The Window of Change

Nils Lid Hjort / Stability and Change 2022-2023, CAS

ChangeTrend Workshop, PRIO, 28-29/iii/2023

Changepoints and regime shifts ...

Standard framework for changepoints: observations $y_1, \ldots, y_n$ follow the model $f(y, \theta)$, with parameters

$$\theta_i = \theta_L \text{ if } i \le a,$$
$$\theta_i = \theta_R \text{ if } i \ge a + 1.$$

There's a large literature on

(i) testing whether the world has been constant (no changepoint);

(ii) spotting the changepoint a (if it's there);

(iii) constructing confidence statements;

(iv) assessing how different θ_R is from θ_L.

Applications abound & multiply – and find uses not merely for 'change of visible level' but for inner-working parameters (has a regression coefficient β_4 for education level changed over time, in relation to democracy?).

... but changes often take time

This talk: Window of Change (a, b):

$$\theta_i = \theta_L \text{ if } \; i \le a,$$
$$\theta_i = \theta_R \text{ if } \; i \ge b,$$
$$\theta_i = \text{in between if } a < i < b.$$

Need to assume something for the transition window from Equilibrium A to Equilibrium B. A natural start is

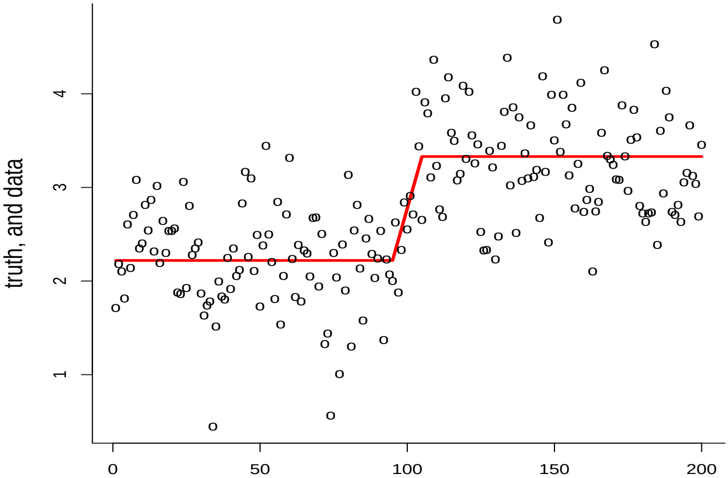
$$\theta_i = \theta_L + \frac{i-a}{b-a}(\theta_R - \theta_L) \quad \text{for } i = a, \ldots, b.$$

How to estimate and reach inference for the Window of Change (a, b), along with θ_L, θ_R?

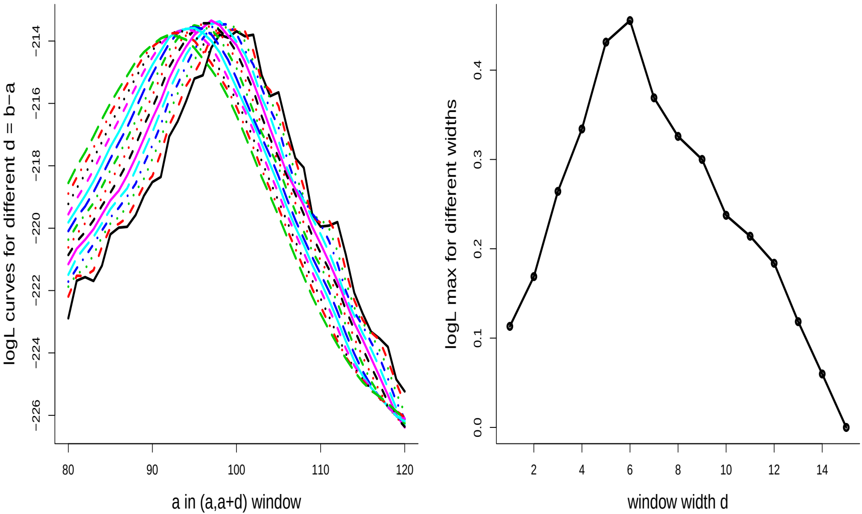
Answers: (i) log-likelihood analyses; (ii) Bayes with MCMC.

Story I: A stylised illustration: transition from A to B

200 normal observations; going from $\theta_L = 2.22$ to $\theta_R = 3.33$ over time window $[a, b] = [95, 105]$. Can we estimate this from data?



Left: log-likelihoods $\ell(a, a+d)$ in a, for fixed widths $d = 1, \ldots, 15$. Right: log-lik maxima $\ell_{\max}(d)$ over d.



Then Bayes MCMC. With a well-chosen prior $\pi(a)\pi(d)\pi(\theta_L,\theta_R)$, in terms of $d = b - a$, the posterior is

$$\pi(a, b, \theta_L, \theta_R \,|\, \mathrm{data}) \propto \pi(a)\pi(d)\pi(\theta_L, \theta_R) \exp\{\ell(a, b, \theta_L, \theta_R)\}$$

in terms of log-likelihood $\ell(\mathrm{parameter})$.

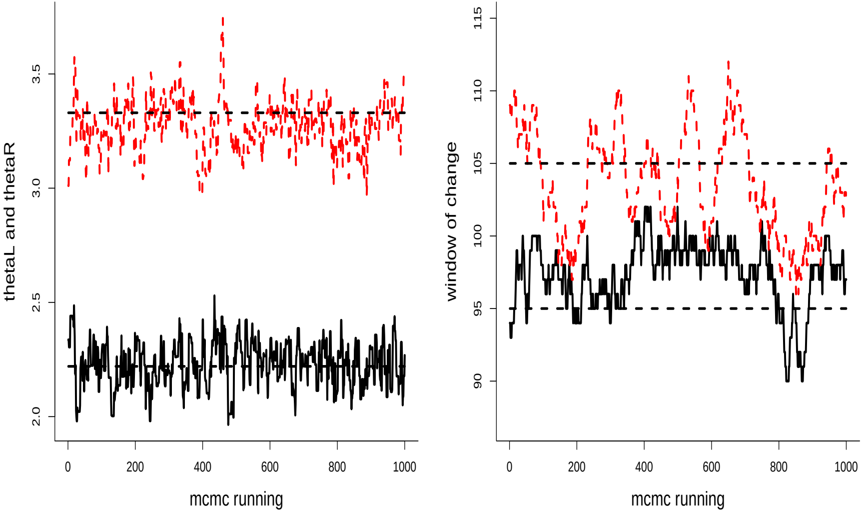
I construct a Markov chain of outcomes $(\theta_L, \theta_R, a, b)$ in my computer. From $\mathrm{old} = (\theta_L, \theta_R, a, b)$ I propose $\mathrm{next} = (\theta'_L, \theta'_R, a', b')$, with a gentle symmetric push for θ_L, θ_R, whereas $a' - a = -1, 0, 1$ and $b' - b = -1, 0, 1$ with equal probabilities $(1/3)^2 = 1/9$. I accept with probability

$$\begin{aligned} \mathrm{pr} &= \min(1, \exp(\delta)), \\ \delta &= \ell(\mathrm{next}) - \ell(\mathrm{old}) + \log \pi(\mathrm{next}) - \log \pi(\mathrm{old}). \end{aligned}$$

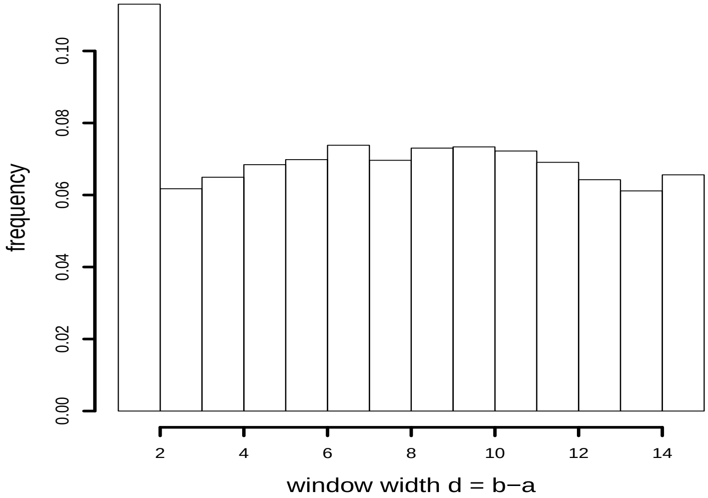
Book-keeping care to ensure $\mathrm{pr} = 0$ for non-windows, etc. Then by MCMC theory this produces simulations from the genuine posterior distribution.

It's good clean fun to see the (a, b) chain on your computer screen.

Bayes MCMC, with flat prior for $d = b - a$ on $1, \ldots 15$.
It works – but easier for θ_L, θ_R than for $[a, b]$. Difficult to get window right, even with good data. Prior for d matters. Here $d_{\mathrm{true}} = 10$.

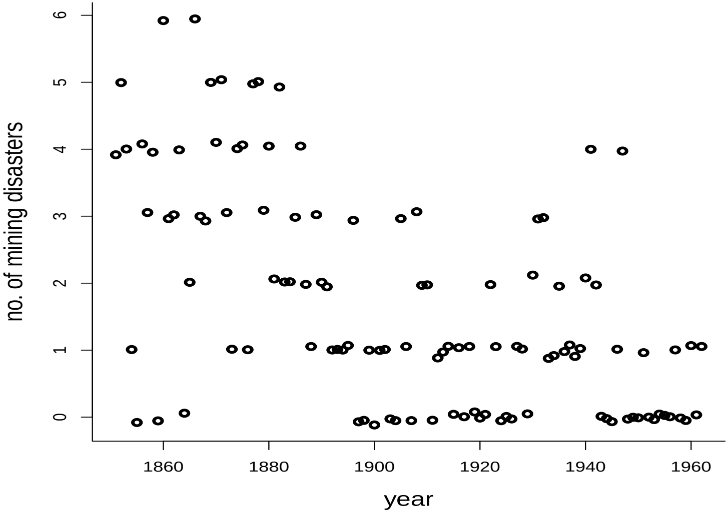


Flat prior for window width d on $1, \ldots, 15$; long MCMC to read off $\pi(d \mid \mathrm{data})$. Here $d_{\mathrm{true}} = 10$. Prior not easily 'washed out by data'.



Story II: British mining disasters

No. of disasters, from 1851 to 1962 (from Jarrett, 1979). The Poisson level has diminished from $\theta_L \approx 3.0$ to $\theta_R \approx 1.0$ ... but about when, and how quickly?



Poisson model with window of change

I take $y_i \sim \mathrm{Pois}(\theta_i)$ with

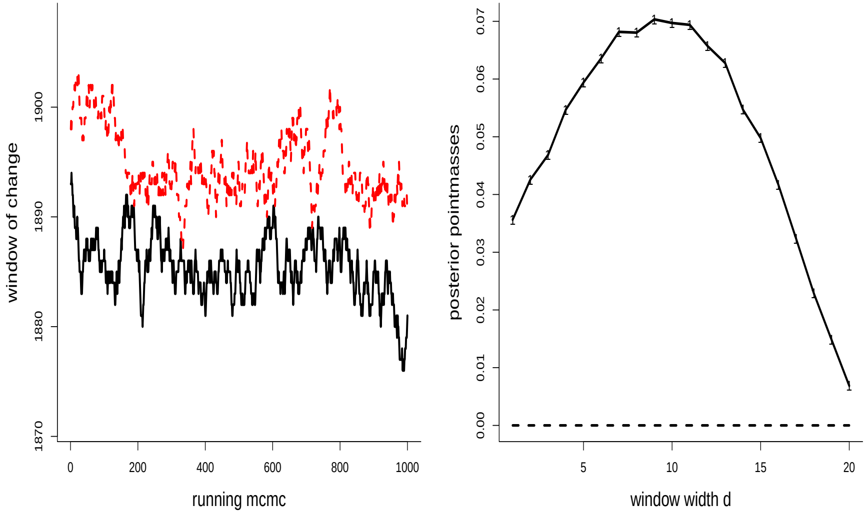
$$\begin{aligned} \theta_i &= \theta_L \text{ if } \ i \leq a, \\ \theta_i &= \theta_R \text{ if } \ i \geq b, \\ \theta_i &= \text{linear in between} \text{ if } a < i < b. \end{aligned}$$

Several papers in the literature have aimed for simple changepoint, i.e. $b = a + 1$.

Their story: change from 1891 to 1892 !, from $\theta_L \doteq 3.25$ to $\theta_R \doteq 0.88$.

My story: gradual change from 1889 to 1898 !

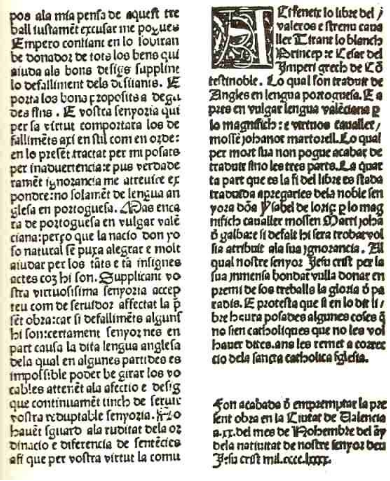
I'm using Bayes with prior $1/d$ on $1, \ldots, 20$, and build my MCMC running in $(\theta_L, \theta_R, a, b)$, reading off $\pi(d \,|\, \mathrm{data})$, position of window $[a, a + d]$, etc. – Reasonably similar frequentist results, with log-likelihoods, but prior on d matters.


window of change
1870
1880
1890
1900
0
200
400
600
800
1000
running mcmc
posterior pointmasses
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
5
10
15
20
window width d

Story III: When (and how quickly) did Author B take over for Author A?

Tirant lo Blanch is the world's first novel, written in Catalan, c. 1460–1464, and published in València in 1490. Somewhere in the sequence of 487 chapters, Martí Joan de Galba took over for Joanot Martorell.

But where, precisely? And did the change take place instantly (from chapter 371 to chapter 372, claim Cunen, Hermansen, Hjort, JSPI 2018), or did it take a few chapters?


pos ala mia pensa de aquest treball iustamēt excusar me porues
Empero confiant en lo souiran be donador de tots los bens qui aiuda als bons desigs suppline lo defalliment dels desitiants. E porta los bons propòsits a degudes fins. E vostra senyoria qui per sa virtut comportara los defallimēts axí en stil com en orde: en lo presēt tractat per mi posats per inaduertencia: e pus verdaderamēt ignorancia me atreuire expondre: no solamēt de lengua anglesa en portoguesa. Mas encara de portoguesa en vulgar valēciana: perço que la nacio don yo so natural se puxa alegrar e molt aiudar per los tāts e tā insignes actes coz hi son. Supplicant vostra virtuosissima senyoria accepteu com de seruidor affectat la presēt obra: car si defallimēts alguns hi son: certament senyor nes en part causa la dita lengua anglesa dela qual en algunes paraules es impossible poder be girar los vocables attenēt ala afectio e desig que continuamēt tinch de seruir vostra redubtable senyoria. E no hauēt sguard ala ruditat dela ordinacio e diferencia de sentēcies així que per vostra virtut la comu

Acabasse lo libre del valeros e strenu caualler Tirant lo blanch Princep e Cesar del Imperi grech de Cōstantinoble. Lo qual fon tradiut de Angles en lengua portoguesa. E apres en vulgar lengua valēciana per lo magnifich e virtuos caualler mossē johanot martorell. Lo qual per mort sua non pogue acabar de traduir sino les tres parts. La quarta part que es la fi del libre es stada traduida apregaries dela noble senyora dōa Ysabel de loriç per lo magnifich caualler mossen Marti johā dē galba: e si defalt hi sera trobat vol sia atribuit ala sua ignorancia. Al qual nostre senyor Jesu crist per la sua immensa bondat vulla donar en premi de sos treballs la gloria dē paradis. E protesta que si en lo dit libre haura posades algunes coses que no sien catholiques que no les vol hauer dites. ans les remet a correctio dela sancta catholica sglesia.

Fon acabada dē emprēptar la present obra en la Ciutat de Valencia a .xx. del mes de Nohembre del any dela natiuitat de nostre senyor deu Jesu crist mil.cccc.lxxxx.

What can we look for?

What can we measure and monitor, chapter for chapter?

From Chapter XII:

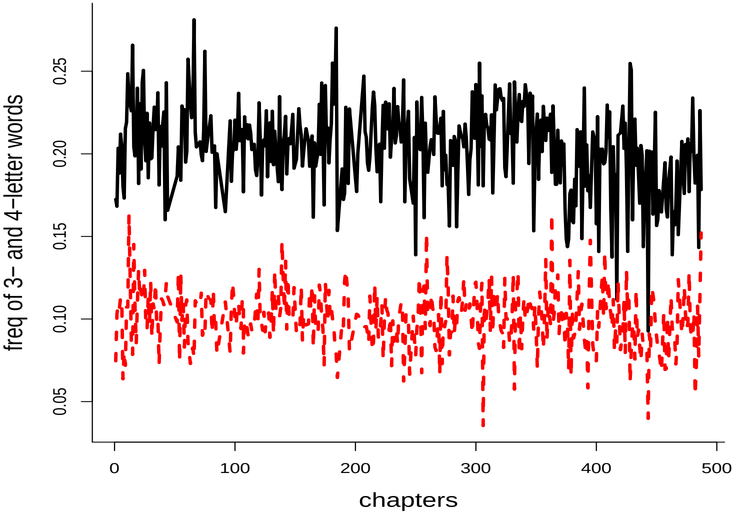
E lo rey dix que era molt content. E en la scura nit lo virtuós hermità mudà's les vestidures que tenia apparellades de moro, e per la porta falsa del castell isqué molt secretament, que per negú no fonch vist ne conegut, e posà's dins lo camp dels moros.

From Chapter CCCCLXXII:

¡Despullau a mi daurades robes y dels palaus leven les riques porpres! ¡Cobriu-me prest de hun aspre scilici, visten-se tots de fort y negra màrrega, sonen ensemps les campanes sens orde, dolga's tothom de tanta pèrdua, per a rahonar la qual ma lengua és feta scaça!

See Céline Cunen's talk: We go for word lengths and their proportions, chapter by chapter.

We compute and examine frequencies $\widehat{p}_1, \ldots, \widehat{p}_{10}$ through chapters $1, \ldots, 487$. Here 3-letter and 4-letter words. Where is the change?



The multinomial-Dirichlet window of change model for word lenghts

In chapter i there are m_i words, sorted into $y_{i,1}, \ldots, y_{1,10}$ of lengths 1, ..., 10. My model takes

$$f_i = \int \frac{m_i!}{y_{i,1}! \cdots y_{i,10}!} p_1^{y_{i,1}} \cdots p_{10}^{y_{i,10}} \, \mathrm{Dir}(\mathrm{d}p)$$
$$= \frac{m_i!}{y_{i,1}! \cdots y_{i,10}!} \frac{\Gamma(k_i p_{i,0,1} + y_{i,1}) \cdots \Gamma(k_i p_{i,0,10} + y_{i,10})}{\Gamma(k_i p_{i,0,1}) \cdots \Gamma(k_i p_{i,0,10})} \frac{\Gamma(k_i)}{\Gamma(k_i + m_i)}$$

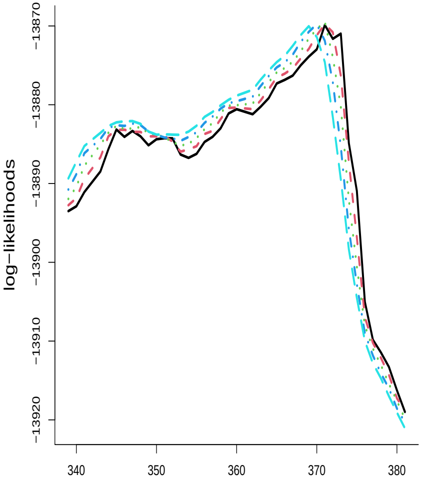
for chapter i, with the Window of Change setup:

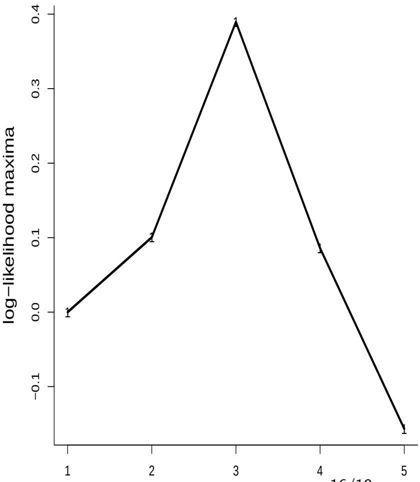
(k_L, p_L) up to a; (k_R, p_R) after $a + d$; linear interpolation inside $(a, a + d)$.

For given $(a, a + d)$, need to optimise over $1 + 9 + 1 + 9 = 20$ parameters, and over the full dataset.

Did it take ... 3 chapters to settle in?

So change sets in at c. Chapter 371 of the 487. I'm checking chapter windows $[a, a+1], \ldots, [a, a+5]$: but no, $\widehat{d} = 3$ is not significant.





Remarks

- Estimation and inference for (a, b) window inherently more difficult than for a single changepoint. Methods are harder to construct; harder to analyse well; and precision is lower.

- There's room for good Bayesian methods (with MCMC), but prior for $d = b - a$ is crucial, and matters more than for other components. Use context & knowledge. Also easy to read off how much θ has changed.

- In Statistical Sightings of Better Angels, Céline and I did changepoint analysis for CoW battle deaths for 95 interstate wars, 1824 to 2004, and found $\widehat{a} = 1953$ (with a confidence curve). Wish to attempt Window of Change methods there too. I then need some more careful book-keeping code for dealing with non-equal time differences. See also Dennis Christensen's talk.

- Straight Bayesian flat priors for a and for d on $1,\ldots,d_{\mathsf{max}}$ make sense ... but have difficulties. Posterior will be tempted to push the windows to one of the two sides:

$$\pi(a,d\,|\,\mathrm{data})\approx\pi(a,d)\exp\{\ell_{\mathrm{prof},\mathsf{max}}(a,d)\}\frac{1}{\sqrt{c(n-c)}},$$

  with c the midpoint of $(a,a+d)$. So there are certain mathematical differences between non-Bayes and Bayes here; 'bigger than we are used to'.

- For what happens inside the window of change I've posited simple linear change from θ_L to θ_R. I think (a) this is ok, (b) other attempts at more sophistication might not change results much, as long as the window of change is not a very long one.

(Some) references

BP Carlin, AE Gelfand, AFM Smith (1992). Hierarchical Bayesian analysis of change point processes. *Applied Statistics.*

G Claeskens, NL Hjort (2008). *Model Selection and Model Averaging.* CUP.

C Cunen, GH Hermansen, NL Hjort (2018). Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference.*

C Cunen, NL Hjort (2022). Combining information across diverse sources: newine the II-CC-FF paradigm. *Scandinavian Journal of Statistics.*

C Cunen, NL Hjort, HM Nygård (2020). Statistical Sightings of Better Angels. *Journal of Peace Research.*

NL Hjort, EAa Stoltenberg (2023). *Statistical Inference: 666 Exercises, 66 Stories (and Solutions to All).* CUP.

RG Jarrett (1979). A note on the intervals between coal-mining diasters. *Biometrika.*

M Jaudet, N Iqbal, SMC Mirza (2009). Change-point analysis of coal mining disasters data in various time resolutions. US Dept of Energy.

M Jullum, NL Hjort (2017). Parametric or nonparametric? The FIC approach. *Statistica Sinica*.

T Schweder, NL Hjort (2016). *Confidence, Likelihood, Probability.* CUP.