EGT3
ENGINEERING TRIPOS PART IIB

Friday 7 May 2021 1.30 to 3.10

Module 4M24

COMPUTATIONAL STATISTICS AND MACHINE LEARNING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet and at the top of each answer sheet.*

STATIONERY REQUIREMENTS
Write on single-sided paper.

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed.
You are allowed access to the electronic version of the Engineering Data Books.

10 minutes reading time is allowed for this paper at the start of the exam.

The time taken for scanning/uploading answers is 15 minutes.

Your script is to be uploaded as a single consolidated pdf containing all answers.

1 Let $X_1, \ldots, X_N$ denote iid random variables with mean μ and variance σ^2.

(a) Show that the Monte Carlo estimate of the mean is unbiased. [10%]

(b) A generalised Monte Carlo estimate for μ is given as

$$\hat{\mu} = \sum_{n=1}^{N} w_n X_n$$

where $w_1, \ldots, w_N$ are deterministic real valued weights.

Derive the mean and variance of the estimate $\hat{\mu}$. [30%]

(c) What conditions are required on $w_1, \ldots, w_N$ for the generalised Monte Carlo estimator to be an unbiased estimate of the mean μ? [10%]

(d) What choice of weights $w_1, \ldots, w_N$ gives a minimum variance unbiased estimator? [50%]

2 Bayesian Machine Learning often needs Monte Carlo estimates with respect to distributions of the form:

$$p(u, v|y)$$

where y is the data, and u, v are the model variables.

(a) Give a detailed description of the Gibbs sampler for such a distribution and derive the acceptance probability of the sampler. [40%]

(b) Suppose $y_1, \ldots, y_N$ are iid random variables and each y_i are Gaussian distributed with mean μ and precision τ:

$$y_i \sim \mathcal{N}(\mu, \tau^{-1})$$

The Bayesian model has priors on the mean and precision given by:

$$\mu \sim \mathcal{N}(0, w^{-1})$$
$$\tau \sim \text{Gamma}(\alpha, \beta)$$

Where α, β, w are fixed. The form of the Gamma distribution is given by:

$$\text{Gamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau}$$

Denoting $y_1, \ldots, y_N$ as $\mathbf{y}$, define the Gibbs sampler for $p(\mu, \tau|\mathbf{y})$, giving both relevant full conditional distributions. [60%]

3 The Lebesgue measure in the intervals $[a_d, b_d]$ in $d = 1, \dots, D$ takes the form

$$\prod_{d=1}^{D}(b_d - a_d)$$

Noting that the Lebesgue measure is Finite, Monotonic, and Translation Invariant

(a) Show that in a complete infinite dimensional Hilbert space, the Lebesgue measure is not well defined. [30%]

(b) The standard Gaussian measure in $\mathbb{R}$ is given as:

$$g(B) = \frac{1}{\sqrt{2\pi}} \int_B \exp\left(-\frac{x^2}{2}\right) dx$$

Consider $\mathbb{R}^\infty$, under what conditions will the product measure

$$\mu = \prod_{k=1}^{\infty} g$$

be well defined? What subspace of $\mathbb{R}^\infty$ does this represent? [15%]

(c) Given the form of the Hellinger integral

$$H(\mu, \nu) = \int \sqrt{d\mu d\nu}$$

Show that for two Gaussian measures with Lebesgue densities $\mu = \mathcal{N}(0, a)$, $\nu = \mathcal{N}(0, b)$, the Hellinger integral is given by

$$H(\mu, \nu) = (ab)^{\frac{1}{4}} \sqrt{\frac{2}{a + b}}$$

[35%]

Hint: Note that the Radon-Nikodym derivative defines the Lebesgue density

$$\frac{d\mu}{dx} = \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{x^2}{2a}\right)$$

(d) Using the result in (c), determine whether the product measures

$$\prod_{k=1}^{\infty} \mathcal{N}(0, a_k) \qquad \prod_{k=1}^{\infty} \mathcal{N}(0, \beta a_k)$$

are equivalent or singular for specific values of β, where $a_k = a$. [20%]

4 Consider the sequence of functions $\{f_n\}$ defined by

$$f_n(x) = nx(1 - x^2)^n \qquad x \in [0, 1]$$

(a) Find the value of the following limit

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx$$

and determine whether

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = \int_0^1 \lim_{n\to\infty} f_n(x)dx$$

is satisfied. [15%]

(b) Discuss the types of convergence of a sequence of functions for which limit and integral operations can be exchanged. [15%]

(c) (i) The continuous functions $C[0, 1]$ are defined as

$$f_n(x) = \begin{cases} 1 & 0 \le x \le \frac{1}{2} \\ 1 - 2n\left(x - \frac{1}{2}\right) & \frac{1}{2} \le x \le \frac{1}{2}\left(1 + \frac{1}{n}\right) \\ 0 & \frac{1}{2}\left(1 + \frac{1}{n}\right) \le x \le 1 \end{cases}$$

Sketch the functions and determine whether they form a Cauchy sequence in $L^2[0, 1]$. [50%]

(ii) What is the limiting function as $n \to \infty$, $f(x)$? Sketch this limiting function. Are the functions $f_n(x)$ and $f(x)$ elements of $L^2[0, 1]$? Does this form a complete Hilbert space? [20%]

END OF PAPER

THIS PAGE IS BLANK