Monte Carlo Methods for Optimization
Statistics 202C
Professor: Mark S. Handcock

Project 1

Due Friday, April 24, 2020

In this homework we consider two problems based on situations introduced in the lectures and textbook, **BZ**. Solutions require computation and the analysis of the results.

1) *Importance sampling and the effective number of samples:* In a 2-dimensional plane, suppose the target distribution $\pi(x, y)$ is a symmetric Gaussian with mean $\mu = (2, 2)$ and standard deviation $\sigma = 1$. Suppose we use an approximate distribution $g(x, y)$ as the trial density which is a Gaussian with mean $\mu_0 = (0, 0)$ with standard deviation σ_0. So

$$\pi(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}[(x-2)^2+(y-2)^2]}, \qquad g(x, y) = \frac{1}{2\pi\sigma_0^2} e^{-\frac{1}{2\sigma_0^2}[x^2+y^2]}$$

We wish to estimate the quantity $\theta = \int \sqrt{y^2 + x^2} \cdot \pi(x, y) \cdot dxdy$.

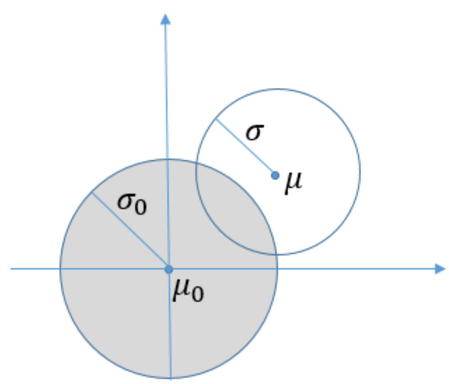


Figure 1: Graphical representation of the sample space

In this question, we will compare the effectiveness of three alternative reference probabilities used in importance sampling.

Alternative 1 Compute $\hat{\theta}_1$: estimate θ by drawing n_1 samples directly from $\pi(x, y)$. Since the two dimensions are independent, we can sample x and y from the 1D marginal Gaussians.

Alternative 2 Compute $\hat{\theta}_2$: estimate θ by drawing n_2 samples from $g(x, y)$ with $\sigma_0 = 1$.

Alternative 3 Compute $\hat{\theta}_3$: estimate θ by drawing n_3 samples from $g(x, y)$ with $\sigma_0 = 4$.

a) Plot $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ over n (increasing n so that they converge) in one figure to compare the convergence rates. Before running the experiment, try to guess whether Alternative 3 is more effective than Alternative 2. *Hint:* You can use a log plot at a few points $n = 10, 100, 1000, 10000, \ldots$

b) Estimate the "effective sample size" for each alternative. In Lecture 2 Supplement we defined the effective sample size as:

$$\text{ess}^*(n) \equiv \frac{n\mathbb{V}_\pi(\hat{\theta})}{\mathbb{V}_g(\hat{\theta})}$$

In Lecture 2 Supplement and **BZ**, eq. (2.6), we suggested the approximation

$$\text{ess}^*(n) \approx \frac{n}{1 + \mathbb{V}_g(\omega_n)} \approx \frac{n}{1 + S_g^2(\omega_n)} \equiv \text{ess}(n),$$

where $S_g^2(\omega)$ is the sample variance of the importance weights ω. In this question we assess how good this approximation to $\text{ess}^*(n)$ is. Since the samples in Alternative 1 are each "effective" samples of size 1 as they are directly drawn from the target distribution, we use $\text{ess}^*(n_1) = n_1$ as the truth and compare the effective sample sizes for Alternative 2 and Alternative 3 to Alternative 1.

Compute $\text{ess}(n)$ and $\text{ess}^*(n)$ for Alternatives 2 and 3.

Plot $\text{ess}^*(n_2)$ and $\text{ess}(n_3)$. Plot $\text{ess}^*(n_2)$ over $\text{ess}(n_2)$, and $\text{ess}(n_3)$ over $\text{ess}^*(n_3)$.

Discuss your results.

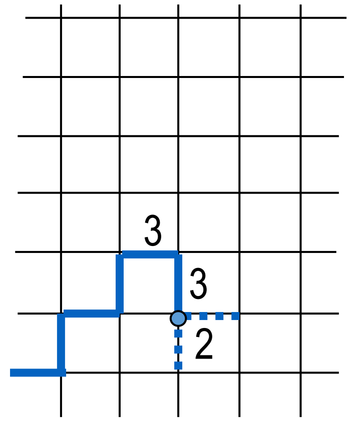
2) *Estimating the number of Self-Avoiding-Walks in an $(n+1)\times(n+1)$ grid:* This problem refers to the problem of counting the number of Self-Avoiding-Walks as described in Example 1.3 of **BZ**.

Suppose we always start from position $(0, 0)$, i.e. the lower-left corner. We design a trial (reference) probability $p(r)$ for a Self-Avoiding-Walk (SAW) $r = (r_1, r_2, \ldots, r_N)$ of varying length N. Then we sample a number M SAWs from $p(r)$, and the estimation process is described in the left panel of Figure 2.

Your grade will be based on the quality of results and analysis of different designs. The textbook shows results obtained by prior year students. At each step, the trial probability $p(r)$ can choose to stop (terminate the path) or walk to the left/right/up/down as long as it does not intersect itself. Each option is associated with a probability (of your design) and these probabilities sum to 1 at each point.

a) What is the total number, K, of SAWs for $n = 10$? *Hint*: Try $M = 10^7$ to 10^8. To clarify: a square is considered a 2×2 grid with $n = 1$. Plot K against m (in a log-log plot) and monitor whether the Sequential Importance Sampling (SIS) process has converged. Try to compare at least three different designs for $p(r)$ and see which is more efficient. For example, you may start from a path that you find before multiple times, as long as you compute the $p(r)$ correctly.

$$\begin{aligned}
\mathsf{K} &= \sum_{r \in \Omega_{n^2}} 1 = \sum_{r \in \Omega_{n^2}} \frac{1}{p(r)} p(r) \\
&= E\left[\frac{1}{p(r)}\right] \\
&\approx \frac{1}{M} \sum_{i=1}^{M} \frac{1}{p(r_i)}
\end{aligned}$$



$p(r) = \prod_{t=1}^{N} \frac{1}{n_t}$ where n_t is the number of options to move from step t.

Figure 2: Graphical representation of the lattice. The left panel provides the equations for the primary estimator and the right the spatial path followed by a representative r

b) What is the total number of SAWs that start from $(0,0)$ and end at (n,n)? Here you can still use the same sampling procedure above, but only record the SAWs which successfully reach (n,n). As we noted, the true number is 1.5687×10^{24}.

Bonus: If you get the result closest to this number in class, you will be awarded 3 bonus points.

c) For each experiment in a) and b), plot the distribution of the lengths, N, of the SAWs in a histogram (Think: Do you need to weight the SAWs in calculating the histogram?) and visualize (print) the longest SAW that you find.