MONTE CARLO CONDITIONING ON A SUFFICIENT STATISTIC UNDER NONUNIQUENESS

N. B. Potadar

Department of Mathematics/Ramnarain Ruia College/Mumbai-19

Abstract: General formulae appropriate for Monte Carlo simulation of conditional expectations of functions of random variable given a sufficient statistic are known in literature. Sampling from conditional distributions depends heavily on the way auxiliary variable and parameter distributed jointly. We argue that this joint distribution of auxiliary variable and parameter gives rise to heuristic to get correct distribution of underlaying random variable.

Introduction: In statistics we have two different, but not exclusive, paradigms. One is about models which are parametric and other is nonparametric. In parametric regime of statistics models are parametrized and we can think of drawing samples from the population. In nonparametric regime we think of order statistics, methods like artificial neural networks and wavelets etc. Nonparametric methods, although, are not independent of parameters altogether they significantly avoid use of parameters. In other words statistical analysis of any model in one way or other depend on the form of underlaying distribution. Assuming some form of distribution means assuming dependence of distribution on a parameter. This parameter may be a scalar or a vector, usually not known in advance and we have to depend on estimator of parameter known as statistic.

In statistics objective of variety of simulations is estimation of statistic which frequently expressed as estimation of an expectation of the form $E(g(X))$ where X is a random variable. Suppose $f(x)$ is the joint density of X. Then

$$E\big(g(X)\big) = \int g(x)f(x)dx$$

Let $T(x)$ is a statistic based on a random sample x, we may be interested in computing quantities like mean and variance based on expectations, like above, depending on values of $T(x)$. It is one of the most important question in statistics is how to carry out evaluations of integrals involved in computations of expectations. Monte Carlo methods are precisely developed to perform these kinds of calculations of expectations. Monte Carlo methods essentially use drawing samples from known distributions and then calculating mean gives rise to sought expectations. These calculations are based on some known statistic which approximates corresponding population parameter. It is however a question the extent to which sample drawn resembles to original distribution. Lindqvist and Taraldsen have suggested some methods which makes drawn samples using statistic indeed correspond to original distribution.

Work of Lindqvist and Taraldsen: Consider a random variable X along with a sufficient statistic T. In the following we give one heuristic to get a random sample for X if given a sufficient statistic T. We adopt the notations and framework given in Lindqvist and Taraldsen. In their paper Lindquist and Taraldsen it is assumed that a random variable U (referred as auxiliary variable in abstract) is given with known distribution such that (X,T) for given population parameter θ can be simulated using U. In other words there exist functions χ and τ such that given θdistribution of $(\chi(U,\theta),\ \tau(U,\theta))$ equals the joint distribution of (X,T). Their approach is to first draw U, then determine parameter value $\hat{\theta}$ such that $\tau\big(U,\hat{\theta}\big) = t$ and then use $X_t = \chi(U,\hat{\theta})$ as the sought sample. Engen and Lillegard (1997) have shown that in general $X_t = \chi(U,\hat{\theta})$ may not have the correct distribution when $\hat{\theta}$ is not uniquely determined by t and u from the equation $\tau(u,\theta) = t$ even when X_t is uniquely determined. Their claim is X_t is distributed like X given t. Let $f(u)$ be a density of U. Let θ be distributed like Θ. Conditional density of $\tau(U,\theta)$ given $U = u$ is denoted by $W_t(u)$. Note that as U and Θ are independent, $W_t(u)$, as a function of t, is $\tau(u,\theta)$ for fixed u.

Then it follows that for an arbitrary continuous function $\phi(x)$,

$$E\{\phi(X)|T = t\} = \frac{E(Z_tW_t)}{E(W_{t)}}$$

Where

$$Z_t(u) = E(\phi\{\chi(u,\Theta)\}|\tau(u,\Theta) = t).$$

Basic advantage of these formulae is that quantities involved are functions of u and hence they can be simulated with a judicious choice of Θ. We will comment on choice of Θ afterwards. Thus, problem of simulating X using Monte Carlo method have a solution if $\tau(u,\theta) = t$ has a unique solution for fixed u. Lindqvist and Taraldsen have further treated the case in which $\tau(u,\theta)$ depends on u only through some function of u.

Heuristic in Case of Nonuniqueness: Most important question is that how to proceed if $\tau(u,\theta) = t$ doesn't have a unique solution. Suppose now that $\tau(u,\theta) = t$ has two solutions for fixed u denoted by θ_1 and θ_2. In order to handle such cases we propose that iterations of random samples can be used. Consider now first solution θ_1. Corresponding to this solution we have distribution of $\chi(u,\theta_1)$. This expression is evaluated for fixed u. Now we take iterates of χ. Denote random sample for u by u_0

$u_1 = \chi(u_0,\theta_1)$. $u_1 = \chi(u_0,\theta_1)$, …. $u_n = \chi(u_{n-1},\theta_1)$

This sequence of random variables asymptotically goes to distribution of X given a sufficient statistic T.

Conclusion: Lindqvist and Taraldsen have successfully treated the case when $\tau(u,\theta) = t$ for fixed u have unique solution. They also have considered some cases in which there is no unique solution to underlaying equation for θ in terms of t. We have proposed one heuristic which is mainly of theoretical concern. However with the help of modern powerful computing techniques validity of such an heuristic can be tested.

References:

1. G. Csella, R Berger (2002). Statistical Inference. Duxbury Adavanced Series.
2. P. G. Siddheshwar, B. N. Veena, Linear And Non-Linear Analyses Of Brinkman- Bènard Convection In A Nanoliquid-Saturated Enclosure; Mathematical Sciences International Research Journal : Issn 2278-8697 Volume 6 Issue 1 (2017), Pg 74-77
3. B. Lindqvist, G.Taraldsen (2005). Monte Carlo Conditioning on Sufficient Statistics. Biometrika. 92.2
4. N.Rajasekhar Reddy, Althaf Ali, Y.Suresh Babu, Enabling Smart Cloud Services To The Smart Cities Via Cars Architecture; Mathematical Sciences International Research Journal : Issn 2278-8697 Volume 6 Issue 1 (2017), Pg 78-85
5. Engen, S. Lillegard, M.(1997) Stochastic Simulations Conditioned on Sufficient Statistics. Biometrika 84, 235-40
6. N. Jansirani, V.Kamalnasir, M-Extended Ωω-Local Languages and Tiling Systems; Mathematical Sciences International Research Journal : Issn 2278-8697 Volume 6 Issue 1 (2017), Pg 86-89

***