Hidden semi-Markov Models

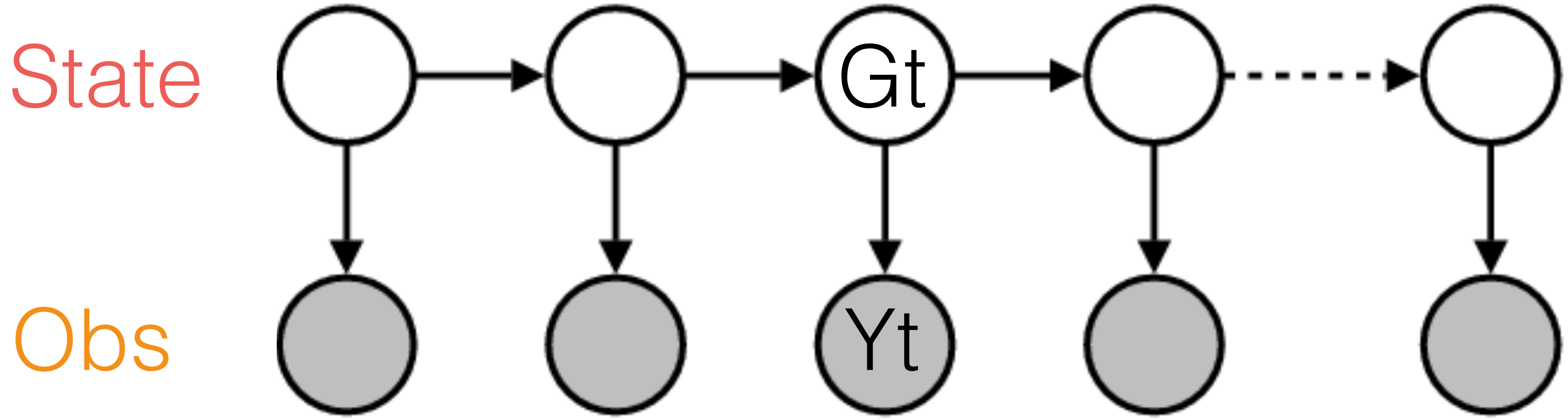
Kevin Murphy

Colin Lea
Summer 2014

Notation

Yt = Observation at time t
Gt = labels at time t
q = idx of state
l = duration

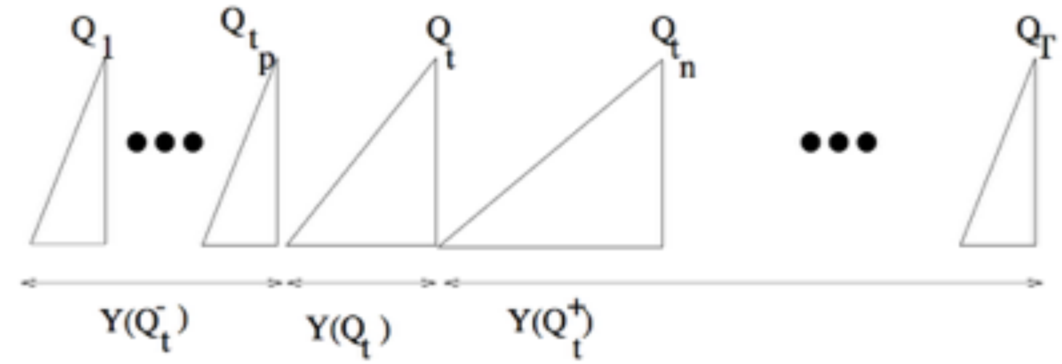
Y(Gt) = observation of segment Gt



HMM

Today

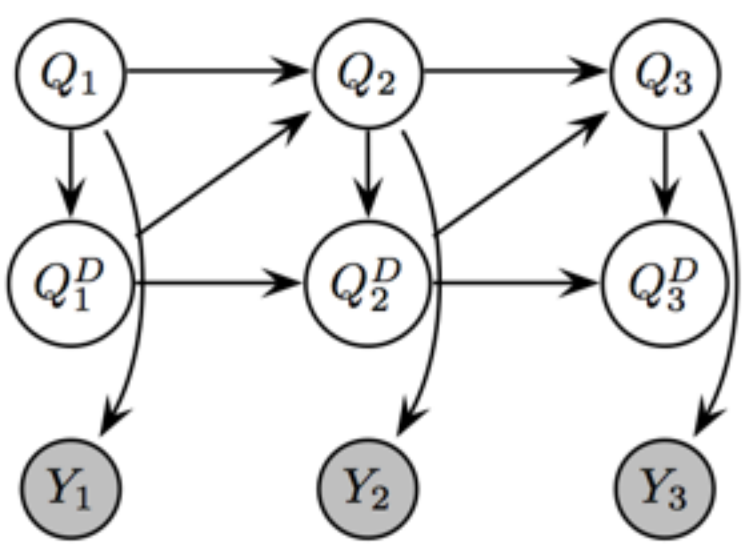
Explicit Duration HMM



P(Obs | state, length):　　　　P(Obs_i | state):

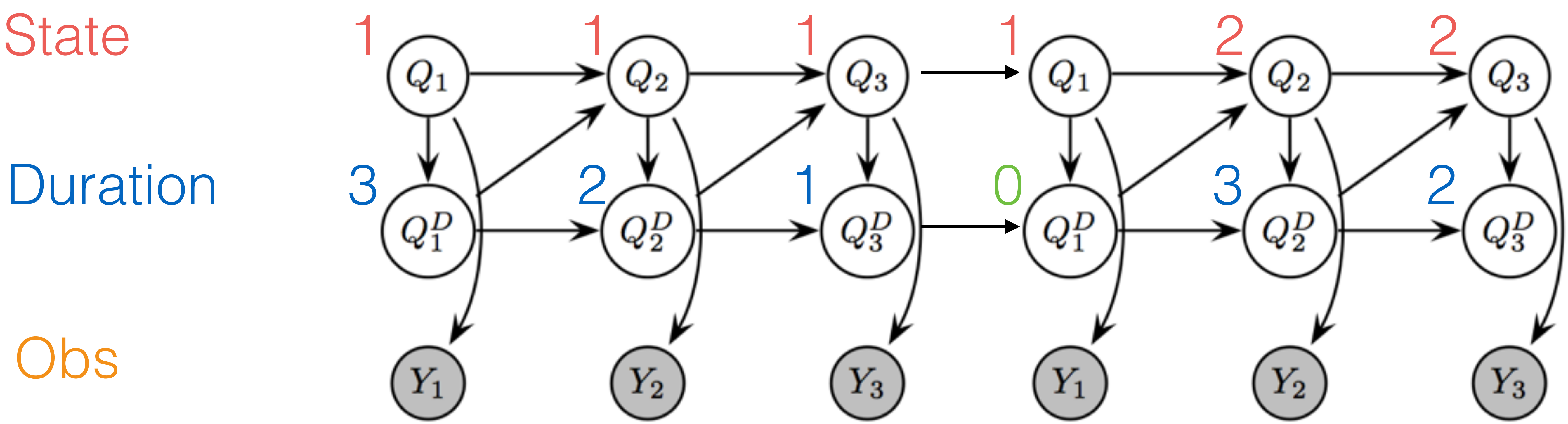
$$P(Y(G_t)|q,l) = \prod_{i=t-l+1}^{t} P(y_i|q)$$

Segment HMM



Variable Duration HMM (1/2)

Add duration variable
Decrement consequent nodes



CPD:

P(StateT | StateT-1, DurationT-1):

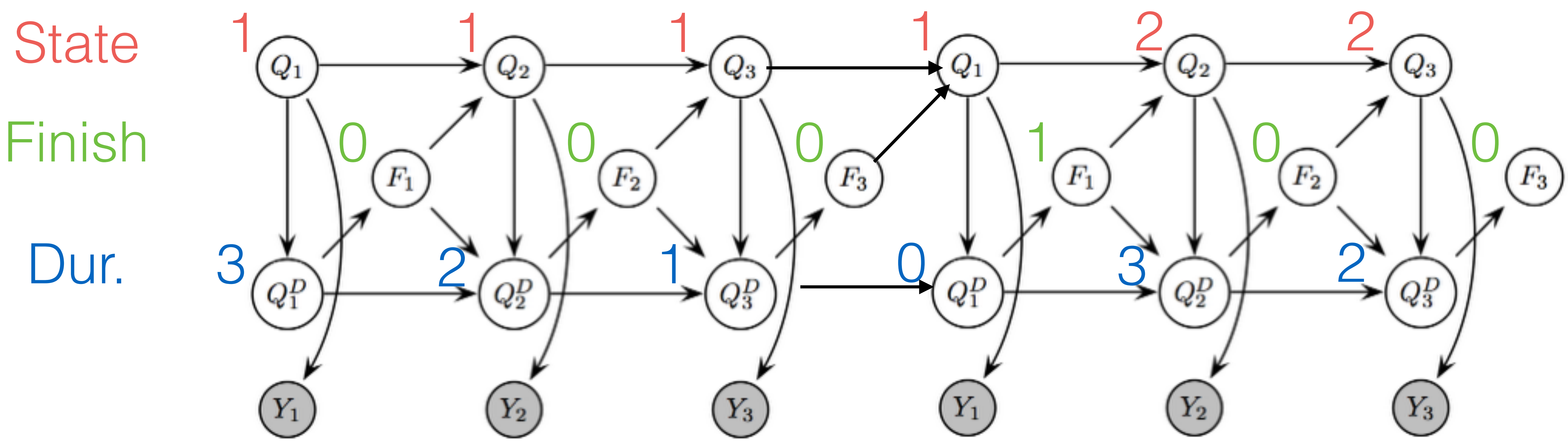
$$P(Q_t = j | Q_{t-1} = i, Q_{t-1}^D = d) = \begin{cases} \delta(i,j) & \text{if } d > 0 \text{ (remain in same state)} \\ A(i,j) & \text{if } d = 0 \text{ (transition)} \end{cases}$$

P(DurationT | DurationT-1, StateT):

$$P(Q_t^D = d' | Q_{t-1}^D = d, Q_t = k) = \begin{cases} p_k(d') & \text{if } d = 0 \text{ (reset)} \\ \delta(d', d-1) & \text{if } d > 0 \text{ (decrement)} \end{cases}$$

Variable Duration HMM (2/2)

Add explicit Finish indicator



CPD:

P(StateT | StateT-1, FinishT-1):

$$P(Q_t = j | Q_{t-1} = i, F_{t-1} = f) = \begin{cases} \delta(i,j) & \text{if } f = 0 \text{ (remain in same state)} \\ A(i,j) & \text{if } f = 1 \text{ (transition)} \end{cases}$$

P(DurT | DurT-1, StateT, FinishT-1=1):

$$P(Q_t^D = d' | Q_{t-1}^D = d, Q_t = k, F_{t-1} = 1) = p_k(d')$$

P(DurT | DurT-1, StateT, FinishT-1=0):

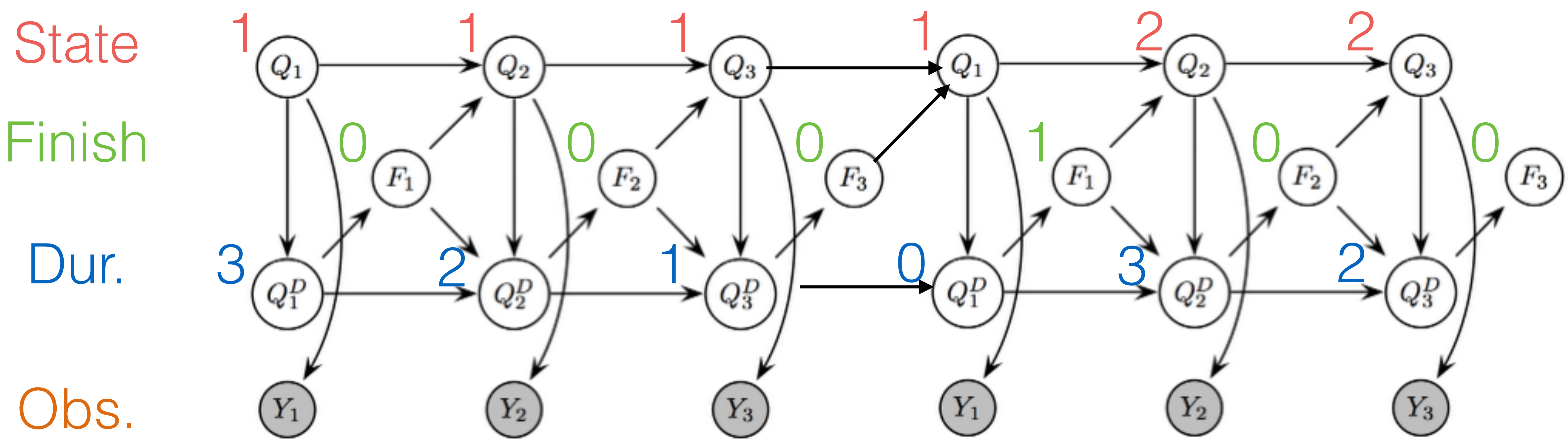
$$P(Q_t^D = d' | Q_{t-1}^D = d, Q_t = k, F_{t-1} = 0) = \begin{cases} \delta(d', d-1) & \text{if } d > 0 \\ \text{undefined} & \text{if } d = 0 \end{cases}$$

P(FinishT=1 | DurT):

$$P(F_t = 1 | Q_t^D = d) = \delta(d, 0)$$

Variable Duration HMM (2/2)

Add explicit Finish indicator



P(Obs | StateT, length): P(ObsT | StateT):

$$P(y_{1:l}|Q_t = k, l) = \prod_{t=1}^{l} P(y_t|Q_t = k)$$

All observations are independent given states

Segments as HMMs

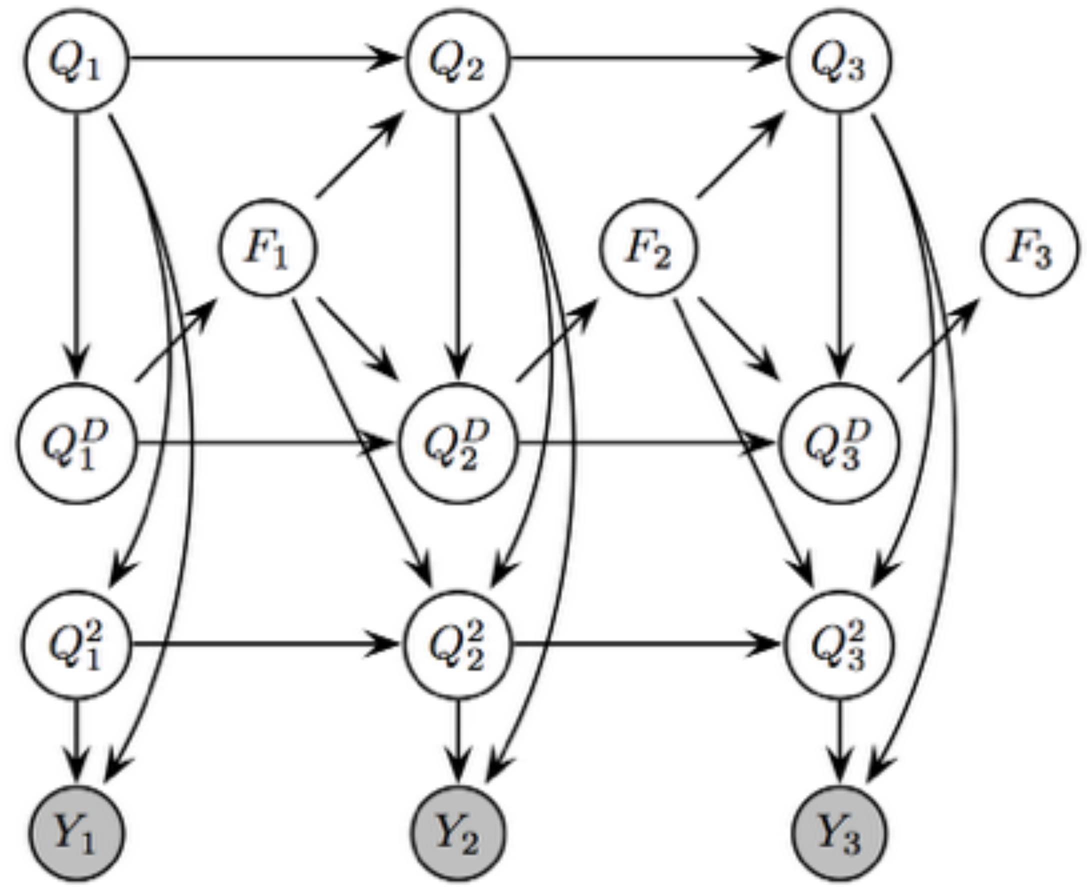
Higher level state

Finish

Duration

Lower level state

Observation



P(Obs | StateT, length): prior P(Obs1 | State1T, State2,1): State2Transition P(ObsT | State1T, State2,1):

$$P(y_{1:l}|Q_t = k, l) = \sum_{q_{1:l}} \pi_k(q_1) P(y_1|Q_t = k, Q_1^2 = q_1) \prod_{\tau=2}^{l} A_k(q_{\tau-1}, q_\tau) P(y_t|Q_t = k, Q_\tau^2 = q_\tau)$$

CPD:

P(State2T | State2T-1, State1T, FinishT-1):

$$P(Q_t^2 = j|Q_{t-1}^2 = i, Q_t = k, F_{t-1} = f) = \begin{cases} \pi_k^2(j) & \text{if } f = 0 \text{ (reset)} \\ A_k^2(i,j) & \text{if } f = 1 \text{ (transition)} \end{cases}$$

Inference

Forwards Backwards Algorithm

(Start with traditional HMM on board)
max P(Gt | Y)

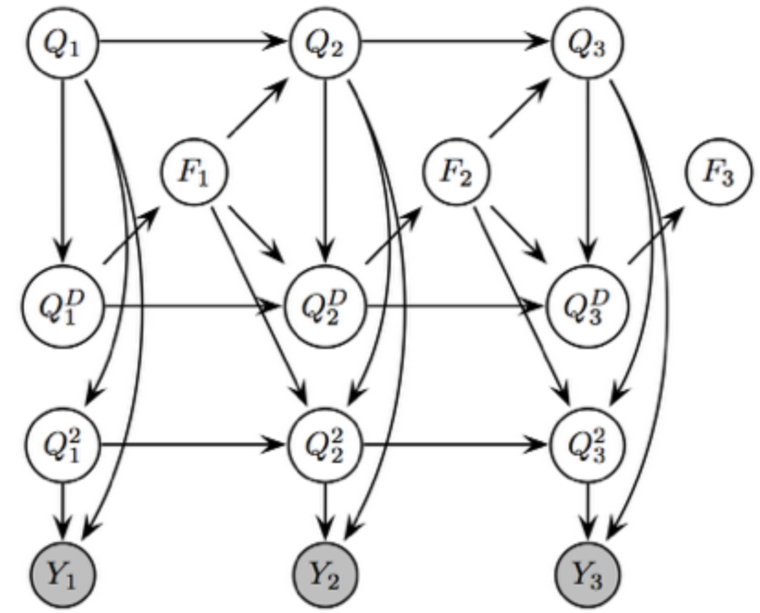
P(StateT, ObsT, ObsT-1)

$$\alpha_t(g) \stackrel{\text{def}}{=} P(G_t = g, Y(G_t), Y(G_t^-))$$

P(ObsT | StateT)

$$\beta_t(g) \stackrel{\text{def}}{=} P(Y(G_t^+)|G_t = g)$$

Inference (Forwards)



Random Variables:

tp = Prev time

tn = Next time

P(StateT, Obs) P(StateT, StateT-1, ObsT-1, ObsT)

$$\alpha_t(g) = \sum_{g'} P(G_t = g, G_{t_p} = g', Y(G_t^-), Y(G_t))$$

P(ObsT | StateT, ~~StateT-1~~, ~~ObsT-1~~) P(StateT, StateT-1, ObsT-1)

$$= \sum_{g'} P(Y(G_t)|G_t = g, \cancel{G_{t_p} = g'}, \cancel{Y(G_t^-)})P(G_t = g, G_{t_p} = g', Y(G_t^-))$$

P(ObsT | StateT) P(StateT | StateT-1, ~~ObsT-1~~) P(StateT-1, ObsT-1)

$$= \sum_{g'} P(Y(G_t)|G_t = g)P(G_t = g|G_{t_p} = g', \cancel{Y(G_t^-)})P(G_{t_p} = g', Y(G_t^-))$$

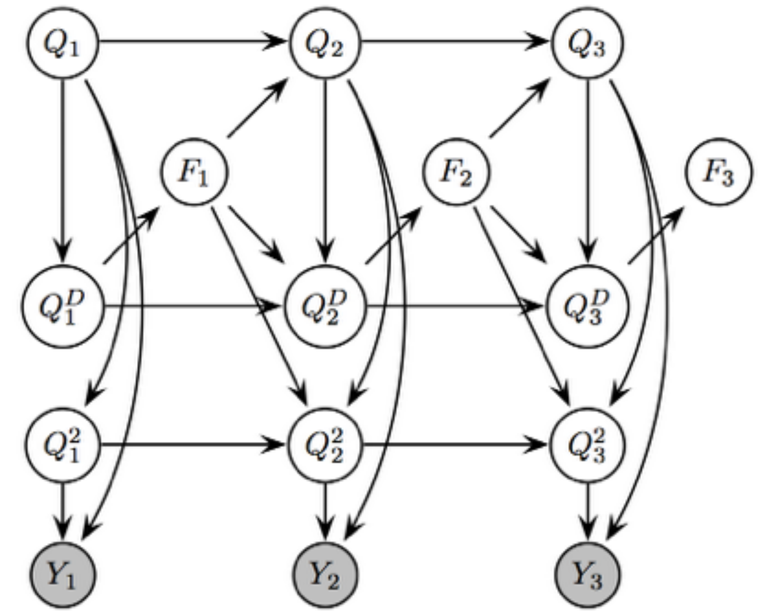
P(ObsT | StateT) P(StateT | StateT-1) P(StateT-1, ObsT-1)

$$= \sum_{g'} P(Y(G_t)|G_t = g)P(G_t = g|G_{t_p} = g')P(G_{t_p} = g', Y(G_t^-))$$

P(ObsT | StateT) P(StateT | StateT-1) P(StateT-1, ObsT-1)

$$= O_t(g) \sum_{g'} P(g|g')\alpha_{t_p}(g')$$

Inference (Backwards)



P(ObsT | StateT) P(ObsT+1, ObsT, StateT+1 | StateT)

$$\beta_t(g) = \sum_{g'} P(Y(G_{t_n}^{+}), Y(G_{t_n}), G_{t_n} = g' | G_t = g)$$

P(ObsT+1 | ~~ObsT~~, StateT+1, ~~StateT~~) P(ObsT+1, StateT+1 | StateT)

$$= \sum_{g'} P(Y(G_{t_n}^{+}) | \cancel{Y(G_{t_n})}, G_{t_n} = g', \cancel{G_t = g}) P(Y(G_{t_n}), G_{t_n} = g' | G_t = g)$$

P(ObsT+1 |StateT+1) P(ObsT+1 | StateT+1, ~~StateT~~) P(StateT+1 | StateT)

$$= \sum_{g'} P(Y(G_{t_n}^{+}) | G_{t_n} = g') P(Y(G_{t_n}) | G_{t_n} = g', \cancel{G_t = g}) P(G_{t_n} = g' | G_t = g)$$

P(ObsT+1 |StateT+1)P(ObsT+1 | StateT+1)P(StateT+1 | StateT)

$$= \sum_{g'} \beta_{t_n}(g') O_{t_n}(g') P(g'|g)$$

Inference (Approach 2)

Problem! tn and tp are random variables!

P(StateT, DurT, FinishT, Obs1:T)

$$\alpha_t(q,l) \stackrel{\text{def}}{=} P(Q_t = q, L_t = l, F_t = 1, y_{1:t})$$
$$= P(y_{t-l+1:t}|q,l) \sum_{q'} \sum_{l'} P(q,l|q',l')\alpha_{t-l}(q',l')$$

P(StateT, DurT | StateT-1, DurT-1)

P(Obs |StateT, DurT)

P(StateT-1, DurT-1, FinishT-1, Obs1:T-1)

P(Obs_t:T | StateT, DurT, FinishT)

$$\beta_t(q,l) = P(y_{t+1:T}|Q_t = q, L_t = l, F_t = 1)$$
$$= \sum_{q'} \sum_{l'} \beta_{t+l'}(q',l') P(y_{t+1:t+l'}|q',l') P(q',l'|q,l)$$

P(Obs | StateT+1, DurT+1, FinishT+1) P(Obs |StateT+1, DurT+1)

P(StateT+1, DurT+1 | StateT, DurT)

Learning

Same as HMM but per-segment

Prior

P(Obs | State1, Finish0)

$$\hat{\pi}_i \propto P(Q_1 = i | y_{1:T}) \propto P(Q_1 = i) P(y_{1:T} | Q_1 = i, F_0 = 1) = \pi_i \beta_0^*(i)$$

State transition matrix

P(StateT, ObsT, ObsT-1) P(StateT+1 | StateT) P(ObsT+1 | StateT+1)

$$\hat{A}_{ij} \propto \sum_{t=1}^{T-1} \alpha_t(i) A_{ij} \beta_t^*(j)$$

β_t^* = probability of seeing future evidence given that we start in state i at t + 1

Observation matrix

P(StateT | Obs)

$$\hat{B}_{i,k} \propto \sum_{t:Y_t=k} P(Q_t = i | y_{1:T})$$

P(StateT+1, FinishT | Obs) P(StateT+1, FinishT | Obs)

$$\sum_{t=1}^{T} P(Q_t = i | y_{1:T}) = \sum_t \sum_{\tau < t} P(Q_{t+1} = i, F_t = 1 | y_{1:T}) - P(Q_t = i, F_t = 1 | y_{1:T})$$