Supplementary material

The following theorem (see Mood, 1974) is the key to the observation 2:

Theorem: If $\Theta_n = \theta_n(X_1, \ldots, X_n)$ is the maximum likelihood estimator of θ for a random sample of size n from the density $f(x, \theta)$, then Θ_n is approximately normally distributed with the mean θ and variance

$$1/n\varepsilon_\theta\left[\left[\frac{\partial}{\partial\theta}\log f(x,\theta)\right]^2\right]. \qquad 12$$

Observe that, from the equation 1, it follows that

$$P(X = a) = \frac{e^{-E}E^a}{a!} = \frac{e^{-Kmne^{-\lambda c}}(Kmne^{-\lambda c})^a}{a!}. \qquad 13$$

Thus, the theorem implies that $\hat{K}$ is normally distributed with mean K and variance

$$1/N\varepsilon\left[\left(\frac{\partial}{\partial K}\log\frac{e^{-Kmne^{-\lambda c}}(Kmne^{-\lambda c})^a}{a!}\right)^2\right] = \frac{K}{Nmne^{-\lambda c}} \qquad 14$$

The same argument can be applied to derive the parameters for the distribution of the maximum likelihood estimates of λ (Altschul *et al.*, 2001).

The inequality 11 follows from the fact that the expected number of locally optimal sub-alignments with score $\geq c$ is strictly less than the size of the dynamic programming matrix, i.e. $E = Kmne^{-\lambda c} < mn$.