บทความวิจัย

การศึกษาเปรียบเทียบวิธีการคำนวณทฤษฎีบทของเบส์ โดยใช้แบบแผนภาพรูปต้นไม้กับการใช้ตารางแจกแจงความถี่

ชะเอม สายทอง*

*โปรแกรมวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา 1061 ถนนอิสรภาพ แขวงหิรัญรูจี เขตธนบุรี กรุงเทพฯ 10600

แนวคิดทฤษฎีบทของเบส์

การทำนายเหตุการณ์ในอนาคตอย่างหนึ่งเมื่อกำหนดเหตุการณ์มาให้หรือที่เกิดก่อนในการคำนวณค่าความน่าจะเป็นซึ่งเป็นเรื่องที่ทำได้ง่าย เพราะว่าเหตุการณ์ที่กล่าวถึงเป็นเหตุการณ์ที่เกิดต่อเนื่องกันเช่น ถ้าโฆษณาสินค้าย่อมหมายถึงความต้องการที่จะทำให้สินค้านั้นขายได้ดี ดังนั้นค่าความน่าจะเป็นที่สินค้าจะขายได้ดีย่อมมีค่ามากขึ้น หรือถ้าสภาพอากาศในวันนี้มีเมฆมากและความชื้นในอากาศสูงย่อมมีผลให้เกิดฝนตกมากกว่า เรื่องเหล่านี้เป็นการคำนวณโดยใช้ความน่าจะเป็นแบบมีเงื่อนไขซึ่งมีหลักการดังต่อไปนี้

บทนิยาม 1 ความน่าจะเป็นของเหตุการณ์ A เมื่อกำหนดเหตุการณ์ B ให้ แทนด้วยสัญลักษณ์ $P(A|B)$ และ $P(A|B)$ มีค่าดังนี้

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ เมื่อ } P(B) > 0$$

โดยที่ $P(A|B)$ อ่านว่าความน่าจะเป็นของเหตุการณ์ A เมื่อกำหนดเหตุการณ์ B จากบทนิยามจัดสมการ $P(A|B) = \frac{P(A \cap B)}{P(B)}$ เสียใหม่จะได้สูตรสำหรับการหาค่าความน่าจะเป็นแบบมีเงื่อนไขคือ $P(A \cap B) = P(B)\,P(A|B)$

ถ้ากำหนดเหตุการณ์ย่อยๆ เกิดขึ้นก่อนแล้วเหตุการณ์ตามมามีจริง ให้กลับไปทำนายเหตุการณ์ย่อยๆ ที่มาก่อนว่าเหตุการณ์ใดมีความน่าจะเป็นที่เกิดขึ้นมีค่าเท่าไร ตัวอย่างเช่น ผู้จัดการห้างสรรพสินค้าแห่งหนึ่งได้โฆษณาสินค้าในร้านไว้หลายทางเช่น ทางวิทยุ ทางโทรทัศน์ ทางหนังสือพิมพ์ และทางเครือข่ายคอมพิวเตอร์ (internet) ต่อมามีลูกค้าเข้ามาซื้อสินค้า ผู้จัดการต้องการทราบว่าที่ลูกค้าซื้อสินค้า

ในเรื่องของโหราศาสตร์หรือเรียกอย่างชาวบ้านว่าหมอดูก็เป็นอีกตัวอย่างหนึ่ง หมอดูเป็นเรื่องของการคิดที่ใช้วิธีการให้เหตุผลเชิงอุปนัยอย่างหนึ่ง คืออ้างเหตุที่มีหรือที่เกิดขึ้นซ้ำๆ กัน แล้วนำไปทำนายคนที่มาให้ทำนายว่าเข้ากับเหตุการณ์หรือข้อมูล วัน เดือน ปีเกิด ของคนที่มีเหตุการณ์ที่เป็นตัวอย่างมาแล้ว จึงทำนายไปตามนั้น ซึ่งในบางครั้งก็อาจจะไม่จริงก็ได้ การทำนายไปแบบนั้นเป็นเพียงความน่าจะเป็นมากกว่า ยกตัวอย่างเช่น ชายคนหนึ่งมีสีหน้าเศร้า เดินเข้าไปหาหมอดูเพื่อต้องการให้หมอดูทำนาย ซึ่งหมอดูมักจะรู้โดยประสบการณ์ว่าผู้ที่มาให้ดูมักจะมีเรื่องทุกข์ร้อนแน่นอน หมอดูจะต้องถามวัน เดือน ปีเกิด แน่นอน และเปิดตำราซึ่งเขียนไว้โดยอาศัยวิธีการให้เหตุผลเชิงอุปนัย จากนั้นจึงทำนายไปว่าสาเหตุของการมีเรื่องทุกข์ร้อนเป็นทางใดเช่น อาจจะเป็นทางที่เขาเจ็บป่วย มีอุบัติเหตุเกิดขึ้น โดนไล่ออกจากงาน หรือมีหนี้สินรุงรัง เป็นต้น โดยหมอดูจะทำนายไปในทางใดทางหนึ่ง การทำนายของหมอดูนี้อาจจะจริงหรือไม่ก็ได้ เป็นเรื่องของความน่าจะเป็นมากกว่า ถ้าบังเอิญทำนายไปตรงกับความจริงของชายคนนั้น ก็เรียกว่าความน่าจะเป็นมีสูง ทำให้หมอดูทายแม่นมาก

ตัวอย่างอื่นๆ เช่น สมมุติว่ามีรถทัศนาจรคันหนึ่งเกิดอุบัติเหตุชนราวสะพานรถยนต์เสียหลักตกลงไปในแม่น้ำมีผู้บาดเจ็บและตายหลายคน ต่อมาทราบว่าคนขับไปตายที่โรงพยาบาล ตำรวจและแพทย์ได้ช่วยกันตรวจสถานที่เกิดอุบัติเหตุ พิสูจน์ศพคนขับรถสอบถามผู้ที่รอดชีวิต จึงได้ตั้งประเด็นไว้ว่าคนขับเมาสุรา หลับใน หรือเบรครถยนต์ชำรุด ปัญหาคือ เมื่อทราบว่ามีอุบัติเหตุ คนขับตาย ความน่าจะเป็นที่คนขับเมาสุรา ความน่าจะเป็นที่คนขับหลับใน หรือความน่าจะเป็นที่เบรครถยนต์ชำรุด อย่างไหนจะมีค่าสูงกว่ากัน นี่เป็นปัญหาที่ต้องหาข้อมูลหลักฐานมายืนยัน การคำนวณปัญหาอย่างนี้การใช้ทฤษฎีบทของเบส์เป็นทางหนึ่งที่จะช่วยได้

ทฤษฎีบทของเบส์ (Bayes theorem) คิดขึ้นโดยนักคณิตศาสตร์ชาวอังกฤษชื่อโทมาส เบส์ (Thomas Bayes) ซึ่งมีชีวิตอยู่ระหว่าง ค.ศ. 1702-1761 ได้ให้คำตอบในการหาค่าความน่าจะเป็นว่าเหตุการณ์ใดๆ (ที่เป็นเหตุการณ์ย่อยๆ) เกิดขึ้นด้วยความน่าจะเป็นเท่าไรเมื่อทราบเหตุการณ์ที่เกิดขึ้นในภายหลัง ซึ่งเรียกว่าความน่าจะเป็นแบบมีเงื่อนไข ผลงานของเบส์เกี่ยวกับเรื่องนี้ได้ตีพิมพ์เผยแพร่เมื่อ ค.ศ. 1765 หลักพื้นฐานของทฤษฎีบทของเบส์จะต้องรู้ความหมายของผลแบ่งกั้น (partition) ซึ่งมีหลักการดังต่อไปนี้

บทนิยาม 2. $B_1, B_2, B_3, \ldots, B_n$ จะเป็นเหตุการณ์ที่เป็นผลแบ่งกั้นของปริภูมิตัวอย่าง S ถ้าเงื่อนไขต่อไปนี้เป็นจริง

1) $B_1, B_2, B_3, \ldots, B_n$ เป็นเหตุการณ์ไม่เกิดร่วมกัน

2) $S = \bigcup_{i=1}^{n} B_i$

3) $P(B_i) > 0$ เมื่อ $i = 1, 2, \ldots, n$

ตัวอย่าง 1. ในการโยนลูกเต๋าลูกหนึ่งที่มี 6 หน้า แต่ละหน้าเขียนเลข 1, 2 , …, 6 ไว้ตามลำดับ ดังนั้น

$S = \{1, 2, \ldots, 6\}$ ถ้าให้ $B_1 = \{1, 2\}$ $B_2 = \{3\}$ และ $B_3 = \{4, 5, 6\}$ จะได้ว่า B_1, B_2 และ B_3 เป็นผลแบ่งกั้นของ S เพราะว่า

1) B_1, B_2 และ B_3 เป็นเหตุการณ์ที่ไม่เกิดร่วม

2) $B_1 \cup B_2 \cup B_3 = S$

3) $P(B_i) > 0$ เมื่อ $i = 1, 2, 3$

ทฤษฎีบทของเบส์ กล่าวไว้ดังต่อไปนี้

ทฤษฎีบทของเบส์ ถ้า $B_1, B_2, \ldots, B_n$ เป็นเหตุการณ์ n เหตุการณ์ที่เป็นผลแบ่งกั้นของปริภูมิตัวอย่าง S แล้ว สำหรับเหตุการณ์ A ใดๆ ใน S จะได้

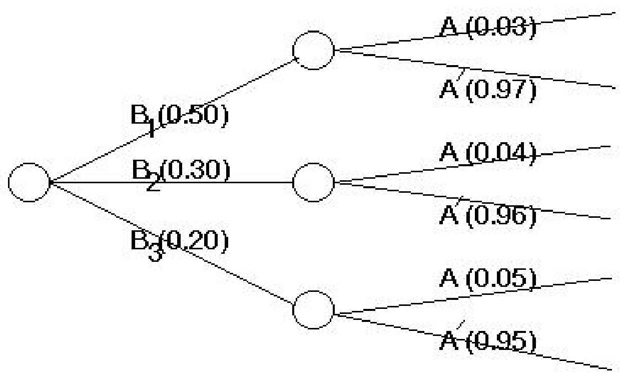
$$P(B_j | A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

การคำนวณความน่าจะเป็นตามทฤษฎีบทของเบส์ มีหลายวิธีในที่นี้จะเสนอแบบที่เป็นตัวช่วยคำนวณ 2 รูปแบบคือ (1) แผนภาพรูปต้นไม้ และ (2) ตารางแจกแจงความถี่ ซึ่งทั้ง 2 แบบมีวิธีการคำนวณดังตัวอย่างต่อไปนี้

ตัวอย่าง 2. สมมุติว่าเครื่องจักร ก, ข และ ค ของโรงงานแห่งหนึ่งมีกำลังผลิตสินค้าได้ 50, 30 และ 20 เปอร์เซ็นต์ ของจำนวนสินค้าที่ผลิตได้จากโรงงานตามลำดับ และปรากฏว่าสินค้าดังกล่าวเมื่อส่งออกไปจำหน่าย มีสินค้าที่ชำรุดที่ผลิตมาจากเครื่องจักร ก, ข และ ค เป็นจำนวน 3, 4 และ 5 เปอร์เซ็นต์ตามลำดับ ถ้าสุ่มเลือกสินค้าขึ้นมา 1 ชิ้น จงหา (1) ความน่าจะเป็นที่ได้สินค้าที่ชำรุด (2) ถ้าหยิบสินค้าขึ้นมา 1 ชิ้นและพบว่าเป็นสินค้าที่ชำรุด จงหาความน่าจะเป็นที่ได้สินค้านั้นมาจากการผลิตโดยเครื่องจักร ก

วิธีทำ

วิธีที่ 1. ใช้แผนภาพรูปต้นไม้ ให้ B_1, B_2 และ B_3 เป็นเหตุการณ์ที่เลือกได้สินค้าที่ผลิตจากเครื่องจักร ก, ข และ ค ตามลำดับ A เป็นเหตุการณ์ที่เลือกได้สินค้าที่ชำรุด ดังนั้น A' คือเหตุการณ์ที่เลือกได้สินค้าที่ไม่ชำรุด สร้างแผนภาพรูปต้นไม้แสดงเหตุการณ์ที่ต่อเนื่องกันดังภาพที่ 1



ภาพที่ 1. แผนภาพรูปต้นไม้แสดงเหตุการณ์ที่ต่อเนื่องกัน

1) ตามหลักการความน่าจะเป็นแบบมีเงื่อนไข ดังนั้น

P(A) = ความน่าจะเป็นที่ได้สินค้าที่ชำรุดที่ผลิตจากเครื่องจักร ก, ข หรือ ค

$= \sum_{i=1}^{3} P(B_i)P(A|B_i)$

$= (0.50)(0.03) + (0.30)(0.04) + (0.20)(0.05)$

$= 0.037$

นั่นคือ ความน่าจะเป็นที่ได้สินค้าที่ชำรุดเท่ากับ 0.037

2) จากทฤษฎีบทของเบส์จะได้

$P(B_1|A)$ = ความน่าจะเป็นที่ได้สินค้ามาจากการผลิตโดยเครื่องจักร ก ถ้าสินค้าที่หยิบมา 1 ชิ้น พบว่าเป็นสินค้าที่ชำรุด

$$= \frac{P(B_1)P(A|B_1)}{\sum_{i=1}^{3} P(B_i)P(A|B_i)}$$

$$= \frac{(0.50)(0.03)}{0.037}$$

$= 0.4054$

วิธีที่ 2. ใช้ตารางแจกแจงความถี่ สมมุติว่ามีสินค้าที่สำรวจ 1,000 ชิ้น คำนวณหาความถี่ของเหตุการณ์ที่เกี่ยวข้องกัน ได้ดังตารางที่ 1

ตารางที่ 1. ตารางแจกแจงความถี่ของสินค้าที่สำรวจ 1,000 ชิ้น

จำนวนสมาชิกของ	A	A'	รวม
B_1	15	485	500
B_2	12	288	300
B_3	10	190	200
รวม	37	963	1,000

1) P(A) = ความน่าจะเป็นที่ได้สินค้าที่ชำรุดที่ผลิตจากเครื่องจักร ก, ข หรือ ค

$= \frac{37}{1,000}$

$= 0.037$

2) $P(B_1|A)$ = ความน่าจะเป็นที่ได้สินค้ามาจากการผลิตโดยเครื่องจักร ก ถ้าสินค้าที่หยิบมา 1 ชิ้น พบว่าเป็นสินค้าที่ชำรุด

$= \frac{15}{37}$

$= 0.4054$

ตัวอย่าง 3. ในสวนสาธารณะแห่งหนึ่งมีม้านั่งอยู่จำนวน 1,000 ตัว ในจำนวนนี้เป็นม้านั่งที่ทำด้วยไม้ 40% ที่เหลือเป็นม้านั่งที่ทำด้วยหิน ถ้าในแต่ละวัน 4% ของม้านั่งที่ทำด้วยไม้มีผู้คนนั่ง และ 1% ของม้านั่งที่ทำด้วยหินมีผู้คนนั่ง ถ้าท่านเห็นคนนั่งม้านั่งอยู่คนหนึ่งไม่ทราบว่าเป็นม้านั่งที่ทำด้วยไม้หรือทำด้วยหิน จงหาความน่าจะเป็นที่ม้านั่งที่คนนั่งอยู่นั้นเป็นม้านั่งที่ทำด้วยไม้

วิธีทำ ให้ A เป็นเหตุการณ์ที่พบคนนั่งม้านั่ง

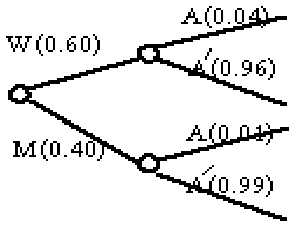
ดังนั้น A' เป็นเหตุการณ์ที่ไม่พบคนนั่งม้านั่ง

W เป็นเหตุการณ์ที่ที่เลือกได้ม้านั่งที่ทำด้วยไม้

M เป็นเหตุการณ์ที่ที่เลือกได้ม้านั่งที่ทำด้วยหิน

ถ้าต้องการหาความน่าจะเป็นที่ม้านั่งทำด้วยไม้เมื่อเห็นคนนั่งอยู่ นั่นคือต้องการหาค่าของ P(W|A)

วิธีที่ 1. ใช้แผนภาพรูปต้นไม้ สร้างรูปต้นไม้แสดงเหตุการณ์ได้ดังภาพที่ 2



ภาพที่ 2. รูปต้นไม้แสดงเหตุการณ์

จากทฤษฎีบทของเบส์จะได้

$$P(W|A) = \frac{P(W)\,P(A|W)}{P(W)\,P(A|W) + P(M)\,P(A|M)}$$

$$= \frac{(0.60)(0.04)}{(0.60)(0.04) + (0.40)(0.01)}$$

$$= \frac{0.024}{0.028}$$

$$= 0.8571$$

วิธีที่ 2. ใช้ตารางแจกแจงความถี่ สมมุติว่ามีผู้คนเข้ามาในสวนสาธารณะจำนวน 1,000 คน สร้างตารางแจกแจงความถี่แสดงเหตุการณ์ที่เกี่ยวข้องกันได้ดังตารางที่ 2

ตารางที่ 2. ตารางแจกแจงความถี่ของผู้คนเข้ามาในสวนสาธารณะจำนวน 1,000 คน

จำนวนสมาชิกของ	A	A'	รวม
W	24	576	600
M	4	396	400
รวม	28	972	1,000

จากตารางที่ 2 จะได้

$$P(W|A) = \frac{\text{จำนวนสมาชิกของ W ที่อยู่ใน A}}{\text{จำนวนสมาชิกของ A}}$$

$$= \frac{24}{28}$$

$$= 0.8571$$

การศึกษาเปรียบเทียบวิธีการคำนวณโดยใช้แบบแผนภาพรูปต้นไม้กับการใช้ตารางแจกแจงความถี่

เนื่องจากผู้เขียนบทความนี้ได้สอนรายวิชาความน่าจะเป็นและสถิติเบื้องต้น และรายวิชาสถิติเชิงคณิตศาสตร์ เห็นว่าการสอนในเรื่องทฤษฎีของเบส์นักศึกษาเข้าใจยาก โดยเฉพาะในเรื่องการคำนวณค่าความน่าจะเป็นในทฤษฎีของเบส์ จึงได้สร้างรูปแบบการคำนวณ 2 วิธีคือ การคิดคำนวณทฤษฎีบทของเบส์โดยใช้แบบแผนภาพรูปต้นไม้ กับแบบการใช้ตารางแจกแจงความถี่ การคำนวณค่าความน่าจะเป็นดังกล่าวน่าจะเป็นตัวช่วยที่ทำได้ง่าย จึงได้ทดลองนำวิธีการทั้ง 2 ไปทดลองสอนให้นักศึกษาใช้คิดคำนวณ และหลังจากนั้นได้ทดสอบนักศึกษาที่เรียน จำนวน 16 คน ผลการทดสอบได้คะแนนดังตารางที่ 3 ในที่นี้ X แทนคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้ และ Y แทนคะแนนสอบโดยการคำนวณแบบตารางความถี่ โดยตั้งสมมุติฐานว่า นักศึกษาจะสอบได้คะแนนโดยการคำนวณแบบตารางแจกแจงความถี่สูงกว่าคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้ จากนั้นได้ทำการวิเคราะห์ข้อมูลโดยใช้ค่าสถิติ t สำหรับข้อมูลคู่และไม่เป็นอิสระแก่กัน

การวิเคราะห์ข้อมูล

กำหนดให้ d คือผลต่างระหว่างคะแนนสอบจากการคำนวณแบบตารางความถี่ กับคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้ ซึ่งในที่นี้ $d = Y - X$

1) สมมุติฐานที่ตั้ง

Ho : $\mu_d = 0$ (คะแนนสอบโดยการคำนวณแบบตารางความถี่มีค่าเท่ากับคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้)

Ha : $\mu_d > 0$ (คะแนนสอบโดยการคำนวณแบบตารางความถี่มีแนวโน้มจะสูงกว่าคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้)

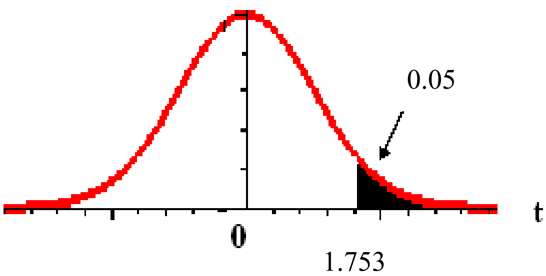
ตารางที่ 3. ผลการทดสอบการคำนวณทฤษฎีบทของเบส์จากนักศึกษาจำนวน 16 คน

คนที่	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X	4	2	0	0	0	18	0	0	4	3	8	5	3	15	6	6
Y	9	5	2	4	2	19	10	10	3	0	2	0	0	16	10	5

2) เลือกค่าสถิติที่ใช้ เนื่องจากข้อมูลคู่ และไม่เป็นอิสระแก่กัน ใช้ค่า t สำหรับการทดสอบ

จากสูตร $t_{cal} = \dfrac{\bar{d}-\mu_d}{\frac{S_d}{\sqrt{n}}}$

3) กำหนดระดับนัยสำคัญที่ $\alpha = 0.05$ จากตารางการแจกแจง t ที่ระดับนัยสำคัญ 0.01 และระดับขั้นแห่งความเสรี (df) = 16 – 1 = 15 ได้ค่าวิกฤติ t เท่ากับ 1.753 แสดงดังภาพที่ 4



ภาพที่ 4. ค่าวิกฤติ t

4) การคำนวณ จะหาค่าสถิติ t จากสูตร

$t_{cal} = \dfrac{\bar{d}-\mu_d}{\frac{S_d}{\sqrt{n}}}$ ดังนั้นจึงต้องหาค่าของ $\bar{d}$ และ S_d สร้างตารางแสดงค่าของ d และ d^2 ซึ่ง $d = Y - X$ ได้ดังตารางที่ 4 คำนวณได้ $\sum d = 23$ และ $\sum d^2 = 357$ ใช้ค่าที่ได้คำนวณค่า $\bar{d}$ และ S_d ได้ดังต่อไปนี้

$$\bar{d} = \frac{\Sigma d}{n} = \frac{23}{16} = 1.44$$

$$S_d = \sqrt{\frac{n(\Sigma d^2)-(\Sigma d)^2}{n(n-1)}}$$

$$= \sqrt{\frac{16(357)-(23)^2}{16(16-1)}} = 4.65$$

ตารางที่ 4. ค่าของ d และ d^2 ของนักศึกษาที่ถูกทดสอบจำนวน 16 คน

นักศึกษาคนที่	X	Y	$d = Y - X$	d^2
1	4	9	5	25
2	2	5	3	9
3	0	2	2	4
4	0	4	4	16
5	0	2	2	4
6	18	19	1	1
7	0	10	10	100
8	0	10	10	100
9	4	3	– 1	1
10	3	0	– 3	9
11	8	2	– 6	36
12	5	0	– 5	25
13	3	0	– 3	9
14	15	16	1	1
15	6	10	4	16
16	6	5	– 1	1
รวม			23	357

คำนวณได้ค่า $t_{cal} = \dfrac{\bar{d}-\mu_d}{\frac{S_d}{\sqrt{n}}}$

$$= \frac{1.44-0}{\frac{4.65}{\sqrt{16}}} = 1.24$$

5) ผลสรุป การคำนวณได้ $t_{cal} = 1.24$ เปรียบเทียบกับค่าวิกฤติ ได้ $1.24 < 1.753$ จึงยอมรับ Ho ที่ว่า $\mu_d = 0$

การแปลผล

ผลจากการวิเคราะห์ข้อมูล สรุปได้ว่าคะแนนสอบโดยการคำนวณแบบตารางความถี่ไม่สูงไปกว่าคะแนนสอบโดยการคำนวณแบบใช้รูปต้นไม้อย่างมีนัยสำคัญ