é

QUESTION 1

5 points

What is an estimator and what is the goal of estimation? Provide an example.

QUESTION 2

4 points

In this and the next two questions, let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be independent, unbiased estimators for an unknown parameter θ, with respective variances σ_1^2 and σ_2^2.

Show that $\widehat{\theta}_3 = \lambda\widehat{\theta}_1 + (1-\lambda)\widehat{\theta}_2$ is also an

unbiased estimator of θ for all values of λ.

QUESTION 3

5 points

Calculate the variance of $\widehat{\theta}_3$ and show that the value of λ that minimises this variance is

$$\lambda = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

QUESTION 4

8 points

Save Ar

Let $\overline{X}_1$ and $\overline{X}_2$ be the means of two independent random samples of sizes n_1 and n_2 from the independent distributions $N(\mu,\sigma_1^2)$ and $N(\mu,\sigma_2^2)$, respectively, where the values of σ_1^2 and σ_2^2 are known. Find the value of λ which minimises the variance of the estimator $\widehat{\mu}=\lambda\overline{X}_1+(1-\lambda)\overline{X}_2$.

QUESTION 5

8 points

Explain the idea behind Ordinary Least Squares method of estimation. Illustrate it with a stylised figure.

QUESTION 6

3 points

Which of the expressions below correspond to the statement: *the probability that it is Monday, given that it is raining?*
Note: there may be more than one such expression.

- ☐ a. $P(rain \mid Monday)$
- ☐ b. $P(Monday \mid rain)$
- ☐ c. $P(rain \mid Monday)\ P(Monday)$
- ☐ d. $P(rain \mid Monday)\ P(Monday)\ /\ P(rain)$
- ☐ e. $P(Monday \mid rain)\ P(rain)\ /\ P(Monday)$

QUESTION 7

2 points

When working on this question and the two that follow it, imagine that there are two species of panda. Both are equally common in the wild and live in the same kinds of places. They look exactly alike and eat the same food, but they differ in the number of offspring they produce. Species A gives birth to twins 10% of the time, otherwise producing only a single infant. Species B gives birth to twins 20% of the time, otherwise producing single infants.

Suppose further that you are managing a captive panda breeding program. You have a new female panda of unknown species, and she has just given birth to twins.

Compute the probability that the panda is from species A, given that she has produced twins. Give your answer to three decimal places.

QUESTION 8

2 points

Using the same details as in the previous problem, compute the probability that the panda is from species B, given that she has produced twins. Here too, you should give your answer to three decimal places.

QUESTION 9

2 points

Finally, find the probability that the panda's next pregnancy will produce twins. Here too, you should give your answer to three decimal places.

QUESTION 10

6 points Save Answ

The next two questions concern the kernel density estimators (KDE) given by the formula

$$f(s) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{w} K\left(\frac{s - x_j}{w}\right)$$

where $\{x_1, \ldots, x_n\}$ is an observed data set and the x_j are ordinary real numbers. Explain how a KDE is used. Your answer should include a discussion of all the elements in the formula above.

QUESTION 11

2 points

Consider a KDE where the data are $\{-1, 1, 2\}$, the kernel is the standard normal distribution, $K(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$, and $w = 1$: evaluate $f(s)$ at $s = 0$.

QUESTION 12

3 points

This question and the one that follows at concern a survey of Data Science students done to determine the proportion p who prefer Python to R. Suppose that k out of the N students surveyed preferred Python and consider the problem of doing Bayesian inference for p with the following ingredients:

- **Prior on** p: a Beta distribution, which has probability density function $f(p) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\,\Gamma(\beta_0)} p^{\alpha_0 - 1} (1-p)^{\beta_0 - 1}$
- **Likelihood:** The Binomial distribution, which has probability mass function $g(k) = \frac{N!}{k!\,(N-k)!} p^k (1-p)^{N-k}$

Which choice of the parameters α_0 and β_0 corresponds to a uniform, uninformative prior?

- ○ $\alpha_0 = 0, \beta_0 = 1$
- ○ $\alpha_0 = \beta_0 = 1$
- ○ $\alpha_0 = 1, \beta_0 = 0$
- ○ $\alpha_0 = \beta_0 = 0$

QUESTION 13

3 points

Which of the following is true:

- ○ The posterior is a Binomial distribution.
- ○ The posterior is a Beta distribution with shape parameters α' and β' given by $\alpha' = \alpha_0 + k + 1$ and $\beta' = \beta_0 + N + 1$.
- ○ The posterior is a Beta distribution with shape parameters α' and β' given by $\alpha' = \alpha_0 + k$ and $\beta' = \beta_0 + N - k$.
- ○ The posterior is a Normal distribution with mean and variance given by $\mu = k/N$ and $\sigma^2 = k(N-k)/N^2$.

QUESTION 14

7 points Save Answ

The figure below is part of an exploratory data analysis on a collection of measurements made while human subjects, some of whom suffered from Alzheimer's disease, performed 25 writing and drawing tasks. These were designed for the early detection of Alzheimer's disease and the measurements summarised below are either times—`air_time23` is, for example, the time for which the subject held the pen in the air above the page while performing task 23—while others—such as `max_x_extension21`—have to do with the size and shape of the marks the subjects made.

What insights can you draw from this figure? Which measures do you think are most likely to distinguish the healthy subjects from the Alzheimer's patients and why?