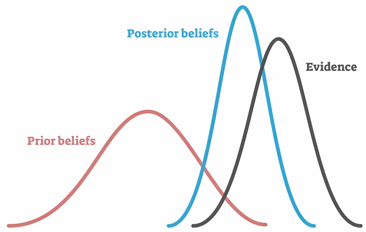


Z. Wei

Objective Bayesian v3_20230905

Motivation

(1) Let the data speak for themselves. (2) Cromwell's rule. 🌙🧀🧑‍🚀

"… the search of priors with a minimal impact on the corresponding posterior analysis, …" (p. 627)

Consonni, G., Fouskakis, D., Liseo, B., & Ntzoufras, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, *13*, 627–679.

Types of Priors, no unanimity, often a misnomer (p. 628)

- Noninformative (Objective)

  e.g., Jeffreys, $\pi(\boldsymbol{\theta}) \propto \sqrt{|\mathcal{I}(\boldsymbol{\theta})|}$;

  cannot be directly used on the discrete parameter space;
  (How to interpret?) invariant to transformations of the parameters;
  observations can be independent or dependent.

  $B(n, p)$. $\pi(p) = \text{Beta}(0.5,\ 0.5)$, U-shape, conjugate

  $\mathcal{N}(\mu, \sigma^2)$. $\pi(\mu) \propto 1$

  Stan: implicit

  JAGS: `mu ~ dunif(-100, 100)`

  $Y = (X - a)/(b - a) \sim \text{Beta}(1,\ 1)$, maybe slightly faster

  `mu ~ dnorm(0,` precision $\tau =$ `0.0001)`

  $$\pi(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi(\tau) \propto \frac{1}{\tau}, \quad \pi(\sigma) \propto \frac{1}{\sigma}, \quad \pi(\ln \sigma) \propto 1$$

  Stan: `target += -log(sigma);`

  JAGS: `sigma ~ dgamma(0.0001,` rate = `0.0001)`

  `log.sigma ~ dunif(-10, 10)`

  e.g., (Bernardo's) reference, maximizing the expected Kullback–Leibler divergence;

  observations are independent and identically distributed.
  Is there always a reference prior? Is it unique?

- Weakly informative (Vague), diffuse
- Informative (Subjective)

Examples:

Flat: You know nothing, often improper (may not integrate or sum to 1)

- Super-vague but proper (not usually recommended): $\mathcal{N}(0,\ 1000^2)$
- Weakly informative: $\mathcal{N}(0,\ 10^2)$
- Generic weakly informative: $\mathcal{N}(0,\ 1)$
- Specific informative (context-based): $\mathcal{N}(0.4,\ 0.5^2)$ or a scaling $0.4 + 0.5 \cdot \mathcal{N}(0,\ 1)$

An improper prior may lead to an improper posterior.

e.g., Haldane, $\pi(p) = p^{-1}(1-p)^{-1}$ for $B(n, p)$.

Default, e.g., g-priors (and hyperpriors) in the *anovaBF* function of the "*BayesFactor*" **R** package.

$$g \sim \text{Scale-inv-}\chi^2(\nu,\ h^2), \quad g \sim \text{Inv-Gamma}\left(\nu/2,\ \text{scale} = \nu h^2/2\right), \quad g^{-1} \sim \text{Gamma}\left(\nu/2,\ \text{rate} = \nu h^2/2\right)$$

Prefer to choose priors for biologically meaningful parameters, e.g., p instead of logit(p) (Meredith, 2021).

Virtues for data analysis (p. 629) + **Sensitivity analysis**, Bayesian model checking and model evaluation

Instead of objectivity, think about:	Instead of subjectivity, think about:
Transparency	Multiple perspectives
Consensus	Context dependence
Impartiality	
Correspondence to observable reality	

<u>Example:</u> mean of population distribution of $\ln(\mathrm{BVA}_j^{\mathrm{latent}}/50)$, centered at 0 because the mean of the boundary value analysis (BVA) values in the population should indeed be near 50. We set the prior standard deviation to 0.2 which is close to $\ln(60/50) \approx 0.18$ to indicate that we are pretty sure the mean is between 40 and 60 (Gelman & Hennig, 2017, p. 979).

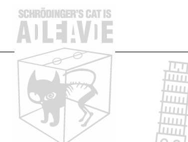
Criteria for Objective Bayesian Model Selection (p. 640-641)

- Basic

Consistency
- Model selection consistency
- Information consistency
- Intrinsic consistency

- Predictive matching

Invariance
- Measurement invariance
- Group invariance



DISCUSSIONS

1. What are the cons of being data-driven?
   Keywords: quality of data, correlation ≠ causation, overfitting, privacy concerns, thought experiment.

2. Do priors play different roles in i) estimation or **prediction** and ii) model selection or model comparison?
   Keywords: conformal prediction, multicollinearity, suppression, confounding, prior-hacking.

3. The prior should ideally be chosen before data collection.
   ```
   "lwr"=mean(data)-6*sd(data), "upr"=mean(data)+6*sd(data)
   ```
   https://discourse.mc-stan.org/u/dexterw/activity/topics

4. Improper priors for estimation may not guarantee proper posterior.
   Keywords: multivariate, heteroscedasticity, Jeffreys priors for σ_i , MCMC divergence warning.
   $$\pi(\mu,\, \sigma_1^2, \cdots, \sigma_a^2) \propto \prod_{i=1}^{a} \sigma_i^{-2}$$

5. Effect-size parametrization and measurement invariance, $Y_{ij} = \mu + \sigma_\epsilon(t_i + b_j) + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, \sigma_\epsilon^2)$.

   https://rpubs.com/sherloconan/1024130

$$\mathscr{M}^*_{\text{full}}:\ \ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon^*_{ijk}$$
$$\pi(\mu) \propto 1$$
$$(\alpha_1^\star, \cdots, \alpha_{a-1}^\star) = (\alpha_1, \cdots, \alpha_a) \cdot \mathbf{Q}, \quad \mathbf{I}_a - a^{-1}\mathbf{J}_a = \mathbf{Q} \cdot \mathbf{Q}^\top, \quad \cdots$$
$$\alpha_i^\star \mid g_A \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, g_A), \quad \beta_j^\star \mid g_B \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, g_B), \quad (\alpha\beta)_{ij}^\star \mid g_{AB} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\, g_{AB})$$
$$g_A \sim \text{Scale-inv-}\chi^2(1,\, h_A^2), \quad g_B \sim \text{Scale-inv-}\chi^2(1,\, h_B^2), \quad g_{AB} \sim \text{Scale-inv-}\chi^2(1,\, h_{AB}^2)$$
$$\epsilon_k = (\epsilon^*_{11k}, \epsilon^*_{21k}, \epsilon^*_{12k}, \epsilon^*_{22k})^\top \sim \mathcal{N}_4(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} \sim \mathscr{W}^{-1}(\mathbf{I}_{4\times 4},\, 5)$$