Bootstrap and Resampling Methods, Spring 2025

Homework exercise 4

Due date: 1 July 2025

1. **MCMC on normal distribution.**
   For a standard normal distribution $N(0,1)$, implement a Metropolis algorithm using a $N(0,\sigma^2)$ "random walk" sampling function for different values of σ (in this mode, $x_{\text{new}} = x_{(t)} + N(0,\sigma^2)$). Start all chains from a random draw from $N(0,1)$.

   (a) Empirically estimate the autocorrelation curves for $\sigma \in \{1, 1.5, 2, 2.5, 3, 5, 10\}$. Plot them (as a function of time) and discuss which one is best and why.

   (b) With the same values of σ, calculate estimates of $E(X^4)$ for $X \sim N(0,1)$, using MCMC with $n = \{1000, 10000, 100000\}$ samples.

   (c) Calculate the true expectation (integration in parts...) or find its true value, and compare to your results above.

   (* Extra credit) Can you design a different MCMC algorithm, not using the normal random walk sampling, that will give guaranteed better estimates of $E(X^4)$?
   Hint: Think how to accomplish the lowest possible level of dependence between consecutive samples.

2. **Gibbs sampling on a binormal example.**
   Consider a simple bivariate normal distribution:

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

   Implement a Gibbs sampling approach to generate data according to this model using only univariate random draws. That is, write down the conditional distribution X_2 given X_1 and vice versa, and use this to performa Gibbs sampling. Perform 100 repetitions of random-scan sampling (where you choose X_1 or X_2 randomly), and 100 of systematic-scan sampling (where you alternate between drawing X_1 given X_2 and vice versa). Run all 200 of the Gibss sampling experiments for 10000 iterations each, and use the resulting sequence to generate 200 estimates of

$$P(1 \leq X_2 \leq 2 | 1 \leq X_1 \leq 2)$$

   Comparing to the true value of this probability, evaluate the performance of the two Gibbs samplers in terms of bias and variance.

3. **Importance sampling to calculate integral (problem 24.1 of All of statistics).**

Consider estimating $I = \int_1^2 \frac{e^{-(x^2/2)}}{\sqrt{2\pi}} dx$.

(a) Estimate I using "basic Monte Carlo": sampling $X_i \sim U(1,2)$ and averaging the values of $h(X_i) = \frac{e^{-(X_i^2/2)}}{\sqrt{2\pi}}$. Use $N = 10^4$ samples and estimate empirically the standard error of your Monte Carlo estimate.

(b) Find an analytical expression for the standard error of your estimate in (a). Compare to the estimated error in (a).
Hint: use the cumulative standard normal distribution where needed.

(c) Estimate I using importance sampling: Take g to be $N(1.5, v^2)$ with $v \in \{0.1, 1, 10\}$. Compute the estimated standard error in each case with $N = 10^5$ samples, and compare them to the results in (a),(b).

(d) (* +5) Calculate the true standard errors for the importance sampling schemes in (c).

(e) Find the optimal importance sampling function g^*. What is the standard error using g^*?