Approximate Bayesian Computation for Maximum Likelihood Estimation in Hidden Markov Models (ABC MLE in HMM's)

Sinan Yıldırım*, Tom Dean, Sumeetpal S. Singh, Ajay Jasra

* Cambridge University Statistical Laboratory

Warwick
20 September 2012

Outline

ABC MLE approaches in HMMs

The Hidden Markov model (HMM)

HMM is a time series model comprised of with two processes

$$\{X_t \in \mathcal{X}, Y_t \in \mathcal{Y}\}_{t\geq 1}$$

$\{X_t\}_{t\geq 1}$ is the hidden Markov process with initial and transition densities μ_θ, f_θ

$$X_1 \sim \mu_\theta(x), \quad X_t|(X_{1:t-1} = x_{1:t-1}) \sim f_\theta(x|x_{t-1}),$$

$\{Y_t\}_{t\geq 1}$ is the conditionally independent observation process.

$$Y_t|(\{X_i\}_{i\geq 1} = \{x_i\}_{i\geq 1}, \{Y_i\}_{i\neq t} = \{x_i\}_{i\neq t}) \sim g_\theta(y|x_t).$$

We assume that the model is parametrised by $\theta \in \Theta \subseteq \mathbb{R}^d$, Θ compact.
The actual **observed data** is $\hat{Y}_1, \hat{Y}_2, \ldots$ assumed to be generated by $\theta^* \in \Theta$

ABC in HMMs

- Given $\hat{Y}_1, \hat{Y}_2, \ldots \hat{Y}_n$ generated from the HMM $\{X_t, Y_t\}_{t \geq 1}$, we seek for the MLE of θ^* which maximises the log-likelihood of the observations:

$$\theta_{\text{MLE}} = \arg\max_{\theta \in \Theta} \log p_\theta(\hat{Y}_{1:n})$$

$$p_\theta(\hat{Y}_{1:n}) = \int \mu_\theta(x_1) g_\theta(\hat{Y}_1|x_1) \prod_{t=2}^{n} f_\theta(x_t|x_{t-1}) g_\theta(\hat{Y}_t|x_t) dx_{1:t}.$$

- We are interested in MLE in HMMs where g_θ is intractable:
  - either not analytically available,
  - or prohibitive to calculate
- However, we can sample from $g_\theta(\cdot|x)$ as follows: There is a density ν_θ on $\mathcal{U}$ and a function $\tau_\theta : \mathcal{U} \times \mathcal{X} :\rightarrow \mathcal{Y}$ such that
  - Draw $U \sim \nu_\theta(\cdot|x)$, ν_θ rather simple.
  - $Y = \tau_\theta(U, x) \sim g_\theta(\cdot|x)$
- The assumption of ν_θ and τ_θ is the core of ABC.

Standard ABC MLE for HMM

The ABC approximation to the likelihood of $\hat{Y}_{1:n}$ for some fixed $\theta \in \Theta$ is

$$\mathbb{P}_\theta\left(Y_1 \in B^\epsilon_{\hat{Y}_1}, \ldots, Y_n \in B^\epsilon_{\hat{Y}_n}\right) \propto p^\epsilon_\theta(\hat{Y}_{1:n})$$
$$= \int_{\mathcal{X}^n} \mu_\theta(x_1) g^\epsilon_\theta(\hat{Y}_1|x_1)\left[\prod_{t=2}^n f_\theta(x_{t-1}, x_t) g^\epsilon_\theta(\hat{Y}_t|x_t)\right] dx_{1:n}$$

$B^\epsilon_y \subseteq \mathcal{Y}$ the ball around y with radius ϵ and the perturbed observation density is

$$g^\epsilon_\theta(y|x) = \frac{1}{|B^\epsilon_y|}\int g_\theta(u|x)\mathbb{I}_{B^\epsilon_y}(u)du$$

Standard ABC MLE: $\theta_{\text{ABC MLE},n} = \arg\max_{\theta\in\Theta} p^\epsilon_\theta(\hat{Y}_{1:n})$.

- We are maximising the likelihood of observations $\hat{Y}_{1:n}$ as if they were generated from the perturbed HMM $\{X_t, Y^\epsilon_t\}_{t\geq 1} = \{X_t, Y_t + \epsilon Z_t\}_{t\geq 1}$.
- This perturbed HMM $\{X_k, Y^\epsilon_k\}_{t\geq 1}$ has transitional laws f_θ and g^ϵ_θ.
- $\theta_{\text{ABC MLE},n}$ converges to a $\theta^\epsilon \neq \theta^*$, the bias is proportional to ϵ.

Noisy ABC MLE for HMM

- Standard ABC MLE maximises the likelihood $\hat{Y}_{1:n}$ under the law of the perturbed HMM $\{X_t, Y_t^\epsilon\}_{t\geq 1}$ with transitional laws f_θ and g_θ^ϵ although $\hat{Y}_{1:n}$ are generated from the HMM $\{X_t, Y_t\}_{t\geq 1}$ with transitional laws f_θ, g_θ.
- This model discrepancy is alleviated by Noisy ABC MLE:
  - Add noise to data:

$$\hat{Y}_t^\epsilon = \hat{Y}_t + \epsilon Z_t, \quad t = 1, \ldots, n \quad \text{where } Z_t \overset{\text{i.i.d.}}{\sim} \mathcal{U}_{B_0^1}$$

  - Maximise the likelihood of the noisy data

$$\theta_{\text{N-ABC MLE},n} = \arg\max_{\theta\in\Theta} p_\theta^\epsilon(\hat{Y}_{1:n}^\epsilon).$$

- Noisy ABC noted by Wilkinson (2008), Fearnhead and Prangle (2010)
- $\theta_{\text{N-ABC MLE},n}$ is asymptotically unbiased but statistically less efficient.

Two extensions

- Use of summary statistics: If the data sequence $\hat{Y}_1, \ldots, \hat{Y}_n$ is too high-dimensional, use

$$\mathcal{S}(\hat{Y}_1, \ldots, \hat{Y}_n) = S(\hat{Y}_1), \ldots, S(\hat{Y}_n)$$

for some function $S(\cdot)$ that maps from $\mathbb{R}^m \to \mathbb{R}^{m'}$, $m' < m$.

  - Markovian structure of the data is preserved.
  - If the mapping $S(\cdot)$ preserves the identifiability of the system, then conclusions continue to hold.

- Smoothed ABC: Using other types of noise can preserve the conclusions (and it is sometimes necessary!) In general,

$$\{X_t, Y_t^\epsilon\}_{t\geq 0} := \{X_t, Y_t + \epsilon Z_t\}_{t\geq 0}$$

where the $\{Z_t\}_{t\geq 0}$ are such that $Z_t \overset{i.i.d.}{\sim} \kappa$, κ is a centred kernel density, say Gaussian. Than the perturbed observation density will be

$$g_\theta^\epsilon(y|x) = \int g_\theta(u|x)\frac{1}{\epsilon}\kappa\left[\frac{1}{\epsilon}(y-u)\right] du.$$

Implementing ABC MLE

Implementing ABC MLE

Recall

- Standard ABC MLE: Maximise $p_\theta^\epsilon(\hat{Y}_{1:n})$ under the law of HMM $\{X_t, Y_t^\epsilon\}_{t\geq 1}$ with transitional laws f_θ and g_θ^ϵ.
- Noisy ABC MLE: Maximise $p_\theta(\hat{Y}_{1:n}^\epsilon)$ under the law of HMM $\{X_t, Y_t^\epsilon\}_{t\geq 1}$ with transitional laws f_θ and g_θ^ϵ.

But how to implement these ABC MLE ideas?
ABC is based on sampling Y_t. So, how about working with the HMM $\{(X_t, Y_t), Y_t^\epsilon\}_{t\geq 1}$ with

- either $Y_t^\epsilon = \hat{Y}_t$ (ABC MLE)
- or $Y_t^\epsilon = \hat{Y}_t^\epsilon$ (noisy ABC MLE)

This HMM may be OK for SMC filtering (by sampling Y_t we can get rid of having to calculating g_θ); but its transitional law is still intractable - hence of limited use.

Implementing ABC MLE

Recall the intermediate variables U_t and $\tau_\theta(U_t, x_t)$ to sample from $g_\theta(\cdot|x_t)$. Construct the modified HMM $\{R_t = (X_t, U_t), Y_t^\epsilon\}_{t\geq 1}$ where

$$U_t \overset{\text{i.i.d}}{\sim} \nu_\theta, \quad Y_t^\epsilon = \tau_\theta(X_t) + \epsilon Z_t, \quad Z_t \overset{\text{i.i.d}}{\sim} \kappa.$$

- This HMM has tractable initial and transitional densities π_θ, q_θ and h_θ:

$$\pi_\theta(r) = \mu_\theta(x)\nu_\theta(u|x), q_\theta(r'|r) = f_\theta(x'|x)\mu_\theta(u'|x'),$$
$$h_\theta(y|r) = \frac{1}{\epsilon}\kappa\left[\frac{1}{\epsilon}(y - \tau_\theta(r))\right].$$

- Can be shown that MLE for this HMM indeed maximises the ABC likelihood

$$p_\theta^\epsilon(Y_{1:n}^\epsilon) = \int_{(\mathcal{X}\times\mathcal{U})^n} \pi_\theta(r_1)h_\theta^\epsilon(Y_1^\epsilon|r_1)\prod_{t=2}^n q_\theta(r_t|r_{t-1})h_\theta^\epsilon(Y_t^\epsilon|r_t)dr_{1:n}$$

  - If $Y_{1:n}^\epsilon = \hat{Y}_{1:n} \Rightarrow$ Standard ABC MLE,
  - If $Y_{1:n}^\epsilon = \hat{Y}_{1:n}^\epsilon \Rightarrow$ Noisy ABC MLE

Implementing ABC MLE: Gradient ascent MLE

Given $Y^{\epsilon}_{1:n} = y_{1:n}$, we want to maximise $p^{\epsilon}_{\theta}(Y^{\epsilon}_{1:n})$ for the HMM $\{R_t = (X_t, U_t), Y^{\epsilon}_t\}_{t \geq 1}$ whose law is given by

$$\pi_\theta(r) = \mu_\theta(x)\nu_\theta(u|x), \quad q_\theta(r'|r) = f_\theta(x'|x)\mu_\theta(u'|x'), \quad g_\theta(y|r) = \frac{1}{\epsilon}\kappa\left[\frac{1}{\epsilon}(y - \tau_\theta(r))\right].$$

Gradient ascent MLE: Given the estimator θ_i at iteration i,

$$\theta_{i+1} = \theta_i + \gamma_i \nabla_\theta \log p^{\epsilon}_{\theta}(y^{\epsilon}_{1:n})|_{\theta=\theta_i}$$

Step size sequence $\{\gamma_i\}_{i\geq 0}$ satisfy $\sum_i \gamma_i = \infty$, $\sum_i \gamma_i^2$.
By Fisher's identity

$$\nabla_\theta \log p^{\epsilon}_{\theta}(y_{1:n}) = \mathbb{E}_\theta\left[\sum_{t=1}^{n} \underbrace{\nabla_\theta \log q_\theta\left(R_t|\, R_{t-1}\right) + \nabla_\theta \log h_\theta\left(Y^{\epsilon}_t|R_t\right)}_{s_t(R_{t-1},R_t)} \middle| Y^{\epsilon}_{1:n} = y_{1:n}\right]$$

Note: $\nabla_\theta \log h_\theta\left(y|r\right) = \nabla_\theta \log \kappa\left[\frac{1}{\epsilon}(y - \tau_\theta(r))\right]$ requires smooth κ.

Online calculation of the gradient

$$\nabla \log p_\theta^\epsilon(y_{1:n}) = \mathbb{E}_\theta \left[\sum_{t=1}^{n} \underbrace{\nabla \log q_\theta\left(R_t | R_{t-1}\right) + \nabla \log h_\theta\left(Y_t^\epsilon | R_t\right)}_{s_t(R_{t-1}, R_t)} \middle| Y_{1:n}^\epsilon = y_{1:n} \right]$$

Online smoothing of $S_{\theta,n}(r_{1:n}) = \sum_{i=1}^{n} s_\theta(r_{i-1}, r_i)$ is available via a recursion (Del Moral et al, 2009)

$$\begin{aligned} T_{\theta,n}(r_n) &= \mathbb{E}_\theta \left[S_{\theta,n}(R_{1:n}) | R_n = r_n, Y_{1:n}^\epsilon = y_{1:n} \right] \\ &= \mathbb{E}_\theta \left[T_{\theta,n-1}(R_{n-1}) + s_\theta(R_{n-1}, r_n) | Y_{1:n-1}^\epsilon = y_{1:n-1} \right] \qquad (1) \end{aligned}$$

$$\nabla \log p_\theta^\epsilon(y_{1:n}) = \mathbb{E}_\theta \left[T_{\theta,n}(R_n) | Y_{1:n}^\epsilon = y_{1:n} \right] \qquad (2)$$

- Equation (1) requires integration w.r.t $p_\theta(dr_{n-1}|y_{1:n-1}, r_n) = \frac{p_\theta(dr_{n-1}|y_{1:n-1}) f_\theta(r_n|r_{n-1})}{\int p_\theta(dr_{n-1}|y_{1:n-1}) f_\theta(r_n|r_{n-1}) dr_{n-1}}$.
- Equation (2) requires integration w.r.t. $p_\theta(dr_n|y_{1:n})$

SMC approximation to the gradient

- Exact calculation of forward smoothing recursion for is rarely the case.
- A stable SMC approximation are available:
  Assume we run a particle filter for the HMM $\{R_n, Y_n^\epsilon\}_{n\geq 1}$ to obtain approximations to $\{p_\theta(dr_n|y_{1:n})\}_{n\geq 1}$

$$p_\theta^N(dr_n|y_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{R_n^{(i)}}(dr_n), \quad \sum_{i=1}^{N} W_n^{(i)} = 1.$$

  At time we calculate
  - for $i = 1, \ldots, N$

$$T_n^{(i)} = \frac{\left[T_{n-1}^{(j)} + s(R_{n-1}^{(j)} + R_n^{(i)})\right] W_{n-1}^{(j)} f_\theta(R_n^{(i)}|R_n^{(j)})}{\sum_{j'=1}^{N} W_{n-1}^{(j')} f_\theta(R_n^{(i)}|R_n^{(j')})}$$

  - $\nabla^N \log p_\theta^\epsilon(y_{1:n}) = \sum_{i=1}^{N} T_n^{(i)} W_n^{(i)}$
- This algorithm requires $\mathcal{O}(N^2)$ calculations per time n.

Online gradient ascent MLE

- The batch gradient ascent MLE algorithm may be inefficient when n is large
- An alternative to the batch algorithm is online gradient ascent MLE. (Del Moral et al (2011), Poyiadjis et al (2011)): Given $y_{1:n-1}$, assume we have the estimate θ_{n-1}. When y_n is received, we update

$$\theta_n = \theta_{n-1} + \gamma_n \nabla_\theta \log p_\theta(y_n|y_{1:n-1})\Big|_{\theta=\theta_{n-1}}.$$

  - One SMC approximation of $\nabla_\theta \log p_\theta(y_n|y_{1:n-1})$ is (Poyiadjis et al, 2011)

$$\nabla_\theta^N \log p_\theta(y_n|y_{1:n-1}) = \nabla_\theta^N \log p_\theta(y_{1:n}) - \nabla_\theta^N \log p_\theta(y_{1:n-1})$$

  - A (slightly) different approximation uses the filter derivate (Del Moral et al, 2011):
  - Stability of the SMC online gradient ascent algorithm is demonstrated (Del Moral et al, 2011).

Special case: i.i.d. random variables

- $\{Y_k\}_{k\geq 1}$ are i.i.d. w.r.t. $g_\theta(\cdot)$.
- We generate $U \in \mathcal{U}$ from μ_θ, and by applying a transformation function $\tau_\theta : \mathcal{U} \to \mathcal{Y}$ so that $\tau_\theta(U) \sim g_\theta$.
- $p_\theta(Y_n^\epsilon | Y_{1:n-1}^\epsilon) = p_\theta(Y_n^\epsilon) \Rightarrow$ the batch and online update rules reduce to
  - Batch gradient ascent: Given $Y_{1:n}^\epsilon = y_{1:n}$, at iteration i

$$\theta_i = \theta_{i-1} + \gamma_i \sum_{t=1}^{n} \nabla_\theta \log p_\theta(y_t)\Big|_{\theta=\theta^{(i-1)}}$$

  - Online gradient ascent: when $Y_n^\epsilon = y_n$ is received

$$\theta_n = \theta_{n-1} + \gamma_n \nabla_\theta \log p_\theta(y_n)\Big|_{\theta=\theta_{n-1}}.$$

$$\nabla_\theta \log p_\theta(y) = \int \left[\nabla_\theta \log \nu_\theta(u) + \nabla_\theta \log h_\theta(y|u)\right] p_\theta(u|y_n) du, \qquad (3)$$

- The original $\mathcal{O}(N^2)$ algorithm reduces to an $\mathcal{O}(N)$ algorithm $\Rightarrow$ one can use Monte Carlo with more samples.

Controlling stability of the gradient

- If the additional gradients $\nabla_\theta \log q_\theta(r'|r)$ or $\nabla_\theta \log h_\theta(y|r)$ have very high or infinite variances; we expect failure of gradient ascent MLE (e.g. α-stable).
- In particular, assuming $\kappa = \mathcal{N}(0,1)$,

$$\nabla_\theta \log h_\theta(Y^\epsilon|R) = \frac{1}{\epsilon^2}(Y^\epsilon - \tau_\theta(R))\nabla_\theta \tau_\theta(R)$$

  To overcome this problem, we can transform the observations $\hat{Y}_k$ using a one-to-one differentiable function $\psi : \mathcal{Y} \to \mathcal{Y}_\psi$.
- Then, we perform ABC MLE for $\{(X_t, U_t), Y_t^{\psi,\epsilon}\}_{t\geq 1}$ where this time

$$Y_t^{\psi,\epsilon} = \psi(Y_t) + \epsilon Z_t, \quad Z_t \sim^{\text{i.i.d.}} \mathcal{N}(0,1), \quad t \geq 1.$$

- In this case, $h_\theta(y|r)$ changes to $h_\theta^\psi(y|r) = \mathcal{N}(y; \psi(\tau_\theta(r)), \epsilon^2)$.
- We choose ψ such that the gradient of the new log-observation density

$$\nabla_\theta \log h_\theta^\psi(Y^{\psi,\epsilon}|R) = \frac{1}{\sigma_\epsilon^2}\left[Y^{\psi,\epsilon} - \psi(\tau_\theta(R))\right] \nabla_\theta \psi(\tau_\theta(R))$$

  has smaller variance than it would have if no transformation were used.
- In noisy ABC MLE, we obtain the noisy data by $\hat{Y}_t^{\epsilon,\psi} = \psi(\hat{Y}_t) + \epsilon Z_t$.

Numerical examples

MLE for α-stable distribution

$\mathcal{A}(\alpha, \beta, \mu, \sigma)$ is the α-stable distribution where its parameters,

$$\theta = (\alpha, \beta, \mu, \sigma) \in \Theta = (0, 2] \times [-1, 1] \times \mathbb{R} \times [0, \infty),$$

are the shape, skewness, location, and scale parameters, respectively. We generate from $\mathcal{A}(\alpha, \beta, \mu, \sigma)$ by sampling $U = (U^{(1)}, U^{(2)})$, where $U^{(1)} \sim \text{Unif}(-\pi/2, \pi/2)$ and $U^{(2)} \sim \text{Exp}(1)$ independently, and setting

$$Y := \tau_\theta(U) := \sigma t_{\alpha,\beta}(U) + \mu.$$

$$t_{\alpha,\beta}(U) = \begin{cases} S_{\alpha,\beta} \frac{\sin(\alpha(U^{(1)} + B_{\alpha,\beta}))}{(\cos(U^{(1)}))^{1/\alpha}} \left(\frac{\cos(U^{(1)} - \alpha(U^{(1)} + B_{\alpha,\beta}))}{U^{(2)}} \right)^{(1-\alpha)/\alpha}, & \alpha \neq 1 \\ X = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta U^{(1)} \right) \tan U^{(1)} - \beta \log \left(\frac{U^{(2)} \cos U^{(1)}}{\frac{\pi}{2} + \beta U^{(1)}} \right) \right], & \alpha = 1. \end{cases}$$

$$B_{\alpha,\beta} = \frac{\tan^{-1}\left(\beta \tan \frac{\pi\alpha}{2}\right)}{\alpha} \quad S_{\alpha,\beta} = \left(1 + \beta^2 \tan^2 \frac{\pi\alpha}{2}\right)^{1/2\alpha}$$

MLE for α-stable distribution

- We want to use noisy ABC MLE with Gaussian κ for $\mathcal{A}(\alpha, \beta, \mu, \sigma)$
- Variance of the $\mathcal{A}(\alpha, \beta, \mu, \sigma)$ is infinity, unless $\alpha = 2$; hence the gradients $\nabla_\theta \log h_\theta(\hat{Y}_t^\epsilon | U_t)$ are not stable when $\hat{Y}_t^\epsilon = \hat{Y}_t + \sigma_\epsilon Z_t$.
- Instead, we propose using the transformation $\psi = \tan^{-1}$ to have

$$\hat{Y}_t^{\epsilon,\psi} = \tan^{-1}(\hat{Y}_t) + \epsilon Z_t$$

to make the gradient ascent algorithm stable. Then we have

$$h_\theta^\psi(y|u) = \mathcal{N}(y; \tan^{-1}[\tau_\theta(u)], \epsilon^2)$$
$$\nabla_\theta \log h_\theta^\psi(y|u) = \frac{1}{\epsilon^2}\left[y - \tan^{-1}[\tau_\theta(u)]\right] \frac{\nabla_\theta \tau_\theta(u)}{1 + \tau_\theta(u)^2}.$$

online gradient ascent noisy ABC MLE for α-stable

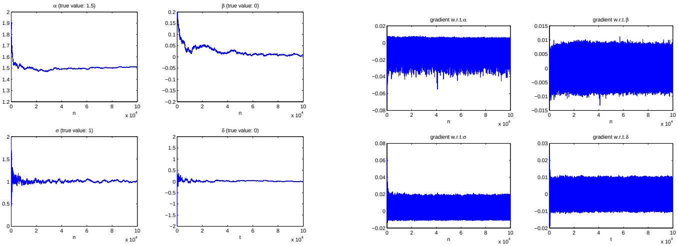

Figure: On the left: Online estimation of α-stable parameters from a sequence of i.i.d. random variables transformed with $\tan^{-1}(\cdot)$ using online gradient ascent noisy ABC MLE. True parameters are $\theta = (1.5, 0, 0, 1)$. On the right: Gradient of incremental likelihood for the α-stable parameters.

ABC MLE vs Noisy ABC MLE: Bias

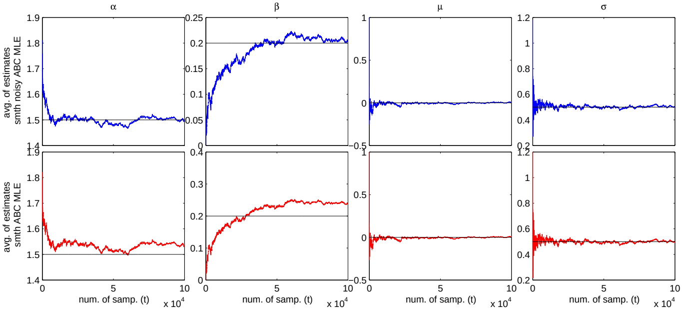

Figure: Online gradient ascent estimates (averaged over 50 runs) using noisy smoothed ABC MLE and smoothed ABC MLE. For the noisy smoothed ABC MLE, a different noisy data sequence is used in each run. True parameters $(\alpha, \beta, \mu, \sigma) = (1.5, 0.5, 0, 0.5)$ are indicated with a horizontal line.

MLE for *g-and-k* distribution

The *g-and-k* distribution is determined by (A, B, g, k, c) and is defined by its quantile function Q_θ, which is the inverse of the cumulative distribution function F_θ

$$Q_\theta(u) = F_\theta^{-1}(u) = A + B\left[1 + c\frac{1 - e^{-g\phi(u)}}{1 + e^{-g\phi(u)}}\right]\left(1 + \phi(u)^2\right)^k \phi(u), \quad u \in (0, 1).$$

where $\phi(u)$ is the u'th standard normal quantile. The parameters $\theta = (g, k, A, B) \in \Theta = \mathbb{R} \times (-0.5, \infty) \times \mathbb{R} \times [0, \infty)$ are the skewness, kurtosis, location, and scale parameters, and c is usually fixed to 0.8. Note that Q_θ in is differentiable w.r.t. θ, so the gradient ascent algorithms are applicable.

Online gradient ascent noisy ABC MLE for *g-and-k* distribution

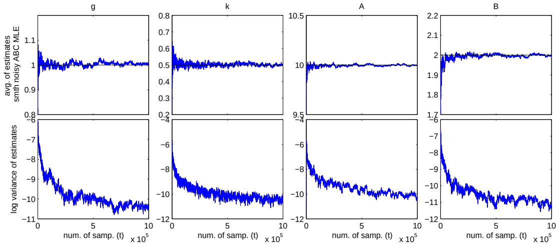

Figure: Mean and the variance (over 50 runs) of online gradient ascent estimates using noisy ABC MLE. Same noisy data sequence is used in each run. True parameters $(g, k, A, B) = (2, 0.5, 10, 2)$ are indicated with a horizontal line.

Batch gradient ascent noisy ABC MLE for *g-and-k* distribution

Batch gradient ascent ABC MLE algorithm on 500 data sets of $n = 1000$ i.i.d. samples from the same *g-and-k* distribution.

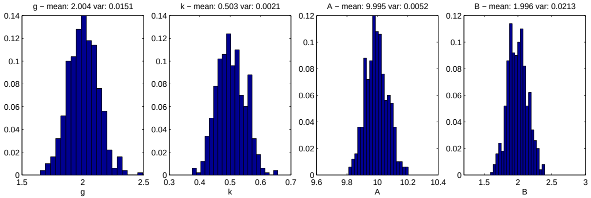

Figure: Approximate distributions (histograms over 20 bins) of the estimates for 500 different data sets with $\theta = (2, 0.5, 10, 2)$

The mean and variance of the MLE estimates for (g, k, A, B) are $(2.004, 0.503, 9.995, 1.996)$ and $(0.0151, 0.0021, 0.0052, 0.0213)$ respectively.

HMM example: The stochastic volatility model with symmetric α-stable returns

- The model for $\{X_t, Y_t\}_{t\geq 1}$ is:

$$X_1 \sim \mathcal{N}(0, \frac{\sigma_x^2}{1-\phi^2}), \quad X_k = \phi X_{k-1} + \sigma_x V_k, \quad V_k, \sim \mathcal{N}(0,1), \quad k \geq 2,$$
$$Y_k|(X_k = x_k) \sim e^{x_k/2}\mathcal{A}(\alpha, 0, 0, 1), \quad t \geq 1. \tag{4}$$

- Noisy ABC MLE for $\hat{Y}_k^{\epsilon} = \tan^{-1}(\hat{Y}_k) + \epsilon Z_k, \quad Z_k \sim \mathcal{N}(0,1)$.
- The densities π_θ, q_θ, and h_θ of the HMM $\{R_k = (X_k, U_k), Y_k^{\epsilon}\}_{k\geq 1}$

$$\pi_\theta(r) = \mathcal{N}(x; 0, \sigma_x^2/(1-\phi^2))\frac{1}{\pi}\mathbb{I}_{[-\pi/2,\pi/2]}(u^{(1)})\mathbb{I}_{[0,\infty)}(v)e^{-u^{(2)}},$$
$$q_\theta(r'|r) = \mathcal{N}(x'; \phi x, \sigma_x^2)\frac{1}{\pi}\mathbb{I}_{[-\pi/2,\pi/2]}(u'^{(1)})\mathbb{I}_{[0,\infty)}(u'^{(2)})e^{-u'^{(2)}},$$
$$h_{\theta,\psi}(y|r) = \mathcal{N}(y; \tan^{-1}(e^{x/2}t_{\alpha,0}(u)), \sigma_\epsilon^2),$$

HMM example: The stochastic volatility model with symmetric α-stable returns: online gradient ascent noisy ABC MLE

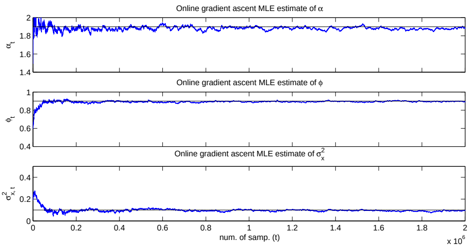

Figure: Online estimation of SVαR parameters using online gradient ascent algorithm to implement noisy ABC MLE. True parameter values $\theta = (1.9, 0.9, 0.1)$ are indicated with a horizontal line.

HMM example: The stochastic volatility model with symmetric α-stable returns: online EM noisy ABC MLE (α known)

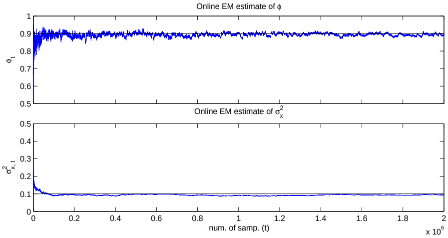

Figure: Online estimation of SVαR parameters using the online EM algorithm to implement noisy ABC MLE. True parameter values $\theta = (1.9, 0.9, 0.1)$ are indicated with a horizontal line.

Conclusion

Conclusion

- Noisy ABC MLE for HMMs is a consistent method for parameter estimation
- SMC implementations are practical: Gradient ascent, EM, etc.
- Stability should be concerned.
- Future work: sharper error analysis, new methods, new applications ...?