ADVANCED WORKBOOK

Task for: DLMDSAS01– Advanced Statistics

Note on copyright and plagiarism:
Please take note that IU Internationale Hochschule GmbH holds the copyright to the examination tasks. We expressly object to the publication of tasks on third-party platforms. In the event of a violation, IU Internationale Hochschule is entitled to injunctive relief. We would like to point out that every submitted written assignment is checked using a plagiarism software. We therefore suggest not to share solutions under any circumstances, as this my give rise to the suspicion of plagiarism.

The workbook bases on the output of the parameter generator. The produced numbers are ξ_1 to ξ_{20} and are used in the assignment. Only start the assignment after generating the numbers (which is done personally).

Task 1: Basic Probabilities and Visualizations (1)

Please provide the requested visualization as well as the numeric results. In both cases, please provide how you realized these (calculations, code, steps…) and why it is the appropriate tools. Do not forget to include the scale of each graphics so a reader can read the numbers represented.

- If ξ_1 is 0: A vote with outcome for or $against$ follows a Bernoulli distribution where $P(\text{vote} = "for") = \xi_2$. Represent the proportion of “for” and “against” in this single Bernoulli trial using a graphic and a percentage. Can an expectation be calculated? Justify your answer by all necessary hypotheses.

- If ξ_1 is between 1 and 3:
  The number of meteorites falling on an ocean in a given year can be modelled by one of the following distributions. Give a graphic showing the probability of one, two, three… meteorites falling (until the probability remains provably less than 0.5% for any bigger number of meteorites). Calculate the expectation and median and show them graphically on this graphic:
  - If ξ_1 is 1: a Poisson distribution with an expectation of $\lambda = \xi_2$
  - If ξ_1 is 2: a negative binomial distribution with number of successes of $k = \xi_2$ und $p = \xi_3$.
  - If ξ_1 is 3: a geometric distribution counting the number of Bernoulli trials with $p = \xi_2$ until it succeeds.

Task 2: Basic Probabilities and Visualizations (2)

Let Y be the random variable with the time to hear an owl from your room’s open window (in hours). Assume that the probability that you still need to wait to hear the owl after y hours is one of the following:

- If ξ_4 is 0: the probability is given by $\xi_5 \mathrm{e}^{-\xi_6\, y} + \xi_7 \mathrm{e}^{-\xi_8\, y}$
- If ξ_4 is 1: the probability is given by $\xi_5 \mathrm{e}^{-\xi_6\, y^2} + \xi_7 \mathrm{e}^{-\xi_8\, y^8}$
- If ξ_4 is 2: the probability is given by $\xi_5 \mathrm{e}^{-\xi_6 \sqrt{y}} + \xi_7 \mathrm{e}^{-\xi_8 \sqrt[3]{y}}$
- If ξ_4 is 3: the probability is given by $\xi_5 \mathrm{e}^{-\xi_6\, y^2} + \xi_7 \mathrm{e}^{-\xi_8\, y^2}$

Find the probability that you need to wait between 2 and 4 hours to hear the owl, compute and display the probability density function graph as well as a histogram by the minute. Compute and display in the graphics the mean, variance, and quartiles of the waiting times.
Please pay attention to the various units of time!

Task 3: Transformed Random Variables

A type of network router has a bandwidth total to first hardware failure called S expressed in terabytes. The random variable S is modelled by a distribution whose density is given by one of the following functions:

- (if $\xi_9 = 0$): $f_S(s) = \frac{1}{\theta} e^{-\frac{s}{\theta}}$
- (if $\xi_9 = 1$): $f_S(s) = \frac{1}{24\theta^5} s^4 e^{-\frac{s}{\theta}}$
- (if $\xi_9 = 2$): $f_S(s) = \frac{1}{\theta}$ for $s \in [0, \theta]$

with a single parameter θ. Consider the bandwidth total to failure T of the sequence of the two routers of the same type (one being brought up automatically when the first is broken).
Express T in terms of the bandwidth total to failure of single routers S_1 and S_2. Formulate realistic assumptions about these random variables. Calculate the density function of the variable T.
Given an experiment with the dual-router-system yielding a sample $T_1, T_2, \ldots, T_n$, calculate the likelihood function for θ. Propose a transformation of this likelihood function whose maximum is the same and can be computed easily.

An actual experiment is performed, the infrastructure team has obtained the bandwidth totals to failure given by the sequence ξ_{10} of numbers. Estimate the model-parameter with the maximum likelihood and compute the expectation of the bandwidth total to failure of the dual-router-system.

Task 4: Hypothesis Test

Over a long period of time, the production of 1000 high-quality hammers in a factory seems to have reached a weight with an average of ξ_{11} (in g) and standard deviation of ξ_{12} (in g). Propose a model for the weight of the hammers including a probability distribution for the weight. Provide all the assumptions needed for this model to hold (even the uncertain ones)? What parameters does this model have?

One aims at answering one of the following questions about a new production system:

- (if $\xi_{13} = 0$): Does the new system make *more constant* weights?
- (if $\xi_{13} = 1$): Does the new system make *lower* weights?
- (if $\xi_{13} = 2$): Does the new system make *higher* weights?
- (if $\xi_{13} = 3$): Does the new system make *less constant* weights?

To answer this question a random sample of newly produced hammers is evaluated yielding the weights in ξ_{14}.

What hypotheses can you propose to test the question? What test and decision rule can you make to estimate if the new system answers the given question? Express the decision rules as logical statements involving critical values. What error probabilities can you suggest and why? Perform the test and draw the conclusion to answer the question.

Task 5: Regularized Regression

Given the values of an unknown function $f: \mathbb{R} \to \mathbb{R}$ at some selected points, we try to calculate the parameters of a model function using OLS as a distance and a ridge regularization:

- (if $\xi_{15} = 0$): a polynomial model function of thirteen α_i parameters:
  $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{12} x^{12}$

- (if $\xi_{15} = 2$): a polynomial model function of eleven α_i parameters:

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{10} x^{10}$$

Calculate the OLS estimate, and the OLS ridge-regularized estimates for the parameters given the sample points of the graph of f given that the values are y = ξ_{16}.
Provide a graphical representation of the graphs of the approximating functions and the data points.
Remember to include the steps of your computation which are more important than the actual computations.

Task 6: Bayesian Estimates

(following Hogg, McKean & Craig, exercise 11.2.2)
Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a gamma distribution with $\alpha = 3$ and $\beta = 1/\theta$. Suppose we believe that θ follows a gamma-distribution with $\alpha = \xi_{17}$ and $\beta = \xi_{18}$ and suppose we have a trial $(x_1, \ldots, x_n)$ with an observed $\overline{x} = \xi_{19}$.

a) Find the posterior distribution of θ.
b) What is the Bayes point estimate of θ associated with the square-error loss function?
c) What is the Bayes point estimate of θ using the mode of the posterior distribution?