Prove:

We know $\sum_{i=1}^{N} \hat{\alpha}_t(i) = 1$

and if $\hat{\alpha}_t(i) = \prod_{k=1}^{t} \eta_k \alpha_t(i)$

$\therefore \sum_{i=1}^{N} \prod_{k=1}^{t} \eta_k \alpha_t(i) = 1$

$\therefore \sum_{i=1}^{N} \alpha_t(i) \prod_{k=1}^{t} \eta_k = 1$

$\therefore \prod_{k=1}^{t} \eta_k = \frac{1}{\sum_{i=1}^{N} \alpha_t(i)} = \frac{1}{p(O(t)|\lambda)}$

in particular, $\prod_{k=1}^{T} \eta_k = \frac{1}{\sum_{i=1}^{N} \alpha_T(i)} = \frac{1}{p(O|\lambda)}$

$\therefore \hat{\alpha}_t(i) = p(q_t = S_i \mid O(t), \lambda)$

$\because \alpha_t(i) = p(O(t), q_t = S_i / \lambda)$

$\therefore$ Prove (make $\eta = c$, $k = s$, $\therefore$ Prove)

2、 We know : $\beta_T(j) = c_T$ and

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

$$= P(o_{t+1} \cdots o_T \mid i_t = q_i, \lambda)$$

Prove $\hat{\beta}_t(j) = \prod_{s=t}^{T} c_s \beta_t(j)$

Obviously, when $t = T$, Prove.

When $t \neq T$, Assume $\hat{\beta}_t(j) = \prod_{s=t}^{T} c_s \beta_t(j)$

$$\sum_{j=1}^{N} \prod_{s=1}^{T} c_s \beta_1(j) = 1$$

$$\Rightarrow \sum_{j=1}^{N} \beta_T(j) \prod_{s=1}^{T} c_s = 1$$

Cause $\prod_{s=1}^{T} c_s = \frac{1}{\sum_j \alpha_T(j)}$

$$\Rightarrow \frac{\sum_{j=1}^{N} \beta_1(j)}{\sum_{j=1}^{N} \alpha_T(j)} = 1$$, So prove

3. $P(O|\lambda) = \sum_i \alpha_T(i)$ (using formula in (1)(2))

$$\Rightarrow \xi_t(i,j) = \frac{\alpha_t(i)\,\beta_{t+1}(j)\,a_{ij}\,b_j(o_{t+1})}{P(O|\lambda)}$$

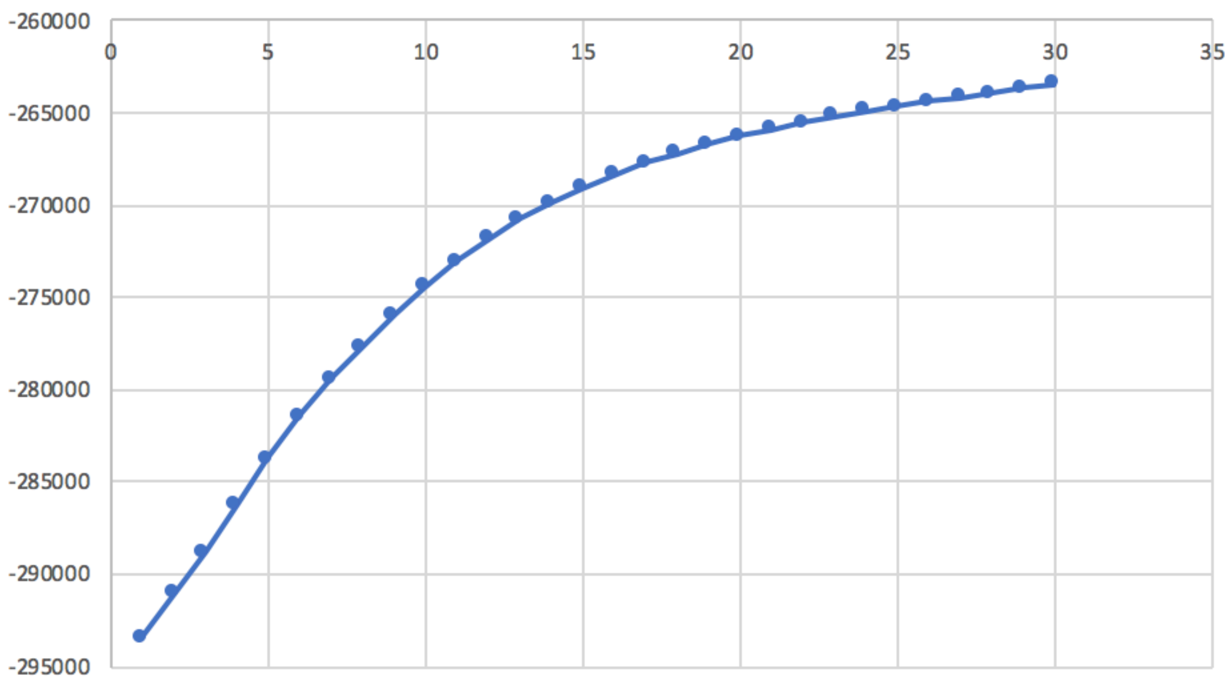
$$= \frac{\alpha_t(i)\,\beta_{t+1}(j)\,a_{ij}\,b_j(o_{t+1})}{\sum_j \alpha_T(j)}$$

$$\sum_j \alpha_T(j) = \frac{1}{\prod_{s=1}^{T} c_s}$$

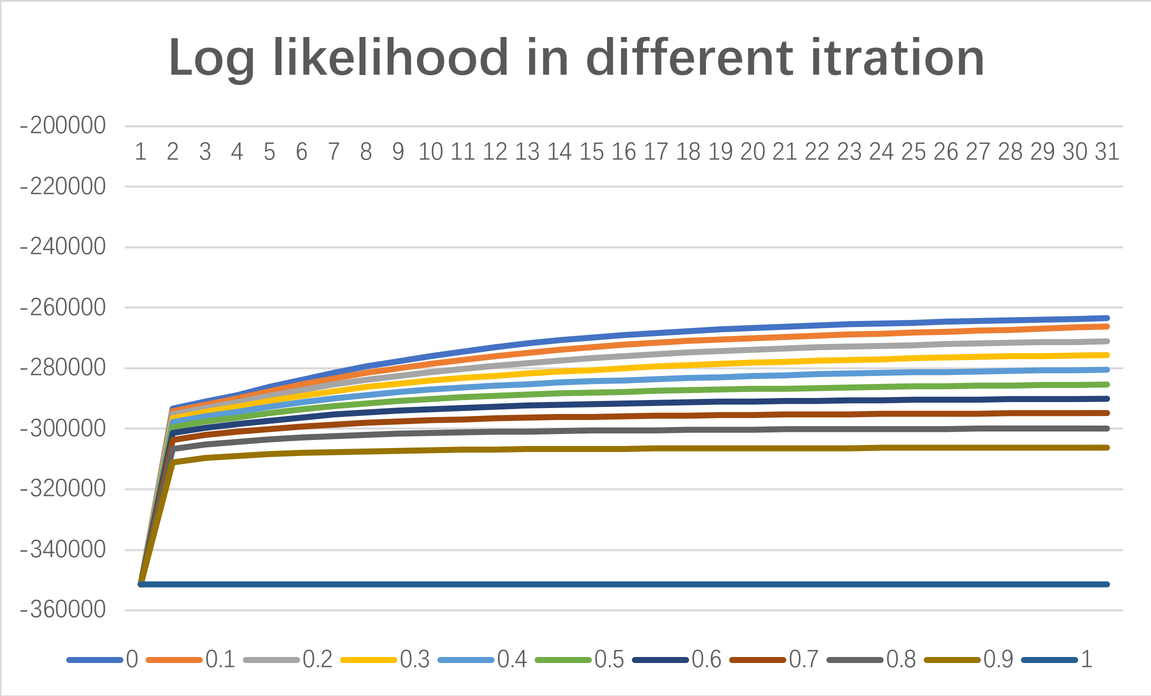
$$\Rightarrow \xi_t(i,j) = \left(\prod_{s=1}^{t} c_s\,\alpha_t(i)\right)\left(\prod_{s=t+1}^{T} c_s\,\beta_{t+1}(j)\right) \cdot a_{ij}\,b_j(o_{t+1})$$

$$= \hat{\alpha}_t(i)\,\hat{\beta}_{t+1}(j)\,a_{ij}\,b_j(o_{t+1})$$

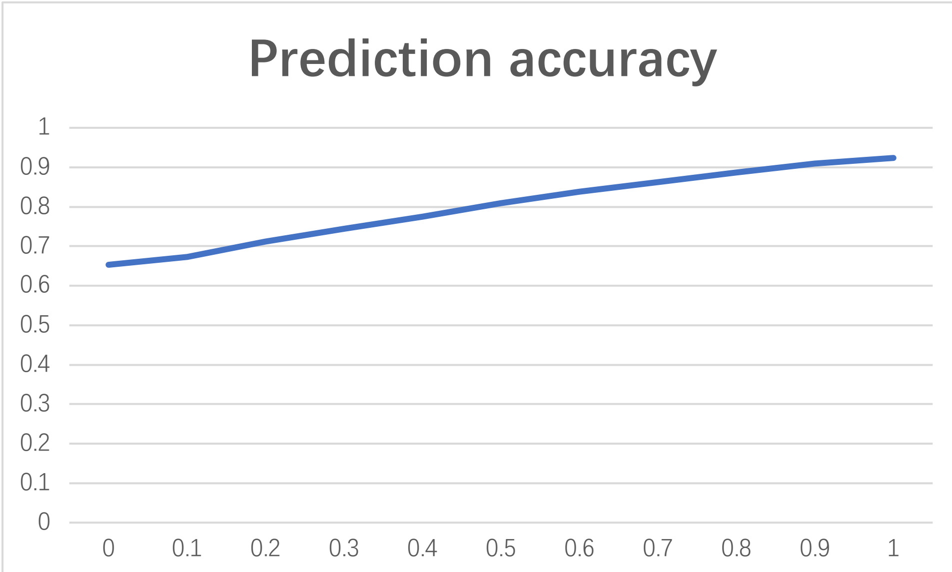
So we could prove that.



-260000
-265000
-270000
-275000
-280000
-285000
-290000
-295000
0
5
10
15
20
25
30
35



x-axis is the number of iterations and the y-axis is the log likelihoods P(D|λ).



x-axis is different mu values and the y-axis is the POS-tagging accuracies.

4.

Because $A = \mu A_L + (1-\mu) A_D$

for A_{ij} in A_L, A_D

We know $\sum_j A_{ij} = 1$ for all of i.

So $\sum_j A_{ij} = \mu \sum_j A^L_{ij} + (1-\mu) \sum_j A^D_{ij}$

$= \mu + 1 - \mu$

$= 1$

And same with β, π

So prove.