筑波大学附属駒場中・高等学校研究報告　第33集　1993年11月

確率・統計におけるシュミレーション

筑波大学附属駒場中高等学校　数学科
深瀬　幹雄

確率・統計におけるシュミレーション

筑波大学附属駒場中高等学校　数学科
深瀬　幹雄

1．はじめに

高校で教科書にある統計の数式や結論の理論的背景を生徒に理解させようと考えると，すぐに高校数学の範囲を越えてしまう。したがって，授業内容では，具体的に統計データを使用しながらも生徒に実感を持たせることが少なく，その結果だけを伝えたり，数式に数値を代入することで統計的な意味を説明することになってしまう。

その結果，生徒にとっては統計的な面白さ，有効性や考え方について理解が深まらずに終わってしまうことがある。この問題点を確率実験やシュミレーションを通して，改善したいと考えた。

2．推定の問題

場合の数や順列・組合せを用いて確率を計算するだけでは，確率の有用性や意味を理解させるのは難しい。実際に，確率を用いて実験，推測することを体験することによって確率・統計についての理解が深まると考える。

ここでは，容器の中に入っている球の個数を推定することを考え，コンピュータで2通りのシュミレーションを行なってみた。

(1)　容器中のN個（推定する数）の中からランダムにn個の球を取りだし，印をつけて容器の中に戻す。次に，よく容器の中をかき回してから，m個の球を取りだした所，m個の中に印の付いた球の個数をkとする。

この確率を$P_k(N)$とすると，

$$P_k(N)={}_mC_k\cdot{}_{N-n}C_{m-k}/{}_NC_m$$

となる。

したがって，$P_k(N)/P_k(N-1)=(N-n)(N-m)/(N-n-m+k)N$

となり，この比は

$(nm/k)\leqq N$のとき

$P_k(N)/P_k(N-1)\leqq 1$

$(nm/k)\geqq N$のとき

$P_k(N)/P_k(N-1)\geqq 1$となる。

したがって，Ｐ$_k$(Ｎ）はＮ＝［ｎｍ／ｋ］で最大となる。これはＮの最尤推定量である。

このことから豆を利用してこの確率実験を行ない，袋の中の豆の総数を推定したことがある。

それでは確率実験の意味については理解させることが出来るが，実際に用いることができる豆の数や実験回数の関係で得られた推定値から確率の有用性を示すことができない場合が多かった。コンピュータを用いることによって，この問題を解決することができる。（図１）は第１，２回に取りだす数を150として，この実験を100回行なった結果である。また，（図２）は第１，２回に取りだす数を300として，実験を100回行なった結果である。100個の推定値のヒストグラムはいずれも推定する数（2500個）を中心に分布している。また，取りだす数を多くすると，推定値が2500の周りに集ってくることも分かる。これによって，推定値の確率的意味の理解を深めることができる。

全体の数（推定数）　:2500
第１回の捕獲数　:150
第２回の捕獲数　:150
試　行　回　数　:100

推定数　:

1730.8	2500.0	2250.0	3750.0	2250.0	1875.0	2250.0	2250.0
1730.8	2250.0	3214.3	2812.5	3750.0	2045.5	1730.8	5625.0
5625.0	2500.0	1250.0	3750.0	2250.0	3750.0	3214.3	2500.0
2250.0	3214.3	2812.5	2045.5	2812.5	2812.5	2500.0	1250.0
2812.5	2812.5	3750.0	4500.0	4500.0	2250.0	2812.5	1607.1
1406.3	1500.0	2250.0	2812.5	3214.3	1875.0	2250.0	1875.0
2500.0	2812.5	5625.0	2812.5	3750.0	3214.3	3214.3	2812.5
4500.0	2045.5	4500.0	3214.3	2250.0	2812.5	3750.0	2250.0
2500.0	3750.0	3750.0	4500.0	2045.5	1875.0	2045.5	2500.0
2500.0	2812.5	2500.0	3750.0	2250.0	2500.0	2812.5	2045.5
2500.0	3750.0	5625.0	2500.0	2250.0	2500.0	2500.0	1875.0
3214.3	2500.0	3750.0	3214.3	2045.5	3214.3	2250.0	1875.0
2045.5	3750.0	2500.0	4500.0				

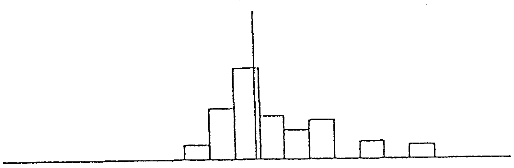
平均推定量　:2833.4

誤差　:333.4

（図１）

最大推定値　:　5625.0
最小推定値　:　1250.0
階級数を入力:
10

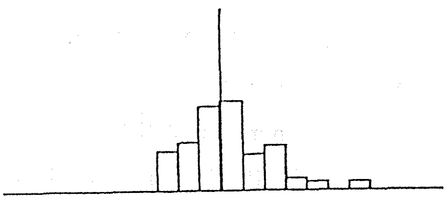
階級幅　:　437.5



（図２）

最大推定値　:　3913.0
最小推定値　:　1875.0
階級数を入力:
10

階級幅　:　203.8



プログラムの機能

①全体の数（推定数）

②第１回目に取りだす数

③第２回目に取りだす数

④実験回数を入力する。

①～④まで入力すると数の推定値，平均推定値，平均推定値と推定数の誤差を表示される。

その後，リターンキーを押すと，最大推定値および最小推定値を表示される。

また，階級数を入力すると，階級幅およびヒストグラムを表示される。

(2)　容器の中に１からＮまでの番号のついた球が入っている。ランダムな復元抜き取りによってｎ個の球を抜き取ったときの最大番号をＸとすると，

$$P\{X \leqq k\} = (k/N)^n$$

である。したがって，Ｘの分布は

$$p_k = P\{X = k\} = P\{X \leqq k\} - P\{X \leqq k-1\}$$
$$= \{k^n - (k-1)^n\} / N^n$$

である。Ｘの期待値Ｅ（Ｘ）は

$$E(X) = \sum_{k=1}^{N} k\, p_k$$

$$= \frac{1}{N^n} \sum_{k=1}^{N} \{k^{n+1} - (k-1)^{n+1} - (k-1)^n\}$$

$$= \frac{1}{N} \{N_{n+1} - \sum_{k=1}^{N} (k-1)^n\}$$

大きいＮに対して，最後の和を積分で評価すると

$$\frac{1}{N} \sum_{k=1}^{N} (k-1)^n \fallingdotseq \frac{1}{N^n} \int_0^N x^n\, dx = \frac{N}{n+1}$$

したがって，

$$E(X) \fallingdotseq nN/(n+1)$$

このことから

$$N \fallingdotseq (n+1)E(X)/n$$

この方法は第２次大戦中に，敵の生産量を推定するのに使用されたといわれる。

この場合の推定値は，総数に比較して取りだすものが少なくてもかなり推定したい数Ｎに近い

ものが得られる。例えば，（図３）にあるように2500個の中から100個程度取りだして推定すると最大誤差で76である。

（図３），（図４）は，コンピュータで実験した結果である。

プログラムの機能

①全体の数（推定数）

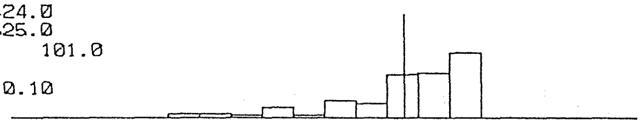
②取りだすデータの数

③実験回数

①～③まで入力されると，最小推定値，最大推定値およびその差が表示される。階級数を入力すると階級幅とヒストグラムが表示される。

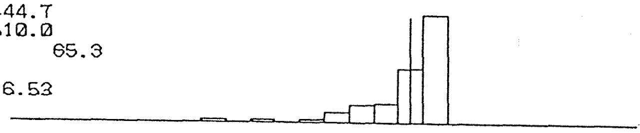
（図３）

```
総数（推定数）を入力してください　：2500
取りだすデータ数を入力してください：100
試行回数を入力してください　　　　：100
  2431.1  2519.9  2494.7  2517.9  2523.0  2477.5  2519.9  2448.2  2462.4  2523.0
  2521.0  2502.8  2507.8  2502.8  2525.0  2499.7  2524.0  2510.9  2519.9  2522.0
  2519.9  2506.8  2503.8  2462.4  2514.9  2477.5  2505.8  2501.8  2514.9  2504.8
  2481.6  2499.7  2491.7  2470.5  2503.8  2507.8  2484.6  2491.7  2511.9  2463.4
  2514.9  2424.0  2502.8  2499.7  2518.9  2510.9  2504.8  2515.9  2523.0  2503.8
  2500.8  2512.9  2511.9  2509.8  2509.8  2507.8  2522.0  2503.8  2507.8  2474.5
  2525.0  2516.9  2497.7  2474.5  2463.4  2502.8  2513.9  2509.8  2475.5  2514.9
  2510.9  2516.9  2513.9  2517.9  2498.7  2503.8  2475.5  2507.8  2519.9  2524.0
  2443.2  2492.7  2492.7  2485.6  2524.0  2503.8  2497.7  2490.7  2438.1  2525.0
  2524.0  2506.8  2463.4  2514.9  2521.0  2487.6  2518.9  2502.8  2513.9  2502.8

最小推定値　：　2424.0
最大推定値　：　2525.0
最大値－最小値　：　101.0
階級数入力　：10
階級幅　　　：　10.10
```



（図４）

```
総数（推定数）を入力してください　：2500
取りだすデータ数を入力してください：250
試行回数を入力してください　　　　：100
  2501.0  2500.0  2499.0  2502.0  2504.0  2508.0  2504.0  2501.0  2508.0  2501.0
  2504.0  2496.9  2506.0  2508.0  2496.9  2501.0  2489.9  2500.0  2504.0  2510.0
  2506.0  2504.0  2504.0  2504.0  2495.9  2501.0  2482.9  2508.0  2507.0  2506.0
  2506.0  2499.0  2481.9  2494.9  2491.9  2507.0  2505.0  2507.0  2488.9  2491.9
  2504.0  2495.9  2506.0  2485.9  2504.0  2507.0  2510.0  2500.0  2499.0  2485.9
  2507.0  2496.9  2508.0  2508.0  2498.0  2501.0  2504.0  2507.0  2444.7  2509.0
  2496.9  2503.0  2507.0  2505.0  2487.9  2493.9  2504.0  2506.0  2477.9  2498.0
  2505.0  2498.0  2507.0  2461.8  2510.0  2506.0  2507.0  2500.0  2504.0  2472.9
  2487.9  2494.9  2508.0  2510.0  2503.0  2510.0  2502.0  2510.0  2486.9  2506.0
  2507.0  2495.9  2506.0  2509.0  2490.9  2483.9  2509.0  2498.0  2508.0  2484.9

最小推定値　：　2444.7
最大推定値　：　2510.0
最大値－最小値　：　65.3
階級数入力　：10
階級幅　　　：　6.53
```



3．eの近似値

実数πの近似値は，乱数を発生させ平面上に点を取り，正方形とそれに内接する円に入る点の個数の比で求めたり，ビュッフォンの針の実験を利用して求める確率実験が行なわれる。

ここでは，順列を利用してeの近似値を求める確率実験を考える。

1からnまでの数字を1つずつかいたカードを
n枚を1列に並べるとき，数字kがk番目の位置にあるならば，この順列はkを不変にするという。どの数字も不変にしない順列を完全順列という。xをn枚のカードに対して存在する完全順列の個数をすると，

$x_1=0,\ x_2=1,$

$x_n=(n-1)x_{n-1}+(n-1)x_{n-2}$……①

が成り立つ。

n枚の作る順列の総数はn！であるから，完全順列の起る確率をp_nとすると，

$p_n=x/n!$

である。

①の両辺をn！で割ると

$p_n=(n-1)p_{n-1}/n+p_{n-2}/n$

$p_1=0,\ p_2=1/2$

よって，

$n(p_n-p_{n-1})=-(p_{n-1}-p_{n-2})$

$p_2-p_1=1/2$

より，$p_3-p_2=-1/3!$

$p_4-p_3=-1/4!$

$p_5-p_4=-1/5!$

•

•

$p_n-p_{n-1}=(-1)^n\cdot 1/n!$

よって，

$p_n=1/2!-1/3!+1/4!-\cdots+(-1)^n\cdot 1/n!$

また，

$e^{-1}=1/2!-1/3!+\cdots+(-1)^n\cdot 1/n!+\cdots$

より，大きなnに対して

$p_n\fallingdotseq e^{-1}$

となり，$e\fallingdotseq 1/p_n$である。

よって，確率実験によってeの近似値を求めることができる。教室での実験では，トランプを用いて実験できる。

試行回数も生徒一人当たりが，10回の実験を行なうことにより，クラス全体では400回程にもなり，eの近似値を求めるもとができる。（図５）では，５枚のカードを100回並べる実験を100回行なうことにより，100個のeの近似値を求めヒストグラムにしたものである。

また，（図６）では，10枚のカードについて行なったものである。

この問題からくじ引きでクラスで座席替えをするとき，少なくとも１人が替える前と同じ座席に決まる確率が１－１／e＝0.632・・・に近くなる。

プログラムの機能

①並べるカードの枚数

②カードを並べる回数

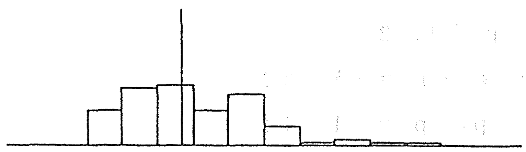
③実験回数

①～③まで入力されると，最大推定値，最小推定値およびその差を表示する。

階級数を入力すると，階級幅およびヒストグラムが表示される。

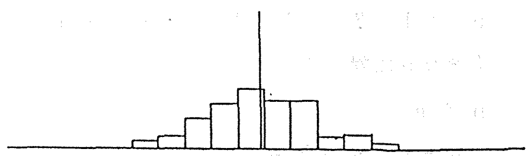
（図５）

何枚のカードを並べますか　: 5
何回試行しますか　　　　　: 100
この実験を何回行ないますか: 100
最大推定値　　　　　: 4.348
最小推定値　　　　　: 2.128
最大推定値ー最小推定値　: 2.220
階級数を入力　　　　　　: 10
階級幅は　　　　　　: 0.222



（図６）

何枚のカードを並べますか　: 10
何回試行しますか　　　　　: 300
この実験を何回行ないますか: 200
最大推定値　　　　　: 3.261
最小推定値　　　　　: 2.206
最大推定値ー最小推定値　: 1.055
階級数を入力　　　　　　: 10
階級幅は　　　　　　: 0.105



4．標本平均の分布

教科書では，標本平均の分布について統計データのヒストグラムを示し，抽出する標本の数が大きくなるとその分布は近似的に正規分布と見なせると説明され，この結果を利用して確率を求める計算をすることになる。これでは公式を暗記し，問題に適用する程度で，その本来の意味も分からないまま終わってしまうことになる。確率実験やコンピュータによるシュミレーションを通して，標本平均値の分布について実感のある理解をさせたいと考えた。

X	−1	0	1	平均0
確率	1/3	1/3	1/3	分散2/3

$\overline{X}=(X_1+X_2)/2$ の分布を作ると、

$\overline{X}$	−1	-1/2	1	1/2	1	平均0
確率	1/9	2/9	3/9	2/9	1/9	分散1/3

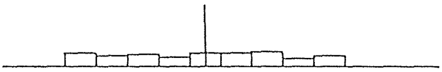
また、$\overline{X}=(X_1+X_2+X_3)/3$ の分布を作る。

$\overline{X}$	−1	-2/3	-1/3	0	1/3	2/3	1	平均0
確率	1/27	3/27	6/27	7/27	6/27	3/27	1/27	分散2/9

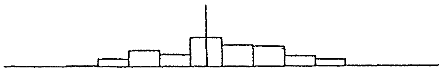
また、これらの分布を図示することで比較する。

（図７）

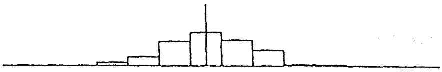
何個の乱数の平均を求めますか ： 1
データ数を入力してください ： 200
階級数を入力してください ： 9
階級幅 ： 1.00



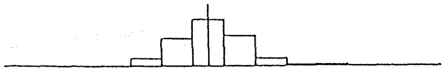
何個の乱数の平均を求めますか ： 2
データ数を入力してください ： 200
階級数を入力してください ： 9
階級幅 ： 1.00



何個の乱数の平均を求めますか ： 4
データ数を入力してください ： 200
階級数を入力してください ： 9
階級幅 ： 1.00

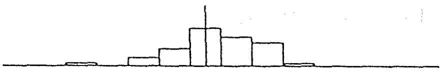


何個の乱数の平均を求めますか ： 8
データ数を入力してください ： 200
階級数を入力してください ： 9
階級幅 ： 1.00

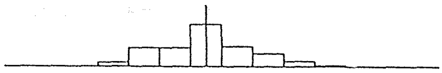


（図８）

何個の乱数の平均を求めますか ： 4
データ数を入力してください ： 50
階級数を入力してください ： 9
階級幅 ： 1.00

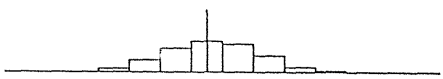


何個の乱数の平均を求めますか ： 4
データ数を入力してください ： 100
階級数を入力してください ： 9
階級幅 ： 1.00



何個の乱数の平均を求めますか ： 4
データ数を入力してください ： 400
階級数を入力してください ： 9
階級幅 ： 1.00

何個の乱数の平均を求めますか ： 4
データ数を入力してください ： 1600
階級数を入力してください ： 9
階級幅 ： 1.00



5．おわりに

4．の実験授業を都立高校で実施したときの生徒の感想をいくつか紹介してみる。

＊感覚的には“そうなりそう”という思いがあるが，実際には一人ではそんなに多く試行することができないから“あれ？”という結果が出ることがある。しかし，今回はクラス全員の集計したグラフができているからわかった。計算機とにらめっこするのはあまり好きではないけれど，けっこう楽しかったと思う。

＊この程度のことなら日常の中でわかっていることなので予想もつけやすく面白くなかった。もっと自分の全然知らない数字を使って実験した方がよいと思う。

＊ふつうの確率の授業では理論ばかりで実用的でないと思われるが，このようなものは実用的でよいと思う。たまには，こういうのもよいと思う。

また，2．と3．については，確率の授業の中では，コンピュータ実験を行なった。実験から得た推定値の分布の様子から統計的推測の意味を生徒に伝えられたと考えている。また，使い方によっては，確率実験の教材開発と共に授業のどの場面でどのような確率実験やコンピュータシュミレーションを提示するかを研究する必要がある。とくに，新教育課程の数学Ｃとの関連で，どのようにコンピュータを活用して統計処理にかかわる内容の理解を深めるかの研究を扱いたい。

参考文献

1）山田　孜：応用確率論，森北出版
2）羽鳥裕久：あたらしい統計学，培風館
3）安斎利洋・伊吹龍：ターボ・グラフィックス，ＪＩＣＣ出版
4）杉山高一：統計教育のための例題資料集