Lügen für Fortgeschrittene
Bayesianische Statistik in der Ökonom(etr)ie

Mathias Moser

Forschungsinstitut Verteilungsfragen
Research Institute Economics of Inequality
WU Wien

AG-Tagung Mathematik
St. Pölten, 17. Februar 2016

Inhalte. Eine Übersicht.

Frequentistische Statistik

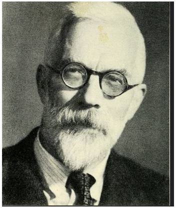
versus

Bayesianische Inferenz

Was wollen die Ökonomen?

Religion Neu:

Reverend Bayes in der Schule



Signifikanz- und Hypothesentests

Ronald Fisher, Neyman, Pearson

Wahrscheinlichkeit: Relative Häufigkeit bei wiederholten Versuchen

---

A-priori Wahrscheinlichkeiten

Thomas Bayes, Laplace, Jeffreys, de Finetti

Wahrscheinlichkeit einer Hypothese gegeben den Daten

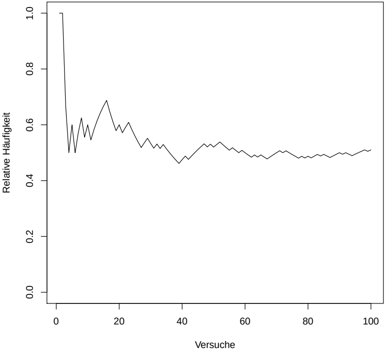
Frequentistische Inferenz I

Wahrscheinlichkeit als Resultat von Experimenten mit Zufallsfehlern (Meßfehler)

Suche nach der relativen Häufigkeit der Eintrittswahrscheinlichkeit,

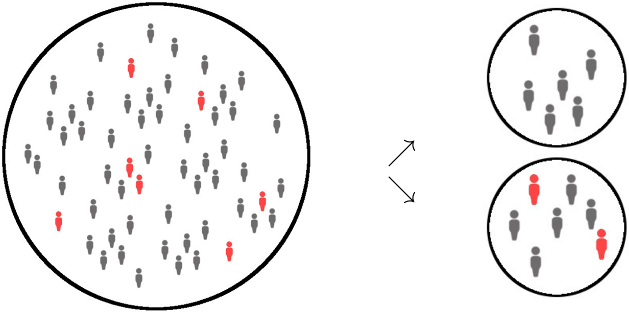
Beispiel:
Kopf (K)/Zahl (Z) beim Münzwurf

$$P(Z) = \frac{\text{Eintritt von } Z}{\text{Anzahl der Experimente}}$$



Frequentistische Inferenz II

Annahme im Experiment: Es gibt einen fixen Parameter θ für die Grundgesamtheit, der a-priori unbekannt ist.
In der Regel beobachten wir nicht die Grundgesamtheit, sondern eine Stichprobe (Sample) davon. Messung ist nicht gleich "wahrem" θ aufgrund von zufälliger Effekte (Messfehler, Samplingfehler).



Frequentistische Inferenz III

Sample: Meta–Experiment
Eine Realisation vieler möglicher Datensätze

Häufige Stichproben $\longrightarrow$ Schätzer für Sample konvergiert gegen wahren Wert der Grundgesamtheit. Es gibt einen wahren Wert — ist (und bleibt) unbekannt.

Konfidenzintervall:
Aussage über das Sample, nicht über Grundgesamtheit. Das wahre θ (ein Wert) hat keine Verteilung, das (mit messfehlern behaftete) Sample schon.

...

Frequentistische Inferenz IV

Schätzen des Parameters: Frequentistische Regression
Kleinstquadarate (OLS)
Maximum-Likelihood Methode:

$$\mathrm{p}(\mathrm{D}|\hat{\theta}) = \frac{1}{\sqrt{2\pi \mathrm{e}^2}} \exp\left[\frac{-(\theta - \hat{\theta})^2}{2\mathrm{e}^2}\right]$$

Estimate $\hat{\theta}$ als bester Punktschätzer: $\mathrm{p}(\mathrm{D}|\hat{\theta}) \longrightarrow$ max.

Bayesianische Inferenz I

Es gibt keinen wahren Wert für θ, sondern "nur" eine Verteilung.
Lerne etwas über $\mathrm{P}(\theta|\mathrm{D})$, gegeben deinem (Vor-)Wissen über $\mathrm{P}(\theta)$ (a-priori Wahrscheinlichkeit)

Frequentistisch

- Daten sind wiederholbare Zufallsziehung
- Parameter sind konstant über Samples
- **Fixe Parameter**

Bayesianisch

- Daten sind eine Realisation/Stichprobe
- Parameter ist unbekannt, probabilistische Beschreibung
- **Fixe Daten**

Konfidenzintervalle

Frequentistisch

- Verwerfen/Akzeptieren H_0 basierend auf α
- Bei wiederholten Experimenten liegt wahrer Parameter in 95% der Fälle im Konfidenzintervall

Bayesianisch

- Wahrscheinliche Werte für Parameter geg. Dichtefunktion
- Für **diese** Daten liegt Parameter zu 95% in diesem Intervall (Integral über Dichte)

Bayesianische Inferenz III

Sollten Sie BayesianerIn sein?

Standardinferenz beschäftigt sich mit Samplewahrscheinlichkeit:

Das 95% Konfidenzintervall enthält den wahren Parameter bevor das Sample bekannt ist. Bei bekanntem Sample ist die Wahrscheinlichkeit 0 oder 1.

Sozialstatistiker meist an post-sample Unsicherheit interessiert:

Wie wahrscheinlich ist der Parameter gegeben den Daten?

Bayes 101

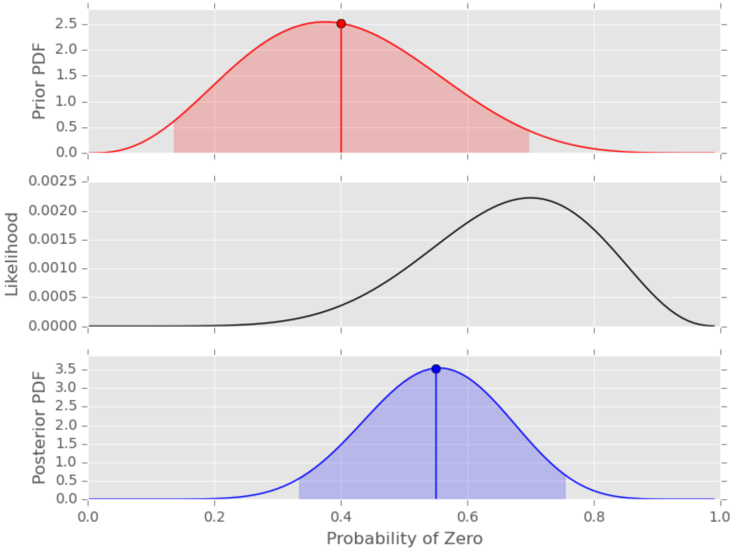
Daten sind fix, der Bevölkerungsdurchschnitt: eine Abstraktion.
Mein Schätzer orientiert sich am Sampledurchschnitt,
wird jedoch von meinem a-priori Wissen beeinflußt.

Satz von Bayes

$$\underbrace{\mathrm{p}(\boldsymbol{\theta}|\mathrm{D})}_{\text{Posterior}} = \frac{\overbrace{\mathrm{p}(\mathrm{D}|\boldsymbol{\theta})}^{\text{Likelihood}} \times \overbrace{\mathrm{p}(\boldsymbol{\theta})}^{\text{Prior}}}{\mathrm{p}(\mathrm{D})} \propto \mathrm{p}(\mathrm{D}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})$$

$\mathrm{p}(\boldsymbol{\theta}|\mathrm{D})$ ist eine Verteilung des Parameters:

Berechnung aller Momente möglich! (inkl. Punktschätzer = Erwartungswert)



Prior PDF
Likelihood
Posterior PDF
Probability of Zero
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025

Subjektivität

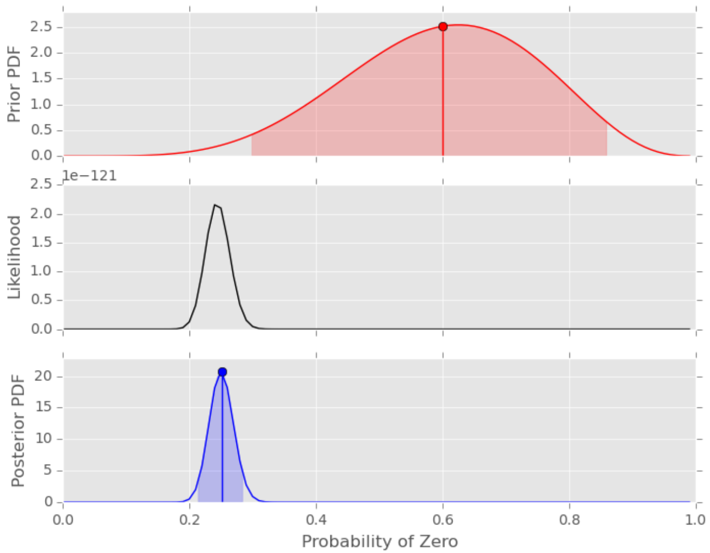
Auswahl der A-Priori Verteilung

Subjektive Einschätzung der Wahrscheinlichkeit eines Ereignisses
Basierend auf Erfahrung, früherer Forschung, Befragungen...

Das ist doch nicht objektiv! ... Naturgemäß.

Explizite Modellierung/Einbeziehung aller (!) Information

Implizite Subjektivität im Frequentismus: Large Sample Properties, Wiederholbare Experimente, ...



Prior PDF
2.5
2.0
1.5
1.0
0.5
0.0
1e−121
Likelihood
2.5
2.0
1.5
1.0
0.5
0.0
Posterior PDF
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
Probability of Zero

> “A frequentist is a person whose long-run ambition is to be wrong 5% of the time.

> “A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.

How-To: Bayesian Analysis I

Inferenz für die Posteriori-Verteilung, das ist Prior $\times$ Likelihood

Aktualisieren der A-Priori Wahrscheinlichkeit: Wir erwarten einen Wert (a-priori), aktualisieren unsere Vermutung durch Bayes Rule und die verfügbaren Daten.
2 Alternativen

(a) Prior und Likelihood gehören zur selben Familie
sind natürlich konjugiert
Analytische Lösung der Posterior,
und damit aller Momente der Verteilung

Bspw. Likelihood: Gauß'sche Verteilung, Prior: Gauß-Gamma
$\rightarrow$ Posterior: Gauß-Gamma

How-To: Bayesian Analysis II

(b) Unterschiedliche Familien, nicht-lineare Modelle, mehrere Parameter
Berechnung des Integrals,

$$\begin{aligned} \mathrm{E}(\boldsymbol{\theta}|\mathrm{D}) &= \int \boldsymbol{\theta}\mathrm{p}(\boldsymbol{\theta}|\mathrm{D})\mathrm{d}\boldsymbol{\theta} \\ &= \frac{\int \boldsymbol{\theta}\mathrm{p}(\mathrm{D}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}{\int \mathrm{p}(\mathrm{D}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \end{aligned}$$

In der Regel: Keine analytische Lösung.
Alternative Strategien: Approximation und **Simulation**

How-To: Bayesian Analysis III

Simulation der Posteriori Verteilung

Erzeugen von unabhängigen Realisationen von $p(\boldsymbol{\theta}|D)$

$\rightarrow$ Monte Carlo Simulation

Bei komplexen Funktionen schwierig, IID (independent, identically distributed) Samples zu bekommen

Markov Chain Monte Carlo: Kette von simulierten Realisationen der Posterior $p(\boldsymbol{\theta}^{t+1}|\boldsymbol{\theta}^{t})$

Wenn MCMC bestimmte Voraussetzungen erfüllt: Konvergenz der Samples zu stationärer Verteilung (Posterior)

$\rightarrow$ Markov Chain Monte Carlo

Endstand: Frequentisten vs. Bayesianer

	Vorteile	**Nachteile**
Frequentisten	"Objektivität"	Interpretation
	Implementierbarkeit	Annahmen
Bayesianer	Interpretation	"Subjektivität"
	Komplexe Modelle	Rechenintensiv
	A-prioris	

Ökonomen als Bayesianer?

Beispiele aus der Ökonom(etr)ie I

Starkes A-Priori Wissen

Finanzmarktanalysen
: Modellierung extremer Ereignisse → grundsätzlich aber "Glaube" an Effizienzmarkthypothese (Prior)

Behavioral Economics
: Beschreibung/Simulation von "Agenten", die neben Verhaltensregeln (Nutzenmaximierung) auch a-priori Wissen über andere Agenten einzbeziehen

Komplexe Modelle
: Flache Likelihood + Expertenwissen

Annahmen
: Annahmen Teil der Modellierung

Beispiele aus der Ökonom(etr)ie II

Unsicherheit

- Selten *ein wahrer* Parameter
- Abhängig von Faktoren jenseits der Modellierbarkeit

$\neq$ Exakte Wissenschaft

- Parameterintervalle
- Beschreibung der Unsicherheit von ökonomischen Prognosen

Beispiele aus der Ökonom(etr)ie III

Lösung Komplexer Modelle

- Einfache, numerische Algorithmen für komplexe Probleme
- Vereinfachung: *sparse models, noisy data* (bspw: Spike&Slab Priors)
- Hochdimensionale Wahrscheinlichkeitsräume (MCMC)
- Klassifikation von großen Datenmengen (*unsupervised learning*)

Warum Sie Bayes unterrichten sollten

a. Ihre SchülerInnen sind bereits BayesianerInnen
(ohne es zu wissen)
b. Es steht im Lehrplan der sechsten Klasse

“ *Wenn der p-Wert klein ist, dann ist es nicht zufällig passiert*

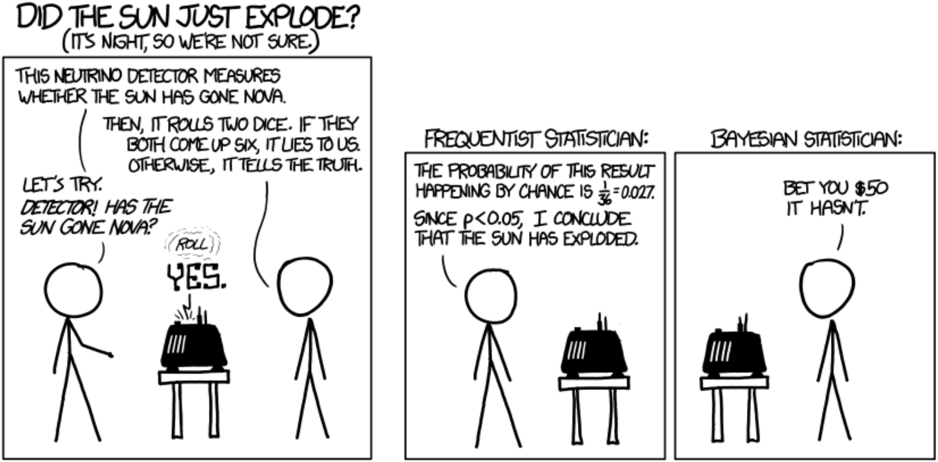
“ *Der p-Wert ist die Wahrscheinlichkeit, dass die Nullhypothese wahr ist*

“ *Der Wert liegt mit 95% Wahrscheinlichkeit im Intervall* $[0, 1]$

“ *In 5% der Fälle ist die Nullhypothese falsch*

Implizite BayesianerInnen

- Intuitiver Fokus auf $\mathrm{p}(\theta|\mathrm{D})$
  ("*Der Wert ist...wahrscheinlich...*")
- Interpretation der Nullhypothese als Verteilung (!)
- Schwieriges Verständnis arbiträrer Signifikanzparameter (α)
- Abstraktion: Grundgesamtheit/Stichprobe/Wiederholtes Ziehen
  $\rightarrow$ Direkte Aussage über Beobachtung
- Lernen, a-priori Wissen zu ignorieren: Würfelspiel, Münzwurf, ...

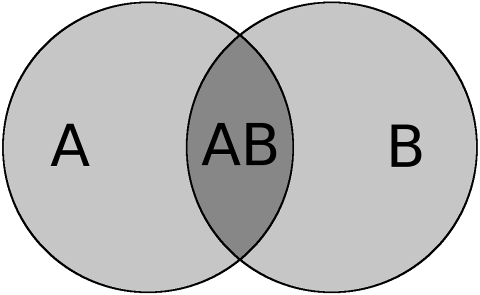

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}=0.027$.
SINCE $p<0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Wie?

Venn Diagramme

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

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$



P(A): Die A-priori Wahrscheinlichkeit

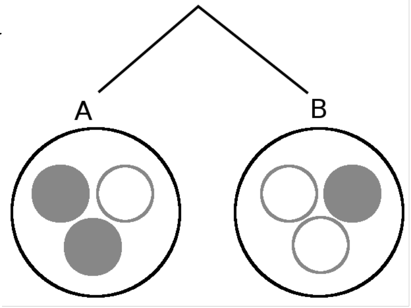
Zweistufiges Sampling: Zwei Urnen, unterschiedliche Wahrscheinlichkeiten, welche Urne wählt der Vortragende?

z.B. Gleiche a-priori Wahrscheinlichkeit

Wahrscheinlichkeit, dass schwarze Kugel aus Urne A?

$$\frac{1/2 \times 2/3}{(1/2 \times 2/3)+(1/2 \times 1/3))}$$

$\rightarrow$ Variieren der a-priori Wahrscheinlichkeit



Vortrags-Posteriori

- Koexistenz Frequentist/Bayesianer
- Hervorheben unterschiedlicher Annahmen
- Anwendungsgebiete: Experiment (vs.) Sozialstatistik
- Inuitiver Zugang zu Statistik?

Literatur

- Holt/Anderson: **Understanding Bayes' Rule**
- Sims (2007): **Bayesian Methods in Applied Econometrics**
- Lam: **Bayesian Statistics in One Hour**
- Ekstrøm: **Frequentist and Bayesian Statistics**