MODERATE AND LARGE DEVIATIONS FOR ERDŐS-KAC THEOREM

BEHZAD MEHRDAD AND LINGJIONG ZHU

Abstract. Erdős-Kac theorem is a celebrated result in number theory which says that the number of distinct prime factors of a uniformly chosen random integer satisfies a central limit theorem. In this paper, we establish the large deviations and moderate deviations for this problem in a very general setting for a wide class of additive functions.

1. Introduction

Let \( V(n) \) be a random integer chosen uniformly from \( \{1, 2, \ldots, n\} \) and let \( X(n) \) be the number of the distinct primes in the factorization of \( V(n) \). A celebrated result by Erdős and Kac \[9, 10\] says that

\[
\frac{X(n) - \log \log n}{\sqrt{\log \log n}} \to N(0, 1),
\]

as \( n \to \infty \). This is a deep extension of Hardy-Ramanujan Theorem (see \[13\]).

In addition, the central limit theorem holds in a more general setting for additive functions. A formal treatment and proofs can be found in e.g. Durrett \[7\].

In terms of rate of convergence to the Gaussian distribution, i.e. Berry-Esseen bounds, Rényi and Turán \[18\] obtained the sharp rate of convergence \( O(1/\sqrt{\log \log n}) \). Their proof is based on the analytic theory of Dirichlet series. Recently, Harper \[14\] used a more probabilistic approach and used Stein’s method to get an upper bound of rate of convergence of the order \( O(\log \log \log n/\sqrt{\log \log n}) \).

In terms of large deviations, Radziwill \[17\] used analytic number theory approach to get a series of asymptotic estimates. Féray et al. \[12\] proved precise large deviations using the mod-Poisson convergence method developed in Kowalski and Nikeghbali \[16\].

In this paper, we study the large deviation principle and moderate deviation principle in the sense of Erdős and Kac. Instead of precise deviations that have been studied in Féray et al. \[12\]. Our large deviations and moderate deviations results are in the sense of Donsker-Varadhan \[19, 3, 4, 5, 6\]. Our proofs are probabilistic and require only elementary number theory results, in constrast to the analytical number theory approach in Radziwill \[17\].

Both large deviations and moderate deviations in our paper are proved for a much wider class of additive functions than what have been studied in Radziwill \[17\] and Féray et al. \[12\].

Date: 8 December 2013. Revised: 8 December 2013.
2000 Mathematics Subject Classification. 60F10,11N37.
Key words and phrases. Erdős-Kac theorem, additive functions, number of distinct prime factors, large deviations, moderate deviations.
Before we proceed, let us introduce the formal definition of large deviations. A sequence \((P_n)_{n \in \mathbb{N}}\) of probability measures on a topological space \(X\) satisfies the large deviation principle with speed \(b_n\) and rate function \(I : X \to \mathbb{R}\) if \(I\) is non-negative, lower semicontinuous and for any measurable set \(A\), we have

\[
\inf_{x \in A^o} I(x) \leq \liminf_{n \to \infty} \frac{1}{b_n} \log P_n(A) \leq \limsup_{n \to \infty} \frac{1}{b_n} \log P_n(A) \leq \inf_{x \in \overline{A}} I(x).
\]

(1.2)

Here, \(A^o\) is the interior of \(A\) and \(\overline{A}\) is its closure. We refer to Dembo and Zeitouni \([2]\) or Varadhan \([20]\) for general background of large deviations and the applications.

Let \(X_1, \ldots, X_n\) be a sequence of \(\mathbb{R}^d\)-valued i.i.d. random vectors with mean 0 and covariance matrix \(C\) that is invertible. Assume that \(E[e^{\theta X_1}] < \infty\), for \(\theta\) in some ball around the origin. For any \(\sqrt{n} \ll a_n \ll n\), a moderate deviation principle says that for any Borel set \(A\),

\[
\frac{1}{2} \inf_{x \in A^o} \langle x, C^{-1} x \rangle \leq \liminf_{n \to \infty} \frac{n}{a_n^2} \log \mathbb{P} \left( \frac{1}{a_n} \sum_{i=1}^{n} X_i \in A \right) \leq \limsup_{n \to \infty} \frac{n}{a_n^2} \log \mathbb{P} \left( \frac{1}{a_n} \sum_{i=1}^{n} X_i \in A \right) \leq \frac{1}{2} \inf_{x \in \overline{A}} \langle x, C^{-1} x \rangle.
\]

(1.3)

In other words, \(\frac{1}{a_n^2} \sum_{i=1}^{n} X_i \in \cdot\) satisfies a large deviation principle with the speed \(a_n\). The above classical result can be found for example in [2]. Moderate deviation principle fills in the gap between central limit theorem and large deviation principle.

In this paper, we are interested to prove both large deviations and moderate deviations for Erdős-Kac theorem for a wide class of additive functions.

2. Main Results

Throughout this paper, we let \(\mathcal{P}\) denote the set of all the prime numbers.

**Assumption 1.** Let \(g\) be a strongly additive, i.e. \(g(p^k) = g(p)\) for all primes \(p\) and integers \(k \geq 1\), and \(g(mn) = g(m) + g(n)\) whenever \(\gcd(m, n) = 1\). In addition, we assume that there exists a probability measure \(\rho\) on \(\mathbb{R}\) so that

- For any \(\theta \in \mathbb{R}\), \(\int_{\mathbb{R}} e^{\theta y} \rho(dy) < \infty\).
- For any \(\theta \in \mathbb{R}\), \(\int_{\mathbb{R}} e^{\theta y} \rho_n(dy) \to \int_{\mathbb{R}} e^{\theta y} \rho(dy)\), where

\[
\rho_n(A) = \frac{\sum_{g(p) \in \mathcal{P}, p \leq n, p \in \mathcal{P}} \frac{1}{p}}{\sum_{p \leq n, p \in \mathcal{P}} \frac{1}{p}},
\]

(2.1)

for any Borel set \(A \subset \mathbb{R}\).

Let \(V(n)\) be a uniformly chosen random integer from \(\{1, 2, \ldots, n\}\) and \(Z_p = 1\) if \(V(n)\) is divisible by \(p\) and \(Z_p = 0\) otherwise. Then, for any strongly additive function \(g\), we have \(g(V(n)) = \sum_{p \leq n, p \in \mathcal{P}} g(p) Z_p\). We have the following large deviations result.

**Theorem 2.** Under Assumption 1, \(\frac{X(n)}{\log \log n} \in \cdot\) satisfies a large deviation principle with speed \(\log \log n\) and rate function

\[
I(x) := \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \int_{\mathbb{R}} (e^{\theta y} - 1) \rho(dy) \right\}.
\]

(2.2)
A celebrated result in number theory due to Hoheisel says that there exists a constant \( \gamma < 1 \) such that \( p_{k+1} - p_k < p_k^\gamma \) for sufficiently large \( k \). This shows that

\[
\sum_{p_k \leq n} \left( \frac{1}{p_k} - \frac{1}{p_{k+1}} \right) = \sum_{p_k \leq n} \frac{p_{k+1} - p_k}{p_k p_{k+1}} \leq C \sum_{p_k \leq n} \frac{1}{p_k^{2-\gamma}}
\]
for some universal constant $C$ and the series is convergent. Therefore, we have

$$\sum_{p_{2k} \leq n} \frac{1}{p_{2k}} \sim \frac{1}{2} \log n$$

and

$$\sum_{p_{2k+1} \leq n} \frac{1}{p_{2k+1}} \sim \frac{1}{4} \log n.$$  Let $0 < \lambda_1 < \lambda_2 < \infty$. Define $g(p_k) = \lambda_1$ if $k$ is odd and $g(p_k) = \lambda_2$ if $k$ is even. Thus, we have

$$\lim_{n \to \infty} \frac{1}{\log \log n} \sum_{p \leq n, p \text{ has odd } \nu} g(p) X(p) = \frac{1}{2} (e^{\theta \lambda_1} - 1) + \frac{1}{2} (e^{\theta \lambda_2} - 1).$$

Hence, we conclude that $(\frac{X(n)}{\log \log n} \in \cdot)$ satisfies a large deviation principle with speed $\log \log n$ and rate function

$$I(x) := \left\{ \begin{array}{ll}
\sup_{\theta \in \mathbb{R}} \left\{ \theta x - \frac{1}{2} (e^{\theta \lambda_1} - 1) - \frac{1}{2} (e^{\theta \lambda_2} - 1) \right\} & \text{if } x \geq 0, \\
+\infty & \text{otherwise.}
\end{array} \right.$$\n
In some special cases, there is an explicit expression for the rate function. For example, consider $\lambda_2 = 2 \lambda_1 \in (0, \infty)$. In this case, the optimal $\theta$ is given by

$$\theta_* = \frac{1}{\lambda_1} \log \left( \frac{-\lambda_1 + \sqrt{\lambda_1^2 + 16 \lambda_1 x}}{4 \lambda_1} \right).$$

Hence, for $x \geq 0$,

$$I(x) = \theta_* x - \frac{1}{2} (e^{\theta_\lambda_1} - 1) - \frac{1}{2} (e^{\theta_\lambda_2} - 1)$$

$$= \frac{x}{\lambda_1} \log \left( \frac{-\lambda_1 + \sqrt{\lambda_1^2 + 16 \lambda_1 x}}{4 \lambda_1} \right) + 1$$

$$- \frac{1}{2} \left( \frac{-\lambda_1 + \sqrt{\lambda_1^2 + 16 \lambda_1 x}}{4 \lambda_1} \right) - \frac{1}{2} \left( \frac{-\lambda_1 + \sqrt{\lambda_1^2 + 16 \lambda_1 x}}{4 \lambda_1} \right)^2.$$\n
Here are some examples in which we can get an explicit expression for the rate function $I(x)$.

**Example 5.** (i) Assume that $\rho(dy) = \frac{1}{2} \delta_{\lambda_1} + \frac{1}{2} \delta_{\lambda_2}$. Then, from Remark 4 (iii), for $x \geq 0$, the rate function $I(x)$ has explicit expression as given in (2.13).

(ii) Assume that $\rho(dy)$ has Poisson distribution with parameter $\lambda > 0$. Then, we have

$$\theta_\lambda x - \int_0^\infty (e^{\theta_\lambda y} - 1) \rho(dy) = \theta_\lambda x + 1 - e^{\lambda (e^{\theta_\lambda} - 1)}.$$\n
Differentiating with respect to $\theta$ and setting the derivative as zero, we get $x = \lambda e^{\theta_\lambda}$. Then, we get the optimal $\theta_* = \log(\frac{1}{\lambda} W(x e^\lambda))$, where $W(z)$ is the Lambert $W$ function defined as $z = W(z) e^{W(z)}$ for any $z \in \mathbb{C}$, see e.g. Corless et al. [1]. Hence, for $x \geq 0$, we have

$$I(x) = x \log \left( \frac{1}{\lambda} W(x e^\lambda) \right) + 1 - e^{W(x e^\lambda) - \lambda}.$$\n
(iii) Assume that $\rho(dy)$ has binomial distribution with parameters $n$ and $\beta$. Then, we get $\int_0^\infty e^{\theta y} \rho(dy) = (1 - \beta + \beta e^{\theta})^n$. Thus, the optimal $\theta_\lambda$ satisfies $x = n (1 - \beta + \beta e^{\theta})^{n-1} \beta e^{\theta_*}$. For $n = 1$, $I(x) = x \log(x/\beta) + \beta - x$. For $n = 2$,

$$I(x) = x \log \left( \frac{-(1 - \beta) + \sqrt{(1 - \beta) + x}}{\beta} \right) + 1 - (1 - \beta)^2 - x^2.$$
Then, we get

\begin{equation}
I(x) = \sqrt{W(x^2)}x + 1 - \frac{x}{\sqrt{W(x^2)}}.
\end{equation}

We conclude this section by stating a moderate rate deviation principle, which fills in the gap between central limit theorem and large deviation principle.

**Theorem 6.** Let \( \mu_n := \sum_{p \leq n} g(p) \) and \( \sigma_n^2 := \sum_{p \leq n} g(p)^2 \). Let \( a_n \) be a positive sequence so that \( \sigma_n \ll a_n \ll \sigma_n^2 \) as \( n \to \infty \). Under Assumption 1, \( \frac{X(n) - \mu_n}{a_n} \) satisfies a large deviation principle with speed \( a_n^2/\sigma_n^2 \) and rate function

\begin{equation}
J(x) := \frac{x^2}{2}.
\end{equation}

3. Proofs

3.1. **Proof of Large Deviation Principle.** In this section, we will prove Theorem 2. The proof consists of a series of superexponential estimates, i.e. Lemma 4, Lemma 5 and Lemma 6, and an application of Gärtner-Ellis theorem (see e.g. Chapter 2 in Dembo and Zeitouni [2]).

Write \( X(n) = \sum_{p \leq n, p \in \mathcal{P}} g(p)Z_p \), where \( Z_p = 1 \) if \( V(n) \) is divisible by \( p \) and 0 otherwise. The first step is to show that \( \sum_{p \leq n, p \in \mathcal{P}} g(p)Z_p \) can be approximated by \( \sum_{|g(p)| \leq C, p \in \mathcal{P}} g(p)Z_p \) in the following sense.

**Lemma 7.** for any \( \epsilon > 0 \),

\begin{equation}
\lim_{C \to \infty} \limsup_{n \to \infty} \frac{1}{\log \log n} \log P \left( \sum_{|g(p)| > C, p \leq n} g(p)Z_p \geq \epsilon \log \log n \right) = -\infty.
\end{equation}

**Proof.** Note that \( \text{(3.1)} \) holds if we can prove the following two estimates,

\begin{equation}
\limsup_{C \to \infty} \limsup_{n \to \infty} \frac{1}{\log \log n} \log P \left( \sum_{|g(p)| > C, p \leq n} g(p)Z_p \geq \epsilon \log \log n \right) = -\infty,
\end{equation}

and

\begin{equation}
\limsup_{C \to \infty} \limsup_{n \to \infty} \frac{1}{\log \log n} \log P \left( \sum_{|g(p)| < -C, p \leq n} g(p)Z_p \leq -\epsilon \log \log n \right) = -\infty.
\end{equation}

Before we proceed, let us define independent random variables \( Y_p, p \in \mathcal{P}, \) so that

\begin{equation}Y_p = \begin{cases} 1 & \text{with probability } \frac{1}{p}, \\ 0 & \text{with probability } 1 - \frac{1}{p}. \end{cases} \end{equation}

For distinct primes \( p_1, p_2, \ldots, p_\ell, \)

\begin{equation}
E[Z_1Z_2 \cdots Z_{p_\ell}] = \frac{1}{n} \left[ \frac{n}{p_1p_2 \cdots p_\ell} \right] = \frac{1}{p_1p_2 \cdots p_\ell} = E[Y_{p_1}Y_{p_2} \cdots Y_{p_\ell}] = E[Y_p].
\end{equation}
Therefore, for any non-negative sequence $\theta_p$, by Taylor’s expansion, we have

\begin{equation}
E\left[e^{\sum_{p \leq n, p \in P} \theta_p Z_p}\right] \leq E\left[e^{\sum_{p \leq n, p \in P} \theta_p Y_p}\right].
\end{equation}

This fact will be used repeatedly later on in the paper.

For $g(p) > C > 0$, for any $\theta > 0$, by Chebychev’s inequality, we have $g(p)\theta > 0$ and

\begin{equation}
\limsup_{n \to \infty} \frac{1}{\log \log n} \log P\left(\sum_{g(p) > C, p \leq n, p \in P} g(p) Z_p \geq \epsilon \log \log n\right)
\end{equation}

\begin{equation}
\leq \limsup_{n \to \infty} \frac{1}{\log \log n} \log E\left[e^{\theta \sum_{g(p) > C, p \leq n, p \in P} g(p) Z_p}\right] - \theta \epsilon
\end{equation}

\begin{equation}
\leq \limsup_{n \to \infty} \frac{1}{\log \log n} \log E\left[e^{\theta \sum_{g(p) > C, p \leq n, p \in P} g(p) Y_p}\right] - \theta \epsilon
\end{equation}

\begin{equation}
= \int_{y > C} (e^{\theta y} - 1) \rho(dy) - \theta \epsilon,
\end{equation}

and the limit goes to $-\theta \epsilon$ as $C \to \infty$. By letting $\theta \to \infty$, we obtained (3.2).

Similarly, for $g(p) < -C < 0$, by taking $\theta < 0$, we get

\begin{equation}
\limsup_{n \to \infty} \frac{1}{\log \log n} \log P\left(\sum_{g(p) < -C, p \leq n, p \in P} g(p) Z_p \leq -\epsilon \log \log n\right)
\end{equation}

\begin{equation}
\leq \int_{y < -C} (e^{\theta y} - 1) \rho(dy) + \theta \epsilon,
\end{equation}

where the limit goes to $\theta \epsilon$ as $C \to \infty$. By letting $\theta \to -\infty$, we obtained (3.3). \hfill \square

Let

\begin{equation}
k_n := n^\frac{1}{\log \log n^{\epsilon}}.
\end{equation}

The second step is to show that $\sum_{p \leq n, |g(p)| \leq C, p \in P} g(p) Z_p$ can be approximated by $\sum_{p \leq k_n, |g(p)| \leq C, p \in P} g(p) Z_p$ in the following sense.

**Lemma 8.** For any $\epsilon > 0$,

\begin{equation}
\limsup_{n \to \infty} \frac{1}{\log \log n} \log P\left(\left|\sum_{p \in A(n, C)} g(p) Z_p\right| \geq \epsilon \log \log n\right) = -\infty,
\end{equation}

where $A(n, C) := \{p : k_n \leq p \leq n, |g(p)| \leq C, p \in P\}$.

**Proof.** For any $\theta > 0$,

\begin{equation}
E\left[e^{\theta \sum_{p \in A(n, C)} g(p) Z_p}\right] \leq E\left[e^{\theta \sum_{p \in A(n, C)} |g(p)| Z_p}\right]
\end{equation}

\begin{equation}
\leq E\left[e^{\theta C \sum_{p \in A(n, C)} Z_p}\right]
\end{equation}

\begin{equation}
\leq E\left[e^{\theta C \sum_{p \in A(n, C)} Y_p}\right].
\end{equation}
Therefore, for any $\theta > 0$, we have
\begin{align}
\log E \left[ e^{\theta |\sum_{p \in A(n,C)} g(p)Z_p|} \right] & \leq \log E \left[ e^{\theta C \sum_{p \in A(n,C)} Y_p} \right] \\
& = \sum_{p \in A(n,C)} \log \left( (e^{\theta C} - 1) \frac{1}{p} + 1 \right) \\
& \leq (e^{\theta C} - 1) \sum_{k_n \leq p \leq n, p \in \mathcal{P}} \frac{1}{p}.
\end{align}

We also have
\begin{align}
\sum_{k_n \leq p \leq n, p \in \mathcal{P}} \frac{1}{p} & \sim \log \log n - \log \log k_n \\
& = \log \log n - \log \left( \frac{1}{(\log \log n)^2} \log n \right) \\
& = 2 \log \log n,
\end{align}

and $\frac{2 \log \log n}{\log n} \to 0$ as $n \to \infty$. Therefore, by Chebychev's inequality,
\begin{align}
\limsup_{n \to \infty} \frac{1}{\log \log n} \log \mathbb{P} \left( \left| \sum_{p \in A(n,C)} g(p)Z_p \right| \geq \epsilon \log \log n \right) & \leq -\epsilon \theta.
\end{align}

This proves (3.10) since it holds for any $\theta > 0$. \hfill \Box

Next, let us show that $E \left[ e^{\theta \sum_{p \leq k_n, |g(p)| \leq C, p \in \mathcal{P}} g(p)Z_p} \right]$ can be approximated by $E \left[ e^{\theta \sum_{p \leq k_n, |g(p)| \leq C, p \in \mathcal{P}} g(p)Y_p} \right]$ in an appropriate way.

**Lemma 9.** For any $\theta \in \mathbb{R}$,
\begin{align}
\lim_{n \to \infty} \frac{1}{\log \log n} \log \mathbb{E} \left| e^{\theta \sum_{p \in B(n,C)} g(p)Z_p} - e^{\theta \sum_{p \in B(n,C)} g(p)Y_p} \right| & = -\infty,
\end{align}

where $B(n,C) := \{ p : p \leq k_n, |g(p)| \leq C, p \in \mathcal{P} \}$.

**Proof.** For any $\theta \in \mathbb{R}$,
\begin{align}
\left| E \left[ e^{\theta \sum_{p \in B(n,C)} g(p)Z_p} \right] - E \left[ e^{\theta \sum_{p \in B(n,C)} g(p)Y_p} \right] \right| & \leq \sum_{r \leq K \log \log n} \frac{|\theta|^r}{r!} E \left[ S^n_r \right] + \sum_{r > K \log \log n} \frac{|\theta|^r}{r!} E \left[ S^n_r \right] + \sum_{r > K \log \log n} \frac{|\theta|^r}{r!} E \left[ \tilde{S}^n_r \right],
\end{align}

where
\begin{align}
S_n := \sum_{p \in B(n,C)} g(p)Z_p, \quad \tilde{S}_n := \sum_{p \in B(n,C)} g(p)Y_p.
\end{align}

We claim that $|E[\tilde{S}^n_r] - E[S^n_r]| \leq \frac{(Ck_n)^r}{n}$. To see this, notice that
\begin{align}
E[\tilde{S}^n_r] & = \sum_{k=1}^r \sum_{r_1! \cdots r_k!} \frac{1}{k!} \sum_{p \leq n} g(p_1)^{r_1} \cdots g(p_k)^{r_k} E[Y_{p_1}^{r_1} \cdots Y_{p_k}^{r_k}] .
\end{align}
We observe that
\begin{equation}
\mathbb{E}[Y_{p_1}^r \cdots Y_{p_k}^r] = \mathbb{E}[Y_{p_1} \cdots Y_{p_k}] = \frac{1}{p_1 \cdots p_k},
\end{equation}
which differs from
\begin{equation}
\mathbb{E}[Z_{p_1}^r \cdots Z_{p_k}^r] = \mathbb{E}[Z_{p_1} \cdots Z_{p_k}] = \frac{1}{n} \left[ \frac{n}{p_1 \cdots p_k} \right],
\end{equation}
by at most \( \frac{1}{n} \). Therefore,
\begin{equation}
|\mathbb{E}[\tilde{S}_n^r] - \mathbb{E}[S_n^r]| \leq \sum_{k=1}^{r} \sum_{r_1} \frac{r!}{r_1! \cdots r_k!} \frac{1}{p_1} \cdots \frac{1}{p_k} \sum_{r_1 + \cdots + r_k}
\end{equation}
\begin{equation}
\leq \frac{1}{n} \left( \sum_{p \in B(n, C)} C \right)^r \leq \frac{(Ck_n)^r}{n}.
\end{equation}

Thus, we can bound the first term in (3.16) by
\begin{equation}
\sum_{r \leq K \log \log n} \frac{\left| \theta \right|^r}{r!} |\mathbb{E}[S_n^r] - \mathbb{E}[\tilde{S}_n^r]| \leq C_0 e^{-\log n} \sum_{r \leq K \log \log n} e^{r(\log |\theta| + \log C + \log k_n + 1) - r \log r} \leq C_0 e^{-\log n} K \log \log n \cdot \max_{r \leq K \log \log n} e^{r(\log |\theta| + \log C + \log k_n + 1) - r \log r}.
\end{equation}

Let \( F(r) := r(\log |\theta| + \log C + \log k_n + 1) - r \log r \). Since \( k_n = n^{1/(\log \log n)} \gg r \), it is straightforward to compute that
\begin{equation}
F'(r) = (\log |\theta| + \log C + \log k_n + 1) - \log r - 1 > 0,
\end{equation}
for any \( r \leq K \log \log n \). Therefore the maximum of \( F(r) \), \( r \leq K \log \log n \) is achieved at \( K \log \log n \) and, hence,
\begin{equation}
\sum_{r \leq K \log \log n} \frac{\left| \theta \right|^r}{r!} |\mathbb{E}[S_n^r] - \mathbb{E}[\tilde{S}_n^r]| \leq C_1 \log \log n \cdot e^{-\log n} \cdot e^{K \log \log n \frac{\log n}{(\log \log n)^2}} \leq e^{-\frac{1}{2} \log n},
\end{equation}
for sufficiently large \( n \).

The second term in (3.16) is bounded above by the third term.
\begin{equation}
\sum_{r > K \log \log n} \frac{\left| \theta \right|^r}{r!} \mathbb{E}[S_n^r] \leq \sum_{r > K \log \log n} \frac{\left| \theta \right|^r}{r!} \mathbb{E}[S_n^r].
\end{equation}

To bound the third term in (3.16), first observe that
\begin{equation}
\mathbb{E}[\tilde{S}_n^r] = \sum_{p_1, p_2, \ldots, p_r \leq k_n} g(p_1) \cdots g(p_r) \mathbb{E}[Y_{p_1} \cdots Y_{p_r}],
\end{equation}
where the sums are over primes $p_1, \ldots, p_\ell \leq k_n$ that may not be distinct. Notice that $k_n \leq n$ and for distinct $p_1, \ldots, p_\ell$, we have

$$\mathbb{E}[Y_{p_1} \cdots Y_{p_\ell}] = \frac{1}{p_1 \cdots p_\ell}.$$  

Therefore, it is not difficult to see that

$$\mathbb{E}[\tilde{S}_n] \leq C^r \left( \sum_{p \leq n, p \in \mathcal{P}} \frac{1}{p} \right) + \left( \sum_{p \leq n, p \in \mathcal{P}} \frac{1}{p} \right)^2 + \cdots + \left( \sum_{p \leq n, p \in \mathcal{P}} \frac{1}{p} \right)^r.$$  

(3.28)

For $r > K \log \log n$,

$$\frac{\theta^r}{r!} \mathbb{E}[\tilde{S}_n^r] \simeq e^{r \log |\theta| + r \log C - r \log r + \log r + (\log \log \log n) r} \leq e^{(\log |\theta| + \log C + 1 - \log K) r + \log r} \leq e^{-\frac{1}{2} (\log K)^r},$$

for sufficiently large $K$. Hence, the third term in (3.16) is bounded above by

$$\sum_{r > K \log \log n} \frac{1}{r} e^{-\frac{r}{2} (\log K)^r} \leq C_2 e^{-\frac{1}{2} \log K \log \log n},$$

where $C_2$ is a positive constant. The proof is completed by letting $K \to \infty$.  

Finally, we are ready to prove Theorem 2.

**Proof of Theorem 2.** Recall that $B(n, c) = \{ p : p \in \mathcal{P}, |g(p)| \leq C, p \leq k_n \}$. For any $\theta \in \mathbb{R}$, we have

$$\log \mathbb{E}[e^{\theta \sum_{p \in B(n, C)} g(p) Y_p}] = \log \prod_{p \in B(n, C)} \mathbb{E}[e^{\theta g(p) Y_p}] = \sum_{p \in B(n, C)} \log \left( \frac{1}{p} e^{\theta g(p)} + 1 - \frac{1}{p} \right).$$

(3.31)

For sufficiently large $p$,

$$\log \left( \frac{1}{p} e^{\theta g(p)} + 1 - \frac{1}{p} \right) + (e^{\theta \lambda} - 1) \frac{1}{p} + O \left( \frac{1}{p^2} \right),$$

(3.32)

and it is well known that $\frac{1}{\log \log n} \sum_{p \in \mathcal{P}, p \leq n} \frac{1}{p} \to 1$ as $n \to \infty$, and we also have

$$\lim_{n \to \infty} \frac{\log \log k_n}{\log \log n} = \lim_{n \to \infty} \frac{\log \log n - 2 \log \log n}{\log \log n} = 1.$$  

(3.33)

Therefore, by Assumption 1 and definition of $\rho_n$ and $\rho$

$$\lim_{n \to \infty} \frac{1}{\log \log n} \log \mathbb{E}[e^{\theta \sum_{p \in B(n, C)} g(p) Y_p}] = \int_{-C}^C (e^{\theta y} - 1) \rho(dy),$$

if $\rho(-C) = \rho(C) = 0$ since convergence of moment generating functions implies weak convergence. Here we can assume that $\rho(-C) = \rho(C) = 0$ since there
are at most countably many atoms for \( \rho \), we can assume that we choose a sequence of \( C \) that goes to infinity so that \( -C \) and \( C \) are not atoms of \( \rho \).

By Lemma 9 we have

\[
(3.35) \quad \lim_{n \to \infty} \frac{1}{\log \log n} \log \mathbb{E} \left[ e^{\theta \sum_{p \in B(n,C)} g(p)Z_p} \right] = \lim_{n \to \infty} \frac{1}{\log \log n} \log \mathbb{E} \left[ e^{\theta \sum_{p \in B(n,C)} g(p)Y_p} \right]
\]

\[= \int_C (e^{\theta y} - 1)\rho(dy).\]

By Gärner-Ellis theorem, see e.g. Dembo and Zeitouni [2], \( \left( \frac{\sum_{p \in B(n,C)} g(p)Z_p}{\log \log n} \right) \in \cdot \) satisfies a large deviation principle with the rate function

\[
(3.36) \quad I_C(x) = \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \int_{-C}^{C} (e^{\theta y} - 1)\rho(dy) \right\}.
\]

Hence, by the approximation estimates developed in Lemma 7 and Lemma 8 \( \left( \frac{X(n)}{\log \log n} \in \cdot \right) \) satisfies a large deviation principle with the rate function

\[
(3.37) \quad I(x) = \lim_{C \to \infty} I_C(x) = \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \int_{-\infty}^{\infty} (e^{\theta y} - 1)\rho(dy) \right\}.
\]

\[
\square
\]

3.2. Proof of Moderate Deviation Principle. We conclude this section by giving a proof of the moderate deviation principle.

Proof of Theorem 2. Recall that \( Y_p \) are independent Bernoulli random variables with parameter \( \frac{1}{p} \) and \( Z_p = 1 \) if \( V(n) \) is divisible by \( p \) and \( Z_p = 0 \) otherwise, where \( V(n) \) is an integer uniformly distributed on \( \{1, 2, \ldots, n\} \).

Let

\[
(3.38) \quad k_n := n^{\frac{\epsilon}{p_n}}.
\]

Similar to the proof of Lemma 7 we can show that, for any \( \epsilon > 0 \),

\[
(3.39) \quad \limsup_{C \to \infty} \limsup_{n \to \infty} \frac{\sigma^2}{a_n^2} \log \mathbb{P} \left( \left| \sum_{g(p) > C, p \in \mathcal{P}} g(p)Z_p \right| \geq \epsilon a_n \right) = -\infty,
\]

and also similar to the proof of Lemma 8 we can show that, for any \( \epsilon > 0 \),

\[
(3.40) \quad \limsup_{n \to \infty} \frac{\sigma^2}{a_n^2} \log \mathbb{P} \left( \sum_{k_n < p \leq n, |g(p)| \leq C, p \in \mathcal{P}} g(p)Z_p \geq \epsilon a_n \right) = -\infty.
\]

Let us define

\[
(3.41) \quad S_n = \sum_{p \in B(n,C)} g(p)Z_p, \quad \text{and} \quad \bar{S}_n = \sum_{p \in B(n,C)} g(p)Y_p,
\]

where we recall that \( B(n, C) = \{ p : p \in \mathcal{P}, p \leq k_n, |g(p)| \leq C \} \).

Let

\[
(3.42) \quad \mu_n^C := \mathbb{E}[\bar{S}_n] = \sum_{p \in B(n,C)} \frac{g(p)}{p},
\]
and recall that
\begin{equation}
\sigma_n^2 := \sum_{p \in P, p \leq n} \frac{g(p)^2}{p}.
\end{equation}

Following the proof of Lemma 11, for any $\theta \in \mathbb{R}$, we can also prove that
\begin{equation}
\lim_{n \to \infty} \frac{\sigma_n^2}{\sigma_n^2 \log \mathbb{E}[e^{\frac{\theta}{\sigma_n^2}(S_n-\mu_C^n)}]} = -\infty.
\end{equation}

Finally, for any $\theta \in \mathbb{R}$,
\begin{equation}
\frac{\sigma_n^2}{\sigma_n^2 \log \mathbb{E}[e^{\frac{\theta}{\sigma_n^2}(S_n-\mu_C^n)}]} = -\frac{\theta}{\sigma_n^2 \mu_n^C} + \frac{\sigma_n^2 \rho}{\sigma_n^2} \sum_{p \in B(n,C)} \log \left( \frac{\frac{\theta}{\sigma_n^2}g(p)}{e^{\frac{\theta}{\sigma_n^2}}} - 1 \right) \frac{1}{p} + 1
\end{equation}

Therefore, by Assumption 11,
\begin{equation}
\mathbb{E}[e^{\frac{\theta}{\sigma_n^2}(S_n-\mu_C^n)}] = \theta^2 \int_{-C}^{C} y^2 \rho(dy) + o(1).
\end{equation}

By letting $C \to \infty$ and applying Gärtner-Ellis theorem, we complete the proof. $\square$

Acknowledgements

The authors are very grateful to Professor S. R. S. Varadhan for helpful discussions and generous suggestions.

References

[1] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. and D. E. Knuth. (1996). On the Lambert W function. Advances in Computational Mathematics 5, 329-359.
[2] Dembo, A. and O. Zeitouni. Large Deviations Techniques and Applications. Springer-Verlag, 2nd Edition, New York, 1998.
[3] Donsker, M. D. and S. R. S. Varadhan. (1975). Asymptotic evaluation of certain Markov process expectations for large time, I. Comm. Pure Appl. Math 28, 1-47.
[4] Donsker, M. D. and S. R. S. Varadhan. (1975). Asymptotic evaluation of certain Markov process expectations for large time, II. Comm. Pure Appl. Math 28, 279-301.
[5] Donsker, M. D. and S. R. S. Varadhan. (1976). Asymptotic evaluation of certain Markov process expectations for large time, III. Comm. Pure Appl. Math 29, 389-461.
[6] Donsker, M. D. and S. R. S. Varadhan. (1983). Asymptotic evaluation of certain Markov process expectations for large time, IV. Comm. Pure Appl. Math 36, 183-212.
[7] Durrett, R. Probability: Theory and Example Cambridge University Press, 4th edition, 2010.
[8] Erdős, P. (1937). Note on the number of prime divisors of integers. J. London Math. Soc. 12, 308-314.
[9] Erdős, P. and M. Kac. (1939). On the Gaussian law of errors in the theory of additive functions. *Proc. Natl. Acad. Sci. USA* **25**, 206-207.

[10] Erdős, P. and M. Kac. (1940). The Gaussian law of errors in the theory of additive number theoretic functions. *American Journal of Mathematics* **62**, 738-742.

[11] Erdős, P. and A. Sárközy. (1980). On the number of prime factors of integers. *Acta Sci. Math.* **42**, 237-246.

[12] Féray, V., Méliot, P.-L., and A. Nikeghbali. (2013). Mod-$\phi$ convergence and precise deviations. *arXiv:1304.2934*

[13] Hardy, G. H. and S. Ramanujan. (1917). The normal number of prime factors of a number $n$. *Quarterly Journal of Mathematics* **48**, 76-92.

[14] Harper, A. J. (2009). Two new proofs of the Erdős-Kac Theorem, with bound on the rate of convergence, by Stein’s method for distributional approximations. *Math. Proc. Camb. Phil. Soc.* **147**, 95-114.

[15] Hoheisel, G. (1930). Primzahlprobleme in der Analysis. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* **33**, 3-11.

[16] Kowalski, E. and A. Nikeghbali. (2010). Mod-Poisson convergence in probability and number theory. *Intern. Math. Res. Not.* **18**, 3549-3587.

[17] Radziwill, M. (2009). On large deviations of additive functions. *arXiv:0909.5274*.

[18] Rényi, A. and P. Turán. (1958). On a theorem of Erdős-Kac. *Acta Arith.* **4**, 71-84.

[19] Varadhan, S. R. S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19**, 261-286.

[20] Varadhan, S. R. S. *Large Deviations and Applications*. SIAM, Philadelphia, 1984.

**Courant Institute of Mathematical Sciences**
**New York University**
**251 Mercer Street**
**New York, NY-10012**
**United States of America**

*E-mail address: mehrdad@cims.nyu.edu, ling@cims.nyu.edu*