Parameter estimation for Markov processes killed at a threshold

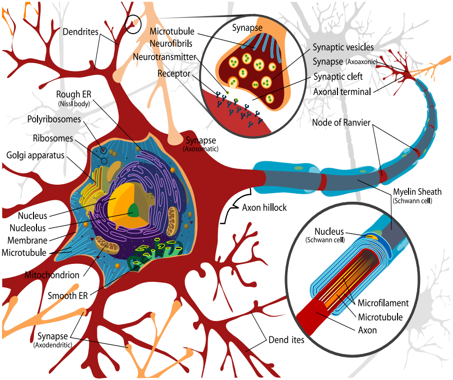
Enrico Bibbona

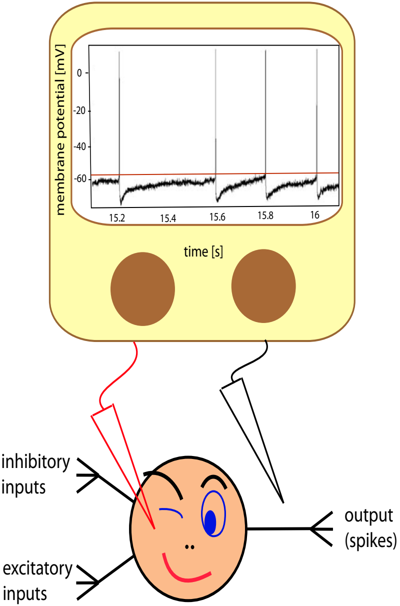
Università di Torino

Angers, 18 June 2010

joint work with S. Ditlevsen (Copenhagen University)

The origin of the problem





A neuron...

... and its membrane potential

The problem: discrete time

The original process

Let X_i be a discrete-time homogeneous Markov chain on $E \subset \mathbb{R}$ (continuous) with transition density

$$f_\theta(x_i|x_{i-1})$$

for given $\theta \in \Theta \subset \mathbb{R}^p$. $x_0 < b$ fixed. We assume $\boxed{\mathbb{E}(T_{x_0,b}) < \infty}$

The problem: discrete time

The original process

Let X_i be a discrete-time homogeneous Markov chain on $E \subset \mathbb{R}$ (continuous) with transition density

$$f_\theta(x_i|x_{i-1})$$

for given $\theta \in \Theta \subset \mathbb{R}^p$. $x_0 < b$ fixed. We assume $\boxed{\mathbb{E}(T_{x_0,b}) < \infty}$

The killed process X^k agrees with X up to first passage time of X to b. At that time it goes to C (and there it stays forever). The state space of X^k is $E_b \cup C$ and its transition is

$$f_\theta^k(y|x) = f_\theta(y|x) \cdot \mathbb{1}_{\{x,y \in E_b\}} + p \cdot \mathbb{1}_{\{y=C\}}$$

with

$$p = \int_{E\cap(b,\infty)} f_\theta(z|x)\,dz \cdot \mathbb{1}_{\{x\in E_b\}} + \mathbb{1}_{\{x=C\}} = 1 - \int_{E_b} f_\theta(z|x)\,dz$$

The problem: discrete time

The original process
Let X_i be a discrete-time homogeneous Markov chain on $E \subset \mathbb{R}$ (continuous) with transition density

$$f_\theta(x_i|x_{i-1})$$

for given $\theta \in \Theta \subset \mathbb{R}^p$. $x_0 < b$ fixed. We assume $\boxed{\mathbb{E}(T_{x_0,b}) < \infty}$

The killed process X^k agrees with X up to first passage time of X to b. At that time it goes to C (and there it stays forever). The state space of X^k is $E_b \cup C$ and its transition is

$$f_\theta^k(y|x) = f_\theta(y|x)\cdot\mathbb{1}_{\{x,y\in E_b\}} + \int_{E\cap(b,\infty)} f_\theta(y|x)\,dy\cdot\mathbb{1}_{\{y=C,\,y<b\}} + \mathbb{1}_{\{y=x=C\}}$$

The likelihood function for n observations from X^k is

$$L_n(X_1,\cdots,X_n,\theta) = \prod_{i=1}^{n} f_\theta^k(X_i|X_{i-1})$$

Example: a simple RW

Let $T_{0,C} = \min\{i : X_i^k = C\}$. We have n observations. The following relations hold

$$U + D = \min\{T_{0,C}, n\} \qquad U - D = \min\{b, X_n^k\}$$

adopting the convention $\min\{b, C\} = b$, and the joint density is

$$L_n(X_1^k, \cdots, X_n^k, \theta) = p^U (1-p)^D$$

and the Maximum Likelihood estimator is

$$\hat{p}_n = \frac{U}{U+D} = \frac{\min\{b, X_n^k\}}{2\min\{T_{0,C}, n\}} + \frac{1}{2}$$

that for $n \longrightarrow \infty$ does not converge to p but to the r.v.

$$\boxed{\hat{p}_\infty = \frac{b}{2\,T_{0,C}} + \frac{1}{2}}$$

Example: a simple RW

with

$$\mathbb{E}(\hat{p}_\infty) = (p^b/b)\mathrm{Hyperg}([b/2, b/2, (b+1)/2], [b+1, b/2+1], 4p-4p^2)$$

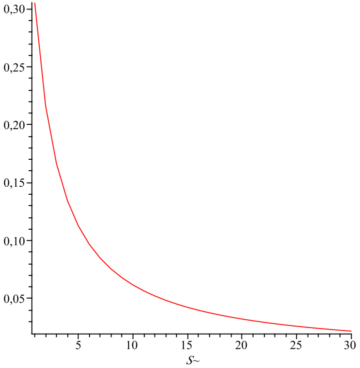


Figure: Relative bias $[\mathbb{E}(\hat{p}_\infty) - p]/p$, plotted against b with $p = 2/3$.

- keeping on observing after the fpt does not add any info
- the ML estimate has an asymptotic bias

Discrete observations of a diffusion

The original process

Let X_t be a solution of the following SODE

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t$$

with state space $E \subset \mathbb{R}$ and transition density

$$f_\theta(x_t|x_s)$$

for given $\theta \in \Theta \subset \mathbb{R}^p$. $x_0 < b$ fixed. We assume $\boxed{\mathbb{E}(T_{x_0,b}) < \infty}$

Discrete observations of a diffusion

The original process

Let X_t be a solution of the following SODE

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t$$

with state space $E \subset \mathbb{R}$ and transition density

$$f_\theta(x_t|x_s)$$

for given $\theta \in \Theta \subset \mathbb{R}^p$. $x_0 < b$ fixed. We assume $\boxed{\mathbb{E}(T_{x_0,b}) < \infty}$

The killed process X^k agrees with X up to first passage time of X through b. At that time it goes to C (and there it stays forever). The state space of X^k is $E_b \cup C$ and its transition is

$$f_\theta^k(x_t|x_s) = f_\theta^b(x_t|x_s) \cdot \mathbb{1}_{\{x_s,x_t \in E_b\}} + G(t-s|x) \cdot \mathbb{1}_{\{x_t=C,\, x_s \in E_b\}} + \mathbb{1}_{\{x_t=x_s=C\}}$$

In other words...

Given that at time s you are in x_s, at time t the process may

- have crossed b with probabililty

$$P_\theta(T \leq (t-s)|x_s) = G_\theta^b(t-s|x) = \int_0^{t-s} g_\theta^b(r|x_s)\,dr$$

- have reached position x_t having been below the threshold for all the time between t and s with a probability density

$$f_\theta^b(x_n|x_n-1) = \frac{\partial}{\partial y} P_\theta(X_t \leq y \;\&\; T > h \,|\, x_{n-1})\Big|_{y=x_n}$$

Discrete observations of a diffusion

Let $(X_1, \cdots, X_n)$ be a sample of observations from a process X^k at times $t_i = ih$. The joint density of the sample is the following likelihood function

$$L_n(X_1, \cdots, X_n, \theta) = \prod_{i=1}^{n} f_\theta^k(X_i|X_{i-1})$$

thus if we observe $(x_1, \cdots, x_p, C, \cdots)$ to estimate θ we minimize

$$f_\theta^b(x_1|x_0) \cdots f_\theta^b(x_p|x_{p-1}) \cdot G_\theta^b(h|x_p)$$

How to calculate $f_\theta^b(x_i|x_{i-1})$?

1. $f_\theta^b(x_i|x_{i-1})$ is know in closed form for brownian motion and a few more cases, it is a joint "probability density" of being in $x_i < b$ and not passing, it may be calculated conditioning:

$$f_\theta^b(x_i|x_{i-1}) = f_\theta(x_i|x_{i-1})(1 - p_i)$$

   where $p_i = P(t_{i-1} < T \leq t_i|x_{i-1}, x_i)$ is the crossing probability for a Bridge process (pinned diffusion)

2. for the probability p_i Caramellino and Baldi (Ann. Appl. Prob., 2005) proposed a large deviation estimate when $h = t_i - t_{i-1} \to 0$ in a very general context
3. for the OU process its expression is compact and looks like the following

$$p_i \approx \exp\left[-\frac{2}{h}(b - x_{i-1})(b - x_i)\right]\Big(1 - h\,\phi(x_{i-1}, x_i, b)\Big)$$

   where ϕ is a rational function.

How to calculate $P_\theta(T \leq h|x_{n-1})$?

$$P_\theta(T \leq h|x_{n-1}) = \int_0^h g_b(r|x_{n-})\,dr$$

The density $g_b(r|x_s)$ of the first passage time satisfies (cf. Buonocore et al., Adv. Appl. Prob. 19, 1987) the following integral equation with regular kernel

$$g_b(r|x_{n-1}) = -2\Psi_b(r|x_{n-1}) + 2\int_0^r d\tau\; g_b(\tau|x_{n-1})\,\Psi_b(r-\tau|b)$$

for a known function $\Psi_b(r|x_{n-1})$.
For $h \to 0$, a not too crude approximation is

$$g_b(r|x_{n-1}) \approx -2\Psi_b(r \mid x_{n-1})$$

and

$$P_\theta(T \leq h|x_{n-1}) = \int_0^h \Psi_b(r \mid x_{n-1})\,dr$$

Example: an OU process

X_t solves

$$dX_t = (-\beta X_t + \mu)\, dt + \sigma\, dW_t$$

we choose

$$N = 10000,\ h = 0.12,\ x_0 = 0, b = 5$$

An exact discetization:

$$X_n = \mathrm{e}^{-\beta h} X_{n-1} + \frac{\mu}{\beta}(1 - \mathrm{e}^{-\beta h}) + Z_n \qquad Z_n \sim N\left(0, \frac{\sigma^2}{2\beta}(1 - \mathrm{e}^{-2\beta h})\right)$$

Results

method	μ	β	σ
true values	4	1	1.5
"naive"	5.24	1.27	1.42
AR(1) approx	4.50	0.89	1.45
full OU model	4.81	1.04	1.43

We improved, but there is still a BIAS

A way out?

If $\left\{(x_0^{(i)}, x_1^{(i)}, \cdots, x_{n_i-1}^{(i)}, C, \cdots)\right\}_{i=1\cdots N}$ are N independent realizations of a killed simple RW that are fully observed up to the killing. I can glue them together as a single trajectory

$$(C, x_1^{(1)}, \cdots, x_{n_i-1}^{(1)}, C, x_1^{(2)}, \cdots, x_{n_i-1}^{(i)}, C, \cdots)$$

of a regenerating process Y_i that restarts with the same law each time it hits C.
The likelihood of this N trajectories collected together is

$$L_N p = p^{\sum_i U_i}(1-p)^{\sum_i D_i}$$

The maximum likelihood estimator is

$$\hat{p}^b = \frac{\sum_{i=1}^N U_i}{\sum_{i=1}^N T_{C,C}^i} = \frac{Nb}{2\sum_{i=1}^N T_{C,C}^i} + \frac{1}{2}.$$

and by standard LLN it converges to

$$\hat{p}_\infty^b = \frac{b}{2\mathbb{E}(T_{C,C}^i)} + \frac{1}{2} = p$$

If $\left\{(x_0^{(i)}, x_1^{(i)}, \cdots, x_{n_i-1}^{(i)}, C, \cdots)\right\}_{i=1\cdots N}$ are N independent realizations, I can glue them together as a single trajectory

$$(C, x_1^{(1)}, \cdots, x_{n_i-1}^{(1)}, C, x_1^{(2)}, \cdots, x_{n_i-1}^{(i)}, C, \cdots)$$

of a regenerating process Y_i which is a Markov chain with transitions

$$f_\theta^Y(y_i|y_{i-1}) = f_\theta^b(y_i|y_{i-1}) \cdot \mathbb{1}_{\{y_i \in E_b\}} + G_\theta^b(h|y_{i-1}) \cdot \mathbb{1}_{\{y_i = C\}},$$

where $f_\theta^b(y|C)$ is interpreted as $f_\theta^b(y|y_0)$ And the likelihood function for n observations of Y_i is known as

$$L_n(Y_1, \cdots, Y_n, \theta) = \prod_{i=1}^{n} f_\theta^Y(Y_i|Y_{i-1})$$

no matter how the observations split into many trajectories of X^k

Estimating functions framework

We are going to prove consistency and asymptotic normality of the MLE for the regenerating process in the framework of estimating functions theory. The score function is just a particular case of an estimating function of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(Y_i, Y_{i-1}, \theta)$$

from which estimators are found solving $G_n(\theta) = 0$.
Estimators solving $G_n(\theta) = 0$ are consistent and asymptotically normal under the following assumptions

- the function g is regular enough
- $G_n(\theta)$ is a martingale
- the Markov chain Y_i allows for a stationary distribution π
- $\pi(\cdot) >> f^Y(\cdot|y)$ for any y in the state space

(Sorensen 1999, Billingsley 1961, ...)

The stationary measure

The probability density of going from C to y in n steps without being killed is

$$_n f^b(y_n|C) = \int_{E_b}\int_{E_b}\cdots\int_{E_b} f^Y(y_1|C)\cdots f^Y(y_n|y_{n-1})\,dy_1\cdots dy_{n-1}$$

(it also holds for $y_n = C$, giving the density of the fpt)

Theorem

If $\mathbb{E}(T_{C,C}) < \infty$, the probability density

$$\pi(y) = \frac{\sum_{n=1}^{\infty} {}_n f^b(y|C)}{\mathbb{E}(T_{C,C})} \qquad (1)$$

is the density of the invariant measure for Y_i

Interpratation: $\pi(A)$ is the probability of ever reaching the set A!
$\pi(y) = 0$ if and only if $_n f^b(y|C) = 0\ \forall n$ so that y cannot be reached at all! We can restrict to $E_C^+ = \{x \in E_b \cup C : \pi(x) > 0\}$

Some definitions

Let us define Q as the measure on $E_C^+ \times E_C^+$ with density

$$Q(x, y) = \pi(x) f^Y(y|x).$$

A function $h : E_C^+ \times E_C^+ \times \Theta \longrightarrow \mathbb{R}^p$ is said to be *locally dominated integrable* w.r.t. Q if for each $\theta' \in \Theta$ there exists a neighbourhood $U_{\theta'}$ of θ' and a non-negative Q-integrable function $g_{\theta'} : E_C^+ \times E_C^+ \longrightarrow \mathbb{R}^p$ such that $|h(x, y, \theta)| \leq g_{\theta'}(x, y)$ for all $(x, y, \theta) \in E_C^+ \times E_C^+ \times U_{\theta'}$.

Theorem: Consistency and AN

Under the following conditions

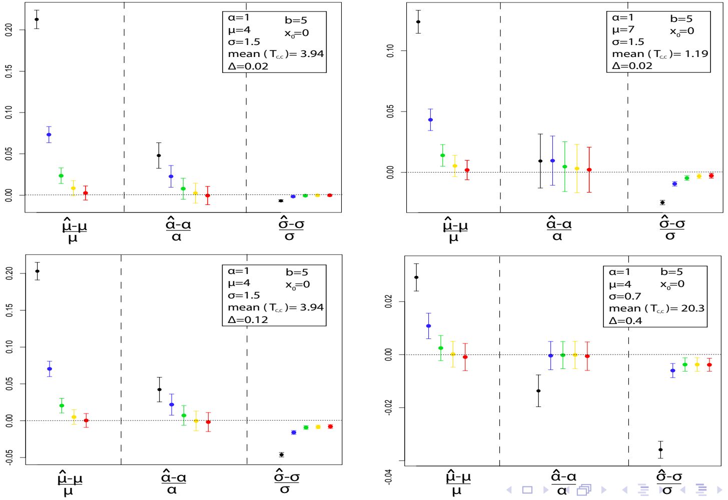
1. The function g is twice continuously differentiable with respect to θ for all $x, y \in E_C^+$
2. the functions $(x,y) \longmapsto g_i(x,y;\theta)$, $(x,y) \longmapsto \partial_{\theta_j} g_i(x,y;\theta)$ and $(x,y) \longmapsto \partial_{\theta_k}\partial_{\theta_j} g_i(x,y;\theta)$ are all locally dominated integrable w.r.t. Q for all $i,j,k \in \{1 \cdots p\}$. Moreover the functions $(x,y) \longmapsto g_i(x,y;\theta)$ are in $L_2(Q)$ for every $i \in \{1 \cdots p\}$ and for all $\theta \in \Theta$
3. the $p \times p$ matrix J_Q with entries $\mathbb{E}_Q\big(\partial_{\theta_j} g_i(\theta_0)\big)$ is invertible
4. G_n is a martingale estimating function
5. $\mathbb{E}(T_{C,C}) < \infty$ and the state space of the process Y does not contain *unreachable points* with $\pi(y) = 0$

for every n an estimator $\hat{\theta}_n$ exists that solves the estimating equation $G_n(\hat{\theta}_n) = 0$ with a probability tending to one as $n \longrightarrow \infty$.
Moreover $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$ as $n \longrightarrow \infty$ and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, J_Q^{-1} V J_Q) \quad \text{with } V = \mathbb{E}_Q\big(g_i^T(\theta_0) g_i(\theta_0)\big)$$

Example

Again an OU Process, we generate 10000 trajectories and we collect them into groups of size 1, 3, 10, 30, 100. We take a global estimate for each group and average.



Thanks

Thanks are due to L. Sacerdote and M. Jacobsen for some enlightening discussions on this topic.