1 Dirichlet-Multinomial posterior predictive

Let $\alpha'_j = \alpha_j + \sum_{y_i \in D} y_i^{(j)}$. The set of α'_j has the same dimensionality as the hyperparameters α_j and needs to be associated with each node in the dendrogram.

The posterior predictive is:[1]

$$f(y \mid D) = \frac{\Gamma(n+1)}{\prod_{j=1}^{K} \Gamma(y^{(j)}+1)} \frac{\Gamma\left(\sum_{j=1}^{K} \alpha'_j\right)}{\prod_{j=1}^{K} \Gamma(\alpha'_j)} \frac{\prod_{j=1}^{K} \Gamma(y^{(j)}+\alpha'_j)}{\Gamma\left(n+\sum_{j=1}^{K} \alpha'_j\right)}$$

We want the log-likelihood:

$$\begin{aligned}
\log f(y \mid D) = \log \Gamma(n+1) - \log \prod_{j=1}^{K} \Gamma(y^{(j)}+1) \\
&+ \log \Gamma\left(\sum_{j=1}^{K} \alpha'_j\right) - \log \prod_{j=1}^{K} \Gamma(\alpha'_j) \\
&+ \log \prod_{j=1}^{K} \Gamma(y^{(j)}+\alpha'_j) - \log \Gamma\left(n+\sum_{j=1}^{K} \alpha'_j\right) \\
&= \log \Gamma(n+1) - \sum_{j=1}^{K} \log \Gamma(y^{(j)}+1) \\
&+ \log \Gamma\left(\sum_{j=1}^{K} \alpha'_j\right) - \sum_{j=1}^{K} \log \Gamma(\alpha'_j) \\
&+ \sum_{j=1}^{K} \log \Gamma(y^{(j)}+\alpha'_j) - \log \Gamma\left(n+\sum_{j=1}^{K} \alpha'_j\right)
\end{aligned}$$

The node-dependent values in this function are n and α'_j. Both of these need to be computed and kept as an attribute for each node. There are further optimizations possible, e.g. by caching sums of α'_j.

[1] `https://people.eecs.berkeley.edu/~stephentu/writeups/dirichlet-conjugate-prior.pdf`