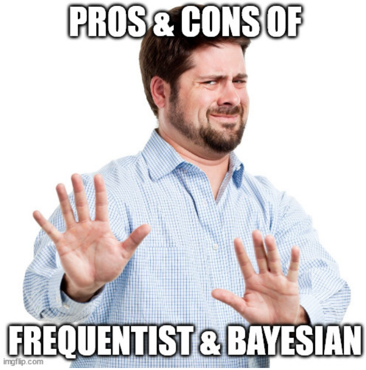
ODE parameter estimation in drug research: a deep dive into a Bayesian inference engine

Yi Zhang

<2023-05-26 Fri>

Outline

NOT to talk about



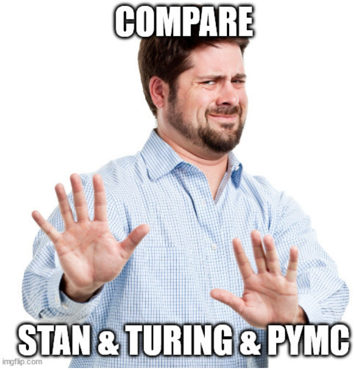


Table of contents

Pharmacokinetics (PK) data

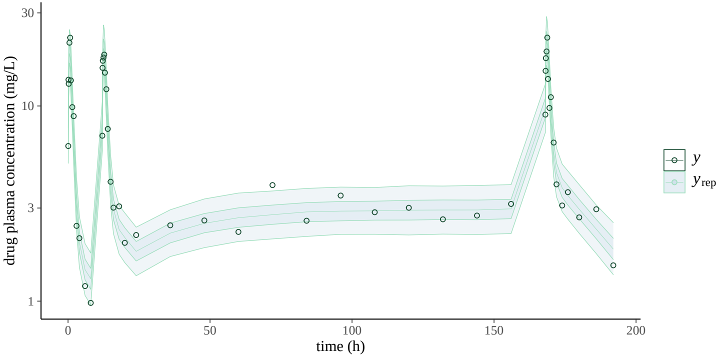
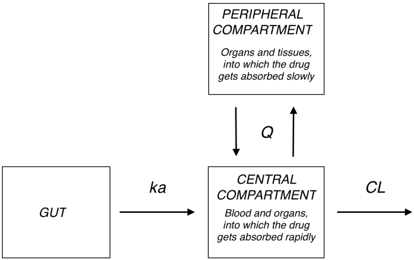


Figure: Subject plasma concentration history (q12hx14).

A Pharmacokinetics (PK) model

$$\frac{\mathrm{d}\hat{y}_{\mathrm{gut}}}{\mathrm{d}t} = -k_a \hat{y}_{\mathrm{gut}}$$
$$\frac{\mathrm{d}\hat{y}_{\mathrm{cent}}}{\mathrm{d}t} = k_a \hat{y}_{\mathrm{gut}} - \left(k_{\mathsf{CL}} + k_{\mathsf{Q,cent}}\right) \hat{y}_{\mathrm{cent}} + k_{\mathsf{Q,peri}} \hat{y}_{\mathrm{peri}}$$
$$\frac{\mathrm{d}\hat{y}_{\mathrm{peri}}}{\mathrm{d}t} = k_{\mathsf{Q,cent}} \hat{y}_{\mathrm{cent}} - k_{\mathsf{Q,peri}} \hat{y}_{\mathrm{peri}}$$



A statistical model

$$
\begin{aligned}
\theta &\equiv \{k_a, k_{\mathrm{CL}}, k_{\mathrm{Q},\cdot}\}, \\
\hat{y}(t) &= \hat{y}(t;\theta), \\
y(t) &\sim \mathrm{Normal}(\hat{y}(t;\theta), \sigma).
\end{aligned}
$$

- play mix & match with two models
- Likelihood $p(y|\theta)$.
- Usually likelihood $p(y|\theta, \sigma)$

Estimate θ

- Maximum likelihood estimation $\theta_{\text{argmax}}(p(y|\theta))$
- Posterior distribution

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\bigcirc}$$

$$\log p(\theta|y) = C + \log p(y|\theta) + \log p(\theta)$$

- Connection to regularization

Prior & likelihood

$$\log p(\theta|y) = C + \log p(y|\theta) + \log p(\theta)$$

Sample the posterior

target (distribution) $= \log p(\theta|y) = \log p(y|\theta) + \log p(\theta)$

- Posterior from an oracle: $\theta \to \log p(\theta|y)$

Sample the posterior

target (distribution) $= \log p(\theta|y) = \log p(y|\theta) + \log p(\theta)$

- Posterior from an oracle: $\theta \to \log p(\theta|y)$
- Every query costs one ODE solution

Sample the posterior

target (distribution) $= \log p(\theta|y) = \log p(y|\theta) + \log p(\theta)$

- Posterior from an oracle: $\theta \to \log p(\theta|y)$
- Every query costs one ODE solution
- Sampling procedure for a distribution: $\theta^{(1)}, \theta^{(2)}, \ldots$

$$\lim_{n \to} \frac{1}{n} \sum f(\theta^{(i)}) = \mathbb{E}(f)$$

Sample the posterior

target (distribution) $= \log p(\theta|y) = \log p(y|\theta) + \log p(\theta)$

- Posterior from an oracle: $\theta \to \log p(\theta|y)$
- Every query costs one ODE solution
- Sampling procedure for a distribution: $\theta^{(1)}, \theta^{(2)}, \ldots$

$$\lim_{n\to} \frac{1}{n} \sum f(\theta^{(i)}) = \mathbb{E}(f)$$

- MCMC: $p(\theta^{(i)}|y, \theta^{(i-1)}, \theta^{(i-2)}, \ldots, \theta^{(1)}) = p(\theta^{(i)}|y, \theta^{(i-1)})$

The Metropolis-Hastings algorithm

Proposal-rejection from $\theta^{(i)}$ to $\theta^{(i+1)}$.

1. Propose $\theta^{(i+1)}$ according to transition density $q(\theta^{(i+1)}, \theta^{(i)})$.
2. Accept $\theta^{(i+1)}$ with probability

$$\alpha(\theta^{(i)}, \theta^{(i+1)}) = \min\left[1, \frac{p(\theta^{(i+1)}|y)q(\theta^{(i)}, \theta^{(i+1)})}{p(\theta^{(i)}|y)q(\theta^{(i+1)}, \theta^{(i)})}\right] \tag{1}$$

otherwise reject.

- Chain generated by M-H has detailed balance with $p(\theta|y)$ as its stationary distribution.
- Convergence in TVD with proper proposal that garantees irreducibility.

Why MCMC?

Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse.
– Sokal, A. D. (1989). "Monte carlo methods in statistical mechanics: foundations and new algorithms."

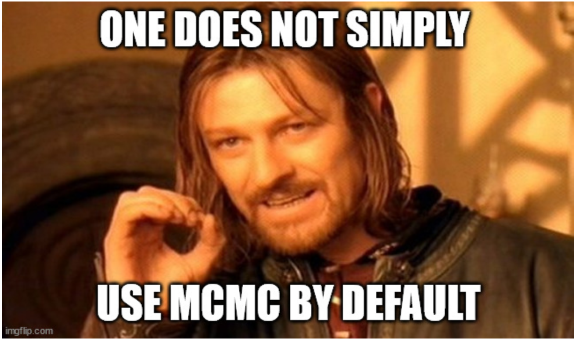
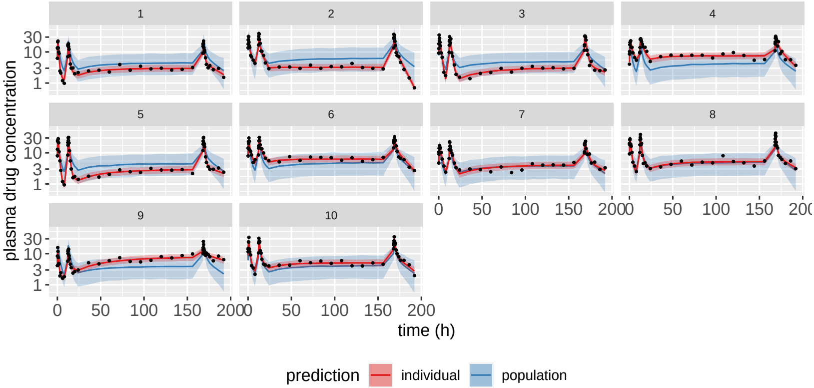


Table of contents

The problem



1
2
3
4
5
6
7
8
9
10
plasma drug concentration
30
10
3
1
0
50
100
150
200
time (h)
prediction
individual
population

Hierarchical model

$$
\begin{aligned}
\theta_0 &\sim \mathrm{Prior}(\cdot), \\
\theta_i &\sim \mathrm{MultiNormal}(\theta_0, \Sigma), \\
y_i &\sim \mathrm{Normal}(\hat{y}_i(\theta_i), \sigma), \\
\theta &= \{\theta_0, \theta_1, \ldots, \theta_n, \sigma\}
\end{aligned}
$$

- Posterior(likelihood) equation is an oracle

Hierarchical model

$$
\begin{aligned}
\theta_0 &\sim \text{Prior}(\cdot), \\
\theta_i &\sim \text{MultiNormal}(\theta_0, \Sigma), \\
y_i &\sim \text{Normal}(\hat{y}_i(\theta_i), \sigma), \\
\theta &= \{\theta_0, \theta_1, \ldots, \theta_n, \sigma\}
\end{aligned}
$$

- Posterior(likelihood) equation is an oracle
- Curse of dimensionality
  - Computational tractability

Hierarchical model

$$
\begin{aligned}
\theta_0 &\sim \text{Prior}(\cdot), \\
\theta_i &\sim \text{MultiNormal}(\theta_0, \Sigma), \\
y_i &\sim \text{Normal}(\hat{y}_i(\theta_i), \sigma), \\
\theta &= \{\theta_0, \theta_1, \dots, \theta_n, \sigma\}
\end{aligned}
$$

- Posterior(likelihood) equation is an oracle
- Curse of dimensionality
  - Computational tractability
  - Concentration of measure (e.g. high dimensional gaussian distribution is like uniform distribution)

Hierarchical model

$$\begin{aligned}
\theta_0 &\sim \mathrm{Prior}(\cdot), \\
\theta_i &\sim \mathrm{MultiNormal}(\theta_0, \Sigma), \\
y_i &\sim \mathrm{Normal}(\hat{y}_i(\theta_i), \sigma), \\
\theta &= \{\theta_0, \theta_1, \ldots, \theta_n, \sigma\}
\end{aligned}$$

- Posterior(likelihood) equation is an oracle
- Curse of dimensionality
  - Computational tractability
  - Concentration of measure (e.g. high dimensional gaussian distribution is like uniform distribution)
- Geometry of posterior

Challenges: high dimensional gaussian distribution

$$p(|\|y_d\|_2 - \sqrt{d}| \geq t) \leq 2\exp\left(-ct^2\right), \forall t \geq 0.$$

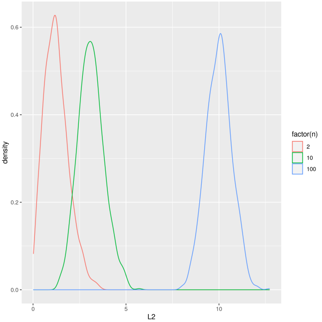
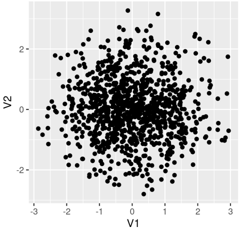
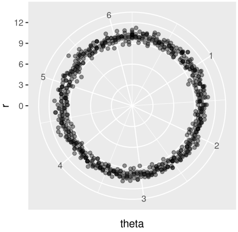


Figure: $\|y_d\|_2, y_d \sim \text{Normal}(0, \mathbb{I}_d)$





Average is not representative.
Random Walk sampler is not efficient.

Challenges: Geometry of posterior

$$
\begin{aligned}
\theta_0 &= 0, \\
\kappa &\sim \text{Normal}(0, 3), \\
\theta_i(k) &\sim \text{Normal}(0, \exp(\kappa/2)), \\
k &= 1, 2, \ldots
\end{aligned}
$$

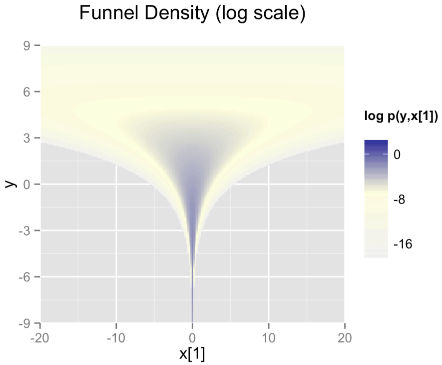


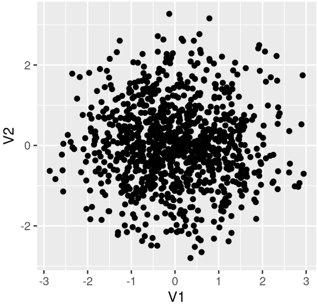
Figure: Neal's funnel

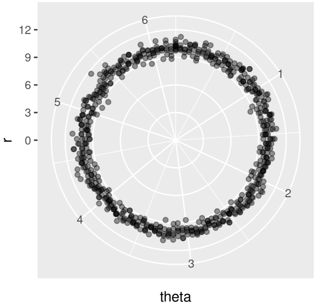
Mode is not representative.
Optimizer is not efficient.

Table of contents

Hamiltonian Monte Carlo

$$H(\theta, r) = -\log p(r, \theta|y) = T(r) + V(\theta) = -\log r - \log p(\theta|y),$$
$$\frac{d\theta}{dt} = \frac{\partial H}{\partial r}, \qquad \frac{dr}{dt} = -\frac{\partial H}{\partial \theta},$$





Hamiltonian Monte Carlo

$$H(\theta, r) = -\log p(r, \theta|y) = T(r) + V(\theta) = -\log r - \log p(\theta|y),$$

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial r}, \qquad \frac{dr}{dt} = -\frac{\partial H}{\partial \theta},$$

Apply M-H to $p(r, \theta)$

$$\alpha((r^{(i)}, \theta^{(i)}), (r^{(i+1)}, \theta^{(i+1)})) = \min\left[1, \frac{p(r^{(i+1)}, \theta^{(i+1)}|y)q()}{p(r^{(i)}, \theta^{(i)}|y)q()}\right]$$
$$= \min\left[1, \frac{p(r^{(i+1)}, \theta^{(i+1)}|y)}{p(r^{(i)}, \theta^{(i)}|y)}\right]$$

$$\alpha(\cdot, \cdot) = \min\left[1, \exp\left(H(\theta^{(i)}, r^{(i)}) - H(\theta^{(i+1)}, r^{(i+1)})\right)\right]$$

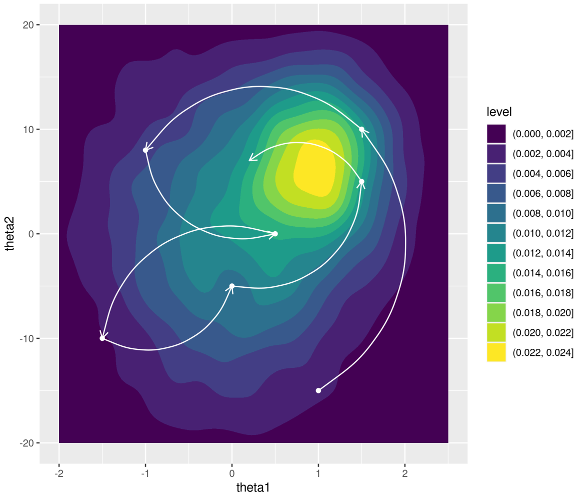
Hamiltonian Monte Carlo

$$p(\theta^{(i)} \to \theta^{(i+1)}) = \min\left[1, \exp\left(H(\theta^{(i)}, r^{(i)}) - H(\theta^{(i+1)}, r^{(i+1)})\right)\right]$$

Proposal $(r^{(i+1)}, \theta^{(i+1)})$:

$$\begin{gathered} r \sim \mathrm{Normal}(0, M), \\ (r^{(i)}, \theta^{(i)}) \to (r^{(i+1)}, \theta^{(i+1)}) \end{gathered}$$

A principled sampler for $p(\theta|y)$



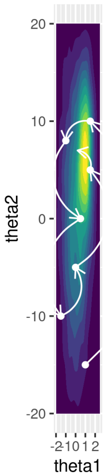


Table of contents

A tale of two ODEs

$\theta^{(\mathrm{i})} = \{\theta^{(\mathrm{i})}{}_{\mathrm{j}}\}$, j=1,2,...,n: $(r, \theta)(\tau^{(i)}) \to (r, \theta)(\tau^{(i+1)})$

$$\begin{cases} \theta_j^{(i)} \to \hat{y}_j(t; \theta_j^{(i)}), \\ y_{jk} \sim \text{Normal}(\hat{y}_{jk}(\theta_j), \sigma), \quad p(y_{jk}|\theta_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\hat{y}_{jk}(\theta_j) - y_{jk})^2}{2\sigma^2}\right] \end{cases}$$

Nested ODE solvers:

$$r^{(i+1/2)} = r^{(i)} - \frac{h}{2}\nabla_\theta \log p(\theta^{(i)}|y), \text{ a step in leapfrog}$$

$$\nabla_\theta \log p(\theta^{(i)}|y) = \nabla_\theta \log p(\theta^{(i)}) + \nabla_\theta \log p(y|\theta^{(i)}),$$

$$\nabla_\theta \log p(y|\theta^{(i)}) = -(\cdots)\sum_{j,k} \nabla_\theta \frac{\hat{y}_{jk}(\theta_j^{(i)}) - y_{jk}}{\sigma^{(i)}} + \ldots,$$

$$\nabla_\theta \hat{y}_{jk} : \text{ forward or adjoint sensitivity solver}$$

Automatic differentiation

For subject $i = 1, 2, \ldots$, assuming single observation:

$$\pi(\theta_1, \theta_2, \ldots) = \log p(\theta|y) = -\frac{(\hat{y}_1 - y_1)^2}{2\sigma^2} - \frac{(\hat{y}_2 - y_2)^2}{2\sigma^2} - \ldots,$$

$$\left[b_1 = \frac{(\hat{y}_1 - y_1)^2}{\sigma^2}, \frac{d\pi}{db_1} = -\frac{1}{2}\right], \quad \ldots,$$

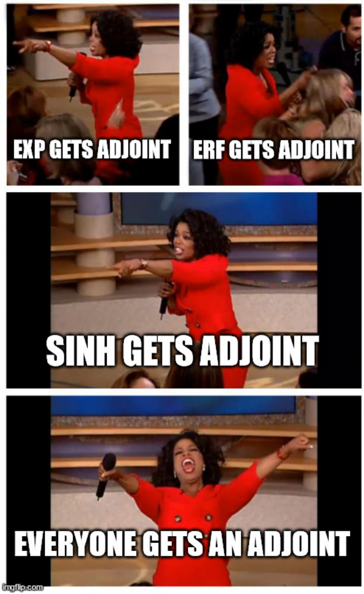
$$\left[a_1 = (\hat{y}_1 - y_1)^2, \frac{d\pi}{da_1} = \frac{d\pi}{db_1}\frac{db_1}{da_1} = \frac{d\pi}{db_1}\frac{1}{\sigma^2}\right], \quad \ldots,$$

$$\left[\hat{y}_1, \frac{d\pi}{d\hat{y}_1} = \frac{d\pi}{da_1}\frac{da_1}{d\hat{y}_1}\right], \quad \ldots,$$

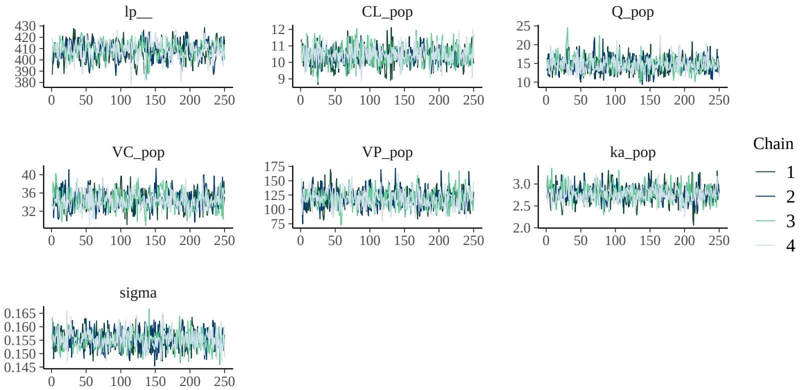
$$\left[\theta_1, \frac{d\pi}{d\theta_1} = \frac{d\pi}{d\hat{y}_1}\frac{d\hat{y}_1}{d\theta_1}\right], \quad \ldots,$$

Automatic differentiation

function	*arguments*	*return*
`abs`	`(real x)`	`real`
`acos`	`(real x)`	`real`
`acosh`	`(real x)`	`real`
`asin`	`(real x)`	`real`
`asinh`	`(real x)`	`real`
`atan2`	`(real x, real y)`	`real`
`atan`	`(real x)`	`real`
`atanh`	`(real x)`	`real`
`cos`	`(real x)`	`real`
`cosh`	`(real x)`	`real`
`cbrt`	`(real x)`	`real`
`ceil`	`(real x)`	`real`
`erf`	`(real x)`	`real`
`erfc`	`(real x)`	`real`
`exp2`	`(real x)`	`real`
`exp`	`(real x)`	`real`
`fdim`	`(real x, real y)`	`real`
`floor`	`(real x)`	`real`
`fma`	`(real x, real y, real z)`	`real`
`fmax`	`(real x, real y)`	`real`
`fmin`	`(real x, real y)`	`real`
`fmod`	`(real x, real y)`	`real`
`hypot`	`(real x, real y)`	`real`
`log`	`(real x)`	`real`
`log10`	`(real x)`	`real`
`log1p`	`(real x)`	`real`
`log2`	`(real x)`	`real`
`pow`	`(real x, real y)`	`real`
`round`	`(real x)`	`real`
`sin`	`(real x)`	`real`
`sinh`	`(real x)`	`real`
`sqrt`	`(real x)`	`real`
`tan`	`(real x)`	`real`
`tanh`	`(real x)`	`real`
`trunc`	`(real x)`	`real`

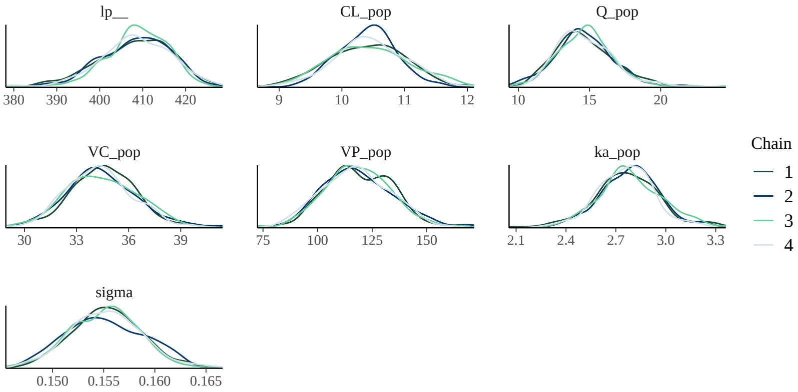


Estimation



lp__
430
420
410
400
390
380
CL_pop
12
11
10
9
Q_pop
25
20
15
10
VC_pop
40
36
32
VP_pop
175
150
125
100
75
ka_pop
3.0
2.5
2.0
sigma
0.165
0.160
0.155
0.150
0.145
0
50
100
150
200
250
Chain
1
2
3
4

Estimation


lp__
380
390
400
410
420
CL_pop
9
10
11
12
Q_pop
10
15
20
VC_pop
30
33
36
39
VP_pop
75
100
125
150
ka_pop
2.1
2.4
2.7
3.0
3.3
sigma
0.150
0.155
0.160
0.165
Chain
1
2
3
4

Misc

- Diagnostics
- Parallel chains
- Within-chain parallellization