



**FAKULTA  
JADERNÁ  
A FYZIKÁLNĚ  
INŽENÝRSKÁ  
ČVUT V PRAZE**



**COMPASS**

INSTITUTE OF PLASMA PHYSICS ASCR

# **BAYESOVSKÉ METODY ANALÝZY DAT A THOMSONŮV ROZPTYL NA TOKAMAKU COMPASS**

**STUDENTSKÁ KONFERENCE FTTF**

**JAN HEČKO**

**ŠKOLITEL: MIROSLAV ŠOS**

**20.-24.1.2020 / MARIÁNSKÁ**



# BAYESOVSKÉ METODY MOTIVACE

- Zpracování dat z experimentu – „Běžný způsob“

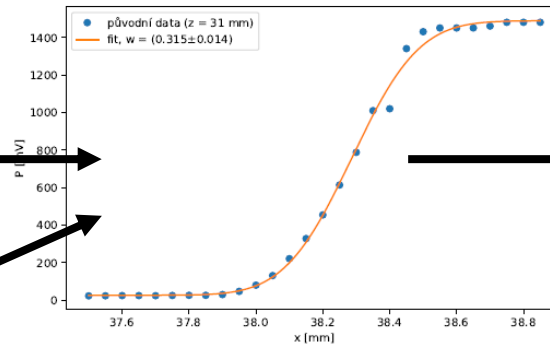
$x$	$P$
37,80	27,0
37,75	26,0
37,70	
37,65	
...	...

Data

Fyzikální  
model

$$P = P(x; w)$$

$w$  ... parametr



Fitování dat modelem

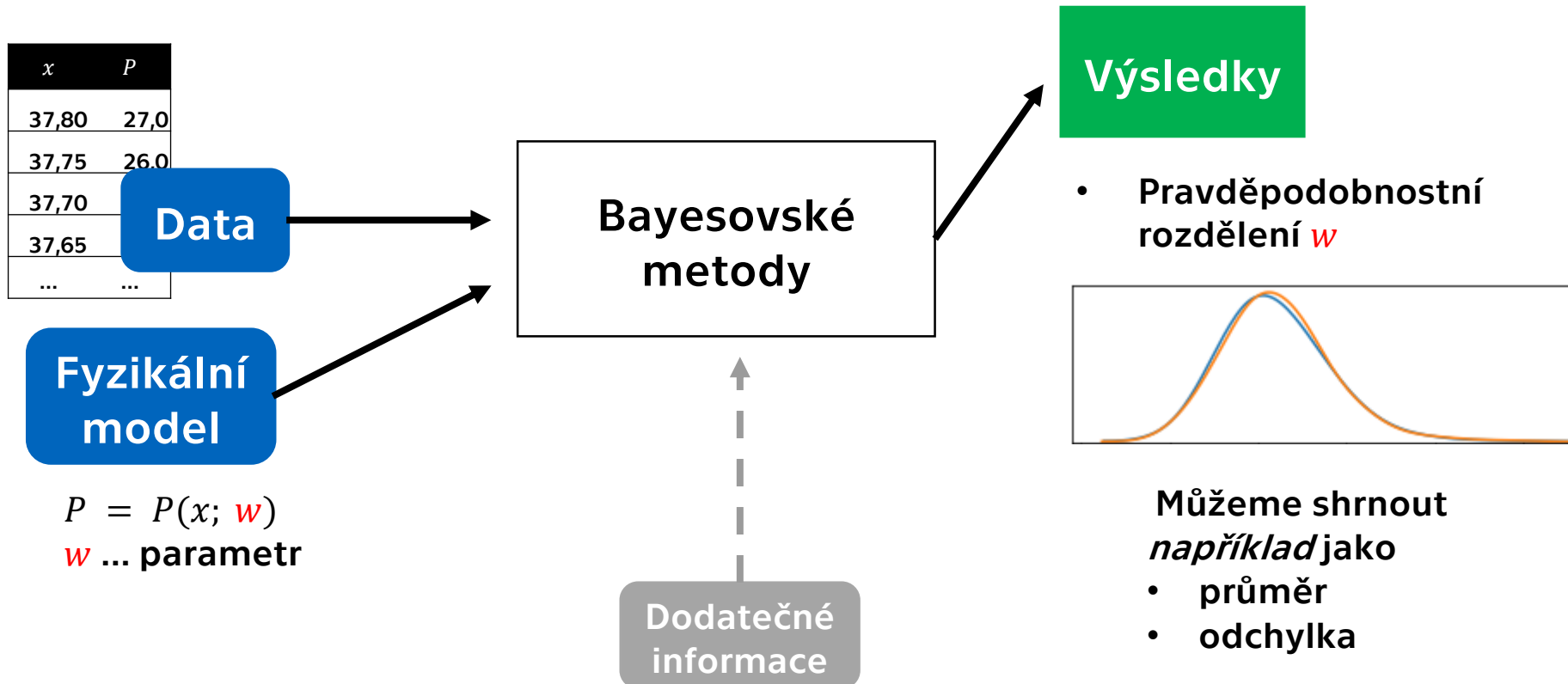
Výsledky

- Průměr  $w$
  - Odchylka  $w$
- = dvě čísla!



# BAYESOVSKÉ METODY MOTIVACE

- Zpracování dat z experimentu – Alternativní přístup





**1. BAYESOVA VĚTA**

**2. PŘÍKLADY**

**3. NUMERIKA**

**4. KNIHOVNA PYMC3 (PYTHON)**

**5. POUŽITÍ PRO THOMSONŮV ROZPTYL**



# 1. BAYESOVA VĚTA

Podmíněná pravděpodobnost:

$A$  ... dnes bude pršet

$$P(A) = ?$$

- Základní odhad „bez pohledu z okna“
- Např. počet deštivých dnů v roce

$$P(A|B) = ?$$

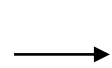
- Získáme novou informaci => přehodnotíme svůj názor
- $B$  ... mraky, fouká vítr, včera také pršelo



# 1. BAYESOVA VĚTA

- $A$  ... Měřená veličina
- $D$  ... Data

Posterior  
probability

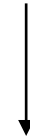


$P(A|D)$

=

$$\frac{P(D|A) \cdot P(A)}{P(D)}$$

Likelihood of  
observations



Prior probability



Normalizing  
constant



$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

## 2. PŘÍKLADY

Dnes bude pršet ...  $A$ . Podle meteorologických statistik:

**Prior:**  $P(A) = \frac{117}{365} \doteq 0,32$

Nová data  $D$ : Jsou venku mraky?

- Podle zkušenosti, v 9 z 10 případů jsou za deště vidět mraky a 3 z 10 případů jsou vidět, i když ten den neprší.

**Likelihood:**  $P(D|A) = 0,9$  a  $P(D|\neg A) = 0,3$

Změna předpovědi na základě dat:

**Posterior:**  $P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)} = \frac{0,9 \cdot 0,32}{0,9 \cdot 0,32 + 0,3 \cdot 0,68} \doteq 0,59$



$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

## 2. PŘÍKLADY

### Lineární regrese

Např.  $a, b > 0$

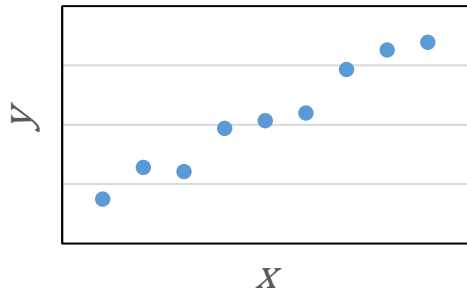
$P(a, b)$  Prior

- Mějme závislost

$$y = ax + b$$

$P(D|a, b)$  Likelihood

- Data z experimentu



- Chceme určit parametry  $a, b$



$P(a, b|D)$  Posterior

= **Výsledek**  
(pravděpodobnostní  
rozdělení  $a, b$ )





$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

## 2. PŘÍKLADY

### Volba **prior**?

- Volíme pravděpodobnostní rozdělení  $A$
- Různé přístupy: Obecný vs. Hodně informativní
- Otázka subjektivity

### Gaussovské procesy

- Pokročilejší **prior**
- Hledané řešení není hodnota ale spojitá funkce
- Např. prostorové profily



$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

## 3. NUMERIKA

Spojité  $A$ : pravděp. rozdělení je možné jen aproximovat.

### Posterior sampling

- MCMC = Markov chain Monte Carlo algoritmy
- Výběr hodnot  $A$  s cílem získat reprezentativní vzorek

### Konkrétní algoritmy

- Metropolis-Hastings ~ 1953 / 2003
- Hamiltonian Monte Carlo (HMC) ~ 1987 / 2010
- No-U-turn Sampler (NUTS) ~ 2011



## **4. KNIHOVNA PYMC3**

### **Rozhraní pro sestavení Bayesovského modelu**

- Implementace různých pravděpodobnostních rozdělení (Uniform, Normal, Gamma, etc.)

### **Implementace MCMC algoritmů**

- Nejběžněji NUTS (autotuning)
- Možnost paralelizace

### **Nástroje pro vizualizaci a analýzu výsledků**

- Související knihovna ArviZ

## Example : Linear fit

First we generate random data:

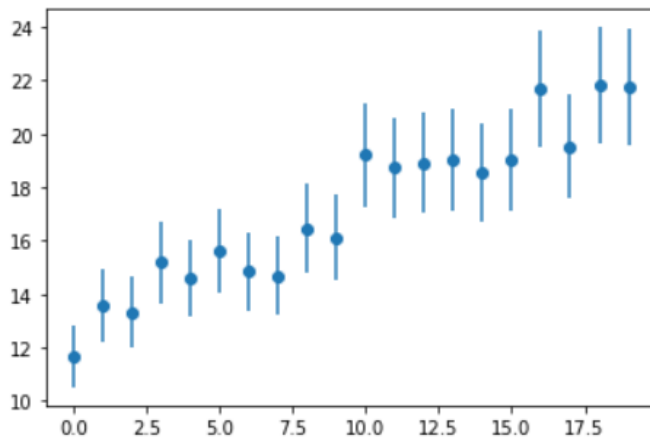
```
[1]: import numpy as np
import matplotlib.pyplot as plt

size = 20

x_data = np.arange(size)
y_data = 0.5 * x_data + 12.5 + 1.0*np.random.normal(size=size)
# y_data[:,2] = 2.5 * x_data[:,2] + 14.5 + 7.0*np.random.rand(size//2)
# Let't assume 10% realtive measurement error
y_err = y_data * 0.1

fig, ax = plt.subplots()
ax.errorbar(x_data, y_data, fmt='o', yerr=y_err/10.0)

plt.show()
```





Then the model is defined:

```
[2]: from pymc3 import Model, Uniform, Normal, Deterministic, sample
      from arviz import plot_trace

      # Set the context of our model
      with Model() as linfit_model:
          # Free RVs (What we want)
          A = Uniform('A', lower=-10, upper=10)
          B = Uniform('B', lower=-20, upper=20)

          # Observed RV (What we measure/observe)
          # - Here we define the relationship between y and A, B
          mu = A * x_data + B # NOTE: Part of a relationship are the `x_data`
          y = Normal('y', mu=mu, sigma=y_err*np.sqrt(len(y_data)), observed=y_data)
```

Sample the model:

```
[3]: with linfit_model:
      trace = sample(draws=4000, tune=1000, chains=4, cores=1)





      # Plot results (estimated probability density functions)
      plot_trace(trace);
```

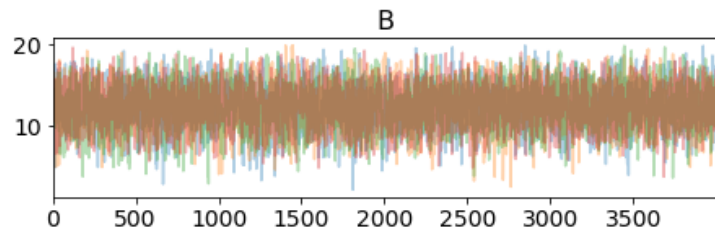
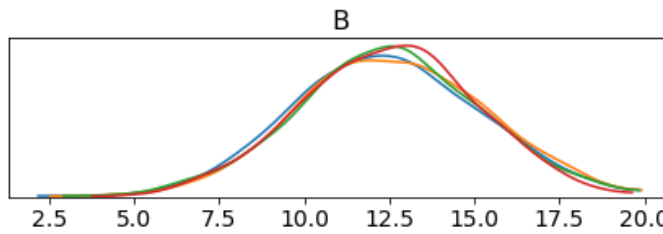
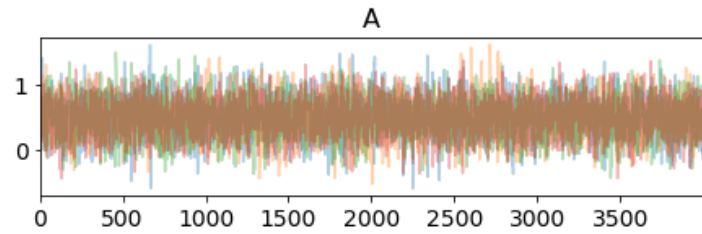
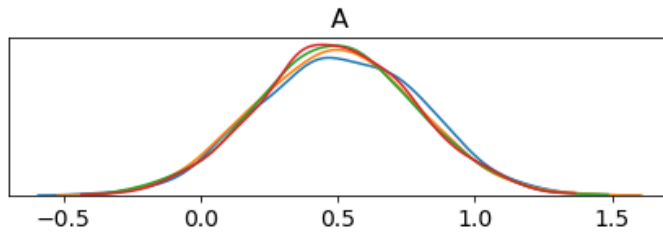
Auto-assigning NUTS sampler...

Initializing NUTS using jitter+adapt\_diag...

Sequential sampling (4 chains in 1 job)

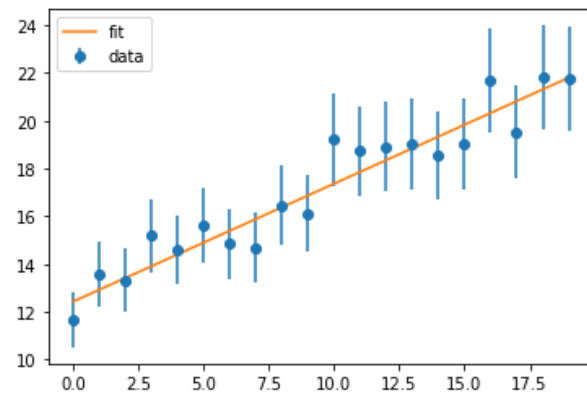
NUTS: [B, A]

100%			5000/5000	[00:04<00:00, 1221.10it/s]
100%			5000/5000	[00:03<00:00, 1399.24it/s]
100%			5000/5000	[00:04<00:00, 1233.42it/s]
100%			5000/5000	[00:03<00:00, 1480.64it/s]



And finally, plot the results:

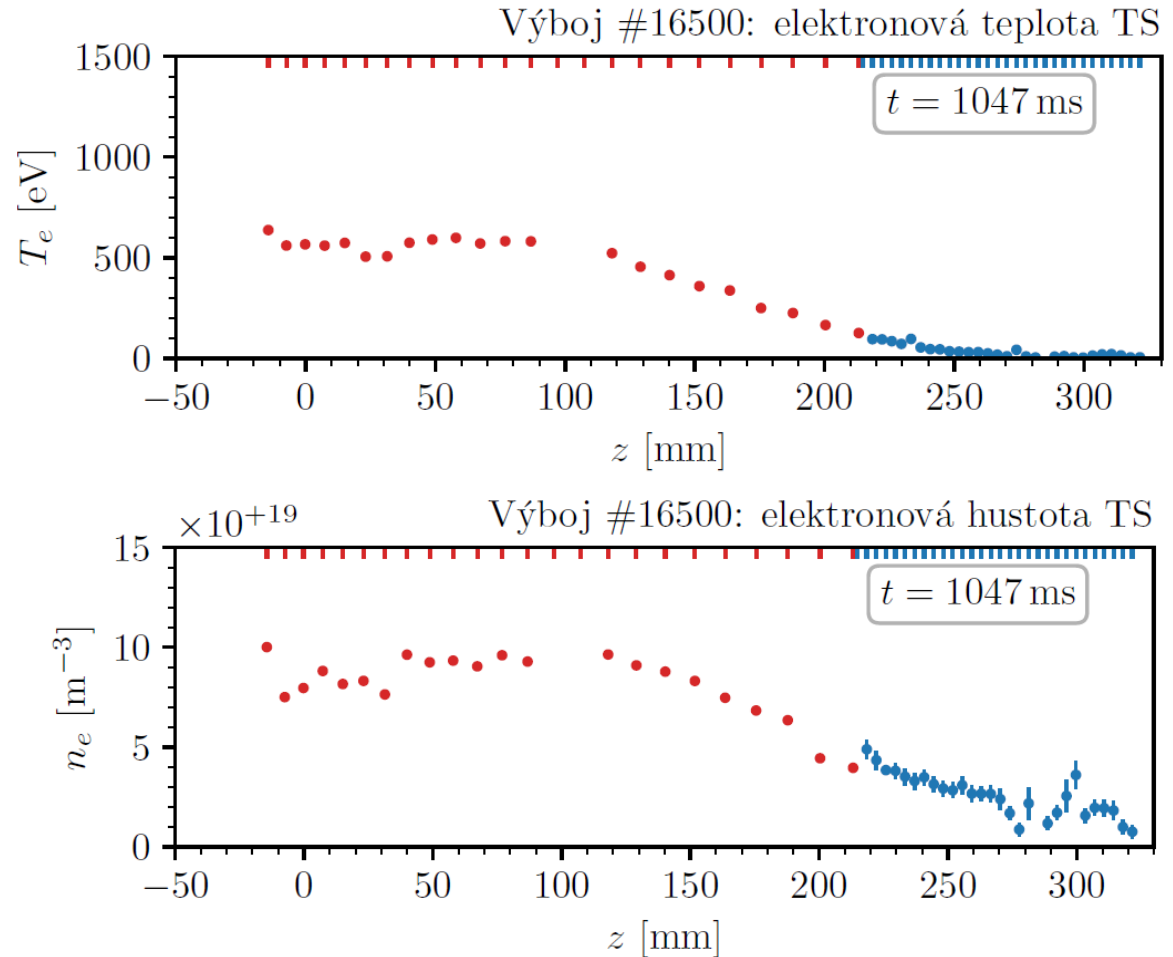
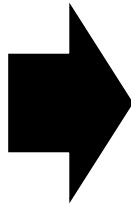
...





# 5. POUŽITÍ PRO THOMSONŮV ROZPTYL (TS)

Analýza  
spektra  
rozptylu







# 5. POUŽITÍ PRO THOMSONŮV ROZPTYL (TS)

Alternativa k současné metodě zpracování naměřených dat a určení profilů  $T_e$  a  $n_e$

Model diagnostiky TS:

- Signál v  $i$ -tém kanálu  $V_{TS}^{(i)} = V_{TS}^{(i)}(n_e, T_e)$   
Likelihood
- Předpoklady pro  $T_e$  a  $n_e$ ?
  - Nezáporné
  - Řádový odhad
  - ...Prior



# 5. POUŽITÍ PRO THOMSONŮV ROZPTYL (TS)

Další možné cíle:

## A) Využití Gaussovských procesů

- Profil je funkce, uvažujeme korelaci bodů v prostoru
- Korelace v časové dimenzi

## B) Sloučení výsledků z různých diagnostik

- Měření  $n_e$ : TS a interferometr



## **ZDROJE**

Salvatier J., Wiecki T.V., Fonnesbeck C. Probabilistic programming in Python using PyMC3, *PeerJ Computer Science*, 2:e55, 2016

- <https://docs.pymc.io/>

MCMC with PyMC3, Fonnesbeck C., PyData London 2019 (conference)

- [https://youtu.be/SS\\_pqgFziAg](https://youtu.be/SS_pqgFziAg)
- [https://github.com/fonnesbeck/mcmc\\_pydata\\_london\\_2019](https://github.com/fonnesbeck/mcmc_pydata_london_2019)

Von Nessi G. T., Hole M. J., Using Bayesian analysis and Gaussian processes to infer electron temperature and density profiles on the Mega-Ampere Spherical Tokamak experiment, *Review of Scientific Instruments*, 84(6), 2013

Fisher R. et al, Bayesian modelling of fusion diagnostics, *Plasma Phys. Control. Fusion*, 45(7), 2003