Bayesian estimation of binary regression models

Applied Bayesian Statistics
Winter Term 2018

Susumu Shikano
GSDS

Last session

- In Bayesian inference the (posterior) probability of individual models is at stake.
- You can obtain the posterior even after observing one single case.
- The posterior also depends on your prior belief.
- For a larger number of observations frequentists and Bayesian have similar results.
- Conjugacy of the beta distribution with the likelihood based on the binomial distribution.

Last session

Example

Last session

Example

- Model: capacity of face-based inference of candidate ideology

Last session

Example

- Model: capacity of face-based inference of candidate ideology
- Parameter of interest: π

Last session

Example

- Model: capacity of face-based inference of candidate ideology
- Parameter of interest: π
- $\rightarrow$ Only one parameter is at stake.

Last session

Example

- Model: capacity of face-based inference of candidate ideology
- Parameter of interest: π
- $\rightarrow$ Only one parameter is at stake.
- In praxis there are multiple parameters.

Last session

Example

- Model: capacity of face-based inference of candidate ideology
- Parameter of interest: π
- $\rightarrow$ Only one parameter is at stake.
- In praxis there are multiple parameters.
- e.g. (bivariate) regression models

Linear Regression Model

Two approaches to obtain posterior

- Conjugacy analysis
    - Conjugacy: The property that the prior and posterior have the same probability form depending on the form of the distribution used to calculate the likelihood.
    - Posterior can be obtained analytically.

Linear Regression Model

Two approaches to obtain posterior

- Conjugacy analysis
  - Conjugacy: The property that the prior and posterior have the same probability form depending on the form of the distribution used to calculate the likelihood.
  - Posterior can be obtained analytically.
- Deriving posterior per Gibbs Sampling
  - Conjugacy is not must.
  - Use of MCMC.

Conjugacy analysis

(Bivariate) linear regression model

$$\begin{aligned} y &= \beta_0 + \beta_1 x + \epsilon \\ \epsilon &\sim N(0, \sigma^2) \end{aligned}$$

$$(y|\beta_0, \beta_1, \sigma^2, x) \sim N(\beta_0 + \beta_1 x, \sigma^2)$$

Conjugacy analysis

(Bivariate) linear regression model

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x + \epsilon \\
\epsilon &\sim N(0, \sigma^2)
\end{aligned}
$$

$$
(y|\beta_0, \beta_1, \sigma^2, x) \sim N(\beta_0 + \beta_1 x, \sigma^2)
$$

The unknown parameters to be estimated

Conjugacy analysis

(Bivariate) linear regression model

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x + \epsilon \\
\epsilon &\sim N(0, \sigma^2)
\end{aligned}
$$

$$(y|\beta_0, \beta_1, \sigma^2, x) \sim N(\beta_0 + \beta_1 x, \sigma^2)$$

The unknown parameters to be estimated

- β_0
- β_1
- σ^2

Conjugacy analysis

How to proceed

- Likelihood?
- Prior?
- Posterior?

Conjugacy analysis

How to proceed

- Likelihood? $\rightarrow$ Normal distribution
- Prior?
- Posterior?

Conjugacy analysis

How to proceed

- Likelihood? $\rightarrow$ Normal distribution
- Prior? $\rightarrow$ Normal-Inverse-Gamma
- Posterior?

Conjugacy analysis

How to proceed

- Likelihood? $\rightarrow$ Normal distribution
- Prior? $\rightarrow$ Normal-Inverse-Gamma
- Posterior? $\rightarrow$ Normal-Inverse-Gamma

Conjugacy analysis

How to proceed

- Likelihood? $\rightarrow$ Normal distribution
- Prior? $\rightarrow$ Normal-Inverse-Gamma
- Posterior? $\rightarrow$ Normal-Inverse-Gamma

Conjugacy!!

Conjugacy analysis

How to proceed

- Likelihood? $\rightarrow$ Normal distribution
- Prior? $\rightarrow$ Normal-Inverse-Gamma
- Posterior? $\rightarrow$ Normal-Inverse-Gamma

Conjugacy!!

Likelihood Function

$$f_N(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) = \prod_i^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \boldsymbol{X}\beta)'(y_i - \boldsymbol{X}\beta)}{2\sigma^2}\right\}$$

Conjugacy analysis

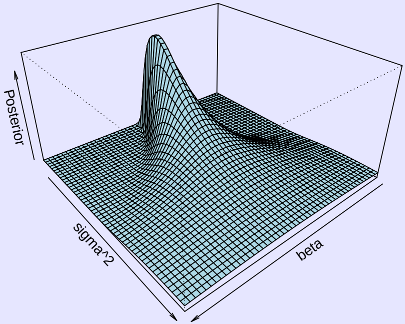
Normal-Inverse-Gamma?

It looks like…

Conjugacy analysis

Normal-Inverse-Gamma?

It looks like…



Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

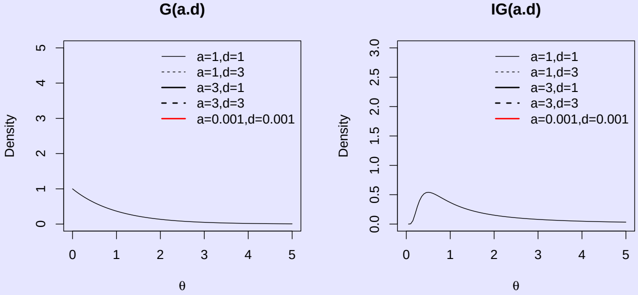
$$f_\Gamma(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp\left(-a\theta\right); \; f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$

Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

$$f_\Gamma(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp(-a\theta)\,; \quad f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$

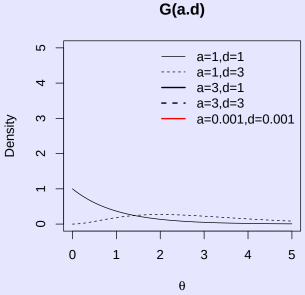


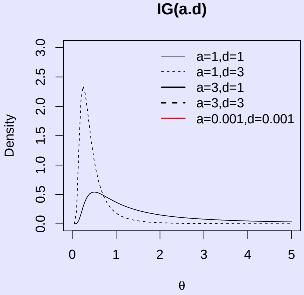
Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

$$f_\Gamma(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp\left(-a\theta\right); \quad f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$



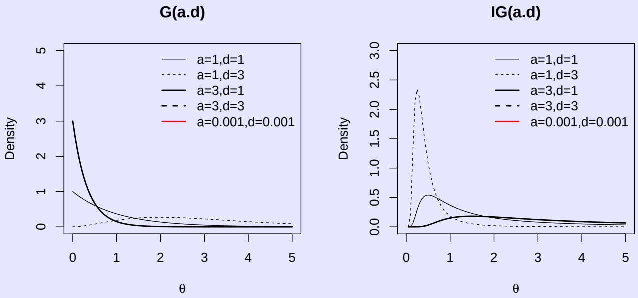


Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

$$f_\Gamma(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp(-a\theta)\,; \quad f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$

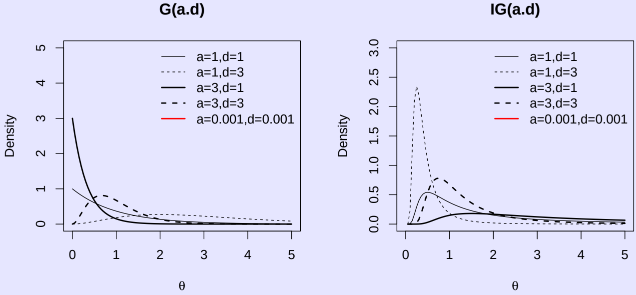


Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

$$f_\Gamma(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp(-a\theta)\,; \quad f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$

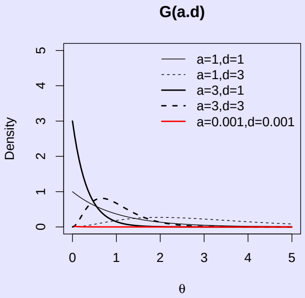


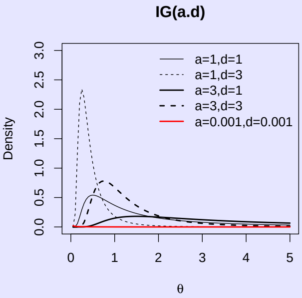
Conjugacy analysis

Inverse-Gamma $IG(a, d)$ with a as rate and d as shape

If $X \sim G(a, d)$, then $1/X \sim IG(a, d)$:

$$f_{\Gamma}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{d-1} \exp(-a\theta)\,; \quad f_{\Gamma^{-1}}(\theta) = \frac{a^d}{\Gamma(d)} \theta^{-d-1} \exp\left(-\frac{a}{\theta}\right)$$





Conjugacy analysis

Prior: Normal-Inverse-Gamma

$$
\begin{aligned}
f_{N-\Gamma-1}(\beta, \sigma^2|\mu, \lambda, a, d) &= f_N(\beta|\mu, \sigma^2/\lambda) f_{\Gamma-1}(\sigma^2|a, d) \\
&= \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\lambda(\beta-\mu)^2}{2\sigma^2}\right\} \underbrace{\frac{a^d}{\Gamma(d)} \sigma^{2(-d-1)} \exp\left(-\frac{a}{\sigma^2}\right)}_{\text{Inverse-Gamma-Dist.}} \\
&= \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \frac{a^d}{\Gamma(d)} \left(\frac{1}{\sigma^2}\right)^{d+1} \exp\left\{-\frac{\lambda(\beta-\mu)^2 + 2a}{2\sigma^2}\right\}
\end{aligned}
$$

Conjugacy analysis

Prior: Normal-Inverse-Gamma

$$
\begin{aligned}
f_{N-\Gamma-1}(\beta,\sigma^2|\mu,\lambda,a,d) &= f_N(\beta|\mu,\sigma^2/\lambda) f_{\Gamma-1}(\sigma^2|a,d) \\
&= \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\lambda(\beta-\mu)^2}{2\sigma^2}\right\} \underbrace{\frac{a^d}{\Gamma(d)} \sigma^{2(-d-1)} \exp\left(-\frac{a}{\sigma^2}\right)}_{\text{Inverse-Gamma-Dist.}} \\
&= \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \frac{a^d}{\Gamma(d)} \left(\frac{1}{\sigma^2}\right)^{d+1} \exp\left\{-\frac{\lambda(\beta-\mu)^2+2a}{2\sigma^2}\right\}
\end{aligned}
$$

The prior is more precisely a multivariate normal-inverse gamma distribution:

$$
\begin{aligned}
f_{N-\Gamma-1}(\boldsymbol{\beta},\sigma^2|\boldsymbol{\mu},\boldsymbol{\Sigma},a,d) &= f_N(\boldsymbol{\beta}|\boldsymbol{\mu},\sigma^2\cdot\boldsymbol{\Sigma}) f_{\Gamma-1}(\sigma^2|a,d) \\
&= \frac{1}{\sqrt{2\pi\sigma^k|\boldsymbol{\Sigma}|}} \frac{a^d}{\Gamma(d)} \left(\frac{1}{\sigma^2}\right)^{d+1} \\
&\quad \times \exp\left\{-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})+2a}{2\sigma^2}\right\}
\end{aligned}
$$

Conjugacy analysis

Deriving posterior

$$
\begin{aligned}
\text{Prior } f(\boldsymbol{\beta}, \sigma^2) &= f_{N-\Gamma^{-1}}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0, a_0, d_0) \\
&= \frac{1}{\sqrt{2\pi\sigma^k|\boldsymbol{\Sigma_0}|}} \frac{{a_0}^{d_0}}{\Gamma(d_0)} \left(\frac{1}{\sigma_0^2}\right)^{d_0+1} \\
&\quad \times \exp\left\{-\frac{(\boldsymbol{\beta}_0 - \boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_0 - \boldsymbol{\mu_0}) + 2a_0}{2\sigma^2}\right\} \\
\text{Likelihood } f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) &= \prod_i^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \boldsymbol{X}\beta)'(y_i - \boldsymbol{X}\beta)}{2\sigma^2}\right\} \\
\text{Posterior } f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2)}{f(\boldsymbol{y})} \\
f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &= f_{N-\Gamma^{-1}}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, a^*, b^*)
\end{aligned}
$$

Conjugacy analysis

Posterior

$$
\begin{aligned}
\text{Posterior } f(\beta, \sigma^2 | \boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\beta, \sigma^2) f(\beta, \sigma^2)}{f(\boldsymbol{y})} \\
&= f_{N-\Gamma-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, a^*, d^*) \\
\text{with} & \\
\boldsymbol{\mu}^* &= (\boldsymbol{\Sigma}_{\mathbf{0}}^{-\mathbf{1}} + \boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{\Sigma}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\mu}_{\mathbf{0}} + \boldsymbol{X}'\boldsymbol{y}) \\
\boldsymbol{\Sigma}^* &= (\boldsymbol{\Sigma}_{\mathbf{0}}^{-\mathbf{1}} + \boldsymbol{X}'\boldsymbol{X})^{-1} \\
a^* &= a_0 + \frac{1}{2}(\boldsymbol{\mu}_{\mathbf{0}}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_{\mathbf{0}} + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\mu}^{*\prime}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\mu}^*) \\
d^* &= d_0 + \frac{n}{2}
\end{aligned}
$$

Conjugacy analysis

Posterior

$$
\begin{aligned}
\text{Posterior } f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) &= \frac{f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2)}{f(\mathbf{y})} \\
&= f_{N-\Gamma^{-1}}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, a^*, d^*) \\
\text{with} & \\
\boldsymbol{\mu}^* &= (\boldsymbol{\Sigma}_{\mathbf{0}}^{\mathbf{-1}} + \mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\Sigma}_{\mathbf{0}}^{\mathbf{-1}} \boldsymbol{\mu}_{\mathbf{0}} + \mathbf{X}'\mathbf{y}) \\
\boldsymbol{\Sigma}^* &= (\boldsymbol{\Sigma}_{\mathbf{0}}^{\mathbf{-1}} + \mathbf{X}'\mathbf{X})^{-1} \\
a^* &= a_0 + \frac{1}{2}(\boldsymbol{\mu}_{\mathbf{0}}' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_{\mathbf{0}} + \mathbf{y}'\mathbf{y} - \boldsymbol{\mu}^{*\prime} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\mu}^*) \\
d^* &= d_0 + \frac{n}{2}
\end{aligned}
$$

Relationship to ML/OLS

If n and/or Σ_0 approach to infinity:

Conjugacy analysis

Posterior

$$\begin{aligned}
\text{Posterior } f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2)}{f(\boldsymbol{y})} \\
&= f_{N-\Gamma^{-1}}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, a^*, d^*) \\
\text{with} & \\
\boldsymbol{\mu}^* &= (\boldsymbol{\Sigma_0^{-1}} + \boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{\Sigma_0^{-1}}\boldsymbol{\mu_0} + \boldsymbol{X}'\boldsymbol{y}) \\
\boldsymbol{\Sigma}^* &= (\boldsymbol{\Sigma_0^{-1}} + \boldsymbol{X}'\boldsymbol{X})^{-1} \\
a^* &= a_0 + \frac{1}{2}(\boldsymbol{\mu_0'}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu_0} + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\mu}^{*\prime}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\mu}^*) \\
d^* &= d_0 + \frac{n}{2}
\end{aligned}$$

Relationship to ML/OLS

If n and/or Σ_0 approach to infinity:

$$\lim_{n\to\infty} \boldsymbol{\mu}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Conjugacy analysis

Posterior is not posterior…

- Joint posterior
- Conditional posterior
- Marginal posterior

Conjugacy analysis

Posterior is not posterior...

- Joint posterior
- Conditional posterior
- Marginal posterior

Marginal posterior of β

By integrating out σ^2, we can obtain a multivariate student t-distribution with a degree of freedom of $\nu = n - k$:

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{y}) &= \int_0^\infty f(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) d\sigma^2 \\ &= f_t(\nu, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \end{aligned}$$

Wrap up

Bayesian regression models

- In conjugacy analysis posterior can be derived analytically.
- The larger *n*/the smaller the dispersion of prior, the more similar results with the maximum likelihood.