~~[illegible]~~

Arithmetic Statistics

1. Introduction

Typical questions

- What is the probability that a random integer is even?

  $\mathbb{P}(x \text{ even} \mid x \in \mathbb{Z}) = \frac{1}{2}$ (?)

- $\mathbb{P}(x \text{ squarefree} \mid x \in \mathbb{Z}) = ?$

- $\mathbb{P}(p \equiv 1 \bmod 4 \mid p \text{ prime}) = ?$

- For a fixed pol. $f \in \mathbb{Z}[X]$,

  $\mathbb{E}(\#\{x \in \mathbb{F}_p \mid f(x) = 0\} \mid p \text{ prime}) = ?$

- For a fixed ell. curve $E/\mathbb{Q}$,

  how does $\# E(\mathbb{F}_p)$ behave for random p?

- $\mathbb{P}(f \text{ irred.} \mid f \in \mathbb{Z}[X] \text{ of deg. } n) = ?$

- $\mathbb{P}(\mathrm{Gal}(f) = S_n \mid$ — " — $) = ?$

- For a fixed number field K,

  $\mathbb{P}(\mathfrak{a} \text{ principal ideal} \mid \mathfrak{a} \subseteq \mathcal{O}_K \text{ ideal}) = ?$

  $\#\{\mathfrak{a} \subseteq \mathcal{O}_K \mid \mathrm{Nm}(\mathfrak{a}) \leq T\} \sim ?$ for $T \to \infty$

- $\mathbb{P}(\mathrm{cl}(K) = 1 \mid K \text{ (random) number field of deg. } n) = ?$

  $\#\{K \text{ number field of deg. } n \mid |\mathrm{disc}(K)| \leq T\} \sim ?$ for $T \to \infty$

- $\mathbb{P}(\mathrm{Gal}(\text{Gal. cl. of } K/\mathbb{Q}) = S_n \mid K \text{ n.f. of deg. } n) = ?$

- $\mathbb{E}(\mathrm{rk}(E) \mid E \text{ ell. curve over } \mathbb{Q}) = ?$

  $\vdots$

~~1.1 Statistics~~ 1.1. Statistics

Let X be a set, $A \subseteq X$ a subset, $f: X \to \mathbb{R}$ a function.

If X is finite (e.g. $X = \mathbb{Z}/n\mathbb{Z}$):

Use e.g. the uniform prob. measure unless specified otherwise.

prob. that random $x \in X$ lies in A:

$$\mathbb{P}(x \in A \mid x \in X) = \frac{\# A}{\# X}$$

expected value of $f(x)$:

$$\mathbb{E}(f(x) \mid x \in X) = \frac{\sum_{x \in X} f(x)}{\# X}$$

(We could also assign weights $w(x) \geq 0$ and let $\mathbb{P}(x \in A \mid x \in X) = \frac{\sum_{x \in A} w(x)}{\sum_{x \in X} w(x)}$.)

If X is countable (e.g. $X = \mathbb{N}, \mathbb{Z}, \{\text{primes}\}, \{\text{n.f.}\}, \ldots$):

Intuitively, we want $\mathbb{P}(x = 1 \mid x \in \mathbb{N}) = \mathbb{P}(x = 2 \mid x \in \mathbb{N}) = \ldots = 0$.

$\Rightarrow \mathbb{P}$ can't be given by a σ-additive probability measure.

Instead, order the elements of X by a fct. $\mathrm{inv}: X \to \mathbb{R}$ such that $X_T := \{x \in X \mid \mathrm{inv}(x) \leq T\}$ is finite for every T.

$$\mathbb{P}(x \in A \mid x \in X) := \lim_{T \to \infty} \mathbb{P}(x \in A \mid x \in X_T)$$

$$\mathbb{P}^{\sup} = \limsup$$

$$\mathbb{P}^{\inf} = \liminf$$

$$\mathbb{E}(f(x) \mid x \in X) := \lim_{T \to \infty} \mathbb{E}(f(x) \mid x \in X_T)$$

(We could again use weights.)

Rmk: If $\# X = \# \mathbb{N}$ (with ~~[illegible]~~ weights 1), then removing finitely many elements from X doesn't change $\mathbb{P}, \mathbb{E}$.

Rmk a) $\mathbb{P}$ is finitely additive:

$$\mathbb{P}(x \in A_1 \sqcup \dots \sqcup A_k) = \mathbb{P}(x \in A_1) + \dots + \mathbb{P}(x \in A_k)$$

if the RHS exists

b) $\mathbb{E}$ is finitely linear

$$\mathbb{E}(\lambda_1 f_1(x) + \dots + \lambda_k f_k(x)) = \lambda_1 \mathbb{E}(f_1(x)) + \dots + \lambda_k \mathbb{E}(f_k(x)).$$

if the RHS exists

We will order $\mathbb{Z}$ or $\mathbb{N} = \{1, 2, \dots\}$ by $\mathrm{inv}(x) = |x|$.

$$\mathbb{P}(x \text{ even} \mid x \in \mathbb{N} \text{ or } \mathbb{Z}) = \lim_{T \to \infty} \mathbb{P}(x \text{ even} \mid 1 \leq x \leq T) = \lim \frac{\lfloor \frac{T}{2} \rfloor}{T} = \frac{1}{2}$$

$\mathbb{P}(x \text{ square}) = 0$

$\mathbb{P}(x \text{ prime}) = 0$ by the prime number theorem

$\mathbb{E}((-1)^x) = 0$

Rmk For any ~~$a \in \mathbb{Z}/m\mathbb{Z}$~~ $f: \mathbb{Z}/m\mathbb{Z} \to \mathbb{R}$,

~~$\mathbb{E}(f(a \bmod n) \mid x \in \mathbb{Z}) = \mathbb{E}(f(x) \mid x \in \mathbb{Z}/n\mathbb{Z})$~~

~~$\mathbb{P}(x \equiv a \bmod m \mid x \in \mathbb{Z}) = \frac{1}{m}$~~.

$$\mathbb{E}(f(x \bmod m) \mid x \in \mathbb{Z}) = \mathbb{E}(f(x) \mid x \in \mathbb{Z}/m\mathbb{Z})$$

More generally, if we order $\mathbb{Z}^n$ by any norm on $\mathbb{R}^n$,

for any ~~$a \in (\mathbb{Z}/m\mathbb{Z})^n$~~, $f: (\mathbb{Z}/m\mathbb{Z})^n \to \mathbb{R}$

~~$\mathbb{P}(x \equiv a \bmod m \mid x \in \mathbb{Z}^n) = \frac{1}{m^n}$~~.

$$\mathbb{E}(f(x \bmod m) \mid x \in \mathbb{Z}^n) = \mathbb{E}(f(x) \mid x \in (\mathbb{Z}/m\mathbb{Z})^n).$$

1.2. Squarefree integers

$$\mathbb{P}(x \text{ squarefree} \mid x \in \mathbb{N}) = ?$$

$$\left.\begin{array}{l} \mathbb{P}(4 \nmid x) = 1 - \frac{1}{4} \\ \mathbb{P}(9 \nmid x) = 1 - \frac{1}{9} \end{array}\right\} \underset{\text{CRT}}{\Rightarrow} \mathbb{P}(4, 9 \nmid x) = \left(1 - \tfrac{1}{4}\right)\left(1 - \tfrac{1}{9}\right)$$

$$\mathbb{P}(4, 9, 25 \nmid x) = \left(1 - \tfrac{1}{4}\right)\left(1 - \tfrac{1}{9}\right)\left(1 - \tfrac{1}{25}\right)$$

$$\vdots$$

$\rightsquigarrow$ Guess: ~~This is calle~~

<u>Thm 1.2.1</u> $\mathbb{P}(x \text{ squarefree}) = \prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} \approx 0.61$

<u>Rmk</u> This process, considering more and more primes is ~~The above~~ called a <u>sieve</u>.

The above argument shows "$\le$":

For any $B > 0$,

$$\mathbb{P}^{\sup}(x \text{ squarefree}) \le \mathbb{P}(p^2 \nmid x \ \forall p \le B) \underset{\text{CRT}}{=} \prod_{p \le B} \left(1 - \frac{1}{p^2}\right)$$

$$\Big\downarrow B \to \infty$$

$$\prod_p \left(1 - \frac{1}{p^2}\right).$$

More generally:

<u>Rmk 1.2.3</u> For every prime p, let $e_p \ge 0$ and $A_p \subseteq (\mathbb{Z}/p^{e_p}\mathbb{Z})^n$.

$$\Rightarrow \mathbb{P}^{\sup}(x \bmod p^{e_p} \in A_p \underbrace{\forall p)}_{\mid x \in \mathbb{Z}^n} \le \prod_p \mathbb{P}(x \in A_p \mid x \in (\mathbb{Z}/p^{e_p}\mathbb{Z})^n).$$

But "$\geq$" is tricky because ~~there is no~~ the CRT with ∞ many primes fails badly:

Exe (sieve theory nightmare: conspiracy of primes)

Let $p_1, p_2, \ldots$ be the prime numbers.

We'd expect $\mathbb{P}(x \not\equiv i \bmod p_i^2 \ \forall i) = \prod_i \mathbb{P}(x \not\equiv i \bmod p_i^2)$

$$= \prod \left(1 - \frac{1}{p_i^2}\right) \approx 0.61.$$

But actually there is no such $x \in \mathbb{N}$ because always $x \equiv x \bmod p_x^2$.

Proof 1 of "$\geq$"

$\mathbb{P}^{\inf}(x \text{ squarefree}) \geq$

$$\underbrace{\mathbb{P}(p^2 \nmid x \ \forall p \leq B)}_{\xrightarrow{B\to\infty} \prod_p \left(1-\frac{1}{p^2}\right)} - \underbrace{\mathbb{P}^{\sup}(p^2 \mid x \text{ for some } p > B)}_{\xrightarrow{B\to\infty} \text{Goal: } 0}$$

[Note: There are ∞ many $p > B$, so we can't use additivity on the RHS!]

Indeed,

$$\mathbb{P}^{\sup}(p^2 \mid x \text{ for some } p > B) = \limsup_{T\to\infty} \underbrace{\mathbb{P}(\ldots \mid 1 \leq x \leq T)}_{\leq \sum_{B<p\leq\sqrt{T}} \mathbb{P}(p^2 \mid x \mid 1 \leq x \leq T)} \leq \frac{1}{B} \xrightarrow{B\to\infty} 0$$

$$\leq \sum_{B<p\leq\sqrt{T}} \underbrace{\mathbb{P}(p^2 \mid x \mid 1 \leq x \leq T)}_{\frac{\lfloor T/p^2 \rfloor}{T} \leq \frac{1}{p^2} \text{ (careful!)}}$$

$$\leq \sum_{B<p\leq\sqrt{T}} \frac{1}{p^2} \leq \frac{1}{B}$$

$\square$

Pf 2 of Thm

Use Möbius inversion:

$$\#\{x \le T \text{ sqfree}\} = \#\{x \le T\} - \#\{x \le T : 4 \mid x\} - \#\{x \le T \mid 9 \mid x\} - \dots$$
$$+ \#\{x \le T \mid 4 \cdot 9 \mid x\} + \dots$$
$$\mp \dots$$

$$= \sum_{1 \le d \le \sqrt{T}} \mu(d) \cdot \underbrace{\#\{x : d^2 \mid x\}}_{\left\lfloor \frac{T}{d^2} \right\rfloor = \frac{T}{d^2} + O(1)}$$

$$= \dots = \underbrace{\left(\sum_{d \ge 1} \frac{\mu(d)}{d^2} \right)}_{\prod_p \left(1 - \frac{1}{p^2}\right)} \cdot T + O(T^{1/2})$$

$\square$

Pf 3 of Thm

Let $a_n := \begin{cases} 1, & n \text{ sqfree} \\ 0, & \text{otherwise} \end{cases}$

$$\sum_{n \ge 1} \frac{a_n}{n^s} = \prod_p \left(1 + \frac{1}{p^s}\right) = \prod_p \frac{1 - \frac{1}{p^{2s}}}{1 - \frac{1}{p^s}} = \frac{\zeta(s)}{\zeta(2s)}$$ has rightmost pole

at $s = 1$ with residue $\frac{1}{\zeta(2)}$.

By Wiener-Ikehara, $\underset{\#\{n \le T \text{ sqfree}\}}{\underset{\parallel}{\sum_{n \le T} a_n}} \sim \frac{1}{\zeta(2)} \cdot T$ for $T \to \infty$

$\square$

Conjecture

Let $f \in \mathbb{Z}[x]$ be a nonconstant polynomial. Then,

$$\mathbb{P}_{x\in\mathbb{Z}}(f(x) \text{ squarefree}) = \prod_p \mathbb{P}_{x\in\mathbb{Z}/p^2\mathbb{Z}}\left(p^2 \nmid f(x)\right).$$

("$\leq$" is trivial)

This is known for:

$\deg(f) \leq 2$ (similar proof)

$\deg(f) = 3$ (Hooley, 1967)

$\deg(f)$ arbitrary assuming the ABC conjecture (Granville, 1998)

~~The upper bound is~~

Notation

$f(X,\varepsilon) \ll_\varepsilon g(X,\varepsilon)$

$\Leftrightarrow f(\dots) = O_\varepsilon(\dots)$

$\Leftrightarrow \exists\, C(\varepsilon)>0 : \forall X: \ |f(X,\varepsilon)| \le C(\varepsilon)\cdot g(X,\varepsilon).$

e.g. $100T^{1/2} \ll T$ for large T

$\lfloor T \rfloor = T + O(1)$

$f \asymp g$ means: $f \ll g$ and $g \ll f$.

$f(x) \sim g(x)$ as $x \to \infty$ means: $\dfrac{f(x)}{g(x)} \xrightarrow{x\to\infty} 1$

$f(x) = o_{x\to\infty}(g(x))$ means: $\dfrac{f(x)}{g(x)} \xrightarrow{x\to\infty} 0$