Department of Astronomy, Astrophysics and Space Engineering

Indian Institute of Technology Indore

Spring Semester	**Academic Year 2020-21**
End Semester Exam	**Course Code : AA 608**
Date: **12 March 2021 (FN)**	Max. duration : **3 hr 30 min**
Max. Marks : **80**	Number of pages: 4

There are 4 pages in this question paper. Please check all of them. All questions are compulsory.

- Please, answer in the most complete and clear way. Do not forget the units!
- When you define a quantity try to give a complete definition: words + formula + units.
- Provide schematic diagrams where necessary to support your answers.
- Numbers in the parentheses at the end of each question represent the allocated marks.

1. The distribution of flux densities of extragalactic radio sources is a power-law with slope $-\alpha$, say, so the likelihood to measure a source flux S is $p(S|\alpha) \propto S^{-\alpha}$, above some (known) instrumental limiting flux density of S_0. In a non-evolving Euclidean universe $\alpha = 3/2$ and departure of α from the value 3/2 is evidence for cosmological evolution of radio sources (we assume measurement errors are negligible). This was the most telling argument against the steady-state cosmology in the early 1960s.
   (a) Given observations of radio sources with flux densities S, what is the most probable value of α, assuming a uniform prior?**(5 Marks)**
   (b) Show that if a single source is observed, and the flux is $2S_0$, that the most probable value of α is 2.44. **(3 Marks)**
   (c) By examining the second derivative of the posterior, estimate the error on α to be 1.44.**(2 Marks)**

2. An astronomical source emits photons with a Poisson distribution, at a rate of λ per second. A telescope detects the photons independently, with probability p. In time t, the source emits M photons, and N are detected.
   (a) Show that the joint probability of N and M is
   $$P(M,N) = \frac{\mu^M}{M!}e^{-\mu}\frac{M!}{N!(M-N)!}p^N q^{(M-N)}$$
   where $\mu = \lambda t$ and $q = 1-p$.**(3 Marks)**
   (b) Marginalise over M to show that
   $$P(N) = \frac{p^N q^{-N} e^{-\mu}}{N!}\sum_{M=N}^{\infty}\frac{(q\mu)^M}{(M-N)!}.$$
   (3 Marks)
   (c) Sum the series to show that N has a Poisson distribution with expectation value $p\mu$ (You can use the following information: $\sum_{i=0}^{\infty}\frac{x^i}{i!} = e^x$). Why could this have been anticipated? **(4 Marks)**

3. (a) What is evidence in the context of Bayes theorem? In the context of Bayesian inference, in what type of problems you have to care about the evidence and when you can safely ignore it? **(3 Marks)**
   (b) Define χ^2 in the context of M independent variables x_i, which are normally distributed with mean μ_i and variance σ_i^2. Define "degrees of freedom" for the same scenario. **(2 Marks)**
   (c) What is "Goodness of fit"? Give a flow chart to ilustrate how would you use χ^2 test to check the "Goodness of fit"? **(2 Marks)**
   (d) The results of measurements of mass (M_z) of Z^0 boson made by four different detectors (L3, OPAL, Aleph and Delphi) at CERN are as follows:

Detector	**Mass in** GeV/c^2
L3	91.161 ± 0.013
OPAL	91.174 ± 0.011
Aleph	91.186 ± 0.013
Delphi	91.188 ± 0.013

Table 1: Mass of Z^0 boson

Using a χ^2 test demonstrate whether these results can be represented by a single number or not. **(3 Marks)**

4. (a) Assume the time-series (photon intensities with time) from a source (a variable star). The time-series is embedded with a signal which comprises a broad periodic feature at 0.1 Hz and a narrow coherent periodic feature at 500 Hz.
    i. What would be the sampling rate if you want to detect the 0.1 Hz signal? What would be the preferred technique to best detect the signal significantly? **(2 Marks)**
    ii. What would be the sampling rate if you want to detect the 500 Hz signal? Can the same technique as the previous question be applied here to detect the signal most significantly? **(2 Marks)**
   (b) How would you approach the same problem if the signal is transient or non-stationary (i.e. evolving frequency)? Give a qualitative answer. **(1 Marks)**
   (c) What is Fast Fourier Transform and why is it preferred? **(2 Marks)**
   (d) What is aliasing? **(1 Marks)**
   (e) Let us assume we are observing a star that exhibits a light curve which a stationary Poissonian time series (having constant signal-to-noise ratio) using a satellite. The satellite has an equatorial orbit of 90 minutes around the earth as it observes the star and therefore our observing satellite is not able to observe the target star all the time. What is the primary challenge in conducting the time-series analysis of the observed light curve? **(2 Marks)**

5. Refer to the dataset ("supernovadata.csv" file) provided to you along with this question paper. This dataset has three (3) columns: redshift (z), the distance modulus (μ) and the errors on the distance modulus ('muerr'). Import the csv file into python / R and perform the following tasks / answer the following questions. You may borrow the syntax for plotting / fitting functions, and import any libraries you like, but the program/s you write must be your own. You must submit ALL programs, plots and written statements.
   (a) Plot μ vs. z, with error bars. Feel free to use any plotting technique. **(3 Marks)**
   (b) Depending on the distributions of z and μ, decide on the simplest possible fitting function that is not trivial. State why you choose this function and write the functional form, with the coefficients you fit. Provide an estimate of goodness of fit – state which one, and why you chose it. **(2+2+1+2+2 Marks)**
   (c) State which function you can fit which might be a better fit – perform the fit and quote the coefficients. **(1+2 Marks)**
   (d) Is the data homoskedastic or heteroskedastic? Show through plots and numbered statements. **(2+5+3 Marks)**

Please see the next two pages for more questions!

6. What information can we extract from the following confusion matrix? Define and calculate the recall, accuracy, precession and false alarm from the confusion matrix.
(5 Marks)

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

7. What are the major advantages and dis-advantages of K-means algorithm? Is it a clustering or classification algorithm? **(2 Marks)**

8. Calculate the Naïve Bayes classification probability for playing golf today = (Sunny, Hot, Normal, False) given the data set as follows: **(8 Marks)**

	Outlook	**Temperature**	**Humidity**	**Windy**	**Play Golf**
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No

8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Hint: Calculate conditional probability of each feature (i.e. P(feature|yes) and P(feature|No) and then use Bayes principle to get P(Yes|today) and P(No|today).