หลักการประมาณแบบเบส์ (Bayes Estimation)

รศ.ดร. วีระชาติ กิเลนทอง
มหาวิทยาลัยหอการค้าไทย

บทนำ

- หลักการพื้นฐานของการประมาณค่า (estimation) โดยพิจารณาวิธีการประมาณค่าทั้งหมด 3 วิธี
  - **วิธีการประมาณค่าแบบเบส์ (Bayes Estimation)**
  - วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation หรือเรียกสั้นๆ ว่า MLE)
  - วิธีการประมาณค่าด้วยโมเมนต์ (Method of Moments หรือเรียกสั้นๆ ว่า MM)

นิยาม: พารามิเตอร์ (Parameter)

Definition

พารามิเตอร์ (parameter) $\boldsymbol{\theta} \in \Theta$ หมายถึงลักษณะเฉพาะ (characteristics) ที่กำหนดการแจกแจงร่วม (joint distribution) ของตัวแปรสุ่มที่สนใจ โดยเรียกเซ็ต Θ ว่าปริภูมิพารามิเตอร์ (parameter space)

นิยาม: แบบจำลองทางสถิติ (Statistical Model)

Definition

แบบจำลองทางสถิติ (statistical model) ประกอบไปด้วย

1. การระบุอย่างชัดเจนว่า "**ตัวแปรสุ่มที่สนใจมีอะไรบ้าง ทั้งที่สามารถสังเกตได้ (observable) และไม่สามารถสังเกตได้ (unobservable)**" แต่กำหนดให้มีอยู่ในทางทฤษฎี เช่น รายได้, อัตราดอกเบี้ย หรือ อัตราผลตอบแทนสุทธิ์ของสินทรัพย์ต่างๆ
2. "**ข้อกำหนดหรือสมการที่บอกถึงความสัมพันธ์ระหว่างตัวแปรสุ่มที่สนใจ**" โดยอาจอยู่ในรูปของการแจกแจงร่วม (joint distribution) สำหรับตัวแปรสุ่มที่สามารถสังเกตได้ (observable random variables)
3. "**การระบุค่าของพารามิเตอร์ (parameter identification) $\boldsymbol{\theta}$**" ที่ทำหน้าที่กำหนดการแจกแจงร่วม (joint distribution) ของตัวแปรสุ่มที่สนใจ โดยในส่วนนี้จะมองว่า **พารามิเตอร์ที่ไม่ทราบค่า (unknown parameters) เป็นค่าคงที่** นั่นคือ สิ่งที่ต้องการทราบค่าเป็นค่าจำนวนจริงหรือจุด ทำให้เรียกการประมาณค่าแบบนี้ว่า **การประมาณค่าแบบจุด (point estimation)** ส่วนการแจกแจงร่วมของตัวแปรสุ่มที่สามารถสังเกตได้ (observable random variables) ก็จะพิจารณาเป็นการแจกแจงร่วมแบบมีเงื่อนไข (joint conditional probability) สำหรับค่าที่เกิดขึ้นจริง (realized value) ของ $\boldsymbol{\theta}$ นั่นคือ $f(\boldsymbol{x}|\boldsymbol{\theta})$
4. สำหรับกรณีที่ "**สมมุติให้พารามิเตอร์ $\boldsymbol{\theta}$ เป็นตัวแปรสุ่ม**" สิ่งที่จำเป็นสำหรับแบบจำลองทางสถิติอีกอย่างหนึ่งคือ **การกำหนดรูปแบบการแจกแจงร่วม (joint distribution) ของพารามิเตอร์ที่ไม่ทราบค่า (unknown parameters)** ส่วนการแจกแจงร่วมของตัวแปรสุ่มที่สามารถสังเกตได้ (observable random variables) ก็จะพิจารณาเป็นการแจกแจงร่วมแบบมีเงื่อนไข (joint conditional probability) สำหรับค่าที่เกิดขึ้นจริง (realized value) ของ $\boldsymbol{\theta}$ นั่นคือ $f(\boldsymbol{x}|\boldsymbol{\theta})$ เช่นเดียวกับกรณีก่อนหน้านี้

นิยาม: การอนุมานทางสถิติ (Statistical Inference)

- โดยทั่วไป เราจำเป็นต้องประมาณค่าพารามิเตอร์เพราะ เราต้องการ **ทดสอบสมมุติฐาน (hypothesis testing)** ที่ต้องการทราบ
- การทดสอบสมมุติฐานจะอยู่ในรูปของข้อความเชิงความน่าจะเป็น (probabilistic statement) ที่เรียกอย่างเป็นทางการว่า **การอนุมานทางสถิติ (statistical inference)**
- ยกตัวอย่างเช่น เราอาจจะสนใจว่า สามารถบอกด้วยความมั่นใจแค่ไหนว่าค่าพารามิเตอร์ θ มีค่ามากกว่าศูนย์ นั่นคือ ต้องการทดสอบสมมุติฐานที่ว่า $\theta > 0$ เป็นต้น

Definition

การอนุมานทางสถิติ (statistical inference) คือกระบวนการ (procedure) ที่สร้างข้อความเชิงความน่าจะเป็น (probabilistic statement) ที่เกี่ยวข้องกับแบบจำลองทางสถิติ (statistical model)

รูปแบบของการอนุมานทางสถิติ

การอนุมานทางสถิติ (statistical inference) แบ่งได้เป็น 4 รูปแบบดังนี้

- **การคาดการณ์ (prediction)** เป็นคาดเดาค่าตัวแปรสุ่มหรือฟังก์ชันของตัวแปรสุ่มที่ยังไม่ทราบค่า
  - ยกตัวอย่างเช่น การคาดการณ์ว่าค่าเฉลี่ยของผลตอบแทนของกองทุนรวมในปีหน้ามีค่าเท่ากับเท่าใด? การคาดการณ์ว่าผลิตภัณฑ์มวลรวม (GDP) ของประเทศในปีหน้าจะขยายตัวร้อยละเท่าใด? หรือ การคาดการณ์ว่าในปีหน้าจะมีเด็กที่มีพัฒนาช้ากว่าวัยร้อยละเท่าใด?
  - เรามักเรียกการคาดการณ์ในกรณีที่สิ่งที่ต้องคาดการณ์หรือพยากรณ์ว่า **การประมาณค่า (estimation)** ซึ่งคือประเด็นหลักของบทนี้
- **การออกแบบการทดลอง (experimental design)** เป็นการกำหนดว่าควรจะเก็บข้อมูลอย่างไรและมากน้อยเพียงใด รวมทั้งรูปแบบการทดลอง
  - ยกตัวอย่างเช่น การทดลองแบบสุ่ม (randomized controlled trials) ที่ดีควรจะมีการออกแบบจำนวนกลุ่มตัวอย่างที่เหมาะสม เพื่อให้มั่นใจได้ว่าการทดสอบสมมุติฐานที่จะดำเนินการมีอำนาจทางสถิติ (statistical power) มากเพียงพอ

รูปแบบของการอนุมานทางสถิติ Con't

- **ปัญหาการตัดสินใจทางสถิติ (statistical decision problems)** หมายถึงกระบวนการตัดสินใจที่ผลที่ตามมาขึ้นอยู่กับค่าของพารามิเตอร์ที่ไม่ทราบค่า (แต่ต้องใช้ข้อมูลและเครื่องมือทางสถิติประมาณค่า) เพื่อให้สามารถตัดสินใจได้อย่างมีหลักการ จึงจำเป็นจะต้องพยายามประมาณค่าพารามิเตอร์ดังกล่าว
  - โดยธรรมชาติแล้ว เราไม่มีทางที่จะแน่ใจได้ว่าพารามิเตอร์ดังกล่าวมีค่าเท่าใดกันแน่ แต่เป็นไปได้ที่จะทราบการแจกแจงของพารามิเตอร์ ดังนั้น การตัดสินใจในกรณีนี้จึงต้องทำภายใต้เงื่อนไขเชิงความน่าจะเป็น ซึ่งขึ้นอยู่กับ **การแจกแจงของตัวประมาณค่า (distribution of estimators)**
  - ยกตัวอย่างเช่น เราจะตัดสินใจซื้อกองทุนรวม A ถ้าคาดว่าอัตราผลตอบแทนไม่น้อยกว่าร้อยละ 3 ดังนั้น ปัญหาทางสถิติในกรณีคือ ความน่าจะเป็นที่อัตราผลตอบแทนของกองทุน A จะมีค่าไม่น้อยกว่าร้อยละ 3 มีค่าเท่าใด?
  - วิธีการอนุมานทางสถิติ (statistical inference) แบบนี้เกี่ยวข้องกับสิ่งที่เรียกว่า **การทดสอบสมมุติฐาน (hypothesis testing)**
- **การอนุมานทางสถิติแบบอื่นๆ** ที่ไม่สามารถจัดอยู่ในสามประเภทแรกได้

นิยาม: ตัวประมาณค่า (Estimator)

Definition

กำหนดให้ $X_1, \ldots, X_n$ เป็นตัวอย่าง (sample) ที่สุ่มมาจากการแจกแจงร่วมกำหนดที่มีพารามิเตอร์ $\theta \in \Theta$ เราเรียกฟังก์ชันจำนวนจริง $\hat{\theta}\left(X_1, \ldots, X_n\right)$ ว่า ตัวประมาณค่า (estimator) ของพารามิเตอร์ θ นอกจากนี้ เราจะเรียก $\hat{\theta}\left(x_1, \ldots, x_n\right)$ ว่าค่าประมาณ (estimate) ของพารามิเตอร์ θ เมื่อประมาณค่าด้วยข้อมูล $x_1, \ldots, x_n$ ซึ่งเป็นค่าที่เกิดขึ้นจริง (realized value) ของ $X_1, \ldots, X_n$

- ในขณะเดียวกัน เรามักเรียกฟังก์ชันของตัวอย่าง $\hat{\theta}\left(X_1, \ldots, X_n\right)$ ว่าค่าสถิติ (statistic) ดังนั้น จึงอาจจะสรุปได้ว่า "**ตัวประมาณค่า (estimator) ก็คือค่าสถิติ (statistic) อย่างหนึ่ง**" นั่นเอง

นิยาม: ฟังก์ชันความน่าจะเป็นก่อนการสังเกต (Prior P.F.)

Definition

กำหนดให้ $\boldsymbol{\theta}$ คือพารามิเตอร์ที่ไม่ทราบค่าซึ่งสมมุติให้เป็นตัวแปรสุ่ม การแจกแจงก่อนการสังเกต (prior distribution) คือการแจกแจง (distribution) ของพารามิเตอร์ θ ที่กำหนดขึ้นก่อนที่จะสังเกตค่าของตัวแปรสุ่มอื่นๆ ซึ่งมักแทนด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนการสังเกต (prior p.d.f.) หรือฟังก์ชันความน่าจะเป็นก่อนการสังเกต (prior p.f.) $h(\boldsymbol{\theta})$

ตัวอย่าง: ฟังก์ชันความน่าจะเป็นก่อนการสังเกต (Prior P.F.)

Example

กำหนดให้ θ แทนความน่าจะเป็นที่จะได้ผลการโยนเหรียญเป็นหัว ซึ่งในที่นี้เป็น**พารามิเตอร์ของแบบจำลองทางสถิติ**ที่สนใจ สมมุติว่าเหรียญหนึ่งอาจจะเป็นเหรียญที่ด้านหนึ่งเป็นหัวส่วนอีกด้านหนึ่งเป็นก้อย หรืออาจจะเป็นเหรียญที่มีแต่หัวทั้งสองด้าน

- จากการสมมุติดังกล่าว θ เป็นไปได้สองค่าคือ $\theta = \frac{1}{2}$ และ $\theta = 1$
- **การแจกแจงก่อน (prior distribution)** สำหรับกรณีคือการแจกแจงของ θ คือ $h\left(\theta = \frac{1}{2}\right) = 0.6$ และ $h\left(\theta = 1\right) = 0.4$
- สังเกตว่า h ต้องรวมแล้ว**เท่ากับหนึ่ง**เพราะเป็นฟังก์ชันความน่าจะเป็นก่อนการสังเกต (prior p.f.)

นิยาม: ฟังก์ชันความน่าจะเป็นหลังการสังเกต (Posterior P.F.)

Definition

กำหนดให้ $\boldsymbol{\theta}$ คือพารามิเตอร์ที่ไม่ทราบค่าซึ่งสมมุติให้เป็นตัวแปรสุ่มและ $X_1, \ldots, X_n$ คือตัวอย่าง (sample) ซึ่งพิจารณาว่าเป็นตัวแปรสุ่ม การแจกแจงหลังการสังเกต (posterior distribution) คือ การแจกแจงแบบมีเงื่อนไขของ $\boldsymbol{\theta}$ เมื่อทราบค่า $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ ซึ่งมักแทนด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นหลังการสังเกต (posterior p.d.f.) หรือฟังก์ชันความน่าจะเป็นหลังการสังเกต (posterior p.f.) $h\left(\boldsymbol{\theta}|\mathbf{x}\right)$

- สังเกตว่าความแตกต่างเชิงสัญลักษณ์ระหว่างการแจกแจงก่อนการสังเกต (prior distribution) $h\left(\boldsymbol{\theta}\right)$ และการแจกแจงหลังการสังเกต (posterior distribution) $h\left(\boldsymbol{\theta}|\mathbf{x}\right)$ คือ "**การที่อันแรกไม่มีเงื่อนไขแต่อันหลังเป็นการแจกแจงแบบมีเงื่อนไข**"

ทฤษฎี: การแปลงการแจกแจงก่อนการสังเกต (prior distribution) ให้เป็นการแจกแจงหลังการสังเกต (posterior distribution)

- ทฤษฎีบทต่อไปนี้ประยุกต์ใช้ ทฤษฎีบทของเบส์ เพื่อ**แปลง**การแจกแจงก่อนการสังเกต (prior distribution) ให้เป็นการแจกแจงหลังการสังเกต (posterior distribution) โดยอาศัยสารสนเทศ (information) ที่ได้จากข้อมูล (data) $\boldsymbol{x} = (x_1, \ldots, x_n)$
- ประเด็นทางเทคนิคที่สำคัญในที่นี้คือหลักการที่ว่า ข้อมูลที่ได้มานั้นมาจากการสุ่มจากการแจกแจงแบบมีเงื่อนไข (conditional distribution) ที่ถูกกำหนดโดยพารามิเตอร์ $\boldsymbol{\theta}$ ซึ่งแทนด้วย $f(\boldsymbol{x}|\boldsymbol{\theta})$

Theorem

กำหนดให้ $X_1, \ldots, X_n$ คือตัวอย่างสุ่มที่เกิดจากการสุ่มเลือกจากการแจกแจง $f(\boldsymbol{x}|\theta)$ และกำหนดให้ $h(\theta)$ แทนการแจกแจงก่อนการสังเกต (prior distribution) ของ θ แล้ว ฟังก์ชันความหนาแน่นของความน่าจะเป็นหลังการสังเกต (posterior p.d.f.) หรือฟังก์ชันความน่าจะเป็นหลังการสังเกต (posterior p.f.) เท่ากับ

$$h(\theta|\boldsymbol{x}) = \frac{f(x_1|\theta)\cdots f(x_n|\theta)\, h(\theta)}{\int_{\tilde{\theta}\in\Theta} f\left(\boldsymbol{x}|\tilde{\theta}\right) h\left(\tilde{\theta}\right) d\tilde{\theta}}, \quad \text{สำหรับ } \theta \in \Theta \tag{1}$$

การพิสูจน์

Proof.

- จากนิยามของการแจกแจงแบบมีเงื่อนไข เราสามารถเขียนได้ว่า

$$h\left(\theta|\mathbf{x}\right)=\frac{f\left(\mathbf{x},\theta\right)}{g\left(\mathbf{x}\right)},\ \text{สำหรับ}\ \theta\in\Theta$$

  โดยที่ $g\left(\mathbf{x}\right)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นตามขอบ (marginal p.d.f.) ซึ่งมีค่าเท่ากับ

$$g\left(\mathbf{x}\right)=\int_{\tilde{\theta}\in\Theta}f\left(\mathbf{x}|\tilde{\theta}\right)h\left(\tilde{\theta}\right)d\tilde{\theta}$$

  ส่วนตัวตั้งสามารถเขียนใหม่ได้เป็น

$$f\left(\mathbf{x},\theta\right)=f\left(\mathbf{x}|\theta\right)h\left(\theta\right)$$

- ในขณะเดียวกัน การที่ $X_1,\ldots,X_n$ เป็นตัวอย่างสุ่ม ช่วยให้สามารถเขียน $f\left(\mathbf{x}|\theta\right)$ ได้เป็น

$$f\left(\mathbf{x}|\theta\right)=f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)$$

- ดังนั้น จึงสามารถสรุปได้ว่า

$$h\left(\theta|\mathbf{x}\right)=\frac{f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right)h\left(\theta\right)}{\int_{\tilde{\theta}\in\Theta}f\left(\mathbf{x}|\tilde{\theta}\right)h\left(\tilde{\theta}\right)d\tilde{\theta}},\ \text{สำหรับ}\ \theta\in\Theta$$

การแปรผันตามส่วน (proportionality) ของการแจกแจงหลังการสังเกต (posterior distribution)

- ประเด็นที่น่าสังเกตอันหนึ่งคือ **ตัวหารในสมการที่ 1 ไม่ขึ้นอยู่กับพารามิเตอร์ θ** เพราะได้อินทิเกรต θ ออกไปหมดแล้วและอันที่จริงพจน์นี้ทำหน้าที่เป็นเพียงแค่ค่าคงที่ที่ทำให้การแจกแจงหลังการสังเกต (posterior distribution) มีคุณสมบัติที่เหมาะสม ซึ่งในที่นี้หมายถึงการที่ผลรวมหรือผลการอินทิเกรตของการแจกแจงหลังการสังเกต (posterior distribution) มีค่าเท่ากับหนึ่งนั่นเอง
- บางครั้งเราอาจจะมองข้ามพจน์นี้ไปได้และเขียน $h\left(\theta|\boldsymbol{x}\right)$ ในรูปของการแปรผันตามส่วน (proportionality) ได้เป็น

$$h\left(\theta|\boldsymbol{x}\right) \propto f\left(x_1|\theta\right)\cdots f\left(x_n|\theta\right) h\left(\theta\right) \tag{2}$$

ตัวอย่าง: การประยุกต์การแปรผันตามส่วนของการแจกแจงหลังการสังเกต (posterior distribution)

Example

กำหนดให้ θ แทนสัดส่วนของสินค้าที่มีตำหนิ (defective items) ที่ยังไม่ทราบค่า และการแจกแจงก่อนการสังเกต (prior distribution) เป็นแบบเอกรูป (uniform distribution) ในช่วง $[0, 1]$ สิ่งที่ต้องการทราบคือการแจกแจงหลังการสังเกต (posterior distribution) ของ θ หลังจากสุ่มตรวจสินค้าทั้งหมด n ชิ้น

- กำหนดให้ X_i แทนผลการตรวจสินค้าชิ้นที่ $i = 1, \ldots, n$ โดยที่ $X_i = 1$ ถ้าสินค้าที่ i มีตำหนิ ไม่เช่นนั้นจะมีค่าเท่ากับศูนย์ นั่นคือ X_i มีการแจกแจงแบบเบอร์นูลลี ซึ่งมีฟังก์ชันความน่าจะเป็น (p.f.) เท่ากับ

$$f(x_i|\theta) = \begin{cases} \theta^{x_i} (1-\theta)^{1-x_i}, & \text{สำหรับ } x_i = 0, 1 \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

- การที่ตัวอย่างที่ได้มาเป็นตัวอย่างสุ่ม (random sample) ทำให้สามารถเขียนได้ว่า

$$f(x_1|\theta) \cdots f(x_n|\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} = \theta^y (1-\theta)^{n-y}$$

  โดยที่ $y = \sum_{i=1}^n x_i$ ส่วนการที่การแจกแจงก่อนการสังเกต (prior distribution) เป็นแบบเอกรูป (uniform distribution) ในช่วง $[0, 1]$ ทำให้เขียนได้ว่า $h(\theta) = 1$

วิธีทำ: การประยุกต์การแปรผันตามส่วนของการแจกแจงหลังการสังเกต (posterior distribution)

Example

- ดังนั้น การแจกแจงหลังการสังเกต (posterior distribution) ของ θ เท่ากับ

$$h(\theta|\mathbf{x}) \propto \theta^{y}(1-\theta)^{n-y}$$

- เมื่อพิจารณาให้ดีแล้วจะเห็นว่าพจน์ด้านขวานั้นเป็นส่วนหนึ่งของการแจกแจงเบต้า (beta distribution) ของ θ ที่มีค่าพารามิเตอร์ $\alpha = y+1$ และ $\beta = n-y+1$ ดังนั้น จึงสามารถกำหนดค่าคงที่ที่ต้องการได้โดยไม่ต้องอินทิเกรต ทำให้ได้การแจกแจงหลังการสังเกต (posterior distribution) ของ θ เป็น

$$h(\theta|\mathbf{x}) = \frac{\Gamma(n+2)}{\Gamma(y+1)\,\Gamma(n-y+1)}\theta^{y}(1-\theta)^{n-y}$$

- สังเกตได้ว่า ในกรณีนี้ เราสามารถเขียนการแจกแจงร่วมในรูปของค่าสถิติ $y = \sum_{i=1}^{n} x_i$ นั่นคือ เราไม่จำเป็นต้องทราบค่า x_i แต่ละค่า สิ่งเดียวที่ต้องการทราบคือผลรวมหรือจำนวนสินค้าที่มีตำหนิเท่านั้น
- หลักการนี้ช่วยให้เราวิเคราะห์ปัญหาที่สนใจได้สะดวกขึ้นมากเพราะค่าผลรวมนี้เป็นค่าสถิติที่รวบรวมเอาสารสนเทศ (information) ทั้งหมดจาก $\mathbf{x}$ ที่จำเป็นในการวิเคราะห์ปัญหานี้ไว้อย่างครบถ้วน เราจะอภิปรายประเด็นนี้อย่างละเอียดในหัวข้อค่าสถิติที่เพียงพอ (sufficient statistics)

การแปลงการแจกแจงได้ด้วยการปรับเป็นลำดับ (sequential updating)

- ในกรณีที่มีตัวอย่างสุ่มมากกว่าหนึ่งตัวอย่าง เราสามารถแปลงการแจกแจงได้ด้วยการปรับเป็นลำดับ (sequential updating) โดยเริ่มการแจกแจงหลังการสังเกต (posterior distribution) ซึ่งเป็นผลจากการสังเกตุ X_1

$$h(\theta|x_1) \propto f(x_1|\theta)\, h(\theta) \tag{3}$$

- หลังจากนั้นจึงคำนวณหาการแจกแจงหลังการสังเกต (posterior distribution) จาก X_1, X_2 โดยใช้ $h(\theta|x_1)$ เป็นการแจกแจงก่อนการสังเกต (prior distribution)

$$h(\theta|x_1, x_2) \propto f(x_2|\theta)\, h(\theta|x_1) \propto f(x_1|\theta) f(x_2|\theta)\, h(\theta) \tag{4}$$

- ดังนั้น การดำเนินการในรูปแบบนี้ต่อไปจนครบทั้งหมด n ครั้งก็จะได้การแจกแจงหลังการสังเกต (posterior distribution)

$$h(\theta|\mathbf{x}) \propto f(x_1|\theta) \cdots f(x_n|\theta)\, h(\theta) = f(\mathbf{x}|\theta)\, h(\theta) \tag{5}$$

- ตรงกับการแจกแจงหลังการสังเกต (posterior distribution) ที่แสดงในสมการ (1) ซึ่งได้จากการปรับเพียงโดยใช้ข้อมูล $X_1, \ldots, X_n$ เพียงครั้งเดียว

ข้อสรุป: การแปลงการแจกแจงได้ด้วยการปรับเป็นลำดับ (sequential updating)

- อย่างไรก็ตาม การปรับเป็นลำดับ (sequential updating) นี้มีความสำคัญอย่างมากในโลกของข้อมูลขนาดใหญ่ (big data) เพราะ **พัฒนาการทางเทคโนโลยีทำให้มีการเพิ่มข้อมูลตลอดเวลา**
- ในขณะเดียวกัน ก็มีความต้องการพยากรณ์ที่ทันท่วงที ซึ่งในที่นี้สามารถทำได้โดยการปรับการแจกแจงทุกครั้งที่มีการรับข้อมูลเข้ามาใหม่ ด้วยหลักการปรับเป็นลำดับ (sequential updating) ซึ่งเป็น **พื้นฐานสำคัญอันหนึ่งของรูปแบบการเรียนรู้ของเครื่องจักร (machine learning)**
- โดยทั่วไป การแจกแจงหลังการสังเกต (posterior distribution) จะมี **ความแตกต่าง** จากการแจกแจงก่อนการสังเกต (prior distribution) โดยสิ้นเชิง
  - ในตัวอย่างที่ผ่านมา เริ่มจากการแจกแจงเอกรูป (uniform distribution) แต่ได้การแจกแจงหลังการสังเกตที่เป็นการแจกแจงแบบเบต้า (beta distribution)
- บางครั้งมีการแจกแจงบางรูปแบบที่ถ้าเริ่มจากการแจกแจงก่อนการสังเกต (prior distribution) แล้วยังคงทำให้ การแจกแจงหลังการสังเกต (posterior distribution) นั้น ยังคง มีการแจกแจงเหมือนเดิม เราเรียกการแจกแจงที่มีคุณสมบัติแบบนี้ว่า "**การแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution)**"
  - ตัวอย่างที่สำคัญอันหนึ่งของการแจกแจงก่อนการสังเกตคู่คือ การแจกแจงปกติ (normal distribution)
- นอกจากนี้ เครื่องมือที่นิยมใช้ในการปรับการแจกแจงหรือปรับความเชื่อ (updating beliefs) ในทางเศรษฐศาสตร์และการเงินคือการกรองแบบคาลแมน (Kalman filtering)

ทฤษฎี: การแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution) ของการแจกแจงปกติ (normal distribution)

Theorem

สมมุติว่า $X_1, \ldots, X_n$ คือตัวอย่างสุ่มที่สุ่มเลือกมาจากการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ ซึ่งไม่ทราบค่า และค่าความแปรปรวนเท่ากับ σ_x^2 ซึ่งทราบค่า และสมมุติว่าการแจกแจงก่อนการสังเกต (prior distribution) ของ μ เป็นการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_0 และค่าความแปรปรวนเท่ากับ σ_0^2 แล้วการแจกแจงหลังการสังเกต (posterior distribution) ของ μ หลังจากทราบค่าของ $X_1, \ldots, X_n$ เป็นการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_1 และค่าความแปรปรวนเท่ากับ σ_1^2 โดยที่

$$\mu_1 = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{x}_n \tag{6}$$

$$\sigma_1^2 = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\sigma_0^2 \tag{7}$$

การพิสูจน์: การแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution) ของการแจกแจงปกติ (normal distribution)

Proof.

- พิจารณาฟังก์ชันความเป็นไปได้ (likelihood function)

$$f(\mathbf{x}|\mu) \;\propto\; \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

- เนื่องจากสิ่งที่เราต้องการจริงๆ คือ รูปแบบของการแจกแจงหลังการสังเกต (posterior distribution) ที่เกี่ยวข้องกับ μ ดังนั้น เราจึงสามารถที่จะละเลยพจน์ที่แยกออกไปและที่ไม่เกี่ยวข้องกับ μ ได้โดยไม่ส่งผลเสียต่อสิ่งที่ต้องการพิสูจน์ โดยเริ่มจาก

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}((x_i-\bar{x}_n)+(\bar{x}_n-\mu))^2 &= \sum_{i=1}^{n}(x_i-\bar{x}_n)^2+\sum_{i=1}^{n}(\bar{x}_n-\mu)^2 \\
&+(\bar{x}_n-\mu)\sum_{i=1}^{n}(x_i-\bar{x}_n) \\
&= \sum_{i=1}^{n}(\mu-\bar{x}_n)^2+\sum_{i=1}^{n}(x_i-\bar{x}_n)^2
\end{aligned}$$

□

การพิสูจน์: การแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution) ของการแจกแจงปกติ (normal distribution) Con't

Proof.

- เนื่องจากพจน์ด้านขวาไม่ขึ้นอยู่กับ μ ดังนั้น เราสามารถสรุปได้ว่าฟังก์ชันความเป็นไปได้ (likelihood function)

$$f(\mathbf{x}|\mu) \;\propto\; \exp\left[-\frac{1}{2\sigma^2} n\left(\mu - \bar{x}_n\right)^2\right]$$

- ในขณะเดียวกัน การแจกแจงก่อนการสังเกต (prior distribution) ของ μ เขียนได้เป็น

$$h(\mu) \;\propto\; \exp\left[-\frac{1}{2\sigma_0^2}\left(\mu - \mu_0\right)^2\right]$$

- ดังนั้น การแจกแจงหลังการสังเกต (posterior distribution) เท่ากับ

$$\begin{aligned} h(\theta|\mathbf{x}) \;&\propto\; \exp\left[-\frac{n}{2\sigma^2}\left(\mu - \bar{x}_n\right)^2\right] \exp\left[-\frac{1}{2\sigma_0^2}\left(\mu - \mu_0\right)^2\right] \\ &= \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}\left(\mu - \bar{x}_n\right)^2 + \frac{1}{\sigma_0^2}\left(\mu - \mu_0\right)^2\right]\right\} \end{aligned}$$

□

การพิสูจน์การแจกแจงก่อนการสังเกตคู่ (conjugate prior distribution) ของการแจกแจงปกติ (normal distribution) Con't

Proof.

- ขั้นตอนต่อไปคือการแยกพจน์ที่ไม่เกี่ยวกับ μ ออกโดยการจัดรูป

$$\begin{aligned}
\frac{n}{\sigma^2}(\mu-\bar{x}_n)^2+\frac{1}{\sigma_0^2}(\mu-\mu_0)^2 &= \frac{n}{\sigma^2}\left(\mu^2-2\mu\bar{x}_n+\bar{x}_n^2\right)+\frac{1}{\sigma_0^2}\left(\mu^2-2\mu\mu_0+\mu_0^2\right)\\
&= \frac{1}{\sigma^2\sigma_0^2}\left[\left(\sigma^2+n\sigma_0^2\right)\mu^2-2\left(\sigma^2\mu_0+n\sigma_0^2\bar{x}_n\right)\mu\right]+\frac{1}{\sigma^2\sigma_0^2}\left[\sigma^2\mu_0^2+n\sigma_0^2\bar{x}_n^2\right]\\
&= \frac{\left(\sigma^2+n\sigma_0^2\right)}{\sigma^2\sigma_0^2}\left[\mu^2-2\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)\mu+\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)^2\right]\\
&\quad+\frac{1}{\sigma^2\sigma_0^2}\left[\sigma^2\mu_0^2+n\sigma_0^2\bar{x}_n^2-\left(\sigma^2+n\sigma_0^2\right)\left(\frac{\sigma^2\mu_0+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)^2\right]\\
&= \frac{1}{\sigma_1^2}(\mu-\mu_1)^2+\frac{n}{\sigma^2+n\sigma_0^2}(\bar{x}_n-\mu_0)^2
\end{aligned}$$

โดยที่ $\mu_1=\frac{\sigma_x^2}{\sigma_x^2+n\sigma_0^2}\mu_0+\frac{n\sigma_0^2}{\sigma_x^2+n\sigma_0^2}\bar{x}_n$ และ $\sigma_1^2=\frac{\left(\sigma^2+n\sigma_0^2\right)}{\sigma^2\sigma_0^2}$ เนื่องจากพจน์ด้านขวาไม่เกี่ยวข้อง μ

- ดังนั้นเราสามารถสรุปได้ว่า

$$h(\theta|\mathbf{x}) \propto \exp\left[-\frac{1}{\sigma_1^2}(\mu-\mu_1)^2\right]$$

ตัวอย่าง: การประมาณค่าแบบเบส์ (Bayes Estimator) ของความเสี่ยงเชิงระบบ (systematic risks) β

Example

พิจารณากองทุน หลักทรัพย์ AIT ในช่วงปี 2016 และสมมุติว่าความเสี่ยงเชิงระบบของหลักทรัพย์มีการแจกแจงปกติ (normal distribution) ที่มีความแปรปรวนคือ 14.56 ($\sigma_\beta^2 = 14.66$) เราสามารถหาได้ว่า ความเสี่ยงเชิงระบบในช่วงปี 2016 คือ 0.91 ($\beta_{2016} = 0.91$) ความแปรปรวนของความเสี่ยงเชิงระบบในช่วงปี 2016 คือ 0.04 ($\sigma_{2016}^2 = 0.04$) และ เมื่อใช้อัตราผลตอบแทนสุทธิ์รายสัปดาห์ทั้งสิ้น 52 สัปดาห์ ($n = 52$) และกำหนดให้ ค่าเฉลี่ยของความเสี่ยงเชิงระบบในช่วงปี 2016 ของทุกหลักทรัพย์ที่อยู่ในตลาดคือ 0.77 ($\bar{\beta} = 0.77$) จงหา ความเสี่ยงเชิงระบบและความแปรปรวนของความเสี่ยงเชิงระบบในช่วงปี 2017 (วิธีการประมาณค่า β แบบนี้เริ่มมาจากงานวิจัยของ Vasicek (1973))

- เราสามารถหา ความเสี่ยงเชิงระบบและความแปรปรวนของความเสี่ยงเชิงระบบในช่วงปี 2017 ได้จากทฤษฎีก่อนหน้านี้
  - ความเสี่ยงเชิงระบบในช่วงปี 2017

$$\beta_{2017} = \frac{\sigma_\beta^2}{\sigma_\beta^2 + n\sigma_{2016}^2}\beta_{2016} + \frac{n\sigma_{2016}^2}{\sigma_\beta^2 + n\sigma_{2016}^2}\bar{\beta} \tag{8}$$

$$= \frac{14.66}{14.66 + 52 \times 0.04} \times 0.91 + \frac{0.04}{14.66 + 52 \times 0.04} \times 0.77 \approx 0.893 \tag{9}$$

  - ความแปรปรวนของความเสี่ยงเชิงระบบในช่วงปี 2017

$$\sigma_{2017}^2 = \frac{\sigma_\beta^2}{\sigma_\beta^2 + n\sigma_{2016}^2}\sigma_{2016}^2 \tag{10}$$

$$= \frac{14.66}{14.66 + 52 \times 0.04} \times 0.04 \approx 0.035 \tag{11}$$

นิยาม: ฟังก์ชันสูญเสีย (loss function)

- ตัวประมาณค่าแบบเบส์ (Bayes Estimator) ของ θ คือตัวประมาณค่า (estimator) $\hat{\theta}\left(X_1, \ldots, X_n\right)$ ซึ่งทำให้ค่าคาดหมายของฟังก์ชันสูญเสีย (loss function) ที่คำนวณโดยใช้ **การแจกแจงหลังการสังเกต (posterior distribution) มีค่าต่ำที่สุด**"

Definition

ฟังก์ชันสูญเสีย (loss function) หมายถึงฟังก์ชันของสองกลุ่มตัวแปร $L\left(\theta, a\right)$ ซึ่งตีความว่าเป็นการสูญเสียทางสถิติถ้าพารามิเตอร์มีค่าเท่ากับ θ แต่ตัวประมาณค่ามีค่าเท่ากับ a

- ในทางปฏิบัติ เรามักจะใช้ "**ค่าคาดหมายของการสูญเสีย (expected loss)** "

$$E\left[L\left(\theta, a\right)|\mathbf{x}\right] = \int_{\Theta} L\left(\theta, a\right) h\left(\theta|\mathbf{x}\right) d\theta \tag{12}$$

เป็นฟังก์ชันเป้าหมาย (objective function) ที่ใช้ในการเลือกตัวประมาณค่าแบบเบส์ (Bayes Estimator)

ตัวประมาณค่าแบบเบส์ (Bayes Estimator)

Definition

กำหนดให้ $L(\theta, a)$ แทนฟังก์ชันสูญเสีย (loss function) และ $\hat{\theta}(x)$ เป็นคำตอบของปัญหาการหาค่าต่ำสุด (minimization problem) ต่อไปนี้

$$E\left[L\left(\theta, \hat{\theta}(x)\right)\right] = \min_{a} E\left[L(\theta, a) \mid x\right] \tag{13}$$

แล้ว ฟังก์ชัน $\hat{\theta}(X)$ คือตัวประมาณค่าแบบเบส์ (Bayes Estimator) ส่วนค่า $\hat{\theta}(x)$ คือค่าประมาณแบบเบส์ (Bayes estimate) ของ θ เมื่อข้อมูลที่ใช้ในการประมาณค่าคือ x

- ตัวประมาณค่าแบบเบส์ (Bayes Estimator) ขึ้นอยู่กับรูปแบบค่าคาดหมายของการสูญเสีย (expected loss) ซึ่งเป็นฟังก์ชันเป้าหมาย (objective function)
- กล่าวอีกนัยหนึ่งได้ว่า การกำหนดรูปแบบฟังก์ชันสูญเสียที่ **แตกต่างกัน** ย่อมนำไปสู่ตัวประมาณค่าที่ **แตกต่างกัน**
- ในขณะเดียวกัน ก็ **ไม่มีทฤษฎีที่บอกได้ว่าควรจะใช้ฟังก์ชันสูญเสียแบบใดดี** ดังนั้น จึงเป็นหน้าที่ของนักวิเคราะห์ที่จะต้องเลือกฟังก์ชันสูญเสียให้เหมาะสม ซึ่งอาจจะต้องอาศัยประสบการณ์เป็นสำคัญ

นิยาม: ฟังก์ชันสูญเสียกำลังสอง (square error loss function)

Definition

ฟังก์ชันสูญเสียกำลังสอง (square error loss function) นิยามได้เป็น

$$L(\theta, a) = (\theta - a)^2 \tag{14}$$

- ทฤษฎีบทต่อไปนี้ระบุว่า ค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) คือตัวประมาณค่าแบบเบส์ (Bayes Estimator) ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบกำลังสอง (square error loss function)

ตัวประมาณค่าแบบเบส์ (Bayes Estimator) เมื่อใช้ ฟังก์ชันสูญเสียกำลังสอง (square error loss function)

Theorem

สมมุติว่าฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบฟังก์ชันสูญเสียกำลังสอง (square error loss function) ดังแสดงในสมการที่ 14 แล้ว ตัวประมาณค่าแบบเบส์ (Bayes Estimator)

$$\hat{\theta}(X) = E[\theta|X] = \int_{\Theta} \theta h(\theta|X)\, d\theta \tag{15}$$

สังเกตว่า ค่าคาดหมายในที่นี้คำนวณจากการแจกแจงหลังการสังเกต (posterior distribution) ของ θ

Proof.

พิจารณาปัญหาการหาค่าคาดหมายของการสูญเสีย (expected loss) ที่ต่ำที่สุดดังต่อไปนี้

$$\min_a E[L(\theta, a)|\mathbf{x}] = \min_a \int_{\Theta} (\theta - a)^2 h(\theta|\mathbf{x})\, d\theta$$

เงื่อนไขอันดับที่หนึ่ง (first-order condition) สำหรับปัญหานี้คือ

$$\left.\frac{\partial E[L(\theta, a)|\mathbf{x}]}{\partial a}\right|_{a=\hat{\theta}(\mathbf{x})} = 0$$

$$\int_{\Theta} \left.\frac{\partial(\theta - a)^2}{\partial a} h(\theta|\mathbf{x})\, d\theta\right|_{a=\hat{\theta}(\mathbf{x})} = -2\int_{\Theta}\left(\theta - \hat{\theta}(\mathbf{x})\right) h(\theta|\mathbf{x})\, d\theta = 0 \Rightarrow \hat{\theta}(\mathbf{x}) = \int_{\Theta} \theta h(\theta|\mathbf{x})\, d\theta = E[\theta|\mathbf{x}]$$

ตัวอย่าง: ฟังก์ชันสูญเสียกำลังสอง (square error loss function)

Example

พิจารณาตัวอย่างสุ่ม $X_1, \ldots, X_n$ ซึ่งสุ่มเลือกมาจากการแจกแจงปกติ (normal distribution) ที่มีค่าคาดหมาย μ และค่าความแปรปรวน σ^2 สมมุติว่าเราทราบค่าความแปรปรวน แต่ไม่ทราบค่าคาดหมาย ดังนั้น จึงต้องการประมาณค่าคาดหมาย μ จากข้อมูลที่มีอยู่ด้วยการประมาณค่าแบบเบส์ (Bayes estimation) สมมุติอีกว่า การแจกแจงก่อนการสังเกต (prior distribution) ของ μ เป็นการแจกแจงแบบปกติ (normal distribution) ที่มีค่าคาดหมายเท่ากับ μ_0 และค่าความแปรปรวนเท่ากับ σ_0

- ทฤษฎีบทล่าสุดระบุว่าตัวประมาณค่าของเบส์ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบกำลังสองคือค่าคาดหมายแบบมีเงื่อนไข (conditional expectation) นั่นคือ

$$\hat{\theta}(X) = E[\theta|X] = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{X}_n$$

- ทฤษฎีบทต่อไปนี้แสดงถึงตัวประมาณค่าแบบเบส์ (Bayes Estimator) ในกรณีที่ฟังก์ชันสูญเสียเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function)

นิยาม: ฟังก์ชันสูญเสียแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function)

Definition

ฟังก์ชันสูญเสียแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) นิยามได้เป็น

$$L(\theta, a) = |\theta - a| \tag{16}$$

Theorem

สมมุติว่าฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) ดังแสดงในสมการที่ 16 แล้ว ตัวประมาณค่าแบบเบส์ (Bayes Estimator) มีค่าเท่ากับ ***ค่ามัธยฐาน (median)*** *ของการแจกแจงหลังการสังเกต (posterior distribution) ของ* θ

- โดยทั่วไป ฟังก์ชันสูญเสียที่แตกต่างกันมักจะนำไปสู่ตัวประมาณค่าที่ต่างกันดังแสดงในทฤษฎีบทที่แล้ว
- ค่าประมาณ (estimate) ที่ได้จากตัวประมาณที่แตกต่างกันมีค่าเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบการแจกแจงหลังการสังเกต

ตัวอย่าง: ที่มีค่าคาดหมายและค่ามัธยฐานมีค่าเท่ากัน

Example

พิจารณาสถานการณ์ที่เหมือนกับตัวอย่างที่ 17 แต่คราวนี้กำหนดให้ฟังก์ชันสูญเสียที่ใช้สำหรับการประมาณค่าเป็นแบบค่าสัมบูรณ์ของผลต่าง (absolute error loss function) ดังนั้น จากทฤษฎีบทที่ 19 ตัวประมาณค่าแบบเบส์ (Bayes Estimator) มีค่าเท่ากับค่ามัธยฐาน (median) ของการแจกแจงหลังการสังเกต (posterior distribution) ของ μ ซึ่งในที่นี้จะมีค่าเท่ากับค่าคาดหมาย นั่นคือ

$$\hat{\theta}(X) = E[\theta|X] = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{X}_n$$

คุณสมบัติ: ความคงเส้นคงวาของตัวประมาณค่าแบบเบส์เมื่อตัวอย่างมีขนาดใหญ่

- **ความคงเส้นคงวา (consistency)** เป็นคุณสมบัติการลู่เข้าของตัวประมาณค่าที่เกิดจากการเพิ่มขึ้นของ ขนาดตัวอย่าง (sample size) จนมีขนาดเข้าใกล้อนันต์ ($n \to \infty$) โดยใช้หลักการลู่เข้าเชิงความน่าจะเป็น (convergence in probability) เป็นเครื่องมือในการวิเคราะห์
- ถึงแม้ว่าในโลกความจริง เราจะไม่เคยมีข้อมูลขนาดอนันต์ แต่หลักการความคงเส้นคงวา (consistency) ก็เป็นเครื่องมือทางสถิติที่มีประโยชน์และสะดวกในการใช้งาน โดยที่บางครั้งอาจจะเป็นเครื่องมือเดียวที่สามารถบอกถึงความแม่นยำของตัวประมาณค่า เพราะไม่สามารถพิสูจน์ในกรณีที่มีตัวอย่างจำกัดได้ (finite sample)
- คุณสมบัติของการลู่เข้าเชิงความน่าจะเป็น (convergence in probability) สามารถส่งผ่านฟังก์ชันที่ต่อเนื่องใดๆ ในขณะที่หลักการหาค่าคาดหมาย (expectation) สามารถส่งผ่านได้เพียงฟังก์ชันเชิงเส้นเท่านั้น

คุณสมบัติ: ความคงเส้นคงวาของตัวประมาณ

Definition

ตัวประมาณค่า $\hat{\theta}$ ของพารามิเตอร์ θ มีความคงเส้นคงว่า (consistent) ถ้า

$$\hat{\theta} \xrightarrow{p} \theta \tag{17}$$

- คำถามคือ **ตัวประมาณค่าแบบเบส์ (Bayes estimator) มีความคงเส้นคงวา (consistent)** หรือไม่?
- คำตอบคือ ภายใต้เงื่อนไขที่ค่อนข้างมาตรฐาน ตัวประมาณค่าแบบเบส์ (Bayes estimator) **มีความคงเส้นคงวา (consistent)**
- แต่การพิสูจน์ความคงเส้นคงวาของตัวประมาณค่าแบบเบส์ (Bayes estimator) อยู่เหนือขอบเขตของวิชานี้ เพราะจำเป็นต้องใช้เทคนิคขั้นสูงของ **ทฤษฎีการวัด (measure theory)**
- แต่อย่างไรก็ตาม ยังสามารถแสดงให้เห็นถึงความคงเส้นคงวา (consistency) ของตัวประมาณค่าแบบเบส์ (Bayes estimator) ได้อย่างไม่ยากเย็น โดยใช้ตัวอย่างต่อไปนี้

ตัวอย่าง: ความคงเส้นคงวาของตัวประมาณ

Example

พิจารณาสถานการณ์ที่เหมือนกับตัวอย่างที่แล้วดังนั้น ตัวประมาณค่าแบบเบส์ (Bayes estimator) เท่ากับ

$$\hat{\theta}(X) = \frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\bar{X}_n$$

- ขั้นตอนต่อไปคือ การตรวจสอบว่า $\hat{\theta}(X)$ ลู่เข้าเชิงความน่าจะเป็นสู่ μ หรือไม่?

$$\begin{aligned} plim_{n\to\infty}\hat{\theta}(X) &= \left[\lim_{n\to\infty}\frac{\sigma_x^2}{\sigma_x^2 + n\sigma_0^2}\right]\mu_0 + \left[\lim_{n\to\infty}\frac{n\sigma_0^2}{\sigma_x^2 + n\sigma_0^2}\right]\left[plim_{n\to\infty}\bar{X}_n\right] \\ &= [0]\,\mu_0 + [1]\left[plim_{n\to\infty}\bar{X}_n\right] = \mu \end{aligned}$$

- โดยที่สมการสุดท้ายเป็นผลมาจากกฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers)

สรุป: คุณสมบัติความคงเส้นคงวาของตัวประมาณค่าแบบเบส์เมื่อตัวอย่างมีขนาดใหญ่

- บทเรียนอีกอย่างหนึ่งที่ได้จากตัวอย่างนี้คือ โดยทั่วไป การแจกแจงก่อนการสังเกต (prior distribution) ที่ **แตกต่างกัน** มักนำไปสู่การแจกแจงหลังการสังเกต (posterior distribution) ที่ **แตกต่างกัน** ซึ่งส่งผลให้ได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) ที่ **แตกต่างกัน** ด้วย
- เมื่อ **ตัวอย่างมีขนาดใหญ่มากพอ** ผลของการแจกแจงก่อนการสังเกต (prior distribution) ดังกล่าว **จะหมดไป** ทำให้ได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) **เหมือนกัน** ไม่ว่าจะเริ่มด้วยการแจกแจงก่อนการสังเกต (prior distribution) แบบใด
- หากพิจารณาจากมุมมองของการปรับเป็นลำดับ (sequential updating) อาจจะสรุปได้ว่า เมื่อเราปรับการแจกแจง (updating distribution) ไปเรื่อยๆ ก็จะได้ตัวประมาณค่าแบบเบส์ (Bayes estimator) ที่เหมือนกัน ไม่ว่าจะเริ่มด้วยการแจกแจงก่อนการสังเกต (prior distribution) แบบใดก็ตาม

ตัวอย่าง: คุณสมบัติความคงเส้นคงวาของตัวประมาณค่าแบบเบส์ เมื่อมีการแจกแจงก่อนการสังเกต (prior distribution) 2 การแจกแจง

Example

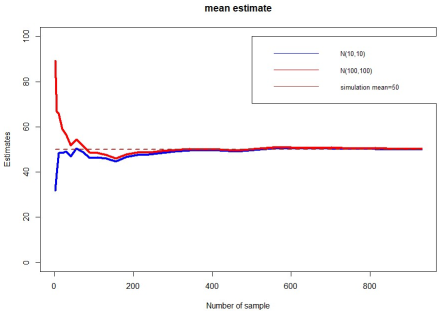
ตัวอย่างต่อไปนี้แสดงการประมาณค่าแบบเบส์ (Bayes estimation) จากการแจกแจงก่อนการสังเกต (prior distribution) ที่แตกต่างกันสองอันคือ

- อันแรกเป็นการแจกแจงปกติที่มีค่าคาดหมาย $\mu_0 = 10$ และค่าความแปรปรวน $\sigma_0^2 = 10$
- ส่วนอันที่สองเป็นการแจกแจงปกติที่มีค่าคาดหมาย $\mu_0 = 100$ และค่าความแปรปรวน $\sigma_0^2 = 100$

ในขณะที่ กลุ่มตัวอย่างที่ใช้ในตัวอย่างนี้สุ่มเลือกมาจากการแจกแจงปกติที่มีค่าคาดหมาย $\mu = 50$ และค่าความแปรปรวน $\sigma^2 = 50$ (ด้วยการจำลอง (simulation) ในคอมพิวเตอร์)

ตัวอย่าง: คุณสมบัติความคงเส้นคงวาของตัวประมาณค่าแบบเบส์ เมื่อมีการแจกแจงก่อนการสังเกต (prior distribution) 2 การแจกแจง

- รูปแสดงค่าประมาณแบบเบส์ (Bayes estimate) สำหรับการแจกแจงก่อนการสังเกต (prior distribution) ทั้งสองการแจกแจง



- บทเรียนที่สำคัญจากตัวอย่างนี้คือ
  - ในช่วงแรกที่ขนาดของตัวอย่างยัง **ไม่มากนัก** ค่าประมาณแบบเบส์ (Bayes estimate) ที่ได้จากการแจกแจงก่อนการสังเกต (prior distribution) ทั้งสองอัน มี **ความแตกต่างกันอย่างชัดเจน**
  - เมื่อตัวอย่างมี **จำนวนมากพอ ความแตกต่างดังกล่าวแทบจะไม่เหลืออยู่เลย**