One Thousand is Unfair, Two Thousand is Fair

FOCUS DRIVEN STATISTICAL
INFERENCE WITH COMPLEX DATA

Nils Lid Hjort

Department of Mathematics, University of Oslo

FocuStat Kombo-Kjøkken
November 2018

The Best Medal-Grabbing Games Ever

'How a Tiny Nation Won the Most Medals (By Far) at the Olympics':

Some (out of many) themes

Plan:

A A binomial process approach to medal counting

B A quick ski sprint story

C 2 x 500 m ('how I changed the Olympics')

D 2 x 1000 m

E Yet other skating-statistical issues

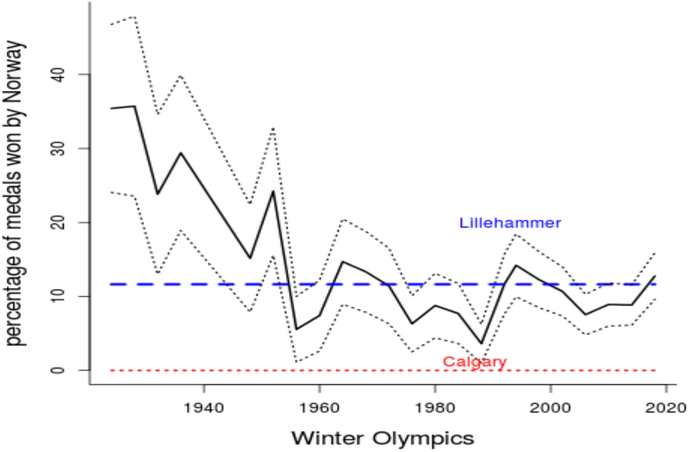
F Science communication

Some (out of many) key concepts:

- Getting hold of (enough) relevant data
- Building focused models for particular purposes
- Combining information, here: across multiple events, and possible other biophysical sources
- Communicating results

A: A binomial process approach to medal counting

14 + 14 + 11 = 39 medals, new Olympic world record ... but with steadily more events per Olympics.



B: Ski sprint interlude

Nikita i Aleksandr – Молодцы! But why only bronze for Northug?

Olympic cross-country ski sprint, Sochi, 2014:

The cross-country skiers must go through
prologue $\Longrightarrow$ quarter-final $\Longrightarrow$ semi-final $\Longrightarrow$ final
With A and B as labels for first and second semi-final:

1	O.V. Hattestad	A	1	M.C. Falla	A
2	T. Peterson	A	2	I.F. Østberg	B
3	E. Jönsson	B	3	V. Fabjan	A
4	A. Gløersen	A	4	A.U. Jacobsen	A
5	S. Ustiugov	B	5	I. Ingemarsdotter	A
6	M. Hellner	A	6	S. Caldwell	B

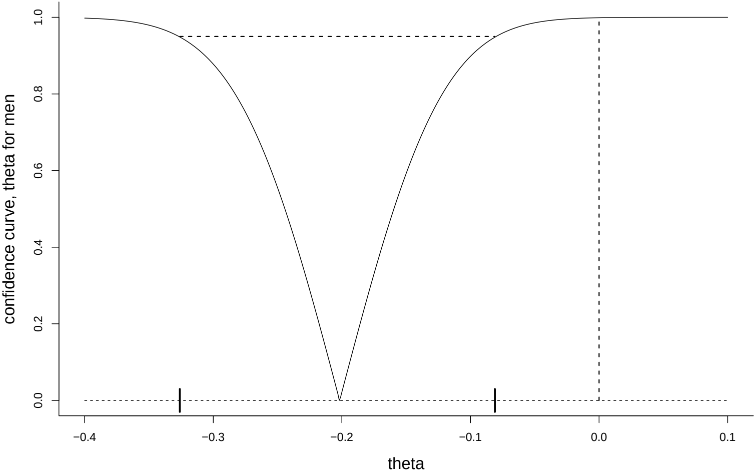
So A skiers appear to do better than B skiers ...

Just by chance? I doubt it (!).

One cannot deduce much from one single event – but statisticians can see (potential) structural differences between A-skiers and B-skiers, by watching a long enough list of events.

Confidence curve for unfairness parameter

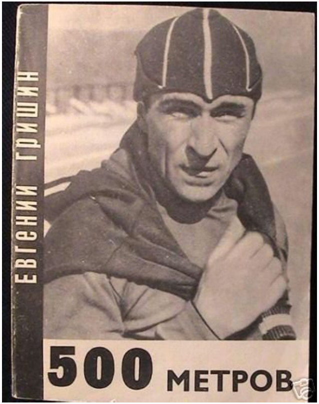
A certain model, with $\theta < 0$ implying A has better chance than B (after watching 57 World Cup and Olympics events):



Combo?, with medical parameters?

C: 2 x 500 (how I changed the Olympics)

"He drew lane with anxious attentiveness – and could not conceal his disappointment: First outer lane! He threw his blue cap on the ice with a resigned movement, but quickly contained himself and picked it up again. With a start in inner lane he could have set a world record, perhaps be the first man in the world under 40 seconds. Now the record was hanging by a thin thread – possibly the gold medal too. At any rate he couldn't tolerate any further mishaps."



What I did (in 1994)

The problem for the 500 m sprint lies with the last turn. Try this, with speed 60 km/h, when fatigue sets in:

$$a = mv^2/r, \quad \text{with } r = 25.0\,\mathrm{m} \text{ (inner)}, \ r = 30.0\,\mathrm{m} \text{ (outer)}.$$

The (average) guy in last inner lane has more trouble than the (average) guy in last outer lane.

I went hunting for the Olympic Unfairness Parameter:

d = difference between what FastGuy would have had with last outer and what he had in last inner.

Not easy to define, conceptually and operationally (it's counterfactual, and depends on both skater and race conditions, etc.).

The statistician finds relevant data; builds a model where the d has a clear role and interpretation; then estimates d; and reaches a conclusion (here: testing $d = 0$, giving confidence intervals, etc.).

Valid thought (but it doesn't work well): Watch & clock lots of races with last inner and compare with lots of races with last outer. Then do a school-book t-test.

The problem is that we're hunting for a Very Tiny Difference (well, if it exists at all), and need a better statistical microscope.

I hunted down (in 1994, very manually) lots of paired results from World Sprint Championships. Skaters run the 500 m on Saturday and Sunday, and change start lanes. I used

$$\begin{aligned} y_{i,1} &= a_1 + bx_{i,1} \pm \tfrac{1}{2}d + \delta_i + \varepsilon_{i,1}, \\ y_{i,2} &= a_2 + bx_{i,2} \pm \tfrac{1}{2}d + \delta_i + \varepsilon_{i,2}, \end{aligned}$$

with $x_{i,1}$ and $x_{i,2}$ passing times after 100 m; with $\delta_i \sim \mathrm{N}(0, \kappa^2)$ (variation from skater to skater); independent $\varepsilon_{i,j} \sim \mathrm{N}(0, \sigma^2)$ (variation from race to race); coefficients a_1, a_2 (ice conditions); b (using 100 m information); and $\pm\frac{1}{2}d$ depending on inner/outer.

I needed $4 + 2 = 6$ parameters to learn well enough about d, for one World Sprint event.

After having watched say 25 top skaters over two days, I reach my estimate $\widehat{d}$ – which is not sufficiently precise. So I needed such data, with championship by championship analysis, giving me $\widehat{d}_1, \ldots, \widehat{d}_k$, for k events.

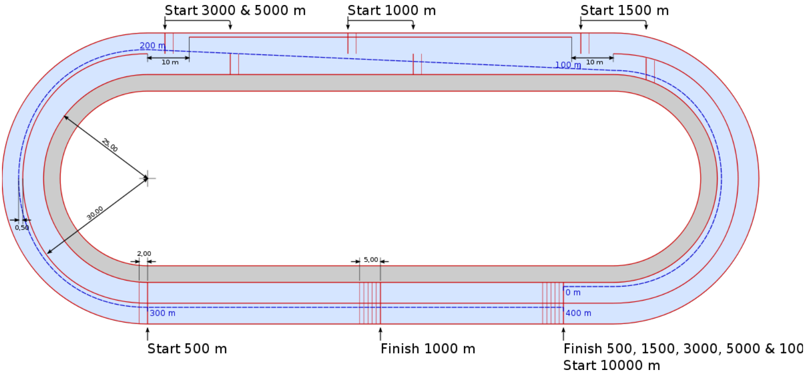
I needed about $k = 10$ to get sufficient precision. A complication is combining summary information across events – since the underlying $d_1, \ldots, d_k$ will not be identical. This calls for meta-analysis.

Conclusion: $\widehat{d}_{\rm grand} = 0.06$, with 99% confidence interval $[0.035, 0.085]$. This is enough to have medals change necks.

I told the Norwegian Skating Association, and proposed 2×500 as the balanced solution; they took it to the Technical Committee for Speed Skating under ISU; 37 nations voted (one nation, one vote!); they took it to the IOC; ... and from 1998 Nagano to 2014 Sochi, the skaters did the 500 m twice.

D: 2 x 1000

inner guy has 3 inners + 2 outers.
outer guy has 3 outers + 2 inners.



Issues: it's asymmetric; first inner guy has more room to accelerate, and an easier finish; might benefit from drafting («om å spise opp en rygg»).

The data

So the 1k is more complex than a 500 m. Again I hunt for the Unfairness Parameter d: the average time difference between inner and outer, across top skaters of the world (without mishaps).

I'm again blessed by the annual World Sprint Championships – lovely to watch!, one kilometre in 67 seconds! (first Norwegian champion since Bjørang 1975 and Rønning 1981), and giving paired results, with one race Saturday and one race Sunday:

1	Lorentzen	o	16.61	41.67	69.21	i	16.65	42.18	69.81
2	Nuis	o	16.57	41.81	68.97	i	16.76	42.16	69.11
3	Verbij	o	16.57	42.03	69.22	i	16.66	42.25	69.47
4	Ihle	o	16.62	42.19	69.73	i	16.74	42.14	69.29
5	Poutala	i	16.59	42.18	69.82	o	16.52	41.97	69.87
6	Whitmore	o	16.59	41.89	69.96	i	16.67	42.13	70.10

The bi-Gaußian linear mixed effects model

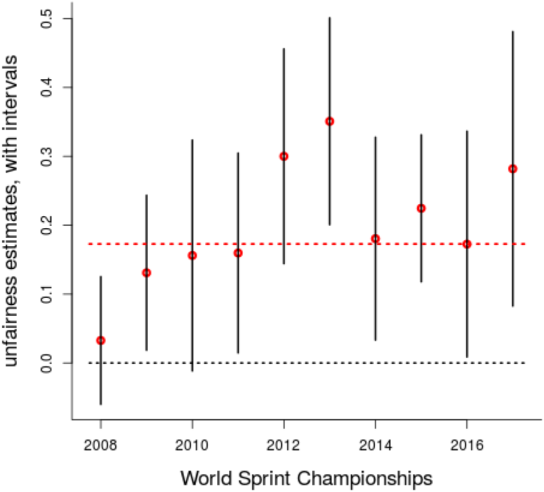
The results, for the Saturday and Sunday races, for skater i:

$$\begin{aligned} y_{i,1} &= a_1 + bu_{i,1} + cv_{i,1} + \tfrac{1}{2}dz_{i,1} + \delta_i + \varepsilon_{i,1}, \\ y_{i,2} &= a_2 + bu_{i,2} + cv_{i,2} + \tfrac{1}{2}dz_{i,2} + \delta_i + \varepsilon_{i,2}, \end{aligned}$$

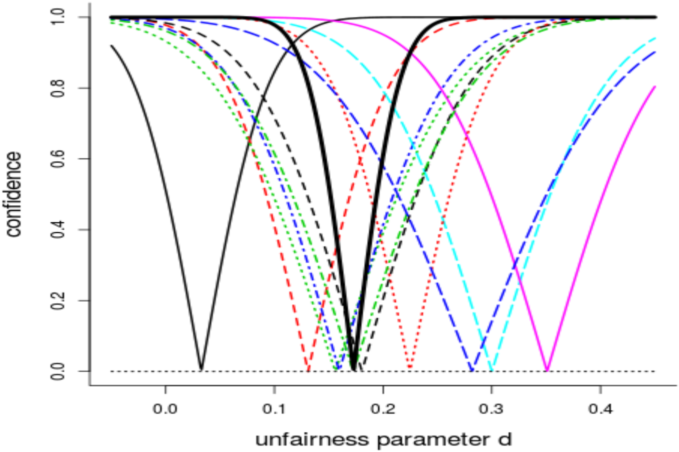
with $(u_{i,1}, u_{i,2})$ passing times after 200 m; $(v_{i,1}, v_{i,2})$ passing times after 600 m; the δ_i taking care of variations skater-to-skater; the $\varepsilon_{i,j}$ variations from race to race for same skater; a_1, a_2, b, c regression fixed effects parameters; $(z_{i,1}, z_{i,2})$ coded as $(-1, 1)$ of $(1, -1)$ for inner-outer starts – and finally d the unfairness parameter.

For each Sprint World Championships I watch (with c. 25 skaters, and with careful outlier screening), I can fit the model, via maximum likelihood, and estimate the $5 + 2 = 7$ parameters, and finally read off the estimate $\widehat{d}$, along with a confidence curve.

For Sprint Championships 2008 to 2017: point estimates $\widehat{d}_j$ and 95% confidence intervals: very clear advantage for Inner Guy:

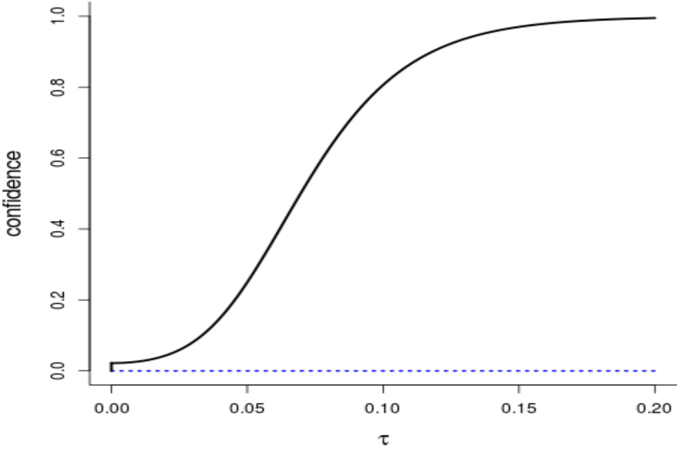


I need meta-analysis for summing-up inference based on $\widehat{d}_1, \ldots, \widehat{d}_k$ from different events – and use II-CC-FF (Independent Inspection, Confidence Conversion, Focused Fusion), from Schweder and Hjort (2016), Cunen and Hjort (2018). Conclusion: $\widehat{d}_{\rm grand} = 0.187$, with 99% interval $[0.108, 0.266]$.



The spread among $d_1, \ldots, d_k$

We have $\widehat{d}_j \mid d_j \sim \mathrm{N}(d_j, \sigma_j^2)$ and take $d_j \sim \mathrm{N}(d_0, \tau^2)$, giving $\widehat{d}_j \sim \mathrm{N}(d_0, \sigma_i^2 + \tau^2)$. The CD $C(\tau)$ has point-mass at zero:



E: Statisticians watching speedskating ... yet other themes

- Who should be allowed to skate the 10k? Statistical prediction methods do better than the ISU rules
- Evolution of records (will they ever stop?)
- The rise and fall of nations (Spengler watches speedskating)
- Doping (does it work?, is it visible?)
- Mass-sport vs. the smaller elites
- Cultural shifts, economics, resources

F: Science communication

(Some) references

Cunen, C., Walløe, L., and Hjort, N.L. (2018). Focused model selection for linear mixed models, with an application to whale ecology.

Cunen, C. and Hjort, N.L. (2018). Combining information across diverse sources: The II-CC-FF paradigm.

Efskind, L.D. (1970). *Soldatens lille røde bok.* Pax, Oslo.

Friesinger, A. (2010). *Mein Leben, mein Sport, meine besten Fitness-Tipps.* Goldmann, Berlin.

Гришин, Е. (1969). 500 метров. Молодая Гвардия.

Hjort, N.L. (1994). Should the Olympic speed skaters run the 500 m twice? Statistical Research Report, Department of Mathematics, University of Oslo.

Hjort, N.L. (2017). The semifinals factor for skiing fast in the finals. *FocuStat Blog Post.*

Hjort, N.L. (2018a). One Thousand is Unfair, Two Thousand is Fair. *FocuStat Blog Post.*

Hjort, N.L. (2018b). The Best Metal-Grabbing Games Ever. *FocuStat Blog Post.*

Holum, D. (1984). *The Complete Handbook of Speed Skating.* High Peaks Cyclery, Lake Placid.

Jansen, D. (1994). *Full Circle.* Villard Books, New York.

Keller, E. (1968). *74 Schritte zum Ziel.* Copress-Verlag Mnchen.

Knausgård, K.O. (2018). *The Hidden Drama of Speedskating.* Essay, The New York Times, 1-Feb-2018.

LeMay Doan, C. (2002). *Going for Gold.* McClelland & Stewart, Toronto.

Rosa, D. and Hjort, N.L. (1999). Who Won? The Annual Speedskating Race of the Burg of Ducks. Speedskating World.

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