Statistics of Modular Symbol

William Stein[1]

University of Washington

wstein@uw.edu

April 7, 2015

[1] Joint work-in-progress with Barry Mazur and Karl Rubin.

Overview

Modular symbols and L-functions

Statistics of modular symbols

Not-so-random walks: sums of modular symbols

Modular symbols associated to an elliptic curve

- *Elliptic curve:* $E/\mathbb{Q}$
- *Period mapping:* integration defines a map $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\} \to \mathbb{C}$ given by $\alpha \mapsto \int_\alpha^\infty 2\pi i f(z) dz$.
- *Homology:* $H_1(E, \mathbb{Z}) \cong \Lambda_E \subset \mathbb{C}$ is the image of all integrals of closed paths in the upper half plane, and $E(\mathbb{C}) \cong \mathbb{C}/\Lambda_E$.
- *Complex conjugation:* $\Lambda_E^+ \oplus \Lambda_E^- \subset \Lambda_E$ has index 1 or 2. Write $\Lambda_E^+ = \mathbb{Z}\omega^+$, where $\omega^+ > 0$ is well defined.
- *Modular symbols:* $[\alpha]_E^+ : \mathbb{P}^1(\mathbb{Q}) \to \mathbb{Q}$ defined by

$$\frac{1}{2\omega^+}\left(\int_\alpha^\infty 2\pi i f(z) dz + \int_{-\alpha}^\infty 2\pi i f(z) dz\right) = [\alpha]_E^+ \cdot \omega^+$$

  (similar for $[\alpha]_E^-$.)
- Explain on the blackboard how the integral above “works” for $\alpha = \infty$ using $f(z) = \sum a_n e^{2\pi i n z}$.

Example

We compute some modular symbols using Sage. Despite the numerical definitions above, the following computations are entirely algebraic.

```
E = EllipticCurve('11a')
s = E.modular_symbol()
s(17/13)
```

```
-4/5
```

Let's compute more symbols with denominator 13:

```
[s(n/13) for n in [-13..13]]
```

```
[1/5, -4/5, 17/10, 17/10, -4/5, -4/5, -4/5, -4/5, -4/5,
-4/5, 17/10, 17/10, -4/5, 1/5, -4/5, 17/10, 17/10, -4/5,
-4/5, -4/5, -4/5, -4/5, -4/5, 17/10, 17/10, -4/5, 1/5]
```

Lots of random-looking rational numbers... patterns...?
Summetry: $[a/M]^+ = [-a/M]^+$ and $[1 + (a/M)]^+ = [a/M]$.

A motivation for considering modular symbols: L-functions

L-series of E: $L(E,s) = \sum a_n n^{-s}$, where $a_p = p + 1 - \#E(\mathbb{F}_p)$.

For each Dirichlet character $\chi : (\mathbb{Z}/M\mathbb{Z})^* \to \mathbb{C}^*$ there is a twisted L-function $L(E,\chi,s) = \sum \chi(n) a_n n^{-s}$. Moreover,

$$\frac{L(E,\chi,1)}{\omega_\chi} = \text{explicit sum involving } \left[\frac{a}{M}\right]_E^{\pm} \text{ and Gauss sums}$$

(The details are not important for this talk...)

Thus statistical properties of the set of numbers

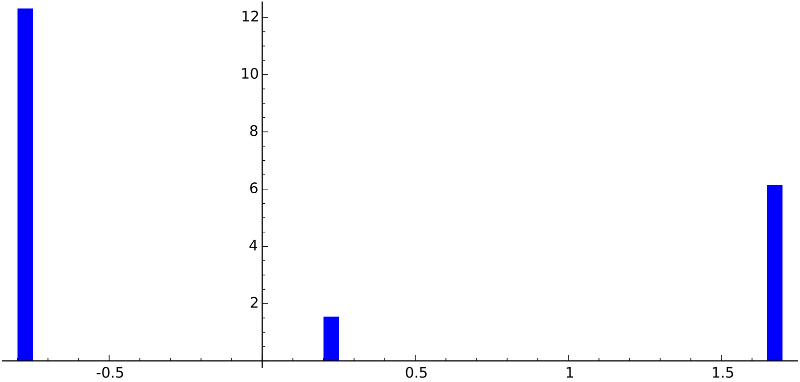
$$Z(M) = \left\{ \left[\frac{a}{M}\right]_E^{+} : a = 0, \ldots, M-1 \right\}$$

are relevant to understanding special values of twists.

Frequency histogram: $M = 13$

```
E = EllipticCurve('11a')
s = E.modular_symbol()
M = 13
print([s(a/M) for a in range(M)])
stats.TimeSeries([s(a/M) for a in range(M)]).plot_histogram()
```

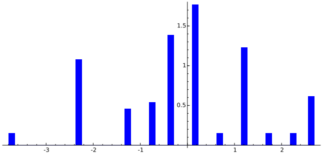
```
[1/5, -4/5, 17/10, 17/10, -4/5, -4/5, -4/5,
-4/5, -4/5, -4/5, 17/10, 17/10, -4/5]
```



Frequency histogram: $M = 100$

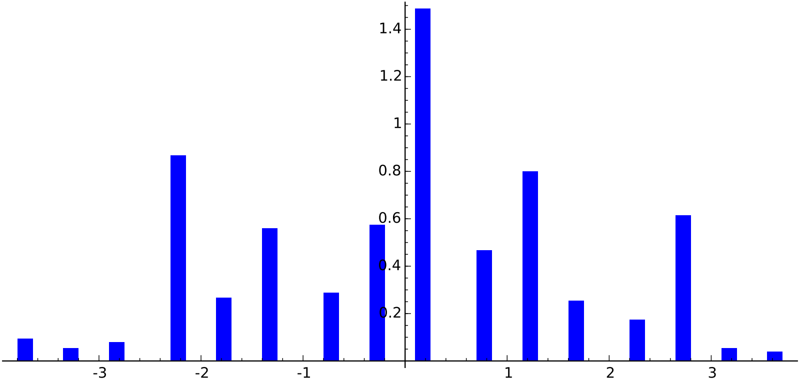
```
E = EllipticCurve('11a'); s = E.modular_symbol()
M = 100; v = [s(a/M) for a in range(M)]; print(v)
stats.TimeSeries(v).plot_histogram()
```

```
[1/5, 1/5, 6/5, 1/5, -3/10, -4/5, 6/5, 1/5, -3/10, 1/5,
1/5, 1/5, -3/10, 1/5, 6/5, 17/10, 11/5, 27/10, 6/5, 1/5,
6/5, 27/10, 6/5, 27/10, -3/10, 7/10, 6/5, 1/5, -3/10,
27/10, 1/5, -23/10, -3/10, 1/5, -13/10, -4/5, -3/10,
-23/10, 6/5, -23/10, -13/10, -23/10, -19/5, -23/10,
-3/10, -4/5, -13/10, -23/10, -3/10, -23/10, -4/5,
-23/10, -3/10, -23/10, -13/10, -4/5, -3/10, -23/10,
-19/5, -23/10, -13/10, -23/10, 6/5, -23/10, -3/10,
-4/5, -13/10, 1/5, -3/10, -23/10, 1/5, 27/10, -3/10,
1/5, 6/5, 7/10, -3/10, 27/10, 6/5, 27/10, 6/5, 1/5,
6/5, 27/10, 11/5, 17/10, 6/5, 1/5, -3/10, 1/5, 1/5,
1/5, -3/10, 1/5, 6/5, -4/5, -3/10, 1/5, 6/5, 1/5]
```



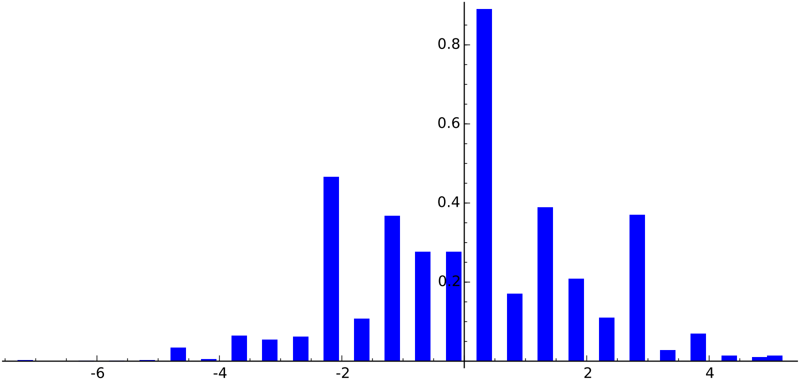
Frequency histogram: $M = 1000$

```
E = EllipticCurve('11a')
s = E.modular_symbol()
M = 1000
stats.TimeSeries([s(a/M) for a in range(M)]).plot_histogram()
```



Frequency histogram: $M = 10000$

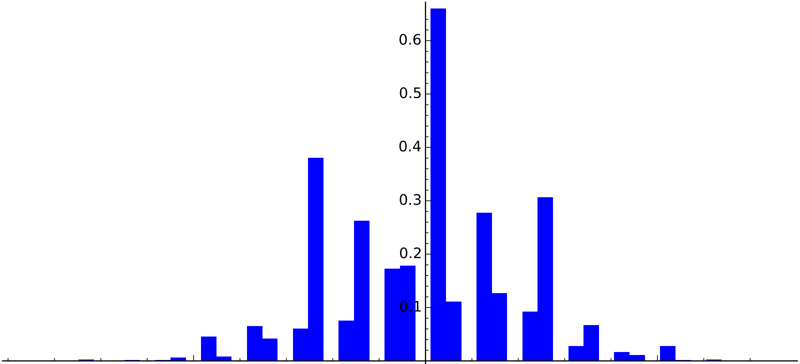
```
E = EllipticCurve('11a')
s = E.modular_symbol()
M = 10000
stats.TimeSeries([s(a/M) for a in range(M)]).plot_histogram()
```



We quickly want **much** larger M in order to see what might happen in the limit, and the code in Sage is way too slow for this...

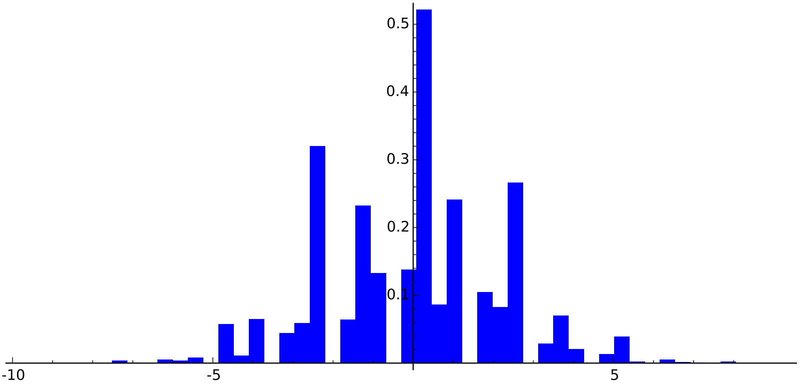
More frequency histograms: need Cython...

```
%load modular_symbol_map.pyx
def ms(E, sign=1):
    g = E.modular_symbol(sign=sign)
    h = ModularSymbolMap(g)
    d = float(h.denom)  # otherwise get int division!
    return lambda a,b: h._eval1(a,b)[0]/d
s = ms(EllipticCurve('11a'))
M = 100000  # the following takes about 1 second
stats.TimeSeries([s(a, M) for a in range(M)]).plot_histogram()
```



More frequency histograms (Cython)

```
s = ms(EllipticCurve('11a'))
M = 1000000  # the following takes about 1 second
stats.TimeSeries([s(a, M) for a in range(M)]).plot_histogram()
```



Note that there are only 38 distinct values in $Z(10^6)$ and 40 in $Z(1500000)$.

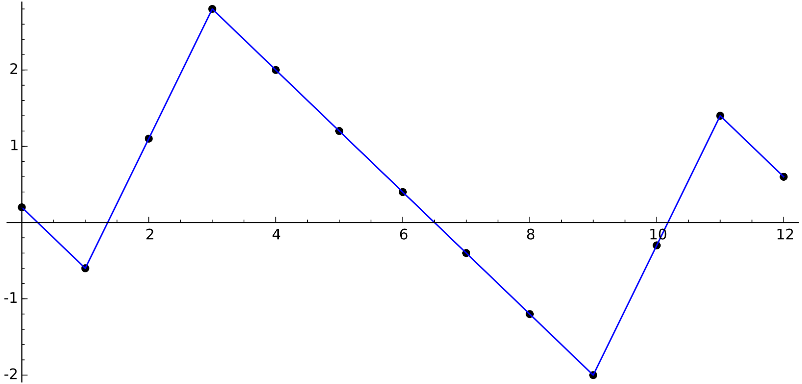
Sorry...

- But I can't tell you "the answer" yet.
- Since I'm not sure what to ask or even if *this* is a good question...
- So let's consider another question.

Return to $M = 13$ and make a random walk

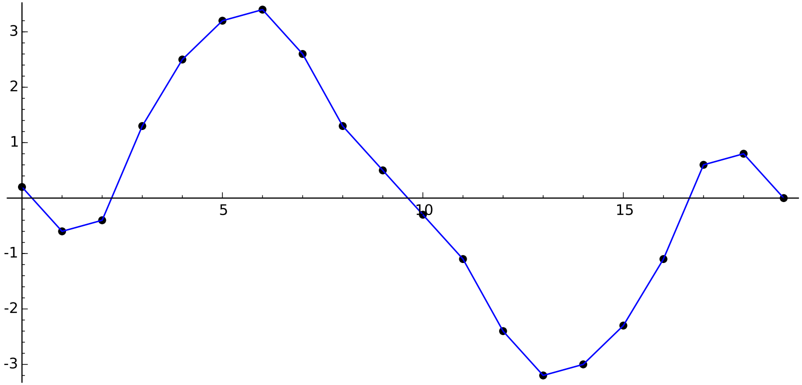
```
E = EllipticCurve('11a')
s = E.modular_symbol()
M = 13; v = [s(a/M) for a in range(M)]; print(v)
w = stats.TimeSeries(v).sums()
w.plot() + points(enumerate(w), pointsize=30, color='black')
```

```
[1/5, -4/5, 17/10, 17/10, -4/5, -4/5, -4/5, -4/5,
-4/5, -4/5, 17/10, 17/10, -4/5]
```

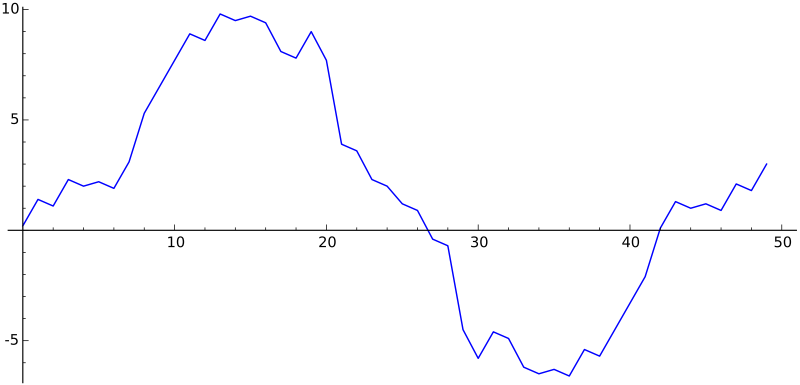


How about $M = 20$?

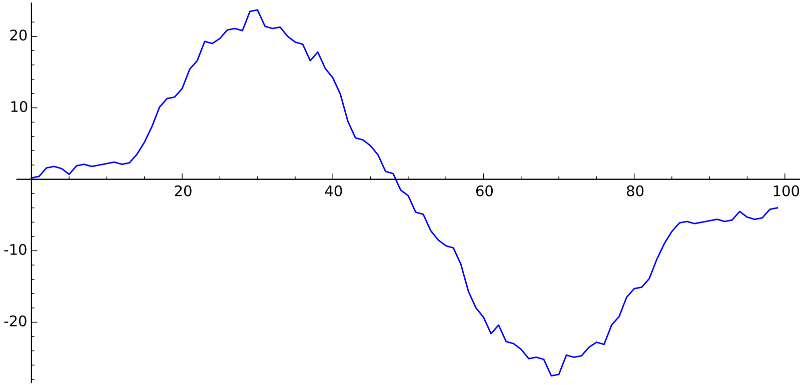
```
s = EllipticCurve('11a').modular_symbol()
M = 20; v = [s(a/M) for a in range(M)]
w = stats.TimeSeries(v).sums()
w.plot() + points(enumerate(w), pointsize=30, color='black')
```



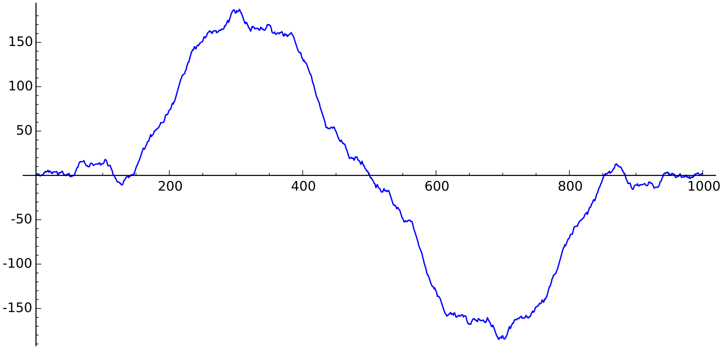
How about $M = 50$?



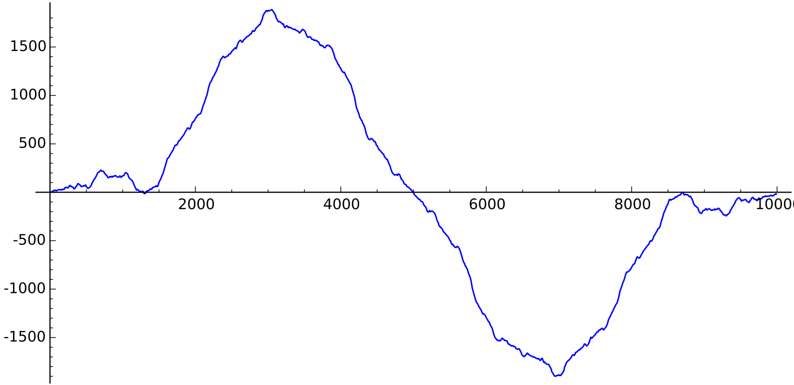
How about $M = 100$?



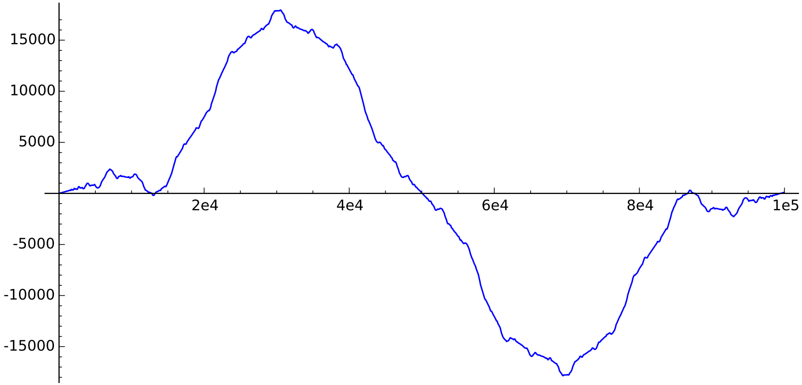
How about $M = 1000$?



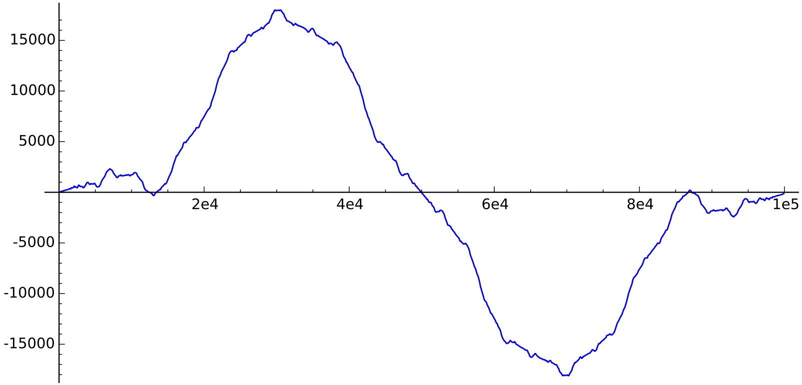
How about $M = 10000$?



How about $M = 100000$?

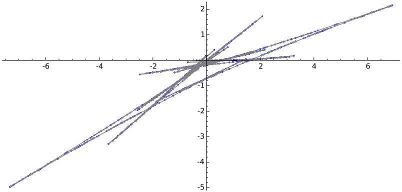


How about $M = 100003$ next prime after 100000?

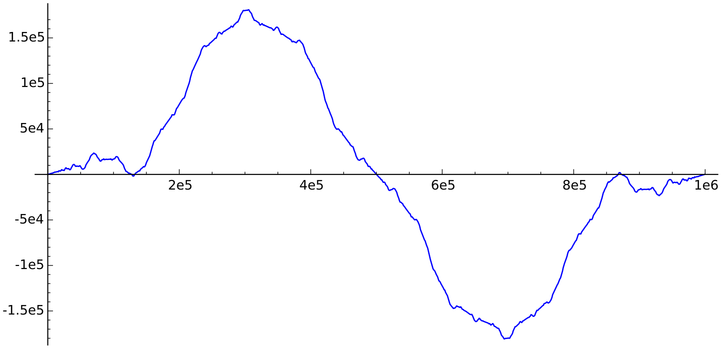


Notice Anything?

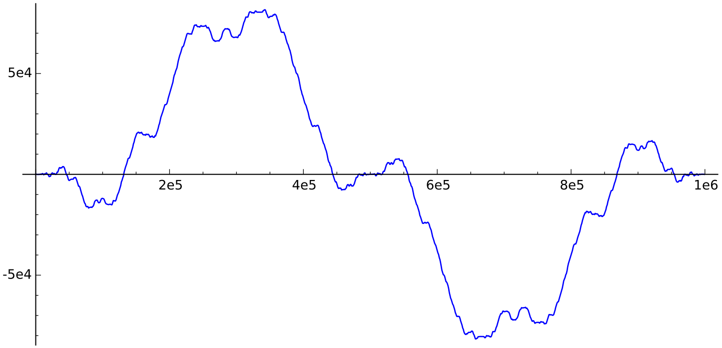
- The pictures all look almost the same, as if they are converging to some limiting function.
- There's a similar pattern (with a different picture) for each elliptic curve.
- There's a similar pattern for the -1 modular symbol $[\alpha]_E^-$.
- And a similar pattern for modular symbols attached to modular newforms with Fourier coefficients in a number field, or of higher weight (we get a multi-dimensional random walk).



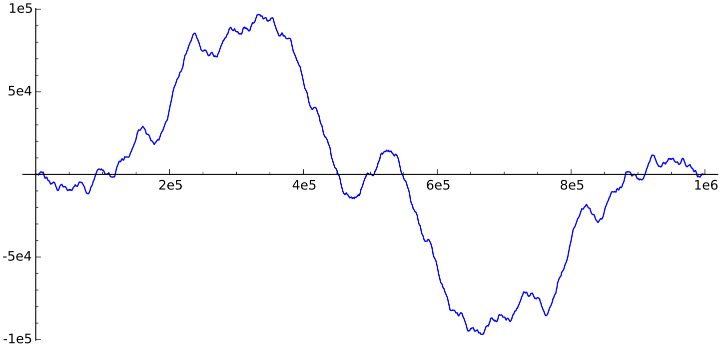
Sum for $M = 10^6$ and $E = 11a$ (rank 0)



Sum for $M = 10^6$ and $E = 37a$ (rank 1)



Sum for $M = 10^6$ and $E = 389a$ (rank 2)



Taking the limit

- *Normalize* the "not so random walk" so it is comparable for different values of M. Consider $f_M : [0,1] \to \mathbb{Q}$ given by

$$f_M(x) = \frac{1}{M} \cdot \sum_{a=1}^{Mx} \left[\frac{a}{M} \right]^+, \qquad \text{(write on board)}$$

  where, by $\sum_{a=1}^{Mx}$ we mean $\sum_{a=1}^{\lfloor Mx \rfloor}$.

- Conjecture (-)

  The limit $f(x) = \lim_{m\to\infty} f_M(x)$ *exists.*

What is the limit?

- Let ω^+ be the least real period as before. (NOTE: This need not be the Ω_E in the BSD conjecture, since when the period lattice is rectangular then $\Omega_E = 2\omega^+$.)
- Let $\sum a_n q^n$ be the newform attached to the elliptic curve E. Then:

Conjecture (-)

$$f(x) = \frac{1}{2\pi\omega^+} \cdot \sum_{n=1}^{\infty} \frac{a_n \sin(2\pi n x)}{n^2}.$$

- We expect a similar conjecture for the -1 modular symbol. What about general newforms of weight at least 2?

Rubin: connections with special values of L-functions

$$g(x) = \frac{1}{2\pi\omega^+} \cdot \sum_{n=1}^{\infty} \frac{a_n \sin(2\pi n x)}{n^2}$$

1. If we integrate $g(x)$ from 0 to $1/2$, (up to scaling) we get essentially $\sum_{n \text{ odd}} \frac{a_n}{n^3}$, which is $L(E,3)$ with the Euler factor at 2 removed, which is positive. This shows at least that $g(x)$ is usually positive.
2. If we evaluate g at $1/4$, (up to scaling) we get

$$\sum_{n=1}^{\infty} \chi(n) \frac{a_n}{n^2} = L(E,\chi,2),$$

where χ is the quadratic Dirichlet character mod 4. So $g(1/4)$ is always positive.

Mazur: “We are integrating.” (non-rigorous argument)

For $\eta > 0$ and $k \in \mathbb{Z}$, there's a complex integral that *approximates the sum of modular symbols* we're considering.
Unjustified conclusion: for each η and k,

$$\frac{1}{M}\sum_{n=1}^{k}\left[\frac{n}{M}\right]^{+} \sim \frac{1}{2\pi\omega^{+}} \cdot \sum_{n=1}^{\infty}\frac{a_n e^{-2\pi\eta}}{n^2} \cdot \sin\left(\frac{2\pi nk}{M}\right).$$

Set $k = Mx$ gives

$$f_M(x) = \frac{1}{M} \cdot \sum_{n=1}^{Mx}\left[\frac{n}{M}\right]^{+} \sim \frac{1}{2\pi\omega^{+}} \cdot \sum_{n=1}^{\infty}\frac{a_n e^{-2\pi\eta}}{n^2} \cdot \sin(2\pi nx).$$

Take the limit as $\eta \to 0$ to get our conjecture.

How? Questions

- How quickly does does $f_M(x)$ converge to the limit in practice?
- How does the following behave

$$d_M = \sqrt{\int_0^1 |f_M(x) - g(x)|^2 dx}?$$

- Are the following errors distributed normally with some mean and standard deviation?

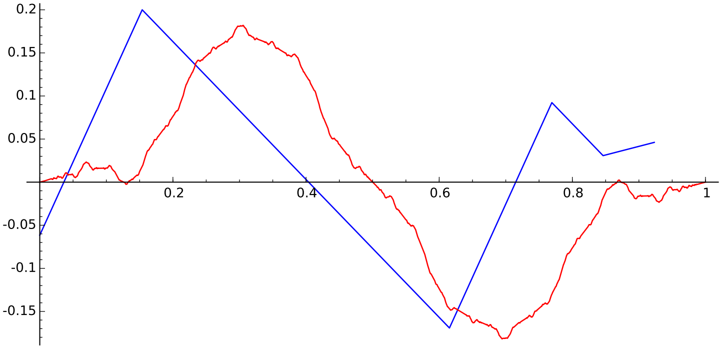
$$\{f_M(i/M) - g(i/M) : i = 0, \dots, M\}$$

Computing $g(x)$ efficiently in Sage

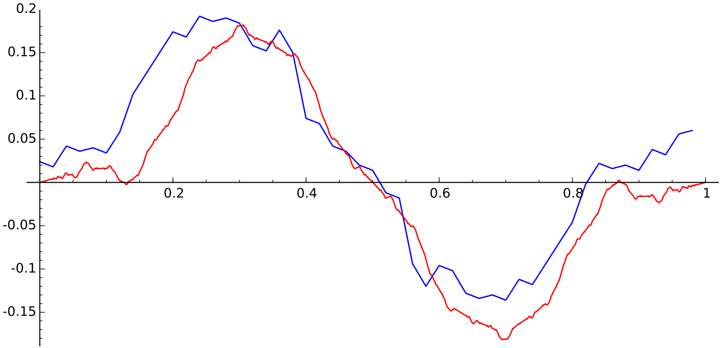
Cython is extremely useful for very efficient numerical approximation of the infinite sum $\sum_{n=1}^{\infty} \frac{a_n \sin(2\pi n x)}{n^2}$:

```
%cython

cdef extern from "math.h":
    float sin(float)

def conj(float x, list a):
    cdef float PI = 3.1415926535897932384626433833
    cdef float s = 0
    cdef long an, n = 1
    for an in a:
        s += an * sin(2*PI*n*x) / (n*n)
        n += 1
    return s
```

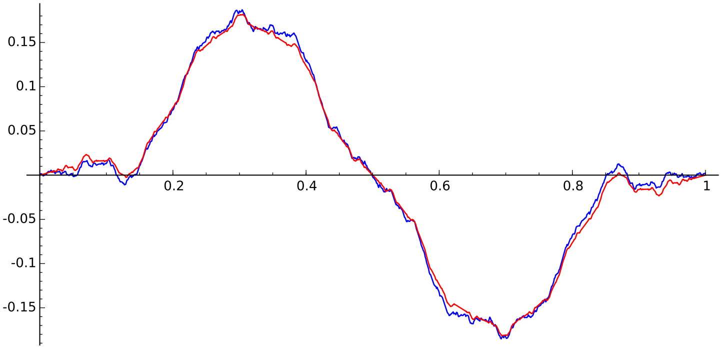
$f_{13}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



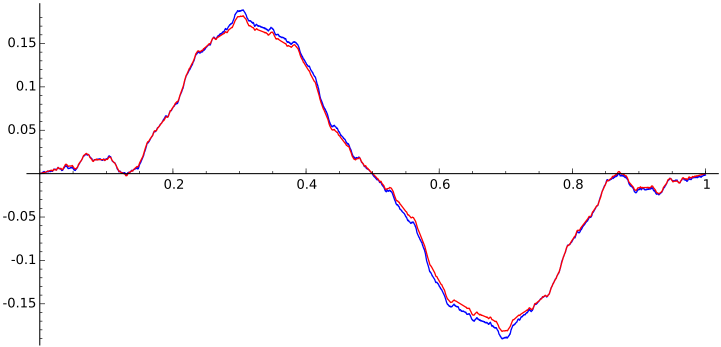
$f_{50}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



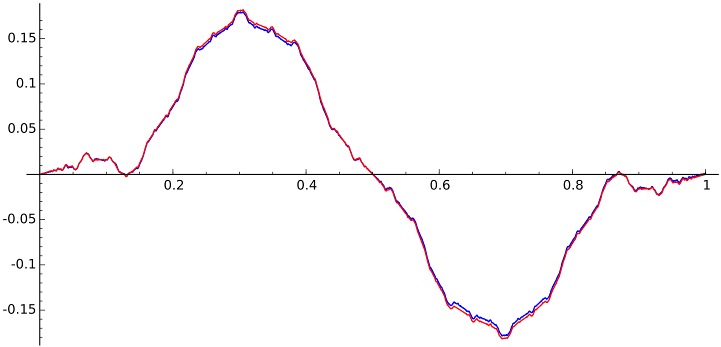
$f_{1000}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



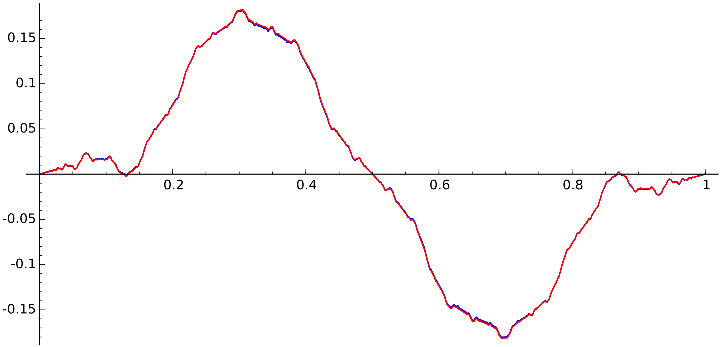
$f_{10000}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



$f_{100000}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



$f_{1000000}(x)$ versus $g(x)$ for $E = 11a$ using 10^4 terms



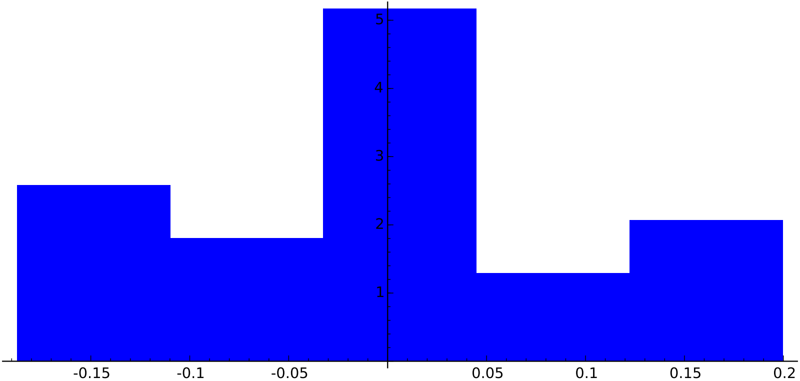
Return to our question about the distribution of errors

Are the following errors distributed normally with some mean and standard deviation?

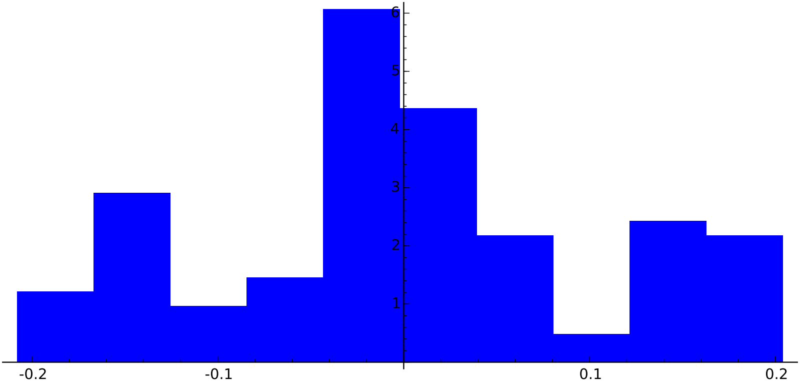
$$\{f_M(i/M) - g(i/M) : i = 0, \ldots, M\}$$

(copy to board)

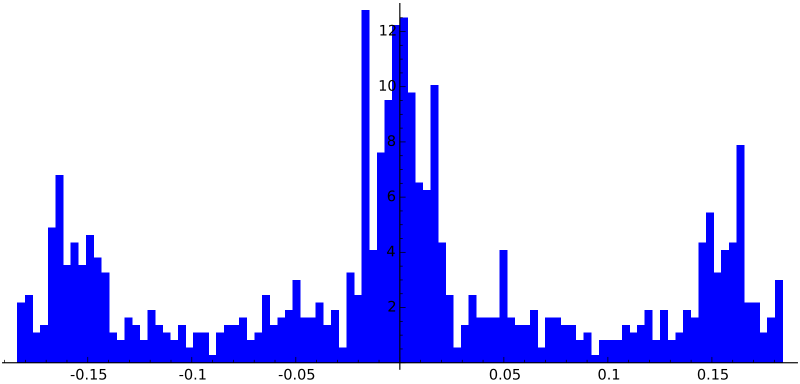
Error term distribution for $M = 50$



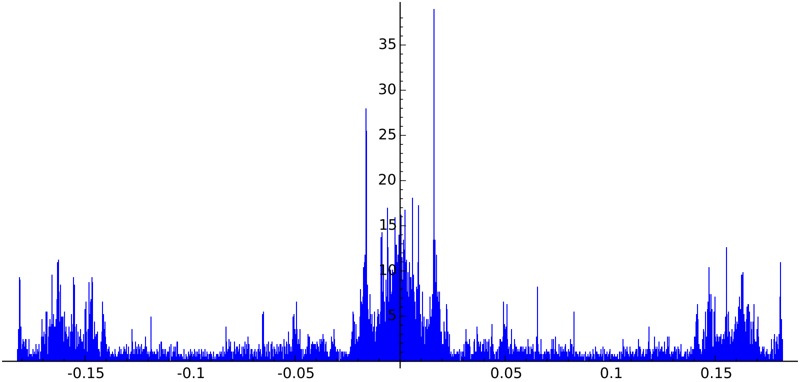
Error term distribution for $M = 100$



Error term distribution for $M = 1000$



Error term distribution for $M = 10000$



Clearly not a normal distribution.

The End