PREVAIL II ANALYSIS CONTINUED

Bayesian workflow, sampling to summarize, and a first look at Stan

August 25, 2019

Jon Fintzi

Biostatistics Research Branch
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Taylor Swift Released a New Album

But I still like my oldies:

- Bayesian inference *always* starts with a model for the **joint distribution** of θ and y:.

$$\pi(\theta, y) = f(y|\theta)\pi(\theta) = \pi(\theta|y)m(y).$$

- **Bayes rule** yields the **posterior distribution**

$$\pi(\theta|y) = \frac{f(y,\theta)}{m(y)} = \frac{f(y|\theta)\pi(\theta)}{m(y)} \propto Likelihood \times Prior.$$

.

- All of the information used in the *update* to our prior is encoded in the **likelihood**,

$$L(\mathbf{y}|\theta) = \prod_{i=1}^{N} f(y_i|y_{1,\ldots,i-1}\theta).$$

- Analysis of PREVAIL II data with Beta-Binomial model:
  - Conjugate prior $\implies$ analytical posterior.
  - Beta(1,1), i.e., Uniform(0,1), prior for probability of 28-day mortality.
  - Distributions of functionals of the posterior, e.g., assess evidence for ZMapp + oSOC more effective than oSOC alone: $\Pr(p_T < p_C|y_T, y_C)$.

Lecture 10 of Statistical Rethinking

Key takeaways:

- Language for modeling:

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma^2), \\
\mu_i &= \beta_0 + \beta_1 x_{i,1} + \beta_p x_{i,p}, \\
\beta &\sim \mathrm{N}(0, 10^2), \\
\sigma &\sim \mathrm{Exponential}(1), \\
x_i &\sim \mathrm{N}(0, 1).
\end{aligned}
$$

- Prior is full of lines.
- Posterior is full of lines.
- Prior and posterior also induce distributions over functionals, e.g., $\pi(R^2|\mathbf{y})$.
- Parameterization is important for interpretation, prior specification, and computation.

This Week

Managing the elephant

- Analysis of PREVAIL II with non-conjugate priors:
  - Defining the model.
  - Prior predictive simulations.
  - Posterior predictive distributions.
- MCMC via `Stan`.

Motivating Example — PREVAIL II Trial

Context:

- 2014–2016 Ebola virus disease (EVD) outbreak in Guinea, Liberia, and Sierra Leone.
- Over 28,000 suspected or confirmed cases and 11,000 fatalities.
- Urgent need to identify effective theraputics to reduce mortality.

Partnership for Research on Ebola Virus in Liberia (PREVAIL) II trial:

- Adaptive trial to determine the effectiveness of ZMapp, and possibly other agents, in reducing Ebola mortality.
- Primary endpoint: 28 day mortality on optimized standard of care (oSOC) vs. ZMapp + oSOC.
- 72 patients enrolled at sites in Liberia, Sierra Leone, Guinea, and the US.
  - Overall mortality: 21/71 died (30%),
  - SOC alone: 13/35 (37%),
  - ZMapp + SOC: 8/36 (22%).
- Modeled as Beta-Binomial with $p_T \sim \text{Beta}(1,1),\ p_C \sim \text{Beta}(1,1)$ (Proschan, 2016).
  - Beta dist. hyper-parameters interpretable as pseudo-observations.
  - $p_T|y_T \sim \text{Beta}(9,29)$ and $p_C|y_C \sim \text{Beta}(14,23)$.

Analysis with Informed Priors

Suppose we had information about the overall odds of death and differences in mortality.

$$\theta_{GM} = \exp\left[\frac{1}{2}\left(\log\left(\frac{p_T}{1-p_T}\right) + \log\left(\frac{p_C}{1-p_C}\right)\right)\right] \sim \text{LogNormal}(\mu_{GM}, \sigma^2_{GM})$$

$$\theta_{OR} = \exp\left[\log\left(\frac{p_T}{1-p_T}\right) - \log\left(\frac{p_C}{1-p_C}\right)\right] \sim \text{LogNormal}(0, \tau^2).$$

- What's with all of the logs and exps?
  - Parameters defined multiplicatively are additive on the log scale.
  - Constraints on parameter space, e.g., $\theta_{OR} \in (0, \infty)$.
- Lose marginal interpretations of Beta priors. How to set hyperparameters for joint priors?
- No more conjugacy, how to compute the posterior?

Rethinking the Model

Bayesian inference **always** starts with a model for the joint distribution of parameters, $\boldsymbol{\theta}$, and data, $\mathbf{Y}$.

- Here, $\boldsymbol{\theta} = (\theta_{GM}, \theta_{OR})$.
- Relevant posterior: $\pi(\theta_{GM}, \theta_{OR}|\mathbf{Y}) \propto L(\mathbf{Y}|\theta_{GM}, \theta_{OR})\pi(\theta_{GM})\pi(\theta_{OR})$.
  - $Y_T \sim \text{Binomial}(N_T, p_T),\ Y_C \sim \text{Binomial}(N_C, p_C)$.
  - $(p_T, p_C) = \left(\text{expit}\left(\log(\theta_{GM}) + 0.5\log(\theta_{OR})\right), \text{expit}\left(\log(\theta_{GM}) - 0.5\log(\theta_{OR})\right)\right)$.
  - **Very super-duper important:** p_T and p_C are functionals of the model parameters. They may be of interest. They are essential to the likelihood. They even have a distribution. But our model is defined in terms of θ_{GM} and θ_{OR}.
- What do our beliefs about θ_{GM} and θ_{OR} imply about $p_T,\ p_C$, and $\mathbf{Y}$?

Rethinking the Model

What do our priors[1] for θ_{GM} and θ_{OR} imply about p_T, p_C, and $\mathbf{Y}$?

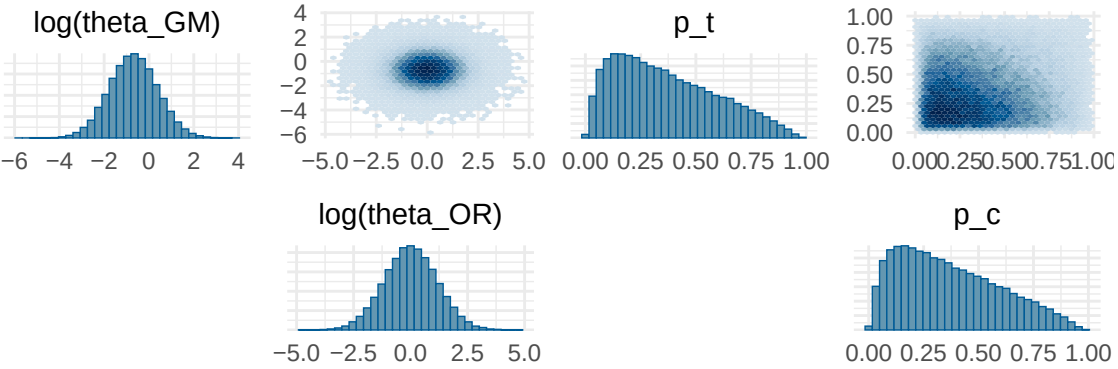
- $\Pr(\theta_{GM} < 0.5) = 0.5$ and $\Pr(\theta_{GM} > 2) = 0.1$
  $\implies \theta_{GM} \sim LogNormal(log(0.5), 1.08^2)$
- $\Pr(\theta_{OR} < 0.25) = 0.1$ and $\Pr(\theta_{OR} > 4) = 0.1 \implies \theta_{OR} \sim LogNormal(0, 1.08^2)$.
- **Interpretation:** the majority of people, roughly 2/3 half the time, will die within 28 days on oSOC. There is a non-negligible chance that ZMapp is much more, or much less, effective than oSOC alone, though we are agnostic to the direction of the effect. It is more likely, however, that the odds of 28 day mortality do not differ hugely.
- **Problem:** humans are really bad at interpretting probabilities (and hence calibrating priors using probabilities). We really only have three heuristics: not gonna happen, I dunno ¯_(ツ)_/¯, and definitely gonna happen I'd bet my firstborn on it.
- Simulate to summarize the prior, *especially when fitting complex models*.

[1] I'm just picking the probabilities that determine the following priors arbitrarily here for the sake of demonstration. Don't read too much into them.

Rethinking the Model

What do our priors for θ_{GM} and θ_{OR} imply about p_T, p_C, and $\mathbf{Y}$?

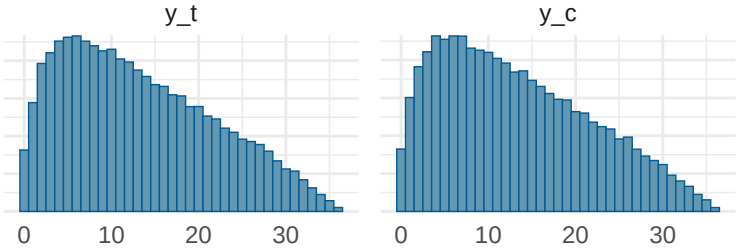
- Prior predictive distribution: (1) draw $\widetilde{\theta}_{GM} \sim \pi(\theta_{GM})$, $\widetilde{\theta}_{OR} \sim \pi(\theta_{OR}) \implies (\widetilde{p}_T, \widetilde{p}_C)$, then (2) simulate outcomes $\widetilde{Y}_T \sim \text{Binomial}(N_T, \widetilde{p}_T)$, and $\widetilde{Y}_C \sim \text{Binomial}(N_C, \widetilde{p}_C)$.



Rethinking the Model

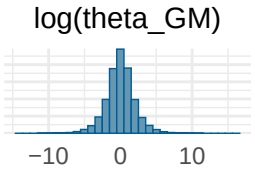
What do our priors for θ_{GM} and θ_{OR} imply about p_T, p_C, and $\mathbf{Y}$?

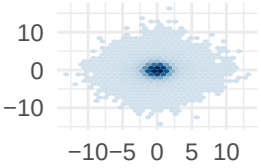
- Prior predictive distribution: (1) draw $\widetilde{\theta}_{GM} \sim \pi(\theta_{GM})$, $\widetilde{\theta}_{OR} \sim \pi(\theta_{OR}) \implies (\widetilde{p}_T, \widetilde{p}_C)$, then (2) simulate outcomes $\widetilde{Y}_T \sim \text{Binomial}(N_T, \widetilde{p}_T)$, and $\widetilde{Y}_C \sim \text{Binomial}(N_C, \widetilde{p}_C)$.
- Summary statistics of interest (90% prior predictive interval): total deaths (8, 48) , deaths per arm (2, 29), absolute difference in deaths per arm (1, 24).

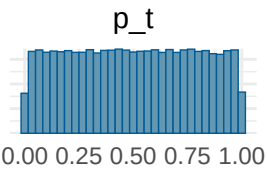


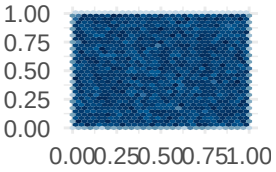
Rethinking the Model

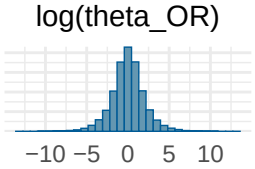
Suppose we had stuck with $p_T \sim \text{Beta}(1,1)$, $p_C \sim \text{Beta}(1,1)$. What are the induced priors on θ_{GM}, θ_{OR}, and $\mathbf{Y}$?

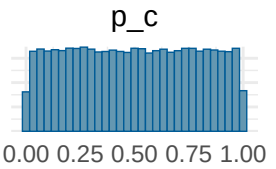








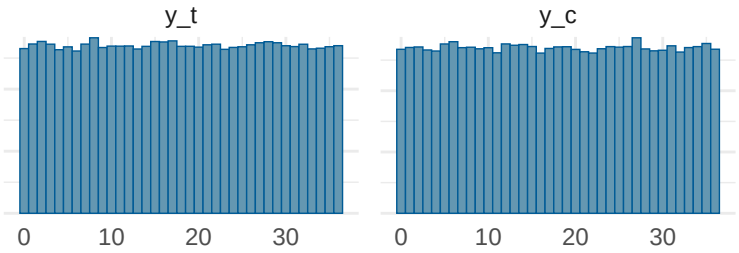




Rethinking the Model

Suppose we had stuck with $p_T \sim \text{Beta}(1,1)$, $p_C \sim \text{Beta}(1,1)$. What are the induced priors on θ_{GM}, θ_{OR}, and $\mathbf{Y}$?

- Summary statistics of interest (90% prior predictive interval): total deaths (11, 61) , deaths per arm (1, 35), absolute difference in deaths per arm (1, 29).



MCMC for Non-conjugate Priors

Posterior distribution no longer analytically available. Let's do that MCMC!
To specify the model in `Stan`, we need to define (at a minimum)

- *Data*: $\mathbf{N}$, $\mathbf{Y}$,
- *Parameters*: θ_{GM}, θ_{OR},
- *Model*: Binomial likelihood, log-normal priors.

Optionally, we can also specify

- *User-defined functions*: e.g., for manipulating data, transforming parameters,
- *Transformed data*: e.g., centered covariates,
- *Transformed parameters*: p_T, p_C,
- *Generated quantities*: functionals of parameters, posterior predictive samples.

Each of these is defined in a block of `Stan` code, which is compiled into `C++`.

`Stan` has incredible documentation and an active (and very supportive) user community that you can lean on if you ever need help with a model.

Let's do that MCMC!

Data:

- The data, but also any quantities that are needed to instatiate objects, e.g., dimensions of matrices.
- Objects here are fixed, no random variables declared here.

```
data {
  int<lower=0> N[2]; // sample sizes per arm
  int<lower=0> y[2]; // numbers of deaths per arm
}
```

Let's do that MCMC!

Parameters:

- Correspond to the variables being sampled in the MCMC.
- Constraints for safety and to tell `Stan` how to transform to avoid boundaries of parameter space.
- `Stan` defines an unnormalized log-probability over unconstrained parameters and adds Jacobians automatically.

```
parameters {
  real<lower=0> theta_GM; // geometric mean odds of death
  real<lower=0> theta_OR; // odds ratio
}
transformed parameters {
  real<lower=0,upper=1> probs[2];
  probs[1] = inv_logit(log(theta_GM) + 0.5 * log(theta_OR)); // p_T
  probs[2] = inv_logit(log(theta_GM) - 0.5 * log(theta_OR)); // p_C
}
```

Let's do that MCMC!

Model:

- Sampling statements define contributions to the log-posterior.
- Includes both priors and likelihood.
- If declare priors for transformed parameters, need to manually add a Jacobian adjustment.

```
model {
  y ~ binomial(N, probs);               // likelihood
  theta_GM ~ lognormal(log(0.5), 1.08); // prior for theta_GM
  theta_OR ~ lognormal(0, 1.08);        // prior for theta_OR
}
```

Let's do that MCMC!

Generated quantities:

- Posterior predictives and derived quantities computed online and returned in MCMC output.
- Later, compute log-likelihoods for model comparison.

```
generated quantities {
  int y_pp[2] = binomial_rng(N, probs); // simulate from posterior predictive
  real rr_pp = probs[1] / probs[2];     // relative risk
  real rd_pp = probs[1] - probs[2];     // risk difference
}
```

Let's do that MCMC!

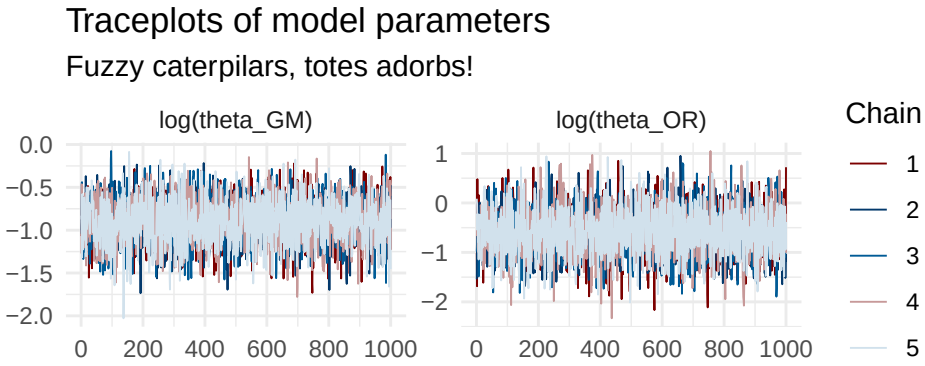
Running the model in R:

- Write `Stan` code into a `.stan` file (RStudio: File > New File > Stan File).
- Probably something similar in Python, I dunno.
- Compile and run as follows:

```
prevailmod = stan_model(file = "prevailmod.stan") # compile model
data = list(N = c(ZMapp = 36, oSOC = 35),         # numbers of participants
            y = c(ZMapp = 8, oSOC = 13))          # deaths per arm
prevailfit = sampling(object = prevailmod,        # compiled model
                      data   = data,              # data
                      chains = 5,                 # number of chains (>1 !!!!)
                      iter   = 2e3)               # number of iterations
```

MCMC Output

`Stan` produces correlated MCMC samples from the posterior via Hamiltonian Monte Carlo.



MCMC Diagnostics

Lots of diagnostics, always check your diagnostics. More here and in the `Stan` reference manual.

```
##      accept_stat__ stepsize__ treedepth__ n_leapfrog__ divergent__
## [1,]     1.0000000  0.9754109           2            3           0
## [2,]     0.7390574  0.9754109           2            3           0
## [3,]     1.0000000  0.9754109           2            3           0
## [4,]     0.9621533  0.9754109           1            3           0
## [5,]     0.8248869  0.9754109           2            3           0
##      energy__
## [1,] 44.60989
## [2,] 45.24771
## [3,] 45.31665
## [4,] 45.22013
## [5,] 43.54042
```

Results

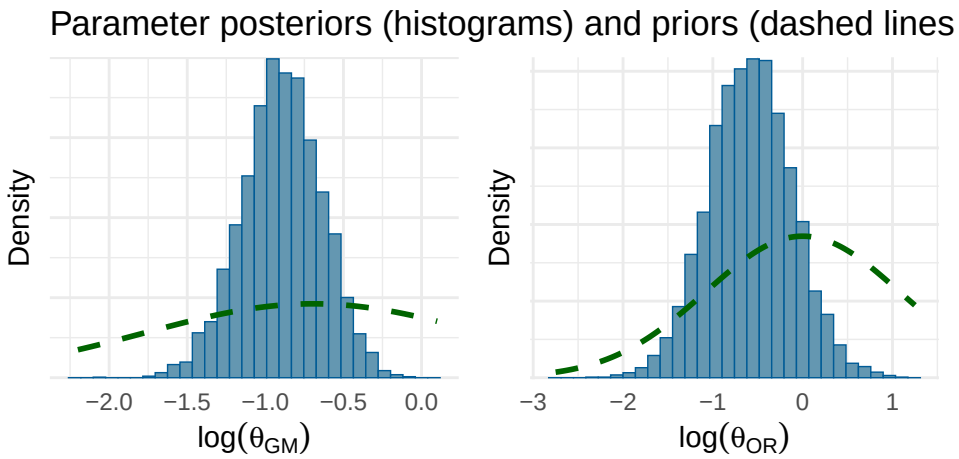
Marginal summaries[2]:

```
##                mean     se_mean         sd       2.5%        25%
## theta_GM  0.4213551 0.001759412 0.10718448  0.2417749  0.3456898
## theta_OR  0.6241425 0.005282141 0.30981109  0.2282034  0.4083772
## probs[1]  0.2399010 0.001075035 0.06349585  0.1278092  0.1946408
## probs[2]  0.3564025 0.001106634 0.07330828  0.2184726  0.3047550
## rr_pp     0.7012247 0.004023744 0.23996935  0.3424526  0.5311668
## rd_pp    -0.1165015 0.001434496 0.09249205 -0.2934280 -0.1799115
##                50%         75%      97.5%    n_eff      Rhat
## theta_GM  0.4092737  0.48563778 0.66043117 3711.323 1.0004128
## theta_OR  0.5584364  0.76411135 1.39572402 3440.123 0.9999599
## probs[1]  0.2355453  0.28230435 0.37421211 3488.555 0.9998798
## probs[2]  0.3548359  0.40579766 0.49962301 4388.314 1.0004480
## rr_pp     0.6647296  0.82959835 1.27363815 3556.729 0.9998410
## rd_pp    -0.1188121 -0.05450594 0.06559438 4157.287 0.9998778
```

[2] In Beta-Binomial model with Uniform(0,1) priors: $p_T = 0.23(0.12, 0.38)$; $p_C = 0.38(0.23, 0.54)$; $RD = -0.14(-0.34, 0.06)$; $RR = 0.62(0.29, 1.24)$; $OR = 0.50(0.18, 1.36)$.

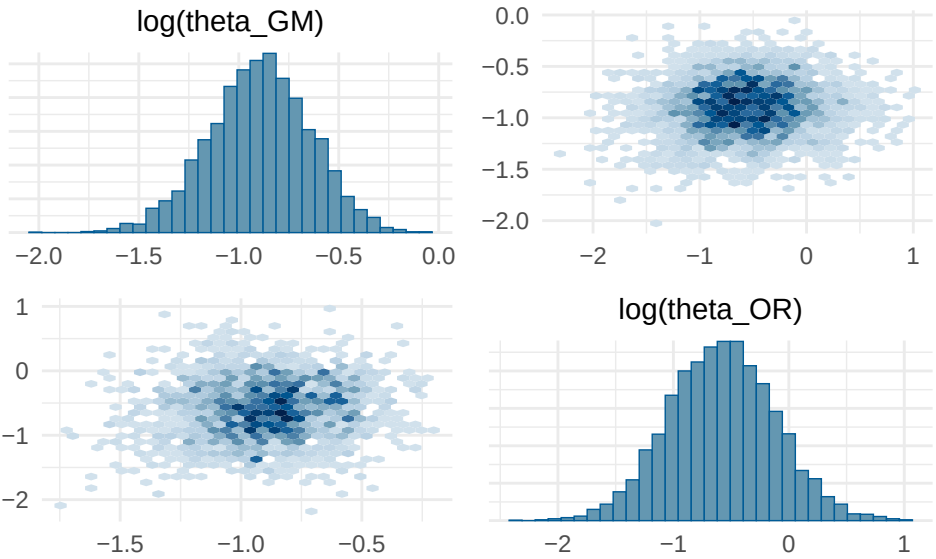
Results

Posterior distributions of model parameters.



Results

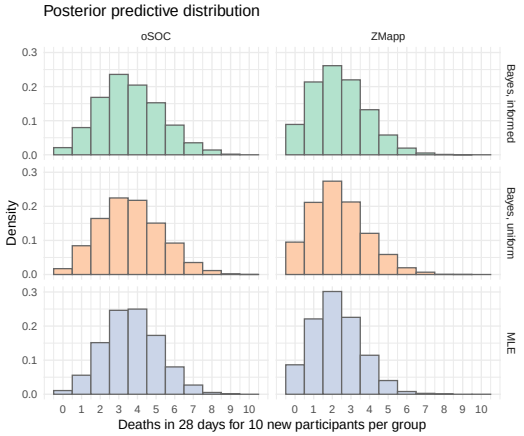
Can do better than marginal summaries, we have access to the *joint* posterior!



Posterior Predictive Distribution

We can sample from the posterior predictive distribution by iteratively drawing $\widetilde{\theta}_{post} \sim \pi(\theta|y)$, then simulating $\widetilde{Y}|\widetilde{\theta}_{post} \sim \text{Binomial}(N, \widetilde{p}_{post})$.

Suppose we want to predict 28 day mortality for 10 new individuals in each group.



Next time

No meeting next week (Labor Day). In two weeks we'll talk about linear regression (for real this time).

- Setting weakly (not weekly) informative priors.
- Assessing failure modes of different priors.
- More Stan.

Watch/rewatch lecture 3 and first half of 4 (SmaRt).

References

M. Betancourt “Calibrating model-based inferences and decisions.” arXiv preprint arXiv:1803.08393 (2018).

M. Betancourt “Toward a principled Bayesian workflow.” `https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html` (2018).

J. Gabry, et al. “Visualization in Bayesian workflow.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.2 (2019): 389-402.