INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

WORKBOOK

Assignments for: DLMDSAS01– Advanced Statistics

WORKBOOK ASSIGNMENT 1: BASIC PROBABILITIES AND VISUALIZATIONS (1)

For each of the distributions below, please provide the requested graphics 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.

a) A vote with outcome for or $against$ follows a Bernoulli distribution where $P(\text{vote} = "for") = 0.27$. Represent the proportion of "for" and "against" in this single Bernoulli trial using a graphics and a percentage. Can an expectation be calculated? Justify your answer.

b) The number of meteorites falling on an ocean in a given year can be modelled by a Poisson distribution with an expectation of $\lambda = 37$. Explain why a Poisson distribution is a natural candidate for this phenomenon. Give a graphic showing the probability of one, two, three… meteorites falling (until the probability is less than 0.5%). Calculate the median and variance and show them graphically on this graphic.

c) 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:

$$0.\overline{3}\mathrm{e}^{-0.5\,y} + 0.\overline{6}\mathrm{e}^{-0.25\,y}$$

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.

WORKBOOK ASSIGNMENT 2: BASIC PROBABILITIES AND VISUALIZATIONS (2)

For each of the distributions below, please provide the requested graphics 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.

a) Consider the variables X and Y. The realization of a sample of size 20 is given below (where X is the first variable and Y is the second):
$(-1.202, 563.024)$, $(2.112, 291.072)$, $(2.827, -893.619)$, $(-0.314, 1321.814)$, $(-1.477, -91.573)$, $(-6.516, 446.336)$, $(-0.920, -111.487)$, $(3.477, -153.165)$, $(-7.273, 1076.221)$, $(2.251, 477.931)$, $(-0.713, 909.696)$, $(-0.853, 226.865)$, $(-3.176, 389.413)$, $(1.913, -47.169)$, $(-1.070, -178.695)$, $(-3.385, 744.486)$, $(-9.506, 362.670)$, $(-7.004, 364.578)$, $(0.504, 324.975)$, $(2.861, -360.571)$
Sketch an appropriate plot that displays the values of these points. Calculate the sample covariance as well as the sample's expectations and variances of X and Y.

b) Consider that a ball is thrown with a random angle $\theta \in [0, 360)$ (in degrees) and a random radius $r \in [0, 1]$ (in meters) both independent and uniform. Calculate the density of the variables X and Y (the cartesian coordinates of the point at angle θ and radius r) as well as their expectation and variance.

WORKBOOK ASSIGNMENT 3: A SIMPLE PARAMETER ESTIMATION

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 an exponential distribution whose density is given by:

$$s(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}}$$

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 following bandwidth totals to failure:

$$9.2, 5.6, 18.4, 12.1, 10.7$$

Estimate the model-parameter with the maximum likelihood and compute the expectation of the bandwidth total to failure of the dual-router-system.

WORKBOOK ASSIGNMENT 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 $971g$ and standard deviation of $15.2\ g$. Propose a model for the weight of the hammers including a probability distribution for the weight. What are the assumptions for this model to hold? What parameters does this model have?

A new production system is configured, and one wants to evaluate if the new system makes *more constant* weights. For this a random sample of newly produced hammers is evaluated yielding the following weights:

$$987, 966, 955, 977, 981, 967, 975, 980, 953, 972$$

What hypothesis can you formulate and what test and decision rule can you make to estimate if the new system produces a more constant weight? Express these assertions as logical statements involving critical values. What error probabilities can you suggest and why? Calculate the p-value. Perform the test and express conclusions.

WORKBOOK ASSIGNMENT 5: SUFFICIENT STATISTICS

(following Larsen & Marx, exercise 5.6.5)
Let $X_1, X_2, \ldots, X_n$ be the random sample of a positive random variable X having the density function:

$$f_X(x;\ \theta) = \frac{1}{5!\ \theta^6} x^5 e^{-\frac{x}{\theta}}$$

with one parameter $\theta \in \mathbb{R}^+$. Find an estimator for θ that is sufficient.

WORKBOOK ASSIGNMENT 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 $\alpha = 3$ and $\beta = 2$:

a) Find the posterior distribution of θ.
b) If the observed $\overline{x} = 18.2$, what is the Bayes point estimate associated with the square-error loss function?
c) What is the Bayes point estimate using the mode of the posterior distribution?

All the best with your workbook!