Statistical Data Analysis
Problem sheet #9 (Optional)

Problem 1: The binomial distribution is given by

$$f(n; N, \theta) = \frac{N!}{n!(N-n)!}\theta^n (1-\theta)^{N-n} ,$$

where n is the number of 'successes' in N independent trials, with a success probability of θ for each trial. Recall that the expectation value and variance of n are $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$, respectively. Suppose we have a single observation of n and using this we want to estimate the parameter θ.

1(a) Find the maximum likelihood estimator $\hat{\theta}$.

1(b) Show that $\hat{\theta}$ has zero bias and find its variance.

1(c) Suppose we observe $n = 0$ for $N = 10$ trials. Find the upper limit for θ at a confidence level of CL = 95% and evaluate numerically.

1(d) Suppose we treat the problem with the Bayesian approach using the Jeffreys prior, $\pi(\theta) \propto \sqrt{I(\theta)}$, where

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

is the expected Fisher information. Find the Jeffreys prior $\pi(\theta)$ and the posterior pdf $p(\theta|n)$ as proportionalities.

1(e) Explain how in the Bayesian approach how one would determine an upper limit on θ using the result from (d). (You do not actually have to calculate the upper limit.)

Explain briefly the differences in the interpretation between frequentist and Bayesian upper limits.

Problem 2: The outcome of a measurement consists of two independent random values, x and y, that follow the pdfs

$$f(x|\theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\theta_1-\theta_2)^2/2\sigma^2} ,$$

$$g(y|\theta_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\theta_2)^2/2\sigma^2} .$$

Consider the standard deviation σ (same for x and y) to be known.

2(a) Write down the log-likelihood function for θ_1 and θ_2.

2(b) Show that the Maximum-Likelihood estimators for θ_1 and θ_2 are

$$\hat{\theta}_1 = x - y ,$$

$$\hat{\theta}_2 = y .$$

2(c) Show that the estimators given above are unbiased. Find their exact variances, their covariance and correlation coefficient.

2(d) Show with the help of a sketch how the standard deviations of $\hat{\theta}_1$ and $\hat{\theta}_2$ can be determined from a contour of the log-likelihood function. Label all of the relevant features.

2(e) Suppose θ_1 is the parameter of interest and we regard θ_2 as a nuisance parameter. Find the profiled value $\hat{\hat{\theta}}_2(\theta_1)$ and using this show that the log of the profile likelihood for θ_1 can be written

$$\ln L_{\rm p}(\theta_1) = -\frac{1}{4}\frac{(x - y - \theta_1)^2}{\sigma^2} + C$$

where C represents terms that do not depend on the unknown parameters.

Show that the variance of $\hat{\theta}_1$ as determined directly from the second derivative of the profile likelihood is the same as found in (c).