MEK4350, fall 2016
Exercises XI

We have seen that for our toy model for the surface elevation at a point

$$\eta(t) = A\cos(\omega_p t) + B\sin(\omega_p t)$$

where A and B are independent identically distributed normal stochastic variables with mean 0 and variance σ^2, we can use the expression $\tilde{S}(\omega_l) = |\hat{\eta}(\omega_l)|^2$ as an estimator for the spectrum $S(\omega_l)$. Here we suppose that $\eta(t)$ is periodic on an interval $0 \leq t < T$, and we have $\omega_l = 2\pi l/T$ and $\omega_p = 2\pi p/T$ where l and p are integers. From Exercises X we know that $\tilde{S}$ has $\chi^2(2)$ distribution, which is the same as the exponential distribution. The estimator is unbiased, i.e. $E[\tilde{S}(\omega_l)] = S(\omega_l)$. However, the standard deviation of the estimator is equal to the spectrum itself $\sqrt{\mathrm{Var}[\tilde{S}(\omega_l)]} = S(\omega_l)$, which means that the quality of the estimate is terrible, some averaging or smoothing will be needed.

Problem 1

Let $\bar{X} = \frac{1}{N}\sum_{n=1}^{N} X_n$ be an estimator for the mean based on N measurements X_n. Suppose that the individual measurements X_n are statistically independent with mean μ and variance σ^2.

Recall that an estimator ($\bar{X}$) is said to be unbiased if the expected value is equal to the true value ($E[\bar{X}] = \mu$).

Is $\bar{X}$ a biased or unbiased estimator for μ?

Compute the variance of $\bar{X}$.

How big should N be in order that the estimate of the mean has standard deviation less than one tenth of the standard deviation of each individual measurement?

Hint 1: This problem is easily done using characteristic functions.

Hint 2: You may assume that all the X_n are Gaussian, but maybe that is not necessary?

Problem 2

Matlab and Python have estimators for standard deviation (std), variance (var), skewness and kurtosis. These estimators can be called in biased or unbiased forms, by means of an optional OPT or FLAG which can be either 0 or 1 or omitted for default value.

Find out how your preferred system behaves by default for these four estimators! Do you consider this behavior to be logical or expected or reasonable?