Large Deviations for Erdős-Kac Theorem

Lingjiong Zhu

University of Minnesota-Twin Cities
Email: zhul@umn.edu

Frontier Probability Days
University of Arizona, Tucson
18-20 May 2014
Joint work with Behzad Mehrdad

Erdős-Kac Theorem

A celebrated result in number theory due to Erdős and Kac says, for a uniformly chosen random integer $\omega(n)$ from $\{1, 2, \ldots, n\}$, let $X(n)$ be the number of the distinct prime factors of $\omega(n)$, then, we have the following central limit theorem,

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

in distribution as $n \to \infty$.

- Where does $\log\log n$ come from?
- In the "ideal world", a uniformly chosen integer is divisible by a prime p with probability $\frac{1}{p}$.
- It is well known that $\sum_{p\in\mathcal{P},p\leq n}\frac{1}{p}\simeq\log\log n$ as $n\rightarrow\infty$.
- In the "real world", a uniformly chosen integer from $\{1,2,\ldots,n\}$ is divisible by p with probability $\frac{\lfloor n/p\rfloor}{n}$.
- The main difficulty is to convince ourselves that the results in the "real world" are comparable with those in the "ideal world".

Indeed, the central limit theorem holds in a more general setting for strongly additive functions.

Let g be strongly additive, i.e. $g(p^k) = g(p)$ for all prime p and $k \in \mathbb{N}$ and $g(mn) = g(m) + g(n)$ whenever $\gcd(m, n) = 1$. Then, under certain additional assumptions of $g(\cdot)$,

$$\frac{X(n) - \mu(g, n)}{\sigma(g, n)} \to N(0, 1), \tag{2}$$

where $X(n) = g(\omega(n)) = \sum_{p \in \mathcal{P}, p \leq n} g(p) Z_p$, where $\mathcal{P}$ stands for the set of all the prime numbers and $Z_p = 1$ if $\omega(n)$ is divisible by p and $Z_p = 0$ otherwise and

$$\mu(g, n) := \sum_{p \in \mathcal{P}, p \leq n} \frac{g(p)}{p}, \quad \sigma^2(g, n) := \sum_{p \in \mathcal{P}, p \leq n} \frac{g(p)^2}{p} \left(1 - \frac{1}{p}\right). \tag{3}$$

The rate of convergence to the Gaussian distribution, i.e. Berry-Esseen bounds, have been obtained in Rényi and Turán, Harper, etc.

Recently, Radziwill used analytic number theory approach and obtained a series of large deviations estimates. Féray et al. proved precise large deviations using the mod-Poisson convergence method developed in Kowalski and Nikeghabli.

In this talk, we are interested to study the Donsker-Varadhan type large deviations. Our methods are probabilistic and our results hold for a much wider class of additive functions $g(\cdot)$.

Before we proceed, let us recall that a sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ on a topological space X satisfies a 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,

$$
\begin{aligned}
-\inf_{x\in A^o} I(x) &\leq \liminf_{n\to\infty} \frac{1}{b_n} \log P_n(A) \\
&\leq \limsup_{n\to\infty} \frac{1}{b_n} \log P_n(A) \leq -\inf_{x\in \overline{A}} I(x). \qquad (4)
\end{aligned}
$$

Theorem (Mehrdad and Z. 2013)

Let g be a strongly additive function and assume that there exists a probability measure $\rho(dy)$ on $\mathbb{R}$ so that for any $\theta \in \mathbb{R}$, $\int_{\mathbb{R}} e^{\theta y}\rho(dy) < \infty$ and $\int_{\mathbb{R}} e^{\theta y}\rho_n(dy) \to \int_{\mathbb{R}} e^{\theta y}\rho(dy)$, where

$$\rho_n(A) := \frac{\sum_{g(p)\in A, p\leq n, p\in\mathcal{P}} \frac{1}{p}}{\sum_{p\leq n, p\in\mathcal{P}} \frac{1}{p}}, \tag{5}$$

for any Borel set $A \subseteq \mathbb{R}$. Then, $\mathbb{P}(\frac{X(n)}{\log\log n} \in \cdot)$ satisfies a large deviation principle with speed $\log\log n$ and rate function

$$I(x) := \sup_{\theta\in\mathbb{R}} \left\{\theta x - \int_{\mathbb{R}} (e^{\theta y} - 1)\rho(dy)\right\}. \tag{6}$$

Corollary

Let $X(n)$ *be the number of the distinct prime factors of a uniformly chosen integer from* $\{1, 2, \ldots, n\}$, *then* $\mathbb{P}(\frac{X(n)}{\log\log n} \in \cdot)$ *satisfies a large deviation principle with speed* $\log\log n$ *and rate function*

$$I(x) := \begin{cases} x\log x - x + 1 & \textit{if } x \geq 0, \\ +\infty & \textit{otherwise.} \end{cases} \tag{7}$$

Sketch of the Proof

The proof consists of a series of superexponential estimates and the approximation by the independent model.

- Approximate $\sum_{p\in\mathcal{P},p\leq n} g(p)Z_p$ by $\sum_{p\in\mathcal{P},p\leq n,|g(p)|\leq C} g(p)Z_p$.
- Approximate $\sum_{p\in\mathcal{P},p\leq n,|g(p)|\leq C} g(p)Z_p$ by $\sum_{p\in\mathcal{P},p\leq k_n,|g(p)|\leq C} g(p)Z_p$, where $k_n = n^{\frac{1}{(\log\log n)^2}}$.
- Approximate $\sum_{p\in\mathcal{P},p\leq k_n,|g(p)|\leq C} g(p)Z_p$ by $\sum_{p\in\mathcal{P},p\leq k_n,|g(p)|\leq C} g(p)Y_p$, where Y_p are independent random variables so that $Y_p = 1$ with probability $\frac{1}{p}$ and $Y_p = 0$ with probability $1-\frac{1}{p}$.
- Finally, establish a large deviation principle for $\left(\frac{\sum_{p\in\mathcal{P},p\leq k_n,|g(p)|\leq C} g(p)Y_p}{\log\log n}\in\cdot\right)$ with rate function $I_C(x)$ and $\lim_{C\to\infty} I_C(x) = I(x)$.

Selberg's Central Limit Theorem

Let t be uniformly distributed on $(T, 2T)$. Then,

$$\frac{\log|\zeta(\frac{1}{2}+it)|}{\sqrt{\frac{1}{2}\log\log T}} \to N(0,1), \tag{8}$$

in distribution as $T \to \infty$.

- There is a conjecture on the precise large deviations for Selberg's Central Limit Theorem.
- Estimates for positive moments of Riemann Zeta function are obtained in Soundararajan, Harper etc. (assuming the Riemann Hypothesis).
- The proof of Selberg's Central Limit Theorem does not assume the Riemann Hypothesis.