

War, Peace .. and the Long Peace (insert '!' or '?' here)



Nils Lid Hjort

Department of Mathematics, University of Oslo,
and [Statistical War and Peace Project](#), Centre for Advanced Research, Oslo

[Akademiet, 26/x/2023](#)

Stability and Change, august '22 til juni '23

H. Hegre og N.L. Hjort ledet CAS-prosjektet **Stability and Change**.
«Deretter redegjør kandidaten for **hensikten med og resultatene av** den vitenskapelige undersøkelsen, maksimalt 30 minutter.»

♠ 45% statsvitere, freds- og konfliktforskere

♠ 45% statistikere (teoretiske og appliedniks)

♠ 10% andre

Statistikere:

datasamlere, datavitenskapere, registerbyggere, registerkopplere, tallknusere, datasyntetikere, sannsynlighetsteoretikere, dataanalytikere, modellbyggere, algoritmikere, maskinlærere, informasjonskondensører, informasjonskombinasjonsfolk, prediksjonskonstruktører, mønstergjenkjennerne, kunstigintelligensianere, metaanalytikere, generalister, generaliserere – med matematikk, filosofi, numerikk, m.m.

Statistikere i Stability and Change

The Two Cultures: 'to explain or to predict'.

Statistikere kan bruke **de samme data** (og delvis de samme modeller) for ulike formål:

Statistiker A arbeider frem nøyaktige modeller, presise analyser, klare tolkninger, forståelse, med god kontroll på analyser og utsagn om verden (stundom for å **endre den**¹, om nødvendig).

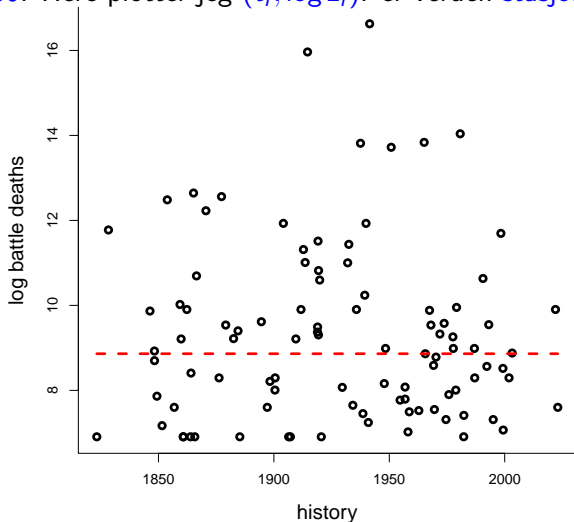
Statistiker B lager modeller som passer ganske godt med store hauger av data, av og til uten klare tolkninger, og leverer gode metoder for **klassifikasjon** («dette er en konflikt av Type III») og for **prediksjon** («her er prognoser for konflikter av Type I, II, III for 2025-2035»).

Stability and Change-prosjektet: ja takk, begge deler.

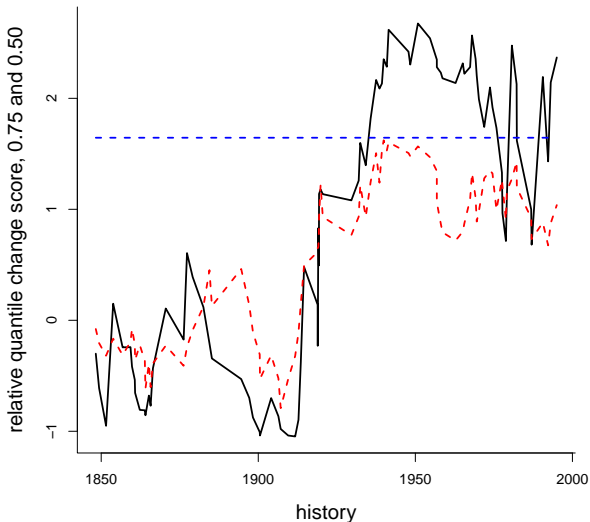
¹es kömmt darauf an, sie zu verändern

Store mellomstatlige kriger, dødstall i kamp

Fra [Correlates of War](#)-databasen, med (t_i, z_i) , starttidspunkt og antall falne, for 97 mellomstatlige kriger fra 1823 til 2023, de med $z_i \geq 1000$. Her plotter jeg $(t_i, \log z_i)$: er verden stasjonær?



Inferens og grafiske metoder fra Cunen og Hjort –
for median $F^{-1}(\frac{1}{2})$ (rød kurve, ikke signifikant),
for øvre kvartil $F^{-1}(\frac{3}{4})$ (sort kurve, signifikant, p-verdi 0.022).



Statistisk modellering av falne-i-krig (og annet)

Empirisk funn, stor litteratur: falne i **tilstrekkelig store kriger** (samt diverse annet, om mord, vold, konflikt) følger en **potens-lov fordeling**, med

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

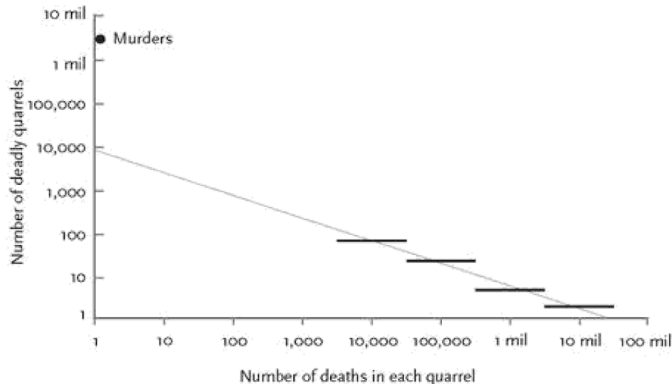


Figure 5.7 from Pinker (2011). Number of deadly conflicts of different sizes.

Vi arbeider med

- a gode statistiske modeller, som beskriver data godt;
- b men også med plausible og tolkbare bakgrunnsmodeller som forklarer (a)-type modeller.

Illustrasjon av (a): falne-i-krig følger en fordeling av denne typen:

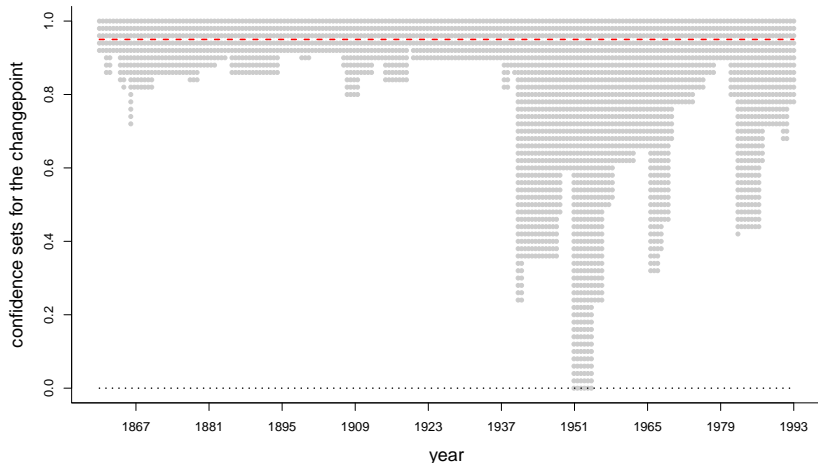
$$\Pr(Z \leq z) = \left[\frac{\{(z - 1000)/\mu\}^\theta}{1 + \{(z - 1000)/\mu\}^\theta} \right]^\alpha$$
$$\approx 1 - \alpha/(z - 1000)^\theta \text{ for large } z.$$

Data (for diverse tidsrekker, med krig, vold, konflikt) brukes for

- å estimere;
- å forstå;
- å sammeligne;
- å predikere;
- (samt, av og til) å intervenere.

En litt mindre krigersk planet ... siden 1950

Via nøyaktig modellering og analyser av falne-i-krig, kan vi estimere **endringpunktet** (overgang til noe mindre krig enn tidligere) ... som er **Korea-krigen 1950**. Et **konfidensplott** viser usikkerheten:



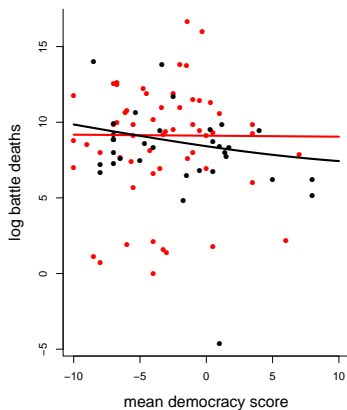
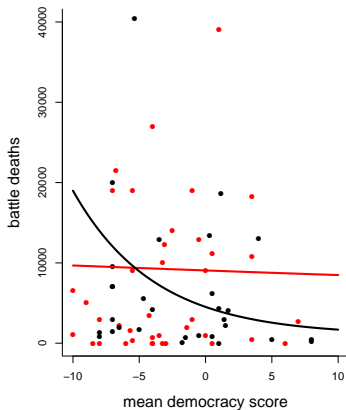
Quo vadis, demokrati (og med hvilke konsekvenser)?

Mange korrelerte temaer og perspektiver:



Har «demokrati-indeks» noen innflytelse?

Hvert land har sine **democracy score** (endeløst tema, også for CAS-prosjektet) – og slike kan brukes inne i våre modeller:



Før Korea: liten innflytelse; etter Korea: **demokrati hjelper (litt)**.

Counting the not counted

'Everyone who has been murdered should be remembered' – says [Human Rights Data Analysis Group](#) Director of Research, Patrick Ball (winner of the Rafto Prize 2021).

Men i krig-og-drama har vi [manglende data](#) (samt data av dårlig kvalitet). Hvordan telle dem som ikke er tallet?

Srebrenica, 1995: hvor mange drepte? Fra Brunborg, Lyngstad, Urdal (2003): Lister med døde, fra [ICRC](#) og [PHR](#):
 $N_{1,1} = 5712$, $N_{1,0} = 1586$, $N_{0,1} = 192$.

Hvordan kan vi estimere

$N_{0,0}$ = antall døde, som ikke er havnet på noen av de to lister?

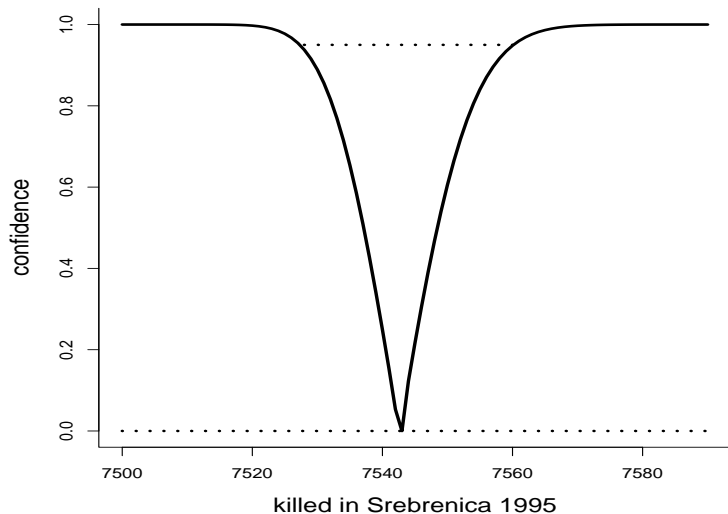
Estimering av [nullboksen](#) (hvor mange er de usynlige) dukker opp i diverse sammenenger. Hjort og andre i CAS-prosjektet har utviklet nye metoder for dette.

Srebrenica 1995

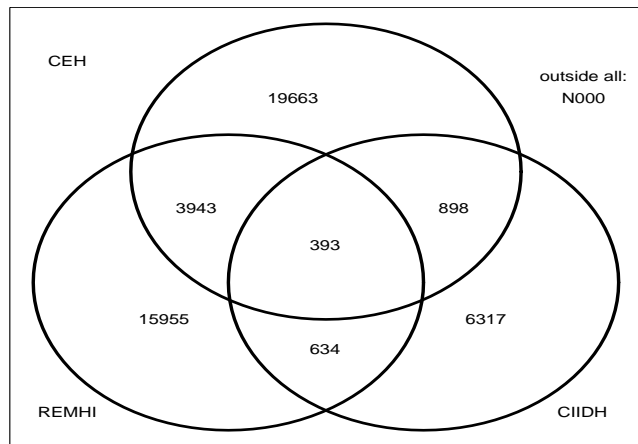
Fra Brunborg, Lyngstad, Urdal (2003): ICRC og PHR har lister:

$N_{1,1} = 5712$, $N_{1,0} = 1586$, $N_{0,1} = 192$.

Estimat 7543; intervall [7528, 7560]; $\hat{p} = 0.967$, $\hat{q} = 0.783$.



Hvor mange ble drept i Guatemala, 1978–1996?

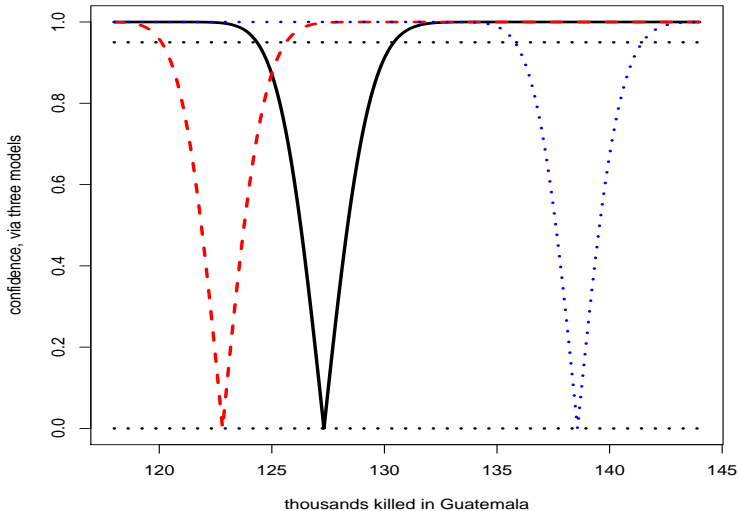


Lister fra REMHI, CEH, CIIDH:

$n_{1,1,1} = 393$, $n_{1,1,0} = 3943$, $n_{1,0,0} = 15955$, $n_{1,0,1} = 634$,

$n_{0,1,1} = 898$, $n_{0,1,0} = 19663$, $n_{0,0,1} = 6317$.

Hvor mange i 000-boksen?



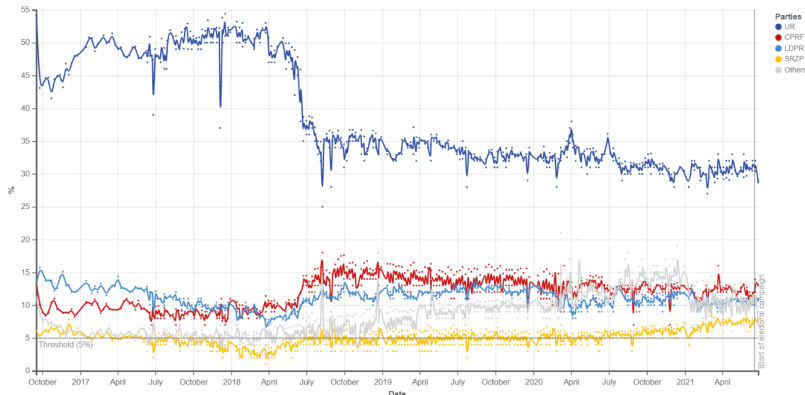
3-para: **138,576**, with 135,794 to 141,453 (width 5,659)
 4-para: **122,812**, with 120,100 to 125,634 (width 5,534)
 5-para: **127,314**, with 124,341 to 130,415 (width 6,074)
 Ball (1999): **132,174** (with a standard error of 6,568?).

Opinion polls – in troubled nations, in troubled times

Opinion polls [\[edit \]](#)

This section is an excerpt from [Opinion polling for the 2021 Russian legislative election § Graphical summary](#). [\[edit \]](#)

Pre-campaign [\[edit \]](#)



January 2019 to September 2021, polling for the different parties

Единая Россия faller i popularitet, mai-juni 2019 – hvorfor?

How to combine (very) different types of information?

Steven Pinker har «fulgt oss», og har interessert seg for om Cunen-Hjort-metoder for **kombinasjon av (svært) ulike informasjonskilder** kan trekke inn Ngram-analyser:

Google Books Ngram Viewer

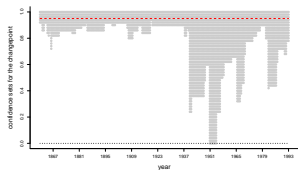
Graph these comma-separated phrases: case-insensitive

between and from the corpus with smoothing of

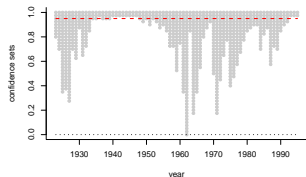
Replaced anti-war with anti - war to match how we processed the books.



Battle deaths data
(hard work leads to this):



'Anti-war' Ngram data
(hard work leads to this):



We have methods for combining also such (very different) information sources:

- ♣ Number of conflicts, their intensity, number of deaths.
- ♣ Levels of aggression (or the opposite), military training, rhetoric (by government, and in society).
- ♣ Protests, reading and interpreting sociale media.
- ♣ Degree of cooperation (export-import), common interests.

Men *hvorf* blir det potenslov-oppførsel?

Since Lewis Fry Richardson (1944, 1948, 1960), it's an established **empirical fact** the deadly quarrels (and other types of violence) exhibit **power-law behaviour**:

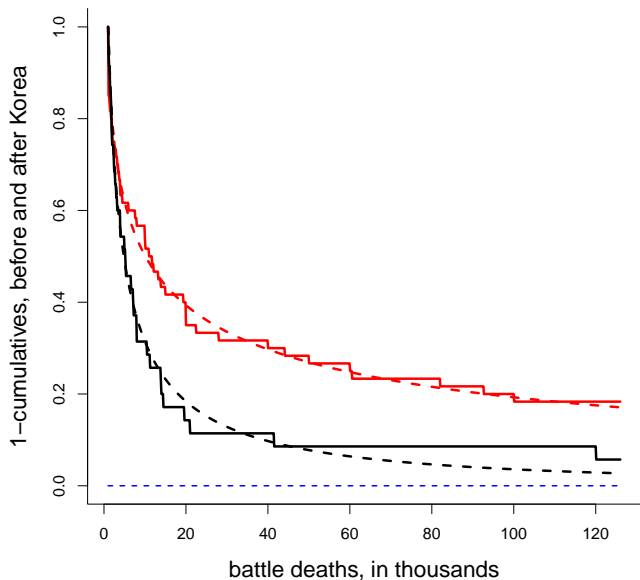
$$\Pr(T \geq t) \approx \frac{c}{t^\theta} \quad \text{for all large } t.$$

This means **long, fat tails** – if a war has 10,000 deads, we should not be statistically shocked if it escalates to 50,000.

But **why?** We can search for **sets of plausible conditions** that **imply** power-law tails. One such example (Nils, 2023): consider

$$Z_n(t) = \sum_{i \leq [nt]} J_i c(i/n)^\alpha U_i \quad \text{for } t \geq 0,$$

with U_i being i.i.d., and J_i being Bernoulli with $p_i = 1/i$: steadily rarer, but steadily bigger. **Theorem:** $Z_n \rightarrow_d Z$, and **time to hit threshold for Z** has power-law tails.



Nils-modell, tilpasset pre-Korea og etter-Korea battle deaths.

Post-CAS

Samarbeidet mellom **Blindern-statistikere** (og venner) og **PRIO-forskerne** (og venner) har **har hatt seg en boost**, med **≈ 60 foredrag gjennom 6 workshops** – det fortsetter, og skal fortsette.

Noen statistikere har deltidsstillinger på PRIO, med prosjekter som kan vokse. Hjort og andre veileder PhD- og masterprosjekter på UiO med CAS-temaer.

For å sikre **signifikant videreutvikling** trenger vi **noen faste rammer** (utover **gode data, gode ideer, gode intensjoner**).

Visse **Stability and Change**-temaer kan komme inn i **INTEGREAT** (nytt Centre of Excellence, Oslo, oppstart høsten 2023) – deriblant **kunnskapsbasert maskinlæring** og **bedre kombinasjon av (svært) ulike informasjonskilder**.

Kan **FFI** kople seg på?