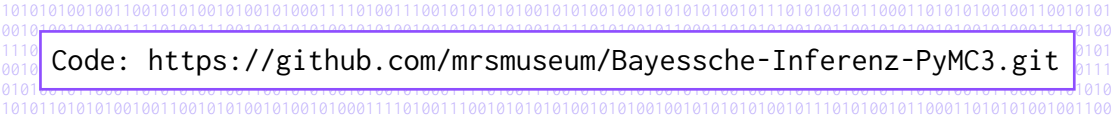
Julius-Maximilians-Universität Würzburg
Institut für deutsche Philologie
Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte
Seminar: Statistische Modellierung (04063632)

Dozent: Herr Dr. Steffen Pielström
Referentin: Madlin Marenec
Wintersemester 2022/23
02.02.2023

Bayes'sche Inferenz / PyMC3



Bayes'sche Inferenz

- benannt nach dem englischen Mathematiker Thomas Bayes (*1701, †1761)
- man geht vorab mit einer bestimmten Erwartung bzw. Vorannahme an statistische Auswertungen heran
- im Experiment selbst wird diese Vorannahme dann überprüft
- nach dem Experiment wird diese Vorannahme durch das Ergebnis des Experiments geupdatet
- statt Ja/Nein Fragen zu stellen wird getestet, wie *wahrscheinlich* Ja ODER Nein ist
- besonders charakteristisch: man modelliert Wahrscheinlichkeits- bzw. Randverteilungen

Bayes'scher Wahrscheinlichkeitsbegriff

Wahrscheinlichkeit im Sinne der Bayes'schen Inferenz ist definiert als…
- die persönliche Überzeugung, dass ein bestimmtes Ereignis in einem bestimmten Experiment eintritt
- Überzeugung, welche sich aus der bisherigen Erfahrung sowie dem Ergebnis des Experiments herausbildet
- Maß, für dessen Glaubwürdigkeit einer Aussage gilt: 0 (falsch, unglaubwürdig) bis 1 (glaubwürdig, wahr)

→ man spricht hierbei vom **Grad persönlicher Überzeugung**

Begriffe

A-priori-Wahrscheinlichkeit (Prior)
Wahrscheinlichkeitsverteilung einer Zufallsgröße aufgrund begründeter Vorannahmen bzw. durch gegebenes Vorwissen $\mathscr{I}$ („vor dem Blick auf die Daten“): $\Pr(\mathscr{M} \mid \mathscr{I})$

Achtung! Der Prior sollte so neutral wie möglich gewählt sein

A-posteriori-Wahrscheinlichkeit (Posterior)
Wahrscheinlichkeitsverteilung nach Beobachtung der Zufallsgröße bzw. unter Einbezug des Vorwissens $\mathscr{I}$ und der Messdaten $\mathscr{D}$ („nach dem Blick auf die Daten“): $\Pr(\mathscr{M} \mid \mathscr{D}, \mathscr{I})$

Likelihood
Wahrscheinlichkeitsverteilung für die Messdaten $\mathscr{D}$, wenn der Modellparameter $\mathscr{M}$ und das Vorwissen $\mathscr{I}$ gegeben sind: $\Pr(\mathscr{D} \mid \mathscr{M}, \mathscr{I})$

Julius-Maximilians-Universität Würzburg
Institut für deutsche Philologie
Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte
Seminar: Statistische Modellierung (04063632)

Dozent: Herr Dr. Steffen Pielström
Referentin: Madlin Marenec
Wintersemester 2022/23
02.02.2023

Satz von Bayes (Bayes Theorem)

$$\underbrace{\Pr(\mathscr{M}\,|\,\mathscr{D},\mathscr{I})}_{\text{Posterior}} = \frac{\overbrace{\Pr(\mathscr{D}\,|\,\mathscr{M},\mathscr{I})}^{\text{Likelihood}}\ \overbrace{\Pr(\mathscr{M}\,|\,\mathscr{I})}^{\text{Prior}}}{\underbrace{\Pr(\mathscr{D}\,|\,\mathscr{I})}_{\text{Normierungsfaktor}}}$$

Implementierung in Python mit PyMC3

- PyMC3: Python-Bibliothek (Open Source) für probabilistische Programmierung → spezialisiert auf fortgeschrittene Markov-Chain-Monte-Carlo- (MCMC) und Variationsanpassungs-Algorithmen
- Warum MCMC? Um mit leistungsstarken Stichprobenalgorithmen die A-Posteriori-Wahrscheinlichkeit zu berechnen (z. B. No-U-Turn Sampler (NUTS), Metropolis-Hastings, Sequenzielle Monte-Carlo-Methoden)
- ArviZ: Python-Bibliothek zur Datenvisualisierung; vor allem für die Interpretation und Visualisierung von Posterior-Verteilungen

Hausaufgaben

Ein Würfel hat sechs Seiten mit je sechs verschiedenen Augenzahlen.
Es soll untersucht werden, mit welcher Wahrscheinlichkeit beim Würfeln die Zahl 6 fällt.

Es gilt:
0 = Misserfolg, es wurde keine 6 gewürfelt
1 = Erfolg, es wurde eine 6 gewürfelt

1. Mit welcher Vorannahme geht ihr in das Experiment hinein? Stellt die A-priori-Wahrscheinlichkeit auf und begründet diese kurz.
2. Generiert nun die Daten für den Würfelwurf. Dabei soll eure A-priori-Wahrscheinlichkeit in der Variable p_six gespeichert werden und es soll zu Anfang 10x gewürfelt werden.

3. Auf Grundlage der generierten Daten implementiert nun das passende PyMC3 Modell. Visualisiert anschließend die A-posteriori-Wahrscheinlichkeit. Was ist der Mittelwert und der Wert für die Standardabweichung?

Beispiel: Münzwurf

PyMC3 installieren

In [1]:
```
#conda create -c conda-forge -n pymc_env "pymc>=4"
#conda activate pymc_env
```

Importe

In [2]:
```
import numpy as np
import scipy.stats as stats
import pymc3 as pm
import arviz as az
```

Daten generieren

In [3]:
```
trials = 4
p = 0.50
```

In [4]:
```
tosses = stats.bernoulli.rvs(p=p, size=trials)
tosses
```

Out[4]:
```
array([0, 1, 0, 1])
```

PyMC3 Modell implementieren

In [5]:
```
with pm.Model() as first_model:

    #Prior defninieren
    prior = pm.Beta('prior', 1, 1)

    #Likelihood definieren
    data = pm.Bernoulli('data', p=prior, observed=tosses)

    #Stichprobenziehung
    trace = pm.sample(1000, random_seed=123, return_inferencedata=True)
```

```
Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [prior]
```

100.00% [4000/4000 00:03<00:00 Sampling 2 chains, 0 divergences]

```
/Users/madlinma/opt/anaconda3/lib/python3.9/site-packages/scipy/stats/_continuous_distns.py:624: RuntimeWarning: ov
erflow encountered in _beta_ppf
  return _boost._beta_ppf(q, a, b)
/Users/madlinma/opt/anaconda3/lib/python3.9/site-packages/scipy/stats/_continuous_distns.py:624: RuntimeWarning: ov
erflow encountered in _beta_ppf
  return _boost._beta_ppf(q, a, b)
Sampling 2 chains for 1_000 tune and 1_000 draw iterations (2_000 + 2_000 draws total) took 14 seconds.
```

Daten visualisieren

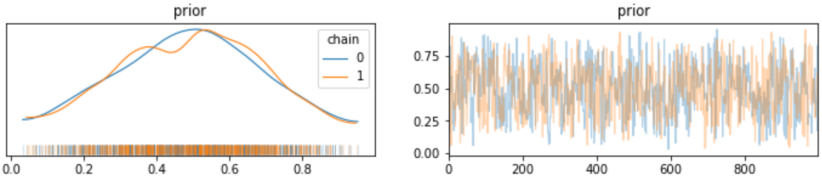
In [6]:
```
az.summary(trace, kind=('stats'))
```

Out[6]:

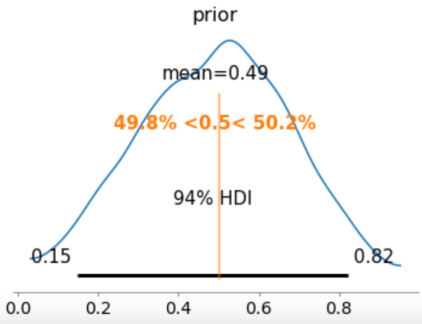
	mean	sd	hdi_3%	hdi_97%
prior	0.491	0.185	0.146	0.823

In [7]:
```
az.plot_trace(trace, rug=True, rug_kwargs={"alpha": .2}, compact=False, legend=True)
```

Out[7]:
```
array([[<AxesSubplot:title={'center':'prior'}>,
        <AxesSubplot:title={'center':'prior'}>]], dtype=object)
```



```
In [8]: az.plot_posterior(trace, ref_val=0.5)
Out[8]: <AxesSubplot:title={'center':'prior'}>
```



Literatur

- Juola, Patrick/Ramsay, Stephen: Six Septembers. Mathematics for the Humanist. Nebraska 2017. Online verfügbar unter: https://digitalcommons.unl.edu/zeabook/55/ [Stand: 30.01.2023].
- McElreath, Richard: Statistical Rethinking. A Bayesian Course with Examples in R and STAN. Boca Raton/Oxon 2020. Online verfügbar unter: https://ebookcentral.proquest.com/lib/ub-wuerzburg/detail.action?docID=6133700 [Stand: 19.01.2023]

Tutorials

- „Statistical Rethinking 2022" Vorlesung von Richard McElreath. Online verfügbar unter: https://www.youtube.com/watch?v=BYUykHScxj8&list=PLDcUM9US4XdMROZ57-OIRtIK0aOynbgZN
- Jupyther Notebooks „Statistical Rethinking: A Bayesian Course Using python and pymc3" zur Vorlesung „Statistical Rethinking 2022". Online verfügbar unter: https://github.com/gbosquechacon/statrethink_course_in_pymc3
- Jupyter Notebook „Probabilistic Programming and Bayesian Methods for Hackers". Online verfügbar unter: https://nbviewer.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter1_Introduction/Ch1_Introduction_PyMC3.ipynb