Chapter 10

隨機亂數的產生與應用

- 瞭解模擬的亂數產生方式
- 將(0,1)亂數轉換成各種機率分佈的隨機亂數
- 產生具有特徵的商業數據
- 合理安排隨機亂數，可降低模擬結果的誤差

What is Randomness? [1]

隨機擲銅板**10**次

- Which is more likely to happen?

  HTTHTHHHTT **HHHHHTTTTT** **HTHTHTHTHT**
- 出現5次正面與5次反面的機率? **0.246**
- 出現7次(含)以上正面或7次(含)以上反面的機率? **0.344**
- 昨天出現7 次正面，今天出現7次(含)以上正面的機率?

最近**10**次打擊表現

- 打擊率三成的選手有0或1支安打的機率？ **0.149**
- 打擊率兩成的選手有4支(含)以上安打的機率？ **0.121**

What is Randomness? [2]

- 你擲銅板**10**次，出現 **HHHHHHHHHH** or **TTTTTTTTTT** 的機率?
- 如果全校所有師生一起擲銅板10次，會有多少人擲出 **HHHHHHHHHH** or **TTTTTTTTTT** ?
- 全校沒有人擲出 **HHHHHHHHHH** or **TTTTTTTTTT** 的機率?
- P(some event will occur) >> P(a specific event will occur)
  P(**40**人的班級至少有兩人的生日為同一天) >**0.5**
- **JFMAMJJASOND MVEMJSUNP**
  聖經密碼：特定頁數的特定位置之字母組合，數以百萬計

What is Randomness? [3]

全校所有師生一起隨機擲銅板10次

- 全校加總後，正面與反面的比例?

最近**100**次打擊表現

- 打擊率三成的選手不到**10**支安打的機率？ **<0.0000004**
- 打擊率兩成的選手有**40**支(含)以上安打的機率？ **<0.000004**

- 因為隨機亂數的變化，模擬過程可能出現極端的情形
- 實驗時間夠長或次數夠多，平均結果仍會接近未知的真實值

隨機亂數產生器的必備條件

- 符合隨機變化的特性
  - Uniform(0,1)亂數必須平均出現在 (0,1)區間各處
  - 經常會出現似乎罕見的情形
- 能迅速產生大量亂數
  - 運算簡易，不影響模擬速度
  - 提供數以億計的亂數
- 能夠控制亂數的產生
  - 可完全重複，可選擇完全不同亂數
- 可通過各種統計檢定
  - 2D與3D繪圖檢驗
  - Chi-square test

10.2 線性同餘產生器

- 選定一個整數Z_0為種子(**seed**)，遞迴運用下列公式：

  Linear Congruential Generator $Z_i = (a\, Z_{i-1} + c)$ **mod** m

- 產生一連串的整數$Z_1, Z_2, Z_3, \ldots$ ，介於0與$m-1$之間。
- Example: $m = 64, a = 21, c = 1$，種子為$Z_0 = 37$

  $(a\, Z_0 + c) = (21\times 37+1) = 778 \rightarrow 778 = 64\times 12+10 \rightarrow Z_1 = 10$
- 將$Z_1, Z_2, Z_3, \ldots$除以m，就轉換成(0,1)亂數。

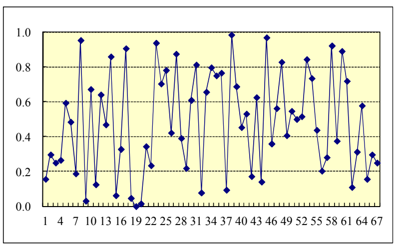
隨機亂數產生器範例

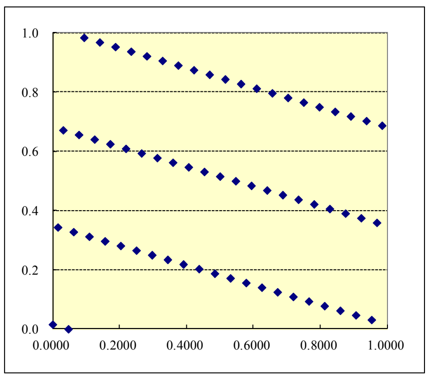
$Z_0=37$	U_i	Z_i	U_i	Z_i	U_i	Z_i	U_i	Z_i	U_i
10	0.1563	55	0.8594	56	0.8750	29	0.4531	54	0.8438
19	0.2969	4	0.0625	25	0.3906	34	0.5313	47	0.7344
16	0.2500	21	0.3281	14	0.2188	11	0.1719	28	0.4375
17	0.2656	58	0.9063	39	0.6094	40	0.6250	13	0.2031
38	0.5938	3	0.0469	52	0.8125	9	0.1406	18	0.2813
31	0.4844	0	0.0000	5	0.0781	62	0.9688	59	0.9219
12	0.1875	1	0.0156	42	0.6563	23	0.3594	24	0.3750
61	0.9531	22	0.3438	51	0.7969	36	0.5625	57	0.8906
2	0.0313	15	0.2344	48	0.7500	53	0.8281	46	0.7188
43	0.6719	60	0.9375	49	0.7656	26	0.4063	7	0.1094
8	0.1250	45	0.7031	6	0.0938	35	0.5469	20	0.3125
41	0.6406	50	0.7813	63	0.9844	32	0.5000	37	0.5781
30	0.4688	27	0.4219	44	0.6875	33	0.5156	10	0.1563

U_i平均值為0.4806，變異數為0.08383

以繪圖方式檢測隨機亂數的品質

$a = 21, c = 1, m = 64$



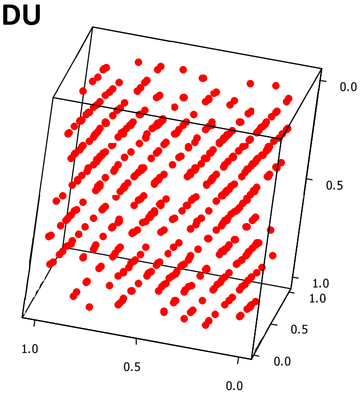


(U_i, U_{i+1})組成二維平面的點座標

三維空間檢驗

FORTRAN的亂數產生子程式**RANDU**

***a* = 2^{16}+3 = 65,539**
***c* = 0**
***m* = 2^{31} = 2,147,483,648**



Chi-square Test

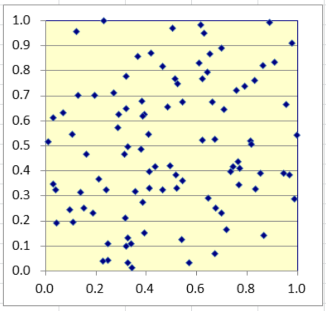
- 產生n個亂數，n>500，將(0,1)區間均分為k個子區間，然後計算落在各子區間的亂數數量f_i，i=1,...,k。
- 計算Chi-Square卡方值： $\chi_{k-1}^{2}=\frac{k}{n}\sum_{i=1}^{k}(f_i-\frac{n}{k})^2$
- 卡方值小，代表亂數是均勻分佈在(0,1)區間。
- Excel的**RAND**函數所產生的亂數，平均而言，約有10%的測試未通過Chi-square test，因此不建議大量使用。
- 微軟公司已經在網頁上承認Excel 2003與Excel 2007的**RAND**函數都有此問題。

早期的ARENA虛擬亂數

$a = 7^5 = 16{,}807$ $c = 0$ $m = 2^{31} - 1 = 2{,}147{,}483{,}647$

A3 =MOD(16807*A2, 2^31-1)

	A	B	C	D	E
1	Seed	Uniform(0,1)			
2	20031015				
3	1653820173	0.77012		0.7701	0.4074
4	874804490	0.4073626		0.4074	0.5430
5	1166016068	0.5429685		0.5430	0.6723
6	1443776001	0.6723106		0.6723	0.5241
7	1125521354	0.5241117		0.5241	0.7457
8	1601433902	0.7457258		0.7457	0.4131
9	887043063	0.4130616		0.4131	0.3266
10	701282367	0.32656		0.3266	0.4948
11	1062487433	0.4947593		0.4948	0.4190
12	899761626	0.4189842		0.4190	0.8667
13	1861289655	0.8667305		0.8667	0.1401
14	300945736	0.1401388		0.1401	0.3125
15	670996267	0.312457		0.3125	0.4646
16	997629072	0.4645572		0.4646	0.8135
17	1746980975	0.8135014		0.8135	0.5182
18	1112825041	0.5181995		0.5182	0.3797
19	815382364	0.379692		0.3797	0.4835
20	1038240241	0.4834683		0.4835	0.6515



Simple LNG LNG Test Arena Test RANDU Distribution Warehouse

The Current (2000) Arena RNG

- **Combined multiple recursive generator**

$A_n = (1403580\, A_{n-2} - 810728\, A_{n-3}) \bmod 4294967087$

$B_n = (527612\, B_{n-1} - 1370589\, B_{n-3}) \bmod 4294944443$

將兩組亂數組合成為

$Z_n = (A_n - B_n) \bmod 4294967087$

$$U_n = \begin{cases} Z_n / 4294967088 & \text{if } Z_n > 0 \\ 4294967087 / 4294967088 & \text{if } Z_n = 0 \end{cases}$$

Seed = a six-vector of first three A_n's, B_n's

10.3 隨機亂數的轉換

$U_i \sim$Uniform(0,1)亂數

If $U_i < p$,	$X = 1$
else,	$X = 0$
If $U_i > 1-p$,	$X = 1$
else,	$X = 0$

$X=1,\ldots, n \qquad P(X=j)=p_j$

$$0 < U_i < p_1 \Rightarrow X = 1$$
$$p_1 < U_i < p_1+p_2 \Rightarrow X = 2$$
$$\vdots$$
$$p_1+\cdots+p_{n-1} < U_i < p_1+\cdots+p_n \Rightarrow X = n$$

產生兩個骰子的點數和

將(0,1)亂數轉換成擲骰子的點數和

方法1: 將(0,1)分割成11個區間，每個區間長度對應到一種結果的機率，例如 (0, 1/36)對應點數2，(1/36, 3/36)對應點數3… (35/36, 1)對應點數12。

方法2: 產生兩個(0,1)亂數U_1, U_2，點數之和就等於

$$\lfloor 6\times U_1 \rfloor+1+\lfloor 6\times U_2 \rfloor+1$$

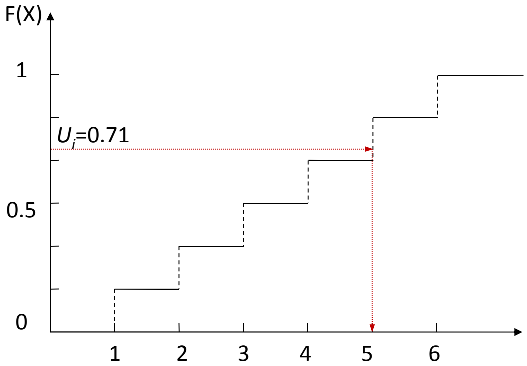
Inverse Transform

$0 \le F(x) \le 1 \Rightarrow U_i = F(x) \Rightarrow x = F^{-1}(U_i)$

P(X=x)=1/6, x=1,…,6

CDF
F(x)=P(X≤x)=⌊x⌋/6,

反函數
$F^{-1}(y)=\lfloor 6\times y \rfloor+1,\ 0<y<1$



Inverse Transform

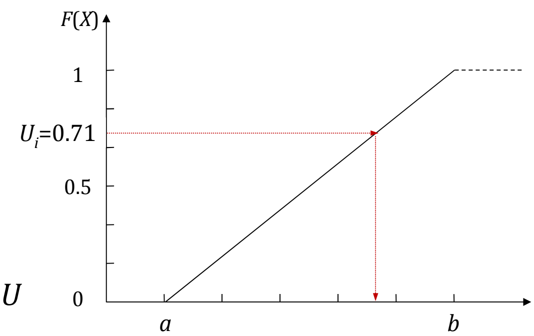
$0 \le F(x) \le 1 \Rightarrow U_i = F(x) \Rightarrow x = F^{-1}(U_i)$

$$F(x) = \frac{x-a}{b-a}, a \le x \le b$$

$$U = F(x) = \frac{x-a}{b-a}$$

$$(b-a)U = x - a$$

$$x = \mathrm{F}^{-1}(U) = a + (b-a)U$$



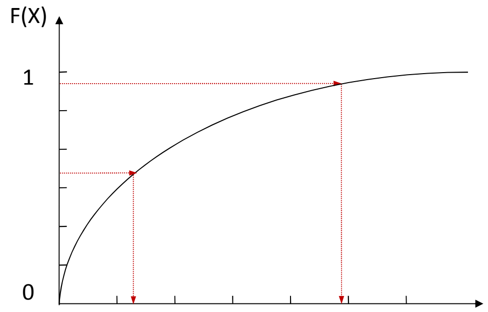
Inverse Transform of Exponentials

$U = F(x) = 1 - e^{-\lambda x}$

$e^{-\lambda x} = 1 - U$

$-\lambda x = \ln(1 - U)$

$x = -(1/\lambda)\ln(1 - U)$



Normal Random Numbers

常態分佈的CDF無法表成函數形式，不適用Inverse Transform

Marsaglia polar method：先產生兩個(0,1)亂數U_1 and U_2

$$Z_1 = \sqrt{-2 \cdot \ln(U_1)} \cdot \cos(2\pi \cdot U_2)$$

$$Z_2 = \sqrt{-2 \cdot \ln(U_1)} \cdot \sin(2\pi \cdot U_2)$$

Z_1 and Z_2為相互獨立的N(0,1)亂數

$\Rightarrow N_i = \sigma \cdot Z_i + \mu$成為N($\mu$, σ^2)的亂數

Bivariate Normal Distribution

Z_1 and Z_2為相互獨立的N(0,1)亂數

$\Rightarrow N_1 = \sigma_1 \cdot Z_1 + \mu_1$為N($\mu_1$, σ_1^2)的亂數

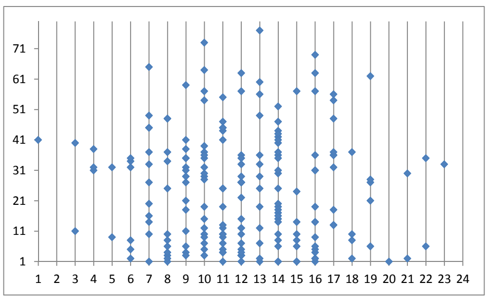
$\Rightarrow N_2 = \sigma_2(\rho\, Z_1 + \sqrt{1-\rho^2}\, Z_2) + \mu_2$為N($\mu_2$, σ_2^2)亂數

N_1 and N_2的相關性為ρ

以亂數模擬隨機現象

模擬物流中心的出貨訂單內容，包括商品儲位與數量，可用來測試揀貨作業方式的效率。

假設某一物流中心的揀貨區共有24排，每排有80個儲位，熱門品項多安排於中間排且靠前方的儲位



以亂數模擬橋牌牌局

East		South		West		North	
0.001838	♥Q	0.007512	♥4	0.013617	♥8	0.043829	♠6
0.061727	♠4	0.061862	♣3	0.066119	♠9	0.094545	♥9
0.109957	♠7	0.113664	♦A	0.117707	♣7	0.123104	♥5
0.185314	♣A	0.188955	♦4	0.212930	♣9	0.218418	♠A
0.257578	♠8	0.260303	♥3	0.267132	♦J	0.299329	♣6
0.350752	♥10	0.351629	♦K	0.361529	♣10	0.367027	♣Q
0.396988	♦2	0.401306	♠2	0.415307	♠10	0.449539	♠3
0.453794	♦7	0.561662	♦9	0.561695	♠J	0.574366	♣K
0.597917	♦5	0.617205	♣J	0.633966	♠5	0.692066	♦10
0.710781	♣2	0.714471	♣8	0.754673	♥A	0.761492	♦3
0.797287	♥K	0.822887	♦Q	0.824660	♣4	0.825003	♣5
0.829509	♠Q	0.840847	♥6	0.859097	♥7	0.861216	♦8
0.897504	♥J	0.911977	♦6	0.912484	♥2	0.956318	♠K

10.5 變異數降低技術

- 亂數轉換自 *U*~U(0,1)，1-*U*~U(0,1)也可轉換成亂數，可能有相反的影響。
- Antithetic Variates是以互為相反的亂數進行兩次模擬，如果使用原本亂數的模擬結果偏高，使用相反亂數的模擬結果可能偏低，平均起來會獲得較準確的結果。
- AV適用於以多次 replications評估個別系統績效的模擬實驗。
- Common Random Numbers透過亂數的同步運用，讓兩種系統在幾乎完全相同的環境下進行模擬，即使是細微的績效差距都是來自於系統的差異。CRN適用比較兩個系統差異。

Comparing Means

以模擬評估兩個系統的營運成本

System A: 22278 ± 360, or [21918, 22638]
System B: 21763 ± 284, or [21479, 22047]

可信賴區間相互重疊，只能判定兩者無明顯差異

	A	B
比賽1	9.8	9.7
比賽2	9.9	9.85
比賽3	9.6	9.55
比賽4	9.65	9.6
比賽5	9.7	9.65

Paired *t*-test

假設 x_i 與 y_i 是相同環境下，兩個隨機現象的觀察值，而x_{i+1}與y_{i+1}是另一組相同環境下的觀察值，我們計算成對觀察到的數值之差距，$x_i - y_i = z_i$

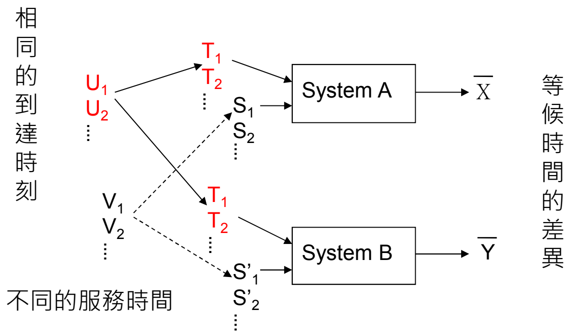
Paired *t*-test System 1 - System 2 = difference

$$\left.\begin{array}{ccccc} x_1 & - & y_1 & = & z_1 \\ x_2 & - & y_2 & = & z_2 \\ & \vdots & & & \vdots \\ x_n & - & y_n & = & z_n \end{array}\right\} \text{Compute confidence interval}$$

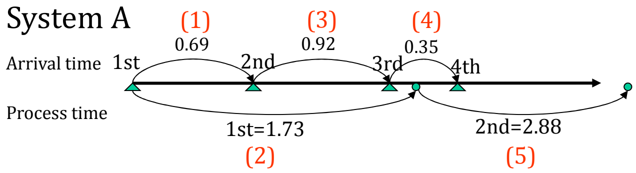
計算Z_i's的平均值與標準差，並建立可信賴區間，如果區間不涵蓋0，代表兩個隨機現象有顯著差異。

Common Random Numbers (CRN)

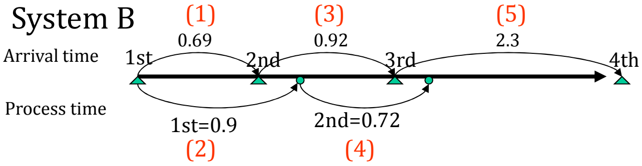
- CRN透過亂數的同步運用，讓兩種系統使用完全相同的(0,1)亂數以進行模擬，任何細微的績效差距都來自系統的差異。



Comparison w/o Synchronization



U(0,1) Expo(1)
(1) 0.5 → 0.69
(3) 0.6 → 0.92
(4) 0.3 → 0.35
U(0,1) Expo(0.8)
(2) 0.75 → 1.73
(5) 0.9 → 2.88



U(0,1) Expo(1)
(1) 0.5 → 0.69
(3) 0.6 → 0.92
(5) 0.9 → 2.3
U(0,1) Unif(0.6,1.0)
(2) 0.75 → 0.9
(4) 0.3 → 0.72

多廠區訂單分配問題

- 假設A夥伴與外商合作改善Shanghai廠的生產問題，處理緊急訂單額外所需時間會從8%降至4%。
- 水果公司希望能降低緊急訂單的流程時間，A夥伴希望能爭取到更多的訂單

Example 7-3: No CRN, 30 replications

	改善前(8%)	改善後(4%)
Product 1 Time	2.468±0.029	2.449±0.030
Product 2 Time	3.527±0.040	3.500±0.044
Order Ratio for A	0.623±0.017	0.614±0.017

CRN亂數設定

每個隨機來源使用不同組別的亂數，確保兩次實驗的訂單資訊完全相同

隨機來源	亂數設定
一般訂單的到達	expo(12,**1**)
一般訂單的廠別	disc(0.2,1,0.4,2,0.6,3,0.8,4,1.0,5,**2**)
一般訂單的處理時間 regular process time	tria(1,2,3,**3**) tria(1,2,3,**4**) tria(1,2,3,**5**) unif(1.5,3.5,**6**) tria(1,2,3,**7**)
Product 1緊急訂單的到達	expo(5,**8**)
Product 1緊急訂單的大小	tria(1.2,2.0,2.6,**9**)
Product 2緊急訂單的到達	expo(7,**10**)
Product 2緊急訂單的大小	tria(2.4,2.8,3.6,**11**)

使用CRN的績效值比較

w/o CRN

	改善前(8%)	改善後(4%)
Product 1 Time	2.468±0.029	2.449±0.030
Product 2 Time	3.527±0.040	3.500±0.044
Order Ratio for A	0.623±0.017	0.614±0.017

Example 10-2 Example 10-3

w. CRN

	改善前(8%)	改善後(4%)
Product 1 Time	2.459±0.038	2.425±0.028
Product 2 Time	3.525±0.045	3.487±0.040
Order Ratio for A	0.638±0.020	0.632±0.015

Statistics模組設定輸出資料檔

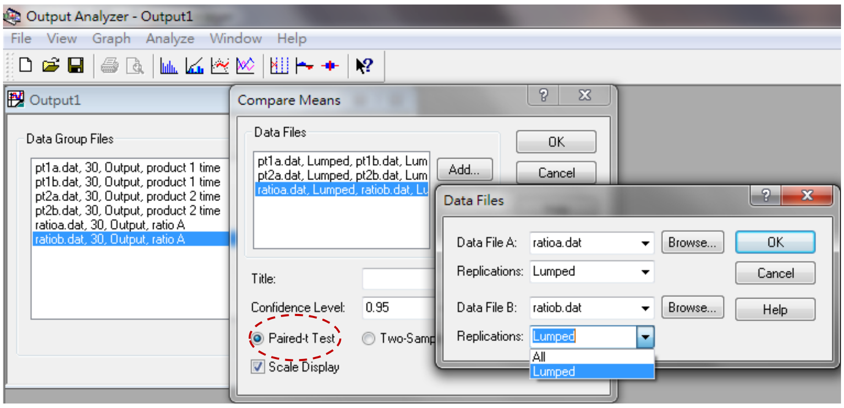
Statistic - Advanced Process

	Name	Type	Expression	Report Label	Output File
1	product 1 time	Output	TAVG(product 1.TotalTime)	product 1 time	pt1a.dat
2	product 2 time	Output	TAVG(product 2.TotalTime)	product 2 time	pt2a.dat
3	ratio A	Output	NC(Record A) / (NC(Record A) + NC(Record B))	ratio A	ratioa.dat
4	avg order 1 time	Output	TAVG(Order 1.TotalTime)	avg order 1 time	
5	avg order 2 time	Output	TAVG(Order 2.TotalTime)	avg order 2 time	
6	avg order 3 time	Output	TAVG(Order 3.TotalTime)	avg order 3 time	
7	avg order 4 time	Output	TAVG(Order 4.TotalTime)	avg order 4 time	
8	avg order 5 time	Output	TAVG(Order 5.TotalTime)	avg order 5 time	

- 改善前的結果: pt1a.dat, pt2a.dat, ratioa.dat
- 改善後的結果: pt1b.dat, pt2b.dat, ratiob.dat
- 輸出模擬結果，使用Output Analyzer進行Paired *t*-test

Paired-t test的設定

- Analyze > Compare Means
- Add…
- Replications → Lumped



使用CRN進行Paired-t test