การทดสอบสมมติฐาน (Hypothesis Testing)

รศ.ดร. วีระชาติ กิเลนทอง
มหาวิทยาลัยหอการค้าไทย

บทนำ

- หัวข้อนี้นำเสนอหลักการและวิธีการที่เกี่ยวข้องกับการทดสอบสมมุติฐาน โดยให้ความสำคัญกับหลักการต่อไปนี้
  - ความสัมพันธ์ระหว่าง Type I and Type II errors และฟังก์ชันอำนาจทางสถิติ (power function)
  - บทบาทของ ฟังก์ชันอำนาจทางสถิติ (power function) ในการกำหนดระดับนัยสำคัญทางสถิติ (significance level) และการคำนวณหาค่า p-value
  - การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

นิยาม: สมมุติฐาน (hypothesis)

Definition

สมมุติฐาน (hypothesis) คือข้อความที่เขียนในรูปของค่าพารามิเตอร์ของประชากรหรือค่าพารามิเตอร์ที่แท้จริง (true parameters) ซึ่งแทนด้วย $\theta \in \Omega$ โดยที่ Ω คือปริภูมิของพารามิเตอร์ (parameter space) หรือเซ็ตของค่าพารามิเตอร์ที่เป็นไปได้ทั้งหมด

- ตัวอย่างเช่น ภายใต้สมมุติฐานว่าง (null hypothesis) ที่ว่า $H_0 : \lambda = \lambda_0$ โดยที่ λ คือพารามิเตอร์ที่แท้จริงของการแจกแจงแบบ Poisson เราจะทราบทันทีว่ากลุ่มตัวอย่างถูกสุ่มมาจากการแจกแจงแบบ Poisson ที่มีค่าพารามิเตอร์เท่ากับ λ_0

นิยาม: สมมุติฐานอย่างง่าย (simple hypothesis) และ สมมุติฐานประกอบ (composite hypothesis)

Definition

สมมุติฐานถูกแบ่งออกเป็น 2 สมมุติฐาน ด้วยการแบ่งส่วน (partition) Ω ออกได้เป็น 2 เซ็ตที่ไม่มีส่วนร่วมต่อกัน (disjoint) Ω_0 และ Ω_1 ดังต่อไปนี้

1. สมมุติฐานว่าง (null hypothesis) คือสมมุติฐานที่เป็นจริงหากค่าพารามิเตอร์ $\theta \in \Omega_0$ ซึ่งแทนด้วย $H_0 : \theta \in \Omega_0$
2. สมมุติฐานทางเลือก (alternative hypothesis) คือสมมุติฐานที่เป็นจริงหากค่าพารามิเตอร์ $\theta \in \Omega_1$ ซึ่งแทนด้วย $H_1 : \theta \in \Omega_1$

Definition

ถ้า Ω_i มีสมาชิกเพียงตัวเดียว (singleton) เราจะเรียกสมมุติฐาน H_i ว่าเป็นสมมุติฐานอย่างง่าย (simple hypothesis) แต่ถ้า Ω_i มีสมาชิกมากกว่าหนึ่งตัว เราจะเรียกสมมุติฐาน H_i ว่าเป็นสมมุติฐานประกอบ (composite hypothesis)

ตัวอย่างสมมุติฐานประกอบ (composite hypothesis)

Example

กองทุน K-USXNDQ มีอัตราผลตอบแทนรายเดือนเท่ากับ R_A ซึ่งมีค่าคาดหมาย (mean) เท่ากับ μ_A ดังนั้น ปริภูมิพารามิเตอร์ (parameter space) ในกรณีนี้เท่ากับ $\Omega = \{\mu_A \in \mathbb{R}\}$ สมมุติว่าเราต้องการตัดสินใจว่าจะลงทุนในกองทุนนี้หรือไม่ โดยมีเงื่อนไขว่า จะลงทุนในกองทุนนี้ หากค่าคาดหมายของอัตราผลตอบแทนมีค่ามากกว่าหรือเท่ากับร้อยละ 1 ต่อเดือน หรือ $\mu_A \geq 0.01$ ดังนั้น สมมุติฐานที่เหมาะสมในกรณีคือ

$$H_0 : \mu_A \geq 0.01 \text{ และ } H_1 : \mu_A < 0.01 \quad (1)$$

ซึ่งในกรณีนี้ $\Omega_0 = \left\{(\mu_A, \mu_B) \in \mathbb{R}^2 : \mu_A \geq \mu_B\right\}$ และ $\Omega_1 = \left\{(\mu_A, \mu_B) \in \mathbb{R}^2 : \mu_A < \mu_B\right\}$ และจะเห็นได้ว่า สมมุติฐานทั้งสองอันเป็นสมมุติฐานประกอบ (composite hypothesis)

- กรณีนี้เป็นการสมมุติว่า การตัดสินใจลงทุนขึ้นอยู่กับค่าคาดหมายของอัตราผลตอบแทนเพียงอย่างเดียว โดยไม่สนใจระดับความเสี่ยงของกองทุน ซึ่งในชีวิตจริงไม่ควรทำอย่างยิ่ง เพราะความเสี่ยงมีความสำคัญมากและเป็นเหตุผลสำคัญที่ทำให้การวิเคราะห์ทางการเงินมีความยุ่งยากและซับซ้อน

ส่วนวิกฤติ (critical region): การเชื่อมโยงสมมุติฐานกับข้อมูล

- กำหนดให้ตัวอย่างสุ่ม (random sample) คือข้อมูลที่ได้จากการสุ่ม $\mathbf{X} = (X_1, \ldots, X_n)$ และปริภูมิตัวอย่าง (sample space) S หมายถึงเซ็ตของค่าของตัวอย่างสุ่มที่เป็นไปได้ทั้งหมด:

$$S = \left\{ \mathbf{x} = (x_1, \ldots, x_n) : x_i \text{ เป็นค่าที่เป็นไปได้ของ } X_i \text{ สำหรับ } i = 1, \ldots, n \right\} \quad (2)$$

โดยที่ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X_i คือค่าจำนวนจริงที่อยู่ในส่วนค้ำจุน (support) ของตัวแปรสุ่ม X_i นั่นเอง

Definition

ส่วนวิกฤติ (critical region) ของกระบวนการทดสอบ $S_1 = \left\{ \mathbf{x} \in S : \text{ ผลการทดสอบสมมุติฐานโดยใช้ข้อมูล } \mathbf{x} \text{ ระบุว่าปฏิเสธ } H_0 \right\}$ หมายถึงเซ็ตของค่าของตัวอย่างสุ่มที่เมื่อนำมาทดสอบสมมุติฐานแล้วจะสรุปได้ว่าปฏิเสธ H_0

ส่วนวิกฤติ (critical region) ในรูปของค่าสถิติ (Statistic)

- เพื่อความสะดวกในการวิเคราะห์ เรามักนิยามส่วนวิกฤติ (critical region) โดยใช้ค่าสถิติสำหรับการทดสอบสมมุติฐาน (test statistic) $T = r(\mathbf{X})$

Definition

กำหนดให้ $\mathbf{X}$ คือตัวอย่างสุ่มที่สุ่มมาจากการแจกแจงที่ขึ้นอยู่กับพารามิเตอร์ θ และสมมุติฐาน

$$H_0 : \theta \in \Omega_0 \text{ และ } H_1 : \theta \in \Omega_1, \tag{3}$$

ส่วนวิกฤติ (critical region) $S_1 = \{\mathbf{x} \in S : r(\mathbf{x}) \in R\}$ ซึ่งเขียนในรูปของค่าสถิติ T

ตัวอย่าง: ส่วนวิกฤติ (critical region) ในรูปของค่าสถิติ (Statistic)

- พิจารณาการทดสอบว่าค่าคาดหมาย μ มีค่าเท่ากับค่าจำนวนจริง μ_0 หรือไม่? นั่นคือ $H_0 : \mu = \mu_0$
- การทดสอบนี้สามารถทำได้โดยใช้ค่าสถิติ $T = |\bar{X}_n - \mu_0|$ โดยที่ $\bar{X}_n = \frac{\sum_{i=1} nX_i}{n}$ และกำหนดให้ส่วนวิกฤติคือช่วง $R = [0, \infty)$
- ดังนั้น เราจะปฏิเสธ $H_0 : \mu = \mu_0$ ถ้าค่าสถิติที่คำนวณได้จากข้อมูลอยู่ในช่วง $[0, \infty)$
- Note: สมมุติฐานอันเดียวกันนี้สามารถทดสอบได้โดยใช้ค่าสถิติที่ต่างออกไป เช่น $T = \frac{\bar{X}_n}{\mu_0}$

คำถามที่สำคัญก็คือ แล้วเราควรจะเลือกใช้กระบวนการทดสอบแบบไหนดี? คำตอบก็คือ ควรจะเลือกวิธีที่มีโอกาสผิดพลาดน้อยที่สุด ซึ่งดูเหมือนจะเป็นข้อความง่ายๆ แต่ที่จริงแล้วเป็นเรื่องที่ยุ่งยากมากทีเดียว

นิยาม:ความผิดพลาดของการทดสอบสมมุติฐาน

Definition

การทดสอบสมมุติฐานสามารถเกิดความผิดพลาดได้ 2 รูปแบบดังต่อไปนี้

1. ความผิดพลาดประเภทที่ 1 (Type I error) หมายถึงความผิดพลาดที่เกิดขึ้นจากการ**ปฏิเสธ** H_0 ทั้งที่ในความเป็นจริงสมมุติฐานนั้นเป็นจริง นั่นคือ $\theta \in \Omega_0$
2. ความผิดพลาดประเภทที่ 2 (Type II error) หมายถึงความผิดพลาดที่เกิดขึ้นจากการ**ไม่ปฏิเสธ** H_0 ทั้งที่ในความเป็นจริงสมมุติฐานนั้นไม่เป็นจริง นั่นคือ $\theta \in \Omega_1$

Summary of Type I and Type II Errors

	When H_0 is True	When H_1 is True
Do not reject H_0	Correct decision $1-\alpha$	Type II error β
Reject H_0	Type I error α	Correct decision $1-\beta$

นิยาม: ฟังก์ชันอำนาจทางสถิติ (power function) ของกระบวนการทดสอบ δ

Definition

ฟังก์ชันอำนาจทางสถิติ (power function) ของกระบวนการทดสอบ δ ที่กำหนดให้ส่วนวิกฤติคือ S_1 สามารถเขียนได้เป็น

$$\pi\left(\theta|\delta\right) = Pr\left(\mathbf{X} \in S_1|\theta\right), \forall\theta \in \Omega \tag{4}$$

หรือ

$$\pi\left(\theta|\delta\right) = Pr\left(T\left(\mathbf{X}\right) \in R|\theta\right), \forall\theta \in \Omega \tag{5}$$

- ฟังก์ชันอำนาจทางสถิติ (power function) เป็นฟังก์ชันของพารามิเตอร์ที่แท้จริงหรือค่าพารามิเตอร์ทางทฤษฎี θ และขึ้นอยู่กับการแจกแจงของตัวอย่างสุ่มที่ขึ้นอยู่กับค่าพารามิเตอร์ θ นอกจากนี้ ยังขึ้นอยู่กับกระบวนการทดสอบผ่านช่องทางส่วนวิกฤติ S_1 หรือหากพิจารณาในรูปของค่าสถิติ T ก็จะขึ้นอยู่กับส่วนวิกฤติ R

ตัวอย่าง: ฟังก์ชันอำนาจทางสถิติ (power function) ของกระบวนการทดสอบ δ

- พิจาณาการทดสอบสมมุติฐาน

$$H_0 : \mu = \mu_0 \text{ และ } H_1 : \mu \neq \mu_0$$

  โดยทราบว่าค่าความแปรปรวนเท่ากับ σ^2

- พิจารณากระบวนการทดสอบ δ ที่ทดสอบโดยใช้ค่าสถิติ $T = |\bar{X}_n - \mu_0|$ และมีส่วนวิกฤติ $R = [c, \infty)$
- คำถามในที่นี้คือ ฟังก์ชันอำนาจทางสถิติ (power function) ของกระบวนทดสอบนี้เท่ากับเท่าใด?
- เนื่องจากตัวแปรสุ่ม X_i มีการแจกแจงแบบปกติที่มีค่าคาดหมายเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 ทำให้สามารถสรุปได้ว่า ค่าคาดหมายของตัวอย่างสุ่ม (sample mean) $\bar{X}_n$ มีการแจกแจงแบบปกติที่มีค่าคาดหมายเท่ากับ μ และค่าความแปรปรวนเท่ากับ $\frac{\sigma^2}{n}$

ตัวอย่าง: ฟังก์ชันอำนาจทางสถิติ (power function) สำหรับการแจกแจงปกติ

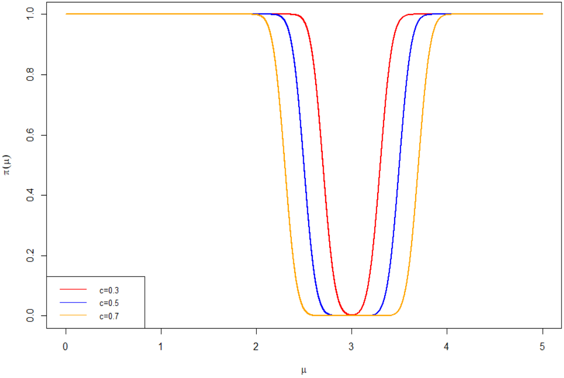
- ดังนั้น ฟังก์ชันอำนาจทางสถิติ (power function) สำหรับค่าพารามิเตอร์ μ ใดๆ เท่ากับ

$$
\begin{aligned}
\pi(\mu|\delta) &= Pr\left(T(\mathbf{X}) \in R|\mu\right) = Pr\left(|\bar{X}_n - \mu_0| \in [c, \infty)\,|\mu\right) \\
&= Pr\left(\bar{X}_n \geq \mu_0 + c|\mu\right) + Pr\left(\bar{X}_n \leq \mu_0 - c|\mu\right) \\
&= Pr\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{\mu_0 + c - \mu}{\frac{\sigma}{\sqrt{n}}}|\mu\right) + Pr\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{\mu_0 - c - \mu}{\frac{\sigma}{\sqrt{n}}}|\mu\right) \\
&= 1 - \Phi\left(\frac{\mu_0 + c - \mu}{\frac{\sigma}{\sqrt{n}}}\right) + \Phi\left(\frac{\mu_0 - c - \mu}{\frac{\sigma}{\sqrt{n}}}\right)
\end{aligned}
$$

- สังเกตว่า ฟังก์ชันอำนาจทางสถิติ (power function) $\pi(\mu|\delta)$ เป็นฟังก์ชันของ μ และค่าของฟังก์ชันนี้ขึ้นอยู่กับค่า c ซึ่งเป็นค่าที่กำหนดรูปแบบของกระบวนการทดสอบ และค่า μ_0 ซึ่งเป็นค่าที่กำหนดรูปแบบของสมมุติฐาน นั่นคือ ค่าของฟังก์ชันอำนาจทางสถิติ (power function) ขึ้นอยู่กับทั้งรูปแบบของสมมุติฐานและรูปแบบของกระบวนการทดสอบ

ฟังก์ชันอำนาจทางสถิติ (power function) ของกระบวนการทดสอบ δ

รูปด้านล่างแสดงค่าฟังก์ชันอำนาจทางสถิติ (power function) สำหรับการทดสอบสมมุติฐานที่กำหนดให้ $\mu_0 = 3$ ด้วยข้อมูลที่มีทั้งหมด $n = 1,000$ ตัวอย่าง และสมมุติให้ $\sigma^2 = 10$ สำหรับกระบวนการทดสอบที่แตกต่างกันซึ่งกำหนดด้วยค่า $c = 0.3, 0.5, 0.7$



ฟังก์ชันอำนาจทางสถิติ (power function) ในอุดมคติ

- ในอุดมคติ ระดับความผิดพลาดประเภทที่ 1 (Type I error) และระดับความผิดพลาดประเภทที่ 2 (Type II error) ควรจะมีค่าเท่ากับศูนย์ทั้งคู่
  1. $\pi\left(\theta|\tilde{\delta}\right) = 1$ สำหรับทุกๆ ค่า $\theta \in \Omega_1$
  2. $\pi\left(\theta|\tilde{\delta}\right) = 0$ สำหรับทุกๆ ค่า $\theta \in \Omega_0$
- ความยุ่งยากในการทดสอบสมมุติฐานเกิดขึ้นจากการที่ระดับความผิดพลาดประเภทที่ 1 (Type I error) และระดับความผิดพลาดประเภทที่ 2 (Type II error) มักจะมีค่าที่สวนทางกัน
- ทางออกคือ การเลือกกระบวนการทดสอบที่มีความสมดุลระหว่างความผิดพลาดทั้งสองประเภท

ระดับนัยสำคัญทางสถิติ (Statistical Significance Level)

- วิธีการที่ได้รับความนิยมอันหนึ่งคือ การเลือกจำนวนจริง $\alpha_0 \in (0,1)$ ที่ทำให้ Type I error สำหรับทุกๆ ค่า $\theta \in \Omega_0$ มีค่าไม่เกิน α_0 ซึ่งหมายถึง การกำหนดให้ Type I error มีค่าไม่เกิน α_0

Definition (ระดับนัยสำคัญ)

กระบวนการทดสอบ (test) ที่สอดคล้องกับเงื่อนไข

$$\pi(\theta|\delta) \leq \alpha_0, \text{ สำหรับทุกค่า } \theta \in \Omega_0 \tag{6}$$

ถูกเรียกว่าการทดสอบที่ระดับ α_0 กล่าวคือ การทดสอบนี้มีระดับนัยสำคัญ (significance level) เท่ากับ α_0 ส่วนขนาดการทดสอบ (size of test) ของกระบวนการทดสอบ δ สามารถนิยามได้เป็น

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta|\delta) \tag{7}$$

- ถ้าขนาดการทดสอบมีค่าเท่ากับ α_0 แล้วกระบวนการทดสอบ δ นั้นจะมีระดับนัยสำคัญมากกว่าหรือเท่ากับ α_0 เสมอ

การออกแบบกระบวนการทดสอบสมมติฐานให้มีระดับนัยสำคัญที่ต้องการ

- สมมุติฐานที่ต้องการทดสอบคือ

$$H_0 : \theta \in \Omega_0 \text{ และ } H_1 : \theta \in \Omega_1$$

- กระบวนการทดสอบที่พิจารณาใช้ค่าสถิติ T และส่วนวิกฤติ $R = [c, \infty)$
- ดังนั้น การออกแบบกระบวนการทดสอบในที่นี้คือการเลือก T และ c ที่ทำให้

$$\sum_{\theta \in \Omega_0} Pr\left(T \geq c|\theta\right) \leq \alpha_0 \tag{8}$$

- Note: ฟังก์ชันอำนาจทางสถิติในที่นี้มีค่าเท่ากับ $Pr\left(T \geq c|\theta\right)$ nonincreasing in c ดังนั้น ค่าที่ต่ำที่สุดของ c จะทำให้ค่าฟังก์ชันอำนาจทางสถิติสูงสุด ดังนั้น ควรจะเลือกค่า c ที่ต่ำที่สุดที่สอดคล้องกับเงื่อนไขด้านบน

ตัวอย่าง: การออกแบบกระบวนการทดสอบสมมติฐานให้มีระดับนัยสำคัญที่ต้องการ

- ต้องการทดสอบสมมุติฐานต่อไปนี้

$$H_0 : \mu = \mu_0 \text{ และ } H_1 : \mu \neq \mu_0$$

  โดยสมมุติว่าทราบค่าความแปรปรวน σ
- พิจารณากระบวนการทดสอบ (test) ที่ใช้ค่าสถิติ $T = |\bar{X}_n - \mu_0|$ และส่วนวิกฤติ $R = [c, \infty)$
- คำถามหลักในตัวอย่างนี้คือ ค่า c ที่ทำให้ขนาดการทดสอบเท่ากับ α_0:

$$Pr\left(|\bar{X}_n - \mu_0| \geq c | \mu = \mu_0\right) = \alpha_0 \tag{9}$$

- ตัวแปรสุ่ม $Y = \bar{X}_n - \mu_0$ แจกแจงแบบปกติด้วยค่าคาดหมาย 0 และค่าความแปรปรวน $\frac{\sigma^2}{n}$ ดังนั้น ตัวแปรสุ่ม

$$Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \tag{10}$$

  จึงมีการแจกแจงปกติมาตรฐานด้วยค่าคาดหมาย 0 และค่าความแปรปรวน 1

ตัวอย่าง: การออกแบบกระบวนการทดสอบสมมติฐานให้มีระดับนัยสำคัญที่ต้องการ

- เราสามารถเขียนสมการที่ (9) ในรูปของตัวแปรสุ่ม Z ได้เป็น

$$Pr\left(\frac{\bar{X}_n-\mu_0}{\frac{\sigma}{\sqrt{n}}}\leq -\frac{c}{\frac{\sigma}{\sqrt{n}}} \text{ และ } \frac{\bar{X}_n-\mu_0}{\frac{\sigma}{\sqrt{n}}}\geq \frac{c}{\frac{\sigma}{\sqrt{n}}}|\mu_0\right)=\alpha_0,$$

$$\Rightarrow Pr\left(Z\leq -\frac{c}{\frac{\sigma}{\sqrt{n}}} \text{ และ } Z\geq \frac{c}{\frac{\sigma}{\sqrt{n}}}\right)=\alpha_0, \tag{11}$$

- ประยุกต์ใช้คุณสมบัติความสมมาตร (symmetric)

$$Pr\left(Z\leq -\frac{c}{\frac{\sigma}{\sqrt{n}}} \text{ และ } Z\geq \frac{c}{\frac{\sigma}{\sqrt{n}}}\right)=2Pr\left(Z\geq \frac{c}{\frac{\sigma}{\sqrt{n}}}\right)=2\left[1-\Phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right)\right]$$

ดังนั้น

$$2\left[1-\Phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right)\right]=\alpha_0\Rightarrow c=\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right) \tag{12}$$

ระดับความน่าจะเป็นที่จะปฏิเสธสมมุติฐานจริง (p-value)

- ค่า p-value คือค่าระดับนัยสำคัญทางสถิติ $\underline{\alpha}_0$ ที่ต่ำที่สุดที่การทดสอบโดยใช้ข้อมูลที่มีอยู่ (observed data) สามารถปฏิเสธสมมุติฐาน H_0 ได้ที่ระดับนัยสำคัญทางสถิตินั้น ซึ่งหมายความว่า ย่อมจะปฏิเสธสมมุติฐาน H_0 สำหรับการทดสอบสมมุติฐานด้วยข้อมูลอันเดียวกันที่ระดับนัยสำคัญทางสถิติ $\alpha_0 \geq \underline{\alpha}_0$ ใดๆ

Definition

ระดับความน่าจะเป็นที่จะปฏิเสธสมมุติฐานจริง (p-value) หมายถึงระดับนัยสำคัญทางสถิติที่ต่ำที่สุด $\underline{\alpha}_0 \equiv \hat{p}$ ที่การทดสอบโดยใช้ข้อมูลที่มีอยู่ (observed data) สามารถปฏิเสธสมมุติฐาน H_0 ได้ที่ระดับนัยสำคัญทางสถิตินั้น

- นักทดลอง (experimenter) ที่ต้องการการทดสอบที่ระดับ α_0 จะปฏิเสธสมมุติฐานว่างก็ต่อเมื่อ p-value, $\hat{p}\left(\mathsf{X}|\delta\right)$, มีค่าไม่เกิน α_0 บางครั้งจึงเรียก p-value ว่า ระดับนัยสำคัญทางสถิติที่คำนวณได้ (observed significance level)

ตัวอย่าง: ระดับความน่าจะเป็นที่จะปฏิเสธสมมุติฐานจริง (p-value)

- พิจารณากระบวนการทดสอบสมมุติฐาน

$$H_0 : \mu = \mu_0 \text{ และ } H_1 : \mu \neq \mu_0$$

ที่ใช้ค่าสถิติ $Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

- สมมุติว่าค่าสถิติ Z จากข้อมูลที่เก็บรวบรวมมาได้มีค่าเท่ากับ $\hat{Z} = -3.14$ ซึ่งทำให้สรุปได้ว่า เราควรจะต้องปฏิเสธ H_0 ที่ระดับนัยสำคัญ α_0 ที่สอดคล้องกับเงื่อนไข

$$|\hat{Z}| = |-3.14| \geq \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)$$

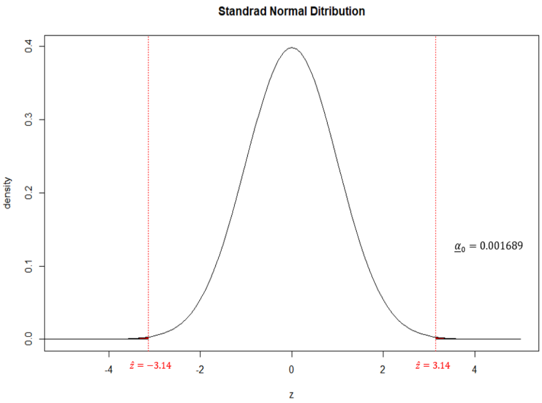
- เนื่องจาก พจน์ทางขวามือมีค่าเพิ่มขึ้นเมื่อค่าระดับนัยสำคัญ α_0 ลดลง ดังนั้น ค่าระดับนัยสำคัญ $\underline{\alpha}_0$ ที่ต่ำสุดที่ทำให้เงื่อนไขข้างบนเป็นจริง จะสอดคล้องกับสมการ

$$|-3.14| = \Phi^{-1}\left(1 - \frac{\underline{\alpha}_0}{2}\right)$$

ซึ่งคำตอบที่ได้ก็คือ $\underline{\alpha}_0 = 0.0017$ ดังนั้น จึงสามารถสรุปได้ว่า ระดับความน่าจะเป็นที่จะปฏิเสธสมมุติฐานจริง (p-value) $\hat{p} = 0.0017$

ตัวอย่าง: ระดับความน่าจะเป็นที่จะปฏิเสธสมมุติฐานจริง (p-value)

รูปด้านล่างแสดงค่า p-value โดยเริ่มจากการคำนวณค่า $\hat{z}$ แล้วจึงคำนวณหาค่า p-value จากพื้นที่ใต้กราฟที่แรเงาด้วยสีแดง



การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

- ค่าสถิติ (statistic) ที่ได้รับความนิยมอันหนึ่งคือ ค่าสัดส่วนความเป็นไปได้ (likelihood ratio statistic)

$$\Lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} \mathscr{L}(\theta|\mathbf{x})}{\sup_{\theta \in \Omega} \mathscr{L}(\theta|\mathbf{x})} \tag{13}$$

- ค่าสัดส่วนความเป็นไปได้ (likelihood ratio statistic) สามารถนำมาใช้ทดสอบสมมุติฐานที่อยู่ในรูปแบบ

$$H_0 : \theta \in \Omega_0 \text{ และ } H_1 : \theta \in \Omega_1$$

โดยที่จะปฏิเสธสมมุติฐาน H_0 ถ้า $\Lambda(\mathbf{x}) \leq \lambda$ สำหรับค่าคงที่ λ ค่าใดค่าหนึ่ง

- การทดสอบสมมุติฐานด้วยวิธีนี้จะปฏิเสธ H_0 ก็ต่อเมื่อค่าสถิติ $\Lambda(\mathbf{x})$ มีค่าต่ำ ซึ่งหมายความว่า ข้อจำกัดที่มาจาก H_0 มีผลทำให้ค่าสูงสุดของฟังก์ชันความเป็นไปได้มีค่าลดลงอย่างมากเมื่อเทียบกับกรณีที่ไม่มีข้อจำกัด ซึ่งมีนัยว่า ค่าพารามิเตอร์ที่แท้จริงน่าจะไม่อยู่ใน Ω_0 จึงควรปฏิเสธสมมุติฐาน

ตัวอย่าง: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

- กำหนดให้ X_i คือตัวแปรสุ่มที่แทนผลของการทดลองยาชนิดใหม่ซึ่งมีพารามิเตอร์ θ แทนความน่าจะเป็นของความสำเร็จในแต่ละครั้ง
- สมมุติว่า Y คือตัวแปรสุ่มที่แทนจำนวนครั้งที่จะสำเร็จ
- ฟังก์ชันความเป็นไปได้สำหรับการทดลองทั้งหมด n ครั้งในรูปของจำนวนครั้งที่สำเร็จคือ

$$\mathscr{L}(\theta|y) = \frac{n!}{y!(n-y)!}\theta^y(1-\theta)^{n-y} \tag{14}$$

- ตัวอย่างนี้ต้องการทดสอบ

$$H_0 : \theta = \theta_0 \text{ และ } H_1 : \theta \neq \theta_0$$

- ในกรณีนี้ $\Omega_0 = \{\theta_0\}$ และ $\Omega = [0,1]$ ดังนั้น ค่าสัดส่วนความเป็นไปได้ (likelihood ratio statistic) เท่ากับ

$$\Lambda(y) = \frac{\theta_0^y(1-\theta_0)^{n-y}}{\sup_{\theta\in[0,1]}\theta^y(1-\theta)^{n-y}} \tag{15}$$

ตัวอย่าง: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

- คำตอบของปัญหาการหาค่าความเป็นไปได้สูงสุดในส่วนของตัวหารคือ ตัวประมาณค่า MLE

$$\hat{\theta}_{MLE}(\mathbf{y}) = \frac{y}{n}$$

และค่าฟังก์ชันความเป็นไปได้สูงสุดเท่ากับ

$$\left(\frac{y}{n}\right)^{y}\left(\frac{n-y}{n}\right)^{n-y}$$

- ดังนั้น ค่าสถิติ $\Lambda(\mathbf{y})$ มีค่าเท่ากับ

$$\Lambda(y) = \frac{\theta_0^{y}(1-\theta_0)^{n-y}}{\left(\frac{y}{n}\right)^{y}\left(\frac{n-y}{n}\right)^{n-y}} = \left(\frac{n\theta_0}{y}\right)^{y}\left(\frac{n(1-\theta_0)}{n-y}\right)^{n-y} \tag{16}$$

ตัวอย่าง: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

- สมมุติเพิ่มเติมอีกว่า มีการทดลองทั้งหมด $n = 10$ ครั้ง และค่า $\theta_0 = 0.6$
- คำถามแรกที่ต้องการตอบคือ ถ้าระดับนัยสำคัญทางสถิติ $\alpha_0 = 0.05$ เราควรจะต้องเลือกค่าคงที่ λ เท่ากับเท่าใด?
- การหาคำตอบในส่วนนี้เริ่มจากนิยามของระดับนัยสำคัญทางสถิติ

$$\alpha_0 \geq \sup_{\theta \in \Omega_0} \pi(\theta|\delta) = \pi(\theta_0|\delta) = Pr(\Lambda(y) \leq \lambda|\theta_0) = Pr(Y|\theta_0) \qquad (17)$$

โดยที่เซ็ต $Y = \{y \in \{0, 1, \ldots, 10\} : \Lambda(y) \leq \lambda\}$ และ

- เนื่องจากการทดลองแต่ละครั้งเป็นอิสระต่อกัน ดังนั้น $Pr(Y|\theta_0)$ จะมีค่าเท่ากับผลรวมของความน่าจะเป็นของแต่ละค่า $y \in Y$

ตัวอย่าง: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests)

- ดังนั้น วิธีการหาค่า λ สามารถทำได้ด้วยการคำนวณหาค่า $\Lambda(y)$ และ $Pr(y|\theta_0)$ สำหรับแต่ละค่าของ y ที่เป็นไปได้ และนำมาเรียงลำดับจากค่า y ที่ให้ค่า $\Lambda(y)$ จากต่ำสุดไปถึงค่าสูงสุด ดังแสดงในตารางที่ 29
- หลังจากนั้นจึงหาค่าผลรวมของ $Pr(y|\theta_0)$ จากตำแหน่งต่ำสุดไปเรื่อยๆ จนได้ค่าที่ใกล้เคียงแต่ไม่เกิน $\alpha_0 = 0.05$ ซึ่งในที่นี้มีค่าเท่ากับ 0.018 และ ผลบวกนี้จบลงที่ตำแหน่งที่ค่า $\Lambda(y) = 0.035$
- ดังนั้น ค่าคงที่ λ ที่ต้องการคือ ค่าจำนวนจริงใดๆ ที่อยู่ระหว่าง $[0.035, 0.104)$

y	0	1	10	2	9	3	8	4
$\Lambda(y)$	0.000	0.004	0.006	0.035	0.104	0.159	0.400	0.444
$Pr(y\|\theta_0)$	0.000	0.002	0.006	0.011	0.040	0.042	0.121	0.111
$Pr(Y\|\theta_0)$	0.000	0.002	0.008	0.018	0.059	0.101	0.222	0.334
$-2\ln\Lambda(y)$	18.326	11.013	10.217	6.696	4.526	3.676	1.830	1.622

- ขนาดของการทดสอบเท่ากับ $\alpha(\delta) = \sup_{\theta\in\Omega_0} \pi(\theta|\delta) = Pr(Y|\theta_0) = 0.018$
- สังเกตว่า เราจะปฏิเสธ H_0 สำหรับค่า $y = 0, 1, 2$ หรือ 10

ทฤษฎี: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests) โดยใช้การแจกแจงที่ลิมิต (Asymptotic Distribution)

Theorem

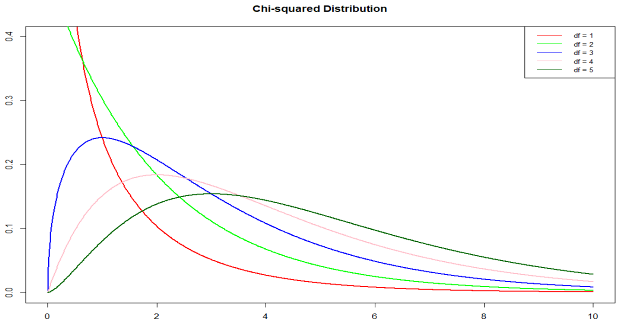
กำหนดให้ Ω เป็นปริภูมิของพารามิเตอร์ $\boldsymbol{\theta}$ ที่ประกอบไปด้วย p พารามิเตอร์ และสมมุติให้เงื่อนไขที่จำเป็นสำหรับการพิสูจน์ว่าการประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) มีการแจกแจงที่ลิมิตเป็นการแจกแจงปกติ พิจารณาสมมุติฐาน H_0 เป็นสมมุติฐานที่กำหนดให้ $h_i(\boldsymbol{\theta}) = \bar{h}_i$ สำหรับ $i = 1, \ldots, k$ โดยที่ $\bar{h}_i$ คือค่าคงที่ ดังนั้น เราสามารถพิสูจน์ได้ว่า ภายใต้ข้อกำหนดของ H_0

$$-2\log\Lambda(\mathrm{X}) \xrightarrow{d} \chi_k^2, \tag{18}$$

- ทฤษฎีบทด้านบนช่วยให้เราสามารถทราบการแจกแจงของ $-2\log\Lambda(\mathrm{X})$ ได้โดยที่ไม่ต้องคำนวณอะไรที่ยุ่งยาก
- ดังนั้น เราจึงนิยมทดสอบสมมุติฐานโดยใช้ค่าสถิติ $-2\log\Lambda(\mathrm{X})$ มากกว่า $\Lambda(\mathrm{X})$
- เราจะปฏิเสธสมมุติฐาน H_0 ถ้า $-2\log\Lambda(\mathrm{X}) \geq \bar{\lambda}$

รูปภาพ: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests) โดยใช้การแจกแจงที่ลิมิต (Asymptotic Distribution)

- เราสามารถกำหนดค่าวิกฤติ $\bar{\lambda}$ สำหรับการทดสอบที่มีระดับนัยสำคัญ α_0 ใดๆ ได้โดยใช้ $-2\log\Lambda(\mathbf{X}) \geq \left(\chi^2\right)^{-1}(1-\alpha_0)$
- สังเกตได้ว่าการทดสอบโดยใช้การแจกแจงแบบไคว์กำลังสอง (Chi square) จะใช้พื้นที่ส่วนท้ายของการแจกแจง (tail area) เพียงด้านเดียว เนื่องจากการแจกแจงแบบไคว์กำลังสองมีส่วนค้ำจุน (support) ที่เป็นบวกเท่านั้น



ตัวอย่าง: การทดสอบสมมุติฐานโดยใช้สัดส่วนความเป็นไปได้ (Likelihood Ratio Tests) โดยใช้การแจกแจงที่ลิมิต (Asymptotic Distribution)

- พิจารณาตัวอย่างที่นำเสนอมาก่อนหน้านี้ แต่คราวนี้จะประยุกต์ใช้ทฤษฎีบทที่ 11 (ถึงแม้ว่าขนาดตัวอย่างจะเล็กมากก็ตาม)
- โดยสำหรับตัวอย่างนี้ $k = 1$ ดังนั้น $-2 \log \Lambda(y) \xrightarrow{d} \chi_1^2$ ซึ่งหมายความว่า เราจะปฏิเสธสมมุติฐาน H_0 ที่ระดับนัยสำคัญ $\alpha_0 = 0.05$ ถ้า
  $-2 \log \Lambda(y) \geq \left(\chi_1^2\right)^{-1}(1 - 0.05) = 3.841$
- เมื่อเปรียบเทียบกับค่า $-2 \log \Lambda(y)$ กับค่าในตารางจะสรุปว่า เราจะปฏิเสธ H_0 ถ้า $y = 0, 1, 2, 9$ และ 10 ซึ่งแตกต่างจากกรณีการคำนวณโดยตรงเล็กน้อย ทั้งนี้เนื่องจากจำนวนตัวอย่างมีจำนวนน้อยมาก ทำให้การประมาณโดยใช้การแจกแจงที่ลิมิต (asymptotic distribution) เกิดความคลาดเคลื่อน

y	0	1	10	2	9	3	8	4
$\Lambda(y)$	0.000	0.004	0.006	0.035	0.104	0.159	0.400	0.444
$Pr(y\|\theta_0)$	0.000	0.002	0.006	0.011	0.040	0.042	0.121	0.111
$Pr(Y\|\theta_0)$	0.000	0.002	0.008	0.018	0.059	0.101	0.222	0.334
$-2 \ln \Lambda(y)$	18.326	11.013	10.217	6.696	4.526	3.676	1.830	1.622