Estimation of a partially observed Ornstein-Uhlenbeck model in anti-cancer therapy

Adeline Samson

Current work in collaboration with D Balvay, CA Cuenod, B Favetto, V Genon-Catalot and Y Rozenholc

MAP5, Université Paris Descartes

UNIVERSITÉ PARIS DESCARTES

Context

Anti-angiogenesis treatments

- Promising anti-cancer therapy
- Need to evaluate effects of drugs *in vivo*
- Estimation of tissue microcirculation parameters
- Difference in microcirculation parameters along time = measure of treatment impact

Acquisition Protocol

Experiment

- Patient with ovary cancer
- Bolus injection of contrast agent
- Dynamic acquisition of gradient-echo MRI

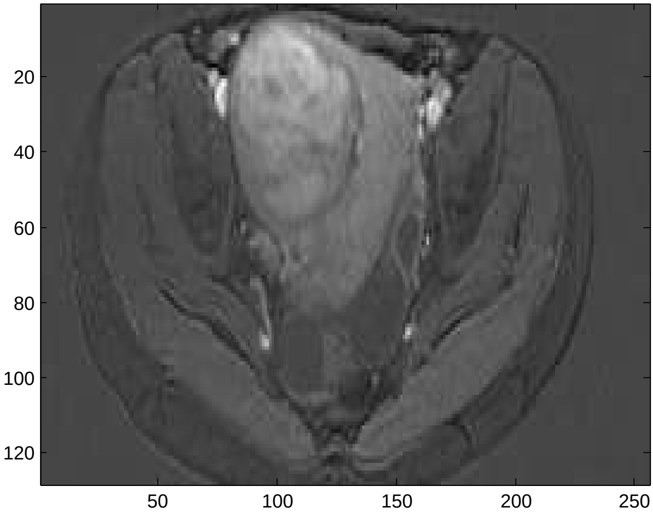
Observation times

- Beginning 10 seconds after injection
- 130 images, every 2.4 seconds

Tissue microcirculation parameters

- Model of contrast agent pharmacokinetic
- Estimation of *in vivo* tissue microcirculation parameters

Example of one image in the sequence



High level of noise

Observation model

Measurement

- Intensity of gray level on a voxel $I(t)$
- Assumption : $I(t)$ *proportional to the quantity of contrast agent in voxel* $Q(t)$

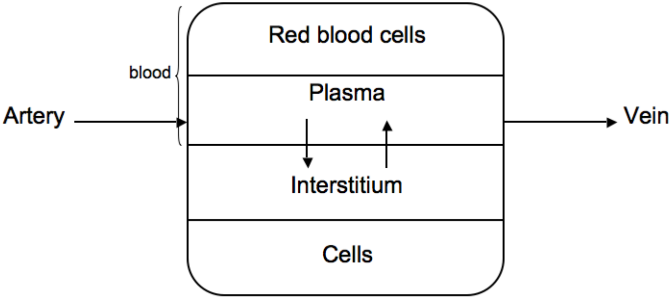
$$I(t) - I(0) = Q(t)$$

Observations

- Discrete times $t_0, \ldots, t_n$
- Noisy observations y

$$\begin{aligned} y_i &= Q(t_i) + \sigma\varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, 1) \end{aligned}$$

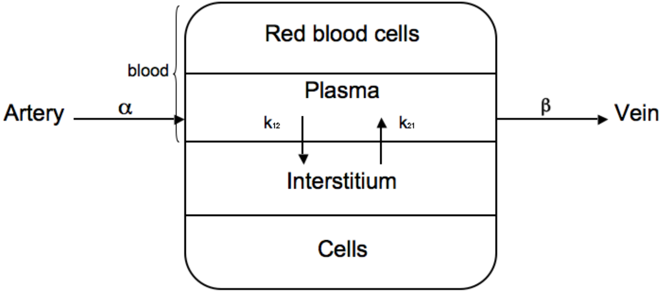
Pharmacokinetic model



Quantity of contrast agent in

- Arterial voxel: Arterial Input Function (AIF)
  - Assumed to be known
- Non arterial voxel $Q(t) = Q_P(t) + Q_I(t)$
  - Plasma: $Q_P(t)$
  - Interstitium: $Q_I(t)$

Two-compartment model



$$Q(t) = Q_P(t) + Q_I(t)$$

$$\begin{array}{rcccc} \frac{dQ_P(t)}{dt} & = & \alpha AIF(t) & -(k_{12}+\beta)Q_P(t) & +k_{21}Q_I(t) \\ \frac{dQ_I(t)}{dt} & = & & k_{12}Q_P(t) & -k_{21}Q_I(t) \end{array}$$

Initial condition at time t_0 : $Q_P(t_0) = Q_I(t_0) = 0$

Stochastic extension

Drawback of ordinary differential equations

- Smooth theoretical model
- Failure to capture
  - Fluctuations in plasma/interstitium permeability
  - Movement of patient (breathing)
- Numerical instability

Stochastic approach

- Random fluctuations around deterministic model
- Keep same interpretation of physiological parameters

Stochastic model

Observations

$$\begin{aligned} y_i &= Q_P(t_i) + Q_I(t_i) + \sigma\varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0,1) \end{aligned}$$

Stochastic model

$$\begin{aligned} dQ_P(t) &= (\alpha AIF(t) - (k_{12}+\beta)Q_P(t) \quad + k_{21}Q_I(t)\,)\,dt \quad + \sigma_1 dW_t^1 \\ dQ_I(t) &= (k_{12}Q_P(t) \quad - k_{21}Q_I(t))\,dt \quad + \sigma_2 dW_t^2 \end{aligned}$$

- W_t^1, W_t^2: independent Brownian motions
- σ_1, σ_2: unknown standard deviations

Objectives

Our aim: Exact Maximum likelihood estimates of our model

- Bi-dimensional Ornstein-Uhlenbeck model
- Partially observed
- Noisy discrete observations

Estimation methods

- Minimization of a contrast (Genon-Catalot and Jacod '93, Kessler '97)
- Martingale estimating functions (Bibby and Sorensen '95)
- Approximation of density distribution (Ditlevsen et al '05, Picchini et al '06)
- ...

⇒ Development of an exact MLE procedure

Maximum likelihood estimation

Two ways of computing the exact likelihood $p(y;\theta)$ with $\theta = (\alpha, \beta, k_{12}, k, \sigma_1, \sigma_2, \sigma)$

1. Direct approach: calculus of joined data $(y_0, \ldots, y_n)$ distribution
2. Discretization approach: Kalman Filter

Matricial formulation

Set $U(t) = (S(t), Q_I(t))'$ with $S(t) = Q_P(t) + Q_I(t)$

$$\begin{aligned} dU_t &= (G\ U_t + F(t))dt + \Sigma dW_t \\ U(t_0) &= U_0 \\ y_i &= (1\ \ 0)\ U(t_i) + \sigma\varepsilon_i \end{aligned}$$

with $\lambda = k_{12}$ and $k = k_{12} + k_{21}$

$$F = \begin{pmatrix} \alpha AIF(t) \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} -\beta & \beta \\ \lambda & -k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_2 \end{pmatrix}$$

Result: G is diagonalizable

- D diagonal matrix of eigenvalues (μ_1, μ_2) of G
- P transit matrix of eigenvectors
- new process $X = P^{-1}U$ in the new basis
- $\Gamma = P^{-1}\Sigma$

Solving the model

SDE in the new basis

$$
\begin{aligned}
dX_t &= (D\ X_t + P^{-1}F(t))dt + \Gamma dW_t \\
X(t_0) &= P^{-1}U_0 \\
y_i &= (1\ \ 1)\ X(t_i) + \sigma\varepsilon_i
\end{aligned}
$$

Result:

- $X(t+h)|X(t) \sim \mathcal{N}_2\left(e^{Dh}X(t) + B\left(t, t+h\right), Q\left(t, t+h\right)\right)$ where

$$
\begin{aligned}
B(t, t+h) &= e^{D(t+h)} \int_t^{t+h} e^{-Ds} P^{-1} F(s) ds \\
Q(t, t+h) &= \left(\frac{e^{(\mu_k + \mu_{k'})h} - 1}{\mu_k + \mu_{k'}} (\Gamma\Gamma')^{kk'} \right)_{1 \le k, k' \le 2}
\end{aligned}
$$

- Stationnary distribution for (X_t) if $\alpha = 0$ $(F(s) = 0)$

1. Direct approach

Continuous model solution

- $X(t)$ bi-dimensional Gaussian process
- Explicit distribution, with covariance matrix of dimension $2n \times 2n$

Computation of likelihood

- y Gaussian vector
- Expectation and variance derived from those of X

Maximization of exact likelihood

- Gradient descent but no direct way of computing the exact gradient of the likelihood
- Numerical method

2. Discretization approach

Discretization of SDE: $X_i = X(t_i)$

$$\begin{aligned} X_i &= A_i X_{i-1} + B_i + \eta_i \\ \eta_i &\sim \mathcal{N}(0, Q_i) \\ X_0 &= x_0 \\ y_i &= H X_i + \sigma \varepsilon_i \end{aligned}$$

with

$$A_i = \exp(D\,(t_i - t_{i-1})), \quad B_i = B(t_{i-1}, t_i), \quad Q_i = Q(t_{i-1}, t_i), \quad H = (1 \;\; 1)$$

$\Rightarrow$ Hidden Markov Model

Likelihood

Computation of exact likelihood of the discrete model

$$\begin{aligned} L(y_0, \ldots, y_n; \theta) &= p(y_0; \theta) \prod_{i=1}^{n} p(y_i | y_0, \ldots, y_{i-1}; \theta) \\ &= L(y_0, \ldots, y_{n-1}; \theta) p(y_n | y_0, \ldots, y_{n-1}; \theta) \end{aligned}$$

From calculus on Gaussian distributions, we deduce

$$y_i | y_0, \ldots, y_{i-1}; \theta \sim \mathcal{N}(m_i(\theta), V_i(\theta))$$

where $m_i(\theta)$ and $V_i(\theta)$ depend on $p(X_i | y_0, \ldots, y_{i-1}; \theta)$

Kalman filter

Kalman filter computes iteratively expectations and variances

- $p(X_i|y_0,\ldots,y_{i-1};\theta)$ (*prediction distribution*)
- $p(X_i|y_0,\ldots,y_i;\theta)$ (*filter distribution*)

We deduce iterative computations of

- Expectation and variance of $y_i|y_0,\ldots,y_{i-1};\theta$

$$m_i(\theta) = F_\theta(m_{i-1}(\theta)) \text{ and } V_i(\theta) = G_\theta(V_{i-1}(\theta))$$

- Log-likelihood $l_{0:i}(\theta) = \log L(y_0,\ldots,y_i;\theta)$

$$\begin{aligned} l_{0:i}(\theta) &= l_{0:i-1}(\theta) + \log p(y_i|y_0,\ldots,y_{i-1};\theta) \\ &= l_{0:i-1}(\theta) - \frac{1}{2}\log(2\pi V_i(\theta)) - \frac{1}{2}\frac{(y_i - m_i(\theta))^2}{V_i(\theta)} \end{aligned}$$

$\Rightarrow$ No inversion of large covariance matrix

Likelihood maximization

Maximization using a gradient descent method

- Need to compute the gradient and hessian of the log-likelihood

Assumption

- Equally spaced time observations $t_i - t_{i-1} = \Delta$
- New parametrization

$$A_i = A(\theta) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix} Q_i = Q(\theta) = \begin{pmatrix} \theta_3 & \theta_5 \\ \theta_5 & \theta_4 \end{pmatrix}$$

Gradient descent

Iterative computation of

- $\partial m_i(\theta)/\partial\theta$ and $\partial V_i(\theta)/\partial\theta$ by deriving

$$\begin{aligned} m_i(\theta) &= F_\theta(m_{i-1}(\theta)) \\ V_i(\theta) &= G_\theta(V_{i-1}(\theta)) \end{aligned}$$

- Gradient and hessian of the log likelihood by deriving

$$l_{0:i}(\theta) = l_{0:i-1}(\theta) - \frac{1}{2}\log(2\pi V_i(\theta)) - \frac{1}{2}\frac{(y_i - m_i(\theta))^2}{V_i(\theta)}$$

$\Rightarrow$ Efficient way to implement the gradient descent method and to obtain maximum likelihood estimate

Theoretical results

Link with an ARMA model

Proposition

Under assumption $\alpha = 0$,

- The process $(y_i)_{i \in \mathbb{Z}}$ is ARMA(2,2)
- The spectral density of $(y_i)_{i \in \mathbb{Z}}$ has an explicit form

$$f(u, \theta) = \frac{\gamma(0)+\gamma(1)2\cos(u)+\gamma(2)2\cos(2u)}{1+(\theta_1+\theta_2)^2+\theta_1^2\theta_2^2-2(\theta_1+\theta_2)(1+\theta_1\theta_2)\cos(u)+2\cos(2u)\theta_1\theta_2}$$

Identifiability problem

- From spectral density, only five parameters (out of 6) are identifiable:

  $(\theta_1 + \theta_2)$,$\theta_1\theta_2$, $\gamma(0)$, $\gamma(1)$, $\gamma(2)$
- From simulations, better results obtained when σ^2 is fixed

Consistency of the MLE

θ_0 the true value of parameter
Asymptotic information matrix defined by

$$\text{for } i,j \in \{1,\dots,5\} \quad I(\theta)_{i,j} = \int_{\mathbb{T}} \frac{\partial}{\partial \theta_i} \log f(u,\theta) \frac{\partial}{\partial \theta_j} \log f(u,\theta) du$$

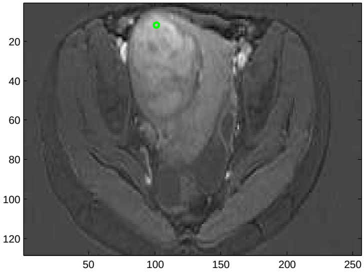
Proposition

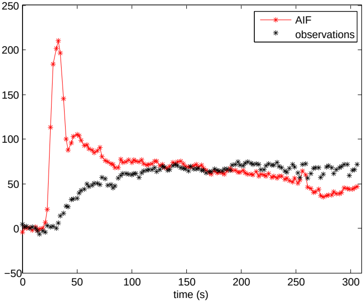
Let $\hat{\theta}_n$ be a maximum likelihood estimator of θ_0 based on $(y_0,\dots,y_n)$. Then, $\hat{\theta}_n \to \theta_0$ a.s. as $n \to \infty$. Moreover, if $I(\theta_0)$ is invertible, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, I^{-1}(\theta_0))$$

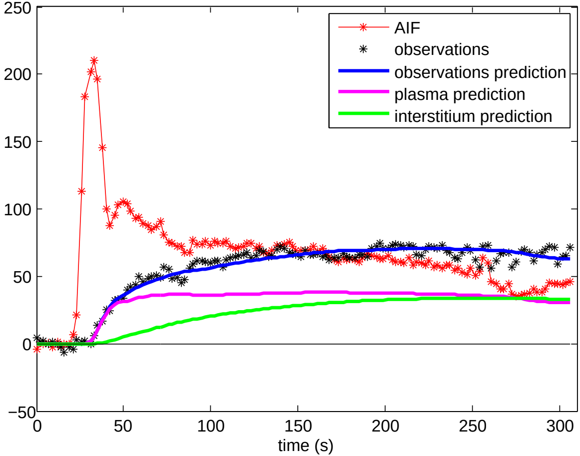
Proof. Times series results (Brockwell and Davis '91).

Application to real data



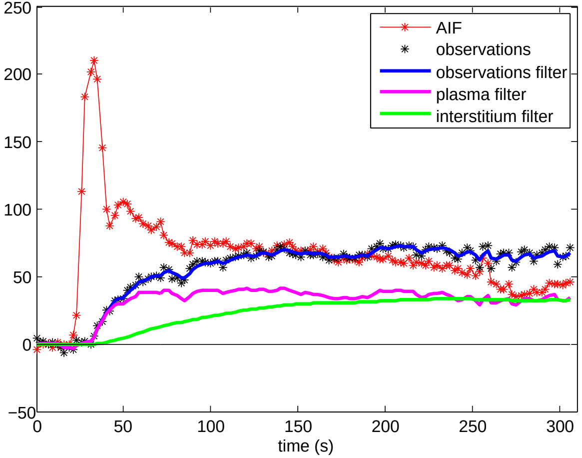


Direct approach



$\alpha = 0.015,\ \beta = 0.024,\ k_{12} = 0.014,\ k_{21} = 0.014,\ \delta = 5.94$
$\sigma_1 = 1.07, \sigma_2 = 0.07, \sigma = 2.96$

Kalman approach



$\alpha = 0.015,\ \beta = 0.024,\ k_{12} = 0.014,\ k_{21} = 0.014,\ \delta = 5.94$
$\sigma_1 = 1.07, \sigma_2 = 0.07, \sigma = 2.96$

Discussion

Conclusion

- Adequate physiological model
- Exact maximum likelihood estimation of model parameters

Perspectives

- Map of the microcirculation estimated parameters on all the voxels
- Expectation-Maximization algorithm with smoother Kalman algorithm (work on progress)
- Extension of SDE model to ensure positive solutions
- Extension to non equidistant observations times
- Extension to *d*-dimensional O-U partially observed process