



ISSN 0803-8953

Tilfeldig gang

Nr. 1, årgang 43

6 2026

Utgitt av Norsk statistisk forening

Redaksjonelt

Artikler

Det ultimate femtusenmeterskeiserrennet. <i>Nils Lid Hjort</i>	3
Anyone for chess? Analysing chess ratings above high thresholds. <i>Nils Lid Hjort</i>	6

Tankenøtter

Pusleri nr. 64. <i>Jostein Lillestøl</i>	16
Pusleri nr. 63 - løsning. <i>Jostein Lillestøl</i>	17

Møter og konferanser

The 40th International Workshop on Statistical Modelling i Oslo	20
---	----

Meldinger

Nytt fra Oslo senter for biostatistikk og epidemiologi (OCBE)	21
Nytt fra NTNU	21

Fra redaksjonen

I denne utgaven av TG skriver Nils Lid Hjort om (minst) to av sine mange diller, nemlig skøyter og sjakk og Jostein Lillestøl byr på nok et pusleri. I tillegg finner du meldinger fra OCBE og NTNTU.

Redaksjonen takker bidragsyterne til denne utgaven og ønsker både dem og alle lesere en god sommer!

Sist, men ikke minst, oppfordres alle medlemmer til å bidra med tekst og/eller tips til neste utgave av TG.

Ta kontakt via verahk@uio.no.

Vennlig hilsen,

Vera Kvisgaard

DET ULTIMATE FEMTUSENMETERSKEISERENNET

Nils Lid Hjort

MATEMATISK INSTITUTT, UNIVERSITETET I OSLO

ETTER GULL & GULL I HEERENVEEN-VM

KRONIKK (PLUSS BONUS), DAG OG TID, 13. MARS 2026



Figure 1: Når du renn femtusen, hugs dei delikate tyrannosaurus rex-armrørsler i svingane.

Me lærte på skulen at det var synd på Nederland. Landet er flatt, det har 33 gonger så mange innbyggjarar per kvadratkilometer som Noreg, og det er uråd å staka seg ut frå desse kvadratkilometrane til ein skog for å gå på ski.

Men dei greier seg fint på andre vis, vart me minte um, me som var på Thialf stadion 5.–8. mars i det skeiseshistoriske året 2026. Alt går på skjener, isen er blank, folk er høgreiste og tolerante, me får bananar og eple gratis i vatningspausane, dei heiar på Sander Eitrem og Ragne Wiklund og ser på oss med noko eg tolkar som vyrdnad, når me, med hjarta på rette staden, syng «Ja, vi elsker» to gonger på rad i B-dur. Berre tankane fór raskare enn Tjalve, som me veit.

Det er visstnok haugar av folk som kan ligatabellar og lagoppstillingar utanåt. Fire herlege døgn med likesinna på skeisetur peikar mot eit anna sjikt, nokre meir introverte nerdar, med eit breitt utsyn over alle sportslege tema, til frukost og kvelds på hotellet.

Her er hundre års kulturhistorie (kven var eigenleg fyrst under 39.0, Lepjosjkin, Muratov eller Börjes? og korleis greidde den demokratiske tribunefolkerepublikken å vinna over oppmannen og juryen på Bislett 3. januar 1969?); sosiologi (ulike skilnader å spora mellom norske, nederlandske, polske og italienske fans); psykologi (ja, X er verdsmeister på kvar trening, men hovudet låser seg i galne banar når det er meisterskap); idrettsfag (Y trenar gale, Z får ikkje pendelen til å svinga); og spådomskunst (Eitrem kjem til å læppa Stolz, stol på meg).

Me er òg kunstløpdommarar og gjev karakterar til stilikonet Sander, armane låste på ryggen, og til Ragne, som har byrja med nokre delikate tyrannosaurus rex-armrørsler i svingane. Og driv

sjølvsagt med statistikk og tabellar og nasjonsrekordar i minst 400 meter-lange banar.

Å sjå på skeiser er som kjend også ei mental utfordring. Me skal ikkje berre gle oss over estetikk, over fart, over den som vinn, over dei jublande røyster. Me må ha tal- og poengkontroll undervegs: Poengsummer og differansar må reknast ut i sanntid – og sjekkast med dei andre rundt oss på tribunen. Eitrem, Stolz, Jilek, Semirunniy spelar ein strykekvartett i fire satsar med bartókske kompleksitetar.

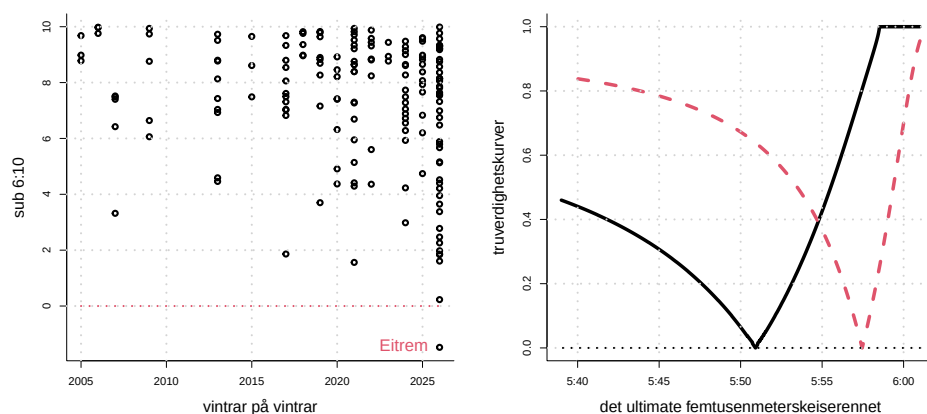


Figure 2: Venstre: samlede renn under 6.10, 129 før denne OL-sesongen pluss ein himmelsk hærske med 47 nye. Nedre datapunkt er på lista over nasjonale stoltheiter. Høgre: statistiske konfidenskurver for kva verdsrekorden vil verta, om tjue år, sett fra før sesongen byrja (raudt) og i den posteitremiske verd (sort). Punktestimatet er 5.51,38, men med ei betydeleg og asymmetrisk uvisse.

Eg er sjølv professor i matematisk statistikk og kan tillata meg også litt idiosynkratiske prosjekt, der eg lèt mine såkalla kulturelle interesser koma inn på banane mine. Lesaren må gjerne finna *Six-Minute Man Sander Eitrem 5:58.52 – First Man below the 6:00.00 Barrier* og *The Best Metal-Grabbing Games Ever: How a Tiny Nation Won the Most Medals (By Far)* der ute på verdsveven, jamsides den storfelte og nærast episke Klassekampen-kronikken frå desember 2025; vitja gjerne òg FocuStat-sidone. Her skal eg svara på spørsmålet som er på sin plass: Kva vert 5000-verdsrekorden um 20 år?

Datamaterialet mitt, henta frå eigne skuffer, er alle femtusenmeterrenn i verdshistoria under 6.10: dei 126 mellom 2005 og 2025 og den himmelske hærsken av 47 nye slike renn denne olympiske sesongen.

Eg treng ein halvtime med krit og tavle for å forklara alle detaljane, komplett med greske bokstavar og matematiske utrekningar frå nærskyldte område, men tre nykelmoment er desse:

(1) Det å modellera og analysa kva som skjer langt ute i halane av ei fordeling, er monaleg vanskelegare enn å halda seg til det normale, det som skjer på midten.

Dessutan må ein vera varsamare med alt som heiter hypotesar. Her skal me opp på hovudfagsnivå, og eg brukar variantar av Gumbel-fordelingar (han med *Vier Jahre Lüge*, 1919, og *Vier Jahre Politischer Mord*, 1922, før han vart ekstremmatematikar).

(2) Det er lettare å levera eit presist punktestimat for ein parameter inne i ein modell enn for

eit utfall frå framtida, som alltid vil ha sin innebygde variasjon.

(3) Etter fagleg strev finn eg estimatet 5.51,38 for det ultimate femtusenmeterskeiserennet. Men her er det monaleg, og dessutan asymmetrisk, uvisse, som me ser av konfidenskurvene i figuren – den raude prikkete linja er frå den zweigske verda av i går, den svarte er frå den posteitremske. Det er innanfor det statistisk fullt mogelege at rekorden framleis heiter 5.58,52 til OL i Narvik 2042.

NASJONSREKORDAR I SUB-6.10-SAMSKIPNADEN, PER 8. MARS 2026

5.58.52 EITREM Sander	NOR Inzell	26 0124	1
6.00.23 LOUBINEAUD Timothy	FRA SLC	25 1114	1
6.01.56 van der POEL Nils	SWE SLC	21 1203	1
6.01.84 DAWSON Casey	USA Calgary	25 1121	1
6.01.86 BLOEMEN Ted-Jan	CAN SLC	17 1210	1
6.01.98 JILEK Metodej	CZE Inzell	26 0124	2
6.02.98 ROEST Patrick	NED SLC	24 0128	1
6.04.21 FARTHOFFER Alexander	AUT SLC	25 1114	1
6.04.23 GHIOTTO Davide	ITA SLC	24 0128	2
6.06.75 GROSS Gabriel	GER SLC	25 1114	3
6.07.04 LEE Seung-Hoon	KOR Calgary	13 1110	3
6.07.81 SEMIRUNNIY Vladimir	POL Inzell	26 0124	6
6.08.36 LIU Hanbin	CHN Calgary	25 1121	4
6.08.64 TROFIMOV Sergej	RUS SLC	21 1203	1
6.08.76 SWINGS Bart	BEL SLC	21 1203	4
6.08.83 SASAKI Shomu	JPN SLC	25 1114	4
6.09.68 MICHAEL Peter	NZL SLC	17 1210	7

Me vonar at DEN og BLR og KAZ og SUI og EST og LAT kan verte med dei òg, om ikkje for lenge. Skeisaren frå UZB med 6.08,96 (nieuw baanrecord!, i Heerenveen!), vert ikkje rekna med i denne samanhengen, av di han den gongen representerte eit anna land, som me kjem i hug.



Figure 3: Det hev sjølvsagt vore rom for S. Eitrem's «Overtru og trollskap hjå dei gamle grekarar» i bokhylla mi på Blindern (frå 1921, og med ei ny umsetjing av Lukans «Løgnhalsen»), sidan S. Eitrem fyrst synte lovnad i verdsklasse på 5000 meter i 2021, det vil segja hundre år seinare.

Anyone for chess?

Analysing chess ratings above high thresholds

Nils Lid Hjort

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSLO

JANUARY 7, 2026

ABSTRACT. Suppose some cleverness score parameter is sufficiently interesting to be defined and then measured, perhaps for different strata of specialists or for the broader population. Such phenomena could have Gaussian distributions, when it comes to all players in a stratum, but when interest focuses on the very tails, for the top few percent, those above certain high thresholds, different models are called for, along with the need to analyse such based on the listed top scores only. In this note I develop such models and tools, and apply them to the top-100 and above 2100 points lists for regular chess ratings, for the currently active 14671 men and 753 women, as given by the FIDE, January 2026. It is argued that even when two or more distributions have close to identical expected values, or medians, even smaller differences in variance may explain gaps for the few very best ones.

Key words: chess ratings, gender gap, from tails to the main bulk, Magnus Carlsen, models for scores above thresholds, participation volume, variance ratio

1 INTRODUCTION: THE SIZES OF MAXIMA



Figure 1: Queen’s Gambit.

The beautiful game of chess is for nearly everyone: just grab the pieces and your nearby board, on your laptop or on your phone, or perhaps your analogue bishops and rooks and knights and pawns, and play on. It takes beautiful minds to play it really wondrously, however. “Not In 10,000 Lifetimes Will We See Chess Like This Again”, proclaims *Epic Chess*, among the many chess experts on the net, with hundreds of thousands of followers, analysing in detail Magnus Carlsen’s complicated march towards winning the World Rapid and World Blitz, in Doha, Christmas week 2025. So statisticians should be interested too.

In this note I study the chess ratings, as per the Elö scale used by the International Chess Federation FIDE in clubs and tournaments all over the world, not on the population scale but for the very clever ones, those with ratings above a high threshold, like $r_0 = 2100$. Before I come to chess, however, with models and analysis methods developed in Section 2, it is useful to build some intuition around extreme scores, under natural and simple sets of assumptions. The aim is partly to identify what the driving factors are, for extreme scores inside different strata.

Consider therefore M_n , the maximum of $X_1, \dots, X_n$, an i.i.d. sample from the $N(\xi, \sigma^2)$. We may think in terms of a stratum of people and their heights, with M_n the height of the tallest. Then, with $c_n \doteq d_n$ indicating $c_n/d_n \rightarrow 1$,

$$E M_n \doteq \xi + \sigma(2 \log n)^{1/2},$$

which captures the essential ways in which the tallest can be very tall (or perhaps not so very tall, after all): (i) ξ should be big; (ii) higher participation, i.e. a bigger n , matters, but only on the very slowly growing $(\log n)^{1/2}$ scale; and (iii) it is also helpful with a not small σ . There is a similar approximation also in the equicorrelation case, where individuals in a stratum or population have ‘something in common’, modelled as $\text{corr}(X_i, X_j) = \rho$ for all $i \neq j$. In that case,

$$E M_n \doteq \xi + (1 - \rho)^{1/2} \sigma(2 \log n)^{1/2}, \quad (1.1)$$

so with more similarity than under pure independence, the maxima follow the same pattern, modulo a reduction factor; more independence between individuals imply bigger extremes.

For illustration, consider countries A and B where something of sufficient interest is recorded for their people have distributions $N(\xi, \sigma_A^2)$ and $N(\xi, \sigma_B^2)$. The ‘winners’ from A and B can be expected to have scores

$$\xi + \sigma_A(2 \log n_A)^{1/2} \quad \text{and} \quad \xi + \sigma_B(2 \log n_B)^{1/2},$$

in terms of population sizes n_A and n_B . The excess ratio, B compared to A, is essentially

$$r = \frac{\sigma_B(\log n_B)^{1/2}}{\sigma_A(\log n_A)^{1/2}}.$$

For the participation ratio, note that if A is e.g. 3 times bigger than B, the sample size ratio is

$$\frac{(\log 3 + \log n_B)^{1/2}}{(\log n_B)^{1/2}} = \left(1 + \frac{\log 3}{\log n_B}\right)^{1/2},$$

indicating that the difference matters, when it comes to the highest-scoring levels, but not by much, when it comes to very high sizes. The tallest Dutch cannot be expected to be much taller than the tallest Norwegians, assuming here these countries have the same grand average and the same grand variances, even if the Dutch population is three times that of the Norwegians. Neither is it entirely impossible that a guy from Lommedalen is taller than all Russians.

In many cases, therefore, the standard deviation ratio matters rather more than the sample sizes. To get even with B, A needs

$$\frac{\log n_B}{\log n_A} = \frac{\sigma_A^2}{\sigma_B^2}, \quad \text{or} \quad n_A = n_B \exp(\sigma_B^2/\sigma_A^2),$$

featuring the crucial variance ratio. For illustration, suppose A and B have $N(100, 15.0^2)$ and $N(100, 16.0^2)$; then B will indeed tend to have the tallest scores. Country A needs to be 15 percent bigger than B in order to make up for the small 6 percent standard deviation increase.

2 A PARAMETRIC MODEL FOR RATINGS OVER THRESHOLD

My two-parameter model for scores above threshold has density of the form

$$f(x, a, \theta) = \Gamma(a+1)^{-1} \exp\left\{-\left(\frac{x-r_0}{\theta}\right)^{1/a}\right\} \frac{1}{\theta}, \quad \text{for } x \geq r_0, \quad (2.1)$$

monotonically decreasing on its domain. We shall see in Section 3 that the fit to chess scores data is very good. Here a is driving the fine aspects of the upper tail, whereas θ is a scale parameter reflecting the variability of the distribution.

2.1 Properties. A little exercise verifies that the integration constant in (2.1) is the right one, and that the model also can also be represented as

$$X = r_0 + \theta V_a^a \quad \text{where } V_a \sim \text{Gamma}(a, 1). \quad (2.2)$$

Formulae for moments and quantiles may be deduced from this. In particular,

$$\mathbb{E} X = r_0 + \theta \frac{\Gamma(2a)}{\Gamma(a)}, \quad \text{Var } X = \theta^2 \left[\frac{\Gamma(3a)}{\Gamma(a)} - \left\{ \frac{\Gamma(2a)}{\Gamma(a)} \right\}^2 \right]. \quad (2.3)$$

Also useful is an explicit formula for the cumulative distribution function (c.d.f.), via (2.2):

$$F(x, a, \theta) = P(r_0 + \theta V_a^a \leq x) = G_a(((x-r_0)/\theta)^{1/a}) \quad \text{for } x \geq r_0, \quad (2.4)$$

in terms of the c.d.f. G_a for the $\text{Gamma}(a, 1)$. Furthermore, solving $F(x, a, \theta) = p$ for x leads to the quantile function

$$F^{-1}(p, a, \theta) = r_0 + \theta G_a^{-1}(p)^a. \quad (2.5)$$

The maxima of samples of players from some given stratum, like all active players in a country, follow the pattern of (1.1), though with an adjusted logarithmic speed, depending on the a . Specifically, with M_n the maximum of an i.i.d. sample from (2.1), we have

$$\mathbb{E} M_n \doteq r_0 + \theta (\log n)^a \quad (2.6)$$

for growing participation volume n . The Gaussian case corresponds to $a = \frac{1}{2}$ and the exponential to $a = 1$; the chess scores have tails between these two classical cases. Formula (2.6) follows from extreme value theory; we may point to Gumbel (1919, 1922, 1958), Embrechts et al. (1997). There is a Gumbel distribution limit for the maximum here, with $(M_n - a_n)/b_n \rightarrow_d G$, the Gumbel distribution with c.d.f. $\exp(-\exp(-g))$, for suitable slowly increasing a_n and b_n , with consequences also for all quantiles; see the details in Remark B of Section 5.

2.2 Estimation and inference. Consider a stratum of interest, with ratings $x_1, \dots, x_n$ for its n individuals. We need to estimate the parameters (a, θ) in the (2.1) model.

Suppose first that the full set $x_1, \dots, x_n$ is available. The log-likelihood is then

$$\ell(a, \theta) = \sum_{i=1}^n \left\{ -\log \Gamma(a+1) - \left(\frac{x_i - r_0}{\theta} \right)^{1/a} - \log \theta \right\}, \quad (2.7)$$

which can be maximised numerically. When needed, for numerical precision or for bigger datasets, the maximisation can be reduced to a function of a alone, namely $\ell(a, \hat{\theta}(a))$, as the log-likelihood for given a is maximised for

$$\hat{\theta}(a) = n^{-1} \sum_{i=1}^n (x_i - r_0)^{1/a}.$$

Intriguingly and importantly, parameters may be estimated also in cases where only the say the top-100 ratings are listed; in various databases for chess and performance sports only such top- k lists are available, for suitable cutoff k . Sorting these top k values as $x_1 > \dots > x_k$, then, one merely knows that the other $n - k$ ratings are inside the $[r_0, x_k]$ box. The statistical information, clearly less complete compared to knowing all the data, can be represented by the likelihood

$$L_k(a, \theta) = F(x_k, a, \theta)^{n-k} f(x_k, \theta) \cdots f(x_1, \theta),$$

i.e. with log-likelihood

$$\ell_k(a, \theta) = (n - k) \log F(x_k, a, \theta) + \sum_{i=1}^k \log f(x_i, a, \theta), \quad (2.8)$$

which again can be maximised to fit the available top- k dataset. This takes knowing the participation volume, i.e. the n , though.

3 CHESS RATINGS, FOR MEN AND WOMEN, PER JANUARY 2026

From the FIDE website I have got hold of the 15,424 players, active as of January 2026, with regular ratings equal to or above the threshold $r_0 = 2100$, with $n_m = 14671$ male and $n_w = 753$ female players. Top-rated Magnus Carlsen has 2840, whereas Hou Yifan with 2613 leads the women table; we note, for further analysis below, that the top-100 rated players are all men.

We can now fit the parameters of the (2.1) density by maximising the log-likelihood (2.7), for the two categories. Approximate standard deviations may also be computed, for the basic model parameters and for smooth functions thereof, via the delta method; see e.g. Hjort & Stoltenberg (2026, Chs. 2, 5). In some detail, for the maximum likelihood (ML) estimators,

$$\begin{pmatrix} \hat{a} \\ \hat{\theta} \end{pmatrix} \approx_d N_2 \left(\begin{pmatrix} a \\ \theta \end{pmatrix}, \hat{J}^{-1} \right),$$

with $\hat{J}$ the observed Fisher information matrix, i.e. minus the Hessian at the ML position. Also, for any focus parameter $\mu = \mu(a, \theta)$, like the mean, the standard deviation, or a quantile, we have

$$\hat{\mu} = \mu(\hat{a}, \hat{\theta}) \approx_d N(\mu, \hat{\tau}^2), \quad \text{with } \hat{\tau}^2 = \hat{c}^t \hat{J}^{-1} \hat{c},$$

where $\hat{c}$ is the gradient of the μ evaluated at the ML. Results are as follows, for the two strata, with estimates and standard errors (estimated standard deviations):

Table I:

	men		women	
a	0.689	0.013	0.612	0.052
theta	209.28	3.49	194.86	12.74
mu	2241.52	0.99	2221.53	3.60
sigma	119.67	1.07	98.31	1.58
median	2210.79	1.05	2198.23	3.98

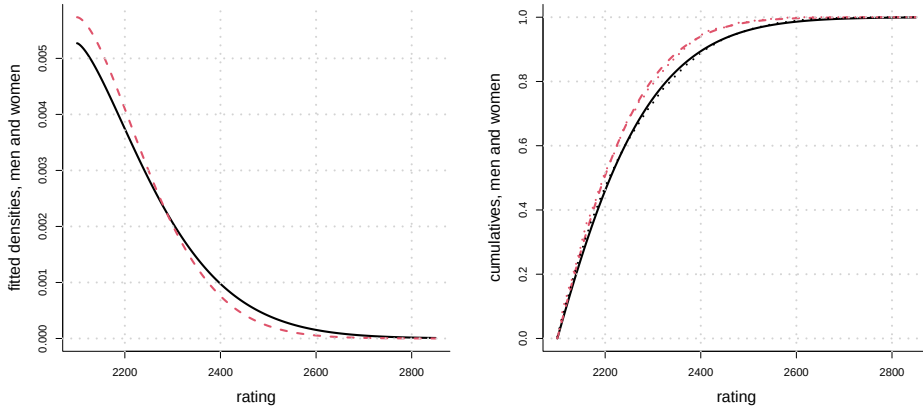


Figure 2: Left panel: fitted densities $\hat{f}_m$ and $\hat{f}_w$, for men (black, full) and women (red, slanted) having rating above 2100; they are not far apart but the right tail is dominated by the men. Right panel: fitted c.d.f.s $\hat{F}_m$ and $\hat{F}_w$, along also with the empirical c.d.f.s (very slightly non-continuous, as dotted curves). The parametric and nonparametric are very close, i.e. the model used is fully adequate.

The estimated densities $\hat{f}_m$ and $\hat{f}_w$ for men and women are displayed in Figure 2, left panel; these are seen to be close for the main bulk of players, but with a significantly fatter right tail for the men. That aspect, which shows up also for c.d.f.s and quantile functions, is not coupled with the higher participation number for men, per se, but by the upper performances of the two groups. A mild caveat, though, is that the estimated curves and quantities for the women group are less precise than those for the men.

In Figure 2, right panel, both the parametric and empirical c.d.f.s are portrayed, i.e.

$$\begin{aligned} \hat{F}_m(x) &= F(x, \hat{a}_m, \hat{\theta}_m), & \hat{F}_w(x) &= F(x, \hat{a}_w, \hat{\theta}_w), \\ F_{\text{emp},m}(x) &= n_m^{-1} \sum_{i=1}^{n_m} I(x_{m,i} \leq x), & F_{\text{emp},w}(x) &= n_w^{-1} \sum_{i=1}^{n_w} I(x_{w,i} \leq x). \end{aligned}$$

We learn from this that the model (2.1) is even surprisingly adequate; formal monitoring of $\hat{F}_w - F_{\text{emp},w}$ shows no significant deviation from zero, etc.

In Figure 3, left panel, we also plot the difference $S_{\text{emp},m} - S_{\text{emp},w}$ (black, wiggly), with $S_{\text{emp},m}(x)$ and $S_{\text{emp},w}(x)$ defined as the remaining fractions above x . So at all levels x the male proportion above it is a few percent more than for the females. The red slanted curve is the corresponding parametrically estimated difference, again indicating a very good fit. The lower and upper blue wiggly curves form a 90 percent confidence band for the $S_m - S_w$ difference, indicating also that the differences noted are statistically significant. Finally, that figure's right panel displays the estimated quantile functions, again with the parametric fits very close to the empirical ones. Visibly and significantly, there is a gender gap at the very top levels (for the populations studied, those above 2100 regular rating, for those active as of January 2026).

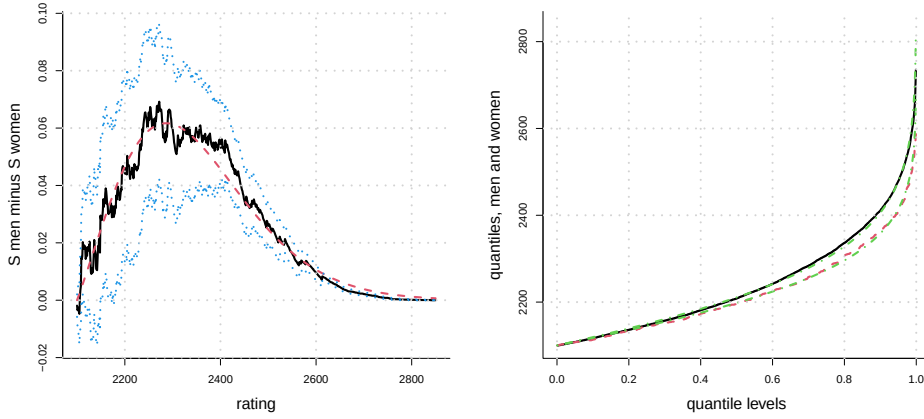


Figure 3: Left panel: the difference $S_m(x) - S_w(x)$, empirical (black, wiggly) and parametrically fitted (red, slanted, smooth), along with a 90 percent confidence band. Right panel: empirical and theoretical quantile functions Q_m (black, full) and Q_w (red, slanted). The model fit is excellent.

4 ASSESSING THE GENDER GAP

We have seen above, see e.g. Table I, that the $a_m = a_w$ hypothesis is not in conflict with the data. For the two standard deviation estimates 119.67 and 98.31, however, with standard errors 1.07 and 1.48, we have

$$t = \frac{\hat{\theta}_m - \hat{\theta}_w}{(\text{se}_m^2 + \text{se}_w^2)^{1/2}} = \frac{21.26}{1.91},$$

indicating that the male variance is much stronger than the female one. This also agrees with our findings and figures above. Whereas Figure 2, left panel, shows the estimated direct densities $\hat{f}_m, \hat{f}_w$, Figure 4, left panel, shows their logarithms, i.e. $\log \hat{f}_m$ and $\log \hat{f}_w$, showing more clearly that the male tail is bigger and fatter than the female tale.

We hence already know well that there are significant differences between the top men and top women players, for regular chess ratings, as of January 2026. It is instructive to build other

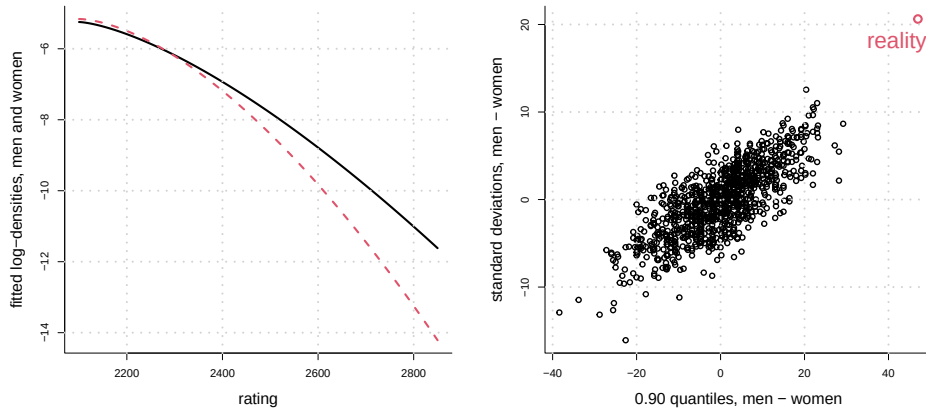


Figure 4: Left panel: the estimated log-densities $\log \hat{f}_m$ and $\log \hat{f}_w$, indicating that the man tail is fatter and longer than the woman tail; more top players are, indeed, men. Right panel: bootstrap points (A^*, B^*) for 1,000 bootstrap null distribution datasets, along with the factual $(A_{\text{obs}}, B_{\text{obs}})$, for A the 0.90 quantile difference and B the standard deviation difference.

formal tests as well. A class of such tests take the following form, using bootstrap samples. Pooling the men and women together, with $n = n_m + n_w = 15424$ chess scores, sample randomly $n_m = 14671$ of these, labelled x_m^* , the remaining $n_w = 753$ labelled x_w^* . From such a random bootstrap split of the full dataset, we may compute discrepancy scores, like

$$A = (F_{\text{emp},m}^*)^{-1}(0.90) - (F_{\text{emp},w}^*)^{-1}(0.90) \quad \text{and} \quad B = \text{sd}(x_m^*) - \text{sd}(x_w^*),$$

the differences of 0.90 quantiles and of standard deviations. Doing this say 1,000 times yields bootstrap null distributions of A and B , under the no-difference hypothesis. This is portrayed in Figure 4, right panel. Then we compute the actually observed $(A_{\text{obs}}, B_{\text{obs}}) = (47.0, 20.6)$, plotted in the right up corner of that figure. We see again, in a clear fashion, that the null hypothesis of gender equality is firmly rejected; the differences A_{obs} and B_{obs} are far outside the null hypothesis range.

5 CONCLUDING REMARKS

In this note I have introduced a two-parameter model for scores above thresholds, and demonstrated that it very adequately describes the statistics of top players in chess. The model, with relevantly estimated parameters, may be used to monitor progress over time, to predict overall level performances in tournaments, and can surely be applied to other top scores phenomena, in biology or otherwise.

To round off this note a few comments are as follows.

Remark A. Same tail index for men and women. Slightly sharper estimates might be given, than those obtained in Table I of Section 3, via estimation in the three-parameter model which takes $a_m = a_w$; model selection analysis, with the AIC and the BIC, cf. Claeskens & Hjort (2008, Chs. 2,3), indicate that this three-parameter reduction has a slightly better explanatory

power than needing two plus two parameters. It is also of interest to see in this fashion that the tail parameters are quite close for the two groups (the null hypothesis $a_m = a_w$ is not refuted), indicating that the primary factor underlying the current gender gap in top chess is not participation, or different skills for the main bulk of top players, but variability.

Remark B. Sample extremes. Suppose $u_1, \dots, u_n$ are an i.i.d. sample for the density $f(u) = \Gamma(a+1)^{-1} \exp(-u^{1/a})$ on the halfline, with tail behaviour dictated by the a parameter. For large u , the survivor function is

$$\bar{F}(u) = 1 - F(u) \doteq cu^{1-1/a} \exp(-u^{1/a}),$$

with the dominant term being the stretched exponential. The tail is Gaussian for $a = \frac{1}{2}$ and exponential for $a = 1$. The sample winner $M_n = \max_{i \leq n} u_i$ will have $\mathbb{E} M_n \doteq (\log n)^a$. More precisely, there is a Gumbel limit,

$$(M_n - b_n)/a_n \rightarrow_d G, \quad \text{with } a_n = a(\log n)^{a-1}, \quad b_n = (\log n)^a,$$

which leads to

$$\mathbb{E} M_n = (\log n)^a + \gamma_e a (\log n)^{a-1} + o((\log n)^{a-1}),$$

with $\gamma_e = 0.5772\dots$ the Euler constant.

For our chess ratings model, then, in view of the representation (2.2), the expected top rating, for a stratum with n players above 2100, is

$$\begin{aligned} \mathbb{E} M_{n,m} &\doteq 2210 + 210.22 (\log n)^{0.68}, \\ \mathbb{E} M_{n,w} &\doteq 2200 + 179.04 (\log n)^{0.61}, \end{aligned}$$

for men and women. These estimates, or predictions for future years or other strata, can be sharpened somewhat in view of Remark A.

Remark C. The Rashomon Syndrome. Statistical modelling and inference can be very good at explaining the statistical essence of what goes on for some perhaps complicated phenomenon studied, in terms of pointing to mechanisms, predicting what happens next, spotting and assessing differences between groups, etc. etc. A sound caveat is that of the so-called Rashomon Syndrome, however; rather different statistical ‘explanations’, from perhaps different models and arguments, might turn out to be explaining the data equally well. So even though the (2.1) model succeeds very well in fitting all such higher ratings chess data, it should not be seen as the canonical ‘explanation’ for what goes on in the minds of the 15,000 most clever chess players of our time.

Remark D. Learning the full iceberg from only seeing the top. As explained with (2.8), the full model can be learned from recording say only the top-100, as long as we also know the participation level, i.e. the size n above the threshold in question. This should be useful when it comes to analysing chess ratings for various strata or groups for which the full datasets are not fully known.

Remark E. Other gender gap literature. There are many articles and stories, published in journals or as blog posts or expert views in social media, pertaining to the gender gap in chess; the extent to which it is present, what likely causes might be, and how to amend, in the direction of less imbalance. A simple search on Google Scholar finds more than a hundred publications. Good articles published on the web include *What gender gap in chess?*, by Wei Ji Ma, who argues that participation is the dominating factor. The present note does not disagree with this, but points to the crucial presence of *statistical variation* within strata.

Remark F. Application to other domains. Clearly, ‘high scores above threshold’ extreme value data abound in other application areas, from finance and meteorology and sports, see e.g. Hjort (2026), to biological traits, like heights and academic skills. Note e.g. via Remark D that if we know the tallest 50 US basketball players, in a season with 500 players, we may actually infer the full distribution.

Thanks. I greatly enjoyed following the Norwegian Broadcasting Corporation’s home channel televising the World Rapid and World Blitz from Doha, Qatar, during the Christmas week of 2025, captained by Torstein Bae, pedagogically guiding us through the jungles of complexities towards Magnus Carlsen’s magnificent wins. I am also indebted to Arne Tjølsen who volunteered assisting me in getting hold of the right data-files from FIDE.

REFERENCES

- CLAESKENS, G. & HJORT, N. L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.
- EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. London: Springer.
- GUMBEL, E. J. (1919). *Vier Jahre Lüge*. Berlin: Verlag Neues Vaterland.
- GUMBEL, E. J. (1922). *Vier Jahre Politischer Mord*. Berlin: Verlag der neuen Gesellschaft.
- GUMBEL, E. J. (1958). *Statistics of Extremes*. Cambridge: Columbia University Press.
- HJORT, N. L. (2026). The stochastic view used in climate sciences: (some) perspectives from (some of) mathematical statistics. In *The Hasselmann Legacy: Stochastic Thinking in Climate Science*, L. Lin & H. von Storch, eds. Berlin: Springer Verlag, pp. xx–xx.
- HJORT, N. L. & STOLTENBERG, E. A. (2026). *Statistical Inference: 600 Exercises and 100 Stories*. Cambridge: Cambridge University Press.

Footnote. After having been absorbed by the most exciting chess in ‘10,000 Lifetimes’ during the Christmas Week of 2025 I was inspired to dig out data, construct and test my models (along with a few discarded attempts, like the Pareto and the truncated normal, neither of which work well here). I then wished to write up a good *FocuStat Blog Post*, as these are sometimes reaching a sizeable international readership (we’ve got fanmail from Steven Pinker, etc.). To my frustration & irritation the University of Oslo has changed the Wondrously Well-Working Vortex `html`, which I and FocuStat comrades have used for ten years, for a Horribly Bad-Working Vortex editing system, where writing math is a nightmare; having the new Vortex system installed has also seriously disturbed some of our previously perfect pitch layouted Blog Posts. I might have to move to Andorra to find a university taking better care of its humble employees.

tankenøtter

Pusleri nr. 64

Ta en vanlig kortstokk med $n=52$ kort, herav de fire Ess. Stokk den godt og legg stokken på bordet med billedsiden ned. Ta så kort, ett etter ett, fra toppen av stokken, snu kortet og legg det ut på bordet med billedsiden opp. Fortsett til det første Ess er kommet opp.

Hva er forventet antall snudde utlagte kort?

Utfordring: Finn flere alternative løsningsmetoder.

Løsning til jostein.lillestol@nhh.no

Pusleri nr. 63 (løsning)

Oppgaven var: To kamerater A og B har en diskusjon gående om mønstre i myntkast. De enes om å se på mønstre i tre etterfølgende kast, der hver har på forhånd valgt sitt favorittmønster blant de åtte mulige mønstrene KKK, KKM, KMK, MKK, KMM, MKM, MMK, MMM.

Myntkastingen starter og pågår inntil ett av de to valgte favorittmønstrene opptrer, og A eller B utropes som vinner. B er grei og lar A ha førstevalget. A mener at de åtte mønstrene er like sannsynlige, og da spiller det ingen rolle hvilket som velges. Han velger KKK fordi det er lett å holde rede på. B tror at KMK er bedre valg ut fra forestillingen om mer balanse mellom K og M heller enn ubalanse. De eksperimenterte og observerte MMKMMKKKKMK. Her ser vi at KKK først inntraff med det åttende kastet, mens KMK først inntraff med det tiende kastet. Altså vant A her. Vi ser også at dersom B isteden hadde valgt MKK ville han ha vunnet.

- Finn vinneroddsene til A og B, og undersøk om MKK er et bedre valg for B.
- A har isteden valgt MKK. Hva bør da B velge?
- Hva er det optimale valget til A?

Løsning: Tankene bak valgene til både A og B har lite for seg. Når A har valgt KKK er MKK det beste valget for B i sannsynlighetsteoretisk forstand, der vi antar rettfærdige myntkast med uavhengighet mellom kastene. Betrakt de fire like sannsynlige resultatene av de to første kastene: KK, KM, MK og MM. Eneste mulighet for at KKK opptrer før MKK er at dette skjer allerede med de tre første kastene. I alle øvrige situasjoner kommer det en M som umuliggjør at KKK kommer før MKK. Sannsynligheten for KKK i de tre første kast er $1/8$, som gir odds lik $o(A:B)=1:7$. B har altså $7/8$ sjanse for å vinne med sitt valg MKK.

Beregning av vinnerodds for B sitt opprinnelige valg KMK og øvrige alternative valg for B er ikke like enkelt. En mulighet er å etablere ligninger og løse disse, her fire ligninger med fire ukjente, en for hver av de fire startsituasjonene for de to første kast. La p, q, r, s være sannsynlighetene for at A vinner, for gitte startsituasjon henholdsvis KK, KM, MK, MM. Den ubetingede sannsynligheten for at A vinner er da gjennomsnittet $(p+q+r+s)/4$. For de betingede sannsynlighetene har vi ved betingning mhp. utfallet av tredje kast:

$$\begin{aligned}p &= 0.5 \cdot 1 + 0.5 \cdot q \\q &= 0.5 \cdot 0 + 0.5 \cdot s \\r &= 0.5 \cdot p + 0.5 \cdot q \\s &= 0.5 \cdot r + 0.5 \cdot s\end{aligned}$$

Løsningen er $(p, q, r, s) = (\frac{3}{5}, \frac{1}{5}, \frac{2}{5}, \frac{2}{5})$ med gjennomsnitt $\frac{2}{5}$. Vinnere sannsynlighetene for A og B er dermed $(\frac{2}{5}, \frac{3}{5})$, dvs. odds $o(A:B)=2:3$.

Tilsvarende beregninger kan gjøres for de øvrige valg B kan ta etter A sitt valg av KKK. Her følger liste over de ulike vinnerodds $o(A:B)$:

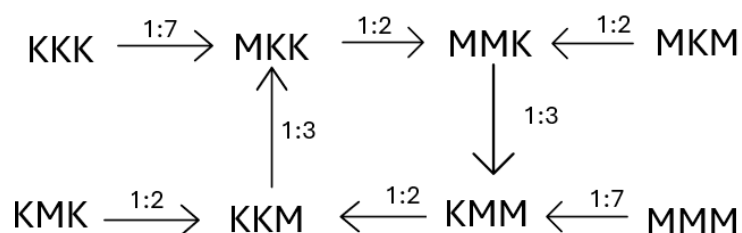
B ->	KKM	MKK	KMK	MKM	KMM	MMK	MMM
A -> KKK	1:1	1:7	2:3	5:7	2:3	3:7	1:1

Vi ser at med A sitt valg KKK kan B aldri gjøre det dårligere enn A i sannsynlighetens forstand. Videre ser vi at MKK er det overlegent beste valget for B, langt bedre enn det opprinnelige mer balanserte valget KMK. Velger B det like balanserte MKM gir det dårligere odds for seg. Hvis B vil, kan han gjøre spillet rettfærdig ved å velge MMM eller KKM. De vrangforestillinger som er tillagt A og B er hermed avdekket. Posisjon er nøkkelen! Å velge først var heller ingen fordel.

Merknad: En alternativ angrepsmåte er å beskrive situasjonen som en Markovkjede med seks tilstander: Fire transiente tilstander KK, KM, MK, MM som representerer utfallet av de to siste kastene, samt to absorberende tilstander svarende til de valgte mønstre. Da er tilgjengelig matriseformler for både sannsynligheter og forventninger uttrykt ved overgangsmatrisen.

Anta at A isteden har valgt den tilsynelatende suverene MKK i håp om at nå er det hans tur å vinne. For B kan da MMK se lovende ut. Er B så heldig at de to første kastene er MM vinner B med en etterfølgende K, og med etterfølgende M er likevel B i en gunstigere situasjon enn A. For de andre mulige utfall av de to første kastene har A tilsynelatende ikke noen tilsvarende fordel som kan oppveie dette. Beregning som ovenfor gir vinnere sannsynlighetene $(1/3, 2/3)$, dvs. odds er $o(A:B)=1:2$. MKK var altså ikke et vinnende førstevalg. Spørsmålet er om MMK var det beste valget B kunne svare med. Beregninger for alternativene viser at slik er det.

Jakten på et optimalt vinnende førstevalg har hittil vært forgjeves. Og med god grunn, det fins ikke! Paradoxsalt nok er det slik at uansett hva A velger, så kan B parere med et alternativ som gir bedre odds enn A. Dette kan skje når alternativene er satt opp mot hverandre parvis. Med 8 ulike mønstre blir det 28 parvise sammenligninger. Resultatet av disse kan ikke ordnes på noe vis fra best til dårligst. Dette er vist med en figur, der hver pil går fra et mulig valg for A (førstevalg) til det derav følgende beste valget for B (andrevalg) med tilhørende odds $o(A:B)$.



Vi ser at blant de åtte mulige førstevalg til A etterfulgt av den optimale responsen fra B, er det fire situasjonene som gir odds 1:2, inkludert den beskrevne MKK versus MMK ovenfor. Blant disse er også en situasjon der førstevalget er KMK, dvs. B sin opprinnelige favoritt, med

optimal responsen KKM. I disse fire situasjonene er altså B ventet å vinne dobbelt så ofte som A. Videre har vi to situasjoner der odds er 1:3, og B er ventet å vinne tre ganger oftere enn A, og endelig to situasjoner der oddsen er hele 1:7. Merk at pilene for de fire mønstrene MKK, MMK, KMM, KKM går i sirkel, og mønstrene kan dermed ikke ordnes i rekkefølge med omsyn til sannsynligheten for å vinne spillet. Ethvert førstevalg (A) kan pareres med et andrevalg (B) som gir bedre odds for B enn for A. På den annen side er ethvert andrevalg i sirkelen det beste for et av førstevalgene i sirkelen. Utenfor sirkelen har vi de fire mønstrene KKK, KMK, MKM og MMM. Som førstevalg pareres disse med et av mønstrene i sirkelen. Samtidig er ikke disse noe beste andrevalg for noe førstevalg overhodet.

Spørsmålet om hva som er optimalt for A å gjøre har nå på sett og vis fått en løsning: Hvis mulig, unngå å velge først! Er du nødt til å velge først, velg en av de fire mønstrene MKK, KMM, KMK, MKM, likegyldig hvilket (men forvent å tape dobbelt så ofte som du vinner).

Til slutt: Dersom A valgte B sin favoritt KMK og B svarte med MKK har vi et rettferdig spill. Men, som vi ser av figuren, B kan vinne dobbelt så ofte som A ved å velge KKM isteden.

Den beskrevne situasjon kalles ofte «Penney's Game», navngitt etter Walter Penney som lanserte den i Journal of Recreational Mathematics tilbake i 1969.

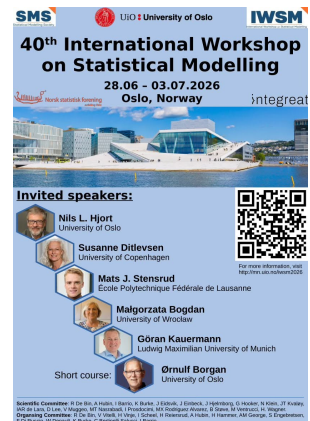
Jostein.lillestol@nhh.no

møter og konferanser

The 40th International Workshop on Statistical Modelling i Oslo

Årets International Workshop on Statistical Modelling (IWSM2026), den førtiende i rekken, arrangeres i Oslo i perioden 29. juni - 2. juli 2026. Konferansen fokuserer på statistiske problemstillinger motivert av virkelige data og løsninger som gir nye bidrag til fagfeltet. Programmet består av omtrent 60 muntlige presentasjoner, komplementert med postervisninger.

For program og mer informasjon, se <https://mn.uio.no/iwsm2026/>.



Nytt fra Oslo senter for biostatistikk og epidemiologi (OCBE)

Nyansatte

- *Forsker:* Brook N. Wolford, Owen M. T. Thomas
- *Postdoc:* Wei H. Deng

Ny Phd-grad

- Wei Hai Deng. *Targeting fatigue, mental health, and quality of life through latent profiles and physical activity interventions in childhood and young adult cancer survivor*. Bedømmelseskomité: Professor Maxime Caru, Penn State Cancer Institute, US; Professor Örjan Ekblom, Karolinska Institutet, Sweden; Associate professor Lise Mørkved Helsing, Institutt for helse og samfunn, UiO. Hovedveileder: Seniorforsker Corina S. Rueegg, OCBE, Oslo Universitetssykehus; Medveiledere: Førsteamanuensis Hanne C. Lie, Avdeling for atferdsmedisin, UiO; Overlege Cecilie E. Kiserud, Nasjonalt kvalitets- og kompetansenettverk for seneffekter etter kreftbehandling, Oslo Universitetssykehus.

Nytt fra NTNU

Nye Mastergrader

- Oscar Holmen. Bayesian Inference for Superposed Renewal Processes. Veileder: Jarle Tufto.
- Maiken Astri Canlas Hatten. Modelling the Phototactic Movement of Salmon Louse Copepodids. Veileder: Thea Bjørnland.
- Morten Johan Rønning. Forward Curve Modeling Using Fractal Ornstein-Uhlenbeck Processes and the Heath-Jarrow- Morton Framework On Evaluating Risk of Swap Contracts under the Physical Measure. Veileder: Jacob Kooter Laading

Nye PhD-grader

- Øyvind Stormark Auestad. Numerical approximation of parabolic stochastic evolution equations revisited. Veiledere: Geir-Arne Fuglstad, Espen Robstad Jakobsen, og Annika Lang (Chalmers).