Empirical Statistical Testing Of Cryptographic PRNGs

Juan Soto

National Institute Of Standards & Technology

soto@nist.gov
Existing Packages

- Stanford University, Donald Knuth
  - Classical Tests
- Florida State University, George Marsaglia
  - DIEHARD
- Queensland University of Technology, Helen Gustafson, Edward Dawson, William Caelli and Lauren Nielsen
  - Crypt-X
- University of Montreal, Pierre L’Ecuyer
  - TestU01 (?)
Project Goals

- The development of a computer package suitable in the assessment of binary stream randomness.
- Applicable to binary streams produced by both hardware and software based PRNGs.
- Warning:
  - No set of statistical tests can certify a generator as appropriate for usage in a particular application.
  - Statistical testing cannot serve as a substitute for cryptanalysis.
Research Team

- The NIST RNG TWG
  - Computer Security Division
    • Miles Smid, James Nechvatal, James Dray, San Vo, Juan Soto
  - Statistical Engineering Division
    • Andrew Rukhin, David Banks, Stefan Leigh, Mark Vangel, Mark Levenson
NIST Test Suite Strengths

- Diverse research team.
- Full scientific documentation provided (each algorithm based on rigorous math).
- More advanced statistical tests.
- Uniform reporting standard (p-value).
Pseudorandom Number Generators

- ANSI X9.17 PRNG (ANSI X9.17)
- FIPS 186 One Way Function Using DES (G-DES)
- FIPS 186 One Way Function Using SHA-1 (G-SHA)
- *Blum-Blum-Shub (BBS)*
- *Micali-Schnorr (MS)*
- Polynomial Congruential (LCG,QCG,CCG)
- Modular Exponentiation (MODEXP)
- Exclusive OR (XOR)
NIST Statistical Test Suite

- Frequency
- Block Frequency
- Cusum
- Runs
- Longest Run Of Ones
- Marsaglia’s Rank*
- Spectral (DFT)
- Template Matchings
- Maurer’s Universal*
- Approximate Entropy
- Random Excursions
- Moving Averages
- Lempel Ziv Complexity
- Linear Complexity*
Evaluation Approaches

- **Analytical**
  - Probability Theory
  - Information Theory
  - Complexity Theory

- **Graphical**
  - Approximate Entropy
  - Spectral Graph
  - Cycle Structure
Evaluation Procedure

- **Null Hypothesis.**
  - Binary stream is random.

- **Compute the test statistic.**
  - Testing is carried out at the bit level.

- **Compute its P-value.**
  - Probability of observing a test statistic at least as extreme as the value actually observed.

- **Compare the P-value to $\alpha$.**
  - **Success** whenever $P$-value $\geq \alpha$. **Failure** otherwise.
  - $\alpha$ is chosen *conservatively* in $(0.001, 0.01]$. 
Numerical Experiments

● **Experiment Parameters**
  – 1,000,000 bits/sequence.
  – 300 binary sequences/generator.

● **PRNGs for which:**
  – flaws were not detected
    • ANSI X9.17, G-DES, G-SHA, BBS, MS, LCG, QCG2
  – flaws were detected
    • QCG1, CCG, XOR, MODEXP
    • Statistically significant results detected at the 0.01 level.
## Pass Rates at 1% Significance Level

<table>
<thead>
<tr>
<th>Statistical Test</th>
<th>G-SHA-1</th>
<th>G-DES</th>
<th>X9.17</th>
<th>BBS</th>
<th>MS</th>
<th>QCG II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>99.67%</td>
<td>99.00%</td>
<td>100.00%</td>
<td>99.00%</td>
<td>99.33%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Block Frequency</td>
<td>99.33%</td>
<td>99.33%</td>
<td>98.67%</td>
<td>100.00%</td>
<td>99.00%</td>
<td>97.67%</td>
</tr>
<tr>
<td>Cusum Forward</td>
<td>99.00%</td>
<td>98.00%</td>
<td>97.67%</td>
<td>97.67%</td>
<td>98.00%</td>
<td>98.00%</td>
</tr>
<tr>
<td>Cusum Reverse</td>
<td>99.33%</td>
<td>97.67%</td>
<td>98.33%</td>
<td>98.33%</td>
<td>98.00%</td>
<td>98.33%</td>
</tr>
<tr>
<td>Runs</td>
<td>98.67%</td>
<td>98.33%</td>
<td>99.67%</td>
<td>99.33%</td>
<td>99.33%</td>
<td>99.67%</td>
</tr>
<tr>
<td>Longest Run Of Ones</td>
<td>98.67%</td>
<td>99.67%</td>
<td>99.67%</td>
<td>99.33%</td>
<td>99.67%</td>
<td>99.33%</td>
</tr>
<tr>
<td>Marsaglia's Rank</td>
<td>98.67%</td>
<td>98.67%</td>
<td>97.67%</td>
<td>100.00%</td>
<td>97.00%</td>
<td>99.33%</td>
</tr>
<tr>
<td>Spectral (DFT)</td>
<td>99.67%</td>
<td>99.33%</td>
<td>99.67%</td>
<td>99.33%</td>
<td>99.33%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Nonoverlapping Template</td>
<td>99.00%</td>
<td>99.33%</td>
<td>99.00%</td>
<td>98.33%</td>
<td>99.00%</td>
<td>99.33%</td>
</tr>
<tr>
<td>Overlapping Template</td>
<td>98.33%</td>
<td>99.33%</td>
<td>98.00%</td>
<td>99.00%</td>
<td>99.67%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Maurer's Universal</td>
<td>98.67%</td>
<td>98.67%</td>
<td>98.67%</td>
<td>99.00%</td>
<td>98.00%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Approximate Entropy</td>
<td>99.00%</td>
<td>98.33%</td>
<td>99.33%</td>
<td>98.67%</td>
<td>100.00%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Random Excursions</td>
<td>99.48%</td>
<td>97.37%</td>
<td>99.48%</td>
<td>100.00%</td>
<td>97.50%</td>
<td>98.91%</td>
</tr>
<tr>
<td>Lempel-Ziv Complexity</td>
<td>99.33%</td>
<td>99.67%</td>
<td>99.67%</td>
<td>99.33%</td>
<td>98.33%</td>
<td>99.67%</td>
</tr>
<tr>
<td>Linear Complexity</td>
<td>98.67%</td>
<td>98.33%</td>
<td>99.33%</td>
<td>98.67%</td>
<td>99.00%</td>
<td>99.00%</td>
</tr>
</tbody>
</table>
Pass Rates at 1% Significance Level

<table>
<thead>
<tr>
<th>Statistical Test</th>
<th>XOR</th>
<th>CCG</th>
<th>MODEXP</th>
<th>QCG I</th>
<th>LCG</th>
<th>BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>99.33%</td>
<td>71.33%</td>
<td>65.00%</td>
<td>58.67%</td>
<td>98.33%</td>
<td>99.33%</td>
</tr>
<tr>
<td>Block Frequency</td>
<td>90.33%</td>
<td>100.00%</td>
<td>99.33%</td>
<td>99.33%</td>
<td>98.67%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Cusum Forward</td>
<td>97.67%</td>
<td>62.67%</td>
<td>58.33%</td>
<td>51.67%</td>
<td>97.67%</td>
<td>98.00%</td>
</tr>
<tr>
<td>Cusum Reverse</td>
<td>99.33%</td>
<td>64.00%</td>
<td>59.00%</td>
<td>51.00%</td>
<td>97.33%</td>
<td>98.33%</td>
</tr>
<tr>
<td>Runs</td>
<td>99.33%</td>
<td>0.00%</td>
<td>99.33%</td>
<td>97.67%</td>
<td>98.33%</td>
<td>98.67%</td>
</tr>
<tr>
<td>Longest Run Of Ones</td>
<td>99.67%</td>
<td>99.00%</td>
<td>99.67%</td>
<td>100.00%</td>
<td>98.67%</td>
<td>99.67%</td>
</tr>
<tr>
<td>Marsaglia's Rank</td>
<td>86.33%</td>
<td>98.33%</td>
<td>98.67%</td>
<td>98.67%</td>
<td>99.67%</td>
<td>98.67%</td>
</tr>
<tr>
<td>Spectral (DFT)</td>
<td>100.00%</td>
<td>83.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>99.33%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Nonoverlapping Template</td>
<td>83.67%</td>
<td>100.00%</td>
<td>98.00%</td>
<td>98.33%</td>
<td>99.00%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Overlapping Template</td>
<td>94.67%</td>
<td>99.67%</td>
<td>99.00%</td>
<td>99.67%</td>
<td>98.67%</td>
<td>99.00%</td>
</tr>
<tr>
<td>Maurer's Universal</td>
<td>68.33%</td>
<td>99.00%</td>
<td>99.00%</td>
<td>98.67%</td>
<td>98.67%</td>
<td>95.00%</td>
</tr>
<tr>
<td>Approximate Entropy</td>
<td>87.67%</td>
<td>0.00%</td>
<td>95.00%</td>
<td>94.33%</td>
<td>99.67%</td>
<td>99.33%</td>
</tr>
<tr>
<td>Random Excursions</td>
<td>98.97%</td>
<td>99.12%</td>
<td>98.26%</td>
<td>100.00%</td>
<td>98.98%</td>
<td>98.95%</td>
</tr>
<tr>
<td>Lempel-Ziv Complexity</td>
<td>99.00%</td>
<td>98.67%</td>
<td>98.67%</td>
<td>99.33%</td>
<td>99.67%</td>
<td>98.33%</td>
</tr>
<tr>
<td>Linear Complexity</td>
<td>0.00%</td>
<td>98.33%</td>
<td>99.67%</td>
<td>99.00%</td>
<td>98.00%</td>
<td>99.67%</td>
</tr>
</tbody>
</table>
Depicts the cycle structure for 3600 binary sequences among 12 PRNGs. Clear discriminant among classes of generators.
Status

- Spring 1998:
  - Release documentation & reference implementation for peer review.

- Summer 1999:
  - Release the statistical test suite and associated documents to the public.

FOR MORE INFO...

http://www.nist.gov/div893/staff/soto/sts.html
Closing Remarks

- Benefits Of Statistical Testing
  - Helps to distinguish between bad PRNGs and good PRNGs.
  - Helps to ensure that the implementation of good PRNGs is in fact producing random looking binary sequences.
  - Helps to evaluate other cryptographic primitives, such as encryption algorithms.
References

