Statistics for Astronomy 2021–2022

RE-EXAMINATION

3 February 2022 (15:00 - 18:00)

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 the following open questions: (8 points/question)

   (a) Given N mutually exclusive hypotheses $\{H_i\}$ that together cover all possible options. Consider the case of calculating $\text{prob}(A \mid I)$ from $\text{prob}(A, H_i \mid I)$ using marginalization, where I is some background information.

   i. Write down Cox's product and sum rules.

   ii. What equation is implied by having N mutually exclusive hypotheses $\{H_i\}$ that together cover all options?

   iii. Use these equations to derive the mathematical formula that describes the discrete marginalization rule.

   (b) In the context of parameter estimation, describe the difference between a scale parameter and a location parameter. What is the appropriate prior for each case?

   (c) For certain parameter estimation problems, the most probable estimate of the parameters can be solved by minimizing the χ^2 function, i.e., by doing a least-squares fit.

   i. Write down the four requirements necessary for least-squares fitting to produce the most probable estimate.

   ii. Consider the case of fitting parameter a in model function $f(x) = a\tan(x^2)$ to a set of N measurements $\{y_i\}$ with corresponding uncertainties $\{\sigma_i\}$ and positions $\{x_i\}$, with $0 < i \leq N$. The position values $\{x_i\}$ can be considered to have absolute certainty (i.e., have no errors). Write down the χ^2 function that needs to be minimized.

   (d) Given a one-dimensional probability density function, describe how you would find and compute the 95% confidence interval.

   (e) Describe the principle of indifference and give a simple example of its application to determine a distribution function.

   (f) Given a continuous uniform distribution function $\text{prob}(x) = \frac{1}{5}$ for $0 \leq x \leq 5$ and $\text{prob}(x) = 0$ otherwise:

   i. Calculate $\langle x \rangle$

   ii. Calculate $\langle x^3 - 1 \rangle$

   iii. Calculate $\text{Var}(x)$, the variance of x.

   (g) Describe what the Levenberg-Macquard algorithm can be used for and how it conceptually works. What two methods does it combine?

   (h) Given that we know $\text{prob}(\text{data} \mid \eta, a, b, c, I)$ and $\text{prob}(\eta \mid I)$, how would we calculate $\text{prob}(\eta \mid \text{data}, I)$ from these two functions?

→ **See next page for questions 2 and 3**

2. Two open questions involving derivations:

   (a) **(14 points)** Assume a coin with unknown bias $H \in [0 \dots 1]$ is thrown N times ($H = 0.5$ for a fair coin, $H = 1$ for a coin that always throws heads). The number of times heads comes up is R. Therefore, $P(R \mid H, N, I) \propto H^R(1-H)^{N-R}$. Start from Bayes' theory and, given R and N, derive these two expressions:

      - H_0, the best estimate of the bias.
      - σ_H^2, the variance of H_0.

   (b) **(12 points)** Consider "the lighthouse problem" (Fig. 1): A lighthouse at sea is at a position α along the coast and a distance β at sea. The shore is lined with photodetectors that record only that a flash has been recorded, but not the angle from which it came. N flashes have been recorded at positions $\{x_i : 0 \leq i < N\}$ along the coast, corresponding with azimuth directions $\{\theta_i : 0 \leq i < N\}$. The relation between a detected position and azimuth direction is given by $\beta \tan\theta_i = x_i - \alpha$. We know that the lighthouse flashes uniformly in the directions of interest: $\text{prob}(\theta_i|\alpha,\beta,I) = \frac{1}{\pi}$ for $-\frac{\pi}{2} \leq \theta_i \leq \frac{\pi}{2}$. Write down the transformation rule, and use it to calculate $\text{prob}(x_i|\alpha,\beta,I)$.

   Hint: the table below contains some of the common trigonometric derivatives.

function	derivative	function	derivative
$\sin x$	$\cos x$	$\arcsin x$	$1/\sqrt{1-x^2}$
$\cos x$	$-\sin x$	$\arccos x$	$-1/\sqrt{1-x^2}$
$\tan x$	$\sec^2 x$	$\arctan x$	$1/(1+x^2)$

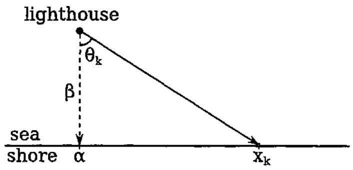

Figure 1: *The Lighthouse problem (see Question 2b)*

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

   (a) In a statistical context, the words "error" and "uncertainty" are synonyms.

   (b) The theory of Ockham's razor can be used to show that conspiracy theories that involve highly specific and tweaked models(/ideas) are highly unlikely.

   (c) Compared to Spearman's rank-order correlation coefficient, the linear correlation coefficient is more robust against outliers.

   (d) The Poisson distribution is given by $\text{prob}(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

   (e) Obtaining $\text{prob}(B \mid I)$ from $\text{prob}(A, B \mid I)$ requires marginalization over parameter A.

   (f) The central limit theory states that distributions that have a well defined mean tend to a uniform distribution for large N.

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

   (h) The mean of the Poisson distribution is equal to its standard deviation.

   (i) MCMC (Markov chain Monte Carlo) methods can be used to analyse the posterior of complex, non-linear models.

   (j) There are four alpacas. Each alpaca spends half of its time eating grass and the other half of its time sleeping or drinking. All of them have different, independent eat-sleep-drink rhythms and durations. The chance that I find all alpacas eating at a particular time is 1/8th.