Asn 1

ZHANG Yan

September 17, 2021

Question 1 (Importance Sampling for Rare Event Probability)

(a) For $a = 2, 4, 6$, calculate $P(X > a) = \int_a^\infty \phi(x)\mathrm{d}x$, where $X \sim N(0, 1)$. With 100000 samples, use the 'Naive' approach (simulate samples from $\phi(x)$), the non-ratio version Importance Sampling (simulate samples from $q(x) = e^{-(x-a)}\mathbf{1}_{[a,\infty)}(x)$) or the ratio version Importance Sampling (simulate samples from $q(x) = 0.5t_4(x) + 0.5e^{-(x-a)}\mathbf{1}_{[a,\infty)}(x)$) to calculate the point estimates and the confidence intervals.

(b) Calculate the optimal proposals of both non-ratio or ratio version Importance Sampling. Compare the IS proposals in (a) with the corresponding optimal proposals by plots.

(c) For the non-ratio version IS, we find that the previous IS proposal too flat compared to the optimal proposal. To improve the performance of Importance Sampling, we assume the proposal family $q(x|b) = be^{-b(x-a)}\mathbf{1}_{[a,\infty)}(x)$. To find the best parameter b, notice that the asymptotic variance can be estimated by

$$\begin{aligned}
\sigma^2(b) &= \mathbb{E}_b[(w_b(X)f(X) - \mu)^2] \\
&= \int \frac{(\phi(x)\mathbf{1}_{[a,\infty)}(x) - \mu q(x|b))^2}{q(x|b)}\mathrm{d}x = \int \frac{(\phi(x)\mathbf{1}_{[a,\infty)}(x))^2}{q(x|b)}\mathrm{d}x - \mu^2 \\
&\approx \hat{S}(b|b_0) - \mu^2 = \frac{1}{n}\sum_{i=1}^n \frac{(\phi(x_i)\mathbf{1}_{[a,\infty)}(x_i))^2}{q(x_i|b)q(x_i|b_0)} - \mu^2, x_i \sim q(x|b_0), i = 1, \ldots, n.
\end{aligned}$$

So, instead of minimizing $\sigma^2(b)$ with respect to b, we can minimize $\hat{S}(b|b_0)$ for a given b_0. Further more, we can even do the optimization iteratively (replace b_0 by the optimal b), which lead to the Adaptive Importance Sampling (AIS) methodology. When $a = 6$, $b_0 = 1$ and $n = 100$, build the AIS algorithm and find the optimal b, draw the log value of asymptotic variance (estimated with 100000 new samples) corresponding to the initial and optimal b's against iteration numbers. When does this algorithm converge?

(d) In many cases, instead of simple $P(X > a)$, we are more interested in estimating $P(h(X) > a)$ for a function $h(x)$ and a large a, which can be really complicated especially when it is a multidimensional problem. For illustration, we return to the original problem of the estimation for $P(X > a)$, $X \sim N(0, 1)$, but this time we consider the gamma family $q(x|b, c) = b^2(x - c)e^{-b(x-c)}\mathbf{1}_{[c,\infty)}(x)$ and try to obtain the optimal values for b and

c automatically. So, consider

$$\hat{S}(b,c|b_0,c_0)=\frac{1}{n}\sum_{i=1}^{n}\frac{(\phi(x_i)\mathbf{1}_{[a^*,\infty)}(x_i))^2}{q(x_i|b,c)q(x_i|b_0,c_0)},x_i\sim q(x|b_0,c_0),i=1,\ldots,n$$

where $a^*=\min(a,a^{(1-\varepsilon)})$, $a^{(1-\varepsilon)}$ is the $(1-\varepsilon)$-quantile of $\{x_i\}$. When $a=6$, $b_0=1$, $c_0=0$, $n=100$ and $\varepsilon=0.1$, build an AIS algorithm to find the optimal b and c, draw the log value of asymptotic variance (estimated with enough new samples) corresponding to the initial and optimal parameters against iteration numbers. When does this algorithm converge? The spirit behind this algorithm actually coincide with that of the Generative Adversarial Networks, think about why (no need to answer this).

Question 2 (Modified Rejection Sampling)

To draw samples from the target distribution $\pi(x)$, remember the procedure of Rejection Sampling (RS): 1) Draw initial samples $\{x_1,\ldots,x_m\}$ from an envelope distribution $q(x)$, and calculate the corresponding ratios or weights $\{w_1,...,w_m\}, w_j=w(x_j)=\pi(x_j)/q(x_j)$; 2) Calculate $C=\sup w(x)$, and accept each sample x_j with the probability $p_j=w_j/C$ to obtain the final samples $\{x_1^*,\ldots,x_n^*\}$. A modified version of RS is to replace C by $\max w_j$. We will assume m to be fixed and n to be random, and check the difference between these two methods numerically.

(a) Obviously the first benefit of replacing $\sup w(x)$ by $\max w_j$ is that it saves efforts to calculate the maximum value especially when the shape of the target or proposal are complicated or non-smooth. Consider drawing samples from $\pi(x)=0.5N(x|-2,0.5^2)+0.5N(x|1,1^2)$ based on $q(x)=t_1(x)$. Draw the plot of $w(x)$ and calculate C.

(b) We know that the expected acceptance rate for the original RS is just $1/C$. As $\max w_j<C$, the modified RS may have higher acceptance rate. Set $m=5,10,20,50,100,200,500,1000$. Under each m, implement these two kinds of Rejection Sampling $K=3000$ times to estimate and compare their acceptance rates and calculate the probability that the modified version has more final samples. Display your results with plots or tables.

(c) The third advantage is that different from $\sup w(x)$, $\max w_j$ can always be obtained even if the tail of $q(x)$ is lighter than that of $\pi(x)$. Consider a simple toy example that $\pi(x)=N(x|0,1^2)$, $q(x)=N(x|0,\sigma^2)$, how little σ can be so that we can still obtain reasonable final samples (at least for very big m) in the sense that its histogram still looks like the target distribution? Use mathematics or simulations to justify your guess, rigorously or intuitively.

(d) Finally, let's talk about some bad things about the modified RS procedure. It is actually a biased method, which means that the final samples do not follow the target distribution. Set $m=5$, $\pi(x)=N(x|0,1^2)$, $q(x)=N(x|0,3^2)$. Try to test the bias with histograms: Run $K=20000$ repetitions for both the RS before and after modification, for the kth repetition, obtain a group of final samples $\{x_{k,1}^*,\ldots,x_{k,n_k}^*\}$ and assign weights $1/n_k$ for each sample in this group (there would be no sample to set the weights if $n_k=0$), finally pool the samples in the K groups to get the weighted samples $\{(x_{1,1}^*,1/n_1),\ldots,(x_{k,i}^*,1/n_k),\ldots,(x_{K,n_K}^*,1/n_K)\}$ whose histogram with little bin width can be viewed as the expectation of the distribution of the final samples (conditional on that

the final sample set is not empty). Compare the two obtained histograms in a single plot, where you should also draw the target density as a reference curve.