Should this Drug be Approved? A Bayesian's Answer with Stan

Konstantinos Vamvourellis[1] Marianne Corvellec[2]

[1]Department of Statistics at LSE

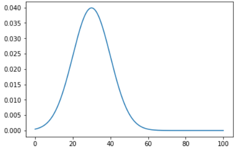
[2]IGDORE

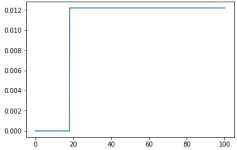
SciPy (Austin, TX)
July 11, 2018

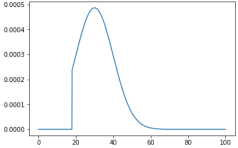
Another Introduction to Bayesian Inference

1. Build a Bayesian statistical model

Data $\mathcal{D}$ Parameter θ







likelihood $p(\mathcal{D}|\theta)$ $\times$ **prior** $p(\theta)$ $=$ **posterior** $p(\theta|\mathcal{D})$

2. Compute expectations

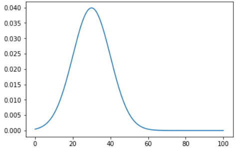
$$\mathbb{E}[a] = \int a(\theta)p(\theta|\mathcal{D})\mathrm{d}\theta$$

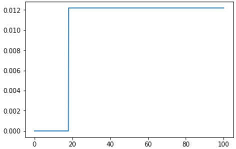
(HMC, etc.)

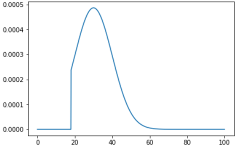
Another Introduction to Bayesian Inference

1. Build a Bayesian statistical model

Data $\mathcal{D}$ Parameter θ







likelihood $p(\mathcal{D}|\theta)$ $\times$ **prior** $p(\theta)$ $=$ **posterior** $p(\theta|\mathcal{D})$

2. Compute expectations

$$\mathbb{E}[a] = \int a(\theta) p(\theta|\mathcal{D}) \mathrm{d}\theta$$

(HMC, etc.)

Clinical Trial Data

Subject ID	Group Type	Glucose	Dyspepsia	Nausea
123	Control	-3.42	1	0
231	Treatment	-4.41	1	0
221	Control	1.2	1	1
222	Treatment	-0.2	0	1
322	Treatment	-2.7	0	0

Clinical Trial Data

Subject ID	Group Type	Glucose	Dyspepsia	Nausea
123	**Control**	**-3.42**	**1**	**0**
231	Treatment	-4.41	1	0
221	**Control**	**1.2**	**1**	**1**
222	Treatment	-0.2	0	1
322	Treatment	-2.7	0	0

Clinical Trial Data

Subject ID	Group Type	Glucose	Dyspepsia	Nausea
123	Control	-3.42	1	0
231	**Treatment**	**-4.41**	**1**	**0**
221	Control	1.2	1	1
222	**Treatment**	**-0.2**	**0**	**1**
322	**Treatment**	**-2.7**	**0**	**0**

Model

- Split data in two groups

$$Y_C = \begin{bmatrix} -3.42 & 1 & 0 \\ 1.2 & 1 & 1 \\ & \dots & \end{bmatrix}, \; Y_T = \begin{bmatrix} -4.41 & 1 & 0 \\ -0.2 & 0 & 1 \\ & \dots & \end{bmatrix}$$

- Model each group separately and retrieve two joint posterior distributions to compare

$$p(\theta_C | Y_C) \,, \;\; p(\theta_T | Y_T)$$

$$p(\{\mu, \sigma, R\}_C | Y_C) \,, \;\; p(\{\mu, \sigma, R\}_T | Y_T)$$

Model

- Split data in two groups

$$Y_C = \begin{bmatrix} -3.42 & 1 & 0 \\ 1.2 & 1 & 1 \\ & \dots & \end{bmatrix}, \; Y_T = \begin{bmatrix} -4.41 & 1 & 0 \\ -0.2 & 0 & 1 \\ & \dots & \end{bmatrix}$$

- Model each group separately and retrieve two joint posterior distributions to compare

$$p(\theta_C | Y_C) \, , \;\; p(\theta_T | Y_T)$$

$$p(\{\mu, \sigma, R\}_C | Y_C) \, , \;\; p(\{\mu, \sigma, R\}_T | Y_T)$$

Model

- Split data in two groups

$$Y_C = \begin{bmatrix} -3.42 & 1 & 0 \\ 1.2 & 1 & 1 \\ & \dots & \end{bmatrix}, \; Y_T = \begin{bmatrix} -4.41 & 1 & 0 \\ -0.2 & 0 & 1 \\ & \dots & \end{bmatrix}$$

- Model each group separately and retrieve two joint posterior distributions to compare

$$p(\boldsymbol{\theta}_C | Y_C) \; , \;\; p(\boldsymbol{\theta}_T | Y_T)$$

$$p(\{\boldsymbol{\mu}, \boldsymbol{\sigma}, R\}_C | Y_C) \; , \;\; p(\{\boldsymbol{\mu}, \boldsymbol{\sigma}, R\}_T | Y_T)$$

Model

- Split data in two groups

$$Y_C = \begin{bmatrix} -3.42 & 1 & 0 \\ 1.2 & 1 & 1 \\ & \dots & \end{bmatrix}, \quad Y_T = \begin{bmatrix} -4.41 & 1 & 0 \\ -0.2 & 0 & 1 \\ & \dots & \end{bmatrix}$$

- Model each group separately and retrieve two joint posterior distributions to compare

$$p(\boldsymbol{\theta}_C | Y_C) \, , \;\; p(\boldsymbol{\theta}_T | Y_T)$$

$$p(\{\boldsymbol{\mu}, \boldsymbol{\sigma}, R\}_C | Y_C) \, , \;\; p(\{\boldsymbol{\mu}, \boldsymbol{\sigma}, R\}_T | Y_T)$$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{\ell(\theta|\mathcal{D})p(\theta)}{p(\mathcal{D})}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference gives us the joint distribution $p(\theta|D)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

Bayesian Inference gives us the joint distribution $p(\theta|D)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- data $\mathcal{D} = Y$
- parameters $\theta = (\mu, \sigma, R)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- data $\mathcal{D} = Y$
- parameters $\theta = (\mu, \sigma, R)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- likelihood $p(\mathcal{D}|\theta) = p(Y|Z, \mu, R, \sigma)$
- prior $p(\theta)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- likelihood $p(\mathcal{D}|\theta) = p(Y|Z, \mu, R, \sigma)$

$$\begin{aligned} p(Y|Z,\mu,\sigma,R) &= f(Y|Z)\cdot\pi(Z|\mu,\Sigma) \\ &= [\prod_{j\in J_b}\prod_{i=1}^{N} h_j^{-1}(Z_{ij})^{Y_{ij}}(1-h_j^{-1}(Z_{ij}))^{(1-Y_{ij})}]\cdot p(Z|\mu,\Sigma) \\ &= [\prod_{j\in J_b}\prod_{i=1}^{N} h_j^{-1}(Z_{ij})^{Y_{ij}}(1-h_j^{-1}(Z_{ij}))^{(1-Y_{ij})}]\cdot N(Z;\mu,\Sigma) \end{aligned}$$

where $\Sigma = \mathrm{diag}(\sigma)\ R\ \mathrm{diag}(\sigma)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- likelihood $p(\mathcal{D}|\theta) = p(Y|Z, \mu, R, \sigma)$
- prior $p(\theta)$

Bayesian Inference

Bayes' rule

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- likelihood $p(\mathcal{D}|\theta) = p(Y|Z, \mu, R, \sigma)$
- prior $p(\theta) = p(\mu)p(\sigma)p(R)$

Bayesian Inference

Bayes' rule

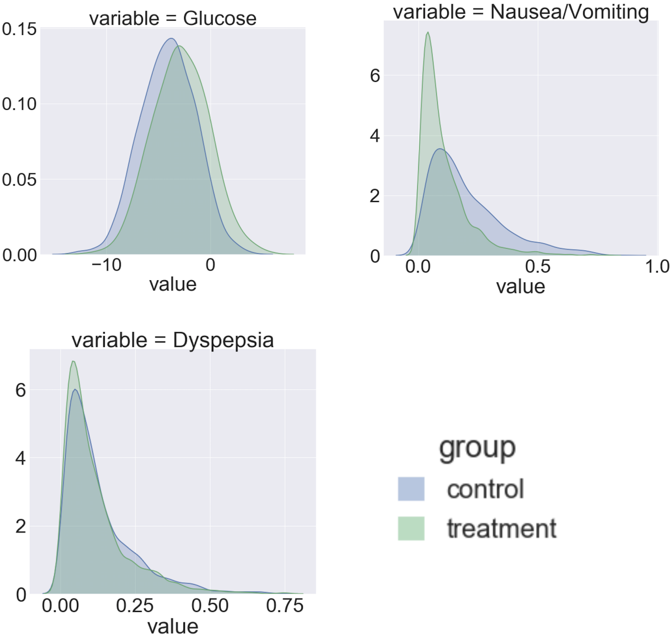
$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

ties together the ingredients of statistical inference and allows information to flow from the data to the parameters

- likelihood $p(\mathcal{D}|\theta) = p(Y|Z, \mu, R, \sigma)$
- prior $p(\theta) = p(\mu)p(\sigma)p(R)$

$$\begin{aligned} \mu &\sim N(0, 10) \\ R &\sim \text{LKJ}(2) \\ \sigma &\sim \text{Half-Cauchy}(0, 2) \end{aligned}$$

Prediction



variable = Glucose
0.15
0.10
0.05
0.00
−10
0
value
variable = Nausea/Vomiting
6
4
2
0
0.0
0.5
1.0
value
variable = Dyspepsia
6
4
2
0
0.00
0.25
0.50
0.75
value
group
control
treatment

What is Stan?

- C++ compiled language for Bayesian Statistical Inference
- Automates the "$p(\theta|\mathcal{D})$" part
- You give the ingredients $\{p(\mathcal{D}|\theta), p(\theta)\}$, it gives you the posterior $p(\theta|\mathcal{D})$

How to use Stan in practice?

- Python interface: PyStan
- Stan resources online
  - **Stan Manual** mc-stan.org/users/documentation/
  - **Case Studies** mc-stan.org/users/documentation/case-studies
  - **Forum** discourse.mc-stan.org
- See SciPy Proceedings paper: *"A Bayesian's journey to a better research workflow"*
- Similar tool: PyMC3

What is Stan?

- C++ compiled language for Bayesian Statistical Inference
- Automates the "$p(\theta|\mathcal{D})$" part
- You give the ingredients $\{p(\mathcal{D}|\theta), p(\theta)\}$, it gives you the posterior $p(\theta|\mathcal{D})$

How to use Stan in practice?

- Python interface: PyStan
- Stan resources online
  - **Stan Manual** mc-stan.org/users/documentation/
  - **Case Studies** mc-stan.org/users/documentation/case-studies
  - **Forum** discourse.mc-stan.org
- See SciPy Proceedings paper:
  "A Bayesian's journey to a better research workflow"
- Similar tool: PyMC3

Our Journey towards Reproducibility

- Definition: recovering results using the *same* materials
- Motivations and aspirations
- Key practices
  - project organization
  - documentation
  - automation
- Interactions with the community
  - porting software
  - sharing data

Thank you!

Scipy Proceedings Paper:
"A Bayesian's journey to a better research workflow"

Konstantinos Vamvourellis
Department of Statistics
k.vamvourellis@lse.ac.uk
github.com/bayesways
personal.lse.ac.uk/vamourel