Evaluating Entropy Source Testing Tool Implementation of NIST SP800-90B Min-entropy Estimation Framework

Random Bit Generation Workshop
December 5-6 2012

Evaluating Entropy Source Testing Tool (est)

- Developed by Australian Defence Signals Directorate
- Implements 800-90B statistical tests
- C code, developed under linux
- command line application

Evaluation

- Mapping of files and functions to sections of the standard
- Code review of routines
- Comparison of test results with internal python implementation

Determining if the data is IID:
Shuffling Tests on Independence and Stability
800-90B Section 9.1.2

800-90B section	est function(s)	est file(s)
9.1.2.1 Compression Score	iid_compression()	iid_compression.c
9.1.2.2 Over/Under Runs Scores	iid_over_under_run()	iid_over_under_run.c
9.1.2.3 Excursion Score	iid_excursion()	iid_excursion.c
9.1.2.4 Directional Runs Scores	iid_derivative()	iid_derivative.c
9.1.2.5 Covariance Score	iid_covariance()	iid_covariance.c
9.1.2.6 Collision Score	iid_collision()	iid_collision.c

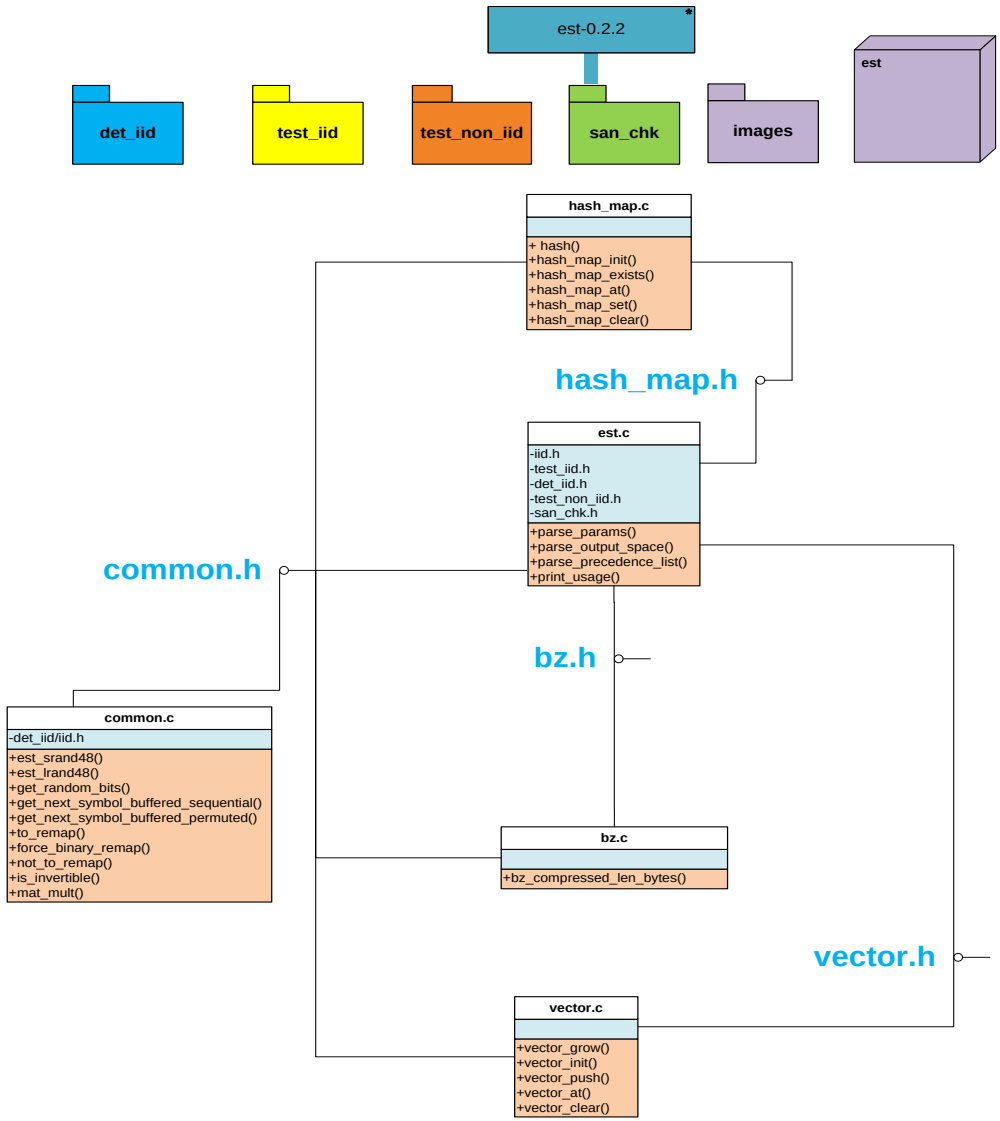
Determining if the data is IID:
Specific Statistical Tests – 800-90B Section 9.1.3

800-90B section	est functions(s)	est file(s)
9.1.3.1 Chi-Square Test	iid_chi_square_tests()	iid_chi.h, iid_driver.c
9.1.3.1.1 Independence test for non-binary data	iid_chi_nb_indep()	iid_chi_nb.c
9.1.3.1.2 Test for Goodness of Fit for Non-Binary Data	iid_chi_nb_fit()	iid_chi_nb.c
9.1.3.1.3 Testing Independence for Binary Data	iid_chi_bin_indep ()	iid_chi_nb.c
9.1.3.1.4 Testing for Stability of Distribution in Binary Data	iid_chi_bin_stab ()	iid_chi_nb.c

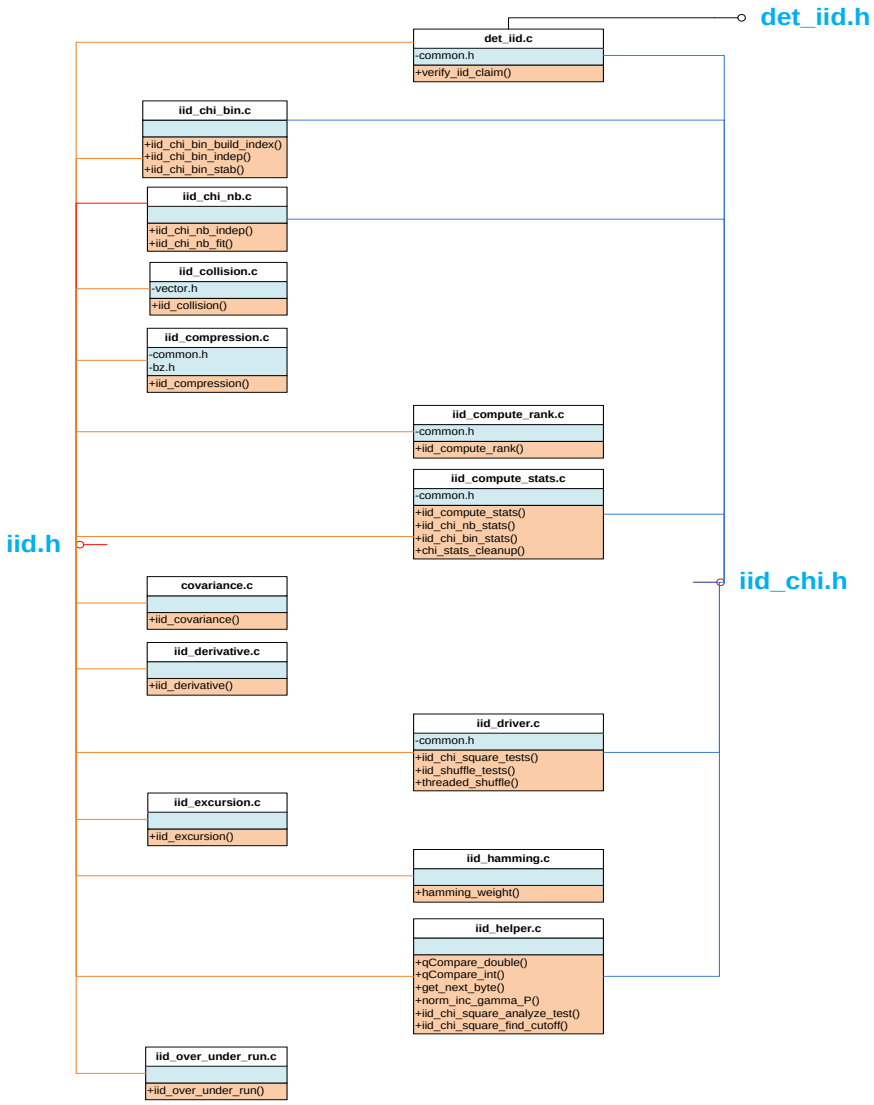
Specific Statistical Tests – 800-90B Section 9.2, 9.3, 9.10

800-90B section	est function(s)	est file(s)
Estimating the Min_Entropy of IID Sources- 800-90B Section 9.2		
9.2 Estimating the Min_Entropy of IID Sources	the_test()	test_iid.h, the_test.c
Estimating the Min_Entropy of non-IID Sources- 800-90B Section 9.3		
9.3.3 The Collision Test	collision_test()	collision_test.c
9.3.4 The Partial Collection Test	partial_collection_test()	partial_collection_test.c
9.3.5 The Markov Test	markov_test()	markov_test.c, test_non_iid.c
9.3.6 The Compression Test	compression_test()	compression_test.c
9.3.7 The Frequency Test	frequency_test()	frequency_test.c
Sanity Checks against Entropy Estimates – 800-90B Section 10		
10.1 Compression Sanity Check	compression_chk()	san_chk.h, compression_chk.c
10.2 Collision Sanity Check	collision_chk()	san_chk.h, collision_chk.c

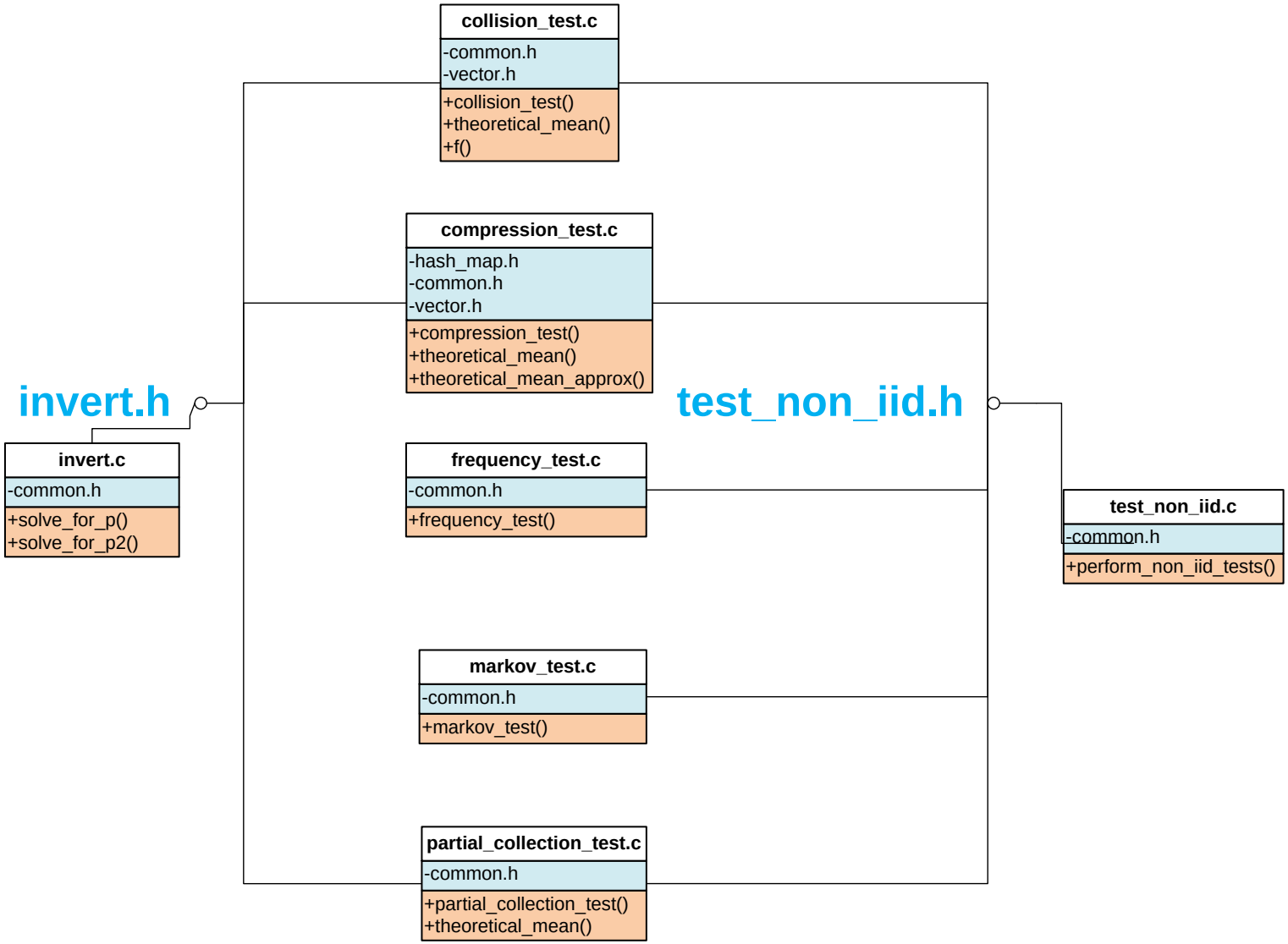
est overview



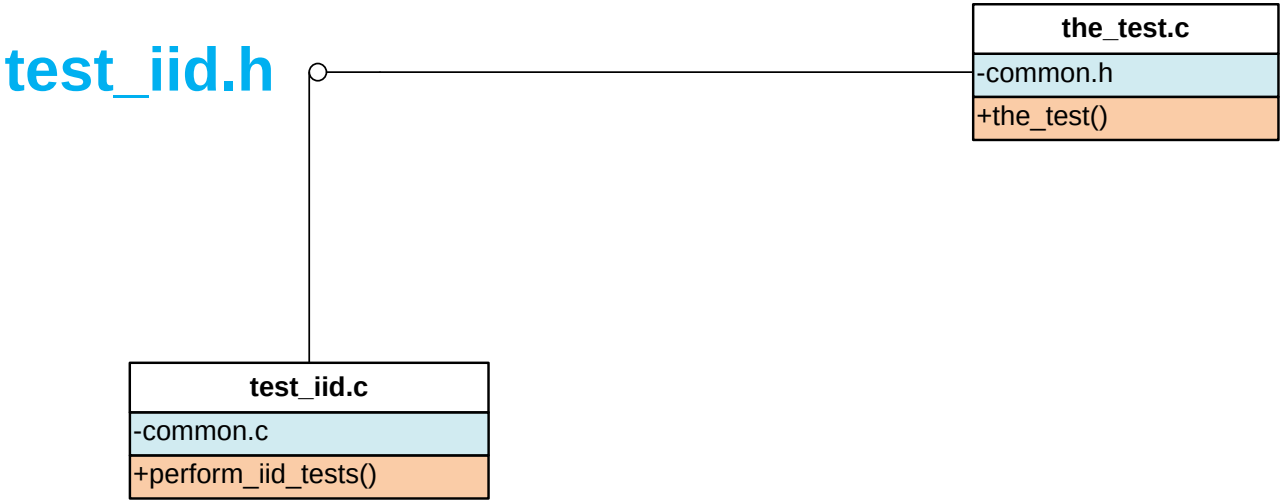
det_iid


det_iid.h
det_iid.c
-common.h
+verify_iid_claim()
iid_chi_bin.c
+iid_chi_bin_build_index()
+iid_chi_bin_indep()
+iid_chi_bin_stab()
iid_chi_nb.c
+iid_chi_nb_indep()
+iid_chi_nb_fit()
iid_collision.c
-vector.h
+iid_collision()
iid_compression.c
-common.h
-bz.h
+iid_compression()
iid_compute_rank.c
-common.h
+iid_compute_rank()
iid_compute_stats.c
-common.h
+iid_compute_stats()
+iid_chi_nb_stats()
+iid_chi_bin_stats()
+chi_stats_cleanup()
iid.h
iid_chi.h
covariance.c
+iid_covariance()
iid_derivative.c
+iid_derivative()
iid_driver.c
-common.h
+iid_chi_square_tests()
+iid_shuffle_tests()
+threaded_shuffle()
iid_excursion.c
+iid_excursion()
iid_hamming.c
+hamming_weight()
iid_helper.c
+qCompare_double()
+qCompare_int()
+get_next_byte()
+norm_inc_gamma_P()
+iid_chi_square_analyze_test()
+iid_chi_square_find_cutoff()
iid_over_under_run.c
+iid_over_under_run()

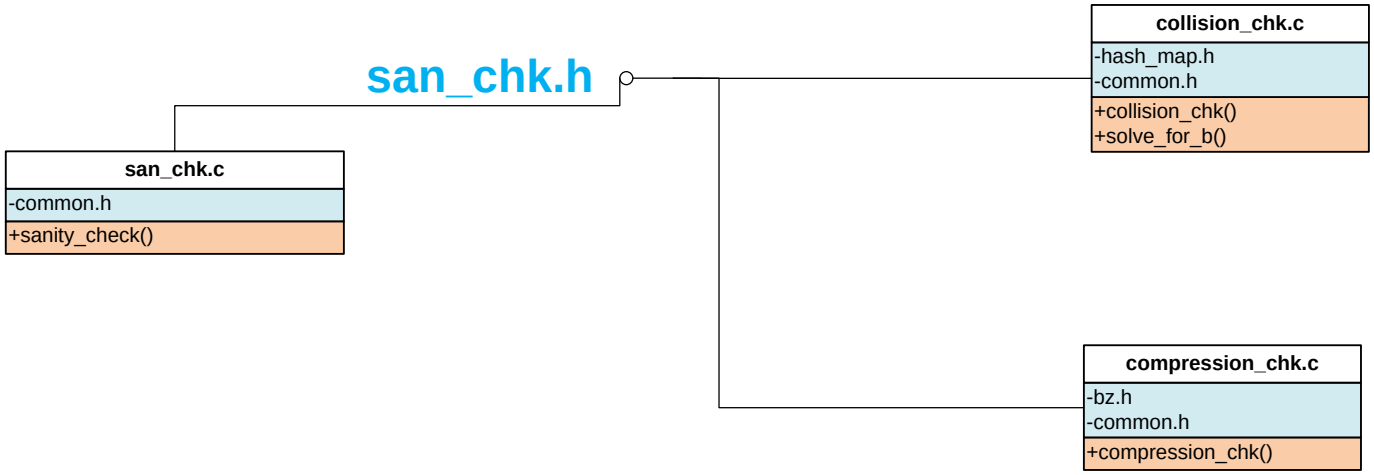
test non-iid


collision_test.c
-common.h
-vector.h
+collision_test()
+theoretical_mean()
+f()
compression_test.c
-hash_map.h
-common.h
-vector.h
+compression_test()
+theoretical_mean()
+theoretical_mean_approx()
invert.h
test_non_iid.h
invert.c
-common.h
+solve_for_p()
+solve_for_p2()
frequency_test.c
-common.h
+frequency_test()
test_non_iid.c
-common.h
+perform_non_iid_tests()
markov_test.c
-common.h
+markov_test()
partial_collection_test.c
-common.h
+partial_collection_test()
+theoretical_mean()

test iid


test_iid.h
the_test.c
-common.h
+the_test()
test_iid.c
-common.c
+perform_iid_tests()

san_chk


san_chk.h
san_chk.c
-common.h
+sanity_check()
collision_chk.c
-hash_map.h
-common.h
+collision_chk()
+solve_for_b()
compression_chk.c
-bz.h
-common.h
+compression_chk()

Comparison with internal implementation

```
# Partial collection test in Draft NIST SP 800-90B (Aug 2012), Section 9.3.4.3
def partial_collection_test(dataset, n):
    # 1. Consider dataset as v non-overlapping data subsets of length n, where
    #    n is the size of the output space
    # 2. Count the number of distinct values seen in each data subset (Ti)
    # 3. Repeat Step 2 until end of dataset is reached. If a minimum of 500
    #    events have not been observed ... (NOTE: not checked here)
    v = len(dataset) // n
    counts = [len(set(dataset[i:i+n])) for i in range(0, v*n, n)]

    # 4. Calculate the sample mean, mu, and the sample std dev, sigma
    mu = sum(counts) / v

    sigma = sum([Ti * Ti for Ti in counts]) / v - (mu * mu)
    sigma = math.sqrt(sigma)

    # 5. Lower bound of confidence interval for the mean
    mu_bar = mu - (1.96 * sigma)/math.sqrt(v)

    # 6. Define one-parameter family of prob distributions...
    # 7. Solve for p s.t. E{Ti} equals mu_bar
    p = solve_for_p(mu_bar, n)

    # 8. The min-entropy is negative log base 2 of p
    min_entropy = -math.log(p, 2.0)

    return p, min_entropy
```

est run on non-iid data

```
[thall@csd bin]$ ./est -v -s 3 -o 2,4,8 -p 1,2,3 -d ../test_data/noniid_3bit.bin -t 20
Performing entropy analysis on 1000000 3-bit symbols from ../test_data/noniid_3bit.bin

Running non-i.i.d. source specific tests ... done

Running sanity checks over min-entropy estimate (1.299970) ... accepted

Non-i.i.d Min-Entropy Estimation Test Results:
Collision test                      1.299970 bits per symbol (from a maximum of 3 bits)
Partial collection test             1.753737 bits per symbol (from a maximum of 3 bits)
Markov test                         1.634499 bits per symbol (from a maximum of 3 bits)
Compression test                    1.667351 bits per symbol (from a maximum of 3 bits)
Frequency test                      2.310948 bits per symbol (from a maximum of 3 bits)

Sanity Check Results:
Compression test                    passed - (285816 bits, 286344 bits, 286520 bits,
286376 bits, 286664 bits, 285992 bits, 285944 bits, 286856 bits, 286928 bits, 286752 bits)
Collision test                      passed - (0 23-symbol values with a count of 3 or
more, 0 colliding 23-symbol values in total)

Final min-entropy is 1.299970 bits per 3-bit symbol.

Tests took 1 second.
```

est run on iid data

```
[thall@csd bin]$ ./est -v -s 3 -o 2,4,8 -p 1,2,3 -d ../test_data/truerand_3bit.bin -i -t 20
Performing entropy analysis on 1000000 3-bit symbols from ../test_data/truerand_3bit.bin

Entropy source claimed as i.i.d. ... confirmed

Running i.i.d. source specific tests ... done

Running sanity checks over min-entropy estimate (2.988550) ... accepted

i.i.d. Determination Results:
Compression score                   passed (3-bit symbols not remapped) - ranks=((676), (105), (952), (127), (562), (8),
(85), (168), (376), (227))
Over/under runs scores              passed (3-bit symbols not remapped) - ranks=((378, 811), (202, 819), (220, 647),
(52, 753), (500, 23), (590, 19), (500, 807), (366, 667), (958, 157), (606, 733))
Excursion score                     passed (3-bit symbols not remapped) - ranks=((647), (341), (757), (285), (310), (6),
(585), (246), (542), (272))
Directional runs scores             passed (3-bit symbols not remapped) - ranks=((366, 500, 85), (393, 240, 666), (373,
947, 819), (635, 500, 528), (241, 500, 566), (10, 260, 319), (14, 790, 733), (7, 251, 79), (8, 500, 275), (132, 241,
798))
Covariance score                    passed (3-bit symbols not remapped) - ranks=((86), (252), (267), (211), (565),
(228), (449), (640), (543), (239))
Collision scores                    passed (3-bit symbols not remapped) - ranks=((500, 23, 500), (500, 21, 500), (500,
31, 500), (500, 650, 500), (500, 747, 500), (500, 13, 500), (500, 8, 500), (500, 452, 500), (500, 447, 500), (500, 487,
500))
Chi-square independence test        passed (3-bit symbols not remapped) - pval=4.796842e-01
Chi-square goodness of fit test     passed (3-bit symbols not remapped) - pval=9.863165e-01

i.i.d Min-Entropy Estimation Test Results:
The test                            2.988550 bits per symbol (from a maximum of 3 bits)

Sanity Check Results:
Compression test                    passed - (311592 bits, 311192 bits, 311328 bits, 311608 bits, 311184 bits, 311448
bits, 311328 bits, 311560 bits, 311424 bits, 311336 bits)
Collision test                      passed - (0 11-symbol values with a count of 3 or more, 0 colliding 11-symbol values
in total)

Final min-entropy is 2.988550 bits per 3-bit symbol.

Tests took 12 minutes and 5 seconds.
```

(Likely) Future Plans

- Continue evaluation as 800-90B nears publication
- Port to Windows to match our other tools
- Develop GUI for ease of use and to assist with reporting requirements

Questions and Discussion

Contact:
tim.hall@nist.gov