7 New formulas and bounds for $g^{-1}(n)$ and its summatory function

7.1 Exact probabilistic bounds on the distributions of component sequences

We have remarked already in the introduction that the relation of the component functions, $g^{-1}(n)$ and $C_k(n)$, to the canonical additive functions $\omega(n)$ and $\Omega(n)$ leads to the regular properties of these functions witnessed in Table T.1. In particular, each of $\omega(n)$ and $\Omega(n)$ satisfies an Erdös-Kac theorem that shows that a shifted and scaled variant of each of the sets of these function values can be expressed through a limiting normal distribution as $n \to \infty$. This extremely regular tendency of these functions towards their average order is inherited by the component function sequences we are summing in the approximation of $M(x)$ stated by Proposition 8.1. In the remainder of this section we establish more technical analytic proofs of related properties of our key sequences, again in the spirit of Montgomery and Vaughan's reference.

Proposition 7.1. *For $|z| < 2$, let the summatory function be defined as*

$$\widehat{A}_z(x) := \sum_{n \leq x} (-1)^{\omega(n)} C_{\Omega(n)}(n) z^{\Omega(n)}.$$

Let the function $F(s,z)$ is defined for $\operatorname{Re}(s) > 1$ and $|z| < |P(s)|^{-1}$ in terms of the prime zeta function by

$$F(s,z) := \frac{1}{1-P(s)z} \times \prod_p \left(1 - \frac{1}{p^s}\right)^z.$$

Then we have that for large x

$$\widehat{A}_z(x) = \frac{x \cdot F(2,z)}{\Gamma(z)} (\log x)^{z-1} + O_z\left(x \cdot (\log x)^{\operatorname{Re}(z)-2}\right), |z| < P(2)^{-1}.$$

Proof. **(TODO)** We know from the proof of Proposition 4.1 that for $n \geq 2$

$$C_{\Omega(n)}(n) = (\Omega(n))! \times \prod_{p^\alpha || n} \frac{1}{\alpha!}.$$

Then we can generate the denominator terms by the Dirichlet series

$$\sum_{n \geq 1} \frac{C_{\Omega(n)}(n)}{(\Omega(n))!} \cdot \frac{(-1)^{\omega(n)} z^{\Omega(n)}}{n^s} = \prod_p \left(1 + \sum_{r \geq 1} \frac{z^{\Omega(p^r)}}{r! \cdot p^{rs}}\right)^{-1} = \exp\left(z \cdot P(s)\right), \operatorname{Re}(s) > 1, z \in \mathbb{C}.$$

By computing a Laplace transform on the right-hand-side of the above with respect to the variable z, we obtain

$$\sum_{n \geq 1} C_{\Omega(n)}(n) \cdot \frac{(-1)^{\omega(n)} z^{\Omega(n)}}{n^s} = \int_0^\infty e^{-t} \exp\left(tz \cdot P(s)\right) dt = \frac{1}{1-P(s)z}, \operatorname{Re}(s) > 1, |z| < |P(s)|^{-1}.$$

It follows that

$$\sum_{n \geq 1} \frac{(-1)^{\omega(n)} C_{\Omega(n)}(n) z^{\Omega(n)}}{n^s} = \zeta(s)^z \times F(s,z),$$

where

$$F(s,z) := \frac{1}{1-P(s)z} \times \prod_p \left(1 - \frac{1}{p^s}\right)^z, \operatorname{Re}(s) > 1, |z| < |P(s)|^{-1}.$$

We adapt the details to the case where this method arises in the first application from [11, §7.4; Thm. 7.18] so that we can sum over our modified function depending on $\Omega(n)$. In fact, we notice that since $|z|^{\Omega(n)} < n$ for $|z| < P(2)^{-1}$, we have the exact DGF

$$\mathcal{H}(s) := \sum_{n \geq 1} \frac{\lambda(n) C_{\Omega(n)}(n)}{n^s},$$

which is absolutely convergent for $\mathrm{Re}(s) \geq 2$. The DGF $\mathcal{H}(s)$ is thus an analytic function of s whenever $\mathrm{Re}(s) \geq 2$, and so we can differentiate it any integer $m \geq 0$ number of times to still obtain an absolutely convergent series of the form

$$\left| \sum_{n\geq 1} \frac{(-1)^{\omega(n)} C_{\Omega(n)}(n) (\log n)^m z^{\Omega(n)}}{n^s} \right| < +\infty, \mathrm{Re}(s) \geq 2, |z| < P(2)^{-1}.$$

Let the function $d_z(n)$ be generated as the coefficients of the DGF $\zeta(s)^z$ for $\mathrm{Re}(s) > 1$, with corresponding summatory function $D_z(x) := \sum_{n\leq x} d_z(n)$. Adopting the notation from the reference, we set $b_z(n) := (-1)^{\omega(n)} C_{\Omega(n)}(n) z^{\Omega(n)}$, let the convolution $a_z(n) := \sum_{d|n} b_z(d) d_z(n/d)$, and define the summatory function $A_z(x) := \sum_{n\leq x} a_z(n)$. The theorem in [11, Thm. 7.17; §7.4] implies that for any $z \in \mathbb{C}$ and $x \geq 2$

$$D_z(x) = \frac{x(\log x)^{z-1}}{\Gamma(z)} + O\left(x \cdot (\log x)^{\mathrm{Re}(z)-2}\right).$$

Then we have that

$$\begin{aligned} A_z(x) &= \sum_{m\leq x/2} b_z(m) D_z(x/m) + \sum_{x/2<m\leq x} b_z(m) \\ &= \frac{x}{\Gamma(z)} \times \sum_{m\leq x/2} \frac{b_z(m)}{m^2} \log\left(\frac{x}{m}\right)^{z-1} + O\left(x \sum_{m\leq x} \frac{|b_z(m)|}{m^2} \times \log\left(\frac{2x}{m}\right)^{\mathrm{Re}(z)-2}\right). \end{aligned}$$

The error term in the previous equation satisfies

$$\begin{aligned} x \sum_{m\leq x} \frac{|b_z(m)|}{m^2} \times \log\left(\frac{2x}{m}\right)^{\mathrm{Re}(z)-2} &\ll x(\log x)^{\mathrm{Re}(z)-2} \sum_{m\leq\sqrt{x}} \frac{|b_z(m)|}{m^2} + x(\log x)^{-(R+2)} \sum_{m>\sqrt{x}} \frac{|b_z(m)|}{m^2} (\log m)^{2R} \\ &\ll x(\log x)^{\mathrm{Re}(z)-2}, |z| \leq R. \end{aligned}$$

In the main term estimate for $A_z(x)$, when $m \leq \sqrt{x}$ we have

$$\log\left(\frac{x}{m}\right)^{z-1} = (\log x)^{z-1} + O\left((\log m)(\log x)^{\mathrm{Re}(z)-2}\right).$$

The remaining main term sum over the interval $m \leq x/2$ corresponds to bounding

$$\begin{aligned} \sum_{m\leq x/2} b_z(m) D_z(x/m) &= x(\log x)^{z-1} \sum_{m\leq x/2} \frac{b_z(m)}{m^2} \\ &\quad + O\left(x(\log x)^{\mathrm{Re}(z)-2} \sum_{m\leq\sqrt{x}} \frac{|b_z(m)|}{m^2} + x(\log x)^{R-1} \sum_{m>\sqrt{x}} \frac{|b_z(m)|}{m^2}\right) \\ &= x(\log x)^{z-1} F(2,z) + O\left(x(\log x)^{\mathrm{Re}(z)-2} \sum_{m\geq 1} \frac{b_z(m)(\log m)^{2R+1}}{m^2}\right). \end{aligned}$$

□

Remark 7.2 (A standard simplifying assumption)**.** Let the constant $\widehat{c} \approx 1.5147$ be defined explicitly as the product of primes

$$\widehat{c} := \frac{1}{6} \times \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right)^{-1}.$$

This constant is related to expressions of the asymptotic densities of the sets

$$N_k(x) := \{n \leq x : \Omega(n) - \omega(n) = k\},$$

for integers $k \geq 0$ in the form of [11, §2.4]

$$N_k(x) = d_k x + O\left(\left(\frac{3}{4}\right)^k \sqrt{x}(\log x)^{4/3}\right), \tag{27a}$$

where for each natural number $k \geq 0$, $d_k > 0$ is an absolute constant that satisfies

$$d_k = \frac{\widehat{c}}{2^k} + O\left(5^{-k}\right). \tag{27b}$$

A hybrid DGF generating function for these densities is given by

$$\sum_{k\geq 0} d_k z^k = \prod_p \left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{p-z}\right). \tag{27c}$$

The limiting distribution of $\Omega(n) - \omega(n)$ is utilized in the proof of Theorem 7.3.

For $m \leq \omega_{\max}$ and $k \leq \Omega_{\max}$, as $n \to \infty$ we expect

$$\mathbb{P}\left(\omega(n) = m | \Omega(n) = k\right) \approx \frac{\omega_{\max} + 1 - k}{\omega_{\max}},$$

so that the conditional distribution of $\omega(n), \Omega(n)$ is not uniform over its bounded range. However, we do as is standard fare in proofs of the more traditional Erdös-Kac theorems require the simplifying assumption that as $n \to \infty$, we expect independently that $\omega(n), \Omega(n)$ are approximately equally likely to assume any values in some bounded $[1, M]$. This means we can treat the difference $\Omega(n) - \omega(n)$ as being approximately randomly distributed over some bounded range of its possible values. For a more rigorous treatment of this underlying principle see [4, 2, 15].

Theorem 7.3. *We have uniformly for $1 \leq k < \log\log x$ that as $x \to \infty$*

$$\widehat{C}_k(x) := \sum_{\substack{n \leq x \\ \Omega(n) = k}} \lambda(n)(-1)^{\omega(n)} C_k(n) \asymp \frac{x}{\log x} \cdot \frac{(-1)^{k-1}(\log\log x + P(2))^k}{k!}\left[1 + O\left(\frac{k}{(\log\log x)^3}\right)\right].$$

Proof. The proof is a similar adaptation of the method of Montgomery and Vaughan we cited in Remark 5.3 to prove our variant of Theorem 3.7. We begin by bounding a contour integral over the error term for fixed large x for $r := \frac{k-1}{\log\log x}$ with $r < P(2)^{-1} \approx 2.21118$:

$$\begin{aligned}
\left|\int_{|z|=r} \frac{x \cdot (\log x)^{-(\operatorname{Re}(z)+2)}}{z^{k+1}} dz\right| &\ll x(\log x)^{-(r+2)} r^{-(k+1)} \ll \frac{x}{(\log x)^2} \cdot \frac{(\log\log x)^{k+1}}{(k-1)^{k+1}} \cdot \frac{1}{e^{k-1}} \\
&\ll \frac{x}{(\log x)^2} \cdot \frac{(\log\log x)^{k+1}}{(k-1)^{3/2}} \cdot \frac{1}{e^{2k}(k-1)!} \\
&\ll \frac{x}{(\log x)^2} \cdot \frac{(\log\log x)^{k-1}}{(k-1)!} \ll \frac{x}{\log x} \cdot \frac{k \cdot (\log\log x)^{k-5}}{(k-1)!}.
\end{aligned}$$

We must find an asymptotically accurate main term approximation to the coefficients of the following contour integral for $r \in [0, z_{\max}]$ where $z_{\max} < P(2)^{-1} \approx 2.21118$:

$$\widetilde{A}_r(x) := -\int_{|z|=r} \frac{x \cdot \exp(-P(2)z)(\log x)^{-z}}{(\log x)\Gamma(1+z) \cdot z^{k+1}(1 + P(2)z)} dz. \tag{28}$$

Finding an exact formula for the derivatives of the function that is implicit to the Cauchy integral formula (CIF) for (28) is complicated significantly by the need to differentiate $\Gamma(1+z)^{-1}$ up to integer order k in the formula. What results in this case is a mess of confluent hypergeometric function approximations depending on k and an extra factor of $(k!)^{-1}$ in the main-most term that *substantially* complicates the formative summation patterns related to the incomplete gamma function in the $\widehat{\pi}_k(x)$ cases from Section 5.2. We can show that provided a restriction on the uniform bound parameter to $1 \leq r < 1$, we can approximate the contour integral in (28) using a sane bounding procedure where the resulting main term is accurate up to a bounded constant factor.

We observe that for $r := 1$, the function $|\Gamma(1 + re^{2\pi\imath t})|$ has a singularity (pole) when $t := \frac{1}{2}$. Thus we restrict the range of $|z| = r$ so that $0 \leq r < 1$ to necessarily avoid this problematic value of t when we parameterize $z = re^{2\pi\imath t}$ as a real integral over $t \in [0, 1]$. Then we can compute the finite extremal values as

$$\min_{\substack{0 \leq r < 1 \\ 0 \leq t \leq 1}} |\Gamma(1 + re^{2\pi\imath t})| = |\Gamma(1 + re^{2\pi\imath t})|\Bigg|_{(r,t) \approx (1, 0.740592)} \approx 0.520089$$

$$\max_{\substack{0 \leq r < 1 \\ 0 \leq t \leq 1}} |\Gamma(1 + re^{2\pi\imath t})| = |\Gamma(1 + re^{2\pi\imath t})|\Bigg|_{(r,t) \approx (1, 0.999887)} \approx 1.$$

This shows that

$$\widetilde{A}_r(x) \asymp -\int_{|z|=r} \frac{x \cdot \exp(-P(2)z)(\log x)^{-z}}{(\log x) \cdot z^{k+1}(1 + P(2)z)} dz, \tag{29}$$

where as $x \to \infty$

$$\frac{\widetilde{A}_r(x)}{-\int_{|z|=r} \frac{x \cdot \exp(-P(2)z)(\log x)^{-z}}{(\log x) \cdot z^{k+1}(1+P(2)z)} dz} \in [1, 1.92275].$$

In particular, this argument holds by an analog to the mean value theorem for real integrals based on sufficient continuity conditions on the parameterized path and the smoothness of the integrand viewed as a function of z.

By induction we can compute the remaining coefficients $[z^k]\Gamma(1 + z) \times \widehat{A}_z(x)$ with respect to x for fixed $k \leq \log\log x$ using the CIF. Namely, it is not difficult to see that for any integer $m \geq 0$, we have the m^{th} partial derivative of the integrand with respect to z has the following expansion:

$$\begin{aligned}
\frac{1}{m!} \times \frac{\partial^{(m)}}{\partial z^{(m)}} \left[\frac{(\log x)^{-z}}{1 + P(2)z}\right]\Bigg|_{z=0} &= \sum_{j=0}^{m} \frac{(-1)^m P(2)^j (\log\log x + P(2))^{m-j}}{(m-j)!} \\
&= \frac{e \cdot (-P(2))^m (\log x)^{\frac{1}{P(2)}}}{m!} \times \Gamma\left(m + 1, 1 + \frac{\log\log x}{P(2)}\right) \\
&\sim \frac{(-1)^m (\log\log x + P(2))^m}{m!}.
\end{aligned}$$

Now by parameterizing the countour around $|z| = r := \frac{k-1}{\log\log x} < 1$ we deduce that the the main term of our approximation corresponds to

$$-\int_{|z|=r} \frac{x \cdot \exp(-P(2)z)(\log x)^{-z}}{(\log x) z^{k+1}(1 + P(2)z)} dz \asymp \frac{x}{\log x} \cdot \frac{(-1)^{k-1}(\log\log x + P(2))^k}{k!}. \qquad \square$$

Lemma 7.4. *We have that as $x \to \infty$*

$$\left|\mathbb{E}\left[\sum_{n \leq x} \lambda(n)(-1)^{\omega(n)} C_{\Omega(n)}(n)\right]\right| \asymp \frac{P(2)}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\log\log x}}.$$

Proof. We observe that

$$\sum_{n \leq x} \lambda(n)(-1)^{\omega(n)} C_{\Omega(n)}(n) = \sum_{k=1}^{\log_2(x)} \sum_{\substack{n \leq x \\ \Omega(n) = k}} \lambda(n)(-1)^{\omega(n)} C_{\Omega(n)}(n) = \sum_{k=1}^{\log_2(x)} \widehat{C}_k(x).$$

We claim that

$$\sum_{k=1}^{\log_2(x)} \widehat{C}_k(x) \asymp \sum_{k=1}^{\log\log x} \widehat{C}_k(x). \tag{30}$$

To prove (30), it suffices to show that

$$\left|\frac{\sum\limits_{\log\log x<k\leq\log_2(x)} \widehat{C}_k(x)}{\sum\limits_{k=1}^{\log\log x} \widehat{C}_k(x)}\right| = o(1), \text{ as } x\rightarrow\infty. \tag{31}$$

We first compute the absolute value of the following summatory function by applying Theorem 7.3 for large $x\rightarrow\infty$:

$$\begin{aligned}\left|\sum_{k=1}^{\log\log x} \widehat{C}_k(x)\right| &\asymp \left|\frac{x}{\log x} - \frac{x\cdot e^{-P(2)}\Gamma(\log\log x, -(\log\log x)+P(2))}{(\log x)^2\cdot\Gamma\left(\log\log x\right)}\right|\left[1+O\left(\frac{1}{(\log\log x)^2}\right)\right] \\ &\sim \frac{x}{\log x} + \frac{x\cdot\sqrt{\log\log x}}{\sqrt{2\pi}(\log\log x+P(2))}(1+o(1)).\end{aligned} \tag{32}$$

We define the following component sums for large x and $0<\varepsilon<1$ so that $(\log\log x)^{\varepsilon\frac{\log\log x}{\log\log\log x}} = o(\log x)$:

$$S_{2,\varepsilon}(x) := \sum_{P(2)^{-1}\log\log x<k\leq\log\log x} \widehat{C}_k(x).$$

Then

$$\sum_{P(2)^{-1}\log\log x<k\leq(\log\log x)^{\varepsilon\frac{\log\log x}{\log\log\log x}}} \widehat{C}_k(x) \gg S_{2,\varepsilon}(x),$$

with equality as $\varepsilon\rightarrow 1$ so that the upper bound of summation tends to $\log x$. To show that (31) holds, observe that whenever $\Omega(n)=k$, we have that $C_{\Omega(n)}(n)\leq k!$. We can bound the sum defined above using Theorem 5.4 for large $x\rightarrow\infty$ as

$$\begin{aligned}S_{2,\varepsilon}(x) &\leq \sum_{\log\log x\leq k\leq\log\log x}\ \sum_{\substack{n\leq x\\ \Omega(n)=k}} C_{\Omega(n)}(n) \ll \sum_{k=\log\log x}^{(\log\log x)^{\varepsilon\frac{\log\log x}{\log\log\log x}}} \frac{\widehat{\pi}_k(x)}{x}\cdot k! \\ &\ll \sum_{k=\log\log x}^{(\log\log x)^{\varepsilon\frac{\log\log x}{\log\log\log x}}} (\log x)^{\frac{k}{\log\log x}-1-\frac{k}{\log\log x}(\log k-\log\log\log x)}\cdot\left(\frac{k}{e}\right)^k\sqrt{2\pi k} \\ &\ll \sum_{k=\log\log x}^{\varepsilon\frac{\log\log x}{\log\log\log x}} (\log x)^{k\frac{\log\log\log x}{\log\log x}-1}\sqrt{k} \ll \frac{1}{(\log x)}\times\int_{\log\log x}^{\varepsilon\frac{\log\log x}{\log\log\log x}} (\log\log x)^t\sqrt{t}\cdot dt \\ &\ll \frac{1}{(\log x)}\sqrt{\frac{\varepsilon\cdot\log\log x}{\log\log\log x}}(\log\log x)^{\frac{\varepsilon\cdot\log\log x}{\log\log\log x}} = o(x),\end{aligned}$$

where $\lim_{x\rightarrow\infty}(\log x)^{\frac{1}{\log\log x}} = e$. By (32) this form of the ratio in (31) clearly tends to zero. If we have a contribution from the terms $\widehat{\pi}_k(x)$ as $\varepsilon\rightarrow 1$, e.g., if x is a power of two, then $C_{\Omega(x)}(x)=1$ by the formula in (9), so that the contribution from this upper-most indexed term is negligible:

$$x=2^k \implies \Omega(x)=k \implies C_{\Omega(x)}(x) = \frac{(\Omega(x))!}{k!} = 1.$$

The formula for the expectation claimed in the statement of this lemma above then follows from (32) by scaling by $\frac{1}{x}$ and dropping the asymptotically lesser error terms in the bound. $\square$

Remark 7.5. The signs of the functions estimated in Theorem 7.3 are dictated by the differences of the prime omega functions as $(-1)^{\Omega(n)-\omega(n)}$. It happens, as we have summarized above, that this distribution is fairly regular with limiting asymptotic densities of the distinct values of the difference between the additive functions. This signedness property, in place of the more natural $\lambda(n)$ weights as appear in Proposition 4.1, is necessary to simplify the DGF expansion we used to obtain the asymptotics for the summatory functions $\widehat{A}_z(x)$ in Proposition 7.1. It also leads to additional cancellation in the corresponding summatory functions and the resulting average order expectations we would obtain from these sums in this raw form.

An exact DGF expression for $\lambda(n)C_{\Omega(n)}(n)$ is in fact very much complicated by the need to estimate the asymptotics of the coefficients of the right-hand-side products

$$\begin{aligned}\sum_{n\geq 1}\frac{\lambda(n)C_{\Omega(n)}(n)z^{\Omega(n)}}{(\Omega(n))!\cdot n^s} &= \prod_p \left(2-\exp\left(-z\cdot p^{-s}\right)\right)^{-1}, \operatorname{Re}(s)>1, |z|<\log 2 \\ &= \exp\left(\sum_{j\geq 1}\sum_p \left(e^{-zp^{-s}}-1\right)^j \frac{1}{j}\right).\end{aligned}$$

It is unclear how to exactly, and effectively, bound the coefficients of powers of z in the DGF expansion defined by the last equation. We use an alternate method in the next corollary to obtain the asymptotics for the actual summatory functions on which we require tight average case bounds.

Corollary 7.6 (Summatory functions of the unsigned component sequences)**.** *We have that for large $x\geq 2$ and $1\leq k\leq \log\log x$*

$$\sum_{\substack{n\leq x\\ \Omega(n)=k}} C_{\Omega(n)}(n) \asymp \frac{3}{2\hat{c}}\cdot\frac{x}{(\log x)^2}\left[\frac{(\log\log x+P(2))^k}{k!}-\frac{(\log\log x+P(2))^{k-1}}{(k-1)!}\right].$$

Proof. We handle transforming our previous results for the sum over the unsigned sequence $C_{\Omega(n)}(n)$ such that $\Omega(n)=k$. The argument basically boils down to approximating the smooth summatory function of $\lambda_*(n) := (-1)^{\Omega(n)-\omega(n)}$ using the weighted densities defined by (27). We then have an integral formula involving the non-sign-weighted sequence that results by again applying ordinary Abel summation (and integrating by parts) in the form of

$$\begin{aligned}\sum_{n\leq x}\lambda_*(n)h(n) &= \left(\sum_{n\leq x}\lambda_*(n)\right)h(x) - \int_1^x \left(\sum_{n\leq t}\lambda_*(n)\right)h'(t)dt \qquad (33)\\ &\asymp \left\{\begin{array}{ll} u_t = L_*(t) & v_t' = h'(t)dt \\ u_t' = L_*'(t)dt & v_t = h(t)\end{array}\right\}\int_1^x \frac{d}{dt}\left[\sum_{n\leq t}\lambda_*(n)\right]h(t)dt.\end{aligned}$$

Let the signed left-hand-side summatory function in (33) for our function be defined by

$$\begin{aligned}\widehat{C}_{k,*}(x) &:= \left|\sum_{\substack{n\leq x\\ \Omega(n)=k}}\lambda(n)(-1)^{\omega(n)}C_{\Omega(n)}(n)\right| \\ &= \frac{x}{\log x}\cdot\frac{(\log\log x+P(2))^k}{k!}\left[1+O\left(\frac{1}{(\log\log x)^2}\right)\right],\end{aligned}$$

where the second equation follows from the proof of Theorem 7.3. Then by differentiating the formula we engineered well for ourselves in (33), and then summing over the uniform range of $1\leq k\leq \log\log x$, we can recover an approximation to the unsigned summatory function for the sequence we need to bound in later results proved in this section.

We handle the sign weighted terms by defining and approximating the asymptotic main term of the following summatory function (*cf.* Table T.2 starting on page 54):

$$\begin{aligned} L_*(t) &:= \sum_{n \leq t} \lambda(n)(-1)^{\omega(n)} = \sum_{j=0}^{\log_2(t)} (-1)^j \cdot \#\{n \leq t : \Omega(n) - \omega(n) = j\} \\ &\sim \sum_{j=0}^{\log_2(t)} \cdot \frac{\hat{c} \cdot t(-1)^j}{2^j} = \frac{2\hat{c} \cdot t}{3} + o(1), \text{ as } t \to \infty. \end{aligned}$$

The approximation to the densities d_k for the difference of the prime omega functions is cited from (27) [11, §2.4]. After applying the formula from (33), we deduce that the unsigned summatory function variant satisfies

$$\begin{aligned} \widehat{C}_{k,*}(x) &= \int_1^x L_*'(t) C_{\Omega(t)}(t) dt \qquad \implies C_{\Omega(x)}(x) \asymp \frac{\widehat{C}_{k,*}'(x)}{L_*'(x)} \\ C_{\Omega(x)}(x) &\asymp \frac{3}{2\hat{c}} \left[\frac{(\log\log x + P(2))^k}{(\log x) k!} \left(1 - \frac{1}{\log x}\right) + \frac{(\log\log x + P(2))^{k-1}}{(\log x)^2 (k-1)!}\right] =: \widehat{C}_{k,**}(x). \end{aligned}$$

So again applying the Abel summation formula, we obtain that

$$\begin{aligned} \sum_{\substack{n \leq x \\ \Omega(n) = k}} C_{\Omega(n)}(n) &\asymp \left| \int \widehat{C}_{k,**}'(x) dx \right| \\ &= \frac{3}{2\hat{c}} \cdot \frac{x}{(\log x)^2} \left[\frac{(\log\log x + P(2))^k}{k!} - \frac{(\log\log x + P(2))^{k-1}}{(k-1)!}\right]. \end{aligned}$$

This proves the stated formula, and it similarly holds uniformly for all $1 \leq k \leq \log\log x$ when x is large. □

Remark 7.7. Notice that even though we are using asymptotic notation ($\gg$ and $\asymp$) that does not preserve constant factors of its operands well (in principle), we are still making an effort to keep the sanctity of the multiplicative constants which we can be certain are exact in our new formulas. This is not an objection to nor ignorance of conventions, but rather a necessity in maintaining tight enough bounds so we can still sum over differences involving these functions within a small window of error. That we are not off by more than, say a factor of 2, as we established in proving Theorem 7.3, increases the accuracy of the next probabilistic results that will be important in bounding $|g^{-1}(n)|$ near its expectation, or average order asymptotics. In particular, we will require a fairly close bound near the expectation of this function in conjunction with the probabilistic statement of the result in Corollary 7.10 below. This means in practice that we are unable to be too imprecise with constant factors as error terms in the differences of $|g^{-1}(n)| - \mathbb{E}|g^{-1}(n)|$ can accumulate and generate non-negligible noise when we apply these results in the next section (n.b.).

Corollary 7.8 (Expectation formulas)**.** *We have that as $n \to \infty$*

$$\mathbb{E}|g^{-1}(n)| \asymp \frac{1}{\hat{c}\sqrt{2\pi}} (\log n)(\log\log n)^{3/2}.$$

Proof. We use the formula from Corollary 7.6 to find $\mathbb{E}[C_{\Omega(n)}(n)]$ up to a small bounded multiplicative constant factor as $n \to \infty$:

$$\begin{aligned} \mathbb{E}[C_{\Omega(n)}(n)] &= \sum_{k=1}^{\log_2(n)} \sum_{\substack{n \leq x \\ \Omega(n) = k}} C_k(n) \\ &\asymp \frac{1}{n} \times \sum_{k=1}^{\log\log n} \frac{3}{2\hat{c}} \cdot \frac{n}{(\log n)^2} \left[\frac{(\log\log n + P(2))^k}{k!} - \frac{(\log\log n + P(2))^{k-1}}{(k-1)!}\right] \end{aligned}$$

$$\asymp \frac{3}{2\hat{c}\sqrt{2\pi}} \cdot \frac{\sqrt{\log\log n}}{\log n}\left[1 + O\left(\frac{1}{\log\log n}\right)\right].$$

This implies that for large x

$$\int \frac{\mathbb{E}[C_{\Omega(x)}(x)]}{x} dx \asymp \frac{1}{\hat{c}\sqrt{2\pi}} \cdot (\log\log x)^{3/2}\left[1 + O\left(\frac{1}{\log\log x}\right)\right].$$

Therefore, citing the formula we derived in the proof of Corollary 6.6, we find that

$$\begin{aligned}
\mathbb{E}|g^{-1}(n)| &= \frac{6}{\pi^2}\left[\mathbb{E}[C_{\Omega(n)}(n)] + \sum_{d<n} \frac{\mathbb{E}[C_{\Omega(d)}(d)]}{d}\right] + O(1) \\
&\asymp \frac{3}{4\sqrt{2}\hat{c}} \cdot \operatorname{erfi}\left(\sqrt{\log\log n}\right) + \frac{1}{\hat{c}\sqrt{2\pi}}(\log n)(\log\log n)^{3/2} \\
&\asymp \frac{3}{4\sqrt{2\pi}\hat{c}} \cdot \frac{(\log n)}{\sqrt{\log\log n}} + \frac{1}{\hat{c}\sqrt{2\pi}}(\log n)(\log\log n)^{3/2}.
\end{aligned}$$

In the previous equation, we have used a known asymptotic expansion of the function $\operatorname{erfi}(z)$ about infinity in the form of [3, §3.2]

$$\operatorname{erfi}(z) = \frac{e^{z^2}}{\sqrt{\pi}}\left(z^{-1} + \frac{1}{2}z^{-3} + \frac{3}{4}z^{-5} + \cdots\right), \text{ as } |z| \to \infty.$$

This proves the claimed formula for the expectation of our key function. □

Theorem 7.9. *Let the mean and variance analogs be denoted by*

$$\mu_x(C) := \log\log x + P(2), \qquad \text{and} \qquad \sigma_x(C) := \sqrt{\mu_x(C)}$$

Set $Y > 0$ and suppose that $z \in [-Y, Y]$. Then we have uniformly for all $-Y \leq z \leq Y$ as $x \to \infty$ that

$$\frac{1}{x} \cdot \#\left\{2 \leq n \leq x : \frac{C_{\Omega(n)}(n) - \mu_x(C)}{\sigma_x(C)} \leq z\right\} = \frac{3e^{P(2)}}{2\widehat{c} \cdot (\log x)}\left[\Phi(z) + O\left(\frac{1}{(\log\log x)^{1/2}}\right)\right].$$

Proof. For large x and $n \leq x$, define the following auxiliary variables:

$$\alpha_n := \frac{C_{\Omega(n)}(n) - \mu_n(C)}{\sigma_n(C)}, \quad \beta_{n,x} := \frac{C_{\Omega(n)}(n) - \mu_x(C)}{\sigma_x(C)}.$$

Let the corresponding densities (whose limiting distributions we must verify) be defined by the functions

$$\Phi_1(x, z) := \frac{1}{x} \cdot \#\{n \leq x : \alpha_n \leq z\},$$

and

$$\Phi_2(x, z) := \frac{1}{x} \cdot \#\{n \leq x : \beta_{n,x} \leq z\}.$$

We first argue that it suffices to consider the distribution of $\Phi_2(x, z)$ as $x \to \infty$ in place of $\Phi_1(x, z)$ to obtain our desired result statement. In particular, the difference of the two auxiliary variables is neglibible as $x \to \infty$ for n, x taken over the ranges that contribute the non-trivial weight to the main term of each density function. We have for $\sqrt{x} \leq n \leq x$ and $C_{\Omega(n)}(n) \leq 2 \cdot \mu_x(C)$ that

$$|\alpha_n - \beta_{n,x}| \ll \frac{1}{\sigma_x(C)} \xrightarrow{x\to\infty} 0.$$

So we naturally prefer to estimate the easier forms of the distribution function $\Phi_2(x, z)$ when x is large, and for any fixed $z \in \mathbb{R}$. That is, we replace α_n by $\beta_{n,x}$ and estimate the limiting densities corresponding to these terms.

We use the formula proved in Corollary 7.6, which holds uniformly for x large when $1 \leq k \leq \log\log x$, to estimate the densities claimed within the ranges bounded by z as $x \to \infty$. We have already proved in Lemma 7.4 (in the signed summatory function case analysis) by applying Theorem 5.4 that to express an accurate asymptotic main term for these values, it suffices to omit the cases of $\Omega(n) = k$ for $k > \log\log x$ where we do not recover uniform formulas on these sums. The rest of our argument follows closely along with the method in the proof of the related theorem in [11, Thm. 7.21; §7.4].

Let $k \geq 1$ be a natural number defined by $k := t + P(2) + \log\log x$. We write the small parameter $\delta_{t,x} := \frac{t}{P(2)+\log\log x}$. When $|t| \leq \frac{1}{2}(P(2) + \log\log x)$, we have by Stirling's formula that

$$\frac{3}{2\hat{c}} \cdot \frac{x}{(\log x)^2} \frac{(\log\log x + P(2))^k}{k!} \sim \frac{3}{2\hat{c}\sqrt{2\pi}} \cdot \frac{x \cdot e^{P(2)+t}}{(\log x)(\log\log x + P(2))^{1/2}} (1+\delta_{t,x})^{-(\log\log x + P(2))(1+\delta_{t,x})+\frac{1}{2}}.$$

We have the uniform estimate $\log(1+\delta_{t,x}) = \delta_{t,x} - \frac{\delta_{t,x}^2}{2} + O(|\delta_{t,x}|^3)$ whenever $|\delta_{t,x}| \leq \frac{1}{2}$. Then we can expand the factor involving $\delta_{t,x}$ in the previous equation as follows:

$$\begin{aligned}(1+\delta_{t,x})^{-(P(2)+\log\log x)(1+\delta_{t,x})+\frac{1}{2}} &= \exp\left(\left(\frac{1}{2} - (P(2)+\log\log x)(1+\delta_{t,x})\right) \times \left(\delta_{t,x} - \frac{\delta_{t,x}^2}{2} + O(|\delta_{t,x}|^3)\right)\right) \\ &= \exp\left(-t + \frac{t-t^2}{2\mu_x(C)} - \frac{(t^2-2t^3)}{4\mu_x(C)^2} + O\left(\frac{|t|^3}{\mu_x(C)^2}\right)\right).\end{aligned}$$

For both $|t| \leq (P(2)+\log\log x)^{1/2}$ and $(P(2)+\log\log x)^{1/2} < |t| \leq (P(2)+\log\log x)^{2/3}$, we see that

$$\frac{t}{P(2)\log\log x} \ll \frac{1}{\sqrt{P(2)+\log\log x}} + \frac{|t|^3}{(P(2)+\log\log x)^2}.$$

Similarly, for $|t| \leq 1$ and $|t| > 1$, we see that both

$$\frac{t^2}{(P(2)+\log\log x)^2} \ll \frac{1}{\sqrt{P(2)+\log\log x}} + \frac{|t|^3}{(P(2)+\log\log x)^2}.$$

Let the error terms in (x,t) be denoted by

$$\widetilde{E}(x,t) := O\left(\frac{1}{\sigma_x(C)}\right) + O\left(\frac{|t|^3}{\mu_x(C)^2}\right).$$

Combining these estimates with the previous computations, we can deduce that uniformly for $|t| \leq (P(2) + \log\log x)^{2/3}$

$$\frac{3}{2\hat{c}} \cdot \frac{x}{(\log x)^2} \frac{(\log\log x + P(2))^k}{k!} \sim \frac{3e^{P(2)}}{2\hat{c}\sqrt{2\pi}} \cdot \frac{x}{(\log x)\sigma_x(C)} \cdot \exp\left(-\frac{t^2}{2\sigma_x(C)^2}\right) \times \left[1 + \widetilde{E}(x,t)\right].$$

Hence, by the formula from Corollary 7.6,

$$\sum_{\substack{n \leq x \\ \Omega(n)=k}} C_{\Omega(n)}(n) = \frac{3e^{P(2)}}{2\hat{c}\sqrt{2\pi}} \cdot \frac{x}{(\log x)\sigma_x(C)} \cdot \exp\left(-\frac{(k-\mu_x(C))^2}{2\sigma_x(C)^2}\right) \times \left[1 + \widetilde{E}(x, k-\mu_x(C))\right].$$

By the argument in the proof of Lemma 7.4, we see that the contributions of these summatory functions for $k \leq P(2) + \log\log x - (P(2)+\log\log x)^{2/3}$ is negligible. We also require that $k \leq \log\log x$ as we have worked out in Theorem 7.3. So we sum over a corresponding range of

$$P(2) + \log\log x - (P(2)+\log\log x)^{2/3} \leq k \leq R \cdot \log\log x,$$

for $R := 1 + \frac{z}{\sigma_x(C)}$ to approximate the stated normalized densities. Then finally as $x \to \infty$, the three terms that result (one main term, two error terms) can be considered to correspond to a Riemann sum for an associated integral. ∎

Corollary 7.10. *Let $Y > 0$ and $z \in [-Y, Y]$. Then uniformly for all $-Y \leq z \leq Y$ as $x \rightarrow \infty$ we have that*

$$\frac{1}{x} \cdot \#\left\{2 \leq n \leq x : |g^{-1}(n)| - \mathbb{E}|g^{-1}(n)| \leq z\right\} = \frac{3e^{P(2)}}{2\widehat{c} \cdot (\log x)} \left[\Phi\left(\frac{\frac{\pi^2}{6} z - \mu_x(C)}{\sigma_x(C)}\right) + O\left(\frac{1}{\sqrt{\log\log x}}\right)\right].$$

Proof. We compute using the argument sketched in the proof of Corollary 6.6 from Section 6.3 that

$$|g^{-1}(n)| - \mathbb{E}|g^{-1}(n)| \sim \frac{6}{\pi^2} C_{\Omega(n)}(n).$$

Then the result follows from Theorem 7.9. We can also compute using Corollary 7.10 that

$$\frac{\pi^2 e^{P(2)}}{4\hat{c} \cdot (\log x) \cdot \sigma_x(C)} \times \int_{-\infty}^{\infty} z \cdot \Phi'\left(\frac{\frac{\pi^2}{6} z - \mu_x(C)}{\sigma_x(C)}\right) dz = \frac{9e^{P(2)}}{\pi^2 \hat{c}(\log x)} \sqrt{\log\log x + P(2)} + o(1) \xrightarrow{x \rightarrow \infty} 0. \tag{34}$$

So we interpret this calculation to mean that the contribution from the sum over $|g^{-1}(n)|$ where $g^{-1}(n)$ is not very close to its average order is essentially negligible. We will use this property in the proof of Theorem 7.12 in the next subsection. □

7.2 Establishing initial lower bounds on the summatory functions $G^{-1}(x)$

Definition 7.11. Let the summatory function $G_E^{-1}(x)$ be defined for $x \geq 1$ by

$$G_E^{-1}(x) := \sum_{n \leq (\log x)^{\frac{7}{3}} (\log\log x)} \lambda(n) \times \sum_{\substack{d|n \\ d>e}} \frac{(\log d)^{\frac{3}{4}}}{\log\log d}. \tag{35}$$

The subscript of E is a formality of notation that does not correspond to an actual parameter or any implicit dependence on E in the function defined above.

Theorem 7.12. *For all sufficiently large integers $x \rightarrow \infty$, we have that*

$$|G^{-1}(x)| \gg |G_E^{-1}(x)|.$$

Proof. First, consider the following upper bound on $|G_E^{-1}(x)|$:

$$\begin{aligned}
|G_E^{-1}(x)| &= \left| \sum_{e \leq n \leq (\log x)^{7/3} (\log\log x)} \lambda(n) \times \sum_{\substack{d|n \\ d>e}} \frac{(\log d)^{\frac{3}{4}}}{\log\log d} \right| \\
&\ll \sum_{e < d \leq (\log x)^{7/3} (\log\log x)} \frac{(\log d)^{\frac{3}{4}}}{\log\log d} \cdot \left\lfloor \frac{(\log x)^{7/3} (\log\log x)}{d} \right\rfloor \\
&\ll (\log x)^{7/3} (\log\log x) \times \int_e^{(\log x)^{7/3} (\log\log x)} \frac{(\log t)^{\frac{3}{4}}}{t \cdot \log\log t} dt \\
&= (\log x)^{7/3} (\log\log x) \times \operatorname{Ei}\left(\frac{7}{4} \log\log\left((\log x)^{7/3} (\log\log x)\right)\right) \\
&\ll (\log x)^{7/3} (\log\log x) (\log\log\log x)^2.
\end{aligned} \tag{36}$$

We need a couple of observations to sum $G^{-1}(x)$ in absolute value and bound it from below. First, as noted in [11, §7.4], the function $\mathcal{G}(z)$ from Theorem 3.6 satisfies

$$\mathcal{G}\left(\frac{k-1}{\log\log x}\right) = 1 + O(1), k \leq \log\log x,$$