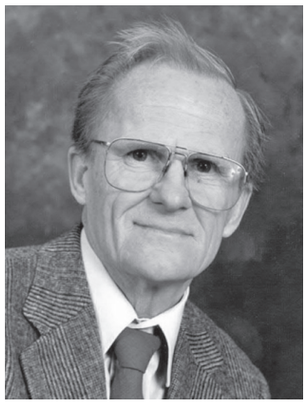
In Memoriam: Jorma J. Rissanen, 1932–2020

Jorma Johannes Rissanen, giant of information theory, control theory and statistics, passed away at age 87 after a long illness in Los Gatos, California, on 9 May 2020. Jorma Rissanen was perhaps best known for making arithmetic coding practical and for inventing the Minimum Description Length (MDL) principle. The importance of both contributions can hardly be overstated. Arithmetic coding is a central part of information theory; the MDL Principle has also had a profound influence on the data sciences: statistics, machine learning and data mining. It has played a central role in making information-theoretic approaches mainstream to those fields.



Born October 20, 1932, Rissanen grew up in Kemi, a small town near the Swedish border in the north-west of Finland. He moved on to study in Helsinki and obtained his Master's degree in electrical engineering in 1956, his Licentiate and Doctor of Technology degrees in control theory and mathematics in 1960 and 1965, respectively, all from the Helsinki University of Technology. Most of the doctorate work was done at a distance, while Rissanen worked for the IBM Nordic Laboratory in Stockholm, Sweden, which he joined in 1960. Eventually this led him to move to IBM's San Jose Research Laboratory in California in 1966. He stayed there until his retirement in 2002, except for the academic year 1973-1974 when he held the chair of control theory in Linköping University, Sweden. Until quite recently, Rissanen remained active in research as a professor emeritus of Tampere University of Technology, where he became part-time professor after receiving an Honory Doctorate degree in 1992.

While Rissanen's early work was mostly in control theory, he was given free hands at IBM to study and work on whatever he found interesting. Continuously expanding his research interests, he thus became, in his own words, a "lifelong professional student", passionate about learning and discovery. This freedom, in combination with Rissanen's highly original thinking and a serendipitous touch of luck connected to his stay in Sweden, led to an Annus Mirabilis: around 1975, Rissanen invented both practical arithmetic coding (first publication 1976) and the MDL Principle (first publication 1978) – in his mind, both concepts were in fact very closely connected. Ironically, though Rissanen did not enjoy his time in Sweden in 1973–74, it played a fundamental role in these discoveries. In his own words:

"[going to Sweden]…was a disastrous move. I found that I don't like the field of control, I don't like to be a professor, and I don't like the climate nor the at that time very socialist Sweden. However, something happened, which maybe could not have happened otherwise: I was exposed to the exciting ideas of Chaitin, Kolmogorov and Martin-Löf [on Kolmogorov complexity and algorithmic randomness], which set my mind in fire….I found that this is what interests me…We returned to IBM San Jose after just one year."

While Rissanen had already made fundamental contributions to control theory, system theory, database theory and numerical analysis, the focus now shifted and productivity accelerated: the years after 1975 saw a steady stream of ground-breaking articles and patents on data compression and modeling, containing one fundamental contribution after the other. What were these contributions about? The main idea of the MDL Principle is that all learning from data can be fruitfully cast in terms of data compression. This goes hand in hand with the philosophy that probability models in statistics should be viewed as *codes*, i.e. languages for describing patterns in the data, and one should never think of them as 'true' or 'generating the data' – as traditional, frequentist, statisticians do. Whereas researchers often do pay lip service to such statements, they then usually go on to design estimators that minimize expected losses, where expectations are taken under some distribution. Even though they interpret such procedures differently, such a path is also routinely followed by Bayesian statisticians – and Rissanen rejected the standard Bayesian approach just as much as the frequentist one. Instead, he derived and insisted on methods that have a clear interpretation *in terms of data only* – a prime example is Shtarkov's normalized maximum likelihood code, achieving an objective required to hold for all possible data sequences at the same time. As he wrote in the monograph *Stochastic Complexity and Statistical Inquiry* (1989),

"We never want to make the false assumption that the observed data actually were generated by a distribution of some kind, say Gaussian, and then go on to analyze the consequences and make further deductions. Our deductions may be entertaining, but quite irrelevant to the task at hand, namely, to learn useful properties from the data."

Once the principle – learning as compression – and the accompanying philosophy – probability models as languages - were in place, Rissanen continued to refine it and to apply it to a variety of problems in statistics, and the results started flowing. We stress that, while Rissanen's radical philosophy served as a continuous source of inspiration, the resulting learning algorithms were often shown to be optimal also in more traditional analyses – to use and enjoy them, one does not have to subscribe to the philosophy.

The first MDL paper (*Modeling by the Shortest Data Description*, 1978, Best Paper Award IFAC 1981) introduced the two-part code and its application to model selection, which remains the most important application area for MDL methods till this day. It included the celebrated $(k/2)\ \log n$ formula – formally equivalent to the BIC, formulated in the same year with a very different, Bayesian motivation by Schwartz. The two-part code turned out be a special case of an (individual sequence) *universal code*, and Rissanen soon found that other, more sophisticated universal codes can be used as well: in 1984 he introduced the predictive plug-in code, also known as the 'prequential plug-in code' given its relation to A.P. Dawid's ideas on 'prequential' statistics. In the same year, he established the general link between sequential prediction and universal coding and formulated one of his most important results, a lower bound on prediction error and codelength which he referred to as a `Grand Cramèr-Rao Bound' (*Universal Coding, Information, Prediction and Estimation*, 1994, IEEE IT Soc. Best Paper Award 1996). The Bayesian mixture universal code was introduced in 1986, jointly with the fundamental concept of *stochastic*

complexity as a measure of the inherent complexity of a statistical model. By now, a whole new theory had emerged, which was eloquently described in the 1989 monograph from which we cited above. The crowning achievement came in 1996 with the paper *Fisher information and stochastic complexity* in which Shtarkov's normalized maximum likelihood code was identified as the ideal universal code to use, and its properties were analyzed. In the mean time, the work on arithmetic coding and data compression was also expanded, resulting in, for example, the data compression algorithm *Context* (1983) that introduced what has later been called variable-length Markov chains.

This remarkable sequence of articles was soon to be followed by an equally remarkable sequence of honors, including aforementioned best paper awards, the IEEE 1993 Richard W. Hamming medal, an IEEE Information Society Golden Jubilee Award for Technical Innovation for the invention of arithmetic coding in 1998, and, in 2009, the most important award in information theory, the IEEE Claude E. Shannon Award. Numerous further honors include two honorary doctorates, several IBM awards and foreign membership of the Finnish Academy of Science and Letters.

As regards his personal life, Rissanen was married for 64 years to Riitta Rissanen (née Aberg), and in his free time he was a passionate fan of football (soccer). In the 1950s, he seriously contemplated a professional football career. He kept playing several times a week during noon breaks at IBM San José from 1966 all the way up until his retirement in 2002 – indeed, as a tribute to Jorma, plans are under way to organize a Jorma Rissanen Soccer Cup at the next ITA conference. Jorma, atypically for the modern scientist, had no patience at all for small talk, networking and the typical conference breaks or receptions with hundreds of people present – still he was very sociable and tremendously enjoyed time in restaurants and bars with much smaller groups of close academic friends. He was a most loyal and inspiring mentor for many younger researchers (including some of the undersigned), providing essential support in building their careers. Jorma Rissanen impressed all those who had the privilege of knowing him with his commitment to stay true to his values and pursue truth through science. He is indeed famous for numerous memorable anecdotes that reflect his intransigent attitude towards science. Some of the stories have been saved for future generations in the Festschrift collection that was compiled in 2008 to honor his 75th birthday. The quotation above on his time in Sweden, taken from an interview in the Festschrift is Vintage Jorma. We strongly encourage everybody who wants to see more of his highly refreshing directness and honesty to take a look at the interview and the entertaining quotes and recollections in the Festschrift. He will be missed.

Peter Grünwald, Petry Myllymäki, Teemu Roos, Ioan Tabus. The first quotation is from the interview *A Conversation with Jorma Rissanen by P. Huuhtanen, E. Liski and S. Puntanen, published in the Festschrift in Honor of Jorma Rissanen (P. Grünwald, P. Myllymäki, I. Tabus, M. Weinberger and B. Yu, eds.), Tampere 2008*.

President's Column *(continued from page 1)*