EGT3
ENGINEERING TRIPOS PART IIB

Wednesday 27 April 2022 9.30 to 11.10

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.*

STATIONERY REQUIREMENTS
Write on single-sided paper.

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

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 $X_1, \ldots, X_N$ be i.i.d. random variables with unknown mean and variance μ and σ^2 respectively.

(a) Show that the variance around the mean μ of a Monte Carlo estimate $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} X_n$ is equal to $\frac{\sigma^2}{N}$. [20%]

(b) For a known function of the random variables $g(X)$ with known mean $m = \mathbb{E}[g(X)]$ and coefficient $c \in \mathbb{R}$:

(i) Write an expression for an unbiased control variate Monte Carlo estimator $\hat{\mu}_{CV}$ of the mean μ. [10%]

(ii) Show that the optimal variance achievable for this estimator is equal to

$$\mathrm{Var}\left(\hat{\mu}_{CV}\right) = \mathrm{Var}\left(\hat{\mu}\right) - \frac{\mathrm{Cov}\left(\hat{\mu}, \hat{m}\right)^2}{\mathrm{Var}\left(\hat{m}\right)}$$

where $\hat{m} = \frac{1}{N}\sum_{n=1}^{N} g(X_n)$.

[30%]

Hint: $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$

(c) The random variable X is distributed uniformly on the unit interval $[0, 1] \subset \mathbb{R}$. Now consider the task of estimating the expectation $\mu = \mathbb{E}[f(X)]$ using a control variate $g(X)$ with known mean $m = \mathbb{E}[g(X)]$. Noting that $\frac{1}{2} - (\log 2)^2 \approx 0.0195$, show that the control variate estimate $\hat{\mu}_{CV}$ of μ, with $f(X) = \frac{1}{1+X}$ and $g(X) = 1 + X$ has optimal variance equal to

$$\mathrm{Var}\left(\hat{\mu}_{CV}\right) = \frac{0.0195}{N} - 12N\left(\frac{1}{N} - \frac{3}{2N}\log 2\right)^2$$

[40%]

2 The set of functions $C_2[-\pi,\pi]$ is defined as the intersection between the set of continuous functions and L_2 integrable functions

$$C_2[-\pi,\pi] = C[-\pi,\pi] \cap L_2[-\pi,\pi]$$

(a) Describe the type of functions that $C_2[-\pi,\pi]$ contains. [10%]

(b) Consider a set of square integrable orthonormal functions $\{\phi_k\}$ defined on $[-\pi,\pi]$, with $\phi_k(x) = \exp(ikx)$ where i is the imaginary unit. The expansion of a function $f \in L_2$ is represented as

$$f(x) = \sum_{k=1}^{\infty} c_k \phi_k(x)$$

where c_k are the Fourier coefficients of f.

(i) Derive an expression for the norm $\|f\|_{L_2}$ of the function $f \in L_2$. [20%]

(ii) Denoting order s weak derivatives as D^s, the Sobolev space W_2^s is defined as

$$W_2^s = \left\{ f \in C_2[-\pi,\pi];\ \|D^s f\|_{L_2}^2 < \infty \right\}$$

Show that the norm in W_2^s is given by

$$\|f\|_{W_2^s}^2 = \sum_{k=1}^{\infty} k^{2s} c_k^2$$

[30%]

(c) Consider an approximation for the function $f(x)$, as in part (b), to be defined as

$$f_N(x) = \sum_{k=1}^{N} c_k \phi_k(x)$$

Derive the following error bound:

$$\epsilon_N(f) = \|f - f_N\|_{L_2}^2 < \frac{1}{N^{2s}} \|f\|_{W_2^s}^2$$

and discuss the implications of the bound on the rate of convergence of the error. [40%]

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 (a) For a time indexed variable $X_t \in \mathbb{R}$ and standard Brownian motion denoted by B_t, show that the following Stochastic Differential Equation (SDE):

$$dX_t = -X_t dt + \tanh(X_t)dt + \sqrt{2}dB_t$$

has an invariant probability measure which has Lebesgue density corresponding to an equally weighted mixture of two Gaussian density functions

$$p(X) = \frac{1}{2}\mathcal{N}(1, 1) + \frac{1}{2}\mathcal{N}(-1, 1)$$

[60%]

Note that $\cosh(X) = \frac{1}{2}(e^X + e^{-X})$, and $\tanh(X) = \frac{\sinh(X)}{\cosh(X)}$

(b) Now suppose we wish to sample from the Gumbel distribution having Lebesgue density:

$$q(X) = \exp\left(-X - \exp(-X)\right)$$

Determine the corresponding Langevin SDE that has $q(X)$ as its invariant density. [40%]

END OF PAPER

THIS PAGE IS BLANK