

Scorestatistik for local Alignments

Utz J. Pape

November 1, 2004

1 Score Histogram

1.1 Simulation

In figure 1, we see the histogram of scores of 1000 sequence alignment of sequences with length 500.

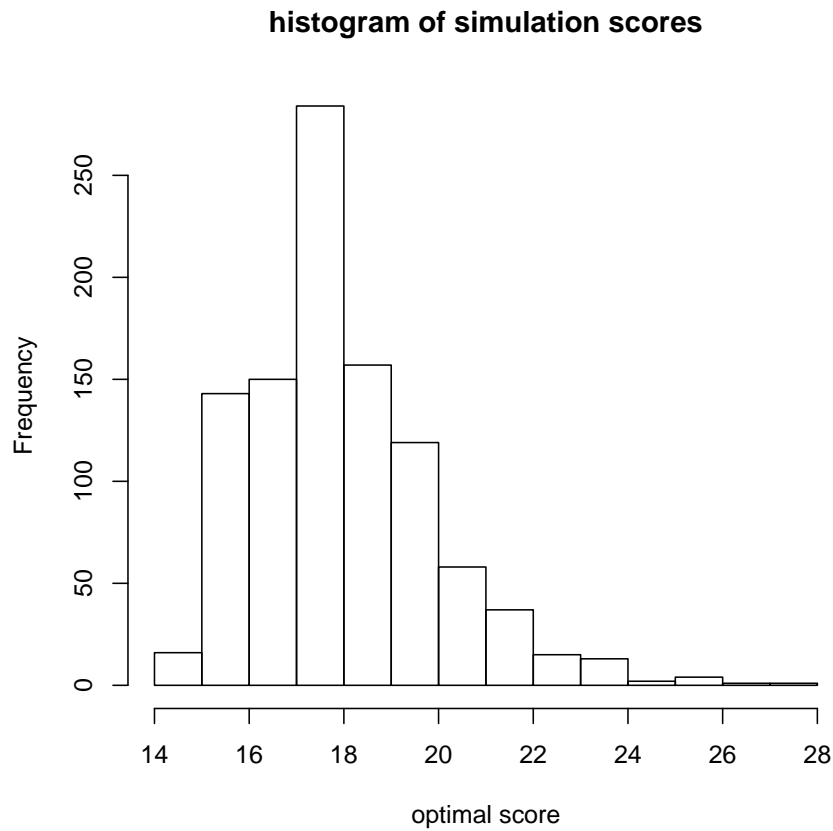


Figure 1: Histogram of null model scores

1.2 Fitting Normal Distribution

Figure 2 shows the fitted normal distribution. The parameters have been estimated with $\mu = 18.45$ and $\sigma = 1.94917913473834$.

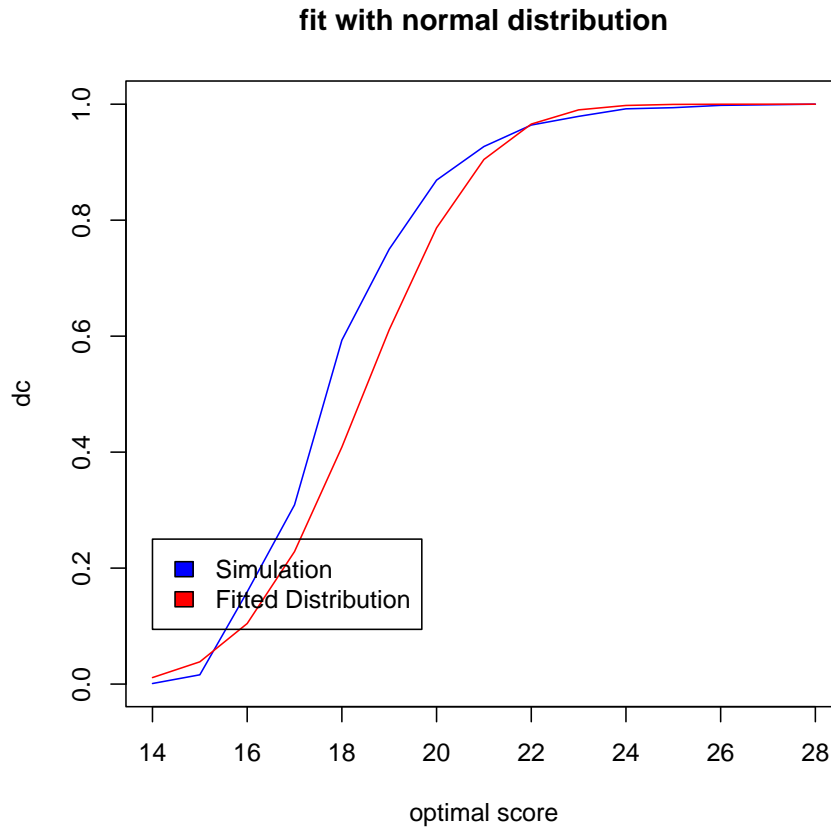


Figure 2: Fitting Normal Distribution

1.3 Fitting Extrem Value Distribution

Figure 3 shows the fitted extreme value distribution, figure 4 shows the same plot where the distributions are two times logarithmic transformed. The parameters have been estimated with $\theta = 1.51976873638691$ and $\xi = 17.5730934391048$.

2 Conserved Sequence

The score of the conserved sequences is 471. The p-value for this score computed by the simulation scores is $<2e-16$. If we use the fitted normal distribution we get a p-value of $<2e-16$. The fit with the extreme value distribution retrieves a p-value of $<2e-16$.

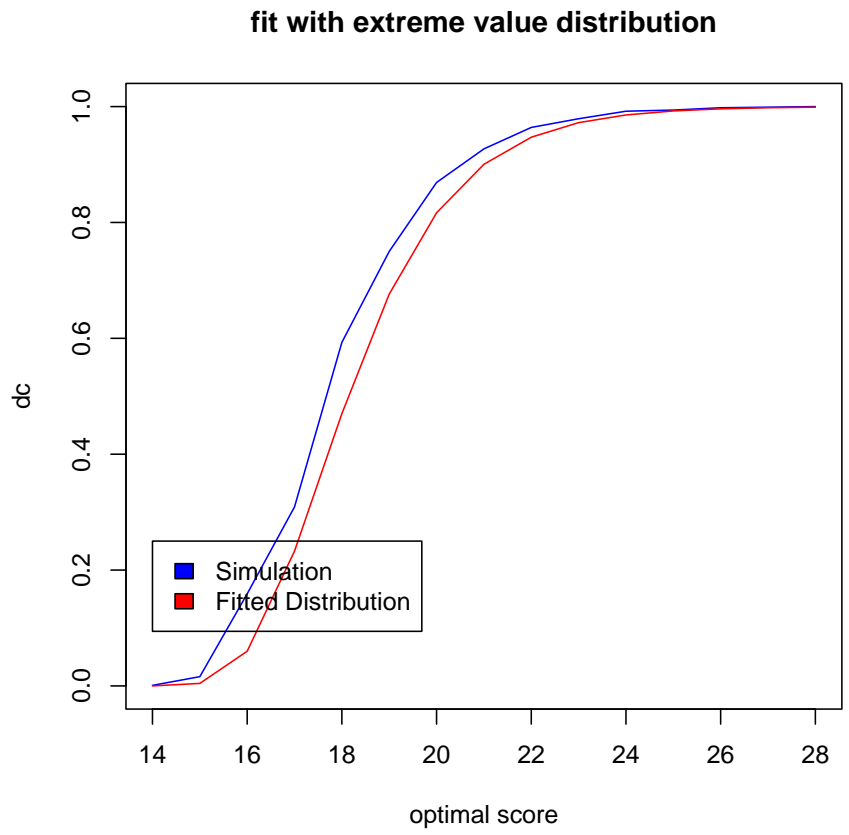


Figure 3: Fitting Extreme Value Distribution

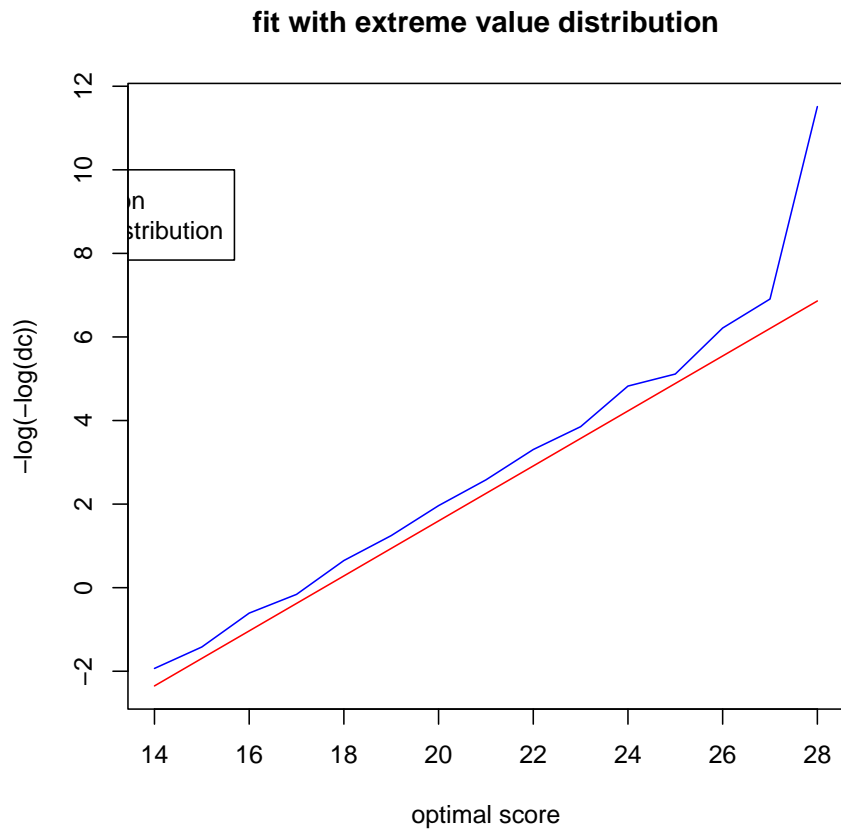


Figure 4: Log-log-Plot: Fitting Extreme Value Distribution