Parameter Estimation Methods for Discretely Observed Markov Processes

Peter W. Glynn

Stanford University

Parameter Estimation Methods

for Discretely Observed Markov Processes

1. Maximum Likelihood

   Martingale Estimating Equations

   Method of Moments

2. A Runge-Kutta Method for Simulating SDE's

3. Simulating Transition Densities

Given: Observed data

Context: A family of models (diffusions, jump-diffusions, etc.)

$$dX(t) = \mu(\theta, t, X(t))\, dt \; + \; \sigma(\theta, t, X(t))\, dB(t)$$

Issues:

1.) Which model fits the data best?

2.) How much uncertainty in the parameter estimators?

3.) How well does the model fit?

General Statistical Principle:

Maximum likelihood achieves maximal (asymptotic) statistical efficiency

Other Issues:

1.) Computational tractability / robustness for computing estimator

2.) Computational tractability / statistical quality of related confidence regions

Continuously Observed Process

Observe $(X(s) : 0 \le s \le t)$

$$[X](t) = \int_0^t \sigma^2(\theta^*, s, X(s))ds \qquad \mathsf{P}_{\theta^*} \text{ a.s.}$$

So, assume

$$dX(t) = \mu(\theta, t, X(t))\, dt \; + \; \sigma(t, X(t))\, dB(t)$$

Then

$$\left[\frac{d\mathsf{P}_\theta}{d\mathsf{P}}\right](t)$$

$$= \exp\left(\int_0^t \frac{\mu(\theta, s, X(s))}{\sigma(s, X(s))} dB(s) - \frac{1}{2}\int_0^t \frac{\mu(\theta, s, X(s))^2}{\sigma(s, X(s))^2} ds\right)$$

$$= \exp\left(\int_0^t \frac{\mu(\theta, s, X(s))}{\sigma(s, X(s))^2} dX(s) - \frac{1}{2}\int_0^t \frac{\mu(\theta, s, X(s))^2}{\sigma(s, X(s))^2} ds\right)$$

Likelihood easily computable
(but optimization may be hard!)

Corresponding Estimating Equation

Maximizer of $L(\theta, t) = \left[\frac{d\mathsf{P}_\theta}{d\mathsf{P}}\right](t)$ satisfies

$$\nabla\mathcal{L}(\widehat{\theta}_t, t) = 0$$

where $\mathcal{L}(\theta, t) = \log L(\theta, t)$.
Then,

$$\nabla\mathcal{L}(\widehat{\theta}_t, t) - \nabla\mathcal{L}(\theta^*, t) = -\nabla\mathcal{L}(\theta^*, t)$$

$$\nabla H(\widehat{\theta}_t, t)\, t^{1/2}\left(\widehat{\theta}_t - \theta^*\right) \approx -t^{1/2}\nabla\mathcal{L}(\theta^*, t)$$

martingale

Martingale CLT ...

$$t^{-1/2}\nabla\mathcal{L}(\theta^*, t) \Rightarrow N(0, C)$$

The covariance matrix C can be estimated via the quadratic variation of $t^{-1/2}\nabla\mathcal{L}(\widehat{\theta}_t, t)$

Discretely Observed Process

Extended Data:

$$(T_1, X(T_1)), \ldots, (T_n, X(T_n))$$

Ordered Data:

$$(X(T_1), \ldots, X(T_n))$$

Unordered Data:

$$\frac{1}{n}\sum_{i=0}^{n} \delta_{X(T_i)}(\cdot)$$

How are T_i's generated?

- No model:

  $T_1, T_2, \ldots$ arbitrary deterministic sequence with $T_n \to \infty$

- Statistical model:

  Independent T_i's (e.g. T_n's renewal)

  Correlated T_n's

  $$\mathsf{P}_\theta\left(N(t+h) - N(t) = 1 \,|\, X, N(u) : u \leq t\right)$$
  $$= \lambda(\theta, t, X(t))\, h \,+\, o(h)$$

Extended Data

T_i's deterministic

$$L_n(\theta) = \prod_{i=1}^{n-1} p\left(\theta, T_i, T_{i+1}, X(T_i), X(T_{i+1})\right)$$

T_i's independent and random

$$\begin{aligned} L_n(\theta) \;=\; & \prod_{i=1}^{n-1} p\left(\theta, T_i, T_{i+1}, X(T_i), X(T_{i+1})\right) \\ & \cdot \prod_{i=1}^{n-1} f\left(\theta, T_{i+1} - T_i\right) \end{aligned}$$

T_i's independent and random

$$\begin{aligned} L_n(\theta) \;=\; & \textstyle\prod_{i=1}^{n-1} p\left(\theta, T_i, T_{i+1}, X(T_i), X(T_{i+1})\right) \\ & \cdot \textstyle\prod_{i=1}^{n-1} \left[\lambda\left(\theta, T_{i+1}, X(T_{i+1})\right)\right. \\ & \qquad \left. \cdot u\left(\theta, T_i, T_{i+1}, X(T_i), X(T_{i+1})\right)\right] \end{aligned}$$

where

$$\begin{aligned} & u(\theta, s, t, x, y) \\ & = E\left[\text{exp}\left(- \textstyle\int_s^t \lambda(\theta, r, X(r))dr\right) \,\middle|\, X(s) = x, X(t) = y\right] \end{aligned}$$

Additional Computational Challenges

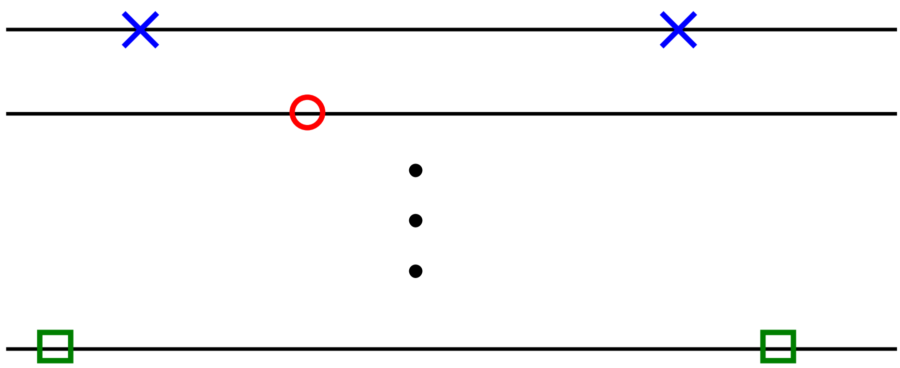
Stochastic Volatility

$$dX(t) = \mu(\theta, t, X(t))\, dt + \sigma(\theta, t, X(t), Z(t))\, dB(t)$$

$$L_n(\theta) = \int \cdots \int \prod_{i=1}^{n-1} \Big[p\left(\theta, T_i, T_{i+1}, (X(T_i), z_i), (X(T_{i+1}), z_{i+1})\right) \Big] \, dz_1 \cdots dz_n$$

Must integrate over unobserved variables

Asynchronous Observations



Martingale Estimating Equations

e.g. $(X(T_1), \ldots, X(T_n))$ ordered data

T_n's Poisson

- $$\sum_{i=1}^{n-1} \Big(u(X(T_{i+1})) - E_{\theta_n}\left[u(X(T_{i+1})) \middle| X(T_i) \right] \Big) \Gamma_i = 0$$

where Γ_i is $\mathcal{F}_i$-measurable

- Conditional Least Squares special case

- Covariance structure easily estimated (quadratic variation)

- Computing

$$g(\theta, x) = E_x^\theta u(X(T))$$

requires solving

$$(\lambda I - A(\theta))\, g(\theta) = \lambda \mu$$

where

$$A(\theta) = \mu(\theta, x) \frac{d}{dx} + \frac{\sigma^2(\theta, x)}{2} \frac{d^2}{dx^2}$$

This computation can be avoided.

- Choose g smooth ... DG 04

$$(\lambda I - A(\theta))\, g \stackrel{\triangle}{=} \lambda u(\theta)$$

- Martingale estimating equation:

$$\sum_{i=1}^{n-1} \Big(u\left(\theta, X(T_{i+1})\right) - g(X(T_i)) \Big)\, \Gamma_i = 0$$

- Covariance structure easily estimated

- Easily extended to doubly stochastic Poisson process

- No known extension to deterministic sampling

Method of Moments

e.g. unordered data

$$\frac{1}{n}\sum_{i=0}^{n}\delta_{X(T_i)}(\cdot)$$

- T_i's deterministic or independent of X;
  X positive recurrent.
- $\frac{1}{n}\sum_{i=1}^{n} u\left(\theta_n, X(T_i)\right) = \int \pi(\theta_n, dx)u(\theta_n, x)$
- H-S 95

$$u(\theta, x) = \left(A(\theta)g\right)(x)$$

- no obvious martingale structure
- computing confidence regions (much) harder

Central Limit Theorem

Assume:

- X is Harris recurrent in continuous time
- $\underline{\lim}\,(T_{i+1} - T_i) > 0$
- $A(\theta^*)V \leq -\varepsilon V$ on K^C
- $g(\theta^*) \leq V^{1/4}$

Then, there exists $s_n \longrightarrow \infty$ such that

$$s_n^{-1} \sum_{i=1}^{n} g(\theta^*, X(T_i)) \Rightarrow N(0,1)$$

Computing Confidence Regions

Simulate n iid copies of X under P_θ, and compute sample variance of

$$\sum_{i=1}^{n} g(\theta, X(T_i))$$

(with $\theta = \theta_n$)

Simulation-based Estimators

1. Method of Moments

$$\text{e.g. Compute} \quad \int \pi(\theta, dx) g(x)$$

2. Martingale Estimating Equations

$$\text{e.g. Compute} \quad E_\theta \left[u(X(T_{i+1})) \middle| X(T_i), T_i \right]$$
$$\text{for} \quad 1 \leq i \leq n-1$$

3. Maximum Likelihood

$$\text{Compute} \quad p\left(\theta, T_i, T_{i+1}, X(T_i), X(T_{i+1})\right)$$
$$\text{for} \quad 1 \leq i \leq n-1$$

Simulating SDE's

Goal: Compute $Ef(X)$ where

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t)$$

Euler scheme:

$$X_h((k+1)h) - X_h(kh)$$

$$= \mu\left(kh, X_h(kh)\right) h + \sigma\left(kh, X_h(kh)\right) h^{1/2} Z_{k+1}$$

Milstein scheme: If $\mu = 0$,

$$X_h((k+1)h) - X_h(kh)$$

$$\begin{aligned} = \int_{kh}^{(k+1)h} \Bigg[& \sigma\left(kh, X_h(kh)\right) + \frac{\partial}{\partial t}\sigma\left(kh, X_h(kh)\right)(t - kh) \\ & + \frac{\partial}{\partial x}\sigma\left(kh, X_h(kh)\right)\sigma\left(kh, X_h(kh)\right) \\ & \cdot \left[B(t) - B(kh)\right] \Bigg] dB(t) \end{aligned}$$

partial derivatives of σ
must be computed

Estimate $Ef(X)$ via $\frac{1}{m}\sum_{j=1}^{m} f_h(X_h^j)$

Mean Square Error:

$$\frac{1}{m}\mathrm{var} f_h(X_h) + (\mathrm{bias}(f_h(X_h)))^2$$

Weak Error: Suppose f smooth

$$Ef(X_h(t)) = Ef(X(t)) + ch^p + o(h^p)$$

p'th order scheme

Euler: $p = 1$

Strong Error:

$$E\max_{0\le s\le t}|X_h(s) - X(s)| = O(h^p)$$

p'th order scheme

Euler: $p = 1/2$

Milstein: $p = 1$

$$h^{-1/2}(X_h(t) - X(t)) \Rightarrow Z(t) \qquad \text{K-P 91}$$

Useful for f Lipschitz:

$|f(x) - f(y)| \le c\|x - y\|$

Total Variation Error

- $$\begin{aligned} d(P,Q) &= \|P-Q\| \\ &= \mathsf{sup}_A \, |P(A)-Q(A)| \end{aligned}$$

- Instead, consider

$$d_h(\mathsf{P}_h,\mathsf{P}) \stackrel{\triangle}{=} \sup_{A\in\mathcal{F}_h} |\mathsf{P}_h(A)-\mathsf{P}(A)|$$

$$\mathcal{F}_h = \sigma\left(X(0), X(h), \ldots, X(h/\lfloor t/h \rfloor)\right)$$

$$= \frac{1}{2}\, E\left| \prod_{k=0}^{n-1} \frac{p_h\left(X(kh), X((k+1)h)\right)}{p\left(h, X(kh), X((k+1)h)\right)} - 1 \right|$$

- Euler scheme: $\quad d_h(\mathsf{P}_h,\mathsf{P}) \to \gamma \in (0,\ 1)$

$$E \prod_{k=0}^{n-1} \frac{p_h\left(X(kh), X((k+1)h)\right)}{p\left(h, X(kh), X((k+1)h)\right)} \;=\; 1$$

$$E \prod_{k=0}^{n-1} \frac{p_h\left(X_h(kh), X_h((k+1)h)\right)}{p\left(h, X_h(kh), X_h((k+1)h)\right)} \;\not\to\; 1$$

Computing Short-Time Asymptotics

for Transition Density:

To illustrate idea, let $d = 1$, $\mu = 0$, $x = 0$, $a(\cdot) = \sigma^2(\cdot)$. Suppose

$$p = p_0 + p_1 + p_2 + \cdots$$

$$\frac{\partial}{\partial t} p = (1/2)\, \frac{\partial^2}{\partial y^2} (a(y)p(t,x,y))$$

If $a(y) = a(0) + a'(0)y + a''(0)\frac{y^2}{2} + \cdots$ then

$$\begin{aligned} \frac{\partial}{\partial t} p_0 &= \tfrac{1}{2}\, a(0) \frac{\partial^2}{\partial y^2} p_0 \\ \frac{\partial}{\partial t} p_1 &= \tfrac{1}{2}\, a(0) \frac{\partial^2}{\partial y^2} p_1 + \frac{a'(0)}{2} \frac{\partial^2}{\partial y^2} (y p_0) \end{aligned}$$

Can be solved in terms of Hermite polynomials

e.g. $\quad p_1(t,0,y) = \frac{\sigma'(0)}{2} t^{1/2} H_3\left(\frac{y}{\sigma(0)t^{1/2}}\right) p_0(t,0,y)$

$$\|p - p_0 - p_1 - \cdots - p_m\| = O\left(t^{\frac{m+1}{2}}\right)$$

Milstein scheme:

$$d_h(\mathsf{P}_h, \mathsf{P}) = O(h^{1/2})$$

Runge-Kutta scheme:

$$\begin{aligned}
\widetilde{Y}_k^1 &= \mu(t_k, X_k)h \\
Y_k^1 &= \sigma(t_k, X_k)h^{1/2}Z_{k1} \\
Y_k^2 &= \sigma\left(t_k + \varepsilon h, X_k + \alpha Y_k^1 + \tilde{\alpha}\widetilde{Y}_k^1\right) h^{1/2} Z_{k2} \\
\widetilde{Y}_k^2 &= \mu\left(t_k + \tilde{\varepsilon} h, X_k + \alpha_1 Y_k^1 + \tilde{\alpha}_1 \widetilde{Y}_k^1\right) h \\
X_{k+1} &= X_k + \widetilde{Y}_k^2 + \beta Y_k^1 + \gamma Y_k^2
\end{aligned}$$

Choose $\varepsilon, \tilde{\varepsilon}, \alpha, \tilde{\alpha}, \alpha_1, \tilde{\alpha}_1, \beta, \gamma$ appropriately

$$d_h(\widetilde{\mathsf{P}}_h, \mathsf{P}) = O(h^{1/2})$$

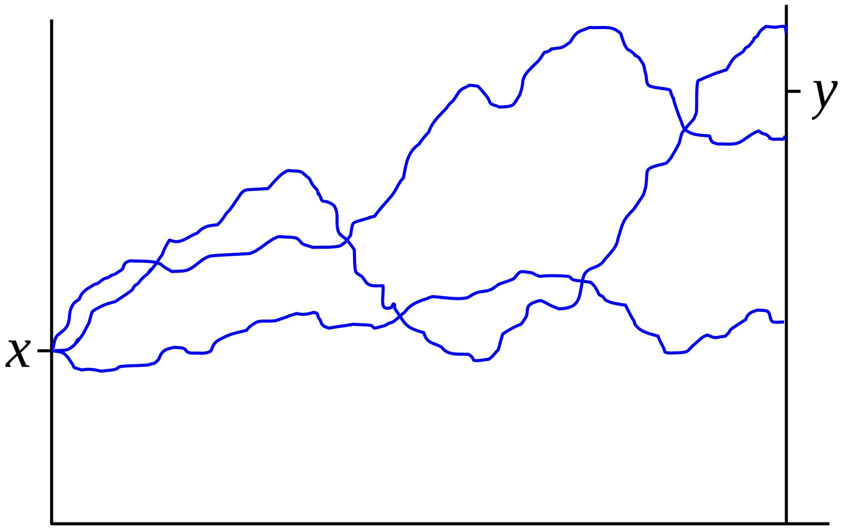
We are working on higher order Runge-Kutta schemes

Computing the Transition Density

Goal: Compute $p(0, 1, x, y)$

1. Simulate the approximating SDE over [0,1] and compute the density at t=1

$$\frac{1}{n}\sum_{i=1}^{n} h_n^{-d}\phi\left(\frac{y - X_h^i(1)}{h_n}\right)$$

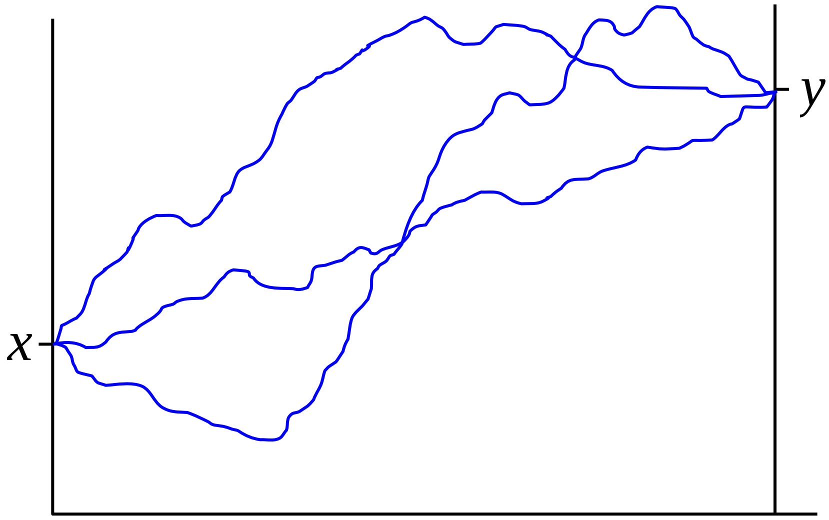


2. Simulate the approximate solution to the SDE over [0,1] (based on m time steps) and compute

$$E\, p\left(h, 0, 1, X_h\left(\frac{m-1}{m}\right), y\right)$$

approximation to transition density

3. Importance sampling via Brownian bridge



$$p_h(0, 1, x, y)$$

$$= \int \cdots \int p_h\left(0, \frac{1}{m}, x, z_1\right) \cdots p_h\left(\frac{m-1}{m}, 1, z_{m-1}, y\right) dz_1 \cdots dz_{m-1}$$

$$= \int \cdots \int \frac{p_h\left(0, \frac{1}{m}, x, z_1\right)}{q\left(0, \frac{1}{m}, x, z_1\right)} \cdots \frac{p_h\left(\frac{m-1}{m}, 1, z_{m-1}, y\right)}{q\left(\frac{m-1}{m}, 1, z_{m-1}, y\right)} \cdot q\left(0, \frac{1}{m}, x, z_1\right) \cdots q\left(\frac{m-1}{m}, 1, z_{m-1}, y\right) dz_1 \cdots dz_{m-1}$$

$$= E\, \frac{p_h\left(0, \frac{1}{m}, x, X(\frac{1}{m})\right)}{q\left(0, \frac{1}{m}, x, X(\frac{1}{m})\right)} \cdots \frac{p_h\left(\frac{m-1}{m}, 1, X(\frac{m-1}{m}), y\right)}{q\left(\frac{m-1}{m}, 1, X(\frac{m-1}{m}), y\right)}$$

4. Markov Chain Monte Carlo

E 01

- Run Markov chain with stationary distribution
$$\frac{p_h\left(0, \frac{1}{m}, x, z_1\right) \cdots p_h\left(\frac{m-1}{m}, 1, z_{m-1}, y\right)}{p_h(0,1,x,y)}$$
using Gibbs sampler

- Compute
$$\left(E\, p_h\left(0, \frac{1}{m}, x, X(\tfrac{1}{m})\right)^{-1} \cdots p_h\left(\frac{m-1}{m}, 1, X(\frac{m-1}{m}), y\right)^{-1}\right)^{-1}$$

Computational Experience

D G 03, E 01

- These methods are computationally feasible for low dimensional problems (e.g. 1-dimensional model with stochastic volatility)

- Other variance reduction methods: common random numbers, stratification, etc.

Maximizing over θ

- Iterative algorithms (K-W, R-M)

- Random search

$$\max_{\theta \in [0,1]^d} \alpha(\theta)$$

  Choose m points $\theta_1, \ldots, \theta_m$

  Simulate n replications at each point

  Choose best point

- Optimal design C-G 04

$$\left. \begin{array}{r} m \sim c^{\frac{d}{d+4}} \\ n \sim c^{\frac{4}{d+4}} \end{array} \right\} \text{rate} \quad c^{-\frac{2}{d+4}}$$

- limit distribution (special case)

$$\mathsf{P}\left(c^{\frac{2}{d+4}}(\widehat{\alpha}(c) - \alpha(\theta^*)) \le x\right)$$

$$\longrightarrow \exp\left(-2^{-d} \int_0^\infty \cdots \int_0^\infty \mathsf{P}\left(N(0,1) > x + y_1 + \cdots + y_d\right) y_1^{-1/2} \cdots y_d^{-1/2}\, dy_1 \cdots dy_d\right)$$

Concluding Remarks

1. Trade-off of statistical efficiency versus computational tractability.

2. Simulation-based algorithms for ML are (potentially) feasible

3. Research issues:

    algorithmic behavior as $n \longrightarrow \infty$

    better optimization tools