

Bayesian Pharmacokinetic Models

Demetri Pananos

March 12, 2020

0.1 Derivation of the Differential Equation For Concentration

A dose of size D is orally administered to a patient. The rate of drug absorption from the gut G into the blood plasma C is proportional to the concentration of drug in the gut. The proportionality constant in this case is k_a (the a stands for “absorption”).

The drug is eliminated from the blood plasma at a rate proportional to the concentration of drug currently in the blood plasma. The proportionality constant in this case is k_e (the e stands for “excretion”).

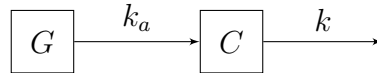


Figure 1: A compartmental diagram for a one compartment pharmacokinetic model. Arrows indicate the direction of flux. Flux is proportional to the concentration in each compartment, with proportionality constant indicated over the arrows.

I can write this system as a differential equation, namely

$$\frac{dG}{dt} = -k_a G(t) \quad G(0) = D/V \quad (1)$$

$$\frac{dC}{dt} = k_a G(t) - k_e C(t) \quad C(0) = 0. \quad (2)$$

Here V is the volume of blood plasma in the body and is assumed to be constant over time. The differential equation for G is uncoupled from the differential equation for C , so I can solve for G and substitute this solution into the differential equation for C . This results in the following differential equation, which describes the concentration of the drug in the blood plasma over time

$$\frac{dC}{dt} = \frac{D}{V} k_a \exp(-k_a t) - k_e C(t) \quad C(0) = 0. \quad (3)$$

Through the relation $V = Cl/k_e$, where Cl is the rate of total body drug clearance, a form more amenable to analysis is obtained. Thus, the dynamics of concentration hereforth are modeled using

$$\frac{dC}{dt} = \frac{D}{Cl} k_e k_a \exp(-k_a t) - k_e C(t) \quad C(0) = 0 \quad (4)$$

which has solution

$$C(t) = \frac{D}{Cl} \frac{k_a k_e}{k_a - k_e} \left(\exp(-k_e t) - \exp(-k_a t) \right) \quad (5)$$

The benefit of this parameterization is that the area under the curve T_∞ is given by D/Cl .

0.2 Non-Dimensionalization Gives Insight For Prior Selection

Non-dimensionalization will become useful as I begin to select priors for the various hyperparameters of our eventual Bayesian model. To begin, I list all the dimensions of the parameters. In the table below M indicates units in mass, V indicates units in volume, and T indicate units in time.

Parameter	Units
D	M
Cl	VT^{-1}
k_a	T^{-1}
k_e	T^{-1}

Let y and τ be non-dimensional concentration and time variables such that

$$C = \frac{Dk_e}{Cl}y \quad (6)$$

$$t = \frac{\tau}{k_a} \quad (7)$$

The non-dimensionalization results in

$$\frac{dy}{d\tau} = \frac{dt}{d\tau} \frac{Cl}{Dk_e} \left(\frac{dC}{dt} \right) \quad (8)$$

$$= \frac{1}{k_a} \frac{Cl}{Dk_e} \left(\frac{D}{Cl} k_e k_a \exp(-k_a t) - k_e C(t) \right) \quad (9)$$

$$= \exp(-\tau) - \alpha y(\tau) \quad (10)$$

The non-dimensionalization constant is $\alpha = k_e/k_a$.

0.3 Bayesian Hierarchical Model With No Covariates

Shown below is a Bayes' net describing the hierarchical structure. In what follows, I further explain and rationalize the model structure.

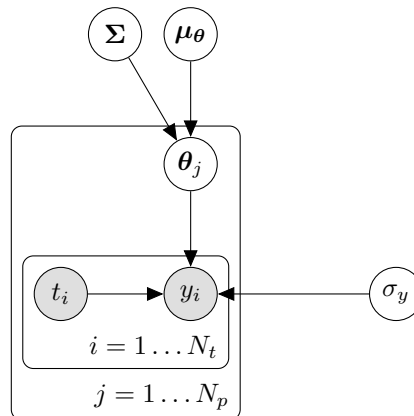


Figure 2: Bayes net for first pharmacokinetic model. Parameters here are shown to be modelled jointly but are assumed independent.

In the Bayes' net above, I have concatenated the parameters Cl , k_e , and k_a into a single random vector θ for economy of thought. The parameters will be independent in the joint distribution. The population level means for these parameters is μ_θ . Since all parameters are taken to be positive, I model these parameters on the log scale and then later exponentiate. The priors for μ_{Cl} and μ_{k_e} are

$$\log(\mu_{Cl}) \sim \mathcal{N}(\cdot, \cdot) \quad (11)$$

$$\log(\mu_{k_e}) \sim \mathcal{N}(\cdot, \cdot) \quad (12)$$

The exception is for $\log(k_a)$. This model is known to display ‘Flip-Flop’ kinetics; a different dynamic regime occurs when the non-dimensionalization parameter crosses 1. The details of absorption for most drugs *should* be well know (for instance, an estimate for time max concentration). Using this information should give insight into if $\alpha < 1$ or $\alpha > 1$. In this case, I assume $\alpha < 1$ implying $k_e < k_a$. The non-dimensionalization thus allows for the following parameterization for $\log(k_a)$

$$\alpha \sim \text{Beta}(\cdot, \cdot) \quad (13)$$

$$\log(k_a) = \log(k_e) - \log(\alpha) > \log(k_e) \quad (14)$$

Since $0 < \alpha < 1$, then $-\log(\alpha) > 0$, and since the logarithm is monotonically increasing, the inequality $k_e < k_a$ is preserved through this parameterization. Furthermore, pharmacokinetic expertise can be leveraged to put an appropriate prior on α .

In the next stage of the hierarchy, the pharmacokinetic parameters are drawn a lognormal density. The conditional density at this stage is coded using a non-central parameterization to help the HMC sampler explore the posterior more easily. To reflect that parameterization, I’ve written the conditional density in the same manner, namely

$$z_j^{Cl} \sim \mathcal{N}(0, 1) \quad (15)$$

$$\sigma_{Cl} \sim \text{Cauchy}(0, 1) \quad (16)$$

$$Cl_j = \exp \left(\log(\mu_{Cl}) + z_j^{Cl} \sigma_{Cl} \right) \quad (17)$$

$$z_j^{k_e} \sim \mathcal{N}(0, 1) \quad (18)$$

$$\sigma_{k_e} \sim \text{Cauchy}(0, 1) \quad (19)$$

$$k_{e,j} = \exp \left(\log(\mu_{k_e}) + z_j^{k_e} \sigma_{k_e} \right) \quad (20)$$

$$z_j^{k_a} \sim \mathcal{N}(0, 1) \quad (21)$$

$$\sigma_{k_a} \sim \text{Cauchy}(0, 1) \quad (22)$$

$$k_{a,j} = \exp \left(\log(\mu_{k_a}) + z_j^{k_a} \sigma_{k_a} \right) \quad (23)$$

Here, I have put half-cauchy priors on the σ . These ought to be changed to something more realistic, but serve as good-enough in scenarios where practitioners are uncertain about noise estimates. Once the pharmacokinetic parameters are drawn, predictions are made by evaluating $C(t)$ with the drawn parameters. The likelihood is then

$$\log(y_{i,j})|t, Cl_j, k_{a,j}, k_{e,j} \sim \mathcal{N}(\log(C(t)), \sigma_y^2) \quad (24)$$

Here, a half Cauchy prior can be placed on σ_y^2 for convenience. However, pharmacokinetic expertise says that the machinery used to detect these drug levels is guaranteed to be within 10% of the true value (according to the manufacturer). This could be used to justify a more precise prior.

0.4 Extensions

The model proposed above can be extended easily to perform regression by specifying a prior over regression parameters

$$\beta_{Cl} \sim p(\beta_{Cl}) \quad (25)$$

$$\log(\mu_{Cl}) = \mathbf{X}\beta_{Cl} \quad (26)$$

The prior in this instance could also serve to regularize coefficient estimates so as to select only predictive variables (think Bayesian lasso). My only worry here is that the effects are likely already very close to 0, which may make selection difficult. Anyway, no way to now know, we could always do yet another simulation.

0.5 Time Delay Effects

Time delays can be modelled as follows. Assume patient j 's first non-zero measurement is observed at τ_j . The concentration at this time is $C_j(\tau_j - \delta_j)$, where δ_j is patient j 's time delay.

Assume each patient has their own time delay which comes from some population delay distribution. Thus, the model structure for the delay is

$$\mu_\delta \sim p(\mu_\delta) \quad (27)$$

$$\kappa_\delta \sim p(\kappa_\delta) \quad (28)$$

$$\delta_j \sim \text{Beta}(\mu_\delta \kappa_\delta, (1 - \mu_\delta) \kappa_\delta) \quad (29)$$

$$t_{j,actual} = t_{j,obs} - \tau_j \delta_j \quad (30)$$

Here, $t_{j,actual}$ is the actual time after absorption of the drug for the j^{th} patient, and p is an appropriate prior for μ_δ and κ_δ .

0.6 Model Evaluation

The goal of these models is to predict concentrations for out of sample patients. I could compare RMSE between models, or I could evaluate the some information criterion. The problem with these approaches is that they utilize *observations* rather than true concentrations to evaluate how well the model predicts on unseen data. After all, observations are noisy, and we aren't so interested in predicting observed concentrations so much as we are interested in predicting actual concentrations.

Sampling pharmacokinetic parameters from the posterior allows us to in essence sample functions since $\hat{C}(t)$ is determined by the parameters. In a simulation study, we can compare the $\hat{C}(t)$ to $C(t)$ directly by computing the ℓ_2 norm

$$\text{Error}^2 = \langle C(t), \hat{C}(t) \rangle = \int_0^\infty |C(t) - \hat{C}(t)|^2 dt \quad (31)$$

Since $C(t)$ is comprised of elementary functions, the evaluation of this integral can be done exactly using computer algebra, and made a function of the parameters for $C(t)$ and $\hat{C}(t)$.

Other pharmacokinetic phenomena can be evaluated similarly. The time to max concentration is

$$t_{max} = \frac{\log(k_a) - \log(k_e)}{k_a - k_e} \quad (32)$$

and so bias and RMSE for this quantity, as well as the max achieved concentration, $C(t_{max})$, and area under the curve, T_∞ , can be evaluated easily.

0.7 What Have Other People Done? Why Do This?

0.7.1 Calling the Kettle Black; I Am The Dogmatic One

I have believed from the start that “Bayes will save us” because

- i) Small samples per patient
- ii) Prior knowledge
- iii) Conditioning on new information

I think it is worth examining these presuppositions in a simulation study. How well do our models perform when we have small samples (as small and as sparse as the personalized medicine clinic has)? How well do weakly informative priors do in terms of out of sample prediction error, but also in terms of bias and RMSE for parameter estimation? When is it worth while to have informative priors? Is it better to collect a bunch of possibly informative covariates, or are we better just observing patients for a little while and letting our model learn from that?

0.7.2 NONMEM Is Dominant

Nearly all the pharmacokinetic papers I’ve read have used NONMEM to model pharmacokinetics. That is well and great, but if free open source tools (like Stan, like PyMC3) are not being used, I think it is important to ask “why not?”. Worst case scenario, these tools do worse than the NONMEM, but that in and of itself would be a useful result not only for the PK community but also for open source. Best case scenario the open source tools do better, which I’m sure the creators of NONMEM would love to know. Most probably case is that they do equally well, in which case everyone wins.

0.7.3 Principled Evaluation of Models

A lot of the PK models I’ve seen often use questionable methods for evaluation (e.g. multiple hypothesis tests looking for $p < 0.05$). It would be nice to introduce a principled way to evaluate these sorts of models for fit. I’m not quite sure how I would do that just yet.