Sequential Monte Carlo Methods for Bayesian Model Selection in Positron Emission Tomography

Yan Zhou, John A.D. Aston and Adam M. Johansen

6th January 2014
Outline

PET compartmental models
- Positron emission tomography (PET)
- Linear Compartmental models
- Plasma input PET compartmental models

Bayesian model selection for PET
- Robust modeling of the error structure
- Biologically informative priors

Sequential Monte Carlo
- Algorithm setting for Bayesian modeling
- Computational challenge
  - Accuracy of estimators
  - Heterogeneous structure and algorithm tuning
  - Computational cost and parallel computing

Results / Conclusions / References

Conclusions
Use compounds labeled with positron emission radionuclides as molecular tracers to image and measure biochemical process in vivo.

One of the few methods available to neuroscientists to study living brains.

Research into diseases where biochemical changes are known to be responsible symptomatic changes.

For example, diagnostic procedure for cancer through fluorodeoxyglucose (\(^{18}\text{F}\)-FDG) tracers.
Linear Compartmental models

- Comprise a finite number of macroscopic subunits called *compartments*.
- Each is assumed to contain homogeneous and well-mixed material.
- Material flows from one compartment to another at a constant rate.
- In PET *total* concentration of material is measured.

These models yield systems of ODEs:

\[
\dot{f}(t) = Af(t) + b(t) \\
\]

\[
f(0) = \xi
\]
N.B. We actually focus on linear compartmental models.
Plasma input PET compartmental models

System,

\[ \dot{C}_T(t) = AC_T(t) + bC_P(t) \]
\[ C_T(t) = 1^T C_T(t) \]
\[ C_T(0) = 0 \]

Solution,

\[ C_T(t) = \int_0^t C_P(t - s) H_{TP}(s) \, ds \]
\[ H_{TP}(t) = \sum_{i=1}^{r} \phi_i e^{-\theta_i t} \]

Parameter of interest,

\[ V_D = \int_0^{\infty} H_{TP}(t) \, dt = \sum_{i=1}^{r} \frac{\phi_i}{\theta_i} \]
Bayesian model selection for PET

- Determine the number of tissue compartments.
- "Mass univariate analysis."
  - Each time course of $C_T(t)$ is analyzed individually.
  - Many: quarter of a million time series per PET scan.
- Data is measured at discrete times $t = t_1, \ldots, t_n$,

$$y_i = C(t_i) + \sqrt{\frac{C(t_i)}{t_i - t_{i-1}}} \varepsilon_i$$

where $\varepsilon_i$ are (iid) errors.
Typical PET Time Courses

Data Set 1

Data Set 2

Data Set 3

Concentration (kBq/mL)

Time (sec)
Robust modeling of the error structure

- Low signal to noise ratio.
- Standard approach (in likelihood-based procedures)
  - Use Normal distributions to model the error.
  - Employ weighted Non-negative Least Squares.
  - Assign (arbitrary) small weights to the most noisy data points.
- Bayesian modeling
  - No justifiable way to bound “weights” with normal errors.
  - Need more robust modeling of the error structure.
- Simple solution:
  Use three-parameter $t$ distribution instead of Normal.
Biologically informative priors [Zhou et al., 2013a]

Starting point:

- Parameters $\phi_{1:r}$ and $\theta_{1:r}$ are functions of the rate constants.
- The matrix $A$ of rate constants obey some simple rules.
- Rate constants are constrained by biophysical considerations.

Key observations: For $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_r$: into the environment.

- In the linear plasma input model, there is one outflow, $k_2$, $\theta_1 \leq k_2$.
- There is also only one inflow $K_1$, $\sum_{i=1}^{r} \phi_i = K_1$.

Biophysical knowledge constrains possible values for $\phi_{1:r}$ and $\theta_{1:r}$.
Sequential Monte Carlo [Del Moral et al., 2006]

- Iteratively generate importance sampling proposal distributions for a sequence $\{\pi_t\}_{t=0}^T$.
- Use MCMC kernels to propose samples

1. Generate $\{X_0^{(i)}\}_{i=1}^N$ from $\pi_0$. Set $\{W_0^{(i)}\}_{i=1}^N$, the importance weights, to $1/N$.

2. For $t = 1, \ldots, T$,
   2.1 Resample if necessary.
   2.2 Generate $\{X_t^{(i)}\}_{i=1}^N$ from $K(x_{t-1}, x_t)$, a $\pi_t$-invariant Markov kernel.
   2.3 Set $W_t^{(i)} \propto W_{t-1}^{(i)} \tilde{w}_t^{(i)}$, where $\tilde{w}_t^{(i)} \propto \pi_t(X_t^{(i)})/\pi_{t-1}(X_t^{(i)})$. 
Algorithm setting for Bayesian modeling

Sequence of distributions,

$$\pi_t(\varphi) \propto \pi_0(\varphi)[L(\varphi|y_1:n)]^{\alpha(t/T)}$$

where $\varphi$ is the parameter vector, $\pi_0$ is the prior and $L$ is the likelihood function.

Markov kernels,

- Update $\phi_{1:r}$ with Normal random walks.
- Update $\theta_{1:r}$ with Normal random walks.
- Update $\lambda$, the scale parameter of the $t$ distributed error, with a Normal random walk on $\log \lambda$.
- Update $\nu$, the degree of freedoms of the $t$ distributed error, with a Normal random walk on $\log \nu$. 
Computational challenge

- Accuracy of estimator
- Heterogeneous structure
- Computational cost
Improve the accuracy of estimators [Zhou et al., 2013b]

- Increase the number of particles.
- Increase the number of intermediate distributions.
- Fast mixing Markov kernels.
  - Multiple MCMC passes each iteration.
  - Adaptive proposal scales for random walks.
- Better specification of intermediate distributions.
  - Place more distributions where $\pi_t$ changes fast when $\alpha(t/T)$ increases.
  - Adaptive specification such that the discrepancy between $\pi_t$ and $\pi_{t-1}$ remain almost constant.
Improve the accuracy of estimators —
adaptive specification of the sequence of distributions

Figure: Variation of the distribution specification parameter $\alpha(t/T)$ when using adaptive algorithms.
Heterogeneous structure and algorithm tuning

We cannot tune the algorithm for each of 250,000 time series.

Figure: Estimates of $V_D$ using selected model

- SMC is more robust compared than (our) MCMC.
- Adaptive strategies.

SMC for PET Model Selection
Y. Zhou, J. A. D. Aston and A. M. Johansen
Computational cost and parallel computing

- SMC can be parallelized naturally in contrast to MCMC.
- SMC can be parallelized more efficient compared to other algorithms, such as population MCMC.
  - We can increase the number of particles freely.
  - Increase the number of distributions in population MCMC come with a cost – global mixing speed.
- Well suited for SIMD architectures, such as GPUs:
  - They perform best when each thread does exactly the same thing.
Results

- Bayesian model selection for simulated data performance considerably better than methods such as AIC and BIC.
  - Higher frequency of selecting the true model.
  - More accurate parameter estimates.
  - Biological informative priors improve the results further (but results are fairly insensitive to the prior).
- Bayesian model selection for real data shows more plausible structures than existing techniques.
  - Voxels with higher volume of distributions ($V_D$) are expected to have higher order models associated with them.
Model selection results using AIC (above) / Bayes factor (below).
Conclusions

SMC is *not* “too computationally demanding” for neuroscience.

- Monte Carlo methods are feasible for large data problems.
- SMC can outperform MCMC even in time-limited settings such as this one.
- Many problems in neuroscience are amenable to similar solutions [Sorrentino et al., 2013, Nam et al., 2012]

Ongoing work on this problem seeks to replace the “mass univariate analysis” approach.
References


