บทที่ 2
การประมาณค่าพารามิเตอร์ด้วยค่าคงที่
(Point Estimate)

2.1 บทนำ

ประโยชน์ของวิชาสถิติที่เป็นส่วนสำคัญก็คือ ใช้ในการประมาณค่าคุณลักษณะของประชากรที่สำคัญ เช่น บริษัทผู้ผลิตยาสีฟันแห่งหนึ่งได้โฆษณาว่า ยาสีฟันของบริษัทสามารถลดอัตราของการเป็นโรคฟันผุได้ถึงร้อยละ 30 นั่นหมายความว่าในประชากรที่ใช้ยาสีฟันชนิดนี้จะลดอาการของโรคฟันผุถึง 30 เปอร์เซ็นต์ หรือถ้ากรมอุตุนิยมวิทยาประกาศว่าอุณหภูมิของเดือนเมษายนปี 2521 จะอยู่ระหว่าง 35 - 45 องศาฟาเรนไฮด์ แม้กระทั่งในการสร้างเครื่องบินนักวิศวกรจะต้องประมาณน้ำหนักเฉลี่ยของผู้โดยสารที่จะบรรทุกได้

ในชีวิตประจำวันเราอาจจะต้องประมาณปริมาณไฟฟ้าที่ใช้ในแต่ละวัน เพื่อที่จะคิดค่าใช้จ่ายที่จะต้องเสียในสิ้นเดือน จากบรรดาปัญหาที่ยกมากล่าวนี้ล้วนแต่จะต้องการศึกษาคุณภาพ (คุณลักษณะ) ของประชากรทั้งสิ้น ไม่ว่าจะเป็นทางด้านการแพทย์ วิศวกร ตลอดจนถึงในชีวิตประจำวัน

สิ่งที่จะตามมาในการศึกษาถึงคุณลักษณะของประชากรในด้านต่าง ๆ เช่น ร้อยละ สัดส่วน ค่าเฉลี่ย เหล่านี้ ไม่ว่าจะศึกษาในด้านการประมาณคุณลักษณะที่กล่าวถึงด้วยตัวคงที่ หรือการประมาณด้วยช่วงเชื่อมั่น ล้วนแต่ต้องการค่าประมาณที่ดีในแง่ของความถูกต้องใกล้เคียงกับค่าจริง ๆ ของประชากรทั้งสิ้น หลักการที่จะตัดสินใจว่าตัวประมาณใดเป็นตัวประมาณที่ดี และวิธีการในการหาค่าตัวประมาณต่าง ๆ จึงเป็นสิ่งที่สำคัญอย่างยิ่งในทางสถิติที่จะต้องศึกษาให้ถ่องแท้

ในการประมาณค่าพารามิเตอร์หรือคุณลักษณะที่เราสนใจของประชากรนั้น แยกออกเป็นสองส่วนคือ

1. การประมาณค่าพารามิเตอร์ด้วยตัวคงที่ (Point Estimation)

2. การประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่น (Interval Estimation)

1. การประมาณค่าพารามิเตอร์ด้วยค่าคงที่ หมายถึงการประมาณค่าคุณลักษณะของประชากรที่เราสนใจด้วยตัวคงที่ (ค่าที่ได้นี้สร้างได้จากค่าของข้อมูลที่เราทำการสังเกตุได้) เช่นการประมาณค่าระดับสติปัญญาเฉลี่ยของเด็กในระดับชั้นมัธยมศึกษาว่ามีค่าเท่ากับ 115 โดยการประมาณค่าเฉลี่ยจากตัวอย่างนักเรียนมัธยม 100 คน หรือประมาณค่าจำนวนหลอดไฟฟ้าที่ผลิตจากโรงเรียนแห่งหนึ่งว่ามีเปอร์เซ็นต์ของหลอดไฟฟ้าที่ไม่ได้มาตรฐานเท่ากับ 0.5 โดยการสุ่มตัวอย่างหลอดไฟฟ้ามา 1,000 หลอด แล้วนำมาทดสอบ เป็นต้น

ตัวเลขเหล่านี้ เช่น สติปัญญาเฉลี่ยของเด็กนักเรียนมีค่าเท่ากับ 115 หรือเปอร์เซ็นต์ของหลอดไฟที่เสียเท่ากับ 0.5 เหล่านี้ล้วนแต่เป็นค่าคงที่ที่เราจะนำไปเป็นค่าประมาณของค่าพารามิเตอร์ของประชากร (ซึ่งเราไม่ทราบ)

2. การประมาณด้วยช่วงเชื่อมั่น คือการประมาณคุณลักษณะที่เราสนใจในประชากรหนึ่ง ๆ ด้วยช่วงเชื่อมั่นแทนที่จะใช้ค่า ๆ เดียวแบบการประมาณด้วยตัวคงที่ตัวเดียว เช่นในกรณีของระดับสติปัญญาเฉลี่ยของนักเรียนชั้นมัธยมเป็น 115 เราอาจจะประมาณเป็นว่าระดับสติปัญญาของนักเรียนชั้นมัธยมจะเป็น 100-130 ด้วยช่วงเชื่อมั่น 95 เปอร์เซ็นต์ก็ได้

เราสังเกตได้ว่าการใช้การประมาณด้วยตัวคงที่และด้วยช่วงเชื่อมั่นต่างก็มีข้อดีและข้อเสียแตกต่างกันดังนี้คือ

- การประมาณด้วยช่วงเชื่อมั่นลดความผิดพลาดลงได้เพราะมีจำนวนค่ามากกว่าในการทำนายค่าของพารามิเตอร์

- ในด้านของข้อเสียในกรณีที่เราใช้ช่วงเชื่อมั่นกว้างไป ค่าของการประมาณก็จะมีขอบเขตกว้างจนไม่ก่อให้เกิดประโยชน์ในแง่ของการนำไปใช้งาน ดังเช่นในกรณีของการประมาณระดับสติปัญญาของนักเรียนชั้นมัธยมศึกษา ว่าด้วยความเชื่อมั่นถึง 100 เปอร์เซ็นต์ระดับสติปัญญาของนักเรียนชั้นมัธยมจะอยู่ในช่วง 50-200 เป็นต้น

2.2 ลักษณะของตัวประมาณค่าที่ดี (Desirable of Good Estimator)

1. ปราศจากความเอียงเฉ (Unbiasedness)

นั่นหมายความว่า ถ้าใช้ $\hat{\theta}$ เป็นตัวประมาณค่าของ θ (คุณลักษณะของประชากร) แล้ว ถ้า $\hat{\theta}$ เป็นตัวประมาณที่ปราศจากความเอียงเฉต่อ θ จะได้ว่า

ค่าคาดหวังของ θ ก็คือค่าเดียวกันกับ θ นั่นเอง

หรือจะเขียนได้ว่า $E(\hat{\theta}) = \theta$

ตัวอย่างเช่น กำหนดให้ $X_1,...,X_n$ เป็นตัวประมาณค่าที่สังเกตได้ของกลุ่มตัวอย่างขนาด n โดยที่การกระจายของ $X_i \sim N(\mu, \sigma^2)$

กำหนดให้ $\hat{\mu} = \overline{X}$ เป็นตัวประมาณค่าของ μ แล้ว

และ $\hat{\sigma}^2 = S^2$

ถ้าหากทั้ง $\hat{\mu}$ และ $\hat{\sigma}^2$ เป็นค่าประมาณที่ปราศจากความอคติต่อ μ และ σ^2 แล้ว เราจะต้องได้ผลลัพธ์ดังนี้คือ

1. $E(\hat{\mu}) = \mu$
2. $E(S^2) = \sigma^2$

เราสามารถพิสูจน์ว่า $\hat{\mu} = \overline{X}$เป็นตัวประมาณค่าที่ปราศจากอคติต่อ μ ได้ดังนี้คือ ในกรณีที่ใช้ S^2 เป็นตัวประมาณค่าของ σ^2 นั้น ถ้าเราสามารถพิสูจน์ได้ว่า $E(S^2) = \sigma^2$ ก็แสดงว่า S^2 เป็นตัวประมาณค่าที่ปราศจากอคติของ σ^2

$$\because \quad \frac{nS^2}{\sigma} \sim x^2(n-1)$$

$$\therefore \quad E\left(\frac{nS^2}{\sigma^2}\right) = (n-1)$$

$$\frac{n}{\sigma^2} E(S^2) = (n-1)$$

$$E(S^2) = \frac{(n-1)}{n}\sigma^2$$

นั่นหมายความว่า S^2 ไม่ใช่ตัวประมาณค่าที่ปราศจากอคติของ σ^2

ตัวอย่างเช่น ถ้าเราใช้ขนาดตัวอย่าง 10 หน่วย

$$E(S^2) = \frac{(10-1)}{10}\sigma^2$$

$$= \frac{9}{10}\sigma^2$$

แต่เราเพิ่มขนาดตัวอย่างให้มากขึ้น $n \to \infty$ เราจะได้ว่า

$$E(S^2) = \lim_{n\to\infty} \frac{(n-1)}{n}\sigma^2$$

$$= \sigma^2$$

นั่นแสดงว่า เมื่อตัวอย่างมีขนาดใหญ่มาก ๆ แล้วตัว S^2 จะเป็นตัวประมาณค่าที่ปราศจากอคติของ σ^2

ในกรณีที่ตัวประมาณค่า

$$E(\overline{X}) = E\left(\sum_{i=1}^{n} X_i/n\right)$$

$$= \frac{1}{n}\sum_{1}^{n} E(X_i) = \mu$$

$\overline{X}$ เป็นตัวประมาณที่ปราศจากความเอียงเฉของ μ เป็นต้น

ในกรณีที่ตัวประมาณมีความเอียงเฉต่อค่าพารามิเตอร์จะได้ว่าความเอียงเฉที่เกิดขึ้น

ความเอียงเฉ (Bias) $= E(\hat{\theta}) - \theta$

เมื่อ $\hat{\theta}$ เป็นตัวประมาณค่าและ θ เป็นตัวพารามิเตอร์

2.3 2. ความแนบนัย (Consistency)

กำหนดให้ $\hat{\theta}$ เป็นตัวประมาณค่าของ θ

ถ้า $E(\hat{\theta} - \theta)^2 \to 0$ เมื่อ $n \to \infty$

จะกล่าวได้ว่า $\hat{\theta}$ เป็นตัวประมาณที่มีลักษณะแนบนัยต่อ θ (หรือ $\hat{\theta}$ เป็น Consistent Estimator)

อาจจะกล่าวอีกนัยหนึ่งได้ว่า $\lim_{n \to \infty} E(\hat{\theta}) = \theta$ และ

$\lim_{n \to \infty} V(\hat{\theta}) = 0$ แล้ว

$\hat{\theta}$ มีคุณสมบัติเป็น Consistent Estimator

ให้พิจารณาจาก Mean Square Error

$$
\begin{aligned}
\text{MSE of } \hat{\theta} &= E(\hat{\theta} - \theta)^2 \\
&= E\{(\hat{\theta} - E(\hat{\theta}) + (E(\hat{\theta}) - \theta))\}^2 \\
&= E(\hat{\theta} - E(\hat{\theta}))^2 + E(E(\hat{\theta}) - \theta)^2 + 2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)
\end{aligned}
$$

เพราะว่า

$$2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) = 2(E(\hat{\theta}) - \theta)E(\hat{\theta} - E(\hat{\theta}))$$

และเนื่องจาก $(E(\hat{\theta}) - \theta)$ เป็นค่าคงที่ และ $E(\hat{\theta} - E(\hat{\theta})) = 0$

ดังนั้น

$$2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) = 0$$

ดังนั้น เราจะได้ว่า

$$
\begin{aligned}
E(\hat{\theta} - \theta)^2 &= E(\hat{\theta} - E(\hat{\theta}))^2 + (\text{Bias})^2 \\
&= V(\hat{\theta}) + \text{Bias}^2
\end{aligned}
$$

$\hat{\theta}$ จะเป็นค่าประมาณที่แนบนัย ก็ต่อเมื่อค่าของความเอียงเฉและความแปรปรวนมีค่าเข้าใกล้ 0 ในขณะที่ ขนาดของตัวอย่างเข้าใกล้ค่าอนันต์ (ในกรณีที่ประชากรนับถ้วนก็คือ ขนาดของตัวอย่างเข้าใกล้จำนวนประชากรนั่นเอง)

นิยาม Mean - squared - error consistency

กำหนดให้ $Y_1, Y_2, \ldots, Y_n, \ldots$ เป็นชุดของตัวประมาณค่าของ θ โดยที่ $Y_i = u_i(X_1, X_2, \ldots, X_n)$ ตัวประมาณค่าชุดนี้ จะถูกเรียกว่า มีคุณสมบัติเป็น mean - squared - error - consistency ของ θ ก็ต่อเมื่อ (if and only if) $\lim_{n\to\infty} E[Y_i - \theta]^2 = 0$ สำหรับทุกค่าของ θ

ตัวอย่างที่ 1 ให้ $X_1, X_2, \ldots, X_n$ คือค่าสังเกตขนาด n โดยที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยและความแปรปรวนของประชากรเป็น μ, σ^2 ตามลำดับ

จะได้ว่า $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$\text{Bias of } \bar{X} = (E(\bar{X}) - \mu) = 0$$

$$V(\bar{X}) = \sigma^2/n \to 0$$

ดังนั้น $\bar{X}$ เป็นค่าประมาณที่มีความแนบนัยต่อค่าพารามิเตอร์ μ

ตัวอย่างที่ 2 ของ biased estimator เช่น ตัว n order statisties Y_n ที่นำมาจาก $X_1, \ldots, X_n$ โดยที่ X_i มีการแจกแจงแบบ Uniform $(0,\theta)$, $0<\theta<\infty$

$$f(x;\theta) = \frac{1}{\theta}\ ; 0<x<\theta, 0<\theta<\infty$$

$$\begin{aligned} g(y_n;\theta) &= n\,[F(y_n)]^{n-1} f(y_n) \\ &= n\left[\int_0^{y_n} \frac{1}{\theta}\,dx\right]\cdot\frac{1}{\theta} \\ &= \frac{n y_n^{n-1}}{\theta^n}\ ; 0<y_n<\theta\ ; 0<\theta<\infty \end{aligned}$$

เนื่องจาก $0<Y_1 \leq Y_2 \leq \ldots \leq Y_n \leq \theta$

ถ้าเราใช้ Y_n เป็นตัวประมาณค่าของ θ

หมายความว่า $\hat{\theta} = Y_n$

$$E(Y_n) = \int_0^\theta y_n \, g(y_n;\theta) \, dy_n$$

$$= \int_0^\theta y_n \frac{n y_n^{n-1}}{\theta^n} dy_n$$

$$= \frac{n}{\theta^n} \int_0^\theta y_n^n \, dy_n$$

$$= \frac{n}{\theta^n} \frac{(y_n^{n+1})}{n+1} \Big|_0^\theta$$

$$E(Y_n) = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1}$$

$$= \frac{n\theta}{n+1}$$

ดังนั้น bias ของ Y_n $= E(Y_n) - \theta$

$$= \frac{n\theta}{n+1} - \theta$$

$$= \frac{n\theta - (n+1)\theta}{n+1}$$

$$= \frac{n\theta - n\theta - \theta}{n+1}$$

$$= \frac{-\theta}{n+1}$$

ถ้ากำหนดขนาดตัวอย่างเท่ากับ 10

$$\text{bias} = \frac{-\theta}{11} \; ; \; 0<\theta<\infty$$

ดังนั้น จะเห็นได้ว่าค่าของ bias จะมีค่ามาก ถ้าค่าของ θ มีค่ามากขึ้น ถ้า $\theta \to \infty$ ค่าของ bias มีค่ามาก แต่ถ้าอยากให้ค่าของ bias ลดลงมาก ๆ จนเข้าใกล้ 0 เราจะแก้ไขได้โดยการเพิ่มขนาดตัวอย่าง $n \to \infty$ การเพิ่มขนาดตัวอย่างจะทำให้ $E(Y_n) = \lim_{n\to\infty} \frac{n\theta}{n+1} \to \theta$ ดังนั้น ถ้าใช้ Y_n เป็นตัวประมาณค่าที่ดีของ θ ก็ต่อเมื่อขนาดตัวอย่างมีค่ามาก

ตัวอย่างที่ 3 สุ่มตัวอย่าง $X_1, X_2, \ldots, X_n$ มาจากประชากรที่มีการแจกแจงแบบปกติ ที่มีพารามิเตอร์ เป็น μ และ σ^2

กำหนดให้ $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ และ $S_n^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$

เป็นตัวประมาณค่าสำหรับ μ และ σ^2 ตามลำดับ

$$E(S_n^2 - \sigma^2)^2 = \frac{1}{n}\left(\mu_4 - \left(\frac{n-3}{n-1}\right)\sigma^4\right)$$

$$\lim_{n\to\infty} E(S_n^2 - \sigma^2)^2 = \lim_{n\to\infty} \frac{1}{n}\left(\mu_4 - \left(\frac{n-3}{n-1}\right)\sigma^4\right)$$

$$= 0$$

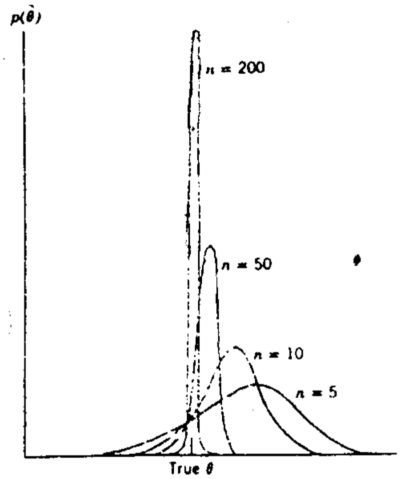
ดังนั้น S_n^2 เป็น mean-squared-errar consistency ของ σ^2

ในทำนองเดียวกัน เราก็สามารถพิสูจน์ได้ว่า $S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n}$ ก็มีคุณสมบัติเป็น mean-squared-errar consistent estimator ของ σ^2 ด้วยเช่นกัน

คุณสมบัติที่ว่าการเป็นค่าประมาณที่แนบนัยต่อพารามิเตอร์ไม่ได้ประกันว่าตัวประมาณใดที่มีคุณสมบัติเช่นนี้จะต้องดีเสมอไป ตัวอย่างเช่นในกรณีที่ประชากรมีการกระจายแบบปกติโดยที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2/n เราจะได้ว่าทั้งค่าเฉลี่ย

ของตัวอย่าง $\bar{X}$ และค่ามัธยฐานจากตัวอย่าง md ล้วนมีคุณสมบัติเป็นค่าประมาณที่แนบนัยต่อ μ ทั้งคู่ แต่เรานิยมที่จะเลือก ค่าเฉลี่ยจากตัวอย่าง $(\bar{X})$ เป็นตัวประมาณค่าของ μ มากกว่า ทั้งนี้เพราะว่านอกจาก $\bar{X}$ จะมีคุณสมบัติเป็นค่าประมาณที่แนบนัยแล้ว ยังมีคุณสมบัติเป็นตัวประมาณที่มีประสิทธิภาพดีที่สุดอีกด้วย (Efficient Estimator)

รูปต่อไปนี้จะแสดงถึงตัวประมาณค่าที่มีคุณสมบัติเป็น consistency โดยจะแสดงถึงฟังก์ชั่นการแจกแจงของตัวประมาณค่า (ในรูปคือ $\hat{\theta}$) ซึ่งจะแจกแจงกันหนาแน่นใกล้ ๆ พารามิเตอร์ θ เมื่อ n มีค่าเพิ่มขึ้น



เราอาจจะพิสูจน์คุณสมบัติด้าน consistency ของตัวประมาณค่าได้อีกวิธีดังนี้ คือ

ถ้ากำหนดให้ $\hat{\theta}$ เป็นตัวประมาณค่าของพารามิเตอร์ θ แล้วเราสามารถหา δ ใด ๆ ซึ่ง δ มีคุณสมบัติเป็นค่าเล็ก ๆ ที่มีมากกว่า 0 โดยที่

$$\lim_{n\to\infty} \Pr\{|\hat{\theta}-\theta| < \delta\} \to 1$$

แล้วจะได้ว่า $\hat{\theta}$ มีคุณสมบัติ consistency

ข้อสังเกต จะเห็นว่าคุณลักษณะนี้ตรงกับเรื่องของ Stochastic Convergence นั่นเอง

ตัวอย่าง การพิสูจน์ว่า $\bar{X}$ เป็น Consistency estimator ของ $E(\bar{X})$ โดยใช้วิธีการนี้ในการพิสูจน์

$$\bar{X} = \sum_{1}^{n} \frac{X}{n}$$

$$\therefore \Pr\{|\hat{\theta} - \theta| < \delta\} = \Pr\{|\bar{X} - E(\bar{X})| < \delta|\}$$

โดยคุณสมบัติของอสมการของเชบบิเชฟ จะได้ว่า

$$\Pr\{|\bar{X} - E(\bar{X})|<\delta\} \geqslant 1 - \frac{V(\bar{X})}{\delta^2}$$

$$\geqslant 1 - \frac{\sigma^2}{n\delta^2}$$

$$\lim_{n\to\infty} \Pr\{|\bar{X} - E(\bar{X})|<\delta\} \geqslant 1 - \lim_{n\to\infty} \frac{\sigma^2}{n\delta^2}$$

$$\geqslant 1$$

ดังนั้น $\lim_{n\to\infty} \Pr\{|\bar{X} - E(\bar{X})|<\delta\} = 1$

นั่นก็คือ $\bar{X}$ เป็น Consistency estimator ของ $E(\bar{X})$

ตัวอย่างที่ 4 กำหนดให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่า 2 ตัวของพารามิเตอร์ θ โดยที่ $E(\hat{\theta}_1) = \theta$ และ $E(\hat{\theta}_2) = .9\theta$ $V(\hat{\theta}_1) = 3$ และ $V(\hat{\theta}_2) = 2$ เราจะเลือกตัวประมาณค่าตัวใดดี

$$MSE(\hat{\theta}_1) = V(\hat{\theta}_1) + (\text{Bias})^2$$

$$= 3 + 0$$

$$MSE(\hat{\theta}_2) = V(\hat{\theta}_2) + (\text{Bias})^2$$

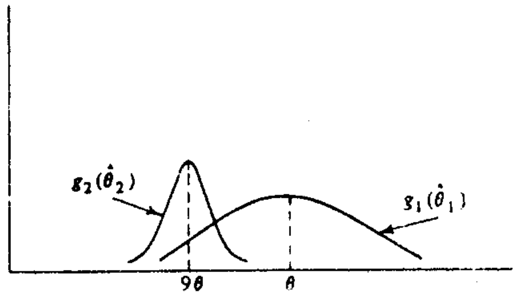
$$= 2 + .01\theta^2$$

พิจารณา $\hat{\theta}_1$ และ $\hat{\theta}_2$

จะเลือก $\hat{\theta}_1$ ถ้า $|\theta|>10$

และ $\hat{\theta}_2$ ถ้า $|\theta|<10$

และจะเลือก $\hat{\theta}_1$ หรือ $\hat{\theta}_2$ ตัวใดก็ได้ ถ้า $|\theta| = 10$



Densities of two estimators of the parameter θ.

3. มีค่าความแปรปรวน (Variance) ต่ำ ๆ

กรณีทั่ว ๆ ไปเมื่อพบว่า มีตัวประมาณค่า ที่ปราศจากอคติต่อพารามิเตอร์ เราก็มักจะเลือกตัวประมาณค่านั้น ๆ ไปใช้เลย แต่ก็มีปัญหาที่เกิดขึ้นมาพบว่า ในบางครั้งเราอาจจะได้ตัวประมาณค่าที่มีคุณสมบัติเป็นตัวประมาณที่ปราศจากอคติต่อพารามิเตอร์หลาย ๆ ตัว หรือไม่ก็ไม่สามารถที่จะหาตัวประมาณค่า ที่ปราศจากอคติต่อพารามิเตอร์ได้เลย

ดังนั้นการจะเลือกตัวประมาณค่าที่ดี เมื่อประสบปัญหาข้างต้น จึงแยกเป็น 2 ประเด็น คือ

ประเด็นที่ 1 กำหนดให้ $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ เป็นชุดของตัวประมาณค่าที่ปราศจากอคติต่อพารามิเตอร์ θ ดังนั้นการที่จะยึดหลักว่า จะเลือกตัวประมาณค่าตัวใดก็ให้ยึดหลักว่า

$$V(\hat{\theta}_i) \leqslant V(\hat{\theta}_j) \; ; \; j = 1,2,\ldots,k \quad i \neq j$$

ในกรณีที่เรามีตัวประมาณค่า 2 ตัว ซึ่งทั้งคู่ต่างก็เป็นตัวประมาณค่าปราศจากอคติต่อพารามิเตอร์ เราอาจเปรียบเทียบประสิทธิภาพของตัวประมาณค่าทั้งสองได้ดังนี้คือ

$$\text{Relative efficiency of } \hat{\theta}_i : \hat{\theta}_j = \frac{V(\hat{\theta}_i)}{V(\hat{\theta}_j)} \times 100$$

4. มีคุณสมบัติเป็นตัว sufficient statistic คุณสมบัติข้อ 4 นี้ เป็นคุณสมบัติเพิ่มเติมนอกเหนือจากคุณสมบัติที่กล่าวมาแล้ว 3 ประการ เรื่อง sufficient statistic นี้จะได้กล่าวโดยละเอียดอีกครั้งในเรื่องของ sufficient statistics

2.4 Sufficient Statistics

ในการทดลองสุ่มตัวแปร X มาโดยใช้ขนาดของตัวอย่าง n โดยที่การกระจายของตัวแปร X มี $f(x;\theta)$ ดังนั้นตัวอย่างที่เราได้ ซึ่งประกอบด้วย $X_1,X_2,\ldots,X_n$ จะเป็นชุดหนึ่งของตัวแปรที่อธิบายคุณลักษณะ (Information) ของตัวพารามิเตอร์ ดังนั้นการศึกษาลักษณะของพารามิเตอร์โดยอาศัยขนาดตัวอย่างนี้ จะต้องศึกษาตัวแปรแต่ละตัวซึ่งครอบคลุมพารามิเตอร์ θ อยู่ เพื่อเป็นการลดความยุ่งยากในการที่จะศึกษาจาก $X_1,X_2,\ldots,X_n$ ทั้งชุด เราจะแปลงจากตัวแปรขนาด n มาสู่ตัวแปรใหม่ โดยที่ $\hat{\theta} = d(x_1,x_2,\ldots,x_n)$

ในที่นี้ $\hat{\theta}$ จะเป็นตัวแปรตัวใหม่ที่ได้จากฟังก์ชั่นของตัวแปร X ชุดเดิม ดังนั้นคุณลักษณะของพารามิเตอร์ที่ถูกเก็บไว้ในตัวแปร $X_1,X_2,\ldots,X_n$ จะมารวมอยู่ใน $\hat{\theta}$

เราจะเรียก $\hat{\theta}$ ที่เกิดมาโดยลักษณะนี้ (เพื่อที่จะนำไปอธิบายถึงคุณลักษณะของตัวพารามิเตอร์) ว่าตัวสถิติที่อธิบายคุณลักษณะของพารามิเตอร์ได้อย่างเพียงพอ (Sufficient Statistics)

จุดประสงค์ในการสร้างตัว Sufficient Statistics ก็เพื่อที่จะนำไปสู่การสร้างให้เป็นตัวประมาณค่าที่ดีแก่ตัวพารามิเตอร์ ความหมายของการเป็นตัว Sufficient Statistics ก็คือตัวสถิติที่มีคุณสมบัติเช่นนี้จะเป็นตัวที่รวบรวมคุณลักษณะของของพารามิเตอร์ไว้ในตัวของมันอยู่แล้ว (กล่าวง่าย ๆ ก็คือ ตัว Sufficient Statistic จะมี Information เพียงพอที่จะอธิบายถึงตัวพารามิเตอร์)

การศึกษาเรื่อง Sufficient Statistic จะแยกออกเป็น 2 กรณีคือ

- กรณีของพารามิเตอร์ตัวเดียว
- กรณีของหลาย ๆ พารามิเตอร์

นิยาม กำหนดให้ $X_1,X_2,\ldots,X_n$ เป็นตัวแปรเชิงสุ่มชุดหนึ่งที่มีฟังก์ชั่นความน่าจะเป็นคือ $f(x;\theta)$ และให้ $\hat{\theta} = d(x_1,\ldots,x_n)$ (เป็นฟังก์ชั่นของ X_i เท่านั้น) เป็นตัวสถิติที่สร้างขึ้น

$\theta^* = d^*(x_1,x_2,\ldots,x_n)$ เป็นตัวสถิติตัวอื่นที่ ไม่ใช่ฟังก์ชั่นของ θ

ถ้าหากว่าเราสามารถหาได้ว่าสำหรับตัวสถิติ θ^* แล้ว ฟังก์ชั่นความน่าจะเป็นของ θ^* ภายใต้เงื่อนไขว่ากำหนด $\hat{\theta}$ แล้ว จะต้องเป็นฟังก์ชั่นที่ที่ไม่เกี่ยวข้องกับ θ แล้ว จะสรุปได้ว่า $\hat{\theta}$ จะเป็น Sufficient Statistics สำหรับพารามิเตอร์ θ

นั่นคือ ถ้า $p(\theta^*/\hat{\theta})$ ไม่เกี่ยวข้องกับ θ แล้ว

$\hat{\theta}$ เป็นตัว Sufficient Statistic สำหรับ θ

สรุป ถ้าเราสร้างเซทของตัวสถิติ (ตัวแปร) ขึ้นมาชุดหนึ่งประกอบด้วย

$\tilde{\theta}\quad \{\hat{\theta},\theta;\ldots,\theta^*\ldots,\}$

โดยที่
$$
\begin{aligned}
\hat{\theta} &= d(x_1,\ldots,x_n)\\
\theta' &= d'(x_1,\ldots,x_n)\\
&\vdots\\
\theta^* &= d^*(x_1,\ldots,x_n)\\
&\vdots
\end{aligned}
$$

และ $X_1\ X_2,\ldots,X_n$ เป็นตัวแปรที่เราสุ่มมา โดยที่มี $f(x;\theta);\ \theta \varepsilon \Omega$

และถ้าเราได้ว่า $p(\hat{\theta}^*,\hat{\theta}',\ldots/\hat{\theta})$ ไม่เกี่ยวข้องกับ θ

จะได้ว่า $\hat{\theta}$ เป็นตัวสถิติที่ใช้อธิบายพารามิเตอร์ได้อย่างเพียงพอ (Sufficient Statistic) สำหรับทฤษฎีที่ใช้ในการที่จะพิสูจน์ว่า ตัวสถิติใดเป็น Sufficient Statistic หรือไม่นั้น เราแยกออกเป็นสองทฤษฎีคือ

ทฤษฎีที่ 1 ทฤษฎีของฟิชเชอร์เนย์แมน (Fisher-Neyman Criterion)

กำหนดตัวแปร $X_1,\ldots,X_n$ สุ่มมาจากตัวแปรที่มี $f(x;\theta)$; $\theta\varepsilon\Omega$ และ $a<x<b$ (ทั้ง a และ b ไม่เกี่ยวข้องกับ θ) กำหนดให้การกระจายร่วมของตัวแปร $X_1,\ldots,X_n$ อยู่ในรูป $g(x_1,\ldots,x_n;\theta)$

$$g(x_1,\ldots,x_n;\theta) = f(x_1;\theta)\ldots f(x_n;\theta)$$

ถ้าหากว่าเราสามารถแยก

$$g(x_1,\ldots,x_n;\theta) = h(\hat{\theta}_1;\theta)k(x_1,\ldots,x_n)$$

โดยที่ $k(x_1,\ldots,x_n)$ เป็นฟังก์ชั่นที่ไม่เกี่ยวข้องกับ θ และ $\hat{\theta}_1 = d_1(x_1,\ldots,x_n)$ จะสรุปว่า $\hat{\theta}_1$ คือตัว sufficient Statistic สำหรับ θ

และ $h(\hat{\theta}_1;\theta)$ คือ marginal pdf. of $\hat{\theta}_1$

พิสูจน์ กำหนดให้ $\hat{\theta}_1$ คือ ฟังก์ชั่นของตัวแปร $X_1,\ldots,X_n$

$$\begin{array}{lcl}
\hat{\theta}_1 = d_1(x_1,\ldots,x_n) & \rightarrow & x_1 = u_1(\hat{\theta}_1,\ldots,\hat{\theta}_n) \\
\hat{\theta}_2 = d_2(x_1,\ldots,x_n) & \rightarrow & x_2 = u_2(\hat{\theta}_1,\ldots,\hat{\theta}_n) \\
\vdots & & \vdots \\
\hat{\theta}_n = d_n(x_1,\ldots,x_n) & \rightarrow & x_n = u_n(\hat{\theta}_1,\ldots,\hat{\theta}_n)
\end{array}$$

การกำหนด one to one transformation จาก $\hat{\theta}_i$ ไปยัง X_i เพื่อที่จะให้ง่ายแก่การพิสูจน์

$$f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)$$

$$= f(u_1(\hat{\theta}_1,\ldots,\hat{\theta}_n);\theta)\ldots f(u_n(\hat{\theta}_1,\ldots,\hat{\theta}_1);\theta)$$

$$= g(\hat{\theta}_1,\ldots,\hat{\theta}_n;\theta)$$

$$g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta) = \frac{g(\hat{\theta}_1,\ldots,\hat{\theta}_n;\theta)}{h(\hat{\theta}_1;\theta)}$$

$$g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta) = \frac{f(u_1(\hat{\theta}_1,\ldots,\hat{\theta}_n);\theta)\ldots f(u_n(\hat{\theta}_1,\ldots,\hat{\theta}_n);\theta)}{h(\hat{\theta}_1;\theta)}$$

$$f(u_1(\hat{\theta}_1,\ldots,\hat{\theta}_n);\theta)\ldots f(u_n(\hat{\theta}_1,\ldots,\hat{\theta}_n);\theta) = h(\hat{\theta}_1;\theta)g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1),\theta)\cdot\frac{1}{|J|}$$

เทียบไปแล้ว $g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta\cdot\frac{1}{|J|}$ ก็คือ $k(x_1,\ldots,x_n)$ ตามทฤษฎี

พิจารณา $g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta)\cdot\frac{1}{|J|}$ จะเห็นว่า $|J|$ จะไม่เกี่ยวข้องกับ θ

เพราะว่าการแปลงรูปจากตัวแปร $X_1,\ldots,X_n$ มายังตัวแปร $\hat{\theta}_1,\ldots,\hat{\theta}_n$ นั้นไม่ได้นำ θ เข้าไปเกี่ยวข้องด้วย

ดังนั้น $|J|$ จึงไม่ขึ้นอยู่กับ θ

ดังนั้นเมื่อ $k(x_1,\ldots,x_n)$ ไม่ขึ้นอยู่กับ θ

และ $g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta)$ ก็ไม่ขึ้นอยู่กับ θ

$\rightarrow$ $\hat{\theta}_1$ เป็น Sufficient Statistic ตามนิยาม การพิสูจน์ย้อนกลับก็พิสูจน์ในทำนองเดียวกัน

ทฤษฎีที่ 2 Factorization Theorem

กำหนดให้สุ่มตัวแปรชุดหนึ่งขนาด n คือ $X_1,\ldots,X_n$ โดยที่การแจกแจงของตัวแปร X คือ $f(x;\theta)$ $\theta\varepsilon\Omega$ $a<x<b$ โดยที่ a และ b ไม่เป็นฟังก์ชั่นของ θ ถ้าเราสร้างฟังก์ชั่นของความน่าจะเป็นร่วมของตัวแปร $X_1,\ldots,X_n$ ได้เป็น $g(x_1,\ldots,x_n;\theta) = f(x_1;\theta)\ldots f(x_n;\theta)$

ถ้าหากว่าเราสามารถจัด $g(x_1,\ldots,x_n;\theta) = h(\hat{\theta};\theta).k(x_1,\ldots,x_n)$

โดยที่ $\hat{\theta} = u(x_1,\ldots,x_n)$ และ $k(x_1,\ldots,x_n)$ เป็นฟังก์ชั่นที่ไม่เกี่ยวข้องกับ θ เราจะสรุปได้ว่า $\hat{\theta}$ เป็นตัวสถิติที่อธิบายพารามิเตอร์ (Sufficient Statistic) ได้อย่างเพียงพอ

พิสูจน์ หลักการพิสูจน์คล้ายคลึงกับทฤษฎีที่ 1 โดยที่ $k(x_1,\ldots,x_n)$ มีหน้าที่เหมือนกับ $g(\hat{\theta}_2,\ldots,\hat{\theta}_n;\hat{\theta}_1,\theta)$ (โดยที่ $\hat{\theta}$ คือ $\hat{\theta}_1$)

ข้อแตกต่างระหว่างทฤษฎีที่ 1 และทฤษฎีที่ 2 ก็คือ $h(\hat{\theta};\theta)$ ในทฤษฎีที่ 1 จะต้องเป็น marginal pdf of $\hat{\theta}$, แต่ $h(\hat{\theta};\theta)$ ในทฤษฎีที่ 2 ไม่ได้บ่งว่า $h(\hat{\theta};\theta)$ จะต้องเป็น marginal pdf ของ $\hat{\theta}$ ดังนั้นการที่จะใช้ ทฤษฎีที่ 1 หรือทฤษฎีที่ 2 ก็อาจจะพิจารณาง่าย ๆ ในขั้นแรกว่า ถ้าเราสามารถหา pdf ของตัวสถิติที่จะพิสูจน์ว่ามีคุณสมบัติเป็น sufficiency เราก็ใช้ทฤษฎีที่ 1 แต่ถ้าการหา pdf. ของตัวสถิติตัวนั้นค่อนข้างจะยุ่งยาก เช่นอยู่ในรูปของผลคูณคือ $\prod_{i=1}^{n} X_i$ หรือ $X_1X_2+X_3$ ดังนี้ ก็ให้ใช้ทฤษฎีที่ 2 เข้าช่วย

ข้อสังเกต ทฤษฎีที่ 1 และที่ 2 มีความแตกต่างในการใช้พอที่จะชี้ให้เห็นได้ก็คือ ในกรณีที่ปัญหากำหนดว่าให้พิสูจน์ว่าตัวสถิติตัวใดเป็นตัวสถิติที่เพียงพอ (Sufficient Statistic) ในการอธิบายพารามิเตอร์และ Marginal Density Function ของตัวสถิตินั้นพอที่จะหาได้ ก็ให้ใช้ทฤษฎีของพิชเชอร์เนยแมน แต่ถ้าในกรณีที่โจทกย์ไม่บอกว่าตัวสถิติที่ให้หาว่าเป็นตัว Sufficient Statistic ของพารามิเตอร์เป็นตัวอะไร ก็ให้ใช้ทฤษฎี Factorization Theorem แต่แนวทางที่บอกไว้นี้มิใช่กฎเกณฑ์ที่ตายตัวเสมอไปในบางกรณีก็ไม่เป็นไปตามแนวทางนี้ โดยยึดถือตามความสะดวกและง่ายแก่การพิสูจน์เป็นสำคัญ

ตัวอย่างที่ 1 กำหนดให้ X_1,X_2,X_3 เป็นตัวอย่างที่สุ่มมาจากการแจกแจงที่เป็น Bernoulli

$$S = u_1(X_1,X_2,X_3) = X_1+X_2+X_3 \qquad T = u_2(X_1,X_2,X_3) = X_1X_2+X_3$$

จงแสดงว่า S เป็นตัว Sufficient Statistic แต่ T ไม่ใช่ตัว Sufficient Statistics

	Values of S	Value of T	$f(x_1,x_2,x_3/S)$	$f(x_1,x_2,x_3/T)$
(0,0,0)	0	0	1	$(1-p)/(1+p)$
(0,0,1)	1	1	1/3	$(1-p)/(1+2p)$
(0,1,0)	1	0	1/3	$p/(1+p)$
(1,0,0)	1	0	1/3	$p/(1+p)$
(0,1,1)	2	1	1/3	$p/(1+2p)$
(1,0,1)	2	1	1/3	$p/(1+2p)$
(1,1,0)	2	1	1/3	$p/(1+2p)$
(1,1,1)	3	2	1	1

พิจารณาค่าที่ได้ในแต่ละสดมภ์

$$f(x;\theta) = p^{x_i}(1-p)^{1-x_i} \; ; \; x_i = 0,1$$

$$f(x_1,x_2,x_3) = f(x_1;\, p)\, f(x_2;\, p)\, f(x_3;\, p)$$

$$= p^{\Sigma x_i}(1-p)^{3-\Sigma x_i}$$

$$f(x_1,x_2,x_3/S) = f(0,1,0/1)$$

$$= \frac{Pr(X_1=0,X_2=1,X_3=0/S=1)}{Pr(S=1)}$$

$$= \frac{(1-p)p(1-p)}{\binom{3}{1}p(1-p)^2} = 1/3$$

$$\mathbf{f(x_1,x_2,x_3/T) = f(0,1,0/0)}$$

$$= \frac{Pr(X_1=0,X_2=1,X_3=0/T=0)}{Pr(T=0)}$$

$$= (1-P)\,P\,(1-P)$$

$$= \frac{(1-p)^2 p}{(1-P)^3 +}$$

$$= \frac{-P}{14P}$$

พิจารณา $f(x_1,x_2,x_3/S)$ จะไม่เกี่ยวข้องกับ p แต่ $f(x_1,x_2,x_3/T)$ จะติดค่าของ p

ดังนั้นจึงสรุปได้ว่า S จึงเป็น Sufficient Statistic ของพารามิเตอร์ p ในขณะที่ T ไม่มีคุณสมบัติที่ว่านี้

การพิสูจน์ดังกล่าวในตัวอย่างเป็นการพิสูจน์ตามคำจำกัดความของตัวสถิติที่มีคุณสมบัตินี้ เราอาจจะใช้ทฤษฎีที่ 1 เข้าช่วยได้เพื่อให้ง่ายเข้าดังนี้

$$g(x_1,x_2,x_3;p) = f(x_1;p)f(x_2;p)f(x_3;p)$$

$$= p^{\Sigma x_i}(1-p)^{3-\Sigma x_i}.$$

$$S = X_1+X_2+X_3 = \Sigma X_1 \sim B(3,p)$$

$$g(s;p) = p^s(1-p)^{3-s} \quad ; \quad s = 0,1,2,3$$

$$k(x_1,x_2,x_3) = 1/\binom{3}{1}$$

$$p^{\Sigma x_i}(1-p)^{3-\Sigma x_i} = p^s(1-p)^{3-s} \cdot \frac{1}{\binom{3}{1}}$$

เนื่องจาก $S = \Sigma X_i$

$$\Longrightarrow p^{\Sigma x_i}(1-p)^{3-\Sigma x_i} = p^s(1-p)^{3-s} \cdot \frac{1}{\binom{3}{1}}$$

ดังนั้น S จึงเป็น Sufficient Statistic สำหรับพารามิเตอร์ p

ตัวอย่างที่ 2 กำหนดให้ $X_1,\ldots,X_n$ เป็นตัวอย่างที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย และความแปรปรวนเป็น μ, 1 ตามลำดับ จงหาตัว Sufficient Statistic สำหรับพารามิเตอร์ p

$$g(x_1,\ldots,x_n;\mu,1) = n(x_1;\mu,1)\ldots,n(x_n;\mu,1)$$

$$= \frac{1}{\sqrt{2\pi}} \exp-(x_1-\mu)^2\ldots\frac{1}{\sqrt{2\pi}} \exp-(x_n-\mu)^2$$

$$= \frac{1}{(2\pi)^{n/2}} \exp-\sum_{i}^{n}(x_i-\mu)^2 \qquad \ldots\ldots\ldots.(1)$$

$$\Sigma(x_i-\mu)^2 = \Sigma((x_i-\bar{x}) - (\bar{x}-\mu))^2$$

$$= \Sigma(x_i-\bar{x})^2 + \Sigma(\bar{x}-\mu)^2$$

$$= \Sigma(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

แทนที่ผลที่ได้ในสมการที่ (1)

$$g(x_1,\ldots,x_n;\mu) = \frac{1}{(2\pi)^{n/2}} \exp(-n(\bar{x}-\mu)^2/2).\exp(\Sigma(x_i-\bar{x}^2)/2)$$

$\hat{\mu} = \bar{x}$ โดยทฤษฎีของ Factorization

$$g(\hat{\mu},\mu) = \frac{1}{(2\pi)^{n/2}} \exp(-n(\bar{x}-\hat{\mu})^2/2)$$

$$k(x_1,\ldots,x_n) = \exp(-\Sigma(x_i-\mu)^2/2)$$ ซึ่งไม่เกี่ยวข้องกับ μ

ดังนั้น $\hat{\mu} = \bar{X}$ จะเป็น Sufficient Statistic;for μ

ทฤษฎีที่ 3 กำหนดให้ $\hat{\theta} = d(x_1,\ldots,x_n)$ เป็นตัว Sufficient Statistic สำหรับ θ โดยที่เราสุ่มตัวอย่างมาจากประชากรที่มีฟังก์ชั่นของความน่าจะเป็นคือ $f(x;\theta)$ ถ้าหากว่า u ที่สร้างขึ้นจากฟังก์ชั่นของ θ โดยที่มีข้อแม้ว่าต้องเป็น Single Valued Inverse แล้วเราจะได้ว่า

$\tilde{\theta} = u(\hat{\theta})$ ก็จะเป็น Sufficient Statistic สำหรับพารามิเตอร์ θ ด้วยเช่นกัน และ $\hat{\theta}$ ก็จะเป็นตัว Sufficient Statistic สำหรับ $u(\theta)$ ด้วย

พิจารณาจากตัวอย่างที่เราได้ว่า $\overline{X}$ เป็นตัว Sufficient Statistic สำหรับ μ ถ้าเรากำหนดให้ $\tilde{\mu} = \overline{X} + 8$ หรือ $n\overline{X}$ เราก็จะได้ว่า $\tilde{\mu}$ เป็นตัวสถิติที่มีคุณสมบัติเป็น Sufficiency สำหรับค่า μ และยังได้อีกว่า $\overline{X}$ เป็นตัวสถิติที่ Sufficiency สำหรับ 5μ หรือ $3\mu + 1$ ฯลฯ อีกด้วย

ตัวอย่างที่ 3 ให้ $X_1,\ldots,X_n$ เป็นตัวแปรชุดหนึ่ง ที่มาจากประชากรที่มีการแจกแจงเป็นปกติ $N(0,\sigma)^2$

ดังนั้นฟังก์ชั่นการแจกแจงร่วมก็คือ

$$g(x_1,\ldots,x_n) = (1/(2\sigma)^{n/2})\exp - (\Sigma x_i^2/2\sigma^2)$$

กำหนดให้ $h(\hat{\sigma}^2;\sigma^2) = (1/(2\sigma^2)^{n/2}\exp - (\Sigma x_i^2/2\sigma^2)$

และ $k(x_1,\ldots,x_n) = 1$

จากทฤษฎีของ Factorization จะได้ว่า $\hat{\sigma}^2 = \Sigma x_i^2$ เป็นตัว sufficient Statistic สำหรับพารามิเตอร์ σ^2

Sufficient Statistics; More Than One Parameter

จากทฤษฎีและตัวอย่างที่กล่าวมาแล้วนั้นได้กล่าวถึงเฉพาะในกรณีของการแจกแจงที่มีพารามิเตอร์ตัวเดียว หรือในกรณีที่มีสองพารามิเตอร์ของการแจกแจงปกติ เราก็จะกำหนดค่าใดค่าหนึ่งของพารามิเตอร์มาให้ทราบ จึงกลายเป็นกรณีของพารามิเตอร์ตัวเดียวไปเช่นกัน

ถ้าเราขยายขนาดของพารามิเตอร์ให้มากออกไป เช่นกำหนดให้ ตัวแปรชุดหนึ่งคือ $X_1,\ldots,X_n$ ที่มีประชากรมีฟังก์ชั่นของความน่าจะเป็นเป็น $f(x;\theta_1,\ldots,\theta_k)$ จะเห็นว่าจำนวนพารามิเตอร์มีมากขึ้นเป็น $\theta_1,\ldots,\theta_k$ (k พารามิเตอร์) การที่จะหาตัวสถิติที่มีคุณสมบัติเป็น Sufficiency สำหรับ $\theta_1,\ldots,\theta_k$ จะขอกล่าวถึงโดยการใช้เซทแทนว่าเซทของ Sufficient Statistic สำหรับเซทของพารามิเตอร์ $\theta_1,\ldots,\theta_k$ เช่นกำหนดว่า

$$\hat{\theta}_1 = d_1(x_1,\ldots,x_n)$$

$$\hat{\theta}_2 = d_2(x_1,\ldots,x_n)$$

$$\vdots \qquad \vdots$$

$$\hat{\theta}_m = d_m(x_1,\ldots,x_n)$$

$\hat{\theta}_1,\ldots,\hat{\theta}_m$ เป็นตัวสถิติ m ตัว ถ้าเราได้ว่าฟังก์ชั่นการแจกแจงของ $X_1,\ldots,X_n$ ภายใต้เงื่อนไขว่า กำหนดค่า $\hat{\theta}_1,\ldots,\hat{\theta}_m$ แล้ว จะไม่ขึ้นอยู่กับ $\theta_1,\ldots,\theta_k$

จะสรุปได้ว่า $(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ เป็นเซทของ Sufficient Statistics สำหรับพารามิเตอร์ $(\theta_1,\ldots,\theta_k)$

จะสามารถอธิบายโดยใช้สัญลักษณ์ได้คือ $g(x_1,\ldots,x_n/\hat{\theta}_1,\ldots,\hat{\theta}_m,\theta_1,\ldots,\theta_k)$ จะเป็นฟังก์ชั่นที่ไม่เกี่ยวข้องกับจากเซท $(\theta_1,\ldots,\theta_k)$

ในเรื่องที่เกี่ยวข้องกับค่า m นี้จะกล่าวถึงอีกครั้งในตอนที่ว่าด้วยคุณสมบัติของตัวสถิติที่มีคุณสมบัติว่า Sufficiency

สำหรับฟังก์ชั่นที่มีหลาย ๆ พารามิเตอร์นั้น ก็ยังคงสามารถใช้ทฤษฎีทั้งหลายที่เกี่ยวข้องกับเรื่อง Sufficient Statistic ได้ เพียงแต่เราจะต้องขยายขอบเขตของตัวพารามิเตอร์และตัวสถิติเพิ่มขึ้นตามลักษณะของแต่ละปัญหา เช่นทฤษฎีของ Factorization

pdf of X_1; $f(x_1;\theta_1,\ldots,\theta_k)$

$$g(x_1,\ldots,x_n;\theta_1,\ldots,\theta_k) = h(\theta_1,\ldots,\theta_m;\theta_1,\ldots,\theta_k)k(x_1,\ldots,x_n)$$

ถ้าหากว่า $k(x_1,\ldots,x_n)$ เป็นฟังก์ชั่นที่ปราศจากเซทของ $(\theta_1,\ldots,\theta_k)$ แล้วจะได้ว่า $(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ เป็นเซทของ Sufficient Statistics สำหรับเซทของพารามิเตอร์ $(\theta,\ldots,\theta_k)$ เป็นต้น

ตัวอย่างที่ 4 ให้ $X_1,\ldots,X_n$ เป็นตัวแปรสุ่มที่มาจาก ประชากร $N(\mu,\sigma^2)$ โดยที่ $-\infty<\mu<\infty$ $,0<\sigma^2<\infty$ ให้หา Sufficient Statistics สำหรับ μ และ σ^2

$$g(x_1,\ldots,x_n;\mu,\sigma^2) = N(x_1;\mu,\sigma^2)\ldots N(x_n;\mu,\sigma^2)$$

$$= (1/(\sigma\sqrt{2\pi})^n)\exp - \Sigma(x_i-\mu)^2/2\sigma^2)$$

$$= (1/(\sigma\sqrt{2\pi})^n)\exp - (\Sigma(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)/2\sigma^2$$

$$= (1/(\sigma\sqrt{2\pi})^n)\exp\{-(n-1)\hat{\theta}_1/2\sigma^2 + n(\hat{\theta}_2-\mu)^2/2\sigma^2\}.1$$

$$\hat{\theta}_1 = \Sigma(x_i-\bar{x})^2/(n-1)\ ,\ \hat{\theta}_2 = \bar{x}$$

$$g(x_1,\ldots,x_n;\mu,\sigma^2) = (1/(\sigma\sqrt{2\pi})^n\exp\{-(n-1)\hat{\theta}_1/2\sigma^2 + n(\hat{\theta}_2-\mu)^2/2\sigma^2\}$$

$$k(x_1,\ldots,x_n) = 1$$

เนื่องจาก $k(x_1,\ldots,x_n)$ เป็นฟังก์ชั่นที่ไม่เกี่ยวข้องกับ μ และ σ^2 ดังนั้นจึงสรุปได้ว่า $(\hat{\theta}_1, \hat{\theta}_2)$ เป็นเซทของสอง Sufficient Statistics สำหรับพารามิเตอร์ μ และ σ^2

คุณสมบัติของ Sufficient Statistics

กำหนดให้ $\hat{\theta}$ เป็นตัว Sufficient Statistic สำหรับ θ แล้วเราจะได้ว่า

1. $\hat{\theta}$ ไม่จำเป็น ต้องเป็นตัวประมาณที่ปราศจากความเอียงเฉต่อค่า θ

พิจารณาจากตัวอย่างที่ 1 จะพบว่า $X_i \sim$ Bernoulli (p)

ถ้าเราสุ่มตัวอย่างมาขนาด 3 คือ X_1, X_2, X_3 เมื่อ $T = X_1+X_2+X_3$; T จะมีคุณสมบัติเป็นตัว Sufficient Statistic สำหรับ p

สามารถพิสูจน์ได้ว่า $E(T) \neq p$

ส่วนในตัวอย่างที่ 2 ได้ว่า $\bar{X}$ เป็นSufficient Statistic สำหรับ μ ในขณะเดียวกันก็พิสูจน์ได้ว่า $E(\bar{X}) = \mu$

จากข้อสรุป ตัวอย่างที่ 1 และที่ 2 ก็คือ Sufficient Statistic จึงไม่มีคุณสมบัติว่าจะต้องปราศจากความเอียงเฉต่อพารามิเตอร์ เสมอไป

2. ทุกฟังก์ชั่นของการแจกแจงไม่จำเป็นจะต้องหาตัว Sufficient Statistic ได้เสมอไป ตัวอย่างของ Distribution ที่เราไม่สามารถจะหา Sufficient Statistic ได้ก็คือ Cauchy Distribution

$X_i \sim$ Cauchy (θ)

$$f(x;\theta) = \frac{1}{\pi(1+(x-\theta)^2)} \; ; \; -\infty<x<\infty$$

(ลองทำเป็นแบบฝึกหัด)

ตัวอย่างที่ 5

กำหนดให้ X เป็นตัวแปรเชิงสุ่ม ซึ่งมีการแจกแจง

$$f(x;\theta) = \frac{1}{\theta}, 0<x<\theta$$

จงแสดงให้เห็นว่า Y_n เป็นตัวสถิติอันดับที่ n จะเป็น Sufficient Statistics สำหรับ θ

$$\sum_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}$$

$$g(y_n,\theta) = n\,[F(y_n)]^{n-1} f(y_n)$$

$$= n\,[\int_0^{y_n} \frac{1}{\theta}\, dx\,]^{n-1} \frac{1}{\theta}$$

$$= \frac{ny_n^{n-1}}{\theta^n} \; ; \; 0<y_n<\theta$$

$$\frac{1}{\theta^n} = \frac{ny_n^{n-1}}{\theta^n} \cdot \frac{1}{ny_n^{n-1}}$$

โดยทฤษฎีที่ 1

$$\sum_{i=1}^{n} f(x_i, \theta) = g(y_n,\theta)\, k(x_1,\ldots,x_n)$$

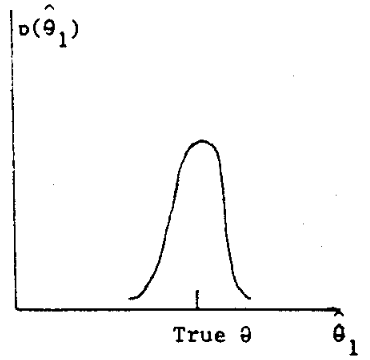
$$k(x_1,\ldots,x_n) \;=\; \frac{1}{ny_n^{n-1}}$$ ซึ่งไม่ขึ้นอยู่กับพารามิเตอร์ θ

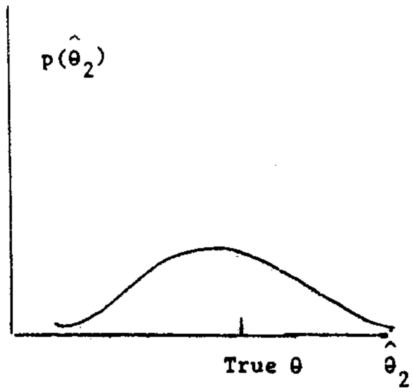
ดังนั้น Y_n จึงเป็น sufficient statistic สำหรับ θ

3. ประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ (Efficiency)

คุณสมบัติข้อนี้ก็คือ การนำบรรดาตัวประมาณค่าทั้งหลายมาเปรียบเทียบกันในแง่ของความแปรปรวนนั่นเอง เช่น ถ้ามี $\hat{\theta}_1$, $\hat{\theta}_2$ ทั้งคู่ต่างใช้เป็นตัวประมาณค่าของ θ หน้าที่ของเราก็คือเลือกตัวประมาณค่าที่มีค่าของความแปรปรวนน้อย ๆ นั่นเอง

ให้พิจารณาจากรูปประกอบ





จากรูปได้ว่า $\hat{\theta}_1$ จะมีประสิทธิภาพ (more efficient) มากกว่า $\hat{\theta}_2$ ทั้งนี้เพราะ $\hat{\theta}_1$ มีความแปรปรวนน้อยกว่า $\hat{\theta}_2$ นั่นเอง เราวัดประสิทธิภาพได้ดังนี้

1. ประสิทธิภาพของ $\hat{\theta}_1$ เมื่อเปรียบเทียบกับ $\hat{\theta}_2$ (Relative efficiency of $\hat{\theta}_1$ compared to $\hat{\theta}_2$)

$$= \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}\ 100\%$$

ในแง่ของ Relative efficiency ถ้าค่าที่ได้เท่ากับ 100% ก็แปลว่า $\hat{\theta}_1$ กับ $\hat{\theta}_2$ มีประสิทธิภาพเท่ากัน ถ้าค่าที่ได้เกิน 100% แปลว่า $\hat{\theta}_2$ ดีกว่า $\hat{\theta}_1$ ถ้าค่าที่ได้ต่ำกว่า 100% แปลว่า $\hat{\theta}_2$ ด้อยกว่า $\hat{\theta}_1$

ในกรณีที่ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ต่างก็ไม่ใช่ค่าประมาณที่ปราศจากความเอียงเฉต่อพารามิเตอร์ θ การเปรียบเทียบประสิทธิภาพของ $\hat{\theta}_1$ ต่อ $\hat{\theta}_2$ จะกลายเป็นข้อ 2

2. ประสิทธิภาพของ $\hat{\theta}_1$ เมื่อเปรียบเทียบกับ $\hat{\theta}_2$ (Relative efficiency of $\hat{\theta}_1$ compare to $\hat{\theta}_2$)

$$= \frac{E(\hat{\theta}_2 - \theta)^2}{E(\hat{\theta}_1 - \theta)^2}$$

ผลที่ได้รับก็อภิปรายได้เช่นเดียวกับ ข้อ 2

ตัวอย่างที่ 1 สุ่มตัวอย่างขนาด n มาจากประชากรที่มีการกระจายปกติ ที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 จะพบว่าตัวสถิติที่เราสร้างขึ้น เช่น ค่าเฉลี่ยจากตัวอย่าง $\overline{X}$ และค่ามัธยฐานจากตัวอย่าง md เพื่อใช้เป็นตัวประมาณค่าของพารามิเตอร์ μ นั้น ทั้ง $\overline{X}$ และ md ต่างก็เป็นค่าประมาณที่ปราศจากความเอียงเฉต่อ μ ทั้งคู่ โดยที่เราทราบว่าความแปรปรวนของ $\overline{X}$ คือ σ^2/n และความแปรปรวนของ md คือ (ประมาณ) $(\pi/2)(\sigma^2/n)$

ดังนั้นประสิทธิภาพของค่าเฉลี่ยจากตัวอย่างเมื่อเปรียบเทียบกับค่ามัธยฐาน

$$= \frac{(\pi/2)(\sigma^2/n)}{\sigma^2} = \pi/2 = 157\%$$

สรุปได้ว่า $\overline{X}$ มีประสิทธิภาพสูงกว่า md ถึง 57% เราจึงเลือก $\overline{X}$ เป็นตัวประมาณค่าของ μ แทนที่จะใช้ md

พิจารณาในกรณีทั่ว ๆ ไป ในบรรดาตัวประมาณค่าทั้งหลายที่เราสร้างขึ้นกำหนดให้เป็น $\{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k\}$ เป็นเซทที่ประกอบด้วยตัวประมาณค่าของ θ (พารามิเตอร์ของประชากร) ซึ่งมีคุณสมบัติว่าเป็นตัวประมาณค่าที่ปราศจากความเอียงเฉต่อ θ

นั่นคือ $E(\hat{\theta}_i) = \theta \quad ; \quad i = 1,2,\ldots,k$

การที่จะเลือก $\hat{\theta}_i$ ตัวใดตัวหนึ่งในเซทนั้นมาเป็นตัวประมาณค่าของพารามิเตอร์ θ โดยใช้คุณสมบัติในเรื่องของประสิทธิภาพเป็นเครื่องวัดนั้น เราอาจใช้วิธีการง่าย ๆ ดังนี้คือ เลือก $\hat{\theta}_j$ ซึ่ง

$$V(\hat{\theta}_j) \leqslant V(\hat{\theta}_i) \quad ; \quad i = 1,2,\ldots,k \; ; \; i \neq j$$

นั่นก็คือจะได้ตัวประมาณค่าที่มีความแปรปรวนน้อยที่สุด (ต่ำที่สุด) ในบรรดาตัวประมาณค่าทั้งหลายในเซทของตัวประมาณค่านั้น (หรืออีกนัยหนึ่งก็คือตัวประมาณค่าที่มีประสิทธิภาพสูงที่สุดนั่นเอง)

หมายเหตุ เราเรียกว่า $\hat{\theta}_j$ เป็นตัวประมาณที่มีประสิทธิภาพสูงที่สุด (Efficient-statistic) ในเรื่องนี้จะกล่าวอีกครั้งในเรื่องของ Minimum Lower Bound (Minimum Variance Unbiased Estimator)

โดยทั่วไปจะใช้สูตร

$$\eta = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

หรือจะเทียบเป็นร้อยละ

$$\eta = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} \times 100\%$$

ตัวอย่างที่ 2 กำหนดให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าสำหรับพารามิเตอร์ θ โดยที่เราสร้าง $\hat{\theta}_1$ และ $\hat{\theta}_2$ มาจากกลุ่มตัวอย่าง $X_1,\ldots,X_n$ ที่มาจากประชากรที่มีการแจกแจงแบบปกติมีค่าความแปรปรวนคือ σ^2 กำหนดให้ $\hat{\theta}_1$ เป็นตัวประมาณค่าที่ปราศจากความเอียงเฉต่อ θ โดยมี $V(\hat{\theta}_1) = \frac{2\sigma^4}{n}$ และ $\hat{\theta}_2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$ โดยที่ $\hat{\theta}_2$ เป็นตัวประมาณค่าที่ปราศจากความเอียงเฉต่อ θ เช่นกัน

$$V(\hat{\theta}_1) = \frac{2\sigma^4}{n}$$

$$V(\hat{\theta}_2) = \frac{1}{(n-1)^2} V\left[\Sigma(X_i - \overline{X})^2\right]$$

$$= \frac{\sigma^4}{n-1} V(\chi^2_{n-1})$$

$$= \frac{2\sigma^4}{n-1}$$

$\therefore$ Relative efficiency ของ $\hat{\theta}_2 : \hat{\theta}_1 = \eta = \dfrac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)} = \dfrac{n-1}{n}$

ถ้ากำหนดขนาดตัวอย่าง = 10

$$\eta = \frac{9}{10} \times 100 = 90\%$$

ดังนั้นแสดงว่า $\hat{\theta}_1$ ดีกว่า $\hat{\theta}_2$

ในกรณีที่เราเพิ่มขนาดตัวอย่างให้ใหญ่ขึ้น

$$\lim_{n\to\infty} \eta = \lim_{n\to\infty} \frac{n-1}{n} = 1$$

เราอาจจะเรียกได้ว่า $\hat{\theta}_2$ มีคุณสมบัติเป็น assymtotically most efficient estimator ของ σ^2 ถ้าหากว่า $\hat{\theta}_1$ มีคุณสมบัติเป็น efficient estimator ของ σ^2

นั่นคือ $\hat{\theta}_2$ จะเป็นตัวประมาณค่า ที่มีประสิทธิภาพดีที่สุด เท่าเทียมกับ $\hat{\theta}_1$ ถ้าขนาดตัวอย่างใหญ่ขึ้น

แบบฝึกหัด

1. กำหนดให้สุ่มตัวอย่างมาขนาด n จากประชากรที่มีการแจกแจงแบบพัวซอง จงพิสูจน์ว่า ΣX_i จะเป็นตัว Sufficient Statistic สำหรับพารามิเตอร์ ตัวสถิติตัวนี้มีคุณสมบัติว่าปราศจากความเอียงเฉหรือไม่
2. กำหนดให้ฟังก์ชั่นความน่าจะเป็นของตัวแปรสุ่ม X คือ

$$f(x;\theta) = (\theta/2)^{|x|}(1-\theta)^{1-|x|}\ ;\ x = -1,0,1,\ 0<\theta<1$$

   a. ตัวแปร X เป็นตัว Sufficient Statistic หรือไม่
   b. ตัวแปร $|X|$ เป็นตัว Sufficient Statistic หรือไม่
3. กำหนดให้สุ่มตัวอย่างมาขนาด n จากประชากรที่มีการแจกแจงแบบปกติ $(N0,\theta)$; $\theta=\sigma^2$ จงพิสูจน์ว่าตัวสถิติต่อไปนี้มีคุณสมบัติว่าเป็นตัว Sufficient Statistic หรือไม่
   a. ΣX
   b. $\Sigma |X|$
   c. ΣX^2 เป็นตัวสถิติที่ปราศจากความเอียงเฉต่อพารามิเตอร์หรือไม่
4. ให้ $X_1,\ldots,X_n$ เป็นตัวแปรเชิงสุ่มชุดหนึ่งที่สุ่ม จากประชากรที่มีการแจกแจงแบบ Geometric นั่นคือ $f(x;\theta) = (1-\theta)^x\theta;\ x=0,1,2,\ldots;\ 0<\theta<1$ จงแสดงว่า ΣX_i เป็นตัว Sufficient Statistic สำหรับ θ
5. ให้ $X_1,\ldots,X_n$ เป็นตัวแปรเชิงสุ่มที่สุ่มมาจากประชากรที่มีการแจกแจงแบบ Beta โดยที่มีพารามิเตอร์ $\alpha=\theta>0$ และ $\beta=2$ จงแสดงว่าผลคูณของ $X_1,\ldots,X_n$ จะเป็นตัว Sufficient Statistic สำหรับ θ
6. กำหนดให้ $X_1,\ldots,X_n$ มีคุณสมบัติว่าเป็นอิสระต่อกัน และมีการแจกแจงเหมือนกัน จงแสดงให้เห็นว่า

$$\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$$

   เป็นตัวประมาณค่าที่ปราศจากความเอียงเฉต่อความแปรปรวนของประชากร

7. กำหนดให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าของพารามิเตอร์ θ โดยที่ $E(\hat{\theta}_1) = \theta$ และ $E(\hat{\theta}_2) = \theta + .5$ ความแปรปรวนของ $\hat{\theta}_1$ และ $\hat{\theta}_2$ มีค่าเป็น 6 และ 4 ตามลำดับ จงเปรียบเทียบดูว่า ตัวประมาณค่าตัวใดที่ดีที่สุด

8. กำหนดให้ $X_1, X_2, \ldots, X_n$ เป็นชุดของตัวแปรเชิงสุ่มที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ $N(0,\theta); 0<\theta<\infty$ จงแสดงว่า ΣX_i^2 เป็นตัว sufficient statistic สำหรับพารามิเตอร์ θ

9. จงพิสูจน์ว่าผลบวกของตัวแปรเชิงสุ่ม ที่มาจากประชากรที่มีการแจกแจงแบบพัวซองที่มีพารามิเตอร์เป็น $\theta(0<\theta<\infty)$ มีคุณสมบัติเป็นตัว sufficient statistic

10. กำหนดให้ $X_1, X_2, \ldots, X_n$ เป็นชุดของตัวแปรสุ่ม ที่สุ่มมาจาก geometric distribution ซึ่งมี pdf.

$$f(x;\theta) = (1-\theta)^x\theta,\ x = 0,1,2,\ldots,\ 0<\theta<1$$

จงแสดงว่า $\sum_{i=1}^{n} X_i$ เป็นตัว sufficient statistic

จงแสดงว่า ผลบวกของตัวแปรเชิงสุ่มขนาด n ที่สุ่มมาจากประชากร ที่มีการแจกแจงแบบแกมม่า ซึ่งมี pdf.

$f(x;\theta) = (1/\theta)e^{-x/\theta},\ 0<x<\infty,\ 0<\theta<\infty$ เป็นตัว sufficient statistic สำหรับ θ

11. ในงานวิจัยเกี่ยวกับการพัฒนาทางเศรษฐกิจนักเศรษฐศาสตร์ 2 คน ได้ทำการประมาณค่าเฉลี่ย μ (ค่าใช้จ่ายเฉลี่ยที่ใช้ไปในการซื้อสิ่งของบริโภค) นักเศรษฐศาสตร์ทั้งสองต่างมีแนวความคิดเป็นอิสระต่อกัน คนที่หนึ่งใช้ $\bar{X}_1$ เป็นค่าประมาณ คนที่สองใช้ $\bar{X}_2$ เป็นค่าประมาณ μ เป็นที่น่าสังเกตว่านักเศรษฐศาสตร์คนแรกทำงานค่อนข้างละเอียดกว่าคนที่สอง โดยที่เขาได้ค่าส่วนเบี่ยงเบนมาตรฐานจาก $\bar{X}_1$ ต่ำกว่าคนที่สองอยู่ 1/5 เท่าของส่วนเบี่ยงเบนมาตรฐานจาก $\bar{X}_2$ ถ้าหากว่าเราต้องการที่จะรายงานผลที่ได้โดยการนำผลที่ได้ของนักเศรษฐศาสตร์ทั้งสองมาร่วมกันเพื่อที่จะใช้ในการประมาณค่าเฉลี่ยที่ใช้ไปในการซื้อสิ่งอุปโภคของประชากร จากข้อสรุปทั้งสี่ที่ได้มาโดยนักสถิติให้เรียงลำดับของวิธีทั้งสี่โดยเรียงตามลำดับประสิทธิภาพของตัวประมาณค่าในแต่ละวิธี

1. $\hat{\mu}_1 = \frac{1}{2}(\overline{X}_1 + \overline{X}_2)$

2. $\hat{\mu}_2 = \frac{4}{5}\overline{X}_1 + \frac{1}{5}\overline{X}_2$

3. $\hat{\mu}_3 = \frac{5}{6}\overline{X}_1 + \frac{1}{6}\overline{X}_2$

4. $\hat{\mu}_4 = \overline{X}_1$

12. สุ่มตัวอย่างขนาด n ให้พิจารณาตัวประมาณค่าต่อไปนี้ (พารามิเตอร์คือ μ) (n = 2)

$$\overline{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

$$W = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

a. จงแสดงว่า ทั้งคู่เป็นตัวประมาณที่ปราศจากความเอียงเฉ

b. ให้เปรียบเทียบประสิทธิภาพของ W กับ $\overline{X}$ ตัวประมาณค่าใดดีกว่ากัน

13. โดยการใช้ตัวอย่างขนาด n หน่วย จงพิจารณาตัวประมาณค่าทั้งสองที่กำหนดให้เพื่อใช้ในการประมาณค่าพารามิเตอร์ μ

$$\overline{X}_{100} = \frac{1}{100}(X_1 + X_2 + \ldots + X_{100})$$

$$\overline{X}_{90} = \frac{1}{90}(X_1 + X_2 + \ldots + X_{90})$$

ให้เปรียบเทียบประสิทธิภาพของ $\overline{X}_{100}$ กับ $\overline{X}_{90}$

14. จงอภิปรายข้อความต่อไปนี้

ในปัญหาที่ 2 และ 3 จะเห็นได้ว่าเป็นตัวอย่างที่แสดงถึงการเปรียบเทียบประสิทธิภาพของตัวประมาณค่า ตัวอย่างในปัญหาที่ 3 เป็นตัวอย่างที่เห็นได้ชัดเจน เพราะว่าการคิดตัวประมาณโดยใช้ค่าสังเกตุเพียง 90 ค่าเพื่อหาค่า $\overline{X}_{90}$ ก็เท่ากับการทิ้งประโยชน์ที่ควรจะได้รับจากอีก 10 ข้อมูลทิ้งไปนั่นเอง ส่วนในปัญหาที่ 2 ข้อเสียของค่าประมาณ W ก็คือเราถ่วงน้ำหนักให้กับ X_1, X_2 ไม่เหมาะสม จากข้อเสียทั้งสองก็จะทำ

ให้ได้รับผลลัพธ์ที่ไม่เหมาะสมเช่นเดียวกัน (การทิ้งข้อมูลไป 10 หน่วย กับการวิเคราะห์ที่ผิดพลาดก็จะได้รับผลลัพธ์คือได้ตัวประมาณค่าที่ไม่ดีเช่นกัน) หรือจะกล่าวอีกนัยหนึ่งก็คือการที่เราใช้วิธีการวิเคราะห์ไม่ดีก็เปรียบเสมือนการทิ้งข้อมูลไปโดยไม่ใช้ประโยชน์นั่นเอง

จากเหตุผลที่เรายกมาเปรียบเทียบ ถ้าให้ท่านช่วยแนะนำกับนักวิจัยในการวิจัยชิ้นหนึ่งที่ต้องเสียเงิน 100,000 \$ ในการเก็บข้อมูลและ 100 \$ สำหรับการวิเคราะห์ข้อมูล

15. ชาวนาคนหนึ่งมีที่ดินของตนเองเป็นรูปสี่เหลี่ยม เขาต้องการประมาณพื้นที่ของที่ดินแปลงนี้ (เป็นรูปสี่เหลี่ยมด้านเท่า) โดยการวัดความยาวของที่ดินแปลงนี้ความยาวที่วัดได้คือ X_1 (ปรากฏว่าค่าที่วัดไว้มีความผิดพลาด) โดยที่ X_1 มีการแจกแจงแบบปกติมีค่า μ ใกล้เคียงกับ 200 ค่า $\sigma^2 = 20$ เพื่อจะแก้ไขข้อผิดพลาดที่เกิดขึ้นเขาทำการวัดความยาวด้านใหม่คือ X_2 หลังจากที่ได้ข้อมูลทั้งสองครั้งแล้ว ชาวนาผู้นี้จะเลือกตัวสถิติตัวใดที่จะเหมาะที่สุดในการประมาณพื้นที่ของเขา

a. $\left(\dfrac{X_1+X_2}{2}\right)^2$

b. $\dfrac{X_1^2 + X_2^2}{2}$

16. กำหนดให้ X เป็นตัวแปรเชิงสุ่มที่มีค่าคาดหมายเป็น μ และความแปรปรวนเป็น σ^2 ถ้าสุ่มตัวอย่างมาขนาด n คือ $X_1,\ldots,X_n$ เราจะสามารถสร้างตัวประมาณค่าของ σ^2 ได้หลายทางด้วยกัน ทางหนึ่งที่ได้ก็คือ $c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$ ซึ่งถ้าหากเราสามารถเลือกค่าของ c ให้เหมาะจะได้ตัวประมาณค่านี้เป็นตัวประมาณที่ปราศจากความเอียงเฉต่อ μ ให้หาค่า c ที่จะเป็นไปได้ในกรณีนี้

17. $X_1,X_2,\ldots,X_n$ คือกลุ่มตัวอย่างขนาด n ที่สุ่มมาจากตัวแปร X เราทราบว่า $\sum_{i=1}^{200} X_i = 300$ และ $\sum_{i=1}^{200} X_i = 3754$ จากค่าที่กำหนดมาให้นี้จงสร้างค่าประมาณที่ปราศจากความเอียงเฉต่อ E(X) และ V(X)

18. ตัวประมาณค่าเป็นตัวแปรชนิดหนึ่งหรือไม่ ทำไมจึงเป็นเช่นนั้นยกเหตุผลประกอบ

19. จากข้อความต่อไปนี้จงอธิบายและแยกให้เห็นความแตกต่าง

a. สถิติเชิงพรรณนากับสถิติเชิงอนุเคราะห์

b. Estimate และ Estimator

20. จงอธิบายและแสดงเหตุผลในการคำนวณประกอบด้วยว่า ทำไม $\overline{X}$ จึงมีคุณสมบัติดังต่อไปนี้ (X_i มาจากประชากรที่มีการแจกแจงปกติมีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2)

a. เป็นค่าประมาณที่มีความแนบนัยต่อพารามิเตอร์ (Consistent Estimator)

b. เป็นค่าประมาณที่ปราศจากความเอียงเฉ (Unbiased Estimator)

c. เป็นค่าประมาณที่มีประสิทธิภาพที่สุด (An Efficient Estimator)

d. เป็นค่าประมาณที่บอกลักษณะของตัวพารามิเตอร์ได้อย่างเพียงพอ (A Sufficient Estimator)

21. $X_1, X_2, \ldots, X_n$ เป็นกลุ่มตัวอย่างที่มีการแจงใด ๆ โดยที่มีโมเมนต์รอบที่ k สามารถหาค่าได้คือ $E(X)^k$ จงแสดงว่า $\frac{1}{n}\sum_{i=1}^{n} X_i^k$ เป็นค่าประมาณที่แนบนัยต่อ $E(X^k)$

22. $\overline{X}$ เป็นค่าเฉลี่ยของตัวอย่าง โดยที่กลุ่มตัวอย่างได้มาจากประชากรที่มีการแจกแจงแบบทวินาม B(1,p) จงแสดงว่า $\overline{X}$ เป็นค่าประมาณที่อธิบาย p ได้อย่างเพียงพอ พร้อมกันนั้นให้แสดงด้วยว่า $\overline{X}$ มีคุณสมบัติที่เป็นตัวประมาณที่มีประสิทธิภาพที่สุดด้วย

23. $X_1, X_2, \ldots, X_n$ เป็นกลุ่มตัวอย่างที่มาจากประชากรที่มีการแจกแจงแบบ Exponential โดยที่มีฟังก์ชั่นความน่าจะเป็นคือ $f(x;\theta) = \theta . e^{-\theta x}$; $x>0$ (θ เป็นค่าบวก) จงแสดงว่าค่าเฉลี่ยจากตัวอย่าง $\overline{X}$ เป็นตัวประมาณที่อธิบายพารามิเตอร์ได้อย่างพอเพียง (Sufficient Statistic for θ) และจะได้ว่า $((n-1)/n)/\overline{X}$ จะเป็นค่าประมาณที่ปราศจากความเอียงเฉต่อ θ

24. ถ้า $\overline{X}$ และ S^2 ซึ่งเป็นค่าเฉลี่ยและความแปรปรวนจากตัวอย่าง ที่มีขนาด n โดยที่ประชากรมีการแจกแจงแบบปกติ $N(\mu, \sigma)^2$ จงแสดงว่า ตัวแปรสองมิติ $(\overline{X}, S^2)$ จะเป็นค่าประมาณที่เพียงพอ (Sufficient Statistics) สำหรับเวคเตอร์ของพารามิเตอร์ (μ, σ^2)

25. กำหนดให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าสำหรับพารามิเตอร์ θ โดยที่ทั้ง $\hat{\theta}_1$ และ

$\hat{\theta}_2$ ต่างก็เป็นตัวประมาณค่าที่มีประสิทธิภาพที่สุดทั้งคู่ จงแสดงว่าสหสัมพันธ์ระหว่าง $\hat{\theta}_1$ และ $\hat{\theta}_2$ มีค่าเท่ากับ 1

26. จงแสดงว่า สำหรับตัวประมาณค่าใด ๆ ของความแปรปรวน ซึ่งได้จากการแจกแจงปกติ โดยที่ตัวประมาณค่าเหล่านั้นมีคุณสมบัติว่าเป็นตัวประมาณที่ปราศจากความเอียงเฉแล้ว ตัวประมาณค่าเหล่านี้จะมีค่าของความแปรปรวนต่ำที่สุดคือ $2\sigma^4/n$ โดยกำหนดให้ $(X_i \sim N(\mu,\sigma^2)$

27. กำหนดให้ $\hat{\theta}_1$ เป็นตัวประมาณค่าที่มีประสิทธิภาพที่สุดสำหรับ θ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าตัวอื่นที่มีคุณสมบัติว่าเป็นตัวประมาณที่ปราศจากความเอียงเฉต่อ θ โดยที่ความแปรปรวนของ $\hat{\theta}_1$ และ $\hat{\theta}_2$ คือ σ^4, $k\sigma^2 (k>1)$ ตามลำดับ จงแสดงว่า สหสัมพันธ์ระหว่าง $\hat{\theta}_1$ และ $\hat{\theta}_2$ มีค่าเท่ากับ $1/\sqrt{k}$

28. ถ้ากำหนดให้ T เป็นตัวประมาณที่เพียงพอต่อ θ(Sufficient for θ) แล้ว จงแสดงว่าสำหรับฟังก์ชั่นชนิด one to one แล้วเราจะได้ว่า $\phi(T)$ ก็จะมีคุณสมบัติเป็นตัวประมาณค่าที่เพียงพอสำหรับ θ ด้วย (ให้พิสูจน์ผลที่ได้กับกรณีต่อไปนี้ด้วย $2\bar{X}+3$ เป็นตัวประมาณที่เพียงพอต่อ $2\mu + 3$ จากตัวอย่างขนาด n ที่มาจากประชากรที่มีการแจกแจงปกติ $N(\mu, \sigma^2)$)

29. กำหนดให้ T เป็นตัวประมาณค่าที่มีความแนบนัยต่อ θ (Consistent Estimator for θ) จงแสดงว่า $nT/(n-1)$ ก็จะมีคุณสมบัติ Consistent ต่อ θ ด้วย

30. จงแสดงว่าผลคูณของตัวแปรเชิงสุ่ม ที่สุ่มมาจากประชากรที่มีการแจกแจงแบบ beta ซึ่งมีพารามิเตอร์ คือ $\alpha = \theta > 0$ และ $\beta = 2$ มีคุณสมบัติเป็นตัว sufficient statistics สำหรับพารามิเตอร์ θ

31. ในวงจรไฟฟ้าอันหนึ่งประกอบด้วยชิ้นส่วนย่อย ๆ 3 ชิ้นต่อกัน วงจรอันนี้จะไม่ทำงานถ้าหากชิ้นส่วนหนึ่งส่วนใดชำรุด โดยที่ชิ้นส่วนแต่ละชิ้นเป็นอิสระจากกัน โดยอายุการใช้งานมีการแจกแจงแบบเอ็กโพเนนเชียล มีพารามิเตอร์เป็น λ กำหนดให้ T_i เป็นอายุการใช้งานของแต่ละชิ้นส่วน โดยที่ $t_1 \leq t_2 \leq t_3$

ก. จงหาค่าเฉลี่ยอายุการใช้งานของวงจรไฟฟ้านี้