EGT3
ENGINEERING TRIPOS PART IIB

Thursday 01 May 2025 14.00 to 15.40

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.*

STATIONERY REQUIREMENTS
Write on single-sided paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

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

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

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) By providing a derivation, 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 The probability of data $y \in \mathbb{R}$ given model parameters $\theta \in \mathbb{R}^D$ is denoted by $p(y|\theta)$. Consider a probabilistic model for the data that includes a variable $x \in \mathbb{R}$ which is unobserved and has probability density denoted as $p(x)$, such that

$$p(y|\theta) = \int_{\mathbb{R}} p(y, x|\theta) dx.$$

It is assumed that $p(y, x|\theta)$ is a sufficiently regular function that integration and derivative operators can be interchanged. The parameters of the model, θ, can be estimated by maximising the probability of the data y using iterative gradient based optimisation methods. For example

$$\theta_{n+1} = \theta_n + \delta \nabla_\theta \log p(y|\theta_n)$$

where θ_n denotes the parameter values at the n'th iteration of the optimisation procedure, and δ is the step size of the gradient-based optimisation procedure.

(a) Noting that

$$\nabla_\theta \log p(y|\theta_n) = \frac{\nabla_\theta p(y|\theta_n)}{p(y|\theta_n)}$$

show that

$$\nabla_\theta \log p(y|\theta_n) = \mathbb{E}_{x|y,\theta} \{\nabla_\theta \log p(y, x|\theta_n)\}$$

where $\mathbb{E}_{x|y,\theta} \{\cdot\}$ denotes the expectation with respect to the posterior $p(x|y, \theta)$. [45%]

(b) Provide a Metropolis-Hastings based Markov chain algorithm that will yield samples from the posterior $p(x|y, \theta)$ which uses a Langevin diffusion based proposal mechanism. [40%]

(c) Describe a Monte Carlo estimator for $\nabla_\theta \log p(y|\theta)$ that can be used in the gradient based optimisation procedure described above. [15%]

3 Bayes' rule for probability measures on $\mathbb{R}^D$ is

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mathbb{P}(u)$$

where each $\mathbb{P}$ is a probability measure defined with respect to the Lebesgue measure.

(a) (i) Is this definition of Bayes' rule in $\mathbb{R}^D$ suitable as a means of changing from prior to posterior in an infinite dimensional complete Hilbert space $\mathcal{H}$? [10%]

(ii) Use the Radon-Nikodym derivative to provide a definition of Bayes' rule suitable for a Hilbert space $\mathcal{H}$. [10%]

(b) Based on a standard Gaussian measure on $\mathbb{R}$

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

where dx denotes the Lesbesgue measure.

(i) Define a prior reference measure for Bayes rule on $\mathcal{H}$ which takes the form

$$\mu^0 = \prod_{k=1}^{\infty} g$$

and detail what conditions need to be satisfied for μ^0 to be well defined (finite) and describe the subspace of $\mathbb{R}^\infty$ that μ^0 is defined on. [30%]

(ii) From the definition of Bayes' rule in $\mathcal{H}$ and using $\mu^0 = \mathcal{N}(0, C)$, where C is a trace class covariance operator, with a proposal $\mathcal{N}(u, \beta^2 C)$, where β is a constant, derive an expression for the Metropolis-Hastings acceptance ratio. Discuss the implications on the performance of the method in infinite dimensional space. [30%]

(iii) Suggest an alternative proposal mechanism that would resolve the issues highlighted in part (b)(ii). [20%]

4 The Hyperbolic-Secant distribution defined on $\mathbb{R}$ has a probability density function with respect to Lebesgue Measure which is given as

$$p(x) = \frac{1}{2}\operatorname{sech}\left(\frac{\pi x}{2}\right)$$

where $x \in \mathbb{R}$ and $\operatorname{sech}(x) = \frac{2}{\exp(x)+\exp(-x)}$.

(a) Verify that there exists a Langevin Diffusion defined on $\mathbb{R}$ whose invariant density is that of the Hyperbolic-Secant distribution and derive the corresponding equation describing the Langevin Diffusion. Hint: Show that the Hyperbolic-Secant distribution is log-concave. [40%]

(b) Write out an Unadjusted Langevin Algorithm (ULA) which will converge to a biased version of the Hyperbolic-Secant distribution. [20%]

(c) The Metropolis Adjusted Langevin Algorithm (MALA) takes ULA as a proposal mechanism and removes the bias by applying an Accept-Reject step. Define the corresponding MALA acceptance probability for the Hyperbolic-Secant distribution target. [40%]

END OF PAPER

THIS PAGE IS BLANK