Multi-resolution Inference of Stochastic Models from Partially Observed Data

Samuel Kou
Department of Statistics
Harvard University

Joint with Ben Olding

Stochastic Differential Equations

- Models based on stochastic differential equations (SDE) are widely used in science and engineering.
- General form

$$dY_t = \mu\left(Y_t, \boldsymbol{\theta}\right) dt + \sigma\left(Y_t, \boldsymbol{\theta}\right) dB_t$$

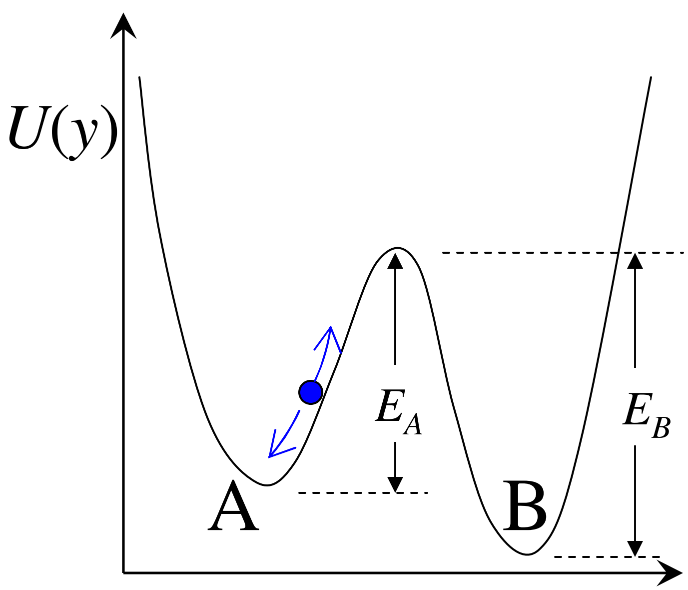
Y_t : the process, $\boldsymbol{\theta}$: the underlying parameters

Stochastic Differential Equations

- Models based on stochastic differential equations (SDE) are widely used in science and engineering.
- General form

$$dY_t = \mu\left(Y_t, \boldsymbol{\theta}\right) dt + \sigma\left(Y_t, \boldsymbol{\theta}\right) dB_t$$

Y_t : the process, $\boldsymbol{\theta}$: the underlying parameters



y: reaction coordinate

- Example 1. In chemistry and biology.

A reversible enzymatic reaction **A↔B** typically modeled as

$$dY_t = -U'(Y_t)dt + \sigma dB_t$$

$\boldsymbol{\theta}$: the energy barrier heights etc.

Stochastic Differential Equations

- Models based on stochastic differential equations (SDE) are widely used in science and engineering.
- General form

$$dY_t = \mu\left(Y_t, \boldsymbol{\theta}\right) dt + \sigma\left(Y_t, \boldsymbol{\theta}\right) dB_t$$

  Y_t : the process, $\boldsymbol{\theta}$: the underlying parameters

- Example 2. In finance and economics.

  Feller process (a.k.a. CIR process) has been used to model interest rates

$$dY_t = \gamma(\mu - Y_t)dt + \sigma\sqrt{Y_t}\,dB_t$$

parameters

Stochastic Differential Equations

- Models based on stochastic differential equations (SDE) are widely used in science and engineering.
- General form

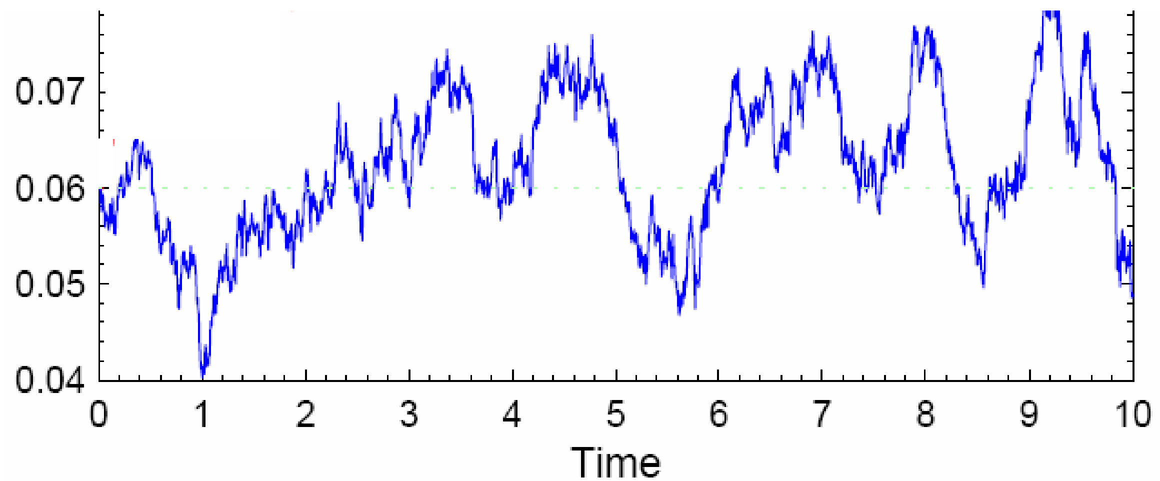
$$dY_t = \mu\left(Y_t, \boldsymbol{\theta}\right) dt + \sigma\left(Y_t, \boldsymbol{\theta}\right) dB_t$$

Y_t : the process, $\boldsymbol{\theta}$: the underlying parameters

- Example 2. In finance and economics.

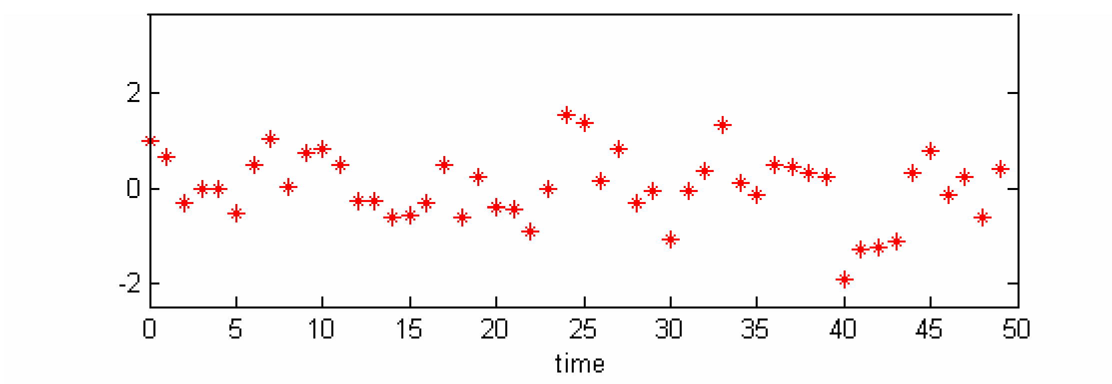
  Feller process (a.k.a. CIR process) has been used to model interest rates

$$dY_t = \gamma(\mu - Y_t)dt + \sigma\sqrt{Y_t}\,dB_t$$



Statistical Inference

- Given a stochastic model, infer the parameter values from data
- Major complication: the continuous-time model is only observed at discrete time points
  Example: (i) Biology or chemistry experiments can track movement of molecules only at discrete camera frames
  (ii) Finance or economics, interest rates, price index, etc. only observed daily, weekly or monthly



Likelihood Inference

- Data $(Y_1,t_1), (Y_2,t_2),\ldots(Y_n,t_n)$ from $dY_t = \mu\left(Y_t,\boldsymbol{\theta}\right)dt + \sigma\left(Y_t,\boldsymbol{\theta}\right)dB_t$
- Likelihood

$$L(\mathbf{Y}|\boldsymbol{\theta}) = f(Y_1|\boldsymbol{\theta})\prod_i f(Y_{i+1}|Y_i,\triangle t_i,\boldsymbol{\theta}), \qquad \triangle t_i = t_{i+1} - t_i$$

$f(y|x,t,\theta)$: transition density

$$\frac{\partial f(y|x,t,\boldsymbol{\theta})}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y,\boldsymbol{\theta})f(y|x,t,\boldsymbol{\theta})] - \frac{\partial}{\partial y}[\mu(y,\boldsymbol{\theta})f(y|x,t,\boldsymbol{\theta})]$$

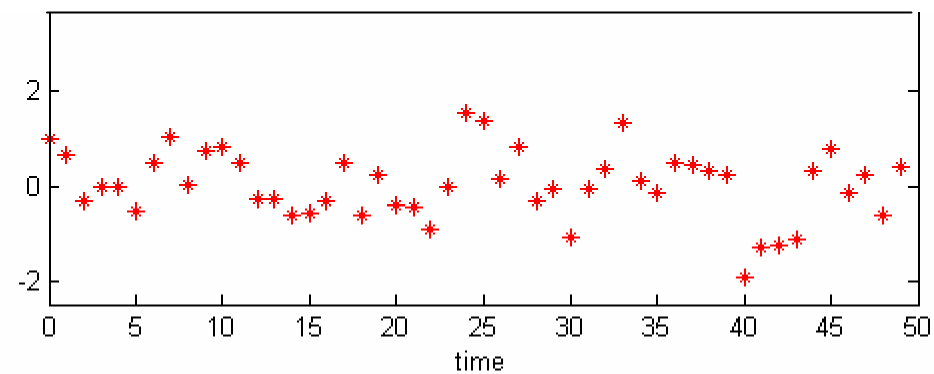
- In most cases, f does not permit analytical form; solving a PDE numerically is not feasible either

The Euler Approximation

- Idea: approximate an SDE $dY_t = \mu\left(Y_t, \boldsymbol{\theta}\right) dt + \sigma\left(Y_t, \boldsymbol{\theta}\right) dB_t$ by a *difference* equation

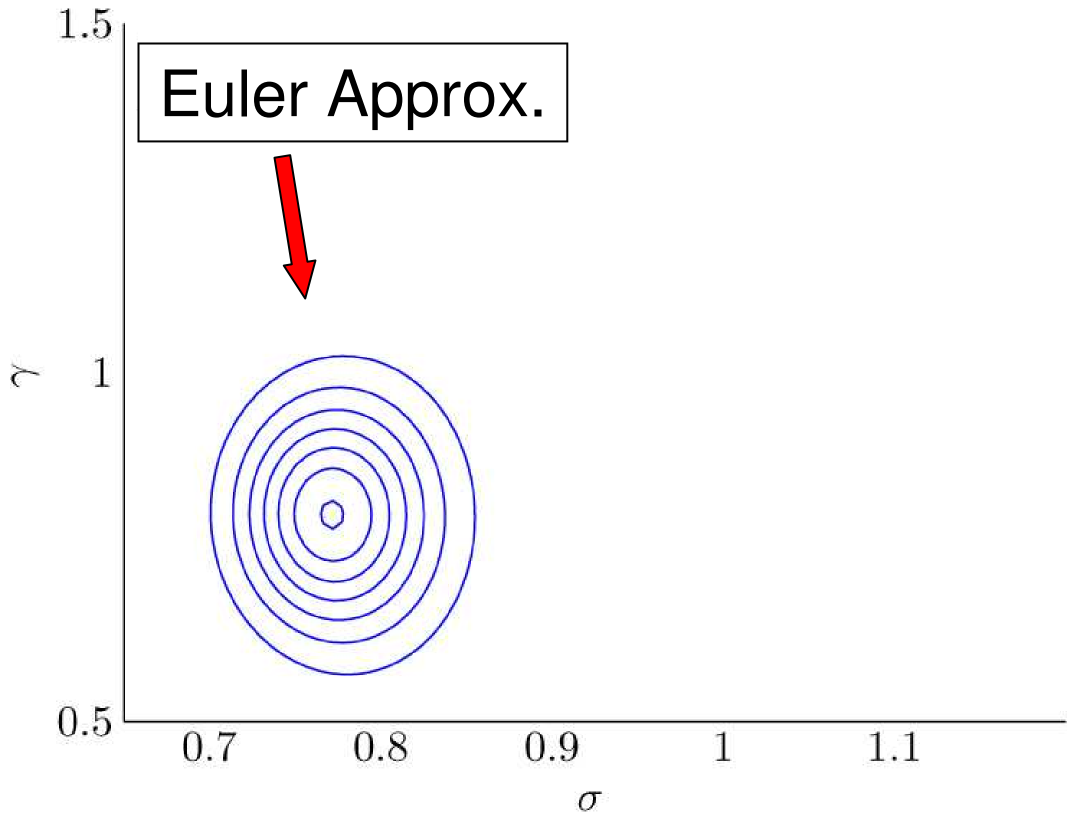
$$Y(t + \Delta t) - Y(t) = \mu(Y_t, \boldsymbol{\theta})\Delta t + \sigma(Y_t, \boldsymbol{\theta})\sqrt{\Delta t} Z_t$$

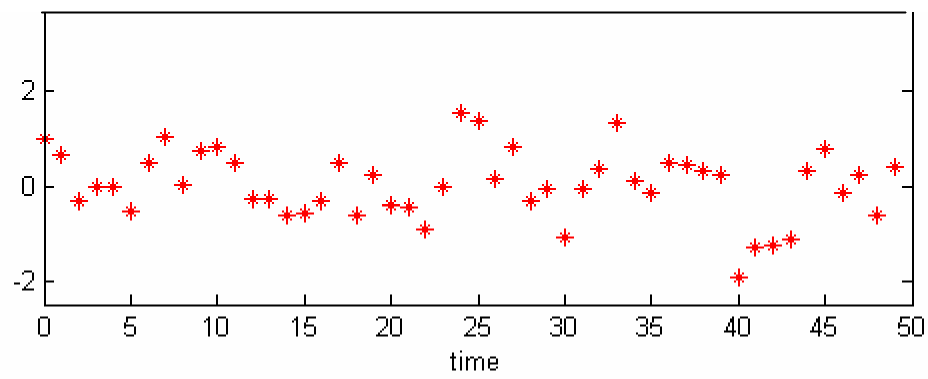
- Obtain approx likelihood from the difference eqn
- Works well *only* if Δt is small



Generated from
Ornstein-Uhlenbeck process

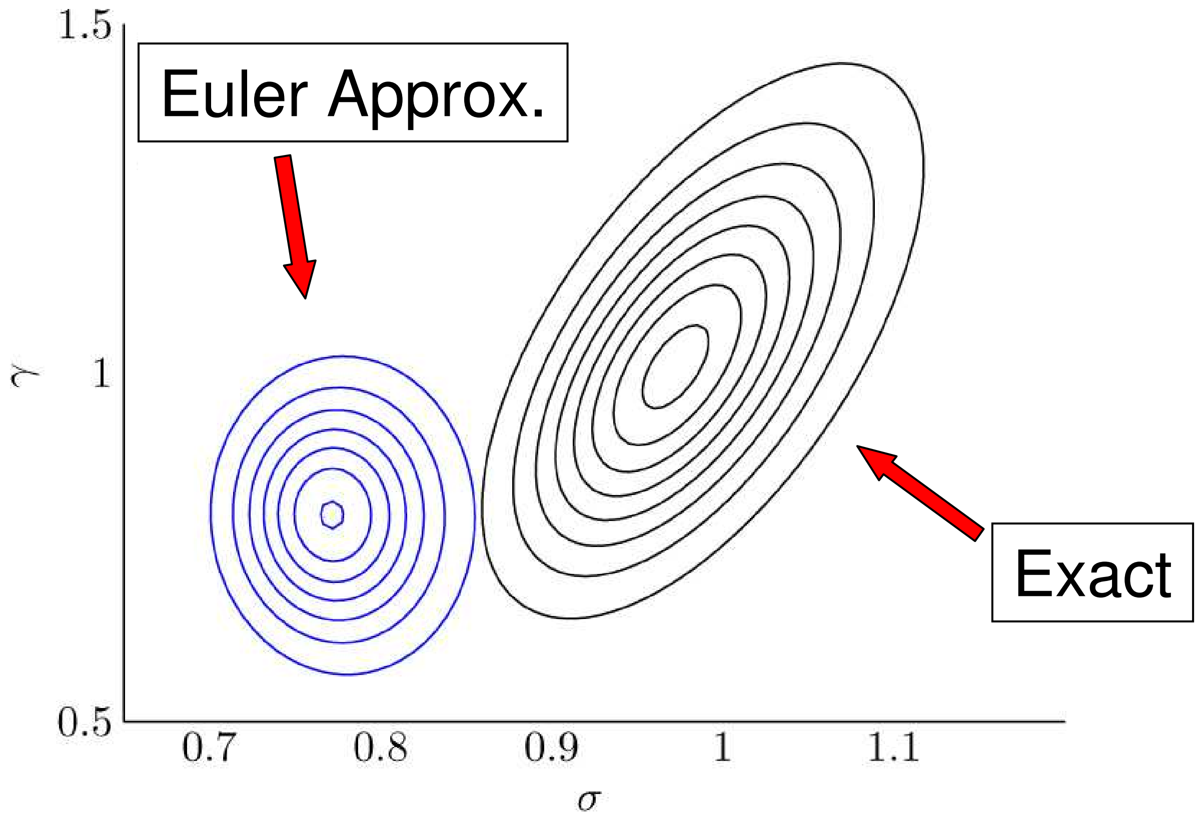
$$dY_t = \left(\mu - Y_t\right)\gamma dt + \sigma dB_t$$

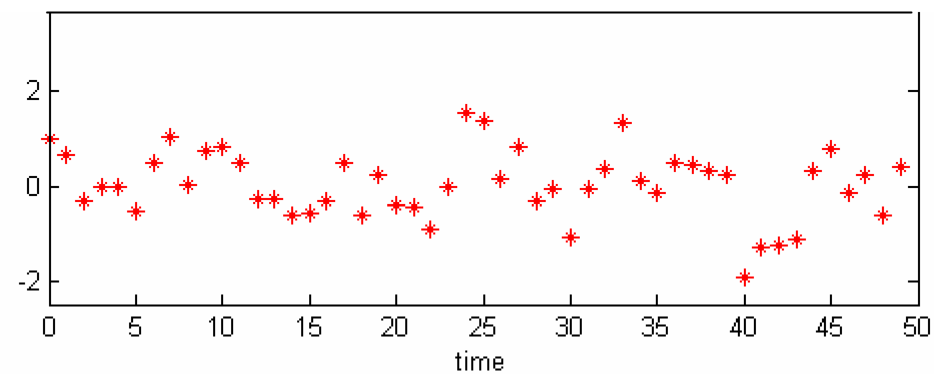




Generated from
Ornstein-Uhlenbeck process

$$dY_t = (\mu - Y_t)\,\gamma dt + \sigma dB_t$$



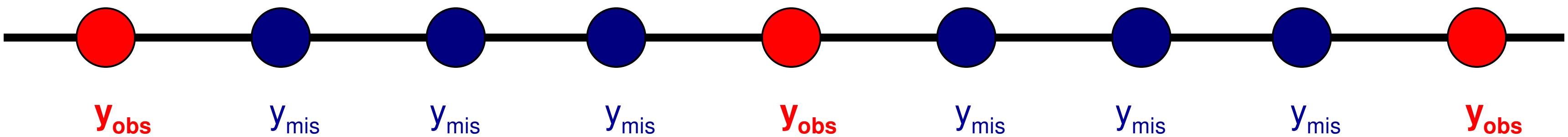


Generated from
Ornstein-Uhlenbeck process

$$dY_t = (\mu - Y_t)\,\gamma dt + \sigma dB_t$$

Bayesian Data Augmentation

- If Δt is not "sufficiently small"
  - Choose a Δt small enough so that the Euler approximation is appropriate.



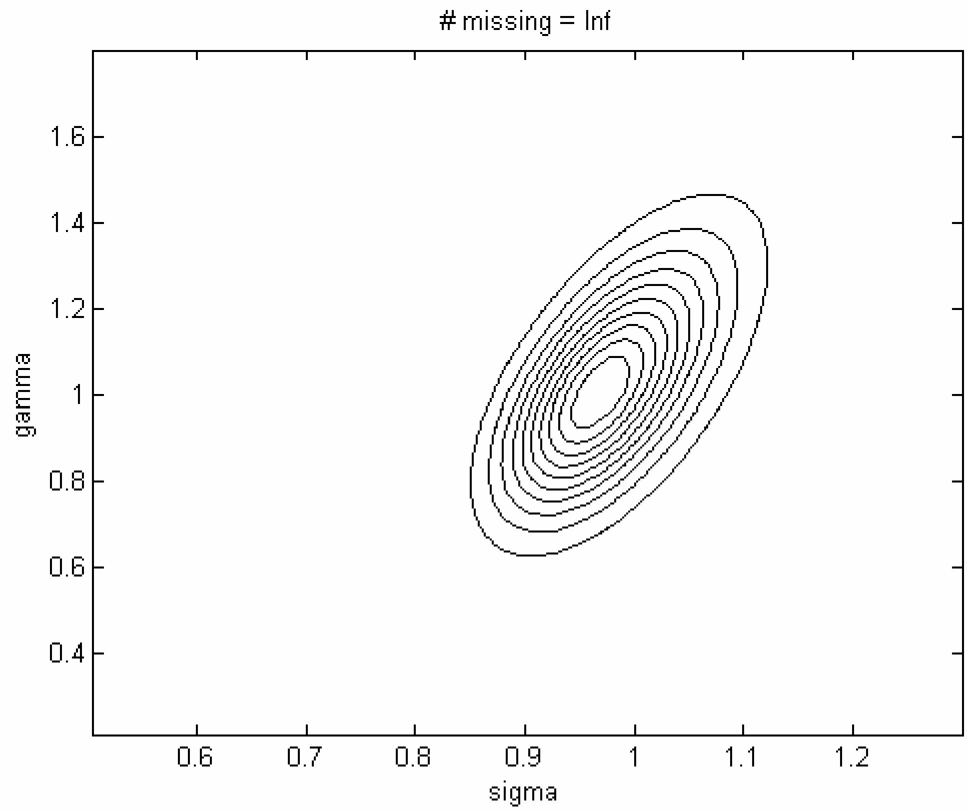
  - Treat the unobserved values of Y_t as *missing data*.

- $P(\theta \mid y_{obs}) \propto \int P(y_{obs}, y_{mis} \mid \theta)\pi(\theta)dy_{mis}$
- Data augmentation:

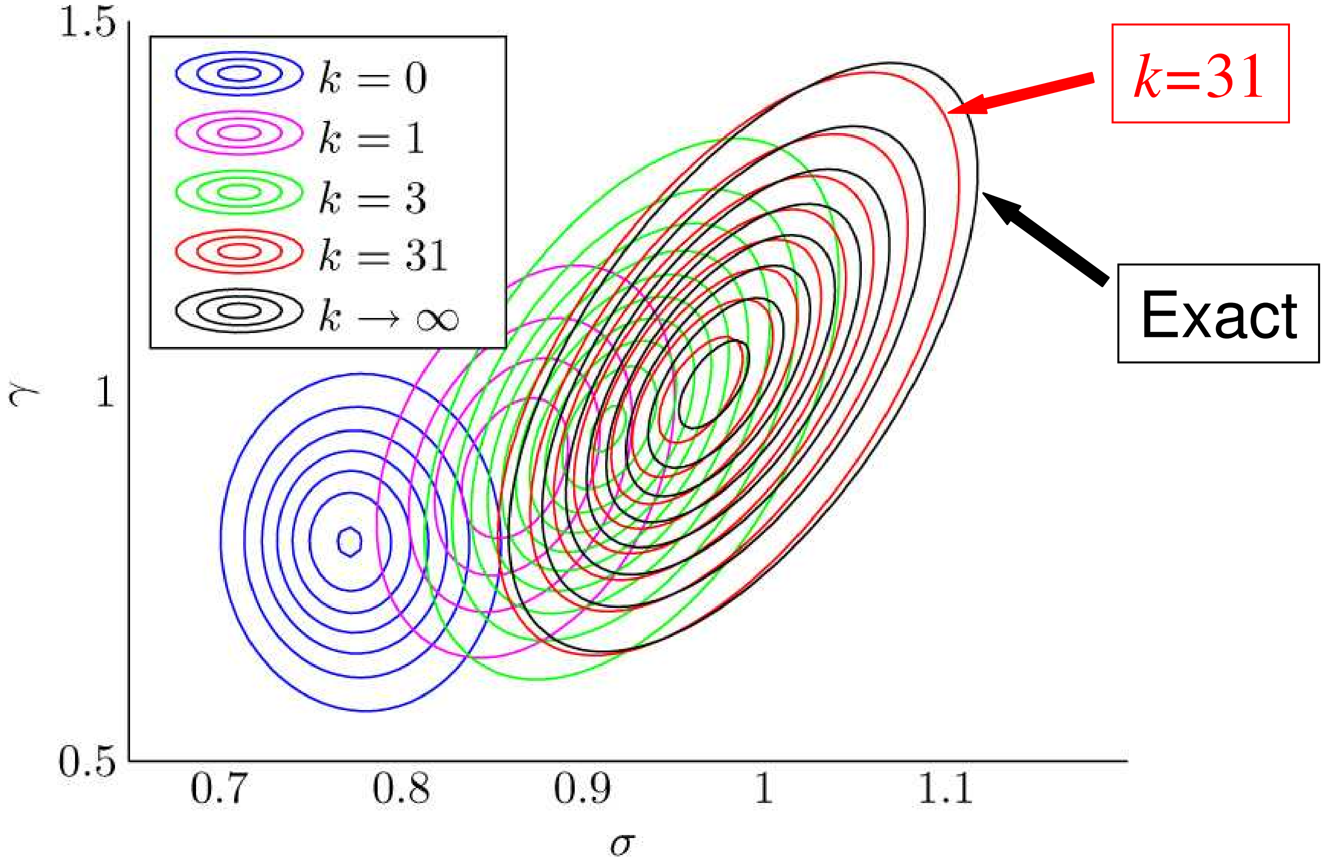
$$P(\theta, y_{mis} \mid y_{obs}) \propto P(y_{obs}, y_{mis} \mid \theta)\pi(\theta)$$

- Use Monte Carlo to perform the augmentation

Bayesian Data Augmentation (ctd)



Bayesian Data Augmentation (ctd)



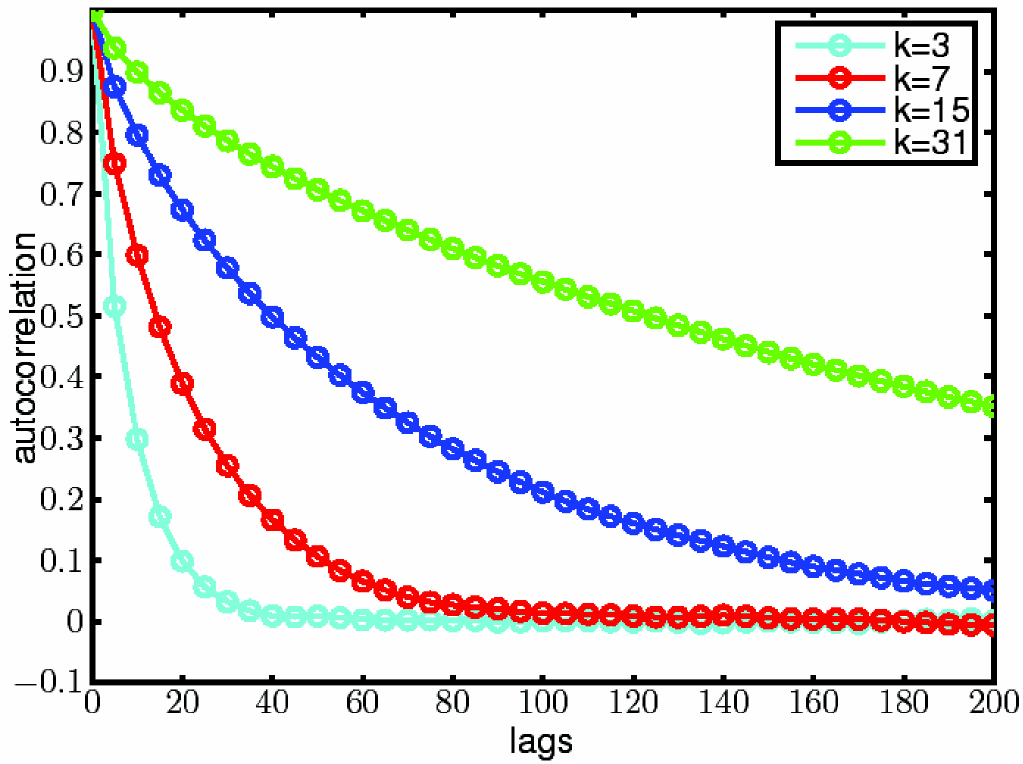
Idea appeared simultaneously in stats & econ literature in late 1990s:

Elerian, Chib, Shephard (2001); Eraker (2001); Jones (1998)

Monte Carlo: Not that easy

- The smaller the Δt , the more accurate the approximation
- However, the smaller the Δt, the more missing data we need to augment: dimensionality goes way up!
- The missing data are dependent as well!

very *slow* convergence of the Gibbs sampler at small Δt



Monte Carlo: Not that easy

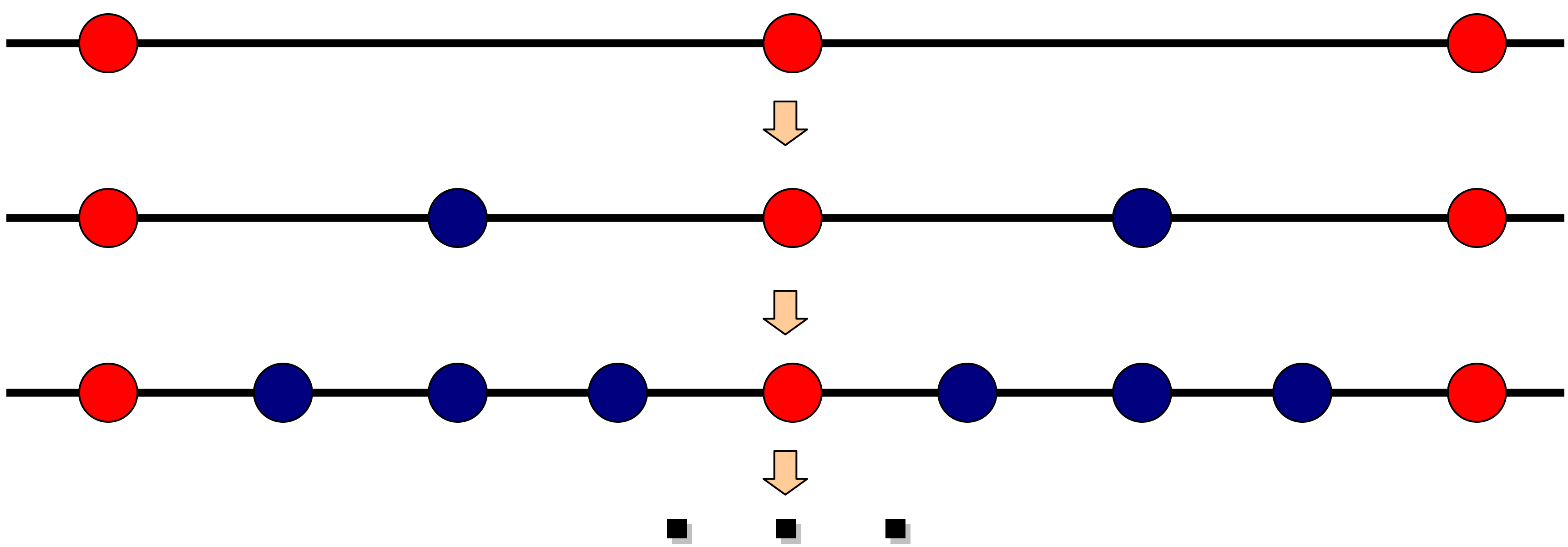
- The smaller the Δt , the more accurate the approximation
- However, the smaller the Δt, the more missing data we need to augment: dimensionality goes way up!
- The missing data are dependent as well!
- The dilemma:
  - Low resolution (big Δt) runs quickly, but result inaccurate
  - High resolution (small Δt) good approximation, but painfully slow

Multi-resolution Idea

- Utilize the strength of different resolutions, while avoid their weakness
- Simultaneously work on multiple resolutions

  “rough” approximations quickly locate the important regions

  “fine” approximations get jump start, and then accurately explore the space



Multi-resolution sampler

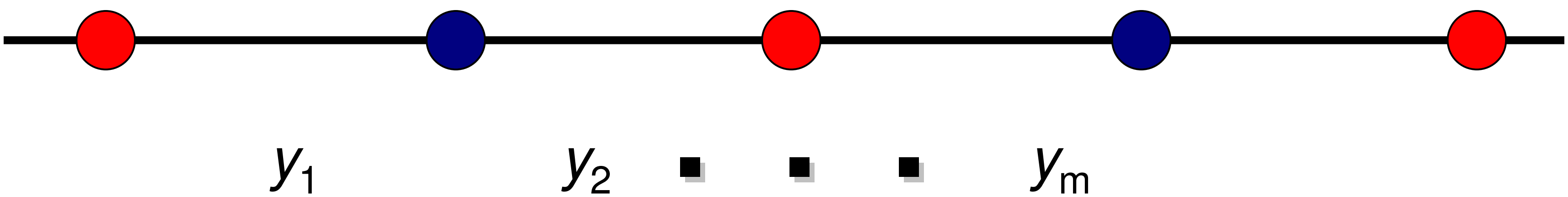
- Consider multiple resolutions (i.e., approximation levels) together. Associate each level with a Monte Carlo chain
- Start from the lowest level with a MC (such as Gibbs sampler); record the results
- Move on to the 2nd level
  - In each MC update, with prob p do Gibbs
  - With prob $1-p$, draw y from previous lower level chain augment y to (y, y') by $\boxed{\text{“upsampling”}}$ accept (y, y') with probability

$$r = \min\{1, \frac{L^{(k+1)}(y, y')L^{(k)}(y_{old})T(y_{old} \to y'_{old})}{L^{(k+1)}(y_{old}, y'_{old})L^{(k)}(y)T(y \to y')}\}$$

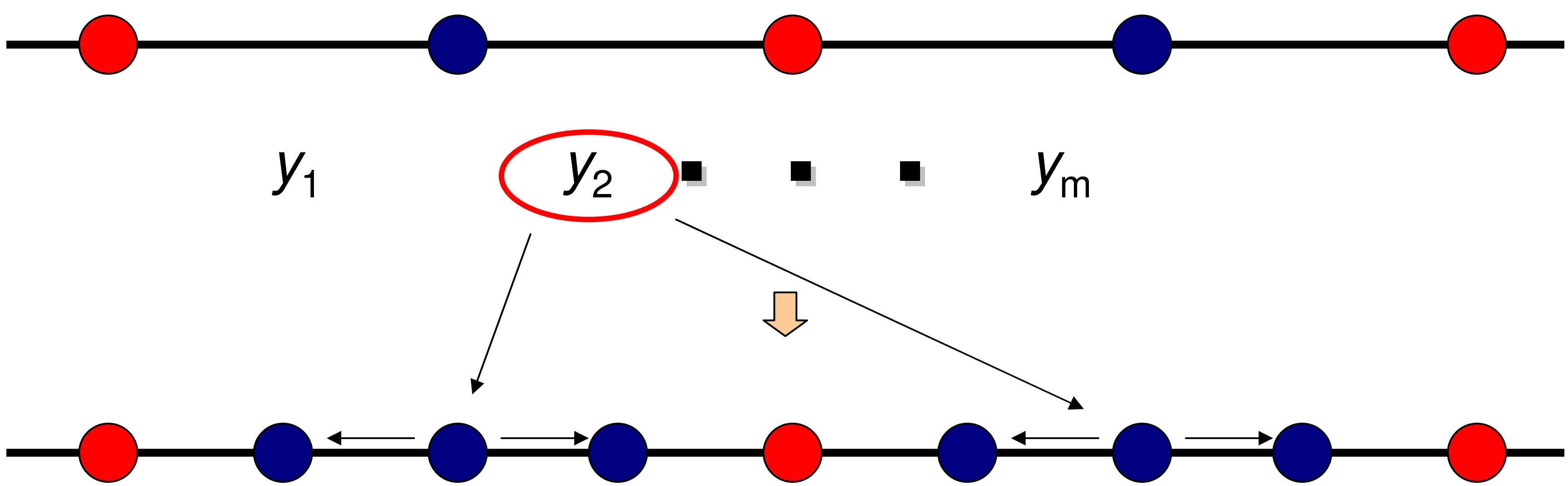
- Move on to the 3rd level ……

Likelihood at level k

A Pictorial Guide



A Pictorial Guide



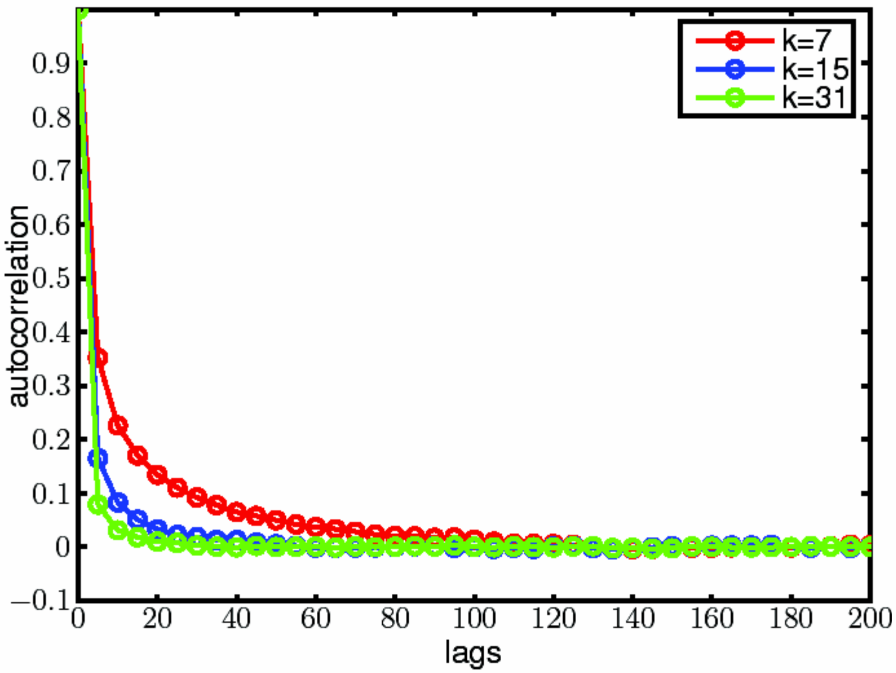
with prob p Gibbs

with prob $1\text{-}p$

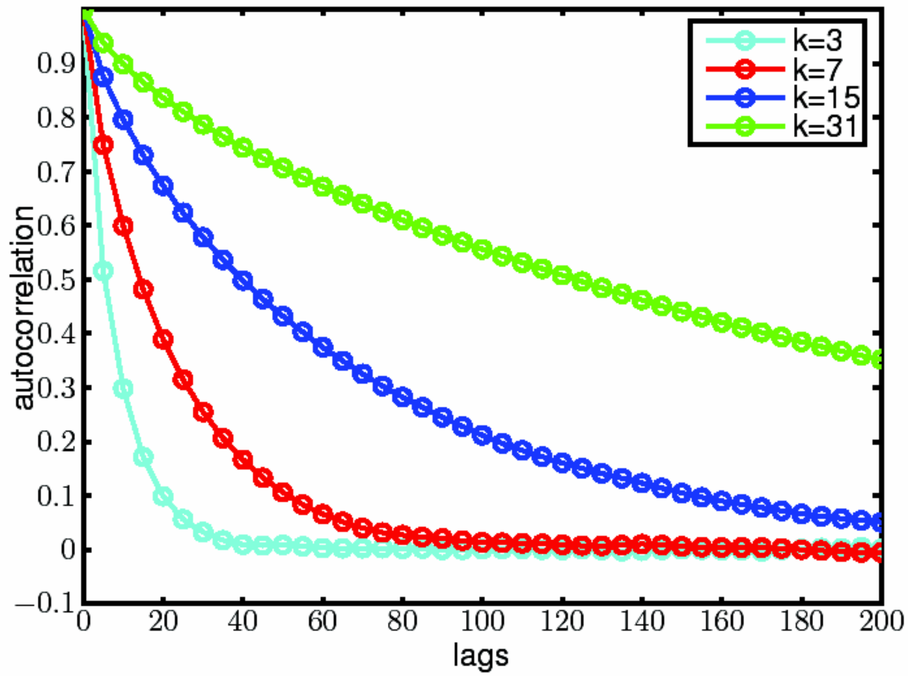
accept with $r = \min\{1, \dfrac{L^{(k+1)}(y, y')L^{(k)}(y_{old})T(y_{old} \to y'_{old})}{L^{(k+1)}(y_{old}, y'_{old})L^{(k)}(y)T(y \to y')}\}$

The comparison

multi-resolution



vanilla Gibbs



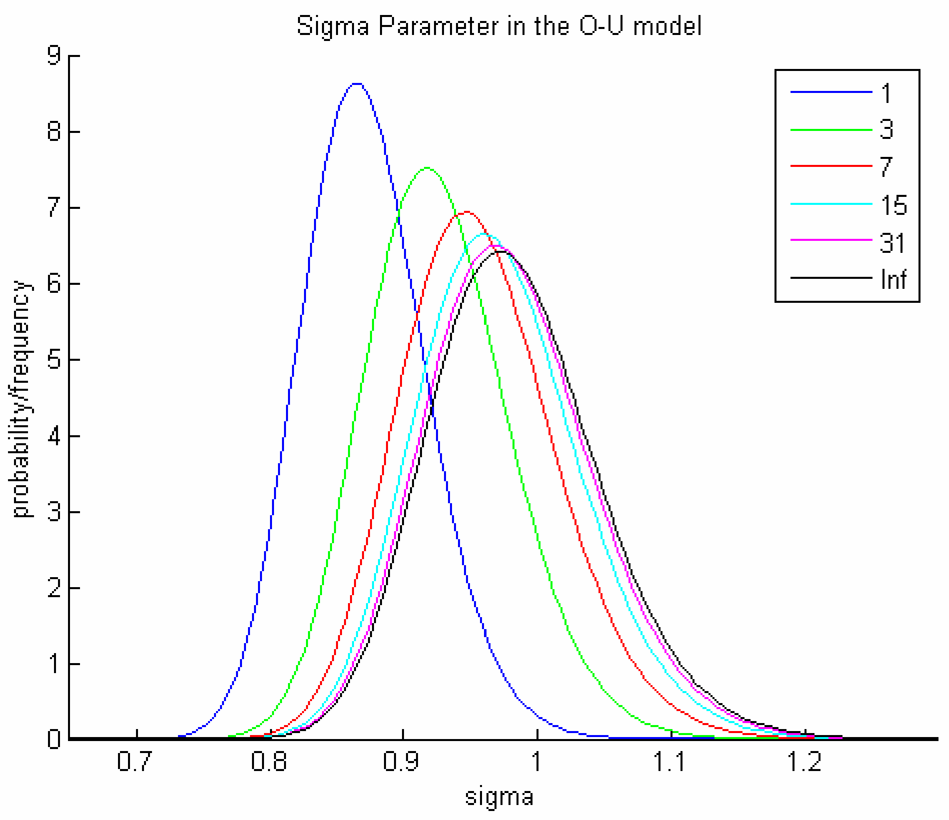
Multi-resolution Inference

- Observation: A by-product of the Multiresolution sampler is that we obtain multiple approximations to the same distribution

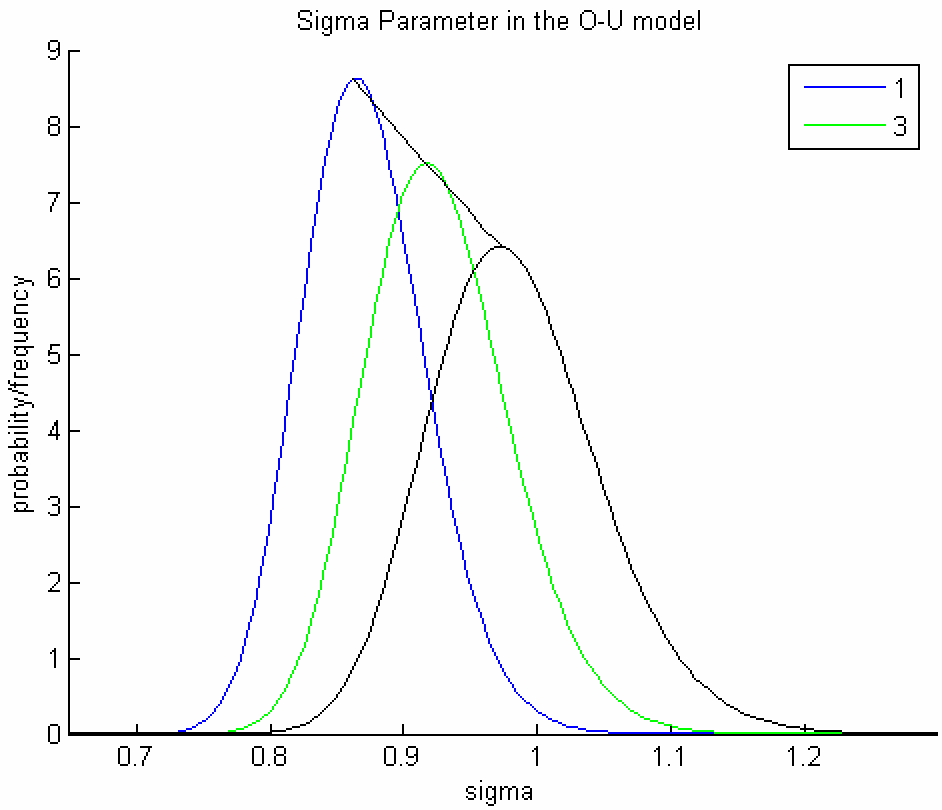
- Question: Can we combine them together for inference, instead of using *only* the finest resolution?

- Idea: Look for trend from successive approximations and leap forward

Illustration



Leap forward:
the multiresolution extrapolation



Richardson Extrapolation

- Richardson (1927)
- If $A = \lim_{h \to 0} A(h)$ is what we want

With resolution h

$$A(h) = A + a_0 h^{k_0} + a_1 h^{k_1} + a_2 h^{k_2} + \cdots$$
$$= A + a_0 h^{k_0} + O(h^{k_1})$$

But for resolution $h/2$

$$A(\frac{h}{2}) = A + a_0 \left(\frac{h}{2}\right)^{k_0} + O(h^{k_1})$$

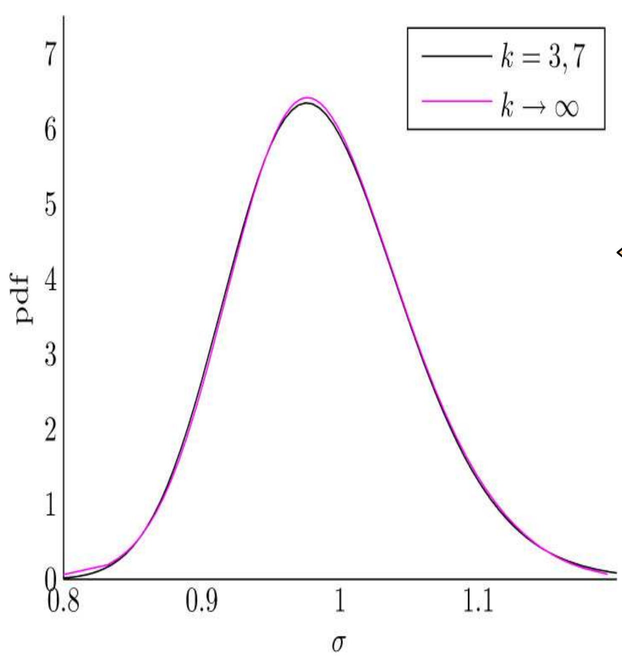
$\Rightarrow$

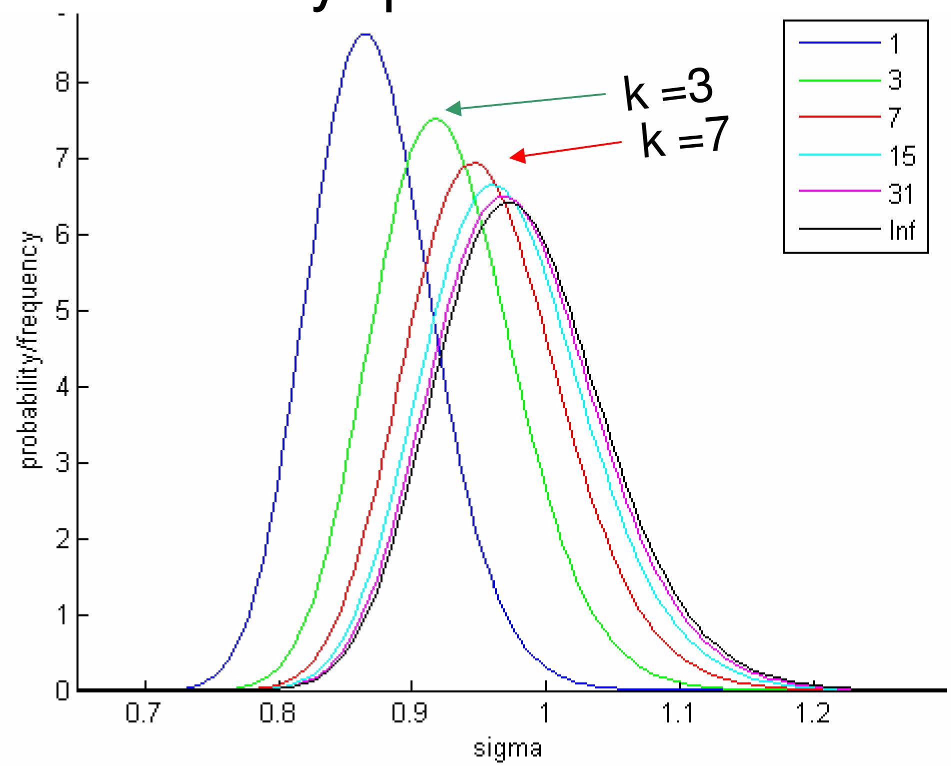
$$\tilde{A}(h) \equiv \frac{2^{k_0} A(\frac{h}{2}) - A(h)}{2^{k_0} - 1} = A + O(h^{k_1})$$

is an order of magnitude better!

Multiresolution Extrapolation

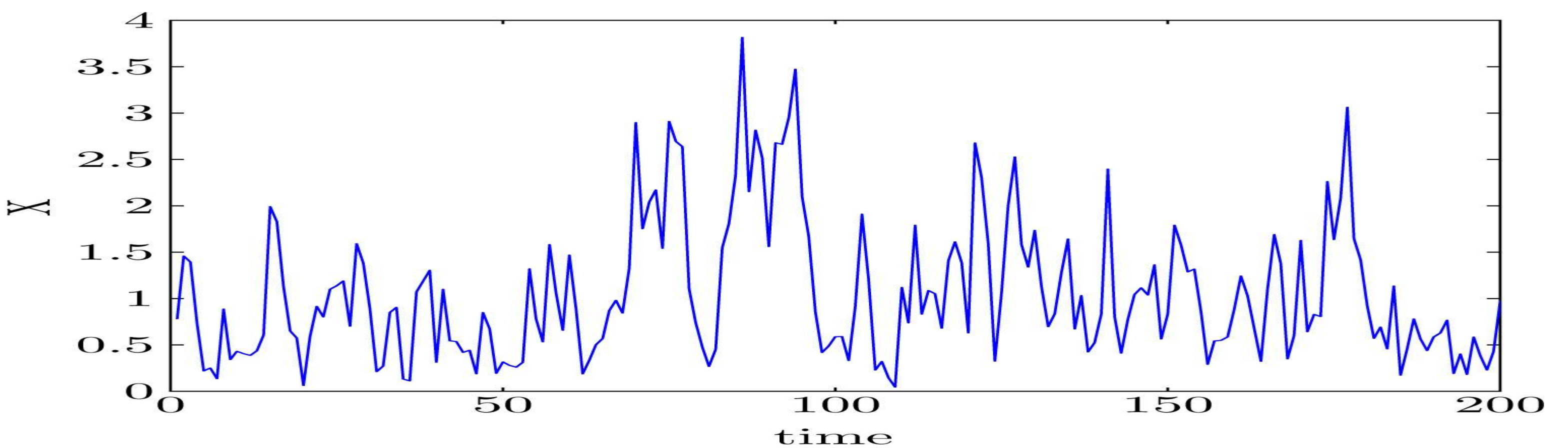
- We have multiple posterior distributions from the multi-resolution sampler
- Extrapolate the entire distribution by quantiles





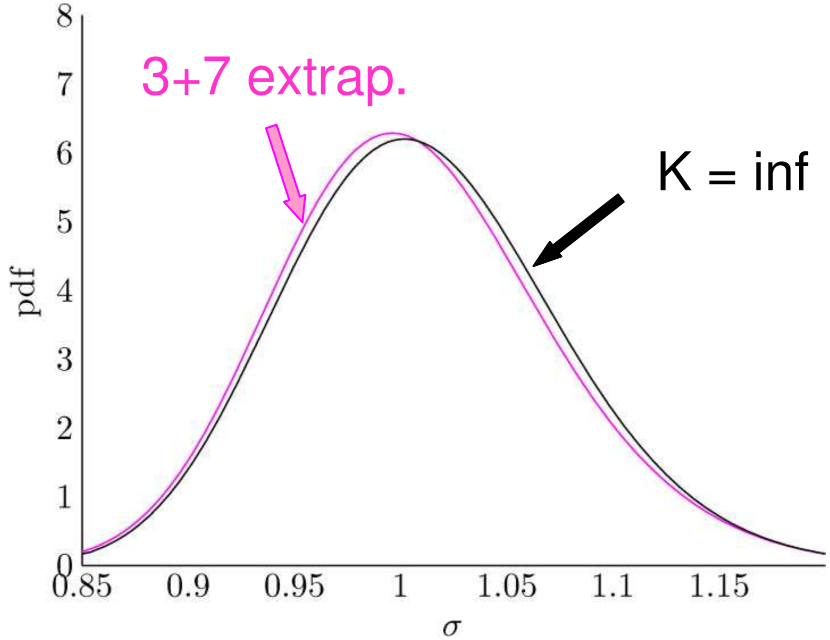
Inference of GCIR process

- $dY_t = \gamma(\mu - Y_t)dt + \sigma Y_t^{\psi} dB_t, \quad \theta = (\gamma, \mu, \sigma, \psi)$
- Model for interest rate, bond rate, exchange rate
- No *analytical* solution for the transition density
- The data

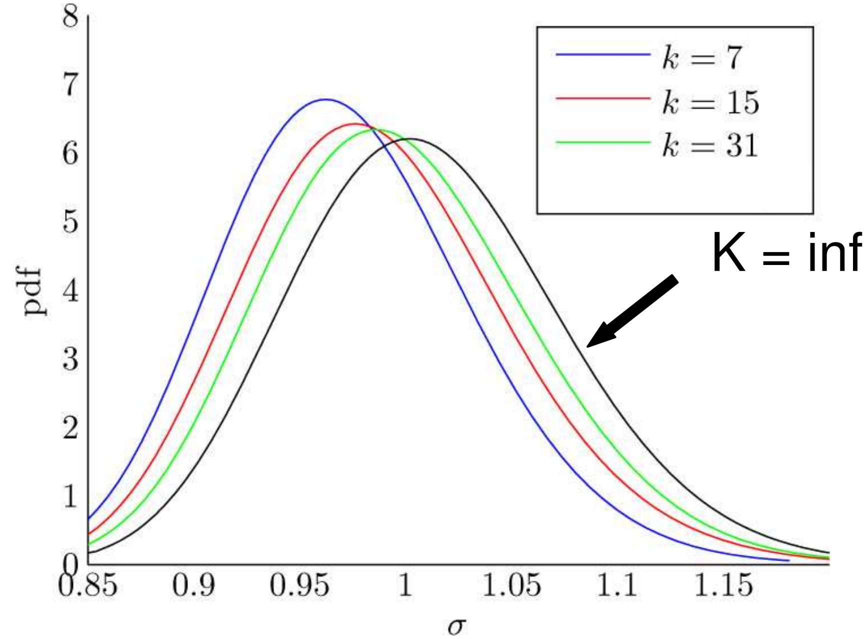


Result

- Posterior distribution



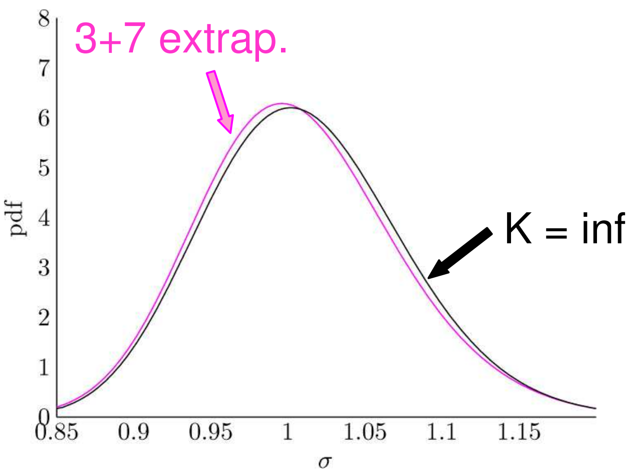
multiresolution

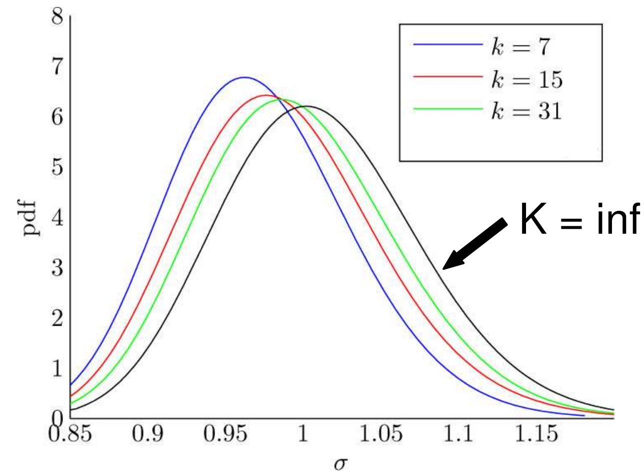


vanilla Gibbs

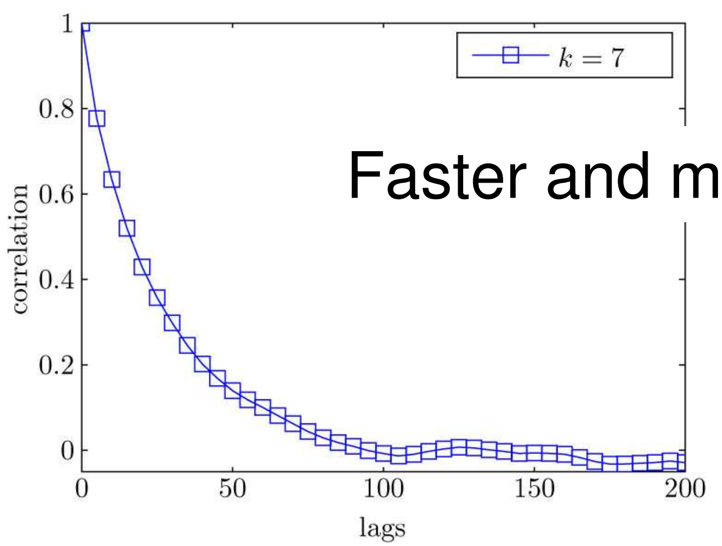
Result

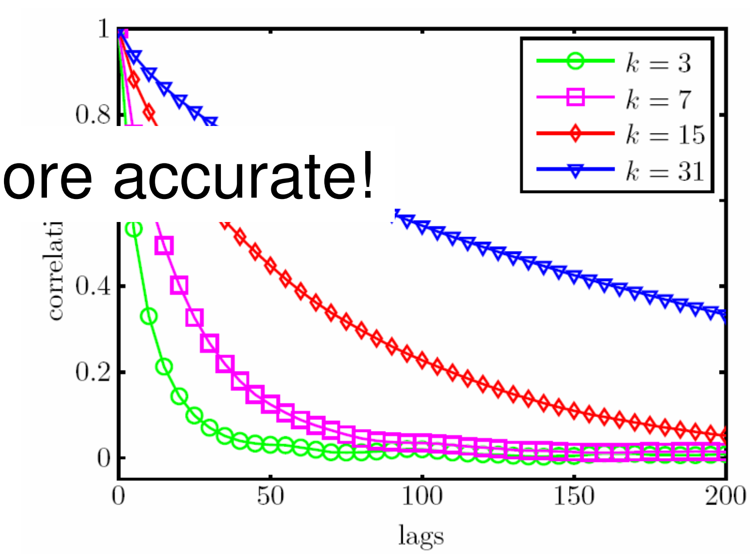
- Posterior distribution





- Autocorrelation plots





Faster and more accurate!

Result (continued)

- In Bayesian analysis, use MC samples to approximate posterior quantities of interest (eg., mean, median, etc.)

$$\bar{\theta} \rightarrow E(\theta \mid Y_{obs})$$

- Use quantiles from MC sample to construct interval estimate

$$\hat{Q}^{(\alpha)}(\theta) \rightarrow Q^{(\alpha)}(\theta \mid Y_{obs})$$

- Compare ratio of Mean Square Error given same time budget:

σ Ratio Estimates	
	$\widehat{R} \pm \sqrt{\widehat{Var}(\widehat{R})}$
$Q_{0.05}$	32 ± 6.2
$Q_{0.25}$	17 ± 3.8
$Q_{0.5}$	11 ± 2.0
$Q_{0.75}$	11 ± 1.8
$Q_{0.95}$	18 ± 2.4

ψ Ratio Estimates	
	$\widehat{R} \pm \sqrt{\widehat{Var}(\widehat{R})}$
$Q_{0.05}$	13 ± 2.0
$Q_{0.25}$	10 ± 1.7
$Q_{0.5}$	10 ± 1.5
$Q_{0.75}$	16 ± 2.7
$Q_{0.95}$	38 ± 6.0

Application 1: Eurodollar rate

- 3-month Eurodollar deposit rate

- Use GCIR model

Eurodollar rate

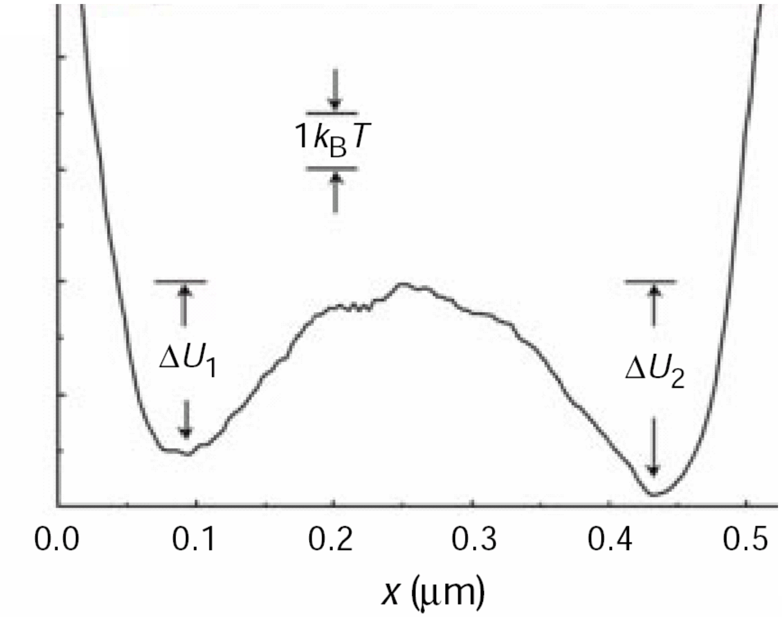
- Posterior mean, median and interval

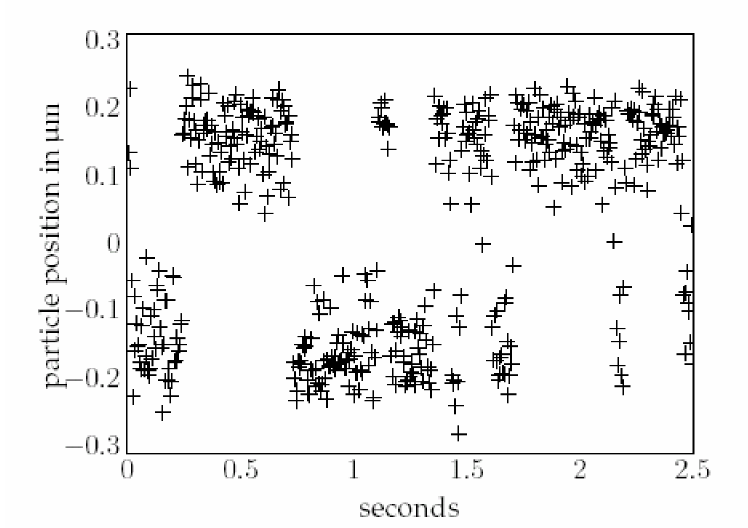
$$dY_t = \gamma(\mu - Y_t)dt + \sigma Y_t^{\psi} dB_t, \quad \theta = (\gamma, \mu, \sigma, \psi)$$

	γ	σ	ψ	μ
mean	9.6776	2.4489	0.84196	4.1417
median	9.6848	2.4492	0.8397	4.1403
2.5%	8.6437	2.3055	0.77723	3.9852
97.5%	10.6421	2.5909	0.91921	4.3034
5%	8.832	2.3292	0.7854	4.0101
95%	10.4808	2.5688	0.90466	4.2772

Application 2: Inference of Optically-Trapped Particle Data

- McCann et al. (1999): Data of a particle in a bistable trap





$$dY_t = -\left(4Y_t^3 + cY_t^2 - 4\beta^2 Y_t - c\beta^2\right)\gamma dt + \sigma dB_t$$

- Again compare ratio of Mean Square Error given same time budget:

γ Ratio Estimates	
	$\widehat{R} \pm \sqrt{\widehat{Var}(\widehat{R})}$
$Q_{0.05}$	2.4 ± 0.65
$Q_{0.25}$	2.3 ± 0.64
$Q_{0.5}$	2.2 ± 0.59
$Q_{0.75}$	2.1 ± 0.54
$Q_{0.95}$	1.8 ± 0.45

c Ratio Estimates	
	$\widehat{R} \pm \sqrt{\widehat{Var}(\widehat{R})}$
$Q_{0.05}$	1.6 ± 0.34
$Q_{0.25}$	2.2 ± 0.45
$Q_{0.5}$	3.1 ± 0.66
$Q_{0.75}$	3.0 ± 0.59
$Q_{0.95}$	2.8 ± 0.59

Discussion

- We introduce the multi-resolution framework
- Efficient Monte Carlo with the multi-resolution sampler
- Accurate inference with the multi-resolution extrapolation
- Extendible to higher dimensions
- Extendible to state space (HMM) models

References

- SØrensen (2004) – Survey paper
- Elerian et al. (2001) – Original Gibbs paper
- Roberts & Stramer (2001) – SDE transformations
- Kloeden & Platen (1992) – Book on Numerical Solution of SDEs

Acknowledgement

- Ben Olding
- Jun Liu
- Xiaoli Meng
- Wing Wong

- NSF, NIH

Thank you!

Extrapolation Theorem

- Assuming:
  - The diffusion & volatility functions $\mu(\bullet)$ and $\sigma^2(\bullet)$ have linear growth
  - $\mu(\bullet)$ and $\sigma^2(\bullet)$ are twice continuously differentiable with bounded derivatives
  - $\sigma^2(\bullet)$ is bounded from below
- Then for any integrable function $g(\boldsymbol{\theta})$:

$$E_{\varepsilon,w}(g(\boldsymbol{\theta})|Y^{(k)}(\boldsymbol{t}) \simeq \boldsymbol{y}) - E_{\varepsilon,w}(g(\boldsymbol{\theta})|Y(\boldsymbol{t}) \simeq \boldsymbol{y}) = \frac{K_g}{k+1} + \mathcal{O}(\frac{1}{k+1})$$

Extrapolation Corollary

- Taking $g(\boldsymbol{\theta})$ to be an indicator function, then if a posterior cdf F has non-zero derivative at all points, its quantiles can be expanded as:

$$F_{\varepsilon,w}^{-1}(z|Y^{(k)}(\boldsymbol{t}) \simeq \boldsymbol{y}) = F_{\varepsilon,w}^{-1}(z|Y(\boldsymbol{t}) \simeq \boldsymbol{y}) + \frac{C}{k+1} + \mathcal{O}(\frac{1}{k+1})$$