單元 17: 樣本期望值與變異數
(課本 §7.1)

令 θ 爲一隨機模型中欲估計的量.

問. 如何估計 θ?

答. 模擬此模型, 使得產生的隨機變數 X (稱作輸出值, output data) 滿足

$$E(X) = \theta$$

亦即, 建構一模型使得 θ 爲所生成的輸出值 X (乃一隨機變數) 的期望值.

接著, 根據強大數法則 (SLLN), 重覆此模擬 k 次, 得獨立且同分布的輸出值

$$X_1, X_2, \ldots, X_k$$

並以

$$\frac{1}{k}\sum_{i=1}^{k} X_i \approx \theta$$

其中模擬次數 k 要夠大.

問. 到底 k 爲多少, 才算夠大?

答. 由探討 θ 的估計量 (estimator)

$$\frac{1}{k}\sum_{i=1}^{k} X_i$$

的性質, 可獲知決定 k 的方法, 詳述如下.

令隨機變數

$$X_1, X_2, \ldots, X_n$$

爲獨立同分布, 且共同的期望值

$$E(X_i) = \theta < \infty$$

與變異數

$$\mathsf{Var}(X_i) = \sigma^2 < \infty$$

但均未知.

定義

$$\overline{X} \stackrel{\mathsf{def}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i$$

稱作樣本期望值, 會隨著 X_i 而改變, 故爲一隨機變數, 常用來作爲期望值 θ 的估計.

幾個基本性質爲

(i) 樣本期望值的期望值

$$E(\overline{X}) = \theta$$

亦即, 樣本期望值 $\overline{X}$ 爲 θ 的不偏估計量 (unbiased estimator).

<證> 根據期望值的線性性質, 亦即, 線性組合的期望值等於期望值的線性組合, 以及同方布, 得

$$\begin{aligned} E(\overline{X}) &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta = \theta \end{aligned}$$

得證.

(ii) 樣本期望值的變異數

$$\mathsf{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

亦即, 當 n 愈大時, $\mathsf{Var}(\overline{X})$ 愈小, 乃表示 $\overline{X}$ 愈集中在 $E(\overline{X}) = \theta$ 附近, 而導出 $\overline{X}$ 爲一個好的估計量.

<證> 首先, 根據變異數的純量乘積性質, 得

$$\begin{aligned}\mathrm{Var}(\overline{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)\end{aligned}$$

接著, 根據獨立性以及同分布, 由上式得

$$\begin{aligned}\mathrm{Var}(\overline{X}) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

得證.

問. 如何精確地描述 $\overline{X}$ 集中在 θ 的附近?

答. 一個方法是柴比雪夫不等式 (Chebyshev's inequality)

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\mathrm{Var(X)}}{\epsilon^2}$$

只要求期望值與變異數存在即可, 故適用性廣, 但比較粗糙, 爲一保守的敍述, 請自行推導代入樣本期望值 $\overline{X}$ 的結果.

另一個較常用並更準確的方法爲性質

(iii) 中央極限定理 (Central Limit Theorem, CLT). 當 $n \to \infty$,

$$\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N(0,1)$$

其中向右箭頭上的 $\mathcal{D}$ 表示分布收斂 (convergence in distribution), 亦即, 當 $n \to \infty$ 時, 對任意的實數 x,

$$P\left(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \le x\right) \to P(Z \le x) \overset{\text{表成}}{=} \Phi(x)$$

也就是說, 當 n 夠大時,

$$\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \overset{\mathcal{D}}{\approx} Z \sim N(0,1)$$

應用: 試問 $\overline{X}$ 偏離期望值 θ 在 c 個標準差, $c\frac{\sigma}{\sqrt{n}}$, 之內的機率爲何? 如圖示.

<解> 當 n 夠大時, 由 CLT,

$$\begin{aligned} P\left(|\overline{X} - \theta| \le c\frac{\sigma}{\sqrt{n}}\right) &= P\left(\left|\frac{\overline{X} - \theta}{\sigma/\sqrt{n}}\right| \le c\right) \\ &\approx P(|Z| \le c) \\ &= \Phi(c) - \Phi(-c) \\ &= 2\Phi(c) - 1 \end{aligned}$$

其中最後一個等號成立乃根據 Z 的 pdf 的對稱性, 得

$$\Phi(-c) = 1 - \Phi(c)$$

所致, 如圖示.

若 $c = 1.96$, 則透過查表或數學軟體, 得

$$\Phi(1.96) \approx 0.975$$

故

$$\begin{aligned} P\left(|\overline{X} - \theta| \le 1.96\frac{\sigma}{\sqrt{n}}\right) &\approx 2\Phi(1.96) - 1 \\ &\approx 2(0.975) - 1 \\ &= 0.95 \end{aligned}$$

亦相當於

$$P\left(|\overline{X} - \theta| > 1.96\frac{\sigma}{\sqrt{n}}\right) \approx 0.05$$

亦即, 樣本期望值 $\overline{X}$ 偏離 θ 在 1.96 個標準差之外的機率近似於 0.05.

但 σ^2 未知; 即使 n 選定了, 偏離度 $1.96\frac{\sigma}{\sqrt{n}}$ 仍未知. 因此, 仍然無法描述 $\overline{X}$ 是如何地集中在 θ 附近, 或者說無法精確地描述以 $\overline{X}$ 估計 θ 是如何地好. 所以, 需要針對變異數 $\sigma^2 = E[(X_i - \theta)^2]$ 估計, 如下述.

定義

$$S^2 \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

稱作樣本變異數, 會隨著 X_i 而改變, 故爲一隨機變數, 常用來作爲變異數 σ^2 的估計.

幾個基本性質爲

(i) 樣本變異數的期望值

$$E(S^2) = \sigma^2$$

亦即, 樣本變異數 S^2 爲 σ^2 的不偏估計量 (unbiased estimator).

<證> 首先, 經由展開及化簡, 得

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2) \\
&= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \qquad (1)
\end{aligned}$$

其中第三個等號的第二項乃根據樣本期望值 $\overline{X}$ 的定義

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

得

$$\sum_{i=1}^{n} X_i = n\overline{X}$$

所致.

接著, 根據樣本變異數的定義, (1) 式, 以及期望值的線性性質, 得

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right] \quad (2) \end{aligned}$$

又因爲 $X_1, X_2, \ldots, X_n$ 爲同分布, 故由 (2) 式, 得

$$E(S^2) = \frac{1}{n-1}\left[nE(X_1^2) - nE(\overline{X}^2)\right] \quad (3)$$

此外, 對於任一隨機變數 Y,

$$E(Y^2) = \mathsf{Var}(Y) + [E(Y)]^2$$

故由 (3) 式, 得

$$\begin{aligned} E(S^2) &= \frac{n}{n-1}\Big\{\mathsf{Var}(X_1) + [E(X_1)]^2 \\ &\qquad -\mathsf{Var}(\overline{X}) - [E(\overline{X})]^2\Big\} \\ &= \frac{n}{n-1}\left(\sigma^2 + \theta^2 - \frac{\sigma^2}{n} - \theta^2\right) \\ &= \frac{n}{n-1}\left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2 \end{aligned}$$

得證.

以

$$S \overset{\text{def}}{=} \sqrt{S^2}$$

稱作樣本標準差 (sample standard deviation), 取代 σ, 可得性質

(ii) Slutsky 定理. 在某些條件下 (自行參考 Hogg and Craig 第 5 版, 第 271 頁的註解), 當 $n \to \infty$,

$$\frac{\overline{X} - \theta}{S/\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N(0,1)$$

亦即, 當 n 夠大時,

$$\frac{\overline{X} - \theta}{S/\sqrt{n}} \overset{\mathcal{D}}{\approx} Z$$

註 1. 根據性質 (ii), 在 n 選定下,

$$S = \left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}$$

可由資料求得, 故偏離度 $1.96\frac{S}{\sqrt{n}}$ 可確定. 因此, 可以精確地描述用 $\overline{X}$ 估計 θ 是如何地好, 也就是說, 以樣本期望值 $\overline{X}$ 估計未知的期望值 θ 時, 有 95% 的機會不會偏離 $1.96\frac{S}{\sqrt{n}}$.

註 2. 反之, 若先選定誤差 (或稱偏離度, deviation), 則可決定樣本的大小 n, 如下述.

若要求的誤差 (偏離度) 爲 $1.96d$, 亦即, 要求在 95% 的機率下,

$$\text{標準差} = \frac{\text{誤差}}{1.96} = d$$

時, 選定 n 滿足

$$\frac{S}{\sqrt{n}} < d$$

則

$$P(|\overline{X} - \theta| \le 1.96d) \quad \ge \quad P\left(|\overline{X} - \theta| \le 1.96\frac{S}{\sqrt{n}}\right)$$

如圖示.

再經由標準化以及 Slutsky 定理, 由上式得

$$\begin{aligned} P(|\overline{X}-\theta| \le 1.96d) &\ge P\left(\left|\frac{\overline{X}-\theta}{S/\sqrt{n}}\right| \le 1.96\right) \\ &\approx P(|Z| \le 1.96) \\ &= 0.95 \end{aligned}$$

因此, 至少有 95% 的信心可確信, 當樣本大小 n 滿足

$$\frac{S}{\sqrt{n}} < d$$

時, 以 $\overline{X}$ 估計 θ 的誤差會小於或等於 $1.96d$.

由於 Slutsky 定理中的常態近似, 故樣本數 (sample size) 不可太小. 因此, 決定何時停止模擬的演算法如下述.

演算法:

(1) 選擇一適當的 d 作爲標準差, 亦相當於, 在 95% 的信心下,

$$d = \frac{\text{給定的誤差}}{1.96}$$

(2) 至少生成 100 個資料.

(3) 繼續模擬, 直至生成的 k 個資料 ($k \geq 100$) 滿足

$$\frac{S}{\sqrt{k}} < d$$

才停止, 其中

$$S = \left[\frac{1}{k-1} \sum_{i=1}^{k} (X_i - \overline{X})^2 \right]^{1/2}$$

(4) 以

$$\overline{X} = \frac{1}{k} \sum_{i=1}^{k} X_i$$

估計 θ.

例 1. 在單一服務者佇列系統中, 若下午 5 點以後不接受服務, 並欲估計超過時間 T_p (單位: 秒). 試問需要模擬多少次, 以致於可以 95% 地確信誤差在 15 秒內?

<解> 根據前述經驗, 95% 的信心乃相當於要求

$$\text{誤差 } 15 \text{ 秒} = 1.96d$$

故標準差

$$d = \frac{15}{1.96}$$

接著, 由演算法知, 模擬次數

$$k \geq 100$$

且滿足

$$\frac{S}{\sqrt{k}} < \frac{15}{1.96}$$

亦相當於

$$1.96\frac{S}{\sqrt{k}} < 15$$

註 3. 避免每得到一筆資料就從頭計算 $\overline{X}$ 與 S^2, 可採用如下的迭代公式:

(i) 令初始值 $\overline{X}_0 = 0$ 且 $S_1^2 = 0$.

(ii) 對於 $j \geq 0$,

$$\overline{X}_{j+1} = \overline{X}_j + \frac{X_{j+1} - \overline{X}_j}{j+1}$$

(iii) 對於 $j \geq 1$,

$$S_{j+1}^2 = \left(1 - \frac{1}{j}\right) S_j^2 + (j+1)(\overline{X}_{j+1} - \overline{X}_j)^2$$

<證> (i) 合理的初始值. 因爲沒有任何資料時, 一個合理的資料期望值設定爲 $\overline{X}_0 = 0$; 當有一筆資料時, 資料的期望值就是此筆資料, 故此筆資料對於資料期望值而言, 無任何的偏離, 因而 $S_1^2 = 0$ 是一個合理的設定.

(ii) 對於 $j \geq 0$, 根據樣本期望值的定義,

$$\begin{aligned}
\overline{X}_{j+1} &= \frac{1}{j+1} \sum_{i=1}^{j+1} X_i \\
&= \frac{1}{j+1} \left(\sum_{i=1}^{j} X_i + X_{j+1} \right) \\
&= \frac{1}{j+1} (j\overline{X}_j + X_{j+1}) \\
&= \frac{1}{j+1} \left[(j+1)\overline{X}_j + X_{j+1} - \overline{X}_j \right] \\
&= \overline{X}_j + \frac{X_{j+1} - \overline{X}_j}{j+1}
\end{aligned}$$

得證.

註. 當 $j = 0$ 時, 由迭代公式 (ii), 得

$$\overline{X}_1 = \overline{X}_0 + \frac{X_1 - \overline{X}_0}{1} = X_1$$

恆成立, 無論 $\overline{X}_0$ 爲何數. 因此, 除了迭代公式 (i) 中的合理假設 $\overline{X}_0 = 0$ 外, 事實上可假設 $\overline{X}_0$ 爲任何實數.

(iii) 對於 $j \geq 1$, 根據樣本變異數的定義,

$$\begin{aligned} S_{j+1}^2 &= \frac{1}{j}\left[\sum_{i=1}^{j+1}(X_i - \overline{X}_{j+1})^2\right] \\ &= \frac{1}{j}\left[\sum_{i=1}^{j+1}(X_i - \overline{X}_j + \overline{X}_j - \overline{X}_{j+1})^2\right] \end{aligned}$$

接著, 將上式中的小括號展開並整理, 得

$$S_{j+1}^2 = \frac{1}{j}(\Sigma_1 + \Sigma_2 + \Sigma_3) \qquad (4)$$

其中

$$\Sigma_1 = \sum_{i=1}^{j+1}(X_i - \overline{X}_j)^2$$

且

$$\Sigma_2 = 2(\overline{X}_j - \overline{X}_{j+1})\sum_{i=1}^{j+1}(X_i - \overline{X}_j)$$

以及

$$\Sigma_3 = (j+1)(\overline{X}_j - \overline{X}_{j+1})^2$$

又

$$\begin{aligned}\Sigma_1 &= \sum_{i=1}^{j}(X_i - \overline{X}_j)^2 + (X_{j+1} - \overline{X}_j)^2 \\ &= (j-1)S_j^2 + (j+1)^2(\overline{X}_{j+1} - \overline{X}_j)^2\end{aligned}$$

其中第二個等號的第一項成立乃根據樣本變異數 S_j^2 的定義, 第二項成立乃根據迭代公式 (ii) 所導出的

$$X_{j+1} - \overline{X}_j = (j+1)(\overline{X}_{j+1} - \overline{X}_j)$$

所致. 另外, 再根據上式以及樣本期望值 $\overline{X}_j$ 所導出的

$$\sum_{i=1}^{j}(X_i - \overline{X}_j) = j\overline{X}_j - j\overline{X}_j = 0$$

得

$$\begin{aligned}\Sigma_2 &= 2(\overline{X}_j - \overline{X}_{j+1})\sum_{i=1}^{j}(X_i - \overline{X}_j) + \\ &\qquad 2(\overline{X}_j - \overline{X}_{j+1})(X_{j+1} - \overline{X}_j) \\ &= 2(\overline{X}_j - \overline{X}_{j+1})(j+1)(\overline{X}_{j+1} - \overline{X}_j) \\ &= -2(j+1)(\overline{X}_{j+1} - \overline{X}_j)^2\end{aligned}$$

最後, 將 Σ_1, Σ_2 以及 Σ_3 代入 (4) 式, 並提出公因式

$$(j+1)(\overline{X}_{j+1}-\overline{X}_j)^2$$

得

$$\begin{aligned} S_{j+1}^2 &= \frac{1}{j}[(j-1)S_j^2+(j+1)(\overline{X}_{j+1}-\overline{X}_j)^2\cdot \\ &\qquad\qquad (j+1-2+1)] \\ &= \frac{1}{j}[(j-1)S_j^2+j(j+1)(\overline{X}_{j+1}-\overline{X}_j)^2] \\ &= \left(1-\frac{1}{j}\right)S_j^2+(j+1)(\overline{X}_{j+1}-\overline{X}_j)^2 \end{aligned}$$

得證.

註 4. 若欲估計某一特定事件發生的機率 p, 可設法建構此隨機模型, 並令輸出資料爲

$$X=\begin{cases} 1, & \text{若此事件發生,} \\ 0, & \text{若此事件不發生.} \end{cases}$$

則

$$E(X)=1\cdot P(\text{事件發生})=1\cdot p=p$$

就是欲估計的值.

因此, 重複模擬 k 次, 得

$$X_1, X_2, \ldots, X_k$$

並以

$$\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i \approx p$$

因爲

$$\mathsf{Var}(X) = p(1-p)$$

所以不需用樣本變異數估計, 而可用

$$\overline{X}_k(1-\overline{X}_k)$$

估計, 並依然得

Slutsky 定理. 當 $k \to \infty$ 時,

$$\frac{\overline{X}_k - p}{\sqrt{\overline{X}_k(1-\overline{X}_k)/k}} \xrightarrow{\mathcal{D}} Z$$

優點: 省掉算樣本變異數 S_k^2, 而只需計算樣本期望值 $\overline{X}_k$ 即可.

具備何時停止功能的對應

演算法:

(1) 選擇適當的 d 作標準差.

(2) 至少生成 100 筆資料.

(3) 繼續模擬, 直至生成的 k 筆資料 ($k \geq 100$) 滿足

$$\sqrt{\overline{X}_k(1-\overline{X}_k)/k} < d$$

才停止, 其中

$$\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$$

(4) 以 $\overline{X}_k$ 估計 p.