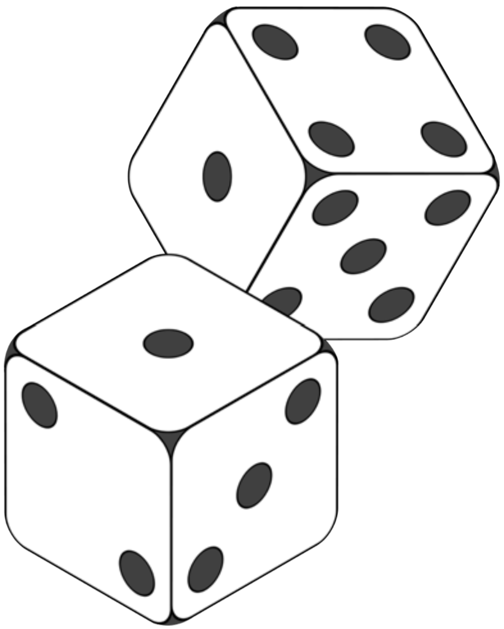
Introduction to Bayesian Methods - 1

Edoardo Milotti
Università di Trieste and INFN-Sezione di Trieste

Point your browser to

http://wwwusers.ts.infn.it/~milotti/Didattica/Bayes/Bayes.html

The nature of probabilities



In a dice throwing game one defines probabilities of different events by counting the outcomes

Examples:
- with one die, the probability of getting a 4 is 1/6
- with two dice, the probability of getting two 4's is 1/36
- with two dice, the probability of getting one 4 AND one 5 is 1/18
- with two dice, the probability of getting one 4 OR one 5 is ??

Die 1	Die 2	# of outcomes
4	Not 4 or 5	4
5	Not 4 or 5	4
Not 4 or 5	4	4
Not 4 or 5	5	4
4	4	1
4	5	1
5	4	1
5	5	1
		Total: 20

NB, if 4 and 5 were independent, we would have

P(4 OR 5) = P(4) + P(5) = 1/3 + 1/3 = 2/3

$$p = \frac{20}{36} = \frac{5}{9} < \frac{2}{3}$$

Die 1	Die 2	# of outcomes
4	Not 4 or 5	4
5	Not 4 or 5	4
Not 4 or 5	4	4
Not 4 or 5	5	4
4	4	1
4	5	1
5	4	1
5	5	1
		Total: 20

Outcomes = elementary events

Composite events contain many elementary events

We usually assume that elementary events are all equally likely.

This is not true for biased dice.

Coin tossing as a model of randomness

Sequence #1

T HHHHT T T T HHHHT HHHHHHHHT T THHT T HHHHHT T T T T T HHT HHT HHH T
T T HT T HHHHT HT T T HT T T HHT T T T HHHHHHT T T HHT T HHHT HHHHHT T T T
T HT T T HHT T HT T HHT T T HHT T T HH THHT HHT T T T T HHT HHHHHHT HT HT T
HT HT T HHHT T HHT HT HHHHHHHT THT T HHHT HHT T HT T T T T T HHHT HHH

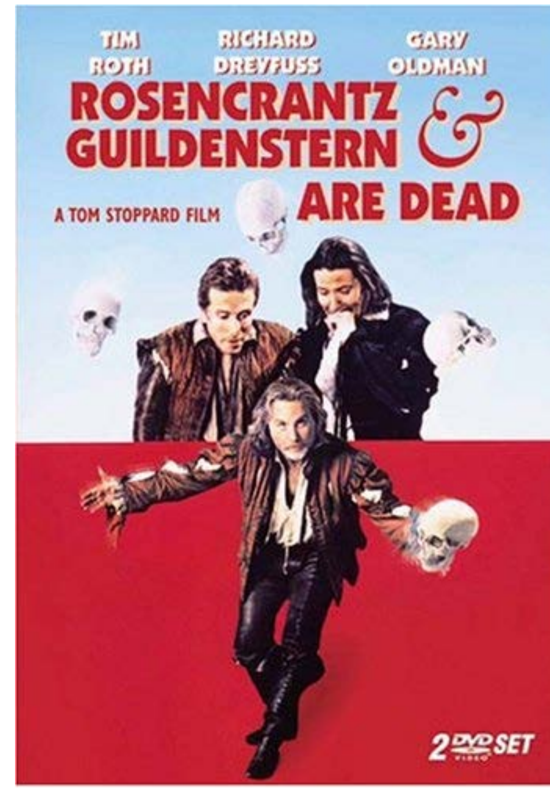
Sequence #2

T HT HT T T HT T T T T HT HT T T HT T HHHT HHT HT HT HT T T T HHT T HHT T HHH T
HHHT T HHHT T T HHHT HHHHT T T HT HT HHHHT HT T T HHHT HHT HT T T HHT H
HHT HHHHT T HT HHT HHHT T T HT HHHT HHT T T HHHT T T HHHT HT HHHHT H
T T HHT T T T HT HT HT T HT HHT T HT T THT T T T HHHHT HT HHHT T HHHHHT HH

One of these sequences has been obtained with a real coin, the other one is artificial. Which one belongs to the true coin?

Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads, again.

Tom Stoppard’s classic play *Rosencrantz and Guildenstern Are Dead* opens with two Elizabethan players, some well-stocked prop moneybags, and the flip of a coin that lands as heads. Again. And again. And again.



In Stoppard’s scene, the bit actors Rosencrantz and Guildenstern kill time during a production of Shakespeare’s *Hamlet* by betting on coin tosses.

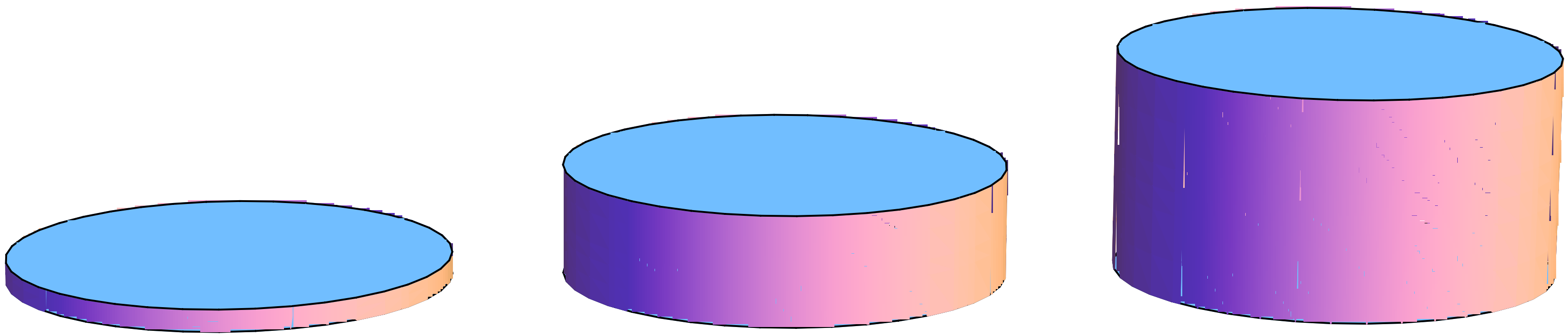
Guildenstern flips a florin and Rosencrantz predicts that it will land as heads. It does. Guildenstern spins another coin and it lands as heads again.

After Rosencrantz has successfully bet heads 77 times in a row, Guildenstern proclaims that, “A weaker man might be moved to re-examine his faith, if in nothing else at least in the law of probability.” He ends up flipping heads 92 times in a row.

What are the odds of such a thing happening?

Now, consider “coins” with different aspect ratio *r*

(aspect ratio = thickness/diameter)



r = 0.05　　r=0.25　　r = 0.5

How do these coins land on heads, tails, sides? When is the probability of landing on the side equal to the probability of landing on heads or tails?

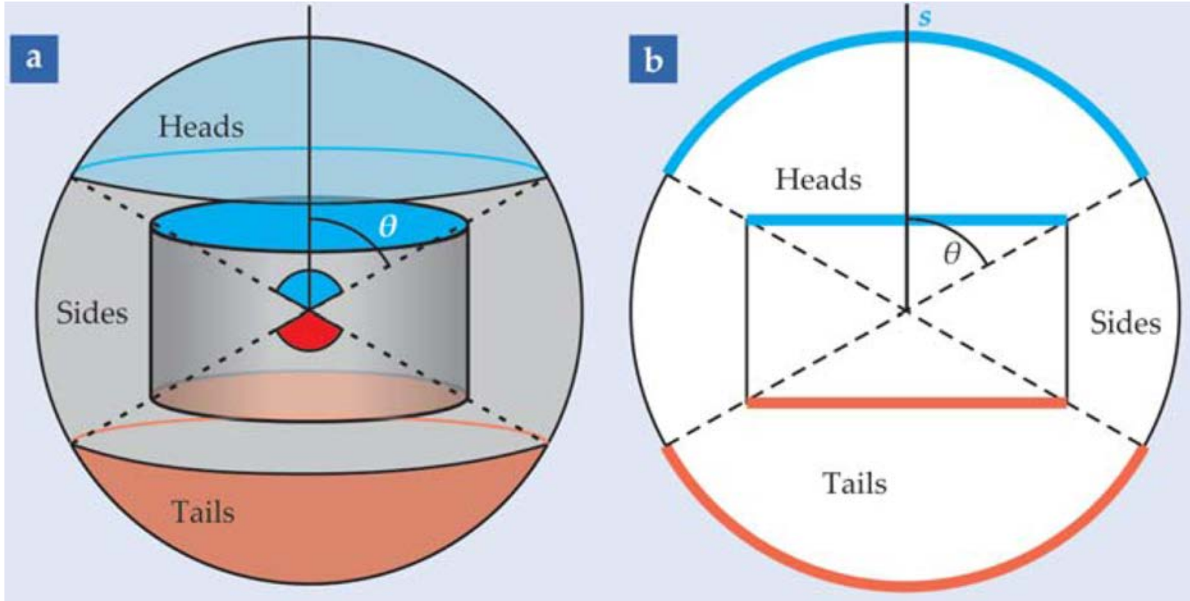
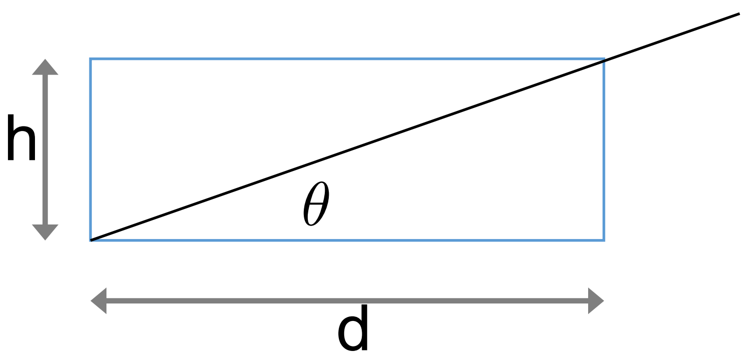


figure from Mahadevan and Yong, "Probability, physics, and the coin toss", Phys. Today, July 2011, pp. 66-67

a. Von Neumann's answer: consider solid angles subtended by heads, tails, sides

$$2\pi \times \int_0^{\theta_0} \sin\theta d\theta = 2\pi(1-\cos\theta_0)$$

$$\begin{aligned}\Omega_{\text{heads}} = \Omega_{\text{tails}} = \Omega_{\text{sides}} &= 4\pi/3 \\ \Rightarrow\ 2\pi(1-\cos\theta_0) &= 4\pi/3 \\ \Rightarrow\ \frac{h}{\sqrt{h^2+d^2}} = \frac{r}{\sqrt{r^2+1}} &= 1/3 \\ \Rightarrow\ r = 1/2\sqrt{2}\end{aligned}$$



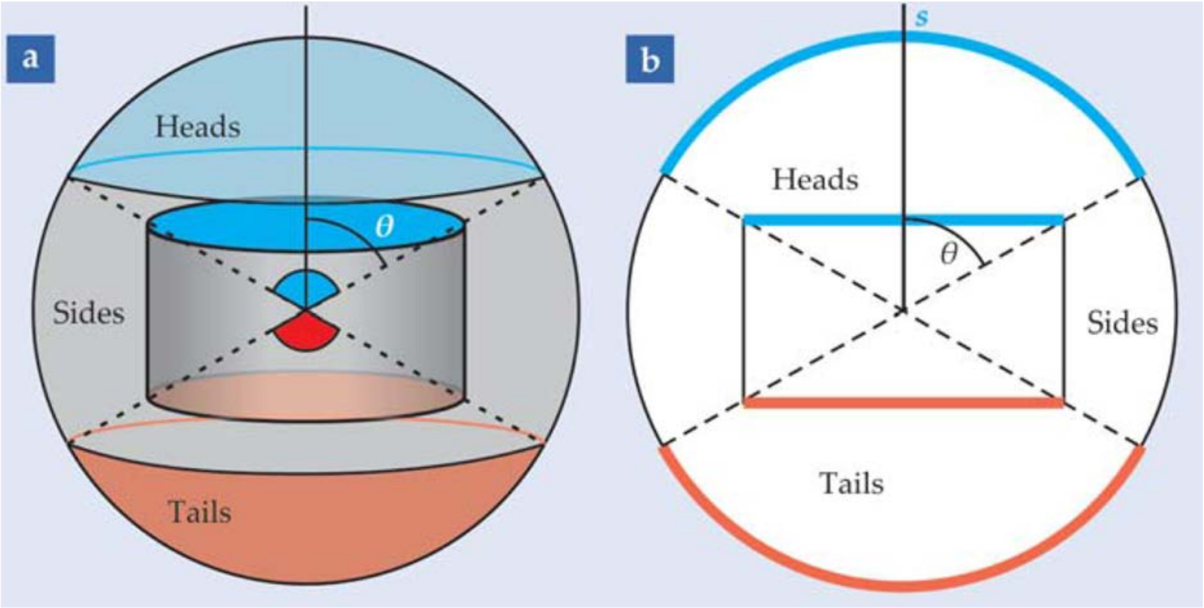


figure from Mahadevan and Yong, "Probability, physics, and the coin toss", Phys. Today, July 2011, pp. 66-67

b. alternative answer: consider *angles* subtended by heads, tails, sides (rotation about axis through center of coin, and parallel to faces)

$$
\begin{aligned}
\theta_{\text{heads}} &= \theta_{\text{tails}} = \theta_{\text{sides}} = \pi/3 \\
&\Rightarrow \cos\theta_0 = 1/2 \\
&\Rightarrow \frac{h}{\sqrt{h^2+d^2}} = \frac{r}{\sqrt{r^2+1}} = 1/2 \\
&\Rightarrow r = 1/\sqrt{3}
\end{aligned}
$$

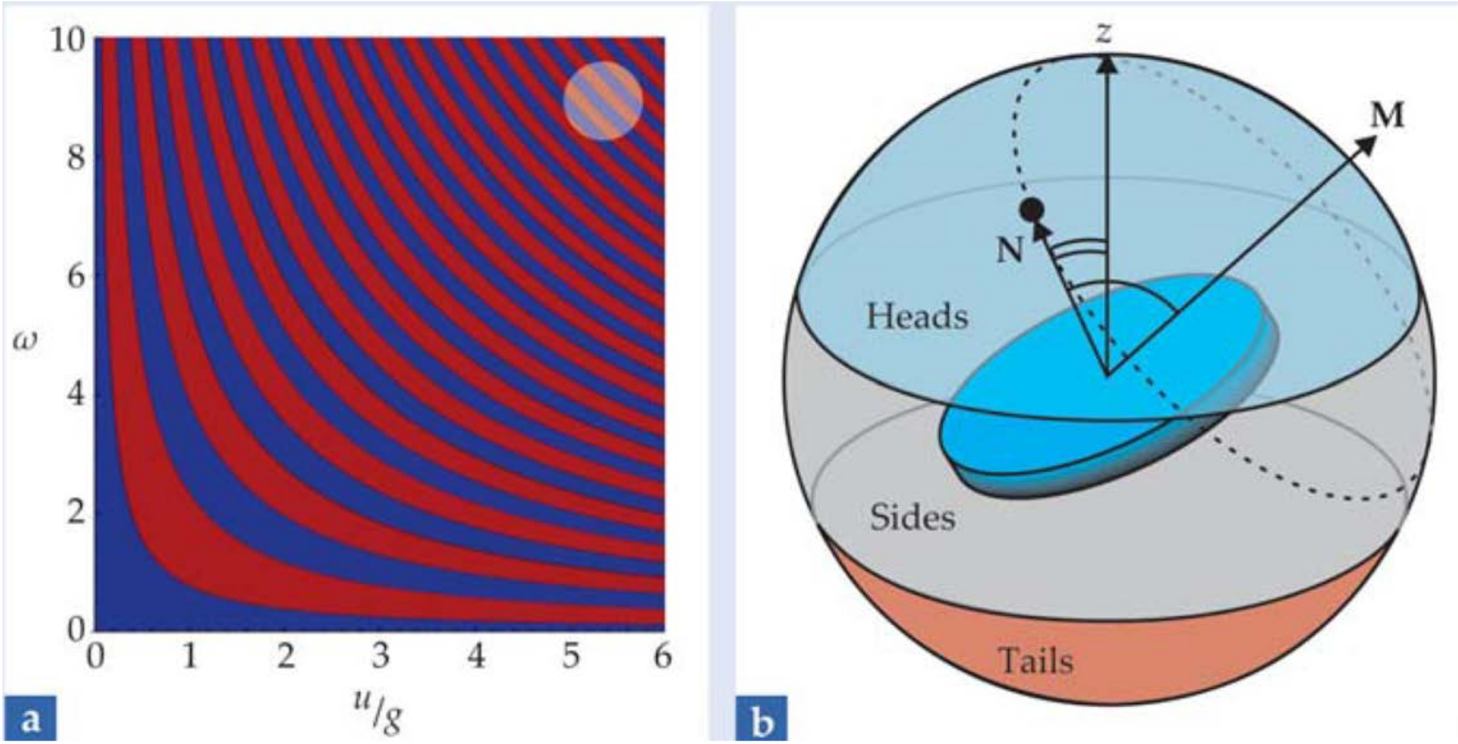
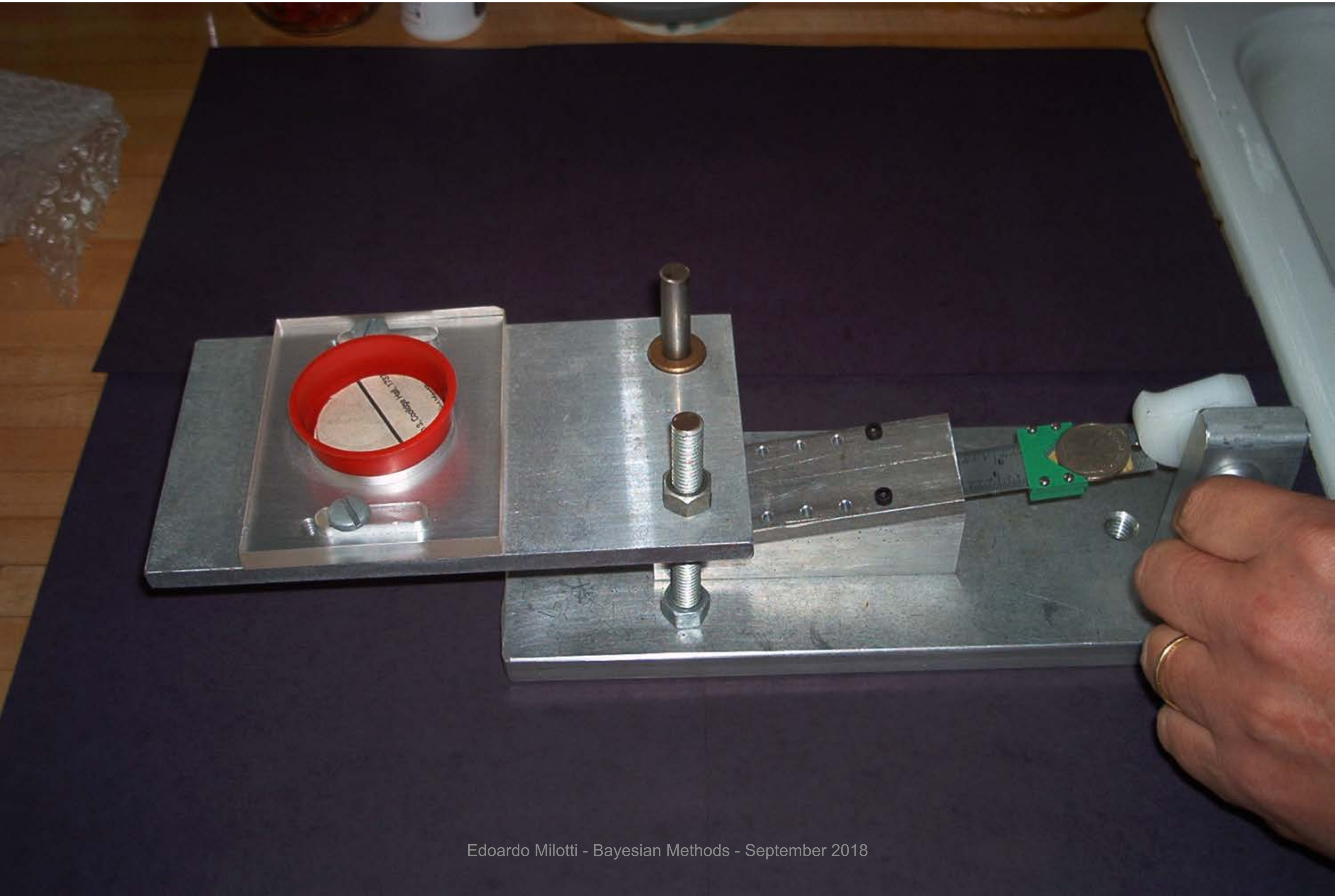


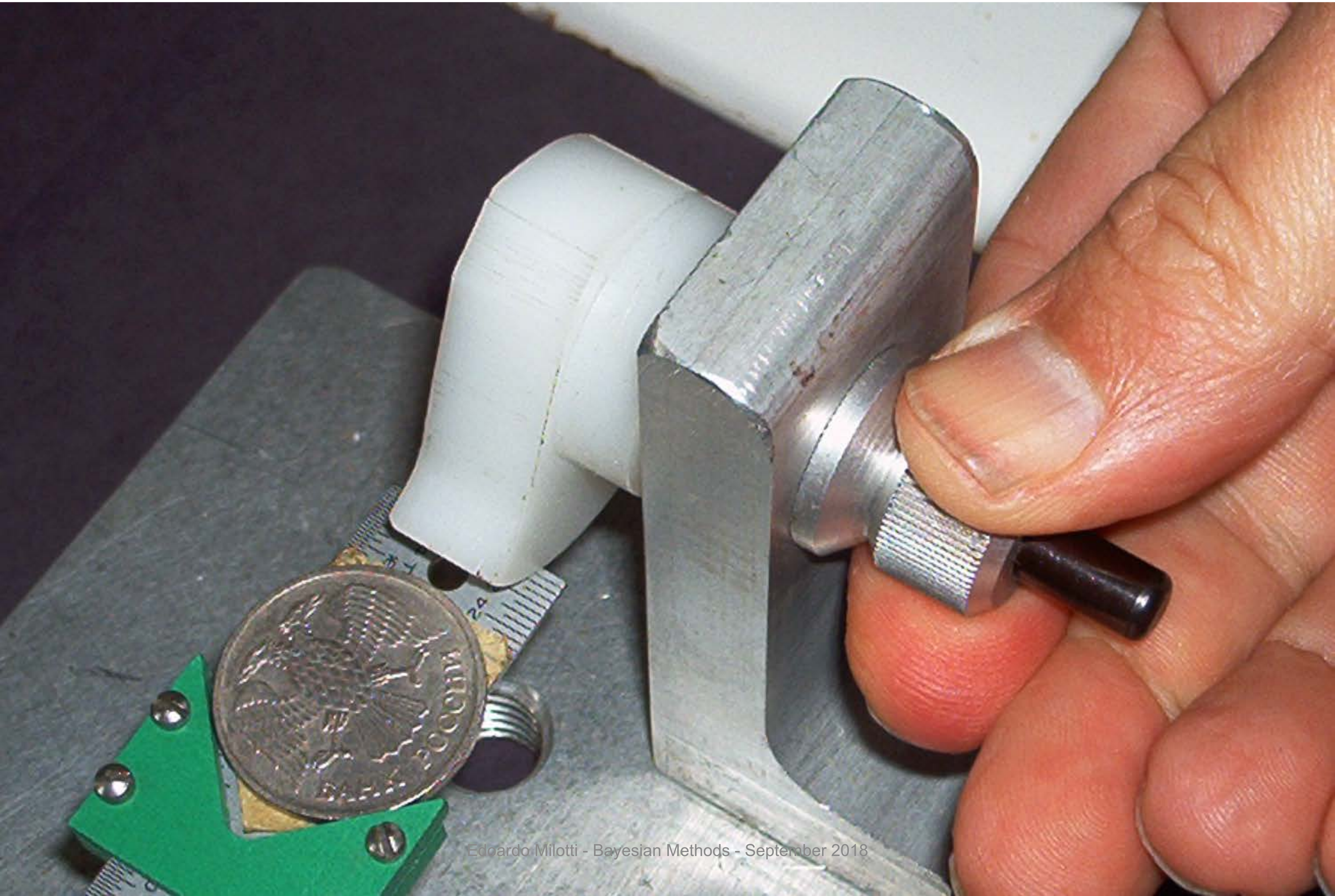
figure from Mahadevan and Yong, "Probability, physics, and the coin toss", Phys. Today, July 2011, pp. 66-67

In 1986 J. B. Keller analyzed the infinitely thin coin and found that coin toss is not random for finite rotation speed and vertical speed (rotation axis as in previous case b)

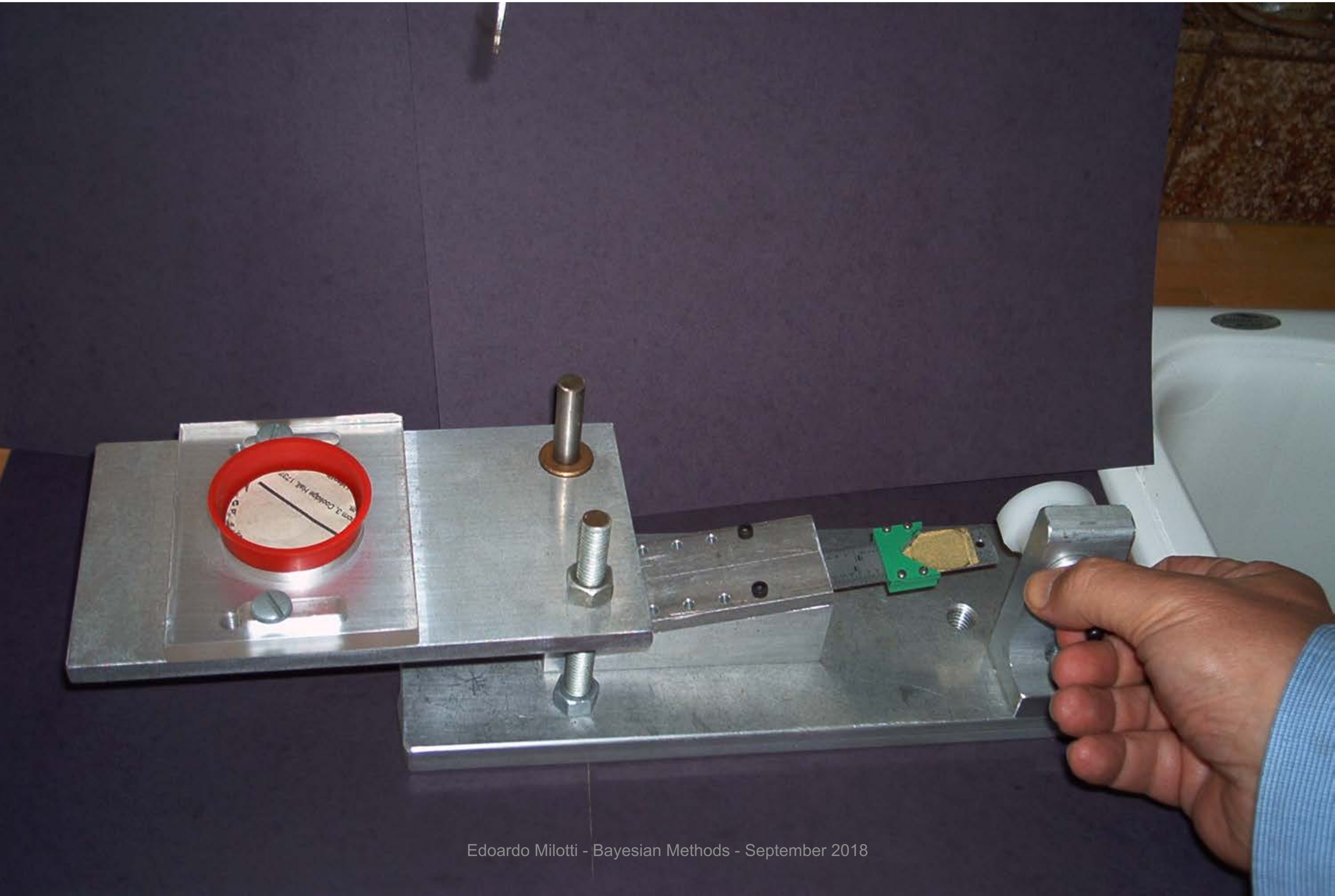
Coin tossing machine (Diaconis, Holmes and Montgomery 2007)



Coin tossing machine (Diaconis, Holmes and Montgomery 2007)



Coin tossing machine (P. Diaconis, S. Holmes and R. Montgomery 2007)



Coin tossing machine (Diaconis, Holmes and Montgomery 2007)

... Coin-tossing is a basic example of a random phenomenon. However, naturally tossed coins obey the laws of mechanics (we neglect air resistance) and their flight is determined by their initial conditions. Figure 1 a-d shows a coin-tossing machine. The coin is placed on a spring, the spring released by a ratchet, the coin flips up doing a natural spin and lands in the cup. With careful adjustment, the coin started heads up always lands heads up – one hundred percent of the time. We conclude that coin-tossing is ‘physics’ not ‘random’. ...

(Diaconis, Holmes and Montgomery, “Dynamical bias in the coin toss”, *SIAM Rev.* **49** (2007) 211)

Therefore, the assumed randomness of coin toss – and in general, of complex mechanical processes – is related to the difficulty in determining the outcome, both because of the complex and often unknown dynamics, and because of the uncertain initial conditions.

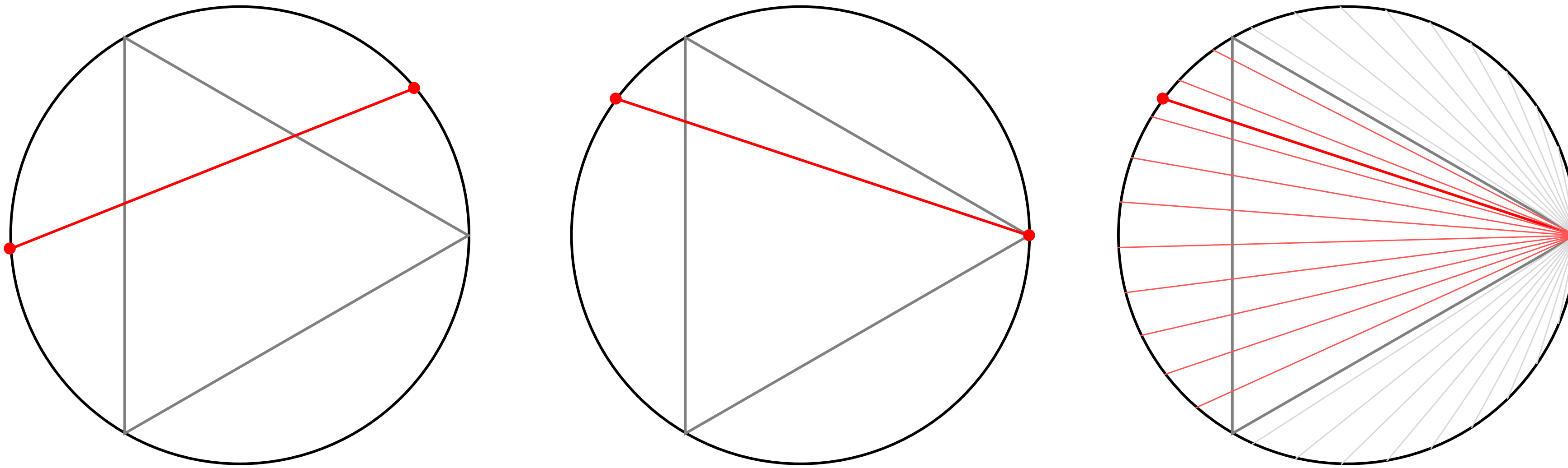
Thus – at least in this case – probabilities are a measure of our own ignorance rather than an intrinsic property of the physical system.

Bertrand's paradox and the ambiguities of probability models

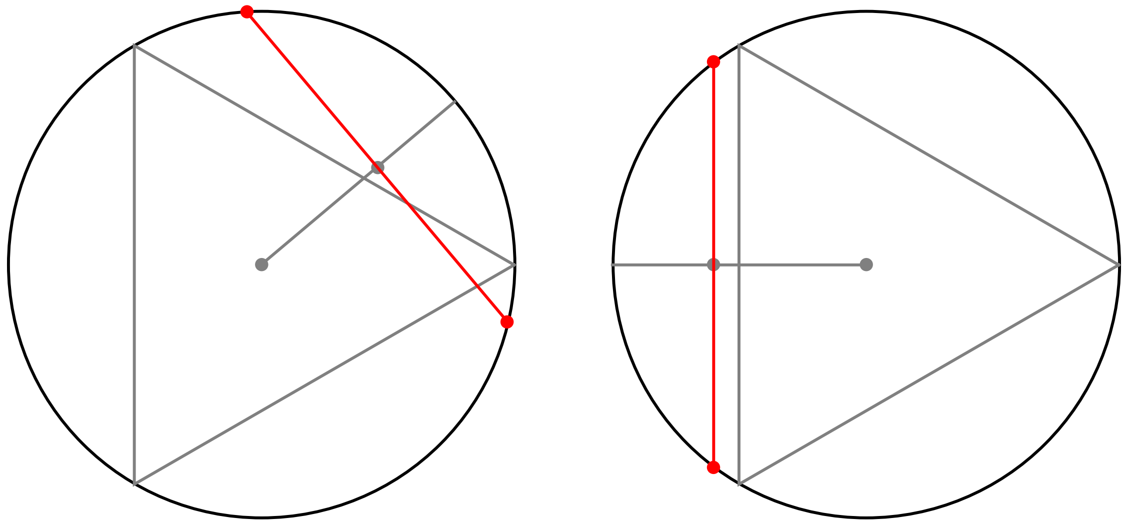
Bertrand's paradox goes as follows:

"consider an equilateral triangle inscribed inside a circle, and suppose that a chord is chosen at random. What is the probability that the chord is longer than a side of the triangle?"

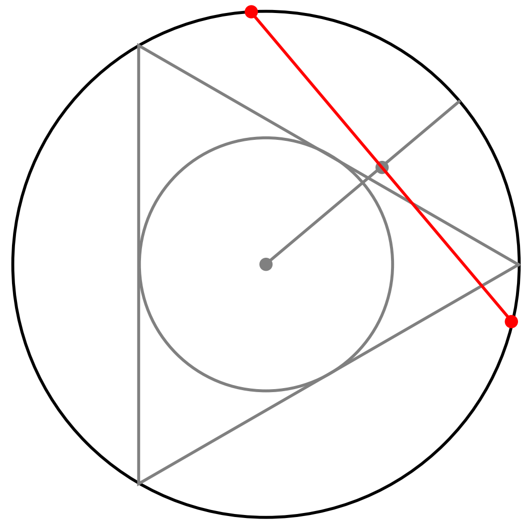
(Bertrand, 1889)



Solution: we take two random points on the circle (radius *R)*, then we rotate the circle so that one of the two points coincides with one of the vertices of the inscribed triangle. Thus a random chord is equivalent to taking the first point that defines the chord as one vertex of the triangle while the other is taken “at random” on the circle. Here “at random” means that it is uniformly distributed on the circumference. Then only those chords that cross the opposite side of the triangle are actually longer than each side. Since the subtended arc is 1/3 of the circumference, the probability of drawing a random chord that is longer than one side of the triangle is 1/3.



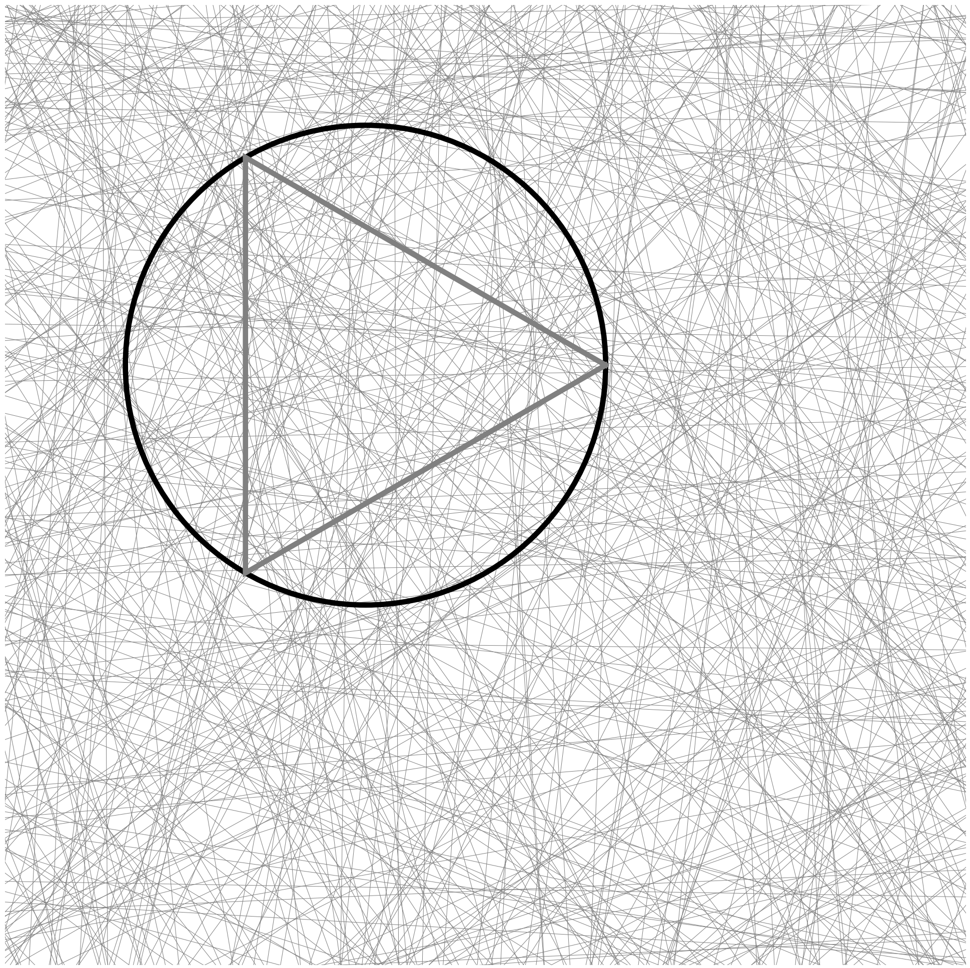
Solution 2: we take first a random radius, and next we choose a random point on this random radius. Then, we take the chord through this point and perpendicular to the radius. When we rotate the triangle so that the radius is perpendicular to one of the sides, we see that half of the points give chords longer than one side of the triangle, therefore the probability is 1/2.

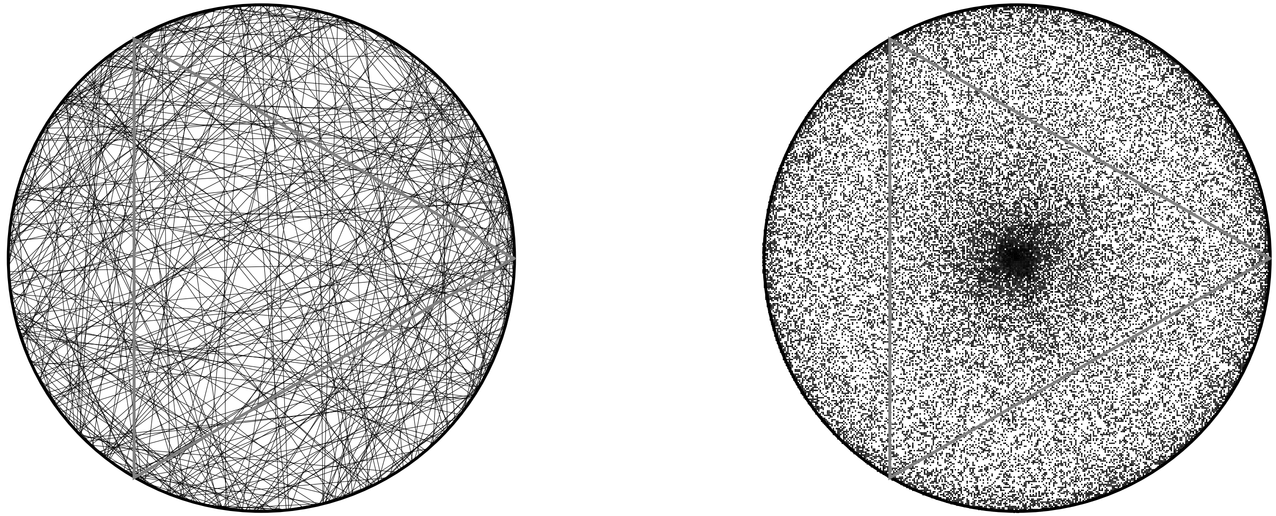


Solution 3: we take the chord midpoints located inside the circle inscribed in the triangle, and we obtain chords that are longer than one side of the triangle. Since the ratio of the areas of the two circles is 1/4, we find that now the probability of drawing a long chord is just 1/4.

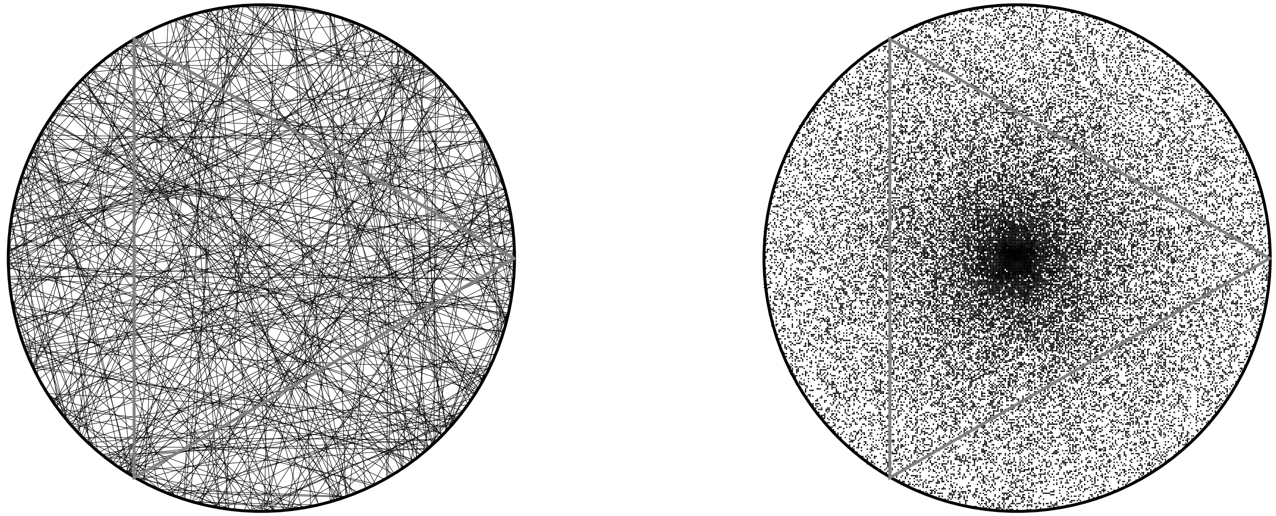
At least 3 different "solutions": which one is correct, and why?

Now we widen the scope of the problem and we consider the distribution of chords in the plane



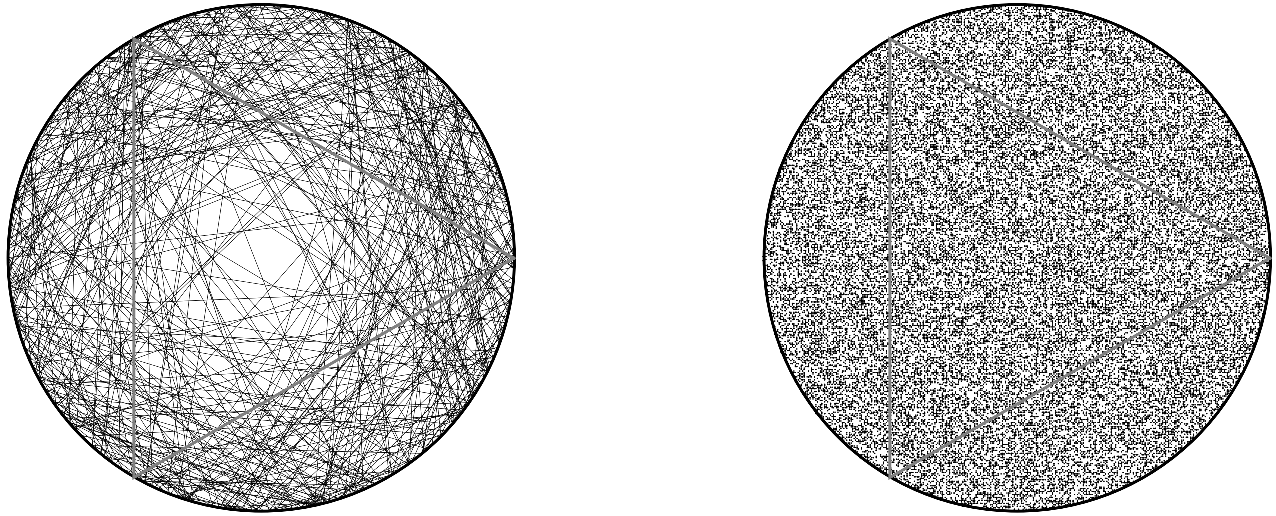


Distribution 1: distribution of chords (left panel) and of midpoints (right panel) in the first solution of Bertrand’s paradox (the left panel shows 400 chords, the right panel shows 100000 midpoints).

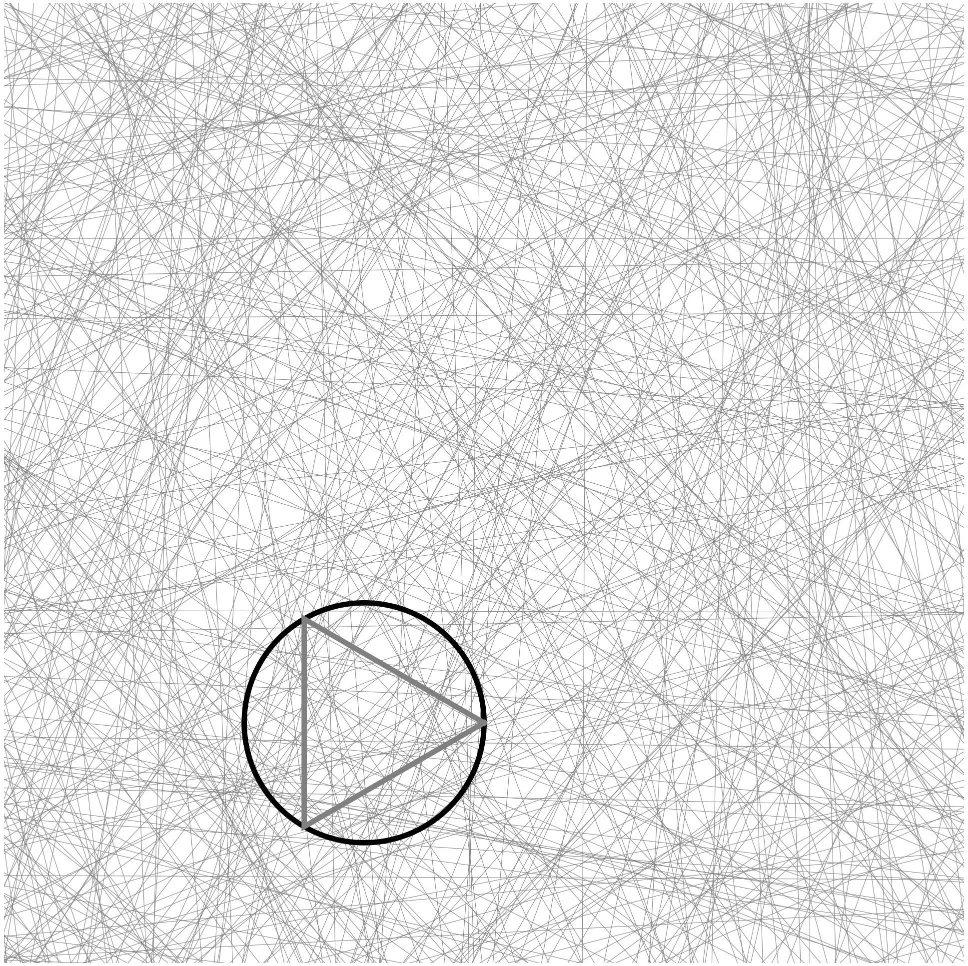


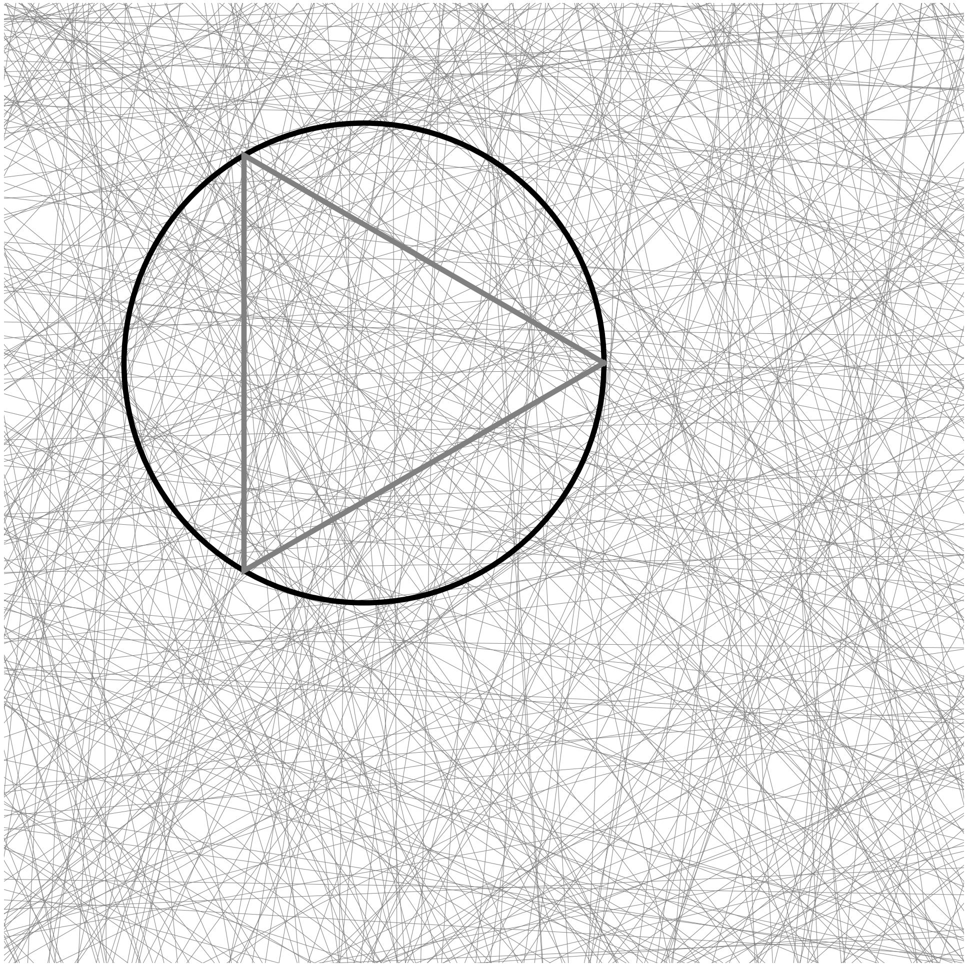
Distribution 2: Distribution of chords (left panel) and of midpoints (right panel) in the second solution of Bertrand's paradox (the left panel shows 400 chords, the right panel shows 100000 midpoints).

In this case it is very easy to find the radial density function of chord centers, since here we take first a random radius, and next we choose a random point (the center) on this random radius.



Distribution 3: Distribution of chords (left panel) and of midpoints (right panel) in the third solution of Bertrand’s paradox (the left panel shows 400 chords, the right panel shows 100000 midpoints). Notice that while the distribution of midpoints is uniform, the distribution of the resulting chords is distinctly non-uniform.





Hidden assumptions (Jaynes):

- rotational invariance
- scale invariance
- translational invariance

Now let

$$f(r, \theta)$$

be the probability density of chord centers

Rotational invariance

In a reference frame which is at an angle α with respect to the original frame, i.e., the new angle $\theta' = \theta - \alpha$, the distribution of centers is given by a different distribution function $g(r, \theta') = g(r, \theta - \alpha)$. Since we require rotational invariance

$$f(r, \theta) = g(r, \theta - \alpha)$$

with the condition $g(r, \theta)|_{\alpha=0} = f(r, \theta)$, and this must hold for every angle α, so the only possibility is that there is no dependence on θ, and $f(r, \theta) = g(r, \theta) = f(r)$.

Scale invariance

When we consider a circle with radius R, the normalization of the distribution $f(r)$ is given by the integral

$$\int_0^{2\pi}\int_0^R f(r)rdrd\theta = 2\pi\int_0^R f(r)rdr = 1$$

The same distribution induces a similar distribution $h(r)$ on a smaller concentric circle with radius aR $(0 < a < 1)$, such that $h(r)$ is proportional to $f(r)$, i.e., $h(r) = Kf(r)$, and

$$1 = 2\pi\int_0^{aR} h(u)udu = 2\pi\int_0^{aR} Kf(u)udu = 2\pi K\int_0^{aR} f(u)udu$$

i.e.,

$$K^{-1} = 2\pi\int_0^{aR} f(u)udu$$

and

$$f(r) = 2\pi h(r)\int_0^{aR} f(u)udu$$

inside the smaller circle.

Now we invoke the assumed scale invariance: the probability of finding a center in an annulus with radii r and $r + dr$ in the original circle, must be equal to the probability of finding a center in the scaled down annulus,

$$h(ar)(ar)d(ar) = f(r)rdr$$

and therefore

$$a^2 h(ar) = f(r)$$

Equation

$$a^2 h(ar) = f(r)$$

can also be rewritten in the form

$$h(r) = \frac{1}{a^2} f\left(\frac{r}{a}\right) \tag{1}$$

and inserting this into equation

$$f(r) = 2\pi h(r) \int_0^{aR} f(u) u du$$

we find

$$a^2 f(ar) = 2\pi f(r) \int_0^{aR} f(u) u du \tag{2}$$

We solve equation

$$a^2 f(ar) = 2\pi f(r) \int_0^{aR} f(u) u du$$

taking first its derivative with respect to a: the relation that we find must hold for all a's, and therefore also for $a = 1$ (no scaling), and we find the differential equation

$$r f'(r) = \left(2\pi R^2 f(R) - 2\right) f(r)$$

i.e.,

$$r f'(r) = (q - 2) f(r)$$

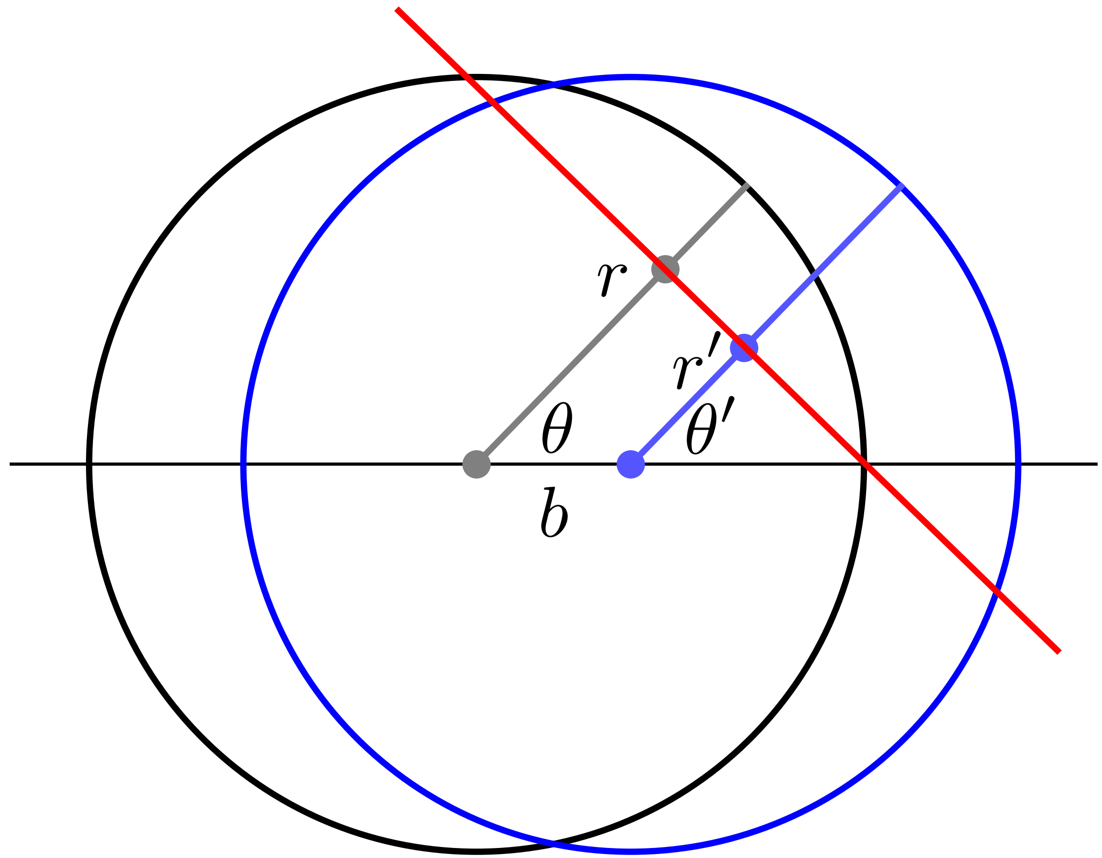
where the constant $q = 2\pi R^2 f(R)$ is unknown. However, we can still solve the equation and find

$$f(r) = A r^{q-2}$$

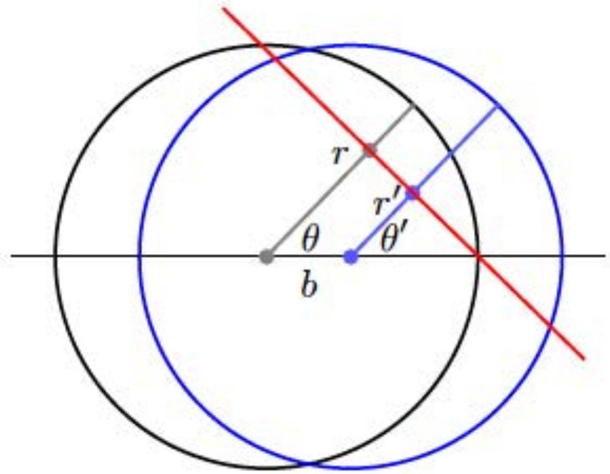
The constant A is easy to find from the normalization condition: $A = q/2\pi R^q$, and therefore

$$f(r) = \frac{q r^{q-2}}{2\pi R^q}$$

Translational invariance



Geometrical construction for the discussion of translational invariance. The original circle (black) is crossed by a straight line (red) which defines the chord. The translated circle is shown in blue.



This circle is displaced by the amount b, and the new radius and angle that define the midpoint of the chord are

$$r' = |r - b\cos\theta|$$
$$\theta' = \theta \ \ (\text{if } r \geq b\cos\theta) \quad \text{or} \quad \theta' = \theta + \pi \ \ (\text{if } r < b\cos\theta)$$

Now consider a region Γ surrounding the midpoint in the original circle, which is transformed into a region Γ' by the translation. The probability of finding a chord with the midpoint in the region Γ is

$$\int_\Gamma f(r) r dr d\theta = \int_\Gamma \frac{q r^{q-1}}{2\pi R^q} dr d\theta = \frac{q}{2\pi R^q} \int_\Gamma r^{q-1} dr d\theta$$

Likewise, the same probability for the translated circle is

$$\frac{q}{2\pi R^q} \int_{\Gamma'} (r')^{q-1} dr' d\theta' = \frac{q}{2\pi R^q} \int_\Gamma |r - b\cos\theta|^{q-1} dr d\theta \qquad (3)$$

where the Jacobian of the transformation is 1. Equating these expressions, we see that the integrand must be a constant, and therefore $q = 1$, and

$$f(r, \theta) = \frac{1}{2\pi R r} \quad (r \le R;\ \ 0 \le \theta < 2\pi)$$

Therefore

$$f(r,\theta) = f(r) = C/r$$

$$\Rightarrow \quad (\text{normalization}) \quad 1 = \int_C f(r) 2\pi r dr = 2\pi C R$$

$$\Rightarrow \quad f(r) = \frac{1}{2\pi r R}$$

Using this distribution, we find that the probability of finding a midpoint inside the circle with radius $R/2$ – i.e., the probability of finding a chord longer than the side of the triangle in Bertrand's paradox – is

$$\int_0^{2\pi} d\theta \int_0^{R/2} f(r,\theta) r dr = 2\pi \int_0^{R/2} \frac{1}{2\pi R r} r dr = \frac{1}{2}$$

which corresponds to the second alternative in the previous discussion of Bertrand's paradox.

Lesson drawn from Bertrand's paradox:

probability models depend on physical assumptions, they are not God-given. We define the elementary events on the basis of real-world constraints, derived from our own experience.

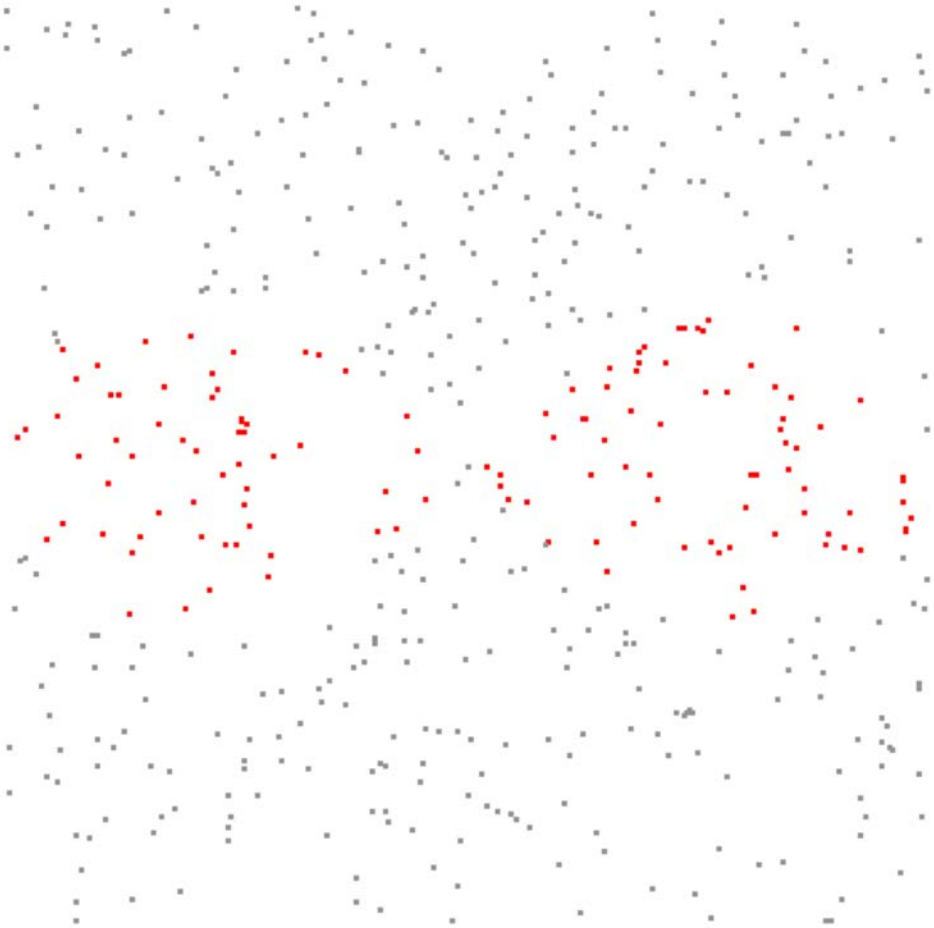
Probabilities as a measure of “reasonable expectation”, and their relationship with statistical inference. (Cox, 1946)

- We construct – explicitly or implicitly – probabilistic theoretical models to understand measurements (the most common such model is the Gaussian model)
- We utilize the empirical probability distributions to infer the parameter values of physical models

What if we "measure" a mathematical constant instead of a physical parameter?

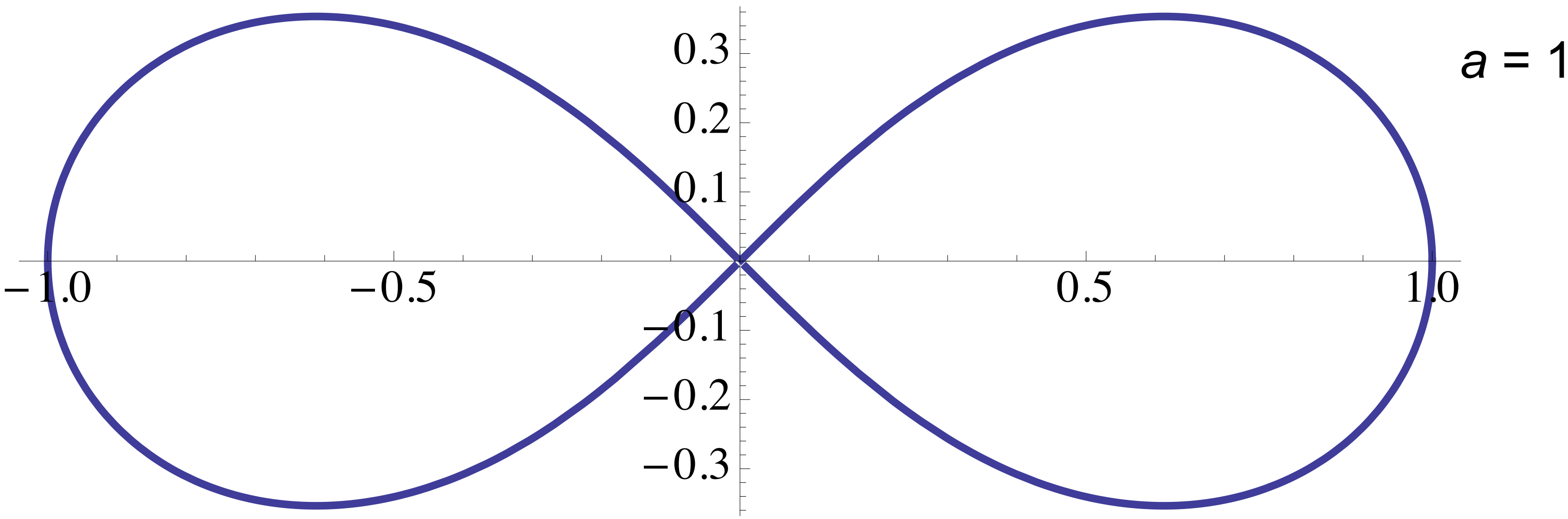
Example:

area of Bernoulli's lemniscate obtained with a Monte Carlo simulation.



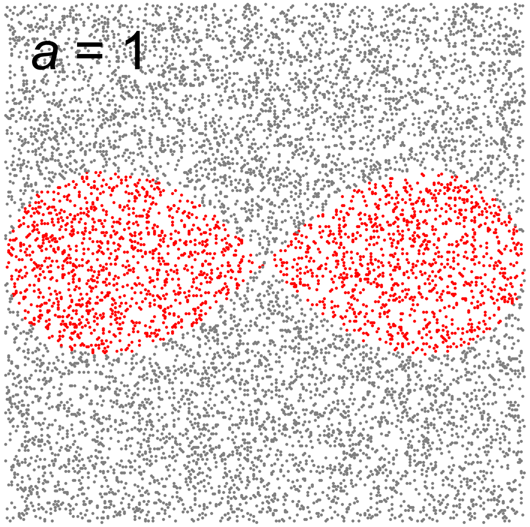
Parametric equation of Bernoulli's lemniscate

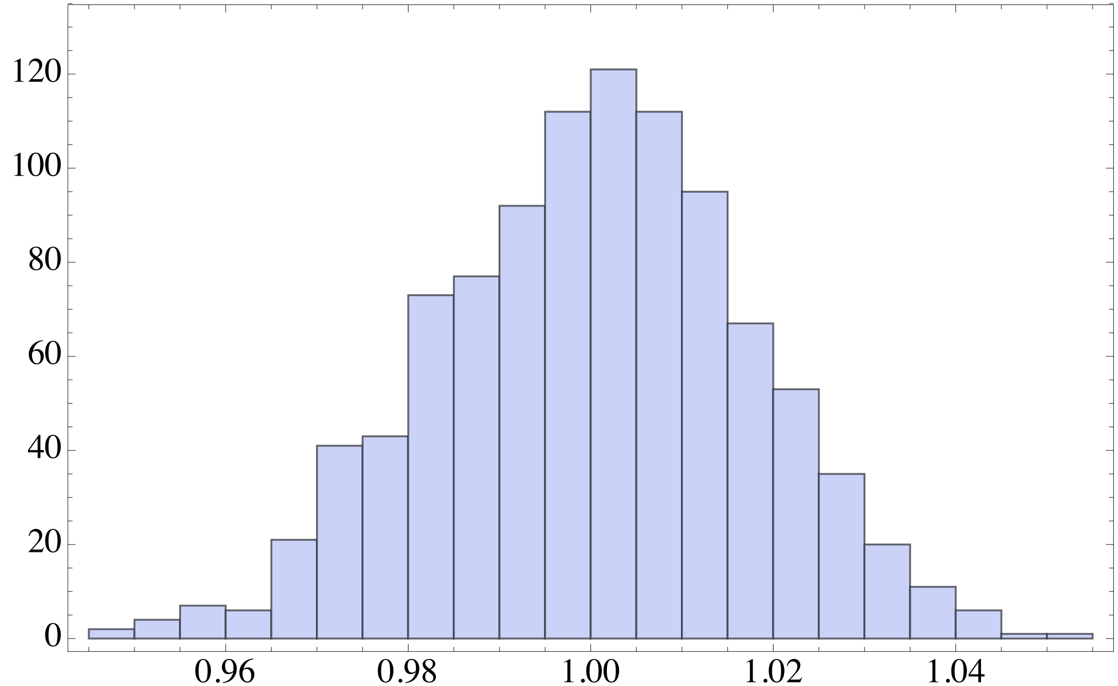
$$r = a\sqrt{\cos 2\theta}$$



What is its area?

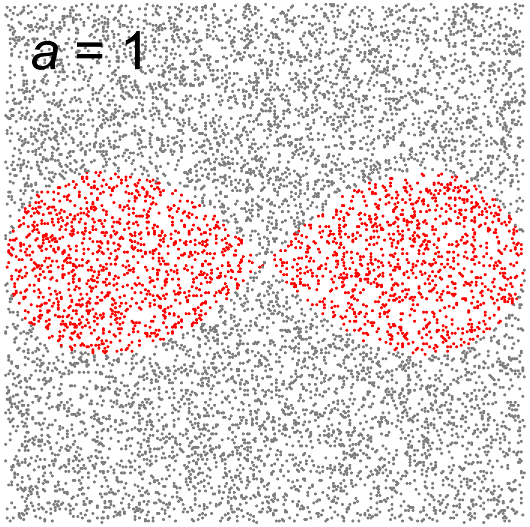
Empirical Monte Carlo distribution of the area estimate

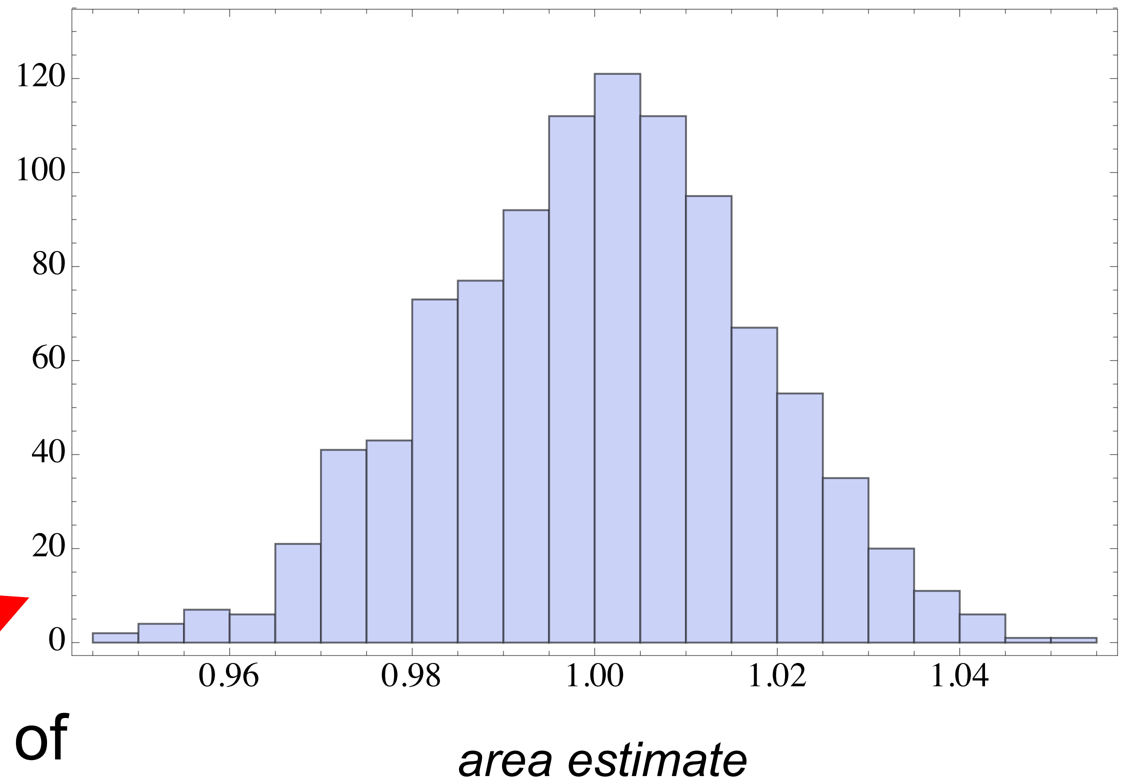




area estimate

Empirical Monte Carlo distribution of the area estimate





a probability distribution of
a mathematical constant???

Frequentist view: this is the distribution of an estimate, it does not make sense to talk of the distribution of a constant.

Bayesian view: while in this case the value to be estimated is unmistakably "true", this is not a real experiment where the model itself is not certain, and probability applies to it as well.

We can start from the Bayesian "reasonable expectation" and use it unambiguously as probability: indeed Cox showed that any reasonable measure of "reasonable expectation" must behave just like common probability.

AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1 JANUARY–FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. COX
The Johns Hopkins University, Baltimore 18, Maryland

Boolean algebra (symbolic logic)

a, **b**, **c** ... propositions (true or false)

Basic operations

OR: **a** v **b**

AND: **a**·**b**

NOT: **~a**

Truth tables

a	b	a v b
T	T	T
T	F	T
F	T	T
F	F	F

a	b	a · b
T	T	T
T	F	F
F	T	F
F	F	F

a	~a
T	F
F	T

Combinations of propositions

$$\sim\sim a = a, \qquad (1)$$

$$a \cdot b = b \cdot a, \qquad (2) \qquad a \vee b = b \vee a, \qquad (2')$$

$$a \cdot a = a, \qquad (3) \qquad a \vee a = a, \qquad (3')$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c = a \; b \cdot c, \qquad (4)$$

$$a \vee (b \vee c) = (a \vee b) \vee c = a \vee b \vee c, \qquad (4')$$

$$\sim(a \cdot b) = \sim a \vee \sim b, \qquad (5)$$

$$\sim(a \vee b) = \sim a \cdot \sim b, \qquad (5')$$

$$a \cdot (a \vee b) = a, \qquad (6) \qquad a \vee (a \cdot b) = a. \qquad (6')$$

The combination rules are not all independent, e.g., consider

$$\sim(\mathbf{a}\cdot\mathbf{b}) = \sim\mathbf{a} \vee \sim\mathbf{b} \qquad \text{and} \qquad \sim(\mathbf{a} \vee \mathbf{b}) = \sim\mathbf{a}\cdot\sim\mathbf{b}$$

When we assume the first, and utilizing $\sim\sim\mathbf{a} = \mathbf{a}$, we can deduce the second:

$$\sim(\mathbf{a} \vee \mathbf{b}) = \sim(\sim\sim\mathbf{a} \vee \sim\sim\mathbf{b}) = \sim\sim(\sim\mathbf{a}\cdot\sim\mathbf{b}) = \sim\mathbf{a}\cdot\sim\mathbf{b}$$

Now let

$$p(\mathbf{b}|\mathbf{a})$$

denote any *measure of reasonable credibility (credibility for short)* of proposition **b** when **a** is known to be true, and let *F* be a function that combines credibilities

$$p(\mathbf{c}\cdot\mathbf{b}|\mathbf{a}) = F[p(\mathbf{c}|\mathbf{b}\cdot\mathbf{a}), p(\mathbf{b}|\mathbf{a})]$$

While *p* is still quite arbitrary, *F* is constrained by the algebra of propositions.

Now we derive a functional equation for *F* from

$$
\begin{aligned}
p(\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}) &= p\left((\mathbf{d}\cdot\mathbf{c})\cdot\mathbf{b}|\mathbf{a}\right) \\
&= F[p(\mathbf{d}\cdot\mathbf{c}|\mathbf{b}\cdot\mathbf{a}), p(\mathbf{b}|\mathbf{a})] \\
&= F\left[F[p(\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a}), p(\mathbf{c}|\mathbf{b}\cdot\mathbf{a})], p(\mathbf{b}|\mathbf{a})\right]
\end{aligned}
$$

and also

$$
\begin{aligned}
p(\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}) &= p\left(\mathbf{d}\cdot(\mathbf{c}\cdot\mathbf{b})|\mathbf{a}\right) \\
&= F[p(\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a}), p(\mathbf{c}\cdot\mathbf{b}|\mathbf{a})] \\
&= F[p(\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a}), F[p(\mathbf{c}|\mathbf{b}\cdot\mathbf{a}), p(\mathbf{b}|\mathbf{a})]]
\end{aligned}
$$

Therefore, setting

$$x = p(\mathbf{d}|\mathbf{c} \cdot \mathbf{b} \cdot \mathbf{a})$$
$$y = p(\mathbf{c}|\mathbf{b} \cdot \mathbf{a})$$
$$z = p(\mathbf{b}|\mathbf{a})$$

we find the functional equation

$$F[x, F[y, z]] = F\left[F[x, y], z\right]$$

It is easy to see by substitution that the equation

$$F[x, F[y, z]] = F\,[F[x, y], z]$$

has the solution

$$C\ f(F[p, q]) = f(p) f(q)$$

where C is an arbitrary constant and f is an arbitrary single-variable function. It can be shown that this is also the general solution if F has continuous second derivatives.

(**homework!**)

Given the arbitrariness of *f*, we take the identity function, so that

$$\begin{aligned} C\ p(\mathbf{c}\cdot\mathbf{b}|\mathbf{a}) &= C\ F[p(\mathbf{c}|\mathbf{b}\cdot\mathbf{a}), p(\mathbf{b}|\mathbf{a})] \\ &= p(\mathbf{c}|\mathbf{b}\cdot\mathbf{a})p(\mathbf{b}|\mathbf{a}) \end{aligned}$$

Then, when we let **c = b**, and we assume that credibility ranges from 0 (no credibility) to 1 (certainty), and therefore

$$p(\mathbf{a}|\mathbf{a}) = p(\mathrm{certainty}) = 1$$

we find

$$C\ p(\mathbf{b}\cdot\mathbf{b}|\mathbf{a}) = C\ p(\mathbf{b}|\mathbf{a}) = p(\mathbf{b}|\mathbf{b}\cdot\mathbf{a})p(\mathbf{b}|\mathbf{a}) = p(\mathbf{b}|\mathbf{a})$$

and therefore $C = 1$.

Thus we have found that credibility satisfies the condition

$$p(\mathbf{c} \cdot \mathbf{b}|\mathbf{a}) = p(\mathbf{c}|\mathbf{b} \cdot \mathbf{a})p(\mathbf{b}|\mathbf{a})$$

however this is not yet enough, because if we took a power law instead of the identity, we could still satisfy all the conditions and find, e.g., a condition like

$$p(\mathbf{c} \cdot \mathbf{b}|\mathbf{a})^m = p(\mathbf{c}|\mathbf{b} \cdot \mathbf{a})^m \, p(\mathbf{b}|\mathbf{a})^m$$

Can we do better?

We have used the properties of logical AND, but not yet those of logical NOT and OR ...

Taking a negated proposition we expect to find the relationship

$$p(\tilde{}\mathbf{b}|\mathbf{a}) = S[p(\mathbf{b}|\mathbf{a})]$$

and therefore we find a functional equation

$$p(\mathbf{b}|\mathbf{a}) = p(\tilde{}\tilde{}\mathbf{b}|\mathbf{a}) = S[S[p(\mathbf{b}|\mathbf{a})]]$$

which, however, is not restrictive enough ...

Now we note that

$$
\begin{aligned}
S[p(\mathbf{c} \vee \mathbf{b}|\mathbf{a})] &= p(\sim(\mathbf{c} \vee \mathbf{b})|\mathbf{a}) = p(\sim\mathbf{c} \cdot \sim \mathbf{b}|\mathbf{a}) \\
&= p(\sim\mathbf{c}|\sim\mathbf{b} \cdot \mathbf{a})p(\sim\mathbf{b}|\mathbf{a}) \\
&= S[p(\mathbf{c}|\sim\mathbf{b} \cdot \mathbf{a})]S[p(\mathbf{b}|\mathbf{a})]
\end{aligned}
$$

and also that

$$
\begin{aligned}
p(\mathbf{c}|\tilde{}\mathbf{b} \cdot \mathbf{a}) &= \frac{p(\mathbf{c} \cdot \tilde{}\mathbf{b}|\mathbf{a})}{p(\tilde{}\mathbf{b}|\mathbf{a})} = \frac{p(\tilde{}\mathbf{b} \cdot \mathbf{c}|\mathbf{a})}{p(\tilde{}\mathbf{b}|\mathbf{a})} \\
&= \frac{p(\tilde{}\mathbf{b}|\mathbf{c} \cdot \mathbf{a})\, p(\mathbf{c}|\mathbf{a})}{p(\tilde{}\mathbf{b}|\mathbf{a})} \\
&= \frac{S[p(\mathbf{b}|\mathbf{c} \cdot \mathbf{a})]\, p(\mathbf{c}|\mathbf{a})}{S[p(\mathbf{b}|\mathbf{a})]}
\end{aligned}
$$

And finally we find

$$p(\mathbf{c}|\tilde{}\mathbf{b}\cdot\mathbf{a})) = \frac{S[p(\mathbf{b}|\mathbf{c}\cdot\mathbf{a})]\; p(\mathbf{c}|\mathbf{a})}{S[p(\mathbf{b}|\mathbf{a})]} = S\left[\frac{S[p(\mathbf{c}\vee\mathbf{b}|\mathbf{a})]}{S[p(\mathbf{b}|\mathbf{a})]}\right]$$

or alternatively

$$S\left[\frac{p(\mathbf{c}\cdot\mathbf{b}|\mathbf{a})}{p(\mathbf{c}|\mathbf{a})}\right]\frac{p(\mathbf{c}|\mathbf{a})}{S[p(\mathbf{b}|\mathbf{a})]} = S\left[\frac{S[p(\mathbf{c}\vee\mathbf{b}|\mathbf{a})]}{S[p(\mathbf{b}|\mathbf{a})]}\right]$$

This results hold for all propositions, and if we let **b = c · d** we find

$$S\left[\frac{p(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})}{p(\mathbf{c}|\mathbf{a})}\right]\frac{p(\mathbf{c}|\mathbf{a})}{S[p(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})]} = S\left[\frac{S[p(\mathbf{c}|\mathbf{a})]}{S[p(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})]}\right]$$

Introducing the auxiliary variables

$$x = p(\mathbf{c}|\mathbf{a}); \quad y = S[p(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})]$$

we obtain a compact form for the functional equation for *S*

$$x\, S\left[\frac{S[y]}{x}\right] = y\, S\left[\frac{S[x]}{y}\right]$$

It is easy to see by substitution that the equation

$$x\, S\left[\frac{S[y]}{x}\right] = y\, S\left[\frac{S[x]}{y}\right]$$

has the solution

$$S[p] = (1 - p^m)^{1/m}$$

It can be shown that this is also the general solution if *S* is twice differentiable.

(**homework!**)

Again, this means that

$$p(\tilde{}\mathbf{b}|\mathbf{a}) = S[p(\mathbf{b}|\mathbf{a})] = (1 - p(\mathbf{b}|\mathbf{a})^m)^{1/m}$$
$$\Rightarrow \quad p(\mathbf{b}|\mathbf{a})^m + p(\tilde{}\mathbf{b}|\mathbf{a})^m = 1$$

and again, whatever the value of *m*, credibility satisfies the usual probability rule. Since the choice of *m* is conventional we take *m* = 1.

Summarizing, we have the following collection of assumptions and rules:

$$
\begin{aligned}
p(\text{certainty}) &= 1 \\
p(\text{impossibility}) &= 0 \\
p(\mathbf{b}|\mathbf{a}) + p(\tilde{\ }\mathbf{b}|\mathbf{a}) &= 1 \\
p(\mathbf{c} \cdot \mathbf{b}|\mathbf{a}) &= p(\mathbf{c}|\mathbf{b} \cdot \mathbf{a})p(\mathbf{b}|\mathbf{a})
\end{aligned}
$$

and from these all the usual rules of probability follow.

Therefore we can take probabilities as measures of credibility.

Probability in Quantum Mechanics

Probability in QM has a fundamental role. This is highlighted by the *Bell's inequalities*.

Consider two identical objects, i.e., objects that share the same general properties, and assume that

1. **the properties are multivalued**, and that their values are predetermined and are not influenced by measurement (an example of such a property could be the color of a set of "identical" balls: the color can take different values such as red or blue);

2. **locality holds**, i.e., when the objects are spatially separated, the determination of the value of the property of one object does not influence the other one.

We assume that these two objects share 3 different binary properties, A, B, and C, which can take the values 0 and 1, and we experiment with them.

We enclose each of them in an opaque box, so that we cannot observe them directly, and because of locality, we know that a measurement on object 1 cannot influence a measurement on object 2.

We can also consider probabilities for each particular property value or combination of property values, such as

$$P(A_1 = B_2)$$

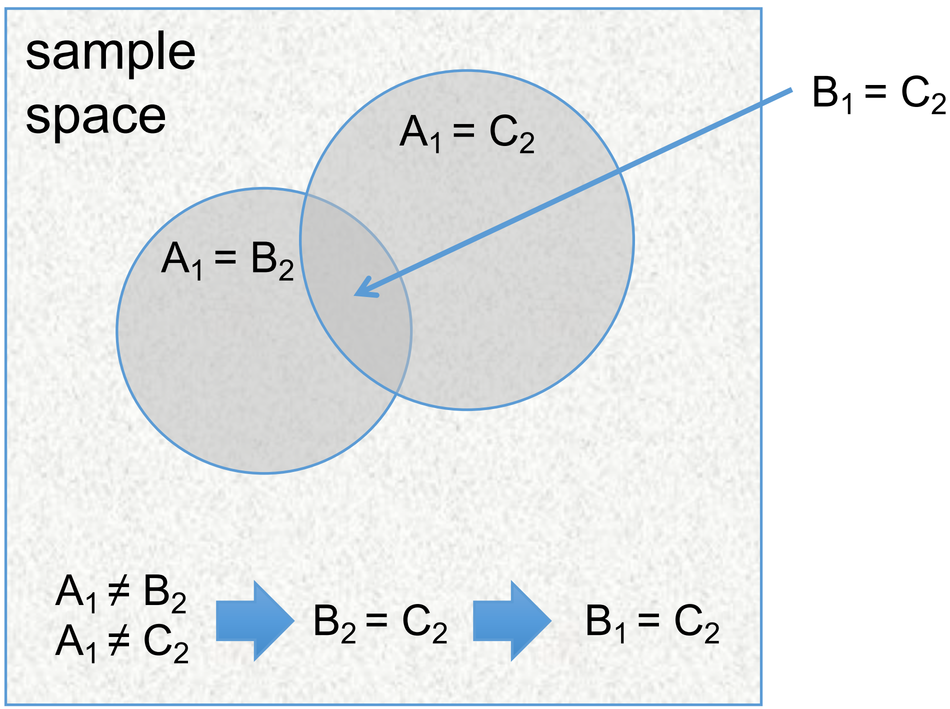
which is that the probability that the value of property A for object 1 is the same as the value of property B for object 2, i.e., that they are both 0 or both 1.

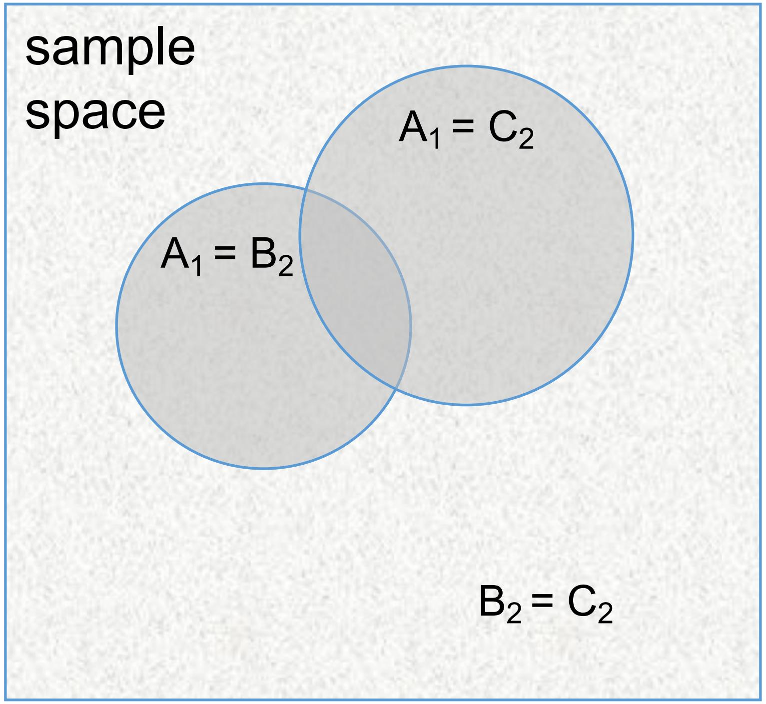
If we assume that the two objects share exactly the same property values for each property, then

$$P(A_1 = A_2) = P(B_1 = B_2) = P(C_1 = C_2) = 1$$

From this assumption we deduce that

$$P(A_1 = B_2) + P(A_1 = C_2) + P(B_1 = C_2) \geq 1$$





The overlap region shared by the two regions marked $A_1 = B_2$ and $A_1 = C_2$ clearly contains events such that $B_2 = C_2$.

The same happens in the external region where $A_1 \neq B_2$ and $A_1 \neq C_2$, because of the binary nature of the properties A, B and C.

However $B_1 = B_2$ by assumption, therefore $B_2 = C_2$ implies $B_1 = C_1$, and finally we obtain the inequality

$$P(A_1 = B_2) + P(A_1 = C_2) + P(B_1 = C_2) \geq 1$$

The inequality is violated by QM

Consider two two-level systems in an entangled state

$$|\phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

and the two-valued properties A, B, and C obtained by projecting the entangled state onto rotated orthogonal bases

$$A : \begin{cases} |a_0\rangle = |0\rangle \\ |a_1\rangle = |1\rangle \end{cases}$$

$$B : \begin{cases} |b_0\rangle = \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \\ |b_1\rangle = \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle \end{cases}$$

$$C : \begin{cases} |c_0\rangle = \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle \\ |c_1\rangle = \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle \end{cases}$$

It is easy to check that these are orthogonal bases and that

$$|\phi^+\rangle = \frac{|a_0 a_0\rangle + |a_1 a_1\rangle}{\sqrt{2}} = \frac{|b_0 b_0\rangle + |b_1 b_1\rangle}{\sqrt{2}} = \frac{|c_0 c_0\rangle + |c_1 c_1\rangle}{\sqrt{2}}$$

so that

$$\langle a_0 a_0 | \phi^+\rangle = \langle a_1 a_1 | \phi^+\rangle = \frac{1}{\sqrt{2}}$$

and therefore

$$P(A_1 = 0 \text{ and } A_2 = 0) = P(A_1 = 1 \text{ and } A_2 = 1) = 1/2$$

$$\Downarrow$$

$$P(A_1 = A_2) = 1$$

Similar calculations hold for properties B and C, and thus

$$P(A_1 = A_2) = P(B_1 = B_2) = P(C_1 = C_2) = 1$$

holds.

However, when we write the vectors A as linear combinations of the B's

$$\begin{aligned} |a_0\rangle &= \tfrac{1}{2}|b_0\rangle + \tfrac{\sqrt{3}}{2}|b_1\rangle \\ |a_1\rangle &= \tfrac{\sqrt{3}}{2}|b_0\rangle - \tfrac{1}{2}|b_1\rangle \end{aligned}$$

we find

$$|\phi^+\rangle = \frac{|a_0\rangle(|b_0\rangle + \sqrt{3}|b_1\rangle) + |a_1\rangle(\sqrt{3}|b_0\rangle - |b_1\rangle)}{2\sqrt{2}}$$

This means that

$$\langle a_0 b_0 | \phi^+ \rangle = \langle a_1 b_1 | \phi^+ \rangle = \frac{1}{2\sqrt{2}}$$

and therefore

$$P(A_1 = 0 \text{ and } B_1 = 0) = P(A_1 = 1 \text{ and } B_2 = 1) = 1/8$$

$$\Downarrow$$

$$P(A_1 = B_2) = 1/4$$

this can be replicated to find

$$P(A_1 = C_2) = P(B_1 = C_2) = 1/4$$

and therefore

$$P(A_1 = B_2) + P(A_1 = C_2) + P(B_1 = C_2) = 3/4 < 1$$

which violates Bell's inequality

The inequality presented here is just one of many versions that have been produced since Bell's discovery in 1964.

The violation is surprising, it is confirmed by experiments, and it indicates that there is something amiss in our understanding of the physical world.

It is still unclear how all this comes about, what is the origin of the violation.

Probabilities and inference: the case of the Phoenix virus

Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements

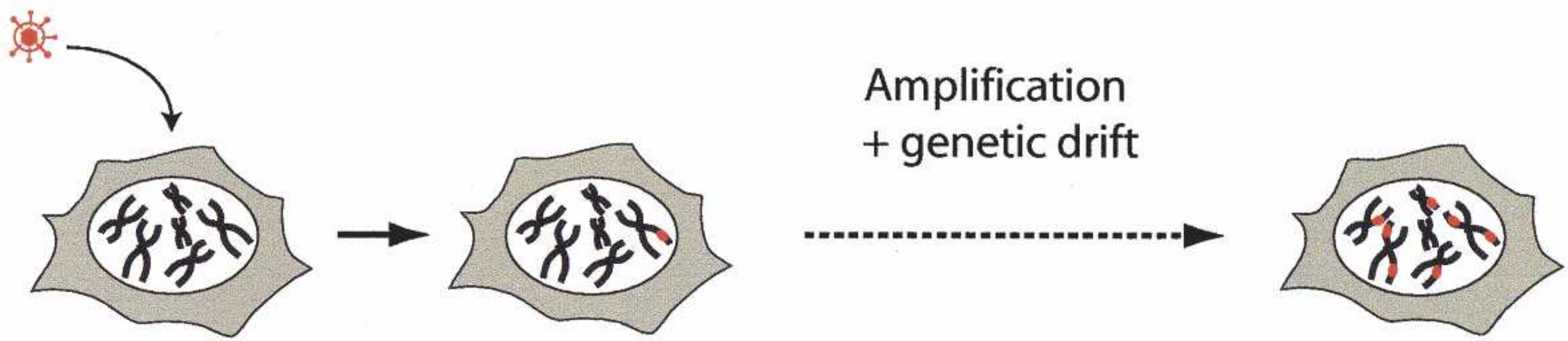
Marie Dewannieux,[1,3] Francis Harper,[2,4] Aurélien Richaud,[1,4] Claire Letzelter,[1] David Ribet,[1] Gérard Pierron,[2] and Thierry Heidmann[1,5]

[1]Unité des Rétrovirus Endogènes et Eléments Rétroïdes des Eucaryotes Supérieurs, UMR 8122 CNRS, Institut Gustave Roussy, 94805 Villejuif Cedex, France; [2]Laboratoire de Réplication de l'ADN et Ultrastructure du Noyau, UPR1983 Institut André Lwoff, 94801 Villejuif Cedex, France

Human Endogenous Retroviruses are expected to be the remnants of ancestral infections of primates by active retroviruses that have thereafter been transmitted in a Mendelian fashion. Here, we derived in silico the sequence of the putative ancestral "progenitor" element of one of the most recently amplified family—the HERV-K family—and constructed it. This element, *Phoenix*, produces viral particles that disclose all of the structural and functional properties of a bona-fide retrovirus, can infect mammalian, including human, cells, and integrate with the exact signature of the presently found endogenous HERV-K progeny. We also show that this element amplifies via an extracellular pathway involving reinfection, at variance with the non-LTR-retrotransposons (LINEs, SINEs) or LTR-retrotransposons, thus recapitulating ex vivo the molecular events responsible for its dissemination in the host genomes. We also show that in vitro recombinations among present-day human *HERV–K* (also known as *ERVK*) loci can similarly generate functional HERV-K elements, indicating that human cells still have the potential to produce infectious retroviruses.

1548 Genome Research
www.genome.org

16:1548–1556 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org



Phoenix, the ancestral HERV-K(HML2) retrovirus

To construct a consensus HERV-K(HML2) provirus, we assembled all of the complete copies of the 9.4-kb proviruses that are human specific (excluding those with the 292-nt deletion at the beginning of the *env* gene) and aligned their nucleotide sequence to generate the consensus in silico, taking for each position the most frequent nucleotide.

* provirus = virus genome integrated into DNA of host cell

Name	GenBank Accession Number	FirstPosition	Last position	Orientation
K104	AC116309	123,567	114,122	—
K108-1	AC072054	47,417	37,947	—
K108-2	AC072054	38,914	29,443	—
K109	AC055116	139,321	148,740	+
K113	AY037928	1	9,472	+
K115	AY037929	1	9,463	+
	Y178333	1	8,629	+
	AP000776	101,084	110,549	+
	AC025420	37,159	46,615	+

This table provides the GenBank coordinates of the human endogenous HERV-K copies used to generate the *Phoenix* provirus.

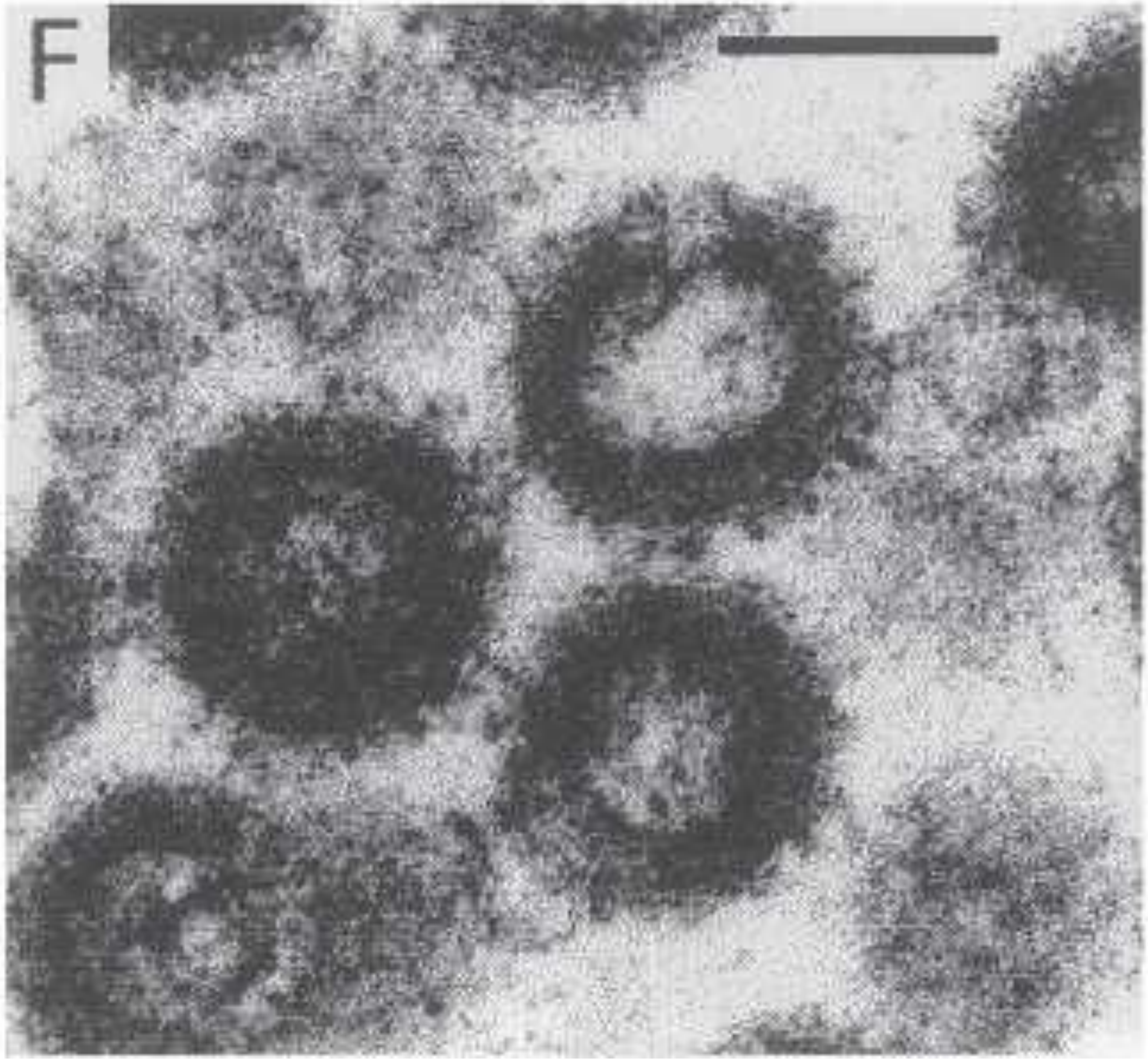


Image of representative particles obtained after transfection with an expression vector for the *Phoenix pro* mutant. Scale bar 100 nm.

An extremely short history of early Bayesianism

- Rev. Thomas Bayes discovered an early form of Bayes' theorem (second half of 18^{th} century)
- Price discovered the theorem inside Bayes' unpublished notes (end 18^{th} century)
- Laplace reinvented a version of the theorem and later expanded it after studying the Bayes' notes (around 1800)
- Laplace successfully applied the theorem to many experimental data analysis problems (until about 1820)
- Laplace was sometimes ridiculed by people who did not understand some of his approaches
- Laplace discovered the basic version of the Central Limit Theorem and in his later life he abandoned the Bayes theorem in favour of frequency-based methods (until about 1830)
- After the death of Laplace, Bayes' theorem was nearly forgotten and cornered to the darkest parts of statistics (crossing the desert ...)

Conditional probabilities and Bayes' Theorem

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

Joint probability and conditional probabilities

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' theorem: a purely logical statement

$$P(H|D) = \frac{P(D|H)}{P(D)}P(H)$$

Bayes' theorem again: now as an inferential statement

Likelihood

Prior distribution

$$P(H|D) = \frac{P(D|H)}{P(D)} P(H)$$

Posterior distribution

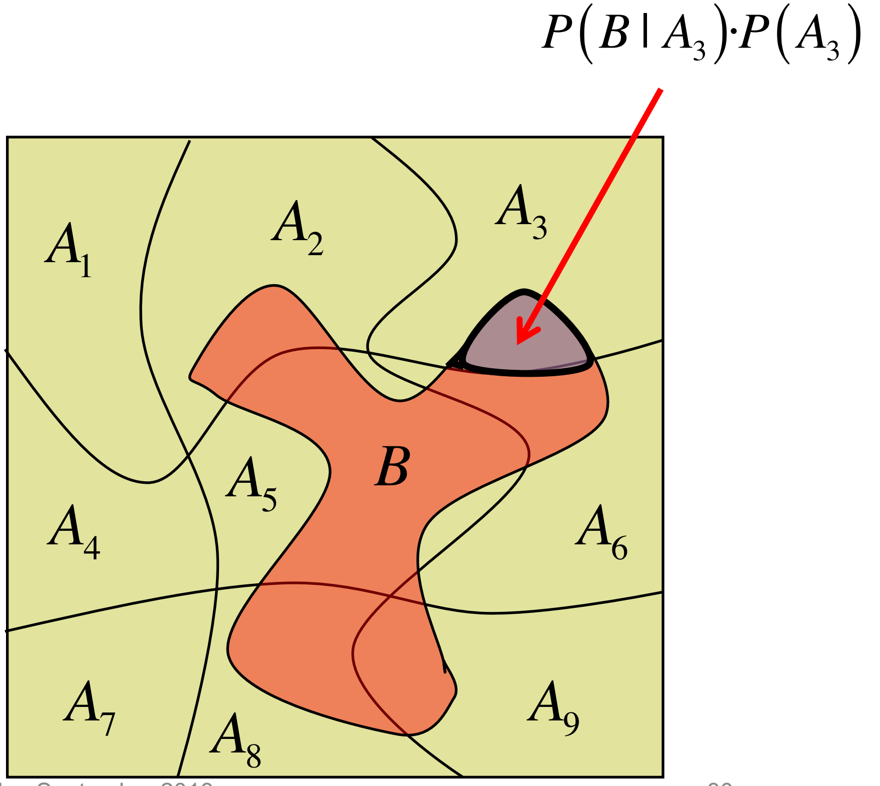
Evidence

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$P(A_k \mid B) = \frac{P(B \mid A_k) \cdot P(A_k)}{P(B)} \qquad k = 1, \ldots, N$$

if the events A_k are mutually exclusive, and they fill the universe

$$P(B) = \sum_{k=1}^{N} P(B \mid A_k) \cdot P(A_k)$$



$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$+$$

$$P(B) = \sum_{k=1}^{N} P(B \mid A_k) \cdot P(A_k)$$

$$\Downarrow$$

$$P(A_k \mid B) = \frac{P(B \mid A_k) \cdot P(A_k)}{\sum_{k=1}^{N} P(B \mid A_k) \cdot P(A_k)}$$

$$P(H_k|D) = \frac{P(D|H_k)}{\sum_j P(D|H_j)P(H_j)} P(H_k)$$

MAP estimates

Check the webpage:

http://wwwusers.ts.infn.it/~milotti/Didattica/Bayes/Bayes.html

References:

- L. Mahadevan and E. H. Yong, "Probability, physics, and the coin toss", Phys. Today, July 2011, pp. 66-67
- P. Diaconis, S. Holmes and R. Montgomery, "Dynamical bias in the coin toss", *SIAM Rev.* **49** (2007) 211
- G. D'Agostini, Rep. Prog. Phys. **66** (2003) 1383
- V. Dose, Rep. Prog. Phys. **66** (2003) 1421
- W. C. Thompson and E. L. Schulman, Law and Human Behavior **11** (1987) 167
- M. Botje, lecture notes: http://www.nikhef.nl/~h24/bayes/