Particle Filtering and Change Detection

Namrata Vaswani

http://www.cfar.umd.edu/~namrata

Introduction

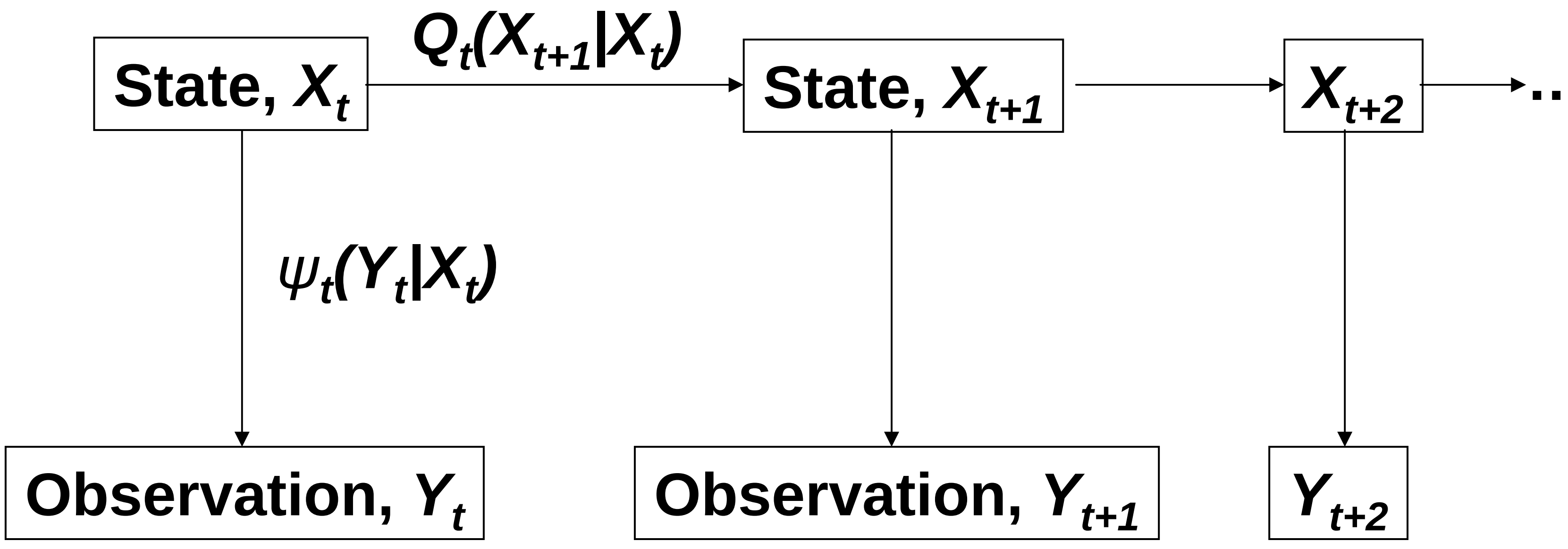
- A particle filter approximates the optimal nonlinear filter as the no. of particles (Monte Carlo samples) goes to infinity
- A.k.a. “Bayesian bootstrap filtering” [Gordon,Salmond, Smith], “Sequential Monte Carlo” [Fearnhead], “Condensation algorithm” [Isard, Blake]
- Given a state space model (or HMM)

$$Y_t = h(X_t) + w_t \;: \qquad \psi_{\mathrm{t},\mathrm{Y}_t}(x_t) \propto p(Y_t \mid x_t)$$

$$X_t = f(X_{t-1}) + n_t \;: \qquad Q_t(x_{t-1}, dx_t) = \Pr(X_t \in dx_t \mid x_{t-1})$$

- Assume distribution of Y_t conditioned on X_t is absolutely continuous, i.e. the pdf $p(Y_t|x_t)$ exists.
- Nonlinear filtering problem: find $Pr(X_t \in dx_t \,|\, Y_{1:t})$, denoted by $\pi_t(dx_t)$ for each t. Here “ϵ” *denotes* “belongs to”

The General HMM



The Optimal Nonlinear Filter

The Bayes' Recursion at time t

$\pi_{t-1}(dx_{t-1})------->\pi_t(dx_t)$ runs as follows:

$$\pi_{t-1}(dx_{t-1})$$

$$\downarrow$$

$$\pi_{t|t-1}(dx_t) = \int Q_t(x_{t-1}, dx_t)\pi_{t-1}(dx_{t-1}),$$

(Most general: $Q_t(x_{t-1}, Y_{1:t-1}, dx_t)$)

$$\downarrow$$

$$\pi_t(dx_t) = \frac{\psi_{t,Y_t}(x_t)\pi_{t|t-1}(dx_t)}{\int \psi_{t,Y_t}(x_t)\pi_{t|t-1}(dx_t)}$$

Possible Solutions

- Kalman filter: linear/Gaussian sys & obs models
  - Extended Kalman Filter: First order Taylor series approx of *f* & *h* at each time step

- Grid based filtering: discrete state space
  - Approx. Grid based filtering: fixed discretization of the state space

- Gaussian Sum Filter: Second order linearization

- Particle filtering

The Basic Particle Filter
[Gordon, Salmon, Smith]

Aim : Evaluate filtering distribution, $\pi_t^N(dx) = \pi_{t|t}^N(dx) = \Pr(X_t \in dx | Y_{1:t}),\ \forall t$

Also get prediction distribution, $\pi_{t|t-1}^N(dx) = \Pr(X_t \in dx | Y_{1:t-1}),\ \forall t$

1. Initialization : Generate Monte Carlo samples from initial prior, $\pi_{0|0}$

$$\pi_{0|0}^N(dx) = \sum_{i=1}^N \delta_{x_0^{(i)}}(dx), \qquad x_0^{(i)} \sim \pi_{0|0}(dx)$$

2. Prediction : Generate samples from prior state transition kernel at t

$$\pi_{t|t-1}^N(dx) = \sum_{i=1}^N \delta_{\tilde{x}_t^{(i)}}(dx), \qquad \tilde{x}_t^{(i)} \sim Q_t(\bullet | x_{t-1}^{(i)})$$

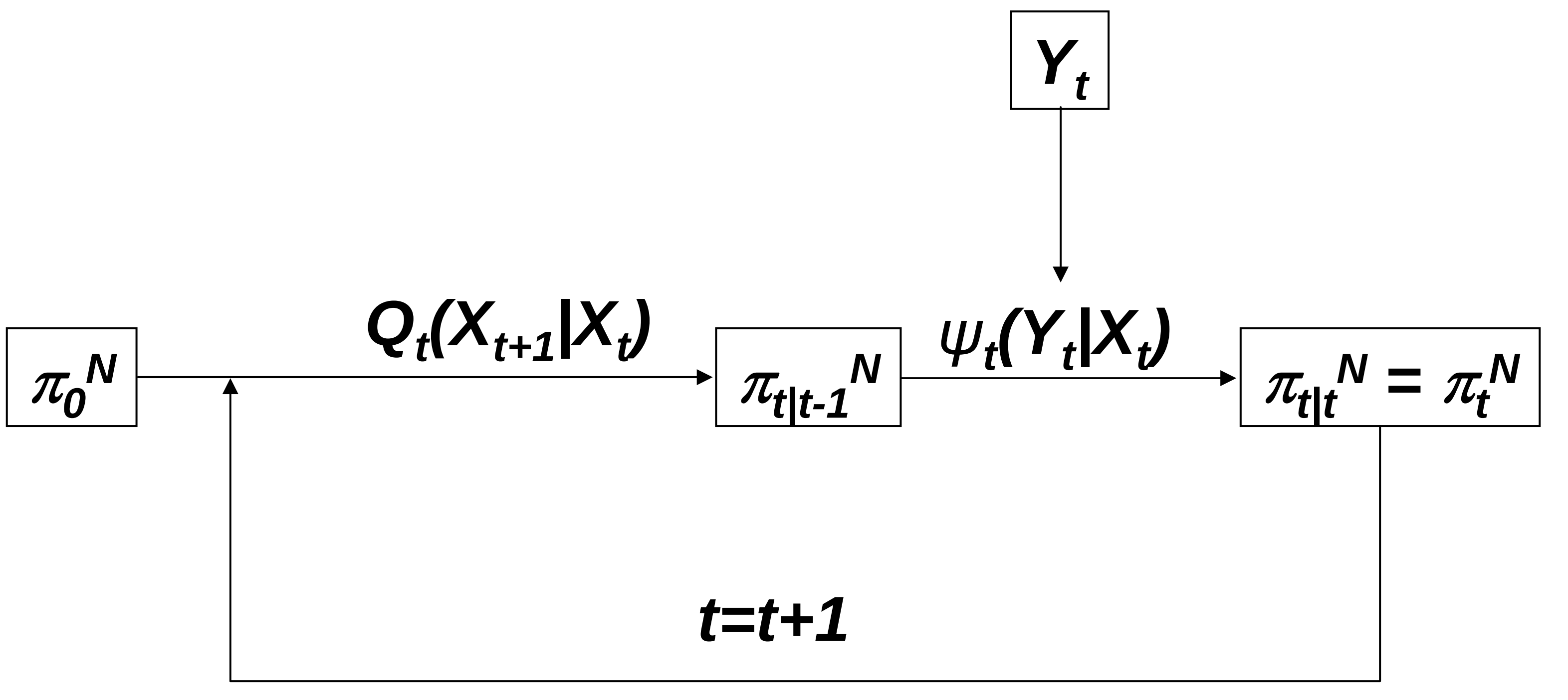
3. Update : Weight each sample by probability of obs. given sample

(a) : $$\tilde{\pi}_{t|t}^N(dx) = \sum_{i=1}^N w_t^{(i)} \delta_{\tilde{x}_t^{(i)}}(dx), \qquad w_t^{(i)} = \frac{\psi_t(Y_t | \tilde{x}_t^{(i)})}{\sum_{i=1}^N \psi_t(Y_t | \tilde{x}_t^{(i)})}$$

(b) : $$\pi_{t|t}^N(dx) = \sum_{i=1}^N \delta_{x_t^{(i)}}(dx), \qquad x_t^{(i)} \sim Multinomial(\{\tilde{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^n)$$: Resample step

4. Set t = t + 1, Go to step 2

Block Diagram for a PF



Some PF Modifications

- Do not resample at each step
- Likelihood PF: sample from observation likelihood, weight by prior [Arulampalam et al]
- Rao Blackwellized PF: if part of the state and observation model is linear, perform Kalman filtering for that part [Chen, Liu]
- Regularised PF [Musso, Oudjane, LeGland]
  - Resample from a kernel density approx. of the posterior
  - Projection PF: Let posterior be approximated by a certain parametric family (e.g. exponential), learn its parameters using the empirical dist., and resample from it [Azimi-Sadjadi & Krishnaprasad]

Change detection problem

- Given a sequence of observations Y_1, $Y_2, \ldots Y_t$ find out if a change occurred at some $t_c <= t$
- Treat as a sequential hypothesis testing problem where t_c is a nuisance parameter

Change detection in Nonlinear systems

- Known change parameters
  - CUSUM on LRT of Y_t given $Y_{1:t-1}$: run *t* PFs
  - Modification of CUSUM with non-growing complexity with *t*
  - Multimode system→ detect change in mode

- Unknown change parameters
  - CUSUM on Generalized LRT ($Y_{1:t}$)
    - Finite parameter set: run *Mt* PFs at time *t*
    - Parameter set is continuous: perform GLRT using ML parameter estimation : relies on accuracy of the ML estimate of the parameter which is based on accuracy of posterior, has been suggested but not implemented (as far as I know!)

  - Model Validation type approaches:
    - Testing if $\{u^j = Pr(Y_t < y^j | Y_{1:t-1})\}$ are uniformly distributed
    - Tracking Error (TE): error b/w Y_t & its prediction
    - Negative log-likelihood of observations (OL), $-logP(Y_t|Y_{1:t-1},H_0)$

Slow Change, Unknown parameters

- All these approaches use observation statistics
- Do not detect slow changes since PF is stable and hence is able to “track a slow change”: error in posterior is small
- For slow change detection, we attempt to use the fact that PF is stable and hence estimate of posterior is “correct” (error small). We propose to use:

  $ELL = E[-Log\ Likelihood(X_t)|Y_{1:t}]$

  $= E[-log\ p_t^0 (X_t)|Y_{1:t}]$

  $= E_\pi[-log\ p_t^0(X_t)]$

  =Kerridge Inaccuracy b/w π_t (posterior) & p_t^0 (prior)

Motivation for ELL

- General Hidden Markov Model (HMM): Markov state sequence $\{X_t\}$, Observation sequence $\{Y_t\}$

$$Y_t = h(X_t) + w_t,\ X_t = f(X_{t-1}) + n_t,\ \{w_t\}, \{n_t\} \text{ indep.}$$

- Finite duration change in system model which causes a permanent change in probability distribution of state
- Slow change: Tracking Error & Observation Likelihood do not detect slow changes. Use distribution of $X_t|Y_{1;t}$
- Change parameters unknown: use *Log-Likelihood*(X_t)
- State is partially observed: use posterior expectation of $LL(X_t)$ given observations, $E[LL(X_t)|Y_{1:t})]$ = **ELL**
- Nonlinear dynamics: Particle filtered estimate of ELL

ELL

- Expected Log Likelihood (ELL)
  - ELL = Kerridge Inaccuracy b/w π_t (posterior) and p_t^0 (prior)

    $$ELL(Y_{1:t}) = E[-\log p_t^0(X_t) | Y_{1:t}] = E_\pi[-\log p_t^0(X_t)] = K(\pi_t : p_t^0)$$

- A sufficient condition for "detectable changes" using ELL
  - $E[ELL(Y_{1:t}^0)] = K(p_t^0 : p_t^0) = H(p_t^0)$, $E[ELL(Y_{1:t}^c)] = K(p_t^c : p_t^0)$
  - Chebyshev Inequality: With false alarm & miss probabilities of 0.11, ELL detects all changes s.t.

    $$K(p_t^c : p_t^0) - H(p_t^0) > 3 [\sqrt{Var\{ELL(Y_{1:t}^c)\}} + \sqrt{Var\{ELL(Y_{1:t}^0)\}}]$$
  - *Set threshold* $= H(p_t^0) + 3\sqrt{Var\{ELL(Y_{1:t}^0)\}}$

- Drastic Change: ELL does not work, use OL or TE
  - OL: Neg. log of current observation likelihood given past

    $$OL = -\log [Pr(Y_t | Y_{0:t-1}, H_0)] = -\log[<\pi_{t|t-1}, \psi_t>]$$
  - TE: Tracking Error. If white Gaussian observation noise, TE $\approx$ OL

ELL & OL: Slow & Drastic Change

- ELL fails to detect drastic changes: large approx. error
  - Approximating posterior for changed system observations using a PF optimal for unchanged system: error large for drastic changes
  - OL relies on the error introduced due to the change to detect it

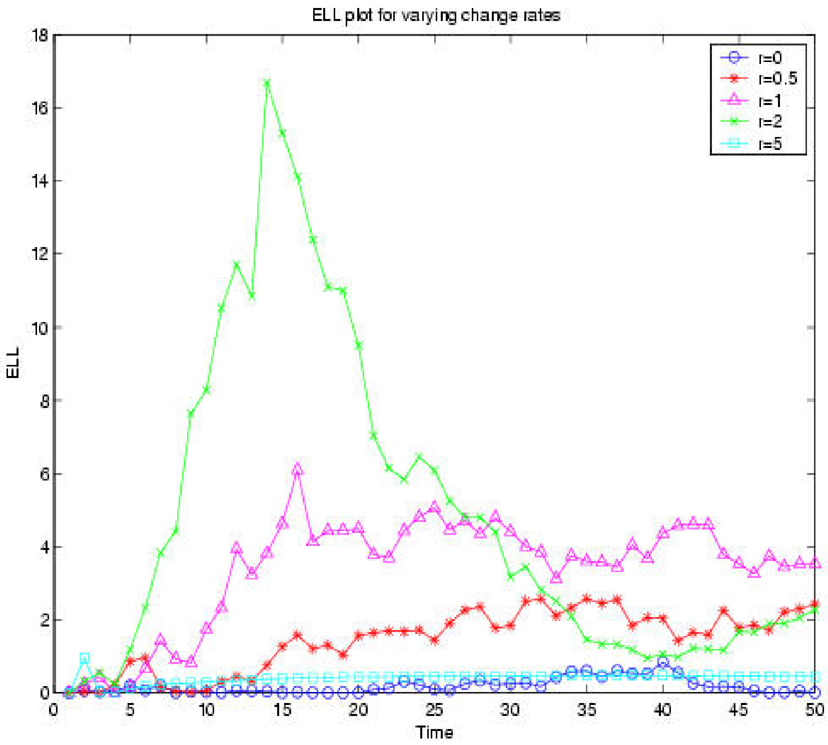
- OL fails to detect slow changes
  - Particle Filter tracks slow changes "correctly"
  - Assuming change till *t-1* was tracked "correctly" (error in posterior small), OL only uses change introduced at *t*, which is also small
  - ELL uses total change in posterior till time *t* & the posterior is approximated "correctly" for a slow change: so ELL detects a slow change when its total magnitude becomes "detectable"

- ELL detects change before loss of track, OL detects after
- OL relies on the fact that there is large error in posterior introduced by the change. ELL relies on the fact that PF is stable and hence error in posterior is small

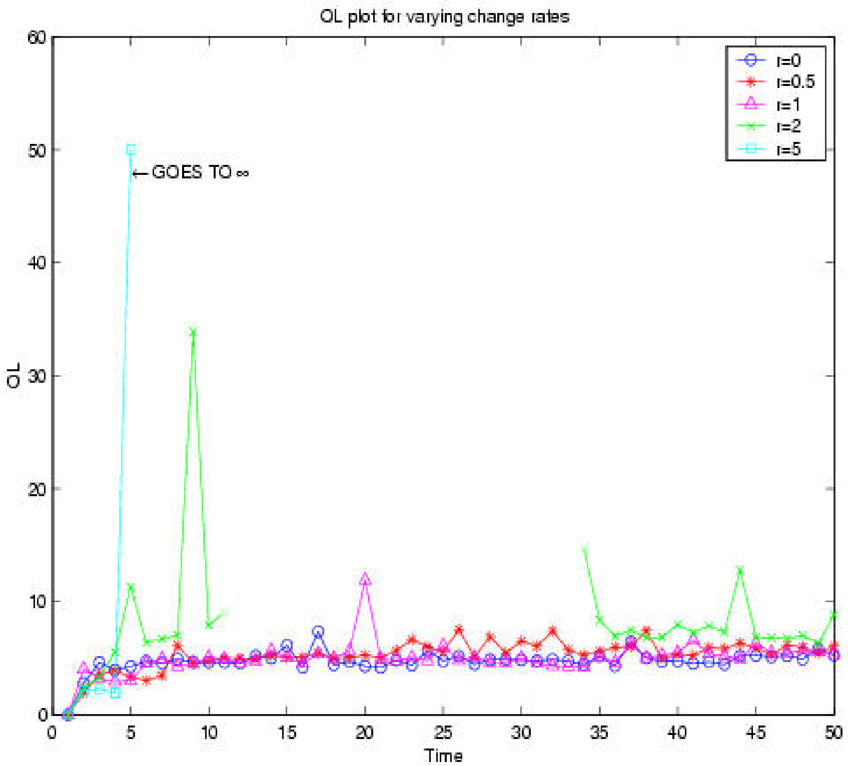
A Simulated Example

- $Y_t = X_t^3 + w_t$, w_t: truncated Gaussian
- $X_t = X_{t-1} + n_t \ + r_t\sigma, \ n_t \sim N(0,\sigma^2)$
- No change: $r_t = 0$, Change: r_t not equal to 0 from $t=5$ to $t=15$

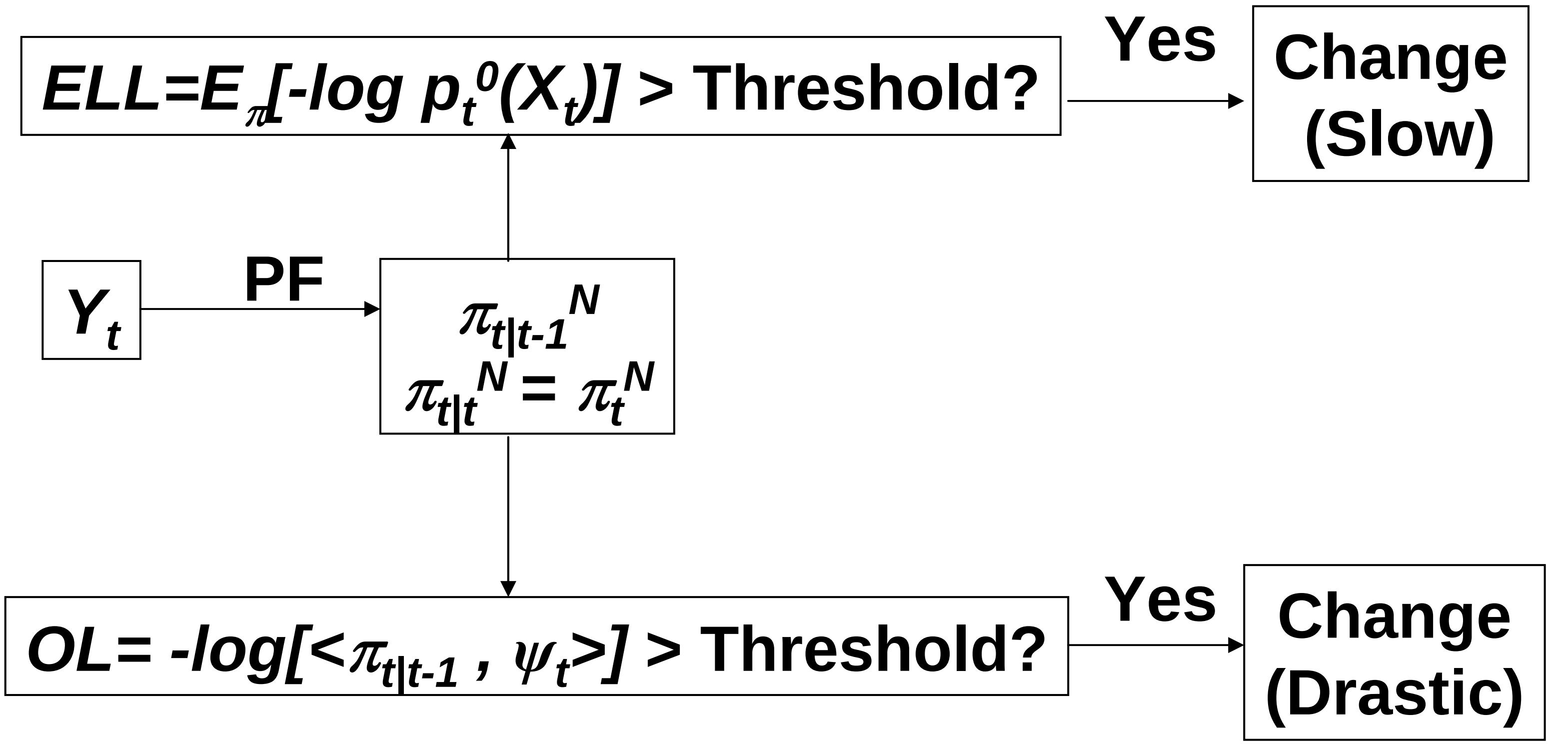
ELL



OL



Change Detection



Approximation Errors

- Total error < Bounding error + Exact filtering error + PF error
  - Bounding error: Stability results hold only for bounded fn's but LL is unbounded. So approximate LL by $min\{-log\ p_t^0(X_t),M\}$ (Huber's M-estimate of LL)
  - Exact filtering error: Error b/w exact filtering with changed system model & with original model. Evaluating $\pi_t^{c,0}$ (using Q_t^0) instead of $\pi_t^{c,c}$ (using Q_t^c)
  - PF Error: Error b/w exact filtering with original model & particle filtering with original model. Evaluating $\pi_t^{c,0,N}$ which is a Monte Carlo estimate of $\pi_t^{c,0}$

Complementary Behavior of ELL & OL

- ELL approx. error, $e_t^{c,0}$, is upper bounded by an increasing function of $OL_k^{c,0}$, $t_c < k < t$

$$e_t^{c,0} \leq \sum_{k=t_c}^{t} \exp(OL_k^{c,0})\,\omega(D_{Q,k}, {}^{1}\!/_{\varepsilon_k})$$

- Implication: Assume "detectable" change i.e. $ELL^{c,c}$ large
  - OL fails => $OL_k^{c,0}, t_c<k<t$ small => ELL error, $e_t^{c,0}$ small=> $ELL^{c,0}$ large => ELL detects
  - ELL fails => $ELL^{c,0}$ small =>ELL error, $e_t^{c,0}$ large => at least one of $OL_k^{c,0}, t_c<k<t$ large => OL detects

Applications

- Abnormal activity detection, Detecting motion disorders in human actions, Activity Segmentation
- Neural signal processing: detecting changes in stimuli
- Acoustics: Direction of Arrival estimation, detect when model changes
- Congestion Detection
- Video Shot change or Background model change detection
- System model change detection in target tracking problems without the tracker losing track