Centre for Robotics

Bayesian quickest change detection and insufficiently informative measurements

Prof Jason J. Ford

Science and Engineering Faculty

Queensland University of Technology

Talk contains work and concepts developed in collaboration with the following: John Lai, Onvaree Techakesari, Dragan Nesic, Tim Molloy, Justin Kennedy, Jasmin Martin, Troy Bruggemann and Valeri Ugrinovskii.

This work was (partial supported) by Boeing, Queensland government, and the ARC.
Acknowledge continued support from the Queensland University of Technology (QUT) through the Centre for Robotics.

Centre for Robotics

Contents

A talk in 3 parts

Introduction to Quickest Change Detection (QCD)

What it is useful for?

Examples of Past and future applications

QCD solved as an optimal stopping problem.

Deep drive into the role of modelling assumptions

QUT Centre for Robotics

What is Quickest Change Detection?

Quickly detecting change (or faults) is important in many disciplines including mechanical engineering, chemical engineering, aerospace engineering and automotive systems.

In many situations it is advantageous to monitor a signal of an engineering system for the purpose of quicky alerting in the event of a change in behaviour (e.g. it becomes broken).

Mathematically, consider sequentially observing a system process whose statistical distribution changes, at some unknown time ν, from

- having probability density $b^1(.)$ to
- having probability density $b^2(.)$.

Quickest Change Detection (QCD) is an optimal stopping problem where the task is to quickly declare the change has occurred to minimize detection delay subject a false alarm criteria.

Change in the statistics

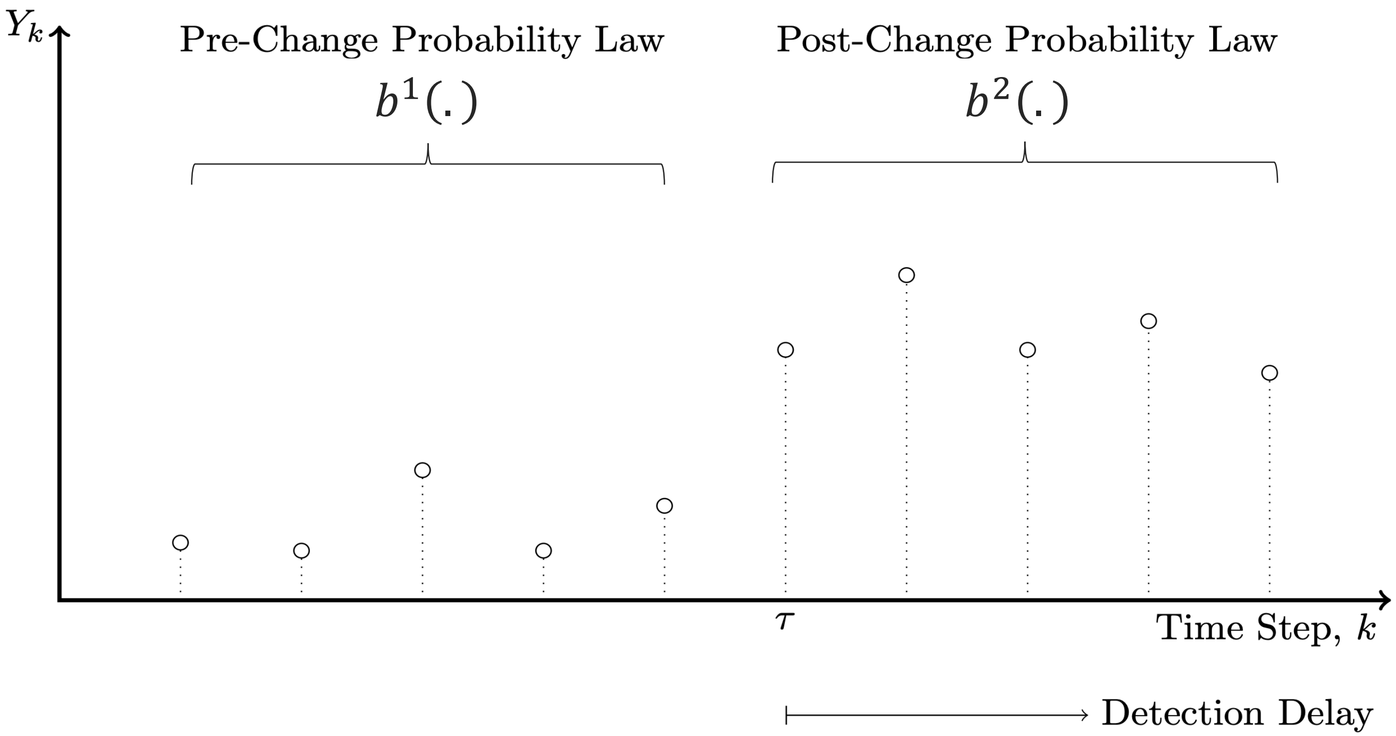
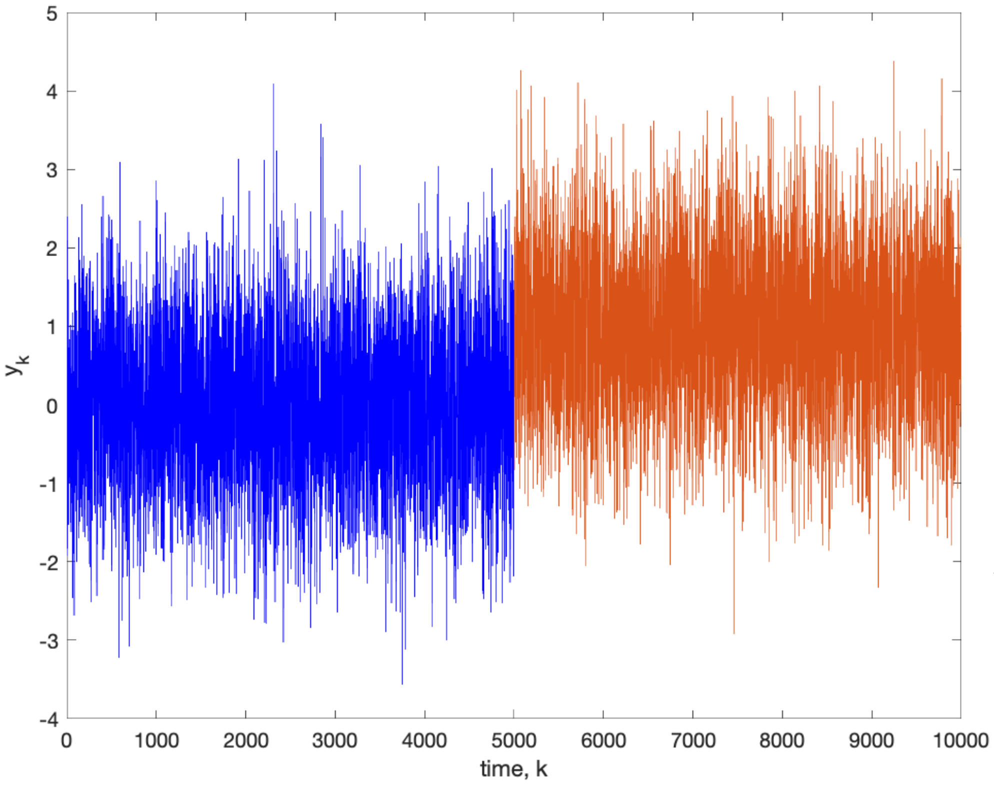


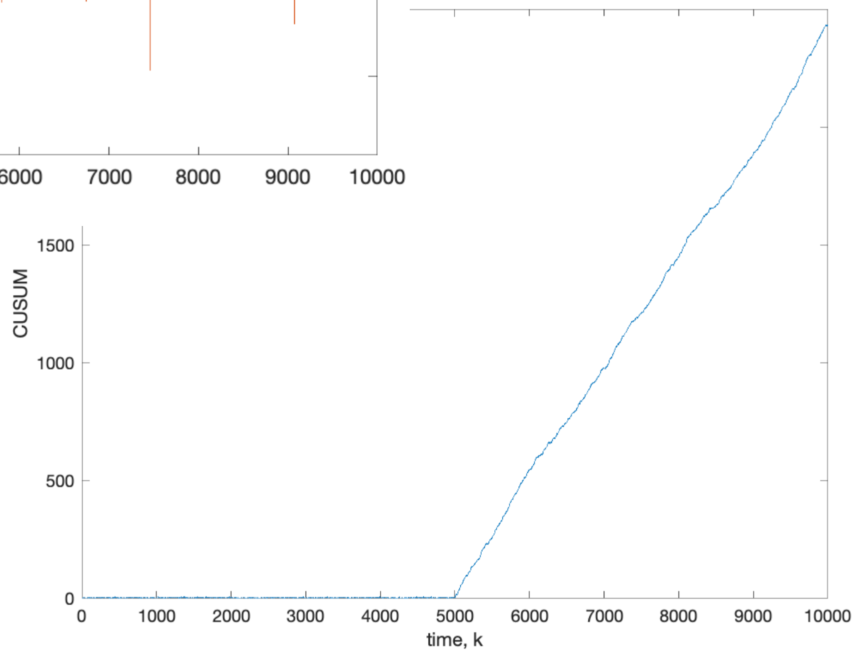
Image credit:
Molloy

Very general problem that can be considered in any application with sequential data

Toy illustrative example

- The signal y_k is
  - pre change $b^1(.)$: unit variance Gaussian r.v. with mean =0, and
  - post change $b^2(.)$: unit variance Gaussian r.v. with mean =1.
- You are sequentially watching y_k and want to alert a change with short delay whilst managing risk of false alarm.
- Here change occurred a 5000 (post change marked in red for emphasis).
- CUSUM QCD approach clearly detects.





Example: UAS fault detection

- Building fault resilient autonomous system like requires system with ability to self-detect fault or anomaly conditions and switch to recovery model.
- Aircraft avionics have a wealth of information: GPS/INS navigation data, airspeed etc.
- Currently investigate what measurements and what fault detection tools.
- Nice dataset to play with Air Lab Failure and Anomaly (ALFA) Dataset:
  - http://theairlab.org/alfa-dataset/

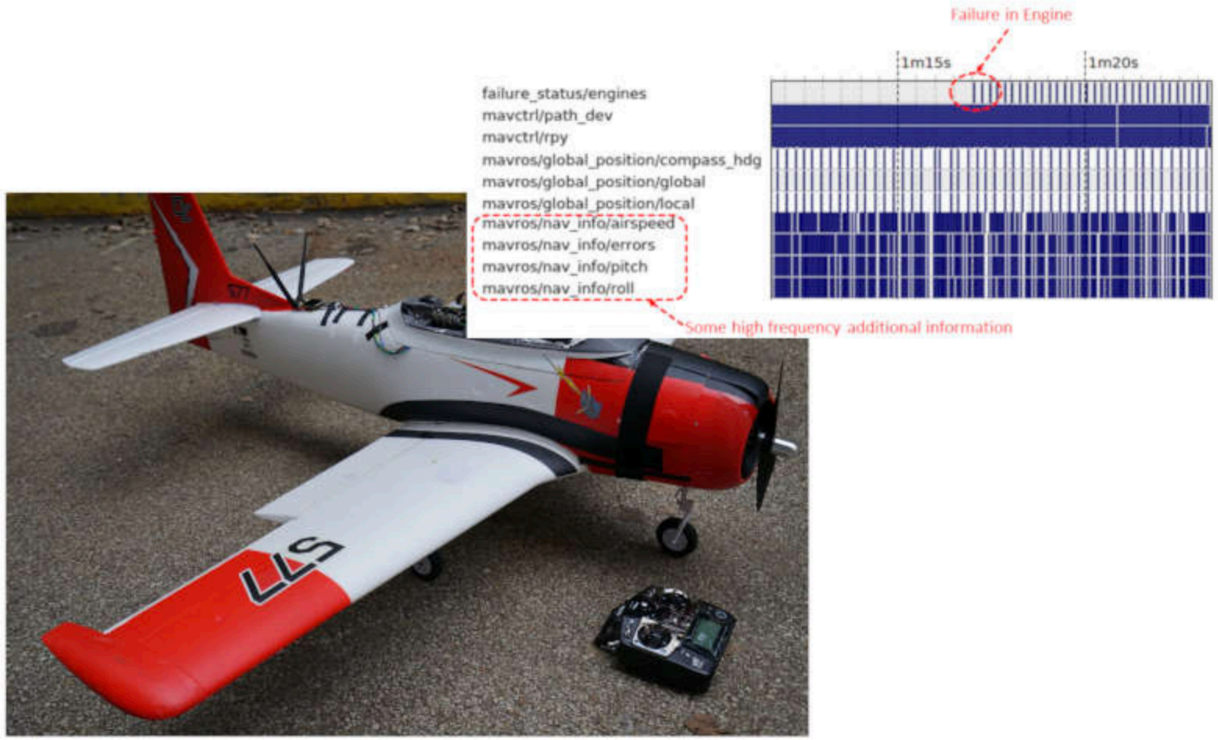
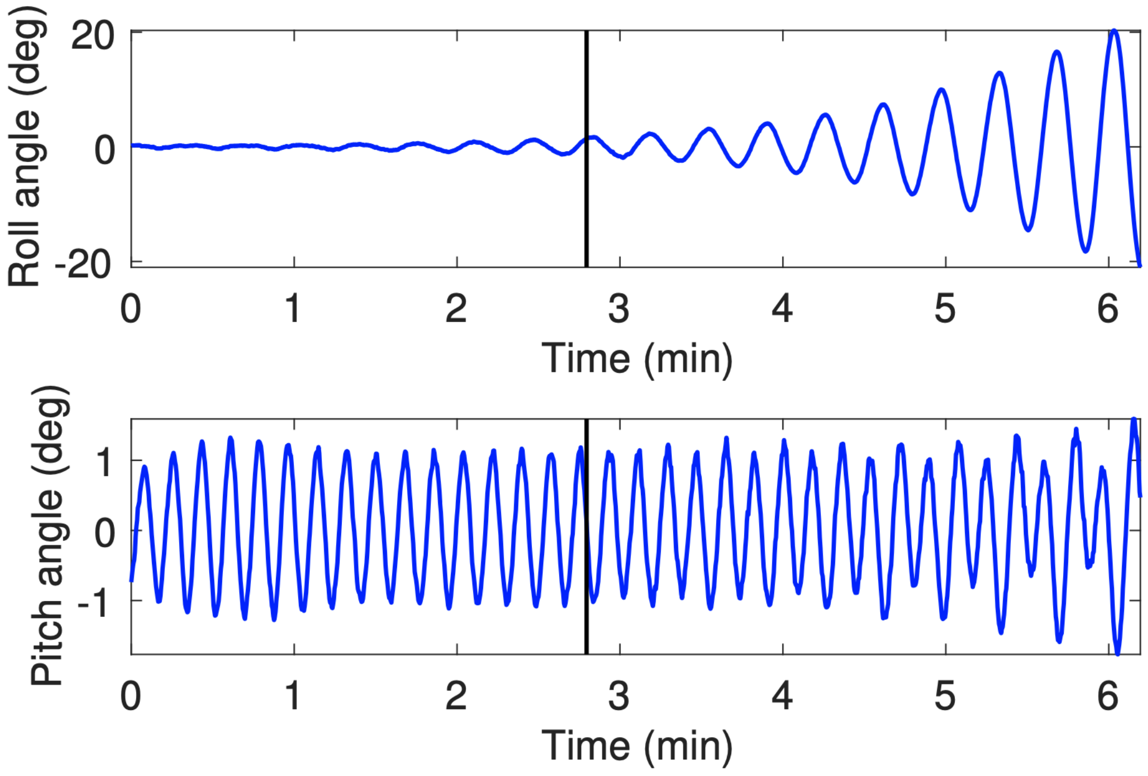


Image credit: http://theairlab.org/alfa-dataset/

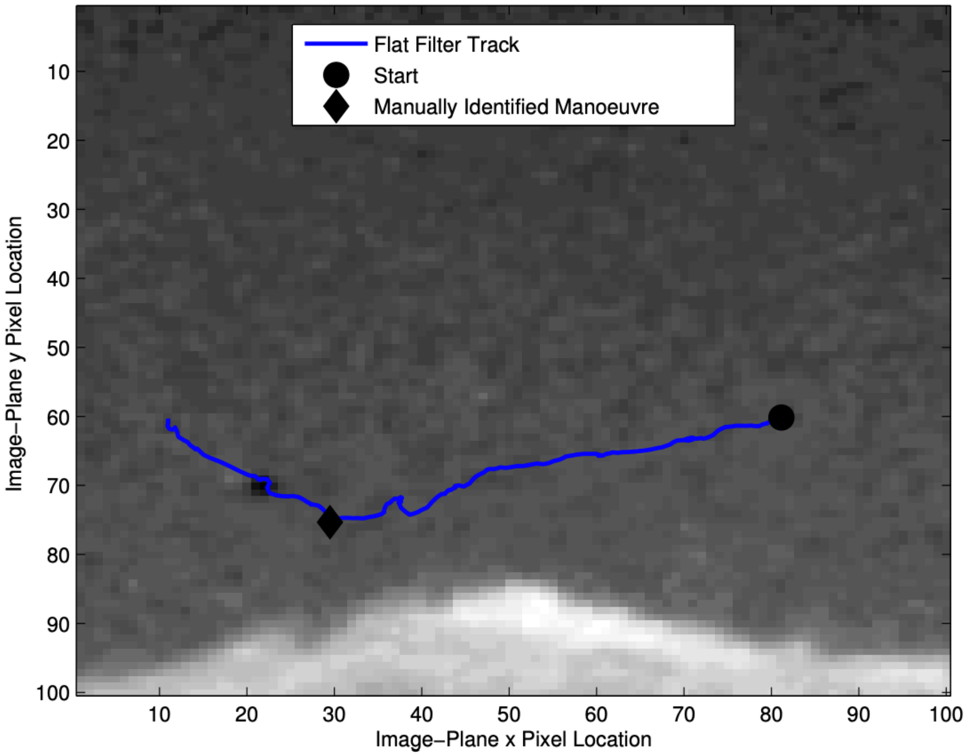
Detection of parametric roll resonance (Maritime system) application

- Parametric roll resonance is a phenomenon where the wave encounter frequency is twice the natural roll frequency of the ship can lead to unsafe roll motion amplification.
- Quick detection would allow preventative action to avoid capsizing or damage to the ship and crew.
- There are effects present in roll, pitch and heave axis.



J. Kennedy, J. Ford, T. Perez and F. Valentinis, “Detection of parametric roll resonance using Bayesian discrete-frequency model selection”, CAMS 2018.

Aircraft heading change detection



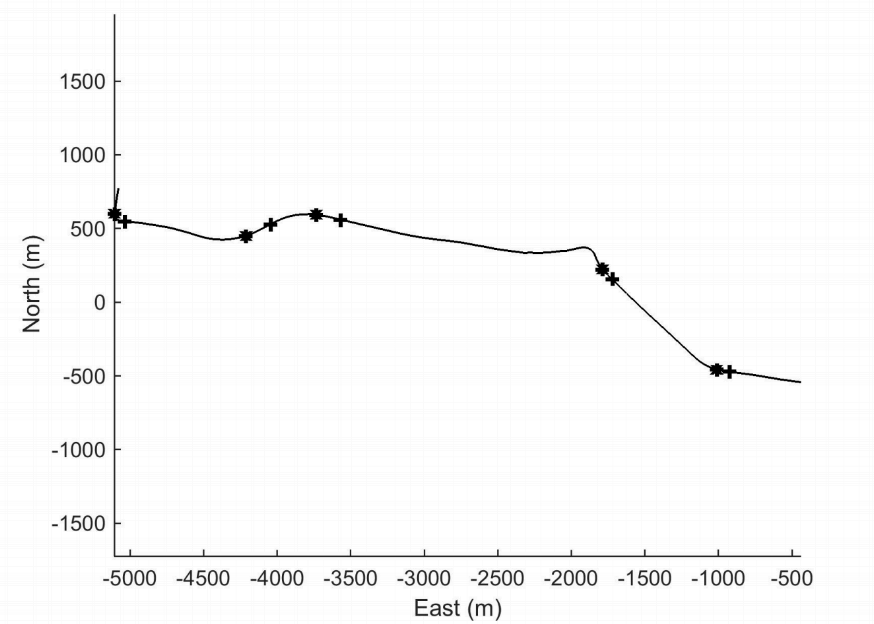
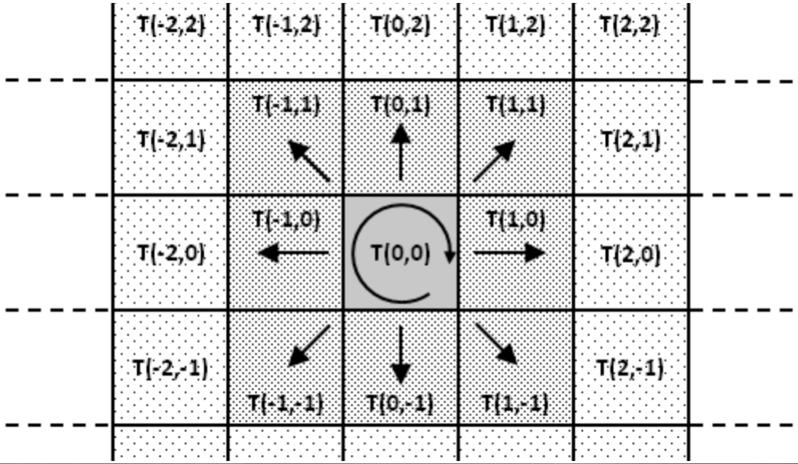


Fig. 3. Ground track of vehicle travelling East to West with true changes + and change detections *.

Image Based:
Timothy L. Molloy and Jason J. Ford, “HMM Relative Entropy Rate Concepts for Vision-based Aircraft Manoeuvre Detection”, AuCC 2013.

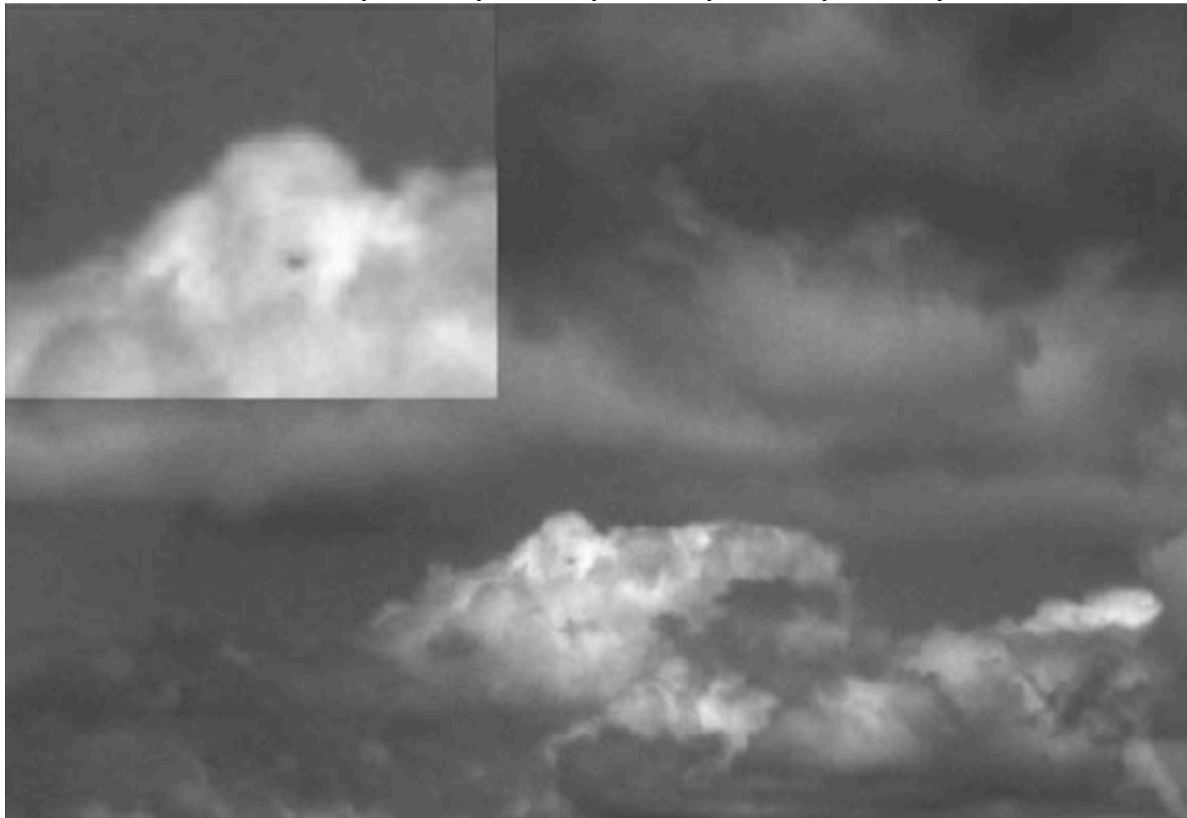
Heading info based:
Troy S. Bruggemann and Jason J. Ford, “Coordinated Change Detection for UAV Formations”, AuCC 2016.

Lai, et al., Relative Entropy Rate Based Multiple Hidden Markov Model Approximation, IEEE Trams SP, 2010.

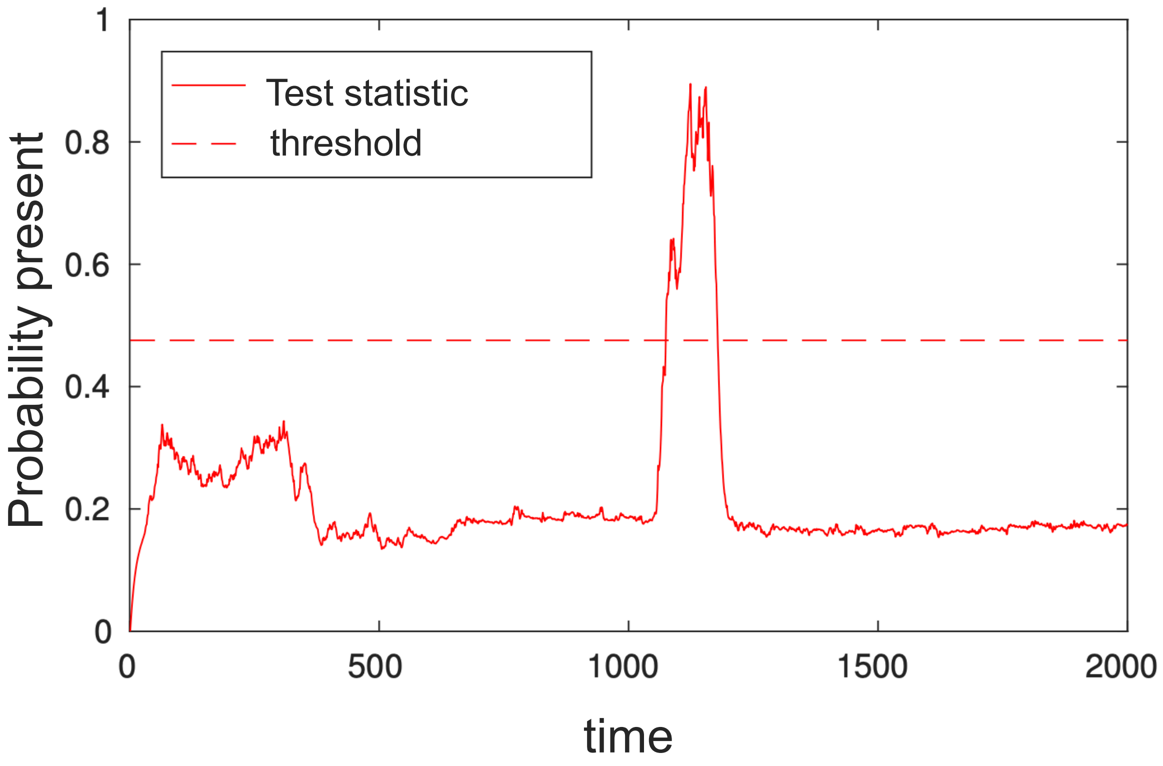


Vision based aircraft detection

- UAS need to replicate human pilot's role in mid-air collision avoid, detect 12.5s before.
- UAS have limited carriage capability.
- Vision seems most viable path.
- But, at this range, aircraft occupies small number of pixels, and lots of aircraft like artefacts.
- Posed as a QCD problem
  - consider no aircraft $b^1(.)$ and aircraft present $b^2(.)$ densities. Very low SNR between these densities.
  - We include within a Hidden Markov Model (HMM) to model temporal characteristics (i.e. that aircraft will persist across frames).



Example of test statistic from aircraft detection data



- This data is from a real vision-based aircraft detection sequence.
- QCD techniques used to trigger an alert (this is a low SNR event so QCD important tool).
- The threshold (dash line) provides a tradeoffs delay and false alarms.
- Notes:
  - A slightly target looking object relates to earlier cause a slight raise in the test statistic.
  - The large increase corresponds to (true case) of an aircraft present.
  - The test statistic goes down once aircraft has passed.

QUT Centre for Robotics

Improved detection of weak signals (SNR boosting)

In 2017-9 developed better change detection techniques for weak signal cases (with Jasmin Martin and Tim Molloy).

The result: Principled QCD and models for low SNR vision-based aircraft detector.

The result: detection now >2.3 km, low SNR. Now exceeding human level performance.

Plus deep learnt boosting by few hundred metres.

Quickest Detection of Intermittent Signals With Application to Vision-Based Aircraft Detection

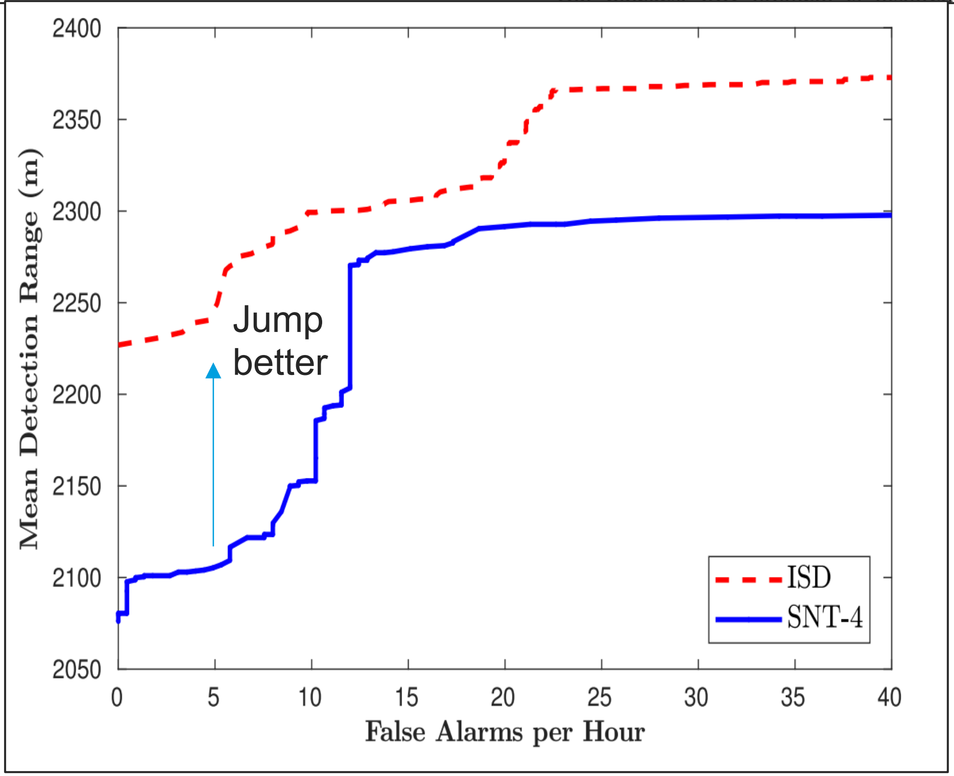
Jasmin James, Jason J. Ford, and Timothy L. Molloy

***Abstract*—In this brief, we consider the problem of quickly detecting changes in an intermittent signal that can (repeatedly) switch between a normal and an anomalous state. We pose this intermittent signal detection (ISD) problem as an optimal stopping problem and establish a quickest ISD rule with a threshold structure. We develop bounds to characterize the performance of our ISD rule and establish a new filter for estimating its detection delays. Finally, we examine the performance of our ISD rule in both a simulation study and an important vision-based aircraft detection application where the ISD rule demonstrates improvements in detection range and false alarm rates relative to the current state-of-the-art aircraft detection techniques.**

***Index Terms*—Bayesian quickest change detection, change detection, filtering, sense and avoid.**

posed in the past decade. Incipient fault detection seeks to identify slow drifts in system parameters [7]; multicyclic detection seeks to identify a distant change in a stationary regime where detection procedures are reset after each false alarm [8]; quickest transient detection seeks to identify a change that occurs once for a period of time and then disappears [9], [10]; and quickest detection under transient dynamics that seeks to identify a persistent change which does not happen instantaneously, but after a series of transient phases [11]. In this brief, we consider a new quickest ISD problem where a change can repeatedly appear and disappear over time.

Our quickest ISD problem is inspired by the important



A range of change detection approaches compared

- Fault detection (FD). Filter based on model of fault-free system
  - Chi-squared test on the filter residue
  - Test on the model evidence (filter model posterior)
- Classic Bayesian QCD (BQCD). Non-ergodic Bayesian change model
  - Shiryaev rule (test on the no-change posterior)
- Intermittent Signal Detection (ISD). Ergodic Bayesian change model
  - Test on a signal's posterior
- Non-Bayesian QCD (NBQCD). Non-random change event model
  - The famous CUSUM test (likelihood ratio type quantity)
  - There are two criteria: Lorden and Pollak
- There are robust versions of these (i.e. versions with minimax criteria)

QUT Centre for Robotics

When to use what technique?

The simplest approach that might work, sorted by SNR and before/after pdf knowledge.

From simplest to more powerful

- Fault detection (FD)
- Bayesian QCD (BQCD)
- Intermittent Signal Detection (ISD)
- Non-Bayesian QCD (NBQCD)

SNR (residue)	pdfs known	pdfs partially known	Intractable models*
> 6 dB (High)	anything	anything	FD
2 to 6 dB (Medium)	FD	FD	FD
-10 to 2 dB (Low)	BQCD	Robust BQCD	open problem
< -10 dB (Very low)	NBQCD or ISD	Robust NBQCD or ISD	open problem

Note:

SNR (residue) means the residue SNR after you in exploited all the known signal structure.

*Not clear now to use BQCD, ISD or NBQCD on intractable models. Some work done.

Centre for Robotics

(Bayesian) QCD as an optimal stopping problem

The hire problem

- Bayesian Quickest Change Detection is an optimal stopping problem.
- Optimal stopping problems are an important sub-class of optimal control problems.
- In a general informal sense, their solutions can be described by discrete time dynamic programming equations; which rarely have closed-form solutions.

Bayesian Quickest Change Detection

Here “Bayesian” refers to the model of change time event.

In the classic Bayesian version of the problem, **we assume the change time $\nu \geq 0$ is a random variable with geometric prior, $\pi_k = (1-\rho)^k \rho$, with $\rho \in (0,1)$.** Also assume, that at the change time ν the statistical distribution of a sequential observed process changes from

- i.i.d random variable with probability density $b^0(.)$ to
- i.i.d random variable with probability density $b^1(.)$.

The $P_\pi(.)$ and $E_\pi[.]$ that follow are measures and expectation operations arising from assumed geometry prior.

Bayesian Quickest Change Detection

We are interested at designing a stopping rule $\tau \geq 0$. For a considered stopping rule τ, let us define PFA (probability of false alarm) as

$$PFA(\tau) = P_{\pi}(\tau < \nu).$$

We are interested in designing a stopping rule τ (declaring a change) that solves the optimization problem

$$\inf_{PFA(\tau)\leq\alpha} E_{\pi}[\tau - \nu | \tau > \nu].$$

That is, minimizing the average detection delay subject to a constraint on false alarm performance.

This problem can be re-written as unconstrained optimization of $\mathrm{J}(\tau) = cE_{\pi}[(\tau-\nu)^{+}] + PFA(\tau)$,

where $(\tau-\nu)^{+} = \max(\tau-\nu, 0)$.

Dynamic programming sketch

Importantly, we can equivalently write our BQCD solution as the stopping rule τ that optimises the cost

$$J(\tau, \widehat{X}_0^1) = E_\pi[c\Sigma_{\ell=0}^{\tau-1}(1 - \widehat{X}_\ell^1) + \widehat{X}_\tau^1 | \widehat{X}_0^1]$$

where $\widehat{X}_0^1$ is the prior that no change has occurred at time 0, and $\widehat{X}_k^1 \triangleq P_\pi(no\ change|y_o, y_1, \ldots, y_k)$.

We can define the value function as, for $\hat{X} \in [0,1]$,

$$V(\hat{X}) \triangleq inf_\tau J(\tau, \hat{X}).$$

In this case $V(\hat{X})$ satisfies a variational inequality type discrete recursion

$$V(\hat{X}) = \min\left(c(1-\hat{X}) + E\left[V\left(\hat{X}^+(\hat{X},y)\right)\middle|\hat{X}\right], \hat{X}\right)$$

Red =stop,
Green = continue

where $\hat{X}^+(\hat{X}, y)$ is the 1 step ahead of the conditional pdf of the no change event given the measurement y (i.e. one step ahead of the filter).

The value function encodes the optimal solution, but it is challenging to solve this recursion directly.

Solution: A threshold test

Remarkably, it can be shown that the optimal rule solving the BQCD problem can be written as a simple threshold check on the no-change posterior $\hat{X}_k^1 \triangleq P_\pi(no\ change | y_o, y_1, \dots, y_k)$. That is, the optimal rule is

$$\tau^* = \inf\{k : \hat{X}_k^1 \leq h\}$$

where h is selected to control PFA.

This $\hat{X}_k^1$ can be exactly computed using simple filter recursion as the observations arrive, and the complicated DP equation avoided. The posterior filter is a simple 2 state HMM filter.

Optimal rule is a threshold test initially established by Shiryaev (1963), but an elegant modern treatment can be found in V. Krishnamurphy, Partially observed Markov decision processes, Cambridge University Press, 2016.

Convenient recursion for no-change posterior

The no-change posterior can be written as the scalar recursion

$$\hat{X}_k^1 = N_k(1-\rho)b^1(y_k)\hat{X}_k^1$$

where normalization factor given by

$$N_k^{-1} = b^2(y_k) + (1-\rho)(b^1(y_k) - b^2(y_k))\hat{X}_{k-1}^1.$$

Proof: Let 1st Markov chain state denote no change yet and 2nd state denote change has occurred. Build left-to-right transition matrix with ρ etc. elements, $b^0(.)$ etc. in observation equation. Then $\hat{X}_k^1$+ $\hat{X}_k^2$=1 means simple algebra on the hidden Markov model filter recursions leads to the above scalar recursion.

Helps make the posterior recursion look somewhat scalar and linear and open to analysis.

Jason J. Ford, Jasmin James, Timothy L. Molloy, On the informativeness of measurements in Shiryaev's Bayesian quickest change detection, Automatica, 2020.

Deep Drive

The role of model assumptions

Jenkins (1976) “All models are wrong, but some are useful”

QUT Centre for Robotics

Empirical science's basic principle is that knowledge should be extracted from observations

One debate is the role of prior knowledge and models in this process

My journey with model error started in ‘98

What do your estimates mean when your model is wrong?

– from my PhD Thesis, 1998.

This simple question has inspired much of my career!

7.3 Future Research Directions

In this thesis several open questions have arisen about identification of hidden Markov models.

1. If the “true model in the model set” assumption is relaxed then what do our model estimates mean? Will our estimates be maximum likelihood estimates? Can maximum likelihood estimation on a set of HMMs of a particular model order be performed? Consider estimation of an HMM of order N in the following situations:

   (a) the data is generated by an HMM of an order different than N.

   (b) the data is generated by a linear system.

   (c) the data is generated by a non-linear system.

2. Is there a concept of a minimal representation? When will high-order HMMs be ‘well’ approximated by low-order HMMs?

Ford, Jason (1998) *Adaptive hidden Markov model estimation and applications.* PhD thesis, Australian National University.

QUT Centre for Robotics

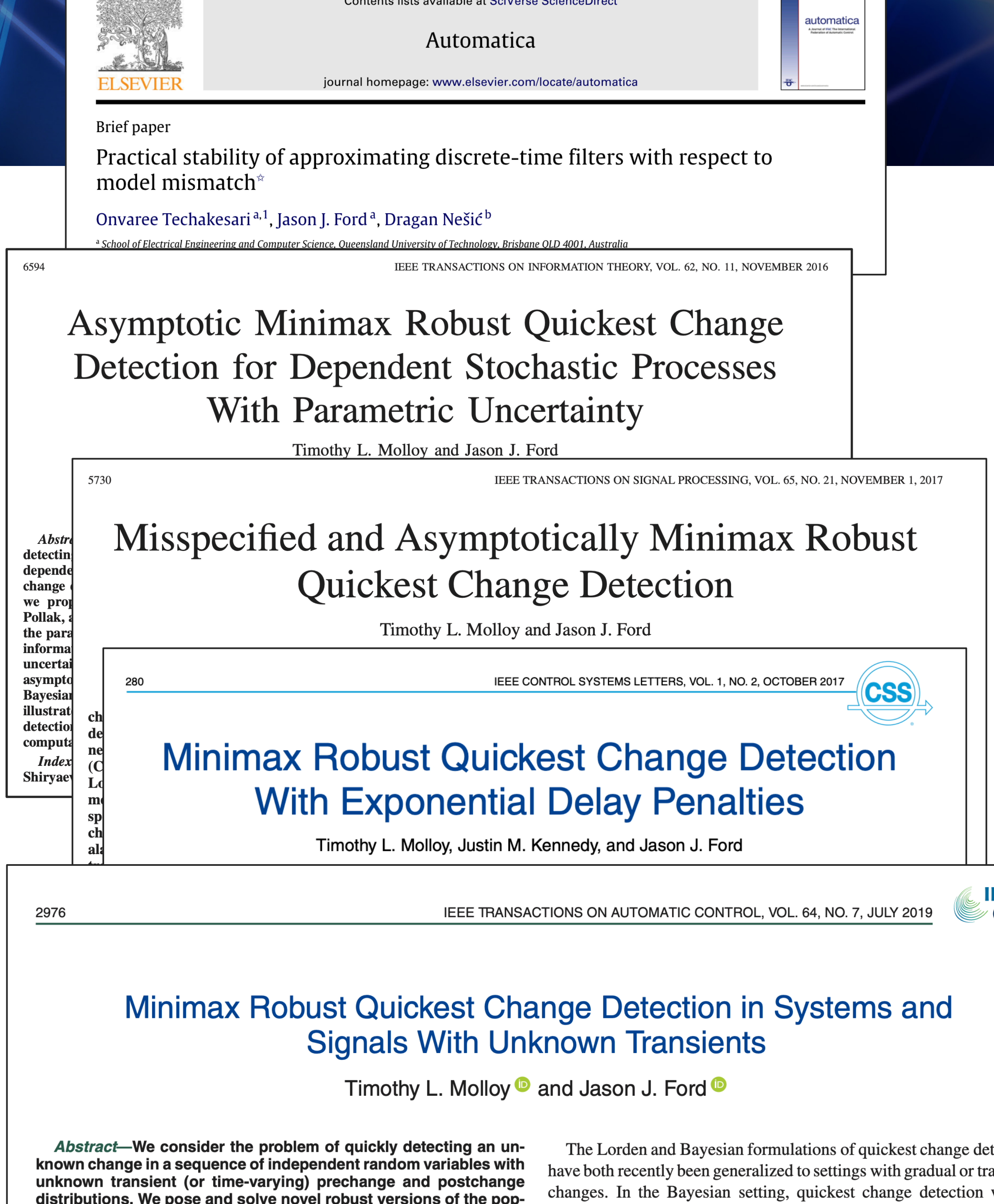
Quantify and managing error via information theoretic tools

Aim: investigate relative entropy (AKA Kullback Leibler distance) to quantify model error.

Program of research (2011-19):

- Techakesari et al., Automatica 2011 (first progress on my PhD question – thought I was done).
- Then Molloy, et al. in the "* minimax robust quickest change detection *" series of IEEE papers† appearing in
  - IEEE Trans IF 2016,
  - IEEE Trans SP 2017,
  - IEEE CSL 2017, and
  - IEEE Trans AC 2019.

† contain much other goodness (RE only small part).


Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Practical stability of approximating discrete-time filters with respect to model mismatch

Onvaree Techakesari [a,1], Jason J. Ford [a], Dragan Nešić [b]

6594 IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 62, NO. 11, NOVEMBER 2016

Asymptotic Minimax Robust Quickest Change Detection for Dependent Stochastic Processes With Parametric Uncertainty

Timothy L. Molloy and Jason J. Ford

5730 IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 65, NO. 21, NOVEMBER 1, 2017

Misspecified and Asymptotically Minimax Robust Quickest Change Detection

Timothy L. Molloy and Jason J. Ford

280 IEEE CONTROL SYSTEMS LETTERS, VOL. 1, NO. 2, OCTOBER 2017

Minimax Robust Quickest Change Detection With Exponential Delay Penalties

Timothy L. Molloy, Justin M. Kennedy, and Jason J. Ford

2976 IEEE TRANSACTIONS ON AUTOMATIC CONTROL, VOL. 64, NO. 7, JULY 2019

Minimax Robust Quickest Change Detection in Systems and Signals With Unknown Transients

Timothy L. Molloy and Jason J. Ford

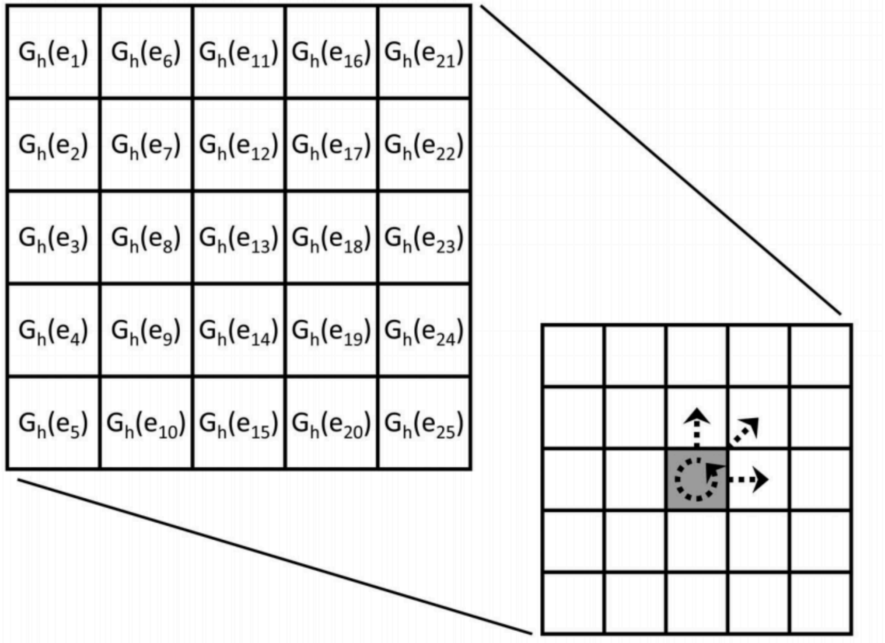
What didn’t we know in 2018?

Since 2008 we had been using a physically unrealistic motion model within our HMM filter engine to make our vision-based aircraft detector work.

Bad model but ergodic and detector worked. Was a hack! (BQCD involves a left-to-right HMM, so not ergodic).

In 2019 we developed new ergodic QCD for our detector, better physical match, and achieved better detection performance.

Nice, but still didn’t have a mathematical explanation why BQCD didn’t work!



We enumerate image pixel locations $G_h(e_i)$ for $i = 1, 2, \ldots, N$ in a column-wise manner when considering the image frame as a 2D pixel grid

The transition probability matrix characterises the way the target moves between one image frame and the next.

Transitions that would otherwise cross the image boundary will “wrap-around” to the opposite image boundary

Lai, Et al. Vision-Based Estimation of Airborne Target Pseudobearing Rate using Hidden Markov Model Filters, IEEE AES 2013.

What was wrong with BQCD?

- In ’94-5 Bengio showed RNNs and left-to-right HMMs have difficulty temporal modelling because of vanishing dependence in time!
- Further, using HMMs Bengio showed more structure allowed learning longer dependences.
- We had observed that Bayesian QCD didn’t work in low SNR (Bayesian QCD involves left-to-right HMMs 🤔).
- What is the interplay between low SNR and left-to-right models?

We now show the consequences of robust latching; i.e., vanishing gradient.

Theorem 4: If the input u_t is such that a system remains robustly latched on attractor X after time 0, then $\frac{\partial a_t}{\partial a_0} \to 0$ as $t \to \infty$.

Proof: See the Appendix.

Bengio, Et al., Learning long-term dependencies with gradient descent, IEEE Trans NN. 1994

Diffusion of Credit in Markovian Models

Yoshua Bengio*
Dept. I.R.O., Université de Montréal,
Montreal, Qc, Canada H3C-3J7
bengioy@IRO.UMontreal.CA

Paolo Frasconi
Dipartimento di Sistemi e Informatica
Universitá di Firenze, Italy
paolo@mcculloch.ing.unifi.it

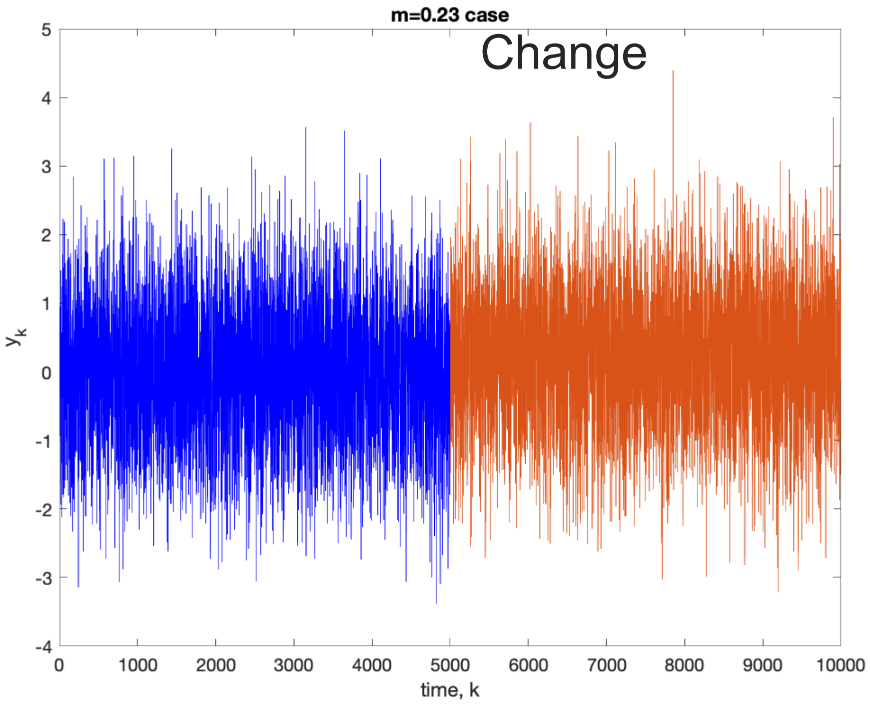
Abstract

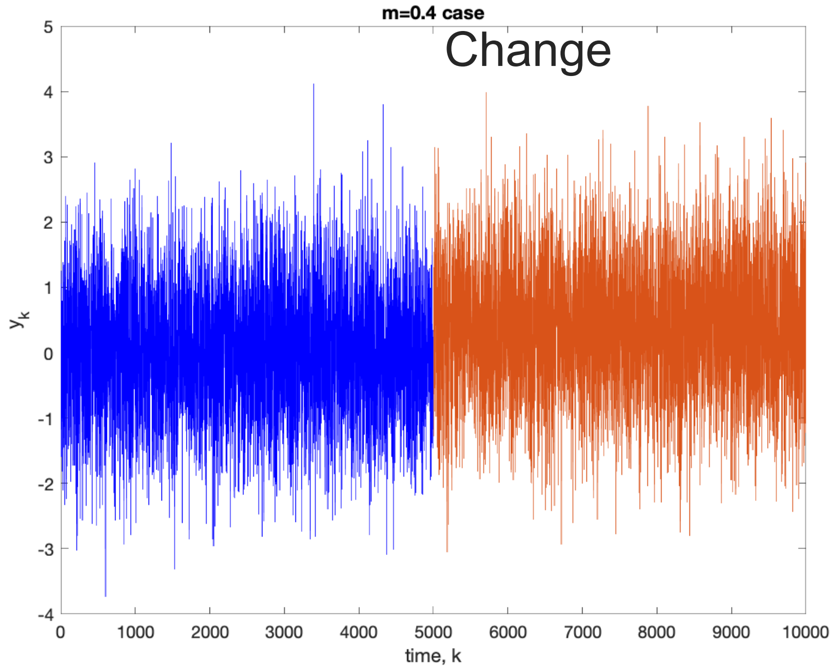
This paper studies the problem of diffusion in Markovian models, such as hidden Markov models (HMMs) and how it makes very difficult the task of learning of long-term dependencies in sequences. Using results from Markov chain theory, we show that the problem of diffusion is reduced if the transition probabilities approach 0 or 1. Under this condition, standard HMMs have very limited modeling capabilities, but input/output HMMs can still perform interesting computations.

Bengio, Et al., Diffusion of Credit in Markovian Models, NIPS, 1995.

Example: Gaussian models

Let us consider $b^1(y_k)$ and $b^2(y_k)$ unit variance Gaussian with slightly different post change means

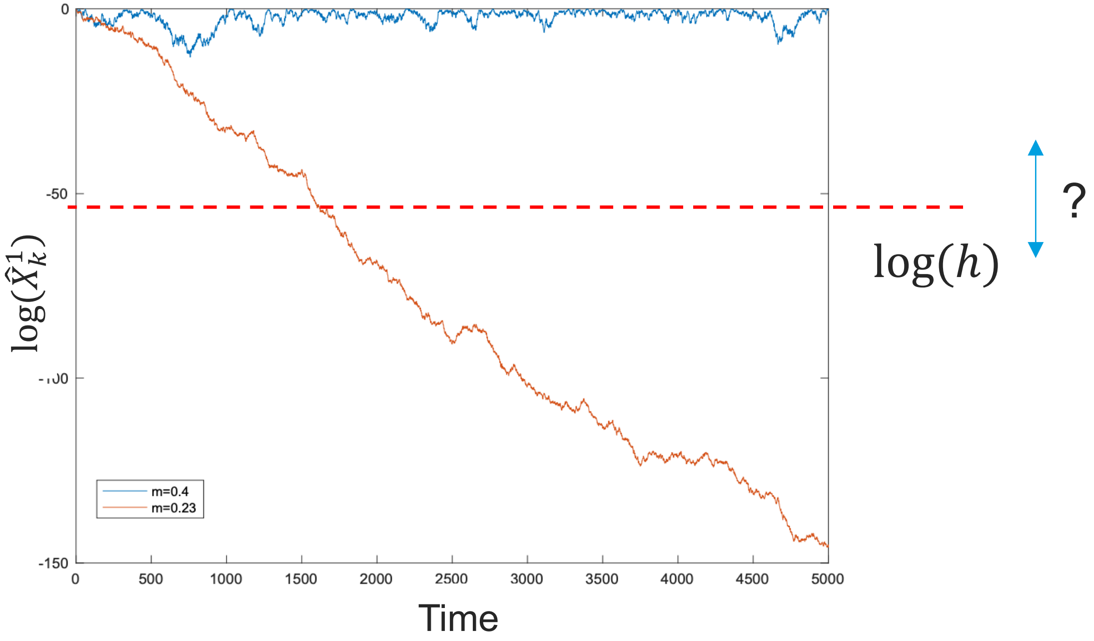




Example: Gaussian models

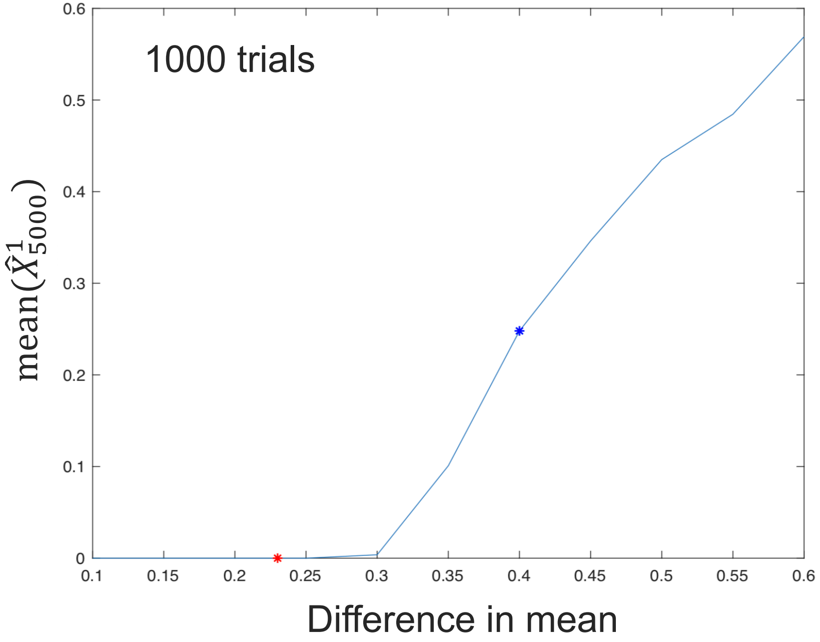
Consider $\rho = 0.05$. Simulation before the change event. Red is case with post change mean shift of 0.23 (not informative) and Blue is case with post change mean shift of 0.4 (expected case)

There is a critical value $\mathrm{m_c}$ below which the detector breaks.



Below $\mathrm{m_c} = 0.32$ the behavior of $\hat{X}_{5000}^1$ changes becoming increasingly convinced a change has occurred?

Can't meaningfully set a detection threshold h



What is the mechanism here?

For some insight consider the case when $\mathrm{b}^1(\mathrm{y_k}) = \mathrm{b}^2(\mathrm{y_k})$ (no measurement information). Here no change posterior recursion become exponential decay of

$$\hat{X}_k^1 = (1-\rho)^k$$

and hence the test statistic becomes increasing confident change has occurred (even when no change).

Informally, this $(1-\rho)^k$ mechanism is dominating $\hat{X}_k^1$ and when $\mathrm{b}^1(\mathrm{y_k})$ and $\mathrm{b}^2(\mathrm{y_k})$ are not sufficiently different (the prior model is too strong to overcome).

Interestingly, there is a critical point where suddenly the measurements are strong enough!

Definition: a weak practical super-martingale

A super-martingale has the property $E_\pi[\mathrm{X_{k+1}}|\mathrm{X_k}] \leq \mathrm{X_k}$ (does not trend upwards). We introduced a new super-martingale concept.

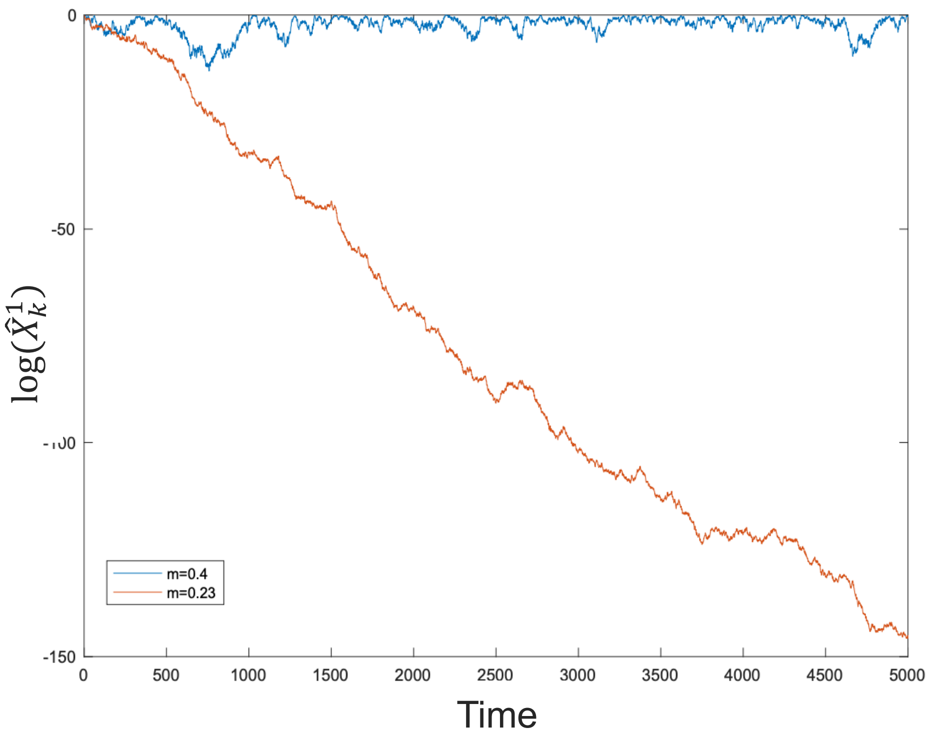
The no change posterior $\log(\hat{X}_k^1)$ is a *weak practical super-martingale* if for any arbitrarily small $\delta_p > 0$ there exists a $h_s > 0$ such that if $\hat{X}_k^1 < h_s$ implies

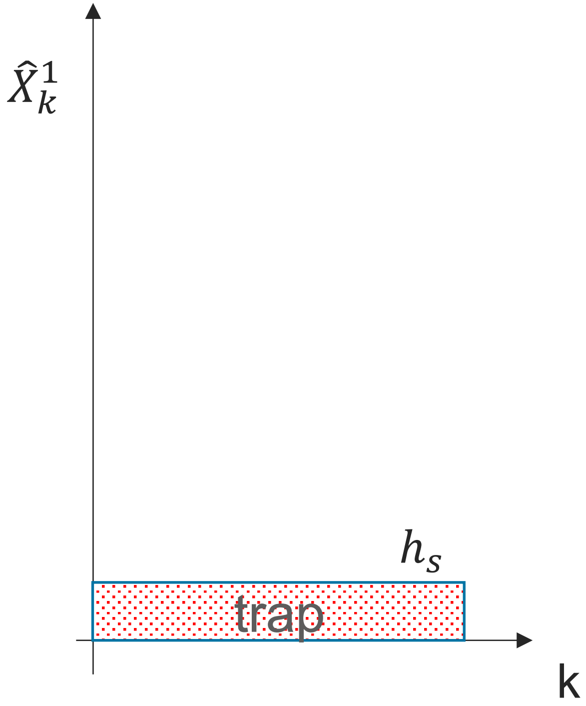
$$P_\pi\left(\text{for all } n \geq k, \underbrace{E_\pi\left[\log(\hat{X}_{n+1}^1)\middle|\log(\hat{X}_n^1)\right] < \log\left(\hat{X}_n^1\right)}\right) > 1 - \delta_p$$

Basically, to any level of probabilistic certainty, $1 - \delta_p$, there existed practical interval $\hat{X}_k^1 < h_s$ that $\hat{X}_k^1$ trends down (in a weak sense).

Note: “weak” as holding with $P_\pi > 1 - \delta_p$, “practical” as holding for any “$\delta_p > 0$ there exists a $h_s > 0$”, and “super-martingale” the green part.

A weak practical super-martingale when an interval trap $\hat{X}_k^1 < h_s$ exists





If you reach the trap, $< 1 - \delta_p$ probability of not trending upwards.

Lemma: A no change posterior update bound

Writing the log of the recursion in $\hat{X}_k^1$ is useful

$$\log(\hat{X}_k^1) = \log M_k + \log(\hat{X}_{k-1}^1)$$

where $M_k \triangleq N_k(1-\rho)b^1(y_k)$ as properties of M_k let us start to access the behavior of $\hat{X}_k^1$ via an additive mechanism.

The following $\hat{X}_{k-1}^1$ dependent bound on M_k exists. For any $\delta > 0$, then is a $\mathrm{h}_\delta > 0$ such that for any $\hat{X}_{k-1}^1 < h_\delta$ we have

$$E_\pi\left[\log M_k \,|\hat{X}_{k-1}^1\right] < \log(1-\rho) + \mathrm{D}\left(\mathrm{b}^1(\mathrm{y_k})\middle|\mathrm{b}^2(\mathrm{y_k})\right) + \delta$$

Note: Relative entropy $\mathrm{D}\left(\mathrm{b}^1(\mathrm{y_k})\middle|\mathrm{b}^2(\mathrm{y_k})\right)$ is a pseudo distance measure between densities and $\log(1-\rho) < 0$.

Lemma proof sketch

Recall:

$M_k \triangleq N_k(1-\rho)b^1(y_k)$

Lemma: The following $\hat{X}^1_{k-1}$ dependent bound on M_k exists. For any $\delta > 0$, then is a $\mathrm{h}_\delta > 0$ such that for any $\hat{X}^1_{k-1} < h_\delta$ we have

$$E_\pi\left[\log M_k \,|\hat{X}^1_{k-1}\right] < \log(1-\rho) + \mathrm{D}\left(\mathrm{b}^1(\mathrm{y_k})\middle|\mathrm{b}^2(\mathrm{y_k})\right) + \delta$$

Proof sketch:

- We re-express $\log(\mathrm{N}_k) = -\log\left(\mathrm{b}^2(\mathrm{y_k})\right) + \gamma_k$ where γ_k can be monotonically bounded by $\hat{X}^1_{k-1}$.
- We can then re-express $E_\pi\left[\log M_k \,|\hat{X}^1_{k-1}\right] = \log(1-\rho) + \mathrm{E}_\pi\left[\log\left(\frac{\mathrm{b}^1(\mathrm{y_k})}{\mathrm{b}^2(\mathrm{y_k})}\right)\middle|\, \hat{X}^1_{k-1}\right] + \delta$ where δ can be monotonically bounded by $\hat{X}^1_{k-1}$.
- We then show $\mathrm{E}_\pi\left[\log\left(\frac{\mathrm{b}^1(\mathrm{y_k})}{\mathrm{b}^2(\mathrm{y_k})}\right)\middle|\, \hat{X}^1_{k-1}\right]$ is bounded by the relative entropy $\mathrm{D}\left(\mathrm{b}^1(\mathrm{y_k})\middle|\mathrm{b}^2(\mathrm{y_k})\right)$.
- As here δ is monotonically bounded by $\hat{X}^1_{k-1}$, we can always find the $\mathrm{h}_\delta > 0$ s.t. the lemma result holds.

Key Result: Theorem on Insufficiently informative measurements

If $\mathrm{b}^1(\mathrm{y_k})$ and $\mathrm{b}^2(\mathrm{y_k})$ are insufficiently informative in the sense

$$\mathrm{D}(\mathrm{b}^1(\mathrm{y_k})|\mathrm{b}^2(\mathrm{y_k})) < \log\left(\frac{1}{1-\rho}\right)$$

then $\log(\hat{X}_k^1)$ is a *weak practical super-martingale* as defined earlier.

Proof sketch:

- Previous lemma and the above condition, means if $\hat{X}_{k-1}^1 < h_\delta$ then $\log(\hat{X}_k^1)$ heads down in a super-martingale looking sense (it remains to show if continues that direction at future steps).
- A tunneling argument can be used to allow application of a maximal inequality for positive super-martingales. For any h_δ select, I pick an auxiliary positive super-martingale to tunnel.
- Then follows for any $\delta_p > 0$ there exists a $h_s > 0$ such that an interval trap $\hat{X}_k^1 < h_s$ must exist and the theorem claim holds.

Outline of proof (part 1)

Noting $\log(\hat{X}_k^1) = \log M_k + \log(\hat{X}_{k-1}^1)$, then taking $E_\pi[.\,|\hat{X}_{k-1}^1]$ gives

$$E_\pi\left[\log\left(\hat{X}_k^1\right) \,\middle|\, \log\left(\hat{X}_{k-1}^1\right)\right] = E_\pi\left[\log M_k \,\middle|\, \log\left(\hat{X}_{k-1}^1\right)\right] + \log(\hat{X}_{k-1}^1)$$

Then previous lemma gives if $\mathrm{D}(\mathrm{b}^1(\mathrm{y_k})|\mathrm{b}^2(\mathrm{y_k})) < \log\left(\frac{1}{1-\rho}\right)$, then for any $\delta > 0$, then is a $\mathrm{h}_\delta > 0$ such that for any $\hat{X}_{k-1}^1 < h_\delta$ we have that $E_\pi\left[\log M_k \,\middle|\, \log\left(\hat{X}_{k-1}^1\right)\right]$<0 and hence

$$E_\pi\left[\log\left(\hat{X}_k^1\right) \,\middle|\, \log\left(\hat{X}_{k-1}^1\right)\right] < \log(\hat{X}_{k-1}^1)$$

This looks somewhat like a super-martingale, once in $\hat{X}_{k-1}^1 < h_\delta$ we expect to tread downward. But is there is chance of leaving $\hat{X}_n^1 < h_\delta$ for some $n > k$?

Outline of proof (part 2)

For any $\delta_p > 0$ you select there is a $h_\delta > 0$ such that the posterior bound holds.

Select smaller internals h_s and h_m where $h_m < h_s < h_\delta$. Then define an auxiliary positive super-martingale

$$\mathrm{Z_k} = \max\left(\log\left(\frac{\hat{X}_k^1}{h_m}\right), 0\right).$$

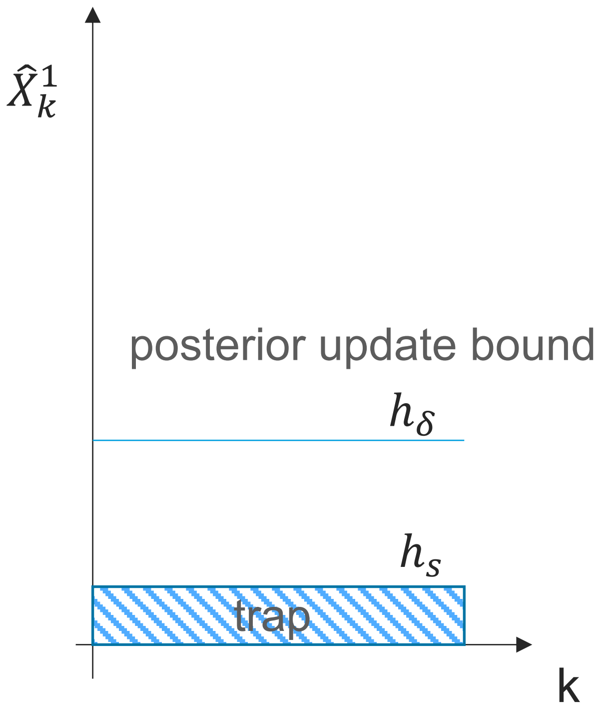
Using maximal inequality for positive super-martingales gives

$$\mathrm{P}_\pi\left(\max_{\{\mathrm{n} \geq k\}} Z_n \geq \log h_\delta / h_m\right) \leq \frac{E_\pi[Z_k]}{\log h_\delta / h_m}.$$

After a few substitutions, we can obtain for complimented event set,

$$\mathrm{P}_\pi\left(\max_{\{\mathrm{n} \geq k\}} \log(\hat{X}_k^1) < \log h_\delta\right) > 1 - \frac{h_s / h_m}{h_\delta / h_m}.$$

It then follows, for any $\delta_p > 0$ we can then select h_s and h_m so that $\frac{h_s/h_m}{h_\delta/h_m} \leq \delta_p$ and remain in $\hat{X}_k^1 < h_\delta$ and the theorem claim holds.



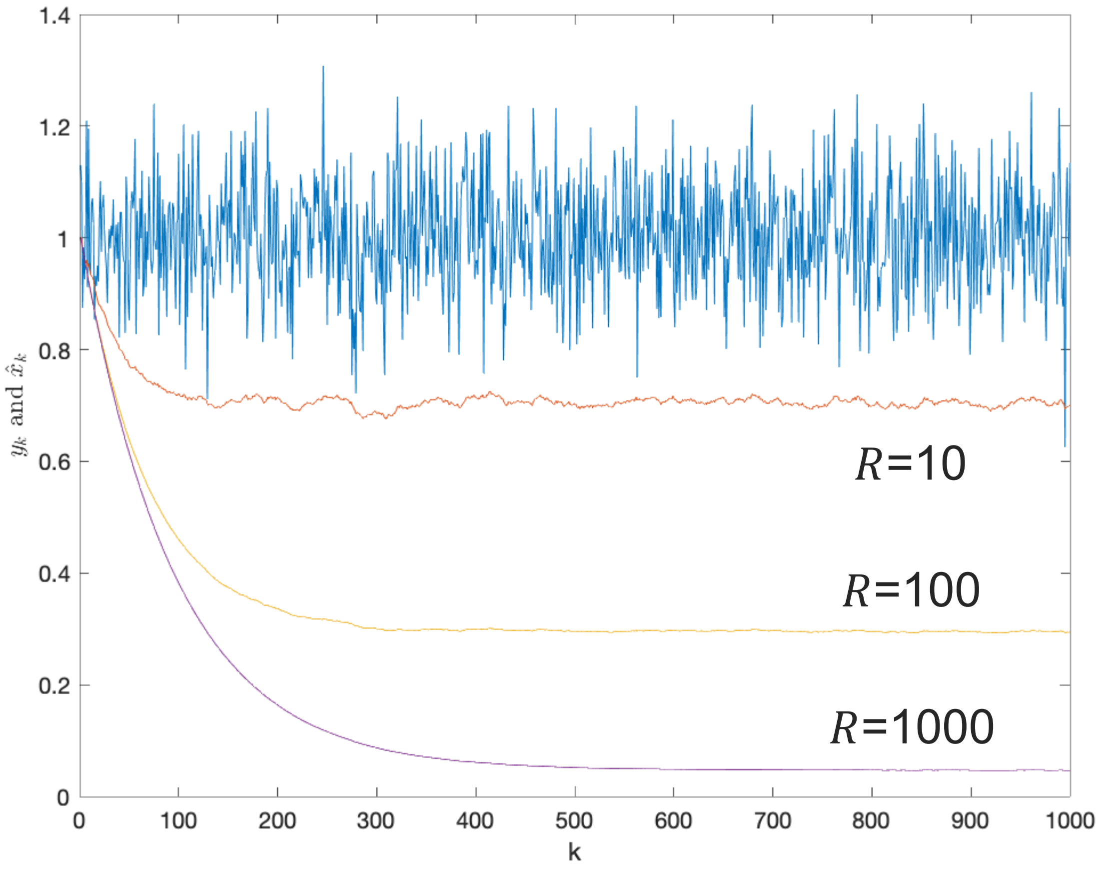
If you reach the trap, $< 1 - \delta_p$ probability of escaping $\hat{X}_k^1 < h_\delta$.

QUT Centre for Robotics

I know Kalman filters. Convince me of this sorcery!

- Imagine a simple scalar process $x_k \in R$ stable to the origin, $A < 1$, and small process variance. Consider w.l.o.g. an output matrix $\mathrm{C} = 1$.
- Imagine a (mismatched) sequence of measurements y_k having nonzero mean (w.l.o.g.) $M = 1$,
  - We will now imagine the Kalman filter estimates $\hat{x}_k$ arising under difference assumed measurement variances R. Informally, as $R \uparrow$, $\hat{x}_k \to 0$ (as the Kalman gain $\mathrm{K_k} \to 0$, measurements have less impact and the $A < 1$ drives $\hat{x}_k \to 0$).
  - Whilst the $\hat{x}_k$ will still be influenced by y_k (having mean $M > 0$), but the $\hat{x}_k$ will get closer to 0 as $R \uparrow$ (i.e. the influence of y_k reduces as $R \uparrow$).

$\mathrm{A} = 0.99, \mathrm{C} = 1, \mathrm{Q} = 0.1$ and R_{true}=0.1.



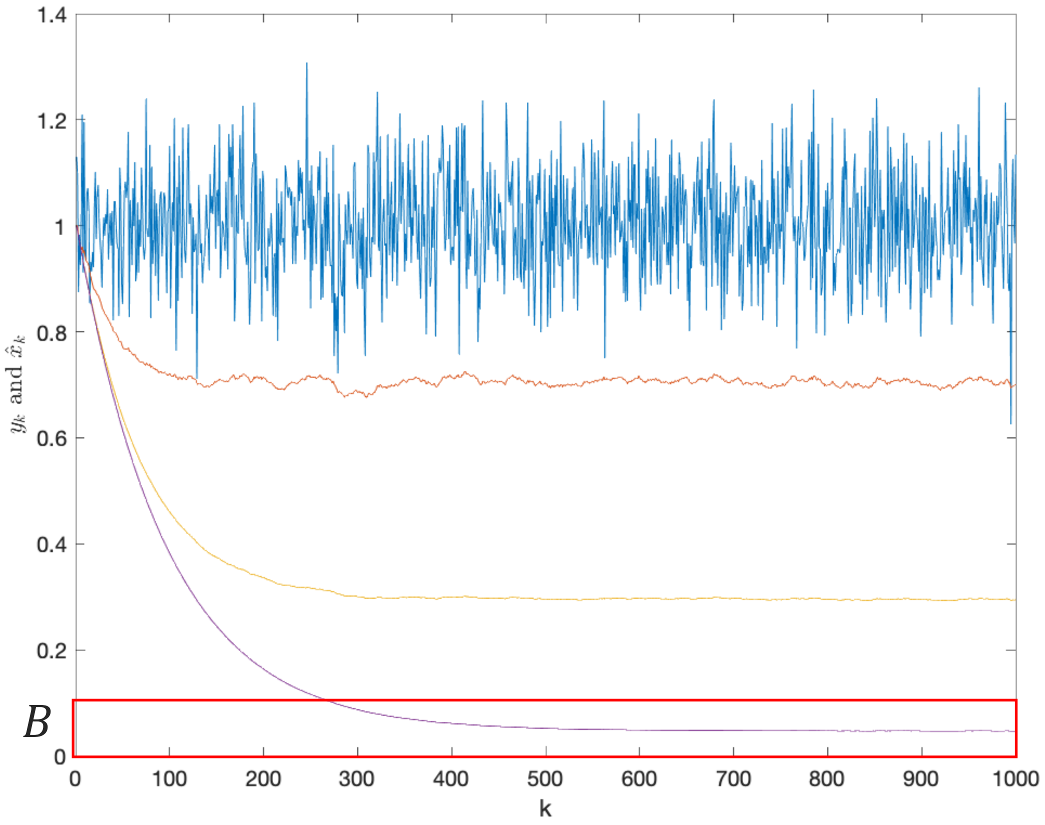
Similar phenomenon if $R = R_{true}$ fixed and Q reduced

I know Kalman filters. Convince me of this sorcery!

I let you select a small ball B (interval) at the origin. Whatever B you select, I claim there will be a critical R_c such that for any $R > R_c$ then $\hat{x}_k$ can become trapped in B (in a probabilistic sense).

The issue is not that the Kalman filter/HMM filter loses optimality, but that measurements are so weak (relative to the model) that I wonder if we **lose practical observability**.

$R_c < 1000$



QCD fixes and general thoughts

- Fixes for QCD. There are other QCD approaches that don’t suffer the same issues.
  - Behaviour arises due to Bayesian non-ergodic change model.
  - [Non-Bayesian] e.g. Pollak or Lorden’s criterion (well known CUSUM algorithm).
  - [Ergodic Bayesian change model] e.g. Our quickest intermittent signal detection problem.
- General thoughts
  - We would expect similar phenomenon to arise in more complex Bayesian QCD or filter problems involving non-ergodic models with weak observations.
  - But my Kalman filter example shows also artefacts present in ergodic models.
- Take away message: Be careful with models in low SNR situations.

Thank you.

Questions