乱数

宮部賢志 明治大学数学科

2017 年 4 月 20 日　横浜サイエンスフロンティア高校

はじめに

くじ *BIG*

- くじ BIG は，サッカーの試合結果を予想する宝くじ．
- 14 試合について，勝ち (1)，負け (2)，その他 (0) の 3 種類の結果がある．
- 予想はコンピュータによりランダムに選ばれる．
- スポーツ振興投票の実施等に関する法律第 9 条の規定により、19 歳未満の者のくじ購入および譲受、当せん金の受け取りは禁止されている．(wikipedia より)

ある人がこの BIG をある日に 5 口，次の日に 10 口，購入したところ，最初の 5 口が完全に一致した．

奇跡が起きたのか？

本当にランダムならば，このようなことが起こる確率は，$\frac{1}{3^{70}}$．これは宇宙誕生 (138 億年前) から現在まで毎秒 100 口ずつ (3 万円) 買い続ける，ということを 10^{10} 回繰り返して 1 回起こるくらいの確率．

考えられるのは…

- たんなる偶然
- 不正
- 擬似乱数の不適切な扱い

主催者は「非常に確率は低いものの、販売している全体の数の中でたまたま結果が重複する可能性は否定しきれない」として，返金はしないとのこと．

今日の話

1. ランダム性は便利
乱数がどのように使われているか.
乱数は便利である数学的理由は何か.

2. 乱数を検定する
「確率は低いが可能性は否定できない」と言っていたら，何も結論できない.
不自然かどうかを数学的に判定する方法を紹介する.

3. 擬似乱数を作る
乱数を計算機で作るには，高度な数学が使われる.
乱数の利用で注意すべきことを説明し，BIG への教訓とする.

ランダム性は便利！

乱数は何に使われているか *1*

宝くじ

誰にとっても予測不可能であることが，公平を担保する．「戦争に行く人をくじで決める」となれば，「本当に公平なのか？」という疑問が湧く．

ゲーム

予測不可能性が，面白さを生む (場合もある)．

乱数は何に使われているか2

ランダム化比較試験

薬 A と薬 B ではどちらがよく効くかをどのように判定したら良いか？ランダムにサンプルを振り分けることにより，様々な誤差が打ち消し合う．

乱択アルゴリズム

乱数を使った計算で，高い確率で正しい結果を得る．典型的な場合を選択することで，例外を排除する．

乱数は何に使われているか*3*

暗号

情報を「規則のない列」に紛れ込ませる

ランダムであれば持つ性質

大数の法則

平均は期待値に近づく.

サイコロを n 回投げて，その和を n で割ると，n が十分大きければ，だいたい 3.5 くらいになる.

BIG の予想で当たる人もいれば，当たらない人もいる．多くの人が当たれば，主催者は損をするのではないか？

もし予想がランダムならば，そういうことは滅多に起こらない．toto や BIG では，当選金額の合計はかなり正確に予測できる．ある人は「パチンコ屋が倒産しないのは，大数の法則のおかげ」と言った.

結論

誤解「ランダムは役に立たない」

$\Downarrow$

事実「ランダムは役に立つ」

乱数を検定する

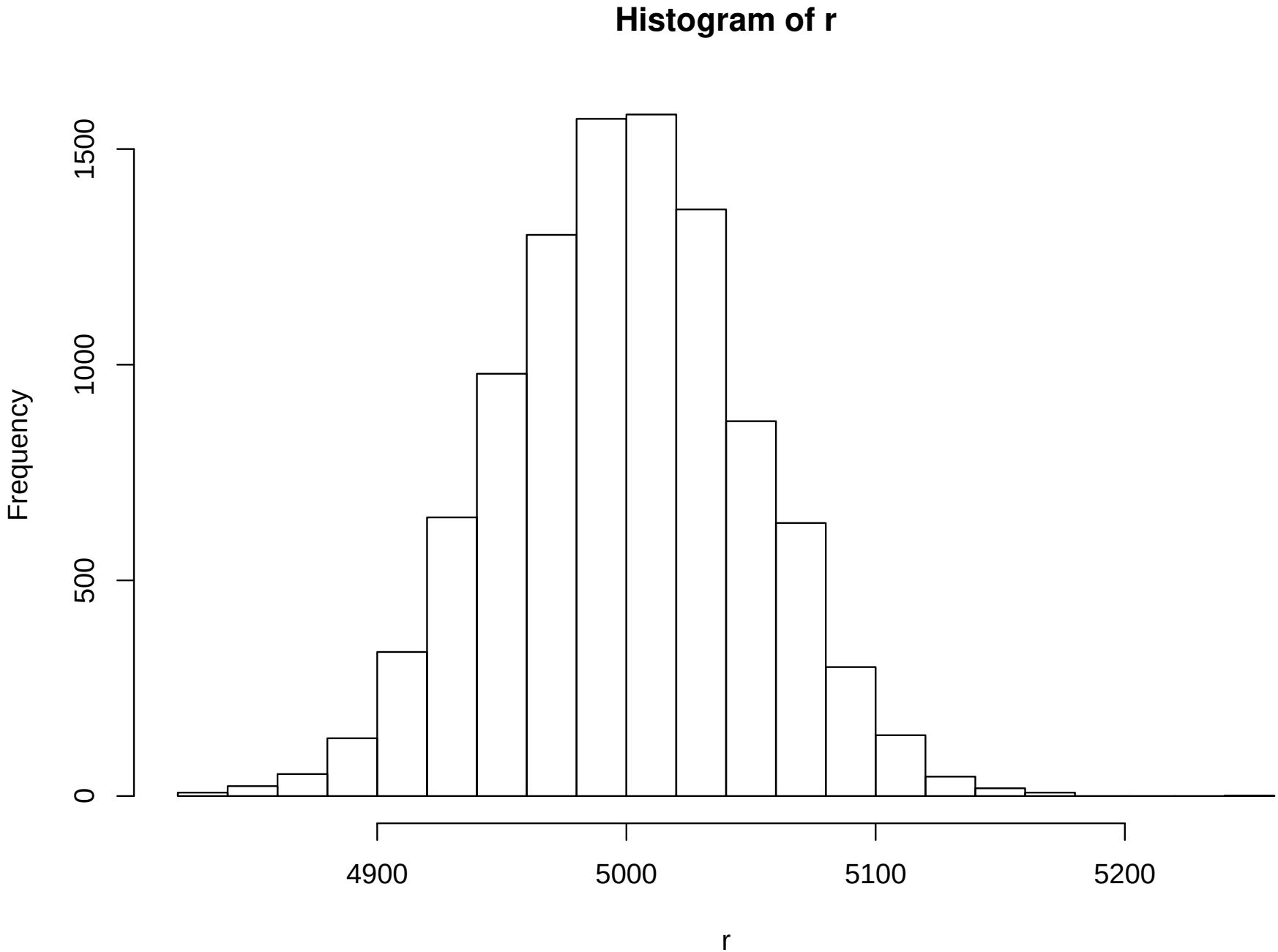
近似

コインを投げて表の回数を数える.
高校数学では

10 回投げて 5 回表が出る確率は？ $\frac{{}_{10}C_5}{2^{10}}$

大学数学では

10000 回投げて表が $4900 \sim 5100$ 回の確率は？


Histogram of r
Frequency
0
500
1000
1500
4900
5000
5100
5200
r

正規分布

計算機によるシミュレーションから，だいたい5000回近くに集まっていることが分かる．これは大数の法則の結論でもある．
それだけでなく，その散らばり具合もまた,「よくある形」をしている．この形を正規分布 (ガウス分布) と言い，この密度関数は,

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

と書かれる．

中心極限定理

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \approx \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

計算機による1万回のシミュレーションでは，9560回であった．

一方，$\mu = \frac{1}{2}$，$\sigma^2 = \frac{1}{4}$ より，$4900 \sim 5100$ 回となる確率は，約 0.9544997 であることが計算できる．

すなわち，約95%の確率でこの中に入る．それならば，この中に入らなければ，歪んだコインであったと言えるだろう．

統計的仮説検定

- 仮にコインが公平であったとしよう.
- このとき，表の回数の期待値が c 回で，それを中心に $a \sim b$ 回出る確率が 95% であったとする.
- この中に入っていなければ，コインが公平であるとは信じられない.
- この中に入っていれば，コインが公平であったとしても，不自然さは見当たらない.

背理法の確率版である.

科学的な主張

仮説検定は「唯一絶対の正しいもの」とは言えない．
しかし，科学の世界で広く使われている方法であり，これ以外に反する方法で議論するのであれば，仮説検定よりも優れていることを何らかの方法で示す必要があり，難しい．科学的に何かを主張するのであれば，最低限知っておくべきことである．

実際，文系でデータ分析などをする人たちは，統計にずいぶん苦しめられているようである…

結論

誤解「数学では答えがただ一つだが,
現実には誤差がある
(から数学は役に立たない)」

⇓

事実「数学で誤差を表現することができ,
それは現実世界の判断の材料になる」

擬似乱数を作る

擬似乱数とは

乱数の有用性は昔から知られていたが，計算機で乱数を作る試みは，当然のことながら，計算機が作られるようになってからのこと．
計算機はプログラムされたとおりにしか動かないので，予測不可能な乱数は作れない．
フォン・ノイマンは「漸化式で乱数を作るのはある種の罪」と言っている．

それでも，短い時間で大量の乱数が必要になったことから，乱数に見える列，擬似乱数の研究が始まった．
長く使われてきたのは線型合同法であり，その発想の元であるフェルマーの小定理を紹介する．

フェルマーの小定理

定理 1. p を素数とし，a と p は互いに素であるとすると，

$$a^{p-1} \equiv 1 \mod p$$

証明. $\{1, 2, \ldots, p-1\}$ は $\{a, 2a, \ldots, (p-1)a\}$ と p を法として等しい．なぜなら，$ia \equiv ja$ なら，a と p が互いに素なので，$i \equiv j$ より矛盾する．すべてをかけ合わせると，$(p-1)! \equiv a^{p-1}(p-1)!$ となるが，p と $(p-1)!$ は互いに素なので，$a^{p-1} \equiv 1 \mod p$. □

例

$a = 5,\ p = 7$ とすると，

$$a \equiv 5,\ a^2 \equiv 4,\ a^3 \equiv 6,\ a^4 \equiv 2,\ a^5 \equiv 3,\ a^6 \equiv 1.$$

しかも，この並びはランダムに見えて，乱数に使えそう！

線型合同法

$$X_{n+1} = (A \times X_n + B) \mod M$$

古くから使われている例が,

$$X_{n+1} = (1103515245 \times X_n + 12345) \mod 2^{32}$$

このような漸化式による乱数の生成は，種 (seed) を保存しておけば，全く同じ乱数列を生成することができるというメリットもある.

一方で，線型合同法には多くの短所が知られている．例えば，上の例の場合，最下位のビットは必ず 0 と 1 が交互に現れる.

問題と改善

BIG くじの問題における予想される原因

- 不適切な初期化 (種選び) の問題
- 古い乱数生成法の不適切な使用

改善策

- メルセンヌツイスターなどの改善された乱数生成法を使う
- 初期化の方法と頻度に注意

結論

誤解「なんとなく動いているから良さそう」

$\Downarrow$

事実「背景にある数学をきちんと理解して使わないと，
大きな問題が起こることがある」

終わり

ありがとうございました.

明治大学
MEIJI UNIVERSITY