Same setup as in Felsenstein.

Outline

Input:T_σ, $\mathcal{D}$
Output: All marginal distributions $P(V_u|\mathcal{D},T)$ for $u \in V \setminus L(T)$

This requires two passes of "message passing" (see N. Friedman Probabilistic graphical models: Principles and techniques, D Koller - 2009 - MIT Press: Cambridge, MA, USA).

Message

The messages and the DP algorithm to compute them is closely related to Felsenstein.

Consider any tree edge $B \to A$ between inner nodes A and B.

Let $\mathcal{D}_{|u}$ for any $u \in V \setminus L(T)$ denote the data restricted to leaves below u.
Let $\mathcal{D}_{|-u}$ for any $u \in V \setminus L(T)$ denote the data restricted to all other leaves not below u.

Upward message:

$$\begin{aligned}\delta\uparrow_{A,B} = &\sum_a P(A=a|B) \prod_{C\in child(A)} \delta\uparrow_{C,A=a} \\ &= \sum_a P(A=a|B)\alpha(A,a) \\ &= P(D_{|A}|B)\end{aligned}$$

with $\delta\uparrow_{A,B=b} = P(D_{|A}|B=b)$. If $P_{B,A}$ is the transition matrix along edge $B \to A$ then $\delta\uparrow_{A,B} = P_{B,A}\alpha(A,\cdot)$.
Initial case (where A is leaf): $\delta\uparrow_{A,B} = P_{B,A}\mathcal{D}_A$.

Downward message:

$$\begin{aligned}\delta\downarrow_{B,A} = &\left(\prod_{C\in child(B)\setminus\{A\}} \delta\uparrow_{C,B=b}\right) \sum_p P(B|parent(B)=p)\delta\downarrow_{parent(B)=p,B} \\ &= \frac{\alpha(B,\cdot)}{\delta\uparrow_{A,B}} \sum_p P(B|parent(B)=p)\delta\downarrow_{parent(B)=p,B} \\ &= \frac{\alpha(B,\cdot)}{\delta\uparrow_{A,B}} P_{parent(B),B}{}^T \delta\downarrow_{parent(B),B} \\ &= P(D_{-A},B)\end{aligned}$$

with $\delta\downarrow_{B=b,A} = P(D_{-A},B=b)$.
Initial case (where B is root): $\delta\downarrow_{B,A} = P(B)\frac{\alpha(B,\cdot)}{\delta_{A,B}}$.
with α computed by Felsenstein.

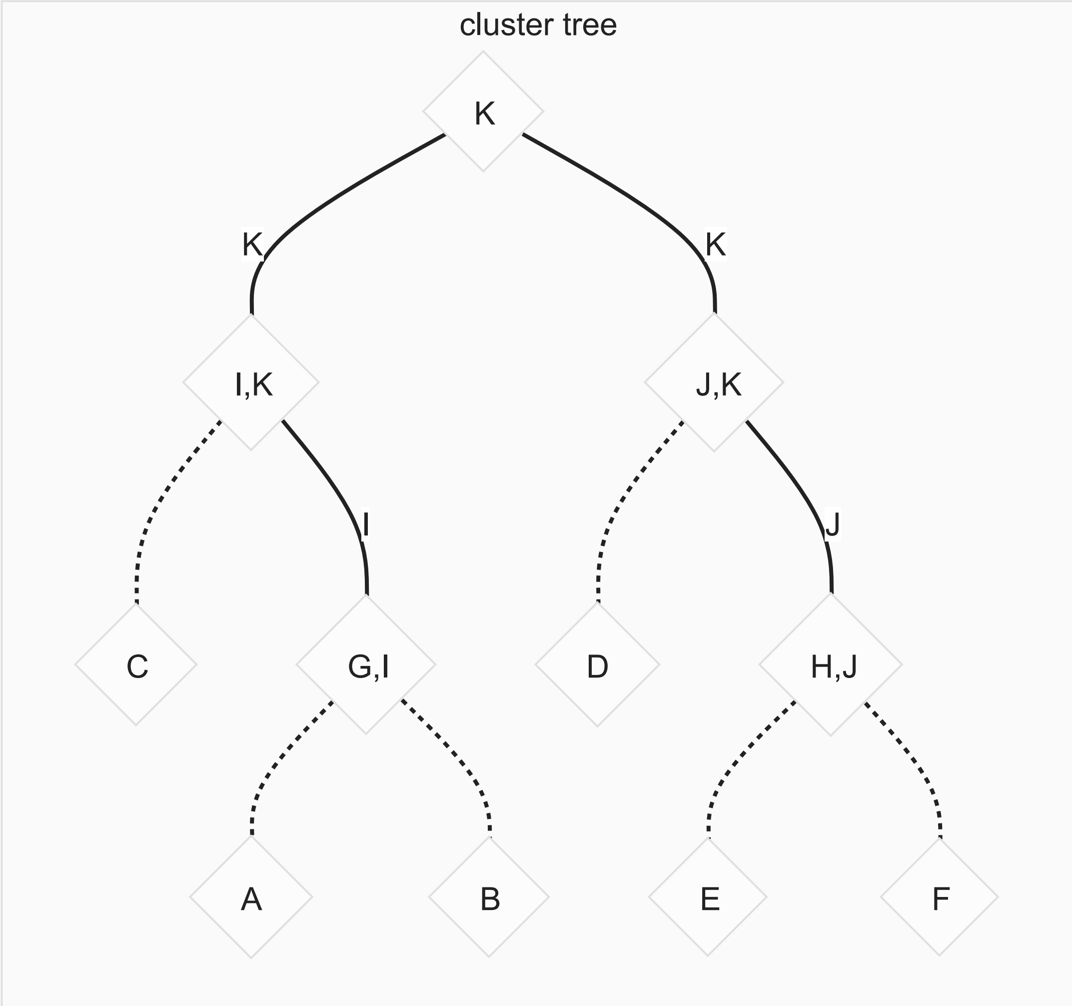
Beliefs:

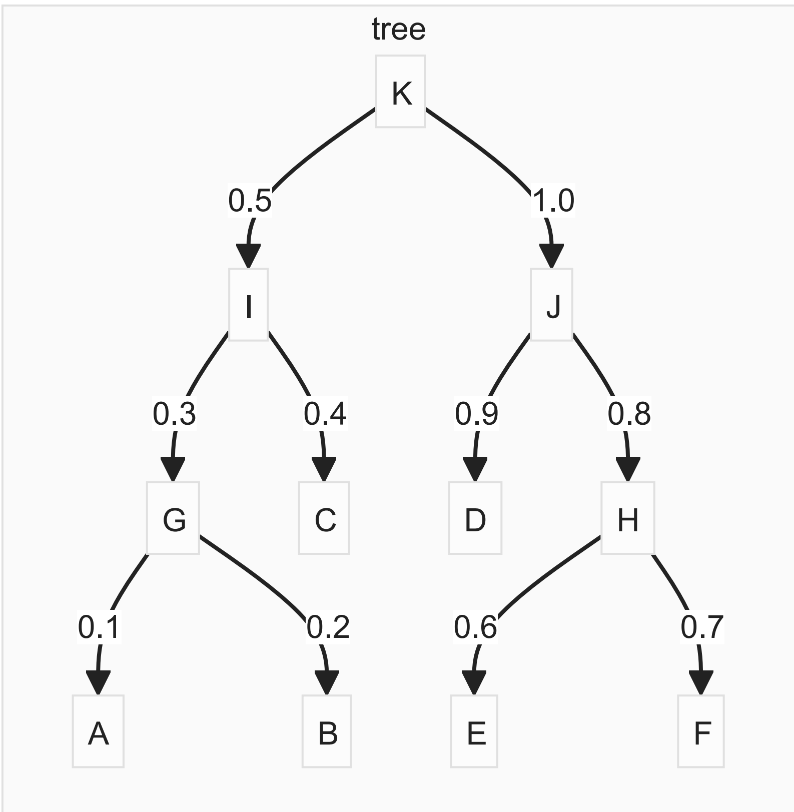
$$\begin{aligned}\beta_X = &\left(\prod_{C\in child(X)} \delta\uparrow_{C,X}\right) \sum_p P(X|parent(X)=p)\delta\downarrow_{parent(X)=p,X} \\ &= \alpha(X,\cdot)\sum_p P(X|parent(X)=p)\delta\downarrow_{parent(X)=p,X} \\ &= \alpha(X,\cdot)P_{parent(X),X}{}^T\delta\downarrow_{parent(X),X} \\ &= P(\mathcal{D},X)\end{aligned}$$

for $X \neq root$ and $\beta_{root} = \alpha(root,\cdot)\pi$.

Example

Distribution $P(A,B,C,D,E,F,G,H,I,J,K) = P(K)P(I|K)P(J|K)P(C|I)P(G|I)P(A|G)P(B|G)P(D|J)P(H|J)P(E|H)P(F|H))$.
We have evidence $e = \{A,B,C,D\}$ so they disappear in the cluster tree.





$\mathcal{D}$

	A	B	C	D	E	F
1	1	0	0	0	1	1
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0

Model: Jukes-Cantor
$P_{i,i}^{\tau} = \frac{1}{4} + \frac{3}{4}\exp(-\frac{4}{3}\tau)$
$P_{i,j}^{\tau} = \frac{1}{4} - \frac{1}{4}\exp(-\frac{4}{3}\tau)$

Upward pass:

$\delta\uparrow_{A,G} = (0.90638, 0.031207, 0.031207, 0.031207)^T$
$\delta\uparrow_{B,G} = (0.058518, 0.824446, 0.058518, 0.058518)^T$
$\delta\uparrow_{C,I} = (0.103338, 0.103338, 0.689985, 0.103338)^T$
$\delta\uparrow_{D,J} = (0.174701, 0.174701, 0.174701, 0.475896)^T$
$\delta\uparrow_{E,H} = (0.586997, 0.137668, 0.137668, 0.137668)^T$

$\delta\uparrow_{F,H} = (0.544931, 0.15169, 0.15169, 0.15169)^T$
$\alpha(G,\cdot) = (0.05303954, 0.02572849, 0.00182617, 0.00182617)^T$
$\alpha(H,\cdot) = (0.31987286, 0.02088286, 0.02088286, 0.02088286)^T$
$\delta\uparrow_{G,I} = (0.04234655, 0.02403941, 0.0080172, 0.0080172)^T$
$\delta\uparrow_{H,J} = (0.17280427, 0.06990572, 0.06990572, 0.06990572)^T$
$\alpha(I,\cdot) = (0.00437601, 0.00248418, 0.00553175, 0.00082848)^T$
$\alpha(J,\cdot) = (0.03018908, 0.0122126, 0.0122126, 0.03326785)^T$
$\delta\uparrow_{I,K} = (0.00385493, 0.00288363, 0.0044483, 0.00203356)^T$
$\delta\uparrow_{J,K} = (0.02413692, 0.01939837, 0.01939837, 0.02494847)^T$
$\alpha(K,\cdot) = (9.30461370e-05, 5.59377217e-05, 8.62897693e-05, 5.07342107e-05)^T$

Downward pass and Beliefs:

$\delta\downarrow_{K,I} = (0.00603423, 0.00484959, 0.00484959, 0.00623712)^T$
$\delta\downarrow_{K,J} = (0.00096373, 0.00072091, 0.00111207, 0.00050839)^T$
$\delta\downarrow_{I,G} = (0.00059633, 0.00053348, 0.00356204, 0.0006071)^T$
$\delta\downarrow_{J,H} = (0.00015068, 0.0001395, 0.00015751, 0.00035334)^T$

$\beta_G = (4.43660557e-05, 2.04372134e-05, 5.15791894e-06, 1.54072264e-06)^T$
$\beta_H = (5.85991776e-05, 3.74529033e-06, 3.87472667e-06, 5.28214031e-06)^T$
$\beta_I = (2.52526318e-05, 1.28245382e-05, 2.85575679e-05, 4.86720826e-06)^T$
$\beta_J = (2.60383153e-05, 9.75177438e-06, 1.10109992e-05, 2.47007606e-05)^T$
$\beta_K = (2.32615342e-05, 1.39844304e-05, 2.15724423e-05, 1.26835527e-05)^T$

Marginals:

$P(G|\mathcal{D}) = (0.62048769, 0.28582751, 0.0721368, 0.02154799)^T$
$P(H|\mathcal{D}) = (0.81955362, 0.0523807, 0.05419097, 0.07387471)^T$
$P(I|\mathcal{D}) = (0.35317405, 0.17935929, 0.39939567, 0.06807099)^T$
$P(J|\mathcal{D}) = (0.36416282, 0.13638493, 0.15399601, 0.34545625)^T$
$P(K|\mathcal{D}) = (0.32532723, 0.19558108, 0.30170421, 0.17738748)^T$