Limiting distributions of Betti numbers

Georgia Tech. April 11, 2011

F. Rodriguez Villegas

X cplx manifold of dimension d

$b_i(X) := \dim H^i(X, \mathbb{C})$

Betti numbers. We will assume X connected but not necessarily compact.

Our manifolds will all have $b_i(X)=0$ for i odd so we consider

$$P_X(q) := \sum_{i=0}^{d} b_{2i}(X)\, q^i$$

the (even) Poincaré polynomial of X.

If X is compact then by Poincaré duality we have

$$q^d P_X(q^{-1}) = P_X(q)$$

We will discuss examples of sequences of manifolds $X_1, X_2, \ldots$ for which

$$i \longmapsto b_{2i}(X_n),$$

appropriately scaled, converges to a limiting distribution.

The question, vaguely formulated, is what is the distribution of the Betti numbers of random manifolds? (This question is mentioned by Reineke, Kontsevich)

In order to compute $P_X(q)$ we will use the following fact. Suppose X is a (smooth connected) algebraic variety $/\mathbb{C}$ such that

$$\# \mathcal{X}(\mathbb{F}_q) = C(q)$$

for all base changes $\phi : R \to \mathbb{F}_q$

$\mathbb{C} \supseteq R$ a finitely generated algebra $/\mathbb{Z}$

$\mathcal{X}/R$ a spreading out of X,

for some polynomial C indep of ϕ.

Assume that the MHS of X is pure (for example, X is smooth and projective)

Then

$$P(q) = q^d C(q^{-1})$$

Hard Lefschetz theorem implies that for X smooth & projective we also have

$$b_0 \le b_2 \le \cdot \le b_{2[d/2]}$$

Example 1

Fix $k \in \mathbb{N}$ and take

$$X = Gr_k^{k+n}$$

the Grassmannian parametrizing dim k subspaces of a fixed $(n+k)$-dim'l space.

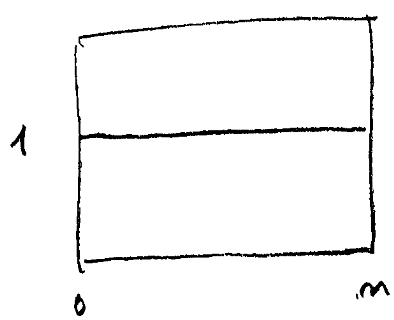
Then

$$\# X(\mathbb{F}_q) = \begin{bmatrix} n+k \\ k \end{bmatrix}$$

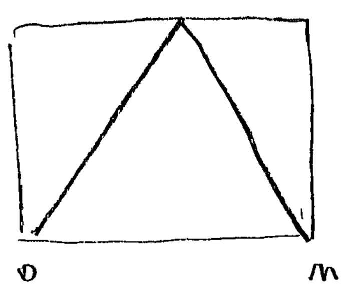
the q-binomial coefficient. If we plot the coefficients, say for $n = 200$

$\underline{k=1}$ projective space $\mathbb{P}^n$

unimodality follows from Hard Lefschetz.



$\underline{k=2}$



$$\begin{bmatrix} n+k \\ k \end{bmatrix} = \frac{(q^{n+k}-1)}{(q^{k}-1)} \cdots \frac{(q^{n+1}-1)}{(q-1)}$$

Each factor looks like

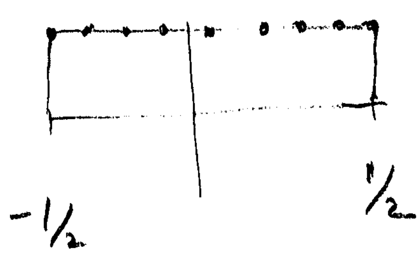
$$\frac{q^{\frac{n+j}{2}} - q^{-\frac{(n+j)}{2}}}{q^{j/2} - q^{-j/2}} \cdot q^{n/2}$$

If $n = jm$ this equals

$$q^{n/2} \cdot \sum_{i=0}^{m} q^{j(i-m/2)}$$

Dividing by $q^{n/2}$ and replacing q by $q^{1/n}$ we get

$$\sum_{i=0}^{m} q^{i/m - 1/2}$$



Discrete uniform measure in $[-1/2, 1/2]$.

Taking product of polynomials corresponds to convolution of measures. So we essentially have the k-th convolution of the uniform measure in $[-1/2, 1/2]$

for n large. By the law of large numbers for large k this is approximately Gaussian.

Example 2

Consider the Hilbert scheme $X^{[n]}$ of n points on $X = \mathbb{C}^2$. Parametrizes 0-dim'l subschemes of X of length n.

By a formula of Göttsche the counting polynomial

$$C_n(q) := \# X^{[n]}(\mathbb{F}_q)$$

is given by $q^n A_n(q)$, where A_n is the coeff of T^n in the infinite product

$$\prod_{m \geq 1} (1 - qT^m)^{-1} = 1 + qT + (q^2+q)T^2 + (q^3+q^2+q)T^3 + \cdots$$

Explicitly,

$$A_n(q) = \sum_{|\lambda| = n} q^{\ell(\lambda)},$$

where λ runs over all partitions of n and $\ell(\lambda)$ is the length

E.g. $n = 5$

λ	$\ell(\lambda)$
1 1 1 1 1	5
2 1 1 1	4
2 2 1	3
3 1 1	3
3 2	2
4 1	2
5	1

$$A_5(q) = q^5 + q^4 + 2q^3 + 2q^2 + q$$

Erdös and Lehner proved that the coefficients of A_n, when appropriately scaled, converge to the Gumbel distribution. More precisely,

$$\#\{\lambda \mid |\lambda| = n,\ \ell(\lambda) = k\} \approx g\left(\frac{k - a_n}{b_n}\right)$$

where

$$g(x) = e^{-x - e^{-x}}$$

$$\begin{cases} a_n = \frac{1}{c}\, n^{1/2} \log(4n/c^2) \\ b_n = 2n^{1/2}/c \end{cases} \qquad c = \pi\sqrt{2/3}$$

The Gumbel distribution has a universality property. It is the limiting distribution of properly normalized maxima of a sequence of identically distributed, independent random variables. This is a sort of central limit theorem but for max instead of sum.

Example 3

Let Γ be a finite graph. Given a choice of non-negative integers for each vertex of Γ (a dimension) Nakajima defined an associated hyperkähler manifold X_Γ. In the case that all dimensions equal one (Hausel-Sturmfels) the counting polynomial of X_Γ is (up to a power of $q^{\dim X_\Gamma}$) the reliability polynomial of Γ. This is an invariant of graphs which is the specialization

$$q^{\dim X_\Gamma} T_\Gamma(1, q) = C(q)$$

of the Tutte polynomial $T_\Gamma(x, y)$ of the graph.

Consider the case $\Gamma = K_n$, the complete graph on n vertices. Let $A_n(q) = T_{K_n}(1,q)$. It follows from the defn of the Tutte polynomial that

$$A_n(q+1) = \sum_{k \geq 0} c_{n,k}\, q^k$$

has coefficients

$$c_{n,k} = \#\{\text{connected graphs on } n \text{ labeled vertices of genus } k\}$$

If we let

$$A_n(q) = \sum_{k \geq 0} a_{n,k}\, q^k$$

then

$$c_{n,k} = \sum_{i \geq 0} a_{n,i} \binom{i}{k}$$

These are the factorial moments of the discrete distribution $i \mapsto a_{n,i}$.

The asymptotics of $c_{n,k}$ for fixed k and $n \to \infty$ was computed by Wright (of Hardy and Wright fame)

$$C_{n,k} \sim w_K \, n^{n-2+\frac{3}{2}k}\left(1+O(n^{-1/2})\right)$$

for certain constants w_K.

There is a unique distribution w/ this moments: the Airy distribution. This distrib appears in several different contexts. (Brownian excursions, analysis of algorithms, ...)

Growth function on finite abelian gp V

~~take V vector space of dim d/Fq~~

Let $S = \{v_1, \ldots, v_N\}$ be a set of generators for V. Define

$$|v| := \min\left\{ \sum_{i=1}^{N} a_i \;\middle|\; v = \sum_{i=1}^{N} a_i v_i,\ a_i \geq 0 \right\}$$

Distance from the origin to v in the directed Cayley graph w/ these generators.

For a graph take $V = \partial \alpha C$, $S = \{ v_1 - v_0, \ldots, v_{n-1} - v_0 \}$

Enumerator $\displaystyle\sum_{v \in V} q^{|v|}$.

Take for example K_n then

$$\mathrm{Jac} \simeq (\mathbb{Z}/n)^{n-2}$$

$$S = \{e_1 + w, \dots, e_{n-2} + w, w\}$$

where $w = (1, \dots, 1)$

For a graph Enumerator $(q) = q^{\deg} C(q^{-1})$

what do we get in general for $\{v_1, \dots, v_N\}$ generators of $V =$

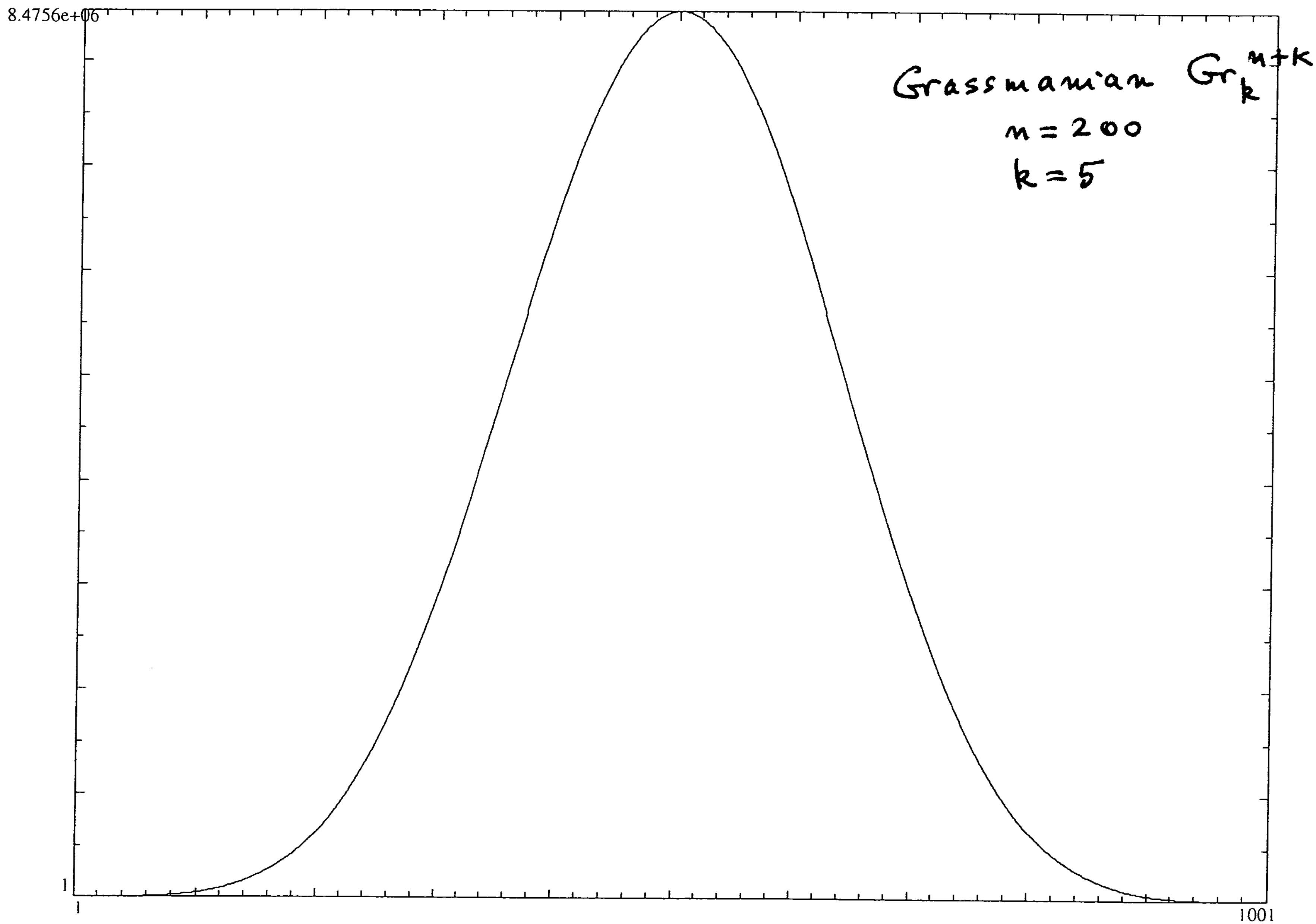
V vector space of dim d / $\mathbb{F}_q$

$N = d + r$

$\underline{r = 0}$ $\longrightarrow$ Gaussian

$\underline{r = 1}$ $\longrightarrow$ Airy?

$\underline{r > 1}$ $\longrightarrow$?


8.4756e+06
1
1
1001
Grassmanian Gr_k^{n+k}
$n = 200$
$k = 5$

S. Majumdar, A. Comtet J. Stat. Phys. 119 (2005)

where $b_k = 2\alpha_k^3/27$ and $U(a, b, z)$ is the confluent hypergeometric function.[4] The function $f(x)$ has the asymptotic tails,[3,5]

$$f(x) \sim x^{-5} e^{-2\alpha_1^3/27x^2} \quad \text{as} \quad x \to 0$$
$$f(x) \sim e^{-6x^2} \quad \text{as} \quad x \to \infty. \tag{3}$$

A plot of this function, obtained by evaluating the sum in Eq. (2) using the Mathematica, is provided in Fig. 1.

So, why would anyone care about such a complicated function? The reason behind the sustained interest and study[3,5–8] of this function $f(x)$ seems to be the fact that it keeps resurfacing in a number of seemingly unrelated problems, mostly in computer science and graph theory [for a list of such problems see ref. 6]. For example, the function $f(x)$ describes the distribution of the cost of the construction of a linear table for data storage using the linear probing with random hashing algorithm.[6] The function $f(x)$ also describes the distribution of the total path length in Catalan trees.[3] The generating function for the number of inversions in trees involves the Airy distribution function $f(x)$.[9] Also, the moments M_n's of the function $f(x)$ appear in the enumeration of the connected components in a random graph.[10,11] Recently, it has been conjectured and subsequently tested numerically that the asymptotic distribution of the area of two dimensional self-avoiding polygons is also given by the Airy distribution function $f(x)$.[12] Besides, numerical evidence suggests

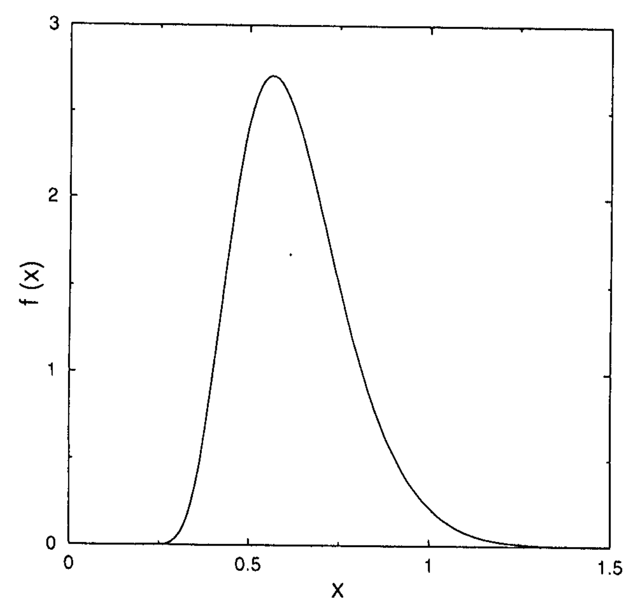
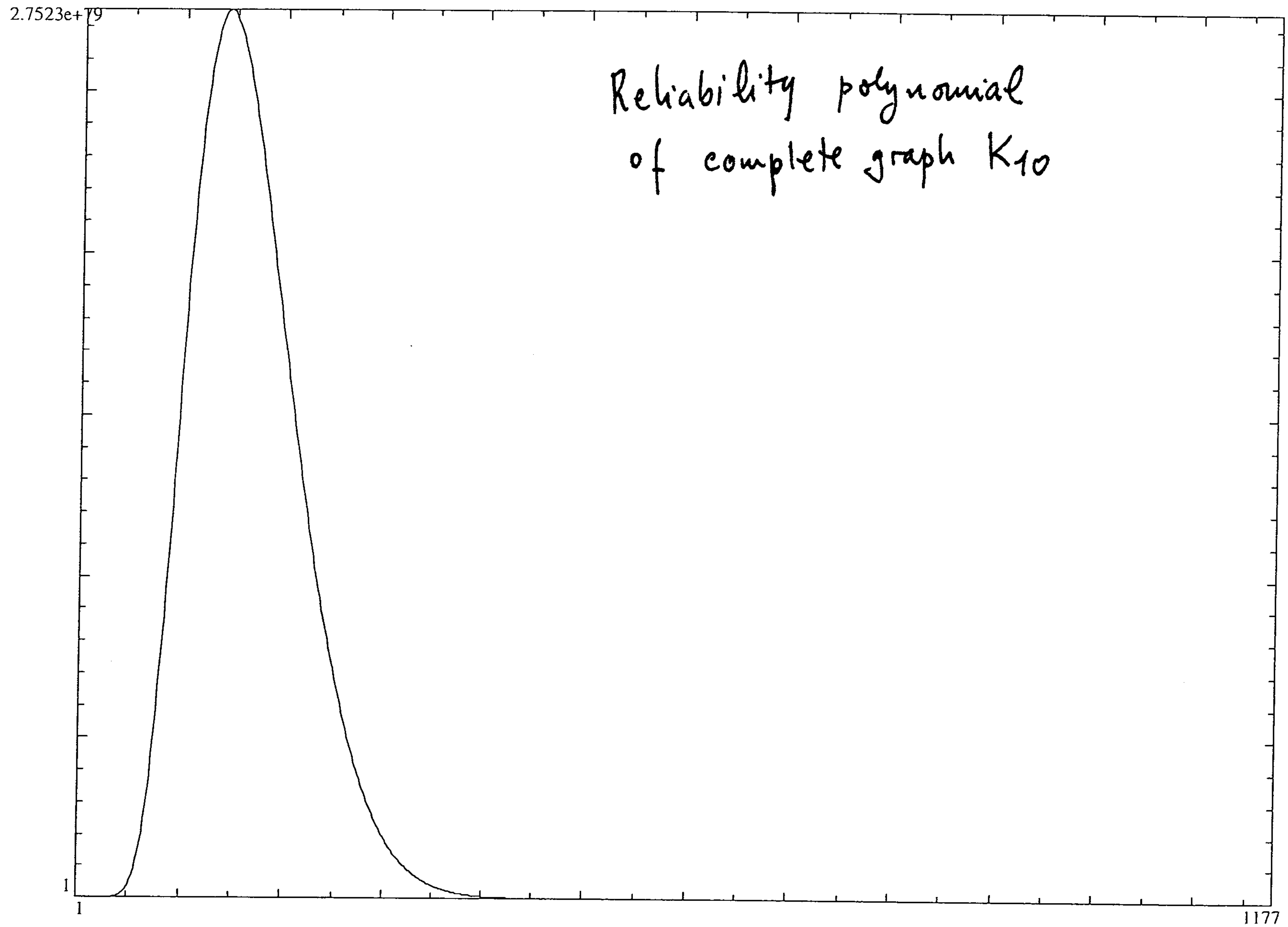
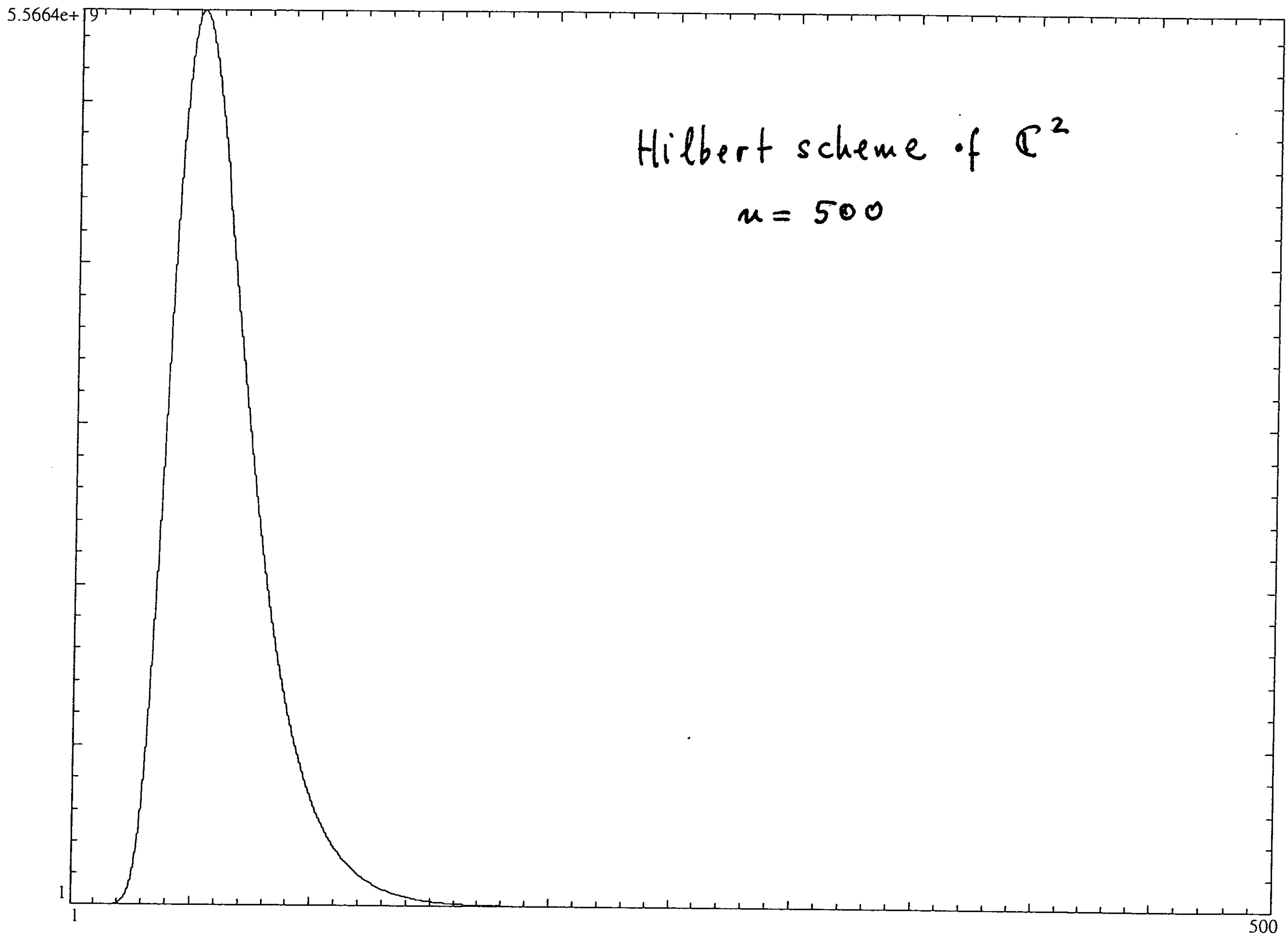
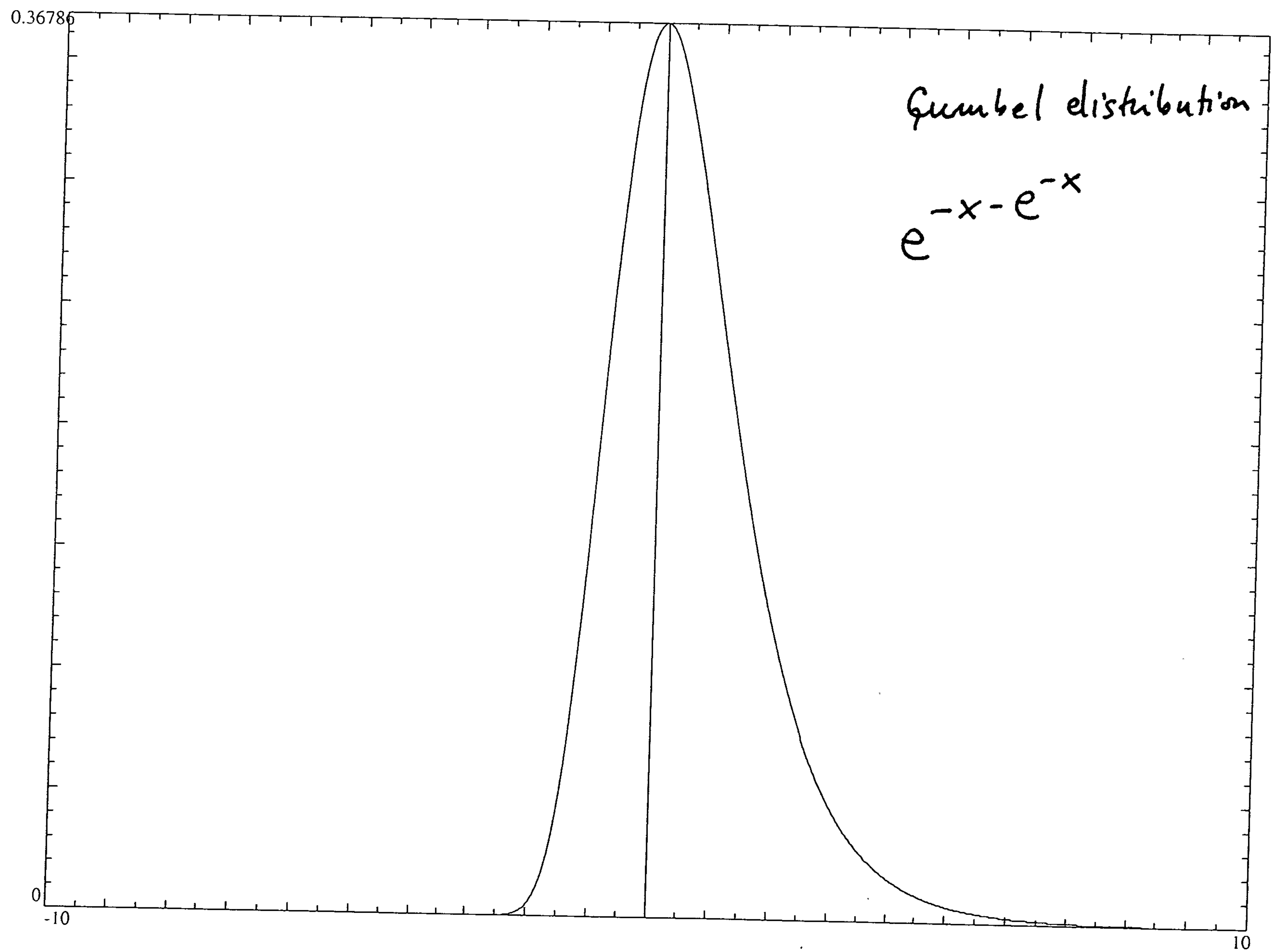


Fig. 1. A Mathematica plot of the Airy distribution function $f(x)$ in Eq. (2).


2.7523e+79
1
1
1177
Reliability polynomial
of complete graph K_{10}


5.5664e+19
1
1
500
Hilbert scheme of $\mathbb{C}^2$
$n = 500$



Gumbel distribution
$e^{-x-e^{-x}}$
0.36786
0
-10
10