Macroeconometrics

Lecture 4 Bayesian Estimation

Tomasz Woźniak

Department of Economics
University of Melbourne

Bayes' rule

Useful distributions

Likelihood function

Prior distribution

Posterior distribution

Compulsory reading:
Woźniak (2021) Posterior derivations for a simple linear regression model, Lecture notes

Greenberg (2008) Chapter 4: Prior Distributions, Introduction to Bayesian Econometrics

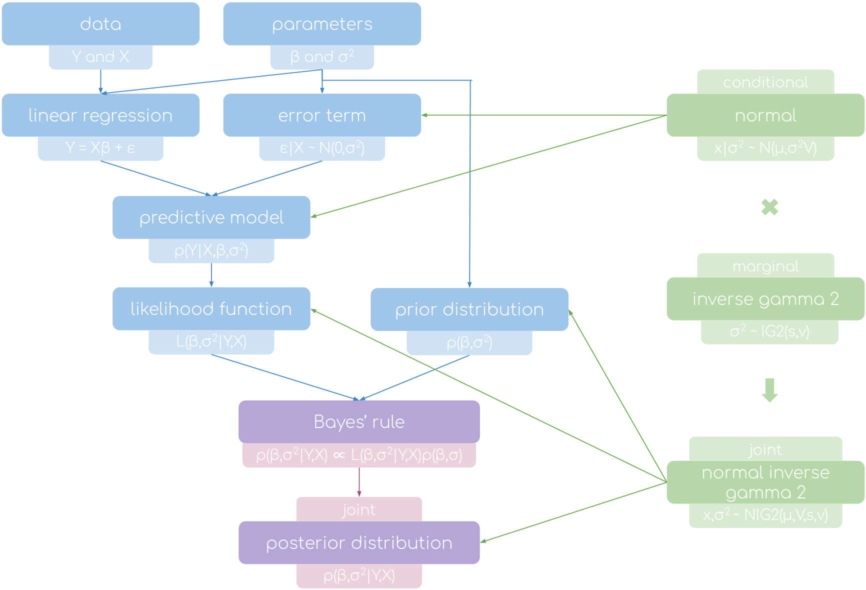
Materials:
An R file `L3 grphs.R` for the reproduction of graphs
Selected quotations from Sims (2012) – Bayesian vs. frequentist approach

Objectives.

- To learn the basics of Bayesian inference
- To familiarise with essential probability distributions
- To derive Bayesian estimation results for a simple model

Learning outcomes.

- Using basic Bayesian terminology
- Identifying parameters of transformed distributions
- Recognising functional form of kernels of basic distributions


data
Y and X
parameters
β and σ²
linear regression
Y = Xβ + ε
error term
ε|X ~ N(0,σ²)
conditional
normal
x|σ² ~ N(μ,σ²V)
predictive model
ρ(Y|X,β,σ²)
marginal
inverse gamma 2
σ² ~ IG2(s,ν)
likelihood function
L(β,σ²|Y,X)
prior distribution
ρ(β,σ²)
Bayes' rule
ρ(β,σ²|Y,X) ∝ L(β,σ²|Y,X)ρ(β,σ)
joint
normal inverse gamma 2
x,σ² ~ NIG2(μ,V,s,ν)
joint
posterior distribution
ρ(β,σ²|Y,X)

Bayes' rule

Bayes' rule

$$p(\theta|Y) = \frac{p(Y|\theta)\,p(\theta)}{p(Y)}$$

$p(\theta|Y)$ - posterior distribution of parameters θ given data Y

$p(Y|\theta)$ - sampling distribution of data Y given parameters θ

$p(\theta)$ - prior distribution of parameters θ

$p(Y)$ - marginal data density

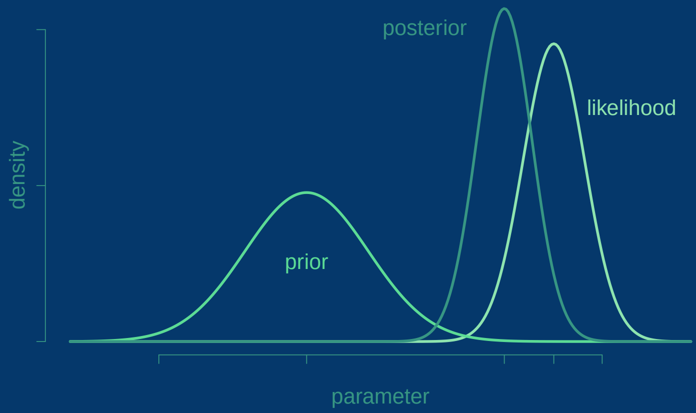
Bayes' rule

$$p(\theta|Y) \propto L(\theta|Y)\, p(\theta)$$

Learning mechanism.
The prior information about the parameters is updated by the information contained in the data and represented by the likelihood function resulting in the posterior distribution.

Likelihood principle.
All the information about the parameters of the model included in the data is captured by the likelihood function.

Bayes' rule



Bayes' rule

Joint distribution of data and parameters.

$$p(Y|\theta)\, p(\theta) = p(Y,\theta) = p(\theta|Y)\, p(Y)$$

The joint distribution of data and parameters is decomposed into:

Inputs: likelihood function $p(Y|\theta)$ and prior distribution $p(\theta)$

Outputs: posterior distribution $p(\theta|Y)$ and marginal data density $p(Y)$

Useful distributions

Multivariate normal distribution

Let an $N \times 1$ real-valued random vector X follow a multivariate normal distribution:

$$X \sim \mathcal{N}_N(\mu, \Sigma)$$

with the mean vector μ and the covariance matrix Σ.

pdf.

$$\mathcal{N}_N(\mu, \Sigma) = (2\pi)^{-\frac{N}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right\}$$

Moments.

$$\mathbb{E}(X) = \mu, \text{ and } Var(X) = \Sigma$$

Inverse gamma 2 distribution

Let a positive real-valued scalar random variable x follow an inverse gamma 2 distribution:

$$x \sim \mathcal{IG}2(s, \nu)$$

with the shape parameter $\nu > 0$ and the scale parameter $s > 0$.

pdf.

$$\mathcal{IG}2(s, \nu) = \Gamma\left(\frac{\nu}{2}\right)^{-1}\left(\frac{s}{2}\right)^{\frac{\nu}{2}} x^{-\frac{\nu+2}{2}} \exp\left\{-\frac{1}{2}\frac{s}{x}\right\}$$

Moments.

$$\mathbb{E}(x) = \frac{s}{\nu - 2}, \text{ for } \nu > 2, \;\; Var(x) = \frac{2}{\nu - 4}\left[\mathbb{E}(x)\right]^2, \text{ for } \nu > 4$$

$$mode = \frac{s}{\nu + 2}$$

Normal inverse gamma 2 distribution

$$p\left(X|\sigma^2\right) = \mathcal{N}_N\left(\mu, \sigma^2\Sigma\right)$$
$$p\left(\sigma^2\right) = \mathcal{IG}2(s, \nu)$$

Then, (X, σ^2) follow a normal inverse gamma 2 distribution:

$$p(X, \sigma^2) = p(X|\sigma^2)p(\sigma^2) = \mathcal{NIG}2_N\left(\mu, \Sigma, s, \nu\right)$$

pdf.

$$\mathcal{NIG}2\left(\mu, \Sigma, s, \nu\right) = c_{nig2}^{-1}\left(\sigma^2\right)^{-\frac{\nu+N+2}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\left[s + (X-\mu)'\Sigma^{-1}(X-\mu)\right]\right\}$$

$$c_{nig2} = \Gamma\left(\frac{\nu}{2}\right)\left(\frac{s}{2}\right)^{-\frac{\nu}{2}}(2\pi)^{\frac{N}{2}}\det(\Sigma)^{\frac{1}{2}}$$

Moments.

$$\mathbb{E}\left(X\right) = \mu, \text{ for } \nu > 1, \;\; Var\left(X\right) = \frac{s}{\nu-2}\Sigma, \text{ for } \nu > 2$$

$$\mathbb{E}\left(\sigma^2\right) = \frac{s}{\nu-2}, \text{ for } \nu > 2, \;\; Var\left(\sigma^2\right) = \frac{2}{\nu-4}\left[\mathbb{E}\left(\sigma^2\right)\right]^2, \text{ for } \nu > 4$$

Normal inverse gamma 2 distribution

Kernel of the $\mathcal{NIG}2$ distribution.

$$\mathcal{NIG}2\left(\mu, \Sigma, s, \nu\right) \propto \left(\sigma^2\right)^{-\frac{\nu+N+2}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(X-\mu)'\Sigma^{-1}(X-\mu)\right\} \exp\left\{-\frac{1}{2}\frac{s}{\sigma^2}\right\}$$

Normal inverse gamma 2 distribution

Generating random numbers from the $\mathcal{NIG}2$ distribution.

$$p\left(\beta, \sigma^2\right) = p\left(\beta|\sigma^2\right) p\left(\sigma^2\right)$$

$$p\left(\beta|\sigma^2\right) = \mathcal{N}\left(\mu, \sigma^2\Sigma\right)$$
$$p\left(\sigma^2\right) = \mathcal{IG}2\left(s, \nu\right)$$

To draw S draws from the $\mathcal{NIG}2$ distribution...

Step 1: Draw independently S draws from the $\mathcal{IG}2\left(s, \nu\right)$. Collect these draws in sequence $\left\{\sigma^{2(s)}\right\}_{s=1}^{S}$

Step 2: For each $\sigma^{2(s)}$ sample a corresponding draw of $\beta^{(s)}$ from $\mathcal{N}\left(\mu, \sigma^{2(s)}\Sigma\right)$

Return: $\left\{\beta^{(s)}, \sigma^{2(s)}\right\}_{s=1}^{S}$ as draws from the target distribution.

Likelihood function

Likelihood function

A simple linear regression model.

$$
\begin{aligned}
Y &= \beta X + E \\
E|X &\sim \mathcal{N}\left(\mathbf{0}_T, \sigma^2 I_T\right) \\
&\downarrow \\
Y|X &\sim \mathcal{N}\left(\beta X, \sigma^2 I_T\right)
\end{aligned}
$$

The likelihood function.

$$
L\left(\theta | Y, X\right) = (2\pi)^{-\frac{T}{2}} \left(\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y - \beta X)'(Y - \beta X)\right\}
$$

Likelihood function

The likelihood function as the $\mathcal{NIG}2$ distribution.

$$L(\theta|Y,X) \propto \left(\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y-\beta X)'(Y-\beta X)\right\}$$

$$= \left(\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y-\hat{\beta}X+\hat{\beta}X-\beta X)'(Y-\hat{\beta}X+\hat{\beta}X-\beta X)\right\}$$

$$= \left(\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\left[(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})+(Y-\hat{\beta}X)'(Y-\hat{\beta}X)\right]\right\}$$

$$= \left(\sigma^2\right)^{-\frac{T-3+1+2}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right\} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y-\hat{\beta}X)'(Y-\hat{\beta}X)\right\}$$

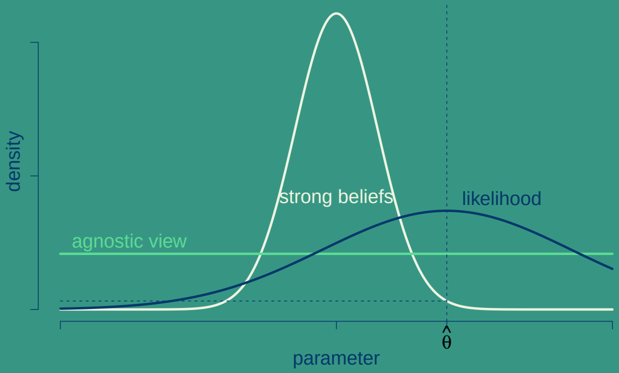
The result.

$$L(\theta|Y,X) = \mathcal{NIG}2\left(\mu=\hat{\beta}, \Sigma=(X'X)^{-1}, s=(Y-\hat{\beta}X)'(Y-\hat{\beta}X), \nu=T-3\right)$$

where $N=1$.

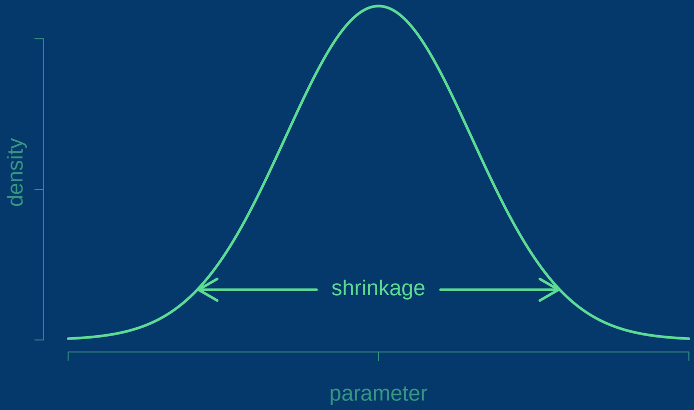
Prior distribution

Prior distribution



A prior distribution formalizes researcher's beliefs regarding the parameters of the model before seeing the data.

Prior distribution



Natural-conjugate prior distribution

A natural-conjugate prior distribution is of the same form as the distribution of the parameters implied by the likelihood function.

$$p\left(\beta, \sigma^2\right) = p\left(\beta|\sigma^2\right) p\left(\sigma^2\right)$$

$$p\left(\beta|\sigma^2\right) = \mathcal{N}\left(\underline{\beta}, \sigma^2\underline{\sigma}_{\beta}^2\right)$$

$$p\left(\sigma^2\right) = \mathcal{IG}2(\underline{s}, \underline{\nu})$$

Then, (β, σ^2) follow a priori a normal inverse gamma 2 distribution:

$$p\left(\beta, \sigma^2\right) = \mathcal{NIG}2_N\left(\underline{\beta}, \underline{\sigma}_{\beta}^2, \underline{s}, \underline{\nu}\right)$$

pdf.

$$\mathcal{NIG}2_{(N=1)}\left(\underline{\beta}, \underline{\sigma}_{\beta}^2, \underline{s}, \underline{\nu}\right) \propto \left(\sigma^2\right)^{-\frac{\underline{\nu}+3}{2}} \mathbf{exp}\left\{-\frac{1}{2}\frac{1}{\sigma^2}\frac{1}{\underline{\sigma}_{\beta}^2}(\beta-\underline{\beta})'(\beta-\underline{\beta})\right\}\mathbf{exp}\left\{-\frac{1}{2}\frac{\underline{s}}{\sigma^2}\right\}$$

Posterior distribution

Posterior distribution

$$p\left(\beta, \sigma^2 | Y, X\right) \propto L\left(Y|X, \beta, \sigma^2\right) p\left(\beta, \sigma^2\right)$$
$$= L\left(Y|X, \beta, \sigma^2\right) p\left(\beta|\sigma^2\right) p\left(\sigma^2\right)$$

Kernel of posterior distribution.

$$p\left(\beta, \sigma^2 | Y, X\right) \propto \left(\sigma^2\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right\} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}(Y - \hat{\beta}X)'(Y - \hat{\beta}X)\right\}$$
$$\times \left(\sigma^2\right)^{-\frac{\underline{\nu}+3}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\frac{1}{\underline{\sigma}_\beta^2}(\beta - \underline{\beta})'(\beta - \underline{\beta})\right\} \exp\left\{-\frac{1}{2}\frac{\underline{s}}{\sigma^2}\right\}$$

Posterior distribution

Kernel of posterior distribution.

$$p\left(\beta, \sigma^2 | Y, X\right) \propto \left(\sigma^2\right)^{-\frac{\underline{\nu}+T+3}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\left[\frac{1}{\underline{\sigma}^2_{\beta}}(\beta-\underline{\beta})'(\beta-\underline{\beta}) + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta}) \right.\right.$$
$$\left.\left. + \ \underline{s} + (Y-\hat{\beta}X)'(Y-\hat{\beta}X)\right]\right\}$$

After derivations, the expression in the square parentheses can be shown to have the following form:

$$\frac{1}{\underline{\sigma}^2_{\beta}}(\beta-\underline{\beta})'(\beta-\underline{\beta}) + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta}) + \underline{s} + (Y-\hat{\beta}X)'(Y-\hat{\beta}X)$$
$$= \overline{\sigma}_{\beta}^{-2}(\beta-\overline{\beta})'(\beta-\overline{\beta}) + \underline{s} + \underline{\beta}^2\underline{\sigma}_{\beta}^{-2} - \overline{\beta}^2\overline{\sigma}_{\beta}^{-2} + Y'Y$$

where expressions for $\overline{\beta}$ and $\overline{\sigma}^2_{\beta}$ are given on the next slides.

Posterior distribution

After plugging in the expression, the kernel of the posterior distribution takes the form of:

$$p\left(\beta, \sigma^2 | Y, X\right) \propto \left(\sigma^2\right)^{-\frac{\overline{\nu}+3}{2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2}\frac{1}{\overline{\sigma}_\beta^2}(\beta - \overline{\beta})'(\beta - \overline{\beta})\right\} \exp\left\{-\frac{1}{2}\frac{\overline{s}}{\sigma^2}\right\}$$

in which we recognize the kernel of the normal inverse gamma 2 distribution.

Posterior distribution

$$p\left(\beta, \sigma^2 | Y, X\right) = \mathcal{NIG}2_{(N=1)}\left(\overline{\beta}, \overline{\sigma}_\beta^2, \overline{s}, \overline{\nu}\right)$$

$$\begin{aligned}
\overline{\sigma}_\beta^2 &= \left(\underline{\sigma}_\beta^{-2} + X'X\right)^{-1} \\
\overline{\beta} &= \overline{\sigma}_\beta^2 \left(\underline{\sigma}_\beta^{-2}\underline{\beta} + X'Y\right) \\
\overline{s} &= \underline{s} + \underline{\sigma}_\beta^{-2}\underline{\beta}^2 - \overline{\sigma}_\beta^{-2}\overline{\beta}^2 + Y'Y \\
\overline{\nu} &= \underline{\nu} + T
\end{aligned}$$

Posterior distribution

The posterior mean of β.

$$
\begin{aligned}
\overline{\beta} &= \overline{\sigma}_\beta^2 \left(\underline{\sigma}_\beta^{-2} \underline{\beta} + X'Y \right) \\
&= \overline{\sigma}_\beta^2 \underline{\sigma}_\beta^{-2} \underline{\beta} + \overline{\sigma}_\beta^2 X'X (X'X)^{-1} X'Y \\
&= \frac{\underline{\sigma}_\beta^{-2}}{\underline{\sigma}_\beta^{-2} + X'X} \underline{\beta} + \frac{X'X}{\underline{\sigma}_\beta^{-2} + X'X} \hat{\beta} \\
&= \omega \underline{\beta} + (1 - \omega) \hat{\beta}
\end{aligned}
$$

The posterior mean of β is the weighted average between the prior mean $\underline{\beta}$ and the MLE $\hat{\beta}$.

Posterior distribution

$$\overline{\beta} = \frac{\underline{\sigma}_{\beta}^{-2}}{\underline{\sigma}_{\beta}^{-2} + X'X}\underline{\beta} + \frac{X'X}{\underline{\sigma}_{\beta}^{-2} + X'X}\hat{\beta}$$

$$= \frac{\frac{\underline{\sigma}_{\beta}^{-2}}{T}}{\frac{\underline{\sigma}_{\beta}^{-2}}{T} + \frac{X'X}{T}}\underline{\beta} + \frac{\frac{X'X}{T}}{\frac{\underline{\sigma}_{\beta}^{-2}}{T} + \frac{X'X}{T}}\hat{\beta}$$

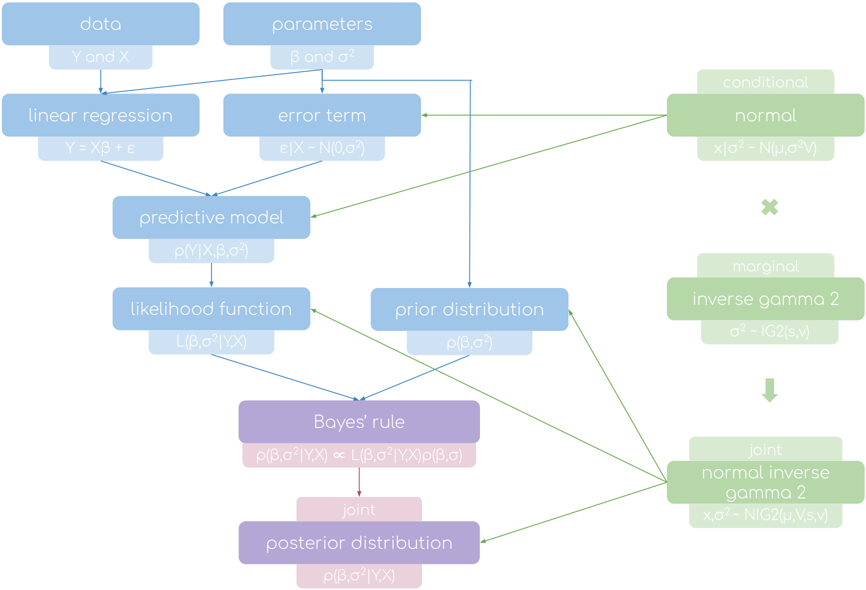
Limits.

$\lim_{T\to\infty} \frac{\underline{\sigma}_{\beta}^{-2}}{T} = 0$ – as $\underline{\sigma}_{\beta}^{-2}$ is a constant

$\lim_{T\to\infty} \frac{X'X}{T} = \sigma_X^2$ – σ_X^2 is the second non-central moment of X

The posterior mean of β when $T \to \infty$.

$$\lim_{T\to\infty} \overline{\beta} = \hat{\beta}$$



data
Y and X
parameters
β and σ²
linear regression
Y = Xβ + ε
error term
ε|X ~ N(0,σ²)
predictive model
p(Y|X,β,σ²)
likelihood function
L(β,σ²|Y,X)
prior distribution
p(β,σ²)
Bayes' rule
p(β,σ²|Y,X) ∝ L(β,σ²|Y,X)p(β,σ)
joint
posterior distribution
p(β,σ²|Y,X)
conditional
normal
x|σ² ~ N(μ,σ²V)
marginal
inverse gamma 2
σ² ~ IG2(s,v)
joint
normal inverse gamma 2
x,σ² ~ NIG2(μ,V,s,v)

Bayesian estimation

For a linear Gaussian regression:

Likelihood function has a form of a Normal inverse gamma 2 distribution for the parameters of the model

Normal inverse gamma 2 distribution for the parameters is the naturally-conjugate prior distribution leading to...

Normal inverse gamma 2 posterior distribution

Asymptotically Bayesian estimation converges to the MLE