Simple Application of Approximate Bayesian Computation (ABC) in Modelling Tumor Growth

Chun-Hei (Sammuel) Lam

Submitted in part fulfilment of the requirements
for the first year research project M1R at
Imperial College London

2019 Summer

Contents

Scenario

A research (Zanfardino et al, 2013) reported that *Simvastatin* (a class of drug) is able to *inhibit* the growth of Melanoma Cells.

They have observed a

- Delay in tumor development
- Back to Scenario

To justify the result they have carried out an *in vivo* experiment:

1. 42 mice were injected with melanoma cells
2. 22 mice were treated with *Simvastatin* (in 'experimental condition') while 20 mice weren't treated (in 'control condition').
3. The tumor volume measures were made every 3 days starting after 10 days from cell injection.
4. The experiment was conducted for 20 days before the tumor became too large inducing animal sufferings.

Day after injection (t/days)	Volume of Tumor (Control) ($V_1/(mm^3)$)	Volume of Tumor (Experimental) ($V_2/(mm^3)$)
11	176 ± 72	81 ± 42
13	618 ± 231	218 ± 98
16	1176 ± 305	400 ± 206
19	1893 ± 398	682 ± 267
22	2293 ± 373	853 ± 292
25	2485 ± 311	950 ± 299
28	2780 ± 43	1613 ± 324

Table: Average Volume of Tumor with Standard Error

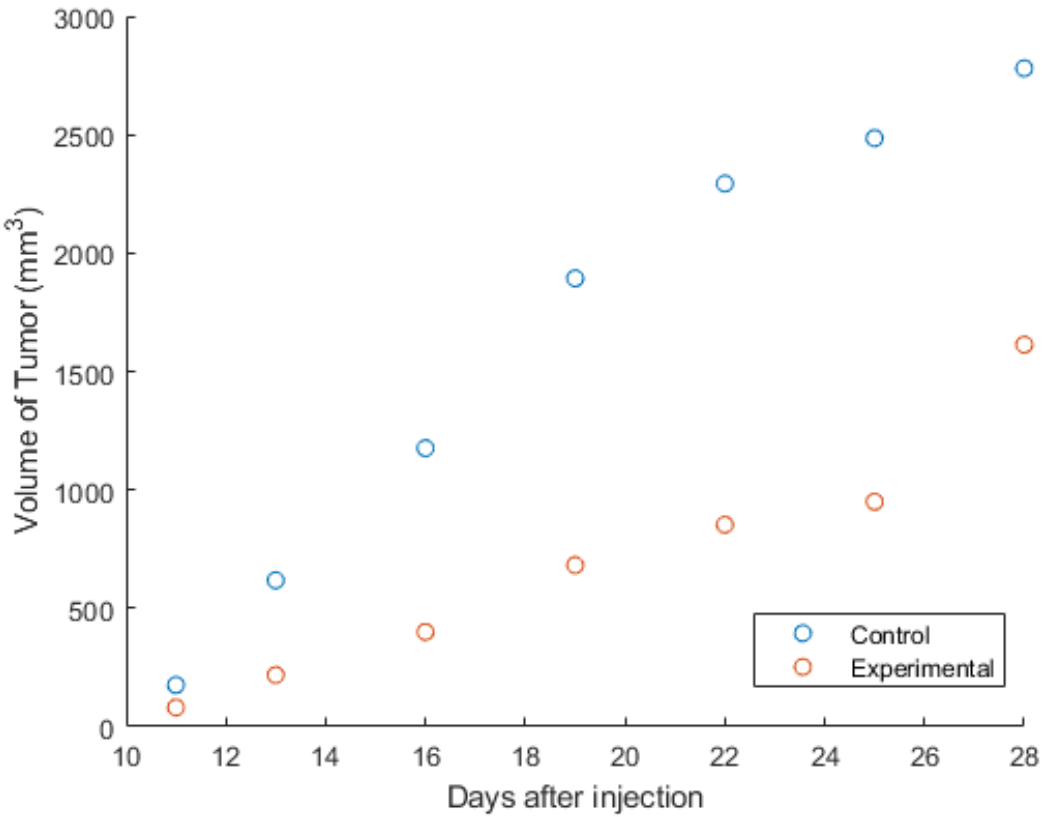


Figure: Raw Data

t-Test

Let:
$\vec{Y}^{(1)}$ (sample mean $\bar{Y}^{(1)} = V_1$, sample SD s_1, size $n_1 = 20$) and
$\vec{Y}^{(2)}$ (sample mean $\bar{Y}^{(2)} = V_2$, sample SD s_2, size $n_2 = 22$).
be the samples of volume of tumor in control and experimental condition taken on day t respectively.

In high school we have learnt how to perform Hypothesis Test on population means of two sets of data μ_1, μ_2.

We start by assuming the linear model $E(\vec{Y}_i^{(1)}) = \mu_1$ and $E(\vec{Y}_i^{(2)}) = \mu_2$ and the data are normally distributed.

Moreover, we have

$$\begin{aligned}\text{RSS} &= \sum_{i=1}^{n_1}((\vec{Y}^{(1)})_i)_i - \bar{Y}^{(1)}))^2 + \sum_{i=1}^{n_2}((\vec{Y}^{(2)})_i - \bar{Y}^{(2)})^2 \\ &= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\end{aligned}$$

We also have

$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix}$$

Let $n = n_1 + n_2$. A corollary of Fischer Cochran Theorem states that if $\vec{c} \in \mathbb{R}^2$, then

$$\frac{\vec{c}^T\hat{\beta} - \vec{c}^T\vec{\beta}}{\sqrt{\vec{c}^T(X^TX)^{-1}\vec{c}\frac{\text{RSS}}{n-2}}} \sim t_{n-2}$$

Substituting $\vec{c} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ one have

$$\frac{(\bar{Y}^{(1)} - \bar{Y}^{(2)}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$$

We may thus perform one-sided hypothesis test of $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$ respectively. We set $\alpha = 0.05$ and see if the p-values under H_0 are smaller than α.

Day after injection (t/days)	p-value	accept/reject
11	0.1967	reject
13	0.0999	reject
16	0.0405	accept
19	0.0161	accept
22	0.0047	accept
25	0.0012	accept
28	0.0019	accept

Table: Results of t-test

From t-test we have seen that after some days from injection, the size of tumor in control condition is smaller than that in experimental condition. However, there are three possible reasons for this:

1. Delay in tumor development.
2. Reduction of tumor volume (as $t \to \infty$)
3. Both (1) and (2)

Therefore we need a way to quantify the 'speed' of tumor development and 'volume of tumor in long term'.

Gompertz Growth

Biologists have developed a model for Tumor Growth in general (Gompertz, 1824, Retsky, 1990). It is commonly written (Dennis, 2014) as

$$\frac{dV^{(i)}(t)}{dt} = \theta^{(i)} V^{(i)}(t)(\ln V_{\infty}^{(i)} - \ln V^{(i)}(t))$$

where $i = 1$ indicates control condition and $i = 2$ indicates experimental condition, $V_{\infty}^{(i)}, \theta^{(i)} > 0$.

Immediately we see that there is an equilibrium at $V^{(i)} = 0$ (unstable) and $V^{(i)} = V_\infty^{(i)}$ (stable).

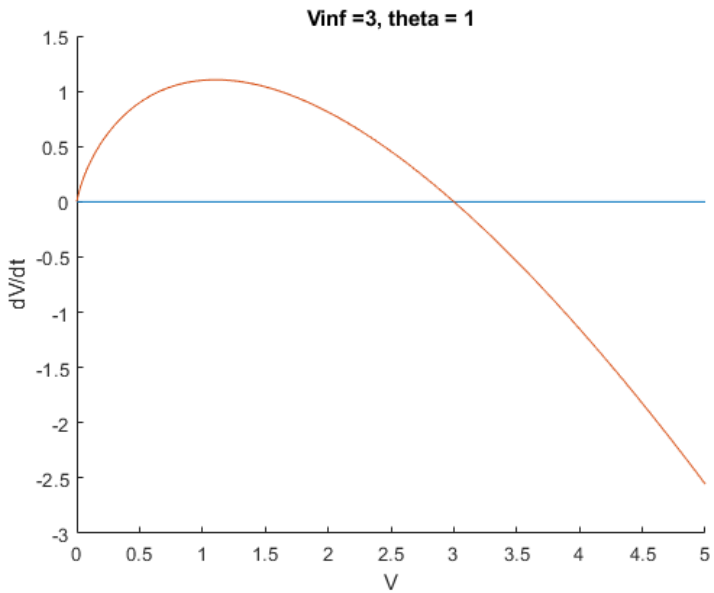


Figure: $\frac{dV}{dt}$ against V, $V_\infty^{(i)} = 3$, $\theta^{(i)} = 1$

Here $\theta^{(i)}$ determines the speed of growth...

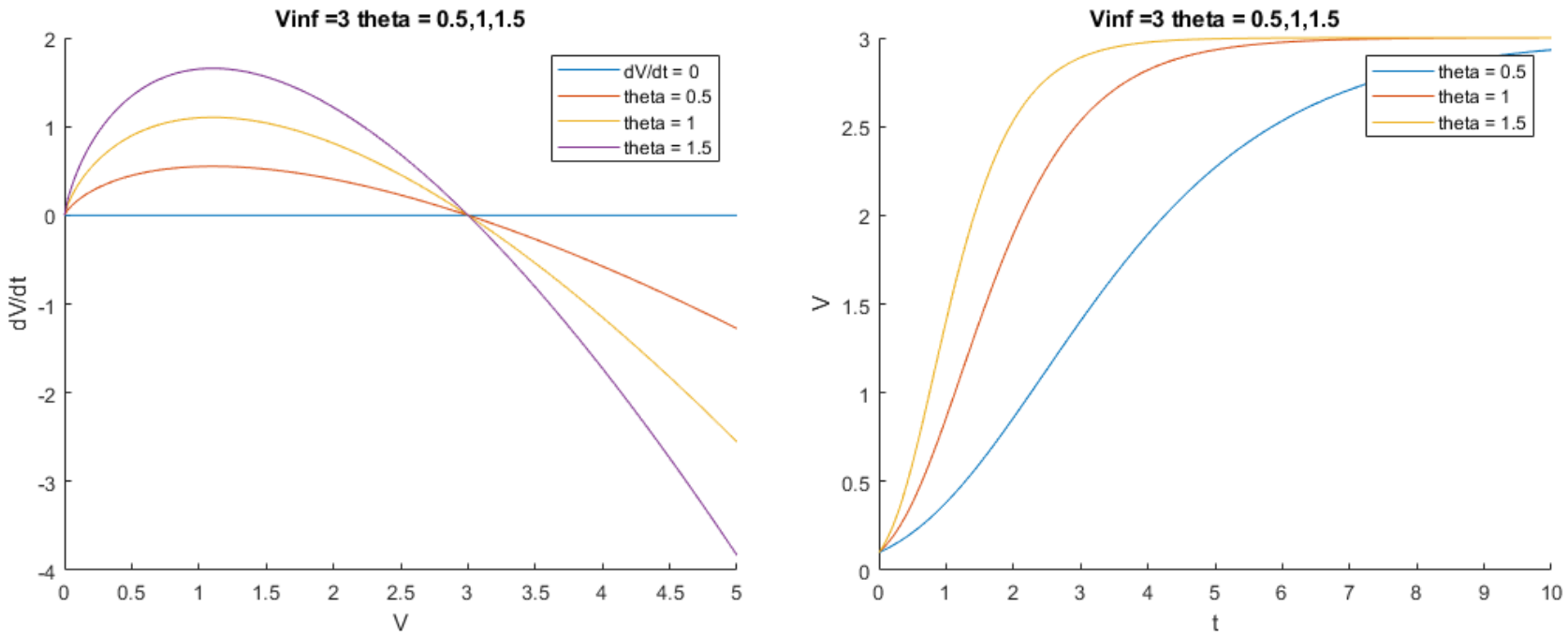


Figure: $V_\infty^{(i)} = 3$, $\theta^{(i)} = 0.5, 1, 1.5$

... while $V_\infty^{(i)}$ determines the volume of tumor in long term.

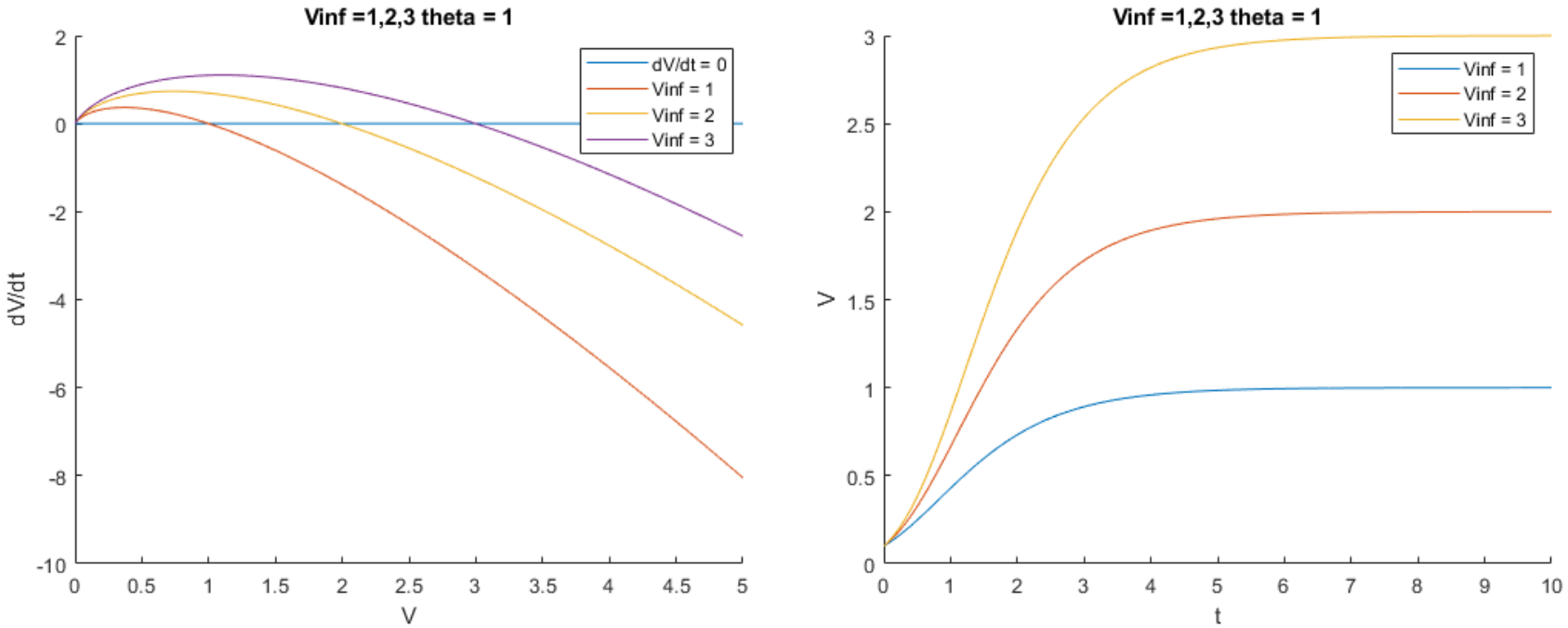


Figure: $V_\infty^{(i)} = 1, 2, 3$, $\theta^{(i)} = 1$

Solution to Gompertz Growth

$$\frac{dV^{(i)}(t)}{dt} = \theta^{(i)} V^{(i)}(t)(\ln V_\infty^{(i)} - \ln V^{(i)}(t))$$

$$\Rightarrow \frac{1}{V^{(i)}(t)} \frac{1}{\ln V_\infty^{(i)} - \ln V^{(i)}(t)} \frac{dV^{(i)}(t)}{dt} = \theta^{(i)}$$

$$\Rightarrow \int \frac{d(\ln V_\infty^{(i)} - \ln V^{(i)}(t))}{\ln V_\infty^{(i)} - \ln V^{(i)}(t)} = \int -\theta^{(i)} dt$$

$$\Rightarrow \ln(\ln V_\infty^{(i)} - \ln V^{(i)}(t)) = -\theta^{(i)} t + C$$

$$\Rightarrow \ln V_\infty^{(i)} - \ln V^{(i)}(t) = Ae^{-\theta^{(i)} t}$$

Writing $V^{(i)}(0) = V_0^{(i)}$ we have $A = \ln V_\infty^{(i)} - \ln V_0^{(i)}$ and

$$\ln V_\infty^{(i)} - \ln V^{(i)}(t) = \left(\ln V_\infty^{(i)} - \ln V_0^{(i)}\right) e^{-\theta^{(i)} t}$$

$$\ln V^{(i)}(t) = \ln V_\infty^{(i)} + \left(\ln V_0^{(i)} - \ln V_\infty^{(i)}\right) e^{-\theta^{(i)} t}$$

$$V^{(i)}(t) = V_\infty^{(i)} e^{\left(\ln V_0^{(i)} - \ln V_\infty^{(i)}\right) e^{-\theta^{(i)} t}}$$

Good news: We have an analytical form of ODE!

Bad news: We can't use simple way to estimate the parameters $V_0^{(i)}$, $V_\infty^{(i)}$ and $\theta^{(i)}$. It is not just a matter of least square, we need a rather sophisticated way!

Frequentist and Bayesian Approach

Ordinary (Frequentists) Approach:

- $V_0^{(i)}$, $V_\infty^{(i)}$ and $\theta^{(i)}$ are **unknown constants** that need to be estimated.
- *Problem (for me):* It is very hard to construct distribution on a pivot involving the unknown constants and their estimators (it is not linear regression anymore!)

Our (Bayesian) Approach:

- $V_0^{(i)}$, $V_\infty^{(i)}$ and $\theta^{(i)}$ are **variables**. The data are **constants.**
- It is easier to construct predictive intervals.
- *Problem (for me): Later*

Bayesian Approach

Given some prior knowledge on variables (usually *prior distribution* $(V_0^{(i)}, V_\infty^{(i)}, \theta^{(i)}) := \vec{\Theta}^{(i)}$ with pdf $f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)})$.), and (fixed) observations $\mathcal{D}$.

Aim: Wish to find the *posterior distribution* $\vec{\Theta}^{(i)}|\mathcal{D}$ with pdf $f_{\vec{\Theta}^{(i)}|\mathcal{D}}(\vec{\theta}^{(i)}|\mathcal{D})$.

Assumptions

- All individual observations are **independent** - in such case we can perform ABC separately on $\vec{V}^{(1)}$ and $\vec{V}^{(2)}$.
- The mean of measurements now become an actual observation with *negligible* error - this *oversimplifies* our discussion.

Bayes Theorem

$$f_{\vec{\Theta}^{(i)}|\mathcal{D}}(\vec{\theta}^{(i)}|\mathcal{D}) = \frac{f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)})}{f(\mathcal{D})}$$
$$= \frac{f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)})}{\int_R f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)}) d\vec{\theta}^{(i)}}$$

Here $f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)})$ is called the *likelihood*, which is from the model itself. (Calderhead, 2010)

Ingredient: Prior Distribution

- We know that $V_0^{(i)}$ must be greater than 0 and smaller than the observation 11 days after injection, so let's assume its (marginal) be $V_0^{(i)} \sim U[0, V^{(i)}(11)]$.
- We know that $V_\infty^{(i)}$ should be greater than the last observation taken $V_{\text{last}}^{(i)}$ (which is also maximum among data taken). A guess of prior is $V_\infty^{(i)} \sim U[V_{\text{last}}^{(i)}, 3V_{\text{last}}^{(i)}]$

Ingredient: Prior Distribution

- We only know that $\theta^{(i)} > 0$. Therefore we wish the prior to be heavy tailed - a possible choice is the 'modded Cauchy' with pdf

$$f_{\Theta^{(i)}}(\theta^{(i)}) = \begin{cases} \frac{2}{\pi(1+(\theta^{(i)})^2)} & \theta^{(i)} > 0 \\ 0 & \theta^{(i)} \leq 0 \end{cases}$$

(**Remark:** We may also use some 'improper prior', something that I don't want to even mention.)

Ingredient: Prior Distribution

- Assume three variables are independent, so that joint pdf of prior is

$$f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)}) = \frac{1}{\pi V^{(i)}(11) V^{(i)}_{\text{last}}(1 + (\theta^{(i)})^2)}$$

with joint support $[0, V^{(i)}(11)] \times [V^{(i)}_{\text{last}}, 3V^{(i)}_{\text{last}}] \times (0, \infty)$.

Ingredient: Likelihood

We assume an individual datum $(t_j, V_j^{(i)})$ are simulated from

Normal Distribution with mean $\hat{V}^{(i)}(t_j) = V_\infty^{(i)} e^{\left(\ln V_0^{(i)} - \ln V_\infty^{(i)}\right) e^{-\theta^{(i)} t_j}}$ as obtained from solution to Gompertz Model and standard deviation $\sigma^{(i)}$.

Hence the likelihood is

$$f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) = \prod_{j=1}^{n} \left(\frac{1}{\sigma^{(i)}\sqrt{2\pi}} e^{-\frac{(V_j - \hat{V}^{(i)}(t_j))^2}{2(\sigma^{(i)})^2}} \right)$$

Ingredient: Likelihood

How do we determine $\sigma^{(i)}$?

We may include $\sigma^{(i)}$ in our prior, but that will make our computation a lot complicated. Instead we just set $\sigma^{(i)}$ = mean of absolute uncertainties in different conditions.

Warning: I may have generated data which make no sense (when $V < 0$.) However these data will most likely be excluded from our simulations, and $\sigma^{(i)}$ is *nuisance* (usually does not affect our analysis).

Main Problem of Bayesian Analysis

$$f_{\vec{\Theta}^{(i)}|\mathcal{D}}(\vec{\theta}^{(i)}|\mathcal{D}) = \frac{f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)})}{\int_R f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)}) d\vec{\theta}^{(i)}}$$

Problem (for me): It is very hard to evaluate the integral at the bottom.

With the help of computing power, we may actually simulate samples from the posterior distribution.

Rejection-Based ABC

Rubin (1984) suggests the following algorithm:

1. Obtain a sample $\vec{\theta}_k^{(i)}$ from $\vec{\Theta}^{(i)}$.
2. Obtain a sample $\tilde{D}$ from $\mathcal{D}|\vec{\Theta}^{(i)}$
3. If $\tilde{D}$ and actual data $\mathcal{D}$ are 'close enough', then we include $\vec{\theta}_k^{(i)}$ in our sample of $\vec{\Theta}^{(i)}|\mathcal{D}$.
4. Repeat (1)-(3) until the sample of $\vec{\Theta}^{(i)}|\mathcal{D}$ is large enough.

Justification of Rejection-Based ABC

- Step 1 is a simulation of $f_{\vec{\Theta}^{(i)}}(\vec{\theta}^{(i)})$.
- Step 2/3: the probability of accepting the sample $\vec{\theta}_k^{(i)}$ is proportional to the likelihood $\mathcal{D}|\vec{\Theta}^{(i)}$

Overall, the algorithm simulate the posterior distribution up to a normalizing constant.

Step 1 - Obtaining samples from prior

Since I have assumed that all parameters are independent, we can sample $V_0^{(i)}$, $V_\infty^{(i)}$, $\theta^{(i)}$ one by one.

- The first two parameters can be simulated by obtaining sample from $u \sim U[0,1]$, and apply the transformation $x = a + (b-a)u$, where a and b are minimum and maximum elements of support respectively.
- The third parameter can be simulated by obtaining sample from $u \sim U[0,1]$, and apply the transformation $\theta^{(i)} = |\tan(\pi(u-0.5))|$.

The justification is based on technique of transformation of continuous random variable taught in M1S. For instance, let $U \sim U[0,1]$ and $\Theta^{(i)} = |\tan(\pi(U-0.5))|$, then

$$\begin{aligned}
F_{\Theta^{(i)}}(\Theta^{(i)} \leq \theta^{(i)}) &= F_{\Theta^{(i)}}(|\tan(\pi(U-0.5))| \leq \theta^{(i)}) \\
&= 2F_{\Theta^{(i)}}(0 \leq \pi(U-0.5) \leq \tan^{-1}\theta^{(i)}) \\
&= 2(F_{\Theta^{(i)}}(\pi(U-0.5) \leq \tan^{-1}\theta^{(i)}) - 1) \\
&= 2F_{\Theta^{(i)}}\left(U \leq \frac{1}{\pi}\tan^{-1}\theta^{(i)} + 0.5\right) - 1 \\
&= \frac{2}{\pi}\tan^{-1}\theta^{(i)} \\
f_{\Theta^{(i)}}(\theta^{(i)}) &= \frac{2}{\pi(1+(\theta^{(i)})^2)} \qquad \theta^{(i)} > 0
\end{aligned}$$

Step 2 - Obtaining samples from likelihood

This could be achieved by obtaining samples from standard normal distribution and perform a standard scale-location transformation.

Step 3 - Comparison of Simulated/Actual Data

Recall that the data simulated are $\tilde{D} = \{\vec{t}, \tilde{V}^{(i)}\}$, where $\vec{t}$ is from actual data. We are only required to compare $\tilde{V}^{(i)}$ with $\vec{V}^{(i)}$ from actual data.

We may consider the data 'closed enough' if $\| \tilde{V}^{(i)} - \vec{V}^{(i)} \| < \epsilon$, where $\|.\|$ is the Euclidean Norm and ϵ is the tolerance to be chosen.

Actual Code in Matlab

```
function posterior = ABCFunctions(t,V,deltaV,M)
    % t is the input, V is the response,
    % M is the number of samples from prior
     k=1;
     posterior = zeros(3,M); %Pre=allocation
     while k <= M
        % Step 1 - Sample from Prior
        thetak = [V(1)*rand, V(end)+2*V(end)*rand, abs(tan(pi*(rand-0.5)))];

        % Step 2 - Sample from likelihood
        MeanVhat = Gompertz(thetak(1), thetak(2), thetak(3), t);
        SDVhat = mean(deltaV);
        Vsim = randn(1,7).*SDVhat + MeanVhat;

        % Step 3 - Accept/Reject
        if norm(V-Vsim) < 500
            posterior(:,k) = thetak;
            k = k+1;
        end
    end
end
```

Convergence of ABC (Webster, 2016)

Let's say we have performed ABC with sample size n and obtain a sample $\{\vec{\theta}_1^{(i)}, ..., \vec{\theta}_n^{(i)}\}$ from $\vec{\Theta}^{(i)}|\mathcal{D}$. Let $Y_n = \frac{1}{n}\sum_{k=1}^{n} h(\vec{\theta}_k^{(i)})$, where $h(.)$ is a vector-to-scalar function such that $E_{\Theta^{(i)}}(h(\vec{\theta}^{(i)})) < \infty$. Then

- By strong law of large numbers, the sequence of variables $Y_1, Y_2, ...$ **converges almost surely** to $E(Y_n)$. In other words, we have

$$P\left(\lim_{n\to\infty} Y_n \to E_{Y_n}(Y_n)\right) = 1$$

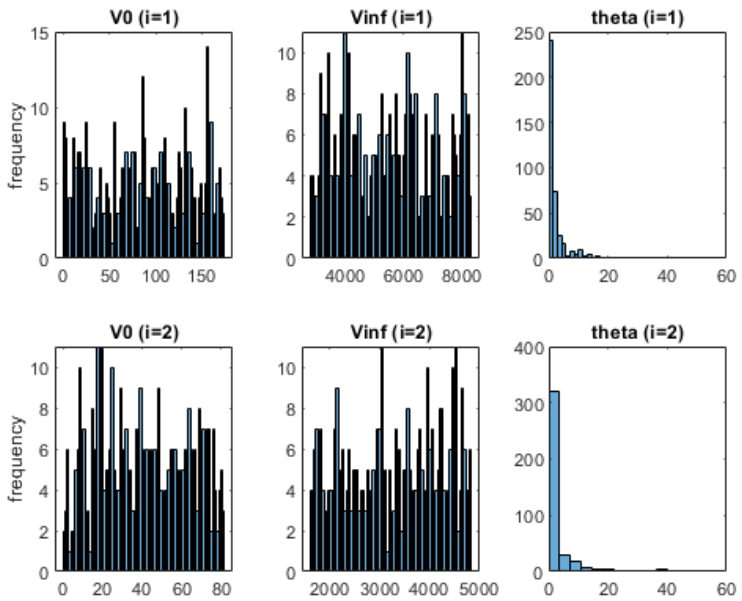
- $\lim_{\epsilon\to 0} E_{Y_n}(Y_n) = E_{\vec{\Theta}^{(i)}|\mathcal{D}}(h(\vec{\Theta}^{(i)})|\mathcal{D})$ for almost all $\mathcal{D}$.

Convergence of ABC (Webster, 2016)

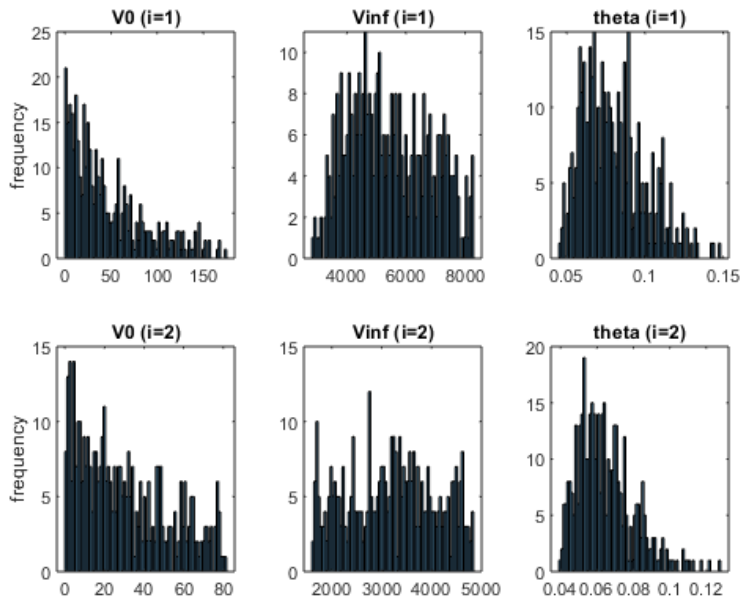
The theorem guarantees that we are certain that Y_n will become $E(Y_n)$ and will stay there forever. It also guarantees that when δ is small than the sample mean obtained from sample is approximately equal to the expectation for almost all data.

Bad News: In fact $E_{\Theta(i)}(h(\vec{\theta}^{(i)}))$ is unbounded if $h \equiv 1$. Indeed we need to change our prior to ensure convergence, but let's do that later...

Actual Simulation



(a) Samples from Prior



(b) Samples from Posterior

Analysis of Posterior Distribution

- The posterior are data-driven (we can see a change of shape of distribution)
- $V_0^{(i)}$ is really small (the posterior seemed to be positively skewed.)
- Further action: We should limit our range of prior (??)

Problems of Rejection-Based ABC

It is computationally inefficient. There are a lot of rejected points. Moreover, it is very hard to repeat Bayesian Analysis.

Is there a wiser way to choose the points?

Metropolitan Hasting Algorithm

Metropolitan et al. (1953) and Hasting (1970) has suggested a way to obtain sample from a general probability distribution $\pi(\vec{\theta}^{(i)})$: Initialize a starting point $\vec{\theta}_1^{(i)}$ and repeat the following

1. Given $\vec{\theta}_n^{(i)}$. Propose a new point $\vec{\theta^*}^{(i)}$ by sampling from *proposal* distribution $T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$.
2. Set $\vec{\theta}_{n+1}^{(i)} = \vec{\theta^*}^{(i)}$ with probability $A := A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$, and $\vec{\theta}_{n+1}^{(i)} = \vec{\theta}_n^{(i)}$ with probability $1 - A$, where

$$A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)}) := \min\left(\frac{\pi(\vec{\theta^*}^{(i)})T(\vec{\theta}_n^{(i)}|\vec{\theta^*}^{(i)})}{\pi(\vec{\theta}_n^{(i)})T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})}, 1\right)$$

Justification of Metropolitan Hasting Algorithm

- If $\pi(\vec{\theta^*}^{(i)}) \geq \pi(\vec{\theta}_n^{(i)})$ then an update will probably happen. This allows the frequency of certain points included in posterior sample proportional to its posterior density.
- Even if $\pi(\vec{\theta^*}^{(i)}) < \pi(\vec{\theta}_n^{(i)})$ there may be an update. This is to introduce noises on the sample.
- The correction factor $\frac{T(\vec{\theta}_n^{(i)}|\vec{\theta^*}^{(i)})}{T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})}$ is included so that the T can now be asymmetrical ($T(\vec{\theta}_n^{(i)}|\vec{\theta^*}^{(i)}) \neq T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$) while not violating the **balance condition**.

Balance Condition

Let $p(\vec{\theta}^{(i)})$ be the (average) probability of moving to point $\vec{\theta}^{(i)}$. The Metropolitan Hasting Algorithm satisfies the **detailed balance condition**

$$p(\vec{\theta}^{(i)})p(\vec{\theta^*}^{(i)}|\vec{\theta}^{(i)}) = p(\vec{\theta^*}^{(i)})p(\vec{\theta}^{(i)}|\vec{\theta^*}^{(i)})$$

by noting that $p(\vec{\theta^*}^{(i)}|\vec{\theta}^{(i)}) = T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$. This implies the **balance condition**, that

$$p(\vec{\theta^*}^{(i)}) = \int p(\vec{\theta}^{(i)})p(\vec{\theta^*}^{(i)}|\vec{\theta}^{(i)})d\vec{\theta}^{(i)}$$

The average probability of reaching $\vec{\theta^*}^{(i)}$ from other points is equal to the probability that $\vec{\theta^*}^{(i)}$ is chosen.

Application of Metropolitan Hasting Algorithm

We may apply to the algorithm to $\pi(\vec{\theta}^{(i)}) = f_{\vec{\Theta}^{(i)}|\mathcal{D}}(\vec{\theta}^{(i)}|\mathcal{D})$. Then the accepting probability is

$$A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)}) := \min\left(\frac{f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta^*}^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta^*}^{(i)}) T(\vec{\theta}_n^{(i)}|\vec{\theta^*}^{(i)})}{f_{\mathcal{D}|\vec{\Theta}^{(i)}}(\mathcal{D}|\vec{\theta}_n^{(i)}) f_{\vec{\Theta}^{(i)}}(\vec{\theta}_n^{(i)}) T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})}, 1\right)$$

We have avoided the integral in the denominator! In fact, it is computationally more efficient than the Rejection-Based ABC

Metropolitan Hasting ABC

In fact, we may combine the Metropolitan Hasting Algorithm with Rejection-Based ABC to simulate the posterior distribution: Initialize a starting point $\vec{\theta}_1^{(i)}$ and repeat the following

1. Given $\vec{\theta}_n^{(i)}$. Propose a new point $\vec{\theta^*}^{(i)}$ by sampling from *proposal* distribution $T(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$.
2. Obtain a sample $\tilde{D}$ from $\mathcal{D}|\vec{\theta}_n^{(i)}$.
3. **If $\tilde{D}$ and actual data $\mathcal{D}$ are 'close enough'**, then set $\vec{\theta}_{n+1}^{(i)} = \vec{\theta^*}^{(i)}$ with probability $A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$, and $\vec{\theta}_{n+1}^{(i)} = \vec{\theta}_n^{(i)}$ with probability $1 - A(\vec{\theta^*}^{(i)}|\vec{\theta}_n^{(i)})$.
4. **Otherwise** set $\vec{\theta}_{n+1}^{(i)} = \vec{\theta}_n^{(i)}$

In our case, ...

We let $T \sim \mathcal{N}(\vec{\theta}^{(i)}, \Sigma)$, where $\Sigma = \text{diag}(10, 500, 1)$. In such case T is symmetric and we don't need to include the correction factor

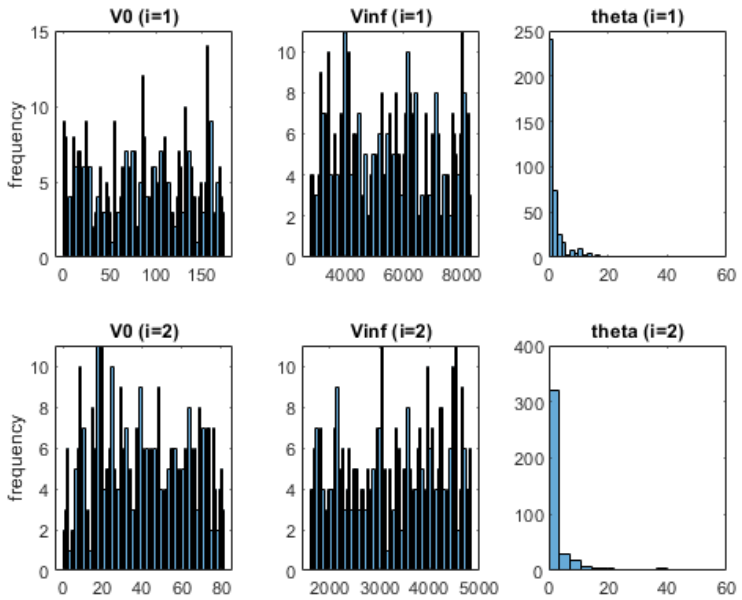
Actual Code in Matlab (Part I)

```
function posterior = MetroHasting(t,V,deltaV,N)
% t is input variable,  V is response variable
% N is number of sample required,  Step is the sd of proposal function

% ===== Preparation =====
priorpdf = @(theta) 1/(V(1)*V(end)*(1+(theta(3))^2)); % prior pdf
MeanVhat = @(theta) Gompertz(theta(1), theta(2), theta(3), t); % mean of
    likelihood
SDVhat = mean(deltaV); %SD of likelihood
likelihood = @(theta) prod(normpdf(V, MeanVhat(theta), SDVhat)); %likelihood
target = @(theta) likelihood(theta)*priorpdf(theta); %target function

% ===== Initialization =====
theta = [V(1)*rand, V(end)+2*V(end)*rand, abs(tan(pi*(rand-0.5)))]; % Starting
    Points
posterior = zeros(3,N); % Pre-allocation
```

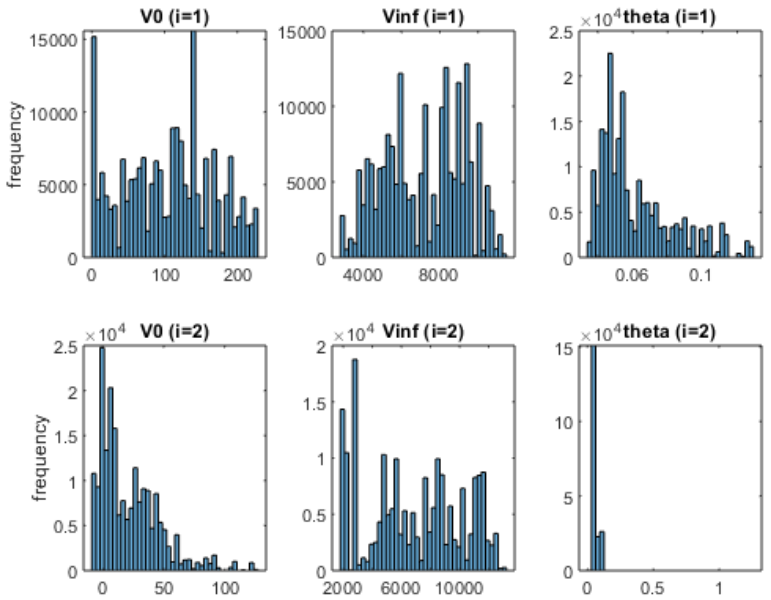
Actual Code in Matlab (Part II)

```
% ===== Loop =====
for i = 1:N

    % Propose New Point
    protheta = [theta(1) + 10*randn, theta(2) + 500*randn, theta(3) + randn];

    % Simulation for Rejection
    Vsim = randn(1,7)*SDVhat + MeanVhat(protheta);

    % Rejection
    dist = norm(V-Vsim);
    acc = min(1, target(protheta)/target(theta));

    if (dist < 1000) && (rand < acc)
            posterior(:,i) = protheta;
            theta = protheta;
    else
            posterior(:,i) = theta;
    end
end
```

Actual Simulation

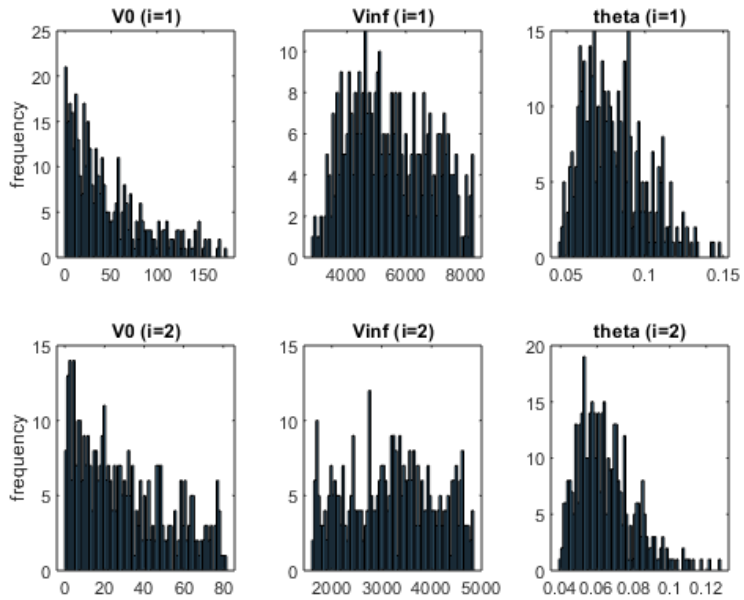


(a) Samples from Prior

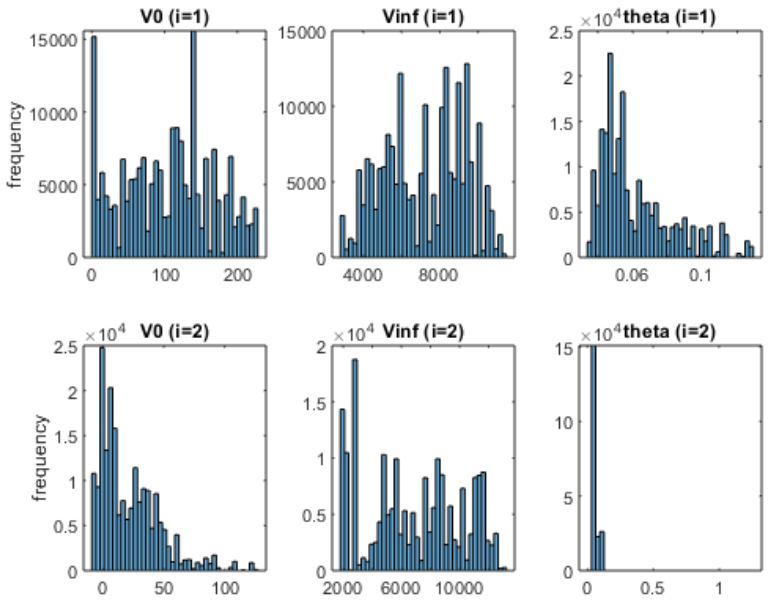


(b) Samples from Posterior (n = 200000)

Actual Simulation



(a) Samples from Posterior (Rejection-Based ABC)



(b) Samples from Posterior (Metropolitan Hasting ABC)

Trade Off

Sometimes we are unlucky - we encounter the same problem as in Rejection-Based ABC. In that case Metropolitan Hasting ABC is meaningless.

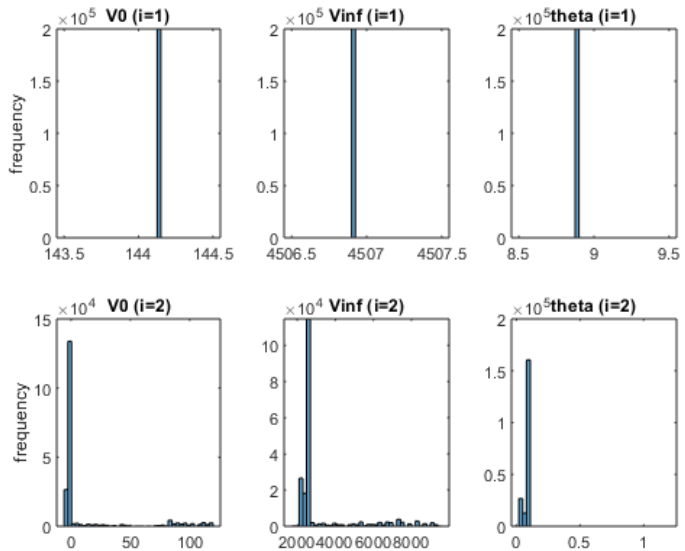


Figure: Samples from Prior

Predictive Distribution

Given $t = t^*$, then the probability density function of $V^{(i)}(t^*) = V^*$ is:

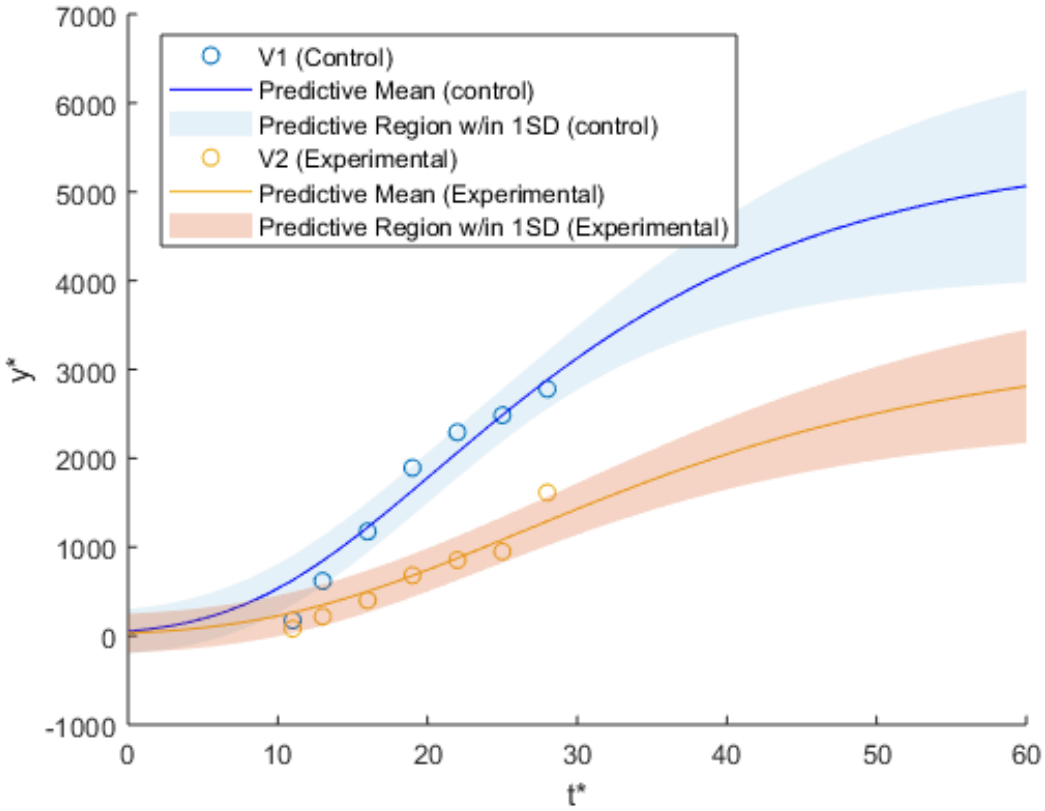
$$\begin{aligned} f(V^*|t^*, \mathcal{D}) &= \int_R f(V^*|t^*, \vec{\theta}^{(i)}) f_{\Theta^{(i)}|\mathcal{D}}(\vec{\theta}^{(i)}|\mathcal{D}) d\vec{\theta}^{(i)} \\ &= E_{\Theta^{(i)}|\mathcal{D}}(f(V^*|t^*, \vec{\theta}^{(i)})) \end{aligned}$$

We can thus evaluate the integral by using a Crude Monte Carlo (finding the mean of $f(.)$ evaluated at the samples of posterior distribution.) We are particularly interested with the mean and SD of predictive distribution.

Actual Code in Matlab

```
function [expvec, sdvec] = predictive(posterior,deltaV,eval)
% 'posterior' refers to sample from posterior
% 'eval' refers to values to be evaluated

expvec = zeros(1,length(eval));
sdvec = zeros(1,length(eval));

for i = 1:200
    samplematrix = zeros(length(posterior),400);
    for j = 1:length(posterior)
        thetak = posterior(:,j);
        MeanVhat = Gompertz(thetak(1), thetak(2), thetak(3), eval(i));
        SDVhat = mean(deltaV);
        samplematrix(j,:) = randn(1,400).*SDVhat + MeanVhat;
    end
    expvec(i) = mean(samplematrix,'all');
    sdvec(i) = sqrt(var(samplematrix,1,'all'));
end
end
```

Actual Plot


V1 (Control)
Predictive Mean (control)
Predictive Region w/in 1SD (control)
V2 (Experimental)
Predictive Mean (Experimental)
Predictive Region w/in 1SD (Experimental)
y*
t*
7000
6000
5000
4000
3000
2000
1000
0
-1000
0
10
20
30
40
50
60

Back to Scenario

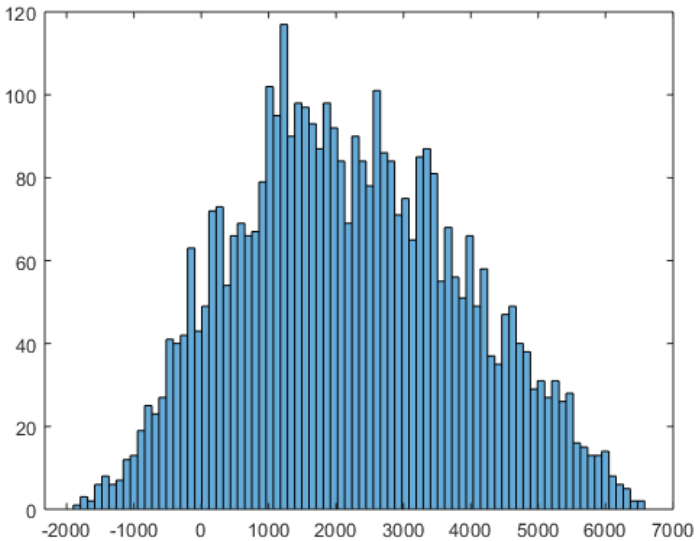
Now we need to test:

1. Whether there are delay in tumor development ($H_0 : \theta^{(1)} = \theta^{(2)}, H_1 : \theta^{(1)} > \theta^{(2)}$).
2. Whether there are reduction in volume of tumor at equilibrium ($H_0 : V_\infty^{(1)} = V_\infty^{(2)}, H_1 : V_\infty^{(1)} > V_\infty^{(2)}$).

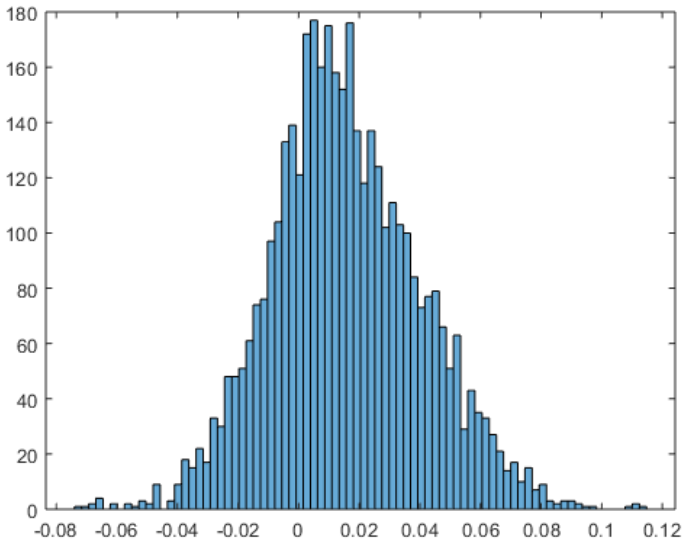
However we do not have enough information. We will just test the statements crudely using marginal distributions.

Monte Carlo Test (or is it?)

A rather rough way to perform hypothesis testing (*) on (1) is to obtained an ordered sample of $\theta^{(1)} - \theta^{(2)}$ and see the empirical probability that $\theta^{(1)} - \theta^{(2)} \leq 0$, similar for (2).



(a) Samples from $V_{\infty}^{(1)} - V_{\infty}^{(2)}$



(b) Samples from $\theta^{(1)} - \theta^{(2)}$

(*) I am not referring to actual hypothesis testing.
Here we have

- $p(\theta^{(1)} - \theta^{(2)} > 0) = 0.0125$,
- $p(V_\infty^{(1)} - V_\infty^{(2)} > 0) = 0.7398$

How do we interpret the probability? We can directly say that there is 90% chances for H_1 to be true in (1), while there is 74% chances for H_1 to be true in (2).

We are actually not doing hypothesis tests! Moreover, I had made a rather unreasonable assumption that the variables in posterior distributions are **independent**. I need to go back to M2S2 lectures again...

Conclusion

To conclude, we have had an excursion on various ways to do Bayesian Approach. We have witnessed the power of computer in doing simulations and statistics. Let us discover more how Bayesian Approach affect our analysis.