

Krig, fred, og statistikere



Nils Lid Hjort (med Céline Cunen og Håvard Nygård)

Department of Mathematics, University of Oslo

Faglig-pedagogisk dag, 1-xi-18

Tidligere fagligpedagogiske foredrag:

2002: Hvorfor skøytesprinterne må gå 500 meter to ganger.

2006: Stille flyter ikke Don: En kranjel mellom Nobel-prisvinnere (Солженицын против Шолохова).

2013: Hvem skal få lov til å gå titusenmeter?

2014: Skisprint: hvordan semifinalene virker inn på finalene.

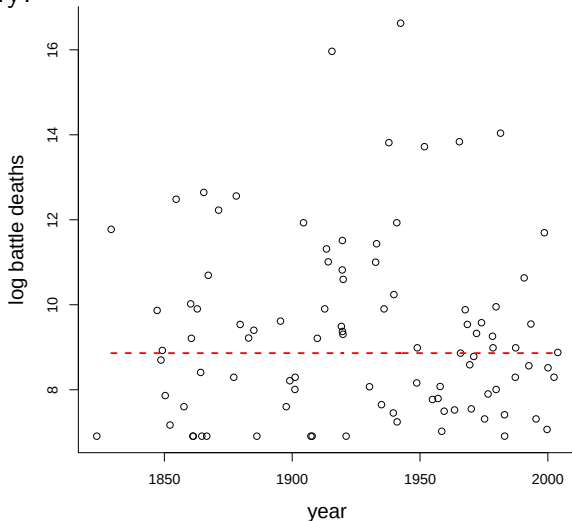
2015: Når tok Forfatter To over for Forfatter En, i verdens første roman (Tirant lo Blanch, 1460)?

2017: Hvaler, politikere, og statistikere.

2018: And now for something completely different ...

Wars over time, $\log(\text{battle deaths})$:

The CoW dataset, with (t_i, z_i) for 95 interstate wars from 1823 to 2003, with $z_i \geq 1000$. Here with $(t_i, \log z_i)$: does the world look stationary?



Some (out of many) themes

Plan:

- A The Long Peace (yes or no?, Pinker or Taleb?)
- B Testing for stationarity against a change-point
- C Finding change-point τ , with confidence curve $cc(\tau)$
- D Power laws, parametric models for $f(z)$, $z \geq 1000$
- E Results: we point to Korea 1950 (and to Vietnam 1965)
- F Including relevant covariates (e.g. democracy scores)
- G Discussion & remarks

Some (out of many) key words:

- ▶ modelling the past (why?)
- ▶ methods for assessing precision of change-point estimates
- ▶ careful modelling (with powerlaws for tails)
- ▶ combining (diverse sources of) information
- ▶ model selection, via FIC
- ▶ a (partial) list of further statistical issues

A: The Long Peace?

Some war, peace, conflict studies (quantitative and qualitative) conclude that yes!, the world **has become a bit more peaceful** than earlier. This also involves **statistical predictions** of frequency and sizes of **[godforbid]** future wars.

So there are proponents of **The Long Peace** (e.g. the famous Steven Pinker), commonly understood to mean **after WWII**.

Not everyone agrees – for several valid reasons. (i), such statements depend on technicalities and definitions of **wars** and their **consequences**, and one risks counting things which change under our feet. (ii), **Clever Statisticians A** disagree (somewhat) with **Clever Statisticians B**, even over the same data.

Some claim **the world of wars has been (and remains) more or less constant**, regarding frequency and volume (**Taleb** is constantly disagreeing with **Pinker**, etc.). **Clauset** says **he must wait 150 years** until he can be sure The Long Peace is in place. – Our estimate is lower: **0 years**.

B: Testing constancy vs. there being a change-point

Consider independent observations $y_1, \dots, y_n$ over time. Under H_0 , they are i.i.d. The alternative is that there is some τ with

$$y_1, \dots, y_\tau \sim f_L, \quad y_{\tau+1}, \dots, y_n \sim f_R,$$

with f_R different from f_L .

There are many tests for H_0 . Suppose ϕ is a focus parameter, with estimators

$$\hat{\phi}_{a,b} \approx_d N(\phi, \kappa_{a,b}^2 / (b - a + 1))$$

for each stretch $y_a, \dots, y_b$. With $\hat{\phi}_L = \hat{\phi}_{1,\tau}$ and $\hat{\phi}_R = \hat{\phi}_{\tau+1,n}$, let

$$Z_n(\tau) = \frac{\hat{\phi}_L - \hat{\phi}_R}{\{\hat{\kappa}_L^2/\tau + \hat{\kappa}_R^2/(n - \tau)\}^{1/2}}.$$

(i) We can plot it; (ii) we can read off $\max |Z_n(\tau)|$; (iii) we can derive (limiting) null distribution.

At each τ , $Z_n(\tau)$ compares past with future, and is approximately a $N(0, 1)$ under the null. There is a well-defined limit null distribution for maxima, in terms of a Brownian motion process W :

$$Z_{n,\max} = \max_{c \leq \tau/n \leq d} |Z_n(\tau)| \rightarrow_d Z_{\max} = \max_{c \leq s \leq d} |Z(s)|,$$

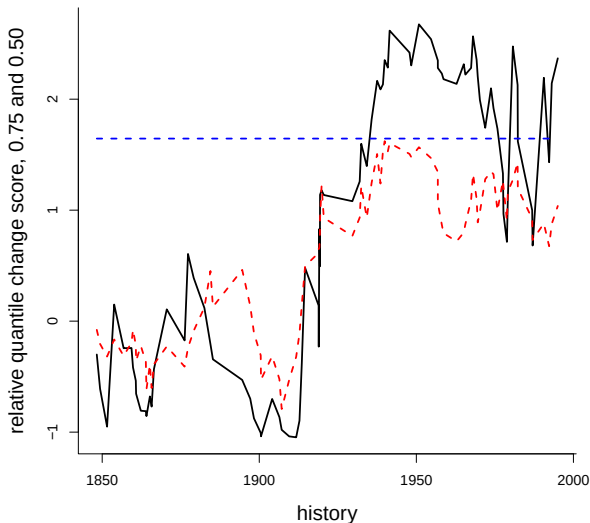
where

$$Z(s) = \sqrt{1-s} \frac{W(s)}{\sqrt{s}} - \sqrt{s} \frac{W(1) - W(s)}{\sqrt{1-s}}.$$

To construct a plot of $Z_n(\tau)$, and to read off local and global p-values, we need to start with a focus parameter:

Different foci lead to different plots (of course) and to different p-values. – We could have no change in mean, but a change in standard deviation, etc.

Two Z_n plots – for $\phi = F^{-1}(\frac{1}{2})$ (red curve, not significant), for $\phi = F^{-1}(\frac{3}{4})$ (black curve, p-value 0.022).



C: Spotting the change – and assessing its uncertainty

So H_0 of constancy is rejected in favour of there being a change-point (for several foci ϕ , but not all).

Natural setup, for change-point searching (and assessment): first build a reasonable $f(y, \theta)$ model for the y_i . Then postulate

$$y_1, \dots, y_\tau \sim f(y, \theta_L), \quad y_{\tau+1}, \dots, y_n \sim f(y, \theta_R).$$

If $\dim(\theta) = p$, then this model has $2p + 1$ parameters.

The log-likelihood profile is

$$\begin{aligned} \ell_{n,\text{prof}}(\tau) &= \max\{\ell_{1,\tau}(\theta_L) + \ell_{\tau+1,n}(\theta_R) : \text{all } \theta_L, \theta_R\} \\ &= \ell_{1,\tau}(\hat{\theta}_L) + \ell_{\tau+1,n}(\hat{\theta}_R) \\ &= \sum_{i \leq \tau} \log f(y_i, \hat{\theta}_L) + \sum_{i \geq \tau+1} \log f(y_i, \hat{\theta}_R). \end{aligned}$$

This leads to (i) the maximum likelihood estimate $\hat{\tau}$; (ii) the deviance function

$$D(\tau, y) = 2\{\ell_{n,\text{prof}}(\hat{\tau}) - \ell_{n,\text{prof}}(\tau)\}.$$

For regular models and parameters, the [classic Wilks theorem](#) for profiled log-likelihoods says that $D(\phi_0, y) \approx_d \chi_1^2$ at the true parameter, which leads to the Schweder–Hjort [confidence curve](#)

$$\text{cc}(\phi, y_{\text{obs}}) = \Pr_{\phi}\{D(\phi, Y) < D(\phi, y_{\text{obs}})\} = \Gamma_1(D(\phi, y_{\text{obs}})),$$

with the defining property that

$$\Pr_{\phi_0}\{\text{cc}(\phi_0, Y) \leq \alpha\} \doteq \alpha \quad \text{for all } \alpha \in (0, 1).$$

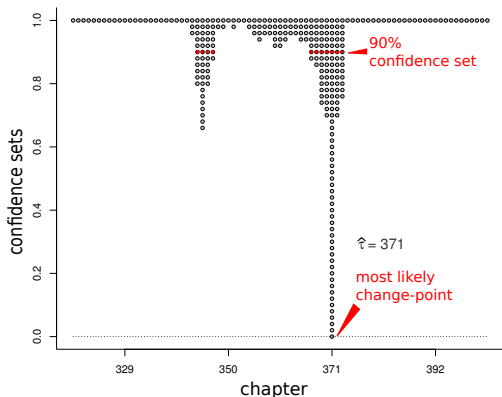
We can read off confidence intervals at all levels!

This recipe can [not be used](#) for change-points, however. Instead we use

$$\text{cc}(\tau, y_{\text{obs}}) = \Pr_{\tau}\{D(\tau, Y) < D(\tau, y_{\text{obs}})\} = K_{\tau}(D(\tau, y_{\text{obs}})),$$

with $K_{\tau}(x)$ the c.d.f. of the random $D(\tau, Y)$. This requires simulations for each candidate τ .

We summarise our inference about τ as a confidence curve. It could look like this (from a completely different dataset):



Recall that τ is a discrete parameter, so we don't quite have the usual confidence intervals, but confidence sets.

D: Modelling the fatalities distribution

The literature is in reasonable agreement: the **big wars** follow a distribution with a **power-law behaviour**, with

$$\Pr(Z \geq z) \approx c/z^\theta \quad \text{for all } z \geq z_0.$$

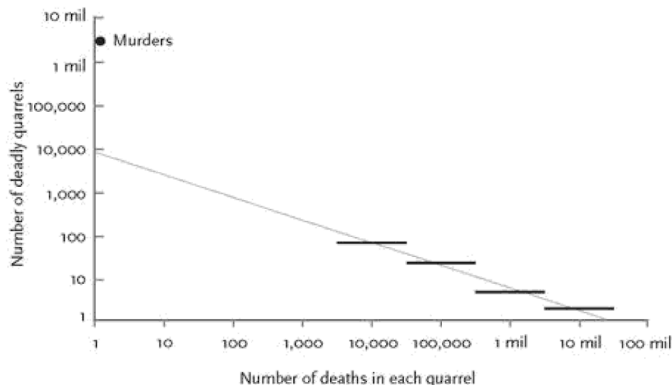


Figure 5.7 in Pinker (2011). Numbers of deadly quarrels of different sizes.

Several authors have worked along lines of (i) finding a threshold z_0 (Clauset, 2018, finds $z_0 = 7061$); then (ii) estimating power-index θ for wars bigger than z_0 .

Our approach is instead to model all wars. Among several candidates, we have used the inverse Burr, with

$$\begin{aligned}\Pr(Z \leq z) &= \left[\frac{\{(z - 1001)/\mu\}^\theta}{1 + \{(z - 1001)/\mu\}^\theta} \right]^\alpha \\ &\approx 1 - \alpha/(z - 1001)^\theta \text{ for large } z,\end{aligned}$$

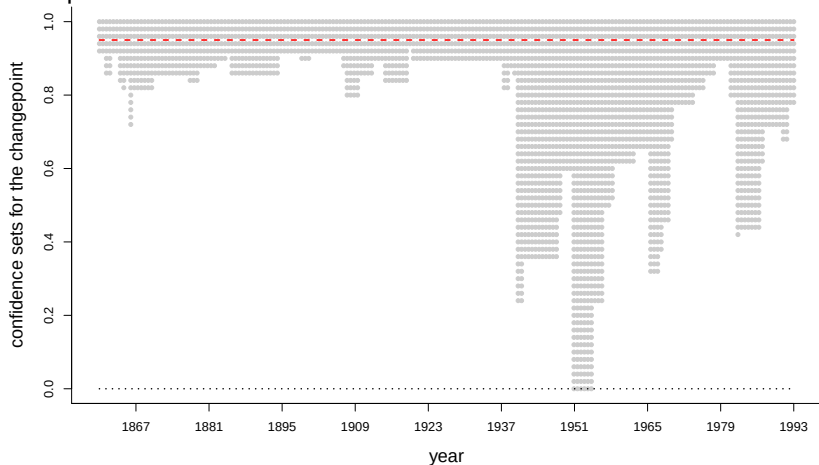
so it has the power-law tail behaviour, with power index θ .

We will use the inverse Burr distribution in our change-point framework:

$$\begin{aligned}y_1, \dots, y_\tau &\sim \text{InvBurr}(\mu_L, \theta_L, \alpha), \\ y_{\tau+1}, \dots, y_n &\sim \text{InvBurr}(\mu_R, \theta_R, \alpha).\end{aligned}$$

E: Results: Korea!

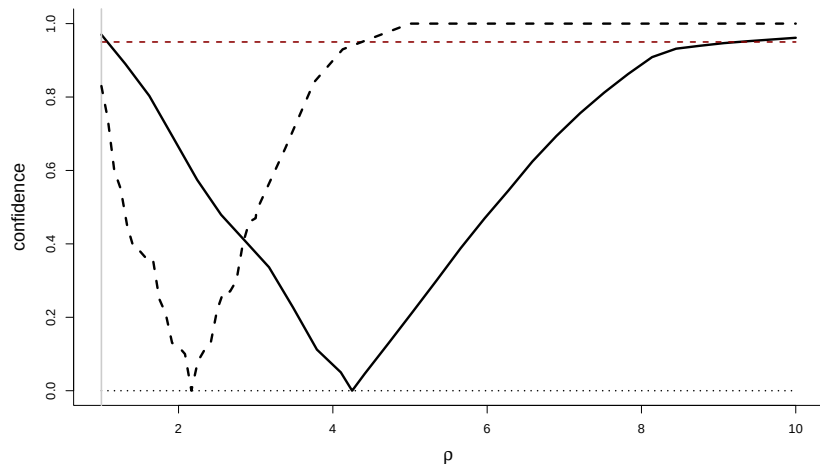
The point estimate is $\hat{\tau} = 1950.48 \Rightarrow$ the Korean War.



But ... in what direction is the change happening, and how big is it?

Confidence curve for the degree of change

Let $\rho_1 = \phi_{0.50,L}/\phi_{0.50,R}$ (dashed) and $\rho_2 = \phi_{0.75,L}/\phi_{0.75,R}$ (fully drawn),



At $\rho = 1$ we are testing H_0 : the world has been constant.

F: Incorporating covariate information

We can incorporate covariates in the change-point model. We have looked at the democracy score for each war: the mean democracy level for the countries involved, measured in the year before the war started.

Our model is then extended to

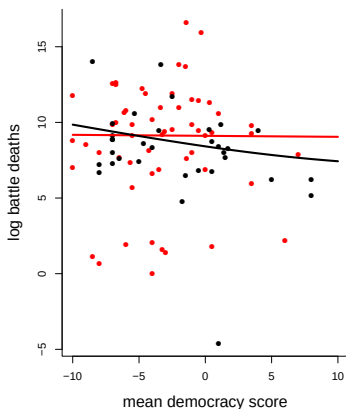
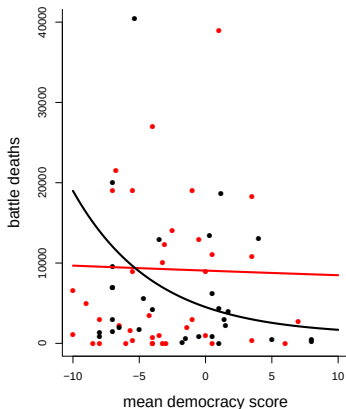
$$\begin{aligned}y_1, \dots, y_\tau &\sim \text{InvBurr}(\mu_{L,i}, \theta_L, \alpha), \\ y_{\tau+1}, \dots, y_n &\sim \text{InvBurr}(\mu_{R,i}, \theta_R, \alpha),\end{aligned}$$

with

$$\mu_{L,i} = \mu_{L,0} \exp(\beta_L x_i) \quad \text{and} \quad \mu_{R,i} = \mu_{R,0} \exp(\beta_R x_i).$$

Here $\mu_{L,0}$ and $\mu_{R,0}$ represent the baseline parameters whereas β_L and β_R reflect the effect of democracy before and after the change-point.

The effect of democracy:



Before Korea there is no effect of democracy (red line); after Korea there is a pacifying effect of democracy (black line).

G: Discussion and remarks

There are quite a few contributions investigating more-or-less the same questions with more-or-less the same data:

- ▶ Cirillo and Taleb (2016): cannot reject H_0 : the world is constant.
- ▶ Clauset (2018): cannot reject H_0 : the world is constant.
- ▶ Spagat and van Weezel (2018): some evidence for a change, in 1950;
- ▶ Cunen, Hjort, Nygård (2018): some evidence for a change, around 1950.

These conclusions are not necessarily incompatible – not-rejecting- H_0 does not imply that the null hypothesis is true.

But we have sharper tools, with stronger detection power.

What is driving the differences in the conclusions?

There are differences in modelling choices and normalisation or not with world population, but we believe that the main difference lies in the **focus**, or the lack of thereof.

- ▶ Cirillo and Taleb (2016): **cannot reject** that **the same (stationary) model** has been at work over the entire time-period;
- ▶ Clauset (2018): **cannot reject** the hypothesis that the data after WWII were generated from the **model** fitted to the data before WWII;
- ▶ Spagat and van Weezel (2018): **reject the hypothesis** that the probability of a **large war** has remained the same before and after 1950;
- ▶ Cunen, Hjort and Nygård (2018): **reject the hypothesis** that the **75% quantiles of war sizes** (and for all quantiles above 60%) are the same before and after the change-point.

What does this all mean?

- ▶ There might be good reasons to believe that **something has changed** after WWII.
- ▶ Conclusions depend on choices of models and framework, and on the **focus** of the analysis (the precise form of the question you ask).
- ▶ This is only a beginning. Whether there has been a **statistically significant trend** or not, the patterns observed in incidences and sizes of wars are worth **exploring and explaining**, and we therefore need to move from the simple models used here (with independence!), to more complex and theoretically driven models.

Should you feel reassured?

World War 3? 'Long peace' UNDER THREAT as huge conflict is OVERDUE, claims mathematician

THE WORLD could be on the brink of a huge global war because the statistics claim the "long peace" that has been enjoyed since 1945 might be an anomaly, a leading mathematician has declared.

By THOMAS MACKIE

Express (UK)

Cunen, Hjort, Nygård (politely & respectfully) disagree! The world has become (somewhat) more peaceful.

Other analyses in our [Statistical Sightings of Better Angels](#) (2018) paper:

- ▶ Per capita counting instead.
- ▶ Waiting times $t_i - t_{i-1}$.
- ▶ Change-point analyses with other models.
- ▶ Focused model selection.
- ▶ Combining (diverse sources of) information.

Combining information

Steven Pinker likes our work – and suggested to us:

Google Books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive

between and from the corpus: with smoothing of [Search lots of books](#)

Replaced **anti-war** with **anti - war** to match how we processed the books.

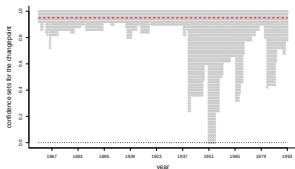


Soft data? Can we combine this information with the battle deaths analysis? What assumptions are required? Do we really believe this is the **same** change-point?

We have **methods for accomplishing this** (but one has to ‘live with one’s assumptions’).

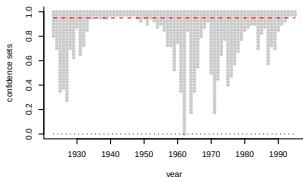
Battle death data

$$\ell_{\text{prof},b}(\tau)$$



“Anti-war” data

$$\ell_{\text{prof},n}(\tau)$$

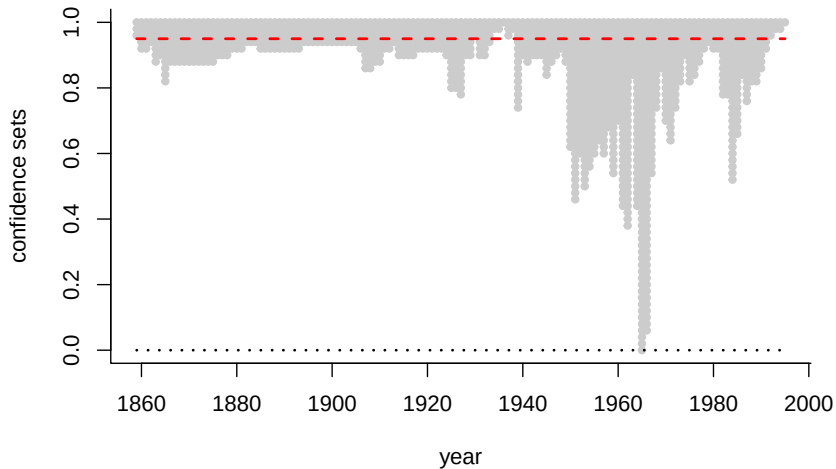


So we can simply use

$$\ell_{\text{fuse}}(\tau) = \ell_{\text{prof},b}(\tau) + \ell_{\text{prof},n}(\tau),$$

but here we do not use the Wilks theorem to produce the final confidence curve.

Combined confidence curve



Point estimate is 1965! But...

Statisticians Against War

Nils Lid Hjort (professor), Céline Cunen (PhD and soon PostDoc), Håvard Nygård (PRIO), Jens Kristoffer Haug (master student), Gudmund Hermansen (PostDoc, now førsteamanuensis BI), perhaps another master student, ...



... reading [Война и мир](#) and Pinker.

(Photos from two different articles about our work in [Titan](#). Soon more in [Apollon](#).)

(Some) references

Cirillo, P. and Taleb, N.N. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*.

Claeskens, G. and Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press.

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

Cunen, C., Hjort, N.L., Nygård, H. (2018). Statistical sightings of better angels. Submitted for publication.

Green, P.J., Hjort, N.L., and Richardson, S. (2003). *Highly Structured Stochastic Systems*. Oxford University Press.

Hjort, N.L. (2018). [Towards a More Peaceful World](#) [insert '!' or '?' here]. FocuStat blogpost.

Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Pinker, S. (2011). *The Better Angels of Our Nature*.

Schweder, T. and Hjort, N.L. (2016). *Confidence, Likelihood, Probability*. Cambridge University Press.