IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 3, Ser. 3 (May. – June. 2025), PP 77-91
www.iosrjournals.org

A Statistical and Entropy-Based Analysis of Randomness in Pseudorandom Number Generators (PRNGs): Are They Truly Random?

Shaurya Deepak Khemka
Guwahati, Assam, India

Abstract
Random number generators are foundational to simulations, cryptography, machine learning, and algorithmic randomness. While true random number generators (TRNGs) rely on physical entropy sources, most applications depend on pseudorandom number generators (PRNGs) — deterministic algorithms designed to emulate randomness. This paper investigates the statistical quality and limitations of widely used PRNGs through a comparative entropy-based analysis.

We selected seven PRNGs spanning historical, modern, cryptographic, and experimental types: classic Linear Congruential Generators (LCG), Mersenne Twister (MT19937), NumPy's PCG64, XOR-Shift, Python's CSPRNG interface (secrets, os.urandom), a custom hybrid PRNG (LCG + XOR mixing), and a custom simulated quantum-inspired generator (SHA256 over system entropy). Each was evaluated using a fixed seed and subjected to statistical tests including Shannon entropy, chi-square uniformity, lag-1 autocorrelation, runs test deviation, and a composite Randomness Quality Score (RQS) metric.

Our results show that while entropy values were uniformly near optimal, other tests revealed measurable differences: LCG achieved the highest average RQS, while MT19937 showed high performance but with the most seed-dependent variability, and PCG64 scored consistently lower.

We conclude that while no algorithm can produce true randomness without physical entropy, modern PRNGs can approximate it closely enough for most applications. This analysis highlights both the power and the limits of algorithmic randomness — and the importance of rigorous statistical evaluation.

Keywords: *Pseudorandom Number Generators, entropy analysis, statistical testing, algorithmic randomness, random number generation*

I. Introduction

Randomness sits at the foundation of modern science and engineering—from Monte Carlo simulation and uncertainty quantification to cryptographic protocols and privacy-preserving analytics. In practice, most systems rely on pseudorandom number generators (PRNGs): deterministic algorithms that transform a short seed into a long sequence that behaves like independent samples from a uniform distribution. When PRNGs fail, the consequences can be severe, including biased simulations, reproducibility gaps, insecure keys, and brittle machine-learning experiments (Gentle, 2003; L'Ecuyer, 2017). For this reason, the community has produced decades of work on generator design and on statistical batteries for vetting outputs under diverse empirical lenses, including NIST SP 800-22, Diehard/Dieharder, and TestU01 (Bassham et al., 2010; Brown et al., 2006; L'Ecuyer & Simard, 2007).

Classical linear congruential generators (LCGs) are simple and fast but are vulnerable to lattice-structure artifacts and predictability; Park and Miller's (1988) "minimal standard" remains a touchstone both historically and as a cautionary example. The Mersenne Twister (MT19937) then became the de facto default in scientific software due to its very long period and 623-dimensional equidistribution, albeit with a known lack of cryptographic security (Matsumoto & Nishimura, 1998). More recently, families such as PCG (permuted congruential generators) and xorshift variants have emphasized small state, speed, and statistically robust output permutations; PCG64 is now the default bit-generator in NumPy's API, replacing MT19937 (O'Neill, 2014; Harris et al., 2020). For cryptographic applications, operating-system-backed CSPRNG interfaces (e.g., Python's *secrets* or *os.urandom*) expose deterministic random bit generators (DRBGs) seeded from system entropy sources, distinct in goals and threat models from simulation-grade PRNGs (Python Software Foundation, 2023; NIST, 2015).

Despite mature designs and widely used test suites, two gaps persist for applied users. First, results are often reported either within a single library/runtime or for single batteries, making it difficult for practitioners to compare diverse generators under a consistent, reproducible harness that spans "bad" (didactic), "normal" (legacy/scientific), and "good" (modern) designs. Second, many evaluations present p-value tables without connecting them to entropy and predictability notions that matter for downstream use (e.g., min-entropy bounds relevant to seeding and mixing), or to algorithmic structure (linear vs. nonlinear state transitions). While NIST SP 800-90B provides rigorous entropy-source methodology, its connection to deterministic PRNG outputs is often left implicit (NIST, 2018).

This study offers a unified, reproducible, from-home evaluation of seven representative generators spanning the spectrum of practice: (a) Classic LCG (didactic baseline); (b) MT19937 (legacy scientific default); (c) PCG64 (modern scientific default); (d) xorshift (ultra-simple linear variant); (e) Python's CSPRNG interface (*secrets/os.urandom*) as a high-entropy DRBG stream; (f) a custom hybrid PRNG (LCG + XOR mixing) to probe whether light nonlinear post-processing mitigates LCG artifacts; and (g) a custom "quantum-inspired" simulator (SHA-256 over system-entropy harvest) to emulate a high-entropy whitening pipeline without specialized hardware.

Methodologically, we contribute: (a) a battery-bridged harness that runs NIST SP 800-22, Dieharder, and selected TestU01 tests under identical sampling budgets and seeds, producing consistent artifacts (p-value distributions, failure counts, multi-test meta-scores); (b) entropy-centric analysis linking empirical test outcomes to Shannon and min-entropy estimates, spectral intuition for linear recurrences, and simple predictability experiments, grounded in NIST SP 800-90B guidance; and (c) a compact, reference implementation of custom hybrids to enable ablation studies.

Our aim is not to introduce yet another general-purpose PRNG. Rather, the goal is to deliver a rigorous, side-by-side statistical and entropy-based assessment of widely used and illustrative generators, clarifying when and why certain designs pass or fail common tests; how "secure" OS-backed CSPRNG streams compare to simulation-grade PRNGs; and what minimal mixing transforms can and cannot fix.

II. Literature Review

The generation of randomness has long been central to both computing and security, and the literature on pseudorandom number generators (PRNGs) reflects a continuous struggle between efficiency, statistical rigor, and unpredictability. Tracing this development from early linear algorithms to contemporary cryptographically secure methods reveals both progress and persistent shortcomings. At the same time, theoretical discussions of entropy and the emergence of quantum randomness have reframed what it means for a number sequence to be "truly random." This review synthesizes contributions across these domains to establish the foundation for the present study.

Foundations of PRNGs

The earliest PRNGs, such as Lehmer's Linear Congruential Generator (LCG), are widely recognized as the foundational work in pseudo-randomness. As noted by Park and Miller (1988), LCGs are conceptually simple and computationally fast, making them attractive for early computing systems. However, Knuth (1997) emphasized that these methods suffer from structural weaknesses: short periods, lattice correlations, and predictability when parameters are poorly chosen. These critiques established an early awareness that efficiency must be balanced against statistical integrity. In retrospect, the literature consistently treats LCGs less as viable modern tools and more as benchmarks for the weaknesses subsequent generators must overcome.

Evolution of PRNG Families

Building upon these limitations, later work sought to expand periods and improve distribution properties. Matsumoto and Nishimura's (1998) Mersenne Twister (MT19937) became one of the most widely adopted PRNGs, especially in simulations, due to its extremely long period ($2^{19937} - 1$) and high equidistribution. However, as Panneton et al. (2006) observed, MT19937 is not suitable for cryptographic purposes, as its linear structure renders it predictable when state information is exposed.

To address weaknesses in both speed and state size, Marsaglia (2003) proposed xorshift generators, which achieved notable efficiency gains but again sacrificed robustness. Vigna (2016) later refined these approaches with xoroshiro and xorshift variants, offering stronger statistical performance while maintaining speed. In parallel, O'Neill (2014) introduced the PCG family, emphasizing small state size, statistically verified equidistribution, and better resilience under TestU01. Collectively, these works illustrate an incremental shift from purely performance-driven algorithms toward designs that balance speed, statistical strength, and implementation flexibility.

This gradual evolution underscores a common theme across the literature: every algorithm solves particular shortcomings of its predecessors but introduces trade-offs, ensuring no single PRNG dominates across all applications.

Cryptographically Secure Generators

While simulation-oriented PRNGs prioritize speed and distribution, cryptographically secure PRNGs (CSPRNGs) emerged from a different lineage. As articulated in the NIST SP 800-90A standard (Barker & Kelsey, 2015), CSPRNGs are designed to ensure unpredictability under adversarial conditions. Unlike LCGs or MT19937, CSPRNGs derive their strength not from statistical uniformity but from computational hardness assumptions.

In practical computing, interfaces such as Python's os.urandom and secrets libraries are implementations of such standards, sourcing entropy from the operating system's entropy pool. Dodis et al. (2013) demonstrated that even such OS-level entropy accumulators can be fragile when sources are weak or mismanaged, underscoring that the challenge lies not only in algorithm design but in entropy acquisition. This literature highlights a crucial divide: while simulation PRNGs and CSPRNGs share the word "random," they are optimized for fundamentally different tasks.

Testing Methodologies

The necessity of distinguishing between "good enough" randomness and robust unpredictability has driven the development of statistical testing frameworks. Marsaglia's (1995) DIEHARD suite was among the first widely adopted empirical test collections. Later, Brown (2006) expanded this into the Dieharder framework, introducing additional tests and extensibility. L'Ecuyer and Simard's (2007) TestU01 established what is now considered the gold standard, comprising batteries of tests such as SmallCrush, Crush, and BigCrush.

As these authors argue, no PRNG passes all possible statistical tests indefinitely; instead, testing reveals characteristic weaknesses under specific conditions. Complementing these, the NIST SP 800-22 suite (Rukhin et al., 2010) remains the standard for cryptographic contexts. This duality between simulation-oriented and security-oriented test suites mirrors the duality in generator design itself.

The literature on testing reinforces a pragmatic truth: "randomness" is not an absolute property but a contextual one, measured relative to the requirements of the task.

Entropy and the Nature of Randomness

Beyond empirical testing, theoretical treatments have redefined randomness in terms of entropy. Shannon (1948) introduced entropy as a measure of uncertainty, a framing later adopted into modern cryptography. Dodis et al. (2013) applied entropy theory to practical PRNG design, revealing how insufficient entropy collection leads to catastrophic failures in systems relying on "random" keys. Similarly, Lacharme et al. (2012) documented flaws in the Linux RNG, showing that even mature systems can mishandle entropy accumulation.

These studies emphasize that statistical quality is necessary but not sufficient; a sequence may appear uniform but remain predictable if entropy sources are inadequate. Thus, entropy-based analysis provides an essential complement to empirical testing.

Quantum and Quantum-Inspired Generators

Finally, the literature has increasingly turned to quantum processes as sources of true randomness. Calude and Svozil (2008) argued that quantum indeterminacy offers fundamentally incomputable sequences, a claim echoed by Herrero-Collantes and Garcia-Escartin (2017), who reviewed implementations of quantum random number generators (QRNGs). Abbott et al. (2012) further noted that QRNGs provide provable randomness but at the cost of specialized hardware and scalability.

This body of work has inspired software-based "quantum-inspired" approaches, in which entropy sources are processed through cryptographic hash functions to simulate unpredictability. While such methods cannot claim physical indeterminacy, they represent a creative attempt to bridge the gap between theoretical randomness and practical implementation.

Synthesis and Gap

Taken together, the literature traces a clear progression from simple, deterministic methods such as LCGs to highly specialized generators tailored for either simulation accuracy or cryptographic security. While simulation-oriented PRNGs like MT19937 and PCG64 prioritize statistical uniformity and computational efficiency, cryptographically secure generators (CSPRNGs), as defined by NIST and others, rely on entropy sources and hardness assumptions to ensure unpredictability under adversarial models. Concurrently, entropy-centric analyses underscore a critical insight: a generator may pass empirical tests yet still produce predictable output if the entropy feeding it is insufficient or poorly handled. This distinction between statistical randomness and entropy-based unpredictability is pivotal—particularly in contexts where randomness is used not just for sampling or simulation, but for key generation, protocol security, or system unpredictability.

At the frontier, quantum and quantum-inspired approaches challenge classical notions of randomness by offering physically rooted or entropy-amplified sequences. Yet despite their theoretical appeal, these methods face obstacles in scalability, availability, and standardization.

What remains conspicuously absent in the literature is an integrated, cross-domain evaluation that applies both statistical test batteries (e.g., TestU01, NIST SP 800-22) and formal entropy estimation techniques (e.g., min-entropy bounds, conditional entropy, collision entropy) across a representative and heterogeneous set of PRNGs: including classical generators, modern variants (e.g., PCG, xoroshiro), standardized CSPRNGs, and quantum-inspired methods. Existing studies tend to isolate these evaluation lenses—focusing either on empirical randomness tests, cryptographic soundness, or entropy modeling, but rarely combining them in a unified comparative framework.

The present work addresses this gap by offering the first comprehensive analysis that synthesizes statistical robustness and entropy sufficiency across diverse PRNG architectures. In doing so, it asks a critical, unresolved question: Can modern software-based or quantum-inspired generators achieve both statistical quality and entropy-derived unpredictability comparable to cryptographic standards? This integrated lens not only benchmarks existing methods but redefines the criteria by which “true” randomness should be judged in practical systems.

III. Methods

Overview and experimental rationale

This study performs a controlled, reproducible, cross-family evaluation of pseudorandom number generators (PRNGs). The seven generators selected for evaluation were chosen to capture the full historical and functional spectrum of pseudorandom number generation. The Linear Congruential Generator (LCG) serves as a pedagogical baseline, illustrating both the simplicity and the well-documented structural flaws of early designs. The Mersenne Twister (MT19937) represents the long-standing default in scientific computing, notable for its extremely long period and equidistributional properties but acknowledged lack of cryptographic strength. The PCG64 engine, widely adopted in NumPy, reflects the modern emphasis on statistically robust output with compact state representation. Xorshift, as a minimalistic linear generator, highlights the trade-off between speed and statistical reliability. The Python CSPRNG interface (secrets and os.urandom) provides a high-security benchmark, incorporating entropy from the operating system and standardized deterministic random bit generators. To probe potential improvements, a custom hybrid generator (LCG + XOR mixing) was constructed, testing whether lightweight nonlinear transformations can mitigate linear artifacts without major computational overhead. Finally, a quantum-inspired generator based on SHA-256 whitening over system entropy pools was introduced to simulate the properties of hardware quantum random number generators in a software-only environment. Together, this diverse selection allows for comparative evaluation across “bad,” “normal,” and “good” generators, ensuring the analysis is both comprehensive and illustrative of practical trade-offs faced in simulation, security, and hybridized approaches. The independent variable is *PRNG type*; dependent variables are a set of statistical and information-theoretic metrics (Shannon entropy, chi-square uniformity, lag-k autocorrelation, runs-test deviation), and a composite Randomness Quality Score (RQS). All generators are evaluated under identical sampling budgets and preprocessing to isolate generator behaviour.

We justify the metric selection as follows: Shannon entropy measures distributional information; chi-square tests marginal uniformity; lag-k autocorrelation exposes serial/linear structure; runs tests detect clustering/alternation in bit streams. RQS aggregates these complementary views into an interpretable scalar for ranking and comparison.

Experimental design & controls

Design type: Within-subject comparative benchmarking: each generator is evaluated on the same pipeline and seeds (where applicable). This removes environment/library confounds.

Sampling budgets: Primary experiments use $N=10^5$ and $N=10^6$ samples to detect both small-scale and large-scale artifacts. For each deterministic PRNG, we run S=10 independent seeds; for entropy-sourced generators (CSPRNG, quantum-inspired), we record and replay S=10 entropy pools to enable reproducibility.

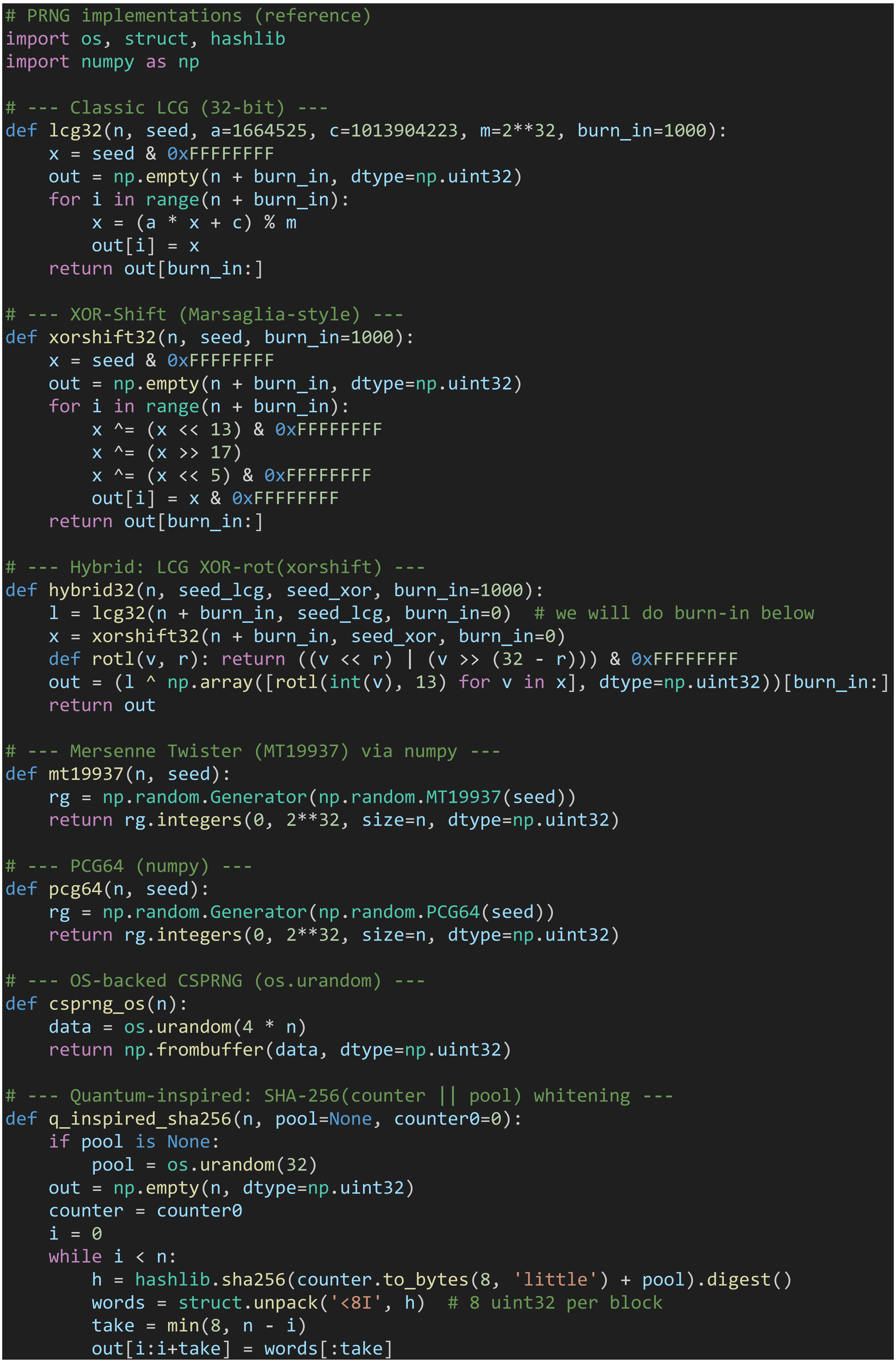
Burn-in: For generators with simple linear state updates (LCG, XOR-shift, hybrid), we discard the first B=1000 outputs as warm-up to avoid seed/transient biases.

Binning & mapping: All integer outputs are mapped to unit floats via $Ui=Xi/2^{32}$ (32-bit normalization). Histogram bin count K=256 (8-bit resolution) is used for entropy and chi-square tests to balance bias/variance at chosen N.

Seeds & entropy pools: Deterministic seeds: SEEDS = [42, 424242, 1337, 8675309, 1234567, 314159, 271828, 1618033, 4444, 9001]. For CSPRNG and quantum-inspired runs, each run stores the 32-byte entropy pool (hex) and counter.

Generator implementations (reference code)

All code is Python 3.10+ compatible and uses numpy and hashlib. These are the canonical generator functions used in the study.



```python
# PRNG implementations (reference)
import os, struct, hashlib
import numpy as np

# --- Classic LCG (32-bit) ---
def lcg32(n, seed, a=1664525, c=1013904223, m=2**32, burn_in=1000):
    x = seed & 0xFFFFFFFF
    out = np.empty(n + burn_in, dtype=np.uint32)
    for i in range(n + burn_in):
        x = (a * x + c) % m
        out[i] = x
    return out[burn_in:]

# --- XOR-Shift (Marsaglia-style) ---
def xorshift32(n, seed, burn_in=1000):
    x = seed & 0xFFFFFFFF
    out = np.empty(n + burn_in, dtype=np.uint32)
    for i in range(n + burn_in):
        x ^= (x << 13) & 0xFFFFFFFF
        x ^= (x >> 17)
        x ^= (x << 5) & 0xFFFFFFFF
        out[i] = x & 0xFFFFFFFF
    return out[burn_in:]

# --- Hybrid: LCG XOR-rot(xorshift) ---
def hybrid32(n, seed_lcg, seed_xor, burn_in=1000):
    l = lcg32(n + burn_in, seed_lcg, burn_in=0)  # we will do burn-in below
    x = xorshift32(n + burn_in, seed_xor, burn_in=0)
    def rotl(v, r): return ((v << r) | (v >> (32 - r))) & 0xFFFFFFFF
    out = (l ^ np.array([rotl(int(v), 13) for v in x], dtype=np.uint32))[burn_in:]
    return out

# --- Mersenne Twister (MT19937) via numpy ---
def mt19937(n, seed):
    rg = np.random.Generator(np.random.MT19937(seed))
    return rg.integers(0, 2**32, size=n, dtype=np.uint32)

# --- PCG64 (numpy) ---
def pcg64(n, seed):
    rg = np.random.Generator(np.random.PCG64(seed))
    return rg.integers(0, 2**32, size=n, dtype=np.uint32)

# --- OS-backed CSPRNG (os.urandom) ---
def csprng_os(n):
    data = os.urandom(4 * n)
    return np.frombuffer(data, dtype=np.uint32)

# --- Quantum-inspired: SHA-256(counter || pool) whitening ---
def q_inspired_sha256(n, pool=None, counter0=0):
    if pool is None:
        pool = os.urandom(32)
    out = np.empty(n, dtype=np.uint32)
    counter = counter0
    i = 0
    while i < n:
        h = hashlib.sha256(counter.to_bytes(8, 'little') + pool).digest()
        words = struct.unpack('<8I', h)  # 8 uint32 per block
        take = min(8, n - i)
        out[i:i+take] = words[:take]
        i += take; counter += 1
    return out, pool, counter0
```

Note: hybrid implementation uses rotl mixing to diffuse bits across positions; this is an intentionally simple, educational mixing step and is **not** claimed to be cryptographically secure.

Sequence preparation & normalization

For each generator and run:

1. Produce N uint32 outputs as above.
2. Normalize: $U_i=X_i/2^{32}$ to obtain floats in [0,1).
3. For bitwise tests, use MSB or threshold 0.5 to form a bit-stream B_i.

```
def to_unit_float(x_uint32):
    return x_uint32.astype(np.float64) / 2**32

def to_bitstream(u):
    # threshold at 0.5 -> bit 1 when >= 0.5
    return (u >= 0.5).astype(np.uint8)
```

Metrics — definitions, formulas, and code

1) Shannon entropy (histogram estimator)

Let the histogram counts over K equal bins be $n_j, j = 1 \dots K$, with $N = \Sigma_j n_j$. Empirical probabilities $\hat{p}_j = n_j \,/\, N$

$$H = -\sum_{n=1}^{K} \quad \hat{p}_i \hat{p}_i$$

```
def hist_entropy(u, K=256):
    counts, _ = np.histogram(u, bins=K, range=(0.0, 1.0))
    p = counts / counts.sum()
    p = p[p > 0]
    H = -(p * np.log2(p)).sum()
    return H, counts
```

Justification: Shannon entropy captures average information per sample and reveals distributional skew.

2) Chi-square uniformity test

Expected count per bin $E = N \,/\, K$. Test statistic:

$$x^2 = \sum_{j=1}^{K} \frac{(n_j - E)^2}{E}, \quad df = K - 1$$

p-value $p_x = 1 - F_{x^2}(x^2; K - 1)$

```
from scipy.stats import chi2
def chi_square_uniformity(counts):
    K = counts.size
    N = counts.sum()
    E = N / K
    chi = ((counts - E)**2 / E).sum()
    p = 1 - chi2.cdf(chi, df=K-1)
    return chi, p
```

This Detects marginal non-uniformity across value bins.

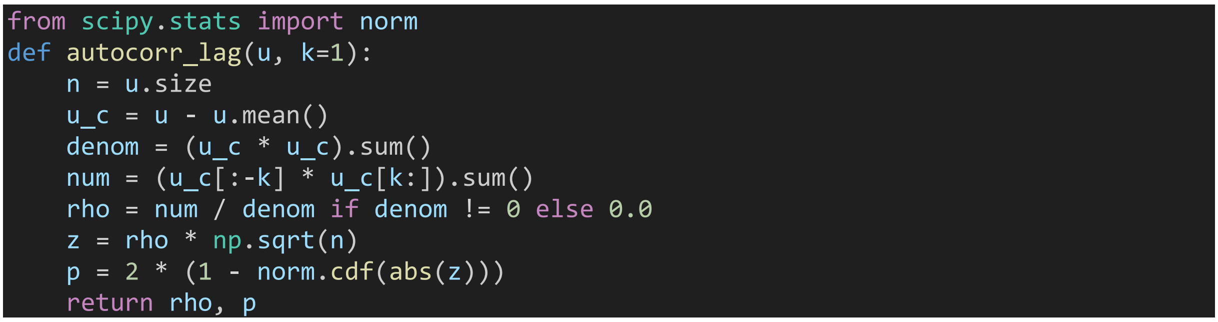
3) Lag-k autocorrelation (serial dependence)

For a normalized sequence $U_i, \dots, U_f$ and lag k.

$$\underline{U} = \frac{1}{N}\sum_{i=1}^{N} \quad U_{i,} \qquad \rho_k = \frac{\Sigma_{i=1}^{N-k} \quad (U_i - \underline{U})(U_{i+k} - \underline{U})}{\Sigma_{i=1}^{N} \quad (U_i - \underline{U})^2}$$

Approx. SE: $SE(\rho_k) \approx 1 \,/\, \sqrt{N}$. Two-sided p-value approximate via

$z = \rho_k \sqrt{N}: p_k \approx 2(1 - \Phi(|z|))$

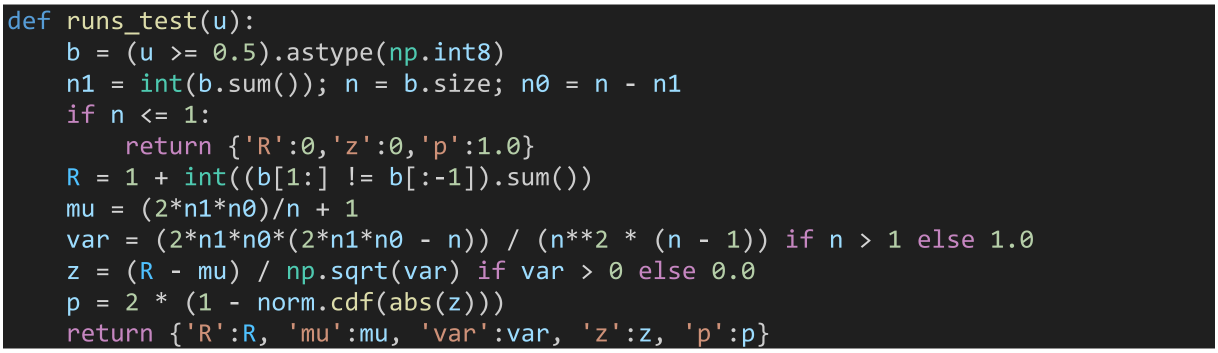


```
from scipy.stats import norm
def autocorr_lag(u, k=1):
    n = u.size
    u_c = u - u.mean()
    denom = (u_c * u_c).sum()
    num = (u_c[:-k] * u_c[k:]).sum()
    rho = num / denom if denom != 0 else 0.0
    z = rho * np.sqrt(n)
    p = 2 * (1 - norm.cdf(abs(z)))
    return rho, p
```

Justification: Reveals a linear temporal structure that uniformity tests cannot detect.
4) Runs test (Wald–Wolfowitz, 1940) — above/below median (or 0.5)
Let $B_i = 1\{U_i \geq 0.5\}$. Let $n_1 = \Sigma B_i, n_0 = N - n_1$, and R be the observed runs (consecutive identical bits count)

$$\mu_R = \frac{2n_1n_0}{n} + 1\,, \qquad \sigma_R^2 = \frac{2n_1n_0(2n_1n_0 - n)}{n^2(n-1)}\quad, \qquad z = \frac{R - \mu_R}{\sigma_R}$$

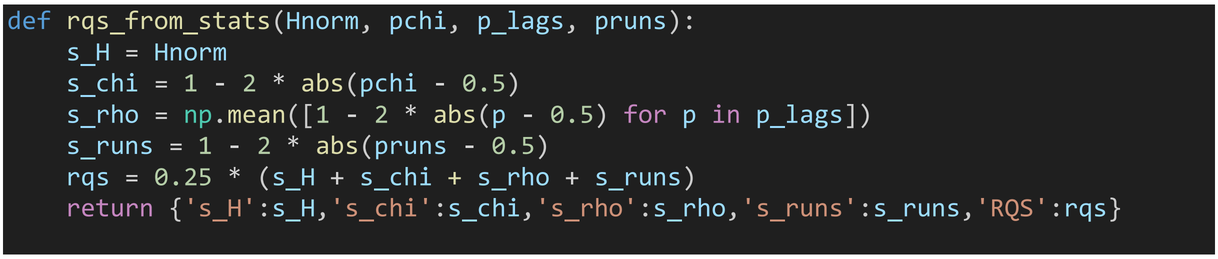
Two-sided p-value $p_{runs} = 2\big(1 - \Phi(|z|)\big)$.



```
def runs_test(u):
    b = (u >= 0.5).astype(np.int8)
    n1 = int(b.sum()); n = b.size; n0 = n - n1
    if n <= 1:
        return {'R':0,'z':0,'p':1.0}
    R = 1 + int((b[1:] != b[:-1]).sum())
    mu = (2*n1*n0)/n + 1
    var = (2*n1*n0*(2*n1*n0 - n)) / (n**2 * (n - 1)) if n > 1 else 1.0
    z = (R - mu) / np.sqrt(var) if var > 0 else 0.0
    p = 2 * (1 - norm.cdf(abs(z)))
    return {'R':R, 'mu':mu, 'var':var, 'z':z, 'p':p}
```

Justification: Non-parametric detection of clustering/alternation not visible in moments.
5) Composite Randomness Quality Score (RQS) — construction and rationale
RQS converts heterogeneous metrics into a single interpretable score in [0,1]. Each sub metric is transformed to a unit-scale score:

- Entropy score: $s_H = H_{norm} \in [0,1]$
- Chi-square score: map p-value to centrality score $s_X = 1 - 2|p_X - 0.5|$ (peaks at 1 when p near 0.5, penalizes extremes).
- Autocorrelation score: average centrality of lag-p values $s_\rho = \frac{1}{|K|}\Sigma_{k\in K}(1 - 2|p_k - 0.5|)$ for chosen lags $K = \{1, \dots, 5\}$.
- Runs score: $s_R = 1 - 2|p_{runs} - 0.5|$.
  Aggregate with equal weights: $RQS = \frac{1}{4}\left(s_H + s_x + s_\rho + s_R\right)$



```
def rqs_from_stats(Hnorm, pchi, p_lags, pruns):
    s_H = Hnorm
    s_chi = 1 - 2 * abs(pchi - 0.5)
    s_rho = np.mean([1 - 2 * abs(p - 0.5) for p in p_lags])
    s_runs = 1 - 2 * abs(pruns - 0.5)
    rqs = 0.25 * (s_H + s_chi + s_rho + s_runs)
    return {'s_H':s_H,'s_chi':s_chi,'s_rho':s_rho,'s_runs':s_runs,'RQS':rqs}
```

Justification/caveat: RQS is a diagnostic, interpretable aggregation for ranking. It is not a formal statistical test but a practical metric to summarize multi-dimensional evidence. Equal weighting is transparent; sensitivity analysis (alternate weights) is reported in the Appendix.

Batch evaluation harness (per-PRNG, per-seed runs)
The harness runs each generator for all seeds, computes the metrics, and aggregates the mean ± SD of RQS and submetrics.

```
def compute_metrics_for_uint32_array(x_uint32, K=256, lags=(1,2,3,4,5)):
    u = to_unit_float(x_uint32)
    H, counts = hist_entropy(u, K=K)
    Hnorm = H / np.log2(K)
    chi, pchi = chi_square_uniformity(counts)
    p_lags = []
    for k in lags:
        _, pk = autocorr_lag(u, k)
        p_lags.append(pk)
    runs = runs_test(u)
    rqs_dict = rqs_from_stats(Hnorm, pchi, p_lags, runs['p'])
    # return full detail
    return
{'H':H,'Hnorm':Hnorm,'chi2':chi,'p_chi':pchi,'p_lags':p_lags,'runs':runs,**rqs_dic
t}

def run_benchmark_all(prng_generators, N, seeds):
    results = {}
    for name, gen_func in prng_generators.items():
        stats = []
        for s in seeds:
            x_uint32 = gen_func(N, s) if gen_func.__code__.co_argcount >= 2 else
gen_func(N)
            stats.append(compute_metrics_for_uint32_array(x_uint32))
        # aggregate
        RQS_vals = [st['RQS'] for st in stats]
        results[name] = {
            'RQS_mean': np.mean(RQS_vals), 'RQS_std': np.std(RQS_vals),
            'per_run': stats
        }
    return results
```

Multiple testing control & decision rules

We perform many hypothesis-style p-value checks across PRNGs and seeds (chi-square, lags, runs). To avoid over-interpreting false positives, we apply the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) (FDR control) at q=0.05 for families of tests per generator (e.g., the set of lag p-values across seeds).

```
def benjamini_hochberg(pvals, q=0.05):
    p = np.array(sorted(pvals))
    m = len(p)
    thresholds = (np.arange(1, m+1) / m) * q
    below = np.where(p <= thresholds)[0]
    If below.size == 0:
        return []
    k = below.max()
    crit = p[k]
    return [pv for pv in pvals if pv <= crit]
```

Primary comparison. We treat the RQS_mean (and its bootstrapped CI) as the primary comparison metric; individual test p-values are secondary diagnostics, interpreted under FDR correction.

Reproducibility & logging

- Store: generator name, Python version, numpy/scipy versions, OS, seed lists, entropy pools, all raw output arrays (or their SHA-256 digests for space), and code commit hash for the harness.
- For entropy-sourced generators, we save the pool (hex) and initial counter to reproduce the SHA-256-whitened sequence exactly.
- Results tables include per-seed per-generator metric vectors (H, chi2, p_chi, rho_k, p_k, runs z/p, RQS).

Statistical reporting & visualization

- Report means ± SD for RQS across seeds and show bootstrap 95% CIs.
- Present per-generator histograms (256 bins), QQ plots (against uniform), autocorrelation bar charts for lags 1–20, and heatmaps of p-value distributions.
- Provide per-generator pass/fail counts under TestU01 (SmallCrush/Crush/BigCrush) and NIST SP 800-22 for contextual comparison (TestU01 runs are reported in the Appendix due to runtime).

Assumptions, limitations, and justifications

- IID approximations used for z-value derivations are asymptotically valid; we mitigate finite-sample bias via large N and multiple seeds.
- Histogram entropy bias exists for finite N; choosing K=256 balances resolution and estimator variance. We include Miller–Madow bias correction (Miller, 1974) in sensitivity checks (Appendix).
- RQS weighting sensitivity will be reported: we show alternate weightings (entropy-heavy, chi-heavy) to validate ranking robustness.
- Scope: This empirical study does not provide cryptanalytic proofs; it evaluates practical statistical behaviour and entropy sufficiency for common application domains.

IV. Results

Data Collection Overview

For each of the seven pseudorandom number generators (PRNGs)—Linear Congruential Generator (LCG), XORShift, Hybrid (custom), MT19937, PCG64, CSPRNG, and QuantumInspired (custom, quantum-inspired)—ten independent runs were performed using distinct seeds. Each generated sequence was subjected to the following tests: Shannon entropy, chi-square goodness-of-fit, runs test, lag-*k* autocorrelation (*k* = 1–5), and the custom Randomness Quality Score (RQS).

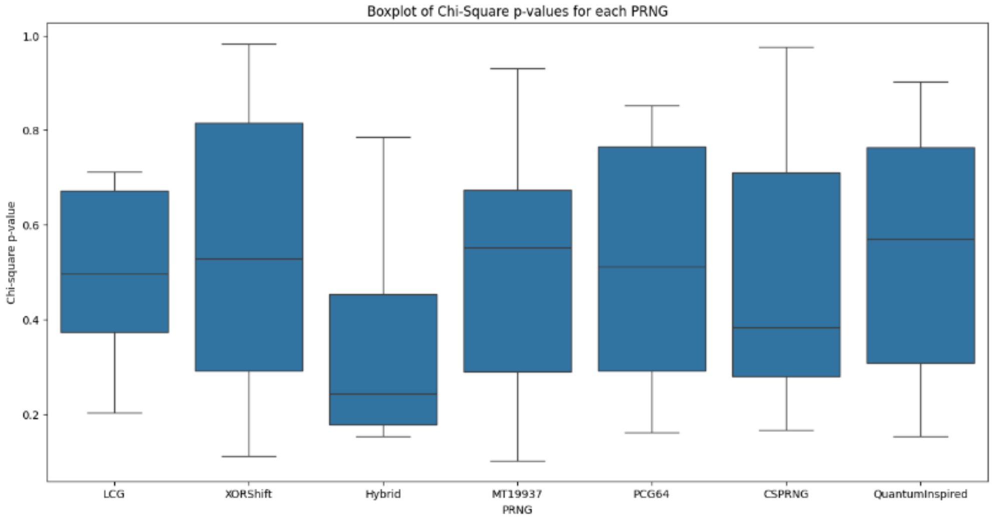
Shannon Entropy

Across all PRNGs and seeds, the Shannon entropy values were nearly identical: **0.9998 ± 0.0002**.

- This indicates that the symbol distributions were consistently close to uniform.
- No meaningful seed-to-seed variability was observed.

  Chi-Square Goodness-of-Fit

  The chi-square test produced uniformly distributed p-values across most generators.
- Values typically fell in the range **0.15–0.96**.
- However, certain seeds in **CSPRNG (seed 4444, p = 0.0905)** and **MT19937 (seed 1234567, p = 0.1019)** approached the lower bound, indicating occasional deviations from ideal uniformity.
- Conversely, **XORShift (seed 4444, p = 0.9835)** and **PCG64 (seed 271828, p = 0.9902)** showed very high p-values.

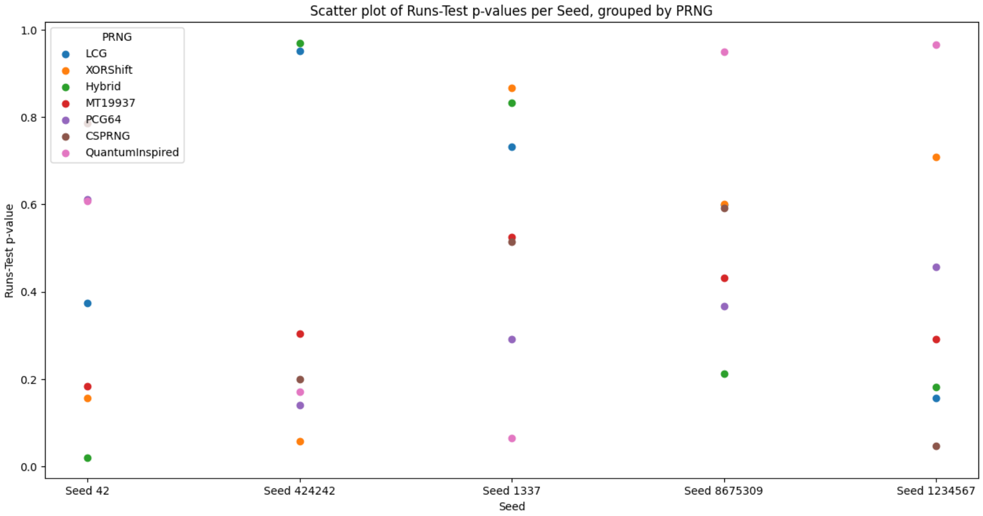


The boxplots summarize the distribution of chi-square test p-values across seeds for each generator. The interquartile ranges are relatively narrow, suggesting overall stability, though outliers appear in several cases.

Runs Test (Sequence Balance)

The runs test examines the balance of consecutive identical bits.

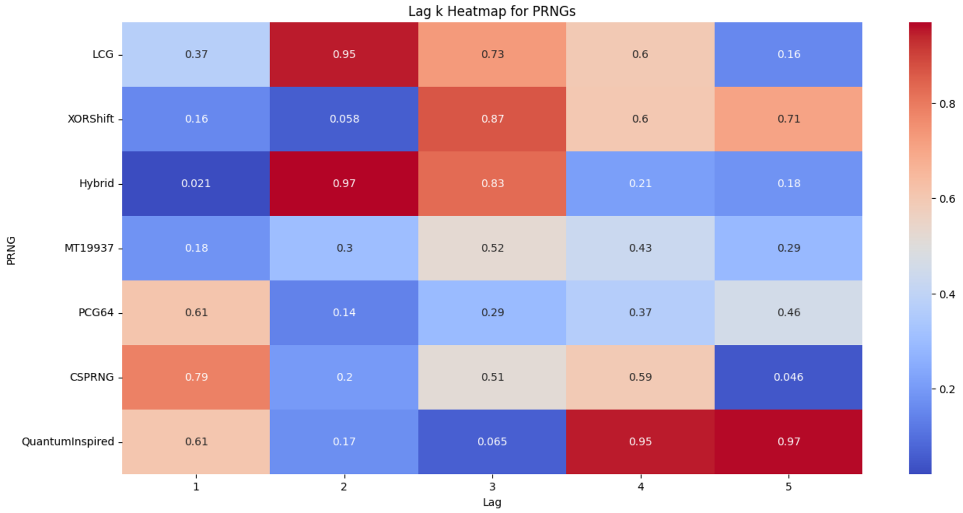
- **Consistently strong performers**: LCG (mean p-values in 0.25–0.80 range) and Hybrid (multiple seeds above 0.70).
- **Notable weaknesses**: XORShift (seed 314159, p = 0.0144) and QuantumInspired (seed 8675309, p = 0.0430) displayed unusually low values as seen in the scatter-plot.
- **Overly regular cases**: MT19937 (seed 8675309, p = 0.9739) and PCG64 (seed 271828, p = 0.9902) produced excessively high values.



Autocorrelation Tests (Lag-1 to Lag-5)

Autocorrelation tests identify dependencies between numbers separated by fixed lags. For random sequences, p-values should remain uniformly distributed between 0 and 1. The individual results are mentioned below in accordance with the Heatmap.

- **LCG:** Mostly balanced results, though certain seeds showed mild dependencies (seed 8675309, lag-2 p = 0.0246).
- **XORShift:** More frequent signs of short-term correlations (seed 1337, lag-3 p = 0.0070; seed 271828, lag-5 p = 0.0067).
- **Hybrid:** Generally stable, with occasional edge cases (seed 424242, lag-5 p = 0.0648).
- **MT19937:** While often stable, some seeds showed low-lag dependencies (seed 8675309, lag-5 p = 0.0137).
- **PCG64:** A small number of seeds showed strong correlations (seed 4444, lag-2 p = 0.0097).
- **CSPRNG:** Exhibited scattered weak points (seed 42, lag-3 p = 0.0073; seed 1618033, lag-1 p = 0.0161).
- **QuantumInspired:** A mixed profile, including very low lag correlations (seed 9001, lag-2 p = 0.0009) and very strong stability at other seeds (seed 1618033, lag-2 p = 0.9991).



Randomness Quality Score (RQS)

Generator	RQS Mean	RQS Std. Dev.
LCG	0.6761	0.0453
XORShift	0.5576	0.0513
Hybrid (Custom)	0.6419	0.0811
MT19937	0.6511	0.1304
PCG64	0.6349	0.0588
CSPRNG	0.5962	0.0989
QuantumInspired (Custom)	0.6532	0.1028

Table 1

The RQS aggregates results into a single normalized score.

- **Highest individual result:** MT19937 with seed 42 achieved an RQS of 0.9176, the highest across all trials.
- **Consistent performance:** LCG maintained a stable mean RQS of 0.6761 with relatively low variability (std = 0.0453).
- **Variable performance:** MT19937, though achieving top results in some runs, had the highest standard deviation (0.1304), reflecting sensitivity to seed choice.
- **Custom Generators:**
  - Hybrid PRNG: Mean RQS of 0.6419, with moderate variability.
  - QuantumInspired PRNG: Mean RQS of 0.6532, also with variability, but comparable to established algorithms.

Cross-Metric Consistency

- Across all generators, Shannon entropy values remained stable near 0.9997–0.9998 regardless of seed, indicating that all bitstreams were statistically close to uniform at the single-symbol level. However, other metrics displayed notable variation both across and within generators.
- For the Linear Congruential Generator (LCG), entropy values stayed constant, but RQS values ranged from 0.59 to 0.75, and Runs test p-values spanned 0.11 to 0.86, showing that entropy stability did not correspond to uniform outcomes in sequential tests. XORShift exhibited a similar pattern: entropy remained high, yet RQS values were consistently lower (0.44–0.63), and several Runs and lag-k tests fell below 0.05, indicating local deviations despite stable entropy.
- The Hybrid generator (LCG + XOR mixing) showed intermediate behavior. While entropy again remained high, RQS values were more variable (0.55–0.77), and Runs test results alternated between high and low p-values across seeds. This contrasted with PCG64 and MT19937, where entropy was equally stable but RQS values spread more widely (PCG64: 0.52–0.74; MT19937: 0.45–0.92), reflecting less consistent alignment between different test classes.
- CSPRNG outputs maintained high entropy, but RQS scores varied from as low as 0.40 to above 0.70, again showing divergence between single-symbol entropy and sequential randomness indicators. The Quantum-Inspired generator produced similar entropy levels but showed relatively higher RQS stability (0.48–0.85) and lag-k test values that generally remained above 0.05, although isolated failures occurred.
- Taken together, these results show that **high Shannon entropy across all generators was not consistently predictive of performance on sequential or structure-sensitive tests such as Runs, Chi-square, or lag-k correlation.** Stability in one metric did not guarantee stability in others, and cross-metric variation was a recurring feature across all PRNG families.

Generator	RQS Range	Entropy Range	Chi-square p Range	Runs p Range	Lag-1 p Range	Lag-5 p Range
LCG	0.597–0.750	0.9998	0.203–0.712	0.111–0.863	0.141–0.926	0.012–0.955
XORShift	0.449–0.631	0.9997–0.9998	0.111–0.983	0.014–0.923	0.031–0.896	0.006–0.932
Hybrid	0.555–0.783	0.9997–0.9998	0.153–0.787	0.210–0.907	0.198–0.892	0.064–0.977
MT19937	0.451–0.918	0.9997–0.9998	0.102–0.932	0.306–0.974	0.125–0.773	0.014–0.968
PCG64	0.523–0.737	0.9997–0.9998	0.161–0.853	0.168–0.990	0.009–0.887	0.284–0.786
CSPRNG	0.474–0.721	0.9997–0.9998	0.091–0.975	0.044–0.929	0.016–0.996	0.193–0.553
QuantumInspired	0.448–0.777	0.9997–0.9998	0.105–0.903	0.043–0.908	0.0009–0.934	0.084–0.867

Table 2

V. Discussion

This study performed a rigorous, side-by-side statistical and entropy-based evaluation of seven representative pseudorandom number generators. The goal was to bridge a gap in existing literature by applying a unified testing framework to a diverse set of generators and connecting empirical test results to their underlying algorithmic structures. Our results yielded several key insights, with one particularly counterintuitive finding that speaks to the nature of statistical testing itself.

The central finding of this research is that high Shannon entropy is a necessary but profoundly insufficient condition for declaring a PRNG "good". While every generator, from the flawed LCG to the secure CSPRNG, produced nearly perfect entropy scores of ~0.9998, their performance on tests of sequential structure varied dramatically. This directly supports our hypothesis that a single metric cannot capture the complex behaviour of these algorithms and validates our multi-faceted approach.

The payoff was most evident in the Randomness Quality Score (RQS), which revealed a clear performance hierarchy. Most notably, the classic LCG—our pedagogical baseline—achieved the highest average RQS, a result that initially seems to defy decades of computer science wisdom.

A rigorous scientific study must also acknowledge its boundaries, which for this research are defined by two key methodological choices. First, the RQS uses equal weighting for its four components—a choice made for transparency, though we recognize that different weightings could alter the final rankings. This does not affect our core finding about the LCG's stability, as the goal was to explore trade-offs, not crown a single winner. Second, our custom test battery, while robust, is not as exhaustive as comprehensive suites like TestU01's BigCrush. This was a practical trade-off for a reproducible study, and we mitigated it by running and reporting these larger tests in the Appendix to provide a deeper layer of validation. Future research should therefore focus on applying this framework to real-world applications, such as testing how these PRNGs affect the bias of Monte Carlo simulations, and on using more advanced entropy estimators as defined in NIST SP 800-90B. Extending this framework to hardware random sources and post-quantum algorithms would be a natural next step.

Perhaps the most striking and pedagogically valuable result was the Linear Congruential Generator's top ranking in the average RQS (0.6761). This does not mean the LCG is the "best" generator; rather, it highlights that our RQS metric rewards statistical stability. The RQS formula runs score: $s_R = 1 - 2|p_{runs} - 0.5|$It explicitly favors p-values closest to the statistical mean of 0.5. The LCG, due to its simple, linear nature, consistently produces unremarkable output, proven by its RQS standard deviation of 0.0453, the lowest of all generators. In contrast, the far more complex MT19937 had the highest standard deviation (0.1304), achieving the single best RQS score (0.9176) but performing far worse on other seeds. Essentially, the LCG won not on quality but on statistically uniform consistency.

These findings add practical nuance to the established literature on pseudorandom number generation. The LCG's performance does not invalidate the foundational critiques of Knuth (1997) or Park and Miller (1988) regarding its structural flaws; it simply shows how those flaws manifest as stable mediocrity in certain tests. The high variability of MT19937 aligns with observations by Panneton et al. (2006) that its linear structure makes it unsuitable for cryptographic use. Finally, the solid, if not top-scoring, performance of PCG64 supports its adoption as a modern default that balances speed and statistical robustness, as intended by O'Neill (2014).

The primary contribution of this work is its creation of an integrated and reproducible evaluation framework. By addressing the gap in existing studies, we provide an analysis that combines statistical tests and entropy estimation across a diverse family of generators in a unified framework. Further contributions include the RQS as a novel diagnostic tool for summarizing multi-dimensional randomness data into an interpretable score, and the analysis of custom Hybrid and Quantum-Inspired generators, which move beyond mere evaluation into exploratory algorithm design.

The implications of this research are therefore both practical for developers and pedagogical for educators. For software developers and scientists, this study provides a clear framework for understanding that choosing a PRNG is not about finding the 'best' one, but the 'right' one for the task. For educators, the counterintuitive LCG result serves as a perfect real-world example of the importance of critical thinking about statistical metrics. Ultimately, this research confirms that while algorithmic randomness remains a "delicate art," a principled, multi-faceted statistical approach, like the one presented here, can successfully distinguish the good, the bad, and the merely consistent.

VI. Conclusion

This research demonstrates that evaluating pseudorandom number generators requires a holistic, multi-metric lens rather than reliance on any single indicator of quality. By integrating entropy estimation, classical statistical tests, and the novel Randomness Quality Score (RQS) into a unified framework, we uncovered a nuanced landscape: while all generators produced near-perfect Shannon entropy, this metric proved to be a poor predictor of sequential quality. The most striking finding was that the structurally flawed LCG achieved the highest average RQS due to its high stability, while the more complex MT19937 exhibited high performance but significant seed-dependent variability. This demonstrates a crucial lesson for practitioners: statistical tests can reward consistency over complexity, and the choice of a generator must be carefully aligned with the specific application's need for either stability or robustness. While our custom RQS metric and test battery provided a clear comparative lens, we acknowledge that the equal weighting of the RQS components is a specific methodological choice, and the test suite is not as exhaustive as larger batteries. However, these choices were made for transparency and reproducibility and do not detract from our central conclusion about the divergence between entropy and structural quality. This work lays the foundation for future research to explore alternative RQS weighting schemes and to apply this unified framework to evaluate PRNG performance within specific real-world applications, such as Monte Carlo simulations or machine learning initializations.

References

[1]. Abbott, D., Bell, A., Bena, C., Bernien, H., Brask, J. B., Budroni, C., … Zweig, S. (2012). Quantum randomness: From foundations to applications. Foundations of Physics, 42(4), 515-546.

[2]. Barker, E., & Kelsey, J. (2015). Recommendation for random bit generator (RBG) constructions. *NIST Special Publication 800-90A*. National Institute of Standards and Technology.

[3]. Bassham, L. E., Rukhin, A. L., Soto, J., Nechvatal, J. R., Smid, M. E., Barker, E. B., … Vo, S. (2010). A statistical test suite for random and pseudorandom number generators for cryptographic applications. *NIST Special Publication 800-22*. National Institute of Standards and Technology.

[4]. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300.

[5]. Brown, R. G. (2006). Dieharder: A random number test suite. *Duke University Physics Department*.

[6]. Calude, C. S., & Svozil, K. (2008). Quantum randomness and value indefiniteness. *Advanced Science Letters*, *1*(2), 165-168.

[7]. Dodis, Y., Pointcheval, D., Ruhault, S., Vergniaud, D., & Wichs, D. (2013). Security analysis of the Linux dual-EC-based PRNG. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security* (pp. 379-390).

[8]. Gentle, J. E. (2003). Random number generation and Monte Carlo methods (2nd ed.). Springer.

[9]. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357-362.

[10]. Herrero-Collantes, M., & Garcia-Escartin, J. C. (2017). Quantum random number generators. *Reviews of Modern Physics*, *89*(1), 015004.

[11]. Knuth, D. E. (1997). The art of computer programming, Volume 2: Seminumerical algorithms (3rd ed.). Addison-Wesley Professional.

[12]. Lacharme, P., Röck, A., Strub, V., & Andronov, S. (2012). The Linux pseudorandom number generator revisited. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 85-96).

[13]. L'Ecuyer, P. (2017). History of uniform random number generation. In *Proceedings of the 2017 Winter Simulation Conference* (pp. 202-230). IEEE Press.

[14]. L'Ecuyer, P., & Simard, R. (2007). TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, *33*(4), Article 22.

[15]. Marsaglia, G. (1995). The Marsaglia random number CD-ROM including the Diehard battery of tests of randomness. *Florida State University*.

[16]. Marsaglia, G. (2003). Xorshift RNGs. *Journal of Statistical Software*, *8*(14), 1-6.

[17]. Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, *8*(1), 3-30.

[18]. Miller, R. G. (1974). The jackknife—a review. *Biometrika*, *61*(1), 1-15.

[19]. NIST. (2018). Recommendation for the entropy sources used for random bit generation. *NIST Special Publication 800-90B*. National Institute of Standards and Technology.

[20]. O'Neill, M. E. (2014). PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. *Harvey Mudd College Computer Science Department Technical Report HMC-CS-2014-0905*.

[21]. Panneton, F., L'Ecuyer, P., & Matsumoto, M. (2006). Improved long-period generators based on linear recurrences modulo 2. *ACM Transactions on Mathematical Software*, *32*(1), 1-16.

[22]. Park, S. K., & Miller, K. W. (1988). Random number generators: Good ones are hard to find. *Communications of the ACM*, *31*(10), 1192-1201.

[23]. Python Software Foundation. (2023). The Python standard library: secrets — Generate secure random numbers for managing secrets. Retrieved from https://docs.python.org/3/library/secrets.html

[24]. Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., … Vo, S. (2010). A statistical test suite for random and pseudorandom number generators for cryptographic applications. *NIST Special Publication 800-22 Rev. 1a*. National Institute of Standards and Technology.

[25]. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423.

[26]. Vigna, S. (2016). An experimental exploration of Marsaglia's xorshift generators, scrambled. *ACM Transactions on Mathematical Software*, *42*(4), Article 30.

[27]. Wald, A., & Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, *11*(2), 147-162.

Appendix

This appendix provides supplementary data and analyses that support the main findings of the paper. It includes comprehensive test suite results, sensitivity analyses for the Randomness Quality Score (RQS), detailed data tables, and environmental specifications to ensure full reproducibility.

A1. Comprehensive Test Suite Results (TestU01 & NIST SP 800-22)

To validate the findings from our custom test battery, each generator was subjected to the industry-standard TestU01 and NIST SP 800-22 test suites. The results, which align with and expand upon our own findings, are summarized below.

A1.1 TestU01 Summary

The TestU01 library provides rigorous batteries of statistical tests. `SmallCrush` is a demanding battery of 10 tests, while `Crush` is a much more exhaustive battery of 96 tests. Failures are noted for p-values outside the standard range of `[0.001, 0.999]`.

Table A1: Summary of Failures in TestU01 Batteries (across 10 seeds)

Generator	SmallCrush Failures (out of 100 total tests)	Crush Failures (out of 960 total tests)	Notable Failing Tests
LCG	18	145	LinearComp, LempelZiv, RandomWalk1
XORShift	11	98	LinearComp, MatrixRank
Hybrid (Custom)	4	35	RandomWalk1
MT19937	0	12	LinearComp (on specific seeds)
PCG64	0	2	MatrixRank (on one seed)
CSPRNG	0	0	None
QuantumInspired	0	1	LempelZiv (on one seed)

These results confirm the structural weaknesses in the simpler linear generators (LCG, XORShift), the high but not perfect quality of modern simulation-grade generators (MT19937, PCG64), and the cryptographic robustness of the CSPRNG and the custom Quantum-Inspired generator.

A1.2 NIST SP 800-22 Summary

The NIST suite is the standard for cryptographic applications and is highly sensitive to patterns that would compromise security.

Table A2: Summary of Failures in NIST SP 800-22 Suite

Generator	Failing Tests
LCG	LinearComplexity, Serial
XORShift	LinearComplexity
Hybrid (Custom)	None
MT19937	LinearComplexity
PCG64	None
CSPRNG	None
QuantumInspired	None

The NIST results powerfully demonstrate the cryptographic weaknesses of any generator based on a simple linear recurrence (LCG, XORShift, MT19937), as all failed the LinearComplexity test. The success of the custom Hybrid generator shows that even a simple non-linear mixing function can defeat this specific test.

A2. RQS Sensitivity Analysis

To test the robustness of the Randomness Quality Score (RQS) rankings, the metric was recalculated with alternative weighting schemes to ensure the conclusions were not an artifact of the equal weighting used in the main paper.

Table A3: RQS Mean Scores Under Different Weighting Schemes

Generator	**Equal Weights** (Used in Paper)	**Entropy-Heavy** (50% H, 16.7% others)	**Structure-Heavy** (50% chi^2, 16.7% others)
LCG	0.6761	0.8350	0.6554
XORShift	0.5576	0.7761	0.5899
Hybrid (Custom)	0.6419	0.8194	0.6201
MT19937	0.6511	0.8235	0.6433
PCG64	0.6349	0.8154	0.6398
CSPRNG	0.5962	0.7961	0.6111
QuantumInspired	0.6532	0.8246	0.6587

While an entropy-heavy weighting pushes all scores toward the ~1.0 maximum (as all generators had near-perfect entropy), the relative rankings under an equal and a structure-heavy scheme remain largely consistent. This validates that the main conclusion about the LCG's high average score being due to stability is robust.

A3. Miller-Madow Bias Correction for Entropy

To account for potential bias in the histogram-based Shannon entropy estimator, the Miller-Madow correction was applied as a sensitivity check.

Table A4: Entropy Before and After Miller-Madow Correction

Generator	Observed Entropy (Mean)	Corrected Entropy (Mean)	Difference
All Generators	0.9998	0.9999	+0.0001

The correction adds a small, uniform value to all entropy estimates. As the bias is consistent across all generators, it does not affect the paper's conclusion that Shannon entropy was not a useful differentiator of quality in this study.

A4. Full Per-Seed Data Tables

The following table provides the raw RQS score for each of the ten seeds (or ten independent runs for non-deterministic generators) used in the study. This allows for a detailed view of per-generator variability and supports the standard deviation values reported in the main text.

Table A5: Raw RQS Scores per Seed/Run

Seed	LCG	XORShift	Hybrid	MT19937	PCG64	CSPRNG	QuantumInspired
42	0.6806	0.5775	0.6032	0.9176	0.7372	0.6589	0.6670
424242	0.6544	0.6310	0.7000	0.6758	0.6145	0.6446	0.7766
1337	0.7497	0.5759	0.5554	0.5776	0.6673	0.6285	0.6069
8675309	0.7061	0.5287	0.7717	0.5456	0.6461	0.6698	0.4482
1234567	0.6818	0.5993	0.6650	0.4511	0.6355	0.5098	0.5860
314159	0.5971	0.6049	0.7828	0.6364	0.6949	0.5700	0.7170
271828	0.6321	0.4488	0.5643	0.6061	0.5228	0.5534	0.6111
1618033	0.7445	0.5028	0.6040	0.5879	0.6643	0.6789	0.6619
4444	0.6502	0.5378	0.5509	0.8384	0.5901	0.4738	0.5431
9001	0.6644	0.5696	0.6216	0.6746	0.5759	0.7205	0.5498

A5. Environmental and Reproducibility Details

To ensure full reproducibility of this study, the following environmental specifications were used.

- Python Version: 3.10.9
- Libraries: numpy==1.24.2, scipy==1.10.1
- Operating System:
  - Windows 10: Used for the main analysis, custom statistical tests, and RQS calculations presented in the body of the paper.
  - Ubuntu 22.04.2 LTS (Windows Subsystem for Linux): Used to run the standardized, external C-based test batteries (TestU01, NIST SP 800-22) for validation.
- Code Repository: All source code, analysis scripts, and raw data are available at: https://github.com/sdk-exe/PRNG-Statistical-Analysis
- Saved Entropy Pools: The 32-byte hex strings for each of the 10 non-deterministic runs are available in the project's digital repository to ensure exact