MATH70093– Computational Statistics
Assessed coursework —Autumn 2025

Preparing your Coursework

1. **Hand-out date**: December 1st 2025 - 9am.
2. **Hand-in date**: December 12th 2025 - noon.
3. **Please use the provided Rmarkdown template file** to write your report. Code should be provided in the appendix (which will be done automatically by the template provided). The appendix will not count towards the page limit. Ensure your submitted file has tidy and well documented code chunks.
4. The report should be properly structured, and should be written using complete sentences. Marks are given both for the content of the report (correctness of code, numerical answers, etc.) and the quality of the presentation (clarity of plots, explanations, etc.).
5. The main part of the report (i.e., excluding the code appendix) should not exceed 20 pages including all explanations, mathematics and plots.
6. At the beginning of your report **you must include the following statement**:
   "I, [insert CID], certify that this assessed coursework is my own work, unless otherwise acknowledged, and includes no plagiarism. I have not discussed my coursework with anyone else except when seeking clarification with the module lecturer via email or on MS Teams. I have not shared any code underlying my coursework with anyone else prior to submission. "
7. All coding should be done in 'Rmarkdown'. Please ensure your submitted file has **tidy and well documented code chunks** in appendix.

Submitting your coursework

1. Submit **via Blackboard** before the deadline a pdf version of your report.
2. The name of your submitted file should begin with **CW_** followed by your CID, e.g. if I were to submit a coursework, I would submit: **CW_00830053.pdf**.
3. Please note your report will be checked for plagiarism.

1. *(50% of total marks)* Consider the following target density:

$$f(x) = k\left\{\exp[-3(x-2)] + \exp[-(x-30)^2] + \exp[-(x-20)^2/0.01]\right\} 1_{\{x\geq 2\}}$$

where k is a normalising constant and $1_{\{\cdot\}}$ denotes the indicator function.

(a) Implement a random-walk Metropolis-Hastings algorithm with normal noise to sample from this target distribution. Investigate the impact of the variance of the random noise on the convergence and mixing quality of the chain. Discuss your findings.

(b) Devise and implement a Metropolis-Hastings algorithm which does not propose values of x outside of the support of f. Precisely describe this algorithm with mathematical equations. Verify graphically that the MCMC algorithm has reached convergence.

(c) Devise and implement a parallel tempering algorithm to sample from f. Provide a pseudo-code describing the algorithm you constructed. Precisely state the temperatures you are using as well as the transition kernels for each chain. Verify graphically that the sampler is exploring the full space and correctly sampling from the target distribution.

(d) Compare the three algorithms (from part (a), (b) and (c)) in terms of convergence, mixing properties of the resulting Markov Chains and computational costs.

(e) Suppose that we want to estimate $E(X)$ for X following a distribution with p.d.f. f. Denote by $\hat{I}_n^{(a)}$,$\hat{I}_n^{(b)}$ and $\hat{I}_n^{(c)}$ the estimate of $E(X)$ using a chain of length n produced by the MCMC algorithms from questions (a), (b) and (c) respectively. Write down the equation of a 95% confidence interval for $\hat{I}_n^{(a)}$,$\hat{I}_n^{(b)}$ and $\hat{I}_n^{(c)}$. Produce three plots featuring $\hat{I}_n^{(a)}$, $\hat{I}_n^{(b)}$ and $\hat{I}_n^{(c)}$ respectively as a function of n along with a 95% confidence interval for these estimates. Assess how the three estimates differ from each other, and compare them to the reference value of $E[X]$ obtained by numerical integration.

2. *(50% of total marks)* In this question we explore quantile regression, using the following asymmetric loss

$$\ell_\tau(y, q) = \begin{cases} \tau(y-q), & y > q, \\ (\tau-1)(y-q), & y \le q, \end{cases}$$

for a quantile level $\tau \in (0,1)$. Let $R_\tau(q; x) = \mathbb{E}\big[\ell_\tau(Y, q) \mid X = x\big]$ denote the conditional risk. It is known that $q^*_\tau(x) := \arg\min_q R_\tau(q; x)$ is the conditional τ-quantile of $Y \mid X = x$. We therefore use empirical risk minimisation with this loss to learn an estimator $f_\tau(x)$ for the conditional τ-quantile.

We use the `Boston` dataset from the `MASS` package in R. Let the response be `medv` (median house price). Our goal is to predict the conditional quantiles of `medv` based on the features `lstat` and `rm`. Randomly split the data into a training set (70%) and a test set (30%). Throughout this question, you must implement the loss yourself. You may use generic optimisation routines such as `optim`, but you may not use pre-built quantile regression or distributional regression packages. In each part, explain how you performed the optimisation (the choice of optimiser, the initial state, any associated settings, and how convergence was assessed).

(a) Consider the baseline linear model $f^{(0)}_\tau(x) = \beta_{0,\tau} + \beta_{1,\tau}\,\texttt{lstat} + \beta_{2,\tau}\,\texttt{rm}$.

- Implement the loss and its empirical mean in R. Fit the parameters $(\beta_{0,\tau}, \beta_{1,\tau}, \beta_{2,\tau})$ for each $\tau \in \{0.1, 0.5, 0.9\}$ by minimising the empirical loss over the training set using `optim` and report the estimated coefficients.
- For each τ evaluate the empirical loss on the test set.
- The empirical coverage

$$\widehat{C}^{(0)}_\tau = \frac{1}{n_{\text{test}}} \sum_{i \in \text{test}} \mathbf{1}\{Y_i \le f^{(0)}_\tau(X_i)\},$$

  measures the proportion of test responses whose actual value lies *below* the predicted τ-quantile. Since a true τ-quantile satisfies

$$\Pr\big(Y \le q_\tau(X)\big) = \tau,$$

  we ideally expect $\widehat{C}^{(0)}_\tau \approx \tau$ for a well-calibrated model. For each τ, compute the empirical coverage and compare it with the target level τ.

(b) Extend the model by introducing additional quadratic terms $\texttt{lstat}^2$, $\texttt{rm}^2$ and the interaction term $\texttt{lstat} \times \texttt{rm}$. Let $f^{(1)}_\tau$ be the resulting linear model.

- Fit $f^{(1)}_\tau$ for each $\tau \in \{0.1, 0.5, 0.9\}$ by minimising the empirical loss over the training set and report the coefficients.
- Evaluate the test loss and empirical coverage

(c) Introduce an ℓ_2 (ridge) penalty on the coefficients of the nonlinear model. For a given $\lambda \ge 0$, define the ridge-penalised empirical risk

$$\widehat{R}^{\text{ridge}}_{\tau,\lambda}(\beta) = \frac{1}{n_{\text{train}}} \sum_{i \in \text{train}} \ell_\tau\big(Y_i, f^{(1)}_\tau(X_i; \beta)\big) + \lambda \|\beta\|_2^2.$$

- For each $\tau \in \{0.1, 0.5, 0.9\}$ and a set of penalty values λ within $[10^{-4}, 10^{-1}]$, fit the ridge-penalised model over the training set, using k-fold cross-validation on the empirical risk, within the training set to select λ for each τ.
- For the selected λ, report the ridge-penalised coefficients $\hat{\beta}^{\text{ridge}}_{\tau,\lambda}$ and the test loss and empirical coverage

(d) Discuss and interpret the relative performances of each model, in terms of (i) coefficient magnitudes, (ii) test loss, (iii) overfitting and (iv) empirical coverage. In particular, explain the reason for the outcomes you have observed.

(e) The true quantiles never cross, i.e. $q_\tau(x) < q_{\tau'}(x)$ for all x and $0 \le \tau < \tau' \le 1$.

- Will the quantile prediction models you have trained automatically satisfy a similar relationship in general? Explain your answer.
- Suggest an approach for training a quantile prediction model for quantiles $\tau = \{0.1, 0.5, 0.9\}$, which discourages such crossings. Provide details of the associated optimisation problem, the loss and any other relevant details. Note that you do not need to implement this, just describe it.