Modeling Antibody Levels Post SARS-CoV-2 Infection

Layan S. Wahdan, Arpan Das

Mason Experimental Geometry Lab

December 6th 2024

Introduction and Goals

Using SARS-CoV-2 antibody data from convalescent patients obtained from a longitudinal study by Yang et al:

- Model SARS-CoV-2 antibody data with higher dimensionality and capture time dependence of antibody response.
- Select variables for better data separation and biological significance.
- visualize patterns in the variability in immune response within a population of patients infected with SARS-CoV-2 over time.

Definition (Antibodies)

- Proteins that bind to viral antigens.
- Mark antigens for destruction or block cell entry.
- Quantitatively measure immune response.
- IgG and IgA: immunoglobulin G and A, respectively; types of antibodies.
- Anti-RBD and N: Indicates the antibody targets the receptor binding domain 'spikes' or the nucleocapsid surface on the virus, respectively

Definition (Maximum Likelihood Estimation)

- Optimizes the parameters of a probabilistic distribution function to maximize the likelihood of observing a given set of data, assuming an underlying distribution.
- Numerically, the parameters are identified by minimizing the negative logarithm of the likelihood function (the product of the pdf evaluated at each point in the dataset).
- As opposed to a deterministic model, probabilistic models are better suited to capturing variability in natural phenomena.

Let $\vec{x}$ = a list of data points $(x_1 \ldots x_n)$

Let $P(\vec{x}, \vec{p})$ = model with data $\vec{x}$ and list of parameters $\vec{p}$

Optimize the negative log likelihood over the parameters $\vec{p}$:

$$\min\left(-\ln L(\vec{p})\right) = \min\left(-\sum_{i=1}^{n} \ln P(x_i, \vec{p})\right)$$

Selecting the Gamma Distribution

- Following Bedekar et. al. (1), we select a gamma distribution to model immune response.
- Frank, 2009 (2) indicates that many biological phenomena follow this distribution.
- This distribution has two parameters: shape and scale.

$$f(r, a, b) = \frac{r^{a-1} e^{-r/b}}{\Gamma(a) b^a} \tag{1}$$

Model Details

- Figure 1 depicts a convalescent population antibody levels over time.
- the gamma distribution equation is modified to reflect the time dependence

Time-dependent gamma distribution:

$$a(t) = \frac{\theta_1 t}{1 + (\theta_2 t^2)} + a \tag{2}$$

$$f(r, a(t), b) = \frac{r^{a(t)-1} e^{-r/b}}{\Gamma(a(t)) b^{a(t)}} \tag{3}$$

- The modifications seen above are done to naturally show that at time t = 0 the equation will simplify to:

$$a(0) = a$$

which simplifies to the negative distribution

- Similarly another biological phenomenon that is accounted for is

$$\lim_{t \to \infty} a(t) = a$$

which demonstrates that as time approaches infinity the positive results will decay to similar levels as the negative distribution

- Figure 2 plots the distribution of nucleocapsid antibodies against receptor binding domain antibodies
- We use the assumption that these antibodies are independent of one another to be able to multiply their respective probability density functions and obtain the equation:

$$f(x, y, a_x, b_x, a_y, b_y) = \gamma(x, a_x, b_x) * \gamma(y, a_y, b_y) \tag{4}$$

- When the two antibodies are plotted we see a linear correlation in the data which is also represented by the gamma distribution in Figure 2 naturally
- Though we assume independence, the correlation of the data does not require any additional change of variables to uncorrelated variables to account for it.
- Figure 3 plots three antibodies: anti-RBD IgG, anti-N IgG, and anti-RBD IgA
- This represents the first step in generating 3 dimensional models similar to what was done for the 2 dimensional model in Figure 2

Graphics

Figure 1: Time Dependent Probability Model of Anti-RBD IgG

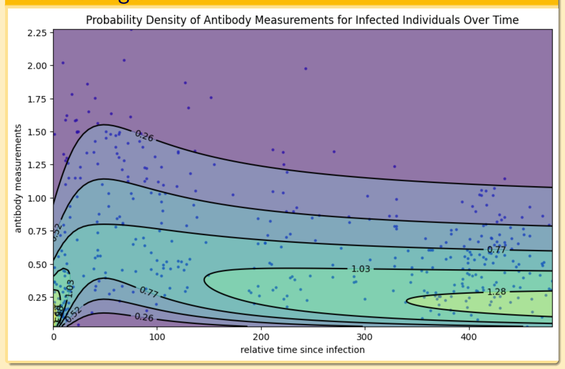


Figure 2: 2D Probability Model of Anti-RBD IgG and Anti-N IgG

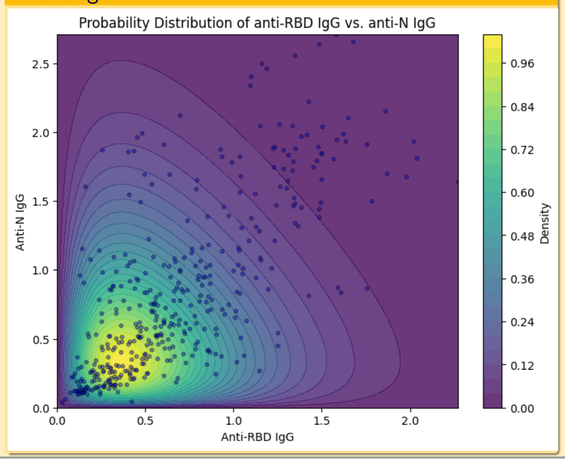
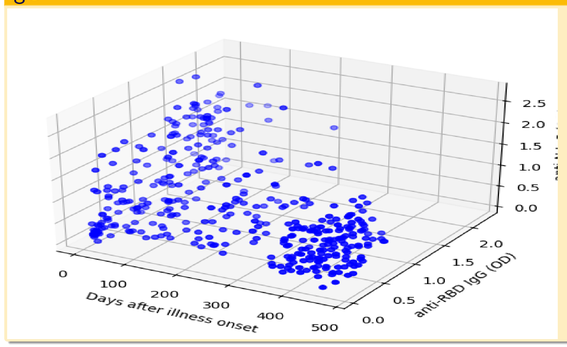


Figure 3: 3D Scatterplot of Anti-RBD IgG and IgA and Anti-N IgG



Conclusions/Future Work

- Expand dimensionality to 3 dimensions, and later 4 dimensions, for better data separation and understanding the structure of the data by identifying biologically significant variables.
- Prevalence Estimation to determine the percentage of the population that has certain characteristics of interest.
- Predict positive versus negative cases.

Acknowledgments

We would like to thank our faculty mentor, Dr. Rayanne Luke, and our graduate mentor Kelsey Ellis for their guidance and constant communication with us throughout the challenging process of conducting research, finding data, and generating models and visualizations.

References

1. Bedekar, P., Kearsley, A. J., and Patrone, P. N. (2023). Prevalence estimation and optimal classification methods to account for time dependence in antibody levels. Journal of Theoretical Biology, 559, 111375. https://doi.org/10.1016/j.jtbi.2022.111375
2. Frank, S. A. (2009). The common patterns of nature. Journal of Evolutionary Biology, 22(8), 1563–1585. https://doi.org/10.1111/j.1420-9101.2009.01775.x
3. Luke, R. A., Kearsley, A. J., Pisanic, N., Manabe, Y. C., Thomas, D. L., Heaney, C. D., and Patrone, P. N. (2023). Modeling in higher dimensions to improve diagnostic testing accuracy: Theory and examples for multiplex saliva-based SARS-CoV-2 antibody assays. PLOS ONE, 18(3), e0280823. https://doi.org/10.1371/journal.pone.0280823