Assignment I: Reducing the variance of Monte Carlo Estimators.

Computational Statistics
Instructor: Luiz Max de Carvalho

October 14, 2020

Hand-in date: 04/11/2020.

Background

Suppose we are interested in an integrable function $\varphi(\cdot)$ and its expectation under a probability distribution with law π, $I = \int_{\mathbb{X}} \varphi(x)\pi(x)dx$. Monte Carlo is a large class of sampling algorithms to approximate integrals. It depends on constructing an estimator

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i), \tag{1}$$

where $X_i \sim \pi$. The estimator in (1) can be improved in a number of ways. In particular, the procedure known as *Rao-Blackwellisation* allows one to obtain estimators that have lower variance with the same bias, i.e., are more efficient.

Questions

1. Show that, if $\pi(x) = \int_{\mathbb{X}} g(x \mid y)\varphi(y)\, dy$, then both

$$\hat{I}_n^{\text{MC}} = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i),\ X_i \sim \pi, \text{ and}$$

$$\hat{I}_n^{\text{RB}} = \frac{1}{n}\sum_{i=1}^{n} E[\varphi(X_i) \mid Y_i],\ Y_i \sim \varphi,$$

   converge to $E_\pi[\varphi(X)]$.

2. For each of the following cases, generate random variables X_i and Y_i and compute the average and variance of the "vanilla" Monte Carlo and Rao-Blackwellised estimators:

   (a) $X \mid Y \sim \text{Poisson}(Y)$, $Y \sim \text{Gamma}(a, b) \implies X \sim \text{Negative-Binomial}$;

   (b) $X \mid Y \sim \text{Normal}(0, Y)$, $Y \sim \text{Gamma}(a, b) \implies X \sim \text{Generalised-t}$;

   (c) $X \mid Y \sim \text{Binomial}(n, Y)$, $Y \sim \text{Beta}(a, b) \implies X \sim \text{Beta-Binomial}$;

3. Propose a Rejection Control (see Question 5 in exercise sheet 1) algorithm to sample from each of the targets in the previous question;

4. Compute (empirically) the n_{eff} for the results in the previous question. How does the algorithm perform in each case? Do you have to fiddle with the value of $c > 0$ in order to obtain better performance? Discuss.

5. Compare the variance of the Rejection Control estimate with the variance of the "vanilla" Monte Carlo estimator and the Rao-Blackwellised estimator;

6. What other improvements (in terms of variance and bias) would you propose? Can you come up with an algorithm that dominates all algorithms studied here for the targets under consideration?