Un enfoque probabilístico para la conjetura de Bateman-Horn

59º Congreso Nacional

Presenta: Lorenzo Antonio Alvarado Cabrera

Instituto de Matemáticas, UNAM | 2026

Parte I:
La conjetura de Bateman-Horn

Consideremos

$$f(n) = n^2 + 1$$

¿Puede tomar valores primos?

n	$f(n) = n^2 + 1$	Factorización
1	2	Primo
2	5	Primo
3	10	$2 \cdot 5$
4	17	Primo
5	26	$2 \cdot 13$
6	37	Primo
7	50	$2 \cdot 5^2$
8	65	$5 \cdot 13$

Surgen restricciones:

$$\text{Si } n \text{ impar } \Rightarrow \ n^2 + 1 \text{ par}$$

pero abra mas, ya que por ejemplo

$$n = 8 \text{ par } \Rightarrow \ 8^2 + 1 = 65 = \text{compuesto}$$

¿Existen infinitos primos de la forma n^2+1?

¿Existen infinitos enteros n tales que n y $n+2$ sean primos?

Se estudia de forma simultanea los polinomios:

$$f_1(n) = n \quad \text{y} \quad f_2(n) = n+2$$

Las parejas $(3,5)$, $(5,7)$, $(11,13)$ y $(17,19)$ son ejemplos de primos gemelos

No hay prueba de su infinitud

Podemos dar una formulacion general para familias de polinomios

Dada una familia de polinomios $f_1,\ f_2,\ldots,\ f_k \in \mathbb{Z}[x]$ queremos estudiar los enteros n para los cuales los valores

$$f_1(n),\ f_2(n),\ldots,\ f_k(n)$$

son todos simultaneamente primos

Dos objetivos:

- Infinitud: Demostrar que hay infinitos n con esta propiedad
- Conteo asintotico: Estimar la cantidad de estos hasta un limite x

El caso de un polinomio lineal es el primer antecedente al problema

Teorema (Dirichlet 1837 y posterior)

Sean $a, b \in \mathbb{Z}$ con $a \geq 1$ y $(a, b) = 1$. Entonces la progresion aritmetica

$$an + b$$

contiene una infinidad de primos.

Las condiciones:

- $a \geq 1$
- $(a, b) = 1$

Son para evitar obstrucciones en la primalidad

Teorema (Dirichlet 1837 Reescrito)

Sean $a, b \in \mathbb{Z}$ con $a \geq 1$ y $(a, b) = 1$. Entonces el polinomio

$$f_1(n) = an + b$$

genera una infinidad de primos.

Este es un resultado de infinitud

En 1896 llego el primer resultado de densidad

Teorema (TNP 1896)

Si $\pi(x) = \#\left\{p \leq x\colon p \text{ primo}\right\}$, entonces $\pi(x) \sim \int_2^x \frac{1}{\log(t)}\,dt := li(x)$

Teorema (TNP Reescrito)

Sea $f_1(n) = n$ y definimos $\pi(x) = \#\left\{n \leq x\colon f_1(n) \text{ es primo}\right\}$, entonces

$$\pi(x) \sim \int_2^x \frac{1}{\log(t)}\,dt$$

Y se tiene la version de densidad para Dirichlet:

Teorema (Dirichlet)

Sean $a, b \in \mathbb{Z}$ con $a \geq 1$ y $(a,b) = 1$. Consideremos $f_1(n) = an + b$ y definimos

$$\pi_{a,b}(x) = \#\left\{n \leq x\colon f(n) \text{ es primo}\right\}$$

entonces

$$\pi_{a,b}(x) \sim \frac{a}{\varphi(a)} li(x) = \prod_{p|a}\left(1 - \frac{1}{p}\right)^{-1} li(x)$$

Sin embargo atacar mas de un polinomio o uno cuadratico permanese inalcanzable

Conjetura de los primos gemelos (Hardy y Littlewood)

Si $f_1(n) = n$ y $f_2(n) = n + 2$ entonces

$$\pi_{(f_1,f_2)}(x) = \#\left\{n \leq x\colon f_1(n),\ f_2(\mathrm{n}) \text{ son primos}\right\} \sim C \cdot \int_2^x \frac{1}{(\log t)^2}\, dt$$

con $C = 2\prod_{p\geq 3}\left(1 - \frac{1}{(p-1)^2}\right) \approx 0.6601....$ *(Metodo del Circulo)*

¿Como generalizamos estas ideas?

Definición (Familia adecuada)

Sean $f_1,\ f_2,\ldots,\ f_k \in \mathbb{Z}[x]$ de grado $h_1, h_2, \ldots, h_k$ respectivamente, tales que:

⊗ Cada f_i es irreducible sobre $\mathbb{Q}$ y tiene coeficiente principal positivo.

⊗ El producto $f := f_1 \cdots f_k$ no tiene divisores primos fijos

(no existe un primo p que divida a $f(n)$ para todo n)

en tal caso decimos que la familia $\underline{f} := (f_1, \ldots, f_k)$ es adecuada

¿Cual es la probabilidad de todos los f_i sean primos simultaneamente?

$$n \text{ es primo con probabilidad } \approx \frac{1}{\log n} \Rightarrow f_i(n) \text{ es primo con probabilidad } \approx \frac{1}{\log f_i(n)}$$

si ser primo es un evento independiente "tendriamos" que

$$\approx \frac{1}{\log f_1(n) \cdots \log f_k(n)} \approx \frac{1}{h_1 \log n \cdots h_k \log n} = \frac{1}{h_1 \cdots h_k} \frac{1}{(\log n)^k}$$

$\therefore$ sumando sobre todos los posibles n hasta x

$$\pi_{\underline{f}}(x) = \#\left\{n \le x : \ f_1(n), \ldots, f_k(n) \text{ son todos primos}\right\}$$

$$\approx \sum_{n \le x} \frac{1}{h_1 \cdots h_k} \frac{1}{(\log n)^k} \approx \frac{1}{h_1 \cdots h_k} \int_2^x \frac{1}{(\log t)^k} dt$$

$$"\pi_{\underline{f}}(x) \sim \frac{1}{h_1 \cdots h_k} \int_2^x \frac{1}{(\log t)^k} dt"$$

sin embargo esto no es cierto, los eventos de ser primos no son independientes

$$f_1(n) = n, \ \ f_2(n) = n + 2$$

En 1962 *Paul T. Bateman y Roger A. Horn* analizaron las obstrucciones de forma heuristica y conjeturaron lo siguiente:

Conjetura de Bateman-Horn

Sean $f_1,\ f_2,\ldots,\ f_k \in \mathbb{Z}[x]$ de grado $h_1, h_2, \ldots, h_k$ respectivamente, familia adecuada. Si

$$\pi_{\underline{f}}(x) = \#\left\{n \leq x :\ f_1(n), \ldots, f_k(n) \text{ son todos primos}\right\}$$

entonces

$$\pi_{\underline{f}}(x) \sim \frac{C(\underline{f})}{h_1 \cdots h_k} \int_2^x \frac{1}{(\log t)^k} dt$$

donde $C(\underline{f})$ es la denominadad Constante de Bateman-Horn, definida como:

$$C(\underline{f}) := \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{N_f(p)}{p}\right)$$

con $N_f(p) := \#\left\{a \pmod p :\ f(a) \equiv 0 \pmod p\right\}$

Para $k = 2$ con $f_1(n) = n,\ f_2(n) = n + 2$

$$N_f(p) := \#\left\{a \pmod p :\ a(a+2) \equiv 0 \pmod p\right\} = \begin{cases} 1 & \text{si } p = 2 \\ 2 & \text{si } p > 2 \end{cases}$$

$$\therefore\ C(\underline{f}) := \prod_p \left(1 - \frac{1}{p}\right)^{-2} \left(1 - \frac{N_f(p)}{p}\right) = \left(1 - \frac{1}{2}\right)^{-2} \left(1 - \frac{1}{2}\right) \prod_{p \geq 3} \left(1 - \frac{1}{p}\right)^{-2} \left(1 - \frac{2}{p}\right)$$

$$= 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right)$$

$$\pi_{\underline{f}}(x) \sim C(\underline{f}) \int_2^x \frac{1}{(\log t)^2} dt$$

Teorema (Bateman-Stemmler)

Sean $f_1,\ f_2, \ldots,\ f_k \in \mathbb{Z}[x]$ de grado $h_1, h_2, \ldots, h_k$ respectivamente, familia adecuada.

Entonces para x suficientemente grande

$$\pi_{\underline{f}}(x) \leq (2^k k! + o(1)) \frac{C(\underline{f})}{h_1 \cdots h_k} \int_2^x \frac{1}{(\log t)^k} dt$$

Parte II:
La medidad zeta

Definimos una medidad de probabilidad en los enteros

Definición (Distribución zeta)

Para $s > 1$, definimos la variable aleatoria N_s tal que:

$$\mathbb{P}[N_s = n] := \frac{1}{\zeta(s)} \frac{1}{n^s}, \quad n = 1, 2, \ldots$$

- $\sum_{n=1}^{\infty} \mathbb{P}[N_s = n] := \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{n^s} = 1$

- Sus momentos positivos pueden calcularse explícitamente: $\mathbb{E}[N_s{}^q] = \frac{\zeta(s-q)}{\zeta(s)}, \quad s > q + 1$

En particular, la esperanza es finita únicamente para $s > 2$, y la varianza para $s > 3$.

- Se tiene que $\mathbb{P}[N_s \leq M] = \frac{1}{\zeta(s)} \sum_{n=1}^{M} \frac{1}{n^s} \underset{s \to 1^+}{\to} 0$, es decir, $N_s \overset{\mathbb{P}}{\to} +\infty$

¿Existe alguna normalización de N_s que produzca una distribución límite no trivial?

Sea $Y_s = (s-1)\log(N_s)$, entonces se tiene

$$Y_s \xrightarrow[s\to 1^+]{D} \mathrm{Exp}(1)$$

Cranston y Peltzer trabajaron en una versión del célebre teorema de Erdős–Kac en este contexto probabilístico. Si $\omega(n)$ denota el número de factores primos distintos de n y

$$P(s) := \sum_p \frac{1}{p^s}$$

entonces

$$\frac{\omega(N_s) - P(s)}{\sqrt{P(s)}} \xrightarrow[s\to 1^+]{D} N(0,1)$$

Una propiedad util de esta medida es que:

$$\mathbb{P}[d \mid N_s] = \sum_{m=1}^{\infty} \mathbb{P}[N_s = md] = \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{(md)^s} = \frac{1}{d^s}$$

de modo que si $p_1, .., p_r$ son primos distintos:

$$\mathbb{P}[p_1 \mid N_s, ..., p_r \mid N_s] = \mathbb{P}[p_1 \cdots p_r \mid N_s] = \frac{1}{(p_1 \cdots p_r)^s} = \mathbb{P}[p_1 \mid N_s] \cdots \mathbb{P}[p_r \mid N_s]$$

Mas aun, si $v_p(n)$ denota la valuacion $p-$adica de $n,$ entonces:

$$\boxed{v_p(N_s) \sim Geo(p^{-s})}$$

siendo independientes entre cada primo p

Para estudiar nuestros polinomios necesitaremos además una propiedad sobre las clases residuales. Para cualquier módulo fijo d, tenemos el límite:

$$\mathbb{P}[N_s \equiv r \pmod d] = \frac{1}{\zeta(s)} \sum_{n \equiv r \pmod d} \frac{1}{n^s} = \frac{1}{\zeta(s)} \sum_{m=0}^{\infty} \frac{1}{(md+r)^s} \underset{s \to 1^+}{\sim} (s-1)\frac{1}{d(s-1)} = \frac{1}{d}$$

$$\therefore \quad \lim_{s \to 1^+} \mathbb{P}[N_s \equiv r \pmod d] = \frac{1}{d}$$

por lo tanto, se sigue inmediatamente que

$$\lim_{s \to 1^+} \mathbb{P}[d \mid f(N_s)] = \frac{N_f(d)}{d}$$

Una vez definida la distribución zeta, necesitamos escoger una variable aleatoria que detecte la primalidad de los valores de nuestros polinomios.

Funcion de Von Mangold:

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^k \\ 0 & \text{e.o.c} \end{cases}$$

De esta manera, para la familia $\underline{f} = (f_1, \ldots, f_k)$ el producto

$$\prod_{i=1}^{k} \Lambda(f_i(n))$$

es no nulo solo si todos los polinomios son primos (o potencias de primos)

Esta idea generaliza a la funcion ψ de Chebyshev $\quad \psi(x) = \sum_{n \leq x} \Lambda(n)$

Definición

Sean $f_1,\ f_2, \ldots,\ f_k \in \mathbb{Z}[x]$ de grado $h_1, h_2, \ldots, h_k$ respectivamente, familia adecuada.
Definimos

$$\psi_{\underline{f}}(x) = \sum_{n \leq x} \Lambda(f_i(n)) \cdots \Lambda(f_i(n))$$

Teorema (Bateman-Stemmler)

Bateman-Horn es cierta si, y solo si, $\psi_{\underline{f}}(x) \sim C(\underline{f})x$

Acotar potencias primas mayores y sumacion de Abel

Nuestra propuesta consiste en evaluar estos coeficientes en el índice aleatorio N_s

Es decir, definimos la esperanza

$$\Omega_{\underline{f}}(s) := \mathbb{E}\left[\prod_{i=i}^{k} \Lambda(f_i(N_s))\right]$$

¿Es finita? ¿Como se comporta cuando $s \to 1^+$?

$$\Omega_{\underline{f}}(s) = \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{k} \Lambda(f_i(n))}{n^s} \ll \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{k} \log(f_i(n))}{n^s} \ll \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{\log^k(n)}{n^s} < \infty$$

$\therefore$ la esperanza $\Omega_{\underline{f}}(s)$ es finita para todo $s > 1$

Proposición

$$\text{Si } \psi_{\underline{f}}(x) \sim C(\underline{f})x \;\Rightarrow\; \lim_{s \to 1^+} \Omega_{\underline{f}}(s) = C(\underline{f})$$

Esto muestra que nuestra esperanza conserva la informacion artimetica del problema

¿Qué información podemos recuperar si logramos demostrar directamente ese límite?

$$\lim_{s\to 1^+} \Omega_{\underline{f}}(s) = C(\underline{f})$$

Teorema

Se tiene que

$$\lim_{s\to 1^+} \Omega_{\underline{f}}(s) = C(\underline{f}) \quad \Leftrightarrow \quad \sum_{n\leq x} \frac{\prod_{i=1}^{k} \Lambda(f_i(n))}{n} \sim C(\underline{f}) \log x$$

Dem.- *Stieltjes* y *Teorema Tauberiano de* Karamata

Si bien es un resultado mas "debil", nos preserva la informacion de infinitud

Corolario

Si $\lim_{s\to 1^+} \Omega_{\underline{f}}(s) = C(\underline{f})$, entonces, existen infinitos índices n en los que todos los polinomios $f_1(n),\ldots,f_k(n)$ toman valores primos simultaneamente

Parte III:
Acercamientos a la prueba

Identidad de Golomb

Sea $k \geq 1$ y enteros $a_1, .., a_k > 1$ coprimos dos a dos. Entonces se tiene:

$$\prod_{i=1}^{k} a_i = \frac{(-1)^k}{k!} \sum_{d | a_1 \cdots a_k} \mu(d) \log^k(d)$$

$\prod_{i=i}^{k} \Lambda(f_i(N_s))$ para poder aplicarlo necesitaremos nuevas condiciones para nuestros polinomios

Definición (Familia admisible)

Sean $f_1,\ f_2, \ldots,\ f_k \in \mathbb{Z}[x]$ de grado $h_1, h_2, \ldots, h_k$ respectivamente, tales que:

- ⊗ Cada f_i es irreducible sobre $\mathbb{Q}$ y tiene coeficiente principal positivo.
- ⊗ El producto $f := f_1 \cdots f_k$ no tiene divisores primos fijos
- ⊙ El $f_i(n) > 1$ para todo $i = 1, \ldots, k$ y $n \geq 1$
- ⊙ Para cualquier par $i \neq j$, y todo entero $n \geq 1$, $(f_i(n), f_j(n)) = 1$

$\{n, n+2\}$ es adecuada $\leftrightarrow$ $\{4n-1, 4n+1\}$ es admisible

Proposición (Marc Hindry y Tanguy Rivoal)

Si la conjetura de Bateman-Horn es cierta para familias admisibles, enotnces lo es para familias adecuadas

Sea $\underline{f} = (f_1, \ldots, f_k)$ admisible, entonces

$$\Omega_{\underline{f}}(s) := \mathbb{E}\left[\prod_{i=i}^{k} \Lambda(f_i(N_s))\right] = \mathbb{E}\left[\frac{(-1)^k}{k!} \sum_{d \mid f(N_s)} \mu(d) \log^k(d)\right]$$

$$= \frac{(-1)^k}{k!} \mathbb{E}\left[\sum_{d \geq 1} \mu(d) \log^k(d) 1_{\{d \mid f(N_s)\}}\right]$$

$$\underset{(*)}{=} \frac{(-1)^k}{k!} \sum_{d \geq 1} \mu(d) \log^k(d) \mathbb{E}\left[1_{\{d \mid f(N_s)\}}\right]$$

$$\therefore \quad \Omega_{\underline{f}}(s) = \frac{(-1)^k}{k!} \sum_{d=1}^{\infty} \mu(d) \log^k(d) \mathbb{P}\left[d \mid f(N_s)\right]$$

Observación. –

$$\lim_{s\to 1^+} \Omega_{\underline{f}}(s) = \lim_{s\to 1^+} \frac{(-1)^k}{k!} \sum_{d=1}^{\infty} \mu(d)\log^k(d)\mathbb{P}\left[d \mid f(N_s)\right]$$

$$" = " \frac{(-1)^k}{k!} \sum_{d=1}^{\infty} \mu(d)\log^k(d) \lim_{s\to 1^+} \mathbb{P}\left[d \mid f(N_s)\right]$$

$$= \frac{(-1)^k}{k!} \sum_{d=1}^{\infty} \frac{\mu(d)\log^k(d) N_f(d)}{d} \underset{(\Delta)}{=} C(\underline{f})$$

(Δ) = *Hinry* y *Rivoal*

¿Que hacemos entonces?