

Probability depends on status of information

- ⇒ Absolute probability makes no sense
(even in coin tossing!)
- ⇒ only conditional probability

$$P(E|I)$$

If I changes then also P changes:

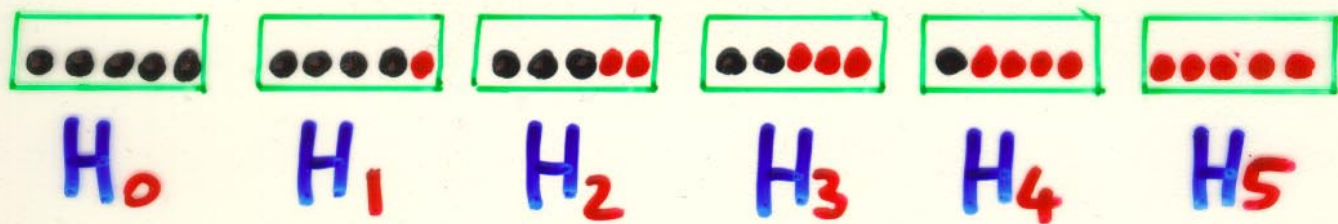
- In many cases: "irrationally"

↳ ≡ "not easy to model"

- In most of scientific problems:
updated through Bayes' Theorem

$$P(E|I) = \frac{P(E \cap I)}{P(I)}$$

Six boxes problem



- Symmetric status of preparation
- Choose one box:

→ Which one is it? H_0, H_1, \dots, H_5 ?

→ If I extract one ball, what will come out?

→ Red: E_1 ?

→ Black: E_2 ?

- I am uncertain on H_i , but $\boxed{U_{i=0}^5 H_i = \Omega}$
- I am uncertain on E_i , but $\boxed{E_1 \cup E_2 = \Omega}$ ^{certainty}
- In general, I am uncertain about the "constituents" ("which box and which outcome?")

$$E_1 \wedge H_0, E_1 \wedge H_1, E_1 \wedge H_2, \dots, E_1 \wedge H_5$$

$$E_2 \wedge H_0, E_2 \wedge H_1, E_2 \wedge \bar{H}_2, \dots, E_2 \wedge H_5$$

Since both E_i and H_j form complete classes of hypotheses

$$\Rightarrow E_i = \bigcup_j (E_i \cap H_j) \quad \text{"}E_i \text{ whatever } H_j \text{ might be"}$$

$$\Rightarrow H_j = \bigcup_i (E_i \cap H_j)$$

$$\Rightarrow \begin{cases} P(E_i) = \sum_j P(E_i \cap H_j) = \sum_j P(E_i | H_j) \cdot P(H_j) \\ P(H_j) = \sum_i P(E_i \cap H_j) = \sum_i P(H_j | E_i) \cdot P(E_i) \end{cases}$$

Easiest thing to evaluate?

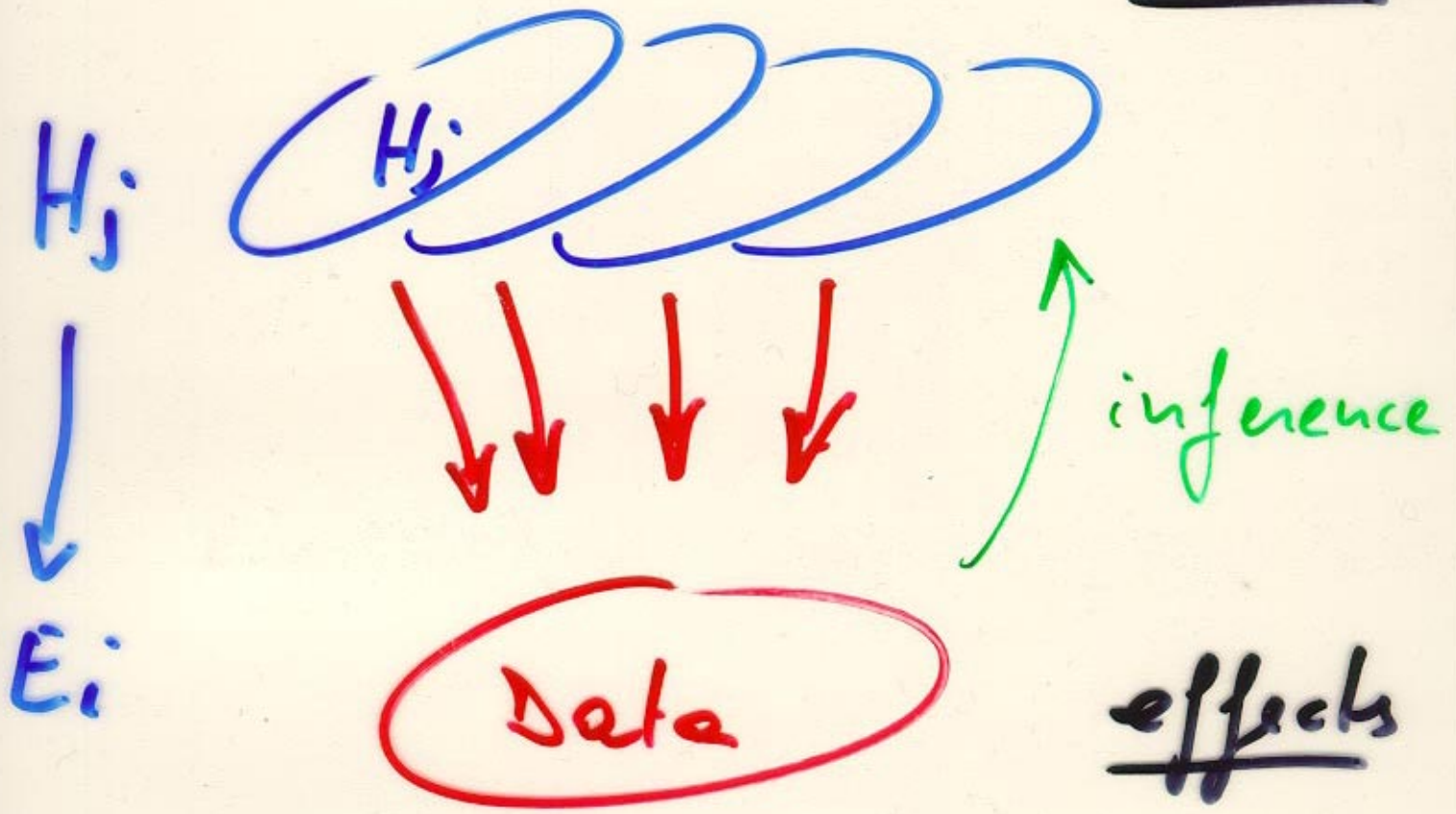
$$\rightarrow P(E_i | H_j)$$

E_i plays the role of observable: effects

H_j plays the role of physical hypotheses ("true values") \rightarrow cause of the effects

\Rightarrow we see effects \rightarrow guess causes
("infer")

causes



Inference : $P(H_j | \text{data})$

Poincaré:

Now, these problems are classified as probability of causes, and are the most interesting of all from their scientific applications. I play at *écarté* with a gentleman whom I know to be perfectly honest. What is the chance that he turns up the king? It is $\frac{1}{8}$. This is a problem of the probability of effects. I play with a gentleman whom I do not know. He has dealt ten times, and he has turned the king up six times. What is the chance that he is a sharper? This is a problem in the probability of causes. It may be said that it is the essential problem of the experimental method.

130) 18)

-216-
-91-

116(35)

→ after an effect has been observed,
we tend to believe to some
causes more than to others

→ repeating the experiment
the beliefs about the different
causes are constantly
updated

⇒ how?

$$\rightarrow P(H_j | E^{(1)})$$

$$P(H_j | E^{(1)} \wedge E^{(2)})$$

⋮

$$P(H_j | \underbrace{E^{(1)} \wedge E^{(2)} \wedge \dots \wedge E^{(n)}})$$

entire set of observations

$$P(E_i | H_j)$$

Red

Black

$$P(\underline{E}_1 | H_0) = 0$$

$$P(E_2 | H_0) = 1$$

$$P(\underline{E}_1 | H_1) = 1/5$$

$$P(E_2 | H_1) = 4/5$$

...

$$P(\underline{E}_1 | H_j) = j/5$$

$$P(E_2 | H_j) = \frac{5-i}{5}$$

$$P(\underline{E}_1 | H_5) = 1$$

$$P(E_2 | H_5) = 0$$

Before first observation : what shall we observe?

$$\rightarrow P(E_1) ?$$

$\frac{1}{2}$ by symmetry

\rightarrow let's evaluate it in a general way
valid also when the problem might (will!)
become asymmetric

$$P(E_1) = \sum_j P(E_1 \cap H_j) = \sum_j P(H_j) \cdot P(E_1 | H_j)$$

$\underbrace{P_0(H_j)}_{\rightarrow 1/6}$ because initial symmetry

$$P(E_1 | \text{initial symmetry}) = \frac{1}{6} \times \left(\frac{0+1+2+3+4+5}{5} \right)$$

Let us now make the experiment

Outcome \rightarrow Black \leftarrow
(E_2)

What can we say now about H_j
in the light of the observation $E^{(1)} = E_2$?

(We don't really need to wait: the conclusions
can be evaluate in advance under the hypothesis
that $E^{(1)}$ will be observed)

$$P(H_j | E_2) = \frac{P(E_2 | H_j)}{P(E_2)} \quad P(E_2) \neq 0$$

“after” the observation

“before” the observation

- “under the condition E_2 ”
- “under the hypothesis
that E_2 is true”

Rewriting the r.h.s. of the previous formula, making appearing the "easy things":

$$\sum_j P(H_j | E_i) = \frac{P(E_i \wedge H_j)}{P(E_i)} = \frac{P(E_i | H_j) \cdot P(H_j)}{\sum_j P(E_i | H_j) \cdot P(H_j)}$$

Bayes' Theorem

Written in this way, we don't need any more the denominator (just a normalisation factor)

$$P(H_j | E_i) \propto P(E_i | H_j) \cdot P_0(H_j)$$

↑
"after"
"posterior"
"final"

↑
"likelihood"

↑
"before"
"prior"
"initial"

$$P(H_j | \text{Information})$$

	H_0	H_1	H_2	H_3	H_4	H_5	$P(E_1)$
Initial	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0.5
E_1	0	0.07	0.13	0.20	0.27	0.33	0.73
E_2	0	0.02	0.07	0.16	0.29	0.45	0.82

E_1							0.73
E_2	0	0.20	0.30	0.30	0.20	0	0.5

E_2	0.33	0.27	0.20	0.13	0.07	0	0.27
E_2	0.45	0.29	0.16	0.07	0.02	0	0.18

E_2							0.27
E_1	0	0.20	0.30	0.30	0.20	0	0.5

$$P(H_j | E_1, E_2) = P(H_j | E_2, E_1)$$

A scatola nr. 1
pw = 0.200

H₀ H₁ H₂ H₃ H₄ H₅

67

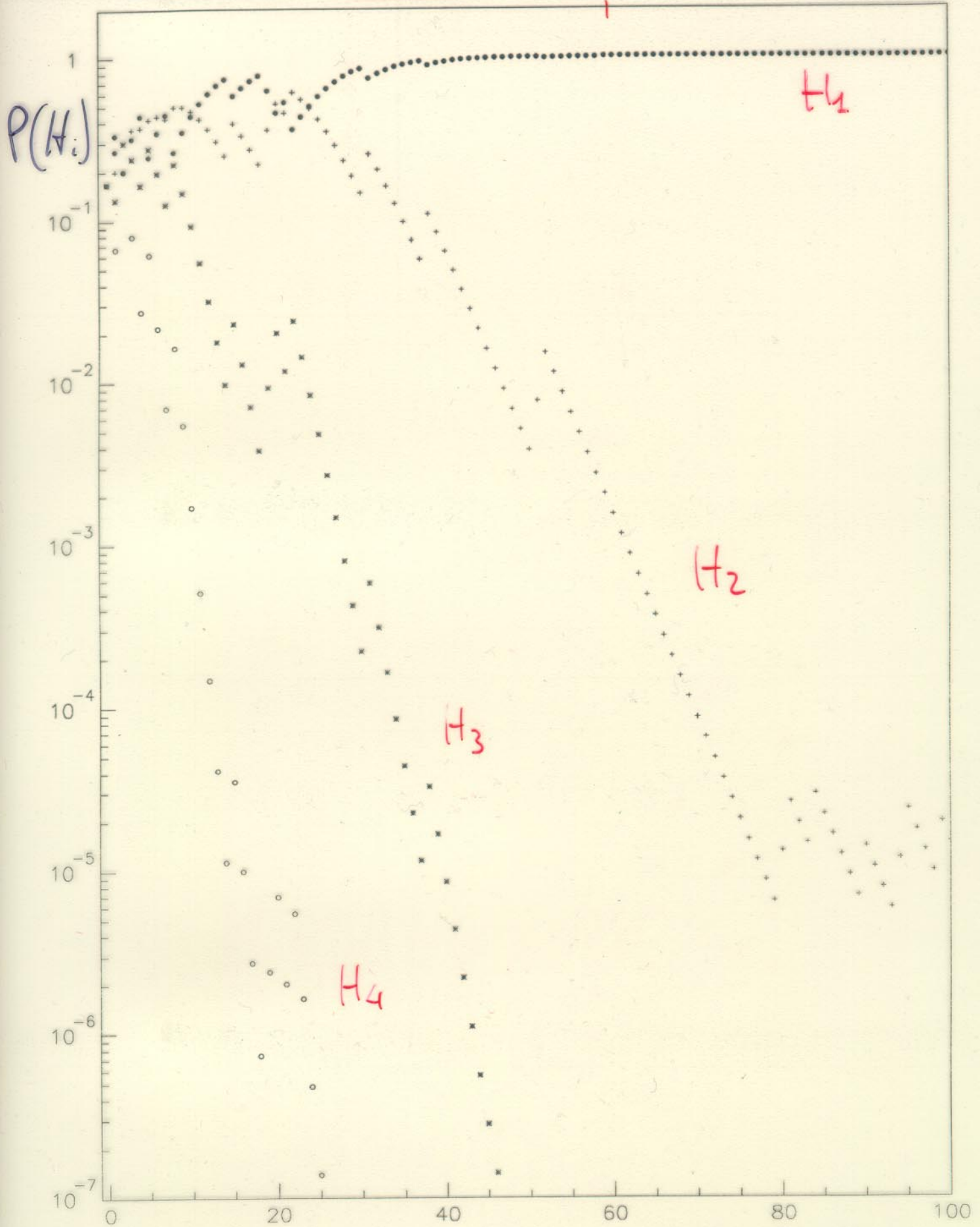
1	0	(0, 1)	0.3333	0.2667	0.2000	0.1333	0.0667	0.0000
R	→	↓	0.0000	0.2000	0.3000	0.3000	0.2000	0.0000
3	0	(1, 2)	0.0000	0.3200	0.3600	0.2400	0.0800	0.0000
4	0	(1, 3)	0.0000	0.4384	0.3699	0.1644	0.0274	0.0000
5	1	(2, 3)	0.0000	0.2462	0.4154	0.2769	0.0615	0.0000
6	0	(2, 4)	0.0000	0.3459	0.4378	0.1946	0.0216	0.0000
7	0	(2, 5)	0.0000	0.4452	0.4226	0.1252	0.0070	0.0000
8	1	(3, 5)	0.0000	0.2628	0.4990	0.2218	0.0164	0.0000
9	0	(3, 6)	0.0000	0.3495	0.4976	0.1474	0.0055	0.0000
10	0	(3, 7)	0.0000	0.4381	0.4678	0.0924	0.0017	0.0000
11	0	(3, 8)	0.0000	0.5243	0.4199	0.0553	0.0005	0.0000
12	0	(3, 9)	0.0000	0.6047	0.3632	0.0319	0.0001	0.0000
13	0	(3, 10)	0.0000	0.6771	0.3050	0.0179	0.0000	0.0000
14	0	(3, 11)	0.0000	0.7401	0.2501	0.0098	0.0000	0.0000
15	1	(4, 11)	0.0000	0.5830	0.3939	0.0231	0.0000	0.0000
16	0	(4, 12)	0.0000	0.6550	0.3320	0.0130	0.0000	0.0000
17	0	(4, 13)	0.0000	0.7194	0.2735	0.0071	0.0000	0.0000
18	0	(4, 14)	0.0000	0.7752	0.2210	0.0038	0.0000	0.0000
19	1	(5, 14)	0.0000	0.6309	0.3597	0.0094	0.0000	0.0000
20	1	(6, 14)	0.0000	0.4577	0.5219	0.0204	0.0000	0.0000
21	0	(6, 15)	0.0000	0.5326	0.4555	0.0118	0.0000	0.0000
22	1	(7, 15)	0.0000	0.3601	0.6159	0.0240	0.0000	0.0000
23	0	(7, 16)	0.0000	0.4317	0.5539	0.0144	0.0000	0.0000
24	0	(7, 17)	0.0000	0.5053	0.4862	0.0084	0.0000	0.0000
25	0	(7, 18)	0.0000	0.5780	0.4171	0.0048	0.0000	0.0000
26	0	(7, 19)	0.0000	0.6471	0.3502	0.0027	0.0000	0.0000
27	0	(7, 20)	0.0000	0.7102	0.2883	0.0015	0.0000	0.0000
28	0	(7, 21)	0.0000	0.7660	0.2332	0.0008	0.0000	0.0000
29	0	(7, 22)	0.0000	0.8138	0.1858	0.0004	0.0000	0.0000
30	0	(7, 23)	0.0000	0.8536	0.1462	0.0002	0.0000	0.0000
31	1	(8, 23)	0.0000	0.7445	0.2550	0.0006	0.0000	0.0000
32	0	(8, 24)	0.0000	0.7954	0.2043	0.0003	0.0000	0.0000
33	0	(8, 25)	0.0000	0.8383	0.1615	0.0002	0.0000	0.0000
34	0	(8, 26)	0.0000	0.8737	0.1262	0.0001	0.0000	0.0000
35	0	(8, 27)	0.0000	0.9022	0.0978	0.0000	0.0000	0.0000
36	0	(8, 28)	0.0000	0.9248	0.0752	0.0000	0.0000	0.0000
37	0	(8, 29)	0.0000	0.9425	0.0575	0.0000	0.0000	0.0000
38	1	(9, 29)	0.0000	0.8913	0.1087	0.0000	0.0000	0.0000
39	0	(9, 30)	0.0000	0.9162	0.0838	0.0000	0.0000	0.0000
40	0	(9, 31)	0.0000	0.9358	0.0642	0.0000	0.0000	0.0000
41	0	(9, 32)	0.0000	0.9511	0.0489	0.0000	0.0000	0.0000
42	0	(9, 33)	0.0000	0.9629	0.0371	0.0000	0.0000	0.0000
43	0	(9, 34)	0.0000	0.9719	0.0281	0.0000	0.0000	0.0000
44	0	(9, 35)	0.0000	0.9788	0.0212	0.0000	0.0000	0.0000
45	0	(9, 36)	0.0000	0.9840	0.0160	0.0000	0.0000	0.0000
46	0	(9, 37)	0.0000	0.9879	0.0121	0.0000	0.0000	0.0000
47	0	(9, 38)	0.0000	0.9909	0.0091	0.0000	0.0000	0.0000
48	0	(9, 39)	0.0000	0.9932	0.0068	0.0000	0.0000	0.0000
49	0	(9, 40)	0.0000	0.9949	0.0051	0.0000	0.0000	0.0000
50	0	(9, 41)	0.0000	0.9962	0.0038	0.0000	0.0000	0.0000
51	1	(10, 41)	0.0000	0.9923	0.0077	0.0000	0.0000	0.0000
52	1	(11, 41)	0.0000	0.9848	0.0152	0.0000	0.0000	0.0000
53	0	(11, 42)	0.0000	0.9885	0.0115	0.0000	0.0000	0.0000
54	0	(11, 43)	0.0000	0.9914	0.0086	0.0000	0.0000	0.0000
55	0	(11, 44)	0.0000	0.9935	0.0065	0.0000	0.0000	0.0000
56	0	(11, 45)	0.0000	0.9951	0.0049	0.0000	0.0000	0.0000
57	0	(11, 46)	0.0000	0.9963	0.0037	0.0000	0.0000	0.0000
58	0	(11, 47)	0.0000	0.9973	0.0027	0.0000	0.0000	0.0000
59	0	(11, 48)	0.0000	0.9979	0.0021	0.0000	0.0000	0.0000
60	0	(11, 49)	0.0000	0.9985	0.0015	0.0000	0.0000	0.0000
61	0	(11, 50)	0.0000	0.9988	0.0012	0.0000	0.0000	0.0000
62	0	(11, 51)	0.0000	0.9991	0.0009	0.0000	0.0000	0.0000
63	0	(11, 52)	0.0000	0.9993	0.0007	0.0000	0.0000	0.0000
64	0	(11, 53)	0.0000	0.9995	0.0005	0.0000	0.0000	0.0000
65	0	(11, 54)	0.0000	0.9996	0.0004	0.0000	0.0000	0.0000

(A)

BOX 1

60

100



Extractions

Let us consider simulation (A)
with $n=50$ (~~9 White~~^{Red} + 41 Black)

\Rightarrow which H_i ?

Consider binomial process with

$$p_i \equiv P(W | H_i) = \frac{i}{5}$$

X : "# White in n extractions"

$$n=50, x=9$$

$$P(H_i | x, n) = \frac{f(x | B_{n, p_i}) \cdot P_0(H_i)}{\sum_j f(x | B_{n, p_j}) \cdot P_0(H_j)}$$

$$P(H_i | x=9, n=50) = \frac{50!}{9! 41!} \left(\frac{1}{5}\right)^9 \left(\frac{4}{5}\right)^{41} \cdot \frac{1}{6}$$

($\sum_j \dots$)

$$\sum_j \frac{1 \cdot 50!}{6 \cdot 9! 41!} \left[\left(\frac{0}{5}\right)^9 \left(\frac{5}{5}\right)^{41} + \left(\frac{1}{5}\right)^9 \left(\frac{4}{5}\right)^{41} + \left(\frac{2}{5}\right)^9 \left(\frac{3}{5}\right)^{41} + \dots + \left(\frac{5}{5}\right)^9 \left(\frac{0}{5}\right)^{41} \right]$$

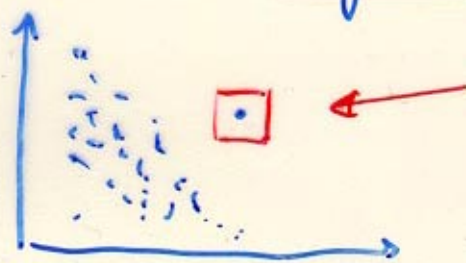
$$\boxed{\rightarrow 0.996}$$

⇒ The conclusions of the Bayesian inference depend only on

- observations
- previous knowledge

⇒ They don't depend on the way how data are treated (if done consistently with Bayesian updating procedure)

⇒ In particular, they don't depend on what the experimentalist wanted to observe before taken data



conventional approach:
the window must have been chosen before

⇒ Nice paper

J.O. Berger & D.A. Berry, "Illusion of objectivity"
Am. Scientist, 76 ('88) 158

Similar problem (forget for the moment the previous!)

- A box could contain 0, 1, 2, 3, 4, 5 ~~white~~ ^{Red} balls out of 5 total balls.
- What is the probability of H_0, \dots, H_5 ?
- Extract and reintroduce balls, as in the previous problem:
 $\rightarrow P(H_i | E_n^{(w)})$?

Same solution? It could be, but it could as well not be!

The 5 balls could have been chosen randomly from a large amount of balls, of which 50% ~~are~~ ^{are} white and 50% black

$$P_0(H_i) = (H_i | E_0) = \frac{5!}{i!(5-i)!} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{5-i}$$

	0	1	2	3	4	5
$P_0(H_i)$	0.031	0.156	0.313	0.313	0.156	0.031

$P(H_j | \mu_0)$

	H_0	H_1	H_2	H_3	H_4	H_5	
Initial	0.03	0.16	0.31	0.31	0.16	0.03	0.5
E_1	0	0.06	0.25	0.38	0.25	0.06	0.60
E_2	0	0.02	0.16	0.38	0.33	0.10	0.67
E_1							0.60
E_2	0	0.13	0.38	0.38	0.13	0	0.5

⇒ Differences on conclusions on $P(H_i | E_n)$?

⇒ What is the right solution?

⇒ It makes no sense to speak about the "right solution":

→ Priors depend on the status of knowledge (subjective)

→ after many "extractions" the conclusions will depend very little from priors

→ "practically" objective conclusions

Poincaré (Science et Hypothèse)

An effect may be produced by the cause a or by the cause b . The effect has just been observed. We ask the probability that it is due to the cause a . This is an à posteriori probability of cause. But I could not calculate it, if a convention more or less justified did not tell me in advance what is the à priori probability for the cause a to come into play—

"On se rappelle constamment de parallèles
en physique."

"Les lois ne nous sont connues
que par leurs effets qu'on observe.
Chercher à en déduire les lois qui
sont des causes, c'est résoudre un
problème de probabilité des causes."

- Calcul des probabilités -

Let's consider again

$$P(E^{(2)} = E_1 \mid E^{(1)} = E_2) = 0.27$$

It was calculated from

$$P(E_1 \mid E^{(1)} = E_2) = \sum_j \underbrace{P(E_1 \mid H_j)}_{\substack{\text{in principle} \\ P(E_1 \mid H_j, E^{(1)} = E_2)}} \cdot P(H_j \mid E^{(1)} = E_2)$$

but independent...

$$= 0 \times 0.33 + \frac{1}{5} \times 0.27 + \frac{2}{5} \times 0.2 + \frac{3}{5} \times 0.13 + \frac{4}{5} \times 0.067 + 1 \times 0 = \underline{\underline{0.267}}$$

What means? "we believe 27% that $E^{(2)}$ will be Red, given the available information" (already said...) \Rightarrow

Note

- 1) 27% has no interpretation in terms of favorable/possible events (rather easy to understand)
- 2) 27% has no interpretation in terms of long run frequency either (and this need some discussion...)

⇒ Thinking of measuring $P(E_1 | E^{(1)} = E_2)$ makes no sense!

If we imagine to make an experiment to measure the frequency for " $n \rightarrow \infty$ " we shall get, most likely 0%, or 20%, or 40%, or 60%, or 80%, but not 27%

⇒ "The Bayesian conclusion is clearly wrong"



But who said that 27% means
frequencies?

⇒ 27% is simply the rational belief
that the next ball will be red,
based on the best use of all
available information in hand
nothing more!

Does it mean that if the experiment
is made and $\approx 20\%$ frequency
is obtained, ^{with $n=100$ trials} a Bayesian will
defend his position?

|| **YES**, in the sense that for him
|| $P(E^{(2)} = E_1 | E^{(1)} = E_2)$ was **27%!**
(as "head" tossing a coin was $\frac{1}{2}$
before the coin was tossed)

BUT →

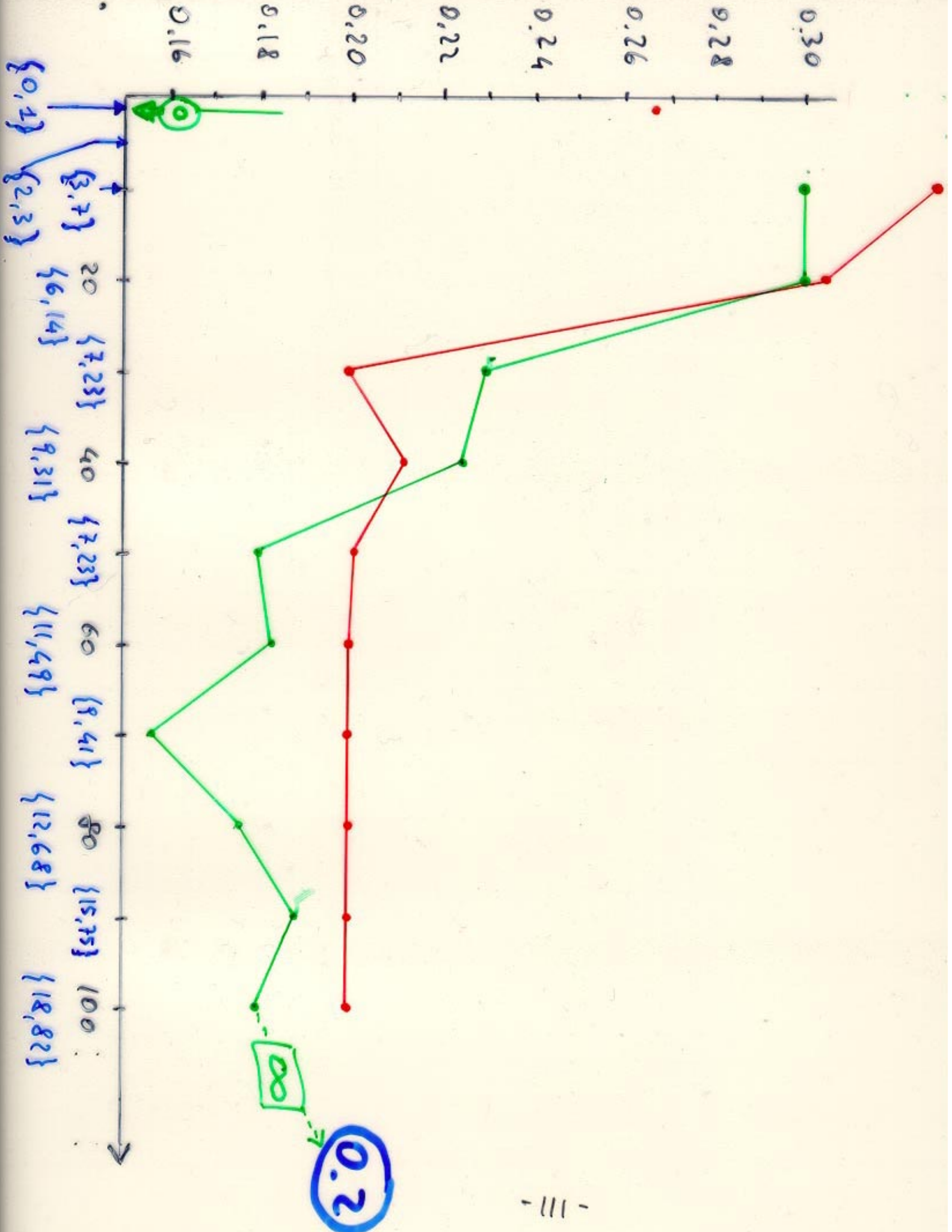
Now, the interesting question is rather

$$P(E^{\underline{n+1}} = E_1 \mid \frac{\#E_1}{\#E_2} = 0.2, n=100)$$

and this is a different story!

Probabilities

n	f/n	Bayesian	frequentist
0	-	50%	?
1	0/1	27%	0
5	2/5	43%	40%
10	3/10	33%	30%
20	6/20	31%	30%
30	7/30	20.0%	23.3%
40	9/40	21.3%	22.5%
50	9/50	20.1%	18.0%
60	11/60	20.03%	18.3%
70	11/70	20.002%	15.7%
80	12/80	20.000%	17.6%
90	15/90	20.000%	18.8%
100	18/100	20.000%	18.0%
		∇	↓ "n = ∞" ↓ 20%



$\{0, 1\}$
 $\{3, 7\}$
 $\{2, 3\}$
 20 $\{7, 23\}$
 $\{6, 14\}$
 40 $\{7, 23\}$
 $\{9, 31\}$
 60 $\{7, 23\}$
 $\{11, 49\}$
 80 $\{9, 41\}$
 $\{12, 68\}$
 100 $\{15, 75\}$
 $\{18, 82\}$

0.2

∞

Does it violate Bernoulli's theorem?
("law of large numbers")

$$\lim_{n \rightarrow \infty} f_n \rightarrow p$$

⇒ If we think to $n \rightarrow \infty$ ^{situations} in which we are exactly in the same status of knowledge given by $E^{(1)} = E_2$, i.e. we are interested to

$$P(E^{(2)} = E_2 \mid E^{(1)} = E_2)$$

there is nothing wrong:

$$\lim_{n \rightarrow \infty} f_n = 27\%$$

(try with a little Monte Carlo simulation)

⇒ We should not confuse the event we are interested to with

$$E^{(2)} = E_1 \mid H = H_1,$$

which is something different.