TSFS06 Diagnos och övervakning
Föreläsning 8 - Change detection

Erik Frisk

Institutionen för systemteknik
Linköpings universitet
erik.frisk@liu.se

2020-04-28

Översikt

Change detection

Denna föreläsning baseras på

1. ett inledande kapitel i boken: "Detection of Abrupt Changes: Theory and Application"
   av
   M. Basseville, I.V. Nikiforov
   I kursen ingår **ej** hela kapitlet. Avsnitt som kan ses som överkurs är: 2.2.2, 2.2.4, 2.2.6, 2.3, 2.4.2, 2.4.3.2, 2.5, 2.6 (men läs dem gärna ändå).
   För den som är intresserad av sådan här teori finns hela boken elektroniskt tillgänglig via länk från kurshemsidan.
2. Avsnitt 4.7 i textkompendiet

Change detection

Tidigare i kursen såg hypotestesterna ut liknande

$$H^0 : \theta \in \Theta_0$$
$$H^1 : \theta \in \Theta_0^C = \Theta_1$$

dvs, antingen $\theta \in \Theta_0$ eller $\theta \in \Theta_1$ hela tiden.
I detta avsnitt antas att "hopp-tillfället" beaktas, dvs. att hypoteserna ser ut liknande:

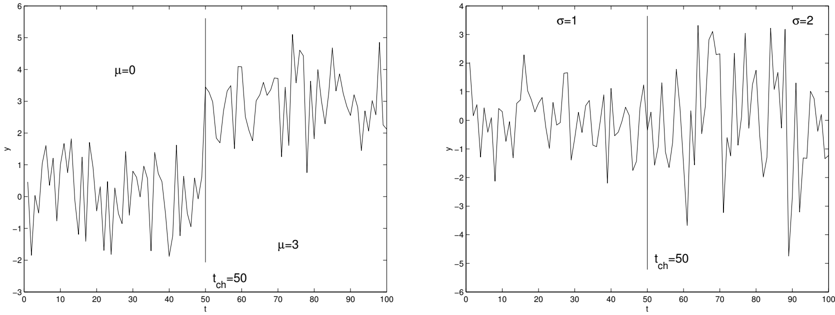
$$H^0 : \theta \in \Theta_0, \quad \forall t$$
$$H^1 : \theta \in \begin{cases} \Theta_0, & t < t_{ch} \\ \Theta_0^C = \Theta_1, & t \geq t_{ch} \end{cases}$$

Egentligen ingen skillnad mot tidigare, endast att parametrarna som testas inkluderar hopptiden

Hur passar detta in i det stora hela?

- Teststorhetskapitlet där man beaktar hopptiden
- Klassiskt statistiskproblem, väl utvecklad teori
- Allmänbildning
- Här det enklaste fallet (oberoende data)
- Förändringar i olinjära dynamiska modeller följer samma princip men kräver lite mer invecklade algoritmer

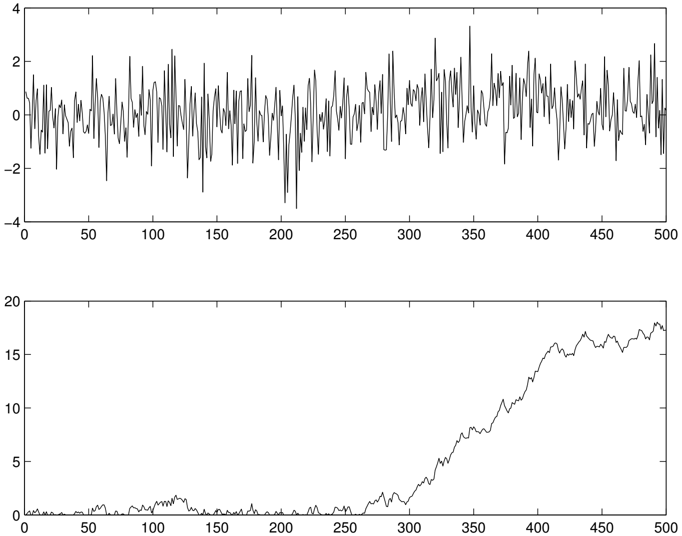
'Change detection'-problemet



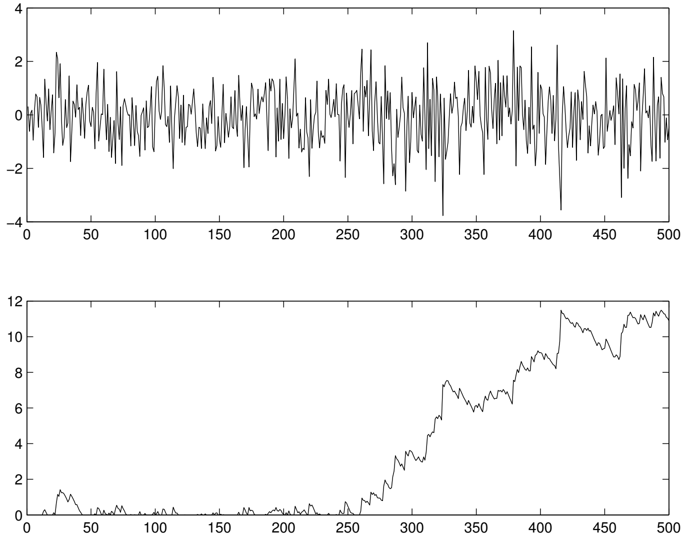
Ta hänsyn till hopp-tiden då en parameter hoppar från ett värde till ett annat dvs $\dot{\theta} \neq 0$.

$$\theta = \begin{pmatrix} \mu \\ t_{ch} \end{pmatrix}, \qquad \theta = \begin{pmatrix} \sigma \\ t_{ch} \end{pmatrix}$$

Var byter mätsignalen nivå?



Var byter mätsignalen varians?



Likelihood-funktionen

Definition (Likelihood-funktion)

Låt $f(\mathbf{z}|\theta)$ vara täthetsfunktionen för $\mathbf{Z} = [Z_1,\ Z_2,\ \dots\ Z_n]$. Då är likelihood-funktionen, givet att $\mathbf{z} = \mathbf{Z}$ observeras, funktionen som definieras av

$$L(\theta|\mathbf{z}) = f(\mathbf{z}|\theta)$$

En teststorhet kan formas som:

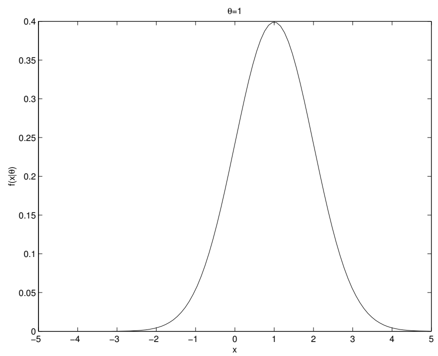
$$T(\mathbf{z}) = \max_{\theta \in \Theta^0} L(\theta|\mathbf{z})$$

Not: likelihood-funktionen är en funktion av θ medans täthetsfunktionen är en funktion av $\mathbf{z}$.

Tolkning: Hur sannolikt är det att observera det utfall vi observerat.

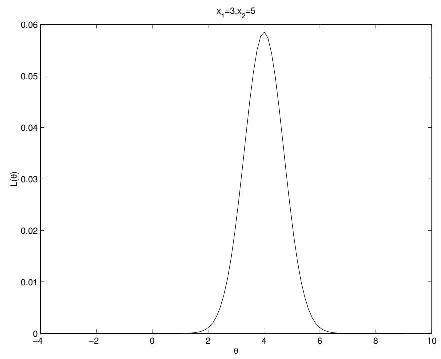
Likelihood-funktionen

Modell: $Z \sim N(\theta, 1)$
Två oberoende observationer: $z_1 = 3, z_2 = 5$



Täthetsfunktion

$$f(z_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z_i-\theta)^2}{2\sigma^2}}$$



Likelihood-funktionen

$$L(\theta) = f(3|\theta)f(5|\theta)$$

Likelihood-kvot

Antag man har hypoteserna

$$H^0 : \theta = \theta_0$$
$$H^1 : \theta = \theta_1$$

och man känner fördelningsfunktionen för data givet θ, $f(z|\theta) = L(\theta)$.

Eftersom Likelihood-funktionen säger något om hur sannolikt ett visst värde på θ är så är det rimligt att förkasta nollhypotesen om

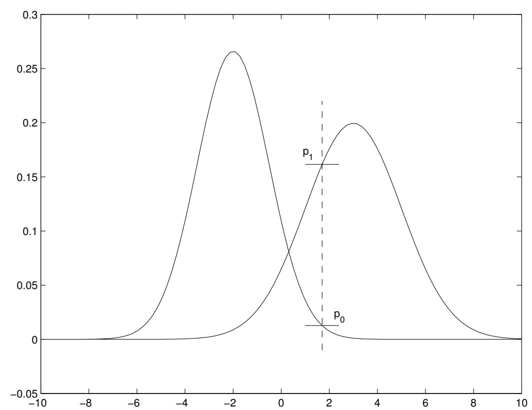
$$L(\theta_1) > L(\theta_0)$$

dvs. att kvoten

$$\frac{L(\theta_1)}{L(\theta_0)} = \frac{f(z|\theta_1)}{f(z|\theta_0)}$$

är större än en given tröskel.

Likelihood-kvot



Rimligt att välja H_1 om

$$p_1 > p_0$$

Starka kopplingar till normalisering av teststorheter baserat på likelihood-funktionen.

Likelihood-kvot/ratio

Betrakta ett hypotestest med enkla hypoteser:

$$H^0 : \theta = \theta_0$$
$$H^1 : \theta = \theta_1$$

Log-likelihood kvoten blir då (byter notation för att passa utdelad text):

$$s(y) = \log \frac{L(\theta_1|y)}{L(\theta_0|y)} = \log \frac{P_{\theta_1}(y)}{P_{\theta_0}(y)}$$

Huvudegenskapen som gör den användbar

$$E_{\theta_1}(s) > 0, \qquad E_{\theta_0}(s) < 0$$

dvs, en förändring i parametern θ visar sig som en förändring av tecken hos s.

Neyman-Pearson lemma

Antag hypoteserna

$$H^0 : \theta = \theta_0$$
$$H^1 : \theta = \theta_1$$

där pdf för observationerna är den kända fördelningsfunktionen $f(z|\theta_i)$ i de två fallen.

En lite "slarvig" formulering av Neyman-Pearson lemma är då:
Den bästa tänkbara teststorheten (tänk styrkefunktion) för dessa hypoteser är

$$T(z) = \frac{f(z|\theta_1)}{f(z|\theta_0)}$$

Finns generaliserade resultat för nollhypoteser som inte är singeltons.

Slutsats: LR är bra att sikta på om man har en god statistisk modell över sina observationer.

Översikt

- *Change detection*
  - *Likelihood-funktionen*
  - *Likelihood-kvot och hypotestester*
- *Algoritmer för change detection*
  - *θ_1 känd - CUSUM*
  - *θ_1 okänd - GLR*
- *CUSUM och residualer*
- *Sammanfattning*

Algoritmer för Change Detection

Delar in algoritmerna i två kategorier:

$$H^0 : \theta = \theta_0$$
$$H^1 : \theta = \theta_1$$

1. θ_1 känd
   - Shewhart control chart
   - Geometrical Moving Average
   - Finite Moving Average
   - Filtered Derivative Algorithm
   - ⇒ CUSUM (CUmulative SUM)
   - ...
2. θ_1 ej känd
   - ⇒ GLR (Generalized Likelihood Ratio)
   - ...

Alla dessa detektionsalgoritmer har gemensamt att likelihood-kvoten är en grundsten.

Shewhart control chart

Samla in "batchar" med N data och beräkna teststorheten enligt:

$$T = \sum_{t=1}^{N} s_t, \quad s_t = \log \frac{P_{\theta_1}(y_t)}{P_{\theta_0}(y_t)}$$

Varför summera på det där viset?
Ett grundantagande som ofta görs är att y_t:na är oberoende, då gäller att

$$\log \frac{L(\theta_1|y_1,\dots,y_N)}{L(\theta_0|y_1,\dots,y_N)} = \log \frac{P_{\theta_1}(y_1,\dots,y_N)}{P_{\theta_0}(y_1,\dots,y_N)} =$$
$$= \log \prod_{t=1}^{N} \frac{P_{\theta_1}(y_t)}{P_{\theta_0}(y_t)} = \sum_{t=1}^{N} \log \frac{P_{\theta_1}(y_t)}{P_{\theta_0}(y_t)} = \sum_{t=1}^{N} s_t = T$$

Vilket också förklarar varför log-likelihood ratio ofta används istället för likelihood-ratio.

Ex.: Hopp hos medelvärde för normalfördelad variabel

Mätsignal y_t som är en oberoende sekvens av variabler med fördelning $N(\mu, \sigma)$, σ känd.

$$H^0 : \mu = \mu_0$$
$$H^1 : \mu = \mu_1$$

$$s_t = \log \frac{P_{\theta_1}(y_t)}{P_{\theta_0}(y_t)} = \log \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t-\mu_1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t-\mu_0)^2}{2\sigma^2}}} = \dots = \frac{\mu_1 - \mu_0}{\sigma^2} \left(y_t - \frac{\mu_0 + \mu_1}{2} \right)$$

Teststorheten blir då ett ganska naturligt och intuitivt mått (tolka):

$$T = \sum_{t=1}^{N} s_t = N \frac{\mu_1 - \mu_0}{\sigma^2} \left(\bar{y} - \frac{\mu_0 + \mu_1}{2} \right)$$

Ex.: Hopp hos varians för normalfördelad variabel

Mätsignal y_t som är en oberoende sekvens av variabler med fördelning $N(\mu, \sigma)$.

$$H^0 : \sigma = \sigma_0$$
$$H^1 : \sigma = \sigma_1$$

$$s_t = \log \frac{P_{\theta_1}(y_i)}{P_{\theta_0}(y_t)} = \log \frac{\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t-\mu)^2}{2\sigma_1^2}}}{\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(y_t-\mu)^2}{2\sigma_0^2}}} = \dots =$$
$$= \log \frac{\sigma_0}{\sigma_1} - (y_t - \mu)^2 \frac{1}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right)$$

Teststorheten blir då ett kopplad till en variansskattning $\hat{\sigma}^2$ av y_t.

$$T = \sum_{t=1}^{N} s_t = \frac{N}{2} \log \frac{\sigma_0^2}{\sigma_1^2} - \frac{N}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right) \hat{\sigma}^2$$

Exempel forts.

"Tanken" bakom teststorheten är inte helt enkel att tolka. Vi kan verifiera att den uppfyller de grundläggande villkoren på teststorheten.

Väntevärdet av teststorheten under nollhypotesen är

$$E_{\Theta_0}\{T\} = \frac{N}{2} \log \frac{\sigma_0^2}{\sigma_1^2} - \frac{N}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right) \sigma_0^2 = -\frac{N}{2} \left(\frac{\sigma_0^2}{\sigma_1^2} - \log \frac{\sigma_0^2}{\sigma_1^2} - 1 \right) < 0$$

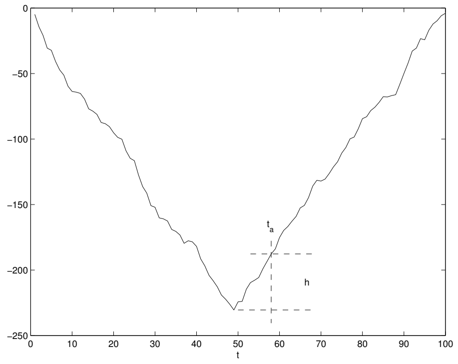
Sista olikheten eftersom $x - \log x \geq 1$ och likhet endast då $x = 1$, dvs då $\sigma_1 = \sigma_0$.

Under mothypotesen blir motsvarande räkningar enligt samma olikhet

$$E_{\Theta_1}\{T\} = \frac{N}{2} \log \frac{\sigma_0^2}{\sigma_1^2} - \frac{N}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right) \sigma_1^2 = \frac{N}{2} \left(\frac{\sigma_1^2}{\sigma_0^2} - \log \frac{\sigma_1^2}{\sigma_0^2} - 1 \right) > 0$$

CUSUM - intuitiv härledning

Hittills har ej hopp-tiden beaktats, det är detta som görs i CUSUM algoritmen.
Hur uppför sig den kumulativa summan $\sum s_i$ typiskt runt tiden för förändringen?



Intuitivt borde man därför bilda test-storheten enligt:

$$T_t = S_1^t - m_t$$

där

$$S_1^t = \sum_{i=1}^{t} s_i$$

$$m_t = \min_{1 \le j < t} S_1^j$$

CUSUM - härledning

$$H^0 : \theta = \theta_0, \quad \forall t$$

$$H^1 : \theta = \begin{cases} \theta_0 & t < t_{ch} \\ \theta_1 & t \ge t_{ch} \end{cases}$$

Likelihood ratio och oberoende ger:

$$L_1^t(t_{ch}) = \frac{P_{H^1}(y)}{P_{H^0}(y)} = \frac{\prod_{i=1}^{t_{ch}-1} P_{\theta_0}(y_i) \prod_{i=t_{ch}}^{t} P_{\theta_1}(y_i)}{\prod_{i=1}^{t} P_{\theta_0}(y_i)} = \prod_{i=t_{ch}}^{t} \frac{P_{\theta_1}(y_i)}{P_{\theta_0}(y_i)}$$

Log-likelihood ratio:

$$\log L_1^t(t_{ch}) = \sum_{i=t_{ch}}^{t} \log \frac{P_{\theta_1}(y_i)}{P_{\theta_0}(y_i)} = \sum_{i=t_{ch}}^{t} s_i = S_{t_{ch}}^t$$

Problem: Hopp-tiden t_{ch} är okänd.
Standardlösning inom statistikfältet: gör en ML-skattning av t_{ch} och sätt in i formeln.

CUSUM, forts.

ML-skattningen av t_{ch} blir då:

$$\hat{t}_{ch} = \arg \max_{1 \le t_{ch} \le t} \log L_1^t(t_{ch}) = \arg \max_{1 \le t_{ch} \le t} S_{t_{ch}}^t$$

och uttrycket för teststorheten

$$T_t = S_{\hat{t}_{ch}}^t = \max_{1 \le t_{ch} \le t} S_{t_{ch}}^t = \max_{t_{ch}} S_1^t - S_1^{t_{ch}-1} = S_1^t - \min_{t_{ch}} S_1^{t_{ch}-1}$$

vilket är samma uttryck som resonerades fram tidigare:

$$T_t = S_1^t - m_t$$

där

$$S_1^t = \sum_{i=1}^{t} s_i, \quad m_t = \min_{1 \le j < t} S_1^j$$

CUSUM, forts.

Teststorheten beräknas som

$$T_t = S_1^t - m_t$$

där

$$S_1^t = \sum_{i=1}^{t} s_i, \quad m_t = \min_{1 \le j < t} S_1^j$$

Ett lite enklare, och vanligare, sätt är istället beräkna

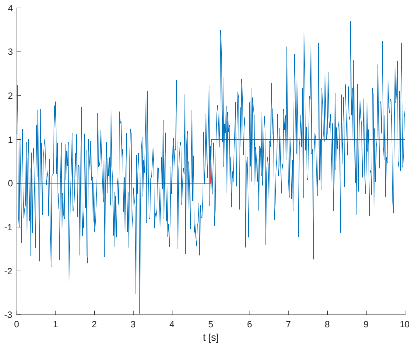
$$T_t' = \max(0, T_{t-1}' + s_t), \quad T_0' = 0$$

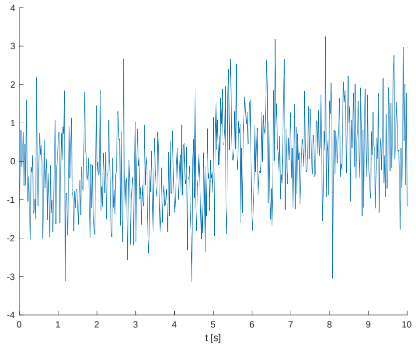
Att besluten från de två olika algoritmerna är samma ges av

Proposition

$$T_t' = \max(0, T_t)$$

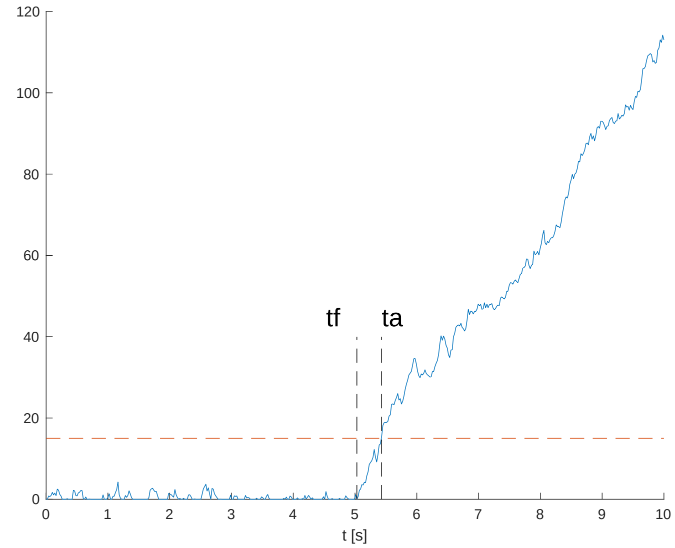
CUSUM exempel: förändring i väntevärde



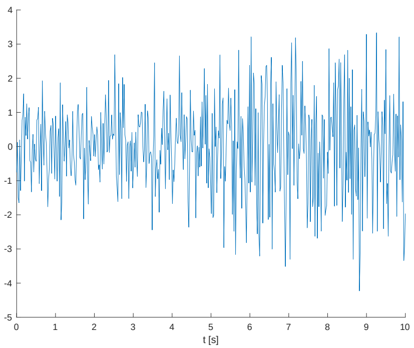


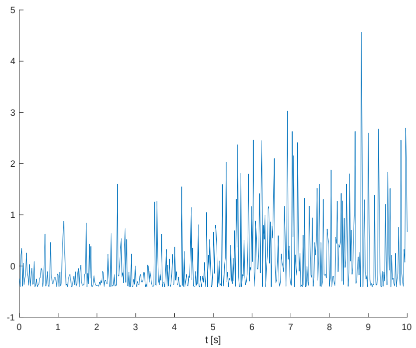
$$s_t = \log \frac{p_1(y_t)}{p_0(y_t)}$$

CUSUM exempel: förändring i väntevärde



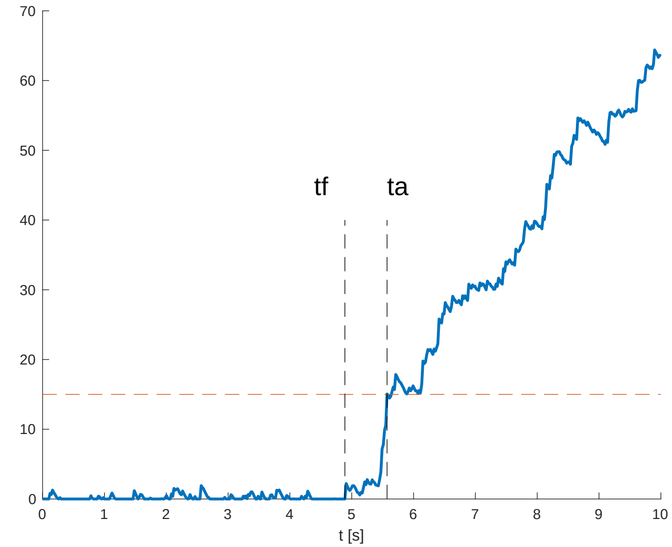
CUSUM exempel: förändring i varians





$$s_t = \log \frac{p_1(y_t)}{p_0(y_t)}$$

CUSUM exempel: förändring i väntevärde



GLR (Generalized Likelihood Ratio)

CUSUM krävde kunskap om det nya värdet på variabeln θ. Vad gör vi i det (normala) fallet då θ_1 är okänt?

Svar: Hanterar det på exakt samma sätt som den okända hopptiden hanterades i CUSUM, dvs. gör en ML-skattning av θ_1.

$$T_t = \max_{\theta_1, t_{ch}} \log \frac{P_{H_1}(y)}{P_{H_0}(y)}$$

Log-likelihood ratio blir alltså en funktion av både t_{ch} och θ_1:

$$S_{t_{ch}}^t(\theta_1) = \sum_{i=t_{ch}}^{t} s_i = \sum_{i=t_{ch}}^{t} \log \frac{P_{\theta_1}(y_i)}{P_{\theta_0}(y_i)}$$

ML-skatta genom att maximera över båda variablerna

$$T_t = \max_{1 \le t_{ch} \le t} \max_{\theta_1} S_{t_{ch}}^t(\theta_1)$$

Implementation och ytterligare detaljer

- Principen enligt förra bilden men den dubbla maximeringen kan vara svår att göra effektiv.
- Mer information om GLR (bland annat implementationsdetaljer) och liknande algoritmer hittas i
  - "Adaptive Filtering and Change Detection", Fredrik Gustafsson, Wiley, 2000.
  - "Detection of Abrupt Changes - Theory and Application", Michèle Basseville and Igor V. Nikiforov, previously published by Prentice-Hall, 1993. (Nu tillgänglig på nätet, se kurshemsidan)
  - Kay, Steven M. Fundamentals of statistical signal processing, volume II: detection theory."(1998).

Översikt

- Change detection
  - Likelihood-funktionen
  - Likelihood-kvot och hypotestester
- Algoritmer för change detection
  - θ_1 känd - CUSUM
  - θ_1 okänd - GLR
- CUSUM och residualer
- Sammanfattning

CUSUM och residualer

Grunden för CUSUM var att man hade en s.k. score-function s_t med egenskapen

$$E_{H_0}\{s_t\} < 0, \quad E_{H_1}\{s_t\} > 0$$

då CUSUM-algoritmen blev

$$T_t = \max(0, T_{t-1} + s_t), \quad T_0 = 0$$

Tidigare har s_t varit log-likelihood kvoten.

Antag vi inte har den statistiska kunskapen utan har genererat residualer. Kan CUSUM ändå vara intressant?

Residualer har ju typiskt egenskapen

$$r_t = \begin{cases} \text{brusig runt 0} & \text{under } H_0 \\ \text{signifikant skild från 0} & \text{under } H_1 \end{cases}$$

CUSUM och residualer, forts.

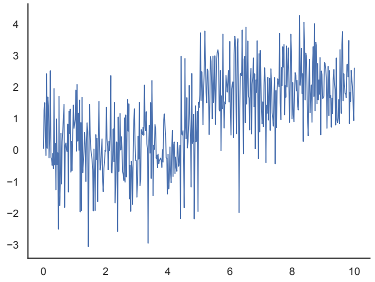
Det är enkelt åtgärdat, definiera till exempel

$$s_t = |r_t| - \nu$$

där ν är, i runda slängar, i storleksordning lika som $r(t)$:s största värde under H_0.

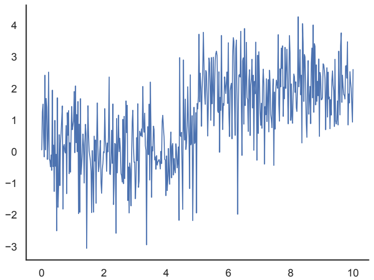
- ν kallas en driftsfaktor
- ν är ett mått på modellosäkerheten
- adaptivitet kan med fördel infogas i ν, stort ν då systemet opererar i arbetsområden med stor modellosäkerhet etc.
- olinjärt filter som visar sig vara ett användbart alternativ till linjära lågpassfilter för efterbehandling av residualer.

CUSUM och residualer, exempel

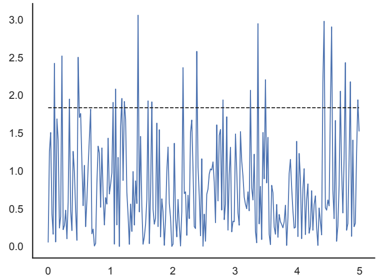


Residual $r(t)$

CUSUM och residualer, exempel

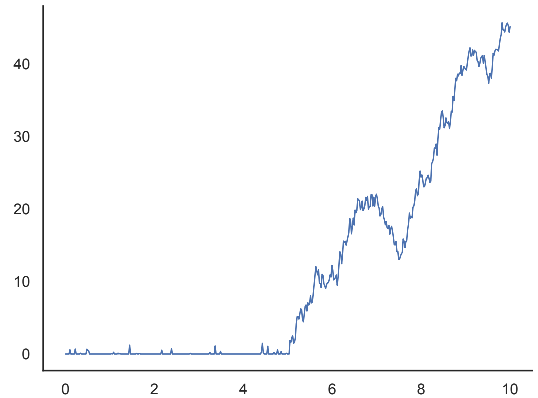


Residual $r(t)$



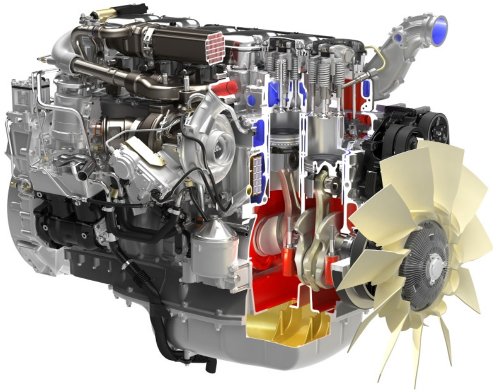
$|r(t)|$ i felfritt fall

CUSUM och residualer, exempel



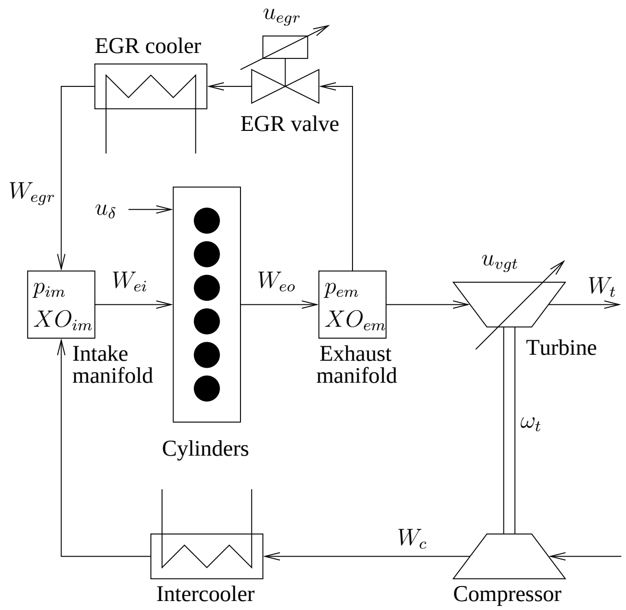
$$T_t = \max(0, T_{t-1} + |r_t| - \nu), \quad T_0 = 0$$

Dieselmotor i SCANIA-lastbil med EGR/VGT



- Alla modellvariabler är kopplade via turbo och EGR
- Ett fel påverkar därför typiskt alla mätta variabler
- Viktigt att systematiskt kunna utnyttja ovanstående
- Svårmodellerat system

Dieselmotor i SCANIA-lastbil med EGR/VGT



Dynamisk modell av motorn

Motorn har tydligt dynamiskt beteende och kan beskrivas av ett system av ordinära differentialekvationer

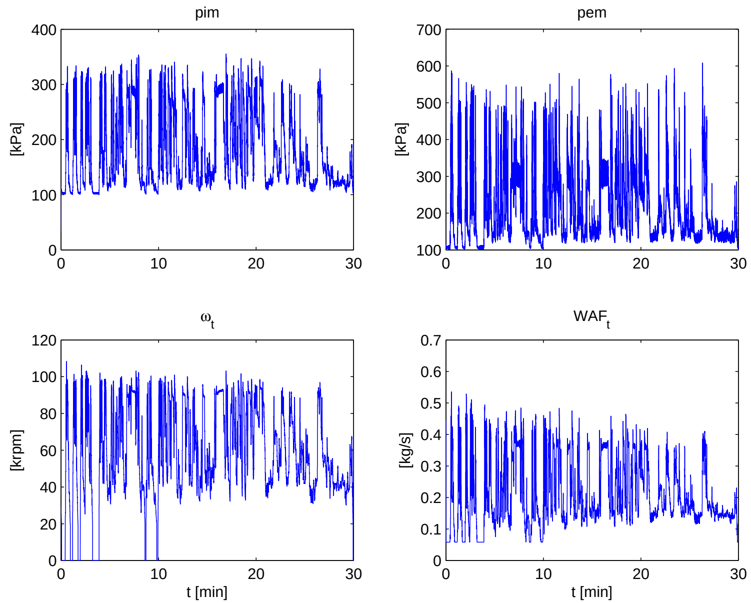
$$\frac{d}{dt}x = g(x, u)$$
$$y = h(x)$$

där tillstånden, mätsignalerna, och styrsignalerna är

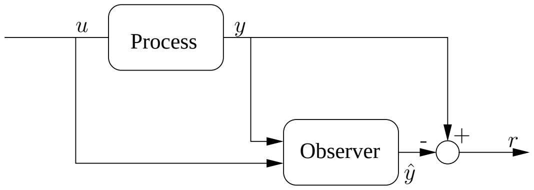
$$x = (p_{im}, p_{em}, \omega_t), \quad y = (p_{im}, p_{em}, \omega_t, W_{af}), \quad u = (\delta, EGR, VGT)$$

- En modell som bygger på grundläggande fysikaliska principer, innehåller mappar som kalibrerats via experiment i testcell.
- Tydligt olinjär
- Används för reglerdesign
- Hyfsat korrekt över motorns hela arbetsområde
- Tydliga bias och osäkerheter på sina ställen

European Transient Cycle - experimentdata

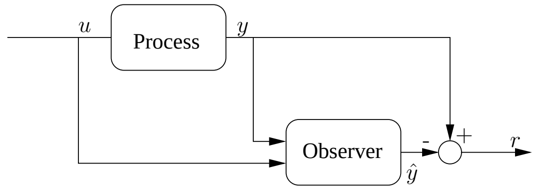


Residualgenerator via estimering



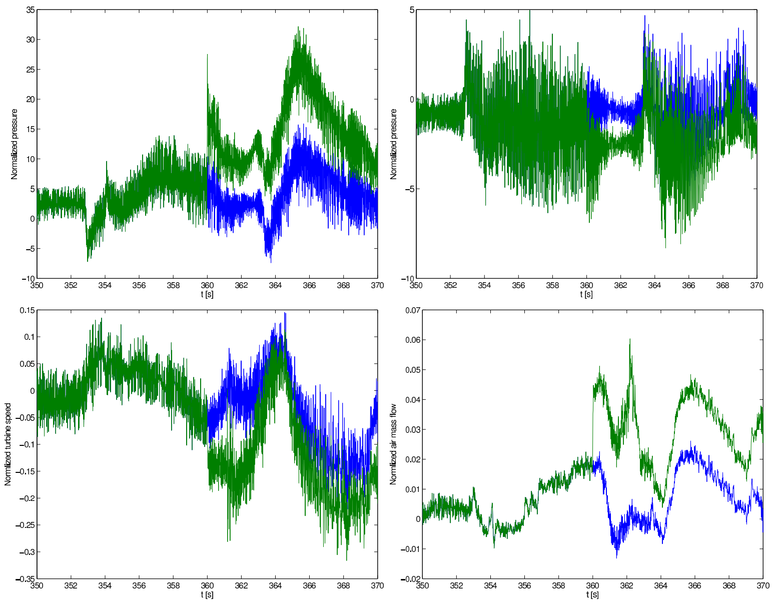
- Kandidatmetod att användas i en toolbox är stokastiska filter som till exempel: Extended Kalman Filter (EKF)/Unscented Kalman Filter (UKF)/Particle Filter (PF)
- "plug and play"
- Inga garantier

Residualgenerator via estimering

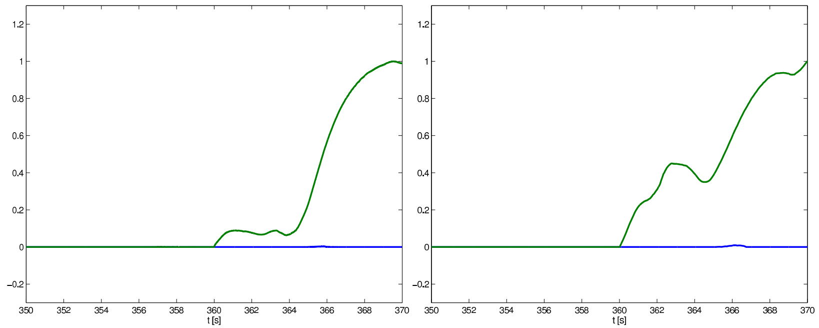


$$\hat{x} = \text{EKF(observations)}$$
$$r_1 = p_{im} - \hat{p}_{im} \quad \text{(Intake manifold pressure)}$$
$$r_2 = p_{em} - \hat{p}_{em} \quad \text{(Exhaust manifold pressure)}$$
$$r_3 = \omega_t - \hat{\omega}_t \quad \text{(Turbine speed)}$$
$$r_4 = W_{af} - \hat{W}_{af} \quad \text{(Air mass flow past air-filter)}$$

Residualer - 10% fel i trycksensor på lastbilsmotor



CUSUM och residualer



- Applicera CUSUM på två av residualerna från förra bilden
- Enkelt att göra ett pålitligt beslut

- tiden för förändring tas med i beaktande
- här enkla fall och grundalgoritmer
- teorin bygger på likelihood kvoter - optimal
- CUSUM - om θ_1 är känd
  GLR - om θ_1 ej är känd
- alternativ till efterbehandling av residualer, oavsett hur de är genererade. Enkelt att införa och tolka adaptiv tröskel.

TSFS06 Diagnos och övervakning
Föreläsning 8 - Change detection

Erik Frisk

Institutionen för systemteknik
Linköpings universitet
erik.frisk@liu.se

2020-04-28