Finding the MLE of the F81 model Consider a dataset consisting of 18-base long sequence from a single species with: $n_A = 8$, $n_C = 1$, $n_G = 3$, and $n_T = 6$. The bases are not equally frequent in this sample, but can we reject the null that this sample was drawn from a model in which all of the base frequencies are 0.25?

The Jukes-Cantor model assumes that the equilibrium base frequencies are all equal: $\pi_A = \pi_C = \pi_G = \pi_T = 1/4$.

Felsenstein's 1981 model allows the equilibrium base frequencies to be free parameters estimated from the data. If $2\left[\ln(L(F81)) - \ln(L(JC))\right] \geq 7.8147$, then we can reject the null that the sample comes from the JC model. Where does "7.814" come from? That is the $\alpha = 0.05$ critical value from the χ^2_3 distribution (the "chi-square distribution with 3 degrees for freedom).

Where does $df = 3$ come from? When comparing a simple model to a richer model that contains it (the simple model is nested inside the richer model by making constraints on the parameters of the rich model), then twice the log of the likelihood ratio (the LRT statistic) is distributed as a chi-squared variate with the degrees of freedom equal to the number of parameters that are free to vary in the rich model, but constrained in the simple model.

π_A, π_C, and π_G are free to vary to fit the data in the F81 model. Once you know those three, you know $\pi_T = 1 - \pi_A - \pi_C - \pi_G$. So the difference in the number of free parameters is 3 for this model choice problem.

It is important that we compare the maximized likelihoods when calculating this LRT statistic.

So how do we know that we have the maximize likelihood?

$$\begin{aligned}
L(F81) = \Pr(D \mid F81, \pi_A, \pi_C, \pi_G) &= \prod_{i=1}^{18} \Pr(D_i \mid F81, \pi_A, \pi_C, \pi_G) \\
&= \Pr(A|\pi_A)^{n_A} \Pr(C|\pi_C)^{n_C} \Pr(G|\pi_G)^{n_G} \Pr(T|\pi_A, \pi_C, \pi_G)^{n_T} \\
&= \pi_A^{n_A} \pi_C^{n_C} \pi_G^{n_G} (1 - \pi_A - \pi_C - \pi_G)^{n_T} \\
\ln L(F81) &= \ln\left[\pi_A^{n_A}\right] + \ln\left[\pi_C^{n_C}\right] + \ln\left[\pi_G^{n_G}\right] + \ln\left[(1 - \pi_A - \pi_C - \pi_G)^{n_T}\right] \\
&= n_A \ln \pi_A + n_C \ln \pi_C + n_G \ln \pi_G + n_T \ln\left[(1 - \pi_A - \pi_C - \pi_G)\right]
\end{aligned}$$

To find the maximum likelihood estimates (MLEs) of the parameters we take the derivatives of the likelihood (or log-likelihood) functions with respect to the parameters and then we solve for the parameter values that cause these derivative to be 0.

It is often easier to do this on the log scale. Recall that:

$$\frac{d \ln[f(u)]}{du} = \left(\frac{1}{f(u)}\right)\left(\frac{df(u)}{du}\right)$$

So:

$$\begin{aligned}
\frac{\partial \ln L(F81)}{\partial \pi_A} &= \frac{n_A}{\pi_A} + \left(\frac{n_T}{1-\pi_A-\pi_C-\pi_G}\right)\left(\frac{\partial(1-\pi_A-\pi_C-\pi_G)}{\partial \pi_A}\right) \\
&= \frac{n_A}{\pi_A} + \left(\frac{n_T}{1-\pi_A-\pi_C-\pi_G}\right)(-1) \\
&= \frac{n_A}{\pi_A} - \frac{n_T}{1-\pi_A-\pi_C-\pi_G} \\
&= \frac{n_A}{\pi_A} - \frac{n-n_A-n_C-n_G}{1-\pi_A-\pi_C-\pi_G}
\end{aligned}$$

Similarly:

$$\begin{aligned}
\frac{\partial \ln L(F81)}{\partial \pi_C} &= \frac{n_C}{\pi_C} - \frac{n-n_A-n_C-n_G}{1-\pi_A-\pi_C-\pi_G} \\
\frac{\partial \ln L(F81)}{\partial \pi_G} &= \frac{n_G}{\pi_G} - \frac{n-n_A-n_C-n_G}{1-\pi_A-\pi_C-\pi_G}
\end{aligned}$$

It is a bit cumbersome to go through all of the steps, but note that if we choose the following estimates:

$$\begin{aligned}
\hat{\pi}_A &= \frac{n_A}{n} \\
\hat{\pi}_C &= \frac{n_C}{n} \\
\hat{\pi}_G &= \frac{n_G}{n}
\end{aligned}$$

We find that all 3 of the derivatives are 0 for this choice of estimators:

$$\begin{aligned}
\frac{\partial \ln L(F81)}{\partial \pi_A} &= \frac{n_A}{\hat{\pi}_A} - \frac{n-n_A-n_C-n_G}{1-\hat{\pi}_A-\hat{\pi}_C-\hat{\pi}_G} \\
&= \frac{n_A}{\frac{n_A}{n}} - \frac{n-n_A-n_C-n_G}{1-\frac{n_A}{n}-\frac{n_C}{n}-\frac{n_G}{n}} \\
&\quad \text{mulitply numerator and denominator by } n \\
&= \frac{nn_A}{n_A} - \frac{n(n-n_A-n_C-n_G)}{n-n_A-n_C-n_G} \\
&= n-n=0
\end{aligned}$$

Note that it is often easier to find the MLE by reparameterizing. A+G are the purines (denoted R) and C+T are the pyrimidines (denoted Y). So we could express the F81 models as:

$$\begin{aligned}
\pi_R &= \Pr(A \text{ or } G) \\
\pi_{AGR} &= \Pr(A \mid R) \\
\pi_{CGY} &= \Pr(C \mid Y) \\
\pi_A &= \pi_R \pi_{AGR} \\
\pi_G &= \pi_R(1 - \pi_{AGR}) \\
\pi_C &= (1 - \pi_R)\pi_{CGY} \\
\pi_T &= (1 - \pi_R)(1 - \pi_{CGY})
\end{aligned}$$

In this parameterization:

$$\begin{aligned}
\ln L(F81) &= n_A \ln[\pi_R \pi_{AGR}] + n_C \ln[(1 - \pi_R)\pi_{CGY}] + n_G \ln[\pi_R(1 - \pi_{AGR})] + n_T \ln[(1 - \pi_R)(1 - \pi_{CGY})] \\
&= n_A \ln \pi_R + n_A \ln \pi_{AGR} + n_C \ln(1 - \pi_R) + n_C \ln \pi_{CGY} + \ldots \\
&\quad \ldots + n_G \ln \pi_R + n_G \ln(1 - \pi_{AGR}) + n_T \ln(1 - \pi_R) + n_T \ln(1 - \pi_{CGY}) \\
&= (n_A + n_G) \ln \pi_R + (n_C + n_T) \ln(1 - \pi_R) + \ldots \\
&\quad \ldots + n_A \ln \pi_{AGR} + n_G \ln(1 - \pi_{AGR}) + n_C \ln \pi_{CGY} + n_T \ln(1 - \pi_{CGY})
\end{aligned}$$

This does not look easier until we differentiate with respect to our new parameters, and solve each to find the value that gives a derivative of 0:

$$\begin{aligned}
\frac{\partial \ln L(F81)}{\partial \pi_R} &= \frac{n_A + n_G}{\pi_R} - \frac{n_C + n_T}{1 - \pi_R} \\
0 &= \frac{n_A + n_G}{\hat{\pi}_R} - \frac{n_C + n_T}{1 - \hat{\pi}_R} \\
\frac{n_C + n_T}{1 - \hat{\pi}_R} &= \frac{n_A + n_G}{\hat{\pi}_R} \\
(n_C + n_T)\hat{\pi}_R &= (n_A + n_G)(1 - \hat{\pi}_R) \\
(n_C + n_T)\hat{\pi}_R &= (n_A + n_G) - (n_A + n_G)\hat{\pi}_R \\
(n_A + n_G + n_C + n_T)\hat{\pi}_R &= (n_A + n_G) \\
\hat{\pi}_R &= (n_A + n_G)/(n_A + n_G + n_C + n_T) = (n_A + n_G)/n
\end{aligned}$$

Similar arguments from:

$$\begin{aligned}
\frac{\partial \ln L(F81)}{\partial \pi_{AGR}} &= \frac{n_A}{\pi_{AGR}} - \frac{n_G}{1 - \pi_{AGR}} \\
\frac{\partial \ln L(F81)}{\partial \pi_{CGY}} &= \frac{n_C}{\pi_{CGY}} - \frac{n_T}{1 - \pi_{CGY}}
\end{aligned}$$

lead to the MLEs:

$$\begin{aligned}
\hat{\pi}_{AGR} &= n_A/(n_A + n_G) \\
\hat{\pi}_{CGY} &= n_C/(n_C + n_T)
\end{aligned}$$

In terms of the orginal parameterization:

$$\begin{aligned}
\hat{\pi}_A &= \hat{\pi}_R\hat{\pi}_{AGR} = \left(\frac{n_A + n_G}{n}\right)\left(\frac{n_A}{n_A + n_G}\right) = \frac{n_A}{n} \\
\hat{\pi}_G &= \hat{\pi}_R(1 - \hat{\pi}_{AGR}) = \left(\frac{n_A + n_G}{n}\right)\left(\frac{n_G}{n_A + n_G}\right) = \frac{n_G}{n} \\
\hat{\pi}_C &= (1 - \hat{\pi}_R)\hat{\pi}_{CGY} = \left(\frac{n_C + n_T}{n}\right)\left(\frac{n_C}{n_C + n_T}\right) = \frac{n_C}{n} \\
\hat{\pi}_T &= (1 - \hat{\pi}_R)(1 - \hat{\pi}_{CGY})\left(\frac{n_C + n_T}{n}\right)\left(\frac{n_T}{n_C + n_T}\right) = \frac{n_T}{n}
\end{aligned}$$

This is just one demonstration of a general property of ML: it is invariant to reparameterization. We can reparameterize whenever we feel like it to make it easier to find MLEs. We can simply invert the transformation involved in the reparameterization to get the MLEs in terms of the original parameterization. The maximized likelihood will be the same in either parameterization.