# Computer Intensive Statistics: APTS 2025–26 Assessment

Andi Wang (andi.wang@warwick.ac.uk),

July 2026

The intention of these exercises is to provide an opportunity for you to demonstrate what you can do. They are intended to include many of the disparate aspects of *Computer Intensive Statistics* and it is not envisaged that anyone should be asked to answer all of them (although, of course, you're welcome to do so if you find it interesting). Discuss which questions you should attempt with your supervisor. It is anticipated that most students will attempt about two of these questions and won't expend much more than one day of effort on the task.

1. Choose a multimodal univariate probability density and:
    1. Implement an exact slice sampler for this density. You'll need to identify the level sets either analytically or numerically and to sample from these level sets.
    2. Implement a random walk Metropolis algorithm for which your chosen density is the invariant distribution.
    3. Implement a Metropolised slice sampler for this density.
    4. Compare the three algorithms which you have implemented, taking into account the quality of the approximation and the computational cost of the various approaches.
2. Exchange Algorithm. The goal of this exercise is to implement the Exchange Algorithm on a simple example, which is a simplified version of the Ising model example discussed in the lectures. Model definition: we imagine $N$ particles a discrete 1d lattice, each with a spin $x = (x_1, \ldots, x_N) \in \{-1, +1\}^N$. The likelihood of a configuration given inverse temperature $\theta \geq 0$ is
$$f(x|\theta) = \frac{1}{Z(\theta)} \exp\left(\theta \sum_{i=1}^{N-1} (x_i \cdot x_{i+1})\right).$$
We are interested to infer $\theta$ from a given realisation of $x$. So we put a prior $\theta \sim \text{Exp}(1)$. To be concrete, let's set $N = 20$, and true parameter $\theta_{\text{true}} = 0.8$. We will need to be able to generate synthetic data $x \sim f(x|\theta)$.
    1. Suppose we fix all positions of $x$ except $x_i$: what is the (categorical) conditional distribution of $x_i$ on $\{-1, +1\}$ given all the others $x_{-i}$? Implement a random scan Gibbs sampler to sample from $f(x|\theta)$ for any fixed $\theta \geq 0$.
    2. Run your Gibbs sampler for $\theta_{\text{true}} = 0.8$ for large number (e.g. 10,000 iterations) from any initialisation (e.g. i.i.d. uniform spins) to obtain an observation $x^{\text{obs}}$. (Note it is technically possible to sample exactly from $f(x|\theta)$ by calculating all $2^N$ probabilities.)
    3. Consider using a random walk proposal (with standard deviation 0.1) on $\theta$, and implement the Exchange algorithm. For the inner step of sampling from $f(x|\theta)$, you can use your Gibbs sampler for a modest number of steps (e.g. 50) started from $x = x^{\text{obs}}$. (Technically this does invalidate the sampler as the Exchange algorithm requires exact samples to be correct - there are ways to do exact sampling based on coupling from the past but that goes beyond the scope of this exercise.) Check that the resulting histograms are sensible.

4. As an alternative, implement a naive (and wrong!) random walk Metropolis chain targeting the posterior where we ignore the intractable $Z(\theta)$ term - that is, we imagine the likelihood to be

$$\tilde{f}(x|\theta) = \exp\left(\theta \sum_{i=1}^{N-1} (x_i \cdot x_{i+1})\right).$$

Compare the resulting traceplots of the Exchange algorithm vs the naive Metropolis–Hastings.

3. Theoretical exercise on the involution framework (pseudo-marginal methods). Pseudo-marginal MCMC, formally introduced by Andrieu and Roberts (2003) was one of the significant breakthroughs in MCMC, substantially expanding the scope of what MCMC could be applied to. In this question we will flesh out some of the details which were touched on in the lectures. Recall the model setting: we wish to perform Bayesian inference on posterior

$$\pi(\theta|y) \propto p(\theta) f(y|\theta)$$

on $\Theta$, where $f(y|\theta)$ is an intractable likelihood which cannot be evaluated. However we assume we have access to a family of nonnegative random variables, indexed by $\theta \in \Theta$, $W \sim Q_\theta$ with the property that $\mathbb{E}_{Q_\theta}[W] = f(y|\theta)$, in other words we have an unbiased estimator for the likelihood.

1. We consider the following extended target distribution:

$$\pi(\theta, w|y)\mathrm{d}\theta \propto p(\theta) w Q_\theta(\mathrm{d}w)\mathrm{d}\theta.$$

Show that this possesses the exact posterior $\pi(\theta|y)$ as a marginal. We will now build an MCMC sampler to target this joint distribution.

2. We assume we have some proposal kernel for $\theta$, $q(\theta'|\theta)$. Given the current position $(\theta, w)$, at each iteration of pseudomarginal MCMC, we will simulate a proposal $\theta' \sim q(\theta'|\theta)$ and a new unbiased estimator $w' \sim Q_{\theta'}(\cdot)$. Write down the measure $\mu(\mathrm{d}\theta, \mathrm{d}w, \mathrm{d}\theta', \mathrm{d}w')$ on the augmented state space $\mathcal{E} = \Theta \times [0, \infty) \times \Theta \times [0, \infty)$, which has $\pi(\theta, w|y)$ as its first marginals and the remaining $\theta', w'$ are generated according to this pseudo-marginal proposal.

3. We take involution $\phi(\theta, w, \theta', w') = (\theta', w', \theta, w)$. Show that the resulting Radon–Nikodym derivative $r = \frac{\mathrm{d}(\phi \circ \mu)}{\mathrm{d}\mu}$ is given by

$$r(\theta, w, \theta', w') = \frac{p(\theta') w' q(\theta|\theta')}{p(\theta) w q(\theta'|\theta)}.$$

(Hint: for this involution, the Jacobian is simply 1.) This shows that a valid acceptance probability is

$$1 \wedge \frac{p(\theta') w' q(\theta|\theta')}{p(\theta) w q(\theta'|\theta)},$$

as given in the lectures, which is equivalent to the standard MH acceptance probability where the intractable likelihoods have been replaced by unbiased estimators. (Note that for correctness of the algorithm it is crucial that we carry around the pair $(\theta, w)$: in particular if we reject a move, we keep the current value $w$ in the denominator rather than re-drawing it at each step. Unfortunately this can lead to "sticky" behaviour which is a known issue of pseudo-marginal methods.)

4. Consider the Ising model on a graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$ which was briefly mentioned in lectures. Let $m = |\mathcal{V}|$. Recall that the distribution over the $\pm 1$-valued binary variables attached to each vertex has probability mass function

$$p(x_1, \ldots, x_m) = \frac{1}{Z} \exp\left(J \sum_{(i,j) \in \mathcal{E}} x_i x_j\right),$$

where $Z$ is a normalising constant and the sum is over all adjacent vertices with the graph. Consider the *ferromagnetic* case in which the coupling strength $J > 0$.

This question concerns the relationship between the Ising model and a related bond percolation model and a data augmentation strategy which allows a very efficient algorithm to be implemented for this model.

Consider adding a Bernoulli variable to every edge in the graph such that there is a $U_{i,j}$ associated with every $(i, j) \in \mathcal{E}$. Conditional upon the value of $\boldsymbol{X} = (X_1, \ldots, X_m)$, these variables are mutually independent. If $X_i \neq X_j$ then $U_{i,j} = 0$, otherwise $U_{i,j} \sim \mathsf{Ber}\,(1 - \exp(-2J))$. This may seem rather a peculiar thing to do at first, but the interpretation is that bonds are introduced at random between those adjacent vertices which take common values; the stronger the coupling strength the greater the probability that adjacent like vertices are bonded.

1. Write down the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{U}$ where $\boldsymbol{U} = \{U_{i,j} : (i, j) \in \mathcal{E}\}$.
2. Simplify your expression to express the probability distribution (up to a constant of proportionality) in terms of: the number of unlike adjacent vertices, the number of bonded like adjacent vertices and the number of unbonded like adjacent vertices implied by $\boldsymbol{X}$ and $\boldsymbol{U}$.
3. What is the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{U}$?
4. Identify a Gibbs Sampling algorithm in which one iteratively updates first the entirety of $\boldsymbol{X}$ and then the entirety of $\boldsymbol{U}$.
5. Why might the Gibbs Sampler described here be expected to outperform the naïve Gibbs sampling strategy for the Ising model?

5. (For those who want a more explicitly statistical exercise.) Bootstrap methodology was not discussed in the lectures but it detailed in the notes, and remains a popular computer intensive method. Consider the application of bootstrap methods to a simple random sample of size $n = 100$ obtained from a standard normal population:

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{N}\,(0, 1).$$

1. Simulate such a sample. Let us call it $\boldsymbol{x}^\star = (x_1^\star, \ldots, x_n^\star)$.
2. We know that the most immediately intuitive estimator of the population variance,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

   is biased.

   (i) What is the bias in this case?
   (ii) Use a bootstrap method to estimate the bias of this estimator using $\boldsymbol{x}^\star$ as the real sample.
   (iii) Repeat (ii) a number of times to assess the bootstrap method in this case.

3. Population kurtosis is another quantity which we might wish to estimate from sample data. One simple estimator of this quantity might be:

$$\hat{k} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n(\hat{\sigma}^2)^2}.$$

   (i) Estimate $\hat{k}$ for $\boldsymbol{x}^\star$.
   (ii) Estimate the bias of $\hat{k}$ using a bootstrap method.
   (iii) Obtain a 95% confidence interval (try a bootstrap percentile interval, or a more sophisticated method if you prefer) for the population kurtosis.

6. If your PhD involves a model[1] for which you have been conducting inference by other means, try implementing one or two simple Markov chain Monte Carlo algorithms in order to obtain Bayesian estimates of the unknown parameters (you'll need to choose some prior distributions if you don't already have these). Contrast these estimates with those obtained with whatever other methods you've been using.

[1] If it's a complicated model you might want to consider a simplification of that model, this exercise isn't intended to take a very long time and implementing good MCMC algorithms for complex models can be very time consuming.