Learning for Decision Making

Why? DM selects policy $\{\pi(a_t | s_{t-1})\}$

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

- Optimization needs to expectation E
- Dynamic programming works with

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

Learning primarily is to provide this *predictor*

Mathematics Only Transforms Its Inputs

Inputs?

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

relating a hidden variable h_t to observed s_t

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

Prior distribution $p(h_1 | d^0)$

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

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

Transformation Uses Rules for Probabilities

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

Marginal probability $P(\alpha) = \int P(\alpha,\beta)\, d\beta$

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

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

Bayesian Learning

Predictor

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

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

observation model x h_t-estimate

$$p(h_{t+1}|a_{t+1},d^t) = \int p(h_{t+1}|a_{t+1},h_t)\ p(h_t|a_{t+1},d^t) dh_t$$

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)

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

Time updating

$$p(h_{t+1}|a_{t+1},d^t) = \int p(h_{t+1}|a_{t+1},h_t)p(h_t|d^t)\ dh_t$$

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

Prior probability initiates $p(h_1|d^0) = p(h_1)$

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

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

Free variables obtained via learning!

Black-box modelling:

Probabilities from a parametrized dense set

Choice well-approximating m. by learning!

Example: Moving a Point of Mass μ

h'_t =[position, speed, acceleration](tT), T small

$$h_{t+1} \approx \begin{pmatrix} 1 & T & 0 \\ 0 & 1 & T \\ 0 & 0 & 0 \end{pmatrix} h_t + \begin{pmatrix} 0 \\ 0 \\ \mu^{-1} \end{pmatrix} a_{t+1}$$

a_{t+1} applied force

$$h_{t+1} \approx A\, h_t + B\, a_{t+1}$$

$$s_t \approx \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} h_t + \begin{pmatrix} 0 \end{pmatrix} a_t$$

$$s_t \approx Ch_t + Da_{t+1}$$

 Kalman Filtering: Google 660 000 items

Application samples

Process control temperature, pressure, flow rate

Tracking systems radar, navigation, GPS, video

Transportation vehicle flow, position, high-way state

Military weapons navigation

Business econometry, prediction, marketing

Medicine brain imaging, epidemics …

KF = Bayesian Filtering with Gaussian Models

Observation model (hidden h_t vs. seen s_t)

$$p(s_t | a_t, s_{t-1}, h_t) = N_{st}(Ch_t + Da_t, Q), \; Q>0$$

Equation form (wide spread)

$$s_t = Ch_t + Da_t + e_t, \qquad e_t \sim N_{st}(0, Q),$$

has to be complemented by the assumption:

Noise e_t conditionally independent of h_t, a_t

Observation model = hidden h_t vs. seen s_t

typical construction by Taylor expansion

$$s_t \approx \Psi(h_t, a_t) \approx C\, h_t + D\, a_t$$

- Noise: observation & approximation errors
- C, D, Q data dependent (no trouble)
- C, D, Q rarely fully known (big trouble)
- Gauss: Limit th., maximum entropy principle

  Questionable but feasibility dominates

Time-evolution m. = unseen changes of h_t

$$p(h_{t+1} | a_{t+1}, h_t) = N_{st} (Ah_t + Ba_{t+1}, R), \; R>0$$

Equation form (wide spread)

$$h_{t+1} = Ah_t + Ba_{t+1} + w_{t+1}, \qquad w_{t+1} \sim N(0, R),$$

has to be complemented by assumption:

State noise w_{t+1} independent of e_t, h_t, a_{t+1}

Time-evolution m. = unseen changes of h_t

typical construction by Taylor expansion

$$h_{t+1} \approx \Psi(h_t, a_{t+1}) \approx Ah_t + Ba_{t+1}$$

Ψ choice: a hard use of domain knowledge

- Noise: observation & approximation errors
- A, B, R data dependent (no trouble)
- A, B, R rarely fully known (big trouble)
- Gauss: Limit th., max. entropy, feasibility

Prior distribution = guess of h_0 *before* using

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

$p(h_1) = N_{h1} (h_{1|0} , P_{1|0}), \; P_{1|0} > 0$

Often flat prior: $h_{0|0} = 0, P_{0|0}$ = big diagonal

Wide spread but unreasonable for DM

(strong transient oscillations)

Knowledge elicitation: "fictitious" data d^t, $t<1$

Overfitting danger: care pays back to DM

Theorem (KF): $p(h_{t+i} \mid d^t) = N(h_{t+i|t}, P_{t+i|t})$, $i = 0,1$

Predictor $p(s_{t+1} \mid a_{t+1}, d^t) = N(s_{t+1|t}, Q_{t+1|t})$

Outline of the proof by induction:

- Data updating, Bayes' r.= a function product with in exp(quadratic form in h_t). Square completion gives Gauss its moments. h_t-independent factor cancels

- Time updating, integrand function product with in exp(quadratic form in h_t). Square completion in h_t gives $\int$Gauss dh_t exp(quadratic form in h_{t+1})

Square completion (free c: compare sides):

$$c'\Omega c - 2\, c'\psi + \rho = (c-\check{c})'\Omega(c-\check{c}) + \Lambda$$

$$\check{c} = \Omega^{-1}\psi,\ \Lambda = \rho - \psi'\Omega^{-1}\psi \qquad (\ ^{-1} \text{ critical})$$

Matrix inversion lemma helps $R,P > 0$

$(P^{-1} + C'R^{-1}C)^{-1} = P – PC\ (R + C'P^{-}C)^{-1}CP$ consider C = [redacted]

Data updating: mean $h_{t|t} = h_{t|t-1} + G_t \, \varepsilon_{t|t-1}$

Prediction error $\varepsilon_{t|t-1} = s_t - C h_{t|t-1} - D a_t$

Kalman gain $G_t = P_{t|t-1} C' (C P_{t|t-1} C' + Q)^{-1}$

Covariance: $P_{t|t} = P_{t|t-1} - G_t \, C \, P_{t|t-1}$

- $\varepsilon_{t|t-1}$ intuitive: $h_{t|t-1}$ instead h_t in observ. mod.
- $P_{t|t} \leq P_{t|t-1}$ (numeric: Choleski decomposition)
- C, D, Q, $P_{1|0}$ strong influence

Time updating: mean $h_{t+1|t} = Ah_{t|t} + Ba_{t+1}$

Covariance: $P_{t+1|t} = AP_{t|t}A' + R$

- Covariances and gains can be precomputed
- $A, B, R, P_{1|0}$ strong influence
- Formulae valid for past-data dependent A, B
  exploited non-linear filtering
  Extended Kalman filter $(A,B,C,D)_t$: Taylor
  expansion of nonlinear $E[s_t|a_t, d^{t-1}]$ at $h_{t|t-1}$

Prediction of s_{t+1}: mean $s_{t+1|t} = Ch_{t|t} + Da_{t+1}$

covariance $Q_{t+1|t} = Q + C\,P_{t|t}\,C'$

- $Q_{t+1|t} > Q$: $C\,P_{t|t}\,C'$ reflects uncertainty on h_t
- For data-independent matrices, only mean values are data dependent; covariance not
- Prediction of observations & hiddens need sufficient statistic (information state) $h_{t+i|t}, P_{t+i|t}$, i=0,1, $p(s_t | s_{t-1}, a_t) \neq p(s_t | a_t, d^{t-1})$

- KF updates sufficient statistic: don't-spoil it!

  E.g., if $0 \leq h_t$ & $h_{t|t} \leq 0$ do not touch $h_{t|t}$

- KF updates posterior probabilities

  $\Rightarrow$ EKF too much local and often diverges

  $\Rightarrow$ unscented KF: approximates high probability area, HPA)

  $\Rightarrow$ particle filters: approximate probabilities on a grid concentrated in HPA

- KF as point estimator of h_t generates $h_{t|t}$
- Linear estimator *min*(square error) $\Rightarrow$ KF like
- Luenberger observers KF structure with KF gain replaced by another one optimized with respect to stability and transients

- Dependence of results on A,B,C,D studied. E.g. with $C = 0$ nobody can learn h_t :

  $$p(h_{t+1} | d^t) = p(h_{t+1}) = \int p(h_{t+1} | a_{t+1}, h_t)\, p(h_t) dh_t$$

- Generally, rank of observability matrix $[C', A'C', \ldots A'^{\dim(h)-1}C']$ matter (Cayley-Hamilton)
- For data-dependent matrices, observability can be enhanced or spoilt by your actions!

$$h_{t+1} \approx \begin{pmatrix} 1 & T & 0 \\ 0 & 1 & T \\ 0 & 0 & 0 \end{pmatrix} h_t + \begin{pmatrix} 0 \\ 0 \\ \mu^{-1} \end{pmatrix} a_{t+1}$$

Moving mass μ

$$h_{t+1} \approx Ah_t + Ba_{t+1}$$

$$s_t \approx \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} h_t + \begin{bmatrix} 0 \end{bmatrix} a_t$$

$$s_t \approx Ch_t + Da_{t+1}$$

$$[C', A'C', A'^2 C'] = \begin{pmatrix} 1 & 1 & 1 \\ 0 & T & 2T \\ 0 & 0 & T^2 \end{pmatrix}$$

rank = 3

observable h

Observed speed $C = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$

rank = 2

unobservable h

- Filtering for discrete-valued case provides model for POMDP: dimensionality curse
- Mixed data case formally possible but elaborated for particular cases
- Black-box model approximation by normal mixtures use KF as basic building block
- Cooperating KF filters used for data fusion and coping with high-dimensional h_t

Historical Readings

Origin: R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems", Transactions of the ASME, Journal of Basic Engineering, 35-45, 1960.

Nice classics: A.M. Jazwinski. Stochastic Processes and Filtering Theory. Academic Press, NY, 1970.

Square-root: G.J. Bierman. Factorization Methods for Discrete Sequential Estimation. Ac. Press, NY, 1977.

Nice reading (you are responsible for assumptions)

T. Bohlin. Interactive System Identication: Prospects and Pitfalls. Springer, NY, 1991.

Observers: D.G. Luenberger. Observers for multivariable system. IEEE Trans. AC, 2, 190 – 197, 1996

Unscented filters: S.J. Julier et al . A new approach for the nonlinear transformation of means and covariances in linear filters. IEEE Trans. on AC, 5(3):477-482, 2000.

Particle filters: A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: 15 years later. In Handbook of Nonlinear Filtering. Oxford U. Press, 2011.

Bayesian Estimation

Parameter: time-invariant hidden variable

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

Predictor:

$$p(s_{t+1} | a_{t+1}, d^t) = \int p(s_t | a_t, s_{t-1}, \theta)\, p(\theta | d^t)\, d\theta$$

Data updating (Bayes rule)

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

Prior probability initiates: $p(\theta | d^0) = p(\theta)$

Recursive Least Squares (Gauss Model)

Parametric (observation) m. (hidden θ vs. s_t)

$p(s_t \mid a_t, s_{t-1}, \theta) = N_{st}(\theta' \Psi_t, Q)$, regression vector

$\Psi_t = \Psi_t(a_t, s_{t-1})$, Q known (can be relaxed)

Equation form (wide spread)

$$s_t = \theta \Psi_t + e_t, \qquad e_t \sim N_{st}(0, Q),$$

Noise e_t conditionally independent of h_t, a_t

Prior probability $p(\theta | d^0) = N(\theta_1, P_1)$

Theorem (RLS special KF): $p(\theta | d^t) = N(\theta_t , P_t)$

Predictor $p(s_{t+1} \mid a_{t+1} , d^t) = N(s_{t+1|t} , Q_{t+1})$

Data updating: mean $\theta_t = \theta_{t-1} + G_t \varepsilon_t$

Prediction error $\varepsilon_t = s_t - \theta'_{t-1} \Psi_{t;}$

Kalman gain $G_t = P_{t-1} \Psi_t (\Psi_t' P_{t-1} \Psi_t + Q)^{-1}$

Covariance: $P_t = P_{t-1} – G_t \Psi_t' \Psi_t$

Prediction: mean $s_{t|t-1} = \theta'_{t-1} \Psi_{t,}$

covariance $Q_t = Q(1 + \Psi_t' P_{t-1} \Psi_t)$

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

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

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 !

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