Bootstrap and Resampling Methods, Spring 2021

Homework exercise 4

Due date: 13 June 2021 (day of last class)

Submission format: Please include your code in your submission as an appendix.

1. **Questions on Assaf's presentation on testing isotropy**

   (a) Explain why we cannot use regular non-parametric bootstrap to test this hypothesis. Try to give two distinct reasons, one related to the covariance and sampling mechanism of this specific problem, and the other related to the general problem of hypothesis testing with bootstrap.

   (b) Assume that instead of the null hypothesis of isotropy we want to test the null hypothesis of anisotropy in the canonical direction (i.e the axis of the ellipse are in directions 0 and $\pi/2$). Note this null hypothesis includes in it isotropy as a special case. Suggest a change to the parametric bootstrap algorithm on slide 24 to perform this.

   (c) (* Extra credit) Python code implementing the standard test Assaf designed is available from the class home page. Implement the test from the previous item, generate appropriate data under the null and alternative and demonstrate its performance.
   Note: This problem has been updated on 7/6 with corrected code

2. **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.

3. **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.