1. (6P) Distribution of the mean for Normal distributed random variables.
   Show: If $\bar{X}$ is the mean of a random sample of size $n$ from a Normal distribution with mean $\mu$ and variance $\sigma$, then $\bar{X}$ is Normal distributed with mean $\mu$ and variance $\sigma^2/n$.
   Hint: Use the characteristic function.
   Remark: This theorem was used several times in class, without proof so far.

2. (6P) Parameter errors for the multivariate distribution.
   For the bi-variate distribution with covariance matrix
   \[
   C = \begin{pmatrix} \sigma_{X_1}^2 & \rho \sigma_{X_1} \sigma_{X_2} \\ \rho \sigma_{X_1} \sigma_{X_2} & \sigma_{X_2}^2 \end{pmatrix},
   \]
   show analytically that the error in parameter $X_1$ is given by $\sigma_{X_1}$. Parameter error is defined as in the lecture, i.e., the uncertainty in the parameter irrespectively of what values the other parameters have.

3. (15P) Confidence regions for the multivariate distribution.
   Show in two ways: The quantity $y = (\theta - \mu)^T C^{-1} (\theta - \mu)$ with $m$-dimensional random vector $\theta$ is $\chi^2_m$ distributed.
   a) (5P) Proof by diagonalizing $C$ suitably.
   b) (7P) Proof by calculating the characteristic function of the random variable $y$ and compare to the characteristic function of the $\chi^2$ distribution (derive the latter as well).
   c) (3P) Calculate the change in $\chi^2$ required to determine the 95% confidence region in a 5-parameter fit.
   Remark: We assume here that the covariance matrix $C$ and mean vector $\mu$ are precisely known. If they have to be estimated from data, $y$ is only approximately $\chi^2_m$ distributed (caveat for small number of data points).

4. (16P) Numerical simulations based on Q4c from sheet 2.
   a) (3P) Estimate the covariance matrix from the set of parameters $(\hat{a}, \hat{b})_i$ estimated in Q4c from sheet 2. Estimate the correlation coefficient $\rho$ and the parameter errors $\sigma_a, \sigma_b$.
   b) (2P) Perform a fit to one of the data sets of Q4c from sheet 2 with NonlinearModelFit and compare your result from 4a) with the covariance matrix quoted by that function.
   c) (2P) For the same data, determine the covariance matrix from the Hessian matrix and compare to previous results.
   d) (4P) In Q4c of sheet 2 you have also used linear regression to fit to that data. Calculate the covariance matrix according to $C^{-1} = A^T \Sigma^{-1} A$.
   e) (5P) Determine the 68% confidence region and count the $(\hat{a}, \hat{b})_i$ inside that region. What percentage is inside the confidence region?

5. (10P) Bootstrap
   In general, you do not know the underlying physics, as we assumed in the generation of the synthetic data around the “true” $y(x) = a_0 + b_0 x$. Instead, you are confronted with one data set only. But you can estimate the covariance matrix (and other quantities) from that set by using the data themselves as starting point for the generation of ensembles.
   For this, proceed as follows:
Generate synthetic data around the data points of the same data set used in 4b), 4c), 4d). Repeat, to generate multiple ensembles of synthetic data.

For every ensemble, perform a fit with the hypothesis $y = a + bx$ and determine $\hat{a}, \hat{b}, \hat{\chi}^2$.

From the set of $(\hat{a}, \hat{b})_i$, estimate the covariance matrix and compare to the results of 4).

Determine the distribution of the $\hat{\chi}^2$ by showing the histogram of $\hat{\chi}^2$ together with the expected theoretical distribution. Also, do the same with the distribution of the $\chi^2$ determined in Q4c from sheet 2. Discuss the difference.

Remark: The bootstrap method does not assume linearity in the parameters and can be applied to determine confidence regions and parameter errors for non-linear fits.