Bayesian Inference

- Ingredients of Bayesian Analysis:
  - Likelihood function \( p(Y|\phi) \)
  - Prior density \( p(\phi) \)
  - Marginal data density \( p(Y) = \int p(Y|\phi)p(\phi)d\phi \)

- Bayes Theorem:
  \[
p(\phi|Y) = \frac{p(Y|\phi)p(\phi)}{p(Y)}
  \]
• Consider AR(1) model:

\[ y_t = y_{t-1}\phi + u_t, \quad u_t \sim iidN(0,1). \]

• Let \( x_t = y_{t-1} \). Write as

\[ y_t = x_t'\phi + u_t, \quad u_t \sim iidN(0,1), \]

or

\[ Y = X\phi + U. \]

We can easily allow for multiple regressors. Assume \( \phi \) is \( k \times 1 \).

• Notice: we treat the variance of the errors as know. The generalization to unknown variance is straightforward but tedious.

• Likelihood function:

\[ p(Y|\phi) = (2\pi)^{-T/2} \exp \left\{ -\frac{1}{2} (Y - X\phi)'(Y - X\phi) \right\}. \]
A Convenient Prior

- Prior:

  \[ \phi \sim N \left( 0_{k \times 1}, \tau^2 I_{k \times k} \right), \quad p(\phi) = (2\pi \tau^2)^{-k/2} \exp \left\{ -\frac{1}{2\tau^2} \phi' \phi \right\} \]

- Large \( \tau \) means diffuse prior.
- Small \( \tau \) means tight prior.
Deriving the Posterior

- **Bayes Theorem:**
  \[
p(\phi|Y) \propto p(Y|\phi)p(\phi)
  \]
  \[
  \propto \exp\left\{-\frac{1}{2}[(Y - X\phi)'(Y - X\phi) + \tau^{-2}\phi'\phi]\right\}.
  \]

- **Guess:** what if \( \phi|Y \sim N(\bar{\phi}_T, \bar{V}_T) \). Then
  \[
p(\theta|Y) \propto \exp\left\{-\frac{1}{2}(\phi - \bar{\phi}_T)'\bar{V}_T^{-1}(\phi - \bar{\phi}_T)\right\}.
  \]

- **Rewrite exponential term**
  \[
  Y'Y - \phi'X'Y - Y'X\phi + \phi'X'X\phi + \tau^{-2}\phi'\phi
  \]
  \[
  = Y'Y - \phi'X'Y - Y'X\phi + \phi'(X'X + \tau^{-2}I)\phi
  \]
  \[
  = \left(\phi - (X'X + \tau^{-2}I)^{-1}X'Y\right)'\left(X'X + \tau^{-2}I\right)
  \]
  \[
  \times \left(\phi - (X'X + \tau^{-2}I)^{-1}X'Y\right)
  \]
  \[
  + Y'Y - Y'X(X'X + \tau^{-2}I)^{-1}X'Y.
  \]
Deriving the Posterior

- Exponential term is a quadratic function of $\phi$.
- Deduce: posterior distribution of $\phi$ must be a multivariate normal distribution

$$
\phi | Y \sim N(\phi_T, \tilde{V}_T)
$$

with

$$
\phi_T = (X'X + \tau^{-2}I)^{-1}X'Y \\
\tilde{V}_T = (X'X + \tau^{-2}I)^{-1}.
$$

- $\tau \rightarrow \infty$:

$$
\phi | Y \approx N(\hat{\phi}_{mle}, (X'X)^{-1}).
$$

- $\tau \rightarrow 0$:

$$
\phi | Y \approx \text{Pointmass at 0}
$$
Marginal Data Density

- Plays an important role in Bayesian model selection and averaging.
- Write

\[
p(Y) = \frac{p(Y|\theta)p(\theta)}{p(\theta|Y)}
\]

\[
= \exp \left\{ -\frac{1}{2} [Y'Y - Y'X(X'X + \tau^{-2}I)^{-1}X'Y] \right\}
\times (2\pi)^{-T/2} |I + \tau^2X'X|^{-1/2}.
\]

- The exponential term measures the goodness-of-fit.
- \(|I + \tau^2X'X|\) is a penalty for model complexity.
• We will often abbreviate posterior distributions $p(\phi|Y)$ by $\pi(\phi)$ and posterior expectations of $h(\phi)$ by

$$
\mathbb{E}_\pi[h] = \mathbb{E}_\pi[h(\phi)] = \int h(\phi)\pi(\phi)d\phi = \int h(\phi)p(\phi|Y)d\phi.
$$

• We will focus on algorithms that generate draws $\{\phi^i\}_{i=1}^N$ from posterior distributions of parameters in time series models.

• These draws can then be transformed into objects of interest, $h(\phi^i)$, and under suitable conditions a Monte Carlo average of the form

$$
\bar{h}_N = \frac{1}{N} \sum_{i=1}^N h(\phi^i) \approx \mathbb{E}_\pi[h].
$$

• Strong law of large numbers (SLLN), central limit theorem (CLT)…
• In the simple linear regression model with Gaussian posterior it is possible to sample directly.

• For \( i = 1 \) to \( N \), draw \( \phi^i \) from \( N(\bar{\phi}, \bar{V}_{\phi}) \).

• Provided that \( \nabla_{\pi}[h(\phi)] < \infty \) we can deduce from Kolmogorov’s SLLN and the Lindeberg-Levy CLT that

\[
\bar{h}_N \overset{a.s.}{\to} \mathbb{E}_{\pi}[h] \\
\sqrt{N} \left( \bar{h}_N - \mathbb{E}_{\pi}[h] \right) \implies \mathcal{N}(0, \nabla_{\pi}[h(\phi)]).
\]
• The posterior expected loss associated with a decision $\delta(\cdot)$ is given by

$$\rho(\delta(\cdot)|Y) = \int_{\Theta} L(\theta, \delta(Y)) p(\theta|Y) d\theta.$$ 

• A Bayes decision is a decision that minimizes the posterior expected loss:

$$\delta^*(Y) = \arg\min_d \rho(\delta(\cdot)|Y).$$

• Since in most applications it is not feasible to derive the posterior expected risk analytically, we replace $\rho(\delta(\cdot)|Y)$ by a Monte Carlo approximation of the form

$$\bar{\rho}_N(\delta(\cdot)|Y) = \frac{1}{N} \sum_{i=1}^{N} L(\theta^i, \delta(\cdot)).$$

• A numerical approximation to the Bayes decision $\delta^*(\cdot)$ is then given by

$$\delta^*_N(Y) = \arg\min_d \bar{\rho}_N(\delta(\cdot)|Y).$$
Inference

- Point estimation:
  - Quadratic loss: posterior mean
  - Absolute error loss: posterior median

- Interval/Set estimation $\mathbb{P}_\pi \{ \theta \in C(Y) \} = 1 - \alpha$:
  - highest posterior density sets
  - equal-tail-probability intervals
• Example:

\[ y_{T+h} = \theta^h y_T + \sum_{s=0}^{h-1} \theta^s u_{T+h-s} \]

• \( h \)-step ahead conditional distribution:

\[ y_{T+h}\mid (Y_{1:T}, \theta) \sim N\left(\theta^h y_T, \frac{1 - \theta^h}{1 - \theta}\right). \]

• Posterior predictive distribution:

\[ p(y_{T+h}\mid Y_{1:T}) = \int p(y_{T+h}\mid y_T, \theta)p(\theta\mid Y_{1:T})d\theta. \]

• For each draw \( \theta^i \) from the posterior distribution \( p(\theta\mid Y_{1:T}) \) sample a sequence of innovations \( u_{T+1}^i, \ldots, u_{T+h}^i \) and compute \( y_{T+h}^i \) as a function of \( \theta^i, u_{T+1}^i, \ldots, u_{T+h}^i, \) and \( Y_{1:T}. \)
Model Uncertainty

- Assign prior probabilities $\gamma_{j,0}$ to models $M_j$, $j = 1, \ldots, J$.
- Posterior model probabilities are given by

$$\gamma_{j,T} = \frac{\gamma_{j,0} p(Y|M_j)}{\sum_{j=1}^{J} \gamma_{j,0} p(Y|M_j)},$$

where

$$p(Y|M_j) = \int p(Y|\theta(j), M_j) p(\theta(j)|M_j) d\theta(j).$$

- Log marginal data densities are one-step-ahead predictive scores:

$$\ln p(Y|M_j) = \sum_{t=1}^{T} \ln \int p(y_t|\theta(j), Y_{1:t-1}, M_j) p(\theta(j)|Y_{1:t-1}, M_j) d\theta(j).$$

- Model averaging:

$$p(h|Y) = \sum_{j=1}^{J} \gamma_{j,T} p(h_j(\theta(j))|Y, M_j).$$