The Econometrics of DSGE Models

Giuseppe Ragusa
Luiss University

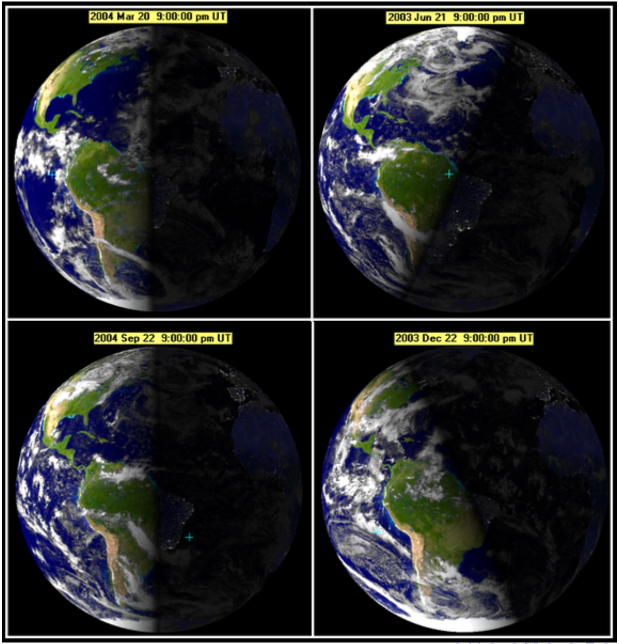
EIEF
Lecture 7: Twilight zone of DSGE models (II)

May 12, 2016

Twilight Zone

*The **twilight zone** or “grey line” is a moving line that separates the illuminated day side and the dark night side of a planetary body*

Twilight Zone



Twilight Zone of DSGE Model

- Hybrid approach to DSGE estimation
- Forecasting performances of DSGE models
- Go beyond the linear approximation
- Identification

Hybrid approach

- The basic idea is to combine an unrestricted VAR and DSGE model
- The combination is indexed by a scalar parameter λ
  - $\lambda \to \infty$ the model converges to the DSGE model
  - $\lambda \to 0$ the model converges to the unrestricted VAR
- The "best" hybrid model is the one associated with the $\hat{\lambda}$, the value of λ the correspond to the highest value of the marginal likelihood.

Hybrid approach

Details of the procedure

- We start with a VAR, whose likelihood is given by

$$L(Y|\Phi,\Sigma)$$

where Φ and Σ are parameters, e.g.

$$Y_t = \Phi Y_{t-1} + u_t, \quad E[u_t u_t'] = \Sigma$$

- Given priors $p(\Phi,\Sigma)$ the posterior is formally given

$$p(\Phi,\Sigma|Y) \propto L(Y|\Phi,\Sigma)p(\Phi,\Sigma)$$

- Suppose now that priors depends on "hyper-parameters"

$$p(\Phi,\Sigma|\theta,\lambda)$$

where θ are the structural parameters of a DSGE model

Hybrid approach

Details of the procedure

- The joint posterior

$$p(\Phi, \Sigma, \theta | Y, \lambda) = \frac{L(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta, \lambda)}{\int_{\Phi,\Sigma} L(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta, \lambda)d(\Phi, \Sigma)}$$

- Notice that

$$p(Y|\theta, \lambda) = \int_{\Phi,\Sigma} L(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta, \lambda)d(\Phi, \Sigma),$$

  is the marginal likelihood, which can be evaluated (under "conditions") analytically for given values of θ and λ.

- We search for the "best" λ

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda} p(Y|\theta, \lambda)$$

Hybrid approach

Details of the procedure

- Conditions:
  - Likelihood:

$$L(Y|\Phi,\Sigma) = \prod_{t=1}^{T} p(Y_t|Y_{t-1},\Phi,\Sigma),$$

    where the $p(\cdot|\cdot,\Phi,\Sigma)'s$ are is normal densities
  - Priors

$$p(\Phi,\Sigma|\theta,\lambda) \propto \underbrace{p(\Phi|\Sigma,\theta,\lambda)}_{\text{Normal}} \; \underbrace{p(\Sigma|\theta,\lambda)}_{\text{Inverted Wishart}}$$

  - The hyperparameters of the prior distribution of Φ and Σ are obtained from the (linearized DSGE model) with (structural) parameter θ
- Under these conditions
  - The posterior can be simulated using the Gibbs sampling (marginals are normal and inverted Wishart)

DSGE models: forecasting performances

Dynamic stochastic general equilibrium (DSGE) models use modern macroeconomic theory:

1. to explain comovements of aggregate time series over the business cycle
2. to forecast future values of economic aggregates
3. to perform policy analysis

How well these models performs in terms of forecast performances?

DSGE models: forecasting performances

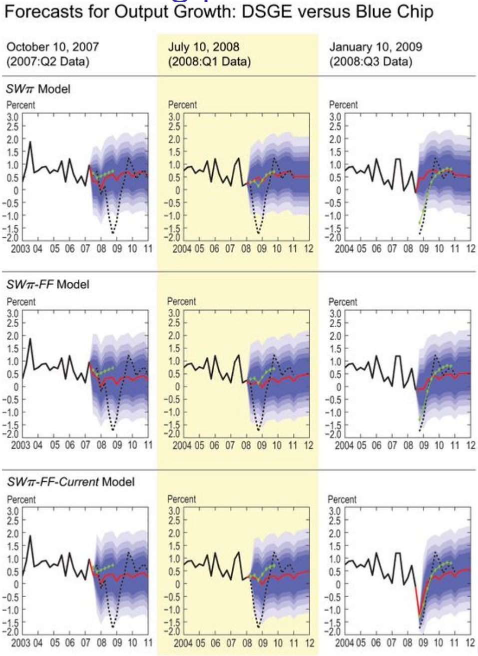
- Dynamic stochastic general equilibrium (DSGE) models have been trashed, bashed, and abused during the Great Recession and after
- One of the many reasons for the bashing was the models' alleged inability to forecast the recession itself.
- Oddly enough, there's little evidence on the forecasting performance of DSGE models during this turbulent period.

DSGE models: forecasting performances

"DSGE Model-Based Forecasting," prepared for Elsevier's Handbook of Economic Forecasting, Del Negro, Schorfheide, Herbst

- Findings:
  - what it really matters what information you feed into your model: Feed in the right information, and even a dingy DSGE model may not do so poorly at forecasting the recession.
  - compared with the "Blue Chip Economic Consensus" forecasts, DSGE models do about the same, if not better, in fall 2007, in summer 2008, before the Lehman crisis, and at the beginning of 2009–provided one incorporates up-to-date financial data into the DSGE model.

DSGE models: forecasting performances



Nonlinearities

- DSGE models are essentially nonlinear representation of our economic reality
- When estimated a DSGE is not actually solved, but rather the solution is approximated by a log-linear function whose coefficients are nonlinear functions of model parameters.
- What passes for the DSGE model is actually the driving processes passed through a filter.
- How accurate is this approximation?

Perturbation Methods

Consider the prototypical DSGE model

$$E_t[f(y_{t+1}, y_t, x_{t+1}, x_t)] = 0$$

where

- x_t: denotes predetermined (or *state*) variables

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$$

  where x_{1t} endogenous ***predetermined*** states and x_{2t} exogenous state
- y_t: denotes predetermined (or *control*) variables
- x_0: the initial condition for the economy is given

Example: Neoclassical growth model

- Households

$$\max E_0 \sum_{t=1}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}, \quad 0 < \beta < 1, \text{ and } \gamma \neq 1$$

- The period-by-period budget constraint

$$A_t k_t^{\alpha} = c_t + k_{t+1} - (1-\delta) k_t$$

  where k_t is the capital stock. In period t the capital is **predetermined.**

- The variable A_t denotes exogenous technological change

$$(A_{t+1} - 1) = \rho (A_t - 1) + \sigma \varepsilon_{t+1}$$

Example: Neoclassical growth model

- The Lagrangian of the household's optimization problem

$$\mathscr{L} = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \frac{c_t^{1-\gamma}}{1-\gamma} + \lambda_t \left[A_t k_t^{\alpha} - c_t - k_{t+1} + (1-\delta) k_t \right] \right\}$$

- The first order optimality conditions are

$$c_t^{-\gamma} = \beta E_t c_{t+1}^{-\gamma} [\alpha A_{t+1} k_{t+1}^{\alpha-1} + 1 - \delta]$$

$$A_t k_t^{\alpha} = c_t + k_{t+1} - (1-\delta) k_t$$

- k_t is endogenously **predetermined**, so it belongs to x_{1t}
- A_t is exogenous state, so it belongs to x_{2t}
- A_t denotes exogenous technological change
- c_t is the control variable

Example: Neoclassical growth model

- Let
$$x_t = \begin{pmatrix} k_t \\ A_t \end{pmatrix}, \quad y_t = c_t$$
then
$$E_t f(y_{t+1}, y_t, x_{t+1}, x_t) = E_t \begin{bmatrix} y_t^{-\gamma} - \beta y_{t+1} \left(\alpha x_{2t+1} x_{1t+1}^{\alpha-1} + 1 - \delta \right) \\ y_t + x_{t+1} - x_{2t} x_{1t}^{\alpha} - (1-\delta) x_{1t} \\ (x_{2t+1} - 1) - \rho (x_{2t} - 1) \end{bmatrix}$$

Policy functions

- The solution to models belonging to the class described previously is given by

$$y_t = g(x_t)$$
$$x_{t+1} = h(x_t) + \eta \sigma \varepsilon_{t+1}$$

- The matrix η is

$$\eta = \begin{bmatrix} 0 \\ \tilde{\eta} \end{bmatrix}$$

  with the dimension of 0 equal to the dimension of x_{1t}.

- That is,

$$y_t = g(x_t)$$
$$x_{1t+1} = h_1(x_t)$$
$$x_{2t+1} = h_2(x_t) + \tilde{\eta} \sigma \varepsilon_{t+1}$$

Perturbation methods

- The key idea of perturbation methods is to express the solution as function of the state vector **and** of σ, the amount of uncertainty in the economy

$$y_t = g(x_t, \sigma)$$
$$x_{t+1} = h(x_t, \sigma) + \eta \sigma \varepsilon_{t+1}$$

- Given this interpretation, a perturbation methods finds *a local* approximation of the functions g and h
- By local approximation, we mean an approximation that is valid at a particular point $(\bar{x}, \bar{\sigma})$

Perturbation methods

Taking a series approximation of of the function g and h around this point we have

$$\begin{aligned} g(x,\sigma) &= g(\bar{x},\bar{\sigma}) + g_x(\bar{x},\bar{\sigma})(x-\bar{x}) + g_\sigma(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma}) \\ &+ \frac{1}{2}g_{xx}(\bar{x},\bar{\sigma})(x-\bar{x})^2 + \frac{1}{2}g_{\sigma\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})^2 \\ &+ \frac{1}{2} + g_{x\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})(x-\bar{x}) + \cdots \end{aligned}$$

and

$$\begin{aligned} h(x,\sigma) &= h(\bar{x},\bar{\sigma}) + h_x(\bar{x},\bar{\sigma})(x-\bar{x}) + h_\sigma(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma}) \\ &+ \frac{1}{2}h_{xx}(\bar{x},\bar{\sigma})(x-\bar{x})^2 + \frac{1}{2}h_{\sigma\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})^2 \\ &+ \frac{1}{2} + h_{x\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})(x-\bar{x}) + \cdots \end{aligned}$$

Perturbation methods

- Let

$$\begin{aligned} F(x,\sigma) &= E_t f(y_{t+1}, y_t, x_{t+1}, x_t) \\ &= E_t f(g(h(x,\sigma)+\eta\sigma\varepsilon'), g(x,\sigma), h(x,\sigma)+\eta\sigma\varepsilon', x) \end{aligned}$$

- Notice that

$$F(x,\sigma) = 0$$

  for all x and σ

- This simple results, together with the fact that

$$\frac{\partial F(x,\sigma)}{\partial^i x \partial^j \sigma} = 0, \quad \forall x, \sigma, i, j$$

  is used to identify $h_x, g_x, h_\sigma, g_\sigma, h_{\sigma x}, g_{\sigma x}, \ldots$

Perturbation methods

- What is a good local point? The steady state of the economy, which coincides with the case $\sigma = 0$
- Notice that at the deterministic steady-state of the economy ($\sigma = 0$)

$$g(\bar{x}, 0) = \bar{y}$$
$$h(\bar{x}, 0) = \bar{x}$$

- The reason why the steady state is particularly convenient is that in most cases is possible to solve for the steady state
- With the steady state values at hand, you can find the derivatives of F

Perturbation methods - linear case

- If we limit ourselves to the linear approximation of the policy functions, we have

$$g(x,0) = g(\bar{x},0) + g_x(\bar{x},0)(x-\bar{x}) + g_\sigma(\bar{x},0)\sigma$$
$$h(x,0) = h(\bar{x},0) + h_x(\bar{x},0)(x-\bar{x}) + h_\sigma(\bar{x},0)\sigma$$

- We know that

$$g(\bar{x},0) = \bar{y}, \quad h(\bar{x},0) = \bar{x}$$

- The remaining unknown coefficient are the four first derivatives: $g_x(\bar{x},0)$, $h_x(\bar{x},0)$, $g_\sigma(\bar{x},0)$, $h_\sigma(\bar{x},0)$
- We have

$$F_\sigma(\bar{x},0) = f_{y'}[g_x h_\sigma + g_\sigma] + f_y g_\sigma + f_{x'} h_\sigma$$
$$F_x(\bar{x},0) = f_{y'} g_x h_x + f_y g_x + f_{x'} h_x + f_x$$

Perturbation methods - linear case

- Solving the linear system of equation we have

$$0 = F_\sigma(\bar{x}, 0) \implies [f_{y'} g_x + f_{x'} \quad f_{y'} + f_y] \begin{pmatrix} h_\sigma \\ g_\sigma \end{pmatrix} = 0$$

- This is *linear* and *homogenous* in g_σ and h_σ system of equations
- If a unique solution exists, we have that

$$h_\sigma = 0, \quad g_\sigma = 0$$

- **Important theoretical results: in general, up to first order, one need not to correct the constant term of the approximation of the policy function for the size of the variance of the shocks**
- In the linear approximation, certainty equivalence holds — the policy function is independent of of the variance covariance matrix of ε_t

Perturbation methods - linear case

- To find g_x and h_x, observe that

$$[f_{x'} \quad f_{y'}] \begin{pmatrix} I \\ g_x \end{pmatrix} h_x = -[f_x \quad f_y] \begin{pmatrix} I \\ g_x \end{pmatrix}$$

- Let $A = [f_{x'} \quad f_{y'}]$ and $B = -[f_x \quad f_y]$. Both A and B are known
- Let P the matrix of eigenvector of h_x be such

$$h_x P = P\Lambda$$

where Λ is diagonal. Let also

$$Z = \begin{pmatrix} I \\ g_x \end{pmatrix} P$$

- Then, from

$$[f_{x'} \quad f_{y'}] \begin{pmatrix} I \\ g_x \end{pmatrix} h_x P = -[f_x \quad f_y] \begin{pmatrix} I \\ g_x \end{pmatrix} P,$$

it follows

$$AZ\Lambda = BZ$$

Perturbation methods - linear case

- We can map the above problem into a generalized eigenvalue problem

$$AZ\Lambda = BZ$$

- A and B can be written as (in Julia `V,D = eig(B,A)`)

$$AV\Lambda = BV$$

then

$$A[V_1 \quad V_2]\begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} = B[V_1 \quad V_2]$$

- Comparing terms,

$$\Lambda = D_{11}$$

and

$$\begin{pmatrix} I \\ g_x \end{pmatrix} P = V_1 \equiv \begin{bmatrix} V_{11} \\ V_{12} \end{bmatrix}$$

- Thus,

$$h_x = V_{11} D_{11} V_{11}^{-1}$$

$$g_x = V_{12} V_{11}^{-1}$$

Perturbation methods - linear case

- Problem with first order approximation
  - In many economic application we are interested in finding the effect of uncertainty on the economy
    - e.g., up to first order, the mean of the rate of return of all assets must be the same; we cannot study risk premia with linear approximated models
    - e.g., how uncertainty affect welfare cannot be studied with linear approximation; any two policies that give rise to the same steady state yield, up to first order, the same level of welfare

Second order perturbation methods

Consider approximating g and h up to second order up to a point $(\bar{x}, \bar{\sigma})$:

$$\begin{aligned}
g(x,\sigma) &= g(\bar{x},\bar{\sigma}) + g_x(\bar{x},\bar{\sigma})(x-\bar{x}) + g_\sigma(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma}) \\
&\quad + \frac{1}{2} g_{xx}(\bar{x},\bar{\sigma})(x-\bar{x})^2 + \frac{1}{2} g_{\sigma\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})^2 \\
&\quad + \frac{1}{2} g_{x\sigma}(\bar{x},\bar{\sigma})(x-\bar{x})(\sigma-\bar{\sigma}) + \frac{1}{2} g_{\sigma x}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})(x-\bar{x})
\end{aligned}$$

and

$$\begin{aligned}
h(x,\sigma) &= h(\bar{x},\bar{\sigma}) + h_x(\bar{x},\bar{\sigma})(x-\bar{x}) + h_\sigma(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma}) \\
&\quad + \frac{1}{2} h_{xx}(\bar{x},\bar{\sigma})(x-\bar{x})^2 + \frac{1}{2} h_{\sigma\sigma}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})^2 \\
&\quad + \frac{1}{2} h_{x\sigma}(\bar{x},\bar{\sigma})(x-\bar{x})(\sigma-\bar{\sigma}) + \frac{1}{2} h_{\sigma x}(\bar{x},\bar{\sigma})(\sigma-\bar{\sigma})(x-\bar{x})
\end{aligned}$$

Note: This notation holds only for the univariate case. For the general case the notation is messier.

Second order perturbation methods

The unknown are

$$\begin{aligned} h_{xx}(\cdot,\cdot), &\quad g_{xx}(\cdot,\cdot) \\ h_{x\sigma}(\cdot,\cdot), &\quad g_{x\sigma}(\cdot,\cdot) \\ h_{\sigma x}(\cdot,\cdot), &\quad g_{\sigma x}(\cdot,\cdot) \\ h_{\sigma\sigma}(\cdot,\cdot), &\quad g_{\sigma\sigma}(\cdot,\cdot) \end{aligned}$$

These coefficient can be identified by taking the derivative of $F(x,\sigma)$ with respect to x and σ and evaluating them at $(\bar{x},0)$

Second order perturbation methods

It can be easily shown that the h_{xx}, g_{xx}, $h_{\sigma\sigma}$, and $g_{\sigma\sigma}$ are the solution to systems of linear equation
Similarly with the first order perturbation methods, it can be shown that

$$g_{\sigma x}(\bar{x}, 0) = 0$$

and

$$h_{\sigma x}(\bar{x}, \sigma) = 0$$

These result represent an important theoretical results.

> *Up to second order, the coefficient of the policy function on the terms that are linear in the state vector do not depend on the size of the variance of the underlying shocks.*

Second order perturbation methods

Another way to state this result is the following:

$$g(x,\sigma) = g(\bar{x},0) + g_x(\bar{x},0)(x-\bar{x}) + \frac{1}{2}g_{xx}(\bar{x},0)(x-\bar{x})^2 + \frac{1}{2}g_{\sigma\sigma}(\bar{x},\bar{\sigma})\sigma^2$$

$$h(x,\sigma) = h(\bar{x},0) + h_x(\bar{x},0)(x-\bar{x}) + \frac{1}{2}h_{xx}(\bar{x},0)(x-\bar{x})^2 + \frac{1}{2}h_{\sigma\sigma}(\bar{x},\bar{\sigma})\sigma^2$$

Up to second order the variance affect of σ on the determinist version of the model is

$$\frac{1}{2}g_{\sigma\sigma}(\bar{x},\bar{\sigma})\sigma^2 \quad \text{and} \quad \frac{1}{2}h_{\sigma\sigma}(\bar{x},\bar{\sigma})\sigma^2$$

which can be regarded as constant term.

Second order perturbation methods

Nice things about second order perturbation methods:

- The second order solution captures at least the constant effect of the variance on the policy function
- It involves only linear operations

Not so nice things:

- The coefficients of the policy function do not respond to the variance of the shocks

N-th perturbation methods

- It is straightforward to apply the method described so far to finding higher than second-order approximation to the policy function
  - Given the first- and second-order terms in the Taylor expansion of h and g, the **third** order terms can be identified by solving system of linear equations.
  - Given the first- and second- and third-order terms in the Taylor expansion of h and g, the **fourth** order terms can be identified by solving system of linear equations.
  - ......

Estimate DSGE models with higher order term

Given

$$y_t = \underbrace{g(x_t, \sigma; \theta)}_{\text{replace approximation}} + u_t$$

$$x_{t+1} = \underbrace{h(x_t, \sigma; \theta)}_{\text{replace approximation}} + \eta \sigma \varepsilon_{t+1}$$

There are three main solutions to estimate θ (structural parameter!)

- Extended Kalman Filter
- Unscented Kalman Filter
- Particle Filter (generically Sequential Monte Carlo methods, SMC)

SMC

Needed:

1. SMC assume x_t and the observations y_t can be modeled in this form:
2. $x_0, x_1, \ldots$ is a first order Markov process such that

$$x_t | x_{t-1} = x \sim f_\theta(\cdot | x)$$

   and with an initial distribution $p(x_0)$.
3. The observations $y_0, y_1, \ldots$ are conditionally independent (conditionally on x_t) and their marginal probability density function is

$$y_t | x_1, x_2, \ldots, x_t = x, \ldots, x_T \sim g_\theta(\cdot | x)$$

SMC

Sequential MC are known in the literature as:

- bootstrap filtering
- condensation algorithm
- particle filtering
- interacting particle approximation
- survival of the fittest

Particle Filter in economics

- Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez. "Estimating macroeconomic models: A likelihood approach." The Review of Economic Studies 74.4 (2007): 1059-1087.
- Flury, Thomas, and Neil Shephard. "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models." Econometric Theory 27.5 (2011): 933.
- Fernandez-Villaverde, Jesus. "The econometrics of DSGE models." SERIEs 1.1-2 (2010): 3-49.

SMC

The aim is to perform Bayesian inference on θ which is based on the likelihood

$$p_\theta(x_{1:T}, y_{1:T}) = \mu(x_0) \prod_{t=1}^{T} f_\theta(x_t|x_{t-1}) \prod_{t=1}^{T} g_\theta(y_t|x_t),$$

which given a prior $p(\theta)$ gives the posterior distribution

$$p(\theta, x_{1:T}|y_{1:T}) \propto p_\theta(x_{1:T}, y_{1:T}) p(\theta).$$

- For nonlinear models, $p_\theta(x_{1:T}, y_{1:T})$ and $p(\theta, x_{1:T}|y_{1:T})$ do not usually admit closed form expressions, making inference in difficult to perform in practice.
- It is thus necessary to resort to approximations
  - SMC methods are a class of algorithms to approximate sequentially the sequence of distributions $p_\theta(x_{1:T}|y_{1:T}) \propto p_\theta(x_{1:T}, y_{1:T})$

The basic sequential Monte Carlo algorithm

1. *Step 1:* Approximate $p_\theta(x_1|y_1)$

- Use the importance density $q_\theta(x_1|y_1)$ and use Importance Sampling
  - generate x_1^k, $k = 1, \ldots, N$ (these are the particles) and ascribe weights w_1^k which take into account the discrepancy between q_θ and p_θ

$$w_1^k = \frac{p_\theta(y_1|x_1^k)p(x_1^k)}{q_\theta(x_1^k|y_1)}, \quad k = 1, \ldots, K$$

  and

$$W_1^k = \frac{w_1^k}{\sum_{k=1}^K w_1^k}, \quad k = 1, \ldots K$$

- Notice that

$$\frac{1}{K}\sum_{k=1}^{K} w_1^k \xrightarrow[K\to\infty]{p} \int \frac{p_\theta(x_1, y_1)}{q_\theta(x_1|y_1)} q_\theta(x_1|y_1) dx_1 = p_\theta(y_1)$$

and for $\delta_{x_1^k} = 1(x - x_1^k)$

$$\hat{p}_\theta(x_1|y_1) := \sum_{k=1}^{K} W_1^k \delta_{x_1^k} \xrightarrow{p} \int \delta_{x_1^k} \frac{p_\theta(x_1^k, y_1)}{p_\theta(y_1)} dx_1^k = p_\theta(x_1 = x|y_1)$$

Identity

The "identity"

$$
\begin{aligned}
p_\theta(x_{1:2}|y_{1:2}) &= \frac{p_\theta(y_1, y_2, x_1, x_2)}{r_\theta(y_{1:2})} \\
&\propto p_\theta(y_1, y_2, x_1, x_2) \\
&= r_\theta(y_2|y_1, x_1, x_2) r_\theta(y_1|x_1, x_2) r_\theta(x_1, x_2) \\
&= g_\theta(y_2|x_2) r_\theta(y_1|x_1) r(x_1) f_\theta(x_2|x_1) \\
&= g_\theta(y_2|x_2) f_\theta(x_2|x_1) \left[\frac{p_\theta(x_1|y_1) r_\theta(y)}{r_\theta(x_1)} r(x_1) \right] \\
&\propto g_\theta(y_2|x_2) f_\theta(x_2|x_1) p_\theta(x_1|y_1)
\end{aligned}
$$

The basic sequential algorithm

1. *Step 2: Use $\tilde{x}_1^k$, which are distributed according to $p_\theta(x_1|y_1)$, to sample from $p_\theta(x_1|y_1)f_\theta(x_1|x_2)g_\theta(y_2|x_2)$*
   - New importance density $q_\theta(x_2|x_1, y_2)$
   - Sample indexes A_1^k from W_1^k
   - $x_2^k \sim q_\theta(x_2|x_1^{A_1^k}, y_2), \quad k = 1, \dots, K$, set $x_{1:2}^k = (x_1^k, x_2^k)$ and

$$
\begin{aligned}
w_2^k &= \frac{p_\theta(x_{1:2}^k, y_{1:2})}{p_\theta(x_1^k, y_1)q(x_2^k|y_2, x_1^k)} \\
&= \frac{f_\theta(x_2^k|x_1^k)g_\theta(y_2|x_2^k)p_\theta(y_1|x_1^k)p(x_1^k)}{p_\theta(y_1|x_1^k)p(x_1^k)q(x_2^k|y_2, x_1^k)} \\
&= \frac{f_\theta(x_2^k|x_1^k)g_\theta(y_2|x_2^k)}{q(x_2^k|y_2, x_1^k)}
\end{aligned}
$$

   and

$$
W_2^k = \frac{w_t^k}{\sum_{k=1}^K w_t^k}, \quad k = 1, \dots, K
$$

The basic sequential algorithm

- Notice that

$$
\begin{aligned}
\frac{1}{K}\sum_{k=1}^{K} w_2^k \xrightarrow[K\to\infty]{p} \int \frac{f_\theta(x_2|x_1)g_\theta(y_2|x_2)}{q(x_2|y_2,x_1)} q_\theta(x_2|y_2,x_1)p_\theta(x_1|y_1)dx_2dx_1 \\
&= \int f_\theta(x_2|x_1)g_\theta(y_2|x_2)dx_2 p_\theta(x_1|y_1)dx_1 \\
&= \int g_\theta(y_2|x_1)p_\theta(x_1|y_1)dx_1 \\
&= p_\theta(y_2|y_1)
\end{aligned}
$$

The basic sequential algorithm

The SMC sequentially "estimates"

$$p_\theta(x_{1:t}|y_{1:t}), \quad t = 1, \dots, T$$

but also provides an approximation to

$$p_\theta(y_t|y_{1:t-1}), \qquad t = 1, \dots, T$$

which in turns provides

$$p_\theta(y_{1:T}) = p(y_1) \prod_{t=2}^{T} p_\theta(y_t|y_{t-1})$$

Can be estimated

$$\hat{p}_\theta(y_{1:T}) = \hat{p}_\theta(y_1) \prod_{t=2}^{T} \hat{p}_\theta(y_t|y_{t-1})$$

Alternative estimation

- Indirect Inference
- Approximate Bayesian Computation (ABC)
- Going back to drawing board (Gallant, Giacomini, Ragusa; 2013)