

On Just How Surprised We Ought to Have Been When Bolt Set his World Records



Nils Lid Hjort

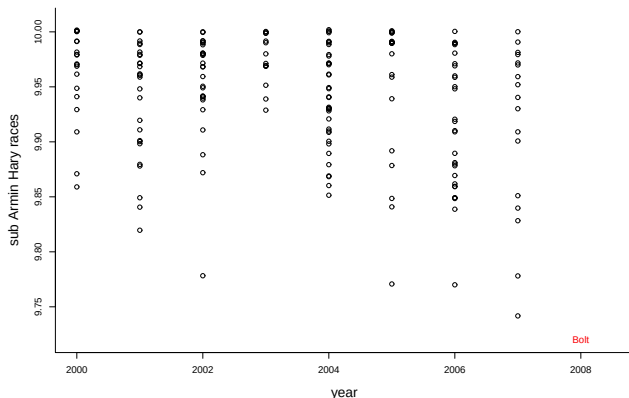
Department of Mathematics, University of Oslo

Det 18. norske statistikarmøtet, June 2015, Solstrand

The problem: Quantifying surprise

On 31 May 2008, **Usain Bolt** did 9.72. How surprised were we?

I spent the following night tracking down (and then analysing) the 195 **sub-10.00-races** recorded from seasons 2000, 2001, ..., 2007.



General problem formulation:

Suppose $y_1, \dots, y_n$ are observed, from some distribution. How surprised are we, if in fact

$$y_{n+1} \geq \max_{i \leq n} y_i + w,$$

for some threshold w ? Answer: **It depends** (!).

- A A **parametric model** for extreme values
- B A formula for $p(w) = \Pr\{\max(Y_1, \dots, Y_N) \geq w\}$ with N Poisson, with estimation
- C **Confidence distribution** for associated surprise measures
- D Concluding remarks (and **further questions**)

A: The extreme value distribution

With r_i 100 m running times, I work with

$$y_i = 10.005 - r_i$$

and care only about sub-Armin-Hary races, i.e. $y_i > 0$.

Theorems of Fisher, Tippet, Gnedenko imply (under some conditions) that these very fast races follow the c.d.f.

$$G(y, a, \sigma) = 1 - (1 - ay/\sigma)^{1/a}$$

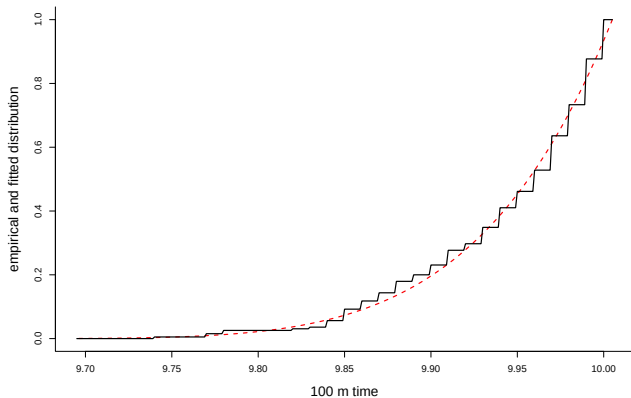
for some (a, σ) . For the $n = 195$ (**bona fide**) races of 2000–2007, can hence form the **log-likelihood function**

$$\ell_n(a, \sigma) = \sum_{i=1}^n \{-\log \sigma + (1/a - 1) \log(1 - ay_i/\sigma)\}.$$

This leads to maximum likelihood **estimates** (and **estimates of their precision**).

The **two-parameter extreme value model** fits data very well:

$$\hat{a} = 0.1821 \text{ (sd} = 0.0701\text{)}, \hat{\sigma} = 0.0745 \text{ (sd} = 0.0074\text{)}.$$



B: Distribution of best race in a season: a formula

For a season with N top races (below Hary threshold), consider

$$p = p(a, \sigma) = \Pr\{\max(Y'_1, \dots, Y'_N) \geq w\}.$$

With $N \sim \text{Pois}(\lambda)$, we find

$$p = p(a, \sigma) = 1 - \exp\{-\lambda(1 - aw/\sigma)^{1/a}\}.$$

I use $\lambda = 195/8 = 24.375$, rate of top races per year. For each threshold w we may **estimate** p and its approximate **standard error**.

With $w = 10.005 - 9.72 = 0.285$, for 31 May 2008, I find $\hat{p} = 0.035$:

3.5% probability of seeing a **9.72 or better in the course of 2008**, as judged from 1 January 2008.

C: Full confidence distribution for $p = p(a, \sigma)$

Traditional approach: ML is approximately normal, so with **delta method** $0.035 \pm 1.96 \widehat{\text{sd}}$ would be a 0.95 interval, etc. This doesn't work well here, in spite of $n = 195$, and in spite of $(\widehat{a}, \widehat{\sigma})$ being approximately binormal: $p = p(a, \sigma)$ is not close to linear in this part of the parameter space.

Better (both for approximation quality and for representing uncertainty): the **confidence distributions** of Schweder and Hjort (**Confidence, Likelihood, Probability**, 2015). I compute the **confidence curve**

$$\text{cc}(p) = \Gamma_1(D_n(p))$$

via the profiled deviance:

$$D_n(p) = 2\{\ell_n(\widehat{a}, \widehat{\sigma}) - \ell_{n,\text{prof}}(p)\}.$$

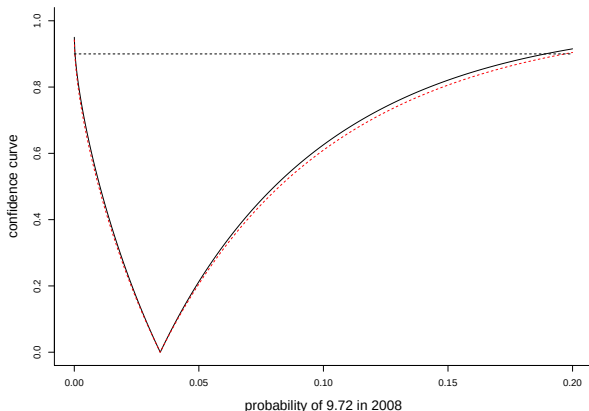
The $\text{cc}(p)$ curve gives both the estimate (3.5%) and each confidence interval. These are **highly skewed**; the 90% interval is [0%, 18.9%], etc.

Even better with a mean correction to make distribution of $D_n(p)/(1 + \varepsilon_n)$ closer to χ_1^2 . For the case of

$$p = \Pr\{\text{seeing a 9.72 or better in 2008}\}$$

(as happened on May 31),

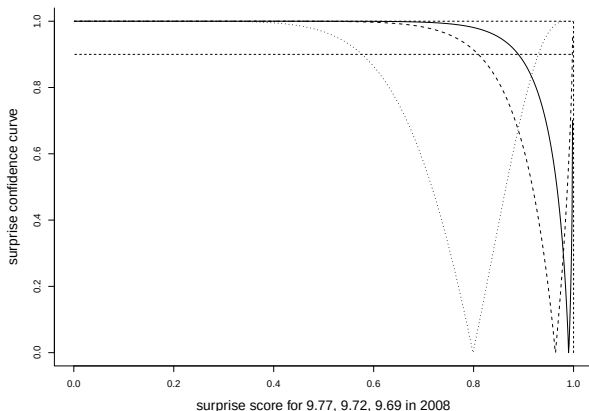
$$cc^*(p) = \Gamma_1(D_n(p)/1.07).$$



Can now carry out these computations each time I wish to have my surprise level analysed. My **shock barometer** is

$$q = 1 - p, \quad \text{with} \quad p = \Pr\{\text{seeing what I just saw}\}$$

(both **shock estimate** and **confidence curve**), perhaps made into percent. For **Gay's 9.77** (June 2008), **Bolt's 9.72** (May 2008), **Bolt's 9.69** (August 2008):



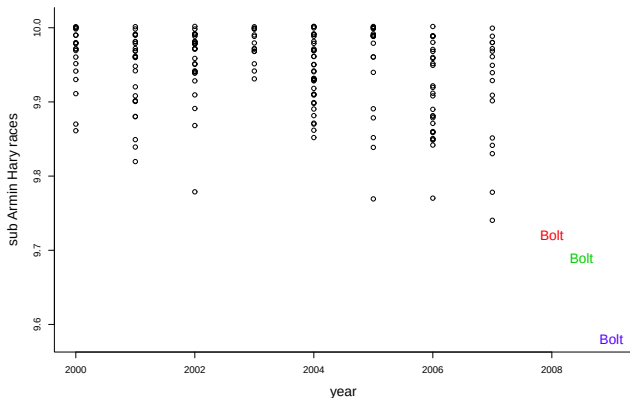
Shock barometer (set up and calibrated as of January 2008):

Gay 9.77 (June 2008): shock = 79.7

Bolt 9.72 (May 2008): shock = 96.5

Bolt 9.69 (August 2008): shock = 99.3

Bolt 9.58 (August 2009): shock = beamonesque 100.0



D: Concluding remarks (and further questions)

- (a) Theory & assumptions: Races are independent and the world was approximately stationary from 2000 to 2007. This is ok, even with multiple races for the same individuals (extension of Gnedenko theorems etc.).
- (b) Theorems relate to races-below-threshold, and I used Armin Hary. Fine-tuning does not change much.
- (c) Theory is harder (and needs more fine-tuning) when there is a trend over time, $y_i = m(x_i, \beta) + \varepsilon_i$ etc.
- (d) The shock barometer (with confidence curve) may be used to compute and compare degree of wondrousness of different world records.

(e) Modelling top results from different events jointly: **hierarchical modelling**, empirical Bayes. This would give **better understanding** and **improved precision** of the different shock barometers.

(f) Applications of theory and methods **abound** for other areas: finance, waves, floods, wind, weather, engineering.

(g) The theory and methods may be used to predict world (and national) record evolution. When will Bøkkø or Pedersen run the 1500 m faster than Rodal ran the 800 m?

