Information Theory for Maximum Likelihood Estimation of Diffusion Models

Hwan-sik Choi
Department of Economics
Binghamton University

Nov. 2014
Motivating Questions

Suppose we have a misspecified diffusion model.

• Does the maximum likelihood estimator (MLE) converge to some limits asymptotically?

• Do the limits have any meaning?

• Are the usual inference procedures based on likelihood ratios valid?

• What is the consequence of partial (drift or diffusion function) misspecication?

• What is the impact of having to use an approximate transition density when the exact density is unknown in closed-form?
Asymptotic Framework: \( T \to \infty, \Delta \to 0 \)

Stationary True diffusion \( X \) : 
\[
dX_t = \mu_0(X_t) \, dt + \sigma_0(X_t) \, dW_t
\]

Parametric Model \( M(\theta) \) : 
\[
dX_t = \mu(X_t; \alpha) \, dt + \sigma(X_t; \beta) \, dW_t
\]

\( M(\theta) \) is misspecified \( \Rightarrow \) There is no \( \theta = (\alpha, \beta) \) such that 
\[
\mu(\cdot; \alpha) = \mu_0(\cdot) \quad \text{and} \quad \sigma(\cdot; \beta) = \sigma_0(\cdot).
\]

Data: \( X_0, X_\Delta, X_{2\Delta}, \ldots, X_{i\Delta}, \ldots, X_{N\Delta} \) with \( T = N\Delta \).

\[
\begin{align*}
\Delta & \quad \ldots \quad T \\
\hline
0 & \quad \ldots \quad N \Delta \\
\end{align*}
\]

\[
N = \frac{T}{\Delta}, \quad T = N\Delta, \quad \Delta = \frac{T}{N}.
\]

Information theoretic foundation for MLE \( \hat{\theta}_{T,\Delta} \) as 
\[
T \to \infty \quad \text{and} \quad \Delta \to 0.
\]
Main Contribution

The new foundation provides a unified framework for both correctly specified and misspecified diffusion models.

• MLE converges to the pseudo-true parameter $\theta^*$ closest to the true process.

• In what sense, is $M(\theta^*)$ the closest diffusion to $X$?
  $\Rightarrow$ Provide two-tier information criteria $C^p$ and $C^S$ for diffusions.

• Curvatures of $C^p$ and $C^S = \text{Asymptotic variances of MLE for correct models.}$

• Geometric interpretation for ML estimation is possible.
White (1982) and K-L Information Criterion

KLIC from a measure $P$ to $Q$: $\text{KLIC}(P, Q) \equiv \int \log \left( \frac{dP}{dQ} \right) dP$.

MLE in the discrete-time framework:

- Sample $(X_1, \ldots, X_n)$ from a measure $P$ with a density $\prod_{i=1}^{n} p(x_i)$.
- Let $Q(\theta)$ be a model with density $\prod_{i=1}^{n} q(x_i; \theta)$.
- MLE maximizes $\frac{1}{n} \sum_{i=1}^{n} \log q(X_i; \theta)$ or minimizes
  \[ \hat{\theta}_n = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \log \frac{p(X_i)}{q(X_i; \theta)} \approx \arg\min_{\theta} \text{KLIC}(P, Q(\theta)) \]

- Pseudo-true value $\theta^* = \text{plim}_{n \to \infty} \hat{\theta}_n = \arg\min_{\theta} \text{KLIC}(P, Q(\theta))$. 
True transition = $p_0(t, X_0, X_t)$; transition of $M(\theta) = p(t, X_0, X_t; \theta)$.

Define KLIC from $p_0$ to $p \Rightarrow$ Expand it around $t = 0$:

$$D_t(\theta) \equiv \mathbb{E} \log \frac{p_0(t, X_0, X_t)}{p(t, X_0, X_t; \theta)} = C^P(\theta) + C^S(\theta) t + o(t).$$

Primary criterion $C^P(\theta) \equiv \lim_{t \to 0} D_t(\theta)$,

Secondary criterion $C^S(\theta) \equiv \lim_{t \to 0} \frac{\partial D_t(\theta)}{\partial t}$. 
Primary Criterion

Let \( f(t, x, y) = t \log \frac{p_0(t, x, y)}{p(t, x, y; \theta)} \) and \( \mathcal{A} \) be the infinitesimal generator of \( X \).

\[
C^P(\beta) \equiv \lim_{t \to 0} D_t(\theta) = \mathbb{E}(\mathcal{A}f)_{t \to 0, y \to x}(X_0)
\]

\[
= \frac{1}{2} \mathbb{E} \left[ \left( \frac{\sigma_0^2(X_0)}{\sigma^2(X_0; \beta)} - 1 \right) - \log \left( \frac{\sigma_0^2(X_0)}{\sigma^2(X_0; \beta)} \right) \right].
\]

\( C^P(\theta) \) depends on \( \sigma(\cdot; \beta) \) only \( \Rightarrow \) Write it as \( C^P(\beta) \).

\( C^P(\beta) \geq 0 \) and \( C^P(\beta) = 0 \) iff \( \sigma^2(\cdot; \beta) = \sigma_0^2(\cdot) \) regardless of \( \mu(\cdot; \alpha) \).
Secondary Criterion: Derivative of $D_t$ at $t = 0$

We can show similarly that

$$C^S(\theta) \equiv \lim_{t \to 0} \frac{\partial D_t(\theta)}{\partial t} = \frac{1}{2} \mathbb{E}(A^2f)_{t \to 0, y \to x}(X_0).$$

It depends on both $\alpha$ and $\beta$, and approximate transition (if used).

If $\sigma = \sigma_0$, $C^S(\theta) \geq 0$. When $\sigma(\cdot; \beta)$ is misspecified,

- it can be negative and
- is minimized at $\alpha^* \neq \alpha_0$ even if $\mu(\cdot; \alpha)$ is correctly specified.
Explicit Form of Secondary Criterion

Notation: $\sigma = \sigma(X_0; \beta)$, $f_{yyy} = \lim_{t \to 0, y \to x} \partial^3 f / \partial y^3$, etc.

$$C_S(\alpha; \beta) = \frac{1}{2} \mathbb{E} \left[ \left\{ \mu_0 \left( \mu_0 + \frac{\sigma_0^2}{2} \right) + \sigma_0^2 \left( \mu_0/x + \frac{\sigma_0^2}{4} \right) \right\} \left( -\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right) \right.$$

$$+ \sigma_0^2 \left( \mu_0 + \frac{\sigma_0^2}{2} \right) f_{yyy} + \frac{\sigma_0^4}{4} f_{yyyy} + f_{tt} + 2\mu_0 f_{yt} + \sigma_0^2 f_{yyt} \right].$$

The choice of an approximate transition density matters.

- 5 terms above depend on approximate transition.
- They are identical for the exact and Aït Sahalia’s closed-form approximate transitions.
Asymptotics for MLE

Under $\Delta^3 T \to 0$ as $\Delta \to 0$ and $T \to \infty$,

$$\hat{\beta} \overset{p}{\to} \beta^* \equiv \arg\min_\beta C^P(\beta), \quad \hat{\alpha} \overset{p}{\to} \alpha^* \equiv \arg\min_\alpha C^S(\alpha; \beta^*).$$

- $C^P = 0$ iff $\sigma(\cdot; \beta)$ is correct $\Rightarrow$ $\hat{\beta}$ is consistent regardless of $\mu(\cdot; \alpha)$.
- $C^P = C^S = 0$ iff both $\sigma(\cdot; \beta) = \sigma_0$ and $\mu(\cdot; \alpha) = \mu_0$.
- Approximate transition $\Rightarrow$ $C^S(\theta)$ changes $\Rightarrow$ $\alpha^*$ changes $\Rightarrow$ Same $\alpha^*$ for exact and Aït-Sahalia’s approx.
- Incorrect $\sigma(\cdot; \beta)$ changes $\alpha^*$ $\Rightarrow$ Bad $\sigma$ leads to bad $\mu$ estimation.
  $\Rightarrow$ If $\mu(\cdot; \alpha)$ is correct, use the Euler approx. to get $\alpha_0$ because $\alpha^*$ does not depend on $\sigma(\cdot; \beta^*)$. 
First Order Condition for $C^P(\beta)$ and $C^S(\alpha; \beta^*)$

For diffusion function estimator $\hat{\beta}$,

$$EAf_{/\beta} = \mathbb{E}\left[\frac{1}{2} \left(-\sigma^2 + \sigma_0^2\right) (\sigma^{-2})_{/\beta}\right] = 0.$$

For drift estimator $\hat{\alpha}$, consider three cases.

(1) Exact and Aït-Sahalia’s approximate transition densities:

$$EA^2f_{/\alpha} = \mathbb{E}\left[\frac{2\mu/\alpha}{\sigma^2} (\mu - \mu_0) + \left(1 - \frac{\sigma_0^2}{\sigma^2}\right) \left(\frac{\mu/\alpha}{x} - \frac{\mu/\alpha\sigma^2/\sigma_x}{\sigma^2}\right)\right] = 0.$$

If $\sigma \neq \sigma_0$ but $\mu$ is correctly specified (partial misspecification), the correct $\mu = \mu_0(x; \alpha_0)$ would not satisfy the FOC.

$\Rightarrow \hat{\alpha} \xrightarrow{p} \alpha^* \neq \alpha_0$ ($\hat{\alpha}$ is inconsistent).
(2) Milstein approximation: \( \mathbb{E} \left[ \frac{2\mu/\alpha}{\sigma_0^2} (\mu - \mu_0) - \left(1 - \frac{\sigma_0^2}{\sigma^2}\right) \frac{3\mu/\alpha\sigma^2}{2\sigma^2} \right] = 0, \)

(3) Euler approximation: \( \mathbb{E} \left[ \frac{2\mu/\alpha}{\sigma_0^2} (\mu - \mu_0) \right] = 0. \)

For (2) Milstein approx., if \( \sigma \neq \sigma_0 \), \( \hat{\alpha} \) is inconsistent.

But for (3) Euler approx., if \( \sigma \neq \sigma_0 \), \( \hat{\alpha} \) is consistent.

\( \Rightarrow \) **Robustness of Euler approx.**

If \( \sigma = \sigma_0 \), then all FOC are equal. \( \Rightarrow \) \( \hat{\alpha} \) is consistent for all (1)-(3).
Curvatures of $C^P$ and $C^S$ are Asymp Variances

For fixed $\Delta > 0$, Fisher information matrix $= \text{Second derivative of KLIC:}$

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left(0, \left(\frac{\partial^2 D_\Delta(\theta_0)}{\partial \theta \partial \theta^T}\right)^{-1}\right) \text{ as } T \rightarrow \infty.$$

Under $\Delta \rightarrow 0$ and $T \rightarrow \infty$, for correctly specified diffusions:

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N \left(0, \left(\frac{\partial^2 C^P(\beta_0)}{\partial \beta \partial \beta^T}\right)^{-1}\right) \quad \sqrt{T}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N \left(0, \left(\frac{\partial^2 C^S(\alpha_0; \beta_0)}{\partial \alpha \partial \alpha^T}\right)^{-1}\right)$$

For misspecified diffusions: No Bartlett equality, and asymp. variance is the sandwich matrix as in the discrete time framework.
Fiber Bundle Structure of Diffusion Model

- $\sigma$-orbit: the space of all diffusion functions with equal divergence in $C^P$ from the true diffusion function $\sigma_0$.
- Under $C^P$, the closest member of the diffusion functions in $B(\beta)$ is shown as $\beta^*$.
- $C^S$ measures the divergence from $(\mu_0, \sigma_0)$ to an element of $\pi^{-1}(\beta^*)$.
- The minimizer of $C^S$ is $\alpha^*$.

$\theta^* = (\alpha^*, \beta^*)$ is obtained by the sequential application of two projections.