

# Statistical Computing and Simulation HW3


授課教授：余清祥教授

學生:　統計碩一　106354003　林健宏

　　　統計碩一　106354012　曹立諭

# 目錄

# Question 01.

Experiment with as many variance reduction techniques as you can think of to apply the problem of evaluating $P(X > 1)$for $X \sim \text{Cauchy}$.

$$\text{實際}\theta \approx P(X > 1) = 0.25$$

以 Monte-Carlo Integration 為基準，並使用其他變異數縮減方法去降低估計的誤差，使用方法有以下幾種：Hit or miss、Antithetic Variate、Importance Sampling、Control variate 以及 Stratified Sampling

（其中用於 Control variate 的另一函數為$\frac{1}{1+x}$, x = 0 ~ 1，Stratified Sampling 將定義域分為 5 層並平均分配抽樣數，用於 Importance Sampling 的另一函數為$\frac{1}{x^2}$ $,x = 0\sim1$。）

每一個方法皆取 1000 個隨機樣本生成一個$\hat{\theta}$，再生成 1000 樣本 $\hat{\theta}$，並運用其平均去估計真實的$\theta$及計算樣本變異數。

在此以表格方式呈現

| | MonteCarlo | Hit or miss | Antithetic | Importance | Control | Stratified(5) |
|---|---|---|---|---|---|---|
| $E(\hat{\theta})$ | 0.24997910 | 0.2498500 | 0.24999466 | 0.24997706 | 0.25017903 | 0.25001367 |
| $var(\hat{\theta})$ | 0.00000272 | 0.0001932 | 0.00000004 | 0.00000244 | 0.00000983 | 0.00000009 |

另外亦有探討抽樣樣本數的問題，分別取 100、1000、10000 樣本數，去探討樣本平均以及樣本變異數之間的關係。

在此以表格方式呈現

| $n = 100$ | Monte-Carlo | Hit or miss | Antithetic | Importance | Control | Stratified(5) |
|---|---|---|---|---|---|---|
| $E(\hat{\theta})$ | 0.25028315 | 0.24937000 | 0.25002894 | 0.24989915 | 0.24878815 | 0.2500383 |
| $var(\hat{\theta})$ | 0.00002736 | 0.00185055 | 0.00000041 | 0.00002551 | 0.00009001 | 0.0000008 |

| $n = 1000$ | Monte-Carlo | Hit or miss | Antithetic | Importance | Control | Stratified(5) |
|---|---|---|---|---|---|---|
| $E(\hat{\theta})$ | 0.24997910 | 0.2498500 | 0.24999466 | 0.24997706 | 0.25017903 | 0.25001367 |
| $var(\hat{\theta})$ | 0.00000272 | 0.0001932 | 0.00000004 | 0.00000244 | 0.00000983 | 0.00000009 |

| $n = 10000$ | Monte-Carlo | Hit or miss | Antithetic | Importance | Control | Stratified(5) |
|---|---|---|---|---|---|---|
| $E(\hat{\theta})$ | 0.25001537 | 0.24978600 | 0.2499985 | 0.24997728 | 0.25001918 | 0.25000063 |
| $var(\hat{\theta})$ | 0.00000025 | 0.00001847 | 0.0000001 | 0.00000024 | 0.00000099 | 0.00000001 |

結論：

實際$\theta = P(X > 1) = 0.25$，由上表可知，每個方法都估計的數值相差無幾，只有 Antithetic Variate 與 Stratified Sampling 的樣本變異數來得非常小，而 Control variate 與 Hit or miss 來得相對大滿多，其中 Control variate 可能與取得函數有關，若是能取得更適合的函數，也許可以將樣本變異數降低更多。

而抽樣樣本數也是一個問題，當樣本數由小至大會發現樣本數的樣本變異數會下降很快，會近乎接近0。因此，在本題使用的估計方法以Antithetic Variate與Stratified Sampling為優，樣本數則是越多誤差越小。

# Question 02.

Hammersley and Handscomb (1964) used the integration of $\theta = \int_0^1 \frac{e^x-1}{e-1}dx$ on (0,1)

as a test problem of variance reduction techniques (which is about 0.4180233).
Achieve as large a variance reduction as you can. (They achieved 4 million.)

$$實際\ \theta = \int_0^1 \frac{e^x-1}{e-1}dx \approx 0.4180233$$

以 Monte-Carlo Integration 為基準，並使用其他變異數縮減方法去降低估計的誤差，使用方法有以下幾種： Antithetic Variate、Importance Sampling、Control variate 以及 Stratified Sampling

（其中用於 Control variate 的另一函數為$e^x - 1$, $x = 0 \sim 1$，Stratified Sampling 將定義域分為 5 層並平均分配抽樣數，用於 Importance Sampling 的另一函數為 $q(x) = \frac{4}{\pi(1+x^2)},\ 0 < x < 1$。）

每一個方法皆取 1000 個隨機樣本生成一個$\hat{\theta}$，再生成 1000 樣本 $\hat{\theta}$，並運用其平均去估計真實的$\theta$及計算樣本變異數。

| | Monte-Carlo | Antithetic | Importance | Control | Stratified(5) |
|---|---|---|---|---|---|
| $E(\hat{\theta})$ | 0.41787975 | 0.41805356 | 0.41824956 | 0.4180233 | 0.41804309 |
| $var(\hat{\theta})$ | 0.00008453 | 0.00000281 | 0.00015477 | 9.253717e-36 | 0.00000345 |

結論：

實際 $\theta = \int_0^1 \frac{e^x-1}{e-1}dx \approx 0.4180233$，由上表可知，每個方法都估計的數值相差無幾，其中 Control variate 的樣本變異數相對於其他的方法來得小很多，近乎接近 0。

而 Importance Sampling 的樣本變異數則是相對於其他方法來得大，這可能與選取的函數有關（本題使用$q(x) = \frac{4}{\pi(1+x^2)},\ 0 < x < 1$），若是能取得更適合的函數，這樣或許可以將樣本變異數降下來。因此，在本題中所使用的估計方法以 Control variate 為優，而樣本數與第一題一樣，取樣的樣本數越多，則樣本變異數會越小。

# # Question 03.

Let $X_i, i = 1,2,3,4,5$ be independent exponential random variables each with mean 1,
and consider the quantity $\theta$ defined by $\theta = P(\sum_{i=1}^{5} iX_i \geq 21.6)$.
Propose at least three simulation methods to estimate $\theta$ and compare their variances.

# # Question 04.

First, simulate 100 observations from $\text{Beta}(2,3)$ and then use 3 density estimating methods to smooth the observations. You need to specify the parameters in the smoothing methods, and compare the results.

# # Question 05.

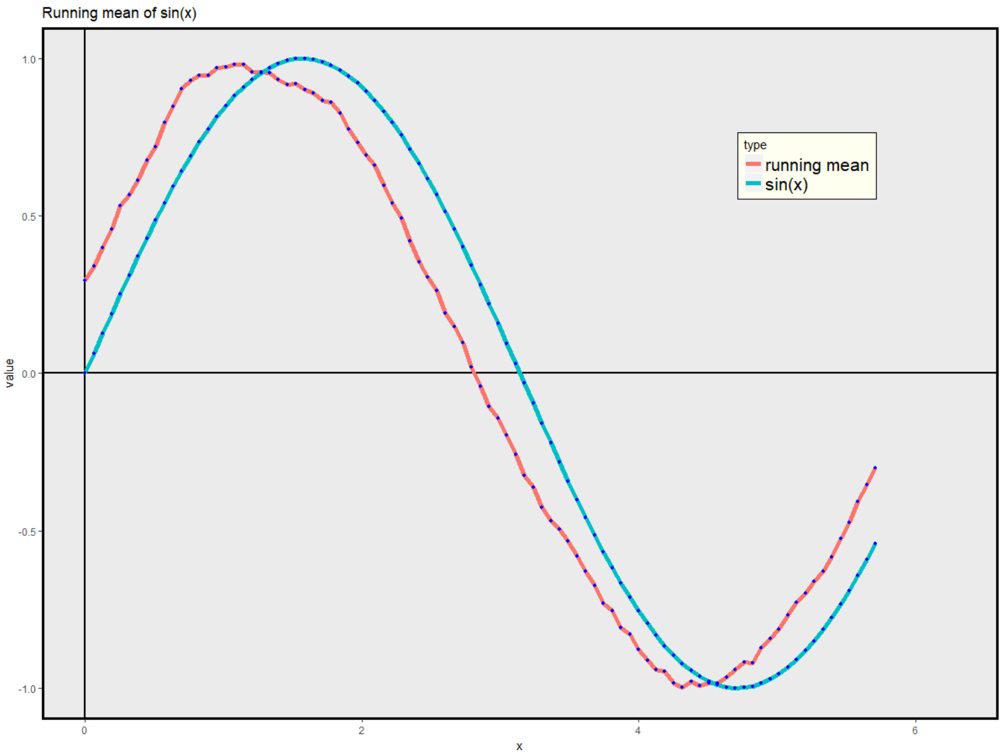
Let $x$ be 100 equally spaced points on $[0,2\pi]$ and let $y_i = \sin x_i + \epsilon_i$ with $\epsilon_i \sim N(0,0.09)$. Apply at least 3 linear smoothers and compare the differences, with respect to mean squares error (i.e., $\text{bias}^2$ and variance) from 1,000 simulation runs.

(a) Kernel smooth

(b) Spline smooth

(c) Lowess smooth

(d) Running mean smooth



## ❐ 附錄（R code）

Github : https://github.com/CaoCharles/Statistical-Computing-and-Simulation-HW3
R Markdown :

1.

2.

3.

4.

5.