APPENDIX F

This appendix describes the possibility of using standard maximum likelihood estimation together with an overdispersed discrete distribution other than the negative binomial distribution. To derive this distribution, the Poisson distribution with null dispersion, which is given by

$$f_P\ (x/k) = \frac{k^{x} e^{-k}}{x!} \quad ; \quad f\colon\ \mathbb{N}\to\mathbb{R} \tag{F.1}$$

where the only parameter k represents both the mean and the variance, is used as a starting point. To extend the tail to the right, a second parameter $\alpha \in [0,1]$ can be introduced as the exponent of the denominator. For $\alpha<1$, the mean of the new distribution will then be smaller than the variance. But to describe a proper probability function, the sum over all observations must be equal to 1, which makes it necessary to weigh the expression by the sum of all of its values for all x. The new distribution with two parameters is then given by

$$f(x/k,\alpha) = \frac{\dfrac{k^{x} e^{-k}}{(x!)^{\alpha}}}{\displaystyle\sum_{x=0}^{\infty} \frac{k^{x} e^{-k}}{(x!)^{\alpha}}} \quad ; \quad f\colon\ \mathbb{N}\to\mathbb{R} \quad , \tag{F.2}$$

where the length of the right tail, which describes the degree of overdispersion, increases as α decreases.

With increasing k and small α it is very time consuming to calculate the probability values during the optimization of the parameters. To reduce the optimization time it is helpful to create a large table that containes the denominator in equation F.2 for various combinations of k and α, and to interpolate the value for a specific combination of k and α from that table. A similar table can be created for all other necessary distribution values, for example for the mean and for the standard deviation.

While the procedure proved to be too time consuming in the current analysis with the current computer equipment (75 Mhz Pentium), it might still be useful for other analyses that use a smaller data set, or when the available computer is fast enough. In addition, it shows that there is a possibility to avoid both the Poisson and the negative binomial distributions if the data does not seem to fit either one.