Durham University

EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2025	MATH4337-WE01

Title:
Uncertainty Quantification IV

Time:	2 hours	
Additional Material provided:		
Materials Permitted:		
Calculators Permitted:	No	Models Permitted: Use of electronic calculators is forbidden.

Instructions to Candidates:	Answer all questions. Section A is worth 40% and Section B is worth 60%. Within each section, all questions carry equal marks. Write your answer in the white-covered answer booklet with barcodes. Begin your answer to each question on a new page.

Revision:	

ED01/2025

SECTION A

Q1 We wish to design the first run of an expensive 1D computer model $f(x)$ over the input range $x \in [0, 1]$. We set up a standard Bayes Linear emulator, specifying a prior expectation $\mathrm{E}[f(x)] = 0$ and a squared exponential prior covariance structure $\mathrm{Cov}[f(x), f(x')] = \sigma^2 \exp\{-|x - x'|^2/\theta^2\}$, for given fixed $\sigma, \theta > 0$.

(a) Find an expression for the general emulator variance $\mathrm{Var}_D[f(x)]$ at input x assuming a single run has been performed at $x^{(1)}$, even though we have not performed the run yet, and hence do not currently know $D = f(x^{(1)})$.

(b) Hence find the location of the first run $x^{(1)}$ that would minimise the integrated emulator variance over the interval $[0, 1]$ i.e. find

$$\underset{x^{(1)} \in [0,1]}{\arg\min} \int_0^1 \mathrm{Var}_D[f(x)]\, dx, \qquad \text{where } D = f(x^{(1)})$$

(c) Comment on whether your answer to **Q1**(b) is intuitive.

Q2 We wish to history match an expensive 1D computer model $f(x)$ over the input range $x \in \mathcal{X}_0 = [0, 1]$. We set up a standard Bayes Linear emulator, specifying an exponential prior covariance structure $\mathrm{Cov}[f(x), f(x')] = \sigma^2 \exp\{-|x - x'|/\theta\}$, with $\sigma^2 = 2^2$, $\theta = 1/3$, and prior expectation $\mathrm{E}[f(x)] = 1$. A single run is performed at the location $x^{(1)} = 0.8$, yielding the output $D = f(x^{(1)}) = -1$.

(a) Find an expression for the emulator expectation $\mathrm{E}_D[f(x)]$ and variance $\mathrm{Var}_D[f(x)]$.

(b) The real system y corresponding to $f(x)$ is measured, yielding the observation $z = 0.25$. Measurement error and model discrepancy are specified to be $\mathrm{Var}[e] = 0.4^2$ and $\mathrm{Var}[\epsilon] = 0.3^2$ respectively. Using your results from **Q2**(a), explicitly construct the corresponding implausibility measure $I(x)$ for use in the history match.

(c) Define the non-implausible region and give a reasonable implausibility cutoff c. Justify your choice of cutoff.

SECTION B

Q3 We have a computer model $f(x)$ with 2D input $x \in \mathcal{X}_0 = [0,1]^2$ and scalar output $f \in \mathbb{R}$. In preparation for emulation, we specify the usual squared exponential covariance structure:

$$\text{Cov}[f(x), f(x')] \;=\; \sigma^2 \exp\{-||x - x'||^2/\theta^2\}$$

for given fixed $\sigma, \theta > 0$.

Imagine that we have performed n runs at the locations $x^{(i)} = (0, b_i)$, $i = 1, \dots, n$ with $b_i \in [0,1]$ yielding output $D = (D_1, \dots, D_n)^T = (f(x^{(1)}), \dots, f(x^{(n)}))^T$. We specify the prior expectation of $f(x)$ as the general expression $\text{E}[f(x)]$. We now wish to emulate the model at the input location $x = (a, b_1)$.

(a) By considering the above covariance structure, show that for $x = (a, b_1)$ we have

$$\text{Cov}[f(x), D] \;=\; e^{-a^2/\theta^2}\text{Cov}[f(x^{(1)}), D]$$

(b) Use this to evaluate exactly the emulator expectation $\text{E}_D[f(x)]$ at the point $x = (a, b_1)$, and show that the only run that it depends on is the first one.

(c) Find $\text{Var}_D[f(x)]$ at $x = (a, b_1)$ and comment on its form.

(d) For many expensive computer models $f(x)$ there exists a *known boundary* $\mathcal{K}$ located at $\mathcal{K} \equiv \{x_1 = 0, x_2 \in [0,1]\}$, where the function $f(x)$ is analytically solvable and hence known. Hence we can generate arbitrarily large numbers of runs of the form $\{f(x) : x \in \mathcal{K}\}$. Use the above results to argue that we can still update the emulator analytically with respect to the information contained on $\mathcal{K}$, and hence find the emulator adjusted expectation $\text{E}_K[f(x)]$ and variance $\text{Var}_K[f(x)]$ where K represents a vector of outputs of a large number of n runs evaluated on $\mathcal{K}$. You must specify carefully any particular points in K that are required.

(e) Show that the emulator adjusted covariance can be written in the separable form:

$$\text{Cov}_K[f(x), f(x')] \;=\; R(x_1, x_1')r(x_2, x_2')$$

for suitable functions $R(x_1, x_1')$ and $r(x_2, x_2')$, which you should identify, where K again represents a vector of outputs of a large number of n runs evaluated on the boundary $\mathcal{K}$.

(f) Examine the behaviour of $\text{Cov}_K[f(x), f(x')]$ in the limit where both $x_1, x_1' \to \infty$ but where the difference $x_1 - x_1'$ remains finite. Comment on your answer.

Q4 A Bayes linear emulator with basis functions $g_j(x)$, unknown coefficients β_j and weakly stationary process $u(x)$, takes the form:

$$f(x) \;=\; \sum_{j=1}^{p} \beta_j g_j(x) \;+\; u(x) \;=\; g(x)^T \beta \;+\; u(x)$$

Prior specifications for $\beta \equiv (\beta_1, \ldots, \beta_p)^T$ and $u(x)$ are given by:

$$\mathrm{E}[\beta] = \mu_\beta, \qquad \mathrm{Var}[\beta] = \Sigma_\beta, \qquad \mathrm{E}[u(x)] = 0,$$

$$\mathrm{Cov}[u(x), u(x')] = \sigma^2 c(x - x'), \qquad \mathrm{Cov}[\beta, u(x)] = 0,$$

where $c(x - x')$ represents the usual squared exponential covariance structure with $\sigma > 0$ and where Σ_β is of full rank. We define the model output from n runs to be $D = \left(f(x^{(1)}), \ldots, f(x^{(n)})\right)^T$ and also $U = \left(u(x^{(1)}), \ldots, u(x^{(n)})\right)^T$ with $\mathrm{Var}[U] \equiv \Omega$ where Ω is an $n \times n$ covariance matrix with individual elements $\Omega_{ij} = \sigma^2 c(x^{(i)} - x^{(j)})$. We also define the $n \times p$ design matrix X with elements $X_{ij} = g_j(x^{(i)})$.

(a) With the above specifications, show that we can write $D \;=\; X\beta + U$.

(b) Show that the variance of the weakly stationary process $u(x)$ adjusted by the run data D is given by:

$$\mathrm{Var}_D[u(x)] \;=\; \sigma^2 - a(x)^T \Omega^{-1} a(x) + a(x)^T \Omega^{-1} X \mathrm{Var}_D[\beta] X^T \Omega^{-1} a(x)$$

where you must define the function $a(x)$ but can assume that $\mathrm{Var}_D[\beta] = (X^T \Omega^{-1} X + \Sigma_\beta^{-1})^{-1}$. *Hint*: you can use the following matrix identity which states that for matrices A, B, C, G of appropriate dimension:

$$(A + BCG)^{-1} \;=\; A^{-1} \;-\; A^{-1} B\, (C^{-1} + G A^{-1} B)^{-1} G A^{-1}$$

(c) Comment briefly on your answer to **Q4**(b) in relation to a simple emulator without regression terms.

(d) Show that the covariance of the regression terms with the weakly stationary process, adjusted by the run data D, is given by:

$$\mathrm{Cov}_D[g(x)^T \beta, u(x)] \;=\; -\, g(x)^T \mathrm{Var}_D[\beta] X^T \Omega^{-1} a(x)$$

Hint: you can use the following matrix identity, which states that for matrices A, B, C, G of appropriate dimension:

$$AB\, (GAB + C)^{-1} \;=\; (BC^{-1}G + A^{-1})^{-1} BC^{-1}$$

(e) Use the results of **Q4**(b) and **Q4**(d) to find an expression for the emulator variance $\mathrm{Var}_D[f(x)]$ and hence show that

$$\mathrm{Var}_D[f(x^{(1)})] = 0$$

where $x^{(1)}$ is the location of the first run. Comment on the importance of this answer in relation to emulator evaluation at all known runs.