

Statistical Testing of Randomness

(Yesterday, Today, and Possibly Tomorrow)



Jan Krhovják



BUSLab & LaBAK

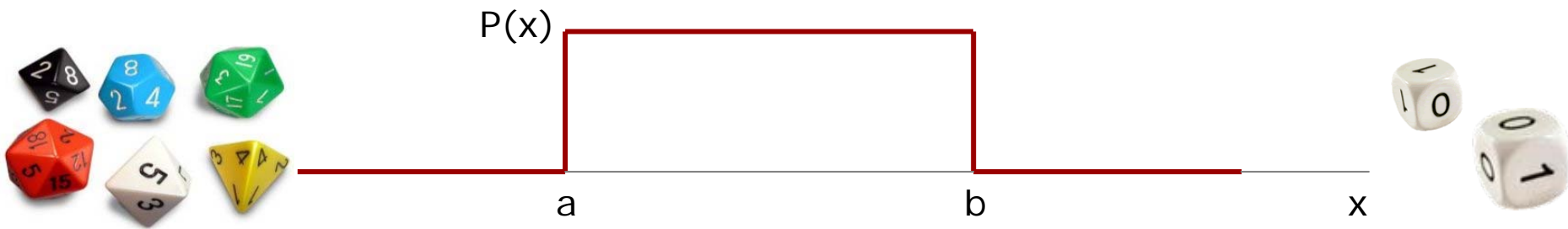
Faculty of Informatics, Masaryk University, Brno

LaBAK & KD Lab Seminar, Cikhákj, Spring 2007

Randomness and Unpredictability



- A random number is
 - an instance of an unbiased random variable
 - the output produced by a uniformly distributed random process (follows a probability distribution)
- Random sequence is
 - a sequence of random variables
 - properties as sample drawn from uniform (rectangular) distribution
- Uniform (rectangular) continuous distribution
 - $P(x) = 1/(b-a)$ for $a \leq x \leq b$; 0 otherwise



Basic Idea Behind the Statistical Testing

- Statistical testing of random sequences
- Particular tests are based on test statistic
 - Expected value of some test statistic is known for the reference distribution
 - Generated random stream is subjected to the same test
 - Obtained value is compared against the expected value
- Boundless number of statistical test can be constructed
 - Some of them are accepted as the de facto standard
 - NIST battery consists of 15 such tests (e.g. frequency test)
 - Generators that pass such tests are considered “good”
 - Absolute majority of generated sequences must pass

Statistical Hypothesis Testing (Basics)

- Null hypothesis (H_0) – denotes test hypothesis
 - H_0 = the sequence being tested is random
- Alternative hypothesis (H_A) – negates H_0
 - H_A = the sequence is not random
- Each test is based on some test statistic (TS)
 - TS is quantity calculated from our sample data
 - TS is random variable/vector obtained from transformation of random selection
 - TS has mostly standard normal or chi-square (χ^2) as reference distributions
- After each applied test must be derived conclusion that rejects or not rejects null hypothesis

Statistical Hypothesis Testing (Errors)

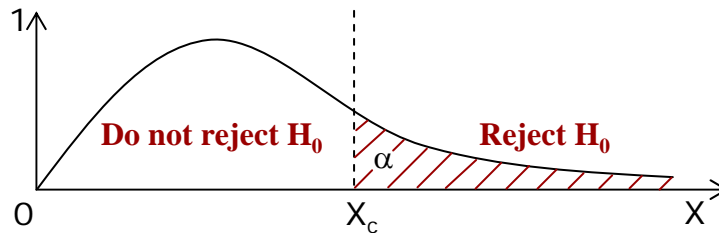
- Conclusion generation procedure and errors

Real situation	Conclusion	
	H_0 is not rejected	H_0 is rejected
H_0 is true	good decision	type I error
H_0 is not true	type II error	good decision

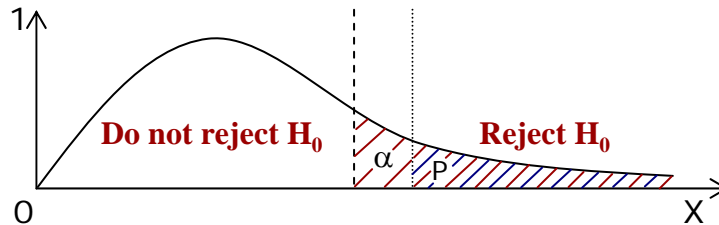
- Probability of type I error (α) = level of significance
 - Set prior the test; typically between 0.0001 and 0.01
 - Nonrandom sequence, produced by “good” generator
- Probability of type II error (β)
 - Random sequence, produced by “bad” generator
- α and β are related to each other and to sample size

Statistical Hypothesis Testing (Evaluation)

- Critical value – threshold between rejection and non-rejection regions
- Two (quite similar) ways of testing
 - $\alpha \Rightarrow$ critical value; test statistic \Rightarrow value; compare



- Test statistic \Rightarrow P-value; α ; compare



- NIST battery uses P-values
 - $P \leq \alpha \Rightarrow$ reject H_0
 - $P > \alpha \Rightarrow$ do not reject H_0

Frequency (Monobit) Test

- Basic idea
 - Number of zeros and ones expected in the truly random sequence should be the same
- Description
 - Length of the bit string: n
 - Sequence of bits: $\varepsilon = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$
 - Test statistic: $s_{\text{obs}} = |S_n|/\sqrt{n}$
 - $S_n = X_1 + X_2 + \dots + X_n$, where $X_i = 2\varepsilon_i - 1$ (conversion to ± 1)
 - The absolute value \Rightarrow half normal distribution
- Example (for $n = 10$)
 - $\varepsilon = 1011010101$; $S_n = 1-1+1+1-1+1-1+1-1+1 = 2$
 - $s_{\text{obs}} = |2|/\sqrt{10} = 0.632455532$; P-value = 0.5271
 - For $\alpha = 0.01$: $0.5271 > 0.01 \Rightarrow \varepsilon$ "is random"

NIST Testing Strategy

1. Select (pseudo) random number generator
2. Generate sequences
 - a) Generate set of sequences or one long sequence
 - i. Divide the long sequence to set of subsequences
3. Execute statistical tests
 - a) Select the statistical tests
 - b) Select the relevant input parameters (how?)
4. Examine (and analyse) the P-values
 - a) For fixed α a certain percentage are expected to failure
5. Assign Pass/Fail

Interpretation of Empirical Results

- Examination of the proportion of sequences that pass the statistical test
 - Determined by confidence interval
 - $p' \pm 3 \cdot \sqrt{(p' \cdot (1-p')/n)}$, where $p' = 1 - \alpha$; n is sample size
 - If proportion falls outside \Rightarrow data are non-random
- Check for uniformity of the distribution of obtained P-values
 - Level of significance $\alpha = 0.0001$ (hardwired)
 - Test statistic: $\chi^2 = \sum (o_i - e_i)^2 / e_i$
 - o_i is observed number of P-values in i^{th} subinterval
 - e_i is expected number of P-values in i^{th} subinterval
 - Sample size multiplied by probability of occurrence in each subinterval (i.e. for sample size n it is $n/10$)
 - Total P_T -value is calculated and compared to α
 - P_T -value $> 0.0001 \Rightarrow$ sequence is uniformly distributed

Machine Learning and Random Numbers

- Some problems of classical approach
 - Length of subsequences
 - Parameters setting
 - Independency of tests
- Can machine learning (ML) help?
 - Improved/unified statistical testing
 - Learning of existing ML algorithms should fail
 - Design of new ML algorithm for finding of regularities
 - Visualization techniques
 - Predictions and regression methods
 - Genetic programming
 - Support vector machines