Statistical & Numerical Methods
Exam
11 April 2014

DIRECTIONS: ALLOW 3 HOURS. WRITE YOUR NAME AND STUDENT NUMBER AT THE TOP OF EVERY PAGE OF YOUR SOLUTIONS. Please explain clearly all of the steps that you used to derive a result. Please make certain that your handwriting is readable to someone besides yourself.

1. Answer briefly (at least 4–5 sentences) the following questions: **10 points/question**

   (a) Describe two methods for finding the root of a 1D non-linear real function. Mention the drawbacks and advantages of each method.

   (b) Describe why many data points are needed to determine their underlying probability distribution function.

   (c) Suppose you have a model that has several parameters, A, B, C, and D, and that you can estimate $\mathrm{prob}(A, B, C, D|\{\mathrm{data}\}, I)$. Name and describe the process by which you could determine $\mathrm{prob}(B|\{\mathrm{data}\}, I)$.

   (d) Describe the basic assumptions of the least-squares method for fitting a model to data. Describe the use of the χ^2 statistic in the least-squares method. Describe (at least) two methods to determine the uncertainties (confidence intervals) of the inferred parameters of the model.

   (e) Describe the difference between accuracy and precision, and relate these terms to random uncertainties and systematic uncertainties.

   (f) You are given a function $Z = f(X, Y)$ and you know that X and Y have Gaussian probability distribution functions, with uncertainties σ_X and σ_Y, and that X and Y are independent. What is the variance σ_Z^2 of Z? Determine σ_Z^2 when $Z = 3X + Y^2$.

   (g) Describe how you would numerically find the value of an improper integral. Provide a few examples of such integrals. For those examples, describe the steps would you take and the algorithms you would use to find the solutions.

2. A derivation: **4 points/question**. Explain your steps clearly and show your work.

   (a) Write down Cox's product rule for the probabilities $\mathrm{prob}(X, Y|I)$ and $\mathrm{prob}(Y, X|I)$.

   (b) Solve these two equations for $\mathrm{prob}(X|Y, I)$. This is Bayes' theorem.

   (c) Write the equation you derived in part (b) again, calling X the "hypothesis" and Y the "data". What is the name of each term on each side of this equation?

   (d) Set $P = \mathrm{prob}(X|Y, I)$. How would you determine the best estimate X_0 of X, given P? How would you prove that X_0 is indeed the best estimate of X?

   (e) Describe the importance of Bayes' theorem for data analysis (no more than 3–4 sentences).

3. True/false questions – mark T for a true statement or F for a false statement on your exam paper: **1 point/question**

   (a) A spline is a 4th order polynomial.

   (b) A Poisson distribution is an appropriate likelihood function for binned data when you know the expected signal in each bin.

   (c) A Kolmogorov-Smirnov test is the appropriate test when comparing two distributions generated from binned data, for example when asking whether two luminosity functions have been drawn from the same distribution.

   (d) The best estimate of a parameter drawn from a (one-dimensional) Gaussian distribution based on data with equal uncertainties is the mean of the individual data points.

   (e) It is possible to use the rejection method to generate random numbers following a Gaussian distribution.

   (f) The singular value decomposition method can be used to find the inverse of an $N \times M$ matrix.

   (g) The spline method does not require knowledge of the 2nd derivative of the function to be interpolated.

   (h) The leapfrog algorithm is time-reversible, and can be applied to find the solution to all types of ordinary differential equations.

   (i) The appropriate prior for a location parameter is a constant in the allowed range.

   (j) A binomial distribution is the appropriate probability distribution function when you know both the expected value and expected variance of an experiment.