Bayesian Econometrics
Tutorial 01 - Introduction

Tutor: Richard Schnorrenberger

richard.schn@stat-econ.uni-kiel.de

Institute for Statistics and Econometrics
Kiel University

Winter Term 2024/25

Review the Concepts and Proofs

- 1. How can Bayes' rule be used to learn about unknown parameters?
- 2. What is a prior density, a posterior density, and a likelihood function?
- 3. Why is a natural conjugate prior helpful?
- 4. How can we assess the influence of the prior on the posterior results?
- 5. How can we compare different models using the Bayesian approach?
- 6. What is the marginal likelihood of a model M_i?
- 7. What is the prior odds ratio, the posterior odds ratio, and the Bayes factor?
- 8. What does a predictive density tell us?

A Change of Perspective

From a frequentist perspective ...

- the sample information $\mathbf{w} = (\mathbf{y}, \mathbf{x})$ is random and comes from the joint distribution with PDF $f(\mathbf{w}; \boldsymbol{\theta})$;
- the unknown true parameter $\boldsymbol{\theta}$ is treated as a fixed constant;
- randomness arises from the sampling distribution of the estimator $\hat{\theta}$. Hence, $\hat{\theta}$ is a random variable with sampling distribution in repeated samples.

In Bayesian econometrics ...

- the unknown parameter $\boldsymbol{\theta}$ itself is regarded as a random variable with a posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ reflecting all the information at hand;
- the observed data $\mathbf{y}$ is treated as fixed;
- the *prior belief* or *knowledge* an individual accumulated about $\boldsymbol{\theta}$ is formally incorporated into the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$.

Exercise 1

Decision theory (Exercise 1 of Koop's textbook). **In a formal decision theoretic context, the choice of a point estimator of θ is made by defining a loss function and choosing the point estimator which minimizes the expected loss. Thus, if $C(\tilde{\theta}, \theta)$ is the loss associated with choosing $\tilde{\theta}$ as a point estimator of θ, then we would choose that $\tilde{\theta}$ which minimizes $\mathrm{E}[C(\tilde{\theta}, \theta)|y]$, where the expectation is taken with respect to the posterior of θ. For the case where θ is a scalar, show the following:**

- (a) ***Squared error loss*. If $C(\tilde{\theta}, \theta) = (\tilde{\theta} - \theta)^2$ then $\tilde{\theta} = \mathrm{E}(\theta|y)$.**

Remarks:

- The loss function $C(\tilde{\theta}, \theta)$ is a non-decreasing function of the sampling error $(\tilde{\theta} - \theta)$.
- The *Bayes point estimate* $\tilde{\theta} = \tilde{\theta}(y)$ of the unknown parameter θ is obtained by

$$\min_{\tilde{\theta}} \mathrm{E}_{\theta|y}[C(\tilde{\theta}, \theta)] = \min_{\tilde{\theta}} \int_{-\infty}^{\infty} C(\tilde{\theta}, \theta) p(\theta|y) d\theta$$

Exercise 1

- (b) ***Asymmetric linear loss*. If**

$$C(\tilde{\theta}, \theta) = \begin{cases} c_1 |\tilde{\theta} - \theta| & \textbf{if } \tilde{\theta} \leq \theta \\ c_2 |\tilde{\theta} - \theta| & \textbf{if } \tilde{\theta} > \theta \end{cases}$$

where $c_1 > 0$ and $c_2 > 0$ are constants, then $\tilde{\theta}$ is the $\frac{c_1}{c_1 + c_2}$th quantile of $p(\theta|y)$. Recall Leibniz' general rule for differentiation of an integral:

$$\frac{\partial}{\partial \tilde{\theta}} \int_{g(\tilde{\theta})}^{h(\tilde{\theta})} f(\theta, \tilde{\theta}) d\theta = \int_{g(\tilde{\theta})}^{h(\tilde{\theta})} \frac{\partial f}{\partial \tilde{\theta}} d\theta + f(h(\tilde{\theta}), \tilde{\theta}) \frac{\partial h}{\partial \tilde{\theta}} - f(g(\tilde{\theta}), \tilde{\theta}) \frac{\partial g}{\partial \tilde{\theta}}.$$

- (c) ***All-or-nothing loss*. If**

$$C(\tilde{\theta}, \theta) = \begin{cases} c & \textbf{if } \tilde{\theta} \neq \theta \\ 0 & \textbf{if } \tilde{\theta} = \theta \end{cases}$$

where $c > 0$ is a constant, then $\tilde{\theta}$ is the mode of $p(\theta|y)$.

Exercise 2

Let $y = (y_1, \ldots, y_N)'$ **be a random sample with** y_i **drawn from a Gamma distribution with parameters** $1/\theta$ **and** 2 **and density** $p(y_i|\theta) = f_G(y_i|\theta^{-1}, 2)$**. (As you may see, this is equal to an exponential distribution.) Assume a Gamma prior for** θ**,** $p(\theta) = f_G(\theta|\underline{\mu}, 2\underline{\nu})$**, where** $\underline{\mu}$ **and** $\underline{\nu}$ **are prior hyperparameters.**

Note that the gamma density is given by:

$$f_G(y|\mu, \nu) = c_G^{-1}\, y^{\frac{\nu-2}{2}} \exp(-\frac{y\nu}{2\mu}),$$

where $c_G^{-1} = (\frac{\nu}{2\mu})^{\frac{\nu}{2}}\, \Gamma(\frac{\nu}{2})^{-1}$ and $0 < y < \infty$ (see Koop's Appendix B.2).

- (a) **Derive** $p(\theta|y)$ **and** $\mathrm{E}(\theta|y)$**.**
- (b) **What happens to** $\mathrm{E}(\theta|y)$ **as** $\underline{\nu} \to 0$**? In what sense is such a prior noninformative?**

Exercise 3

Let $y = (y_1, \ldots, y_N)'$ be a Bernoulli random sample where

$$p(y_i|\theta) = \begin{cases} \theta^{y_i}(1-\theta)^{1-y_i} & \textbf{if } y_i = 0 \textbf{ or } 1 \\ 0 & \textbf{otherwise} \end{cases}$$

- (a) **Derive the posterior for θ assuming a uniform prior, $\theta \sim U(0,1)$. Find $\mathrm{E}(\theta|y)$. What happens if the sample size increases?**
- (b) **Repeat part (a) assuming a Beta prior of the form**

$$p(\theta) = \begin{cases} B(\underline{\alpha}, \underline{\beta})^{-1}\theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\beta}-1} & \textbf{if } 0 \leq \theta \leq 1 \\ 0 & \textbf{otherwise} \end{cases}$$

where $B(\underline{\alpha}, \underline{\beta})$ is the beta function.