

# Distance Between Probability Distributions on Distinct Sets and Applications to Data Compression and Model Reduction

**Professor Mathukumalli Vidyasagar**  
Cecil & Ida Green Chair in Systems Biology Science  
Erik Jonsson School of Engineering & Computer Science  
The University of Texas at Dallas

Thursday, August 18, 2011, 11:00

Boğaziçi University  
Department of Electrical and Electronics Engineering  
Yorgo I Stefanopulos Meeting Lounge (KB 217)

*Contact: Mehmet Akar, Boğaziçi Üniversitesi*

## Abstract

Suppose we have a high-order probability distribution, with  $n$  components, and a low-order distribution, with  $m$  components. How can we define a "distance" between them? If both distributions have the same number of components, then defining a distance is easy, but not when we compare probability vectors with different dimensions. In this talk I will show how, by using the notion of Maximum Mutual Information (MMI), we can compute the distance between two distributions of possibly unequal dimensions. This approach can be used right now for data compression, by quantizing a random variable with many values into another with fewer values with minimum loss of information. The same approach can in principle be used to approximate high-order Markov chains by lower-order Markov chains, and high-resolution images by lower-resolution images. However, some mathematical problems need to be solved first. I will describe those problems.

**Mathukumalli Vidyasagar** was born in Guntur, India on September 29, 1947. He received the B.S., M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin in Madison, in 1965, 1967 and 1969 respectively. Between 1969 and 1989, he was a Professor of Electrical Engineering at various universities in the USA and Canada. His last overseas job was with the University of Waterloo, Waterloo, ON, Canada, where he served between 1980 and 1989. In 1989 he returned to India as the Director of the newly created Centre for Artificial Intelligence and Robotics (CAIR) in Bangalore, under the Ministry of Defence, Government of India. Between 1989 and 2000, he built up CAIR into a leading research laboratory with about 40 scientists and a total of about 85 persons, working in areas such as flight control, robotics, neural networks, and image processing. In 2000 he moved to the Indian private sector as an Executive Vice President of India's largest software company, Tata Consultancy Services. In the city of Hyderabad, he created the Advanced Technology Center, an industrial R&D laboratory of around 80 engineers, working in areas such as computational biology, quantitative finance, e-security, identity management, and open source software to support Indian languages. In 2009 he retired from TCS at the age of 62, and joined the Erik Jonsson School of Engineering & Computer Science at the University of Texas at Dallas, as a Cecil & Ida Green Chair in Systems Biology Science. In March 2010 he was named as the Founding Head of the newly created Bioengineering Department. His current research interests are in the application of stochastic processes and stochastic modeling to problems in computational biology, control systems and quantitative finance.