The strong law of large numbers and computability

M. Neri

Department of Computer science
University of Bath, Bath, United Kingdom
kejineri.github.io

Large deviations in the strong law of large numbers

Let $X_1, X_2, \ldots$ are pairwise independent identically distributed real-valued random variables with 0 expected value and $\mathbb{E}(X_1) < \infty$. Define $S_n = \sum_{i=1}^{n} X_i$. The strong law of large numbers states that,

$$\frac{S_n}{n} \to 0$$

almost surely.

Large deviations in the strong law of large numbers

$$\frac{S_n}{n} \to 0 \text{ almost surely} \iff \forall \varepsilon \, P_{n,\varepsilon} \to 0$$

where,

$$P_{n,\varepsilon} = \mathbb{P}(\max_{m \geq n} |\frac{1}{n} S_n| > \varepsilon)$$

Large deviations in the strong law of large numbers

- The quantitative content of this theorem has been studied extensively in the probability literature, in the form of studying the asymptotics of P_{n,ε_n}, for certain sequences of real numbers $\{\varepsilon_n\}$

Large deviations in the strong law of large numbers

- The quantitative content of this theorem has been studied extensively in the probability literature, in the form of studying the asymptotics of P_{n,ε_n}, for certain sequences of real numbers $\{\varepsilon_n\}$
- Examples of results in this space include the works of Strassen, Siegmund and Fill (1967, 1975 and 1983 respectfully) who calculated P_{n,ε_n} up to asymptotic equivalence to simple functions, for various classes of sequences $\{\varepsilon_n\}$ (including a constant sequence) under strong assumptions about $\{X_n\}$. Furthermore, their bounds depend heavily on the distribution of the random variable.

Large deviations in the strong law of large numbers

- The quantitative content of this theorem has been studied extensively in the probability literature, in the form of studying the asymptotics of P_{n,ε_n}, for certain sequences of real numbers $\{\varepsilon_n\}$
- Examples of results in this space include the works of Strassen, Siegmund and Fill (1967, 1975 and 1983 respectfully) who calculated P_{n,ε_n} up to asymptotic equivalence to simple functions, for various classes of sequences $\{\varepsilon_n\}$ (including a constant sequence) under strong assumptions about $\{X_n\}$. Furthermore, their bounds depend heavily on the distribution of the random variable.
- In addition, in 2018 Luzia was able to produce an upper bound for $P_{n,\varepsilon}$ that does not depend on the distribution of the random variables, for $0 < \varepsilon \leq 1$ by only assuming further that $Var(X_1) < \infty$.

Computability in the strong law of large numbers

- In 2010 Gács demonstrates how the almost sure convergence of $\frac{S_n}{n} \to 0$ is effective given X_1 has a computable distribution and $\mathbb{E}(|X_1|)$ is computable.

Computability in the strong law of large numbers

- In 2010 Gács demonstrates how the almost sure convergence of $\frac{S_n}{n} \to 0$ is effective given X_1 has a computable distribution and $\mathbb{E}(|X_1|)$ is computable.
- Furthermore, Gács demonstrates that any general rate of almost sure convergence for $\frac{S_n}{n} \to 0$ must depend on $\mathbb{E}(|X_1|)$

Computability in the strong law of large numbers

- In 2010 Gács demonstrates how the almost sure convergence of $\frac{S_n}{n} \to 0$ is effective given X_1 has a computable distribution and $\mathbb{E}(|X_1|)$ is computable.
- Furthermore, Gács demonstrates that any general rate of almost sure convergence for $\frac{S_n}{n} \to 0$ must depend on $\mathbb{E}(|X_1|)$
- They do this by the construction of a random variable X_1 with a computable distribution but $\mathbb{E}(|X_1|)$ noncomputable

Kolmogorov's strong law of large numbers

Dropping the identical distribution assumption means we must include additional assumptions to ensure that we can conclude $\frac{S_n}{n} \to 0$ almost surely.

Theorem (Kolmogorov's strong law of large numbers)

If $\{X_n\}$ *are independent and*

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty \tag{1}$$

then $\frac{S_n}{n} \to 0$ *almost surely*

Pairwise independent

Unlike the identical distribution case, we cannot reduce the independence condition to pairwise independent random variables. Csörgő and Tandori showed,

Theorem (Csörgő-Tandori 1983)

If $\{X_n\}$ *are pairwise independent, (1) holds and*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(|X_k|) = O(1) \tag{2}$$

then, $\frac{S_n}{n} \to 0$ *almost surely.*

Aim of talk

- There does not appear to be any bounds for P_{n,ε_n} in the same spirit as the likes of Siegmund, Fill, and Luzia in the case where we do not assume the random variables are identically distributed. Furthermore, the computability theory of these Strong laws of large numbers has not been studied.

Aim of talk

- There does not appear to be any bounds for P_{n,ε_n} in the same spirit as the likes of Siegmund, Fill, and Luzia in the case where we do not assume the random variables are identically distributed. Furthermore, the computability theory of these Strong laws of large numbers has not been studied.
- I hope to present the progress I have made in studying $P_{n,\varepsilon}$ for random variables which are not identically distributed, analytically and from a computability point of view.

Computable convergence

Recall the standard definition of Cauchy convergence,

$$\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

Computable convergence

Recall the standard definition of Cauchy convergence,

$$\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

A natural question to ask is, from a proof of convergence can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that $\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$

Computable convergence

Recall the standard definition of Cauchy convergence,

$$\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

A natural question to ask is, from a proof of convergence can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that $\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$

Specker showed this was not always possible, through his famous construction of a monotone sequence of rational numbers converging to a non-computable number

Computable convergence

One can show that, over classical logic, Cauchy convergence is equivalent to,

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \, \forall i, j \in [n, n + g(n)](|a_i - a_j| \leq \varepsilon)$$

Computable convergence

One can show that, over classical logic, Cauchy convergence is equivalent to,

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \, \forall i, j \in [n, n + g(n)](|a_i - a_j| \leq \varepsilon)$$

This is known as the No-counterexample interpretation of the definition of Cauchy convergence. A bound on n depending on ε and g is known as a rate of metastability and the extraction of such rates are standard results in applied proof theory. See Proof Interpretations and their Use in Mathematics, by Kohlenbach for detailed discussions and examples.

Computable Probabilistic Convergence

In the same spirit, one can talk about rates of almost sure convergence and metastable rates of almost sure convergence.
If $\{Y_n\}$ is a sequence of random variables, then Y_n converges to 0 almost surely simply means,

$$\mathbb{P}(\{Y_n\} \text{ converges to } 0) = 1$$

It is clear that there is no direct computational interpretation that can be given to this definition, however one can show that it is equivalent to,

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \, \forall n \geq N \mathbb{P}(\max_{m \geq n} |Y_m| > \lambda_1) \leq \lambda_2$$

We call a solution to the computational interpretation a rate of almost sure convergence

Computable Probabilistic Convergence

Furthermore, we can also obtain a metastable notion of almost sure uniform convergence

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \forall K : \mathbb{N} \to \mathbb{N} \, \exists N \mathbb{P}(\max_{N \leq m \leq K(N)} |Y_m| > \lambda_1) \leq \lambda_2$$

This notion of probabilistic convergence was first studied by Avigad, Gerhardy and Towsner in in 2007, with explicit rates being extracted for the pointwise Ergodic theorem.

Computability and the strong law of large numbers

We construct a sequence of computable independent random variables $\{X_n\}$ such that

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < V$$

For some rational V and

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(|X_k|) = O(1)$$

But, $\frac{S_n}{n}$ does not converge to 0 with a computable rate of almost sure convergence. Or equivalently, it is not the case that $P_{n,\varepsilon}$ converges to 0 with a computable rate of convergence, for every rational $\varepsilon > 0$.

Metastable Csörgő and Tandori

Suppose, $\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2}$ converges with rate of metastability Ψ. If we have $A, V > 0$ such that,

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < V$$

and, for all n

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(|X_k|) < A$$

Then, $\frac{1}{n}S_n$ converges to 0 almost surely, with rate of metastable almost sure convergence,

$$\Lambda(\lambda_1, \lambda_2, K) = \lfloor \alpha^{\max\{\Phi_1^{\alpha, \frac{\lambda_1}{3\alpha}}(\frac{\lambda_2}{4}, R_{\frac{\lambda_2}{2}, H}), \Phi_2^{\alpha, \frac{\lambda_1}{3\alpha}}(\frac{\lambda_2}{4}, S_{\frac{\lambda_2}{2}, H})\}} \rfloor$$

Metastable Csörgő and Tandori

Where

$$\alpha = \min\{|\frac{3A}{3A-2\lambda_1}|, \frac{3A+2\lambda_1}{3A}\}$$

$$H(n) = \lfloor log_\alpha(K(\lfloor \alpha^n \rfloor))\rfloor$$

$$\Phi_1^{\alpha,\varepsilon}(\lambda, H) = \lfloor log_\alpha(\Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{8\alpha^4}, F_H))\rfloor$$

$$\Phi_2^{\alpha,\varepsilon}(\lambda, H) = \lfloor log_\alpha(\max\{\Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{32}, h),$$

$$\lceil\sqrt{\frac{32V\Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{32}, h)^2\alpha^4}{\lambda\varepsilon^2(\alpha^2-1)}}\rceil\})\rfloor$$

$$F_H(n) = \lfloor \alpha^{H(\lfloor log_\alpha(n)\rfloor)+1}\rfloor$$

$$h(n) = F_H(\max\{n, \lceil\sqrt{\frac{32Vn^2\alpha^4}{\lambda\varepsilon^2(\alpha^2-1)}}\rceil\})$$

Rates of convergence for Csörgő and Tandori

Suppose, $\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2}$ converges with rate of convergence Ψ.For all $\varepsilon, \lambda > 0$ and $\forall n \geq \Lambda(\frac{\lambda}{2})$

$$P_{n,\varepsilon} = \mathbb{P}(\max_{m \geq n} |\frac{1}{n} S_n| > \varepsilon) \leq \lambda$$

where, $\Lambda(\frac{\lambda}{2}) =$

$$\max\{\alpha\Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{288\alpha^6}), \Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{1152\alpha^6}), \sqrt{\frac{2304\Psi(\frac{\lambda\varepsilon^2(\alpha^2-1)}{1152\alpha^6})^2 V\alpha^6}{\lambda\varepsilon^2(\alpha^2-1)}}\}$$

with $\alpha = 1 + \frac{2\varepsilon}{3A}$. Thus, $P_{n,\varepsilon}$ converges to 0 with rate of convergence given above.

Improving known rates

In 2018 Luzia showed that if $\{X_n\}$ is a sequence of pairwise independent, identically distributed random variables with $\mathbb{E}(X_1) = 0$, $Var(X_1) = \sigma^2 < \infty$ and $\mathbb{E}(|X_1|) = \tau < \infty$ then for all $\beta > 1$

$$\mathbb{P}(\max_{m \geq n} |\frac{1}{m} S_m| > \varepsilon) = O(\frac{log(n)^{\beta-1}}{n})$$

Observe that Luzia's assumptions are stronger than those made by Csörgő and Tandori's. Furthermore, Csörgő and Tandori's proof simplifies tremendously under these stronger assumptions and analysing this simplified proof allows us to deduce that,

$$\mathbb{P}(\max_{m \geq n} |\frac{1}{m} S_m| > \varepsilon) = O(\frac{1}{n})$$

Improving known rates

More precisely, for all $\varepsilon, \lambda > 0$ and $\forall n \geq \Phi(\frac{\lambda}{2})$

$$P_{n,\varepsilon} = \mathbb{P}(\max_{m \geq n} |\frac{1}{n} S_n| > \varepsilon) \leq \lambda$$

where, $\Phi(\frac{\lambda}{2}) = \frac{32\sigma^2\alpha^3}{\lambda\varepsilon^2(\alpha-1)}$, with $\alpha = 1 + \frac{\varepsilon}{\tau}$.

Concluding remarks

- Can we obtain asymptotic equivalence or better bounds if we assume stronger conditions, in the case where we still do not assume identical distribution?

Concluding remarks

- Can we obtain asymptotic equivalence or better bounds if we assume stronger conditions, in the case where we still do not assume identical distribution?
- Can we give a computational interpretation to the Menšov and Rademacher theorem states that if $\{X_n\}$ are pairwise uncorrelated and

$$\sum_{n=2}^{\infty} \frac{\mathbb{E}(X_n^2)(log(n))^2}{n^2} < \infty$$

then the conclusion of the strong law of large numbers? Observe that pairwise uncorrelated is a weaker condition than pairwise independence (for random variables with finite second moment). Does reducing to the case of identical distribution with finite variance may give better bounds?