Exercise 3

29. Oktober 2018

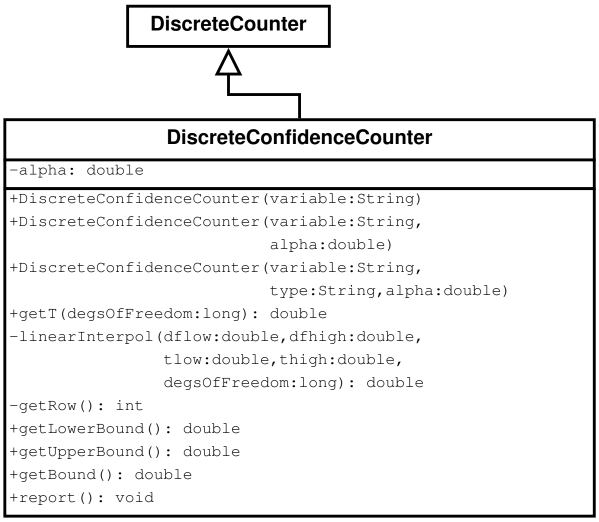
Abgabe: 5. November 2018, 12:00 Uhr

Problem 3.1: *Implementation and test of confidence intervals*

This problem provides an implementation for confidence intervals based on the Student-t and the Normal distribution.

1. What is the meaning of a two-sided confidence interval with $\alpha = 0.05$? Provide the $t_{n-1,(1-\alpha/2)}$ values for $n = 2, 3, 11, 101$. 5 Points

2. Provide an implementation in the subclass *DiscreteConfidenceCounter* of the *Discrete-Counter* class which calculates the lower and upper bound of a confidence interval for the mean of a given set of samples! 25 Points



Unless otherwise specified, a default value of 0.05 for *alpha* should be chosen.

getT should return the $t_{n-1,(1-\alpha/2)}$ value for a given number n of samples (see Section 4.4 in the course syllabus). Mind the connection between the number of samples and the degrees of freedom. The class already provides an array with the quantiles for the Student-t[1] [2] distribution for $alpha = \{0.01, 0.05, 0.10\}$ and
$df = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000000\}$.

[1] http://people.richland.edu/james/lecture/m170/tbl-t.html
[2] https://en.wikipedia.org/wiki/Student's_t-distribution#Table_of_selected_values

As arbitrary values for *alpha* and *degsOfFreedom* can be chosen, additional means must be provided:

- If the quantile contains an exact value for a given *alpha* and *degsOfFreedom* (e.g. for $alpha = 0.01$ and $degsOfFreedom = 100$), this value can be returned directly.
- If *degsOfFreedom* is out of the values range, the maximum number for *degsOfFreedom* in the quantile array should be chosen.
- *getRow* should select the most suitable quantile row for a value of *alpha*. Choose the quantiles for 0.01 if $alpha < 0.05$, the quantiles for 0.05 if $0.05 \leq alpha < 0.1$, and the quantiles for 0.1 otherwise!
- If *degsOfFreedom* is between two entries in the array, *linearInterpol* should return a linearly interpolated value.

getBound should be implemented using Formula 4.8 (see Section 4.4 in the course syllabus), which is used by *getLowerBound* and *getUpperBound*. *report* should output *alpha*, *t* as well as the lower and upper bound on the console.

3. Generate samples distributed according to a Normal distribution (see provided code) with $mean = 10$ and $c_{var} \in \{0.25, 0.5, 1, 2, 4\}$! Choose sample sizes of $n_{samples} \in \{5, 10, 50, 100\}$ items and calculate the 90% and 95% confidence intervals for the mean of these samples. Perform these experiments $n_{experiments} = 500$ times and calculate the fraction of experiments whose confidence intervals contain the real mean! Organize the results in a table and comment on them. See *src/study/DCCTest.java*! — 20 Points

4. Perform the same experiments with the k-Erlang, Exponential, and Hyperexponential distribution where appropriate! Organize the results in a table and comment on them. See *src/study/DCCTest.java*! — 20 Points

Problem 3.2: *Composed Random Variables*

Provide derivations, either by showing the complete calculation path or by using the scaling properties of random variables (see Section 1.3.6 in the course syllabus). If chosen latter, shortly refer to the applied principle of scaling!

1. Derive $E[X]$ and $VAR[X]$ of the bernoulli and binomial distribution. — 10 Points
2. Given are $E[X] = {}^{1-p}/_{p}$ and $VAR[X] = {}^{1-p}/_{p^2}$ of the geometric distribution. Derive $E[X]$ and $VAR[X]$ for the negative binomial distribution with parameter s! — 10 Points
3. Given are $E[X] = {}^{1}/_{\lambda}$ and $VAR[X] = {}^{1}/_{\lambda^2}$ of the exponential distribution. Derive $E[X]$, $VAR[X]$ and $c_{var}[X]$ for the k-Erlang distribution! — 10 Points

Total: 100 Points