Likelihood-based estimation in Dynare

Willi Mutschler

Dynare Team

Motivation

estimation toolbox provides functionality to estimate parameters by

- Maximum Likelihood
- Bayesian Markov-Chain Monte-Carlo (MCMC)

and then do all sorts of analysis!

Bayesian Primer

Bayesian Primer

- We want to use data to learn about something *unknown* (parameters) given something *known* (data)
- Classical/Frequentist/ML approach:
  - parameters are fixed quantity ("true value")
  - estimating function (e.g. mean) is a random variable
  - estimator is best in the sense of having highest probability of being close to true parameter

Bayesian Primer

- Bayesian approach:
  - parameters are random variables characterized by a probability distribution (posterior = prior x likelihood)
  - prior information: subjective beliefs about how likely different parameter values are (BEFORE seeing data)
  - likelihood: sample information (AFTER seeing data)

Bayesian Primer

- Bayesian simulation methods provide easy way to characterize estimation uncertainty in the form of posterior distributions
- Incorporating non-model information into estimation typically makes estimates more precise and feasible

Bayes' rule

- $p(A, B) = p(A|B)p(B)$ and $p(A, B) = p(B|A)p(A)$ gives:

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

- Also for continuous distributions:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$$

- posterior $p(\theta|y)$ is product of likelihood $p(y|\theta)$ and prior $p(\theta)$ divided by marginal data density $p(y)$

How to compute likelihood?

Linear Gaussian State-Space System

$$y_t = g_x y_{t-1} + g_u u_t \quad \text{[Transition Equation]}$$

$$d_t = \bar{d} + H y_t + e_t \quad \text{[Measurement Equation]}$$

- y_t: model variables in deviation from steady-state (state variables)
- d_t: subset of model variables that are observable (control variables)
- $\bar{d} = H\bar{y}$: steady-state of observable variables
- u_t: shock vector (innovations), $u_t \sim N(0,\Sigma_u)$
- e_t: measurement error (noise), $e_t \sim N(0,\Sigma_e)$

Kalman Filter

- As u_t and e_t are Gaussian, so is y_t and $d_t = \bar{d} + H(g_x y_{t-1} + g_u u_t) + e_t$
- Mean and covariance of d_t depend on mean and covariance of unobserved state variables y_{t-1}, so we cannot directly construct Gaussian likelihood
- Kalman filter is a recursive algorithm that provides "best" estimates for the mean and covariance matrix of unobserved y_t by backing these out from observed data
- At higher-order: particle filter (not easy to handle computationally) or nonlinear Kalman filter (not as general)

What priors?

Prior distributions

- Ideally: probalistic representation of our beliefs on the parameter before seeing the data
- Realistically: informed by some (other) observations or studies
- We need multivariate prior on θ (not on g_x and g_u, these are functions of θ)
- In practice: we specify independent priors for each θ_i; but often parameter range narrows down prior choice

Typical choices

- Normal distribution:
  - unbounded support, symmetric.
  - typically used for feedback parameters or when sign of parameter is unknown.
- Uniform distribution:
  - lower bound and upper bound
  - flat: all points are equally likely
- Beta distribution:
  - bounded support between 0 and 1 (there is also generalized version)
  - often used for autoregressive parameters, discount factor, Calvo probabilities
  - very flexible shape

Typical choices

- Gamma distribution:
  - support: $[0,\infty)$
  - often used for feedback parameters or (large) variances
- Inverse Gamma distribution:
  - support: $(0,\infty)$
  - often used for (small) variances

How to obtain posterior?

$$p(\theta | d^T) \propto p(d^T | \theta) p(\theta)$$

- Posterior is typically not analytical, no closed-form
- But: we are more interested in objects of posterior than posterior itself:

$$E[\theta | d^T] = \int_{-\infty}^{\infty} \theta p(\theta | d^T) d\theta$$

$$V[\theta | d^T] = \int_{-\infty}^{\infty} \theta^2 p(\theta | d^T) d\theta - (E(\theta | d^T))^2$$

- IF we had iid draws from posterior, we could simply use Law of Large Numbers:

$$E[f(\theta) | d^T] = \int_{-\infty}^{\infty} f(\theta) p(\theta | d^T) d\theta \approx \frac{1}{S} \sum_{s=1}^{S} f(\theta_s)$$

Monte Carlo Integration

$$E[f(\theta)\,|\,d^T] = \int_{-\infty}^{\infty} f(\theta)p(\theta\,|\,d^T)d\theta \approx \frac{1}{S}\sum_{s=1}^{S} f(\theta_s)$$

- Monte Carlo integration: Replace integral by sum over S draws from the posterior
- How good is this estimate? Central Limit Theorem gives asymptotic normality for iid draws.
- Posterior sampling algorithms:
  - Direct sampling
  - Importance sampling
  - Metropolis Hastings
  - Gibbs Sampling
  - Sequential Monte-Carlo (Particle)

Metropolis-Hastings

- Key idea: we cannot draw from posterior directly, but we can evaluate posterior distribution
  - draw from a stand-in proposal distribution
  - evaluate both proposal and target density
  - re-weight the draws based on comparison of densities
- Intuition:
  - Jump always uphill
  - Jump downhill with some probability to visit the whole domain of the posterior

Metropolis-Hastings

- Generic algorithm:
  - Start with a vector θ_0
  - For $j = 1,...,S$
    - Generate a candidate $\tilde{\theta}$ from a proposal $q(\theta \mid \theta_{j-1})$, typically $N(\theta_{j-1}, \Sigma)$
    - Calculate acceptance probability: $\alpha = \dfrac{p(\tilde{\theta} \mid d^T)}{p(\theta_{j-1} \mid d^T)}$
    - With probability $min(\alpha,1)$ accept the jump from θ_{j-1} to $\tilde{\theta}$

Metropolis-Hastings

- Generic algorithm (continued)
  - With complementary probability don't accept jump, but:
    - draw uniformly distributed variable r between 0 and 1
    - if $r \leq \alpha$, set $\theta_j = \tilde{\theta}$ (accept)
    - if $r > \alpha$, set $\theta_j = \theta_{j-1}$ (don't accept)

Illustration
Metropolis Hastings

Metropolis-Hastings

- Algorithm returns a so-called Markov Chain $\theta_0, \theta_1, \ldots, \theta_S$
- This is what we need for Monte Carlo integration:

$$E[f(\theta) \,|\, d^T] = \int_{-\infty}^{\infty} f(\theta) p(\theta \,|\, d^T) d\theta \approx \frac{1}{S} \sum_{s=1}^{S} f(\theta_s)$$

- BUT: Only valid for iid draws! Our draws are initially highly correlated, so:
  - Get rid of burnin (say 30-50% of S)
  - Do Monte-Carlo Integration on last 50-70% of draws
  - Test convergence

Metropolis-Hastings

- Test whether draws are iid, i.e. the chain is said to be converged
  - Geweke (1992) convergence test
  - Brooks and Geman (1998) pooled diagnostic
  - Raftery Lewis (1992) diagnostic

Which proposal

- Proposal distribution is key: typically re-centered normal $N(\theta_j, c\Sigma)$ or student's t
- Theoretically: anything goes for Σ, e.g. identity matrix, prior covariance matrix, random numbers; eventually chain will converge no matter what, but speed depends on Σ
- Best practice: try to initialize at the posterior mode (so sampler ist already uphill)
  - run numerical optimization to find the mode (or other pre-sampling to get in the vicinity of the mode)
  - initialize Σ at the inverse hessian (not guaranteed)
  - don't overdo your efforts, just be close to the mode

Scaling Factor

- Typically we scale the proposal covariance matrix to target a certain acceptance rate (i.e. the percentage of times a move is made); optimal values between 25-45%
- We want to visit the whole region of the parameter space
  - If scaling factor too high -> low acceptance probability
  - If scaling factor too low -> high acceptance probability
  - In both cases: highly autocorrelated draws