Math 833 # Introduction: integrable probability and branching graphs. Week 1

Central question of probability theory:
asymptotic behavior of large stochastic systems.

Approach of integrable probability: focus on special exactly-solvable or integrable cases, where the systems can be described by means of exact formulas for observables

a vague notion

## Benefits:

1. Formulas allow much finer asymptotic results
2. Once we understood integrable case, attacking generic systems becomes simpler: we have a base for predictions and conjectures.

Example 1: You are given a sequence of random variables $\xi_1, \xi_2, \ldots, \xi_n$ (on the same probability space)

Question: what is the distribution of their sum

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n \;?$$

In general, this is a hard question. Even if $\xi_i$ are independent, you either need to compute $(n-1)$ convolutions (integrals over $\mathbb{R}$) or make the Fourier transform (= compute characteristic function), multiply the results, and then make the inverse Fourier transform.

However, there are special nice cases.

We show one example here – another one will appear in mid-semester homework.

**Theorem 1:** Let $\xi_i$ be independent Bernoulli random variables

$$\xi_i = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p. \end{cases}$$

Then $S_n = \xi_1 + \ldots + \xi_n$ is a Binomial random variable

$$\text{Prob}\,(S_n = k) = p^k (1-p)^{n-k} \binom{n}{k}, \qquad 0 \le k \le n.$$

---

**Proof:** $S_n = k$ means that $k$ out of $n$ $\xi_i$ came $\xi_i = 1$ and $n-k$ others came $\xi_i = 0$. This has probability $p^k (1-p)^{n-k}$. Binomial coefficient $\binom{n}{k}$ arises from counting all ways to choose which $\xi_i$ are equal to 1. $\blacksquare$

---

The proof is not difficult. But we had to somehow guess to look into i.i.d. Bernoulli case! For generic $(\xi_1, .. \xi_n)$ this is much harder!

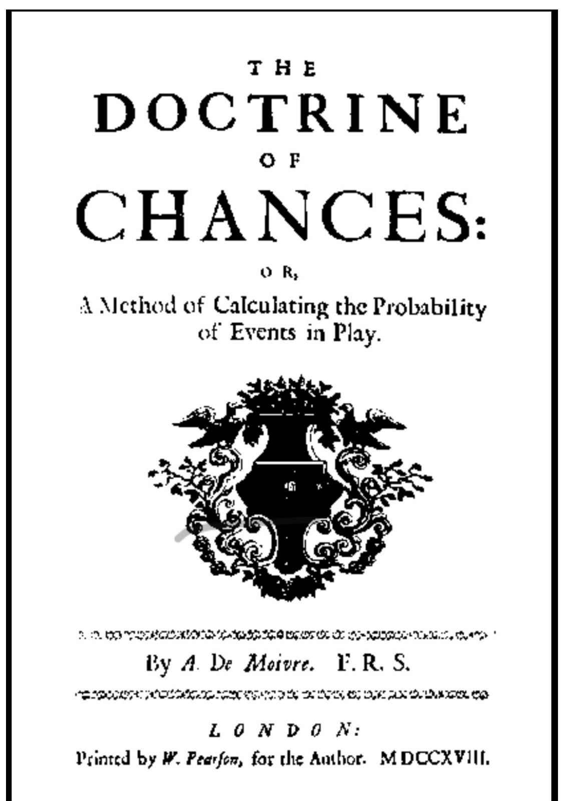
Is this useful for asymptotic questions? Very much!

**Theorem 2:**

$$\lim_{n\to\infty} \frac{S_n - np}{\sqrt{p(1-p)}} \overset{d}{=} N(0,1)$$

convergence in distribution

Gaussian random variable of mean 0 and variance 1.

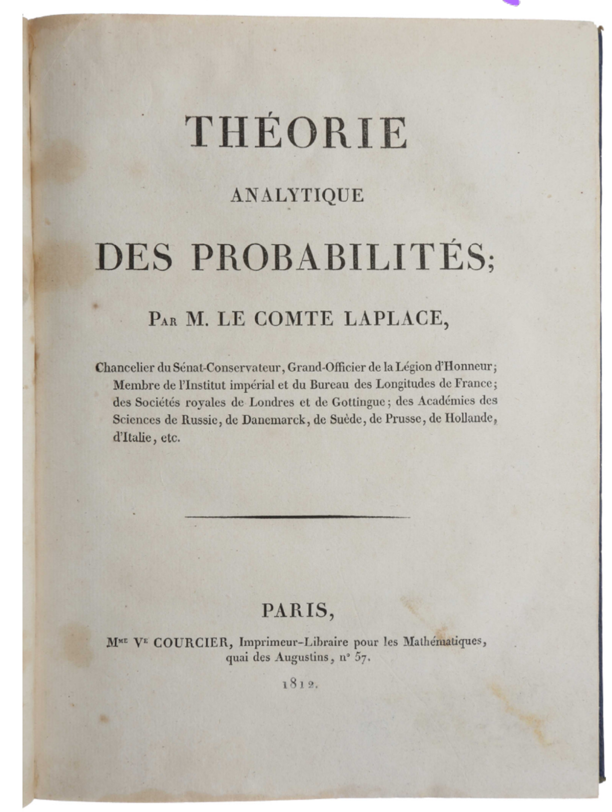


THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability of Events in Play.

By A. De Moivre. F. R. S.

LONDON:
Printed by W. Pearson, for the Author. MDCCXVIII.

← First proofs rely on Theorem 1 and asymptotic expansion of factorials in $\binom{n}{k}$ →



THÉORIE
ANALYTIQUE
DES PROBABILITÉS;
Par M. LE COMTE LAPLACE,

Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

PARIS,
Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.
1812.

- This was the very first instance of Central Limit Theorem
- This was the way Gaussian distribution first appeared.
- Later greatly extended (independent $\xi_i$, weakly-dependent, martingales, etc)

**Example 2:** Take $N \times N$ real symmetric random matrix and suppose that the joint law of its matrix elements is known.

**Question:** What is the distribution of its $N$ (real) eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ ?

In general, this is a very hard problem. Eigenvalues are zeros of characteristic polynomial, which is a degree $N$ polynomial with coefficients quite non-trivially depending on the matrix elements.

**However,** for some special choices of the (random) matrix elements, the answer is quite simple.

**Theorem 3:** Let $X$ be $N\times N$ matrix with independent Gaussian $N(0,1)$ matrix elements. Set $M=\frac{X+X^t}{2}$

Then the joint law of eigenvalues $\lambda_1\le\dots\le\lambda_N$ of $M$ has density proportional to

$$\prod_{1\le i<j\le N}(\lambda_j-\lambda_i)\prod_{i\ge 1}^{N}\exp\left(-\frac{\lambda_i^2}{2}\right)$$

Gaussian Orthogonal Ensemble (GOE)

- At $N=1$ this is the Gaussian density
- General $N$ proof will be given later in this class
- There are similar computations for $XX^t$ and for complex and quaternion matrices
- Origin: Representation theory (1920s), Multidimensional statistics (1930s), High energy physics (1950s)

Is this useful for asymptotis statements? Very much!

**Theorem 4:** Let $\lambda_1 \le \dots \le \lambda_N$ be eigenvalues of GOE.

Then

$$\lim_{N\to\infty} \sqrt{2}\, N^{1/6}\left(\lambda_N - \sqrt{2N}\right) = F_1$$

Convergence in distribution

Note scaling!

$\beta = 1$ (GOE)

Tracy-Widom distribution

On orthogonal and symplectic matrix ensembles

Craig A. Tracy & Harold Widom

Communications in Mathematical Physics **177**, 727–754 (1996) | Cite this article

ELSEVIER

Nuclear Physics B

Volume 402, Issue 3, 9 August 1993, Pages 709-728

The spectrum edge of random matrix ensembles

P.J. Forrester

Proofs are based on the technology of Pfaffian point processes applied to the result of Theorem 3.

- This was the very first appearance of the Tracy-Widom distribution
- Later $F_1$ was shown to be the scaling limit of largest eigenvalues of very general random matrices.

Example 3: Take any probability measure $\mu$ on $\mathbb{R}$.

Consider monic orthogonal polynomials $p_n(x)$:

$$\begin{cases} p_n(x) = x^n + \dots \\ \int_{\mathbb{R}} p_n(x)\, p_m(x)\, \mu(dx) = 0, \text{ unless } n = m. \end{cases}$$

Question: what are they explicitly?

There is a general recurrent algorithm:

$$p_n(x) = x^n + \sum_{k=0}^{n-1} c_n^k \cdot p_k(x), \quad \text{where}$$

$$c_n^k = - \frac{\int x^n p_k(x)\, \mu(dx)}{\int p_k(x) p_k(x)\, \mu(dx)}$$

Check that this works!

But there are many integrations involved and the final answer is unclear.

However, there are much better special cases!

**Theorem 5:** Suppose $\mu = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$ [Gaussian weight]

Then $p_n(x)$ are Hermite polynomials

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \left(\frac{\partial}{\partial x}\right)^n e^{-\frac{x^2}{2}} = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)!} \frac{x^{n-2m}}{2^m}$$

**Proof:** For $k<n$ we have

$$\int H_n(x)\, x^k e^{-\frac{x^2}{2}} dx = (-1)^n \int x^k \left(\frac{\partial}{\partial x}\right)^n e^{-\frac{x^2}{2}} dx \overset{\text{(integrate by parts)}}{=}$$

$$= (-1)^{n-1} \int k\, x^{k-1} \left(\frac{\partial}{\partial x}\right)^{n-1} e^{-\frac{x^2}{2}} dx = \underset{\text{(integrate by parts again)}}{\dots} = 0$$

$\blacksquare$

---

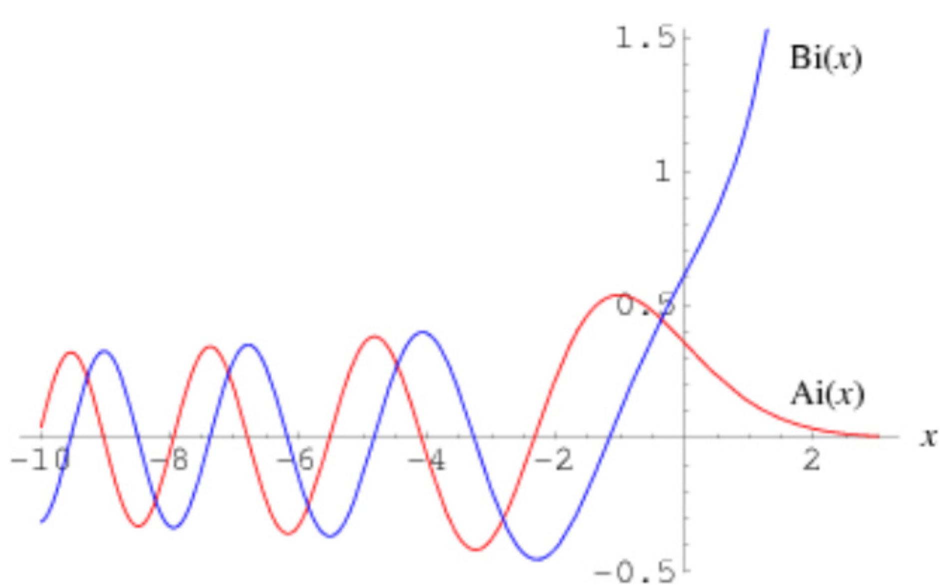
Proof is not hard, but is not going to work for generic $\mu$. We somehow needed to guess that this weight or polynomials are nice.

And again there are asymptotic statements. For instance:

**Theorem 6:** Let $x_1^n < x_2^n < \ldots < x_n^n$ be $n$ roots of Hermite polynomial $H_n(x)$. Then for $k=0,1,2,\ldots$

$$\lim_{n\to\infty} n^{1/6}\left(x_{n-k}^n - 2\sqrt{n}\right) = a_k$$

$a_k$ — $i$-th largest zero of the Airy function $Ai(x)$




$Ai(x)$ is a solution to

$f''(x) = x\, f(x)$, which decays as $x \to +\infty$.

- Theorem is not hard to prove using the formula for Hermite polynomials
- The same $a_k$ appear in largest zeros of very general orthogonal polynomials. (but this is much harder to show)

We saw so far:

- Among sequences of random variables, i.i.d. Bernoulli are especially nice.
- Among random matrices, Gaussian Orthogonal ensemble is especially nice.
- Among polynomials, Hermite polynomials are especially nice

---

These are exactly-solvable or integrable cases.

How do you find them?

---

Often some luck is involved in the first discoveries.
Yet we want to develop structural guiding principles.

Our approach to the hunt on integrable stochastic systems:

Let mathematics do the work!

In this class we obtain such systems as <u>answers</u> to classification problems.

They involve:

- Large dimension $N$ or even $N = \infty$.
- Invariance under groups of symmetries.

Eventually symmetries give rise to integrability.

## Back to example 1. I.i.d. Bernoulli as an answer.

**De Finetti's theorem.** Let $\xi_1, \xi_2, \ldots$ be an infinite sequence of random variables taking values $0/1$, such that its distribution is invariant under all finite permutations of coordinates.

Then $\{\xi_i\}$ is a mixture of i.i.d. Bernoulli's.

---

This means that the laws of $\{\xi_i\}$ are parameterized by measures $\mu$ on $[0,1]$. In order to sample $\xi_i$, one first samples $p \in [0,1]$ according to $\mu$ and then samples $\{\xi_i\}$ as a sequence of i.i.d. Bernoulli with $\mathrm{Prob}(\xi_i = 1) = p$.

---

In particular ergodic (or extreme) invariant measures are i.i.d. Bernoulli

do not have to understand these words — will be explained in class

## Back to example 2 — GOE as an answer

Consider infinite real-symmetric matrices $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & \cdots \\ \vdots & & \ddots \end{pmatrix}$

Any such matrix can be conjugated by an element $u$ of the infinite-dimensional orthogonal group $O(\infty)$

$$u = \begin{pmatrix} \boxed{\begin{matrix} N\times N \\ \text{orthogonal} \\ \text{matrix} \end{matrix}} & 0 \\ 0 & \begin{matrix} 1 & & \\ & 1 & \\ & & 1 \\ & & & \ddots \end{matrix} \end{pmatrix}$$

**Question:** What are all possible infinite random real-symmetric matrices with $O(\infty)$-invariant distribution?

**Answer:** There is a complete and concise classification and GOE is one of them.

[stay tuned to the class in order to learn about others!]

Back to example 3. Hermite polynomials as an answer

Consider a sequence of polynomials $p_n(x) = x^n + \dots$, $n=0,1,2,\dots$

It is called an Appell sequence if for each $n$

$$p_{n-1}(x) = \frac{1}{n}\frac{\partial}{\partial x} p_n(x)$$

---

Question: What are all possible Appell sequences of real-rooted polynomials?

all zeros of $p_n(x)$ are real

---

Answer: There is a (short) complete list and Hermite polynomials $p_n(x) = H_n(x)$ are in this list.

[Full theorem is coming next week!]

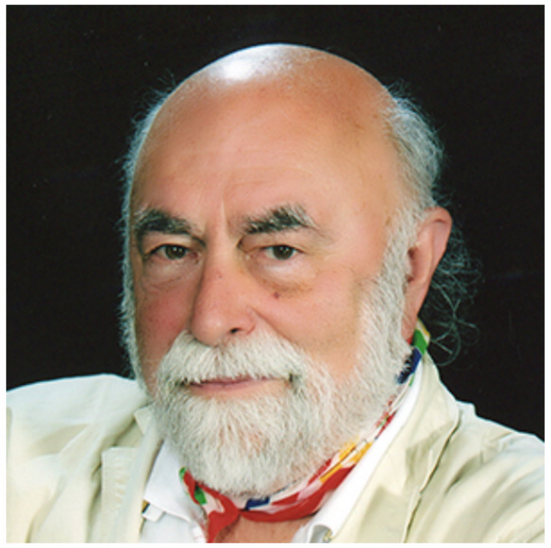
Summary:

Classification problems involving large dimensions and symmetries give rise to many distinguished stochastic systems of integrable probability.

A unifying framework for such problems has a name

identification of boundaries of branching graphs.

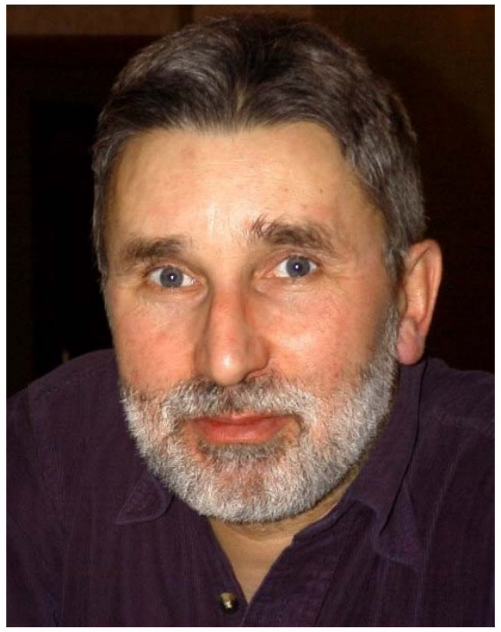
The setting is largely developped by russian mathematicians



Anatoly Vershik

and

Sergei Kerov



This is the main topic of our class.