Selecting Prior Parameters

Lin Zhang

Department of Biostatistics
School of Public Health
University of Minnesota

Specifying a beta prior for binomial data

- Question: In a safety study B with N patients, we want to test if the rate of NO AEs at 3 months is higher than 85%.
- Model: Let $\theta = \Pr(\text{Patient does not experience AE})$ and $X = \#$ patients with no AE ("success"). Assuming independent patients,

$$X|\theta \sim Binomial(N, \theta)$$

- Using a conjugate prior $Beta(\alpha, \beta)$, then the posterior is

$$\theta|X \sim Beta(X + \alpha, N - X + \beta)$$

- Question: How do we determine the prior parameters α and β?

Using historical data

- We have access to the 1-month data from a previous safety study A:

	No AE	AE	total
count (%)	110 (94)	7 (6)	117

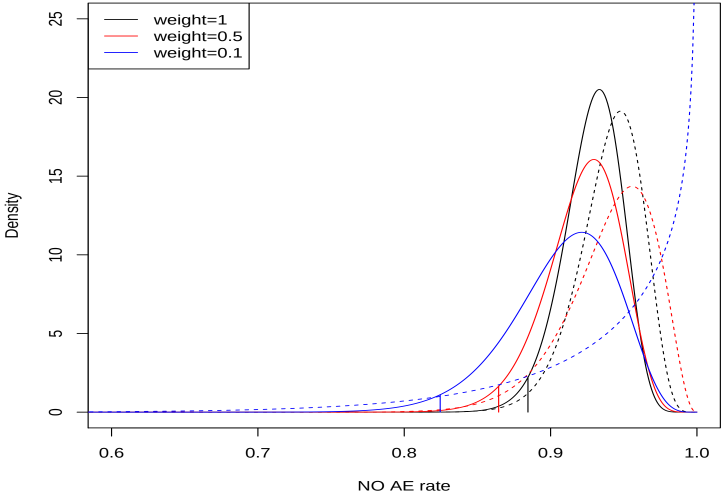
  How to turn this into a prior?

- Consider three priors:
  - $Beta(110, 7)$
  - $Beta(110 * 0.5, 7 * 0.5)$
  - $Beta(110 * 0.1, 7 * 0.1)$

  Assuming you observe $X = 45$: (1)plot the priors and the corresponding posteriors in one figure, (2) calculate the 95% credible intervals, and (3) make conclusions.

- Question: Do you have different conclusions?

Prior and Posterior Plots



Effective sample size for a beta prior

- Using all historical data is clearly too strong! We need to consider prior effective sample size.
- **Effective sample size:** the approximate number of subjects worth of data found in the prior.
- Recall that a beta prior results in the posterior $Beta(X + \alpha, N - X + \beta)$

  $\Rightarrow$ The prior is essentially adding α success and β failures to the data!

  $\Rightarrow$ Effective sample size for the beta prior is $\alpha + \beta$.
- Back to our example, we probably want to downweight the historical data to reduce prior effective sample size so that the prior does not dominate the posterior!

Example: Heart Valves Study

- Goal: Show that the thrombogenicity rate (TR) is less than two times the objective performance criterion (OPC), $\mathbf{0.038}$.

- Data: Assume we observe y events from a total of 200 patient-years of follow-up from the study. In addition, in a previous study on a similar product in St. Jude, 98 events were observed from a total of 5891 patient-years of follow-up.

- Likelihood Model: Let $T = 200$ denote the total number of patient-years of follow up, and θ be the TR per year. We assume the number of thrombogenicity events $Y \sim Poisson(\theta T)$:

$$f(y|\theta) = \frac{e^{-\theta T}(\theta T)^y}{y!} \ .$$

- Prior: Assume a $Gamma(\alpha, \beta)$ prior for θ:

$$p(\theta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \ , \theta > 0 \ .$$

Heart Valves Study

- The gamma prior is conjugate with the likelihood, so the posterior emerges in closed form:

$$\begin{aligned} p(\theta|y) &\propto \theta^{y+\alpha-1} e^{-\theta(T+1/\beta)} \\ &\propto Gamma(y+\alpha, (T+1/\beta)^{-1}) \,. \end{aligned}$$

- The study objective is met if

$$P(\theta < 2\theta_0 \mid y) \geq 0.95 \,,$$

with $\theta_0 = OPC = 0.038$.

- Question: What values of α and β should we choose?

Heart Valves Study

- Prior selection: Prior can be determined based on the desired mean and variance!
- A gamma distribution has mean $M = \alpha\beta$ and variance $V = \alpha\beta^2$. This means that for a desired value of mean M and variance V, we can solve for α and β as

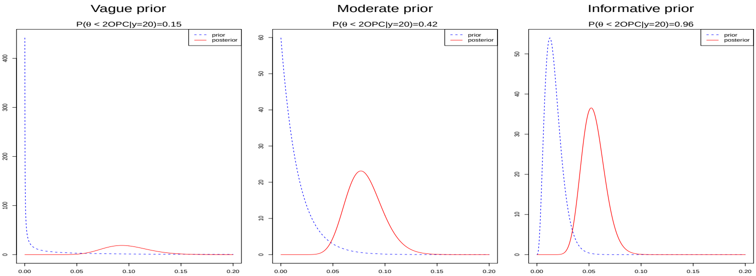
$$\alpha = M^2/V \text{ and } \beta = V/M \ .$$

- A few possibilities for prior parameters:
  1. Suppose we set $M = \theta_0 = 0.038$ and $\sqrt{V} = 2\theta_0$ (so that 0 is 0.5 standard deviations below the mean). Then $\alpha = 0.25$ and $\beta = 0.152$, a rather vague prior.
  2. Suppose we set $M = 98/5891 = .0166$, the overall value from the St. Jude studies, and $\sqrt{V} = M$ (so 0 is one sd below the mean). Then $\alpha = 1$ and $\beta = 0.0166$, a moderate (exponential) prior.
  3. Suppose we set $M = 98/5891 = .0166$ again, but set $\sqrt{V} = M/2$. This is a rather informative prior.

Heart Valves Study

- Suppose we observe $y = 20$ for the current study.
  - Plot the prior and posterior distributions.
  - Calculate $P(\theta < 2\theta_0 \mid y)$.
- Is the objective $P(\theta < 2\theta_0 \mid y) \geq 0.95$ met for each prior?
- What if you observe $y = 3$?

Heart Valves Study



The study objective is not met with the "bad" data – *unless* the posterior is "rescued" by the informative prior (lower right corner).