ORIE6217/CS6384: Applied Bayesian Data Analysis for Research

Lecture 3: Priors and multivariate models

Nikhil Garg

Announcements

Lecture plan

- Priors $P(\theta)$
- Multivariate models
  - Normal distribution
  - Regression

Priors

Priors

- Conjugate prior (BDA3 p. 35)
- Noninformative prior (BDA3 p. 51)
- Proper and improper prior (BDA3 p. 52)
- Weakly informative prior (BDA3 p. 55)
- Informative prior (BDA3 p. 55)
- Prior sensitivity (BDA3 p. 38)

Conjugate prior

- Prior and posterior have the same form
  - only for exponential family distributions (plus for some irregular cases)
- Used to be important for computational reasons, and still sometimes used for special models to allow partial analytic marginalization (Ch 3)
  - with dynamic Hamiltonian Monte Carlo used e.g. in Stan no computational benefit

Beta prior for Binomial model

- Prior

$$\text{Beta}(\theta|\alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- Posterior

$$\begin{aligned} p(\theta|y, n, M) &\propto \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \end{aligned}$$

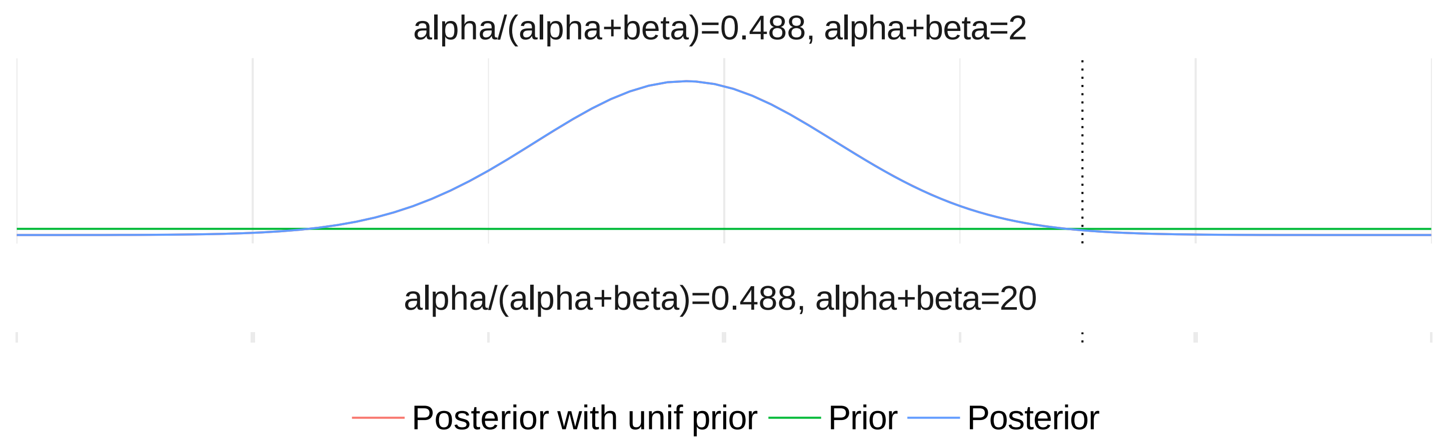
  after normalization

$$p(\theta|y, n, M) = \text{Beta}(\theta|\alpha + y, \beta + n - y)$$

- $(\alpha - 1)$ and $(\beta - 1)$ can considered to be number of prior observations
- Uniform prior when $\alpha = 1$ and $\beta = 1$

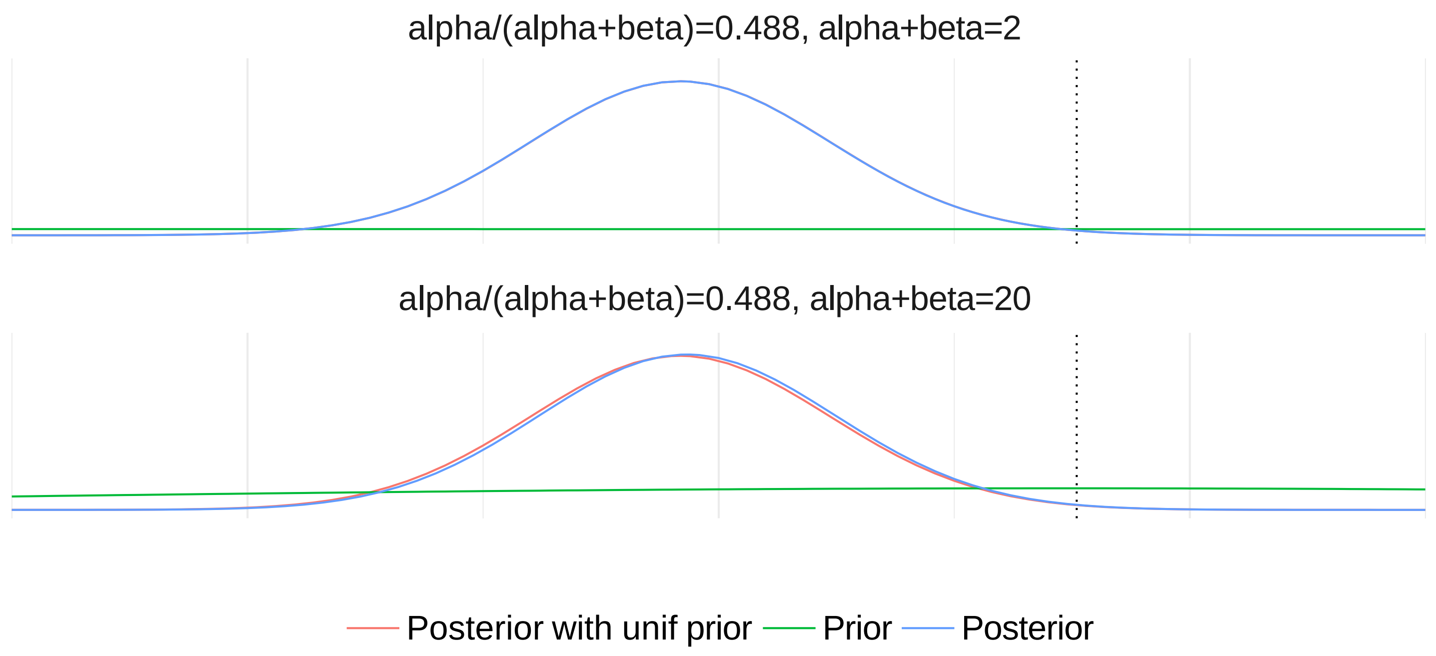
Placenta previa

- Beta prior centered on population average 0.485



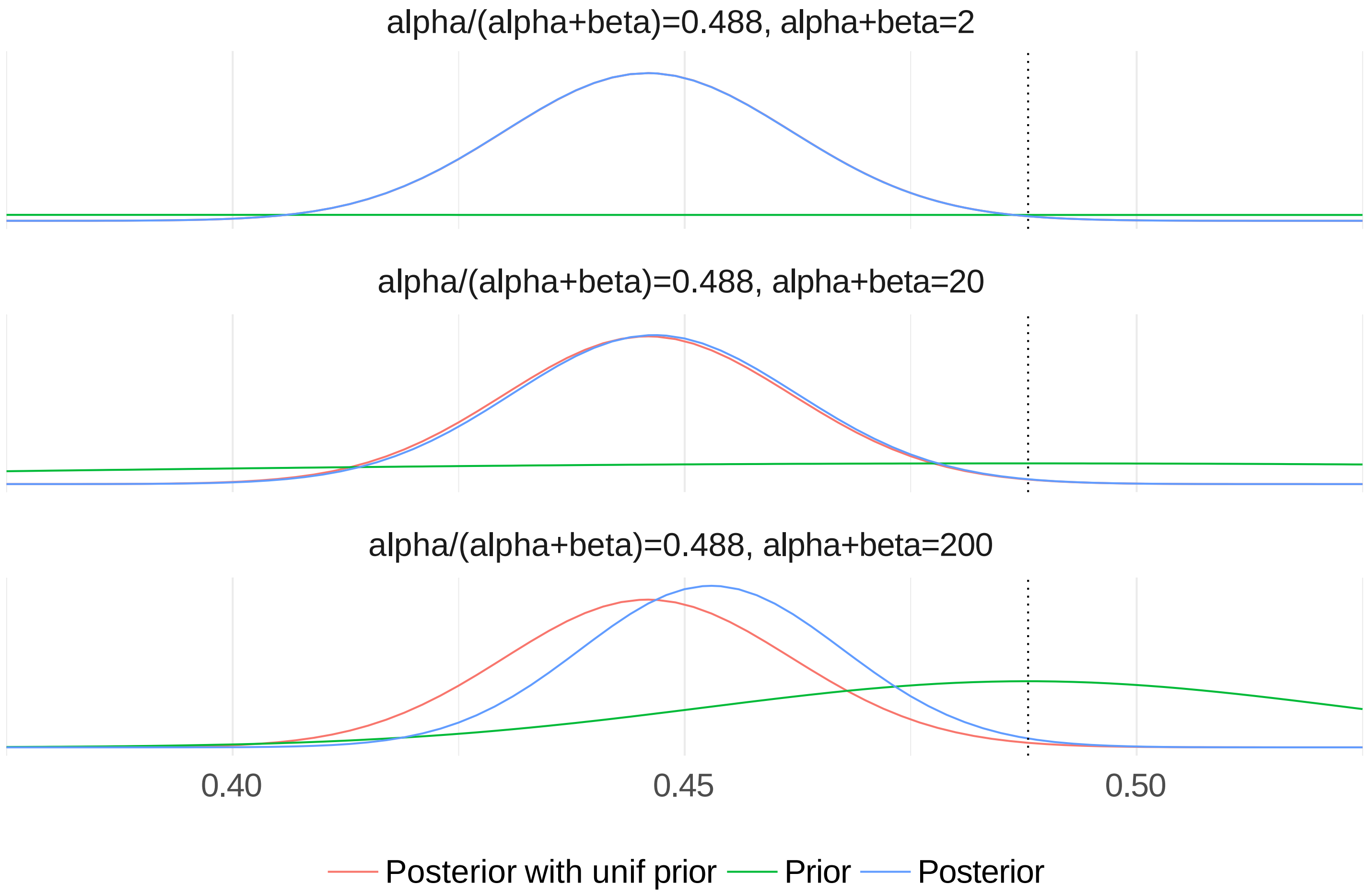
Placenta previa

- Beta prior centered on population average 0.485



Placenta previa

- Beta prior centered on population average 0.485



Noninformative prior, proper and improper prior

- Vague, flat, diffuse, or noninformative
  - try to “to let the data speak for themselves”
  - flat is not non-informative
  - flat can be stupid
  - making prior flat somewhere can make it non-flat somewhere else
- Proper prior has $\int p(\theta) = 1$
- Improper prior density doesn’t have a finite integral
  - the posterior can still sometimes be proper

Weakly informative priors

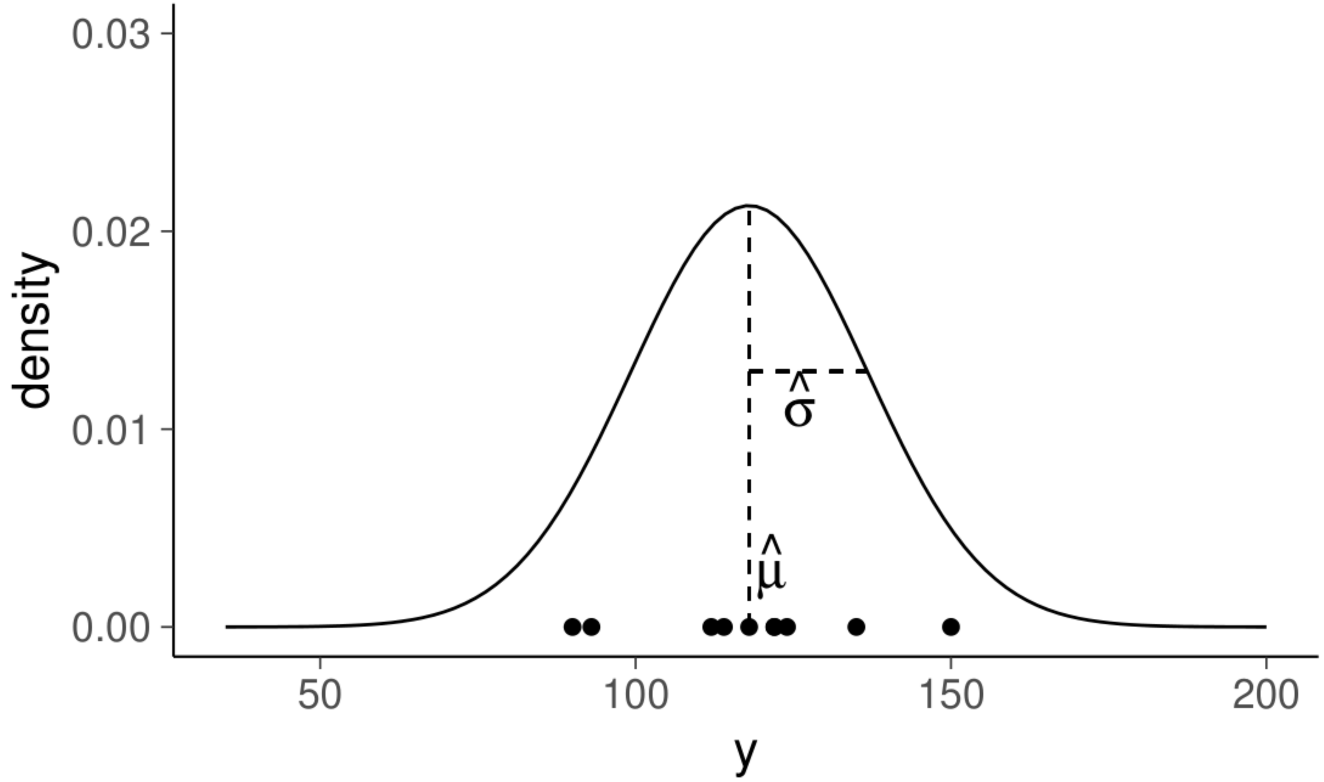
- Weakly informative priors produce computationally better behaving posteriors
  - quite often there's at least some knowledge about the scale
  - useful also if there's more information from previous observations, but not certain how well that information is applicable in a new case uncertainty

Weakly informative priors

- Weakly informative priors produce computationally better behaving posteriors
  - quite often there’s at least some knowledge about the scale
  - useful also if there’s more information from previous observations, but not certain how well that information is applicable in a new case uncertainty
- Construction
  - Start with some version of a noninformative prior distribution and then add enough information so that inferences are constrained to be reasonable.
  - Start with a strong, highly informative prior and broaden it to account for uncertainty in one’s prior beliefs and in the applicability of any historically based prior distribution to new data.
- Stan team prior choice recommendations https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Multivariate models

Normal distribution



$$p(y \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$

Calculate posterior for normal distribution with uninformative prior

Marginal posterior $p(\sigma^2 \mid y)$ (easier for σ^2 than σ)

$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \\
&\qquad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \\
&\qquad\qquad \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) d\theta = 1 \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\
p(\sigma^2 \mid y) &= \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)
\end{aligned}
$$

=> Posterior for σ^2

Known mean

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n, v)$$

$$\text{where} \quad v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2$$

Unknown mean

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n-1, s^2)$$

$$\text{where} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Posterior for μ

Marginal posterior $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$

$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Transformation

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{and} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

Recognize gamma integral $\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

$$\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2}$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

$$p(\mu \mid y) = t_{n-1}(\mu \mid \bar{y}, s^2/n)$$ Student's t

This is a lot of math!

Let’s see what this looks like in [Stan] code

Regression

We can also do regression models in Stan

Logistic regression definition:

$$y \sim \text{Bernoulli}\left(\text{logit}^{-1}(\alpha + \beta * x)\right)$$

Where

$$\text{logit}^{-1}(z) = \frac{1}{1 + \exp(-z)}$$

α: intercept

$\beta \in R^d$: coefficient vector

Let’s see what this looks like in code

What about priors for regression?

Priors are equivalent to regularization!

- Normal prior $\sigma = 1/\sqrt{\lambda} \Leftrightarrow$ L2 regularization
- Laplace prior $\Leftrightarrow$ L1 regularization

Read more: https://bjlkeng.github.io/posts/probabilistic-interpretation-of-regularization/

[Look at code]

Questions?