統計計算與模擬

政治大學統計系余清祥

2024年4月23日～4月30日

第六單元 ：蒙地卡羅積分、縮減變異數

如何整理大數據的分析結果

- 分析大數據的策略與傳統資料分析並無不同，僅在大量、時效(速度)的需求及差異，特別需要資料所屬領域的知識支持，尤其定義研究目的、量化目標變數（標的）。
- 藉由大數據探討某個議題，如同撰寫論文、報告，具備幾個必要的元素。

→即使研究主題相同，因為切入角度、研究素材（資料）、方法理論（研究者專業）等，使得研究方向、甚至研究結論會大異其趣。

報告撰寫的幾個關鍵要素

- 一篇研究報告包含至少三個要素：

<u>**動機與目標**</u>：問題背景及動機、問題的重要性及其影響、具體(或量)研究目標；

<u>**文獻探討**</u>：相關研究方法及參考文獻、現有方法優勢及限制、本篇研究貢獻(區隔)；

<u>**本研究特色**</u>：研究方法及素材、主要研究發現（及其意涵）、本文適用時機及限制。

- 註：研究發現的價值除了創新之外，也希望能夠具有實質意涵及影響。

- Numerical integration

→Also named as "quadrature," it is related to the evaluation of the integral

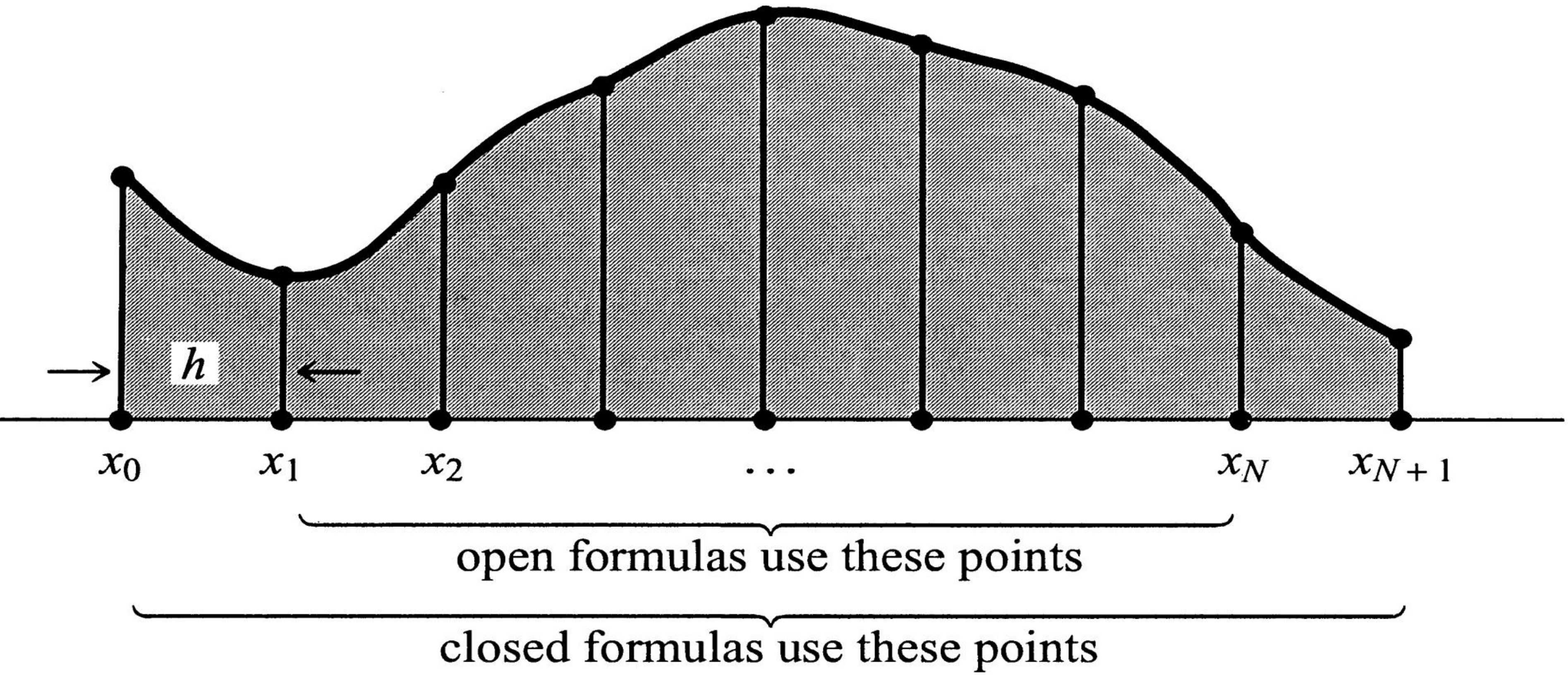
$$I = \int_a^b f(x)dx$$

is precisely equivalent to solving for the value $I \equiv y(b)$ the differential equation

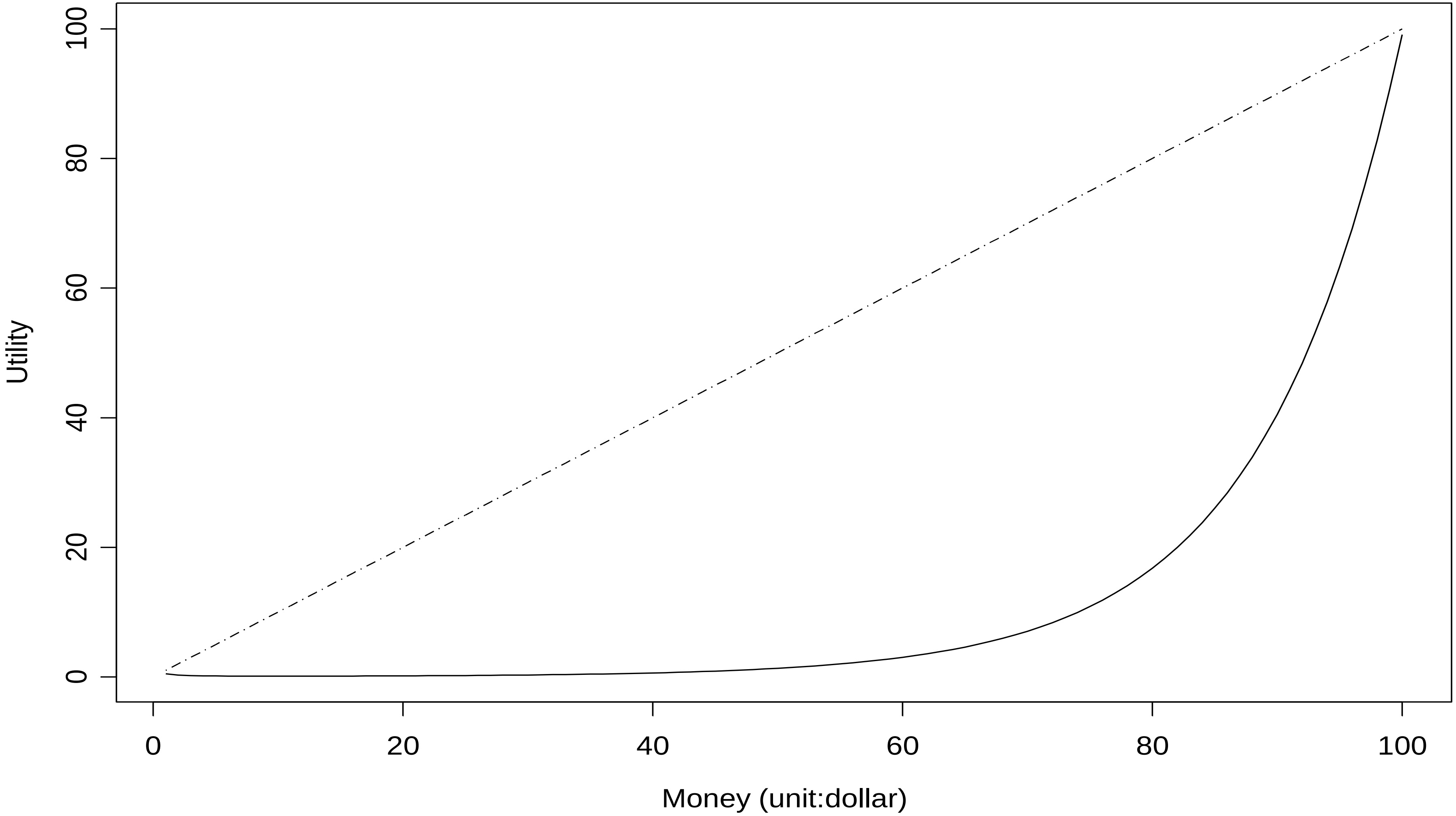
$$\frac{dy}{dx} = f(x)$$

with the boundary condition $y(a) = 0$.

→Classical formula is known at equally spaced steps. The differences are on if the end points, i.e., *a* and *b*, of function *f* are used. If they are (are not), it is called a closed (open) formula. (Only one is used is called semi-open).



Question: What is your intuitive idea of calculating an integral, such as Gini's index (in a Lorenz curve)?



→We shall denote the equal spaced points as $a = x_0, x_1, \ldots, x_N, x_{N+1} = b,$ which are spaced apart by a constant step *h*, i.e.,

$$x_i = x_0 + ih, \quad i = 0, 1, \cdots, N+1.$$

We shall focus on the integration between any two consecutive *x*'s points, i.e., (x_i, x_{i+1}), and the integration between *a* and *b* can be divided into integrating a function over a number of smaller intervals.

- Closed Newton-Cotes Formulas

→ *Trapezoidal rule*:

$$\int_{x_1}^{x_2} f(x)dx = h\left[\frac{1}{2}f(x_1)+\frac{1}{2}f(x_2)\right]+O(h^3 f'').$$

Here the error term *O*( ) signifies that the true answer differs from the estimate by an amount that is the product of some numerical coefficient times h^3 times f''.

In other words, the error of the integral on *(a,b)* using Trapezoidal rule is about $O(n^{-2})$.

→ *Simpson's rule:*

$$\int_{x_1}^{x_3} f(x)dx = h\left[\frac{1}{3}f_1 + \frac{4}{3}f_2 + \frac{1}{3}f_3\right] + O(h^5 f^{(4)}).$$

→ *Simpson's $\frac{3}{8}$ rule:*

$$\int_{x_1}^{x_4} f(x)dx = h\left[\frac{3}{8}f_1 + \frac{9}{8}f_2 + \frac{9}{8}f_3 + \frac{3}{8}f_4\right] + O(h^5 f^{(4)}).$$

- *Extended Trapezoidal rule:*

$$\int_{x_1}^{x_N} f(x)dx = h[\frac{1}{2}f_1 + f_2 + f_3 +$$

$$\cdots + f_{N-1} + \frac{1}{2}f_N] + O\left(\frac{(b-a)^2 f''}{N^2}\right).$$

→ Extended formula of order $1/N^3$:

$$\int_{x_1}^{x_N} f(x)dx = h\Big[\frac{5}{12}f_1 + \frac{13}{12}f_2 + f_3 + f_4$$

$$\cdots + f_{N-2} + \frac{13}{12}f_{N-1} + \frac{5}{12}f_N\Big] + O\left(\frac{1}{N^3}\right).$$

→ Extended Simpson's rule:

$$\int_{x_1}^{x_N} f(x)dx = h\Big[\frac{1}{3}f_1 + \frac{4}{3}f_2 + \frac{2}{3}f_3 + \frac{4}{3}f_4$$

$$\cdots + \frac{2}{3}f_{N-2} + \frac{4}{3}f_{N-1} + \frac{1}{3}f_N\Big] + O\left(\frac{1}{N^4}\right).$$

Variance Reduction

- Since the standard error reduces at the rate of $1/\sqrt{n}$, we need to increase the size of experiment to f^2 if a factor of f is needed in reducing the standard error.

→ However, larger sample size means higher cost in computer.

- We shall introduce methods for reducing standard errors, including Importance sampling, Control and Antithetic variates.

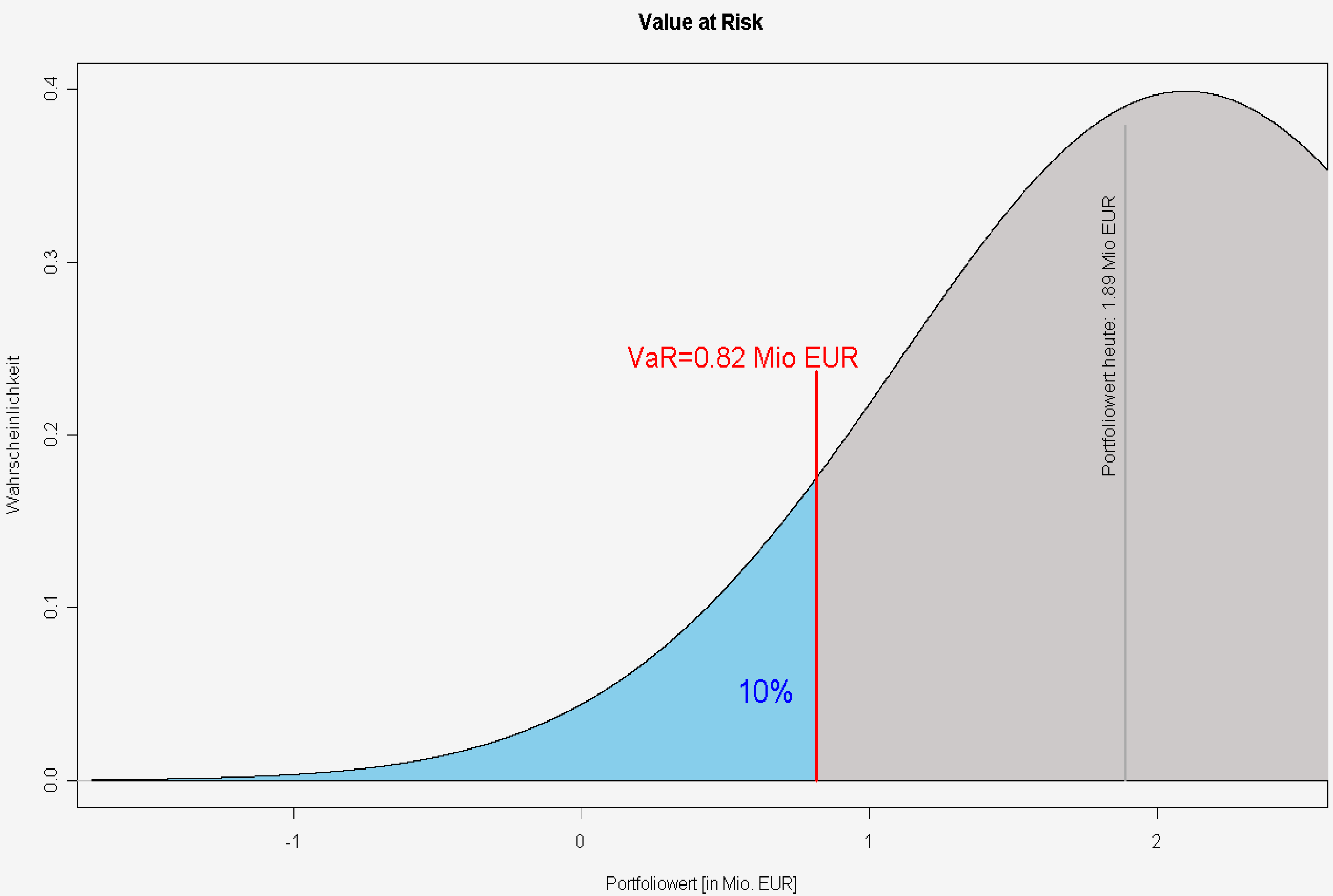
縮減變異數的實例

- Value at Risk (VaR)

→定義$\alpha=\beta-$VaR，表示損失會超過α的機率不超過β。這與統計的信心水準一樣，常見的β值有0.90、0.95、0.99三種數值，在給定機率值β下，找出最小可能的資產值α，或是最大的資產損失α。

註：模擬次數與估計值的精確度成正比，但每次模擬通常需要不少時間，不易由增加模擬次數達成精確度的需求。

10%-value of Risk of a Normally Distributed Portfolio



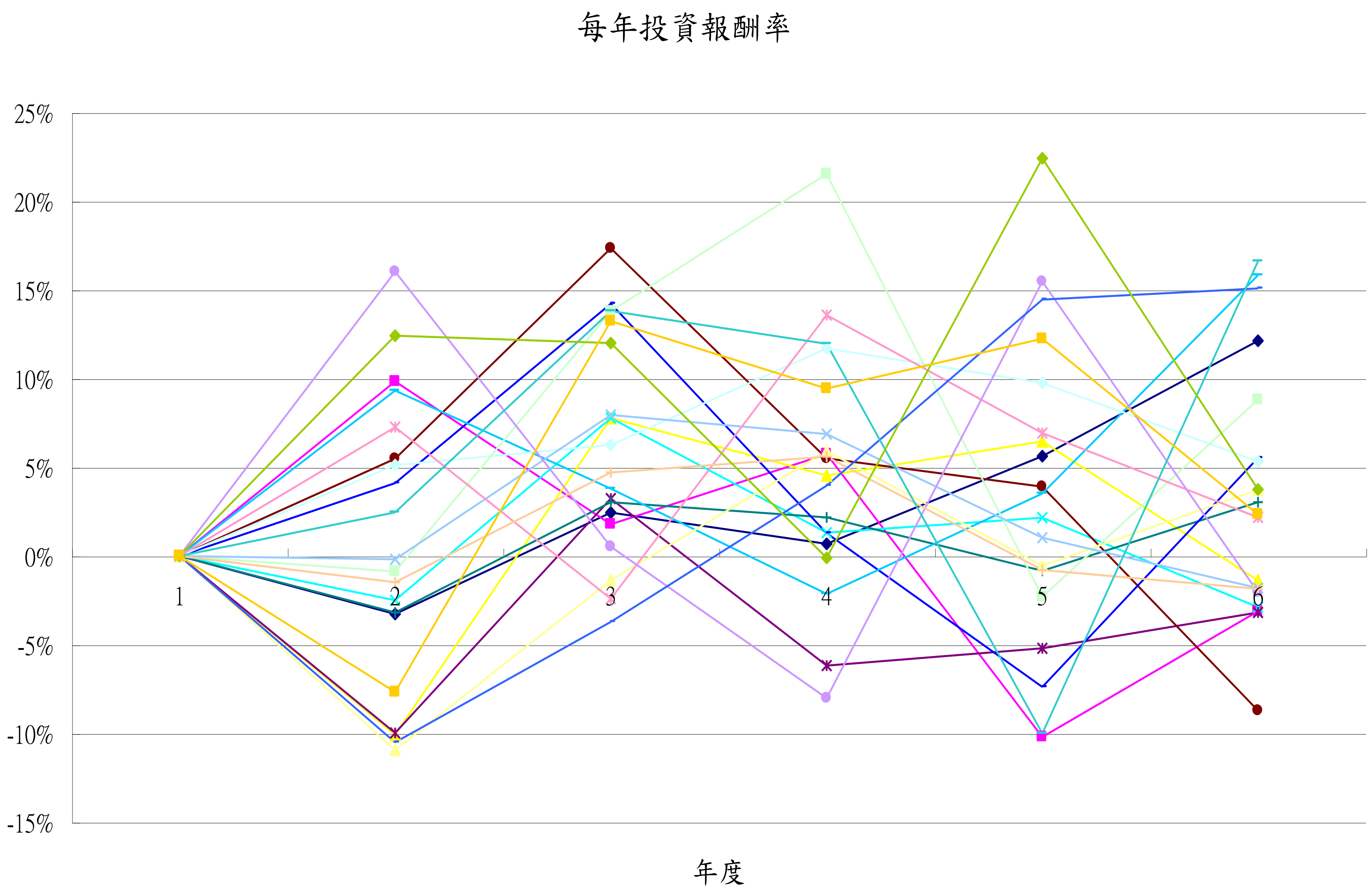
→目標在求得NPV(淨現值)會小於0的機率。如果20次的電腦模擬得出12次的結果小於0，因此機率值的估計值為0.6，計算可得該機率值95%信賴區間約為(0.38, 0.82)。

→如果模擬次數為100次，則95%信賴區間縮至(0.50,0.70)。

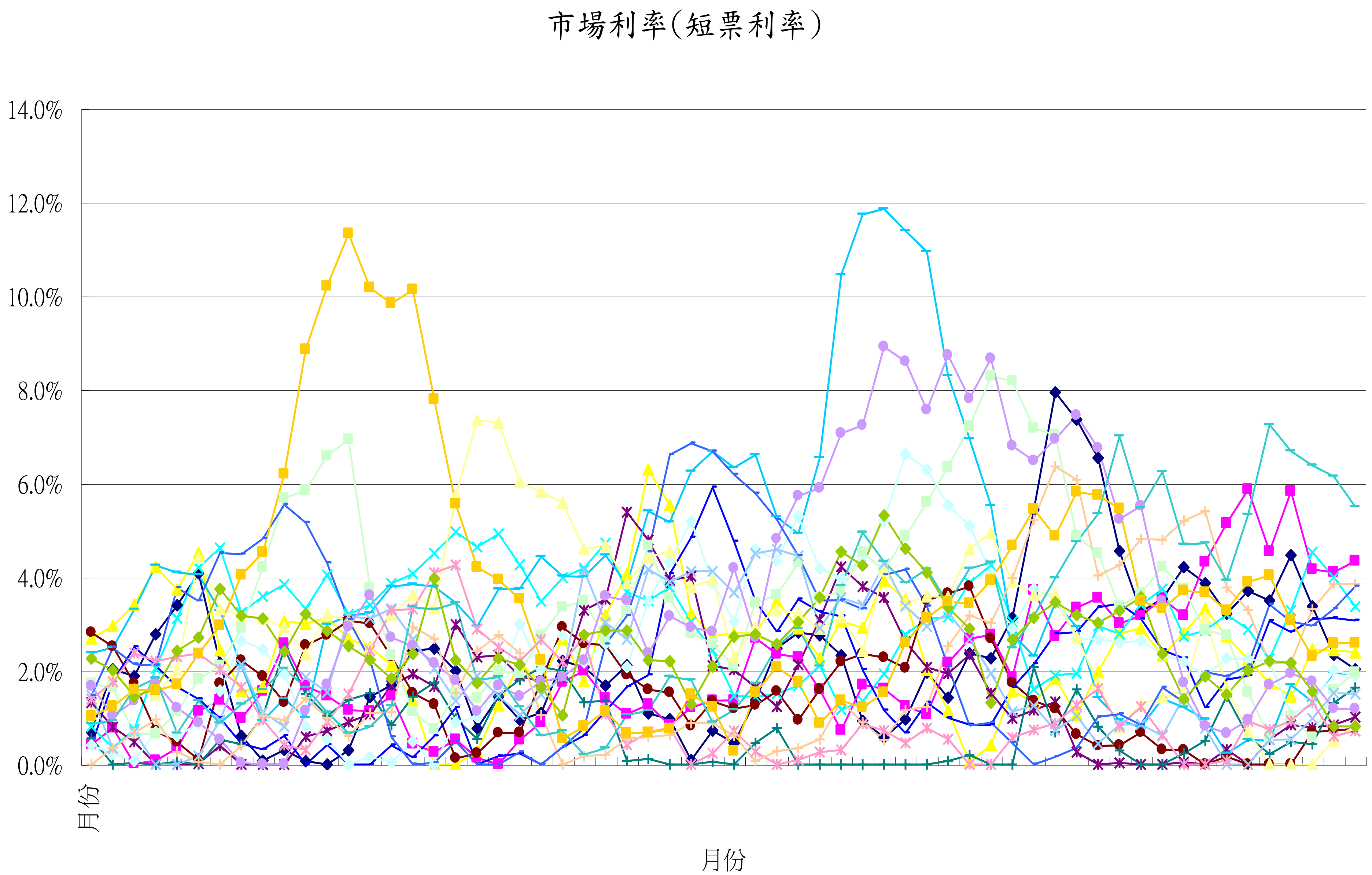
→如果模擬次數為1,000次，則95%信賴區間縮至(0.57,0.63)，似乎精確度已足夠。

註：精確度為1%時，需要10,000次電腦模擬！

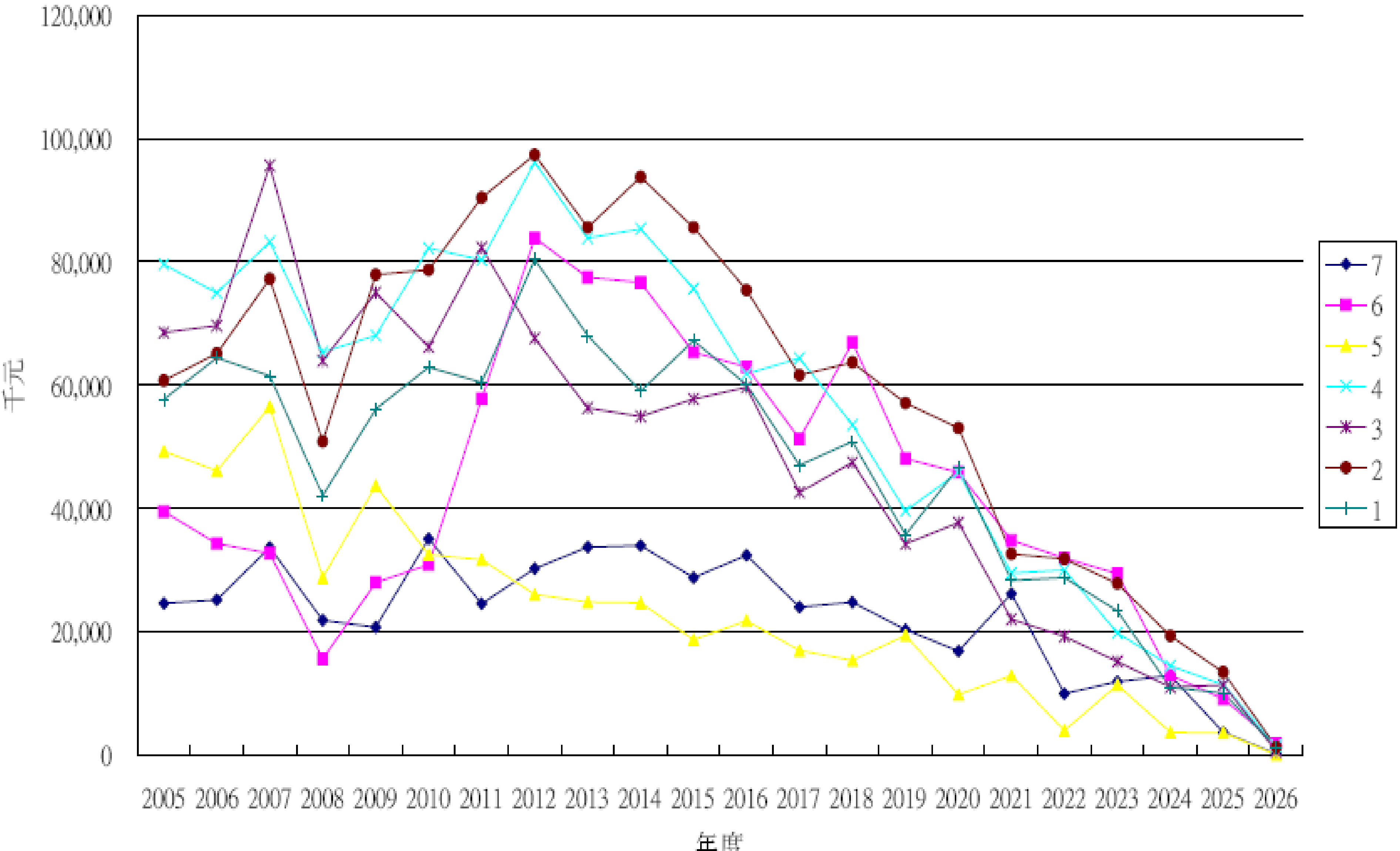
投資情境模擬



投資情境模擬

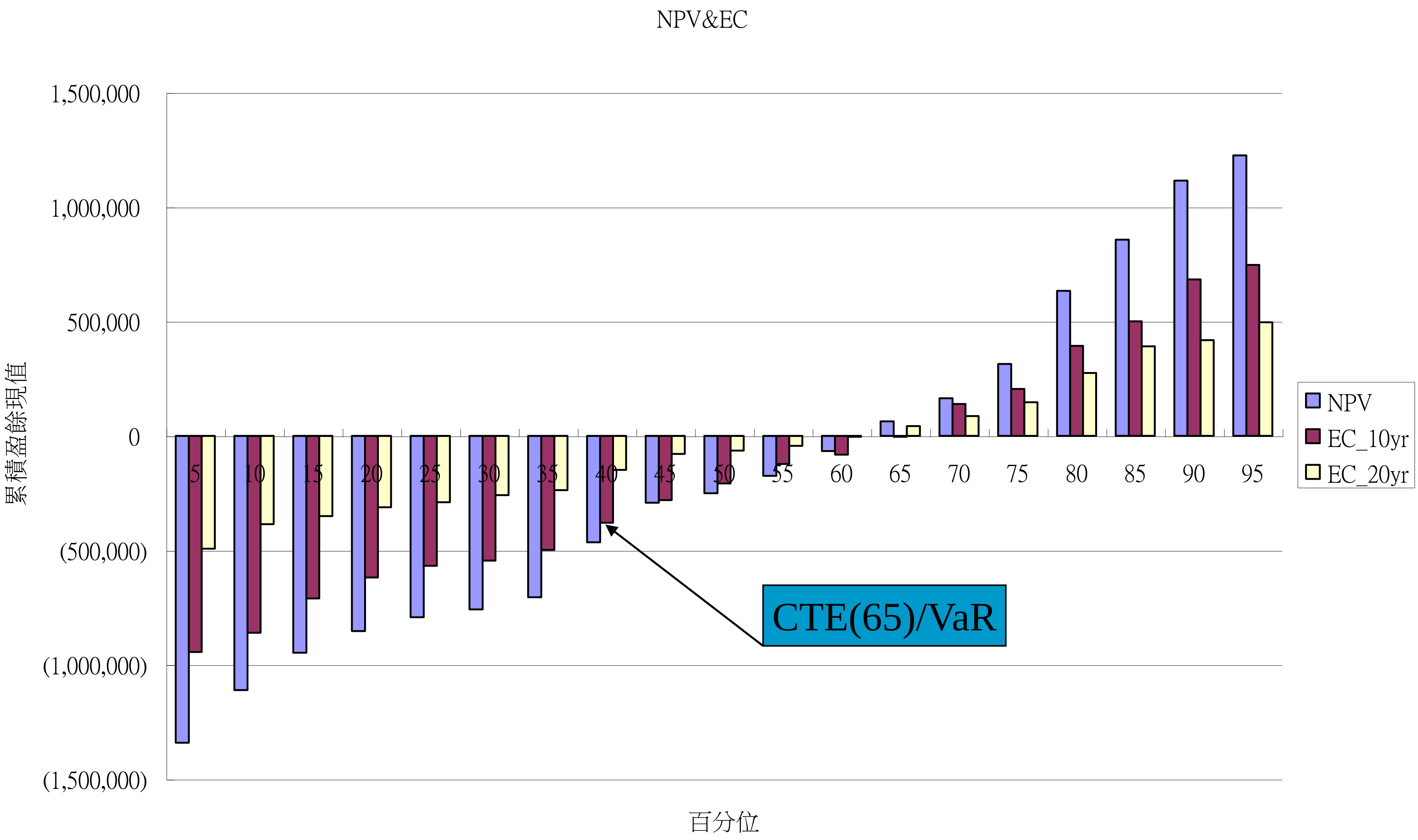


資產負債現金流量預測



註：上述為New York 7，7種最有可能的情境。

依投資情境推算投資報酬與風險、經濟資本額以及風險資本額。



Monte-Carlo Integration

→ Suppose we wish to evaluate $\theta = \int \phi(x) f(x)\,dx$

After sampling $x_1, x_2, \ldots, x_n$ independently from f and form

$$\begin{cases} \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i) \\ Var(\hat{\theta}) = \frac{1}{n}\int [\phi(x) - \theta]^2 f(x)\,dx \end{cases}$$

i.e., the precision of $\hat{\theta}$ is proportion to $1/\sqrt{n}$.

(In numerical integration, n points can achieve the precision of $\mathrm{O}(n^{-4})$.)

Question: We can use Riemann integral to evaluate definite integrals. Then why do we need Monte Carlo integration?

→ As the number of dimensions k increases, the number of points n required to achieve a fair estimate of integral would increase dramatically, i.e., proportional to n^k.

→ Even when the value k is small, if the function to integrated is irregular, then it would inefficient to use the regular methods of integration. (Also, what happen if the range of integration is not close, e.g.$(0,\infty)$)

Numerical vs. Monte Carlo Integration

- Example 1. If C is a Cauchy deviate, then estimate θ = P(C > 2) = 0.5 – P(0 < C < 2).

→ If $X \sim$ Cauchy, i.e., $f(x) = \dfrac{1}{\pi(1+x^2)}$,

$$\theta = 1 - F(2) = \frac{1}{2} - \pi^{-1}\tan 2 \cong 0.1476.$$

→Choosing various values for *n*:

	100	1,000	10,000	100,000
Numerical	0.15013	0.14784	0.14761	0.14759
Monte Carlo	0.12	0.151	0.1460	0.14702

Multivariate Cauchy distribution

A random vector $X = (X_1, \ldots, X_k)^T$ is said to have the multivariate Cauchy distribution if every linear combination of its components $Y = a_1 X_1 + \cdots + a_k X_k$ has a Cauchy distribution. That is, for any constant vector $a \in \mathbb{R}^k$, the random variable $Y = a^T X$ should have a univariate Cauchy distribution.[23] The characteristic function of a multivariate Cauchy distribution is given by:

$$\varphi_X(t) = e^{ix_0(t) - \gamma(t)},$$

where $x_0(t)$ and $\gamma(t)$ are real functions with $x_0(t)$ a homogeneous function of degree one and $\gamma(t)$ a positive homogeneous function of degree one.[23] More formally:[23]

$$x_0(at) = ax_0(t),$$
$$\gamma(at) = |a|\gamma(t),$$

for all t.

An example of a bivariate Cauchy distribution can be given by:[24]

$$f(x, y; x_0, y_0, \gamma) = \frac{1}{2\pi} \left[\frac{\gamma}{((x - x_0)^2 + (y - y_0)^2 + \gamma^2)^{1.5}} \right].$$

History of Monte Carlo Integration

- Modern Monte Carlo method was born in 1940s when Stanislaw Ulam, John von Neumann and others started to use random numbers to examine physics from the stochastic perspective.

→One of the most effective use of Monte Carlo method is to evaluate definite integrals which analytically would be too difficult to find. (For example, the CDF values of normal, t, and other distributions.)

- Example 1. Estimate $\theta = P(C > 2)$, given that C is a Cauchy deviate.

→ There are quite a few ways to estimate θ, depending on the simulation setting.

(1) $n\hat{\theta} \sim B(n,\theta) \Rightarrow Var(\hat{\theta}) = \dfrac{\theta(1-\theta)}{n} \cong \dfrac{0.126}{n}$.

(2) Compute $\theta = \dfrac{1}{2}P(|C|>2)$ by setting

$$\phi(x) = \frac{1}{2}I(|x|>2),\ and$$

$$2n\hat{\theta} \sim B(n,2\theta) \Rightarrow Var(\hat{\theta}) = \frac{\theta(1-2\theta)}{2n} \cong \frac{0.052}{n}.$$

(3) Since $1-2\theta=\int_{-2}^{2} f(x)dx = 2\int_{0}^{2} \frac{dx}{\pi(1+x^2)}$,

let $X_i \sim U(0,2)$, $\phi(x)=2f(x)$, then we can get

$$Var(\hat{\theta}) \cong \frac{0.028}{n}.$$

(4) Let y = $1/x$ in (3), then

$$\theta = \int_{2}^{\infty} \frac{dx}{\pi(1+x^2)} = \int_{0}^{\frac{1}{2}} \frac{y^{-2}dy}{\pi(1+y^{-2})} = \int_{0}^{\frac{1}{2}} f(y)dy.$$

Using transformation on $\phi(y)=\frac{1}{2}f(y)$ and

$Y_i \sim$ U(0,1/2), we have $Var(\hat{\theta}) \cong \frac{9.3\times 10^{-5}}{n}$.

Note: The largest reduction is $Var(\hat{\theta}) \cong \frac{1.1\times 10^{-9}}{n}$.

- Antithetic Variate:

→Suppose Z^* has the same dist. as Z but is negatively correlated with Z. Suppose we estimate θ by $\tilde{\theta} = \frac{1}{2}(Z + Z^*)$. $\tilde{\theta}$ is unbiased and

$$Var(\tilde{\theta}) = [2Var(Z) + 2Cov(Z, Z^*)]/4$$
$$= \tfrac{1}{2}Var(Z)[1 + Corr(Z, Z^*)].$$

If Corr(Z,Z*) < 0, then a smaller variance is attained. Usually, we set $Z = F^{-1}(U)$ and $Z^* = F^{-1}(1-U) \Rightarrow Corr(Z,Z^*) < 0.$

- Note: Consider the integral $\theta = \int_0^1 g(x)dx,$

which is usually estimated by $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} g(U_i),$

while the estimate via antithetic variate is

$$\tilde{\theta} = \frac{1}{2n}\sum_{i=1}^{n}\left[g(U_i) + g(1-U_i)\right].$$

→ For symmetric dist., we can obtain perfect negative correlation. Consider the Bernoulli dist. with $P(Z = 1) = 1 - P(Z = 0) = p$. Then

$$\max\left[\frac{-p}{1-p}, \frac{-(1-p)}{p}\right] \le Corr(Z, Z^*) \le 1.$$

→The general form of antithetic variate is

$$Corr(Z,Z^*) = \frac{(2p-1)-p\times p}{p(1-p)} = \frac{-(1-p)}{p}, \; if \;\; p \geq 0.5.$$

→Thus, it is of no use to let $Z^* = 1 - Z$ if $p \simeq 0$ or $p \simeq 1$, since only a minor reduction in variance is possible.

Note: The idea of antithetic variate is like "duplicating" the number of random numbers.

- Example 1(continued). Use Antithetic variate to reduce variances of (3) and (4).

	(3)	(3)a	(4)	(4)a
Ave.	.1457	.1477	.1474	.1476
Var.	.0286	5.9e-4	9.6e-5	3.9e-6

Note: The antithetic variate on (4) does not reduce as much as the case in control variate.

Example 2. Want to estimate π. We simulate U(0,1) random numbers U_i, $i = 1, 2$, and set

$$I = \begin{cases} 1, & \text{if } U_1^2 + U_2^2 \le 1 \\ 0, & \text{Otherwise.} \end{cases}$$

→The value π can be estimated by 4*E(I) and also by antithetic variate. Simulate 1,000 times, with 10,000 U's, we found that

	Mean	Variance
4*E(I)	3.142468	2.5844e–4
Antithetic	3.141529	9.6713e–5

- Conditioning:

→ In general, we have

$$Var[E(Z\,|\,W)] = Var(Z) - E[Var(Z\,|\,W)] \le Var(Z).$$

→ Example 2 (Conti.). Want to estimate π. We did it before by $V_i = 2U_i - 1,\ i = 1, 2$, and set

$$I = \begin{cases} 1, & if\ V_1^2 + V_2^2 \le 1 \\ 0, & Otherwise. \end{cases}$$

We can improve the estimate $E(I)$ by using $E(I|V_1)$.

$$E[I \mid V_1 = v] = P\{V_1^2 + V_2^2 \le 1 \mid V_1 = v\}$$

$$= P\{v^2 + V_2^2 \le 1 \mid V_1 = v\} = P\{V_2^2 \le 1 - v^2\}$$

$$= \int_{-(1-v^2)^{1/2}}^{(1-v^2)^{1/2}} (\frac{1}{2})\, dx = (1 - v^2)^{1/2}.$$

Thus, $E(I \mid V_1) = (1 - V_1^2)^{1/2}$ has mean π/4 (check this!) and a smaller variance. Also, the conditional variance equals to

$$Var[(1 - V_1^2)^{1/2}] = Var[(1 - U^2)^{1/2}] = E(1 - U^2) - (\frac{\pi}{4})^2$$

$$= \frac{2}{3} - (\frac{\pi}{4})^2 \cong 0.0498.$$

smaller than $Var(I) = (\frac{\pi}{4})(1 - \frac{\pi}{4}) \cong 0.1686.$

- Stratified Sampling:

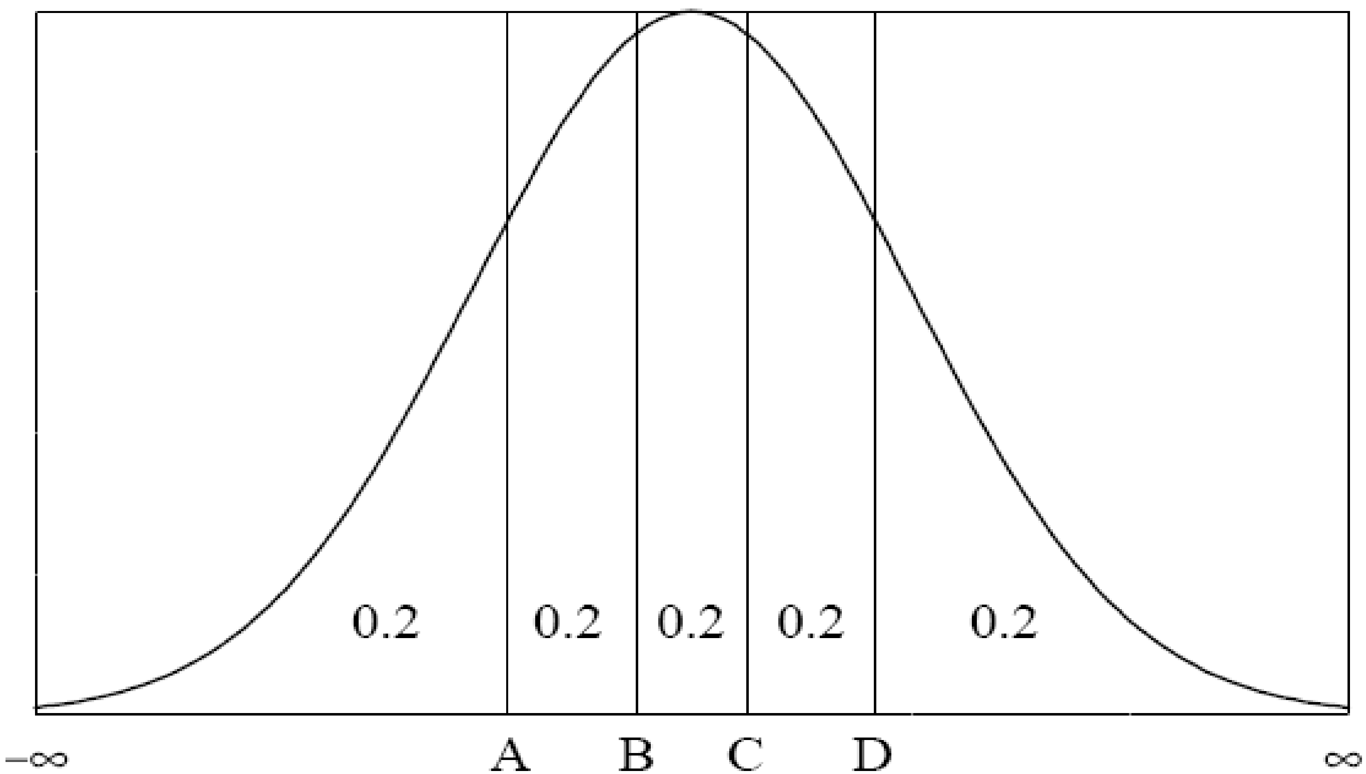
→Without loss of generality, assume the study region is (0,1) and $A_1, A_2, \ldots, A_n$ is a partition of (0,1). Instead of randomly selecting *n* obs. from (0,1), we select one obs. from each interval and have

$$Eg(X) = \sum_{i=1}^{n} E(g(x) \mid x \in A_i) \cdot P(x \in A_i).$$

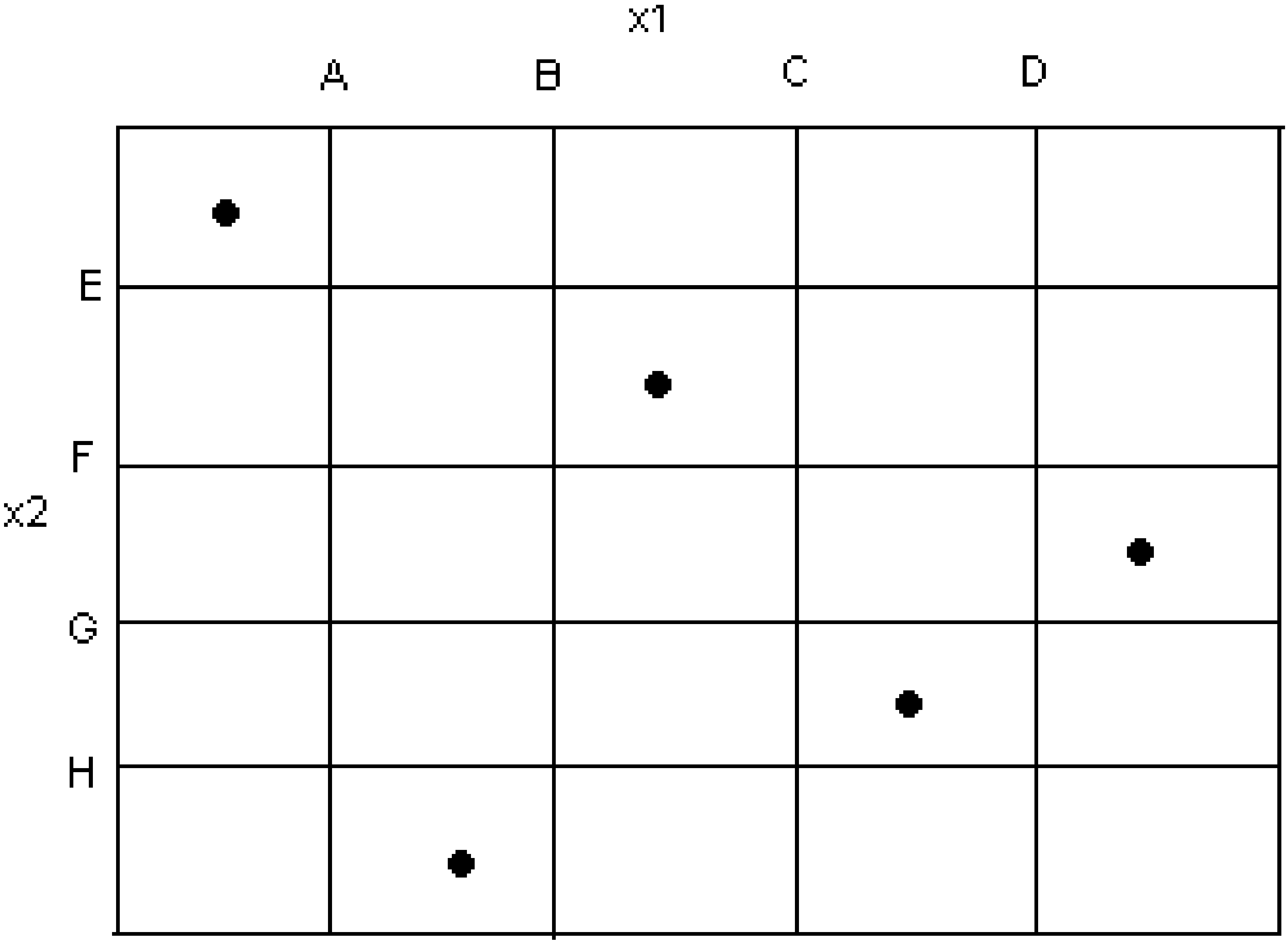
Note: For convenience, let $A_i = (\frac{i-1}{n}, \frac{i}{n})$.

- Latin Hypercube Sampling with Dependence

→This is a type of stratified Monte Carlo sampling. The range of each variable is portioned into N non-overlapping intervals of equal probability 1/N.

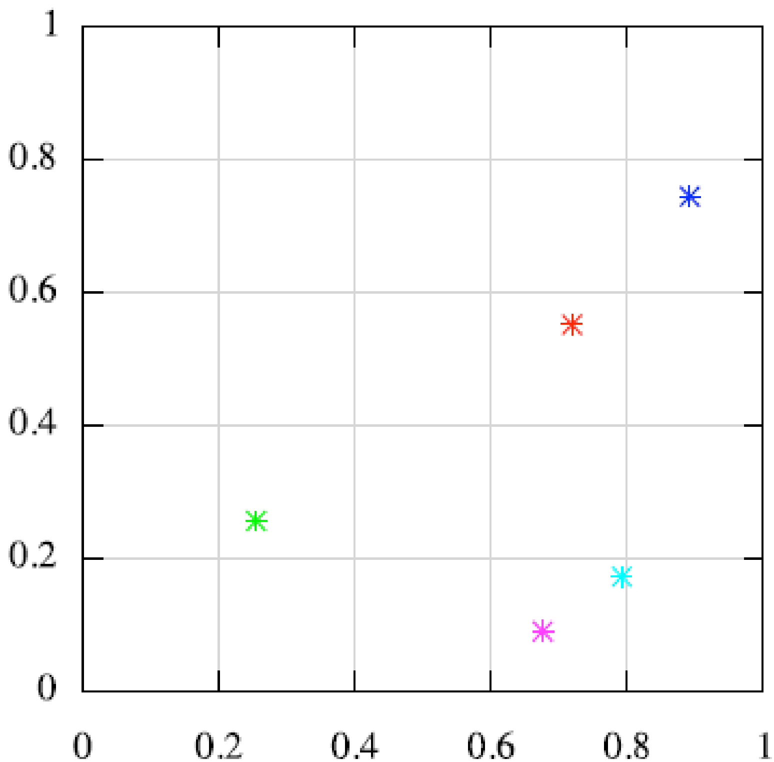


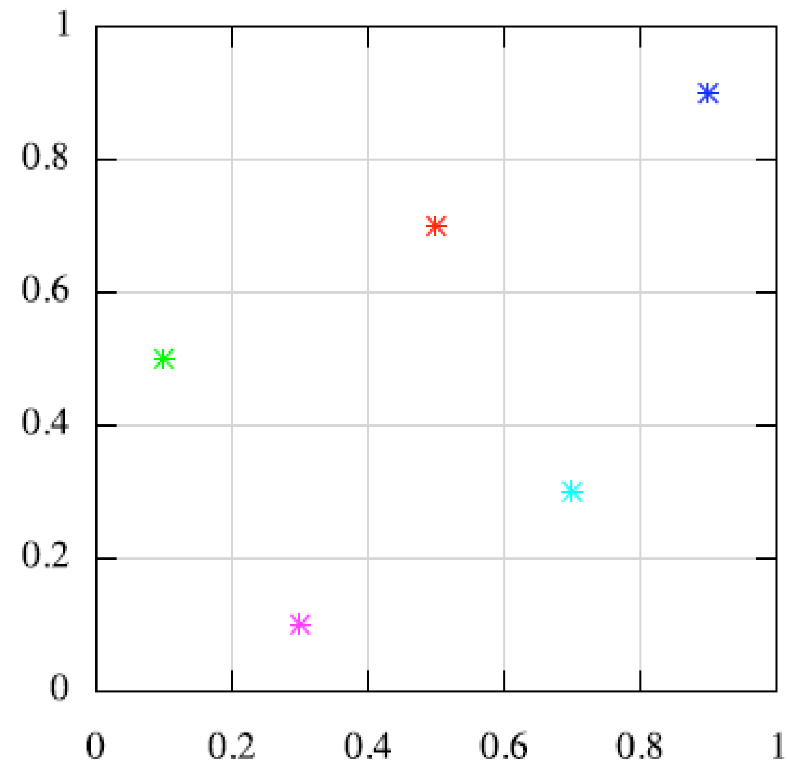
Pairing x1 with x2 when N=5



Latin Hypercube Sampling

- Consider 5 random numbers on $[0,1]^2$, correlated with a Gaussian copula with parameter 0.5:





- Ranks in each dimension: $\{(3{,}4), (1{,}3), (5{,}5), (2{,}1), (4{,}2)\}$

Permutation in each dimension

- Idea: LHS, but instead of choosing a random permutation, choose a particular one.

Definition (Rank statistic)

Let $X_1, \dots, X_n$ be i.i.d. random variables with continuous distribution function. Reorder them such that $X^{(1)} < \dots < X^{(n)}$ $\mathbf{P}$–a.s.. The index of X_i within $X^{(1)}, \dots, X^{(n)}$ is the **i-th rank statistic**, given by

$$r_{i,n}(X_1, \dots, X_n) := \sum_{k=1}^{n} \mathbf{1}_{\{X_k \le X_i\}}.$$

- LHSD sample

$$V_{i,n}^j := \frac{r_{i,n}(U_1^j, \ldots, U_n^j) - 1}{n} + \frac{\eta_{i,n}^j}{n}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, d,$$

  with $\eta_{i,n}^j$ random variables taking values in $[0, 1]$.

- $(V_{1,n}^j, \ldots, V_{n,n}^j)$ is a stratified sample in each dimension j.
- Choices for $\eta_{i,n}^j$:
  - $\eta_{i,n}^j := 1/2$ (just capture joint distribution)
  - ...

- LHS a special case of LHSD.
- LHSD sampling strategy already suggested by [Stein, 1987], but not analysed.

A simple way to get an estimate to the variance of average is called replicated Latin hypercube sampling [7]. The idea is to produce several independent LHS samples and then compare the variances between samples. If the number of samples is α and size of them M, N=$\alpha\cdot$M. If

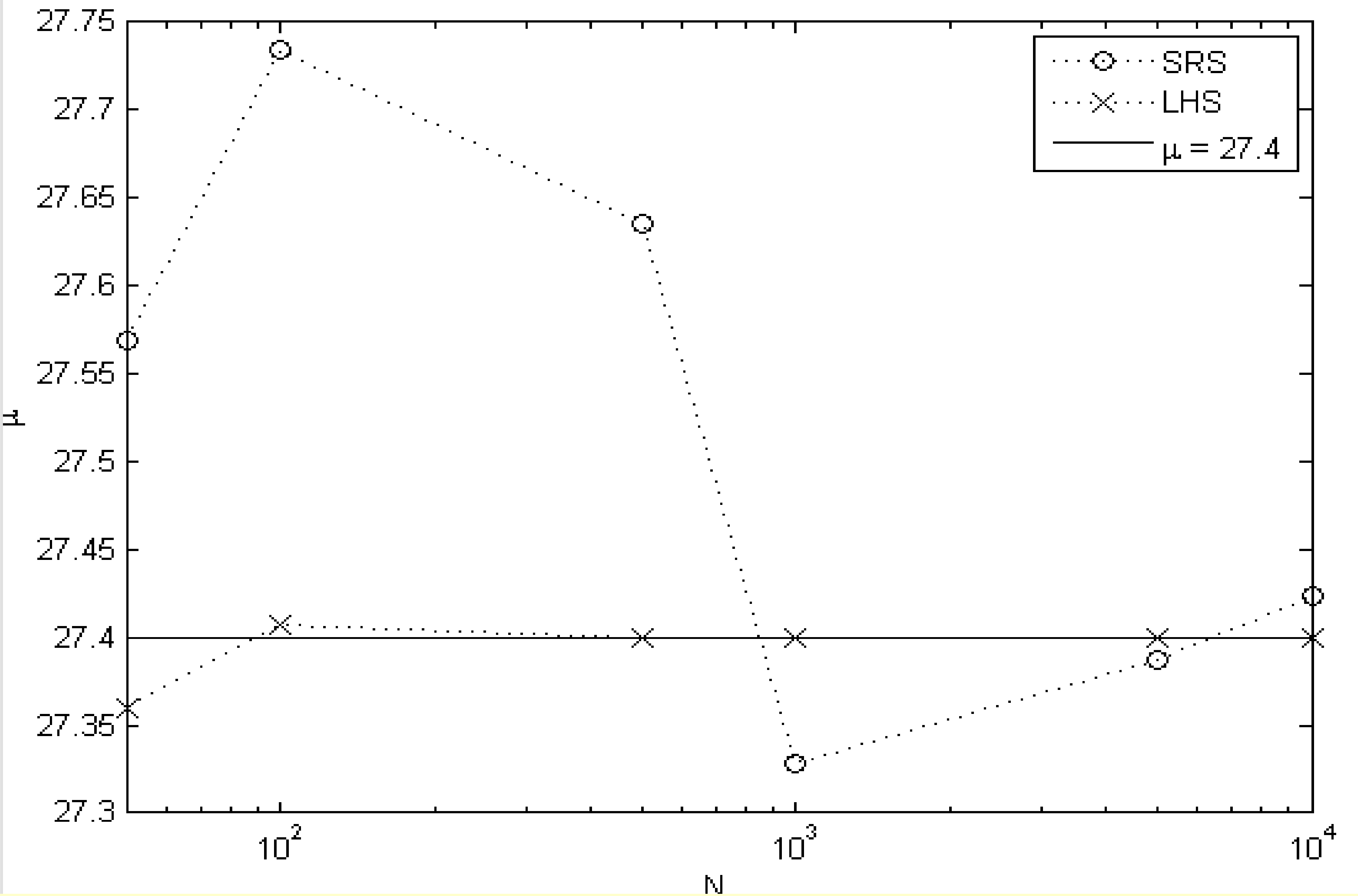
$$\overline{h}_i = M^{-1}\sum_{j=1}^{M} h(\mathbf{X}_j^i), \tag{15}$$

then an unbiased estimator for the variance is

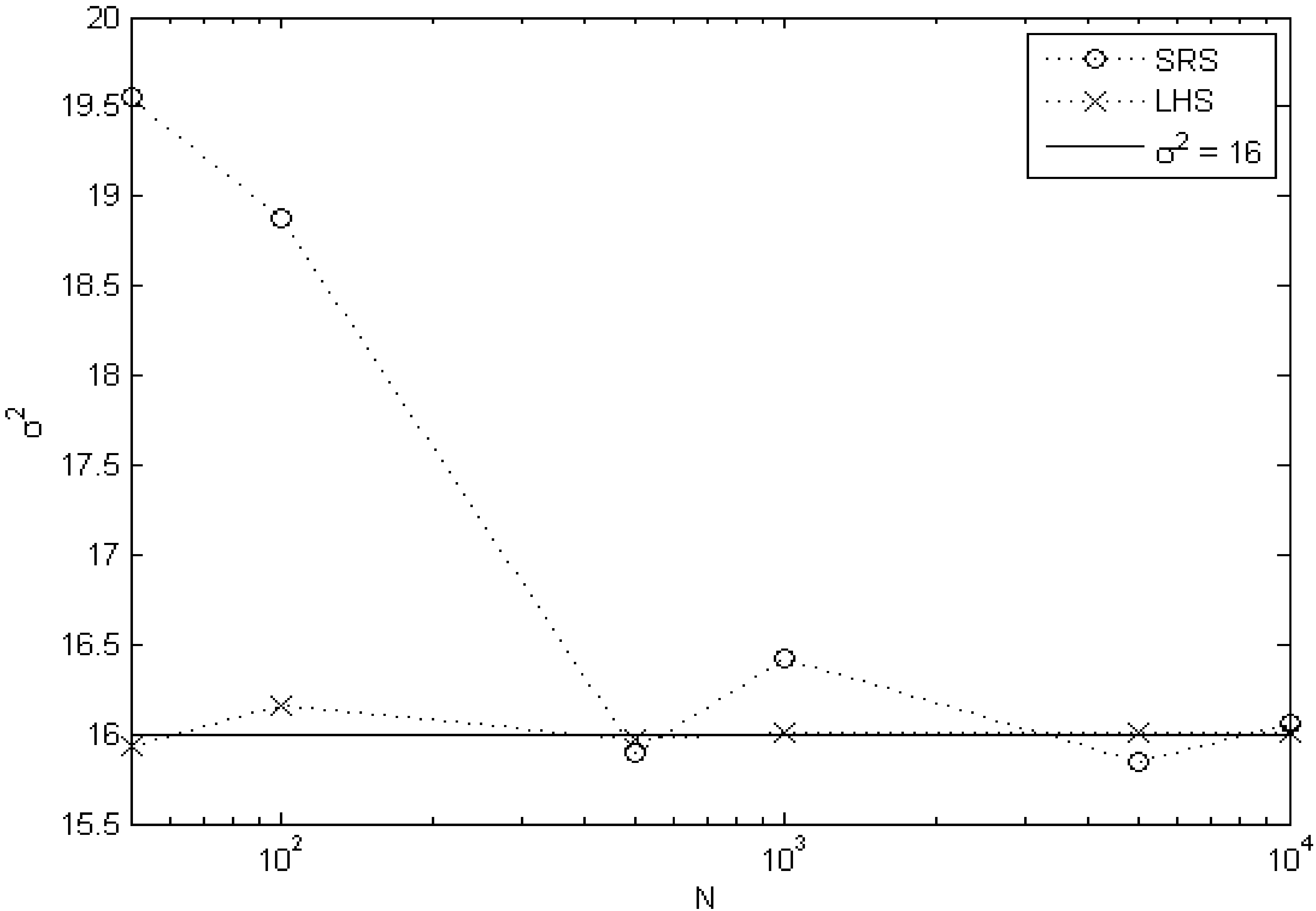
$$\overline{var}(\alpha^{-1}\sum_{i=1}^{\alpha}\overline{h}_i) = \frac{1}{\alpha(\alpha-1)}[\sum_{i=1}^{\alpha}\overline{h}_i^2 - (\frac{\sum_{i=1}^{\alpha}\overline{h}_i}{\alpha})^2]. \tag{16}$$

This is exactly same as the sample variance of mean for normal sampling. Now if α is too small, the estimate of variance is not precise. On the other hand, increasing α while N is fixed increases this estimator of variance. However, as long as ratio $\frac{M}{K}$ is large, the increase in variance is small.

Example. Sampling with Lognormal Distribution (LHS)



Sample means as function of sample size N



Sample Variances as Function of Sample Size N

- Importance Sampling:

→ In evaluating $\theta = E\phi(X)$, some outcomes of X may be more important.

(Take θ as the prob. of the occurrence of a rare event ⇒ produce more rare event.)

→ Idea: Simulate from p.d.f. g (instead of the true f). Let $\psi = \frac{\phi f}{g}$ & $\hat{\theta}_g = \frac{1}{n}\sum \psi(x_i)$, $X_i \sim g$.

Here, $\hat{\theta}_g$ is unbiased estimate of θ, and

$$Var(\hat{\theta}_g) = \frac{1}{n}\int [\psi(X) - \theta]^2 g(x)dx = \frac{1}{n}\int (\frac{\phi f}{g} - \theta)^2 g(x)dx$$

can be very small if $\frac{\phi f}{g}$ is close to a constant.

- Example 1 (continued)

We select g so that $\{g > 0\} = \{|\phi f| > 0\} = \{X > 2\}$.

For x >2, $f(x) = \dfrac{1}{\pi(1+x^2)}$ is closely matched by

$g(x) = \dfrac{2}{x^2}$, i.e., sampling $X = 2/U$, $U \sim$ U(0,1),

and let $\psi(x) = \dfrac{1}{2} \cdot \dfrac{f(x)}{g(x)} = \dfrac{x^2}{2\pi(1+x^2)} = \dfrac{1}{2\pi(1+x^{-2})}$.

By $\hat{\theta}_g = \dfrac{1}{n} \sum \psi(x_i)$, this is equivalent to (4).

We shall use simulation to check.

	(1)	(2)	(3)	(4)
10,000 runs (s.e.)	.1461	.1484	.1474	.1475
	.1248	.1043	.0283	9.6e-5
50,000 runs (s.e.)	.1464	.1470	.1467	.1476
	.1250	.1038	.0285	9.6e-5
100,000 runs (s.e.)	.1455	.1462	.1477	.1476
	.1243	.1034	.0286	9.5e-5
Ratio	1	1.2	4.3	1300

Note: We can see that all methods have unbiased estimates, since $\theta \cong 0.1476$. The format of cells are the estimate (top) and s.e. (bottom).

- Example 3. Evaluate the CDF of standard normal dist.: $\Phi(x)=\int_{-\infty}^{x}\phi(y)dy=\int_{-\infty}^{x}\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy.$

→ Note that $\phi(y)$ has similar shape as the logistic, i.e., $g(y)=\frac{\pi e^{-\pi y/\sqrt{3}}}{\sqrt{3}(1+e^{-\pi y/\sqrt{3}})^2} \rightarrow \mu=0, \sigma^2=1.$

$\therefore \Phi(x)=\int_{-\infty}^{x}\frac{k\phi(y)}{g(y)}\cdot\frac{g(y)}{k}dy,$ k is a normalization constant, i.e., $\frac{g(y)}{k}=\frac{\pi e^{-\pi y/\sqrt{3}}(1+e^{-\pi y/\sqrt{3}})}{\sqrt{3}(1+e^{-\pi y/\sqrt{3}})}.$

$\Rightarrow$ We can therefore estimate $\Phi(x)$ by

$$\hat{\theta} = \frac{1}{n} \cdot \frac{1}{1+e^{-\pi x/\sqrt{3}}} \cdot \sum_{i=1}^{n} \frac{\phi(y_i)}{g(y_i)}.$$

where y_i's is a random sample from $g(y)/k$.

In other words,

$$U \sim U(0,1) \Rightarrow U = F(Y) = \frac{1+e^{-\pi x/\sqrt{3}}}{1+e^{-\pi Y/\sqrt{3}}},$$

$$\therefore Y = -\frac{\sqrt{3}}{\pi} \log\left[(1+e^{-\pi x/\sqrt{3}})U^{-1} - 1\right]$$

x	$\hat{\Phi}(x)$ n=100	$\hat{\Phi}(x)$ n=1000	$\hat{\Phi}(x)$ n=5000	$\Phi(x)$
-2.0	.0222	.0229	.0229	.0227
-1.5	.0652	.0681	.0666	.0668
-1.0	.1592	.1599	.1584	.1587
-0.5	.3082	.3090	.3081	.3085

Note: The differences between estimates and the true values are at the fourth decimal.

- Control Variate:

→ Suppose we wish to estimate $\theta = E(Z)$ for some $Z = \phi(X)$ observed in a process of X. Take another observation $W = \varphi(X)$, which we believe varies with Z and which has a known mean. We can then estimate θ by averaging observations of $Z - (W - E(W))$.

For example, we can use "sample mean" to reduce the variance of "sample median."

- Example 4. We want to find median of Poisson(11.5) for a set of 100 observations.

→ Similar to the idea of bivariate normal distribution, we can modify the estimate as

$$median(X) - \rho(median, mean) \times (sample\ mean - grand\ average)$$

As a demonstration, we repeat simulation 20 Times (of 100 obs.) and get

Median	Ave.=11.35	Var.=0.3184
Control	Ave.=11.35	Var.=0.2294

- Control Variate (continued)

→In general, suppose there are p control variates $W_1, \ldots, W_p$ and Z generally varies with each W_i, i.e.,

$$\hat{\theta} = Z - \beta_1(W_1 - EW_1) - \cdots - \beta_p(W_p - EW_p)$$

and $\hat{\theta}$ is unbiased.

For example, consider $p = 1$, i.e.,

$$Var(\hat{\theta}) = Var(Z) - 2\beta Cov(Z, W) + \beta^2 Var(W)$$

and the min. is attained when $\beta = \dfrac{Cov(Z, W)}{Var(W)}$.

$\Rightarrow$ Multiple regression of Z on $W_1, \ldots, W_p$.

- Example 1(continued). Use control variate to reduce the variance. We can modify (3)

$$\hat{\theta}_3 = \frac{1}{2}-[f(x)+0.15(x^2-\tfrac{4}{3})-0.025(x^4-\tfrac{16}{5})]$$

$$where\ X \sim U(0,2) \Rightarrow Var(\hat{\theta}_3) \cong \frac{6.3\times 10^{-4}}{n}$$

Note: The correlation coefficients of x^2 and x^4 were obtained via the multiple regression model with *f(x)* as the dependent variable and x^2 & x^4 as the independent variables. Also, the values of x can be chosen as *0.01, 0.02, 0.03, …, 1.99, 2.00.*

Question: How do we find the estimates of correlation coefficients between *Z* and *W*'s?

→ For example, in the previous case, the range of x is (0,2) and so we can divide this range into $x = 0.01, 0.02, \ldots, 2$. Then fit a regression equation on $f(x)$, based on x^2 and x^4. The regression coefficients derived are the correlation coefficients between f(x) and x's. Similarly, we can add the terms x and x^3 if necessary.

Question: How do we handle multi-collineaity?

- Example 1(continued).

  Similarly, the modified version of (4) is

$$\hat{\theta}_4 = f(x) + 0.312(x^2 - \tfrac{1}{12}) - 0.233(x^4 - \tfrac{1}{80})]$$

$$\text{where } X \sim U(0, \tfrac{1}{2}) \Rightarrow Var(\hat{\theta}_4) \cong \frac{3.8 \times 10^{-6}}{n}.$$

Note: These two estimates are the control variate version of (3) and (4) in the previous case.

Based on 10,000 simulation runs, we can see the effect of control variate from the following table:

	(3)	(3)c	(4)	(4)c
Ave.	.1465	.1475	.1476	.1476
Var.	.0286	6.2e-4	9.6e-5	1.1e-9

Note: The control variate of (4) can achieve a reduction of 1.1×10^{8} times.

Mortality Rates of the Elderly Male in Taiwan

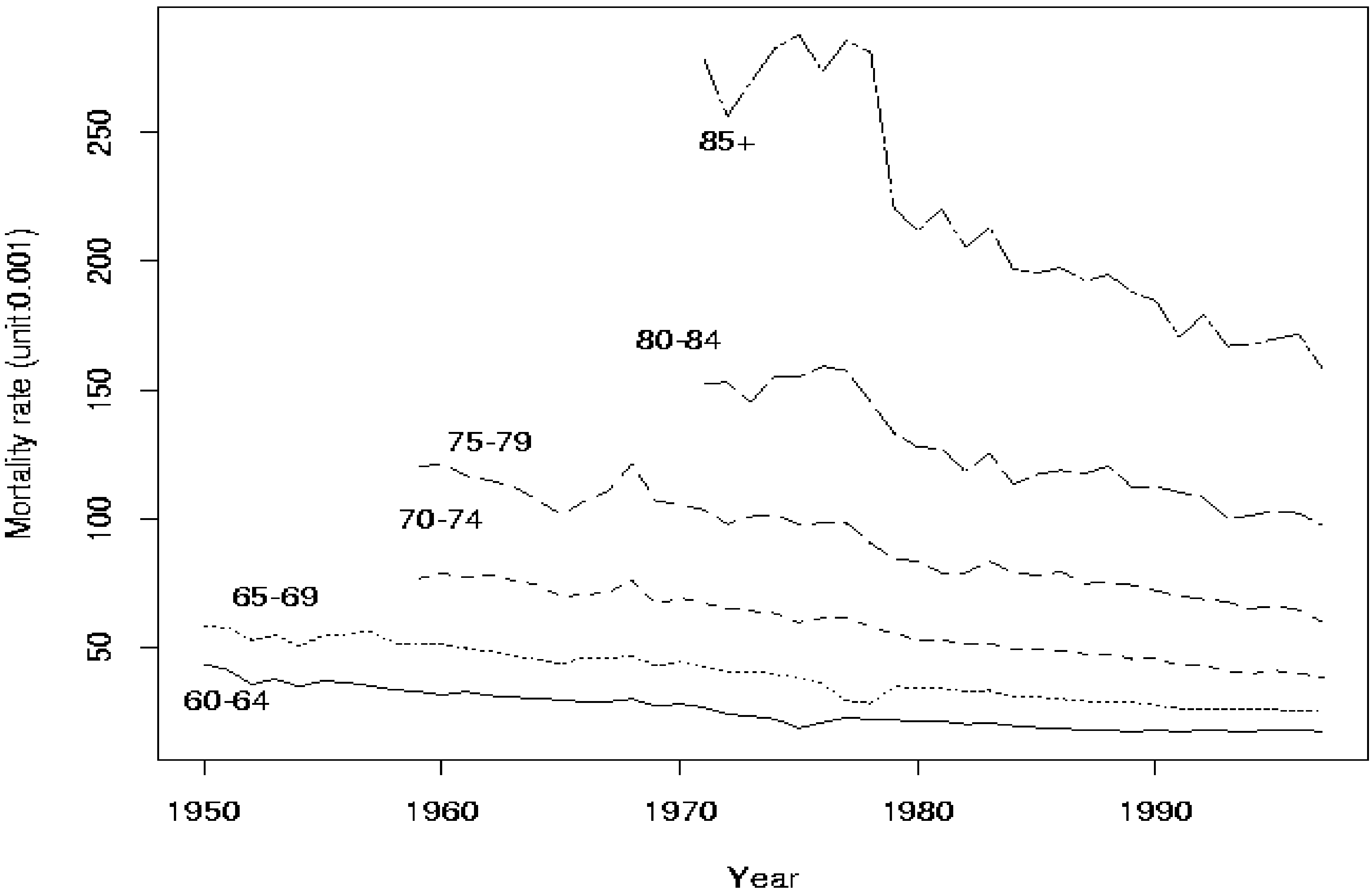


Table 1. Variance Reduction for the Ratio Method

$(\mu_x, \mu_y, \rho, \sigma_x, \sigma_y) = (2, 3, 0.95, 0.2, 0.3)$

	$\hat{\mu}_Y(1)$	$\hat{\mu}_Y(2)$
$n = 34$	3.000408	3.002138
$m = 9$	(.0026533)	(.0021510)
$n = 23$	3.000246	3.000802
$m = 20$	(.0038690)	(.0023129)

$$\hat{\mu}_Y(1) = \sum_{i=1}^{n} y_i / n \qquad \hat{\mu}_Y(2) = \bar{x}_{n+m} \times \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}$$