TMA4245 Statistics
Exam August 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Suppose that the total change in global mean temperature over the next 50 years can be written as a sum $W = X + Y$ where $X \sim N(0, \sigma_X^2)$ is the change as a result of natural fluctuations in the climate and $Y \sim N(\mu, \sigma_Y^2)$ is the change resulting from anthropogenic CO_2-emissons. The change as a result of natural fluctuation thus has a variance of σ_X^2 but a zero expected value. The parameters μ and σ_Y^2 represents the expectation and variance associated with the effect of anthropogenic influences based on climate models. We assume that X and Y are independent stochastic variables. Given that CO_2-emissions are kept at the present level, we know that $\mu = 2°\text{C}$ og at $\sigma_Y = 1.5°\text{C}$. In addition, we assume that $\sigma_X = 0.5°\text{C}$.

a) Find the probability that $X > 1°\text{C}$ and the probability that the total change in mean temperature becomes larger than $5°\text{C}$.

Suppose that the cost of damages associated with climate change are at its minumum provided that the mean temperature remains unchanged ($W = 0$) and that both an increase and a decrease in global mean temperature will be costly. To model this we approximate the annual total costs resulting from a change W in mean temperature by a second order Taylor polynomial $g(W) = aW^2$. Assume that $a = 1.1 \cdot 10^{12}$ USD/$°C^2$.

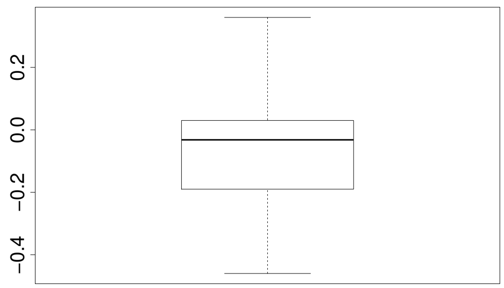
b) Find the expected value of the annual damage costs $g(W)$ as a result of a change in mean temperature W expressed as a function a, μ, σ_X and σ_Y.

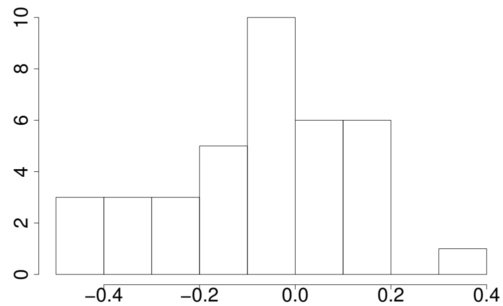
Suppose that we through a reduction in CO_2-emissons by a factor of two from the present level are able reduce μ and σ_Y, also by a factor of two. By how much does this reduce the expected damage costs? Does the reduction depend on the magnitude of natural climate fluctuations?

Oppgave 2

In the world championship in single distance speed skating, each skater does two runs, one with the inner lane in the last curve and one with the outer lane in the last curve. The final position of each skater is decided based on the total time used in the two runs. A similar rule is employed in the Olympic games. The rule was introduced at the world championship in Hamar in 1995. Before this, each skater did only one run. The rationale behind the rule is a small average advantage of finishing in the outer lane; negotiating the last curve at high speed is typically more difficult in the inner than in the outer lane.

In a championship with n skaters, let Y_i and Z_i denote the time for skater i for the runs with





Figur 1: Boxplot and histogram of the differences d_i for the 37 skaters in 500 meter for men in Sochi that did not fall. In the boxplot the differences are along the second axis, and in the the histogram along the first axis.

the last curve in the outer and inner lane respectively. Additionally, let X be the number out of the n skaters that have their fastest time in the run with the last outer lane, that is, X is the number of skaters for which $Y_i < Z_i$. We assume that X has a binomial distribution, that is, $P(X = x) = b(x; n, p)$ where $p = P(Y_i < Z_i)$.

a) List which assumptions that must be met in the situation discussed above if the assumption of X following a binomial distribution is to be true.

For $n = 20$ and $p = 0.7$, find the probabilities

$$P(X \leq 10) \quad \text{and} \quad P(X \geq 8 | X \leq 10).$$

b) Write down the likelihood function for p and use this to show that the maximum likelihood estimator of p is

$$\widehat{p} = \frac{X}{n}.$$

Show that $\widehat{p}$ is an unbiased estimator of p and that $\text{Var}(\widehat{p}) = p(1-p)/n$.

In the following we shall use the results from the 500 meter event in the Olympic games in Sochi in Russia in February 2014 to examine if there is any reason to believe in a last outer lane advantage. In this event there were $n = 39$ skaters that completed both runs, and $x = 24$ out of these had their fastest time in the run with last outer lane.

In derivations that follow, you may use approximations if necessary. If so, justifications for these must be given.

c) Formulate a hypothesis test for the above situation. Specify H_0 and H_1, choose a suitable test statistic, and derive a decision rule given that the level of significance is $\alpha = 0.05$.

What is the conclusion of the test based on the results from Sochi?

Also compute the p-value of the test based on the Sochi results.

In the remaining part of this problem, we shall denote the above test as Test 1. An alternative test for the same situation, denoted Test 2, can be formulated by defining the differences

$$D_i = Y_i - Z_i \quad \text{for } i = 1, 2, \ldots, n,$$

where Y_i and Z_i as before are the times of skater i in the run with last curve in outer and inner lane respectively. A test can then be constructed using $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ as a starting point.

d) Formulate a hypothesis test for the above situation based on $\bar{D}$. Specify H_0 and H_1, choose a suitable test statistic, and derive a decision rule given that the level of significance is α.

Out of the 39 skaters in Sochi, two fell in one of their runs. If we discard these results, the remaining results for Sochi gives $n = 37$, $\sum_{i=1}^{37} d_i = -2.654$ and $\sum_{i=1}^{37} d_i^2 = 1.552$. Based on these data, what is the conclusion of Test 2 when $\alpha = 0.05$? If necessary, round the degrees of freedom to the nearest value given in the table.

The power of Test 2 (the probability of rejecting H_0 given that H_1 is correct) equals 0.67 for $n = 37$, $E(D_i) = -0.07$ and $\text{Var}(D_i) = 0.2^2$. These values for $E(D_i)$ and $\text{Var}(D_i)$ are estimates obtained from the results for the 37 skaters who did not fall. Note also that these values of $E(D_i)$ and $\text{Var}(D_i)$ imply $p = P(Y_i < Z_i) = 0.64$ provided that each D_i follows a normal distribution.

e) Find the power of Test 1 for $n = 39$ and $p = 0.64$.

Figure 1 shows a boxplot and a histogram of d_i for the 37 skaters in Sochi who did not fall. Based on the conclusions of Test 1 and Test 2, their power, and the plots in Figure 1, discuss how you would conclude overall in the given situation. The two skaters in the Sochi who fell, both fell in the run in which they had the last inner curve. Does this have any implications for your conclusion?

Oppgave 3

To assess the accuracy of a new measuring procedure, n measurements are made of the same uknown quantity. Let $X_1, X_2, \ldots, X_n$ denote the resulting measurements, and assume that these represent a random sample from a normal distribution with expected value μ and standard deviations σ. Here, our interest is in the value of σ, but we will assume that μ is also unknown.

Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. From the curriculum it is then known that $S^2(n-1)/\sigma^2$ follows a chi-square distribution with $n-1$ degrees of freedom.

a) Derive a $100(1-\alpha)\,\%$ confidence interval for σ^2.

Derive a $100(1-\alpha)\,\%$ confidence interval also for σ.

b) Suppose that Y follows a chi-square distribution with v degrees of freedom such that the probability density function is given by

$$f(y) = \frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})} y^{\frac{v}{2}-1} e^{-\frac{y}{2}}.$$

Show that

$$E(\sqrt{Y}) = \frac{\sqrt{2}\,\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})}.$$

c) Use the result in the previous point to examine whether

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

is a biased estimator of σ. If so, suggest a new estimator of σ corrected for bias.

An estimator $\hat{\theta}$ that over- and underestimates θ with equal probabilities is referred to as median-unbiased. Thus, for continuously distributed median-unbiased estimators we have $P(\hat{\theta} \leq \theta) = 1/2$. Suggest a median-unbiased estimator of σ.

Fasit

1. **a**) 0.023,0.029 **b**) $5.11 \cdot 10^{12}$

2. **a**) 0.048, 0.98 **c**) Accept H_1, 0.075 **d**) Reject H_0 **e**) 0.545