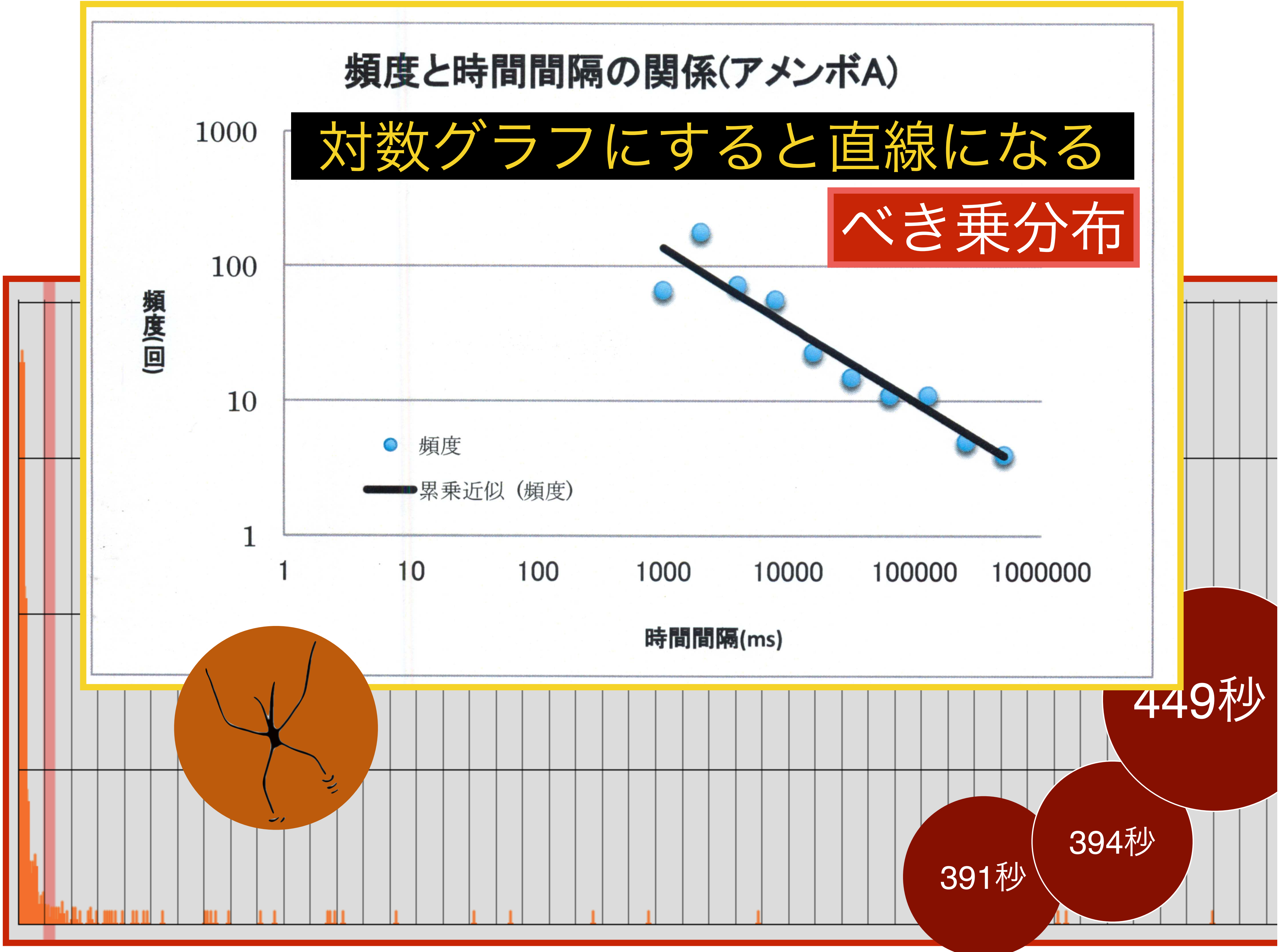
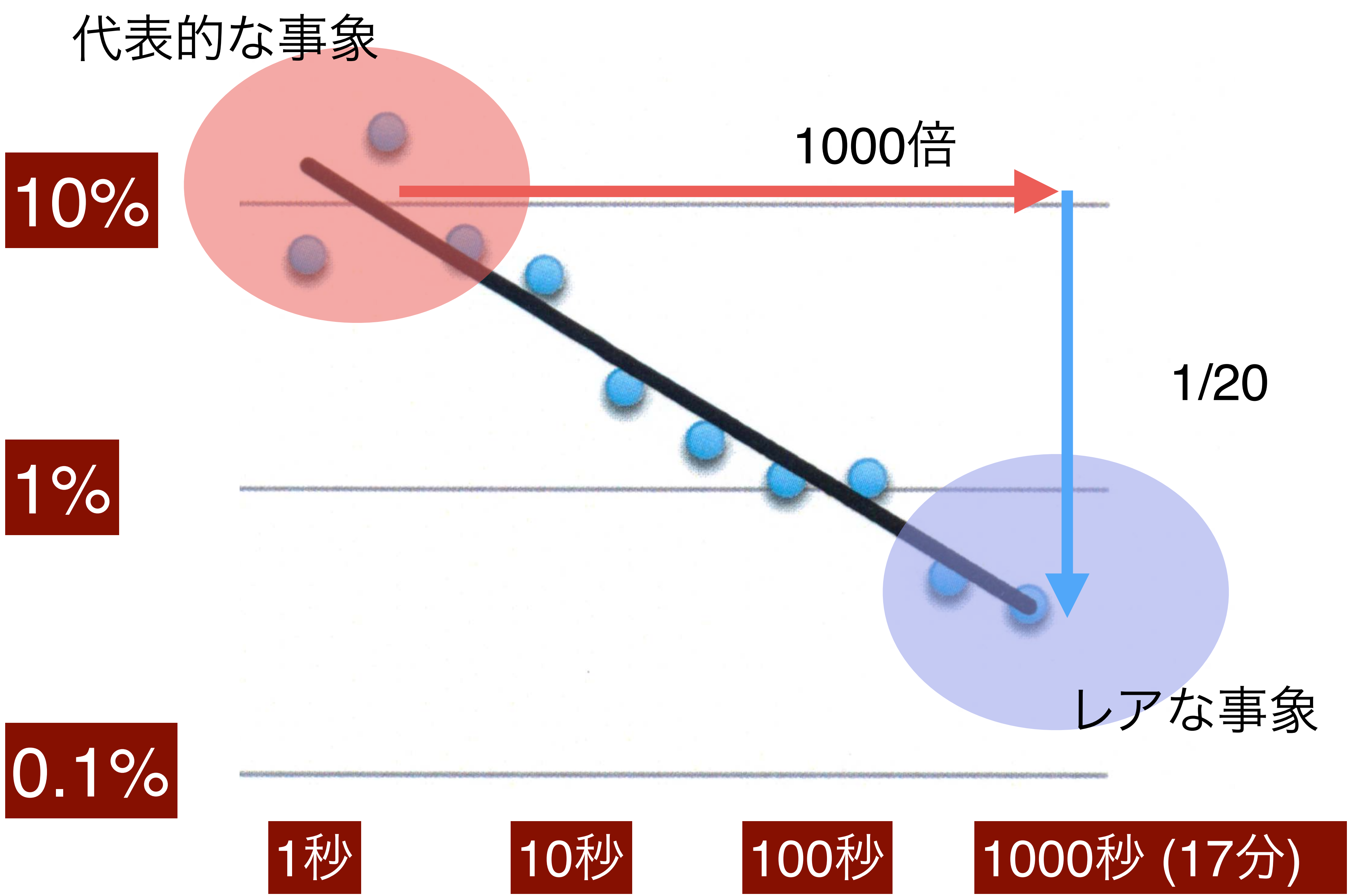
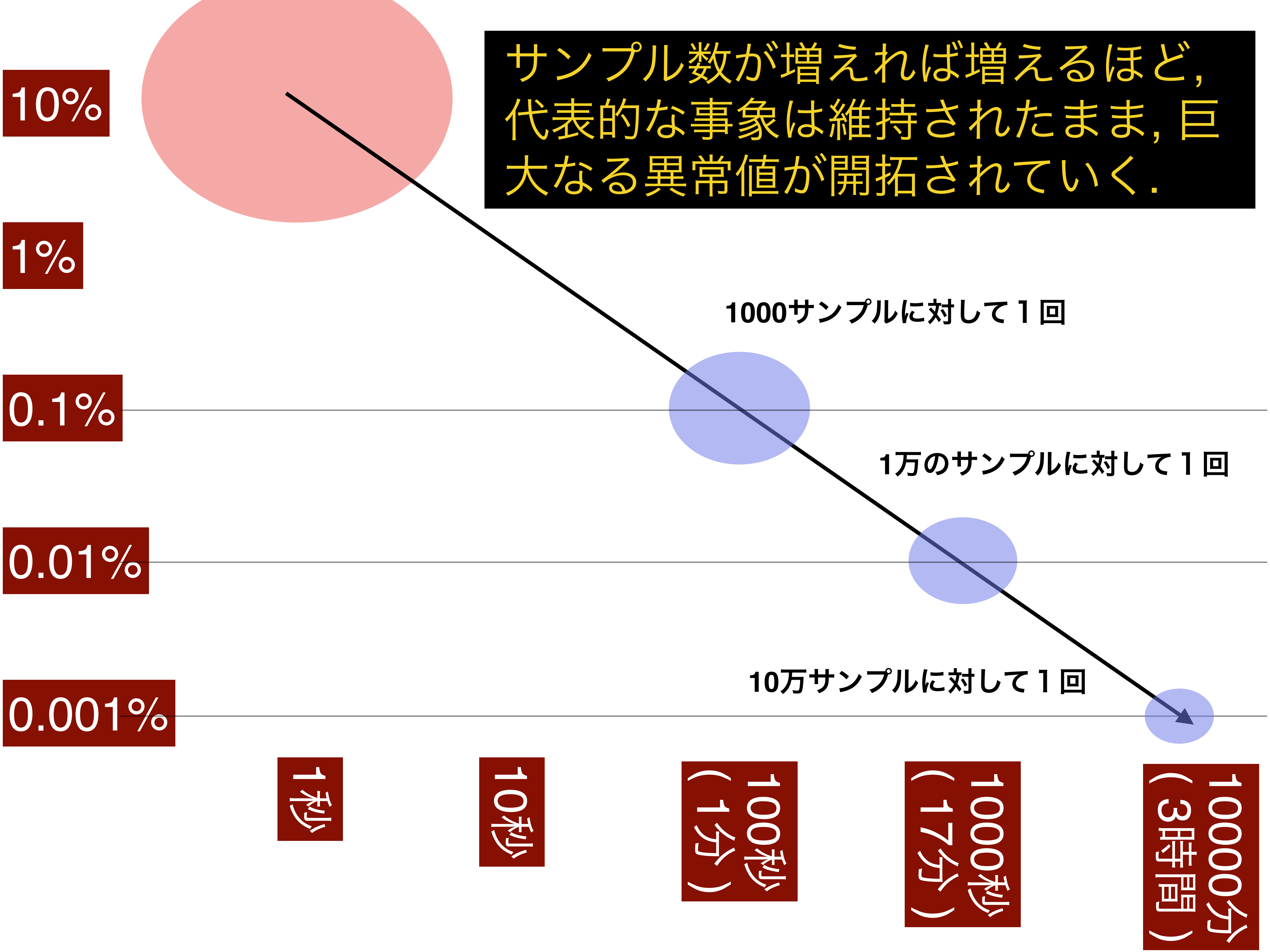
べき乗分布

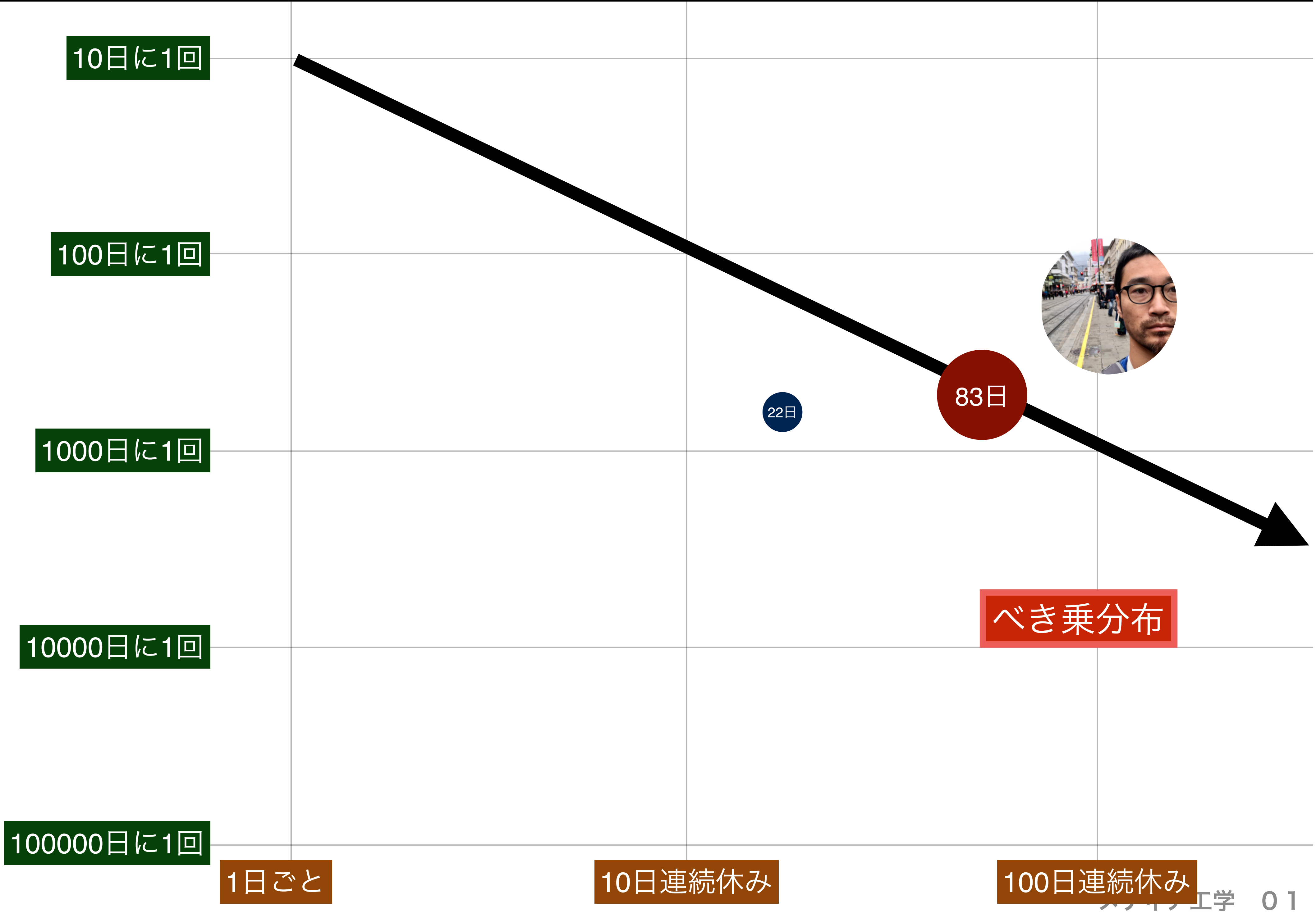


頻度と時間間隔の関係(アメンボA)
対数グラフにすると直線になる
べき乗分布
1000
100
10
1
頻度(回)
頻度
累乗近似（頻度）
1
10
100
1000
10000
100000
1000000
時間間隔(ms)
449秒
394秒
391秒

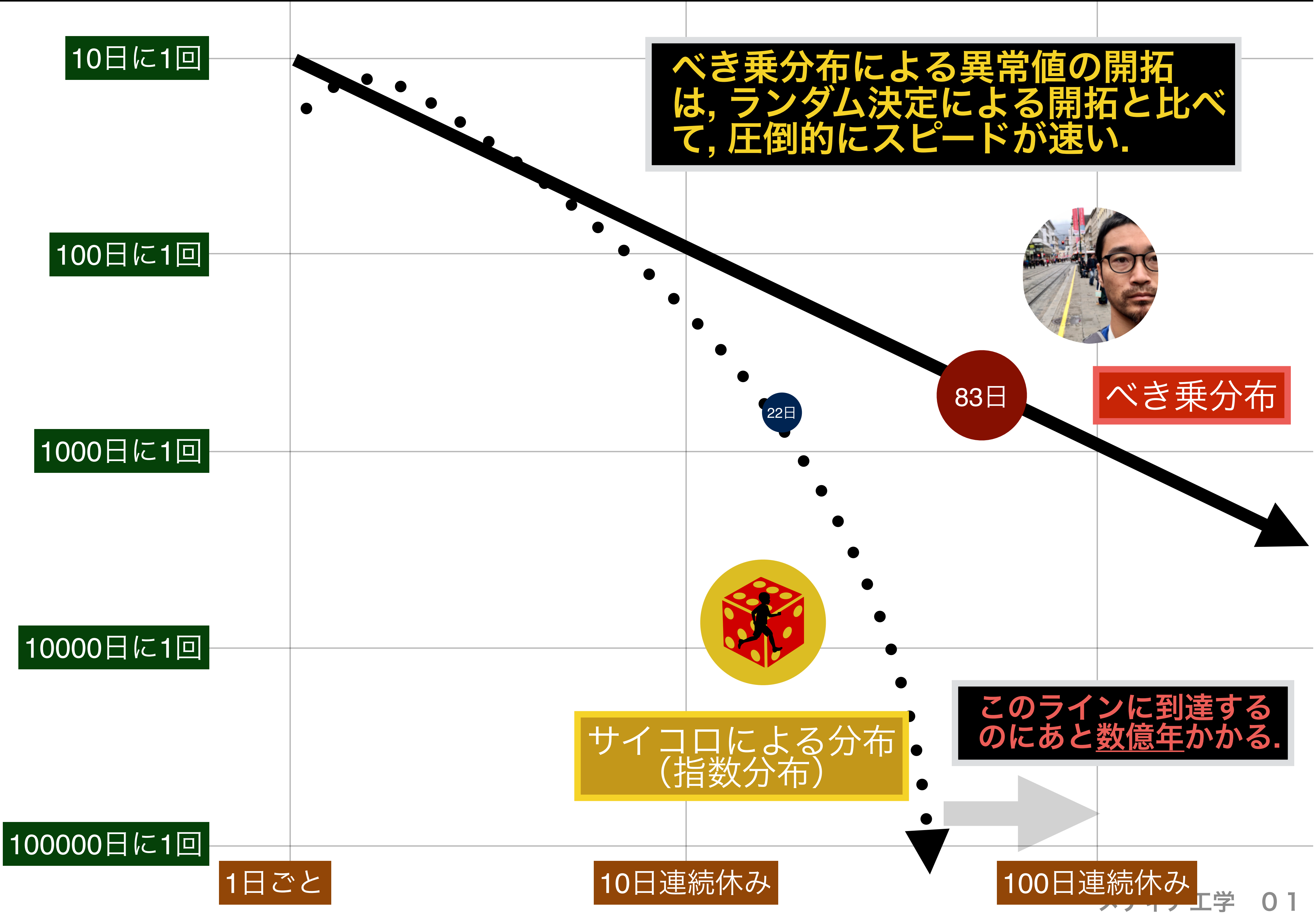

代表的な事象
1000倍
10%
1/20
1%
レアな事象
0.1%
1秒
10秒
100秒
1000秒 (17分)


サンプル数が増えれば増えるほど，代表的な事象は維持されたまま，巨大なる異常値が開拓されていく．
10%
1%
0.1%
0.01%
0.001%
1000サンプルに対して１回
1万のサンプルに対して１回
10万サンプルに対して１回
1秒
10秒
100秒（1分）
1000秒（17分）
10000分（3時間）

べき乗分布の補足


10日に1回
100日に1回
1000日に1回
10000日に1回
100000日に1回
1日ごと
10日連続休み
100日連続休み
22日
83日
べき乗分布

べき乗分布の補足

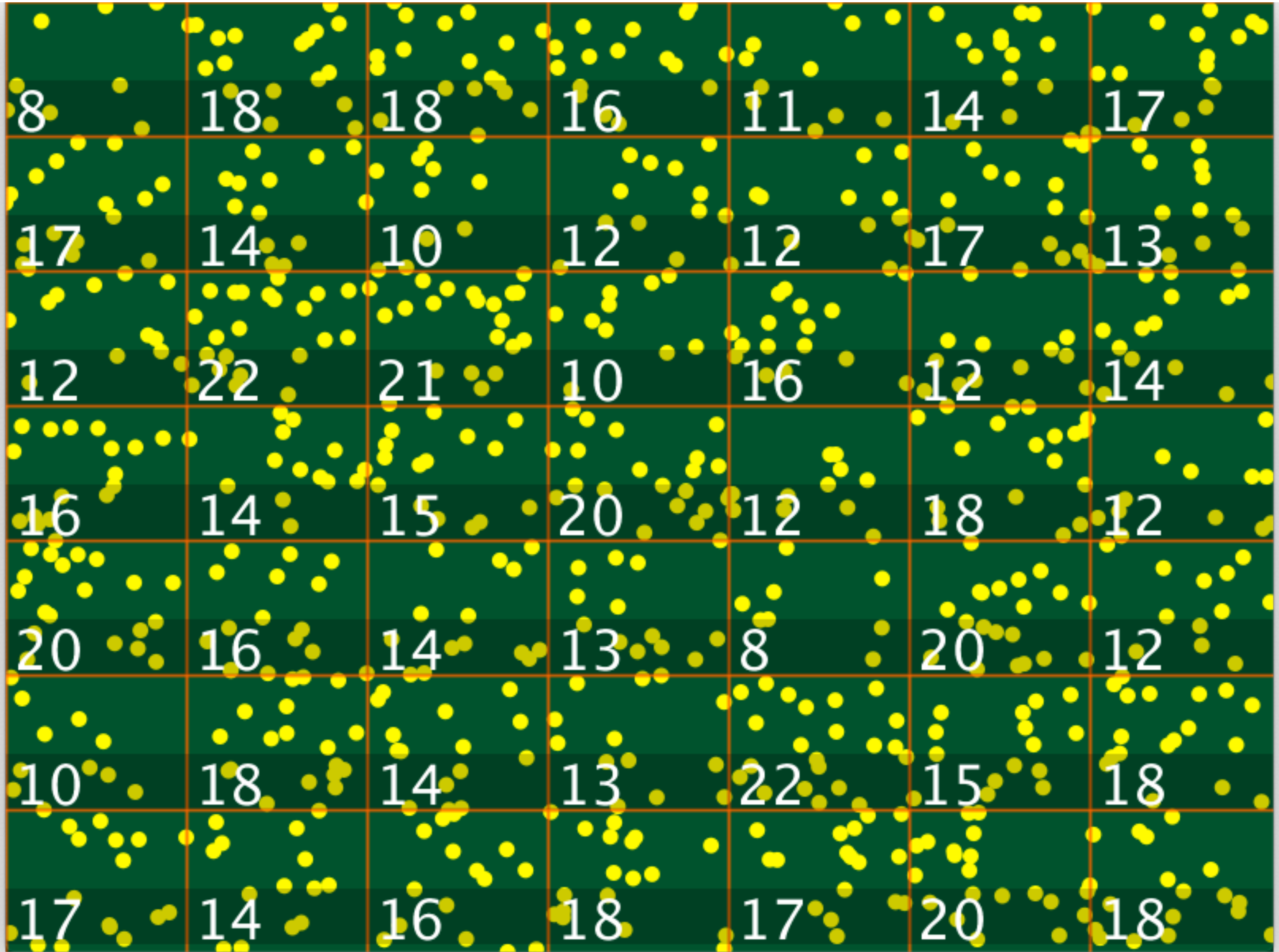


10日に1回
100日に1回
1000日に1回
10000日に1回
100000日に1回
1日ごと
10日連続休み
100日連続休み
べき乗分布による異常値の開拓は, ランダム決定による開拓と比べて, 圧倒的にスピードが速い.
83日
22日
べき乗分布
サイコロによる分布（指数分布）
このラインに到達するのにあと数億年かかる.

ランダム関数によって作られる確率分布の数学的性質

愛知県のゾンビ市町村

745万人を49（7 x 7）市町村へ割り当てる

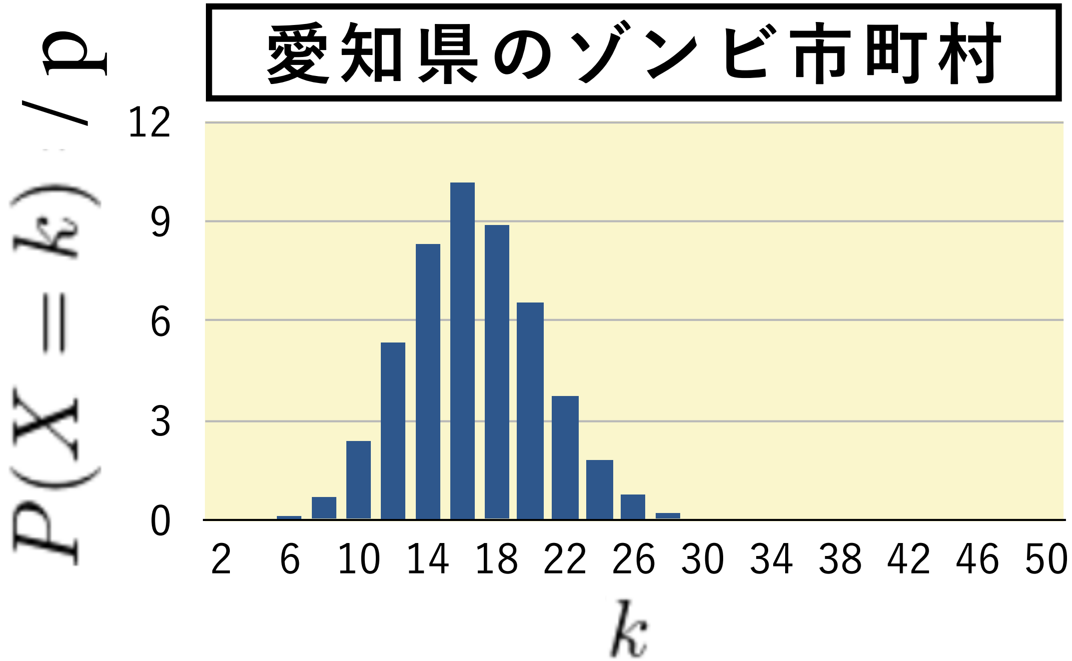


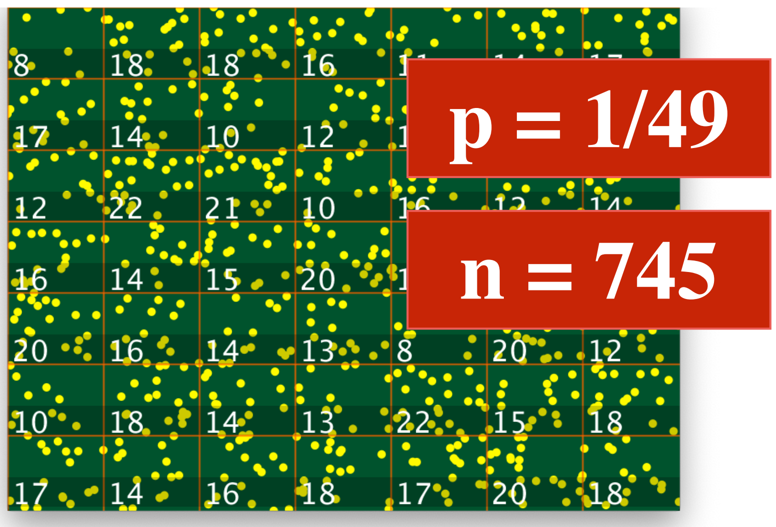
二項分布

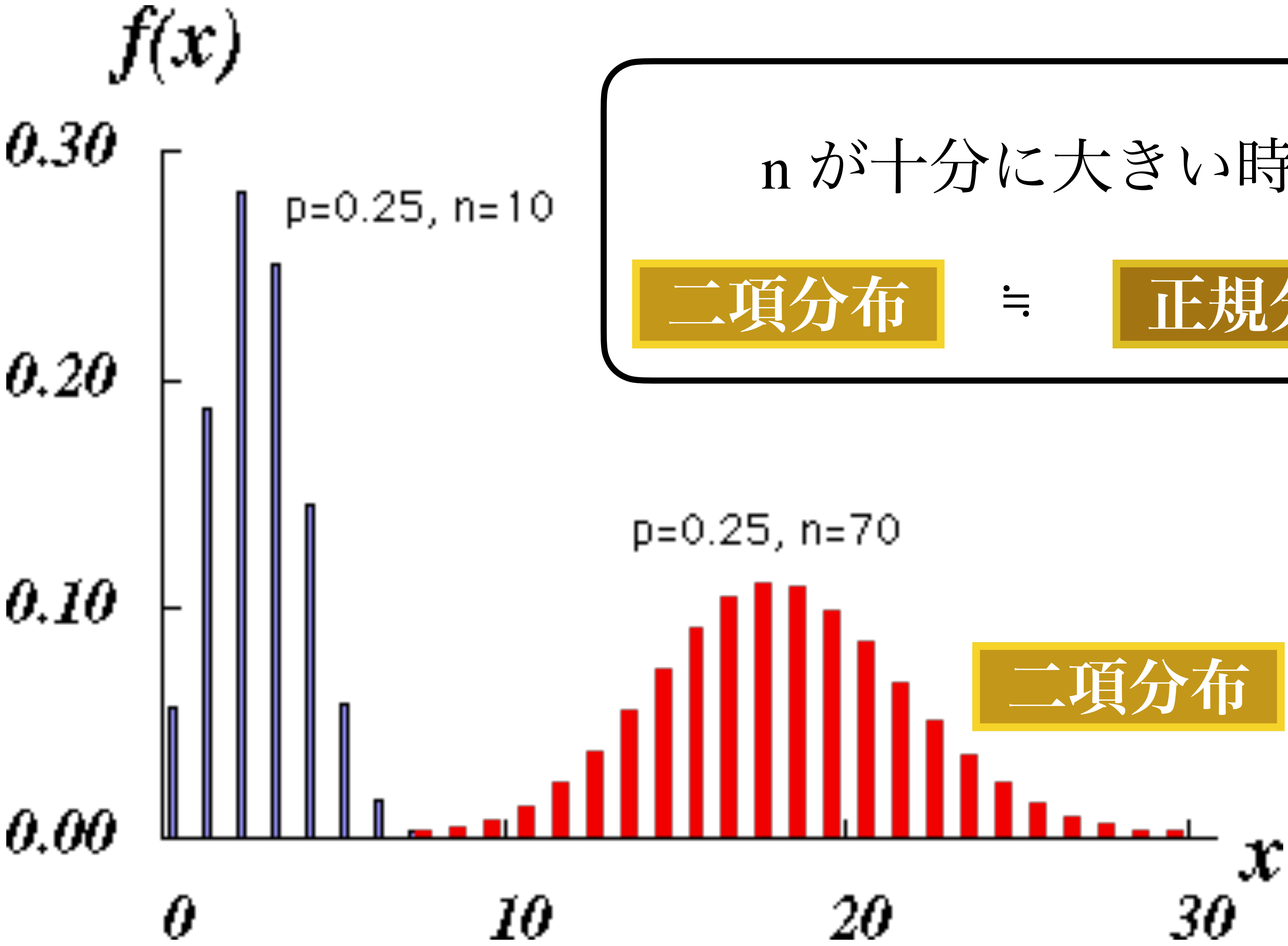
一回の成功確率が「p」の試行を「n」回行う.

このとき,「n」回中,「k」回成功する確率は,

$$P(X=k) = {}_n\mathrm{C}_k p^k (1-p)^{n-k} \quad (k = 0, 1, 2, \cdots, n)$$







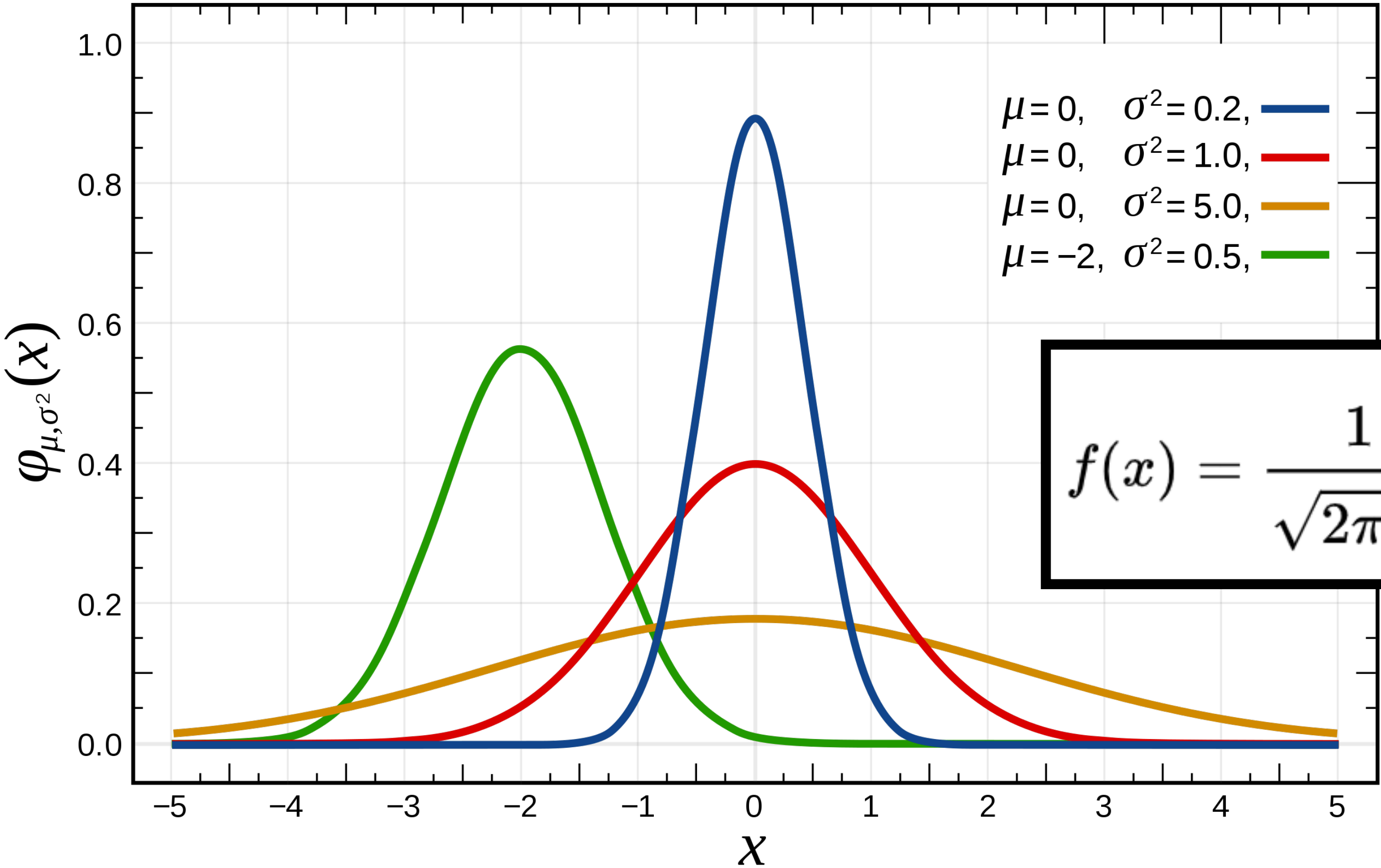
n が十分に大きい時,

二項分布 ≒ 正規分布

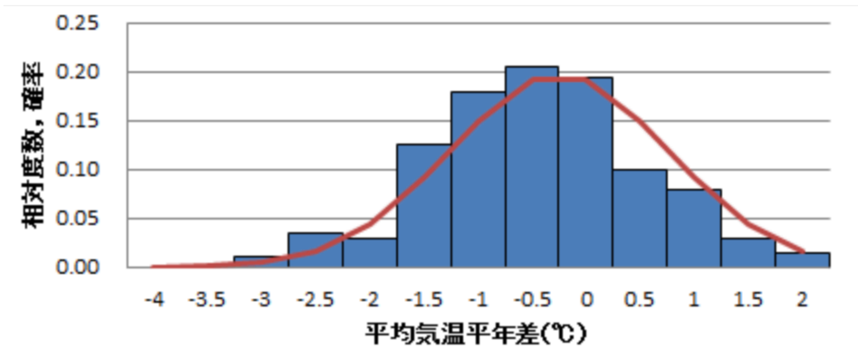
https://www.weblio.jp/content/%E4%BA%8C%E9%A0%85%E5%88%86%E5%B8%83

正規分布

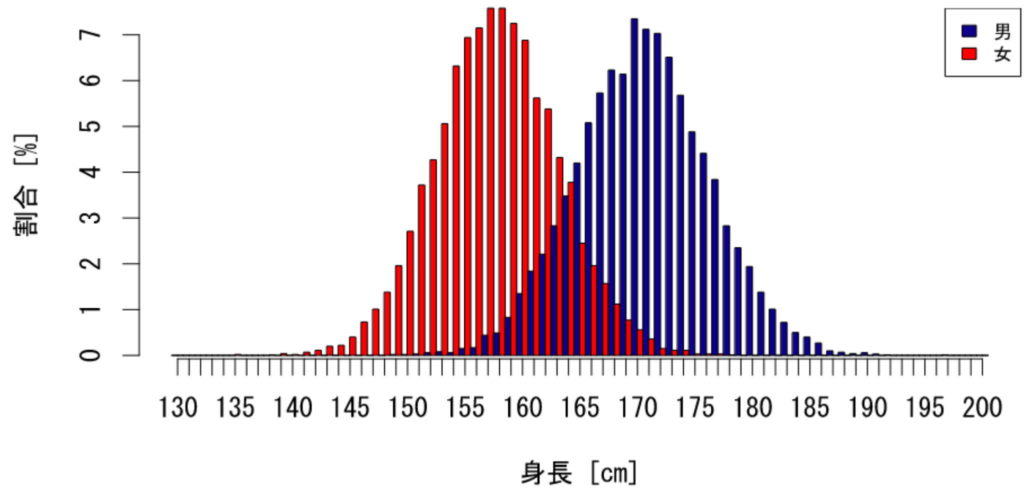
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



平均気温・平年差（6-8月）

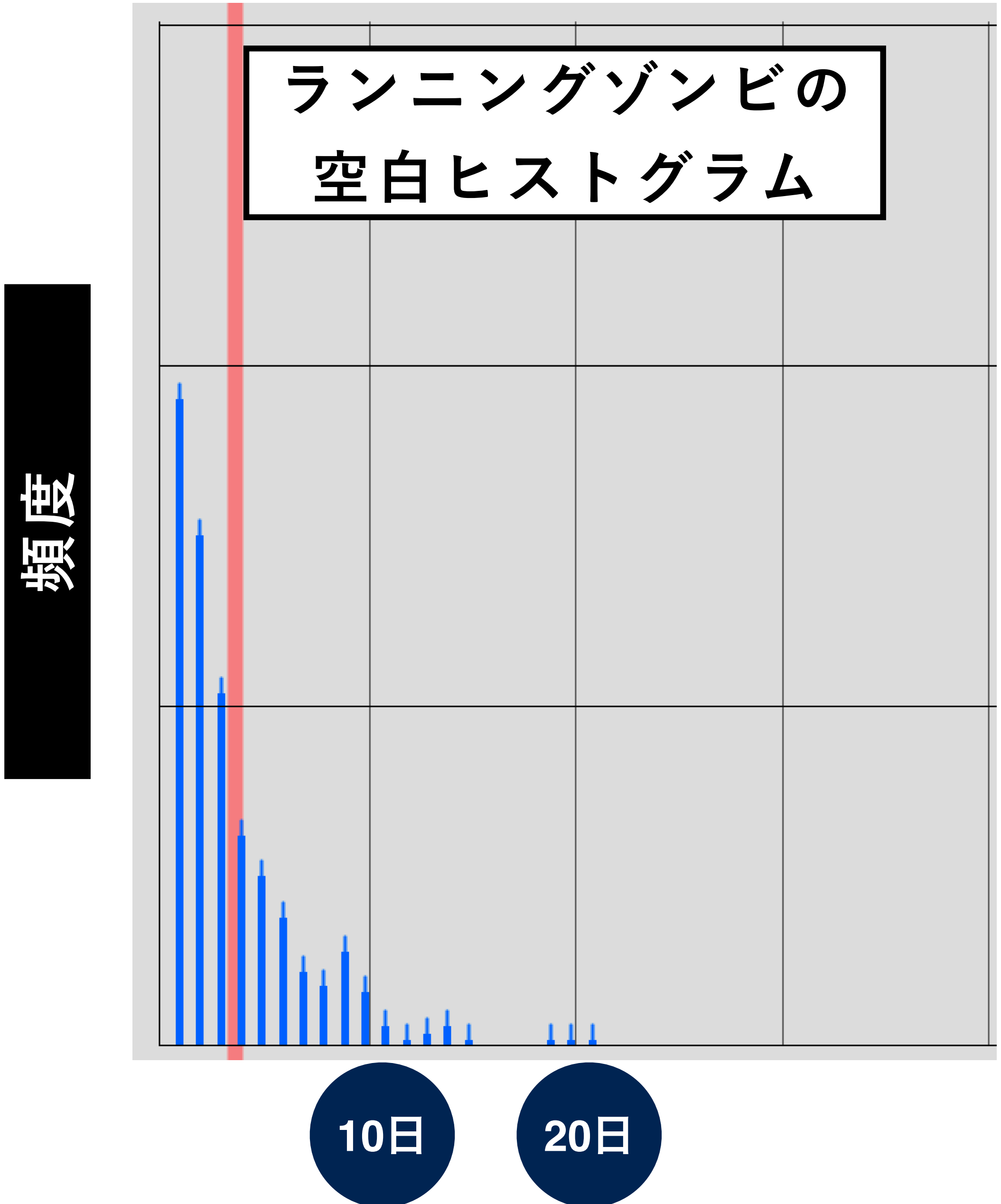


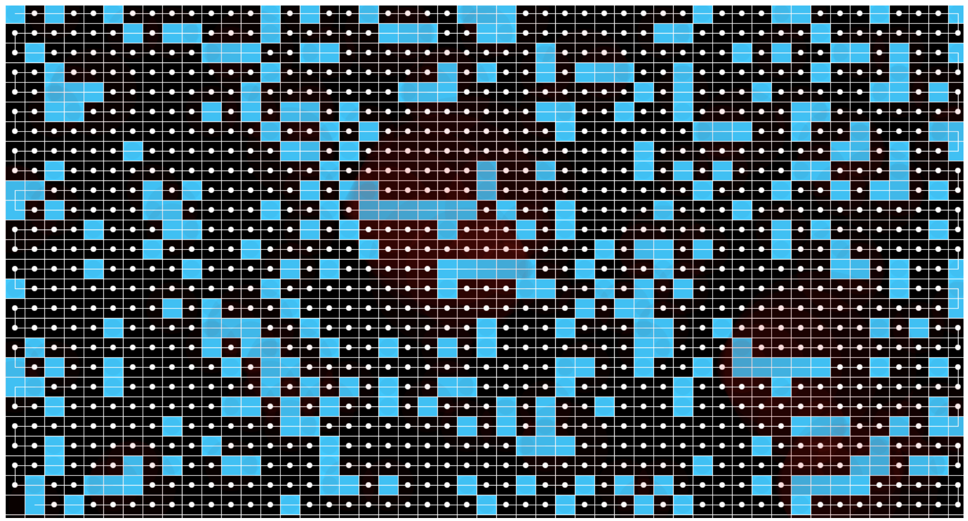
17歳の身長分布（日本）



2012/10/13 - 2016/04/08

（走った日）326 / 1273 = 0.256





指数分布

単位時間あたり平均「λ」回起きる事象について,

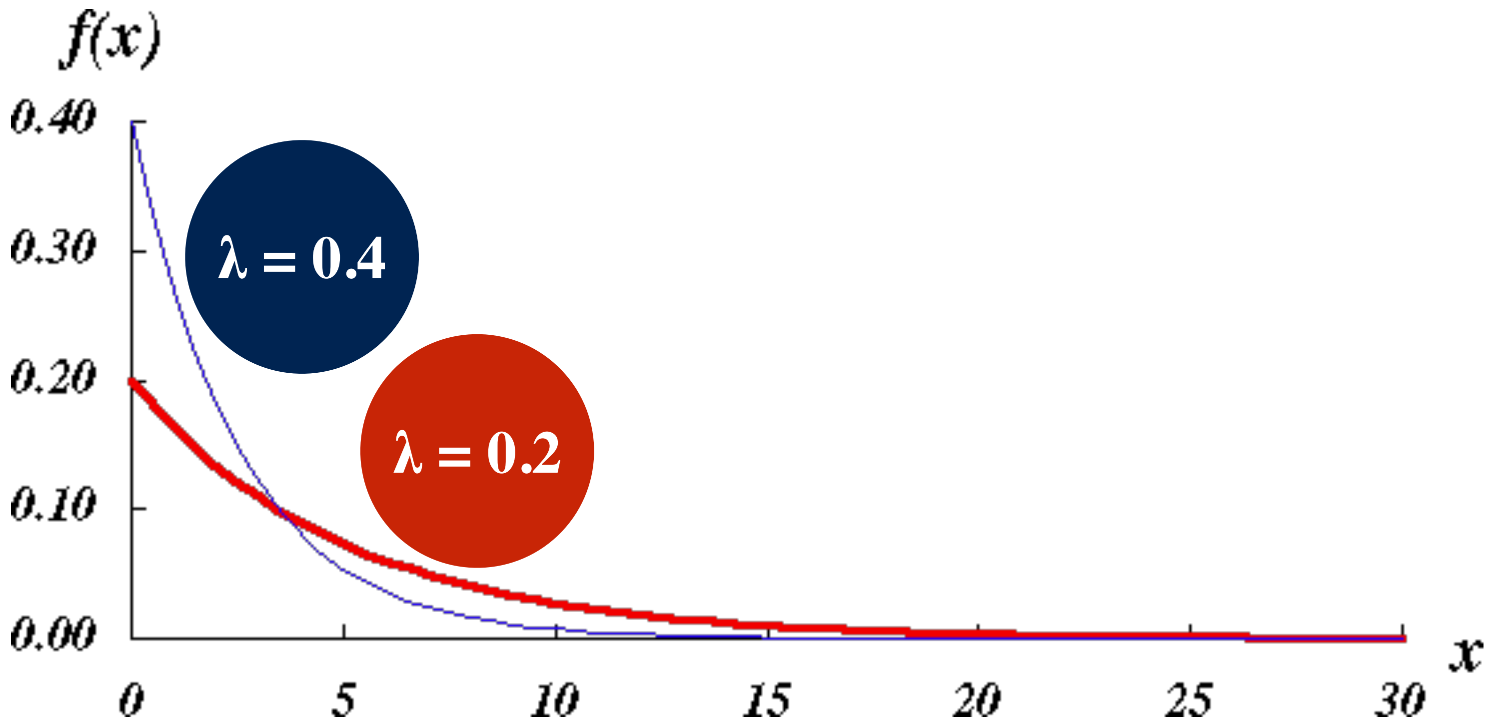
（ある時間において）その事象が生じてから, 次の事象が生じるまでの時間が「x」となる確率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

指数分布

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

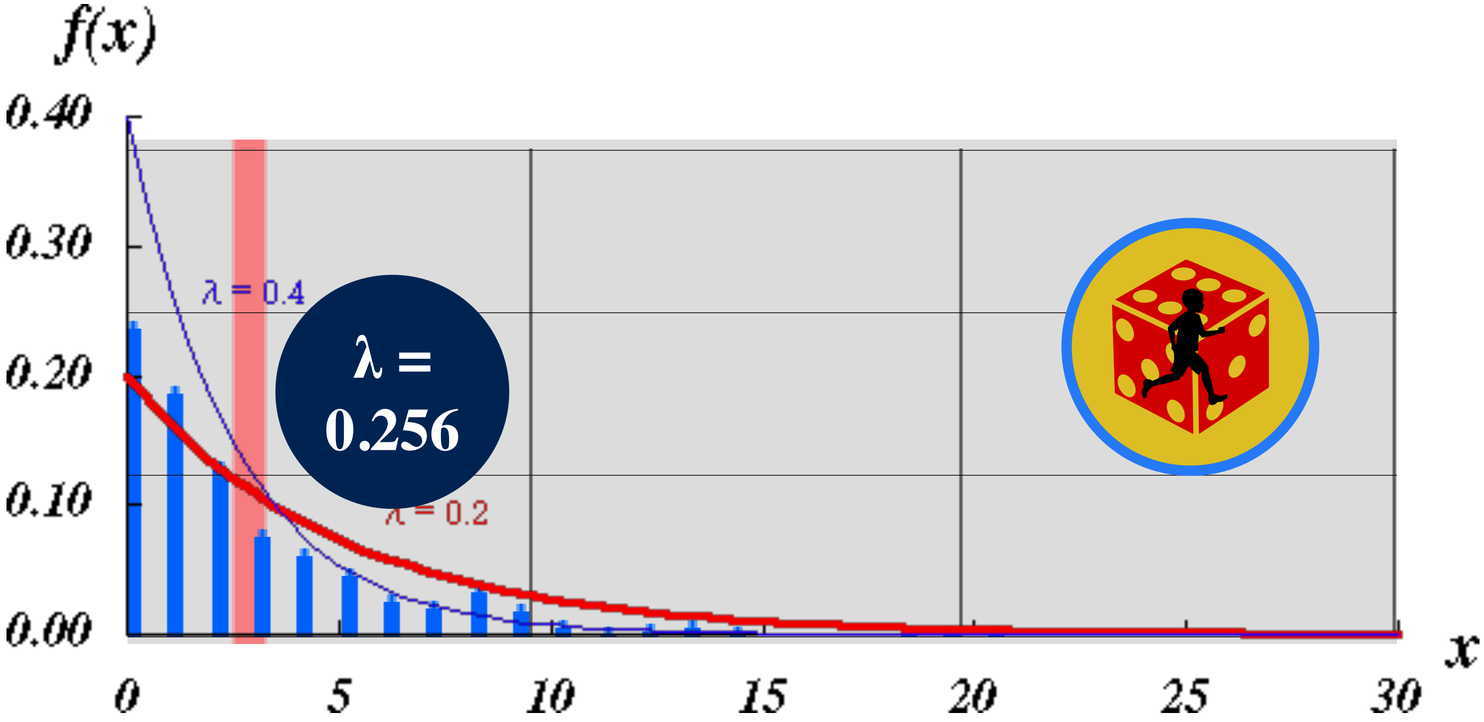
一日あたり平均「λ = 0.256」回起きる事象について，（ある時間において）その事象が生じてから，次の事象が生じるまでの時間が「x（日）」となる確率密度



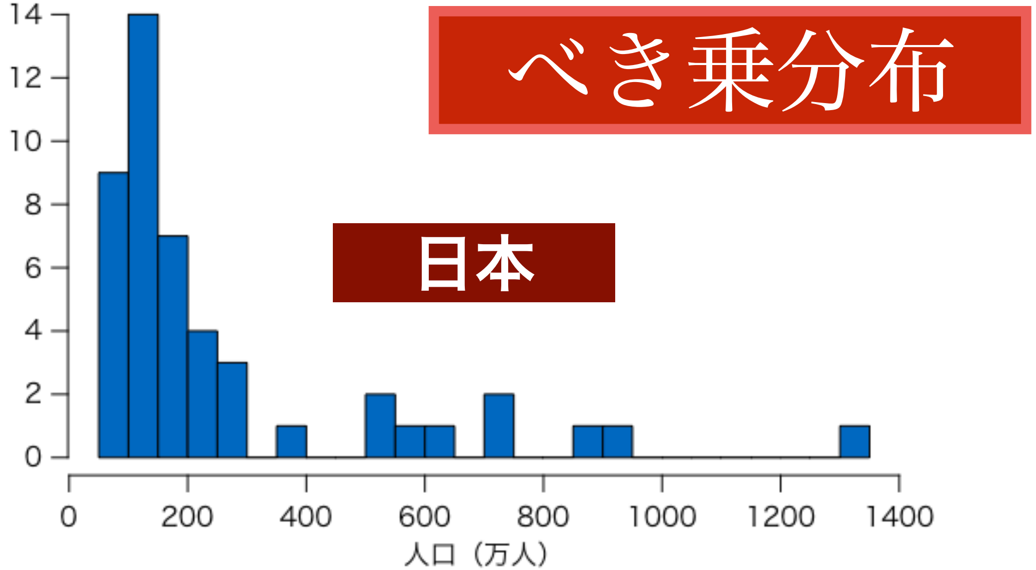
指数分布

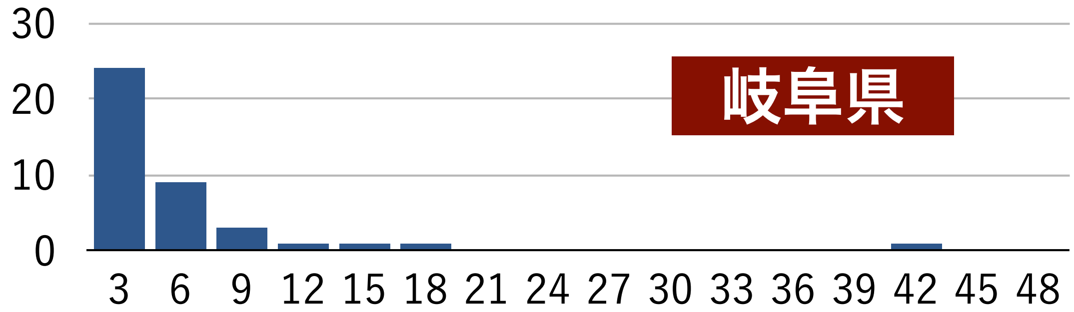
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

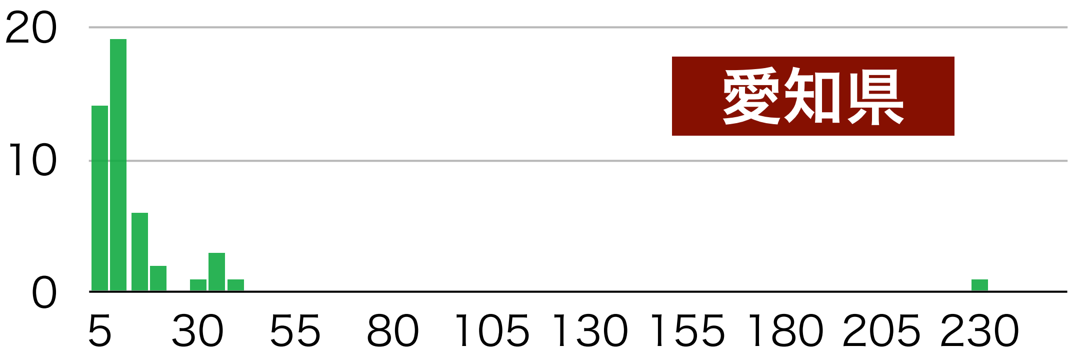
一日あたり平均「λ = 0.256」回起きる事象について，（ある時間において）その事象が生じてから，次の事象が生じるまでの時間が「x（日）」となる確率密度

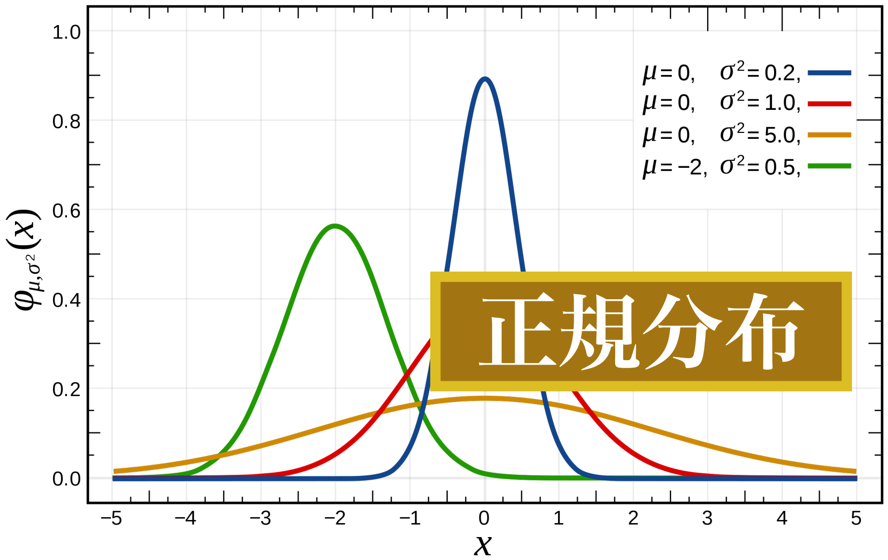


空間に関する分布

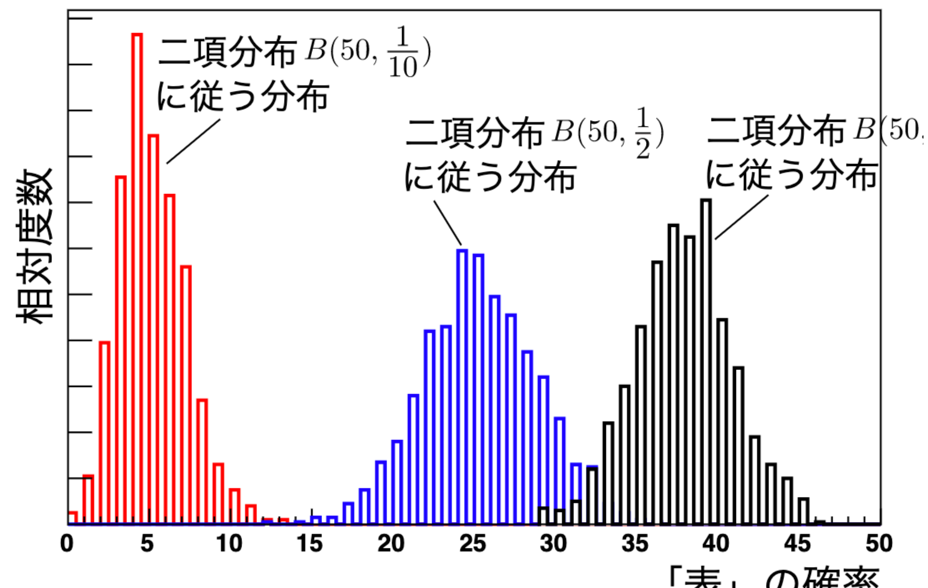






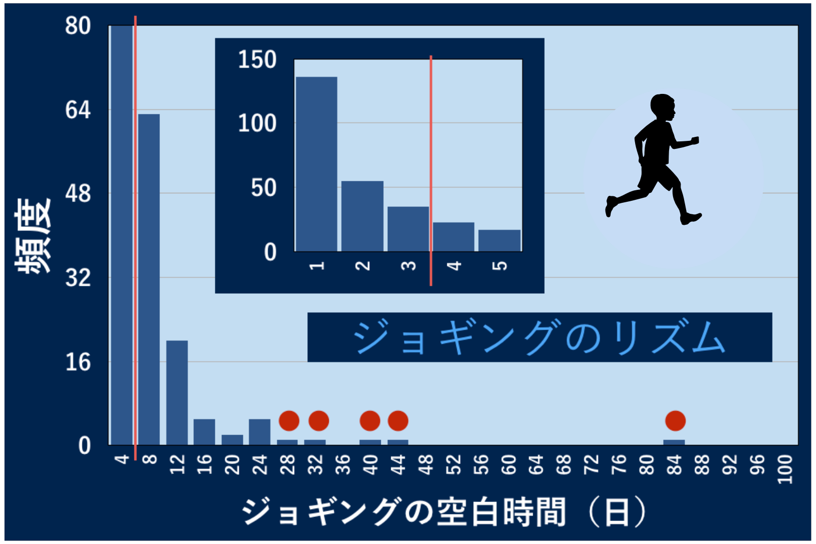


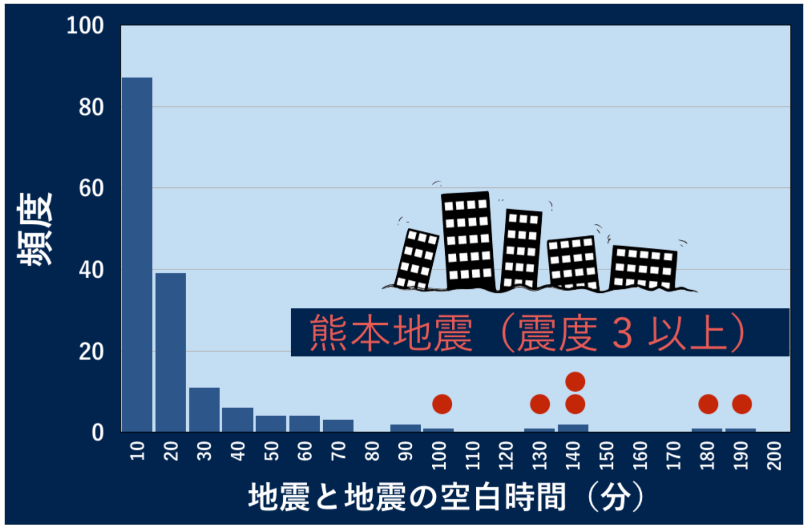
二項分布



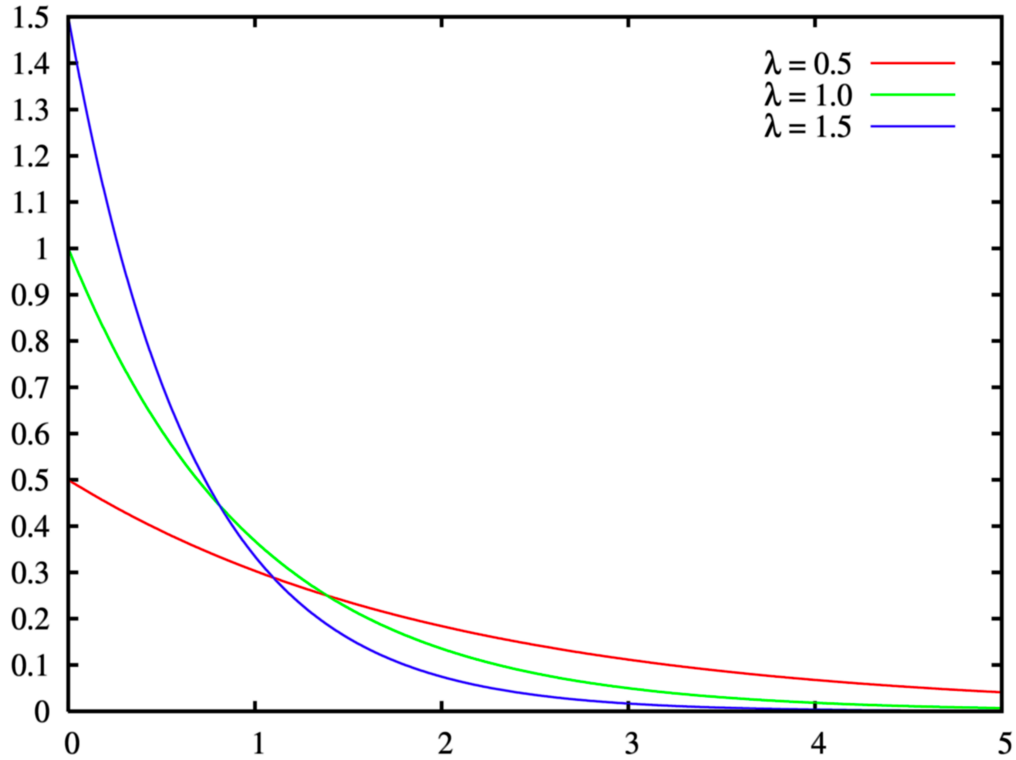
時間に関する分布（空白ヒストグラム）

べき乗分布





指数分布



$\boxed{\text{空間分布の非ランダム性の指標}}$

🔶 1. 分散比（Index of Dispersion / Variance-to-Mean Ratio）

概要：

- 各マスに含まれるたんぽぽの数（=セル値）の**分散**を、その**平均**で割ったもの。
- 記号で書くと：

$$\mathrm{ID} = \frac{\mathrm{Var}(x)}{\mathrm{Mean}(x)}$$

解釈：

- **ID ≈ 1**：ポアソン分布（完全ランダム配置）の特徴 → ランダム
- **ID > 1**：分布が**クラスタ化**している（偏りあり）
- **ID < 1**：分布が**均質化**している（偏りが少ない）

たんぽぽの場合の応用：

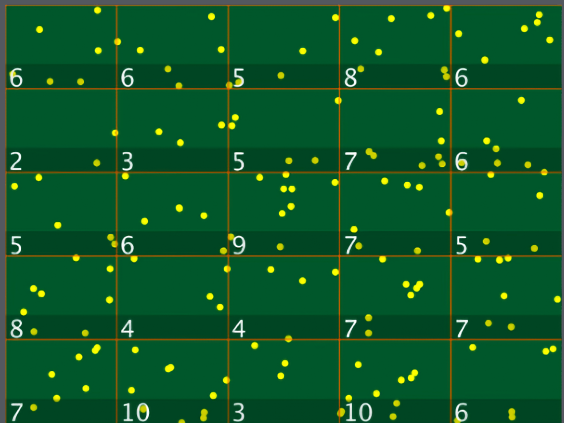
- 各マスに何本たんぽぽがあるかを数える → 25個の値を得る
- それらの分散と平均を計算 → 比率が **1 を大きく超えていたら** 偏りあり

空間分布の分散比（ID）算出(例)
セル値から平均・分散へ

算出の流れ

ID = Index of Dispersion（分散比）
各マスの個体数（セル値）の分散を，その平均で割ったもの

① 観察領域を25マスに分割し，各セルの個体数を数える

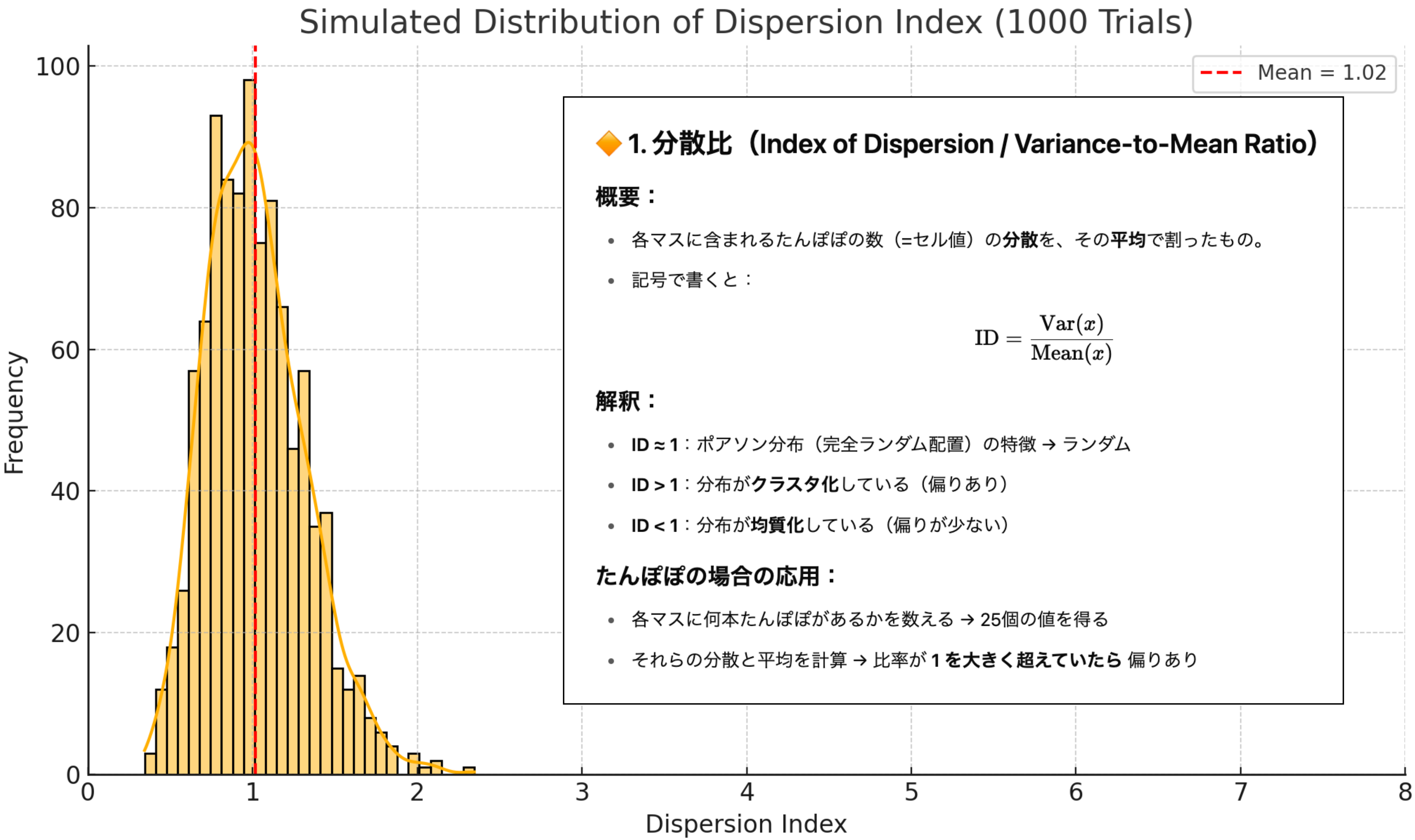


セル値列（25マスの個体数）
6, 6, 5, 8, 6,
2, 3, 5, 7, 6,
5, 6, 9, 7, 5,
8, 4, 4, 7, 7,
7, 10, 3, 10, 6

② 25個のセル値を得る：　$x_1, x_2, x_3, \ldots, x_{25}$

セル値列 $x_1, x_2, \ldots$	→	平均 μ = Mean(x)	→	分散 Var(x)	→	ID ID = Var(x) / Mean(x)

分散比（分散/平均）の分布



🔶 1. 分散比（Index of Dispersion / Variance-to-Mean Ratio）

概要：

- 各マスに含まれるたんぽぽの数（=セル値）の**分散**を、その**平均**で割ったもの。
- 記号で書くと：

$$\mathrm{ID} = \frac{\mathrm{Var}(x)}{\mathrm{Mean}(x)}$$

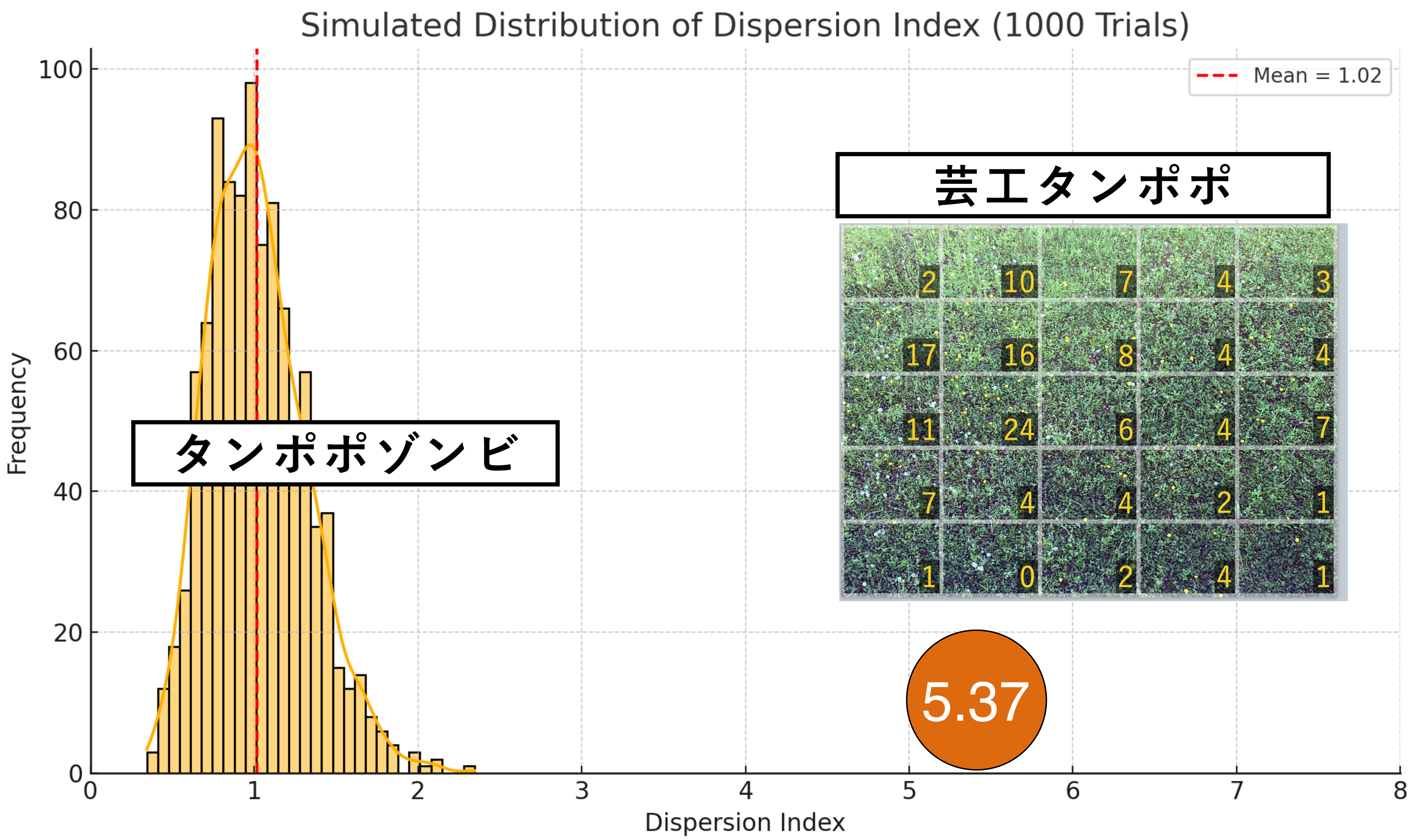
解釈：

- **ID ≈ 1**：ポアソン分布（完全ランダム配置）の特徴 → ランダム
- **ID > 1**：分布が**クラスタ化**している（偏りあり）
- **ID < 1**：分布が**均質化**している（偏りが少ない）

たんぽぽの場合の応用：

- 各マスに何本たんぽぽがあるかを数える → 25個の値を得る
- それらの分散と平均を計算 → 比率が **1 を大きく超えていたら** 偏りあり

分散比（分散/平均）の分布


Simulated Distribution of Dispersion Index (1000 Trials)
Mean = 1.02
Frequency
Dispersion Index
0
20
40
60
80
100
0
1
2
3
4
5
6
7
8
タンポポゾンビ
芸工タンポポ
2 10 7 4 3
17 16 8 4 4
11 24 6 4 7
7 4 4 2 1
1 0 2 4 1
5.37

時間分布の非ランダム性の指標

CV（変動係数, coefficient of variation）

CV の定義

$$\mathrm{CV} = \frac{\text{標準偏差}}{\text{平均}} = \frac{\sigma}{\mu}$$

解釈の目安

- CV = 0：完全に規則的
- CV = 1：ランダム（ポアソン過程）
- CV > 1：ばらつきが大きい（バースト的）

時間的事象のCV算出
IEIから平均・標準偏差へ

算出の流れ

IEI = Inter-Event Interval（事象間隔）
隣り合う事象の「時間差」を並べたもの

① まず, 事象が起きた時刻を時間軸上に並べる

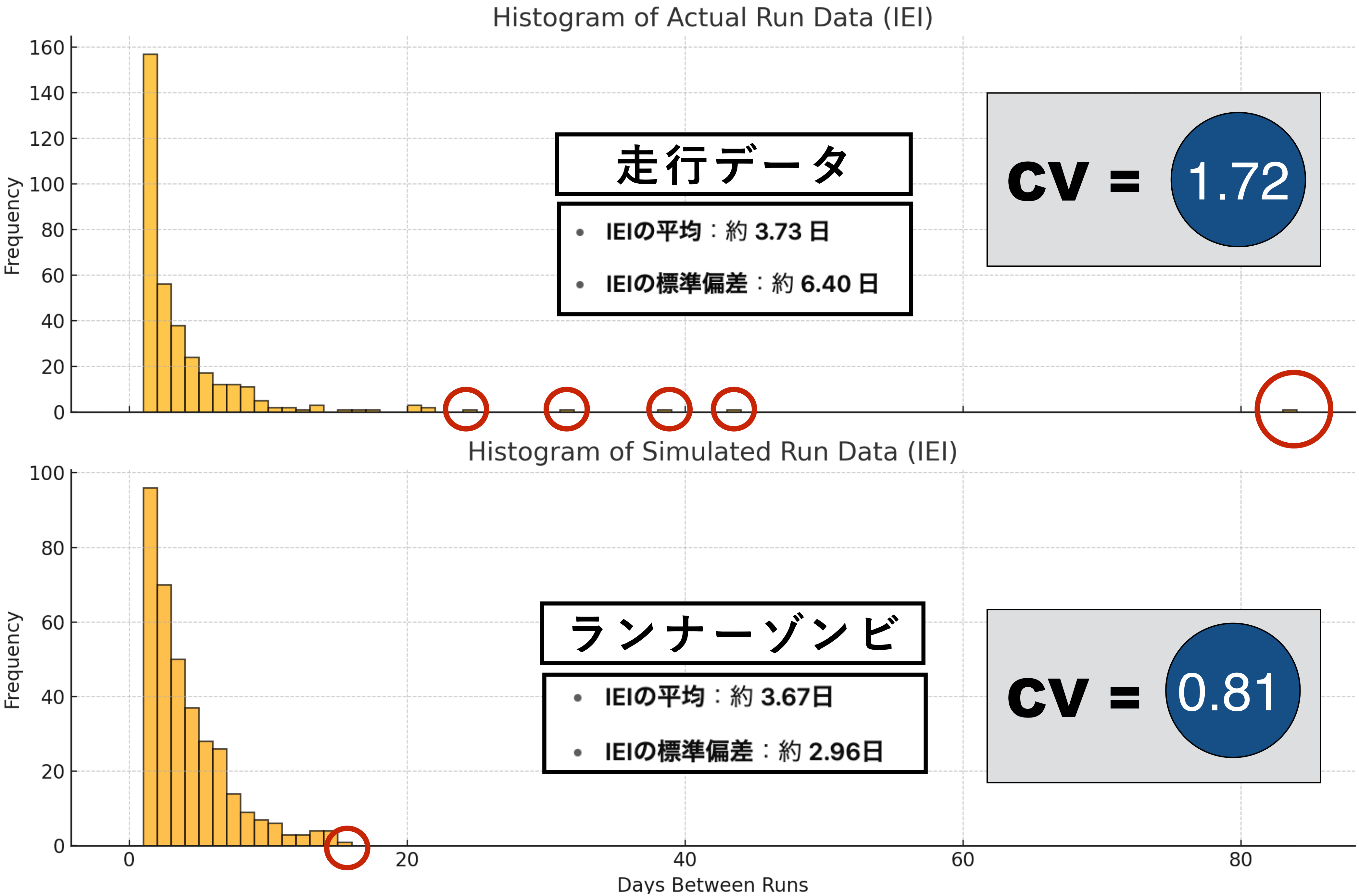
② 隣同士の差を取る： $IEI_i = t_{i+1} - t_i$

IEI列	→	平均	→	標準偏差	→	CV
$IEI_1, IEI_2, \ldots$		μ = Mean(IEI)		σ = SD(IEI)		CV = σ / μ

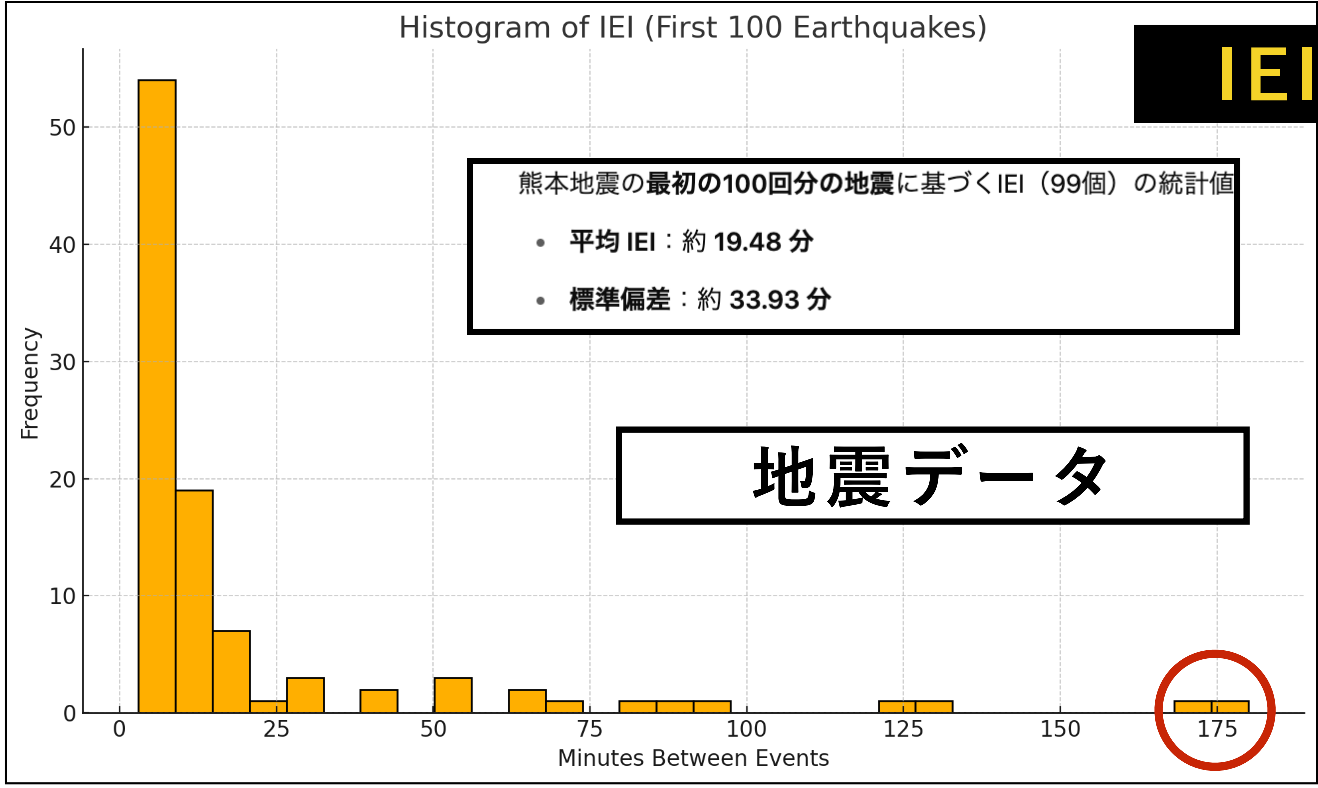
③ IEIのばらつきを, IEIの平均で割る。
→ 平均間隔に対して「どれくらい不規則か」を表す。

※ 本授業では小数表記：CV = 1 は CV = 100% と同じ意味

IEI(日数)のヒストグラム



IEIのヒストグラム

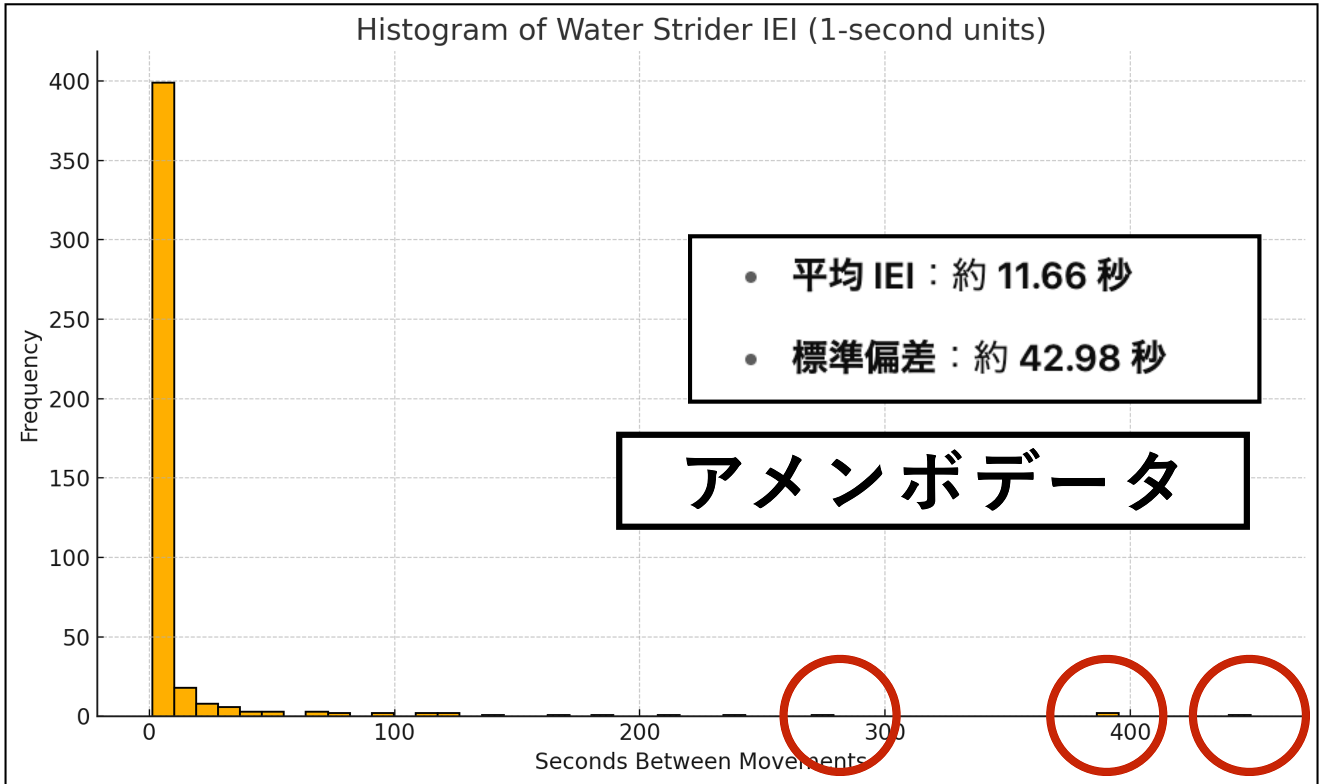


熊本地震の**最初の100回分の地震**に基づくIEI（99個）の統計値

- **平均 IEI**：約 **19.48 分**
- **標準偏差**：約 **33.93 分**

地震データ

CV = 1.74



- **平均 IEI**：約 **11.66 秒**
- **標準偏差**：約 **42.98 秒**

アメンボデータ

CV = 3.69