Philosophy and Methodology of Sciences

Tomasz Placek

Probability

(lecture 7 and 8)

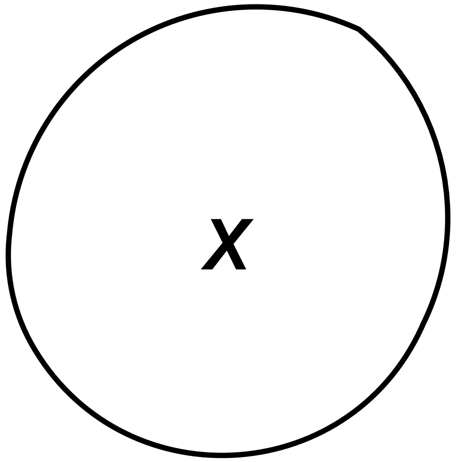
Project co-financed by the European Union under the European Social Fund

What are probabilities?

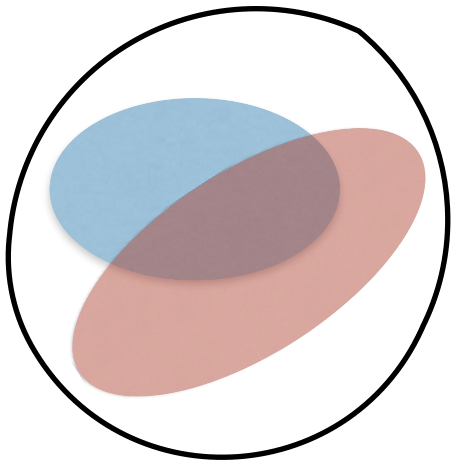
Easy formally: a normalized to unity, countable measure on a Boolean algebra (field of sets)

Begin with a set *X*

Take *some* family *F* of subsets of *X* that (1) contains *X*, (2) is closed under finite intersection and (3) countable union and (4) complementation. Such an *F* is called sigma-field (of subsets of X) or an algebra (of subsets of X).

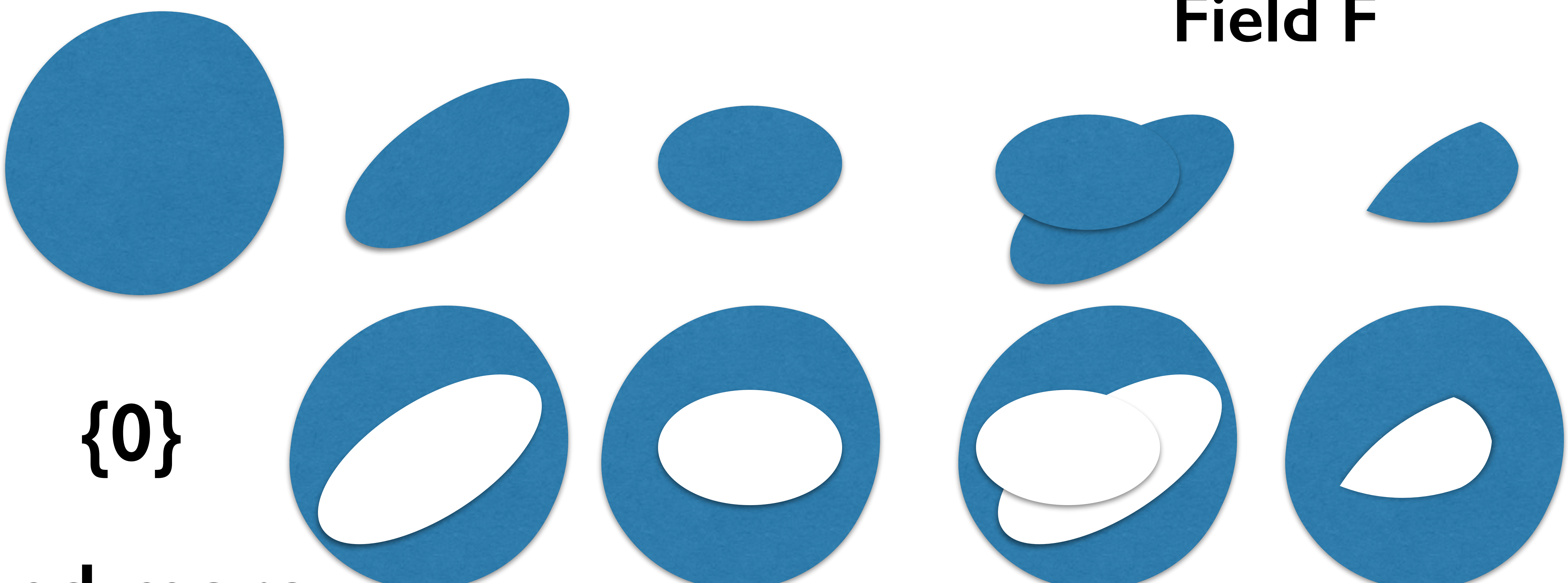


Set X



Generating subsets

Field F



and more ..

Axioms for the probability measure

Ax. 1: $0 \leq p(E) \leq 1$ for every $E \in F$;

Ax 2: $p(X) = 1$

Ax 3: for a sequence $E_1, E_2, E_3, \ldots$ pairwise disjoint elements of *F* (ie, such that $E_i \cap E_j = 0$ if $i \neq j$):

$p(\bigcup_i E_i) = \sum_i p(E_i)$

Ax 2 reads: p is normalized to unity, Ax 3 reads: p is countably additive.

In sum:

a classical probability space is a triple $<X, F, p>$, where X is a set, *F* a sigma-field on X, and $p: F \to [0, 1]$ satisfies Ax 1-3.

We said “classical”. Other name: Kolmogorovean (after A.V. Kolmogorov, who gave the axioms above in 1930’s).

There are non-classical theories of probability. Typically, they alter the algebraic part, i.e,, *X, F.*

Most known: quantum probability (defined on an orthomodular lattice)

They come with a conceptual price. In quantum case, rejection of the “union law”.

Price is conceptual: you cannot see the law violated in experiment

Math of probability calculus is easy, interpretation of probability is hard.

Interpretation: what is probability about, what do probability statements mean?

As probability statements occur in sciences, a scientist might want to know what they mean.

Criteria for a good interpretation:

- math: compatibility with the / a mathematical probability calculus (ideally, derivability of all Kolmogorov's axioms, but weaker results are also welcome.)
- knowability - an interpretation should permit that we learn numerical values of probability
- reasonability: an interpretation should make sense of applications of probability. E.g., of statistical experiments in science.

Examples of probability statements:

- The probability that this coin lands head up is 0.5

- The probability that this atom will decay within next hour is 0.1789

- Given the present data, the probability that Copernicus had curly hair is below 30 percent

- It’s twice more likely that Wisła will win with Jagiellonia (than it looses or draws) - I’m ready to bet two to one that Wisła will win.

-The probability of a rain in Krakow tomorrow is between 10 and 20 percent.

Intuitions:

- The probability that this coin lands head up is 0.5

(a) It's about statistics. Roughly 50 percent of tosses with this coin end with heads up. "This coin" (particular) surreptitiously refers to "a long run of tosses"

(b) The coin + the tossing setup is an indeterministic device. c + t has a disposition to produce heads (tails).

Probability is a measure of this disposition or potency to produce heads (tails)

Same diagnoses (but a different emphasis) of the next statement, i.e,

- The probability that this atom will decay within next hour is 0.1789.

- Given the present data, the probability that Copernicus had curly hair is below 30 percent

A semi-logical numerical relation between two statements, say, an evidence and a hypothesis: how much the former confirms the latter. Generalization of deducibility:

if in theory T, h is deducible from e, then c(e, h) = 1;

if in theory T, non-h is deducible from e, then c(e, h)=0.

This relation (logical probability, degree of confirmation) is supposed to be as objective as logical deducibility.

It’s twice more likely that Wisła will win with Jagiellonia (than it looses or draws). I’m ready to bet two to one that Wisła will win.

My degree of belief. Well, a rational agent’s degree of beliefs.

-The probability of a rain in Krakow tomorrow is between 10 and 20 percent.

Tricky. Statistics? Logical probability? Weighted degrees of beliefs of meteorology experts?

Interpretations-present contenders are the versions of:

- logical (relation between data and confirmation).
- frequentism (hypothetical, or limiting)
- degree of rational beliefs (Bayesianism, after Thomas Bayes)
- degree of possibility (propensity interpretation)

Historical contenders:

- classical (Laplacean)
- frequentism (actual)

Plan: first about a historical contender:

- classical (Laplacean)

Then about:

- logical (relation between data and confirmation).
- frequentism (actual and hypothetical);
- Bayesianism;
- (propensity interpretation - left for determinism lecture)

Classical (Laplacean) interpretation

Illustration: throwing a die.

- Events: "1", …"6" but "even", "odd", "divisible by 3", etc.
- Find "basic events". Presumably "1", …"6".
- In the absence of evidence, assign equal probability, i.e., 1/6.
- For an event, say "odd number of dots" ask in which basic event this can be realized (1, 3, 5) and calculate the fraction:

  # (favorable basic events)/ #(basic events),

  which is 3/6 in our case

Classical (Laplacean) interpretation

In the absence of any evidence / balanced evidence: probability is shared equally among all the reduced (basic) possible outcomes. Probability of an event is the fraction of the total number of possibilities in which the event occurs.

"The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought." Laplace 1884

Problems for classical interpretation

“cases equally possible”:

- if possible means probable, analysis is circular
- if possible means conceivable (for me), we get degrees of belief;
- if possible means possible, we have a substantial thesis about degrees of possibility.

A die might be unfair, our procedure will then go awry. Could we still then believe that at some level, there are basic equiprobable events?

This is the spirit of the (notorious) indifference principle. Classical int. is committed to it.

Even if the indifference p. were correct, how to discern basis events?

- Bertrand paradoxes (cubes variously described)
- physical statistics (Maxwell, Bose-Einstein, Fermi)

Criteria: some (minor) problems with extending to infinite probability spaces, but otherwise the criteria are met.

But think of making sense of a statistical experiment.

Logical interpretation (Łukasiewicz 1908, Carnap 1950). (Below about Carnap’s version)

Recall the boundary condition. Let T be a language with a deducibility relation. Then we require:

if h is deducible from e in T, then $c(e, h) = 1$;

if non-h is deducible from e in T, then $c(e, h)=0$.

Task: define a unique c for every pair of sentences.

A simple language with of a finite number of logically independent monadic predicates (naming properties) and countably many individual constants and the usual logical connectives.

Consider the strongest (consistent) statements that can be made in that language.

These are conjunctions of complete descriptions of each individual; each description is a conjunction containing exactly one occurrence (negated or unnegated) of each predicate of the language. These are called *state descriptions*.

A probability measure *m(.)* on state descriptions extends to all sentences of our language, as any sentence is a disjunction of state desc.

A measure *m* can be used to define confirmation function *c*(. , .):

$$c(h, e) = \frac{m(h \,\&\, e)}{m(e)}$$

Ideology: it should not be relevant which names enter into a state description. Consider thus a set S of states descriptions, such that any sd from S can be transformed into a sd from S by a permutation of names. Such an S is called *structure description*.

Illustrations (after Hajek). For 1 predicate F and 3 names.

1. $Fa \,\&\, Fb \,\&\, Fc$
2. $\neg Fa \,\&\, Fb \,\&\, Fc$
3. $Fa \,\&\, \neg Fb \,\&\, Fc$
4. $Fa \,\&\, Fb \,\&\, \neg Fc$
5. $\neg Fa \,\&\, \neg Fb \,\&\, Fc$
6. $\neg Fa \,\&\, Fb \,\&\, \neg Fc$
7. $Fa \,\&\, \neg Fb \,\&\, \neg Fc$
8. $\neg Fa \,\&\, \neg Fb \,\&\, \neg Fc$

These are state desc.

They yield these structure descs:

$\{1\}$, "Everything is F";

$\{2, 3, 4\}$, "Two Fs, one $\neg F$";

$\{5, 6, 7\}$, "One F, two $\neg F$s"; and

$\{8\}$, "Everything is $\neg F$".

State description	*Structure description*	*Weight*	*m**
1. $Fa.Fb.Fc$	I. Everything is F	1/4	1/4
2. $\neg Fa.Fb.Fc$			1/12
3. $Fa.\neg Fb.Fc$	II. Two Fs, one $\neg F$	1/4	1/12
4. $Fa.Fb.\neg Fc$			1/12
5. $\neg Fa.\neg Fb.Fc$			1/12
6. $\neg Fa.Fb.\neg Fc$	III. One F, two $\neg F$s	1/4	1/12
7. $Fa.\neg Fb.\neg Fc$			1/12
8. $\neg Fa.\neg Fb.\neg Fc$	IV. Everything is $\neg F$	1/4	1/4

Measure *m** assigns same value to each state description, hence 1/4. Each state des, receives its (equal) share of the value for the state description.

State description	*Structure description*	*Weight*	*m**
1. $Fa.Fb.Fc$	I. Everything is F	1/4	1/4
2. $\neg Fa.Fb.Fc$			1/12
3. $Fa.\neg Fb.Fc$	II. Two Fs, one $\neg F$	1/4	1/12
4. $Fa.Fb.\neg Fc$			1/12
5. $\neg Fa.\neg Fb.Fc$			1/12
6. $\neg Fa.Fb.\neg Fc$	III. One F, two $\neg F$s	1/4	1/12
7. $Fa.\neg Fb.\neg Fc$			1/12
8. $\neg Fa.\neg Fb.\neg Fc$	IV. Everything is $\neg F$	1/4	1/4

Consider hypothesis $h = Fc$, true in 4 of the 8 state descriptions, hence $m^*(h) = 1/2$. Suppose we learned that $F(a)$ — this is our evidence e. We have: $m^*(h \,\&\, e) = 1/4+1/12 = 1/3$ and $m^*(e) = 1/2$. Thus:

$$c^*(h, e) = \frac{m^*(h \,\&\, e)}{m^*(e)} = 2/3.$$

*m** and *c** are merely examples: Carnap is not committed to the equal share/equal share rule.

So the task is to put forward such a set of rules on *m* that they yield a single measure *m** and *c** (or a nice set of these)

Technical problem.

Criteria:

- axioms derived;
- probabilities can be calculated (if rules for *m** are known)
- making sense of applications: we need to re-describe what a role of experiments is.

Frequency interpretations

Lingo:

attribute in a class - e.g., heads up in the class of tosses of that coin

The *relative frequency* (rf) of an attribute in a class = the proportion of members of the class that have the attribute.

Actual frequentism

$f^a_X(O)$ = r iff, in all actual outcomes of X, O has an rf of r , where X- class and O - attribute.

Interpretation:
Probability of O (wrt to actual outcomes of X) = $f^a_X(O)$

Problems:

1. typically, probability is not relativized to actual outcomes of some class
2. the two, probability of O and $f^a{}_X(O)$ are typically different, if the number of actual outcomes of X is small.

Solution: go infinite (limit + modality)

Hypothetical frequentism

Let S be an infinite sequence of outcomes of X. Let $rf_n(O)$ be the rf of O in the first *n* elements of S.

If $rf_n(O) \rightarrow r$ as $n \rightarrow 1$, r is called the limiting rf of O in S.

$f^{\infty}_{X}(O) = r$ iff repeating X infinitely would produce a sequence in which O has a limiting rf of r.

$f^{\infty}_{X}(O) = r$ iff repeating X infinitely would produce a sequence in which O has a limiting rf of r.

Here we go above strict empiricism

Hypothetical frequentism

Let S be an infinite sequence of outcomes of X. Let $rf_n(O)$ be the rf of O in the first *n* elements of S.

If $rf_n(O) \rightarrow r$ as $n \rightarrow 1$, r is called the limiting rf of O in S.

$f^{\infty}{}_{X}(O) = r$ iff repeating X infinitely would produce a sequence in which O has a limiting rf of r.

$f^{\infty}{}_{X}(O) = r$ iff repeating X infinitely would produce a sequence in which O has a limiting rf of r.

Here we go above strict empiricism

It may happen that limiting rf does not exists, e.g.,

HTHHTTHHHHTTTTHHHHHHHHTTTTTTTT….

$rf_n(H)$ =

1,1/2,2/3,3/4,3/5,3/6,4/7,5/8,6/9,7/10,7/11,7/12,7/13,7/14,8/15,9/16,10/17,11/18,12/19,13/20,14/21,15/22…. - it oscillates between 1/3 and 2/3

Postulate outright that a limiting rf exists in S. This excludes the HT sequence above

Next, we ascertain probability (limiting rf) on the basis of some sampling of X (e.g, we may take every second element of X).

A place selection is a rule for selecting a subsequence of a sequence in which the decision whether to retain the **n**th element does not depend on the value of that or any subsequent element of the sequence.

A collective is an infinite sequence of outcomes in which each attribute has a limiting rf that is insensitive to place selection.

A collective is an infinite sequence of outcomes in which each attribute has a limiting rf that is insensitive to place selection.

Then HTHTHTHTHTHT …. is not a collective !!!!

For place selection “pick everything” rf(H) = 1/2 whereas place selection “pick every second” rf(H) = 0.

von Mises / Church axioms for a collective:

Axiom of Convergence: the limiting relative frequency of any attribute in S exists.

Axiom of Randomness: the limiting relative frequency of each attribute in S is the same in any recursively specifiable infinite subsequence of S.

Final word:

$f^m{}_X(O) = r$ iff repeating X infinitely would produce a collective in which O has a limiting rf of r.

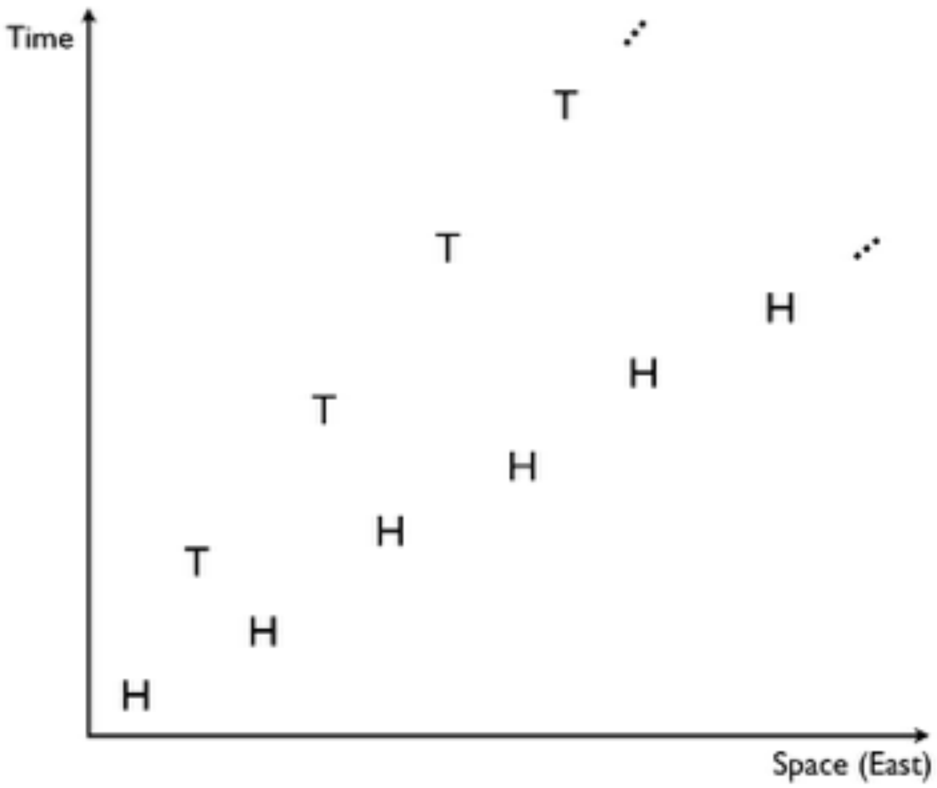
Probability of O in X = r iff $f^m{}_X(O) = r$ iff repeating X infinitely would produce a collective in which O has a limiting rf of r.

Problem with *would* (allegedly, counterfactual)

Jeffrey: there isn't a single collective that would result; many sequences could result.

That's uncharitable reading (no emphasis on single, no counterfactual) "Would" invites us to consider, e.g., possible infinite sequences of results of a coin tossing. Among them, there would be collectives. But do all collectives have the same r?

Problem with an ordering of X (after Hajek)



time: HHTHHTHHT… r(H) = 2/3

east: HTHTHTHTH… r(H) = 1/2

Responses:

- bite the bullet, X comes with an ordering, X = "A's coin tosses, executed on a train, ordered wrt "towards the east";
- observe that none of the above sequence is a collective; hope (and prove) that nothing like that can occur in collectives

Finally, our criteria of success

- applicability: met perfectly for statistical experiments
- knowability: our estimation of r can diverge from r by any margin. But, the larger the margin, the less likely the divergence (appeal to LLargeNumbers) and (by Bayes Th) we may learn how likely our estimation is good.
- math: nightmare. In a sequence, there exist r(A), r(B) but not r(A and B), which suggests fr. probability is not defined on a field … But the examples are not collectives… Further, fr. prob violate countable additivity…

Bayesianism

Logic imposes constraints on our rationality and how we should reason.

Our set of beliefs should not be inconsistent.
Having asserted all premises of a valid rule of inference, we should assert a conclusion of this rule.

But we rarely simply accept beliefs. Typically, we have beliefs with some degrees, measuring how much certainty we have.

How to think of inconsistencies of sets our beliefs. E.g., is it inconsistent to believe, for 60 percent, that Gagarin was an automaton and also believe, for 70 percent, that he was not an automaton?

Inconsistencies aside, are our degrees of belief probabilities?

(to be such, they need to satisfy the axioms)

To answer “yes”, one needs three arguments:

1. Degrees of beliefs *can* be represented by probabilities
2. Degrees of beliefs *must* be represented by probabilities
3. There is a way to elicit *numerical* degrees of beliefs (=probabilities) from human agents.

Ad. 1 Our beliefs are expressible in a language. A language in question can be thought of as equipped with a deducibility relation. In this case, there is an equivalence relation defined on formulas of the language. If it is reflexive, symmetric, and transitive, it yields equivalence classes of the set of formulas - the so-called Lindenbaum algebra.Elements of Boolean algebra can be thought of as propositions. Now, if the Lindenbaum algebra for a given language is Boolean (wrt logical operation - connectives of that language), it can have probabilities assigned.

Ad. 1 Continued

Drawback: since probabilities are assigned to propositions, we get same probability of equivalent formulas. This is the logical omniscience problem of probabilism.

Ad. 2 Leave it for a while - we need acquaintance with de Finetti’s bets first.

Ad 3. Ramsay’s idea: your degrees of belief drive your actions.

A’s degree of belief that *p* can be captured as a maximal amount of money A is ready to pay for a bet that pays 1 unit if *p* turns out to be true.

(or, a minimal amount for which A is ready to sell a bet that pays 1 unit if *p* turns out to be false).

In a slogan, degrees of belief = betting quotients

How to get “maximal amount of money you are ready to pay”? We tend (rationally!) to spend as little as possible

Some terminology: I’m putting X to Y that H (=it will rain tomorrow). X and Y are known as your odds that H is true.

Stake is what you and your opponent invest in a bet, that is, $S = X + Y$

We want to express your potential loss, X, in terms of *betting quotient* q and the stake S: $X = qS$. Hence Y (your potential gain) is $(1-q)S$.

The bet is fully characterized by two numbers, q and S, and a sentence, H.

Ad 3. Continued.

How to get “maximal amount of money you are ready to pay”? We tend (rationally!) to spend as little as possible

de Finetti’s symmetrised bets. To elicit your degree of belief that *H,* you (1) choose your odds for the bet that *H* is true. You come next to a bookie who (2) decides how large is the stake and (3) decides as well if you are going to bet for the truth, or against the truth, of *H.*

The bet (q, S, H) is fair for you if you are ready to pay qS for truth of H as well as (1-q)S for the falsity of H.

Your degree of belief that H is true is the betting quotient q of a fair (for you) de Finetti's bet (q, S, H)

Terminology: A's fair betting quotient for H = betting quotient for a (fair for A) bet for the truth of H

Pay-offs for de Finetti's bets

	Bet on the truth of H	Bet against the truth of H
H is true	(1-q)S	-(1-q)S
H is false	-qS	qS

(2) Our degrees of beliefs must be probabilities (must satisfy the probability axioms)

The Dutch Book

Dutch book is a set of bets you buy from a bookie and that is quite peculiar: no matter what happens you make a net loss.

That is, on some bets you may win, on some you may lose, but in total you lose.

The Dutch Book more precisely:

Let F be a set of sentences closed under classical logical connectives and closed restrictively under conditionalization stroke and let assume that an agent is willing to bet (simultaneously) on arbitrarily many bets expressible by sentences from F. A set of bets expressible by (some) sentences from F is called Dutch Book if it guarantees an agent's sure loss if the agent buys these bets.

Dutch Book theorem:

Let an agent have a degree of belief for any sentence in F, i.e., a fair betting quotient for any such a sentence. Then: if there is no Dutch Book on any subset of F, the agents betting quotients on sentences from F satisfy the probability axioms.

Proof (an idea)

Argue for the contraposition: assume that the agent's fair betting quotients violate a probability axiom. Argue then by cases, wrt the axioms, by exploiting the violation to construct a Dutch Book.

Reverse Dutch Book theorem

If someone's fair betting quotients for a subset of F satisfy all the probability axioms, there is no Dutch book for this person.

Considerable power of the two theorems. But there is more.

Subjectivism mitigated.

Agents can have completely different initial degrees of beliefs. How then could bayesianism apply to objective, or social matters, like confirmation of theories, or a community's opinion?

Convergence theorems:

Suppose that two agents begin their investigation with different degrees of belief in a hypothesis, but later witness (and accept) the same series of evidence. Suppose further that they agree about the hypothesis' implications: what evidence one should expect if the hypothesis is true.

Then, as a result of being exposed to the same evidence, their degrees of belief in the hypothesis converge.

How do our criteria for successful interpretation fare?

Math is fine.

We know our degrees of belief (just try to bet)

We need to do some massaging wrt applications. Hint: all experiments, theories, etc. are to help us to produce nice degrees of belief (rather than true descriptions of our world).

Since “nice” in “nice degrees of belief” will perhaps be explained by appeal to a good fit to the world (because of the convergence theorems), this is not awfully bad.

Literature:

- A. Hajek's entry in http://plato.stanford.edu/entries/probability-interpret/
- P. Maher's notes on von Mises, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.175.6281&rep=rep1&type=pdf
- A.Hajek "Fifteen Arguments Against Hypothetical Frequentism", http://philrsss.anu.edu.au/people-defaults/alanh/papers/fifteen.pdf
- An entry on Dutch Books in Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/dutch-book/
- A. Hajek SCOTCHING DUTCH BOOKS?, http://www.hss.caltech.edu/~jiji/