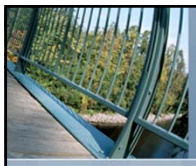
統計の分析と利用

乱数とシミュレーション

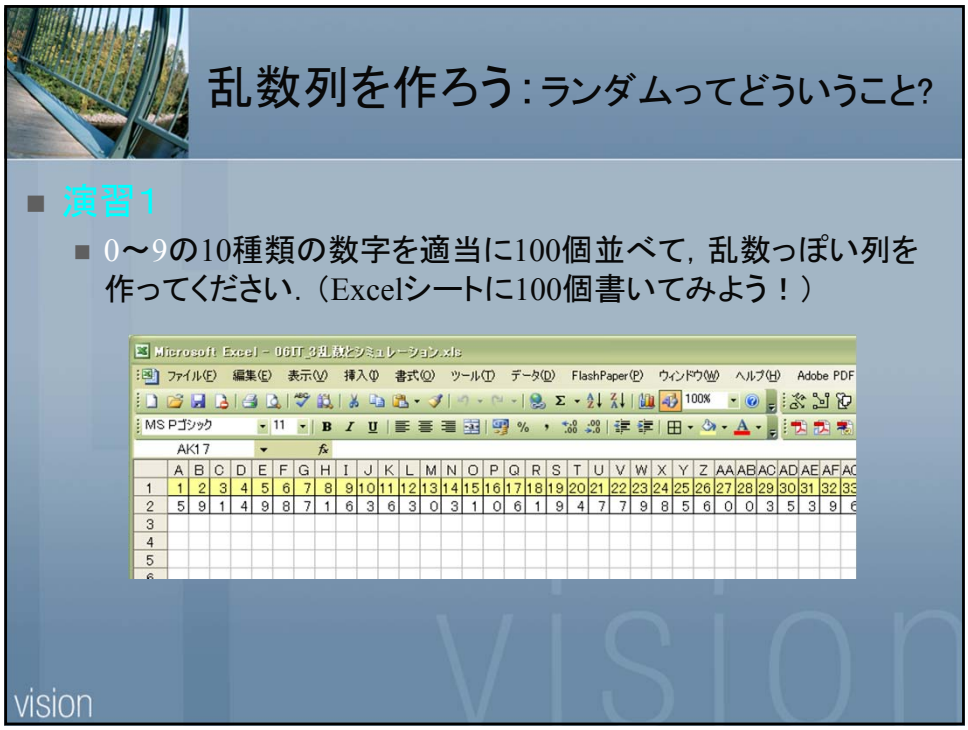
情報学部 堀田敬介

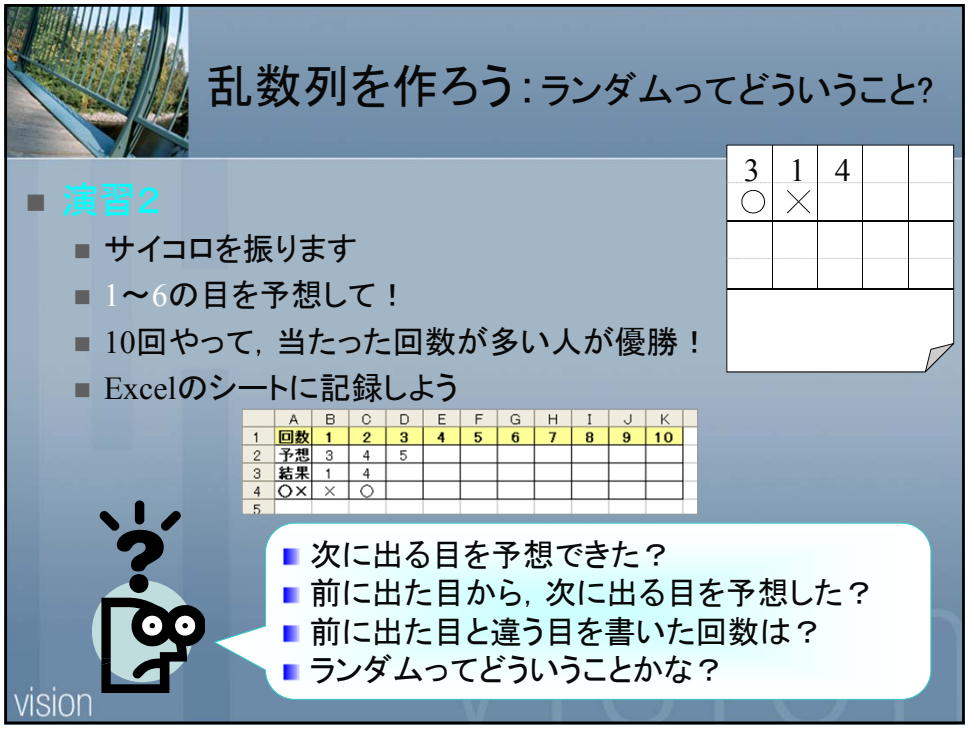
2010/1/22,Fri.

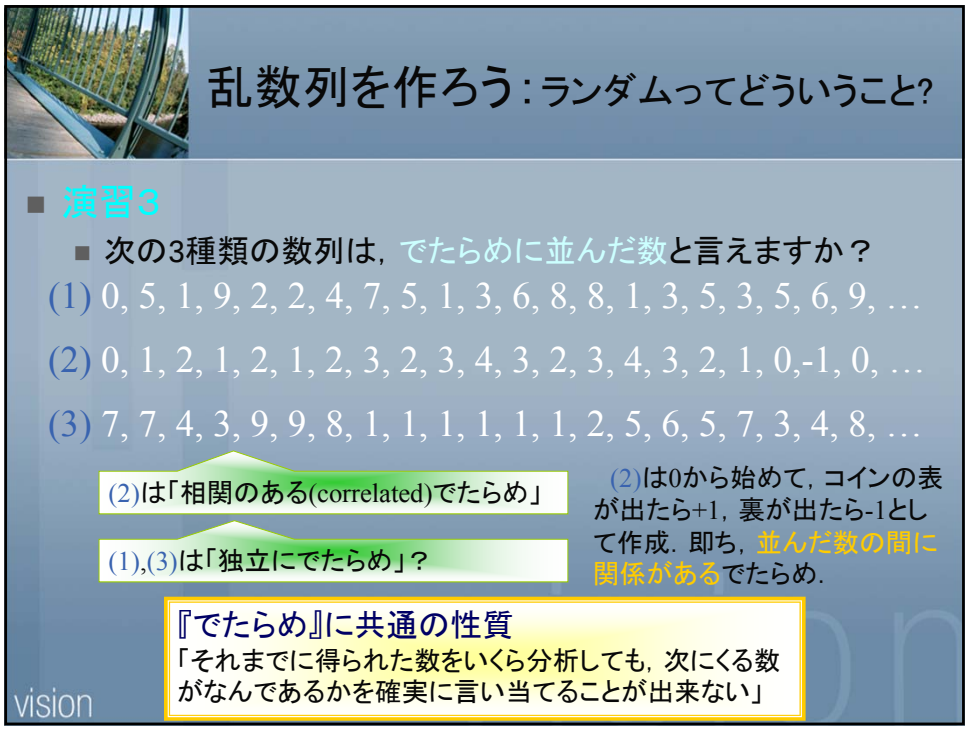


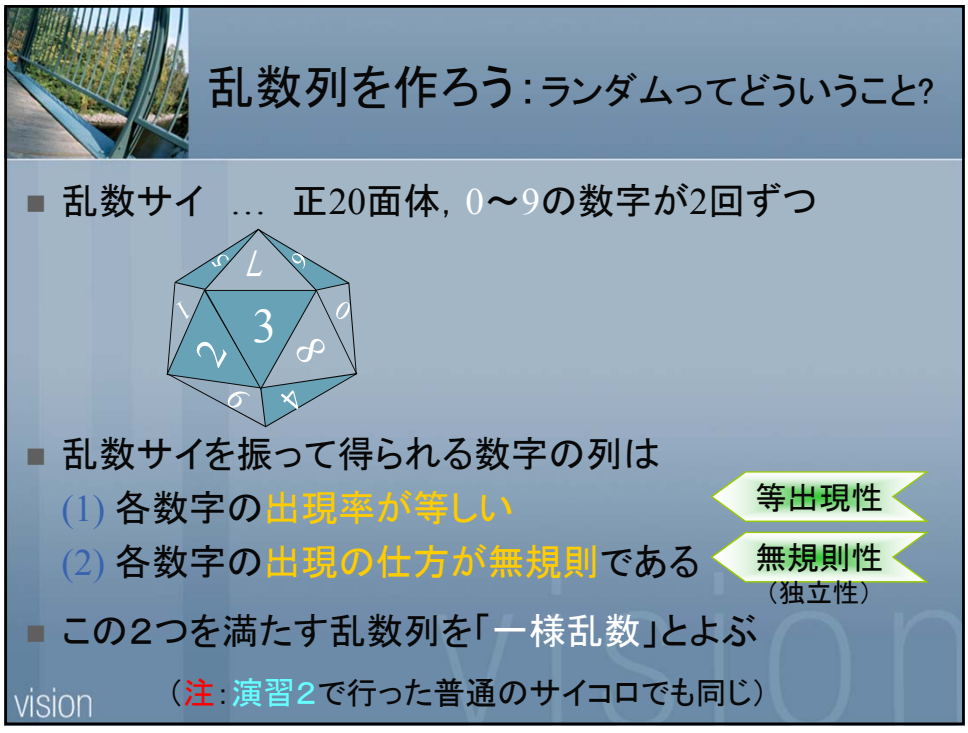
Contents

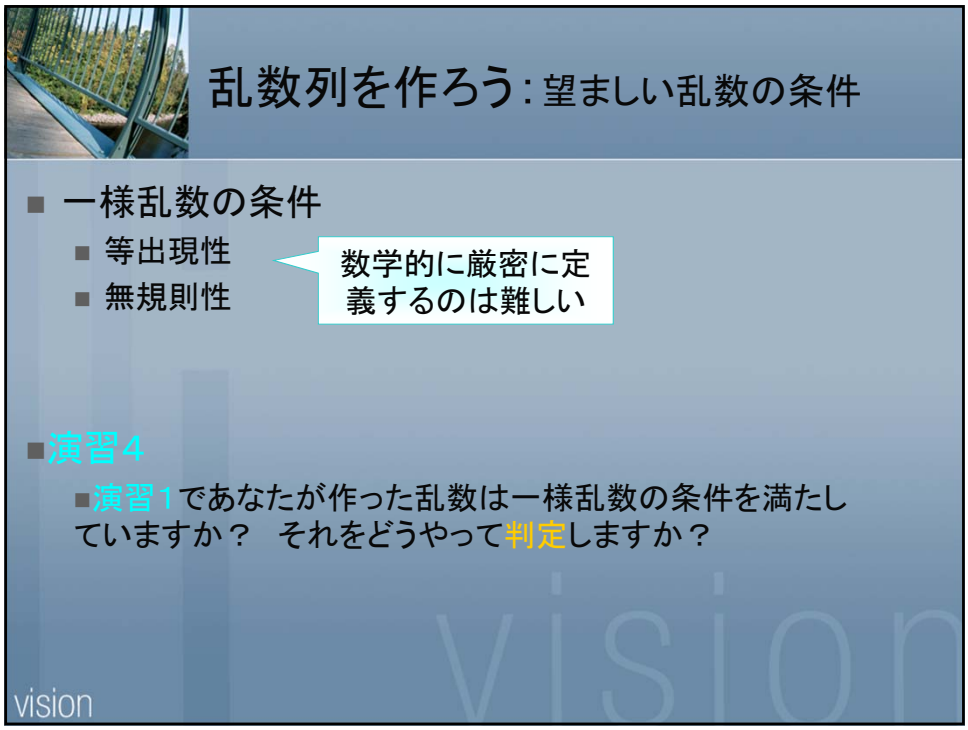
- 乱数列を作ろう
  - 乱数って何？　ランダムってどういうこと？
    - 乱数列の条件
    - 乱数列の検定法
  - 擬似乱数列の生成
    - 擬似乱数列の生成法の望ましい条件
    - 一様乱数以外の乱数を生成してみよう

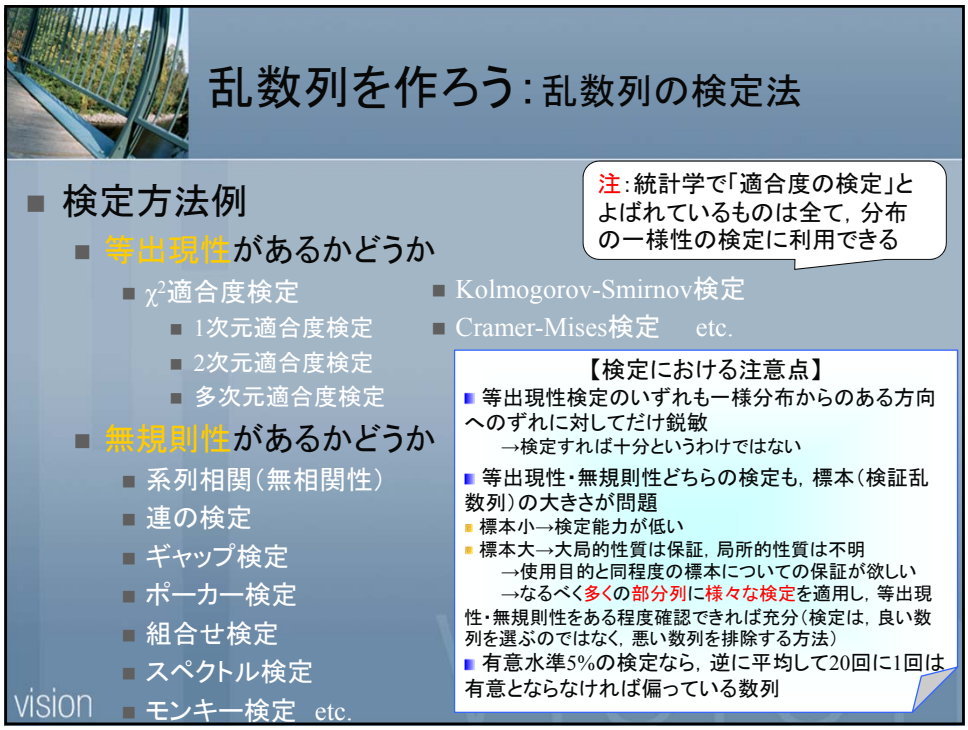

乱数列を作ろう：ランダムってどういうこと?
演習1
0～9の10種類の数字を適当に100個並べて，乱数っぽい列を作ってください．（Excelシートに100個書いてみよう！）
vision

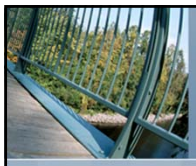

乱数列を作ろう：ランダムってどういうこと?
演習2
サイコロを振ります
1～6の目を予想して！
10回やって，当たった回数が多い人が優勝！
Excelのシートに記録しよう
3 1 4
次に出る目を予想できた？
前に出た目から，次に出る目を予想した？
前に出た目と違う目を書いた回数は？
ランダムってどういうことかな？
vision


乱数列を作ろう：ランダムってどういうこと?
演習3
次の3種類の数列は，でたらめに並んだ数と言えますか？
(1) 0, 5, 1, 9, 2, 2, 4, 7, 5, 1, 3, 6, 8, 8, 1, 3, 5, 3, 5, 6, 9, …
(2) 0, 1, 2, 1, 2, 1, 2, 3, 2, 3, 4, 3, 2, 3, 4, 3, 2, 1, 0,-1, 0, …
(3) 7, 7, 4, 3, 9, 9, 8, 1, 1, 1, 1, 1, 1, 2, 5, 6, 5, 7, 3, 4, 8, …
(2)は「相関のある(correlated)でたらめ」
(1),(3)は「独立にでたらめ」？
(2)は0から始めて，コインの表が出たら+1，裏が出たら-1として作成．即ち，並んだ数の間に関係があるでたらめ．
『でたらめ』に共通の性質
「それまでに得られた数をいくら分析しても，次にくる数がなんであるかを確実に言い当てることが出来ない」
vision


乱数列を作ろう：ランダムってどういうこと?
乱数サイ … 正20面体，0～9の数字が2回ずつ
乱数サイを振って得られる数字の列は
(1) 各数字の出現率が等しい
(2) 各数字の出現の仕方が無規則である
等出現性
無規則性
（独立性）
この2つを満たす乱数列を「一様乱数」とよぶ
（注：演習2で行った普通のサイコロでも同じ）
vision

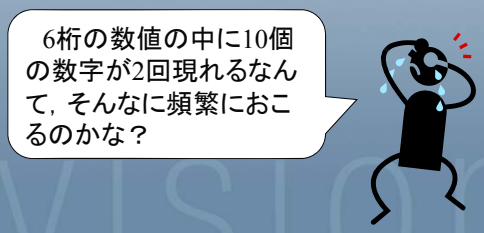

乱数列を作ろう：望ましい乱数の条件
一様乱数の条件
等出現性
無規則性
数学的に厳密に定義するのは難しい
演習４
演習１であなたが作った乱数は一様乱数の条件を満たしていますか？　それをどうやって判定しますか？
vision


乱数列を作ろう：乱数列の検定法
検定方法例
等出現性があるかどうか
χ2適合度検定
1次元適合度検定
2次元適合度検定
多次元適合度検定
無規則性があるかどうか
系列相関（無相関性）
連の検定
ギャップ検定
ポーカー検定
組合せ検定
スペクトル検定
モンキー検定　etc.
Kolmogorov-Smirnov検定
Cramer-Mises検定　etc.
注：統計学で「適合度の検定」とよばれているものは全て，分布の一様性の検定に利用できる
【検定における注意点】
等出現性検定のいずれも一様分布からのある方向へのずれに対してだけ鋭敏
→検定すれば十分というわけではない
等出現性・無規則性どちらの検定も，標本（検証乱数列）の大きさが問題
標本小→検定能力が低い
標本大→大局的性質は保証，局所的性質は不明
→使用目的と同程度の標本についての保証が欲しい
→なるべく多くの部分列に様々な検定を適用し，等出現性・無規則性をある程度確認できれば充分（検定は，良い数列を選ぶのではなく，悪い数列を排除する方法）
有意水準5%の検定なら，逆に平均して20回に1回は有意とならなければ偏っている数列
vision

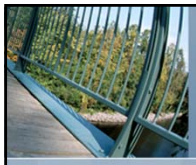


Coffee Break !

- 0, 1, …, 9 の10個の数字で6桁の乱数を作ります.
  【000000, 000001, 000002, …, 999998, 999999】
  - **Question**: 0, 1, …, 9 の数のうち，少なくとも1つがちょうど2回現れる確率を求めなさい.



vision



乱数列を作ろう：乱数列の検定法

- χ^2適合度検定：1次元適合度検定

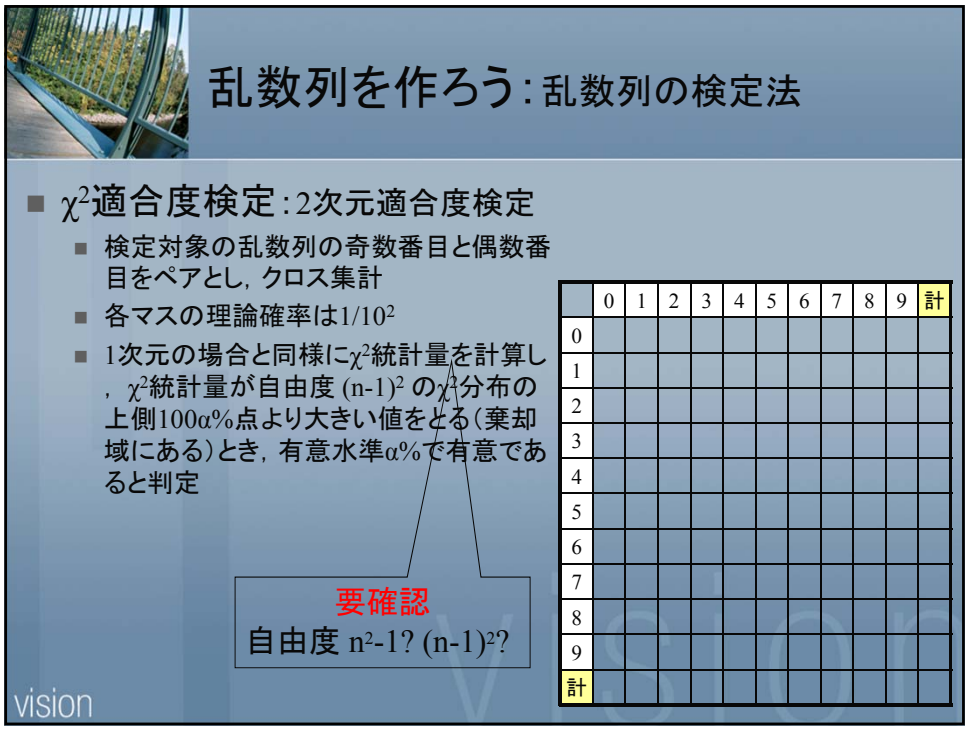
	0	1	2	3	4	5	6	7	8	9	計
頻度	9	11	11	6	14	5	9	7	12	16	100
理論確率	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1
理論値	10	10	10	10	10	10	10	10	10	10	100

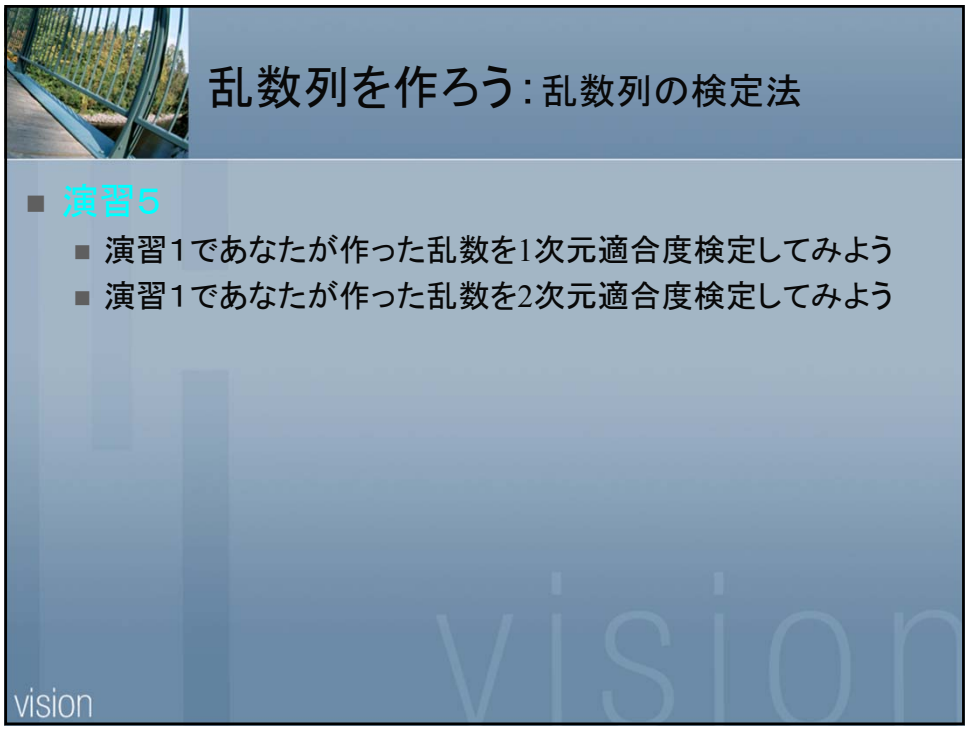
- χ^2統計量

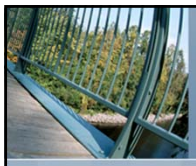
$$\chi^2 := \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \cdots + \frac{(16-10)^2}{10} \quad > \chi^2_\alpha(9)\,?$$

χ^2統計量が，自由度 n-1 のχ^2分布の上側100α%点より大きい値をとる（棄却域にある）とき，有意水準α%で有意であると判定

vision → 帰無仮説「観測度数が理論度数に適合する」が棄却される


乱数列を作ろう：乱数列の検定法
χ²適合度検定：2次元適合度検定
検定対象の乱数列の奇数番目と偶数番目をペアとし，クロス集計
各マスの理論確率は1/10²
1次元の場合と同様にχ²統計量を計算し，χ²統計量が自由度 (n-1)² のχ²分布の上側100α%点より大きい値をとる（棄却域にある）とき，有意水準α%で有意であると判定
要確認
自由度 n²-1? (n-1)²?
0 1 2 3 4 5 6 7 8 9 計
0
1
2
3
4
5
6
7
8
9
計
vision

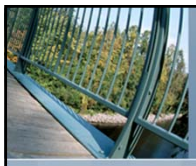

乱数列を作ろう：乱数列の検定法
演習５
演習１であなたが作った乱数を1次元適合度検定してみよう
演習１であなたが作った乱数を2次元適合度検定してみよう
vision



乱数列を作ろう：乱数列の検定法

- 無規則性：ポーカー検定（パーティション検定）
  - $0 \leqq j < n$ について，5個の連続する整数 $\{x_{5j}, x_{5j+1}, \ldots, x_{5j+4}\}$ を取り出し，ポーカーの役のどれに当たるかで7通りに分類する
    〔ブタ，ワンペア，ツーペア，スリーカード，フルハウス，フォーカード，ファイブカード〕
  - 各カテゴリに属する5項組について，χ^2検定を行う

vision



乱数列を作ろう：乱数列の検定法

- 無規則性：連の検定

0, 5, 1, 9, 2, 2, 4, 7, 5, 1, 3, 6, 8, 8, 1, 3, 5, 3, 5, 6, 9

＋－＋ －／＋＋－－＋＋＋／－＋＋－＋＋＋

  - 上昇連（連続上昇）… 左右両端と $x_j > x_j+1$ の間に縦棒

0, 5, 1, 9, 2, 2, 4, 7, 5, 1, 3, 6, 8, 8, 1, 3, 5, 3, 5, 6, 9

2 2 4 1 5 3 4

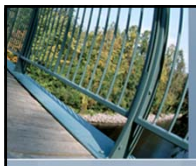
  - 下降連（連続下降）… 左右両端と $x_j < x_j+1$ の間に縦棒

0, 5, 1, 9, 2, 2, 4, 7, 5, 1, 3, 6, 8, 8, 1, 3, 5, 3, 5, 6, 9

1 2 3 1 3 1 1 3 1 2 1 1 1

    - 上昇連，下降連それぞれ，連の長さで6通りに分類
      〔c1:長さ1の連の個数，…，c5:長さ5の連の個数，c6:長さ6以上の連の個数〕
      し，連の検定を行う．複雑なので詳細は省略．

vision

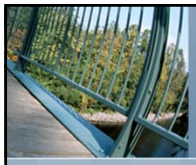


乱数列を作ろう：乱数列の検定法

- 無規則性：目で見るテスト

- 演習６
  - 演習１であなたが作った乱数を目で見よう（Mathematica利用）

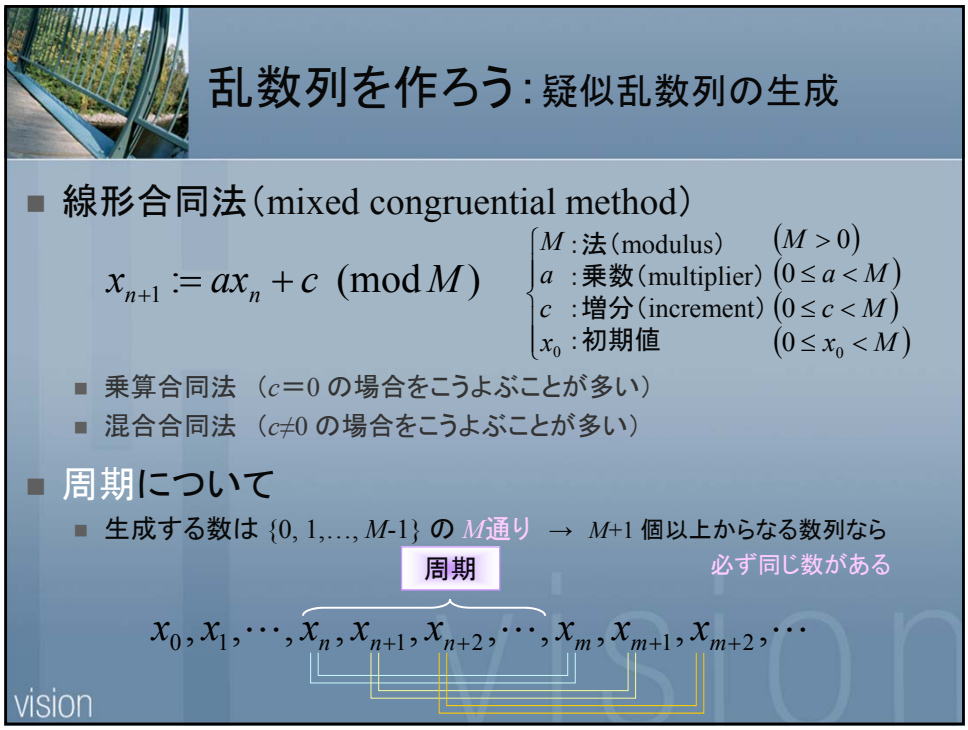
vision

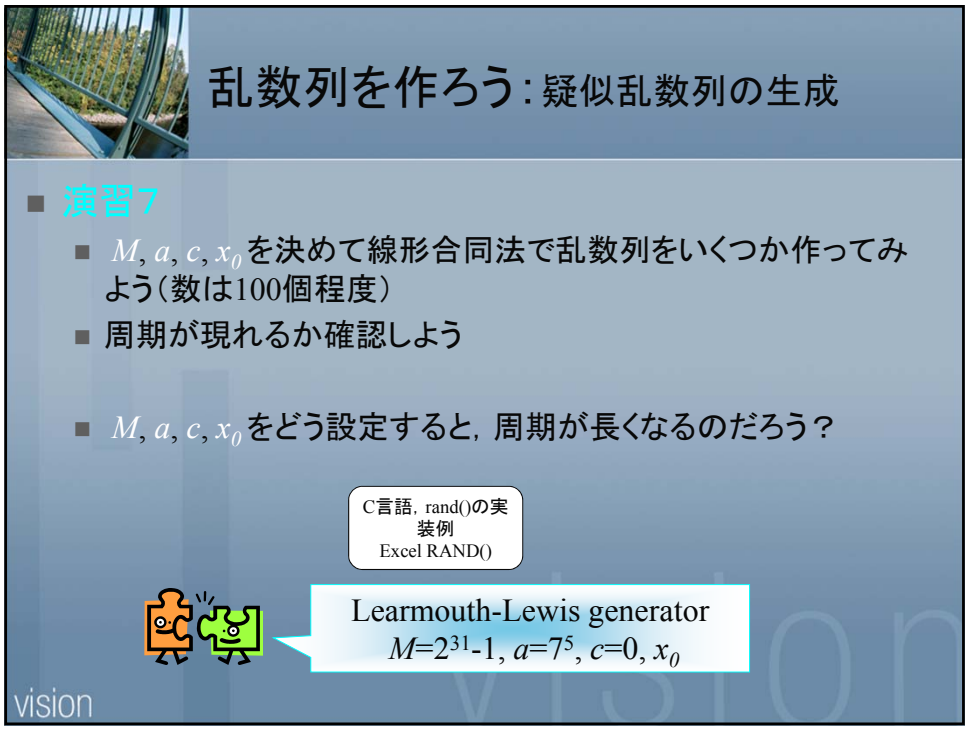


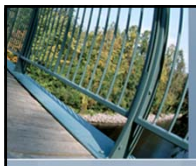
乱数列を作ろう：疑似乱数列の生成

- （一様）擬似乱数列（pseudo-random sequence）の生成
  - 線形合同法（mixed congruential method） … Lehmer(1948)
    - 乗算合同法（multiplicative congruential method）
    - 混合合同法（mixed congruential method）
  - M系列法，最大周期列法（maximum length sequence, MLS）
  - 平方採中法（middle-square method） … von Neumann(1946)
- 一様乱数以外の擬似乱数列の生成
  - 逆変換法
  - 正規乱数の生成

vision


乱数列を作ろう：疑似乱数列の生成
線形合同法（mixed congruential method）
$x_{n+1} := ax_n + c \pmod{M}$
M：法（modulus） $(M > 0)$
a：乗数（multiplier） $(0 \le a < M)$
c：増分（increment） $(0 \le c < M)$
x_0：初期値 $(0 \le x_0 < M)$
乗算合同法 （$c=0$ の場合をこうよぶことが多い）
混合合同法 （$c \neq 0$ の場合をこうよぶことが多い）
周期について
生成する数は $\{0, 1, \ldots, M-1\}$ の M通り → $M+1$ 個以上からなる数列なら
必ず同じ数がある
周期
$x_0, x_1, \cdots, x_n, x_{n+1}, x_{n+2}, \cdots, x_m, x_{m+1}, x_{m+2}, \cdots$
vision


乱数列を作ろう：疑似乱数列の生成
演習7
M, a, c, x_0 を決めて線形合同法で乱数列をいくつか作ってみよう（数は100個程度）
周期が現れるか確認しよう
M, a, c, x_0 をどう設定すると，周期が長くなるのだろう？
C言語，rand()の実装例
Excel RAND()
Learmouth-Lewis generator
$M=2^{31}-1, a=7^5, c=0, x_0$
vision



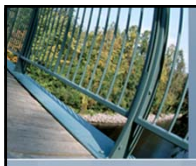
乱数列を作ろう：疑似乱数列の生成

- 線形合同法（mixed congruential method）
  - 最長の周期にするには？

$$x_{n+1} := ax_n + c \pmod{M}$$

  - M はなるべく大きくしたい（なぜか？）
  - $a \geqq 2$ であることが必要（a=0, a=1 としてはいけない）（なぜか？）
  - 周期が最長となる a=c=1 としてはいけない（なぜか？）
- 演習8
  - M=7 のとき，a, c, x_0 を決めて最長周期列を見つけて！

vision



乱数列を作ろう：疑似乱数列の生成

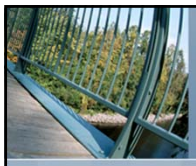
- 線形合同法（mixed congruential method）

$$x_{n+1} := ax_n + c \pmod{M}$$

  - 定理：M, a, c, x_0 で定義する線形合同数列が最長周期を持つ必要十分条件は
    (1) c と M は互いに素
    (2) a-1 は M を割り切る全ての素数 p の倍数
    (3) m が4の倍数なら，a-1 も4の倍数
    である．

補足：c=0 の場合の最大周期は c≠0 の場合の最大周期よりも小さい．また，c=0 のときの最大周期とそれを達成する条件はC.F. Gauss によって証明されている(1801).
なお，Mが素数なら M-1の周期を得る．

vision

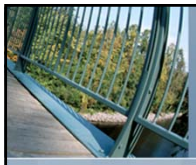


乱数列を作ろう：疑似乱数列の生成

- 線形合同法の欠陥と結晶構造

$$x_{n+1} := ax_n + c \pmod{M}$$

vision



乱数列を作ろう：疑似乱数列の生成

- M系列法，最大周期列法（MLS）
  - ランダムなビット列を生成
  - 3つの初期値を設定 ex) $(x_0, x_1, x_2)=(0, 1, 1)$

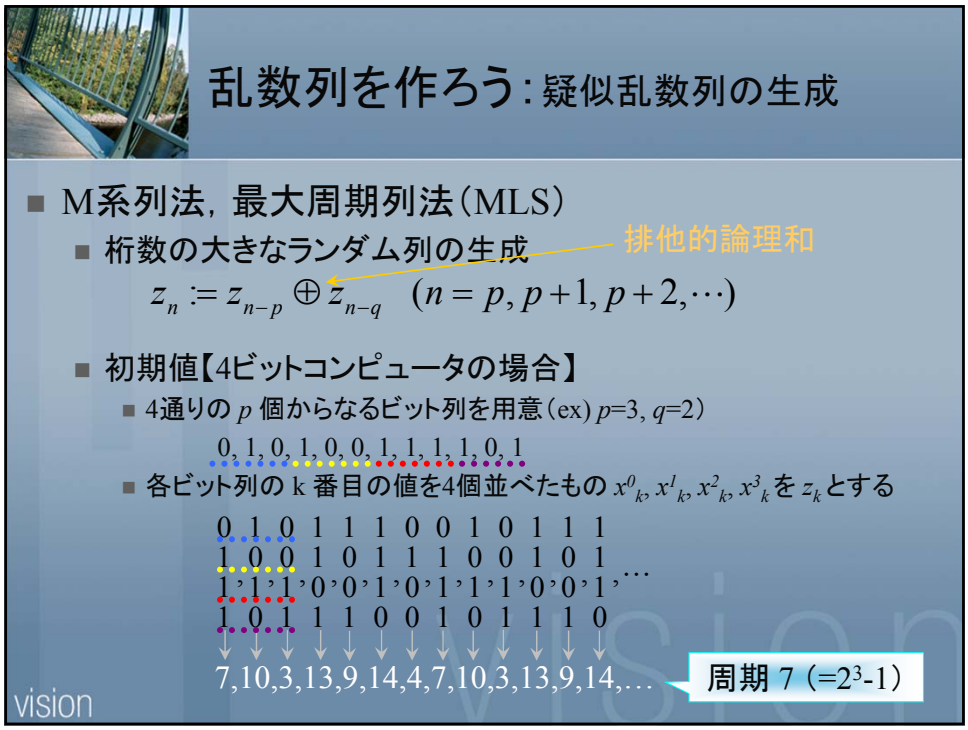
$$x_3 := \begin{cases} 0 & if\ x_0 + x_1 : even \\ 1 & if\ x_0 + x_1 : odd \end{cases}, \quad x_4 := \begin{cases} 0 & if\ x_1 + x_2 : even \\ 1 & if\ x_1 + x_2 : odd \end{cases}, \quad \cdots$$

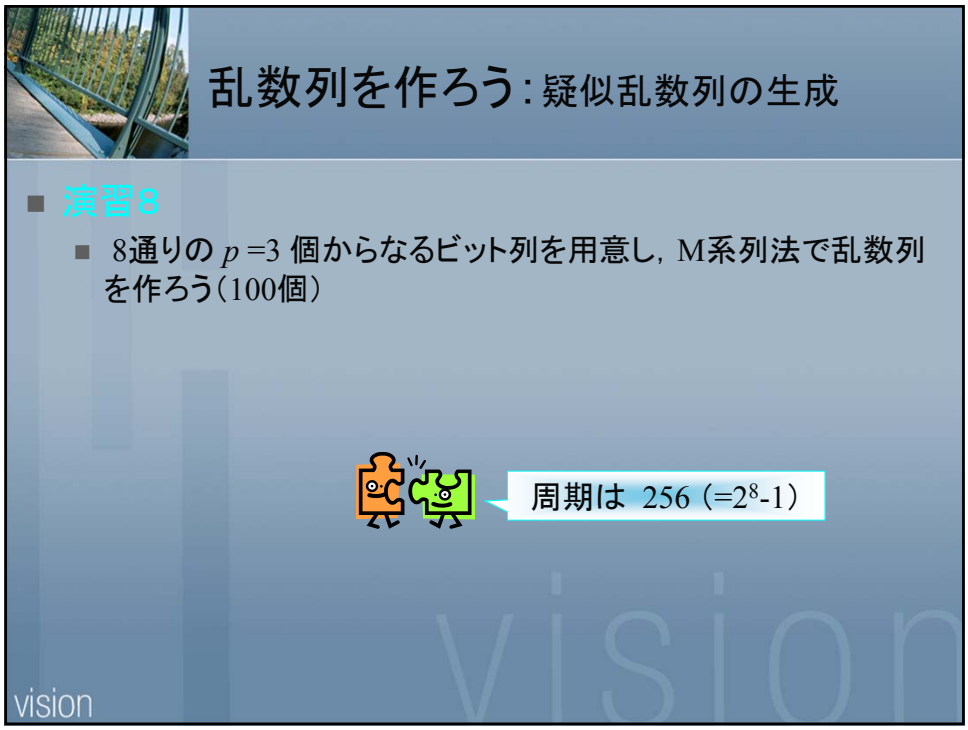
0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, …

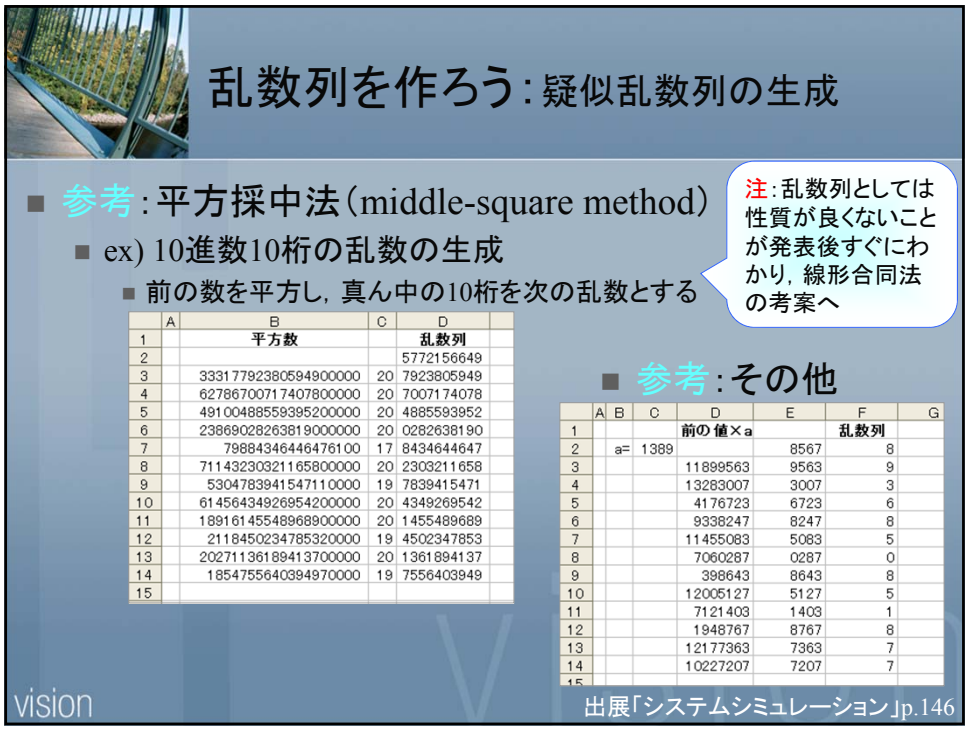
- M系列，最大周期列
  - p 個の初期値 $x_0, x_1, x_2, \ldots, x_p$ から出発し，適当な数 q $(<p)$ を決めて作った数列 ＝ M系列（周期 2^p-1）

$$x_n := x_{n-p} + x_{n-q} \pmod{2}$$

vision

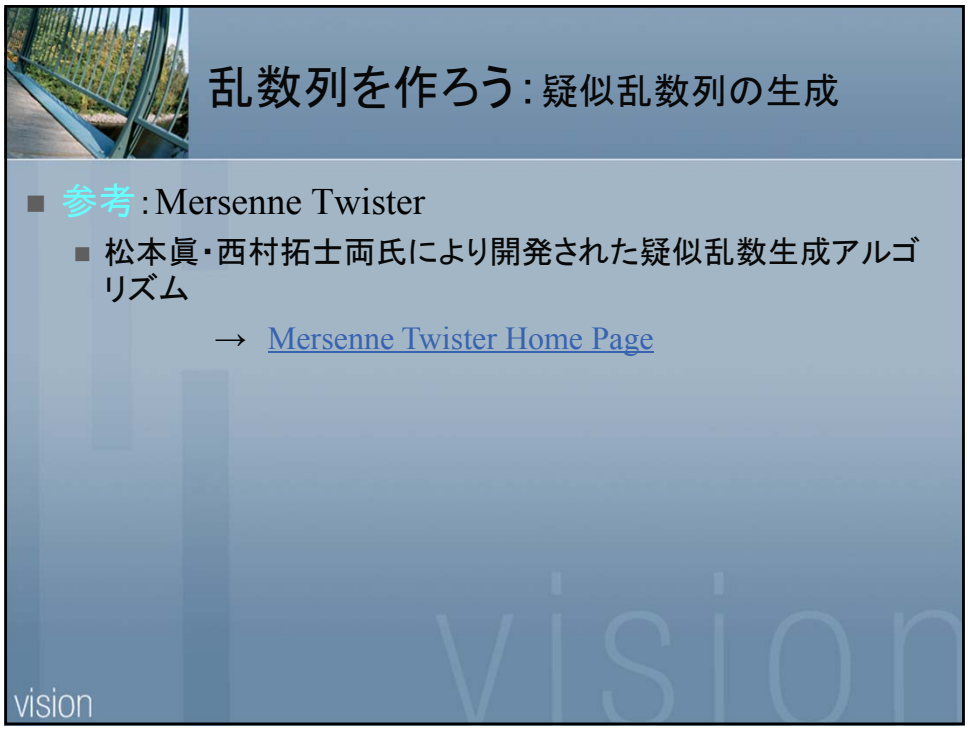

乱数列を作ろう：疑似乱数列の生成
M系列法，最大周期列法（MLS）
桁数の大きなランダム列の生成
排他的論理和
$z_n := z_{n-p} \oplus z_{n-q} \quad (n = p, p+1, p+2, \cdots)$
初期値【4ビットコンピュータの場合】
4通りの p 個からなるビット列を用意（ex) p=3, q=2）
0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1
各ビット列の k 番目の値を4個並べたもの $x^0{}_k, x^1{}_k, x^2{}_k, x^3{}_k$ を z_k とする
0 1 0 1 1 1 0 0 1 0 1 1 1
1 0 0 1 0 1 1 1 0 0 1 0 1
1,1,1,0,0,1,0,1,1,1,0,0,1,…
1 0 1 1 1 0 0 1 0 1 1 1 0
7,10,3,13,9,14,4,7,10,3,13,9,14,…
周期 7（=2³-1）
vision

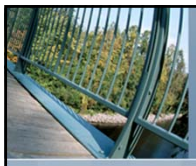

乱数列を作ろう：疑似乱数列の生成
演習8
8通りの p =3 個からなるビット列を用意し，M系列法で乱数列を作ろう（100個）
周期は 256（=2⁸-1）
vision



	A	B	C	D
1		平方数		乱数列
2				5772156649
3		33317792380594900000	20	7923805949
4		62786700717407800000	20	7007174078
5		49100488559395200000	20	4885593952
6		23869028263819000000	20	0282638190
7		7988434644647610 0	17	8434644647
8		71143230321165800000	20	2303211658
9		5304783941547110000	19	7839415471
10		61456434926954200000	20	4349269542
11		18916145548968900000	20	1455489689
12		2118450234785320000	19	4502347853
13		20271136189413700000	20	1361894137
14		1854755640394970000	19	7556403949
15				

	A	B	C	D	E	F	G
1				前の値×a		乱数列	
2		a=	1389		8567	8	
3				11899563	9563	9	
4				13283007	3007	3	
5				4176723	6723	6	
6				9338247	8247	8	
7				11455083	5083	5	
8				7060287	0287	0	
9				398643	8643	8	
10				12005127	5127	5	
11				7121403	1403	1	
12				1948767	8767	8	
13				12177363	7363	7	
14				10227207	7207	7	
15							





Coffee Break !

- 皆さん（学生）に，年齢が若い順に通し番号を付けます．また，学籍番号が若い順にも通し番号を付けます．
  - **Question**：2つの番号が同じになる人は，平均何人いますか？

何人の学生がいるかによるのかな？人数が少ないほど，同じ番号になる人が少ないのかな？

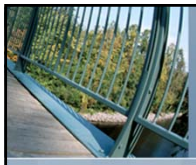
《ヒント》

$$U_i = \begin{cases} 1 & （年齢\ i\ 番目の学生が学籍番号\ i\ 番目） \\ 0 & （o.w.,\ \text{i.e.,}年齢\ i\ 番目の学生が学籍番号\ i\ 番目ではない） \end{cases}$$

$X_n = \sum_{i=1}^{n} U_i$ とし，$E(X_n)$ を考えればOK！

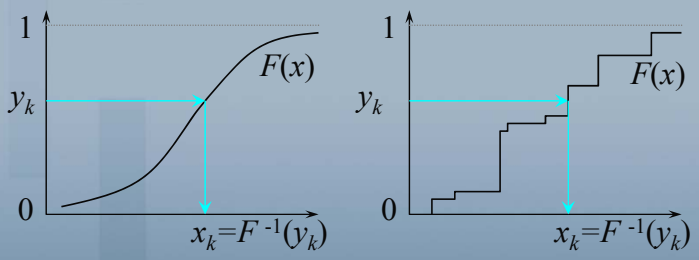
ただし，各 U_i は独立ではないので，X_n は二項分布には従わないよ

vision



乱数列を作ろう：一様乱数以外の乱数

- 逆変換法（逆関数法）
  - （累積）分布関数 $F(x)$ に従う乱数列をつくる

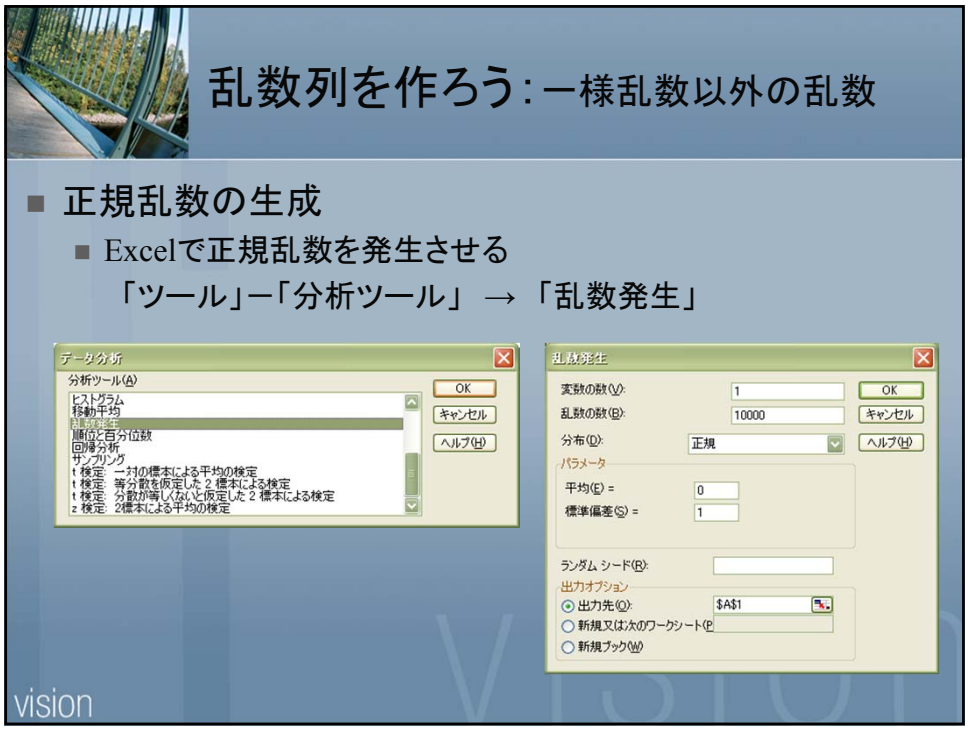


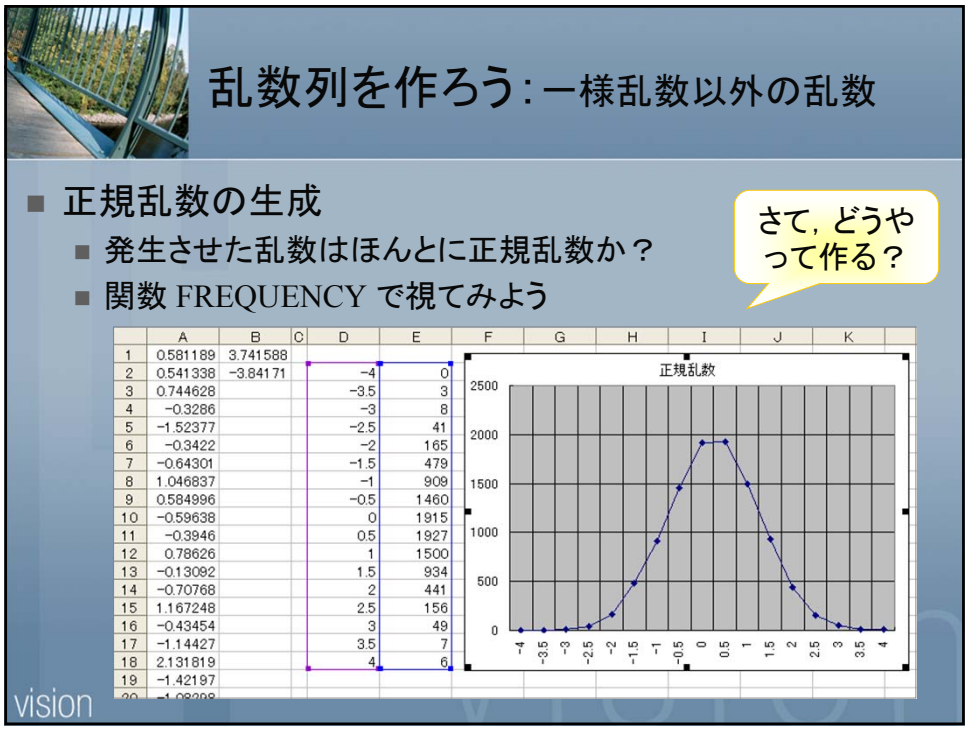
(0,1)-一様分布に従う確率変数を y_k とすると，確率変数 x_k が分布 $F(x)$ に従う

y_k が一様分布に従うため

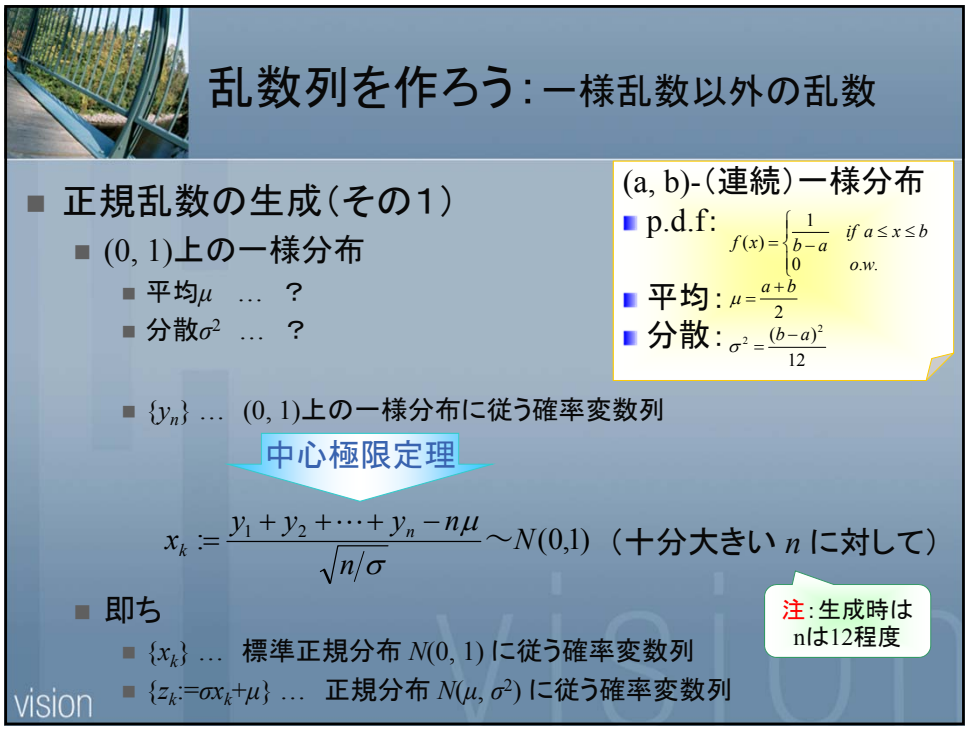
$$\left[\because P\{x_k \le u\} = P\{F^{-1}(y_k) \le u\} = P\{y_k \le F(u)\} = F(u)\right]$$

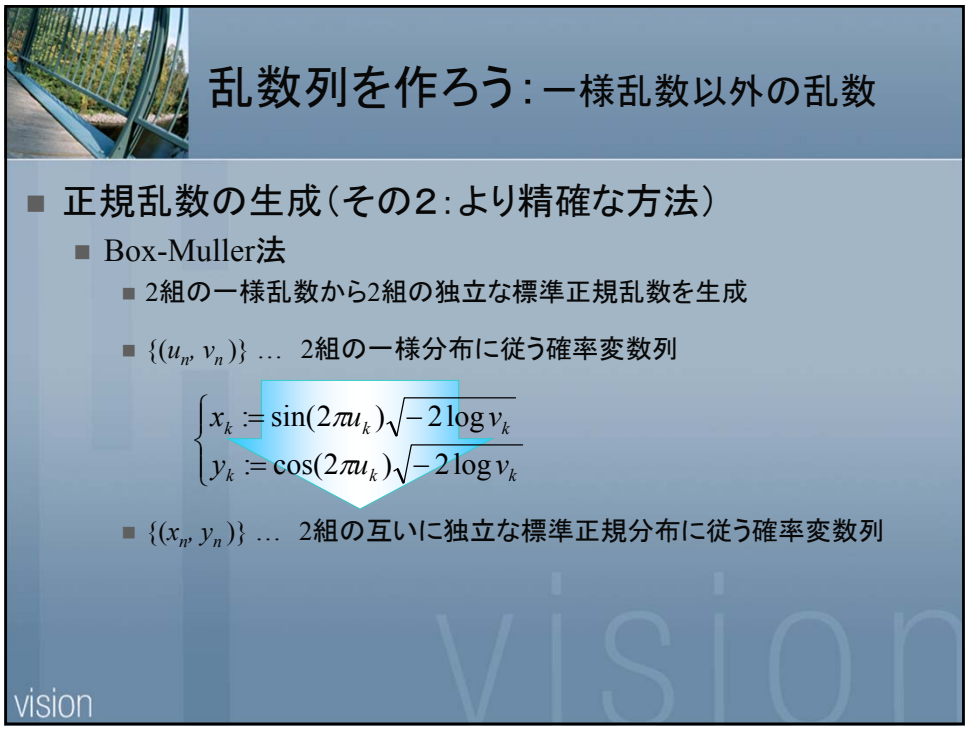
vision

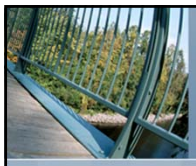




	A	B	C	D	E
1	0.581189	3.741588			
2	0.541338	−3.84171		−4	0
3	0.744628			−3.5	3
4	−0.3286			−3	8
5	−1.52377			−2.5	41
6	−0.3422			−2	165
7	−0.64301			−1.5	479
8	1.046837			−1	909
9	0.584996			−0.5	1460
10	−0.59638			0	1915
11	−0.3946			0.5	1927
12	0.78626			1	1500
13	−0.13092			1.5	934
14	−0.70768			2	441
15	1.167248			2.5	156
16	−0.43454			3	49
17	−1.14427			3.5	7
18	2.131819			4	6
19	−1.42197				


乱数列を作ろう：一様乱数以外の乱数
正規乱数の生成（その1）
(0, 1)上の一様分布
平均μ … ？
分散σ² … ？
{y_n} … (0, 1)上の一様分布に従う確率変数列
中心極限定理
x_k := (y_1 + y_2 + ⋯ + y_n − nμ) / √(n/σ) ～N(0,1) （十分大きい n に対して）
即ち
{x_k} … 標準正規分布 N(0, 1) に従う確率変数列
{z_k:=σx_k+μ} … 正規分布 N(μ, σ²) に従う確率変数列
(a, b)-（連続）一様分布
p.d.f: f(x) = 1/(b−a) if a ≤ x ≤ b, 0 o.w.
平均：μ = (a+b)/2
分散：σ² = (b−a)²/12
注：生成時は nは12程度
vision

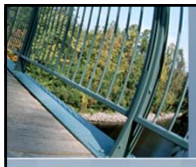

乱数列を作ろう：一様乱数以外の乱数
正規乱数の生成（その2：より精確な方法）
Box-Muller法
2組の一様乱数から2組の独立な標準正規乱数を生成
{(u_n, v_n)} … 2組の一様分布に従う確率変数列
x_k := sin(2πu_k)√(−2 log v_k)
y_k := cos(2πu_k)√(−2 log v_k)
{(x_n, y_n)} … 2組の互いに独立な標準正規分布に従う確率変数列
vision



乱数列を作ろう：疑似乱数列の生成

- 演習9
  - 平均50，分散100の正規乱数を100個作ろう．その際，一様乱数12個で一つの正規乱数になるようにすること
  - Box-Muller法により，正規乱数を作り，Mathematica で2次元平面上にプロットしてみよう

vision



参考文献

- 森戸晋・逆瀬川浩孝「システムシミュレーション」朝倉書店（2000）
- 森雅夫・松井知己「オペレーションズ・リサーチ」朝倉書店（2004）
- D.E.Knuth「準数値算法 3.乱数」サイエンス社（1981（初版1969））
- D.E.Knuth「The Art of Computer Programming 2日本語版」ASCII（2004）〔注：上記の新訳版〕
- 伏見政則「乱数」東京大学出版会（1989）
- 山内・森口・一松「電子計算機のための数値計算法Ⅰ」倍風館（1965）
- 関根智明・高橋磐郎・若山邦紘「シミュレーション」日科技連（1976）
- G.Blom, L.Holst, D.Sandell「確率論へようこそ」シュプリンガー（初版1995,新装版2005）
- E.Beltrami「ランダム～数学における偶然と秩序～」青土社（2002）
- Mersenne Twister Home Page http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html（2006年2月現在）
- NtRand: Numerical Technologies Random Generator for Excel http://numtech.com/NtRand/（2006年2月現在）

vision