Monte Carlo transition density and likelihood approximation

Umberto Picchini
www.math.ku.dk/~umberto

Department of Mathematical Sciences
University of Copenhagen, Denmark

Middelfart 3-16 August 2008

Numerical approximation of SDEs

- Researchers are interested in using differential equations to modelize the dynamics of physical phenomena.
- However analytic solutions to general systems of differental equations are often unavailable
- Computer based numerical strategies are necessary to overcome such difficulty

- SDEs solutions can be obtained numerically using approximations based on stochastic Itô-Taylor expansion, obtained through iterated use of Itô's formula, see Kloeden and Platen (1992).
- The easiest approximation is the stochastic version of the Euler method for ODEs, namely the Euler-Maruyama approximation.

Consider the one-dimensional Itô SDE

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \qquad X_0 = x_0$$

where

- W is an m-dimensional standard Wiener process
- $f : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^{1\times m}$ are known functions.

Consider the Itô SDE above on $[t_0, T]$: for a given discretization $t_0 < t_1 < \cdots < t_n < \cdots < t_N = T$ of $[t_0, T]$, the *Euler–Maruyama approximation* is a continuous time stochastic process satisfying the iterative scheme

$$y_{n+1} = y_n + hf(y_n, t_n) + g(y_n, t_n)\Delta W_n \qquad y_0 = x_0, \quad n = 0, 1, \ldots, N-1$$

where $y_n = y(t_n)$, $h = t_{n+1} - t_n$ is the *stepsize*, $\Delta W_n = W(t_{n+1}) - W(t_n) \sim \mathcal{N}(0, h)$ with $W(t_0) = 0$, and $\mathcal{N}$ is the normal distribution.

Another approximation method is the *Milstein scheme*, which is given by

$$y_{n+1} = y_n + hf(y_n) + g(y_n)\Delta W_n + \frac{1}{2}g(y_n)g'(y_n)((\Delta W_n)^2 - h), \qquad y_0 = x_0$$

where the superscript $'$ denotes differentiation with respect to X.
When g is constant the Euler-Maruyama and the Milstein scheme coincide.
Def: an approximation y of X is said to *converge strongly* to X with order p if exists $C > 0$ (independent of the stepsize h) and $\delta > 0$ such that

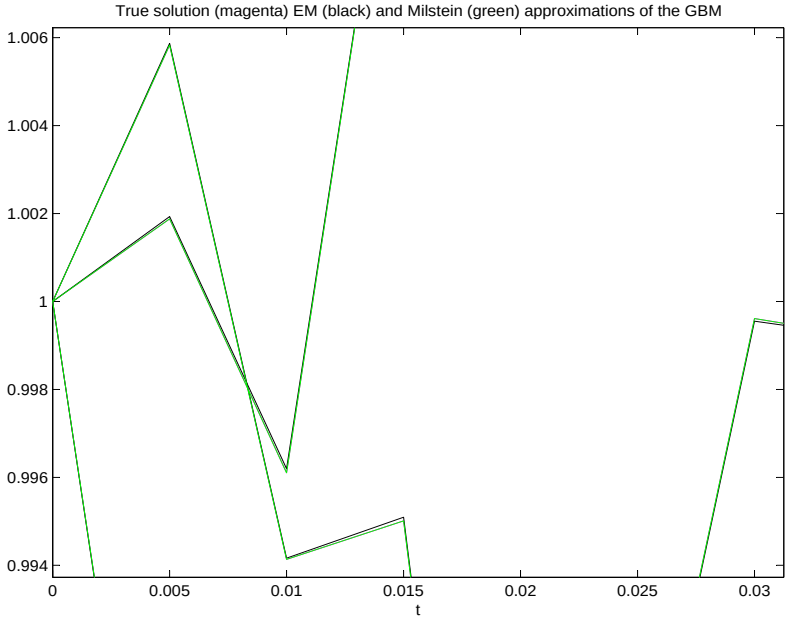
$$\mathbb{E}(|y_N - X_{t_N}|) \leq Ch^p, \qquad h \in (0, \delta).$$

The higher the order the better the approximation.
The Euler-Maruyama scheme has strong order of convergence 0.5 while the Milstein scheme has order 1.

Consider for example, the output you get by running the Matlab program in `demo_GBM.m`, which produces three realizations of the Geometric Brownian Motion using the exact, the Euler-Maruyama and the Milstein solutions.

Notice that the exact and the Milstein solutions cannot be distinguished at this scale.

...and are undistinguishable also at a much smaller scale...



Monte Carlo estimation of SDEs: basic setup

Consider the following d-dimensional Itô SDE:

$$dX_t = b(X_t, t; \theta)dt + \sigma(X_t, t; \theta)dW_t$$

$$X_t = (X_t^{(1)}, ..., X_t^{(d)})^T$$

$$W_t = (W_t^{(1)}, ..., W_t^{(m)})^T$$

$$b : \mathbb{R}^d \times \mathbb{R}^+ \times \Theta \to \mathbb{R}^d$$

$$\sigma : \mathbb{R}^d \times \mathbb{R}^+ \times \Theta \to \mathbb{R}^{d \times m}$$

where $b(\cdot)$ and $\sigma(\cdot)$ are known functions depending on an unknown finite-dimensional parameter vector $\theta \in \Theta$.
We assume that $b(\cdot)$ and $\sigma(\cdot)$ satisfy the usual conditions (Lipschitz, linear growth) for the existence of a unique t-solution of the SDE.
Here T denotes transposition.

Suppose that the following observations are available from process X at discrete time points $\{t_0, t_1, ..., t_n\}$:

$$X_0, X_1, ..., X_n.$$

The central inferential problem for diffusion processes is to estimate θ from the available observations.

As for any parametric problem, maximum likelihood is the preferred method for estimating parameters given a model and a set of data.

The maximum likelihood estimator (MLE) of θ can be calculated if the transition densities $p(X_{i+1}; X_i, \theta)$ are known ($i = 0, 1, ..., n-1$).

Since X is Markovian then the loglikelihood of θ is given by

$$l_n(\theta) = \sum_{i=0}^{n-1} \log p(X_{i+1}; X_i, \theta)$$

and

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} l_n(\theta)$$

Under mild regularity conditions $\hat{\theta}_{MLE}$ is consistent, asymptotically normally distributed and asymptotically efficient as $n \to +\infty$ (Dacunha Castelle-Florens Zmirnou (1986)).

The problem is that the transition density $p(X_{i+1}; X_i, \theta)$ is often unknown!

Thus approximations are needed, e.g.:

1. simulate many times the numerical solution of the process to approximate the transition density via Monte Carlo techniques *(numerically intensive!)*
2. derive a closed-form Hermite expansion to the transition density (Aït-Sahalia (2002,2008))*(very fast but not always applicable)*;
3. solve numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988))*(numerically intensive!)*;

Today we consider only (1).

Monte Carlo based approximation

A Monte Carlo approximation to p consists in:

1. approximating the SDE solution by numerical techniques (Euler-Maruyama, Milstein, stochastic Runge-Kutta, etc.), and
2. averaging Monte-Carlo replicas of functionals of such approximations.

Merits: often applicable over very general, nonlinear, multidimensional SDEs;
Drawbacks: numerically intensive; not always efficient (from a statistical point of view).

- One of the most important MC methods for the estimation of transition densities has been suggested by Pedersen (1995) and, indipendently, by Santa-Clara (1995) (published as article in Brandt and Santa-Clara (2002)).
- Other important contributions are e.g. Durham-Gallant (2002) and Nicolau (2002).
- In the following we consider the contribution by Pedersen and Brandt & Santa-Clara.

The Pedersen/Brandt/Santa-Clara Monte Carlo method

Consider two consecutive observations (X_i, X_{i+1}) from the d-dimensional process X recorded at times (t_i, t_{i+1}) respectively. Without loss of generality, assume $t_{i+1} - t_i = 1$ and divide this interval into M subintervals of length $h = 1/M$.
Denote with $\hat{X}_{t_i+(m+1)h}$ the Euler-Maruyama (EM, shortly) approximation of X_t at time $t_i + (m+1)h$, for $m = 0, 1, ..., M-1$:

$$\hat{X}_{t_i+(m+1)h} = \hat{X}_{t_i+mh} + b(\hat{X}_{t_i+mh}, t_i+mh; \theta)h + \sigma(\hat{X}_{t_i+mh}, t_i+mh; \theta)\sqrt{h}\varepsilon_{t_i+(m+1)}$$

with $\hat{X}_{t_i} = X_{t_i}$.
By definition, the ε's are (independent) $\sim \mathcal{N}(0,1)$ and thus the EM scheme produce a Gaussian process which is such that $\hat{X}_t \to X_t$ weakly as $h \to 0$ under the following assumptions.

Def: a numerical approximation $\hat{X}(h)$ of X, corresponding to a given time-discretization with stepsize h, converges weakly to X at time T if

$$\lim_{h \downarrow 0} |\mathbb{E}X_T - \mathbb{E}\hat{X}_T(h)| = 0.$$

The following are necessary to assure both the weak convergence of the EM scheme to the true solution **and** the applicability of the MC transition density estimation method.
A1: $b(\cdot)$ (drift) and $\sigma(\cdot)$ (diffusion) infinitely differentiable with continuous and bounded derivatives of all orders.
A2: (covariance) matrix $\sigma\sigma^T$ positive defined.
A3: the parameter space Θ is compact and contains the true θ_0;
A4: the likelihood function $L_n(\theta)$ is twice continously differ. in θ in a neighbor. of θ_0; $\mathbb{E}([\partial L_N(\theta)/\partial\theta][\partial L_N(\theta)/\partial\theta^T])$ has full rank and is bounded for all $\theta \in \Theta$
There are a couple of more assumptions which, however, in this context it is not strictly necessary to mention.

Since the EM scheme between two adjacent instants $[t_i + mh, t_i + (m+1)h]$ defines a Gaussian process, we have that the transition density at $\hat{X}_{t_i+(m+1)h} = y$ given $\hat{X}_{t_i+mh} = x$ is a multivariate gaussian pdf:

$$q_M(y, t_i + (m+1)h | x, t_i + mh; \theta) = \phi(y; x + b(x, t_i + mh; \theta) \cdot h, h \cdot V(x, t_i + mh; \theta))$$

where $V = \sigma(\cdot)\sigma(\cdot)^T$.

- However q_M is just an approximation of $p(y, t_i + (m+1)h | x, t_i + mh; \theta)$ **but** $q_M \to p$ as $h \to 0$.
- Unfortunately the relation above can only give the transition density between two adjacent points: instead, we are interested in the density between two non-adjacent points, say between X_{t_i+mh} and $X_{t_i+(m+j)h}$ (with $j = 2, 3, ..., M-m$).
- **remember:** our goal is to compute the density between X_{t_i} and $X_{t_{i+1}}$, which correspond to the case $m = 0$ and $j = M$.

To reach our goal we need to compute the *multi-step-ahead* transition density, which can be defined recursively: that is

$$
\begin{aligned}
&q_M(y, t_i + (m+j)h | x, t_i + mh; \theta) = \\
&\quad \int_{\mathbb{R}^d} \Big\{ q_M(y, t_i + (m+j)h | z, t_i + (m+j-1)h; \theta) \\
&\times \quad q_M(z, t_i + (m+j-1)h | x, t_i + mh; \theta) \Big\} dz \\
&= \quad \int_{\mathbb{R}^d} \Big\{ \phi(y; z + b(z, t_i + (m+j-1)h; \theta) \cdot h, h \cdot V(z, t_i + (m+j-1)h; \\
&\times \quad q_M(z, t_i + (m+j-1)h | x, t_i + mh; \theta) \Big\} dz
\end{aligned}
$$

and the last term in the integrand can be computed in the same way...in the end we get a convolution of $M-1$ integrals which in general is not solvable neither analytically nor numerically (via quadrature techniques) when M increases.

The idea is to treat the integral as an expectation of the function ϕ w.r.t. the random variable z:

$$
\begin{aligned}
&q_M(y, t_i + (m+j)h | x, t_i + mh; \theta) = \\
&= \quad \int_{\mathbb{R}^d} \phi(y; z + b(z, t_i + (m+j-1)h; \theta) \cdot h, h \cdot V(z, t_i + (m+j-1)h; \theta \\
&\times \quad f(z) dz = \mathbb{E}_z(\phi(\cdot))
\end{aligned}
$$

where $f(z) = q_M(z, t_i + (m+j-1)h | x, t_i + mh; \theta)$.

The expectation can be approximated via Monte Carlo simulations from the density f.

Very interesting...but in practice?!

- This means that we can generate a large number R of random draws z_r $(r = 1, ..., R)$ using the EM scheme
- using the $z_1, ..., z_R$ we approximate the density q_M by averaging all the ϕ's evaluated at the different z_r's, i.e.

$$
\begin{aligned}
&q_M(X_{i+1}, t_{i+1} | X_i, t_i; \theta) \approx \\
&\quad \hat{q}_{M,R}(X_{i+1}, t_{i+1} | X_i, t_i; \theta) = \\
&\quad \frac{\sum_{r=1}^{R} \phi(X_{i+1}; z_r + b(z_r, t_i + (M-1)h; \theta) \cdot h, h \cdot V(z_r, t_i + (M-1)h; \theta)}{R}
\end{aligned}
$$

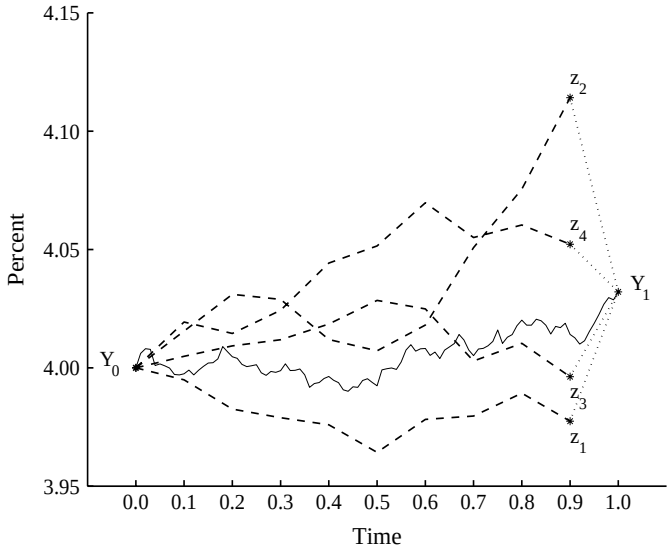


Fig. 1. Approximating the transition densities. This figure illustrates the approximation of the transition densities of a diffusion. The solid line represents the unobserved continuous-time sample path of a univariate diffusion. The four dashed lines represent incomplete ten-step Euler discretizations.

In the end we get:

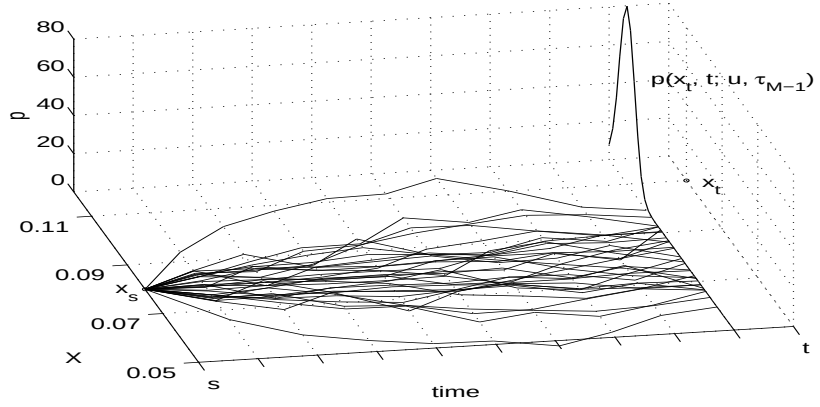
$$l_n(\theta) \approx l_{n,M,R}(\theta) = \sum_{i=0}^{n-1} ln(\hat{q}_{M,R}(X_{i+1}, t_{i+1}|X_i, t_i; \theta))$$

and by denoting with $\hat{\theta}_{n,M,R} = \arg\max_{\theta\in\Theta} l_{n,M,R}(\theta)$ we have

$$\hat{\theta}_{n,M,R} \to \hat{\theta}_n, \qquad M, R \to \infty, \quad R^{1/2}/M \to 0$$
$$\hat{\theta}_n \to \theta_0, \qquad n \to \infty$$

under the stated assumptions.

- **Merits**: "easy" to implement; applicable with multidimensional SDEs;
- **Drawbacks**: numerically intensive (requires large M and R values); there is no specified criteria to sample from the $f(z)$ distribution efficiently, i.e. importance sampling techniques are needed (see the figure below and the improvements by Durham and Gallant (2002)).



Let's consider some computer simulations...

Software for SML

1. R package: S.M. Iacus's **sde**, `http://cran.r-project.org/web/packages/sde/` with companion monography available from Springer;
2. MATLAB package: U. Picchini's **SDE Toolbox**, `http://sdetoolbox.sourceforge.net/`
3. ???

References

- Y. Aït-Sahalia (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1), 223-262.
- Y. Aït-Sahalia (2008). Closed-form likelihood expansion for multivariate diffusions. *Ann. Stat*, 36(2), 906-937.
- M.W. Brandt and P. Santa-Clara (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 63, 161-210.
- P.E. Kloeden and E. Platen (1992). *Numerical solution of stochastic differential equations*. Springer.
- A. Lo (1988). Maximum likelihood estimation of generalized Ito processes with discretely-sample data. *Econometric Theory*, 4, 231-247.

References

- A.R. Pedersen (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Stat.*, 22, 55-71.
- P. Santa-Clara (1995). Simulated likelihood estimation of diffusions with an application to the short term interest rate. Ph.D. dissertation, INSEAD.

Modelization of the euglycemic hyperinsulinemic clamp by SDEs

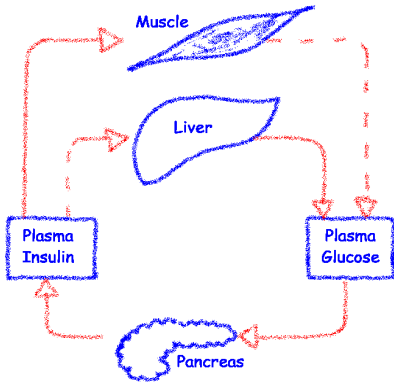
Umberto Picchini
www.math.ku.dk/~umberto

Department of Mathematical Sciences
University of Copenhagen, Denmark

Middelfart 3-16 August 2008

The Physiological Problem: Feedback Loop

Glucose and Insulin

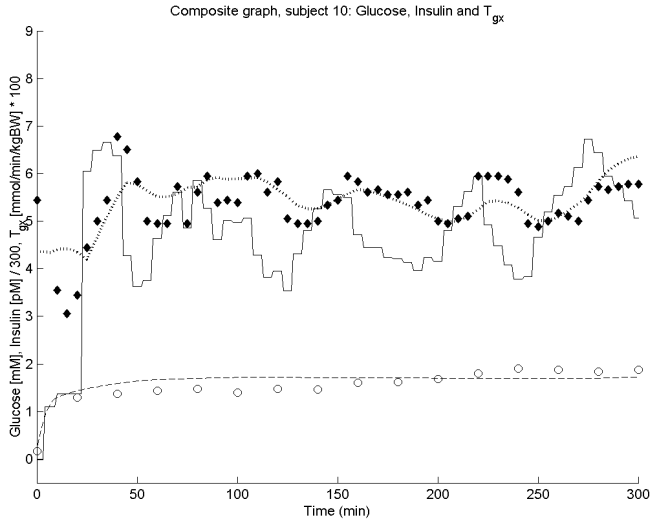


The Physiological Problem: EHC procedure

The Euglycemic Hyperinsulinemic Clamp (EHC, DeFronzo et.al. (1979)) is the diabetological gold-standard for the assessment of *insulin sensitivity*.
In this procedure

1. insulin concentrations are rapidly raised to a high value by means of an insulin bolus injection and maintained at this level during the experiment (2-5 hrs) by means of a constant insulin infusion
2. glucose concentrations are maintained close to a “target” (basal) level by means of variable rate glucose infusions
3. both glycemia and insulinemia are sampled during the experiment

Observed glycemias (♦) and external glucose infusion rate (solid line).



Notice how the external infusion try to keep the glycemia around the "normal" (basal) value, i.e. around the first observation.

A ODE model: Picchini et al. (2006) *J. Math. Biol.*

We firstly hypothesize a system of ODEs explaining glycemia and insulinemia dynamics:

$$\begin{aligned}
\frac{dG(t)}{dt} &= \frac{(T_{gx}(t-\tau_g)+T_{gh}(t))}{V_g} - T_{xg}\frac{G(t)}{0.1+G(t)} - K_{xgl}G(t)I(t)\\
\frac{dI(t)}{dt} &= \frac{(T_{iG}G(t)+T_{ix}(t))}{V_i} - K_{xi}I(t)\\
T_{gh}(t) &= T_{ghmax}\exp(-\lambda G(t)I(t))
\end{aligned}$$

where

$$G(0) = G_b, \quad I(0) = I_b,$$

$$T_{gh}(0) = T_{ghb} = T_{ghmax}\exp(-\lambda G_b I_b),$$

$$T_{gx}(s) = 0 \quad \forall s \in [-\tau_g, 0] \qquad \textit{and} \qquad T_{ix}(0) = T_{ixb}$$

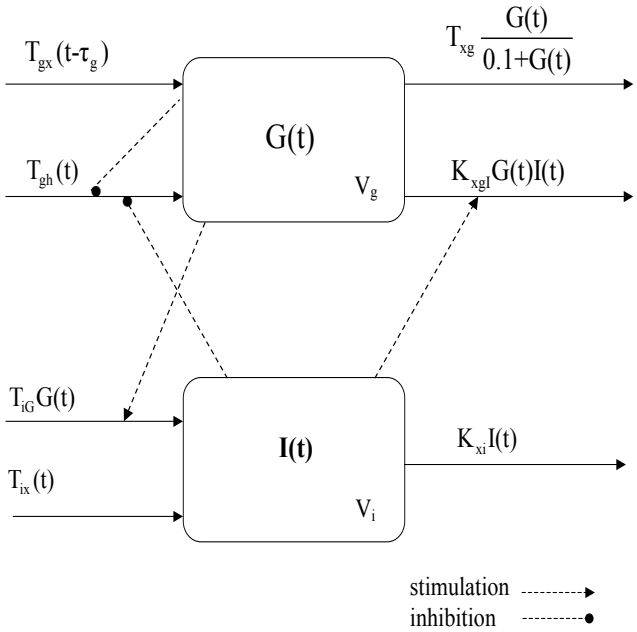


Figure: Schematic representation of the model.

We considered two different procedure to estimate the model parameters $\theta = (G_b, I_b, K_{xgl}, K_{xi}, T_{ghmax}, V_g, V_i, \tau_g, \lambda)$:

1. estimate the parameters separately for each subjects;
2. perform a *population* estimation approach.

For the sake of this lecture we will consider the second approach only, which is more interesting to motivate the use of SDEs in this experiment.

Population estimation

In this analysis we want to estimate from the data both:

- the structural model parameters considered in $\theta = (G_b, I_b, K_{xgl}, K_{xi}, T_{ghmax}, V_g, V_i, \tau_g, \lambda)$ and
- the coefficient of variations for the glycemia/insulinemia measurements $\xi = (CV_G, CV_I)$.

The best way to estimate the CV's is by the use of repeated measurements over the subjects, when available.
In this case repeated measurements are not available, and thus we pool together the data obtained on all the subjects in order to estimate ξ.

Population estimation: notation

We have glycemia and insulinemia data from 15 subjects. Let θ^i be the individual model parameters for subject i and consider the complete measurement vector for subject i: $y^i = (y^i_{G,1}, ..., y^i_{G,n^i_G}, y^i_{I,1}, ..., y^i_{I,n^i_I})$ for the sequence of data for subject i; n^i_G and n^i_I are the numbers of glycemia and insulinemia measurements for subject i respectively. Then we consider the following error-model:

$$y^i = f^i(\theta^i) + \varepsilon^i, \qquad i = 1, ..., 15$$

such that

$$\mathbb{E}(\varepsilon^i|\theta^i) = \mathbf{0}, \quad Cov(\varepsilon^i|\theta^i) = \Omega^i(\theta^i, \xi)$$

with $f^i(\cdot) = (f^i_{G,1}, ..., f^i_{G,n^i_G}, f^i_{I,1}, ..., f^i_{I,n^i_I})$ representing the numerical solution of the ODE system for subject i; ε^i has the same dimensions of y^i and f^i.

We assume that the functional form of $\Omega^i(\cdot,\cdot)$ and the intra-individual covariance parameter $\xi = (CV_G, CV_I)$ are the same across individuals.

$$\Omega^i(\theta^i, \xi) = \begin{pmatrix} \Omega^i_G & 0 \\ 0 & \Omega^i_I \end{pmatrix} \qquad i = 1, ..., 15$$

$$\Omega^i_G = \begin{pmatrix} CV_G^2(f^i_{G,1})^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & CV_G^2(f^i_{G,n_{iG}})^2 \end{pmatrix},$$

and similarly

$$\Omega^i_I = \begin{pmatrix} CV_I^2(f^i_{I,1})^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & CV_I^2(f^i_{I,n_{il}})^2 \end{pmatrix}$$

We used these structures to perform the following GLS (General Least Squares) algorithm:

1. in $m = 15$ separate regressions, obtain preliminary estimates $\hat{\theta}^{(p),i}$ for the parameters of each subject, $i = 1, ..., m$;
2. use residuals from these preliminary fits to estimate ξ by minimizing the pseudolikelihood of ξ for the ith individual

$$\begin{aligned} \sum_{i=1}^m PL^i(\hat{\theta}^{(p),i}, \xi) &= \sum_{i=1}^m \log|\Omega^i(\hat{\theta}^{(p),i}, \xi)| \\ &+ (y^i - f^i(\hat{\theta}^{(p),i}))'(\Omega^i)^{-1}(\hat{\theta}^{(p),i}, \xi)(y^i - f^i(\hat{\theta}^{(p),i})) \end{aligned}$$

   Form estimated weight matrices

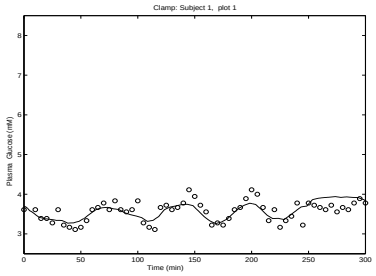
$$\hat{\Omega}^i(\hat{\theta}^{(p),i}, \hat{\xi})$$

3. using the estimated weight matrices from step 2, re-estimate the θ^i's by m separate minimizations: for individual i, minimize in θ^i

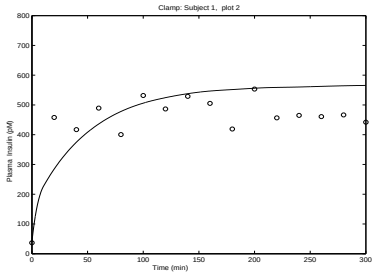
$$(y^i - f^i(\theta^i))'(\hat{\Omega}^i)^{-1}(y^i - f^i(\theta^i))$$

   Treating the resulting estimators as new preliminary estimators, return to step 2.

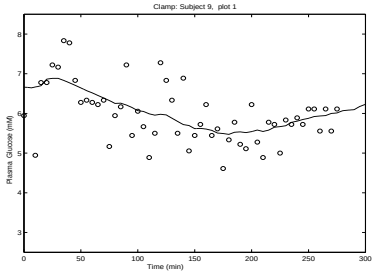
In this way we obtained reasonable fits:



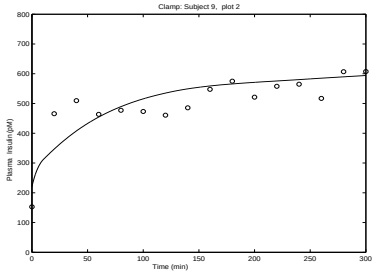
(a) Subject 1, fitted glycemias



(b) Subject 1, fitted insulinemias



(c) Subject 9, fitted glycemias



(d) Subject 9, fitted insulinemias

... but

we found that $\widehat{CV}_G = 0.071$ and $\widehat{CV}_I = 0.170$, which are too high to be compatible with measurement error, especially if compared with commonly accepted values, e.g. $(CV_G, CV_I) = (0.015, 0.07)$ in Bergman et al. (1979).

This finding has prompt us to consider an additional source of noise, besides measurement error, explaining the variation of the observations around the predicted curve.

Modeling the EHC by SDEs

Why do we model glycemia and insulinemia values by means of SDEs?

- We hypothesize that some physiological process (e.g. glucose tissue absorption) is perturbed by random *noise*.
- This noise represents the overall action of many factors (e.g. sudden changes in physical activity or emotional stresses), each with a small individual effect, affecting glucose tissue absorption
- This noise is superimposed to a non-random (drift) term, representing the most relevant and well-recognized factors affecting glycemia
- We will show how to modelize and identify both system noise and measurement error, where the latter influences the measurement values but not the course of the underlying process.

A SDE model of the EHC

- We let the insulin dependent glucose tissue uptake rate K_{xgI} vary randomly as $(K_{xgI} - \xi(t))$, where $\xi(\cdot)$ is a gaussian white noise process;
- the system noise $\xi(t)dt$ can be written as $\sigma dW(t)$ and σ is an intensity factor.

We obtain the following SDE model of glycemia $G(t)$ and insulinemia $I(t)$ dynamics

$$\begin{aligned} dG(t) &= \left[\frac{(T_{gx}(t-\tau_g) + T_{gh}(t))}{V_g} - T_{xg}\frac{G(t)}{0.1+G(t)} - K_{xgI}G(t)I(t)\right]dt \\ &\quad +\sigma G(t)I(t)dW(t) \\ dI(t) &= \left[\frac{(T_{iG}G(t) + T_{ix}(t))}{V_i} - K_{xi}I(t)\right]dt \\ T_{gh}(t) &= T_{ghmax}\exp(-\lambda G(t)I(t)) \end{aligned}$$

Some Complications...

...but

Partially Observed System

$G(t)$ is measured every 5 min, whereas $I(t)$ every 20 min; this means that we deal with a *partially observed system* $\Rightarrow$ the transition densities cannot be computed at each observation time-point

Measurement Error

We allow the measurements to be generated with non-negligible measurement error, e.g. for a generic subject his/her jth observation (glycemia or insulinemia) is modelized as: $y_j = X_j(\theta) + \varepsilon_j$ with $\varepsilon_j \sim \mathcal{N}(0, \Sigma_j(\xi))$, with the ε's independent of $W(\cdot)$, thus...

Non-Markovianity

... the observations y_j are Non-Markovian

Estimation strategy

1. Consider **all the past history** of the observed process
2. build the likelihood function of (θ, ξ) separately for each subject

$$L(\theta, \xi) = \prod_{j=1}^{n} p_{j|j-1}(y_j | y_0, y_1, \ldots, y_{j-1}, \theta, \xi)$$

where we use the notation $p_{j|j-1}$ to denote the conditional density function of y_j given $y_0, y_1, ..., y_{j-1}$ and $n+1$ is the total number of measurements (glycemia and insulenimia) available for the considered subject

3. for each subject, maximize the corresponding likelihood to obtain the individual MLE estimates of (θ, ξ)

How To? (A.R. Pedersen (2001))

Since measurement errors are stochastically independent, and independent of the diffusion process, we have

$$\begin{aligned} L(\theta, \xi) &= \int \left[\prod_{j=0}^{n} \phi_j(y_j; x_j, \xi) \right] \lambda(x_1, \ldots, x_n; \theta) dx_1 \cdots dx_n \\ &= \mathbb{E}_\theta \prod_{j=0}^{n} \phi_j(y_j; X_{t_j}; \xi) \end{aligned}$$

where λ denotes the joint pdf of $X_{t_1}, \ldots, X_{t_n}$, $\mathbb{E}_\theta$ denotes expectation with respect to the distribution of $X_{t_1}, \ldots, X_{t_n}$ for the indicated parameter values, and

$$\phi_j(y_j; x_j, \xi) = |2\pi\Sigma_j(\xi)|^{-1/2} \exp\left(-\frac{1}{2} \left[y_j - x_j(\theta) \right]' \Sigma_j^{-1}(\xi) \left[y_j - x_j(\theta) \right] \right)$$

Error model

We used the following error-model $y_j = X_j(\theta) + \varepsilon_j$ with $\varepsilon_j \sim \mathcal{N}(0, \Sigma_j(\xi))$ where

- $X_j(\theta) = X(t_j) = (G_j, I_j)$ is the (numerical) solution of the SDEs at sampling time t_j;
- $$\Sigma_j(\xi) = \begin{pmatrix} CV_G^2 G_j^2 & 0 \\ 0 & CV_I^2 I_j^2 \end{pmatrix}$$

and $\xi = (CV_G, CV_I)$ contains the glucose and insulin coefficient of variations.

So, by simulating *many* (R) trajectories X^r of X,

$$L(\theta, \xi) \simeq \frac{1}{R} \sum_{r=1}^{R} \prod_{j=0}^{n} \phi_j(y_j | X_{t_j}^r(\theta), \xi), \qquad (R \to +\infty)$$

- The $X_{t_j}^r$ can be simulated using a standard algorithm (e.g. Euler-Maruyama, Milstein)
- the independent Wiener increments should be simulated, initially, and kept fixed in all subsequent calculations of the Monte Carlo approximation of the likelihood function.

Difficulties

This simulation-based approach is HIGHLY time consuming since:

1. the number of simulations R should be at least of the order of *thousands* (we choose $R = 2000$)
2. the integration stepsize should be "small" enough (we choose $h = 0.1$)
3. the larger the number of parameters to be optimized the slower the convergence for the optimization procedure
4. the likelihood approximation procedure must be performed for every infinitesimal variation of (θ, ξ) during the optimization procedure

Our Strategy

We optimize only the σ parameter in the diffusion part.

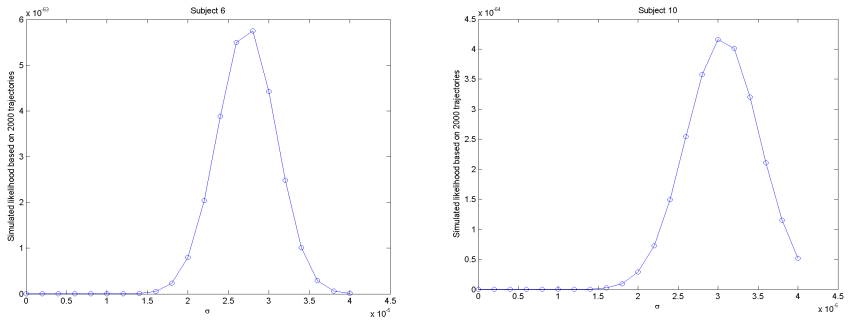
We plugged into the SDE the parameter estimates obtained for the ODE model (corresponding to the SDE's drift) for different levels of measurement error;

$$(CV_G, CV_I) \in \{(0.015, 0.07), (0.02, 0.10), (0.03, 0.10), (0.03, 0.15), (0.04, 0.15), (0.05, 0.15), (0.15, 0.30)\}.$$

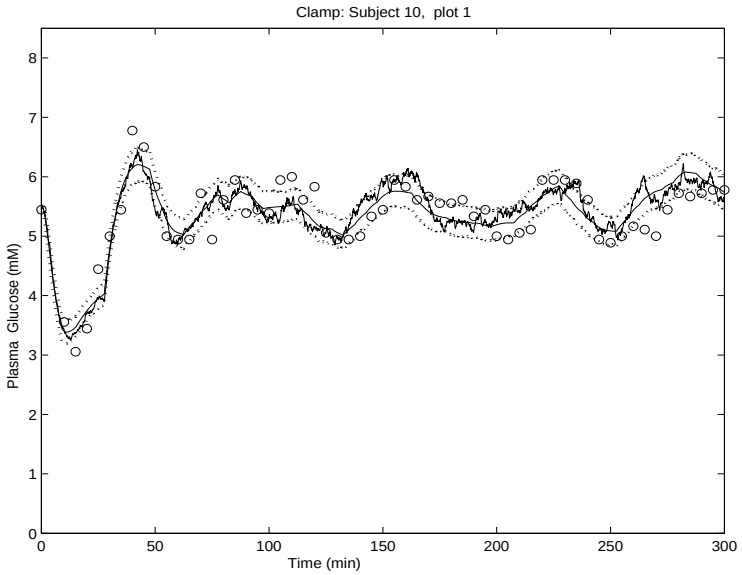
We also used the **GLS** population estimates of the measurement error variance;

Then, we **fixed** both the parameters entering the drift and the measurement error variances and, conditionally to these, only σ was optimized.

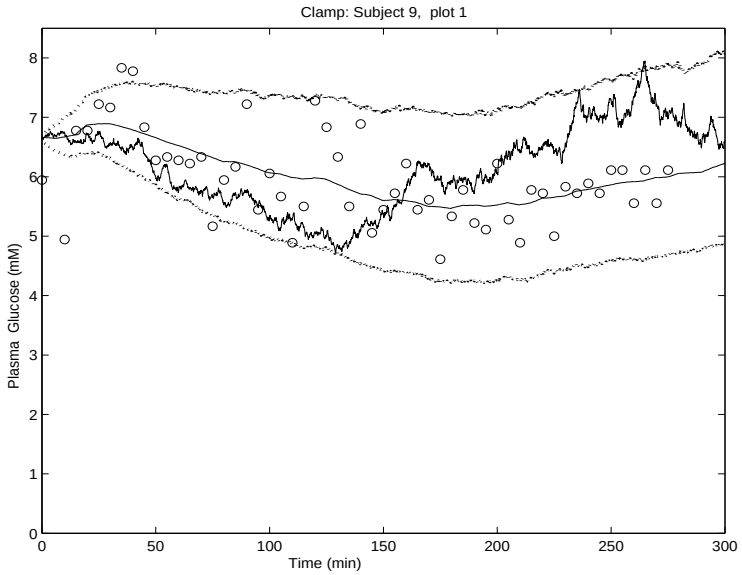
Two typical likelihoods of σ



Some Results: plasma glucose (subject 10)



Some Results: plasma glucose (subject 9)



Estimates of the diffusion coefficient

Here are the 7 estimates of σ corresponding to the 7 hypothesized increasing levels of measurement error (proportional to (CV_G, CV_I))

Subjects	$\hat{\sigma}^{(1)}$	$\hat{\sigma}^{(2)}$	$\hat{\sigma}^{(3)}$	$\hat{\sigma}^{(4)}$	$\hat{\sigma}^{(5)}$	$\hat{\sigma}^{(6)}$	$\hat{\sigma}^{(7)}$
1	1.60E-5	1.78E-5	2.25E-5	1.59E-5	2.10E-5	2.25E-5	0
2	NA	1.38E-5	1.47E-5	1.38E-5	1.38E-5	1.15E-5	2.88E-7
3	2.39E-5	4.55E-5	5.71E-5	2.54E-5	3.95E-5	2.58E-5	0
4	NA	1.00E-5	1.00E-5	1.00E-5	1.00E-5	0.95E-5	3.68E-8
5	1.83E-5	1.97E-5	2.00E-5	1.93E-5	1.93E-5	1.93E-5	0.91E-5
6	2.72E-5	2.65E-5	2.68E-5	2.71E-5	2.72E-5	2.73E-5	5.29E-8
7	0.80E-5	0.80E-5	0.80E-5	0.80E-5	2.35E-8	2.35E-8	1.47E-7
8	0.72E-5	0.76E-5	0.73E-5	0.72E-5	0.60E-5	0.42E-5	2.12E-7
9	NA	2.42E-5	2.65E-5	2.50E-5	2.60E-5	2.69E-5	4.77E-7
10	3.08E-5	3.04E-5	3.04E-5	3.04E-5	3.00E-5	2.88E-5	4.25E-7
11	0.62E-5	0.59E-5	3.68E-8	3.68E-8	2.35E-8	3.68E-8	4.77E-7
12	1.44E-5	0.98E-5	1.36E-5	1.41E-5	1.47E-5	1.53E-5	8.47E-7
13	1.23E-5	0.82E-5	0.86E-5	0.74E-5	0.84E-5	0.87E-5	2.12E-7
14	1.73E-5	1.65E-5	1.64E-5	1.62E-5	1.56E-5	0.12E-5	7.21E-8
15	1.87E-5	1.23E-5	1.44E-5	1.86E-5	1.47E-5	1.50E-5	0.84E-5

In this table we notice that the σ estimates are stable when considered in a reasonable region of the coefficient of variations values, that is when considered in $(CV_G, CV_I) \in [0.02, 0.05] \times [0.10, 0.15]$ (i.e. from $\hat{\sigma}^{(2)}$ to $\hat{\sigma}^{(6)}$).

Monte Carlo Confidence Intervals

Subjects		$\hat{\sigma}^{(1)} \cdot 10^{-5}$	$\hat{\sigma}^{(4)} \cdot 10^{-5}$
1	estimates (mean [95% CI])	1.60 (1.27 [0.61, 1.85])	1.59 (1.39 [0.45, 2.63])
	Skewness	-0.2684	0.6751
	Kurtosis	2.216	4.552
6	estimates (mean [95% CI])	2.72 (2.26 [0.89, 3.01])	2.71 (2.25 [1.15, 3.27])
	Skewness	-0.676	-0.273
	Kurtosis	3.142	2.670
10	estimates (mean [95% CI])	3.08 (1.92 [1.10, 3.17])	3.04 (2.06 [0, 4.08])
	Skewness	0.787	0.286
	Kurtosis	3.084	3.961

Conclusions

1. (from the ODE model) the level of error around the predicted curve is very large, in particular it is much larger than the (0.015,0.07) commonly accepted levels of measurement error in *in vitro* repeated testing of the same laboratory preparation.
2. (SDE model) for any reasonable level of observation error, the estimated diffusion has more or less the same value. For "reasonable" it is here meant larger than pure measurement error and smaller than the total error around the expected trajectory as estimated by GLS. Having excluded these extreme cases, it can be seen that, in the present situation, the estimation of the diffusion is very robust to changes in the likely value of the observation error.
3. the diffusion coefficient estimates are generally strictly positive: this means that the dynamical process which most likely represents the glycemia time-course (given the estimated ODE) is a stochastic process with a non-negligible system noise.

Reference

U. Picchini, A. De Gaetano and S. Ditlevsen (2006). Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *Journal of Mathematical Biology*, 53(5), 771-796.

Efficient estimation of a new SDE model of the EHC

Umberto Picchini
www.math.ku.dk/~umberto

Department of Mathematical Sciences
University of Copenhagen, Denmark

Middelfart, 3-16 August 2008

In the first lecture we considered the problem of approximating the transition density $p_X(X_{i+1}; X_i, \theta)$ of a Markovian process X. At least three categories of possible strategies can be considered:

1. simulating many times the numerical solution of the process to approximate the transition density via Monte Carlo techniques *(numerically intensive!)*
2. deriving a closed-form Hermite expansion to the transition density (Aït-Sahalia (2002,2008), Egorov(2003))*(very fast but not always applicable)*;
3. solving numerically the Kolmogorov partial differential equations satisfied by the transition density (Lo (1988))*(numerically intensive!)*;

Today we consider (2).

The idea (Ait-Sahalia (2002))

With *closed-form approximation* we mean that a mathematical expression for an approximation of p_X is *explicitly given*. Thus:

- there is no need to perform simulations of process trajectories;
- the computations needed to obtain an estimator for θ are extremely fast (just like when the exact p_X is available);
- the approximation of the likelihood obtained with this method is extremely accurate, especially if compared to Monte Carlo based methods (see e.g. Durham-Gallant (2002) and Jensen-Poulsen (2002))

- consider the one-dimensional time-homogeneous Ito SDE:

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

- transform the original SDE into an equivalent SDE with unit diffusion

$$dY_t = \mu_Y(Y_t, \theta)dt + dW_t$$

- "standardize" the new process Y_t to obtain another process Z_t having a transition density p_Z close to a Gaussian one;
- After obtaining an approximation for p_Z, we can get the corresponding approximation for p_Y and finally p_X using the Jacobian formula.

Let's start with the transformation $X_t \to Y_t$.

- consider the function

$$Y_t = \gamma(X_t, \theta) = \int^{X_t} \frac{1}{\sigma(u, \theta)} du \quad \Rightarrow \quad X_t = \gamma^{-1}(Y_t, \theta)$$

- using the Ito's formula on our SDE (where $\gamma(\cdot)$ plays the role of the function g) we get

$$dY_t = \mu_Y(Y_t, \theta)dt + dW_t$$

as desired, where

$$\mu_Y(Y_t, \theta) = \frac{\mu(\gamma^{-1}(Y_t, \theta))}{\sigma(\gamma^{-1}(Y_t, \theta))} - \frac{1}{2}\frac{\partial \sigma}{\partial X_t}(\gamma^{-1}(Y_t, \theta));$$

notice that the vector θ is the same vector of the original SDE;

- the transition density of Y is much closer to a Gaussian distribution than that of X (thanks to the unit diffusion), i.e. when $\Delta = t_{i+1} - t_i \to 0$ the magnitude of the tails of p_Y is similar to the magnitude of the tails of a Gaussian distribution (Prop. 2 in Ait-Sahalia (2002)).

Notation: we denote with $p_W(w|w_0)$ the transition density of a generic process W at w conditionally on w_0.

With a further transformation $Z = \Delta^{-1/2}(Y - y_0)$ we get a process Z which is close enough to a $\mathcal{N}(0,1)$ variable to make it possible to create a convergent series of expansions for its density p_Z around a $\mathcal{N}(0,1)$, see the next slides for details.

By denoting with $p_Y(y|y_0;\theta)$ the conditional density of $Y_{t+\Delta}|Y_t$, we have the following relation

$$p_Z(z|y_0;\theta) = \Delta^{1/2} p_Y(\Delta^{1/2}z + y_0|y_0;\theta)$$

and now we are going to show how to approximate p_Z.

For a Hilbert space $L^2(P)$ with measure P, density ϕ and inner product $< \cdot, \cdot >$ a density $g(w)$ on the real line can be represented by an expansion with respect to an orthogonal base $\{H_1, H_2, ...\}$, i.e.:

$$g = \sum_{j=0}^{\infty} \frac{< g, H_j >}{< H_j, H_j >} H_j, \qquad < g, H_j >= \int_{-\infty}^{+\infty} g(w) H_j(w) \phi(w) dw.$$

For any density $p(w)$ on the real line if we expand the ratio $p(w)/\phi(w)$ we get

$$\frac{p(w)}{\phi(w)} = \sum_{j=0}^{+\infty} \frac{< p/\phi, H_j >}{< H_j, H_j >} H_j \Rightarrow p(w) = \phi(w) \sum_{j=0}^{\infty} \eta^{(j)} H_j$$

where

$$\eta^{(j)} = \frac{< p/\phi, H_j >}{< H_j, H_j >} = \frac{\int_{-\infty}^{+\infty} p(w) H_j(w) dw}{\int_{-\infty}^{+\infty} H_j(w) H_j(w) \phi(w) dw}$$

Now if we choose P to be the standard normal distribution and $\{H_1, H_2, ...\}$ as the Hermite base, we have

$$H_j(w) = \phi(w)^{-1} \frac{d^j}{dw^j} \phi(w) \quad (\textit{Hermite polynomials})$$

where $\phi(w)$ is the standard normal pdf and we have the following properties:

1. $(j+1)H_j(w) = (d/dw)H_{j+1}(w)$
2. $\int_{-\infty}^{+\infty} H_i(w) H_j(w) \phi(w) dw = \begin{cases} i!, & \textit{if } j = i \\ 0, & \textit{otherwise} \end{cases}$

and we can write the transition density for Z as:

$$p_Z(z|y_0;\theta) = \phi(z) \sum_{j=0}^{+\infty} \eta_Z^{(j)}(y_0;\theta) H_j(z)$$

where (see the previous relations)

$$\eta_Z^{(j)}(y_0;\theta) = \frac{1}{j!} \int_{-\infty}^{+\infty} p_Z(z|y_0) H_j(z) dz$$

In practice we trunk the expansion to an integer J

$$p_Z^{(J)}(z|y_0;\theta) = \phi(z)\sum_{j=0}^{J}\eta_Z^{(j)}(y_0;\theta)H_j(z)$$

thus obtaining the corresponding approximations for the transition densities of Y and X, using the Jacobian formula:

$$p_Y^{(J)}(y|y_0;\theta) = \Delta^{-1/2}p_Z^{(J)}(\Delta^{-1/2}(y-y_0)|y_0;\theta)$$

$$p_X^{(J)}(x|x_0;\theta) = \sigma(x;\theta)^{-1}p_Y^{(J)}(y|y_0;\theta)$$

Notice that (Ait-Sahalia (2002), Theorem 1) $p_X^{(J)}(x|x_0;\theta) \to p_X(x|x_0;\theta)$ as $J \to +\infty$ for every $\theta \in \Theta$ and for every (x, x_0) in D_X^2 (where D_X is the state space of X).

Of course the main problem is computing the $\eta_Z^{(j)}$'s: we note that

$$\begin{aligned}
\eta_Z^{(j)}(y_0;\theta) &= 1/j!\int_{-\infty}^{+\infty} H_j(z)p_Z(z|y_0,\theta)dz \\
&= 1/j!\int_{-\infty}^{+\infty} H_j(z)\Delta^{1/2}p_Y(\Delta^{1/2}z+y_0|y_0,\theta)dz \\
&= 1/j!\int_{-\infty}^{+\infty} H_j(\Delta^{-1/2}(y-y_0))p_Y(y|y_0;\theta)dy \\
&= 1/j!\mathbb{E}[H_j(\Delta^{-1/2}(Y_{t+\Delta}-y_0))|\,Y_t = y_0]
\end{aligned}$$

As usual we may calculate the expectation using, e.g. Monte Carlo methods, but in this case we want to find some more elegant and efficient solution.

Ait-Sahalia (2002) expands the expectation using a Taylor series of order K in powers of Δ and thus, in the end, he gets Taylor approximations of $\eta_Z^{(j)}$ which we denote by $\eta_Z^{(j,K)}$ $(j = 1, ..., J)$, e.g. the following are the first 3 coefficients for the case $K = 3$:

$$\begin{aligned}
\eta_Z^{(1,3)} &= -\mu_Y\Delta^{1/2} - (2\mu_Y\mu_Y^{[1]} + \mu_Y^{[2]})\Delta^{3/2}/4 \\
&\quad - (4\mu_Y\mu_Y^{[1]2} + 4\mu_Y^2\mu_Y^{[2]} + 6\mu_Y^{[1]}\mu_Y^{[2]} + 4\mu_Y\mu_Y^{[3]} + \mu_Y^{[4]})\Delta^{5/2}/24, \\
\eta_Z^{(2,3)} &= (\mu_Y^2 + \mu_Y^{[1]})\Delta/2 + (6\mu_Y^2\mu_Y^{[1]} + 4\mu_Y^{[1]2} + 7\mu_Y\mu_Y^{[2]} + 2\mu_Y^{[3]})\Delta^2/12 \\
&\quad + (28\mu_Y^2\mu_Y^{[1]2} + 28\mu_Y^2\mu_Y^{[3]} + 16\mu_Y^{[1]3} + 16\mu_Y^3\mu_Y^{[2]} + 88\mu_Y\mu_Y^{[1]}\mu_Y^{[2]} \\
&\quad + 21\mu_Y^{[2]2} + 32\mu_Y^{[1]}\mu_Y^{[3]} + 16\mu_Y\mu_Y^{[4]} + 3\mu_Y^{[5]})\Delta^3/96, \\
\eta_Z^{(3,3)} &= -(\mu_Y^3 + 3\mu_Y\mu_Y^{[1]} + \mu_Y^{[2]})\Delta^{3/2}/6 - (12\mu_Y^3\mu_Y^{[1]} + 28\mu_Y\mu_Y^{[1]2} \\
&\quad + 22\mu_Y^2\mu_Y^{[2]} + 24\mu_Y^{[1]}\mu_Y^{[2]} + 14\mu_Y\mu_Y^{[3]} + 3\mu_Y^{[4]})\Delta^{5/2}/48,
\end{aligned}$$

where $\mu_Y^{[k],m}$ denotes $(\partial^k\mu_Y(y_0;\theta)/\partial y_0^k)^m$.

Using the coefficients $\eta_Z^{(j,K)}$ we can finally write an *explicit* closed-form approximation for p_Z:

$$p_Z^{(J,K)}(z|y_0;\theta) = \phi(z)\sum_{j=0}^{J}\eta_Z^{(j,K)}(y_0;\theta)H_j(z)$$

and using the Jacobian formula we can obtain the corresponding $p_Y^{(J,K)}$ and $p_X^{(J,K)}$,

$$p_Y^{(J,K)}(y|y_0;\theta) = \Delta^{-1/2}p_Z^{(J,K)}(\Delta^{-1/2}(y-y_0)|y_0;\theta)$$

$$p_X^{(J,K)}(x|x_0;\theta) = \sigma(x;\theta)^{-1}p_Y^{(J,K)}(y|y_0;\theta),$$

thus if we have a sample $X_0, ..., X_n$ of observations from the process X we get

$$l_n^{(J,K)}(\theta) = \sum_{i=0}^{n-1}\log p_X^{(J,K)}(X_{i+1}|X_i;\theta)$$

Merits:

1. Accurate approximations of the true transition density (an order of approximation $K = 1$ or 2 is often sufficient);
2. Fast computations for the approximated transition density;
3. Theory available for multidimensional SDEs (both time-homogeneous and time-inhomogeneous, see also Egorov et al. (2003))

Drawbacks:

1. In practice it may be difficult to apply the method to multidimensional SDEs, especially when the noise is *non-reducible* (Ait-Sahalia (2008));
2. The tails of the true transition density may be estimated inadequately when Δ is not "small enough" (Stramer-Yan (2007))
3. a symbolic calculus program may be necessary

Let's see some simulations...

In the previous lecture we considered a SDE model of the EHC for glycemia and insulinemia dynamics (Picchini *et al.* (2006))

$$\begin{aligned} dG(t) &= \left[\frac{(T_{gx}(t-\tau_g)+T_{gh}(t))}{V_g} - T_{xg}\frac{G(t)}{0.1+G(t)} - K_{xgl}G(t)I(t)\right]dt \\ &\quad +\sigma G(t)I(t)dW(t) \\ dI(t) &= \left[\frac{(T_{iG}G(t)+T_{ix}(t))}{V_i} - K_{xi}I(t)\right]dt \end{aligned}$$

but we noticed that **estimating ALL the 12 parameters (structural + measurement error-covariance) by simulating thousands of trajectories to approximate the transition densities was numerically unfeasible**, especially if bootstrap approaches are needed to compute confidence intervals.

Here we simplify the model to a one-dimensional SDE, restricting attention to glucose dynamics after the steady state of insulin concentration has been reached and disregarding measurement error.

Consider the glycemia state variable X_t, we simplify the diffusion part of the SDE in order for it to not contain insulinemias I_t, since

after 40 min from the start of the EHC, I_t should be nearly constant, so we fix $I_t \equiv I^*$ and in the sequel we assume $t \geq t_0 = 40$

$$dX_t = \mu(X_t, t)dt + \sigma(X_t)dW_t, \qquad t \geq t_0$$

where

$$\begin{aligned} \mu(X_t, t) &= \frac{T_{gx}(t-\tau_g)+T_{ghnet}}{V_g} - K_{xgl}I^*X_t, \\ \sigma(X_t) &= \sigma I^* X_t \\ T_{gx}(t) &= \sum_{\nu_j \leq t} \frac{(\lambda_j - \lambda_{j-1})\cdot(t-\nu_j)^5}{\nu_j + (t-\nu_j)^5}, \qquad t > 0, \quad \lambda_0 = 0, \quad j = 1, ..., m, \end{aligned}$$

The function $T_{gx}(t)$, which modelizes the glucose infusion rates $\{\lambda_j\}_j$ measured at times $\{\nu_j\}_j$, depends explicitly on t and so the SDE is *time-inhomogeneous*.

The extension of the Ait-Sahalia's method for time-inhomogeneous diffusions is given in Egorov et al. (2003): in this case an expansion of order 2 (in powers of Δ) for the transition density p_Z of Z is given by:

$$p_Z^{(2)}(z, t|y_s, s) = \phi(z) \sum_{k=0}^{4} \beta_k^{(2)}(t, y_s, s) H_k(z)$$

$$\begin{aligned}
\beta_0^{(2)}(t, y_s, s) &= 1 \\
\beta_1^{(2)}(t, y_s, s) &= -\Delta^{1/2}\psi - \frac{\Delta^{3/2}}{4}(2\psi_{01} + 2\psi\psi_{10} + \psi_{20}) \\
\beta_2^{(2)}(t, y_s, s) &= \frac{\Delta}{2}(\psi^2 + \psi_{10}) + \frac{\Delta^2}{12}(6\psi\psi_{01} + 6\psi^2\psi_{10} \\
&+ 4\psi_{10}^2 + 4\psi_{11} + 7\psi\psi_{20} + 2\psi_{30}) \\
\beta_3^{(2)}(t, y_s, s) &= -\frac{\Delta^{3/2}}{6}(\psi^3 + 3\psi\psi_{10} + \psi_{20}) \\
\beta_4^{(2)}(t, y_s, s) &= \frac{\Delta^2}{24}(\psi^4 + 6\psi^2\psi_{10} + 3\psi_{10}^2 + 4\psi\psi_{20} + \psi_{30})
\end{aligned}$$

where the ψ's are partial derivatives of μ_Y w.r.t. y and s.

Using the Jacobian transformation, given $p_Z^{(2)}$ it is straightforward to retrieve $p_X^{(2)}$, and then

$$L_n^{(2)}(\theta) = \sum_{i=1}^{n} \ln p_X^{(2)}(x_i, t_i|x_{i-1}, t_{i-1})$$

is the order 2 Hermite approximation of the likelihood function.
So we estimate ALL parameters entering the SDE

$$\theta = \{K_{xgI}, T_{ghnet}, V_g, \sigma\}$$

(τ_g has been fixed to 1 min.)
via

$$\theta^{(2)} = \arg\min_{\theta} -L_n^{(2)}(\theta),$$

which is a consistent estimate of θ.

$\hat{K}_{xgI} \times 10^4$ [min^{-1}/pM]	$\hat{T}_{ghnet}$ [mmol/min/kgBW]	$\hat{V}_g$ [L/kgBW]	$\hat{\sigma} \times 10^5$ [pM]$^{-1}$[min]$^{-1/2}$
2.951 [2.874, 3.028]	0.175	0.463 [0.451, 0.474]	7.542 [7.395, 7.688]
5.782 [5.658, 5.906]	0.085	0.117 [0.115, 0.120]	12.213 [12.023, 12.402]
5.553 [5.519, 5.587]	0.346	0.211 [0.210, 0.213]	5.878 [5.856, 5.900]
3.594 [3.472, 3.714]	1.051	0.99 [0.952, 1.028]	13.891 [13.588, 14.194]
2.531 [2.455, 2.607]	0.192	0.371 [0.359, 0.383]	6.717 [6.585, 6.849]

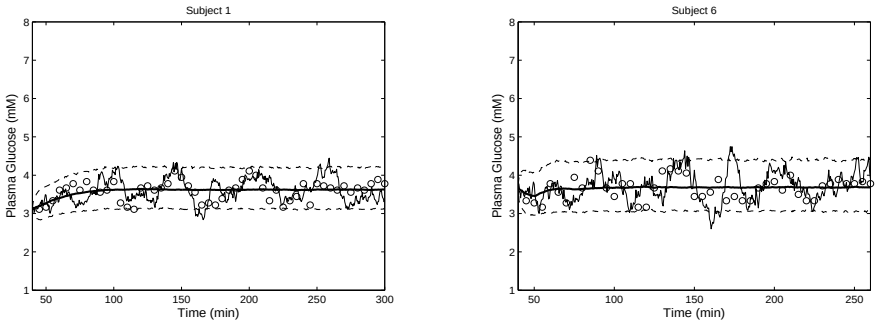


Figure: Observed glycemias (o), mean and 95% confidence intervals of 2000 trajectories.

Improvements and limitations

- In the previous lecture a two-dimensional SDE model with 12 parameters considering simultaneously the dynamics of glucose and insulin was analyzed (Picchini *et al.*, JMB 2006) but...
- parameter estimation proved difficult (**several hours** for the estimations of the diffusion coefficient σ only).
- In the present work **all** the parameters can be estimated in few seconds (the time required for a single evaluation of the likelihood function is of the order of **milliseconds**)...
- using a single common PC.
- limitations: we needed to consider a much simpler model in order to apply the closed-form estimation method; furthermore measurement error has not been modelized.

References

- Y. Aït-Sahalia (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1), 223-262.
- Y. Aït-Sahalia (2008). Closed-form likelihood expansion for multivariate diffusions. *Ann. Stat*, 36(2), 906-937.
- G.B. Durham and A.R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics*, 20(3), 297-316.
- A.V. Egorov, H. Li and Y. Xu (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics* 114, 107-139.
- B. Jensen and R. Poulsen (2002). Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques. *Journal of Derivatives*, 9(4), 18-32.

References

- U. Picchini, S. Ditlevsen and A. De Gaetano (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology*, 25(2), 141-155.