單元 23: 分層取樣
(課本 §8.4)

一有限容量佇列系統 (a finite capacity queueing system) 的假設爲

(1) 每天結束時, 倒空系統, 第二天重新開始.

(2) 顧客的到達時間 $\sim \text{PP}(\lambda_i)$, 其中對於 $i = 1, 2, 3$,

$$P(\text{到達率} = \lambda_i) = \frac{1}{3}$$

且天與天之間的到達率相互獨立.

(3) 隨機變數 L 爲一天的顧客流失數.

問. 如何估計 $E(L)$?

答 1. 一個直接的作法爲, 先考慮下述的

演算法 1:

(1) 生成一仿隨機數 U.

(2) 令

$$I = \text{Int}(3U) + 1$$

亦即,

$$I \sim \text{unif}(\{1, 2, 3\})$$

(3) 以 λ_I 爲到達率, 生成卜松過程, 並求得原始輸出資料 L.

接著, 重覆演算法 1 n 次, 得

$$L_1, \ldots, L_n$$

最後, 根據 SLLN, 以原始估計量 (raw estimator)

$$\overline{L} = \frac{1}{n} \sum_{i=1}^{n} L_i$$

模擬 $E(L)$.

答 2. 因爲此系統的顧客到達率不是事前就確定的, 而是先根據一機率分布的選取而確定, 故可根據此特性, 改良

前述未用到此特性的演算法 1. 首先, 對於 $i = 1, 2, 3$, 令

$$L_i^* \stackrel{\text{def}}{=} L|\lambda_i$$

表示在給定到達率 λ_i 時, 一天的顧客流失數.

接著, 考慮輸出資料爲

$$\frac{1}{3}L_1^* + \frac{1}{3}L_2^* + \frac{1}{3}L_3^*$$

則

(1) 輸出資料的期望值

$$E\left[\frac{1}{3}(L_1^* + L_2^* + L_3^*)\right] = E(L)$$

亦即, 輸出資料

$$\frac{1}{3}(L_1^* + L_2^* + L_3^*)$$

爲 $E(L)$ 的不偏估計量.

(2) 針對變異數,

$$\frac{1}{3}\text{Var}(L_1^* + L_2^* + L_3^*) \leq \text{Var}(L) \qquad (1)$$

由此可導出輸出資料的變異數

$$
\begin{aligned}
& \text{Var}\left[\frac{1}{3}(L_1^* + L_2^* + L_3^*)\right] \\
= \ & \frac{1}{9}\text{Var}(L_1^* + L_2^* + L_3^*) \leq \frac{1}{3}\text{Var}(L) \\
< \ & \text{Var}(L)
\end{aligned}
$$

故有較小的變異數, 比原始資料變異數的 $\frac{1}{3}$ 還小.

為何如此?

(1) 令隨機變數 I 為到達率的下標, 則根據假設,

$$
I \sim \text{unif}(\{1,2,3\})
$$

因此, 以隨機變數 I 條件化, 並根據 I 的分布, L_i^* 的定義, 以及期望值的線性性質, 得

$$
\begin{aligned}
E(L) &= \sum_{i=1}^{3} E(L|I=i)P(I=i) \\
&= \sum_{i=1}^{3} E(L_i^*) \cdot \frac{1}{3} \\
&= E\left(\frac{1}{3}L_1^* + \frac{1}{3}L_2^* + \frac{1}{3}L_3^*\right)
\end{aligned}
$$

得證.

(2) 由條件變異數的公式

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$$

知,

$$\text{Var}(X) \geq E[\text{Var}(X|Y)]$$

因為條件變異數公式的等號右邊二項均為非負.

因此, 根據上式, 隨機變數的函數的期望值公式, 條件變異數的定義, 以及 I 的分布, 得

$$\begin{aligned}\text{Var}(L) &\geq E[\text{Var}(L|I)] \\ &= \sum_{i=1}^{3} \text{Var}(L|I=i)P(I=i) \\ &= \frac{1}{3}\sum_{i=1}^{3} \text{Var}(L|I=i)\end{aligned}$$

又因為 I 為到達率的下標, 故事件 $\{I=i\}$ 乃相當於事件 $\{\lambda=\lambda_i\}$, 再根據 L_i^* 的定義, 以及天與天之間的獨立性, 由上式得

$$\begin{aligned}\text{Var}(L) &\geq \frac{1}{3}\sum_{i=1}^{3} \text{Var}(L|\lambda_i) \\ &= \frac{1}{3}\text{Var}(L_1^* + L_2^* + L_3^*)\end{aligned}$$

得證.

所以, 可考慮以

$$\frac{1}{3}(L_1^* + L_2^* + L_3^*)$$

作爲模擬 $E(L)$ 的輸出資料, 並得對應的

演算法 2:

(1) 對於 $i = 1, 2, 3$, 各自生成 $\frac{n}{3}$ 個

$$L_i^{*1}, \ldots, L_i^{*(n/3)}$$

(2) 對於 $i = 1, 2, 3$, 令

$$\overline{L_i^*} = \frac{3}{n} \sum_{k=1}^{n/3} L_i^{*k}$$

亦即, $\overline{L_i^*}$ 是 L_i^* 的樣本期望值.

(3) 以分層估計量 (stratified estimator)

$$\frac{1}{3}(\overline{L_1^*} + \overline{L_2^*} + \overline{L_3^*})$$

模擬 $E(L)$.

比較. 演算法 1 需生成 n 個原始輸出資料 L, 且原始估計量 $\overline{L}$ 的變異數

$$\text{Var}(\overline{L}) = \frac{1}{n}\text{Var}(L)$$

演算法 2 需對於 $i = 1, 2, 3$, 各自生成 $\frac{n}{3}$ 個分層輸出資料 L_i^*, 共 n 個分層輸出資料, 且分層估計量

$$\frac{1}{3}(\overline{L_1^*} + \overline{L_2^*} + \overline{L_3^*})$$

的變異數

$$\begin{aligned}
&\text{Var}\left[\frac{1}{3}(\overline{L_1^*} + \overline{L_2^*} + \overline{L_3^*})\right] \\
&\quad = \frac{1}{9}\text{Var}(\overline{L_1^*} + \overline{L_2^*} + \overline{L_3^*}) \\
&\quad = \frac{1}{9} \cdot \frac{3}{n}\text{Var}(L_1^* + L_2^* + L_3^*) \\
&\quad = \frac{1}{n} \cdot \frac{1}{3}\text{Var}(L_1^* + L_2^* + L_3^*) \\
&\quad \le \frac{1}{n}\text{Var}(L) = \text{Var}(\overline{L})
\end{aligned}$$

其中第二等號成立乃因爲對於 $i = 1, 2, 3$, $\overline{L_i^*}$ 分別是大小爲 $\frac{n}{3}$ 的 L_i^* 的樣本期望值, 故經由累加和的運算性質,

得

$$\overline{L_1^*}+\overline{L_2^*}+\overline{L_3^*}$$

是大小爲 $\frac{n}{3}$ 的

$$L_1^*+L_2^*+L_3^*$$

的樣本期望值, 並根據樣本期望值的變異數公式所致; 第四個小於或等於符號乃根據 (1) 式所致.

因此, 以演算法 2 模擬, 會得到較小的變異數.

例 1. 一無限多服務員佇列系統 (an infinite server queueing system) 的假設爲

(1) 顧客的到達時間乃根據一個卜松過程 (PP), 其

$$\text{到達率}=\begin{cases}12, & \text{好日子},\\ 4, & \text{一般日子}.\end{cases}$$

且好日子與一般日子乃相互獨立, 以及

$$P(\text{好日子})=\frac{1}{2}$$

(2) 在所有的日子, 顧客接受的

$$\text{服務時間}\sim \exp(1)$$

(3) 在 $t = 10$ 時, 系統關閉, 所有未完成服務的顧客均需離開. 令 L 爲在隨機選中的一天內, 未完成服務的顧客數.

試比較 $E(L)$ 的原始估計量與分層估計量間的差異.

<解> 令隨機變數

$$L_g \stackrel{\text{def}}{=} L|\text{好日子} = L|(\lambda = 12)$$

表示在好日子的未完成服務的顧客數, 且隨機變數

$$L_o \stackrel{\text{def}}{=} L|\text{一般日子} = L|(\lambda = 4)$$

表示在一般日子的未完成服務的顧客數.

接著, 敍述二個事實

$$L_g \sim \text{Poisson}[12(1 - e^{-10})]$$

以及

$$L_o \sim \text{Poisson}[4(1 - e^{-10})]$$

並分別探討下述的兩種情況.

情況 1. 原始估計量.

以日子的種類條件化, 並根據 L_g 與 L_o 的定義, 日子種類的分布, 以及上述的事實, 得 L 的期望值

$$\begin{aligned}
E(L) &= E(L|\text{好日子})P(\text{好日子}) + \\
&\qquad E(L|\text{一般日子})P(\text{一般日子}) \\
&= \frac{1}{2}E(L_g) + \frac{1}{2}E(L_o) \\
&= \frac{1}{2}(12+4)(1-e^{-10}) \\
&\approx \frac{1}{2}(12+4) = 8
\end{aligned}$$

又根據變異數的公式, 上述的事實, 以及卜松分布的變異數等於期望值的性質, 得 L_g^2 的二階動差

$$\begin{aligned}
E(L_g^2) &= \mathrm{Var}(L_g) + [E(L_g)]^2 \\
&= E(L_g) + [E(L_g)]^2 \\
&= 12(1-e^{-10}) + [12(1-e^{-10})]^2 \\
&\approx 12 + (12)^2 = 156
\end{aligned}$$

同理, 得 L_o 的二階動差

$$\begin{aligned}
E(L_o^2) &= \mathrm{Var}(L_o) + [E(L_o)]^2 \\
&= 4(1-e^{-10}) + [4(1-e^{-10})]^2 \\
&\approx 4 + 4^2 = 20
\end{aligned}$$

最後, 以日子的種類條件化, 並根據 L_g 與 L_o 的定義, 日子種類的分布, 以及上述 L_g 與 L_o 的二階動差, 得 L 的二階動差

$$\begin{aligned} E(L^2) &= E(L^2|\text{好日子})P(\text{好日子}) + \\ &\qquad E(L^2|\text{一般日子})P(\text{一般日子}) \\ &= \frac{1}{2}E(L_g^2) + \frac{1}{2}E(L_o^2) \\ &\approx \frac{1}{2}(156 + 20) = 88 \end{aligned}$$

故, L 的變異數

$$\begin{aligned} \text{Var}(L) &= E(L^2) - [E(L)]^2 \\ &\approx 88 - 8^2 = 24 \end{aligned}$$

因此, 模擬的過程爲, 生成原始輸出資料

$$L_1, \ldots, L_n$$

得原始估計量

$$\overline{L} = \frac{1}{n}\sum_{i=1}^{n} L_i$$

以及

$$\text{Var}(\overline{L}) = \frac{1}{n}\text{Var}(L) \approx \frac{24}{n} \qquad (2)$$

情況 2. 分層估計量.

因爲

$$P(\text{好日子}) = \frac{1}{2}$$

故可考慮以

$$\frac{1}{2}(L_g + L_o)$$

作爲模擬 $E(L)$ 的輸出資料.

又根據變異數的純量乘積公式, 好日子與一般日子的獨立性假設, 上述的事實, 以及卜松分布的性質, 得分層輸出資料的變異數

$$\begin{aligned} \text{Var}\left[\frac{1}{2}(L_g + L_o)\right] &= \frac{1}{4}[\text{Var}(L_g) + \text{Var}(L_o)] \\ &= \frac{1}{4}[E(L_g) + E(L_o)] \\ &\approx \frac{1}{4}(12 + 4) = 4 \end{aligned} \qquad (3)$$

因此, 分層模擬法爲, 生成 $\frac{n}{2}$ 個

$$L_{g,1}, \ldots, L_{g,n/2}$$

以及 $\frac{n}{2}$ 個

$$L_{o,1}, \ldots, L_{o,n/2}$$

得分層估計量

$$\frac{1}{2}\overline{L_g}+\frac{1}{2}\overline{L_o}$$

其中

$$\overline{L_g}=\frac{2}{n}\sum_{i=1}^{n/2}L_{g,i}$$

且

$$\overline{L_o}=\frac{2}{n}\sum_{i=1}^{n/2}L_{o,i}$$

另外, 根據累加和的運算性質以及 (3) 式, 得

$$\begin{aligned}\text{Var}\left[\frac{1}{2}(\overline{L_g}+\overline{L_o})\right] &= \frac{2}{n}\text{Var}\left[\frac{1}{2}(L_g+L_o)\right]\\ &\approx \frac{2}{n}\cdot 4=\frac{8}{n}\end{aligned}\tag{4}$$

最後, 綜合 (2) 式與 (4) 式, 得

$$\text{Var}(\text{原始估計量})\approx 3\cdot\text{Var}(\text{分層估計量})$$

註 1. 推廣. 設有 k 種到達率

$$\lambda_i,\ i=1,\ldots,k$$

且對於 $i = 1, \ldots, k$,

$$P(\text{到達率} = \lambda_i) = p_i$$

以及

$$\sum_{i=1}^{k} p_i = 1$$

對於 $i = 1, \ldots, k$, 令隨機變數

$$L_i^* = L|\lambda_i$$

分層法 (stratified approach) 爲

(1) 對於 $i = 1, \ldots, k$, 各自生成

$$n_i \stackrel{\text{def}}{=} np_i$$

個 L_i^*, 並求樣本期望值

$$\overline{L_i^*}$$

(2) 以分層估計量

$$\widehat{\theta} \stackrel{\text{def}}{=} \sum_{i=1}^{k} p_i \overline{L_i^*}$$

估計 $\theta = E(L)$.

爲何可行?

(a) $\hat{\theta}$ 爲 θ 的不偏估計量.

根據期望值的線性性質, 樣本期望值 $\overline{L_i^*}$ 的不偏性, 隨機變數 L_i^* 的定義, 到達率的分布, 以及期望值的條件化公式,

$$\begin{aligned}
E(\hat{\theta}) &= \sum_{i=1}^{k} p_i E(\overline{L_i^*}) = \sum_{i=1}^{k} p_i E(L_i^*) \\
&= \sum_{i=1}^{k} E(L|\lambda_i) P(\text{到達率} = \lambda_i) \\
&= E(L) = \theta
\end{aligned}$$

得證.

(b) $\hat{\theta}$ 有較小的變異數.

根據樣本期望值 $\overline{L_i^*}$ 間的獨立性, 變異數的純量乘積公式, 樣本期望值 $\overline{L_i^*}$ 的變異數公式, 得

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}) &= \sum_{i=1}^{k} p_i^2 \mathrm{Var}(\overline{L_i^*}) \\
&= \sum_{i=1}^{k} p_i^2 \cdot \frac{1}{n_i} \mathrm{Var}(L_i^*)
\end{aligned}$$

再根據 n_i 的定義, 以及 L_i 的定義, 由上式得

$$\begin{aligned}
\mathsf{Var}(\widehat{\theta}) &= \sum_{i=1}^{k} p_i^2 \cdot \frac{1}{np_i}\mathsf{Var}(L_i^*) \\
&= \frac{1}{n}\sum_{i=1}^{k} p_i \mathsf{Var}(L_i^*) \\
&= \frac{1}{n}\sum_{i=1}^{k} \mathsf{Var}(L|\lambda_i)P(\text{到達率} = \lambda_i) \quad (5)
\end{aligned}$$

又令隨機變數 I 滿足, 對於 $i = 1, \ldots, k$,

$$P(I = i) = p_i$$

且

$$\{I = i\} = \{\text{到達率} = \lambda_i\}$$

則將 (5) 式改寫並根據條件變異數的期望值公式, 得

$$\begin{aligned}
\mathsf{Var}(\widehat{\theta}) &= \frac{1}{n}\sum_{i=1}^{k} \mathsf{Var}(L|I = i)P(I = i) \\
&= \frac{1}{n}E[\mathsf{Var}(L|I)] \quad (6)
\end{aligned}$$

最後, 根據熟知的條件期望值的變異數與原隨機變數的變異數的公式

$$\mathsf{Var}(L) = E[\mathsf{Var}(L|I)] + \mathsf{Var}[E(L|I)]$$

知, 上式等號右邊的兩項均爲非負,故由 (6) 式, 上式, 以及樣本期望值的變異數公式, 得

$$\mathrm{Var}(\widehat{\theta}) \leq \frac{1}{n}\mathrm{Var}(L) = \mathrm{Var}(\overline{L})$$

得證.

因此, 分層估計量

$$\widehat{\theta} = \sum_{i=1}^{k} p_i \overline{L_i^*}$$

是 θ 的不偏估計量, 並較原始估計量 $\overline{L}$ 有較小的變異數, 而得上述可行的推廣型分層法.

註 2. 通常 $\mathrm{Var}(L_i^*)$ 未知, 所以可用

$$\sum_{i=1}^{k} p_i^2 \cdot \frac{S_i^2}{n_i}$$

估計

$$\mathrm{Var}(\widehat{\theta}) = \sum_{i=1}^{k} p_i^2 \cdot \frac{\mathrm{Var}(L_i^*)}{n_i}$$

其中 S_i^2 爲 L_i^* 的樣本變異數.

問. 取 $n_i = np_i$ 時, $\mathrm{Var}(\widehat{\theta})$ 是否爲最小?

此問題乃相當於如何選取 n_i 使得

$$\widehat{\theta} = \sum_{i=1}^{k} p_i \overline{L_i^*}$$

有最小的變異數?

答. 對於 $i = 1, \ldots, k$, 令任意的 $n_i \geq 0$ 且滿足

$$\sum_{i=1}^{k} n_i = n$$

則生成 n_i 個 L_i^* 後, 並考慮

$$\sum_{i=1}^{k} p_i \overline{L_i^*}$$

得

(a) 隨機變數

$$\sum_{i=1}^{k} p_i \overline{L_i^*}$$

爲 $E(L)$ 的不偏估計量.

爲何如此? 因爲 $\overline{L_i^*}$ 爲 L_i^* 的樣本期望值, 故根據期望值的線性性質, 樣本期望值的不偏性, 以及期望值的條件

化公式,

$$
\begin{aligned}
E\left(\sum_{i=1}^{k} p_i \overline{L_i^*}\right) &= \sum_{i=1}^{k} p_i E(\overline{L_i^*}) \\
&= \sum_{i=1}^{k} p_i E(L_i^*) \\
&= E(L)
\end{aligned}
$$

得證.

(b) 變異數

$$
\text{Var}\left(\sum_{i=1}^{k} p_i \overline{L_i^*}\right) = \sum_{i=1}^{k} p_i^2 \cdot \frac{\text{Var}(L_i^*)}{n_i}
$$

此乃因爲 $\overline{L_i^*}$ 間的獨立性, 變異數的線性組合公式, 以及樣本期望值的變異數公式所致.

因此, 綜合上述的 (a) 與 (b) 得知, 對於 $i = 1, \ldots, k$, 模擬 S_i^2 估計 Var(L_i^*) 後, 對任意非負的

$$
n_1, \ldots, n_k
$$

滿足

$$
\sum_{i=1}^{k} n_i = n
$$

均得到

$$\sum_{i=1}^{k} p_i \overline{L_i^*}$$

爲 $E(L)$ 的不偏估計量, 且

$$\text{Var}\left(\sum_{i=1}^{k} p_i \overline{L_i^*}\right) \approx \sum_{i=1}^{k} p_i^2 \cdot \frac{S_i^2}{n_i}$$

所以, 前述如何選取 n_i 的問題, 乃相當於一個最佳化問題 (optimization problem):
最小化 (minimize)

$$\sum_{i=1}^{k} p_i^2 \cdot \frac{S_i^2}{n_i}$$

受制於 (subject to)

$$n_i \geq 0 \text{ 且 } \sum_{i=1}^{k} n_i = n$$

因此, 根據最佳化問題而解得的 n_i, 所產生的估計量

$$\sum_{i=1}^{k} p_i \overline{L_i^*}$$

會比取 $n_i = np_i$, 有較小的變異數.

例. 接續例 1 的無窮服務員佇列系統, 其中到達率的種類數 $k = 2$ 且產生的機率均為 $p = \frac{1}{2}$. 對於 $0 \le x \le 1$, 生成 $n_1 = nx$ 個 L_g 及 $n_2 = n(1-x)$ 個 L_o.

又根據前述的兩個事實

$$L_g \sim \text{Poisson}[12(1-e^{-10})]$$

以及

$$L_o \sim \text{Poisson}[4(1-e^{-10})]$$

得

$$S_1^2 \approx \text{Var}(L_g) \approx 12$$

且

$$S_2^2 \approx \text{Var}(L_o) \approx 4$$

故, 根據註 2, 最佳的 n_1 與 n_2 乃相當於求

$$\min_{0\le x\le 1}\left[\left(\frac{1}{2}\right)^2 \frac{S_1^2}{nx} + \left(\frac{1}{2}\right)^2 \frac{S_2^2}{n(1-x)}\right]$$

代入 S_1^2 與 S_2^2 的估計值, 消去不影響極值的

$$\left(\frac{1}{2}\right)^2 \frac{1}{n}$$

後, 亦相當於求

$$\min_{0\leq x\leq 1}\left(\frac{12}{x}+\frac{4}{1-x}\right)$$

亦相當於再除以 4, 並令

$$f(x)=\frac{3}{x}+\frac{1}{1-x}$$

且由

$$\begin{aligned} f'(x) &= -\frac{3}{x^2}+\frac{1}{(1-x)^2} \\ &= \frac{-3+6x-2x^2}{x^2(1-x)^2}=0 \end{aligned}$$

得臨界數 (critical numbers)

$$x=\frac{-6\pm\sqrt{36-24}}{-4}=\frac{3\mp\sqrt{3}}{2}$$

其中 "$+$" 不合, 因爲大於 1, 不在求極值的範圍內.

再根據 f' 的分子爲一開口向下的拋物線, 可計算出 f' 在此一臨界數所分隔的兩個子區間內的符號及對應的遞增遞減性爲

$\left(0,\frac{3-\sqrt{3}}{2}\right)$: $f'=\frac{(-)}{(+)}=(-)$, 遞減;

$\left(\frac{3-\sqrt{3}}{2}, 1\right)$: $f' = \frac{(+)}{(+)} = (+)$, 遞增.

故, 根據一階導函數檢定法,

$$f(x) = \frac{3}{x} + \frac{1}{1-x}$$

在

$$x = \frac{3-\sqrt{3}}{2} \approx 0.634$$

有最小值.

因此, 生成 $(0.634)n$ 個 L_g 以及 $(0.366)n$ 個 L_o, 並以

$$\frac{1}{2}\overline{L_g} + \frac{1}{2}\overline{L_o}$$

估計 $E(L)$ 會有最小的變異數.

定義. 設隨機變數 X 的值需先由另一隨機變數 S, 其已知分布爲

$$P(S=i),\ i = 1, \ldots k$$

所決定, 則不先模擬 S, 而先對於 $i = 1, \ldots, k$, 分別模擬

$$X|S=i$$

並求出對應的樣本期望值 $\overline{X_i}$ 且以

$$\sum_{i=1}^{k} \overline{X_i} P(S = i)$$

估計

$$E(X) = \sum_{i=1}^{k} E(X|S = i)P(S = i)$$

的方法稱爲分層取樣法 (stratified sampling).

註. 由上述期望值的條件期望值表示法知, 分層取樣法的關鍵就在於將條件期望值 $E(X|S = i)$ 以 $X|S = i$ 的樣本期望值 $\overline{X_i}$ 所估計與取代.

例 2. 試以分層取樣法模擬

$$\theta = \int_0^1 h(x)dx$$

<解> (a) 一般方法. 令隨機變數

$$U \sim \text{unif}(0, 1)$$

則根據隨機變數的函數的期望值公式,

$$E[h(U)] = \int_0^1 h(u) \cdot 1du = \theta$$

剛好就是欲估計的值.

故根據蒙地卡羅法, 考慮

演算法 1:

(1) 生成一個仿隨機數 U.

(2) 生成輸出資料 $h(U)$.

接著, 重覆演算法 1 n 次, 得

$$h(U_1), \ldots, h(U_n)$$

最後, 根據 SLLN, 以原始估計量 (raw estimator)

$$\overline{h(U)} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} h(U_i)$$

估計 $E[h(U)] = \theta$.

(b) 分層取樣. 將區間 $(0, 1]$ n 等分, 得

$$(0, 1] = \bigcup_{i=1}^{n} \left(\frac{i-1}{n}, \frac{i}{n} \right]$$

則上述演算法 1 中的步驟 (1) 生成 U, 乃相當於隨機選取一個子區間及其中的一值 U; 步驟 (2) 生成 $h(U)$, 乃相當於在選中的子區間內製造 $h(U)$. 此乃類似於前述例子, 先隨機選定到達率, 再求得對應輸出資料的兩步驟模擬過程, 故可考慮以分層取樣的方式改良演算法 1.

如何分層? 因爲共有 n 個子區間, 且每一子區間被選中的機率均爲 $\frac{1}{n}$, 故可分成 n 層如下,

(1) 對於 $i = 1, \ldots, n$, 定義子區間的指標 (index)

$$I = i, \text{ 若 } \frac{i-1}{n} < U \le \frac{i}{n}$$

由此導出

$$P(I = i) = P\left(\frac{i-1}{n} < U \le \frac{i}{n}\right) = \frac{1}{n}$$

故

$$I \sim \text{unif}(\{1, \ldots, n\})$$

如所求.

(2) 對於 $i = 1, \ldots, n$, 給定 $I = i$ 時, 定義

$$U_i^* \stackrel{\text{def}}{=} U | I = i$$

則顯然地 (自行驗證),

$$U_i^* \sim \text{unif}\left\{\left(\frac{i-1}{n}, \frac{i}{n}\right]\right\} \tag{7}$$

爲何如此? 對於任意固定的 $i = 1, \ldots, n$, 以及對於所有的 $-\infty < x < \infty$, 根據 U_i^* 的定義, I 的定義, 以及條件機率的定義, 得分布函數

$$\begin{aligned} P(U_i^* \leq x) &= P(U \leq x | I = i) \\ &= P\left(U \leq x \,\middle|\, \frac{i-1}{n} < U \leq \frac{i}{n}\right) \\ &= \frac{P\left(U \leq x, \frac{i-1}{n} < U \leq \frac{i}{n}\right)}{P\left(\frac{i-1}{n} < U \leq \frac{i}{n}\right)} \end{aligned} \tag{8}$$

另外, 根據 U 在 $(0,1)$ 的均匀分布, (8) 的分母

$$P\left(\frac{i-1}{n} < U \leq \frac{i}{n}\right) = \frac{i}{n} - \frac{i-1}{n} = \frac{1}{n} \tag{9}$$

故, 當 $x \leq \frac{i-1}{n}$ 時, 事件

$$\left(U \leq x, \frac{i-1}{n} < U \leq \frac{i}{n}\right) = \emptyset$$

故由 (8) 式與 (9) 式, 得

$$P(U_i^* \leq x) = n \cdot P(\emptyset) = n \cdot 0 = 0$$

當 $\frac{i-1}{n} < x \leq \frac{i}{n}$ 時, 事件

$$\left(U \leq x, \frac{i-1}{n} < U \leq \frac{i}{n}\right) = \left(\frac{i-1}{n} < U \leq x\right)$$

故由 (8) 式, (9) 式, 以及 U 的均勻性, 得

$$\begin{aligned} P(U_i^* \leq x) &= n \cdot P\left(\frac{i-1}{n} < U \leq x\right) \\ &= n\left(x - \frac{i-1}{n}\right) \end{aligned}$$

當 $x > \frac{i}{n}$ 時, 事件

$$\left(U \leq x, \frac{i-1}{n} < U \leq \frac{i}{n}\right) = \left(\frac{i-1}{n} < U \leq \frac{i}{n}\right)$$

故由 (8) 式, (9) 式, 以及 U 的均勻性, 得

$$\begin{aligned} P(U_i^* \leq x) &= n \cdot P\left(\frac{i-1}{n} < U \leq \frac{i}{n}\right) \\ &= n \cdot \frac{1}{n} = 1 \end{aligned}$$

因此, 綜合上述, 得 U_i^* 的分布函數

$$P(U_i^* \leq x) = \begin{cases} 0, & \text{若 } x \leq \frac{i-1}{n}, \\ n\left(x - \frac{i-1}{n}\right), & \text{若 } \frac{i-1}{n} < x \leq \frac{i}{n}, \\ 1, & \text{若 } x > \frac{i}{n}. \end{cases}$$

最後, 將分布函數對 x 微分, 得 U_i^* 的 pdf

$$f_{U_i^*}(x) = \begin{cases} n, & \text{若 } \frac{i-1}{n} < x < \frac{i}{n}, \\ 0, & \text{其它.} \end{cases}$$

故,

$$U_i^* \sim \text{unif}\left\{\left(\frac{i-1}{n}, \frac{i}{n}\right]\right\}$$

得證.

(3) 以分層估計量

$$\frac{1}{n}\sum_{i=1}^{n} h(U_i^*)$$

估計 θ.

對於 $i = 1, \ldots, n$, 如何生成 U_i^*? 令

$$X \sim \text{unif}(0, 1)$$

則

$$\frac{1}{n}(i - 1 + X) \sim \text{unif}\left(\frac{i-1}{n}, \frac{i}{n}\right) \tag{10}$$

(自行驗證), 故得對應的

演算法 2:

(1) 生成 n 個仿隨機數

$$U_1, \ldots, U_n$$

(2) 以分層估計量

$$\widehat{\theta} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} h\left[\frac{1}{n}(i-1+U_i)\right]$$

估計 θ.

註. 對於 $i = 1, \ldots, n$, 根據分布,

$$\frac{1}{n}(i-1+U_i) \sim \text{unif}\left(\frac{i-1}{n}, \frac{i}{n}\right)$$

扮演 U_i^* 的角色, 且 U_i^* 間相互獨立. 另外, 分層估計量 $\widehat{\theta}$ 更接近 (或就是) 黎曼和 (Rieman sum).

為何分層估計量 $\widehat{\theta}$ 是一個估計 θ 的好的估計量?

首先, 根據期望值的線性性質, I 的分布, 得

$$\begin{aligned} E(\widehat{\theta}) &= \frac{1}{n} \sum_{i=1}^{n} E\left\{h\left[\frac{1}{n}(i-1+U_i)\right]\right\} \\ &= \sum_{i=1}^{n} E\left\{h\left[\frac{1}{n}(i-1+U_i)\right]\right\} P(I=i) \end{aligned}$$

再根據 (7) 式與 (10) 式, 亦即, 對於 $i = 1, \ldots, n$,

$$\frac{1}{n}(i - 1 + U_i)$$

與

$$U|I = i$$

有相同的分布, 以及條件化的公式, 由上式得

$$\begin{aligned} E(\widehat{\theta}) &= \sum_{i=1}^{n} E[h(U)|I = i]P(I = i) \\ &= E[h(U)] = \theta \end{aligned}$$

亦即, 分層估計量 $\widehat{\theta}$ 爲不偏估計量.

又根據 U_i 間的獨立性, 獨立隨機變數線性組合的變異數公式, I 的分布, 上述的同分布性質, 以及隨機變數函數的期望值公式, 得

$$\begin{aligned} &\mathrm{Var}(\widehat{\theta}) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left\{h\left[\frac{1}{n}(i - 1 + U_i)\right]\right\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\left\{h\left[\frac{1}{n}(i - 1 + U_i)\right]\right\} P(I = i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}[h(U)|I = i]P(I = i) \\ &= \frac{1}{n} E\{\mathrm{Var}[h(U)|I]\} \end{aligned}$$

其中最後一個等號成立乃因爲 $\mathsf{Var}[h(U)|I]$ 爲一 I 的函數所致.

再根據一再應用的變異數不等式, 以及樣本期望值的變異數公式, 由上式得

$$\begin{aligned} \mathsf{Var}(\widehat{\theta}) &\leq \frac{1}{n}\mathsf{Var}[h(U)] \\ &= \mathsf{Var}[\overline{h(U)}] \end{aligned}$$

所以, 分層估計量

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} h\left[\frac{1}{n}(i-1+U_i)\right]$$

比原始估計量

$$\overline{h(U)} = \frac{1}{n}\sum_{i=1}^{n} h(U_i)$$

有較小的變異數.

例 3. 因爲

$$\begin{aligned} \frac{\pi}{4} &= \text{在 } [0,1] \text{ 上, 函數 } y = \sqrt{1-x^2} \text{ 所圍出} \\ &\quad\ \text{的四分之一單位圓的面積} \\ &= \int_0^1 \sqrt{1-x^2}dx \\ &= E\left(\sqrt{1-U^2}\right) \end{aligned}$$

其中

$$U \sim \text{unif}(0,1)$$

故由例 2, 得 π 的分層估計量

$$\frac{4}{n}\sum_{i=1}^{n}\sqrt{1-\left[\frac{1}{n}(i-1+U_i)\right]^2}$$

比原始估計量

$$\frac{4}{n}\sum_{i=1}^{n}\sqrt{1-U_i^2}$$

亦可由條件化法求得, 有較小的變異數, 其中

$$U_1,\ldots,U_n \overset{\text{iid}}{\sim} \text{unif}(0,1)$$

由於分層估計量中,

$$\sqrt{1-\left[\frac{1}{n}(i-1+U)\right]^2}$$

的單調性質, 事實上爲 U 的遞減函數, 故可用對偶變數法改良, 得更好的估計量

$$\begin{aligned}\widehat{\pi} \;=\; & \frac{4}{n}\sum_{i=1}^{n}\frac{1}{2}\left\{\sqrt{1-\left[\frac{1}{n}(i-1+U_i)\right]^2}+\right.\\ & \left.\sqrt{1-\left[\frac{1}{n}(i-1+(1-U_i))\right]^2}\right\}\end{aligned}$$

經化簡, 得

$$\widehat{\pi} = \frac{2}{n}\sum_{i=1}^{n}\left[\sqrt{1-\left(\frac{i-1+U_i}{n}\right)^2}+\sqrt{1-\left(\frac{i-U_i}{n}\right)^2}\right]$$

一個模擬結果如下:

n	$\widehat{\pi}$
5	3.161211
10	3.148751
100	3.141734
500	3.141615
1000	3.141601
5000	3.141593

練習題 1. 試以 $\widehat{\pi}$ 模擬 π, 並驗證上述的模擬結果.

練習題 2. 試用模擬結果比較 $\widehat{\pi}$, 分層估計量, 以及原始估計量.