Bitte gesamtes Projekt abgeben

1	2	Σ
55	30	85/100

Tobias Hille tobias.hille@student.uni-tuebingen.de 3905597
Robin Schmidt rob.schmidt@student.uni-tuebingen.de 4255055

Modellierung & Simulation I

Serie 03

Problem 3.1.1

A two-sided confidence interval means, that with a probability of $1-\alpha$ (confidence level) a sample Z_n is within the interval $[-z_{1-\frac{\alpha}{2}}; +z_{1-\frac{\alpha}{2}}]$. In comparison to a one-sided confidence interval, bounds are desired above and below the sample mean. For $\lim\limits_{n\to\infty}$ the standard normal distribution is used, while for a small number of samples the Student-t distribution is used. The following values were collected using a suitable table (`https://en.wikipedia.org/wiki/Student_t-distribution#Table_of_selected_values`), where the degrees of freedom correspond to $n-1$:

Student-t distribution	
n	$t_{n-1,(1-\alpha/2)}$
2	12.706
3	4.303
11	2.228
101	1.984

✓ 5/5

Problem 3.1.2

please see provided code ✓ 25/25

Problem 3.1.3

Counter wird nicht resettet! -10

Table 1: Normalverteilung

(a) $c_{var} = 0.25$

		α	
		0.1	0.05
samples	5	1.0	1.0
	10	1.0	1.0
	50	0.98	1.0
	100	1.0	1.0

(b) $c_{var} = 0.5$

		α	
		0.1	0.05
samples	5	0.926	1.0
	10	0.994	1.0
	50	0.966	1.0
	100	1.0	0.998

(c) $c_{var} = 1$

		α	
		0.1	0.05
samples	5	1.0	1.0
	10	0.718	0.938
	50	0.996	1.0
	100	0.654	0.998

(d) $c_{var} = 2$

		α	
		0.1	0.05
samples	5	0.996	0.848
	10	1.0	0.81
	50	0.872	1.0
	100	0.956	0.986

(e) $c_{var} = 4$

		α	
		0.1	0.05
samples	5	1.0	1.0
	10	1.0	1.0
	50	0.988	0.952
	100	0.77	0.59

Discussion:
With a different seed (other than 2) the values are quite different. Because of this, outliers like $c_{var} = 1, \alpha = 0.1, samples = 10$ or $c_{var} = 2, \alpha = 0.1, samples = 50$) can be associated with the noise that is created by only using 500 runs per configuration. The majority of the other results following

the mathematical properties of the interaction between the Student-t and normal distribution. That is, for small sample sizes and normal distributed random variables the confidence intervals of the student-t distribution estimate the true mean.

10/20

Problem 3.1.4

these configurations are not possible:
$c_{var} = 0.25$ (unsupportedOperationException)
$c_{var} = 0.5$ (unsupportedOperationException)
$c_{var} = 2$ (unsupportedOperationException)
$c_{var} = 4$ (unsupportedOperationException)

Table 2: Exponentialverteilung

(a) $c_{var} = 1$

		α	
		0.1	0.05
samples	5	0.938	1.0
	10	0.976	1.0
	50	0.862	1.0
	100	0.974	1.0

Table 3: k-Erlangverteilung

(a) $c_{var} = 0.25$

		α	
		0.1	0.05
samples	5	0.998	1.0
	10	0.998	0.996
	50	0.916	0.948
	100	1.0	1.0

(b) $c_{var} = 0.5$

		α	
		0.1	0.05
samples	5	0.112	0.996
	10	0.774	0.992
	50	0.856	1.0
	100	0.996	1.0

(c) $c_{var} = 1$

		α	
		0.1	0.05
samples	5	0.972	0.998
	10	0.972	0.986
	50	0.966	0.638
	100	1.0	0.948

(d) $c_{var} = 2$

		α	
		0.1	0.05
samples	5	0.0	0.0
	10	0.0	0.0
	50	0.0	0.0
	100	0.0	0.0

(e) $c_{var} = 4$

		α	
		0.1	0.05
samples	5	0.0	0.0
	10	0.0	0.0
	50	0.0	0.0
	100	0.0	0.0

It is not possible to generate a hyperexponential distribution with the given mean and covariance values (unsupportedOperationException). −5

Discussion: If another seed (than 2) is used, the values fluctuate a lot. Because of this, the outliers (i.e. Exponential $c_{var} = 1, \alpha = 0.1, samples = 50$ or k-Erlang $c_{var} = 1, \alpha = 0.05, samples = 50$) can be associated with the noise that is created by only using 500 runs per configuration. The fraction of "correct" confidence intervals in the other simulation is with the exception of k-Erlang distribution with $c_{var} = 2$ and $c_{var} = 4$ almost always near 1. The latter can be explained by the long tail of the distribution which causes great fluctuation of the observed mean, such that the higher and lower bound do not contain the real mean. 15/20

Problem 3.2.1

For a bernoulli distributed random variable X with $P(X = 1) = p$ and $P(X = 0) = q$ the expected value is:

$$\begin{aligned} E(X) &= \sum_i i \cdot x(i) \\ &= P(X = 1) \cdot 1 + P(X = 0) \cdot 0 \\ &= 1p + 0q \\ &= p \end{aligned}$$

For a bernoulli distributed random variable X with $P(X = 1) = p$ and $P(X = 0) = q$ the variance is:

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= 1^2 p + 0^2 q - (1p + 0q)^2 \\ &= p - p^2 \\ &= p \cdot (1 - p) \\ &= pq \end{aligned}$$

Because the binomial distribution consists of n bernoulli experiments, the derivation of the expected value and the variance need to be multiplied by the factor n like:

$$\begin{aligned} E(X) &= n \cdot \sum_i i \cdot x(i) \\ &= n \cdot (P(X=1) \cdot 1 + P(X=0) \cdot 0) \\ &= n \cdot (1p + 0q) \\ &= n \cdot p \quad \checkmark \end{aligned}$$

10/10

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= n \cdot p - (n \cdot p)^2 \\ &= n \cdot p(1-p) \\ &= n \cdot pq \quad \checkmark \end{aligned}$$

Problem 3.2.2

The negative binomial distribution is the sum of s random variables, which are geometrically distributed:

$$NegBin(s,p) \sim \sum_{i=0}^{s} Geom(p)$$

Using the scaling properties of random variables (2.27c) and the expected value of the geometric distribution with $E_{geo}(X) = \frac{1-p}{p}$ leads us to the following:

$$\begin{aligned} E_{neg}(X) &= \sum_{i=1}^{s} E_{geo}(Y_i) \\ &= \sum_{i=1}^{s} \frac{1-p_i}{p_i} \\ &= s \cdot \frac{1-p}{p} \quad \checkmark \end{aligned}$$

Because $p_1 = p_2 = p_3 = ... = p_s$ the expected value for the negative binomial distribution is $E_{neg}(X) = s \cdot \frac{1-p}{p}$.

The derivation for the variance uses the scaling properties of random variables (2.27d) and a similar approach with the variance of the geometric distribution

$Var_{geo}(X) = \frac{1-p}{p^2}$:

$$\begin{aligned} Var_{neg}(X) &= \sum_{i=1}^{s} Var_{geo}(Y_i) \\ &= \sum_{i=1}^{s} \frac{1-p_i}{p_i^2} \\ &= s \cdot \frac{1-p}{p^2} \quad \checkmark \end{aligned}$$

10/10

Problem 3.2.3

The k-Erlang distribution is the sum of k random variables, which are exponentially distributed. Using the scaling properties of random variables (2.27c) and the expected value of the exponential distribution with $E_{exp}(X) = \frac{1}{\lambda}$ leads us to the following:

$$\begin{aligned} E_{erl}(X) &= \sum_{i=1}^{k} E_{exp}(Y_i) \\ &= \sum_{i=1}^{k} \frac{1}{\lambda} \\ &= \frac{k}{\lambda} \quad \checkmark \end{aligned}$$

The derivation for the variance uses the scaling properties of random variables (2.27d) and a similar approach with the variance of the exponential distribution $Var_{exp}(X) = \frac{1}{\lambda^2}$:

$$\begin{aligned} Var_{erl}(X) &= \sum_{i=1}^{k} Var_{exp}(Y_i) \\ &= \sum_{i=1}^{k} \frac{1}{\lambda^2} \\ &= \frac{k}{\lambda^2} \quad \checkmark \end{aligned}$$

Using the general formular for the covariance $c_{var} = \frac{\sigma}{E(X)}$ and the derived solutions from above shows:

$$c_{var} = \frac{\sigma_{erl}}{E_{erl}(X)} = \frac{\sqrt{Var_{erl}(X)}}{E_{erl}(X)}$$

$$= \frac{\sqrt{\frac{k}{\lambda^2}}}{\frac{k}{\lambda}}$$

$$= \frac{\sqrt{k} \cdot \frac{1}{\lambda}}{k \cdot \frac{1}{\lambda}}$$

$$= \frac{\sqrt{k}}{k} = \frac{1}{\sqrt{k}}$$ √

10/10