Learning for Decision Making

- *Why*? .... DM selects policy $\{\pi(a_t | s_t)\}$

$$\max_{\pi} E[\Sigma_t\, r(s_t, a_t, s_{t-1})]$$

- Optimization needs to expectation E
- Dynamic programming works with

$$p(s_t \mid a_t, s_{t-1})$$

Learning primarily is to provide this *predictor*

Mathematics Only Transforms Its Inputs

- *Inputs*?

  Observation model $p(s_t \mid a_t, s_{t-1}, h_t)$

  relating a hidden variable h_t to observed s_t

  Time-evolution model $p(h_t \mid a_t, h_{t-1})$

  Prior distribution $p(h_0)$

  Observed data $d^t = (s_t, a_t, s_{t-1}, a_{t-1}, \ldots, s_1, a_1)$

- *Output*? predictor $p(s_{t+1} \mid a_{t+1}, s_t)$

Transformation Uses Rules for Probabilities

- Given joint probability $P(\alpha, \beta) \geq 0$, normalized to unit sum, determines

Marginal probability $P(\alpha) = \Sigma_\beta P(\alpha,\beta)$

Conditional probability $P(\alpha|\beta) = P(\alpha,\beta)/P(\beta)$

$\Leftrightarrow$ Chain rule $P(\alpha,\beta) = P(\alpha|\beta)P(\beta)$

Bayesian Learning

Predictor $p(s_{t+1}|a_{t+1},s_t,d^t)$

$$=\Sigma_{ht}\, p(s_t|a_t,s_{t-1}\,,h_t,d^t)\; p(h_t|d^t)$$

$$=\Sigma_{ht}\, p(s_t|a_t,s_{t-1}\,,h_t) \times p(h_t|\, d^t)$$

observation m. x h_t-estimate

$$p(h_{t+1}|a_{t+1},d^t)=\Sigma_{ht}\, p(h_{t+1}|a_{t+1},h_t) \times p(h_t|a_{t+1},d^t)$$

h_{t+1} predictor=time evolution m. h_t-estimate

Natural: h_t, a_{t+1} conditionally independent

Filtering – Evolution of Hidden Variables

Data updating (Bayes rule)

$p(h_t|d^t) = c \times p(s_t|a_t,s_{t-1,}h_t) \times p(h_t|d^{t-1})$

Time updating

$p(h_{t+1}|a_{t+1},d^t)=\Sigma_{ht}\, p(h_{t+1}|a_{t+1},h_t) \times p(h_t|d^t)$

Valid if : h_t, a_{t+1} conditionally independent

Prior probability initiates $p(\theta|d^{-1}) = p(\theta)$

Bayesian Estimation

Parameter: time-invariant hidden variable

$$h_{t+1} = h_t = \theta \qquad p(h_{t+1} | a_{t+1}, h_t) = \delta(h_{t+1}, h_t)$$

Predictor: $p(s_{t+1} | a_{t+1}, d^t)$

$$= \Sigma_{ht}\, p(s_t | a_t, s_{t-1}, \theta) \times p(\theta | d^t)$$

Data updating (Bayes rule)

$$p(\theta | d^t) = c \times p(s_t | a_t, s_{t-1}, \theta) \times p(\theta | d^{t-1})$$

Prior probability initiates: $p(\theta | d^{-1}) = p(\theta)$

Where observation, time-evolution models
and prior probability come from?

Domain-knowledge modelling: physics,
sociology, economy, engineering

Choice of free variables via learning

Black-box modelling: Probabilities
embedded in a parametrized dense set

Choice well-approximating via learning

Example: Markov Chain given by Finite S, A

$p(s_t = s' | a_t = a, s_{t-1} = s, \theta) = \theta(s' | a,s) \geq 0,$

$\Sigma_{s'} \theta(s' | a,s) = 1,$ on A, S

Prior: $\mathrm{Dirichlet}(V_0) = c\Pi_{s',a,s} [\theta(s' | a,s)]^{V_0(s' | a,s)-1}$

Posterior: $p(\theta | d^t) = \mathrm{Dirichlet}(V_t)$

Bayes: $V_t(s'_t | a_t, s_t) = V_{t-1}(s'_t | a_t, s_t)+1$

Predictor: needs hyper-state s_t, V_t !

$p(s_{t+1} | a_{t+1}, d^t) = p(s_{t+1} | a_{t+1}, s_t, V_t) = cV_t(s'_t | a_t, s_t)$