Bayesian Optimal Experimental Design of Clinical Studies

Tobias Freidling

DPMMS, University of Cambridge

March 22, 2020

supervised by Thomas Moxon, Unilever

Physiologically based pharmacokinetic (PBPK) model

$$V_1 \frac{\mathrm{d}C_1}{\mathrm{d}t} = -\left[c_1 + V_1(k_{12} + k_{13} + k_{10})\right] C_1 + k_{21} V_2 C_2 + k_{31} V_3 C_3 + I(t)$$

$$V_2 \frac{\mathrm{d}C_2}{\mathrm{d}t} = -\left[c_2 + k_{21} V_2\right] C_2 + k_{12} V_1 C_1$$

$$V_3 \frac{\mathrm{d}C_3}{\mathrm{d}t} = -\left[c_3 + k_{31} V_3\right] C_3 + k_{13} V_1 C_1$$

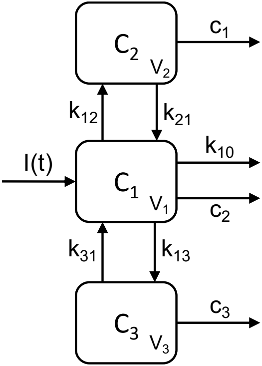


Figure: Schematic of the 3-compartment model for Remifentanil

PBPK model

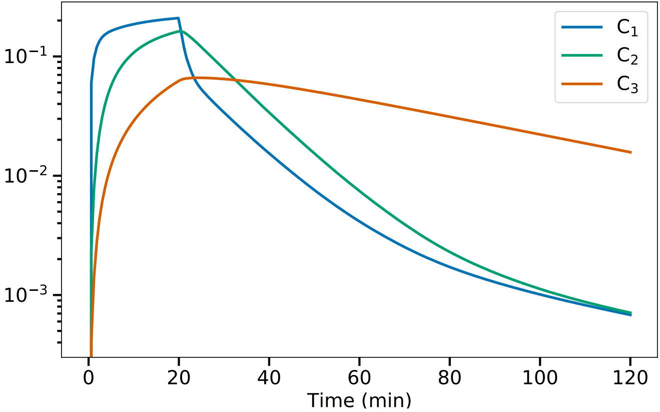


Figure: Concentrations in the 3 compartments

Bayesian 3-Compartment Model

Let $d = (t_1, \ldots, t_p)$, $\bar{\theta} = (k_{10}, k_{12}, \ldots)$ and $C(d, \bar{\theta}) \in \mathbb{R}^{3p}$ the solution of the ODE system evaluated at d.

Bayesian 3-Compartment Model

Let $d = (t_1, \ldots, t_p)$, $\bar{\theta} = (k_{10}, k_{12}, \ldots)$ and $C(d, \bar{\theta}) \in \mathbb{R}^{3p}$ the solution of the ODE system evaluated at d.

Bayesian model:

$$p(y|\theta, d) \sim \mathcal{N}\big(\log C(d, \bar{\theta}), \sigma^2 \mathrm{Id}\big), \qquad p(\sigma^2) \sim \mathrm{Gamma}(2, 31),$$

Gaussian prior for volumes (V-parameters) and flow rates (k-parameters), and log-normal prior for clearance rates (c-parameters). We denote the prior distribution $p(\theta)$, where $\theta = (\bar{\theta}, \sigma^2)$.

Bayesian Optimal Experimental Design

Expected information gain (EIG):

$$\mathrm{EIG}(d) := \mathbb{E}_{p(y|d)}\Big[\mathrm{D}_{\mathrm{KL}} \left(p(\theta|y,d) \,\|\, p(\theta) \right) \Big]$$
$$= \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)}{p(\theta)} \right] = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)}{p(y|d)} \right].$$

Bayesian Optimal Experimental Design

Expected information gain (EIG):

$$\begin{aligned}\mathrm{EIG}(d) &:= \mathbb{E}_{p(y|d)}\Big[\mathrm{D}_{\mathrm{KL}} \left(p(\theta|y,d) \,\|\, p(\theta) \right) \Big] \\ &= \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)}{p(\theta)} \right] = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)}{p(y|d)} \right].\end{aligned}$$

Nested Monte Carlo (NMC) estimator:

$$\hat{\mu}_{\mathrm{NMC}}(d) = \frac{1}{N} \sum_{n=1}^{N} \log \frac{p(y_n|\theta_{n,0},d)}{\frac{1}{M}\sum_{m=1}^{M} p(y_n|\theta_{n,m},d)},$$

with iid samples $\theta_{n,m} \sim p(\theta)$ and $y_n \sim p(y|\theta_{n,0},d)$.
NMC converges to EIG from above but is computationally expensive.

Variational Marginal

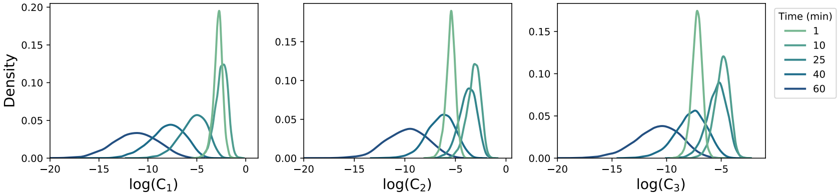
We approximate $p(y|d)$ by $q(y|d, \phi)$ and compute the estimate

$$\hat{\mu}_{\text{marg}}(d) = \frac{1}{N} \sum_{n=1}^{N} \log \frac{p(y_n|\theta_n, d)}{q(y_n|d, \phi)},$$

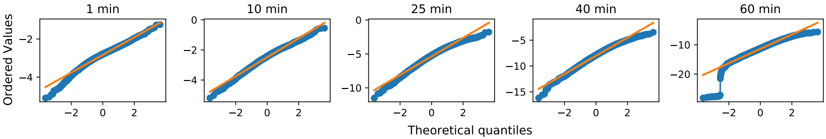
where $y_n, \theta_n \sim p(y, \theta|d)$ is an iid sample.

The variational marginal estimate is asymptotically an upper bound to EIG that is sharp if and only if $q(y|d, \phi) = p(y|d)$.

Marginal distribution



(a) Distribution of the log-concentrations in the three compartments



(b) P-P plot of a Gaussian distribution against $\log(C_1)$

Gaussian and KDE approximation

Gaussian: $q(y|d,\phi) \sim \mathcal{N}(\mu, \Sigma), \phi = (\mu, \Sigma)$

We simulate data points from the model, estimate $\hat{\mu}$ and $\hat{\Sigma}$ and use $q(y|d,\phi) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$.

Gaussian and KDE approximation

Gaussian: $q(y|d,\phi) \sim \mathcal{N}(\mu, \Sigma), \phi = (\mu, \Sigma)$

We simulate data points from the model, estimate $\hat{\mu}$ and $\hat{\Sigma}$ and use $q(y|d,\phi) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$.

Kernel density estimation (KDE):

$$q(y|d,\phi) := \frac{1}{n \det(H)} \sum_{i=1}^{n} K\left(H^{-1}(y - Y_i)\right),$$

where $Y_1, \ldots, Y_n$ are samples from the model. K is a nonnegative function that integrates to 1 and H is the bandwith matrix.

Gaussian and KDE approximation

Gaussian: $q(y|d,\phi) \sim \mathcal{N}(\mu, \Sigma), \phi = (\mu, \Sigma)$

We simulate data points from the model, estimate $\hat{\mu}$ and $\hat{\Sigma}$ and use $q(y|d,\phi) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$.

Kernel density estimation (KDE):

$$q(y|d,\phi) := \frac{1}{n \det(H)} \sum_{i=1}^{n} K\left(H^{-1}(y - Y_i)\right),$$

where $Y_1, \ldots, Y_n$ are samples from the model. K is a nonnegative function that integrates to 1 and H is the bandwith matrix.

We consider factorised (diagonal Σ and H) and non-factorised approximations.

Empirical Evaluation of Estimators

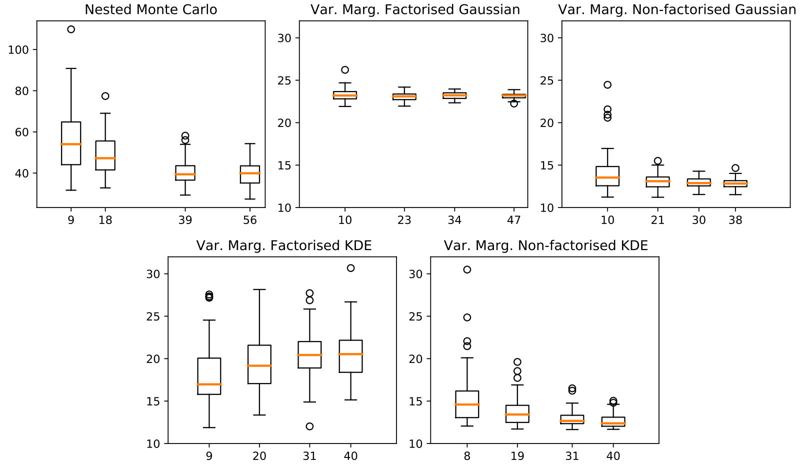


Figure: Box plot of different EIG-estimators for the design $d = (1, 10, 25, 40, 60)$; average runtimes [s] on x-axis

Optimisation

- Constraints on design d: at least 5 minutes between two measurements
- No gradient information $\nabla_d \text{EIG} \rightarrow$ Simulated Annealing or Differential Evolution
- Instable estimates of EIG ☹
- Grid search or list of test designs

EIG	Design	EIG	Design
13.69	(30, 55, 60)	11.93	(5, 45, 60)
13.02	(35, 55, 60)	11.87	(20, 50, 60)
12.75	(35, 45, 60)	11.86	(20, 55, 60)
12.04	(30, 50, 60)	11.85	(40, 55, 60)
11.93	(50, 55, 60)	11.84	(30, 50, 55)

Table: The 10 best designs for three measurements according to grid search; variational marginal estimator with non-factorised Gaussian approximation

Outlook

- Hyper-parameter tuning and sensitivity w.r.t. prior distributions
- Multiple participants with potentially different designs
- New EIG-estimators
- EIG estimation for a subset of parameters
- Optimisation algorithm for instable function evaluations

The end

Thank you!