Exercise set 6

Textbook:	Exercises 9.13 (9.15 in 8. edition) and 9.76.
Note on maximum likelihood:	Exercise 1

Exercise 1: Continuation of exercise 2, exercise set 4.

Recall exercise 2 exercise set 4. In that exercise we did show that the best unbiased estimator was: $\hat{\mu} = \frac{6}{7}\bar{X} + \frac{1}{7}\bar{Y}$.

a) Start with the estimator $\hat{\mu}$ and derive a 95% confidence interval for μ.
Calculate a numerical answer when the measurements gave $\bar{x} = 6.12$ and $\bar{y} = 6.05$.

Exercise 2: Continuation of exercise 3, exercise set 4.

Recall exercise 3 on exercise set 4.

a) Use the estimator $\hat{\mu}$ as starting point, and derive a 95% confidence interval for the oil concentration μ. (Tip: Remember that X_1 and X_2 follow a normal distribution and that σ_1 and σ_2 are assumed known.)
Calculate the interval numerically when $x_1 = 4.41$, $x_2 = 4.19$, $\sigma_1 = 0.5$ and $\sigma_2 = 0.2$.

Exercise 3:

For all kinds of sea structures, wave loads need to be taken into account. Interest is typically centred around the maximum wave height that could occur during a certain period of time.

In this exercise we shall consider a situation where a certain operation needs to be done at an oil installation in the North Sea. The operation takes 3 hours and requires that the maximum wave height during the operation must not exceed 4.0 meters.

To decide whether it is safe to start the operation n measurements of wave heights, $X_1, \ldots, X_n$, are done. We assume that the wave situation is constant during the time interval the measurements are done such that the wave heights can be assumed to be random variables from the same distribution. We further assume that this distribution is

$$f_X(x) = \frac{2x}{\theta} e^{-x^2/\theta} \quad , x > 0.$$

This distribution is called a *Rayleigh*-distribution and is often used to model wave heights.

a) Assume in this point that $\theta = 3$. What is the probability that a randomly chosen wave is having a wave height larger than 4 meters?
What is the probability p that two consecutive waves both are having a wave height larger than 4 meters if we assume that the wave heights are independent?
If the heights of two consecutive waves are positively correlated, how will this influence the value of p? (Argue whether p will increase, decrease or be unchanged.)

In practise the parameter θ is unknown and shall be estimated from the measurements $X_1, \ldots, X_n$, which we now assume are independent.

b) Show that the maximum likelihood estimator of θ, $\hat{\theta}$, is given by: $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

c) Use transformation results to explain that $Z = 2X^2/\theta$ has a χ_2^2-distribution. Find $\mathrm{E}(\hat{\theta})$ and $\mathrm{Var}(\hat{\theta})$.

d) Use the first result from point c) as starting point and explain what the distribution of $2n\hat{\theta}/\theta$ becomes.
Derive a $(1-\alpha)100\%$ confidence interval for θ. Calculate the interval when $\alpha = 0.05$ and the observations give

$$n = 50, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i = 0.911, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 1.047$$

e) Use the result from point d) to derive and calculate a 95% confidence interval for μ where $\mu = \mathrm{E}(X) = 0.5\sqrt{\theta\pi}$

f) Calculate the approximate 95% Wald confidence interval for θ. Compare with the exact interval in d) and comment.

Let $Y = \max(X_1, \ldots, X_m)$ be the largest of m assumed independent wave heights in the forthcoming 3 hour interval.

g) Find an expression for the cumulative distribution function of Y, $F_Y(y)$.
Use this expression to show that if y_c is the value such that

$$P(Y > y_c) = \frac{1}{100}$$

then

$$y_c = \sqrt{-\theta \ln\left[1 - \left(1 - \frac{1}{100}\right)^{1/m}\right]}.$$

Assume that $m = 900$. The operation is only started if the measurements give basis to claim that $y_c < 4$. One way to decide this is to calculate a 90% confidence interval for y_c and only start the operation if the upper limit of this interval is less than 4. (Actually, this is equivalent to doing a hypothesis test on 5% level of the claim that $y_c < 4$.)

h) Do the measurements reported in point d) give basis to claim that $y_c < 4$? Decide this by making a 90% confidence interval for y_c.

<u>**Some answers:**</u>

9.13/9.15: [47.72, 49.28]; **9.76**: [1.12, 2.33]

Note 1 a) $\hat{\beta} = \bar{X}$; b) $\mathrm{E}(\hat{\beta}) = \beta$ and $\mathrm{Var}(\hat{\beta}) = \beta^2/n$; c) $(\hat{\beta} - z_{\alpha/2}\hat{\beta}/\sqrt{n}, \hat{\beta} + z_{\alpha/2}\hat{\beta}/\sqrt{n})$;
d) [157,2391] and [622,3924]; e) [921,1627] and [983,1716]

1: a) [6.05, 6.17]; **2**: a) [3.86,4.58] ;

3 a) 0.0048 and 0.000023; c) θ and θ^2/n; d) [0.81, 1.41] ; e) [0.80, 1.05]; f) [0.76, 1.34];
g) $[1 - e^{y^2/\theta}]^m$; h) [3.10, 3.91], yes