1. History

Let $\Delta_m = \liminf_{n\to\infty} \frac{p_{n+m} - p_n}{\log n}$, $\Delta = \Delta_1$

Possible approximations of the twin prime conj.

(i) Let $d_n = p_{n+1} - p_n$ $\Delta := \liminf_{n\to\infty} \frac{d_n}{\log n} < 1$

since by the PNT: $\pi(x) \sim \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}$

$$\frac{1}{N}\sum_{n=1}^{N} \frac{d_n}{\log n} \longrightarrow 1 \quad \text{as } N \to \infty$$

(ii) Small Gaps Conjecture: $\Delta = 0$

Hardy-Littlewood (1926) GRH $\longrightarrow \Delta \leq 2/3$

Erdős (1940) $\Delta < 1 \Leftrightarrow \Delta \leq 1 - c_0, c_0 > 0$

⋮

Bombieri-Davenport (1966) $\Delta < 0.4665 < 1/2$

⋮

H. Maier (1988) $\Delta < 1/4$

⋮

Goldston – J.P. – Yildirim (2005–2008) $\Delta = 0$
(GPY)

GPY 2005–2009 $\liminf_{n\to\infty} \frac{d_n}{\sqrt{\log n}\,(\log\log n)} < \infty$

J. P. 2011–2013 $d_n \ll_\varepsilon (\log n)^{3/7+\varepsilon}$ inf. often (i.o.)

Remark $\ll_{A,\varepsilon}$ means $\leq C(A,\varepsilon)$

BOUNDED GAPS CONJECTURE: $\exists C \quad d_n \leq C$ i.o.

Y Zhang (2013-2014) True with $C = 7 \cdot 10^7$

Polymath 8A (T. Tao) " $C = 4680$

J. Maynard (2013-?) " $C = 600$

Polymath 8B (T. Tao) " $C = 246$

Theorem (Maynard/Tao) (2013-?) $\liminf_{n\to\infty} (p_{n+m} - p_n) \ll e^{4m}$

2. Conditional results (notation: p, $\mathcal{P}$)

<u>Def</u> primes have an admissible level ϑ of distribution if $\forall A, \varepsilon > 0$

$$(*)\quad \sum_{q \le X^{\vartheta-\varepsilon}} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \ (\mathrm{mod}\ q) \\ p \le X}} \log p - \frac{X}{\varphi(q)} \right| \ll_{A,\varepsilon} \frac{X}{\log^A X}.$$

<u>Theorem</u> GPY (2005-2009) If $\exists \vartheta > 1/2$ admiss. level then the Bounded Gap Conjecture is true $C = C(\vartheta)$

<u>Theorem</u> (Bombieri – Vinogradov 1965/66) $\vartheta = 1/2$ is adm.

<u>Theorem</u> (Y. Motohashi – J.P. 2008 + arXiv 2006):

"A smoothed GPY sieve".

If $(*)$ is true for (a $\vartheta > 1/2$ for) smooth moduli q, i.e. if $\exists b > 0$ $p \mid q \Rightarrow p \le q^b$ (or even X^b is sufficient) and one can replace $\max\limits_{\substack{a \\ (a,q)=1}}$ by $\max\limits_{\substack{a \\ (a,q)=1\ \mathcal{P}(a) \equiv 0 (q)}}$ then the Bounded Gaps Conjecture is true $C = C(\vartheta, b)$

3.) Heuristics behind GPY

Def. $\mathcal{H} = \mathcal{H}_k = \{h_i\}_{i=1}^{k}$ $0 \le h_1 < h_2 < \ldots < h_k$ is admissible if $\mathcal{H}$ does not cover all residue classes mod p for any prime p

$L_i(n) = a_i n + b_i$ is an admissible system $(1 \le i \le k)$ if $\prod_{i=1}^{k} L_i(n)$ has no fixed prime divisor

Dickson C. (1904) If $\{L_i(n)\}_{i=1}^{k}$ is admissible then $\{L_i(n)\}_{i=1}^{k} \in \mathcal{P}^k$ i.o. (for infinitely many n)

Hardy - Littlewood (1923) C. If $\mathcal{H}_k$ is admissible then

$$\sum_{\substack{n \le X \\ \{n+h_i\}_{i=1}^{k} \in \mathcal{P}^k}} 1 \sim \mathfrak{S}(\mathcal{H}) \frac{X}{\log^k X} \qquad \mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} > 0$$

$\nu_p(\mathcal{H}) = \#$ res. classes covered by $\mathcal{H}$ mod p

DHL(k) Conjecture: $\mathcal{H}_k$ adm. $\Longrightarrow$ $\{n+h_i\}_{i=1}^{k} \in \mathcal{P}^k$ i.o.

(This is the twin prime conj. if $k=2$ $\mathcal{H}_2 = \{0,2\}$ $(j, j+2)$)

DHL(k, 2) Conjecture $\mathcal{H}_k$ adm. $\Longrightarrow$ $\{n+h_i\}_{i=1}^{k}$ contains at least two primes i.o.

DHL(k, 2) is true for any k $\Longrightarrow$ Bounded Gap C. is true

<u>Def</u> Elliott - Halberstam conj. (EH, 1966)

$\vartheta = 1$ is an admissible level

<u>Theorem</u> GPY (2005-2009) $EH \Rightarrow DHL(6,2) \Rightarrow$

$d_n = p_{n+1} - p_n \leq 16$ i.o.

Basic idea of GPY: Let us attack $DHL(k,2)$ for large k by trying to find weights

$$a_n = a_n(\mathcal{H}) \geq 0 \quad \text{satisfying for any } N > N_0$$

$$\frac{1}{N} \sum_{n=N+1}^{2N} a_n \sum_{i=1}^{k} \chi_{\mathcal{P}}(n+h_i) > \frac{1}{N} \sum_{n=N+1}^{2N} a_n \qquad \chi_{\mathcal{P}}(m) = \begin{cases} 1 & \text{if } m \in \mathcal{P} \\ 0 & \text{if } m \notin \mathcal{P} \end{cases}$$

(i) $a_n = 1$ $\quad k/\log N < 1$. Let $P_{\mathcal{H}}(n) = \prod_{i=1}^{k} (n+h_i)$

(ii) $a_n = \begin{cases} 1 & n+h_1 \in \mathcal{P} \\ 0 & \text{o.w.} \end{cases}$ $\longrightarrow$ we get equality

(iii) $a_n = \begin{cases} 1 & \{n+h_i\}_{i=1}^{k} \in \mathcal{P}^k \\ 0 & \text{o.w.} \end{cases}$ $\quad LHS = k \times RHS$ tautology

(iv) $a_n = \sum_{d \mid P_{\mathcal{H}}(n)} \mu(d) \log^k \frac{P_{\mathcal{H}}(n)}{d} =: \Lambda_k(P_{\mathcal{H}}(n))$ equivalent with (iii)

(v) $a_n := \sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \le R}} \mu(d) \log^k R/d$ truncated sum

$a_n < 0$ is possible

(vi) $a_n := \left(\sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \le R}} \mu(d) \log^k R/d \right)^2 \ge 0$ (this is

actually Selberg's k-dimensional sieve — an approximation to $\Lambda_k(P_{\mathcal{H}}(n))$, the approximative characteristic function of $\{n+h_i\}_{i=1}^k \in \mathcal{P}^k$.

First chanceful choice

$$\frac{LHS}{RHS} = \vartheta - o(1) \quad \text{with } R = N^{\vartheta/2 - o(1)}$$

Unfortunately this is < 1 even for $\vartheta = 1$, although Heath-Brown showed in the 80's EH $\to \Delta = 0$

(vii) Successfull choice for SMALL GAPS CONJ (GPY)

Promising but unsuccessfull unconditional choice for DHL$(k,2)$ for large k, i.e. Bounded Gaps Conj.

$$a_n = \left(\sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \le R}} \mu(d) \log^{k+\ell} R/d \right)^2 \quad \text{if } \ell \to \infty, \ \frac{k}{\ell} \to \infty$$

$$\frac{LHS}{RHS} = 2\vartheta \left(1 - \underbrace{O\left(\frac{1}{\ell} + \frac{\ell}{k}\right)}_{O\left(\frac{1}{\sqrt{k}}\right) \text{ if } \ell \asymp \sqrt{k}}\right) > 1 \text{ if } \vartheta > 1/2$$

This implies:

<u>Theorem</u> (GPY): $\vartheta > \frac{1}{2} \longrightarrow DHL(k,2)$ is true for $k > C_0(\vartheta)$

<u>Theorem</u> Replacing $\log^{k+\ell} R/d$ by any piecewise continuous function $F(d)$ we obtain ...

$$\frac{LHS}{RHS} < 2\vartheta \quad \left(\leq 1 \text{ with } \vartheta = \tfrac{1}{2}\right) \quad \text{Soundararajan for } F \in C^{\infty}[0,1]$$

$$\frac{LHS}{RHS} < 2\vartheta\left(1 - \frac{c}{k^{2/3}}\right) \quad \text{Conrey–Farkas–Rivesz–Pink}$$

General belief: it is impossible to show $DHL(k,2)$ for large k, therefore Bounded Gaps Conjecture without showing $\vartheta > \frac{1}{2}$, or more precisely (MP) without the weaker condition $\vartheta_{smooth} > \frac{1}{2}$

<u>Theorem</u> Zhang (2013): $(*)$ is true with $\vartheta = \frac{1}{2} + \frac{1}{584}$ $b = \frac{1}{292}$

Remark: The optimal weights would be that of (iii)

$$a_n = \begin{cases} 1 & \{n+h_i\} \in \mathcal{P}^k \\ 0 & \text{o.w.} \end{cases}$$ tautology $\Leftrightarrow$ HL conj.

So it is "hopeless" to find them.

3. Sketch of evaluation of $\frac{1}{N}\sum_{n=N+1}^{2N} a_n$ from (vii)

More generally, let $h_R(x) = \frac{\left(\log R/x\right)^{k+\ell}}{(k+\ell)!}$

Let $f(p) = \gamma_p(\mathcal{H}_k) = k + O\left(\frac{1}{p}\right)$ $\quad \mathcal{H}_k = \{h_i\}_{i=1}^{k}$

f compl. multiplicative

$$S := \frac{1}{N} \sum_{n\sim N} \left(\sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \le R}} \mu(d) h_R(d) \right)^2$$

$n \sim N$: $n \in (N, 2N]$

$P_{\mathcal{H}}(n) = \prod_{i=1}^{k} (n + h_i)$

$\sum{}'$ = sq. free

$$S = \sum_{\substack{d \le R \\ e \le R}} h_R(d) h_R(e) \mu(d) \mu(e) \cdot \frac{1}{N} \sum_{\substack{n \sim N \\ d \mid P_{\mathcal{H}}(n) \\ e \mid P_{\mathcal{H}}(n)}} 1$$

$$= \sum_{\substack{d \le R \\ e \le R}} \frac{\mu(d)\mu(e) f([d,e])}{[d,e]} h_R(d) h_R(e) + O\left(R^2 \log^{4k} R\right)$$

Let $(d,e) = u$ $\quad d = um$, $e = un$, $(m,n) = 1$

$$S \sim \sum_{u \le R} \sum_{\substack{m,n \le R/u \\ (m,n)=1}}{}' \frac{\mu(um)\mu(un)}{umn} f(u) f(m) f(n) h_R(um) h_R(un)$$

But $\sum_{\beta|m,\ \beta|n} \mu(\beta) = \begin{cases} 1 & \text{if } (m,n)=1 \\ 0 & \text{o.w.} \end{cases}$. Let $m=\beta m'$, $n=\beta n'$

$$S \sim \sum_{u\le R}{}' \sum_{\beta \le R/u}{}' \mu(\beta)\frac{f(u)f(\beta^2)}{u\beta^2}\left(\sum_{m'\le R}\frac{\mu(u\beta m')}{m'}f(m')h_R(u\beta m')\right)^2$$

Let us group terms according to the value $u\beta =: \gamma$ $(m' \to m)$

$$S \sim \sum_{\gamma\le R}{}' \frac{f(\gamma)}{\gamma}\left(\sum_{\beta|\gamma}\frac{\mu(\beta)f(\beta)}{\beta}\right)\left(\sum_{\substack{m\le R/\gamma \\ (m,\gamma)=1}}\frac{\mu(\gamma)\mu(m)}{m}f(m)h_R(\gamma m)\right)^2$$

Let us denote the inner sum by $J = J(\gamma, R/\gamma)$, let $c>0$

$$F(s,\gamma) := \sum_{(m,\gamma)=1}{}' \frac{\mu(m)f(m)}{m^{s+1}} = \prod_{p\nmid\gamma}\left(1-\frac{f(p)}{p^{s+1}}\right) \approx \prod_{p\nmid\gamma}\left(1-\frac{1}{p^{s+1}}\right)^k$$

$$= \zeta(s+1)^{-k}\prod_{p|\gamma}\left(1-\frac{f(p)}{p^{s+1}}\right)^{-1}$$

Identity: $\frac{1}{2\pi i}\int_{(c)}\frac{x^s\,ds}{s^{j+1}} = \begin{cases}\frac{(\log x)^j}{j!} & \text{if } x\ge 1\\ 0 & \text{if } x\le 1\end{cases}$ $\Big|$ $(c) = \{s;\ \mathrm{Re}\, s = c\}$

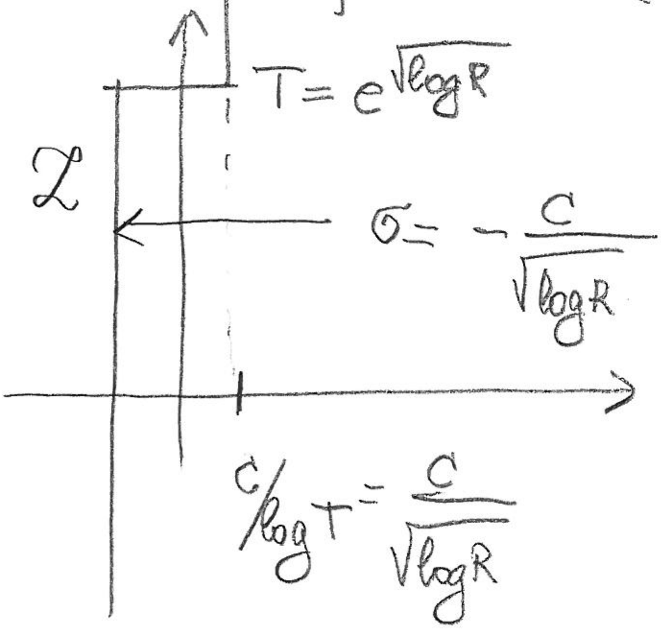
$$J = J(\gamma, R/\gamma) = \frac{1}{2\pi i}\int_{(c)} F(s,\gamma)\left(\frac{R}{\gamma}\right)^s \cdot \frac{ds}{s^{k+l+1}}$$

$$F(s) := \zeta(s+1)^{-k}G(s) \Longleftrightarrow G(s) = G(s,\gamma) = F(s)\zeta(s+1)^k$$

$$G(s) = \prod_p\left(1-\frac{f(p)}{p^{s+1}}\right)\left(1-\frac{1}{p^{s+1}}\right)^{-k}\prod_{p|\gamma}\left(1-\frac{f(p)}{p^{s+1}}\right)^{-1} \ll e^{k(\log\gamma)^{-\sigma}}$$

<u>Prop.</u> $G(s)$ is regular for $\sigma > -\frac{1}{3}$ since $f(p) = k + O\left(\frac{1}{p}\right)$

Let us transform $(c) = (c - i\infty, c + i\infty)$ as follows



On the new path $\mathcal{L}'$ we have

(i) $\zeta^{-1}(s+1) = O\left(\log(|t|+2)\right)$

Simplification: in case of

$R/\gamma > e^{(\log R)^{2/3}}$

we have

$$J \sim \operatorname*{Res}_{s=0} \sim G(0,\gamma) \cdot \operatorname*{Res}_{s=0} \left\{ \left(\frac{R}{\gamma}\right)^s \cdot \frac{1}{s^{\ell+1}} \right\} = \frac{G(0,\gamma)\left(\log R/\gamma\right)^{\ell}}{\ell!}$$

$$G(0,\gamma) = \prod_p \left(1 - \frac{f(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \prod_{p \mid \gamma} \left(1 - \frac{f(p)}{p}\right)^{-1} = \mathfrak{S}(\mathcal{H}) \prod_{p \mid \gamma} \left(1 - \frac{f(p)}{p}\right)^{-1}$$

$$S \sim \sum_{\gamma \le R}{}' \frac{f(\gamma)}{\gamma} \prod_{p \mid \gamma} \left(1 - \frac{f(p)}{p}\right)^{-1} \mathfrak{S}^2(\mathcal{H}) \left(\frac{\log R/\gamma}{(\ell!)^2}\right)^{2\ell}$$

<u>Lemma</u>: $$\sum_{\substack{\gamma \le R \\ (\gamma, r) = 1}}{}' \prod_{p \mid \gamma} \frac{f(p)}{p - f(p)} \sim c_r \frac{(\log x)^k}{k!} = c_r \int_1^x \frac{(\log v)^{k-1}}{v (k-1)!} dv$$

where $$c_r = \underbrace{\prod_p \left(1 - \frac{f(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{k}}_{\mathfrak{S}^{-1}(\mathcal{H})} \cdot \prod_{p \mid r} \left(1 - \frac{f(p)}{p}\right)$$

$$S \sim \frac{(2\ell)!}{(\ell!)^2}\mathfrak{S}(\mathcal{H}) \int_1^{R^{1-o(1)}} \frac{(\log v)^{k-1}}{v(k-1)!}\frac{(\log R/v)^{2\ell}}{(2\ell)!}\,dv \sim \binom{2\ell}{\ell}\mathfrak{S}(\mathcal{H})\, B \cdot (\log R)^{k+2\ell}$$

where $v = R^y$, $\frac{dv}{dy} = \log R \cdot v$ and

$$B = \int_0^1 \frac{y^{k-1}(1-y)^{2\ell}}{(k-1)!\,(2\ell)!}\,dy = \frac{1}{(k+2\ell)!} \quad \left(\begin{array}{c}\text{Euler's}\\ \text{Beta integral}\end{array}\right)$$

This yields the evaluation of the sums of the weights, i.e. $S = \frac{1}{N}\sum_{n \sim N} a_n$.

If the primes are "statistically" uniformly distributed for $q \le R := N^{\vartheta/2-\varepsilon}$, then a very similar argument gives the evaluation of

$$S^*(h_i) = \frac{1}{N}\sum_{n \sim N} a_n \chi_p(n+h_i)\log n \text{ for } h_i \in \mathcal{H}$$

since in this case we have "statistically"

$$\sum_{\substack{n+h_i \equiv a \,(\mathrm{mod}\, q)\\ n \sim N}} \chi_p(n+h_i)\log n \sim \begin{cases} \frac{1}{\varphi(q)}\sum_{n\sim N} 1 & \text{if } (a,q)=1 \\ O(1) & \text{if } (a,q)\neq 1\end{cases}$$

Let us forget for a minute the error term in the Bombieri – Vinogradov theorem.

Then we have the following changes at the beginning of our evaluation of $S^*(h_i)$ [compared with S]

(i) we have $\nu_p(\mathcal{H}) - 1 = k - 1 + O\left(\frac{1}{p}\right)$ residue classes mod p for n, satisfying $n + h_j \equiv 0 \pmod p$, namely $n \equiv -h_1, \ldots, -h_{i-1}, -h_{i+1}, \ldots h_k$ since in case of $n + h_i \equiv 0\ (p)$, $p \mid q \leq R \leq N^{\vartheta/2 - \varepsilon}$ $n + h_i \notin \mathcal{P} \iff \chi_{\mathcal{P}}(n + h_i) = 0$ [compared with $\nu_p(\mathcal{H}) = k + O\left(\frac{1}{p}\right)$]

(ii) in theses residue classes we have an accumulation of primes with a factor $\dfrac{q}{\varphi(q)} = \prod_{p \mid q} \dfrac{p}{\varphi(p)} = \prod_{p \mid q} \dfrac{p}{p-1}$ [compared with 1]

This corresponds to the replacement of $f(p) = \nu_{\mathcal{H}}(p) = k + O\left(\frac{1}{p}\right)$

by

(iii) $f^*(p) = (\nu_{\mathcal{H}}(p) - 1)\dfrac{p}{p-1} = \left(k - 1 + O\left(\frac{1}{p}\right)\right)\dfrac{p}{p-1}$

Therefore k has to be replaced by $k-1$, consequently l has to be replaced by $l+1$ (since $k+l$ remains unchanged),

$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \dfrac{\nu_p(\mathcal{H})}{p}\right)\left(1 - \dfrac{1}{p}\right)^{-k}$ has to be replaced by (iii) by

$$\mathfrak{S}^*(\mathcal{H}) = \prod_p \left(1 - \frac{\nu_p(\mathcal{H}) - 1}{p} \cdot \frac{p}{(p-1)}\right)\left(1 - \frac{1}{p}\right)^{-(k-1)} = \prod_p \frac{p - \nu_p(\mathcal{H})}{p}\left(1 - \frac{1}{p}\right)^{-k}$$

thus $\mathfrak{S}(\mathcal{H}) = \mathfrak{S}^*(\mathcal{H}) \iff$ no change is necessary

The effect of the error terms in the BV theorem:

$$E(N,q) := \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \,(\mathrm{mod}\, q) \\ p \sim N}} \log p - \frac{N}{\varphi(q)} \right| \text{ then by BV}$$

$$\sum_{q \le N^{\vartheta-\varepsilon} = R} E(N,q) \ll_{A,\varepsilon} \frac{N}{\log^A N}$$

If $\tau_j(n) = \sum_{d_1 \cdots d_j = n} 1 = j^{\omega(n)}$ (generalized divisor function)

then the error in the evaluation of $S^*(h_i)$ is at most

$$\frac{1}{N} \sum_{q \le R}' E(N,q) \sum_{\substack{d,e \\ q=[d,e]}} \tau_k(q) \ll \frac{1}{N} \sum_{q \le R} E(N,q) \tau_{3k}(q)$$

$$\ll \frac{1}{N} \left\{ \sum_{q \le R} E(N,q) \sum_{q \le R} E(N,q) \tau_{3k}^2(q) \right\}^{\frac{1}{2}} \ll$$

$$\ll \frac{1}{N} \cdot \frac{N}{\log^A N} \sum_{q \le N} \frac{N \log N \, \tau_{9k^2}(q)}{q} \ll \frac{\log^{C(k)} N}{\log^A N} \ll \frac{1}{(\log N)^{A/2}}$$

which is negligible. This implies for $h_i \in \mathcal{H}$

$$S^*(h_i) \sim \binom{2\ell+2}{\ell+1} \mathfrak{S}(\mathcal{H}) \frac{1}{(k-1+2\ell+2)!} (\log R)^{k-1+2\ell+2}$$

$$\sum_{i=1}^{k} S^*(h_i) \sim 4\left(1 - \frac{1}{2\ell+1}\right)\left(1 - \frac{2\ell+1}{k+2\ell+1}\right) \frac{\log R}{\log N} \cdot S = 2\vartheta S \left(1 - O\left(\frac{1}{\sqrt{k}}\right)\right)$$

where $\frac{\log R}{\log N} = \frac{\vartheta}{2} - \varepsilon$

However, this is still slightly less than one prime on average if n runs between N and $2N$ among $\{n+h_i\}_{i=1}^{k}$. (For $\vartheta = 1/2$) even worse than the trivial weighting (ii) $a_n = \begin{cases} 1 & \text{if } n+h_1 \in \mathcal{P} \\ 0 & \text{o.w.} \end{cases}$ which yields exactly one prime on average. i.

Let us observe that if we consider primes in the interval $(n, n+H]$, $H = \eta \log N$, η fixed, but arbitrarily small then we have in the

(i) found already $1 - O\left(\frac{1}{\sqrt{k}}\right)$ primes on average over n

(ii) but we have still for every n (i.e. on average too) $\eta \log N$ candidates which naively promises η primes on average and $\eta > \frac{c}{\sqrt{k}}$ if k is large, η is fixed, positive.

In fact, the same procedure gives in case of $h_0 \notin \mathcal{H}$

If $\mathcal{H}_0 = \mathcal{H} \cup h_0$ then compared with $S^*(h_i)$, $h_i \in \mathcal{H}$

(i) $\nu_p(\mathcal{H}) - 1$ has to be replaced by $\nu_p(\mathcal{H}_0) - 1$

k and so l remains the same, we have a factor $\frac{p}{p-1}$ so we obtain

$$S^*(h_0) \sim \mathfrak{S}(\mathcal{H} \cup h_0) \binom{2l}{l} \frac{(\log R)^{k+2l}}{(k+2l)!} \sim S \cdot \frac{\mathfrak{S}(\mathcal{H} \cup h_0)}{\mathfrak{S}(\mathcal{H})}$$

To finish we have to show that

$$(*)\sum_{h_0=1,\, h_0\notin \mathcal{H}}^{H} \frac{\mathfrak{S}(\mathcal{H}\cup h_0)}{\mathfrak{S}(\mathcal{H})} \gg C(k) H \quad \text{if } H\to\infty,\ k \text{ fixed}$$

In the original work this was done by Gallagher's theorem

$$(**)\sum_{\substack{\mathcal{H}\subset[1,H]\\ |\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H}) \sim \frac{H^k}{k!} \quad \left(\text{i.e. } \mathfrak{S}(\mathcal{H})=1 \text{ on average over } \mathcal{H}\subset[1,H]\right)$$

if k is fixed and $H\to\infty$. This yields $(*)$ on average over all $\mathcal{H}$. (with $1+o(1)$ in place of $\gg$)

Alternative solution (J.P.): simpler argument

<u>Remark</u> It is enough to prove $(*)$ for one k-tuple $\mathcal{H}$ for any k (or even for at least one sufficiently large k).

Let $\mathcal{H}=\left\{ i\cdot \prod_{p\le 2k} p \right\}_{i=1}^{2k}$. Then we have $\nu_p(\mathcal{H})=1$ for $p\le 2k$, hence

Let $\nu_p(\mathcal{H})=\nu_p$, $\nu_p(\mathcal{H}\cup h_0)=\nu_p'$, $2\mid h$, $p\neq 2$

Then $\mathfrak{S}(\mathcal{H}\cup h_0) > \mathfrak{S}(\mathcal{H})$ for $2\mid h$.

Proof: For $2<p\le 2k$ we have

$$\frac{1-\frac{\nu_p'}{p}}{1-\frac{\nu_p}{p}} \ge \frac{1-\frac{2}{p}}{1-\frac{1}{p}} = \frac{p-2}{p-1}$$

$$\prod_{p\le 2k} \frac{p-2}{p-1}\left(1-\frac{1}{p}\right)^{-1} = \prod_{p\le 2k}\left(1-\frac{1}{p-1}\right)^2 \ge \prod_{p=3}^{\infty}\left(1-\frac{1}{(p-1)^2}\right) = C_0 = 0.66\ldots$$

For $p>2k$

$$\prod_{p>2k} \frac{1-\nu_p'/p}{1-\nu_p/p}\cdot\frac{1}{1-\frac{1}{p}} \ge \prod_{p>2k}\frac{1-\frac{\nu_p+1}{p}}{1-\frac{\nu_p+1}{p}+\frac{\nu_p}{p^2}}$$

$$\geq \prod_{p \geq 2k} \left(1 + O\left(\frac{k}{p^2}\right)\right) \geq 1 + O\left(k \sum_{p > 2k} \frac{1}{p^2}\right)$$

$$\geq 1 + O\left(\frac{1}{\log k}\right)$$

So we have for $k > k_0$, $2 \mid k$ $\quad \dfrac{\sigma(\mathcal{H} \cup h_0)}{\sigma(\mathcal{H})} > 2c_0 + O\left(\dfrac{1}{\log k}\right)$
since we have an extra factor $\left(1 - \frac{1}{p}\right)^{-1} = 2$ for $p = 2$ ∎