High-Dimensional Probability and Applications in Data Science

Roman Vershynin

1. The curse of dimensionality. Probability can help: the Monte-Carlo method.
2. Convexity. Caratheodory theorem. Approximate Caratheodory theorem. Proof by a probabilistic method, the empirical method of Maurey.
3. Applications of Approximate Caratheodory Theorem for financial portfolios and factor analysis. Covering numbers are usually exponential in the dimension. Polytopes with few vertices have small covering numbers.
4. Volumes of polytopes (Carl-Pajor's theorem). Milman's hyperbolic intuition in high dimensions. Concentration inequalities: from 68-95-99.7 rule to general gaussian tails.
5. Toward concentration of sums of independent random variables: Markov's and Chebyshev's inequalities. The error in the central limit theorem is too large (Berry-Esseen theorem). Hoeffding's inequality stated.
6. Hoeffding's inequality proved. Estimation of the mean: the median-of-means estimator.
7. Variants of Hoeffding's inequality: two-sided and for bounded distributions. Poisson limit theorem. Chernoff's inequality.
8. Small and large deviations: Gaussian and Poisson regimes. Random graphs as models of networks. Erdos-Renyi model. Phase transitions. Regularity of dense random graphs.
9. Irregularity of sparse random graphs. Instance vs. uniform guarantees of probabilistic results. Geometric discrepancy
10. Proof of the discrepancy theorem.
11. Spaces of random variables. Normed and Hilbert spaces. Lp spaces. Cauchy-Schwarz, Holder's and Jensen's inequalities.
12. Orlicz spaces. Subgaussian properties.
13. Equivalence of subgaussian properties. Subgaussian distributions and subgaussian norm.
14. Subgaussian Hoeffding's inequality. Subexponential distributions. Bernstein's inequality.
15. The thin shell phenomenon. Dimension reduction with Johnson-Lindenstrauss lemma.

16. Combinatorial optimization. Examples: Ising model, clustering, max-cut. Spectral relaxations. Semidefinite relaxations.

17. Gram matrices. Semidefinite relaxation of max cut: Goemans-Williamson's agorithm.

18. Grothendieck's inequality. Tensor calculus. Krivine's bound.

19. A first exposure to machine learning: binary classification. Support vector machine. The kernel trick: kernelizing machine learning algorithms. Radial basis function kernel.

20. The soft-margin SVM and kernel SVM. What functions are kernels? Mercer's condition. Artificial neural networks. Large width limits. Kernels arising from neural networks. Neural tangent kernel.

21. A refresher in linear algebra: spectral and singular value decompositions; Frobenius norm.

22. The operator norm. Random vectors in high dimensions. Covariance matrix. Normal distribution in high dimensions. Principal component analysis.

23. The covariance estimation problem. Sample covariance matrix. Reduction to stochastic processes. Brownian motion. Epsilon-nets. Computing the operator norm on an epsilon-net.

24. Covariance estimation for normal distributions. Spectrum perturbation theory: Weyl's and Davis-Kahan's inequalities. Implications for PCA.

25. Random matrices. Proof of Wigner's semicircle law using the Stieltjes transform approach.

26. Proof of Marchenko-Pastur law.

27. Review of fundamental laws of random matrix theory: Semicircle law, Marchenko-Pastur law, circular law, Bai-Yin law, Tracy-Widom law. Spike models. Joint eigenvalue distribution; eigenvalue repulsion. Universality. Connections to physics (Wigner surmise) and number theory (Montgomery's pair correlation conjecture).

28. Functional calculus. Loewner order. Matrix monotonicity of functions 1/x and log(x).

29. Lieb's trace inequality. Matrix Hoeffding inequality.

30. Matrix Bernstein inequality. Applications to networks (started): community detection in stochastic block models.

31. Applications to networks continued. A spectral algorithm for community recovery: guarantees for the stochastic block model.

32. Recent developments on community detection for the stochastic block model. Modern visualization techniques: MDS, Isomap, t-SNE, UMAP (play with it). What do numbers look like?

33. Low-dimensional paradigm. The effective rank and effective dimension. Covariance estimation for low-dimensional data.

34. Mathematical foundations of machine learning. Supervised learning. Loss functions. Overfitting problem. Hypothesis space. Risk; empirical risk; empirical risk minimization. Generalization error.

35. A generalization bound. VC dimension. The VC dimension of half-lines, intervals, half-planes, circles, rectangles, linear classifiers, polynomial classifiers, and neural networks.

36. VC theory: Pajor's lemma; Sauer-Shelah lemma.

37. Applications of Sauer-Shelah lemma for counting regions in hyperplane arrangements. Empirical processes. The uniform law of large numbers (stated).

38. Proof of the uniform law of large numbers: Symmetrization; Rademacher complexity. Applications: Glivenko-Cantelli theorem; VC generalization bound.

39. Gaussian and subgaussian stochastic processes. Dudley's intergral inequality.

40. Application of Dudley's inequality: the uniform law of large numbers for Lipschitz functions. Lipschitz regression.

41. Isoperimetric inequalities. Concentration of measure for Lipschitz functions via Gaussian isomepimetric inequality.

July 4, 2025