Bayesian Inference in Univariate and Multivariate Time Series

Giovanni Ricco

3rd November 2023

Bayesian Inference

Frequentist Inference

1. Formulate a parametric model as a collection of probability distributions of all possible realisation of the data Z conditional on different values of the model parameters $\theta \in \Theta$

   Model: $p(Z|\theta)$

2. Collect the data z and treat them as realisations of Z and insert them into the family of distributions

   Likelihood: $\mathcal{L}(z|\theta) = p(z|\theta)$

3. Estimate the parameters by maximising the likelihood of the data

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(z|\theta)$$

Bayesian Inference

Let A and B be two events, the joint probability:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Bayes Rule:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)} \propto P(B|A)P(A)}$$

- $P(A|B)$ is a conditional probability, i.e. the **posterior probability** of A given B
- $P(B|A)$ is also a conditional probability, it can be interpreted as the **likelihood** of A given a fixed B
- $P(A)$ and $P(B)$ are the probabilities of observing A and B unconditionally: can be interpreted as the **prior** probability and **marginal probability**

Bayesian Inference

1. Formulate a parametric model as a collection of probability distributions of all possible realisation of the data Z conditional on different values of the model parameters $\theta \in \Theta$

   Model: $p(Z|\theta)$

2. Organise the belief about θ into a (prior) probability distribution over Θ

   Prior: $p(\theta)$

3. Collect the data z and treat them as realisations of Z and insert them into the family of distributions

   Likelihood: $\mathcal{L}(z|\theta) = p(z|\theta)$

4. Use the Bayes theorem to calculate the new belief about θ

   Posterior: $p(\theta|z) \propto \mathcal{L}(z|\theta)p(\theta)$

A Simple Example

A Simple Example

- T independent observations $\mathbf{y} = (y_1, y_2, ..., y_T)'$ drawn from a normal distribution with unknown mean μ and known variance σ^2

$$y_t = \mu + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

- Posterior for μ

$$\underbrace{p(\mu|\mathbf{y}, \sigma)}_{posterior} \propto \underbrace{p(\mu)}_{prior} \underbrace{p(\mathbf{y}|\mu, \sigma)}_{likelihood\ function}$$

- Likelihood

$$\begin{aligned} p(\mathbf{y}|\mu, \sigma) &= (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{T}(y_i - \mu)^2\right] \\ &= (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{T}{2\sigma^2}[s^2 + (\mu - \bar{\mu})^2]\right] \end{aligned}$$

where $\bar{\mu} = \frac{1}{T}\sum_{i=1}^{T} y_i$ and $s^2 = \frac{1}{T}\sum_{i=1}^{T}(y_i - \bar{\mu})^2$

A Simple Example: Non informative Prior

- **Which prior?** Let's assume we have no information on μ

$$p(\mu) \propto 1$$

  all values are equiprobable

- The prior is not well defined: improper prior

$$\int_{-\infty}^{\infty} p(\mu) d\mu = \infty$$

- However, by using an improper prior with a likelihood function, by use of Bayes' theorem the resulting posterior pdf is proper

A Simple Example: Non informative Prior

- Likhelihood $\Longrightarrow$ Posterior

$$\underbrace{p(\mu|\mathbf{y},\sigma)}_{\textit{posterior}} \propto \underbrace{p(\mathbf{y}|\mu,\sigma)}_{\textit{likelihood function}}$$

$$\begin{aligned} p(\mu|\mathbf{y},\sigma) &= (2\pi\sigma^2)^{-T/2}\exp\left[-\frac{T}{2\sigma^2}[s^2+(\mu-\bar{\mu})^2]\right] \\ &\propto \exp\left[-\frac{1}{2\left(\sigma^2/T\right)}(\mu-\bar{\mu})^2\right] \propto \mathcal{N}\left(\bar{\mu},\sigma^2/T\right) \end{aligned}$$

- Posterior normally distributed with mean $\bar{\mu}$ and variance σ^2/T

A Simple Example: Informative Prior

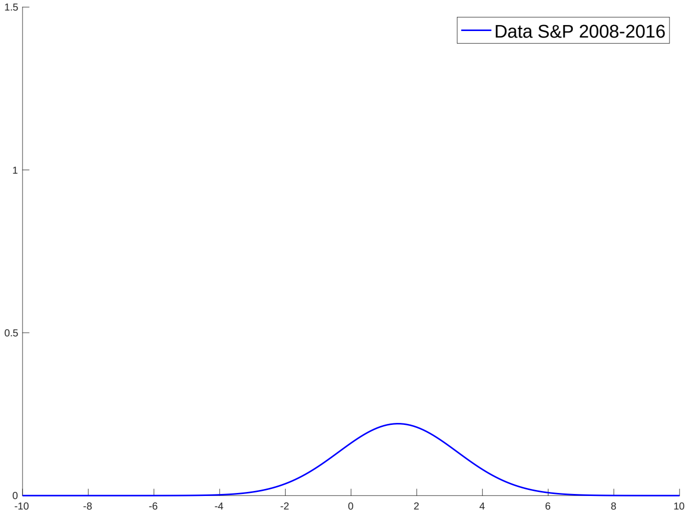
- **Which prior?** we can use as prior the posterior from a 'dummy sample'
- T_d independent (dummy) observations $\mathbf{y}_d$ (observed before sample), which we believe are drawn from the same distribution

$$p(\mu) = p(\mu|\mathbf{y}_d, \sigma) \sim \mathcal{N}\left(\bar{\mu}_d, \sigma^2 / T_d\right)$$

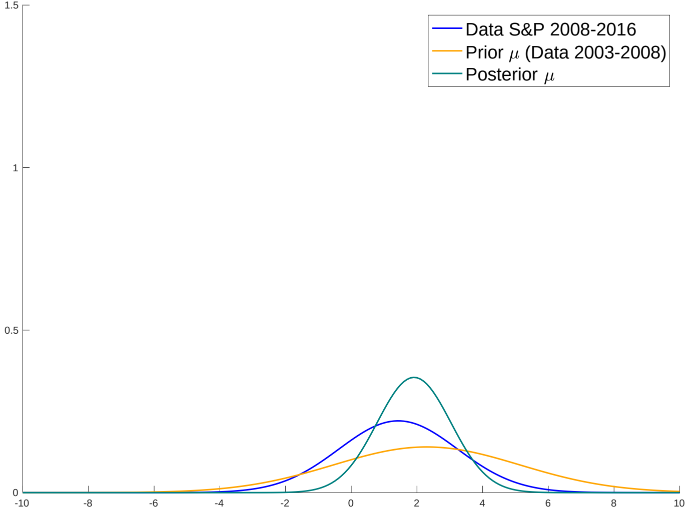
- The new posterior is:

$$\begin{aligned} p(\mu|\mathbf{y}, \sigma) \propto p(\mu)p(\mathbf{y}|\mu, \sigma) &= p(\mathbf{y}_d|\mu, \sigma)p(\mathbf{y}|\mu, \sigma) \\ \propto (2\pi\sigma^2)^{-(T_d+T)/2} \exp&\left[-\frac{1}{2\sigma^2}\left(T_d(\mu - \bar{\mu}_d)^2 + T(\mu - \bar{\mu})^2\right)\right] \\ &\sim \mathcal{N}\left(\frac{1}{T+T_d}(T_d\bar{\mu}_d + T\bar{\mu}), \frac{1}{(T+T_d)}\sigma^2\right) \end{aligned}$$

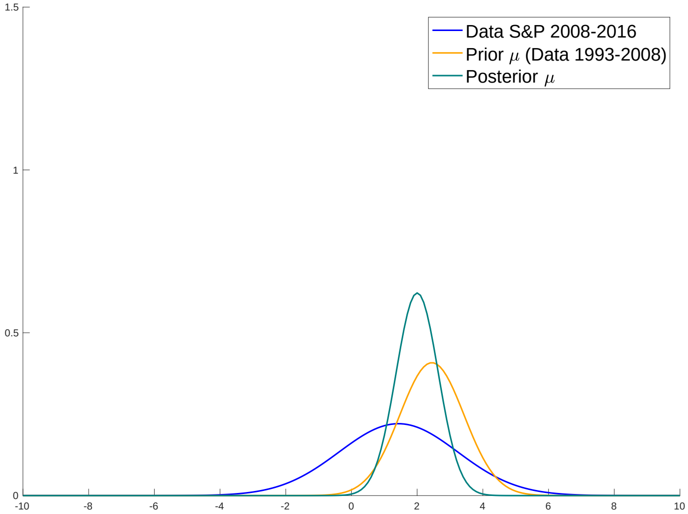
S&P 500 – $p(\mu|\mathbf{y}, \sigma)$



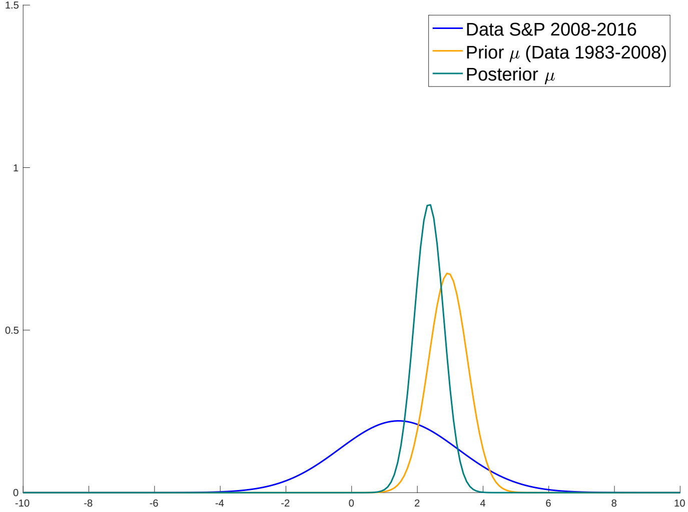
S&P 500 – $p(\mu|\mathbf{y}, \sigma)$



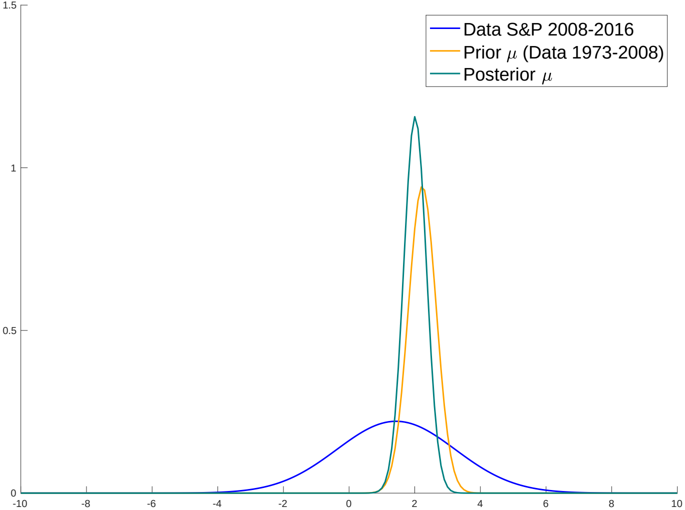
S&P 500 – $p(\mu|\mathbf{y}, \sigma)$



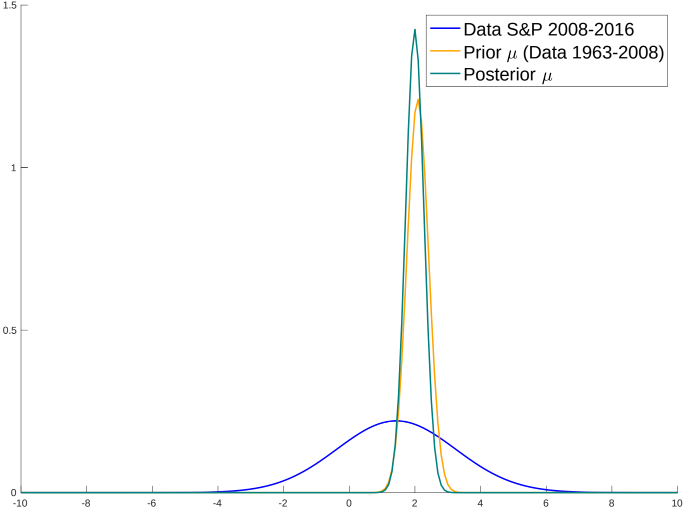
S&P 500 – $p(\mu|\mathbf{y}, \sigma)$



S&P 500 – $p(\mu|\mathbf{y},\sigma)$



S&P 500 – $p(\mu|\mathbf{y},\sigma)$



A Simple Example (revisited)

- T independent observations $\mathbf{y} = (y_1, y_2, ..., y_T)'$ drawn from a normal distribution with unknown mean μ and unknown variance σ^2

$$y_t = \mu + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

- **Diffuse/Non-Informative** priors

$$p(\mu, \sigma) = p(\mu)p(\sigma) \propto \frac{1}{\sigma}$$

$\Longrightarrow$ a priori μ, $\log \sigma$ uniformly distributed

A Simple Example (revisited)

- **Observation**: we consider $\theta = log \sigma$ with a flat prior pdf

$$p(\theta) \propto 1$$

- To obtain the prior distribution for σ we need to change the integration measure

$$dp(\theta) = dp(\sigma)\frac{d\theta}{d\sigma} = \frac{1}{\sigma}d\sigma$$

- Hence

$$p(\theta)d\theta \propto 1 \cdot \frac{1}{\sigma}d\sigma$$

A Simple Example (revisited)

- Joint posterior

$$\underbrace{p(\mu,\sigma|\mathbf{y})}_{\textit{posterior}} \propto \underbrace{p(\mu,\sigma)}_{\textit{prior}} \underbrace{p(\mathbf{y}|\mu,\sigma)}_{\textit{likelihood function}}$$

$$p(\mu,\sigma|\mathbf{y}) \quad \propto \quad \sigma^{-(T+1)} \exp\left[-\frac{T}{2\sigma^2}[s^2+(\mu-\bar{\mu})^2]\right]$$

A Simple Example (revisited)

- **Conditional** posterior for μ

$$p(\mu|\mathbf{y},\sigma) = \mathcal{N}\left(\bar{\mu}, \sigma^2/T\right)$$

- **Marginal** posterior for μ

$$\begin{aligned} p(\mu|\mathbf{y}) &= \int_0^\infty p(\mu,\sigma|\mathbf{y})d\sigma \\ &\propto \left[Ts^2 + T(\mu-\bar{\mu})^2\right]^{-T/2} \end{aligned}$$

Student t distribution

- **Marginal** posterior for σ

$$p(\sigma|\mathbf{y}) = \int_0^\infty p(\mu,\sigma|\mathbf{y})d\mu \propto \sigma^{-T}\exp\left(-\frac{Ts^2}{2\sigma^2}\right)$$

Inverse Gamma distribution

Gamma Distribution

Gamma Distribution

$$p(x|\alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}$$

for $\alpha > 0$ shape and $\beta > 0$ rate

Mean: $\frac{\alpha}{\beta}$

Mode: $\frac{\alpha-1}{\beta}$ for $\alpha \geq 1$, 0 for $\alpha < 1$

Variance: $\frac{\alpha}{\beta^2}$

Inverse Gamma Distribution

Inverse Gamma Distribution

$$p(y|\alpha,\beta) = \frac{\beta^{\alpha} y^{-\alpha-1}}{\Gamma(\alpha)} e^{-\beta/y}$$

it is obtained from the Gamma Distribution via the transformation

$$Y = g(X) = \frac{1}{X}$$

Mean: $\frac{\beta}{\alpha-1}$ for $\alpha > 1$

Mode: $\frac{\beta}{\alpha+1}$

Variance: $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$

A Simple Example (revisited)

- **Informative Priors:** the posterior obtained from a dummy sample $\mathbf{y}_d$ of size T_d

$$p(\mu, \sigma) = p(\mu|\sigma)p(\sigma)$$

$$p(\mu|\sigma) = p(\mu|\mathbf{y}_d, \sigma) \sim \mathcal{N}\left(\bar{\mu}_d, \sigma^2 / T_d\right)$$

$$p(\sigma) = p(\sigma|\mathbf{y}_d) \propto \sigma^{-T_d} \exp\left(-\frac{T_d s_d^2}{2\sigma^2}\right)$$

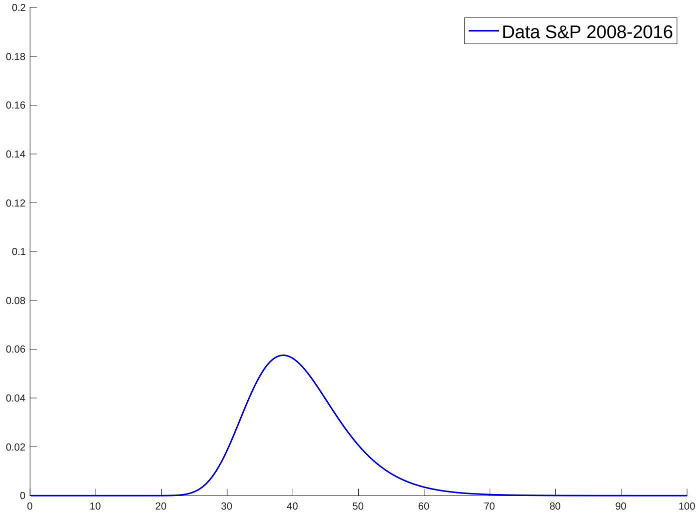
A Simple Example (revisited)

- The **Posterior** $p(\mu, \sigma|\mathbf{y}) = p(\mu|\sigma, \mathbf{y})p(\sigma|\mathbf{y})$ is given by:

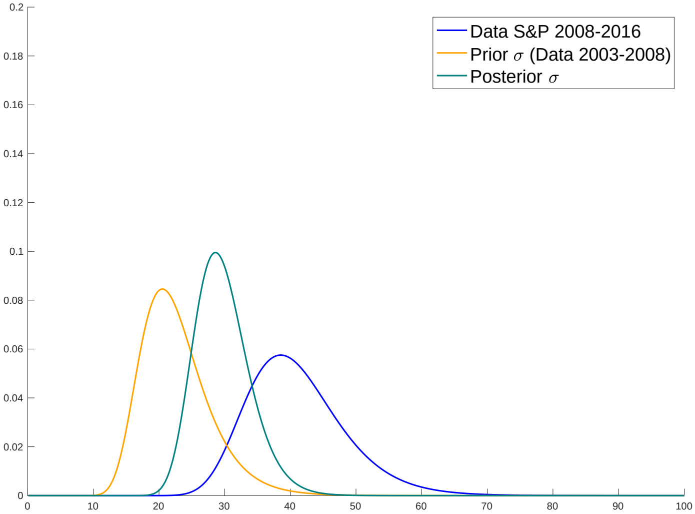
$$p(\mu|\mathbf{y}, \sigma) = \mathcal{N}\left(\frac{1}{T + T_d}(T_d\bar{\mu}_d + T\bar{\mu}), \frac{1}{(T + T_d)}\sigma^2\right)$$

$$p(\sigma|\mathbf{y}) \propto \sigma^{-(T+T_d)} \exp\left(-\frac{(T + T_d)s^2}{2\sigma^2}\right)$$

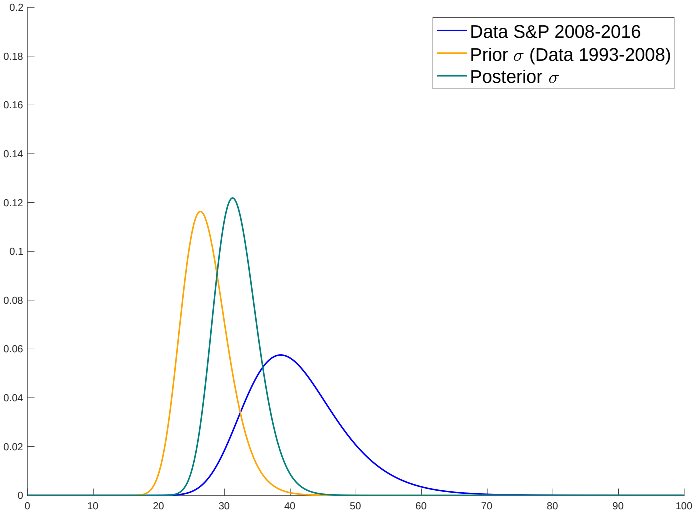
S&P 500 – $p(\sigma|\mathbf{y})$



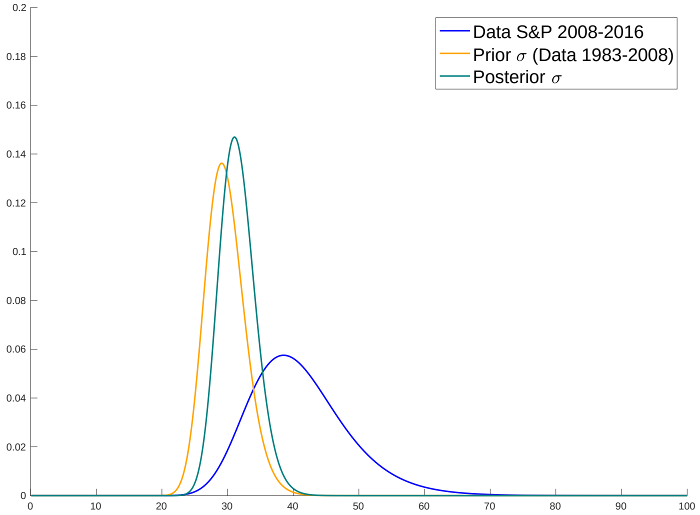
S&P 500 – $p(\sigma|\mathbf{y})$



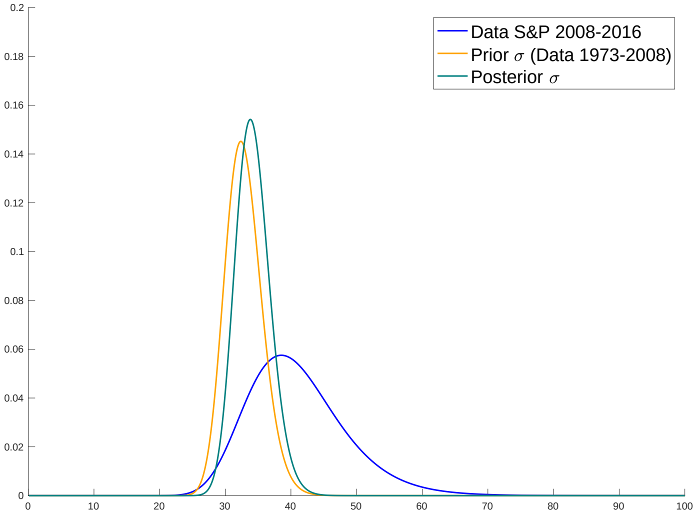
S&P 500 – $p(\sigma|\mathbf{y})$



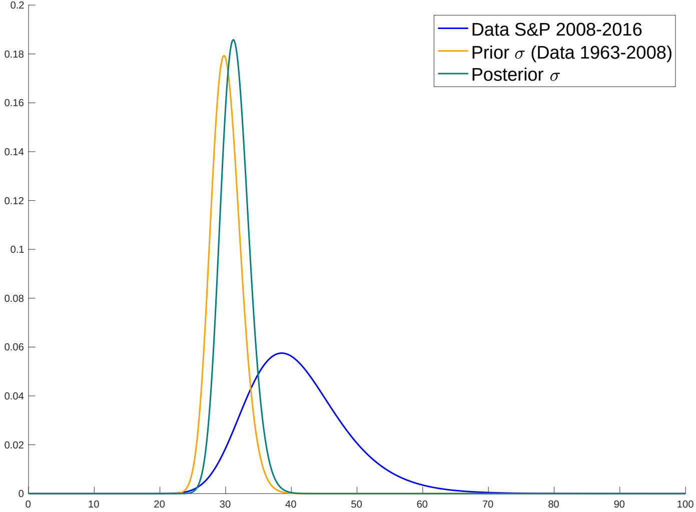
S&P 500 – $p(\sigma|\mathbf{y})$



S&P 500 – $p(\sigma|\mathbf{y})$



S&P 500 – $p(\sigma|\mathbf{y})$



Univariate Regression Model

Linear Regression: The Model

$$\underset{T\times 1}{Y} = \underset{T\times k}{X} \underset{k\times 1}{\beta} + \underset{T\times 1}{\epsilon} \qquad \epsilon \sim N(0, \sigma^2 I_T)$$

- The parameters of the models are:

$$\theta = (\beta', \sigma) \;\; and \;\; \Theta = \mathbb{R}^k \cup \mathbb{R}^+$$

- The probability of Y given the regressors X and the parameters θ is given by:

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

Linear Regression: The Model

We can rewrite as

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

where

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$s^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/\nu$$

and

$$\nu = T - k$$

Linear Regression: The Prior

- Suppose that σ^2 is known.
- We assume a non-informative prior on the regression coefficient: uniform distribution over the real line

$$p(\beta_i|\sigma^2) = 1 \quad -\infty < \beta_i < \infty \quad \forall i = 1, ..., k$$

$$p(\beta|\sigma^2) \propto 1$$

- **Remark**: The prior is improper

$$\int_{-\infty}^{\infty} p(\mu) d\mu = \infty \neq 1$$

Linear Regression: The Posterior

- Collect the data y, x (realisation of Y, X). Use the data to evaluate the likelihood

$$\mathcal{L}(y|x,\beta,\sigma) = p(y|x,\beta,\sigma)$$

- ... and update the belief from the Bayes rule

$$\begin{aligned} p(\beta \mid y, x, \sigma^2) &\propto \mathcal{L}(y|x,\beta,\sigma)p(\beta,\sigma) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{T/2} exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\beta-\hat{\beta})'x'x(\beta-\hat{\beta})]\right\} \end{aligned}$$

Linear Regression: The Posterior

- Up to a scale

$$p(\beta|y,x,\sigma^2) \propto \left|\frac{x'x}{\sigma^2}\right|^{1/2} exp\left\{-\frac{1}{2}\left[(\beta-\hat{\beta})'\frac{x'x}{\sigma^2}(\beta-\hat{\beta})\right]\right\}$$

- Hence we get

$$\beta|\sigma^2 \sim \mathcal{N}(\hat{\beta},(x'x)^{-1}\sigma^2)$$

- **Remark**: Same result as with Maximum likelihood

The Case of Unknown σ^2

- Let's consider now the case for unknown σ^2 and add a prior:

$$p(\log \sigma^2) \propto 1 \quad \Rightarrow \quad p(\sigma^2) \propto \frac{1}{\sigma^2}$$

- The Posterior Likelihood is

$$\begin{aligned}
p(\beta, \sigma^2 \mid y, x) \;&\propto\; \mathcal{L}(y|x, \beta, \sigma)p(\beta, \sigma) \\
&\propto\; \underbrace{\left|\frac{x'x}{\sigma^2}\right|^{1/2} exp\left\{-\frac{1}{2\sigma^2}(\beta - \hat{\beta})'x'x(\beta - \hat{\beta})\right\}}_{p(\beta|y,x,\sigma^2)} \\
&\quad \times \underbrace{\left(\frac{1}{\sigma^2}\right)^{\nu/2+1} exp\left\{-\frac{1}{2\sigma^2}\nu s^2\right\}}_{p(\sigma^2|x,y)}
\end{aligned}$$

Bayesian Regression with Improper Prior

The Model:

$$\underset{T\times 1}{Y} = \underset{T\times k}{X}\ \underset{k\times 1}{\beta} + \underset{T\times 1}{\epsilon} \qquad \epsilon \sim N(0, \sigma^2 I_T)$$

The Priors:

$$p(\beta|\sigma^2) \propto 1$$
$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

The Posterior:

$$\sigma^2|x, y \sim \mathcal{IG}(\nu s^2, \nu)$$
$$\beta|x, y, \sigma^2 \sim \mathcal{N}(\hat{\beta}, (x'x)^{-1}\sigma^2)$$

Single Regression Models: Deriving $p(\sigma^2|\beta, x, y)$

Question: What is the posterior of σ^2 given β, i.e. $p(\sigma^2|\beta, y, x)$?

The Posterior:

$$
\begin{aligned}
p(\beta, \sigma^2|x, y) &\propto p(Y|X, \beta, \sigma^2)p(\beta|\sigma^2)p(\sigma^2) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{T/2} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \times \left(\frac{1}{\sigma^2}\right) \\
&= \left(\frac{1}{\sigma^2}\right)^{T/2+1} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}
\end{aligned}
$$

Single Regression Models: Deriving $p(\sigma^2|\beta, x, y)$

Hence one gets the posterior distribution:

$$p(\sigma^2|\beta, y, x) = \mathcal{IG}(\nu \tilde{s}_\beta^2, \nu)$$

where

$$\tilde{s}_\beta^2 = \frac{(Y - X\beta)'(Y - X\beta)}{T}$$

and

$$\nu = T$$

The Marginal Posterior Distribution of β

Problem: We found the conditional distribution of β given σ^2, and

$$p(\beta|x, y, \sigma^2)$$

the conditional distribution of σ^2 given β

$$p(\sigma^2|x, y, \beta).$$

How do we get the marginal posterior distribution of β

$$p(\beta|x, y)$$

and the joint posterior distribution

$$p(\beta, \sigma^2|x, y)?$$

The Marginal Posterior Distribution of β

- **Gibbs Sampler:**
  1. Generate $(\sigma^2)^{(j)}$ by drawing from $p(\sigma^2|x, y, \beta^{(j-1)})$
  2. Generate $\beta^{(j)}$ a drawing from $p\left(\beta|x, y, (\sigma^2)^{(j)}\right)$

  Starting from arbitrary $\beta^{(0)}$, repeating step (1) and (2) to obtain $\beta^{(j)}, \sigma^{(j)}, j = 1, ..., J$

- For large J we obtain independent draws the joint distribution $p(\beta, \sigma^2|x, y)$. Common practice is to discard the initial draws.

- Approximate moments and quantiles from the empirical distribution of the generated parameters.

- **Remark:** The algorithm can be used to approximate the distribution of any function of the parameters

Bayesian Regression with Informative Priors

- Use the **posterior from a dummy sample** y_d, x_d of length T_d as a prior for the sample y, x of length T generated by the same model
- The **posterior** from the sample y_d, x_d using the **uninformative prior** is,

$$p(\sigma^2) = p(\sigma^2|x_d, y_d) = \mathcal{IG}(\nu_d s_d^2, \nu_d)$$

$$p(\beta|\sigma^2) = p(\beta|x_d, y_d, \sigma^2) = \mathcal{N}(\hat{\beta}_d, (x_d' x_d)^{-1}\sigma^2)$$

where

$$\nu_d = T_d - k$$

$$\hat{\beta}_d = (x_d' x_d)^{-1}(x_d' y_d)$$

and

$$s_d^2 = (y_d - x_d\hat{\beta}_d)'(y_d - x_d\hat{\beta}_d)/\nu_d$$

Informative Priors

Combining prior and likelihood

$$
\begin{aligned}
\mathcal{L}(\beta,\sigma|y,x,y_d,x_d) &\propto \mathcal{L}(y|x,\beta,\sigma)p(\beta|x_d,y_d,\sigma^2)p(\sigma^2|x_d,y_d) \\
&\propto \underbrace{\left(\frac{1}{\sigma^2}\right)^{T/2} exp\left\{-\frac{1}{2\sigma^2}(y-x\beta)'(y-x\beta)\right\}}_{\mathcal{L}(y|x,\beta,\sigma)} \\
&\quad \underbrace{\times\left(\frac{1}{\sigma^2}\right)^{k/2} exp\left\{-\frac{1}{2\sigma^2}(\beta-\hat{\beta}_d)'x_d'x_d(\beta-\hat{\beta}_d)\right\}}_{p(\beta|x_d,y_d,\sigma^2)} \\
&\quad \underbrace{\times\left(\frac{1}{\sigma^2}\right)^{\nu_d/2+1} exp\left\{-\frac{1}{2\sigma^2}\nu_d s_d^2\right\}}_{p(\sigma^2|x_d,y_d)} \\
&\propto \left(\tfrac{1}{\sigma^2}\right)^{(T^*+1)/2} exp\left\{-\tfrac{1}{2\sigma^2}[(y^*-x^*\beta)'(y^*-x^*\beta)]\right\}
\end{aligned}
$$

Augmented data: $y^* = (y_d' \ , \ y')' \qquad x^* = (x_d' \ , \ x')' \qquad T^* = T + T_d$

Informative Priors

Hence

$$\beta|\sigma, x, y \sim \mathcal{N}(\hat{\beta}, (x^{*\prime}x^*)^{-1}\sigma^2)$$

$$\sigma^2|x, y \sim \mathcal{IG}(\nu s^2, \nu)$$

where

$$\hat{\beta} = (x^{*\prime}x^*)^{-1}x^{*\prime}y^*$$

$$s^2 = (y^* - x^*\hat{\beta})'(y^* - x^*\hat{\beta})$$

and

$$\nu = T^* - k$$

Remark: $(y^* - x^*\beta)'(y^* - x^*\beta) = \nu s^2 + (\beta - \hat{\beta})'x^{*\prime}x^*(\beta - \hat{\beta})$

Informative Priors

$$\beta|\sigma, x, y \sim \mathcal{N}(\hat{\beta}, (x'x + x_d'x_d)^{-1}\sigma^2)$$
$$\sigma^2|x, y \sim \mathcal{IG}(\nu s^2, \nu)$$

where

$$\hat{\beta} = (x^{*\prime}x^*)^{-1}x^{*\prime}y^* = (x'x + x_d'x_d)^{-1}(x_d'y_d + x'y)$$
$$s^2 = (y - x\hat{\beta})'(y - x\hat{\beta}) + (y_d - x_d\hat{\beta})'(y_d - x_d\hat{\beta})$$
$$\nu = T + T_d - k$$

Remark: If we had pooled the two samples and used a diffuse prior, the resulting posterior would had been exactly the same.

Conjugate Priors

- Natural conjugate priors are the priors such that the posteriors distributions are the same distributional family of the priors
- In the Normal regression model the natural conjugate prior is the Normal-inverted Gamma (Wishart) prior
- It has the desirable property that the prior can be generated by a sample generated by the same model

General NIG(W) priors:

$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, V_0\sigma^2)$$

$$\sigma^2 \sim \mathcal{IG}(\nu_0\sigma_0^2, \nu_0)$$

Conjugate Prior

The Normal-inverted Wishart prior can be implemented by using 'artificial' dummy observations

Idea: Need to find x_d and y_d such that:

$$(x_d' x_d)^{-1} x_d' y_d = \beta_0$$

and

$$(x_d' x_d)^{-1} = V_0$$

$$(y_d - x_d \beta_0)'(y_d - x_d \beta_0) = \nu_0 \sigma_0^2$$

and

$$T_d - k = \nu_0$$

Conjugate Prior

The posterior is:

$$\beta|x, y, \sigma^2 \sim \mathcal{N}(\beta_1, V_1\sigma^2)$$

where

$$\beta_1 = (x'x + V_0^{-1})^{-1}(V_0^{-1}\beta_0 + x'y)$$

and

$$V_1 = (x'x + V_0^{-1})^{-1}$$

Conjugate Prior

Example: Simple prior:

$$\beta|\sigma^2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{\tau^2} I_k\right)$$

for σ^2 given

- Can be implemented by using additional dummy observations.
- Need to find x_d and y_d such that:

$$(x_d' x_d)^{-1} x_d' y_d = 0 \qquad \text{and} \qquad x_d' x_d = \tau^2 I_k$$

- Set:

$$x_d = \tau I_k \qquad y_d = 0_{k \times 1}$$

Conjugate Prior

The posterior is:

$$\beta|x, y, \sigma^2 \sim \mathcal{N}(\beta_1, V_1\sigma^2)$$

where

$$\beta_1 = (x'x + \tau^2 I_k)^{-1} x'y$$

and

$$V_1 = (x'x + \tau^2 I_k)^{-1}$$

- τ is a tightness parameter, controls the weight we give to the prior.
- $\tau \to \infty \implies$ posterior $=$ prior ('dogmatic')
- $\tau \to 0 \implies$ OLS ('uninformative')

Multivariate Regression Model

The Wishart Distribution

Let

$$Z_t = (Z_{1,t}, ..., Z_{m,t}) \sim i.i.d. \mathcal{N}(0, H)$$

Define

$$S = \sum_{t=1}^{\nu} Z_t Z_t'$$

S/ν is the sample covariance matrix of Z_t based on a sample of size ν.

S has a Wishart distribution, with scale H and ν degrees of freedom

$$S \sim \mathcal{W}(H, \nu)$$

The Inverse Wishart Distribution

Σ has Inverted Wishart distribution with scale Ψ and ν degrees of freedom

$$\Sigma \sim \mathcal{IW}(\Psi, \nu)$$

if

$$\Sigma^{-1} \sim \mathcal{W}(\Psi^{-1}, \nu)$$

$$p(\Sigma) \propto |\Psi|^{\nu/2} |\Sigma|^{-(\nu+m+1)/2} exp \left\{ -\frac{1}{2} tr \left[\Sigma^{-1} \Psi \right] \right\}$$

Remark: In the univariate case ($m = 1$) we have the Gamma distribution and the Inverted Gamma distribution (χ^2 is in the class)

Multivariate Regression

The Model:

$$\underset{T\times m}{Y} = \underset{T\times k}{X} \underset{k\times m}{\beta} + \underset{T\times m}{\epsilon} \qquad vec(\epsilon) \sim N(0, \Sigma \otimes I_T)$$

The Prior:

$$p(\beta|\Sigma) \propto 1$$

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$$

Multivariate Regression

Result 1: The posterior of β given Σ has the form:

$$vec(\beta)|x, y, \Sigma \sim \mathcal{N}(vec(\hat{\beta}), \Sigma \otimes (x'x)^{-1})$$

Result 2: The posterior of Σ has the form

$$\Sigma|x, y \sim \mathcal{IW}(\nu S, \nu) \qquad \nu = T - k$$

Result 3: The posterior of Σ given β

$$\Sigma|\beta, x, y \sim \mathcal{IW}(T\tilde{S}_{\beta}, T)$$

Multivariate Regression

For any matrix A write $vecA = \vec{A}$

The model:

$$\underset{T\times m}{Y} = \underset{T\times k}{X} \underset{k\times m}{\beta} + \underset{T\times m}{\epsilon}$$

can be rewritten as

$$\underset{Tm\times 1}{\vec{Y}} = \underset{Tm\times mk}{(I_m \otimes X)} \underset{mk\times 1}{\vec{\beta}} + \underset{Tm\times 1}{\vec{\epsilon}} \qquad \vec{\epsilon} \sim N(0, \Sigma \otimes I_T)$$

Multivariate Regression

The Likelihood:

$$\mathcal{L}(y|x,\beta,\Sigma) = p(y|x,\beta,\Sigma) \propto$$

$$\propto \underbrace{|\Sigma \otimes I_T|^{-1/2} exp\left\{-\frac{1}{2}(\vec{y} - (I_m \otimes x)\vec{\beta})'(\Sigma \otimes I_T)^{-1}(\vec{y} - (I_m \otimes x)\vec{\beta})\right\}}_{①}$$

$$\propto \underbrace{|\Sigma|^{-(T-k)/2} exp\left\{-\frac{1}{2}(\vec{y} - (I_m \otimes x)\widehat{\vec{\beta}})'(\Sigma \otimes I_T)^{-1}(\vec{y} - (I_m \otimes x)\widehat{\vec{\beta}})\right\}}_{②} \times$$

$$\underbrace{\left|\Sigma \otimes (x'x)^{-1}\right|^{-1/2} \; exp\left\{-\frac{1}{2}(\widehat{\vec{\beta}} - \vec{\beta})' \left(\Sigma \otimes (x'x)^{-1}\right)^{-1} (\widehat{\vec{\beta}} - \vec{\beta})\right\}}_{③}$$

Multivariate Regression

where

$$\widehat{\vec{\beta}} = \left((I_m \otimes x)'(\Sigma \otimes I_T)^{-1}(I_m \otimes x)\right)^{-1}(I_m \otimes x)'(\Sigma \otimes I_T)^{-1}\vec{y} = vec\hat{\beta}$$

$$\hat{\beta} = (x'x)^{-1}x'y$$

We observe that

$$① = |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}tr\Sigma^{-1}\tilde{S}_\beta T\right\}$$

where

$$\tilde{S}_\beta = (y - x\beta)'(y - x\beta)/T$$

$$② = |\Sigma|^{-(T-k)/2} \; exp\left\{-\frac{1}{2}tr\Sigma^{-1}S(T-k)\right\}$$

where

$$S = (y - x\hat{\beta})'(y - x\hat{\beta})/(T-k)$$

Multivariate Regression

Using the priors:

$$p(\beta|\Sigma) \propto 1$$

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$$

we obtain the posteriors:

from ③: $vec(\beta)|x, y, \Sigma \sim \mathcal{N}(vec(\hat{\beta}), \Sigma \otimes (x'x)^{-1})$

from ②: $\Sigma|x, y \sim \mathcal{IW}(\nu S, \nu) \qquad \nu = T - k$

from ①: $\Sigma|\beta, x, y \sim \mathcal{IW}(T\tilde{S}_\beta, T)$

Bayesian Estimation of a VAR

VAR Model:

$$y_t = c + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \quad u_t \sim \mathcal{N}(0, \Sigma)$$

More compactly:

$$A(L) y_t = c + u_t \quad u_t \sim \mathcal{N}(0, \Sigma)$$

This is just a multivariate regression