Workshop on the Mathematical Modelling of Variant Replacement of Infectious Diseases Pathogens

# Maximum Likelihood Estimation of Model Parameters

Kimihito Ito

International Institute for Zoonosis Control

Hokkaido University

# Variant Replacement Model

- The frequency of variant $A_i$ at time $t$ can be presented using $R_{RI}$ of $A_i$ w.r.t $a$ as follows:

$$q_{A_i}(t) = \frac{k_i \sum_{j=1}^{l} g_{A_i}(j)\, q_{A_i}(t-j)}{\sum_{j=1}^{l} g_a(j) q_a(t-j) + \sum_{i}^{n} k_i \sum_{j=1}^{l} g_{A_i}(j) q_{A_i}(t-j)} \quad (6)$$

where $g_a(j), g_{A_1}(j), \dots g_{A_n}(j)$ are the generation time distribution of $a, A_1, \dots, A_n$.

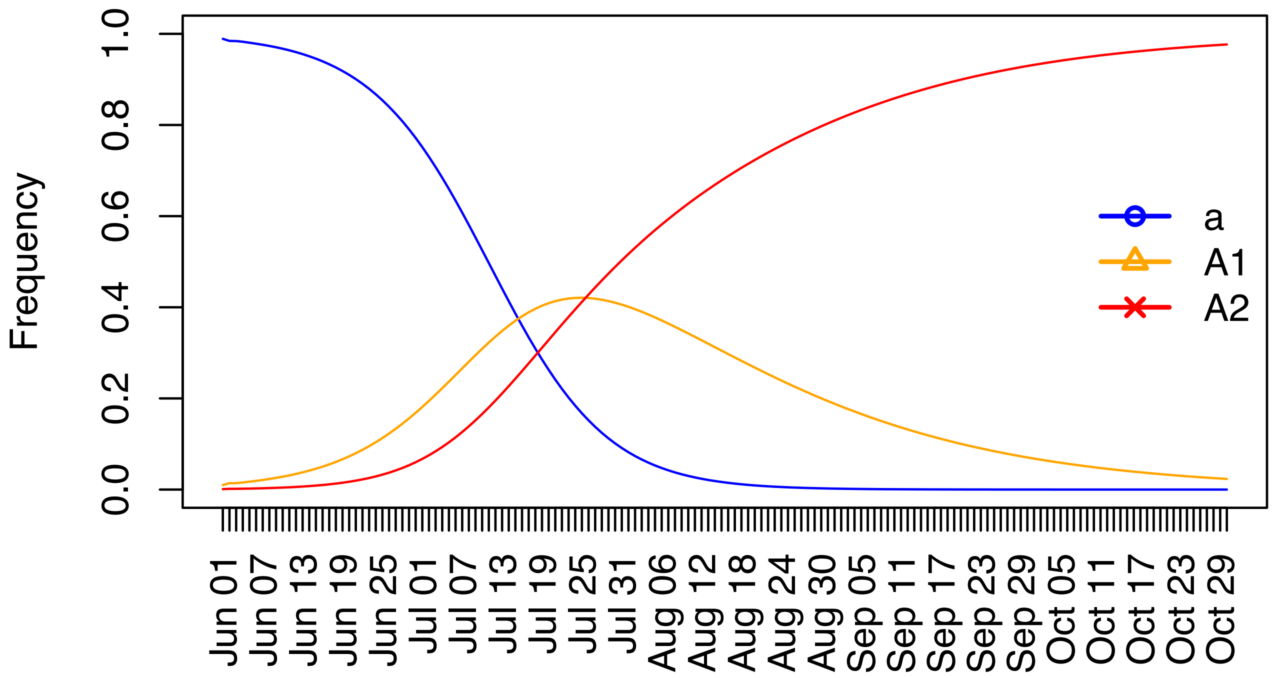
Note that the formula doesn't contain $R_a(t)$ or $I(t)$

# Calculation of Trajectory

- Consider we have variants $a, A_1, \ldots, A_n$
  Suppose we know
  - values of $k_1, \ldots, k_n$, $R_{RI}$ of $A_1, \ldots, A_n$ w.r.t $a$,
  - values of $c_1, \ldots, c_n$, $GT_R$ of $A_1, \ldots, A_n$ w.r.t $a$,
  - values of $q_{A_1}(t_s), \ldots, q_{A_n}(t_s)$, the relative frequencies of $A_1, \ldots, A_n$ at time $t_s$
- Applying the Equation (6) recursively, we can calculate the relative frequencies $q_{A_1}(t), \ldots, q_{A_n}(t)$ for any $t > t_s$ (by assuming $q_{A_i}(t) = q_{A_i}(t_s)$ for any $t < t_s$).

# Example

- We have the baseline $a$ and subjects $A_1$ and $A_2$.
- The generation time of all variants follows the gamma distribution with $\alpha = 4.03$ and $\theta = 0.735$.
- Let $k_1 = 1.4$ and $k_2 = 1.6$.
- Let $q_{A_1}(t_s) = 0.01$ and $q_{A_2}(t_s) = 0.005$ where $t_0$ is June 1.




Calculated frequencies of variants

# Observations

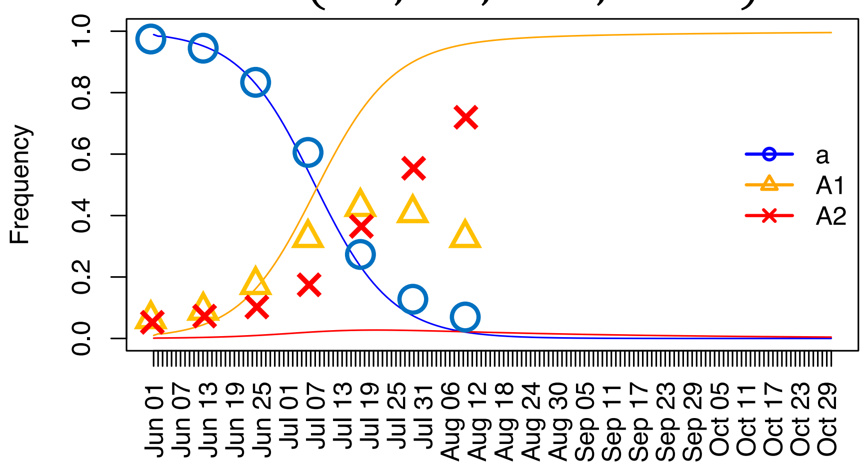
- We use counts of observed variants as observation data to estimate parameters. (Note that observations are not the relative frequencies!)

| **date_from** | **date_till** | **number of *a*** | **number of $A_1$** | **…** | **number of $A_n$** |
|---|---|---|---|---|---|
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $t_h$ | $u_h$ | $N_a(o_h)$ | $N_{A_1}(o_h)$ | $\cdots$ | $N_{A_n}(o_h)$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

# Maximum likelihood estimation
## of Parameters

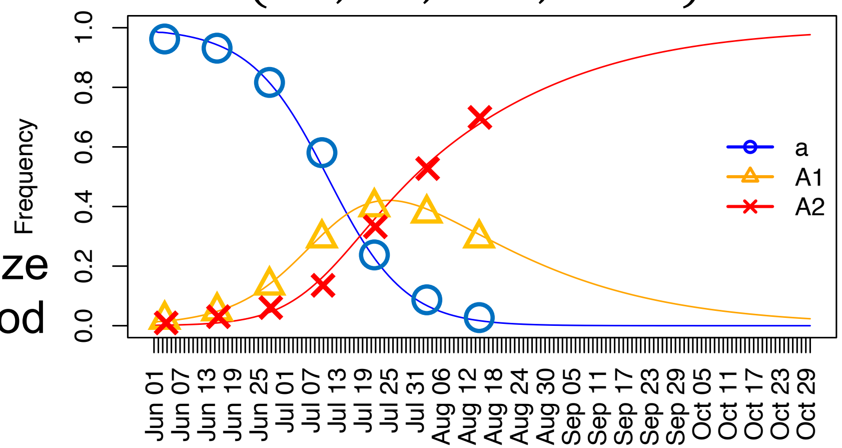
- Find the parameters
  $k_1, \ldots, k_n, q_{A_1}(t_s), \ldots, q_{A_n}(t_s)$
  that maximize the probabilities of observing actual counts.
- $c_1, \ldots, c_n$, are also esitmated if needed

$$(k_1, k_2, q_{A_1}(t_s), q_{A_2}(t_s)) = (1.5, 1.4, 0.01, 0.005)$$




Maximize likelihood

$$(k_1, k_2, q_{A_1}(t_s), q_{A_2}(t_s)) = (1.4, 1.6, 0.01, 0.005)$$



# Likelihood Function (multinomial)

- The probability that $A_1, \dots, A_n$ and $a$ were observed $N_{A_1}(o_h), \dots, N_{A_n}(o_h)$, and $N_a(o_h)$ times at period $o_h$ follows the multinomial distribution of $q_{A_1}(o_h), \dots, q_{A_n}(o_h)$.
- The likelihood function is given as follows.

$$L\left(c_1, \dots, c_n, k_1, \dots, k_n, q_{A_1}\left(t_{A_1}\right), \dots, q_{A_n}\left(t_{A_n}\right)\right)$$

$$= \prod_{h=1}^{L} \left( \frac{N(o_h)!}{N_a(o_h)!\, N_{A_1}(o_h)! \cdots N_{A_n}(o_h)!} q_a(o_h)^{N_a(o_h)} q_{A_1}(o_h)^{N_{A_1}(o_h)} \dots q_{A_n}(o_h)^{N_{A_n}(o_h)} \right)$$

# Likelihood Function (Dirichlet Multinomial)

- The probability that $A_1, \ldots, A_n$ and $a$ were observed $N_{A_1}(o_h), \ldots, N_{A_n}(o_h)$, and $N_a(o_h)$ times at period $o_h$ follows the a Dirichlet multinomial distribution with parameters $q_a\ (o_h)D, q_{A_1}(o_h)D, \ldots, q_{A_n}(o_h)D$, where D is a non-negative integer.
- The likelihood function is given as follows.

$$L\left(c_1, \ldots, c_n, k_1, \ldots, k_n, q_{A_1}\left(t_{A_1}\right), \ldots, q_{A_n}\left(t_{A_n}\right), D\right)$$

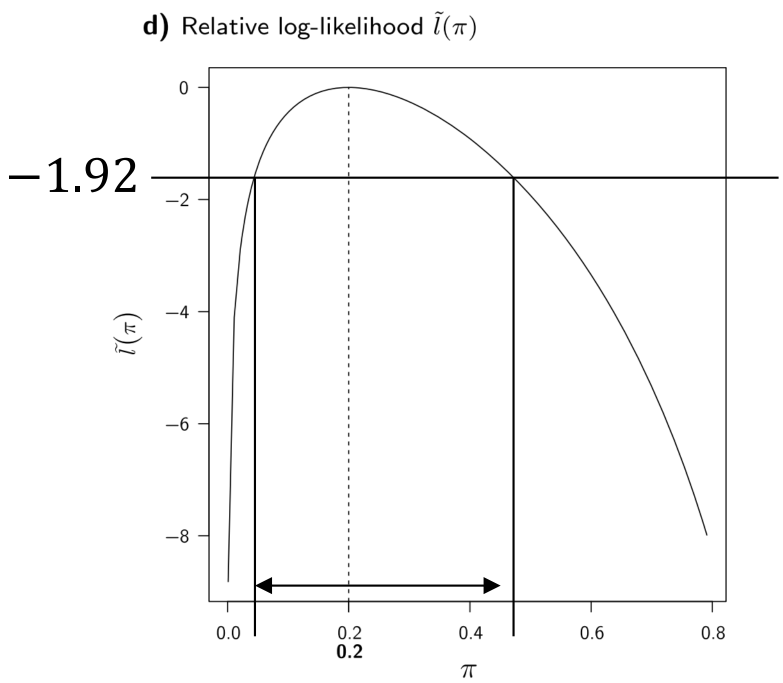
$$= \prod_{h=1}^{L} \left( \frac{\Gamma(D)\Gamma(N(o_h)+1)}{\Gamma(N(o_h)+D)} \frac{\Gamma(N_a(o_h)+q_a(o_h)D)}{\Gamma(q_a(o_h)D)\Gamma(N_a(o_h)+1)} \prod_{i=1}^{n} \frac{\Gamma\left(\Gamma\left(N_{A_i}(o_h)+q_{A_i}(o_h)D\right)\right)}{\Gamma\left(q_{A_i}(o_h)D\right)\Gamma\left(N_{A_i}(o_h)+1\right)} \right)$$

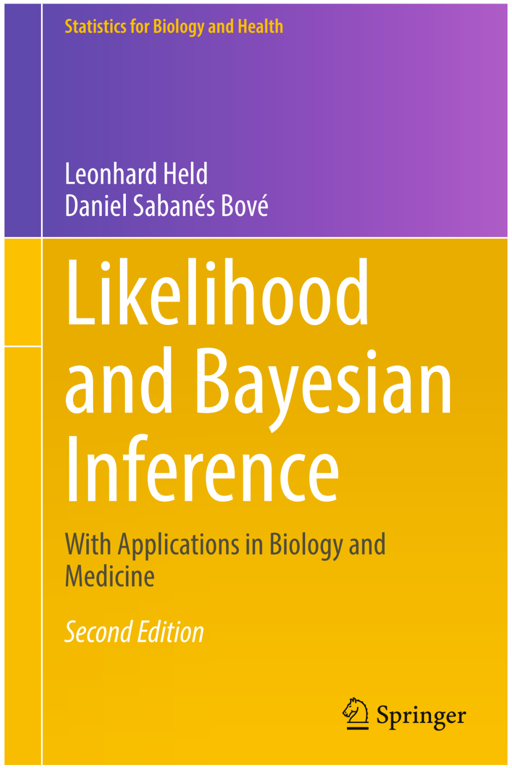
# Multinomial vs Dirichlet Multinomial

- The multinomial sample model assumes the sampled population is always the same.
- Dirichlet Multinomial allows additional errors in sampling. It can be useful when samples are drawn from sub-populations where the relative frequencies of variants are different.

# The 95% Confidence Intervals of Parameters

- The 95% confidence intervals (CIs) of parameters can be estimated using the profile likelihood method (Held & Sabanes Bove, 2020).



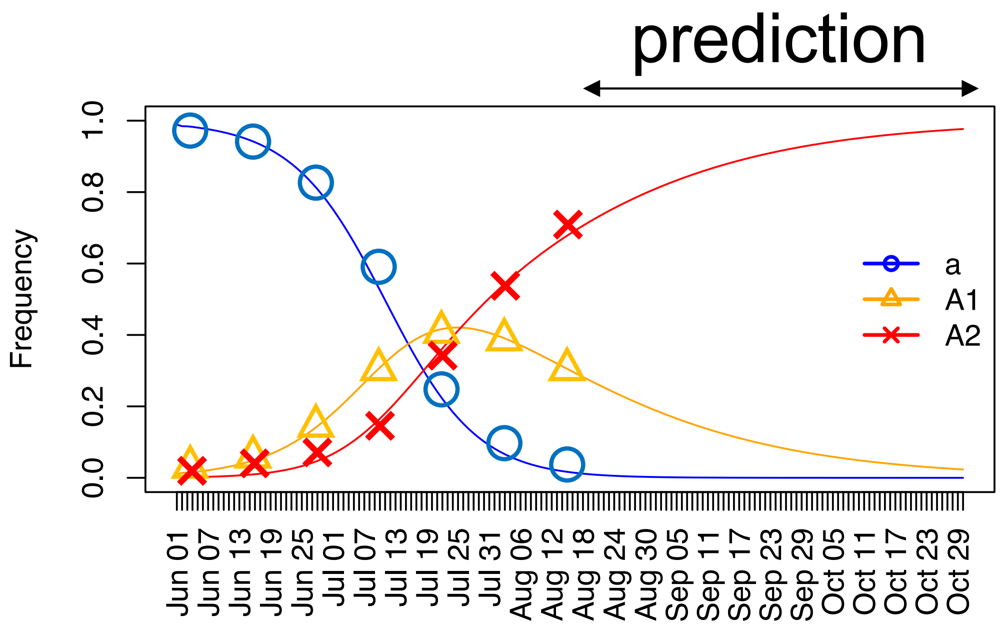



# Prediction of Frequencies of Variants in the futre

- Substituting parameters in the model with the maximum likelihood estimations of parameters, we can calculate the maximum likelihood estimations of variant frequencies $q_{A_1}(t), \ldots, q_{A_n}(t)$, and at $t > t_s$.

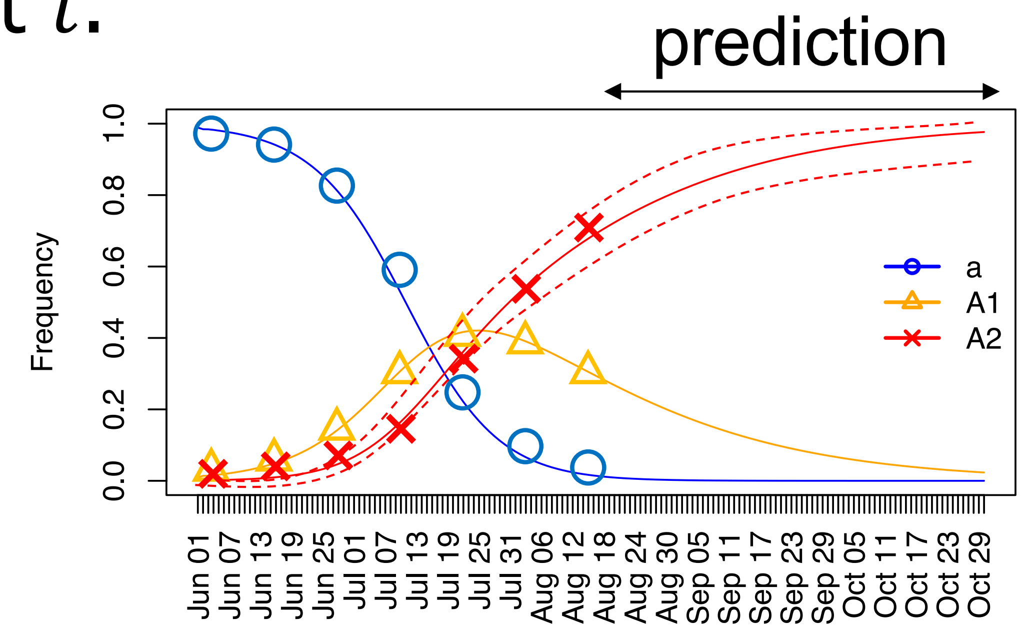
The maximum likelihood esitmation

$$(k_1, k_2, q_{A_1}(t_s), q_{A_2}(t_s)) = (1.4, 1.6, 0.01, 0.005)$$



# Estimation of Frequencies of Variants

- For each variant, we calculate the minimum and maximum of the relative frequencies at $t$ using combinations of parameters within the 95% confidence region. The minimum and maximum give the lower and upper bound of 95%CIs of relative frequencies at $t$.



# Summary

- By maximizing the likelihood function, we can estimate the relative reproduction numbers among variants.