Submitted to the Annals of Statistics

A CONVERSATION WITH ESTATE V. KHMALADZE

By Hira L. Koul and Roger Koenker

Michigan State University and University of Illinois, Urbana-Champaign

x

Estate V. Khmaladze was born in Tbilisi, Georgia, on October 20, 1944. He earned his B.Sc. degree from the Javakhishvili Tbilisi State University in 1964, majoring in physics. and his PhD in mathematics in 1971 and Doctor of Physical and Mathematical Sciences in 1988, both from the Moscow State University. From 1972 to 1990 he held appointments at the Razmadze Mathematical Institute in Tbilisi and interim appointments at the V. A. Steklov Mathematical Institute in Moscow. From 1990 to 1999 he was head of the Department of Probability and Mathematical Statistics of the Razmadze Institute. From 1996 to 2001 he was on the faculty of the Department of Statistics of the University of New South Wales. Since 2002 he holds the Chair in Statistics in the School of Mathematics and Statistics of Victoria University of Wellington, New Zealand. He is a Fellow of the Royal Society of New Zealand and of the Institute of Mathematical Statistics. In 2013 he was awarded Javakhishvili Medal from Tbilisi I. Javakhishvili State University and was elected to be a Foreign Member of the Georgian Academy of Sciences in 2016. As the conversation reveals Khmaladze's research ranges widely over statistical topics and beyond.

The conversation began in the old building of I. Javakhishvili Tbilisi State University during a conference on probability theory and mathematical statistics, September 6-12, 2015, and continued in the Research Center of Ilia University, Stephantsminda, during the subsequent workshop, 12-16, September, Georgia. Mount Kazbegi, 5047m, with its white summit was occasionally visible not too far away. In what follows, the questions are put in italics while the Estate's answers appear in the standard font.

Estate, tell us about the place and date of your birth, the place you grew up, a bit about your family, and your early schooling. What were the professions of your parents?

AMS 2000 subject classifications: Primary 62G10; secondary 62G20.

[1] *Corresponding author. Hira Koul.*

Key words and phrases. Khmaladze transform. Asymptotically distribution-free GOF tests.

I was born in Tbilisi, on October 20, 1944. My father was Georgian and mother was Armenian with some German ancestry. My father was a civil engineer by profession, and his speciality was tunnels and bridges. During World War II he was with the army – as everyone else – and afterwards he would tell me stories about how he opened spaces for tunnels in the mountains, making them very large because he knew that way they would be more stable, and then how he had to put up a false casing so that the troops could go through without worry. Later he became a well known civil engineer in Georgia.

My mother was a person of very broad mind. Talking to her was very interesting. She could have had a very good career in front of her, but she had the stigma of being a daughter of the "enemy of the people." In 1937, my grandfather, her father, Gurgen Dandurov, was arrested and died soon thereafter from a heart attack. He was the deputy head of Trans-Caucasus railway. Imagine Tbilisi in those days, only a couple of dozen cars were driving around, one was his. It was an important position. But he came to a disagreement with Lavrenti Beria, the party leader in Georgia at that time. Of course he was arrested.

The next day my mother, then 16, was kicked out of the apartment along with her old grandmother, and a KGB officer was moved in. Many years later, when it was safe to apologize, the wife of this KGB officer apologized with tears to my mother saying "Oh dear, it was not our fault." And indeed, it wasn't.

I went to school in Tbilisi from 1951 until 1961. I have very fond memories of the school. It was in a blue collar district, but my mother said "the morals are better there and the teachers are very good." And they were. My mathematics teacher was a man called Herman Fercher, a repatriated German – Germans were forcibly evacuated from Georgia, where they had been settled a long time ago, to Soviet Central Asia. After the war they were allowed to return to Georgia. German Nikolaevich, as he was called by us, challenged my friend and me to solve only *-marked exercises in the math textbook through various grades, the kind of "difficult" ones. But neither of us was a nerd, and we were not thinking of becoming mathematicians – at that time we wanted to become chemists. My friend became an electronics engineer, and here I am, a statistician.

Where did you go to college and for graduate studies? At what time did you develop an interest in mathematics, and in particular in probability theory and statistics?

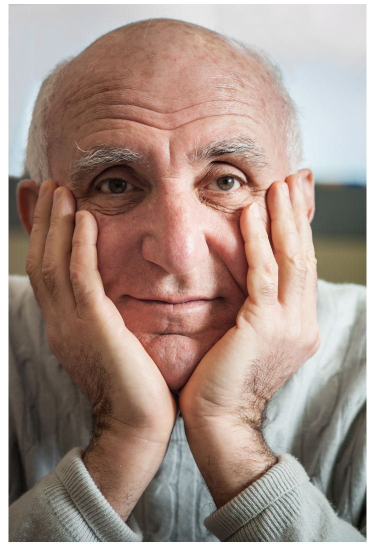

FIG 1. *Estate Khmalazde with his friend and neighbor Elguja Khucishvili in 1959 in Tbilisi, left, and in his office in Wellington, right.*

I went to I. Javakhishvili Tbilisi State University for B.Sc., but I never wanted to be a mathematician. The Faculty of Mathematics at the university where we are sitting now was very strong professionally with a very solid education. But it was somehow gloomy, without flair; a little bit dull to my taste. The Faculty of Physics, on the other hand, was so bright, physics was strong, reading books and knowing who William Faulkner and Thomas Mann were was necessary, and listening on the radio to Willis Conover's *Time for Jazz* and *Music USA* was very common. After high school I joined the Faculty of Physics. I am still half physicist in my soul.

However, about the time that I got my B.Sc. in physics in 1964, Tbilisi State University established a new Faculty of Cybernetics. I was all right as a physics student, but "cybernetics" sounded so mysterious and attractive, so promising of big discoveries, that I betrayed physics. Nobody really knew what "physical cybernetics," the major I enrolled in, was, but there was a very good course in probability theory. The person who taught us is worth a separate story, a very colorful person. His name was Amiran Toronjadze. The favorite pupil of the astrophysicist W.A. Ambarzumian, he had a legendary reputation in the Abastumani Observatory, but quarreled with the management and moved to Tbilisi. He had great influence on my learning

Fig 2. *Estate Khmaladze with Elizbar Nadaraya, at the opening of the conference in Tbilisi, 2015 September.*

probability.

I secured an M.S. degree in Cybernetics and Applied Mathematics in 1966 and started as Junior Researcher at the new Institute of Applied Mathematics. There was a culture of "seminars" – not just in the sense of a single talk, and not as a sequence of talks, but rather as a "research group." If you belonged to A's seminar it meant you belonged, in a more or less broad sense, to A's research group. We also had our seminar in probability. At that time Doob's *Stochastic Processes* and Loève's *Probability Theory* were newly translated books for us. We were studying both books very thoroughly – every page, every exercise, and it was good schooling. The spirit was such that we wouldn't think of saying "oh, it is too much."

What are your impressions of your first years as a researcher?

We are talking about 1966-68. Life was cheerful and after seminars we would go for beer. "We" included me, Kacha Dzhapharidze and Rezo Chitashvili. They were 3 years older, which was something when you are 22, and Rezo already had a reputation of a genius, which he very much deserved. Another person who would need a separate story. Kacha often talked about getting away from Tbilisi. He eventually left in 1967 and became a PhD student of Akiva Yaglom in Moscow. Then, after a few years, he went much further and ended up at the CWI in Amsterdam. In 1968 I also left for Moscow, and that is how my stint at the V. A. Steklov Mathematical Institute started.

But I did not go until Rezo Chitashvili had pushed me into studying the then very fresh book of Erich Lehmann on testing; again, all exercises on all pages. I still think, as I thought then, that it was a great book. In two months Rezo and I knew the book: "now you're ready" he told me. While we are still on books and education in statistics, H. Cramér's *Mathematical Methods of Statistics* needs to be mentioned – it was so very good. And the first book for me, when I still was at the university, was Gnedenko's *Theory of Probability.*

Maybe I should tell you about my "stint" at Steklov Institute? This wasn't a usual arrangement in those years of the Soviet Union – half a year in Tbilisi, half a year in Moscow. Actually, the flight between Tbilisi and Moscow was 2 hours. Here the flight to Sydney or Melbourne, takes 3 hours and nobody is thinking much about it – just our neighbors. But back then some special arrangements were needed. In this respect I was lucky: the head of department in Tbilisi was Gvanji Mania, and the deputy-director at Steklov was Yuri Prohorov; and they were mates. It was a very good and long friendship. So, they both thought that it was a good idea if I spent time in both places.

Gvanji Mania was unusual person: very clever, good and broad-minded. I do not want to sound pompous, but people like him are, actually, great enablers of progress in the world. He cared about everybody – from the smallest clerk to the members of Academy, everybody needed his word, his advice and support. Mail addressed to just "G. Mania, Tbilisi," would reach him alright. There are many stories one can remember, about Gvanji – in our everyday Tbilisi life, in Bakuriani Conferences (Bakuriani is winter resort in Georgia, where we ran conferences in probability theory for about 20 years), in our meetings in Moscow. But – maybe, next time.

What was your motivation to work on what is now known as the Khmaladze transformation?

Well, in 1975 or 1976 we were in Vilnius, at one of the Vilnius conferences, and Longin Bol'shev, my scientific "boss," gave a talk on chi-square statistics: just Pearson's chi-square statistics for continuous observations grouped into class-intervals, but with parameters estimated by these continuous observations, not class frequencies. For example, the expected value estimated by a sample mean, not by mid-points of the class-intervals times the frequencies. Then, the chi-square statistic is not asymptotically chi-square distributed. What Bol'shev suggested was that instead of sticking to the form of the statistics, even though very traditional, one should consider a different quadratic form, one which again will have a chi-square distribution. He was very happy to tell us, his pupils, that Kolmogorov liked his talk quite a lot

(Estate laughs). The next obvious question was how to modify parametric ω^2-statistics, which Bol'shev charged us with. However, it soon became very obvious that you should not work on separate statistics, but rather try to transform the parametric empirical process itself.

In 1979 I published a paper, Khmaladze (1979), where the class of such transformations was suggested. It was based on sequences of Fourier coefficients from the parametric empirical process, which are asymptotically standard normal. Using them, one can then construct any Gaussian process one wishes. These Fourier coefficients were similar to the "components" of Durbin, Knott and Taylor (1975), but did not require spectral decomposition of the covariance operator of the parametric empirical process. I liked them, and I still think they are very easy to construct, but nobody noticed their existence apart from my Soviet colleagues. It was also then that I realized that whenever you estimate a finite-dimensional parameter, it does not really matter how, using MLE or not, what you get asymptotically is a projection of Brownian bridge – either an orthogonal or skew projection. It looked sufficient in the late 70's and 80's to say that if you estimate a parameter, asymptotically you get a different Gaussian process, full stop! But a lot can be gained by understanding that the limiting process is actually a projection. It explained, for example, how it can be that you estimate a parameter and, as a result, you gain power.

This all, however, was in the background for my 1981 paper. At the time, the existence of a connection between the theory of empirical processes and the theory of semi-martingales was not known. Longin Bol'shev and Albert Shiryaev were sitting two doors apart from each other and were good friends for many years, but did not suspect the existence of any connections.

In August 1978 Bol'shev died, a great loss for me, personally and scientifically. When my "martingale approach" paper, submitted to *Theoriya Veroyatnostei*, came up at the meeting of the editorial board, it was not Bol'shev, but Shiryaev who presented it. Statistics was "orphaned" and Shiryaev was looking as if he would take care of it, to my great detriment. (*laughs*).

I heard Shiryaev liked the paper very much, but that drew negative reaction from several others in the probability and statistics community in the Soviet Union. Somehow, it was split between those who would say that the paper contains a "great discovery" and those who would say that "this cannot be true, there must be a mistake there." This latter point of view was somewhat slowly, but still evolving. At some stage, after three or four years, a very interesting position was taken by Yuri Rozanov, who certainly knew

FIG 3. *Left: Roger Koenker, Hira Koul, Estate Khmalazde on the way to Kazbegi*

the version of innovation theory in terms of Volterra operators in Hilbert spaces. This theory is equivalent to the innovation theory for Gaussian semimartingales. “This is not new for me,” he would say. Of course it was not – it would be as if you arrived to an unknown island and told the aborigines that you have discovered their island. They would laugh at you. (*laughs*). But was it not a discovery for the outside world?

Sorry to interrupt, did these things affect you, so to say, personally?

Oh, yes. Under the former Soviet system, without the degree of Doctor of Sciences, your chances for becoming, say, university professor were very remote. Any advancement would be very difficult. And is salary also a “personal” thing? (*laughs*).

Anyway, my friends were promoted, defended their Doctor of Sciences degrees, and I was sitting, working and searching for what else I could find in connections between the theory of martingales and orthodox statistics. In this way “Martingale limit theorems for divisible statistics” appeared. But there was no outside movement. Even my family members were puzzled. My father in law, himself a prominent physicist and a member of the Academy, would say “No, no, I still think Stasik (my nickname) is good” (*laughs*).

Things changed with completely unexpected support from Alexander A. Borovkov. I was in Novosibirsk in 1988, giving a seminar talk, desperate, formally requesting that the Institute of Mathematics there say something, good or bad, to end the suspense. Borovkov had had a stroke a month earlier and was not coming to the institute, but he came to my talk, and after the seminar he said "Well, I thought it would not be possible [to achieve this distribution freeness], but here it is, on the white board." And as he was the one who initially opposed, many things changed after that.

Have you seen the film about the 1980's "Blues Brothers"? Great musical, boisterous, funny. One of the main characters there says from time to time, in baritone: "God acts in mysterious ways." Indeed! (*laughs*).

However, the citations were still very scant, basically until the papers of Koning (1989, 1994), and Koenker and Xiao (2002), who liked the approach and advanced it.

If you ask me what was the main insight behind this transformation, I would say it was the idea that a little bit of the "future" could be included in the "past," and that the Doob-Meyer decomposition for the Brownian motion with respect to this "enriched" filtration, technically very simple, will lead to the Doob-Meyer decomposition of the projected Brownian bridge with respect to its natural filtration. The third fact, that the term in the limiting process associated with estimation of the parameters becomes annihilated, was very pleasant and useful, but came as an unforeseen gift. You know how it goes: if you hit the log with your axe in the natural place it will split almost without your effort. (*laughs*).

What is your memory of the two papers with Hira on fitting an error distribution in regression models?

You shouldn't have asked what I remember about that 2004 paper, because what I remember is that I cursed him, silently but often. He wanted me to work quicker and quicker, and I'd like to sit quietly and think. (*laughs*).

What I also remember is Hira's visit to Sydney. It was in May-June 2000. How after a day's work at the University of New South Wales, we would come back home, because Hira lived nearby, for some dinner, and how a bottle would appear, somehow, on the table and how afterwards I would drive him home using back streets. The work, however, was progressing. (*laughs*).

I'd rather not comment on the content of the paper much. We both knew that the transformation would work. The sport was, however, to persuade others and to present the whole picture of empirical processes in parametric regression in one unified text. Sometimes the empirical processes will be parameter free – you have to estimate the unknown parameter, but this will not change the asymptotic distribution of the empirical process, and sometimes it will not be parameter free. It is elaborated within a geometric framework in that paper.

It only seems to me a little long for a paper – several different aspects are presented in the same place. It was rather more a memoir, than a paper, in the old-fashioned meaning of the word, as in old mathematical memoirs. For a normal paper it was too long, 40 pages.

In the 2009 paper, the question we looked at was "can we provide an asymptotically distribution-free test when the estimation of the parameter cannot be of the order $1/\sqrt{n}$?" The problem was of fitting an error distribution F in the nonparametric regression model $Y_i = m(X_i) + e_i$, with i.i.d. F errors $e_i, i = 1, \cdots, n$. The regression function $m(x)$ did not have a prescribed parametric form, it was some function and was estimated nonparametrically by an estimator $\hat{m}_n(x)$. Of course, we could incorporate parameters in the distribution F of the errors, but it would look distracting, more or less a frill.

To provide a test for this problem one would have to deal with the nonparametric residuals $\hat{e}_i = Y_i - \hat{m}_n(X_i)$ and with their empirical distribution function $\hat{F}_n(y)$, or, rather, with the empirical process

$$\hat{v}_n(y) = \sqrt{n}[\hat{F}_n(y) - F(y)].$$

From the start, we had in front of us very useful results of Akritas and van Keilegom (2001) and Müller, Schick and Wefelmeyer (2007, 2009), stating that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\hat{m}_n(X_i) - m(X_i)] = O_P(1).$$

Given this result one can justify the asymptotic expansion

$$\hat{v}_n(y) = v_n(y) - f(x)R_n + o_p(1),$$

with R_n almost equal to the normalized sum in the display above. Now, this R_n is something of a sore point here, and in testing problems in general: $O_P(1)$ is nice, but you will use your estimator and I may use my estimator, and they will change the distribution of the resulting R_n and, therefore, the limit distribution of the process $\hat{v}_n$. Now imagine somebody who could say that they do not care for extra coding and transformations and would rather use computer simulations. The answer to this then could be that yes, by all means, use the simulations. But do not forget to calculate your estimator $\hat{m}_n$ for any new sample you generate. And to incorporate the details of the estimator you used. Wouldn't it be nice, however, if the term $f(x)R_n$ somehow disappears? (*laughs*).

But indeed it can and it does disappear in the transformation proposed in the 2009 paper:

$$w_n(y) = \hat{v}_n(y) - K(x, \hat{v}_n) = \hat{F}_n(y) - K(x, \hat{F}_n) + o_P(1),$$

and on the right hand side there is no F, which is immaterial, and no R_n, which is useful; that happened because the term $f(x)R_n$ was annihilated by the transformation in the middle. Of course, the estimator $\hat{m}_n$ contributes to the empirical distribution function $\hat{F}_n$, but the manifestation of its main influence, the linear term in R_n in $\hat{v}_n$, is not present in the transformed process w_n any more. Couldn't we say that this is convenient?

FIG 4. *Roger Koenker, Hira Koul, Estate Khmaladze and Robert Mnatsakanov in Stephantsminda*

Would you tell us something about the circumstances that brought you to Australia and then to New Zealand?

In 1990 I returned to Tbilisi for good. I was back in my beloved city, facing lots of challenges in the new situation in Georgia, but with no salary at all. One pay day, it was a warm Autumn afternoon, I was standing in front of our building, the Razmadze Institute, and our deputy director was coming out, "Oh, hi," he said, "this bloody government does not give a damn about its people – his language was always plain – again no salary this month."

In 1991 a local version of a civil war broke out. Although I am saying "a local version," the bullets were not made softer for Tbilisi, and lots of bearded guys were running around with very real machine-guns. In the evenings, after my wife Mzia and I had spent a couple of hours with our friends, we would walk home – about 5 km across the city in empty streets with very scant or no illumination. Well, as a matter of fact, there was some illumination, it was almost cheerful to watch – these red dotted lines of tracer bullets across the sky. Very soon you realized that these bearded guys did not mind you at all. They would ignore you. They would only mind each other, and so you were safe – unless you started running, screaming and doing totally unnecessary things like that. You also realized that although feelings and emotions were running high, still you were more of an observer than a sufferer.

True, there was no money around, and if you ask me how on earth did people survive, I wouldn't be able to tell you. Many didn't. To see good, normal people begging in the street – yes, there was a lot of trouble. But watching all this you had a feeling that you're within some general process that has its own laws, maybe its own intrinsic logic. As for a medical doctor it must be interesting to observe acute cases, I also remember, we wanted so much to understand what it was that we faced. If you say "it was so bad" it will not be enough. One can, perhaps, describe some characteristics of this state of a society, but overall, I did not reach an understanding.

Fig 5. *Estate Khmaladze with Willem van Zwet, in Keukenhof, NL, in 2012.*

With a little help from Bernard Silverman, Terry Lyons invited me to spend 1992-93 academic year at the Unversity of Edinburgh. It was a very happy year: I liked teaching, I like Terry Lyons very much, and our daughter Mariam was born in Edinburgh in June 1993. We have loved Edinburgh ever since. It is not clear why I did not make any real effort to stay in the West, only one or two feeble attempts at the end. But the fact is that we were not personally ready to abandon Tbilisi, so many friends, relatives, neighbors, all this dense tissue of life. So, we went back to Tbilisi with some savings of three or four thousand pounds, which made us feel like the Rothschilds for a year.

Talking about money, in about 1994-95 we happened to be recipients of an International Science Foundation grant, or Soros grant, and also of an INTAS grant. This latter one was organized for post-Soviet probability and statistics by Willem van Zwet. So, we had our probabilistic Mr Soros. In St. Petersburg recipients were those centered around Ildar Ibragimov, in Moscow it was Dmitri Chibisov, in Kiev – I do not remember, it was either Skorohod or Koroljuk, in Tbilisi it was me. Not me, of course, but the whole group of us who worked in probability and statistics in the Georgian Academy. It was a great help and gave us a chance to support many others. Our own salaries were restricted to something like $100-150 a month, but it was possible to buy a desktop computer, for example, which we wiped the dust from every day (*laughs*), and to pay network expenses. I am glad I can say this publicly now, because I don't think we have said "thank you" before.

By the winter 1995, however, all resources were exhausted. Electricity was intermittent, we did not have heating, and when you touched the wall it was quite cold. We were paying quadruple price for bread, or else you would have to stand the whole night in a queue. Oh, these queues – I was not standing there but I saw them at the bakery on the corner of my cul-de-sac. One scene I remember vividly: a guy, who I knew was just a normal hooligan, was shouting, frightening everybody, pushing himself over everybody's heads. Women or not women, who cared. He probably thought that he was able and strong: "fittest to survive." And he was getting his bread quicker than the others. A few years later, when we visited during my sabbatical from University of New South Wales, I learned that the poor bugger had died.

Anyway, we could not pay quadruple prices also for milk, butter and everything else. And as I said, it was very cold inside with no electricity. Although we were not going to let up – my mother would play something cheerful on our old German fortepiano, and we would dance with our baby, but I could not produce anything in mathematics: no electricity, no chance to prepare a manuscript. And then I wrote to my colleagues in Australia, with whom I had had a long standing agreement that "some time" I would spend there, say, a year. So, I wrote and said "if you still want to invite me, invite me now." They replied "no money for a visiting professor, but there is a position; would you apply?" I applied, there was a telephone interview, and we went. The way we had to arrange our visas and everything else was also interesting. But maybe it is enough on this topic; I only mention our first impression of Australia – the huge open sky.

Maybe I should also comment on the general resilience that people suddenly find in themselves: no electricity, yes, and sugar quite scarce, yes, but we had active conferences, we created the Georgian Statistical Association – something we talked about for years earlier without doing anything, we established a speciality in Actuarial Science in Tbilisi State University, and the first PhD in this subject graduated not long afterwards, we sent students to do the PhD abroad, which was not happening in other fields, we ran seminars. All of us wanted this. It is difficult to completely blunt the spirit, they will not degrade . . . , but now this sounds like lyrics of a song. (*laughs*).

And how did you end up in Wellington?.

Yes, it would be quite a long jump, if it were from Tbilisi. But it was from Sydney that I came, as we know, not Tbilisi. What attracted me very much was that the Wellington vacancy was the position of David Vere-Jones, who had retired not long before. Actually, David "retired from teaching," because I was already appointed when he and Daryl Daley continued working on the two volume edition of their famous book on point processes – a whole new volume was created.

I had never met David before and I did not see him even during my first seminar and then the interview in Victoria University. I imagined him tall, self-confident white male, member of some privileged club and with a large country house. David is rather of medium build, there is no club, privileged or not, there is a country house, not big but very pleasant – we had many good meetings there, talking and eating fruit from his trees. I am very fond of David, and we talk often.

Recently, the definition of derivative for the set-valued functions and the differentiation of sets were introduced and some of their properties were investigated in your work. What was your motivation for introducing such concepts in differential geometry?

Frankly, I was almost forced into this direction through my interest in the change-set problem. One version of the change-set problem is like this: you have a point process, given by random locations $X_1, \dots, X_n$, at which you observe random marks $Y_1, \dots, Y_n$; the distribution of these marks is some

Fig 6. *David Vere-Jones in Estate Khmaladze's home in Wellington, October 2012.*

P_K if the location belongs to the set K and is some "grey level" distribution P_0 if X_i is outside K. For example, K is a pollution site and Y is the concentration of the pollutant. The set K is unknown and is the parameter of interest.

Considerable work had been done on the estimation of the change-set, but surprisingly little was done, practically nothing, about testing hypotheses concerning K. We can understand why if we consider the local situation: there is a family of sets K_t, such that as $t \to 0$ this K_t converges to K and we wish to test for K against K_t, when the number of observations n increases and K_t converges to K at the same time. What is this K_t? It is a set-valued function, continuous, say, in Hausdorff metric, at least at K, and it describes a possible small deviation from K in some direction. But then, in what direction? And this "direction" should be a derivative of K_t in t, right? But such a notion did not exist at the time.

Shouldn't we be interested, in principle, in many different directions? That is, in many different set-valued functions, which are like layers of an onion, enveloping the core K.

For a number of years I was just talking to my friends, like John Einmahl, and like my old friend and co-author Robert Mnatsakanov, already in about 1997-98. Very soon the local Steiner formula came into the view, and the

Fig 7. *Estate Khmaladze (in yellow), little Mariam and Wolfgang Weil (in blue), meeting truant sheep in Makara, near Wellington, in 2003*

book of Rolf Schneider on *Brunn – Minkowski Theory* became favorite reading, but mostly, as I said, it was talk. you talk, and talk, and somehow it helps: you convince yourself that what you want to do is natural.

The work really started in February 2004, in a small room in the Institute of Mathematics in Karlsruhe which served as a little extra library. With Wolfgang Weil we had just finished work on the asymptotic theory for local empirical processes in shrinking neighborhoods of the boundary ∂K. This neighborhood is where all of the symmetric differences $K_t \Delta K$ would live. Each such symmetric difference would shrink and disappear somewhere in the boundary, but where would the traces live? On the boundary? – No, they cannot. They live, we said, on what we called a normal cylinder $\partial K \times \mathbb{R}$, which is based on the boundary ∂K. We had a concept of the "local magnification map," which mapped the shrinking neighborhood $(\partial K)_\varepsilon$ onto this normal cylinder. So, the Borel σ-algebra of these shrinking sets was mapped on the Borel σ-algebra of "stable" sets on the cylinder.

Mapping of σ-algebra onto σ-algebra was all very good, but in 2004 we were not able to say what the limits of symmetric differences $K_t \Delta K$ were for any particular set-valued function. To say a bit more, we had functional limit theorems for local point process on $(\partial K)_\varepsilon$, but we did not have a one-dimensional limit theorem. I had never seen such a situation before.

There in Karlsruhe, Wolfgang resisted my attempt to drag him into the work on actual differentiation. He brought to me the book of J-P. Aubin and M. Frankovska on *Set-valued Analysis* and said apologetically: “You see, there are already two chapters on differentiation.”

Indeed there were. Suggested by convex problems, the derivatives were mostly understood as tangent cones to the graph of the set-valued function. A very fruitful and beautiful concept. In research publications, derivatives were also understood as affine (semi-affine, quasi-affine) mappings, if they approximated K_t well enough. The works of Zvi Artstein and other geometers would tell you much in this direction. Another approach, which looked so very natural, was to use the indicator function of the set $K_t \Delta K$. Certainly it should converge, after division by t, to a generalized function on ∂K. Why not use this generalized function as a derivative? However, it would not be good either. It would be too coarse a language: many, actually, infinitely many, set-valued functions, which for our statistical purposes we would need to distinguish, would lead to the same generalized function.

The work started in earnest later in Wellington and the findings were reported in Khmaladze (2007). I was very proud of myself, saying it is not every day a statistical problem is developed into a new result in geometry or analysis. But one should not be too proud – it is not us who create what we publish; it existed in the body of mathematics already; we only discover it.

I must add I was very lucky to have Lucy Kozeratska, from Edmonton, visiting Wellington for 2-3 weeks. Her interests have been in convex analysis and optimization, and she was great help for me in literature search and its evaluation.

Later we proved a Gaussian limit theorem for local empirical processes on $(\partial K)_\varepsilon$ with John Einmahl and essentially extended the notion of the derivative with Wolfgang Weil – we can now differentiate in the neighborhood of a bounded compact set and we can let it split and bifurcate.

What led you to your recent work on unitary transformations and goodness of fit testing for discrete and continuous distributions?

It was Ritei Shibata who asked me during his visit to Wellington why is it that the theory of goodness of fit tests for continuous distributions was so diverse, with so many different tests, while for the discrete distributions we have only one test, the Pearson chi-square goodness of fit test, if we do not count others which are asymptotically equivalent to it.

Of course, I was tempted to answer quickly, with the banality of “usual” explanations. But Shibata was not in a hurry and was not going to press me for the answer or comments. So, I had time to understand why the situation really was as it was. It may be that we associate the expression “distribution-free" too strongly with the time transformation $t = F(x)$. Disjoint events and their probabilities can be defined in any probability space. There is no hope that anything specific like time transformations can help where the notion of time is absent.

The main fact behind the contents of the paper Khmaladze (2013) is quite simple; one can even say it is the same fact which makes the Pearson chi-square statistics asymptotically distribution-free. One only looks at it slightly differently. Let the vector of probabilities $p = (p_i)_{i=1}^m$ denote an m-dimensional discrete distribution. Consider the corresponding "components" of the chi-square statistics $Y_{in} = (\nu_{in} - np_i)/\sqrt{np_i}$. Now let $X = (X_i)_{i=1}^m$ be the vector of independent standard normal random variables – a very homogeneous object – and let us consider its projection parallel to the vector, which we denote somewhat wrongly $\sqrt{p}$:

$$Y_p = X - \langle\sqrt{p}, X\rangle\sqrt{p}, \quad \sqrt{p} = (\sqrt{p_i})_{i=1}^m.$$

The fact is, that this vector Y is the weak limit of the vector $Y_n = (Y_{in})_{i=1}^m$. Therefore, in the problem of testing p to be true distribution, we will end-up with Y_p, but in the problem of testing for another distribution q we will end up with Y_q, and both are projections. But if so, one can then map one projection into another, and therefore, map a problem of testing p into a problem of testing q, and map both to the problem of testing yet another r. And one can choose this r in any way one wants, and make it standard; for example, one can choose it to be the uniform m-dimensional distribution.

Such a clean and almost obvious point of view. Makes you feel as if you stole something from others. But if I did, I stole it from myself as well, because I also was thinking all my life that no other distribution-free test, I should say – no other "sensible" distribution-free test, except chi-square, exists. Now we have a whole class: rotate a Y_p into Y_r and take any functional from Y_r as a test statistic. If I may say this, the whole work showed that there can be surprisingly strong inertia in our thinking.

When I said "map one projection into another" I certainly meant a unitary transformation of Y_p to Y_q. Extension of this to the empirical processes in continuous time suddenly brought strange results, quite unexpected for me. Denote $v_F(\phi), \phi \in L_2(F)$, the function parametric F-Brownian bridge. Then what I am talking about is the unitary transformation of this process, defined as

$$U^* v(\phi) = v_F(U\phi),$$

where U is the unitary operator on $L_2(F)$. Even without telling you any specific result, which comes from this construction, as soon as we have the family of random linear functionals, $v_F(\phi)$ in ϕ, shouldn't there be a linear operator nearby? Shouldn't we do something with these ϕ's?

Without providing too much detail, what actually came out of this work, is that we need to recognize the existence of a huge class of Brownian bridges, different from what we know as the F-Brownian bridge, many of them looking quite unusual, and the unitary transformations of any one of them to any other. Thus, statistical testing problems for distributions in $\mathbb{R}^d$ can be mapped into one another. Since this is an interview and one can speak somewhat loosely – it is an illusion that we have many different testing problems, for different distributions F, although connected through a common general approach. What we actually have is one single problem. I exaggerate, but there is some truth in this exaggeration. As one particular result, give me the F-Brownian bridge with F a continuous distribution in $\mathbb{R}^d$ with a rectangular support, and I will give you the standard Brownian bridge on $[0, 1]^d$. Well, I better not go into this more and hope that the paper, accepted in 2014, will be eventually published. [1]

You recently discovered a property of the Ornstein-Uhlenbeck process in connection with an application in financial mathematics that you thought was somewhat strange.

Oh, yes. I still don't understand it. Is the sum of two independent log-normal random variables again a log-normal random variable? You certainly will say "no." But try numerical simulations (*laughs*). This case again taught me that there is a theoretical truth and also a numerical truth, and they not necessarily are the same.

Sometime in 2008, the National Australia Bank asked me to evaluate their risk management methodology. I was not in Australia, but I was not too far away from Australia, and they knew I have interests in financial mathematics and that I am half-physicist in my soul. So, they asked me. What was prominent in their approach, Frishling and Lauder (2006), were integrals of the form $\int_0^T e^{S_t} dt$, where S_t was the Ornstein-Uhlenbeck process. In Brownrigg and Khmaladze (2011) we looked at the marginal distributions for these integrals, and to our great surprise these were log-normal. Could not be, but they really were. To be precise, the difference, between the exact distribution and its log-normal approximation was only in the third decimal place. Even if you have only two independent $\mathcal{N}(0, 1)$-random variables ξ_1 and ξ_2 and consider the sum $e^{\xi_1} + e^{\xi_2}$, you will discover that the distribution of this sum is surprisingly well approximated by a log-normal distribution.

At the beginning of your scientific career you were involved in interesting

[1] Now published as Khmaladze(2016)

projects in the area of human genetics and electrophysiology. Please, tell us about the problems you and your colleagues studied.

Yes, there was a paper, my first paper in statistics, which was published with my old friend, Rezo Chitashvili, together with a then young geneticist Teimuraz Lezhava. It suggested a model for what is known as association between the so-called acrocentric chromosomes in human somatic cells. The phenomena was later given great importance, because too many of these associations is a clear indicator of Down syndrome in the individual. Any maternity hospital will take blood samples from pregnant women and count the associations.

It was a terrible model, with so many combinatorial counts, but it did one thing: when we started there was a very good paper of J. O. Irvin, an outstanding British statistician of the previous generation, based on the data of the geneticist Patricia Jacobs. Lezhava was telling us that Jacobs was also very famous. What Irvin used as an elementary act was "association between two chromosomes." But what we said was that the association happens, actually, between satellites of the chromosomes, tiny protrusions from the short arms of the acrocentric chromosomes. This was a "structural" assumption about these associations, and changed the probability distribution of the associations. The model fitted well, but as I said it looked complicated. Lezhava then continued these studies, including gerontological aspects of it. You know, there are very old people in the mountain regions of Georgia. So, he studied them and compared with similar groups in other ethnithities. In Japan, in particular.

Multinomial schemes with the so-called LNRE property were introduced and classified in your 1987 CWI report. In one of your latest papers the concept of diversity and fragmentation were introduced and studied. Are there connections between these works? What kind of practical problems can be solved within the framework of such schemes?

LNRE stands for "large number of rare events." I invented this as an alternative to the term which was much in use before the late 70's and early 80's: "relatively small samples." It was often abbreviated to "small samples," which was somewhat misleading. The sample sizes people had in mind were, actually, quite large, hundred of thousands or more, but the number of different outcomes, or events, was also quite large. So the sample size, large in itself, was not sufficient to allow each event to be seen many times. So, relative to the number of different events it was small. To call this situation "small samples" was not good, I think.

Regarding the phrase "theory of LNRE" – I am not using it often because I am only moving towards this theory, quite slowly – I am tempted to replace LNRE by "diversity." Mathematically the situation can be viewed as a triangular array of occupation problems. For each n we have $N = N_n$ boxes in which we throw, independently, n balls with probabilities $p_n = \{p_{1n}, \ldots, p_{Nn}\}$. As $n \to \infty$, N_n also tends to ∞, while all $p_{in} \to 0$. One can imagine partitioning the interval $[0,1]$ into a large number of small disjoint intervals in zillions of ways, and this will produce our vector of probabilities p_n. In many mathematical constructions, based on these partitions, the final result does not depend on the partitions as soon as, say, $\max_i p_{in} \to 0$. However, within our occupancy problems, a whole world of possibilities opens up, and for different partitions very different things can happen. Some examples can be seen in niche allocation models as in, e.g., Magurran (2010).

One question is how do we treat the individual frequencies of balls in different boxes? We treat them as a statistical ensemble, as in statistical physics one treats different particles – not individually, but by grouping particles according to, say, their impulses and coordinates. We too, count the number of boxes with frequencies equal any given k, regardless of what boxes they are. One needs classification of different triangular arrays, and this is what was done in my 1987 Report. It was strange to see that in earlier classifications the arrays which led to, say, Zipf's Law were absent and ignored.

FIG 8. *Estate with daughter Mariam (left) and wife Mzia in Wellington.*

The classification I am talking about was given in terms of various forms of regularity observed in the frequencies. What should be the corresponding behavior of the underlying probabilities needs to be investigated. Of course this behavior is different from what we see in frequencies. Necessary and sufficient conditions were given in the report, and now I think this can be extended to cover other interesting cases. There is still a large space, open for further research. If I have time enough, maybe I will manage to show that, within LNRE, the expected values $\{np_{1n}, \ldots, np_{Nn}\}$ should behave as a triangular array of asymptotically negligible random variables, studied as infinite divisibility. This follows from one result in the fragmentation paper Khmaladze (2011): if the probabilities p_n were created through the fragmentation process, then

$$R_n(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(np_{in} > z)} \to 0, \text{ for all } z > 0.$$

The result was of technical use, but it gives you a hint: I am not saying that np_{iN}-s are random, and I am not saying we will sum them up (*laughs*). But why can't we nonetheless use the apparatus of infinite divisibility?

What was your motivation in writing the book Statistical Methods with Applications to Demography and Life Insurance?

I have a feeling we are tired and need some refreshment. (*laughs*). Let us do something about it and let the book look after itself.

Addendum, from Estate. Looking back on this text, I see with some surprise that many of my life-long friends and good colleagues are not named here. I regret this, but I think I know how this happened: the questions have been mostly about my research, as they are supposed to be, and not about my life, while my friends have not been necessarily my co-authors. Maybe, they should have been.

References

Akritas, M. G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.

Artstein, Z. (1995) A calculus of set-valued maps and set-valued evolution equations, *Set-Valued Anal.*, **3**, 213-261

Aubin, J.-P. and Frankowska, H. (1980) *Set-valued analysis*, Birkhäuser.

Brownrigg, R. and Khmaladze, É. V. (2011). Strange facts about the marginal distributions of processes based on the Ornstein-Uhlenbeck process. *Risk: Journal of Computational Finance*, **15**, 105–119.

Chitashvili, R., Khmaladze, É. V. and Lezava, T. (1972). Use of the Mathematical Satellite Model of determining Frequencies of Association of Acrocentric Chromosomes Depending on Human Age. *Internat. J. Bio-Medical Computing*, **3, N3**, 181–199.

Durbin, J., Knott, M., and Taylor, C. C. (1975). Components of Cramr - von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B*, **37**, 216-237.

Einmahl, John H. J. and Khmaladze, É. V. (2011). Central limit theorems for local empirical processes near boundaries of sets. *Bernoulli*, **17**, 545–561.

Frishling, V. and Lauer, M. (2006) Global and Regional Stress Test Program 2006. *National Australia Bank,* Internal Working Paper.

Khmaladze, É. V. (1979). The use of Omega-square tests for testing parametric hypotheses. *Theory Probab. Appl.* **24**, 283–302.

Khmaladze, É. V. (1981). Martingale approach to nonparametric goodness of fit tests. *Theory Probab. Appl.* **26**, 240–258.

Khmaladze, É. V. (1983). Martingale limit theorems for divisible statistics. *Theory Probab. Appl.* **28**, 530–549.

Khmaladze, É. V. (1987). On the theory of large number of rare events. *CWI Report*, Amsterdam.

Khmaladze, É. V. (2007). Differentiation of sets in measure. *J. Math. Anal. Appl.* **334**, 1055–1072.

Khmaladze, É. V. (2011). Convergence properties in certain occupancy problems including the Karlin–Rouault law. *Journal of Applied Probability*, **48**, 1095–1113

Khmaladze, É. V. (2013). Note on distribution free testing for discrete distributions. *Ann. Statist.* **41**, 2979–2993.

Khmaladze, É. V. (2013). *Statistical methods with application to demography and life insurance.* CRC Press (Chapman and Hall books).

Khmaladze, É. V. (2016). Unitary transformations, empirical processes and distribution free testing, *Bernoulli.* **22**, 563-588 (published online – 2014).

Khmaladze, É. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.*, **32**, 995–1034.

Khmaladze, É. V. and Koul, H. L. (2009), Goodness-of fit problem for errors in nonparametric regression: distribution free approach, The Annals of Statistics, 37, 1365-1385

Khmaladze, É. V. and Lezava, T. (1988a). Aneuploidy in Human Lymphocytes in Extreme Old Age. *Proc. Japan Acad.*, **64, B, N5**, 128–130.

Khmaladze, É. V. and Lezava, T. (1988b). Characteristics of Cis- and Transorientation Chromatid Types of Association in Human Extreme Old Age. *Proc. Japan Acad.*, **64, B, N5**, 131–134.

Khmaladze, É. V. and Weil, W. (2008). Local empirical processes near boundaries of convex bodies. *Ann. Inst. Statist. Math.* **60**, 813–842.

Khmaladze, É. V. and Weil, W. (2014). Differentiation of sets–the general case. *J. Math. Anal. Appl.* **413**, 291–310.

Koenker, R. and Xiao, Z. (2002). Inference on the Quantile Regression Process. *Econometrica*, **70**, 1583–1612.

Koning, A. J. (1992). Approximation of stochastic integrals with applications to goodness-of-fit tests. *Ann. Statist.* **20**, 428–454

Koning, A. J. (1994). Approximation of the basic martingale. *Ann. Statist.* **22**, 565–679

Magurran, A. (2004), *Measuring Biological Diversity*, Wiley–Blackwell.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semi-parametric regression. *Statist. Decisions*, **25**, 1–18.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.* **79**, 957–964.

Schneider, R. (1993) *Convex Bodies: the Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications, **44**, Cambridge University Press, Cambridge.

Hira L. Koul
Department of Statistics
and Probability
Michigan State University
East Lansing, MI 48824-1027
U. S. A.
E-mail: koul@stt.msu.edu

Roger Koenker
Department of Economics
University of Illinois
Urbana, IL 61801
U.S.A.
E-mail: rkoenker@uiuc.edu