Bayesian Econometrics Primer
&
Estimation of linearized DSGE models

Stéphane Adjemian

stephane.adjemian@univ-lemans.fr

June, 2019

Introduction

- Full information Bayesian approach to linear(ized) DSGE models.
- Basically, this approach allows to incorporate prior knowledge about the model and its parameters in the inference procedure.
- We first present the Bayesian approach in the case of a simple linear model for which closed form expressions are available.
- And we discuss issues specific to the estimation of DSGE models.

Outline

ML Estimation

- A model ($\mathcal{M}$) defines a joint probability distribution parameterized (by $\theta_{\mathcal{M}}$) function over a sample of variables (say $\mathcal{Y}_T$):

$$f(\mathcal{Y}_T|\theta_{\mathcal{M}}, \mathcal{M}) \tag{1}$$

- The parameters $\theta_{\mathcal{M}}$ can be estimated by confronting the model to the data through:
  - Some moments of the DGP.
  - The probability density function of the DGP (all the moments).
- The first approach is a method of moments, the second one is a likelihood approach.
- Basically, a ML estimate for $\theta_{\mathcal{M}}$ is obtained by maximizing the density of the sample with respect to the parameters (we seek the value of $\theta_{\mathcal{M}}$ that maximizes the "probability of occurence" of the sample given by the Nature).
- In the sequel, we will denote $\mathcal{L}(\theta) = f(\mathcal{Y}_T|\theta)$ the likelihood function, omitting the indexation with respect to the model when not necessary.

ML Estimation

A simple static model

- As a first example, we consider the following model:

$$y_t = \mu_0 + \epsilon_t \tag{2-a}$$

where $\epsilon_t \underset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and μ_0 is an unknown finite real parameter.

- According to this model, y_t is normally distributed:

$$y_t|\mu_0 \sim \mathcal{N}(\mu_0, 1)$$

and $\mathbb{E}[y_t y_s] = 0$ for all $s \neq t$.

- Suppose that a sample $\mathcal{Y}_T = \{y_1, \ldots, y_T\}$ is available. The likelihood is defined by:

$$\mathcal{L}(\mu) = f(y_1, \ldots, y_T|\mu)$$

- Because the ys are iid, the joint conditional density is equal to a product of conditional densities:

$$\mathcal{L}(\mu) = \prod_{t=1}^{T} g(y_t|\mu)$$

ML Estimation

A simple static model

- Since the model is linear and Gaussian:

$$\mathcal{L}(\mu) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t-\mu)^2}{2}}$$

- Finally we have:

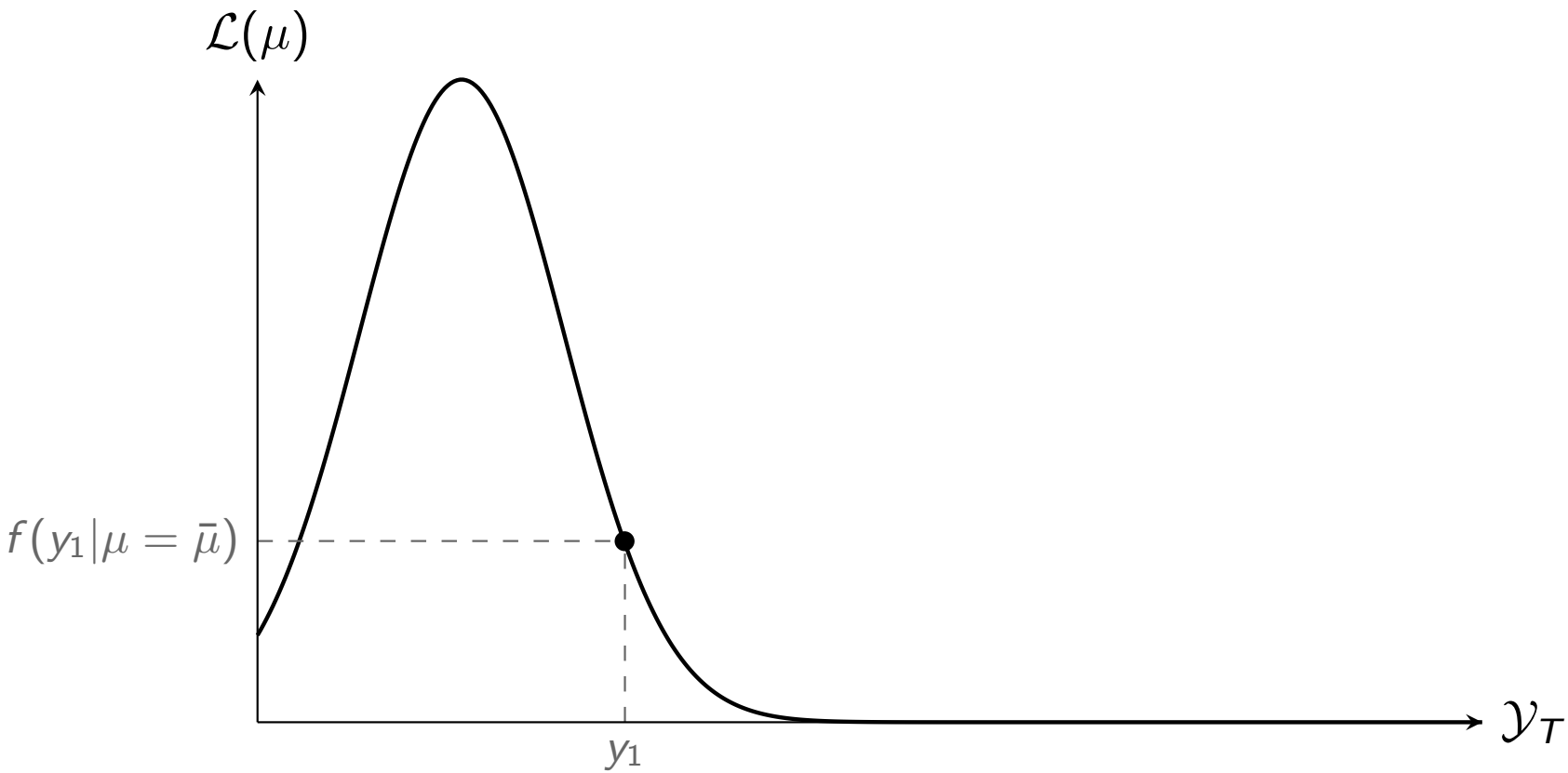
$$\mathcal{L}(\mu) = (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu)^2} \qquad \text{(2-b)}$$

- Note that the likelihood function depends on the data and the unknown parameter (otherwise we would have an identification issue).

- Suppose that $T = 1$ (only one observation in the sample). We can graphically determine the ML estimator of μ in this case.

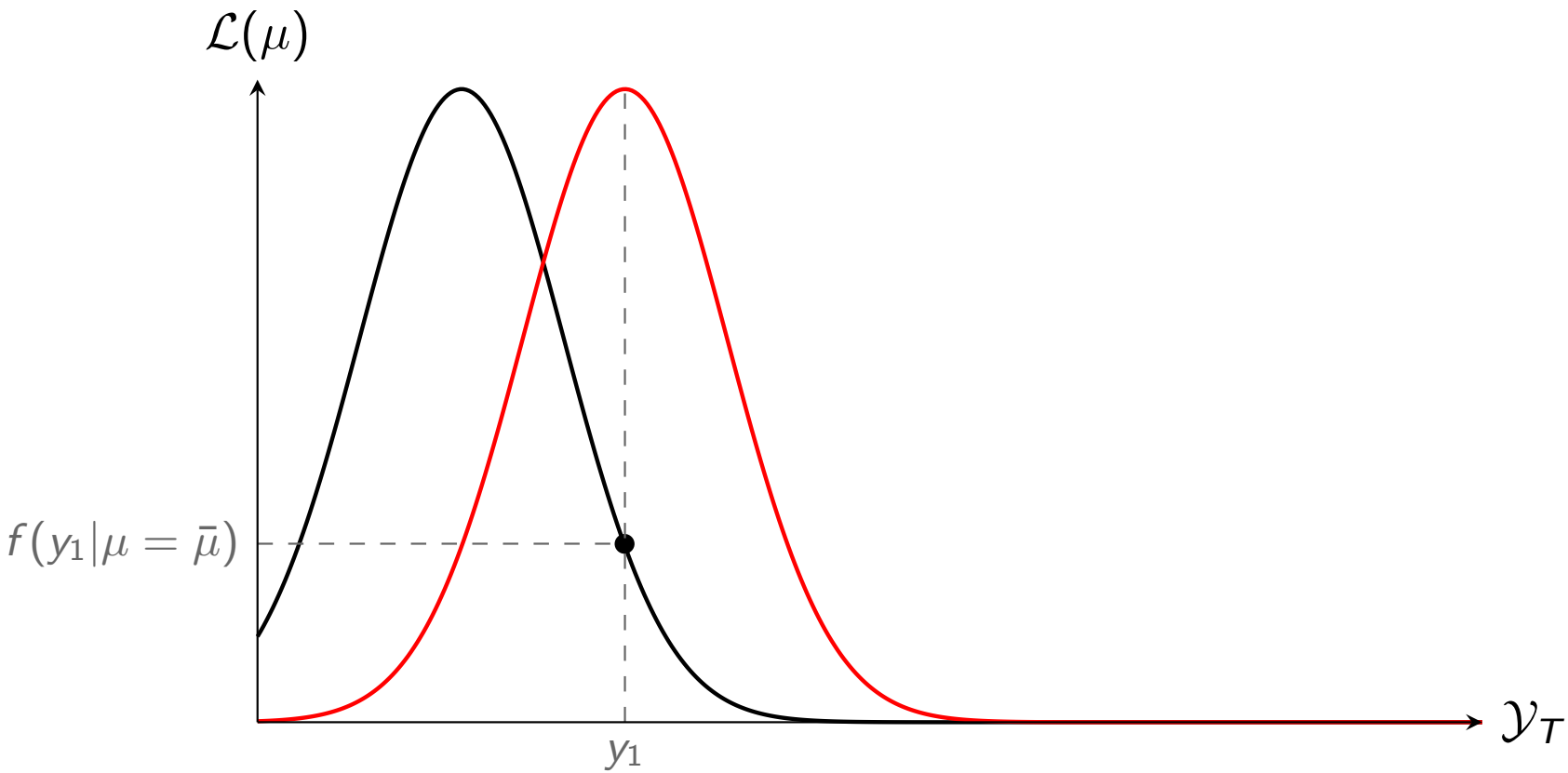
ML Estimation

A simple static model (cont'd)



ML Estimation

A simple static model (cont'd)



Clearly, the value of the density of y_1 conditional on μ, *ie* the likelihood, is maximized for $\mu = y_1$: for any $\bar{\mu} \neq y_1$ we have $f(y_1|\mu = \bar{\mu}) < f(y_1|\mu = y_1)$

ML Estimation

A simple static model (cont'd)

$\Rightarrow$ If we have only one observation, y_1, the Maximum Likelihood estimator is equal to the observation: $\widehat{\mu} = y_1$.

- This estimator is unbiased and its variance is 1.
- More generally, one can show that the maximum likelihood estimator is equal to the sample mean:

$$\widehat{\mu}_T = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{2-c}$$

- This estimator is unbiased and its variance is given by:

$$\mathbb{V}\left[\widehat{\mu}_T\right] = \frac{1}{T} \tag{2-d}$$

- Because $\mathbb{V}\left[\widehat{\mu}\right]$ goes to zero as the sample size goes to infinity, we know that this estimator converges in probability to the true value μ_0 of the unknown parameter:

$$\widehat{\mu}_T \xrightarrow[T \to \infty]{\text{proba}} \mu_0$$

ML Estimation

A simple dynamic model

- Suppose that the data are generated by an AR(1) model:

$$y_t = \varphi y_{t-1} + \epsilon_t$$

with $|\varphi| < 1$ and $\epsilon_t \underset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right)$.

- In this case, y_t depends (directly) on y_{t-1} and also on $y_{t-2}, y_{t-3}, \ldots$.
- It is no more legal to write the likelihood as the as a product of marginal densities of the observations.
- The joint density of $\mathrm{y} \equiv (y_1, y_2, \ldots, y_T)'$ is given by:

$$f(\mathrm{y}) = (2\pi)^{-\frac{H}{2}} |\Sigma_y|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathrm{y}'\Sigma_y^{-1}\mathrm{y}}$$

with

$$\Sigma_y = \frac{\sigma_\epsilon^2}{1-\varphi^2} \begin{pmatrix} 1 & \varphi & \varphi^2 & \ldots & \ldots & \varphi^{T-1} \\ \varphi & 1 & \varphi & \varphi^2 & \ldots & \varphi^{T-2} \\ \vdots & & & & & \\ \varphi^{T-1} & \varphi^{T-2} & \ldots & \ldots & \varphi & 1 \end{pmatrix}$$

under the assumption of stationarity $\Rightarrow$ Likelihood function.

Bayes theorem

- Let A and B be two events.
- Let $\mathbb{P}(A)$ and $\mathbb{P}(B)$ be the marginal probabilities.
- Let $\mathbb{P}(A \cap B)$ be the joint probability of A and B.
- The Bayes theorem states that the probability of B conditional on A is given by:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Splitting a joint probability

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) \quad \text{or} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

Conditioning inversion

$$\Rightarrow \mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

ML Estimation

A simple dynamic model (cont'd)

- The inverse of the covariance matrix, Σ_y, can be factorized as $\Sigma_y^{-1} = \sigma_\epsilon^{-2} L'L$ with:

$$L = \begin{pmatrix} \sqrt{1-\varphi^2} & 0 & 0 & \dots & 0 & 0 \\ -\varphi & 1 & 0 & \dots & 0 & 0 \\ 0 & -\varphi & 1 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & & & & -\varphi & 1 \end{pmatrix}$$

  a $T \times T$ matrix.

- First manifestation of the Bayes theorem.

- The (ecact) likelihood function can be written equivalently as:

$$\mathcal{L}(\varphi, \sigma_\epsilon^2) = (2\pi)^{-\frac{T-1}{2}} \sigma_\epsilon^{-(T-1)} e^{-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^{T} (y_t - \varphi y_{t-1})^2} \times (2\pi)^{-\frac{1}{2}} \left(\frac{\sigma_\epsilon^2}{1-\varphi^2} \right)^{-\frac{1}{2}} e^{-\frac{1-\varphi^2}{2\sigma_\epsilon^2} y_1^2}$$

  — Product of conditional densities, $y_t | y_{t-1}$ (the conditional likelihood) and the marginal density of the initial condition y_1.

Prior and posterior beliefs

- We assume that we are able to characterize our prior knowledge about a parameter with a probability density function.
- Let $p_0(\theta)$ be the prior density characterizing our beliefs about the vector of parameters θ.
- Our aim is to update our (prior) beliefs about θ with the sample information $(\mathcal{Y}_T)$ embodied in the likelihood function, $\mathcal{L}(\theta) = f(\mathcal{Y}_T|\theta)$.
- We define the posterior density, $p_1(\theta|\mathcal{Y}_T)$, which represents our updated beliefs.
- By the Bayes theorem we have:

$$p_1(\theta|\mathcal{Y}_T) = \frac{g(\theta, \mathcal{Y}_T)}{p(\mathcal{Y}_T)}$$

where g is the joint density of the sample and the parameters, and

$$p_1(\theta|\mathcal{Y}_T) = \frac{f(\mathcal{Y}_T|\theta)p_0(\theta)}{p(\mathcal{Y}_T)}$$

Prior and posterior beliefs (cont'd)

- The posterior density is given by:

$$p_1(\theta|\mathcal{Y}_T) = \frac{\mathcal{L}(\theta)p_0(\theta)}{p(\mathcal{Y}_T)}$$

- The denominator does not depend on the parameters $\rightarrow$ the posterior density is proportional (w.r.t θ) to the product of the likelihood and the prior density (the posterior kernel):

$$p_1(\theta|\mathcal{Y}_T) \propto \mathcal{L}(\theta)p_0(\theta)$$

- Posterior inference about the parameters only requires $\mathcal{L}(\theta)p_0(\theta)$.

- The denominator is the marginal density of the sample. Because a density has to sum up to one, we have:

$$p(\mathcal{Y}_T) = \int f(\mathcal{Y}_T|\theta)p_0(\theta)\mathrm{d}\theta$$

The marginal density is a weighted average of the likelihood function $\rightarrow$ will be used later for model comparison.

Prior and posterior beliefs

A simple static model (cont'd, with informative prior)

- For the sake of simplicity, we choose a Gaussian prior for the parameter μ, with prior expectation μ_0 and prior variance σ_μ^2:

$$p_0(\mu) = \frac{1}{\sigma_\mu \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\mu^2}(\mu-\mu_0)^2}$$

- The smaller is σ_μ^2, the more informative is the prior.

- The posterior density is proportional to the product of the prior density and the likelihood:

$$p_1(\mu|\mathcal{Y}_T) \propto \frac{1}{\sigma_\mu \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\mu^2}(\mu-\mu_0)^2} (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu)^2}$$

- One can show that the RHS expression is proportional to a Gaussian density.

Prior and posterior beliefs

A simple static model (cont'd, with informative prior)

- The likelihood can be equivalently written as:

$$\mathcal{L}(\mu) = (2\pi)^{-\frac{T}{2}} e^{-\frac{1}{2}\left(\nu s^2 + T(\mu - \widehat{\mu})^2\right)}$$

with $\nu = T - 1$ and

$$s^2 = \frac{1}{\nu} \sum_{t=1}^{T} (y_t - \widehat{\mu})^2$$

- s^2 and $\widehat{\mu}$ are sufficient statistics: they convey all the necessary sample information regarding the inference w.r.t μ.

- The likelihood is proportional to a Gaussian density centered on $\widehat{\mu}$ with variance $1/T$.

- The posterior density:

$$p_1(\mu|\mathcal{Y}_T) \propto \frac{1}{\sigma_\mu \left(\sqrt{2\pi}\right)^{T+1}} e^{-\frac{1}{2\sigma_\mu^2}(\mu - \mu_0)^2 - \frac{\nu}{2} s^2 - \frac{T}{2}(\mu - \widehat{\mu})^2}$$

Prior and posterior beliefs

A simple static model (cont'd, with informative prior)

- Previous expression can be simplified by omitting all the multiplicative terms not related to μ (this is legal because we are interested in a proportionality w.r.t μ):

$$p_1(\mu|\mathcal{Y}_T) \propto e^{-\frac{1}{2\sigma_\mu^2}(\mu-\mu_0)^2-\frac{T}{2}(\mu-\widehat{\mu})^2}$$

- We develop the quadratic forms and remove all the terms appearing additively (under the exponential function); we obtain:

$$p_1(\mu|\mathcal{Y}_T) \propto e^{-\frac{1}{2}\left(\sigma_\mu^{-2}+T\right)\left(\mu-\frac{T\widehat{\mu}+\mu_0\sigma_\mu^{-2}}{T+\sigma_\mu^{-2}}\right)^2}$$

- We recognize the expression of a Gaussian density (up to a scale parameter that does not depend on μ).

Let $A(\mu) = \frac{1}{\sigma_\mu^2}(\mu - \mu_0)^2 + T(\mu - \widehat{\mu})^2$. We establish the last expression of the posterior kernel by rewriting $A(\mu)$ as:

$$
\begin{aligned}
A(\mu) &= T(\mu - \widehat{\mu})^2 + \frac{1}{\sigma_\mu^2}(\mu - \mu_0)^2 \\
&= T\left(\mu^2 + \widehat{\mu}^2 - 2\mu\widehat{\mu}\right) + \frac{1}{\sigma_\mu^2}\left(\mu^2 + \mu_0^2 - 2\mu\mu_0\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\mu^2 - 2\mu\left(T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0\right) + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\left[\mu^2 - 2\mu\frac{T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0}{T + \frac{1}{\sigma_\mu^2}}\right] + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&= \left(T + \frac{1}{\sigma_\mu^2}\right)\left[\mu - \frac{T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0}{T + \frac{1}{\sigma_\mu^2}}\right]^2 + \left(T\widehat{\mu}^2 + \frac{1}{\sigma_\mu^2}\mu_0^2\right) \\
&\qquad - \frac{\left(T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0\right)^2}{T + \frac{1}{\sigma_\mu^2}}
\end{aligned}
$$

In the last equality, the two last additive terms do not depend on μ and can be therefore omitted.

Prior and posterior beliefs

A simple static model (cont'd, with informative prior)

- The posterior distribution is Gaussian with (posterior) expectation:

$$\mathbb{E}[\mu] = \frac{T\widehat{\mu} + \frac{1}{\sigma_\mu^2}\mu_0}{T + \frac{1}{\sigma_\mu^2}}$$

  and (posterior) variance:

$$\mathbb{V}[\mu] = \frac{1}{T + \frac{1}{\sigma_\mu^2}}$$

- As soon as the amount of prior information is positive ($\sigma_\mu^2 < \infty$) the posterior variance is less than the variance of the maximum likelihood estimator ($1/T$).

- The posterior expectation is a convex combination of the maximum likelihood estimator and the prior expectation.

Prior and posterior beliefs

A simple static model (cont'd, with informative prior)

- The Bayesian approach can be interpreted as a bridge between the calibration approach ($\sigma_\mu^2 = 0$, infinite amount of prior information) and the ML approach ($\sigma_\mu^2 = \infty$, no prior information):

$$\mathbb{E}[\mu] \xrightarrow[\sigma_\mu^2 \to 0]{} \mu_0$$

and

$$\mathbb{E}[\mu] \xrightarrow[\sigma_\mu^2 \to \infty]{} \widehat{\mu}$$

- The more important is the amount of information in the sample, the smaller will be the gap between the posterior expectation and the ML estimator.

Prior and posterior beliefs

A simple dynamic model (cont'd, flat prior)

- For a real parameter, the prior is proportional to 1.

$$p_0(\varphi) \propto 1$$

- For a positive parameter, the prior of the log of the prior is proportional to 1.

$$p_0(\log \sigma_\epsilon^2) \propto 1 \quad \Leftrightarrow \quad p_0(\sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon}$$

Flat prior for AR(1) model

$$p_0\left(\varphi, \sigma_\epsilon^2\right) \propto \frac{1}{\sigma_\epsilon}$$

$\hookrightarrow$ Parameters φ and σ_ϵ^2 are a priori independent.
$\hookrightarrow$ This prior is not a pdf!

Prior and posterior beliefs

A simple dynamic model (cont'd, flat prior)

- Use the conditional likelihood (omitting the marginal density of the first observation).

- The posterior density is:

$$p_1(\varphi, \sigma_\epsilon^2) \propto \underbrace{\sigma_\epsilon^{-1}}_{\text{Prior}} \underbrace{(2\pi)^{-\frac{T-1}{2}} \sigma_\epsilon^{-(T-1)} e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t-\varphi y_{t-1})^2}}_{\text{Conditional likelihood}}$$

$$\propto \sigma_\epsilon^{-T} e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t-\varphi y_{t-1})^2}$$

Posterior density

$$\varphi|\sigma_\epsilon^2, \mathcal{Y}_T \sim \mathcal{N}\left(\frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}, \frac{\sigma_\epsilon^2}{\sum_{t=2}^{T} y_{t-1}^2}\right)$$

$$\sigma_\epsilon^2|\mathcal{Y}_T \sim \mathcal{IG}\left(T-2, \sum_{t=2}^{T} \widehat{\epsilon}_t^2\right)$$

Point estimate

- The outcome of the Bayesian approach is a (posterior) probability density function.
- But people generally expect much less information: a point estimate is often enough for most practical purposes (a single value for each parameter with a measure of uncertainty).

$\Rightarrow$ We need to reduce a distribution to a "representative" point.

- Let $L(\theta, \widehat{\theta})$ be the loss incurred if we choose $\widehat{\theta}$ while θ is the true value.

- The idea is to choose the value of θ that minimizes this loss... But the true value of θ is obviously unknown, so we minimize the (posterior) expected loss instead:

$$\theta^\star = \arg\min_{\widehat{\theta}} \mathbb{E}\left[L(\theta, \widehat{\theta})\right] = \arg\min_{\widehat{\theta}} \int L(\theta, \widehat{\theta}) p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta$$

- The choice of the loss function is purely arbitrary, for each loss we will obtain a different point estimate.

Point estimate

Quadratic loss function (L_2 norm)

- Suppose that the loss function is quadratic:

$$L(\theta, \widehat{\theta}) = (\theta - \widehat{\theta})' \Omega (\theta - \widehat{\theta})$$

where Ω is a symmetric positive definite matrix. Note that this function returns a (real) scalar.

- The (posterior) expectation of the loss is:

$$\begin{aligned}
\mathbb{E}\left[L(\theta, \widehat{\theta})\right] &= \mathbb{E}\left[(\theta - \widehat{\theta})' \Omega (\theta - \widehat{\theta})\right] \\
&= \mathbb{E}\left[\left(\theta - \mathbb{E}\theta - \left(\widehat{\theta} - \mathbb{E}\theta\right)\right)' \Omega \left(\theta - \mathbb{E}\theta - \left(\widehat{\theta} - \mathbb{E}\theta\right)\right)\right] \\
&= \mathbb{E}\left[(\theta - \mathbb{E}\theta)' \Omega (\theta - \mathbb{E}\theta)\right] + (\widehat{\theta} - \mathbb{E}\theta)' \Omega (\widehat{\theta} - \mathbb{E}\theta)
\end{aligned}$$

- Noting that the first term does not depend on the choice variable, $\widehat{\theta}$, the expected loss is trivially minimized when $\widehat{\theta}$ is equal to the (posterior) expectation of θ:

$$\theta^\star = \mathbb{E}[\theta]$$

$\Rightarrow$ If the loss is quadratic the optimal point estimate is the posterior expectation.

Point estimate

Absolute value loss function (L_1 norm)

- Suppose that the (univariate) loss function is:

$$L(\theta, \widehat{\theta}) = |\theta - \widehat{\theta}|$$

- The (posterior) expectation of the loss is:

$$\begin{aligned}\mathbb{E}\left[L(\theta, \widehat{\theta})\right] &= \int_{-\infty}^{\infty} |\theta - \widehat{\theta}| p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta \\ &= \int_{-\infty}^{\widehat{\theta}} (\widehat{\theta} - \theta) p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta + \int_{\widehat{\theta}}^{\infty} (\theta - \widehat{\theta}) p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta\end{aligned}$$

- F.O.C.

$$\begin{aligned}&\int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta - \int_{\theta^\star}^{\infty} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = 0 \\ \Rightarrow\ & 2 \int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = \int_{-\infty}^{\infty} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta \\ \Leftrightarrow\ & \int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = \frac{1}{2} \quad (\rightarrow \text{posterior median})\end{aligned}$$

Point estimate

Generalized bsolute value loss function (quantiles)

- Suppose that the (univariate) loss function is:

$$L(\theta, \widehat{\theta}) = \begin{cases} \alpha(\theta - \widehat{\theta}) & \text{if } \theta \geq \widehat{\theta} \\ \beta(\widehat{\theta} - \theta) & \text{if } \theta < \widehat{\theta} \end{cases}$$

  with $\beta \neq \alpha$ two positive parameters.

- The (posterior) expectation of the loss is:

$$\mathbb{E}\left[L(\theta, \widehat{\theta})\right] = \beta \int_{-\infty}^{\widehat{\theta}} (\widehat{\theta} - \theta) p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta + \alpha \int_{\widehat{\theta}}^{\infty} (\theta - \widehat{\theta}) p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta$$

- F.O.C.

$$\beta \int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta - \alpha \int_{\theta^\star}^{\infty} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = 0$$

$$\Rightarrow (\beta + \alpha) \int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = \alpha \int_{-\infty}^{\infty} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta$$

$$\Leftrightarrow \int_{-\infty}^{\theta^\star} p_1(\theta|\mathcal{Y}_T) \mathrm{d}\theta = \frac{\alpha}{\alpha + \beta} \quad (\rightarrow \text{any posterior quantile})$$

Marginal density of the sample

- If we are only interested in inference about the parameters, the marginal density of the data, $p(\mathcal{Y}_T)$, can be omitted.
- We already saw that the marginal density of the data is:

$$p(\mathcal{Y}_T) = \int f(\mathcal{Y}_T|\theta)p_0(\theta)\mathrm{d}\theta$$

  - The marginal density of the sample acts as a constant of integration in the expression of the posterior density.
  - The marginal density of the sample is an average of the likelihood function (for different values of the estimated parameters) weighted by the prior density.

$\Rightarrow$ The marginal density of the sample is a measure of fit, which does not depend on the parameters (because we integrate them out).

- Note that, theoretically, it is possible to compute the marginal density of the sample (conditional on a model) without estimating the parameters.

Marginal density of the sample

A simple static model (cont'd)

- Suppose again that the sample size is $T = 1$. The likelihood is given by:

$$f(\mathcal{Y}_T|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_1-\mu)^2}$$

- The marginal density is then given by:

$$\begin{aligned} p(\mathcal{Y}_T) &= \int_{-\infty}^{\infty} f(y_1|\mu) p_0(\mu) \mathrm{d}\mu \\ &= (2\pi\sigma_\mu)^{-1} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((y_1-\mu)^2 + \frac{(\mu-\mu_0)^2}{\sigma_\mu^2}\right)} \mathrm{d}\mu \\ &= \frac{1}{\sqrt{2\pi(1+\sigma_\mu^2)}} e^{-\frac{(y_1-\mu_0)^2}{2(1+\sigma_\mu^2)}} \end{aligned}$$

- We can directly obtain the same result by noting that y_1 is the sum of two Gaussian random variables: $\mathcal{N}(0,1)$ and $\mathcal{N}\left(\mu_0, \sigma_\mu^2\right)$.

Marginal density of the sample

Model comparison

- Suppose we have two models $\mathcal{A}$ and $\mathcal{B}$ (with two associated vectors of deep parameters $\theta_{\mathcal{A}}$ and $\theta_{\mathcal{B}}$) estimated using *the same sample* $\mathcal{Y}_T$.
- For each model $\mathcal{I} = \mathcal{A}, \mathcal{B}$ we can evaluate, at least theoretically, the marginal density of the data conditional on the model:

$$p(\mathcal{Y}_T|\mathcal{I}) = \int f(\mathcal{Y}_T|\theta_{\mathcal{I}}, \mathcal{I}) p_0(\theta_{\mathcal{I}}|\mathcal{I}) d\theta_{\mathcal{I}}$$

  by integrating out the deep parameters $\theta_{\mathcal{I}}$ from the posterior kernel.
- $p(\mathcal{Y}_T|\mathcal{I})$ measures the fit of model $\mathcal{I}$. If we have to choose between models $\mathcal{A}$ and $\mathcal{B}$ we will select the model with the highest marginal density of the sample.
- Note that models $\mathcal{A}$ and $\mathcal{B}$ need not to be nested (for instance, we do not require that $\theta_{\mathcal{A}}$ be a subset of $\theta_{\mathcal{B}}$) for the comparison to make sense, because the compared marginal densities do not depend on the parameters. The classical approach (comparisons of likelihoods) by requiring nested models is much less obvious.

Marginal density of the sample

Model comparison (cont'd)

- Suppose we have a prior distribution over models $\mathcal{A}$ and $\mathcal{B}$: $p(\mathcal{A})$ and $p(\mathcal{B})$.

- Again, using the Bayes theorem we can compute the posterior distribution over models:

$$p(\mathcal{I}|\mathcal{Y}_T) = \frac{p(\mathcal{I})p(\mathcal{Y}_T|\mathcal{I})}{\sum_{\mathcal{I}=\mathcal{A},\mathcal{B}} p(\mathcal{I})p(\mathcal{Y}_T|\mathcal{I})}$$

- This formula may easily be generalized to a collection of N models.

- In the literature posterior odds ratio, defined as:

$$\frac{p(\mathcal{A}|\mathcal{Y}_T)}{p(\mathcal{B}|\mathcal{Y}_T)} = \frac{p(\mathcal{A})}{p(\mathcal{B})} \frac{p(\mathcal{Y}_T|\mathcal{A})}{p(\mathcal{Y}_T|\mathcal{B})}$$

  are often used to discriminate between different models. If the posterior odds ratio is large ($>$100) we can safely choose model $\mathcal{A}$.

Marginal density of the sample

Model comparison (cont'd)

- Note that we do not necessarily have to choose one model.
- Even if a model has a smaller posterior probability (or marginal density) it may provide useful informations in some directions (or frequencies), so we should not discard this information.
- An alternative is to mix the models.
- If these models are used for forecasting inflation, we can report an average of the forecasts weighted by the posterior probabilities, $p(\mathcal{I}|\mathcal{Y}_T)$, instead of the forecasts of the best model (in terms of marginal density) $\rightarrow$ Bayesian averaging.

Predictive density

- We often seek to use the estimated model to do inference about unobserved variables.
- The most obvious example is the forecasting exercise.
- In the Bayesian context the density of an unobserved variable (for instance the future growth of GDP) given the sample, is called a predictive density.
- Let $\tilde{y}$ be a vector of unobserved variables. The joint posterior density of $\tilde{y}$ and θ is:

$$p_1(\tilde{y}, \theta | \mathcal{Y}_T) = g(\tilde{y} | \theta, \mathcal{Y}_T) p_1(\theta | \mathcal{Y}_T)$$

- The posterior predictive density is obtained by integrating out the parameters:

$$p(\tilde{y} | \mathcal{Y}_T) = \int g(\tilde{y} | \theta, \mathcal{Y}_T) p_1(\theta | \mathcal{Y}_T) \mathrm{d}\theta$$

The posterior predictive density is the average of the density of $\tilde{y}$ knowing the parameters weighted by the posterior density of the parameters.

Predictive density

A simple static model (cont'd)

- Suppose that we want to do inference about the out of sample variable y_{T+1} (forecast).

- The density of y_{T+1} conditional on the sample and on the parameter is:

$$g(y_{T+1}|\mu, \mathcal{Y}_T) \propto e^{-\frac{1}{2}(y_{T+1}-\mu)^2}$$

  Note that this conditional density does not depend on $\mathcal{Y}_T$ because the model is static (for an autoregressive model, at least y_T would appear under the quadratic term).

- Remember that the posterior density of μ is:

$$p_1(\mu|\mathcal{Y}_T) \propto e^{-\frac{1}{2\mathbb{V}[\mu]}(\mu-\mathbb{E}[\mu])^2}$$

  where $\mathbb{E}[\mu]$ and $\mathbb{V}[\mu]$ are the posterior first and second order moments obtained earlier.

Predictive density

A simple static model (cont'd)

- The predictive density for y_{T+1} is given by:

$$p(y_{T+1}|\mathcal{Y}_T) = \int g(y_{T+1}|\mu, \mathcal{Y}_T) p_1(\mu|\mathcal{Y}_T) \mathrm{d}\mu$$
$$\propto \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y_{T+1}-\mu)^2 - \frac{1}{2\mathbb{V}[\mu]}(\mu-\mathbb{E}[\mu])^2} \mathrm{d}\mu$$

- The terms under the exponential in the last expression can be rewritten as:

$$\left[1 + \frac{1}{\mathbb{V}[\mu]}\right] \left(\mu - \frac{y_{T+1} + \mathbb{V}[\mu]^{-1}}{1 + \mathbb{V}[\mu]^{-1}}\right)^2 + \frac{\mathbb{V}[\mu]^{-1}}{1 + \mathbb{V}[\mu]^{-1}} (y_{T+1} - \mathbb{E}[\mu])^2$$

Predictive density

A simple static model (cont'd)

- By substitution in the expression of the predictive density for y_{T+1} we obtain:

$$
\begin{aligned}
p(y_{T+1}|\mathcal{Y}_T) &\propto \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[1+\frac{1}{\mathbb{V}[\mu]}\right]\left(\mu-\frac{y_{T+1}+\mathbb{V}[\mu]^{-1}}{1+\mathbb{V}[\mu]^{-1}}\right)^2-\frac{1}{2}\frac{\mathbb{V}[\mu]^{-1}}{1+\mathbb{V}[\mu]^{-1}}(y_{T+1}-\mathbb{E}[\mu])^2}\,\mathrm{d}\mu \\
&\propto e^{-\frac{1}{2}\frac{\mathbb{V}[\mu]^{-1}}{1+\mathbb{V}[\mu]^{-1}}(y_{T+1}-\mathbb{E}[\mu])^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[1+\frac{1}{\mathbb{V}[\mu]}\right]\left(\mu-\frac{y_{T+1}+\mathbb{V}[\mu]^{-1}}{1+\mathbb{V}[\mu]^{-1}}\right)^2}\,\mathrm{d}\mu \\
&\propto e^{-\frac{1}{2}\frac{1}{1+\mathbb{V}[\mu]}(y_{T+1}-\mathbb{E}[\mu])^2}
\end{aligned}
$$

- Unsurprisingly, we recognize the Gaussian density:

$$y_{T+1}|\mathcal{Y}_T \sim \mathcal{N}\left(\mathbb{E}[\mu], 1+\mathbb{V}[\mu]\right)$$

- We would have obtained directly the same result by first noting that y_{T+1} is the sum of two Gaussian random variables: $\mathcal{N}\left(\mathbb{E}[\mu], \mathbb{V}[\mu]\right)$ (for the estimated parameter) and $\mathcal{N}\left(0, 1\right)$ (for the error term).

Predictive density

Point prediction

- For reporting our forecast, we may want to select one point in the predictive distribution.
- We proceed as for the point estimate by choosing an arbitrary loss function and minimizing the posterior expected loss.
- Usually the expectation of the predictive distribution is reported (rationalized with a quadratic loss function).

Estimation of DSGE models

Structural form

- We suppose that the DSGE model can be cast in the following form:

$$\mathbb{E}_t \left[\mathcal{F}_\theta (y_{t+1}, y_t, y_{t-1}, \varepsilon_t) \right] = 0 \tag{2}$$

with $\varepsilon_t \sim \text{iid}\left(0, \Sigma_\varepsilon\right)$ is a random vector $(r \times 1)$ of structural innovations, $y_t \in \Lambda \subseteq \mathbb{R}^n$ a vector of endogenous variables, $\mathcal{F}_\theta : \Lambda^3 \times \mathbb{R}^r \to \Lambda$ a real function in $\mathcal{C}^2$ parameterized by a real vector $\theta \in \Theta \subseteq \mathbb{R}^q$ gathering the deep parameters of the model.

- The model is stochastic, forward looking and non linear.

- We want to estimate (a subset of) θ. For any estimation approach (indirect inference, simulated moments, maximum likelihood,...) we need first to solve this model.

Estimation of DSGE models

Reduced form

- In the sequel we consider a local linear approximation around the deterministic steady state of the non linear model.

- The solution of the linearized model (the reduced form) is:

$$y_t = \bar{y}(\theta) + A(\theta)\left(y_{t-1} - \bar{y}(\theta)\right) + B(\theta)\varepsilon_t$$

  where the steady state, $\bar{y}(\theta)$, and matrices $A(\theta)$ and $B(\theta)$ are nonlinear functions of the deep parameters.

- The unconditional covariance matrix of y solves:

$$\Sigma_y = A(\theta)\Sigma_y A(\theta)' + B(\theta)\Sigma_\varepsilon B(\theta)'$$

  and the autocovariance function is defined as:

$$\Gamma_h = A(\theta)\Gamma_{h-1} \quad \forall h \geq 1$$

  with $\Gamma_0 = \Sigma_y$ and $\Gamma_{-h} = \Gamma_h'$.

Estimation of DSGE models

Likelihood

- Let $\mathcal{Y}_T = y_1^\star, y_2^\star, \ldots, y_T^\star$ be the sample, with

$$y_t^\star = Z y_t$$

where Z is a $p \times n$ selection matrix.

- The likelihood is the density of the sample. If the structural innovations are Gaussian:

$$\mathcal{L} = (2\pi)^{-\frac{pT}{2}} |\Sigma_{\mathbf{y}^\star}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}^\star - \bar{\mathbf{y}}^\star)' \Sigma_{\mathbf{y}^\star}^{-1} (\mathbf{y}^\star - \bar{\mathbf{y}}^\star)}$$

with $\mathbf{y}^\star = (\mathbf{y}_1^{\star\prime}, \mathbf{y}_2^{\star\prime}, \ldots, \mathbf{y}_T^{\star\prime})'$ a $Tp \times 1$ vector, and

$$\Sigma_{\mathbf{y}^\star} = \begin{pmatrix} \Gamma_0^\star & \Gamma_1^\star & \Gamma_2^\star & \cdots & \cdots & \cdots & \Gamma_{T-1}^\star \\ \Gamma_1^{\star\prime} & \Gamma_0^\star & \Gamma_1^\star & \Gamma_2^\star & \cdots & \cdots & \Gamma_{T-2}^\star \\ \Gamma_2^{\star\prime} & \Gamma_1^{\star\prime} & \Gamma_0^\star & \Gamma_1^\star & \cdots & \cdots & \Gamma_{T-3}^\star \\ \vdots & & & & & & \\ \Gamma_{T-1}^{\star\prime} & \Gamma_{T-2}^{\star\prime} & \cdots & \cdots & \cdots & \Gamma_1^{\star\prime} & \Gamma_0^\star \end{pmatrix}$$

where $\Gamma_h^\star = Z \Gamma_h Z'$. $\rightarrow$ Symmetric block Toeplitz matrix.

Estimation of DSGE models

Kalman filter

- The well known Kalman filter bayesian recursive algorithm can be used to evaluate the likelihood:

$$\begin{aligned} v_t &= y_t^\star - \bar{y}(\theta)^\star - Z\hat{y}_t \\ F_t &= ZP_tZ' + \mathbb{V}[\eta] \\ K_t &= A(\theta)P_tA(\theta)'F_t^{-1} \\ \hat{y}_{t+1} &= A(\theta)\hat{y}_t + K_tv_t \\ P_{t+1} &= A(\theta)P_t(A(\theta) - K_tZ)' + B(\theta)\Sigma_\varepsilon B(\theta)' \end{aligned}$$

  for $t = 1, \ldots, T$, with initial condition $\hat{y}_0 = y_0^\star - \bar{y}^\star$ and P_0 given by the ergodic distribution of $y^\star$.

- The (log)-likelihood is:

$$\ln \mathcal{L} = -\frac{Tp}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}|F_t| - \frac{1}{2}v_t'F_t^{-1}v_t$$

- F_t is full rank only if we have at least as many innovations as observed variables.

Estimation of DSGE models

Posterior distribution

- We know that the posterior density is proportional to the likelihood times the prior density.

- Because we do not have closed form expressions for the reduced form of the model and the likelihood, we cannot obtain analytical results about the posterior density.

- Posterior inference can be done by considering:
  1. Asymptotic (Gaussian) approximation of the posterior density.
  2. Simulation based methods (exact up to the randomness inherent to these methods).

Simulation based posterior inference

- We need a simulation approach if we want to obtain exact results (*ie* not relying on asymptotic approximation).

- Noting that:

$$\mathbb{E}\left[\varphi(\psi)\right] = \int_{\Psi} \varphi(\psi) p_1(\psi|\mathcal{Y}_T) \mathrm{d}\psi$$

  we can use the empirical mean of $\left(\varphi(\psi^{(1)}), \varphi(\psi^{(2)}), \ldots, \varphi(\psi^{(n)})\right)$, where $\psi^{(i)}$ are draws from the posterior distribution to evaluate the expectation of $\varphi(\psi)$. The approxomation error goes to zero when $n \to \infty$.

- We need to simulate draws from the posterior distribution.

  $\Rightarrow$ Metropolis-Hastings algorithm.

Simulation based posterior inference

A simple example – a –

- Imagine we want to obtain some draws from a $\mathcal{N}(0, 4)$ distribution...
- But we are only able to draw from $\mathcal{N}(0, 1)$ and we don't realize that we should simply multiply by 2 the draws from a standard normal distribution.
- The idea is to build a stochastic process whose limiting distribution is $\mathcal{N}(0, 4)$.
- We define the following AR(1) process:

$$x_t = \rho x_{t-1} + \epsilon_t$$

with $\epsilon_t \sim \mathcal{N}(0, 1)$, $|\rho| < 1$ and $x_0 = 0$.

- We just have to choose ρ such that the asymptotic distribution of $\{x_t\}$ is $\mathcal{N}(0, 4)$.

Simulation based posterior inference

A simple example – b –

We have:

- $x_1 = \epsilon_1 \sim \mathcal{N}(0, 1)$
- $x_2 = \rho\epsilon_1 + \epsilon_2 \sim \mathcal{N}\left(0, 1 + \rho^2\right)$
- $x_3 = \rho^2\epsilon_1 + \rho\epsilon_2 + \epsilon_3 \sim \mathcal{N}\left(0, 1 + \rho^2 + \rho^4\right)$
- ...
- $x_T = \rho^{T-1}\epsilon_1 + \rho^{T-2}\epsilon_2 + \cdots + \epsilon_T \sim \mathcal{N}\left(0, 1 + \rho^2 + \ldots \rho^{2(T-1)}\right)$
- ...
- And asymptotically

$$x_\infty \sim \mathcal{N}\left(0, \frac{1}{1-\rho^2}\right)$$

So that $\mathbb{V}_\infty[x_t] = 4$ iff $\rho = \pm\frac{\sqrt{3}}{2}$.

Simulation based posterior inference

A simple example – c –

- If we simulate enough draws from this Gaussian autoregressive stochastic process, we can replicate the targeted distribution.
- In this case it is very simple because we know exactly the targeted distribution **and** we are able to obtain some draws from its standardized version.
- This is far from true with DSGE models. For instance, we even don't have an analytical expression for the posterior density.

Simulation based posterior inference

Metropolis–Hastings algorithm

1. Choose a starting point Ψ^0 (usually the posterior mode) and run a loop over 2-3-4.

2. Draw a *proposal* $\Psi^\star$ from a *jumping* distribution

$$J(\Psi^\star|\Psi^{t-1}) = \mathcal{N}(\Psi^{t-1}, c \times \Omega_m)$$

3. Compute the acceptance ratio

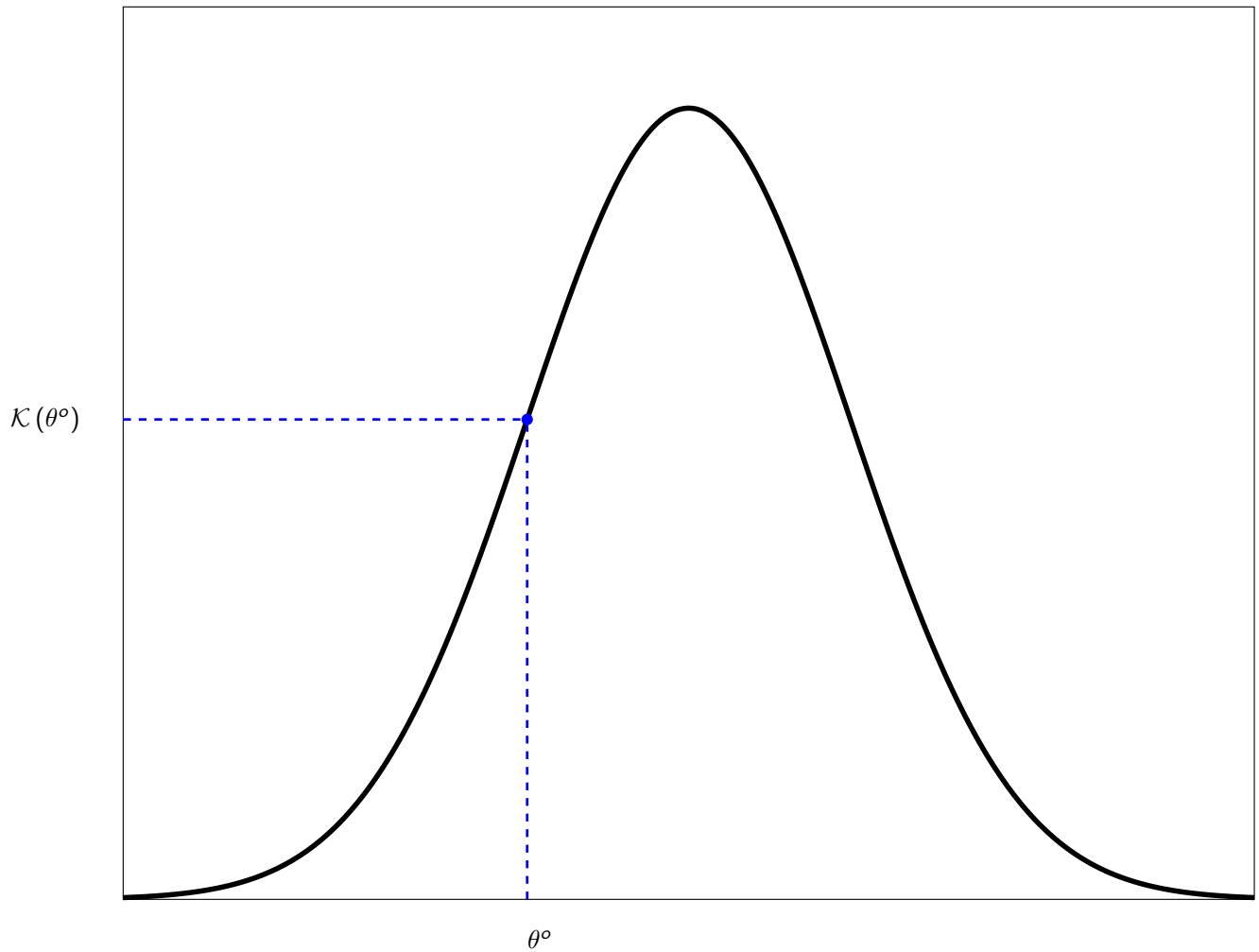
$$r = \frac{p_1(\Psi^\star|\mathcal{Y}_T)}{p_1(\Psi^{t-1}|\mathcal{Y}_T)} = \frac{\mathcal{K}(\Psi^\star|\mathcal{Y}_T)}{\mathcal{K}(\Psi^{t-1}|\mathcal{Y}_T)}$$

4. Finally

$$\Psi^t = \begin{cases} \Psi^\star & \text{with probability } \min(r, 1) \\ \Psi^{t-1} & \text{otherwise.} \end{cases}$$

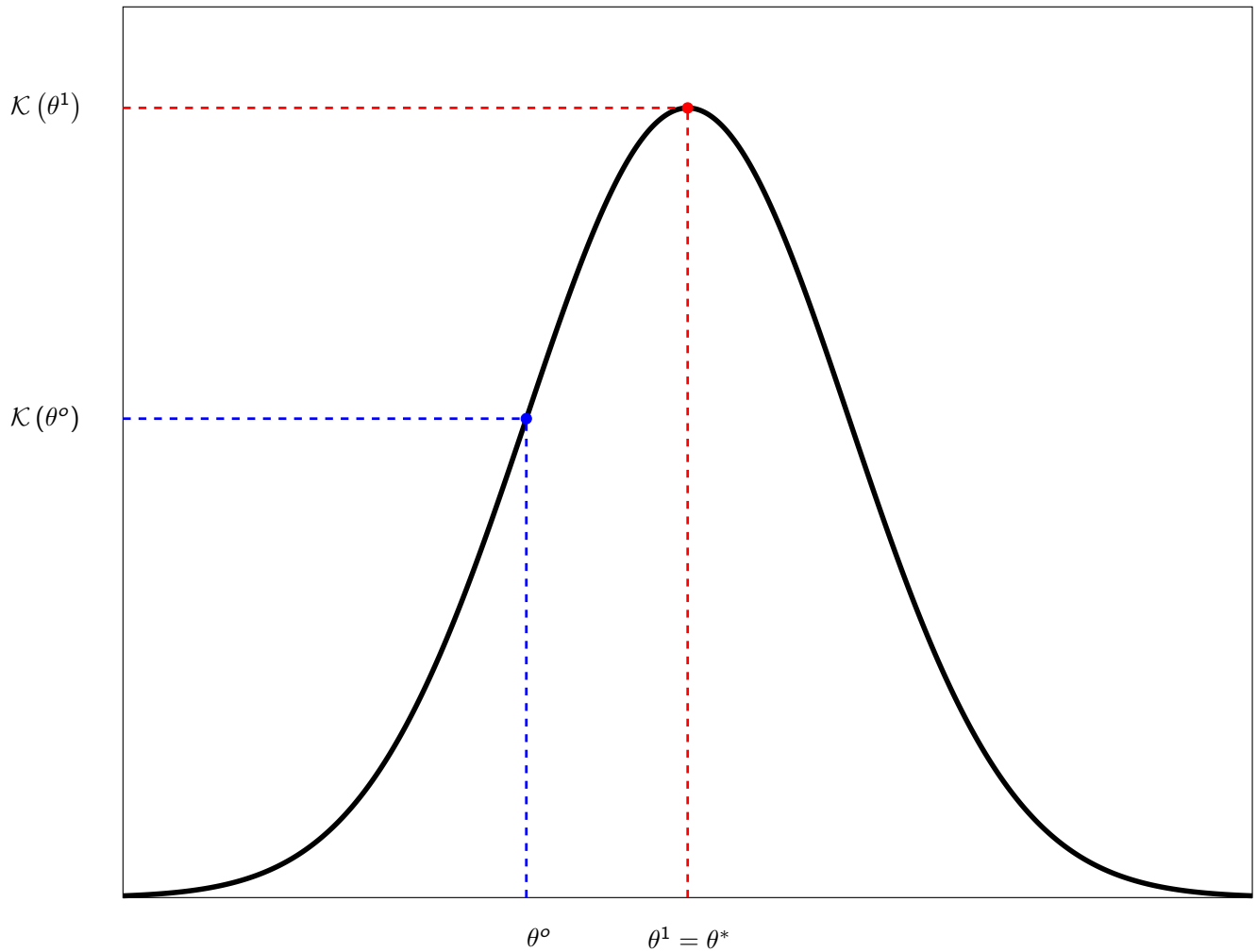
Simulation based posterior inference

Metropolis–Hastings algorithm illustration – a –



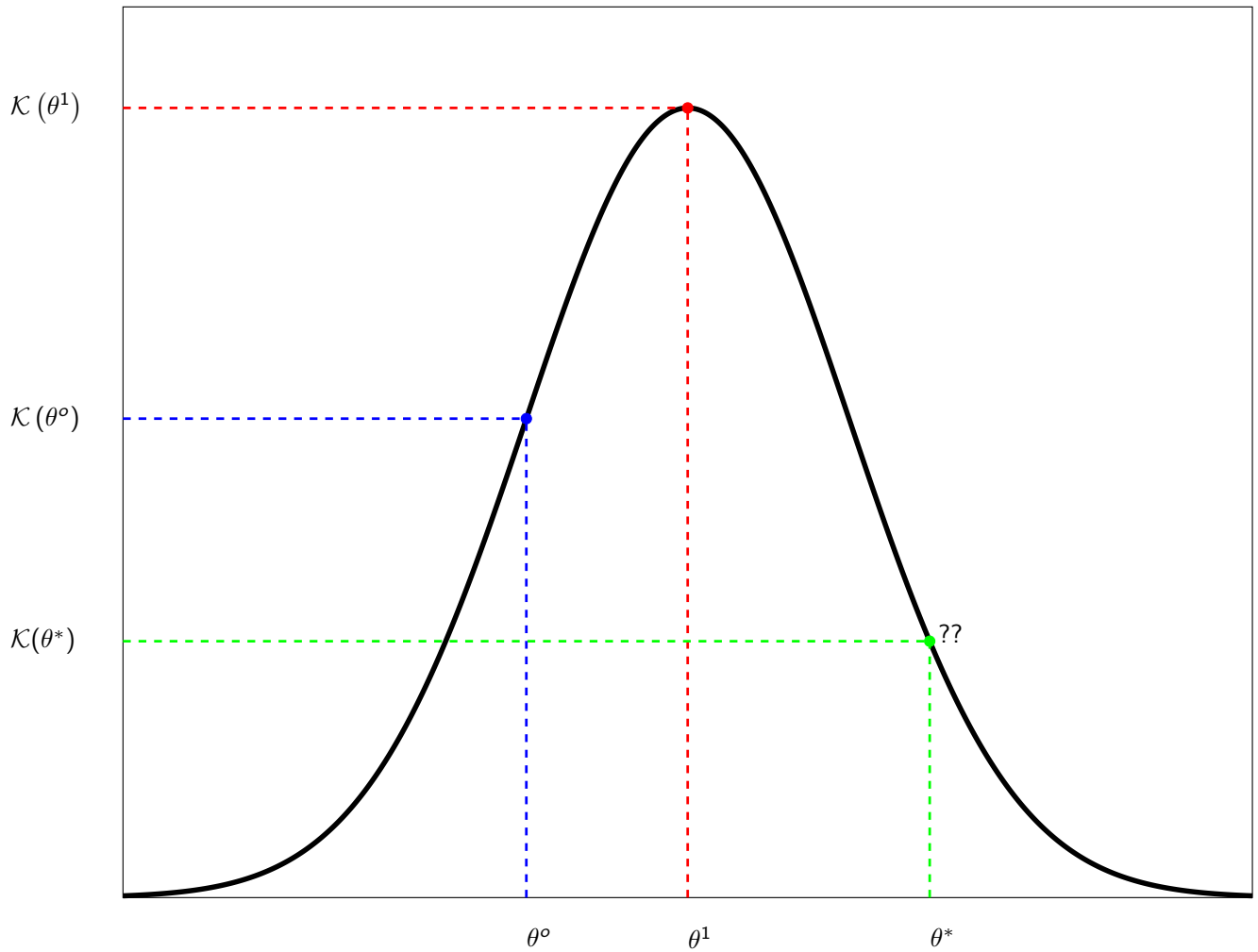
Simulation based posterior inference

Metropolis–Hastings algorithm illustration – b –



Simulation based posterior inference

Metropolis–Hastings algorithm illustration – c –



Simulation based posterior inference

Tuning of the Metropolis–Hastings algorithm

- How should we choose the scale factor c (variance of the jumping distribution)?
- We define the acceptance ratio as the number of accepted draws over the number of proposals:

$$R = \frac{\#\text{ accepted draws}}{\#\text{ proposals}}$$

  - We do not want to have R close to 0 because it means that we reject almost all the proposals.
  - We neither want to have R close to 1 because it means that we accept almost all the proposals and that most likely the jumping distribution proposes too small jumps.
- In the literature, authors generally target an acceptance ratio around one third, although there is no rational for that choice in the case of DSGE models.
- The scale factor, c, must be adjusted to match this target:
  - If R is too low, c should be reduced (why?).
  - If R is too high, c should be increased (why?).

Simulation based posterior inference

Convergence of the Metropolis–Hastings algorithm

- Another issue is the number of draws required to obtain an accurate estimation of the posterior distribution.
- We need to asses the convergence of the Metropolis Hastings algorithm.
- The estimated posterior distribution should be:
  (i) stable when we increase the number of simulations.
  (ii) independent of the initial condition.
- To check that the estimated posterior distribution satisfy these two properties, we can run multiple Monte Carlo Markov Chains and verify that:
  (i) Moments are constant if the number of simulations is increased.
  (ii) Pooled and Within moments are identical.
- This approach is implemented in Dynare (see the manual).

Marginal density estimation

- The marginal density of the sample may be written as:

$$p(\mathcal{Y}_T|\mathcal{A}) = \int_{\Psi_\mathcal{A}} p(\mathcal{Y}_T, \psi_\mathcal{A}|\mathcal{A}) \mathrm{d}\psi_\mathcal{A}$$

- ... or equivalently:

$$p(\mathcal{Y}_T|\mathcal{A}) = \int_{\Psi_\mathcal{A}} \underbrace{p(\mathcal{Y}_T|\Psi_\mathcal{A}, \mathcal{A})}_{\text{likelihood}} \underbrace{p_0(\psi_\mathcal{A}|\mathcal{A})}_{\text{prior}} \mathrm{d}\psi_\mathcal{A}$$

- We face an integration problem.

Marginal density estimation

Asymptotic approximation

- For DSGE models we are unable to compute this integral analytically or with standard numerical tools (because of the curse of dimensionality).

- We assume that the posterior distribution is not too far from a gaussian distribution. In this case we can approximate the marginal density of the sample.

- We have (omitting the conditioning on the model):

$$p(\mathcal{Y}_T) \approx (2\pi)^{\frac{n}{2}} |\mathcal{H}(\psi^*)|^{\frac{1}{2}} p(\mathcal{Y}_T|\psi^*) p_0(\psi^*)$$

- This approach gives accurate estimation of the marginal density if the posterior distribution is uni-modal.

Marginal density estimation

A first simulation based method

- We can estimate the marginal density using a Monte-Carlo approach

$$\widehat{p}(\mathcal{Y}_T) = \frac{1}{B}\sum_{b=1}^{B} p(\mathcal{Y}_T|\psi^{(b)})$$

  where $\psi^{(b)}$ is sampled from the prior distribution.

- $\widehat{p}(\mathcal{Y}_T) \underset{B\to\infty}{\longrightarrow} p(\mathcal{Y}_T)$.

- But this method is highly inefficient, because:
  - $\widehat{p}(\mathcal{Y}_T)$ may have a huge variance (even infinite in some pathological cases).
  - We are not using simulations already done to obtain the posterior distribution (*ie* Metropolis-Hastings draws).

Marginal density estimation

Harmonic mean – a –

- For any probability density function f, we have:

$$\mathbb{E}\left[\frac{f(\psi)}{p_0(\psi)p(\mathcal{Y}_T|\psi)}\middle|\psi\right] = \int_\Psi \frac{f(\psi)p_1(\psi|\mathcal{Y}_T)}{p_0(\psi)p(\mathcal{Y}_T|\psi)}\mathrm{d}\psi$$

- Using the definition of the posterior density:

$$\int_\Psi \frac{f(\psi)}{p_0(\psi)p(\mathcal{Y}_T|\psi)} \frac{p_0(\psi)p(\mathcal{Y}_T|\psi)}{\int_\Psi p_0(\psi)p(\mathcal{Y}_T|\psi)\mathrm{d}\psi}\mathrm{d}\psi$$

- Finally

$$\mathbb{E}\left[\frac{f(\psi)}{p_0(\psi)p(\mathcal{Y}_T|\psi)}\middle|\psi\right] = \frac{\int_\Psi f(\psi)\mathrm{d}\psi}{\int_\Psi p_0(\psi)p(\mathcal{Y}_T|\psi)\mathrm{d}\psi}$$

Marginal density estimation

Harmonic mean – b –

- So that

$$p(\mathcal{Y}_T) = \mathbb{E}\left[\frac{f(\psi)}{p_0(\psi)p(\mathcal{Y}_T|\psi)}\middle|\psi\right]^{-1}$$

- This suggests the following estimator of the marginal density

$$\widehat{p}(\mathcal{Y}_T) = \left[\frac{1}{B}\sum_{b=1}^{B}\frac{f(\psi^{(b)})}{p_0(\psi^{(b)})p(\mathcal{Y}_T|\psi^{(b)})}\right]^{-1}$$

- Each drawn vector $\psi^{(b)}$ comes from the Metropolis - Hastings monte-carlo simulations.

Marginal density estimation

Harmonic mean – c –

- The preceding proof holds if we replace $f(\theta)$ by 1
  $\looparrowright$ Simple Harmonic Mean estimator. But this estimator may also have a huge variance.

- The density $f(\theta)$ may be interpreted as a weighting function, we want to give less importance to extremal values of θ.

- Geweke (1999) suggests to use a truncated gaussian function (modified harmonic mean estimator).

Marginal density estimation

Harmonic mean – d –

$$\overline{\psi} = \frac{1}{B}\sum_{b=1}^{B}\psi^{(b)}$$

$$\overline{\Omega} = \frac{1}{B}\sum_{b=1}^{B}(\psi^{(b)} - \overline{\psi})'(\psi^{(b)} - \overline{\psi})$$

- For some $p \in (0,1)$ we define

$$\widetilde{\Psi} = \left\{\psi : (\psi^{(b)} - \overline{\psi})'\overline{\Omega}^{-1}(\psi^{(b)} - \overline{\psi}) \leq \chi^2_{1-p}(n)\right\}$$

- ... And take

$$f(\psi) = p^{-1}(2\pi)^{-\frac{n}{2}}|\overline{\Omega}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\psi-\overline{\psi})'\overline{\Omega}^{-1}(\psi-\overline{\psi})}\mathbb{I}_{\widetilde{\Psi}}(\psi)$$

Posterior inference

Credible set

- A synthetic way to characterize the posterior distribution is to build something like a confidence interval.

- We define:

$$P(\psi \in C) = \int_C p(\psi) \mathrm{d}\psi = 1 - \alpha$$

  is a $100(1-\alpha)\%$ credible set for ψ with respect to $p(\psi)$ (for instance, with $\alpha = 0.2$ we have a 80% credible set).

- A $100(1-\alpha)\%$ highest probability density (HPD) credible set for ψ with respect to $p(\psi)$ is a $100(1-\alpha)\%$ credible set with the property

$$p(\psi_1) \geq p(\psi_2) \quad \forall \psi_1 \in C \text{ and } \forall \psi_2 \in \bar{C}$$

Posterior inference

Density

- To obtain a view of the posterior distribution we can estimate the marginal posterior densities (for each parameter of the model).

- We use a non parametric estimator:

$$\hat{f}(\psi) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{\psi - \psi^{(i)}}{h}\right)$$

  where N is the number of draws in the metropolis, ψ is a point where we want to evaluate the posterior density, $\psi^{(i)}$ is a draw from the metropolis, $K(\bullet)$ is a kernel (gaussian by default in Dynare) and h is a bandwidth parameter.

- In Dynare the bandwidth parameter is, by default, optimally chosen considering the Silverman's rule of thumb (controling for the repetitions in the MCMC draws).

Posterior inference

Predictive density

- Knowing the posterior distribution of the model's parameters, we can forecast the endogenous variables of the model.

- We define the posterior predictive density as follows:

$$p(\tilde{\mathbf{Y}}|\mathcal{Y}_T) = \int_{\Psi} p(\tilde{\mathbf{Y}}, \psi|\mathcal{Y}_T)\mathrm{d}\psi$$

where, for instance, $\tilde{\mathbf{Y}}$ might be y_{T+1}. Knowing that $p(\tilde{\mathbf{Y}}, \psi|\mathcal{Y}_T) = p(\tilde{\mathbf{Y}}|\psi, \mathcal{Y}_T)p_1(\psi|\mathcal{Y}_T)$ we have:

$$p(\tilde{\mathbf{Y}}|\mathcal{Y}_T) = \int_{\Psi} p(\tilde{\mathbf{Y}}|\psi, \mathcal{Y}_T)p_1(\psi|\mathcal{Y}_T)\mathrm{d}\psi$$

Posterior inference

Predictive density

- In practice, we just have to
  1. sample vectors of parameters from the posterior distribution (using the MCMC draws),
  2. compute the forecast (also IRFs if needed) for each vector of parameters.

- In the end of this process we obtain an empirical posterior distribution for the forecasts (IRFs).

- If our loss function is quadratic, we can then report a point forecast by computing the mean of this empirical posterior distribution.

Posterior inference

Integration

- More generally, the MCMC draws can be used to estimate any moments of the parameters (or function of the parameters).

- We have

$$\begin{aligned} \mathbb{E}\left[h(\psi)\right] &= \int_{\Psi} h(\psi) p(\psi|\mathcal{Y}_T) \mathrm{d}\psi \\ &\approx \frac{1}{N} \sum_{i=1}^{N} h\left(\psi^{(i)}\right) \end{aligned}$$

where $\psi^{(i)}$ is a metropolis draw and h is any continuous function.

Estimation in pratice

- Declare the set of observed variables with `varobs`.
- Declare the priors with the `estimated_params` block (available priors: uniform, gamma, inverse gamma, gaussian, beta and weibull).
- Use the estimation command
  - Estimates the posterior mode (used as an initial condition for MCMC),
  - Runs the MCMC,
  - Computes posterior moments.
- Possible trouble with the first step if the hessian matrix is not positive $\Rightarrow$ The estimated posterior mode is not a mode. $\Rightarrow$ Try other optimization algorithms and/or different initial conditions.