ทฤษฎีตัดสินใจทางสถิติ กับปัญหาการประมาณค่า

Statistical Decision Theory approach to Estimation Problems

ปัญหาการประมาณค่าในการวิเคราะห์การตัดสินใจจะเกี่ยวกับการกะประมาณสภาวการณ์นอกบังคับ โดยการใช้ข้อมูลข่าวสารจากตัวอย่างสุ่มของค่าสังเกต เรากำหนดฟังก์ชันของค่าสังเกตเหล่านี้ และทำการตัดสินใจเกี่ยวกับสภาวการณ์นอกบังคับที่ไม่ทราบค่า โดยอาศัยค่าของฟังก์ชันนั้น ตามปกติฟังก์ชันนั้นเรียกว่า ตัวประมาณค่า (Estimator) และค่าของตัวประมาณเหล่านี้จากค่าสังเกตของตัวอย่างที่กำหนดให้นั้นจะเรียกว่า ค่าประมาณ (Estimate) ของสภาวการณ์นอกบังคับที่ไม่ทราบค่า โดยปกติเราต้องการตัวประมาณค่าที่มีคุณสมบัติว่า ค่าประมาณที่คำนวณได้จากตัวอย่างจะพ้องกัน (coincide) กับสภาวการณ์นอกบังคับที่แท้จริงของปัญหาตัดสินใจที่เราสนใจ อย่างไรก็ตามอาจจะเป็นไปไม่ได้เพราะว่าปัญหาตัดสินใจส่วนมากเราไม่ทราบสภาวการณ์นอกบังคับที่แท้จริง และโดยทั่วไปค่าประมาณจะแตกต่างกันในตัวอย่างต่าง ๆ ซึ่งจะทำให้มีความคลาดเคลื่อน (error) บางอย่างในการกะประมาณสภาวการณ์นอกบังคับที่แท้จริง

การประมาณค่าสภาวการณ์นอกบังคับที่แท้จริงโดยการวิเคราะห์การตัดสินใจนั้น เราจะนำฟังก์ชันค่าเสียโอกาสเกี่ยวกับตัวประมาณค่า d และสภาวการณ์ที่แท้จริง w มาเกี่ยวข้องด้วย ฟังก์ชันนี้จะสะท้อนถึงค่าเสียโอกาสของการหาค่าประมาณที่มีความคลาดเคลื่อน และสมมติว่าฟังก์ชันนี้มีรูปฟอร์มเป็นดังนี้

$$L(d,w) = c(w)\, g(d-w)$$

ในเมื่อฟังก์ชัน $g(d-w) = 0$ สำหรับ $d = w$ และจะเพิ่มขึ้นเมื่อตัวประมาณค่า d บ่ายเบนจากสภาวการณ์นอกบังคับที่แท้จริง w. ให้ $p(w/x_1, x_2, \ldots, x_n)$ เป็นการแจกแจงน่าจะเป็นที่ปรับปรุงของการแจกแจงน่าจะเป็นก่อนทดลอง $p(w)$ ของ $w \in W$ หลังจากสังเกตค่าสังเกตจากตัวอย่าง $X_i = x_i (i=1,2,\ldots,n)$ แล้วสำหรับฟังก์ชันเสียโอกาสที่กล่าวไว้ข้างบนนี้ เราใช้การแจกแจงน่าจะเป็นที่ปรับปรุง $p(w/x_1, x_2, \ldots, x_n)$ นั้นมาคำนวณมาตราวัดแบบเบยส์สำหรับตัวประมาณค่า d และเลือกค่าประมาณ d^* ที่ทำให้มาตราวัดแบบเบยส์ต่ำสุด เราจะเรียกค่าประมาณ d^* นี้ว่า ค่าประมาณแบบเบยส์ (Bayes Estimate)

เราจะสมมติว่ามีฟังก์ชันค่าเสียโอกาสในรูปฟอร์มต่าง ๆ และหาค่าประมาณแบบเบยส์ของ w เมื่อ w มีการแจกแจงน่าจะเป็นก่อนทดลองต่าง ๆ กัน จงสังเกตว่าเซทของการตัดสินใจในปัญหาประมาณค่าจะเป็นตัวประมาณค่าของ w และฉะนั้นกลุ่มตัดสินใจ (Decision Space) จะพ้องกับกลุ่มสภาวการณ์ W

1. ปัญหาการประมาณค่า เมื่อฟังก์ชันค่าเสียโอกาสเป็นแบบเส้นตรง (Linear Loss Function)

ประการแรกจะพิจารณาตัวอย่างง่าย ๆ และพิจารณาค่าประมาณแบบเบยส์ของ w ที่เกี่ยวพันกับปัญหา ต่อไปจึงจะพัฒนาการวิเคราะห์ทั่ว ๆ ไป

สมมติว่าบริษัทผู้ผลิตวิทยุทรานซิสเตอร์ผลิตวิทยุพิเศษมาแบบหนึ่ง ความต้องการของวิทยุแบบนี้ในเดือนใด ๆ อยู่ระหว่าง 7 และ 11 บริษัทสนใจที่จะประมาณจำนวนวิทยุที่ควรสต๊อคไว้เพื่อให้พอกับความต้องการแต่ละเดือน สมมติบริษัทได้กำไรสุทธิ 260 บาทต่อเครื่องที่ขายได้ และจะขาดทุน 600 บาท สำหรับวิทยุที่ขายไม่ได้ในเดือนนั้น นั่นคือค่าเสียโอกาส 600 บาทต่อหน่วยเมื่อสต๊อคของไว้มากไป และค่าเสียโอกาส 260 บาท ต่อหน่วยเมื่อสต๊อคของไว้ไม่พอกับความต้องการ ให้ d เป็นค่าประมาณของจำนวนหน่วยที่ต้องสต๊อคไว้ และ w เป็นจำนวนหน่วยที่ลูกค้าต้องการ ดังนั้นฟังก์ชันค่าเสียโอกาส L สำหรับปัญหาที่กล่าวมานี้ สามารถกำหนดได้เป็น

$$\begin{aligned} L(d, w) &= 600(d - w), \quad w \leq d \\ &= 260(w - d), \quad w > d \end{aligned}$$

ค่าของ L สำหรับค่าประมาณ และความต้องการต่าง ๆ กัน กำหนดได้เป็นดังนี้

ตาราง 1 ค่าเสียโอกาสของค่าประมาณต่าง ๆ กัน

ความต้องการวิทยุ		7	8	9	10	11
สต๊อคที่กะไว้	7	0	260	520	780	1040
	8	600	0	260	520	780
	9	1200	600	0	260	520
	10	1800	1200	600	0	260
	11	2400	1800	1200	600	0

บริษัทผู้ผลิตวิทยุได้กำหนดฟังก์ชันมวลน่าจะเป็นก่อนทดลองของความต้องการวิทยุไว้เป็น

ตาราง 2 ฟังก์ชันมวลน่าจะเป็นก่อนทดลอง

w	7	8	9	10	11
$f(\cdot)$	.15	.19	.23	.27	.16

สมมติบริษัทเชื่อว่า ต้องหาข้อมูลข่าวสารเพิ่มเติมโดยการสุ่มตัวอย่างเพื่อปรับปรุงฟังก์ชันมวลน่าจะเป็นก่อนทดลองข้างบนนั้น และใช้ฟังก์ชันที่ปรับปรุงแล้วไปกะประมาณจำนวนหน่วยที่ต้องการสต๊อคไว้ สมมติว่าค่าสังเกตจากตัวอย่างมีการแจกแจงปัวซองที่มีพารามิเตอร์ w ซึ่งเป็นจำนวนเฉลี่ยของวิทยุที่ต้องการในเดือนใด ๆ เมื่อ $w \in W = \{7, 8, 9, 10, 11\}$ ดังนั้นถ้า $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่ม แล้วฟังก์ชันน่าจะเป็น $f(x_i/w)$ สำหรับ $X_i = x_i (i = 1, 2, \ldots, n)$ กำหนดไว้ด้วย

$$f(x_i/w) \quad = \quad e^{-w} w^{x_i}/x_i! \; , \qquad x_i = 0, 1, \ldots$$

เพื่อความสะดวกให้ s_1, s_2, และ s_3 เป็นสามเหตุการณ์ที่กำหนดไว้ดังนี้

$$s_1 = \{0 \le x_i \le 7\} \; , \quad s_2 = \{8 \le x_i \le 10\} \; , \quad s_3 = \{x_i \ge 11\}$$

ดังนั้นเราสามารถคำนวณความน่าจะเป็นของ $s_i \; (i = 1, 2, 3)$ ด้วย $f(x_i/w)$ จะได้ผลดังตารางต่อไปนี้

ตาราง 3 ความน่าจะเป็นของ $s_i \quad (i = 1, 2, 3)$

	S	s_1	s_2	s_3
W	7	.5987	.3028	.0985
	8	.4530	.3629	.1841
	9	.3239	.3821	.2940
	10	.2202	.3628	.4170
	11	.1432	.3167	.5401

ถ้าสมมติว่าเหตุการณ์ s_1 เกิดขึ้น เรามีวิธีพิจารณาค่าประมาณของจำนวนหน่วยที่ต้องสต๊อคไว้โดยอาศัยข้อมูลข่าวสารนี้ นั่นคือเราต้องปรับปรุงฟังก์ชันมวลน่าจะเป็นก่อนทดลอง โดยกำหนดว่า s_1 เกิดขึ้นแล้ว แล้วก็ใช้ฟังก์ชันที่ปรับปรุงนั้นไปคำนวณมาตราวัด

แบบเบยส์ของค่าประมาณต่าง ๆ กัน เราจะเลือกค่าประมาณที่ทำให้มาตราวัดแบบเบยส์น้อยที่สุด ความน่าจะเป็นที่ปรับปรุงสำหรับ $w_j \in W$ โดยกำหนด s_i $(i=1,2,3)$ ให้นั้นคำนวณจากทฤษฎีของเบยส์

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

ซึ่งเราจะได้ $p(w_j / s_i)$ ตามตารางต่อไปนี้

<u>ตาราง</u> 4 การแจกแจงน่าจะเป็นที่ปรับปรุง $p(w_j / s_i)$

S	s_1	s_2	s_3
$p(w_1/s_i)$	.2699	.1294	.0468
$p(w_2/s_i)$	.2588	.1963	.1106
$p(w_3/s_i)$	.2239	.2505	.2136
$p(w_4/s_i)$	.1786	.2793	.3559
$p(w_5/s_i)$	.0688	.1445	.2731

เราสามารถคำนวณมาตราวัดแบบเบยส์ของค่าประมาณ d เมื่อกำหนด s_i $(i=1,2,3)$ เป็นผลทดลองของตัวอย่าง ซึ่งเราจะแทนด้วย $B(d/s_i)$ และกำหนดไว้ดังนี้

$$\begin{aligned} B(d/s_i) &= \sum_{w \in W} L(d,w)\, p(w/s_i) \\ &= \sum_{w \le d} 600(d-w)\, p(w/s_i) + \sum_{w > d} 260(w-d)\, p(w/s_i) \end{aligned}$$

สำหรับ $d = 7$ และผลทดลองของตัวอย่างเป็น s_1 เราจะได้ $B(d=7/s_1)$ ดังนี้

$$\begin{aligned} B(d=7/s_1) &= 0(.2699) + 260(.2588) + 520(.2239) \\ &\quad + 780(.1786) + 1040(.0688) \\ &= 394.60 \end{aligned}$$

สำหรับมาตราวัดแบบเบยส์สำหรับ d และ s_i ใด ๆ เราจะคำนวณได้ดังนี้

<u>ตาราง</u> 5 มาตราวัดแบบเบยส์ $B(d/s_i)$

S	s_1	s_2	s_3
$B(d=7/s_i)$	394.6	549.4	701.4

	s_1	s_2	s_3
$B(d=8/s_i)$	<u>366.7</u>	<u>400.7</u>	481.7
$B(d=9/s_i)$	561.4	420.8	<u>357.1</u>
$B(d=10/s_i)$	948.6	656.4	416.1
$B(d=11/s_i)$	1489.4	1132.1	781.3

จากการเปรียบเทียบมาตรารวัดแบบเบยส์ เราจะเห็นว่าค่าประมาณแบบเบยส์จะเป็น $d_1^* = 8$, $d_2^* = 8$, และ $d_3^* = 9$ นั่นคือถ้า s_1 เป็นผลทดลองของตัวอย่างแล้วบริษัทควรจะสต๊อควิทยุไว้ 8 เครื่อง ถ้า s_2 เป็นผลทดลอง จะสต๊อคไว้ 8 เครื่อง และถ้า s_3 เป็นผลทดลอง บริษัทจะสต๊อควิทยุไว้ 9 เครื่อง

ในการประมาณค่าแบบนี้ เราสามารถใช้การวิเคราะห์แบบเพิ่ม (Marginal or Incremental Analysis) มาช่วยหาค่าประมาณแบบเบยส์ได้

สำหรับฟังก์ชั่นมวลน่าจะเป็นทางเดียว (marginal P.M.F.) ของ s_i ($i=1,2,3$) คือ $p(s_i) = \sum_{w_j \in W} p(s_i/w_j)p(w_j)$ เราคำนวณได้เป็น

$$p(s_1) = .3327, \quad p(s_2) = .3509, \quad p(s_3) = .3164$$

จากผลอันนี้เราสามารถคำรวณการเสี่ยงแบบเบยส์ (Bayes Risk) หรือ $B(d^*)$ ได้ด้วยสมการ

$$B(d^*) = \sum_{s_i \in S} B(d^*/s_i)\, p(s_i)$$

ดังนั้น $B(d^*) = 366.7(.3327) + 400.7(.3509) + 357.1(.3164)$
$= 376.5$

ฉะนั้นในการประมาณจำนวนหน่วยที่จะสต๊อค เราจะสมมติว่าค่าเสียโอกาส $L(d,w)$ เพิ่มขึ้นเป็นแบบเดียวกัน เมื่อตัวประมาณค่า d ห่างจากค่าจริง w ฟังก์ชั่น L จะสะท้อนถึงความแตกต่างกัน (discrepancy) ระหว่างตัวประมาณค่า d และสภาวการณ์ที่แท้จริง w ดังนั้นหลังจากรวบรวมข้อมูลข่าวสารจากตัวอย่าง เราได้ปรับปรุงฟังก์ชั่นน่าจะเป็นก่อนทดลอง และเลือกค่าประมาณแบบเบยส์ที่ทำให้มาตรารวัดแบบเบยส์น้อยที่สุด

การวิเคราะห์นี้จะใช้ได้กับปัญหาที่มีการแจกแจงน่าจะเป็นก่อนทดลองแบบต่อเนื่อง

ด้วย ต่อไปเราจะพิจารณากรณีต่อเนื่องนี้

ให้ $$L(d,w) = \ell_o(d-w)\ , \quad w \le d$$
$$= \ell_u(w-d)\ , \quad w > d$$

ในเมื่อ ℓ_o และ ℓ_u เป็นค่าเสียโอกาส เนื่องจากการประมาณสต็อคมากไป และน้อยเกินไปตามลำดับ

และให้ $f(w/x)$ เป็นฟังก์ชั่นหนาแน่นน่าจะเป็นที่ปรับปรุงของฟังก์ชั่นหนาแน่นน่าจะเป็นก่อนทดลองที่กำหนดไว้ $f(w)$ หลังจากสังเกต $X = x$ แล้วมาตราวัดแบบเบยส์ $B(d/x)$ ของตัวประมาณค่า d ตาม $f(w/x)$ กำหนดไว้เป็น

$$\begin{aligned}B(d/x) &= \int_{-\infty}^{\infty} L(d,w) f(w/x)\,dw \\ &= \ell_o\int_{-\infty}^{d}(d-w) f(w/x)\,dw + \ell_u\int_{d}^{\infty}(w-d) f(w/x)\,dw \\ &= \ell_o d F(d/x) - \ell_u d[1 - F(d/x)] \\ &\quad - \ell_o\int_{-\infty}^{d} w f(w/x)\,dw + \ell_u\int_{d}^{\infty} w f(w/x)\,dw\end{aligned}$$

ในเมื่อ $$F(d/x) = \int_{-\infty}^{d} f(w/x)\,dw$$

เราพิจารณาค่าประมาณแบบเบยส์ได้โดยหาค่า d^* ที่ทำให้ $B(d/x)$ มีค่าน้อยที่สุด จากวิธีการแคลคูลัส เราได้

$$\begin{aligned}0 &= \ell_o\{F(d^*/x) + d^*F'(d^*/x)\} \\ &\quad - \ell_u\{(1 - F(d^*/x)) - d^*F'(d^*/x)\} \\ &\quad - \ell_o d^* f(d^*/x) - \ell_u d^* f(d^*/x) \\ &= \ell_o F(d^*/x) - \ell_u + \ell_u F(d^*/x)\end{aligned}$$

{เพราะว่า $f(w/x)$ เป็นฟังก์ชั่นหนาแน่นน่าจะเป็นแบบต่อเนื่อง เราจึงได้ $F'(d^*/w) = f(d^*/x)$}

เพราะฉะนั้น $$F(d^*/x) = \ell_u/(\ell_u + \ell_o)$$

ดังนั้นในปัญหาที่มีฟังก์ชั่นค่าเสียโอกาสเป็นเส้นตรง ค่าประมาณแบบเบยส์พิจารณาได้ว่าเป็นแฟรคไทล์ (Fractile) หรือควอนไทล์ (Quantile) ที่ $\ell_u/(\ell_u+\ell_o)$ ของ

ฟังก์ชั่นหนาแน่นน่าจะเป็นที่ปรับปรุง $f(w/x)$ นั่นคือค่าประมาณแบบเบยส์ d^* จะขึ้นอยู่กับ ℓ_u และ ℓ_o เมื่อ ℓ_u เพิ่มขึ้น d^* ก็จะเพิ่มขึ้นด้วย เพราะว่าการเพิ่มขึ้นของ ℓ_u สะท้อนถึงค่าเสียหาย หรือค่าชดเชย (penalty) ที่มากขึ้นของทุกหน่วยที่ไม่พอกับความต้องการ ในทำนองเดียวกัน การเพิ่มขึ้นของ ℓ_o จะสะท้อนถึงการสูญเสีย (loss) ของสิ่งที่เหลือ เพราะฉะนั้น d^* จะลดลง

สำหรับมาตราวัดแบบเบยส์ของ d^* หลังจากสังเกต $X=x$ หาได้ดังนี้

$$B(d^*/x) = \int_{-\infty}^{\infty} L(d^*,w) f(w/x)\,dw$$

$$\begin{aligned}
B(d^*/x) &= \ell_o d^* F(d^*/x) - \ell_u d^*(1 - F(d^*/x)) \\
&\quad - \ell_o \int_{-\infty}^{d^*} w f(w/x)\,dw + \ell_u \int_{d^*}^{\infty} w f(w/x)\,dw \\
&= (\ell_u + \ell_o) d^* F(d^*/x) - \ell_u d^* - \ell_o \int_{-\infty}^{d^*} w f(w/x)\,dw \\
&\quad + \ell_u \int_{d^*}^{\infty} w f(w/x)\,dw \\
&= \ell_u \int_{d^*}^{\infty} w f(w/x)\,dw - \ell_o \int_{-\infty}^{d^*} w f(w/x)\,dw \\
&= \ell_u \int_{-\infty}^{\infty} w f(w/x)\,dw - (\ell_u + \ell_o) \int_{-\infty}^{d^*} w f(w/x) \\
&= \ell_u E(w/x) - (\ell_u + \ell_c) \int_{-\infty}^{d^*} w f(w/x)\,dw \\
&= \ell_u E(w/x) - (\ell_u + \ell_o) E^{d^*}(w/x)
\end{aligned}$$

ในเมื่อ $E^{d^*}(w/x)$ เรียกว่า ค่าคาดหวังบางส่วน (Partial Expectation) ของ w และกำหนดไว้ว่า

$$E^{d^*}(w/x) = \int_{-\infty}^{d^*} w f(w/x)\,dw$$

เราสามารถสรุปผลต่าง ๆ ได้ดังทฤษฎีต่อไปนี้

<u>ทฤษฎี</u> 1 ให้ $L(d,w)$ เป็นฟังก์ชั่นค่าเสียโอกาส ที่กำหนดไว้ดังนี้

$$\begin{aligned}
L(d,w) &= \ell_o(d - w), \quad w \le d \\
&= \ell_u(w - d), \quad w > d
\end{aligned}$$

ในเมื่อ ℓ_o และ ℓ_u เป็นค่าเสียโอกาสเนื่องจากสต็อคไว้ หรือกะประมาณไว้มากและน้อยเกินไป ตามลำดับ และให้ $f(w/x)$ เป็นฟังก์ชั่นหนาแน่นน่าจะเป็นที่ปรับปรุงของ $f(w)$

หลังจากสังเกต $X = x$ แล้วค่าประมาณ d^* ที่สอดคล้องสมการ

$$F(d^*/x) = l_u/(l_u + l_o)$$

จะเป็นค่าประมาณแบบเบยส์ของ w ที่มีมาตรารวัดแบบเบยส์ $B(d^*/x)$ ดังนี้

$$B(d^*/x) = l_u E(w/x) - (l_u + l_o) E^{d^*}(w/x)$$

ในเมื่อ $$E^{d^*}(w/x) = \int_{-\infty}^{d^*} w f(w/x)\, dw$$

<u>ตัวอย่าง</u> กำหนด

$$L(d, w) = 50(d - w), \quad w \leq d$$
$$= 20(w - d), \quad w > d$$

สมมติฟังก์ชันหนาแน่นน่าจะเป็นก่อนทดลองของ w เป็นแบบปกติที่มีค่าเฉลี่ย μ และความเที่ยงตรงที่ทราบ η ให้ x เป็นค่าสังเกตของตัวอย่างจากการแจกแจงปกติ มีค่าเฉลี่ยไม่ทราบค่า และความเที่ยงตรงที่ทราบค่า h_0 แล้วเราจะได้ว่าฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุงของ w เมื่อ $X=x$ จะเป็นแบบปกติ มีค่าเฉลี่ย μ' และความเที่ยงตรง $h_0 - \eta$ ในเมื่อ

$$\mu' = \frac{h_0 x + \mu}{h_0 + \eta}$$

นั่นคือฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุง $p(w/x)$ ของ w เมื่อ $X = x$ กำหนดไว้ดังนี้

$$p(w/x) = N(w; \mu', h_0 + \eta)$$
$$= \sqrt{\frac{h_0+\eta}{2\pi}}\, e^{-\frac{h_0+\eta}{2}(w-\mu')^2}, \quad -\infty < \mu < \infty$$

สำหรับค่าประมาณแบบเบยส์ d^* ของ w พิจารณาจาก

$$F(d^*/x) = l_u/(l_u + l_o)$$
$$= 20/(20 + 50) = 0.29$$

หรือ $$\int_{-\infty}^{d^*} N(w; \mu', h_0 + \eta)\, dw = .29$$

การเสี่ยงแบบเบยส์ของ d^* กำหนด $X = x$ ให้นั้น จะหาได้จากสมการ

$$B(d^*/x) = l_u E(w/x) - (l_u + l_o) E^{d^*}(w/x)$$

เมื่อ $$E^{d^*}(w/x) = \int_{-\infty}^{d^*} w N(w; \mu', h_0 + \eta)$$

$$
\begin{aligned}
E^{d^*}(w/x) &= \int_{-\infty}^{d^*}(w-\mu'-\mu')\,N(w;\mu',h_0+\eta)\,dw \\
&= \int_{-\infty}^{d^*}(w-\mu')\,N(w;\mu',h_0+\eta)\,dw+\mu' l_u/(l_u+l_0)
\end{aligned}
$$

ดังนั้น

$$
\begin{aligned}
B(d^*/x) &= -(l_u+l_0)\int_{-\infty}^{d^*}(w-\mu')\,N(w;\mu',h_0+\eta)\,dw \\
&= -(l_u+l_0)\left\{\frac{-1}{\sqrt{h_0+\eta}}\,\psi\left(\sqrt{h_0+\eta}\,(d^*-\mu')\right)\right\} \\
&= (l_u+l_0)\frac{1}{\sqrt{h_0+\eta}}\,\psi\left(\sqrt{h_0+\eta}\,(d^*-\mu')\right)
\end{aligned}
$$

2. <u>ปัญหาการประมาณค่า เมื่อฟังก์ชั่นค่าเสียโอกาสเป็นแบบกำลังสอง</u> (Quadratic Loss Function)

บางครั้งฟังก์ชั่นค่าเสียโอกาสในปัญหาตัดสินใจที่เกี่ยวกับการประมาณค่า จะเป็นฟังก์ชั่นแบบกำลังสอง ซึ่งกำหนดไว้ดังนี้

$$L(d,w) = c(d-w)^2, \quad c > 0$$

ฟังก์ชั่น L จะเป็นศูนย์ (vanishes) สำหรับ $d=w$ และเพิ่มขึ้นเป็นแบบเดียวกัน เมื่อ d เคลื่อนห่างจาก w ฟังก์ชั่นเสียโอกาสส่วนมากที่เป็นศูนย์เมื่อ $d=w$ นั้นประมาณได้ดีด้วยฟังก์ชั่นกำลังสองดังที่ระบุไว้ข้างบนนี้ การวิเคราะห์เชิงคณิตศาสตร์ที่เกี่ยวกับฟังก์ชั่นแบบนี้จะง่าย เมื่อเทียบกับฟังก์ชั่นเสียโอกาสแบบอื่น ๆ ต่อไปเราจะพิจารณาตัวประมาณค่าแบบเบยส์ เมื่อฟังก์ชั่นเสียโอกาสเป็นแบบกำลังสอง ดังที่กล่าวมาข้างบนนี้

ให้ $p(w/x)$ แทนการแจกแจงน่าจะเป็นที่ปรับปรุงของ w เมื่อ $X=x$ แล้วค่าประมาณแบบเบยส์ d^* ได้จากการทำให้มาตรารวัดแบบเบยส์ $B(d/x)$ มีค่าน้อยที่สุด แต่อย่างไรก็ตาม

$$
\begin{aligned}
B(d/x) &= E[L(d,w)] = E[c(d-w)^2] \\
&= cE\{(d-E(w/x))-(w-E(w/x))\}^2 \\
&= c\{(w-E(w/x))^2+E(d-E(w/x))^2 \\
&\qquad -2E(d-E(w/x))(w-E(w/x))\} \\
&= cV(w/x)+[d-E(w/x)]^2 \\
&\qquad -2(d-E(w/x))\,E(w-E(w/x)) \\
&= cV(w/x)+(d-E(w/x))^2
\end{aligned}
$$

จากข้อสมมติที่ว่า $c > 0$ เราจะเห็นว่า เทอมแรกไม่เป็นลบ เทอมที่สองก็ไม่เป็นลบด้วย และจะเป็นศูนย์ก็ต่อเมื่อ $d = E(w/x)$ นี่แสดงว่า $B(d/x)$ จะน้อยที่สุดก็ต่อเมื่อ $d = E(w/x)$ ดังนั้นค่าประมาณแบบเบยส์ของ w ก็คือ $d^* = E(w/x)$ และค่าเสียโอกาสคาดหวังของ d^* หลังจากสังเกต $X=x$ จะเป็น $cV(w/x)$ ผลที่ได้มานี้จะสรุปได้ดังทฤษฎีต่อไปนี้

<u>ทฤษฎี</u> 2 ให้ $L(d,w) = c(d-w)^2$ เป็นฟังก์ชันเสียโอกาส และให้ $p(w/x)$ เป็นการแจกแจงน่าจะเป็นที่ปรับปรุงหลังจากสังเกต $X=x$ ของการแจกแจงก่อนทดลองที่กำหนดให้ $p(w), w\in W$ สมมติค่าเฉลี่ย $E(w/x)$ ของการแจกแจงที่ปรับปรุงมีค่าเป็นจำนวนจำกัด แล้วค่าประมาณแบบเบยส์ d^* ของ w กำหนดว่า $X=x$ คือ $E(w/x)$ และค่าเสียโอกาสคาดหวัง $B(d^*/x)$ จะเป็น $cV(w/x)$

ถ้า $E(w/x)$ หาค่าไม่ได้ ปัญหาตัดสินใจก็จะหาค่าประมาณแบบเบยส์ไม่ได้เช่นกัน สำหรับ $V(w/x)$ นั้นอาจจะหาค่าไม่ได้ ถึงแม้ $E(w/x)$ หาค่าได้ อย่างไรก็ตามปัญหาส่วนมากนั้นเราสมมติว่าทั้งสองอย่างหาค่าได้

<u>ตัวอย่าง</u> ให้ฟังก์ชันเสียโอกาส และการแจกแจงน่าจะเป็นก่อนทดลองสำหรับปัญหาตัดสินใจกำหนดไว้ด้วย

$$L(d,w) = (d-w)^2$$

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

สมมติฟังก์ชันน่าจะเป็น $f(x/w)$ หลังจากสังเกต $\Sigma x_i = x$ นั้นเป็นแบบทวินาม นั่นคือ

$$f(x/w) = \binom{n}{x} w^x(1-w)^{n-x} \quad , \quad x = 0,1,\ldots,n \quad 0 < w < 1$$

เราจะหาค่าประมาณแบบเบยส์ d^* ของ w และ $V(w/x)$ เมื่อกำหนด $\Sigma x_i = x$ ให้ เราจะได้ว่าการแจกแจงน่าจะเป็นที่ปรับปรุง $p(w/x)$ หลังจากสังเกต $\Sigma x_i = x$ นั้นเป็นแบบบีทะ มีพารามิเตอร์ $\alpha + x$ และ $\beta + n - x$ นั่นคือ

$$p(w/x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} w^{\alpha+x-1}(1-w)^{\beta+n-x-1}$$

จากการแจกแจงนี้ เราหาค่าเฉลี่ย และความแปรปรวนได้เป็น

$$E(w/x) = \frac{\alpha+x}{\alpha+\beta+n}, \quad V(w/x) = \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$$

เพราะฉะนั้นค่าประมาณแบบเบยส์ และค่าเสียโอกาสคาดหวังของค่าประมาณแบบเบยส์หลังจากสังเกต $\Sigma x_i = x$ คือ

$$d^* = \frac{\alpha+x}{\alpha+\beta+n}, \quad B(d^*/x) = \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$$

<u>ตัวอย่าง</u> ให้ฟังก์ชันเสียโอกาสเป็น $L(d,w) = (d-w)^2$ และการแจกแจงน่าจะเป็นก่อนทดลอง $p(w)$ เป็นแบบแกมมะ มีพารามิเตอร์ $\alpha > 0$ และ $\nu \geq 1$

$$p(w;\alpha,\nu) = \frac{\alpha}{\Gamma(\nu)}(\alpha w)^{\nu-1} e^{-\alpha w}$$

สมมติว่าเราเลือกค่าสังเกตของตัวอย่างเพียงหนึ่งจากการแจกแจงปัวซอง ดังนั้นฟังก์ชันน่าจะเป็น $f(x/w)$ สำหรับ $X=x$ กำหนดไว้ดังนี้

$$f(x/w) = e^{-w} w^x / x!, \quad x = 0,1,\ldots$$

สำหรับการแจกแจงน่าจะเป็นที่ปรับปรุง สามารถพิจารณาได้ว่าเป็น

$$p(w/x;\alpha,\nu) = \frac{\alpha+1}{\Gamma(\nu+x)}[(\alpha+1)w]^{\nu+x-1} e^{-(\alpha+1)w}$$

จากการแจกแจงนี้เราคำนวณค่าเฉลี่ย และความแปรปรวนได้เป็น

$$E(w/x) = \frac{\nu+x}{\alpha+1}, \quad V(w/x) = \frac{\nu+x}{(\alpha+1)^2}$$

เพราะฉะนั้นค่าประมาณแบบเบยส์ d^* และ $B(d^*/x)$ จากการสังเกต $X=x$ กำหนดไว้เป็น

$$d^* = \frac{\nu+x}{\alpha+1}, \quad B(d^*/x) = \frac{\nu+x}{(\alpha+1)^2}$$

<u>ตัวอย่าง</u> ให้ฟังก์ชันเสียโอกาสเป็น $L(d,w) = (d-w)^2$ และฟังก์ชันหนาแน่นน่าจะเป็นก่อนทดลอง $p(w)$ เป็นแบบปกติ มีค่าเฉลี่ย μ และความเที่ยงตรงที่ทราบค่า h

สมมติว่าสุ่มตัวอย่างจากการแจกแจงปกติ ที่ไม่ทราบค่าเฉลี่ย w และความเที่ยงตรงที่ทราบค่า h_0 หลังจากสังเกต $X=x$ แล้วเราจะได้ฟังก์ชันน่าจะเป็น $f(x/w)$ ที่กำหนดไว้ดังนี้

$$f(x/w) = \sqrt{\frac{h_0}{2\pi}}\, e^{-\frac{h_0}{2}(x-w)^2}$$

สำหรับการแจกแจงน่าจะเป็นที่ปรับปรุง $p(w/x)$ ก็จะเป็นแบบปกติ มีค่าเฉลี่ย μ' และความเที่ยงตรง η' ในเมื่อ

$$\mu' = \frac{h_0 x + \eta_0 \mu}{h_0 + \eta_0}, \quad \eta' = h_0 + \eta_0$$

เพราะฉะนั้นค่าประมาณแบบเบยส์ และค่าเสียโอกาสคาดหวัง กำหนดไว้ดังนี้

$$d^* = \frac{h_0 x + \eta_0 \mu}{h_0 + \eta_0}, \quad B(d^*/x) = \frac{1}{h_0 + \eta_0}$$

การวิเคราะห์ในตัวอย่าง 2 ตัวอย่างท้ายนี้ จะเกี่ยวข้องกับการสังเกตค่าสังเกตของตัวอย่างสุ่มเพียงหนึ่งเท่านั้น แต่เราสามารถขยายเป็นการสังเกตตัวอย่างสุ่มจำนวนจำกัดใด ๆ $X_i = x_i$ $(i=1,2,..,n)$ ก็ได้

สำหรับฟังก์ชันเสียโอกาสที่กล่าวไว้ว่าเป็น

$$L(d,w) = c(d-w)^2, \quad c > 0$$

นั้นสามารถกำหนดไว้เป็นแบบฟอร์มทั่วไปได้ดังนี้

$$L(d,w) = c(w)(d-w)^2$$

ในเมื่อ $c(w) > 0$ สำหรับทุก ๆ $w \in W$ ตัวอย่างเช่น $L(d,w)$ อาจจะเป็น

$$L(d,w) = (1/w)(d-w)^2$$

หรือ

$$L(d,w) = (1/w(1-w))(d-w)^2$$

3. <u>การกำหนดตัวอย่างที่เหมาะสม</u> (Optimal Sample Size)

การวิเคราะห์ที่ผ่านมา เราสนใจแต่จะพิจารณาค่าประมาณแบบเบยส์ของ w และค่าเสียโอกาสคาดหวังของมันสำหรับตัวอย่างสุ่มขนาด n ต่อไปเราจะพิจารณาขนาดตัวอย่างสุ่มที่เหมาะสม ตามปกติจะมีค่าใช้จ่ายของการสุ่มตัวอย่าง $c(n)$ ด้วย ซึ่งเราจะกำหนดให้เป็น

$$c(n) = K + kn, \quad K \geq 0, k \geq 0$$

ในเมื่อ K แทนค่าใช้จ่ายในการทดลองซึ่งคงที่ และ k แทนค่าใช้จ่ายแปรผันของหน่วยทดลอง

จากทฤษฎี 2 เราทราบว่าค่าเสียโอกาสคาดหวังของค่าประมาณแบบเบยส์ d^* ของ w หลังจากสังเกต $X = x$ คือ

$$B(d^*/x) = V(w/X = x)$$

ในเมื่อฟังก์ชั่นเสียโอกาสกำหนดไว้เป็น

$$L(d,w) = (d-w)^2$$

เพราะฉะนั้นถ้าเราพิจารณาตัวอย่างสุ่ม $X_1, X_2, .., X_n$ ที่มีขนาด n โดยที่ $X_i = x_i$ $(i=1,2,..,n)$ แล้วเราจะได้ว่า การเสี่ยงแบบเบยส์ $B(d^*)$ ของค่าประมาณแบบเบยส์กำหนดไว้ด้วยสมการ

$$\begin{aligned} B(d^*) &= E\{B(d^*/x_1, x_2, .., x_n)\} \\ &= E\{V(d^*/x_1, x_2, .., x_n)\} \end{aligned}$$

เราสามารถคำนวณการเสี่ยงทั้งหมดของค่าประมาณแบบเบยส์ d^* ของ w หลังจากพิจารณาตัวอย่างสุ่ม $X_1, X_2, .., X_n$ ซึ่งเราจะแทนด้วย $T(n)$ และกำหนดไว้ว่า

$$\begin{aligned} T(n) &= B(d^*) + c(n) \\ &= E\{V(w/x_1, x_2, .., x_n)\} + K + kn \end{aligned}$$

จาก $T(n)$ เราสามารถพิจารณาขนาดตัวอย่างที่เหมาะสม คือ n_0 ได้โดยวิธีการของแคลคูลัส ลองพิจารณาตัวอย่างที่สองในตอน 2 ที่แล้วมานั้น ถ้าตัวอย่างสุ่ม $X_1, X_2, .., X_n$ ขนาด n แล้วเราจะได้ว่า สำหรับ $X_i = x_i (i=1,2,..,n)$ นั้น ค่าประมาณแบบเบยส์ d^* ของ w และค่าเสียโอกาสคาดหวังของมัน $B(d^*/x_1, x_2, .., x_n)$ กำหนดไว้ด้วยสมการ

$$d^* = \frac{\nu + \Sigma x_i}{\alpha + n}, \quad B(d^*/x_1, x_2, .., x_n) = \frac{\nu + \Sigma x_i}{(\alpha+n)^2}$$

ดังนั้นการเสี่ยงแบบเบยส์ของ d^* หรือ $B(d^*)$ คือ

$$\begin{aligned} B(d^*) &= E\{B(d^*/x_1, x_2, .., x_n)\} \\ &= E\left\{\frac{\nu + \Sigma x_i}{(\alpha+n)^2}\right\} \\ &= \frac{1}{(\alpha+n)^2}\left\{\nu + \sum E(x_i)\right\} \end{aligned}$$

แต่ $E(X_i) = E\{E(X_i/w)\}$ และโดยที่ X_i เป็นแบบปัวซอง เราจึงได้ $E(X_i/w) = w$ $(i=1,2,..,n)$ เพราะฉะนั้น

$$E(X_i) = E(w) = \nu/\alpha$$

โดยที่ w มีการแจกแจงก่อนทดลองแบบแกมมะ เราจึงได้

$$B(d^*) = \frac{1}{(\alpha+n)^2}\{\nu + \frac{n\nu}{\alpha}\}$$
$$= \frac{\nu}{\alpha(\alpha+n)}$$

สมมติว่าฟังก์ชันค่าใช้จ่ายในการสุ่มตัวอย่างเป็น $c(n) = K + kn$ ตามที่กล่าวมาแล้ว เราก็สามารถหาฟังก์ชันการเสี่ยงทั้งหมด $T(n)$ ได้เป็น

$$T(n) = B(d^*) + c(n)$$
$$= \frac{\nu}{\alpha(\alpha+n)} + [K + kn]$$

เราสามารถหาค่า n ที่ทำให้ $T(n)$ มีค่าต่ำสุดได้ โดยวิธีการแคลคูลัส นั่นคือจะได้สมการ

$$\frac{\nu}{\alpha(\alpha+n_0)^2} = k \quad \text{หรือ} \quad (\alpha+n_0)^2 = \frac{\nu}{k\alpha}$$

$$n_0 = (\nu/k\alpha)^{1/2} - \alpha$$

สำหรับ n_0 นั้นต้องไม่เป็นลบ แต่ถ้าค่า n_0 ที่ได้จากสมการนี้เป็นลบก็ให้ถือว่า $n_0 = 0$ ซึ่งในกรณีเช่นนี้ก็ไม่จำเป็นต้องสุ่มตัวอย่างขนาดใด ๆ เลย ในทำนองเดียวกัน ค่าของ n_0 นั้นจะต้องเป็นจำนวนเต็ม ซึ่งเราพิจารณาเลขจำนวนเต็มที่ใกล้ที่สุด แต่น้อยกว่าหรือเท่ากับ n_0 เมื่อได้ขนาดตัวอย่างที่เหมาะสมแล้ว จึงทำการทดลอง หรือสุ่มตัวอย่างตามขนาดที่หาได้นั้น สำหรับปัญหาที่กล่าวมานั้น ขนาดตัวอย่างที่เหมาะสมจะเป็น

$$n_0 = (\nu/k\alpha)^{1/2} - \alpha$$

4. <u>วิธีประมาณค่าแบบน่าจะเป็นสูงสุดทั่วไป</u> (Generalized Maximum Likelihood Estimation Method)

เท่าที่พิจารณาปัญหาการประมาณค่าที่ผ่านมานั้น เราสมมติว่าฟังก์ชันเสียโอกาส $L(d,w)$ นั้นมี แต่บางปัญหาไม่มีฟังก์ชันเหล่านี้ สำหรับปัญหาที่ไม่มีฟังก์ชันเสียโอกาส ค่า

ประมาณของ w สามารถหาได้ด้วยวิธีประมาณค่าที่เรียกว่า การประมาณแบบน่าจะเป็นสูงสุดทั่วไป วิธีนี้เป็นการเลือกค่าของสภาวการณ์นอกบังคับ w ที่ทำให้ฟังก์ชันน่าจะเป็นที่ปรับปรุง $p(w/x_1, x_2, .., x_n)$ นั้นมีค่าสูงสุด ให้ $w(x)$ เป็นค่าหลังจากสังเกต $X = x$ ดังนั้น ถ้า $w(x)$ มีได้สำหรับทุกค่า $w \in W$ ที่ $p(w(x)/x) \geqslant p(w/x)$ แล้ว $w(x)$ จะเรียกว่า ค่าประมาณแบบน่าจะเป็นสูงสุดทั่วไป และ $w(X)$ เรียกว่า ตัวประมาณค่าแบบน่าจะเป็นสูงสุดทั่วไปของ w ลองพิจารณาตัวอย่างที่แสดงถึงวิธีประมาณค่าแบบน่าจะเป็นสูงสุดทั่วไป

<u>ตัวอย่าง</u> สมมติว่าฟังก์ชันหนาแน่นน่าจะเป็นก่อนทดลองแบบบีทะ สำหรับ w มีพารามิเตอร์ α และ β และฟังก์ชันน่าจะเป็นทวินามแบบจุด หลังจากสังเกตค่าของตัวอย่าง $X_i = x_i (i=1, 2, .., n)$ แล้วเราจะได้ฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุง $p(w/x_1, x_2, .., x_n)$ ดังนี้

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

ในเมื่อ $\Sigma x_i = r$

ค่า $w(X_1, X_2, .., X_n)$ เมื่อ $X_i = x_i \ (i=1,2,..,n)$ ที่สอดคล้องกับ

$$p(w(x)/x) \geqslant p(w/x)$$

นั้น ได้จากการหาอนุพันธ์ $p(w/x_1, x_2, .., x_n)$ หรือ $\ln p(w/x_1, x_2, .., x_n)$ โดยการเทียบกับ w และให้อนุพันธ์นั้นเท่ากับศูนย์ แล้วเราจะได้

$$\frac{d}{dw}[\ln p(w/x_1, x_2, .., x_n)] = \frac{\alpha+r-1}{w} - \frac{\beta+n-r-1}{1-w}$$

และ $$w = \frac{\alpha+r-1}{\alpha+\beta+n-2}$$

ค่าของ w นี้จะทำให้ฟังก์ชันหนาแน่นน่าจะเป็นที่ปรับปรุงสูงสุด เพราะฉะนั้นเราจึงใช้ค่า $w(x_1, x_2, .., x_n)$ นี้เป็นค่าประมาณแบบน่าจะเป็นสูงสุดทั่วไปของ w