

Importance Sampling Based on the Kernel Density Estimator

XueGao Zhang

Name of organization: Department of Business Administration, Yunnan Normal University Business School
Address: No. 1 Business School Road, Kunming High-tech Zone
Kunming, China
E-mail: 853638670@qq.com

Abstract—Importance Sampling is an unbiased sampling method used to sample random variables from different densities than originally defined. The importance sampling densities should be constructed to pick up ‘important’ random variables to improve the estimation of a interesting statistics. In this article, we present an importance sampling in which its density function is constructed from the kernel density estimators. This method can generate a sufficient number of samples, and then increase the accuracy of the probability estimate.

Keywords—Kernel Density Estimator; Importance Sampling

I. INTRODUCTION

Importance Sampling is an unbiased sampling method used to sample random variables from different densities than originally defined. The importance sampling densities are constructed to pick ‘important’ values of input random variables to improve the estimation of a statistical response of interest. The use of input sampling densities will result in biased estimators if they are applied directly to the simulation results. However, the simulation results are weighted to correct for the use of the importance sampling densities, and this ensures that the importance sampling estimators are unbiased. In practice, Importance sampling can be challenging to implement efficiently, especially in a general framework that will allow solutions for many classes of problems[1]. One of the most commonly used uncertainty quantification methods is Latin Hypercube Sampling, a stratified sampling technique which places samplings in equi-probability bins throughout the space. We are often got the request such as: “I have run an LHS sample of size 50 and I want to take 50 more samples that are more tailored to the failure region.” A major concern is that even if we are able to provide an importance sampling density that will preferentially sample in the failure region, the resulting failure probability estimate may not be significantly improved due to the small number of samples that are able to be performed.

In this article, I present a new importance sampling, which based on the kernel density estimators. Simulation indicates its robustness and accuracy. This paper is organized as follows: section 2 provides background on importance sampling, including a discussion of current approaches; Section 3 discusses the importance sampling density based on kernel density estimators that we are using;

Section 4 give the simulations of this new methods. And Section 5 provides a summary.

II. IMPORTANCE SAMPLING BACKGROUND

Accurate computation of high-dimensional integrals is common to many engineering and scientific application. Monte Carlo methods have been commonly used for many years to approximation the expectation of functions of random variables. That is, when calculating the expectation:

$$E(r(X)) = \int r(x) f_X(x) dx \quad (1)$$

Where $r(X)$ is a response function, X is a multidimensional random variable, the estimator of $E(r(X))$ is:

$$\hat{E}_n(r(X)) = \frac{1}{n} \sum_{i=1}^n r(x_i) \quad (2)$$

Note that many quantities of interest can be cast as expectations, for example:

$$\text{Prob}(r(X) \in A) = E[I_{\{A\}}(r(X))] \quad (3)$$

where, $I_{\{A\}}$ is a indicator function, namely $I_{\{A\}} = 1$ when $r(X) \in A$, $I_{\{A\}} = 0$ when $r(X) \notin A$.

The purpose of importance sampling is to sample the random variables from a different distribution than the original distribution of interest and use those samples to calculate an estimator $\hat{E}_n(r(X))$, with the goal of reducing the variance in the estimate. To do this, the Monte Carlo estimate must be weighted appropriate. If $h_X(x)$ is the new distribution, the new estimate is derived as follows:

$$E(r(x)) = \int r(x) \frac{f(x)}{h(x)} h(x) = \hat{E}_n \left[r(x) \frac{f(x)}{h(x)} \right] = \frac{1}{n} \sum_{i=1}^n \frac{r(x_i) f(x_i)}{h(x_i)}, x_i \sim h(X) \quad (4)$$

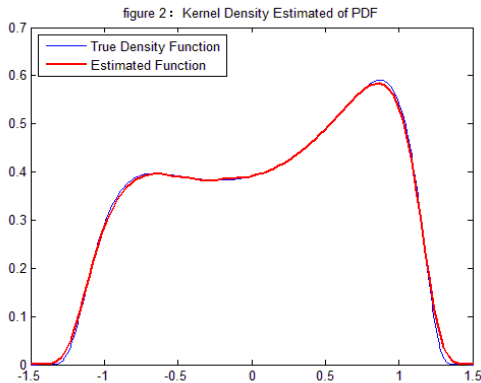
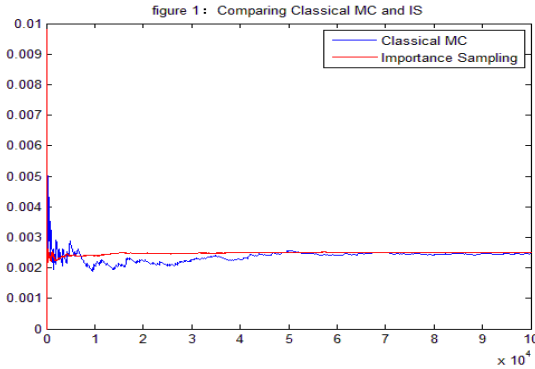
Where $h(x)$ is importance sampling density, $f(x)/h(x)$ is called the weight function. when $h(x) \propto r(x)f(x)$, the variance of the importance sampling estimator is minimized.

Much of the work in importance sampling has been finding minimum variance estimators for specific cases. The fundamental issue in implementing importance sampling is the choice of the new distribution which encourages the important regions of the input variables. If

you have a good distribution, the payoff is a much lower simulation cost. There are various measure used to calculate the goodness of the importance sampling scheme. One is the ratio of the variance obtained by Monte Carlo vs the variance of the importance sampling result $\sigma_{MC}^2 / \sigma_{IS}^2$; another measure is the ratio of the number of samples required by each scheme, given the same output variance N_{MC} / N_{IS} . In general, a good importance sampling function should be as follows:

1. $h(x) > 0$ whenever $r(x)f(x) \neq 0$.
2. $h(x)$ should be close to being proportional to $r(x)f(x)$.
3. It should be relatively easy to generate samples from $h(x)$ and also to calculate the density $h(x)$.

Some standard approaches to determining $h(x)$ including scaling, where the original random variable X is scaled by linearly shifted to put more probability density into a particular region. Another classical approach is to assume that $h(x)$ belongs to a parametric distribution family, for example, determining the mean and variance for $h(x)$ if it is assumed to be normal, often these parameters are obtained by optimizing the variance of importance sampling estimator \hat{E}_h .



III. IMPORTANCE SAMPLING DENSITY AND KERNEL DENSITY ESTIMATORS

Kernel Density Estimation (KDE) is a technique used to estimate the density of a random variable X given n independent samples $X_1, X_2, X_3, \dots, X_n$ of it [2], if one considers the discrete distribution obtained from sampled:

$$F_n^{\{discrete\}}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i < x) \quad (5)$$

The KDE can be viewed as a smoothed version of this estimator. let $K(\cdot)$ be a probability density function ($K(\cdot) > 0$, and $\int K(\cdot) dx = 1$) then KDE is

$$f_n^{\{KDE\}}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (6)$$

Where h is known as the bandwidth parameter, which controls the influence of each sample in providing a density estimate at a near-by point. Small h corresponds to a small region of influence; a large h to a large one. In the case of estimating a hitting probability, we only want points in the target set to contribute to the approximated density, so we use:

$$f_n^{\{KDE\}}(x; h) = \frac{1}{nh} \sum_{i=1}^n K(r(X_i) < c) K\left(\frac{x - X_i}{h}\right) \quad (7)$$

In this paper, we chose to use the optimal bandwidth obtained by minimizing the asymptotic mean-squared error of the importance sampling density. The optimal bandwidth was calculated according to the following formula [3]:

$$h_{opt} = \left\{ \frac{dR(K_d)}{\sigma_K^4} A J^{-1} \right\}^{1/(d+4)} n^{-1/(d+4)} \quad (8)$$

Where full kernel is product kernel of the marginal densities $K_d(x) = K(x_1)K(x_2)\dots K(x_d)$, σ_k^2 is the variance of the marginal kernel function $K(\cdot)$. $R(K_d)$ and A are given in equation 10.

$$\sigma_K^2 = \int x_1^2 K(x_1) dx_1$$

$$R(K_d) = \int K_d(x) dx; \quad (9)$$

$$A = \frac{1}{n \hat{f}_{init}} \sum_{i=1}^n I(x_{init}) f(x_{init})$$

In theoric, The kernel function K can be any probability density function, but, as we use this density in importance sampling, we need it to have support over the entire domain. Therefore, we pick the Gaussian kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (10)$$

this has the effect of making the density a mixture of Gaussian, which will lead to an efficient implementation in the Importance Sampling algorithm [4][5].

IV. SIMULATIONS

In this section, We will give two examples.

Example 1 comparing Classical MC and IS. We calculate: $\theta = \Pr ob\{X > 6\}$ $X \sim \exp(1)$. Red line indicates the results which was from Importance Sampling, blue line display results from Classical MC. Clearly, the Convergence rate of IS is faster than MC methods , it shows that an extra digit of accuracy on a result requires less replications.

Example 2: Kernel Density Estimated of PDE. Let $f(x) = \arctan(x)$. The blue line is the true density functions of $f(x)$, and the red line is an Estimated plot using by KED. .

V. CONCLUSION

This paper has proposed an importance sampling

approach based on kernel density estimators. Although many nonparametric approaches to importance sampling in the literature use some type of density estimation, we tried to develop a new approach ,which could be implemented for any type of simulation problem. we tested this approach on two tests problems and found that improtance sampling using this approach is reasonably robust.

- [1] A.Owen. Y.Zhou. Safe and Effective Importance Sampling ,Journal of American Statistical Association, Vol.95, N0.449, Theory and Methods ,March 2000.
- [2] Wand, M.P. and Jones, M.C Kernel Smoothing. Chapman and Hall, 1994
- [3] Zhang, P. "Nonparametric Importance Sampling." Journal of the American Statistical Association, Vol.91, no 435. pp 1245-1253. 1996
- [4] Chen, J-C, Lu, D Sadowsky, J. Yao K. 1993. On importance sampling in digital communication -Part I: Fundamentals. IEEE J. Selectd Areas Comm. II, 289-299.
- [5] Madras, N and Piccioni, M. 1999. Importance Sampling for Families of Distributions. The annals of Applied Probability, 9, 1202-1225