Piet Groeneboom

Delft Institute of Applied Mathematics
TU Delft
p.groeneboom@tudelft.nl

Column   Piet takes his chance

Chernoff's law and Jante's law

Piet Groeneboom regularly writes a column on statistical topics in this magazine.

Some history

In my column "Chernoff's distribution and the bootstrap" [8] I complained about the fact that in the Wikipedia article [14] Chernoff's original paper, where 'Chernoff's distribution' (which could also be called 'Chernoff's law') was discussed, was not mentioned, whereas three papers written by me were mentioned.

I do not know whether this complaint reached the Wikipedia authors, but the Wikipedia item on Chernoff's distribution has been drastically changed since I wrote that column. The section on 'History' now starts with the line: "Groeneboom, Lalley and Temme [4] state that the first investigation of this distribution was probably by Chernoff in 1964 [5] who studied the behavior of a certain estimator of a mode."

The Wikipedia article does not summarize the history completely accurately from our paper, but since the law of Jante [16] prohibits me from writing in the Wikipedia article, I thought that it might be a good idea to write a column about the history while I am still alive. Henry Daniels [15] and Tony Skyrme [17], who both contributed to this history, are no longer alive.

However, I cannot tell the full story on the characterization of Chernoff's distribution in this column without breaking Jante's law several times. I will indicate that I am breaking the law by BJL (Breaking Jante's Law).

My own dealings with Chernoff's distribution started in the academic year 1982/1983, when I was invited to give a lecture in the so-called Neyman–Kiefer conference in Berkeley. I spent January till July 1983 at the Mathematical Research Sciences Institute (MSRI) in Berkeley, sharing the office with Steve Lalley and Thomas Sellke. It was not always clear that Thomas was present in our office, but he could suddenly become visible rising from below his desk, where he was taking naps.

The invitation to speak at this conference had been somewhat strange: the famous statistician from Berkeley who invited me had told me: "Some people seem to think that you should be invited to speak at the Neyman–Kiefer conference." Perhaps he was applying Jante's first commandment: "You're not to think you are anything special." Or Jante's second commandment: "You're not to think you are as good as we are." The people who witnessed this invitation were appalled. But anyway, I accepted the kind invitation and talked about my investigations with respect to the (almost surely unique) location of the maximum of two-sided

Brownian motion minus a parabola. So I started with a result from Chernoff's paper:

Theorem 1 (Chernoff [1]). *Let W be two-sided 1-dimensional Brownian motion, originating from zero. The density f_Z of $Z = \operatorname{argmax}\{W(t) - t^2\}$ is given by*:

$$f_Z(s) = \frac{1}{2}\partial_2 u(-s,s^2)\,\partial_2 u(s,s^2),$$

where $u(s,x)$ solves the (heat) equation:

$$\frac{\partial}{\partial s}u(s,x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}u(s,x).$$

subject to:

$$u(s,x) = 1,\ x \geq s^2, \qquad u(s,x) \to 0,\ x \to -\infty.$$

and $\partial_2 u(t,x)$ denotes the first derivative with respect to the second argument x.

The original computations of this density were based on numerically solving Chernoff's heat equation. This was done by Chernoff himself and also Willem van Zwet, both with the help of people from numerical mathematics, and by myself in 1982 at the Mathematical Centre, Amsterdam (now CWI), again with the help of numerical mathematics people. The latter mathematicians (in particular, Ben Sommeijer) noticed the instability of the solutions in the region where the time argument s is negative, if the then rather fashionable 'multigrid method' for the solution of partial differential equations was used. This phenomenon was explained in [4]:

$$\partial_2 u(-s,s^2) \sim c_1 \exp\left\{-\frac{2}{3}s^3 - cs\right\},\ \ s \to \infty,$$

where $c \approx 2.9458$ and $c_1 \approx 2.2638$. This fast decay entails that a numerical solution of this partial differential equation on a grid will not give a really accurate solution.

However, an analytic characterization that can be used for accurate numerical computations is given in the following theorem.

Theorem 2 [2,4]. *The probability density f of the location of the maximum of the process $t \mapsto W(t) - t^2, t \in \mathbb{R}$, is given by*

$$f(s) = \frac{1}{2}g(s)g(-s),$$

where

$$g(s) = \frac{1}{2^{2/3}\pi}\int_{-\infty}^{\infty}\frac{e^{-ius}}{\mathrm{Ai}(i2^{-1/3}u)}du,$$

where Ai is the Airy function.

The density f_Z of Z can now be computed efficiently by two lines in Mathematica, see Figure 1.

Theorem 2 is also given in [11] and [9]. The proof in the latter paper seems at present the easiest way to obtain the result. The distribution of the maximum itself was studied in [5,6,12].

The Airy functions enter (in my approach to the problem) via the Cameron–Martin–Girsanov formula and the Feynman–Kac. The proof of Chernoff's result [1] and an exposition of how the Airy

```
In[1]:= f[x_] := (1 / (2 * Pi)) * 2^(1 / 3) *
         Re[NIntegrate[Exp[-I * u * x] / AiryAi[I * 2^(-1 / 3) * u], {u, -10, 10}]]

In[2]:= g[x_] := (1 / 2) * f[x] * f[-x]

In[3]:= Plot[g[x], {x, -2, 2}]
```
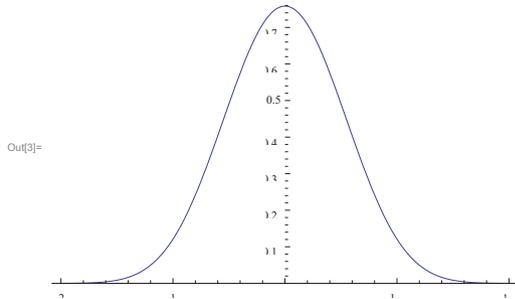

Figure 1   Density of $Z = \operatorname{argmax}\{W(t) - t^2\}$.

functions enter in [9] is given in [7] (my corresponding 2018 lecture in Banff, Canada, is also still on the internet, BJL).

The Airy functions appear in a lot of different contexts. Percy Deift says about this in his lectures on Riemann–Hilbert problems [3]: "Special functions are important because they provide *explicitly solvable models* for a vast array of phenomena in mathematics an physics. By 'special functions' I mean Bessel functions, Airy function, Legendre functions, and so on. If you have not met up with these functions, be assured, sooner or later, you surely will."

Now, on the day of my lecture at the Neyman–Kiefer conference in 1983 and after I had delivered my lecture, a well-known Dutch statistician and a Berkeley statistician suggested that we should walk to a coffeeshop nearby. On the way to the coffeeshop the Dutch statistician kept saying "this is unimportant". I wondered what this meant and why he was saying it (Jante's first commandment again?). The Berkeley statistician, on the other hand, kept saying: "This limit result of Chernoff for the mode cannot be right." Why couldn't it be right? Because Hasminskii had shown that the minimax rate for estimating the mode was $n^{1/5}$, so how could an estimate of the mode converge at rate $n^{1/3}$ (this was Chernoff's limit result in the paper)?

The answer is of course rather simple. The minimax calculation corresponds to a rather pessimistic view of the world, where the convergence has to be uniform over whole neighborhoods of distribution functions. If one allows distribution functions that are sufficiently unpleasant in such neighborhoods one cannot get a faster rate than $n^{1/5}$. But Chernoff's (correct) result is for a fixed underlying distribution which is sufficiently well-behaving.

Further history
After returning from MSRI to CWI I wrote down my derivation leading (among other things) to Theorem 2 above in the CWI report [4]. The British statistician Henry Daniels, who around the same time also had derived an analytical expression with Airy functions for Chernoff's distribution, together with the physicist Tony Skyrme, got hold of my report and handed this to David Kendall. David Kendall was at the same Institute of Pure Mathematics and Mathematical Statistics at Mill Lane in Cambridge.

There was a lot of contact between Nico Temme and myself on the analytical aspects (we were both at CWI). Nico also wrote a paper on an analytical aspect of the problem [13]. So Nico and I certainly did not work independently on this problem, as stated now in the present Wikipedia article.

In 1984 I moved to the Mathematics Institute of the University of Amsterdam, where I received in 1985 a letter and a cheque from Cambridge, telling me that the Rollo Davidson Prize had been awarded to me (BJL, but I must say it because it is relevant for the further history). I later heard that Terry Lyons was the other winner that year.

I did not know who Rollo Davidson was, so I was not aware of the existence of this prize. The accompanying letter was written by Peter Whittle, but I later understood that David Kendall was the main person who had decided on this award. In fact, I received a letter from David Kendall, saying: "You cannot imagine how happy I felt when I read your report" (BJL again, I fear). I wondered when I read this letter whether I had felt like that myself in the past and realized that I had indeed felt like that sometimes, but not very often. Remarkably, I had felt like that reading a paper of my co-winner Terry Lyons.

But to return to what happened after I was awarded with the Rollo Davidson Prize: I sent in my CWI report to the journal *Probability Theory and Related Fields*. I now know who the two referees were. I first received a referee report of Steve Lalley (my office mate at MSRI and co-author in [9]). He also separately wrote me a letter, which basically suggested that I should introduce a stopping time argument in the Cameron–Martin–Girsanov part of the proof. That seemed a good idea, but I didn't do it. I kept his letter, though (something I usually don't do). In fact I did nothing, I put the report in a drawer and forgot about it.

A year later, I received another referee report. This was rather remarkable, because I already got the news that my paper was accepted with the first referee report a year earlier. This report was written by Rudi Lerche and he complained that the paper started off rather nicely, but then there was this awful appendix full of technical stuff on complex analysis computations, wasn't there some way to avoid that? I fully agreed with him, but didn't know how to do it.

Nothing happened for a while, in fact for a few years. Now and then I received letters from people asking: "What the hell is happening to this paper of yours that is now announced for years to appear in *Probability Theory and Related Fields*?" I had no answer, except that I received this prize for it and that I did not work on it (I got interested in other things in the mean time and felt no big urge to work on it because I had already been awarded with this prize for it).

And then finally, in 1987 or 1988 I received a letter from Mrs. Zassenhaus in Paris, who was handling the publishing of *Probability Theory and Related Fields*: "Mr. Groeneboom, are you still interested in your own paper? If so, send it in and we will publish it immediately." This seemed an easy way out for me and that is what I did. I sent in the original version without any change, and so it was finally published in 1989.

But after all these events I kept this nagging feeling that I should do something about the awful appendix in my original paper. And that I should use the suggestion of Steve Lalley to introduce the stopping time argument. So around 2013 I wrote to my old friends Steve Lalley and Nico Temme that I had some new ideas on how to tackle the problem without using my appendix in the original paper. Since this is a column, I will not go into the details here (which are explained in [9]), but the basic idea is that two expressions, involving integrals over the Airy func-

tions Ai and Bi have to be equal (something I could not prove in the eighties) and that they have to be equal because they satisfy the same partial differential equation. And then finally we need the maximum principle to pin everything down via the boundary conditions.

So we wrote a paper on this and I submitted it to the *Annals of Probability*. This seemed to me a safe place to send it, because Steve had been main editor of this journal and one would think that this would give him some credit. But that is not how things work nowadays. We waited half a year (not an unusual waiting time) and then I reminded the editor of our paper. Again some waiting, after which I received a letter from him: "Yes, we had some telephone communications about your paper and the upshot is that we were not very enthusiastic." No Associate Editor's report or referee reports. Often the Associate Editor (AE) and referees don't have a clue what the paper is about and then start saying this kind of thing. Or: "You proved this before, didn't you? So why a new proof?" Students can also say: "You already have a proof, why do you want another proof?" I do not know how many proofs Euclid's prime number theorem has in the last edition of *Proofs from THE BOOK*, but in my (4th) edition it has six.

Then I sent it to the *Journal of Mathematical Analysis and Applications* (Nico's suggestion). Again a considerable waiting time, but then I received a letter from the editor that it had been accepted (without any referee reports). Nico remarked: "Well, at least we do not have to thank the referees." The vagaries of publishing...

Marcel Proust submitted his work 'Á la recherche du temps perdu' to the *Nouvelle Revue Française* (I talked about this in my previous column in connection with the 'sonate de Vinteuil'). This was turned down by this publisher, for which André Gide was responsible. Marcel Proust then published it on his own account. André Gide apologized a year later (usually this is not what the AE and referees do, but of course they are anonymous). It is a matter of keeping faith if one really believes in one's work and of having the means to push it through.

There is a difference in style in the paper of Henry Daniels and Tony Skyrme on one hand and my paper and the paper of Steve Lalley, Nico Temme and myself on the other hand. For example, Daniels and Skyrme say: "It should be possible to establish analytically that

$$\frac{1}{2\pi i} \int\limits_{-i\infty}^{i\infty} \frac{dw}{\mathrm{Ai}^2(w)} = 1.\text{"}\tag{1}$$

(Part of the second to last formula of the paper.) This is British style, on the continent we are not allowed to say this. We say in [9, Appendix D]: "We have not been able to find a direct reference for the nice relation (1)" and then prove it (the proof is not very long, though).

A conference in Tashkent and Henry Daniels

Around this time, there was the first World Congress of the Bernoulli Society in Tashkent (1986). I was asked to visit the refuseniks in Moscow and was planning to do that on my way back from Tashkent. The refuseniks were (mostly) Jews who had asked permission to emigrate to another country (e.g., Israel), a request that had been turned down, usually under the pretext that this was because of security reasons (if the wife of a male refusenik

had been working at an official bureau, she had had access to 'classified information'). Even the parents of the refuseniks could lose their jobs because they "hadn't raised their children properly".

Certain British statisticians were planning to visit the refuseniks on their way to Tashkent. They made more noise about their plans than I did. There was absolutely nothing illegal in visiting the refuseniks, but the way the refuseniks were treated was not exactly propaganda for the Soviet Union. So there was the threat from Moscow that the whole conference in Tashkent would be canceled if the refuseniks would be visited. The British statisticians subsequently canceled their plans. But I did not cancel my plans (taking a lot of precautionary measures that had been advised to me by the organization that asked me to make the visit). In contrast with the way the refuseniks were treated, the Tashkent conference was a big propaganda event. I thought that canceling the conference was an empty treat.

In the plane for invited speakers and high officials from Moscow to Tashkent I was asked: "Piet, are you still planning to visit the refuseniks in Moscow on your way back to Holland?" After I answered "Yes, I am", I was told: "Be aware of the fact that this will have very serious consequences for your position in the Bernoulli Society!" At the time I was curious what this meant: would I be thrown out of the Bernoulli Society? Years later, during a night in Oberwolfach it became clear what had been meant. I 'was up' for an appointment in the council of the Bernoulli Society and the person who asked me this question would prevent (and indeed had prevented) this to happen. I was completely unaware of all this machinery behind the scenes.

These differences of opinion on the visit to the refuseniks could conceivably be between Henry Daniels and myself. Also, sometimes one gets the impression that mathematicians are involved in a race who will first get a result and this could be a further difficulty. But we actually became friends (I believe).

Henry Daniels played the piano but also a British invention, the concertina (credited to Sir Charles Wheatstone). The concertina is a kind of 'accordion'. In Oberwolfach he could fill in the missing violin and (if I remember correctly) the viola part in string quartets, playing on his concertina. He could fake a vibrato by moving the concertina in a special way. Henry and I even played Bach's violin-oboe concerto (don't remember which pianist played the orchestral part) in Oberwolfach, where Henry played the oboe part on his concertina and I played the violin part. In [10] I described the music event which took place at my house on a very rainy day during the ISI (International Statistical Institute) meeting in Amsterdam, 1985.

Later Henry invited me to come to Cambridge, where I stayed at King's College and attended the high table dinner with him. This was my first high table dinner (later I attended a similar event in Oxford; there were some differences, for example still more drinking in Oxford, but it might depend on the college) and it was exactly as I had expected. Men hanging in club chairs in a kind of antechamber with a glass of sherry before the dinner started and then this dinner at a table separate from the students' tables.

Henry was not so keen on these events and said something to me of the sort: "Glad you are with me, I always feel a bit out of place!" I know this feeling myself also all too well, so was glad to be of some assistance to him here. During the day we played music at his house. He claimed that one's rank in the hierarchy in King's college was measured by the amount of keys one had to different doors. Don't know whether this is really true.

He was a very kind man. I still remember that in the subway (underground) in Tashkent young men would offer him their seat (he was in his eighties then) and I was wondering whether a similar event would be possible in the Amsterdam subway. My guess was that this would not happen there.   ⌐····

References

1 H. Chernoff, Estimation of the mode, *Ann. Inst. Statist. Math.* 16 (1964), 31–41.

2 H. E. Daniels and T. H. R. Skyrme, The maximum of a random walk whose mean path has a maximum, *Adv. in Appl. Probab.* 17 (1985), 85–99.

3 Percy Deift, Riemann–Hilbert problems, 2019, arXiv:1903.08304.

4 P. Groeneboom, *Brownian Motion with a Parabolic Drift and Airy Functions*, CWI Technical Report Department of Mathematical Statistics-R 8413, CWI, 1984.

5 P. Groeneboom, The maximum of Brownian motion minus a parabola, *Electron. J. Probab.* 15, no. 62 (2010), 1930–1937.

6 P. Groeneboom and N. M. Temme, The tail of the maximum of Brownian motion minus a parabola, *Electron. Commun. Probab.* 16 (2011), 458–466.

7 Piet Groeneboom, Chernoff's distribution and differential equations of parabolic and Airy type, Talk at meeting *Shape-Constrained Methods: Inference, Applications, and Practice*, Banff, 28 January to 2 February 2018.

8 Piet Groeneboom, Chernoff's distribution and the bootstrap, *Nieuw Archief voor Wiskunde* 5/22 (2021), 241–244.

9 Piet Groeneboom, Steven Lalley and Nico Temme, Chernoff's distribution and differential equations of parabolic and Airy type, *J. Math. Anal. Appl.* 423 (2015), 1804–1824.

10 Piet Groeneboom, Jan van Mill and Aad van der Vaart, In Memoriam Kobus Oosterhoff: Statistics as both a purely mathematical activity and an applied science, *Nieuw Archief voor Wiskunde* 5/18 (2017), 55–59.

11 S. Janson, Moments of the location of the maximum of Brownian motion with parabolic drift, *Electron. Commun. Probab.* 18, no. 15 (2013), 1–8.

12 S. Janson, G. Louchard and A. Martin-Löf, The maximum of Brownian motion with parabolic drift, *Electron. J. Probab.* 15, no. 61 (2010), 1893–1929.

13 N. M. Temme, A convolution integral equation solved by Laplace transformations, in *Proceedings of the International Conference on Computational and Applied Mathematics* (Leuven, 1984), Vol. 12/13, 1985, pp. 609–613.

14 Wikipedia, Chernoff's distribution, https://en.wikipedia.org/wiki/Chernoff%27s_distribution, 2023.

15 Wikipedia, Henry Daniels, https://en.wikipedia.org/wiki/Henry_Daniels, 2023.

16 Wikipedia, The Law of Jante, https://en.wikipedia.org/wiki/Law_of_Jante, 2023.

17 Wikipedia, Tony Skyrme, https://en.wikipedia.org/wiki/Tony_Skyrme, 2023.