Gaussian Processes with Applications

Wenbo V. Li
University of Delaware
E-mail: wli@math.udel.edu
Harbin, May, 2006

Abstract: The theory of Gaussian processes plays a fundamental role in diverse areas of mathematics and engineering. This topic course is intended to introduce graduate students in mathematics and other scientific disciplines to some powerful Gaussian methods and their applications. Our endeavor throughout will be to give attractive examples in various parts of mathematics. Many open problems and conjectures at the undergraduate/graduate level will be discussed.

After covering basic theory, we will present some selected topics from the list below, depending on the interests of students.

Basic theory: Review of basic facts of positive definite matrices and probability; Fundamental elements of Gaussian random vectors and processes; Isoperimetric Gaussian inequalities with applications; Gaussian representations and comparisons; Special series representations for (fractional) Brownian motion.

Selected topics: Entropy, information and central limit theorem; Logarithmic Sobolev inequalities; transportation of measures; Metric entropy of compact sets, reproducing kernel Hilbert space and small ball probabilities; Random polynomials and matrices (real and complex); Kac-Rice formula for level crossing; Applications in geometric functional analysis, approximation theory and complexity.

Some Quotes

"We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it. ... It is remarkable that this science, which originated in the consideration of games of chance, should have become the most important object of human knowledge. ... The most important questions of life are, for the most part, really only problems of probability." So said the famous French mathematician and astronomer (the "Newton of France") Pierre Simon, Marquis de Laplace. Although many people might feel that the famous marquis, who was also one of the great contributors to the development of probability, might have exaggerated somewhat, it is nevertheless to nearly all scientists, engineers, medical practitioners, jurists, and industrialists. In fact, the enlightened individual had learned to ask not "is it so?" but rather "What is the probability that it is so?"
——The first paragraph of preface in the textbook " A First Course in Probability" (7th ed.) by Sheldon Ross.

The most natural measure in a finite-dimensional linear space is, of course, Lebesgue measure. However, it is not a probability measure (the measure of the whole space is infinity, not 1), and it fails to exist in infinite dimension.

The most natural probability measure in a linear space (of finite or infinite dimension) is Gaussian measure.

The modern theory of Gaussian measures lies at the intersection of the theory of random processes, functional analysis, and mathematical physics and is closely connected with diverse applications in quantum field theory, statistical physics, financial mathematics, and other areas of sciences. The study of Gaussian measures combines ideas and methods from probability theory, nonlinear analysis, geometry, linear operators, and topological vector spaces in a beautiful and nontrivial way. (Preface, p. xi.) V.I. Bogachev, "Gaussian measures", AMS 1998.

“Gaussian random variables and processes always played a central role in the probability theory and statistics. The modern theory of Gaussian measures combines methods from probability theory, analysis, geometry and topology and is closely connected with diverse applications in functional analysis, statistical physics, quantum field theory, financial mathematics and other areas.” —-R. Latala, On some inequalities for Gaussian measures. Proceedings of the International Congress of Mathematicians (2002), 813-822. arXiv:math.PR/0304343.

“It is of course impossible to even think the word Gaussian without immediately mentioning the most important property of Gaussian processes, that is concentration of measure.” —-M. Talagrand, See page 189 of “Mean field model for spin glasses: a first course”, Lecture Notes in Math. **1816** (2003), 181-285.

“Gaussian inequalities are always one of the most important parts in both theory and applications of Gaussian measures/vectors/processes.” —-W. Li. “Gaussian inequalities and conjectures”, (2006+).

“The theory of random matrices makes the hypothesis that the characteristic energies of chaotic systems behave locally as if they were the eigenvalues of a matrix with randomly distributed elements. ... but when the complications increase beyond a certain point the situation becomes hopeful again, for we are no longer required to explain the characteristics of every individual state but only their average properties, which is much simpler.” —-M.L. Mehta, Preface and Introduction to his book “Random matrices”, second edition, Academic Press, 1991.

“Moreover a mathematical problem should be difficult in order to entice us, yet not completely inaccessible, lest it mock our efforts. It should be to us a guidepost on the mazy path to hidden truths, and ultimately a reminder of our pleasure in the successful solution.” —David Hilbert, 1900.

“It is not knowledge, but the act of learning, not possession but the act of getting there which generates the greatest satisfaction.” —Carl Friedrich Gauss (1777-1855)

Some literature

•Ledoux, M. and Talagrand, M. (1991). *Probability on Banach Spaces*, Springer, Berlin.

•Lifshits, M.A. (1995). *Gaussian Random Functions*. Kluwer Academic Publishers, Boston.

•Yurinsky, V. (1995). Sums and Gaussian Vectors, *Lecture Notes in Math.* **1617**, Springer-Verlag.

•Ledoux, M. (1996). Isoperimetry and Gaussian Analysis, *Lectures on Probability Theory and Statistics, Lecture Notes in Math.* **1648** 165–294, Springer-Verlag.

•V.I. Bogachev, "Gaussian measures", AMS 1998.

•Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities, *Lecture Notes in Mathematics* **1709**, 120-216.

•Li, W.V. and Shao, Q-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. Handbook of Statistics, Vol. **19**, *Stochastic processes: Theory and methods*, 533-598, Elsevier, New York.

•Ledoux, M. (2001). *The concentration of measure phenomenon*, Mathematical Surveys and Monographs 89, AMS.

•Talagrand, M. (2003). *Spin Glasses: a Challenge to Mathematicians*, Cavity and Mean Field Models. A Series of Modern Surveys in Mathematics, Vol. 46.
•Guionnet, A. (2004). Large deviations and stochastic calculus for large random matrices, Probability Surveys **1**, 72-172.
•Ledoux, M. (2005). Deviation inequalities on largest eigenvalues, Summer School on the Connections between Probability and Geometric Functional Analysis, Jerusalem.
•Sheffield, S. (2004+). Gaussian free fields for mathematicians, arXiv:math.PR/0312099.
•Bai, Z. and Silverstein, J. (2005+). Spectral analysis of Large dimensional random matrices, 435 pages.
•Marcus, M. and Rosen, J. (2006+). *Markov processes, Gaussian processes and local times*, 500+ pages.
•Li, W.V. (2006+). *Gaussian inequalities and conjectures*, 30+ pages.

Notion: If X and Y have the same d.f, i.e. $\mathbb{P}(X \leq z) = \mathbb{P}(Y \leq z)$ for all $z \in \mathbb{R}$, we say X and Y *equal in distribution* and write $X \stackrel{d}{=} Y$.

Ex: (Sum of two fair dice). Roll two fair and "standard" dice with values $\{1, 2, 3, 4, 5, 6\}$ and let X be the sum of the face-up values. Next roll two fair but "non-standard" dice with values $\{1, 2, 2, 3, 3, 4\}$ on one and $\{1, 3, 4, 5, 6, 8\}$ on the other, and let Y be the sum of the face-up values. Then $X \stackrel{d}{=} Y$.

Uniqueness and discovery: Generating function.

Review: Normal (Gaussian) Random Variables

A real valued r.v X is *normal* or *Gaussian* with parameter μ and σ^2, $\sigma > 0$, if its p.d.f is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty.$$

• One can show that $\int_{-\infty}^{\infty} f(x)dx = 1$.
• The expectation $\mathbb{E}\, X = \mu$.
• The variance Var$(X) = \sigma^2$ and SD$(X) = \sigma$.
• We use $X \sim N(\mu, \sigma^2)$ to denote normal r.v. with mean μ and variance σ^2.
• When $\mu = 0$ and $\sigma = 1$, $N(0,1)$ is called standard normal and denoted by ξ with c.d.f

$$\Phi(x) = \mathbb{P}(\xi \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt = \int_{-\infty}^{x} \phi(t) dt$$

and $\Phi(-x) = 1 - \Phi(x) = \bar{\Phi}(x)$.
• Standardization: If $X \sim N(\mu, \sigma^2)$, then

$$\xi = \frac{X - \mu}{\sigma} \sim N(0,1).$$

Four ways to show $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$.

I. Double integral, polar-coordinates substitution.
II. Double integral, scale transformation.
III. Differentiation of an integral. *Monthly*, 1990, p39.
IV. Limit representation for e^{-x^2}, Wallis formula.

Thm: For $x > 0$, we have

$$e^{x^2/2} \int_x^\infty e^{-y^2/2} dy \downarrow 0$$

$$\int_x^\infty e^{-y^2/2} dy \sim \frac{1}{x} e^{-x^2/2}$$

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \le \frac{x}{1+x^2} \le e^{x^2/2} \int_x^\infty e^{-y^2/2} dy \le \frac{1}{x}$$

$$\int_x^\infty e^{-y^2/2} dy \approx e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3} + \cdots + (-1)^k \frac{(2k-1)!!}{x^{2k+1}}\right).$$

$$\bar{\Phi}(x) = \int_x^\infty \phi(t) dt < \frac{1}{2} \exp\{-x^2/2\}.$$

Idea of Pf: Integration by parts.

- Any better estimates?

Estimate of Ito-Mckean (1974, p17):

$$\frac{2}{\sqrt{x^2+4}+x} \leq e^{x^2/2}\int_x^\infty e^{-y^2/2}dy < \frac{2}{\sqrt{x^2+2}+x}.$$

Stirling's Formula (De Moivre 1730)

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi}e^{-n}n^{n+1/2}} = 1,$$

i.e. $n! \sim \sqrt{2\pi}e^{-n}n^{n+1/2}$.

A Little History: Stirling's formula was found by Abraham de Moivre and published in "*Miscellenea Analytica*" in 1730. It was later refined, but published in the same year, by J. Stirling in "*Methodus Differentialis*" along with other little gems of thought. For instance, therein, Stirling computes the area under the Bell Curve:

$$\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi},$$

by showing that $\Gamma(1/2) = \sqrt{\pi}$. Note that

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt, \quad n! = \Gamma(n+1)$$

and $\Gamma(1/2) = \int_0^\infty e^{-x^2/2}dx$.

Tusnády's Inequality

In one of the most important probability papers of the last forty years, Komlós, Major and Tusnády (1975) sketched a proof for a very tight coupling of the standardized empirical distribution function with a Brownian bridge, a result now often referred to as the KMT, or Hungarian, construction. Their coupling greatly simplifies the derivation of many classical statistical resultssee Shorack and Wellner [(1986), Chapter 12 et seq.], for example.

At the heart of the KMT method [with refinements as in the exposition by Csörgő and Révész (1981), Section 4.4] lies the quantile coupling of the Bin$(n, 1/2)$ and $N(n/2, n/4)$ distributions, which may be defined as follows. Let Y be a random variable distributed $N(n/2, n/4)$. Find the cutpoints $-\infty = \beta_0 < \beta_1 < \cdots < \beta_n < \beta_{n+1} = \infty$ for which

$$\mathbb{P}(\text{Bin}(n, 1/2) \geq k) = \mathbb{P}(Y > \beta_k), \quad k = 0, 1, \cdots, n.$$

When $\beta_k < Y \leq \beta_{k+1}$, let X take the value k. Then X has a Bin$(n, 1/2)$ distribution.

Symmetry considerations show that $\beta_{n-k+1} = n - \beta_k$, so that it suffices to consider only half the range for k. The usual normal approximation with continuity correction suggests that $\beta_k \simeq k - 1/2$, which, if true, would bound $|X - Y|$ by a constant that does not change with n. Of course, such an approximation for all k is too good to be true, but results almost as good have been established. The most elegant version appeared in the unpublished dissertation (in Hungarian) of Tusnády (1977), whose key inequality may be expressed as the assertion

$$k - 1 \le \beta_k \le \frac{3n}{2} - \sqrt{2n(n-k)}, \quad \text{for } n/2 \le k \le n.$$

As explained by Csörgő and Révész [(1981), Section 4.4], Tusnády's inequality implies that

$$|X - n/2| \le |Y - n/2| + 1 \quad \text{and} \quad |X - Y| \le 1 + \xi^2/8,$$

where ξ denotes the standardized variable $(2Y - n)/\sqrt{n}$. They also noted that Tusnády's proof of the inequality above was "elementary", but "not at all simple."

Bretagnolle and Massart [(1989), Appendix] published another proof of Tusnády's inequality–an exquisitely delicate exercise in elementary calculus and careful handling of Stirling's formula to approximate individual Binomial probabilities. With no criticism intended, it is noted in Carter and Pollard (2004) that the proof is quite difficult. More recently, Dudley [(2000), Chapter 1] and Massart (2002) have reworked and refined the Bretagnolle/Massart calculations. Clearly, there is a continuing perceived need for an accessible treatment of the coupling result that underlies the KMT construction.

Very recently, an elementary proof of a result that sharpens the Tusnády inequality, modulo constants is found Carter and Pollard (2004). Their method uses the beta integral representation of Binomial tails, simple Taylor expansion and some novel bounds for the ratios of normal tail probabilities $\bar{\Phi}(x+y)/\bar{\Phi}(x)$.

Ref: Carter, A. and Pollard, D. (2004). Tusnady's inequality revisited, *Ann. Stat.* **32**, 2731-2741.

Lemma: If $Y \geq 0$ and $p > 0$, then

$$\mathbb{E}\,(Y^p) = \int_0^\infty p y^{p-1} \mathbb{P}(Y > y) dy.$$

Pf. 1: Use Fubini's Thm. Standard argument. Details in class.

Pf. 2: Use the integration by parts. Discovery argument. Details in class.

- Fubini's Thm $\leftrightarrow$ integration by parts?!
- For $p = 1$ and $Y \geq 0$, $\mathbb{E}\,Y = \int_0^\infty \mathbb{P}(Y > y) dy$.
- If $H(x) = \int_{-\infty}^x h(y) dy$ with $h(y) \geq 0$, then

$$\mathbb{E}\,H(Y) = \int_{-\infty}^\infty h(y) \mathbb{P}(Y > y) dy.$$

Basic Properties of Variance

Def: The variance of a r.v. X is defined by

$$\text{Var}(X) = \mathbb{E}\,(X-\mu)^2, \quad \text{where} \quad \mu = \mathbb{E}\,X.$$

It is a measurement of the spread of X around its mean $\mu = \mathbb{E}\,X$. The standard deviation is $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

- $\text{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$.
- For any $a, b \in \mathbb{R}$, $\text{Var}(aX+b) = a^2\text{Var}(X)$.
- For any $a \in \mathbb{R}$,

$$\mathbb{E}\,(X-a)^2 \geq \mathbb{E}\,(X-\mu)^2 = \text{Var}(X).$$

- For any $a \in \mathbb{R}$,

$$\begin{aligned} \text{Var}(X) &\geq \text{Var}\left(\max(a, X)\right) \\ \text{Var}(X) &\geq \text{Var}\left(\min(a, X)\right). \end{aligned}$$

•For any C-Lipschitz function f, i.e. $|f(x)-f(y)| \leq C|x-y|$ for all $x, y \in \mathbb{R}$,

$$\text{Var}(f(X)) \leq C^2\text{Var}(X).$$

Key: Representation and randomization.

•Let U and V be independent and have uniform distributions on $(0, 1)$. Let $\Theta = 2\pi U$, $Z = -\log V$, $R = \sqrt{2Z}$, $X_1 = R \cos \theta$ and $X_2 = R \sin \theta$. Then X_1 and X_2 are independent and have standard normal distributions. This gives a way to simulate standard normals when your computer can only give you uniforms.

•The r.v $\frac{1}{2}(\xi_1^2 + \xi_2^2) \sim \text{expo}(1)$ where ξ_1, ξ_2 are ind. $N(0, 1)$.

Review: De Moivre-Laplace CLT

Let $X_1, X_2, \cdots$ be i.i.d with $\mathbb{P}(X_1 = \pm 1) = 1/2$ and $S_n = \sum_{i=1}^n X_i$.

The De Moivre-Laplace Thm: If $a < b$, then as $n \to \infty$,

$$\mathbb{P}\left(a \le S_n/\sqrt{n} \le b\right) \to \int_a^b (2\pi)^{-1/2} e^{-x^2/2} dx$$

Ideas of Pf:

- Exact computation:

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n}$$

- Local CLT: If $2k/\sqrt{2n} \to x$, then

$$\mathbb{P}(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}$$

- Approximation:

$$
\begin{aligned}
& \mathbb{P}\left(a \le S_{2n}/\sqrt{2n} \le b\right) \\
= & \sum_{m\in[a\sqrt{2n},b\sqrt{2n}]\cap 2\mathbb{Z}} \mathbb{P}(S_{2n}=m) \\
\approx & \sum_{x\in[a,b]\cap(2\mathbb{Z}/\sqrt{2n})} (2\pi)^{-1/2}e^{-x^2/2}\cdot(2/n)^{1/2} \\
\approx & \int_a^b (2\pi)^{-1/2}e^{-x^2/2}dx
\end{aligned}
$$

Ex: The proof above is very special but of historical interests. Another special case that can be treated with Stirling's formula is $T_n = Y_1 + \cdots + Y_n \sim$ Poisson(n) where $Y_i \sim$ Poisson(1) are i.i.d, i.e. $\mathbb{P}(T_n = k) = e^{-n}n^k/k!$. We have

$$\mathbb{P}(T_n = k) \sim (2\pi n)^{-1/2}e^{-x^2/2}$$

if $(k-n)/\sqrt{n} \to x$, and

$$\mathbb{P}\left(a \le (T_n - n)/\sqrt{n} \le b\right) \to \int_a^b (2\pi)^{-1/2}e^{-x^2/2}dx.$$

Review: Characteristic Functions

Def: The *characteristic function* (ch.f.) of a r.v. X is $\phi(t) = \phi_X(t) = \mathbb{E}\, e^{\mathrm{i}tX} = \mathbb{E}\,(\cos tX + \mathrm{i} \sin tX)$. It is the Fourier transform of the density. The importance of the transform is the series representation of the function and the determination of properties from the transformed function. The ch.f. uniquely determines the distribution.

• **Normal:** $f(x) = (\sqrt{2\pi}\sigma)^{-1} \exp(-(x-u)^2/2\sigma^2)$ and $\phi(t) = \exp(iut - \sigma^2 t^2/2)$.

Physics Proof vs Math Proof: Details in class.

•For $\xi \sim N(0,1)$, $\mathbb{E}\,\xi^{2n} = (2n)!/(2^n \cdot n!)$.

Review: Central Limit Theorems (CLT)

Thm (CLT for i.i.d. sequences): Let $X_1, \cdots$ be i.i.d. with $\mathbb{E}\, X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$. Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow N(0,1).$$

• Pairwise ind. is good enough for SLLN but it is not enough for CLT.

Here are a list of common methods to show CLT:

• Exact computation
• Ch.f method
• Moments method
• Cumulants method, $\log\phi(t) = \sum_{j=0}^{\infty} \kappa_j (it)^j / j!$
• Stein-Chen method
• Martingale method
• Empirical distribution method
• Diffusion limit

Basic Gaussian Properties

The density and distribution function of the standard Gaussian (normal) distribution on the real line $\mathbb{R}$ are

$$\phi(x) = (2\pi)^{-1} \exp\{-x^2/2\} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

Let γ_n denote the canonical Gaussian measure on $\mathbb{R}^n$ with density function

$$\phi_n(x) = (2\pi)^{-n/2} \exp(-|x|^2/2)$$

with respect to Lebesgue measure, where $|x|$ is the Euclidean norm of $x \in \mathbb{R}^n$. We use μ to denote a centered Gaussian measure throughout. All results for γ_n on $\mathbb{R}^n$ can be used to determine the appropriate infinite dimensional analogue by a classic approximation argument. Details later.

Basic Facts for d-dim Gaussian Density

Def: The (non-degenerate) d-dimensional r.v $X = (X_1, \cdots, X_d) \in \mathbb{R}^d$ is normal or Gaussian with mean $m = (m_1, \cdots, m_d) \in \mathbb{R}^d$ and covariance matrix $\Sigma_{d\times d}$ if its density is

$$\begin{aligned} f_X(x) &= f(x_1, \cdots, x_d) \\ &= \frac{1}{(2\pi)^{d/2}(\det(\Sigma))^{1/2}} \cdot \\ &\qquad \exp\left\{-\frac{1}{2}\langle (x-m), \Sigma^{-1}(x-m)^t\rangle\right\} \end{aligned}$$

where the covariance matrix $\Sigma_{d\times d} = (\Sigma_{ij})_{d\times d}$ is a (strictly) positive definite symmetric matrix.

- $\langle (x-m), \Sigma^{-1}(x-m)^T\rangle = (x-m)\Sigma^{-1}(x-m)^T$.
- $\int_{\mathbb{R}^d} f_X(x)dx = 1$.
- $\mathbb{E}\, X = (\mathbb{E}\, X_1, \cdots, X_d) = (m_1, \cdots, m_d) = m$.
- $\Sigma_{ij} = (\text{Cov}(X_i, X_j)_{d\times d} = (\mathbb{E}\,(X_i - \mathbb{E}\, X_i)(X_j - \mathbb{E}\, X_j))_{d\times d}$, $1 \le i, j \le d$.
- What about $\det(A) = 0$ or infinite dimensional extension? Need a different point of view.

Review: Positive Definite Matrices

Def: A symmetric matrix $A_{n\times n}$ is (strictly) *positive definite, or p.d* if $\mathbf{x}A\mathbf{x}^t > 0$ for every nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$. A symmetric matrix $A_{n\times n}$ is *positive semi-definite* if $\mathbf{x}A\mathbf{x}^t \geq 0$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$.

•A is p.d iff A^{-1} is p.d.
•A is p.d iff $\det(A_k) > 0$ for all $1 \leq k \leq n$, where $A_k = (a_{ij})_{1\leq i,j\leq k}$.
•Let $S = \{\mathbf{u}_1, \cdots, \mathbf{u}_d\}$ be an ordered basis for a finite-dimensional inner product vector space. Let $C = [c_{ij}]_{d\times d}$, $c_{ij} = (\mathbf{u}_i, \mathbf{u}_j)$. Then the matrix C is positive definite.

•A symmetric matrix $A_{n\times n}$ is positive definite iff all the eigenvalues of A are positive.

A Useful Continuity Argument

To show det(A) > 0 or det(A_k) > 0 for p.d A, we first note that it is easily seen det(A) $\neq 0$. If we should have det(A) $= 0$, we could choose a nontrivial set of values x_j so that $A\mathbf{x}^t = \mathbf{0}$ and we would have $\mathbf{x}A\mathbf{x}^t = 0$, a contradiction. Next we employ a continuity argument that can be often be used in similar situations.

Consider the matrix $\lambda I + (1-\lambda)A$, where $0 \leq \lambda \leq 1$ and I is the identity matrix. Clearly this matrix is p.d. if A is p.d. Hence

$$f(\lambda) = \det(\lambda I + (1-\lambda)A) \neq 0.$$

Since the determinant function $f(\lambda)$ is continuous and $f(1) > 0$, it forces det(A) $= f(0) > 0$.

•For p.s.d. matrix $A_{n\times n}$, det(A_k) ≥ 0 for all $1 \leq k \leq n$.

Representation as a Sum of Squares

For p.d matrix A, the following representation makes the positivity of $Q(\mathbf{x}) = \mathbf{x}A\mathbf{x}^t$ obvious.

Thm: (Lagrange, Beltrami) If $\det(A_k) \neq 0$ for all $1 \leq k \leq n-1$, then

$$Q(\mathbf{x}) = \sum_{k=1}^{n} \frac{\det(A_k)}{\det(A_{k-1})} y_k^2$$

where $\det(A_0) = 1$, $y_k = x_k + \sum_{j=k+1}^{n} b_{kj} x_j, \quad 1 \leq k \leq n$, and the b_{kj} are rational functions of the a_{ij}.

•Idea of Pf: Inductive (projective or Gram-Schmidt procedure or QR factorization) argument.

•An immediate consequences of the representation is the fundamental characterization: A is p.d iff $\det(A_k) > 0$ for all $1 \leq k \leq n$.

•For p.d A, $A = QDQ^t$ where Q is orthogonal, D is diagonal matrix with (positive) eigenvalues λ_k of A on the diagonal. Equivalently, there is an orthogonal transform $\mathbf{x}^t = Q\mathbf{w}^t$ such that $Q(\mathbf{x}) = \mathbf{x}A\mathbf{x}^t = \sum_{k=1}^{n} \lambda_k w_k^2$.

P.s.d. Matrix and Inner Product

For any real matrix $M_{n\times m}$, the matrix $A_{n\times n} = MM^t$ is p.s.d. and $|A| = \det(A) = |M|^2$ if $n = m$. If we write $M = \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(n)} \end{pmatrix}$ where $x^{(i)}$ are row vectors in $\mathbb{R}^m$, then $MM^t = ((x^{(i)}, x^{(j)}))$. If M is nonsingular, then MM^t is p.d. In general, for any vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$ in an inner product space, the matrix $M = ((\mathbf{u}_i, \mathbf{u}_j))_{n\times n}$ is p.s.d and it is p.d if the vectors form a base.

•The matrix

$$\left(\int_a^b f_i(x) f_j(x) dx\right)_{n\times n}$$

is p.s.d and we have the determinant representation

$$\begin{aligned} &\det\left(\int_a^b f_i(x) g_j(x) dx\right)_{n\times n} \\ = \ &\frac{1}{n!}\int_{[a,b]^n} \det(f_i(x_j))_{n\times n} \cdot \det(g_i(x_j))_{n\times n} d\mathbf{x} \end{aligned}$$

where $d\mathbf{x}$ denotes $dx_1 \cdots dx_n$.

Representation of $\det(A)$ for p.d Matrix

From the Gaussian density formula,

$$\frac{\pi^{n/2}}{\det(A)^{1/2}} = \int_{\mathbb{R}^n} e^{-(x,Ax)} dx$$

which can be evaluated by the change of variables $y_k = x_k + \sum_{j=k+1}^n b_{kj} x_j, \quad 1 \le k \le n$.

Thm: If A and B are $n \times n$ p.d matrices and $0 \le \lambda \le 1$, then

$$|\lambda A + (1-\lambda)B| \ge |A|^\lambda \cdot |B|^{1-\lambda}$$

Pf: Gaussian rep. and Holder's inequality.

Thm: (Hadamard's inequality for p.d matix). Define $A_{rs} = (a_{ij})_{r \le i,j \le s}$ for a given p.d. matrix $A = (a_{ij})_{n \times n}$. Then $|A_{1n}| \le |A_{1k}| \cdot |A_{k+1,n}|$. In particular, $|A| \le \prod_{i=1}^n a_{ii}$.

Pf: Gaussian representation and symmetric sums.

Thm: (Hadamard's inequality). For any matrix $M = (x_{ij})_{n \times n}$, $|\det(M)| \le \prod_{i=1}^n \left(\sum_{j=1}^n x_{ij}^2 \right)^{1/2}$. The study of the equality leads to the famous Hadamard conjecture.

Pf: Make a p.d matrix from M.

Complex Matrices with p.d Real Part

With bit more effort, we can establish the result:
Thm: If $A_{n\times n}$ and $B_{n\times n}$ are real, symmetric matrices with A p.d, then

$$\int_{\mathbb{R}^n} e^{-(x,(A+\mathrm{i}B)x)}dx = \frac{\pi^{n/2}}{(\det(A+\mathrm{i}B))^{1/2}}$$

in which the principle values of the square roots are understood to be used.

•Ostrowski and Taussky Inequality: Under the assumption of the above theorem,

$$|\det(A+\mathrm{i}B)| \geq |\det(A)|,$$

and the sign of equality holds iff B is the zero matrix.

Siegel's Representation

The classical integral of Euler is

$$\int_0^\infty e^{xy} x^{s-1} dx = \Gamma(s) y^{-s}, \quad \mathrm{Re}(s), \mathrm{Re}(y) > 0.$$

an extensive generalization, due to Siegel (1935), is

$$\begin{aligned} &\int_{X>0} e^{-\mathrm{tr}(XY)} |X|^{s-(n+1)/2} dV \\ = \; &\pi^{n(n-1)/4} \frac{\Gamma(s)\Gamma(s-1/2)\cdots\Gamma(s-(n-1)/2)}{|Y|^s} \end{aligned}$$

Here X and Y are p.d. matrices of order n, tr(X) denotes the trace of X,

$$dV = \prod_{1 \le i \le j \le n} dx_{ij}$$

and the integration is over the region where X is p.d. The real part of s is to be greater than $(n-1)/2$.

Permanent Rep. via Complex Gaussian r.v's

We denote by $\gamma_n^{\mathbb{C}}$ the canonical Gaussian measure on $\mathbb{C}^n$, that is, the measure with density function $(2\pi)^{-n}\exp(-(|x|^2+|y|^2)/2)$, where $|x|$ is the Euclidean norm.

If $A=(a_{ij})$ is an $n\times n$ matrix, then the permanent of A is defined as

$$\mathrm{per}(A)=\sum_{\sigma\in[n]} a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$$

where $[n]$ denotes the set of permutations of the first n natural numbers. More information on the permanent can be found in the book by Mint (1978), *Permanents, Encyclopedia of mathematics and its applications*. In particular, it presents a proof of the following theorem.

Thm: (Lieb). Let $A=(a_{ij})$ be a p.s.d Hermitian $n\times n$ matrix, and let $A=\begin{pmatrix} B & C \\ C^* & D\end{pmatrix}$ where B is $k\times k$ and D is $(n-k)\times(n-k)$, then

$$\mathrm{per}(A)\geq \mathrm{per}(B)\cdot\mathrm{per}(D).$$

Let $p = (p_1, \cdots, p_n)$ be an n-tuple of nonnegative integers. We denote the degree of p by $|p| = p_1 + \cdots + p_n$, and the product of factorials $p_1!p_2!\cdots p_n!$ by $p!$. If p is an n-tuple we denote by z^p the monomial $z_1^{p_1}\cdots z_n^{p_n}$ of degree $|p|$.

Prop: Denote by $X_j : \mathbb{C}^n \to \mathbb{C}$ the coordinate functions $X_j(z) = z_j$. Then the monomials $(X^q)(q \in \mathbb{N}^n)$ form an orthogonal system in $L_2(\gamma_n^{\mathbb{C}})$.

Pf: Since $X_i, \cdots, X_n$ are independent standard complex gaussian r.v's, we only need to prove that $(X_1^n)_{n\in\mathbb{N}}$ are orthogonal system. If $m > k$ we have

$$\begin{aligned}\mathbb{E}\,(X_1^m\overline{X}_1^k) &= \int_{\mathbb{C}} z^m\overline{z}^k d\gamma_1^{\mathbb{C}}(z) \\ = \ & \frac{1}{2\pi}\int_{\mathbb{R}^2}(x^2+y^2)^k(x+\mathtt{i}y)^h e^{-(x^2+y2)/2}dxdy,\end{aligned}$$

where $h = m - k > 0$. Changing to polar coordinates

$$\mathbb{E}\,(X_1^m\overline{X}_1^k) = \frac{1}{2\pi}\int_0^\infty\int_0^{2\pi} r^{2k+h+1}e^{\mathtt{i}h\theta}e^{-r^2/2}d\theta dr = 0$$

If $m = k$ we obtain

$$\begin{aligned}\mathbb{E}\,(|X_1|^m) &= \int_0^\infty r^{2m+1}e^{-r^2/2}dr \\ &= \int_0^\infty (2u)^m e^{-u}du = 2^m m!.\end{aligned}$$

The cornerstone is the following relation between complex Gaussian r.v's and the permanent.

Thm: (Arias-de-Reyna) Let $(\mathbf{a}_j)_{j=1}^n$ and $(\mathbf{b}_j)_{j=1}^n$ be vectors in $\mathbb{C}^d$. Then

$$\begin{aligned} & 2^n \mathrm{per}(\langle \mathbf{a}_j, \mathbf{b}_k \rangle) \\ = & \int_{\mathbb{C}^d} \langle \mathbf{a}_1, \mathbf{z} \rangle \langle \mathbf{z}, \mathbf{b}_1 \rangle \cdots \langle \mathbf{a}_n, \mathbf{z} \rangle \langle \mathbf{z}, \mathbf{b}_n \rangle d\gamma_d^{\mathbb{C}}(\mathbf{z}) \\ = & \mathbb{E} \prod_{j=1}^n \langle \mathbf{a}_j, X \rangle \cdot \langle X, \mathbf{b}_j \rangle. \end{aligned}$$

The rep. gives us the following inequality.

Thm: Let $Y = (Y_j)_{j=1}^n$ be a complex Gaussian vector, then

$$\mathbb{E}\left(|Y_1^{p_1} \cdots Y_n^{p_n}|^2\right) \geq \mathbb{E}\left(|Y_1^{p_1}|^2\right) \cdots \mathbb{E}\left(|Y_n^{p_n}|^2\right).$$

Definitions and some simple properties

A real valued random variable X is a Gaussian random variable if it has characteristic function

$$\mathbb{E}\, e^{i\lambda X} = \exp\{im\lambda - \frac{\sigma^2\lambda^2}{2}\}$$

for some real numbers m and σ. It follows, by differentiating with respect to λ and then setting $\lambda = 0$, that

$$\mathbb{E}\,(X) = m \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2.$$

An $\mathbb{R}^n$ valued random variable ξ is a Gaussian random variable if (y, ξ) is a real valued Gaussian random variable for each $y \in \mathbb{R}^n$. This is equivalent to saying that it has characteristic function

$$\begin{aligned}\phi_\xi(y) &:= \mathbb{E}\, e^{\mathrm{i}(y,\xi)} = \mathbb{E}\, e^{\mathrm{i}\xi y^t} \\ &= \exp\{\mathrm{i}\mathbb{E}\,((y,\xi)) - \frac{\mathrm{Var}((y,\xi))}{2}\}\end{aligned}$$

for each $y \in \mathbb{R}^n$.

Setting

$$\mathbb{E}\,\xi_j = m_j \quad \text{and} \quad \mathbb{E}\,(\xi_j - m_j)(\xi_k - m_k) = \Sigma_{j,k}$$

Then we can rewrite

$$\phi_\xi(y) = \exp\{\mathrm{i}my^t - \frac{y\Sigma y^t}{2}\}$$

where $m, y \in \mathbb{R}^n$ and $\Sigma = \{\Sigma_{j,k}\}_{j,k=1}^n$ is a symmetric $n \times n$ matrix with real components. For this reason m is called the mean vector, or simply the mean of ξ and Σ the covariance matrix of ξ. We note that the distribution of an $\mathbb{R}^n$ valued Gaussian random variable is completely determined by its mean vector and covariance matrix. The rank of Σ equals the dimension of the subspace of $\mathbb{R}^n$ which supports the distribution of ξ.

The covariance matrix Σ in the definition of Gaussian random vector is positive definite. This is easy to see, since

$$\sum_{j,k=1}^n a_j a_k \Sigma_{j,k} = \mathbb{E}\left(\sum_{j=1}^n a_j(\xi_j - m_j)\right)^2 .$$

Let ξ be a $\mathbb{R}^n$ valued Gaussian random variable with mean m and covariance matrix Σ and let $U : \mathbb{R}^n \to \mathbb{R}^p$ be a linear transformation. Then $\eta = (U\xi^t)^t$ is a $\mathbb{R}^p$ valued Gaussian random variable with mean $(Um^t)^t$ and covariance matrix $U\Sigma U^t$. This follows since

$$\phi_\eta(z) = \mathbb{E}\, e^{\mathrm{i}\eta z^t} = \mathbb{E}\, e^{\mathrm{i}\xi(zU)^t} = \phi_\xi(zU).$$

When Σ is strictly positive definite its rank is equal to n. In this case the probability distribution of ξ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^n$ and has probability density function $f(x)$ given early.

Let $\xi \sim N(0, \Sigma) \in \mathbb{R}^n$ and $a \in \mathbb{R}^n$. Then in dimension one

$$\mathbb{P}(\alpha \le (a, \xi) \le \beta) = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_\alpha^\beta e^{-u^2/2\sigma^2}\, du$$

where $\sigma^2 = \mathbb{E}\,((a, \xi))^2 = a\Sigma a^t$. Using this we easily see that

$$\mathbb{E}\, e^{(a,\xi)} = \exp\{\frac{a\Sigma a^t}{2}\}.$$

A remarkable property of Gaussian r.v's is that uncorrelated ones are independent.

Thm: Let ξ be an $\mathbb{R}^n$ valued Gaussian random variable with mean vector m and assume that

$$\mathbb{E}\,(\xi_j - m_j)(\xi_k - m_k) = 0 \qquad j \neq k$$

Then $\xi_1, \dots, \xi_n$ are independent.

Pf: In this case

$$\phi_\xi(y) = \prod_{j=1}^{n} \exp\{i m_j y_j - \frac{\Sigma_{j,j} y_j^2}{2}\}.$$

Ex: If $X \sim N(0, \Sigma)$ in $\mathbb{R}^n$ and X^* is an ind. copy of X, then for every θ, the rotation of angle θ of the vector (X, X^*), i.e.

$$(X \sin\theta + X^* \cos\theta, X \cos\theta - X^* \sin\theta)$$

has the same distribution as (X, X^*).

•Let $X_1, X_2, \cdots$ be ind. $N(\mu, \sigma^2)$ r.v's and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Show that for any $m < n$, $\overline{X}_n$ is ind. of $\overline{X}_n - \overline{X}_m$.

•Let X and Y be ind. $N(0,1)$ r.v's. Show that

$$\frac{2XY}{\sqrt{X^2+Y^2}} \quad \text{and} \frac{X^2 - Y^2}{\sqrt{X^2+Y^2}}$$

are ind. $N(0,1)$ r.v's. Can you generalize this result?

The Sample Mean and the Sample Variance

Let $X_1, X_2, \cdots$ be ind. $N(\mu, \sigma^2)$ r.v's and let

$$\overline{X} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Then the joint and marginal distributions of the sample mean and the sample variance for i.i.d normal samples are given by (1), (3) and (4) in the following useful result.

Thm: Assume $X_1, \cdots, X_n$ are ind. and $X_i \sim N(\mu, \sigma^2)$. Then
(1). $\overline{X} \sim N(\mu, \sigma^2/n)$;
(2). $\overline{X}$ is independent of the random vector
$(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$, i.e. $\overline{X}$ is ind. of each of $(X_i - \overline{X})$, $1 \leq i \leq n$.
(3). $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$;
(4). $\overline{X}$ and S^2 are independent;
(5). $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$;
(6). $\mathbb{E}\left(\overline{X}(X_j - \overline{X})\right) = 0$ for any $1 \leq j \leq n$.
Pf of (3): Use Ch.f and the identity

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2.$$

A function $\Sigma(s,t)$ on $T \times T$ is a positive definite function if for all $n \geq 1$ and $t_1, \ldots, t_n$ in T, the $n \times n$ matrix Σ, defined by $\Sigma_{j,k} = \Sigma(t_j, t_k)$ is a positive definite matrix. A function $\{\phi(u), u \in T\}$ is positive definite if the function $\Sigma(s,t) := \phi(s-t)$ is positive definite.

Ex: The following functions are positive definite.

(i). Brownian motion. $\Sigma(s,t) = \min(s,t)$ with $T = [0, \infty)$.

(ii). Two sided BM. $\Sigma(s,t) = \frac{1}{2}(|s| + |t| - |s-t|)$ with $T = (-\infty, \infty)$.

(iii). Brownian bridge. $\Sigma(s,t) = \min(s,t) - st$ with $T = [0,1]$.

(iv). O-U process. $\Sigma(s,t) = e^{-\sigma|t-s|}$ with $T = [0, \infty)$ and $\sigma > 0$.

(v). Fractional Brownian motion. $\Sigma(s,t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H})$ with $T = (-\infty, \infty)$ or $T = [0, \infty)$, where $0 < H \leq 1$ is the Hust index.

(vi). L-process. $\Sigma(s,t) = \frac{2st}{s+t}$ with $T = [0, \infty)$.

(vii). Shifted BM: $\Sigma(s,t) = (s-t)^2 - |s-t| + \frac{1}{2}$ with $T = [0,1]$

(viii). Slepian process: $\Sigma_a(s,t) = \max(0, a - |s-t|)$ with $T = [0, \infty)$ and $a > 0$.

(ix). Median process: $\Sigma(s,t) = \sqrt{st} \sin^{-1}\left(\frac{\min(s,t)}{\sqrt{st}}\right)$ with $T = [0, \infty)$.

Gaussian Processes with Arbitrary Index Set

A real valued stochastic process $\{X(t), t \in T\}$, (T is some index set), is a Gaussian process if its finite dimensional distributions are Gaussian. It is characterized by its mean function m and its covariance kernel Σ which are given by

$$m(t) = \mathbb{E}\, X(t), \Sigma(s,t) = \mathbb{E}\,(X(t) - m(t))(X(s) - m(s)).$$

Note that $\Sigma(s,t)$ is a positive definite function on $T \times T$.

Conversely, and this is the route we generally take, given a real valued function $m(t)$ on T and a positive definite function $\Sigma(s,t)$ on $T \times T$, we obtain a real valued Gaussian process $\{X(t), t \in T\}$, with mean function m and covariance kernel Σ. Note that we do have a consistent family of finite dimensional Gaussian distributions on the finite subsets of T.

Ep: All examples given in the last page are Gaussian processes.

Ep: The following scaling properties can be easily checked via the correlation functions. We assume the constant $c > 0$.

(i). For Brownian motion $W(t)$ on $T = [0, \infty)$, $\{W(ct), t \in T\} =^d \{c^{1/2}W(t), t \in T\}$ and the time-inversion $\{tW(\frac{1}{t}), t \in T\} =^d \{W(t), t \in T\}$. Similar properties also hold for the L-process.

(ii). For fBM $B_H(t)$ with $T = (-\infty, \infty)$ or $T = [0, \infty)$, $\{B_H(ct), t \in T\} =^d \{c^H B_H(t), t \in T\}$.

The fact that there is a one–one correspondence between mean zero Gaussian processes and positive definite functions allows us to very easily obtain important properties of positive definite functions. Let $\Sigma(s,t)$, $s, t \in T$ be a positive definite function and let $\{X(t), t \in T\}$ be a mean zero Gaussian process with covariance Σ. Then because $\Sigma(s,t) = \mathbb{E}\, X(s)X(t)$ it follows from the Cauchy-Schwarz inequality that

$$\Sigma(s,t) \le (\Sigma(s,s)\Sigma(t,t))^{1/2}\,.$$

Ep: Let 0 be a point in T. Then

$$\Sigma(s,t) - \frac{\Sigma(s,0)\,\Sigma(t,0)}{\Sigma(0,0)}$$

is also a positive definite function on $T \times T$. This is because it is the covariance of

$$\eta(t) = \mathbb{E}\,(X(t)|X(0)) = X(t) - \frac{\Sigma(t,0)}{\Sigma(0,0)}X(0).$$

Note that η is a Gaussian process.

Let $(\Omega, \mathbb{P})$ denote the probability space of a Gaussian process, $\{X(t), t \in T\}$. This process exists in $L^2(d\mathbb{P})$ with the inner product given by $(X(s), X(t)) = \mathbb{E}\,X(s)X(t)$, $s, t \in T$. Note that $\eta(t)$ is the projection of $X(t)$, upon the orthogonal complement of $X(0)$ with respect to $L^2(d\mathbb{P})$.

Ep: The α–potential density

$$u^{\alpha}(x,y) = \int_0^{\infty} e^{-\alpha t} p_t(x,y)\,dt,$$

of a strongly symmetric Borel right process with transition probability $p_t(x,y)$, is positive definite. This can be easily seem by using Chapman-Kolmogorov equation to rewrite

$$\begin{aligned} p_t(x,y) &= \int_{-\infty}^{\infty} p_{t/2}(x,z) p_{t/2}(z,y)\,dz \\ &= \int_{-\infty}^{\infty} p_{t/2}(x,z) p_{t/2}(y,z)\,dz. \end{aligned}$$

Therefore, to every such process we can associate a mean zero Gaussian process with covariance equal to its α–potential density $u^{\alpha}(x,y)$. We call these Gaussian processes, associated processes. Not all Gaussian processes are associated processes.

• See the coming book "Markov Processes, Gaussian Processes and Local Times" of Marcus and Rosen (2006) for a deep study of various isomorphism theorems between local time of symmetric Markov process and the associated Gaussian process.

• Are there any useful connections between Graph Laplacian and the associated Gaussian vectors? Any explanations for the Matrix-tree theorem?

Partition of Covariance Matrix

Def: The generalized inverse of a symmetric matrix $A_{n\times n}$ is the symmetric matrix

$$A^{-} = \sum_{k=1}^{k} \lambda_i^{-1}\mathbf{e}_i\mathbf{e}_i^t$$

where $\lambda_i \neq 0$ are non-zero eigenvalues of A and $\mathbf{e}_i$ are (column) unit eigenvectors of A.

•From spectral representation theorem, $A = \sum_{k=1}^{k} \lambda_i\mathbf{e}_i\mathbf{e}_i^t$ and hence $AA^{-} = A^{-}A = \sum_{k=1}^{k} e_ie_i^t$, which is just the orthogonal projection onto $R(A)$. Also, $A^{-}AA^{-} = A^{-}$ and $AA^{-}A = A$.

Thm: Consider the covariance matrix of two random vectors $X_1 \in \mathbb{R}^n$ and $X_2 \in \mathbb{R}^m$,

$$\text{Cov}(X_1, X_2) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}.$$

Then $X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2$ and X_2 are uncorrelated, and

$$\text{Var}(X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}$$

where $\Sigma_{21} = \Sigma_{12}^t$.

•Assume $(X, Y) \sim N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ in $\mathbb{R}^2$. Find $\mathbb{E}\, X^m Y^n$ for positive integer m and n.

Conditional Distributions

The basic result here gives the conditional distribution of one normal random vector given another normal random vector. It is this result that underlines many of the important distributional and independence properties of the normal and related distributions

Consider Gaussian r.v $X = (X_1, X_2) \sim N((\mu_1, \mu_2), \Sigma)$ with

$$\text{Cov}(X) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}.$$

Thus $X_i \sim N(\mu_i, \Sigma_{ii})$ for $i = 1, 2$.

Thm: Let $\mathcal{L}(X_1 | X_2 = x_2)$ denote the conditional dis. of X_1 given $X_2 = x_2$. Then

$$\begin{aligned} & \mathcal{L}(X_1 | X_2 = x_2) \\ = \ & N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{12}^t). \end{aligned}$$

Here Σ_{22}^{-} denotes the generalized inverse of Σ_{22}.
Pf: Use the fact that $X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2$ and X_2 are ind.
•Find the conditional distribution

$$\mathcal{L}((X_1, \cdots, X_n)^t | A_{m\times n} X_{n\times 1} = \mathbf{d}_{m\times 1})$$

where $X = (X_1, \cdots, X_n)^t \sim N(0, \Sigma)$ in $\mathbb{R}^n$ and rank$(A_{m\times n}) < n$.

•The conditional mean $\mathbb{E}(X_1|X_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2)$ is an affine function of x_2 (affine means a linear transformation, plus a constant vector so zero does not necessarily get mapped into zero).
•The conditional covariance $\text{Cov}(X_1|X_2 = x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{12}^{t}$ is the same as the unconditional covariance $\text{Cov}(X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2)$.
•The conditional mean and the conditional covariance determines the conditional distribution of X_1 given $X_2 = x_2$.

Ex: Assume $(X, Y, Z) \sim N(0, \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix})$ in $\mathbb{R}^3$. Then

$$\begin{aligned} \mathbb{E}(Z|X, Y) &= \frac{(\rho_2 - \rho_1\rho_2)X + (\rho_2 - \rho_1\rho_3)Y}{1 - \rho_1^2} \\ \text{Var}(Z|X, Y) &= \frac{1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3}{1 - \rho_1^2} \end{aligned}$$

Ex: Let (X, Y, Z) be as above. Then

$$\begin{aligned} \mathbb{P}(X > 0, Y > 0) &= \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho_1 \\ \mathbb{P}(X > 0, Y > 0, Z > 0) &= \frac{1}{8} + \frac{1}{4\pi}\sum_{i=1}^{3}\sin^{-1}\rho_i \end{aligned}$$

•Assume $(X,Y) \sim N(0, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix})$ in $\mathbb{R}^2$. Show that:

(a). $\mathbb{E}(X|Y) = \frac{\rho\sigma_1}{\sigma_2}Y = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}Y$,

(b). $\text{Var}(X|Y) = \sigma_1^2(1-\rho^2) = (1-\rho^2(X,Y))\text{Var}(X)$,

(c). $\mathbb{E}(X|X+Y=z) = \frac{(\sigma_1^2+\rho\sigma_1\sigma_2)z}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}$,

(d). $\text{Var}(X|Y) = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}$

•Let X and Y be ind. $N(0,1)$ r.v's, and let $Z = X+Y$. Find the distribution and density of $\mathcal{L}(Z|X>0, Y>0)$. Show that

$$\mathbb{E}(Z|X>0, Y>0) = 2\sqrt{2/\pi}.$$

Ex: Let $X \sim N(0,1)$ and let $a > 0$. Then the r.v $Y = \begin{cases} X & \text{if} |X| < a \\ -X & \text{if} |X| \geq a \end{cases}$ has the $N(0,1)$ distribution. But the pair (X,Y) is not a bivariate normal distribution.

Ex: Assume $(X,Y) \sim N(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ in $\mathbb{R}^2$. Then

$$\mathbb{E}\,(\text{sign}(X)\text{sign}(Y)) = \frac{2}{\pi}\arcsin \mathbb{E}\,(XY).$$

Pf: Write $Y = \rho X + \sqrt{1-\rho^2}\xi = X\sin\alpha + \xi\cos\alpha$ with $\alpha = \arcsin\rho$.

$\bullet$This simple representation can be sued to prove the Grothendieck inequality with the best known constant in the real case.

Ex: Assume (X,Y) is defined above and define $M = \max(X,Y)$. Then $\mathbb{E}\,M = \sqrt{(1-\rho)/\pi}$ and $\mathbb{E}\,M^2 = 1$.

Ex: Assume X_1, X_2, X_3 are ind. $N(\mu,\sigma^2)$. Then

$$\max(X_1,X_2,X_3) - \min(X_1,X_2,X_3) =^d \frac{\sqrt{3}}{2}\sum_{i=1}^{3}|X_i - \overline{X}|.$$

The Grothendieck inequality

There are more general formulations but this is the simplest one.

Thm: For any $n \in \mathbb{N}$, there is $c(n) > 1$ such that for any finite matrix $(a_{jk})_{n\times n}$ with

$$\left|\sum_{j,k=1}^{n} a_{jk} t_j s_k\right| \le \sup_{1\le j,k\le n} |t_j|\cdot|s_k|$$

and any sequence of vectors $(x_j)_{j=1}^n$, $(y_k)_{k=1}^n$ in an m-dimensional Hilbert space H one has

$$\left|\sum_{j,k=1}^{n} a_{jk}\langle x_j, y_k\rangle\right| \le c(n) \sup_{1\le j,k\le n} \|x_j\|\cdot\|y_k\|$$

Denoting the best constant $c(n)$ in the real or complex case by $K_G^{\mathbb{R}}(n)$ or $K_G^{\mathbb{C}}(n)$, and $K_G^{\mathbb{K}} = \sup_{n\in\mathbb{N}} K_G^{\mathbb{K}}(n)$ for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, one actually has $K_G^{\mathbb{R}} > K_G^{\mathbb{C}}$. The best known values are due to Krivine (1979) for $\mathbb{K} = \mathbb{R}$ and Haagerup (1987) for $\mathbb{K} = \mathbb{C}$. The exact values of $\mathbb{K}_G^{\mathbb{R}}$ and $\mathbb{K}_G^{\mathbb{C}}$ are still unknown, though the Krivine upper estimate $K_G^{\mathbb{R}} \le \frac{\pi}{2\log(1+\sqrt{2})} \approx 1.782$ has been conjectured to be exact.

Ideas of Pf: Gaussian representation and series expansion.

•The well-known Grothendieck inequality (1956) has important applications in Banach space theory and, in particular, the theory of p-summing operators and the study of unconditional convergence in Banach spaces, Lindenstrauss and Pelczynski (1968). Its generalizations including the noncommutative version for C^*-algebras by Pisier and Haagerup.

•Comme Appelé du Néant–As If Summoned from the Void: The Life of Alexandre Grothendieck, AMS Notices, two parts, Oct. and Nov. 2004, by Allyn Jackson. The life and mathematics of one of the most influential living twentieth-century mathematicians is explored.

•Grothendieck, A. (1956). Résumé de la théorie métrique des produits tensoriels topologiques, *Bol. Soc. Mat. Sao Paolo*, **8**, 1-79.

•Krivine, J. (1979). Constantes de Grothendieck et fonctions de type positif sur les spheres, *Adv. in Math.* **31**, 16-30.

•Haagerup, U. (1987). A new upper bound for the complex Grothendieck constant, *Israel J. Math.* **60**, 199-224.

•König, H. (1991). On the complex Grothendieck constant in the n-dimensional case, *Proc. Conf. Banach Spaces in Stobl*, 181-198.

•Grothendieck's Constant:
http://mathworld.wolfram.com/GrothendiecksConstant.html

Maximizing Quadratic Programs: Extending Grothendieck's Inequality

M. Charikar and A. Wirth, Princeton University

Abstract: This paper considers the following type of quadratic programming problem. Given an arbitrary matrix A, whose diagonal elements are zero, find $x \in \{-1,1\}^n$ such that $x^T Ax$ is maximized. Our approximation algorithm for this problem uses the canonical semidefinite relaxation and returns a solution whose ratio to the optimum is in $O(1/\log n)$. This quadratic programming problem can be seen as an extension to that of maximizing $x^T Ay$ (where y's components are also ± 1). Grothendieck's inequality states that the ratio of the optimum value of the latter problem to the optimum of its canonical semidefinite relaxation is bounded below by a constant. The study of this type of quadratic program arose from a desire to approximate the maximum correlation in correlation clustering. Nothing substantive was known about this problem; we present an $O(1/\log n)$ approximation, based on our quadratic programming algorithm. We can also guarantee that our quadratic programming algorithm returns a solution to the MAXCUT problem that has a significant advantage over a random assignment.

Moment Identity for Products

Here is an important moment identity for products of Gaussian random variables.

Thm: Let $X = \{X_i\}_{i=1}^k$ be an $\mathbb{R}^k$ valued Gaussian random variable with mean zero. When $k = 2m+1$ is odd,

$$\mathbb{E} \prod_{i=1}^{2m+1} X_i = 0.$$

When $k = 2m$ is even,

$$\mathbb{E} \prod_{i=1}^{k} X_i = \sum_{D_1 \cup \cdots \cup D_m = \{1,\ldots,2m\}} \prod_{i=1}^{m} \mathsf{Cov}(D_i)$$

where the sum is over all pairings $(D_1, \ldots, D_m)$ of $\{1, \ldots, 2m\}$, i.e., over all partitions of $\{1, \ldots, 2m\}$ into disjoint sets each containing two elements, and where

$$\mathsf{Cov}(\{i,j\}) := \mathsf{Cov}(X_i, X_j) := \mathbb{E}\,(X_i X_j).$$

Ex: For any mean zero Gaussian (X_1, X_2, X_3, X_4),

$$\begin{aligned}
& \mathbb{E}\,(X_1 X_2 X_3 X_4) \\
= \; & \mathbb{E}\,X_1 X_2 \cdot \mathbb{E}\,X_3 X_4 + \mathbb{E}\,X_1 X_3 \cdot \mathbb{E}\,X_2 X_4 \\
& + \mathbb{E}\,X_1 X_4 \cdot \mathbb{E}\,X_2 X_3.
\end{aligned}$$

Pf: We use the moment generating function

$$\mathbb{E}\,\exp\{\sum_{i=1}^{k}\lambda_i X_i\} = \exp\{\frac{1}{2}\mathbb{E}\,(\sum_{i=1}^{k}\lambda_i X_i)^2\}.$$

Clearly

$$\frac{\partial}{\partial\lambda_1}\cdots\frac{\partial}{\partial\lambda_k}\mathbb{E}\,\exp\{\sum_{i=1}^{k}\lambda_i X_i\}\Big|_{\lambda_1=\cdots=\lambda_k=0} = \mathbb{E}\,(\prod_{i=1}^{k}X_i).$$

Also

$$\begin{aligned}&\frac{\partial}{\partial\lambda_1}\cdots\frac{\partial}{\partial\lambda_k}\exp\{\frac{1}{2}\mathbb{E}\,(\sum_{i=1}^{k}\lambda_i X_i\}^2)\\ =\ &\sum_{n=0}^{\infty}\frac{1}{2^n n!}\frac{\partial}{\partial\lambda_1}\cdots\frac{\partial}{\partial\lambda_k}(\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\mathrm{Cov}(X_i,X_j))^n\end{aligned}$$

It is easy to see that

$$\frac{\partial}{\partial\lambda_1}\cdots\frac{\partial}{\partial\lambda_k}(\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\mathrm{Cov}(X_i,X_j))^n\Big|_{\lambda_1=\cdots=\lambda_k=0}$$

is zero when $n \neq k/2$. Thus, in particular, when $k = 2m+1$ is odd, $\mathbb{E}\,\prod_{i=1}^{2m+1}X_i = 0$.

When $k = 2m$ is even and $n = k/2 = m$, the partials is not zero only for those terms in

$$(\sum_{i=1}^{2m} \sum_{j=1}^{2m} \lambda_i \lambda_j \mathsf{Cov}(X_i, X_j))^m$$

in which each λ_i, $i = 1, \ldots, 2m$ appears only once. The terms with this property form a pairing, say $(D_1, \ldots, D_m)$ of $\{1, \ldots, 2m\}$. For this pairing it is equal to $\prod_{i=1}^m \mathrm{Cov}(D_i)$. Finally, it is easy to see that there are $2^m m!$ terms corresponding to each pairing of $\{1, \ldots, 2m\}$. This establishes the result.

Conj: For any mean zero Gaussian $(X_1, \cdots, X_n)$,

$$\mathbb{E} \prod_{k=1}^n X_j^2 \geq \prod_{k=1}^n \mathbb{E} X_j^2.$$

•Show the conjecture is true for $n = 2, 3$.

More general form of the so called diagram formula and its applications can be found in following references:

•Major, P. *Multiple Wiener-Itô integrals*, Lecture Notes in Math., **849**, Springer, 1981.

•Arcones, M. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22**, 2242–2274.

•Houdré, C. and Pérez-Abreu, V. (1994). Chaos expansions, multiple Wiener-Itô integrals and their applications.

Exponential Moments Via Gaussian

Here is a well-known inequality which is proved by introducing an independent Gaussian r.v.

Fact: For $0 \leq \lambda < 1/2$ and $\sum x_i^2 = 1$,

$$\mathbb{E} \exp\left\{\lambda \left(\sum x_i \varepsilon_i\right)^2\right\} \leq \mathbb{E} \exp\left\{\lambda \left(\sum x_i \xi_i\right)^2\right\} = \frac{1}{\sqrt{1-2\lambda}}$$

where ε_i are ind. r.v's with $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$ and ξ_i are ind. $N(0,1)$ r.v's. This is an example of comparison between ± 1 and Gaussian.

Pf: First notice that for any fixed ε_i, $i = 1, \cdots n$,

$$\exp\left\{\lambda \left(\sum x_i \varepsilon_i\right)^2\right\} = \mathbb{E}_{\xi} \exp\left\{\sqrt{2\lambda}\left(\sum x_i \varepsilon_i\right)\xi\right\}$$

where ξ is a standard Gaussian and $\mathbb{E}_{\xi}$ denotes expectation with respect to ξ. Using the independence we see that

$$\begin{aligned}
\mathbb{E} \exp\left\{\lambda \left(\sum x_i \varepsilon_i\right)^2\right\} &= \mathbb{E}\,\mathbb{E}_{\xi} \exp\left\{\sqrt{2\lambda}\left(\sum x_i \varepsilon_i\right)\xi\right\} \\
&= \mathbb{E}_{\xi} \prod \mathbb{E}_{\varepsilon_i} \exp\left\{\sqrt{2\lambda} x_i \varepsilon_i \xi\right\} \\
&= \mathbb{E}_{\xi} \prod \cosh\left(\sqrt{2\lambda} x_i \xi\right) \\
&\leq \mathbb{E}_{\xi} \prod \exp\{\lambda x_i^2 \xi^2\} \\
&= \mathbb{E}_{\xi} \exp\{\lambda \xi^2\} = \frac{1}{\sqrt{1-2\lambda}}
\end{aligned}$$

where we have used the easily verified inequality $\cosh(x) \leq \exp\{x^2/2\}$.

Other use of the fact that one can majorize the expression by a similar expression for appropriate Gaussian's can be found in references below:

•Talagrand, M. (1998), The Sherrington-Kirkpatrick model: a challenge for mathematicians. Probability Theory and Related Fields, **110**, 109-176.
•de la Peña, V. and Giné, E. (1999). *Decoupling, from Dependence to Independence*.
•Achlioptas, D. (2003). Database-friendly random projections. Proc. of PODS 01, 274-281.
•Talagrand, M. (2003), *Spin glasses: A challenge for mathematicians* .

Ex: On the index set $T = \{-1, +1\}^n$ consider the Gaussian process

$$X(\sigma_1, \cdots, \sigma_n) = -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} \xi_{ij} \sigma_i \sigma_j$$

where (ξ_{ij}) is a family of $n(n-1)/2$ i.i.d $N(0,1)$. They model disorder in a system of n spins $\sigma_1, \cdots, \sigma_n = \pm 1$, namely, $X(\sigma_1, \cdots, \sigma_n)$ is the energy of the spin configuration $\sigma_1, \cdots, \sigma_n$. This is the Sherrington-Kirkpatrick model for spin glasses, well-known in statistical physics. It ignores the geometric location of atoms, assuming that all pairs interact in the same way ('mean field approximation').

Quadratic Functionals

Ref: X. Chen and W.V. Li (2003), Quadratic functionals and small ball probabilities for the m-fold integrated Brownian motion, *Annals of Probability*, **31**, 1052-1077.

Abstract: Let the Gaussian process $X_m(t)$ be the m-fold integrated Brownian motion for positive integer m. The Laplace transform of the quadratic functional of $X_m(t)$ is found by using an appropriate self-adjoint integral operator. The result is then used to show the power of a general connection between small ball probabilities for Gaussian process. The connection is discovered by introducing an independent random shift. Various interplay between our results and principal eigenvalues for non-uniform elliptic generators on an unbounded domain are discussed.

Basic ideas: Let $W(t)$, $t \geq 0$, be the standard Brownian motion starting at 0. Denote by $X_0(t) = W(t)$ and

$$X_m(t) = \int_0^t X_{m-1}(s)ds, \quad t \geq 0, \quad m \geq 1.$$

the m-fold integrated Brownian motion for positive integer m. Upon integration by parts we also have the representation

$$X_m(t) = \frac{1}{m!}\int_0^t (t-s)^m dW(s), \quad m \geq 0.$$

To motivate and show how we find a different operator to work with, we introduce an auxiliary independent Brownian motion $\widetilde{W}(t)$. By a stochastic Fubini theorem, we have

$$\begin{aligned}
\phi(\theta) &= \mathbb{E}\exp\Big(-\frac{\theta^2}{2}\int_0^1 X_m^2(t)\,dt\Big) \\
&= \mathbb{E}\exp\Big(i\theta\int_0^1 X_m(t)\,d\widetilde{W}(t)\Big) \\
&= \mathbb{E}\exp\Big(i\theta\int_0^1\int_0^t X_{m-1}(s)\,ds\,d\widetilde{W}(t)\Big) \\
&= \mathbb{E}\exp\Big(i\theta\int_0^1 X_{m-1}(s)(\widetilde{W}(1)-\widetilde{W}(s))\,ds\Big) \\
&= \mathbb{E}\exp\Big(i\theta\int_0^1 X_{m-1}(s)\widetilde{W}(1-s)\,ds\Big)
\end{aligned}$$

since $\{\widetilde{W}(1) - \widetilde{W}(1-t);\ 0 \le t \le 1\}$ is Brownian motion independent of $\{X_{m-1}(t);\ 0 \le t \le 1\}$.

Next we can write the the convolution in the form

$$
\begin{aligned}
& \int_0^1 X_{m-1}(u)\widetilde{W}(1-u)\,du \\
= & \int_0^1 \int_0^u \frac{1}{(m-1)!}(u-s)^{m-1} dW(s) \int_0^{1-u} d\widetilde{W}(t)\,du \\
= & \int_0^1 \int_0^1 \frac{1}{(m-1)!} \\
& \left(\int_0^1 1_{(0,u)}(s)(u-s)^{m-1} 1_{(0,1-u)}(t)\,du \right) dW(s) d\widetilde{W}(t) \\
= & \int_0^1 \int_0^1 K(s,t) dW(s) d\widetilde{W}(t)
\end{aligned}
$$

where

$$
m!K(s,t) = \max(0, 1-s-t)^m = (1-s-t)_+^m.
$$

More details can be found in the paper.

•Let ξ_i be i.i.d $N(0,1)$. Show the following is true for $n = 2, 3, 4$.

$$
\mathbb{E} \prod_{1 \le i < j \le n} (\xi_i - \xi_j)^2 = \prod_{j=1}^{n} j!
$$

Can you prove the result for any $n \geq 2$?

Isoperimetry

The main geometric property of both measures (Lebesgue and Gaussian) is an isoperimetric inequality. For Lebesgue measure it is classical (J. Steiner 1842, H. Schwarz 1884). Among all bodies of a given volume, a ball minimizes the surface area. For Gaussian measure, an isoperimetric inequality has appeared independently in the work

•V.N. Sudakov, B.S. Tsirelson, "Extremal properties of half-spaces for spherically invariant measures", Zapiski LOMI **41** (1974), 14-24 (Russian); Journal of Soviet Mathematics **9**, (1978), 9-18 (English).
•C. Borell, "The Brunn-Minkowski inequality in Gauss space", Invent. Math. **30** (1975), 207-216.

Statement: Among all sets of a given Gaussian measure, a half-space minimizes the Gaussian measure of a neighborhood.

Many (better) proofs have appeared later, see A. Ehrhard (1983), M. Ledoux (1994), S. Bobkov (1997), C. Borell (2005).

The Classical Isoperimetric Inequality on $\mathbb{R}^n$

The classical isoperimetric inequality in $\mathbb{R}^n$ asserts that among all compact sets A in $\mathbb{R}^n$ with smooth boundary ∂A and with fixed volume, Euclidean balls are the ones with the minimal surface measure. In other words, whenever $\mathrm{Vol}_n(A) = \mathrm{Vol}_n(B)$ where B is a ball (and $n > 1$),

$$\mathrm{Vol}_{n-1}(\partial A) \geq \mathrm{Vol}_{n-1}(\partial B)$$

There is an equivalent formulation of this result in terms of isoperimetric neighborhoods or enlargements which in particular avoids surface measures and boundary considerations; namely, if $A_r = A + rB$ denotes the (closed) Euclidean neighborhood of A of order $r \geq 0$, and if B is as before a ball with the same volume as A, then, for every $r \geq 0$,

$$\mathrm{Vol}_n(A_r) = \mathrm{Vol}_n(B_r).$$

The equivalence between follows from the Minkowski content formula

$$\mathrm{Vol}_{n-1}(\partial A) = \liminf_{r \to 0} \frac{1}{r} \left(\mathrm{Vol}_n(A_r) - \mathrm{Vol}_n(A)\right)$$

(whenever the boundary ∂A of A is regular enough). Actually, if we take the latter as the definition of $\mathrm{Vol}_{n-1}(\partial A)$, it is not too difficult to see the equivalence for every Borel set A (see an argument for the Gaussian case later). The simplest proof of this isoperimetric inequality goes through the Brunn-Minkowski inequality which states that if A and B are two compact sets in $\mathbb{R}^n$, then

$$\mathrm{Vol}_n(A+B)^{1/n} \geq \mathrm{Vol}_n(A)^{1/n} + \mathrm{Vol}_n(B)^{1/n}.$$

To deduce the isoperimetric inequality from the Brunn-Minkowski inequality, let $r_0 > 0$ be such that $\mathrm{Vol}_n(A) = \mathrm{Vol}_n(B(0, r_0))$. Then,

$$\begin{aligned}
\mathrm{Vol}_n(A_r)^{1/n} &= \mathrm{Vol}_n(A + B(0,r))^{1/n} \\
&\geq \mathrm{Vol}_n(A)^{1/n} + \mathrm{Vol}_n(B(0,r))^{1/n} \\
&= (r_0 + r)\mathrm{Vol}_n(B(0,1))^{1/n} \\
&= \mathrm{Vol}_n(B(0,r_0)_r)^{1/n}.
\end{aligned}$$

For a briefly sketch of the proof of the Brunn-Minkowski inequality and related reference, see Ledoux (1996).

The Classical Isoperimetric Inequality on $\mathbb{S}_\rho^{n-1}$

The classical isoperimetric inequality on spheres expresses that among all subsets with fixed volume on a sphere, geodesic balls (caps) achieve the minimal surface measure. In other words, whenever $\sigma_{n-1}^\rho(A) = \sigma_{n-1}^\rho(B)$ where B is a cup on the sphere $\mathbb{S}_\rho^{n-1} \subset \mathbb{R}^n$ of radius $\rho > 0$,

$$\sigma_{n-1}^\rho(A_r) = \sigma_{n-1}^\rho(B_r)$$

where $A_r = \{x \in \mathbb{S}^{n-1} : d(x, A) \leq r\}$, $d(\cdot,\cdot)$ is the geodesic distance on $\mathbb{S}^{n-1}$, and σ_{n-1}^ρ is the normalized rotation invariant measure on $\mathbb{S}_\rho^{n-1}$. This inequality has been established independently by E. Schmidt (1948) and P. Lévy (1951) for sets with smooth boundaries. Schmidt's proof is based on the classical isoperimetric rearrangement or symmetrization techniques due to J. Steiner Lévy's argument, which applies to more general types of surfaces, uses the modern tools of minimal hypersurfaces and integral currents. His proof has been generalized to Riemannian manifolds with positive Ricci curvature by M. Gromov (1980), see Ledoux (1996).

The Gaussian Isoperimetric Inequality

Gaussian Isoperimetric Inequality states that for all measurable sets in $\mathbb{R}^n$ with the same canonical Gaussian measure $\gamma = \gamma_n$, half–spaces achieve the minimal surface measure with respect to γ.

Let A be a Borel set in $\mathbb{R}^n$. Define $A_r = \{x + rB, x \in A\}$ where B is the open unit ball in $\mathbb{R}^n$. The surface measure of A is defined as

$$\gamma_S(\partial A) = \liminf_{r \to 0} \frac{1}{r}(\gamma(A_r) - \gamma(A)).$$

Let H be the half–space $\{x \in \mathbb{R}^n : (x, u) < a\}$, for some $u \in \mathbb{R}^n$ with $|u| = 1$ and $a \in [-\infty, \infty]$. Let $A \subset \mathbb{R}^n$ be such that $\gamma(A) = \gamma(H)$. The isoperimetric inequality just referred to is that

$$\gamma_S(\partial A) \geq \gamma_S(\partial H).$$

It is convenient to express the result only in terms of A. Since $\gamma(H) = \Phi(a)$ and $\gamma_S(\partial H) = \phi(a)$ we can rewrite the inequality as

$$\gamma_S(\partial A) \geq \phi(a) = \phi \circ \Phi^{-1}(\gamma(A)) \qquad (*)$$

with equality when $A = H$.

It is more useful for us to have an integrated version of this isoperimetric inequality.

Thm: (Borell, Sudakov-Tsirelson) Let A be a measurable subset of $\mathbb{R}^n$ such that

$$\gamma(A) = \gamma(H) = \Phi(a), \qquad -\infty \leq a \leq \infty.$$

Then, for $r \geq 0$

$$\gamma(A + rB) \geq \gamma(H + rB) = \Phi(a + r) \qquad (**)$$

where B is the open unit ball in $\mathbb{R}^n$.

Standard Reduction: Let $\mathcal{C}$ denote the sets in $\mathbb{R}^n$ that are finite unions of open balls. Note that for $C \in \mathcal{C}$, the $\liminf$ in (*) is actually a limit, so that $\dfrac{d}{dr}\gamma(C_r) = \gamma_S(\partial C_r)$. We show that both (*) and (**) hold if one of them holds for all $C \in \mathcal{C}$.

(1). Consider the function $v(r) = \Phi^{-1}(\gamma(C_r))$, $r \geq 0$. It follows from (*) that for $C_r \in \mathcal{C}$

$$v'(r) = \frac{\gamma_S(\partial C_r)}{\phi \circ \Phi^{-1}(\gamma(C_r))} \geq 1.$$

Therefore $\int_0^r v'(u)\,du \geq r$ and consequently $v(r) \geq v(0) + r$. Since $v(0) = a$ we get (**) for sets in $\mathcal{C}$.

(2). Suppose that $A \subset \mathbb{R}^n$ is open. Then we can find an increasing sequence of sets C_n in $\mathcal{C}$ such that $C_n \subset A$ and $\gamma(C_n)$ increases to $\gamma(A)$. Applying (**) to each C_n and taking the limit we see that (**) holds for open sets.

(3). Now let A be a measurable set in $\mathbb{R}^n$. Let $\rho > 0$. Then A_ρ is an open set and

$$\gamma(A_{\rho+r}) = \gamma((A_\rho)_r) \geq \Phi(a_\rho + r),$$

where a_ρ is such that $\gamma(A_\rho) = \Phi(a_\rho)$. Taking the limit as ρ goes to zero we get $\gamma(A_r) \geq \Phi(\tilde{a} + r)$, where $\tilde{a}$ is such that $\Phi(\tilde{a}) = \lim_{\rho\to 0} \gamma(A_\rho) \geq \gamma(A)$. Thus $\tilde{a} \geq a$ and we get (**) in the general case.

(4). The general case of (**) implies the general case of (*).

Poincaré's Limit

The Gaussian isoperimetric inequality may be considered as the limit of the isoperimetric inequality on the spheres $\mathbb{S}_\rho^{n-1}$ with $\rho = \sqrt{n}$ It has been known for a long time that the normalized measures $\sigma_{n-1}^{\sqrt{n}}$ on $\mathbb{S}_{\sqrt{n}}^{n-1}$, projected on a fixed subspace $\mathbb{R}^d$, converge when n goes to infinity to the canonical Gaussian measure γ_d on $\mathbb{R}^d$. To be more precise, denote by $\Pi_{n,d}$, $n > d$, the projection from $\mathbb{R}^n$ onto $\mathbb{R}^d$.

Fact: For every Borel set A in $\mathbb{R}^d$,

$$\lim_{n\to\infty} \sigma_{n-1}^{\sqrt{n}} \left(\Pi_{n,d}^{-1}(A) \cap \mathbb{S}_{\sqrt{n}}^{n-1} \right) = \gamma_d(A).$$

This fact is commonly known as Poincaré's lemma or Poincaré's limit although it does not seem to be due to H. Poincaré, see Diaconis and Freedman (1987). The convergence is better than only weak convergence of the sequence of measures $\Pi_{n,d}^{-1}(\sigma_{n-1}^{\sqrt{n}})$ to γ_d. Simple analytic or probabilistic proofs may be found in the literature.

Idea of Pf:. Let $(\xi_i)_{i\geq 1}$ be i.i.d $N(0,1)$. For every integer $n \geq 1$, set $R_n^2 = \xi_1^2 + \cdot + \xi_n^2$. Now,

$$\frac{\sqrt{n}}{R_n} \cdot (\xi_1, \cdots, \xi_n) =^d \sigma_{n-1}^{\sqrt{n}},$$

and thus $\frac{\sqrt{n}}{R_n} \cdot (\xi_1, \cdots, \xi_d)$ is equal in distribution to $\Pi_{n,d}(\sigma_{n-1}^{\sqrt{n}})$, $n > d$. Since $R_n^2/n \to 1$ almost surely by the strong law of large numbers, we already get the weak convergence result. The fact is however stronger since convergence is claimed for every Borel set. See Ledoux (1996) for more details. In particular, the proof is easily modified to actually yield uniform convergence of densities on compacts sets and in the variation metric.

•As we have seen, caps are the extremal sets of the isoperimetric problem on spheres. Now, a cap may be regarded as the intersection of a sphere and a half-space, and, by Poincaré's limit, caps will thus converge to half-spaces. This is why half-spaces are the extremal sets of the isoperimetric problem for Gaussian measures.

A Functional Inequality

Thm: For all Lipschitz functions, including sufficiently smooth functions $f : \mathbb{R}^n \longrightarrow [0,1]$

$$\mathcal{U}(\int f\, d\gamma) \le \int \sqrt{\mathcal{U}^2(f) + |\bigtriangledown f|^2}\, d\gamma$$

where $|\bigtriangledown f|$ denotes the Euclidean length of the gradient of f.
Pf of (*): This is based on the above functional inequality for set A in $\mathcal{C}$ (so $\gamma(\partial A) = 0$). Set

$$f_r(x) := (1 - \frac{\text{dist}(x, A)}{r})^+, \quad r > 0.$$

Let $\bar{A}$ be the closure of A. If $x \in \bar{A}$, $f_r(x) = 1$ for all $r > 0$. If $x \in \bar{A}^c$, $\lim_{r\to 0} f_r(x) = 0$. Thus $\lim_{r\to 0} f_r(x) = I_{\bar{A}}(x)$ and $\lim_{r\to 0}\mathcal{U}(f_r(x)) = 0$. Moreover, $|\bigtriangledown f_r| = 0$ on A and on the complement of the closure of A_r and $|\bigtriangledown f_r| \le 1/r$ everywhere. Take the limit inferior as $r \to 0$ in the functional inequality. Since $\gamma(A) = \gamma(\bar{A})$ and $\gamma(\partial A_r) = 0$ for all r, we get

$$\begin{aligned}
\mathcal{U}(\gamma(A)) &\le \liminf_{r\to 0} \int \mathcal{U}(f_r)\, d\gamma + \liminf_{r\to 0} \int |\bigtriangledown f_r|\, d\gamma \\
&= \liminf_{r\to 0} \int |\bigtriangledown f_r|\, d\gamma \\
&\le \liminf_{r\to 0} \frac{1}{r}(\gamma(A_r) - \gamma(A)) = \gamma_S(\partial A)
\end{aligned}$$

The Ornstein-Uhlenbeck (Hermite) Semigroup

To do this we use the Ornstein–Uhlenbeck or Hermite semigroup with invariant measure $\gamma = \gamma_n$. For $f \in L^1(\gamma)$ set

$$\begin{aligned} P_t f(x) &= \int_{\mathbb{R}_n} f(e^{-t}x + (1 - e^{-2t})^{1/2} y)\, d\gamma(y) \\ &= \mathbb{E}\, f(e^{-t}x + (1 - e^{-2t})^{1/2} Y) \end{aligned}$$

where $Y \sim N(0, I_n)$, $x \in \mathbb{R}^n$, $t > 0$. We note the following properties of $(P_t)_{t \geq 0}$.

Lemma: (i). The operators P_t are contractions on the functions spaces $L^p(\gamma)$ for all $p \geq 1$ and the operator norm $\|P_t\| = 1$.
ii). For all sufficiently smooth integrable functions f and g and every $t > 0$

$$\int f P_t g\, d\gamma = \int g P_t f\, d\gamma.$$

(iii). $(P_t)_{t \geq 0}$ is a semi–group of operators, i.e. $P_s \circ P_t = P_{s+t}$. P_0 is the identity and for any $f \in L^1(\gamma)$, $\lim_{t \to \infty} P_t f = \int f\, d\gamma$.
(iv). For $f \in L^1(\gamma)$, $\mathbb{E}\, P_t f(Y) = \mathbb{E}\, f(Y)$ and hence γ is an unique invariant measure for P_t.

Idea of Pf: For any $0 \le \alpha \le 1$ set $\tilde{\alpha} = (1-\alpha^2)^{1/2}$. Then let $\alpha = e^{-t}$ and Y be a standard normal random variable. We can write

$$P_t f(x) = \mathbb{E}\,(f(\alpha x + \tilde{\alpha} Y)).$$

Let X be a $N(0,1)$ independent of Y and note that $Z = \alpha X + \tilde{\alpha} Y$ is also a $N(0,1)$ We have

$$\begin{aligned} \|P_t f\|_p &= (\mathbb{E}_X(|\mathbb{E}_Y(f(\alpha X + \tilde{\alpha} Y))|^p))^{1/p} \\ &= (\mathbb{E}\,(|\mathbb{E}\,(f(Z)\,|\,Y)|^p))^{1/p} \\ &\le (\mathbb{E}\,(\mathbb{E}\,(|f(Z)\,|^p|\,Y)))^{1/p} = \|f\|_p \end{aligned}$$

which is (i). (In the last line we use the conditional Hölder's inequality).
To obtain (ii) we note that

$$\int g P_t f \, d\gamma = \mathbb{E}\, g(X) f(\alpha X + \tilde{\alpha} Y) = \mathbb{E}\, g(X) f(Z).$$

Here (X, Z) is a Gaussian random variable, $\mathbb{E}\, X^2 = \mathbb{E}\, Z^2 = 1$ and $\mathbb{E}\,(XZ) = \alpha$. Clearly then $\mathbb{E}\, g(X) f(Z) = \mathbb{E}\, g(Z) f(X)$ which is *(ii)*.
When $\beta = e^{-s}$ we have that

$$\begin{aligned} P_t \circ P_s f(x) &= \mathbb{E}\,(P_s f(\alpha x + \tilde{\alpha} Y)) \\ &= \mathbb{E}\,(f(\beta \alpha x + \beta \tilde{\alpha} Y + \tilde{\beta} X)). \end{aligned}$$

Since $\beta \tilde{\alpha} Y + \tilde{\beta} X$ is normal with variance $1 - (\alpha\beta)^2$ we see that $P_s \circ P_t = P_{s+t}$. The rest is easy.

Infinitesimal Operator

Let L be the infinitesimal operator for the semi–group $(P)_{t\geq 0}$. That is,

$$Lf = \lim_{t\to 0+} \frac{P_t f - f}{t}$$

and thus L satisfies

$$\frac{d}{dt} P_t f = P_t L f = L P_t f$$

for all sufficiently smooth functions $f \in \mathbb{R}^n$. One can check that

$$Lf(x) = \triangle f(x) - (x, \triangledown f(x)).$$

It follows, by repeatedly integrating by parts on each component of $\mathbb{R}^n$, that

$$-\int f(x)(Lg(x))\, d\gamma(x) = \int (\triangledown f(x), \triangledown g(x))\, d\gamma(x).$$

Let Z be a standard n-variate normal vector and define $\Phi h = \mathbb{E}\, h(Z)$.

Lemma: Let $h : \mathbb{R}^n \to \mathbb{R}$ have three bounded derivatives. Then

$$g(x) = -\int_0^\infty (P_t h(x) - \Phi h) dt$$

solves

$$\text{tr} D^2 g(x) - x \cdot \bigtriangledown g(x) = h(s) - \Phi h,$$

and for any k-th partial derivative we have the bound

$$\left| \frac{\partial^k}{\prod_{j=1}^k \partial x_{i_j}} g(x) \right| \leq \frac{1}{k} \|D^k h\|.$$

Further, for any $y \in \mathbb{R}^n$ and positive definite $n \times n$ matrix Σ, f defined by the change of variable

$$f(x) = g(\Sigma^{-1/2} 1/2 (x - y))$$

solves

$$\text{tr} \Sigma D^2 f(x) - (x - y)) \cdot \bigtriangledown f(x) = h(\Sigma^{-1/2}(x - y))) - \Phi h,$$

and hence

$$\left| \frac{\partial^k}{\prod_{j=1}^k \partial x_{i_j}} f(x) \right| \leq \frac{n^k}{k} \|\Sigma^{-1/2}\|^k \|D^k h\|.$$

Idea of Pf: One can follow Barbour (1990) to show that g is a solution, and that under the assumptions above, by dominated convergence,

$$D^k g(x) = -\int_0^\infty e^{-kt}\mathbb{E}\,(D^k h(e^{-t}x + (1 - e^{-2t})^{1/2} Z) dt.$$

The Lemma now follows by straightforward calculations.

•Barbour, A.D. (1990) Steins method for diffusion approximations, Probab. Th. Rel. Fields **84** 297-322.

•Götze, F. (1991) On the rate of convergence in the multivariate CLT. Annals of Probability, **19**, 724-739.

•Goldstein, L and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings *Journal of Applied Probability*, **33**, 1-17.

Proof of the Functional Inequality

Let $0 \leq f \leq 1$ be a smooth function on $\mathbb{R}^n$. We assume that $0 < P_t f < 1$. This is the case unless f is equal to one or zero on a set of γ measure one, in which case the inequality is satisfied. To verify the inequality when $0 < P_t f < 1$ it suffices to show that the function

$$F(t) = \int \sqrt{\mathcal{U}^2(P_t f) + |\bigtriangledown P_t f|^2}\, d\gamma$$

is non–increasing in $t \geq 0$, since if this is true then $F(\infty) \leq F(0)$ which implies the inequality by property (iii) of $(P_t)_{t \geq 0}$.

Showing that the derivative of F is less that or equal to zero is an exercise in elementary calculus.

Gaussian Logarithmic Sobolev Inequality

The Gaussian logarithmic Sobolev inequality, due to L. Gross (1967, 1975), has numerous applications in analysis, geometry and stochastics. In the following formulation, we use the convention $f^2(x)\log f(x) = 0$ if $f(x) = 0$

Thm: For every function $f \in W^{2,1}(\gamma_n)$,

$$\int_{\mathbb{R}^n} f^2 \log f d\gamma_n \leq \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n + \frac{1}{2}\int_{\mathbb{R}^n} f^2 d\gamma_n \log \int_{\mathbb{R}^n} f^2 d\gamma_n$$

and equality holds for Gaussian density.
Idea of Pf: Suppose that $f \in C_0^\infty(\mathbb{R}^n$ and $f \geq c > 0$. Put $g = f^2$. Then it is equivalent to show

$$\int_{\mathbb{R}^n} g \log g d\gamma_n - \int_{\mathbb{R}^n} g d\gamma_n \log \int_{\mathbb{R}^n} g d\gamma_n \leq \frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla g|^2}{g} d\gamma_n$$

Note that $P_t g \geq c^2$. Since we have

$$P_t g \log P_t g \to \int_{\mathbb{R}^n} g d\gamma_n \log \int_{\mathbb{R}^n} g d\gamma_n$$

the left side above can be represented as, by the semigroup property,

$$
\begin{aligned}
& -\int_0^\infty \left(\frac{d}{dt}\int_{\mathbb{R}^n} P_tg \log P_tg d\gamma_n\right) dt \\
= & -\int_0^\infty \left(\int_{\mathbb{R}^n} (LP_tg) \log P_tg d\gamma_n\right) dt \\
& \quad -\int_0^\infty \left(\int_{\mathbb{R}^n} P_tg \cdot \frac{1}{P_tg}\frac{d}{dt} P_tg d\gamma_n\right) dt \\
= & -\int_0^\infty \left(\int_{\mathbb{R}^n} (LP_tg) \log P_tg d\gamma_n\right) dt \\
= & \int_0^\infty \left(\int_{\mathbb{R}^n} \langle \triangledown P_tg, \triangledown (\log P_tg)\rangle d\gamma_n\right) dt \\
= & \int_0^\infty \left(\int_{\mathbb{R}^n} \frac{1}{P_tg} | \triangledown P_tg|^2 d\gamma_n\right) dt
\end{aligned}
$$

next we use the identity $\triangledown P_tg = e^{-t}P_t(\triangledown g)$ and Holder's inequality,

$$
\frac{1}{P_tg}|P_t \partial_{x_i} g|^2 \leq P_t\left(\frac{1}{g}(\partial_{x_i} g)^2\right)
$$

which finishes the proof since $\int P_th d\gamma = \int g d\gamma$ for any $h \in C_0^\infty(\mathbb{R}^n)$.

•Gross, L. (1993). Logarithmic Sobolev inequalities and contractivity properties of semigroups. *Lecture Notes in Math.* **1563**, 54-88.

Perelman's Proof of Poincaré Conjecture

In two places in the proof of Perelman (2002+), section 3 and section 10, explicit constructions lead to contradiction of the Gaussian logarithmic Sobolev inequality.

•Grisha Perelman (2002+). The entropy formula for the Ricci flow and its geometric applications, math.DG/0211159

•Grisha Perelman (2003+). Ricci flow with surgery on three-manifolds, math.DG/0303109

•John W. Morgan (2005). Recent progress on the Poincaré conjecture and the classification of 3-manifolds *Bull. Amer. Math. Soc.* **42**, 57-78.

•Notes and commentary on Perelman's Ricci flow papers, http://www.m
research/ricciflow/perelman.html

Gaussian Methods:
Tools, Applications and Problems

Lectures given in Peking University, July, 2005

Abstract: In this series of lectures, we will provide an overview of some very power Gaussian methods. Basic Gaussian tools will be introduced in the context of various applications to diverse areas of probability and mathematics. The main goal is to show the beauty and the usefulness of the field.

The theory of Gaussian vectors/processes is of fundamental importance in probability and its development should be centered on (i) applications of the existing methods to a variety of fields, (ii) new techniques and problems motivated by current interests of advancing knowledge.

•The next 26 pages come from page 43-54, 70-83 of my Beijing lecture notes.

Fundamental Gaussian Inequalities

Gaussian inequalities, whose goal, loosely speaking, consists of searching for an inequality between dependent (complicated) and independent (simpler) structures that becomes an equality in certain (possibility limiting) cases. We will present several recent conjectures for Gaussian measure/vectors and their correctness in lower dimension.

- Isoperimetric inequalities: Gaussian Isoperimetric inequalities; Ehrhard's inequality; Shift inequalities; S-inequality; B-inequality; etc.
- Comparison inequalities: Anderson's inequality; Slepian's inequality; Gordon's min-max inequalities; Reverse Slepian type inequalities; etc.
- Correlation inequalities and conjectures: Sidak inequality; Weak correlation inequality; etc.
- Concentration and deviation inequalities: Dudley, Fernique, Berman, Talagrand, etc.
- Functional inequalities: Logarithmic Sobolev inequality; Bobkov's inequality; etc.

• Gaussian Isoperimetric inequality: For any Borel set A in $\mathbb{R}^n$ and a half space $H = \{x \in \mathbb{R}^n : \langle x, u \rangle \leq a\}$ such that $\gamma_n(A) \geq \gamma_n(H) = \Phi(a)$ for some real number a and some unit vector $u \in \mathbb{R}^n$, we have for every $r \geq 0$

$$\gamma_n(A + rU) \geq \gamma_n(H + rU) = \Phi(a + r),$$

where U is the unit ball in $\mathbb{R}^n$ and $A + rU = \{a + ru : a \in A, u \in U\}$.

• The above isoperimetric inequality has played a fundamental role in topics such as integrability and upper tail behavior of Gaussian semi-norms, deviations and concentrations, and small ball probabilities.

• There are at least four different proofs.

• The result is due independently to Borell (1975) and Sudakov and Tsirelson (1974). The standard proof is based on the classic isoperimetric inequality on the sphere and the fact that the standard Gaussian distribution on $\mathbb{R}^n$ can be approximated by marginal distributions of uniform laws on spheres in much higher dimensions. The approximation procedure, so called Poincare limit, can be found in Ledoux (1996), chapter 1.

A direct proof based on the powerful Gaussian symmetrization techniques is given by Ehrhard (1983). This also led him to a rather complete isoperimetric calculus in Gauss space, see Ehrhard (1984, 1986). In particular, he obtained the following remarkable Brunn-Minkowski type inequality with both sets A and B convex.

• Ehrhard's inequality: For any Borel sets A and B of $\mathbb{R}^n$, and $0 \leq \lambda \leq 1$,

$$\begin{aligned} & \Phi^{-1} \circ \gamma_n(\lambda A + (1-\lambda)B) \\ \geq \ & \lambda \Phi^{-1} \circ \gamma_n(A) + (1-\lambda)\Phi^{-1} \circ \gamma_n(B). \end{aligned}$$

where $\lambda A + (1-\lambda)B = \{\lambda a + (1-\lambda)b : a \in A, b \in B\}$.

• The case of one convex set and one Borel set is due to Latala (1996). A special case was studied in Kuelbs and Li (1995).

• Borell (2004+): The above inequality holds for any Borel sets A and B.

Prob: Find a direct proof in $\mathbb{R}$ which can be extended to $\mathbb{R}^n$ by known method.

Ehrhard's inequality is a delicate result, which implies the isoperimetric inequality for Gaussian measures and has some other interesting consequences as well. It also improve upon the more classical so called log-concavity of Gaussian measures:

$$\log \mu(\lambda A + (1-\lambda)B) \geq \lambda \log \mu(A) + (1-\lambda) \log \mu(B).$$

A proof of the above result may be given using again the Poincare limit on the classical Brunn-Minkowski inequality on $\mathbb{R}^n$; see Ledoux and Talagrand (1991) for details.

Other than using addition of sets as enlargement, multiplication to a set can also be considered. The following result is due to Landau and Shepp (1970).

• For any convex set A in $\mathbb{R}^n$ and a half space $H = \{x \in \mathbb{R}^n : \langle x, u \rangle \leq a\}$ such that $\gamma_n(A) \geq \gamma_n(H) = \Phi(a)$ for some $a \geq 0$ and some unit vector $u \in \mathbb{R}^n$, one has for every $r \geq 1$

$$\gamma_n(rA) \geq \gamma_n(rH) = \Phi(ra),$$

where $rA = \{rx : x \in A\}$.

The proof is based on the Brunn-Minkowski inequality on the sphere without using the Poincare limit. An application is the exponential square integrability of the norm of a Gaussian measure.

For a symmetric convex set A, the following was conjectured by Shepp in 1969 (so called S-conjecture) and proved recently by Latała and Oleszkiewicz (1999).

• S-inequality: Let μ be a centered Gaussian measure on a separable Banach space E. If A is a symmetric, convex, closed subset of E and $S \subset E$ is a symmetric strip, i.e., $S = \{x \in E : |x^*(x)| \leq 1\}$ for some $x^* \in E^*$, the dual space of E, such that $\mu(A) = \mu(S)$, then $\mu(tA) \geq \mu(tS)$ for $t \geq 1$ and $\mu(tA) \leq \mu(tS)$ for $0 \leq t \leq 1$.

Open: Bounds of $\gamma_n(A + rU)$ for symmetric and convex set A. Talagrand (1992, 1993) has provided very sharp upper and lower estimates for $\gamma_n(A+rU)$ when r is large and A is convex symmetric. In particular, the estimates relate to the small ball problem and its link with metric entropy discovered by Kuelbs and Li (1992).

• Shift inequality of Kuelbs and Li (1998): If A is a Borel subset of $\mathbb{R}^n$, $h \in \mathbb{R}^n$, $H_- = \{x : \langle x, h\rangle \le a\}$, and $H_+ = \{x : \langle x, h\rangle \ge b\}$ where a and b are such that $\gamma_n(A) = \gamma_n(H_-) = \gamma_n(H_+)$, then

$$\gamma_n(H_+ + h) \le \gamma_n(A + h) \le \gamma_n(H_- + h).$$

More generally, if $A \subseteq \{x : a \le \langle x, h\rangle \le d\}$. and $S_- = \{x : a \le \langle x, h\rangle \le b\}$, $S_+ = \{x : c \le \langle x, h\rangle \le d\}$, are such that $\gamma_n(S_-) = \gamma_n(S_+) = \gamma_n(A)$, then $\gamma_n(S_+ + h) \le \gamma_n(A + h) \le \gamma_n(S_- + h)$.

• **Monotone Shift Conj:** For any convex set $A \subset \mathbb{R}^n$ and any unit vector $u \in \mathbb{R}^n$, the function

$$\lambda(t) = \frac{\gamma_n(tu + A) - \gamma_n(tu + H^+_{A,u})}{\gamma_n(tu + H^-_{A,u}) - \gamma_n(tu + H^+_{A,u})}$$

is decreasing on $\mathbb{R}^1$, where

$$\begin{aligned} H^-_{B,u} &= \{x \in \mathbb{R}^n : \langle x, u\rangle < -\theta\} \\ H^+_{B,u} &= \{x \in \mathbb{R}^n : \langle x, u\rangle > \theta\} \end{aligned}$$

with $\gamma_n(A) = \gamma_n(H^-_{A,u}) = \gamma_n(H^+_{A,u}) = \Phi(-\theta)$. Note that $0 \le \lambda(t) \le 1$ by the shift inequality.
Yes for $n = 1$ and "almost done" via symmetrization.

• **Gaussian Half Space Conj:** Let $(X, Y) = (X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ be centered joint Gaussian r.v. in $\mathbb{R}^{m+n}$. Then for any two Borel sets $A \subset R^m$ and $B \subset R^n$,

$$
\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &\le \sup_H \mathbb{P}(X \in H_A, Y \in H_B) \\
\mathbb{P}(X \in A, Y \in B) &\ge \inf_H \mathbb{P}(X \in H_A, Y \in H_B)
\end{aligned}
$$

where H_A and H_B are half-spaces with $\mathbb{P}(X \in A) = \mathbb{P}(X \in H_A)$, and $\mathbb{P}(Y \in A) = \mathbb{P}(Y \in H_B)$. The sup and inf are taken over all such half spaces.

Yes for $n = 2$ by using monotone shift inequality ($n = 1$ case) and symmetrization.

Outline: Let U and V be joint Gaussian random variables in $\mathbb{R}^1$ with mean zero and $\mathbb{E}\,U^2 = \mathbb{E}\,V^2 = 1$, $\mathbb{E}\,(UV) = r$. Write $V = rU + \sqrt{1-r^2}\xi$, and $\xi \sim N(0,1)$, ind. of U. Then for any two convex sets $A, B \subset R^1$, $\mathbb{P}(U \in A, V \in B) = \mathbb{P}(U \in A, rU + \sqrt{1-r^2}\xi \in B)$.

Conditioning on U, we have for $h = rU$,

$$\begin{aligned}
\mathbb{E}_h \mu(h+B) &= \mathbb{E}_h(\lambda(B,h)\mu(h+H_B^-)) \\
&\quad + \mathbb{E}_h((1-\lambda(B,h))\mu(h+H_B^+)) \\
&\leq \mathbb{E}_h \lambda(B,h) \cdot \mathbb{E}_h \mu(h+H_B^-) \\
&\quad + \mathbb{E}_h (1-\lambda(B,h)) \cdot \mathbb{E}_h \mu(h+H_B^+) \\
&\leq \max(\mathbb{E}_h \mu(h+H_B^-), \mathbb{E}_h \mu(h+H_B^+))
\end{aligned}$$

Gaussian Independent-marginal Conjecture

Conj: The upper and lower independent constants for the multivariate normal distribution $f(x)$ above are achieved when their independent-marginals are one dimensional normal densities.

Yes for $n = 2$ proved in Li (2004+).

It is not an easy problem to find explicitly the variance of those independent-marginals assuming the conjecture above. In fact, the characterization and determination are the topic of study in Leung, Li and Rakesh (2005+).

In fact,

$$g(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(D)}} \exp(-xD^{-1}x^T/2)$$

for some diagonal matrix D, $D > 0$. Hence we are interested in: *Minimize* $\det(D)$ *over diagonal matrices* $D \geq A$

• Concentration and deviation inequalities

Let f be Lipschitz function on $\mathbb{R}^n$ with

$$\|f\|_{Lip} = \sup\{|f(x) - f(y)|/|x-y| : x, y \in \mathbb{R}^n\}.$$

Denote further by M_f a median of f for μ and by $\mathbb{E}_f = \int f d\mu(x)$ for the expectation of f. Then

$$\mu(|f - M_f| > t) \leq \exp\{-t^2/2\,\|f\|^2_{Lip}\}$$

and

$$\mu(|f - \mathbb{E}_f| > t) \leq 2\exp\{-t^2/2\,\|f\|^2_{Lip}\}$$

Another version of the above result can be stated as follows. Let $\{X_t, t \in T\}$ be a centered Gaussian process with

$$d(s,t) = (\mathbb{E}\,|X_s - X_t|^2)^{1/2}, \quad s, t \in T$$

and $\sigma^2 = \sup_{t \in T} \mathbb{E}\, X_t^2$. Then for all $x > 0$, we have

$$\mathbb{P}\Big(\sup_{t \in T} X_t - \mathbb{E}\sup_{t \in T} X_t \geq x\Big) \leq \exp\Big(-\frac{x^2}{2\sigma^2}\Big).$$

There are several other inequalities of various flavor given by Dudley (1967), Fernique (1972) and Berman (1985), etc.

Comparison inequalities

Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be independent centered Gaussian random vectors.

The basic comparison identity: Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be a function with bounded second derivatives. Then

$$\mathbb{E}\, f(X) - \mathbb{E}\, f(Y) = \frac{1}{2} \int_0^1 \sum_{1 \le i,j \le n} (\mathbb{E}\, X_i X_j - \mathbb{E}\, Y_i Y_j)$$

$$\cdot \mathbb{E}\, \frac{\partial^2 f}{\partial x_i \partial x_j}((1-\lambda)^{1/2} X + \lambda^{1/2} Y) d\lambda.$$

Fernique type inequality: If

$$\mathbb{E}\,(X_i - X_j)^2 \ge \mathbb{E}\,(Y_i - Y_j)^2 \quad \text{for} \quad 1 \le i, j \le n$$

then

$$\mathbb{E} \max_{1 \le i \le n} X_i \ge \mathbb{E} \max_{1 \le i \le n} Y_i$$

and

$$\mathbb{E}\, f(\max_{i,j}(X_i - X_j)) \ge \mathbb{E}\, f(\max_{i,j}(Y_i - Y_j))$$

for every non-negative convex increasing function f on $\mathbb{R}^+$.

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

$$\mathbb{E}\min|X_i| \geq \mathbb{E}\min_{1\leq i\leq n}|\widehat{X}_i|$$

where $\widehat{X}_i$ are ind. centered Gaussian with $\mathbb{E}\,\widehat{X}_i^2 = \mathbb{E}\,X_i^2$.
Yes for $n = 2, 3$.

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

$$\mathrm{Var}(\min_{1\leq i\leq n}|X_i|) \leq \min_{1\leq i\leq n}\mathrm{Var}(|X_i|).$$

Yes for $n = 2$ and i.i.d case.

It is known that

$$\begin{aligned}&\max\left(\mathrm{Var}(\max_{1\leq i\leq n}|X_i|), \mathrm{Var}(\max_{1\leq i\leq n}X_i)\right)\\ \leq\ &\max_{1\leq i\leq n}\mathrm{Var}(X_i).\end{aligned}$$

• For any real valued f with $|f(x) - f(y)| \leq |x - y|$ and any r.v X, we have

$$\mathrm{Var}(f(X)) \leq \mathrm{Var(X)}.$$

In particular, we can talk $f(x) = \min(x, a)$ for any real number a.

Slepian's lemma: If $\mathbb{E}\, X_i^2 = \mathbb{E}\, Y_i^2$ and $\mathbb{E}\, X_i X_j \le \mathbb{E}\, Y_i Y_j$ for all $i, j = 1, 2, ..., n$, then for any x,

$$\mathbb{P}\left(\max_{1\le i\le n} X_i \le x\right) \le \mathbb{P}\left(\max_{1\le i\le n} Y_i \le x\right).$$

Other interesting and useful extensions of Slepian's inequality, involving min-max, etc, can be found in Gordon (1985).

Reverse Slepian type Inequality: Let $\{\xi_i, 1 \le i \le n\}$ and $\{\eta_i, 1 \le i \le n\}$ be two normal random vectors with mean zero and variance one. Assume that $\mathbb{E}\, \xi_i \xi_j \ge \mathbb{E}\, \eta_i \eta_j \ge 0$ for $1 \le i, j \le n$. Then for $u \ge 0$

$$\mathbb{P}\left(\max_{1\le i\le n} \xi_i \le u\right) \le \mathbb{P}\left(\max_{1\le i\le n} \eta_i \le u\right) \cdot$$
$$\cdot \prod_{1\le i<j\le n} \left(\frac{\pi - 2\,\mathrm{arcsin}(\mathbb{E}\, \eta_i\eta_j)}{\pi - 2\,\mathrm{arcsin}(\mathbb{E}\, \xi_i\xi_j)}\right)^{\exp\{-u^2/(1+\mathbb{E}\, \xi_i\xi_j)\}}.$$

Shift Inequalities

Anderson's inequality: For every convex symmetric set A,

$$\mu(A + x) \le \mu(A)$$

which follows easily from log-concavity of Gaussian measure.

Measure of shifted balls: For any $f \in H_\mu$ and $r > 0$,

$$\exp\{-|f|_\mu^2/2\} \le \frac{\mu(x : \|x - f\| \le r)}{\mu(x : \|x\| \le r)} \le 1.$$

Furthermore, as $\varepsilon \to 0$,

$$\mu(x : \|x - f\| \le \varepsilon) \sim \exp\{-|f|_\mu^2/2\} \cdot \mu(x : \|x\| \le \varepsilon).$$

The upper bound follows from Anderson's inequality. The lower bound follows from the Cameron-Martin formula. Various sharp refinements and applications are given in Kuelbs, Li and Talagrand (1994), Kuelbs, Li and Linde (1994), and Kuelbs and Li (1998).

Smooth Analysis of Simplex Method for Linear Programming

Simplex method for linear programming:

$$\max B^T x \quad s.t. \quad Ax \leq y$$

- Worst case analysis: exponential.
- Average (Gaussian for A) case analysis: polynomial.
- Widely used in practice.

Smooth analysis of simplex method:

$$\max B^T x \quad s.t. \quad (A + \sigma G)x \leq y$$

where $G = (g_{ij})$, $1 \leq i, j \leq n$, with i.i.d normal g_{ij}.

Edelman (1988):

$$\mathbb{P}\left(\|G^{-1}\| > t\right) \leq \frac{\sqrt{n}}{t}.$$

with the best constant. Here $\|A\| = \max_{\|x\|=1} \|Ax\|$ denotes the operator norm of A.

Sanker, Spielman and Teng (2002):

$$\mathbb{P}\left(\|(G+A)^{-1}\| > t\right) \le \frac{1.823\sqrt{n}}{t}$$

Gaussian Perturbation Conj:

$$\mathbb{P}\left(\|(G+A)^{-1}\| > t\right) \le \mathbb{P}\left(\|G^{-1}\| > t\right)$$

This is a part of understanding how things behavior under perturbation, such as $Ax = b$ for the input (A, b).

Note that

$$\|M^{-1}\| = \frac{1}{d(M, \mathcal{S})}$$

so this is really a small value problem. Here the distance

$$\begin{aligned} d(M, \mathcal{S}) &= \inf_{S \in \mathcal{S}} d(M, S) \\ &= \inf_{\det(s_{ij})=0} \left(\sum_{i,j=1}^{d} (m_{ij} - s_{ij})^2 \right)^{1/2}. \end{aligned}$$

The following result was proved very recently in Cordero-Erausquin, Fradelizi and Maurey (2004+) by using transportation method.

• B-inequality: For any symmetric convex set K and positive numbers a and b,

$$\mu(aK) \cdot \mu(bK) \le \mu^2(\sqrt{ab}K)$$

This is equivalent to the log-concavity of the function $t \to \gamma_n(e^t K)$ on $\mathbb{R}$.

• Let A and B be two $n \times n$ random real matrices, $A = ((a_{ij})), B = ((b_{ij}))$.

Assume that the a_{ij} are i.i.d. standard Gaussian and the same for the b_{ij}. On the other hand, A and B may be dependent.

Conj:

$$E\left(|\det(A)\det(B)|\right) \ge \left[E\left(|\det(A)|\right)\right]^2$$

Moreover, equality holds if and only if A and B are independent.

Correlation inequalities

The Gaussian correlation conjecture: For any two symmetric convex sets A and B in a separable Banach space E and for any centered Gaussian measure μ on E,

$$\mu(A \cap B) \geq \mu(A)\mu(B).$$

An equivalent formulation: If $(X_1, \ldots, X_n)$ is a centered, Gaussian random vector, then

$$\begin{aligned} & \mathbb{P}\left(\max_{1\leq i\leq n} |X_i| \leq 1\right) \\ \geq \; & \mathbb{P}\left(\max_{1\leq i\leq k} |X_i| \leq 1\right)\mathbb{P}\left(\max_{k+1\leq i\leq n} |X_i| \leq 1\right) \end{aligned}$$

for each $1 \leq k < n$.

- Sidak inequality: The above holes for $k = 1$ or any slab B.

The weaker Correlation inequality: For any $0 < \lambda < 1$, any symmetric, convex sets A and B,

$$\mu(A \cap B)\mu(\lambda^2 A + (1-\lambda^2)B) \geq \mu(\lambda A)\mu((1-\lambda^2)^{1/2}B).$$

In particular,

$$\mu(A \cap B) \geq \mu(\lambda A)\mu((1-\lambda^2)^{1/2}B)$$

and

$$\mathbb{P}(X \in A, Y \in B) \geq \mathbb{P}\Big(X \in \lambda A\Big)\mathbb{P}\Big(Y \in (1-\lambda^2)^{1/2}B\Big)$$

for any centered joint Gaussian vectors X and Y.

The varying parameter λ plays a fundamental role in applications, see Li (1999). It allows us to justify

$$\mu(A \cap B) \approx \mu(A) \quad \text{if} \quad \mu(A) \ll \mu(B).$$

Note also that

$$\mu(\cap_{i=1}^m A_i) \geq \prod_{i=1}^m \mu(\lambda_i A_i)$$

for any $\lambda_i \geq 0$ with $\sum_{i=1}^m \lambda_i^2 = 1$.

For the weaker correlation inequality established in Li (1999), here is a very simple proof given in Li and Shao (2001). Let $a = (1 - \lambda^2)^{1/2}/\lambda$, and (X^*, Y^*) be an independent copy of (X, Y). Then $X - aX^*$ and $Y + Y^*/a$ are independent. Thus, by Anderson inequality

$$
\begin{aligned}
& \mathbb{P}(X \in A, Y \in B) \\
\geq\ & \mathbb{P}(X - aX^* \in A, Y + Y^*/a \in B) \\
=\ & \mathbb{P}(X - aX^* \in A)\mathbb{P}(Y + Y^*/a \in B) \\
=\ & \mathbb{P}\Big(X \in \lambda A\Big)\mathbb{P}\Big(Y \in (1 - \lambda^2)^{1/2} B\Big).
\end{aligned}
$$

• Richard Hamming "You and Your Research" Transcription of the Bell Communications Research Colloquium Seminar, March 7, 1986. http://www.cs.virginia.edu/
~robins/YouAndYourResearch.html

Consider the sums of two centered Gaussian random vectors X and Y in a separable Banach space E with norm $\|\cdot\|$.

Theorem: If X and Y are independent and

$$\begin{aligned}
\lim_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|X\| \le \varepsilon) &= -C_X, \\
\lim_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|Y\| \le \varepsilon) &= -C_Y
\end{aligned}$$

with $0 < \gamma < \infty$ and $0 \le C_X, C_Y \le \infty$. Then

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|X+Y\| \le \varepsilon) \le -\max(C_X, C_Y) \\
&\liminf_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|X+Y\| \le \varepsilon) \\
&\quad \ge -\left(C_X^{1/(1+\gamma)} + C_Y^{1/(1+\gamma)}\right)^{1+\gamma}.
\end{aligned}$$

Theorem: If two joint Gaussian random vectors X and Y, *not* necessarily independent, satisfy

$$\begin{aligned}
\lim_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|X\| \le \varepsilon) &= -C_X, \\
\lim_{\varepsilon\to 0} \varepsilon^\gamma \log \mathbb{P}(\|Y\| \le \varepsilon) &= 0
\end{aligned}$$

with $0 < \gamma < \infty$, $0 < C_X < \infty$. Then

$$\lim_{\varepsilon \to 0} \varepsilon^{\gamma} \log \mathbb{P}\left(\|X + Y\| \le \varepsilon\right) = -C_X.$$

For any $0 < \delta < 1$, $0 < \lambda < 1$,

$$
\begin{aligned}
& \mathbb{P}\left(\|X+Y\| \leq \varepsilon\right) \\
\geq\ & \mathbb{P}\left(\|X\| \leq (1-\delta)\varepsilon,\ \|Y\| \leq \delta\varepsilon\right) \\
\geq\ & \mathbb{P}\left(\|X\| \leq \lambda(1-\delta)\varepsilon\right) \cdot \mathbb{P}\left(\|Y\| \leq (1-\lambda^2)^{1/2}\delta\varepsilon\right).
\end{aligned}
$$

Thus

$$
\liminf_{\varepsilon \to 0} \varepsilon^{\gamma} \log \mathbb{P}\left(\|X+Y\| \leq \varepsilon\right) \geq -(\lambda(1-\delta))^{-\gamma} C_X
$$

and the lower bound follows by taking $\delta \to 0$ and $\lambda \to 1$.

For the upper bound, we have

$$
\begin{aligned}
& \mathbb{P}\left(\|X\| \leq \frac{\varepsilon}{(1-\delta)\lambda}\right) \\
\geq\ & \mathbb{P}\left(\|X+Y\| \leq \frac{\varepsilon}{\lambda},\ \|Y\| \leq \delta \cdot \frac{\varepsilon}{(1-\delta)\lambda}\right) \\
\geq\ & \mathbb{P}\left(\|X+Y\| \leq \varepsilon\right) \cdot \mathbb{P}\left(\|Y\| \leq (1-\lambda^2)^{1/2}\delta \frac{\varepsilon}{(1-\delta)\lambda}\right).
\end{aligned}
$$

Thus

$$
\limsup_{\varepsilon \to 0} \varepsilon^{\gamma} \log \mathbb{P}\left(\|X+Y\| \leq \varepsilon\right) \leq -(\lambda(1-\delta))^{\gamma} C_X
$$

and the upper bound follows by taking $\delta \to 0$ and $\lambda \to 1$.

• The generic chaining or majorizing measure: Consider a metric space (T,d) and a process $(X_t)_{t\in T}$. The basic relation is

$$X_t - X_{t_0} = \sum_{j=1}^{\infty} X_{s_j(t)} - X_{s_{j-1}(t)}$$

which decomposes the increments as one moves from t_0 to t along the increasing "chain" $s_j(t) \in T_j$ such that $s_j(t) = s_j(s)$ implies $s_{j-1}(t) = s_{j-1}(s)$ and $s_j(v) = v$ for any $v \in T_j$. For $\sum_{j=1}^{\infty} \sum_{v\in T_j} w_j(v) \le 1$

$$\begin{aligned} &\mathbb{P}\left(\sup_{t\in T} \left|X_t - X_{t_0}\right| \ge u\right) \\ \le\ & \mathbb{P}\left(\sum_{j=1}^{\infty} \sum_{v\in T_j} \left|X_v - X_{s_{j-1}(v)}\right| \ge u\right) \\ \le\ & \sum_{j=1}^{\infty} \sum_{v\in T_j} \mathbb{P}\left(\left|X_v - X_{s_{j-1}(v)}\right| \ge w_j(v) u\right) \end{aligned}$$

Metric entropy condition can be used to control the size of T_j.

• Kac-Rice formula for level crossing and the excursion probability $\mathbb{P}(\sup_{t\in T} X_t \ge u)$.

• Series expansions: Let E be a separable Banach space and let X be an E–valued random variable. Then the following are equivalent.
(i). X is centered Gaussian.
(ii). There exist a separable Hilbert space H and an operator $u : H \to E$ such that $\sum_{j=1}^{\infty} \xi_j u(f_j)$ converges a.s. in E for one (each) ONB $(f_j)_{j=1}^{\infty}$ in H and

$$X \stackrel{d}{=} \sum_{j=1}^{\infty} \xi_j u(f_j)$$

where ξ_j are i.i.d. $N(0, 1)$.
(iii). There are $x_1, x_2, \ldots$ in E such that $\sum_{j=1}^{\infty} \xi_j x_j$ converges a.s. in E and

$$X \stackrel{d}{=} \sum_{j=1}^{\infty} \xi_j x_j \ .$$

Note that $\sum_{j=1}^{\infty} \xi_j u(f_j)$ converges a.s. implies that the operator u is compact and the RKHS $H_\mu = u(H)$.

• Orthogonal (Karhunen-Loeve) expansion; The spectral representations stationary process, etc.

• Gaussian randomization and de-randomization.

Reproducing Kernel Hilbert Space (RKHS)

Let (T, d) be a separable metric space and let Γ be a continuous covariance kernel on $T \times T$. Then there exists a separable Hilbert space $H(\Gamma)$ of continuous real valued functions on T such that

$$\Gamma(t, \cdot) \in H(\Gamma), \quad t \in T$$

and

$$(f(\cdot), \Gamma(t, \cdot)) = f(t), \quad f \in H(\Gamma) \quad t \in T$$

where $(\cdot, \cdot)$ denotes the inner product on $H(\Gamma)$. In addition, $H(\Gamma)$ is uniquely determined as a Hilbert space by Γ.

- Consider

$$S = \left\{ \sum_{j=1}^{n} a_j \Gamma(t_j, \cdot), a_1, \ldots, a_n \text{ real}, t_1, \ldots, t_n \in T, n \geq 1 \right\}$$

On S we define the bilinear form

$$\left(\sum_{j=1}^{n} a_j \Gamma(t_j, \cdot), \sum_{k=1}^{m} a_k \Gamma(t_k, \cdot) \right) = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j a_k \Gamma(t_j, s_k)$$

Note that if $f(t) = \sum_{j=1}^{n} a_j \Gamma(t_j, t)$, then $f(t) = (f(\cdot), \Gamma(t, \cdot))$. If $f \in S$ then $(f, f) \geq 0$, because Γ is positive definite. Also $(f, f) = 0$ implies that $f \equiv 0$ since

$$|f(t)|^2 = |(f, \Gamma(t, \cdot))|^2 \leq (f, f)(\Gamma(t, \cdot), \Gamma(t, \cdot)) = 0.$$

Thus we have an inner product on S. $H(\Gamma)$ is the completion (closure) of S under the inner product. We call (B, H, μ) a abstract Wiener space in general and here $B = C(T)$, $H = H(\Gamma)$, $\mu = \mathcal{L}(X)$.

- Isomorphism between $H(\Gamma)$ and

$$\begin{aligned} \mathcal{L}_2(\mathbb{P}) \;&=\; \text{Closure of } \{ \sum_{j=1}^{n} a_j X(t_j), \\ &\qquad a_1, \cdots, a_n \text{ real}, t_1, \cdots, t_n \in T, n \geq 1 \} \end{aligned}$$

in $L^2(\Omega, \mathcal{F}, P)$ containing $X = \{X(t), t \in T\}$. Consider the map

$$\Theta \left(\sum_{j=1}^{n} a_j \Gamma(t_j, \cdot) \right) = \sum_{j=1}^{n} a_j X(t_j)$$

from S given into $\mathcal{L}_2(P)$. Then Θ is linear, one to one, and norm preserving. It extends to all of $H(\Gamma)$ with range equal to $\mathcal{L}_2(P)$.

Our aim will now be to extend the isoperimetric and concentration inequalities to the setting of an infinite dimensional Gaussian measure μ. Let us mention however before that the fundamental inequalities are the ones in finite dimension and that the infinite dimensional extensions we will present actually follow in a rather classical and straightforward manner from the finite dimensional case. The main tool will be the concept of abstract Wiener space and reproducing kernel Hilbert space which will define the isoperimetric neighborhoods or enlargements in this framework. We follow essentially C. Borell (1976) in the construction below. Let μ be a mean zero Gaussian measure on a real separable Banach space E. Consider then the abstract Wiener space factorization [Gr1], [B-C], [Ku], [Bo3] (for recent accounts, cf. [Bog], [Lif3]),

$$E^* \rightarrow^{j} L_2(\mu) \rightarrow^{j^*} E$$

where $j(h) = h(X)$ and $j^*j(h) = \mathbb{E}\,(Xh(X)) = \int xh(x)\mu(dx)$.

Define now the reproducing kernel Hilbert space H of μ as the subspace $j^*(L_2(\mu))$ of E.

Karhunen-Loeve Expansions for Gaussian Process

Consider a centered Gaussian process $\{X_t, a \le t \le b\}$ with continuous covariance function $\sigma(s,t) = \mathbb{E}\, X_s X_t$. By Mercer's Thm, there exist eigenvalues $\lambda_n > 0$ and a complete orthonormal bases (eigenfunctions) $e_n(t)$ of

$$\lambda f(t) = \int_0^1 \sigma(s,t) f(s) ds.$$

In addition,

$$\sigma(s,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t)$$

and the series converges absolutely and uniformly in $[a,b] \times [a,b]$. The Karhunen-Loeve expansion for X_t is

$$X_t = \sum_{n=1}^{\infty} \lambda_n^{1/2} \xi_n e_n(t)$$

and the series converges a.e. and in $L^2(\Omega)$.

•Find the K-L expansion for the Brownian bridge process B_t^0 on $[0,1]$ with $\mathbb{E}\, B_t^0 = 0$ and $\mathbb{E}\, B_s^0 B_t^0 = \min(s,t) - st$.

Brownian Motion

Brownian motion (BM) is a process of tremendous practical and theoretical significance. Intuitively, Brownian motion corresponds to the concept of a homogeneous, continuous time, continuous random walk. One way to visualize Brownian paths is to consider a simple random walk on the real line, in which the walker starts at 0 and moves up or down by an amount $\sqrt{dt}$ after each timeinterval of duration dt. To be more precise, the continuous path $Z_n(t)$, builded from linear interpolation of simple random walks, "converges" to Brownian motion $B(t)$ where

$$Z_n(t) = \frac{1}{\sqrt{n}}\Big(S_{[nt]} + (nt - [nt])X_{[nt]+1}\Big)$$

and $S_0 = 0$, $S_k = \sum_{i=1}^{k} X_i$, X_i are i.i.d with $\mathbb{P}(X_i = \pm 1) = 1/2$.

• The existence of BM was first rigorously proved by Wiener (1923). So the BM is also called the Wiener process.

Def: A continuous time stochastic process $\{B_t, 0 \le t < \infty\}$ on $\mathbb{R}$ is called a standard Brownian Motion (BM) if it has the following properties:

(i). $B_0 = 0$.

(ii). The increments of B_t are independent; i.e. for any finite set of times $t_0 < t_1 < \cdots < t_n$, the r.v's B_{t_0}, $B_{t_1} - B_{t_0}, \cdots, B_{t_n} - B_{t_{n-1}}$ are ind.

(iii). For any $0 \le s \le t$, the increment $B_t - B_s$ has normal distribution with mean 0 and variance $t - s$.

(iv). With probability one, the paths of B_t are continuous.

Def: A n-dimensional Brownian motion started at $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ is given by

$$B_t^x = (x_1 + B_t^{(1)}, \cdots, x_n + B_t^{(n)})$$

where $\{B_t^{(j)}\}$, $1 \le j \le n$ are ind. standard BMs in $\mathbb{R}$.

• **Via a measurable mapping from $C[0,1]$ to $[0,1]$:** The Wiener measure on $C[0,1]$ is induced by the Lebesgue measure on $[0,1]$. This is the original very complected construction of N. Wiener (1923).

• **Via finite dimensional extension:** For $0 \le t_1 \le \cdots \le t_k$,

$$\begin{aligned} & \nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) \\ = \ & \mathbb{P}(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_{t_k}) \\ = \ & \int_{F_1 \times \cdots \times F_k} p(t_1; 0, x_1) p(t_2 - t_1; x_1, x_2) \cdots \\ & \quad p(t_k - t_{k-1}; x_{k-1}, x_k) dx_1 \cdots dx_k \end{aligned}$$

for $F_i \in \mathcal{B}(\mathbb{R})$, where the transition density

$$p(t; x, y) = (2\pi t)^{-1/2} \exp\Big(-(x-y)^2/(2t)\Big).$$

• **Via Gaussian process:** $\mathbb{E}\, B_t = 0$ and covariance function $\sigma(s,t) = \mathbb{E}\,(B_s B_t) = \min(s,t)$.

- **Via Karhunen-Loeve expansion:** On $[0, 1]$,

$$B(t) = \sqrt{2} \sum_{n=1}^{\infty} \xi_n \cdot \frac{\sin((n - \frac{1}{2})\pi t)}{\pi(n - 1/2)}$$

and the series converges a.e. and in $L^2(\Omega)$.

• **Via Donsker's invariance principle:** Let X_i be i.i.d. r.v's with $\mathbb{E}\, X_i = 0$, $\mathrm{Var}(X_i) = 1$. Define the interpolation process $\{Z_n(t), 0 \le t \le 1\}$

$$Z_n(t) = \frac{1}{\sqrt{n}}\Big(S_{[nt]} + (nt - [nt])X_{[nt]+1}\Big)$$

with $S_0 = 0$, $S_k = \sum_{i=1}^k X_i$. Then $Z_n(t)$ converges weakly in the space $C[0,1]$ with the sup-norm, i.e. for any continuous functional $\psi : C[0,1] \to \mathbb{R}$ w.r.t. the sup-norm topology,

$$\lim_{n\to\infty} \mathbb{P}(\psi(Z_n(t)) \le x) = \mathbb{P}(\psi(B(t)) \le x).$$

Ex: For $\psi(f)$ given by $\sup_{0\le t\le 1} f(t)$, $\sup_{0\le t\le 1} |f(t)|$, $\sup_{0\le s,t\le 1} |f(t)-f(s)|$ or $\int_0^1 |f(t)|^p dt$, $p > 0$, we have

$$\begin{aligned}
n^{-1/2} \max_{1\le k\le n} S_n &\Longrightarrow \sup_{0\le t\le 1} B_t \\
n^{-1/2} \max_{1\le k\le n} |S_n| &\Longrightarrow \sup_{0\le t\le 1} |B_t| \\
n^{-1/2}\left(\max_{1\le k\le n} S_n - \min_{1\le k\le n} S_n\right) &\Longrightarrow \sup_{0\le s,t\le 1} |B_t - B_s| \\
\int_0^1 |Z_n(t)|^p &\Longrightarrow \int_0^1 |B_t|^p dt
\end{aligned}$$

• **Via random sum of integrals of orthogonal functions:** Let $\{\phi_n(t)\}_{n\geq 1}$ be any complete orthonormal system in $L^2(0,1)$. Then

$$B_t = \sum_{n=1}^{\infty} \xi_n \int_0^t \phi_n(u)du \quad a.s.$$

and the series converges uniformly in t.

• For ONB $\phi_0(t) = 1$, $\phi_n(t) = \sqrt{2}\cos(n\pi t)$, $n \geq 1$, in $L^2(0,1)$, we obtain

$$B(t) = t\xi_0 + \sqrt{2}\sum_{n=1}^{\infty} \xi_n \cdot \frac{\sin(n\pi t)}{\pi n}$$

This is related to K-L expansion of $B_0(t)$.

• For ONB $\phi_n(t) = \sqrt{2}\sin(n\pi t)$, $n \geq 1$, in $L^2(0,1)$, we have

$$B_t = \sqrt{2}\sum_{n=1}^{\infty} \xi_n \frac{1}{\pi n}(1 - \cos n\pi t)$$

• For ONB $\left\{1, \sqrt{2}\sin(2\pi nt), \sqrt{2}\cos(2\pi nt)\right\}$ in $L^2(0,1)$, we have

$$\begin{aligned} B_t &= t\xi_0 \\ &\quad + \sqrt{2}\sum_{n=1}^{\infty}\left(\frac{\xi_n}{2\pi n}(1-\cos 2\pi nt) + \frac{\xi_n'}{2\pi n}\sin 2\pi nt\right) \end{aligned}$$

• **Via Haar System and Wavelet:** This is a special case of the representation of BM in terms of random sum of integrals of orthogonal functions. Consider a "mother wavelet"

$$H(t) = \begin{cases} 1 & \text{for} \quad 0 \le t < 1/2 \\ -1 & \text{for} \quad 1/2 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Define $H_0(t) = 1$ and for $n = 2^j + k$,

$$H_n(t) = 2^{j/2} H(2^j t - k), \quad j \ge 0, 0 \le k < 2^j$$

by scaling and translating the mother wavelet $H(\cdot)$. The Haar system $\{H_n(t), n \ge 0\}$ is an ONB in $L^2[0,1]$. Hence

$$B_t = \sum_{n=0}^{\infty} \lambda_n \xi_n \Delta_n(t), \quad 0 \le t \le 1$$

where $\lambda_n \Delta_n(t) = \int_0^t H_n(u) du$, $\lambda_0 = 1, \Delta_0(t) = t$ and for $n = 2^j + k$, $j \ge 0, 0 \le k < 2^j$, $\lambda_n = 2^{-1-j/2}$, $\Delta_n(t) = \Delta(2^j t - k)$,

$$\Delta(t) = \int_0^t H(u) du = \begin{cases} 2t & \text{for} \quad 0 \le t < 1/2 \\ 2(1-t) & \text{for} \quad 1/2 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that for fixed j, $\Delta_{2^j+k}(t)$ have disjoint supports for $0 \le k < 2^j$, and are bounded between 0 and 1.

Fractional Brownian Motion

Def: A centered Gaussian random field (processes) $X = \{X(t), t \in \mathbb{R}^N\}$ with values in $\mathbb{R}^d$ is called a **fractional Brownian motion (fBm)** of index H, $0 < H < 1$, if $\forall s, t \in \mathbb{R}^N$

$$\mathbb{E}\left(X_i(s)X_j(t)\right) = \frac{1}{2}\delta_{i,j}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right),$$

where $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise.

•X is also called an (N, d, H)-fractional Brownian motion. Its coordinate processes $X_1, \ldots, X_d$ are independent $(N, 1, H)$-fractional Brownian motions.

•If $N = 1$ and $H = 1/2$, then X is the ordinary d-dimensional Brownian motion.

•If $N > 1$ and $H = 1/2$, X was first introduced by P. Lévy (1948) as an N-parameter Brownian motion.

•If $N = 1$, then $(1, 1, H)$-fBM is the only H-self-similar Gaussian process with stationary increments.

Thm: For any constant $H \in (0,1]$, $(N,1,H)$-fractional Brownian motion exists.

Pf: We only need to show the function

$$R(s,t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}), \quad (s,t) \in \mathbb{R}^N \times \mathbb{R}^N$$

is non-negative definite, i.e. $\forall n \geq 1$, $t_1, \ldots, t_n \in \mathbb{R}^N$ and $\forall a_1, \ldots, a_n \in \mathbb{R}$, we have

$$\sum_{i=1}^n \sum_{j=1}^n a_i R(t_i, t_j) a_j \geq 0.$$

Recall that for any $c > 0$, the function $t \to e^{-c|t|^{2H}}$ is the characteristic function of an isotropic (2H)-stable vector in $\mathbb{R}^N$, so it is non-negative definite. This follows from Lévy-Khintchine formula for Lévy process, and in this special case,

$$\mathbb{E}\, e^{\mathrm{i}\langle t, Y\rangle} = e^{-c|t|^{2H}}, \quad t \in \mathbb{R}^N.$$

For index set $t \in \mathbb{R}$, $\exp(-|t|^{\alpha})$ is a ch.f. for $0 < \alpha \leq 2$ follows from a nice argument due to Frank Spitzer. It is based on the fact

$$1 - (1-x)^{\alpha/2} = \sum_{n=1}^{\infty} c_n x^n$$

where $c_n = (-1)^{n+1}\binom{\alpha/2}{n}$, $c_n \geq 0$ (here we used $\alpha \leq 2$) and $\sum_{n=1}^{\infty} c_n = 1$.

Back to the proof of non-negative definiteness. Let $a_0 = -\sum_{j=1}^{n} a_i$ and $t_0 = 0$, then we can rewrite

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i R(t_i, t_j) a_j = -\sum_{i=0}^{n} \sum_{j=0}^{n} a_i |t_i - t_j|^{2H} a_j.$$

Note that

$$0 \leq \sum_{i,j=0}^{n} a_i e^{-c|t_i - t_j|^{2H}} a_j = \sum_{i,j=0}^{n} a_i (e^{-c|t_i - t_j|^{2H}} - 1) a_j.$$

Since n, $t_1, \ldots, t_n$, $a_1, \ldots, a_n$ are all fixed, by letting $c \to 0$, we have

$$\begin{aligned} & \sum_{i,j=0}^{n} a_i (e^{-c|t_i - t_j|^{2H}} - 1) a_j \\ = \ & -c(1 + o(1)) \sum_{i,j=0}^{n} a_i |t_i - t_j|^{2H} a_j \end{aligned}$$

Hence we can choose $c > 0$ small so that

$$\sum_{i,j=0}^{n} a_i |t_i - t_j|^{2H} a_j \leq 0.$$

This finishes the proof.

•For a different proof without using the positive definiteness of the function $t \rightarrow e^{-c|t|^{2H}}$, see Ossiander and Waymire (1989).

Prop: If $X = \{X(t), t \in \mathbb{R}^N\}$ is an (N, d, H)-fBm, then the following invariance properties hold.
(i). $\forall t > 0$, the process $\{c^{-H} X(t), t \in \mathbb{R}^N\}$ is also an (N, d, H)-fBm.
(ii). $\forall t_0 \in \mathbb{R}^N$, then the process $\{X(t + t_0) - X(t_0), t \in \mathbb{R}^N\}$ is an (N, d, H)- fBm.
(iii). $\forall A \in O(\mathbb{R}^N) =:$ the group of all orthogonal linear operators, $\{X(At), t \in \mathbb{R}^N\}$ is an (N, d, H)-fBm.
(iv). $\forall C \in O(\mathbb{R}^N)$, $\{CX(t),\, t \in \mathbb{R}^N\}$ is an (N, d, H)-fBm.
•Motivated by (iii) above, we define a random field $X = \{X(t), t \in \mathbb{R}^N\}$ is **an operator-self-similar** if there exists an linear operator A on $\mathbb{R}^N$ such that for $\forall c > 0$

$$\{X(c^A t), t \in \mathbb{R}^N\} \overset{d}{=} \{cX(t), t \in \mathbb{R}^N\}$$

where

$$c^A = e^{A \log c} =: \sum_{n=0}^{\infty} \frac{(\log c)^n}{n!} A^n.$$

The linear operator A is called the self-similarity exponent of X.
HW 28: If X is (N, d, H)-fBm, what is its self-similarity exponent A?

Stochastic Integral Representations for FBM

All representations are useful in studying various properties of fractional Brownian motion and other related Gaussian random fields.

Moving average representation of fBM: We consider the case $N = 1$ first.

Thm: Let $H \in (0, 1]$ be given and let $B = \{B(t),\, t \in \mathbb{R}\}$ be Brownian motion, then the Gaussian process $X = \{X(t), t \in \mathbb{R}\}$ defined by

$$X(t) = \frac{1}{c_1(H)} \int_{-\infty}^{\infty} \left[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right] dB(s)$$

is a $(1, 1, H)$-fBm. In the above $s_+ = \max(s, 0)$ and $c_1(H)$ is the normalizing constant given by

$$c_1(H)^2 = \int_0^{\infty} \left[(1+s)^{H-1/2} - s^{H-1/2} \right]^2 ds + \frac{1}{2H}.$$

When $H = \frac{1}{2}$, the above integral is understood as $\int_0^t dB(s)$ if $t > 0$, and $\int_{-t}^0 dB(s)$ if $t < 0$.

Pf: Let $f_t(x)$ denote integrand in the representation of $X(t)$. We first verify that

$$\int_{-\infty}^{\infty} f_t(x)^2 dx < \infty.$$

If $H = \frac{1}{2}$, then

$$\int_{-\infty}^{\infty} f_t(x)^2 dx = \int_0^{|t|} dx < \infty$$

Suppose now $0 < H < 1$ and $H \neq \frac{1}{2}$. Noting that if $t > 0$

$$f_t(x) = \begin{cases} (t-x)^{H-1/2} - (-x)^{H-1/2}, & -\infty < x \leq 0; \\ (t-x)^{H-1/2}, & 0 < x < t; \\ 0, & x \geq t \end{cases}$$

If $t < 0$, then

$$f_t(x) = \begin{cases} (t-x)^{H-1/2} - (-x)^{H-1/2}, & -\infty < x \leq t; \\ -(-x)^{H-1/2}, & t < x < 0; \\ 0, & x \geq 0 \end{cases}$$

So as $x \to -\infty$,

$$f_t(x) \sim (H - \frac{1}{2})(-x)^{H-\frac{3}{2}} t.$$

Therefore the integrand $f_t(x) \in L^2(-\infty, +\infty)$ and the process X_t is well defined.

Next we verify that X_t is a standard fBm. Clearly $X(0) = 0$ and $\mathbb{E}(X(t)) = 0$. For all $t > 0$,

$$
\begin{aligned}
& \mathbb{E}(X(t)^2) \\
= \ & \frac{1}{c_1(H)^2} \int_{-\infty}^{\infty} f_t^2(x) dx \\
= \ & \frac{1}{c_1(H)^2} \Big[\int_{-\infty}^{0} \Big((t-x)^{H-1/2} - (-x)^{H-1/2} \Big)^2 dx \\
& \qquad + \int_0^t (t-x)^{2H-1} \, dx \Big] \\
= \ & \frac{t^{2H}}{c_1(H)^2} \Big[\int_{-\infty}^{0} \Big((1-x)^{H-1/2} - (-x)^{H-1/2} \Big)^2 dx \\
& \qquad + \int_0^1 (1-x)^{2H-1} dx \Big] \\
= \ & t^{2H}
\end{aligned}
$$

Similarly we have $\mathbb{E}(X(t)^2) = |t|^{2H}$ for all $t < 0$. Furthermore, for

$s, t \in \mathbb{R}$ and $s < t$,

$$
\begin{aligned}
& \mathbb{E}\Big(X(t) - X(s)\Big)^2 \\
= \; & \frac{1}{c_1(H)^2} \int_{-\infty}^{\infty} \Big((t-x)_+^{H-1/2} - (s-x)_+^{H-1/2}\Big)^2 \, dx \\
= \; & \frac{1}{c_1(H)^2} \int_{-\infty}^{\infty} [(t-s-x)_+^{H-1/2} - (-x)_+^{H-1/2}]^2 \, dx \\
= \; & |t-s|^{2H}.
\end{aligned}
$$

Therefore X is a standard fBm of index H.

•For $t > 0$, the representation can be written as

$$\begin{aligned} X(t) &= \frac{1}{c_1(H)}\Bigg[\int_{-\infty}^{0}\left[(t-s)^{H-1/2} - (-s)^{H-1/2}\right] dB(s) \\ &\qquad + \int_0^t (t-s)^{H-1/2}\, dB(s)\Bigg] \\ &= X_1(t) + X_2(t) \end{aligned}$$

It is easy to verify that both X_1 and X_2 are H-self-similar *independent* Gaussian processes. But they do not have stationary increments. The process $X_2 = \{X_2(t), t \in \mathbb{R}\}$ is also called the Liouville fBm.

•The representation $X(t)$ is called a moving average representation of fBm. Such representations are not unique, for example, the process Y defined by

$$Y(t) = c_2(H)^{-1} \int_{-\infty}^{\infty} \left(|t-s|^{H-1/2} - |s|^{H-1/2}\right) dB(s)$$

is also a $(1, 1, H)$-fBm, where $c_2(H) > 0$ is a normalizing constant.

An Extension to Multiparameter Case

Thm: Let $H \in (0, 1]$ be a constant and let W be an independently scattered Gaussian random measure on $\mathbb{R}^N$ with Lebesgue measure as its control measure. Then the Gaussian random field $X = \{X(t),\, t \in \mathbb{R}^N\}$ defined by

$$X(t) = \frac{1}{c_3(H)} \int_{\mathbb{R}^N} \left(|t-s|^{H-N/2} - |s|^{H-N/2} \right) dW(s),$$

is an $(N, 1, H)$-fBm, where $c_3(H) > 0$ is a normalizing constant given by

$$c_3(H)^2 = \int_{\mathbb{R}^N} \left(|\theta + z|^{H-N/2} - |z|^{H-N/2} \right)^2 dz.$$

Pf: First we show that $\forall t \in \mathbb{R}$, $X(t)$ is well defined. It is clear that $X(0) = 0$. So we assume $t \neq 0$ and show

$$\int_{\mathbb{R}^N} \left(|t-s|^{H-N/2} - |s|^{H-N/2} \right)^2 ds < \infty$$

Take $T > 2|t|$ and write above integral as

$$\begin{aligned} & \int_{B(0,T)} \left(|t-s|^{H-N/2} - |s|^{H-N/2} \right)^2 ds \\ & + \int_{B(0,T)^c} \left(|t-s|^{H-N/2} - |s|^{H-N/2} \right)^2 ds \\ = \; & I_1 + I_2. \end{aligned}$$

Since

$$\left(|t-s|^{H-N/2} - |s|^{H-N/2}\right)^2$$
$$\leq \ 2\left(|t-s|^{-(N-2H)} + |s|^{-(N-2H)}\right)$$

We see that $I_1 < \infty$. On the other hand, for all $s \in \mathbb{R}^N$ with $|s|$ large, we have

$$\left||t-s|^{H-N/2} - |s|^{H-N/2}\right| \leq c\,|s|^{H-N/2-1},$$

where $c > 0$ is a constant depending on T, H and N only. So

$$I_2 \leq c^2 \int_{B(0,T)^c} |s|^{-(N+2-2H)} ds < \infty.$$

In the second step we show that for all $s, t \in \mathbb{R}^N$,

$$\mathbb{E}\,(X(s)X(t)) = c_3(H)^2\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right)$$

For the purpose, we only need to show

$$\mathbb{E}\left(X(t) - X(s)\right)^2 = c_3(H)\,|s-t|^{2H}.$$

In fact, by change of variables we have

$$
\begin{aligned}
& \text{LHS of Repsentation} \\
= & \int_{\mathbb{R}^N} \left(|t-x|^{H-N/2} - |s-x|^{H-N/2} \right)^2 dx \\
= & \int_{\mathbb{R}^N} \left(|t-s+y|^{H-N/2} - |y|^{H-N/2} \right)^2 dx \\
= & |t-s|^N \int_{\mathbb{R}^N} |t-s|^{2H-N} \\
& \qquad \left(\left| \frac{t-s}{|t-s|} + z \right|^{H-N/2} - |z|^{H-N/2} \right)^2 dz \\
= & |t-s|^{2H} \int_{\mathbb{R}^N} \left(|\theta + z|^{H-N/2} - |z|^{H-N/2} \right)^2 dz
\end{aligned}
$$

where $\theta \in S_N$. Note that the last integral is independent of θ and is denoted by $c_3(H)^2$. This completes the proof.

Here is another representation for fBm. It is very useful in defining stochastic integration w.r.t fBm and in the prediction theory.

Thm: The Gaussian process $X = \{X(t), t \geq 0\}$ defined by

$$X(t) = \int_0^t K_H(t,s)\, dB(s) \qquad \forall t \geq 0,$$

where $K_H(t,s)$ is the kernel given by

$$K_H(t,s) = \left(\frac{t}{s}\right)^{H-\frac{1}{2}} (t-s)^{H-\frac{1}{2}} - (H - \frac{1}{2}) s^{\frac{1}{2}-H}$$

is a $(1,1,H)$-fBm. Moreover, we have the following prediction formula: $\forall 0 < t < T$,

$$\mathbb{E}\left(X(T) | X(s), s \leq t\right) = \int_0^t K_H(T,s)\, dB(s).$$

Steps of a Proof:

(I). Verify

$$\int_0^t K_H^2(t,s)ds < \infty$$

(II). Show that $\forall s < t$

$$\mathbb{E}\left(X(t) - X(s)\right)^2 = C_H'' \left|t-s\right|^{2H}$$

(III). Observe that $\forall T > 0$

$$\left\{\int_0^t K_H(T,s)\,dB(s) : 0 \le t \le T\right\}$$

is a martingale. Therefore

$$\begin{aligned}&\mathbb{E}\left(\int_0^t K_H(T,s)\,dB(s)\right|\\ &\qquad\left.\int_0^r K_H(T,s)\,dB(s),\, 0 \le r \le u\right)\\ = \ & \int_0^u K_H(T,s)\,dB(s),\end{aligned}$$

which implies the prediction formula. We refer to Norros et al. (1999) for details.

Harmonizable representation of fBM

Def: A real valued Gaussian process $W(A)$ indexed by $\mathcal{B}(\mathbb{R}^N)$, the Borel sets on $\mathbb{R}^N$, is called scattered Gaussian random measures on $\mathbb{R}^N$ with control measures ν on $\mathbb{R}^N$ if

$$\mathbb{E}\, W(A) = 0, \quad A \in \mathcal{B}(\mathbb{R}^N)$$

and

$$\mathbb{E}\, W(A)W(B) = \nu(A \cap B), \quad A, B \in \mathcal{B}(\mathbb{R}^N).$$

•Note that $W(A)$ and $W(B)$ are ind. if $A \cap B = \emptyset$.
EX: Brownian sheets.

Thm: Let W and W' be two independently scattered Gaussian random measures on $\mathbb{R}^N$ with Lebesgue measure as their control measures. Then for any constant $H \in (0,1)$, the Gaussian random field $X = \{X(t), t \in \mathbb{R}^N\}$ defined by

$$\begin{aligned} X(t) \;&=\; c_4(H)^{-1} \left\{ \int_{\mathbb{R}^N} \frac{1 - \cos\langle t, x\rangle}{|x|^{H+N/2}}\, dW(x) \right. \\ &\qquad \left. + \int_{\mathbb{R}^N} \frac{\sin\langle t, x\rangle}{|x|^{H+N/2}} dW'(x) \right\}, \end{aligned}$$

is an$(N,1,H)$-fBm. In the above $c_4(H) > 0$ is a normalizing constant given by

$$c_4(H)^2 = 2 \int_{\mathbb{R}^N} \frac{1 - \cos\langle \mathbf{1}, x\rangle}{|x|^{2H+N}}\, dx,$$

where $\mathbf{1} = (1, 0, \ldots, 0) \in \mathbb{S}^{N-1}$.

Pf: Both stochastic integrals are well defined since

$$\int_{\mathbb{R}^N} \frac{(1 - \cos\langle t, x\rangle)^2}{|x|^{2H+N}}\, dx < \infty$$

and

$$\int_{\mathbb{R}^N} \frac{\sin^2\langle t, x\rangle}{|x|^{2H+N}}\, dx < \infty.$$

Next we show $\mathbb{E}\,(X(s) - X(t))^2 = |s - t|^{2H}$. It follows from the representation that

$$
\begin{aligned}
&\mathbb{E}\Big(X(s)-X(t)\Big)^2\\
={}& c_4(H)^{-2}\Big[\int_{\mathbb{R}^N}\frac{(\cos\langle s,x\rangle-\cos\langle t,x\rangle)^2}{|x|^{2H+N}}\,dx\\
&\qquad+\int_{\mathbb{R}^N}\frac{(\sin\langle s,x\rangle-\sin\langle t,x\rangle)^2}{|x|^{2H+N}}\,dx\Big]\\
={}& 2c_4(H)^{-2}\int_{\mathbb{R}^N}\frac{1-\cos\langle t-s,x\rangle}{|x|^{2H+N}}dx\\
={}& |s-t|^{2H},
\end{aligned}
$$

where the last equality follows from the definition of $c_4(H)$.

Def: A complex valued random variable $X = X_1 + \mathrm{i}X_2$ is called Gaussian if (X_1, X_2) is a Gaussian vector in $\mathbb{R}^2$.

•The distribution of X is completely determined by the following five parameters

$$\mathbb{E}(X_1),\ \mathbb{E}(X_2),\ \mathbb{E}(X_1^2),\ \mathbb{E}(X_2^2),\ \mathrm{Cov}(X_1, X_2)$$

•If $\mathbb{E}(X) = 0$, then we say X is centered.

•A complex Gaussian random variable X is called *symmetric* if $\forall \theta \in \mathbb{C}$ with $|\theta| = 1$,we have $X \stackrel{d}{=} \theta X$.

Prop: Let $X = X_1 + \mathrm{i}X_2$ be a complex Gaussian random variable, then the following are equivalent:
(i). X is symmetric.
(ii). $\mathbb{E}(X) = \mathbb{E}(X^2) = 0$.
(iii). X_1, X_2 are independent centered Gaussian random variables with the same variance.

Pf: We just prove (i)$\Leftrightarrow$(ii). Assume X is symmetric, then we verify directly that $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 0$. Conversely if $\mathbb{E}(X) = \mathbb{E}(X^2) = 0$, we must have

$$\mathbb{E}(X_1) = 0,\ \mathbb{E}(X_2) = 0,\ \mathbb{E}(X_1^2) = \mathbb{E}(X_2^2),$$
$$\mathrm{Cov}(X_1, X_2) = 0$$

Take any $\theta = \theta_1 + \mathrm{i}\theta_2 \in \mathbb{C}$ with $|\theta| = 1$, we see that the real and imaginary parts of θX are given by

$$(\theta_1 X_1 - \theta_2 X_2, \theta_2 X_1 + \theta_1 X_2) = \begin{pmatrix} \theta_1 & -\theta_2 \\ \theta_2 & \theta_1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

Since the above matrix is orthogonal, we have $\theta X \stackrel{d}{=} X$. This proves that X is symmetric.

Def: We say that a system of complex valued random variables $X = \{X(t), t \in I\}$ is Gaussian (i.e., $X = \{X(t), t \in I\}$ is a complex Gaussian process) if for all $n \geq 1$, $c_1, \ldots, c_n \in \mathbb{C}$ and all $t_1, \ldots, t_n \in T$, the random variable $\sum_{i=1}^{n} c_i X(t_i)$ is complex Gaussian.

Thm: Given a measure space $(E, \mathcal{E}, \nu)$, let

$$\mathcal{E}_0 = \{A \in \mathcal{E}, \nu(A) < \infty\}.$$

Then there exists a complex Gaussian process $\tilde{W} = \{\tilde{W}(A), A \in \mathcal{E}_0\}$ such that

(i). $\mathbb{E}\left(\tilde{W}(A)\right) = 0$, $\forall A \in \mathcal{E}_0$.

(ii). $\mathbb{E}\left(\tilde{W}(A)\overline{\tilde{W}(B)}\right) = \nu(A \cap B)$, $\forall A, B \in \mathcal{E}_0$.

(iii). $\tilde{W}(-A) = \overline{\tilde{W}(A)}$, $\forall A \in \mathcal{E}_0$.

• $\tilde{W}$ is called a complex Gaussian random measure with control measure ν.

• The random measure $\tilde{W}$ is not independently scattered.

Ex: Consider a measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$. Let W_1 and W_2 be two independently scattered Gaussian random measures on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \mu)$. Assume W_1 and W_2 are independent. Define

$$\tilde{W}_1(A) = \begin{cases} W_1(A), & \text{if } A \in \mathcal{B}(\mathbb{R}_+) \\ W_1(-A), & \text{if } A \in \mathcal{B}(\mathbb{R}_-) \end{cases}$$

$$\tilde{W}_2(A) = \begin{cases} W_2(A), & \text{if} A \in \mathcal{B}(\mathbb{R}_+) \\ -W_2(-A), & \text{if} A \in \mathcal{B}(\mathbb{R}_-) \end{cases}$$

Both $\tilde{W}_1, \tilde{W}_2$ can be extended to $\mathcal{B}(\mathbb{R})$ in a natural way. Then $\tilde{W}$ defined by

$$\tilde{W}(A) = \tilde{W}_1(A) + \mathtt{i}\tilde{W}_2(A)$$

is a complex Gaussian random measure.

Ex: The method in the Example above can be extended to $\mathbb{R}^N$ for $N \geq 2$. Let $\mathbb{R}^N_{1,+} = \{(t_1, \cdots, t_N) \in \mathbb{R}^N, t_1 \geq 0\}$. Let W_1 and W_2 be two independently scattered Gaussian measures on $\mathbb{R}^N_{1,+}$ with control measure $\frac{1}{2}\lambda_N$. We assume that W_1 and W_2 are independent. Define:

$$\tilde{W}_1(A) = \begin{cases} W_1(A) & \text{if } A \in \mathcal{B}(\mathbb{R}^N_{1,+}), \\ W_1(-A) & \text{if } A \in \mathcal{B}(\mathbb{R}^N_{1,+}) \end{cases}$$

$$\tilde{W}_2(A) = \begin{cases} W_2(A) & \text{if } A \in \mathcal{B}(\mathbb{R}^N_{1,+}), \\ W_2(-A) & \text{if } A \in \mathcal{B}(\mathbb{R}^N_{1,+}) \end{cases}$$

Then $\tilde{W}(A) = \tilde{W}_1(A) + \mathtt{i}\tilde{W}_2(A)$ is a complex Gaussian random measure described.

A Harmonizable Representation for fBm

Thm: Let $\tilde{W}$ be a complex Gaussian random measure on $\mathbb{R}^N$ with Lebesque measure as its control measure. For $\forall H \in (0,1)$, the Gaussian random field $X = \{X(t), t \in \mathbb{R}^N\}$ defined by

$$X(t) = c_4(H)^{-1} \int_{\mathbb{R}^N} \frac{1 - e^{\mathrm{i}\langle t,x\rangle}}{|x|^{H+\frac{N}{2}}} \tilde{W}(dx),$$

where $c_4(H)$ is the constant defined early, is an $(N, 1, H)$-fBm.

Pf: First note that $X(t)$ is well defined for every $t \in \mathbb{R}^N$ and

$$\int_{\mathbb{R}^N} \frac{|1 - e^{\mathrm{i}\langle t,x\rangle}|^2}{|x|^{2H+N}} \, dx < \infty$$

and clearly $X(0) = 0$.

To finish the proof, we only need to verify $\mathbb{E}\Big(X(t) - X(s)\Big)^2 = |t-s|^{2H}$. This follows from the proof of in the real case.

Strongly Locally Non-deterministic Gaussian Fields

As an application we now prove that fBM is "strongly locally non-deterministic"

Thm: [Pitt, 1978] Let $X = \{X(t), t \in \mathbb{R}^N\}$ be an $(N, 1, H)$-fBm. Then there exists a constant $c > 0$ such that for all $t \in \mathbb{R}^N \backslash \{0\}$ and $0 \le r \le |t|$, we have:

$$\mathrm{Var}\Big(X(t)\Big|X(s) : |s-t| \ge r\Big) \ge c\, r^{2H}.$$

•We say that fBm X is strongly locally r^{2H}-non-determinism; see Xiao (2005) for a more general definition and its applications.

Pf: We use the Hilbert space setting for conditional variance. Then

$$\begin{aligned} &\mathrm{Var}\Big(X(t)\Big|X(s) : |s-t| \ge r\Big) \\ =\ & \inf \mathbb{E}\left[\Big(X(t) - \sum_{j=1}^{n} a_j X(s_j)\Big)^2\right], \end{aligned}$$

where the infimum is taken over all $n \ge 1$, $a_1, ..., a_n \in \mathbb{R}$ and all $s_1, ..., s_n \in \mathbb{R}^N$ such that $|s_j - t| \ge r$ $(j = 1, \ldots, n)$.

Hence, it is sufficient to show that there exists a constant $c > 0$ such that

$$J := \mathbb{E}\left(X(t) - \sum_{j=1}^{n} a_j X(s_j)\right)^2 \geq c\, r^{2H}$$

for all choices of $n \geq 1, \{a_j\}$ and $\{s_j\}$ as above.

By the harmonizable representation we can write

$$\begin{aligned} & X(t) - \sum_{j=1}^{n} a_j X(s_j) \\ = & \int_{\mathbb{R}^N} \left[1 - e^{\mathrm{i}\langle t,x\rangle} - \sum_{j=1}^{n} a_j \Big(1 - e^{\mathrm{i}\langle s_j,x\rangle}\Big)\right] \frac{d\widetilde{W}(x)}{|x|^{H+N/2}}. \end{aligned}$$

It follows that

$$\begin{aligned} & \mathbb{E}\left(X(t) - \sum_{j=0}^{n} a_j X(s_j)\right)^2 \\ = & \int_{\mathbb{R}^N} \left| e^{\mathrm{i}\langle t,x\rangle} - \sum_{j=0}^{n} a_j e^{\mathrm{i}\langle s_j,x\rangle}\right|^2 \frac{dx}{|x|^{2H+N}}, \end{aligned}$$

where $s_0 = 0$ and $a_0 = 1 - \sum_{j=1}^{n} a_j$.

We introduce a "bump" function $g \in C_c^\infty(\mathbb{R}^N)$ such that $g(0) = 1$, $0 \le g(x) \le 1$ and $g(x) = 0$ if $|x| \ge 1$. Let $\widehat{g}$ be the Fourier transform of g, i.e.,

$$\widehat{g}(\xi) = \int_{\mathbb{R}^N} e^{\mathtt{i}\langle \xi, x\rangle} g(x)\, dx.$$

Then $\widehat{g}(\xi)$ is in $C^\infty(\mathbb{R}^N)$ and decays rapidly (i.e., faster than any power rate) to 0 as $\xi \to \infty$. Define $g_r(x) = r^{-N} g(\frac{x}{r})$ and then $\widehat{g}_r(\xi) = \widehat{g}(r\xi)$.

Now, the Fourier inversion formula gives

$$\begin{aligned}
&\int_{\mathbb{R}^N} \Big(- e^{\mathtt{i}\langle t, x\rangle} + \sum_{j=0}^{n} a_j e^{\mathtt{i}\langle s_j, x\rangle} \Big) \widehat{g}_r(x)\, e^{-\mathtt{i}\langle t, x\rangle}\, dx \\
= \ & -\int_{\mathbb{R}^N} \widehat{g}_r(x)\, dx + \sum_{j=0}^{n} a_j \int_{\mathbb{R}^N} e^{-\mathtt{i}\langle t - s_j, x\rangle} \widehat{g}_r(x)\, dx \\
= \ & -(2\pi)^d r^{-N} + (2\pi)^d \sum_{j=0}^{n} a_j g_r(t - s_j) \\
= \ & -(2\pi)^d r^{-N}.
\end{aligned}$$

The last equality follows from the fact that $|s_j - t| \ge r$ $(j = 1, \ldots, n)$.

Squaring both sides of above and using the Cauchy-Schwarz inequality, we have:

$$
\begin{aligned}
& (2\pi)^{2d}\, r^{-2N} \\
= \;& \left| \int_{\mathbb{R}^N} \Big(e^{\mathtt{i}\langle t,x\rangle} - \sum_{j=0}^{n} a_j e^{\mathtt{i}\langle s_j,x\rangle} \Big) \widehat{g}_r(x) e^{-\mathtt{i}\langle t,x\rangle}\, dx \right|^2 \\
\leq \;& \left(\int_{\mathbb{R}^N} \left| e^{\mathtt{i}\langle t,x\rangle} - \sum_{j=0}^{n} a_j e^{\mathtt{i}\langle s_j,x\rangle} \right|^2 \frac{dx}{|x|^{2H+N}} \right) \\
& \cdot \left(\int_{\mathbb{R}^N} \left| \widehat{g}_r(x) \right|^2 \cdot |x|^{2H+N} dx \right) \\
= \;& \mathbb{E} \left(X(t) - \sum_{j=0}^{n} a_j X(s_j) \right)^2 \\
& \cdot \int_{\mathbb{R}^N} \left| \widehat{g}(rx) \right|^2 \cdot |x|^{2H+N} dx \\
= \;& \mathbb{E} \left(X(t) - \sum_{j=0}^{n} a_j X(s_j) \right)^2 \cdot r^{-2H-2N} \\
& \cdot \int_{\mathbb{R}^N} \left| \widehat{g}(y) \right|^2 \cdot |y|^{2H+N}\, dy \\
= \;& C\, \mathbb{E} \left(X(t) - \sum_{j=0}^{n} a_j X(s_j) \right)^2 r^{-2H-2N},
\end{aligned}
$$

where $C > 0$ is a constant. This yields the result.

Spectral Analysis of Fractional Gaussian Noise

We start with some general definition and results.

Let $\{X_n, n \in \mathbb{Z}^d\}$ be a weakly stationary stochastic process with values in $\mathbb{C}$. We assume that $\mathbb{E}(X_n) = 0, \forall n \in \mathbb{Z}^d$. Then the covariance function $r(n) = \mathbb{E}(X(0)\overline{X(n)}), \forall n \in \mathbb{Z}^d$ satisfies the following conditions:
(i). $r(-n) = \overline{r(n)}$.
(ii). $r(n), n \in \mathbb{Z}^d$ is nonnegative definite, i.e. $\forall M, \forall a_j \in \mathbb{C}, j = 1, ..., M$,

$$\sum_{j=1}^{M} \sum_{k=1}^{M} r(j-k) a_j \overline{a_k} \geq 0.$$

Thm: (Hergoltz). If $r(n), n \in \mathbb{Z}$ is a non-negative definite function with values in $\mathbb{C}$, then there exists a unique positive finite Borel measure μ on $[-\pi, \pi]$ such that $\mu\{-\pi\} = 0$ and

$$r(n) = \int_{-\pi}^{\pi} e^{\mathrm{i}nx} \, \mu(dx).$$

A proof of the following Bochner's theorem can be found in Janson (1997).

Thm: (Bochner). If $\{r(n), n \in \mathbb{Z}^d\}$ is a non-negative definite function with values in $\mathbb{C}$, then there exists a unique positive finite Borel measure on $(-\pi, \pi]^d$ such that

$$r(n) = \int_{(-\pi,\pi]^d} e^{\mathtt{i}\langle n,x\rangle}\, \mu(dx), \quad \forall\ n \in \mathbb{Z}^d.$$

If $\{r(t), t \in \mathbb{R}^N\}$ is a non-negative definite function with values in $\mathbb{C}$, then there exists a unique positive and finite Borel measure μ on $\mathbb{R}^N$ so that

$$r(t) = \int_{\mathbb{R}^N} e^{\mathtt{i}\langle t,x\rangle}\, \mu(dx).$$

Applying Bochner's Theorem, we derive the following representation theorem for stationary Gaussian processes.

Thm:(I). Let $\{X_n, n \in \mathbb{Z}^d\}$ be a centered stationary Gaussian process taking values in $\mathbb{R}$. Then there exists a unique, symmetric Borel measure on $(-\pi,\pi]^d$ such that $r(n) = \int_{(-\pi,\pi]^d} e^{\mathrm{i}\langle n,x\rangle}\,\mu(dx)$ holds. Let $\tilde{W}$ be the centered, complex Gaussian random measure with μ as its control measure. Then

$$X_n = \int_{(-\pi,\pi]^d} e^{\mathrm{i}<n,x>}\tilde{W}(dx), \quad \forall\ n \in \mathbb{Z}^d.$$

(II). Let $X = \{X(t), t \in \mathbb{R}^N\}$ be a centered, stationary Gaussian random field with values in $\mathbb{R}$. Then there exist a unique symmetric positive and finite Borel measure μ on $\mathbb{R}^N$ and a complex Gaussian random measure $\tilde{W}$ on $\mathbb{R}^N$ with control measure μ such that $r(t) = \int_{\mathbb{R}^N} e^{\mathrm{i}\langle t,x\rangle}\,\mu(dx)$ holds and

$$X(t) = \int_{\mathbb{R}^N} e^{\mathrm{i}\langle t,x\rangle}\,\tilde{W}(dx).$$

In the above, μ is called the spectral measure of X. If $\mu << \lambda_N$, then its density function is called the spectral density of X.

Ex: Let $B^H = \{B^H(t), t \in \mathbb{R}\}$ be a standard fractional Brownian motion with values in $\mathbb{R}$ with index H ("standard" means $\mathbb{E}\,(B^H(1)^2) = 1$).

Define $Y_n = B^H(n+1) - B^H(n), \forall n \in \mathbb{Z}$, then $\{Y_n, n \in \mathbb{Z}\}$ is a centered stationary Gaussian process (time series) with covariance function

$$r(n) = \mathbb{E}\,(Y_0 Y_n) = \frac{1}{2}(|n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H}).$$

$\{Y_n, n \in \mathbb{Z}\}$ is called the fractional Gaussian noise (fGn) of index H.

By the moving average and harmonizable representations of fBm we have the following representations for fGn:

$$Y_n = \frac{1}{c_1(H)} \int_{-\infty}^{n+1} \left[(n+1-x)^{H-1/2} - (n-x)_+^{H-1/2} \right] dB(x)$$

and

$$Y_n = \frac{1}{c_4(H)} \int_{-\infty}^{\infty} \frac{1-e^{ix}}{|x|^{H+1/2}} e^{inx}\, \widetilde{W}(dx).$$

The following theorem concerns the spectral density of fGn.

Thm: The fractional Gaussian noise, $\{Y_n, n \in \mathbb{Z}\}$, has a spectral density given by

$$h(x) = \frac{1}{c_4^2(H)} |e^{\mathrm{i}x} - 1|^2 \sum_{k \in \mathbb{Z}} \frac{1}{|x + 2k\pi|^{2H+1}}, \qquad \forall\ x \in (-\pi, \pi]$$

In particular, we have

$$h(x) \sim \frac{1}{c_4^2(H)} |x|^{1-2H} \qquad \text{as } x \to 0.$$

•Note that, if $H > 1/2$, the spectral density $h(x)$ has a singularity at $x = 0$. This is equivalent to the fact that the fractional Gaussian noise $\{Y_n,\, n \in \mathbb{Z}\}$ is long-range dependent, i.e. $\sum_{n=1}^{\infty} |r(n)| = \infty$. This motivates the two (essentially equivalent) definitions of long-range dependence for a stationary process in terms of the correlation function or the spectral density.

Observe that, when if $H > 1/2$, the classical "central limit theorem" does not hold for $\{Y_n\}$. That is, if you wish to have

$$\frac{1}{d_N}\sum_{n=0}^{N-1} Y_n \Rightarrow N(0, \sigma^2),$$

then you can not take $d_N = \sqrt{N}$. Instead, we must take $d_N \simeq N^H$.

In the following, we give a more general distributional limit theorem for stationary Gaussian sequences. Other related results can be found in Davydov (1970), Taqqu (1975) for moving average processes, Enriquez (2004) for correlated random walks.

Thm: Let $\{Z_j, j \in \mathbb{Z}\}$ be a stationary sequence of Gaussian random variables with mean 0 and covariance function $r(j)$ satisfying the following conditions: for a constant $H \in (0,1)$,

(i). if $H > 1/2$, then $r(j) \sim C_1\, j^{2H-2}$ as $j \to \infty$, where $C_1 > 0$ is a constant;
(ii). if $H = 1/2$, then $\sum_{j\in\mathbb{Z}} |r(j)| < \infty$ and $\sum_{j\in\mathbb{Z}} r(j) = C_2 > 0$;
(iii). if $0 < H < \frac{1}{2}$, then $r(j) \sim C_3\, j^{2H-2}$ as $j \to \infty$, where $C_3 < 0$ and $\sum_{j\in\mathbb{Z}} r(j) = 0$.

Then the finite dimensional distributions of the processes Z_N defined by

$$Z^{(N)}(t) = N^{-H} \sum_{j=1}^{[Nt]} Z_j, \quad 0 \le t \le 1\}$$

converge to those of $\{\sigma_0 B^H(t), 0 \le t \le 1\}$, where B^H is a standard fractional Brownian motion and

$$\sigma_0 = \begin{cases} C_1\, H^{-1}(2H-1)^{-1} & \text{if} \quad H > \frac{1}{2} \\ C_2 & \text{if} \quad H = \frac{1}{2} \\ -C_3\, H^{-1}(1-2H)^{-1} & \text{if} \quad H < \frac{1}{2}. \end{cases}$$