้าของลูกค้าเป็นแบบเอ็กซ์โพเนนเชียล เวลาจนกว่าลูกค้าจะโทรเข้าครั้งต่อไปก็มีการแจกแจงแบบเอ็กซ์โพเนนเชียลเช่นกัน รูปที่ 4.6 แสดงกราฟ PDF ที่อัตราการเกิดเหตุการณ์ λ ต่าง ๆ ค่า λ สูงหมายถึงเหตุการณ์เกิดขึ้นถี่โดยเฉลี่ย สังเกตว่าหน่วยของ λ และ x นั้นสัมพันธ์กัน กล่าวคือเมื่อนำมาคูณกันแล้วผลคูณ λx ไม่มีหน่วย เช่น ถ้า λ มีหน่วย “ต่อชั่วโมง” x ต้องมีหน่วย “ชั่วโมง” คุณสมบัติอื่น ๆ แสดงในตารางที่ 4.8

4.2.9 การแจกแจงแบบแกมมา

การแจกแจงแบบแกมมา (Gamma distribution) เป็นการแจกแจงที่มีความยืดหยุ่น (Flexibility) สูงมาก กล่าวคือมีพารามิเตอร์แค่ 2 ตัวแต่มีรูปแบบของกราฟ PDF ได้หลากหลาย เบ้ซ้าย เบ้ขวา หรือสมมาตร (ดังแสดงในรูปที่ 4.7) มักใช้การแจกแจงแบบแกมมาจำลองค่าที่ไม่ติดลบ เช่น เวลาในการทำงานใดงานหนึ่งให้เสร็จ อาทิ การบริการลูกค้าหรือการซ่อมเครื่องจักร ตารางที่ 4.9 แสดงคุณสมบัติของการแจกแจง กำหนดให้ k คือพารามิเตอร์เกี่ยวกับรูปทรง (Shape parameter) θ เป็นพารามิเตอร์เกี่ยวกับขนาด (Scale parameter)



รูปที่ 4.6: กราฟ PDF ของการแจกแจงแบบเอ็กซ์โพเนนเชียล

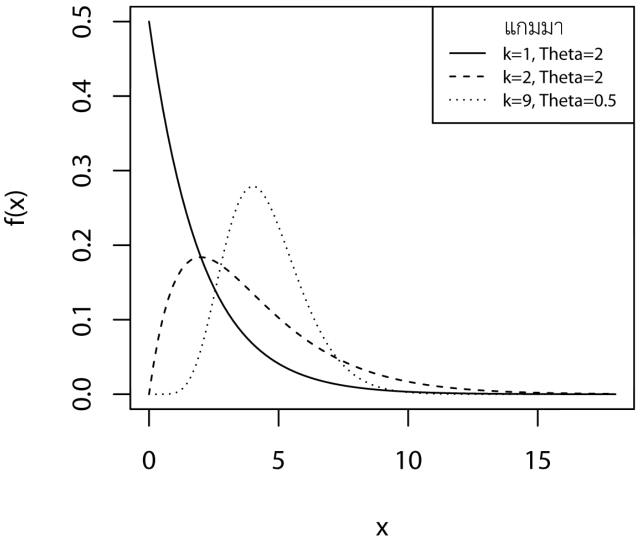
ตารางที่ 4.8: คุณสมบัติของการแจกแจงแบบเอ็กซ์โพเนนเชียล

ค่าสถิติ	เอ็กซ์โพเนนเชียล	ฟังก์ชันเอ็กซ์เซล
พารามิเตอร์	$\lambda > 0$	
ช่วงของค่าที่เป็นไปได้ (Φ)	$[0, \infty)$	
PDF $f(x)$	$\lambda e^{-\lambda x}$	EXPONDIST(x, λ, FALSE)
CDF, $F(x)$	$1 - e^{-\lambda x}$	EXPONDIST(x, λ, TRUE)
ค่าเฉลี่ย	$1/\lambda$	
ความแปรปรวน	$(1/\lambda)^2$	

และ $k, \theta > 0$ นิยามฟังก์ชันแกมมาในตารางที่ 4.9 ดังนี้

$$\Gamma(k) = \begin{cases} (k-1)! & \text{ถ้า } k \text{ เป็นจำนวนเต็มบวก} \\ \int_0^\infty e^{-t} t^{k-1} dt & \text{ถ้า } k \text{ เป็นจำนวนจริงบวก} \end{cases} \tag{4.1}$$

$$\gamma(k, x) = \int_0^x t^{k-1} e^{-t} dt \tag{4.2}$$



รูปที่ 4.7: กราฟ PDF ของการแจกแจงแบบแกมมา

ตารางที่ 4.9: คุณสมบัติของการแจกแจงแบบแกมมา

ค่าสถิติ	แกมมา	ฟังก์ชันเอกซ์เซล
พารามิเตอร์	$\theta, k > 0$	
ฐาน (Φ)	$[0, \infty)$	
PDF, $f(x)$	$x^{k-1} \frac{e^{-x/\theta}}{\Gamma(k)\theta^k}$	GAMMADIST(x, k, θ, FALSE)
CDF, $F(x)$	$\frac{\gamma(k, x/\theta)}{\Gamma(k)}$	GAMMADIST(x, k, θ, TRUE)
ค่าเฉลี่ย	$k\theta$	
ความแปรปรวน	$k\theta^2$	

4.2.10 การแจกแจงแบบเออร์แลง

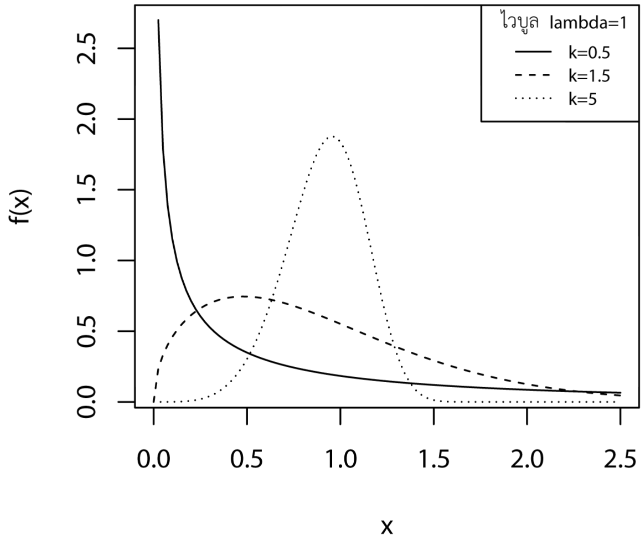
การแจกแจงแบบเออร์แลง (Erlang distribution) จำลองผลบวกของตัวแปรสุ่มแบบเอ็กซ์โพเนนเชียล k ตัวที่เป็นอิสระต่อกันและแต่ละตัวมีค่าเฉลี่ยเท่ากัน แทนด้วย $1/\lambda$ เช่น สมมติว่าเวลาระหว่างการโทรเข้าของลูกค้ามีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย 3 นาที เวลาระหว่างการโทรเข้ามาทุก ๆ 10 สายมีการแจกแจงแบบเออร์แลง ($k = 10, \lambda = 1/3$ นาที) การแจกแจงแบบเออร์แลงเป็นกรณีพิเศษของการแจกแจงแบบแกมมา คือ พารามิเตอร์ k ของการแจกแจงแบบแกมมาเป็นจำนวนจริงบวก แต่ของการแจกแจงแบบเออร์แลงเป็นจำนวนเต็มบวก

4.2.11 การแจกแจงแบบไวบูล

การแจกแจงแบบไวบูล (Weibull distribution) มักใช้เพื่อจำลองอายุการใช้งานที่การเสียอาจมาจากหลายสาเหตุ และอุปกรณ์นั้นเสียทันทีเมื่อเหตุใดก็ได้เกิดขึ้น เช่น บอยเลอร์ (Boiler) หรือหม้อไอน้ำจะเสียเมื่อท่อเล็ก ๆ (Tube) ภายในอุดตันมากพอ การแจกแจงแบบไวบูลสามารถจำลองอายุการใช้งานที่อัตราการเกิดความล้มเหลวเพิ่มขึ้น, เท่าเดิม, หรือลดลงได้ เช่น อายุการใช้งานของดิสก์ไดรฟ์ที่กราฟอัตราการเกิดความล้มเหลวกับเวลาเป็นรูปอ่างอาบน้ำ (Bathtub) คือ ลดลงในช่วงแรกของการใช้งาน ค่อนข้างคงที่หลังจากใช้งานไปได้ช่วงหนึ่ง และเพิ่มสูงขึ้นเนื่องจากการสึกหรอที่เกิดจากการใช้งาน นอกจากนี้ธุรกิจประกันภัยยังใช้การแจกแจงแบบไวบูลเพื่อจำลองมูลค่าเคลมประกัน รูปที่ 4.8 แสดงกราฟ PDF ซึ่งคล้ายกับการแจกแจงแบบแกมมา กำหนดให้ k เป็นพารามิเตอร์เกี่ยวกับรูปทรง และ λ เป็นพารามิเตอร์เกี่ยวกับขนาด ตารางที่ 4.10 แสดงรายละเอียดของการแจกแจงแบบไวบูล

ตารางที่ 4.10: คุณสมบัติของการแจกแจงแบบไวบูล

ค่าสถิติ	ไวบูล	ฟังก์ชันเอกซ์เซล
พารามิเตอร์	$\lambda, k > 0$	
ฐาน (Φ)	$[0, \infty)$	
PDF, $f(x)$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{(-x/\lambda)^k}$	WEIBULL(x, k, λ, FALSE)
CDF, $F(x)$	$1 - e^{(-x/\lambda)^k}$	WEIBULL(x, k, λ, TRUE)
ค่าเฉลี่ย (μ)	$\lambda\Gamma(1 + 1/k)$	
ความแปรปรวน	$\lambda^2\Gamma(1 + 2/k) - \mu^2$	



รูปที่ 4.8: กราฟ PDF ของการแจกแจงแบบไวบูล

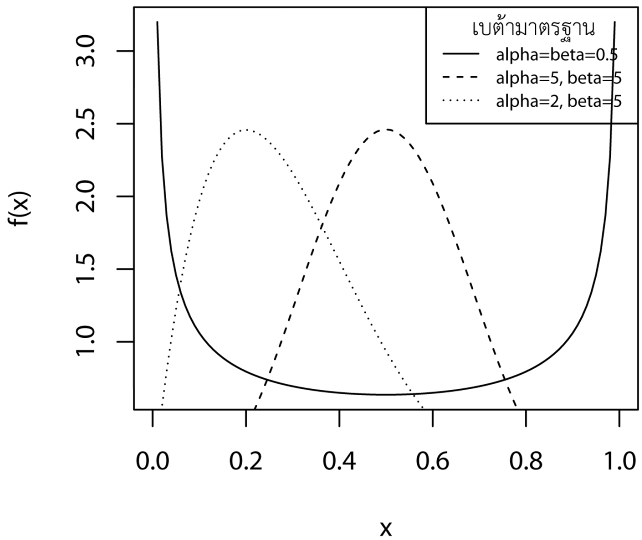
4.2.12 การแจกแจงแบบเบต้า

การแจกแจงแบบเบต้า (Beta distribution) เป็นการแจกแจงที่มีความยืดหยุ่นสูงเช่นกัน ใช้สำหรับจำลองข้อมูลที่มีขอบเขตบนและล่างตรึงไว้ นิยามด้วยพารามิเตอร์ 4 ตัว ดังนี้: α, β เป็นพารามิเตอร์เกี่ยวกับรูปทรง, a เป็นขอบเขตล่าง, และ b เป็นขอบเขตบน รูปที่ 4.9 แสดงกราฟ PDF ตารางที่ 4.11 แสดงรายละเอียดของการแจกแจง นิยามฟังก์ชันเบต้าดังนี้

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

4.2.13 การแจกแจงแบบสม่ำเสมอ

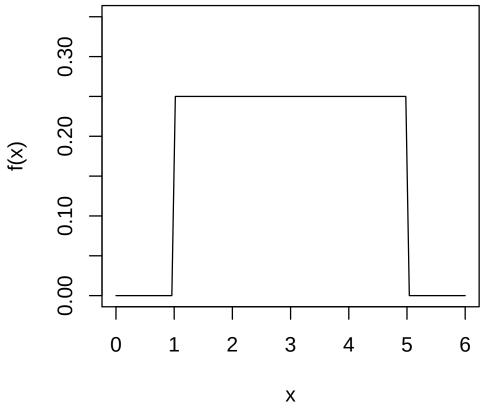
การแจกแจงแบบสม่ำเสมอมีทั้งแบบค่าต่อเนื่องและค่าไม่ต่อเนื่อง ใช้จำลองกรณีที่ผู้วิเคราะห์ไม่รู้ข้อมูลอื่น ๆ เกี่ยวกับตัวแปรสุ่มนอกจากขอบเขตบนและขอบเขตล่าง ถือได้ว่าเป็นการจำลองความไม่แน่นอนอย่างบริบูรณ์ เพราะให้ทุกค่าที่เป็นไปได้มีโอกาสเกิดเท่า ๆ กัน ตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอแบบต่อเนื่องในช่วง [0,1] สำคัญมากสำหรับการสร้างตัวเลขสุ่มที่มีการแจกแจงอื่น ๆ รูปที่ 4.10ก และ 4.10ข แสดงกราฟ PDF และ CDF สำหรับการแจกแจงแบบสม่ำเสมอแบบต่อเนื่อง และรูปที่ 4.11 สำหรับแบบไม่ต่อเนื่อง คุณสมบัติอื่น ๆ เกี่ยวกับการแจกแจงสม่ำเสมอแสดงในตารางที่ 4.12 ถึงแม้ว่าไม่มีฟังก์ชันเอกซ์เซลสำหรับคำนวณ PDF และ CDF ผู้ใช้ก็สามารถคำนวณเองได้ง่าย ๆ จากสูตรในตารางที่ 4.12



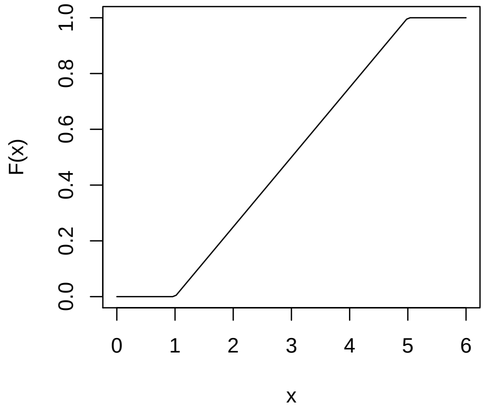
รูปที่ 4.9: กราฟ PDF ของการแจกแจงแบบเบต้ามาตรฐาน $(a = 0, b = 1)$

ตารางที่ 4.11: คุณสมบัติของการแจกแจงแบบเบต้า

ค่าสถิติ	เบต้า	ฟังก์ชันเอกซ์เซล
พารามิเตอร์	$\alpha, \beta > 0, a \leqslant b$	
ฐาน (Φ)	$[a, b]$	
PDF, $f(x)$	$\frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha,\beta)(b-a)^{\alpha+\beta-1}}$	-
CDF, $F(x)$	$\frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt}{B(\alpha,\beta)}$ เฉพาะ $a = 0, b = 1$	BETADIST(x, α, β, a, b)
ค่าเฉลี่ย (μ)	$a + (b-a)\frac{\alpha}{\alpha+\beta}$	
ความแปรปรวน	$(b-a)^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

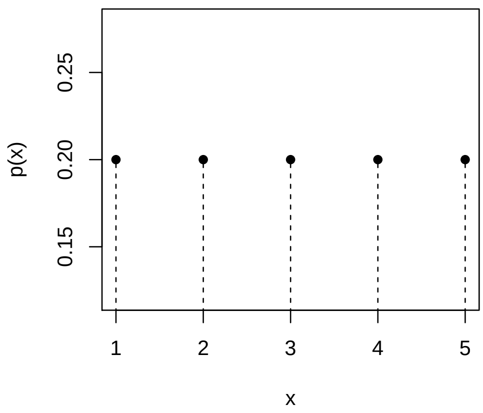


(ก) PDF

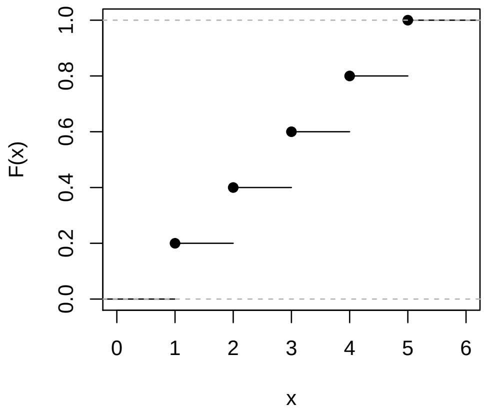


(ข) CDF

รูปที่ 4.10: การแจกแจงแบบสม่ำเสมอแบบต่อเนื่องเมื่อ $a = 1$ และ $b = 5$



(ก) PMF



(ข) CMF

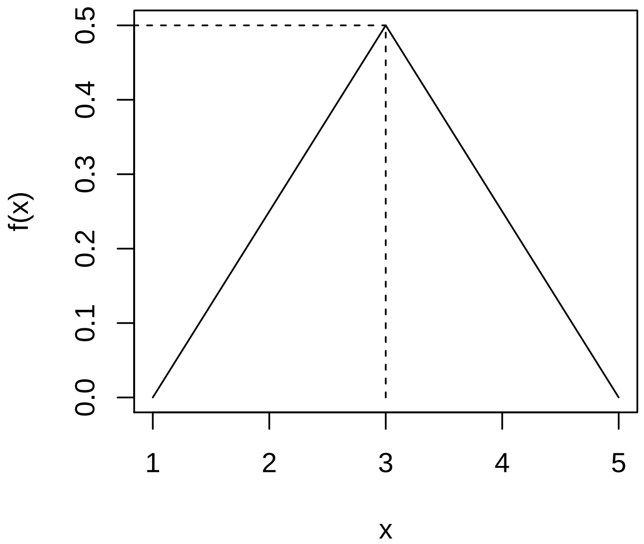
รูปที่ 4.11: การแจกแจงแบบสม่ำเสมอแบบไม่ต่อเนื่อง เมื่อ $a = 1, b = 5$ และ $n = 5$ ในตารางที่ 4.12

ตารางที่ 4.12: คุณสมบัติของการแจกแจงแบบสม่ำเสมอ

ค่าสถิติ	สม่ำเสมอแบบต่อเนื่อง	สม่ำเสมอแบบไม่ต่อเนื่อง
พารามิเตอร์	a และ b เป็นจำนวนจริง	a และ b เป็นจำนวนเต็ม
ฐาน (Φ)	$[a, b], a \leqslant b$	$\{a, a+1, \ldots, b\}$
PDF, $f(x)$	$\frac{1}{b-a}$	$\frac{1}{b-a+1}$
CDF, $F(x)$	$\frac{x-a}{b-a}$	$\frac{x-a+1}{b-a}$
ค่าเฉลี่ยและค่ามัธยฐาน	$(a+b)/2$	
ความแปรปรวน	$(b-a)^2/12$	

4.2.14 การแจกแจงแบบสามเหลี่ยม

การแจกแจงแบบสามเหลี่ยม (Triangular distribution) ใช้จำลองกระบวนการที่รู้เพียงขอบเขตล่าง (a), ขอบเขตบน (b), และค่าที่เป็นไปได้มากที่สุด (c) รูปที่ 4.12 แสดงกราฟ PDF ตารางที่ 4.13 แสดงรายละเอียดของการแจกแจง



รูปที่ 4.12: กราฟ PDF ของการแจกแจงแบบสามเหลี่ยม เมื่อ $a = 1, b = 3$ และ $c = 5$ ในตารางที่ 4.13

เมื่อพิจารณาพื้นฐานทางกายภาพของการแจกแจงมาตรฐานและเลือกการแจกแจงที่น่าจะสอดคล้องกับข้อมูลแล้ว จะต้องประมาณค่าพารามิเตอร์สำหรับการแจกแจงนั้น

ตารางที่ 4.13: คุณสมบัติของการแจกแจงแบบสามเหลี่ยม

ค่าสถิติ	สามเหลี่ยม
พารามิเตอร์	$a \in (-\infty, \infty), a \leqslant c \leqslant b$
ฐาน (Φ)	$[a, b]$
PDF, $f(x)$	$\frac{2(x-a)}{(b-a)(c-a)}, a \leqslant x \leqslant c;\ \frac{2(b-x)}{(b-a)(c-a)}, c < x \leqslant b$
CDF, $F(x)$	$\frac{(x-a)^2}{(b-a)(c-a)}, a \leqslant x \leqslant c;\ 1 - \frac{(b-x)^2}{(b-a)(c-a)}, c < x \leqslant b$
ค่าเฉลี่ย	$\frac{a+b+c}{3}$
ความแปรปรวน	$\frac{a^2+b^2+c^2-ab-ac-bc}{18}$

4.2.15 การประมาณค่าพารามิเตอร์สำหรับการแจกแจงมาตรฐาน

สมมติว่าผู้วิเคราะห์เลือกการแจกแจงแบบปกติเพื่อจำลองข้อมูล จากตารางที่ 4.6 แสดงว่าการแจกแจงนี้มีพารามิเตอร์ 2 ตัวคือ ค่าเฉลี่ยประชากร μ และส่วนเบี่ยงเบนมาตรฐาน σ ดังนั้นจึงต้องประมาณค่าพารามิเตอร์ (แทนด้วย $\hat{\mu}$ และ $\hat{\sigma}$) สองค่านี้ให้เข้ากับชุดข้อมูลที่มี โดยกำหนดให้ $\hat{\mu}$ แทนด้วยค่าเฉลี่ยจากกลุ่มตัวอย่าง (สมการที่ 2.16) และ $\hat{\sigma}$ แทนด้วยส่วนเบี่ยงเบนมาตรฐานจากกลุ่มตัวอย่าง (สมการที่ 2.17)

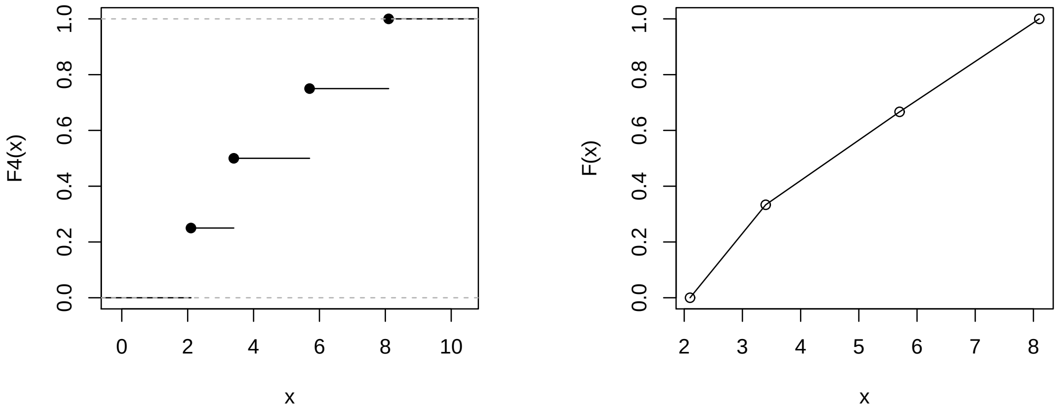
Law (2006) มีตารางแสดงสูตรคำนวณเพื่อประมาณค่าพารามิเตอร์สำหรับการแจกแจงมาตรฐานด้วยวิธีความควรจะเป็นสูงสุด (Maximum likelihood) นอกจากนี้ยังมีวิธีอื่น ๆ เช่น วิธีจับคู่โมเมนต์ (Moment matching method) และวิธีกำลังสองที่น้อยที่สุด (Method of least squares) ในเชิงทฤษฎี วิธีที่ใช้ประมาณค่ามีผลต่อค่าพารามิเตอร์ที่ได้ แต่ความแปรปรวนของข้อมูลเชิงสุ่มมักจะกลบ (Overwhelm) ความแตกต่างในวิธีที่เลือกใช้ ดังนั้นประเด็นเรื่องวิธีประมาณค่าพารามิเตอร์ที่เลือกใช้จึงไม่เป็นประเด็นสำคัญนัก และถึงอย่างไรในเชิงปฏิบัติก็ใช้ซอฟต์แวร์ประมาณค่าพารามิเตอร์อยู่ดี จึงสำคัญที่ซอฟต์แวร์ที่ใช้ประมาณค่าด้วยขั้นตอนวิธี (Algorithm) ที่เสถียรในเชิงตัวเลข ผู้อ่านควรระลึกไว้เสมอว่าไม่มีการแจกแจงที่แท้จริงที่ให้เราค้นพบ การแจกแจงมาตรฐานเป็นเพียงการจำลองข้อมูลจริงด้วยสมการคณิตศาสตร์เท่านั้น

เมื่อไม่สามารถกำหนดการแจกแจงมาตรฐาน (เช่น หัวข้อที่ 4.2.2-4.2.14) ให้สอดคล้องกับข้อมูลที่มีได้ อาจเลือกใช้การแจกแจงตามตัวอย่าง (Empirical distribution)

4.2.16 การแจกแจงเชิงประสบการณ์หรือการแจกแจงตามตัวอย่าง

การแจกแจงนี้เหมาะกับกรณีที่ลักษณะของข้อมูลไม่ตรงกับพื้นฐานทางกายภาพของการแจกแจงมาตรฐาน หรืออาจมีข้อมูลจำนวนน้อยเกินกว่าที่จะแยกแยะได้ว่าการแจกแจงที่มีพารามิเตอร์แบบใดเหมาะสมที่สุด การแจกแจงตามตัวอย่างใช้ข้อมูลที่มีเพื่อสร้างเลขสุ่มโดยให้ทุก ๆ ค่ามีความน่าจะเป็นที่จะได้รับเลือกเท่ากัน

หมายความว่าค่าที่สุ่มได้เป็นค่าที่เก็บมาได้ซึ่งมีจำนวนจำกัด หากต้องการค่าที่ต่อเนื่อง สามารถประมาณค่าในช่วง (Interpolate) ด้วยการเทียบบัญญัติไตรยางศ์ รูปที่ 4.13ก แสดง CDF ของการแจกแจงตามตัวอย่างแบบไม่ต่อเนื่องสำหรับข้อมูล {2.1, 3.4, 5.7, 8.1} มี 4 ค่า จึงให้ค่าความน่าจะเป็นสำหรับแต่ละค่าเท่ากับ 1/4 รูปที่ 4.13ข แสดง CDF ของการแจกแจงตามตัวอย่างแบบต่อเนื่องซึ่งมี 3 ช่วง จึงให้ความน่าจะเป็นของแต่ละช่วงเป็น 1/3



(ก) แบบไม่ต่อเนื่อง (ข) แบบต่อเนื่อง

รูปที่ 4.13: การแจกแจงตามตัวอย่างสำหรับข้อมูล {2.1, 3.4, 5.7, 8.1}

ข้อดีของการแจกแจงตามตัวอย่าง คือ มีคุณสมบัติทางทฤษฎีรองรับ เมื่อขนาดของข้อมูลเข้าใกล้อนันต์ (Infinity) การแจกแจงตามตัวอย่างจะลู่เข้าสู่”การแจกแจงที่แท้จริง” ที่รู้ว่ามีอยู่ แต่ไม่รู้รูปแบบ และผู้ใช้ไม่ต้องเลือกการแจกแจงมาตรฐาน แต่ข้อเสียคือค่าที่สุ่มได้จะไม่เกินค่าที่สูงที่สุดในข้อมูลที่มี และไม่ต่ำกว่าค่าที่ต่ำที่สุดในข้อมูลที่มี ทั้งนี้ยังมีวิธีแก้ไขโดยสร้างค่าที่อยู่นอกเหนือขอบเขตของข้อมูลที่มีอยู่ ด้วยการประมาณหางซ้ายและหางขวาด้วยการแจกแจงแบบเอ็กซ์โพเนนเชียล (Hill, 1975)

ถึงกระนั้นก็ดี ไม่ควรใช้การแจกแจงตามตัวอย่างพร่ำเพรื่อ เพราะการแจกแจงนี้ไม่สามารถจำลองค่าสุดโต่ง (Extreme values) ได้ดีนักโดยเฉพาะเมื่อมีข้อมูลจำนวนน้อย นอกจากนี้ หากต้องการทดสอบแบบจำลองสถานการณ์ด้วยการเปลี่ยนค่าเฉลี่ยหรือส่วนเบี่ยงเบนมาตรฐาน ก็ทำได้ยาก ประเด็นนี้เห็นได้ชัดหากเปรียบเทียบกับการแจกแจงแบบปกติ ที่มีพารามิเตอร์สำหรับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานชัดเจน การแจกแจงมาตรฐานยังมีข้อดีตรงที่สามารถปรับเรียบความขยุกขยิก (Smoothing) ที่มีอยู่ในข้อมูลจริงที่เป็นข้อมูลเชิงสุ่มได้อีกด้วย

เมื่อได้การแจกแจงมาตรฐานที่ประมาณค่าพารามิเตอร์แล้ว (เช่น การแจกแจงแบบปกติที่มี $\hat{\mu} = 20.1$ และ $\hat{\sigma} = 5.93$ ในตัวอย่างที่ 4.1) ก่อนนำไปใช้ในแบบจำลอง ผู้วิเคราะห์ควรประเมินว่าการแจกแจงนี้สอดคล้องกับ

ข้อมูลที่มีหรือไม่

4.3 การพิจารณาความสอดคล้องของการแจกแจง

การวิเคราะห์มี 2 รูปแบบ คือ **การทดสอบเชิงตัวเลขด้วยสมมติฐานทางสถิติ** เช่น การทดสอบไคสแควร์ (Chi-square test) และ**การวิเคราะห์ด้วยกราฟ** เช่น เปรียบเทียบ PDF กับฮีสโตแกรม ควรวิเคราะห์ทั้งสองด้านเพราะการทดสอบสมมติฐานสรุปความไม่พอดีของการแจกแจงความน่าจะเป็นในการจำลองข้อมูล ด้วยตัวเลขค่าสรุปทางสถิติ (Summary statistics) แต่กราฟแสดงให้เห็นว่าความไม่พอดีดังกล่าวนี้เกิดขึ้นที่บริเวณใด เป็นช่วงของค่าที่สำคัญหรือไม่

4.3.1 การทดสอบสมมติฐาน

การทดสอบความสอดคล้องของการแจกแจงมีสมมติฐานหลัก (H_0) และสมมติฐานรอง (H_1) ดังนี้

H_0: ข้อมูลที่มีมาจากการแจกแจงที่เลือก
H_1: ข้อมูลที่มีไม่ใช่มาจากการแจกแจงที่เลือก

เช่น การแจกแจงที่เลือกสำหรับข้อมูลในตัวอย่างที่ 4.1 ซึ่งจะตามมาภายหลัง เป็นการแจกแจงแบบปกติที่มี $\hat{\mu}$ = 20.1 และ $\hat{\sigma}$ = 5.93 (พารามิเตอร์ประมาณจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง) สมมติฐานหลักคือข้อมูลชุดนี้มีการแจกแจงแบบปกติที่มี $\hat{\mu}$ = 20.1 และ $\hat{\sigma}$ = 5.93

การทดสอบสมมติฐานทั่วไปมีขั้นตอนคล้ายกัน กล่าวคือ หลังจากกำหนด H_0 และ H_1 แล้ว จะคำนวณค่าสถิติจากข้อมูลที่มี ค่าสถิตินี้บ่งชี้ว่าจะยอมรับ (ไม่สามารถปฏิเสธ) หรือปฏิเสธ H_0 อีกพารามิเตอร์หนึ่งที่ผู้วิจัยกำหนดคือระดับนัยสำคัญ (Significance level, α) ซึ่งคือความน่าจะเป็นของความผิดพลาดที่ปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นจริง มักใช้ $\alpha = 0.05$ หรือ 5% พิจารณายอมรับหรือปฏิเสธ H_0 ได้ 2 วิธี ดังนี้

1. เปรียบเทียบค่าสถิติกับค่าวิกฤต (Critical value) ซึ่งขึ้นกับขนาดข้อมูลและระดับนัยสำคัญ ถ้าค่าสถิติ (หรือในบางกรณี ใช้ค่าสัมบูรณ์ของค่าสถิติ) น้อยกว่าค่าวิกฤตจะยอมรับ H_0 ถ้าไม่เช่นนั้นจะปฏิเสธ H_0

2. นำค่าสถิติไปคำนวณค่า p-value ซึ่งคือความน่าจะเป็นแบบมีเงื่อนไขที่ถ้า H_0 เป็นจริง ระดับนัยสำคัญที่แทบจะปฏิเสธ H_0 ภายใต้ค่าสถิติที่มีคือ p-value หาก p-value มากกว่าระดับนัยสำคัญ (α) จะยอมรับ H_0 ถ้าไม่เช่นนั้นจะปฏิเสธ H_0

การทดสอบสมมติฐานมีข้อจำกัดในด้านสุดโต่ง 2 ทาง กล่าวคือ หากมีจำนวนข้อมูลไม่มาก (เช่น น้อยกว่า 10) การทดสอบมักจะ**ไม่ปฏิเสธ**การแจกแจงความน่าจะเป็นมาตรฐาน ไม่ว่าจะเลือกอันใด เพราะการทดสอบ

ไม่สามารถแยกแยะความแตกต่างระหว่างการแจกแจงได้ ในทางตรงข้ามหากมีข้อมูลจำนวนมาก (เช่น หลักหลายร้อย) การทดสอบมักจะปฏิเสธแทบทุกการแจกแจงมาตรฐาน เพราะเห็นความแตกต่างระหว่าง PDF (หรือ PMF) ของการแจกแจงที่เลือกกับข้อมูลจริงอย่างชัดเจน ดังนั้นจึงไม่ควรถือเอาผลการทดสอบสมมติฐานเป็นคำสิ้นสุด แต่ควรมองว่าเป็นหลักฐานอีกชิ้นหนึ่งเพื่อประกอบการตัดสินใจ ด้วยข้อจำกัดสองด้านนี้ ในเชิงปฏิบัติ ผู้วิเคราะห์ควรพิจารณาเปรียบเทียบ PDF (หรือ PMF) และฮีสโตแกรมเพิ่มเติมด้วย

การทดสอบสมมติฐานที่จะกล่าวถึง 2 ชนิดนี้ เป็นการทดสอบที่มีใช้แพร่หลายในซอฟต์แวร์จำลองสถานการณ์สำเร็จรูป เช่น อารีน่า มี Input Analyzer และ @Risk มี BestFit นอกจากนี้แล้วสามารถใช้ซอฟต์แวร์ทางสถิติทำการทดสอบเหล่านี้ เช่น Minitab หรือ R ซึ่งเป็นฟรีแวร์ (อ่านเพิ่มเติมได้จาก Ricci 2005) ผู้วิเคราะห์ควรรู้หลักการของการทดสอบ, ข้อจำกัด, และการตีความผลที่ได้

การทดสอบแบบไคสแควร์

การทดสอบนี้มาจากพื้นฐานความคิดที่ให้ค่าตัวเลขสำหรับความสอดคล้องของการแจกแจง ด้วยการเปรียบเทียบกราฟ PDF ของการแจกแจงที่เลือกกับฮีสโตแกรมซึ่งเป็นตัวแทนของข้อมูลจริง กำหนดให้ O_i เป็นจำนวนข้อมูลจริงที่อยู่ในช่วง i ของฮีสโตแกรม เช่น ในรูปที่ 4.14 ช่วงที่ 1 ระหว่างค่า 10-12 มีข้อมูล 2 ค่าที่อยู่ในช่วงนี้ และ E_i เป็นค่าคาดหมายของจำนวนข้อมูลที่น่าจะอยู่ในช่วงนี้ ที่คำนวณได้จาก PDF ที่เลือก เช่น ในรูปที่ 4.14 เมื่อใช้การแจกแจงแบบปกติที่ $\hat{\mu} = 20.1$ และ $\hat{\sigma} = 5.93$ ช่วงที่ 1 มีจุดกึ่งกลางที่ 11 ซึ่งมีค่าคาดคะเนของจำนวนข้อมูลที่อยู่ในช่วงนี้ 1.359 หาก PDF ใกล้เคียงกับฮีสโตแกรม ผลต่างระหว่าง $(O_i - E_i)^2$ ควรมีค่าน้อย หากไม่ใกล้เคียงกัน ผลต่างจะมาก ยกกำลังสองเพื่อให้ผลต่างทุกค่ามีเครื่องหมายบวก สมการที่ (4.3) แสดงสูตรที่ใช้คำนวณค่าสถิติไคสแควร์ เมื่อ k เป็นจำนวนช่วงในฮีสโตแกรม

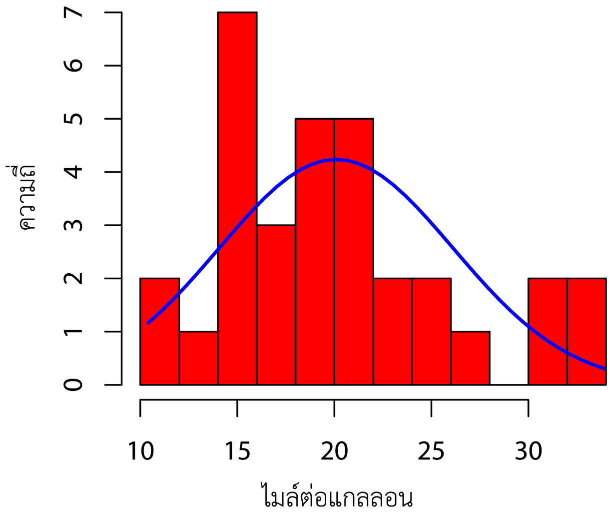
$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{4.3}$$

ตัวอย่างที่ 4.1. ข้อมูลจำนวนไมล์ที่รถวิ่งได้จากน้ำมัน 1 แกลลอน `mpg` ในชุดข้อมูล `mtcars` ของซอฟต์แวร์ `R` มี 32 ค่า ดังนี้

21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4 10.4 14.7
32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7 15.0 21.4

เมื่อจำลองด้วยการแจกแจงแบบปกติที่มี $\hat{\mu}$ = 20.1 และ $\hat{\sigma}$ = 5.93 การทดสอบแบบไคสแควร์ที่ใช้ 4 ช่วงให้ χ^2 = 3.17 และค่า p-value = 0.0794 เนื่องจาก p-value $> \alpha$ = 0.05 จึงไม่สามารถปฏิเสธการแจกแจงแบบปกติที่ $\hat{\mu}$ = 20.1 และ $\hat{\sigma}$ = 5.93 ได้

□



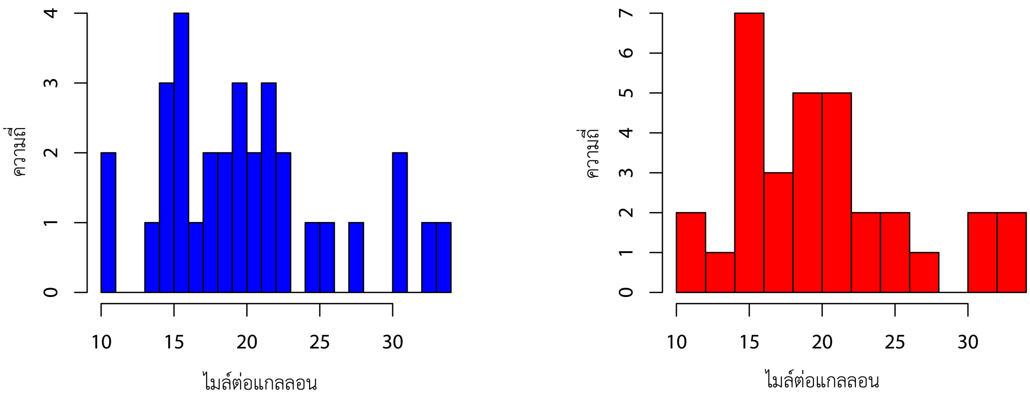
รูปที่ 4.14: การเปรียบเทียบฮีสโตแกรมกับ PDF ของการแจกแจงแบบปกติสำหรับข้อมูล `mpg` ในชุด `mtcars` ของซอฟต์แวร์ `R`

เนื่องจากการทดสอบไคสแควร์มีพื้นฐานมาจากฮีสโตแกรม ค่าสถิติที่คำนวณได้จึงไว (Sensitive) ต่อจำนวนช่วง (Bin หรือ Interval) ของฮีสโตแกรม เช่น ในตัวอย่างที่ 4.1 เมื่อใช้ 4 ช่วง ค่า χ^2 = 3.17 และค่า p-value = 0.079 $>$ α จึงไม่สามารถปฏิเสธ H_0 ที่นัยสำคัญ α=5% แต่เมื่อใช้ 5 ช่วง ค่า χ^2 = 6.37 และค่า p-value = 0.043 $<$ α จึงปฏิเสธ H_0 นอกจากนี้แล้วจำนวนช่วงหรือจำนวนแท่งนี้ยังส่งผลต่อรูปลักษณ์ของฮีสโตแกรมด้วย จำนวนช่วงมาก (รูปที่ 4.15ก) อาจทำให้ฮีสโตแกรมดูมีรายละเอียดมากแต่ขรุขระเกินไป แต่หากจำนวนช่วงน้อย (รูปที่ 4.15ข) ก็อาจทำให้ฮีสโตแกรมดูไม่มีรายละเอียด หยาบเกินไป จึงแนะนำให้ลองเปลี่ยนค่าจำนวนช่วงเพื่อดูว่าผลการทดสอบไคสแควร์ไวต่อค่าจำนวนช่วงเพียงใด Banks et al. (2005) แนะนำว่าเมื่อจำนวนข้อมูล (n) มากกว่า 100 ค่าขึ้นไป ควรใช้จำนวนช่วงระหว่าง $\sqrt{n}$ ถึง $n/5$ กล่อง

เนื่องจากการทดสอบไคสแควร์มีข้อจำกัดเรื่องความไวต่อจำนวนช่วงของฮีสโตแกรม จึงมีการคิดค้นการทดสอบอื่น ๆ เพื่อแก้ไขประเด็นนี้ เช่น การทดสอบแบบโคโมโกรอฟ-สเมอร์นอฟ (Kolmogorov-Smirnov, KS)

การทดสอบแบบโคโมโกรอฟ-สเมอร์นอฟ

การทดสอบแบบโคโมโกรอฟ-สเมอร์นอฟมีประโยชน์มากสำหรับกรณีที่มีจำนวนข้อมูลน้อย การทดสอบ KS ประมาณค่าความไม่พอดีของการแจกแจงในการจำลองข้อมูลด้วยค่าที่มากที่สุดของค่าสัมบูรณ์ของความแตก



(ก) จำนวนช่วง=20 (ข) จำนวนช่วง=12

รูปที่ 4.15: การเปรียบเทียบฮีสโตแกรมของข้อมูล mpg เมื่อจำนวนช่วงต่างกัน

ต่างระหว่าง CDF ของการแจกแจงตามตัวอย่าง (Empirical CDF, $\widehat{F}(x)$) และ CDF ของการแจกแจงมาตรฐาน $F(x)$ เมื่อ n คือขนาดข้อมูล ดังแสดงในสมการที่ (4.4)

$$D = \max_{1 \leqslant i \leqslant n} |F(x_i) - \widehat{F}(x_i)| \tag{4.4}$$

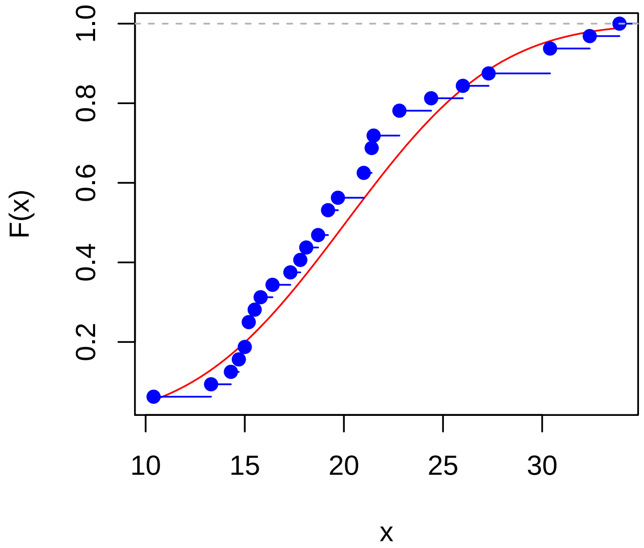
รูปที่ 4.16 เปรียบเทียบ CDF เชิงประสบการณ์จากข้อมูลตัวอย่างและ CDF ของการแจกแจงปกติ ค่า D ในสมการที่ (4.4) คือ ระยะห่างตามแนวดิ่งที่กว้างที่สุดระหว่างจุดและเส้นกราฟ

ตัวอย่างที่ 4.2. ใช้ข้อมูลจากตัวอย่างที่ 4.1 และจำลองด้วยการแจกแจงแบบปกติเช่นเดิม จะได้ $D = 0.126$, p-value = 0.687 เนื่องจาก p-value $> \alpha$ = 0.05 จึงไม่สามารถปฏิเสธการแจกแจงแบบปกติที่ $\hat{\mu}$ = 20.1 และ $\hat{\sigma}$ = 5.93 ได้ ซอฟต์แวร์ เช่น Input Analyzer ในอารีน่า ไม่รายงานค่า p-value สำหรับการทดสอบ KS หากค่านั้นมากกว่า 0.15 หากผู้วิเคราะห์ต้องการค่า p-value ก็สามารถใช้ซอฟต์แวร์ทางสถิติ เช่น R ได้ (อ่านวิธีใช้ได้จาก Ricci 2005)

□

เนื่องจากการทดสอบ KS พิจารณาความแตกต่างที่มากที่สุด (สมการที่ 4.4) จึงอาจไวต่อค่าสุดโต่ง มีการทดสอบ Anderson-Darling ที่คำนวณค่าสถิติด้วยการถัวเฉลี่ยผลต่าง $|F(x_i) - \widehat{F}(x_i)|$ ไม่พิจารณาแค่ค่าที่มากที่สุดดังเช่นการทดสอบ KS (อ่านรายละเอียดเพิ่มเติมใน Law 2006)

การทดสอบสมมติฐานในหัวข้อ 4.3.1 สรุปความไม่พอดีของการแจกแจงความน่าจะเป็นสำหรับจำลองข้อมูลจริง ด้วยตัวเลขค่าสรุปทางสถิติ (Summary statistics) แต่กราฟแสดงให้เห็นว่าความไม่พอดีดังกล่าวนี้เกิดขึ้นที่จุดใด เป็นช่วงของค่าที่สำคัญหรือไม่



รูปที่ 4.16: การเปรียบเทียบ CDF ของการแจกแจงตามตัวอย่าง (จุด) กับ CDF ของการแจกแจงแบบปกติ (เส้น) สำหรับข้อมูล `mpg` ในชุด `mtcars` ของซอฟต์แวร์ R

4.3.2 การพิจารณาด้วยกราฟ

กล่าวถึงกราฟ 2 ชนิด ดังนี้

กราฟความหนาแน่น-ฮีสโตแกรม

รูปที่ 4.14 แสดงตัวอย่างของกราฟความหนาแน่น-ฮีสโตแกรมนี้ กราฟความหนาแน่น (PDF) แสดงการแจกแจงมาตรฐานที่เลือก ฮีสโตแกรมก็เป็นฟังก์ชันความหนาแน่นเช่นกัน แต่เป็นความหนาแน่นจากกลุ่มตัวอย่างที่เป็นข้อมูลจริง สิ่งที่พิจารณาในกราฟความหนาแน่น-ฮีสโตแกรมคือดูว่ากราฟ PDF ใกล้เคียงกับฮีสโตแกรมหรือไม่ เนื่องจากจำนวนช่วงหรือจำนวนแท่งนี้ยังส่งผลต่อรูปลักษณ์ของฮีสโตแกรม จึงแนะนำให้ทดลองเปลี่ยนค่าจำนวนช่วง นอกจากเปรียบเทียบกราฟ PDF กับฮีสโตแกรมในภาพรวมแล้ว หากมีช่วงของข้อมูลที่สนใจ เช่น สำหรับข้อมูลจำนวนไมล์ที่วิ่งได้ต่อแกลลอน (mpg) ในตัวอย่างที่ 4.1 สมมติว่าช่วงของค่าที่ผู้วิเคราะห์สนใจ คือ 30-35 หรือหางด้านขวา ในรูปที่ 4.14 จะเห็นว่าการแจกแจงแบบปกติยังจำลองได้ไม่ดีนัก ผู้วิเคราะห์อาจจะเลือกการแจกแจงอื่นที่จำลองพฤติกรรมสุดโต่งได้ดีกว่า เช่น การแจกแจงแบบล็อกนอร์มอล

ควอนไทล์พล็อต (Q-Q)

พื้นฐานความคิดของควอนไทล์พล็อต คือ ถ้าการแจกแจงมาตรฐานที่เลือกสอดคล้องกับข้อมูลจริง ข้อมูลเทียมที่สร้างมาจากการแจกแจงมาตรฐานไม่ควรดูแตกต่างจากข้อมูลจริงมากนัก เช่น ในตัวอย่างที่ 4.1 ใช้การแจกแจงปกติด้วย $\widehat{\mu}$ = 20.1 และ $\widehat{\sigma}$ = 5.93 หรือ N(20.1, 5.93) ข้อมูลเทียมที่สร้างมาจาก N(20.1, 5.93) ควรใกล้เคียงกับข้อมูล mpg ในตัวอย่าง สมมติให้ n เป็นจำนวนข้อมูล การเปรียบเทียบนี้ใช้แผนภาพกระจายของคู่ลำดับที่มาจาก

- **ข้อมูลจริง**ที่เรียงจากน้อยไปมาก

$$X_{[1]} \leqslant X_{[2]} \ldots \leqslant X_{[n]}$$

- **ข้อมูลเทียม**ที่สร้างมาจาก N(20.1, 5.93) โดยใช้ตัวผกผันของ CDF (Inverse CDF), $\widehat{F}^{-1}(\cdot)$

$$\widehat{F}^{-1}\left(\frac{0.5}{n}\right), \widehat{F}^{-1}\left(\frac{1.5}{n}\right), \ldots, \widehat{F}^{-1}\left(\frac{n-0.5}{n}\right)$$

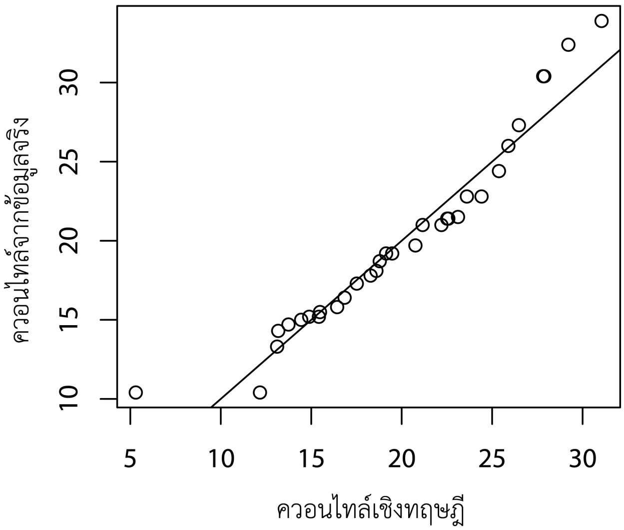
รูปที่ 4.17 แสดงควอนไทล์พล็อตสำหรับตัวอย่างที่ 4.1 เปรียบเทียบกับเส้น 45° (เป็นกราฟของฟังก์ชัน $x = y$) เพื่อช่วยเทียบเคียงว่าข้อมูลจริงกับข้อมูลเทียมใกล้เคียงกันหรือไม่ กราฟของควอนไทล์ที่ค่อนข้างเป็นเส้นตรง (อาจไม่ใช่ 45°) แสดงว่า*ชนิด*ของการแจกแจงที่เลือกนั้นสอดคล้องกับข้อมูล หากกราฟของควอนไทล์นี้ใกล้เคียงกับเส้น 45° บ่งบอกว่าใช้ค่าพารามิเตอร์ถูกต้อง รูปที่ 4.17 ยืนยันผลของการทดสอบสมมติฐานว่า N(20.1, 5.93) นั้นสอดคล้องกับข้อมูล เพราะควอนไทล์พล็อตใกล้เคียงกับเส้น 45° แต่ N(20.1, 5.93) ไม่เหมาะกับข้อมูลสำหรับค่าสุดโต่งทั้งสองด้าน คือ หางซ้ายและหางขวาเพราะค่าควอนไทล์เทียมที่สร้างมาจาก N(20.1, 5.93) และของจริงต่างกันมาก (ห่างจากเส้น 45°) ความแตกต่างที่ตำแหน่งเหล่านี้ไม่สามารถเห็นได้จากการทดสอบสมมติฐาน

ข้อดีของควอนไทล์พล็อตคือไม่ขึ้นอยู่กับการแบ่งกลุ่มของข้อมูลดังที่เป็นข้อจำกัดของฮีสโตแกรม ในกรณีที่จำนวนข้อมูลไม่มากควอนไทล์พล็อตเหมาะสมกว่ากราฟความหนาแน่นและฮีสโตแกรม นอกจากนี้แล้วควอนไทล์พล็อตยังสามารถแสดงได้ว่าการแจกแจงมาตรฐานที่เลือกนั้นไม่พอดีกับข้อมูลจริงที่ช่วงค่าใด ซอฟต์แวร์ทางสถิติ เช่น R มีคำสั่งสำหรับทำกราฟนี้โดยเฉพาะ (อ่านวิธีทำจาก Ricci 2005)

เนื้อหาทั้งหมดที่นำเสนอในหัวข้อ 4.3 ตั้งอยู่บนสมมติฐานที่ว่ามีข้อมูลจริง แต่ในบางกรณี ผู้วิเคราะห์อาจไม่มีข้อมูล เช่น ในกรณีที่ระบบที่สนใจยังไม่มีอยู่จริง หัวข้อถัดไปจะกล่าวถึงการจำลองตัวแปรนำเข้าในกรณีนี้

4.4 การจำลองตัวแปรนำเข้าเมื่อไม่มีข้อมูล

เมื่อไม่มีข้อมูลจริงสำหรับจำลองตัวแปรนำเข้า ผู้วิเคราะห์จำเป็นต้องหาแหล่งข้อมูลอื่น: ข้อมูลหรือมาตรฐานเชิงวิศวกรรม, ข้อบังคับทางกฎหมาย, หรือการจัดลำดับโดยองค์กร อาจสามารถให้ค่ากลาง ๆ ได้ เช่น ความเร็ว



รูปที่ 4.17: ควอนไทล์พล็อตสำหรับข้อมูล `mpg` ในชุด `mtcars` ของซอฟต์แวร์ R

ของรถบรรทุกบนทางหลวง, ความเห็นของผู้เชี่ยวชาญ เช่น เจ้าหน้าที่ที่หน้างานหรือซัพพลายเออร์ที่ขายเครื่องจักรหรืออุปกรณ์อาจให้ค่าประมาณของปริมาณงาน (Throughput) ที่ไหลออกจากเครื่อง, ข้อจำกัดทางกายภาพหรือทางสามัญสามารถกำหนดขอบเขตได้ เช่น ความเร็วในการเดินของมนุษย์ นอกจากนี้แล้วพื้นฐานทางกายภาพของกระบวนการสุ่มยังช่วยเสนอแนะการแจกแจงความน่าจะเป็นที่สอดคล้องได้ เช่น อาจใช้ตัวแปรสุ่มแบบไวบูลเพื่อจำลองอายุขัยของอุปกรณ์ดังที่ได้กล่าวไปแล้วในหัวข้อที่ 4.2.1 วิธีต่อไปนี้สามารถนำมาใช้จำลองข้อมูลนำเข้าในกรณีที่ไม่มีข้อมูลจริงได้

4.4.1 วิธีจุดพัก

วิธีจุดพัก (Breakpoint method) มีประโยชน์สำหรับจำลองตัวแปรนำเข้าที่มีค่าที่เป็นไปได้หลากหลาย เช่น ปริมาณขายใน 1 ปีหรือจำนวนชั่วโมงทำงานล่วงเวลา (Overtime, OT) ทั้งหมด แบบจำลองความน่าจะเป็นที่ใช้ขึ้นอยู่กับข้อมูลที่มี (ความละเอียดของแบบจำลองเรียงจากน้อยไปมาก) ดังนี้

- **ขอบเขตบนและขอบเขตล่าง** เช่น ยอดขายสินค้า Om ไม่ต่ำกว่า 1000 (ขอบเขตล่าง) แต่ไม่เกิน 5000 หน่วย (ขอบเขตบน)
  การแจกแจงที่ใช้ คือ สม่ำเสมอแบบต่อเนื่อง (สำหรับค่าสุ่มที่ต่อเนื่อง) หรือ สม่ำเสมอแบบไม่ต่อเนื่อง (สำหรับค่าสุ่มที่ไม่ต่อเนื่อง) โดยกำหนดให้ขอบเขตล่าง $a = 1000$ และขอบเขตบน $b = 5000$ สำหรับ PDF หรือ PMF ในตารางที่ 4.12

ข้อสังเกต แบบจำลองนี้ไม่ค่อยเหมาะสมในเชิงปฏิบัตินัก เพราะให้ค่าทุกค่าในช่วง $[a, b]$ มีความน่าจะเป็นที่จะเกิดขึ้นเท่า ๆ กัน หมายความว่าค่าสุดโต่งมีโอกาสเกิดเท่า ๆ กับค่าตรงกลาง ซึ่งอาจไม่สมจริง ทั้งนี้หากต้องการให้แบบจำลองสามารถรองรับความไม่แน่นอนสูง (Conservative) หรือไม่สามารถหาข้อมูลเพิ่มเติมได้ ก็อาจใช้การแจกแจงแบบสม่ำเสมอได้

- **ขอบเขตบน, ขอบเขตล่าง, และค่าที่เป็นไปได้มากที่สุด** เช่น ยอดขายสินค้า Om ไม่ต่ำกว่า 1000 แต่ไม่เกิน 5000 หน่วย และค่าที่เป็นไปได้มากที่สุด คือ 3500

  การแจกแจงที่ใช้ คือ สามเหลี่ยม โดยกำหนดให้ a = 1000, b = 5000 และ c = 3500 สำหรับ PDF ในตารางที่ 4.13

  ข้อสังเกต สำหรับกรณีที่ไม่มีข้อมูลจริง การแจกแจงสามเหลี่ยมดีกว่าการแจกแจงแบบสม่ำเสมอ เพราะให้สารสนเทศที่มากกว่า

- **ขอบเขตบน, ขอบเขตล่าง, และจุดพัก 2-3 จุดระหว่างค่าขอบเขต** (ค่าของตัวแปรสุ่มและความน่าจะเป็นที่น้อยกว่าค่านั้น) เช่น ยอดขายสินค้า Om ไม่ต่ำกว่า 1000 แต่ไม่เกิน 5000 หน่วย และ

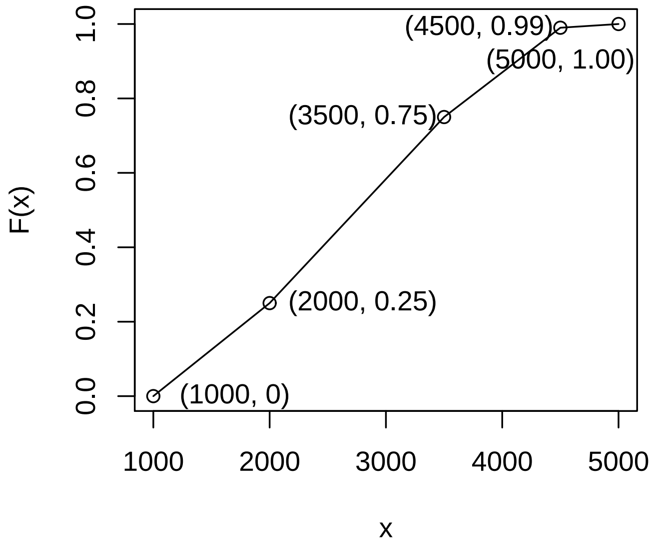
ค่ายอดขาย (x)	**โอกาสที่ยอดขายจริง** $\leqslant x$
2000	25%
3500	75%
4500	99%

  การแจกแจงที่ใช้ คล้ายการแจกแจงตามตัวอย่าง แต่ความน่าจะเป็นของแต่ละช่วงไม่เท่ากัน ดังแสดงด้วย CDF ในรูปที่ 4.18

  ข้อสังเกต ให้ระบุจุดพักเท่าที่สามารถมั่นใจว่าเชื่อถือได้ ปกติมีจุดพักอย่างมาก 3 ค่าในกรณีที่ไม่มีข้อมูลจริง; พยายามกำหนดค่าจุดพักใกล้ค่าสุดโต่ง (ค่าต่ำมาก ๆ หรือค่าสูงมาก ๆ) หากเป็นไปได้ เพราะค่าสุดโต่งมักเป็นค่าที่แปลกกว่าค่าทั่วไป; หากไม่ได้คำตอบเวลาสัมภาษณ์เก็บข้อมูล อาจทดลองเปลี่ยนคำถาม โดยมากแล้วคนทั่วไปจะสามารถประมาณความน่าจะเป็นที่จะเกินค่าใดค่าหนึ่ง (สมมติ x) ได้ง่ายกว่าให้บอกความน่าจะเป็นที่จะไม่เกินค่า x

4.4.2 วิธีค่าเฉลี่ยและความแปรปรวน

วิธีนี้เหมาะกับตัวแปรนำเข้าที่มีค่าที่เป็นไปได้หลากหลาย เช่น ยอดขายรวมของชาเขียวที่ร้านสะดวกซื้อในรอบไตรมาส จำนวนชั่วโมงรวมที่เครื่องจักรหยุดพักเพื่อซ่อมบำรุง นอกจากนี้แล้วยังเป็นประโยชน์ในกรณีที่ความแปรปรวนมีค่าเป็นสัดส่วนของค่าเฉลี่ย แบบจำลองต่อไปนี้เรียงลำดับจาก Information น้อยจนถึงมาก ดังนี้



รูปที่ 4.18: CDF ของแบบจำลองความน่าจะเป็นสำหรับวิธีจุดพัก

- **ค่าเฉลี่ยและความแปรปรวนรอบค่าเฉลี่ย** เช่น ปีที่แล้วยอดขายสินค้าเสื้อยืดโลโก้มหาวิทยาลัยคือ 1,000 ตัว ปีนี้คาดว่ายอดขายจะเพิ่มขึ้น 15% โดยมีความแกว่งประมาณ 5% มากหรือน้อยกว่าค่าเฉลี่ย ตัวแปรสุ่มที่จำลองคือยอดขายปีนี้
  การแจกแจงที่ใช้ คือ ปกติที่ $\mu = 1000 + (0.15 \times 1000) = 1150$ และ $\sigma = 0.05 \times 1150 = 57.5$
  ข้อสังเกต อาจต้องปัดค่าที่สุ่มได้จากการแจกแจงแบบปกติให้เป็นจำนวนเต็ม นอกจากนี้หากความแกว่งมากกว่า 33% ของค่าเฉลี่ย ค่าที่สุ่มได้อาจติดลบ หากตัวแปรนำเข้าเป็นค่าที่ไม่ควรติดลบ อาจต้องตรวจสอบค่าที่สุ่มได้และเปลี่ยนค่าให้เป็นศูนย์หรือค่าบวกอื่น เช่น ในไมโครซอฟท์เอกซ์เซลหากต้องการปัดค่าลบให้เป็นศูนย์ สมมติให้ `B1` เก็บค่าที่สุ่มมาจากการแจกแจงแบบปกติ ใช้ `if(B1< 0, 0, B1)`

- **ค่าเฉลี่ย, ความแปรปรวนรอบค่าเฉลี่ย, ขอบเขตบน, และขอบเขตล่าง** เช่น ปีที่แล้วยอดขายสินค้าเสื้อยืดโลโก้มหาวิทยาลัยที่สหกรณ์เป็น 1,000 ตัว ปีนี้คาดว่ายอดขายจะเพิ่มขึ้น 15% โดยมีความแกว่งประมาณ 5% มากหรือน้อยกว่าค่าเฉลี่ย แต่ยอดขายจะไม่ต่ำกว่า 800 ตัวและไม่เกิน 1,500 ตัว
  การแจกแจงที่ใช้ คือ ปกติที่ μ = 1150, σ = 57.5 และค่าที่น้อยที่สุด 800 ส่วนค่าที่มากที่สุด 1500 ตัว
  ข้อสังเกต สามารถใช้ฟังก์ชัน `min()` และ `max()` เพื่อตรวจสอบค่าที่สุ่มได้ว่าอยู่ในช่วงที่ต้องการหรือไม่ เช่นในไมโครซอฟท์เอกซ์เซลและพ็อพทูล ใช้ `min(1500, max(dNormalDev(1150, 57.5), 800))`

4.4.3 วิธีจำลองผลลัพธ์แบบไม่ต่อเนื่อง

วิธีนี้ใช้สำหรับจำลองเหตุการณ์ไม่ต่อเนื่อง เช่น ได้ขายหรือไม่ได้ขาย สินค้าพร้อมที่จะส่งในไตรมาสที่ 1, 2 หรือ 3 ข้อมูลที่ต้องการคือความน่าจะเป็นที่จะเกิดเหตุการณ์แต่ละอย่างโดยที่ความน่าจะเป็นรวมเป็น 100%

ตัวอย่างที่ 4.3. ผู้ประกอบการรายหนึ่งมีโอกาส 80% ที่จะได้วางขายสินค้าที่ห้างค้าปลีกขนาดใหญ่: ถ้าได้วางขาย มีโอกาส 10% ที่จะเริ่มขายในไตรมาสแรก, 50% ที่จะเริ่มขายในไตรมาสที่สอง, และ 40% ที่จะเริ่มขายในไตรมาส 3 ยอดขายมีมูลค่า 250,000 บาทต่อไตรมาสที่ได้ขาย ตารางที่ 4.14 แสดงแบบจำลองบนเอกซ์เซลที่ใช้พ็อพทูลเพื่อสุ่มค่า เซลล์ `B6` เก็บค่าผลลัพธ์ว่าได้เซ็นสัญญาขายหรือไม่ ซึ่งจำลองด้วยตัวแปรสุ่มแบบเบอร์นูลลี แต่พ็อพทูลมีแต่การแจกแจงแบบทวินาม จึงใช้การแจกแจงแบบทวินามเมื่อสุ่ม 1 ครั้งแทนเบอร์นูลลี หรือจะใช้ `dDiscreteDev` ที่มีผลที่เป็นไปได้ 2 ค่าคือ 0 และ 1 ด้วยความน่าจะเป็น 0.2 และ 0.8 ตามลำดับก็ได้; เซลล์ `B7` แสดงลำดับไตรมาสที่เริ่มขาย ซึ่งจำลองด้วยแบบจำลองพหุนาม; ยอดขายรวมทั้งปีอยู่ในเซลล์ `B8` โดยต้องตรวจสอบเงื่อนไขว่าได้เซ็นสัญญาหรือไม่ ในทางตรรกศาสตร์ ค่าบวกคือ”จริง” และค่าศูนย์คือ”เท็จ” ถ้าได้ขาย ยอดขายขึ้นอยู่กับไตรมาสที่ได้เริ่มขาย เช่น ถ้าได้เริ่มขายตั้งแต่ไตรมาสแรก จำนวนไตรมาสที่ได้ขายคือ 5-1 = 4 เป็นต้น

□

ตารางที่ 4.14: แบบจำลองพ็อพทูลสำหรับตัวอย่างที่ 4.3

	A	B	C
1	ไตรมาส	ความน่าจะเป็น	
2	1	0.1	
3	2	0.5	
4	3	0.4	
5			
6	ได้เซ็นสัญญาขาย?	1	`B6=dBinomialDev(1, 0.8)`
7	ไตรมาสที่เริ่มขาย	2	`B7=dDiscreteDev(A2:A4, B2:B4)`
8	ยอดขายรวมต่อปี	750,000	`B8=IF(B6, (5-B7)*250000,0)`

เมื่อไม่มีข้อมูลจริงสำหรับกำหนดตัวแปรนำเข้าในแบบจำลอง จำเป็นที่จะต้องทำการวิเคราะห์ความไว

4.4.4 การวิเคราะห์ความไว

คำแนะนำคือให้ใส่ใจพิจารณาส่วนเบี่ยงเบนมาตรฐาน และขอบเขตต่าง ๆ ให้ทดลองเปลี่ยนค่าเหล่านี้ว่าจะส่งผลกระทบต่อเอาท์พุททที่เป็นดัชนีชี้วัดมากหรือไม่ เช่น วิธีจุดพักในหัวข้อ 4.4.1 การแจกแจงสามเหลี่ยมที่มีค่าพารามิเตอร์สำหรับขอบเขตล่างคือ a และขอบเขตบนคือ b ผู้วิเคราะห์อาจเลื่อนค่า a และ b แล้วสังเกตดูผลกระทบที่มีต่อเอาท์พุท สำหรับการแจกแจงแบบปกติในหัวข้อ 4.4.2 อาจเพิ่มหรือลดส่วนเบี่ยงเบนมาตรฐานเพื่อดูว่ามีความเปลี่ยนแปลงในเอาท์พุทอย่างเด่นชัดหรือไม่ ในกรณีที่มีตัวแปรนำเข้าหลายตัว อาจเลือกวิเคราะห์ความไวเฉพาะกับตัวแปรนำเข้าที่สำคัญที่ส่งผลต่อเอาท์พุทมาก

4.5 บทส่งท้าย

บทนี้กล่าวถึงการเลือกแบบจำลองความน่าจะเป็นในกรณีที่มีข้อมูลจริงและไม่มีข้อมูล ใช้ได้สำหรับชุดข้อมูลที่อนุมานได้ว่าเป็นอิสระต่อกันและไม่มีแนวโน้ม หากข้อมูลมีค่าสหสัมพันธ์ในตัวเองหรือขึ้นอยู่กับเวลา ผู้วิเคราะห์ควรใช้แบบจำลองอื่น สามารถอ่านเพิ่มเติมเบื้องต้นได้จาก Banks et al. (2005) และ Law (2006) ผู้เขียนจบบทนี้ด้วยข้อผิดพลาดที่พบบ่อย เพื่อให้ผู้อ่านได้พิจารณาว่าเข้าข่ายเหล่านี้บ้างหรือไม่

ความผิดพลาดที่พบบ่อย

- กรณีที่ซอฟต์แวร์สำเร็จรูปสามารถเสนอรูปแบบการแจกแจงได้ ผู้ใช้เชื่อซอฟต์แวร์โดยไม่ทราบว่าการแจกแจงได้รับเลือกด้วยเกณฑ์อะไร และไม่ได้พิจารณาว่าเกณฑ์นั้น ๆ เหมาะสมหรือไม่กับข้อมูลที่มี นอกจากนี้ ผู้ใช้มักเลือกการแจกแจงนั้นโดยไม่พิจารณาปัจจัยอื่น ๆ ประกอบ เช่น ลักษณะทางกายภาพ, กราฟความหนาแน่น-ฮีสโตแกรมและควอนไทล์พล็อต, และค่าสถิติที่ได้จากการทดสอบสมมติฐาน เช่น $p-$value

- ไม่แยกแยะว่าข้อมูลใดเป็นข้อมูลนำเข้า ข้อมูลใดเป็นข้อมูลนำออก เช่น ในระบบแถวคอย ข้อมูลนำเข้าพื้นฐานมีจำนวนที่ลูกค้ามาถึงในหนึ่งหน่วยเวลา, เวลาในการให้บริการ, และจำนวนผู้ให้บริการ ตัวอย่างของข้อมูลนำออก เช่น เวลารอคอยของลูกค้าจนกว่าจะได้รับบริการ, หรือเวลารวมในระบบ ผู้วิเคราะห์จำเป็นต้องกำหนดการแจกแจงของข้อมูลนำเข้าเพื่อใช้จำลองความไม่แน่นอนในแบบจำลอง และใช้ข้อมูลนำออกสำหรับตรวจสอบความถูกต้องของตัวแบบด้วยการเปรียบเทียบค่าจริงที่เก็บได้จากหน้างาน กับเอาท์พุทที่ได้จากแบบจำลอง (หัวข้อ 9.3)

แบบฝึกหัด

4.1. จงเลือกคำตอบที่ถูกต้อง อาจมีมากกว่าหนึ่งข้อ

(i) สิ่งใดต่อไปนี้ขึ้นอยู่กับการแบ่งกลุ่มข้อมูล

a. การทดสอบแบบไคสแควร์

b. การทดสอบแบบโคโมโกรอฟ-สเมอร์นอฟ

c. กราฟความหนาแน่น-ฮีสโตแกรม

d. ควอนไทล์พล็อต

(ii) เมื่อมีข้อมูลจำนวนมาก มีนัยว่าอย่างไร

a. มีโอกาสที่จะเห็นค่าสุดโต่งมากกว่า

b. สามารถค้นพบการแจกแจงที่แท้จริงได้ง่ายขึ้น

c. อาจเป็นได้ว่าไม่มีการแจกแจงใดสอดคล้องกับข้อมูล

d. การแจกแจงที่เลือกจะเป็นตัวแทนของกระบวนการสุ่มที่แท้จริงได้ดีขึ้น

(iii) เมื่อมีข้อมูลจำนวนน้อย มีนัยว่าอย่างไร

a. การทดสอบความสอดคล้องของการแจกแจงทำงานได้ดีกว่าเมื่อมีข้อมูลจำนวนมาก

b. ไม่ควรใช้การจำลองสถานการณ์

c. การแจกแจงความน่าจะเป็นหลายอันดูน่าจะสอดคล้อง

d. เราควรลองใช้แบบจำลองความน่าจะเป็นหลาย ๆ อันเพื่อพิจารณาว่าผลการจำลองจะเปลี่ยนแปลงมากหรือไม่

4.2. อายุการใช้งานของอะไหล่ชิ้นหนึ่งมีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย 5 ปี และอะไหล่นี้ถูกใช้งานอย่างต่อเนื่องมาแล้ว 5 ปี ให้คำนวณความน่าจะเป็นที่อะไหล่นี้จะยังใช้ต่อไปได้อีก 5 ปี

4.3. พิจารณาการทดสอบความแข็งแรงของการเชื่อมที่รอยต่อจนกว่าจะแตกหัก การแตกหักที่จุดเชื่อมเกิดขึ้นด้วยสัดส่วน 80% และอีก 20% เกิดที่แท่งโลหะ มีการทดสอบชิ้นงานหลายชิ้น โดยที่การทดสอบเหล่านี้เป็นอิสระต่อกัน ให้กำหนดการแจกแจงและพารามิเตอร์ที่เหมาะสมสำหรับค่าสุ่มที่สนใจดังต่อไปนี้ และคำนวณคำตอบที่เป็นตัวเลข

(i) การแตกหักที่จุดเชื่อมเกิดขึ้นครั้งแรก เมื่อทดสอบครั้งที่ 4

(ii) การแตกหักครั้งที่ 3 เกิดขึ้นเมื่อทดสอบครั้งที่ 6

4.4. ข้อมูลต่อไปนี้เป็นจำนวนครั้งที่เส้นด้ายขาดต่อเครื่องทอหนึ่งเครื่อง (ทุกเครื่องใช้ด้ายที่มีความยาวเท่ากัน) ดังนี้

26	30	54	25	70	52	51	26	67	18	21	29	17	12	18	35
30	36	36	21	24	18	10	43	28	15	26	27	14	29	19	29
31	41	20	44	42	26	19	16								

(i) สร้างฮีสโตแกรมสำหรับข้อมูลนี้

(ii) ใช้การทดสอบความเหมาะสมของการแจกแจง เพื่อเลือกการแจกแจงมาตรฐานสำหรับข้อมูลนี้ ให้ใช้ α = 5% สังเกตว่าข้อมูลนี้เป็นข้อมูลแบบไม่ต่อเนื่องเพราะเป็นจำนวนครั้ง ระบุการแจกแจงและพารามิเตอร์

(iii) สร้างควอนไทล์พล็อต โดยเปรียบเทียบข้อมูลตามตัวอย่างกับค่าสุ่มที่สร้างมาจากการแจกแจงที่เลือกในข้อ (4.4.ii)

4.5. ข้อมูลเหล่านี้เป็นเวลาในการประกอบชิ้นส่วน 1 ชิ้น (นาที) ดังนี้

5.1	4.9	4.7	4.6	5.0	5.4	4.6	5.0	4.4	4.9	5.4	4.8	4.8	4.3	5.8
5.7	5.4	5.1	5.7	5.1	5.4	5.1	4.6	5.1	4.8	5.0	5.0	5.2	5.2	4.7
4.8	5.4	5.2	5.5	4.9	5.0	5.5	4.9	4.4	5.1					

(i) สร้างฮีสโตแกรมสำหรับข้อมูลนี้

(ii) ใช้การทดสอบความเหมาะสมของการแจกแจง เพื่อเลือกการแจกแจงมาตรฐานสำหรับข้อมูลนี้ ให้ใช้ α = 5% สังเกตว่าข้อมูลนี้เป็นข้อมูลแบบต่อเนื่อง ระบุการแจกแจงและพารามิเตอร์

(iii) ใช้การทดสอบแบบโคโมโกรอฟ-สเมอร์นอฟ เพื่อเลือกการแจกแจงมาตรฐานสำหรับข้อมูลนี้ ให้ใช้ α = 5% ระบุการแจกแจงและพารามิเตอร์

(iv) สร้างควอนไทล์พล็อต โดยเปรียบเทียบข้อมูลตามตัวอย่างกับการแจกแจงที่เลือกในข้อ (4.5.ii)

4.6. มีข้อมูล 100 ค่าที่มีค่าเฉลี่ยตัวอย่าง 5.1 และความแปรปรวนจากตัวอย่างที่ 25.5 สันนิษฐานว่าข้อมูลนี้มาจากการแจกแจงที่ค่อนข้างสมมาตร ให้ประมาณความน่าจะเป็นที่ค่าสุ่มที่จะเก็บได้ค่าต่อไปอยู่ในช่วง 4.5 ถึง 5.5

5

การวิเคราะห์ผลที่ได้จากแบบจำลองเมื่อพิจารณา 1 ระบบ

สถิติเป็นพื้นฐานที่จำเป็นสำหรับการวิเคราะห์ผลที่ได้จากแบบจำลอง เพราะค่าเอาท์พุทของแบบจำลองเป็นค่าสุ่ม สืบเนื่องมาจากแบบจำลองมีตัวแปรนำเข้าเป็นค่าสุ่ม ดังนั้นเอาท์พุทของแต่ละรอบทำซ้ำจึงไม่เหมือนกัน หากไม่พิจารณาความไม่แน่นอนในค่าเอาท์พุทนี้ การตีความอาจผิดพลาดหรือนำไปสู่การตัดสินใจที่ไม่เหมาะสม วิธีการวิเคราะห์ผลขึ้นอยู่กับประเภทของการจำลองสถานการณ์ในแง่กรอบเวลา (Time frame) จุดมุ่งหมายหลักของการวิเคราะห์ผลคือการควบคุมค่าความผิดพลาดในการประมาณค่าดัชนีชี้วัดด้วยแบบจำลอง บทนี้มีหัวข้อหลัก ๆ ดังนี้

- การออกแบบการทดลอง ในแง่ของจำนวนรอบทำซ้ำที่ควรใช้ประมวลผล
- การวิเคราะห์ผลที่ได้จากการประมวลผล

เครื่องมือทางสถิติที่ใช้ในการวิเคราะห์ผลขึ้นอยู่กับลักษณะของข้อมูลว่าเป็นอิสระต่อกัน (Independent) และมีการแจกแจงเดียวกัน (Identically distributed) หรือไม่ ซึ่งเป็นผลที่ตามมาจากประเภทของการจำลองสถานการณ์ที่เลือก ดังนั้นผู้เขียนจึงเริ่มที่หัวข้อนี้

5.1 ประเภทของการจำลองสถานการณ์

การแยกแยะว่าการจำลองสถานการณ์เป็นชนิดใดขึ้นอยู่กับวัตถุประสงค์ของการศึกษาว่าสนใจการดำเนินงานของระบบใน**ช่วงเวลาจำกัด**หรือ**ช่วงเวลาไม่จำกัด** ดังนี้

การจำลองสถานการณ์แบบสิ้นสุด (Terminating simulation) ซึ่งมีสภาวะของระบบที่เวลาเริ่มต้นและเวลาสิ้นสุดที่ชัดเจนและแน่นอน ดังนั้นความยาวของรอบทำซ้ำจึงจำกัดและมีการนิยามอย่างแน่นอน เช่น ธนาคารที่เปิดทำการตอน 8.30 น. ไม่มีลูกค้ารอภายใน พนักงานที่หน้าเคาน์เตอร์ทุกคนพร้อมที่จะให้บริการและธนาคารปิดเวลา 15.30 น. หรือเมื่อลูกค้าคนสุดท้ายออกจากธนาคาร; โครงการเริ่มในวันที่เซ็นสัญญาไว้ และจะสิ้นสุดเมื่อกิจกรรมทุกอย่างได้เสร็จสิ้นลง; พอร์ตโฟลิโอการลงทุนที่สนใจในผลตอบแทนในรอบ 1 ไตรมาส; ดัชนีชี้วัดที่เป็นเอาท์พุทมักขึ้นอยู่กับสภาวะของระบบที่เวลาเริ่มต้นและเวลาสิ้นสุด ดังนั้นการตั้งค่าสภาวะเริ่มต้นจึงเป็นสิ่งสำคัญ เช่น สมมติว่าต้องการประมาณเวลารอคอยของลูกค้าที่ธนาคารในช่วงที่ลูกค้าเข้ามามากระหว่างพักเที่ยง แล้วผู้วิเคราะห์จำลองระบบแค่เฉพาะ 12-13 น. และตั้งค่าจำนวนลูกค้าในระบบเป็นศูนย์ ณ เวลาที่แบบจำลองเริ่มต้นที่ 12 น. ค่าเวลารอคอยของลูกค้าที่ประมาณได้อาจต่ำเกินจริง วิธีแก้ไขคืออาจเริ่มประมวลผลแบบจำลองตั้งแต่เวลาที่ธนาคารเปิดจริงคือ 9 น. แต่ให้แบบจำลองเก็บค่าสถิติเฉพาะช่วง 12-13 น.

การจำลองสถานการณ์แบบไม่สิ้นสุด (Non-terminating simulation) หรือ**ที่กำหนดให้เกิดสภาวะคงตัว** (Steady-state simulation) สนใจที่การดำเนินงานของระบบในระยะยาว ยาวจนกระทั่งสภาวะของระบบที่จุดเริ่มต้นไม่ส่งผลต่อเอาท์พุท และสภาวะของระบบที่เวลาสิ้นสุดมักไม่นิยามชัดเจน เช่น ระบบการ

ผลิตที่ทำงานตลอด 24 ชั่วโมง; ระบบการผลิตที่ทำงาน 2 กะต่อวันโดยที่สภาวะสิ้นสุดของวันนี้เป็นสภาวะเริ่มต้นของวันถัดไป; ธุรกรรมของร้านสะดวกซื้อที่เปิดตลอดเวลา, ระบบอินเตอร์เน็ต, ห้องฉุกเฉินที่โรงพยาบาล, หรือศูนย์แจ้งเหตุด่วนเหตุร้าย ในการใช้งานจริงผู้วิเคราะห์ต้องแน่ใจว่าระยะเวลาการประมวลผลยาวเพียงพอที่ผลกระทบของสภาวะเริ่มต้นแทบจะไม่ส่งผลต่อเอาท์พุท

บางครั้งผู้วิเคราะห์ไม่สามารถจำแนกชนิดของการจำลองชัดเจน ระบบหลายประเภทสามารถจำลองได้ทั้งแบบสิ้นสุดและแบบให้เกิดสภาวะคงตัว แต่จะเป็นชนิดใดขึ้นอยู่กับจุดประสงค์ของดัชนีชี้วัดที่ต้องการประมาณ ตารางที่ 5.1 แสดงตัวอย่างของกรณีเช่นนี้ ทั้งนี้การวิเคราะห์ทางสถิติสำหรับการจำลองสถานการณ์แบบเวลาสิ้นสุดง่ายกว่าแบบไม่สิ้นสุด จึงเป็นที่นิยมมากกว่า แต่การจำลองสถานการณ์แบบไม่สิ้นสุดก็เหมาะสมกับระบบบางชนิดที่ทำงานตลอด 24 ชั่วโมงจริง ๆ หรือเมื่อต้องการออกแบบให้ระบบสามารถรองรับความต้องการในช่วงขีดสุด (Peak) เช่น เพื่อกำหนดจำนวนเจ้าหน้าที่ห้องฉุกเฉินให้รองรับช่วงเทศกาลสงกรานต์หรือปีใหม่ได้ หรือระบบคอมพิวเตอร์เซิร์ฟเวอร์ช่วงที่มีผู้ใช้มาก ๆ เช่น ช่วงเย็น แต่ระบบที่ออกแบบสำหรับรองรับช่วงขีดสุดอาจมีขีดความสามารถ (Capacity) สูงเกินไปหากว่าช่วงขีดสุดเป็นช่วงสั้น ๆ

เนื้อหาส่วนใหญ่ในบทนี้จะกล่าวถึงเรื่องการวิเคราะห์ผลสำหรับการจำลองแบบสิ้นสุด และอธิบายเรื่องการวิเคราะห์ผลสำหรับการจำลองแบบไม่สิ้นสุดเล็กน้อยในหัวข้อที่ 5.4 เพื่อให้ผู้อ่านเข้าใจการประมาณค่าดัชนีชี้วัดจากเอาท์พุท จึงขออธิบายชนิดของกระบวนการสุ่มในหัวข้อถัดไป

ตารางที่ 5.1: ตัวอย่างเปรียบเทียบประเภทการจำลองสถานการณ์

สถานการณ์ทางกายภาพ	**การจำลองแบบสิ้นสุด**	**การจำลองแบบ ให้เกิดสภาวะคงตัว**
แถวคอยที่มีผู้ให้บริการ 1 คน	เวลารอคอยเฉลี่ยจนกว่าจะได้รับบริการของลูกค้า 25 คนแรก เมื่อระบบเริ่มด้วยว่างงาน	เวลารอคอยจนกว่าจะได้รับบริการ เฉลี่ยระยะยาว
ระบบการผลิต	จำนวนชิ้นที่ผลิตได้ต่อวันเฉลี่ยเมื่อมีงานค้างอยู่ในระบบตอนเริ่มต้น	จำนวนชิ้นที่ผลิตได้ต่อวันเฉลี่ย ต่อวันเฉลี่ย
ระบบประเมินความน่าเชื่อถือ	อายุการใช้งานเฉลี่ยหรือความน่าจะเป็นที่อายุการใช้งานเกิน 1 ปี เมื่อชิ้นส่วนทุกชิ้นใหม่และใช้งานได้	อาจไม่เหมาะ

5.2 กระบวนการสุ่ม

เอาท์พุทของแบบจำลองเป็นกระบวนการสุ่มซึ่งแยกได้เป็น 2 ชนิด ตามลักษณะการเกิดขึ้นของข้อมูล ดังนี้

กระบวนการสุ่มแบบเวลาไม่ต่อเนื่อง (Discrete-time random process) หรือ**ข้อมูลเชิงนับ** (Tally data) เป็นข้อมูลที่มีลำดับชัดเจน: ที่หนึ่ง, ที่สอง, ที่สาม, และอื่น ๆ เช่น เวลารอคอยของลูกค้าแต่ละคน, จำนวนครั้งที่ส่งสินค้าตรงเวลาในแต่ละเดือน, ผลกำไรในแต่ละไตรมาส, เวลาในระบบของงานซ่อมแต่ละงาน ตัวอย่างของกระบวนการสุ่มแบบเวลาไม่ต่อเนื่องแสดงในรูปที่ 5.1 ซึ่งเป็นเวลาในระบบของลูกค้า 10 คน ดัชนีชี้วัดที่สามารถคำนวณได้จากข้อมูลนี้ก็เช่น

$$\begin{aligned}
\text{เวลาเฉลี่ยในระบบ } \overline{W}(M) &= \frac{\sum_{i=1}^{M} W_i}{M} \qquad (5.1)\\
\text{สัดส่วนของลูกค้าที่อยู่ในระบบเกิน 1.5 หน่วยเวลา} &= \frac{1}{M}\sum_{i=1}^{M} I\{W_i \geqslant 1.5\},\\
\text{เมื่อ } I\{W_i \geqslant 1.5\} &= \begin{cases} 1 & \text{ถ้า } W_i \geqslant 1.5\\ 0 & \text{ถ้า } W_i < 1.5\end{cases}
\end{aligned}$$

ฟังก์ชัน $I\{E\}$ มีชื่อว่า**ฟังก์ชันบ่งชี้** (Indicator function) ซึ่งให้ผลเป็นค่าจริงทางตรรกะหรือ 1 ถ้าเหตุการณ์ E เป็นจริง และให้ผลเป็นค่าเท็จทางตรรกะหรือ 0 ถ้าเหตุการณ์ E เป็นเท็จ สังเกตว่าสัดส่วนของลูกค้าที่อยู่ในระบบเกิน 1.5 หน่วยเวลาคือจำนวนลูกค้าที่มีเวลาในระบบเกิน 1.5 หารด้วยจำนวนลูกค้าทั้งหมดในกลุ่มตัวอย่าง

ตัวอย่างที่ 5.1. จากรูปที่ 5.1 ซึ่งกราฟของเวลาในระบบของลูกค้า 10 คนดังนี้

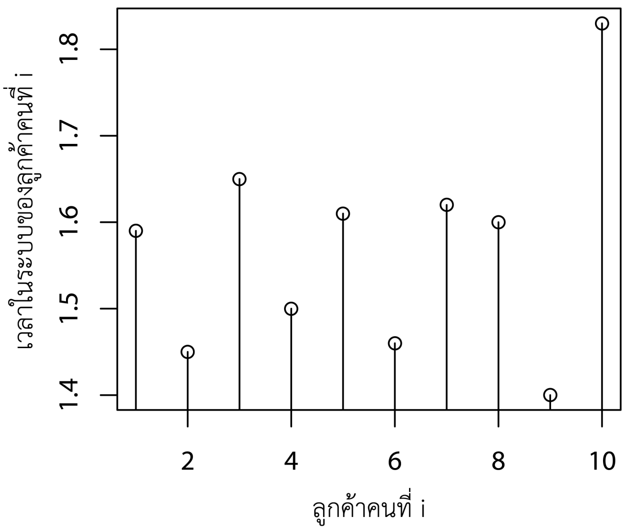
$$1.595, 1.452, 1.657, 1.504, 1.607, 1.465, 1.621, 1.601, 1.401, 1.833$$

สามารถคำนวณดัชนีชี้วัดได้ดังนี้

$$\begin{aligned}
\text{เวลาเฉลี่ยในระบบ } \overline{W}(10) &= \frac{1}{10}\sum_{i=1}^{10} W_i\\
&= \frac{1}{10}\sum_{i=1}^{10}(1.595 + 1.452 + \ldots + 1.833)\\
&= 1.574 \text{ หน่วยเวลา}\\
\text{สัดส่วนของลูกค้าที่อยู่ในระบบ} &= \frac{1}{M}\sum_{i=1}^{M} I\{W_i \geqslant 1.5\}
\end{aligned}$$

$$\begin{aligned} \text{เกิน 1.5 หน่วยเวลา} &= \frac{1}{10}\sum_{i=1}^{10}(I\{1.595 \geqslant 1.5\} + I\{1.452 \geqslant 1.5\} + \ldots \\ &\quad +I\{1.833 \geqslant 1.5\}) \\ &= \frac{1}{10}\sum_{i=1}^{10}(1+0+\ldots+1) = 0.7 \end{aligned}$$

□



รูปที่ 5.1: เวลาในระบบของลูกค้า (กระบวนการสุ่มแบบเวลาไม่ต่อเนื่อง)

กระบวนการสุ่มแบบเวลาต่อเนื่อง (Continuous-time random process) หรือข้อมูลที่เกี่ยวเนื่องกับเวลา (Time-persistent data) ไม่มีลำดับที่ ๆ ชัดเจน เช่น จำนวนคนไข้ในระบบที่เวลา t, จำนวนพนักงานที่กำลังทำงานที่เวลา t, จำนวนเครื่องจักรที่เสีย ณ เวลา t ตัวอย่างของกระบวนการสุ่มแบบเวลาต่อเนื่องแสดงในรูปที่ 5.2 ซึ่งเป็นกราฟจำนวนลูกค้าในระบบ ณ เวลา t บางดัชนีชี้วัดที่สามารถคำนวณได้จากข้อมูลนี้ก็เช่น

$$\text{จำนวนลูกค้าในระบบเฉลี่ย } \bar{L} = \frac{1}{T}\int_0^T L(t)dt \tag{5.2}$$

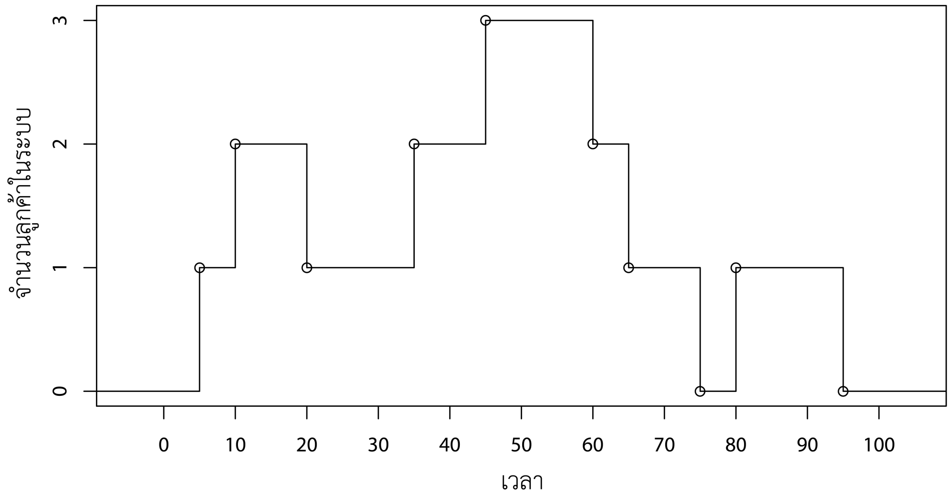
$$\text{ส่วนเบี่ยงเบนมาตรฐานของ } L, S_L = \sqrt{\frac{1}{T}\int_0^T (L(t)-\bar{L})^2 dt} \tag{5.3}$$

$$\text{สัดส่วนของเวลาที่มีลูกค้าอย่างน้อย 2 คนในระบบ} = \frac{1}{T}\int_0^T I\{L(t)\geqslant 2\}dt, \tag{5.4}$$

$$\text{เมื่อ } I\{L(t)\geqslant 2\} = \begin{cases} 1 & \text{ถ้า } L(t)\geqslant 2 \\ 0 & \text{ถ้า } L(t) < 2 \end{cases}$$

สังเกตว่าการหาค่าเฉลี่ยสำหรับกระบวนการสุ่มแบบไม่ต่อเนื่องใช้การบวก (เช่น สมการที่ 5.1) แต่การ

คำนวณค่าเฉลี่ยสำหรับเวลาต่อเนื่องใช้การหาปริพันธ์หรือหาพื้นที่ใต้กราฟ (เช่น สมการที่ 5.2) เพราะการหาปริพันธ์คือการบวกเมื่อจำนวนค่าที่นำมารวมกันเป็นอนันต์ ตัวอย่างที่ 5.2 แสดงการคำนวณจำนวนลูกค้าในระบบเฉลี่ย



รูปที่ 5.2: จำนวนลูกค้าในระบบที่เวลาต่าง ๆ (กระบวนการสุ่มแบบเวลาต่อเนื่อง)

ตัวอย่างที่ 5.2. จากรูปที่ 5.2 สามารถคำนวณดัชนีชี้วัดจากพื้นที่ใต้กราฟ เมื่อ $T = 100$ ได้ดังนี้

$$\text{จำนวนลูกค้าในระบบเฉลี่ย } \bar{L} = \frac{140}{100} = 1.4 \tag{5.5}$$

$$\text{สัดส่วนของเวลาที่มีลูกค้าอย่างน้อย 2 คนในระบบ} = \frac{40}{100} = 0.4 \text{ หรือ } 40\% \tag{5.6}$$

ในสมการ (5.5) พื้นที่ใต้กราฟในรูปที่ 5.2 คือ 140 และในสมการ (5.6) พื้นที่ใต้กราฟของฟังก์ชันบ่งชี้ (5.4) คือ 40 □

กระบวนการที่เวลาต่อเนื่องที่สำคัญอีกอันหนึ่ง คือ สัดส่วนของเวลาที่ทำงานจริง (Utilization, ρ) ในกรณีที่มีผู้ให้บริการ 1 คน สามารถคำนวณ ρ ได้จากการคำนวณพื้นที่ใต้กราฟของฟังก์ชัน $B(t)$ ที่มีนิยามดังนี้

$$B(t) = \begin{cases} 1 & \text{ถ้าผู้ให้บริการกำลังทำงาน ณ เวลา } t \\ 0 & \text{ถ้าผู้ให้บริการว่าง ณ เวลา } t \end{cases}$$

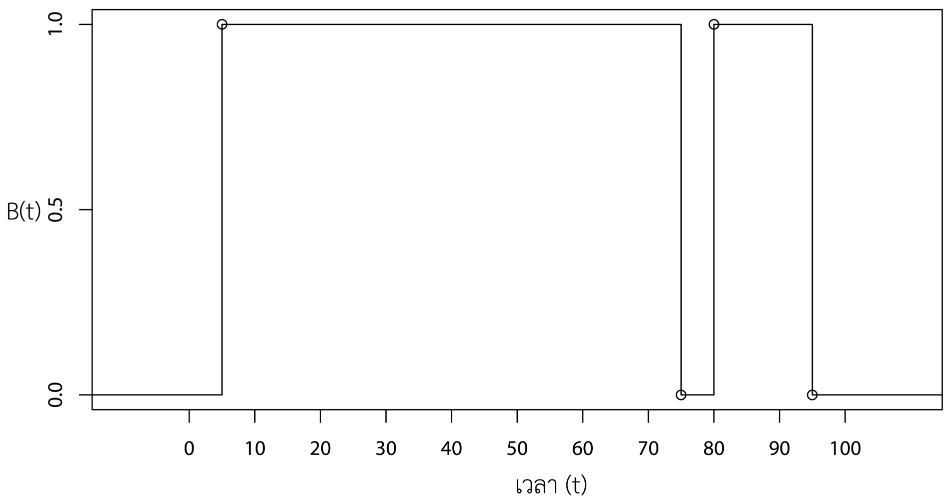
รูปที่ 5.3 แสดงกราฟของ $B(t)$ ซึ่งนำมาคำนวณ ρ ได้ดังนี้

$$\rho = \frac{1}{T}\int_0^T B(t)dt$$

หากมีผู้ให้บริการมากกว่า 1 คน สัดส่วนของเวลาที่ทำงานจริงจะคำนวณจากสัดส่วนของจำนวนผู้ให้บริการที่ทำงาน ณ เวลา t, $S(t)$, เฉลี่ย เมื่อเทียบกับจำนวนผู้ให้บริการทั้งหมด (n) ดังนี้

$$\rho = \frac{1}{n}\left(\frac{1}{T}\int_0^T S(t)dt\right)$$

ตัวอย่างที่ 5.3 แสดงวิธีการคำนวณสัดส่วนของเวลาที่ทำงานจริง



รูปที่ 5.3: ฟังก์ชันที่แสดงว่าผู้ให้บริการทำงานอยู่หรือไม่ (1 คือกำลังทำงานอยู่ และ 0 คือว่าง)

ตัวอย่างที่ 5.3. จากรูปที่ 5.3 สามารถคำนวณสัดส่วนของเวลาที่ทำงานจริงได้ดังนี้

$$\rho = \frac{1}{100}((75-5)+(95-80)) = 85\%$$

□

ในการจำลองแบบเหตุการณ์สิ้นสุด การทดลองก็คือการประมวลผลหลาย ๆ รอบทำซ้ำจากเวลา 0 ถึงเวลาที่สิ้นสุด T_E หัวข้อต่อไปกล่าวถึงประเด็นหลักที่ต้องคำนึงถึง

5.3 การวิเคราะห์ผลสำหรับการจำลองสถานการณ์แบบสิ้นสุด

ประเด็นที่ผู้วิเคราะห์ควรตอบตนเองได้ มีดังนี้

- ดัชนีชี้วัดใดที่ต้องการ

- จะประเมินความผิดพลาดในค่าประมาณของดัชนีชี้วัดอย่างไร
- จะประเมินผลด้วยกี่รอบทำซ้ำ

สมมติว่าแบบจำลองประเมินเวลาส่งมอบโครงการอันหนึ่ง (Y) ถึงแม้ผู้วิเคราะห์ได้กำหนดแล้วว่าดัชนีชี้วัดหลักคือ Y ค่าสถิติที่เกี่ยวข้องกับ Y มีหลากหลาย เช่น

1. **ค่าเฉลี่ย** $\mu = \mathrm{E}(Y)$
2. **ความน่าจะเป็นที่จะเสร็จทันกำหนด** $\Pr\{Y \leqslant c\}$
3. **ควอนไทล์หรือเปอร์เซ็นไทล์** ซึ่งคือค่า γ ที่ทำให้สมการที่ $\Pr\{Y \leqslant \gamma\} = p$ เป็นจริง มองอีกนัยหนึ่งคือระดับการดำเนินการที่เกิดขึ้นได้ด้วยความน่าจะเป็นที่ผู้วิเคราะห์กำหนด

ค่าสถิติเหล่านี้ประมาณได้จากเอาท์พุทของแบบจำลอง หากทำ R รอบทำซ้ำ จะได้เอาท์พุทเวลาที่ทำโครงการเสร็จ R ค่า คือ $Y_1, Y_2, \ldots, Y_R$ เนื่องจากแต่ละรอบทำซ้ำใช้เลขสุ่มที่เป็นอิสระต่อกัน ค่าเอาท์พุทเหล่านี้จึง**เป็นอิสระต่อกัน** เพราะว่าใช้แบบจำลองเดียวกันในทุก ๆ รอบการประมวลผล ค่าเอาท์พุทเหล่านี้จึง**มีการแจกแจงเดียวกัน** ความเป็นอิสระต่อกันและมีการแจกแจงเดียวกัน (Independent and identically distributed, i.i.d.) เป็นข้อสมมติที่สำคัญมากของเครื่องมือวิเคราะห์ทางสถิติพื้นฐานที่คนทั่วไปรู้จักและที่ใช้เป็นส่วนใหญ่ในหนังสือเล่มนี้ ไม่ว่าจะเป็นการทดสอบสมมติฐาน (หัวข้อที่ 4.3.1) หรือการคำนวณค่าสรุปทางสถิติ เช่น ค่าเฉลี่ย, หรือส่วนเบี่ยงเบนมาตรฐาน (หัวข้อที่ 2.5.1) ดังนั้นผู้วิเคราะห์จึงควรตรวจสอบข้อสมมติ i.i.d. ว่าเป็นจริงโดยประมาณสำหรับเอาท์พุทที่มีหรือไม่ ตัวอย่างของเอาท์พุทจากแบบจำลองที่มีลักษณะ i.i.d. เช่น

- **แต่ละรอบทำซ้ำให้เพียงค่า ๆ เดียว** เช่น ในตัวอย่างการวางแผนงบประมาณของบริษัทไอทีในหัวข้อที่ 3.2.2 แต่ละรอบทำซ้ำให้ค่าใช้จ่ายรวม 5 ปี จากตารางที่ 3.7ข จะเห็นว่า $Y_1 = 5.34 \times 10^6, Y_2 = 6.09 \times 10^6, \ldots, Y_{99} = 5.23 \times 10^6$, และ $Y_{100} = 5.30 \times 10^6$
- **ค่าเฉลี่ยของค่าย่อยในรอบทำซ้ำนั้น ๆ** เช่น ในตัวอย่างปัญหาสินค้าคงคลัง (หัวข้อที่ 3.4) ตารางที่ 3.15ข แสดงผลของแต่ละรอบทำซ้ำ คอลัมน์ `Var 2` เป็นค่าสต๊อกเฉลี่ยซึ่งเป็นค่าที่คำนวณจากระดับสต๊อกตอนสิ้นวันของทุก ๆ วันในรอบทำซ้ำนั้น $Y_1 = 28.03, Y_2 = 26.17, \ldots, Y_{99} = 31.40$, และ $Y_{100} = 29.77$

สิ่งที่พึงระวังคือเอาท์พุทที่เกิดขึ้นในรอบทำซ้ำเดียวกันอาจ**ไม่เป็นอิสระต่อกัน** เช่น สำหรับตัวอย่างสินค้าคงคลังในหัวข้อที่ 3.4 ระดับสินค้าคงคลังของสิ้นวันที่ $t+1$ คือ ระดับสินค้าคงคลังของวันที่ t บวกด้วยสินค้าที่เพิ่งมาส่งวันนั้น หักลบด้วยจำนวนสินค้าที่ขายไป ดังนั้นระดับสินค้าคงคลังในวันที่ t และของวันที่ $t+1$ จึงไม่เป็นอิสระต่อกัน หรือมีความสัมพันธ์กัน (Correlation) นอกจากนี้เอาท์พุทที่เกิดขึ้นภายในรอบทำซ้ำเดียวกัน

ก็อาจ**มีการแจกแจงแตกต่างกัน** เช่น เวลารอของลูกค้าตอนเริ่มเปิดธนาคารมักจะสั้นกว่าตอนที่ธนาคารได้เปิดทำการไปแล้วระยะหนึ่ง เพราะตอนที่เพิ่งเปิดยังไม่มีลูกค้ารอและพนักงานทุกคนว่าง

หัวข้อต่อไปจะกล่าวถึงการประมาณค่าสำหรับค่าสรุปทางสถิติที่ผู้วิเคราะห์มักสนใจ

การประมาณค่าแบบจุดและแบบช่วง

สมมติว่าค่าดัชนีชี้วัดที่สนใจเป็น Y ผู้วิเคราะห์สามารถรายงานค่าประมาณของค่าสถิติสำหรับ Y เช่น $\mu = \mathrm{E}(Y)$ ได้ 2 แบบ คือ **ค่าที่จุดเดียว** (Point estimator) เช่น $\hat{\mu}$ (สัญลักษณ์ $\hat{}$ บนค่าสถิติแสดงว่าค่านั้นเป็นค่าประมาณ) หรือ**ช่วงประมาณ** (Interval estimator) เช่น ช่วงความเชื่อมั่น ผู้เขียนจะกล่าวถึงค่าประมาณทั้งสองแบบสำหรับแต่ละค่าสถิติของ Y ในหัวข้อที่ 5.3.1-5.3.5 ต่อไปนี้ สมมติว่าผู้วิเคราะห์มีค่าเอาท์พุทแบบเชิงนับจากแบบจำลอง R ค่า: $Y_1, Y_2, \ldots, Y_R$

5.3.1 ค่าเฉลี่ย

ค่าเฉลี่ยหรือค่าคาดคะเนของตัวแปรสุ่ม Y แทนด้วยสัญลักษณ์ $\mathrm{E}[Y]$ หรือ μ_Y ซึ่งสามารถรายงานค่าประมาณได้ดังนี้

- **ค่าที่จุดเดียว** คือ ค่าเฉลี่ยจากเอาท์พุท:

$$\hat{\mu}_Y = \bar{Y} = \frac{1}{R}\sum_{i=1}^{R} Y_i \tag{5.7}$$

  สมการที่ (5.2) แสดงสูตรคำนวณค่าเฉลี่ยสำหรับข้อมูลที่เกี่ยวเนื่องกับเวลา (Time-dependent data)

- **ช่วงประมาณ**ของค่าเฉลี่ย คือ **ช่วงความเชื่อมั่น** (Confidence interval) ผู้วิเคราะห์คำนวณค่าเฉลี่ยด้วยสมการที่ (5.7) และส่วนเบี่ยงเบนมาตรฐานของเอาท์พุทด้วยสมการต่อไปนี้

$$S_Y = \sqrt{\frac{1}{R-1}\sum_{i=1}^{R}(Y_i - \bar{Y})^2} \tag{5.8}$$

  สมการที่ (5.3) แสดงสูตรคำนวณส่วนเบี่ยงเบนมาตรฐานสำหรับข้อมูลที่เกี่ยวเนื่องกับเวลา หลังจากกำหนดระดับนัยสำคัญ (Significance level, α) ที่มักใช้ 95% ผู้วิเคราะห์ค้นหาค่าสถิติ t ซึ่งขึ้นกับองศาอิสระ (Degree of freedom) ในที่นี้คือ $R-1$ และระดับนัยสำคัญ α แล้วจึงนำมาคำนวณครึ่งช่วงความเชื่อมั่น (Half width) ดังนี้

$$\text{Half width } = t_{1-\alpha/2,R-1}\frac{S_Y}{\sqrt{R}} \tag{5.9}$$

สัดส่วน $S_Y/\sqrt{R}$ เรียกว่า**ค่าความคลาดเคลื่อนมาตรฐาน** (Standard error) เป็นความเบี่ยงเบนเฉลี่ยที่ $\bar{Y}$ ห่างจาก μ ซึ่งเป็นค่าจริงที่ไม่ทราบค่า ซอฟต์แวร์สำเร็จรูปสำหรับการจำลองสถานการณ์ เช่น อารีนารายงานค่าครึ่งช่วงความเชื่อมั่นที่นัยสำคัญ 95% ดังนั้นหากผู้วิเคราะห์ต้องการช่วงความเชื่อมั่นที่นัยสำคัญอื่น จะต้องย้อนไปคำนวณค่าส่วนเบี่ยงเบนมาตรฐานก่อน

เมื่อได้ครึ่งช่วงความเชื่อมั่นจากสมการที่ (5.9) แล้ว จึงคำนวณช่วงความเชื่อมั่น ดังนี้

$$\bar{Y} - \text{ Half width } \leqslant \mu \leqslant \bar{Y} + \text{ Half width} \tag{5.10}$$

สมการที่ (5.9) มีนัยว่าหากต้องการลดความกว้างของช่วงความเชื่อมั่น ผู้วิเคราะห์ไม่สามารถควบคุม S_Y และ $t_{1-\alpha/2,R-1}$ ได้โดยตรง จึงเหลือแต่การกำหนดจำนวนรอบทำซ้ำ (R) หากต้องการให้ช่วงความเชื่อมั่นแคบลง ให้เพิ่มจำนวนรอบทำซ้ำ (R) แต่ทั้งนี้ความกว้างของช่วงความเชื่อมั่นลดลงอย่างช้า ๆ เมื่อเพิ่ม R เป็น $4R$ ก็สามารถลดความกว้างของช่วงความเชื่อมั่นได้ครึ่งหนึ่งโดยประมาณ ตัวอย่างที่ 3.2 แสดงขั้นตอนการคำนวณช่วงความเชื่อมั่น

ถึงแม้ว่าผู้วิเคราะห์ต้องการรายงาน $\hat{\mu}$ ด้วยค่า ๆ เดียว จำนวนเลขนัยสำคัญ (Significance figure) ก็เป็นเรื่องสำคัญที่ควรใส่ใจ ความกว้างของช่วงความเชื่อมั่นเป็นดัชนีที่ระบุว่าควรรายงาน $\hat{\mu}$ ด้วยจำนวนจุดทศนิยมกี่ตำแหน่ง เช่น ในตัวอย่าง 3.2 ช่วงความเชื่อมั่นของค่าคาดคะเนคือ 5.665 $\pm$ 0.0466 ค่าความกว้างของช่วงความเชื่อมั่นมีตัวเลขไม่ใช่ศูนย์ตัวแรกอยู่ที่จุดทศนิยมตำแหน่งที่สอง ดังนั้นจึงควรรายงาน $\hat{\mu}$ เพียงแค่ถึงจุดทศนิยมอันดับที่สอง คือ 5.67 เพราะจุดทศนิยมตำแหน่งอื่น ๆ นั้นไม่ต่างจากศูนย์อย่างมีนัยสำคัญ ตัวอย่างที่ 5.4 แสดงการรายงานผลการประมวลผลด้วยจำนวนเลขนัยสำคัญที่เหมาะสม

ตัวอย่างที่ 5.4. แบบจำลองให้ช่วงความเชื่อมั่นของต้นทุน 3,713,801 $\pm$ 766,218 บาท ถ้าหากต้องการรายงานแค่ค่าเฉลี่ยเพียงค่าเดียว ควรรายงานตัวเลขใด

คำตอบ 3.7 ล้านบาทเพราะครึ่งช่วงความเชื่อมั่นมีหลักที่สำคัญหลักแรกคือหลักแสน

□

ช่วงความเชื่อมั่นขึ้นอยู่กับข้อมูลที่นำมาคำนวณ ถ้าประมวลผลอีกครั้งด้วยเลขสุ่มชุดใหม่ หรือต้นกำเนิดของเลขสุ่ม (Seed) ใหม่ ก็จะได้เอาท์พุทอีกชุดหนึ่ง ซึ่งเมื่อนำมาสร้างช่วงความเชื่อมั่นก็จะเป็นอีกช่วงหนึ่งซึ่งไม่ใช่ช่วงความเชื่อมั่นอันเดิม ถึงแม้ว่าจะประมวลผลจากแบบจำลองเดียวกันก็ตาม ช่วงความเชื่อมั่นเป็นช่วงที่ครอบค่าเฉลี่ยที่แท้จริง μ ที่เราไม่ทราบค่า ด้วยความน่าจะเป็น $1-\alpha$ **ไม่ใช่**ช่วงที่ครอบคลุม 95% ของข้อมูล (ช่วงดังกล่าวคือ**ช่วงการทำนาย** ซึ่งจะอธิบายเพิ่มเติมในหัวข้อที่ 5.3.2)

สูตรช่วงความเชื่อมั่น (สมการที่ 5.9-5.10) ตั้งอยู่บนข้อสมมติที่ว่า $Y_1, Y_2, \ldots, Y_R$ ที่นำมาคำนวณมีการแจกแจงแบบปกติ ซึ่งอาจเป็นจริงโดยประมาณหาก Y_i เองก็เป็นค่าเฉลี่ยด้วย ตามที่ระบุด้วยทฤษฎีแนวโน้มเข้าสู่ศูนย์กลาง (Central limit theorem, Theorem 4.1 ใน Law 2006) เช่น Y_i เป็นเวลาเฉลี่ยที่ลูกค้ารอ

ในระบบของทั้งรอบทำซ้ำที่ i แต่ข้อสมมติที่ว่าข้อมูลมีการแจกแจงแบบปกติอาจไม่จริงหากค่าเอาท์พุทเป็นค่าสุดโต่ง (Extreme value) เช่น ค่ามากที่สุดของเวลาที่ลูกค้ารอในรอบทำซ้ำที่ i

ช่วงความเชื่อมั่นเป็นค่าสรุปทางสถิติ แต่หากต้องการกล่าวอ้างถึงอนาคต ควรใช้ช่วงการทำนายหรือควอนไทล์

5.3.2 ช่วงการทำนาย

ควอนไทล์สามารถระบุช่วงของค่าตัวแปรสุ่มที่อาจจะเกิดขึ้นได้ในอนาคต ค่าทางสถิติอีกอย่างหนึ่งที่คล้ายกันคือช่วงการทำนาย ซึ่งเป็นช่วงที่ครอบคลุมค่าของดัชนีชี้วัดที่อาจเกิดขึ้นในอนาคตด้วยความน่าจะเป็นสูง มีสูตรการคำนวณดังนี้

$$\bar{Y} - t_{1-\alpha/2,R-1} S_Y \sqrt{1+\frac{1}{R}} \leqslant Y \leqslant \bar{Y} + t_{1-\alpha/2,R-1} S_Y \sqrt{1+\frac{1}{R}} \tag{5.11}$$

โดยที่ $\bar{Y}$ คำนวณจากสมการที่ (5.7) และ S_Y คำนวณจากสมการที่ (5.8) ช่วงการทำนายต้องรองรับความไม่แน่นอนที่เกิดตามธรรมชาติของกระบวนการสุ่ม และความไม่แน่นอนจากการประมาณค่า μ ด้วย $\bar{Y}$ ซึ่งสะท้อนผ่านพจน์ 1 และพจน์ $1/R$ ในรากที่สองของสมการที่ (5.11) ตามลำดับ ช่วงการทำนายจึงกว้างกว่าช่วงความเชื่อมั่น ตัวอย่างที่ 5.5 แสดงการคำนวณช่วงการทำนาย

ตัวอย่างที่ 5.5. แบบจำลองที่ประมวลผล 120 รอบให้ค่าเฉลี่ยของเวลาครบรอบ (Cycle time) 5.80 ชั่วโมงและมีส่วนเบี่ยงเบนมาตรฐาน 1.60 ชั่วโมง ให้คำนวณและตีความช่วงการทำนายที่นัยสำคัญ 95%

ค่าสถิติ $t_{(1-0.05/2,120-1)} = t_{0.975,119} = 1.98$ อาจได้มาโดยการใช้สูตรเอกซ์เซล `=tinv(0.05, 119)` (เพราะคำสั่งเอกซ์เซลต้องการความน่าจะเป็นที่สองหาง) แทนค่าในสมการที่ (5.11) ได้ดังนี้

$$5.80 \pm 1.98 \times 1.60\sqrt{1+\frac{1}{120}} = 5.80 \pm 3.647 = [2.2, 9.4] \text{ ชั่วโมง}$$

ความน่าจะเป็นที่เวลาครบรอบที่จะเกิดขึ้นในอนาคตจะมีค่าอยู่ในช่วง [2.2, 9.4] ชั่วโมง ด้วยความน่าจะเป็น 95% หรือมีความน่าจะเป็น 5% ที่ค่าในอนาคตจะอยู่นอกช่วงนี้

□

ช่วงความเชื่อมั่นและช่วงการทำนายแตกต่างกันตรงที่ช่วงความเชื่อมั่นเป็นการอนุมานจากข้อมูลในอดีต ส่วนช่วงการทำนายกล่าวถึงอนาคต

ความแตกต่างระหว่างช่วงความเชื่อมั่นและช่วงการทำนาย

ช่วงความเชื่อมั่นบอกถึงค่าคลาดเคลื่อนการประมาณ (Estimation error) หรือช่วงที่เป็นไปได้สูงของพารามิเตอร์ $\mu_Y = \mathrm{E}[Y]$ ที่ต้องการประมาณ แต่ช่วงการทำนายบอกถึงช่วงที่เป็นไปได้ของ Y ที่จะประมวลผลหรือสุ่มได้

ในอนาคต ดังนั้นเมื่อเปรียบเทียบสูตรช่วงความเชื่อมั่น (สมการที่ 5.9) และช่วงการทำนาย (สมการที่ 5.11) จะเห็นว่าช่วงการทำนายกว้างกว่าเพราะมีความไม่แน่นอนของอนาคตเพิ่มด้วย นอกจากนี้แล้วเมื่อเพิ่มขนาดข้อมูล R ความกว้างของช่วงความเชื่อมั่นจะลดลงเรื่อย ๆ และลู่เข้าสู่ค่าที่แท้จริงคือ μ แต่ช่วงการทำนายจะเริ่มคงตัวที่ช่วง $\mu \pm z_{(1-\alpha/2)}\sigma$ เมื่อ σ เป็นส่วนเบี่ยงเบนมาตรฐานประชากรของ Y หรือ $\sigma^2 = \mathrm{E}[(Y-\mu_Y)^2]$

นอกจากค่าเฉลี่ยของดัชนีชี้วัด ผู้วิเคราะห์สามารถพิจารณาความน่าจะเป็นของดัชนีชี้วัดได้

5.3.3 ความน่าจะเป็น

ความน่าจะเป็นแสดงค่าโอกาสที่ดัชนีชี้วัดที่สนใจ Y จะน้อยกว่าค่าเป้าหมาย c หรือมากกว่าค่าเป้าหมาย c เช่น ความน่าจะเป็นที่เวลารอคอยของคนไข้ตั้งแต่เข้าโรงพยาบาลจนกระทั่งได้พบแพทย์ (Y) ไม่เกิน 30 นาที ($c = 30$ นาที)

$$p = \Pr\{Y \leqslant c\}$$

ความน่าจะเป็นที่เวลารอคอยของลูกค้าไม่เกินเป้าหมายแบบนี้แสดง**ระดับการให้บริการ** (Service level) สำหรับระบบบริการ เช่น โรงพยาบาล ธนาคาร หรือศูนย์ให้บริการลูกค้าทางโทรศัพท์ (Call center) ความน่าจะเป็นสามารถประมาณด้วยค่า ๆ เดียวหรือเป็นช่วง ดังนี้

- **ค่าที่จุดเดียว** หรือ $\hat{p}$ ประมาณด้วยจำนวนเอาท์พุท Y_i ที่ไม่เกิน R หารด้วยจำนวนข้อมูลทั้งหมด

$$\hat{p} = \frac{1}{R}\sum_{i=1}^{R} I\{Y_i \leqslant c\}, \text{เมื่อ } I\{Y_i \leqslant c\} = \begin{cases} 1 & \text{ถ้า } Y_i \leqslant c \\ 0 & \text{ถ้า } Y_i > c \end{cases}$$

  เช่น ตัวอย่างที่ 5.1 ประมาณสัดส่วนของลูกค้าที่อยู่ในระบบเกิน 1.5 หน่วยเวลา

  เพื่อหาค่า $\sum_{i=1}^{R} I\{Y_i \leqslant c\}$ จากซอฟต์แวร์สำเร็จรูป อาจต้องนิยามตัวแปรเพิ่มในแบบจำลองสำหรับเก็บค่านี้โดยเฉพาะหรืออาจเพิ่มโมดูลสำหรับนับ หรืออาจประมาณอย่างคร่าว ๆ จากฮีสโตแกรม

- **ช่วงประมาณ**ของความน่าจะเป็นก็เป็นช่วงความเชื่อมั่นเช่นกัน กำหนดให้ $z_{1-\alpha/2}$ เป็นควอนไทล์ของการแจกแจงแบบปกติมาตรฐาน (ค่าเฉลี่ย = 0 และความแปรปรวน = 1) ที่ความน่าจะเป็นที่หางซ้ายหรือหางขวาเป็น $\alpha/2$ สมมติว่า Z มีการแจกแจงปกติมาตรฐาน ($Z \sim \mathrm{N}(0,1)$) คำนิยามของ $z_{1-\alpha/2}$ คือ

$$\Pr\{-z_{1-\alpha/2} \leqslant Z \leqslant z_{1-\alpha/2}\} = 1-\alpha \text{ หรือ } \Pr\{Z \leqslant -z_{1-\alpha/2}\} = \Pr\{Z \geqslant z_{1-\alpha/2}\} = \alpha/2$$

  เช่น $z_{1-0.05/2} = 1.96$ ผู้อ่านสามารถหา $z_{1-\alpha/2}$ ได้จากตารางท้ายหนังสือสถิติพื้นฐานทั่วไป, ใช้ฟังก์ชัน `=NORMSINV(1-α/2)` ในเอกซ์เซล, หรือค้นจากอินเตอร์เน็ต ช่วงความเชื่อมั่นของค่าประมาณความน่าจะเป็น $\hat{p}$ มีสูตรดังนี้

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{R}} \leqslant p \leqslant \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{R}} \tag{5.12}$$

ช่วงความเชื่อมั่นนี้เหมาะสมก็ต่อเมื่อขนาดข้อมูล R สูงพอที่จะทำให้ $t_{1-\alpha/2,R-1} \approx z_{1-\alpha/2}$ คือประมาณ $R \geqslant 200$

ตัวอย่างที่ 5.6 แสดงการคำนวณความน่าจะเป็นที่ดัชนีชี้วัดจะมีค่าเกินเป้าหมาย

ตัวอย่างที่ 5.6. ใช้ตัวอย่างบริษัทมายาในหัวข้อที่ 3.2.2 สมมติว่าบริษัทมีงบประมาณสำหรับทุนการศึกษา 5.5 ล้านบาทและต้องการทราบโอกาสที่จะไม่เกินงบฯ จากการประมวลผล 100 รอบทำซ้ำ มี 24 รอบที่ให้ค่าใช้จ่ายไม่เกิน 5.5 ล้านบาท ดังนั้น

$$\hat{p} = \frac{24}{100} = 24\%$$

และครึ่งช่วงความเชื่อมั่นที่นัยสำคัญ 95% คือ

$$z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{R}} = 1.96 \times \sqrt{\frac{0.24 \times (1-0.24)}{100}} = 0.084$$

ดังนั้นช่วงความเชื่อมั่นคือ $[0.24 - 0.084, 0.24 + 0.084] = [0.156, 0.324]$ หรือควรรายงานแค่ถึงจุดทศนิยมตำแหน่งที่สอง เพราะครึ่งช่วงความเชื่อมั่นมีหลักแรกที่ตำแหน่งนั้น กล่าวคือควรรายงานผลเป็น [0.16, 0.32] โอกาสที่ค่าใช้จ่ายจะไม่เกินงบฯ อยู่ในช่วง 16% ถึง 32%

□

ควอนไทล์เป็นด้านตรงข้ามของความน่าจะเป็นที่เพิ่งพิจารณาไปในหัวข้อที่แล้ว

5.3.4 ควอนไทล์

ผู้วิเคราะห์ระบุความน่าจะเป็นที่สนใจ p และประมาณค่าควอนไทล์ $\hat{\gamma}$ ซึ่งแสดงทางทฤษฎีได้ดังนี้

$$\Pr\{Y \leqslant \gamma\} = p$$

ควอนไทล์หรือเปอร์เซ็นไทล์ที่ p โดยที่ $0\% < p < 100\%$, ให้ขอบเขตของค่าตัวแปรสุ่มที่อาจเป็นได้ภายใต้ความน่าจะเป็น p ที่ผู้ใช้ระบุ เช่น เปอร์เซ็นไทล์ที่ 90 ของต้นทุนเป็นค่าที่มีโอกาสเพียง 10% ที่ต้นทุนจะเกินค่านี้

- **ค่าที่จุดเดียว** คำนวณได้ด้วยการเรียงข้อมูลจากน้อยไปมาก และประมาณค่าควอนไทล์ที่ p โดยที่ $0 < p < 1$, ด้วยข้อมูลลำดับที่ Rp ดังที่ได้อธิบายไปแล้วในหัวข้อที่ 2.5.1 ตัวอย่างที่ 2.15-2.16 แสดงการคำนวณควอร์ไทล์ซึ่งมีได้แค่ 3 แบบคือควอร์ไทล์ที่หนึ่ง, ที่สอง, และที่สาม เพราะเป็นควอนไทล์ที่ 25%, 50% และ 75% ตามลำดับ ดังนั้นในตัวอย่างต่อไปนี้จึงให้ผู้อ่านทดลองคำนวณควอนไทล์เอง และตรวจคำตอบจากเฉลย ตัวอย่างที่ 5.7 แสดงการคำนวณและการตีความควอนไทล์

ตัวอย่างที่ 5.7. ให้ประมาณค่าควอนไทล์ที่ 80% จากข้อมูลเวลาที่ประกอบชิ้นงานเสร็จ 1 ชิ้น (X) 10 ค่า (หน่วย: นาที) ดังนี้

$$12.2 \;\; 8.8 \;\; 9.7 \;\; 7.1 \;\; 12.9 \;\; 10.1 \;\; 8.9 \;\; 5.6 \;\; 12.5 \;\; 9.5$$

และให้อธิบายความหมายของค่าที่คำนวณได้

หากไม่ใช้การประมาณค่าในช่วง (Interpolate) ควอนไทล์ที่ 80% คือ 12.2 นาที แต่หากประมาณค่าในช่วง จะได้คำตอบใกล้เคียงกัน คือ 12.26 นาที นัยของค่าควอนไทล์ที่ 80% คือค่าเวลาที่ประกอบชิ้นงานไม่เกิน 12.2 นาที ด้วยความน่าจะเป็น 80% หรือมีโอกาส 20% ที่อาจมากกว่า 12.2 นาที

ควอนไทล์สามารถให้ช่วงของค่าที่จะเกิดในอนาคต เช่น ในตัวอย่าง 5.7 ควอนไทล์ที่ 20% คือ 7.1 นาที ดังนั้น

$$\Pr\{7.1 \leqslant X \leqslant 12.2\} = \Pr\{X \leqslant 12.2\} - \Pr\{X \leqslant 7.1\} = 0.8 - 0.2 = 0.6$$

ดังนั้นความน่าจะเป็นที่ค่า X จะตกอยู่ช่วง [7.1, 12.2] คือ 60%

□

- **ช่วงประมาณ**ของควอนไทล์ใช้ประโยชน์จากช่วงความเชื่อมั่นของความน่าจะเป็น (สมการที่ 5.12) กล่าวคือ เพื่อคำนวณ**ขอบเขตของนัยสำคัญ** (Confidence limit) ของ p ดังนี้

$$\begin{aligned} \text{ขอบเขตล่าง: } p_\ell &\leftarrow p - z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{R}} \\ \text{ขอบเขตบน: } p_u &\leftarrow p + z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{R}} \end{aligned}$$

สมมติให้ข้อมูลที่เรียงจากน้อยไปมากแล้วแสดงได้ดังนี้

$$Y_{[1]} \leqslant Y_{[2]} \leqslant \cdots \leqslant Y_{[R]}$$

ในตัวอย่าง 5.7 $Y_{[1]} = 5.6$ และ $Y_{[10]} = 12.9$ เมื่อได้ p_ℓ และ p_u แล้ว ค่าช่วงความเชื่อมั่นของควอนไทล์สามารถประมาณได้จากข้อมูลลำดับที่ Rp_ℓ และ Rp_u หรือ $[Y_{[Rp_\ell]}, Y_{[Rp_u]}]$

วิธีนี้เหมาะสมเมื่อมีจำนวนข้อมูล R มากพอ ซึ่งควรมากแค่ไหนก็ขึ้นอยู่กับค่า p: หาก p ยิ่งสุดโต่ง (ใกล้ 0% หรือ 100% มาก ๆ) ยิ่งต้องการจำนวนข้อมูลมากขึ้นเพื่อให้ได้ช่วงความเชื่อมั่นของควอนไทล์ที่ใช้ประโยชน์ได้จริง ทั้งนี้หากไม่มีข้อมูลดิบเพื่อนำมาเรียงลำดับ ก็สามารถใช้ฮีสโตแกรมที่สร้างกราฟความน่าจะเป็นสะสม (เช่นรูปที่ 4.16) โดยค่าประมาณของขอบเขตควอนไทล์คือค่าแกนนอนที่ความน่าจะเป็นสะสมที่ p_ℓ และ p_u ตัวอย่างที่ 5.8 แสดงการคำนวณช่วงความเชื่อมั่นของควอนไทล์

ตัวอย่างที่ 5.8. มีเอาท์พุทจากแบบจำลอง 1000 ค่าที่ต้องการนำมาประมาณควอนไทล์ที่ 80% และช่วงความเชื่อมั่นของควอนไทล์ที่นัยสำคัญ 95% ดังนั้นจึงสามารถประมาณค่าควอนไทล์ที่ 80% จากข้อมูลที่เรียงจากน้อยไปมากแล้วลำดับที่ $Rp = 1000 \times 0.8 = 800$ หรือ $Y_{[800]}$ และ

$$\begin{aligned} p_\ell &= 0.8 - z_{1-0.05/2}\sqrt{\frac{0.8 \times (1-0.8)}{1000}} = 0.7751 \\ p_u &= 0.8 + z_{1-0.05/2}\sqrt{\frac{0.8 \times (1-0.8)}{1000}} = 0.8248 \end{aligned}$$

ดังนั้นขอบเขตของช่วงความเชื่อมั่นของควอนไทล์ที่ 80% คือ ข้อมูลลำดับที่ $Y_{[Rp_\ell]}$ และ $Y_{[Rp_u]}$ หรือ $Y_{[1000\times 0.7751]}$ และ $Y_{[1000\times 0.8248]}$ ปัดให้เป็นจำนวนเต็มคือข้อมูลลำดับที่ 775 และ 825

เพื่อให้เข้าใจมากขึ้น จะแสดงตัวอย่างที่เป็นตัวเลข สมมติว่าเอาท์พุทบางส่วนมีดังนี้

ลำดับที่	...	**775**	...	**800**	...	**825**	...
ค่าเอาท์พุท	...	188.96	...	212.03	...	256.79	...

ดังนั้นควอนไทล์ที่ 80% คือ 212.03 และช่วงความเชื่อมั่นของควอนไทล์ที่ 80% ที่นัยสำคัญ 95% คือ [188.96, 256.79]

□

โดยปริยาย ซอฟต์แวร์สำเร็จรูปมักรายงานค่าสรุปของดัชนีชี้วัด แต่ไม่แสดงค่าเอาท์พุทของทุก ๆ รอบทำซ้ำ จึงอาจไม่สามารถใช้สูตรที่กล่าวมาคำนวณความน่าจะเป็นและควอนไทล์ ทั้งนี้ ผู้วิเคราะห์สามารถประมาณความน่าจะเป็นและควอนไทล์ได้จากค่าสรุปทางสถิติ

ความน่าจะเป็นและควอนไทล์จากค่าสรุปทางสถิติ

ผู้ใช้สามารถกำหนดให้ซอฟต์แวร์เขียนค่าดัชนีชี้วัดที่ทุก ๆ รอบทำซ้ำออกมาได้โดยใช้คำสั่งเฉพาะ เช่น ในอารีนามีโมดูล `Read/Write` หากว่าผู้วิเคราะห์มีเพียงค่าเฉลี่ย $\bar{Y}$ และครึ่งช่วงความเชื่อมั่น (`hw`) ให้ย้อนกลับไปคำนวณส่วนเบี่ยงเบนมาตรฐาน S_Y ด้วยสมการที่ (5.9) ก่อน สมมติว่าเป็น `hw` จากซอฟต์แวร์อารีนาและ R เป็นจำนวนรอบทำซ้ำ

$$S_Y = \frac{\mathtt{hw} \times \sqrt{R}}{t_{1-0.05/2,R-1}} \tag{5.13}$$

แล้วจึงประมาณค่าความน่าจะเป็นและควอนไทล์โดยอาศัยการแจกแจงแบบปกติ เมื่อผู้วิเคราะห์กำหนดค่า c สามารถประมาณค่าความน่าจะเป็น $\widehat{p}$ ได้ดังนี้

$$\widehat{p} \approx \Pr\left\{Z \leqslant \frac{c-\bar{Y}}{S_Y}\right\} \tag{5.14}$$

โดยที่ Z เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่ค่าเฉลี่ย 0 และความแปรปรวน 1 ซึ่งความน่าจะเป็น (5.14) สามารถคำนวณได้จากเอกซ์เซลด้วยสูตร `=normsdist(`$(c-\bar{Y})/S_Y$`)` ส่วนค่าประมาณของควอนไทล์ $\widehat{\gamma}$ เมื่อผู้วิเคราะห์กำหนดค่า p ก็มีสูตรใกล้เคียงกัน ดังนี้

$$\widehat{\gamma} \approx \bar{Y} + z_p S_Y \tag{5.15}$$

โดยที่ z_p เป็นควอนไทล์ของการแจกแจงแบบปกติที่ค่าเฉลี่ย 0 และความแปรปรวน 1 ซึ่ง z_p สามารถคำนวณได้จากเอกซ์เซลด้วยสูตร `=normsinv(`p`)` ตัวอย่างที่ 5.9 แสดงการคำนวณค่าทางสถิติอื่น ๆ จากค่าสถิติที่อารีนารายงาน

ตัวอย่างที่ 5.9. ผู้วิเคราะห์ประมวลผลแบบจำลองระบบซ่อมบำรุงด้วย 100 รอบทำซ้ำ แต่ละรอบยาว 2 สัปดาห์ ดัชนีชี้วัดที่สนใจคือเวลารอเพื่อรับการซ่อมแซมเฉลี่ย, ความน่าจะเป็นที่เวลาเวลารอมากกว่า 1 วัน, ควอนไทล์ที่ 75% ของเวลารอคอย, และช่วงของเวลารอที่เป็นไปได้ในวันพรุ่งนี้ ผลสรุปจากอารีนาในหมวด `Queues` รายงาน `Repair.Queue Wait Time` = 0.5756 วัน และ `Half Width` = 0.09 วัน

สมมติให้ Y แทนเวลารอ ก่อนจะคำนวณความน่าจะเป็นหรือควอนไทล์ ใช้สมการที่ (5.13) เพื่อคำนวณ S_Y โดยที่ $t_{1-0.05/2,100-1} = 1.9842$ ดังนี้

$$S_Y = \frac{0.09 \times \sqrt{100}}{1.9842} = 0.4536 \text{ วัน}$$

ใช้สมการที่ (5.14) เพื่อประมาณความน่าจะเป็นที่เวลารอรับการซ่อมมากกว่า 1 วัน คือ $1-\widehat{p}$ โดยที่

$$\widehat{p} = \Pr\left\{Z \leqslant \frac{1-0.5756}{0.4536}\right\} = \Pr\{Z \leqslant 0.9356\} = 0.8253$$

ดังนั้นความน่าจะเป็นที่เวลารอมากกว่า 1 วันคือ $1-0.8253 = 0.1747$ ใช้สมการที่ (5.15) เพื่อประมาณควอนไทล์ โดยที่ $z_{0.75} = 0.6745$ ดังนี้

$$\widehat{\gamma} \approx 0.5756 + (0.6745 \times 0.4536) = 0.8816 \text{ วัน}$$

สำหรับช่วงการทำนายของเวลารอคอย แทนค่าในสมการที่ (5.11) ได้ดังนี้

$$0.5756 \pm 1.9842 \times 0.4536\sqrt{1+\frac{1}{100}} \to 0.5756 \pm 0.9045 = [-0.3289, 1.4801] \text{ วัน}$$

เนื่องจากเวลารอคอยไม่ติดลบ จึงปรับช่วงการทำนายเป็น $[0, 1.48]$ วัน

ในรายงาน ควรแสดงค่าดัชนีชี้วัดดังนี้

- เวลารอเพื่อรับการซ่อมแซม**เฉลี่ย**อยู่ในช่วง 0.58 ± 0.09 วัน

- ความน่าจะเป็นที่เวลารอคอยมากกว่า 1 วันคือ 0.17 หรือ 17% (ปัดให้เหลือจำนวนเลขนัยสำคัญเท่ากับค่าเฉลี่ย)
- เวลารอคอยไม่เกิน 0.88 วันด้วยความน่าจะเป็น 75% หรือมากกว่า 0.88 วันด้วยความน่าจะเป็น 25%
- เวลารอเพื่อรับการซ่อมในวันพรุ่งนี้หรืองานซ่อมอื่น ๆ ในอนาคตจะมีค่าอยู่ในช่วง $[0, 1.48]$ วัน ด้วยความน่าจะเป็น 95%

□

ค่าความแม่นยำของการประมาณค่าดัชนีชี้วัดขึ้นอยู่กับขนาดของข้อมูล ซึ่งคือจำนวนรอบทำซ้ำในการจำลองสถานการณ์

5.3.5 การกำหนดจำนวนรอบทำซ้ำ

ซอฟต์แวร์สำเร็จรูปมักตั้งค่าจำนวนรอบทำซ้ำโดยปริยาย เช่น ในพ็อพทูลตั้งไว้ที่ 100 รอบและในอารีนาตั้งไว้ที่ 1 รอบ แต่จริง ๆ แล้ว ผู้ใช้ควรตั้งค่าจำนวนรอบทำซ้ำเอง เพราะเป็นปัจจัยการทดลองที่สำคัญ ที่ส่งผลต่อค่าความผิดพลาดในการประมาณ เฉกเช่นเดียวกับการทดลองทางกายภาพหรือการเก็บข้อมูลจริง ยิ่งจำนวนข้อมูลมาก ความผิดพลาดในการประมาณยิ่งน้อยลง ทั้งนี้ ก็มิใช่ว่าผู้วิเคราะห์จะสามารถประมวลผลมากเท่าใดก็ได้เสมอไป เพราะอาจมีข้อจำกัดด้านเวลา ดังนั้นจึงควรกำหนดจำนวนรอบทำซ้ำที่เหมาะสมสำหรับใช้ประกอบการตัดสินใจ

จำนวนรอบทำซ้ำขึ้นอยู่กับระดับความแม่นยำ (Precision) ที่ต้องการ เช่น $\pm$10,000 บาท, $\pm$ 1 ชั่วโมง หรืออาจเป็นเปอร์เซ็นสัดส่วนของค่าเฉลี่ยก็ได้ เช่น ภายใน 5% ของค่าเฉลี่ยต้นทุน สมมติให้ระดับความแม่นยำที่ต้องการเป็น ε จำนวนรอบทำซ้ำที่ควรใช้คือ R ที่ทำให้ครึ่งช่วงความเชื่อมั่น (สมการที่ 5.9) ใกล้เคียงกับหรือน้อยกว่า ε การประมวลผลแบบจำลองจึงทำเป็น 2 ช่วง ดังนี้

- **ช่วงทดสอบ** (Trial run) เนื่องจากไม่ทราบส่วนเบี่ยงเบนมาตรฐานในสมการที่ (5.9) ผู้วิเคราะห์ต้องประมวลผลเบื้องต้นด้วย R_0 รอบทำซ้ำเพื่อประมาณความแปรปรวนของเอาท์พุท เมื่อได้ครึ่งช่วงความเชื่อมั่น ให้ใช้สมการที่ (5.13) เพื่อคำนวณส่วนเบี่ยงเบนมาตรฐานในช่วงทดสอบ (S_0) จากนั้นจึงคำนวณขอบเขตล่างของจำนวนรอบทำซ้ำที่ต้องการ ดังนี้

$$R \geqslant \left(\frac{z_{1-\alpha/2} S_0}{\varepsilon}\right)^2 \tag{5.16}$$

  เมื่อ $z_{1-\alpha/2}$ เป็นควอนไทล์ของการแจกแจงแบบปกติที่ค่าเฉลี่ย 0 และส่วนเบี่ยงเบนมาตรฐาน 1 เริ่มจากขอบเขตล่างของ R ที่ได้จากสมการที่ (5.16) เพิ่ม R จนกว่าจะได้จำนวนจริงบวกที่น้อยที่สุดที่ทำให้

สมการต่อไปนี้เป็นจริง

$$R \geqslant \left(\frac{t_{1-\alpha/2,R-1}S_0}{\varepsilon}\right)^2 \tag{5.17}$$

สังเกตว่าค่าควอนไทล์ $t_{1-\alpha/2,R-1}$ เป็นฟังก์ชันของ R เมื่อ R เพิ่มขึ้น ค่า $t_{1-\alpha/2,R-1}$ ลดลง

หากไม่ต้องการคำนวณ S_0 อาจใช้ครึ่งช่วงความเชื่อมั่นจากช่วงทดสอบ (h_0) โดยตรงจากสูตรประมาณ (Kelton et al., 2010) ดังนี้

$$R \cong \frac{h_0^2}{\varepsilon^2} R_0 \tag{5.18}$$

สูตรนี้สมมติว่าค่าสถิติ t ที่จำนวนรอบทำซ้ำต่าง ๆ มีค่าใกล้เคียงกัน: $t_{1-\alpha/2,R_0-1} \approx t_{1-\alpha/2,R-1}$ และ S_0 มีค่าใกล้เคียงกับส่วนเบี่ยงเบนมาตรฐานใหม่ หลังจากที่ประมวลผลเพิ่ม

- **ช่วงประมวลผลเพิ่มเติม** ประมวลผลแบบจำลองเพิ่ม $R - R_0$ รอบ หากใช้เวลาไม่นานในการประมวลผล อาจประมวลผลใหม่ทั้งหมด R รอบ แล้วนำเอาท์พุทที่ได้มาสร้างช่วงความเชื่อมั่นด้วยสมการที่ (5.7)-(5.10)

ตัวอย่างที่ 5.10 แสดงการคำนวณจำนวนรอบทำซ้ำเพื่อให้ได้ความแม่นยำที่ต้องการ

ตัวอย่างที่ 5.10. ใช้แบบจำลองศูนย์ให้บริการทางโทรศัพท์เพื่อประมาณค่าสัดส่วนเวลาที่ทำงานจริง ρ ภายในช่วงเวลา 2 ชั่วโมงแรกของวันทำงาน ช่วงทดสอบประมวลผลด้วย $R_0 = 4$ รอบ ได้ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของ ρ คือ $S_0 = 0.072$ ผู้บริหารระบุว่าต้องการทราบ ρ ให้แม่นยำภายใน ± 0.04 สำหรับช่วงความเชื่อมั่นที่นัยสำคัญ 95% ดังนั้น $\varepsilon = 0.04$ สมการที่ (5.16) ให้ขอบเขตล่างของจำนวนรอบทำซ้ำที่ควรมี

$$\left(\frac{z_{1-\alpha/2}S_0}{\varepsilon}\right)^2 = \left(\frac{z_{1-0.05/2} \times 0.072}{0.04}\right)^2 = 12.14$$

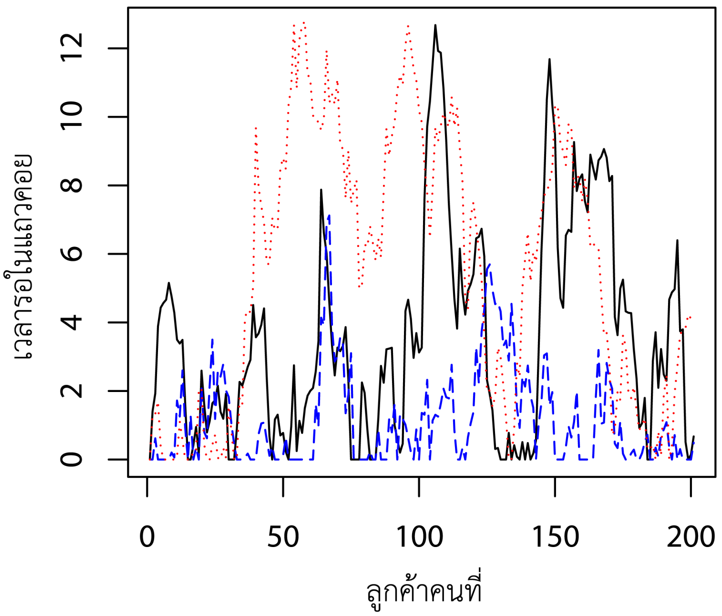
ดังนั้นจึงเริ่มตรวจสอบอสมการที่ (5.17) ด้วย $R = 13$ และเพิ่ม R จนกระทั่งอสมการนี้เป็นจริงค่าแรก พบว่าได้ $R = 15$ ดังนั้นจึงต้องประมวลผลเพิ่มอีก $R - R_0$ คือ 11 รอบ หรืออาจเริ่มใหม่ทั้ง 15 รอบก็ได้ เมื่อได้ช่วงความเชื่อมั่นแล้ว ควรตรวจสอบอีกครั้งว่าครึ่งช่วงความเชื่อมั่นมีค่าน้อยกว่าเป้าหมายที่ต้องการ ε หรือไม่ ถ้ายังมากกว่า ε ต้องเพิ่มจำนวนรอบทำซ้ำอีก

□

ในเชิงปฏิบัติ ผู้อ่านสามารถใช้เอกซ์เซลเทมเพลทชื่อ `SimulationTools.xls` เพื่อคำนวณช่วงความเชื่อมั่น, ช่วงการทำนาย, ควอนไทล์, และจำนวนรอบทำซ้ำที่ควรใช้ ดาวน์โหลดได้จาก `http://www.bcnn.net/` ส่วนที่เหลือของบทนี้จะกล่าวถึงการวิเคราะห์ผลสำหรับการจำลองสถานการณ์แบบไม่สิ้นสุด

5.4 การวิเคราะห์ผลสำหรับการจำลองสถานการณ์แบบไม่สิ้นสุด

ในหัวข้อนี้ วัตถุประสงค์ของการใช้แบบจำลองคือเพื่อประมาณสมรรถนะของระบบในช่วงการวางแผนระยะยาว หลังจากที่ผลกระทบจากสภาวะเริ่มต้น (Initial condition) สลายไป รูปที่ 5.4 แสดงเวลารอคอยในแถวของลูกค้าแต่ละคนจาก 3 รอบทำซ้ำของระบบแถวคอยอย่างง่ายที่มี 1 แถวและผู้ให้บริการ 1 คน โดยที่ ณ เวลาศูนย์ในแบบจำลองระบบไม่มีลูกค้าและผู้ให้บริการว่างอยู่ เห็นชัดเจนว่าในช่วงแรก ๆ เวลารอคอยน้อยและเพิ่มขึ้นเมื่อระบบทำการไปได้สักระยะหนึ่ง ถึงแม้ว่าระบบจะเข้าสู่สภาวะคงตัว เวลารอคอยก็ยังเป็นค่าสุ่มเพราะระบบนี้มีความไม่แน่นอนในการมาถึงของลูกค้าและเวลาให้บริการ ตัวอย่างง่าย ๆ นี้ช่วยชี้ให้เห็นว่าหากแบบจำลองเป็นแบบสถานการณ์ไม่สิ้นสุด อคติที่เกิดจากสภาวะเริ่มต้น (Initial-condition bias) อาจเป็นประเด็นสำคัญ ซึ่งสามารถทำให้เหลือน้อยที่สุดได้หลายวิธี ดังจะกล่าวถึงในหัวข้อที่ 5.4.1–5.4.3



รูปที่ 5.4: เวลารอคอยของลูกค้าสำหรับ 3 รอบทำซ้ำของระบบแถวคอยแบบ M/M/1 ที่มีอัตราลูกค้ามาถึง = 1 และอัตราการให้บริการ = 1.111

วิธีแรกเพื่อลดอคติที่เกิดจากสภาวะเริ่มต้นคือวิธีตั้งค่าสภาวะเริ่มต้นของแบบจำลองอย่างฉลาด

5.4.1 วิธีตั้งค่าสภาวะเริ่มต้นของแบบจำลองอย่างฉลาด

เป้าหมายของวิธีตั้งค่าสภาวะเริ่มต้นของแบบจำลองอย่างฉลาด (Intelligent initialization) คือให้สถานะของระบบเมื่อแบบจำลองเริ่มประมวลผลใกล้เคียงกับสภาวะคงตัวของระบบจริง ตัวอย่างของตัวแปรระบบ เช่น ระบบสินค้าคงคลังมีระดับสินค้าคงคลังหรือจำนวนสินค้าที่สั่งซื้อแล้วแต่ยังไม่มาถึง ส่วนในระบบแถวคอยมีจำนวนลูกค้าในแถวคอยและจำนวนผู้ให้บริการที่ว่างอยู่หรือกำลังทำงานอยู่ การตั้งค่าตัวแปรระบบที่เวลาเริ่มต้นทำได้หลายวิธี เช่น

- หากมีข้อมูลจากระบบที่มีอยู่จริง ใช้ข้อมูลเหล่านั้นมาตั้งค่าตัวแปรระบบ
- หากไม่มีระบบนี้อยู่จริง อาจใช้แบบจำลองอีกชนิดหนึ่ง เช่น ระบบแถวคอย (บทที่ 6) เพื่อกำหนดค่าตัวแปร

เนื่องจากวิธีนี้ขึ้นอยู่กับระบบที่ศึกษา ผู้เขียนจึงไม่อธิบายเพิ่มเติม แต่จะกล่าวถึงวิธีที่สามารถใช้ได้ทั่วไปในหัวข้อต่อไป

5.4.2 วิธีตัดข้อมูลช่วงแรกและประมวลผลหลายรอบ

หากช่วงแรกที่แบบจำลองประมวลผลเป็นระยะอุ่นเครื่อง (Warmup period) สมมติเรียกว่าเป็น $[0, T_0]$ วิธีตัดข้อมูลช่วงแรกและประมวลผลหลายรอบ (Replication-deletion approach) ขจัดอคติในสภาวะเริ่มต้นโดยไม่นำสถานะของระบบในช่วงเวลาอุ่นเครื่องมาคำนวณดัชนีชี้วัด หรืออีกนัยหนึ่งคือเริ่มเก็บค่าสถิติตั้งแต่เวลา T_0 จนถึง T_E ที่แบบจำลองสิ้นสุด แล้วนำเอาท์พุทมาวิเคราะห์เหมือนเอาท์พุทจากแบบจำลองที่เวลาสิ้นสุด (หัวข้อที่ 5.3)

คำถามที่ตามมาคือควรกำหนด T_0 เป็นเท่าใด หาก T_0 น้อยเกินไป ระบบอาจยังอุ่นเครื่องไม่เสร็จ แต่หาก T_0 สูงเกินไป ก็จะทิ้งผลของการประมวลผลมากเกินจำเป็นและทำให้ต้องประมวลผลเพิ่มหลัง T_0 นาน วิธีกำหนดค่า T_0 ที่เป็นที่นิยมคือพิจารณากราฟ เช่น รูปที่ 5.4 ค่าที่นำมาสร้างกราฟมีหลายรูปแบบ เรียงลำดับตามความเหมาะสม จากพอใช้ได้จนถึงเหมาะที่สุด ดังนี้

1. **ค่าที่สังเกตได้จากแต่ละรอบทำซ้ำ** ดังเช่นในรูปที่ 5.4 เป็นกราฟเวลารอคอยจนกว่าจะได้รับบริการของลูกค้าคนที่ j ในรอบทำซ้ำที่ r เมื่อ $1 \leqslant r \leqslant 3$ กำหนดด้วยสัญลักษณ์ Y_{rj} จึงมีอนุกรมเวลา (Time series) 3 ชุด คือ $\{Y_{r1}, Y_{r2}, \ldots, Y_{r200}\}$ สำหรับแต่ละรอบทำซ้ำ r ในรูปที่ 5.4 จึงมีเส้น 3 เส้น สำหรับแต่ละเส้นของรอบทำซ้ำ r นี้ ผู้วิเคราะห์ประมาณ T_{0_r} ด้วยสายตา คือ หลังจาก T_{0_r} ระบบพอเข้าสู่สภาวะสุ่มที่คงตัว ค่า T_0 ที่นำมาใช้เป็นค่า T_{0_r} ที่มากที่สุด นั่นคือ

$$T_0 = \max_r T_{0_r} \tag{5.19}$$

2. **ค่าเฉลี่ยสะสมของแต่ละรอบทำซ้ำ** เป็นค่าที่ซอฟต์แวร์สำเร็จรูป เช่น อารีนา นำมาใช้เมื่อสั่งให้แสดงกราฟขณะที่กำลังประมวลผลแบบจำลอง (Real time) ค่าที่นำมาสร้างกราฟคือค่าเฉลี่ยสะสมเท่าที่มีทั้งหมด คือ ตั้งแต่เริ่มประมวลผลจนถึง ณ ขณะนั้น

$$\bar{Y}_{ri} = \frac{1}{i}\sum_{j=1}^{i} Y_{rj}$$

เช่น $\bar{Y}_{r3} = \frac{1}{3}(Y_{r1} + Y_{r2} + Y_{r3})$ เมื่อได้กราฟอนุกรมเวลาของแต่ละรอบทำซ้ำ ให้ประมาณ T_{0_r} ด้วยสายตา คล้ายกับวิธี 1 ด้านบน และใช้ค่าที่มากที่สุดเป็นจุดตัดจริง (สมการที่ 5.19)

ข้อดีของวิธีนี้คือสะดวกสำหรับใช้กับซอฟต์แวร์สำเร็จรูปเพราะสร้างกราฟได้ทันที ไม่ต้องนำเอาท์พุทมาสร้างกราฟข้างนอก (ในซอฟท์แวร์สำเร็จรูปแกนนอนของกราฟเป็นเวลา) แต่ข้อเสียคือเป็นวิธีที่รอบคอบ (Conservative) เพราะพิจารณาค่าเฉลี่ยตั้งแต่เริ่มต้นจนถึงขณะนั้น ๆ จึงทำให้ตัดข้อมูลมากเกินไป

3. **ค่าเฉลี่ยอองซามเบิ้ล** (Ensemble average) นำข้อมูลลำดับเดียวกันมาเฉลี่ยข้ามรอบทำซ้ำ ดังนี้

$$\bar{Y}_{\cdot j} = \frac{1}{R}\sum_{r=1}^{R} Y_{rj} \tag{5.20}$$

เช่น $\bar{Y}_{\cdot 2} = \frac{1}{3}(Y_{12} + Y_{22} + Y_{32})$ รูปที่ 5.5 แสดงค่าเฉลี่ยอองซามเบิ้ลนี้ ผู้วิเคราะห์ประมาณ T_0 ด้วยสายตาจากกราฟ เมื่อเปรียบเทียบรูปที่ 5.4 และรูปที่ 5.5 จะเห็นว่ากำหนด T_0 จากกราฟของค่าเฉลี่ยอองซามเบิ้ลง่ายกว่าจากกราฟของแต่ละรอบทำซ้ำเพราะกราฟเรียบกว่า ดูเหมือนว่าควรกำหนดให้ T_0 ที่ประมาณลูกค้าคนที่ 50 (สมมติด้วยสัญลักษณ์ ω) ซึ่งหากต้องการกำหนดค่า T_0 ด้วยเวลา ก็อาจประเมินจากอัตราที่ลูกค้ามาถึง (λ) ดังนี้

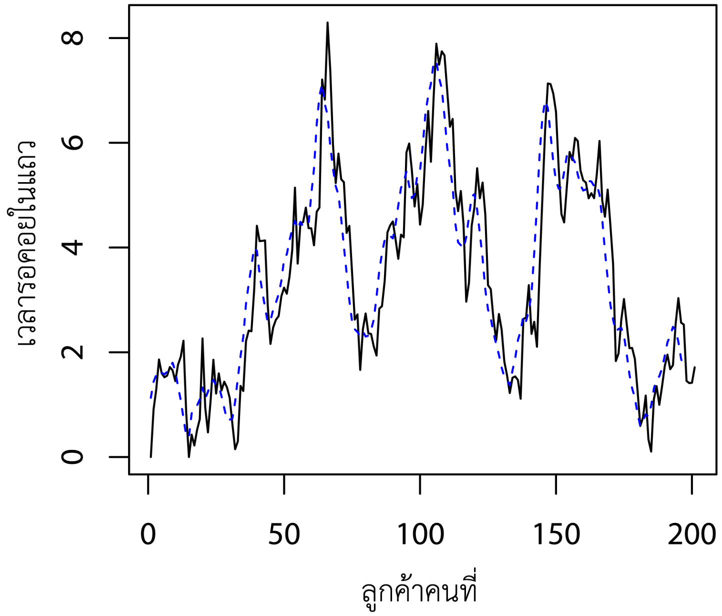
$$T_0 = \omega/\lambda$$

วิธีนี้เป็นวิธีที่แนะนำหากคำนวณค่าเฉลี่ยอองซามเบิ้ลได้ ผู้อ่านอาจใช้ฟังก์ชัน `Mean Plot` ใน `SimulationTools.xls` เพื่อสร้างกราฟค่าเฉลี่ยแบบนี้ได้

4. **ค่าเฉลี่ยอองซามเบิ้ลที่ปรับเรียบ** (Smoothed ensemble average) นำค่าเฉลี่ยอองซามเบิ้ล (สมการที่ 5.20) มาปรับเรียบ อาจด้วยวิธีค่าเฉลี่ยแบบเคลื่อนที่ (Moving average) เพื่อลดความแปรปรวนของค่าที่พล็อตและให้เห็นแนวโน้มของกราฟชัดเจนขึ้น แล้วประมาณค่า T_0 จากกราฟนั้น รูปที่ 5.5 แสดงค่าเฉลี่ยอองซามเบิ้ลที่ปรับเรียบด้วยวิธีค่าเฉลี่ยแบบเคลื่อนที่โดยใช้ 5 พจน์

เมื่อเลือก T_0 แล้ว จำเป็นต้องกำหนด T_E หรือเวลาที่หยุดประมวลผลแบบจำลอง แนวทางที่แนะนำใน Banks et al. (2005) คือ ควรประมวลผลเพิ่ม ยาวเป็นอย่างน้อย 10 เท่าของช่วงอุ่นเครื่อง คือ

$$T_E - T_0 \geqslant 10T_0$$



รูปที่ 5.5: เวลารอคอยเฉลี่ยของ 3 รอบทำซ้ำของข้อมูลในรูปที่ 5.4 (เส้นทึบเป็นค่าเฉลี่ยอองซามเบิ้ลและเส้นประเป็นค่าเฉลี่ยอองซามเบิ้ลที่ปรับเรียบด้วยวิธีค่าเฉลี่ยแบบเคลื่อนที่โดยใช้ 5 พจน์)

หรือ $T_E \geqslant 11T_0$ ถ้าเวลาในการประมวลผลไม่จำกัด จำนวนรอบทำซ้ำ R ขึ้นอยู่กับค่าความแม่นยำที่ต้องการ โดยใช้ขั้นตอนเหมือนกับสำหรับแบบจำลองแบบสิ้นสุด (หัวข้อที่ 5.3.5) แต่หากเวลามีจำกัด ควรตั้งค่า R อย่างน้อย 25 รอบและกำหนด $T_E \approx 11T_0$

หลังจากประมวลผล ค่าสถิติที่สามารถคำนวณจากเอาท์พุทมีดังนี้

- **ค่าเฉลี่ย** ใช้ค่าเฉลี่ยกลุ่มตัวอย่างเป็นค่าประมาณแบบจุดและช่วงความเชื่อมั่นเป็นช่วงประมาณ (หัวข้อที่ 5.3.1)

- **ความน่าจะเป็น** ค่าเอาท์พุทจากแต่ละรอบทำซ้ำเป็นสัดส่วน เช่น หากต้องการประมาณความน่าจะเป็นที่ลูกค้าได้รับการบริการภายใน 3 นาที ค่าเอาท์พุทจากแบบจำลองควรเป็น

$$\widehat{p}_i = \frac{\text{จำนวนลูกค้าที่ได้รับบริการภายใน 3 นาทีในรอบทำซ้ำที่ } i}{\text{จำนวนลูกค้าที่รับบริการทั้งหมดในรอบทำซ้ำที่ } i}$$

  นำเอาท์พุท $\{\widehat{p}_1, \widehat{p}_2, \ldots, \widehat{p}_R\}$ มาคำนวณค่าเฉลี่ยด้วยสมการที่ (5.7) และช่วงความเชื่อมั่นด้วยสมการที่ (5.9)

- **ควอนไทล์** เอาท์พุทจากแต่ละรอบทำซ้ำเป็นค่าประมาณควอนไทล์ภายในรอบทำซ้ำนั้น ๆ เช่น ควอนไทล์ที่ 90% ของเวลารอคอยจนกว่าจะได้รับบริการของรอบทำซ้ำที่ i สังเกตว่าควอนไทล์เหล่านี้ $\{\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots \widehat{\gamma}_R\}$ เป็นอิสระต่อกันและมีการแจกแจงเดียวกัน จึงสามารถนำมาคำนวณค่าเฉลี่ยด้วยสมการที่ (5.7) และช่วงความเชื่อมั่นด้วยสมการที่ (5.9)

วิธีตัดข้อมูลช่วงแรกและประมวลผลหลายรอบเป็นวิธีที่นิยมใช้ในกรณีทั่วไปที่เวลาในการประมวลผล 1 รอบไม่ยาวนัก แต่หากแบบจำลองมีขนาดใหญ่ ใช้เวลาประมวลผลนาน ควรเลือกวิธีประมวลผลเพียงรอบเดียว

5.4.3 วิธีประมวลผลเพียงรอบเดียว

จุดอ่อนของวิธีตัดข้อมูลช่วงแรกและประมวลผลหลายรอบ (หัวข้อที่ 5.4.2) คือ ตัดข้อมูลช่วงอุ่นเครื่องสำหรับทุก ๆ รอบทำซ้ำ หากแบบจำลองมีขนาดใหญ่ ใช้เวลาประมวลผลนาน การทิ้งข้อมูลในช่วงอุ่นเครื่องอาจทำให้สูญเปล่ามาก วิธีประมวลผลเพียงรอบเดียว (Single-replication design) ประมวลผลเพียง 1 รอบทำซ้ำยาว ๆ เพื่อจะตัดช่วงอุ่นเครื่องเพียงครั้งเดียว เนื่องจากการจำลองสถานการณ์แบบไม่สิ้นสุดให้ความสำคัญกับสมรรถนะของระบบในระยะยาวอยู่แล้ว

ความยุ่งยากที่ตามมาคือข้อมูลในรอบทำซ้ำเดียวกันมีความสัมพันธ์กัน ซึ่งส่งผลต่อค่าประมาณความแปรปรวนและสะท้อนต่อไปยังช่วงความเชื่อมั่น ข้อมูลที่มีความสัมพันธ์เชิงบวกมีแนวโน้มกระจุกตัวกัน ทำให้

ช่วงความเชื่อมั่นแคบเกินจริง และดูเหมือนว่าค่าประมาณจากแบบจำลองแม่นยำ ทั้ง ๆ ที่จริง ๆ แล้วอาจไม่เป็นเช่นนั้น ความแปรปรวนของค่าเฉลี่ยจากกลุ่มตัวอย่าง $\{Y_1, Y_2, \ldots, Y_R\}$, $\mathrm{Var}(\bar{Y})$, เกี่ยวข้องกับความแปรปรวนของประชากรดังนี้

$$\mathrm{Var}(\bar{Y}) = \frac{1}{R^2}\sum_{i=1}^{R}\sum_{j=1}^{R}\mathrm{Cov}(Y_i, Y_j) = \frac{1}{R^2}\sum_{i=1}^{R}\mathrm{Cov}(Y_i, Y_i) + \frac{1}{R^2}\sum_{i=1}^{R}\sum_{j=1, i\neq j}^{R}\mathrm{Cov}(Y_i, Y_j) \tag{5.21}$$

สมการที่ (2.11) แสดงนิยามของความแปรปรวนร่วม $\mathrm{Cov}(Y_i, Y_j)$ สังเกตว่า

- ถ้า Y_i, เป็นอิสระต่อกัน $\mathrm{Cov}(Y_i, Y_j) = 0$ สำหรับ $i \neq j$
- ถ้ามาจากการแจกแจงเดียวกัน $\mathrm{Cov}(Y_i, Y_i) = \sigma_Y^2$ สำหรับทุก ๆ i

จากคุณสมบัติ i.i.d. นี้ สมการที่ (5.21) กลายเป็น $\mathrm{Var}(\bar{Y}) = \sigma_Y^2/R$ ซึ่งเป็นที่มาของการประมาณ $\mathrm{Var}(\bar{Y})$ ด้วย S_Y^2/R เมื่อ S_Y^2 เป็นความแปรปรวนของกลุ่มตัวอย่าง แต่เมื่อข้อมูลไม่เป็นอิสระต่อกัน และอาจไม่ได้มาจากการแจกแจงเดียวกันแล้ว

$$\mathrm{Var}(\bar{Y}) \neq \frac{S_Y^2}{R}$$

ดังนั้นจึงต้องปรับข้อมูลเพิ่มเติม เพื่อให้ความสัมพันธ์เหลือน้อยที่สุด หนึ่งในวิธีการที่ใช้คือ**การใช้ค่าเฉลี่ยกลุ่ม** (Batch mean) แทนค่าเอาท์พุทพื้นฐาน คำนวณดังนี้

$$\underbrace{Y_1, Y_2, \ldots, Y_d}_{\text{ตัดช่วงอุ่นเครื่อง}}, \underbrace{Y_{d+1}, Y_{d+2}, \ldots, Y_{d+m}}_{\bar{Y}_1}, \underbrace{Y_{d+m+1}, \ldots, Y_{d+2m}}_{\bar{Y}_2}, \ldots, \underbrace{Y_{d+(k-1)m+1}, \ldots, Y_{d+km}}_{\bar{Y}_k} \tag{5.22}$$

แนวความคิดคือข้อมูลที่เกิดขึ้นในช่วงเวลาห่างกันมีความสัมพันธ์กันน้อยกว่าข้อมูลที่เกิดขึ้นใกล้ ๆ กัน เช่น $\bar{Y}_1$ ห่างจาก $\bar{Y}_2$ มากกว่า Y_1 กับ Y_2 ดังนั้นหากขนาดของกลุ่ม (m) ใหญ่พอ ค่าเฉลี่ยกลุ่มที่ได้คงเกือบเป็นอิสระต่อกัน ดัชนีทางสถิติที่ใช้ตรวจสอบคือค่าอัตสหสัมพันธ์ (Autocorrelation) Lag เป็นหนึ่ง หรือ $\mathrm{Corr}(Y_i, Y_{i+1})$ (อ่านคำนิยามได้จากสมการที่ 2.15) ซึ่งมักจะมีขนาดสูงกว่าที่ Lag อื่น ๆ ขนาดของกลุ่มควรสูงพอที่ทำให้ค่าอัตสหสัมพันธ์ที่ Lag หนึ่งนี้ใกล้เคียงศูนย์ แต่มีข้อจำกัดตรงที่ขนาดของกลุ่มสัมพันธ์กับจำนวนของค่าเฉลี่ยกลุ่มที่ได้ (k) การศึกษาเชิงประจักษ์ (Schmeiser, 1982) พบว่าจำนวนค่าเฉลี่ยกลุ่มควรอยู่ในช่วง 10 ถึง 30 ค่า ถ้า k ต่ำกว่า 10 จะมีจำนวนค่าเฉลี่ยกลุ่มน้อยเกินไป แต่หากมีมากกว่า 30 ก็แทบไม่เพิ่มประโยชน์ใด ๆ

นอกจากค่าเฉลี่ยกลุ่มแล้ว ยังสามารถคำนวณค่าสถิติอื่น ๆ เช่น ความน่าจะเป็นหรือควอนไทล์ภายในกลุ่มได้ และนำค่าเหล่านี้มาวิเคราะห์เหมือนค่าเฉลี่ยกลุ่ม คือ คำนวณค่าเฉลี่ย (สมการที่ 5.7) และช่วงความเชื่อมั่น (สมการที่ 5.9) สำหรับข้อมูลที่เกี่ยวเนื่องกับเวลา เช่น จำนวนลูกค้าในระบบ หรือจำนวนพนักงานที่ทำงานอยู่ การหาค่าเฉลี่ยกลุ่มคำนวณเป็นค่าเฉลี่ยเวลา คล้ายกับสมการที่ (5.2)

หัวข้อสุดท้ายกล่าวถึงประเด็นที่ควรพิจารณาหากใช้แบบจำลองเดียวกันเพื่อประมาณค่าดัชนีชี้วัดหลายตัว

5.5 การประมาณค่าดัชนีชี้วัดหลายตัวของระบบเดียว

หากการวิเคราะห์พิจารณาระบบเดียวแต่ใช้ประมาณค่าดัชนีชี้วัดหลายตัว สมมติว่าเป็น $\theta_i, i = 1, 2, \ldots, K$ เช่น θ_1 เป็นค่าคาดหมายเวลาที่ลูกค้าอยู่ในระบบ, θ_2 เป็นค่าคาดหมายของสัดส่วนของเวลาที่พนักงานทำงานจริง, θ_3 เป็นค่าคาดหมายของสัดส่วนของลูกค้าที่รอเพื่อรับบริการนานกว่าเป้าหมาย ค่าความผิดพลาดในการประมาณค่าดัชนีชี้วัดเหล่านี้ก็ยังระบุด้วย**ช่วงความเชื่อมั่นสำหรับแต่ละดัชนีชี้วัด**เช่นเดิม แต่เนื่องจากช่วงความเชื่อมั่นเหล่านี้ถูกพิจารณาพร้อมกัน **ค่านัยสำคัญรวม** (Overall confidence level) จึงน้อยกว่าค่านัยสำคัญของแต่ละอัน ช่วงความเชื่อมั่นที่มีนัยสำคัญ $1 - \alpha$ มีนัยว่าโอกาสที่จะผิดพลาด (คือ ไม่ครอบคลุมค่าเฉลี่ยประชากรที่แท้จริง) ไม่เกิน α ถ้ามีช่วงความเชื่อมั่น C อัน แต่ละอันมีนัยสำคัญ $1 - \alpha_i$ จะได้ว่า

$$\text{ความน่าจะเป็นของความผิดพลาดชนิดใด ๆ} \leqslant \sum_{i=1}^{C} \alpha_i$$

อสมการของบอนเฟอร์โรนี่ (Bonferroni inequality) ระบุว่า ความน่าจะเป็นที่ทุกช่วงความเชื่อมั่นถูกต้องคือส่วนที่เหลือของความน่าจะเป็นที่ผิดพลาดทางใดทางหนึ่ง ดังนั้นจึงได้ว่าความน่าจะเป็นที่ทุก ๆ ช่วงความเชื่อมั่นไม่พลาดคือ

$$\text{ความน่าจะเป็นที่ค่าเฉลี่ยประชากร } \mu_i \text{ อยู่ในช่วงความเชื่อมั่น สำหรับทุก ๆ ช่วง } i \geqslant 1 - \sum_{i=1}^{C} \alpha_i \tag{5.23}$$

หากผู้วิเคราะห์ใช้ผลการทดลอง 1 ชุดเพื่อประมาณค่าดัชนีชี้วัด C ตัว ให้ใช้สมการที่ (5.23) เพื่อกำหนดขอบเขตล่างของนัยสำคัญรวม

หากทุกช่วงความเชื่อมั่น**เป็นอิสระต่อกัน** โดยประมวลผลทีละชุดสำหรับแต่ละดัชนีชี้วัด นัยสำคัญรวมของทุกช่วงความเชื่อมั่นเมื่อพิจารณาพร้อมกัน C ช่วง หรือ

$$\text{ความน่าจะเป็นที่ค่าเฉลี่ยประชากร } \mu_i \text{ อยู่ในช่วงความเชื่อมั่น สำหรับทุก ๆ ช่วง } i = (1-\alpha_1)(1-\alpha_2)\cdots(1-\alpha_C) \tag{5.24}$$

สังเกตว่าด้านขวาของสมการที่ (5.24) มากกว่าด้านขวาของสมการที่ (5.23)

ตัวอย่างที่ 5.11 แสดงการนำสมการที่ (5.24) ไปใช้

ตัวอย่างที่ 5.11. แบบจำลองระบบให้บริการลูกค้าทางโทรศัพท์ถูกใช้เพื่อประมาณดัชนีชี้วัดหลายตัวของระบบปัจจุบันระบบเดียว เช่น ค่าคาดหมายเวลาที่ลูกค้าอยู่ในระบบ หรือค่าคาดหมายของสัดส่วนของเวลาที่พนักงานทำงานจริง หากพิจารณาดัชนีชี้วัด 4 ตัว แต่ละตัวประมาณจากผลการทดลอง 1 ชุด จึงมีการทดลอง 4 อันที่เป็นอิสระต่อกัน ประมาณแต่ละดัชนีชี้วัดที่นัยสำคัญ 95% ค่านัยสำคัญรวมคือ $0.95^4 = 0.81$ ซึ่งน้อย

กว่า 0.95 ที่เป็นนัยสำคัญของแต่ละช่วงความเชื่อมั่น แต่หากใช้การทดลอง 1 ชุดเพื่อประมาณค่าดัชนีชี้วัด 4 ตัว จากสมการที่ (5.23) นัยสำคัญมีค่ารวมอย่างน้อย $1 - (0.05 \times 4) = 0.80$

□

จากตัวอย่างที่ 5.11 สังเกตว่าเมื่อทำการทดลองเพียงชุดเดียวเพื่อประมาณดัชนีชี้วัดหลายตัว นัยสำคัญรวมลดลงแบบเชิงเส้นกับจำนวนดัชนีชี้วัดที่พิจารณา หากมีดัชนีชี้วัด 10 ตัวที่ประมาณด้วยนัยสำคัญ 95% นัยสำคัญรวมเหลือเพียง $1 - (10 \times 0.05) = 0.5$ ซึ่งน้อยเกินไปสำหรับการนำไปวิเคราะห์จริง ในทางกลับกัน หากต้องการให้นัยสำคัญรวม 95% แต่ละช่วงความเชื่อมั่นจะต้องมีนัยสำคัญ $1-0.05/10 = 1-0.005 = 99.5\%$ ค่านัยสำคัญที่เพิ่มขึ้นส่งผลต่อความกว้างของช่วงความเชื่อมั่น ในสูตร (5.9) มีพจน์ $t_{1-\alpha/2,\nu}$ เช่น ที่ $\alpha = 0.05$ ค่าสถิติ $t_{1-0.05/2,100} = 1.984$ เปรียบเทียบกับค่าสถิติ $t_{1-0.005/2,100} = 2.871$ ถ้าพจน์อื่น ๆ ในสูตร (5.9) มีค่าเกือบคงที่ ค่าครึ่งช่วงความเชื่อมั่นเมื่อนัยสำคัญ 99.5% กว้างกว่าที่นัยสำคัญ 95% ถึง 1.45 เท่า (เมื่อองศาอิสระเป็น 100)

ด้วยเหตุนี้จึงไม่ควรพิจารณาช่วงความเชื่อมั่นของดัชนีชี้วัดหลาย ๆ ตัวพร้อมกัน อาจมีดัชนีชี้วัดหลัก 2-3 ตัว แล้วให้ที่เหลือเป็นดัชนีชี้วัดรองที่พิจารณาภายหลัง นอกจากนี้ควรใช้การทดลอง 1 ชุดแยกสำหรับแต่ละดัชนีชี้วัด และการทดลองเหล่านี้เป็นอิสระต่อกัน เพราะนัยสำคัญที่ได้สูงกว่า เช่น หากพิจารณาดัชนีชี้วัด 10 ตัวพร้อมกัน นัยสำคัญที่ได้คือ $0.95^{10} = 0.6$ ซึ่งมากกว่ากรณีที่ใช้การทดลองเพียง 1 อันเพื่อประมาณดัชนีชี้วัดทุกตัว เพราะเหลือนัยสำคัญเพียง 0.5 คอมพิวเตอร์ในปัจจุบันประมวลผลเร็วมาก เอื้อให้ประมวลผลได้หลาย ๆ รอบ จึงควรใช้การทดลองที่เป็นอิสระต่อกันสำหรับแต่ละดัชนีชี้วัด การสรุปที่รัดกุมเพราะค่านัยสำคัญที่เพิ่มขึ้นช่วยให้ใช้แบบจำลองที่พัฒนาถูกใช้อย่างคุ้มค่ากว่า

ผู้วิเคราะห์มักใส่ใจการพัฒนาแบบจำลองคอมพิวเตอร์แต่ละเลยการออกแบบการทดลองและการวิเคราะห์ผล จึงทำให้ผลการวิเคราะห์ที่ได้ผิดพลาดหรือทำให้เข้าใจผิด (Misleading) หัวข้อต่อไปกล่าวถึงความผิดพลาดที่ผู้เขียนพบเป็นประจำ

5.6 ความผิดพลาดที่พบบ่อย

ตัวอย่างความผิดพลาดที่ผู้เขียนพบบ่อย เช่น

- รายงานเพียงแต่ค่าเฉลี่ย ไม่ระบุค่าคลาดเคลื่อนมาตรฐานหรือครึ่งช่วงความเชื่อมั่น (หัวข้อที่ 5.3)
- รายงานค่าเชิงตัวเลขด้วยจำนวนจุดทศนิยมมากเกินความแม่นยำของข้อมูลนำเข้า โปรแกรมคอมพิวเตอร์ให้ค่าใดก็คัดลอกมาวางในรายงาน โดยไม่คำนึงถึงจำนวนหลักของเลขนัยสำคัญ (Significant digits หรือ Significant figures) ของข้อมูลนำเข้า (หัวข้อที่ 5.3)
- พิจารณาเพียงแค่ค่าเฉลี่ยของดัชนีชี้วัด แต่ไม่คำนึงถึงเกณฑ์ทางสถิติที่บ่งบอกระดับความเสี่ยง เช่น

ความน่าจะเป็น (หัวข้อที่ 5.3.3) หรือควอนไทล์ (หัวข้อที่ 5.3.4)

- เมื่อพิจารณาถึงเหตุการณ์ที่จะเกิดขึ้นในอนาคต ใช้ช่วงความเชื่อมั่นแทนช่วงการทำนาย (หัวข้อที่ 5.3.2)
- ประมวลผลด้วยจำนวนรอบทำซ้ำที่กำหนดโดยไม่มีกฎเกณฑ์ (Arbitrary) ตามแต่จริตผู้วิเคราะห์ แต่ไม่ได้กำหนดจำนวนรอบทำซ้ำที่ควรทำเพื่อให้ได้ความแม่นยำที่ต้องการ (หัวข้อที่ 5.3.5)
- ในรายงานไม่ระบุเงื่อนไขที่ใช้ในการประมวลผล เช่น จำนวนรอบทำซ้ำที่ใช้ประมวลผล สภาวะเริ่มต้น ความยาวของรอบทำซ้ำ (Run length)
- หากเป็นการจำลองแบบสถานการณ์ไม่สิ้นสุด ผู้วิเคราะห์ไม่พิจารณาช่วงอุ่นเครื่อง (หัวข้อที่ 5.4)
- ในกรณีที่มีดัชนีชี้วัดหลายตัว ไม่พิจารณานัยสำคัญรวม (Overall confidence level หัวข้อที่ 5.5)

หัวข้อ 1.6 กล่าวถึงสถานการณ์ที่อาจไม่เหมาะที่จะใช้การจำลองสถานการณ์เป็นเครื่องมือวิเคราะห์

แบบฝึกหัด

5.1. ผู้วิเคราะห์ประมวลผลแบบจำลองแผนกผู้ป่วยนอกของโรงพยาบาลของรัฐแห่งหนึ่ง 10 รอบทำซ้ำ เอาท์พุทเป็นเวลารอของผู้ป่วยจนกว่าจะพบแพทย์เฉลี่ย (หน่วย: นาที) แทนด้วยสัญลักษณ์ W_q ดังนี้

$$45.27 \;\; 43.82 \;\; 46.06 \;\; 48.73 \;\; 51.36 \;\; 46.58 \;\; 50.48 \;\; 47.48 \;\; 46.64 \;\; 45.42$$

ประมาณค่าสถิติต่อไปนี้และอธิบายค่าที่คำนวณได้

(i) ช่วงความเชื่อมั่นสำหรับ W_q ที่นัยสำคัญ 95%

(ii) ค่าควอนไทล์ที่ 90%

(iii) ความน่าจะเป็นที่ $W_q \geqslant 50$ นาที

(iv) ช่วงของค่าเวลารอของผู้ป่วยคนถัดไป ด้วยความน่าจะเป็น 95%

(v) คำนวณจำนวนรอบทำซ้ำที่ต้องการเพื่อประมาณค่าเวลารอให้แม่นยำภายใน 1 นาที

5.2. พิจารณาข้อมูลดังต่อไปนี้ ซึ่งเป็นค่าที่สังเกตได้ของตัวแปรสุ่ม X

88.5	89.9	93.4	90.1	91.1	91.2	84.3	89.2
94.7	98.8	96.1	89.3	90.5	86.7	93.2	92.3
84.3	88.3	89.6	91.1	100.3	94.2	88.6	88.9
90.1	90.4	90.4	92.2	87.6	90.8	88.7	89.8
89	91.2	91.6	83.4	92.7	90.1	92.7	92.7
89.8	90.6	90.7	91	87.9	91.8	89.3	93.3
91.6	92.2	88.6	88.2	93	88.4	91	86.7
90.3	87.7	88.3	88.5	94.4	92.6	87.5	91
90	91.1	94.2	93.3	90.4	93.7	87.8	90.9
91.5	86.7	85.3	87.4	91.2	96.5	88.3	89.9
91.8	89.7	92.2					

ประมาณค่าสถิติต่อไปนี้

(i) ความน่าจะเป็นที่ $X \geqslant 90$ และช่วงความเชื่อมั่นของความน่าจะเป็นนี้ที่นัยสำคัญ 95%

(ii) ค่า β ที่ทำให้สมการต่อไปนี้เป็นจริง

$$\Pr\{X \leqslant \beta\} = 0.8$$

และช่วงความเชื่อมั่นของ β ด้วยความน่าจะเป็น 90%

5.3. ซอฟต์แวร์จำลองสถานการณ์ประมาณค่าเฉลี่ยเวลาในระบบของงานซ่อม Y ได้ 118.86 นาที และครึ่งช่วงความเชื่อมั่นที่นัยสำคัญ 95% เท่ากับ 2.1816 นาที โดยใช้ 10 รอบทำซ้ำ ให้ประมาณค่าสถิติต่อไปนี้

(i) ความน่าจะเป็นที่เวลาในระบบของงานซ่อมเกินเป้าหมายคือ 2 ชั่วโมง

(ii) ค่า β ที่ทำให้สมการต่อไปนี้เป็นจริง

$$\Pr\{Y \geqslant \beta\} = 0.25$$

5.4. ถ้าหากเราใช้แบบจำลองระบบผู้ป่วยนอกเพื่อประมานดัชนีชี้วัด 4 ตัวดังนี้ เวลาที่คนไข้รอคอยทั้งหมดในระบบ, เวลาที่คนไข้อยู่ในระบบ, เวลาที่คนไข้สุดท้ายออกจากระบบ, และสัดส่วนของเวลาที่พยาบาลทำงานจริง โดยคำนวณช่วงความเชื่อมั่นจากผลการประมวลผลชุดเดียวกันที่มีขนาดตัวอย่าง 100 รอบทำซ้ำ ต้องการให้นัยสำคัญรวมเท่ากับ 95 % ช่วงความเชื่อมั่นของแต่ละดัชนีชี้วัดจะต้องมีนัยสำคัญเท่าใดและให้ระบุค่าสถิติ t

5.5. ต่อเนื่องจากแบบฝึกหัดข้อที่ 5.4. หากมีช่วงความเชื่อมั่นของดัชนีชี้วัด 4 ตัวนี้แล้ว โดยคำนวณจากข้อมูลที่เป็นอิสระต่อกัน แต่ละช่วงความเชื่อมั่นมีนัยสำคัญ 95% นัยสำคัญรวมจะเป็นเท่าใด

5.6. ต่อเนื่องจากแบบฝึกหัดข้อที่ 5.4. หากมีช่วงความเชื่อมั่นของดัชนีชี้วัด 4 ตัวนี้แล้ว โดยคำนวณจากข้อมูลที่มาจากชุดเดียวกัน แต่ละช่วงความเชื่อมั่นมีนัยสำคัญ 95% นัยสำคัญรวมจะเป็นเท่าใด

6

แบบจำลองระบบแถวคอย

ระบบแถวคอย (Queueing models หรือ Waiting line models) ที่เห็นได้ทั่วไป เช่น ที่เคาน์เตอร์ชำระเงินที่ซุปเปอร์มาร์เก็ต รถรอจ่ายเงินที่ด่านเก็บเงินทางด่วน คนไข้รอเข้าตรวจที่โรงพยาบาล เมื่อนำแถวคอยมาเชื่อมต่อกันได้เครือข่ายของแถวคอย (Network of queues) ตัวอย่างในอุตสาหกรรมมีสายการประกอบผลิตภัณฑ์ต่าง ๆ เช่น รถยนต์ เครื่องใช้ไฟฟ้า หรือการให้บริการผู้ป่วยนอกของโรงพยาบาล เพราะประกอบด้วยแถวคอยย่อย ๆ เช่น ตรวจสัญญาณชีพ เอ็กซเรย์ พบแพทย์ และจ่ายยา ระบบแถวคอยมีสิ่งที่ต้องการรับบริการ เรียกรวม ๆ ว่า **ลูกค้า** (Customer) ซึ่งอาจเป็นสิ่งไม่มีชีวิตหรือไม่มีรูปลักษณ์ทางกายภาพก็ได้ เช่น คำสั่งซื้อของลูกค้าทางอีเมล์ การบริการทำโดย**ผู้ให้บริการ** (Server) อาจเห็นแถวคอยด้วยตาหรืออาจไม่เห็นก็ได้ เช่น จัดลูกค้าด้วยบัตรคิว หรือให้ลงรายชื่อ (Waiting list) โดยทั่วไปแบบจำลองแถวคอยใช้เพื่อแสดงพฤติกรรมของระบบแถวคอย, กำหนดระดับการให้บริการที่ระบบทำได้, และใช้เพื่อเปรียบเทียบระบบแถวคอยหลาย ๆ ตัวเลือก

แบบจำลองแถวคอยใช้เพื่อเปรียบเทียบกับผลที่ได้จากแบบจำลองสถานการณ์ โดยแบบจำลองแถวคอยที่จะกล่าวถึงในบทนี้เป็นแบบอย่างง่าย ที่สมมติว่าเวลาที่เป็นค่าสุ่มมีการแจกแจงแบบเอ็กซ์โพเนนเชียล มีลูกค้าเพียงชนิดเดียว มีความอดทนรอได้ไม่จำกัด คือ ไม่ออกจากระบบเมื่อเข้าสู่แถวแล้ว มีแถวคอยเพียงแถวเดียว แต่อาจมีผู้ให้บริการหลายคน ซึ่งทำงานได้เหมือนกันทุกคน คือ เวลาให้บริการมีการแจกแจงเดียวกัน ผู้ให้บริการทำงานได้ตลอดเวลา ไม่พิจารณาเวลาพักหรือเวลาเสีย จึงทำให้แบบจำลองจัดการได้ง่ายทางคณิตศาสตร์ (Mathematically tractable) และประมวลผลได้รวดเร็ว

ลำดับการนำเสนอและตัวอย่างส่วนใหญ่ในบทนี้มีพื้นฐานจากเอกสารประกอบการสอนบทที่ 13 หัวข้อเรื่อง Queueing Theory ในหนังสือ Ragsdale (2008) เสริมด้วยตัวอย่างจากหนังสืออื่น ๆ หรือที่ผู้เขียนแต่งขึ้นเอง ตำราเล่มนี้ว่าด้วยการจำลองสถานการณ์จึงไม่เน้นที่แบบจำลองแถวคอย และจะไม่แสดงสูตรคณิตศาสตร์หรือการพิสูจน์สูตรที่เป็นผลลัพธ์ของแบบจำลอง สิ่งเหล่านี้ผู้อ่านที่สนใจสามารถค้นได้จากอินเตอร์เน็ตโดยใช้คำสำคัญ “Queuing models tutorial” หรือหนังสือเรียนการวิจัยดำเนินงาน เช่น Ragsdale (2008) หรืออ่านบทแนะนำการประยุกต์ใช้แบบจำลองแถวคอย Ashley (2002) บทนี้ใช้เครื่องมือคำนวณที่เป็นเอกซ์เซลแอดอิน ชื่อ `QTP.xla` หรือ `Queueing toolpak` ซึ่งพัฒนาโดย ศ.ดร. Armann Ingolfsson และ ดร. Fraser Gallopp จาก University of Alberta ประเทศแคนาดา ดาวน์โหลดได้จาก `http://apps.business.ualberta.ca/aingolfsson/qtp/download.htm`

6.1 ลักษณะของแบบจำลองแถวคอย

หัวข้อนี้อธิบายถึงลักษณะพื้นฐานที่มีในทุก ๆ แถวคอย

6.1.1 กระบวนการมาถึง

กระบวนการมาถึง (Arrival process) แสดงด้วยจำนวนลูกค้าที่มาถึงตั้งแต่เวลา 0 จนถึงเวลา t แสดงด้วยตัวแปรสุ่ม $X(t)$ ซึ่งมักจำลองด้วยการแจกแจงแบบปัวซอง (หัวข้อที่ 4.2.4) ที่มีพารามิเตอร์เพียง 1 ตัว คือ**อัตรา**การมาถึง λ ที่เป็นค่าเฉลี่ย เช่น $\lambda = 5$ ต่อชั่วโมง สังเกตว่าจำนวนลูกค้าที่มาถึงเฉลี่ยภายในช่วงเวลา $[0, t]$ คือ λt ดังนั้นฟังก์ชันมวลความน่าจะเป็นในตารางที่ 4.5 จึงเป็น

$$\Pr\{X(t) = x\} = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \tag{6.1}$$

สังเกตว่าหน่วยของ λ และ t สัมพันธ์กัน คือ ต้องหักล้างกันเมื่อนำมาคูณกัน หรือ λt ไม่มีหน่วย ถ้าจำนวนลูกค้ามาถึงต่อหนึ่งหน่วยเวลามีการแจกแจงแบบปัวซอง เวลาระหว่างการมาถึง (Interarrival time) มีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยพารามิเตอร์เดียวกัน เช่น ถ้าอัตราการมาถึง = 5 ต่อชั่วโมง เวลาระหว่างการมาถึงเฉลี่ย = 1/5 ชั่วโมง หรือ 12 นาที กล่าวคือมีลูกค้ามาถึงทุก ๆ 12 นาทีโดยเฉลี่ย แต่ไม่ใช่แน่นอนทุก ๆ 12 นาที พารามิเตอร์ λ ประมาณได้โดยเฉลี่ยจากจำนวนลูกค้ามาถึงต่อหนึ่งหน่วยเวลา เช่น 1 ชั่วโมง, 30 นาที, หรือ 15 นาที เพราะการเก็บข้อมูลจำนวนครั้งที่มาถึงต่อหนึ่งหน่วยเวลาอาจง่ายกว่าเวลาระหว่างการมาถึง ตัวอย่างที่ 6.1 แสดงการประมาณค่า λ

ตัวอย่างที่ 6.1. ข้อมูลต่อไปนี้เป็นจำนวนครั้งที่ลูกค้าโทรเข้าภายในช่วงเวลา 1 ชั่วโมง ตั้งแต่ 6.00 น. ถึง 18.00 น. ของวันทำงานวันหนึ่ง สมมติให้ $X(t)$ เป็นจำนวนลูกค้ามาถึงใน 1 ชั่วโมง และใช้การแจกแจงแบบปัวซองจำลอง $X(t)$

$$4 \quad 4 \quad 8 \quad 3 \quad 2 \quad 5 \quad 3 \quad 6 \quad 3 \quad 8 \quad 6 \quad 2$$

สังเกตว่าในสมการที่ (6.1) มีพารามิเตอร์ λ ซึ่งสามารถประมาณได้จากค่าเฉลี่ยกลุ่มตัวอย่าง ดังนี้

$$\hat{\lambda} = \frac{1}{12}(4+4+8+3+2+5+3+6+3+8+6+2) = 4.5 \text{ ต่อชั่วโมง}$$

□

กระบวนการสุ่มอีกอันหนึ่งในระบบแถวคอยคือกระบวนการให้บริการ

6.1.2 กระบวนการให้บริการ

กระบวนการให้บริการ (Service process) ว่าด้วยจำนวนผู้ให้บริการและเวลาในการให้บริการ (Service time, T) ซึ่งคือระยะเวลาที่ลูกค้าได้รับการบริการ (ไม่รวมเวลารอคอยในแถว) นิยมจำลอง T ด้วยการแจกแจงแบบเอ็กซ์โพเนนเชียล (หัวข้อที่ 4.2.8) ซึ่งมีความน่าจะเป็นดังนี้

$$\Pr\{t_1 \leqslant T \leqslant t_2\} = \int_{t_1}^{t_2} \mu e^{-\mu t} dt = e^{-\mu t_2} - e^{-\mu t_1}, t_1 \leqslant t_2 \tag{6.2}$$

ซึ่งอาศัยพารามิเตอร์เพียงหนึ่งตัว คือ อัตราการให้บริการ (Service rate, μ) เช่น อัตราที่พนักงานรับโทรศัพท์ให้บริการลูกค้าคือ 7 ต่อชั่วโมง นั่นคือเวลาเฉลี่ยในการให้บริการคือ $1/\mu$ หรือในตัวอย่างพนักงานรับโทรศัพท์คือ 1/7 ชั่วโมง พารามิเตอร์ μ ประมาณได้จากเวลาเฉลี่ยที่ใช้ในการบริการลูกค้า 1 คน ดังแสดงในตัวอย่างที่ 6.2

ตัวอย่างที่ 6.2. ข้อมูลต่อไปนี้เป็นเวลาที่แพทย์ให้คำปรึกษากับคนไข้ 1 คน (หน่วย: นาที)

$$6.72 \;\; 3.27 \;\; 8.25 \;\; 6.26 \;\; 7.08 \;\; 8.62 \;\; 7.97 \;\; 6.91 \;\; 5.57 \;\; 7.02$$

เวลาให้คำปรึกษาเฉลี่ยคือ 6.77 นาที ดังนั้น $\hat{\mu}$ = 1/6.77 ต่อนาที หรือ 60/6.77 = 8.87 ต่อชั่วโมง

นอกเหนือจากกระบวนการไหลเข้าและกระบวนการให้บริการ ระบบแถวคอยยังมีรายละเอียดอื่น ๆ

6.1.3 ลักษณะอื่น ๆ

แถวคอยมีรายละเอียดสำคัญอื่น ๆ เช่น

- **ระเบียบแถวคอย (Queue discipline)** เช่น มาก่อนได้ก่อน (First come first serve, FCFS), มาก่อนได้ทีหลัง (Last come first serve, LCFS) หรือลูกค้ามีลำดับความสำคัญต่างกัน (Priority) บทนี้จะกล่าวถึงเฉพาะแถวคอยแบบ FCFS
- **ความจุของแถวคอย** (Queue capacity) แทนด้วยสัญลักษณ์ c หากมีลูกค้ารอในแถว c คนแล้ว ลูกค้าคนใหม่จะเข้ามาไม่ได้และจะถูกปฏิเสธจากระบบ เช่น ระบบให้บริการทางโทรศัพท์ที่มีคู่สายจำกัด หากมีลูกค้ารออยู่ในระบบมากเกินความจุ ลูกค้าที่พยายามโทรเข้าจะโทรเข้าอีกไม่ได้ และจะได้รับสัญญาณไม่ว่างแทน หากไม่ได้ระบุค่าความจุที่ชัดเจน ความจุของแถวคอยจะสมมติว่ามีค่าสูงมาก จนถือว่าไม่จำกัด (Infinity)
- **จำนวนของลูกค้าที่เป็นไปได้ทั้งหมด** (Size of calling population) เช่น ระบบซ่อมบำรุงเครื่องจักรในโรงงานมีจำนวนลูกค้าที่เป็นไปได้ทั้งหมดคือจำนวนเครื่องจักรที่โรงงานมีทั้งหมด หากไม่ได้ระบุค่าความจุที่ชัดเจน จะสมมติว่าจำนวนของลูกค้าที่เป็นไปได้มีค่าสูงมาก จนถือว่าไม่จำกัด

เพื่อให้การนิยามระบบแถวคอยกระชับและชัดเจน จึงมีการคิดค้นการใช้สัญลักษณ์เพื่อกำหนดลักษณะ

6.1.4 การใช้สัญลักษณ์แบบเคนดอล

Kendall (1953) จึงได้นำเสนอสัญลักษณ์เคนดอล (Kendall notation) ซึ่งแสดงด้วยพารามิเตอร์ 3 ตัวหลัก ดังนี้

a/b/c

พารามิเตอร์ a แสดงกระบวนการมาถึง (หัวข้อที่ 6.1.1) ซึ่งใช้สัญลักษณ์ M สำหรับกระบวนการปัวซอง เพราะเวลาระหว่างการมาถึงเป็นเอ็กซ์โพเนนเชียลหรือไม่ขึ้นกับอดีต (Memoryless) ซึ่งคือมาร์โคเวียน (Markovian) นอกจากนี้ยังใช้ D สำหรับเวลาระหว่างการมาถึงที่เป็นค่าคงที่ (Deterministic) และสัญลักษณ์ G (General) สำหรับเวลาระหว่างการมาถึงแบบทั่วไปที่ระบุเพียงแค่ค่าเวลาเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

พารามิเตอร์ b แสดงการแจกแจงของเวลาการให้บริการ (หัวข้อที่ 6.1.2) ซึ่งอาจเป็น M, D หรือ G เหมือนพารามิเตอร์

พารามิเตอร์ c แสดงจำนวนผู้ให้บริการ

เช่น M/M/3 เป็นระบบแถวคอยที่มีการมาถึงแบบปัวซอง เวลาให้บริการมีการแจกแจงแบบเอ็กซ์โพเนนเชียล และมีผู้ให้บริการ 3 คน

ถึงแม้ว่าระบบแถวคอยอาจแตกต่างกัน ดัชนีชี้วัดหลัก ๆ ที่เป็นที่สนใจเกี่ยวเนื่องกับเวลารอคอย คุณภาพการให้บริการ และความคับคั่งในระบบ

6.2 ดัชนีชี้วัดของระบบแถวคอย

ดัชนีชี้วัดของระบบมีสองด้านซึ่งถ่วงดุลกันอยู่ คือ ด้านลูกค้าและด้านความคุ้มค่าในการใช้งานของระบบ หากไม่ต้องการให้ลูกค้ารอนาน ต้องใช้จำนวนผู้ให้บริการมาก แต่หากสูงเกินไป ผู้ให้บริการก็ว่างงานเป็นส่วนใหญ่ และเกิดต้นทุนสูง ดังนั้นประเด็นพื้นฐานของระบบแถวคอยคือเลือกจำนวนผู้ให้บริการหรือลักษณะการให้บริการ เพื่อถ่วงดุลระหว่างสองด้านนี้

ดัชนีชี้วัดสำหรับระบบแถวคอยส่วนใหญ่เป็นค่าคาดหมาย (Expected value) ที่คำนวณจากค่าความน่าจะเป็นที่มีลูกค้า n คนในระบบ (P_n) สำหรับทุก ๆ $n = 0, 1, \ldots$ ดังนั้นการคำนวณที่เกี่ยวข้องกับระบบแถวคอยจึงเริ่มจาก P_n ผู้เขียนแสดงสูตรคำนวณดัชนีชี้วัดและคำสั่งใน `QTP` แอดอินควบคู่ไปด้วย กำหนดให้

* λ เป็นอัตราการมาถึงในสมการที่ (6.1)
* μ เป็นอัตราการให้บริการของผู้ให้บริการ 1 คน ตามที่แสดงในสมการที่ (6.2) โดยที่ λ และ μ ต้องมีหน่วยเดียวกัน เช่น ต่อชั่วโมง หรือต่อวัน
* s เป็นจำนวนผู้ให้บริการ

สูตรดัชนีชี้วัดและคำสั่ง `QTP` ที่แสดงในหัวข้อนี้สำหรับกรณีที่ความจุของแถวคอยและจำนวนของลูกค้าที่เป็นไปได้ทั้งหมดไม่จำกัด

ค่าความน่าจะเป็นที่มีลูกค้า n คนในระบบ P_n เช่น $P_1, P_2, \ldots$

คำสั่ง QTP: `=QTPMMS_PrState(`n, λ, μ, s`)`

สัดส่วนที่ผู้ให้บริการทำงาน (Utilization, ρ) ซึ่งคือเวลาที่บริการลูกค้าหารด้วยเวลาที่เข้างานทั้งหมด คำนวณได้ดังนี้

$$\rho = \frac{\lambda}{s\mu}$$

คำสั่ง QTP: `=QTPMMS_Util(`λ, μ, s`)`

ความน่าจะเป็นที่ระบบว่าง (P_0) ไม่มีลูกค้าในระบบ ผู้ให้บริการจึงว่าง

คำสั่ง QTP: `=QTPMMS_PrEmpty(`λ, μ, s`)`

จำนวนลูกค้าที่รอใน<u>แถวคอย</u>เฉลี่ย (L_q) เริ่มมีลูกค้าในแถวเมื่อมีลูกค้าในระบบมากกว่าจำนวนผู้ให้บริการ

$$L_q = \sum_{i=s+1}^{\infty} (i-s)P_i \tag{6.3}$$

คำสั่ง QTP: `=QTPMMS_Lq(`λ, μ, s`)`

จำนวนลูกค้าที่รอใน<u>ระบบ</u>เฉลี่ย (L) รวมที่รอในแถวและที่กำลังได้รับบริการอยู่

$$L = \sum_{i=1}^{\infty} iP_i \tag{6.4}$$

คำสั่ง QTP: `=QTPMMS_L(`λ, μ, s`)`

เวลารอเฉลี่ยใน<u>แถวคอย</u> (W_q) ซึ่งคำนวณจาก L_q (สมการที่ 6.3) ด้วย **กฎของลิตเติ้ล** (Little's Law)

$$W_q = \frac{L_q}{\lambda} \tag{6.5}$$

สังเกตว่า L_q ไม่มีหน่วย ถ้า λ มีหน่วย “ต่อชั่วโมง” W_q จะมีหน่วย “ชั่วโมง”

คำสั่ง QTP: `=QTPMMS_Wq(`λ, μ, s`)`

เวลารอเฉลี่ยใน<u>ระบบ</u> (W) ซึ่งคำนวณจาก L (สมการที่ 6.4) ด้วย กฎของลิตเติ้ล เพราะเวลาเฉลี่ยในระบบ (W) เป็นผลบวกของเวลารอเฉลี่ยในแถว (W_q) และเวลาได้รับบริการเฉลี่ย ($1/\mu$)

$$W = W_q + \frac{1}{\mu} \tag{6.6}$$

$$W = \frac{L}{\lambda} \tag{6.7}$$

คำสั่ง QTP: `=QTPMMS_W(`λ, μ, s`)`

ความน่าจะเป็นที่ลูกค้าที่เข้ามาต้องรอ (P_w):

$$P_w = \sum_{i=s}^{\infty} P_i = 1 - \sum_{i=0}^{s-1} P_i$$

คำสั่ง QTP: `=QTPMMS_PrWait(`λ, μ, s`)`

หัวข้อต่อไปจะกล่าวถึงแบบจำลองแถวคอยที่พื้นฐานและง่ายที่สุด บางครั้งเรียกว่าระบบแถวคอยแบบมาร์คอฟ

6.3 แบบจำลอง M/M/s

แบบจำลองนี้มีกระบวนการมาถึงแบบปัวซองด้วยอัตรา λ ต่อหนึ่งหน่วยเวลา เวลาให้บริการมีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยอัตรา μ ต่อหน่วยเวลา ลูกค้ารอในแถวคอยแถวเดียวและได้รับบริการแบบมาก่อนได้ก่อน (FCFS) โดยผู้ให้บริการที่ว่างคนแรก มีผู้ให้บริการที่เหมือนกัน s คน เงื่อนไขความเสถียรคืออัตราการมาถึงไม่เกินอัตราที่ระบบสามารถให้บริการได้ กล่าวคือ

$$\lambda < s\mu \tag{6.8}$$

หาก $\lambda \geqslant s\mu$ แถวคอยจะยาวขึ้นเรื่อย ๆ และสูตรคณิตศาสตร์ของแบบจำลองจะเป็นโมฆะ สำหรับการใช้เพื่อตรวจสอบความถูกต้องของแบบจำลองสถานการณ์ ค่า λ และ μ ถูกประมาณจากข้อมูลที่นำมาสร้างแบบจำลอง ดังที่แสดงด้วยตัวอย่างที่ 6.1-6.2 ตัวอย่างที่ 6.3 แสดงการคำนวณดัชนีชี้วัด

ตัวอย่างที่ 6.3. ศูนย์สนับสนุนลูกค้าของบริษัทซอฟต์แวร์แห่งหนึ่งมีพนักงานดูแลลูกค้าประจำ 1 คน การโทรเข้าของลูกค้าไม่แน่นอน มีการแจกแจงแบบปัวซองด้วยอัตราเฉลี่ย 5 ต่อชั่วโมง พนักงานให้บริการลูกค้าด้วยอัตราเฉลี่ย 7 ต่อชั่วโมงโดยประมาณ แต่เวลาที่ให้บริการจริงเป็นค่าสุ่มที่มีการแจกแจงแบบเอ็กซ์โพเนนเชียล ผู้จัดการศูนย์ฯ ได้รับการร้องเรียนจากลูกค้าว่าต้องรอสายนาน เธอจึงอยากรู้ว่าเวลารอคอยเฉลี่ยขณะนี้มีค่าเท่าใด หากเวลารอคอยเฉลี่ยมากกว่า 5 นาที เธอต้องการกำหนดจำนวนพนักงานดูแลลูกค้าที่ทำให้เวลารอคอยเหลือไม่เกิน 2 นาทีหรือน้อยกว่า

หากจำลองระบบปัจจุบันด้วยแบบจำลอง M/M/1 จะใช้พารามิเตอร์ดังนี้ $\lambda = 5$/ชั่วโมง, $\mu = 7$/ชั่วโมง และ $s = 1$ สังเกตว่าค่าพารามิเตอร์เหล่านี้ทำให้เงื่อนไข (6.8) เป็นจริง เมื่อใช้สูตร `=QTPMMS_Wq(`$5, 7, 1$`)` ได้เวลารอคอยในแถว 0.36 ชั่วโมง หรือ 21.43 นาที ซึ่งสูงกว่าเป้าหมาย 5 นาทีมาก จึงทดลองเพิ่มจำนวนพนักงานอีก 1 คน เป็น 2 คน เมื่อใช้สูตร `=QTPMMS_Wq(`$5, 7, 2$`)` พบว่าเวลารอคอยในแถวเหลือ 1.25 นาที ตามเป้าหมายที่ผู้จัดการต้องการ

นอกเหนือจากเวลารอคอยในแถว ยังสามารถพิจารณาสัดส่วนการทำงานจริงของพนักงาน ระบบปัจจุบันมี $\rho = 71\%$ คำนวณด้วยสูตร =QTPMMS_Util($(5, 7, 1)$) แต่ระบบที่มีพนักงาน 2 คนมีสัดส่วนการทำงานจริงลดลงเหลือ 36%

□

แบบจำลองในหัวข้อที่แล้วสมมติว่าแถวคอยไม่มีขีดจำกัด หรือขีดจำกัดเป็นอนันต์ หัวข้อนี้จะกล่าวถึงกรณีแถวคอยมีขีดจำกัด

6.4 แบบจำลอง M/M/s ที่มีขีดจำกัดของแถวคอย

บางสถานการณ์มีขนาดของพื้นที่รอคอยจำกัด เช่น ระบบโทรศัพท์ในตัวอย่างที่ 6.3 อาจสามารถรองรับคู่สายรอเรียกได้ไม่เกิน 5 สาย ณ ขณะใดขณะหนึ่ง ถ้ามีสายเข้ามาขณะที่มีคู่สายรอเรียกอยู่แล้ว 5 สาย ลูกค้าคนนั้นจะได้รับสัญญาณสายไม่ว่าง วิธีลดจำนวนลูกค้าโทรไม่เข้าคือการเพิ่มจำนวนคู่สายที่รอเรียก แต่ถ้าลูกค้าโทรเข้าได้แต่ต้องรอนาน ลูกค้าอาจชอบการโทรไม่เข้ามากกว่า ตัวอย่างที่ 6.4 แสดงการคำนวณดัชนีชี้วัดของระบบแถวคอยแบบนี้

ตัวอย่างที่ 6.4. ต่อเนื่องจากตัวอย่างที่ 6.3 แต่ระบบรองรับคู่สายรอเรียกได้ไม่เกิน 5 สาย ผู้จัดการต้องการทราบผลกระทบของการเพิ่มจำนวนพนักงานดูแลลูกค้าจาก 1 คน เป็น 2 คน ในแง่

* เวลาที่ลูกค้ารอในแถวจนกว่าได้รับบริการ
* อัตราที่สายโทรไม่เข้า

การคำนวณในตัวอย่างนี้ใช้สูตรเหมือนในตัวอย่างที่ 6.3 แต่เพิ่มขีดจำกัดของแถวคอยที่ 5 สาย เป็นพารามิเตอร์สุดท้ายในสูตร เช่น สำหรับคำนวณเวลาที่ลูกค้ารอในแถวใช้สูตร =QTPMMS_Wq($5, 7, 1, \mathbf{5}$) อัตราที่โทรไม่เข้า คือ อัตราที่สายโทรเข้าคูณความน่าจะเป็นที่ระบบเต็มเมื่อพนักงานทุกคนกำลังบริการลูกค้าอยู่และแถวคอยเต็มขีดความจุ ดังนั้นความน่าจะเป็นที่ระบบเต็มคือ P_6 เมื่อมีพนักงานให้บริการ 1 คน และ P_7 เมื่อมีพนักงาน 2 คน

ตารางที่ 6.1 แสดงผลการคำนวณ สังเกตว่าขนาดความจุของแถวคอยเป็นปัจจัยหนึ่งที่ส่งผลต่อสมรรถนะของระบบแถวคอย สังเกตว่าเมื่อมีพนักงานให้บริการ 1 คน เวลารอคอยในแถวเมื่อกำหนดขีดจำกัดของแถวคอย (13.55 นาที) มีค่าน้อยกว่าเมื่อให้ลูกค้ารอคอยได้ไม่จำกัด (21.43 นาที) ในตัวอย่างที่ 6.3 แต่เวลารอคอยที่ลดลงนี้แลกมาด้วยการที่ลูกค้าบางส่วนโทรไม่เข้า ถึงแม้ว่าจะกำหนดขีดจำกัดของแถวคอย พนักงานเพียง 1 คนก็ยังไม่เพียงพอที่จะลดเวลารอคอยให้ต่ำกว่าเป้าหมายที่ 5 นาทีได้ จึงยังต้องใช้พนักงานรับสาย 2 คน

□

ตารางที่ 6.1: การเปรียบเทียบสมรรถนะของระบบที่จำนวนพนักงานต่าง ๆ สำหรับตัวอย่างที่ 6.4

ดัชนีชี้วัด	**ผู้ให้บริการ 1 คน**	**ผู้ให้บริการ 2 คน**
เวลาที่ลูกค้ารอในแถว	0.23 ชม. (13.55 นาที)	0.02 ชม. (1.22 นาที)
สูตร QTP	=QTPMMS_Wq$(5, 7, 1, 5)$	=QTPMMS_Wq$(5, 7, 2, 5)$
อัตราที่สายโทรไม่เข้า (ต่อชั่วโมง)	0.21	0.004
สูตร QTP	=QTPMMS_PrState$(\mathbf{6}, 5, 7, 1, 5)$	=QTPMMS_PrState$(\mathbf{7}, 5, 7, 2, 5)$
สัดส่วนการทำงานจริงของพนักงาน	68%	36%
สูตร QTP	=QTPMMS_Util$(5, 7, 1, 5)$	=QTPMMS_Util$(5, 7, 2, 5)$

นอกเหนือจากการกำหนดขีดจำกัดของแถวคอยแล้ว ขนาดของกลุ่มลูกค้าหรือประชากรอาจมีจำนวนไม่มาก จึงไม่สามารถสมมติว่าเป็นอนันต์ได้

6.5 แบบจำลอง M/M/s ที่มีขนาดประชากรจำกัด

ระบบนี้มีผู้ให้บริการ s คนโดยที่แต่ละคนให้บริการด้วยอัตรา μ ต่อหนึ่งหน่วยเวลา แต่เวลาในการให้บริการจริงมีการแจกแจงแบบเอ็กซ์โพเนนเชียล มีประชากรที่สมมติว่าเหมือนกันที่อาจเป็นลูกค้า N คน รูปแบบการมาถึงของลูกค้า**แต่ละคน**มีการแจกแจงแบบปัวซองด้วยอัตรา λ ต่อหนึ่งหน่วยเวลา ลูกค้ารอในแถว ๆ เดียว และได้รับการบริการแบบมาก่อนได้ก่อน ตัวอย่างของระบบที่มีจำนวนลูกค้าที่เป็นไปได้จำกัดนี้ เช่น ระบบซ่อมบำรุงเครื่องจักรในโรงงาน หรือระบบการซ่อมบำรุงคอมพิวเตอร์ในออฟฟิศ โดยที่**ลูกค้า**คือเครื่องจักรหรือคอมพิวเตอร์ที่เสีย **ผู้ให้บริการ**คือช่างซ่อม

แบบจำลองนี้ไม่มีใน `Queuing toolpak` แต่มีในอีกเอกซ์เซลเทมเพลทหนึ่งคือ `Q.xls` ซึ่งพัฒนาโดย ศ.ดร. David W. Ashley สามารถดาวน์โหลดได้จาก `www.stat.ncsu.edu/people/reiland/courses/st501/Q.xls` หรืออาจสืบค้นด้วยคำสำคัญ `Q.xls` เทมเพลทนี้แยกแบบจำลองออกเป็นเวิร์คชีท ๆ ที่จะใช้คือ `M/M/s with finite arriving population` ตัวอย่างที่ 6.5 แสดงวิธีการกำหนดค่าพารามิเตอร์และการคำนวณดัชนีชี้วัด

ตัวอย่างที่ 6.5. โรงกลึงแห่งหนึ่งมีเครื่องจักรที่เหมือนกัน 10 เครื่อง จำนวนครั้งที่เครื่องจักรเหล่านี้เสียสามารถจำลองด้วยการแจกแจงแบบปัวซองที่มีอัตราเฉลี่ย 0.01 ครั้งต่อ 1 ชั่วโมงการทำงาน เกิดต้นทุน 800 บาทต่อชั่วโมงต่อเครื่องจักรที่ไม่สามารถทำงานได้ โรงกลึงนี้มีช่างซ่อมบำรุง 1 คน เวลาในการซ่อม (ไม่รวมเวลารอ) มีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย 8 ชั่วโมงต่อเครื่อง (หรืออีกนัยหนึ่งคืออัตราการบริการเป็น 1/8 ต่อชั่วโมง) ผู้บริหารต้องการทราบผลกระทบของการเพิ่มจำนวนช่างซ่อมอีก 1 คน ที่มีต่อ

เวลาที่ใช้ในการนำเครื่องจักรกลับมาใช้อีกครั้งหนึ่ง (เวลาซ่อมและเวลารอจนกว่าจะได้รับการซ่อม) ช่างซ่อมมีค่าแรง 125 บาทต่อชั่วโมง

ตัวอย่างนี้แสดงการถ่วงดุลระหว่างคุณภาพการบริการ (วัดด้วยเวลาในการส่งมอบเครื่องจักรคืน) และค่าใช้จ่ายของผู้ให้บริการ ตารางที่ 6.2 เปรียบเทียบสมรรถนะของระบบเมื่อมีช่างซ่อม 1 คน และ 2 คน ค่าเสียโอกาสคำนวณจากจำนวนเครื่องจักรเสียในแผนกซ่อมคูณด้วยค่าเสียโอกาสต่อชั่วโมงต่อเครื่อง จะเห็นว่าต้นทุนรวมลดลงเมื่อเพิ่มจำนวนช่างจาก 1 คน เป็น 2 คน อาจวิเคราะห์เพิ่มเติมเพื่อหาจำนวนช่างที่ทำให้ต้นทุนต่ำที่สุด เมื่อเพิ่มช่างซ่อมเป็น 3 คน ขนาดของค่าเสียโอกาสที่ลดลงน้อยกว่าค่าแรงพนักงานที่เพิ่มขึ้น สะท้อนด้วยต้นทุนรวมที่สูงขึ้น ดังนั้นจำนวนช่างที่ทำให้ต้นทุนต่ำที่สุดคือ 2 คน

ตารางที่ 6.2: การเปรียบเทียบสมรรถนะของระบบที่จำนวนพนักงานต่าง ๆ สำหรับตัวอย่างที่ 6.5

ดัชนีชี้วัด	**ช่าง 1 คน**	**ช่าง 2 คน**	**ช่าง 3 คน**
เวลาที่เครื่องจักรอยู่ที่ห้องซ่อม (W, ชั่วโมง)	17.986	8.828	8.080
จำนวนเครื่องจักรเสียในแผนกซ่อม (L)	1.524	0.811	0.748
ค่าเสียโอกาสของเครื่องจักร (บาทต่อชั่วโมง)	1,219.51	648.96	598.07
ค่าแรงพนักงาน (บาทต่อชั่วโมง)	125	250	375
ต้นทุนรวม (บาทต่อชั่วโมง)	1,344.51	**898.96**	973.07

□

แบบจำลอง M/M/s จำกัดการแจกแจงของเวลาสุ่มว่าเป็นแบบเอ็กซ์โพเนนเชียลเท่านั้น หัวข้อต่อไปพิจารณาแบบจำลองที่เวลาระหว่างการมาถึงและเวลาให้บริการมีการแจกแจงทั่วไป

6.6 แบบจำลอง G/G/s

แบบจำลอง G/G/s เปิดให้เวลาระหว่างการมาถึงและเวลาในการให้บริการมีการแจกแจงใด ๆ ก็ได้ ผู้ใช้เพียงแค่ระบุอัตราการมาถึง (λ) อัตราการให้บริการ (μ) ส่วนเบี่ยงเบนมาตรฐานของเวลาระหว่างการมาถึง (σ_a) และส่วนเบี่ยงเบนมาตรฐานของเวลาให้บริการ (σ_s) เช่น หากเวลาในการให้บริการเป็นค่าคงที่ ส่วนเบี่ยงเบนมาตรฐานของเวลาให้บริการคือศูนย์เพราะไม่มีความแปรปรวน หรือในระบบที่ลูกค้านัดหมาย (Appointment) ส่วนเบี่ยงเบนมาตรฐานของเวลาระหว่างการมาถึงก็เป็นศูนย์เช่นกันเพราะได้กำหนดไว้ล่วงหน้าแล้ว แบบจำลอง G/G/s จะกลายเป็นแบบจำลอง M/M/s หากการมาถึงเป็นกระบวนการปัวซอง (เวลาระหว่างการมาถึงเป็นเอ็กซ์โพเนนเชียล) และเวลาในการให้บริการมีการแจกแจงแบบเอ็กซ์โพเนนเชียล แบบจำลอง G/G/s

นี้ดูใช้ได้หลากหลายก็จริง แต่มีข้อจำกัดตรงที่ดัชนีชี้วัดที่คำนวณจากแบบจำลอง G/G/s เป็นค่าประมาณ ต่างจากค่าที่ได้จากแบบจำลอง M/M/s ที่เป็นค่าตรง (Exact) สำหรับกรณีที่ความจุของแถวคอยหรือขนาดจำนวนประชากรจำกัด ตัวแบบ G/G/s มีความซับซ้อนกว่าของแบบจำลอง M/M/s มาก จึงไม่มีสูตรคำนวณใน Queuing Toolpak กรณีที่ความจุของแถวคอยและขนาดจำนวนประชากรไม่จำกัด พจน์ที่ต้องระบุในคำสั่งมีดังนี้

$$\texttt{QTPGGS\_L}(\lambda, \mu, s, , \sigma_a, \sigma_s)$$

ตัวอย่างที่ 6.6 แสดงการใช้แบบจำลองนี้เพื่อคำนวณดัชนีชี้วัด

ตัวอย่างที่ 6.6. (M/D/1) ซิปปี้ลูปเป็นร้านเปลี่ยนน้ำมันเครื่องรถยนต์ที่เปิดทำการ 10 ชั่วโมงต่อวันและ 6 วันต่อสัปดาห์ ได้กำไร 100 บาทต่อคัน จำนวนรถของลูกค้าเข้าใช้บริการสามารถจำลองด้วยกระบวนการปัวซองด้วยค่าเฉลี่ย 3.5 คันต่อชั่วโมง เวลาที่ใช้ในการเปลี่ยนน้ำมันเครื่องด้วยมือเฉลี่ยประมาณ 15 นาที (0.25 ชั่วโมง) ด้วยส่วนเบี่ยงเบนมาตรฐาน 2 นาที (หรือ 2/60 ชั่วโมง) สมมติว่าเปลี่ยนน้ำมันเครื่องได้ทีละคัน

มีเครื่องเปลี่ยนน้ำมันอัตโนมัติราคา 500,000 บาทในท้องตลาด ผู้ขายอ้างว่าเครื่องนี้ใช้เวลาเปลี่ยนน้ำมันเครื่อง 3 นาทีต่อคัน เจ้าของร้านซิปปี้ลูปต้องการศึกษาผลกระทบของเครื่องเปลี่ยนน้ำมันอัตโนมัตินี้ในแง่ธุรกิจและประมาณเวลาคืนทุนของเครื่องนี้

หากการมาถึงเป็นกระบวนการปัวซอง เวลาระหว่างการมาถึงมีการแจกแจงแบบเอ็กซ์โพเนนเชียลซึ่งมีค่าเฉลี่ยเท่ากับส่วนเบี่ยงเบนมาตรฐาน ดังนั้นระบบปัจจุบันจึงมี $\lambda = 3.5, \mu = 4, s = 1, \sigma_a = 1/3.5, \sigma_s = 2/60$ และเวลามีหน่วยชั่วโมง

เมื่อใช้เครื่องจักรเปลี่ยนน้ำมันเครื่อง จะกำหนดส่วนเบี่ยงเบนมาตรฐานของเวลาให้บริการเป็นศูนย์ เพราะสมมติว่าค่าในแต่ละครั้งใกล้เคียงกันมาก เพื่อการวิเคราะห์ สมมติเพิ่มว่าเมื่อใช้เครื่องอัตโนมัติ ลูกค้าจะเข้าร้านมากขึ้นเนื่องจากเห็นว่าร้านบริการได้เร็วและแถวคอยสั้น อัตราที่เพิ่มนี้เป็นอัตราที่ทำให้จำนวนลูกค้าในระบบ (L) เท่าเดิม ระบบปัจจุบันมี $L \approx 3.99$ อัตราการมาถึงที่ทำให้ระบบที่นำเสนอนี้มี L เท่ากันคือประมาณ 17.535 คันต่อชั่วโมง (โดยการลองผิดลองถูกบนเอกซ์เซล)

ตารางที่ 6.3: การเปรียบเทียบการเปลี่ยนน้ำมันเครื่องด้วยมือและด้วยเครื่อง (ตัวอย่างที่ 6.6)

ดัชนีชี้วัด	**ระบบปัจจุบัน**	**เครื่องอัตโนมัติ**
เวลารอ (W_q, ชั่วโมง)	0.89	0.18
สูตร QTP	=QTPGGS_W$(3.5, 4, 1, , 1/3.5, 2/60)$	=QTPGGS_W$(17.535, 20, 1, , 1/17.535, 0)$
จำนวนรถในร้าน (L)	3.99	3.99

ในการวิเคราะห์เบื้องต้น จะไม่พิจารณาอัตราคิดลด, การเสื่อมราคาของเครื่องจักร, หรือค่าใช้จ่ายอื่น ๆ

เวลาคืนทุนสามารถคำนวณได้ดังนี้

$$
\begin{aligned}
\text{อัตราที่ลูกค้าเข้าร้านเพิ่มขึ้น} &= 17.535 - 3.5 = 14.035 \text{ คันต่อชั่วโมง} \\
\text{กำไรที่เพิ่มขึ้น} &= 100 \times 14.035 = 1403.5 \text{ บาทต่อชั่วโมง} \\
&= 1403.5 \times 10 = 14035 \text{ บาทต่อวัน} \\
&= 14035 \times 6 = 84210 \text{ บาทต่อสัปดาห์} \\
\text{เวลาคืนทุน} &= 500000/84210 = 5.94 \text{ สัปดาห์}
\end{aligned}
$$

ตัวอย่างนี้อาจมีข้อสมมติที่ไม่สมจริงหลายประการ แต่ประเด็นหลักของตัวอย่างนี้คือแสดงให้เห็นว่าการคำนวณเบื้องต้นสามารถทำได้ง่ายและรวดเร็ว หากตัวเลือกที่พิจารณามีแนวโน้มว่าจะดี ก็สมควรที่จะวิเคราะห์อย่างละเอียดขึ้นเพิ่มเติม □

ในกรณีแถวคอยที่พิจารณารองรับลูกค้าที่มีเวลาให้บริการหลายแบบ ผู้วิเคราะห์ยังสามารถใช้แบบจำลองแถวคอยอย่างง่ายเพื่อประมาณสมรรถนะระบบ โดยต้องประมาณค่าอัตราการให้บริการของระบบ

การประมาณค่าอัตราการให้บริการเมื่อมีงานหลายชนิด

ผู้วิเคราะห์สามารถประมาณอัตราการให้บริการรวม จากเวลาการให้บริการเฉลี่ยย่อย ๆ สมมติว่ามีลูกค้า n ชนิดและแต่ละชนิดมีอัตราการใช้บริการ μ_i และมีสัดส่วนของงานที่แถวคอยนั้นเป็น f_i โดยที่ $\sum_{i=1}^{n} f_i = 1$ อัตราการให้บริการเฉลี่ย ($\bar{\mu}$) คำนวณจากเวลาการให้บริการเฉลี่ยได้ดังนี้

$$\frac{1}{\bar{\mu}} = \sum_{i=1}^{n} f_i \left(\frac{1}{\mu_i}\right)$$

เช่น มีงาน 2 ชนิดที่มี $\mu_1 = 1$ และ $\mu_2 = 2$ และ $f_1 = f_2 = 0.5$ จะได้

$$\frac{1}{\bar{\mu}} = \frac{0.5}{1} + \frac{0.5}{2} = 3/4$$

ดังนั้น $\bar{\mu} = 4/3$

หัวข้อที่ผ่านมากล่าวถึงแถวคอยทีละระบบ ในแง่ปฏิบัติ ระบบที่พิจารณามักเป็นเครือข่ายแถวคอย

6.7 เครือข่ายแถวคอย

ระบบจริงมีความซับซ้อนกว่าแถวคอยเพียงอันเดียว กล่าวคือเป็นแถวคอยหลายระบบเชื่อมต่อกันเป็นเครือข่าย เช่น สายการประกอบ (Assembly line), Flow shop หรือ Job shop หัวข้อนี้สมมติว่าแถวคอยแต่ละ

อันมีความจุของแถวคอยไม่จำกัด เพื่อไม่ต้องพิจารณากรณีที่แถวคอยต้นน้ำไม่สามารถปล่อยงานออกเพราะแถวคอยถัดไปมีงานเต็ม (Blocking) แถวคอยบางอันอาจมีลูกค้าเข้ามาจากภายนอก แถวคอยที่ j ให้บริการลูกค้าด้วยอัตราเดียวคือ μ_j แต่ว่าแต่ละงานไม่จำเป็นต้องมีลำดับงานเดียวกัน งานที่ออกจากแถวคอย i อาจไปสู่แถวคอย j ด้วยความน่าจะเป็น p_{ij} เช่น งานที่ไม่ผ่านการตรวจสอบคุณภาพอาจถูกนำไปทิ้งหรืออาจนำไปแก้ไข วิธีการวิเคราะห์เครือข่ายแถวคอยคือคำนวณอัตราการมาถึงสุทธิของแต่ละแถวคอย แล้วจึงวิเคราะห์แต่ละแถวคอยทีละระบบ หลักการคืออัตราการมาถึงแถวคอยต้องเท่ากับอัตราการไหลออก นอกจากนี้ยังมีเงื่อนไขความเสถียรของแถวคอย คือ อัตราการมาถึงของแถวคอย j (λ_j) ต้องไม่เกินอัตราการทำงานของแถวคอยนั้น ($s_j\mu_j$) เมื่อ s_j เป็นจำนวนผู้ให้บริการของแถวคอย j และ μ_j เป็นอัตราการให้บริการของผู้ให้บริการแต่ละคน นั่นคือ

$$\lambda_j \leqslant s_j\mu_j$$

ทฤษฎีของเครือข่ายแจ็กสัน (Jackson network) กล่าวว่า หากกระบวนการมาถึงของงานจากภายนอกเป็นกระบวนการปัวซอง และแถวคอย j มีผู้ให้บริการที่เหมือนกัน s_j หน่วย โดยที่เวลาในการให้บริการมีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยอัตรา μ_j และมีความจุของแถวคอยไม่จำกัด แถวคอยนั้นจะมีพฤติกรรมเหมือนแถวคอย M/M/c_j (Kleinrock, 1975)

หัวข้อนี้จะกล่าวถึงการคำนวณอัตราการมาถึงสุทธิ ตามด้วยตัวอย่างการคำนวณเครือข่ายแถวคอย

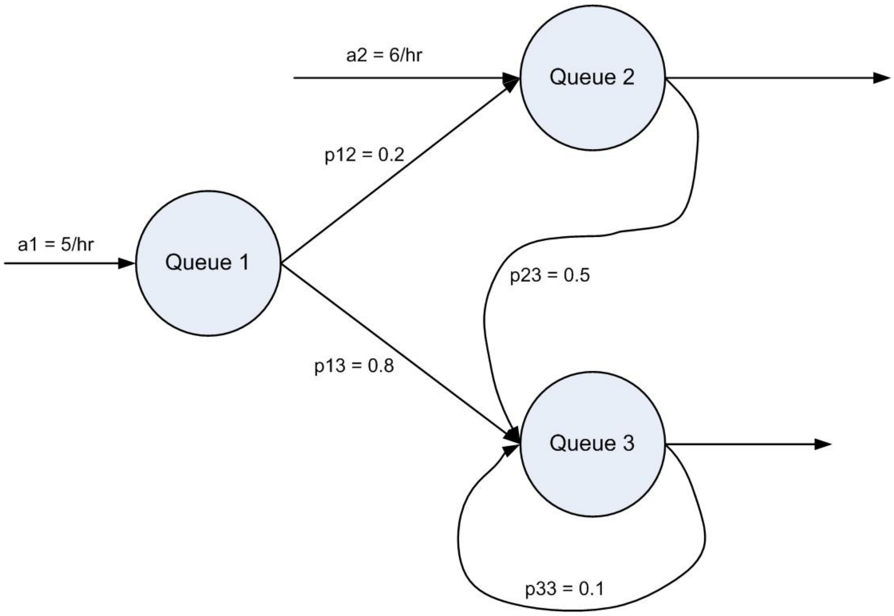
การคำนวนอัตราการมาถึงสุทธิ

การคำนวณอัตราการมาถึงสุทธิอาศัยหลักการง่าย ๆ ว่าอัตราการมาถึงต้องเท่ากับอัตราการไหลออก และอัตราการมาถึงสุทธิเป็นผลบวกของอัตราการมาถึงย่อย ๆ ซึ่งอาจเป็นการไหลกลับ (Feedback) ของแถวคอยนั้นเองก็ได้ ให้มองแถวคอยว่าเป็นกล่องดำ (Black box) คือไม่ต้องพิจารณาจำนวนผู้ให้บริการที่มี กำหนดให้ λ_j เป็นอัตราการมาถึงสุทธิของแถวคอย j; a_j เป็นอัตราการมาถึงจากภายนอก; p_{ij} เป็นความน่าจะเป็นที่ลูกค้าออกจากแถวคอย i และไหลสู่แถวคอย j; เครือข่ายแถวคอยนี้มีแถวคอย n ระบบ

$$\lambda_j = a_j + \sum_{i=1}^{n} \lambda_i p_{ij} \tag{6.9}$$

สังเกตว่ามีหนึ่งสมการสำหรับหนึ่งแถวคอย จึงมี n สมการและ n ตัวแปรที่ไม่ทราบค่า (Unknown variable) เมื่อพิจารณาร่วมกันจึงเป็นระบบของสมการเชิงเส้น (System of linear equations) ที่สามารถใช้ Solver ที่มีในเอกซ์เซลเพื่อแก้สมการได้

ตัวอย่างที่ 6.7. ในรูปที่ 6.1 มีแถวคอย 3 ระบบ แถวคอยที่ 1 และ 2 มีงานจากภายนอกเข้าระบบ ส่วนแถวคอยที่ 3 มีการไหลกลับ



รูปที่ 6.1: เครือข่ายแถวคอยสำหรับตัวอย่าง 6.7

อัตราการมาถึงแต่ละแถวคอยสุทธิสามารถคำนวณโดยใช้สมการที่ (6.9) ดังนี้

$$\begin{aligned}
\lambda_1 &= a_1 = 5 \text{ /hr} \\
\lambda_2 &= a_2 + \lambda_1 p_{12} = 6 + (0.2 \times 5) = 7 \text{ /hr} \\
\lambda_3 &= \lambda_1 p_{13} + \lambda_2 p_{23} + \lambda_3 p_{33} \\
&= (5 \times 0.8) + (7 \times 0.5) + (\lambda_3 \times 0.1) \\
0.9\lambda_3 &= 7.5 \\
\lambda_3 &= 8.33 \text{ /hr}
\end{aligned}$$

ในตัวอย่างนี้ สามารถแก้สมการได้ทีละสมการและแทนค่าพจน์ที่ทราบค่าตามลำดับ ในบางกรณีอาจต้องใช้ Solver ช่วยแก้เพื่อให้รวดเร็วและผิดพลาดน้อยกว่าแก้ด้วยมือ

□

ตัวอย่างที่ 6.8 แสดงการคำนวณดัชนีชี้วัดสำหรับเครือข่ายแถวคอย

ตัวอย่างที่ 6.8. มีแถวคอย 2 ระบบเชื่อมต่อกันเป็นอนุกรม แต่ละแถวคอยมีผู้ให้บริการ 1 คน เวลาในการให้บริการเป็นอิสระต่อกันและมีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย 3 นาทีที่แถวคอยที่ 1 และ 4 นาทีที่แถวคอยที่ 2 แถวคอยแรกมีงานเข้าระบบที่มีลักษณะเป็นกระบวนการปัวซองด้วยอัตรา 10 ต่อชั่วโมง

ให้คำนวณดัชนีต่อไปนี้

* ค่าเฉลี่ยจำนวนงานที่แถวคอย 1 และแถวคอย 2
* ความน่าจะเป็นที่ผู้ให้บริการทั้งสองคนว่าง
* จำนวนงานในระบบเฉลี่ยและเวลาในระบบเฉลี่ย

เนื่องจากแถวคอยสองระบบนี้ต่อกันเป็นอนุกรมและไม่มีงานถูกตีกลับ อัตราที่งานมาถึงสุทธิจึงเท่ากัน $\lambda_1 = \lambda_2 = 10$/ชม. อัตราการให้บริการเป็นส่วนกลับของเวลาให้บริการเฉลี่ย จึงได้ว่า $\mu_1 = 20$/ชม. และ $\mu_2 = 15$/ชม. ด้วยข้อสมมติของโจทย์ทำให้สามารถจำลองระบบแถวคอยที่ 1 และ 2 เป็น M/M/1 อาจใช้คำสั่ง `=QTPMMS(...)` หรือคำนวณด้วยมือจากสูตรก็ได้ ได้ผลดังนี้

- ค่าเฉลี่ยจำนวนงานที่แถวคอย 1 (L_1) = 1 และแถวคอย 2 (L_2) = 2
- ความน่าจะเป็นที่ผู้ให้บริการทั้งสองคนว่าง = ความน่าจะเป็นที่แถวคอยที่ 1 ว่าง $\times$ ความน่าจะเป็นที่แถวคอยที่ 2 ว่าง = $(1/2) \times (1/3) = 1/6$
- จำนวนงานในระบบเฉลี่ย = $L_1 + L_2 = 3$ เวลาในระบบเฉลี่ยคือผลบวกของเวลาในแถวคอยที่ 1 และเวลาในแถวคอยที่ 2 = $0.1 + 0.2 = 0.3$ ชม.

□

6.8 บทส่งท้าย

ในการนำแบบจำลองแถวคอยไปใช้กับปัญหาจริง ผู้วิเคราะห์ควรระลึกไว้ว่าแบบจำลองเหล่านี้ให้ผลลัพธ์ที่เป็นค่าเฉลี่ยระยะยาวคือที่สภาวะคงตัว (Steady state) และไม่ได้นำผลกระทบจากช่วงที่ระบบเพิ่งเริ่มปฏิบัติการ (Start-up effects) เช่น ขณะที่ธนาคารเพิ่งเปิด ไม่มีลูกค้าอยู่ในธนาคาร มาคำนวณ ผู้ใช้ตัวแบบจึงควรทราบว่าแบบจำลองนั้นประเมินสมรรถนะระบบจริงไปในทิศทางใด ค่าจริงน่าจะต่ำกว่าที่แบบจำลองประเมินหรือสูงกว่า แบบจำลอง M/M/s มีเวลาระหว่างการมาถึงและเวลาให้บริการที่มีการแจกแจงแบบเอ็กซ์โพเนนเชียล ซึ่งมีสัมประสิทธิ์การแปรผัน (Coefficient of variation, c_v) เท่ากับหนึ่ง c_v คำนวณจากส่วนเบี่ยงเบนมาตรฐานหารด้วยค่าเฉลี่ย หากสถานการณ์ที่พิจารณามีสัมประสิทธิ์การแปรผัน**น้อยกว่า** 1 คาดว่าสถานการณ์จริงจะคับคั่ง**น้อยกว่า**ที่แบบจำลองให้ค่าไว้ คือ ค่าเวลารอคอยจริงหรือจำนวนลูกค้าในระบบน่าจะน้อยกว่าค่าที่แบบจำลองประเมิน ในทางตรงกันข้ามหากมีสัมประสิทธิ์การแปรผัน**มากกว่า** 1 คาดว่าสถานการณ์จริงจะคับคั่ง**มากกว่า** คือ ค่าเวลารอคอยจริงหรือจำนวนลูกค้าในระบบน่าจะมากกว่าค่าที่แบบจำลองประเมิน

แบบฝึกหัด

6.1. เมื่อโทรเข้าศูนย์แจ้งเหตุ 191 โอกาสที่จะไม่มีคนรับสายต้องต่ำมาก ๆ ถึงแม้ว่าพนักงานรับสายอาจใช้เวลานานมาก ๆ ในการพูดคุยแต่ละสาย ศูนย์ฯ แยกสายโทรเป็น 2 ชนิดคือ **ให้ข้อมูล** (เวลาพูดในสายประมาณ 1 นาที) และ**แจ้งเหตุด่วน** (เวลาพูดในสายประมาณ 15 นาที) ประมาณ 20% ของสายที่เข้ามาเป็นชนิดแจ้งเหตุด่วน ในช่วงที่ยุ่งมาก ๆ สายเข้ามาที่ศูนย์ฯ ด้วยอัตรา 30 ต่อชั่วโมง ทางศูนย์ฯ ต้องการจ้างพนักงานให้เพียงพอที่จะไม่มีประชาชนรอสาย จริง ๆ แล้วถ้าพนักงานรับสายทุกคนรับสายอยู่ สายที่เข้ามาจะถูกโอนไปยังสถานีตำรวจซึ่งอยู่นอกขอบเขตของแบบจำลองนี้ ให้สมมติว่ากระบวนการโทรเข้าเป็นปัวซองและเวลาให้บริการมีการแจกแจงแบบเอ็กซ์โพเนนเชียล และระบบปัจจุบันมีพนักงานรับสาย 4 คน

ให้คำนวณดัชนีต่อไปนี้

(i) อัตราให้บริการของพนักงานรับสาย

(ii) ค่าเฉลี่ยจำนวนพนักงานรับสายที่ทำงานจริง

(iii) ความน่าจะเป็นที่สายจะถูกโอนไปสถานีตำรวจ

(iv) อัตราที่ศูนย์รับสาย

(v) อัตราที่สายถูกโอนไปสถานีตำรวจ

6.2. แผนกซ่อมและตรวจสอบแห่งหนึ่งประกอบด้วยสองสถานี สถานีซ่อมมีช่าง 2 คน และสถานีตรวจสอบคุณภาพมีช่าง 1 คน ช่างซ่อมแต่ละคนทำงานด้วยอัตรา 3 ต่อชั่วโมง ช่างตรวจสอบทำงานด้วยอัตรา 8 ต่อชั่วโมง ประมาณ 10% ของชิ้นงานไม่ผ่านการตรวจสอบและถูกส่งกลับไปที่แผนกซ่อม (สมมติว่าเป็นสัดส่วนนี้ถึงแม้ว่าชิ้นงานจะถูกซ่อมหลายครั้งแล้ว) งานมาถึงระบบด้วยอัตรา 5 ต่อชั่วโมง สามารถจำลองรูปแบบที่งานมาถึงด้วยกระบวนการปัวซองและเวลาให้บริการด้วยการแจกแจงแบบเอ็กซ์โพเนนเชียล ให้คำนวณ

(i) เวลาที่งานรอที่สถานีซ่อมและเวลารอที่สถานีตรวจสอบ

(ii) อัตราที่งานไหลเข้ามากที่สุดที่ระบบจะรองรับได้โดยยังไม่ต้องจ้างคนเพิ่ม

6.3. ระบบแถวคอยแบบ M/M/1 มีอัตราการมาถึง $\lambda = 0.4$ ต่อชั่วโมง และอัตราการให้บริการ $\mu = 0.5$ ต่อชั่วโมง

(i) ให้ใช้แบบจำลองแถวคอยคำนวณค่าเฉลี่ยของเวลารอคอยของลูกค้าในแถว
ถ้าผู้วิเคราะห์สร้างแบบจำลองสถานการณ์ของระบบนี้ แล้วพบว่าค่าเฉลี่ยของเวลารอคอยของลูกค้าในแถวคือ 6.32 ชั่วโมง

(ii) ให้เหตุผลอย่างน้อย 2 ประการสำหรับความแตกต่างของค่าเฉลี่ยจากตัวแบบแถวคอยและจากแบบจำลองสถานการณ์

(iii) ระบุข้อดีของแบบจำลองสถานการณ์เมื่อเปรียบเทียบกับแบบจำลอง M/M/s อย่างน้อย 2 ข้อ

6.4. ผู้โดยสารเดินมาที่เคาน์เตอร์เช็คอินที่สนามบินด้วยอัตรา 0.625 ต่อนาที แต่จำนวนผู้โดยสารที่มาจริงไม่คงที่ สมมติว่าเวลาระหว่างการมาถึงมีการแจกแจงแบบเอ็กซ์โพเนนเชียล ที่หน้าเคาน์เตอร์เช็คอิน ผู้โดยสารรอในแถวเดียวจนกว่าเจ้าหน้าที่คนใดคนหนึ่งจะว่าง มีเจ้าหน้าที่ 5 คน เวลาเช็คอินประมาณ 5 นาที ด้วยความแปรปรวนเฉลี่ย $\pm$ 25% รอบค่าเฉลี่ยนี้ เมื่อเช็คอินเสร็จ ผู้โดยสารสามารถเดินไปที่ประตูขึ้นเครื่องได้ ให้ใช้แบบจำลองแถวคอยเพื่อประมาณเวลาเฉลี่ยที่ผู้โดยสารอยู่ที่เคาน์เตอร์เช็คอิน ให้ระบุชนิดของแบบจำลองที่ใช้

6.5. ห้องเก็บเครื่องมือรองรับการเข้าใช้บริการแบบปัวซอง มีเวลาการให้บริการแบบเอ็กซ์โพเนนเชียล และให้บริการช่างแมคคานิคเป็นจำนวนมาก โดยเฉลี่ย จะมีช่าง 1 คนเข้ามาใช้บริการทุกๆ 4 นาที และพนักงานห้องเก็บเครื่องมือใช้เวลาประมาณ 3 นาทีในการให้บริการช่าง 1 คน ค่าแรงของพนักงานห้องเก็บเครื่องมือวันละ 300 บาท และค่าแรงของช่างแมคคานิควันละ 600 บาท คุณคิดว่าทางโรงงานควรจ้างพนักงานห้องเก็บเครื่องมือเพิ่มอีกคนหรือไม่ ให้แสดงเหตุผลที่เป็นตัวเลขประกอบข้อเสนอแนะของคุณ

7

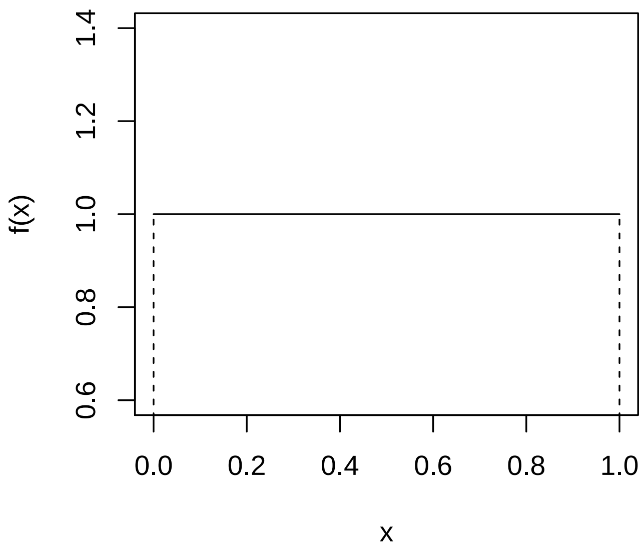
การสร้างค่าสุ่ม

บทที่ผ่านมาถือว่าแบบจำลองเป็นกล่องปิดทึบสีดำ (Black box) ผู้ใช้กำหนดตัวแปรนำเข้า ตรรกะความสัมพันธ์ระหว่างตัวแปร และดัชนีชี้วัดที่เป็นเอาท์พุท บทนี้จะอธิบายว่าความไม่แน่นอนในตัวแบบเกิดขึ้นได้อย่างไร โดยแบ่งออกเป็นสองส่วน หัวข้อที่ 7.1 กล่าวถึงการสร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอมาตรฐาน ซึ่งก็คือการแจกแจงแบบสม่ำเสมอที่อยู่ในช่วง [0, 1] มักเรียกว่า**การสร้างเลขสุ่ม** (Random number generation) เพราะค่าสุ่มแบบสม่ำเสมอมาตรฐานนี้เป็นพื้นฐานในการสร้างค่าสุ่มที่มีการแจกแจงแบบอื่น ๆ ซึ่งจะกล่าวถึงในหัวข้อที่ 7.2 เรื่อง**การสร้างค่าตัวแปรสุ่ม** (Random variate generation)

7.1 การสร้างเลขสุ่มแบบสม่ำเสมอมาตรฐาน

ฟังก์ชันที่สร้างค่าเลขสุ่มแบบสม่ำเสมอมาตรฐานนี้ให้ค่าสุ่มที่มีค่ามากกว่า 0 แต่ไม่เกิน 1 เช่น ฟังก์ชัน `RAND()` ในเอกซ์เซล สมการที่ (7.1) กำหนดฟังก์ชันความหนาแน่นความน่าจะเป็น (PDF) ของการแจกแจงนี้ มีค่าเฉลี่ยที่ $1/2$ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ $1/\sqrt{12}$ รูปที่ 7.1 แสดงกราฟของ PDF

$$f(x) = \begin{cases} 1, & \text{สำหรับ } 0 \leqslant x \leqslant 1 \\ 0, & \text{สำหรับ } x < 0 \text{ หรือ } x > 1 \end{cases} \tag{7.1}$$



รูปที่ 7.1: PDF ของการแจกแจงแบบสม่ำเสมอมาตรฐาน

วิธีที่ใช้สร้างเลขสุ่มให้**เลขสุ่มเทียม** (Pseudo-random number) ไม่ใช่เลขสุ่มจริง ๆ โดยใช้ขั้นตอนวิธี (Algorithm) สร้างลำดับตัวเลข (Sequence) หรือมองได้ว่าเป็นตารางตัวเลขที่มีอันดับก่อนหลังที่ตายตัว เช่น

ตารางที่ A.1 ใน Banks et al. (2005) แสดงทศนิยมของเลขสุ่มดังนี้ $0.94737, 0.87259, 0.63856, \ldots$ แบบจำลองดึงเลขสุ่มทีละตัวจากตารางโดยการเรียกฟังก์ชัน เมื่อใดก็ตามที่ใช้ตัวเลขจนสิ้นสุดตารางก็จะกลับมาเริ่มที่ต้นตารางใหม่และวนอย่างนี้ไปเรื่อย ๆ เขียนเป็นลำดับได้ดังนี้

$$u_1, u_2, u_3, \ldots, u_{P-1}, u_P, u_1, u_2, \ldots$$

โดยที่ P คือ**คาบ** (Period) หรือจำนวนตัวเลขสุ่มในลำดับ เหตุผลที่ใช้ขั้นตอนวิธีทางคณิตศาสตร์แทนที่จะใช้วิธีทางกายภาพ (เช่น การแตกตัวของอะตอมกัมมันตภาพรังสี) ก็เพื่อให้ได้เลขสุ่มที่มีคุณสมบัติที่ต้องการ ดังนี้

- **ความสม่ำเสมอ** (Uniformity): ตัวเลขที่สร้างได้กระจายตัวอย่างสม่ำเสมอในช่วง 0 ถึง 1 ไม่กระจุกตัวอยู่ที่ใดที่หนึ่งหรือว่างเว้นช่วงใดช่วงหนึ่ง

- **ความเป็นอิสระต่อกัน** (Independence): เลขสุ่มค่าหนึ่ง ๆ เป็นอิสระต่อค่าอื่น ๆ (ไม่มีสหสัมพันธ์)

- **ประสิทธิภาพและความประหยัด** (Efficiency and Economy): วิธีที่ใช้สร้างเลขสุ่มประมวลผลได้เร็วและไม่อาศัยหน่วยความจำของคอมพิวเตอร์มากนักเนื่องจากแบบจำลองต้องใช้เลขสุ่มจำนวนมาก ขั้นตอนวิธีที่ใช้สร้างค่าตัวแปรสุ่มสามารถโยกย้ายไปยังเครื่องคอมพิวเตอร์อื่น ๆ ได้ง่าย (Portable)

- **ความทำซ้ำได้** (Reproducibility): สามารถสร้างเลขสุ่มชุดใด ๆ ซ้ำได้ คุณสมบัตินี้มีประโยชน์สำหรับการเปรียบเทียบทางเลือก ให้ทุกทางเลือกได้เผชิญกับความไม่แน่นอนอย่างเดียวกัน เพื่อให้เอาท์พุทที่ประมวลผลได้จากแบบจำลองสะท้อนความแตกต่างของทางเลือกได้ชัดขึ้น

- **สายเลขสุ่มหลาย ๆ สาย** (Multiple streams): สายเลขสุ่ม (Random number stream) ทำงานเสมือนเป็นตัวสร้างเลขสุ่ม (Random number generator) ที่แยกกัน แต่ละสายนำไปใช้สำหรับแต่ละแหล่งความแปรปรวนในแบบจำลอง เช่น ขบวนการมาถึง ขบวนการให้บริการ

หัวข้อที่ 7.1.1-7.1.3 จะกล่าวถึงขั้นตอนวิธีที่ใช้สร้างเลขสุ่มแบบสม่ำเสมอ ซึ่งพัฒนาเพิ่มเติมจากวิธีก่อนหน้าเป็นลำดับเพื่อให้ได้คาบที่ยาวขึ้นและมีคุณสมบัติทางสถิติที่ดี (แหล่งอ้างอิงหลักคือ Nelson 2013)

7.1.1 Multiplicative Congruential Generator (MCG)

วิธี MCG ใช้เลขสุ่มตัวหนึ่ง z_{i-1} สร้างเลขสุ่มตัวถัดไป z_i ด้วยสูตรดังนี้

$$z_i = a z_{i-1} \bmod m \tag{7.2}$$

$$u_i = \frac{z_i}{m} \tag{7.3}$$

ในสมการที่ (7.2) ตัวปฏิบัติการ mod คือจำนวนเต็มที่เป็นเศษเหลือจากการหารด้วย m เช่น $11 \bmod 3 = 2$ สังเกตว่า z_i เป็นจำนวนเต็มที่อยู่ใน $\{0, 1, 2, \ldots, m-1\}$ จึงต้องหารด้วย m (สมการที่ 7.3) เพื่อให้ได้ค่าที่อยู่ในช่วง $[0, 1]$

กรณีทั่วไปของ MCG คือ Linear congruential generator (LCG) ซึ่งมีพจน์ c บวกเพิ่มในสมการที่ (7.2) จึงได้สมการที่ (7.4) ข้างล่างนี้ และอาศัยสมการที่ (7.3) เพื่อให้ค่าที่ได้ u_i อยู่ในช่วง $[0, 1]$

$$z_i = (az_{i-1} + c) \bmod m \tag{7.4}$$

ตัวอย่างที่ 7.1 แสดงการใช้สูตร (7.2)-(7.3)

ตัวอย่างที่ 7.1. (ตัวอย่างจากหัวข้อที่ 6.5 ใน Nelson 2013) พิจารณา MCG ดังต่อไปนี้

$$z_i = (630{,}360{,}016 z_{i-1}) \bmod (2^{31} - 1) \tag{7.5}$$

$$u_i = \frac{z_i}{2^{31} - 1} \tag{7.6}$$

กำหนดให้ $z_0 = 281{,}629{,}770$ ใช้สมการที่ (7.5)-(7.6) อย่างต่อเนื่อง จะได้

$$\begin{aligned}
z_1 &= (630{,}360{,}016 z_0) \bmod (2^{31} - 1) = 405{,}335{,}040 \\
u_1 &= \frac{z_1}{m} = \frac{405{,}335{,}040}{2^{31} - 1} = 0.1887 \\
z_2 &= (630{,}360{,}016) z_1 \bmod (2^{31} - 1) = 1{,}422{,}605{,}440 \\
u_2 &= \frac{z_2}{m} = \frac{1{,}422{,}605{,}440}{2^{31} - 1} = 0.6624 \\
&\vdots \quad \vdots \quad \vdots
\end{aligned}$$

□

มีข้อสังเกตเกี่ยวกับวิธี MCG ดังนี้

* คาบที่มากที่สุดที่เป็นได้ของ MCG (สมการที่ 7.2-7.3) คือ $m - 1$ เพราะเลขสุ่ม u_i ที่ได้จากสมการที่ (7.3) คือ $\{0, 1/m, 2/m, \ldots, (m-1)/m\}$ เพราะไม่ใช้ 0 เพราะทำให้ค่าต่อมาเป็น 0; หากใช้ LCG (สมการที่ 7.4) คาบที่มากที่สุดที่เป็นได้คือ m
* ต้องการให้คาบมีค่าสูง เพื่อให้เลขสุ่มที่สร้างได้เป็นจุดถี่ ๆ ในช่วง [0,1] หากคอมพิวเตอร์เป็นแบบ 32-บิต จำนวนเต็มสูงสุดที่เครื่องรองรับได้คือ $m = 2^{31} - 1$ แต่การเลือกค่าพารามิเตอร์อื่น ๆ เช่น a และ z_0 ให้เหมาะสมก็สำคัญต่อขนาดของคาบ

* จากสมการที่ (7.2) สามารถพิสูจน์ได้ว่า

$$z_i = a^i z_0 \bmod m = (a^i \bmod m) z_0 \bmod m$$

ดังนั้นตัวสร้างค่าตัวแปรสุ่มเพียงตัวเดียวที่มีคาบยาวพอสามารถใช้เป็น**ตัวสร้างค่าตัวแปรสุ่มเสมือน**หลาย ๆ ตัวได้ โดยที่แต่ละอันเริ่มต้นที่**ต้นกำเนิด** (Seed) ที่ห่างกัน z_0 (Nelson, 2013)

วิธี MRG พัฒนาต่อเนื่องจาก MCG (หัวข้อที่ 7.1.1) กล่าวคือ MCG ใช้เฉพาะเลขสุ่มตัวล่าสุดในการสร้างค่าต่อไป แต่ MRG ใช้ K ค่าล่าสุด เพื่อให้ได้คาบที่ยาวขึ้น

7.1.2 Multiple Recursive Generator (MRG)

วิธี MRG มีขั้นตอนดังนี้

$$\begin{aligned} z_i &= (a_1 z_{i-1} + a_2 z_{i-2} + \cdots + a_K z_{i-K}) \bmod m \qquad (7.7) \\ u_i &= \begin{cases} \frac{z_i}{m+1} & z_i > 0 \\ \frac{m}{m+1} & \text{สำหรับกรณีอื่น ๆ} \end{cases} \end{aligned}$$

คาบที่มากที่สุดที่เป็นไปได้ของ MRG ข้างต้นนี้คือ $m^K - 1$ เพราะว่าแต่ละ z_i ในสมการที่ (7.7) มีค่าที่เป็นไปได้ m ค่า และไม่ใช้ $(0, 0, \ldots, 0)$ แต่ทั้งนี้คาบที่แท้จริงก็ขึ้นอยู่กับพารามิเตอร์ $(a_1, a_2, \ldots, a_K)$ และต้นกำเนิดที่ใช้ $(z_0, z_1, \ldots, z_{K-1})$ ซึ่งสังเกตว่ามี K ตัว

7.1.3 ตัวสร้างค่าตัวแปรสุ่มรวม

ตัวสร้างค่าตัวแปรสุ่มรวม (Combined Generator) ใช้ MRG หลาย ๆ ตัวพร้อมกันซึ่งสามารถเพิ่มคาบได้อีก สมมติว่ามี MRG J ตัวและตัวที่ j ให้เลขสุ่มลำดับที่ i ด้วยสมการที่ (7.8) แล้วนำมารวมกันด้วยสมการที่ (7.9)

$$\begin{aligned} z_{i,j} &= (a_{1,j} z_{i-1,j} + a_{2,j} z_{i-2,j} + \cdots + a_{K,j} z_{i-K,j}) \bmod m, 1 \leqslant j \leqslant J \qquad (7.8) \\ z_i &= (\delta_1 z_{i,1} + \delta_2 z_{i,2} + \ldots + \delta_J z_{i,J}) \bmod m_1 \qquad (7.9) \\ u_i &= \begin{cases} \frac{z_i}{m_1+1} & z_i > 0 \\ \frac{m_1}{m_1+1} & \text{สำหรับกรณีอื่น ๆ} \end{cases} \end{aligned}$$

ถ้าตัวสร้างค่าตัวแปรสุ่มที่ j มีมอดูลัสเท่ากับ m_j และมีคาบเป็น $P_j = m_j^K - 1$ คาบสูงสุดที่เป็นไปได้ของตัวสร้างค่าตัวแปรสุ่มรวม (สมการที่ 7.9) คือ $P_1 P_2 \cdots P_K / 2^{J-1}$ ซึ่งมากกว่าของ MRG แต่ละตัวมาก ผู้ใช้ต้องกำหนดค่าต้นกำเนิด (Seed) ให้กับทุก ๆ MRG ด้วย จึงมีต้นกำเนิดทั้งหมด $J \times K$ ค่า ทั้ง MRG และตัวสร้าง

ค่าตัวแปรสุ่มรวมมีหลายสายเลขสุ่มซึ่งสามารถทำงานเป็นตัวสร้างค่าตัวแปรสุ่มเสมือนได้ ตัวอย่างที่ 7.2 มี MRG 2 ตัว ($J = 2$) โดยที่แต่ละตัวใช้เลขสุ่ม 3 ตัวล่าสุด ($K = 3$) เพื่อสร้างตัวต่อไป ตัวอย่างที่ 7.2 แสดงตัวสร้างค่าตัวแปรสุ่ม `MRG32k3a` ที่ใช้ในซอฟต์แวร์อารีน่าตั้งแต่เวอร์ชั่น 5 เป็นต้นไป

ตัวอย่างที่ 7.2. (`MRG32k3a` จาก L'Ecuyer 1999 หรือ Nelson 2013) `MRG32k3a` มีขั้นตอนวิธีดังนี้

$$\begin{aligned} \text{MRG ตัวที่ 1: } z_{i,1} &= (1{,}403{,}580 z_{i-2,1} - 810{,}728 z_{i-3,1}) \bmod (2^{32} - 209) \\ \text{MRG ตัวที่ 2: } z_{i,2} &= (527{,}612 z_{i-1,2} - 1{,}370{,}589 z_{i-3,2}) \bmod (2^{32} - 22{,}853) \\ z_i &= (z_{i,1} - z_{i,2}) \bmod (2^{32} - 209) \end{aligned}$$

ต้นกำเนิดของ MRG ตัวที่ 1 คือ $(z_{0,1}, z_{1,1}, z_{2,1})$ แต่ละตัวต้องเป็นจำนวนเต็มในช่วง $[0, m_1 - 1]$ และไม่เป็น 0 พร้อมกันทุกตัว เช่นเดียวกันสำหรับ MRG ตัวที่ 2 คือ $(z_{0,2}, z_{1,2}, z_{2,2})$ แต่ละตัวต้องเป็นจำนวนเต็มในช่วง $[0, m_2 - 1]$ และไม่เป็น 0 พร้อมกันทุกตัว การทำงานของตัวสร้างค่าตัวแปรสุ่มนี้ไม่ขึ้นกับชุดต้นกำเนิดที่เลือกใช้ คาบของ `MRG32k3a` ประมาณ $2^{191} \approx 3 \times 10^{57}$ ซึ่งยาวมาก ๆ □

หัวข้อต่อไปอธิบายถึงการเลือกใช้เลขสุ่มเทียมให้เหมาะสมในแบบจำลองสถานการณ์

7.1.4 การใช้เลขสุ่มเทียมที่เหมาะสม

หากต้องการประมวลผลแบบจำลองที่สถานะเดียว ไม่จำเป็นต้องกำหนดสายเลขสุ่ม แต่โดยมากแบบจำลองมักใช้เพื่อการเปรียบเทียบทางเลือกต่าง ๆ เช่น ในการปรับปรุงประสิทธิภาพระบบ ทางเลือกใหม่ถูกพิจารณาเทียบกับระบบปัจจุบัน สายเลขสุ่มมีไว้เพื่อให้การเปรียบเทียบทางเลือกด้วยแบบจำลอง (Comparison via simulation) ชัดเจนขึ้น คล้ายกับที่ให้นักวิ่งลงแข่งพร้อมกันเพื่อให้เผชิญสภาพสนาม เช่น ลม และอุณหภูมิแบบเดียวกัน หากทุกทางเลือกพบกับแหล่งความไม่แน่นอนเดียวกัน ความแตกต่างในเอาท์พุทน่าจะเกิดจากความแตกต่างในโครงสร้างของทางเลือกเป็นหลัก ไม่ใช่เพียงความแตกต่างที่เป็นค่าสุ่ม กรอบความคิดนี้มีชื่อว่า **การใช้เลขสุ่มร่วมกัน** (Common random number) ซึ่งกำหนดให้เลขสุ่มเทียมหนึ่ง ๆ ถูกใช้เพื่อจุดประสงค์เดียวกันในแบบจำลอง โดยกำหนดสายเลขสุ่มเฉพาะ (ผ่านทางต้นกำเนิดสายเลขสุ่ม) ให้แต่ละขบวนการสุ่ม เช่น การมาถึงของลูกค้ามีต้นกำเนิดเลขสุ่มเป็น 10 การให้บริการลูกค้ามีต้นกำเนิดเลขสุ่มเป็น 11 ซอฟต์แวร์อารีนาให้ผู้ใช้กำหนดต้นกำเนิดเองได้ โดยเพิ่มเป็นพารามิเตอร์สุดท้ายของฟังก์ชันสร้างการแจกแจง เช่น `NORM(`μ, σ`, stream number)`

มีผู้แย้งว่าต้นกำเนิดตัวสร้างค่าตัวแปรสุ่มควรเป็นค่าไม่แน่นอน ไม่ใช่ค่าคงที่ เพื่อให้เอาท์พุทที่ได้เป็นค่าสุ่มจริง ๆ ความคิดเช่นนี้มองข้ามวัตถุประสงค์ของการจำลองสถานการณ์ว่ากระทำเพื่อประมาณคุณสมบัติของระบบซึ่งมีความไม่แน่นอนอยู่ จึงต้องอาศัยการออกแบบการทดลองเพื่อให้ดัชนีชี้วัดที่ประมาณได้แม่นยำ

เพียงพอสำหรับการตัดสินใจ การใช้ตัวสร้างค่าตัวแปรสุ่มที่ให้เลขสุ่มซ้ำเดิมได้ช่วยให้การเปรียบเทียบทางเลือกชัดเจนขึ้น จึงให้ประโยชน์มากกว่าการสุ่มต้นกำเนิดสายตัวแปรสุ่ม

เลขสุ่มเทียมที่มีการแจกแจงแบบสม่ำเสมอมาตรฐาน (Uniform[0,1]) เป็นตัวตั้งต้นเพื่อสร้างตัวแปรสุ่มที่มีการแจกแจงอื่น ๆ ดังจะกล่าวถึงในหัวข้อถัดไป

7.2 การสร้างค่าตัวแปรสุ่ม

จุดประสงค์ของหัวข้อนี้เพื่อให้ผู้อ่านได้เข้าใจพื้นฐานและเห็นภาพขั้นตอนวิธีการสร้างค่าตัวแปรสุ่ม กำหนดให้เลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอมาตรฐาน แทนด้วยสัญลักษณ์ U หรือ $U \sim \text{unif}(0,1)$ ซึ่งใช้สร้างตัวแปรสุ่ม X ที่มีการแจกแจงอื่น ๆ สมมติว่าเป็น F_X แทนด้วยสัญลักษณ์ว่า $X \sim F_X$

7.2.1 วิธีแปรรูปฟังก์ชันผกผัน

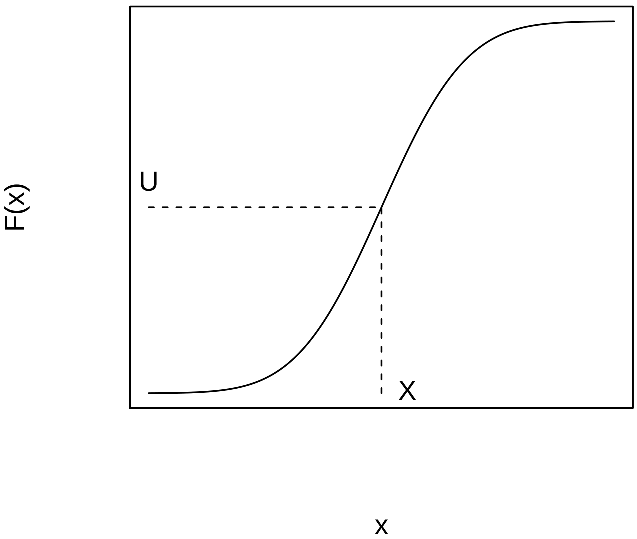
รูปที่ 7.2 แสดงกรอบความคิดของวิธีแปรรูปฟังก์ชันผกผัน (Inverse transform method) คือ แทนค่า $U \sim \text{unif}(0,1)$ เข้าไปในฟังก์ชันผกผันของความน่าจะเป็นสะสม (Inverse CDF) หรือ F_X^{-1} แล้วได้ค่าที่สังเกตได้ (Realization) ของตัวแปรสุ่ม X วิธีแปรรูปฟังก์ชั่นผกผันใช้ได้สำหรับตัวแปรสุ่มแบบต่อเนื่องและไม่ต่อเนื่อง แต่การพิสูจน์ที่จะแสดงต่อไปนี้ใช้ได้สำหรับ X ที่เป็นตัวแปรสุ่มแบบต่อเนื่อง โดยเริ่มจากฟังก์ชันการกระจายความน่าจะเป็นสะสมดังนี้

$$
\begin{aligned}
\Pr\{X \leqslant a\} &= \Pr\{F_X^{-1}(U) \leqslant a\} && (7.10)\\
&= \Pr\{U \leqslant F_X(a)\} && (7.11)\\
&= F_X(a) && (7.12)
\end{aligned}
$$

สมการที่ (7.10) มาจากวิธีที่สร้าง X มาจาก U สมการที่ (7.11) ใช้คุณสมบัติของฟังก์ชันผกผันและความไม่ลดลง (Non-decreasing) ของ CDF; สมการที่ (7.12) เป็นจริงเพราะ U มีการแจกแจงแบบสม่ำเสมอมาตรฐาน ตัวอย่างที่ 7.3 แสดงการหาฟังก์ชันผกผันของ CDF และการใช้วิธีแปรรูปฟังก์ชันผกผัน

ตัวอย่างที่ 7.3. ตัวแปรสุ่มแบบเอ็กซ์โพเนนเชียล พิจารณา CDF ของการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย $1/\lambda$ ดังนี้

$$F(x) = 1 - e^{-\lambda x}$$



รูปที่ 7.2: การสร้างตัวแปรสุ่มด้วยวิธีแปรรูปฟังก์ชันผกผัน

ก่อนใช้วิธีแปรรูปฟังก์ชันผกผัน ต้องคำนวณ $F^{-1}(u)$ ด้วยการเขียนสมการที่ x ในรูปของ u:

$$\begin{aligned} u &= 1 - e^{-\lambda x} \\ e^{-\lambda x} &= 1 - u \\ \ln(e^{-\lambda x}) &= \ln(1-u) \\ x &= -\frac{1}{\lambda}\ln(1-u) \end{aligned} \tag{7.13}$$

ขั้นตอนในการสร้างตัวแปรสุ่มที่มีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย $1/\lambda$ มีดังนี้

1. สร้างตัวแปรสุ่ม U ที่มีการแจกแจงแบบสม่ำเสมอมาตรฐาน: $U \sim$ unif$(0,1)$ อาจด้วยขั้นตอนวิธีในหัวข้อที่ 7.1.1-7.1.3 สมมติได้ $u = 0.1887$
2. แทนค่า U ในสมการที่ (7.13) เพื่อให้ได้ X ใช้ตัวอย่างต่อเนื่องจากด้านบนโดยสมมติว่า $\lambda = 1$ จะได้ว่า

$$x = -\ln(1 - 0.1887) = 0.2091$$

3. ให้ผล x คือ 0.2091

□

การแจกแจงอื่น ๆ ที่สามารถใช้วิธีแปรรูปฟังก์ชันผกผันได้คือการแจกแจงที่มี CDF ผกผันในรูปแบบปิด (Closed form) เช่น การแจกแจงแบบไวบูล (รายละเอียดของการแจกแจงในหัวข้อที่ 4.2.11), การ

แจกแจงแบบสม่ำเสมอ (หัวข้อที่ 4.2.13), การแจกแจงแบบสามเหลี่ยม (หัวข้อที่ 4.2.14), และการแจกแจงตามตัวอย่างแบบต่อเนื่องและไม่ต่อเนื่อง (หัวข้อที่ 4.2.16) ดังแสดงในตัวอย่างที่ 7.4

ตัวอย่างที่ 7.4. ตัวแปรสุ่มที่มีการแจกแจงตามตัวอย่างแบบต่อเนื่อง ฐานข้อมูลซ่อมบำรุงของบริษัทเก็บค่าระยะเวลาที่ใช้ในการซ่อมชิ้นส่วนชนิดหนึ่ง 100 ค่า ดังแสดงในตาราง 7.1 เนื่องจากตัวแปรนี้เป็นระยะเวลาจึงควรจำลองด้วยการแจกแจงแบบต่อเนื่อง แต่มีข้อมูลเพียง 4 ช่วงจึงไม่สามารถจำลองด้วยการแจกแจงแบบพาราเมตริกได้ จึงเลือกใช้การแจกแจงตามตัวอย่างแบบต่อเนื่อง ที่ใช้การประมาณค่าในช่วงเชิงเส้น (Linear Interpolation) เพื่อสุ่มค่าระหว่างขอบบนและขอบล่าง เส้นตรงของแต่ละช่วงมีความชันดังแสดงในสมการที่ (7.14)

$$a_i = \frac{\text{Upper bound}_i - \text{Lower bound}_i}{F(\text{Upper bound}_i) - F(\text{Lower bound}_i)} \tag{7.14}$$

เช่น ในตารางที่ 7.1 ความชันของช่วงที่ 2 คือ

$$a_2 = \frac{1 - 0.5}{0.41 - 0.31} = \frac{0.5}{0.1} = 5.0$$

เมื่อได้ความชันแล้ว จึงสร้างค่าตัวแปรสุ่ม ดังนี้

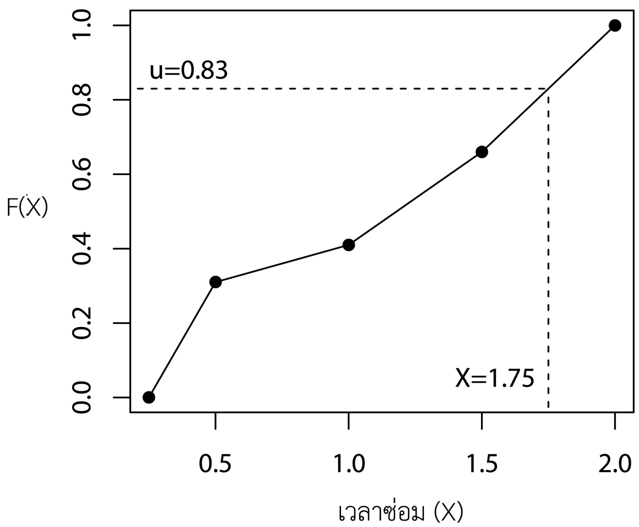
1. สร้างตัวแปรสุ่ม U ที่มีการแจกแจงแบบสม่ำเสมอมาตรฐาน: $U \sim \text{unif}(0, 1)$ สมมติได้ $u = 0.83$
2. หาช่วงของ CDF ที่ครอบค่า u นี้ คือ i ที่ $F(\text{Lower bound}_i) \leqslant u \leqslant F(\text{Upper bound}_i)$ เช่น $u = 0.83$ อยู่ในช่วงที่ 4 (ดูรูปที่ 7.3 ประกอบ)
3. แทนค่า u ในสมการที่ $\text{Lower bound}_i + a_i(u - F\text{Lower bound}_i))$ คือ $1.5 + 1.47 \times (0.83 - 0.66) = 1.75$
4. ให้ผล x คือ 1.75

ตารางที่ 7.1: ความถี่ของระยะเวลาซ่อมสำหรับตัวอย่างที่ 7.4

ช่วงที่ i	**ช่วงเวลา (ชั่วโมง)**	**ความถี่**	**ความถี่สัมพัทธ์**	**ความถี่สะสม**	**ความชัน** a_i **จากสูตร (7.14)**
1	$0.25 \leqslant x \leqslant 0.5$	31	31/100	0.31	0.81
2	$0.5 < x \leqslant 1.0$	10	10/100	0.41	5.0
3	$1.0 < x \leqslant 1.5$	25	25/100	0.66	2.0
4	$1.5 < x \leqslant 2.0$	34	34/100	1.00	1.47

□

จุดเด่นของวิธีแปรรูปฟังก์ชันผกผันคือเป็นวิธีที่ไม่สิ้นเปลือง เพราะใช้ตัวแปรสุ่ม U เพียงหนึ่งตัวสำหรับ



รูปที่ 7.3: การสร้างตัวแปรสุ่มแบบต่อเนื่องด้วยด้วยวิธีแปรรูปฟังก์ชันผกผันในตัวอย่างที่ 7.4

ตัวแปรสุ่มที่มีการแจกแจงที่ต้องการหนึ่งตัว ข้อด้อยคืออาศัยฟังก์ชันผกผันของ CDF สำหรับตัวแปรสุ่มแบบต่อเนื่อง ซึ่งการแจกแจงความน่าจะเป็นบางอันไม่มี เช่น การแจกแจงแบบปกติหรือแกมมา ตัวอย่างที่ 7.5 แสดงการใช้วิธีแปรรูปฟังก์ชันผกผันสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง

ตัวอย่างที่ 7.5. **ตัวแปรสุ่มแบบพหุนามหรือแบบตามตัวอย่างที่ไม่ต่อเนื่อง** สมมติว่าตัวแปรสุ่ม X มีค่าที่เป็นไปได้ r ค่า คือ $x_1 < x_2 < \ldots < x_r$ ด้วยความน่าจะเป็น

$$\Pr\{X = x_i\} = p_i, 1 \leqslant i \leqslant r$$

ซึ่งมี CDF ดังนี้

$$F_X(x_j) = \sum_{i=1}^{j} p_i$$

ด้วยวิธีแปรรูปฟังก์ชันผกผัน เมื่อสร้างค่า U แล้ว จึงนำไปค้นหาค่า x_i ซึ่งก็คือ $\min\{x_i : F_X(x_i) \geqslant U\}$

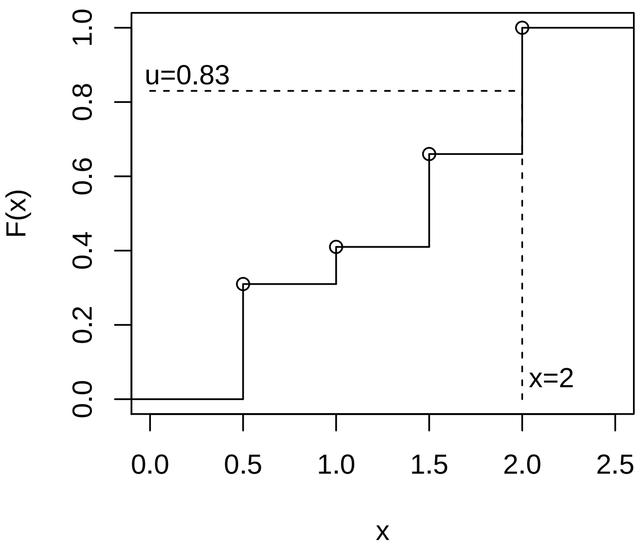
□

ตัวอย่างที่ 7.6. ต่อเนื่องจากตัวอย่างที่ 7.5 แต่ใช้ตัวเลขแทนที่สัญลักษณ์

รูปที่ 7.4 แสดงการสร้างตัวแปรสุ่มจากฟังก์ชันผกผันของ CDF สมมติว่า $u = 0.83$ ตัวแปรสุ่ม X ที่สร้างได้คือ 2 กำหนดตำแหน่งของค่า $u = 0.83$ บนแกนตั้งของกราฟในรูปที่ 7.4 แล้วลากตามแนวนอนจนกว่าจะชนกราฟ CDF เมื่อชนแล้ว ให้ลากลงมาตามแนวดิ่ง คืออ่านค่า x นั้น ๆ จะได้ 2 □

ตารางที่ 7.2: การแจกแจงตามตัวอย่างสำหรับตัวอย่างที่ 7.6

i	x_i	p_i	$F_X(x_i)$
1	0.5	0.31	0.31
2	1.0	0.10	0.41
3	1.5	0.25	0.66
4	2.0	0.34	1.00



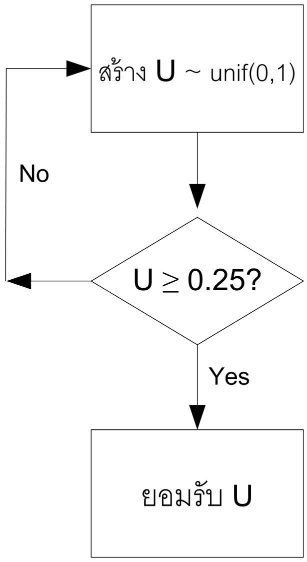
รูปที่ 7.4: การสร้างตัวแปรสุ่มแบบไม่ต่อเนื่องด้วยด้วยวิธีแปรรูปฟังก์ชันผกผันในตัวอย่างที่ 7.6

หากไม่สามารถใช้วิธีแปรรูปฟังก์ชันผกผันสำหรับการแจกแจงที่สนใจ มีวิธีอื่น เช่น วิธียอมรับ-ปฏิเสธในหัวข้อต่อไป

7.2.2 วิธียอมรับ-ปฏิเสธ

วิธีนี้เป็นประโยชน์โดยเฉพาะอย่างยิ่งในกรณีที่ไม่มี CDF ผกผันในรูปแบบสมการปิด ตัวอย่างของ CDF ผกผันในรูปแบบปิดคือสมการที่ (7.13) เรียกอีกอย่างหนึ่งว่าวิธีทำให้เหลือน้อยลง (Thining) คือ สร้างเลขสุ่มที่ยังไม่ใช่การแจกแจงที่ต้องการ แล้วใช้เงื่อนไขบางอย่างคัดให้เหลือเฉพาะค่าตัวแปรสุ่มที่มีการแจกแจงที่ต้องการ ตัวอย่างที่ 7.7 แสดงการประยุกต์ใช้กรอบความคิดนี้

ตัวอย่างที่ 7.7. วิธียอมรับ-ปฏิเสธ ตัวแปรสุ่มที่ต้องการมีการแจกแจงแบบสม่ำเสมอต่อเนื่องในช่วง $[0.25, 1]$ หรือ $X \sim$ unif$(0.25, 1)$ เลขสุ่มที่รู้วิธีสร้างคือ $U \sim$ unif$(0, 1)$ รูปที่ 7.5 แสดงการไหลของขั้นตอนวิธี เมื่อสร้างเลขสุ่ม U แล้วจึงตรวจสอบว่ามีค่าอย่างน้อย 0.25 หรือไม่ ถ้าใช่ก็*ยอมรับ*ค่านั้น ถ้าไม่ใช่ก็*ปฏิเสธ*ค่าที่สร้างมาแล้วกลับไปสร้างค่าใหม่



รูปที่ 7.5: การสร้างตัวแปรสุ่มแบบสม่ำเสมอต่อเนื่องด้วยด้วยวิธียอมรับ-ปฏิเสธในตัวอย่างที่ 7.7

□

ตัวอย่างที่ 7.7 แสดงให้เห็นชัดว่าประสิทธิภาพของวิธีนี้ขึ้นอยู่กับว่าปฏิเสธค่าเลขสุ่มที่สร้างถี่เพียงใด ถ้าปฏิเสธหลายครั้งจนกว่าจะได้ค่าสุ่มที่ต้องการค่าหนึ่ง วิธียอมรับ-ปฏิเสธไม่เหมาะสำหรับตัวแปรสุ่มนั้น การแจกแจงอื่น ๆ ที่ใช้วิธียอมรับ-ปฏิเสธได้คือตัวแปรสุ่มแบบปัวซองที่อัตราคงที่และที่อัตราไม่คงที่ (หัวข้อที่

4.2.4-4.2.5) และการแจกแจงเบต้า (หัวข้อที่ 4.2.4-4.2.5 อ่านรายละเอียดเพิ่มเติมใน Nelson 2013

วิธีสุดท้ายที่จะกล่าวถึงใช้คุณสมบัติเฉพาะของการแจกแจง

7.2.3 วิธีอาศัยคุณสมบัติเฉพาะ

วิธีนี้ใช้คุณสมบัติของการแจกแจง อาจเป็นในแง่ความสัมพันธ์ระหว่างการแจกแจงต่าง ๆ เช่น ตัวแปรสุ่มแบบปกติที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน $\sigma : X \sim \mathrm{N}(\mu, \sigma)$ สามารถแสดงในรูปตัวแปรสุ่มแบบมาตรฐาน ($Z \sim \mathrm{N}(0, 1)$) ได้ ดังนี้

$$X = \mu + \sigma Z$$

ดังนั้นถ้ารู้วิธีสร้างค่าตัวแปรสุ่ม Z ก็สามารถสร้างค่าตัวแปรสุ่ม X สำหรับ μ และ σ ใด ๆ ได้ นอกจากนี้แล้ว $Y = e^X$ ก็ยังมีการแจกแจงแบบล็อกนอร์มอล (หัวข้อที่ 4.2.7)

Nelson (2013) แนะนำว่าควรใช้คุณสมบัติพิเศษสร้างค่าตัวแปรสุ่มก็ต่อเมื่อเป็นวิธีที่ตรง (Exact) กล่าวคือ ให้ CDF ที่ต้องการในเชิงตัวเลขอย่างแม่นยำ

นอกจากวิธีสร้างค่าตัวแปรสุ่มที่ได้นำเสนอ 3 วิธีข้างต้น ยังมีวิธีอื่น ๆ อีก เช่น Convolution และ Composition ผู้อ่านสามารถศึกษาเพิ่มเติมใน Forbes et al. (2010) หรือ Law (2006) หากต้องการข้อมูลเพิ่มเติมหรือต้องการทราบวิธีที่เหมาะสมสำหรับการแจกแจงพาราเมตริกที่สนใจ

7.3 บทส่งท้าย

ผู้เขียนคาดว่าผู้อ่านตำราเล่มนี้ส่วนใหญ่ใช้ซอฟต์แวร์สำเร็จรูปเพื่อสร้างแบบจำลอง และคงไม่ได้พัฒนาแบบจำลองจากภาษาพื้นฐาน เช่น C++ หรือ Java จึงไม่ได้กล่าวถึงรายละเอียดด้านการพัฒนาคำสั่งสร้างค่าตัวแปรสุ่มจากภาษาพื้นฐานมากนัก ทั้งนี้หากผู้อ่านมีความสามารถในการพัฒนาแบบจำลองด้วยภาษาพื้นฐานแหล่งอ้างอิงที่แนะนำมีดังนี้

Barry Nelson พัฒนาชุดคำสั่งบนภาษา VBA และ MATLAB เพื่อประกอบหนังสือ Nelson (2013) `http://users.iems.northwestern.edu/~nelsonb/IEMS435/index.html`

Pierre L'Ecuyer พัฒนาชุดคำสั่ง SSJ สำหรับการพัฒนาแบบจำลอง (Simulation library) ด้วยภาษา Java `http://www.iro.umontreal.ca/~simardr/ssj/indexe.html`

Standard Template Library (STL) เป็นชุดคำสั่งสำหรับภาษา C++ สำหรับการคำนวณทางคณิตศาสตร์ สถิติและความน่าจะเป็น `http://www.sgi.com/tech/stl/`

แบบฝึกหัด

7.1. พิจารณาตัวสร้างค่าตัวแปรสุ่มดังต่อไปนี้

$$X_{i+1} = (16807X_i) \mod (2^{31} - 1)$$

ถ้า $X_0 = 543210$ ให้คำนวณค่า X_2

7.2. ในฤดูการแข่งขันหนึ่ง ทีมฟุตบอล มก.ยูไนเต็ดชนะด้วยความน่าจะเป็น 0.75 สมมติว่าหากทีมนี้ชนะในเกมที่ i ลูกทีมจะมั่นใจ ซึ่งทำให้ชนะในเกมถัดไป $i+1$ ด้วยความน่าจะเป็น 0.85 แต่ถ้าแพ้ในเกมที่ i ลูกทีมจะเสียกำลังใจ ซึ่งทำให้แพ้ในเกมถัดไป $i+1$ ด้วยความน่าจะเป็น 0.5

(i) ผลแพ้ชนะของแต่ละการแข่งขันเป็นอิสระต่อกันหรือไม่ ให้อธิบาย

(ii) ผู้อ่านสามารถจำลองผลแพ้ชนะของการแข่งขันได้ดังนี้ สมมติว่า p_i เป็นความน่าจะเป็นที่ทีมจะชนะในเกมที่ i และ $U_1, U_2, \ldots$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันที่มีการแจกแจงแบบสม่ำเสมอในช่วง [0, 1] ที่สามารถใช้วิธีในหัวข้อที่ 7.1.1-7.1.3 สร้างได้ แบบจำลองสถานการณ์ของเราจะให้ผลเป็นชนะถ้า $U_i < p_i$ ให้ใช้เลขสุ่มแบบสม่ำเสมอมาตรฐานต่อไปนี้สำหรับจำลองผลการแข่งขัน 10 เกมแรก

0.874 0.168 0.958 0.674 0.374 0.656 0.773 0.203 0.142 0.139

(iii) ตัวแบบนี้สมเหตุสมผลหรือไม่ ให้อธิบาย

7.3. ให้ใช้เอกซ์เซลหรือโปรแกรมคณิตศาสตร์อื่นเพื่อสร้าง $U \sim$ unif$(0,1)$ จากตัวสร้างตัวแปรสุ่ม MCG ต่อไปนี้

$$\begin{aligned} z_i &= (11z_{i-1}) \mod 16, z_0 = 2, \\ u_i &= \frac{z_i}{16} \end{aligned}$$

จำนวน 1000 ค่า แล้วสร้างฮีสโตแกรม

7.4. กำหนดให้ PDF ของตัวแปรสุ่ม X เป็นดังนี้

$$f_X(x) = 2x, 0 \leqslant x \leqslant 1$$

(i) ระบุ CDF $(F_X(x))$

(ii) ระบุฟังก์ชันผกผันของ CDF $(F_X^{-1}(u))$

(iii) ใช้วิธีแปรรูปฟังก์ชันผกผันสร้าง X จากตัวแปรสุ่ม $0.2031, 0.8819, 0.5506$

7.5. จำนวนเที่ยวขนส่งที่ท่าบรรทุกของเป็นค่า 0, 1 หรือ 2 เที่ยว

x	$p(x)$
0	0.5
1	0.3
2	0.2

(i) ให้สร้างกราฟ CDF (ดังเช่นในรูปที่ 7.4)

(ii) สร้าง X จากตัวแปรสุ่ม $0.2031, 0.8819, 0.5506$

7.6. กำหนดให้มีตัวสร้างเลขสุ่มแบบสม่ำเสมอมาตรฐาน $U \sim$ unif$(0, 1)$ ให้แสดงวิธีสร้างค่าเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง $[a, b]$ จาก U นี้ สำหรับ $a < b$ และ a และ b ใด ๆ

8

การเปรียบเทียบทางเลือก

การวิเคราะห์ด้วยการจำลองสถานการณ์มักทำเพื่อเปรียบเทียบตัวเลือกหรือสถานะของระบบ 2 ทางหรือมากกว่า เช่น ในโครงการปรับปรุงประสิทธิภาพการให้บริการ ระบบใหม่ที่ปรับปรุงแล้วหลาย ๆ ตัวเลือกถูกเปรียบเทียบกับระบบปัจจุบันที่เป็นอยู่ อาจเป็นในแง่ระดับการให้บริการซึ่งนิยามในมิติของเวลารอเฉลี่ยจนกว่าจะได้รับการบริการ หรือความน่าจะเป็นที่ลูกค้าจะได้รับบริการภายในระยะเวลาเป้าหมาย เนื่องจากแบบจำลองมีความไม่แน่นอน ค่าเอาท์พุทของดัชนีชี้วัดจึงเป็นค่าสุ่ม และต้องอาศัยการวิเคราะห์ทางสถิติเพื่อสรุปว่าความแตกต่างของเอาท์พุทนี้มีนัยสำคัญทางสถิติ (Statistical significance) หรือไม่ หรือเป็นเพียงเสียงรบกวนที่เกิดจากความไม่แน่นอน โดยทั่วไปแบบจำลองประมาณความแตกต่างในดัชนีชี้วัดได้ดีกว่าประมาณค่าดัชนีชี้วัดของระบบเดียวโดด ๆ (Absolute performance) เพราะแบบจำลองของทุกตัวเลือกอยู่ภายใต้ข้อสมมติชุดเดียวกัน แบบจำลองสถานการณ์ก็เป็นเช่นนั้น

ในเชิงปฏิบัติ ย่อมมีดัชนีชี้วัดที่สนใจหลายตัว บทนี้พิจารณาเฉพาะการเปรียบเทียบทางเลือกด้วยเกณฑ์ดัชนีชี้วัดเพียงตัวเดียว เช่น เวลาครบรอบ (Cycle time) หรือค่าใช้จ่ายรวม บทนี้แบ่งออกเป็น 2 ส่วนหลัก หัวข้อที่ 8.1 กล่าวถึงการเปรียบเทียบ 2 ทางเลือกซึ่งซับซ้อนน้อยที่สุด และหัวข้อที่ 8.2 ครอบคลุมกรณีที่มีมากกว่า 2 ทางเลือก

8.1 การเปรียบเทียบ 2 ทางเลือก

เพื่อให้ชัดเจนขึ้น สมมติว่าการวิเคราะห์ด้วยแบบจำลองเปรียบเทียบ 2 ทางเลือก คือ ดีไซน์ที่ 1 และดีไซน์ที่ 2 ในแง่เวลานำในการส่งมอบของดีไซน์ที่ 1 (θ_1) เมื่อเทียบกับของดีไซน์ที่ 2 (θ_2) การทดลองด้วยการจำลองสถานการณ์ (Simulation experiment) สามารถทำได้ 2 ลักษณะ ดังนี้

1. **สุ่มอย่างเป็นอิสระต่อกัน** (Independent sampling) เลขสุ่มที่ใช้ในแบบจำลองของดีไซน์ที่ 1 และดีไซน์ที่ 2 เป็นอิสระต่อกัน

2. **ใช้เลขสุ่มร่วมกัน** (Common random number, CRN) คือ เลขสุ่มเดียวกันถูกนำไปใช้จำลองสถานการณ์ของดีไซน์ทั้งสองแบบ หากทำอย่างถูกต้อง วิธีนี้สามารถทำให้การเปรียบเทียบทางเลือกคมชัดกว่าการสุ่มอย่างเป็นอิสระต่อกัน

หัวข้อที่ 8.1.1 กล่าวถึงการเปรียบเทียบโดยให้แต่ละทางเลือกประสบกับสถานการณ์ใกล้เคียงกันเท่าที่จะทำได้

8.1.1 การใช้เลขสุ่มร่วมกัน

ภาพรวมของวิธีนี้คือให้แต่ละทางเลือกเผชิญกับแหล่งความไม่แน่นอนเดียวกัน (เช่น อุปสงค์ของสินค้า เวลาในการให้บริการ เวลาจัดส่ง จำนวนเครื่องจักรที่เสีย) การใช้เลขสุ่มร่วมกันมีนัยว่าความแตกต่างในการดำเนิน

งานของระบบที่มีเกณฑ์เป็นดัชนีชี้วัด เช่น ระดับการให้บริการ (Service level) หรือปริมาณผลผลิตต่อวัน (Throughput) มาจากความแตกต่างในองค์ประกอบของทางเลือกเป็นหลัก ไม่ใช่แตกต่างเพราะตัวแปรนำเข้าที่ไม่แน่นอน

หากพิจารณาวิธีการใช้เลขสุ่มร่วมกันในแง่คณิตศาสตร์ ความแตกต่างของดัชนีชี้วัดที่ต้องการคือ $\theta_1 - \theta_2$ ซึ่งใช้แบบจำลองประมาณด้วยส่วนต่างของค่าเฉลี่ยจากตัวอย่างคือ $\bar{Y}_1 - \bar{Y}_2$ และระบุค่าความผิดพลาดในการประมาณด้วยช่วงความเชื่อมั่นซึ่งอาศัย $\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)$ ที่แยกพจน์ได้ดังนี้

$$\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2) = \mathrm{Var}(\bar{Y}_1) + \mathrm{Var}(\bar{Y}_2) - 2\mathrm{Cov}(\bar{Y}_1, \bar{Y}_2) \tag{8.1}$$

หากประมวลผลดีไซน์ 1 และ 2 อย่างเป็นอิสระต่อกัน จะได้ $\mathrm{Cov}(\bar{Y}_1, \bar{Y}_2) = 0$ นั่นคือ

$$\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)_{\text{ind}} = \mathrm{Var}(\bar{Y}_1) + \mathrm{Var}(\bar{Y}_2) \tag{8.2}$$

แต่หากทำเทคนิคใช้เลขสุ่มร่วมกันแล้วได้ผล ค่าความแปรปรวนร่วมจะเป็นบวก ซึ่งจากสมการที่ (8.1) จะทำให้

$$\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)_{\text{crn}} < \mathrm{Var}(\bar{Y}_1) + \mathrm{Var}(\bar{Y}_2) \tag{8.3}$$

จากสมการที่ (8.2) และ (8.3) จะได้ว่า

$$\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)_{\text{crn}} < \mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)_{\text{ind}}$$

หมายความว่าเมื่อใช้ตัวแปรสุ่มร่วมกัน ช่วงความเชื่อมั่นของ $(\theta_1 - \theta_2)$ จะ**แคบกว่า**เมื่อสุ่มอย่างเป็นอิสระต่อกัน ทำให้สามารถสรุปได้ง่ายขึ้นว่า $(\theta_1 - \theta_2)$ แตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ เช่น สมมติว่าเมื่อใช้เลขสุ่มร่วมกัน ช่วงความเชื่อมั่นที่ได้คือ $4 \pm 1 = [3, 5]$ นาที ซึ่งแสดงว่าดีไซน์ 1 ต่างจากดีไซน์ 2 อย่างมีนัยสำคัญ แต่เมื่อสุ่มอย่างเป็นอิสระต่อกันได้ช่วงความเชื่อมั่น $4 \pm 5 = [-1, 9]$ นาที ซึ่งคร่อมศูนย์ หมายความว่า $(\theta_1 - \theta_2)$ เป็นไปได้ทั้งบวกและลบ จึงสรุปได้เพียงว่าดีไซน์ 1 และ 2 <u>ไม่</u>ต่างกันอย่างมีนัยสำคัญ ตัวอย่างที่ 8.1 แสดงการประยุกต์ใช้วิธีการใช้เลขสุ่มร่วมกัน

ตัวอย่างที่ 8.1. (ดัดแปลงจากตัวอย่างที่ 12.1 ใน Banks et al. 2005) ผู้บริหารศูนย์ให้บริการลูกค้าทางโทรศัพท์พิจารณาการออกแบบระบบ 2 แบบ ดังนี้

ตัวเลือกที่ 1 ระบบปัจจุบันที่มีพนักงานรับโทรศัพท์ 4 คนสำหรับแผนกลูกค้าในกรุงเทพและอีก 3 คนสำหรับแผนกลูกค้าต่างจังหวัด

ตัวเลือกที่ 2 ควบรวม 2 แผนก โดยมีพนักงาน 7 คนสามารถรับสายของลูกค้าได้ทั้งสองแบบ

ดัชนีชี้วัดที่สนใจคือเวลาที่ลูกค้าอยู่ในระบบ กล่าวคือ

- θ_1 คือ ค่าคาดหมายของเวลารอเฉลี่ยในสายของลูกค้าทั้งสองชนิดใน**ระบบปัจจุบัน**
- θ_2 คือ ค่าคาดหมายของเวลารอเฉลี่ยในสายของลูกค้าทั้งสองชนิดใน**ระบบที่นำเสนอ**

ขบวนการสุ่มที่เป็นค่านำเข้ามี 4 อัน ได้แก่ เวลาระหว่างที่ลูกค้าโทรเข้า, ชนิดของลูกค้า (กรุงเทพหรือต่างจังหวัด), เวลาสนทนาของลูกค้ากรุงเทพ, และเวลาสนทนาของลูกค้าต่างจังหวัด การใช้เลขสุ่มร่วมกันทำได้โดยกำหนดสายเลขสุ่มเฉพาะด้วยการระบุจุดกำเนิด (Seed) ของสายเลขสุ่มสำหรับแต่ละขบวนการนำเข้า ผู้อ่านอาจศึกษาวิธีการกำหนดจุดกำเนิดจากคู่มือการใช้ซอฟต์แวร์ที่ใช้อยู่

□

การเปรียบเทียบทำเพื่อให้เห็นความแตกต่าง ซึ่งต้องการคำอธิบายเพิ่มเติม

8.1.2 ความแตกต่างระหว่างตัวเลือก

ความแตกต่างระหว่างตัวเลือกมี 2 ชนิด คือ

ความแตกต่างในเชิงปฏิบัติ (Practical difference) ความแตกต่างที่มีความหมายในเชิงปฏิบัติขึ้นอยู่กับปัจเจกบุคคล ที่เป็นผู้ตัดสินใจว่าความแตกต่างเพียงใดมีขนาดใหญ่พอที่คุ้มค่าความเปลี่ยนแปลง เช่น ลดรอบเวลาการทำงาน (Cycle time) ได้ 10 นาที, เพิ่มผลตอบแทนพอร์ตการลงทุน 20,000 บาท, หรือลดจำนวนลูกค้าที่ไม่สามารถล็อกอินเข้าระบบออนไลน์ได้ 100 คน

ความแตกต่างทางสถิติ (Statistical difference) ขึ้นอยู่กับความแปรปรวนของเอาท์พุทที่เป็นดัชนีชี้วัด สมมติว่าพิจารณาช่วงความเชื่อมั่นของ $\theta_1 - \theta_2$ ผลที่เป็นไปได้มี 3 กรณี ดังแสดงในรูปที่ 8.1: กรณีที่ 8.1ก มีความเป็นไปได้สูงที่ $\theta_1 < \theta_2$, กรณีที่ 8.1ข มีความเป็นไปได้สูงที่ $\theta_1 > \theta_2$, หรือกรณีที่ 8.1ค เอาท์พุทที่มียังไม่เพียงพอที่จะสรุปได้ จึงต้องประมวลผลแบบจำลองเพิ่ม

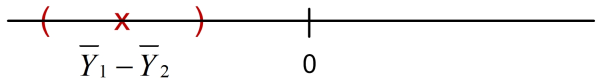
หัวข้อต่อไปกล่าวถึงช่วงความเชื่อมั่น 2 แบบ ขึ้นอยู่กับว่าเอาท์พุทจากแบบจำลองของตัวเลือกที่ 1 และ 2 เป็นอิสระต่อกันหรือใช้เลขสุ่มร่วมกัน

8.1.3 ช่วงความเชื่อมั่นของส่วนต่างของค่าเฉลี่ย

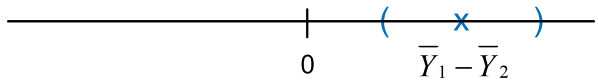
กรณีที่เป็นอิสระต่อกัน ใช้ช่วงความเชื่อมั่นแบบ t สำหรับตัวอย่างสองชุด แต่ถ้าใช้เลขสุ่มร่วมกัน จะใช้ช่วงความเชื่อมั่นแบบ t สำหรับคู่ตัวอย่าง

ช่วงความเชื่อมั่นแบบ t สำหรับตัวอย่างสองชุด (Two-sample t confidence interval)

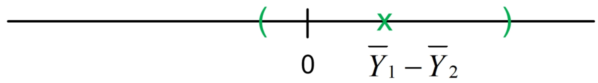
เมื่อประมวลผลแบบจำลองของตัวเลือกที่ 1 และตัวเลือกที่ 2 อย่างเป็นอิสระต่อกัน เอาท์พุทที่เป็นดัชนีชี้วัดจึงเป็นอิสระต่อกัน สมมติให้เอาท์พุทจากแบบจำลองของตัวเลือกที่ 1 คือ Y_{j1} มีเอาท์พุท R_1 ค่า และของตัว



(ก) ช่วงความเชื่อมั่น**น้อยกว่า**ศูนย์



(ข) ช่วงความเชื่อมั่น**มากกว่า**ศูนย์



(ค) ช่วงความเชื่อมั่น**คร่อม**ศูนย์

รูปที่ 8.1: กรณีที่เป็นไปได้สำหรับช่วงความเชื่อมั่นของ $\theta_1 - \theta_2$

เลือกที่ 2 คือ Y_{j2} จำนวน R_2 ค่า สามารถสร้างช่วงความเชื่อมั่นสำหรับ $\theta_1 - \theta_2$ โดยมีลำดับการคำนวณดังนี้

1. **ค่าเฉลี่ยและความแปรปรวนของแต่ละชุดตัวอย่าง** สำหรับเอาท์พุทของตัวเลือกที่ 1

$$\begin{aligned} \bar{Y}_{\cdot 1} &= \frac{1}{R_1}\sum_{j=1}^{R_1} Y_{j1} \\ S_1^2 &= \frac{1}{R_1 - 1}\sum_{j=1}^{R_1}(Y_{j1} - \bar{Y}_{\cdot 1})^2 \end{aligned} \tag{8.4}$$

   ทำซ้ำสำหรับเอาท์พุทของตัวเลือกที่ 2 ได้ค่า $\bar{Y}_{\cdot 2}$ และ S_2^2 ซอฟต์แวร์อารีน่าไม่รายงาน S^2 แต่ให้ครึ่งช่วงความเชื่อมั่นที่นัยสำคัญ 95% ผู้วิเคราะห์สามารถคำนวณ S^2 เองโดยใช้สมการที่ (5.13)

2. **ค่าความคลาดเคลื่อนมาตรฐาน** (Standard error)

$$\text{s.e.}(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = \sqrt{\frac{S_1^2}{R_1} + \frac{S_2^2}{R_2}} \tag{8.5}$$

3. **องศาอิสระ** (Degree of freedom), ν, ดังนี้

$$\nu \approx \frac{(S_1^2/R_1 + S_2^2/R_2)^2}{\frac{(S_1^2/R_1)^2}{R_1 - 1} + \frac{(S_2^2/R_2)^2}{R_2 - 1}} \tag{8.6}$$

   ใช้สัญลักษณ์ $\approx$ เพราะสมการที่ (8.6) เป็นค่าประมาณที่มักเรียกว่า**การประมาณของเวลซ์** (Welch's approximation, Welch 1947) ถ้าค่าที่คำนวณได้ไม่เป็นจำนวนเต็มให้ปัดลง

4. **ค่าสถิติ** t สำหรับนัยสำคัญ $1 - \alpha$ และที่องศาอิสระ ν ใช้ตารางค่าสถิติ t เพื่อสืบค้น $t_{1-\alpha/2,\nu}$ หรือใช้สูตรเอกซ์เซล `=tinv(`α, ν`)`

5. **ช่วงความเชื่อมั่น** แทนค่าสมการลงในสูตรนี้

$$\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2} \pm t_{1-\alpha/2,\nu}\text{s.e.}(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) \tag{8.7}$$

ช่วงความเชื่อมั่นที่คำนวณด้วยสมการที่ (8.4)-(8.7) ตั้งอยู่บนข้อสมมติที่ว่าเอาท์พุทมีการแจกแจงแบบปกติที่มีความแปรปรวนไม่เท่ากันและไม่ทราบค่า (Unknown and unequal variances) ตำราเล่มนี้พิจารณาแค่กรณีนี้ เพราะไม่มีเหตุผลใดชวนให้เชื่อว่าเอาท์พุทจาก 2 ตัวเลือกมีความแปรปรวนเท่ากันหรือทราบค่าความแปรปรวน หนังสือสถิติพื้นฐานทั่วไป เช่น Montgomery and Runger (2002) กล่าวถึงกรณีอื่น ๆ ด้วย เช่น ความแปรปรวนเท่ากันแต่ไม่ทราบค่า (Unknown and equal variances) ซึ่งสามารถใช้ค่าความแปรปรวนร่วม (Pooled variance) แทนสมการที่ (8.5) และสูตรอื่นสำหรับองศาอิสระ

ตัวอย่างที่ 8.2 แสดงการใช้สมการที่ (8.4)-(8.7)

ตัวอย่างที่ 8.2. (ดัดแปลงจากตัวอย่างที่ 9.7 ใน Devore 2011) แบบจำลองของสายการบรรจุขนมลงกล่องถูกพัฒนาขึ้นเพื่อประมาณค่าเฉลี่ยของจำนวนกล่องที่บรรจุได้ต่อวัน เปรียบเทียบระบบปัจจุบัน (ดีไซน์ 1) และระบบที่ปรับปรุงแล้ว (ดีไซน์ 2) วัตถุประสงค์ของการวิเคราะห์ก็เพื่อประเมินว่าระบบที่ปรับปรุงแล้วให้ปริมาณงาน (Throughput) ที่สูงกว่าระบบปัจจุบันหรือไม่ ตารางที่ 8.1 แสดงค่าสรุปทางสถิติที่ได้จากแบบจำลอง

ตารางที่ 8.1: ค่าสรุปทางสถิติสำหรับตัวอย่างที่ 8.2

ชนิด (i)	**จำนวนเอาท์พุท** (R_i)	**ค่าเฉลี่ย** ($\bar{Y}_{\cdot i}$)	**ส่วนเบี่ยงเบนมาตรฐาน** (S_i)
ระบบปัจจุบัน (1)	10	2902.8	277.3
ระบบที่ปรับปรุงแล้ว (2)	8	3108.1	205.9

คำนวณค่าความคลาดเคลื่อนมาตรฐานด้วยสมการที่ (8.5) ได้ดังนี้

$$\text{s.e.}(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}) = \sqrt{\frac{277.3^2}{10} + \frac{205.9^2}{8}} = 113.96$$

ใช้สมการที่ (8.6) เพื่อคำนวณองศาอิสระ ได้ดังนี้

$$\nu \approx \frac{(277.3^2/10 + 205.9^2/8)^2}{\frac{(277.3^2/10)^2}{10-1} + \frac{(205.9^2/8)^2}{8-1}} = 15.94$$

ปัดลงก็จะได้ $\nu = 15$ เลือกนัยสำคัญ 95% หรือ $\alpha = 0.05$ แล้วกำหนดค่าสถิติ t ด้วยฟังก์ชันเอกซ์เซล `tinv(0.05, 15) = 2.131` แทนค่าความคลาดเคลื่อนมาตรฐานและ t ในสมการที่ (8.7) ได้ดังนี้

$$(2902.8 - 3108.1) \pm (2.131 \times 113.96) = -205.3 \pm 242.92 = [-448.22, 37.62]$$

สังเกตว่าครึ่งช่วงความเชื่อมั่น (Half width) มีขนาดมากกว่าส่วนต่างของค่าเฉลี่ย ทำให้ช่วงความเชื่อมั่นครอบศูนย์ (กรณีในรูปที่ 8.1ค) จึงไม่สามารถสรุปได้ว่าระบบที่ปรับปรุงแล้วดีกว่าระบบปัจจุบันอย่างมีนัยสำคัญ ถึงแม้ว่าค่าเฉลี่ยของปริมาณงานของระบบปรับปรุงมากกว่าระบบปัจจุบันก็ตาม ตัวอย่างนี้แสดงให้เห็นว่าการพิจารณาเพียงค่าเฉลี่ยไม่เพียงพอ ต้องพิจารณาค่าคลาดเคลื่อนมาตรฐานประกอบด้วย

□

เมื่อใช้ตัวแปรสุ่มร่วมกัน (CRN ในหัวข้อที่ 8.1.1) เอาท์พุทที่ประมวลผลได้ (Y_{j1} และ Y_{j2}) น่าจะมีความสัมพันธ์กันในเชิงบวก (Positive correlation) จึงจะใช้ช่วงความเชื่อมั่นแบบ t สำหรับคู่ตัวอย่าง

ช่วงความเชื่อมั่นแบบ t สำหรับคู่ตัวอย่าง (Paired t confidence interval)

หากมีเอาท์พุทของแต่ละทางเลือก $Y_{ji}, 1 \leqslant i \leqslant 2$, สามารถนำมาคำนวณส่วนต่าง $D_j = Y_{j1} - Y_{j2}$ แล้วพิจารณา D_j นี้แทนเอาท์พุทของแต่ละทางเลือก โดยมีขั้นตอนดังนี้

1. คำนวณส่วนต่าง $D_j = Y_{j1} - Y_{j2}, 1 \leqslant j \leqslant R,$ โดยที่ R เป็นขนาดตัวอย่างของตัวเลือกที่ 1 และตัวเลือกที่ 2
2. คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ D ได้ $\bar{D}$ และ S_D^2
3. ช่วงความเชื่อมั่นของ D ที่นัยสำคัญ $(1-\alpha)$ คือ

$$\bar{D} \pm t_{1-\alpha/2,R-1}\frac{S_D}{\sqrt{R}} \tag{8.8}$$

โดยปริยาย ซอฟต์แวร์สำเร็จรูปอาจไม่รายงานค่าเอาท์พุททีละค่า แต่ให้เป็นค่าสรุปทางสถิติ เช่น ค่าเฉลี่ยและครึ่งช่วงความเชื่อมั่น หากต้องการค่าเอาท์พุท Y_{ji} เป็นค่าแต่ละค่า อาจต้องใช้คำสั่งหรือเพิ่มโมดูล เช่น `ReadWrite` ในอารีน่าซึ่งเขียนค่าตัวแปรลงในไฟล์แบบ `text`

ตัวอย่างที่ 8.3 แสดงการคำนวณช่วงความเชื่อมั่นแบบ t สำหรับคู่ตัวอย่าง

ตัวอย่างที่ 8.3. (ดัดแปลงจากตัวอย่างที่ 9.9 ใน Devore 2011) แผนกตัดเย็บที่โรงงานแห่งหนึ่งมีจำนวนงานไม่ผ่านการตรวจสอบคุณภาพสูงกว่าเป้าหมายของบริษัท วิศวกรอุตสาหการประจำโรงงานเสนอแนวทางปรับปรุง ซึ่งต้องใช้เงินลงทุนมากกว่าที่ผู้บริหารตั้งงบประมาณไว้ การขออนุมัติเป็นกรณีพิเศษอาศัยผลการวิเคราะห์เชิงตัวเลข วิศวกรอุตสาหการจึงพัฒนาแบบจำลองของแผนกนี้ ที่เป็นอยู่ในปัจจุบันและที่นำเสนอ ดัชนีชี้วัดที่เป็นเอาท์พุทคือจำนวนงานเสียต่อกะ ดังแสดงในตารางที่ 8.2

สามารถคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของส่วนต่างได้ดังนี้ $\bar{D} = 6.75$ และ $S_D = 8.234$ ค่า

ตารางที่ 8.2: ค่าสรุปทางสถิติสำหรับตัวอย่างที่ 8.3

ลำดับที่ (j)	1	2	3	4	5	6	7	8
ระบบปัจจุบัน (Y_{j1})	81	87	86	82	90	86	96	73
ระบบที่นำเสนอ (Y_{j2})	78	91	78	78	84	67	92	70
ส่วนต่าง (D_j)	3	-4	8	4	6	19	4	3

ลำดับที่ (j)	9	10	11	12	13	14	15	16
ระบบปัจจุบัน (Y_{j1})	74	75	72	80	66	72	56	82
ระบบที่นำเสนอ (Y_{j2})	58	62	70	58	66	60	65	73
ส่วนต่าง (D_j)	16	13	2	22	0	12	-9	9

ทางสถิติ $t_{0.975,15} = 2.131$ แทนค่าในสูตร (8.8) ได้ดังนี้

$$6.75 \pm 2.131 \left(\frac{8.234}{\sqrt{16}} \right) = 6.75 \pm 4.387 = [2.36, 11.14] \text{ ชิ้นต่อวัน}$$

หลังปรับปรุงคุณภาพ จำนวนชิ้นงานที่เสียลดลงอย่างมีนัยสำคัญ คือ ลดลงมากกว่า 2.36 ชิ้นและไม่เกิน 11.14 ชิ้นต่อวันด้วยความน่าจะเป็น 95% ขั้นต่อไปของการวิเคราะห์คือการคำนวณจุดคุ้มทุนว่าจำนวนชิ้นงานเสียที่ลดลงนี้จะคุ้มค่าการลงทุนภายในระยะเวลาที่ยอมรับได้หรือไม่

□

ในหัวข้อที่ 5.3.5 เมื่อพิจารณา 1 ระบบผู้วิเคราะห์กำหนดจำนวนรอบทำซ้ำที่ขึ้นอยู่กับความแม่นยำที่ต้องการ เช่นเดียวกันเมื่อสร้างความเชื่อมั่นสำหรับส่วนต่าง $\theta_1 - \theta_2$ จำนวนรอบทำซ้ำที่ควรใช้ก็ขึ้นอยู่กับระดับของส่วนต่างที่ต้องการแยกแยะ

การกำหนดจำนวนรอบทำซ้ำ

สมมติว่าผู้วิเคราะห์ต้องการแยกแยะส่วนต่างมากกว่า $\pm\varepsilon$ หลังจากทดลองครั้งแรกด้วย R_0 รอบทำซ้ำ แล้วได้ค่าส่วนเบี่ยงเบนมาตรฐานของส่วนต่าง S_D จำนวนรอบทำซ้ำที่ต้องทำทั้งหมดคือค่า R ที่ทำให้อสมการที่ (8.9) ข้างล่างนี้เป็นจริง (สังเกตว่ามี R อยู่ทั้งสองข้างของสมการ ด้านขวามืออยู่ที่พจน์ t) ค้นหาค่า R ได้ง่าย ๆ โดยการลองผิดลองถูก (Trial and error) บนเอกซ์เซล

$$R \geqslant \left(\frac{t_{1-\alpha/2,R-1} S_D}{\varepsilon} \right)^2 \tag{8.9}$$

หรืออาจจะใช้สูตรประมาณ (8.10) เป็นค่าตั้งต้นสำหรับการลองผิดลองถูก ดังนี้

$$R \geqslant \left(\frac{z_{1-\alpha/2}S_D}{\varepsilon}\right)^2 \tag{8.10}$$

เมื่อ $z_{1-\alpha/2}$ คือควอนไทล์ของการแจกแจงแบบปกติมาตรฐาน $Z \sim \mathrm{N}(0,1)$ ที่ทำให้สมการต่อไปนี้เป็นจริง

$$\Pr\{-z_{1-\alpha/2} \leqslant Z \leqslant z_{1-\alpha/2}\} = 1-\alpha$$

หากใช้ซอฟต์แวร์สำเร็จรูปและไม่มีค่า S_D แต่มีส่วนเบี่ยงเบนมาตรฐานของค่าประมาณดัชนีชี้วัดของแต่ละระบบ (S_1 และ S_2) สามารถประมาณจำนวนรอบทำซ้ำที่ควรใช้จากสูตรต่อไปนี้ ซึ่งอาศัยการประมาณ $S_D^2 \sim S_1^2 + S_2^2$

$$R \geqslant \frac{t^2_{1-\alpha/2,2(R_0-1)}(S_1^2+S_2^2)}{\varepsilon^2} \tag{8.11}$$

สังเกตว่าสูตรนี้อยู่บนข้อสมมติว่าทั้งสองทางเลือกมีขนาดตัวอย่างเท่ากันคือ R_0

ตัวอย่างที่ 8.4 แสดงการคำนวณจำนวนรอบทำซ้ำที่ควรใช้เพื่อให้สามารถแยกแยะทางเลือกได้ตามความแม่นยำที่ต้องการ

ตัวอย่างที่ 8.4. ต่อเนื่องจากตัวอย่างที่ 8.3 ถึงแม้สรุปได้ว่าทางเลือกทั้งสองแตกต่างกันอย่างมีนัยสำคัญ แต่ครึ่งช่วงความเชื่อมั่นมีค่า 4.387 ชิ้นต่อวันเมื่อใช้ $R_0 = 16$ หากต้องการแยกแยะว่าทางเลือกทั้งสองแตกต่างกันอย่างน้อย 2 ชิ้นต่อวันหรือไม่ ($\varepsilon = 2$) จะต้องประมวลผลเพิ่ม โดยใช้สมการที่ (8.10) เพื่อคำนวณจำนวนรอบทำซ้ำเริ่มต้นที่จะทดลอง กำหนดให้ $1-\alpha = 95\%$

$$R \geqslant \left(\frac{1.96 \times 8.234}{2}\right)^2 = 65.11 \text{ ปัดขึ้นได้ } 66$$

เริ่มที่ $R = 66$ แทนค่าในสมการที่ (8.9) ได้ดังนี้

R	$(t_{1-\alpha/2,R-1}S_D/\varepsilon)^2$
66	67.56
67	67.53
68	**67.49**

ดังนั้นจำนวนรอบทำซ้ำที่ควรใช้คือ 68 รอบเพราะเป็นค่า R ที่น้อยที่สุดที่ทำให้สมการที่ (8.9) เป็นจริง หากใช้เวลาประมวลผลนาน ก็อาจจะใช้เอาท์พุท 16 รอบที่มีอยู่แล้วและทำเพิ่มอีก $68-16=52$ รอบ แต่หากใช้เวลาประมวลผลไม่นาน ก็ทำใหม่ทั้ง 68 รอบก็ได้ เมื่อประมวลผลแล้วควรตรวจสอบอีกครั้งว่าครึ่งช่วงความเชื่อมั่นน้อยกว่า ε หรือไม่ ถ้ายังเกินอยู่ก็ให้เพิ่มจำนวนรอบทำซ้ำ

□

ตัวอย่างที่ 8.5. ต่อเนื่องจากตัวอย่างที่ 8.2 หากกำหนด ε เป็น 300 กล่องต่อวัน ก็ไม่จำเป็นต้องประมวลผลเพิ่ม เพราะ ε มากกว่าครึ่งช่วงความเชื่อมั่น แต่ถ้า $\varepsilon = 100$ ต้องประมวลผลเพิ่ม โดยคำนวณจำนวนรอบทำซ้ำจาก**ค่าประมาณ**ของสูตร (8.11) เพราะว่าในตัวอย่างที่ 8.2 จำนวนรอบทำซ้ำของแต่ละทางเลือกไม่เท่ากัน แต่สูตร (8.11) ใช้สำหรับจำนวนรอบทำซ้ำเท่ากัน

$$R \geqslant \frac{t^2_{0.975,10+8-2}(277.3^2 + 205.9^2)}{100^2} = 53.61$$

หรือ 54 รอบทำซ้ำสำหรับแต่ละทางเลือก เมื่อประมวลผลแล้ว ควรตรวจสอบอีกครั้งว่าครึ่งช่วงความเชื่อมั่นไม่เกิน 100 หรือไม่ ถ้ายังเกินอยู่ ก็ต้องเพิ่มจำนวนรอบทำซ้ำอีก

□

สูตรในสมการที่ (8.9)-(8.11) มีเพื่อช่วยกำหนดจำนวนรอบทำซ้ำ ผู้วิเคราะห์จึงไม่ต้องลองผิดลองถูกเพื่อหาค่า R ที่ให้ความแม่นยำที่ต้องการ แต่หากจำสูตรไม่ได้หรือไม่มีค่าสถิติ t หรือ z อาจใช้วิธีที่อ ๆ ด้วยการเพิ่มจำนวนรอบทำซ้ำทีละ 10 จนกว่าจะได้ครึ่งช่วงความเชื่อมั่นที่แคบกว่า ε ที่ต้องการ หากมีข้อจำกัดด้านเวลาในการประมวลผลที่ไม่สามารถประมวลผลตามจำนวนรอบที่ควรทำได้ ให้รันมากรอบที่สุดเท่าที่เวลามี

ส่วนที่เหลือของบทนี้กล่าวถึงการเปรียบเทียบมากกว่า 2 ทางเลือก

8.2 การเปรียบเทียบมากกว่า 2 ทางเลือก

สมมติว่ามี K ทางเลือกหรือดีไซน์ K แบบที่กำลังพิจารณา วิธีเลือกทางเลือกที่ดีที่สุดขึ้นอยู่กับวัตถุประสงค์ของการศึกษา เช่น

การเปรียบเทียบกับเกณฑ์มาตรฐาน (Benchmark) โดยพิจารณาดัชนีชี้วัดของทางเลือกอื่น ๆ $\theta_i, i = 2, 3, \ldots, K$ เมื่อเทียบกับค่าที่เป็นเกณฑ์มาตรฐาน สมมติเป็น θ_1 เช่น ค่าดัชนีชี้วัดของระบบที่มีอยู่: ใช้วิธีสร้าง**ช่วงความเชื่อมั่นของส่วนต่าง** $\theta_i - \theta_1$ สำหรับ $i = 2, 3, \ldots, K$ จึงมี $K - 1$ อัน ด้วยวิธีที่ได้อธิบายไปแล้วในหัวข้อที่ 8.1.3 คือแบบ t สำหรับตัวอย่าง 2 ชุด (หากประมวลผลทุกทางเลือกอย่างเป็นอิสระต่อกัน) หรือแบบคู่ตัวอย่าง (หากใช้เลขสุ่มร่วมกัน) นอกจากนี้แล้วยังมีวิธีของดันเน็ต (Dunnett method, Tamhane and Dunlop 1999) ซึ่งใช้ซอฟต์แวร์สถิติเช่นมินิแทบ (Minitab) หรือ R สร้างได้

การเปรียบเทียบทุกคู่ (All pairwise comparison) $\theta_i - \theta_j$ สำหรับทุก ๆ $i \neq j$: ใช้วิธีสร้างช่วงความเชื่อมั่นของส่วนต่าง หากมี K ทางเลือก จำนวนคู่เปรียบเทียบคือ $\binom{K}{2} = K(K-1)/2$ สำหรับ $\theta_1 - \theta_2, \theta_1 - \theta_3, \ldots, \theta_{K-1} - \theta_K$ คือได้**ช่วงความเชื่อมั่นของส่วนต่าง** $K(K-1)/2$ อัน นอกจากนี้แล้วยังมีวิธีของทูคีย์ (Tukey method) และวิธีของซู (Hsu method) ซึ่งเปรียบเทียบทุกทางเลือกกับตัวที่ดีที่สุด (ค่าคาด

หมายของดัชนีชี้วัดมากที่สุดหรือน้อยที่สุด) อ่านรายละเอียดเพิ่มเติมจาก Bechhofer et al. (1995) ใช้ซอฟต์แวร์สถิติเช่นมินิแทบหรือ R สร้างได้

การเลือกทางเลือกที่ให้ θ_i ที่ดีที่สุด คือมากที่สุดหรือน้อยที่สุด เช่น ต้นทุนต่ำที่สุดหรือปริมาณงานที่ผลิตได้มากที่สุด: แนะนำให้ใช้**วิธีการจัดลำดับและการคัดเลือก** (Ranking and selection) ซึ่งจะกล่าวถึงวิธีคัดกรองตัวเลือก (Subset selection method) และวิธีเลือกตัวเลือกที่ดีที่สุด (Selection of the best method) ในหัวข้อที่ 8.2.3

หัวข้อที่ 8.2.1 จะกล่าวถึงการเปรียบเทียบด้วยช่วงความเชื่อมั่นทั้งสองแบบที่กล่าวมา

8.2.1 ช่วงความเชื่อมั่น

ตัวอย่างที่ 8.6 แสดงการคำนวณช่วงความเชื่อมั่นแบบเปรียบเทียบกับเกณฑ์มาตรฐาน และตัวอย่างที่ 8.7 ใช้ข้อมูลชุดเดียวกันแต่เปรียบเทียบทุกคู่

ตัวอย่างที่ 8.6. ใช้แบบจำลองเพื่อออกแบบผัง (Layout) แผนกผู้ป่วยนอกของโรงพยาบาลแห่งหนึ่ง ดัชนีชี้วัดคือเวลารอคอยของคนไข้จนกว่าจะได้พบแพทย์ (θ_i) มีผังที่พิจารณา 3 แบบ จึงมีดัชนีชี้วัด 3 ตัวที่เปรียบเทียบกัน คือ $\theta_1, \theta_2, \theta_3$ โดยผังแบบที่ 1 เป็นเกณฑ์มาตรฐาน จึงมีทั้งหมด $C = K - 1 = 2$ คู่ และใช้เลขสุ่มร่วมกัน (CRN) ในการทดลองด้วยแบบจำลอง หากต้องการให้นัยสำคัญรวม 95% แต่ละช่วงความเชื่อมั่นของ $\theta_2 - \theta_1$ และ $\theta_3 - \theta_1$ ควรมีนัยสำคัญ $1 - \alpha/2 = 1 - 0.05/2 = 0.975$ (ด้วยสมการที่ 5.23) ค่าสถิติ $t_{0.975,4} = 3.495$ ตารางที่ 8.3 แสดงเอาท์พุทที่ได้จากการประมวลผลแบบจำลอง

ตารางที่ 8.3: ข้อมูลและการคำนวณสำหรับตัวอย่างที่ 8.6

	เวลารอคอยเฉลี่ยจากรอบทำซ้ำที่ j (นาที)									
ทางเลือก	**1**	**2**	**3**	**4**	**5**					
ผังแบบที่ 1	9	11	10	9	15					
ผังแบบที่ 2	20	21	23	17	30					
ผังแบบที่ 3	6	5	8	14	7					
CI						$\bar{\mathrm{D}}$	$\mathrm{S_D}$	Half width	Lower CI	Upper CI
$Y_{j2} - Y_{j1}$	11	10	13	8	15	11.4	2.70	4.22	7.18	15.62
$Y_{j3} - Y_{j1}$	-3	-6	-2	5	-8	-2.8	4.97	7.77	-10.57	4.97

เมื่อพิจารณาช่วงความเชื่อมั่นในตารางที่ 8.3 จะเห็นว่าผังแบบที่ 1 และ 2 ต่างกันอย่างมีนัยสำคัญ โดยเวลารอคอยในผังแบบที่ 1 ต่ำกว่าผังแบบที่ 2 แต่ผังแบบที่ 1 ไม่แตกต่างจาก 3 อย่างมีนัยสำคัญ จึงตัดผังแบบ

ที่ 2 จากการพิจารณาเพิ่มเติม และยังไม่สามารถสรุปได้ว่าผังแบบที่ 1 และแบบที่ 3 ต่างกันอย่างมีนัยสำคัญหรือไม่

□

ตัวอย่างที่ 8.7. ต่อเนื่องจากตัวอย่างที่ 8.6 เมื่อผู้วิเคราะห์ใช้วิธีเปรียบเทียบทุกคู่ จะมีทั้งหมด

$$C = \binom{3}{2} = 3 \text{ คู่}$$

ใช้เลขสุ่มร่วมกัน (CRN) ในการทดลองด้วยแบบจำลอง หากต้องการให้นัยสำคัญรวม 95% แต่ละช่วงความเชื่อมั่นของ $\theta_i - \theta_j$ สำหรับ $i \neq j$ ควรมีนัยสำคัญ $1 - \alpha/3 = 1 - 0.05/3 = 0.98$ หรือ $\alpha = 0.02$ ด้วยอสมการของบอนเฟอร์โรนี่ (5.23) เนื่องจากใช้เลขสุ่มร่วมกันในการประมวลผล จึงใช้ช่วงความเชื่อมั่นแบบ t สำหรับคู่ตัวอย่างที่ (สมการที่ 8.8) ค่าสถิติ $t_{1-0.02/2,4} = 3.747$

	เอาท์พุทที่ j (นาที)									
CI	1	2	3	4	5	$\bar{D}$	S_D	Half width	Lower CI	Upper CI
$Y_{j1} - Y_{j2}$	-11	-10	-13	-8	-15	-11.4	2.70	4.53	-15.93	-6.87
$Y_{j1} - Y_{j3}$	3	6	2	-5	8	2.8	4.97	8.33	-5.53	11.13
$Y_{j2} - Y_{j3}$	14	16	15	3	23	14.2	7.19	12.05	2.15	26.25

เมื่อพิจารณาช่วงความเชื่อมั่น จะเห็นว่าผังแบบที่ 1 และ 2 ต่างกันอย่างมีนัยสำคัญเพราะช่วงความเชื่อมั่นติดลบทั้งช่วง โดยเวลารอคอยในผังแบบที่ 1 ต่ำกว่าผังแบบที่ 2 แต่ผังแบบที่ 1 ไม่แตกต่างจาก 3 อย่างมีนัยสำคัญเพราะช่วงความเชื่อมั่นคร่อมศูนย์ ผังแบบที่ 2 มีเวลารอคอยสูงกว่าผังแบบที่ 3 โดยสรุปผังแบบที่ 2 ให้เวลารอคอยมากที่สุดจึงตัดออกจากการพิจารณา และสรุปว่าผังแบบที่ 1 และแบบที่ 3 ไม่ต่างกันอย่างมีนัยสำคัญ หากต้องการได้ตัวเลือกดีที่สุดเพียงตัวเดียว ควรประมวลผลใหม่โดยพิจารณาแค่สองทางเลือกนี้ เอาท์พุทที่จะประมวลผลใหม่ควรเป็นอิสระกับข้อมูลที่มีแล้วในตารางที่ 8.3

□

สังเกตว่าผลการวิเคราะห์จากตัวอย่างที่ 8.6 และ 8.7 ให้ผลสรุปในทางเดียวกัน แตกต่างกันตรงที่ตัวอย่างที่ 8.6 เน้นที่เกณฑ์มาตรฐาน ให้น้ำหนักที่ผังแบบที่ 1 มากกว่าแบบอื่น ๆ แต่ตัวอย่างที่ 8.7 ไม่เน้นที่ทางเลือกใดทางเลือกหนึ่งโดยเฉพาะ

หากมีจำนวนตัวเลือกมากกว่า 2-3 ตัว การเปรียบเทียบด้วยช่วงความเชื่อมั่นเป็นวิธีที่ไม่มีประสิทธิภาพจึงไม่เหมาะสม เพราะนัยสำคัญรวมของทุกช่วงความเชื่อมั่นอาจเหลือน้อย หรือหากกำหนดนัยสำคัญรวม ช่วงความเชื่อมั่นของแต่ละอันก็มีนัยสำคัญสูงมาก จึงกว้างจนไม่เป็นประโยชน์ในเชิงปฏิบัติ ดังนั้นควรใช้วิธีในหัวข้อถัดไป

การจัดลำดับและการคัดเลือก

ในกรณีที่มีหลายตัวเลือก วิธีที่ควรใช้คือการจัดลำดับและการคัดเลือกซึ่งประกอบด้วยหลายวิธี ๆ ที่เลือกใช้ขึ้นอยู่กับวัตถุประสงค์ที่ต้องการ เช่น

การคัดกรองตัวเลือก แบ่งทางเลือกเป็นสองกลุ่ม กลุ่มหนึ่งคือ**กลุ่มที่อาจจะดีที่สุด**และ**กลุ่มที่ไม่ใช่กลุ่มที่ดีที่สุดแน่นอน** โดยรับรอง (Guarantee) ว่าความน่าจะเป็นที่ตัวเลือกที่ดีที่สุดจะอยู่ในเซตนี้เป็นอย่างน้อย $1-\alpha$ วิธีนี้มีประโยชน์สำหรับการคัดกรองเบื้องต้นเพื่อตัดทางเลือกที่ด้อยอย่างชัดเจนออกจากการพิจารณา

วิธีเลือกตัวเลือกที่ดีที่สุด คัดเลือกตัวเลือก*ที่ดีที่สุด*เพียงตัวเดียวจากทั้งหมด K ทางเลือก "ดีที่สุด คือมีค่าคาดหมายของดัชนีชี้วัดมากที่สุด (เช่น ปริมาณงานต่อวัน การใช้ประโยชน์ทรัพยากร) หรือน้อยที่สุด (เช่น เวลารอคอยหรือจำนวนงานเสีย) โดยรับรองว่าตัวเลือกที่ดีที่สุดที่แท้จริงจะถูกเลือกด้วยความน่าจะเป็นอย่างน้อย $1-\alpha$

พารามิเตอร์อื่นที่ผู้ใช้ต้องกำหนดสำหรับสองวิธีนี้คือ δ หรือ*ความแตกต่างที่มีนัยสำคัญ* (Significant difference) หากทางเลือกที่ดีที่สุดมีค่าคาดคะเนของดัชนีชี้วัดดีกว่าตัวเลือกอื่น ๆ อย่างน้อย δ ทางเลือกนั้นจะถูกเลือกด้วยความน่าจะเป็นอย่างน้อย $1-\alpha$ ผู้วิเคราะห์หรือผู้มีอำนาจตัดสินใจเป็นผู้กำหนด δ ให้สะท้อนระดับความละเอียดที่ต้องการแยกแยะตัวเลือก กล่าวคือ หากทางเลือกต่างกันน้อยกว่า δ ผู้วิเคราะห์มองว่าทางเลือกเหล่านี้ไม่ต่างกันอย่างมีนัยสำคัญ จึงอาจพิจารณาดัชนีชี้วัดรองว่าควรนำทางเลือกใดไปใช้จริง

8.2.2 การคัดกรองตัวเลือก

หัวข้อนี้จะกล่าวถึงวิธีการคัดกรองตัวเลือกในเครื่องมือ **Process Analyzer (PAN)** ที่มาพร้อมกับซอฟต์แวร์อารีน่า วิธีนี้นำเสนอโดย Boesel et al. (2003) PAN เรียกความแตกต่างที่มีนัยสำคัญว่า**ค่าเผื่อผิดพลาด** (Error tolerance) แทนด้วยสัญลักษณ์ δ และมีค่านัยสำคัญ $1-\alpha$ ตายตัวที่ 95% ผู้อ่านศึกษาวิธีการใช้ PAN ได้จากหนังสือ Kelton et al. (2009) หรือ Help menu ของซอฟต์แวร์

ค่าเผื่อส่งผลต่อเซตที่ได้ เมื่อตั้งค่าเผื่อเป็นศูนย์ วิธีของ Nelson et al. (2001) รับรองว่าเซตที่ได้มีตัวที่ดีที่สุดด้วยความน่าจะเป็น $1-\alpha$ เมื่อค่าเผื่อมากกว่าศูนย์ เซตนี้จะมีตัวที่ดีที่สุดหรือตัวที่มีค่าคาดหมายของดัชนีชี้วัดห่างจากตัวที่ดีที่สุดไม่เกิน δ ด้วยความน่าจะเป็น $1-\alpha$ เมื่ออ่านถึงตอนนี้ ผู้ใช้อาจคิดว่าควรตั้งค่าเผื่อเป็นศูนย์เสมอ แต่ในเชิงปฏิบัติแล้วค่าเผื่อที่เป็นศูนย์บีบให้วิธีการคัดกรองเก็บตัวเลือกไว้หลายตัวเพื่อให้สามารถรับรองว่าเซตที่ถูกเลือกจะมีตัวที่ดีที่สุดด้วยความน่าจะเป็นที่ผู้ใช้กำหนด ผลที่ตามมาคืออาจไม่สามารถสรุปได้ชัดเจนว่าใครดีกว่าใคร เช่น ในตัวอย่างที่ 8.8 ข้างล่างเมื่อตั้งค่าเผื่อเป็นศูนย์ วิธีคัดกรองเก็บทั้งผังแบบที่ 1 และผังแบบที่ 3 ซึ่งเป็นข้อสรุปเดียวกันกับเมื่อใช้ช่วงความเชื่อมั่น (ตัวอย่างที่ 8.6-8.7) แต่เมื่อกำหนดให้ค่า

เผื่อเป็น 2.7 นาที เหลือผังแบบที่ 3 อันเดียวที่ถูกเก็บไว้จึงเป็นตัวเลือกที่ดีที่สุดโดยปริยาย ดังนั้นจึงควรตั้งค่าเผื่อให้เหมาะสม

วิธีการคัดกรองตัวเลือกของ Boesel et al. (2003) มีขั้นตอนคร่าว ๆ ดังนี้ คำนวณค่าเฉลี่ยตัวอย่าง (Sample mean) ของแต่ละตัวเลือก ($\bar{Y}_i$ สำหรับทางเลือก $i = 1, 2, \ldots, K$) เก็บตัวเลือกที่มีค่าเฉลี่ยที่ดีที่สุดเอาไว้ หากดัชนีชี้วัดยิ่งมากยิ่งดี จะเก็บตัวเลือกที่มีค่าเฉลี่ยสูงสุด หากดัชนีชี้วัดยิ่งน้อยยิ่งดี จะเก็บตัวเลือกที่มีค่าเฉลี่ยต่ำที่สุด ตัวเลือกอื่น ๆ จะถูกเก็บไว้หากค่าเฉลี่ยไม่ห่างจากค่าเฉลี่ยที่ดีที่สุด*มากจนเกินไป* (อธิบายเพิ่มเติมภายหลัง) พารามิเตอร์ที่เกี่ยวข้องมีดังนี้: K แทนจำนวนทางเลือกที่มี $1-\alpha$ เป็นระดับนัยสำคัญ δ เป็นค่าเผื่อผิดพลาด R_i เป็นจำนวนรอบทำซ้ำหรือขนาดตัวอย่างของทางเลือก i นิยามค่าสถิติ t_i ดังนี้

$$t_i = t_{(1-\alpha)^{1/(K-1)}, R_i - 1} \tag{8.12}$$

ขนาด*หน้าต่าง* (Window) ของคู่ทางเลือก i และ j เรียกว่า W_{ij} คล้ายกับครึ่งช่วงความเชื่อมั่นเฉลี่ย คำนวณดังนี้

$$W_{ij} = \sqrt{\frac{t_i^2 S_i^2}{R_i} + \frac{t_j^2 S_j^2}{R_j}} \tag{8.13}$$

สังเกตว่า $W_{ij} = W_{ji}$ หากดัชนีชี้วัด**ยิ่งมากยิ่งดี** เก็บทางเลือก i ในเซตถ้าเงื่อนไขนี้เป็นจริง

$$\bar{Y}_{\cdot i} \geqslant \bar{Y}_{\cdot j} - (W_{ij} - \delta)^+ \text{ สำหรับทุก ๆ } j \neq i \tag{8.14}$$

โดย $x^+ = \max(x, 0)$ เช่น $(-5)^+ = \max(-5, 0) = 0$ และ $(5)^+ = \max(5, 0) = 5$ หากดัชนีชี้วัด**ยิ่งน้อยยิ่งดี** เก็บทางเลือก i ในเซตถ้าเงื่อนไขนี้เป็นจริง

$$\bar{Y}_{\cdot i} \leqslant \bar{Y}_{\cdot j} + (W_{ij} - \delta)^+ \text{ สำหรับทุก ๆ } j \neq i \tag{8.15}$$

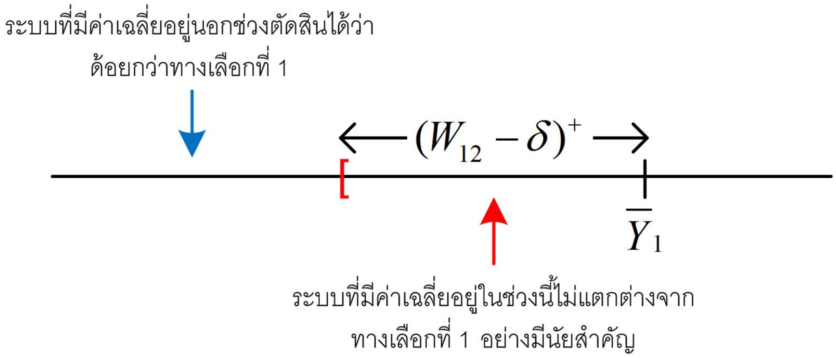
ค่า W_{ij} มีไว้เพื่อให้โอกาสกับทางเลือกซึ่งมีค่าเฉลี่ยไม่แย่มากนัก รูปที่ 8.2 แสดงการเปรียบเทียบระหว่างทางเลือกที่ 1 และ 2 โดยใช้สมการที่ (8.14) สมมติว่าดัชนีชี้วัดยิ่งมากยิ่งดีและทางเลือกที่ 1 มีค่าเฉลี่ยสูงที่สุด $\bar{Y}_1$ หากค่าเฉลี่ยของทางเลือกที่ 2 ตกอยู่ในช่องวงเล็บ ก็จะถูกเก็บไว้ แต่หากตกอยู่นอกวงเล็บ ก็จะถูกคัดออก

ตัวอย่างที่ 8.8 แสดงขั้นตอนการใช้การคัดกรองตัวเลือก ซึ่งแสดงให้เห็นว่าค่าเผื่อมีผลต่อเซตที่ได้

ตัวอย่างที่ 8.8. ข้อมูลจากตัวอย่างที่ 8.6 ค่าพารามิเตอร์ที่เกี่ยวข้องมีดังนี้: ขนาดตัวอย่าง (R_i) เท่ากับ 5 จึงมีค่าองศาอิสระเป็น 4 มีจำนวนตัวเลือก (K) เป็น 3 กำหนดให้นัยสำคัญ $1-\alpha = 0.95$

ในสูตร (8.12) ค่า t_i มีนัยสำคัญที่ต้องการดังนี้

$$t_i = t_{(1-\alpha)^{\frac{1}{K-1}}, R_i - 1} = t_{0.95^{\frac{1}{3-1}}, 5-1} = t_{0.9747, 4} = 2.764$$



รูปที่ 8.2: การเปรียบเทียบทางเลือกที่ 1 และ 2 ด้วยวิธีคัดกรองตัวเลือกโดยใช้สมการที่ (8.14)

ตัวเลือก i	**ผังแบบที่ 1**	**ผังแบบที่ 2**	**ผังแบบที่ 3**
$\bar{Y}_{\cdot i}$	10.8	22.2	8
$S^2_{\cdot i}$	6.2	23.7	12.5

โดยอาจใช้ฟังก์ชัน `=tinv(2*(1-0.9747), 4)` ในเอกซ์เซล คูณสองเพราะฟังก์ชันนี้ใช้ความน่าจะเป็นรวมของหางทั้งสองข้างของ PDF คำนวณค่าความกว้างหน้าต่าง W_{ij} ด้วยสมการที่ (8.13) ดังนี้

$$W_{12} = \sqrt{\frac{2.764^2 \times 6.2}{5} + \frac{2.764^2 \times 23.7}{5}} = 6.759$$

ค่าระหว่างคู่อื่นมีดังนี้ $W_{13} = 5.345$ และ $W_{23} = 7.437$ เมื่อกำหนดค่า $\delta = 0$ แล้วใช้สมการที่ (8.15) เพื่อเปรียบเทียบระบบ ผังแบบที่ 3 มีเวลาเฉลี่ยน้อยที่สุดจึงเก็บไว้ในเซตที่ถูกเลือก พิจารณาผังแบบที่ 1 เปรียบเทียบกับผังแบบที่ 2 และ 3

1 เทียบกับ 2:

$$\begin{aligned} \bar{Y}_{\cdot 1} &\overset{?}{\leqslant} \bar{Y}_{\cdot 2} + (W_{12} - \delta)^+ \\ 10.8 &\overset{\checkmark}{\leqslant} 22.2 + (6.759 - 0)^+ = 28.96 \end{aligned}$$

1 เทียบกับ 3:

$$\begin{aligned} \bar{Y}_{\cdot 1} &\overset{?}{\leqslant} \bar{Y}_{\cdot 3} + (W_{13} - \delta)^+ \\ 10.8 &\overset{\checkmark}{\leqslant} 8 + (5.345 - 0)^+ = 13.345 \end{aligned}$$

เงื่อนไข (8.15) เป็นจริงจึงเก็บผังแบบที่ 1 ไว้ในเซตที่ถูกเลือก เหลือผังแบบที่ 2 ให้พิจารณาต่อ ดังนี้

2 เทียบกับ 1:

$$\bar{Y}_{\cdot 2} \overset{?}{\leqslant} \bar{Y}_{\cdot 1} + (W_{12} - \delta)^{+}$$
$$22.2 \overset{\times}{\leqslant} 10.8 + (6.759 - 0)^{+} = 17.559$$

เงื่อนไข (8.15) ไม่เป็นจริง จึงไม่ต้องตรวจสอบผังแบบที่ 2 เทียบกับแบบที่ 3 ดังนั้นเซตที่ถูกเลือกจึงเหลือ $\{1, 3\}$ แต่หากตั้งค่าเผื่อสูงขึ้น เช่น $\delta = 2.7$ นาที จะเหลือเพียงผังแบบที่ 3 ในเซต ดังนั้นการกำหนดค่าเผื่อที่เหมาะสม (ไม่ใช่แค่ค่าศูนย์) อาจทำให้ได้ข้อสรุปที่ชัดเจน

□

หากจำเป็นต้องใช้วิธีคัดกรองตัวเลือกเพื่อคัดเลือกตัวเลือกที่ดีที่สุดเพียงตัวเดียว ก็ให้เพิ่มขนาดตัวอย่างหรือจำนวนรอบทำซ้ำ (อาจเป็นทีละ 10) จนกว่าเซตที่ถูกเก็บไว้เหลือตัวเลือกที่ดีที่สุดเพียงตัวเดียว หัวข้อถัดไปจะกล่าวถึงวิธีทางสถิติที่ถูกออกแบบมาเพื่อวัตถุประสงค์นี้โดยเฉพาะ

8.2.3 วิธีเลือกตัวเลือกที่ดีที่สุด

วิธีของ Nelson et al. (2001) นี้ต่อยอดมาจากวิธีคัดกรองตัวเลือกที่ได้เพิ่งนำเสนอไปในหัวข้อที่ 8.2.2 สามารถทำได้เมื่อใช้ตัวแปรสุ่มร่วมกัน (CRN ในหัวข้อที่ 8.1.1) หรือเมื่อสุ่มอย่างเป็นอิสระต่อกัน ประกอบด้วยการประมวลผล 2 ขั้น ขั้นแรกเพื่อประมาณความแปรปรวนในเอาท์พุทของแต่ละทางเลือก แล้วใช้ความแปรปรวนนี้ร่วมกับพารามิเตอร์อื่นเพื่อกำหนดจำนวนรอบทำซ้ำ (หรือขนาดตัวอย่าง) ที่ต้องใช้ในการประมวลผลขั้นสุดท้าย หลังจากนั้นทางเลือกที่ดีที่สุดจะถูกเลือกโดยใช้ค่าเฉลี่ยตัวอย่างจากข้อมูลทั้งหมด ถ้าดัชนีชี้วัดยิ่งมากยิ่งดี จะเลือกทางเลือกที่มีค่าเฉลี่ยสูงที่สุด ถ้าดัชนีชี้วัดยิ่งน้อยยิ่งดี จะเลือกทางเลือกที่มีค่าเฉลี่ยต่ำที่สุด

วิธีของ Nelson et al. (2001) รับรองว่าจะเลือกทางเลือกที่ดีที่สุดหรือตัวที่ห่างจากตัวที่ดีที่สุดไม่เกิน δ ด้วยความน่าจะเป็น $1 - \alpha$ เช่น 95% โดย δ และ $1 - \alpha$ เป็นพารามิเตอร์ที่ผู้ใช้กำหนด นอกจากนี้ยังมีจำนวนทางเลือกที่พิจารณา (K) และขนาดตัวอย่างในการประมวลผลขั้นแรก (R_0) ขั้นตอนช่วงแรกของวิธีเลือกตัวเลือกที่ดีที่สุดเหมือนกับวิธีการคัดกรองตัวเลือก แต่ที่ความน่าจะเป็นที่จะเลือกตัวที่ดีที่สุด $1 - 0.5\alpha$ และความน่าจะเป็นที่จะผิดพลาดที่เหลืออีก 0.5α สำหรับการเลือกตัวเลือกที่ดีที่สุด

ขั้นตอนวิธีเลือกตัวเลือกที่ดีที่สุด

สังเกตว่าค่านัยสำคัญที่สมเหตุผลอยู่ในช่วง $1/K < 1 - \alpha < 1$ เพราะหากนัยสำคัญน้อยกว่า $1/K$ ก็ไม่จำเป็นต้องทดลองใด ๆ แค่สุ่มเลือกโดยให้ทุกทางเลือกมีโอกาสที่จะถูกเลือกเท่ากัน ก็ได้นัยสำคัญที่ต้องการแล้ว จำนวนรอบทำซ้ำของการประมวลผลขั้นแรก (R_0) ไม่ควรต่ำกว่า 10 เพราะถ้าต่ำกว่านั้น ค่าความแปรปรวนที่ประมาณได้ S_i^2 จะมีความไม่แม่นยำสูง

ช่วงเริ่มต้น กำหนดค่าคงที่ของไรน็อต (Rinott's constant, h) ที่ $R_0, K,$ และ $1-0.5\alpha$ (ใช้ตารางที่ 8.5 หากไม่มีค่าที่ต้องการ ให้ดูที่ Bechhofer et al. (1995)) และค่าสถิติ

$$t = t_{(1-0.5\alpha)^{\frac{1}{K-1}}, R_0-1} \tag{8.16}$$

สังเกตว่าเป็นสมการเดียวกับสมการที่ (8.12) แต่ต่างกันที่นัยสำคัญ $1-0.5\alpha$ และทุกทางเลือกมี R_0 เท่ากัน

ช่วงประมวลผลขั้นแรก เพื่อคัดทางเลือกที่ด้อยกว่าอย่างชัดเจนออก

1. ประมวลผลแต่ละทางเลือกโดยใช้ R_0 รอบทำซ้ำ; คำนวณค่าเฉลี่ย ($\bar{Y}_i$) และความแปรปรวนจากตัวอย่าง (S_i^2) สำหรับ $i = 1, 2, \ldots, K$
2. คำนวณความกว้างหน้าต่าง W_{ij} ด้วยสมการที่ (8.17) และ t จากสมการที่ (8.16) ดังนี้

$$W_{ij} = t\sqrt{\frac{S_i^2 + S_j^2}{R_0}} \text{ สำหรับ } i \neq j \tag{8.17}$$

3. เมื่อต้องการดัชนีชี้วัดที่**สูง**ที่สุด หากทางเลือกใดผ่านเกณฑ์ (8.14) ให้เก็บทางเลือกนั้นไว้ ในกรณีที่ต้องการดัชนีชี้วัดที่**ต่ำ**ที่สุด หากทางเลือกใดผ่านเกณฑ์ (8.15) ให้เก็บทางเลือกนั้นไว้

   ให้เซต $\mathcal{S}$ เป็นกลุ่มของทางเลือกที่ยังเหลือ หากเซตนี้มีสมาชิกเหลือเพียงตัวเดียว สรุปได้ทันทีว่าทางเลือกนี้ดีที่สุด มิเช่นนั้นดำเนินการต่อในช่วงประมวลผลขั้นสุดท้าย

ช่วงประมวลผลขั้นสุดท้าย พิจารณาเฉพาะทางเลือกที่ยังไม่ถูกคัดออก คือ $i \in \mathcal{S}$

1. คำนวณจำนวนรอบทำซ้ำที่ต้องการของแต่ละทางเลือก R_i ดังนี้

$$R_i = \max\left(R_0, \left\lceil \left(\frac{hS_i}{\delta}\right)^2 \right\rceil\right) \tag{8.18}$$

   เครื่องหมาย $\lceil\cdot\rceil$ แปลว่าปัดขึ้นให้เป็นจำนวนเต็มที่ใกล้ที่สุด เช่น $\lceil 21.37 \rceil = 22$ สังเกตว่าในสูตร (8.18) ค่าเผื่อ δ ต้องมากกว่าศูนย์เพราะเป็นตัวหาร

2. ประมวลผลทางเลือก i เพิ่มเติมด้วยจำนวน $R_i - R_0$ รอบทำซ้ำ หรือถ้าใช้เวลาประมวลผลไม่มากอาจประมวลผลทั้ง R_i รอบใหม่ทั้งหมด
3. คำนวณค่าเฉลี่ยจากข้อมูลทั้งสองช่วง

$$\bar{\bar{Y}}_{\cdot i} = \frac{1}{R_i}\sum_{r=1}^{R_i} Y_{ri} \tag{8.19}$$

   เมื่อต้องการค่าคาดหมายของดัชนีชี้วัดที่**สูง**ที่สุด ให้เลือกทางเลือก i ที่มี $\bar{\bar{Y}}_{\cdot i}$ **สูง**ที่สุด ในกรณีที่ต้องการค่าคาดหมายของดัชนีชี้วัดที่**ต่ำ**ที่สุด ให้เลือกทางเลือก i ที่มี $\bar{\bar{Y}}_{\cdot i}$ **ต่ำ**ที่สุด

เนื่องจากวิธี Nelson et al. (2001) นี้อาศัยค่าคงที่ของไรน็อตซึ่งไม่มีในซอฟต์แวร์สถิติทั่วไปหรือเอกซ์เซล จึงขอแนะนำให้ใช้ฟังก์ชัน `Select the Best` ในเอกซ์เซลเทมเพลทชื่อ `SimulationTools.xls` ที่สามารถดาวน์โหลดได้ฟรีจาก `www.bcnn.net`

ตัวอย่างที่ 8.9 แสดงการคัดเลือกตัวเลือกที่ดีที่สุด

ตัวอย่างที่ 8.9. พิจารณาปัญหาเด็กส่งหนังสือพิมพ์ในหัวข้อที่ 3.3 ตารางที่ 3.9 แสดงแบบจำลองในเอกซ์เซล ตัวแปรตัดสินใจคือจำนวนตั๋วที่ให้จองและดัชนีชี้วัดคือกำไรสุทธิ วัตถุประสงค์คือต้องการกำหนดจำนวนตั๋วที่ให้จองที่ทำให้กำไรสุทธิสูงสุด ตารางที่ 8.4 แสดงทางเลือกที่พิจารณา

ตารางที่ 8.4: ผลการประมวลผลขั้นแรก $R_0 = 30$

ทางเลือก i	จำนวนตั๋วที่ให้จอง	$\bar{Y}_i$	S_i^2
1	18	79,200	2.0390×10^7
2	19	81,600	3.0352×10^7
3	20	81,950	6.8273×10^7
4	21	84,325	6.9272×10^7
5	22	83,025	7.3364×10^7
6	23	81,650	8.8360×10^7

กำหนดค่าพารามิเตอร์ดังนี้: $1-\alpha = 90\%$ หรือ $\alpha = 0.1, \delta = 1{,}000$ บาท, และ $R_0 = 30$ รอบทำซ้ำ การประมวลผลเบื้องต้นให้ค่าสรุปทางสถิติดังแสดงในตารางที่ 8.4 ค่าสถิติ t ในสมการที่ (8.16) คือ

$$t = t_{(1-0.5\times 0.1)^{\frac{1}{6-1}},30-1} = t_{0.99,29} = 2.462$$

และค่าคงที่ของไรน็อต $h(R_0 = 30, K = 6, 1-0.5\alpha = 0.95) = 3.434$ ในช่วงประมวลผลขั้นแรก คำนวณ W_{14} ได้ดังนี้

$$W_{14} = t\sqrt{\frac{S_1^2 + S_4^2}{30}} = 2.462\sqrt{\frac{2.0390\times 10^7 + 6.9272\times 10^7}{30}} = 4,256.33$$

ทางเลือกที่ 1 ถูกคัดออกเพราะไม่ผ่านเกณฑ์ (8.14) เมื่อเปรียบเทียบทางเลือกที่ 1 และทางเลือกที่ 4 ที่มีค่าเฉลี่ยสูงสุด

$$79,200 \overset{\times}{\geqslant} 84,325 - \max(4,256.33 - 1,000, 0) = 84,325 - 3,256.33 = 80,068.67$$

จึงเหลือเพียง $\mathcal{S} = \{2,3,4,5,6\}$ ในการประมวลผลขั้นที่สอง

คำนวณจำนวนรอบทำซ้ำที่ต้องการของทางเลือกที่ 2

$$R_2 = \max\left(30, \left\lceil \frac{3.434^2 \times 3.0352 \times 10^7}{1,000^2} \right\rceil\right) = 358$$

เมื่อใช้ `SimulationTools.xls` จำนวนรอบทำซ้ำที่ต้องการ**เพิ่ม**ทั้งหมดมีดังนี้

$$R_2 - R_0 = 328, R_3 - R_0 = 775, R_4 - R_0 = 787, R_5 - R_0 = 835, R_6 - R_0 = 1012$$

เมื่อประมวลผลเพิ่มแล้ว คำนวณค่าเฉลี่ยของข้อมูลทั้งหมด (สมการที่ 8.19)

$$\bar{\bar{Y}}_{\cdot 2} = 81{,}578, \bar{\bar{Y}}_{\cdot 3} = 83{,}365, \bar{\bar{Y}}_{\cdot 4} = 83{,}190, \bar{\bar{Y}}_{\cdot 5} = 83{,}274, \bar{\bar{Y}}_{\cdot 6} = 83{,}445$$

ทางเลือกที่ดีที่สุดคือทางเลือกที่ 6 หรือจำนวนตั๋วที่ให้จองที่ดีที่สุดคือ 23 ใบ เพราะให้ค่าเฉลี่ยของกำไรสุทธิสูงที่สุด

□

8.3 บทส่งท้าย

การเปรียบเทียบทางเลือกด้วยแบบจำลองมีเนื้อหาอีกมากที่ผู้เขียนยังไม่ได้กล่าวถึง โดยเฉพาะอย่างยิ่งการหาค่าที่เหมาะที่สุด (Optimization) ซึ่งเป็นหัวข้อที่มีงานวิจัยใหม่ ๆ ออกมาอย่างต่อเนื่อง หากผู้อ่านสนใจแนะนำให้เริ่มจากการอ่านบทความในช่วง Advanced tutorial ของการประชุมวิชาการ Winter Simulation Conference (`http://informs-sim.org/`) และคู่มือการจำลองสถานการณ์ เช่น Banks (1998) และ Chung (2004)

ซอฟต์แวร์การจำลองสถานการณ์เชิงพาณิชย์มักมีเครื่องมือการหาค่าที่เหมาะที่สุด ซึ่งโดยมากพัฒนามาจากฮิวริสติกที่ไม่สามารถรับรองว่าคำตอบที่ได้เป็นคำตอบที่เหมาะที่สุดที่แท้จริง ดังนั้นหากใช้เครื่องมือเหล่านี้ ควรทดลองประมวลผลที่พารามิเตอร์หลากหลาย เช่น เวลาในการค้นหา (Run time) หรือจำนวนครั้งที่ประมวลผลต่อหนึ่งคำตอบ เพื่อทดสอบดูว่าคำตอบที่ได้มาทนทานต่อความเปลี่ยนแปลงของพารามิเตอร์ของการทดลองหรือไม่ (Robustness) ผู้ใช้ควรศึกษาจุดเด่นจุดด้อยและข้อจำกัดของเครื่องมือที่ใช้ด้วย

ตารางที่ 8.5: ค่าคงที่ของไรน็อต (ตารางที่ A.12 ใน Banks et al. 2005)

		K								
$1-\alpha$	R_0	2	3	4	5	6	7	8	9	10
	5	2.291	3.058	3.511	3.837	4.093	4.305	4.486	4.644	4.786
	6	2.177	2.871	3.270	3.552	3.771	3.951	4.103	4.235	4.352
	7	2.107	2.758	3.126	3.384	3.582	3.744	3.881	3.999	4.103
	8	2.059	2.682	3.031	3.273	3.489	3.609	3.736	3.845	3.941
	9	2.025	2.628	2.963	3.195	3.372	3.515	3.635	3.738	3.829
	10	1.999	2.587	2.913	3.137	3.307	3.445	3.560	3.659	3.746
	11	1.978	2.556	2.874	3.092	3.258	3.391	3.503	3.598	3.682
	12	1.962	2.531	2.843	3.056	3.218	3.349	3.457	3.551	3.632
	13	1.948	2.510	2.817	3.027	3.186	3.314	3.420	3.512	3.592
0.90	14	1.937	2.493	2.796	3.003	3.160	3.285	3.390	3.480	3.558
	15	1.928	2.479	2.779	2.983	3.138	3.261	3.364	3.453	3.530
	16	1.919	2.467	2.764	2.966	3.119	3.241	3.343	3.430	3.506
	17	1.912	2.456	2.751	2.951	3.102	3.223	3.324	3.410	3.485
	18	1.906	2.447	2.739	2.938	3.088	3.208	3.308	3.393	3.467
	19	1.901	2.438	2.729	2.926	3.075	3.194	3.293	3.378	3.451
	20	1.896	2.431	2.720	2.916	3.064	3.182	3.280	3.364	3.437
	30	1.866	2.387	2.666	2.855	2.997	3.110	3.204	3.284	3.354
	40	1.852	2.366	2.641	2.827	2.966	3.077	3.169	3.247	3.315
	50	1.844	2.354	2.627	2.810	2.948	3.057	3.148	3.225	3.292
	5	3.107	3.905	4.390	4.744	5.025	5.259	5.461	5.638	5.797
	6	2.910	3.602	4.010	4.303	4.533	4.722	4.884	5.025	5.150
	7	2.791	3.424	3.791	4.051	4.253	4.419	4.559	4.681	4.789
	8	2.712	3.308	3.649	3.889	4.074	4.225	4.353	4.463	4.561
	9	2.656	3.226	3.550	3.776	3.950	4.091	4.210	4.313	4.404
	10	2.614	3.166	3.476	3.693	3.859	3.993	4.106	4.204	4.290
	11	2.582	3.119	3.420	3.629	3.789	3.918	4.027	4.121	4.203

ต่อในหน้าถัดไป

ตารางที่ 8.5 – *ต่อจากหน้าที่แล้ว*

$1-\alpha$	R_0	K								
		2	3	4	5	6	7	8	9	10
0.95	12	2.556	3.082	3.376	3.579	3.734	3.860	3.965	4.055	4.135
	13	2.534	3.052	3.340	3.539	3.690	3.812	3.915	4.003	4.080
	14	2.517	3.027	3.310	3.505	3.654	3.773	3.874	3.960	4.035
	15	2.502	3.006	3.285	3.477	3.623	3.741	3.839	3.924	3.998
	16	2.489	2.988	3.264	3.453	3.597	3.713	3.810	3.893	3.966
	17	2.478	2.973	3.246	3.433	3.575	3.689	3.785	3.867	3.938
	18	2.468	2.959	3.230	3.415	3.556	3.669	3.763	3.844	3.914
	19	2.460	2.948	3.216	3.399	3.539	3.650	3.744	3.824	3.894
	20	2.452	2.937	3.203	3.385	3.523	3.634	3.727	3.806	3.875
	30	2.407	2.874	3.129	3.303	3.434	3.539	3.626	3.701	3.766
	40	2.386	2.845	3.094	3.264	3.392	3.495	3.580	3.652	3.716
	50	2.373	2.828	3.074	3.242	3.368	3.469	3.553	3.624	3.687

แบบฝึกหัด

8.1. ผู้วิเคราะห์ใช้แบบจำลองระบบดำเนินงานคลังสินค้า มีวิธีการจัดเก็บสินค้า 4 แบบให้เลือก ดัชนีชี้วัดหลักคือเวลาในการจัดสินค้าต่อหนึ่งใบสั่งซื้อ ใช้ช่วงความเชื่อมั่นแบบทุกคู่ในการเปรียบเทียบ และต้องการให้นัยสำคัญรวม 95% แต่ละช่วงความเชื่อมั่นจะต้องมีนัยสำคัญเท่าใด

8.2. ผู้บริหารโรงพยาบาลแห่งหนึ่งได้รับร้องเรียนจากผู้ป่วยนอกหลายรายว่าเวลาที่อยู่ที่โรงพยาบาลนานเกินไป ผู้บริหารจึงปรึกษาแผนกปรับปรุงระบบธุรกิจ และได้แนวความคิดใหม่ในการวางผังการไหลของคนไข้ ก่อนที่จะเปลี่ยนแปลงหน้างานจริง ผู้บริหารต้องการผลการวิเคราะห์เชิงตัวเลขว่าเมื่อปรับปรุงระบบแล้ว จะทำให้เวลาในระบบของคนไข้เปลี่ยนแปลงอย่างไร แผนกปรับปรุงระบบธุรกิจได้พัฒนาแบบจำลองของแผนกผู้ป่วยนอก และประเมินเวลาในระบบของคนไข้ภายใต้ทั้งสองสถานการณ์ ดังนี้

ชนิด (i)	จำนวนเอาท์พุท (R_i)	ค่าเฉลี่ยเวลาในระบบ ($\bar{Y}_{\cdot i}$, นาที)	ส่วนเบี่ยงเบนมาตรฐาน (S_i, นาที)
ระบบปัจจุบัน (1)	10	136.14	3.59
ระบบที่ปรับปรุงแล้ว (2)	10	112.71	0.79

ระบบที่ปรับปรุงแล้วดีกว่าระบบปัจจุบันอย่างมีนัยสำคัญหรือไม่ ให้ระบุค่านัยสำคัญที่ใช้

8.3. แผนกประกอบอุปกรณ์สุขภัณฑ์อยู่ในระหว่างการปรับปรุงประสิทธิภาพด้วยการลดขั้นตอนการทำงาน ตารางข้างล่างแสดงข้อมูลจากการจับเวลาในการประกอบชิ้นส่วนชิ้นหนึ่ง โดยพนักงานคนที่ j

ข้อมูลที่ j	1	2	3	4	5	6
เวลาประกอบในปัจจุบัน (ชม.)	0.430	0.266	0.567	0.531	0.707	0.716
เวลาประกอบที่ลดขั้นตอนแล้ว (ชม.)	0.415	0.238	0.390	0.410	0.605	0.609

ระบบที่ลดขั้นตอนดีกว่าระบบปัจจุบันอย่างมีนัยสำคัญหรือไม่ ให้ระบุค่านัยสำคัญที่ใช้

8.4. บริษัทแห่งหนึ่งมีรถบรรทุกที่ทำกิจกรรม 3 อย่างซ้ำไปซ้ำมา คือ บรรทุกของขึ้น, ชั่งน้ำหนัก, และวิ่งไปส่งของ สมมติว่ามีรถบรรทุก 8 คันในระบบและที่เวลาศูนย์ (เวลาเริ่มต้น) ทั้ง 8 คันอยู่ที่จุดบรรทุกของขึ้น เวลาในการชั่งน้ำหนักต่อคันมีการแจกแจงแบบสม่ำเสมอในช่วง 1 ถึง 9 นาที เวลาวิ่งไปส่งของมีการแจกแจงแบบเอ็กซ์โพเนนเชียลด้วยค่าเฉลี่ย 85 นาที มีที่ว่างหน้าจุดขึ้นของและหน้าเครื่องชั่งน้ำหนักไม่จำกัด รถบรรทุกทั้งหมดสามารถวิ่งพร้อมกันได้ ผู้บริหารต้องการเปรียบเทียบสองทางเลือก ดังนี้

- **ระบบปัจจุบัน:** เครื่องขึ้นของแบบช้า 2 เครื่อง ซึ่งใช้เวลาบรรจุ 1-27 นาที (การแจกแจงแบบสม่ำเสมอ)
- **ระบบใหม่:** เครื่องขึ้นของแบบเร็ว 1 เครื่อง ซึ่งใช้เวลาบรรจุ 1-19 นาที (การแจกแจงแบบสม่ำเสมอ)

กำหนดให้รอบทำซ้ำยาว 7 วัน ทำงานวันละ 8 ชั่วโมง ดัชนีชี้วัดที่สนใจคือเวลาในระบบ ตั้งแต่รถบรรทุกเข้าคิวบรรจุของจนกระทั่งเสร็จและออกจากเครื่องชั่ง สร้างแบบจำลองสถานการณ์เพื่อเปรียบเทียบสองทางเลือกด้วยวิธีทางสถิติที่สมเหตุสมผล ให้ใช้จำนวนรอบทำซ้ำมากพอที่จะแยกแยะส่วนต่างของเวลาในระบบอย่างน้อย 2 นาที และใช้เลขสุ่มร่วมกันในการเปรียบเทียบสองทางเลือกนี้

8.5. ใช้แบบจำลองสถานการณ์เพื่อเปรียบเทียบสายการประกอบ 3 แบบโดยมีดัชนีชี้วัดคือปริมาณงานที่ประกอบได้ใน 1 ชั่วโมง ประมวลผลแต่ละทางเลือก 15 รอบทำซ้ำ ได้ค่าสรุปทางสถิติดังนี้

$$\bar{Y}_1 = 10, \bar{Y}_2 = 8, \bar{Y}_3 = 5, S_i^2 = 10 \text{ สำหรับทุก ๆ } i$$

กำหนดนัยสำคัญ 95% ให้ใช้วิธีคัดกรองตัวเลือกเพื่อตัดระบบที่ด้อยกว่าอย่างชัดเจนที่ค่าเผื่อผิดพลาด (δ) 2 กรณีดังนี้

(i) $\delta = 0$

(ii) $\delta = 1$

8.6. พิจารณาแบบจำลองการเลือกโครงการลงทุนในหัวข้อที่ 3.5 แบบจำลองในรูปที่ **??** ให้ออกแบบนโยบายการลงทุน 5 แบบ แล้วใช้วิธีการเลือกตัวเลือกที่ดีที่สุดเพื่อเลือกนโยบายที่ทำให้ผลตอบแทนสูงสุด ระบุค่านัยสำคัญและค่าเผื่อ δ ที่ใช้

9

การตรวจสอบแบบจำลอง

บทนี้มีไว้เพื่อตอบคำถามชนิดที่ว่า "รู้ได้อย่างไรว่าแบบจำลองถูกต้อง" โดยจะเริ่มที่ประเด็นที่มักถูกมองข้าม ซึ่งได้ถูกกล่าวถึงไปบ้างแล้วในบทส่งท้ายของบทที่ผ่าน ๆ มา โดยเฉพาะอย่างยิ่งหัวข้อที่ 5.6

9.1 ข้อผิดพลาดที่พบบ่อยในการวิเคราะห์ด้วยแบบจำลองสถานการณ์

ผู้วิเคราะห์มักสนใจในเครื่องมือที่ใช้มากกว่าโจทย์ของระบบที่อยู่ตรงหน้า เช่น อยากใช้การจำลองสถานการณ์เพื่อจำลองระบบ แต่ไม่สามารถตอบตัวเองได้ว่าพัฒนาไปเพื่อประเมินอะไรหรือประมาณดัชนีชี้วัดไหน เมื่อไม่ชัดเจนในวัตถุประสงค์ แบบจำลองที่พัฒนาจึงไม่ตอบสนองต่อการใช้งานจริงเท่าที่ควรจะเป็น ทั้งนี้ผู้วิเคราะห์ก็ไม่ถึงกับเสียเวลาเปล่า ส่วนใหญ่จะเข้าใจระบบที่ศึกษาอยู่มากขึ้นหลังจากที่ได้พัฒนาแบบจำลองแล้ว และเป็นโอกาสดีที่ได้เก็บข้อมูลเกี่ยวกับระบบที่ศึกษาหรือรับทราบถึงความมีหรือไม่มีของข้อมูล

ถึงแม้ว่าประเด็นของการศึกษาจะชัดเจน บางครั้งผู้วิเคราะห์ใช้เวลาส่วนใหญ่เก็บข้อมูลและพัฒนาแบบจำลอง แต่ใช้เวลาน้อยมากในการใช้แบบจำลองให้เป็นประโยชน์ในแง่วิเคราะห์ผลหรือเปรียบเทียบทางเลือก มักพิจารณาเฉพาะค่าเฉลี่ยแต่ไม่ใช้หลักการทางสถิติ ในกรณีที่ใช้ซอฟต์แวร์สำเร็จรูป ผู้ใช้ส่วนใหญ่ไม่เข้าใจหลักการที่อยู่เบื้องหลังเครื่องมือในซอฟต์แวร์ เช่น ไม่รู้ว่าซอฟต์แวร์เลือกการแจกแจงความน่าจะเป็นให้โดยใช้เกณฑ์อะไร หรือว่าพารามิเตอร์หรือผลต่าง ๆ ที่ซอฟต์แวร์ให้มามีความหมายอย่างไร บางครั้งผู้ใช้มั่นใจในซอฟต์แวร์มากจนไม่ถามตัวเองว่าเครื่องมือนั้น ๆ มีข้อจำกัดอย่างไร

อย่างไรก็ดี มีผู้กล่าวว่าอุปสรรคสำคัญของการจำลองสถานการณ์ไม่ใช่การใช้ที่ผิด แต่ผู้เรียนไม่นำไปใช้กับปัญหาจริง ผู้เขียนจึงขอให้มองข้อผิดพลาดที่นำเสนอมาเป็นแนวทาง ไม่ใช่เพื่อบั่นทอนกำลังใจ

หัวข้อต่อไปกล่าวถึงการตรวจสอบแบบจำลอง ซึ่งประกอบด้วย **การทวนสอบ** (Verification) และ**การตรวจสอบความสมเหตุสมผล** (Validation)

9.2 การทวนสอบ

การทวนสอบเป็นการตรวจสอบว่าแบบจำลองสถานการณ์ในคอมพิวเตอร์นั้นตรงกับแบบจำลองในกรอบความคิด (Conceptual model) หรือแบบจำลองในกระดาษหรือไม่ แวดวงคอมพิวเตอร์เรียกขั้นตอนนี้ว่า Debugging คือการตรวจสอบว่าโปรแกรมประมวลผลถูกต้องหรือไม่ เครื่องมือทั่วไปในการทวนสอบมีดังนี้

1. ให้คนอื่นช่วยตรวจสอบแบบจำลอง หรืออธิบายแบบจำลองให้ผู้อื่นฟังอย่างละเอียด ทีละโมดูล (Structured walk through)

2. ตรวจสอบเอาท์พุทของแบบจำลองเมื่อเทียบกับวิธีการประมาณค่าชนิดอื่น เช่น ตัวแบบแถวคอยในบทที่ 6 แน่นอนว่าตัวเลขที่ได้จากสองวิธีจะไม่เท่ากัน แต่ไม่ควรต่างกันมากนัก

3. ประมวลผลแบบจำลองภายใต้สถานการณ์สุดโต่ง เช่น ให้เวลาในการบริการทุกสถานีเป็นค่าคงที่, ค่าน้อยที่สุดหรือค่ามากที่สุด

4. ให้มีการไหลเข้า 1 ครั้งแล้วติดตามสิ่งนี้ไปตลอดทางในระบบ เช่น สังเกตว่าใช้ทรัพยากรในระบบถูกหรือไม่ เวลาที่ใช้ในแต่ละสถานีเป็นอย่างไร มีเส้นทางการเดินถูกหรือไม่

5. เก็บค่าเอาท์พุทอย่างละเอียด เช่น เวลาเข้าและเวลาออกที่ทุกสถานี

6. หากใช้ซอฟต์แวร์สำเร็จรูปที่มีภาพเคลื่อนไหว ให้สังเกตภาพเคลื่อนไหว เช่น มีแถวคอยใดที่ยาวผิดปกติ ควรใช้ภาพที่แตกต่างกันสำหรับแต่ละสถานะของทรัพยากร เช่น เมื่อเครื่องจักรว่าง, เมื่อทำงาน, หรือเมื่อเสีย

7. วิเคราะห์ความไว (Sensitivity analysis) กับพารามิเตอร์สำคัญ ๆ เพื่อทดสอบว่าเอาท์พุทเปลี่ยนตามในทิศทางที่ควรจะเป็นหรือไม่ เช่น สำหรับระบบผู้ป่วยนอก เมื่อเพิ่มอัตราที่คนไข้ไหลเข้า เวลารอคอยในระบบควรเพิ่มขึ้น หรือเมื่อเพิ่มจำนวนแพทย์ เวลารอคอยในระบบควรลดลง

9.3 การตรวจสอบความสมเหตุสมผล

การตรวจสอบความสมเหตุสมผลทำเพื่อให้มั่นใจว่าแบบจำลองสามารถเป็นตัวแทนของระบบที่ศึกษาในระดับความละเอียดที่เหมาะต่อการตัดสินใจ เช่น ผู้วิเคราะห์สนใจลดเวลารอคอยของคนไข้ผู้ป่วยนอก ก็ไม่จำเป็นต้องจำลองส่วนอื่น ๆ ของโรงพยาบาลอย่างละเอียด เช่น คลังยา ห้องฉายรังสี บางแผนกที่โรงพยาบาลจริง ๆ มี แบบจำลองอาจไม่ต้องมี เช่น แผนกซักผ้าและแผนกโรงครัว ดังนั้นควรพัฒนาแบบจำลองและตรวจความสมเหตุสมผลก็ต่อเมื่อนิยามคำถามที่ต้องการตอบอย่างชัดเจน เครื่องมือทั่วไปในการตรวจสอบความสมเหตุสมผลมีดังนี้

1. นำเสนอแบบจำลองต่อผู้ที่มีความคุ้นเคยกับระบบจริงหรือผู้ที่อยู่หน้างาน โดย

   - อธิบายการประมาณและข้อมูลต่าง ๆ ในภาษาที่คนทั่วไปเข้าใจได้ ใช้ศัพท์วิชาการให้น้อย
   - แสดงภาพเคลื่อนไหว (Animation) ประกอบการนำเสนอ ภาพเคลื่อนไหวเป็นเครื่องมือสื่อสารที่ช่วยให้ผู้ที่ไม่คุ้นเคยกับการจำลองสถานการณ์เข้าใจแบบจำลองได้ง่ายขึ้น พอเข้าใจมากขึ้นความเชื่อถือก็มีมากขึ้น
   - แสดงค่าดัชนีชี้วัดที่ประมวลได้จากแบบจำลองให้กับผู้ที่คุ้นเคยระบบจริง เพื่อให้พิจารณาว่าค่าเหล่านั้นสมเหตุสมผลหรือไม่

2. ตรวจสอบความสมเหตุสมผลของการแปลงค่าจากข้อมูลนำเข้าเป็นเอาท์พุท โดย

- พัฒนาแบบจำลองของระบบปัจจุบันที่เป็นอยู่ ก่อนปรับปรุงหรือเปลี่ยนแปลงระบบ
- เก็บค่าดัชนีชี้วัดของระบบจริง เช่น เวลารอคอยในระบบ หรือเวลารวมในระบบ
- เปรียบเทียบค่าดัชนีชี้วัดที่ประมวลผลด้วยแบบจำลองกับค่าที่เก็บได้จริงด้วยการทดสอบสมมติฐานทางสถิติ (หัวข้อที่ 9.3.1) หรือพิจารณาช่วงความเชื่อมั่น (หัวข้อที่ 9.3.2) เน้นว่าให้ตรวจสอบเอาท์พุท ไม่ใช่อินพุทที่ใช้สร้างแบบจำลอง

การตรวจสอบความสมเหตุสมผลเป็นสิ่งที่ควรทำและทำไม่ยาก แต่มักถูกละเลย จึงทำให้แบบจำลองมีความน่าเชื่อถือน้อยลง สำหรับผู้พัฒนาแบบจำลองเอง หากได้ตรวจสอบความสมเหตุสมผลแล้ว จะทำให้ตัวเองมีความมั่นใจในแบบจำลองของตนมากขึ้น หากตรวจสอบแล้วแต่ไม่ผ่าน จะได้ย้อนไปแก้ไขแบบจำลองให้สมจริงมากขึ้น หัวข้อที่ 9.3.1-9.3.2 นำเสนอด้วยตัวอย่างที่ 9.1 ดังนี้

ตัวอย่างที่ 9.1. ในปัจจุบันร้านค้าทางอินเตอร์เน็ตของสาววิศวกรอุตสาหการคนหนึ่งมีเวลาส่งมอบเฉลี่ย 16.2 ชั่วโมงสำหรับคำสั่งซื้อที่มาทางเว็บ เจ้าของร้านต้องการลดเวลาส่งมอบโดยเปลี่ยนแปลงขั้นตอนการจัดการคำสั่งซื้อ ก่อนที่จะพิจารณาแนวทางในการปรับปรุง ผู้วิเคราะห์ต้องการตรวจสอบว่าแบบจำลองของระบบปัจจุบันที่มีสมจริงหรือไม่ ด้วยสองวิธีข้างล่างนี้

□

หัวข้อที่ 9.3.1 อธิบายการทดสอบทางสถิติที่สามารถใช้เพื่อตรวจสอบความสมจริงของแบบจำลอง

9.3.1 การทดสอบสมมติฐานทางสถิติ

สำหรับตัวอย่างที่ 9.1 พิจารณาสมมติฐานดังต่อไปนี้

H_0: เวลาส่งมอบ = 16.2 ชั่วโมง

H_1: เวลาส่งมอบ $\neq$ 16.2 ชั่วโมง

จากการประมวลผล R รอบทำซ้ำ แต่ละรอบให้ค่าเฉลี่ยเวลาส่งมอบ $Y_1, Y_2, \ldots, Y_R$ ซึ่งสามารถนำไปคำนวณค่าเฉลี่ย $\bar{Y}$ (ด้วยสมการที่ 5.7) และส่วนเบี่ยงเบนมาตรฐาน S (ด้วยสมการที่ 5.8) ถ้าเอาท์พุทจากแบบจำลองมีการแจกแจงใกล้เคียงการแจกแจงแบบปกติ สมมติฐานหลัก (H_0) จะถูกปฏิเสธหรือยอมรับสมมติฐานรอง (H_1) หากว่า

$$\frac{|\bar{Y} - 16.2|}{S/\sqrt{R}} > t_{1-\alpha/2,R-1}$$

เช่น ที่นัยสำคัญ 95% หรือ $\alpha = 0.05$ ค่าสถิติ $t_{0.975,29} = 2.045$

การทดสอบสมมติฐานของค่าเฉลี่ยมีข้อจำกัดเหมือนการทดสอบสมมติฐานทั่วไป คือขึ้นอยู่กับขนาดของข้อมูล (R): หาก R สูงพอ สมมติฐานหลักจะถูกปฏิเสธ แต่ถ้า R ต่ำ จะ“ยอมรับ” (ไม่สามารถปฏิเสธ)

สมมติฐานหลัก ดังนั้นเมื่อนำผลของการทดสอบสมมติฐานไปใช้ ควรคำนึงถึงประเด็นเหล่านี้ด้วย หากขณะนี้ยอมรับสมมติฐานหลัก มิได้หมายความว่าแบบจำลองเป็นสิ่งเดียวกับระบบจริง เพราะถ้าขนาดตัวอย่างใหญ่จะปฏิเสธ H_0 ในที่สุด ในทางตรงกันข้าม หากขณะนี้ปฏิเสธสมมติฐานหลัก อาจไม่ปฏิเสธก็ได้หากขนาดตัวอย่างเล็กกว่านี้ คำถามที่ตามมาคือจำนวนรอบทำซ้ำควรเป็นเท่าใด ควรกำหนด R ให้ใหญ่พอที่จะสามารถแยกแยะความแตกต่างในดัชนีชี้วัดที่มากกว่า ε เมื่อ ε คือความแตกต่างที่มีความสำคัญในเชิงปฏิบัติ ดังที่ได้นำเสนอไปแล้วในหัวข้อที่ 5.3.5 หรืออาจใช้**โค้งปฏิบัติการ** (Operating characteristic curve) เพื่อกำหนดขนาดตัวอย่างที่ควรใช้ก็ได้ นอกจากการทดสอบสมมติฐานแล้ว ผู้วิเคราะห์อาจพิจารณาช่วงความเชื่อมั่นประกอบ

นอกเหนือจากการทดสอบสมมติฐานที่ได้นำเสนอไปในหัวข้อที่ 9.3.1 ผู้วิเคราะห์สามารถสร้างช่วงความเชื่อมั่นของค่าเฉลี่ยประชากร

9.3.2 การพิจารณาช่วงความเชื่อมั่น

สมมติว่าตัวอย่างที่ 9.1 มีช่วงความเชื่อมั่นดังนี้

$$[16.0, 17.6] \text{ ชั่วโมง}$$

เปรียบเทียบกับค่าเฉลี่ยของข้อมูลจริงที่ 16.2 ชั่วโมง พิจารณาขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่น ความผิดพลาดที่มากที่สุดคือ

$$\max(16.2 - 16.0, 17.6 - 16.2) = \max(0.2, 1.4) = 1.4 \text{ ชั่วโมง}$$

หากความผิดพลาดที่มากที่สุดน้อยกว่าความแตกต่างที่มีความสำคัญ (ε) แบบจำลองก็เหมาะสมกับการตัดสินใจที่พิจารณา หากความผิดพลาดที่มากที่สุดมากกว่า ε ก็อาจเพิ่มจำนวนรอบทำซ้ำจนกว่าจะสรุปได้ว่าแบบจำลองเหมาะสมหรือไม่เหมาะสม

ต่อเนื่องจากตัวอย่างที่ 9.1 สมมติว่า $\varepsilon = 0.5$ ชั่วโมง พิจารณาช่วงความเชื่อมั่นสองกรณีดังต่อไปนี้

กรณีที่ 1 ช่วงความเชื่อมั่นเป็น $[16.1, 16.5]$ ความผิดพลาดที่มากที่สุดมีค่า $16.5 - 16.2 = 0.3 < \varepsilon$ หรือ 0.5 ชั่วโมง จึงสรุปว่าแบบจำลองเพียงพอต่อการนำไปใช้

กรณีที่ 2 ช่วงความเชื่อมั่นเป็น $[17.2, 17.6]$ ความผิดพลาดที่น้อยที่สุดมีค่า $17.2 - 16.2 = 1 > \varepsilon$ จึงสรุปว่าแบบจำลองไม่เหมาะต่อการนำไปใช้ ควรทบทวนและแก้ไขข้อสมมติที่ใช้ในการพัฒนาแบบจำลอง

9.4 บทส่งท้าย

ผู้อ่านพึงระลึกไว้ว่าการตรวจสอบความเหมาะสมของแบบจำลองพิจารณาดัชนีชี้วัดหรือเอาท์พุท ไม่ใช่ข้อมูลนำเข้าที่ใช้ในการพัฒนาแบบจำลอง นอกจากวิธีที่ได้กล่าวมาแล้ว อาจใช้ข้อมูลเชิงตัวเลขของจริง (แทนการสุ่มจากการแจกแจงความน่าจะเป็น) ในแบบจำลอง เช่น เวลาตามนาฬิกาที่ลูกค้าเข้ารับบริการ เวลาที่ใช้ในการรับบริการ เพื่อดูว่าเอาท์พุทจากแบบจำลอง (เช่น เวลาในระบบของลูกค้า) ใกล้เคียงกับระบบจริงหรือไม่ วิธีนี้ทำได้สะดวกในปัจจุบันเพราะมีการบันทึกข้อมูลด้วยคอมพิวเตอร์อย่างแพร่หลาย ข้อแนะนำอีกประการคือถึงแม้ผู้วิเคราะห์จะตรวจสอบแล้วว่าแบบจำลองเหมาะสมกับระบบจริง แต่เมื่อเปลี่ยนแปลงสภาวะดำเนินงานเพื่อให้ได้แบบจำลองของระบบที่นำเสนอ ผู้วิเคราะห์ไม่สามารถรับรองว่าแบบจำลองนี้จะเป็นตัวแทนที่ดีของระบบที่นำเสนอ เพราะยังไม่ได้ทดสอบความเหมาะสมของระบบใหม่ที่นำเสนอ

แบบฝึกหัด

แบบจำลองจำลองการเติมก๊าซ LPG ลงรถบรรทุก ณ คลังก๊าซแห่งหนึ่ง วัตถุประสงค์ของการศึกษาคือหาวิธีลดเวลาที่รถบรรทุกอยู่ในระบบ แบบจำลองของระบบปัจจุบันให้ค่าเฉลี่ยของเวลาที่รถอยู่ในคลัง (หน่วย: ชั่วโมง) สำหรับ 6 รอบทำซ้ำ ดังนี้

$$2.79 \;\; 1.12 \;\; 2.24 \;\; 3.45 \;\; 3.13 \;\; 2.38$$

ฐานข้อมูลของคลังระบุว่าเวลาเฉลี่ยที่รถก๊าซอยู่ที่คลังประมาณ 3.35 ชั่วโมง กำหนดให้ความแตกต่างที่มีความสำคัญ $\varepsilon = 1$ ชั่วโมง

9.1. ให้ใช้วิธีทดสอบสมมติฐานเพื่อตรวจสอบว่าแบบจำลองนี้เหมาะสำหรับสถานการณ์คลังก๊าซนี้หรือไม่ ที่นัยสำคัญ 95%

9.2. ให้ใช้วิธีสร้างช่วงความเชื่อมั่นเพื่อตรวจสอบว่าแบบจำลองนี้เหมาะสำหรับสถานการณ์คลังก๊าซนี้หรือไม่ที่นัยสำคัญ 95%

9.3. หากขนาดตัวอย่างน้อยเกินกว่าที่จะสรุปได้ว่าเหมาะสมหรือไม่ ให้ระบุจำนวนรอบทำซ้ำที่ควรใช้

เอกสารอ้างอิง

ประไพศรี สุทัศน์ ณ อยุธยา และ พงศ์ชนัน เหลืองไพบูลย์ 2554. *สถิติวิศวกรรม (ฉบับปรับปรุงใหม่)*. สำนักพิมพ์ท้อป.

สงวน ตั้งโพธิธรรม 2538. *การจำลองด้วยคอมพิวเตอร์ = computer simulations*. ไม่เผยแพร่โดยสำนักพิมพ์.

เสกสรร เกียรติสุไพบูลย์ 2555. *การจำลอง = Simulation*. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริจันทร์ ทองประเสริฐ 2540. *การจำลองแบบปัญหา: Simulation*. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

วุฒิชัย วงษ์ทัศนีย์กร 2555. *การวิเคราะห์แบบจำลอง*. จาก `http://wuthichai.ie.engr.tu.ac.th/Wuthichai_Simulation_201205.pdf` [สืบค้นเมื่อ 13 กรกฎาคม 2555].

Albright, S. C., W. Winston, and C. Zappe. 2009. *Data analysis and decision making with microsoft excel, revised*. 3rd ed. USA: South-Western Cengage Learning.

Ashley, D. W. 2002. An introduction to queuing theory in an interactive text format. *INFORMS Transactions on Education* 2 (3).

Banks, J. (Ed.) 1998. *Handbook of simulation - principles, methodology, advances, applications, and practice*. USA: John Wiley & Sons.

Banks, J., J. S. C. II, B. L. Nelson, and D. M. Nicol. 2005. *Discrete-event system simulation*. 4th ed. USA: Prentice Hall Inc.

Barlow, J. F. 2005. *Excel models for business and operations management*. 2nd ed. USA: John Wiley & Sons Ltd.

Bechhofer, R. E., T. J. Santner, and D. Goldsman. 1995. *Design and analysis of experiments for statistical selection, screening and multiple comparisons*. USA: John Wiley and Sons.

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. Using ranking and selection to "clean up" after simulation optimization. *Operations Research* 51 (5): 814–825.

Brailsford, S. C., E. Silverman, S. Rossiter, J. Bijak, R. J. Shaw, J. Viana, J. Noble, S. Efstathiou, and A. Vlachantoni. 2011. Complex systems modeling for supply and demand in health and social care. In *Proceedings of the 2011 Winter Simulation Conference*, ed. S. Jain, R. Creasey, J. Himmelspach, K. White, and M. Fu, 1125–3116. Piscataway, New Jersey, United States of America: Institute of Electrical and Electronics Engineers, Inc.

Breastcancer.org 2012. Genetics. Available via `http://www.breastcancer.org/risk/factors/genetics.jsp` [accessed August 3, 2012].

Chung, C. A. (Ed.) 2004. *Simulation modeling handbook: A practical approach*. CRC Press.

Devore, J. L. 2011. *Probability and statistics for engineering and the sciences*. 8th ed. Duxbury.

FlightStats, Inc. 2012. Available via `http://www.flightstats.com` [accessed August 17, 2012].

Forbes, C., M. Evans, N. Hastings, and B. Peacock. 2010. *Statistical distributions*. 4th ed. Wiley.

Haas, P. 2006. Lecture notes for MS&E 223 (Introduction to Simulation).

Halverson, R. 2011. Lecture notes for CMPS 5333 (Simulation): Data collection for simulation. Available via `http://cs.mwsu.edu/~ranette/CMPS4223-Sim/Simulation.htm` [accessed November 6, 2012].

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3 (5): 1163–1174.

Hood, G. M. 2010a. Poptools version 3.2.5. Available via `http://www.poptools.org/` [accessed August 3, 2012].

Hood, G. M. 2010b. Poptools version 3.2.5. Available via http://www.poptools.org [accessed October 17, 2012].

Kabacoff, R. I. 2012. Quick R: Accessing the power of R. Available via http://www.statmethods.net/graphs/line.html [accessed August 17, 2012].

Kalvelagen, E. 2003. The newsboy problem. Available via http://chentserver.uwaterloo.ca/aelkamel/che720/che725-process-optimization/GAMS-tutorials/erwin/newsboy.pdf [accessed October 24, 2012].

Kelton, W., R. Sadowski, and N. Swets. 2009. *Simulation with Arena*. USA: McGraw-Hill.

Kelton, W. D., R. P. Sadowski, and N. Swets. 2010. *Simulation with Arena*. 5th ed. USA: McGraw-Hill.

Kendall, D. G. 1953. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics* 24 (3): 338–354.

Kleinrock, L. 1975. *Queueing Systems*, Volume I: Theory. Wiley Interscience. (Published in Russian, 1979. Published in Japanese, 1979. Published in Hungarian, 1979. Published in Italian 1992).

Law, A. M. 2006. *Simulation modeling and analysis*. 4th ed. USA: McGraw-Hill.

L'Ecuyer, P. 1999. Good parameter sets for combined multiple recursive random number generators. *Operations Research* 47 (1): 159–164.

L'Ecuyer, P. 2012. SSJ: Stochastic simulation in Java. Available via http://www.iro.umontreal.ca/~simardr/ssj/indexe.html [accessed August 3, 2012].

Luangmul, K., J. Pichitlamken, and W. Weerawat. 2012. A simulation model of a hospital's clinical laboratory. In *Proceedings of the 4th International Conference on Applied Operational Research (ICAOR)*: Tadbir Operational Research Group.

Montgomery, D. C., and G. C. Runger. 2002. *Applied statistics and probability for engineers*. 3rd ed. USA: McGraw-Hill.

Nance, R. E. 1993. A history of discrete event simulation programming languages. *ACM SIGPLAN Notices* 28 (3): 27.

Nelson, B., J. Swann, D. Goldsman, and W. Song. 2001. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Operations Research* 49 (6): 950–963.

Nelson, B. L. 2003. Lecture notes for IE 335 (Discrete-Event System Simulation).

Nelson, B. L. 2013. *Foundations and methods of stochastic simulation*. USA: Springer.

Peringer, P., D. Leska, and D. Martinek. 2011. SIMulation LIBrary for C++. Available via http://www.fit.vutbr.cz/~peringer/SIMLIB/ [accessed August 3, 2012].

Powell, S. G., and K. R. Baker. 2009. *Management science: The art of modeling with spreadsheets*. 3rd ed. USA: John Wiley & Sons Ltd.

R Development Core Team 2008. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ragsdale, C. T. 2008. *Spreadsheet modeling & decision analysis: A practical introduction to management science*. 5th ed. USA: Thomson South-Western.

Ricci, V. 2005. Fitting distributions with R. Available via http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf [accessed November 24, 2012].

Schmeiser, B. W. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.

Seila, A., V. Ceric, and P. Tadikamalla. 2003. *Applied simulation modeling*. USA: Thomson Learning.

Tamhane, A. C., and D. D. Dunlop. 1999. *Statistics and data analysis: From elementary to intermediate*. USA: Pearson.

The Royal Institute of Thailand 2555. ศัพท์บัญญัติราชบัณฑิตยสถาน. จาก http://rirs3.royin.go.th/coinages/webcoinage.php [สืบค้นเมื่อ 19 กรกฎาคม 2555].

Weerawat, W., J. Pichitlamken, and P. Subsombat. 2013. A generic discrete-event simulation model for outpatient clinics in a Thai public hospital. *Journal of Healthcare Engineering* 4 (2): 285–305.

Welch, B. 1947. The generalization of ”student’s” problem when several different population variances are involved. *Biometrika* 34 (1-2): 28–35.

Wikipedia 2012a. Computer simulation — Wikipedia, the free encyclopedia. Available via `http://en.wikipedia.org/wiki/Computer_simulation`. [accessed June 10, 2012].

Wikipedia 2012b. Spreadsheet. Available via `http://en.wikipedia.org/wiki/Spreadsheet` [accessed October 18, 2012].

Winston, W. L. 2000. *Simulation modeling using @risk*. USA: Wadsworth Publishing Company.

Winston, W. L. 2004. *Introduction to probability models: Operations research, volume ii*. 4th ed. USA: Thomson Brooks/Cole.

Wongsammacheep, T., J. Pichitlamken, and W. Weerawat. 2012. A discrete-event simulation model of a health screening center. In *Proceedings of the 4th International Conference on Applied Operational Research (ICAOR)*: Tadbir Operational Research Group.

ดัชนี