**# 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* ~ Cauchy.

|  |  |
| --- | --- |
| (圖一) | (圖二) |
|  |  |

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

使用方法介紹

(1) Monte-Carlo Integration：

P(X1 )可以經由Monte-Carlo Integration模擬，x從uniform(0,1)抽取1000筆資料，並將每筆資料帶入 得到每筆資料的對應值，再取這1000筆資料對應值的平均值為P(X 1 )的估計值。因此P( X > 1)的估計值為0.5 - P(X 1 )的估計值。我們重複這動作1000次取得1000個P( X > 1)的估計值，計算這1000個P( X > 1)的估計值的平均與變異數為0.24997910與0.00000272，此為Monte-Carlo Integration。

(2) Hit or miss

從X ~ Cauchy中隨機抽取1000筆資料，計算幾筆資料超過1的比例即為 P( X > 1)的估計值，重複這個動作直到取得1000筆 P( X > 1)的估計值，最後取這1000筆估計值的平均與變異數，分別為0.2498500與0.000193，此為Hit or miss 估計方法。

(3) Antithetic Variate

做法類似Monte-Carlo Integration，x 從Uniform(0,1)抽取500筆，再將這500筆資料用1 – x的方式生成另外500筆資料，形成1000筆資料。後續步驟雷同Monte-Carlo Integration的步驟，得到P( X > 1)的估計值的平均與變異數為0.24999466與0.00000004，此為Antithetic Variate縮小變異數方法。

(4) Importance Sampling

先選找另一個函數g(x)（此題使用的為），令Y = ，再利用Monte Carlo方法去估計E(Y)。當Y越接近常數，變異數縮減效果越好。因此我們透過此方法，形成1000筆P( X > 1)的估計值，並取這1000筆P( X > 1)的估計值的平均數與變異數為0.24997706與0.00000244，

此為Importance Sampling縮小變異數方法。

(5) Control variate

先找一個與本題函數類似的函數（此題使用g(x) = ），這個函數要能求取期望值，利用  + a( g(x) – E(g(x)))，a = x從uniform(0,1)隨機抽取1000筆，將這1000筆x代入 並取平均即為P(X1 )的估計值，0.5 - 為P( X > 1)的估計值。總共取1000筆P( X > 1)的估計值，並取並取這1000筆P( X > 1)的估計值的平均數與變異數為0.24982097與0.00000983，此為Control variate縮小變異數方法。

(6) Stratified Sampling

做法類似於Monte-Carlo Integration，我們將x的範圍(0,1)切成5等分，即(0, 0.2), (0.2, 0.4), …, (0.8,1)，每做一次P(X 1 )的估計值，要抽取1000個x樣本，這邊將x範圍等分成5等份，因此樣本也平均分配成每一區間各抽取200個樣本並帶入，目的是為了讓樣本的分散程度足夠、代表性足夠，將這5組資料分別取平均後得到   , …, ，再將這5個估計值取平均即為P(X 1 )的估計值 而P( X > 1)的估計值為 0.5 - 。重複上述動作1000次，最後取這1000筆估計值的平均與變異數，分別為0.25001367與0.00000009，此為Stratified Sampling 估計方法。

統整六種方法估計P( X > 1)的估計值樣本平均數與樣本變異數：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Monte-Carlo* | *Hit or miss* | *Antithetic* | *Importance* | *Control* | *Stratified(5)* |
|  | 0.24997910 | 0.2498500 | 0.24999466 | 0.24997706 | 0.24982097 | 0.25001367 |
|  | 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。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Monte-Carlo* | *Hit or miss* | *Antithetic* | *Importance* | *Control* | *Stratified(5)* |
| *with n = 100* | 0.25028315 | 0.24937000 | 0.25002894 | 0.24989915 | 0.24878815 | 0.2500383 |
| *with n = 100* | 0.00002736 | 0.00185055 | 0.00000041 | 0.00002551 | 0.00009001 | 0.0000008 |
| *with n = 1000* | 0.24997910 | 0.2498500 | 0.24999466 | 0.24997706 | 0.25017903 | 0.25001367 |
| *with n = 1000* | 0.00000272 | 0.0001932 | 0.00000004 | 0.00000244 | 0.00000983 | *0.00000009* |
| *with n = 10000* | 0.25001537 | 0.24978600 | 0.2499985 | 0.24997728 | 0.25001918 | 0.25000063 |
| *with n = 10000* | 0.00000025 | 0.00001847 | 0.0000001 | 0.00000024 | 0.00000099 | 0.00000001 |

結論：

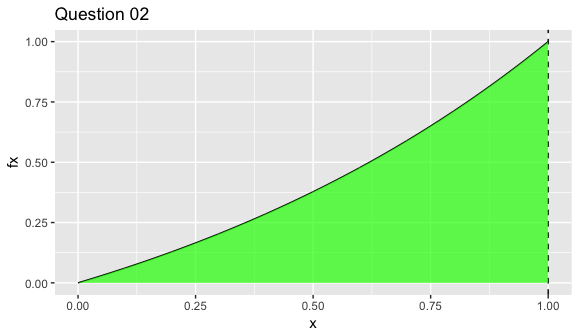
實際 P( X > 1) = 0.25，由上表可知，每個方法都估計的數值相差無幾，只有Antithetic Variate與Stratified Sampling的樣本變異數來得非常小，而Control variate與Hit or miss來得想對大滿多，其中Control variate可能與取得函數有關，若是能取得更適合的函數，便可以將樣本變異數降下來。

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 θ = 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 θ = 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。每一個方法皆取1000個隨機樣本生成一個，再生成1000樣本 ，並運用其平均去估計真實的及計算樣本變異數。

其中本體運用方法方式與第一題類似，而用於Control variate的另一函數為, x = 0 ~ 1，Stratified Sampling將x的範圍0~1平均分為5層並平均分配抽樣數，用於Importance Sampling的另一函數為 。

另外本題亦有多做縮減變異數混搭，在Importance Sampling的方法之中，套入Antithetic Variate抽取x樣本的方法以藉此來縮小只做Importance Sampling的變異數。

統整六種方法估計θ =估計值的樣本平均數與樣本變異數：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Monte-Carlo* | *Antithetic* | *Importance* | *Importance(A)* | *Control* | *Stratified(5)* |
|  | 0.41787975 | 0.41805356 | 0.41824956 | 0.41806185 | 0.41803343 | 0.41804309 |
|  | 0.00008453 | 0.00000281 | 0.00015477 | 0.00004137 | 0.00000134 | 0.00000345 |

結論：

實際 θ = 0.4180233，由上表可知，每個方法估計θ =的數值相差無幾，多落在0.417~0.418附近。接著比較每個方法變異數，以Antithetic Variate*、*Stratified Sampling與Control variate的變異數最小，而Importance Sampling的變異數則是六個之中最大的，可能與Importance Sampling使用的另一函數（）有關，若是能取得更適合的函數，便可以將樣本變異數降下來。在本題中，我們亦有做縮減變異數方法的混搭，透過兩個方法的結合（Importance Sampling與Antithetic Variate），來讓Importance Sampling方法中的變異數進一步縮減，從0.00015477縮減至0.00004137。因此本題以Antithetic Variate*、*Stratified Sampling與Control variate三個方法為優，另外能透夠增加抽樣樣本數來降低所有方法之變異數。

**# Question 03**

Let be independent exponential random variables each with mean 1, and consider the quantity defined by Propose at least three simulation methods to estimate θ and compare their variances.

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以及Stratified Sampling。在使用Monte-Carlo Integration、Antithetic Variate以及Stratified Sampling時，需要的機率分配形式，透過變數變換後，

的機率分配樣子：, y

先計算)的估計值，再透過1 - ) = 的方式取得的估計值。

以Monte-Carlo Integration為例，先從uniform(0,21.6)抽取1000筆樣本，帶入y的機率形式中並乘上21.6後去取平均數即為一個)的估計值，再計算1 - )的估計值即為的估計值。重復此動作1000次，取得1000筆的估計值後，取此1000筆估計值的平均數與變異數，即為0.16960651與0.00018478。

而Antithetic Variate與Stratified Sampling做法類似Question 01的做法，便能取得該方法所估計估計值的平均數與變異數。

Hit or miss方法則是直接從exp(1)分配中抽取5筆資料，並計算的數值，重複取1000筆並計算超過21.6的比例為一個的估計值。重復抽取1000筆的估計值，取其平均數與變異數為0.16906200與0.00015192。

在此以表方式統整四種方法估計 估計值的樣本平均數與樣本變異數：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Monte-Carlo* | *Antithetic* | Stratified(5) | *Hit or miss* |
|  | 0.16810602 | 0.16730692 | 0.16894173 | 0.16816900 |
|  | 0.00019783 | 0.00029715 | 0.00002004 | 0.00013109 |

結論：在本題四種方法中**，**估計的 大多都落在0.168附近，唯獨Antithetic Variate與0.168有一點落差，變異數的部分則是Stratified Sampling最小，Antithetic Variate最大。因此，本題以Stratified Sampling為優。