Ten Lectures on Small Value Probabilities and Applications

L1: Introduction, Overview and Applications

Wenbo V. Li

University of Delaware

wli@math.udel.edu

CBMS Lectures at UAH, June 4-8, 2012

We first define the small value (deviation) probability in several setting, which basically studies the asymptotic rate of approaching zero for rare events that *positive* random variables take smaller values. Benefits and differences of various formulations of small value probabilities are examined in details, together with connections to related fields.

Small Value Probability

Small value (deviation) probability studies the asymptotic rate of approaching zero for rare events that positive random variables take smaller values. To be more precise, let V_n be a sequence of *non-negative* random variables and suppose that some or all of the probabilities

$$\mathbb{P}\left(V_n \le \varepsilon_n\right), \quad \mathbb{P}\left(V_n \le C\right), \quad \mathbb{P}\left(V_n \le (1-\delta)\mathbb{E}\, V_n\right)$$

tend to zero as $n \to \infty$, for $\varepsilon_n \to 0$, some constant $C > 0$ and $0 < \delta \le 1$. Of course, they are all special cases of $\mathbb{P}\left(V_n \le h_n\right) \to 0$ for some function $h_n \ge 0$, but examples and applications given later show the benefits of the separated formulations.

Small Value Probability

Small value (deviation) probability studies the asymptotic rate of approaching zero for rare events that positive random variables take smaller values. To be more precise, let V_n be a sequence of *non-negative* random variables and suppose that some or all of the probabilities

$$\mathbb{P}\left(V_n \le \varepsilon_n\right), \quad \mathbb{P}\left(V_n \le C\right), \quad \mathbb{P}\left(V_n \le (1-\delta)\mathbb{E}\, V_n\right)$$

tend to zero as $n \to \infty$, for $\varepsilon_n \to 0$, some constant $C > 0$ and $0 < \delta \le 1$. Of course, they are all special cases of $\mathbb{P}\left(V_n \le h_n\right) \to 0$ for some function $h_n \ge 0$, but examples and applications given later show the benefits of the separated formulations.

•What is often an important and interesting problem is the determination of just how "rare" the event $\{V_n \le h_n\}$ is, that is, the study of the *small value (deviation) probabilities* of V_n associated with the sequence h_n.

Small Value Probability

Small value (deviation) probability studies the asymptotic rate of approaching zero for rare events that positive random variables take smaller values. To be more precise, let V_n be a sequence of *non-negative* random variables and suppose that some or all of the probabilities

$$\mathbb{P}\left(V_n \le \varepsilon_n\right), \quad \mathbb{P}\left(V_n \le C\right), \quad \mathbb{P}\left(V_n \le (1-\delta)\mathbb{E}\, V_n\right)$$

tend to zero as $n \to \infty$, for $\varepsilon_n \to 0$, some constant $C > 0$ and $0 < \delta \le 1$. Of course, they are all special cases of $\mathbb{P}\left(V_n \le h_n\right) \to 0$ for some function $h_n \ge 0$, but examples and applications given later show the benefits of the separated formulations.

•What is often an important and interesting problem is the determination of just how "rare" the event $\{V_n \le h_n\}$ is, that is, the study of the *small value (deviation) probabilities* of V_n associated with the sequence h_n.

•If $\varepsilon_n = \varepsilon$ and $V_n = \|X\|$, the norm of a random element X on a separable Banach space, then we are in the setting of small ball/deviation probabilities.

Deviations: Large vs Small

- Both are estimates of rare events and depend on one's point of view in certain problems.
- Large deviations deal with a class of sets rather than special sets. And results for special sets may not hold in general.
- Similar techniques can be used, such as exponential Chebychev's inequality, change of measure argument, isoperimetric inequalities, concentration of measure, etc.
- Second order behavior of certain large deviation estimates depends on small deviation type estimates.
- General theory for small deviations has been developed for Gaussian processes and measures.

Concentration of Product Measure

The concentration of measure phenomenon for the product measures has been investigated in depth by Talagrand (1995, 1996) in a remarkable way. His method has been applied to various interesting cases and produced very good concentration inequalities. However, his method is technically too complicated. Hence many people tried to simplify his proof and studied to find an alternative to reproduce and more ambitiously to extend his result.

•One of the successful alternatives is the entropy method. See Ledoux (1996), Massart (2000), Boucheron, Lugosi, Massart (2000, 2003, 2009), Maurer (2006), Kim, Ko and Lee (2006+, Entropy method for the left tail), etc.

Some technical difficulties between small and large values

- Let X and Y be two positive r.v's (not necessarily ind.). Then

$$\begin{aligned}\mathbb{P}(X+Y>t) &\geq \max(\mathbb{P}(X>t), \mathbb{P}(Y>t)) \\ \mathbb{P}(X+Y>t) &\leq \mathbb{P}(X>\delta t)+\mathbb{P}(Y>(1-\delta)t)\end{aligned}$$

but

$$?? \leq \mathbb{P}(X+Y\leq\varepsilon) \leq \min(\mathbb{P}(X\leq\varepsilon), \mathbb{P}(Y\leq\varepsilon))$$

- Moment estimates $a_n \leq \mathbb{E}\, X^n \leq b_n$ can be used for

$$\mathbb{E}\, e^{\lambda X} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\mathbb{E}\, X^n$$

but $\mathbb{E}\, \exp\{-\lambda X\}$ is harder to estimate.

- Exponential Tauberian theorem: Let V be a positive random variable. Then for $\alpha > 0$

$$\log \mathbb{P}(V\leq\varepsilon) \sim -C_V \varepsilon^{-\alpha} \quad \text{as} \quad \varepsilon \to 0^+$$

if and only if

$$\log \mathbb{E}\, \exp(-\lambda V) \sim -(1+\alpha)\alpha^{-\alpha/(1+\alpha)} C_V^{1/(1+\alpha)} \lambda^{\alpha/(1+\alpha)}$$

EX: Simple Random Walks

Let X_i, $i \geq 1$, be i.i.d. random variables with $\mathbb{E}\, X_i = 0$ and $\mathbb{E}\, X_i^2 = 1$, $\mathbb{E}\, \exp(t_0|X_1|) < \infty$ for $t_0 > 0$, and $S_n = \sum_{i=1}^n X_i$. Then as $n \to \infty$ and $x_n \to \infty$ with $x_n = o(\sqrt{n})$

$$\log \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |S_i| \geq x_n \right) \sim -\frac{1}{2} x_n^2$$

and as $n \to \infty$ and $\varepsilon_n \to 0$, $\sqrt{n}\varepsilon_n \to \infty$

$$\log \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |S_i| \leq \varepsilon_n \right) \sim -\frac{\pi^2}{8} \varepsilon_n^{-2}.$$

•**Open:** Find

$$\log \mathbb{P}\left(\max_{1 \leq i \leq n} |S_i| \leq C \right) \sim -??n.$$

Note that $?? \neq \pi^2/8$.

Ex: Let $L_\mu(n)$ be the length of the longest increasing subsequence (or records) in i.i.d sample $\{(X_i, Y_i)\}_{i=1}^n$ with law μ. Then

$$\lim_{n\to\infty} \frac{L_\mu(n)}{\sqrt{n}} = 2J_\mu.$$

The upper tail is known and for $c > 0$

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \log \mathbb{P}\left(L_\mu(n) > (2J_\mu + c)\sqrt{n}\right) = -U_\mu(c).$$

The lower tail is unknown in general, but for $0 < c < 2J_\mu$

$$\log \mathbb{P}\left(L_\mu(n) < (2J_\mu - c)\sqrt{n}\right) \approx -n.$$

See Deuschel and Zeitouni (1999), Aldous and Diaconis (1999), Okounkov (2000), and Li (2004+) for Gaussians.
•**Open:** Find

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(L_\mu(n) < (2J_\mu - c)\sqrt{n}\right).$$

There are some recent work on this direction.

EX: Related formulations for BM

•For one-dim Brownian motion $B(t)$ under the sup-norm, we have by scaling

$$\log \mathbb{P}\left(\sup_{0\le t\le 1}|B(t)|\le\varepsilon\right) = \log \mathbb{P}\left(\sup_{0\le t\le T}|B(t)|\le 1\right) = \log \mathbb{P}(\tau_2 \ge T$$

$$\sim -\frac{\pi^2}{8}\cdot T \sim -\frac{\pi^2}{8}\frac{1}{\varepsilon^2}$$

as $\varepsilon \to 0$ and $T = \varepsilon^{-2} \to \infty$. Here $\tau_2 = \inf\{s : |B(s)| \ge 1\}$ is the first two-sided exit (or passage) time.

•Lower tail and one sided exit time:

$$\mathbb{P}\left(\sup_{0\le t\le 1}B(t)\le\varepsilon\right) = \mathbb{P}\left(\sup_{0\le t\le T}B(t)\le 1\right) = \mathbb{P}(\tau_1 > T)$$

$$= \mathbb{P}(|B(T)|\le 1) \sim (2/\pi)^{1/2}T^{-1/2} = (2/\pi)^{1/2}\varepsilon$$

where $\tau_1 = \inf\{s : B(s) = 1\}$ is the one-sided exit time.

Some Formulations for General Processes

Let $X = (X_t)_{t \in T}$ be a real valued stochastic process (not necessary Gaussian) indexed by T.

The large deviation under the sup-norm:
$\mathbb{P}(\sup_{t \in T}(X_t - X_{t_0}) \geq \lambda)$ as $\lambda \to \infty$.

The small ball (deviation) probability: $\log \mathbb{P}(\|X\| \leq \varepsilon)$ as $\varepsilon \to 0$ for any norm $\|\cdot\|$.

The small ball probability under the sup-norm:
$\mathbb{P}(\sup_{t \in T} |X_t| \leq \varepsilon)$ as $\varepsilon \to 0$.

Two-sided exit problem: $\mathbb{P}(\sup_{t \in T} |X_t| \leq 1)$ as $|T| \to \infty$.

The lower tail probability: $\mathbb{P}(\sup_{t \in T}(X_t - X_{t_0}) \leq \varepsilon)$ as $\varepsilon \to 0$ with $t_0 \in T$ fixed.

One-sided exit problem: $\mathbb{P}(\sup_{t \in T}(X_t - X_{t_0}) \leq 1)$ as $|T| \to \infty$.

• The last two types of probability can also be viewed as the first exit time problems if the process has scaling property. If there is no scaling, then method of proofs are similar in many settings.

Exit Time, Principal Eigenvalue, Heat Equation

Let D be a smooth open (connected) domain in $\mathbb{R}^d$ and τ_D be the first exit time of a diffusion with generator A. For bounded domain D and strong elliptic operator A, by Feynman-Kac formula,

$$\lim_{t\to\infty} t^{-1}\log \mathbb{P}\left(\tau_D > t\right) = -\lambda_1(D)$$

where $\lambda_1(D) > 0$ is the principal eigenvalue of $-A$ in D with Dirichlet boundary condition.

Ex: Brownian motion in $\mathbb{R}^d$ with $A = \Delta/2$. Let $v(x,t) = \mathbb{P}_x\{\tau_D \geq t\}$ Then v solves $\begin{cases} \frac{\partial v}{\partial t} = \frac{1}{2}\Delta v & \text{in } D \\ v(x,0) = 1 & x \in D. \end{cases}$ So this type of results can be viewed as long time behavior of $\log v(x,t)$, which satisfies a nonlinear evolution equation.

• Unbounded domain D and/or degenerated differential operator. Li (2003), Lifshits and Shi (2003), van den Berg (2004), Kuelbs and Li (2004), Bañuelos and Carroll (2005), Bañuelos and K. Bogdan (2005), Bañuelos and DeBlassie (2006).

Conj: The general lower bound in Li (2003) is sharp.

The Wiener-Hopf Equation

The Wiener-Hopf equation

$$H(x) = \int_0^\infty f(x-y)H(y)dy, \quad x \geq 0$$

is still an active area of study, even the existence and uniqueness of a solution.

Spitzer (1956) has obtained a beautiful formula (Spitzer's identity) from which one can (in principle at least) calculate the joint distribution of any pair $(\max_{0\leq j\leq n} S_j, S_n)$ knowing the individual distributions of the first n partial sums, $S_0 = 0, S_k = X_1 + \cdots + X_k$. He then used it in Spitzer (1957, 1960a,b) to study the Wiener-Hopf equation. Here is a typical result.

Let $f(x)$ be the density of X, i,e, $F(x) = \int_{-\infty}^x f(t)dt$. If X is symmetric with characteristic function $\phi(\lambda)$, then

$$\lim_{n\to\infty} n^{1/2}\mathbb{P}(\max_{0\leq k\leq n} S_k \leq x) = \pi^{-1/2}H(x)$$

where $H(x)$ is the unique solution (in the class of functions that are non-decreasing, continuous on the right, with $H(0) > 0$) of the Wiener-Hopf equation

$$H(x) = \int_0^\infty f(x-y)H(y)dy$$

and $H(0^+) = 1$. In addition, the Laplace transform of $H(x)$ is given for $\lambda > 0$ by

$$\begin{aligned}\int_{0^-}^\infty e^{-\lambda x} dH(x) &= 1 + \int_{0^+}^\infty e^{-\lambda x} dH(x) \\ &= \exp\left\{-\frac{1}{2\pi}\int_{-\infty}^\infty \frac{\lambda}{\lambda^2+t^2}\log(1-\phi(t))dt\right\} < \infty.\end{aligned}$$

Moreover, if $\mathbb{E}\, X^2 = \sigma^2 < \infty$, then $H(x)$ has the asymptotic behavior

$$\lim_{x\to\infty} \frac{H(x)}{x} = \frac{\sqrt{2}}{\sigma}.$$

If the variance is infinite, then $H(x) = o(x)$ as $x \to \infty$.

• Li and C-H. Zhang (2010+): Purely probabilistic arguments with bounds on $\mathbb{P}(\max_{0\le k\le n} S_k \le x)$ and $H(x)$ under weaker moment conditions.

Hamiltonian and Partition Function

One of the basic quantity in various physical models is the associated Hamiltonian (energy function) H which is a nonnegative function. The asymptotic behavior of the partition function (normalizing constant) $\mathbb{E}\, e^{-\lambda H}$ for $\lambda > 0$ is of great interests and it is directly connected with the small value behavior $\mathbb{P}(H \leq \epsilon)$ for $\epsilon > 0$ under appropriate scaling.

In the one-dim Edwards model a Brownian path of length t receives a penalty $e^{-\beta H_t}$ where H_t is the self-intersection local time of the path and $\beta \in (0, \infty)$ is a parameter called the strength of self-repellence. In fact

$$H_t = \int_0^t \int_0^t \delta(W_u - W_v) du dv = \int_{-\infty}^{\infty} L^2(t, x) dx$$

It is known, see van der Hofstad, den Hollander and König (2002), that

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\, e^{-\beta H_t} = -a^* \beta_{2/3}$$

where $a^* \approx 2.19$ is given in terms of the principal eigenvalues of a one-parameter family of Sturm-Liouville operators. Bounds on a^* appeared in van der Hofstad (1998).

• Chen and Li (2012+): For the one-dim Edwards model, it is not hard to show

$$\lim_{\varepsilon\to 0} \varepsilon^{2/(p+1)} \log \mathbb{P}\{\int_{-\infty}^{\infty} L^p(1,x)dx \le \varepsilon\} = -c_p$$

for some unknown constant $c_p > 0$. Bounds on c_p can be given by using Gaussian techniques.

•Chen (2010), Chen and Rosinski (2011): Renormalization and asymptotics for physical models.

•Chen, Li, Rosinski and Shao (2011): Large deviations for local times and intersection local times

•There are close connections between SHE/KPZ (finite dimensional distributions) and intersection local times.

• Klenke and Morters (2005): Let $I_{m,n}(B(0,1))$ be the intersection local time of m vs n independent Brownian paths in $\mathbb{R}^d$ for $d = 2, 3$ inside the unit ball $B(0,1) \subset \mathbb{R}^d$. Then

$$\lim_{\varepsilon \to 0} \frac{\log \mathbb{P}(I_{m,n}(B(0,1)) \le \varepsilon)}{-\log \varepsilon} = -\frac{\xi_d(m,n)}{4-d}$$

where $\xi_d(m,n)$ are called the non-intersection exponents. The values $\xi_2(m,n)$ are found by Lawler, Schramm and Werner based on SLE. Much less is known in $\mathbb{R}^3$.

Hitting Probability of a Set

Let X_t, $t \geq 0$, be a fractional Brownian motion on $\mathbb{R}^d$ with index $0 < H < 1$. Then

$$\mathbb{P}(\inf_{1 \leq t \leq 2} |X_t| \leq \varepsilon) \begin{cases} \approx \varepsilon^{d-1/H} & \text{if } d > 1/H \\ > \delta & \text{if } d < 1/H \\ \approx (\log 1/\varepsilon)^{-??} & \text{if } d = 1/H \end{cases}$$

The motivations for extending results classical for Brownian motion to the fractional Brownian motions are not only the importance of these processes, but also the force to find proofs that relay upon general principles at a more fundamental level by moving away from crucial properties (such as the Markov property) of Brownian motion. Fractional Brownian motion might not be an object of central mathematical importance but abstract principles are.

Random graphs

Let $G(n,p)$ be a random graph and $\omega(G)$ denote the number of vertices in the maximum clique of the graph G.

Thm: For $k = o(\log n)$,

$$\mathbb{P}(\omega(G(n,1/2)) \leq k) = \exp(-n^{2+o(1)})$$

Note that a $o(1)$ in the hyper-exponent leaves lots of room! Also, It is not difficult to show that $\omega(G(n,1/2))$ is concentrated at $2\log_2 n$

Thm: Let $\mathbb{P}(X_{ij} = 0) = p = p_n, \quad \mathbb{P}(X_{ij} = 1) = 1 - p$ and

$$H_n \quad = \quad \sum_{1 \leq i < j < k < m \leq n} X_{ij} X_{jk} X_{km} X_{mi}$$

Then the probability of C_4-free is

$$\mathbb{P}(H_n = 0) \begin{cases} \to 1 & \text{if} p = O(n^{-1}) \\ \leq \text{poly. small} & \text{if} p = n^{-2/3} \\ \leq \text{exp. small} & \text{if} p = n^{-1/2} \end{cases}$$

Open: What is the correct cut off behavior?

Statistical Related Applications

- Gao, Li and Wellner (2010): How many Laplace transforms of probability measures are there? Related estimates for the bracket entropy in empirical processes theory are also studied.
- van der Vaart and van Zanten (2008a,b): Statistical applications for Gaussian priors based on Reproducing kernel Hilbert spaces.
- Nikitin and Pusev (2011): Refined estimates for weighted L_2-norm.
- Gine and Nickl (2012+): Non-parametric estimation, using lower bounds for small ball probabilities for m-th integrated BM.

Sequential Analysis

Several stopping times which arise from problems in approximations for sequential point and interval estimation may be written in the form

$$t_c = \inf\{n \geq m \colon S_n < cn^{\alpha} L(n)\},$$

where $S_n = X_1 + \cdots + X_n$, $X_1, X_2, \cdots$ are i.i.d. *positive* r.v's with $\mathbb{E}\, X_1^2 < \infty$, $L(n) = 1 + O(n^{-1})$, $\alpha > 1$, $m \geq 1$ and $c > 0$.

• The probability of stopped early

$$\mathbb{P}(t_c \leq (1-\delta)\mathbb{E}\, t_c) \sim K_{m,\delta} \cdot c^{(m-1)/2\alpha}, \quad c \to 0,$$

which is strongly influenced by the initial sample size m.

• The uniform integrability of $|t_c^*|^r$ in c is determined by the behavior of

$$\mathbb{P}(X_1 \leq x) \quad \text{or} \quad \mathbb{P}(S_m \leq x), \quad x \to 0$$

where $t_c^* = \frac{t_c - \mathbb{E}\, t_c}{\sqrt{\mathrm{Var}(t_c)}} \Rightarrow N(0, 2\alpha^2)$. See Robbins (1959), Chow and Robbins (1965), Starr and Woodroofe (1968, 1972), Woodroofe (1977, 1982), Lai and Siegmund (1977), Yu (1981), Takada (1992), etc.

Smoothness of the Density via Malliavin Matrix

Consider $F = (F^1, \cdots, F^m) : \Omega \to \mathbb{R}^m$ with $F^i \in D^{1,2}$. Then Malliavin Matrix of F is

$$\gamma_F = (\gamma_F^{ij}), \quad \gamma_F^{ij} = \langle DF^i, DF^j \rangle$$

Thm:(Bouleau-Hirsch) If $\det(\gamma_F) > 0$, a.s, then the law of F is absolute continuous.
Thm: (Malliavin) If (1) $F^i \in D^\infty$ and (2) $\mathbb{E}|\det \gamma_F|^{-p} < \infty$ for any $p > 0$, then F has a C^∞ density.
•The condition (ii) is called non-degeneracy for F.
•All these have been extended into theory of SDE and SPDE. It is curial to check the non-degeneracy condition which is small value probability.
•Mueller and Nualart (2008): Regularity of the density for the stochastic heat equation.
•Fei, Hu and Nualart (2011+): convergence of densities.
•Nualart (2010, book): Malliavin Calculus and its Applications.

Smoothness of the Density via Malliavin Matrix

Lemma: Let $M(\omega) = (m_{ij})_{n\times n}$ be a symmetric nonnegative definite random matrix with moments of all order for m_{ij}. If for any $p \geq 2$

$$\sup_{|v|=1} \mathbb{P}(v^T M v \leq \varepsilon) = O(\varepsilon^p), \quad \text{as} \quad \varepsilon \to 0^+,$$

Then $\det(M^{-1}) = (\det M)^{-1} \in L^p$ for all $p > 0$.

•In many applications of Malliavin calculus to the smoothness of the density of the solutions of SPDEs, one needs to show the inverse of the determinant of the Malliavin matrix has moments of all orders, or equivalently, the determinant of the Malliavin matrix has negative moments of all orders.

•In fact, the negative moments estimates

$$\mathbb{E}\, V^{-p} < \infty \quad \text{for any/all} \quad p > 0$$

is equivalent to the upper small value estimates

$$\mathbb{P}(V \leq \varepsilon) \leq C_p \varepsilon^p \quad \text{for any/all} \quad p > 0, \quad \text{as} \quad \varepsilon \to 0.$$

Determinant of Bernoulli Matrices

Let $M_n = (\pm 1)_{n\times n}$ be a random matrix whose entries are i.i.d. Bernoulli random variables with $\mathbb{P}(\pm 1) = 1/2$. This model of random matrices is of considerable interest in many areas, including combinatorics, theoretical computer science and mathematical physics. On the other hand, many basic questions concerning this model have been open for a long time.

•Hadamard's inequality: $|\det(M_n)| \leq n^{n/2}$, with equality if and only if M_n is an Hadamard matrix.

•For typical value, it is easy to show $\mathbb{E}\,(\det M_n)^2 = n!$ and one is led to conjecture that $|\det(M_n)|$ should be of $\sqrt{n!} = e^{-n+o(1)}n^{n/2}$ with high probability.

•Tao and Vu (2006):
$\mathbb{P}(|\det(\pm 1)| \leq \sqrt{n!}\exp(-29\sqrt{(}n\log n))) = o(1)$.

•Conj: $\mathbb{P}(|\det M_n| \leq (1-\delta)\sqrt{n!}) = o(1)$.

Determinant of Bernoulli Matrices: Singularity

For random Bernoulli matrices $(\pm 1)_{n\times n}$ with $\mathbb{P}(\pm 1) = 1/2$,

$$\begin{aligned} p_n = \mathbb{P}(s_{\min}(\pm 1) = 0) &= \mathbb{P}(|\det(\pm 1)| = 0) = \mathbb{P}(|\det(\pm 1)| < 2^{n-1}) \\ &\leq (c + o(1))^n. \end{aligned}$$

•It is easy to show $p_n \geq (1 + o(1))n^2/2^{n-1}$.
•Komlós (1967): $p_n = o(1)$.
•Kahn, Komlós and Szemerédi (1995): $c = 0.999$.
•Tao and Vu (2007): $c = 3/4$
•Bourgain, Wood and Vu (2010+): $c = 1/\sqrt{2}$.
Conj: $c = 1/2$.
•Yes for i.i.d Gaussian under the formulation $\mathbb{P}(|\det(g_{ij})| < 2^{n-1})$ by using the explicit distribution.
•Singularity: It is easy to show

$$\mathbb{P}(|\det M_n| = 0) \geq (1 + o(1))n^2/2^{n-1}.$$

The Smallest Singular Value

"The smallest singular value –the *hard edge* of the spectrum– is generally more difficult and less amenable to analysis by classical methods of random matrix theory than the largest singular value, the 'soft edge'. The difficulty especially manifests itself for square matrices or almost square matrices." —-Rudelson and Vershynin (2010), Proceedings of ICM.

•Edelman (1988) and Szarek (1991): For i.i.d $N(0,1)$,

$$\mathbb{P}\left(\inf_{x\in S^{n-1}} |\Gamma x| \leq c\varepsilon n^{-1/2}\right) \leq C\varepsilon.$$

•Litvak, Pajor, Rudelson, Tomczak-Jaegermann (2005): For $N = (1+\delta)n$ for some $\delta > 0$,

$$\mathbb{P}\left(s_n(\Gamma_{N\times n}) \leq c_1 N\right) \leq \exp(-c_2 N)$$

where γ_{ij} are i.i.d symmetric r.v's with certain assumptions on moments.

•Adamczak, Guedon, Litvak, Pajor, Tomczak-Jaegermann (08, 12):

$$\mathbb{P}\left(\inf_{x\in S^{n-1}}|\Gamma x|\le c\varepsilon n^{-1/2}\right)\le C\min(n\varepsilon,\varepsilon+e^{-c\sqrt{n}})\le C'\varepsilon|\log\varepsilon|^2$$

where Γ is an $n\times n$ matrix with independent columns drawn from an isotropic log-concave probability measure.
•**Conj:** (AGLPT12) $\mathbb{P}\left(\inf_{x\in S^{n-1}}|\Gamma x|\le c\varepsilon n^{-1/2}\right)\le C\varepsilon$.
•Rudelson and Vershynin (08, 10): For i.i.d (mean zero and variance one) subgaussian matrix $A_{n\times n}$,

$$\mathbb{P}(s_{\min}(A)\le\varepsilon n^{-1/2})\le C\varepsilon+c^n,\quad\varepsilon\ge 0$$

where constants $C>0$ and $0<c<1$ depend only on the subgaussian moment of the entries.
•Part of the arguments in the above remarkable (and useful) result are based on covering numbers or ε-net for spare or peaked sets.
•Litvak and Rivasplata (2012): Smallest singular value of sparse random matrices.
•Open: Under suitable condition, find relative simple way to show $\mathbb{P}(s_{\min}(A)\le n^{-logn})=o(1)$.

SVP: Gaussians vs Rademachers

•Three Permutation Problem: Let σ_i, $i = 1, 2, 3$ be three permutations on $[n]$. Is it true that

$$\mathbb{P}\left(\max_{1\le i\le 3}\max_{1\le m\le n}\left|\sum_{j=1}^{m}\varepsilon_{\sigma_i(j)}\right| \le c\right) \ge \frac{1}{2^n}$$

for some absolute constant c, independent of n?

Yes for $N(0,1)$ by the weak Gaussian correlation inequality, but **No** for $\varepsilon_i = \pm 1$, Alantha Newman and Aleksandar Nikolov (2011+).

•Key point: There is no universality in general.

•Assume $\sum_{i=1}^{n} u_i^2 = 1$. Is it true that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\varepsilon_i u_i\right| \le 1\right) \ge \frac{1}{2}??$$

Yes for $N(0,1)$, but **unknown** for $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$. The best known lower bound is $3/8$.

Littlewood-Offord Type Problems

Let $a_1, a_2, \cdots, a_n$ be vectors in $\mathbb{R}^d$ with $|a_i| \geq 1$ for all i. Let η_i be i.i.d r.v's with $\mathbb{P}(\eta_i = 0) = \mathbb{P}(\eta_i = 1) = 1/2$. Then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \eta_i a_i\right| \leq D\right) \leq \frac{c}{\sqrt{n}}.$$

where $D \geq 1$ is a given constant and c depends only on D.
A considerable literature has been devoted to this problem, beginning with Erdos (1945). A variety of tools from extremal set theory and geometry has been used, see Kleitman (1970), Griggs (1980), Frankl and Füredi (1988).
•Inverse Littlewood-Offord type problems and applications to singularity of random symmetric matrices, see Tao and Vu (2006, book), Nguyen and Vu (2011), Nguyen (2011).

Slicing the Cube

A cube of dimension n and side 1 is cut by a hyperplane of dimension $n-1$ through its center. The usual $n-1$ measure of the intersection is bounded between 1 and $\sqrt{2}$. Hensley (1979) and Ball (1988).

Thm: Let U_j be i.i.d uniform on $[-a, a]$. Then for any vector $v = (v_1, \cdots, v_n) \in \mathbb{R}^n$ with $|v| = \left(\sum_{j=1}^n v_j^2\right)^{1/2}$

$$\frac{1}{(1+a^2|v|^2)^{1/2}} \leq \mathbb{P}(|\sum_{j=1}^n v_j U_j| \leq 1) \leq \frac{\sqrt{2}}{(1+a^2|v|^2)^{1/2}}$$

Open: Sharp bounds for ε_k or general symmetric X_j with $\mathbb{E}\, X_j^2 = 1$

Combinatorial Discrepancy

Let $(V, \mathcal{F})$ be a set system, where $V = \{1, \cdots, n\}$. Such a combinatorial structure is often called a *hypergraph*. The discrepancy of a set system $\mathcal{F} \subset 2^V$ is

$$\mathrm{disc}(\mathcal{F}) = \min_{\chi} \max_{A \in \mathcal{F}} \left| \sum_{a \in A} \chi(a) \right|$$

where χ ranges over "two-colorings" $\chi : V \to \{-1, +1\}$.

Thm: Any set system $(V, \mathcal{F})$ such that $|V| = |\mathcal{F}| = n$ has $O(\sqrt{n})$ discrepancy. Some set systems have a matching lower bound. Equivalently,

$$\mathbb{P}\left(\max_{F \in \mathcal{F}} \left| \sum_{v \in F} \varepsilon_v \right| \leq C\sqrt{n} \right) \geq \frac{1}{2^n}$$

and

$$\mathbb{P}\left(\max_{F \in \mathcal{F}} \left| \sum_{v \in F} \varepsilon_v \right| \leq c\sqrt{n} \right) = 0 < \frac{1}{2^n}$$

for some constants $C > c > 0$.

Beck-Fiala **Conj.** (1981): $\mathrm{disc}(\mathcal{F}) \le Ct^{1/2}$ if $|\mathcal{F}| \le t$.
Thm: Let $A = (a_{ij})$, where $a_{ij} = 0$ or 1, be a matrix of size $n \times n$. Then for some $C > 0$

$$\mathbb{P}\left(\max_{1\le m\le n}\max_{1\le k\le n}\left|\sum_{i=1}^{m}\sum_{j=1}^{k} a_{ij}\varepsilon_{ij}\right| \le C(\log n)^4\right) \ge \frac{1}{2^n}.$$

• The Beck-Fiala conjecture implies $C(\log n)^3$ bound.
• There is a lower bound of $\Omega(\log n)$ given in Beck (1981).
Open: Find the correct order of the lower 'cut off' function.
•For more details, see Alon and Spencer: The Probabilistic Method (2000).
Open: Gaussian version.

The k Permutation Conjecture

Let σ_i, $1 \le i \le k$ be k-permutations on $[n]$. Then

$$\mathbb{P}\left(\max_{1\le m\le k}\max_{1\le j\le n}\left|\sum_{i=1}^{j}\varepsilon_{\sigma_m(i)}\right| \le c\sqrt{k}\right) \ge \frac{1}{2^n}$$

for some absolute constant c, independent of n.

The k-permutation Conj: disc$= \Omega(\sqrt{k})$.

- disc$\ge c\sqrt{k}$ via Hadamard matrix.
- disc$\le C(k\log n)$ via the Partial coloring lemma.
- disc$\le C(\sqrt{k}\log n)$ via the entropy method.

Gaussian k-permutation Conj: Under the Gaussian correlation conj. and $h_k = \Omega(\sqrt{k})$.

$$\begin{aligned}\mathbb{P}\left(\max_{1\le m\le k}\max_{1\le j\le n}\left|\sum_{i=1}^{j}\xi_{\sigma_m(i)}\right| \le h_k\right) &\ge \prod_{m=1}^{k}\mathbb{P}\left(\max_{1\le j\le n}\left|\sum_{i=1}^{j}\xi_{\sigma_m(i)}\right| \le h_k\right)\\ &\ge \prod_{m=1}^{k}\exp(-cn/h_k^2) \ge \frac{1}{2^n}\end{aligned}$$

- $h_k = \Omega(k)$ via the weaker Gaussian correlation inequality.

Balancing vectors

Consider an arbitrary pair of symmetric convex bodies, U and V in $\mathbb{R}^n$. Define $\beta(U, V) = \beta_n(U, V)$ as the smallest $r > 0$ satisfying the following: for every sequence $u_1, \cdots, u_n$ of vectors in $U \subset \mathbb{R}^n$ there exists a choice of signs $\varepsilon_1, \cdots, \varepsilon_n = \pm 1$ such that $\sum_{i=1}^n \varepsilon_i u_i \in rV$. And similarly, define $\alpha(U, V) = \alpha_n(U, V)$ as the smallest $r > 0$ such that $\sum_{i=1}^m \varepsilon_i u_i \in rV$ for all $1 \leq m \leq n$.

Clearly, $\beta_n(U, V) \leq \alpha_n(U, V)$.

Reformulation: $\beta_n(U, V)$ is the smallest $r > 0$ such that for any $u_i \in U$, $1 \leq i \leq n$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \varepsilon_i u_i\right\|_V \leq r\right) \geq \frac{1}{2^n}$$

and $\alpha_n(U, V)$ is the smallest $r > 0$ such that for any $u_i \in U$, $1 \leq i \leq n$,

$$\mathbb{P}\left(\max_{1 \leq m \leq n}\left\|\sum_{i=1}^m \varepsilon_i u_i\right\|_V \leq r\right) \geq \frac{1}{2^n}$$

where $\|\cdot\|_V$ is the norm with the unit ball V in $\mathbb{R}^n$.

Let B_p^n denote the unit L_p-ball in $\mathbb{R}^n$.

• $\beta(B_2^n, B_2^n) \le \sqrt{n}$, i.e. for any $u_1, \cdots, u_n \in \mathbb{R}^n$ with $|u_i|_2 \le 1$, there exist $\eta_1, \cdots, \eta_n = \pm 1$ so that

$$|\eta_1 u_1 + \cdots + \eta_n u_n| \le \sqrt{n}.$$

• Komlos Conjecture (197?): $\beta(B_2^n, B_\infty^n) \le C$ for some absolute constant $C > 0$. It is well known that Komlos Conjecture would imply Beck-Fiala Conjecture.

• Beck and Fiala (1981): $\beta(B_1^n, B_\infty^n) \le 2$.

• Spencer (1985, 1986):

$$\begin{aligned} c\sqrt{n} \quad \le \quad & \beta(B_\infty^n, B_\infty^n) \le \alpha(B_\infty^n, B_\infty^n) \le C\sqrt{n} \\ & \beta(B_2^n, B_\infty^n) \le C \log n \end{aligned}$$

• Spencer Conj: $\alpha(B_p^n, B_p^n) \le Cn^{1/2+o(1)}$ for $1 \le p < \infty$.

• Giannopoulos (1997): $\beta(B_2^n, V) \le 6 \log n$ if $\gamma_n(V) \ge 1/2$ where γ_n is the standard n-dimensional Gaussian measure with density $(2\pi)^{-n/2} e^{-\|x\|_2^2/2}$.

• Banaszczyk (1998): $\beta(B_2^n, V) \le C$ if $\gamma_n(V) \ge 1/2$ and in particular, $\beta(B_2^n, B_\infty^n) \le C\sqrt{\log n}$.

Open: All Conj. and results above hold for ξ_k.

Hadamard Conjecture:

There exists an Hadamard matrix H_n, or n by n matrix with every entry ± 1 such that $HH^T = nI$ for every $n = 4m$, $m \geq 1$.

To restate the Hadamard Conjecture, let ε_{ij} be i.i.d random variables with $\mathbb{P}(\varepsilon_{ij} = \pm 1) = 1/2$, $1 \leq i, j \leq n$. Then the equivalent formulation of the Hadamard Conjecture is

$$\mathbb{P}\left(\max_{1 \leq j \neq k \leq n} \left|\sum_{i=1}^{n} \varepsilon_{ij}\varepsilon_{ik}\right| < 1\right) \geq 2^{-n^2}.$$

for $n = 4m$.

Gaussian Hadamard Conjecture: *Let ξ_{ij}, $1 \leq i, j \leq n$, be i.i.d standard normal random variables. Then*

$$\log \mathbb{P}\left(\max_{1 \leq j \neq k \leq n} \left|\sum_{i=1}^{n} \xi_{ij}\xi_{ik}\right| < 1\right) \approx -n^2.$$

The quest for proofs of these conjectures is likely to stimulate and to challenge probabilists for years to come.

Random trigonometric polynomials

Salem and Zygmund (1954): Given complex numbers $z_1, \cdots z_n$, there exists a choice of $+$ and $-$ such that

$$\sup_{0 \le t \le 1} \left| \sum_{k=1}^{n} \pm z_k e^{ikt} \right| \le C \left(\log n \sum_{k=1}^{n} |z_k|^2 \right)^{1/2}$$

where C is an absolute constant.
Kahane (1980): There is a polynomial

$$P(z) = \sum_{k=1}^{n} e^{i\theta_k} z^k$$

such that

$$\sup_{|z|=1} |P(z)| \le \sqrt{n}(1 + O(n^{-19/34}(\log n)^{1/2})).$$

Open: Better estimate?

Some observations

•SVP appears naturally in many problems and they are one of the basic type of probability estimates that are interesting, useful and challenging.

•Are we good enough to find the precise order of the decay rate of SVP?

•In many applications, we only need to find/use weaker estimates. So we will present various alternative approaches to same type of problem.

Applications of small deviation probabilities

- Chung's law of the iterated logarithm
- Lower limits for empirical processes
- Rates of convergence of Strassen's FLIL
- Rates of convergence of Chung type FLIL
- A Wichura type functional LIL
- Fractal Geometry for Gaussian random fields
- Metric entropy estimates
- Capacity in Wiener space
- Natural Rates of escape for infinite dimensional Brownian motions
- Asymptotic evaluation of Laplace transform for large time
- Onsager-Machlup functionals
- Random fractal laws of the iterated logarithm

All are discussed in details in a survey paper of Li and Shao (2001).

Additional Relevant Topics

- Volume of Wiener sausage and fractional Brownian sausage
- Classical and average Kolmogorov widths
- Hypercontractivity and comparison of moments of iterated maxima and minima
- Cascade relations for intersection exponents of planar Brownian motion
- Estimates of principle eigenvalue of (fractional) Laplacian
- Exit time of Brownian motion from unbounded domain, principal eigenvalue, heat equation
- Entropy and quantization of Gaussian measure
- Regularity of density for functionals of Gaussian processes
- Decaying turbulent transport
- Random sum of vectors
- Cube slicing
- Dvoretzky theorem in geometric functional analysis, negative moments of a norm
- Hamiltonian and Partition Function
- The Wiener-Hopf Equation

- Longest increasing subsequences, longest common increasing subsequences
- Determinant of random matrix
- Littlewood and Offord type problems
- Existence in random graphs.
- Combinatorial discrepancy.
- Hadamard conjecture.
- Most visit sites via isomorphism theorems
- Singularity of Burgers equation
- Galton-Watson tree and limit of positive Martingale.
- Gaussian free fields
- Singular values and conditional numbers
- Banach-Mazur distances and projection

- Etc. (please let me know)

References and Activities

Mikhail Lifshits maintains an excellent updated bibliography at
http://www.proba.jussieu.fr/pageperso/smalldev/

•This bibliography contains "all" known published and not yet published articles concerning estimates and asymptotic behaviour of small deviation probabilities. **Last update: Dec. 2011**

•Everyone is encouraged to send corrections and new references.

•This site is devoted to the study of various small value problems in mathematics. Special interest is paid to evaluating small deviation probabilities for stochastic processes and their relationships with metric entropy for operators, as well as applications in - but by no means limited to - functional analysis, random graphs, discrepency theory, geometry in Banach spaces, wavelet decompositions, fractal geometry, quantization, coding theory.

•The articles on applications of small deviation theory (e.g. to the laws of iterated logarithm or to quantization problems) are NOT included unless they contain original small deviation results.

Small Value Theory

We believe a theory of small value phenomenon should be developed and centered on:

- systematically studies of the existing techniques and applications
- applications of the existing methods to a variety of fields
- new techniques and problems motivated by current interests of advancing knowledge.

Typical Small Value Behavior

To make precise the meaning of typical behaviors that positive random variables take smaller values, consider a family of *non-negative* random variables $\{Y_t, t \in T\}$ with index set T. We are interested in evaluation $\mathbb{E} \inf_{t\in T} Y_t$ or its asymptotic behavior as the size of the index set T goes to infinity.

Ex: The first passage percolation indexed by paths.

Ex: Random assignment type problems indexed by permutations.

Conj: (Li and Shao) For any centered Gaussian r.v's $(X_i)_{i=1}^n$,

$$\mathbb{E} \min_{1\le i\le n} |X_i| \ge \mathbb{E} \min_{1\le i\le n} |\widehat{X}_i|$$

where $\widehat{X}_i$ are ind. centered Gaussian with $\mathbb{E}\,\widehat{X}_i^2 = \mathbb{E}\,X_i^2$.

Yes for $n = 2, 3$.

Gordon, Litvak, Schutt and Werner (2006):

$$2\mathbb{E} \min_{1\le i\le n} |X_i| \ge \mathbb{E} \min_{1\le i\le n} |\widehat{X}_i|$$

Expected Lengths of Minimum Spanning Tree (MST)

For a simple, finite, and connected graph G with vertex set $V(G)$ and edge set $E(G)$, we assign a non-negative i.i.d random length ξ_e with distribution F to each edge $e \in E(G)$. The total length of the MST is denoted by

$$L^F_{MST}(G) = \min_T \sum_{e \in T} \xi_e = \sum_{e \in MST(G)} \xi_e.$$

In particular, we use the notation $\mathbb{E}[L^u_{MST}(G)]$ for $U(0,1)$ and $\mathbb{E}[L^e_{MST}(G)]$ for $\exp(1)$.

•Frieze (1985): For complete graph K_n on n vertices,

$$\lim_{n\to\infty} \mathbb{E}[L^e_{MST}(K_n)] = \lim_{n\to\infty} \mathbb{E}[L^u_{MST}(K_n)] = \zeta(3) = \sum_{k=1}^{\infty} k^{-3} = 1.202...$$

See related results in Steele (1987), Frieze and McDiarmid (1989), Janson (1995). Pennose (1998), Beveridge, Frieze McDiarmid (1998), Frieze, Ruszink and Thoma (2000), Fill and Steele (2004), Gamarnik (2005).

Exact Formula

•Steele (2002):

$$\mathbb{E}[L^u_{MST}(G)] = \int_0^1 \frac{(1-t)}{t} \frac{T_x(G; 1/t, 1/(1-t))}{T(G; 1/t, 1/(1-t))} dt,$$

where $T(G : x, y)$ is the Tutte polynomial of G and $T_x(G; x, y)$ is the partial derivative of $T(G; x, y)$ with respect to x.

•Li and X. Zhang (2009): For complete graph K_n,

$$0 < \mathbb{E}[L^e_{MST}(K_n)] - \mathbb{E}[L^u_{MST}(K_n)] = \frac{\zeta(3)}{n} + O\left(n^{-2}\log^2 n\right).$$

Combinatorial Optimization

The TSP (travelling salesman problem, i.e. find the shortest route through a set of points) is the paradigm problem in this area.

- Let $L_n = \min_\sigma \sum_{i=1}^n |X_{\sigma(i)} - X_{\sigma(i+1)}|$ be the shortest tour of n i.i.d uniform points $\{X_1, \cdots, X_n\} \subset [0,1]^d$. Then $\mathbb{E}\, L_n / n^{(d-1)/d} \to \beta(d)$. Find "good" estimates on $\beta(d)$.
- Does the Central Limit Theorem hold, i.e. does the length of the optimal tour have a Normal distribution as n tends to infinity?
- Can one prove anything about the geometric structure of the optimal tour?
- Two-sample matching: There is $0 < c_0 < c_1 < \infty$ such that

$$c_0 \le \frac{\mathbb{E}\, M_n}{\sqrt{n \log n}} \le c_1, \quad c_0 \le \frac{\mathbb{E}\, M_n^*}{n^{-1/2} (\log n)^{3/4}} \le c_1$$

where

$M_n = \min_\sigma \sum_{i=1}^n |X_i - Y_{\sigma(i)}|, \quad M_n^* = \min_\sigma \max_{1 \le i \le n} |X_i - Y_{\sigma(i)}|$ and $\{X_i\}$ and $\{Y_i\}$ are i.i.d uniform samples on $[0,1]^2$. Show the limiting constants exists.