NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

A lemma on conditional Monte Carlo

by

Bo Henry Lindqvist and Gunnar Taraldsen

PREPRINT
STATISTICS NO. 10/2001

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S10-2001.ps

A lemma on conditional Monte Carlo

Bo Henry Lindqvist, Norwegian University of Science and Technology
Gunnar Taraldsen, SINTEF Telecom and Informatics, Norway

9th October 2001

Abstract

In 1993 S.Engen and M.Lillegård published a most interesting approach for doing Monte Carlo simulations conditioned on a sufficient statistic. It turns out that one of their main results is incorrect due to the ignorance of a Borel paradox when conditioning on a zero set. We prove a Lemma which give new versions of the original claims. The formulas obtained are similar in form to the ones presented in the 1956 article by H.F. Trotter and J.W. Tukey who introduced the term "conditional Monte Carlo", but the methods are otherwise rather different.

1 Introduction

We say that a statistic T is sufficient for a parameter θ compared to a statistic X if to each bounded measurable f there exist a measurable g such that $\mathrm{E}^\theta(f(X) \,|\, T = t) = g(t)$ for all θ. This means that g is a version of the conditional expectation for each θ, and we use the notation $\mathrm{E}(f(X) \,|\, T = t) = g(t)$ to denote this particular version of the conditional expectation. This definition is similar to the general sufficiency concept of Blackwell [2], and includes the more standard concept where it is assumed in addition that T is a function of X [3, p.19,p.49] [9, p.84-].

If τ is a measurable function and m is a measure, then the distribution m_τ of τ is defined by $\mathrm{m}_\tau(A) := \mathrm{m}(\tau \in A)$. Here we use the notation $(\tau \in A) := \{x \,|\, \tau(x) \in A\}$, which we prefer compared to the more standard notation $\tau^{-1}(A)$. In the following we will freely use other standard concepts from measure theory [8], probability and statistics [9] without further comment.

Generalizing Engen and Lillegård [4], our basic assumption is that there is a Monte Carlo variable U with a known distribution M, and measurable functions χ, τ such that $(\chi(U,\theta), \tau(U,\theta))$ and (X,T) have equal distributions for each fixed θ. The idea is then that U can be simulated on a computer and the function (χ,τ) makes it possible to simulate observations of the statistic (X,T) for given values of θ. In particular we have $\mathrm{E}^\theta(f(X) \,|\, T = t) = \mathrm{M}(f(\chi^\theta) \,|\, \tau^\theta = t)$. We use the notation τ_u and τ^θ to denote the one-variable functions (sections) defined by $\tau_u(\theta) = \tau^\theta(u) = \tau(u,\theta)$.

The main point in this note is that the original problem of conditioning on τ^θ in effect may be replaced by the problem of conditioning on τ_u. This may be much simpler, and it is in particular so if the equation $\tau(u,\theta) = t$ has a unique solution. In general we have some additional freedom in the choice of the σ-finite distribution π of Θ on the parameter space, and we briefly discuss some natural alternatives.

2 The Lemma

Lemma 1 *Let* m *be the product of two* σ*-finite measures* M *and* π*. Let* τ *be a measurable function such that the distribution* m_τ *of* τ *is* σ*-finite, and such that* $\pi_{\tau_u} \ll \mathrm{m}_\tau$ *for* M *almost all* u*. If* ϕ *is* m*-integrable on sets with finite measure, then there exist measurable* ψ*,* w *and a* σ*-finite measure* μ *such that for almost all arguments*

$$\psi(t,u) = \pi(\phi_u \,|\, \tau_u = t), \quad \pi_{\tau_u}(dt) = w(t,u)\mu(dt)$$

Furthermore we have

$$\mathrm{m}(\phi \,|\, \tau = t) = \frac{\mathrm{M}\,\psi_t w_t}{\mathrm{M}\,w_t}$$

Proof. Using the defining equality $\stackrel{C}{=}$ of conditional expectation (which generalizes to the given assumptions) and the Fubini theorem we calculate

$$\begin{aligned}
\mathrm{m}(f(\tau)\phi) &= \int \pi(f(\tau_u)\phi_u)\ \mathrm{M}(du) \stackrel{C}{=} \int \pi[f(\tau_u)\pi(\phi_u \,|\, \tau_u)]\ \mathrm{M}(du) \\
&= \iint f(t)\pi(\phi_u \,|\, \tau_u = t)w(t,u)\,\mu(dt)\,\mathrm{M}(du) \\
&\stackrel{F}{=} \int f(t)[\int \pi(\phi_u \,|\, \tau_u = t)w(t,u)\ \mathrm{M}(du)]\,\mu(dt).
\end{aligned}$$

The case $\phi = 1$ gives $\mathrm{m}_\tau(dt) = (\mathrm{M}\,w_t)\mu(dt)$, and the above calculation then gives $\mathrm{m}(f(\tau)\phi) = \int f(t)\frac{\mathrm{M}\,\psi_t w_t}{\mathrm{M}\,w_t}\ \mathrm{m}_\tau(dt)$. The calculation holds for all f in the class of indicator functions of sets with finite m_τ-measure, and proves the final claim in the Lemma.

In the above we relied on existence and measurability of ψ and w in applying the Fubini theorem. In the following we put $\mu = \mathrm{m}_\tau$ and $\mu_u = \pi_{\tau_u}$. It follows that (for M-a.e. u) we have $\mu_u \ll \mu$. The Radon-Nikodym theorem gives a w with $\mu_u(dt) = w(u,t)\,\mu(dt)$. Existence of a (jointly) measurable w does unfortunately not follow, so we have to refine this argument.

Define $\nu(g) = \int \mu_u(g_u)\ \mathrm{M}(du)$. Then $\nu \ll \mathrm{M} \times \mu$, and the Radon-Nikodym theorem gives w with $\nu(du, dt) = w(u,t)\ \mathrm{M}(du)\mu(dt)$. If we consider the special case $g(u,t) = g_1(u)g_2(t)$, then it follows that $\int \mu_u(g_2)g_1(u)\ \mathrm{M}(du) = \int \mu(g_2 w_u)g_1(u)\ \mathrm{M}(du)$. We conclude $\mu_u(dt) = w(u,t)\mu(dt)$, and the claim for w and μ is proven.

Let $\nu_u(f) = \pi(\phi_u f(\tau_u))$, and observe that $\nu_u \ll \mu_u$ also gives the corresponding relation for the measures obtained after integration with respect to M. The Radon-Nikodym theorem gives a measurable ψ such that $\int \nu_u(g_u)\ \mathrm{M}(du) = \int \mu_u(g_u\psi_u)\ \mathrm{M}(du)$. As above this allows us to conclude $\mu_u(g_2\psi_u) = \nu_u(g_2) = \pi(\phi_u g_2(\tau_u))$, and the claim $\pi(\psi_u \,|\, \tau_u = t) = \psi(u,t)$ follows. □

The previous proof shows that if we normalize such that $\mathrm{M}\,w_t = 1$, which is always possible, then $\mu = \mathrm{m}_\tau$ and ψ and w are unique. In the remainder of this note we indicate how the previous Lemma can be used to obtain practical Monte Carlo methods for calculations in a statistical setting. A more thorough discussion of this with examples is given in a companion paper [6].

3 Monte Carlo conditioning

If we change the roles of u and θ in the Lemma, and in particular assume that for almost all θ the distribution of τ^θ is dominated by the distribution of τ, then sufficiency gives $\mathrm{m}(\phi(\chi) \,|\, \tau = t) = \mathrm{E}(\phi(X) \,|\, T = t)$. Combining this with the Lemma as stated we get

$$\mathrm{E}(\phi(X) \,|\, T = t) = \mathrm{m}(\phi(\chi) \,|\, \tau = t) = \mathrm{M}\,w_t\psi_t = \sum_i w(t,u_i)\psi(t,u_i),$$

where $\psi(t,u) = \pi(\phi(\chi_u) \,|\, \tau_u = t)$ and $\pi_{\tau_u}(dt) = w(t,u)\,\mathrm{m}_\tau$. The last equality follows from the law of large numbers when $u_1, u_2, \ldots$ are drawn from the distribution of U, and approximation by finite sums gives a Monte Carlo method for the calculation of the conditional expectation. Standard theory on Monte Carlo simulation [7, 10] contains numerous methods for improving on this naive approach of estimating the integral $\mathrm{M}\, w_t \psi_t$, or more generally $(\mathrm{M}\, w_t \psi_t)/\,\mathrm{M}\, w_t$, but we will not pursue this here.

Of course the most straightforward way of computing the conditional distribution of X given $T = t$ would be to fix a parameter value θ_0 and then compute or simulate the conditional distribution under θ_0. In this way the problem is a traditional one. However, such an approach does not make full use of the fact that *any* value for θ_0 will give the same conditional distribution. Thus it is conceivable that the possibility of varying the value of θ in an appropriate manner may be more efficient. The above gives an approach in this direction, where the clue is to equip the parameter space with a σ-finite measure π. If we choose π to be concentrated on $\{\theta_0\}$, [1] then we get nothing new. This motivates the opposite choice, namely to let π be Lebesgue measure in the Euclidean case, or more generally to let π be Haar measure. Another possibility, which also takes the distribution of X into consideration, is to let π be a non informative prior distribution for Θ, or vague prior distributions from Bayesian theory, for example Jeffreys' priors [5]. This is why we found it necessary and natural to introduce conditional expectation also in the case of σ-finite measures. In the remainder of this note we consider cases for which $\mathrm{M}\, w_t \psi_t /\, \mathrm{M}\, w_t$ simplifies further.

Assume that there exist a measurable $\tilde{\chi}$ such that $\tilde{\chi}(\tau(u,\theta), u) = \chi(u,\theta)$. The substitution principle [1] gives $\psi_t(u) = \pi(\phi(\chi_u) \,|\, \tau_u = t) = \phi(\tilde{\chi}(t,u))$. This holds for general ϕ, and we conclude that $\tilde{\chi}(t,V)$ is distributed like X given $T = t$ if V is drawn from the distribution of U with weight w_t. If the equation $\tau(u,\theta) = t$ has a unique solution $\theta = \hat{\theta}(u,t)$, then we may choose $\tilde{\chi}(t,u) = \chi(u, \hat{\theta}(u,t))$. It is claimed by Engen and Lillegård that in this case $\tilde{\chi}_t \sim (X \,|\, T = t)$, but this may be proven wrong by a concrete example. The conclusion is that without further assumptions we can not ignore the weight w_t.

Assume that $(\tau_u = t)$ is countable as a generalization of the unique solution case. Assume furthermore that π is absolutely continuous with respect to Lebesgue measure with density f. Given sufficient smoothness of τ_u it follows that τ_u has density

$$w(t,u) = \sum_{\theta \in (\tau_u = t)} \frac{f(\theta)}{|\partial_\theta \tau(u,\theta)|}$$

with respect to Lebesgue measure. Furthermore the conditional distribution of θ given $\tau_u = t$ is concentrated on $(\tau_u = t)$, with probability $[f(\theta)/\,|\partial_\theta \tau(u,\theta)|]/w(t,u)$ at θ, so we find

$$w(t,u)\psi(t,u) = \sum_{\theta \in (\tau_u = t)} \phi(\chi(u,\theta)) \frac{f(\theta)}{|\partial_\theta \tau(u,\theta)|}.$$

In cases where $(\tau_u = t)$ is not countable, for instance equal to an open set or an interval, we will typically be able to choose π so that $\pi(\tau_u = t) \neq 0$, and the result is again explicit formulas for ψ and w. Mixed cases are also possible corresponding to a decomposition of π_{τ_u} in an atomic and a non atomic part.

Finally we reconsider the unique solution case and will give conditions for when $\tilde{\chi}_t$ is distributed like $X \,|\, T = t$. The derivation will not depend on the Lemma. Consider the following calculation

$$\mathrm{E}^\theta(\phi(X) \,|\, T = t) \overset{M}{=} \mathrm{M}(\phi(\chi^\theta) \,|\, \tau^\theta = t) \overset{B}{=} \mathrm{M}(\phi(\chi^\theta) \,|\, \hat{\theta}^t = \theta) \overset{S_1}{=} \mathrm{M}(\phi(\tilde{\chi}_t) \,|\, \hat{\theta}^t = \theta) \overset{S_2}{=} \mathrm{M}(\phi(\tilde{\chi}_t))$$

[1] Observe however that this special case gives examples where $\pi_{\tau_u} \ll \mathrm{m}_\tau$ is not true, so this assumption is needed in the Lemma!

The equalities $\stackrel{M}{=}$, $\stackrel{S_1}{=}$, and $\stackrel{S_2}{=}$ follows respectively from the Monte Carlo assumptions, the substitution principle, and sufficiency. We have $(\tau^\theta = t) = (\hat{\theta}^t = \theta)$, but the equality $\stackrel{B}{=}$ is not valid in general due to the possibillity of a Borel paradox when $\mathrm{M}(\tau^\theta = t) = 0$. The following pivotal assumptions are however sufficient: (i) Assume that there exist functions $\tilde{\tau}$ and r such that $\tau(u, \theta) = \tilde{\tau}(r(u), \theta)$. (ii) The equation $\tilde{\tau}(v, \theta) = t$ has a unique solution $v = \hat{v}(\theta, t)$. We note that $\hat{v}(\theta, T)$ is a pivotal quantity since it is distributed like $\hat{v}(\theta, \tau(U, \theta)) = r(U)$, so the assumptions give a method for the construction of pivotal quantities. We will now sketch the proof of the equality $\stackrel{B}{=}$ given the added pivotal assumption. The point is to verify that conditioning on $\tau^\theta = t$ can be replaced by conditioning on $\hat{\theta}^t = \theta$. The uniqueness condition on θ ensures that $\tilde{\tau}(v, \theta) = t$ has a unique solution $\tilde{\theta}(v, t)$. In general, conditioning on $Z = z$ can be replaced by conditioning on $f(Z) = f(z)$ if f is one-to-one and measurable. Because of this $t = \tau(U, \theta) = \tilde{\tau}(r(U), \theta)$ can be replaced by $\hat{v}(\theta, t) = r(U)$, and this can be replaced by $\theta = \tilde{\theta}(r(U), t) = \hat{\theta}(U, t)$ which was to be proven. In this pivotal case we have ignored some measurability problems. A full proof for this case is found in a companion paper [6] where it is assumed in addition that the distributions of τ^θ and $\hat{\theta}^t$ are dominated by two respective measures for all θ and t respectively.

References

[1] R.R. Bahadur and P.J. Bickel. Substitution in conditional expectation. *Ann.Math.Statist.*, 39(2):377–378, 1968.

[2] D. Blackwell. Equivalent comparisons of experiments. *Ann.Math.Statist.*, 24:265–272, 1953.

[3] E.L.Lehmann. *Testing statistical hypotheses.* Springer, (second edition) 1997.

[4] S. Engen and M. Lillegård. Stochastic simulations conditioned on sufficient statistics. *Biometrika*, 84(1):235–240, 1997.

[5] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proc.Roy.Soc. A*, 186:453–461, 1946.

[6] B.H. Lindqvist and G. Taraldsen. Monte Carlo conditioning on a sufficient statistic. *manuscript*, 2000.

[7] B. Ripley. *Stochastic simulation.* Wiley, 1987.

[8] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, USA, third edition, 1987.

[9] M.J. Schervish. *Theory of Statistics.* Springer, 1995.

[10] H.F. Trotter and J.W. Tukey. Conditional Monte Carlo for normal samples. *Symposium on Monte Carlo Methods*, pages 64–79, 1956.