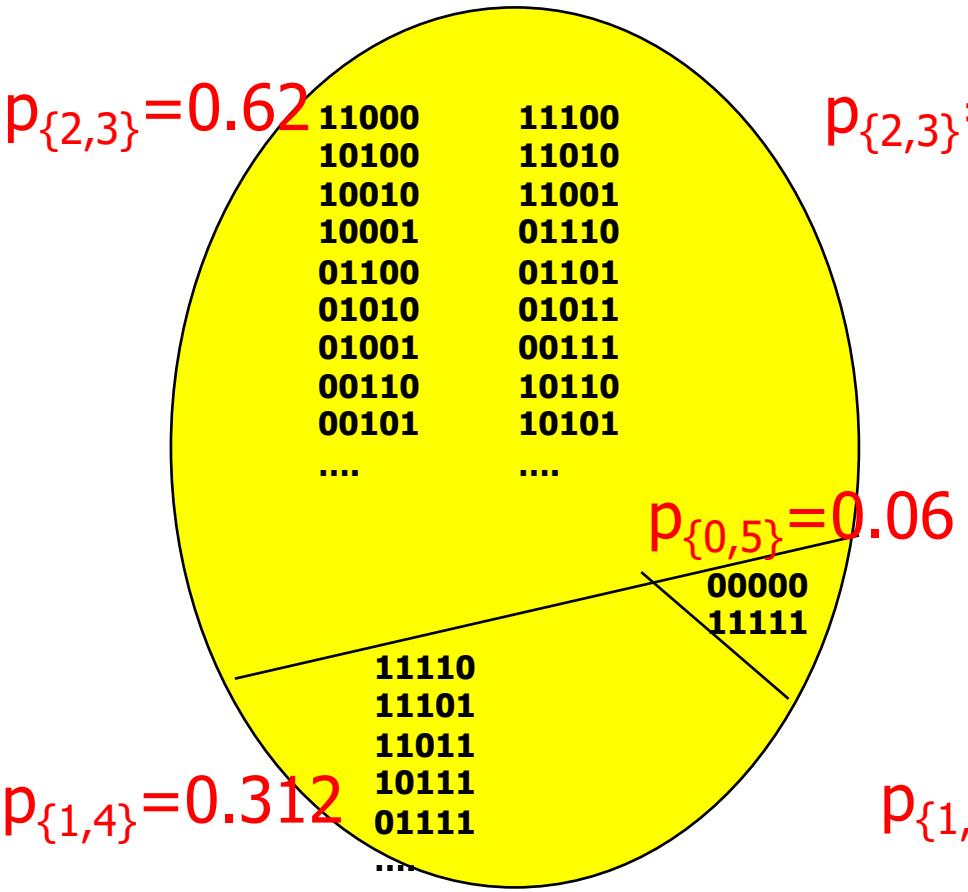
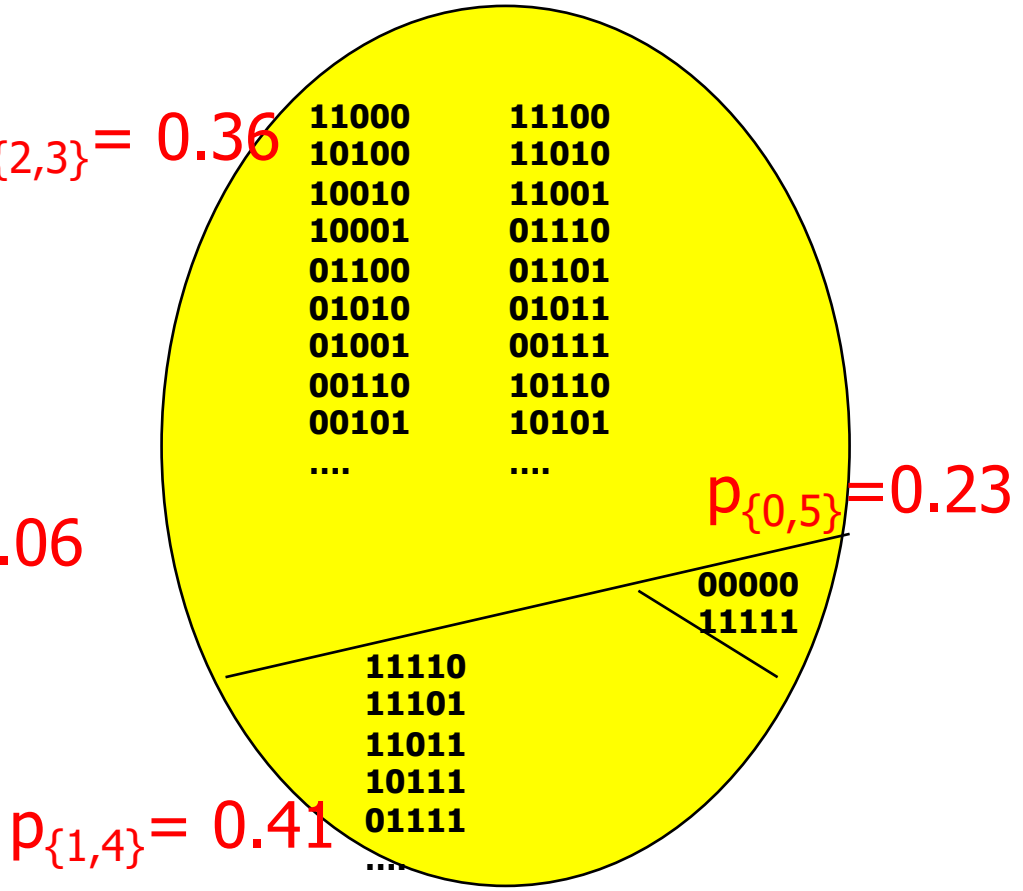
Deciding a sequence is "random"

(generated by a fair coin)

Fair coin



Biased coin: p(0)=3/4



Hypothesis testing &Statistical tests

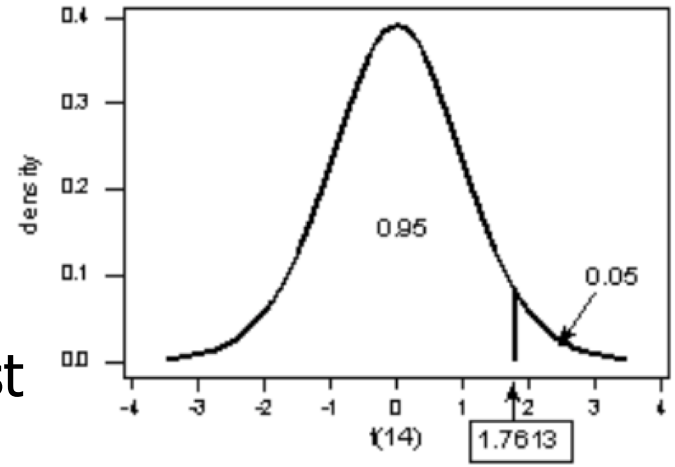
- Given the sequence 00100, which of the two hypotheses hold:

- Null Hypothesis H_0: "the sequence is generated by a fair coin (the generator is a good RNG)"
- Alternative Hypothesis: H_a is "the sequence is not generated by a good generator"

- Use a statistical test to decide:
- Statistical tests rely on test statistics:
  - A "test statistics" is a value that can be calculated for a sequence and has a known distribution under H_0

Statistical tests

- A "test statistics" that,
  - can be calculated for a sequence
  - has a known distribution under H_0
- Choose a significance level α
  - Probability of Type I error
- Using the known distribution of the test statistic, calculate the P-value
- If *P-value* <α, then the null hypothesis is rejected
  - the sequence appears to be random.
- If *P-value* ≥α, the null hypothesis is accepted; i.e.,



NIST Frequency (Monobit) Test

- Calculate test statistics and P-value
- Choose a significance level
- Decide by comparing the p-value with the significance level

For example, if $\varepsilon = 1011010101$, then n=10 and $S_n = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) + 1 + (-1) + 1 = 2$.

(2) Compute the test statistic $s_{obs} = \dfrac{|S_n|}{\sqrt{n}}$

For the example in this section, $s_{obs} = \dfrac{|2|}{\sqrt{10}} = .632455532.$

(3) Compute $P\text{-}value = \textbf{\textit{erfc}}\left(\dfrac{s_{obs}}{\sqrt{2}}\right)$, where ***erfc*** is the complementary error function as defined in Section 5.5.3.

Complementary Error Function

$$erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-u^2}\,du$$

For the example in this section, $P\text{-}value = \textbf{\textit{erfc}}\left(\dfrac{.632455532}{\sqrt{2}}\right) = 0.527089.$

2.1.5 Decision Rule (at the 1% Level)

If the computed *P-value* is < 0.01, then conclude that the sequence is non-random. Otherwise, conclude that the sequence is random.

2.1.6 Conclusion and Interpretation of Results

Since the *P-value* obtained in step 3 of Section 2.1.4 is ≥ 0.01 (i.e., *P-value* = 0.527089), the conclusion is that the sequence is random.

For example, if $\varepsilon = 10$
$(-1) + 1 = 2$.

(2) Compute the test statis

For the example in this

(3) Compute *P-value* = *er*

Section 5.5.3.

For the example in this

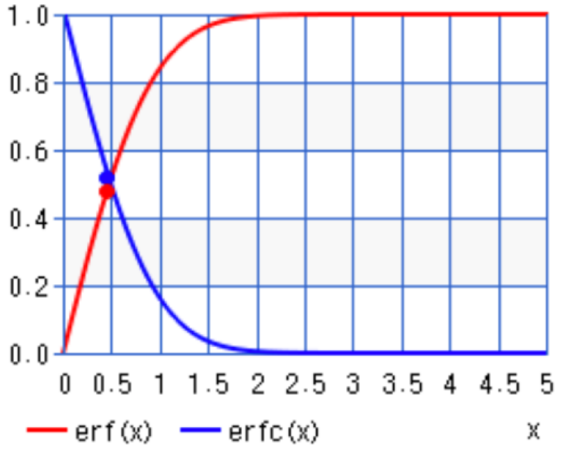
2.1.5 Decision Rule (at th

If the computed *P-value* is < 0.
that the sequence is random.

Error function Calculator

Home / Special Function / Error function

Calculates the error function erf(x) and complementary error function erfc(x).



x 0.632/1.4 real number

Execute Clear Store/Read Print 22digit

Error function	result
● erf(x)	0.476797347254205418 1176
● erfc(x)	0.523202652745794581 8824

2.1.6 Conclusion and Interpretation of Results

Since the *P-value* obtained in step 3 of Section 2.1.4 is ≥ 0.01 (i.e., *P-value* = 0.527089), the conclusion is
that the sequence is random.

NIST Statistical Test Suite

- A statistical framework:

- Randomness is a probabilistic property; that is, the properties of a random sequence can be characterized and described in terms of probability. The likely outcome of statistical tests, when applied to a truly random sequence, is known a priori and can be described in probabilistic terms. There are an infinite number of possible statistical tests, each assessing the presence or absence of a "pattern" which, if detected, would indicate that the sequence is nonrandom. Because there are so many tests for judging whether a sequence is random or not, no specific finite set of tests is deemed "complete." In addition, the results of statistical testing must be interpreted with some care and caution to avoid incorrect conclusions about a specific generator (see Section 4).

Decision using Hypothesis Testing

- A statistical test is formulated to test a specific null hypothesis (H0). For the purpose of this document, the null hypothesis under test is that the sequence being tested is random. Associated with this null hypothesis is the alternative hypothesis (Ha), which, for this document, is that the sequence is not random.

- For each applied test, a decision or conclusion is derived that accepts or rejects the null hypothesis, i.e., whether the generator is (or is not) producing random values, based on the sequence that was produced."

Statistical hypothesis tests define a procedure that controls (fixes) the probability of incorrectly deciding that a default position (null hypothesis) is incorrect. The procedure is based on how likely it would be for a set of observations to occur if the null hypothesis were true. wikipedia

Summary

- Random number generation
  - PRNG
    - Statistical uniformity
    - Unpredictability
      - Computational
  - TRNG
- Tests for evaluating randomness of a string