

# Fit CII Model

William Ruth

## Analysis

In this document, we fit quasi-binomial GLMs in each of the four class size threshold groups. Note that the extreme outlier in the threshold = 100 group has been removed.

Before doing any model fitting, we plot a histogram of the CII, both globally and separately by class size threshold, where the counts are of individual runs of our simulation ( $N = 25627$ ).

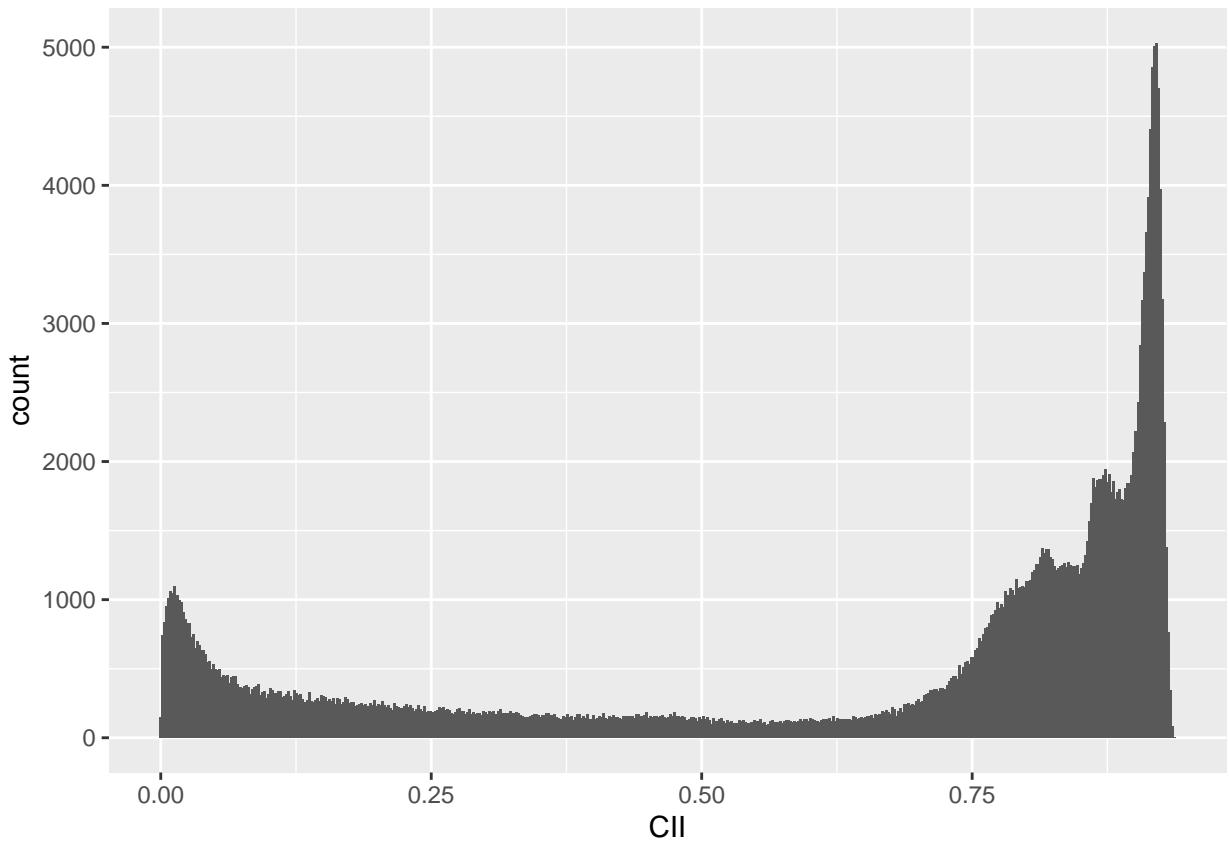


Figure 1: Histogram of CII.

There are clearly different levels of variability between the different threshold groups. This suggests that we should use an extended quasi-likelihood model with different overdispersion parameters in each group.

The structure of our simulation study includes pure replication within each parameter setting (specifically, we have 10 replicates for each parameter combination). This allows us to investigate the pure replication variability, which informs what we expect to see when we fit our GLM. The following plot gives average CII

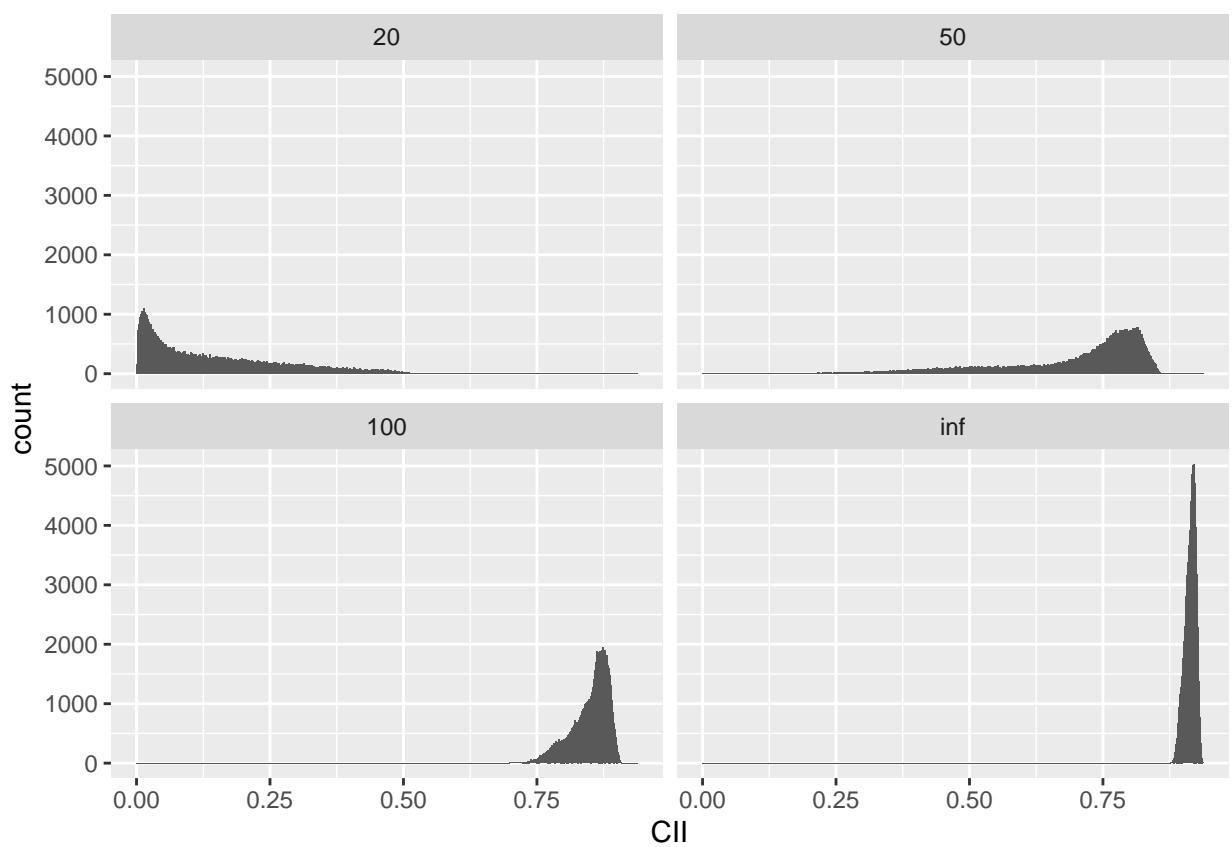


Figure 2: Histograms of CII for each class size threshold with uniform axis scaling.

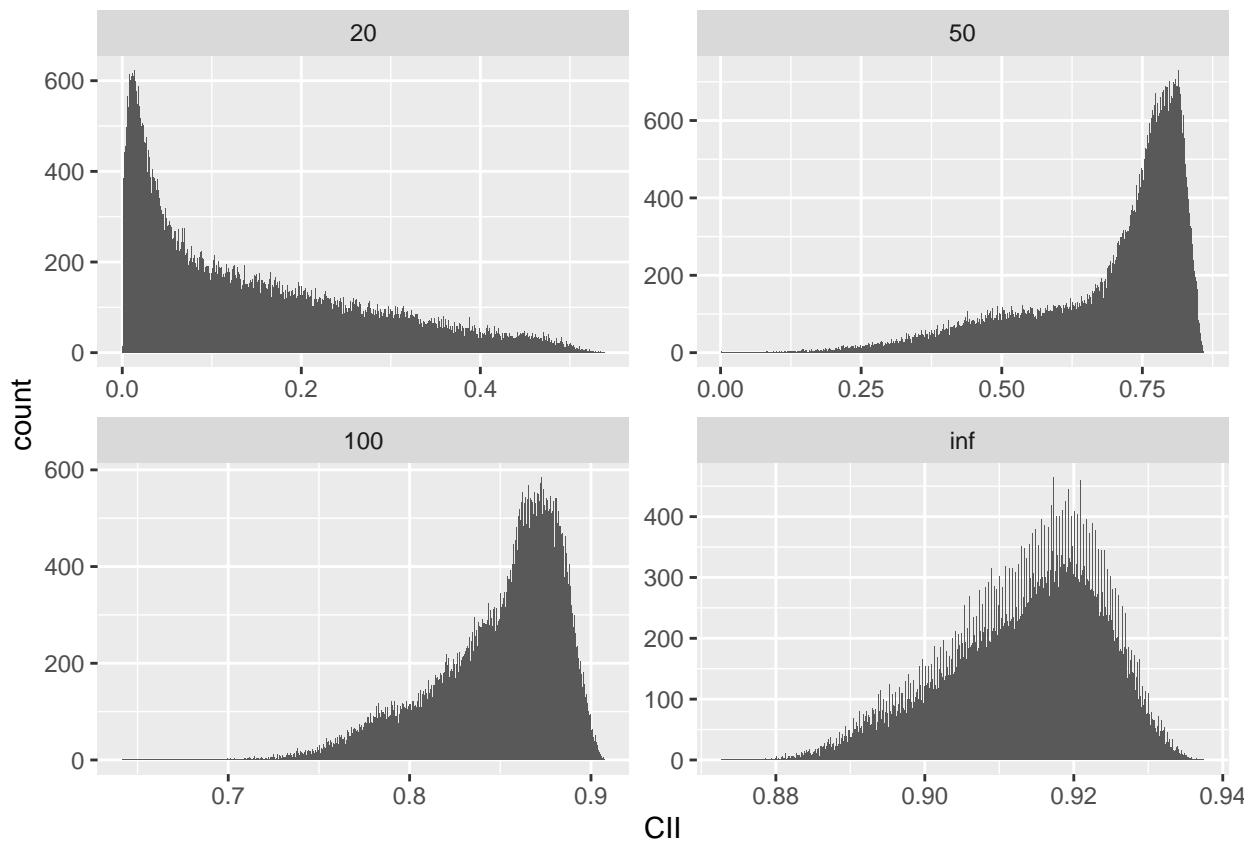
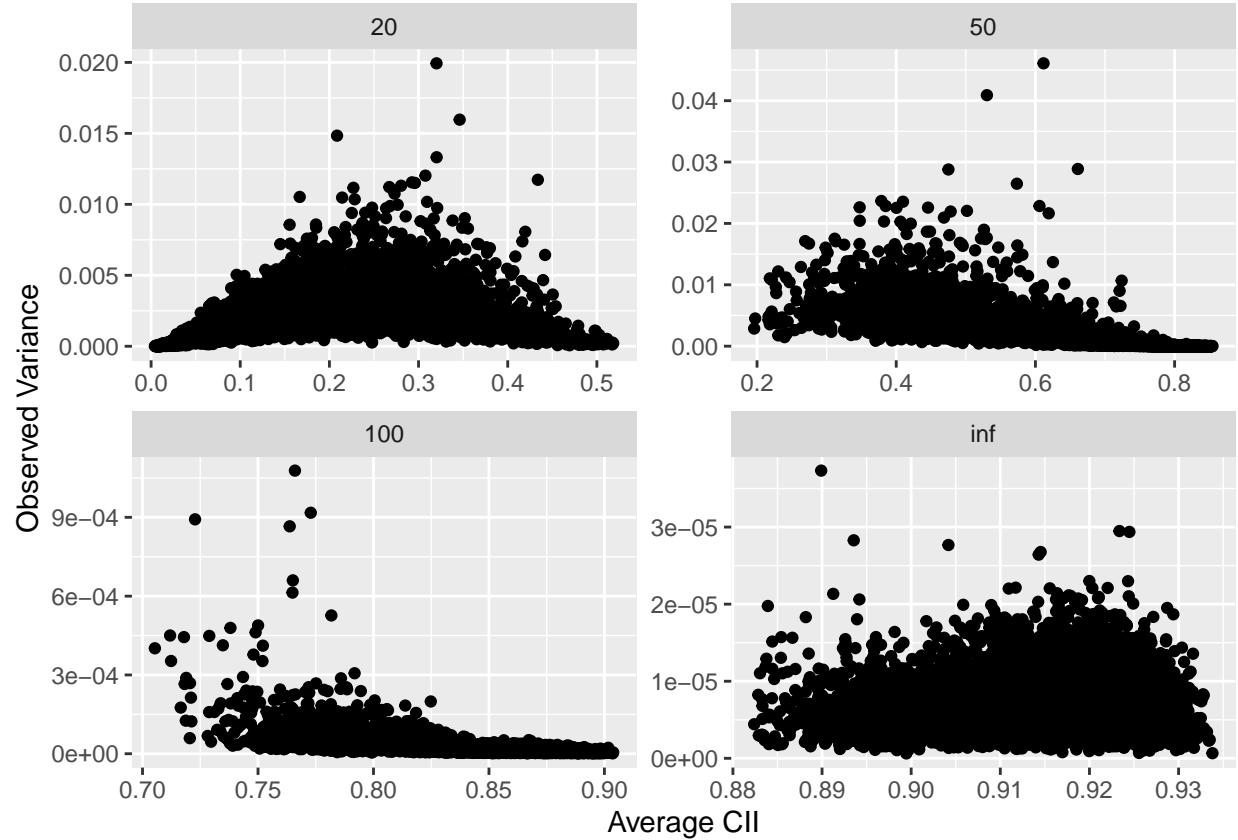
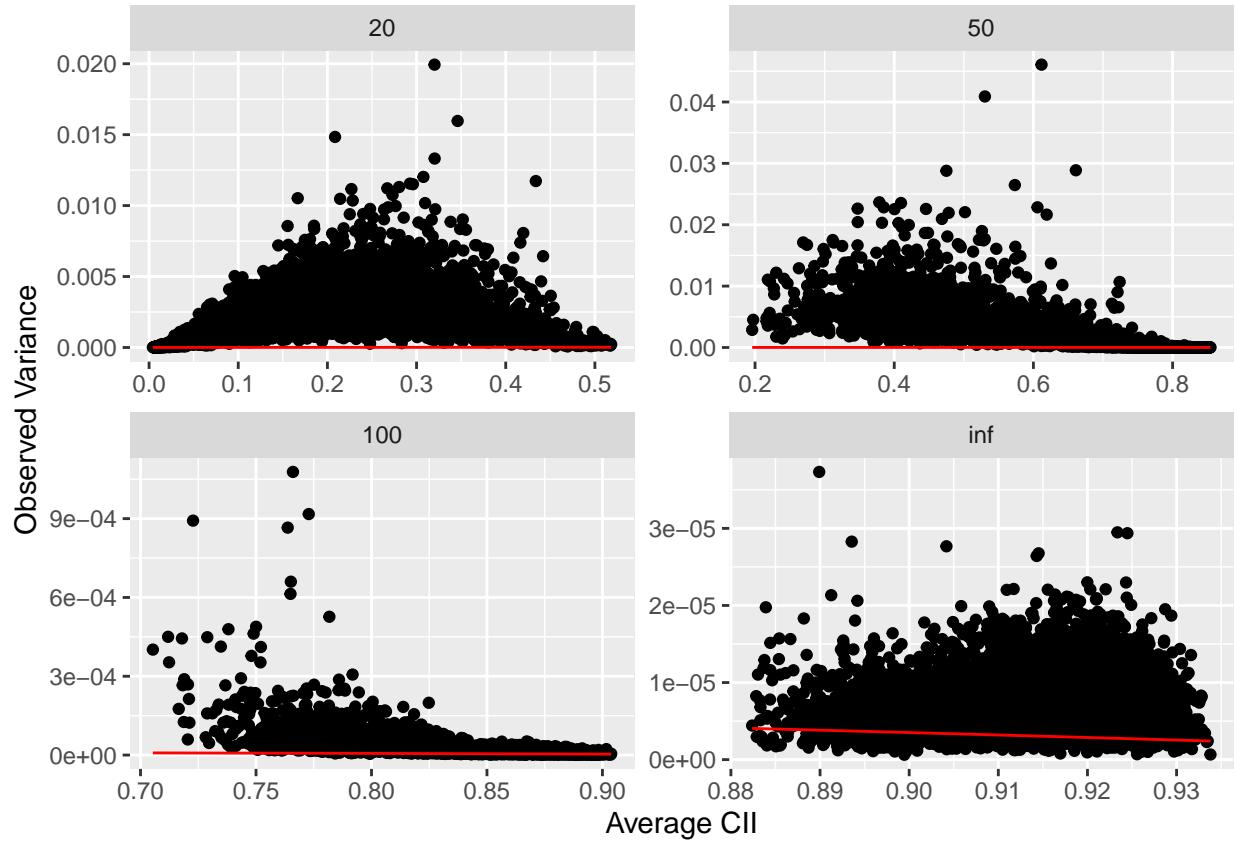


Figure 3: Histograms of CII for each class size threshold with heterogeneous axis scaling.

vs sample variance for each parameter combination (**Footnote:** the average and variance are computed over the 10 replicates within a single parameter combination).



For reference, let's repeat these plots but also add a reference line at the variance which is predicted by the binomial distribution (i.e. at  $p(1 - p)/N$ , where  $N$  is the number of students, `num_students`).



Our data clearly exhibit overdispersion relative to the predictions of a binomial model.

We now proceed to fit the models to which we have been alluding. First, the four models, with one for each level of class size threshold.

## Diagnostics

We now investigate some of the standard diagnostic plots for evaluating GLM fits. Figures 4 and 5 give the deviance and Pearson residuals respectively for each class size threshold.

## Results

We now report some summaries.

### Deviance Changes

To start, we extract the deviance improvement provided by each variable when added to a model already containing the other predictors.

Table 1: Deviance improvements within each class size threshold.

	20	50	100	inf
infect_prop_A	2.31E+06	2.04E+06	2.13E+05	3.00E+04
infect_prop_I1	3.15E+07	1.57E+07	1.37E+06	2.07E+05
infect_param_I2	1.28E+08	1.03E+08	9.93E+06	1.29E+06
advance_prob_E	7.05E+05	6.29E+05	2.20E+04	2.50E+03

	20	50	100	inf
advance_prob_A	3.52E+05	2.62E+05	3.94E+04	6.34E+03
advance_prob_I1	6.81E+06	1.64E+06	3.20E+05	6.38E+04
advance_prob_I2	1.55E+07	1.12E+07	2.12E+06	3.11E+05
E_to_A_prob	5.17E+06	3.40E+06	4.80E+05	7.27E+04

For reference, we give the relative change in deviance compared to the largest in each group.

Table 2: Relative deviance improvements within each class size threshold.

	20	50	100	inf
infect_prop_A	0.018	0.02	0.021	0.023
infect_prop_I1	0.25	0.15	0.14	0.16
infect_param_I2	1	1	1	1
advance_prob_E	0.0055	0.0061	0.0022	0.0019
advance_prob_A	0.0027	0.0025	0.004	0.0049
advance_prob_I1	0.053	0.016	0.032	0.05
advance_prob_I2	0.12	0.11	0.21	0.24
E_to_A_prob	0.04	0.033	0.048	0.056

We also rank the predictors within each group in decreasing order of deviance improvement.

Table 3: Ranked deviance improvements within each class size threshold. 1 has the greatest improvement and 8 has the least.

	20	50	100	inf
infect_prop_A	6	5	6	6
infect_prop_I1	2	2	3	3
infect_param_I2	1	1	1	1
advance_prob_E	7	7	8	8
advance_prob_A	8	8	7	7
advance_prob_I1	4	6	5	5
advance_prob_I2	3	3	2	2
E_to_A_prob	5	4	4	4

## Overdispersion Parameter

Next, we present the fitted overdispersion parameter from each group.

```
##   20    50    100   inf
## "564" "366" "18"   "4"
```

## Global Model

For reference, we repeat the above analysis on the ordinary quasi-likelihood model. That is, we retain the interaction terms for the mean model, but use a single global overdispersion parameter.

The deviance changes for excluding each variable individually from our model (while retaining all others) are as follows.

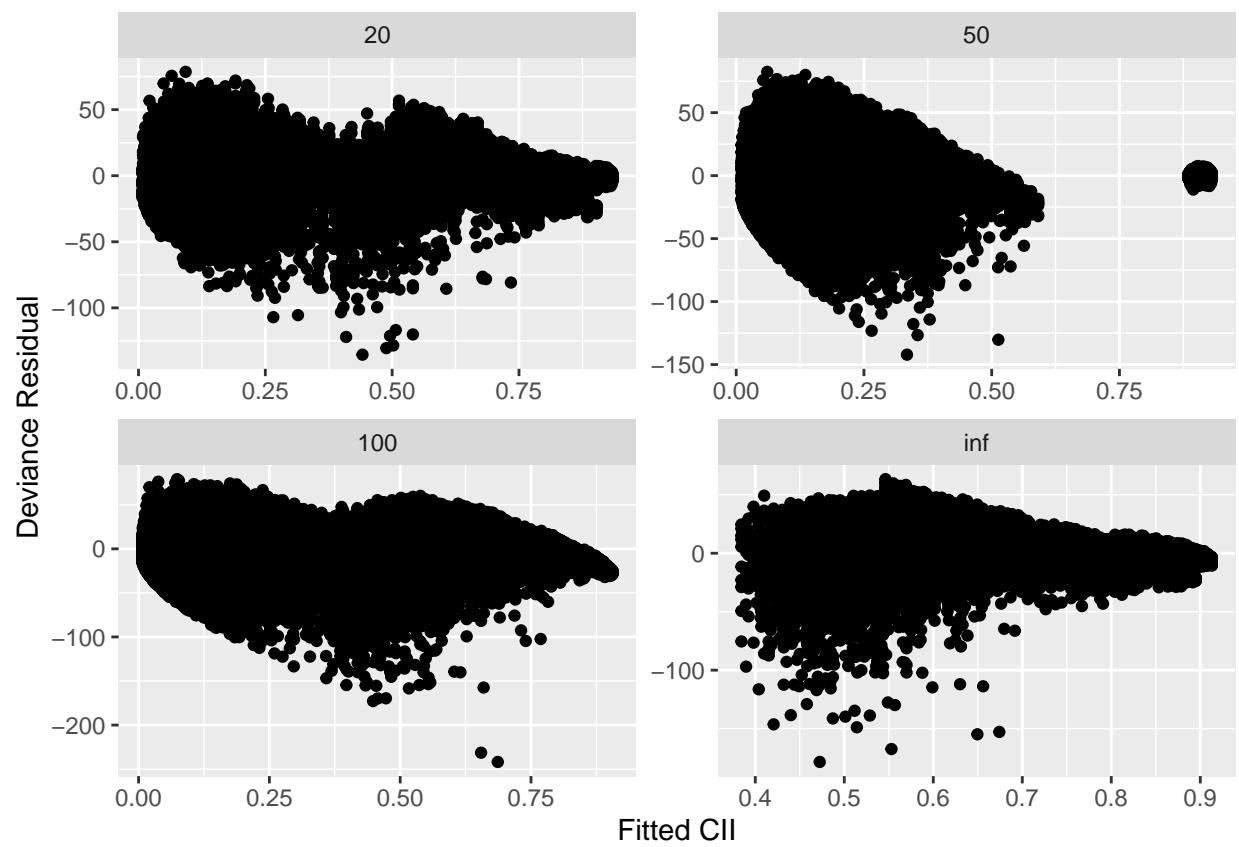


Figure 4: Deviance residuals for each class size threshold.

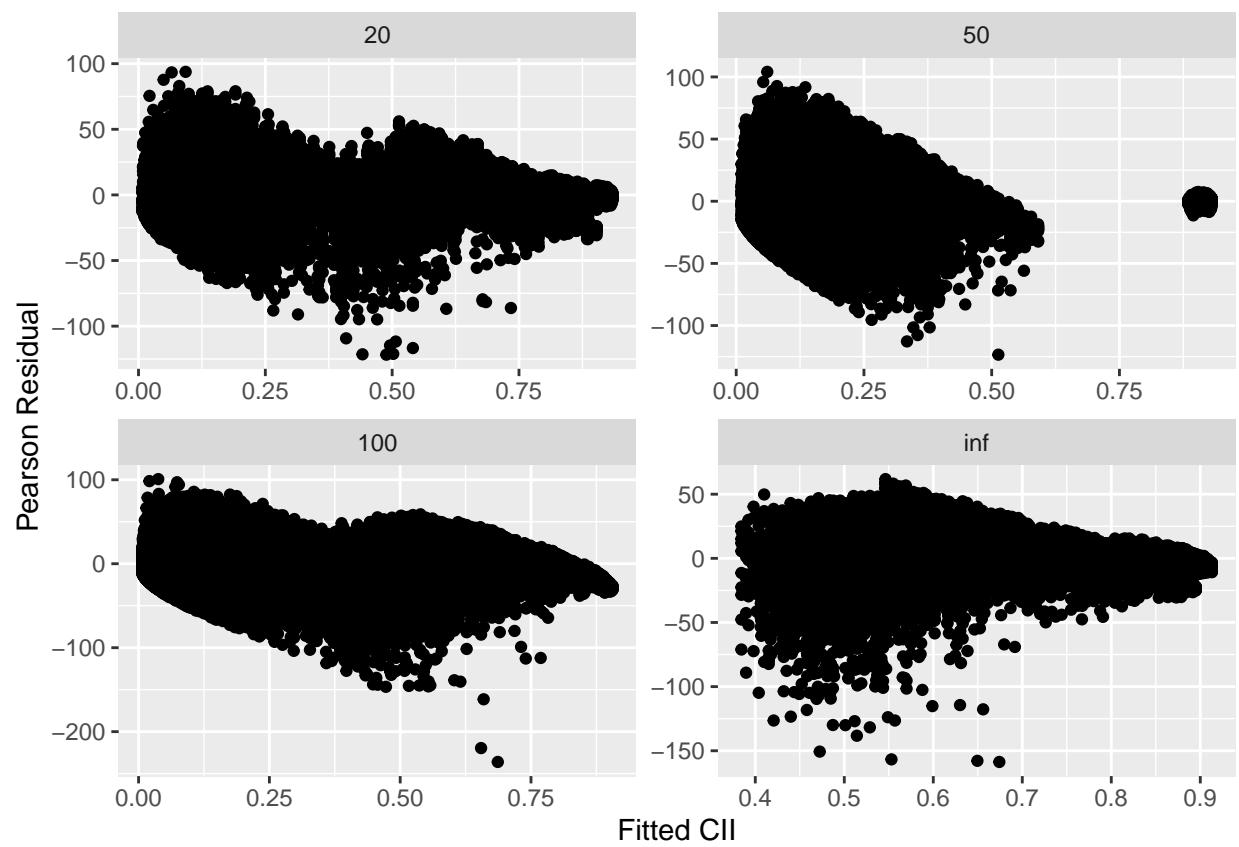


Figure 5: Pearson residuals for each class size threshold.

```

##   infect_prop_A  infect_prop_I1 infect_param_I2  advance_prob_E  advance_prob_A
## "1.04E+06"      "1.42E+07"      "6.47E+07"      "4.10E+05"      "1.38E+05"
## advance_prob_I1 advance_prob_I2    E_to_A_prob
## "3.23E+06"      "5.79E+06"      "2.11E+06"

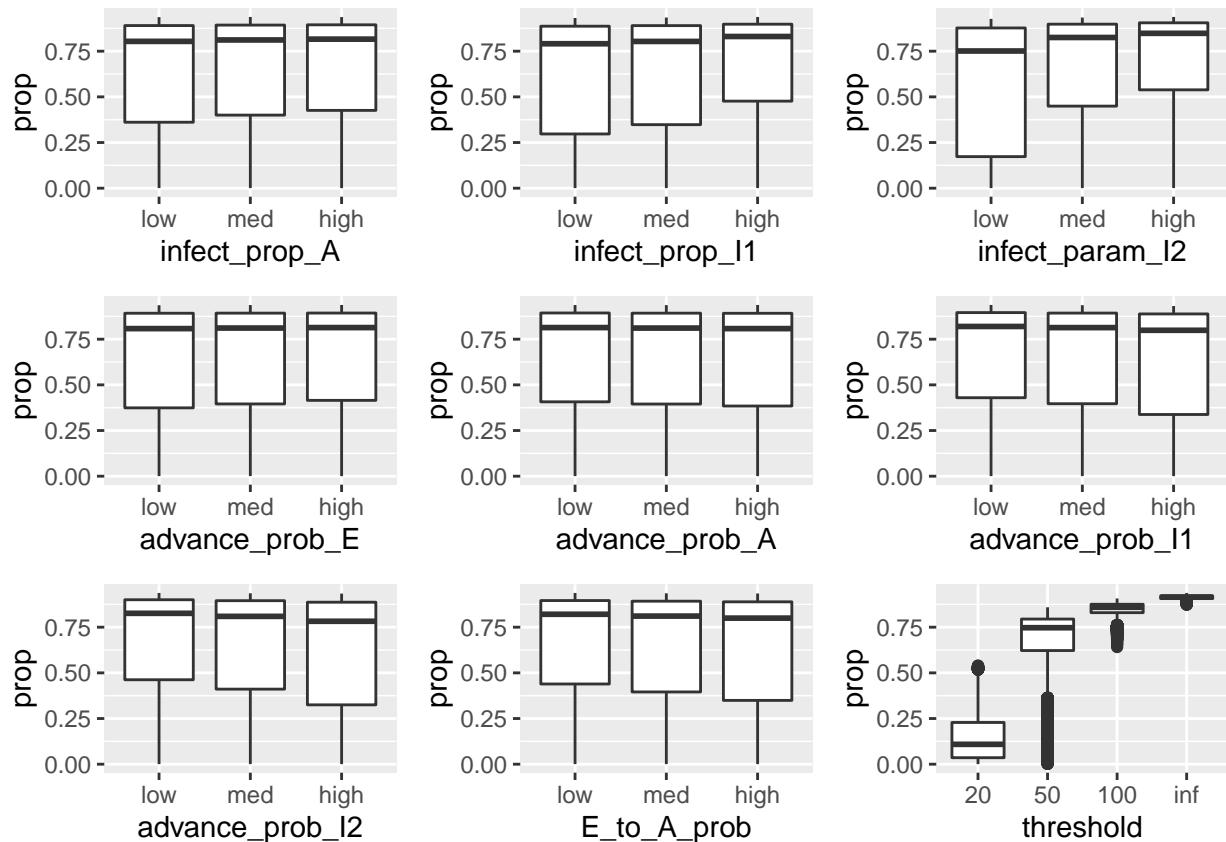
```

Similarly, the relative deviance improvements are:

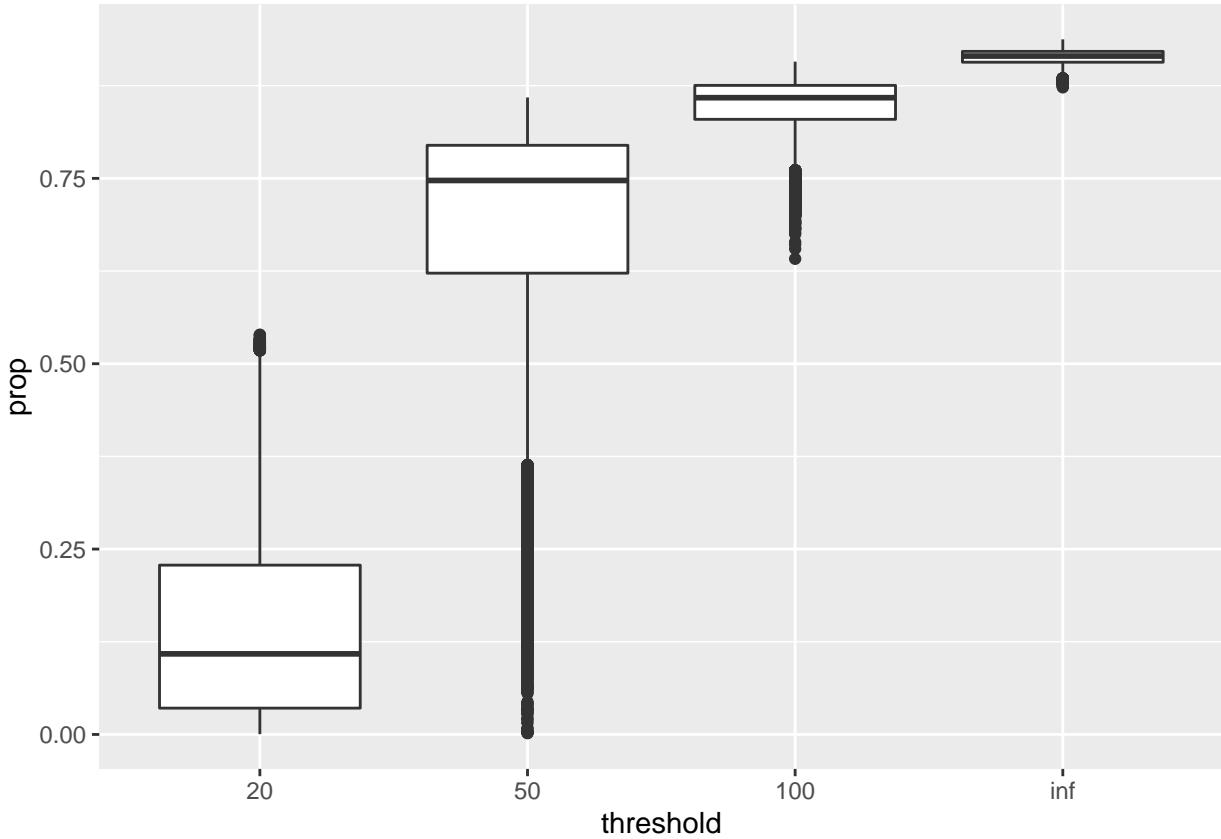
Finally, the fitted global overdispersion parameter is 238, which is the mean of the individual group models' overdispersion parameters.

## Marginal Behaviour of Each Predictor

The following figure gives a sequence of boxplots for the CII in our simulation. Each plot corresponds to a single predictor, and each box contains all simulation runs with that level of the predictor.



The threshold variable clearly has the greatest association with CII, so we reproduce that boxplot **with greater resolution**.



## Discussion

In Tables 1-3, we see that the infectiousness parameter for the symptomatic group has the largest deviance improvement in every model. Recall that the infectiousness parameter for the symptomatic group gives the absolute proportionality constant for infection probability (proportional to  $1 / \sqrt{\text{class size}}$ ). However, the parameters for the presymptomatic and asymptomatic groups are only their relative proportionality constant when compared to the symptomatic group. Thus, increasing or decreasing the infectiousness parameter for the symptomatic group changes the infection probability for all infectious compartments, while the parameters for presymptomatic and asymptomatic only affect their own compartment.

The behaviour of the threshold = 20 group is qualitatively different from the other threshold levels. In Figures 2 and 3, we see that the CIs clump around 0 with a long right tail when threshold is 20, while the other groups have their mass concentrated above 0.5.