LARGE DEVIATIONS

Frank den Hollander

Leiden University & EURANDOM
The Netherlands

Stieltjes afternoon
VU Amsterdam, 5 March 2009

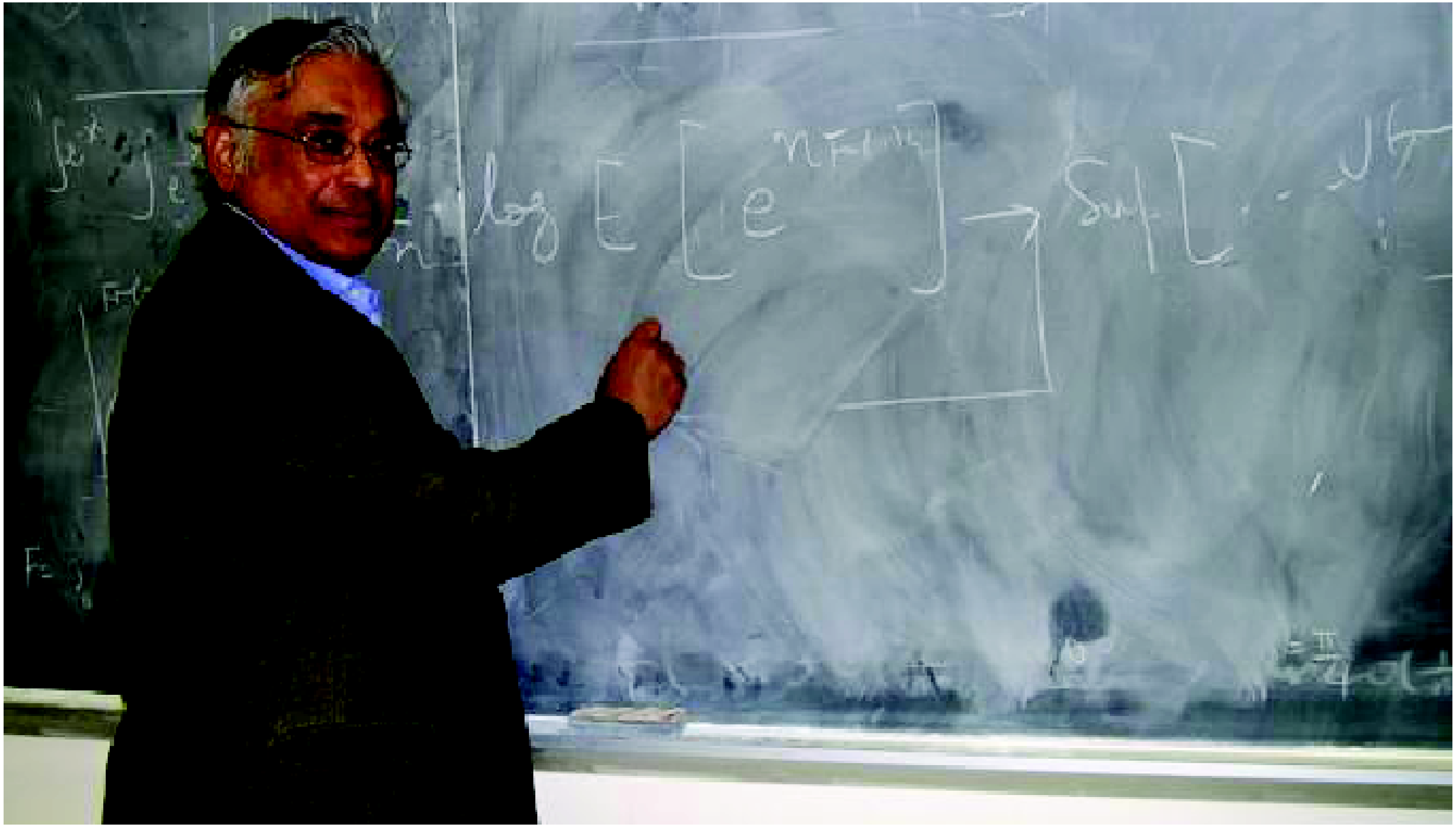
§ ABEL PRIZE 2007

On 22 March 2007 the Norwegian Academy of Sciences announced that the Abel Prize 2007 had been awarded to Srinivasa Varadhan (Courant Institute, New York)

> "... for his fundamental contributions to probability theory, in particular, for the creation of a unified theory of large deviations."

The jury characterized Varadhan's work as being

> "... of great conceptual strength and timeless beauty."



This afternoon offers three lectures that highlight three main directions in Varadhan's work:

• Large deviations	(F. den Hollander)
• Stochastic analysis	(J. van Neerven)
• Hydrodynamic scaling	(F. Redig)

In view of time, only brief sketches can be given.

F. den Hollander, Abelprijs 2007: S.R. Srinivasa Varadhan, Nieuw Archief voor Wiskunde 9 (2008) 192–197.

§ LARGE DEVIATION THEORY

The goal of large deviation theory is to decribe the probability that an empirical average of a large number of random variables deviates significantly from its theoretical average.

As such, large deviation theory provides a refinement of the

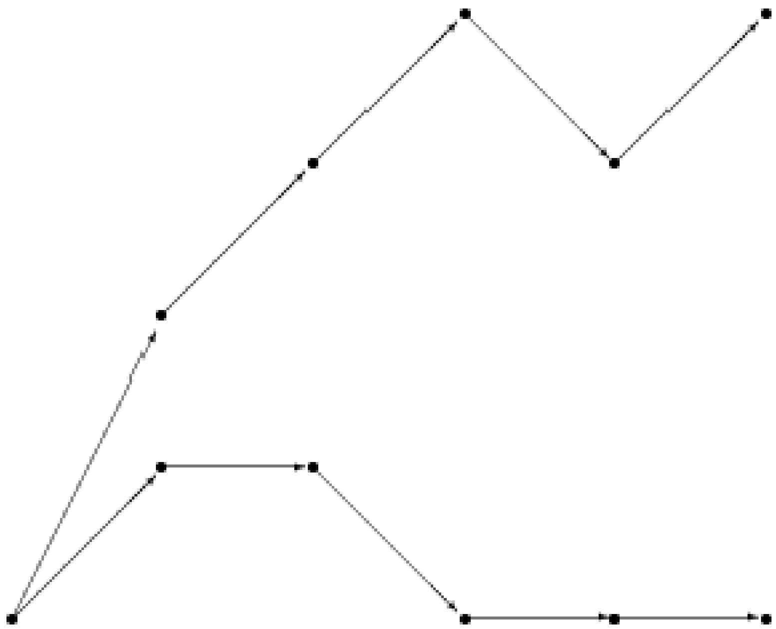
- law of large numbers
- central limit theorem

which are the workhorses of probability theory.

Historical roots of large deviation theory go back to:

- statistical physics
  Boltzmann, Gibbs, Einstein
- probability theory
  Cramér, Sanov
- statistics
  Chernoff, Bahadur
- information theory
  Shannon, Kolmogorov, Sinai

Varadhan is the chief architect of the modern theory.



A typical and an atypical trajectory of a random process.

§ KEY CONCEPTUAL IDEA

Each atypical trajectory of a random process is assigned a cost rate per unit of time. In some sense, this cost rate represents a distance between the trajectory under consideration and the typical trajectory of the random process.

Given the cost rates, the average growth rate of an exponential functional of the random process can be computed. This is done by maximizing "growth rate minus cost rate" over the set of all possible trajectories.

The outcome is an optimal trajectory that dominates the average growth rate of the functional. This optimality links large deviation theory to variational analysis.

Underlying motto:

> A large deviation is done in the least unlikely of all the unlikely ways !!!

§ A PARADIGM EXAMPLE

Let $X = (X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables taking the values 0 and 1 with probability $\frac{1}{2}$ each. Let $S_n = \sum_{i=1}^{n} X_i$, $n \in \mathbb{N}$, denote their partial sums.

- Law of large numbers:

$$\lim_{n\to\infty} \frac{1}{n} S_n = \frac{1}{2} \qquad \text{a.s.}$$

- Central limit theorem:

$$\lim_{n\to\infty} \frac{2}{\sqrt{n}}\left(S_n - \frac{1}{2}n\right) = N(0,1) \qquad \text{in distribution.}$$

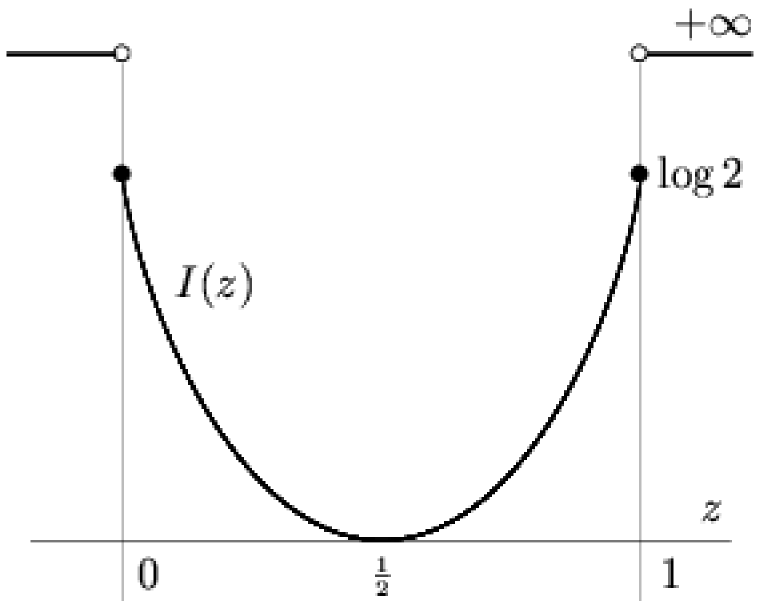
- Large deviation property:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n = \lfloor zn \rfloor) = -I(z), \qquad z \in [0,1],$$

with

$$I(z) = H\Big((1-z, z) \,\Big|\, \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\Big)$$

the Shannon relative entropy of the atypical distribution $(1-z, z)$ w.r.t. the typical distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$.



Graph of $z \mapsto I(z)$

Note that $\frac{1}{n}S_n$ is the empirical average of X after time n. Other functionals of X are interesting as well:

– Empirical distribution:

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \in \mathcal{P}(\{0,1\}).$$

– Empirical process:

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i,X_{i+1},X_{i+2},\ldots)} \in \mathcal{P}\Big(\{0,1\}^{\mathbb{N}}\Big).$$

Both exhibit large deviation behavior, with their own distinctive cost rate function.

Possible generalizations:

- Instead of $\{0,1\}$:
$\mathbb{R}$, $\mathcal{P}(\mathbb{R})$, $\mathcal{P}(\mathbb{R}^{\mathbb{N}})$, $\mathcal{P}(\mathcal{P}(\mathbb{R}))$, $\mathbb{Z}^d$, etc.

- Instead of $X = (X_i)_{i \in \mathbb{N}}$ i.i.d.:
Markov, stationary coding, Gibbs random field, etc.

- Instead of $H(\cdot \mid \cdot)$:
Legendre transform of the cumulant generating function, Dirichlet form of the Markov generator, Kolmogorov-Sinai entropy, etc.

Such generalizations are driven by applications.

§ TWO CORNER STONES OF LD-THEORY

(I) Large Deviation Principle:

A sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ on a Polish space $\mathcal{X}$ is said to satisfy the large deviation principle with rate function $I\colon \mathcal{X} \to [0,\infty]$ if

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -I(C) \qquad \forall\, C \subset \mathcal{X} \text{ closed},$$
$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -I(O) \qquad \forall\, O \subset \mathcal{X} \text{ open},$$

with

$$I(S) = \inf_{x\in S} I(x), \qquad S \subset \mathcal{X}.$$

The rate function I is required to have compact level sets and to not be $\equiv \infty$.

(II) Varadhan's Lemma:

Suppose that $(P_n)_{n\in\mathbb{N}}$ satisfies the LDP with rate function I. Then for any $F\colon \mathcal{X} \to \mathbb{R}$ that is bounded and continuous

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)}\, P_n(dx) = \sup_{x\in\mathcal{X}} [F(x) - I(x)].$$

§ PARADIGM EXAMPLE (continued):

Curie-Weiss model for attractive lattice gas

For the choice

$$\mathcal{X} = \mathbb{R},\ P_n(\cdot) = \mathbb{P}\Big(\frac{1}{n} S_n \in \cdot\ \Big),\ F(z) = \alpha z^2,\ \alpha \in (0,\infty),$$

the combination of LDP+VL yields

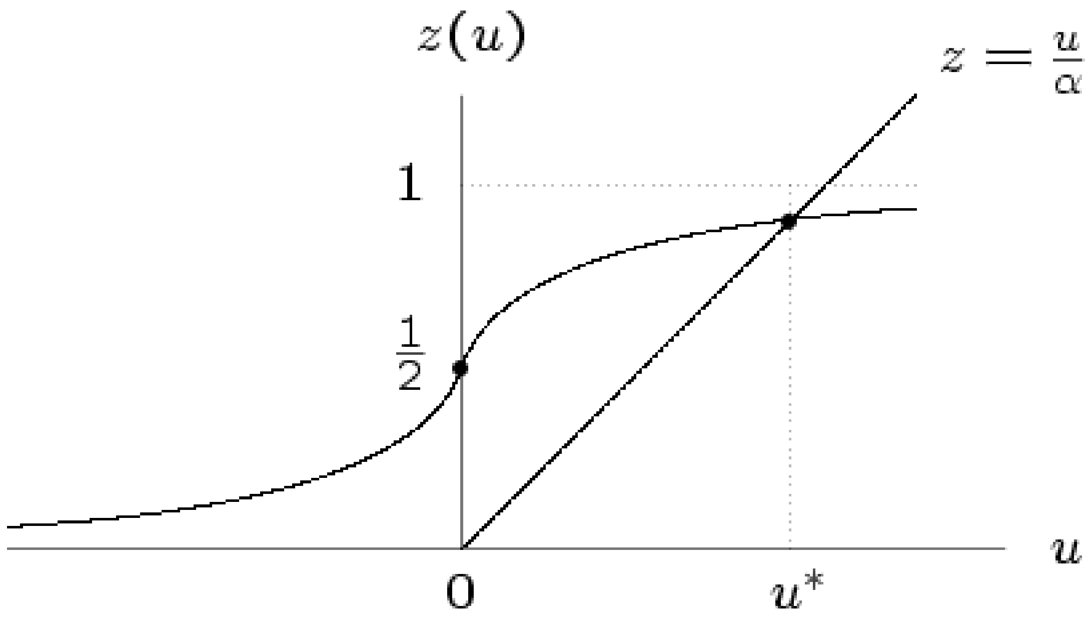
$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\Big(\exp\Big[\frac{\alpha}{n} \sum_{i,j=1}^{n} X_i X_j\Big]\Big) \\ &\quad = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\Big(\exp\Big[\alpha n \Big(\frac{S_n}{n}\Big)^2\Big]\Big) \\ &\quad = \sup_{z\in[0,1]} \Big\{ \alpha z^2 - H\Big((1-z,z)\ \Big|\ \big(\tfrac{1}{2},\tfrac{1}{2}\big)\Big)\Big\}. \end{aligned}$$

The unique maximizer is $z^* = z(u^*)$ with

$$z(u) = \tfrac{1}{2}(1 + \tanh u)$$

and u^* the solution of

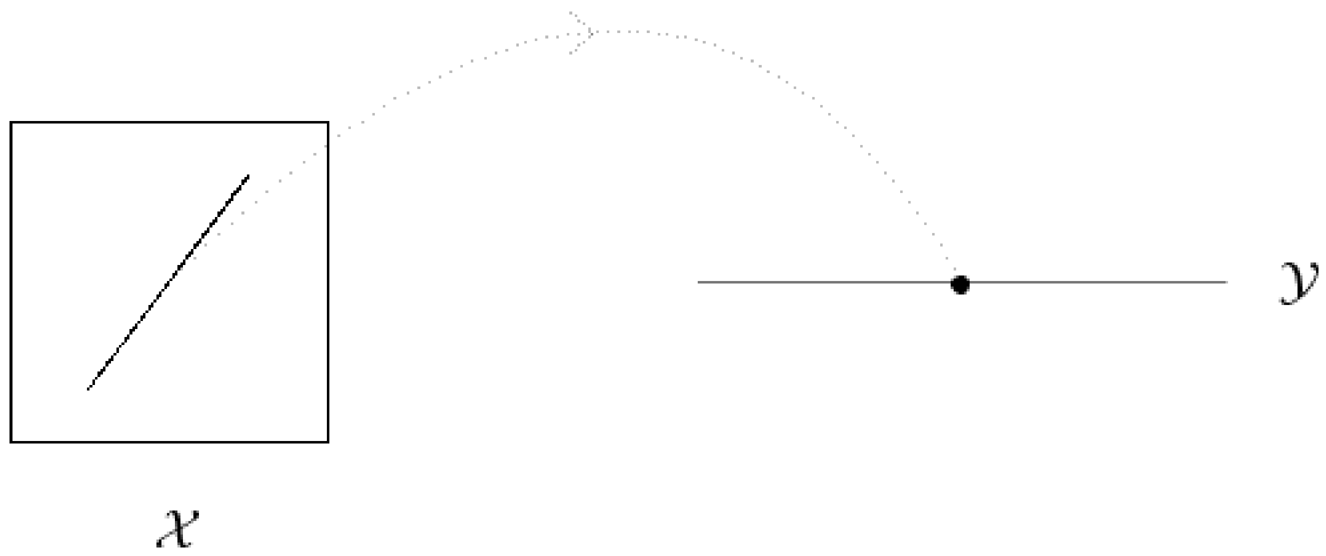
$$z(u) = \tfrac{u}{\alpha}.$$



Large deviation theory systematically builds up an arsenal of theorems based on LDP + VL.

Much of the theory deals with generating new LDP's from old LDP's, e.g. via

- contracting (= sum out information),
- reweighting (= tilt a probability measure),
- projecting (= take a limit of state spaces).



Contraction of the LDP from state space $\mathcal{X}$ to state space $\mathcal{Y}$.

Large deviation theory has been succesfully applied in:

- probability theory, statistics, operations research,
- information theory, ergodic theory, dynamical systems,
- computer science, statistical physics, biology,
- finance, insurance, logistics,
- ...

Rare events may have enormous consequences and therefore may dominate global behavior !!!