Bayesianism & New Experimentalism

S. Huber, L. Ostermann

Outline

- Introduction
- Bayes‘ Theorem
- [Common Probability Blunders]
- Bayesian Statistics: an Example
- Bayesianism
- New Experimentalism
- [Historical perspective on Induction]
- [„The Death Struggle of Skepticism and Inductivism“]

Introduction

Introduction

- One Paradigm in Statistics: Bayesian Statistics
- based on Bayes‘ Theorem, a probability theorem
- therefore connection to (deductive) probability theory.
- HOWEVER: Statistics is induction!
  important means for gain of knowledge.

Introduction

- Deduction vs. Induction

Deduction	Induction
• Conclusion already contained implicitly in premises • Premises true → Conclusion true • General → Special	• Conclusion goes beyond premises • Premises true → Conclusion true **at most** with high probability • Special → General

Introduction

- Deduction vs. Induction: Examples
  - Deductive problem: given that the coin is a fair coin. What is the probability that 100 tosses of the coin will produce 45 heads and 55 tails?
  - Inductive problem: given that 100 tosses of a coin produce 45 heads and 55 tails. What is the probability that the coin is a fair coin?

Introduction

- Deduction vs. Induction

„Why is induction so pervasive and critical in science? Because despite the appearance of being strictly about data and experiments, science is actually almost about unobservables, or, more specifically, about things and times outside the database of actual observations.“

Example: Iron melts at 1808 K.

- Based on particular samples at particular times
- However, taken as general fact

Introduction

- Presuppositions of Induction
  - Obvious: uniformity or regularity of nature

„The method of induction, the inference from yesterday to tomorrow, from here to here, is of course only valid if regularity exists."

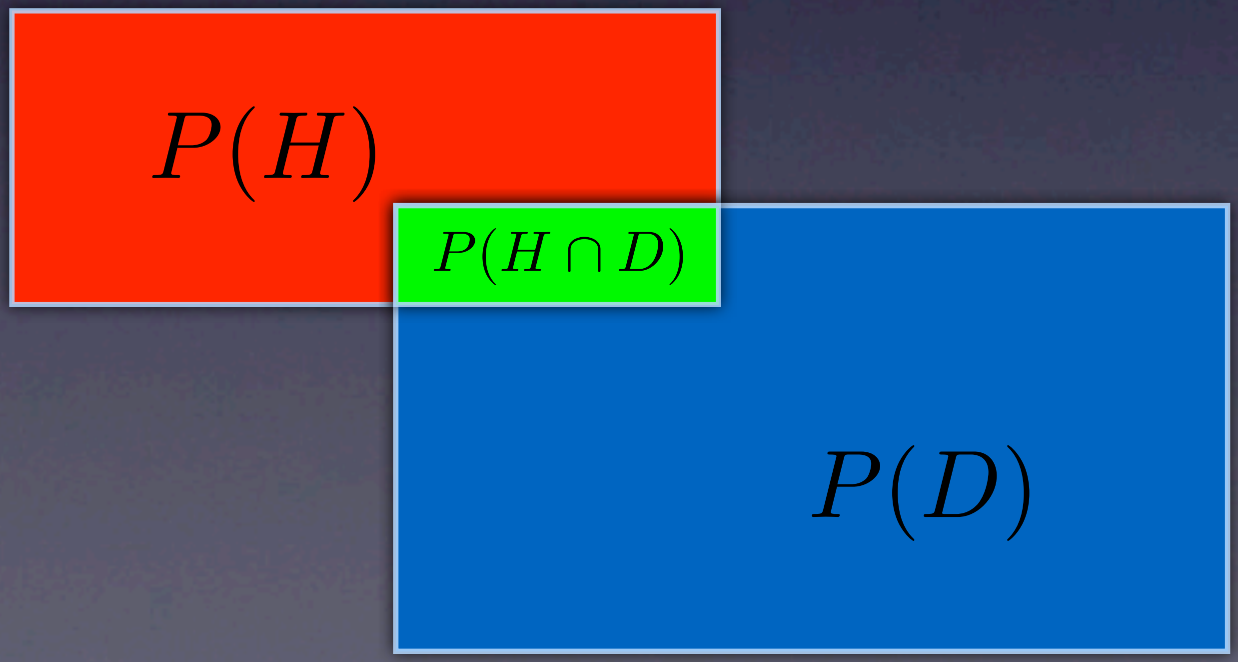
Introduction

- Presuppositions of Induction
  - subtle: parsimony of nature
  - Example: Iron melts at 1808 K.
    - „How can such a statement be true? Well, were nature‘s variety so unlimited that each atom was unique, not merely in its identity but also in its properties then there would be no such things as Iron or Oxygen (or Humans)!“
    - # of chemical elements limited to about 100 → parsimony of nature

Introduction

- Definition: Conditional probability

$$P(H|D) := \frac{P(H \cap D)}{P(D)},\ P(D) \neq 0$$



Bayes‘ Theorem

Bayes' Theorem

$$P(H|D) = \frac{1}{P(D)} P(D|H)\, P(H)$$

Posterior

norm. const.

Likelihood

Prior

Bayes' Theorem

$$P(H|D) = \frac{1}{P(}$$

Posterior: Probability of a hypothesis having utilised the new information considered in the likelihood.

Bayes‘ Theorem

$$\overline{\ \ } \quad P(D|H)\,P($$

$$D)$$

Likelihood: Summarises the data‘s impact on the probability of the the hypothesis; probability of the data **given** the hypothesis.

Bayes' Theorem

$$|H)\ P(H)$$

Prior: Initial probability of a proposition's truth or an event's occurrence, evaluated prior to collecting some particular data or evidence.

Bayes‘ Theorem

$$P(H|D) \sim P(D|H)$$

„[…] appears as a major step in the history of Statistics, being the first inversion of probabilities.“

Bayes' Theorem

- Alternative Forms

Two-Hypothesis-Form	Ratio-Form
$P(H_{1,2}\|D) = \frac{P(D\|H_{1,2})\,P(H_{1,2})}{\sum_{j=1}^{2} P(D\|H_j)\,P(H_j)}$	$\frac{P(H_1\|D)}{P(H_2\|D)} = \frac{P(D\|H_1)\,P(H_1)}{(D\|H_2)\,P(H_2)}$

Bayes' Theorem

- Alternative Forms

Two-Hypothesis-Form	Ratio-Form
$P(H_{1,2}\|D) = \frac{P(D\|H_{1,2})\,P(H_{1,2})}{\sum_{j=1}^{2} P(D\|H_j)\,P(H_j)}$	$\frac{P(H_1\|D)}{P(H_2\|D)} = \frac{P(D\|H_1)\,P(H_1)}{(D\|H_2)\,P(H_2)}$ Note: If there are additional unknown hypotheses no probabilities can be gained from this equation, only odds.

Attention, probability is out to trick you!

Common Probability Blunders

Common Probability Blunders

1. Ignored prior: Suppose a blood test for some rare disease which occurs by chance of one in every 100 000 people. The test is fairly reliable; if you have the disease it will correctly say so with probability 0.95; if you do not have the disease, the test will wrongly say you do with the probability 0.005. If the test says you do have the disease, what is the probability that this is a correct diagnosis?

Common Probability Blunders

1. Answer:

- Given $P(d) = 1/(100\,000)$, $P(pos|d) = 0.95$, $P(pos|\sim d)=0.005$;
- Question: $P(d|pos.) = ?$
- Use Two-Hypotheses-Form: $P(d|pos.) = 0.2\% \ll 95\%$ (common answer)
- Note: A study conducted in a leading American hospital found hat 80% of those questioned gave wrong answers.

Common Probability Blunders

2. Ignored precondition: A family has two children, one of whom is a girl. What is the probability that the other is a girl too?

   Predispositions: $P(b) = P(g) = 1/2$; the sexes are independent from each other.

Common Probability Blunders

2. Answer:

- X = {(b, g), (g, b), (g, g)}, Y = {(g, g)}
- $$P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{1}{4} \cdot \frac{4}{3} = \frac{1}{3}$$
- Note. Common Answer: 1/2

Common Probability Blunders

3. Suppose a court room where the prosecutor is about to convict you for a crime by arguing with the match probability of a DNA test.

Attention, probability is out to trick you (and maybe also the jury)!

Common Probability Blunders

3. „The prosecutor‘s fallacy refers to a confusion of two different probabilities. The «match probability» answers the question «What is the probability that an individual‘s DNA will match the crime sample, given that you are innocent?». But the question that should concern the court is «What is the probability that the suspect is innocent, given a DNA match?». The two queries can have wildly different answers.“

Common Probability Blunders

3. Given P(m|i) = 1/(1 000 000)

1: in dubio pro reo

$$P(i|m) = \frac{P(m|i)\,P(i)}{P(m)}$$

Genetic history and structure of the population of possible perpetrators.

Common Probability Blunders

3. „When that additional genetic information is incorporated into the analysis, typically the evidence is about as strong as the evidence one would get from a match that used half as many genetic markers and that ignored population structure. Because the evidence from the markers combines multiplicatively, the resulting adjustment equates to taking a square root.“

Common Probability Blunders

3. $\rightarrow P(i|m) = 1/1000$

Important: $P(m|i) \neq P(i|m)$

„The arenas in which probability reasoning applies can be coin tossing or weather forecasting, but they can also be law courts and medical establishments.“

Bayesian Statistics: an Example

Bayesian Statistics: an Example

1. Given: fair coin, opaque urne, blue and white marbles
2. Two volunteers A and B
3. A flips the coin without showing it
   i. if heads: 3 blue and 1 white marbles are put into the urne
   ii. if tails: 3 white and 1 blue marbles are put into the urne

Bayesian Statistics: an Example

4. With exception of A everybody else knows only that one of two hypotheses is true:

$$H_B \sim (1w, 3b) \vee H_W \sim (3w, 1b)$$

5. Let the class determine which hypothesis is probably true by the following method:

Bayesian Statistics: an Example

5. …following method:
   i. B mixes the marbles
   ii. B draws one marble, shows it to the class
   iii. The class applies Bayes' Theorem
   iv. B replaces it in the urne
   v. repeat i. - iv. as necessary
   vi. stopping condition:
   posterior probability > 0.999

Bayesian Statistics: an Example

- Common Sense
  - Large number of draws: high probability for conclusion to be true
  - Exact story is unclear: what is a «large» number? How probably is the conclusion true?
  - Exact story could be useful: Lots of data for safety, but is it necessary? → suppose high costs for each observation. Then unnecessary data are wasteful and undesirable.

Bayesian Statistics: an Example

- Quantitatively: Ratio-Form of Bayes‘ theorem

$$\frac{P(H_B|D)}{P(H_W|D)} = \frac{P(D|H_B)\,P(H_B)}{P(D|H_W)\,P(H_W)}$$

- Unitarity condition:

$$P(H_B|D) + P(H_W|D) = 1$$

Bayesian Statistics: an Example

$$P(H_B)/P(H_W) = 1:1 \Rightarrow P(H_B) = P(H_W) = 0.5$$

$$\begin{pmatrix} P(b|H_B) & P(w|H_B) \\ P(b|H_W) & P(w|H_W) \end{pmatrix} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

blue draw	white draw
$\frac{P(b\|H_B)}{P(b\|H_W)} = \frac{0,75}{0.25} = 3$	$\frac{P(w\|H_B)}{P(w\|H_W)} = \frac{0,25}{0.75} = \frac{1}{3}$

For the next draw: posterior → prior

Bayesian Statistics: an Example

# of draw	result	posterior odds	posterior probability
0	(prior)	1:1	0,5
1	w	1:3	0,25
2	b	1:1	0,5
3	w	1:3	0,25
4	b	1:1	0,5
5	b	3:1	0,75
6	b	9:1	0,9
7	b	27:1	0,964286
8	b	81:1	0,987805
9	b	243:1	0,995902
10	w	81:1	0,987805
11	b	243:1	0,995902
12	w	81:1	0,987805
13	b	243:1	0,995902
14	b	729:1	0,998630
15	b	2187:1	0,999543

Bayesian Statistics: an Example

- Question: How long would such an experiment run on average with the above mentioned stopping condition? Good estimate:
  - b → 3:1, w → 1:3, bw → 1:1
  - M(i) = margin of blue draws over white after the i-th draw
  - → $[P(H_B) : P(H_W)](i) = 3^{M(i)}:1$
  - $P > 0.999$ or $P < 0.001 \Leftrightarrow |M| \geq 7$

Bayesian Statistics: an Example

- Question: How long would such an experiment run on average with the above mentioned stopping condition? Good estimate:
  - on average: 4 draws give 3 supporting draws and one against → two will cancel, two will support.
  - Half of the data will cancel
  - → Length of the experiments on average $L = 2M^* = 2 \times 7 = 14$

Bayesian Statistics: an Example

- Question: How long would such an experiment run on average with the above mentioned stopping condition? Good estimate:
  - exact calculations: M*={2, 3, 4, 5} → L = {3.2, 5.6, 7.8, 9.9}
  - M* > 5 → L ~2M*
  - Test this theory for M*=2 (90% confidence) → You will see that it works.

marble.m

```
clear; clc;

sprintf('Welcome to the marble experiment!')

%Anfangsbedingungen: Wahl der 2 Hypothesen und Anzahl der Murmeln N
%Berechnung der Wsken f‚r Farbe unter gegebenen Hypothesen
%N=10;

N=input('Type in the number of marbles:');
b1=input('Type in the number of blue marbles according to the first hypothesis Hb:');
b2=input('Type in the number of blue marbles according to the second hypothesis Hw:');

Nb=[b1 b2]; Pb=[Nb(1)/N Nb(2)/N];
Nw=[(N-Nb(1)) (N-Nb(2))]; Pw=[Nw(1)/N Nw(2)/N];

%Erw‚nschte Verl‰sslichkeit des Ergebnisses in Prozent
V=input('Type in the required confidence V of the result in per cent:');

%M‚nzwurf und F‚llung der Urne, prior Wsken wegen M‚nzwurf f‚r beide
%Hypothesen 1/2
w=round(rand);

if (w ~= 0)
    U=[ones(1,Nb(1)) zeros(1,Nw(1))];
else
    U=[ones(1,Nb(2)) zeros(1,Nw(2))];
end

prior1=0.5; prior0=0.5; prior(1)=prior1/prior0;

%Murmel ziehen und posterior berechnen, maximale Zahl der Iterationen K
K=input('Type in the maximum number of iterations K:');

for k=1:K

    number(k)=k;
    z(k)=U(round(N*rand+0.5));
    schalter1=z(k);

    if (round(N*rand+0.5)==N+0.5)
        sprintf('Sorry! Please try again!')
        break
    end

    if (schalter1 ~= 0)
        post(k)=(Pb(1)/Pb(2))*prior(k);
    else
        post(k)=(Pw(1)/Pw(2))*prior(k);
    end

    prob(k)=post(k)/(post(k)+1);
    schalter2=prob(k);
```

Bayesian Statistics: Inference

- In the following: six variations of the marble experiment that require adjustments or raise difficulties.
- Stimulation of inside into why science often works just fine even when the data, model and scientists are all imperfect.
- Bayesian inference is robust!

Bayesian Statistics: Inference

1. Different preselected confidence:
   - Before: error chance < 1/1000 → L=14
   - Now: error chance < 1/1Mio desired
     → required margin M* = 13 (3^{13} =1.6Mio) →
     L=26
   - General pattern: more data will give greater confidence, relation between experimental work and evidential weight is **exponential**
     (at least for this special case)
   - → reasonable effort can produce virtually certain results: 1/(10^12) → L = 52

Bayesian Statistics: Inference

2. Different priors

   - prior simply different, e.g. 9:1 (different setup)
     - HB starts with 9:1 advantage → M*=5 suffices for HB to win, M*=9 for HW.
     - adjustment of stopping rule
     - shortens the experiments length: L_{neu} = 0.9*10+0.1*18 = 10.8 < 14

Bayesian Statistics: Inference

2. Different priors

- ignorance rather than knowledge of the setup
    - plausible: HB, HW are equally likely (0.5)
    - minimises the maximum possible error
    - analysis most responsive to data: the more the prior is indefinite, the more the analysis should weigh the likelihood in preference to the prior.
    - Most conservative estimates of confidence: L(0.5) > L(prior $\neq$ 0.5)
    - „The penalty for not knowing the real prior is experimental inefficiency.“

Bayesian Statistics: Inference

2. Different priors
   - controversial prior
     - different investigators believing or preferring different priors
     - considering the worst case and adjusting of experiments length will satisfy everyone

Bayesian Statistics: Inference

2. Different priors

   - extremely strong prior

     - priors of 0 and 1 make experimentation and evidence irrelevant, „strongly held error are hard to fix“ (don‘t discuss with fundamentalists)

     - strong prior, e.g. 10000/1, stopping condition: confidence of truth > 99,9% → no experiment needed!

     - It makes sense for strong priors to discourage small experiments.

     - But: small experiments can lead to reconsideration (1/10000 → 1/5)

Bayesian Statistics: Inference

3. Messy data

    - example: observe only every other marble
    - move the original hypotheses closer together
    - increase of experiment‘s length on average and also variability, „[…] Fortunately, sometimes quantity can compensate for quality“
    - (numbers for this example on next slides)

Bayesian Statistics: Inference

4. Different hypotheses

- consider: H1 ~ (3w, 5b), H2 ~(5w, 3b)
- P(b|H1) = 0.625, P(b|H2) = 0.375
  (same numbers for «look at every other marble»)
- more data will be needed: b → 5/3 → M*=14 for 99,9% conf.; 8 draws: 5 good, 3 bad → 2 good remain → every 4 draws will add one to the margin → L=14*4=56
- „Rather similar hypotheses require relatively more work […]. Fortunately for us, in general, the more similar two hypotheses become, the less we care about which is true.“

Bayesian Statistics: Inference

5. Different statistical frameworks

- besides Bayesian statistics: Frequentist statistics
- the smaller an experiment is the more it is relevant which statistical framework one uses. „As more data become available the influence of statistical differences will diminish.“

Bayesian Statistics: Inference

6. Paradigm shifts

- „[…] Asking a question with a hypothesis set that does not include the truth.“
- in case of the marble experiment: HB ~ (3b, 1w), HW ~ (1b, 3w).
  - suppose experimental outcome (57b, 50w)
  - HB true with 99,9%
  - But: peculiar length of experiment!
  - Maybe another hypothesis is true. HE ~ (2b, 2w)
  - again, the needed remedy is more data

Bayesian Statistics: Inference

6. Paradigm shifts

- How can HE be confirmed?
  - frequency of blue (or white) draws shall be nearly 0.5, rather than 0.75 (or 0.25)
  - experiment shall not repeatedly confirm HB, HW
  - if HE is true, then L ~ 49 ≠ 14
  - But: 27% chance that experiments stop at 21 or fewer draws, „This illustrates how perfectly honest and competent science can reach wrong conclusions at an early stage when few data of limited kinds are available.“

Bayesian Statistics: Inference

6. Paradigm shifts

- „The good news is that even if scientists are collecting data to answer a misdirected question, the data are likely, at least eventually, to embarrass the faulty paradigm and thus to precipitate a needed paradigm shift. That is why science works so often: even rather severe mistakes can be remediable. Scientific discovery is like a hike in the woods: you can go the wrong way for a while and yet still arrive at your destination at the end of the day.“

Bayesianism

Bayesianism
Introduction

- Now application to philosophy of science
- Questions:
  - Where does the confidence in science come from?
  - Is the extent to which theories are fallible misplaced or exaggerated?
  - „Why is Newtonian theory fairly reliable, since it has been falsified in a number of ways?“

Bayesianism
Introduction

- The Bayesians enter the stage:
  - no ascription of zero-probabilites to well-confirmed theory
  - use of inductive inference
  - based on Bayes‘ theorem
  - „[The Bayesians] would like to be able to show how and why a high probability can be attributed to Newtonian theory when used to calculate the orbit of Halley‘s comet or a space craft.“

Bayesianism

Objective Bayesianism

- Probabilities represent probabilities that rational agents ought to subscribe to in the light of the objective situation
  - How ascribe objective prior probabilities to hypotheses?
    - Fair: list (?) of all hypotheses and distribute probabilities equally by applying the principle of indifference
    - The list might be infinite → P = 0 → „Popper wins the day“
    - Chalmers: „The problem is insuperable.“ ⇒

Bayesianism
Subjective Bayesianism

- after Howson & Urbach (1989)
- (Prior) probabilities represent subjective degrees of belief
- **Claim:** Consistent interpretation of probability theory can be developed on this basis *and* it can do full justice to science.

Bayesianism
Subjective Bayesianism

- „Whatever the strength of the arguments for attributing zero-probability to all hypotheses and theories, it is simply not the case, argue the subjective Bayesians, that people in general and scientists in particular ascribe zero-probabilities to well-confirmed theories.“
  - People pre-booked trips to observe Halley‘s comet.
  - In their work, scientists take many laws for granted.

Bayesianism
Subjective Bayesianism

- Some arbitrariness in the choice wether probabilities are taken to measure what is reflected in scientific practise (Dorling, 1979) or taken to measure subjective degrees of belief (Howson & Urbach, 1989).
- The latter makes it more clear what the probabilities refer to

Bayesianism
Subjective Bayesianism

- Howson & Urbach‘s claim: subjective Bayesian theory constitutes an objective theory of scientific inference.
  - Probability theory is deductive, logic says nothing about the source of propositions constituting the premises of a deduction.
- But, the Bayesian defence can be taken even a step further.

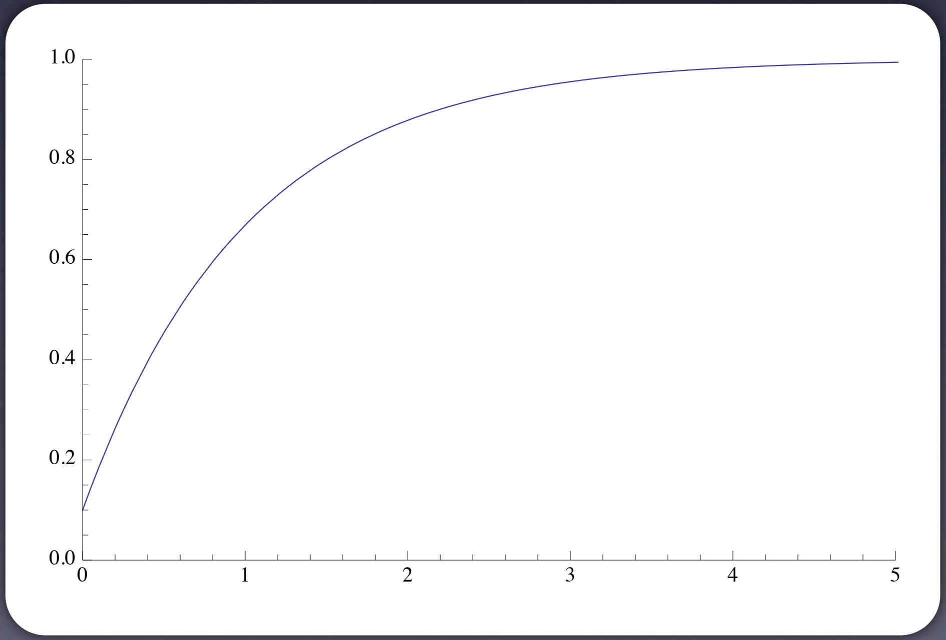
Bayesianism
Subjective Bayesianism

- Deferring beliefs of individuals can be made to converge given the appropriate input of evidence.
- Example: Consider two scientists A, B and some hypothesis H that predicts unexpected data D → P(D|H) = 1.
  A: P(H) high → P(D) high
  B: P(h) low → P(D) low
  ⇒ PA(H|D) ~ PB(H|D)

  what‘s more: positive evidence → scaling up of PB(H|D) → conviction at the end.

Bayesianism
Applications

- Example 1: confirmation of theory several times by the same experiment
  - Suppose: $(T \Rightarrow E) \Rightarrow P(E|T) = 1$; $P(T)$, $P(E_0)$ given
  - With this: $P(T|E_0) = P(T)/P(E_0)$, $P(T|E_1) = P(T|E_0)/P(E_1)$, …



Bayesianism
Applications

- Example 1: confirmation of theory several times by the same experiment
  - Experiment successful ⇒ $P(E_{i>>0}) \sim 1$ ⇒ $P(T|E_{i>>0}) \sim P(T|E_{i-1})$
  - „Once a theory has been confirmed by an experiment, repeating that same experiment under the same circumstances will not be taken by scientists as confirming the theory to as high a degree as the first experiment did.“

Bayesianism
Applications

- Example 2: Prout‘s Hypothesis (1815)
  - Preliminary remark: „Confirmations of a program are important rather than the apparent falsifications, which can be blamed on the assumptions in the protective belt rather than on the hard core [ad Lakatos]. The Bayesians claim to be able to capture the rationale for the strategy.“
  - Observation until 1815: atomic weights of chemical elements divided by the atomic weight of Hydrogen is nearly an integer.

Bayesianism
Applications

- Example 2: Prout‘s Hypothesis (1815)
  - Prout‘s hypothesis: atoms of the elements are made up of whole numbers of Hydrogen atoms.
  - But: 1815 measurement of the mass of Chlorine = 35.83
  - Prout **retained** his hypothesis (hard core), put the blame on some aspect of the measuring process (protective belt)

Bayesianism
Applications

- Example 2: Prout‘s Hypothesis (1815)
  - Bayesians answer why (here we follow the argumentation given by Howson & Urbach as interpreted by Chalmers):
    - h = Prout‘s hypothesis; e = evidence (mass);
      a = background knowledge (confidence in available techniques for measuring atomic weights and the degree of purity of the chemicals involved)
    - Proutians were very convinced (← historical evidence): $P(h) = 0.9$

Bayesianism
Applications

- Example 2: Prout‘s Hypothesis (1815)
  - chemists aware of the problem of impurities and variations in result of different measurements: $P(a) = 0.6$;
  - $P(e) = ?$ supposition: alternative to h was (historically) random distribution of atomic weights. H&U: $P(e|\sim h \,\&\, a) = 0.01$
  - $\Rightarrow P(h|e) = 0.878$; $P(a|e) = 0.073$
  - Note the change of $P(h)$ and $P(a)$

Bayesianism
Applications

- Example 2: Prout‘s Hypothesis (1815)
  - H & U conclude: reasonable response for the Proutians to retain the hypothesis and doubt the measurements
  - They point out: „Nothing much hinges on the absolute value of the numbers as long as they are of the right kind of order to reflect the attitudes of the Proutians as reflected in the historical literature“

Bayesianism
Applications

- Example 3: Ad hoc hypotheses
  - Given: Galileo‘s sights of moon‘s craters
    → moon not spherical
  - claim of rival: moon is still spherical, because there exists a transparent, crystalline substance enclosing the observable moon.
  - Popperian response: this claim is not independently testable and therefore not valid
  - But it is independently testable as the moon landings have shown (Chalmers) → Popper wrong.

Bayesianism
Applications

- Example 3: Ad hoc hypothesis
  - Bayesian response: t = hypothesis that moon is spherical; a = claim of rival (transparent substance)
  - Logic: $P(t \,\&\, a) < P(a) \Rightarrow P(t \,\&\, a) << 1$ if $P(a) << 1$
  - a is implausible → rejection of the rival‘s theory
  - „The theory of Galileo‘s rival could be rejected to the extent that his suggestion was implausible. There is nothing more to it and nothing else needed.“

Bayesianism
Applications

- Example 4: Data used to construct a theory cannot be considered as support for the theory (is this generally true?)
    - Bayesians give a counter example: recall the marble experiment and suppose a theory t_1 that claims all marbles are white.
    - Draw 1,000 marbles and observe that 495 are white.
    - Thus, construct another theory t_2: equal numbers of white and coloured marbles.
    - $P(495w|t_1) << 1 \Rightarrow P(495w|t_2) \sim 1$

      $\Rightarrow$ support for the theory

Bayesianism
Chalmers‘ critique of H & U

- Portray of Bayesian calculus: objectivity made of inference that serves to transform prior probabilities into posterior ones in the light of given evidence.
- It follows: any disagreements in science must have their source in prior probabilities held by scientists (totally subjective, reflecting degrees of belief)
- „Consequently, those of us who raise questions about the relative merits of competing theories and about the sense in which science can be said to progress will not have our questions answered by the subjective Bayesian, unless we are satisfied with an answer that refers to the believes that individual scientists just happen to have started out with.“

Bayesianism
Chalmers‘ critique of H & U

- Problems:
  - Gaining access to a knowledge of private degrees of believes (How shall we ask a dead scientist?)
  - „[…] Implausibility of the idea that we need to gain access to these private believes in order to grasp the sense in which, say, the wave theory of light was an improvement on its predecessor.“
  - Complexity of modern science (e.g. fundamental particle physics): no single person who grasps all aspects of this complex work → whose degree of belief shall be chosen and why?

Bayesianism
Chalmers‘ critique of H & U

- Problems:
  - Very strong believes cannot be shaken be any evidence to the contrary (recall the Proutians: rejection of evidence)
  - subjective Bayesianism cannot identify fundamentalism as bad scientific practise since prior probabilities cannot be judged.
    - „I pointed out here that the original incentive behind Prout‘s hypothesis was the near integral values of a range of atomic weights other than Chlorine, measured by the very techniques which the Proutians have come to regard as so unreliable that they warrant a probability as low as 0.073!“
    - „Does this not show that if scientists are dogmatic enough to begin with they can offset any adverse evidence.“

Bayesianism
Chalmers‘ critique of H & U

- Problems:
  - How is past evidence to count for a theory? (mercury‘s orbit as confirmation for Einstein‘s general relativity)
    - „That probability is not a measure of the degree of belief that a scientist actually has but a measure of a degree of belief they would have had if they did not know what the in fact do know.“

Bayesianism
Chalmers‘ critique of H & U

- Problems:
  - Bayesianism says nothing about the nature «evidence», but:
    - scientists will not respond to some evidential claim by asking the scientist making the claim how strongly he or she believes it.
    - good theory of scientific method will be required to give an account of the circumstances under which evidence can be regarded as adequate
    - experimentalists have plenty of ways of rejecting shoddy work, and not by appealing to subjective degrees of belief.

Bayesianism
Chalmers‘ critique of H & U

- What remains (in view of all these problems)?
  - a theorem of probability calculus, with a status akin to deductive logic
  - this concession serves to bring out the limitation of subjective Bayesianism‘s position:
    - „Their [H&U‘s] theory of scientific method tells us as much about science as the observation that science adheres to the dictates of deductive logic. The vast majority, at least, of philosophers of science would have no problem accepting that science takes deductive logic for granted, but would wish to be told much more.“

Intermezzo

Intermezzo

- So far, all accounts of scientific inference failed to some extent:
  - Positivism: theory dependence of observation; theory always transcend, and so can never be derived from, the evidence (Popper)
  - Popper‘s account: best theories survive the severest tests, but:
    - no clear guidance as to when a theory, rather than some element of background knowledge, should be held responsible for a failed test
    - unable to say something sufficiently positive about theories having survived tests.

Intermezzo

- Lakatos‘s research programs:
  - unable to give grounds for decisions that blame auxiliary assumptions rather than hard core principals
  - too weak to specify when it was time to abandon a research program in favour of another
- Kuhn‘s paradigms: could not give a clear answer to the question of the sense in which a paradigm could be said to be an improvement on the one it replaced (also valid for Lakatos and Popper, but to a smaller extent)

Intermezzo

- Feyerabend: took the theory-dependence movement to its extreme
  - giving up on the idea of special methods and science altogether
  - joining Kuhn in the portrayal of rival theories as incommensurable
- Bayesians: also part of theory-dependence tradition; problems see above.

Intermezzo

Is this it?

New Experimentalism

New Experimentalism

- Tackling the move towards radical theory-dependence at its source
- No return to positivism (i.e., senses provide an unproblematic basis for science)
- secure basis: not observation but experiment
- experiment can have «life of its own» independent of large scale theory
- scientific progress: steady build up of the stock of experimental knowledge

New Experimentalism

«Life of its own»

- 1820: Faraday constructed a primitive electric motor
- sent a sample of this device to his rivals around Europe, complete with instructions
- theory-dependent and fallible?

New Experimentalism

«Life of its own»

- Theory-dependent? Only in a very weak sense:
  - instructions only useful with knowledge about magnets, mercury and electric cells, but: „Nobody need deny the claim that someone who cannot tell the difference between a magnet and a carrot is not in a position to appreciate what counts as an established fact in electromagnetism. It is surely injudicious to use the term «theory» in such a general sense that «carrots are not magnets» becomes a theory.“

New Experimentalism

«Life of its own»

- Fallible?
  - Odd failures are neither surprising nor relevant
  - recent theoretical explanation of Faraday‘s motor differs from both Faraday‘s and Ampere‘s in significant aspects, but: Faraday‘s motor usually works and „it is difficult to comprehend how future advances in theory could somehow lead to the conclusion that electric motors don‘t work.“

New Experimentalism
«Life of its own»

- So, experimental effects that can be produced in a controlled way are not fallible, they are here for keeps.
- Scientific progress: accumulation of such effects
- theory-independent understanding of science‘s growth
- another example: Hertz‘s novel experimental effects (Chalmers, pg. 197)

New Experimentalism

Deborah Mayo on severe experimental testing

- Claim is supported by experiment if various ways in which it could be at fault have been *investigated* and *eliminated*
- Claim is borne out by experiment if severely tested, i.e. tested such that it would be unlikely to pass the test if it were false.
- In the following:; examples that illuminate several features of Mayo‘s position

New Experimentalism

Deborah Mayo on severe experimental testing

1. Snell‘s law

- suppose: very rough experiments, very large margins of error
- results are compatible with law within margins of error
- support?
- DM: No, because due to roughness of experiment, the law would be likely to pass this test even if it were false.

New Experimentalism

Deborah Mayo on severe experimental testing

2. Two cups of coffee in the morning → headaches in the afternoon
    - confirmation of «coffee causes headaches»?
    - DM: No, because elimination of various ways in which the claim could be in error has to be done first
        - Vietnamese beer last night, getting up too early, hard talk
        - controlled experiments to eliminate other possible causes
        - „Seek results that would be most unlikely to occur unless coffee does indeed cause headaches.“

New Experimentalism

Deborah Mayo on severe experimental testing

3. Eddington‘s test of Einstein‘s prediction of the bending of light in a gravitational field

- DM: No confirmation of general theory of relativity
- Only confirmation of Einstein‘s law of gravity
  - There exists a whole class of theories of space-time, all of which predict Einstein‘s law of gravity and hence the results of Eddington‘s experiment → no test of general theory.

New Experimentalism

Deborah Mayo on severe experimental testing

3. Eddington‘s test of Einstein‘s prediction of the bending of light in a gravitational field

    - Observations were in conformity to Einstein‘s law of gravity
    - Alternatives have been considered and eliminated in the experiment
    - Einstein‘s law has been severely tested, general theory not.

- „The growth of scientific knowledge is to be understood as the accumulation and extension of such [severely testable] laws.“

New Experimentalism

Deborah Mayo on severe experimental testing

4. Learning from error (error detection)

- problematic features of Uranus's orbit → problems for Newtonian theory at that time
- source of the trouble could be traced → discovery of Neptune
- Note: learning from error (in that sense) goes beyond Poppers account, in contrast to his slogan „We learn from our mistakes.“

New Experimentalism

Deborah Mayo on severe experimental testing

5. Triggering scientific revolutions

- Black body radiation, radioactive decay, photo-electric effect
- abandonment of classical physics → quantum theory
- scientific revolutions can be rational (in contrast to Kuhn)

New Experimentalism

Deborah Mayo on severe experimental testing

- What about theory? Chalmers, pg. 205ff.
- „There is no doubt that the New Experimentalism has brought philosophy of science down to earth in a valuable way, and that it stands as a useful corrective to some of the excesses of the theory-dominated approach. However, I suggest it would be a mistake to regard it as the complete answer to our question about the character of science. Experiment is not so independent of theory as the emphasis of the previous sections of this chapter might suggest. The healthy and informative focus on the life of experiment should not blind us to the fact that theory has an important life too.“

Historical perspective on Induction

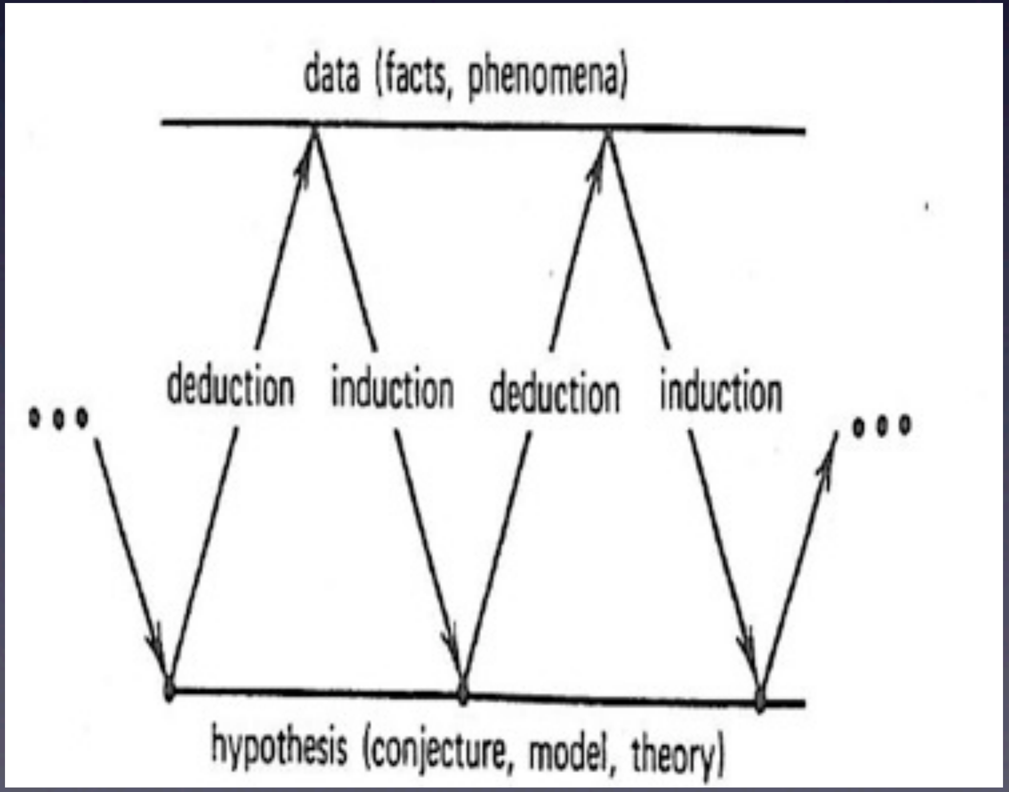
A selective incomplete overview

Historical perspective on Induction

- Aristotle (384-322 BC): three types of επαγωγη (→ lat. inductio → induction):
  - dialectical: „If a skilled pilot is the best pilot and the skilled charioteer is the best charioteer, then in general, the skilled man is the best man in any particular sphere“
  - enumerative: numerous adults have 32 teeth → all adult humans have 32 teeth
  - intuitive: bright side of the moon is always turned to the sun → moons shines because of reflected sunlight.

Historical perspective on Induction

- One of Aristotle‘s most influential contribution to the philosophy of science: the inductive - deductive method.



Historical perspective on Induction

- Epicurus (341-271 BC): discussion of fundamental role of induction in forming concepts and learning language (introduction of the term προληπσις → lat. anticipatio)
- Robert Grosseteste (1168-1253 AD): „Grosseteste‘s contribution was to emphasise the importance of falsification in the search for true causes and to develop the method of verification and falsification into a systematic method of experimental procedure.“
  - Deduction of consequences of a theory beyond its original application and then check of such predictions experimentally
  - Requirement of two metaphysical presuppositions of induction: uniformity and parsimony of nature.

Historical perspective on Induction

- John Duns Scotus (1265-1308 AD): introduction of a Method of Agreement
    - „If circumstances ABCD, ACE, ABEF, ADF all gave rise to the same effect X, then one could conclude that A could be the cause of X.“
- William of Ockham (1285-1347 AD): Method of Difference
    - „ABC gave effect X, AB did not. […] One could conclude that C could be the cause of X.“

Historical perspective on Induction

- Nicholas of Autrecourt (1300-1350 AD): sceptical view, „ […] It cannot be established that a correlation which has been observed to hold [in the past] must continue to hold in the future.“
- Francis Bacon (1561-1626 AD): «anticipation of nature» → «interpretation of nature»
  - „Theories must be larger and wider than the facts from which they are drawn.“
  - Good inductive theories would have predictive success.

Historical perspective on Induction

- Isaac Newotn (1642-1727 AD): „[…] although the arguing from Experiments and Observations by Induction be no Demonstration of general Conclusions, yet it is the best way of arguing which the Nature of Things admits of.“
  - Scientific theories are tentative and potentially revisable in light of future research and experimentation.

Historical perspective on Induction

- John Stuart Mill (1806-1873 AD): Stepwise inductive ascent from detailed observations to general theories
  - „Every deductive inference is as bottom an inductive one“ (see inductive-deductive method, Aristotle).
- Past Century (1901-2000 AD): induction has picked up a common synonym: statistics
  - „Statistics *is* inductive logic.“

„The Death Struggle of Skepticism and Inductivism“

or the story of loss and regain of induction

Induction lost

- Induction has been depicted as a great success so far. But:
  - a tremendous philosophical battle has been fought over induction from the ancient greek sceptics to the present
  - Dozens of books have been written on the so-called problem of induction
  - „The salient feature of attempts to Hume‘s problem is that they have all failed.“ (Hume‘s problem see below)
  - „[Induction is] the glory of science and the scandal of philosophy.“
  - „Hume‘s argument is one of the most robust, if not the most robust , in the history of philosophy.“

Induction lost

- Hume‘s argument builds upon three premises:
  - Any verdict on the legitimacy of induction must result from deductive or inductive arguments, because those are the only kinds of reasoning.
  - A verdict on induction cannot be reached deductively. No inference from the observed to the unobserved is deductive, specifically because nothing in deductive logic can ensure that the course of nature will not change.
  - A verdict cannot be reached inductively. Any appeal to the past successes of inductive logic, such as that bread can continued to be nutritious and that the sun has continued to rise day after day, is but worthless circular reasoning when applied to induction‘s future fortunes.

Induction lost

- Conclusion: Because deduction and induction are the only options, and because neither can reach a verdict on induction, there is no rational justification for induction.

Induction lost

- Another argument by Goodman:
  - consider emeralds examined before time T and suppose that all of them have been green. Then the straight rule of induction claims: all emeralds are green and an emerald will also be green if examined after time T.
  - Introduce the new property «grue»: An object is grue if it is examined before time T and is green or if it is not examined before time T and is blue.

Induction lost

- Goodman concludes: Then scientists examining a sample of emeralds before time T will discover that they are all green, and yet they are also grue. But of course that is a problem, because emeralds examined after time T will be green and hence fail to be grue.
- This problem shows that not all properties are appropriate for application of the straight rule of induction. So how can one decide in a non-arbitrary manner which properties are projectable?

Induction regained

- Hume: We need not fear that sceptical philosophical doubts about induction „should ever undermine the reasonings of common life“, because „Nature will always maintain her rights, and prevail in the end over any abstract reasoning what so ever“, and „Custom [...] is the great guide of human life.“
- Induction is grounded in custom or habit or instinct rather than in philosophical reasoning.
- „So common sense must trump sceptical doubt.“

Induction regained

- „Indeed, when philosophy‘s roots in common sense are not honoured, a characteristic pathology ensues: instead of natural philosophy happily installing science‘s presuppositions once, at the *outset,* by *faith*, in a trifling trinket of common sense knowledge, a death struggle with skepticism gets repeated over and over again for each component of scientific method, including induction. The proper task, «to explain induction», swells to the impossible task, «to defeat skepticism and explain induction». If (Hume‘s) philosophy cannot speak in induction‘s favour, that is because it is a truncated version of philosophy that has exiled animal habit, rather than having accommodated our incarnate human nature as an integral component of philosophy‘s common sense starting points. “

Induction regained

- Hume‘s argument is derived from the concern that the course of nature might change (entrance of skepticism).
  - Sun will not continue to rise, bread will not be nutritious anymore: „This is nothing less than the death fight with skepticism, nothing less than the end of the world.“
  - Not only induction hangs in the balance but also planetary orbits and human life.
- But: „If the course of nature did change, we would not be here to complain.“, „So as long as we are here or we are talking about induction, deep worries about induction are unwarranted“
- ad Goodman: judgements draw on general knowledge of the world and again common sense, „Such broad and diffuse knowledge resists tidy philosophical analysis.“

Literature

- A.F. Chalmers: What is this thing called science? Open University Press, 1999.
- H.G. Gauch: Scientific method in practice. Cambridge University Press, 2003.