Bayesian Data Analysis for Software Engineering

Richard Torkar Carlo A. Furia Robert Feldt
Chalmers and University of Gothenburg, Sweden
USI Lugano, Switzerland

ICSE 2021 Technical Briefings

Part 2: How does Bayesian Data Analysis work?

Bayesian Data Analysis for Software Engineering

Richard Torkar   Carlo A. Furia   Robert Feldt

ICSE 2021 Technical Briefings

$\text{prior} \times \text{likelihood} \propto \text{posterior}$

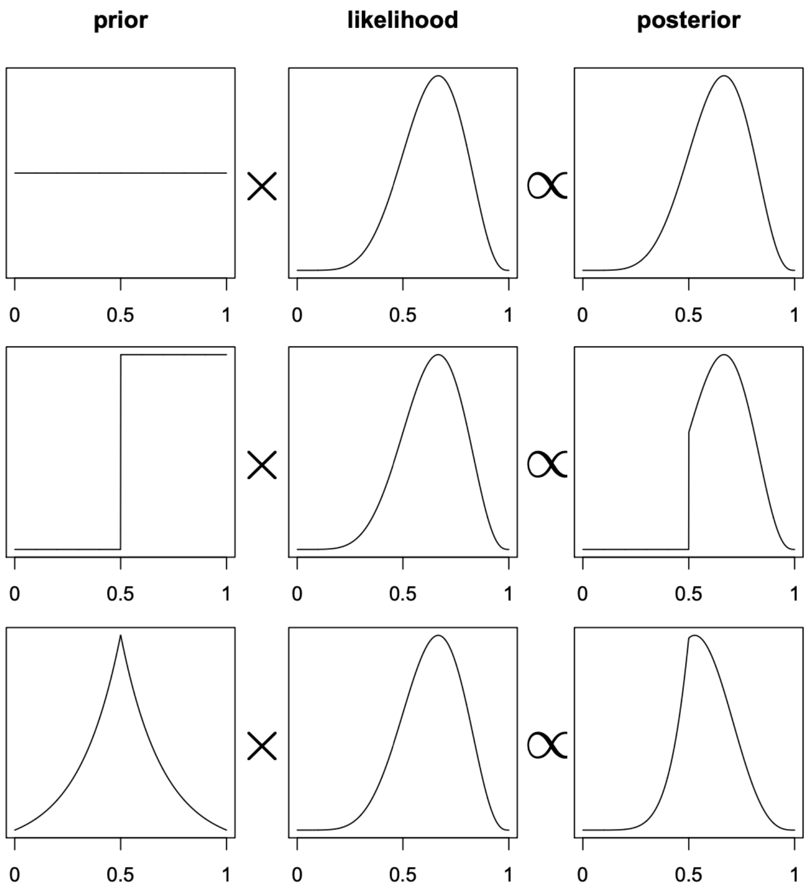


Figure from McElreath, 2020

- Prior knowledge
- The likelihood is our assumptions about the data generation process (maximum entropy)
- Multiplying the two leads to a posterior distribution
- The posterior can be joint (>1 parameter)

Model specifications

$$y_i \sim \text{Normal}(\mu_i,\ \sigma)$$
$$\mu_i = \alpha + \beta_l \bullet x_i$$
$$\alpha \sim \text{Normal}(50, 10)$$
$$\beta_l \sim \text{LogNormal}(-5, 2)$$
$$\sigma \sim \text{Exponential}(1)$$

```
@model function linear_regression(x, y)
    σ ~ Exponential(1)
    α ~ Normal(50, 10)
    βₗ ~ LogNormal(-5, 2)

    N = length(y)
    for n ∈ 1:N
        μ[n] ~ α + βₗ * x
    end

    y ~ Normal(μ, σ)
end
```

turing (Julia)

```
ulam(
  alist(
    y ~ dnorm(mu, sigma),
    mu ← alpha + b_loc * x,
    alpha ~ dnorm(50, 10),
    b_loc ~ dlnorm(-5, 2),
    sigma ~ dexp(1)
    )
)
```

rethinking (R)

```
brm(
    y ~ 1 + LOC,
    data = d,
    family = normal(),
    prior = c(
        set_prior(normal(50, 10), class = Intercept),
        set_prior(lognormal(-5, 2), class = b,
        set_prior(exponential(1), class = sd))
    )
)
```

brms (R)

Step 1

Likelihoods

Likelihoods and their assumptions

(distribution that can happen in most ways is also the one with biggest information entropy)

- How to pick one?
  - Epistemological
    - Maximum entropy
  - Ontological
    - Occurrence of patterns
- Strict falsification not possible
  - Hypotheses are not models
  - Measurement matters

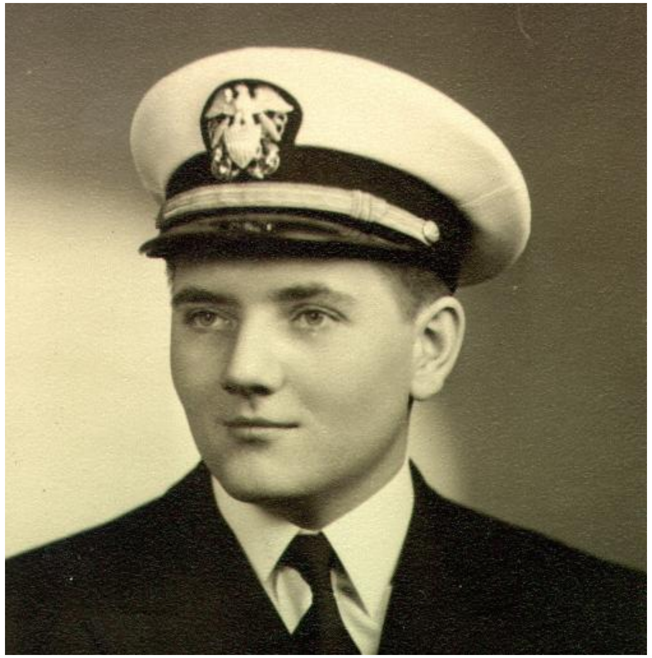
Motoo Kimura (1924–1994)

Step 2

Priors

Priors

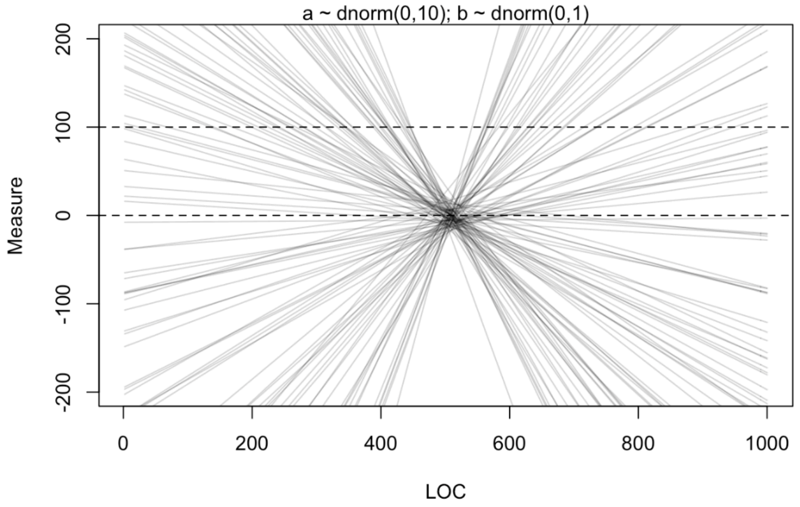
- Infinity is not an option
  - Leads to overfitting
  - Frequentist techniques overfit *a maximum*, generally speaking
- Prior knowledge always exists
- Delimits the multidimensional space a sampler must explore
- Logical principles exist for assigning prior probability distributions

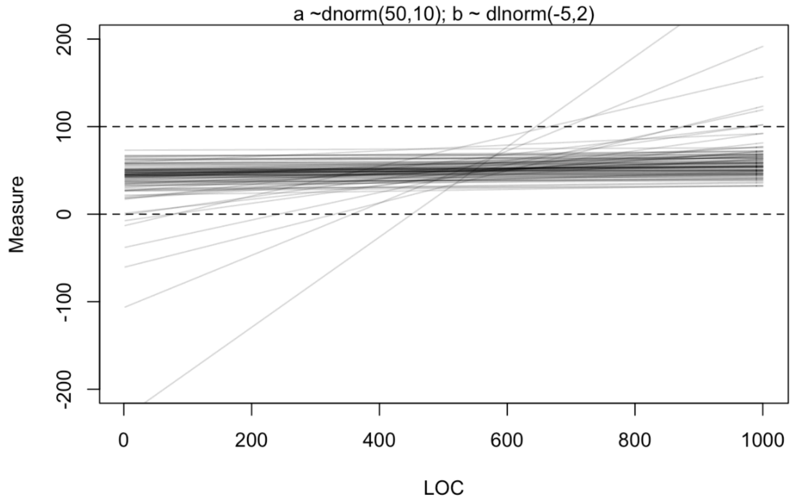


Edwin Thompson Jaynes (1922–1998)

Prior predictive checks

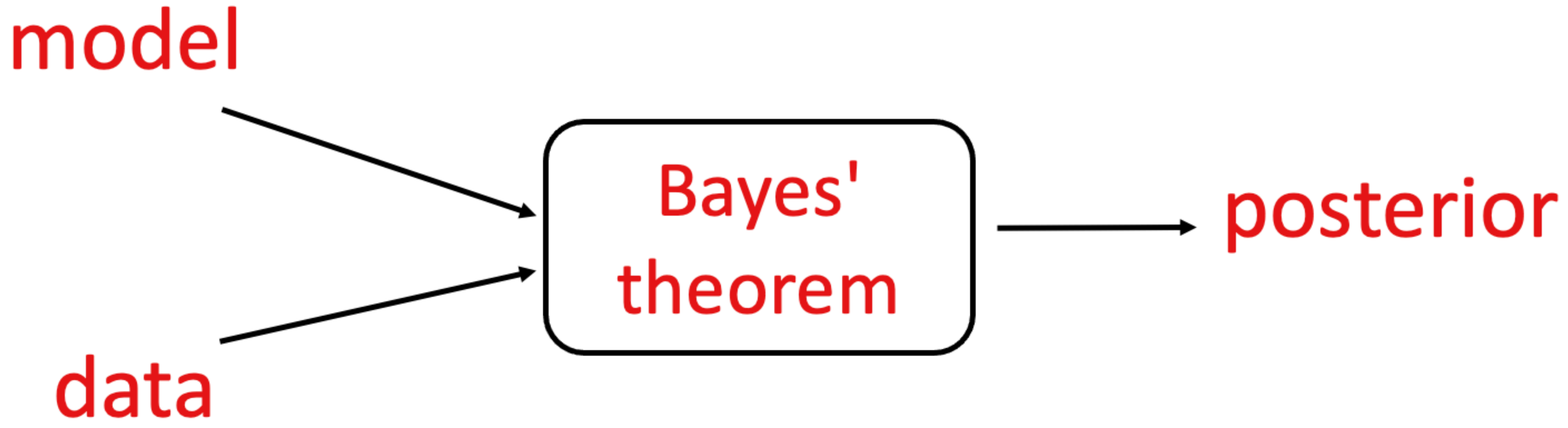
- Generate data from priors only
- Many graphical checks exists
- Purpose is to assess suitability of priors
- Often called sensitivity analysis, i.e., how sensitive is the model to different priors?





Step 3

Calculating the posterior


model
data
Bayes'
theorem
posterior

How do we get the posterior?

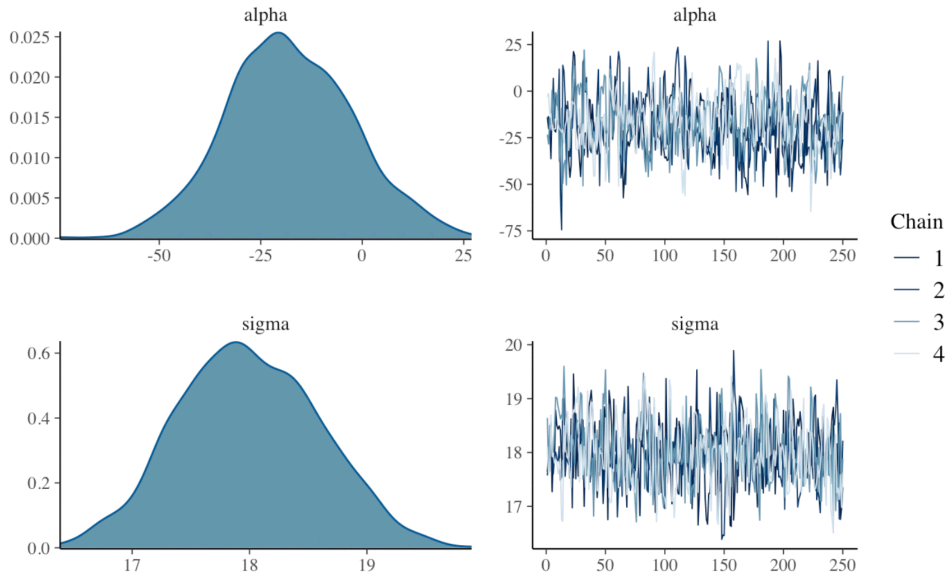
- Grid approximations
- Quadratic approximation
- Sample-Importance-Resample
- Approximate Bayesian Computation
- …
- Markov Chain Monte Carlo
  - Metropolis-Hasting
  - Gibbs
  - **Hamiltonian Monte Carlo**

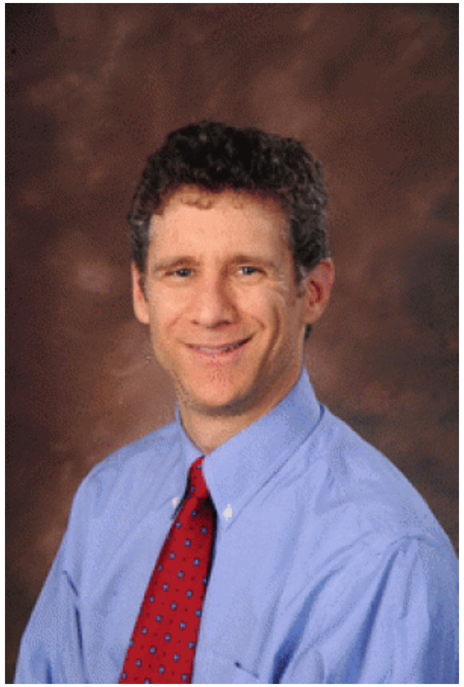
Arianna Rosenbluth (1927–2020)

Stanislaw Ułam (1909–1984)

Diagnostics

- Many diagnostics exists
  - $\hat{R}$
  - **Effective sample sizes**
  - **Traceplots**
  - Divergences
  - E-BFMI
  - Treedepth
- Sampler provides am warnings





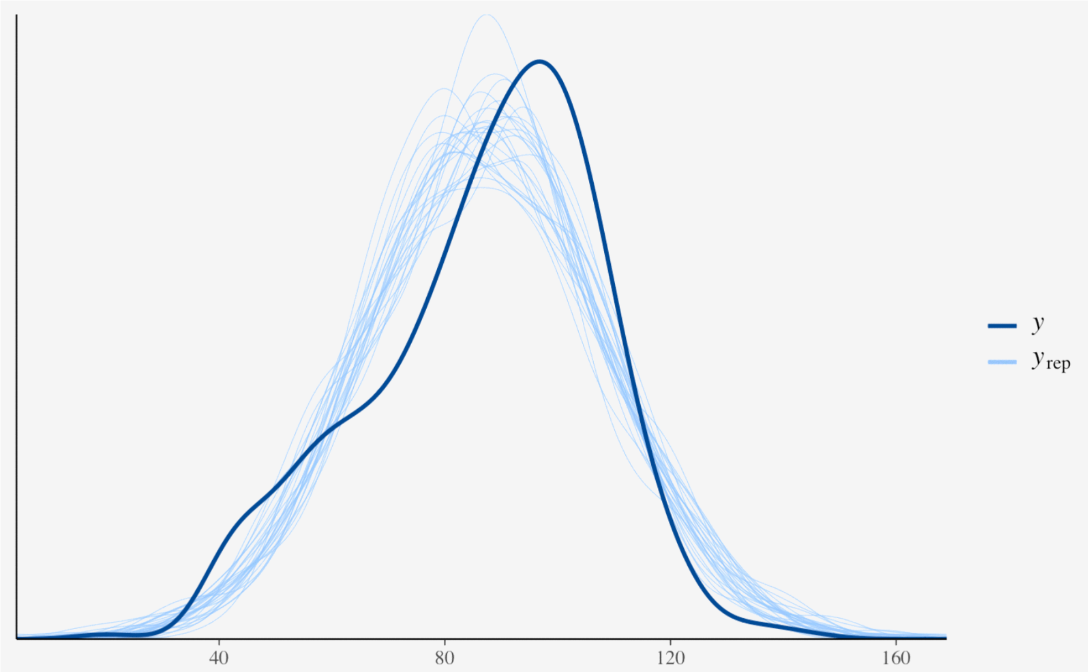
Andrew Gelman, Columbia University, and many in the Stan team

Step 4

Sanity check of the posterior

Posterior predictive checks

- After fitting the model, we
  - compare our empirical data with what the model predicts,
  - check how well we capture variability, zero-inflation, etc., by comparing the predictive error
  - Use plots or even Bayes-R^2 etc.

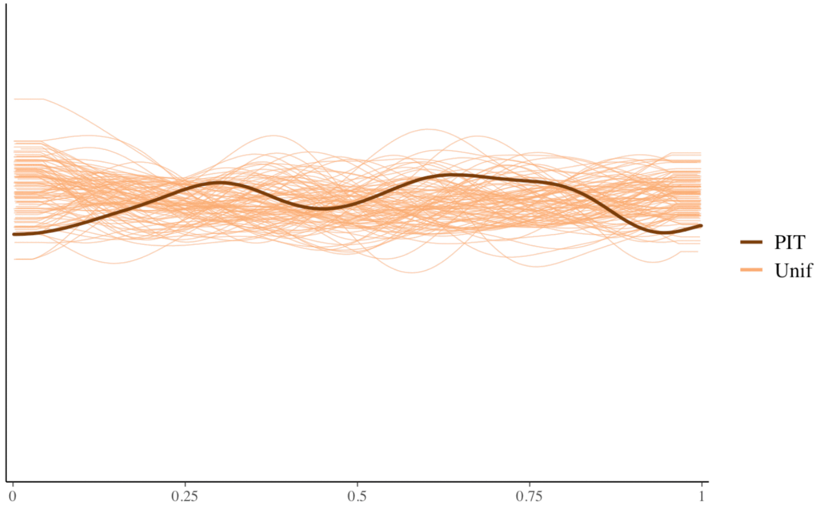


(Step 5)

Model comparisons

Model comparisons

- Traditionally used WAIC/ BIC/DIC etc.
- An information theoretical relative comparison of >1 models
- LOO is state of art
  - Handles all type of likelihoods
  - Compared to WAIC,
    - LOO provides ample diagnostics giving you confidence in results



Aki Vehtari,
Aalto University

Danielle Navarro,
U. New South Wales

Step 6

Compute stuff

Using the posterior

- Compute stuff
  - Intervals
  - Point estimates
- Plot posterior probability distributions
- Simulate

Do you want to know more?

- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan, 2nd Edition*, CRC Press
- Jaynes, E. (2003). *Probability Theory: The Logic of Science* (G. Bretthorst, Ed.). Cambridge: Cambridge University Press
- Navarro, D. J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain and Behavior,* 2:28–34
- Frank, S. A. (2009), The common patterns of nature. *Journal of Evolutionary Biology*, 22:1563–1585
- https://mc-stan.org (`Stan` is state of practice concerning HMC sampling)
- https://mc-stan.org/users/interfaces/brms (`lme4` syntax for models)
- https://mc-stan.org/bayesplot/ (plots galore for Bayesian MLMs)
- https://mc-stan.org/users/interfaces/loo (model comparison)
- https://turing.ml/ (model design in Julia)

License of these slides

Links

- https://github.com/torkar/icse_tutorial (try it yourself)
- http://tiny.cc/bayes-icse21 for more polished videos of what we've talked about!