Statistics and Data Analysis

Error Propagation, Confidence Interval

1. Consider N measurements (x_i, y_i), where the y_i are independent and all have the same error, σ. Assuming a linear dependence is expected, the parameters of the line $y = a_1 + a_2 x$ that fit the data best are obtained by minimizing the sum $\chi^2 = \sum[y_i - (a_1 + a_2 x_i)]^2$, which leads to the simple minimum squares equations:

$$\begin{aligned} a_1 &= (\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i)/\Delta \\ a_2 &= (N \sum x_i y_i - \sum x_i \sum y_i)/\Delta \end{aligned}$$

with $\Delta = N \sum x_i^2 - (\sum x_i)^2$.

   (a) Using the error propagation formula show that the covariance matrix is

$$\mathrm{Cov}(a_1, a_2) = \frac{\sigma^2}{\Delta} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{pmatrix}$$

   Discuss the correlation sign and the case when the parameters are not correlated.

   (b) Find the parameters and errors of the line that best fit the data below, with $\sigma = 0.3$. Plot the data and the fit for $0 \le x \le 5$.

x	2.00	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
y	2.78	3.29	3.29	3.33	3.23	3.69	3.46	3.87	3.62	3.40	3.99

   (c) Plot the result of the fit together with the error. Study the consequences of ignoring the correlation term of the error matrix.

2. Verify the results of exercise 1(c) with the following numerical simulation:

   (a) For each x_i generate a random y_i with a Gaussian distribution $\mathrm{N}(a_1 + a_2 x_i, \sigma)$.

   (b) Fit the new data and predict the y for $x = 0.5$.

   (c) Repeat the steps above 1000 times and construct a histogram with the values of $y(0.5)$. Obtain the standard deviation of these data and compare to the result of 1(c).

3. If the variables $1/p$, λ, ϕ have been measured with errors $\Delta(1/p)$, $\Delta\lambda$, $\Delta\phi$, and with no correlation, what are the errors on $(p_x = p\,cos\lambda\,cos\phi, p_y = p\,cos\lambda\,sin\phi, p_z = p\,sin\lambda)$? What are the correlations in the new variables?

4. Given a normal distribution, what is the sample size, n, needed if the symmetric 95.4% confidence interval for μ shall have a length equal to (a) σ, (b) $\sigma/2$?

5. Six independent observations from $N(\mu, \sigma^2)$ are given by 10.7, 9.7, 13.3, 10.2, 8.9, 11.6. find the symmetric confidence interval for μ corresponding to $\gamma = 0.90$ and $\gamma = 0.95$ if $\sigma = 2$. Discuss the case if σ is unknown.