Development and calibration of tumor models

Ernesto A. B. F. Lima
Emanuelle A. Paixão

Laboratório Nacional de Computação Científica
UNIDADE DE PESQUISA DO MCTI

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

PÁTRIA AMADA BRASIL
GOVERNO FEDERAL

TMG

INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES
ODEN

TACC
TEXAS ADVANCED COMPUTING CENTER

TEXAS
The University of Texas at Austin

THE CENTER FOR COMPUTATIONAL ONCOLOGY

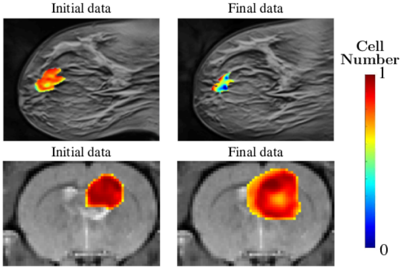
lima@ices.utexas.edu; earantes@lncc.br

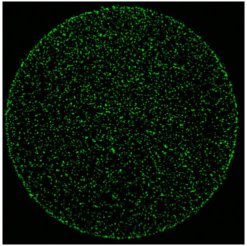
Model Calibration

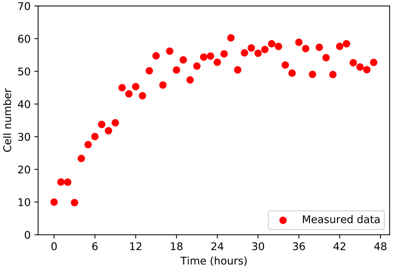
- Calibration Theory:
  - Maximum Likelihood Estimation
  - Bayesian Approach
- EMCEE Python Library[1]

[1] Foreman-Mackey, Daniel, et al. "emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC." arXiv preprint arXiv:1911.07688 (2019).

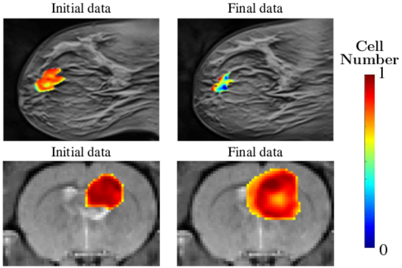
Motivation

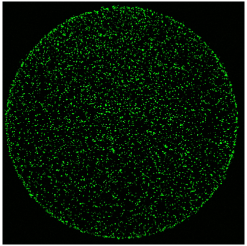

Initial data
Final data
Initial data
Final data
Cell Number
1
0

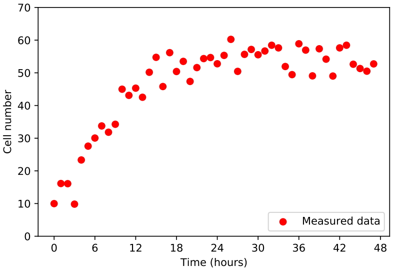



Cell number
Time (hours)
Measured data
0
10
20
30
40
50
60
70
0
6
12
18
24
30
36
42
48

Motivation



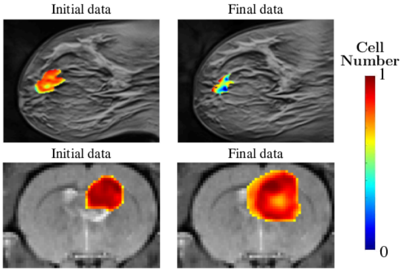


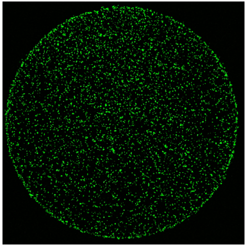


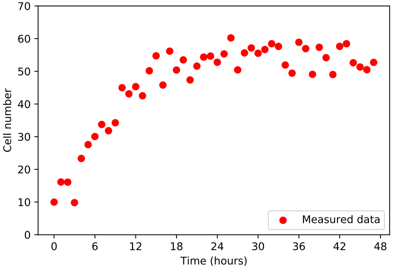
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- N: number of tumor cells;
- r: tumor growth rate;
- K: environment carrying capacity.

Motivation







$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- N: number of tumor cells;
- r: tumor growth rate;
- K: environment carrying capacity.

Model Calibration

It is done to adjust the selected model parameters, such as growth and death rates, to obtain the best fit between the model predicted responses and the available data.

Model calibration: frequentist and Bayesian statistics[2]

Frequentist and maximum likelihood approaches

- there is an unknown but fixed parameter that represents the event;
- maximum likelihood approach: find the parameter that enables the model to deliver the best characterization of the true target distribution.

[2] Oden et al., Encyclopedia of Computational Mechanics Second Edition (2017)

Model calibration: frequentist and Bayesian statistics[2]

Frequentist and maximum likelihood approaches

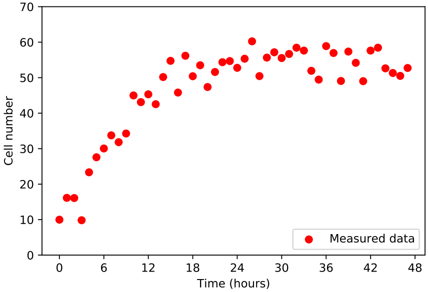
- there is an unknown but fixed parameter that represents the event;
- maximum likelihood approach: find the parameter that enables the model to deliver the best characterization of the true target distribution.

Bayesian approach

- it maps prior beliefs about the parameters to new posterior beliefs in the light of observing the data;
- the estimated parameter is regarded as a random variable, and the true parameter is merely a realization of the random variable.

[2] Oden et al., Encyclopedia of Computational Mechanics Second Edition (2017)

Definitions



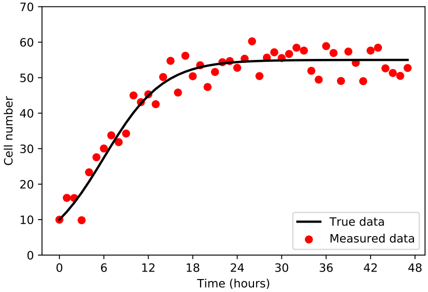
- $\boldsymbol{D}$: measured data;

Mathematical model

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

Definitions



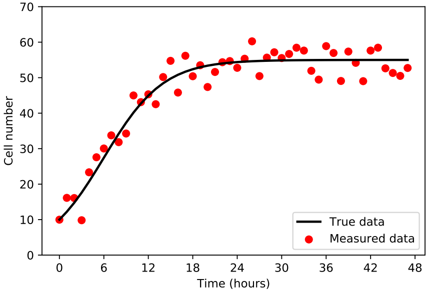
- $\boldsymbol{D}$: measured data;
- $\boldsymbol{T}$: true data;

Mathematical model

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

Definitions



- $\boldsymbol{D}$: measured data;
- $\boldsymbol{T}$: true data;

$$\boldsymbol{T} + \epsilon = \boldsymbol{D},$$

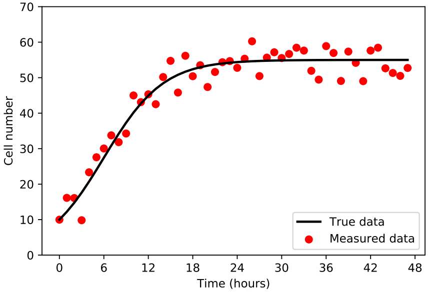
- ϵ: experimental noise;

Mathematical model

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

Definitions



- $\boldsymbol{D}$: measured data;
- $\boldsymbol{T}$: true data;

$$\boldsymbol{T} + \epsilon = \boldsymbol{D},$$

- ϵ: experimental noise;

Mathematical model

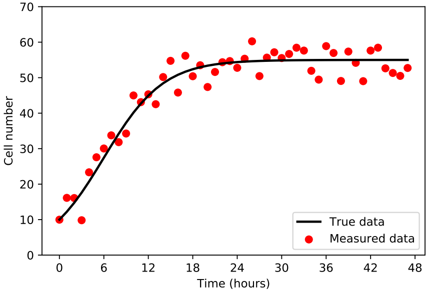
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

$$\boldsymbol{T} - \boldsymbol{Y}(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta}),$$

- γ: model inadequacy.

Definitions



- $\boldsymbol{D}$: measured data;
- $\boldsymbol{T}$: true data;

$$\boldsymbol{T} + \epsilon = \boldsymbol{D},$$

- ϵ: experimental noise;

Mathematical model

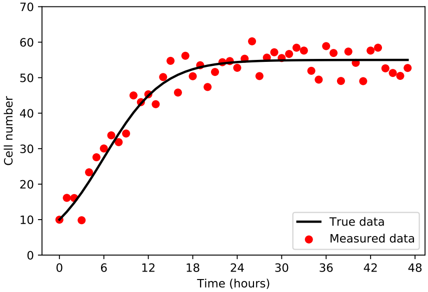
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

$$\boldsymbol{T} - \boldsymbol{Y}(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta}),$$
$$\boldsymbol{T} - \boldsymbol{Y}^h(\boldsymbol{\theta}) = \gamma^h(\boldsymbol{\theta}),$$

- γ: model inadequacy.

Definitions



Mathematical model

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

- $\boldsymbol{\theta}$: vector of model parameters, $\boldsymbol{\theta} = (r, K)$;
- r: tumor growth rate;
- K: environmental carrying capacity;
- $\boldsymbol{Y}(\boldsymbol{\theta})$: model prediction;

- $\boldsymbol{D}$: measured data;
- $\boldsymbol{T}$: true data;

$$\boldsymbol{T} + \epsilon = \boldsymbol{D},$$

- ϵ: experimental noise;

$$\boldsymbol{T} - \boldsymbol{Y}(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta}),$$

$$\boldsymbol{T} - \boldsymbol{Y}^h(\boldsymbol{\theta}) = \gamma^h(\boldsymbol{\theta}),$$

- γ: model inadequacy.

$$\boldsymbol{D} - \boldsymbol{Y}(\boldsymbol{\theta}) = \epsilon + \gamma(\boldsymbol{\theta}),$$

Maximum likelihood estimation

Likelihood

The conditional probability that the data ($\boldsymbol{D}$) is observed for a given set of parameters ($\boldsymbol{\theta}$) is the likelihood $\pi(\boldsymbol{D}|\boldsymbol{\theta})$.

Maximum likelihood estimation

Likelihood

The conditional probability that the data ($\boldsymbol{D}$) is observed for a given set of parameters ($\boldsymbol{\theta}$) is the likelihood $\pi(\boldsymbol{D}|\boldsymbol{\theta})$.

$$\boldsymbol{D} - \boldsymbol{Y}(\boldsymbol{\theta}) = \boldsymbol{\epsilon} + \boldsymbol{\gamma}(\boldsymbol{\theta}),$$

Assuming:

1. the experimental noise is normally distributed ($\boldsymbol{\epsilon} \sim \mathcal{N}(0_{N\times 1}, \sigma^2_{data}\boldsymbol{I}_{N\times N})$);
2. the model inadequacy is normally distributed ($\boldsymbol{\gamma} \sim \mathcal{N}(0_{N\times 1}, \sigma^2_{model}\boldsymbol{I}_{N\times N})$);
3. the variance of the total error (σ) is such as $\sigma^2 = \sigma^2_{data} + \sigma^2_{model}$;
4. the data is normally distributed;

Maximum likelihood estimation

Likelihood

The conditional probability that the data ($\boldsymbol{D}$) is observed for a given set of parameters ($\boldsymbol{\theta}$) is the likelihood $\pi(\boldsymbol{D}|\boldsymbol{\theta})$.

$$\boldsymbol{D} - \boldsymbol{Y}(\boldsymbol{\theta}) = \boldsymbol{\epsilon} + \gamma(\boldsymbol{\theta}),$$

Assuming:

1. the experimental noise is normally distributed ($\boldsymbol{\epsilon} \sim \mathcal{N}(0_{N\times 1}, \sigma^2_{data}\boldsymbol{I}_{N\times N})$);
2. the model inadequacy is normally distributed ($\gamma \sim \mathcal{N}(0_{N\times 1}, \sigma^2_{model}\boldsymbol{I}_{N\times N})$);
3. the variance of the total error (σ) is such as $\sigma^2 = \sigma^2_{data} + \sigma^2_{model}$;
4. the data is normally distributed;

$$\pi(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N_t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(D_i - Y_i(\boldsymbol{\theta}))^2}{2\sigma^2}},$$

- N_t: the number of data points.

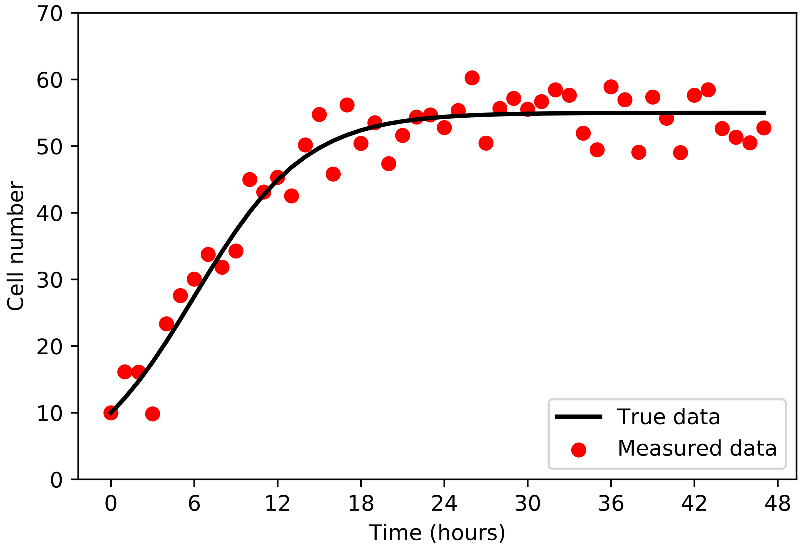
Maximum likelihood estimation

Find $\hat{\boldsymbol{\theta}} \in \Theta$, where Θ is the parameter space, that maximize the likelihood.

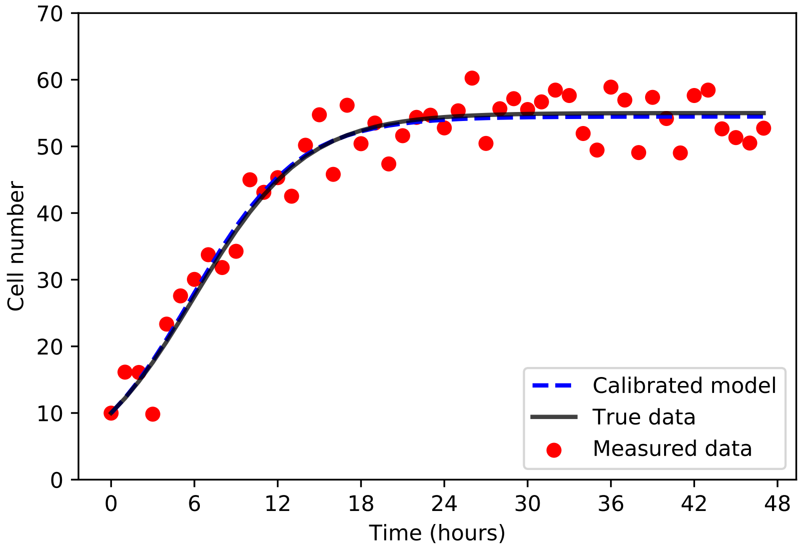
It is customary to work with the more manageable log-likelihood function.

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} [\log \pi(\boldsymbol{D} | \boldsymbol{\theta})]; \\
&= \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \left[-\frac{1}{2} \sum_{i=1}^{N_t} \left(\log(2\pi) + \log(\sigma^2) + \frac{(D_i - Y_i(\boldsymbol{\theta}))^2}{\sigma^2} \right) \right].
\end{aligned}
$$

Maximum likelihood estimation



Maximum likelihood estimation



$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right),$$

$$r = 0.25$$
$$K = 59.14$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$
$$P(A, B) = P(B, A);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$
$$P(A, B) = P(B, A);$$
$$P(B, A) = P(B|A)P(A);$$

Bayesian approach

Given events A and B:

$$
\begin{aligned}
P(A, B) &= P(A|B)P(B); \\
P(A, B) &= P(B, A); \\
P(B, A) &= P(B|A)P(A); \\
P(A|B)P(B) &= P(B|A)P(A);
\end{aligned}
$$

Bayesian approach

Given events A and B:

$$
\begin{aligned}
P(A, B) &= P(A|B)P(B); \\
P(A, B) &= P(B, A); \\
P(B, A) &= P(B|A)P(A); \\
P(A|B)P(B) &= P(B|A)P(A); \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B)};
\end{aligned}
$$

Bayesian approach

Given events A and B:

$$
\begin{aligned}
P(A, B) &= P(A|B)P(B); \\
P(A, B) &= P(B, A); \\
P(B, A) &= P(B|A)P(A); \\
P(A|B)P(B) &= P(B|A)P(A); \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B)};
\end{aligned}
$$

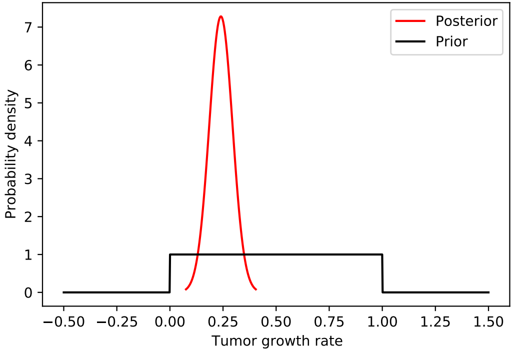
Converting to probability densities π, if A represents the parameter $\boldsymbol{\theta}$ of a model, and B the observational data $\boldsymbol{D}$:

$$
\underbrace{\pi(\boldsymbol{\theta}|\boldsymbol{D})}_{\text{posterior}} = \frac{\overbrace{\pi(\boldsymbol{D}|\boldsymbol{\theta})}^{\text{likelihood}}\,\overbrace{\pi(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{\pi(\boldsymbol{D})}_{\text{evidence}}}; \qquad \pi(\boldsymbol{D}) = \int_{\Theta} \pi(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\theta
$$

Bayesian approach

Given events A and B:

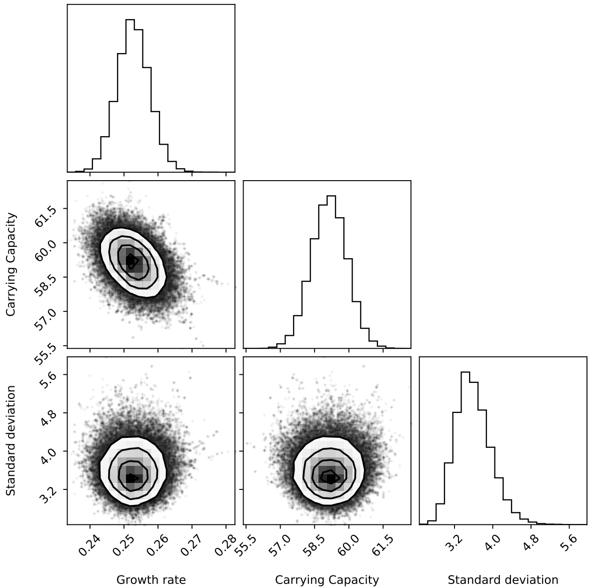
$$P(A, B) = P(A|B)P(B);$$
$$P(A, B) = P(B, A);$$
$$P(B, A) = P(B|A)P(A);$$
$$P(A|B)P(B) = P(B|A)P(A);$$
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)};$$



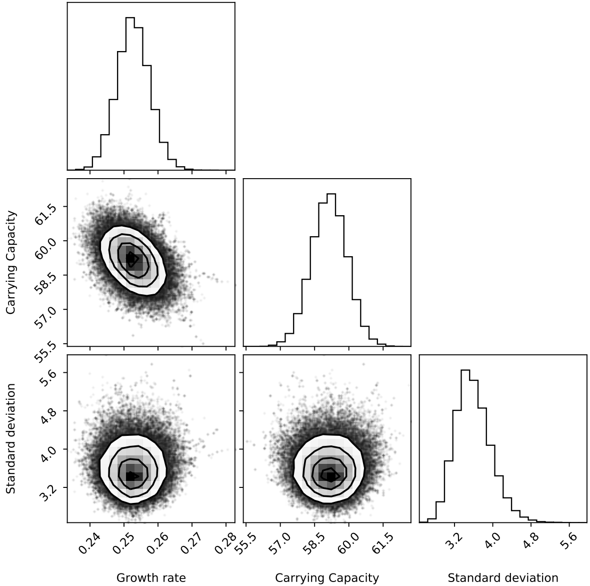
Converting to probability densities π, if A represents the parameter $\boldsymbol{\theta}$ of a model, and B the observational data $\boldsymbol{D}$:

$$\underbrace{\pi(\boldsymbol{\theta}|\boldsymbol{D})}_{\text{posterior}} = \frac{\overbrace{\pi(\boldsymbol{D}|\boldsymbol{\theta})}^{\text{likelihood}}\,\overbrace{\pi(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{\pi(\boldsymbol{D})}_{\text{evidence}}}; \qquad \pi(\boldsymbol{D}) = \int_{\Theta} \pi(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\theta$$

Bayesian approach

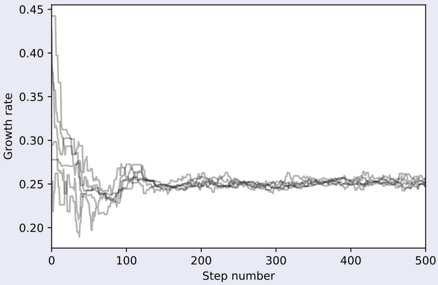


Bayesian approach

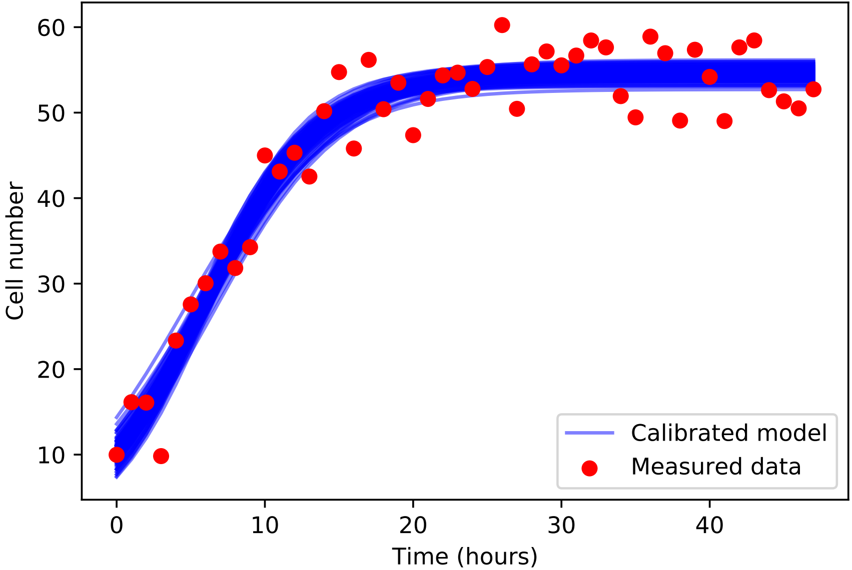


Markov Chain Monte Carlo:

Markov Chain Monte Carlo methods draw samples where the next sample is dependent on the existing sample, called a Markov Chain.



Bayesian approach



Jupyter Notebook - Binder

https://github.com/Ernesto-Lima/ISMCO2021