

A conversation with Nils Lid Hjort

Ørnulf Borgan  | Ingrid K. Glad

Department of Mathematics, University of Oslo, Oslo, Norway

Correspondence

Ørnulf Borgan, Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, NO-0316 Oslo, Norway
Email: borgan@math.uio.no

Funding information

Norges Forskningsråd, Grant/Award Numbers: 237718, 332645

Abstract

Professor (now emeritus) Nils Lid Hjort has through more than four decades been one of the most original and productive statisticians in Norway, contributing to a wide range of topics such as survival analysis, Bayesian nonparametrics, empirical likelihood, density estimation, focused inference, model selection, and confidence distributions. This conversation, which took place at the University of Oslo in December 2023, sheds light on how Nils Hjort's curious and open mind, coupled with a deep understanding, has enabled him to seamlessly navigate between different fields of statistics and its applications. Our aim is to encourage the statistics community to always be on the lookout for unexpected connections in statistical science and to embrace unexpected encounters with fellow statisticians from around the world.

KEYWORDS

history of statistics, Norway

Ingrid: Nils, when we look at your published papers, you have had over 50 co-authors from 10 different countries. What has this broad international contact meant to you?

Nils: I think this has been a very enjoyable dimension of my work, which has suited me very well, even though I wasn't conscious of it in the beginning, and without it being planned. It is a consequence of both how I think and behave and how our field operates. We are exposed to things, we attend conferences, workshops, we meet people, we realize that this or that person is working on something interesting. But there hasn't been a specific algorithm on my part, other than being generally open and interested, in topics and in people.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

Ørnulf: But these people, have you met them at meetings? Or have you seen that they have written something interesting, and then contacted them? Or have they contacted you?

Nils: A handful of people have discovered me, or some common research threads, and come to Oslo for short or slightly longer visits, or invited me to a research week where they are. I would also like to mention Highly Structured Stochastic Systems (HSSS), a fairly large European Science Foundation-supported project from 1993 to 2001, where I was on the steering committee. [HSSS is summarized in Green et al. (2003).] In this project there were active actions and initiatives to help establish academic networks, especially for the younger ones, with mobility, contacts, collaboration. This also strengthened my own networks.

Ingrid: Because it was not obvious that you would become this international? You were rather shy when you were younger?

Nils: Yes, I was famously shy as a child, and remained on the clearly shy side the first years at university. But things have worked out, so now I enjoy talking to people, and I am very approachable.

Ingrid: We will later return to the stories about your transformation from being a “hidden gem” in Oslo to becoming a highly international researcher. Can you start with saying a bit about your family background and early years?

Nils: I am the second of five Hjort brothers, and grew up in Oslo, a biking distance from the university. My mother was an illustrator and housewife and my father was a supreme court lawyer. There was a lot of talk about law and society at home. Two of my brothers became lawyers, while I became a scientist.

Ørnulf: When you were in school, what subjects interested you the most, and what kind of pupil were you?

Nils: I guess I was in the category of “naturally good student,” so school was not much of a challenge for me. I found mathematics easy throughout my school years, and in my last year of high school, I discovered that I also had a talent for more abstract mathematical thinking.

Ørnulf: After high school, you started studying mathematics at the University of Oslo. Were there other subjects you were considering?

Nils: Yes, I was thinking about music, which for me means classical music, as I have an enduring interest in it. I still play the piano a lot, and at one point, I played as many as four instruments, in addition to singing in a choir.

Ingrid: So, you could have become a musician?

Nils: No, even in my optimistic moments, I did not think I could become a musician, a soloist. But I considered studying music history and musicology.

Ørnulf: But you chose mathematics.

Nils: Yes, I chose mathematics, and following that, statistics, which at the time was synonymous with mathematical statistics. I completed my bachelor's degree in 1974 and then started my master's studies.

Ingrid: For your master's thesis, you wrote about the Dirichlet process applied to some nonparametric problems. What led you to choose this topic?

Nils: There are two reasons for that. During my master's studies, I took a course in decision theory with Erik Torgersen. For the exam, I was given the task of reading and presenting the quite fresh Ferguson paper on the Dirichlet process (Ferguson, 1973), which has become a true classic in Bayesian nonparametrics. So I became familiar with these themes. For the calendar year 1976, I went to Tromsø to work as a temporary assistant professor. There, I shared an office with Tore Schweder, who was in the process of developing the field of statistics at the newly established university there. We had a very good time, and I learned a lot from him. Without anything being said or formalized, I started writing about Bayesian nonparametrics around the Dirichlet process. So, that became my master's thesis. But in reality, I had no guidance or feedback along the way, except for Tore asking interestedly from time to time how things were going.

Ørnulf: After you finished your master's degree, you were a research assistant for six years at the Department of Mathematics here in Oslo. Can you tell us a bit about what you were doing during those years? What were your tasks?

Nils: The tasks were primarily teaching and supervision. So I lectured both undergraduate and graduate courses, and I was the supervisor for three to four master's students during those years. It worked out fine, and I enjoyed it. I also learned a lot by attending seminars, with an active mindset.

Ingrid: Did you read a lot?

Nils: Yes, I read a lot and picked up various topics and themes. I obtained a good overview and was well-informed. I also wrote a new compendium for a course on limit theorems in probability theory that was on the curriculum list for twenty years. I felt that I managed it well without too much effort. It contributed to my realization that I could do this; I can write orderly and organized, I can come up with good proofs and examples that are not in other books, along with some tentative research ideas of my own.

Ingrid: Discovering that you were good at this, was it a step towards the realization that maybe you should become a researcher?

Nils: Yes. But in hindsight, it has surprised me that there were no friendly nudges from the seniors for me to also write and publish papers. So, it was only at the end of my research assistant period that I came to the realization that I am good at this, I like it, but I haven't published any papers, contrary to what was really expected of me.

Ørnulf: When the research assistant period ended, you started working at the Norwegian Computing Center (NR).

Nils: Yes, at the end of my six years as a research assistant, I suddenly got a lot of job offers, and one of these was from NR. During the job interview with them, I got the feeling that they had made up their minds before I had even opened my mouth. So, I got a job in the group for statistical analysis of natural resource data, led by Henning Omre, who later became a professor at the Norwegian University of Science and Technology in Trondheim.

Ingrid: Can you tell us a bit about what kind of work you did at NR?

Nils: There was a lot of geostatistics, because all the oil companies needed it. So, we had to learn everything about spatial statistics and covariance functions in dimensions one, two, and three. There was also a lot of classification analysis, from simple discriminant analysis to more advanced techniques. One of the tasks was symbol recognition. Among other challenges, we needed to correctly recognize hand-drawn symbols on thousands of existing maps.

Ørnulf: Did you also work with satellite measurements?

Nils: Yes, there were several projects about remote sensing. It involved both classification and prediction. The thing was that you couldn't just use what was in the textbooks, because we needed methodology for what's called contextual classification. If you know that your three neighbours are in a forest, it increases the chance that you too are in a forest. Building models for this was a big topic, which we needed not only in remote sensing, but also for analysing logs from oil wells.

Ingrid: And you were at NR for about five years?

Nils: Yes, more or less. From the autumn of 1983 to the end of 1988.

Ingrid: Can you tell us what you took with you from your time at NR into your work at the university?

Nils: It might not be very fruitful to think counterfactually, what would have happened to me and my academic career if I had been at the Department of Mathematics all the time? Maybe the answer is "more or less the same." But probably not, because I learned things at NR, both subject matter and the culture of "things have to be done," in a different way than at the university. It definitely broadened my view of what it meant to work with statistics. It wasn't just about inventing more theorems, which in a sense is what I had learned before my time at NR. But it was a nice symbiosis between theory and applications.

Ingrid: But you also published papers?

Nils: Yes, NR was actually eager for us to publish, and preferably as part of our larger projects. Together with Erik Mohn, I had an invited lecture with paper at



FIGURE 1 Nils presents Hjort and Mohn (1987) with overhead projector and handwritten slides at the ISI conference in Tokyo 1987.

the ISI conference in Tokyo 1987, with many topics related to remote sensing (Hjort & Mohn, 1987); see Figure 1. A few years later, Henning and I were honoured to have the so-called SJS paper at the Nordic Statistical Meeting in Odense 1992, with discussants (Hjort & Omre, 1994). Furthermore, I had generous freedom to research and publish outside of the pure NR projects. One opportunity I eagerly seized was when I was invited as a discussant for Per Kragh Andersen's and Ørnulf's lectures on counting process models and life history data at the Nordic Statistical Meeting at Bolkesjø in 1984. It was a perfect opportunity to research different topics, present them, and relate them to the material Per and Ørnulf presented in Andersen and Borgan (1985). The topics of my discussion contribution (Hjort, 1985) eventually became two big papers in *Annals of Statistics* and one in the *Scandinavian Journal of Statistics* (SJS).

Ingrid: So you had some good opportunities to connect with academia in the midst of your time at NR.

Nils: Absolutely. It should also be mentioned that I had a very fruitful year at Stanford, January–December 1985, with a push from Henning. He said, here you go, I've arranged contact with Paul Switzer for you. Have a great year, feel free to do whatever you want, it would be nice if some of it is NR-related, but there's no obligation. During my enjoyable year at Stanford, I talked to a lot of people, and learned a lot about Bayesian nonparametrics, density estimation, and bootstrapping. But I also continued with NR-related topics, like contextual classification.

Ingrid: In the past forty years, you have been one of Norway's most creative and productive researchers in statistical methodology. You have been, and continue to be, a great inspiration for many young researchers. But who has inspired you?

Nils: The matter of inspiration and role models revolves around somewhat unclear parameters, but I will try to answer. First, I must mention you, Ørnulf, and Bo Lindqvist; you were two clever master's students and research assistants a couple of years ahead of me at the university. I would also like to point to Odd Aalen, a couple of years ahead of Ørnulf and Bo. In addition to obvious cleverness and depth, he had the professional gaze turned on, in a different way than others. Jan Hoem, Odd's master's supervisor, should also be mentioned. He had these visible characteristics of being a positive, skilled person, not only with the methodological and theoretical tools, but he would apply them to real demography. I must also mention that at the beginning of my NR period, I was sent to a nice course in Northern Sweden. They had such courses once a year with a new topic and senior person each year. This time, in 1984, it was spatial statistics with Brian Ripley. I thought he was very impressive. I learned many things from Brian and followed him closely in the following years. He also instilled in me the operative thought that I needed to write and publish.

Ingrid: Are there any other international researchers that have been important for you?

Nils: Yes, I had the already mentioned, very pleasant and fruitful year at Stanford, and got to know many of the brilliant researchers there. From Efron and Diaconis to Switzer, Friedman and T.W. Anderson, as well as various interesting guests. When you've been exposed to these researchers you become aware that they're not just talented, but talented in original, creative, and energetic ways. I also became acquainted with people who were younger than these famous seniors, such as Iain Johnstone, Trevor Hastie, Rob Tibshirani, and Art Owen. So yes, we have to include them. Later on I also became aware of several strong Danish professors, Søren Johansen, Steffen Lauritzen, Ole Barndorff-Nielsen, and Niels Keiding, all of whom had this additional layer of natural academic authority.

Ørnulf: In 1989, you were employed as an associate professor in statistics at the Department of Mathematics, University of Oslo, and two years later, you were promoted to professor. You had this position until you became professor emeritus in 2023. In your research, you have been interested in a variety of different topics. We can mention survival analysis, density estimation, Bayesian methods, model selection, confidence distributions, empirical processes, applications in nonstandard areas. You have a very broad scope. What has inspired you to address these different topics?

Nils: I'm not sure if I can answer this very coherently or in a logical manner. I'm not one of those meticulous planners who set a plan and then follow it. I allow myself to be a little more random, open to being exposed to extra things. I've always had a little door where I allow myself to wander off on digressions if something crosses my mind. So, the fact that I've touched on these various topics over the years hasn't been a result of a clearly laid plan.

Ørnulf: When you were doing your master's thesis, you were working with the Dirichlet process. But otherwise, you have quite a few papers on survival analysis early on.

Nils: This has to be linked to Odd Aalen, who was one of the main figures in this period, eventually along with a group of talented researchers in the Nordic countries, with Hoem and you, Ørnulf, and various strong Danes around Niels Keiding. Odd wrote a legendary PhD thesis in Berkeley, 1975. Here he writes about counting processes and martingales and their use in survival and event history analysis, with full control over estimators and inferences, at the right time and with the right emphasis. After Berkeley, he spent some time in Copenhagen before he came back to Norway. Was he employed here?

Ørnulf: I think he went straight to Tromsø. But after Tromsø, he was at the Department of Mathematics here in Oslo for about half a year before he was employed as a biostatistician at the Faculty of Medicine.

Nils: Yes, and during that half-year, he gave a seminar series on counting processes and martingales. It was fun, interesting, broad, and new, and I liked understanding all the technical details. This led me to develop an interest in Aalen theory, to use that term, and how it can be applied in various contexts, including parametric models for survival data. That wasn't something Odd was focused on at that time. Partly inspired by you, Ørnulf, I wrote a couple of papers about parametric survival models. It was also a good opening to do a lot of other things. So that must definitely be on the map here. So, the question of what led to my various "research blocks," we have now identified one, which is survival analysis via Odd, and the branching processes that followed (Figure 2).

Ørnulf: Survival analysis seems to have been a theme that has been running in the background for you all the time. Some aspects of survival analysis have emerged in your research from time to time up to the present day.

Nils: Yes, and the horizon becomes a bit broader if one adds "event history analysis." So the field is significantly broader than just Nelson–Aalen, Kaplan–Meier, and Cox regression. Another block, block number 2, is Bayesian nonparametrics (BNP). I was lucky with the dice, in that I got Ferguson's paper from Torgersen at the right time.

Ørnulf: Then we have to mention one of your most cited papers, where the Beta process is introduced (Hjort, 1990). What was the inspiration for this work?

Nils: Here it is possible to be quite specific. The inspiration can be concretized to a quarter of an hour at a blackboard with Odd Aalen.

Ørnulf: You must give us some details on that!

Nils: Odd said something like: Here is the theory for Kaplan–Meier. What it is based on is that the change $dN(t)$ in the process that counts the number of observed



FIGURE 2 Nils, Sir David Cox, and Odd Aalen meet in Oslo, 2006, in connection with the project “Statistical Analysis of Complex Event History Data” at the Centre for Advanced Study at the Norwegian Academy of Science and Letters.

events is almost binomially distributed, given the past. And if one thinks that $dN(t)$ is binomial, then one quickly gets both Nelson–Aalen and Kaplan–Meier. If one then looks at the binomial formulation, it is not far to try to be Bayesian, at each new time point t . Because the Beta-binomial model is discussed in every Bayesian textbook. It is a long way to go to take this over to the world of stochastic processes. Maybe Odd could have done this if he had wanted to. But he nodded to me and said “look at it a bit.” It became both challenging and demanding, but fun and finally very nice. I presented the essence in my discussion contribution at the Bolkesjø meeting, and then sent the manuscript to *Annals of Statistics*. After a year, *Annals* said “this is great, we want it.” Then it took two more years, because I didn’t have the energy to revise it. This was before the practical convenience of having the glorious TEX on a machine at one’s own work desk.

Ørnulf: This was when one first wrote a manuscript by hand and then typed it on a typewriter?

Nils: Yes, and I thought “oh, do I have to revise this, it takes so long to write everything again.” But I remember that Willem van Zwet, who was the editor of *Annals* at that time, wrote to me and said, hello, it has been two years, I am clearing my desk, you must finish this work. Oops!, I thought, and then I got it done, in long evenings with the electronic typewriter.

Ørnulf: So it took a long time to get this published.

Nils: Yes, a very long time, but it was mostly my own fault. And I absolutely thank van Zwet for his gentle pushing, because he must have thought that this must be published. I am very happy with the Beta process, which for one thing took off in survival analysis, where it provides lovely Bayesian extensions of Nelson–Aalen and Kaplan–Meier, as well as for Cox regression, and more. Secondly, ten years later, the machine learners discovered the Beta process; it has different uses, including as a kernel model in what is called the Indian Buffet Process. The work on the Beta process was my most cited paper on Google Scholar for many years, I know.

Ørnulf: Can you also tell us a bit about other work you have within BNP?

Nils: I have definitely been persistently fond of BNP, where there is good elbow room in the infinite-dimensional spaces. In addition to the Beta process, I have worked on various generalizations of the Dirichlet process, for various purposes, such as semiparametric Bayesian modelling and density estimation. And with Stephen Walker, I have a paper on quantile pyramids (Hjort & Walker, 2009). Since I was among those involved from the beginning, in more and more and larger and larger BNP conferences, four of us got the opportunity to organize a five-week program on BNP at the Isaac Newton Institute, Cambridge, in the summer of 2008. This also led to the book “Bayesian Nonparametrics” (Hjort et al., 2010).

Ørnulf: You mentioned that the work on the Beta process was your most cited paper on Google Scholar for a long time. But what is your most cited work now?

Nils: My most cited work is the book I wrote with Gerda Claeskens on model selection, which we will discuss later. But among the journal papers, my current hit is the one called BHHJ, for Basu, Harris, Hjort and Jones (Basu et al., 1998). BHHJ has become the canonical one-parameter fine-tuning extension of maximum likelihood. If you don’t want to use maximum likelihood because you’re not sure if the model fits well, or if you’re worried about outliers, just add a parameter a and run BHHJ. If a is close to zero, it’s close to maximum likelihood; you’ve just lost a bit of efficiency, but it’s very robust.

Ørnulf: What was the motivation behind this?

Nils: I had a nice little paper for a Prague conference proceedings (Hjort, 1994b). It was discovered by Chris Jones, who has lots of papers on kernel density estimation, but also some other things. Together, we wrote a discussion contribution on minimum L_2 estimation in the *American Statistician* (Jones & Hjort, 1994). Then, a year or two later, Jones realized that Ayan Basu, who is famous in India’s statistical circles, was researching exactly these things, around divergences. So, I was also involved in that game, along also with Ian Harris, a colleague of Chris Jones. Basu and others have a whole book on divergences for statistical inference (Basu et al., 2011), with a lot of BHHJ.

Ingrid: And that became the group with four authors?

Nils: Yes, and it has been very successful. But personally, I think the paper should have been formulated in a slightly different way, and that it should have been a bit

longer. I wrote a model selection section, which was good, but *Biometrika* said no, we only accept shorter papers. But I have no complaints about anything here, and the paper has been a bit of a surprise hit, and is being used in various places, and in steadily new contexts.

Ørnulf: You mentioned Jones and density estimation. Does that mean that your interest in density estimation came before BHHJ?

Nils: Yes, let's sort our blocks here so that we don't get mixed up in the conversation. We have block 1, survival analysis; block 2, Bayesian nonparametrics; and block 3, which includes BHHJ, but also two papers on empirical likelihood and generalizations of it (Hjort et al., 2009, 2018) and other things related to parametric inference, robustness, and such. And now you're asking about block 4, density estimation.

Ingrid: Yes, where does your interest in density estimation come from?

Nils: Let me reconstruct what we now call block 4, everything related to density estimation and such. As mentioned, I had a wonderful year at Stanford in 1985, and each year they had a summer school for everyone who was there. David Scott from Rice University in Texas was the summer school lecturer that year. I found it really exciting to follow, and there was a delightful energy among all these happy, eager summer guests at Stanford. Scott was like the local world champion in density estimation and everything related to bandwidths and such. He wrote a book about this. I found the topic both fun and understandable, and I realized that I could come up with my own ideas.

Ørnulf: So this led to new papers?

Nils: I wrote something that is on the list of Nils papers that should have been published, but I didn't have the energy to revise. I sent a long manuscript to the *Journal of Multivariate Analysis* (JMVA) about frequency polygons and average shifted histograms. For that, David Scott had two newly published papers, one on frequency polygons and one on weighted shifted histograms, but only in dimensions one and two. Then I thought, hello, I can combine these, i.e., the frequency polygons on top of the weighted histograms, and do it directly in dimension d , and better than you, David Scott. That became my paper (Hjort, 1986b).

Ingrid: But it stayed in the drawer?

Nils: JMVA was excited about the work and said, hello, revise, revise, but somehow my energies had leapt on to other themes. So, it's among the things that are in the drawer. But it has been discovered here and there. It's a Stanford technical report, and it's cited in Scott's book and elsewhere, also in big data contexts.

Ørnulf: So, what is your first published work on density estimation?

Nils: It's the work with you, Ingrid (Hjort & Glad, 1995), which is included in your PhD thesis. Do you remember when we started with that?

Ingrid: I met you in 1992 for the first time, at the Nordic Statistical Meeting in Røros, so it must be after that.

Nils: Yes, we got to know each other in Røros, and then we emailed back and forth while I was on sabbatical in Oxford from 1992 to 1993. During this correspondence, we came up with it.

Ingrid: You came up with it, I have to say. I remember at my PhD defence, the first question from the first opponent, Chris Jones, was "Dear Ingrid Glad, how on earth did you manage to come up with this very smart thing?" And oh God, what do I answer now? Then I immediately clarified that it was my co-author Nils Lid Hjort who came up with that idea.

Nils: It was fun and worked out nicely, and it became a very fine paper that has also had a lot of citations.

Ørnulf: But your collaboration with Chris Jones on density estimation, when did that start?

Nils: I think we worked on it a little in parallel with the work on Hjort-Glad. Because when I was at Ingrid's PhD defence, Chris and I were already acquainted and working on several topics.

Ørnulf: I remember he visited you here in Oslo.

Nils: Yes, he was here in Oslo in a cold February week, where he stayed with us and had to borrow my rubber boots. He hadn't been able to imagine, in his English head, what it would be like in wintery Oslo. That's when we finished our work on using local likelihood for estimating a density f (Hjort & Jones, 1996). This has its background, which is academically interesting enough to point out. In a sense, I had come up with Hjort-Jones before, but within the survival analysis framework, where one estimates the hazard function α . One can write down a local likelihood for α , which is easier to write down than for f , and easier to communicate, provided the audience is acquainted with survival analysis. *Biometrika* was interested in this.

Ørnulf: It was submitted to *Biometrika*, but it was never published?

Nils: It was a good paper, and it exists as a technical report from the University of Oslo (Hjort, 1993). *Biometrika* said, "We'll take it, but you need to revise it." For silly reasons, I didn't revise. But that's the background for coming up with Hjort-Jones. I have also done other works on density estimation, but I think the two most important ones are Hjort-Glad and Hjort-Jones. So, this is block 4.

Ingrid: But what about model selection, focused inference, and such topics?

Nils: Yes, the fifth block encompasses everything around the focused information criterion (FIC), model selection, along with various related matters. The background for this can be recapped roughly as follows. It was in the 1990s that I became curious about a topic that later turned into model selection and so on. Everyone learns about parametric models in their studies. But then you may start thinking, how far from your parametric model do you have to be before it really means that the model's results are too poor? How much tolerance is there around a parametric model before it's worth not using it?

Ørnulf: But what was your motivation for starting to think like that?

Nils: It came to me, as they say, and the question is not particularly original as such. But I may have been original in my thoughts around the topic and the methodology that provides precise answers. If you want to work with your parametric $f(x, \theta)$ model, where θ has dimension p , you can say, okay, let's add a γ parameter, of dimension q . You can start by letting q be 1. So, if you like your five-parameter model, you can still try to add a parameter number 6. How close to the null value must parameter number 6 be for your five-parameter model to remain the best? Then you create a framework around this, and as soon as the question is properly posed, it is solvable by applying some maximum likelihood theory. However, it's beyond textbook territory because you have to remember that the model is not entirely true, so there are some additional terms to consider, for both biases and variances. After quite extensive derivations, you get a beautiful answer: the tolerance radius around your favourite model is τ/\sqrt{n} , with an understandable formula for τ .

Ingrid: Is this work published?

Nils: I wrote, in my own biased opinion, a very nice paper which I submitted to the *Journal of the American Statistical Association* (JASA). Again, they said, there's a lot of good stuff here, but we can't accept it as is. Can you revise it? I didn't have the energy. Then I wrote another paper in a corner of this room, which was published in JASA (Hjort, 1994a) and has the slightly odd title, "The exact amount of t-ness that the normal model can tolerate." The answer is that as long as the degrees of freedom is greater than $1.458\sqrt{n}$, normal-based inference is better than t-based inference.

Ørnulf: But what about the other paper?

Nils: JASA wanted t-ness, an interesting special case, but they didn't want the larger and more important parent paper without some editing. So, it ended up in a drawer, and it's a technical report from the University of Oslo (Hjort, 1991). Then several years passed where I didn't think much about this topic. But then I met Gerda Claeskens in Australia in 2000.

Ørnulf: Was she on sabbatical there?

Nils: She had half a year, I think. She went to Canberra, just like me, around the same time as my two months there. She was looking forward to working with Peter Hall. But the first thing Peter did when she arrived was to go to Europe for six weeks.

Ingrid: So there she was.

Nils: She was stuck with me! The group of visitors around Peter Hall talked to each other about what each one was working on. Some of the things Gerda had worked on reminded me of my work on model tolerance. Then I realized that the framework I had used, in addition to answering questions about model tolerance, could actually be used to understand more things on the map by putting it in the right way and calculating asymptotics. In the mentioned framework, it's about working with densities of the form $f(y, \theta, \gamma_0 + \delta/\sqrt{n})$, and calculating the consequences for all your estimated focuses $\hat{\mu} = \mu(\hat{\theta}, \hat{\gamma})$. You get one of these for each candidate model. Then both variances and squared biases have size $1/n$, so they can be combined in the right way and compared relatively easily. This can be used to understand how your $\hat{\theta}$ s and $\hat{\mu}$ s are doing if the model is a bit wrong, you can average over models, and you can find the answer to the following question: the statistician who uses AIC or BIC, how is the final estimator after model selection actually doing? Because, what does Breiman say?

Ingrid: In his paper about the two cultures (Breiman, 2001)?

Nils: No, I'm thinking of an earlier paper. The Quiet Scandal of Statistics, says Breiman (1992), is that statisticians do a lot with their data before they decide, after three days, to use model no. 17. But then they pretend that model no. 17 was given on day no. 1. Then they have pushed a lot of uncertainty under the carpet. But then we could answer questions concerning this uncertainty. The initial thought was to sort out the consequences of post selection estimation, how things are going with model averaging, when is the Quiet Scandal also a Big Scandal. As an afterthought, we realized that given the mathematical framework and the nice answers, we can create a model selection criterion ourselves, which finds the very best model for a given question. That is for the parameter $\mu = \mu(\theta, \gamma)$ you choose to focus on, based on context. Then we have FIC.

Ørnulf: From scandal to FIC! (Figure 3)

Nils: Exactly. It was really fun, and Nils–Gerda was the perfect match at the right time. We had several good and hectic years after those couple of months with Peter Hall in Canberra. Gerda visited me here in Oslo, and I visited her in both Leuven and Texas. It grew under our fingers, after many rounds of chalk and blackboard, and we realized that we had to write two substantial papers, one about frequentist model averaging and one about our new approach to model selection. We finished the FIC paper first and submitted it to JASA. In the introduction, we explained that the methodology requires various results from the other paper, which we planned to submit to *Annals of Statistics*. Frank Samaniego was the editor of JASA at the time, and he and a couple of associate editors quickly said, this is great stuff, we want FIC. But you are referring to another paper that isn't finished – can't you finish it and submit it to us as well? Alright, we said.



FIGURE 3 Nils meets Gerda Claeskens (no. two from the left in front) in Canberra in 2000, the idea of focused information criterion (FIC) is born, and 16 years later researchers from nine countries come together at a three days FICology Workshop in Oslo.

Ørnulf: The papers in JASA on FIC (Claeskens & Hjort, 2003) and frequentist model averaging (Hjort & Claeskens, 2003) are among your most cited works, and they were quickly noticed, right?

Nils: Yes, firstly, they were named “Papers of the Year” in what’s called JASA, Theory and Methods, and they were presented at the Joint Statistical Meeting in San Francisco 2003, with several prominent discussants. People jumped on board and wrote other FIC papers in different contexts, so it was very exciting.

Ingrid: And then there came a book!

Nils: Gerda and I wrote several papers in those years, including one in JASA with ficological methods for Cox regression. And without doing anything, we got several offers to write a book about our new model selection approaches. This eventually became “Model Selection and Model Averaging” (Claeskens & Hjort, 2008), and we received excellent assistance from Cambridge University Press. The book has been well received and has over two thousand citations on Google Scholar.

Ingrid: It seems reasonable that one often wants to focus on a particular quantity. Isn’t it a bit strange that no one had thought about this before Gerda and you?

Nils: There are many statements of the type “the parameter of interest is” in the literature. So, we are not the first to consider it, but we have stated it with greater

emphasis, and with further subsequent methods and formulas, placing it in the right framework. It's helpful and instructive to understand that AIC and BIC, which people use, are not wrong but are in an overall mode. But now we have focused, and I think it's simply a fruitful idea. We are statisticians, analysing lots of data, but it's helpful for us and for the readers of our reports to sharpen our focus for certain quantities.

Ørnulf: You have also written a book with Tore Schweder on confidence distributions. Can you tell us about this work?

Nils: Yes, I have also become very fond of this topic. It comes from Tore, who has been working on this for many years. One can approach the subject from different angles. You can connect it to what we learn in our studies about confidence intervals, and think about confidence intervals for all possible levels from 0.01 to 0.99 at once. Another angle is to connect it to likelihoods. Tore and I have an invited paper in the *Scandinavian Journal of Statistics*. I presented it at the Nordic Statistical Meeting in Grimstad in 2000. In this "Confidence and Likelihood" (Schweder & Hjort, 2002) lies the seeds for many other things.

Ingrid: So, this is where the fiducial comes in?

Nils: Yes, Tore had a liking for Fisher's fiducial distributions. As we all know, Fisher is famous for a long list of things, but in a pertinent historical footnote, he did not succeed with this particular concept. People found errors and lack of coherency. So, for a long time, this was like "Fisher's only blunder." But the train of thought is alive nonetheless, increasingly so, in fact, from around 2000 onwards, with frequent BFF conferences (Bayes, Fiducial, Frequentist, Best Friends Forever). Several heavyweights are involved, including Efron, who has always been a Fisher fan (Figure 4).

Ørnulf: Wasn't Cox also involved in this?

Nils: Sir David was also involved, yes, among the aforementioned heavyweights. Both he and Efron supported the idea that Tore and I should pursue this further. I was introduced to the topic by Tore for the first time in the late 1990s when Tore used it to combine different sources of information in estimating the size of whale populations. The background was that some clever Bayesians had tried something that didn't make complete sense when you looked closer at it. Tore realized that with the Bayesian melding that Raftery and co. had introduced (Poole & Raftery, 2000), you quickly encounter serious problems.

Ørnulf: Melding?

Nils: Yes, melding. They called it Bayesian melding, as in combining or synthesizing; how to combine five posterior distributions, which may have certain common parameters in non-trivial ways. So, Tore proclaimed that this requires a different type of melding, and it should be connected to confidence distributions (CDs) and likelihoods. In a footnote, one understands that if you have many independent sources of



FIGURE 4 BFF, Rutgers 2016: Best Friends Forever, Bayes, Fiducial, Frequentist: Nils and Brad Efron.

information and manage to transcribe each of them into a likelihood, then you are safe. Then you can add up log-likelihood functions and use this to draw conclusions, even when there are common indirect parameters at play for the different sources. But there are lots of complexities around this. There are also various manipulations and connections in the background, which link this to CDs. I cannot easily explain all of this in seven minutes. You almost have to read the entire CLP.

Ørnulf: CLP?

Nils: “Confidence, Likelihood, Probability,” which is Tore’s and my book on this at Cambridge University Press (Schweder & Hjort, 2016). There are connections from confidence to likelihoods, and then you can do a lot around this. Also, about meta-analysis and combining various types of sources of information. But to answer the question you are asking, how did I get into our identifiable block number 6, confidence distributions and more, well, the answer is Tore (Figure 5).

Ingrid: Was it new for you, to sort of step into Tore’s room? While most of what you have said previously has been the opposite, that you have invited people into your rooms?

Nils: Yes, that is a good point. But it is not easy to answer very clearly and uniformly about that. You are right to point out that if we list the Nils’ blocks, this block is a bit different from the others. Because it took me a while to get inside Tore’s room. But then I eventually got into it, it became fun, and then it became a whole book!



FIGURE 5 Nils and Tore Schweder with their CLP book at the ISBA 2016 World Meeting.

Ørnulf: We should also talk a little bit about your interests in speed skating and speed skating statistics. You actually contributed to changing the rules for the Olympic 500 meters, so that the skaters in the period of 1998–2014 had to skate two 500 meters, once starting from the inner lane and once starting from the outer lane.

Nils: My speed skating interests are still going strong. I'm the administrator of the Forum for Speed Skating History, ten thousand enthusiastic nerds, on Facebook. I like the sport of speed skating, I like the numbers, the score sheets, and the epic drama of the overall point scores. In that sense, it was natural for me to look into speed skating data. It started as a small student project in a course where the students learn about different regression analyses, among other things. The results of the project were interesting enough that I obtained the result lists for the 500 meters from about ten sprint World Championships, with information about the inner and outer lanes, passing times after 100 meter, and more. The analysis showed a highly significant deviation from the null hypothesis that everything has been in order since the 1924 Olympics. The estimate for the parameter d , the difference between starting from the inner and outer lane, is as small as $\hat{d} = 0.06$ seconds, but the confidence interval is from 0.04 to 0.08, enough for medals to change necks. And that led to this Olympic change from Nagano 1998.

Ørnulf: Has this work been published?

Nils: It is actually another foolishness on my part, alongside the others I have mentioned earlier, where they wanted my papers but I didn't have the energy to

do the revision, or more accurately my energies had jumped to other themes. My Olympic paper (Hjort, 1994c) was sent to the *American Statistician*, who wanted to publish it. It should have been published there, but something came in the way. I am happy, though, about having used our type of professional tools and insights to various not-so-standard applications, now and then, with wars, whales, Russian and mediaeval literature, in addition to sports statistics.

Ørnulf: It has been very interesting to hear about the background of the various topics you have been interested in. But let's now move on to hear your views on some overarching issues. You are known for being very good at doing asymptotic calculations, so you calculate asymptotics without blinking. At the same time, we are in a world where computing is becoming more and more important. Can you reflect a bit on the role of asymptotics in a world where there is a lot of bootstrapping, computing, and simulations? Is it still important? Or is it more of an academic exercise that the journals require for some papers?

Nils: My view is that it is partly worth the exercise, to see clearly how things work. To find the confidence interval you need for a specific report, you can probably manage it in another way, through bootstrapping or some simulation and so on. But by doing the asymptotics you clarify what assumptions and conditions are needed to make things work. Another case in point is everything around FIC, where I would say that we would not have come up with FIC if we did not have asymptotics to help us clarify the structure and answers.

Ingrid: So there is a two-way bridge here.

Nils: Yes, in several situations. Point 1, we clarify the answers, the essence emerges, the fog lifts. Point 2, we use the answers and the structure of the answers to develop the methods. Then we have used asymptotics in a different way than verifying that a confidence interval has a 95% limit. But it's okay to raise your hand as an eager student and say, "why do I need to learn this when I can bootstrap?" We need to be prepared for such questions, and we should have answers. A partial answer is that sometimes it doesn't work. There are certain things in ficology where the $1/\sqrt{n}$ term makes it so that if you try to bootstrap, you get a slightly wrong answer. Secondly, it is insightful, and you learn how certain estimands are much more difficult than other estimands that are more standard. Sometimes you also come up with things that don't work as you thought. You can suddenly be in the realm of cube root asymptotics, and not in the square root asymptotics terrain you thought you were in.

Ingrid: Can we say that behind the road through your works, as we talked about earlier, behind many of the ideas there, lies the fact that you have such a good insight into calculating asymptotics and seeing things?

Nils: Partially yes. I'm not sure if I'm answering with an exclamation mark, but there is something there. It is possible to come up with BHHJ, so to speak. It is a divergence that generalizes Kullback-Leibler. In that case, I think, great, but we have

to work out some more theory before we understand how it works. It's not enough to give seven examples that make it likely that the idea works.

Ørnulf: Isn't that also the case in BNP? Is that why you are interested in showing Bernstein–von Mises theorems?

Nils: Yes, I think that is part of the package that not everyone has to care about, but some should care about it. It is present in regular Bayes, so to speak. And even more so in BNP because the space is so large and unwieldy, and you don't know how things will go, and you can get certain surprises. The super-principled Bayesian says that I don't care because I have data, I have thought carefully and not only created a good model but also a good prior; the rest follows. End of discussion. That is a legitimate approach. But it is also allowed to ask, what about the next ten people who will use your method? Then you sooner or later come to the question that most Bayesians accept as a valid problem. What are the frequentist properties of Bayesian methods?

Ørnulf: And then Bernstein–von Mises is important to show that it actually works. Are there examples where it goes wrong?

Nils: Yes, definitely. That's how it is in an *Annals of Statistics* paper that I came across when I was at Stanford. Diaconis and Freedman (1986) were the first to discover this clearly. Here are some examples. Everything looks good. We are indirectly or directly in BNP. But if you look at how the $\hat{\theta}$ that comes out of this behaves, it will sometimes fluctuate, or behave improperly. Since people at Stanford realized that I knew a lot about this, I was invited to write a discussion contribution. I used it to its fullest potential (Hjort, 1986a), and came up with some Bayesian semiparametrics, nonparametric envelopes around parametric models. Some of this has to do with Bernstein–von Mises stuff, but in huge spaces. When will your big Bayesian prior lead to things you hadn't thought of, or cause your estimates to behave genuinely differently from what the frequentists get?

Ørnulf: You've worked with both Bayesian and frequentist methods. Some people are staunch frequentists, and some are very Bayesian. What is your role here?

Nils: I'm not sure if I can give a very unified answer to that, or if I should be surprised or apologetic to say that I have no problems being in both camps. I don't think it's a paradox. But to the extent that I am a part-time Bayesian, as I am, I have added as an additional dimension, let me check how it goes with repeated use of a method.

Ørnulf: We should also say some words about your role as a supervisor. We know that you have been a huge source of inspiration and an important mentor for many talented PhD students and even more master's students. But what has the supervision meant for you, both academically and personally?

Nils: I almost always find supervision enjoyable. I mostly think about it as a pleasant conversation with one or more fellow human beings. I become fond of these people, and they become both friends and buddies, to use such words.

Ørnulf: But let's go back to the beginning. You mentioned your six years as a research assistant, where many of the seniors were not very interested in what the young ones were doing. Has this meant anything to you? Have you reflected on the responsibility you have as a role model?

Nils: If I try to consider the question seriously, I mostly think the answer is the slightly uninteresting "it has become like that," because it suited me and my personality and the groups of people around me. But yes, I think about these dimensions more than before, on my own and on behalf of the system. But I hesitate to formalize it as principles. I'm conscious of my role, more than before, and I think that these young, talented people deserve the best of care.

Ingrid: Nils, we can't avoid talking about your teaching. You are famous, and maybe for some students infamous, for giving very well-organized lectures on any topic without any notes.

Nils: And without preparation. I can go to the blackboard and talk continuously for two hours about quite a few topics without having thought about them in advance. Of course, in a different sense I *am* prepared, because I thought about it ten months or seven years ago. We all have our strengths, and possibly among mine is that I remember very well if there are things I have understood and worked on. So I can easily go back to an argument from 1992 and quickly reconstruct what it was. It's a strength in a teaching situation that one can speak cohesively about topics one has understood. But I understand that those who have liked me the most might be the clever students who tolerate a bit of digression, who tolerate that I suddenly talk for four minutes about something completely different, which is not part of the curriculum.

Ingrid: It's not just about tolerating it, but they appreciate being challenged, and that applies to the most capable students, most of the time.

Nils: Yes, but I haven't had the conscious thought in my teaching mind, now for almost fifty years, that when I lecture, I'll only think about the brightest students. Of course, I also think about the brightest, by all means, but I haven't tried to stretch the ambitions to do so. But it's possible that it has turned out that way anyway.

Ingrid: You are now an emeritus professor, Nils. But you are still just as active in research, teaching, and supervision. What motivates you to continue in this way?

Nils: I cannot provide a clear answer to this, although I ponder these questions, more than before. Tentatively, I have the answer in my practical life, that, yes, I continue almost as before, eventually with less volume. The formula "never change a winning team, never change things that work," or something like that, is quite good, then. I enjoy my work. Sometimes thinking alone for half a day, sometimes talking to good colleagues and being exposed to things that one immediately understands and things that one does not immediately understand.

Ørnulf: But do you also have specific things that you are passionate about? Something you would like to explore academically?

Nils: It's not like that with me, as it might be for certain pure mathematicians who have struggled with Theorem 4 for ten years, just waiting for a proof to be found, or for a formidable Unity of Science to finally be summarized and brought into the light. I am working on various articles, with several things in the drawers, but not of this specific type. There are projects I want to complete, from a book with Emil Stoltenberg to a dozen papers that, based on a loose definition, I am working on or have thought enough about, with and without co-authors.

Ingrid: But you still have ideas and inspiration to move forward?

Nils: I think every week that there are enough interesting and challenging problems. Some problems I quickly understand how to approach, while with other problems, I don't know how one can even get started. Often enough, a phase of curiosity comes, which gives me a push. I declare myself sufficiently optimistic when I think about this, and the ideas still come, a source of continuous creative joy.

Ingrid: The privilege, Nils, when you are an emeritus, is that you can choose the fun things.

Nils: Yes!

ACKNOWLEDGMENTS

This work was partially supported by the Research Council of Norway through the Centre for Research based Innovation BigInsight, project number 237718, and the Centre of Excellence Integreat – The Norwegian Centre for knowledge-driven machine learning, project number 332645.

ORCID

Ørnulf Borgan  <https://orcid.org/0000-0002-4687-1202>

REFERENCES

- Andersen, P. K., & Borgan, Ø. (1985). Counting process models for life history data [with discussion]. *Scandinavian Journal of Statistics*, *12*, 97–158.
- Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*, 549–559.
- Basu, A., Shioua, H., & Park, C. (2011). *Statistical inference: The minimum distance approach*. Chapman and Hall/CRC.
- Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. *Journal of the American Statistical Association*, *87*, 738–754.
- Breiman, L. (2001). Statistical modeling: The two cultures [with comments and a rejoinder by the author]. *Statistical Science*, *16*, 199–231.
- Claeskens, G., & Hjort, N. L. (2003). The focused information criterion [with discussion]. *Journal of the American Statistical Association*, *98*, 900–916.
- Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University Press.
- Diaconis, P., & Freedman, D. (1986). On the consistency of Bayes estimates. *Annals of Statistics*, *14*, 1–26.
- Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, *1*, 209–230.

- Green, P. J., Hjort, N. L., & Richardson, S. (Eds.). (2003). *Highly structured stochastic systems*. Oxford University Press.
- Hjort, N. L. (1985). Discussion of P. K. Andersen's and Ø. Borgan's article, 'Counting process models for life history data'. *Scandinavian Journal of Statistics*, *12*, 141–151.
- Hjort, N. L. (1986a). Discussion of P. Diaconis' and D. Freedman's article 'On the consistency of Bayes estimates'. *Annals of Statistics*, *14*, 49–55.
- Hjort, N. L. (1986b). *On frequency polygons and average shifted histograms in higher dimensions* (Technical Report No. 22). Department of Statistics, Stanford University. <https://purl.stanford.edu/nj166wb6803>
- Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life-history data. *Annals of Statistics*, *18*, 1259–1294.
- Hjort, N. L. (1991). *Estimation in moderately misspecified models* (Statistical Research Report No. 8/91). Department of Mathematics, University of Oslo. <http://urn.nb.no/URN:NBN:no-51813>
- Hjort, N. L. (1993). *Dynamic likelihood hazard rate estimation* (Statistical Research Report No. 4/93). Department of Mathematics, University of Oslo. <http://urn.nb.no/URN:NBN:no-51779>
- Hjort, N. L. (1994a). The exact amount of t-ness that the normal model can tolerate. *Journal of the American Statistical Association*, *89*, 665–675.
- Hjort, N. L. (1994b). *Minimum L2 and robust Kullback-Leibler estimation*. In *Transactions of the 12th Prague conference on information theory, statistical decision functions, random processes* (pp. 102–105) Institute of Information Theory and Automation, Czech Academy of Sciences.
- Hjort, N. L. (1994c). *Should the Olympic sprint skaters run 500 meter twice?* (Manuscript). Department of Mathematics, University of Oslo. <http://urn.nb.no/URN:NBN:no-28549>
- Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators [with discussion]. *Journal of the American Statistical Association*, *98*, 879–899.
- Hjort, N. L., & Glad, I. K. (1995). Nonparametric density-estimation with a parametric start. *Annals of Statistics*, *23*, 882–904.
- Hjort, N. L., Holmes, C. C., Müller, P., & Walker, S. G. (Eds.). (2010). *Bayesian nonparametrics*. Cambridge University Press.
- Hjort, N. L., & Jones, M. C. (1996). Locally parametric nonparametric density estimation. *Annals of Statistics*, *24*, 1619–1647.
- Hjort, N. L., McKeague, I. W., & Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Annals of Statistics*, *37*, 1079–1111.
- Hjort, N. L., McKeague, I. W., & Van Keilegom, I. (2018). Hybrid combinations of parametric and empirical likelihoods. *Statistica Sinica*, *28*, 2389–2407.
- Hjort, N. L., & Mohn, E. (1987). *Topics in the statistical analysis of remotely sensed data*. In *Bulletin of the international statistical institute: Proceedings of the 46th session*. Invited paper 21.2. International Statistical Institute.
- Hjort, N. L., & Omre, H. (1994). Topics in spatial statistics [with discussion]. *Scandinavian Journal of Statistics*, *21*, 289–357.
- Hjort, N. L., & Walker, S. G. (2009). Quantile pyramids for Bayesian nonparametrics. *Annals of Statistics*, *37*, 105–131.
- Jones, M. C., & Hjort, N. L. (1994). How to approximate a histogram by a normal density. *American Statistician*, *48*, 353–354.
- Poole, D., & Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, *95*, 1244–1255.
- Schweder, T., & Hjort, N. L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics*, *29*, 309–332.
- Schweder, T., & Hjort, N. L. (2016). *Confidence, likelihood, probability*. Cambridge University Press.

How to cite this article: Borgan, Ø., & Glad, I. K. (2024). A conversation with Nils Lid Hjort. *Scandinavian Journal of Statistics*, 1–22. <https://doi.org/10.1111/sjos.12732>