Hypothesis Testing via Simulation

J. Doi, Cal Poly SLO

Topic 1: Hypothesis Testing (仮説検定) via Simulation [シミュレーションに基づく推論]

METHOD:

- Consider a particular **hypothesis** (仮説)
- Collect **evidence** (証拠)
- Make a **decision** (結論) about the hypothesis based on the evidence

証拠が予想内 → 仮説を棄却しない, 証拠が予想外 → 仮説を棄却する

Example: Suppose we are given a coin and we want to test whether the coin is 'fair' (公正なコイン) or is 'unfair' (不公正なコイン).

- The **hypothesis** that we propose: 仮説 =「コインは公正」
- **Collecting the evidence**: Toss the coin 50 times, record the number of 'head' (表)
- Assuming the coin is fair, we expect 25 'heads'.

# 'heads' (表) out of 50 tosses	Outcome expected/unexpected (予想内/予想外) assuming Hypothesis is True	Our Conclusion (結論)
[24]	**[Expected]**	**[Do not reject hyp.]**
[22]	**[Expected]**	**[Do not reject hyp.]**
[2]	**[Unxpected]**	**[Reject hyp.]**

For the cases of 24 or 22 heads, we would not be inclined to say the hypothesis is false because it seems quite ***expected*** that a fair coin tossed 50 times could yield 24 or 22 heads.

But, for the case of 2 heads, we would be inclined to say the hypothesis is false because it seems quite ***unexpected*** that a fair coin tossed 50 times would yield just 2 heads.

Our reasoning above is the paradigm of hypothesis testing.

これは仮説検定の基本的な考え方です。

Probability of 2 'heads' out of 50 tosses $\fallingdotseq \frac{1}{1{,}000{,}000{,}000{,}000}$ = 1兆分の 1

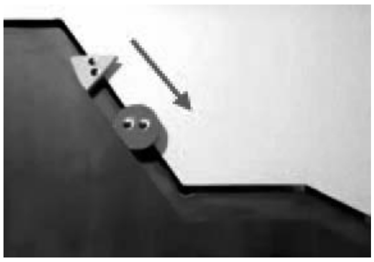
Helper versus Hinderer

Do children less than a year old recognize the difference between helper, friendly behavior as opposed to mean, unhelpful behavior? In a study reported in the November 2007 issue of Nature, researchers asked “Can pre-verbal infants evaluate/judge behavior?”[1].

- Participants: Sixteen 10-month old infants
- Each infant is shown a “climber” character who cannot make it up a hill in 2 tries

- Each infant is then shown two scenarios

Climber is pushed to top by “helper”



Climber is pushed to bottom by “hinderer”

- Each infant is alternately shown the two scenarios several times
- Each infant is then presented with both the helper and hinderer toy to play with

- Researchers record which toy the child selects[2]

RESULTS (実験結果) →

Helper Selected	Hinderer Selected	Total
14	2	16

[1]Hamlin, J., Wynn, K., and Bloom, P. (2007) “Social evaluation by preverbal infants” *Nature* 450, 557–9.
[2]See video link for experiment details: `https://www.youtube.com/watch?v=anCaGBsBOxM`

a. Determine the proportion (比率) of infants who selected the helper toy.

[14/16 = 87.5%]

b. Suggest possible explanations for this result that the researchers observed. Which would be the researchers' conjecture? Can we eliminate any?

[• **Babies have preference for the blue color**] ⟶ **Eliminate**

[• **Babies have preference for the square shape**] ⟶ **Eliminate**

[• **Babies have no preference for either toy**] ⟶ 帰無仮説

[• **Babies have preference for "helper" toy**] ⟶ 対立仮説

c. Suppose for the moment that the researchers' conjecture is wrong, and infants actually have no preference for either type of toy. How many of the 16 would you expect to select the helper toy?

[0.5(16) = 8]

d. Would it be *possible* to have obtained a result as extreme as the researchers found (14 out of 16 for "helper") assuming infants actually have no preference? Would you say this outcome is *likely*?

[It is possible, but not very likely]

e. In your judgment, how many infants, out of the 16, would have to select the helper toy in order for you to be fairly well convinced that the researchers' conjecture is correct, that infants really do have a tendency to prefer the helper toy? Explain.

[Answers vary, but the most common choice is often 12]

	# out of 16 selecting helper							
	9	10	11	12	13	14	15	16
Count			[7]	[22]	[9]	[2]		

f. Describe how we could use a common device to simulate the infants' selection process assuming they had no genuine preference.
何を使えば同じような実験をシミュレーション出来ますか？

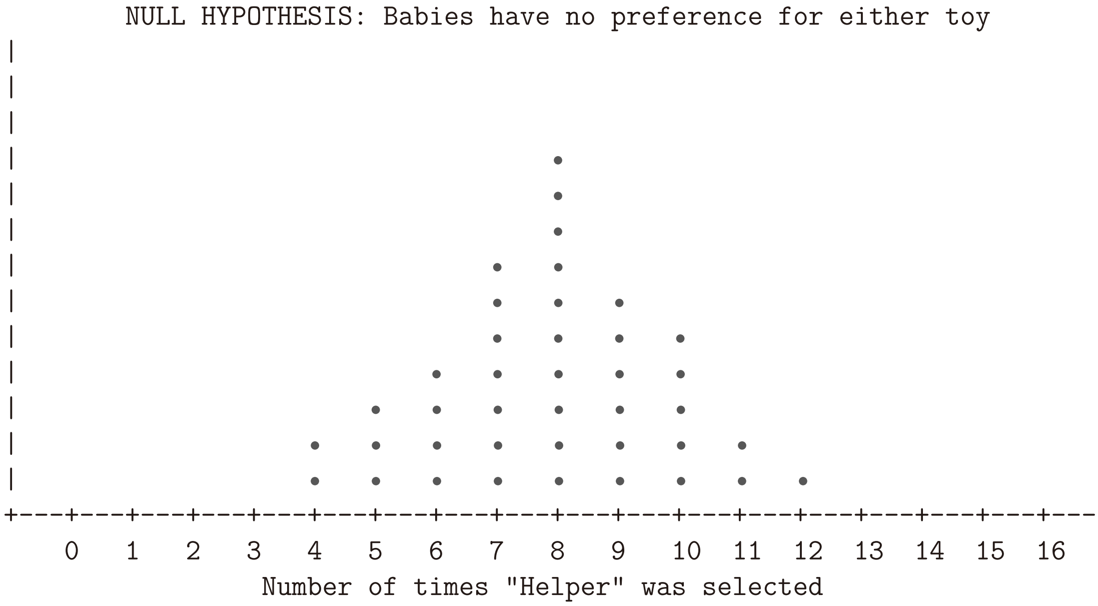
[Toss a coin]

The key question is to determine what results would occur in the long run under the assumption that infants actually have no preference. (We will call this assumption of no preference the **null hypothesis** – 帰無仮説.) We will answer this question by simulating (artificially re-creating) the selection process of 16 infants over and over, *assuming that infants actually have no toy preference.*

g. Toss a coin 16 times and record the outcomes. The number of heads you obtain represents the number of your 16 hypothetical infants who choose the helper toy.

Head = 表 (helper)	Tail = 裏 (hinderer)
[**9**]	[**7**]

h. Combining all our simulation results, produce a corresponding dotplot (ドットプロット).



i. Looking at this dotplot, does it seem that the result obtained by the researchers would have been surprising if in fact the infants had no preference (予想内/予想外)?

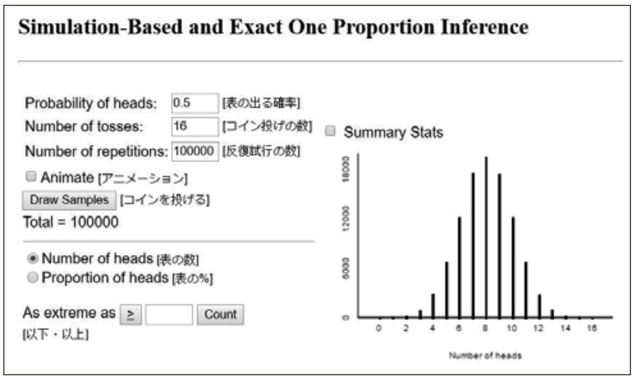
[**Researcher's result is unusual if infants had no preference.**]

Based on this, would you say there's sufficient evidence that the infants do genuinely prefer the helper toy? (「赤ちゃんに好みが無い」仮説：棄却する/棄却しない)

[**There appears to be strong evidence that infants prefer helper toy.**]

We really need to simulate this random selection process hundreds, preferably thousands of times. This would be very tedious and time-consuming with coins, so we'll turn to technology.

j. Use the "One Proportion" applet (www.rossmanchance.com/applets/OneProp/OnePropJPN.html) to simulate the random process of 16 infants making a toy choice, assuming infants have no preference for either toy. Below is a simulation based on 100,000 repetitions.



Where is this distribution centered? Is this what we expected? (予想通り?)

[Centered at 8, as expected.]

k. Report how many of the simulation repetitions produced 14 or more infants choosing the helper toy. Also find the proportion of these repetitions that yielded such an extreme result.

[214 → 214/100000 = 0.00214.]

l. Is this a large or small proportion? Based on this proportion, is this result surprising assuming infants have no preference (予想内/予想外)?

[A very small proportion → a surprising result.]

m. Is there sufficient evidence to suggest infants in general have a genuine preference for the helper toy over the hinderer toy? (「赤ちゃんに好みが無い」仮説：棄却する/棄却しない)

[Very strong evidence infants prefer helper toy.]

SUMMARY: Hypothesis testing process

n. If X = 'number of heads', what distribution is X based on? How would you determine the exact probability that X is greater than or equal to 14, that is $P(X \geq 14)$? [3]

X follows the Binomial distribution, $B(n, p)$ with $n = 16$, $p = .5$

$$\begin{aligned} P(X \geq 14) &= P(X = 14) + P(X = 15) + P(X = 16) \\ &= \frac{16!}{14!2!}0.5^{14}0.5^{2} + \frac{16!}{15!1!}0.5^{15}0.5^{1} + \frac{16!}{16!0!}0.5^{16}0.5^{0} \\ &= 0.00209 \end{aligned}$$

[3] In hypothesis testing, $P(X \geq 14)$ would be called the **p-value** (p-値).
A smaller p-value yields stronger evidence against the null hypothesis: 証拠が予想外 → 仮説を棄却する

Topic 2: Benford's Law

Example 1: Powers of 2

Suppose we look at the first 18 powers of 2^n:

$2^1 = \underline{2}$	$2^{10} = \underline{1}024$
$2^2 = \underline{4}$	$2^{11} = \underline{2}048$
$2^3 = \underline{8}$	$2^{12} = \underline{4}096$
$2^4 = \underline{1}6$	$2^{13} = \underline{8}192$
$2^5 = \underline{3}2$	$2^{14} = \underline{1}6384$
$2^6 = \underline{6}4$	$2^{15} = \underline{3}2768$
$2^7 = \underline{1}28$	$2^{16} = \underline{6}5536$
$2^8 = \underline{2}56$	$2^{17} = \underline{1}31072$
$2^9 = \underline{5}12$	$2^{18} = \underline{2}62144$

Circle the **leading digit** (値(あたい)の最初の桁(けた)) of each number listed above.

- Theoretically, what is the list of all possible leading digits? [**1,2,3,4,5,6,7,8,9**]

- A simple assumption is that the possible leading digit values are equally likely to occur.

  For example, the percent of times the leading digit is '7' should be [**1/9**]

For the 18 values listed above, write the expected counts of leading digits assuming equally likely occurrence. Also write the observed counts.

	Leading Digit								
	1	2	3	4	5	6	7	8	9
Expected (期待)	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]
Observed (実際)	[5]	[4]	[2]	[2]	[1]	[2]	[0]	[2]	[0]

Are the expected and observed counts similar? Explain.

[**Not all observed counts are as expected, higher observed counts at 1 and 2.**]

Benford's Law **J. Doi, Cal Poly SLO**

#	合計
1	[13]
2	[9]
3	[5]
4	[5]
5	[4]
6	[4]
7	[0]
8	[5]
9	[0]

2^1 to 2^{45}

2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
131072
262144
524288
1048576
2097152
4194304
8388608
16777216
33554432
67108864
134217728
268435456
536870912
1073741824
2147483648
4294967296
8589934592
17179869184
34359738368
68719476736
137438953472
274877906944
549755813888
1099511627776
2199023255552
4398046511104
8796093022208
17592186044416
35184372088832

2^1 to 2^{45}

2^{46} to 2^{90}

70368744177664
140737488355328
281474976710656
562949953421312
1125899906842624
2251799813685248
4503599627370496
9007199254740992
18014398509481984
36028797018963968
72057594037927936
144115188075855872
288230376151711744
576460752303423488
1152921504606846976
2305843009213693952
4611686018427387904
9223372036854775808
18446744073709551616
36893488147419103232
73786976294838206464
147573952589676412928
295147905179352825856
590295810358705651712
1180591620717411303424
2361183241434822606848
4722366482869645213696
9444732965739290427392
18889465931478580854784
37778931862957161709568
75557863725914323419136
151115727451828646838272
302231454903657293676544
604462909807314587353088
1208925819614629174706176
2417851639229258349412352
4835703278458516698824704
9671406556917033397649408
19342813113834066795298816
38685626227668133590597632
77371252455336267181195264
154742504910672534362390528
309485009821345068724781056
618970019642690137449562112
1237940039285380274899124224

2^{46} to 2^{90}

#	合計
1	[14]
2	[7]
3	[6]
4	[4]
5	[3]
6	[2]
7	[5]
8	[0]
9	[4]

For these values, write the expected counts of leading digits assuming equally likely occurrence, write the observed counts (partition the work among the students), then convert to percentages. (Suggestion: Tally 1's, then 2's, then 3's, ...)

	Leading Digit								
	1	2	3	4	5	6	7	8	9
Expected (期待)	[10]	[10]	[10]	[10]	[10]	[10]	[10]	[10]	[10]
Observed (実際)	[27]	[16]	[11]	[9]	[7]	[6]	[5]	[5]	[4]

	Leading Digit								
	1	2	3	4	5	6	7	8	9
Expected (期待)%	[11.1]	[11.1]	[11.1]	[11.1]	[11.1]	[11.1]	[11.1]	[11.1]	[11.1]
Observed (実際)%	[30.0]	[17.8]	[12.2]	[10.0]	[7.8]	[6.7]	[5.6]	[5.6]	[4.4]

See display of leading digits array and prevalence of 1's, 2's, 3's, ...

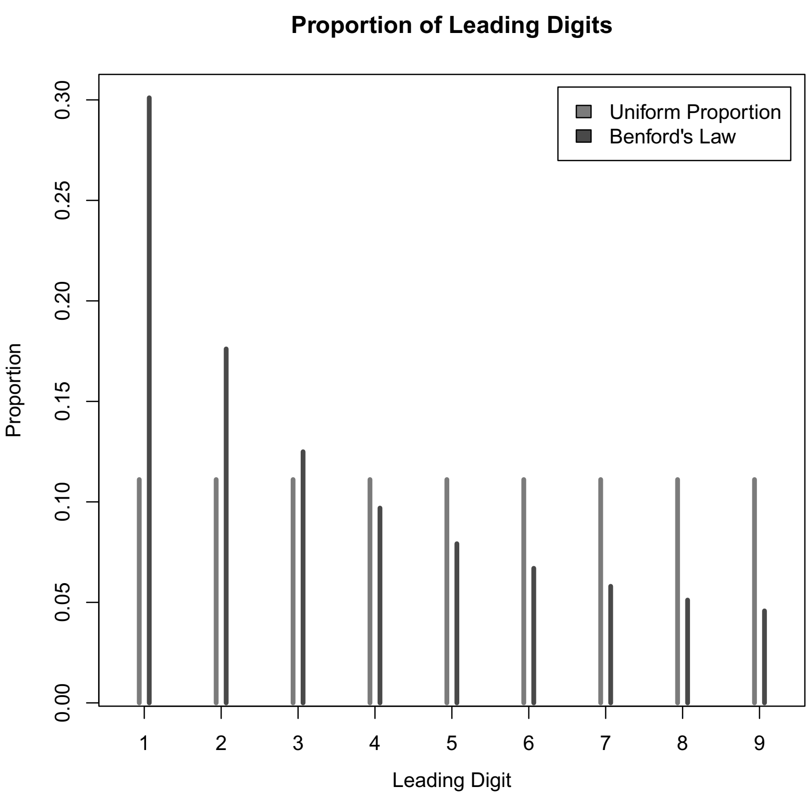
Benford's Law = ベンフォードの法則（ほうそく）

Benford's Law states that, in many situations, the incidence of the leading digit (d) of a set of numbers is well described the following probability function:

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\frac{d+1}{d}$$

d	$P(d)$		
1	$\log_{10}(2/1)$	≒	30.1%
2	$\log_{10}(3/2)$	≒	17.6%
3	$\log_{10}(4/3)$	≒	12.5%
4	$\log_{10}(5/4)$	≒	9.7%
5	$\log_{10}(6/5)$	≒	7.9%
6	$\log_{10}(7/6)$	≒	6.7%
7	$\log_{10}(8/7)$	≒	5.8%
8	$\log_{10}(9/8)$	≒	5.1%
9	$\log_{10}(10/9)$	≒	4.6%

Compare the percentages of the uniform (一様) percentages model to that of Benford's Law:



Benford's Law **J. Doi, Cal Poly SLO**

Previously we determined the observed percentages of leading digits for the first $n = 18$ and $n = 90$ powers of 2. The following table provides the observed percentages of leading digits for more powers of 2:

Leading Digit	**Uniform**	$n = 18$	$n = 90$	$n = 100$	$n = 200$	$n = 300$	**Benford's Law**
1	**11.1**	27.8	30.0	30.0	30.0	30.0	**30.1**
2	**11.1**	22.2	17.8	17.0	18.0	18.0	**17.6**
3	**11.1**	11.1	12.2	13.0	12.0	12.3	**12.5**
4	**11.1**	11.1	10.0	10.0	10.0	9.7	**9.7**
5	**11.1**	5.6	7.8	7.0	8.0	8.0	**7.9**
6	**11.1**	11.1	6.7	7.0	6.5	6.7	**6.7**
7	**11.1**	0.0	5.6	6.0	5.5	5.7	**5.8**
8	**11.1**	11.1	5.6	5.0	5.5	5.0	**5.1**
9	**11.1**	0.0	4.4	5.0	4.5	4.7	**4.6**

Percentage of leading digits for the first n powers of 2
compared to uniform percentages and Benford's Law.

As n increases notice that there is greater agreement of the observed percentages with what Benford's Law suggests.

As it turns out, the agreement of leading digit behavior with Benford's Law is not unique to the sequence of the powers of 2. The sequence of powers of other bases will yield similar behavior. This also applies to additive sequences and even to the sequence of prime numbers.

Cal Poly Shiny Applet for Benford's Law

- Visit `http://shiny.stat.calpoly.edu/BenfordSeq` to use the Shiny web application that applies Benford's Law to other sequences (additive, power, prime number).
- Visit `http://shiny.stat.calpoly.edu/BenfordData` to use the Shiny web application that applies Benford's Law to data examples (US Census, US Stock Exchange ...)

Example 2: Stock Market Data

Closing prices (in JPY) for stocks traded on the Nikkei 225 (日経平均株価) on 15JUN2017[4]:

Name	Closing	Leading Digit
Advantest Corp.	2,011.00	2
Aeon	1,716.00	1
Ajinomoto Co., Inc.	2,468.00	2
Alps Electric	2,928.00	2
Amada	1,267.00	1
ANA Holdings	387.40	3
Aozora Bank	426.50	4
Asahi Glass	935.00	9
Asahi Group Holdings	4,387.00	4
Asahi Kasei Corp.	1,115.00	1
Astellas Pharma Inc.	1,367.20	1
Bridgestone Corp.	4,714.00	4
Canon	3,895.00	3
Casio Computer	1,673.00	1
Central Japan Railway Co.	18,330.00	1
Chiyoda Corp.	644.00	6
Chubu Electric Power Co., Inc.	1,511.50	1
Chugai Pharmaceutical	4,125.00	4
Citizen Holdings	751.00	7
Comsys Holdings Corp.	2,349.00	2
Concordia Financial Group	540.70	5
Credit Saison	2,220.00	2
Dai Nippon Printing	1,236.00	1
Daiichi Sankyo	2,436.50	2
Daikin Industries	10,975.00	1
Dainippon Screen Mfg.	7,969.90	7
Daiwa House Industry	3,809.00	3
Daiwa Securities Group Inc.	653.50	6
DeNA Co	2,437.00	2
Denki Kagaku Kogyo K.K.	550.00	5
Denso Corp.	4,658.00	4
Dentsu Inc.	5,460.00	5
DOWA Holdings	809.00	8
East Japan Railway Co.	11,120.00	1
Ebara Corp.	3,010.00	3
Eisai	6,088.00	6
Familymart Ltd	6,620.00	6
Fanuc Corp.	21,385.00	2
Fast Retailing	36,700.00	3
Fuji Electric	586.00	5
Fuji Heavy Industries	3,694.00	3
Fujifilm Holdings Corp.	3,998.00	3
Fujikura	962.00	9
Fujitsu	788.40	7
Fukuoka Financial Group, Inc.	536.00	5
Furukawa	188.00	1
Furukawa Electric	5,020.00	5
GS Yuasa Corp.	500.00	5
Haseko Corp	1,493.50	1
Hino Motors	1,228.00	1
Hitachi	660.60	6
Hitachi Construction Machinery Co	2,613.00	2
Hitachi Zosen Corp.	539.00	5
Hokuetsu Kishu Paper	900.00	9
Honda Motor	3,057.00	3
IHI Corp.	381.50	3
Inpex Corp.	1,042.00	1
Isetan Mitsukoshi Holdings	1,157.00	1
Isuzu Motors	1,340.50	1
Itochu Corp.	1,583.50	1
J.Front Retailing	1,667.00	1
Japan Tobacco	4,096.00	4
JFE Holdings, Inc.	1,780.50	1
JGC Corp.	1,735.00	1
JTEKT Corp.	1,591.00	1
JX Holdings, Inc.	485.80	4
Kajima Corp.	922.00	9
Kao Corp.	6,887.00	6
Kawasaki Heavy Industries	317.50	3
Kawasaki Kisen Kaisha	278.50	2
KDDI Corp.	3,005.00	3
Keio Corp.	950.00	9
Keisei Electric Railway	2,980.00	2
Kikkoman Corp.	3,670.00	3
Kirin Holdings	2,457.50	2
Kobe Steel	1,025.00	1
Komatsu	2,684.00	2
Konami Corp.	6,360.00	6
Konica Minolta, Inc.	882.00	8
Kubota Corp.	1,907.50	1
Kuraray	2,021.00	2
Kyocera Corp.	6,407.00	6
Kyowa Hakko Kirin	2,020.00	2
Marubeni Corp.	690.70	6
Maruha Nichiro Corp	3,155.00	3
Marui Group	1,718.00	1
Matsui Securities	914.00	9
Mazda Motor	1,500.50	1
Meidensha Corp.	367.00	3
Meiji Holdings	9,270.00	9
Minebea Mitsumi	1,806.00	1
Mitsubishi Chemical Holdings Corp	852.60	8
Mitsubishi Corp.	2,254.00	2
Mitsubishi Electric Corp.	1,566.20	1
Mitsubishi Estate	2,145.00	2
Mitsubishi Heavy Industries	428.30	4
Mitsubishi Logistics Corp.	1,478.00	1
Mitsubishi Materials Corp.	3,092.40	3
Mitsubishi Motors Corp.	716.50	7
Mitsubishi UFJ Financial	715.00	7
Mitsui	1,522.50	1
Mitsui Chemicals, Inc.	563.50	5
Mitsui Engineering & Shipbuilding	153.00	1
Mitsui Fudosan	2,689.50	2
Mitsui Mining and Smelting Co.	410.50	4
Mitsui O.S.K. Lines	319.50	3
Mizuho Financial	196.90	1
MS&AD Insurance Group Holdings	3,908.00	3
NEC Corp.	286.50	2
NGK Insulators	2,151.00	2
Nichirei Corp.	3,380.00	3
Nikon Corp.	1,762.00	1

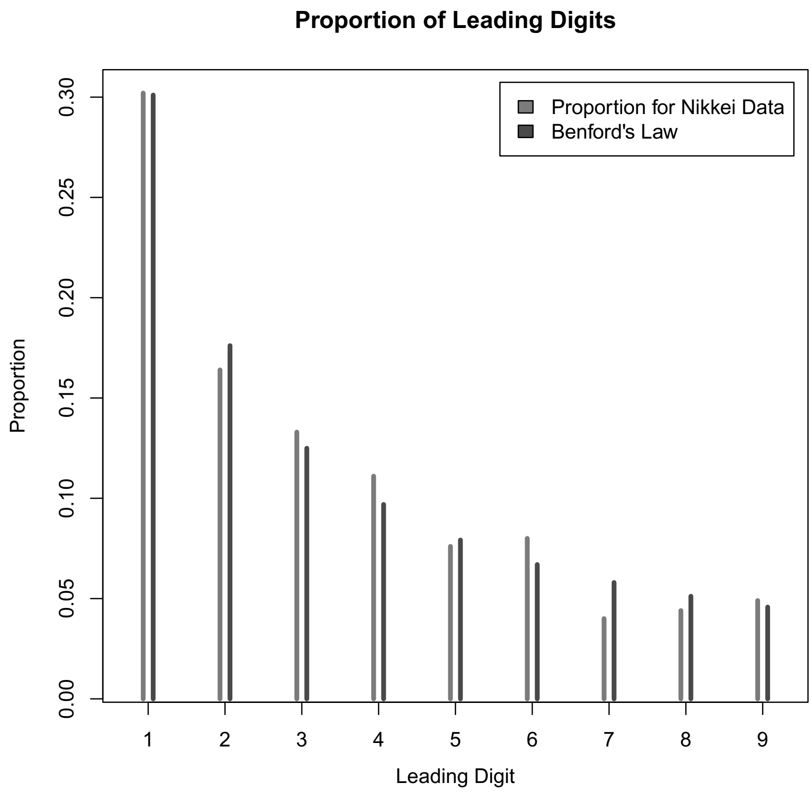
Name	Closing	Leading Digit
Nippon Electric Glass	752.00	7
Nippon Express	695.00	6
Nippon Kayaku	1,586.00	1
Nippon Light Metal Holdings Co.	261.00	2
Nippon Meat Packers, Inc.	3,635.00	3
Nippon Paper Industries	2,247.00	2
Nippon Sheet Glass	843.00	8
Nippon Steel&Sumitomo Metal Corp.	2,341.20	2
Nippon Suisan Kaisha	691.00	6
Nippon Telegraph & Telephone Corp	5,303.00	5
Nippon Yusen K.K	203.50	2
Nissan Chemical Industries	3,735.00	3
Nissan Motor	1,079.20	1
Nisshin Seifun Group Inc.	1,916.00	1
Nisshin Steel Holdings	1,196.00	1
Nisshinbo Holdings Inc.	1,079.00	1
Nitto Denko Co	8,883.00	8
NKSJ Holdings, Inc.	4,315.00	4
Nomura	664.20	6
NSK	1,345.00	1
NTN Corp.	490.00	4
NTT Data Corp.	6,090.00	6
NTT Docomo, Inc.	2,707.00	2
Obayashi Corp.	1,314.00	1
Odakyu Electric Railway	2,300.00	2
Oji Holdings Corp.	571.00	5
Oki Electric Industry	1,589.00	1
Okuma Corp.	1,002.00	1
Olympus Corp.	3,930.00	3
Osaka Gas	457.10	4
Otsuka Holdings Ltd	4,844.00	4
Pacific Metals	281.00	2
Panasonic	1,452.70	1
Pioneer Corp.	200.00	2
Rakuten Inc	1,356.50	1
Resona Holdings, Inc.	587.10	5
Ricoh	956.00	9
Sapporo Holdings	3,290.00	3
Secom	8,548.00	8
Sekisui House	1,977.00	1
Seven & i Holdings	4,786.00	4
Shimizu Corp.	1,172.00	1
Shin-Etsu Chemical	9,956.00	9
Shinsei Bank	184.50	1
Shionogi	6,067.00	6
Shiseido	3,868.00	3
Showa Denko K.K.	2,443.00	2
Showa Shell Sekiyu K.K.	1,061.00	1
SKY Perfect JSAT Holdings Inc.	489.00	4
Softbank Corp.	8,861.50	8
Sojitz Corp.	268.50	2
Sony	4,138.00	4
Sony Financial Holdings Inc.	1,808.00	1
SUMCO Corp.	1,711.00	1
Sumitomo Chemical	586.50	5
Sumitomo Corp.	1,417.00	1
Sumitomo Dainippon Pharma	1,694.00	1
Sumitomo Electric Industries	1,725.00	1
Sumitomo Heavy Industries	710.00	7
Sumitomo Metal Mining	1,330.00	1
Sumitomo Mitsui Financial	4,195.00	4
Sumitomo Mitsui Trust Holdings	3,823.50	3
Sumitomo Osaka Cement	488.00	4
Sumitomo Realty & Development Co.	3,493.00	3
Suzuki Motor Corp.	5,217.00	5
T&D Holdings, Inc.	1,663.50	1
Taiheiyo Cement Corp.	361.50	3
Taisei Corp.	1,025.50	1
Taiyo Yuden	1,643.00	1
Takara Holdings Inc.	1,213.00	1
Takashimaya	1,147.00	1
Takeda Pharmaceutical	5,609.00	5
TDK Corp.	7,270.00	7
Teijin	2,090.00	2
Terumo Corp.	4,405.00	4
The Chiba Bank	791.00	7
The Dai-ichi Life Insurance Co.	1,941.80	1
The Japan Steel Works	1,697.00	1
The Kansai Electric Power Co.	1,600.00	1
The Shizuoka Bank	952.00	9
The Yokohama Rubber	2,154.00	2
Tobu Railway	612.00	6
Toho	3,390.00	3
Toho Zinc	424.00	4
Tokai Carbon	628.00	6
Tokio Marine Holdings, Inc.	4,851.00	4
Tokuyama Corp.	499.00	4
Tokyo Dome Corp.	1,034.00	1
Tokyo Electric Power Co., Inc.	456.00	4
Tokyo Electron	15,935.00	1
Tokyo Gas	608.00	6
Tokyo Tatemono	1,545.00	1
Tokyu Corp.	836.00	8
Tokyu Fudosan	688.00	6
Toppan Printing	1,239.00	1
Toray Industries, Inc.	912.10	9
Toshiba Corp.	317.50	3
Tosoh Corp.	1,059.00	1
TOTO	4,260.00	4
Toyo Seikan Group Holdings	1,849.00	1
Toyobo	203.00	2
Toyota	5,794.00	5
Toyota Tsusho Corp.	3,340.00	3
Trend Micro Inc.	5,670.00	5
Ube Industries	272.00	2
Unitika	82.00	8
West Japan Railway Co.	8,023.00	8
Yahoo Japan	475.50	4
Yamaha Corp.	3,550.00	3
Yamaha Motor Co Ltd	2,823.00	2
Yamato Holdings	2,278.00	2
Yaskawa Electric Corp.	2,362.00	2
Yokogawa Electric Corp.	1,849.00	1

[4] Obtained from `https://www.investing.com/indices/japan-ni225-components`

Benford's Law **J. Doi, Cal Poly SLO**

Here are the percentages by Benford's Law and of the observed leading digits for the closing price of stocks traded on the Nikkei 225 stock exchange for 15JUN2017:

	Leading Digit for Nikkei Data								
	1	2	3	4	5	6	7	8	9
Observed (実際) %	30.2	16.4	13.3	11.1	7.6	8.0	4.0	4.4	4.9
ベンフォードの法則%	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6



Here is a comparison of observed leading digits for closing prices of stocks sold on 15JUN2017 across US stock markets[5]. Note the close agreement to Benford's Law percentages.

Leading Digit	Nikkei 225 Stock Exch.	New York Stock Exch.	NASDAQ Stock Mkt	NASDAQ Capital Mkt	**Benford's Law**
1	30.2	29.1	30.5	35.4	**30.1**
2	16.4	17.7	17.5	17.9	**17.6**
3	13.3	12.1	15.4	12.6	**12.5**
4	11.1	9.8	10.3	8.0	**9.7**
5	7.6	8.2	6.4	6.7	**7.9**
6	8.0	7.4	6.4	5.5	**6.7**
7	4.0	6.1	5.2	5.0	**5.8**
8	4.4	5.9	3.8	3.6	**5.1**
9	4.9	3.7	4.5	5.3	**4.6**

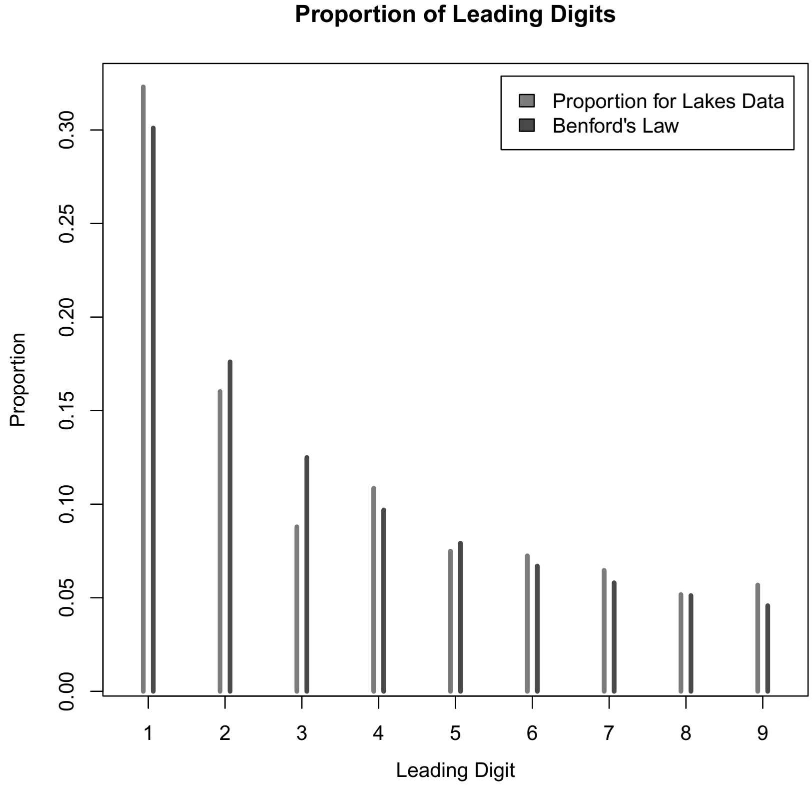
[5]Obtained from `http://online.wsj.com/public/resources/documents/stocksdaily.htm`

Example 3: Surface Area of Lakes in California

The following information is the surface area (表面積 (ひょうめんせき)) of 387 lakes in California. The surface area is provided in acres (1 Acre ≒ 4046.9 meter^2 → 平方メートル)[6]:

#	Lake Name	Acres	Leading Digit
1	Abbotts Lagoon	200	2
2	Almaden Reservoir	62	6
3	Lake Almanor	28257	2
⋮	⋮	⋮	⋮
385	Woodward Reservoir	2427	2
386	Wrights Lake	45	4
387	Lake Yosemite	500	5

	Leading Digit for Lakes Data								
	1	2	3	4	5	6	7	8	9
Observed (実際) %	32.3	16.0	8.8	10.9	7.5	7.2	6.5	5.2	5.7
ベンフォードの法則%	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6



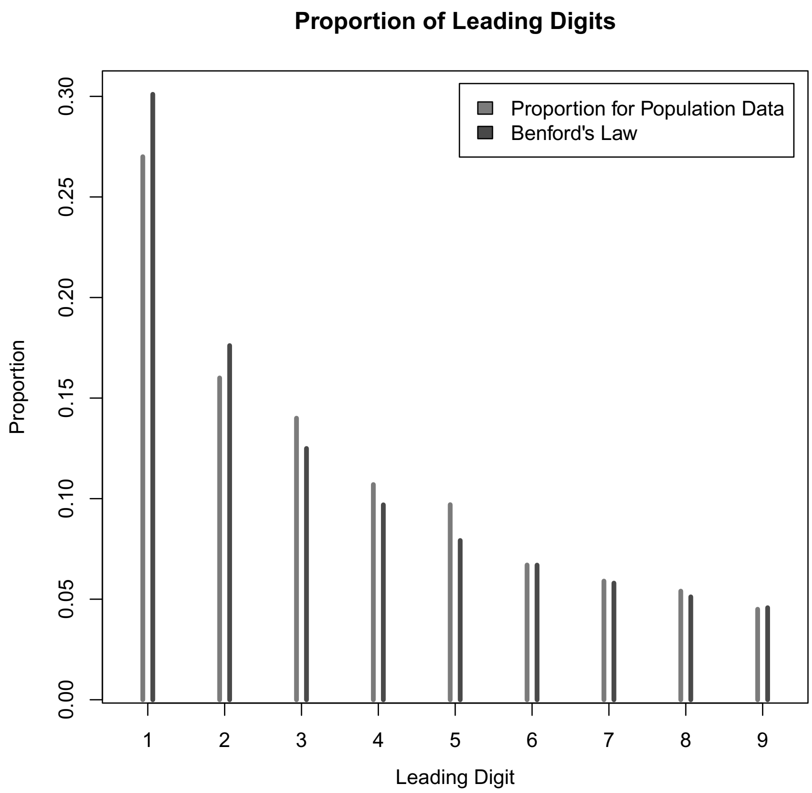
[6]Obtained from `https://en.wikipedia.org/wiki/List_of_lakes_in_California`

Example 4: Population from 2015 Japanese Census

The following Japanese cities population data is based on the 2015 Japanese Census[7]:

#	City Name	Pop'n Size	Leading Digit
1	Hakodate (函館)	265,979	2
2	Otaru (小樽)	121,924	1
3	Asahikawa (旭川)	339,605	3
⋮	⋮	⋮	⋮
966	Naha (那覇)	319,435	3
967	Ishigaki (石垣)	47,564	4
968	Urasoe (浦添)	114,232	1

	Leading Digit for Cities Data								
	1	2	3	4	5	6	7	8	9
Observed (実際) %	27.0	16.0	14.0	10.7	9.7	6.7	5.9	5.4	4.5
ベンフォードの法則%	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6



[7]Obtained via Statistics Bureau Japan (総務省統計局):
`http://www.stat.go.jp/english/data/nenkan/67nenkan/1431-02.html`,
`http://www.stat.go.jp/data/nenkan/index1.html`

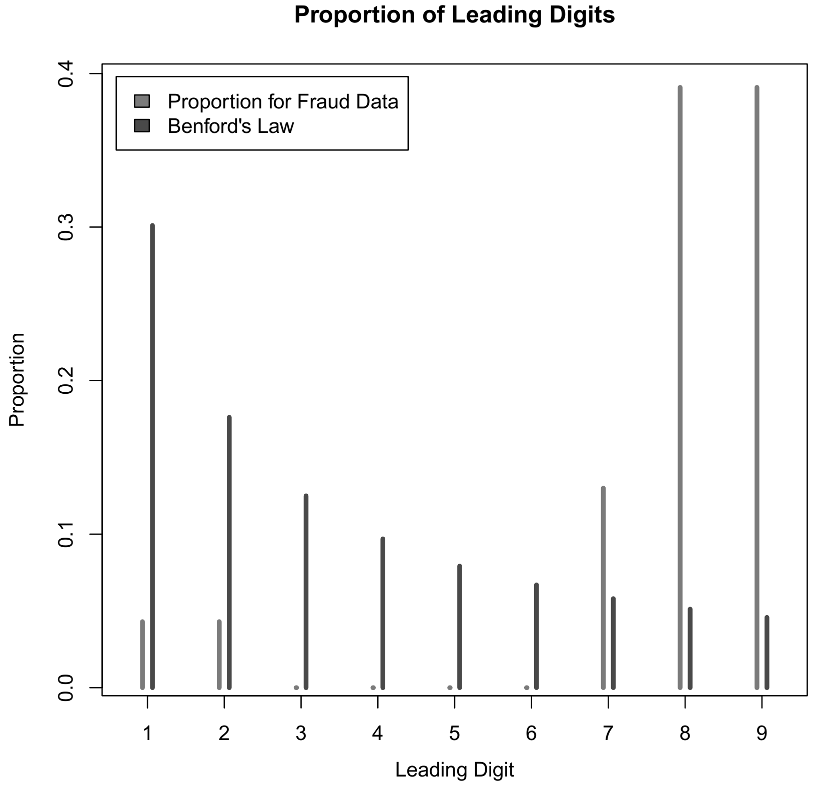
Example 5: Fraud Detection – 不正検出

In 1993, in State of Arizona vs. Wayne James Nelson, the accused was found guilty of trying to defraud the state of nearly $2 million[8]. Nelson was a manager in the office of the Arizona State Treasurer (会計官) and he was accused of writing 23 fraudulent checks.

彼はわざと２３個の嘘の会計を作ってお金を盗もうとしました。

Amounts (in USD) for the 23 Checks			
1,927.48	93,249.11	96,879.27	94,639.49
27,902.31	89,658.17	91,806.47	83,709.28
86,241.90	87,776.89	84,991.67	96,412.21
72,117.46	92,105.83	90,831.83	88,432.86
81,321.75	79,949.16	93,766.67	71,552.16
97,473.96	87,602.93	88,338.72	

	Leading Digit for Fraud Data								
	1	2	3	4	5	6	7	8	9
Observed (実際) %	4.3	4.3	0.0	0.0	0.0	0.0	13.0	39.1	39.1
ベンフォードの法則%	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6



[8]Reference: Mark J. Nigrini (May 1999). "I've Got Your Number". *Journal of Accountancy*. `http://www.journalofaccountancy.com/issues/1999/may/nigrini.html`

Longest Run of Heads/Tails **J. Doi, Cal Poly SLO**

Topic 3: Longest Run of Heads/Tails

Activity: You will be assigned to one of two groups to ...

- (A): Work with a partner and record the results of 30 coin tosses. If your group has 2 people, split the work in half (15 tosses per person). Write down your outcomes in the box below.
- (B): Consult with a partner and create in your minds what you think would be reasonable for a sequence of 30 tosses from a fair coin. Write down your outcomes in the box below.

OUTCOMES: H = Head = 表, T = Tail = 裏

[Students will write their results here]

Example from Group A: Outcomes from actual coin tosses

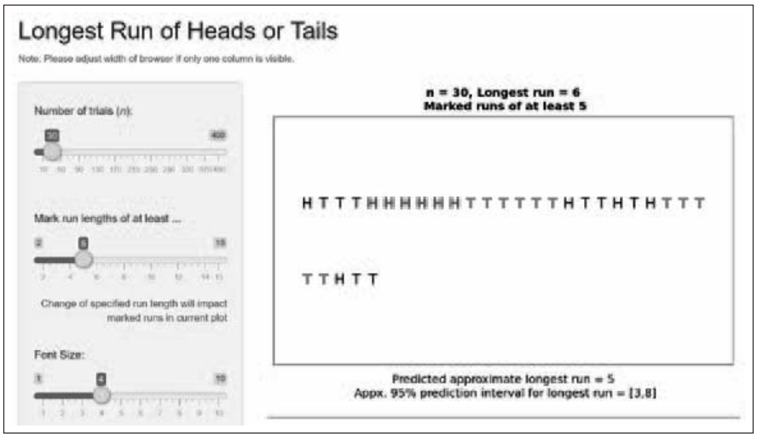
1	2	3	4	5	6	7	8	9	10
H	**H**	**T**	**T**	**T**	**H**	**H**	**H**	**H**	**H**
11	12	13	14	15	16	17	18	19	20
T	**H**	**H**	**T**	**H**	**T**	**H**	**T**	**T**	**H**
21	22	23	24	25	26	27	28	29	30
T	**H**	**H**	**H**	**H**	**T**	**T**	**T**	**T**	**T**

Example from Group B: Outcomes from imagined coin tosses

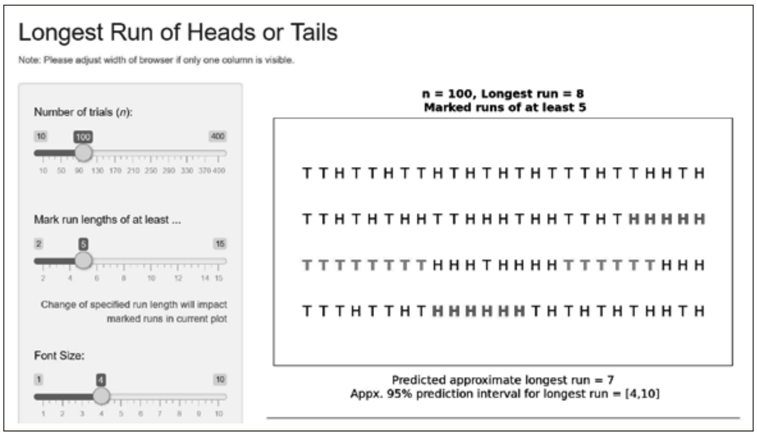
1	2	3	4	5	6	7	8	9	10
T	**H**	**H**	**T**	**T**	**H**	**T**	**H**	**H**	**T**
11	12	13	14	15	16	17	18	19	20
T	**H**	**T**	**H**	**H**	**T**	**T**	**T**	**H**	**T**
21	22	23	24	25	26	27	28	29	30
H	**T**	**H**	**T**	**H**	**H**	**H**	**T**	**H**	**H**

Do you notice any differences between the groups?
What does this tell us about randomness?

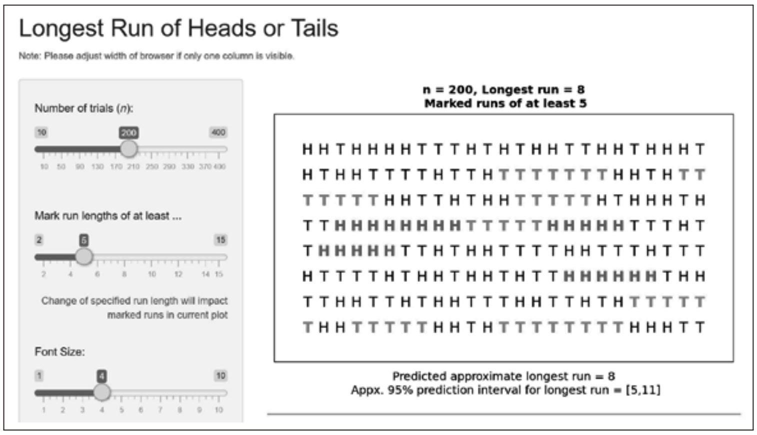
$\boxed{n = 30}$ $\rightarrow$



$\boxed{n = 100}$ $\rightarrow$



$\boxed{n = 200}$ $\rightarrow$

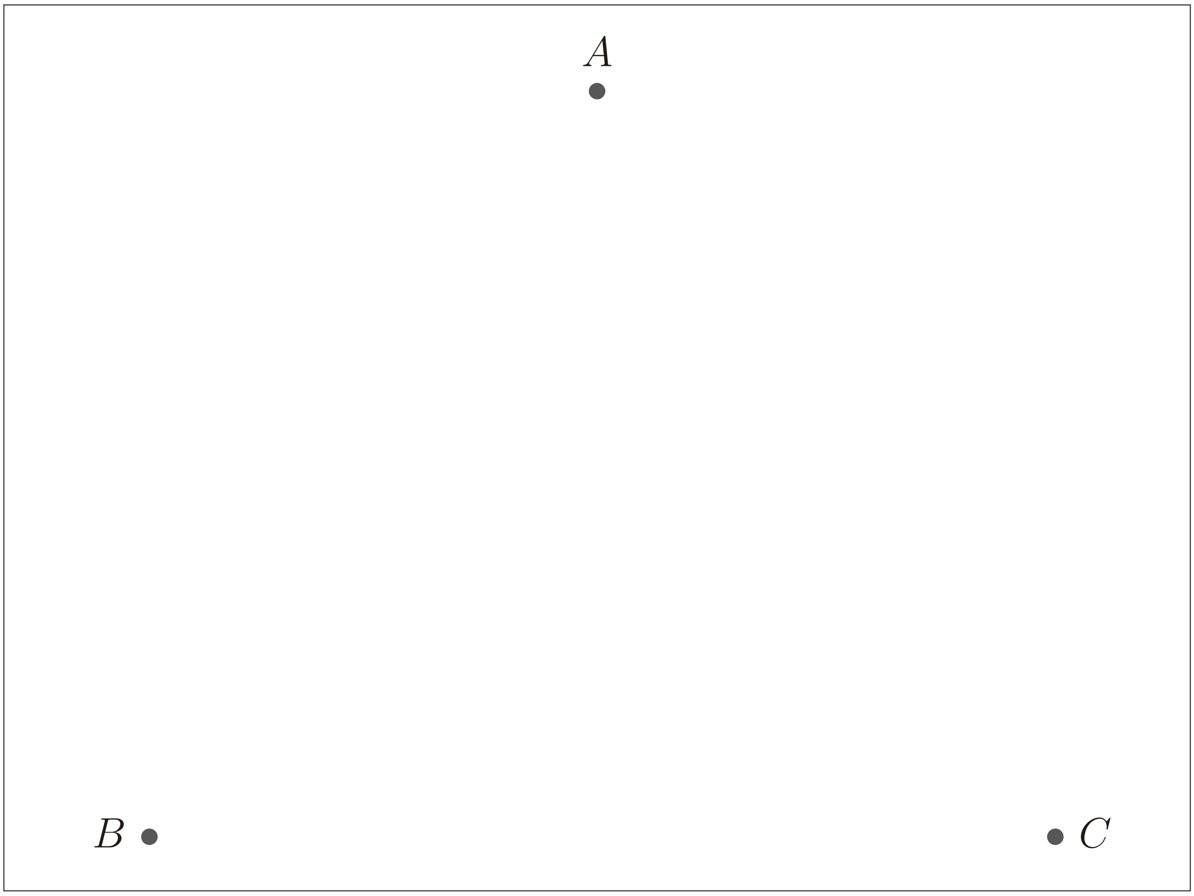


Cal Poly Shiny Applet for Longest Run

Visit `http://shiny.stat.calpoly.edu/LongRun` to use the Shiny web application that simulates random coin flips and the behavior of runs.

Topic 4: Chaos Game

Below, the vertices of a triangle (三角形の頂点) are shown:



Perform the following steps:

a. Mark a point at random in the triangle

b. Pick a vertex at random

c. Mark the halfway point between your point and the selected vertex

d. Pick a vertex at random

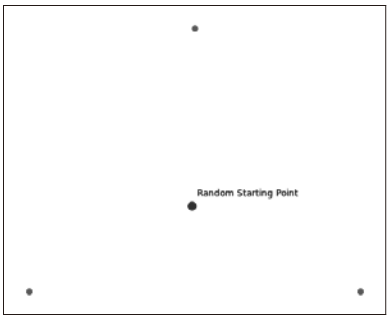
e. Mark the halfway point between your most recent point and the selected vertex

f. Repeat this process 10 times

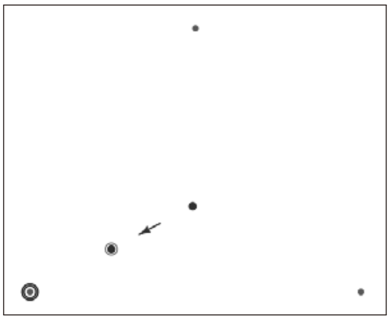
Is a pattern starting to reveal itself? What would happen if we repeated this process many many times? What pattern (if any) would develop?

The graphs below are the results for the number of steps $n = 1, 2, 50, 100, 250, 500, 1000, 2000, 5000$.

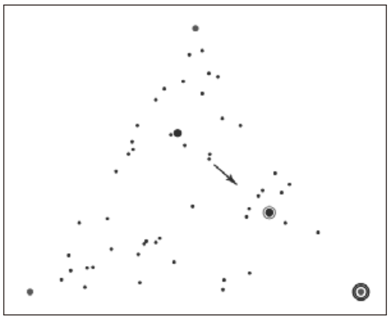
$n = 1$



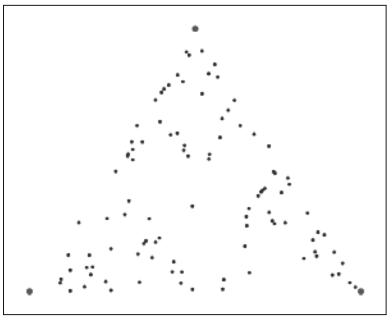
$n = 2$



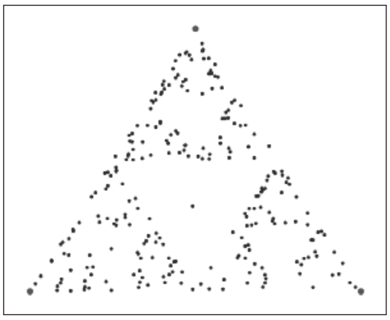
$n = 50$



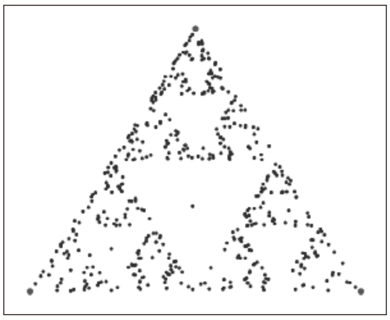
$n = 100$



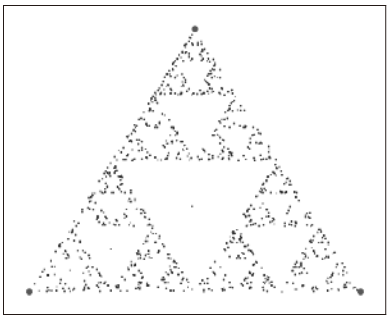
$n = 250$



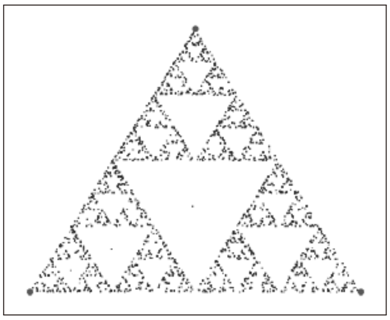
$n = 500$



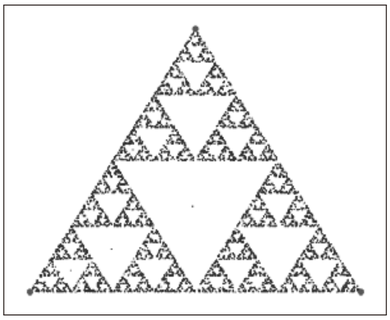
$n = 1000$



$n = 2000$



$n = 5000$



Chaos Theory, Fractal: Sierpinski's Gasket
カオス理論, フラクタル: シェルピンスキーのギャスケット

Cal Poly Shiny Applet for Chaos Game

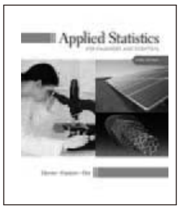
- Visit `http://shiny.stat.calpoly.edu/ChaosGame2D` to use the Shiny web application that simulates the Chaos Game for various shapes (e.g., triangle, square, pentagon).
- Also see `http://shiny.stat.calpoly.edu/ChaosGame3D` for 3-dimensional version.

Topic 5: Statistics Related Work

a. **Textbook**

(2013) Devore, J., Farnum, N., and Doi, J.

- **Applied Statistics for Engineers and Scientists** (3rd ed.) Cengage



b. **Categorical Data Analysis**

- (2014) Schilling, M., and Doi, J. "A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure". The American Statistician, 68, 133–145.
- (2018, in progress) Schilling, M., and Doi, J. "Conf. Intervals for Negative Binomial"

c. **Statistics Education**

- (2016) Doi, J., Potter, G., Wong, J., Alcaraz, A., and Chi, P. "Web Application Teaching Tools for Statistics Using R and Shiny". Tech. Innovations in Stat. Educ., 9(1)
- (2017) Doi, J., Statistics Lecture Translations and Recordings, Center for Statistics and Information, Rikkyo University

Summary

- Baby Helper/Hinderer Experiment (Binomial Distribution pattern)
- Benford's Law (leading digit patterns)
- Longest Run of Heads/Tails (long run patterns)
- Chaos Game (point location patterns)

Common Theme: In randomness there are patterns (ランダムの中に生まれる法則).

A major goal of statistics: To understand the patterns induced by randomness so that we may better answer research questions.

ゴール：ランダムの中に生まれる法則を良く理解する事で研究の問題を解決しようとする事です。

Super Science High School Students = 「将来、科学者や技術者、医者になり、日本の科学技術をリードすることを目標に勉強している生徒です。」

Regardless of what you will study, statistics will play an **important part** in your education and your future profession.

Resources

- **Current Presentation Materials**
  - ▷ Lecture notes (student version, not filled in) → https://tinyurl.com/2018-Doi-notes1
  - ▷ Lecture notes (instructor version, filled in) → https://tinyurl.com/2018-Doi-notes2
  - ▷ Presentation slides → https://tinyurl.com/2018-Doi-JCOTS

- **Previous Presentation Materials on Cal Poly Statistics Deptartment**
  - ▷ (2017) Presentation slides for "Statistical Computing Curriculum and Pedagogy at Cal Poly San Luis Obispo" → https://tinyurl.com/2017-Doi-computing
  - ▷ (2016) Presentation slides for "Department of Statistics at Cal Poly San Luis Obispo: Overview and Recent Changes" → https://tinyurl.com/2016-Doi-StatDept

- **References for Shiny**
  - ▷ Shiny App Teaching Tools Collection Website → https://statistics.calpoly.edu/shiny
  - ▷ (2016) Doi, J., Potter, G., Wong, J., Alcaraz, A., and Chi, P. "Web Application Teaching Tools for Statistics Using R and Shiny". Tech. Innovations in Stat. Educ., 9(1) → https://tinyurl.com/2016-Doi-ShinyPaper
  - ▷ (2016) Presentation slides for "Web Application Tools for Statistics Using R and Shiny" → https://tinyurl.com/2016-Doi-ShinyTalk

---

Jimmy Doi
jdoi@calpoly.edu
https://statistics.calpoly.edu
http://statweb.calpoly.edu/jdoi