

Candidate set sampling: A note on theoretical guarantees

Shifeng Xiong

Academy of Mathematics and Systems Science

Chinese Academy of Sciences, Beijing 100190

xiong@amss.ac.cn

Abstract In this note we introduce a simple numerical sampling method, called candidate set sampling, which is based on an straightforward discretization to the density function. This method requires the knowledge of the density function (up to an unknown normalizing constant) only. Furthermore, candidate set sampling is non-iterative, dimension-free, and easy to implement, with fast convergence and low computational cost. We present its basic convergence properties in the note.

KEY WORDS: Bayesian computation; discretization; Monte Carlo; quasi-Monte Carlo.

1 Introduction

We consider the issue of sampling from a multivariate density function, which plays an important role in Bayesian computation (Ghosh, Delampady, and Samanta 2006), as well as other problems in computational statistics and data analysis, including high-dimensional integration and optimization problems (Fang and Wang 1993). Many numerical methods and algorithms have been developed in the literature, most notably the Markov chain Monte Carlo (MCMC) methods. However, in real applications the MCMC methods may have practical difficulties for complex distributions, as well as other technical problems such as convergence (e.g., O’Hagan, Murphy, and Gormley 2012).

To overcome difficulties encountered in many MCMC algorithms, non-iterative sampling methods have been proposed in the literature. They use some simplified distribution to

approximate the target distribution. Such simplification includes normal distribution via Laplace approximation (Bornkamp 2011) and variation inference (Blei, Kucukelbir, and McAuliffe 2017), mixture of uniform distributions via discretization (Fu and Wang 2002; Wang and Lee 2014), and mixture of normal distributions via Gaussian process interpolation (Joseph 2012), among many others.

In this note we propose another sampling method, called candidate set sampling, which is based on an straightforward discretization to the density function. This method requires the knowledge of the density function (up to an unknown normalizing constant) only. Furthermore, candidate set sampling is non-iterative, dimension-free, and easy to implement, with fast convergence and low computational cost. We present its basic convergence properties in the note.

2 Sampling from a candidate set

We consider the problem of generating random variables from a probability density function $f(\mathbf{x})$, whose support is denoted by $\mathcal{S} \subset \mathbb{R}^d$, up to an (unknown) constant. For unbounded $\mathcal{S}$, we can find a bounded set that contains the significant region of f . Without loss of generality, we set $\mathcal{S} = [0, 1]^d$. Let μ_f denote the probability measure on $[0, 1]^d$ corresponding to f .

For a large integer M , take a candidate set $\mathcal{A}_M = \{\mathbf{a}_i\}_{i=1}^M$, whose elements are uniformly scattered on $[0, 1]^d$. Define a probability measure $\mu_{(f, \mathcal{A}_M)}$ on $\mathcal{A}$ as

$$P(\mathbf{X}_{(f, \mathcal{A}_M)} = \mathbf{a}_i) = p_i = \frac{f(\mathbf{a}_i)}{\sum_{j=1}^M f(\mathbf{a}_j)}, \quad i = 1, \dots, M, \quad (1)$$

where the random vector $\mathbf{X}_{(f, \mathcal{A}_M)}$ follows from $\mu_{(f, \mathcal{A}_M)}$. We can prove that, if $\mathcal{A}_M$ possesses some dense properties, then $\mu_{(f, \mathcal{A}_M)}$ converges to μ_f as $M \rightarrow \infty$. For a sample size m , let $\mathbf{X}_1, \dots, \mathbf{X}_m$ be independently and identically distributed according to (1). The sample $\{\mathbf{X}_i\}_{i=1}^m$ is the desired output of the proposed candidate set sampling approach.

The candidate set $\mathcal{A}_M$ can be generated by randomly sampling from the uniform dis-

tribution on $[0, 1]^d$, quasi-Monte Carlo sequence (Niederreiter 1992), support points (Mak and Joseph 2018), or other space-filling designs (Santner, Williams, and Notz 2018). Here we adopt quasi-Monte Carlo sequences since they yield better approximations to numerical integrals than the Monte Carlo method and are easy to construct deterministically. Let $\mathcal{A}_M$ be a quasi-Monte Carlo sequence satisfying the error bound of integration approximation,

$$\left| \frac{g(\mathbf{a}_1) + \cdots + g(\mathbf{a}_M)}{M} - \int_{[0,1]^d} g(\mathbf{x}) d\mathbf{x} \right| \leq \delta_{M,d} V(g) \quad (2)$$

for any integrable g and sufficiently large M , where $\delta_{M,d}$ is a positive number dependent only on M and d , $\delta_{M,d} \rightarrow 0$ as $M \rightarrow \infty$ for fixed d , and $V(g)$ is the variation of g in the sense of Hardy and Krause; see for instance Niederreiter (1992). In this paper we assume $V(f) < \infty$. The number $\delta_{M,d}$ represents the convergence rate of a quasi-Monte Carlo sequence. For example, the Halton sequence gives $\delta_{M,d} = C_d(\log M)^d/M$, where C_d depends only on d (Dick, Kuo, and Sloan 2013).

Let $d_K(\mu_{(f, \mathcal{A}_M)}, \mu_f) = \sup_{\mathbf{x} \in [0,1]^d} |\mu_{(f, \mathcal{A}_M)}([\mathbf{0}, \mathbf{x}]) - \mu_f([\mathbf{0}, \mathbf{x}])|$ denote the Kolmogorov distance between $\mu_{(f, \mathcal{A}_M)}$ and μ_f , where $[\mathbf{0}, \mathbf{x}] = [0, x_1] \times \cdots \times [0, x_d]$ for $\mathbf{x} = (x_1, \dots, x_d)'$.

Theorem 1. *For sufficiently large M ,*

$$d_K(\mu_{(f, \mathcal{A}_M)}, \mu_f) \leq C \delta_{M,d}, \quad (3)$$

where C is a positive constant independent of M and d .

Proof. Denote $J = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \in (0, \infty)$. Let I represent the indicator function. For any $\mathbf{x} \in [0, 1]^d$, let $J_{\mathbf{x}} = \int_{[\mathbf{0}, \mathbf{x}]} f(\mathbf{x}) d\mathbf{x}$. We have

$$\begin{aligned} \mu_{(f, \mathcal{A}_M)}([\mathbf{0}, \mathbf{x}]) &= \sum_{i=1}^M p_i I(\mathbf{a}_i \in [\mathbf{0}, \mathbf{x}]) \\ &= \left(\frac{1}{M} \sum_{i=1}^M f(\mathbf{a}_i) \right)^{-1} \cdot \frac{1}{M} \sum_{i=1}^M f(\mathbf{a}_i) I(\mathbf{a}_i \in [\mathbf{0}, \mathbf{x}]) \\ &= (J + \Delta_1)^{-1} (J_{\mathbf{x}} + \Delta_2), \end{aligned}$$

where

$$\Delta_1 = \frac{1}{M} \sum_{i=1}^M f(\mathbf{a}_i) - J, \quad \Delta_2 = \frac{1}{M} \sum_{i=1}^M f(\mathbf{a}_i) I(\mathbf{a}_i \in [\mathbf{0}, \mathbf{x}]) - J_{\mathbf{x}}.$$

By (2), $|\Delta_1|, |\Delta_2| \leq \delta_{M,d} V(f)$. Therefore, for sufficiently large M ,

$$\begin{aligned} & \left| \mu_{(f, \mathcal{A}_M)}([\mathbf{0}, \mathbf{x}]) - \mu_f([\mathbf{0}, \mathbf{x}]) \right| = \left| (J + \Delta_1)^{-1} (J_{\mathbf{x}} + \Delta_2) - J^{-1} J_{\mathbf{x}} \right| \\ &= \frac{J|\Delta_1| + J_{\mathbf{x}}|\Delta_2|}{J|J + \Delta_1|} \leq \frac{J|\Delta_1| + J_{\mathbf{x}}|\Delta_2|}{J^2/2} \leq \frac{2V(f)}{J} \delta_{M,d}, \end{aligned}$$

which implies (3). □

Let $\mu_{(m, \mathcal{A}_M)}$ denote the empirical probability measure of the sample $\{\mathbf{X}_i\}_{i=1}^m$ from $\mu_{(f, \mathcal{A}_M)}$ in (1). The following theorem indicates that $\mu_{(m, \mathcal{A}_M)}$ is asymptotically equivalent to the empirical measure of the desired sample from μ_f .

Theorem 2. *If $\delta_{M,d} = O(1/\sqrt{m})$, then $d_K(\mu_{(m, \mathcal{A}_M)}, \mu_f) = O_p(1/\sqrt{m})$.*

Proof. It suffices to prove $d_K(\mu_{(m, \mathcal{A}_M)}, \mu_{(f, \mathcal{A}_M)}) = O_p(1/\sqrt{m})$. By DKW's inequality, for any $\epsilon, z > 0$ and m , there exists a positive constant $C_{\epsilon,d}$, not depending on M , such that $d_K(\mu_{(m, \mathcal{A}_M)}, \mu_{(f, \mathcal{A}_M)}) \leq C_{\epsilon,d} e^{-2(2-\epsilon)mz^2}$, which implies $E \left[\sqrt{m} d_K(\mu_{(m, \mathcal{A}_M)}, \mu_{(f, \mathcal{A}_M)}) \right]^2 = O(1)$ (Shao 1999). This completes the proof. □

3 Sampling in a deterministic way

Usually deterministic approximation, instead of random sampling, via some design strategy can improve the representativeness of $\mu_{(m, \mathcal{A}_M)}$. Here we begin with the minimax distance design (Johnson, Moore, and Ylvisaker 1990) on $[0, 1]$, i.e., $u_i = (2i-1)/(2m)$, $i = 1, \dots, m$, and generate $\{\mathbf{X}_i\}_{i=1}^m$ in the following deterministic manner. Let $q_0 = 0$ and $q_i = \sum_{j=1}^i p_j$ for $i = 1, \dots, M$. Clearly, $u_1 < \dots < u_m$ and $q_0 \leq q_1 \leq \dots \leq q_M$.

For $j = 1$, search from $\{1, \dots, M\}$ in the increasing order, and find the minimal number i_1 such that $u_1 \in [q_{i_1-1}, q_{i_1})$, let $\mathbf{X}_1 = \mathbf{a}_{i_1}$;

for $j = 2$, search from $\{i_1 + 1, \dots, M\}$ in the increasing order, and find the minimal number i_2 such that $u_2 \in [q_{i_2-1}, q_{i_2})$, let $\mathbf{X}_2 = \mathbf{a}_{i_2}$;

$\vdots$

for $j = m$, search from $\{i_{m-1} + 1, \dots, M\}$ in the increasing order, and find the minimal number i_m such that $u_m \in [q_{i_m-1}, q_{i_m})$, let $\mathbf{X}_m = \mathbf{a}_{i_m}$.

If $\{\mathbf{X}_i\}_{i=1}^m$ is generated from this deterministic process, then the empirical measure $\mu_{(m, \mathcal{A}_M)}$ is non-stochastic. We can prove its deterministic convergence properties in the follow-up work.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (grant number 12571276).

References

- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* 112, 859-877.
- Bornkamp, B. (2011), Approximating Probability Densities by Iterated Laplace Approximations, *Journal of Computational and Graphical Statistics*, 20, 656–669.
- Dick, J., Kuo, F. Y., and Sloan, I. H. (2013), High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, 22, 133–288.
- Fang, K. T. and Wang, Y. (1993), *Number-theoretic methods in statistics*. CRC Press.
- Fu, J. and Wang, L., 2002. A random-discretization based Monte Carlo sampling method and its applications. *Methodology and Computing in Applied Probability* 4, 5C25.
- Ghosh, J. K., Delampady, M., and Samanta, T. (2006), *An Introduction to Bayesian Analysis: Theory and Methods*, Springer. New York.
- Johnson, M., Moore, L., and Ylvisaker, D. (1990), Minimax and maximin distance design, *Journal of Statistical Planning and Inference*, 26, 131–148.
- Joseph, V. R. (2012): Bayesian Computation Using Design of Experiments-Based Interpolation Technique, *Technometrics*, 54:3, 209-225

- Mak, S. and Joseph, V. R. (2018), Support points, *The Annals of Statistics*, 46, 2562–2592.
- Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia.
- O’Hagan, A., Murphy, T.B., Gormley, I.C., (2012). Computational aspects of fitting mixture models via the expectation-maximization algorithm. *Computational Statistics and Data Analysis* 56, 3843–3864.
- Santner, T. J., Williams, B. J., and Notz, W. I. (2018). *The Design and Analysis of Computer Experiments*, 2nd Edition, Springer. New York.
- Shao, J. (1999). *Mathematical Statistics*, Springer. New York.
- Wang, L. and Lee, C. H. (2014), Discretization-based direct random sample generation, *Computational Statistics and Data Analysis*, 71, 1001–1010.