Monte Carlo Conditioning on a Sufficient Statistic

Bo H. Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Gunnar Taraldsen
Department of Physiology and Biomedical Engineering
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

February 21, 2000

1 Introduction

Let X and T be a random vectors with distribution indexed by a parameter $\theta \in \Omega_\Theta$. Suppose further that T is sufficient for θ compared to X. By definition this means that the conditional distribution of X given T does not depend on θ. This property is basic in statistical inference, with applications for example in parameter elimination, construction of optimal estimators and in testing (Lehmann 1997, Lehmann and Casella 1998). In practical cases, however, such conditional distributions may be extremely difficult to derive exactly, so there is a strong need for efficient approximations or Monte Carlo algorithms.

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 key feature of this paper is to present an approach in this direction, where the clue is to equip the parameter space Ω_Θ with some suitable σ-finite measure. This idea is from Engen and Lillegård [**?**], whose main result is essentially our Proposition 1.

The basic assumption is that there is a random vector U taking values in a space Ω_U with a known distribution, and 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 (τ, χ) makes it possible to simulate observations of (T, X) for given values of θ.

Now suppose it is given that $T = t$, and that a sample x_t from the conditional distribution of X given $T = t$ is wanted. Since this distribution does not depend on θ, it is reasonable to believe that it can be described in some simple way in terms of the distribution of U, and thus enabling Monte Carlo simulation based on U.

As already mentioned, the basic feature of our general approach is to introduce a σ-finite measure P_Θ on Ω_Θ. Assuming for a moment that this is a probability measure of a random variable Θ, we assume in addition that U, Θ are stochastically independent. The key result is then Theorem 1 which in the setting of probability measures states that the conditional distribution of X given $T = t$ equals the conditional distribution of $\chi(U, \Theta)$ given $\tau(U, \Theta) = t$. This means that samples x_t from the condtional distribution of X given $T = t$ can always be found by sampling (u, θ) from the conditional distribution of (U, Θ) given $\tau(U, \Theta) = t$ and then computing $x_t = \chi(u, \theta)$. The latter conditional distribution will therefore be of main importance in the paper.

2 Assumptions and Notation

As noted in the introduction, we shall find it convenient to supply the parameter space Ω_Θ by a σ-finite measure M_Θ. This leads to the need for modifying the definition of conditional expectation to allow for σ-finite distributions. Thus let μ be a σ-finite measure on some measurable space and let Y and Z be measurable functions defined on this space. We shall define what to mean by the conditional expectation $\mu(Y|Z=z)$ when the distribution μ_Z of Z is σ-finite. Since we will include the case when Y is constant, we do not require Y to be integrable, but require it to be bounded. The *conditional expectation* $\mu(Y|Z=z)$ is defined by requiring

$$\mu(Y\phi(Z)) = \int \mu(Y|Z=z)\phi(z)\mu_Z(dz)$$

to hold for all μ_Z-integrable ϕ. Here we use the common notation μX to mean the integral of X with respect to μ. Let $[Z \in A]$ denote the indicator function of the set $(Z \in A) = \{w|Z(w) \in A\}$. The assumption of σ-finiteness of μ_Z ensures that $A \mapsto \mu(Y[Z \in A])$ is a σ-finite measure. This measure is absolutely continuous with respect to μ_Z, and the Radon-Nikodym Theorem [?, p.121, p.123] gives the conditional expectation as its density: $\mu(Y[Z \in dz]) = \mu(Y|Z=z)\mu_Z(dz)$. This conditional expectation is normalized and has all the other properties of a conventional expectation. If μ is a probability measure it corresponds to the ordinary definition of conditional expectation. In most cases of interest, e.g. when the involved spaces are Borel spaces [?, p.618], the conditional expectation is given by a conditional distribution. We will in the sequel use equality sign between conditional expectations to mean that they are both versions of the conditional expectations and thus equal only a.e.

The general setup of the paper is as follows. Let Ω be a measurable space equipped with a family of probability measures P^θ indexed by a parameter θ in Ω_Θ. Let further Ω_T and Ω_X be measurable spaces and let there be given statistics T and X which are measurable functions $T: \Omega \to \Omega_T$ and $X: \Omega \to \Omega_X$.

We shall assume that T is sufficient for θ compared to X, which means that the conditional expectation $\mathrm{E}^\theta(\phi(X)|T=t)$ can be specified independent of θ. We do in particular not demand T to be a function of X when we define sufficiency of T with respect to X. This is in accordance with the general sufficiency concept of Blackwell (1953), which includes the usual concept where it is assumed that $T = T(X)$.

Let M_U be a probability measure on a measurable space Ω_U, which we shall call the Monte Carlo space. The basic assumption is that there are measurable functions τ and χ such that the distribution of $(\tau(U,\theta), \chi(U,\theta))$ equals the distribution of (T, X) under P^θ for each fixed θ. Moreover, let there be given a σ-finite measure M_Θ on the parameter space Ω_Θ, such that $\tau(U,\Theta)$ has a σ-finite distribution M_τ induced by the product measure $M \equiv M_U \otimes M_\Theta$. Functions of U will be called Monte Carlo variables.

3 General results

The following result, which is essentially the main result of Engen and Lillegård (1997), will be generalized in the subsequent theorem. We state it here since it is in a particularly useful form for applications in discrete models, and since it gives much of the flavour of the general method. The proof is elementary and can be compared with the proof given in [?].

Proposition 1 *Let $\phi: \Omega_X \to R$ be bounded and measurable, and define the Monte Carlo variables Z_t and Y_t by*

$$Z_t = \int \phi(\chi(U,\theta)) \cdot [\tau(U,\theta) = t]\, M_\Theta(d\theta), \quad Y_t = \int [\tau(U,\theta) = t]\, M_\Theta(d\theta).$$

If $M_U Y_t > 0$, then

$$E^\theta(\phi(X)|T=t) = \frac{M_U Z_t}{M_U Y_t}.$$

Proof: First we observe that $M_U Y_t$ is finite since M_τ is σ-finite. The composed function $\phi \circ \chi$ is not more general than χ alone, so in the proof we may assume that $\phi(x) = x$. By considering only θ's such that $¶^\theta(T=t) \neq 0$ it follows that

$$\mathrm{E}^\theta(X|T=t) = \frac{\mathrm{E}^\theta(X[T=t])}{¶^\theta(T=t)} = \frac{\int \mathrm{E}^\theta(X[T=t])\, M_\Theta(d\theta)}{\int ¶^\theta(T=t)\, M_\Theta(d\theta)} = \frac{M(\chi(U,\Theta)[\tau(U,\Theta)=t])}{M(\tau(U,\Theta)=t)}.$$

The above integration is valid since the sufficiency gives that the fraction is independent of θ. The conclusion follows from the Fubini theorem which allows us to change the order of integration since M_U and M_Θ are σ-finite.

The above proof shows in particular that the conditional distribution of X given $T = t$ equals the conditional distribution of $\chi(U, \Theta)$ given $\tau(U, \Theta) = t$. The next theorem shows that this fact holds in general, and is in fact basic to our approach.

Theorem 1 *If $\phi : \Omega_X \to R$ is bounded and measurable, then*

$$E^\theta(\phi(X)|T = t) = M(\phi(\chi(U, \Theta))|\tau(U, \Theta) = t). \tag{1}$$

Proof: As in the proof of Proposition 1 we may ignore the function ϕ. Let $\psi_1(t) = \mathrm{E}^\theta(X|T = t)$ and $\psi_2(t) = M(\chi(U, \Theta)|\tau(U, \Theta) = t)$. We will prove $M(\psi_1(\tau(U, \Theta)) \neq \psi_2(\tau(U, \Theta))) = 0$. Let ξ be a M_τ-integrable function. The claim follows from

$$\begin{aligned} M\xi(\tau(U, \Theta))\psi_2(\tau(U, \Theta)) &= M\xi(\tau(U, \Theta))\chi(U, \Theta) = M[M_U\xi(\tau(U, \Theta))\chi(U, \Theta)] \\ &= M[M_U\xi(\tau(U, \Theta))\psi_1(\tau(U, \Theta))] = M\xi(\tau(U, \Theta))\psi_1(\tau(U, \Theta)). \end{aligned}$$

The next theorem provides a formula for computing the conditional expectation on the right hand side of (1). A specialization to the Euclidean case is given in a subsequent corollary.

As it turns out, our approach depends heavily on solving the equation $\tau(u, \theta) = t$ with respect to θ. Introduce therefore the notation $\Gamma(u, t) = \{\theta : \tau(u, \theta) = t\}$. As in Proposition 1 it is seen that the result can be expressed as a quotient of expected values of Monte Carlo variables.

Theorem 2 *Suppose the conditional distribution of Θ given $\tau(u, \Theta) = t$ for each fixed $u \in \Omega_U$ is given as a regular conditional probability distribution $\rho(d\theta|u, t)$, and which is such that integrals with respect to $\rho(\cdot|u, t)$ are jointly measurable in u, t. Let μ be a σ-finite measure on Ω_T such that $\tau(u, \Theta)$ for fixed u has density $Y_t(u)$ w.r.t. μ. Assume moreover that $Y_t(u)$ is jointly measurable in u, t. Then*

$$E^\theta(\phi(X)|T = t) = \frac{\int\int \phi(\chi(u, \theta))\rho(d\theta|u, t)Y_t(u)M_U(du)}{\int Y_t(u)M_U(du)} \equiv \frac{M_U Z_t}{M_U Y_t}$$

where $Z_t = Y_t \int \phi(\chi(U, \theta))\rho(d\theta|U, t)$

Proof: Let $\nu_u(d\theta, dt)$ be the measure on $\Gamma(u, \cdot) \equiv \{(\theta, t) : \tau(u, \theta) = t\}$ defined by $\int\int h(\theta, t)\nu_u(d\theta, dt) = \int\int h(\theta, t)\rho(d\theta|u, t)\mu(dt)$. Then the joint distribution of $(\Theta, \tau(u, \Theta))$ has the density $Y_t(u)$ with respect to the measure ν_u. Now for any measurable $h : \Omega_U \times \Omega_\Theta \times \Omega_T \to R$ such that $(u, \theta) \mapsto h(u, \theta, \tau(u, \theta))$ is M-integrable, we have by Fubini's theorem,

$$\begin{aligned} &M(h(U, \Theta, \tau(U, \Theta)) \\ &\quad = \int\int h(u, \theta, \tau(u, \theta))M_\Theta(d\theta)M_U(du) \\ &\quad = \int\int h(u, \theta, t)Y_t(u)d\nu_u M_U(du) \\ &\quad = \int\int\int h(u, \theta, t)Y_t(u)\rho(d\theta|u, t)M_U(du)\mu(dt) \end{aligned} \tag{2}$$

In particular, if $h(u, \theta, t)$ is taken as a function of t alone, it follows that the marginal distribution of $\tau(U, \Theta)$ w.r.t. μ is

$$M_\tau(dt) = \int Y_t(u)M_U(du)\mu(dt) \tag{3}$$

To prove the theorem we need by the definition of conditional expectation to show that for any M_τ-integrable function ψ we have

$$M(\phi(\chi(U, \Theta))\psi(\tau(U, \Theta))) = \int \frac{\int\int \phi(\chi(u, \theta))Y_t(u)\rho(d\theta|u, t)M_U(du)}{\int Y_t(u)M_U(du)}\psi(t)M_\tau(dt)$$

By (3) the right hand side equals $\int\int\int \phi(\chi(u,\theta))Y_t(u)\rho(d\theta|u,t)\psi(t)M_U(du)\mu(dt)$ and the result therefore follows directly from (2).

The following corollary specializes to the case where Ω_Θ and Ω_T are subsets of the same Euclidean space R^k. Let $\det \partial_\theta \tau(u,\theta)$ be the determinant of the matrix of partial derivatives of $\tau(u,\theta)$ for fixed u.

Corollary 1 *Let Ω_Θ and Ω_T be open subsets of R^k with positive Lebesgue-measure, and assume $M_\Theta(\theta) = f(\theta)d\theta$ where $d\theta$ means Lebesgue measure on Ω_Θ. Let furthermore $\theta \mapsto \tau(u,\theta)$ be differentiable for each u, and assume that $\Gamma(u,t)$ is a finite set for each fixed u,t.*

Then

$$E^\theta(\phi(X)|T=t) = \frac{\int \sum_{\tilde\theta\in\Gamma(u,t)} \phi(\chi(u,\tilde\theta))|\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\tilde\theta} M_U(du)}{\int \sum_{\tilde\theta\in\Gamma(u,t)} |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\tilde\theta} M_U(du)}$$

Proof: The result follows from the preceding theorem by noting that the density of $\tau(u,\Theta)$ with respect to Lebesgue measure is

$$\sum_{\tilde\theta\in\Gamma(u,t)} |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\tilde\theta}$$

while the conditional distribution of Θ given $\tau(u,\Theta) = t$ is concentrated on $\Gamma(u,t)$ and given by

$$\frac{|\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\theta^*}}{\sum_{\tilde\theta\in\Gamma(u,t)} |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\tilde\theta}};\ \theta^* \in \Gamma(u,t)$$

Remark 1: As remarked earlier, conditional expectations given $T = t$ are to be regarded as measuarable functions of t, with equality between two expressions meaning equality a.e. In applications the measurable sets in Ω_T are typically Borel sets of some topology and often there exist versions of the conditional expectations which are continuous functions of t. In these cases equality a.e. will imply equality everywhere.

Remark 2: The essential feature of Theorem 2 is to evaluate the right hand side of (1) by double expectation by conditioning on U. Formally,

$$M(\phi(\chi(U,\Theta))|\tau(U,\Theta) = t) = E_{U|\tau} M(\phi(\chi(U,\Theta))|U, \tau(U,\Theta) = t) \tag{4}$$

where the outer expectation is with respect to the conditional distribution of U given $\tau(U,\Theta) = t$. With notation as in Theorem 2, this conditional distribution is given as

$$\frac{Y_t(u)M_U(du)}{\int Y_t(u)M_U(du)}$$

while the inner expectation in (4) is $\int \phi(\chi(u,\theta))\rho(d\theta|u,t)$.

Of course we may well change the roles of Θ and U in (4), thus conditioning on $(\Theta, \tau(U,\Theta))$ instead of $(U, \tau(U,\Theta))$. We would then recognize the inner expectation as simply the conditional expectation of $\phi(X)$ given $T = t$. Since this by sufficiency does not depend on θ, the outer expectation becomes trivial and we are back to the left hand side of (1).

This indicates that the essential feature of Theorem 2 is that computations of conditional expectations given $\tau(U,\theta) = t$ for fixed θ are changed into the (often much simpler) problem of computing conditional expectations given $\tau(u,\Theta) = t$ for fixed u. It may in fact be the case that the equation $\tau(u,\theta) = t$ has a single solution for θ (see next section).

Remark 3: The following derivation for the case where both U and Θ are discretely distributed, illustrates the main idea of the theorem. Let for simplicity ϕ be the identity function.

$$
\begin{aligned}
E(X|T=t) &= E(\chi(U,\Theta)|\tau(U,\Theta)=t)\\
&\sum_u\sum_\theta \chi(u,\theta)P(U=u,\Theta=\theta|\tau(U,\Theta)=t)\\
&= \sum_u\sum_\theta \chi(u,\theta)\frac{P(U=u,\Theta=\theta,\tau(U,\Theta)=t)}{P(\tau(U,\Theta)=t)}\\
&= \frac{\sum_u\sum_\theta \chi(u,\theta)P(\Theta=\theta,\tau(U,\Theta)=t|U=u)P(U=u)}{\sum_u P(\tau(U,\Theta)=t|U=u)P(U=u)}\\
&= \frac{\sum_u\sum_\theta \chi(u,\theta)P(\Theta=\theta,\tau(u,\Theta)=t)P(U=u)}{\sum_u P(\tau(u,\Theta)=t)P(U=u)}\\
&= \frac{\sum_u\{\sum_\theta \chi(u,\theta)P(\Theta=\theta|\tau(u,\Theta)=t)\}P(\tau(u,\Theta)=t)P(U=u)}{\sum_u P(\tau(u,\Theta)=t)P(U=u)}
\end{aligned}
$$

where the last formula corresponds to the expression in Theorem 2. Alternative expressions can now be found which correspond to the discussion in the previous remark:

$$
\begin{aligned}
&= \sum_u\{\sum_\theta \chi(u,\theta)P(\Theta=\theta|\tau(u,\Theta)=t)\}P(U=u|\tau(U,\Theta)=t)\\
&= \sum_u E\{\chi(u,\Theta)|\tau(u,\Theta)=t\}P(U=u|\tau(U,\Theta)=t)
\end{aligned}
$$

Remark 4: Proposition 1 follows directly from Theorem 2 by noting that $\rho(d\theta|u,t)$ is now the measure on Ω_Θ with density at θ given by

$$\frac{[\tau(u,\theta)=t]M_\Theta(d\theta)}{Y_t(u)}$$

4 Unique solution for θ

Assumption 1 (Unique θ.) *Assume that there is for each fixed pair (u,t) a unique solution for θ of the equation $\tau(u,\theta)=t$.*

The assumption implies that each $\Gamma(u,t)$ has a single element, which we shall denote by $\hat\theta(u,t)$. It follows that the conditional distribution $\rho(d\theta|u,t)$ is concentrated at $\hat\theta(u,t)$. The Monte Carlo variable $X_t=\chi(U,\hat\theta(U,t))$ will play an important role in the following.

As a consequence of the results of the previous section we get the following result valid under Assumption 1:

Theorem 3 *Make Assumption 1. If $\phi:\Omega_X\to R$ is bounded and measurable, then*

$$E^\theta(\phi(X)|T=t) = M(\phi(X_t)|\tau(U,\Theta)=t). \tag{5}$$

Let μ be a σ-finite measure on Ω_T such that $\tau(u,\Theta)$ for fixed u has density $Y_t(u)$ w.r.t. μ. Assume moreover that $Y_t(u)$ is jointly measurable in u,t. Then

$$E^\theta(\phi(X)|T=t) = \frac{M_U\phi(X_t)Y_t}{M_U Y_t} \tag{6}$$

Suppose that Ω_Θ and Ω_T are open subsets of R^k with positive Lebesgue-measure, and that $M_\Theta(\theta)=f(\theta)d\theta$ where $d\theta$ means Lebesgue measure on Ω_Θ. If $\theta\mapsto\tau(u,\theta)$ is differentiable for each u, then

$$Y_t(u) = |\frac{f(\theta)}{\det\partial_\theta\tau(u,\theta)}|_{\theta=\hat\theta(u,t)} \tag{7}$$

Recall that the simple algorithm given in the introduction, if valid, would imply that X_t under Assumption 1 has the same distribution as X given $T = t$. As already noted in the introduction, this is not the case in general, however. If it is the case, then we shall say for short that "X_t has the correct distribution".

By (5) in Theorem 2 it follows that X_t has the correct distribution if and only if X_t and $\tau(U, \Theta)$ are independent (extending the concept of independence in an obvious way to the case of σ-finite distributions). The formula (6) in Theorem 2 implies, on the other hand, that X_t has the correct distribution whenever X_t is stochastically independent of Y_t. This happens in particular if we can find a measure M_Θ such that $Y_t(u)$ is constant in u. Under the assumptions leading to (7) this corresponds to finding a function $f(\theta)$ such that $Y_t(u)$ becomes constant in u. This is the case provided

$$|\det \partial_\theta \tau(u, \theta)|_{\theta=\hat{\theta}(u,t)} = J(\hat{\theta}(u, t), t) \tag{8}$$

for some function J. Suppose namely that we are interested in the conditional distribution of X given $T = t_0$ for some specific value t_0. Then we may choose

$$f(\theta) = J(\theta, t_0) \tag{9}$$

which leads to a function $Y_t(u)$ satisfying $Y_{t_0}(u) = 1$. Then under continuity conditions (see Remark 1 of the previous section) it follows from (6) that X_t has the correct distribution for all t.

Moreover, if we under condition (10) choose instead $f(\theta) = \psi(\theta)(J(\theta, t_0))$ which gives $Y_{t_0}(u) = \psi(\hat{\theta}(u, t_0))$. Then (6) implies

$$M_U(\phi(X_{t_0})\psi(\hat{\theta}(U, t_0))) = M_U(\phi(X_{t_0}))M_U(\psi(\hat{\theta}(U, t_0)))$$

for all bounded and measurable ϕ and ψ, which finally implies that X_t and $\hat{\theta}(U, t)$ are independent for all t when (10) holds.

Note that the continuity conditions can be avoided under the condition

$$|\det \partial_\theta \tau(u, \theta)|_{\theta=\hat{\theta}(u,t)} = a(t)J(\hat{\theta}(u, t)) \tag{10}$$

in which case we can choose $f(\theta) = J(\theta)$ and get $Y_t(u) = a(t)$. Now (6) directly gives $\mathrm{E}^\theta(\phi(X)|T = t) = M_U\phi(X_t)$.

We close the section by showing that X_t has the correct distribution when a certain pivotal property holds.

Assumption 2 (The pivotal case.) *Assume that there exist measurable functions $r : \Omega_U \to \Omega_V$, $\tilde{\tau} : \Omega_V \times \Omega_\Theta \to \Omega_T$ with $\tau(u, \theta) = \tilde{\tau}(r(u), \theta)$ such that the equation $\tilde{\tau}(v, \theta) = t$ has a unique solution $v = \tilde{v}(\theta, t)$ for all (θ, t).*

In words this means that $\tau(u, \theta)$ depends on u only through $r(u)$, and that for a given θ this $r(u)$ is uniquely determined by t. Moreover, if Assumption 2 holds, then $\tilde{v}(\theta, T)$ is a pivotal quantity in the classical meaning. To see this, note the identity $r(U) = \tilde{v}(\theta, \tau(U, \theta))$, which has a distribution which is not a function of θ. The theorem thus follows since $t = \tilde{\tau}(v, \theta)$ is the unique solution of $v = \tilde{v}(\theta, t)$.

Theorem 4 *Make Assumptions 1 and 2. Then X_t has the correct distribution. Furthermore X_t is independent of $r(U)$ and hence of $\hat{\theta}(U, t)$ and each $\tau(U, \theta)$.*

Proof: We will use the general rule $E(X|Y = y) = E(X|h(Y) = h(y))$, valid whenever h is one-one and measurable, repeatedly below. The idea is to replace the condition $\tau(U, \theta) = t$ by the condition $\hat{\theta}(U, t) = \theta$ and thereby prove that the equality $\overset{?}{=}$ in the introduction actually is an equality under the given assumptions. We have $\tau(U, \theta) = \tilde{\tau}(r(U), \theta)$. The condition $\tau(U, \theta) = t$ can therefore be replaced by $r(U) = \tilde{v}(\theta, t)$. This can in turn be replaced by $\tilde{\theta}(r(U), t) = \theta$. Now $\hat{\theta}(U, t) = \tilde{\theta}(r(U), t)$ and this finishes the proof. The independence result of the theorem follows since $r(U)$ and $\tau(U, \theta)$ generate the same σ-field.

We finally note that under the assumptions of Theorem 4 and the additional assumption of the Jacobi case, condition (10) is always satisfied. Moreover, it is possible to give an explicit formula for the density $f(\theta)$ which makes the $Y_t(u)$ of (7) identically equal to 1.

First it follows from Assumption 2 that $r(u) = \tilde{v}(\hat{\theta}(u,t),t)$. Hence $\tau(u,\theta) = \tilde{\tau}(\tilde{v}(\hat{\theta}(u,t),t),\theta)$, so $\det\partial_\theta\tau(u,\theta) = \det\partial_\theta\tilde{\tau}(\tilde{v}(\hat{\theta}(u,t),t),\theta)$. By substituting $\hat{\theta}(u,t)$ for θ it is seen that (10) holds, and the desired $f(\theta)$ is given by

$$f(\theta) = |\det\partial_\theta\tilde{\tau}(v,\theta)|_{v=\tilde{v}(\theta,t)} \tag{11}$$

5 Monte Carlo Conditioning

The previous sections are devoted to deriving expressions for conditional expectations $\mathrm{E}^\theta(\phi(X)|T=t)$ in terms of expectations of functions of U. The idea is that these expectations can be evaluated by Monte Carlo simulation. Basically, we can estimate $M_U h(U)$ by $(1/N)\sum_{i=1}^N h(u_i)$ where $u_1,\ldots,u_N$ is a computer generated sequence corresponding to N independent realizations of U. Standard theory on Monte Carlo simulation contains methods for improving on this naive approach of estimating $M_U h(U)$. We shall, however, not pursue this matter here.

In Section ... we investigated conditions under which $X_t(u)$ from Assumption 1 directly gives samples from the conditional distribution X given $T=t$, thus not having to explicitly define a measure M_Θ on Ω_Θ. In general, however, we need to specify a measure M_Θ and the choice may influence the effectiveness of the algorithm. It should be stressed that M_Θ in a practical application with given t may well be chosen to depend on the value of t. Equation (10) gives an example of how t may enter the choice of distribution M_Θ.

Another instance where one wants to choose M_Θ according to t is discussed by Engen and Lillegård (1997). In discrete models the sets $\Gamma(u,t)$ are typically intervals (see e.g. Example xxx). The endpoints may however be difficult to find explicitly. Engen and Lillegård (1997) suggest to let M_Θ be counting measure on a grid of points with center at $\theta^*(t)$, the maximum likelihood estimate for θ when $T=t$.

A closer study of the formulae of Theorem 2? and 3? indicate that M_Θ should preferably be chosen to minimize the variation in u of the "weights" $Y_t(u)$. As an example, consider the formula in Theorem 2old where the weights are given by

$$Y_t(u) = |\frac{f(\theta)}{\det\partial_\theta\tau(u,\theta)}_{\theta=\hat{\theta}(u,t)} \tag{12}$$

Here the density $f(\theta)$ should be chosen so that $f(\hat{\theta}(u,t))$ as a function of u is roughly proportional to $|\det\partial_\theta\tau(u,\theta)|_{\theta=\hat{\theta}(u,t)}$. In the pivotal case it follows by Theorem ... that we may choose f such that (12) is constant in u. For example, in the case where X is a sample from $N(\mu,\sigma)$ it will be shown in Example 6 that such an f is given by $f(\mu,\sigma)=1/\sigma$, which is the standard noninformative (improper) prior for this case. This motivates the use of vague prior distributions from Bayesian theory as natural choices for distributions M_Θ, for example Jeffreys' priors (Jeffreys, 1946).

As a simple motivation, consider the (pivotal) case where θ is one-dimensional and $\tau(u,\theta)=r(u)\theta$. This means that T/θ is a pivotal quantity. Assume that $M_U r(U)=1$. In this case $\partial_\theta\tau(u,\theta)=r(u)$ and $\hat{\theta}(u,t)=t/r(u)$. Hence $Y_t(u)$ in (12) becomes constant in u by choosing $f(\theta)=1/\theta$. Assuming that T is the maximum likelihood estimator of θ, then it can be seen that $1/\theta$ is in fact the Jeffreys' prior here. In connection with the use of Jeffreys' priors, it is also interesting to note that the terms $f(\theta)/|\det\partial_\theta\tau(u,\theta)|$ in Theorems 2 and 3? remain unchanged under one-to-one reparametrizations of θ if f is chosen by Jeffreys' rule.

For a chosen measure M_Θ, we are able to compute conditional expectations $E^\theta(\phi(X)|T=t)$ by simulation for given fundtions ϕ. In some cases it may be of interest, however, to simulate direct samples from the conditional distribtuion of X given $T=t$. This was the intention of Algorithm 1 in the introduction. In Section ... we showed how this can be done in certain special cases through $X_t(u)$, but in general it is not straightforward to obtain such samples.

By Theorem 1 we need to sample (u,θ) from the conditional distribution of (U,Θ) given $\tau(U,\Theta)=t$ and then compute $x_t=\chi(u,\theta)$. Theorem 2 implies that the appropriate distribution from which to sample (u,θ) is

$$\frac{Y_t(u)\rho(d\theta|u,t)M_U(du)}{\int Y_t(u)\rho(d\theta|u,t)M_U(du)} \tag{13}$$

In practice it may be easier to start by sampling (u, θ) as follows: (i) Draw u from the distribution M_U. (ii) Given $U = u$, draw θ from the conditional distribution of Θ given $\tau(u, \Theta) = t$. This means that (u, θ) is sampled from the distribution

$$Y_t(u)\rho(d\theta|u,t)M_U(du) \tag{14}$$

It thus remains to modify samples coming from (14) to become samples from (13), for example rejection sampling. This requires, however, that we are able to bound the function $Y_t(u)$ from above. This may be impossible or impractical in many cases. In such cases we are led to using for example Markov Chain Monte Carlo methods (Tierney 1994) or the SIR-algorithm (Rubin, 1988).

6 Examples

We now proceed by giving four examples with the aim of illustrating the theoretical results, as well as indicating their relevance in statistical inference problems.

[Normal distribution] This example illustrates in a well known setting the basic ideas of the paper. Moreover, it gives an example of a case where the algorithm of the introduction holds.

Let $X = (X_1, \dots, X_n)$ be a vector of i.i.d. $N(\mu, \sigma)$ variables and let $T = (\bar{X}, S)$ be the sufficient statistic. Here $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$. For simulation in this model we use a vector $U = (U_1, \dots, U_n)$ of i.i.d. $N(0,1)$ variables and simulate X and T by, respectively,

$$\begin{aligned}
\chi(U,\theta) &= (\mu + \sigma U_1, \dots, \mu + \sigma U_n) \\
\tau(U,\theta) &= (\mu + \sigma \bar{U}, \sigma S_U)
\end{aligned}$$

where $\theta = (\mu, \sigma)$.

Now Assumption 1 holds with

$$\hat{\theta}(U,t) = (\bar{x} - \frac{\bar{U}}{S_U}s, \frac{1}{S_U}s)$$

where $t = (\bar{x}, s)$, and $\bar{U}, S_U$ are similar to $\bar{X}, S$. Also, Assumption 2 holds with $r(U) = (\bar{U}, S_U)$ and hence Theorem 4 implies that samples from the conditional distribution of X given $T = t$ are given by

$$X_t = \chi(U, \hat{\theta}(U,t)) = (\bar{x} + \frac{U_1 - \bar{U}}{S_U}s, \dots, \bar{x} + \frac{U_n - \bar{U}}{S_U}s)$$

Thus X_t has a multivariate t-distribution.

The above formulas show that X_t and $\tau(U,\theta)$ are stochastically independent (use for example Basu's theorem, Lehmann and Casella, 1998, p. 42), likewise that X_t and $\hat{\theta}(U,t)$ are independent. Thus from the discussion following Theorem **??** we can conclude that X_t has the correct distribution even without using Theorem 4.

Further, $\det \partial_t \hat{\theta}(U,t) = 1/S_U$ so (**??**) holds with $J(\theta, t) = \sigma/s$ and we may hence choose $f(\mu,\sigma) = 1/\sigma$ for the density defining M_Θ. Accidentally, this is Jeffreys' suggested joint prior for (μ, σ) when μ and σ are given independent priors (referanse).

[Nonhomogenous Poisson process] This example provides a counterexample to the algorithm of the introduction in the case where $\hat{\theta}(U,t)$ is uniquely determined.

Consider a nonhomogenous Poisson process on the time interval $[0,1]$, with intensity function $\lambda(t) = \exp(\delta + \theta t)$. Let N be the number of events, observed at times $X_1, X_2, \dots, X_N$. It is well known that given $N = n$, the $X_1, X_2, \dots, X_n$ are distributed as the order statistic of n i.i.d. observations from the distribution function on $[0,1]$ given by

$$F(x,\theta) = \int_0^x \lambda(t) / \int_0^1 \lambda(t)dt = \begin{cases} \frac{e^{\theta x}-1}{e^{\theta}-1}, & \theta \neq 0 \\ x, & \theta = 0 \end{cases} \tag{15}$$

This distribution function depends analytically on the parameter $\theta \in R$. The statistic $T = \sum_{i=1}^n X_i$ is sufficient compared to $X = (X_1, \dots, X_n)$.

Let $U = (U_1, U_2, \ldots, U_n)$ be a vector of n i.i.d. random variables uniformly distributed on $[0,1]$. Inversion based on the distribution function (15) gives $(X,T) \sim (\chi(U,\theta), \tau(U,\theta))$ where $\chi(u,\theta) = (\eta(u_1,\theta), \ldots, \eta(u_n,\theta))$, $\tau(u,\theta) = \sum_{i=1}^n \eta(u_i,\theta)$, and

$$\eta(v,\theta) = \begin{cases} \frac{\log(1+(e^\theta-1)v)}{\theta} & \theta \neq 0 \\ v & \theta = 0 \end{cases}$$

The function $\eta(v,\theta)$ is analytic and strictly increasing in θ. This follows from the equation $F(\eta(v,\theta),\theta) = v$ which determined η together with the observation that the distribution function $F(x,\theta)$ is strictly decreasing in θ. Consequently Assumption 1 is fulfilled and $\hat{\theta}(u,t)$ is uniquely determined by $\tau(u, \hat{\theta}(u,t)) = t$.

Consider for a moment the case when $n = 2$. It is then easy to check that the conditional distribution of X_1 given $T = t$ is uniform on $(0,t)$ for $0 < t \leq 1$, and uniform on $(t-1, 1)$ for $1 < t < 2$. Still we can prove that $X_{t,1} = \eta(U_1, \hat{\theta}(U,t))$ is not uniformly distributed for any $0 < t < 2$. This can be demonstrated in general by numerical calculation, but we treat only the special case $t = 1$. Direct calculation gives

$$\eta(U_1, \hat{\theta}(U,t)) = \frac{\log \frac{1-U_1}{U_2}}{\log \frac{1-U_1}{U_2} + \log \frac{1-U_2}{U_1}}$$

which is easily verified to be non-uniformly distributed and thus establishes the announced counter example.

Return now to the case of general n and suppose we want to compute $E(\phi(X)|T=t)$. By Theorem 2 we need to compute $|\det \partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)}$, and we get

$$\partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)} = \frac{e^{\hat{\theta}(u,t)}}{\hat{\theta}(u,t)} \sum_{i=1}^n \frac{u_i}{1 + (e^{\hat{\theta}(u,t)} - 1)u_i} - \frac{t}{\hat{\theta}(u,t)}$$

In principle we can use Theorem 2 with any choice of $f(\theta)$. As suggested in subsection 3.1 we should aim at flattening out the $Y_t(u)$. A straightforward computation gives the square root of Fisher-information to be $(1/\theta^2 - \exp(\theta)/(\exp(\theta)-1)^2)^{1/2}$ which is the Jeffreys' prior here and thus a possible choice for f.

We finally note that since $\theta = 0$ corresponds to the X_i being uniformly distributed on $[0,1]$, we consider in fact the purely probabilistic problem of finding the conditional distribution of n i.i.d. uniform random variables given their sum.

[Logistic regression] In this example we have discrete observations and a continuous parameter. In this case we will typically have interval solutions for θ of the equation $\tau(u,\theta) = t$.

Let $X = (X_1, \ldots, X_n)$ be independent random variabels such that X_i is Bernoulli-distributed with

$$P(X_i = 1) = p_i(\theta, \delta) = \frac{e^{\theta + \delta z_i}}{1 + e^{\theta+\delta z_i}}$$

where $z_1, \ldots, z_n$ are given real covariates and $-\infty < \delta, \theta < \infty$ are parameters. Suppose δ is the parameter of interest. Since $T = \sum_{i=1}^n X_i$ is sufficient for θ for fixed δ, we can eliminate the nuisance parameter θ by conditioning on T.

Thus, consider the value of δ as fixed at δ_0 in the following. We shall see how the method of Proposition **??** can be used to find the conditional distribution of X given $T = t$. For example, this distribution is of interest when testing the null hypothesis $H_0 : \delta = \delta_0$.

Let $U = (U_1, \ldots, U_n)$ be i.i.d. and uniform on $[0,1]$. Then X under θ can be simulated as $\chi(U,\theta) = (\eta_1(U_1,\theta), \eta_2(U_2,\theta), \ldots, \eta_n(U_n,\theta))$ where

$$\eta_i(u_i,\theta) = [u_i < p_i(\theta,\delta_0)], \tag{16}$$

and T is therefore simulated by $\tau(U,\theta) = \sum_{i=1}^n \eta_i(u_i,\theta)$.

From(16) follows that

$$\eta_i(u_i,\theta) = 1 \leftrightarrow \theta > \mathrm{logit}(u_i) - \delta_0 z_i \equiv \psi_i(u)$$

where $\text{logit}(u) = \log(u/(1-u))$. By ordering the $\psi_i(u)$ as $\psi_{(1)}(u) < \psi_{(2)}(u) < \cdots \psi_{(n)}(u)$ it follows that $\tau(u,\theta) = t$ if and only if $\psi_{(t)}(u) < \theta < \psi_{(t+1)}(u)$. Thus the set of θ satisfying $\tau(u,\theta)$ is an interval and

$$Y_t(u) = M_\Theta(\psi_{(t)}(u), \psi_{(t+1)}(u)) \tag{17}$$

where M_Θ is a measure on the real numbers.

Since each $\eta_i(u_i,\theta)$ is constant (either 0 or 1) on the interval $(\psi_{(t)}(u), \psi_{(t+1)}(u))$ it follows that Assumption 1 holds and that we can use Proposition **??** for computing conditinal expectations. Also, we can use the SIR algorithm discussed in the previous section to obtain approximate samples from the conditional distribution.

Given the measure M_Θ the computation of weights Y_t is straightforward since we have the explicit solution (17). For choice of measure, it is by (17) mathematically convenient to use Lebesgue measure. Otherwise one may compute Jeffreys' prior distribution or an approximation of it.

For the special case $\delta_0 = 0$ the conditional distribution of T is of course well known. We shall show that the above derived algorithm simplifies in this case, since we can show that $X_t(u)$ and $Y_t(u)$ are independent. Indeed, now $\psi_i(u) = \text{logit}(u_i)$ so

$$X_t(u) = ([u_1 \le u_{(t)}], [u_2 \le u_{(t)}], \ldots, [u_n \le u_{(t)}])$$

which is clearly stochastically independent of $(u_{(1)}, \ldots, u_{(n)})$. Thus samples from the conditional distribution are given as $X_t(u)$, and it is seen that our algorithm simplifies to a standard one.

It is interesting to note that the algorithm derived in this example also solves the following non-statistical problem. Suppose that $X = (X_1, \ldots, X_n)$ are independent random variabels where X_i is Bernoulli-distributed with $P(X_i = 1) = p_i,\ i = 1, \ldots, n$, and that the conditional distribution of X given $T = \sum_{i=1}^n X_i$ is wanted. Then with $z_i = \text{logit}(p_i)$ for $i = 1, \ldots, n$ and $\delta_0 = 1$ it is seen that this corresponds to $\theta = 0$ in the above statistical setup. The suggested algorithm can therefore be used to solve the present problem as well.

[Censored exponential varables]

In this example we show how Theorem 1 (or 2?) can be applied in a case where the equation $\tau(u,\theta) = t$ may have no solution for θ for given u and t.

Let n units with potential lifetimes $Y_1, Y_2, \ldots, Y_n$ be observed from time 0, but assume that the observation of the ith unit is censored at a given time $c_i > 0$. This means that we observe only $X_i = \min(Y_i, c_i),\ i = 1, \ldots, n$. Suppose $Y_1, \ldots, Y_n$ are i.i.d. exponentially distributed with expected value θ. Then the likelihood of $X_1, \ldots, X_n$ can be written $(1/\theta^R)\exp(-S/\theta)$ where $R = \sum_i [X_i < c_i]$ is the number of noncensored observations and $S = \sum_i X_i$ is the sum of all observations. Here $T = (R, S)$ is sufficient compared to $X = (X_1, \ldots, X_n)$.

It should be remarked that the potential censoring times c_i are assumed known also for the units where $X_i < c_i$. For example this is the case if n machines are observed from possibly different starting points in time, and until a common end point. Let $c_1, \ldots, c_n$ be fixed, known numbers in the following.

In order to simulate X for a given θ we let $U = (U_1, \ldots, U_n)$ be a vector of n i.i.d. exponential random variables with expected value 1 and simulate X by means of $\chi(U,\theta) = (\eta_1(U_1,\theta), \ldots, \eta_n(U_n,\theta))$ where

$$\eta_i(U_i,\theta) = \min(\theta U_i, c_i),\ i = 1, \ldots, n$$

Thus T is simulated by $\tau(U,\theta) = (\gamma(U,\theta), \psi(U,\theta))$ where $\gamma(U,\theta) = \sum_i [\theta U_i < c_i]$ and $\psi(U,\theta) = \sum_i \eta_i(U_i,\theta)$.

Let now $t = (r,s)$ be fixed with $0 < r \le n,\ 0 < s < \sum_i c_i$ and consider the equations $\gamma(u,\theta) = r,\ \psi(u,\theta) = s$ for a given u. Since $\psi(u,\theta)$ is stricly increasing in θ, from 0 to $\sum_i c_i$, there is a unique $\tilde\theta$ which satisfies $\psi(u,\tilde\theta) = s$. However, this $\tilde\theta$ may not solve $\gamma(u,\theta) = r$, even though the solution for θ of the latter equation is an interval. In the cases where indeed $\gamma(u,\tilde\theta) = r$, let $I(u) = \{i_1, \ldots, i_r\}$ be the set of indices i for which $\tilde\theta u_i < c_i$. With this notation we can then express the solution $\tilde\theta$ as

$$\hat\theta(u,(r,s)) = \frac{s - \sum_{i \notin I(u)} c_i}{\sum_{i \in I(u)} u_i} \tag{18}$$

We shall use Theorem 1 for computation of conditional expectations $\mathrm{E}^\theta(\phi(X)|T = t)$. First we need to equip $\Omega_\Theta = R^+$ with a measure. Let this measure be given by a density $f(\theta)$ with respect to Lebesgue measure. We then need to find the density $Y_t(u)$ of $\tau(u,\Theta)$ with respect to the product measure of counting measure and Lebesgue measure. This is found from

$$\begin{aligned} g(u,(r,s))dr\,ds &= M_\Theta\{\theta : \gamma(u,\theta)=r, s\le \psi(u,\theta)\le s+ds\} \\ &= \begin{cases} 0 & \text{if no solution of } \tau(u,\theta)=(r,s) \\ f(\hat\theta(u,(r,s)))ds/\sum_{i\in I(u)} u_I & \text{otherwise} \end{cases} \end{aligned}$$

Furthermore, the $W_t(u)$ in Theorem **??** is given by

$$W_t(u) = \begin{cases} 0 & \text{if no solution of } \tau(u,\theta)=(r,s) \\ X_t(u)\equiv \chi(u,\hat\theta(u,(r,s)) & \text{otherwise} \end{cases}$$

The choice $f(\theta)=1/\theta$ yields in particular the simple weight function

$$Y_t(u) = (s - \sum_{i\notin I(u)} c_i)^{-1}$$

An important special case is when the c_i are all equal. In this case we can sample directly from the conditional distribution of X given $T=t$ by using $X_t(u)=\chi(u,\hat\theta(u,(r,s))$ whenever $\hat\theta(u,(r,s))$ is defined.

It seems clear that the method of this example may be ineffective if n is large, since the probability of drawing a u for which a $\hat\theta$ exists may then be small.

[Multiple solutions for θ]

Consider the distribution on R with density function

$$f(x,\theta) = \begin{cases} \frac{\theta x^{\theta-1}}{a^\theta+b^\theta} & \text{if } 0<x<a \\ \frac{\theta(x-a)^{\theta-1}}{a^\theta+b^\theta} & \text{if } a<x<a+b \end{cases}$$

where $a,b>0$ are given constants and $\theta>0$ is the unknown parameter. It is straightforward to show that if $X=(X_1,\ldots,X_n)$ is a sample from this distribution, then

$$T=\sum_{X_i<a}\log X_i + \sum_{X_i>a}\log(X_i-a)$$

is sufficient for θ.

The given density, and more general versions of it, appear naturally when we study several independent nonhomogeneous Poisson processes with a common intensity function $\lambda(t)=\alpha\theta t^{\theta-1}$ (the so called power law process, see e.g. Crowder et al. 198x), but observed over time intervals with different lengths. As an example, consider two processes, observed over time intervals $(0,a)$ and $(0,b)$, respectively. Suppose we glue together the processes into one process observed on $(0,a+b)$. Then, if we condition on the total number n of events, the event times in the glued process are distributed as the order statistic of n i.i.d. observations from the density $f(x,\theta)$.

For given values of θ we can simulate X from the density $f(x,\theta)$ by ordinary inversion using U with uniform distribution on $(0,1)$. In fact we then simulate X_i by

$$\eta(u_i,\theta) = \begin{cases} u_i(a^\theta+b^\theta))^{1/\theta} & \text{if } u_i<a^\theta/(a^\theta+b^\theta) \\ u_i(a^\theta+b^\theta)-a^\theta)^{1/\theta}+a & \text{if } u_i>a^\theta/(a^\theta+b^\theta) \end{cases}$$

By the above definition of T it is from this also clear how to define $\tau(u,\theta)$ as an appropriate sum of terms of the form $\eta(u_i,\theta)$.

The point of this example is to show that the equation $\tau(u,\theta)=t$ may have a varying number of solutions for θ, depending on the values of u and t. As an illustration we have in Figure xxx plotted the function $\tau(u,\theta)$ as a function of θ when $n=2, a=3, b=1$ and $u=(0.5,0.9)$. It is seen that there are two solutions for θ if $t<0.2$, one solution if $t>1.9$ and no solution if $0.2<t<1.9$. As an example, for $t=0$ we get the solutions $\theta_1=0.925$, $\theta_2=1.801$. Thus the formula of "Corollary" is needed for computation of conditional expectations.

7 Concluding remarks

Si noe om hva vi har gjort....

Andre ting (skal ikke nummereres nødvendigvis i den endelige versjon):

1. The results obtained in this paper are believed to be of greatest value in small sample cases where limit results based on asymptotic normality may not be trustworthy. Note that bootstrapping may be a reasonable "intermediate" alternative in many cases. For example, if θ is a nuisance parameter under the null hypothesis H_0 of a testing problem, the approach of the present paper amounts to sampling conditionally on a sufficient statistic under H_0, while bootstrapping means to sample from an estimated distribution under H_0. The latter approach is thus not exact, but may work well at least for for moderate samples (Efron and Tibshirani, 1993).

2. In particular we have shown that the simple algorithm of the introduction holds if $\hat{\theta}$ is unique and if the model in addition has a certain pivotal structure. Remark that, as noted by Mood et al. (1974), it is always possible to find a pivotal quantity when $X_1, \dots, X_n$ is a sample from a continuous distribution $F(x, \theta)$; simply by taking $-2\sum_{i=1}^n \log F(X_i, \theta))$ which is chi-square distributed with $2n$ degrees of freedom. However, as stated in Theorem **??** our pivotal condition is stronger since it requires that the sufficient statistic can be recovered from the pivotal quantity. For example, this would not be the case in Example 6.

3. In the introduction we mentioned the possibility of constructing statistical problems from probabilistic ones. This can typically be done by introducing exponential families with the correct sufficient statistic. The main problem in practice is, however, to define a simulation algorithm based on a suitable U as in this paper. This problem is discussed for example in Bølviken and Skovlund (1996).

4. The results of Section **??** may be used in theoretical derivations of conditional distributions in particular cases. As an example, in the normal case of Example 6 it follows easily from our setup that the conditional distribution of X given $\bar{X}, S$ is a multivariate t-distribution. Also, Theorem **??** may be applied to find pivotal quantitites, and also to prove independence results.

5. Consider now a Bayesian setting where T is an observed sufficient statistic. Suppose there is given a prior distribtuion for the parameter θ, which we identify with M_Θ. In the notation of the present paper, the posterior distribution of Θ given the observation $T = t$ equals the conditional distribution of Θ given $\tau(U, \Theta) = t$. The latter distribtuion is therefore the marginal distribtuion of Θ in the distribution given by Theorem 2new. For example, under the uniqueness conditions of Theorem ..., the posterior expectation of a function $h(\Theta)$ would be

$$E(h(\Theta)|\tau(U,\Theta) = t) \frac{\int h(\hat{\theta}(u,t)) |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\hat{\theta}(u,t)} M_U(du)}{\int |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\hat{\theta}(u,t)} M_U(du)}$$

   where $f(\theta)$ is interpreted as the prior density for θ.

   In the special case of normal distributions as given in Example 6 we get the posterior expectation of a function $h(\mu, \sigma)$ to be given by

$$\frac{\mathrm{E}(h(\bar{x} - \frac{\bar{U}}{S_U}s, \frac{1}{S_U}s)\frac{1}{S_U} f(\bar{x} - \frac{\bar{U}}{S_U}s, \frac{1}{S_U}s))}{\mathrm{E}(\frac{1}{S_U} f(\bar{x} - \frac{\bar{U}}{S_U}s, \frac{1}{S_U}s))}$$

   This expression exactly corresponds to evaluating the posterior expectation by ordinary importance sampling, drawing (μ, σ) from the distribution of $\hat{\theta}(U, t) \equiv (\bar{x} - \frac{\bar{U}}{S_U}s, \frac{1}{S_U}s)$ which turns out to equal Fisher's fiducial distribution for this case (see e.g. Welsh, 1996).

   The formula can be easily evaluated by simulation by noting that $\bar{U}$ and S_U are independent, $\bar{U} \sim N(0, 1/n)$ and $(n-1)S_U^2 \sim \chi^2_{n-1}$.

Bo Lindqvist
Department of mathematical sciences
Norwegian University of Science and Technology
N-7034 TRONDHEIM

NORWAY

Gunnar Taraldsen
Department of Physiology and Biomedical Engineering
Norwegian University of Science and Technology
Center of Medicine and Technology
N-7005 TRONDHEIM

NORWAY