ทฤษฎีตัดสินใจทางสถิติที่อาศัยการสุ่มตัวอย่าง

Statistical Decision Theory with Sampling

การวิเคราะห์การตัดสินใจแบบชั้นเดียว หรือหลายชั้นที่กล่าวมานั้น เราสมมติว่าผู้ตัดสินใจไม่ได้รับข้อมูลข่าวสารเพิ่มเติมเกี่ยวกับสภาวการณ์นอกบังคับ เพียงแต่สมมติการแจกแจงน่าจะเป็นก่อนทดลองของ W ขึ้น การแจกแจงน่าจะเป็นก่อนทดลองนี้จะแสดงความเชื่อข้อผู้ตัดสินใจในข้อมูลข่าวสารเกี่ยวกับการเกิดขึ้นของสภาวการณ์ขณะที่ทำการตัดสินใจ ต่อไปเราจะละทิ้งข้อตกลงนั้นโดยสมมติว่าผู้ตัดสินใจสามารถจะหาข้อมูลข่าวสารเพิ่มเติมเกี่ยวกับสภาวการณ์นอกบังคับจากการกระทำการทดลองเชิงสุ่มได้ ผู้ตัดสินใจจะสังเกตผลทดลองที่เป็นไปได้ทั้งหมดของการทดลอง และพิจารณาทางเลือกที่เหมาะสมของแต่ละผลทดลองที่เป็นไปได้ การวิเคราะห์เกี่ยวกับการเลือกทางเลือกที่เหมาะสมเช่นนี้เราเรียกว่า "การวิเคราะห์การตัดสินใจด้วยตัวอย่าง" ซึ่งมีแบบของการวิเคราะห์อยู่ 2 แบบ คือ – แบบปกติ (Normal Form) และ แบบขยาย (Extensive Form) แต่ละแบบก็ใช้พิจารณาทางเลือกที่เหมาะสมจากแต่ละผลทดลองที่เป็นไปได้ของการทดลอง การวิเคราะห์ทั้งสองแบบนี้ในเชิงคณิตศาสตร์จะเหมือนกัน แต่วิธีการจะแตกต่างกัน

1. <u>การวิเคราะห์แบบปกติ</u> (Normal Form Analysis)

เราจะพิจารณาตัวอย่างเกี่ยวกับการซื้อชิ้นส่วนเครื่องจักร 1000 ชิ้นที่แล้ว จากตัวอย่างนั้นเรามีกลุ่มทางเลือกเป็น

$$A = \{a_1, a_2\}$$

ในเมื่อ a_1 คือ "ไม่ซื้อ" และ a_2 คือ "ซื้อ" ชิ้นส่วนเครื่องจักร 1000 ชิ้นนั้น และกลุ่มสภาวการณ์เป็น

$$W = \{w_1, w_2\}$$

ในเมื่อ w_1 เป็น "ใช้การไม่ได้เป็น 20 %" และ w_2 เป็น "ใช้การไม่ได้ 50 %" ค่าเสียโอกาสสำหรับ (a_i, w_j) และความน่าจะเป็นก่อนทดลองแสดงได้ดังตาราง

<u>ตาราง</u> 1 ค่าเสียโอกาส และความน่าจะเป็นก่อนทดลอง

W		w_1	w_2
		$P(w_1) = .6$	$P(w_2) = .4$
A	a_1	1600	0
	a_2	0	500

จากข้อมูลข่าวสารเหล่านี้ เราพิจารณาได้ว่าผู้จัดการควรจะเลือกทางเลือก a_2 สมมติว่า เราสามารถที่จะหาข้อมูลข่าวสารเพิ่มเติมเกี่ยวกับการเกิดขึ้นของสภาวการณ์โดยการกระทำการทดลองเชิงสุ่ม สมมติว่าผู้ผลิตจัดหาชิ้นส่วน 1000 ชิ้น (1 ลัง) ให้แก่ผู้จัดการ และผู้จัดการก็สามารถสุ่มตัวอย่างของชิ้นส่วนนั้นมาจำนวนหนึ่ง ซึ่งสามารถหาจำนวนที่ใช้การไม่ได้ในตัวอย่างนั้น ผู้จัดการจะรวมข้อมูลข่าวสารเหล่านี้เพื่อใช้ในการวิเคราะห์พิจารณาหาทางเลือกที่เหมาะสมได้อย่างไร ? เราจะตอบปัญหาด้วยการพัฒนาวิธีวิเคราะห์แบบปกติ ดังนี้

2. <u>กลุ่มผลทดลอง</u> (Sample Description Space)

สมมติว่าผู้จัดการพิจารณาขนาดตัวอย่างสุ่มเพียง 2 ชิ้นเท่านั้น (ปัญหาเกี่ยวกับตัวอย่างขนาด 2 นี้จะเหมาะสมหรือไม่ จะได้อธิบายกันต่อไป) และให้ X_1, X_2 เป็นตัวแปรเชิงสุ่มอิสระสองตัวที่มีคุณสมบัติว่า

สำหรับ $i = 1, 2$; $X_i = 0$ ถ้าชิ้นส่วนใช้การไม่ได้
$= 1$ ถ้าชิ้นส่วนใช้การได้

แล้วตัวอย่างสุ่มที่ผู้จัดการสุ่มได้จึงสามารถแทนได้เป็น X_1, X_2 โดยปกติผู้จัดการก็สนใจที่จะทราบจำนวนชิ้นที่ใช้การไม่ได้ในตัวอย่างนั้น ฉะนั้นในตัวอย่างจะมี $\sum X_i = s$ ถ้าตัวอย่างนั้นมีชิ้นที่ใช้การไม่ได้เป็น s เราจะแทนเซทของผลทดลองที่เป็นไปได้ในตัวอย่างด้วย S ดังนั้นจากปัญหาของเราจะได้

$$S = \{0, 1, 2\}$$

นั่นคือถ้าผู้จัดการพบชิ้นที่ใช้การไม่ได้ 0 เขาจะได้ $\sum X_i = 0$ ถ้าพบ 1 หน่วย $\sum X_i = 1$ และพบ 2 หน่วย $\sum X_i = 2$ เราจะเรียกเซท S ว่า "กลุ่มผลทดลอง" โดยทั่วไป ถ้า $s_1, s_2, \ldots, s_k$ เป็นผลทดลองของตัวอย่างสุ่ม แล้วกลุ่มผลทดลอง S จะกำหนดไว้เป็น

$$S = \{s_1, s_2, \ldots, s_k\}$$

3. <u>ฟังก์ชันน่าจะเป็น</u> (Likelihood Function)

การเกิดขึ้นของผลทดลอง $s_1, s_2, \ldots, s_k$ นั้นขึ้นอยู่กับสภาวการณ์นอกบังคับ W เพราะฉะนั้นสำหรับ w ใน W ที่กำหนดให้ เราจะให้ p_w เป็นฟังก์ชันมวลน่าจะเป็นของผลทดลองใน S ฟังก์ชันนี้จะเรียกว่า "ฟังก์ชันน่าจะเป็น" ของ W ซึ่งจะแทนประจักษ์พยานของการทดลอง

ในตัวอย่างของเรา สำหรับทุก ๆ w ใน w_1 , w_2 เราจะมี

$$p_w(s) \geq 0, \quad s \in \{0, 1, 2\}, \quad \text{และ} \sum p_w(s) = 1$$

สำหรับรูปฟอร์มของ p_w นั้นอาจจะเป็นฟังก์ชันหนาแน่นน่าจะเป็นชนิดต่อเนื่อง หรือเป็นฟังก์ชันมวลน่าจะเป็นถ้าผลทดลองเป็นตัวแปรเชิงสุ่มไม่ต่อเนื่อง ในทำนองเดียวกัน ถ้าการเกิดขึ้นของผลทดลองสอดคล้องกับข้อสมมติหรือข้อตกลงบางอย่าง แล้ว p_w อาจจะเป็นแบบหนึ่งของการแจกแจงชนิดมาตรฐาน (ทั้งแบบต่อเนื่อง หรือไม่ต่อเนื่อง) เช่นถ้าผลทดลองแบ่งประเภทได้เป็น "ใช้ไม่ได้" หรือ "ใช้ได้" และถ้าเราทราบขนาดของตัวอย่าง กับความน่าจะเป็นที่จะเกิดขึ้นของ "ใช้ไม่ได้" แล้วเราสามารถสมมติ p_w ว่าเป็นแบบทวินามได้ จากตัวอย่างของเรา ผู้จัดการได้ดึงชิ้นส่วนมา 2 ชิ้น และชิ้นส่วนสามารถแยกได้เป็น ใช้การได้ หรือใช้การไม่ได้ ตามข้อสมมติที่ว่าลังนั้นอาจจะบรรจุของใช้การไม่ได้ 20 % (w_1) หรือไม่ก็ 50 % (w_2) เพราะฉะนั้นเมื่อตัวอย่างขนาด 2 และความน่าจะเป็นที่จะได้ของที่ใช้การไม่ได้ 1 ชิ้นเป็น p_{w_1} ถ้า w_1 เป็นสภาวการณ์ที่แท้จริง และความน่าจะเป็นนี้คงที่ตลอดการทดลอง ในทางตรงกันข้าม ถ้า w_2 เป็นสภาวการณ์ที่แท้จริง ข้อสมมติที่เหมือนกันก็เป็นจริง ดังนั้นเราสามารถบอกได้ว่า p_{w_1} เป็นแบบทวินาม นั่นคือ

$$p_{w_1}(s) = b(s;\ 2, w_1)$$

= ความน่าจะเป็นแบบทวินามที่จะได้ของที่ใช้การไม่ได้หน่วยจากตัวอย่างขนาด 2 เมื่อกำหนด w_1 ให้

ในทำนองเดียวกัน เราได้ $p_{w_2}(s) = b(s;\ 2, w_2)$

สำหรับตัวอย่างของเรานั้น s จะมีค่าที่เป็นไปได้ดังนี้ $s = 0,\ 1,\ 2$ เพราะฉะนั้น

$$b(0;\ 2, .2) = .64 \qquad b(1;\ 2, .2) = .32$$

$$b(2;\ 2, .2) = .04$$

ทำนองเดียวกัน $b(0;\ 2,.5) = .25$ $\quad b(1;\ 2,.5) = .50$

$$b(2;\ 2,.5) = .25$$

ความน่าจะเป็นเหล่านี้สรุปได้ดังตารางต่อไปนี้

<u>ตาราง</u> 2 ฟังก์ชันน่าจะเป็น

	S	0	1	2
W	w_1	.64	.32	.04
	w_2	.25	.50	.25

โดยอาศัยการสังเกตผลทดลองของการทดลองเชิงสุ่ม ผู้ตัดสินใจก็จะเลือกทางเลือกได้ นี่ก็เป็นลักษณะที่ต่างออกไปอีกลักษณะหนึ่งของการวิเคราะห์การตัดสินใจด้วยการสุ่มตัวอย่าง ในการวิเคราะห์การตัดสินใจแบบขั้นเดียวนั้น ผู้ตัดสินใจจะเลือกทางเลือกโดยไม่ต้องทำการทดลอง แต่ในการวิเคราะห์การตัดสินใจด้วยการสุ่มตัวอย่างนั้นเขาต้องสังเกตผลทดลอง และทำการเลือกทางเลือกโดยใช้การเกิดขึ้นของผลทดลองเป็นหลัก ขบวนการนี้จะให้ลักษณะอื่น ๆ ของการวิเคราะห์ นั่นคือแนวคิดเกี่ยวกับกฎตัดสินใจ หรือฟังก์ชันตัดสินใจ

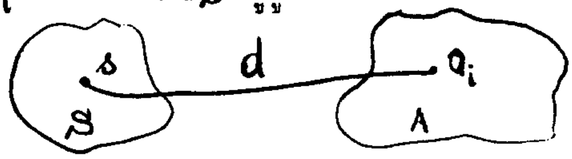
4. <u>กฎตัดสินใจ</u> (Decision Rules)

กฎตัดสินใจเป็นฟังก์ชันที่เกี่ยวพันระหว่างผลทดลองของการทดลองเชิงสุ่ม กับทางเลือกในกลุ่มทางเลือก ดังนั้นกฎตัดสินใจอาจจะคิดว่าเป็นฟังก์ชันที่กำหนดในกลุ่มผลทดลอง S เราจะนิยามกฎตัดสินใจในรูปแบบดังนี้

<u>กฎตัดสินใจ</u>**:** ให้ S เป็นกลุ่มผลทดลอง และให้ A เป็นกลุ่มทางเลือกสำหรับปัญหาตัดสินใจนั้น แล้วฟังก์ชัน d ซึ่งกำหนดใน S และแปลงเข้าไปใน A จะเรียกว่า "กฎตัดสินใจ" โดยสัญลักษณ์จะแทนด้วย

$$d : S \longrightarrow A$$

ดังนั้นกฎตัดสินใจจะเกี่ยวพันระหว่างทางเลือก a_i ใน A กับทุก ๆ ผลทดลอง s ใน S นั่นคือ $d(s) = a_i$ สำหรับ $s \in S$ ดูรูปภาพต่อไปนี้



ให้ D เป็นเซทของกฎตัดสินใจทั้งหมดของปัญหาตัดสินใจ ซึ่งจะเรียกว่า "กลุ่มตัดสินใจ (Decision Space)" เราจะสังเกตเห็นได้ว่าอาจจะมีกฎตัดสินใจจำนวนมากที่เกี่ยวกับระหว่างกลุ่มผลทดลอง S กับกลุ่มทางเลือก A จำนวนจริง ๆ ของกฎตัดสินใจนั้นขึ้นอยู่กับจำนวนของผลทดลองในกลุ่มผลทดลอง และจำนวนทางเลือกในกลุ่มทางเลือก ในตัวอย่างของเรา A มี 2 ทางเลือก และ S มี 3 ผลทดลอง ดังนั้นจึงมีกฎตัดสินใจทั้งหมดจำนวน 2^3 หรือ 8 ดังตารางต่อไปนี้

ตาราง 3 เซทของกฎตัดสินใจทั้งหมดในปัญหาตัวอย่าง

D		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
S	0	a_1	a_1	a_1	a_1	a_2	a_2	a_2	a_2
	1	a_1	a_1	a_2	a_2	a_1	a_1	a_2	a_2
	2	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2

โดยทั่วไป ถ้า $d: S \rightarrow A$ แล้วจะมีจำนวนกฎตัดสินใจทั้งหมดเท่ากับ $[n(A)]^{n(S)}$ ในเมื่อ $n(A)$ แทนจำนวนทางเลือกทั้งหมดใน A และ $n(S)$ แทนจำนวนผลทดลองทั้งหมดใน S ตารางต่อไปนี้จะแสดงถึงกลุ่มพวก (class) ของกฎตัดสินใจทั้งหมดสำหรับ A และ S ต่าง ๆ

ตาราง 4 จำนวนกฎตัดสินใจสำหรับ A และ S ต่าง ๆ

A	S	(A)	n(S)	n(D)
a_1, a_2	s_1, s_2	2	2	4
a_1, a_2	s_1, s_2, s_3	2	3	8
a_1, a_2	s_1, s_2, s_3, s_4	2	4	16
a_1, a_2, a_3	s_1, s_2	3	2	9
a_1, a_2, a_3	s_1, s_2, s_3	3	3	27
a_1, a_2, a_3	s_1, s_2, s_3, s_4	3	4	81
	$\vdots$			
$a_1, a_2, \ldots, a_N$	$s_1, s_2, \ldots, s_M$	N	M	N^M

เราจะเห็นได้ว่ากฎตัดสินใจ d นี้ก็เป็นตัวแปรเชิงสุ่ม เพราะ d กำหนดจากกลุ่มผลทดลอง S ที่เกี่ยวพันกับผลทดลองเชิงสุ่ม แล้วเราสามารถพิจารณาความน่าจะเป็นของแต่ละกฎตัดสินใจใน D ได้ เราได้กล่าวมาแล้วว่ากฎตัดสินใจจะเกี่ยวพันทางเลือกเข้ากับแต่ละผลทดลองใน S ลองพิจารณากฎตัดสินใจ d_7 ของตาราง 3 ซึ่งเป็นกฎตัดสินใจที่เกี่ยวพันทางเลือก a_2 เข้ากับผลทดลองของการทดลอง s ที่เป็น 0 หรือ 1 และทางเลือก a_1 ถ้า s เป็น 2 นั่นคือผู้จัดการควรจะซื้อชิ้นส่วนเครื่องจักร 1000 หน่วยนั้น ถ้าในตัวอย่างสุ่มเขาพบชิ้นส่วนที่ใช้การไม่ได้ 0 หรือ 1 และจะไม่ซื้อ ถ้าพบถึง 2 หน่วย ตามสัญญลักษณ์ของฟังก์ชันตัดสินใจ (Decision Function) แล้วเราจะเขียนได้ว่า

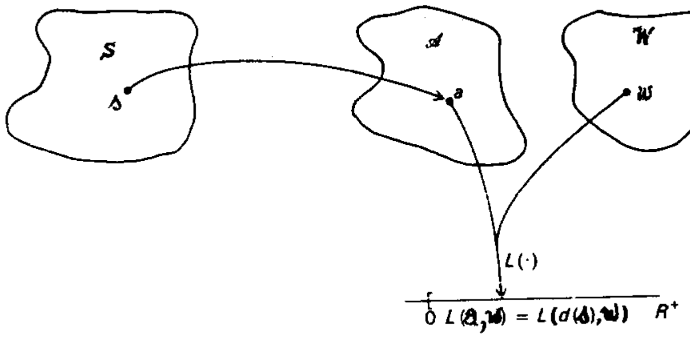
$$
\begin{aligned}
d_7(s) &= a_2 \quad \text{ถ้า } s = 0 \\
&= a_2 \quad \text{ถ้า } s = 1 \\
&= a_1 \quad \text{ถ้า } s = 2
\end{aligned}
$$

ในทำนองเดียวกัน ลองพิจารณากฎตัดสินใจ d_2 กฎตัดสินใจนี้จะแนะนำว่า ให้เลือกทางเลือก a_1 ถ้าพบหน่วยที่ใช้การไม่ได้ในตัวอย่างสุ่มเป็น 0 หรือ 1 หน่วย และเลือกทางเลือก a_2 ถ้าพบ 2 หน่วย ดังนั้นผู้จัดการจะไม่ซื้อถ้าพบหน่วยที่ใช้ไม่ได้เป็น 0 หรือ 1 และจะซื้อ ถ้าพบ 2 หน่วย กฎตัดสินใจนี้จะตรงกันข้ามกับสัญชาตญาณ หรือความรู้สึกของผู้ตัดสินใจ และเราก็หวังว่า ผู้จัดการจะทำตรงกันข้ามกับกฎตัดสินใจนั้น เนื่องจากเราพิจารณากฎตัดสินใจทั้งหมดที่เป็นไปได้ เราจึงรวมกฎตัดสินใจ d_2 ใน D ไว้ โดยเหตุผลเดียวกันเราก็รวม d_1 และ d_8 ไว้ด้วย ซึ่งกฎตัดสินใจทั้งสองจะไม่สนใจผลทดลองของการทดลอง กฎตัดสินใจทั้งสองนี้อาจจะเหมาะสมกับทางเลือกที่เป็นไปได้ของคนที่ขาดสติ ซึ่งลืมดูธรรมชาติแท้จริงของปัญหา

5. <u>ฟังก์ชันการเสี่ยง (Risk Function) ที่เกี่ยวพันกับกฎตัดสินใจ</u>

หลังจากพิจารณาเซทของกฎตัดสินใจทั้งหมดแล้ว งานต่อไปของผู้ตัดสินใจก็คือจะต้องเลือกกฎตัดสินใจที่เหมาะสมที่สุด การวิเคราะห์ของการเลือกกฎตัดสินใจที่ดีที่สุดจะเกี่ยวข้องกับแนวความคิด 2 อย่างคือ — ฟังก์ชันค่าเสียโอกาส และฟังก์ชันการเสี่ยง โดยความจริงแล้วฟังก์ชันทั้งสองนี้จะขึ้นอยู่กับสภาวการณ์นอกบังคับ w ใน W สำหรับ w ใน W ที่กำหนดให้ เราทราบว่า $L(a,w)$ เป็นค่าเสียโอกาส ถ้าผู้ตัดสินใจเลือก a จาก A จากทัศน

ของการวิเคราะห์การตัดสินใจด้วยการสุ่มตัวอย่างนั้นผู้ตัดสินใจจะพิจารณากฎตัดสินใจตามการสังเกตผลทดลองที่ได้จากการทดลอง ดังนั้นเขาพิจารณากฎตัดสินใจ $d(s) = a$ หลังจากสังเกต s ได้ นี่ก็หมายความว่า $L[d(s), w]$ เป็นค่าเสียโอกาสของการพิจารณากฎตัดสินใจ d เพราะว่าเมื่อเขาสังเกตผลทดลอง s ได้ เขาก็จะเลือกทางเลือก a โดยที่ $d(s) = a$ และจะแสดงว่า $L(a,w) = L[d(s),w]$ สำหรับ s บางค่าใน S โดยที่ $d(s) = a$ การวิเคราะห์เช่นเดียวกันนี้แสดงได้ดังรูป



ด้วยเหตุผลเดียวกัน เมื่อกำหนด w เป็นสภาวการณ์ที่แท้จริงให้ แล้วความน่าจะเป็นที่จะทำการตัดสินใจ $d(s)$ จะเป็นเช่นเดียวกับความน่าจะเป็นที่เลือกทางเลือก a ในเมื่อ s เป็นผลทดลองของการทดลองโดยที่ $d(s) = a$ นั่นคือ $p_w[d(s)] = p_w(a)$ สำหรับ s ใน S โดยที่ $d(s) = a$

สำหรับฟังก์ชันค่าเสียโอกาส L นั้นตามเหตุผลที่กล่าวมาตอนแรกจะสามารถแสดงได้ว่าเป็นฟังก์ชันที่กำหนดใน $D \times W$ เนื่องจาก d เป็นตัวแปรเชิงสุ่มที่มีการแจกแจงน่าจะเป็น $p_w(d)$ [หรือเป็น $p_w(a)$ โดยที่ $d(s) = a$ สำหรับ s ใน S] เราจึงสามารถคำนวณค่าเสียโอกาสคาดหวังสำหรับการเลือกกฎตัดสินใจ d และเราจะไม่ลืมที่จะสังเกตว่าค่าเสียโอกาสคาดหวังนั้นกำหนดได้ เมื่อจำกัด w ใน W ค่าเสียโอกาสคาดหวังนี้จะเรียกว่า การเสี่ยงในการตัดสินใจ d

<u>ฟังก์ชันการเสี่ยง</u> : ให้ S, A, W, และ D เป็นกลุ่มผลทดลอง, กลุ่มทางเลือก, กลุ่มสภาวการณ์, และกลุ่มตัดสินใจ สำหรับปัญหาตัดสินใจที่กำหนดให้ ให้ L เป็นฟังก์ชันค่าเสียโอกาสที่กำหนดใน $D \times W$ และ p_w เป็นฟังก์ชันน่าจะเป็นของ w แล้วฟังก์ชันการเสี่ยงของกฎตัดสินใจ d สำหรับ w ใน W ที่กำหนดให้ จะแทนด้วย $R(d,w)$ และกำหนดไว้ว่าเป็นฟังก์ชันค่าจริงใน $D \times W$ โดยที่

$$R(d,w) = \sum_{s \in S} L[d(s),w]\, p_w[d(s)] \quad , \; d \text{ ไม่ต่อเนื่อง}$$
$$= \int_S L[d(s),w]\, p_w[d(s)]\, ds \quad , \; d \text{ ต่อเนื่อง ถ้าหาค่าได้}$$

สำหรับตัวอย่างของเรา d เป็นตัวแปรเชิงสุ่มไม่ต่อเนื่อง เพราะฉะนั้น สำหรับ $w \in W$ ที่กำหนดให้ เราก็ใช้

$$R(d,w) = \sum_{s \in S} L[d(s),w]\, p_w[d(s)]$$
$$= \sum_{a \in A} L(a,w)\, p_w(a)$$

ต่อไปเราจะพูดถึงขบวนการคำนวณ $R(d,w)$ เราทราบว่า $p_w[d(s)] = p_w(a)$ สำหรับ s ใน S โดยที่ $d(s) = a$

เพราะฉะนั้น $p_w(a) = P_w[\{s \mid d(s) = a\}]$

ความน่าจะเป็นเหล่านี้เรียกว่า "ความน่าจะเป็นของทางเลือก (Action Probabilities) ลองพิจารณากฎตัดสินใจ d_2 (จากตัวอย่างของเรา) กฎนี้กล่าวว่าผู้ตัดสินใจควรจะเลือกทางเลือก a_1 เมื่อ s เป็น 0 หรือ 1 และเลือกทางเลือก a_2 เมื่อ s เป็น 2 เพราะฉะนั้นถ้า w_1 เป็นสภาวการณ์ที่แท้จริง ความน่าจะเป็นของทางเลือก a_1 ภายใต้ d_2 จะกำหนดไว้เป็นดังนี้

$$\begin{aligned} p_{w_1}(a_1) &= P_{w_1}(\{s \mid d_2(s) = a_1\}) \\ &= P_{w_1}(\{0, 1\}) \\ &= P_{w_1}(\{0\}) + P_{w_1}(\{1\}) \\ &= .64 + .32 = .96 \end{aligned}$$

แต่ถ้า w_2 เป็นสภาวการณ์ที่แท้จริง เราจะได้

$$\begin{aligned} p_{w_2}(a_1) &= P_{w_2}(\{s \mid d_2(s) = a_1\}) \\ &= P_{w_2}(\{0, 1\}) = .25 + .50 = .75 \end{aligned}$$

ในทำนองเดียวกัน $p_{w_1}(a_2) = P_{w_1}(\{s \mid d_2(s) = a_2\})$
$= P_{w_1}(\{2\}) = .04$
$p_{w_2}(a_2) = P_{w_2}(\{2\}) = .25$

ถ้าเราทำตามวิธีการข้างบนนี้ เราก็จะคำนวณความน่าจะเป็นของทางเลือกในกฎตัดสินใจทั้งหมดจากปัญหาของเราได้ ซึ่งผลคำนวณจะได้ดังตารางต่อไปนี้

ตาราง 5 ความน่าจะเป็นของทางเลือกในปัญหาตัดสินใจ

	S	0	1	2				
W	w_1	.64	.32	.04	$p_{w_1}(a_1)$	$p_{w_1}(a_2)$	$p_{w_2}(a_1)$	$p_{w_2}(a_2)$
	w_2	.25	.50	.25				
d_1		a_1	a_1	a_1	1.00	.00	1.00	.00
d_2		a_1	a_1	a_2	.96	.04	.75	.25
d_3		a_1	a_2	a_1	.68	.32	.50	.50
d_4		a_1	a_2	a_2	.64	.36	.25	.75
d_5		a_2	a_1	a_1	.36	.64	.75	.25
d_6		a_2	a_1	a_2	.32	.68	.50	.50
d_7		a_2	a_2	a_1	.04	.96	.25	.75
d_8		a_2	a_2	a_2	.00	1.00	.00	1.00

ความน่าจะเป็นของทางเลือกจากตาราง 5 นี้ กับค่าเสียโอกาสจากตาราง 1 นั้น เราใช้เพื่อคำนวณการเสี่ยง $R(d_i, W)$ ของทุก ๆ กฎตัดสินใจ d_i โดยกำหนดว่า w เป็นสภาวการณ์ที่แท้จริง ให้ w_1 เป็นสภาวการณ์ที่แท้จริง แล้วการเสี่ยงของ d_i กำหนดให้เป็น

$$R(d_i, w_1) = \sum_{a_i \in A} L(a_i, w_1)\, p_{w_1}(a_i)$$
$$= L(a_1, w_1)\, p_{w_1}(a_1) + L(a_2, w_1)\, p_{w_1}(a_2)$$
$$= 1600\, p_{w_1}(a_1)$$

ในทำนองเดียวกัน ถ้า w_2 เป็นสภาวการณ์ที่แท้จริง เราจะได้

$$R(d_i, w_2) = \sum_{a_i \in A} L(a_i, w_2)\, p_{w_2}(a_2)$$
$$= L(a_1, w_2)\, p_{w_2}(a_1) + L(a_2, w_2)\, p_{w_2}(a_2)$$
$$= 500\, p_{w_2}(a_2)$$

โดยเฉพาะ ถ้าเราพิจารณา d_2 แล้วการเสี่ยงภายใต้ w_1 และ w_2 กำหนดไว้เป็น

$$R(d_2,w_1) = 1600(.96) = 1536.00$$
$$R(d_2,w_2) = 500(.25) = 125.00$$

ด้วยการใช้กระบวนการเดียวกัน เราก็คำนวณการเสี่ยงของกฎตัดสินใจอื่น ๆ ได้ ซึ่งผลที่คำนวณการเสี่ยงแสดงได้ในตารางต่อไปนี้

ตาราง 6 การเสี่ยงของกฎตัดสินใจทั้งหมดภายใต้ w_1 และ w_2

กฎตัดสินใจ d_i	$R(d_i,w_1)$	$R(d_i,w_2)$	$B(d_i)$
d_1	1600.00	0.00	960.00
d_2	1536.00	125.00	971.60
d_3	1088.00	250.00	752.80
d_4	1024.00	375.00	764.40
d_5	576.00	125.00	395.60
d_6	512.00	250.00	407.20
d_7	64.00	375.00	188.40
d_8	0.00	500.00	200.00

6. มาตราวัดแบบเบยส์ของกฎตัดสินใจ และกฎตัดสินใจแบบเบยส์

เราจะใช้การแจกแจงน่าจะเป็นก่อนทดลองของสภาวการณ์เพื่อคำนวณมาตราวัดแบบเบยส์ของกฎตัดสินใจ ลองพิจารณา d_2 ถ้า w_1 เป็นสภาวการณ์ที่แท้จริง แล้วการเสี่ยงที่ใช้ d_2 เป็น 1536.00 ในทำนองเดียวกัน ถ้า w_2 เป็นสภาวการณ์ที่แท้จริง แล้วการเสี่ยงที่ใช้ d_2 เป็น 125.00 อย่างไรก็ตาม w_1 และ w_2 อาจจะเกิดขึ้นด้วยความน่าจะเป็นก่อนทดลองเป็น .60 แล .40 ตามลำดับ เพราะฉะนั้นการเสี่ยงคาดหวังที่ใช้ d_2 เป็น

$$1536.00(.60) + 125.00(.40) = 971.60$$

การเสี่ยงคาดหวังนี้เรียกว่า "มาตราวัดแบบเบยส์ของ d_2" โดยทั่วไปมาตราวัดแบบเบยส์ $B(d_i)$ ของกฎตัดสินใจ d_i สำหรับปัญหาตัดสินใจแบบ 2 สภาวการณ์นั้นกำหนดไว้ว่า

$$B(d_i) = R(d_i,w_1)p(w_1) + R(d_i,w_2)p(w_2)$$

ในเมื่อ $p(w_1)$ และ $p(w_2)$ เป็นความน่าจะเป็นก่อนทดลองของสภาวการณ์ w_1 และ w_2 แนวความคิดนี้สามารถขยายออกไปได้กับกลุ่มสภาวการณ์ W ใด ๆ ซึ่งจะนิยามได้ดังนี้

<u>มาตราวัดแบบเบยส์ของกฎตัดสินใจ</u> : ให้ W และ D เป็นกลุ่มสภาวการณ์ และกลุ่มการตัดสินใจของปัญหาตัดสินใจ ให้ p เป็นการแจกแจงน่าจะเป็นก่อนทดลองใน W แล้ว มาตราวัดแบบเบยส์ $B(d_i)$ ของกฎตัดสินใจ d_i ใน D กำหนดไว้ว่า

$$B(d_i) = \sum R(d_i, w)p(w) \quad \text{ถ้า } p \text{ ไม่ต่อเนื่อง}$$
$$= \int R(d_i, w)p(w)dw \quad \text{ถ้า } p \text{ ต่อเนื่อง (เมื่อหาค่าได้)}$$

ในปัญหาตัวอย่างของเรา มาตราวัดแบบเบยส์ที่คำนวณได้ของทุกกฎตัดสินใจแสดงไว้ในแถวตั้งสุดท้ายของตาราง 6 เราจะเลือกกฎตัดสินใจใน D ที่ทำให้มาตราวัดแบบเบยส์ต่ำสุด กฎตัดสินใจเช่นนี้จะเรียกว่า "กฎตัดสินใจแบบเบยส์ (Bayesian Decision Rule) จากตาราง 6 เราจะเห็นว่า d_7 เป็นกฎตัดสินใจแบบเบยส์สำหรับปัญหาของเรา เพราะฉะนั้นผู้จัดการควรจะเลือกทางเลือก a_2 ไม่ว่าผลทดลองจะเป็น 0 หรือ 1 และควรจะเลือกทางเลือก a_1 เมื่อผลทดลองเป็น 2 นั่นคือเขาควรจะซื้อชิ้นส่วน 1000 ชิ้นนั้นเมื่อตัวอย่างขนาด 2 นี้มีหน่วยที่ใช้การไม่ได้เป็น 0 หรือ 1 และไม่ควรซื้อ ถ้ามีหน่วยที่ใช้ไม่ได้เป็นอย่างอื่น

<u>กฎตัดสินใจแบบเบยส์</u> : ให้ W และ D เป็นกลุ่มสภาวการณ์ และกลุ่มการตัดสินใจของปัญหาตัดสินใจ ให้ $B(d_i)$ เป็นมาตราวัดแบบเบยส์ของ d_i ใน D แล้วกฎตัดสินใจ d_k จะเรียกว่า กฎตัดสินใจแบบเบยส์ ถ้า d_k สอดคล้องกับ

$$B(d_k) = \min_{d_i \in D} B(d_i)$$

โดยสรุป ในรูปแบบปกติของการวิเคราะห์ตัดสินใจด้วยการสุ่มตัวอย่างนั้น เราพิจารณากลุ่มทางเลือก กลุ่มสภาวการณ์ ฟังก์ชันค่าเสียโอกาส และการแจกแจงน่าจะเป็นก่อนทดลองสำหรับปัญหาตัดสินใจ แล้วสำหรับขนาดตัวอย่างที่จำกัด เราก็เลือกตัวอย่างสุ่ม และนิยามกลุ่มผลทดลอง และกลุ่มการตัดสินใจ เราคำนวณการเสี่ยง และมาตราวัดแบบเบยส์ของกลุ่มการตัดสินใจ และเลือกกฎตัดสินใจที่เหมาะสมที่สุดซึ่งจะทำให้มาตราวัดแบบเบยส์มีค่าต่ำสุด เราจะเห็นได้ว่ากฎตัดสินใจที่เหมาะสมซึ่งพิจารณาจากการแจกแจงน่าจะเป็นก่อนทดลองที่สมมติไว้ แต่ถ้าเราพิจารณาการแจกแจงน่าจะเป็นก่อนทดลองอื่น ๆ แล้วกฎตัดสินใจที่เหมาะสมมักจะเปลี่ยนแปลง การวิเคราะห์ในรูปแบบปกติแสดงได้ดังขั้นตอนต่อไปนี้

- กำหนด A, W, L, p

- เลือกตัวอย่างขนาด n และพิจารณากลุ่มผลทดลอง S และกลุ่มการตัดสินใจ D
- สำหรับ w ใน W ที่กำหนด พิจารณาการแจกแจงน่าจะเป็น p_w สำหรับผลทดลองของกลุ่มผลทดลอง S
- สำหรับแต่ละ w ใน W คำนวณความน่าจะเป็นของทางเลือก $p_w(a_i)$
- สำหรับแต่ละ w ใน W และ d_i ใน D คำนวณ $R(d_i,w)$

$$R(d_i,w) = \sum L(a_i,w)p_w(a_i)$$

- สำหรับ d_i ใน D คำนวณมาตราวัดแบบเบยส์ $B(d_i)$

$$B(d_i) = \sum R(d_i,w)p(w)$$

- เลือกกฎตัดสินใจ d_k ที่ $B(d_k) = \min_{d_i \in D} B(d_i)$

ในทำนองเดียวกัน ถ้าเราใช้ฟังก์ชั่นกำไร แทนที่ใช้ฟังก์ชั่นค่าเสียโอกาส เราก็ทำการวิเคราะห์แบบเดิมเพื่อคำนวณมาตราวัดแบบเบยส์ของทุก ๆ กฎตัดสินใจ แต่จะเลือกกฎตัดสินใจแบบเบยส์ที่ทำให้มาตราวัดแบบเบยส์มีค่าสูงสุด

7. <u>การวิเคราะห์แบบขยาย</u> (Extensive Form Analysis)

ดังที่กล่าวมาแล้วในตอนต้น ๆ ว่า การวิเคราะห์แบบขยายของปัญหาการตัดสินใจจะประกอบด้วยการกระทำการทดลองเชิงสุ่ม เพื่อให้ได้ข้อมูลข่าวสารเกี่ยวกับการเกิดขึ้นของสภาวการณ์นอกบังคับ และเลือกทางเลือกที่เหมาะสมสำหรับทุก ๆ ผลทดลองที่เป็นไปได้ของการทดลอง ทางเลือกที่เหมาะสมสำหรับทุก ๆ ผลทดลองที่เป็นไปได้ของการทดลองจะเป็นกฎตัดสินใจแบบเบยส์สำหรับปัญหาตัดสินใจนั้น และจะเป็นเช่นเดียวกับกฎตัดสินใจที่หาได้จากการวิเคราะห์แบบปกติ ต่อไปเราจะอธิบายถึงวิธีเลือกกฎตัดสินใจที่เหมาะสมสำหรับปัญหาตัดสินใจที่ใช้การวิเคราะห์แบบขยาย

ให้ S เป็นกลุ่มผลทดลอง, p เป็นการแจกแจงน่าจะเป็นก่อนทดลองใน W, และ p_w เป็นฟังก์ชั่นน่าจะเป็น การวิเคราะห์แบบขยายนี้การแจกแจงน่าจะเป็นก่อนทดลองจะได้รับการปรับปรุงจากการเกิดขึ้นของผลทดลองใน S สมมติว่า S และ W เป็นแบบไม่ต่อเนื่อง และให้ s_1 เป็นผลทดลองที่สังเกตได้จากการทดลอง สำหรับความน่าจะเป็นที่ปรับปรุงของ w_1

โดยกำหนด s_1 ให้จะแทนด้วย $p(w_1/s_1)$ ซึ่งคำนวณได้จากทฤษฎีเบยส์ดังนี้

$$p(w_1/s_1) = p(s_1/w_1)p(w_1)/\sum_{w_k \in W} p(s_1/w_k)p(w_k)$$

โดยทั่วไปความน่าจะเป็นที่ปรับปรุงของ w_j กำหนด s_i ซึ่งเป็นผลทดลองที่สังเกตได้ จะเป็นดังนี้

$$p(w_j/s_i) = p(s_i/w_j)p(w_j)/\sum_{w_k \in W} p(s_i/w_k)p(w_k)$$

เราใช้ความน่าจะเป็นที่ปรับปรุงนี้คำนวณมาตรารวัดแบบเบยส์สำหรับทุก ๆ ทางเลือก a_i ใน A และเลือกทางเลือกที่เหมาะสมซึ่งมีมาตรารวัดแบบเบยส์ต่ำสุด ขบวนการนี้จะทำซ้ำ ๆ กันสำหรับทุก ๆ ผลทดลองใน S เมื่อขบวนการจบลงเราจะได้ข้อแนะนำที่บอกว่าทางเลือกไหนที่เหมาะสมที่เราควรจะทำสำหรับทุกผลทดลองที่เป็นไปได้จากการทดลอง ข้อแนะนำนี้จะเป็นกฎตัดสินใจแบบเบยส์สำหรับปัญหาตัดสินใจ

จากตัวอย่างของเราจะได้

$$S = \{s_1, s_2, s_3\}$$

ในเมื่อ s_1 เป็นผลทดลองที่แทนหน่วยที่ใช้การไม่ได้ 0 หน่วย s_2 เป็นผลทดลองที่แทนหน่วยที่ใช้การไม่ได้ 1 หน่วย และ s_3 แทน 2 หน่วย

สำหรับ w ใน W นั้น การแจกแจงน่าจะเป็น p_w กำหนดไว้ในตาราง 2 สมมติว่า s_1 เป็นผลทดลองที่สังเกตได้จากการทดลอง ในการปรับปรุงความน่าจะเป็นก่อนทดลอง $p(w_1)$ เราใช้ทฤษฎีเบยส์

$$\begin{aligned} p(w_1/s_1) &= p(s_1/w_1)p(w_1)/\sum p(s_1/w_k)p(w_k) \\ &= .64(.60)/\{.64(.60) + .25(.40)\} \\ &= .7934 \end{aligned}$$

ในทำนองเดียวกัน ความน่าจะเป็นที่ปรับปรุง $p(w_2/s_1)$ สามารถคำนวณได้เป็น

$$\begin{aligned} p(w_2/s_1) &= p(s_1/w_2)p(w_2)/\sum p(s_1/w_k)p(w_k) \\ &= .25(.40)/\{.64(.60) + .25(.40)\} \\ &= .2066 \end{aligned}$$

เรายังทราบอีกว่า $L(a_1, w_1) = 1600$ และ $L(a_1, w_2) = 0$ เพราะฉะนั้นมาตรารวัดแบบเบยส์ของ a_1 เมื่อผลทดลองที่สังเกตได้เป็น s_1 คำนวณได้เป็น

$$\begin{aligned} B(a_1/s_1) &= L(a_1, w_1)p(w_1/s_1) + L(a_1, w_2)p(w_2/s_1) \\ &= 1600(.7934) + 0(.2066) = 1269.44 \end{aligned}$$

ในทำนองเดียวกัน มาตราวัดแบบเบยส์ของ a_2 กำหนด s_1 ให้ จะเป็น

$$B(a_2/s_1) = L(a_2,w_1)p(w_1/s_1) + L(a_2,w_2)p(w_2/s_1)$$
$$= 0(.7934) + 500(.2066) = 103.30$$

จากการเปรียบเทียบมาตราวัดแบบเบยส์ของทั้งสองทางเลือก เราจะเห็นว่าทางเลือก a_2 นั้นเหมาะสมที่สุดสำหรับผลทดลองที่สังเกตได้ s_1 นั่นคือ ถ้าผู้จัดการไม่พบหน่วยที่ใช้การไม่ได้เลยจากตัวอย่างสุ่ม เขาก็ควรจะซื้อชิ้นส่วนเครื่องจักรเหล่านั้น

สมมติว่าผู้จัดการสังเกต s_2 ได้ เราสามารถคำนวณความน่าจะเป็นที่ปรับปรุงได้เป็น

$$p(w_1/s_2) = .32(.60)/\{.32(.60) + .50(.40)\}$$
$$= .4898$$
$$p(w_2/s_2) = .50(.40)/\{.32(.60) + .50(.40)\}$$
$$= .5102$$

และมาตราวัดแบบเบยส์ของ a_1 และ a_2 จะคำนวณได้เป็น

$$B(a_1/s_2) = 783.68$$
$$B(a_2/s_2) = 255.10$$

เพราะฉะนั้นผู้ตัดสินใจควรจะซื้อชิ้นส่วนเหล่านั้น ถ้าเขาพบหน่วยที่ใช้การไม่ได้เพียง 1 หน่วย กระบวนการของการเลือกทางเลือกที่เหมาะสมนี้อธิบายได้ในตารางต่อไปนี้

ตาราง 7 การวิเคราะห์แบบขยาย

		S					S			
		$s_1=0$	$s_2=1$	$s_3=2$	$p(w)$	$p(s_i/w_j)p(w_j)$	$s_1=0$	$s_2=1$	$s_3=2$	$p(w)$
W	w_1	.64	.32	.04	.60	$p(s_1/w_1)p(w_1)$	.3840	.1920	.0240	.60
	w_2	.25	.50	.25	.40	$p(s_1/w_2)p(w_2)$	.1000	.2000	.1000	.40
						$p(s_i)$	.4840	.3920	.1240	1.00
						$p(w_1/s_i) = \frac{p(s_i/w_1)p(w_1)}{\sum p(s_i/w_k)p(w_k)}$	.7934	.4898	.1935	
						$p(w_2/s_i) = \frac{p(s_i/w_2)p(w_2)}{\sum p(s_i/w_k)p(w_k)}$	.2066	.5102	.8065	

	A	a_1	a_2
W	w_1	1600	0
	w_2	0	500

มาตราวัดแบบเบยส์ ของ a_i	a_1	1269.44	783.68	309.60
	a_2	103.30	255.10	403.25
กฎตัดสินใจแบบเบยส์		a_2	a_2	a_1

การเสี่ยงคาดหวังของกฎตัดสินใจแบบเบยส์

$$103.30(.4840) + 255.10(.3920) + 309.60(.1240) = 188.3860$$

การเสี่ยงคาดหวังของกฎตัดสินใจแบบเบยส์แสดงได้เป็นสูตรดังนี้

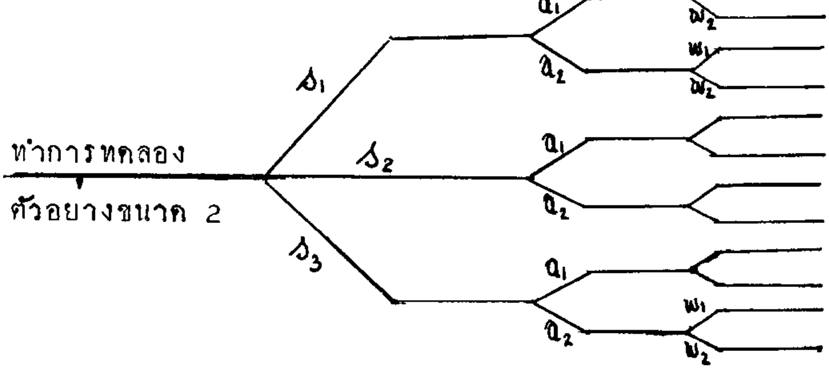
$$E[B(a^\circ/s_i)] = \sum_i B(a^\circ/s_i)p(s_i)$$

ในเมื่อ $B(a^\circ/s_i)$ เป็นมาตราวัดแบบเบยส์ของทางเลือกที่ดีที่สุด (a°) จากผลทดลอง s_i

จากตารางนี้เราจะเห็นว่า ผู้จัดการควรจะเลือกทางเลือก a_2 ถ้า s_1 หรือ s_2 เป็นผลทดลองของการทดลอง และจะเลือกทางเลือก a_1 ถ้า s_3 เป็นผลทดลองของการทดลอง กโลบายนี้อาจจะแสดงได้ด้วยกฎตัดสินใจ d:

$$\begin{aligned} d(s) &= a_2 \quad \text{ถ้า } s = s_1 \\ &= a_2 \quad \text{ถ้า } s = s_2 \\ &= a_1 \quad \text{ถ้า } s = s_3 \end{aligned}$$

จากตาราง เราจะเห็นว่า กฎตัดสินใจนี้ก็คือ d_7 นั่นเอง ดังนั้นตามการวิเคราะห์แบบขยาย กฎตัดสินใจแบบเบยส์ก็คือ d_7 ซึ่งเป็นกฎตัดสินใจเดียวกับที่ได้จากการวิเคราะห์แบบปกติ การวิเคราะห์แบบขยายสามารถอธิบายได้ดังแผนภาพต่อไปนี้



จากที่กล่าวมาทั้งหมด เราจะเห็นว่าโครงสร้าง $(A,S,W,L,\tilde{P})$ จะกำหนดการวิเคราะห์แบบขยาย ในเมื่อ $\tilde{P}$ เป็นการแจกแจงที่ปรับปรุงของสภาวการณ์นอกบังคับ

<u>การวิเคราะห์แบบขยาย</u> : ให้ A, S, และ W เป็นกลุ่มทางเลือก, กลุ่มผลทดลอง, และกลุ่มสภาวการณ์ของปัญหาตัดสินใจ ให้ L เป็นฟังก์ชันค่าเสียโอกาสที่กำหนดใน $A \times W$ ให้ $\tilde{P}$ เป็นการแจกแจงน่าจะเป็นที่ปรับปรุงของสภาวการณ์นอกบังคับ แล้วโครงสร้าง $(A,S,W,L,\tilde{P})$ จะกำหนดการวิเคราะห์แบบขยายของปัญหาตัดสินใจ

จากนิยามข้างบนนี้เราใช้ฟังก์ชันค่าเสียโอกาส แต่เราสามารถใช้ฟังก์ชันผลตอบแทน U ได้ ซึ่งมีการวิเคราะห์เช่นเดียวกัน

ถ้าผู้จัดการตัดสินใจที่จะสุ่มตัวอย่างขนาด 3 หน่วย หรือขนาดใด ๆ ก็ได้ แล้วเราสามารถทำการวิเคราะห์ในรูปแบบปกติ หรือแบบขยายดังที่อธิบายข้างบน เพื่อที่จะพัฒนากฎตัดสินใจแบบเบยส์ได้ การประหยัดในการสุ่มตัวอย่าง และการพิจารณาขนาดตัวอย่างที่เหมาะสมของปัญหาตัดสินใจที่กำหนดให้นั้นจะได้พิจารณาต่อไปนี้

8. <u>วิธีพิจารณาขนาดตัวอย่างที่เหมาะสม</u> (Optimal Sample Size)

เท่าที่ผ่านมาเราไม่ได้พูดถึงขนาดตัวอย่างในการวิเคราะห์ปัญหาการตัดสินใจด้วยการสุ่มตัวอย่าง และเพียงพิจารณาตัวอย่างสุ่มของชิ้นส่วนเพียง 2 เท่านั้น เราอาจจะเลือกชิ้นส่วนมาเป็นร้อย หรือเป็นพันก็ได้ และก็ทำการวิเคราะห์เพื่อจะพิจารณาทางเลือกที่เหมาะสมเช่นที่กล่าวมา อย่างไรก็ตามจะมีค่าใช้จ่ายของการสุ่มตัวอย่าง (Sampling Costs) คือค่าเตรียมการ และค่าใช้จ่ายผันแปร (Set-up and Variable cost) เกี่ยวพันกับแต่ละตัวอย่าง และค่าใช้จ่ายเหล่านี้ต้องเทียบ (weighted) กับค่าของข้อมูลข่าวสารที่จะได้รับด้วยการสุ่มตัวอย่าง ถ้าไม่มีค่าใช้จ่ายแบบนั้น แล้วการตัดสินใจที่เหมาะ ๆ จะเป็นตัวอย่างที่เราไม่ได้ข้อมูลข่าวสารเพิ่มเติมอะไรอีกเลย ยิ่งกว่านั้นมาตราวัดแบบเบยส์ของกฎตัดสินใจที่เหมาะสามารถทำให้น้อยลง เมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะว่าข้อมูลข่าวสารจะได้รับมาก ๆ และทำให้การตัดสินใจดีขึ้น อย่างไรก็ตามเราจะสังเกตว่าการลดลงของมาตราวัดแบบเบยส์ของทางเลือกที่เหมาะจะประสบความสำเร็จเมื่อเพิ่มค่าใช้จ่ายอยู่เรื่อย ๆ เพราะฉะนั้นเราต้องพิจารณาจุดที่อิทธิพลของทั้งสองขัดแย้งกัน เพื่อหาขนาดตัวอย่างที่เหมาะสม ให้ $c(n)$ เป็นฟังก์ชันค่าใช้

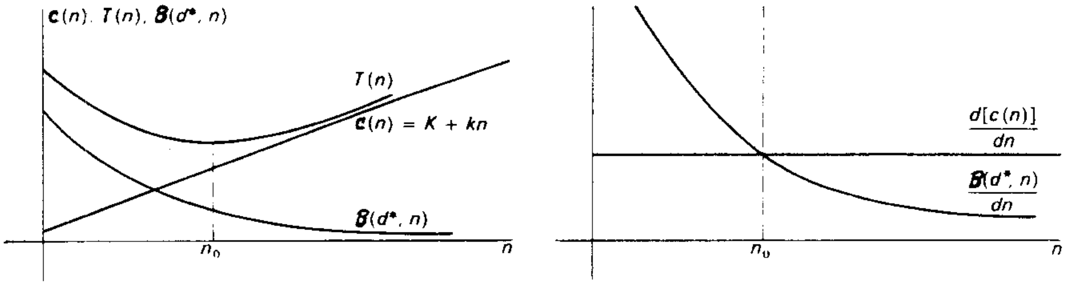
จ่ายของการสุ่มตัวอย่างที่เกี่ยวพันกับปัญหาตัดสินใจ ซึ่งกำหนดไว้ว่า

$$C(n) = K + kn, \quad n \geq 1$$

และให้ $B(d^*;n)$ เป็นมาตราวัดแบบเบยส์ของกฎตัดสินใจที่เหมาะ d^* เมื่อขนาดตัวอย่างเป็น n และให้ $B(a^*)$ เป็นมาตราวัดแบบเบยส์ของ a^* สำหรับปัญหาการวิเคราะห์การตัดสินใจแบบขั้นเดียว แล้วการเสี่ยงทั้งหมด $T(n)$ เมื่อตัวอย่างขนาด n จะกำหนดไว้ว่า

$$T(n) = B(d^*;n) + (K + kn)$$

เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าใช้จ่ายของการสุ่มตัวอย่างจะเพิ่มขึ้นด้วย และ $B(d^*;n)$ จะลดลง โดยที่ $T(n)$ จะเริ่มต้นลดลง และจะเริ่มเพิ่มขึ้นต่อเมื่อหลังจากจุด n_0 ในเมื่อค่าใช้จ่ายที่เพิ่มขึ้น (Marginal cost) ของหน่วยต่อ ๆ ไปที่เลือกสุ่มมาเพียงจะทำให้การลดลงที่ได้ในมาตราวัดแบบเบยส์สมดุลย์ จุด n_0 นี้จะเป็นขนาดตัวอย่างที่เหมาะสม ลองพิจารณารูปต่อไปนี้



ลองพิจารณาตัวอย่างเรื่องชิ้นส่วน ซึ่งเลือกตัวอย่างสุ่มของ 2 ชิ้นส่วน และพบว่า d_7 เป็นกฎตัดสินใจแบบเบยส์ เมื่อ

$$\begin{aligned} d_7(s) &= a_2 \quad \text{ถ้า} \quad s = s_1 = 0 \\ &= a_2 \quad \text{ถ้า} \quad s = s_2 = 1 \\ &= a_1 \quad \text{ถ้า} \quad s = s_3 = 2 \end{aligned}$$

มาตราวัดแบบเบยส์ของ d_7 คำนวณได้เป็น 188.40 ดังนั้นเราจะได้

$$B(d^*;n) = B(d_7;\ 2) = 188.40$$

เพื่อความสะดวกเราจะใช้ d^* แทนกฎตัดสินใจแบบเบยส์ d_7 ดังนั้นเราจะได้ $B(d^*;\ 2) = 188.40$ สมมติว่า $K = 2$ และ $k = 4$ แล้วค่าใช้จ่ายของการสุ่มตัวอย่าง ในเมื่อมีตัวอย่างขนาด 2 กำหนดไว้ดังนี้

$$C(2) = 2 + 4(2) = 10$$

เพราะฉะนั้นการเสี่ยงทั้งหมด $T(n)$ เมื่อตัวอย่างขนาด 2 จะเป็น

$$T(2) = 188.40 + 10 = 198.40$$

เราทราบมาแล้วว่า PEVPI เป็น 200 นั่นคือมันเป็นค่าเสียโอกาสคาดหวังของทางเลือกที่ดีที่สุด ซึ่งขึ้นอยู่กับการแจกแจงน่าจะเป็นก่อนทดลองที่กำหนดให้ เพราะฉะนั้นเราจะได้ผลต่าง

$$200 - 188.40 = 11.60$$

ซึ่งเนื่องมาจากข้อมูลข่าวสารของตัวอย่าง นั่นคือเราสามารถลดค่าเสียโอกาสคาดหวังเท่ากับ 11.60 ถ้าเราสุ่มตัวอย่างขนาด 2 ชิ้นส่วน และเลือก d_7 เป็นกฎตัดสินใจที่เหมาะสม ค่า 11.60 นี้เรียกว่า "ค่าคาดหวังของข้อมูลข่าวสารจากตัวอย่าง (EVSI, Expected Value Of Sample Information) และเราจะให้นิยามต่อไปนี้

<u>ค่าคาดหวังของข้อมูลข่าวสารจากตัวอย่าง</u> : ให้ $B(a^*)$ เป็นมาตราวัดแบบเบยส์ของทางเลือกที่เหมาะสมของปัญหาการตัดสินใจชนิดขั้นเดียว และให้ $B(d^*;n)$ เป็นการเสี่ยงคาดหวังของกฎตัดสินใจแบบเบยส์ของปัญหาการตัดสินใจด้วยการสุ่มตัวอย่างที่สนใจนั้น แล้วค่าคาดหวังของข้อมูลข่าวสารจากตัวอย่าง เมื่อขนาดตัวอย่างเป็น n กำหนดไว้ว่า

$$\begin{aligned} EVSI(n) &= B(a^*) - B(d^*;n) \\ &= PEVPI - B(d^*;n) \quad \text{เพราะ } B(a^*) = PEVPI \end{aligned}$$

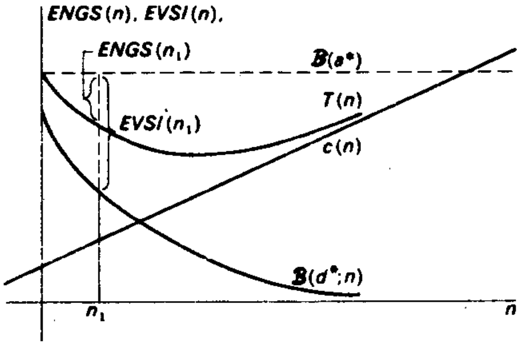
ในทำนองเดียวกันเราจะได้ว่า ผลต่าง $200 - 198.40 = 1.60$ นี้เนื่องมาจากผลตอบแทนสุทธิ (net gain) ในตัวอย่างนั้น เมื่อค่านี้เป็นบวก แล้วมันจะลดลงด้วยจำนวนที่เป็นบวกสำหรับค่าเสียโอกาสคาดหวังของกฎตัดสินใจที่เหมาะสม และเพราะฉะนั้นการสุ่มตัวอย่างจะมีคุณค่า (worthwhile) ในการที่จะลดการเสี่ยง

<u>ผลตอบแทนสุทธิคาดหวังจากตัวอย่าง</u> : สมมติว่ามีข้อสมมติเช่นเดียวกับนิยามของค่าคาดหวังของข้อมูลข่าวสารจากตัวอย่าง (EVSI) แล้วผลตอบแทนสุทธิคาดหวังจากตัวอย่าง (ENGS, Expected net gain from Sample) กำหนดไว้ดังนี้

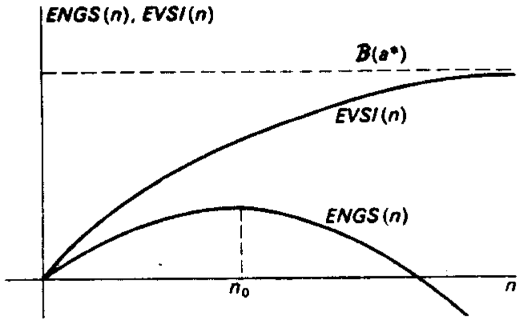
$$ENGS(n) = EVSI(n) - C(n)$$

เมื่อ $C(n)$ เป็นฟังก์ชันค่าใช้จ่ายของการสุ่มตัวอย่าง

ค่าของ EVSI(n) และ ENGS(n) สำหรับปัญหาทั่ว ๆ ไปแสดงได้ดังรูป



จากรูปนี้เราจะเห็นได้ว่าค่าของ EVSI(n) ที่เป็นฟังก์ชันของ n จะเพิ่มขึ้น เมื่อขนาดตัวอย่าง n เพิ่มขึ้น แต่อัตรา (rate) ของการเพิ่มอาจจะลดลงหลังจากค่าของ n บางค่า ในทำนองเดียวกันค่าของ ENGS(n) จะเพิ่มครั้งแรก และถึงจุดสูงสุดของมัน ณ จุด n_0 ซึ่งเป็นขนาดตัวอย่างที่เหมาะสม แล้วจะลดลงเมื่อ $n > n_0$ เราได้ชี้ให้เห็นมาแล้วว่า n_0 เป็นจุดที่ค่าใช้จ่ายที่เพิ่มขึ้นของหน่วยต่อไปที่สุ่มมาจะสมดุลย์กับการลดลงของค่าเสียโอกาสคาดหวัง ดังกราฟ



สำหรับตัวอย่างที่กล่าวมาเราจะได้ค่าของ EVSI(n) และ ENGS(n) สำหรับขนาดตัวอย่างต่าง ๆ กันดังนี้

<u>ตาราง</u> 9 EVSI(n) และ ENGS(n)

ขนาดตัวอย่าง n	$B(d^*,n)$	c(n)	T(n)	EVSI(n)	ENGS(n)
0	200.00	0	200.00		
1	199.99	6	205.99	.006	
2	188.38	10	198.38	11.61	1.61
3	182.62	14	196.62	17.37	3.37

ขนาดตัวอย่าง n	$B(d^*;n)$	$C(n)$	$T(n)$	$EVSI(n)$	$ENGS(n)$
4	163.41	18	181.41	36.59	18.59
5	155.51	22	177.51	44.48	22.48
6	147.54	26	173.54	52.45	26.45
7	131.03	30	181.03	68.96	38.97
8	126.56	34	180.56	73.43	39.42
9	118.89	38	156.89	81.10	43.10
10	106.92	42	148.92	93.07	51.07
11	103.17	46	149.17	96.82	50.82
12	96.27	50	146.27	103.72	53.72
13	87.15	54	141.15	112.84	58.84
14	83.92	58	141.92	116.07	58.07
15	78.09	62	140.09	121.90	59.90
16	71.01	66	137.01	128.98	62.98
17	67.43	70	137.43	132.56	62.56
18	62.66	74	136.66	137.34	63.34
19	59.07	78	137.07	140.93	62.93
20	57.14	82	139.04	142.95	60.95
22	49.61	90	139.61	150.38	60.38
24	43.11	98	141.11	156.88	58.88
26	38.34	106	144.34	161.65	55.65

$$B(d^*;\ 0) = B(a^*)$$

เราจะเห็นได้ว่าผลในตารางนั้นจะแสดงรูปร่างของกราฟทั้งสองที่ผ่านมานั้น จากตารางเราจะเห็นว่าขนาดตัวอย่างที่เหมาะสมจะเป็น 18 ดังนั้นปัญหาการตัดสินใจที่อธิบายมาจากตอนแรก ๆ นั้น ผู้จัดการควรจะสุ่มตัวอย่างเป็นจำนวน 18 ชิ้น และจะพบว่าเขาควรจะซื้อของนั้นถ้าตัวอย่างที่ได้มีชิ้นส่วนที่ใช้การไม่ได้อย่างมาก 7 ชิ้น ในทางตรงกันข้าม ถ้ามีของที่ใช้การไม่ได้มากกว่า 7 ชิ้น แล้วเขาควรจะไม่ซื้อ

ทฤษฎีตัดสินใจทางสถิติที่อาศัยการสุ่มตัวอย่าง เมื่อค่าสังเกตจากตัวอย่าง เป็นแบบไม่ต่อเนื่อง

Statistical Decision Theory with Sampling when Sample Observations are Discrete

ในการวิเคราะห์แบบขยายเราต้องปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลอง ซึ่งกำหนดให้แก่สภาวการณ์นอกบังคับ W และใช้การแจกแจงน่าจะเป็นที่ปรับปรุงแล้วไปเลือกทางเลือกที่ดีที่สุด การปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลองอาจจะซับซ้อนและใช้เวลามาก เมื่อกลุ่มสภาวการณ์มีสภาวการณ์นอกบังคับมาก และเมื่อการแจกแจงน่าจะเป็นของแปรเชิงสุ่มที่กำหนดตามกลุ่มผลทดลองเป็นแบบไม่ต่อเนื่อง ในทำนองเดียวกัน ถ้าการแจกแจงน่าจะเป็นก่อนทดลองเป็นแบบต่อเนื่อง แล้วการวิเคราะห์ที่อธิบายในบทที่แล้วอาจจะไม่เหมาะที่จะปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลอง ยิ่งกว่านั้นข้อกำหนดที่มีเพียง 2 สภาวการณ์นอกบังคับในตัวอย่างที่แล้วมานั้นอาจจะไม่เป็นจริงนัก อย่างไรก็ตามเรากำหนดข้อกำหนดขึ้นก็เพราะว่าการวิเคราะห์จะง่ายกว่า และแสดงการวิเคราะห์แบบขยายได้ชัดแจ้งกว่า ตามปกติเปอร์เซนต์ของเสียอาจจะเป็นค่าใด ๆ ในช่วง (0, 1) ซึ่งในกรณีนี้จะสมมติว่าสัดส่วนของเสียนั้นเป็นแบบต่อเนื่อง ดังนั้นการแจกแจงน่าจะเป็นก่อนทดลองเป็นแบบต่อเนื่อง ในบทนี้เราจะพัฒนาวิธีปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลอง เมื่อมันเป็นแบบต่อเนื่อง และเมื่อค่าสังเกตจากตัวอย่างเป็นแบบไม่ต่อเนื่อง

1. <u>ฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลองแบบบีทะ</u> (Beta Prior Probability Density Function)

ลองพิจารณาตัวอย่างเกี่ยวกับชิ้นส่วนเครื่องจักรที่แล้วมาอีก ในเมื่อกลุ่มทางเลือกกำหนดไว้ว่า $A=\{a_1, a_2\}$ โดยที่ a_1 เป็นทางเลือก "ไม่ซื้อของนั้น" และ a_2 เป็น "ซื้อของ" แทนที่เราจะให้กลุ่มสภาวการณ์เป็นเซทที่มีเพียง 2 สภาวการณ์เท่านั้น เราจะสมมติให้สัดส่วนของเสียมีค่าใด ๆ ในช่วง (0, 1) นั่นคือ $W = (0, 1)$ ตามที่เราพิจารณามาแล้วนั้น ฟังก์ชั่นค่าเสียโอกาส L จะเป็น

$$\begin{aligned} L(a_1,w) &= 3000 - 7000\,w, \quad 0 < w \le .4286 \\ &= 0, \quad .4286 < w < 1.0 \end{aligned}$$

$$L(a_2,w) = 0, \quad 0 < w \leq .4286$$
$$= 7000w - 3000, \quad .4286 < w < 1.0$$

ถ้าสมมติว่าฟังก์ชันหนาแน่นน่าจะเป็นก่อนทดลองเป็นแบบบีทะ (ซึ่งกำหนดให้แก่กลุ่มสภาวการณ์ W) มี $\alpha = 2$ และ $\beta = 2$ ตามฟังก์ชันนี้เราสามารถคำนวณมาตราวัดแบบเบยส์ $B(a_i)$ ของทางเลือก $a_i \in A$ ได้ดังนี้

$$B(a_1) = \int_0^{.4286} (3000 - 7000w)\frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} w(1-w)dw$$
$$= 6\int_0^{.4286} (3000 - 10000w^2 - 7000w^3)dw$$
$$= 433.20$$

$$B(a_2) = \int_{.4286}^{1} (7000w - 3000)\frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} w(1-w)dw$$

$$= 528.20$$

เพราะฉะนั้นผู้จัดการควรจะเลือกทางเลือก a_1 นั่นคือเขาไม่ควรจะซื้อของนั้น

สมมติว่าต้องการจะหาข้อมูลข่าวสารเพิ่มเติมเกี่ยวกับสภาวการณ์นอกบังคับ w และแล้วก็ทำการตัดสินใจเลือกทางเลือกต่าง ๆ ถ้ากำหนดว่าเขาเลือกตัวอย่างสุ่มของ 2 ชิ้น x_1, x_2 และสังเกตจำนวนชิ้นที่เสียในตัวอย่าง สำหรับ x_i นั้นจะมีค่าเป็น 0 หรือ 1 ถ้าชิ้นส่วนนั้นเป็นของที่ใช้การได้ หรือใช้การไม่ได้ ตามปกติเราจะแทนค่าของ X_i ด้วย x_i ฉะนั้นถ้าผู้จัดการพบของที่ใช้การไม่ได้ในตัวอย่างเพียง 1 ชิ้น แล้วเราจะได้ว่า $\sum X_i = \sum x_i = 1$ โดยทั่ว ๆ ไป ถ้าพบ r ชิ้นที่เสีย แล้ว $\sum X_i = \sum x_i = r$ ถ้าสมมติว่าฟังก์ชันน่าจะเป็นของ w เป็นแบบทวินามที่มี $n = 2$ และ $p = w$ (w เป็นความน่าจะเป็นที่จะได้สิ่งที่ต้องการ นั่นคือชิ้นส่วนที่เสีย หรือใช้การไม่ได้) แล้วเราสามารถแสดงได้ว่าฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุงแล้ว $p(w/r)$ ก็จะเป็นฟังก์ชันหนาแน่นน่าจะเป็นแบบบีทะที่มีพารามิเตอร์ $\alpha + r$ และ $\beta + n - r$ ด้วย ในเมื่อ $r = \sum x_i$ นั่นคือ

$$p(w/r) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)} w^{\alpha+r-1}(1-w)^{\beta+n-r-1}, \quad 0 < w < 1$$

สำหรับ $r = 1$ นั่นคือผู้จัดการพบชิ้นส่วนเสีย 1 ชิ้น แล้วจะได้ว่า

$$p(w/r=1) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+1)\Gamma(\beta+n-1)} w^{\alpha}(1-w)^{\beta+n-2}$$

จากปัญหาที่กล่าวมานั้น เราพบว่า $\alpha = 2$, $\beta = 2$, และ $n = 2$ ฉะนั้นฟังก์ชั่นน่าจะเป็นที่ปรับปรุงแล้วของ w หลังจากสังเกต $Y = 1$ จะกำหนดไว้ดังนี้

$$p(w/Y=1) = \frac{\Gamma(6)}{\Gamma(3)\Gamma(3)} w^2(1-w)^2 = 30w^2(1-w)^2, \quad 0 < w < 1$$

เราจะใช้ฟังก์ชั่นที่ปรับปรุงแล้วคำนวณมาตราวัดแบบเบยส์ $B(a_1/Y = 1)$ ของทางเลือก a_i ใน A ดังนั้น

$$B(a_1/Y=1) = 30\int_0^{.4286}(3000 - 7000w)w^2(1-w)^2dw$$
$$= 327.00$$
$$B(a_2/Y=1) = 30\int_{.4286}^{1}(7000w - 3000)w^2(1-w)^2dw$$
$$= 825.20$$

เนื่องจาก $B(a_1/Y = 1) < B(a_2/Y = 1)$ ผู้จัดการควรจะเลือกทางเลือกนั้นคือเขาจะไม่ซื้อของนั้น ถ้าเขาพบชิ้นส่วนที่เสียเพียงหนึ่งจากตัวอย่างสุ่มขนาด 2 ดังนั้นประจักษ์พยานจากตัวอย่างไม่สามารถเปลี่ยนทางเลือกที่เหมาะสมที่สุดภายใต้ฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลองแบบบีทะ ถ้าเขาไม่พบ หรือพบส่วนประกอบที่ใช้การไม่ได้ 2 หน่วยแล้ว เขาอาจจะพบทางเลือกที่เหมาะสมแตกต่างกัน ถ้าพิจารณาตัวอย่างที่มีขนาดต่าง ๆ กัน เราก็จะได้ความจริงเช่นเดียวกัน

จากตัวอย่างข้างบนนี้ ข้อมูลข่าวสารที่สัมพันธ์กันของตัวอย่างสุ่มกำหนดไว้ด้วย $\sum X_i = Y$ ซึ่งก็เป็นข้อมูลข่าสารทั้งหมดที่ได้จากตัวอย่าง ตัวสถิติเช่นนี้ (คือ $\sum X_i = Y$) เราเรียกว่าตัวสถิติพอเพียง (Sufficient statistics) ซึ่งมีค่าอย่างมากในการวิเคราะห์ปัญหาการตัดสินใจ นั่นคือ (1) มันจะให้ข้อมูลข่าวสารที่สัมพันธ์กันทั้งหมดซึ่งตัวอย่างสุ่มจะสามารถให้ได้ และ (2) มันจะมีประโยชน์ในการหาการแจกแจงน่าจะเป็นที่ปรับปรุงแล้ว

ในการกำหนดรูปแบบของฟังก์ชั่นน่าจะเป็นก่อนทดลองชนิดบีทะนั้น ถ้าค่าคงที่ $\alpha = 2$ และ $\beta = 2$ เราก็มีวิธีกำหนดรูปแบบได้จากการพิจารณาค่าเฉลี่ย และความแปรปรวนจากประสบการณ์ จากค่าเฉลี่ย และความแปรปรวน เราก็จะได้ค่าของ α และ β สำหรับ $\alpha = 2$ และ $\beta = 2$ นั้นได้จากค่าเฉลี่ย และความแปรปรวนเป็น 0.5 และ 0.05 ตามลำดับ

เนื่องจากฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลองแบบบีทะของสภาวการณ์นอกบังคับ w ซึ่งมีพารามิเตอร์ α และ β นั้นจะมีค่าเฉลี่ย และความแปรปรวนดังนี้

$$E(w) = \frac{\alpha}{\alpha+\beta} \quad \text{และ} \quad V(w) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

แล้วเราจะได้ว่า

$$\frac{\alpha}{\alpha+\beta} = 0.5 \ , \quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.05$$

จากการแก้สมการทั้งสองนี้เราจะได้ $\alpha = 2$ และ $\beta = 2$ ซึ่งเป็นค่าที่เรากำหนดไว้นั้น

บางครั้งเป็นการยากที่จะกำหนดค่าประมาณของค่าเฉลี่ย และความแปรปรวนในการแจกแจงก่อนทดลองแบบบีทะ ในกรณีเช่นนี้วิธีการที่กล่าวมาจะใช้การไม่ได้ นั่นคือเราจะต้องใช้วิธีการอื่น

2. <u>การแจกแจงน่าจะเป็นที่ปรับปรุงเมื่อค่าสังเกตจากตัวอย่างสุ่มมาจากการแจกแจงทวินาม</u>

ลองพิจารณากลุ่มสภาวการณ์ W ที่เป็นช่วง $(0, 1)$ และสมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นแบบบีทะที่มีพารามิเตอร์ $\alpha > 0$ และ $\beta > 0$ ดังนั้นฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลอง $p(w)$ ของ w จะเป็น

$$p(w) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1}, \quad 0 < w < 1$$

$$\propto w^{\alpha-1}(1-w)^{\beta-1}$$

ในเมื่อ $\propto$ เป็นสัญลักษณ์แสดงสัดส่วนที่บอกว่าฟังก์ชั่น $p(w)/w^{\alpha-1}(1-w)^{\beta-1}$ เป็นค่าคงที่ไม่เกี่ยวกับ w

สมมติว่าผู้ตัดสินใจพิจารณาตัวอย่างสุ่ม $x_1, x_2, .., x_n$ ในเมื่อ x_i เป็นแบบทวินามแบบจุดที่มีพารามิเตอร์ w (ไม่ทราบค่า) แล้วฟังก์ชั่นน่าจะเป็น $f(x_1, x_2, .., x_n/w)$ ของตัวอย่างที่กำหนด w ให้ จะเป็นดังนี้

$$f(x_1, x_2, .., x_n/w) = \prod_{i=1}^{n} f(x_i/w)$$
$$= w^{r}(1-w)^{n-r}$$

ในเมื่อ $\sum x_i = r$

ในการปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลองแบบบีทะนี้จะใช้ทฤษฎีเบยส์ ถ้าเป็นฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลอง และ $f(x_1, x_2, .., x_n/w)$ เป็นฟังก์ชั่นน่าจะเป็น

แล้วฟังก์ชั่นหนาแน่นน่าจะเป็นที่ปรับปรุง กำหนดไว้เป็น

$$p(w/x_1, x_2, .., x_n) = \frac{f(x_1, x_2, .., x_n/w)\, p(w)}{\int_W f(r/w')\, p(w')\, dw'}$$

โดยที่ตัวส่วนจะไม่เกี่ยวข้องกับ w และสามารถแสดงสมการข้างบนนี้เป็น

$$\begin{aligned} p(w/x_1, x_2, .., x_n) &\propto f(x_1, x_2, .., x_n/w)\, p(w) \\ &\propto w^{r}(1-w)^{n-r}\, w^{\alpha-1}(1-w)^{\beta-1} \\ &\propto w^{\alpha+r-1}(1-w)^{\beta+n-r-1} \end{aligned}$$

ในเมื่อ $\sum x_i = r$ เราสามารถหาค่าคงที่ซึ่งแสดงสัดส่วน k โดยสมการ

$$k\int_W w^{\alpha+r-1}(1-w)^{\beta+n-r-1}\, dw = 1$$

แต่ $\int_0^1 w^{\alpha+r-1}(1-w)^{\beta+n-r-1}\, dw = \frac{\Gamma(\alpha+r)\Gamma(\beta+n-r)}{\Gamma(\alpha+\beta+n)}$

ดังนั้น $k = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)}$

และ $p(w/x_1, x_2, .., x_n) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+r)\Gamma(\beta+n-r)} w^{\alpha+r-1}(1-w)^{\beta+n-r-1}$

ซึ่งเป็นแบบบีทะ มีพารามิเตอร์ $\alpha+r$ และ $\beta+n-r$ ในเมื่อ $\sum x_i = r$

<u>ทฤษฎี</u> 1 สมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นการแจกแจงแบบบีทะที่มีพารามิเตอร์ α และ β โดยที่ $\alpha > 0$ และ $\beta > 0$ ให้ $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่ม โดยที่ X_i $(i = 1, 2, .., n)$ มีการแจกแจงทวินามแบบจุด มีพารามิเตอร์ w แล้วการแจกแจงน่าจะเป็นที่ปรับปรุงของ w เมื่อ $X_i = x_i$ $(i = 1, 2, .., n)$ เป็นการแจกแจงบีทะ มีพารามิเตอร์เป็น $\alpha + r$ และ $\beta + n - r$

ในการปรับปรุงการแจกแจงก่อนการทดลองนี้ยังมีกระบวนการอื่นอีกวิธีหนึ่ง วิธีนี้จะง่ายกว่าวิธีที่กล่าวมาแล้ว และจะใช้แนวคิดของตัวสถิติพอเพียง เพื่อพัฒนาการแจกแจงที่ปรับปรุง โดยเฉพาะถ้ามีตัวสถิติพอเพียงของมิติจำกัด (fixed dimension) สำหรับตัวอย่างสุ่ม

$x_1, x_2, \ldots, x_n$ ขนาด n แล้วเราสามารถหากลุ่ม (family) F ของการแจกแจงน่าจะเป็นก่อนทดลองของ w โดยที่การแจกแจงน่าจะเป็นที่ปรับปรุงของ w ต้องเป็นสมาชิกของ F กลุ่ม F จะมีข้อมูลข่าวสารมากพอที่จะอำนวยให้ผู้ตัดสินใจพิจารณาการแจกแจงที่แทนข้อมูลข่าวสารก่อนทดลอง และความเชื่อ ยังจะมีวิธีการทางคณิตศาสตร์ง่าย ๆ ที่จะหาการแจกแจงน่าจะเป็นที่ปรับปรุงสำหรับการแจกแจงก่อนทดลอง และตัวอย่างสุ่มที่กำหนดไว้ ยิ่งกว่านั้นจะถูกปิด (closed) ในแง่ที่ว่า ถ้าการแจกแจงก่อนทดลองเป็นสมาชิกของ F แล้วการแจกแจงที่ปรับปรุงโดยอาศัยตัวอย่างจะเป็นสมาชิกของ F ด้วย ต่อไปนี้เราจะหาเงื่อนไขสำหรับสร้างกลุ่ม F โดยการใช้แนวคิดของตัวสถิติพอเพียง

3. <u>ตัวสถิติพอเพียง และกลุ่มของการแจกแจงที่แปรไป</u> (Sufficient statistics and Conjugate Families of Distributions)

พิจารณาตัวอย่างสุ่ม $x_1, x_2, \ldots, x_n$ ขนาด n ในเมื่อ $x_i (i = 1, 2, \ldots, n)$ มีการแจกแจงทวินามแบบจุด มีพารามิเตอร์ w ความน่าจะเป็นร่วม (Joint likelihood) $p(x_1, x_2, \ldots, x_n/w)$ ของตัวอย่างสุ่ม โดยกำหนด w และ $X_i = x_i (i = 1, 2, \ldots, n)$ เป็นดังนี้

$$p(x_1, x_2, \ldots, x_n/w) = w^{\sum x_i}(1-w)^{n-\sum x_i}, \qquad 0 < w < 1$$

ถ้าเราต้องการทำการอ้างอิงใด ๆ เกี่ยวกับ W และ $p(\cdot/w)$ แล้วเราจะได้จากสมการข้างบนว่าข้อมูลข่าวสารทั้งหมดจากตัวอย่าง $x_1, x_2, \ldots, x_n$ คือ $\sum x_i$ ตัวอย่างเช่น ถ้าสุ่มตัวอย่างของผลิตภัณฑ์ ที่ผลิตมาจากกระบวนการผลิต และเราสนใจจำนวนของเสีย แล้ว $\sum x_i$ จะเป็นจำนวนของเสียทั้งหมดในตัวอย่างนั้น สำหรับลำดับที่ซึ่งของเสีย และของไม่เสียเกิดขึ้นนั้นไม่มีความสำคัญอะไร ดังนั้นผลรวม $\sum x_i$ จะสรุปข้อมูลข่าวสารทั้งหมดจากตัวอย่าง ผลสรุปนั้นเรียกว่า ตัวสถิติพอเพียง

ให้ T เป็นฟังก์ชันใด ๆ ของ $x_1, x_2, \ldots, x_n$ แล้ว T จะเรียกว่า ตัวสถิติพอเพียง การวิเคราะห์ปัญหาตัดสินใจเมื่อใช้ตัวสถิตินี้จะง่ายกว่าใช้ตัวแปรเชิงสุ่ม $x_1, x_2, \ldots, x_n$ เพียงตัวเดียว ยิ่งกว่านั้นถ้า T บรรจุข้อมูลข่าวสารทั้งหมดที่ตัวอย่างจะให้ได้ แล้วเราก็ไม่จำเป็นที่จะหาตัวสถิติอื่น ๆ สมมติว่าเราสังเกต $T(x_1, x_2, \ldots, x_n) = \sum x_i$ ตามตัวอย่างที่กล่าวมา ดังนั้น $\sum x_i$ จะสรุปข้อมูลข่าวสารทั้งหมดของตัวอย่าง ซึ่งเราหวังว่าความน่าจะเป็

ร่วม $p(x_1, x_2, .., x_n/w)$ ของ $X_1, X_2, .., X_n$ โดยกำหนด $T=t$ จะเป็นอิสระกับ w การอธิบายที่กล่าวมานี้ทำให้เรานิยามสถิติพอเพียงได้ต่อไปนี้ นิยามนี้ใช้ได้ทั้งการแจกแจงน่าจะเป็นแบบไม่ต่อเนื่อง และต่อเนื่อง

<u>นิยาม</u> ให้ $f(\cdot/w)$, $w \in W$ เป็นความน่าจะเป็น (likelihood) ของตัวแปรเชิงสุ่ม Y แล้วตัวสถิติ T ของ Y จะเรียกว่าตัวสถิติพอเพียง สำหรับ $p(\cdot/w)$ ถ้าการแจกแจงเงื่อนไขของ Y โดยกำหนด $T=t$ เป็นอิสระกับ w

ถ้า T เป็นตัวสถิติพอเพียง แล้วเราสามารถแสดงได้ว่า $T(X)$ และการแจกแจงน่าจะเป็นที่ปรับปรุงโดยกำหนด $X=x$ นั้นเป็นอย่างเดียวกัน ไม่ว่าการแจกแจงน่าจะเป็นก่อนทดลองจะเป็นแบบใด เราหวังผลเช่นนี้ก็เพราะว่า ถ้า T สรุปข้อมูลข่าวสารทั้งหมดจากตัวอย่าง แล้วการวิเคราะห์ใด ๆ โดยอาศัยตัวสถิติพอเพียงจะมีประสิทธิภาพเช่นเดียวกับการวิเคราะห์ใด ๆ โดยอาศัยค่าสังเกตทั้งหมดของตัวอย่าง

<u>ทฤษฎี</u> 2 ให้ $f(\cdot/w)$, $w \in W$ เป็นฟังก์ชันน่าจะเป็นของตัวแปรเชิงสุ่ม X และให้ T เป็นตัวสถิติพอเพียงสำหรับ $f(\cdot/w)$ แล้วการแจกแจงน่าจะเป็นที่ปรับปรุง $p(w/x)$ ของการแจกแจงก่อนทดลอง $p(w)$ โดยกำหนด $X=x$ นั้นจะขึ้นกับ x ทาง $T(x)$ เท่านั้น

ทฤษฎีนี้ใช้นิยามแนวคิดของตัวสถิติพอเพียงได้ แต่ไม่สะดวกที่จะนำไปประยุกต์สำหรับหาตัวสถิติพอเพียง ทฤษฎีต่อไปนี้จะแก้ไขสถานการณ์นั้น และจะเรียกว่าทฤษฎีตัวประกอบ (Factorization Theorem) ซึ่งจะให้เกณฑ์สำหรับหาตัวสถิติพอเพียง

<u>ทฤษฎี</u> 3 ให้ X เป็นตัวแปรเชิงสุ่มไม่ต่อเนื่อง และให้ $f(\cdot/w)$, $w \in W$ เป็นความน่าจะเป็นของ X แล้วตัวสถิติ $T = T(X)$ จะพอเพียงสำหรับ $f(\cdot/w)$ ก็ต่อเมื่อ $f(x/w)$ สามารถแยกตัวประกอบได้เป็น

$$f(x/w) = h(x)g(T(x), w)$$

สำหรับทุกค่าของ $X=x$ และ $w \in W$ ฟังก์ชัน h จะเป็นบวก และไม่ขึ้นอยู่กับ w และฟังก์ชัน g จะไม่เป็นลบ และขึ้นอยู่กับ x ทาง $T(x)$ เท่านั้น

ทฤษฎีนี้เป็นจริงสำหรับตัวแปรเชิงสุ่มแบบต่อเนื่องด้วย ลองพิจารณาตัวอย่างต่อไปนี้

<u>ตัวอย่าง</u> $p(x_1, x_2, .., x_n / w) = w^{\Sigma x_i}(1-w)^{n-\Sigma x_i}, \quad 0 < w < 1$

ให้ T เป็นตัวสถิติกำหนดไว้ว่า $T(x_1, x_2, .., x_n) = \sum x_i = r$

กำหนด $g(r, w) = w^r(1-w)^{n-r}$ และ $h(x_1, x_2, .., x_n) = 1$

แล้ว $p(x_1, x_2, .., x_n / w) = h(x_1, x_2, .., x_n)\, g(r, w)$

ฟังก์ชั่น h ไม่ขึ้นอยู่กับ w และฟังก์ชั่น g ขึ้นกับ $X_i = x_i\ (i = 1, 2, .., n)$ ทาง $T(x_1, x_2, .., x_n) = \sum x_i = r$ เท่านั้น

เพราะฉะนั้น T เป็นตัวสถิติพอเพียง

ถ้าเราสังเกตมิติ (dimension) ของตัวสถิติพอเพียงในตัวอย่างนี้ เราจะได้เป็น 1 และยังเป็นจริงอีกด้วยว่า มิติจะคงเดิมไม่ว่าขนาดของตัวอย่าง n จะเป็นเท่าใดเมื่อไรก็ตามที่ตัวสถิติมีคุณสมบัติเช่นว่านี้ก็จะเรียกว่ามีมิติจำกัด (fixed dimension)

<u>ตัวอย่าง</u> ให้ $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงปัวซองที่มีพารามิเตอร์ w ฟังก์ชั่นน่าจะเป็นร่วม $f(x_1, x_2, .., x_n/w)$ กำหนด $X_i = x_i\ (i = 1, 2, .., n)$ และ w คือ

$$f(x_1, x_2, \ldots, x_n | w) = e^{-nw}\, w^{\Sigma x_i} / \Pi x_i!$$

ให้ $T(x_1, x_2, .., x_n) = \sum x_i = r$

กำหนด $g(r, w) = e^{-nw}\, w^{\Sigma x_i}$

และ $h(x_1, x_2, .., x_n) = 1/\Pi x_i!$

แล้วเราสามารถแยกตัวประกอบฟังก์ชั่นน่าจะเป็นได้ดังนี้

$$f(x_1, x_2, \ldots, x_n | w) = h(x_1, x_2, \ldots, x_n)\, g(r, w)$$

ผลอันนี้จะเป็นจริงสำหรับทุก ๆ x_i และ $w \in W$ ฉะนั้นตัวสถิติ $T(x_1, x_2, .., x_n) = \sum x_i = r$ จะพอเพียงสำหรับ $f(\cdot / w)$

ต่อไปจะพัฒนาวิธีปรับปรุงการแจกแจงน่าจะเป็นก่อนทดลอง ในเมื่อค่าสังเกตจากตัวอย่าง (และฟังก์ชั่นน่าจะเป็นของตัวอย่าง) ให้ตัวสถิติพอเพียงที่มีมิติจำกัด วิธีนี้จะเกี่ยวกับการหากลุ่ม (family) ของการแจกแจงน่าจะเป็นที่มีคุณสมบัติว่า ถ้าการแจกแจงก่อนทด

ลองอยู่ในกลุ่มนี้แล้ว (สำหรับตัวอย่างขนาดใด ๆ n) การแจกแจงน่าจะเป็นที่ปรับปรุงต้องอยู่ในกลุ่มนี้ด้วย กลุ่มของการแจกแจงที่มีคุณสมบัติเช่นนี้เรียกว่า กลุ่มของการแจกแจงที่แปรไป (Conjugate family of distributions) วิธีการในการสร้างกลุ่มที่แปรไปจะเป็นดังนี้

ให้ $f(x_1, x_2, .., x_n/w)$ เป็นความน่าจะเป็นของตัวอย่างสุ่ม เมื่อกำหนดให้ $w \epsilon W$ ให้ $T(x_1, x_2, .., x_n) = t$ เป็นตัวสถิติพอเพียงที่มีมิติจำกัด โดยทฤษฎีตัวประกอบเราได้

$$\begin{aligned} f(x_1, x_2, .., x_n/w) &\propto g(t, w) \\ &\propto p(w/t) \end{aligned}$$

ดังนั้นฟังก์ชั่นน่าจะเป็นซึ่งพิจารณาว่าเป็นฟังก์ชั่นของ w จะเป็นสัดส่วนกับ $p(w/t)$ จงสังเกตว่า $p(w/t)$ ขึ้นอยู่กับตัวสถิติพอเพียง $T(x_1, x_2, .., x_n) = t$ ลองมาพิจารณากลุ่มของการแจกแจง $\{p(w/t)\}$ สำหรับขนาดตัวอย่างทั้งหมด n และค่าที่เป็นไปได้ทั้งหมด t ของตัวสถิติพอเพียง T เห็นได้ชัดว่าฟังก์ชั่นน่าจะเป็น $f(x_1, x_2, .., x_n/w)$ เป็นสัดส่วนกับ $p(w/t)$ ก็เช่นเดียวกัน กลุ่มนี้จะมีคุณสมบัติที่สำคัญอื่นอีกซึ่งจะแสดงได้ต่อไปนี้

ให้ t และ r เป็น 2 ค่าใด ๆ ของ T ภายใต้ 2 ตัวอย่างสุ่ม $x_1, x_2, .., x_n$ และ $y_1, y_2, .., y_m$ โดยมีคุณสมบัติว่า

$$T(x_1, x_2, .., x_n) = t; \quad T(y_1, y_2, .., y_m) = r$$

ให้ $f(x_1, x_2, .., x_n/w)$ และ $f(y_1, y_2, .., y_m/w)$ เป็นฟังก์ชั่นน่าจะเป็นที่สอดคล้องหรือสมนัยกับตัวอย่างเหล่านั้น เมื่อรวมตัวอย่างทั้งสองแล้วเราจะได้ฟังก์ชั่นน่าจะเป็นร่วมของตัวอย่างรวมเป็นดังนี้

$$\begin{aligned} f(x_1, x_2, .., x_n, y_1, .., y_m/w) &= f(x_1, x_2, .., x_n/w) f(y_1, y_2, .., y_m/w) \\ &\propto p(w/t)p(w/r) \quad \text{ตามสมการที่กล่าวมาแล้ว} \end{aligned}$$

กำหนด $T(x_1, x_2, .., x_n, y_1, .., y_m) = s$ แล้วเราจะได้ว่า

$$f(x_1, x_2, .., x_n, y_1, y_2, .., y_m/w) \propto p(w/z)$$

เพราะฉะนั้น $p(w/z) \propto p(w/t)p(w/r)$

กลุ่มใด ๆ ของการแจกแจงที่สอดคล้องสมการข้างบนนี้จะเรียกว่า ถูกปิดภายใต้การคูณ (closed under multiplication) ดังนั้นเราจะได้กลุ่มของการแจกแจงที่แปรไป $\{p(\cdot/t)\}$ ที่มีคุณสมบัติว่า (สำหรับขนาดตัวอย่างใด ๆ n)

– ฟังก์ชันน่าจะเป็น (likelihood function) เป็นสัดส่วนกับสมาชิกหนึ่งของกลุ่ม และ

– กลุ่มจะถูกปิดภายใต้การคูณ

เพราะฉะนั้นถ้าเราหาการแจกแจงน่าจะเป็นก่อนทดลองจากกลุ่มข้างบนที่สามารถแสดงความเชื่อก่อนทดลอง และข้อมูลข่าวสารของผู้ตัดสินใจโดยกำหนดข้อมูลข่าวสารจากตัวอย่างให้แล้วเราจะได้การแจกแจงน่าจะเป็นที่ปรับปรุงซึ่งสมนัยกันจากกลุ่มเดียวกัน ทั้งนี้เป็นเพราะว่ากลุ่มถูกปิดภายใต้การคูณ และจากความจริงที่ว่า

$$p(w/x_1, x_2, .., x_n) \propto f(x_1, x_2, .., x_n/w)\,p(w)$$
$$\propto p(w/t)\,p(w)$$

ในเมื่อ $T(x_1, x_2, .., x_n) = t$ ผลที่ได้นี้เป็นประโยชน์อย่างมาก และจะให้กฎง่าย ๆ สำหรับหาการแจกแจงน่าจะเป็นก่อนทดลองเหมาะ ๆ เราหากลุ่มของการแจกแจงน่าจะเป็นที่ถูกปิดโดยการคูณ และที่มีฟังก์ชันน่าจะเป็นได้สัดส่วนกับสมาชิกหนึ่งของกลุ่มสำหรับขนาดตัวอย่างใด ๆ และสำหรับค่าใด ๆ ของตัวสถิติพอเพียง แล้วเราเลือกสมาชิกหนึ่งของกลุ่มนี้เป็นการแจกแจงน่าจะเป็นก่อนการทดลอง และแล้วหาการแจกแจงน่าจะเป็นที่ปรับปรุงดังอธิบายข้างบน จงสังเกตว่ากลุ่มจะมีข้อมูลข่าวสารพอที่จะให้ผู้ตัดสินใจหาการแจกแจงซึ่งจะแทนความเชื่อ และข้อมูลข่าวสารก่อนทดลองเกี่ยวกับ $w \in W$ ตัวอย่างเช่น กลุ่มของการแจกแจงน่าจะเป็นบีทะจะสอดคล้องกับข้อกำหนดเหล่านี้ภายใต้การสุ่มทวินาม และในเวลาเดียวกันก็ยืดหยุ่นพอที่จะแทนความเชื่อและข้อมูลข่าวสารก่อนทดลองของผู้ตัดสินใจนั้น เพราะฉะนั้นควรจะพิจารณากลุ่มบีทะเสมือนกลุ่มที่แปรไปของการแจกแจงสำหรับปัญหาตัดสินใจต่าง ๆ พวกหนึ่ง ลองพิจารณาตัวอย่างที่กล่าวมาแล้วนั้น

เรามีตัวอย่างสุ่ม $x_1, x_2, .., x_n$ ในเมื่อ x_i $(i=1, 2, .., n)$ เป็นทวินามแบบจุดที่ มีพารามิเตอร์ w เพราะฉะนั้นฟังก์ชันน่าจะเป็นของตัวอย่างกำหนดไว้ด้วย

$$f(x_1, x_2, .., x_n/w) = w^r(1-w)^{n-r}$$

ในเมื่อ $T = \Sigma x_i = r$

ถ้าเราถือว่า $f(x_1, x_2, .., x_n/w)$ เป็นฟังก์ชั่นของ w เราจะเห็นว่ามันเป็นสัดส่วนกับสมาชิกหนึ่งในกลุ่มของการแจกแจงบีทะ นั่นคือ

$$f(x_1, x_2, .., x_n/w) \quad \propto \quad p(w/r+1,\ n-r+1)$$

ในเมื่อ $p(w/r+1,\ n-r+1)$ เป็นการแจกแจงบีทะ มีพารามิเตอร์ $r+1$ และ $n-r+1$ เช่นเดียวกัน กลุ่มของการแจกแจงบีทะจะถูกปิดภายใต้การคูณ นั่นคือถ้า $p(w/\alpha_1, \beta_1)$ และ $p(\alpha_2/\beta_2)$ ต่างก็มีการแจกแจงบีทะ แล้ว

$$\begin{aligned} p(w/\alpha_1, \beta_2)p(w/\alpha_2, \beta_2) &\propto w^{\alpha_1-1}(1-w)^{\beta_1-1} w^{\alpha_2-1}(1-w)^{\beta_2-1} \\ &\propto w^{\alpha_1+\alpha_2-2}(1-w)^{\beta_1+\beta_2-2} \\ &\propto w^{\alpha_3-1}(1-w)^{\beta_3-1} \\ &\propto p(w/\alpha_3, \beta_3) \end{aligned}$$

ในเมื่อ $\alpha_3 = \alpha_1 + \alpha_2 - 1$ และ $\beta_3 = \beta_1 + \beta_2 - 1$ ดังนั้น $p(w/\alpha_1, \beta_1)p(w/\alpha_2, \beta_2)$ จะมีการแจกแจงบีทะ มีพารามิเตอร์ α_3 และ β_3

ในตัวอย่างนั้นการแจกแจงน่าจะเป็นก่อนทดลองของ w สมมติให้เป็นการแจกแจงบีทะ มีพารามิเตอร์ α และ β เพราะฉะนั้นการแจกแจงที่ปรับปรุง $p(w/x_1, x_2, .., x_n)$ ของ w จะต้องอยู่ในกลุ่ม และกำหนดไว้โดย

$$\begin{aligned} p(w/x_1, x_2, .., x_n) &\propto f(x_1, x_2, .., x_n/w)p(w) \\ &\propto p(w/r+1,\ n-r+1)p(w/\alpha, \beta) \\ &\propto p(w/\alpha+r,\ \beta+n-r) \end{aligned}$$

ดังนั้นการแจกแจงน่าจะเป็นที่ปรับปรุงของ w เมื่อกำหนดค่าสังเกตจากตัวอย่าง $x_1, x_2, ..$ $.., x_n$ ก็จะมีการแจกแจงบีทะ มีพารามิเตอร์ $\alpha + r$ และ $\beta + n - r$ ด้วย ผลที่ได้ก็จะเป็นเช่นเดียวกับที่กล่าวมาแล้ว

กรณีพิเศษของตัวอย่างนี้ก็คือ เมื่อเราสมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นแบบยูนิฟอร์มในช่วง $(0,\ 1)$ การแจกแจงน่าจะเป็นก่อนทดลอง p ของ w ในกรณีนี้ก็คือ

$$p(w) \quad = \quad 1, \qquad 0 < w < 1$$

โดยที่การแจกแจงยูนิฟอร์มเป็นแบบบีทะอย่างหนึ่งซึ่งมีพารามิเตอร์ $\alpha = 1$ และ $\beta = 1$ เรา สามารถเขียนได้เป็น

$$p(w) = p(w/\alpha = 1, \beta = 1)$$

ในเมื่อทางขวามือของสมการจะแทนการแจกแจงบีทะ มี $\alpha = 1$ และ $\beta = 1$ เพราะฉะนั้น ถ้า $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่มที่ $X_i\,(i=1,2,..,n)$ เป็นตัวแปรเชิงสุ่มทวินามแบบจุด แล้วการแจกจแจงน่าจะเป็นที่ปรับปรุง $p(w/x_1, x_2, .., x_n)$ กำหนดว่า $X_i = x_i\,(i=1,2,...,n)$ สามารถพิจารณาจากสมการข้างบนได้โดยการแทน $\alpha = 1$ และ $\beta = 1$ ดังนั้น

$$p(w/x_1, x_2, .., x_n) \propto p(w/r+1, n-r+1)$$

รูปแบบที่สมบูรณ์ของฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุงสามารถเขียนได้เป็น

$$p(w/x_1, x_2, .., x_n) = \frac{\Gamma(n+2)}{\Gamma(r+1)\Gamma(n-r+1)} w^r (1-w)^{n-r}, \quad 0 < w < 1$$

ในเมื่อ $\sum x_i = r$

ในตอนต่อไปเราจะพิจารณาตัวอย่างสุ่มจากการแจกแจงปัวซอง และใช้การวิเคราะห์ที่กล่าวมานั้นสำหรับพิจารณาการแจกแจงน่าจะเป็นที่ปรับปรุงของการแจกแจงน่าจะเป็นก่อนทดลองที่กำหนดไว้ใด ๆ ของ $w \in W$

4. <u>การแจกแจงน่าจะเป็นที่ปรับปรุง เมื่อค่าสังเกตจากตัวอย่างนั้นสุ่มมาจากการแจกแจงแบบปัวซอง</u> (Poisson Distribution)

สมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นแบบแกมมะ (Gamma) ที่มีพารามิเตอร์ α และ ν โดยที่ $\alpha > 0$ และ $\nu \geq 1$ เพราะฉะนั้นฟังก์ชันหนาแน่นน่าจะเป็นก่อนทดลอง $p(w/\alpha, \nu)$ ของ w กำหนดไว้เป็น

$$p(w/\alpha, \nu) \propto w^{\nu-1} e^{-\alpha w}; \quad w > 0, \alpha > 0, \nu \geq 1$$

พิจารณาตัวอย่างสุ่ม $X_1, X_2, .., X_n$ จากการแจกแจงปัวซองที่มีพารามิเตอร์ w ในเมื่อ $w > 0$ ฟังก์ชันน่าจะเป็น $f(x_1, x_2, .., x_n/w)$ ของตัวอย่าง $X_i = x_i\,(i=1,2,...,n)$ คือ

$$\begin{aligned} f(x_1, x_2, .., x_n/w) &= \prod_{i=1}^{n} f(x_i|w) \\ &\propto e^{-nw} w^{\sum x_i} \end{aligned}$$

เราทราบว่า $T(x_1, x_2, .., x_n) = \sum x_i = r$ เป็นตัวสถิติพอเพียง เพราะฉะนั้น

$$f(x_1, x_2, .., x_n/w) \propto w^r e^{-nw}$$
$$\propto p(w/n, r+1)$$

นั่นคือฟังก์ชั่นน่าจะเป็นที่ถือว่าเป็นฟังก์ชั่นของ w นั้นเป็นสัดส่วนกับการแจกแจงแกมมะ มีพารามิเตอร์ n และ $r+1$ ยิ่งกว่านั้นกลุ่มการแจกแจงแกมมะจะถูกปิดภายใต้การคูณ เพราะฉะนั้นการแจกแจงน่าจะเป็นที่ปรับปรุง $p(w/x_1, x_2, .., x_n)$ ของ w เป็นการแจกแจงแบบแกมมะ มีพารามิเตอร์ $\alpha + n$ และ $\nu + r$ เพราะว่า

$$p(w/x_1, x_2, .., x_n) \propto p(w/n, r+1)p(w/\alpha, \nu)$$
$$\propto p(w/n+\alpha, \nu+r)$$

รูปแบบที่สมบูรณ์ของฟังก์ชั่นหนาแน่นน่าจะเป็นที่ปรับปรุงสามารถอธิบายได้ว่าเป็น

$$p(w/x_1, x_2, .., x_n) = \frac{\alpha+n}{\Gamma(\nu+r)}\{(\alpha+n)w\}^{\nu+r-1} e^{-(\alpha+n)w}; \quad \begin{array}{l}\alpha+n>0\\ w>0, \nu+r \geq 1\end{array}$$

ดังนั้นเราจะได้ทฤษฎีต่อไปนี้

<u>ทฤษฎี</u> 4 สมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นการแจกแจงแกมมะ มีพารามิเตอร์ α และ ν โดยที่ $\alpha > 0$ และ $\nu \geq 1$ และสมมติว่า $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงปัวซองที่มีพารามิเตอร์ w แล้วการแจกแจงน่าจะเป็นที่ปรับปรุงของ w เมื่อ $X_i = x_i$ $(i=1, 2, .., n)$ เป็นการแจกแจงแกมมะที่มีพารามิเตอร์ $\alpha+n$ และ $\nu + r$ ในเมื่อ $r = \sum x_i$

ในทำนองเดียวกันสำหรับตัวอย่างสุ่มที่มาจากการแจกแจงทวินามนิเสธ (Negative Binomial distribution) เราจะได้ทฤษฎีต่อไปนี้

<u>ทฤษฎี</u> 5 สมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ w เป็นแบบบีทะ มีพารามิเตอร์ α และ β โดยที่ $\alpha > 0$ และ $\beta > 0$ และสมมติว่า $X_1, X_2, .., X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงทวินามนิเสธที่มีพารามิเตอร์ k และ w โดยที่ k เป็นค่าที่ระบุไว้ ($k > 0$) และค่าของ w ไม่ทราบ แล้วการแจกแจงน่าจะเป็นที่ปรับปรุงของ w เมื่อ

$X_i = x_i$ $(i=1,2,..,n)$ เป็นการแจกแจงแบบปีทะ มีพารามิเตอร์ $\alpha + kn$ และ $\beta + \sum x_i$

การแจกแจงทวินามนิเสธมีฟังก์ชั่นน่าจะเป็นดังนี้

$$f(x) = \binom{x-1}{k-1} w^k (1-w)^{x-k} \; ; \; x = k-1, k-2, \ldots$$

ทฤษฎีต่อไปนี้จะเป็นการสุ่มตัวอย่างจากการแจกแจงพหุนาม (Multinomial) ที่มีฟังก์ชั่นน่าจะเป็นดังนี้

$$f(x_1, x_2, .., x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!} w_1^{x_1} w_2^{x_2} \quad w_k^{x_k} \; ;$$
$$x_i = 0, 1, .., n; \; \sum x_i = n; \; \sum w_i = 1$$

และสมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ W เป็นแบบดิริเชลท (Dirichlet) ที่มีพารามิเตอร์ $\alpha = (\alpha_1, \alpha_2, .., \alpha_k)'$ โดยที่ $\alpha_i > 0$ $(i = 1, 2, .., k)$ นั่นคือมีฟังก์ชั่นหนาแน่นน่าจะเป็น ดังนี้

$$f(w_1, w_2, \ldots, w_k) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \cdots \Gamma(\alpha_k)} w_1^{\alpha_1 - 1} w_2^{\alpha_2 - 1} \cdots w_k^{\alpha_k - 1} \; ;$$
$$w_i > 0 \; (i = 1, 2, \ldots k); \; \sum w_i = 1$$

<u>ทฤษฎี</u> 6 สมมติว่าการแจกแจงน่าจะเป็นก่อนทดลองของ $W = (W_1, W_2, .., W_k)'$ เป็นแบบดิริเชลท มีพารามิเตอร์ $\alpha = (\alpha_1, \alpha_2, .., \alpha_k)'$ โดยที่ $\alpha_i > 0$ $(i=1,2,..,k)$ และสมมติว่าเวคเตอร์เชิงสุ่ม (Random Vector) $X = (X_1, X_2, .., X_k)'$ เป็นตัวอย่างสุ่มจากการแจกแจงพหุนาม ที่มีพารามิเตอร์ n และ W โดยที่ n เป็นค่าเต็มบวกที่ระบุไว้ และค่าของส่วนประกอบของเวคเตอร์ W ไม่ทราบ แล้วการแจกแจงน่าจะเป็นที่ปรับปรุงของ W เมื่อ $X_i = x_i (i=1,2,..,n)$ จะเป็นแบบดิริเชลท มีพารามิเตอร์ $\alpha_0 = (\alpha_1 + x_1, \alpha_2 + x_2, \ldots, \alpha_k + x_k)'$