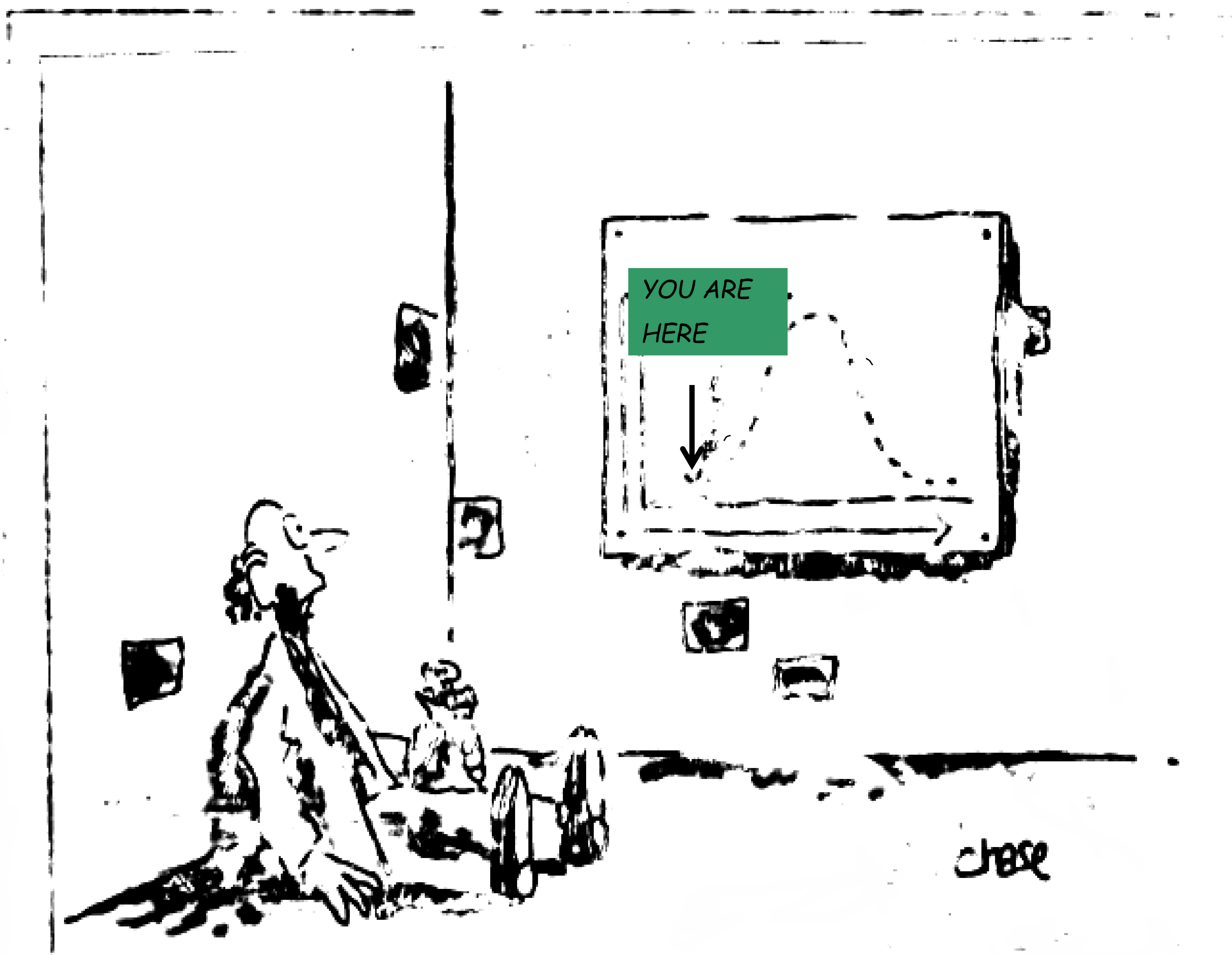
Pitfalls and Paradoxes in the History of Probability Theory

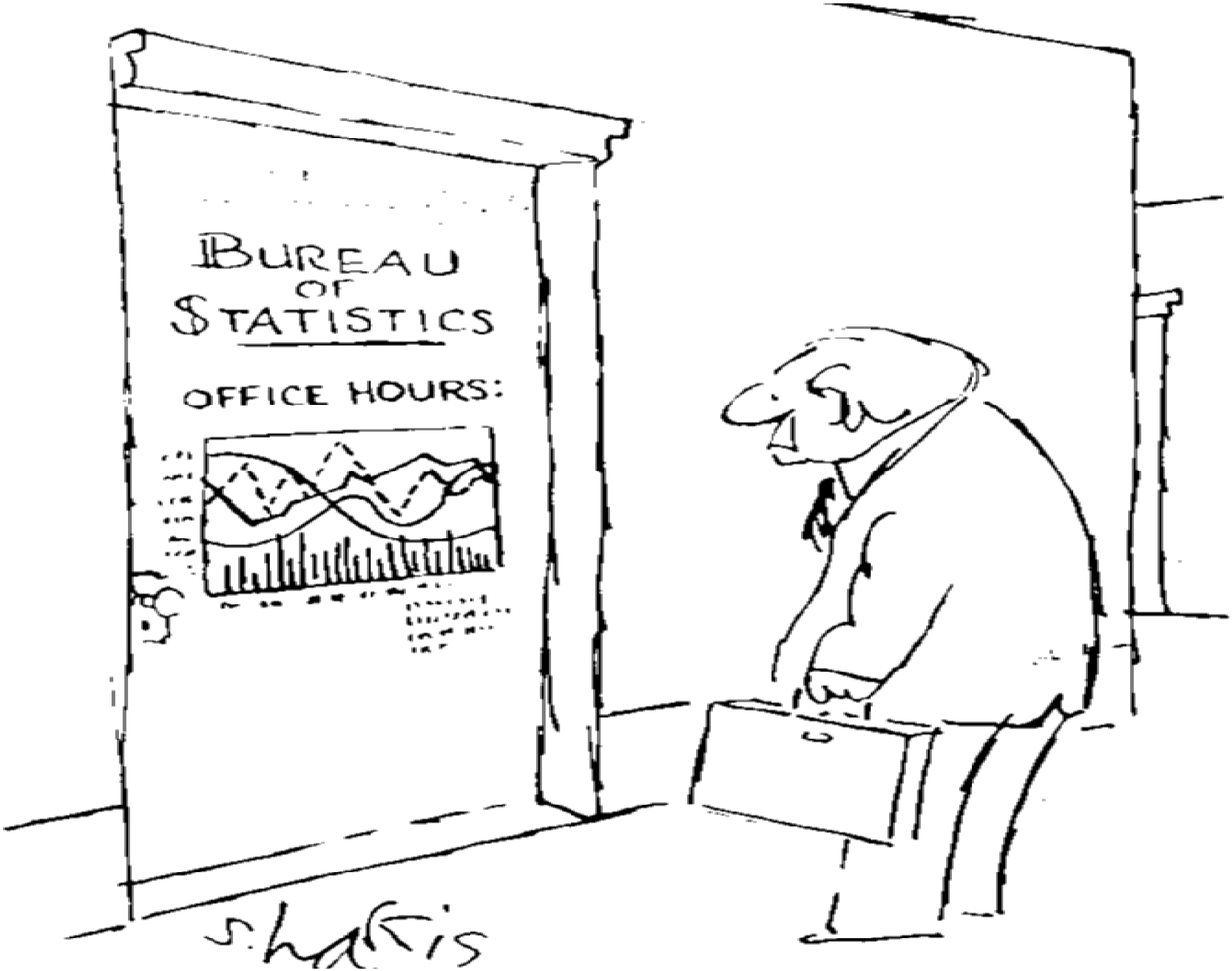
(predicting the unpredictable)

Michael Shlesinger

Office of Naval Research

Gdansk July 12, 2018


YOU ARE
HERE
chase

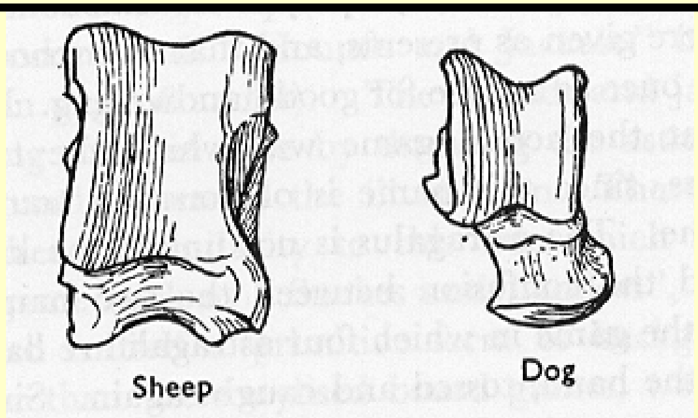

BUREAU OF STATISTICS
OFFICE HOURS:
s.harris

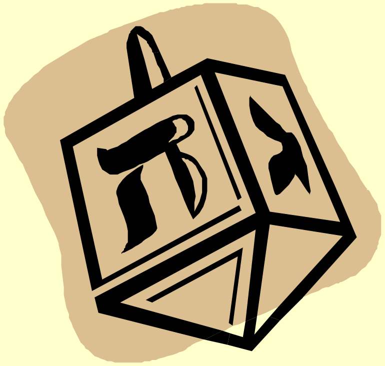
If you have 5 dogs, 3 will be asleep

Throwing the bones
Astrogali (knucklebones)

From Games, Gods, and Gambling, FN David

Al-zar (dice) becomes hazard





upper bone	4
opposite side	3
flat lateral side	1
opposite side	6

4 stable sides

Greeks throws (4 bones)

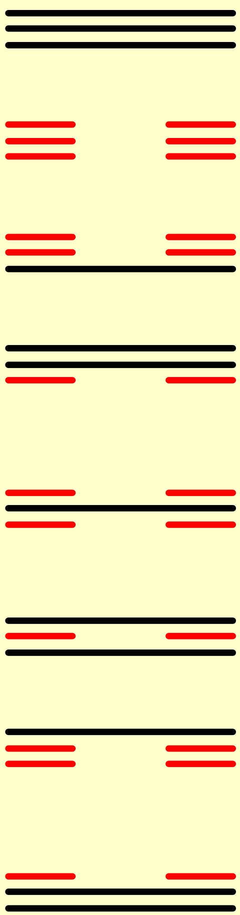
1,3,4,6	Venus
1,1,1,1	Going to the Dogs

Romans throws (5 bones)

1,3,3,4,4	Zeus

I CHING

8 trigrams/ 64 pairs of trigrams



- heaven/strength
- earth/weakness
- activity/thunder
- bending/wind
- pit/water/danger
- Brightness/fire/elegance
- Stop/mountain/firmness
- Pleasure/joyful/collect water

Too many outcomes to check for the persistence of statistical ratios?

An oracle, not a mathematical exercise.

Galileo and Newton

Throw 3 dice. More likely to get a 10 then a 9. But there are six ways to make either number

10	# ways		9	# ways
1. 6 2 2	3	1.	3 3 3	1
2. 5 2 3	6	2.	3 4 2	6
3. 4 2 4	3	3.	3 5 1	6
4. 6 3 1	6	4.	6 2 1	6
5. 4 3 3	3	5.	5 2 2	3
6. 5 4 1	6	6.	4 4 1	3
	27 ways			25 ways

6X6X6=216 total outcomes

Prob (10) =27/216 **prob (9) =25/216**

Would you make this bet, of a 10 before a 9, if you cannot afford to lose?

(risk-benefit analysis)

Yes

No

D. Bernoulli ------------------D’Alembert

controversy

1 in 200 die from smallpox variolation (a type of innoculation)

14 in 200 die in invariolated population

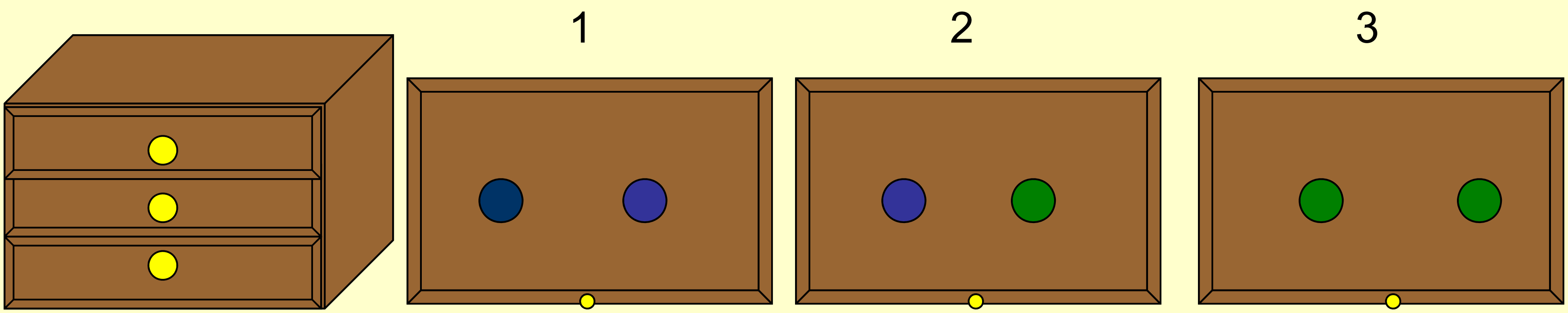
Suppose 200 fatal diseases and 200 innoculations

Each innoculation has probability 1/200 to kill you.

If you elect to take all 200 innoculations does that improve your odds to live or to die?

D’Alembert wrote, “*to enjoy the present and not trouble oneself about the future, is common logic, a logic half good, half bad*”

Talmudic Paradox



Without looking, choose a draw and pick a coin at random, then look to see it is

What is the probability that the other coin in the draw is also

Solution #1: You either picked draw #1 or draw #2, each has probability ½

so probability that the other coin is = ½.

Solution #2:

→ chose draw 1, coin on left, other coin is

→ chose draw 1, coin on right, other coin is

→ chose draw 2, coin on left, other coin is

chose draw 2, coin on right

chose draw 3, coin on left

chose draw 3, coin on right

Probability other coin is =2/3

Bayes' Theorem (1764)

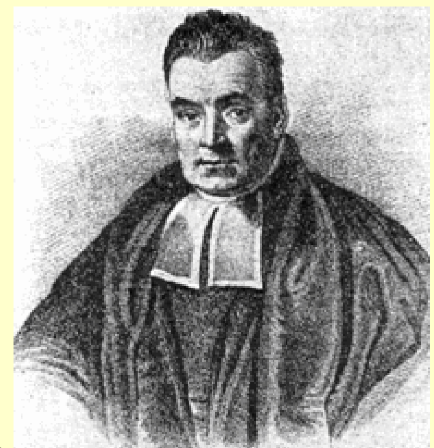
Essay Towards Solving a Problem in the Doctrine of Chances

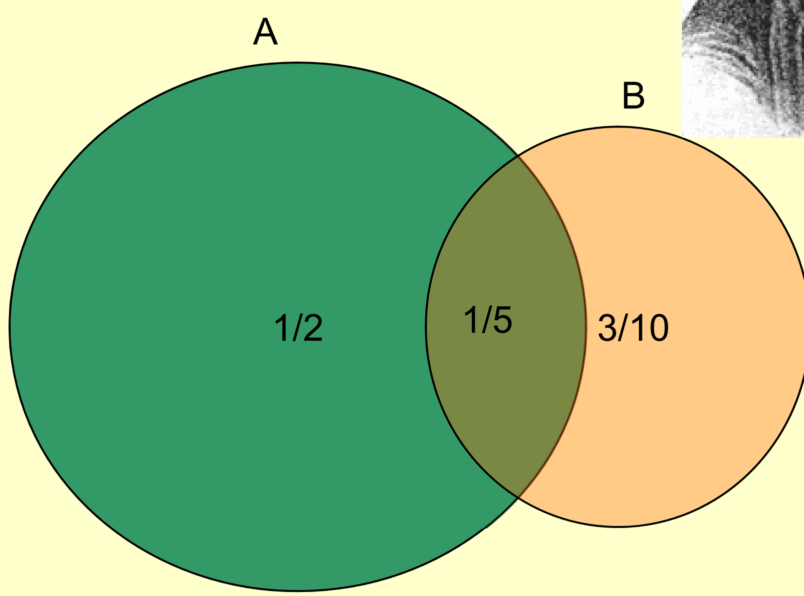
Thomas Bayes 1702-1761

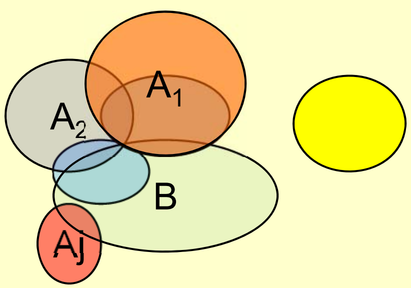
1/5 2/5 1/2 2/7 7/10

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

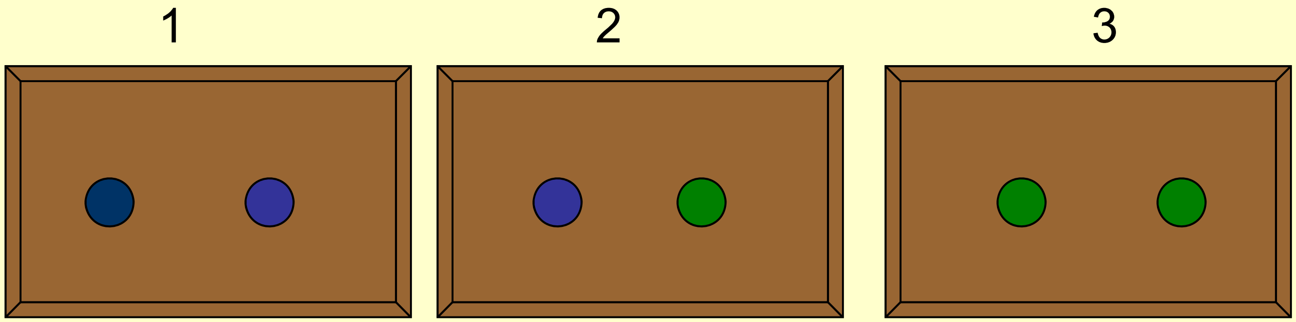
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$







$$P(A_i \cap B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$



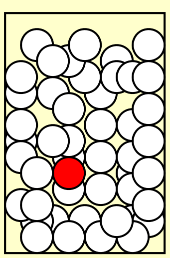
$$P(draw\,1|\,blue) = \frac{P(blue|draw\,1)\,P(draw\,1)}{\sum_{i=1}^{3} P(blue|draw\,i)\,P(draw\,i)}$$

$$P(draw\,1|blue) = \frac{1\bullet 1/3}{1\bullet 1/3 + 1/2\bullet 1/3} = 2/3$$

$$P(draw\,2|blue) = \frac{1/2\bullet 1/3}{1\bullet 1/3 + 1/2\bullet 1/3} = 1/3$$

Employing **Bayes' Theorem** 1764

Urn with N white marbles and 1 red marble



$$P(W_1|R_2) = \frac{P(R_2|W_1)P(W_1)}{P(R_2|W_1)P(W_1) + P(R_2|R_1)P(R_1)} = 1$$

0

statistical inference

Chevalier de Mere - Grande Scandale

(real name Antonie Gombaud)

Throw a 6 with one die, …, advantage in undertaking to do it in 4 throws

(odds in your favor)

Throw 2 sixes in 24 throws, …, a disadvantage

(odds against you)

But 4/6 = 24/36

The start of the famous Pascal-Fermat letters 1654

SOLUTION

Probability (no six in 1 throw) = 5/6

Prob(no 6's in N throws)= $(5/6)^N$

P_6=Prob(at least one 6) = 1-$(5/6)^N$

Probability (no [6,6] in throw of 2 dice) =35/36

Prob(no [6,6] in N throws) = $(35/36)^N$

$P_{[6,6]}$ =Prob(at least one pair of 6's) =1- $(35/36)^N$

N=4 , P_6 = 0.5177

N=24, $P_{[6,6]}$ = 0.4914

More Pascal and Fermat (1654)

A, B, C play a game. A needs 1 win, B and C need 2 wins.
How should the stake be divided if the game is not continued?

Answer: A:B:C=17:5:5

Winning sequences for B
BB prob 1/3X1/3=3/27
BCB prob 1/3X1/3X1/3 =1/27
CBB prob =1/27
Total =5/27

At most 3 more plays with 3^3=27
Possible outcomes even though play could end immediately with a win by A

The 5 winning sequences for player B

BBB
BCB
BBC
CBB
BBA

Bernoulli Trials

m successes in n trials

$$P_m = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}$$

Flip a coin twice, probability no head,

n=2, m=0

HH
HT
TH
TT

$$= \frac{2!}{2!}\left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Probability of not seeing an H =1/4

D'Alembert wrote Probability of not seeing an H =1/3

H

TH

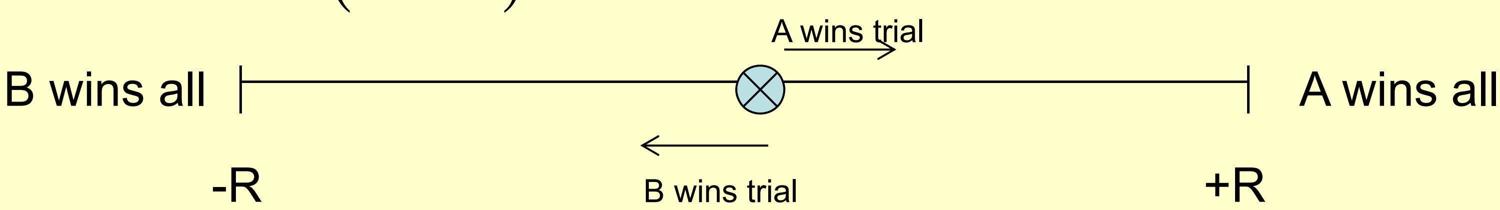
TT

Following Pascal-Fermat (1654), Huygens (1657) wrote a book on the P-F correspondence and presented a problem set with answers.

These problems were solved in Jacob Bernoulli's – *Ars Conjectandi*, published posthumously in 1713 with the help of his nephew, Nicholas Bernoulli. J. Bernoulli also treats the theory of permutations and combinations and applications to civil, moral, and economic questions.

The book was reviewed by Jacob's brother John who called it – *a monster which bears my brother's name.*

$$P_m = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \quad \text{Bernoulli trials}$$



- Bernoulli numbers
- Duration of play N (first, *first* passage time)

$<N(R)> \sim R^2$ J. Bernoulli 1713

$<R^2(N)> \sim N$ A. Einstein 1905

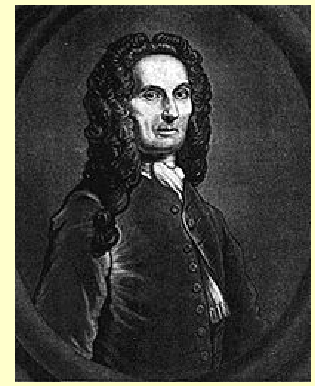
Note: probability of m success in n trials is zero if m>n, a built in diffusion front.

Abraham de Moivre (1667-1754)

(predicted his own date of death seven years in advance)

Imprisoned in France, (edit of Nantes revoked 1685) moved to England, but no real job, despite being the mathematical equal of Jacob Bernoulli and a close friend of Newton.

Worked out of Slaughter's Coffee House (site of 1748 chess championship)
Doctrine of Chances (1718 dedicated to Newton, last edition 1756)
Annuities Upon Lives (1724)) first actuarial book



$$\left(\cos\theta + i\sin\theta\right)^n = \cos n\theta + i\sin n\theta$$

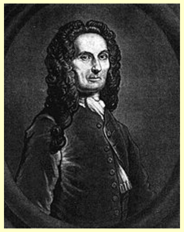
introduced

Introduced (1756 edition) central limit theorem for Bernoulli trials
Introduced generating functions to solve difference equations
In duration of play derives

Sterling approximation $\sqrt{2\pi}$

$$\log(n-1)! = \left(n-\frac{1}{2}\right)\log n - n + \log B + \sum_{r=1}^{\infty}\frac{(-1)^{r-1}B_r}{2r(2r-1)n^{2r-1}}$$

$$B = 1 - \frac{1}{12} + \frac{1}{360} - \Lambda$$



DeMoivre

(1756) 3rd edition of

The Doctrine of Chances

The First Limit Theorem

- For Bernoulli trials using Stirling's approximation, *continuum limit*

" I conclude that if m or n/2 be a quantity infinitely great, then the hyperbolic logarithm of the ratio, which is a term distant from the middle by an interval l, has to the middle term, is -2 l l / n "

Modern notation

$$P_\lambda(n) \sim \exp\left(-2\lambda^2 / n\right)$$

The Gaussian

Johann Carl Friedrich Gauss

THE ANALYST;

MATHEMATICAL MUSEUM.

CONTAINING

NEW ELUCIDATIONS, DISCOVERIES AND IMPROVEMENTS,

IN VARIOUS BRANCHES OF THE

MATHEMATICS,

WITH COLLECTIONS OF QUESTIONS

PROPOSED AND RESOLVED

BY INGENIOUS CORRESPONDENTS

VOL. I

UTILE DULCE.

PHILADELPHIA:

PUBLISHED BY WILLIAM P. FARRAND AND CO.
FRY AND KAMMERER, PRINTERS

1808.

See American Science in the Age of Jefferson, Green 1984

Adrain 1808 (survey meausrements)

Gauss 1809 (astronomical measurements)

ARTICLE XIV.

Research concerning the probabilities of the errors which happen in making observations, &c.

BY ROBERT ADRAIN.

The question which I propose to resolve is this: Supposing AB to be the true value of any quantity, of which the measure by observation or experiment is A*b*, the error being B*b*; what is the expression of the probability that the error B*b* happens in measuring AB?

A ——— *b* B *b*

Let AB, BC, &c. be several successive distances of which the values by measure are A*b*, *bc*, &c. the whole error being C*c*: now supposing the measures A*b*, *bc*, to be given and also the whole error C*c*, we assume as an evident principle that the most probable distances AB, BC are proportional to the measures A*b*, *bc*; and therefore the errors belonging to AB, BC are proportional to their lengths, or to their measured values A*b*, *bc*. If therefore we represent the values of AB, BC, or of their measures A*b*, *bc*, by a, b, the whole error C*c* by E, and the errors of the measures A*b*, *bc* by x, y, we must, for the greatest probability, have the equation

A · B*b* ——— C *c*

$$\frac{x}{a}=\frac{y}{b}.$$

Let X and Y be similar functions of a, x, and of b, y, expressing the probabilities that the errors x, y, happen in the distances a, b; and, by the fundamental principle of the doctrine of chance, the probability that both these errors happen together will be expressed by the product XY. If now we were to determine the values of x and y from the equations $x+y=$E, and XY$=$ maximum, we ought evidently to arrive at the equation $\frac{x}{a}=\frac{y}{b}$: and since x and y are rational functions of the simplest order possible of a, b and E, we ought to arrive at the equation $\frac{x}{a}=\frac{y}{b}$ without the intervention of roots, in other words by simple equations; or, which amounts to the same thing in effect, if there be several forms of X and Y that will fulfil the required condition, we must choose the simplest possible, as having the greatest possible degree of probability.

Let X′, Y′, be the logarithms of X and Y, to any base or modulus *e*: and when XY$=$*max*. its logarithm X′+Y′$=$*max*. and therefore $\dot{X}'+\dot{Y}'=0$, which fluxional equation we may express by $X''\dot{x}+Y''\dot{y}=0$; for as X′ involves only the variable quantity x, its fluxion $\dot{X}'$ will evidently involve only the fluxion of x; in like manner the fluxion of Y′ may be expressed by $Y''\dot{y}$; and from the equation $X''\dot{x}+Y''\dot{y}=0$ we have $X''\dot{x}=-Y''\dot{y}$: but since $x+y=$E we have also $\dot{x}+\dot{y}=0$, and $\dot{x}=-\dot{y}$ by which dividing the equation $X''\dot{x}=-Y''\dot{y}$, we obtain $X''=Y''$.

Now this equation ought to be equivalent to $\frac{x}{a}=\frac{y}{b}$; and this circumstance is effected in the simplest manner possible, by assuming $X''=\frac{mx}{a}$, and $Y''=\frac{my}{b}$; m being any fixed number which the question may require.

Since therefore $X''=\frac{mx}{a}$, we have $X''\dot{x}=\dot{X}'=\frac{mx\dot{x}}{a}$, and taking the fluent, we have $X'=a'+\frac{mx^2}{2a}$.

The constant quantity a' being either absolute, or some function of the distance a.

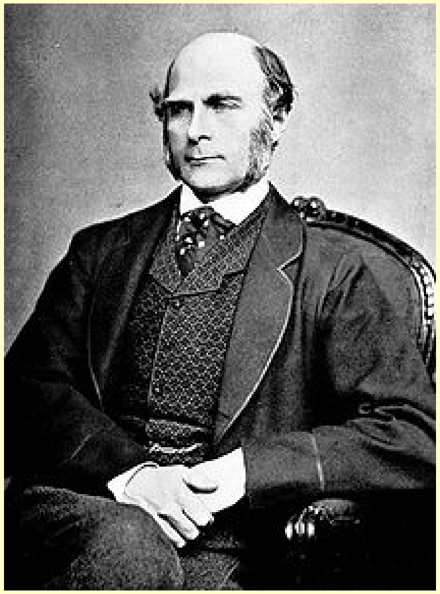
We have discovered therefore, that the logarithm of the probability that the error x happens in the distance a is expressed by $a'+\frac{mx^2}{2a}=X'$, and consequently the probability itself is

$$X=e^{X'}=e^{\left(a'+\frac{mx^2}{2a}\right)}.$$

Such is the formula by which the probabilities of different errors may be compared, when the values of the determinate quantities e, a', and m are properly adjusted. If this probability of the error x be denoted by u, the ordinate of a curve to the abscissa x, we shall have $u=e^{\left(a'+\frac{mx^2}{2a}\right)}$, which is the general equation of *the curve of probability*.

F. Galton

"I know scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expresses by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had know of it, …, The larger the mob and the greater the apparent anarchy, the more perfect its sway."



(~ 1889)

The First Poisson Process

Horse Kicking as Bernoulli Trials (1894)

$$P_k(n) = \text{probability of k successes in n trials}$$

$$= \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k} = \frac{n(n-1)\Lambda\ (n-k+1)}{k!} p^k (1-p)^{n-k}$$

Poisson noted that if

$$p \to 0, \quad n \to \infty, \quad \text{with } np \to \lambda$$

$$P_k(n) \approx \frac{n^k p^k}{k!} \exp(-np)$$

$$\approx \frac{\lambda^k}{k!} e^{-\lambda}$$

German Calvary, 14 Corps, 20 years of data of horse kicking fatalities

14X20 = 280 data points

	1	2	…	14
1875				
:				
1894				

		Poisson λ=0.7
140 cases	no deaths	139 cases
91 cases	1 death	97 cases
32 cases	2 deaths	34 cases

W. S. Gosset ("Student") counted yeast cells/unit volume for Guinness brewing and fit with the Poisson (1904).

Laplace

put the integral into probability

Expressed many problems in the form of picking tickets from an urn with R red and W white tickets, R/W fixed and R and W →infinity

Given: Choose tickets

p where white

q where red

What is the probability that that in choosing more tickets that

m will be white

n will be red

$$\frac{\int_0^1 x^{p+m}(1-x)^{q+n}\,dx}{\int_0^1 x^{p}(1-x)^{q}\,dx}$$

Evaluating these integrals led Laplace to calculate many of the integrals seen in introductory calculus books

Laplace (1749-1827)

Theorie analytique des probabilities

1st edition dedicated to Napolean-le-Grand/fidele sujet, Laplace
2nd edition, no dedication

Includes

- **Generating functions**
- **Laplace transforms**
- **Method of steepest descent**

$$\int_0^\infty \cos(rx)\, e^{-a^2x^2}\, dx = \frac{\sqrt{\pi}}{2a} e^{-\frac{r^2}{4a^2}}$$

$$\int_0^\infty \frac{\sin(rx)}{x}\, dx = \frac{\pi}{2}$$

$$\int_0^\infty \frac{\cos(ax)}{1+x^2}\, dx = \frac{\pi}{2} e^{-a}$$

Inaugural faculty at Ecole Polytechnique

Lagrange, Laplace, Monge, Fourier

↓

Poisson

Louis Bachelier

Theory of Speculation

thesis for Docteur Sciences Mathematiques 1900
dedicated to H. Poincare

Introduced the Chapman-Kolmogorov process

$$P_{z,t+\tau} = \int_{-\infty}^{\infty} P_{x,t} P_{z-x,\tau} dx \qquad c^2 \frac{\partial P}{\partial t} - \frac{\partial^2 P}{\partial x^2} = 0$$

Discovered the radiation of probability. "Each price x during an element of time radiates towards its neighboring price", but incorrectly uses c= lim (x / t) as both tend to zero. The correct limit when change takes places at a finite velocity should arrive at the telegrapher's equation

One needs to take the limit of small x and small t, such that $D=x^2 / t$ = constant. This means the x/t does not converge to a limit, i.e., the velocity is not well defined. Bachelier did not perform the limit process correctly and his thesis was not well received.

Bachelier's analysis would work for the telegraphers' equation with constant velocity c, and rate of change λ of direction and diffusion constant $D=c^2/\lambda$

The Lognormal Distribution McAlister (1879)

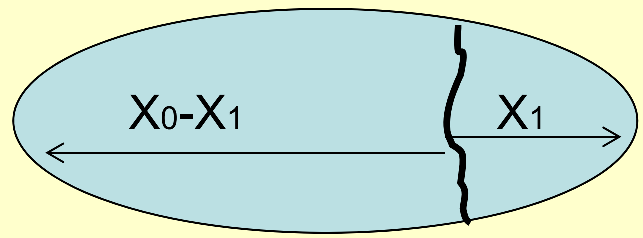
Kolmogorov's (1941) rock crushing for ore separator

$$X_1 = q_1 X_0$$
$$X_2 = q_2 X_1 = q_2 q_1 X_0$$
$$\vdots$$
$$X_n = q_n \cdots q_1 X_0$$

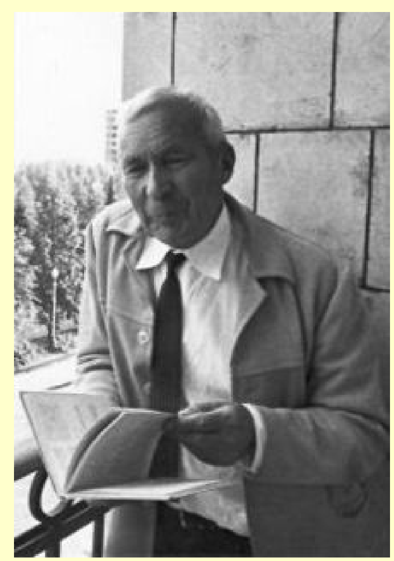
$$X_{n-1} - X_n = q_{n-1} \cdots q_1 (1 - q_n) X_0 = (1 - q_n) X_{n-1}$$

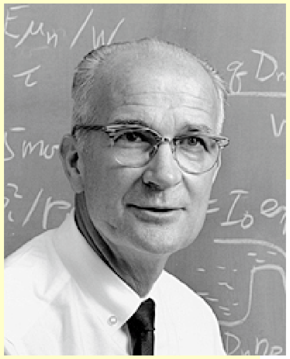
$$\sum_{n=1}^{N} \frac{X_{n-1} - X_n}{X_n} \to \int_{X_N}^{X_0} \frac{dx}{x} = \ln\left(\frac{X_0}{X_N}\right) = \sum_{n=1}^{N} (1 - q_n)$$

$$f(x)\,dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\left[\log\left(\frac{x}{\langle x \rangle} \right) \right]^2 / 2\sigma^2 \right)$$



The RHS is a sum of random variable and for large N it has a Gaussian distribution. If ln (X_N) is normal, then X_N is lognormal





Shockley criteria for bonuses at Bell Labs

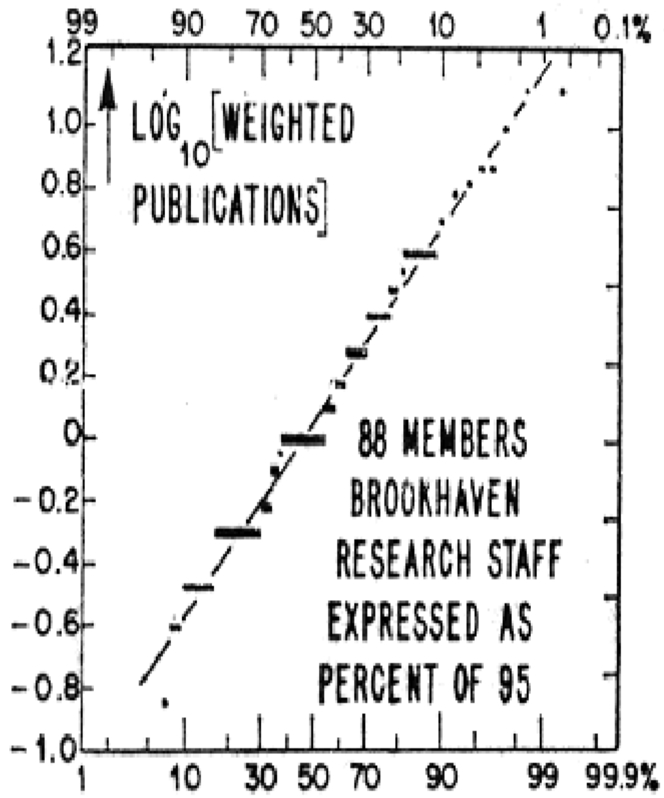


Figure 44 Cumulative distribution of logarithm of "weighted" rate of publication at Brookhaven National Lab. plotted on probability paper.[5]

Product of factors to publish, so judge success according to the logarithm of number of publications

Condorcet

Laplace all in Paris

Poisson

Poisson (1837) **Recherches sur la probabilite des jugements en matiere criminelle et en matiere civile**

(book rejected by his peers, era of the philosophe was over)

St. Petersburg

French activity dormant until

J. Bertrand “Calcul de probabilites” (1879)

H. Poincare “Calcul de probabilites” (1896)

P. Levy “Calcul de probabilites” (1925)

“Theorie de l’addition des variables aleatoires” (1937)

Processes stochastiques et mouvement brownien (1948)

$$P(|X| \geq t) = \int_{|x|\geq t} p(x)\,dx$$

$$\leq \int_{|x|\geq t} \frac{x^2}{t^2} p(x)\,dx$$

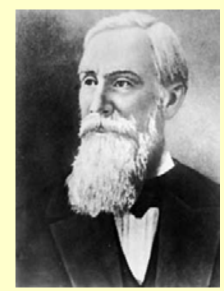
$$= \frac{1}{t^2} \int_{|x|\geq t} x^2 p(x)\,dx$$

$$\leq \frac{1}{t^2} \langle x^2 \rangle$$

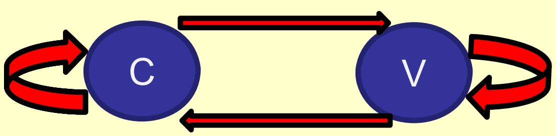
Chebyshev (Ph. D 1849) Theory of Probability (1846)

Students: Markov (**Ischislenie Veroyatnostei**) (1900)

Lyapunov (limit theorems/characteristic functions)



A. A. Markov, An example of statistical investigation of the text *Eugene Onegin* concerning the connection of samples in <u>chains</u> (1913)



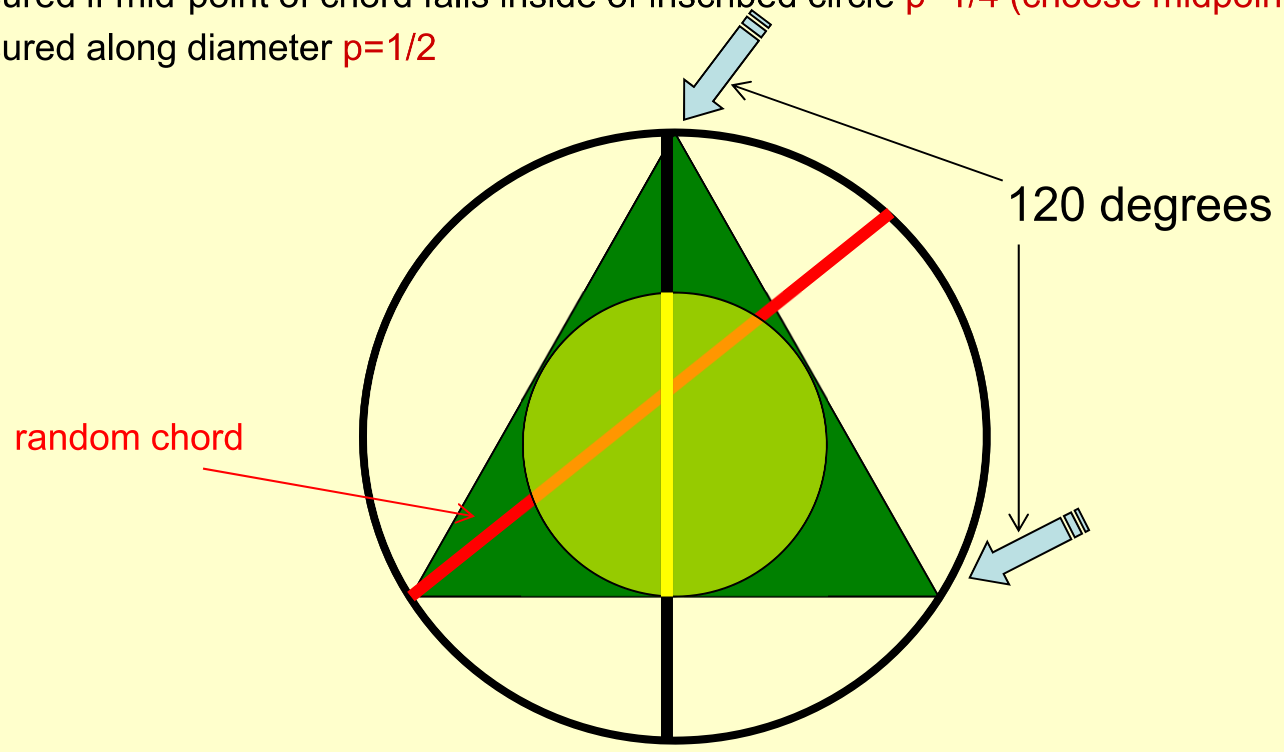
Chebyshev’s inequality

J. Bertrand's Paradox (1888)

What is the probability that a randomly drawn chord, in a circle, will be longer than the side of an inscribed equilateral triangle?

Answer depends on the assumption of what random variable (angle, area, length, … is uniformly distributed.

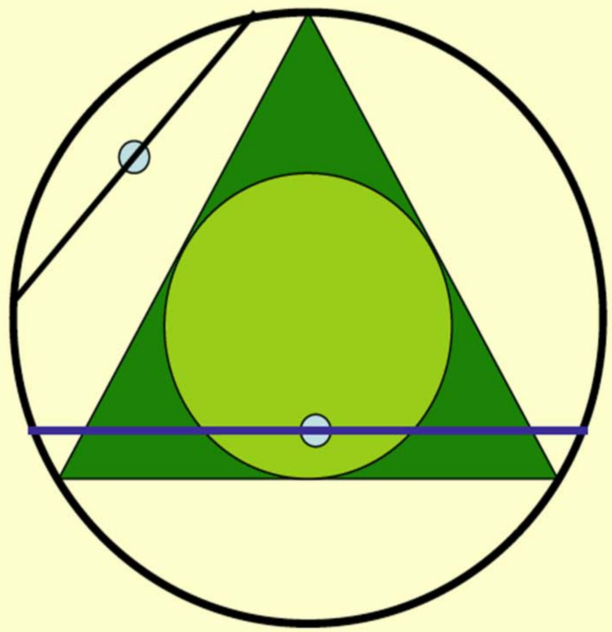
- Measured along circumference p=1/3 (choose points on circumference)
- Measured if mid-point of chord falls inside of inscribed circle p=1/4 (choose midpoint)
- Measured along diameter p=1/2



A DIFFERENT APPROACH

Method of choosing a random point leads to a probability of 1/4 for chord to be larger than side of the triangle

Green circle has ¼ the area of the large circle



If the randomly chosen point lies outside the green circle the chord will be shorter than the side of the triangle.

If the point is inside the green circle the chord will be longer than the side of the triangle.

St. Petersberg Paradox

(a Nicholas Bernoulli problem in Montmort's book)

Flip a coin until a head (H) appears

Possible sequences	P(W)	W(winnings)
H	½	1
TH	¼	2
TTH	1/8	4
$\overbrace{T \cdots T}^{N}H$ (N tails)	$1/2^N$	2^N

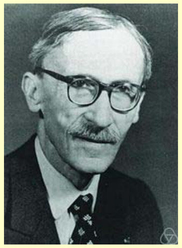
$$<W> = \Sigma\ Wp(W) = 1X1/2 + 2X1/4 + 4X1/8 + \cdots$$

$$= \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots$$

$$= \infty$$

What is the fair ante? Finite or infinite?

This problem is discussed by Daniel Bernoulli in the Commentarii of the St. Petersburg Academy in the 1720's



Levy Flights

Random walks with infinite variance (scale invariance)

$$p(x) = \frac{n-1}{2n}\sum_{j=0}^{\infty} n^{-j}\left(\delta_{x,b^{j}} + \delta_{x,b^{-j}}\right)$$

$$\tilde{p}(k) = \int_{-\infty}^{\infty} e^{ikx} p(x)dx = \frac{n-1}{n}\sum_{j=0}^{\infty} n^{-j}\cos\left(b^{j}k\right)$$

$$\left\langle x^{2}\right\rangle = \sum_{j=0}^{\infty} x^{2} p(x) = \infty \text{ if } b^{2} > n$$

$$\cos\left(b^{j}k\right) = \frac{1}{2\pi i}\oint \Gamma(s)\cos\left(\frac{\pi s}{2}\right) b^{-js}|k|^{-s} ds$$

$$\tilde{p}(k) = \frac{1}{2\pi i}\left(\frac{n-1}{n}\right)\sum_{j=0}^{\infty}\oint \Gamma(s)\cos\left(\frac{\pi s}{2}\right) b^{-js}|k|^{-s} n^{-j} ds$$

$$\tilde{p}(k) = \frac{1}{2\pi i}\oint \frac{|k|^{-s}\Gamma(s)\cos(\pi s/2)}{1-n^{-1}b^{-s}}\mathrm{ds}$$

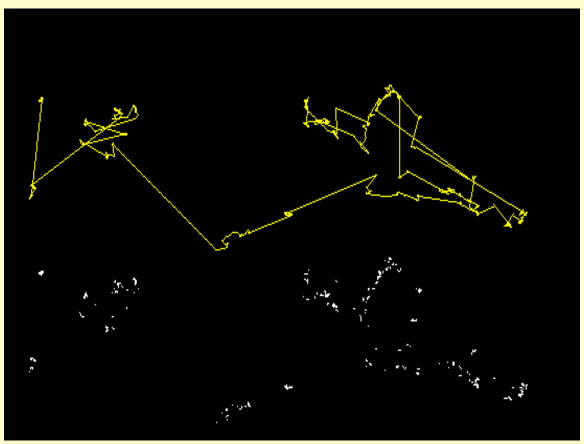
the integrand has simple poles at $\mathrm{s} = 0, -2, -4, \Lambda$

and at $\mathrm{s} = -\frac{\ln n}{\ln b} \pm 2\pi i m/\ln b,\ \ m = 0,1,2,\Lambda$

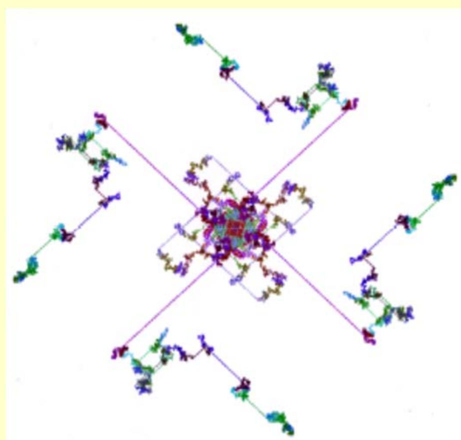
$$\lim_{k\to 0}\tilde{p}(k) = 1-|k|^{\beta}Q(k)+O\left(k^{2}\right) \approx \exp\left(-|k|^{\beta}\right)$$

where Q is periodic in $\ln k$ with period $\ln b$

$$\beta = \frac{\ln n}{\ln b} < 1$$



Random process



Nonlinear dynamical process

Both cases exhibit fractal structure