# Priori oder nicht Priori, das ist die Frage

## Ein Versuch, die Auseinandersetzung zwischen Frequentisten und Bayesianern wissenschaftshistorisch und -philosophisch zu betrachten

Shuai Shao

17. Mai 2014

*„The Theorem itself is a landmark of logical reasoning and the first serious triumph of statistical inference, yet is still treated with suspicion by a majority of statisticians."*

*— Bradley Efron <Bayes' Theorem in the 21th Century>*

Bayes Theorem:

$$\underbrace{f(\theta|x_1,...,x_n)}_{Posteriori-Dichte} = \frac{\overbrace{f(x_1,...,x_n|\theta)}^{Likelihood} \cdot \overbrace{f(\theta)}^{Priori-Dichte}}{f(x_1,...,x_n)}$$

Anders als bei der frequentistischen Statistik, ist der Verteilungsparameter jetzt eine Zufallsvariable anstatt eine feste Größe.

Bayesianischer Hypothesentest: $H_0 : \theta = \Theta_0$ vs. $H_1 : \theta = \Theta_1$

Idee von der Likelihood-Ratio-Test anwenden:

$$\underbrace{\frac{P(H_0|\mathbf{x})}{P(H_1|\mathbf{x})}}_{Posteriori-Odds} = \underbrace{\frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)}}_{Bayes-Faktor} \cdot \underbrace{\frac{P(H_0)}{P(H_1)}}_{Priori-Odds}$$

Nach <The Theory of Probability> (3. Edition, 1961) von Jeffreys:

| BF | Interpretation für die Testentscheidung |
|---|---|
| $< 1$ | Nullhypothese wird gestützt |
| $\in [1, 3]$ | Anzeichen gegen Nullhypothese, aber kaum |
| $\in [3, 10]$ | Beachtliche Anzeichen gegen Nullhypothese |
| $\in [10, 30]$ | Starke Anzeichen gegen Nullhypothese |
| $\in [30, 100]$ | Sehr starke Anzeichnen gegen Nullhypothese |
| $> 100$ | Ausschlaggebene Anzeichnen gegen Nullhypothese |

Jeffreys-Lindley-Paradoxon:

- Harold Jeffreys <The Theory of Probability> (1. Edition, 1939)
  Darlegung die Diskrepanz von Ergebnissen zwischen frequentistischen und bayesianischen Hypothesentesten

- Dennis Lindley <A Statistical Paradox> (1957)
  Ausformulierung dieses Problems als ein Paradoxon

Beispiel: $H_0 : \theta = \theta_0 = 0.5 \text{ vs. } H_1 : \theta \neq 0.5$

$$X \sim Bin(n, \theta)$$

$$n = 98451$$

$$x = \sum_{i=1}^{n} x_i = 49581 \Leftrightarrow n - x = 48870$$

Frequentisten:

$$z(x) := \sqrt{\frac{n}{\theta_0(1-\theta_0)}}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \theta_0\right)$$

$$= \sqrt{\frac{98451}{0.5(1-0.5)}}\left(\frac{49581 - 49225.5}{98451}\right) \approx 2.266$$

$$\Rightarrow p\text{-Wert}: P_{H_0}(|z(X)| \geq |z(x)|) \ll 0.05$$

Ablehnen

Bayesianer:

$$BF(x) = \frac{P(x|H_0)}{P(x|H_1)} = \frac{\binom{n}{x}(\theta_0)^x(1-\theta_0)^{n-x}}{\int_0^1 \binom{n}{x} u^x (1-u)^{n-x} du}$$

$$= \frac{\binom{n}{x}(\theta_0)^x(1-\theta_0)^{n-x}}{\binom{n}{x} B(x+1, n-x+1)} \approx \frac{0.000195}{0.000010} \approx 19$$

Beibehalten

## Historische Entwicklung:

*Quelle: Wikipedia*

- Bis ins 20. Jahrhundert: Bayesianische Statistik

- Fisher (1890-1962) verlässt bayesianischen Rahmen, gründet mathematisierte Statistik mit eignen neuen Begriffen (Suffizienz, Effizienz, Konsistenz usw.) in induktiver Ausrichtung

- Neyman (1894-1981) gründet frequentistische Statistik mit Stochastik im Mittelpunkt in deduktiver Ausrichtung

- Neo-Bayesianer ab 1950: De Finetti, Savage

- Heftige Auseinandersetzung um die richtige Ausrichtung der Statistik bis heute

Ein Beispiel für die unterschiedliche Wahrscheinlichkeitsinterpretation zwischen Frequentisten und Bayesianer:

$$X \sim N(\mu, \sigma^2),\ n = 7,\ \bar{x} = 1,\ s = 1$$

95%-Konfidenzintervall: $\left[\bar{x} \pm t_{1-\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}\right] = \left[1 \pm 2.447 \frac{1}{\sqrt{7}}\right] = \left[1 \pm 0.925\right]$

Interpretation: Wird das gleiche Versuch unendlich oft durchgeführt, so überdeckt das Intervall in 95% der Versuche den wahren Erwartungswert.

Mit Jeffreys-Referenz-Priori-Dichte $f(\mu, \sigma) \propto 1/\sigma$ bekommt man auch das selbe 95%-Kredibilitätsintervall.

Interpretation: Mit der Wahrscheinlichkeit(=subjektiver Glaubensgrad) 0.95 liegt der wahre Erwartungswert in diesem Kreidibilitätsintervall.

Ein Beispiel für die unterschiedliche Wahrscheinlichkeitsinterpretation zwischen Frequentisten und Bayesianer:

$$X \sim N(\mu, \sigma^2),\ n = 7,\ \bar{x} = 1,\ s = 1,\ \sigma = 1.5$$

95%-Konfidenzintervall: $\left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = \left[1 \pm 1.96 \frac{1}{\sqrt{7}}\right] = \left[1 \pm 1.111\right]$

!!!

$$H_0 : \mu \leq 0 \text{ ablehnen} \stackrel{\text{Info}}{\Longrightarrow} H_0 : \mu \leq 0 \text{ beibehalten}$$

Mit Jeffreys-Referenz-Priori-Dichte $f(\mu, \sigma) \propto 1/\sigma$ bekommt man auch das selbe 95%-Kredibilitätsintervall.

???

Wie kann das Intervall durch mehr Information breiter werden?

Ein Beispiel für die unterschiedliche Wahrscheinlichkeitsinterpretation zwischen Frequentisten und Bayesianer:

$$X \sim N(\mu, \sigma^2),\ n = 7,\ \bar{x} = 1,\ s = 1,\ \sigma > 1.5$$

95%-Konfidenzintervall: Lässt sich nicht mehr exakt bestimmen

95%-Kredibilitätsintervall: $[1 \pm 1.45]$

- Ergebnis hängt von der Wahl der Priori ab
- Quelle des Vorwissens?

Einige Ursachen für die Beliebtheit der  frequentistischen Statistik nach <Why isn't everyone a Bayesian?> von Efron (1938-):

- Wesentliche Konzepte von Datensammlung und Experimentdesign wurden von Nicht-Bayesianern entwickelt
- Leicht anwendbare und weitgehend automatisierte Verfahren
- Bei komplexen Problemen ist es umso schwieriger, ein passendes Modell (und die Rechtfertigung dafür) zu finden
- Frequentistische Methoden gelten als wissenschaftlich objektiv

Einige Ursachen für die Beliebtheit der  frequentistischen Statistik nach <Why isn't everyone a Bayesian?> von Efron (1938-):

- ~~Wesentliche Konzepte von Datensammlung und Experimentdesign wurden von Nicht-Bayesianern entwickelt~~

**=> #FiveThirtyEight, #NateSilver**

- ~~Leicht anwendbare und weitgehend automatisierte Verfahren~~

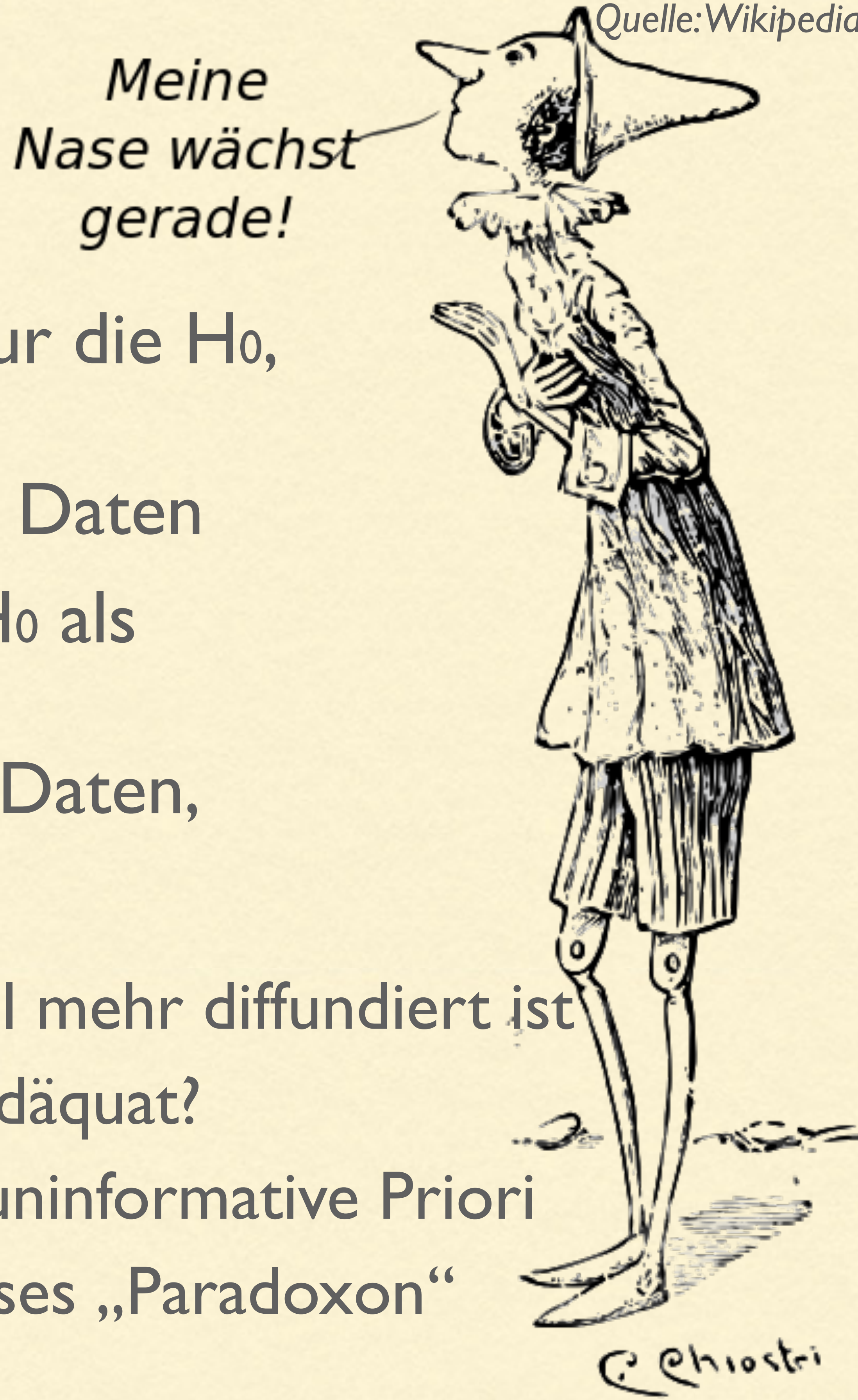
**=> MCMC, INLA**

- Bei komplexen Problemen ist es umso schwieriger, ein passendes Modell (und die Rechtfertigung dafür) zu finden

**=>„All models are wrong, but some are useful" - George Box**

- Frequentistische Methoden gelten als wissenschaftlich objektiv

## Zurück zu Lindley-Paradoxon:
## Ist es überhaupt ein Paradoxon?




Quelle: Wikipedia

$$H_0 : \theta = \theta_0 = 0.5 \text{ vs. } H_1 : \theta \neq 0.5$$

- Frequentistischer Ansatz testet nur die $H_0$, ohne $H_1$ zu referenzieren
  => $H_0$ ist schlechte Erklärung der Daten
- Bayesianischer Ansatz bewertet $H_0$ als vergleich zu $H_1$
  => $H_1$ schlechtere Erklärung der Daten, **gegenüber $H_0$!**
  - $H_0$ wird besser bewertet, da $H_1$ viel mehr diffundiert ist
  - Die Priori ist flach, ist diese Wahl adäquat?
  - Wenn man (objektive Bayesianer) uninformative Priori [0, 1] wählt, dann verschwindet dieses „Paradoxon“

## Was ist Objektivität?

+ Frequentisten gelten als objektiv, da ihre Analysen nur auf Experimente basieren und „kein Vorwissen“ benutzen
    - Modelle der Analyse müssen auch subjektiv ausgewählt werden
    - Experimentdesign hängt stark von Vorwissen und Annahmen ab
    - Fehlerbehaftete Ergebnisse => inwiefern objektiv richtig?

## Bayesianisches Verständnis von Objektivität:

+ Nur abhängig von Modellannahmen (Likelihood und Priori)

+ Jede Wissenschaft besteht aus subjektiven Prozeduren, die jedoch objektiv getestet werden können

Kritischer Rationalismus und Falsifizierbarkeit nach <Logik der Forschung> (1934) von Popper:

- Richtigkeit einer empirisch-wissenschaftlichen Theorie lässt sich nie zeigen, da das Induktionsproblem (die Verallgemeinerung vom Einzelnen auf das Allgemeine) nicht lösbar

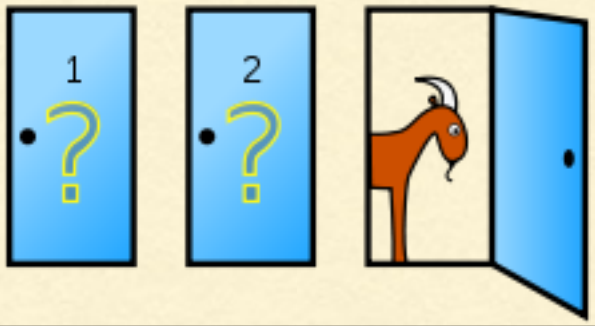
+ Jedoch lassen sich Theorien durch Beobachtungen widerlegen

Logischer Empirismus nach <Logical Fundation of Probability> (1950) von Rudolf Carnap:

Wissenschaftliche Ergebnisse werden mit Wahrscheinlichkeiten versehen. Aus Daten lässt sich der Grad der Bestätigung einer Theorie berechnen.

# Monty-Hall-Dilemma(Das Ziegenproblem/Drei-Türen-Problem):

Nehmen Sie an, Sie wären in einer Spielshow und hätten die Wahl zwischen drei Toren. Hinter einem der Tore ist ein Auto, hinter den anderen sind Ziegen. Sie wählen ein Tor, sagen wir, Tor Nummer 1, und der Showmaster, der weiß, was hinter den Toren ist, öffnet ein anderes Tor, sagen wir, Nummer 3, hinter dem eine Ziege steht. Er fragt Sie nun: ,,Möchten Sie das Tor Nummer 2?" Ist es von Vorteil, die Wahl des Tores zu ändern?



*Quelle: Wikipedia*

$H_1$ : Das Auto ist hinter dem Tor 1

$H_2$ : Das Auto ist hinter dem Tor 2

$H_3$ : Das Auto ist hinter dem Tor 3

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$$

## Bayesianische Bestätigungstheorie (Bayesian Confirmation Theory):

$$P_E(H) := P(H|E) = \frac{P(E|H)\cdot P(H)}{P(E)}$$

$$\begin{aligned} P(E) &= P(E,H) + P(E,\neg H) \\ &= P(E|H)\cdot P(H) + P(E|\neg H)\cdot P(\neg H) \\ \text{oder allgemein:} &= \sum_{i=0}^{n} P(E|H_i)\cdot P(H_i) \end{aligned}$$

E: Das Auto ist nicht hinter dem Tor 3

$H_1$ : Das Auto ist hinter dem Tor 1

$H_2$ : Das Auto ist hinter dem Tor 2

$H_3$ : Das Auto ist hinter dem Tor 3

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$$

$$E := \neg H_3 = H_1 \vee H_2$$

$$P(E|H_1) = 1$$

$$P(E|H_2) = 1$$

$$P(E|H_3) = 0$$

$$P(E) = 1\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} = \frac{2}{3}$$

$$\Rightarrow P_E(H_1) = P_E(H_2) = \frac{1\cdot\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

## Bayesianische Bestätigungstheorie (Bayesian Confirmation Theory):

$$P_E(H) := P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

$$P(E) = P(E, H) + P(E, \neg H)$$

$$= P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)$$

$$\text{oder allgemein: } = \sum_{i=0}^{n} P(E|H_i) \cdot P(H_i)$$

F: Showmaster öffnet Tür 3

$H_1$ : Das Auto ist hinter dem Tor 1

$H_2$ : Das Auto ist hinter dem Tor 2

$H_3$ : Das Auto ist hinter dem Tor 3

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$$

$$P(F|H_1) = \frac{1}{2}$$

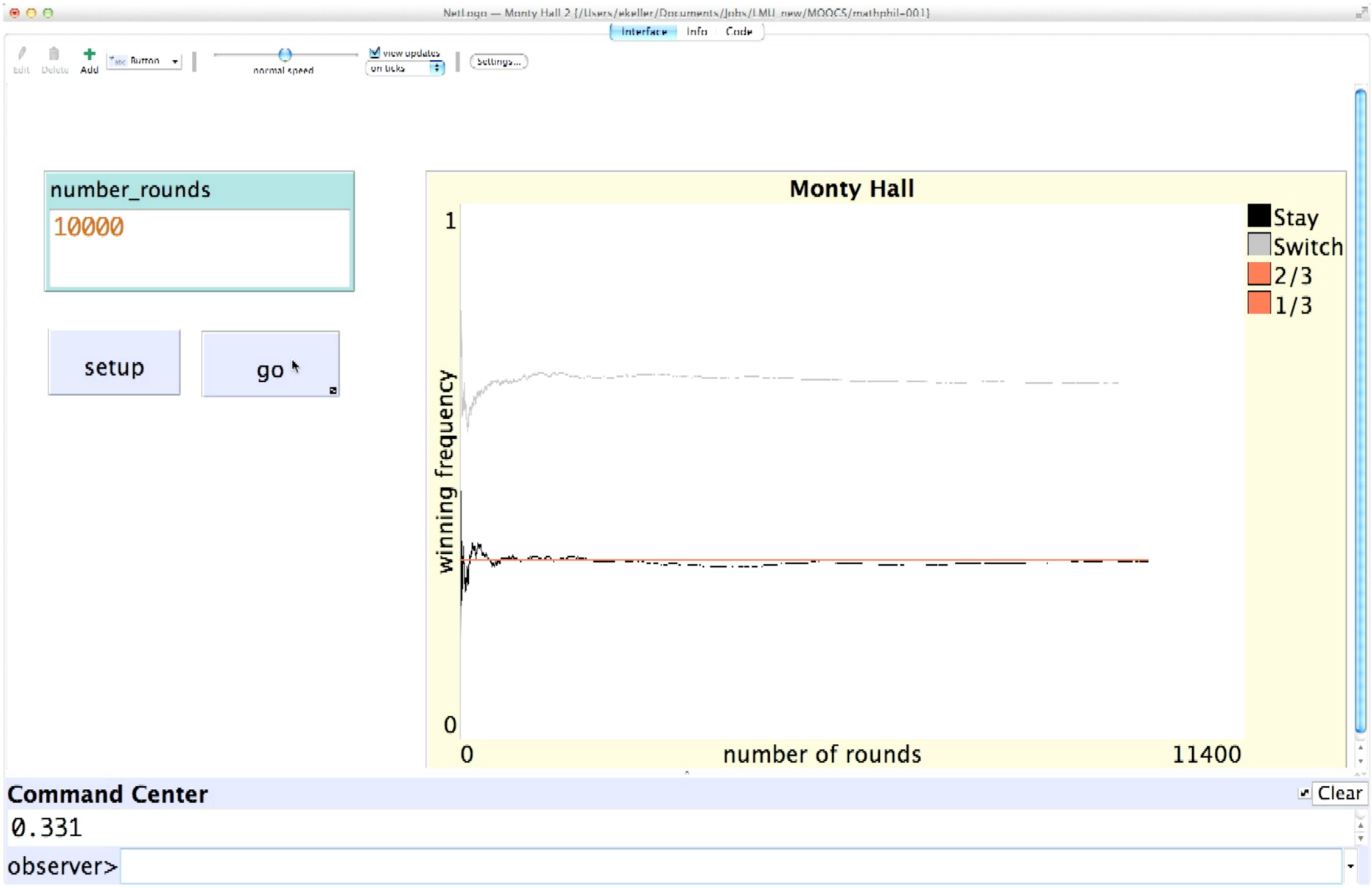
$$P(F|H_2) = 1$$

$$P(F|H_3) = 0$$

$$P(F) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}$$

$$\Rightarrow P_E(H_1) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$\text{und } P_E(H_2) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$


Interface
Info
Code
Edit
Delete
Add
Button
normal speed
view updates
on ticks
Settings...
number_rounds
10000
setup
go
Monty Hall
1
0
winning frequency
Stay
Switch
2/3
1/3
0
number of rounds
11400
Command Center
Clear
0.331
observer>

Bayesianische Bestätigungstheorie (Bayesian Confirmation Theory):

E bestätigt H wenn $P_E$(H) > $P$(H)
E entkräftet H wenn $P_E$(H) < $P$(H)
E ist irrelevant für H wenn $P_E$(H) = $P$(H)

Betrachten wir zuerst die Belege E und F noch mal:

E: Das Auto ist nicht hinter dem Tor 3 vs. F: Showmaster öffnet Tür 3

$$E := \neg H_3 = H_1 \vee H_2$$
$$P(E|H_1) = 1$$
$$P(E|H_2) = 1$$
$$P(E|H_3) = 0$$
$$P(E) = 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{2}{3}$$
$$\Rightarrow P_E(H_1) = P_E(H_2) = \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

$$P(F|H_1) = \frac{1}{2}$$
$$P(F|H_2) = 1$$
$$P(F|H_3) = 0$$
$$P(F) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}$$
$$\Rightarrow P_E(H_1) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$
$$\text{und } P_E(H_2) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Bayesianische Bestätigungstheorie (Bayesian Confirmation Theory):

E bestätigt H wenn $P_E(H) > P(H)$
E entkräftet H wenn $P_E(H) < P(H)$
E ist irrelevant für H wenn $P_E(H) = P(H)$

Betrachten wir zuerst die Belege E und F noch mal:

E: Das Auto ist nicht hinter dem Tor 3 <= F: Showmaster öffnet Tür 3
E ist also logischer Schluss aus F, umgekehrt jedoch nicht! => F>>E

$$E := \neg H_3 = H_1 \vee H_2$$

$$P(E|H_1) = 1$$

$$P(E|H_2) = 1$$

$$P(E|H_3) = 0$$

$$P(F|H_1) = \frac{1}{2}$$

$$P(F|H_2) = 1$$

$$P(F|H_3) = 0$$

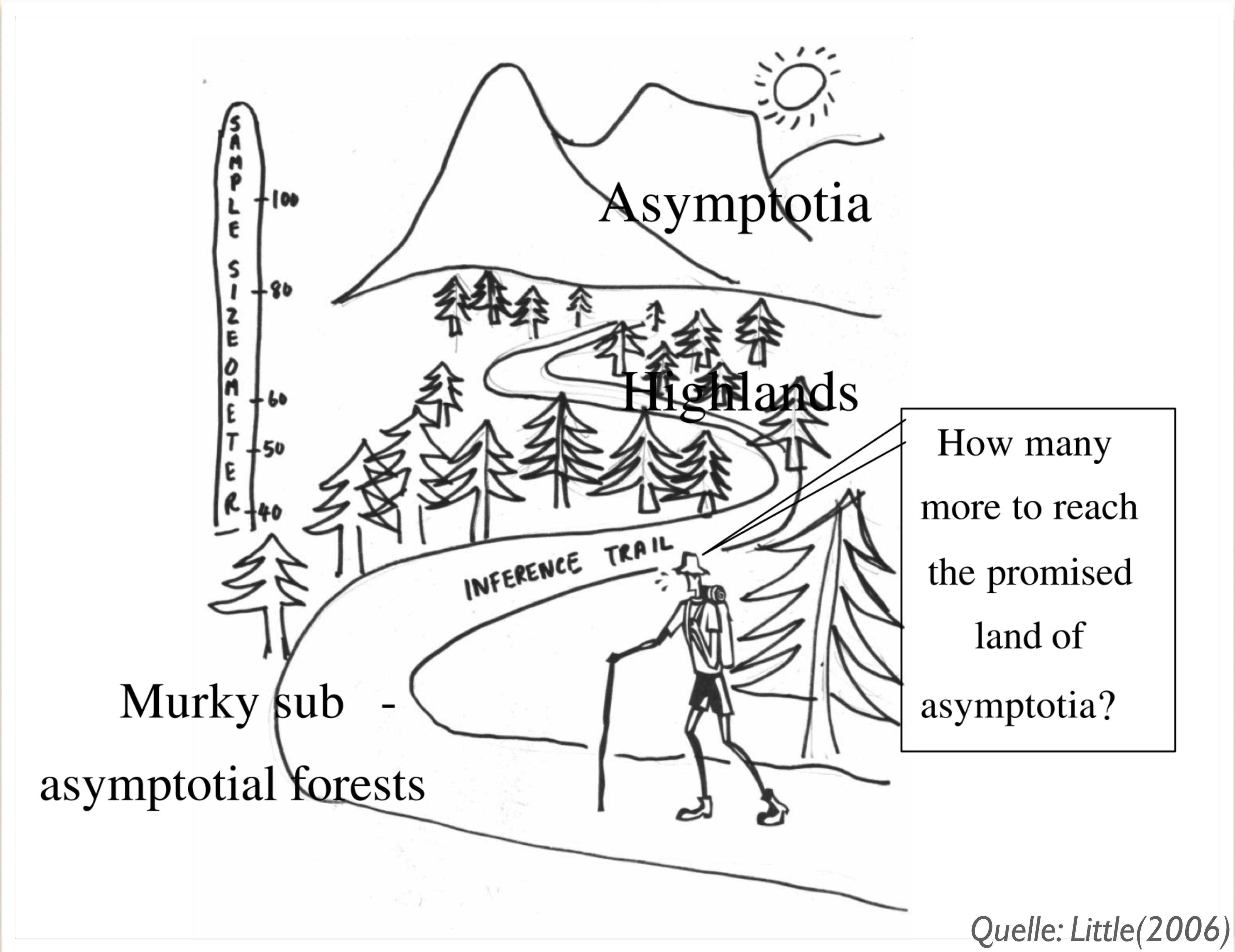
F ist also logisch stärkere Evidenz, die alle relevante Information enthält, während E gewisse Information missachtet.
=> Prinzp der totalen Evidenz

Bayesianische Bestätigungstheorie (Bayesian Confirmation Theory):

$$P_E(H) := P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

E: Das Auto ist nicht hinter dem Tor 3 <= F: Showmaster öffnet Tür 3
E ist also logischer Schluss aus F, umgekehrt jedoch nicht! => F>>E

$$\underbrace{f(\theta|x_1, ..., x_n)}_{Posteriori-Dichte} = \frac{\overbrace{f(x_1, ..., x_n|\theta)}^{Likelihood} \cdot \overbrace{f(\theta)}^{Priori-Dichte}}{f(x_1, ..., x_n)}$$




Quelle: Little(2006)

# Vielen Dank für Ihre Aufmerksamkeit!

**Literaturverzeichnis**:


- Efron, B. (1986) <Why isn’t everyone a Bayesian>, *The American Statistican,Vol/ 40, No. 1*
- Earman, J. (1992) <Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory> *The MIT Press*
- Haigh, J. (2012) <Probability: A Very Short Introduction> *Oxford University Press*
- Howson, C. ;Urbach, P. (2006) <Scientific Reasoning: The Bayesian Approach: The Bayesian Method> 3rd edition, *Open Court Publishing*
- Jeffrey, R. (2004) <Subjective Probability: The Real Thing> *Cambridge University Press*
- Little, R. (2006) <Calibrated Bayes: A Bayes/Frequentist Roadmap>, *The American Statistican,Vol/ 60, No. 3*
- Sprenger, J. (2013) <Testing a Precise Null Hypothesis: The Case of Lindley’s Paradox> *Philosophy of Science*
- Wakefield, J. (2013) <Bayesian and Frequentist Regression Methods>, *Springer*
- *Blog von Andrew Gelman: http://andrewgelman.com*
- *Blog von Christian Robert: http://xianblog.wordpress.com*
- *Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/*