Modeling Antibody Levels Post SARS-CoV-2 Vaccination

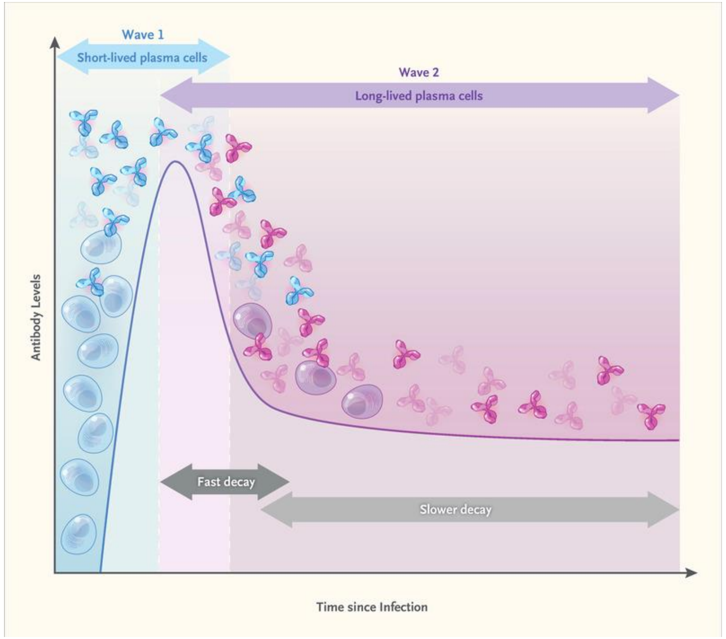
Zihao Zhao, Johns Hopkins University
Rayanne Luke, George Mason University

Background

Antibodies: proteins made by body's immune system to help fight off infections[1]

Immune system learns to make antibody levels through vaccinations

IgG(Immunoglobulin G) antibody: important in fighting infections



[2]

[1]U.S. Food & Drug Administration
[2]New England Journal of Medicine

Motivation & Data

Model the antibody response: antibody level & time (antibody level depend on time)

More than 70,000 samples are utilized from [1]

Total data are divided into 4 classes [1]:

Class1: plausibly previously infected group

Class2: high antibody response group

Class3: medium antibody response group

Class4: low antibody response group

[1] (Wei, 2021)

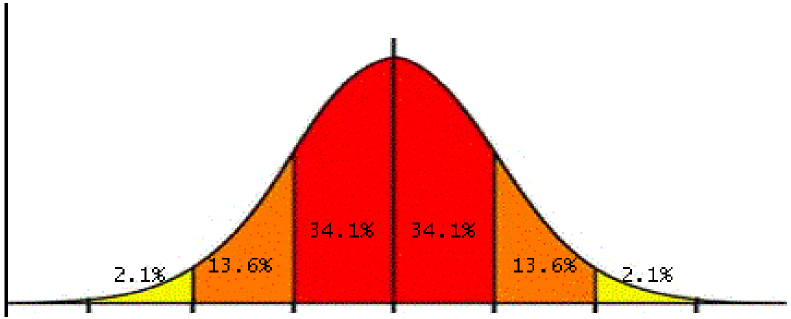
Probabilistic model

Deterministic models provide a single prediction for each input

Probabilistic models: probabilistic characterization of the uncertainty

Variability in antibody response

$$f(x_i;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$



[1]

[1] study.com

Method

MLE(maximum likelihood estimation) is applied to find the best fit parameters.

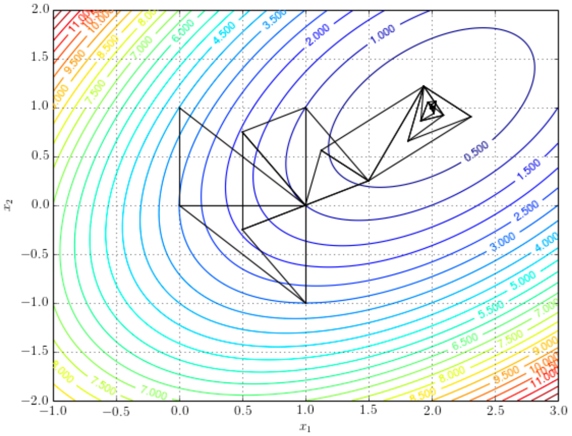
Get the likelihood function

$$L(\theta) = \prod_{i=1}^{n} f_Y(y_i;\theta)$$

Take log sum

Use optimization to find the maximum

Optimization algorithm: Nelder-Mead



[1]

[1] wikipedia

Time independent model

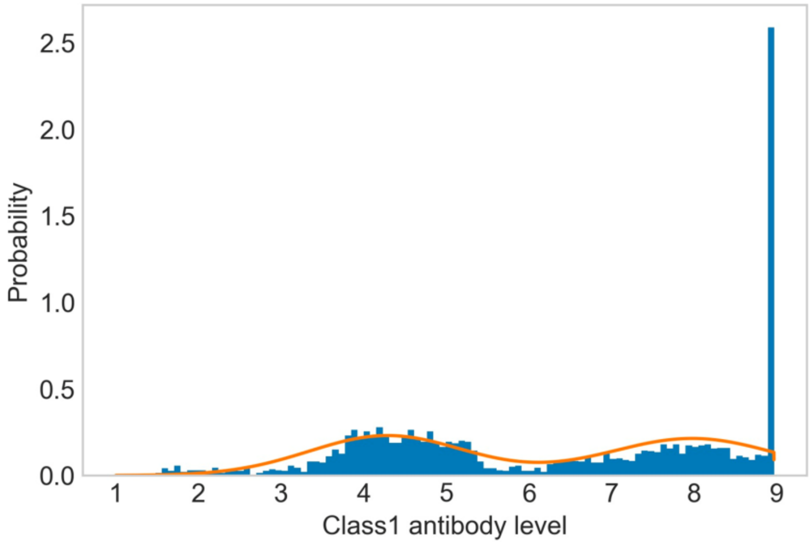
Data censoring:

Data are cut off at 2 ng/ml and 500 ng/ml

Take log to see the distribution of data clearly

Bimodal distribution

$$\frac{\frac{1}{\sqrt{2*\pi}}e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}+\frac{1}{\sqrt{2*\pi}}e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}}{U}$$



Why we need to do time dependent model

Antibody level are dependent on time

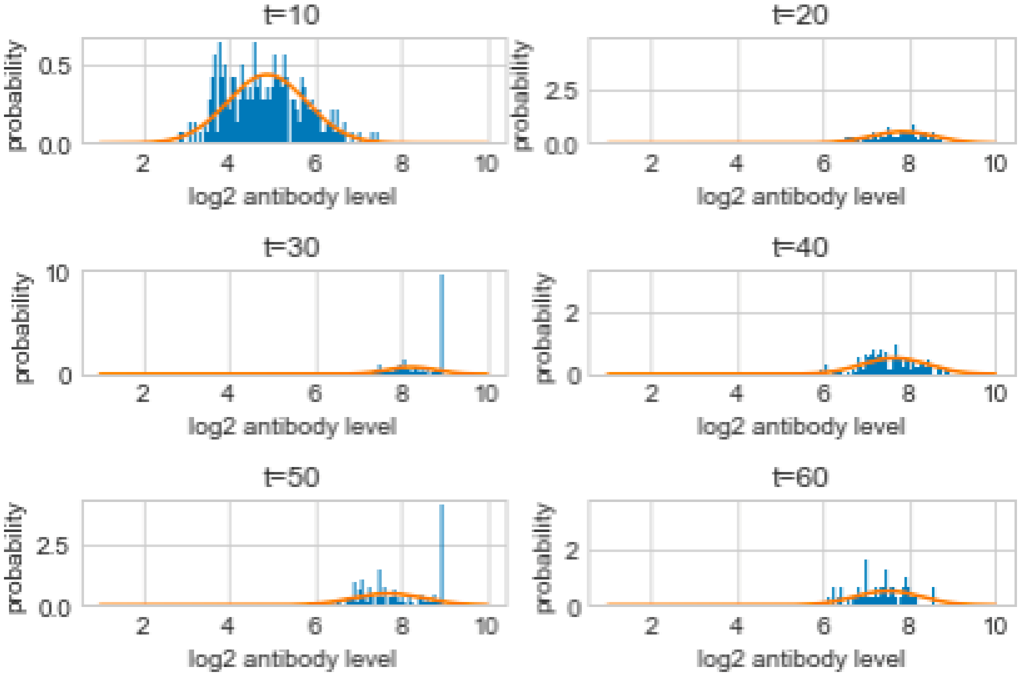
Wei et al. data qualitatively different across classes

Time dependent model

Normal distribution

Proposed model:

$$f(x,t) = \frac{1}{\sigma(t)\sqrt{2 * \pi}} * e^{\frac{-1}{2} * (\frac{x-\mu(t)}{\sigma(t)})^2}$$



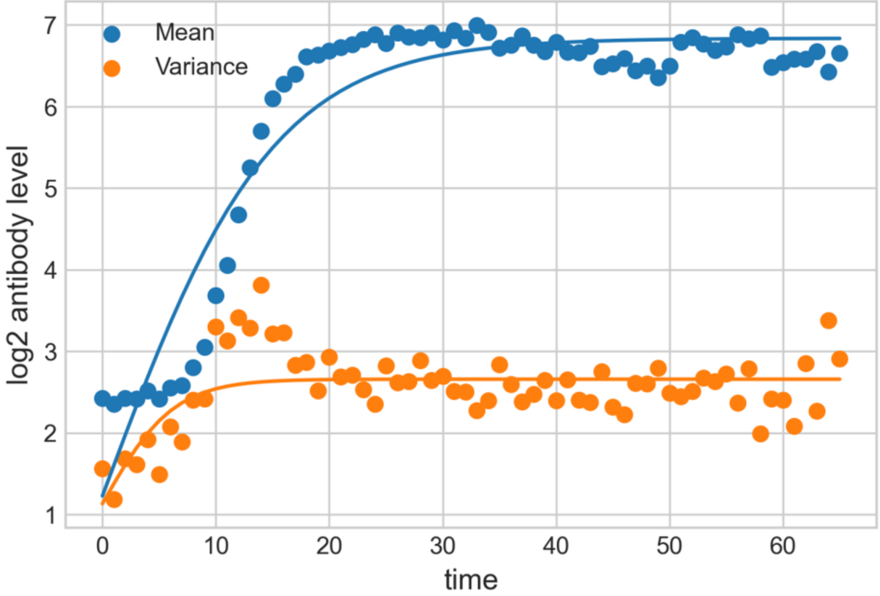
Time dependent model

Logistic function(change in time)

Proposed model for mean and variance:

$$\mu(t) = \frac{c_1}{1 + e^{b_1 t}} + d_1$$

$$\sigma(t) = \frac{c_2}{1 + e^{b_2 t}} + d_2$$

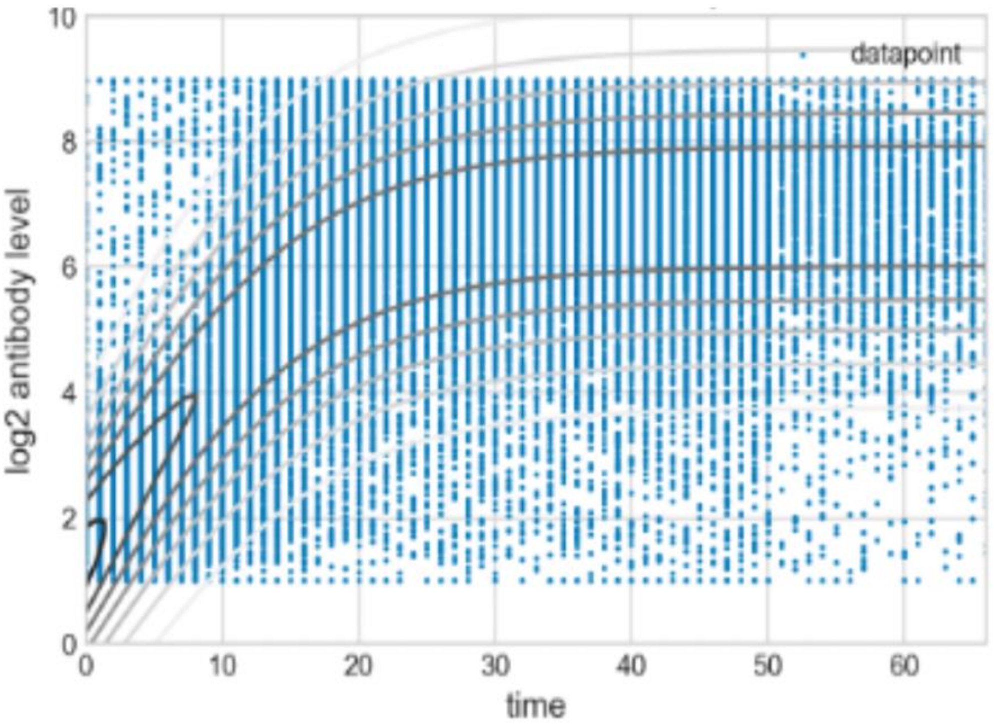


Time dependent model

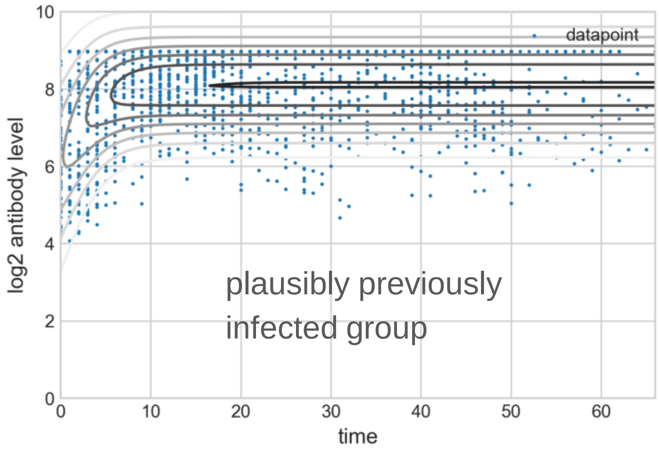
Use MLE to fit the model to the data

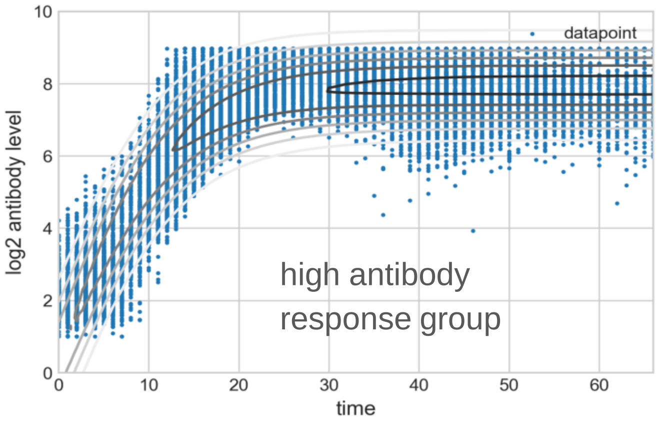
The contour line represents the probabilistic model

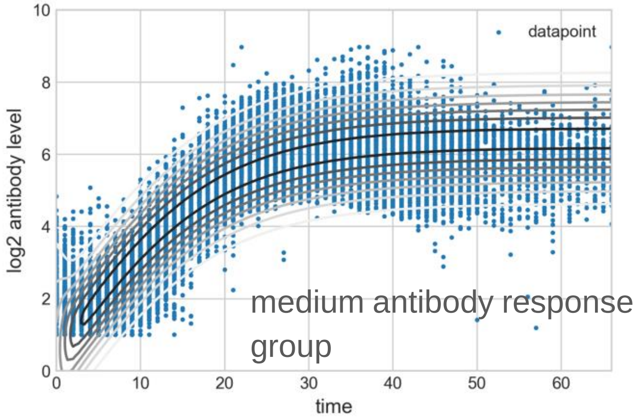
Model for the total data (4 classes combined)

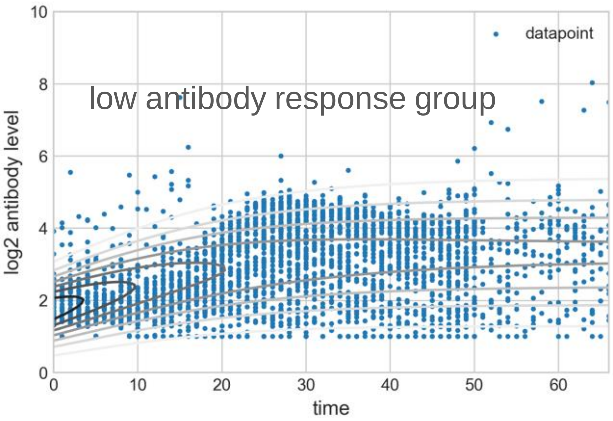


Time dependent model of each class









KL divergence between classes

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right),$$

Class	1	2	3	4	Total data
1		0.89	0.90	**2.30**	0.80
2			0.40	1.65	0.39
3				0.71	**0.14**
4					0.89

Conclusion & Future work

The models for each class: qualitatively different → confirms Wei's classification

Ongoing: Uncertainty quantification

Model-form error and effect of initial guess on optimal parameters

Future work: Extrapolate the antibody level change after 70 days

Future: Suggest optimal timing for boosters

Future: Suggest protocols for individuals in potentially highly vulnerable groups

Thank you

KL divergence over vaccine brand

Pfizer / Astra-Zeneca	1	2	3	4
1	0.46			
2		0.50		
3			0.71	
4				0.70