New Approaches to Uniformity Testing

Winston Li

Rutgers University

Mentor: Periklis Papakonstantinou

June 2, 2025

Uniformity Testing

Randomness

Randomness is a ubiquitous resource in modern algorithms, such as Cryptography, Machine Learning, and MCMC.

Uniformity Testing

Randomness

Randomness is a ubiquitous resource in modern algorithms, such as Cryptography, Machine Learning, and MCMC.

Theory

Given a random variable X over a set Ω, can we determine if this variable is uniform? That is, for all $x \in \Omega$, $\Pr[X = x] = |\Omega|^{-1}$. It turns out, determining if $\Delta(X, U_\Omega) \leq \varepsilon$ requires $\Theta(\varepsilon^{-2} \cdot \sqrt{n})$ samples. [2]

Uniformity Testing

Randomness

Randomness is a ubiquitous resource in modern algorithms, such as Cryptography, Machine Learning, and MCMC.

Theory

Given a random variable X over a set Ω, can we determine if this variable is uniform? That is, for all $x \in \Omega$, $\Pr[X = x] = |\Omega|^{-1}$. It turns out, determining if $\Delta(X, U_\Omega) \leq \varepsilon$ requires $\Theta(\varepsilon^{-2} \cdot \sqrt{n})$ samples. [2]

Practice

Modern randomness test suites (like NIST and Dieharder) use statistical hypothesis testing to detect non-uniform distributions.

Hypothesis Testing

Definition

Given a sequence of bits x, perform a statistical test with hypothesis H_0 if x is random and H_a if x is not. Computing test statistic S for x, find P-value to represent the probability x is random. If P-value is less than significance level α, reject H_0 and find the sequence is non-random. [1]

Hypothesis Testing

Definition

Given a sequence of bits x, perform a statistical test with hypothesis H_0 if x is random and H_a if x is not. Computing test statistic S for x, find P-value to represent the probability x is random. If P-value is less than significance level α, reject H_0 and find the sequence is non-random. [1]

Flaws

A statistical test with significance level α effectively declares α proportion of bit sequences to be non-random. For a random variable X over $\{0,1\}^n$, a test can be identified with a subset of *non-random* values with size $\alpha \cdot 2^n$.

Many non-uniform distributions will fail to be detected by this method of testing. Ex: Suppose $x \in \{0,1\}^n$ is some string that is accepted by all NIST tests. Then the constant distribution over x will be seen as random.

Goals

Overview

- Analyzing uniformity testing through the lens of distribution testing becomes intractible for larger strings of bits.
- Existing test suites fail to detect even simple pathological distributions.

Goals

Overview

- Analyzing uniformity testing through the lens of distribution testing becomes intractible for larger strings of bits.
- Existing test suites fail to detect even simple pathological distributions.

Goals

This project aims to come up with more concise definitions for uniformity testing. Furthermore, we are looking to develop a stronger set of tests compared to existing test suites. Ultimately, we are looking to bridge theory and practice for this field.

References

[1] Lawrence Bassham et al. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. en. 2010. URL: https://www.nist.gov/publications/statistical-test-suite-random-and-pseudorandom-number-generators-cryptographic.

[2] Siu-On Chan et al. "Optimal Algorithms for Testing Closeness of Discrete Distributions". In: *Proceedings of the 2014 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1193–1203. DOI: 10.1137/1.9781611973402.88. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611973402.88.

Acknowledgements

This work was carried out as part of the 2025 DIMACS REU program at Rutgers University, supported by NSF grant CCF-2447342.