PK/PD Inference with Stan

Arya Pourzanjani, Daniel Lee, Eric Novik

Generable, New York, NY, USA

arya,daniel,eric(@generable.com)

1. Overview: Why Stan?

- Support for and ODE model including any PK/PD model
- Population level random effects are more intuitive to implement and interpret than traditional techniques
- More accurate uncertainty estimates leads to safer inference
- Prior information e.g. from previous studies or meta-analysis can be readily included in any model
- More intuitive diagnostics means less statistics expertise required
- Open source with a large, helpful community for answering questions

2. Setting up a model is as easy as entering an ODE, then specifying an error distribution and a population distribution

2.1 Specify an ODE

```
real[] dz_dt(real t,          // time
             real[] z,        // system state
             real[] theta) { // parameters

  // set states
  real A = z[1]; // mass in 1st compartment
  real c = z[2]; // concentration in 2nd compartment

  // set unknown parameters
  real Ka = theta[1]; // absorption rate
  real Cl = theta[2]; // clearance rate of drug
  real V = theta[3];  // volume of blood

  // specify differential equations
  real dA_dt = -Ka*A;
  real dc_dt = (1/V)*(Ka*A - (Cl/V)*(c*V));

  return { dA_dt, dc_dt };
}
```

2.2 Specify the Data You Have

```
data {
  int Nt;       // number of unique measurement times
  real ts[Nt]; // measurement times
  real y_init; // initial drug dose
  real<lower=0> y[Nt]; // concentrations
}
```

2.3 Specify the Parameters to Estimate

```
parameters {
  real<lower = 0> theta[3]; // Ka,Cl,V
  real<lower = 0> sigma;    // measurement error
}
```

2.4 Specify Your Probabilistic Model

```
model {

  // prior information or population level prior
  theta[3] ~ normal(5.1, 0.2); // V close to 5 liters

  // measurement model
  z[Nt,] = integrate_ode_rk45(dz_dt, ts, theta);
  y ~ lognormal(log(z[, 2]), sigma);

}
```

2.5 Simulate Synthetic Data from Your Model

```
generated quantities {
  real y_rep[Nt];
  for (n in 1:Nt) {
    // generate data w/ the same noise distribution
    y_rep[n] = lognormal_rng(log(z[n, 2]), sigma);
  }
}
```

2.6 Extending to a Multi-Patient Population-Level Model

```
model {
  // set up model for each subject
  for (n in 1:NumSubjects) {

    // nth patient follows population level prior
    theta[n,3] ~ normal(mu_V, sigma_v);

    // integrate ode for nth patient
    z[n,Nt,] = integrate_ode_rk45(dz_dt, ts, theta[n]);
    y[n] ~ lognormal(log(z[n,, 2]), sigma);
  }
}
```

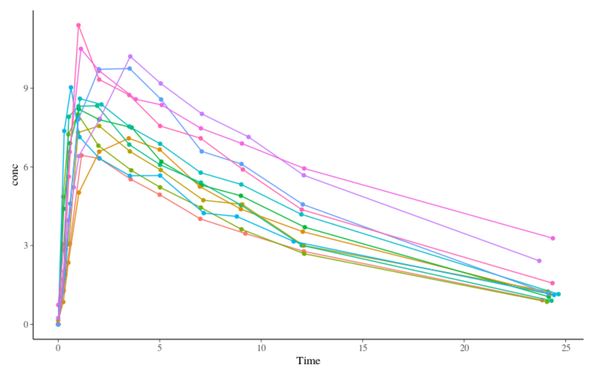

Figure 1: *Concentration of the drug Theophylline measured over time for 12 distinct subjects. Each color represents a unique subject.*

3. Fitting models is done automatically with the state-of-the-art NUTS sampler

Data is noisy, which means estimates are noisy! Stan not only provides uncertainty intervals, but does so more accurately than traditional methods such as asymptotic confidence intervals and Metropolis sampling.

3.1 Efficiently Exploring the Parameter Space

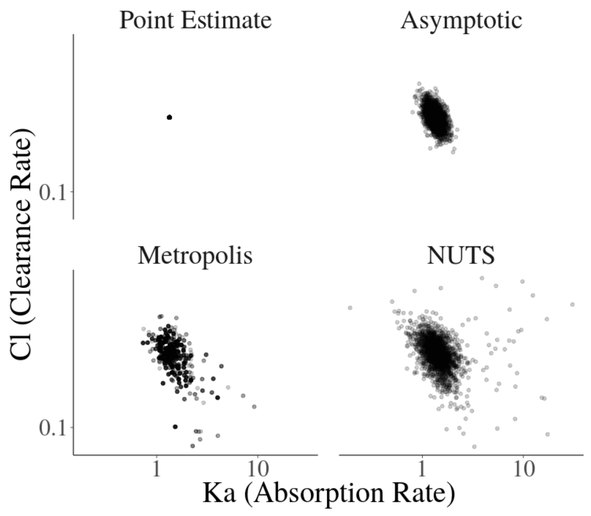

Figure 2: *Unlike asymptotic confidence estimates which assume that the likelihood is Gaussian, NUTS sampling makes no assumptions about the shape of the likelihood function and is able to explore the entirety of parameter space rather than being confined to the best-fitting Gaussian. Furthermore, NUTS is much more efficient than its predecessor Metropolis in that it provides nearly independent samples, reducing inference time by orders of magnitude, especially for large problems.*

3.2 Getting the Right Uncertainty Intervals

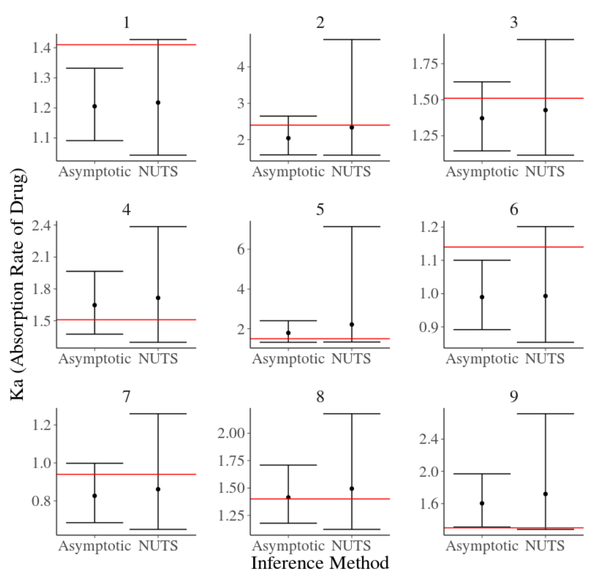

Figure 3: *On simulated data where we know the true value of parameters (red line), it is easy to see how commonly used asymptotic estimates provided in traditional inference software often underestimate the uncertainty inherent in parameter estimations. NUTS better represents the uncertainty inherent in estimates, giving us more confidence in applications where safety and worst-case scenarios are important to quantify*

4. Posterior samples make interpreting fits intuitive

Using posterior samples we can intuitively understand the estimated value of any unknown parameter we are trying to estimate. As an example, if we wanted to know the probability a patient's peak concentration level for a given dose will reach a value greater than 11 g/L, we can simply check the fraction of samples that were over 11, and this value represents a probability.

```
mean(peak.conc.post.samples > 11)
> [1] 0.026
```

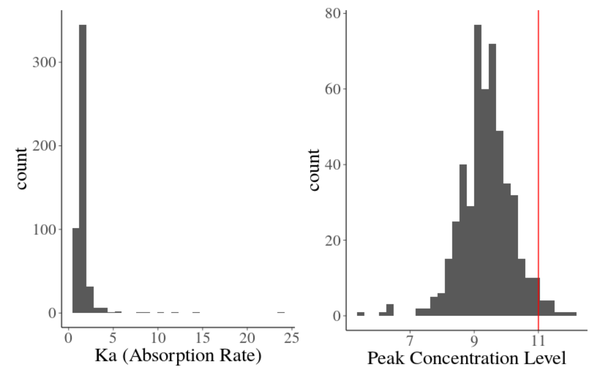

Figure 4: *Posterior distributions for drug absorption rate and peak concentration level estimated from Theophylline data for a single patient. These histograms not only give us an idea of what the value of any unknown parameters can be, but they illustrate the uncertainty range for unknown parameters as well.*

5. Generative Models allow us to easily simulate data, making diagnostics intuitive

The generative modeling framework and Stan allow us to simulate synthetic data from our model making model diagnostics intuitive. Furthermore, by the ability to describe our data generating process allows us to simulate our model under different conditions. For example, we can simulate a subject's concentration trajectory under different dosages in a straight-forward manner. Stan also makes it easy to simulate concentration profiles for patients who have no observed data or very little data, by taking advantage of the population model and the generative framework.

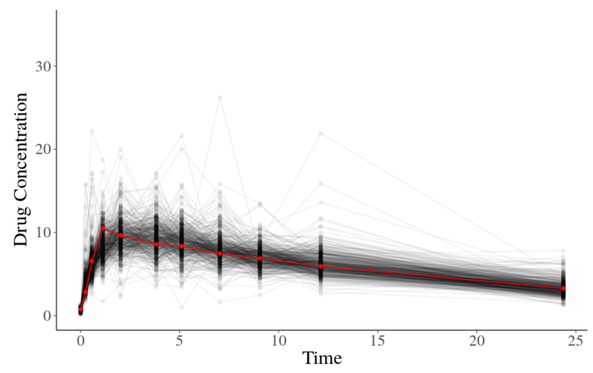

Figure 5: *When simulations of our data seem to match the true data (red line) this indicates that the model captures the subtleties of the data and is thus a good fit.*

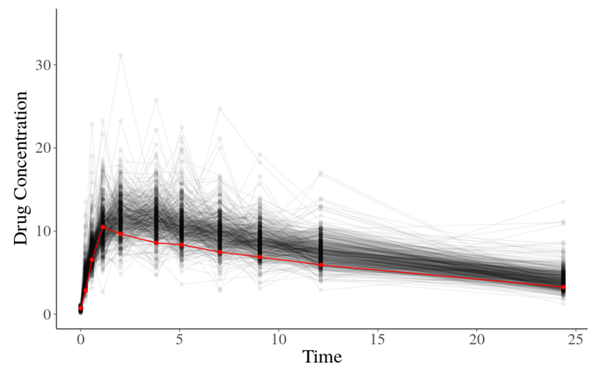

Figure 6: *With Stan, it is straight-forward to take estimated parameters for a patient and use those estimates to generate concentration profiles for that same patient under a different dosage, all while taking in to account the uncertainty in our estimates. In this case we simulate concentration levels when the patients takes a dosage of 400mg as oppose to the original dose of 320 that the experiment was based off of. Compared to the observed concentrations (red line), the higher dosage results in much higher concentrations.*