デタラメさの効用と、１＋１＝０の世界

松本 眞（東京大学数理科学研究科）

2012年1月11日　東京大学学術俯瞰講義

第壱話　デタラメさの効用：意外なところで
　ランダムネスのお世話に

第弐話　デタラメさを生み出すのは意外に難しい：
　（デタラメ禅問答）

第参話　近現代数学によるデタラメさの生成
　（メルセンヌ・ツイスター）

１．デタラメさの効用

最初の例 さいころ：デタラメさ（ランダムネス）
を生成する装置（＝乱数発生器）

1. 振るたびに、１から６までの数を確率1/6で発生する
   （偏りのなさ、一様性）
2. 過去の履歴に依存しない （独立性）
3. 規則性がない （独立性、推測不能性）
4. これから出る目は予見できない
   （予測不能性）
5. どこかで誰かが振ったサイコロの目を、推測できない
   （推測不能性）

ゲーム・ギャンブル：「運」を生み出すためのデタラメさ
ゲーム・ギャンブルでのランダムネスの利用

1. たからくじ　(勝つ戦略なし。)
   予見できないことが大切。偏りがないことも大切。
   回転する数字板に、弓矢を射って番号を決める。
2. ルーレット（掛け方にいろいろあるが、戦略はない。）
3. トランプ：混ぜてランダムネスを発生させる。戦略性があるゲームもある。
4. マージャン：混ぜてランダムネスを発生させる。戦略性もある。

5. 競馬・競輪：見ようによっては、「ただの」ランダムネス発生器と言える。

   予見できないことが大切。偏りがあるが、その「偏り具合」すらはっきりしないところが面白い。

予見できない「幸運」を狙う人々の心に働きかける。

対照的に：囲碁、将棋、チェス、オセロなどのゲームにはランダムネスは全くない。
「運」に左右されない。
（先手後手を決めるときに、ランダムネスを使うことはある。）

きちんとできないからデタラメにやる

きちんとはできないとき、デタラメにやるのは良い作戦

1. 世論調査 内閣支持率

   2012年12月13日付ＮＨＫニュースより抜粋

   「ＮＨＫは、今月９日から３日間、全国の２０歳以上の男女を対象に、コンピューターで無作為に発生させた番号に電話をかける「ＲＤＤ」という方法で、世論調査を行いました。調査の対象となったのは１６４５人で、６１％に当たる１００５人から回答を得ました。それによりますと、野田内閣を「支持する」と答えた人は、先月の調査より８ポイント下がって３７％だったのに対し、「支持しない」と答えた人は、１２ポイント上がって４２％で、内閣発足後３か月で、不支持が支持を上回りました。」

全ての２０歳以上の日本在住の人にきちんと電話をした場合、大体一億人くらいに質問をしなくてはならない。

１日は86400秒しかない。

そこで「無作為に１６４５人抽出する」ことで、全体の傾向をさぐる。

RDD=Random Digit Dialing 電話帳を使わずに、デタラメな電話番号をコンピュータで生成して電話をかける。

調査対象の母集団が大きすぎるとき、無作為に適当な数の対象を抽出する（ランダムサンプリング)。

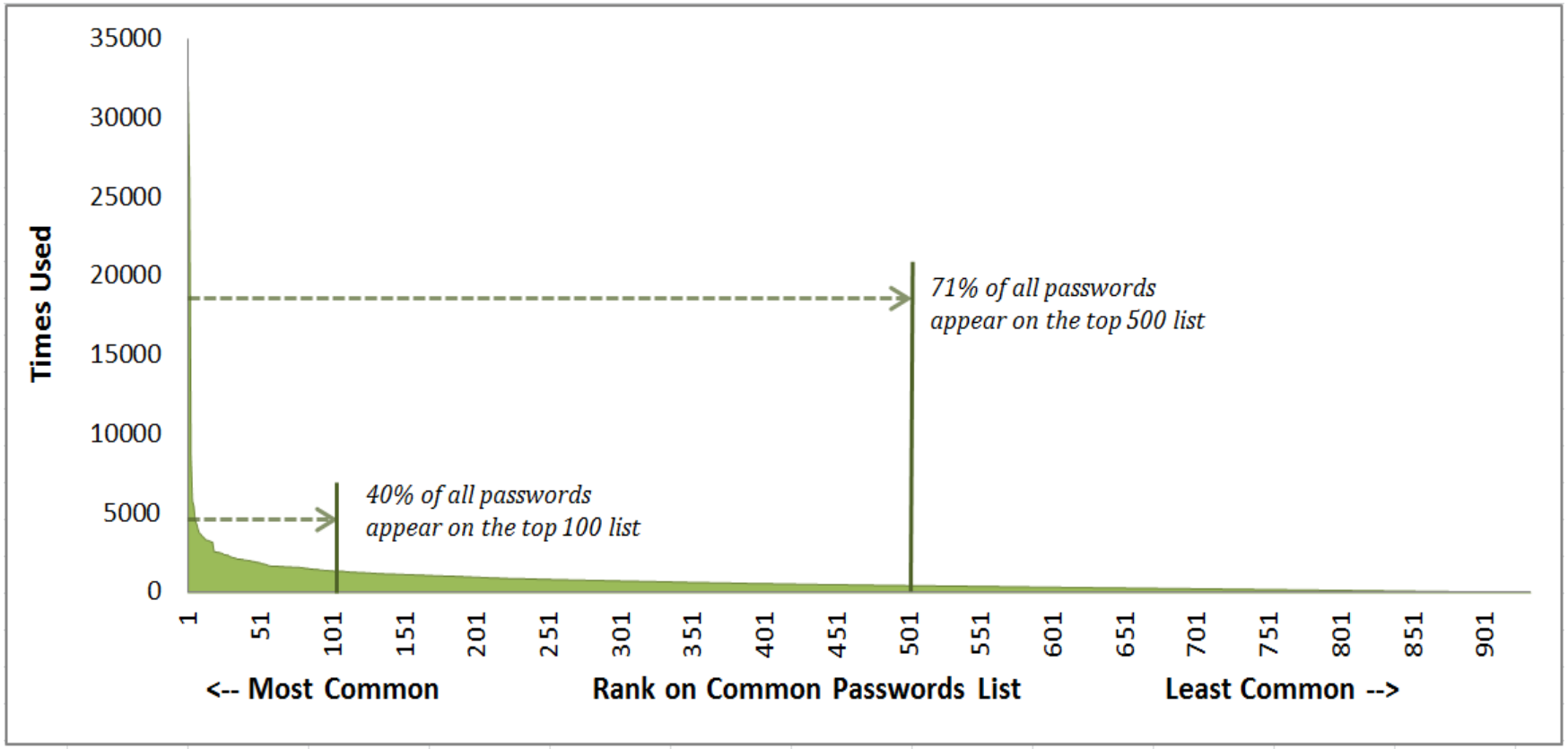
偏りがないことが重要

公平にするためのデタラメさ

1. くじびきなど
2. 予測不能性が必要
3. 偏りのなさ（一様性）が必要
4. 「じゃんけん」もこの仲間である（が、偏りが大きい。
   ランダムネス生成の難しさを垣間見る。）

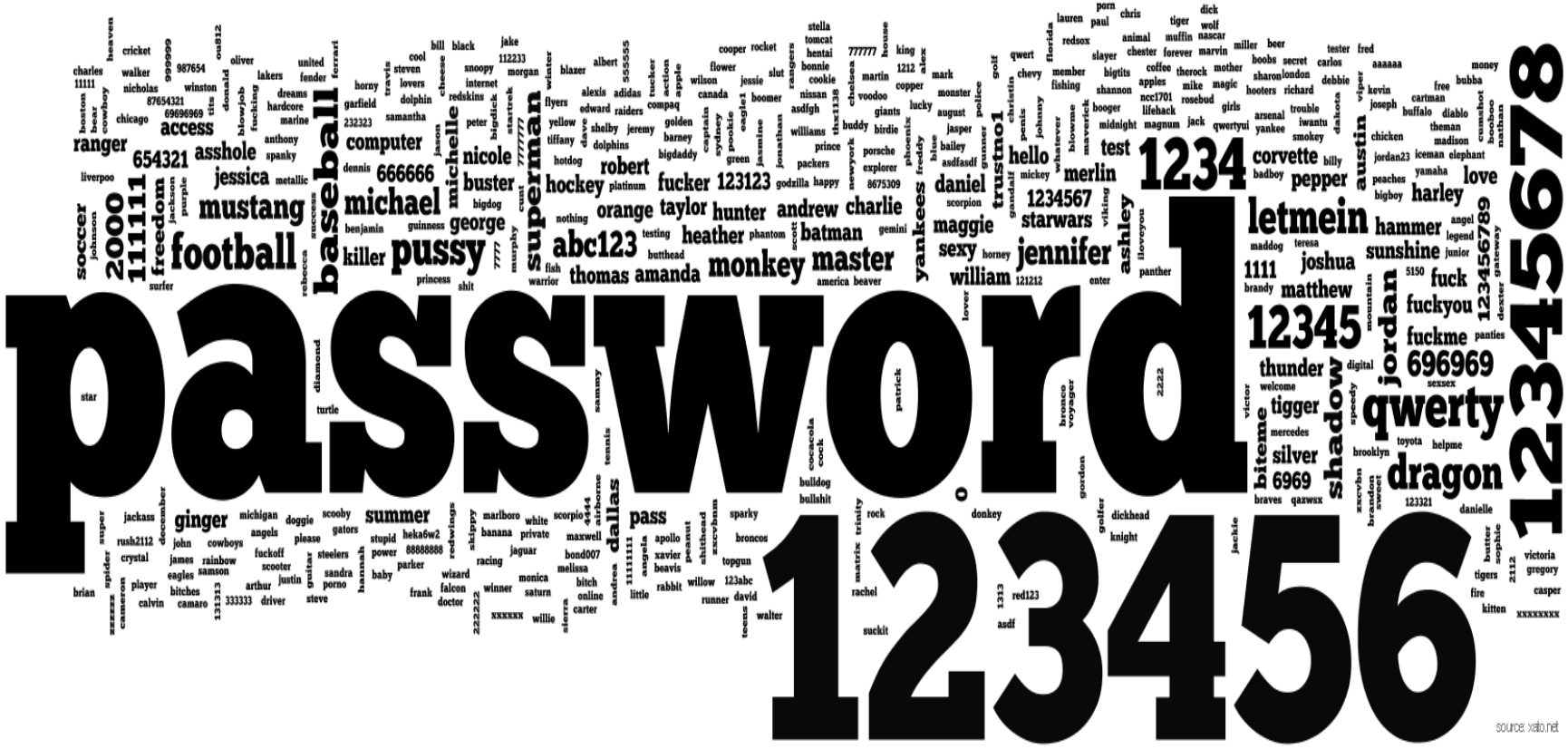
予測・推測を不可能にするためのデタラメさ（暗号乱数）
パスワード：「自分にしかわからない文字列を考え出して、
　パスワードとして利用する。」
しばしば簡単に破られる。以下のデータは下記より引用
http://xato.net/passwords/more-top-worst-passwords

世界で使われているパスワードの40%は、トップの100種類に入る（password:4.7%, 123456:3.8%, 12345678: 1.3%, ... ）
71%は、トップの500種類に入る



‡ 引用元 http://xato.net/passwords/more-top-worst-passwords

各パスワードの面積を使われている度数に比例させてえがいたもの



‡ 引用元 http://xato.net/passwords/more-top-worst-passwords

工夫されたものであっても家族やペットの名前、誕生日を組み合わせたものも多く、第三者が推測しやすい

推測しにくさの点で理想的なパスワードは
　デタラメに選ばれたもの：

たとえばAからZの26種のアルファベットからなる８文字のパスワードを作るなら、26種の文字を8個一様独立にランダムに並べるべきである。
（「記憶しにくい」という大問題は、今回は無視する。）
これなら、26^8のどの文字列も等確率1/208827064576 (約2088億分の1)であらわれ、推測はできない。

問題：どうやってこの8文字を作り出すか？
一つの方法：サイコロを使う

コンピュータの中で、デタラメさを生成できないか？
→次回のテーマ

「フォン・ノイマン型計算機ではデタラメさを生成できない」

「デタラメさとは何か、定義せよ」

完全な暗号法：one time pad

暗号は、ジュリウス・シーザー (BC100-44) の時代にはすでに存在した。

ここでは第一次大戦ですでに利用されたストリーム暗号と、その一つの理想形である one time pad について解説する。

（河東さんの第二回の講義でも取り扱われているが、ここではより鈍長に解説する。）

準備：データの二進化

取り扱われるデータは、全て0-1の列に還元される。

古典的な例：モールス信号（ひらがなやアルファベットを、トンとツーの二種の長さの音の列であらわす）

近代的な例：ASCIIコード：一文字の数字およびアルファベットを、7個の0-1の並び（2進数）に対応づける。

画像も同様。

仮定：

1. AさんがBさんにデータを送りたい
2. データの通信路からは、第三者が内容を傍受しほうだい
3. 第三者に内容を全く洩らしたくない

例：無線通信（第一次世界大戦）
例：インターネット（近現代）
インターネット上を流れるデータは、データが経由していく各コンピュータからはそのまま傍受することができる。

従って、何も工夫しなければ、インターネット上で入力するあらゆる秘匿情報、たとえば

- クレジットカード番号
- パスワード
- 住所・電話番号

などは全て第三者に漏えいしうる。
そこで、このような秘匿情報の通信の際には、暗号化が必要である。

http:　で始まるホームページでは、（原則として）暗号化がされていない。このようなホームページに、秘匿すべき情報を入力するのは第三者への漏えいの可能性が高く危険である。

https:　で始まるホームページでは、Secure Socket Layer (SSL) と呼ばれる暗号化が施されており、通信路からの情報漏えいを防いでいる。

SSLに用いられていることもある、ストリーム暗号の概略を述べる。ここでも「デタラメさ」が中心的役割を果たす。

ストリーム暗号

準備：AさんとBさんは、将来通信するデータの長さ以上の「デタラメな0-1の列」を生成し、「秘密鍵」Kとして、安全な通信路（手渡し、手紙）を使って、二人だけで共有する。
「秘密鍵」Kは他の誰にも知らせず、使用後は廃棄する。
以下、簡単のためにデータの長さは11ビットとする。

鍵共有秘密鍵として、デタラメな（=他人から推測されない）11個の0-1の列をつくる：たとえば

$$K = 10011100101$$

ができたとして、直接手渡しなどで共有する

通信後になって、Aさんがメッセージ

$$M = 01000100010$$

を通信したいとする。
Aさんは、盗聴し放題の通信路を用いて、暗号文(cipher text)

$$C = M + K$$

を送る。ここに、MとKは横11次元のベクトルだとみなし、
足し算はベクトル成分ごとに足す。
ただし、$1{+}1 = 2$は$0, 1$の範疇におさまらないので、$1{+}1 = 0$
という約束で足し算を行う。

$$
\begin{array}{rl}
 & M = 01000100010 \\
+ & K = 10011100101 \\
\hline
 & C = 11011000111
\end{array}
$$

M から C を作ることを**暗号化**という。C を安全でない通信路を用いて B さんに送る。

B さんは、この C に、前に A さんと共有しておいた秘密鍵である K を同様の規則で足すことにより、元の M を求める。このことを**復号化**という。

$$
\begin{array}{rl}
 & C = 11011000111 \\
+ & K = 10011100101 \\
\hline
 & M = 01000100010
\end{array}
$$

- なぜMが復元できるのか？

$$C + K = (M + K) + K = M + (K + K) = M + 0 = M$$

だから。（当たり前ではない。$1 + 1 = 1$と決める数学もあるが、それだと元に戻らない。）

$$\begin{array}{rr} & M = 01000100010 \\ + & K = 10011100101 \\ \hline & C = 11011000111 \\ + & K = 10011100101 \\ \hline & M = 01000100010 \end{array}$$

- Cを傍受した第三者がMを推測することはできないのか?

  Shannon(情報理論の創始者, ATTベル研究所)は上の方式が**安全**であること、すなわち

  「Kが全くデタラメに選んであり、漏えいしなければ、第三者はMを推測することはできない」

  ことを証明した。

より厳密に述べると、

定理(Shannon 1949)
「Kが全くデタラメに選んであり、かつ、このKは漏えいすることなく、一度使われたら二度と使われないならば、第三者はMの長さ以外の情報を推測することはできない」

Shannonの定理の証明の概略：

1. Mを送りたいメッセージとする。
2. KをMと同じ長さの一様ランダムに発生された0-1列とする。
3. このとき、$C = M + K$も（Mがなんであろうと）Mと同じ長さの一様ランダムな0-1列となる。
4. 一様ランダムなCからは、長さ以外の情報は何も得られない。

難しかったので、例をもっと難しくして復習

送りたいデータは、0-1からなる8×11の行列だとする

あらかじめAさんとBさんはサイズ8×11の秘匿鍵K
をデタラメに生成し安全な方法で共有する。
たとえば次を生成したとする。

$$K = \begin{matrix} 11110111011 \\ 10000100000 \\ 10011100101 \\ 11011110010 \\ 10011100000 \\ 11100000100 \\ 10101110000 \\ 11101010111 \end{matrix}$$

あとになって、AさんがBさんに次のメッセージを送りたい
とする。

$$M = \begin{matrix} 00000000000 \\ 00111011100 \\ 01000100010 \\ 01000000010 \\ 00100000100 \\ 00010001000 \\ 00001010000 \\ 00000100000 \end{matrix}$$

このとき、A さんは暗号化文

$$C = M + K = \begin{matrix} 00000000000 \\ 00111011100 \\ 01000100010 \\ 01000000010 \\ 00100000100 \\ 00010001000 \\ 00001010000 \\ 00000100000 \end{matrix} + \begin{matrix} 11110111011 \\ 10000100000 \\ 10011100101 \\ 11011110010 \\ 10011100000 \\ 11100000100 \\ 10101110000 \\ 11101010111 \end{matrix} = \begin{matrix} 11110111011 \\ 10111111100 \\ 11011000111 \\ 10011110000 \\ 10111100100 \\ 11110001100 \\ 10100100000 \\ 11101110111 \end{matrix}$$

を生成し、ダダ漏れの通信路で B さんに送る。

Bさんは受け取ったCとKを比べる：

$$C = \begin{matrix} 11110111011 \\ 10111111100 \\ 11011000111 \\ 10011110000 \\ 10111100100 \\ 11110001100 \\ 10100100000 \\ 11101110111 \end{matrix}, K = \begin{matrix} 11110111011 \\ 10000100000 \\ 10011100101 \\ 11011110010 \\ 10011100000 \\ 11100000100 \\ 10101110000 \\ 11101010111 \end{matrix}, \text{重ねて} \begin{matrix} \mathbf{11110111011} \\ \mathbf{10\textcircled{1}1\textcircled{1}1\textcircled{1}0\textcircled{1}0\textcircled{1}00} \\ \mathbf{10\textcircled{1}1\textcircled{0}0\textcircled{0}1\textcircled{1}1} \\ \mathbf{10\textcircled{1}1\textcircled{1}1\textcircled{1}0\textcircled{0}\textcircled{1}0} \\ \mathbf{10\textcircled{1}1\textcircled{1}1\textcircled{0}0\textcircled{1}00} \\ \mathbf{111\textcircled{1}0\textcircled{0}0\textcircled{0}\textcircled{1}100} \\ \mathbf{10\textcircled{1}0\textcircled{0}1\textcircled{1}0\textcircled{0}000} \\ \mathbf{111\textcircled{0}1\textcircled{0}10111} \end{matrix}$$

・Cだけからは求まらないMが、Kもあれば復元される

Bさんは受け取ったCとKを比べる：

$$C = \begin{matrix} 11110111011 \\ 10111111100 \\ 11011000111 \\ 10011110000 \\ 10111100100 \\ 11110001100 \\ 10100100000 \\ 11101110111 \end{matrix}, K = \begin{matrix} 11110111011 \\ 10000100000 \\ 10011100101 \\ 11011110010 \\ 10011100000 \\ 11100000100 \\ 10101110000 \\ 11101010111 \end{matrix}, C+K = \begin{matrix} 00000000000 \\ 00111011100 \\ 01000100010 \\ 01000000010 \\ 00100000100 \\ 00010001000 \\ 00001010000 \\ 00000100000 \end{matrix}$$

・Cだけからは求まらないMが、Kもあれば復元される
・大量のデータ（例：動画）を暗号化するには
　大量のデタラメさ（大量の乱数）が必要となる

確率的現象をシミュレーションするためのデタラメさ

（モンテカルロシミュレーションと呼ばれる）
もっとも大量のデタラメさを必要とする例

デジタルコンピュータで最初に演算により
「デタラメさ」を生成して使ったのはおそらくvon Neumann
(河東さんの講義第三回にもでてきた、２０世紀最大の数学者・計算機科学者の一人)

第二次大戦中の米国、核爆発の計算機シミュレーションに使われた（マンハッタン計画）。

計算機シミュレーション：現象を数値化・数式化し、計算により計算機の中でその現象を模倣する。

核爆発（連鎖核分裂）：それまで人類は目撃したことのない現象。

物理的・数学的に予言されただけの現象であるため、
実験する前に（後にも）計算機シミュレーションが行われた。

- 一つのウラン原子が、単位時間あたりにある確率で分裂する（確率的な、デタラメさを伴う現象）
- 分裂したウランは数個の中性子をある方向に発射する（これも確率的）。それが当たったウランは分裂する確率が高い。
- 原子の個数は膨大。それぞれに対し、単位時間当たりに「デタラメな数＝乱数」を発生させる必要がシミュレーションではある。

余りに大量なため、乱数表をメモリ内に読み込むことも、
物理的装置を使ってデタラメさを発生することも現実的で
はない。

von Neumannは、コンピュータの演算を使って
「疑似デタラメさ＝**疑似乱数**」
を発生させて、核反応シミュレーションに利用した。

最初の「疑似乱数発生法」とみなされている
（1940年代始めらしい、軍事機密で確かではない）。

モンテカルロシミュレーションは、

- 科学　あらゆる物理現象のシミュレーションや、
- 生物　DNA配列から得られるたんぱく質の構造の予測
- 経済金融における株価の予測など、

確率的要素を持つあらゆる分野で大量に使われている。

モンテカルロシミュレーションのためには、
「予測不可能性」は重要ではなく、
「一様性」「独立性」「高速性」「再現可能性」が重要となる。

再現可能性については、次回。

以上、第壱話：「デタラメさの効用：意外なところでランダムネスのお世話に」（地之章）終

次回予告：「第弐話：デタラメさを生み出すのは意外に難しい：デタラメ禅問答」（天之章）