NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Monte Carlo Conditioning on a Sufficient Statistic

by

Bo Henry Lindqvist and Gunnar Taraldsen

PREPRINT
STATISTICS NO. 9/2001

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S9-2001.ps

Monte Carlo Conditioning on a Sufficient Statistic

BY BO HENRY LINDQVIST AND GUNNAR TARALDSEN

Norwegian University of Science and Technology
and SINTEF Telecom and Informatics

We derive general formulae for computation of conditional expectations of functions $\phi(X)$ given $T = t$, when (X, T) is a pair of random vectors such that T is sufficient compared to X for a parameter θ. The basic assumption is that there is a random vector U with a known distribution and functions χ, τ such that (X, T) under the parameter value θ has the same distribution as the pair $(\chi(U, \theta), \tau(U, \theta))$. The conditional expectations are then expressed in terms of ordinary expectations of functions of U, and are thus well suited for computation by Monte Carlo simulation. The clue is to equip the parameter space by a suitably chosen σ-finite measure. The problem of direct sampling from conditional distributions of X given $T = t$ is also considered. It is shown in particular that when the model satisfies a certain pivotal condition, then this can be done by sampling $\chi(U, \theta)$ where the value of θ is adjusted for each U in such a manner that $\tau(U, \theta) = t$ is kept fixed. Several examples are given in order to demonstrate different cases which may occur in practice.

1. Introduction. Let (X, T) be a pair of random vectors with joint distribution indexed by a parameter θ. Suppose further that T is sufficient for θ compared to X. This means that the conditional distribution of X given T does not depend on θ. The concept of sufficiency is basic in statistical inference, with applications for example in parameter elimination, construction of optimal estimators, and in testing [Lehmann (1986), Lehmann and Casella (1998)]. In practical cases, however, the conditional distributions given T may be 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 conditional expectations of functions $\phi(X)$ given $T = t$ would be to fix a parameter value θ_0 and then compute the conditional expectation under θ_0. In this way the problem is a traditional one. Trotter and Tukey (1956) introduced the term "conditional Monte Carlo". Their idea [see e.g. Ripley (1987), page 136] was to determine a weight function $w_t(\cdot, \cdot)$ such that $\mathrm{E}(\phi(X)|T = t) = \mathrm{E}\phi(X)w_t(X, T)$ for any function $\phi(\cdot)$, thus replacing conditional expectations by ordinary expectations with respect to the original distribution of (X, T) and allowing Monte Carlo computation by sampling from this distribution.

The general method presented in this paper is similar in spirit, as it replaces conditional expectations by ordinary expectations which can be computed by Monte Carlo simulation. The clue is to utilize the fact that the conditional distribution of X given $T = t$ is the same for all θ, and, instead of fixing the value of θ, to equip the parameter space Ω by a σ-finite measure π. This idea is inherent in the paper by Engen and Lillegård (1997), which was the point of departure of the present work.

In order to describe our approach in more detail we introduce next the general notation and assumptions used in the paper. Let $\mathcal{S}$ be a measurable space equipped with a family of probability measures P^θ indexed by a parameter θ in Ω, where Ω is itself a measurable space. We shall consider statistics X and T, which more precisely are measurable functions $X : \mathcal{S} \to \mathcal{S}_X$ and $T : \mathcal{S} \to \mathcal{S}_T$ for measurable spaces $\mathcal{S}_X$ and $\mathcal{S}_T$. Throughout the paper we assume that T is sufficient for θ compared to X, with the precise meaning that to each bounded and measurable function $\phi : \mathcal{S}_X \to R$ there is a measurable function $\psi : \mathcal{S}_T \to R$ which is a version of the conditional expectation $E^\theta(\phi(X)|T = t)$ for all $\theta \in \Omega$. The function $\psi(t)$ will usually be denoted $E(\phi(X)|T = t)$.

AMS 2000 subject classifications. Primary 62B05, 65C05; secondary 62H10, 65C60
Key words and phrases. Sufficiency, conditional distribution, Monte Carlo simulation, pivotal statistic

Following the main idea of Engen and Lillegård (1997) we assume that there is given a random vector U with values in a measurable space $\mathcal{M}$, called the Monte Carlo space, and with a known distribution M, such that (X,T) for given θ can be simulated by means of U. More precisely we assume the existence of measurable functions $\chi : \mathcal{M} \times \Omega \to \mathcal{S}_X$ and $\tau : \mathcal{M} \times \Omega \to \mathcal{S}_T$ such that for each $\theta \in \Omega$ the joint distribution of $(\chi(U,\theta), \tau(U,\theta))$ equals the joint distribution of (X,T) under P^θ. Here the product space $\mathcal{M} \times \Omega$ is equipped with the usual product σ-field. Functions of U will be called Monte Carlo variables. Expectations of Monte Carlo variables $Y \equiv Y(U)$ will be denoted $M(Y)$ and a basic idea of the paper is that such expectations are computable by Monte Carlo simulation.

EXAMPLE 1 - NORMAL DISTRIBUTION. Let $X = (X_1, \ldots, X_n)$ be a vector of i.i.d. $N(\mu, \sigma)$ variables. Then $T = (\bar{X}, S)$ is sufficient compared to X, where $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$. Letting $U = (U_1, \ldots, U_n)$ be i.i.d. $N(0,1)$ variables we can put

$$\chi(U,\theta) = (\mu + \sigma U_1, \ldots, \mu + \sigma U_n)$$

$$\tau(U,\theta) = (\mu + \sigma \bar{U}, \sigma S_U)$$

where $\theta = (\mu, \sigma)$ and $\bar{U}, S_U$ are defined in the obvious way similar to $\bar{X}, S$. $\square$

Now suppose it is given that $T = t$, and that a sample from the conditional distribution of X given $T = t$ is wanted. Since the conditional 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. A suggestive way for this would be to first draw U from the distribution M, then determining a parameter value $\hat{\theta}$ such that $\tau(U, \hat{\theta}) = t$ and using $X_t = \chi(U, \hat{\theta})$ as the desired sample. In this way we indeed get a sample of X with the corresponding T having the correct value t. The question remains, however, whether X_t obtained in this way really is distributed like X given $T = t$.

EXAMPLE 1 (CONTINUED). For given $t = (\bar{x}, s)$ and U there is a unique $\hat{\theta} \equiv \hat{\theta}(U,t)$ with $\tau(U, \hat{\theta}) = t$, namely

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

This leads to the sample

$$X_t = \chi(U, \hat{\theta}(U,t)) = (\bar{x} + \frac{U_1 - \bar{U}}{S_U} s, \ldots, \bar{x} + \frac{U_n - \bar{U}}{S_U} s) \tag{1.1}$$

and it can be shown that the distribution of X_t indeed coincides with the conditional distribution of X given $T = t$. $\square$

Engen and Lillegård (1997) showed by examples that $X_t = \chi(U, \hat{\theta})$ in general may not have the correct distribution when $\hat{\theta}$ is not uniquely determined by t and u from the equation $\tau(u, \theta) = t$. They claimed, however, that X_t will always be distributed like X given $T = t$ when $\hat{\theta}$ is unique as was the case for the normal example above. Unfortunately this claim does not hold in general, as will be shown in Example 3 in subsection 2.2.

In order to see what may go wrong, assume that the equation $\tau(u, \theta) = t$ for each fixed u, t has the unique solution $\theta = \hat{\theta}(u,t)$. Define for fixed t the Monte Carlo variable

$$X_t = \chi(U, \hat{\theta}(U,t)) \tag{1.2}$$

and consider the conditional expectation of $\phi(X)$ given $T = t$ for a given function ϕ. Then for all θ and t we can formally write

$$\begin{aligned} E(\phi(X)|T = t) &\overset{(i)}{=} E(\phi(\chi(U,\theta))|\tau(U,\theta) = t) \\ &\overset{(ii)}{=} E(\phi(\chi(U,\theta))|\hat{\theta}(U,t) = \theta) \end{aligned}$$

$$\stackrel{(iii)}{=} E(\phi(\chi(U,\hat\theta(U,t)))|\hat\theta(U,t)=\theta)$$
$$\stackrel{(iv)}{=} E(\phi(\chi(U,\hat\theta(U,t)))) \equiv E(\phi(X_t))$$

which if correct would imply that X_t has the correct distribution, in the sense that X_t is distributed like the conditional distribution of X given $T=t$.

In fact, the equality (i) follows from sufficiency, (iii) follows by substitution in conditional expectations [Bahadaur and Bickel (1968)], while (iv) follows since the conditional expectation on the left hand side of this equality sign is constant in θ, and hence must equal the unconditional expectation when t is fixed.

Equality (ii) follows formally from equivalence of the events $\{\tau(U,\theta)=t\}$ and $\{\hat\theta(U,t)=\theta\}$. This is unproblematic if these events have positive probability, but the equality may otherwise be invalid, a fact which is known as the Borel paradox [Kolmogorov (1933)]. The key point in proving that X_t has the correct distribution in a particular case is therefore to prove that equality holds at (ii). This is essentially what is done in our Theorem 3, which provides a sufficient condition for X_t to have the correct distribution under uniqueness of $\hat\theta$. The basic condition is that $\tau(u,\theta)$ depends on u only through a function $r(u)$, where the value of $r(u)$ can be uniquely recovered from the equation $\tau(u,\theta)=t$ for given θ and t. This means that there is a function $\tilde\tau$ such that $\tau(u,\theta)=\tilde\tau(r(u),\theta)$ for all (u,θ), and a function $\tilde v$ such that $\tilde\tau(r(u),\theta)=t$ implies $r(u)=\tilde v(\theta,t)$. In this case $\tilde v(\theta,T)$ is a pivotal quantity in the classical meaning, and is moreover invertible according to the definition in Tukey (1957). The validity of (ii) can essentially be proven under this condition, which we shall call the pivotal condition. The basic idea of the proof is that both of the functions $u\mapsto\tau(u,\theta)$ and $u\mapsto\hat\theta(u,t)$ are in one-to-one correspondence with the function $r(u)$ (see precise statement and proof of Theorem 3).

EXAMPLE 1 (CONTINUED). The pivotal condition is satisfied here with $r(U)=(\bar U,S_U)$. This implies that X_t in (1.1) has the correct distribution, as we have already noted. □

For the general case, where the pivotal condition may not be satisfied and there may not even be a unique $\hat\theta$, we introduce the variable Θ with a σ-finite distribution π on Ω. For helping intuition we assume at this stage that π is a probability measure and hence that Θ is an ordinary random variable. We assume furthermore that U and Θ are stochastically independent.

The key result is Theorem 1 which basically states that the conditional distribution of X given $T=t$ equals the conditional distribution of $\chi(U,\Theta)$ given $\tau(U,\Theta)=t$. This is intuitively obvious from the definition of sufficiency, which implies that this holds when Θ is replaced by any fixed value $\theta\in\Omega$.

It follows that conditional expectations $\psi(t)\equiv E(\phi(X)|T=t)$ can be computed from the formula

$$\psi(t)=E(\phi(\chi(U,\Theta)|\tau(U,\Theta)=t) \tag{1.3}$$

The main idea of our method for computation of $\psi(t)$ is to compute (1.3) by first conditioning on U (in addition to $\tau(U,\Theta)$). This leads to the main result of the paper, Theorem 2, which states that (under rather general conditions)

$$E(\phi(X)|T=t)=\frac{\int W_t(u)Z_t(u)M(du)}{\int W_t(u)M(du)} \tag{1.4}$$

where $Z_t(u)$ is the conditional expectation of $\phi(\chi(u,\Theta))$ given $\tau(u,\Theta)=t$ for fixed u, while $W_t(u)$ is the density of the variable $\tau(u,\Theta)$ at t, for fixed u. An informal proof of this, starting from (1.3) is given in subsection 2.7. Thus our method essentially amounts to changing computations of conditional expectations of $\phi(\chi(U,\theta))$ given $\tau(U,\theta)=t$ for fixed θ into the (often much simpler) problem of computing conditional expectations of $\phi(\chi(u,\Theta))$ given $\tau(u,\Theta)=t$ for fixed u. Note the freedom to choose a suitable measure π for Θ. As will become clear, it is convenient to allow π to be σ-finite. This corresponds to the use of improper priors in Bayesian analysis, and will essentially not alter the arguments and results described above. However, in this case we need to replace the assumption of independence of U and Θ by the assumption that the pair (U, Θ) has distribution given by the product measure $m \equiv M \otimes \pi$ on $\mathcal{M} \times \Omega$.

The method simplifies when the equation $\tau(u, \theta) = t$ has a single solution $\theta = \hat{\theta}(u, t)$ for each pair u, t. In this case the condtional distribution of Θ given $\tau(u, \Theta) = t$ is a one point mass at $\hat{\theta}(u, t)$, so we have $Z_t(u) = \phi(\chi(u, \hat{\theta}(u, t))) = \phi(X_t(u))$ with X_t given in (1.2). In this case (1.4) can be interpreted to say that samples from the conditional distribution of X given $T = t$ are obtained as $\chi(V, \hat{\theta}(V, t))$ when V is drawn from the distribution proportional to $W_t(u)M(du)$, i.e. a weighted version of the original distribution M of U. In particular X_t will itself have the correct distribution if there exists a measure π on Ω so that $W_t(u)$ does not depend on u. This will be further discussed in subsection 2.2.

In general, direct sampling from the conditional distribution of X given $T = t$ may not be straightforward. As follows from (1.3), in principle we need to sample (u, θ) from the conditional distribution of (U, Θ) given $\tau(U, \Theta) = t$ and then compute $\chi(u, \theta)$. By (1.4), the appropriate distribution from which to sample (u, θ) is proportional to

$$W_t(u)\rho(d\theta|u, t)M(du) \tag{1.5}$$

where $\rho(d\theta|u, t)$ denotes the conditional distribution of Θ given $\tau(u, \Theta) = t$. In practice it is often useful to start by sampling (u, θ) from the simpler distribution

$$\rho(d\theta|u, t)M(du), \tag{1.6}$$

which means to first draw u from the distribution M and then draw θ from the conditional distribution of Θ given $\tau(u, \Theta) = t$. Samples from (1.5) may then be obtained from this by rejection sampling provided an upper bound for the function $W_t(u)$ is available. Alternatively, one may find it more convenient to use Markov Chain Monte Carlo methods [Tierney (1994)] or the SIR-algorithm of Rubin [see Tanner (1996)].

As will be indicated in Examples 3 and 4, our results may be utilized also for computation of conditional expectations $\mathrm{E}(\phi(X)|T = t)$ in probabilistic settings not involving parameters. In principle such computations may be done by first constructing artificial statistical models for which the conditioning variable T is sufficient. If for example $T \equiv T(X)$ is a function of X, then we can use exponential models like $f(x, \theta) = c(\theta)h(x)e^{\theta T(x)}$, where $h(x)$ is the original density of X. A main problem in practice will be to find a suitable simulation algorithm for the resulting statistical model, as required by our method. This problem is discussed by Bølviken and Skovlund (1996) for the case of one-parameter exponential models.

While the main motivation for the work has been to provide methods for Monte Carlo computation of conditional expectations, the results of the paper may also be useful for deriving theoretical representations of conditional distributions. This applies in particular to cases where the pivotal condition holds. As an example, in subsection 2.1 we derive a simple representation of the conditional distribution of a multinormal sample given the sample mean and sample covariance matrix.

The plan of the paper is as follows. In Section 2 we give several examples and study some typical special cases. The section includes a counter example to the assertion that X_t always has the correct distribution when $\hat{\theta}$ is uniquely determined. Moreover, we give an informal derivation of (1.4) and discuss briefly the relation to Bayesian and fiducial statistics. Section 3 contains statements and proofs of the precise theoretical results of the paper. The section starts by defining conditional expectations for σ-finite measures.

2. Special cases and examples

2.1. *The pivotal case.* In the present subsection we generalize the result for the univariate normal distribution in Example 1 to the multinormal case. As we shall see, the pivotal condition can be verified under a parameterization via Cholesky-decompositions. This leads to a simple recipe for simulation from the conditional distribution.

EXAMPLE 2 - MULTINORMAL DISTRIBUTION. Let $X = (X_1, \dots, X_n)$ be n i.i.d. random vectors with distribution $N_p(\mu, \Sigma)$ which is the p-dimensional multinormal distribution with expectation vector μ and covariance matrix Σ (assumed positive definite). Then $T = (\bar{X}, S)$ is sufficient compared to X, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and $S = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$, where $'$ denotes transpose. Now we reparameterize the model from (μ, Σ) to $\theta \equiv (\mu, A)$, where $\Sigma = AA'$ is the Cholesky-decomposition. Note that A is the unique lower triangular matrix with positive diagonal such that $\Sigma = AA'$ [Golub and van Loan (1996), Theorem 4.2.5, page 143].

Letting $U = (U_1, \dots, U_n)$ be i.i.d. $N_p(0, I)$ vectors (where I is the identity matrix) we can define

$$\chi(U, \theta) = (\mu + AU_1, \dots, \mu + AU_n)$$
$$\tau(U, \theta) = (\mu + A\bar{U}, AS_U A')$$

where $\bar{U}, S_U$ are defined in the same way as $\bar{X}, S$.

Let $t = (\bar{x}, s)$ be a given value of T. Consider the equation $\tau(U, \theta) = t$, which here is equivalent to

$$\mu + A\bar{U} = \bar{x}$$
$$AS_U A' = s$$

These equations are uniquely solved for (A, μ) by $A = cL_U^{-1}$, $\mu = \bar{x} - cL_U^{-1}\bar{U}$, where $s = cc'$ and $S_U = L_U L_U'$ are Cholesky-decompositons. Note that uniqueness follows from the requirement that A is lower triangular with positive diagonal. It follows that

$$\hat{\theta}(U, t) = (\bar{x} - cL_U^{-1}\bar{U}, cL_U^{-1}) \tag{2.7}$$

is the unique solution of $\tau(U, \theta) = t$.

To see that the pivotal condition holds, observe first that $\tau(U, \theta)$ depends on U only through $r(U) = (\bar{U}, S_U)$. Thus we need to show that the equation $(\mu + A\bar{U}, AS_U A') = (\bar{x}, s)$ can be uniquely solved for $(\bar{U}, S_U)$ for all μ, A. This is easily seen to be the case since A is nonsingular.

It follows that we can sample from the conditional distribution of X given $T = t$ by simulating U and computing

$$\chi(U, \hat{\theta}(U, t)) = (\bar{x} + cL_U^{-1}(U_1 - \bar{U}), \dots, \bar{x} + cL_U^{-1}(U_n - \bar{U}))$$

where $s = cc'$ and $S_U = L_U L_U'$ are Cholesky-decompositions.

It is interesting to note that the above argument would not go through if we replaced the Cholesky decompositions by square root matrices (i.e. putting $A = \Sigma^{1/2}$ etc.) $\square$

2.2. *The Euclidean case: Unique solution for θ.* In a large class of models the parameter space Ω and the range $\mathcal{S}_T$ of T are Euclidean with the same dimension, while the equation $\tau(u, \theta) = t$ has a unique solution for θ for all possible u, t. Before stating a general result on conditional expectations of $\phi(X)$ given $T = t$ for this case we give an example which provides a counter example to the assertion that X_t always has the correct distribution when $\hat{\theta}(U, t)$ is uniquely determined.

EXAMPLE 3 - THE COUNTER EXAMPLE. Let $X = (X_1, \dots, X_n)$ be a vector of i.i.d. random variables on $[0, 1]$ with distribution depending on a parameter $\theta \in \mathbb{R}$ and with distribution function

$$F(x, \theta) = \begin{cases} \frac{e^{\theta x} - 1}{e^{\theta} - 1} & \text{if } \theta \neq 0 \\ x & \text{if } \theta = 0 \end{cases} \tag{2.8}$$

for $0 \leq x \leq 1$. The statistic $T = \sum_{i=1}^n X_i$ is sufficient compared to X. Suppose we want to compute conditional expectations of functions of X given $T = t$. Note that since $\theta = 0$ corresponds to the X_i being uniformly distributed on $[0, 1]$, the relevant conditional distribution is that of n i.i.d. uniform random variables given their sum, for which there seems to be no simple expression.

Let $U = (U_1, U_2, \ldots, U_n)$ be i.i.d. from $U(0,1)$. Inversion based on the distribution function (2.8) 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

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

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 the solution $\hat{\theta}(u,t)$ of $\tau(u,\theta) = t$ is unique. The pivotal condition is not satisfied, however, since no function $r(u)$ with the required properties can be found. We show next that, in fact, X_t in (1.2) does not in general have the correct distribution.

Consider the case $n = 2$. It is 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 the first component of X_t, $\eta(U_1, \hat{\theta}(U,t))$, is not uniformly distributed for any $0 < t < 2$. For simplicity we treat only the special case $t = 1$, for which direct calculation gives

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

It is easily verified that this random variable is not uniformly distributed, and this establishes the announced counter example. □

The following result applies, however, to the situation of the above example.

PROPOSITION 1. *Assume that Ω is an open subset of R^k and that $\theta \mapsto \tau(u,\theta)$ for each u is one-to-one and differentiable. Let the measure π on Ω be given by a density $f(\theta)$ w.r.t. Lebesgue measure and define*

$$W_t(u) = \left|\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}\right|_{\theta = \hat{\theta}(u,t)}$$

Assume that $\tau(U,\Theta)$ has a σ-finite distribution m_τ and let X_t be given by (1.2). Then for any bounded real function ϕ we have

$$E(\phi(X)|T = t) = \frac{M(W_t \phi(X_t))}{M(W_t)} \qquad a.e.[m_\tau] \tag{2.9}$$

PROOF. The given expression for $W_t(u)$ is the density of $\tau(u,\Theta)$ w.r.t. Lebesgue measure [Rudin (1987), page 153]. The claim then follows from Theorem 2 of Section 3. □

The previous proposition can be viewed heuristically as a consequence of the informal result (1.4). In order to use the result we first need to choose a measure π on Ω. Note that the requirement that $\tau(U,\Theta)$ has a σ-finite distribution is automatic if π is chosen as a probability measure. The choice of π may clearly influence the efficiency of the resulting computational algorithm. A possible strategy is to try to minimize the variation in u of the "weights" $W_t(u)$. We shall see below that under certain conditions we are able to find an $f(\theta)$ such that $W_t(u)$ is constant in u. For example, when X is a sample from $N(\mu,\sigma)$ we shall see that such an f is given by $f(\mu,\sigma) = 1/\sigma$. This is the standard noninformative (improper) prior for this case and may suggest that vague prior distributions from Bayesian theory should be used as distributions π also in the general case. Further investigations in this direction are, however, beyond the scope of this paper.

EXAMPLE 3 (CONTINUED). Consider again a general n and suppose we want to compute $E(\phi(X)|T = t)$ for a given function ϕ. Using Proposition 1 we need to compute $|\det \partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)}$ and get

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

In principle (2.9) can now be used with any choice of $f(\theta)$ for which the requirements of Proposition 1 hold ($f(\theta)$ may for example be any probability density). $\square$

Proposition 1 implies in particular that $X_t \equiv \chi(U, \hat{\theta}(U,t))$ has the correct distribution provided X_t and W_t are independent. This is trivially the case if $W_t(u)$ does not depend on u, and this happens if there are functions $J(\theta)$ and $a(t)$ so that

$$|\det \partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)} = a(t) J(\hat{\theta}(u,t)) \tag{2.10}$$

for all u, t. Indeed, with $f(\theta) = J(\theta)$ we get $W_t(u) = a(t)^{-1}$ so (2.9) implies that

$$E(\phi(X)|T = t) = M(\phi(X_t)) \tag{2.11}$$

for all ϕ. Moreover, if we define $f(\theta) = g(\theta)J(\theta)$ for an arbitrary bounded function g, we get $W_t(u) = g(\hat{\theta}(u,t))a(t)^{-1}$. Then (2.9) implies

$$M(\phi(X_t)g(\hat{\theta}(U,t))) = M(\phi(X_t))M(g(\hat{\theta}(U,t)))$$

for all bounded and measurable ϕ and g, so we can conclude that X_t and $\hat{\theta}(U,t)$ are independent for fixed t when (2.10) holds.

A more general condition than (2.10) is

$$|\det \partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)} = J(\hat{\theta}(u,t), t) \tag{2.12}$$

for some function $J(\theta, t)$. If under this condition we fix t_0 and define $f_{t_0}(\theta) = J(\theta, t_0)$ we get a function $W_t(u)$ with $W_{t_0}(u) = 1$ for all u. Then under appropriate continuity conditions (which we will not pursue here) we can show that (2.11) holds for each t.

It is interesting to note that condition (2.12) is satisfied under the pivotal condition discussed in the introduction. First, with notation as given there, it follows that $r(u) = \tilde{v}(\hat{\theta}(u,t), t)$ for all u, 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 therefore seen that (2.12) holds, with

$$J(\theta, t) = |\det\partial_\theta\tilde{\tau}(v,\theta)|_{v=\tilde{v}(\theta,t)} \tag{2.13}$$

This indicates that X_t always has the correct distribution under the pivotal condition. That this indeed is so will follow from the precise and more general result and proof of Theorem 3.

Example 1 (Continued). Here we get $|\det \partial_\theta \tau(U,\theta)|_{\theta=\hat{\theta}(U,t)} = S_U = s/\hat{\sigma}(U,t)$ so (2.10) holds with $a(t) = s$, $J(\hat{\theta}(u,t)) = 1/\hat{\sigma}(u,t)$. This gives another proof of the assertion that X_t has the correct distribution. Moreover, it follows that $f(\theta) = J(\theta) = 1/\sigma$ defines a distribution π for which $W_t(u)$ does not depend on u. $\square$

2.3. *The discrete case.* The next proposition applies in particular to cases where T has a discrete distribution for all θ. The result is essentially the main result of Engen and Lillegård (1997).

Here and in the rest of the paper we shall use $[\cdot]$ to denote indicator functions.

Proposition 2. *Let ϕ be a bounded real function and define the Monte Carlo variables Y_t and W_t by*

$$Y_t(u) = \int \phi(\chi(u,\theta)) \cdot [\tau(u,\theta) = t]\, \pi(d\theta), \quad W_t(u) = \int [\tau(u,\theta) = t]\, \pi(d\theta).$$

Then for all t with $0 < M(W_t) < \infty$ we have

$$E(\phi(X)|T = t) = \frac{M(Y_t)}{M(W_t)}$$

PROOF. The proposition is a corollary of Theorem 1 in Section 3, but we give here an elementary proof which we find more transparent. 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 $P^\theta(T = t) > 0$ it follows that

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

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 and π are σ-finite. $\square$

In some cases (e.g. Example 4 below) there is a function $\tilde{X}_t(u)$ such that $\chi(u,\theta) = \tilde{X}_t(u)$ for all θ solving $\tau(u,\theta) = t$. Then the Y_t of Proposition 2 factorizes as $Y_t = \phi(\tilde{X}_t)W_t$. Moreover, if $\tilde{X}_t$ and W_t are independent, then $\tilde{X}_t$ is distributed like the conditional distribution of X given $T = t$ (see Example 4, where this happens in a special case).

In applications of the above proposition, T will typically have a discrete distribution under any θ, while Ω is an open subset of R^k. The sets $\{\theta : \tau(u,\theta) = t\}$ for fixed u, t are then usually sets with positive Lebesgue measure, which in many cases can be found explicitly. Thus W_t may often be computed directly. In these cases the measure π can conveniently be given by densities w.r.t. Lebesgue measure. In other instances, however, the set of θ solving $\tau(u,\theta) = t$ are difficult to find explicitly. For such cases Engen and Lillegård (1997) suggest to let π be counting measure on a grid of points in Ω with center at the maximum likelihood estimate for θ when $T = t$.

EXAMPLE 4 - LOGISTIC REGRESSION. Let $X = (X_1, \ldots, X_n)$ be independent random variables 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 θ by conditioning on T.

Thus, consider the value of δ as fixed at δ_0 in the following. We shall see how Proposition 2 can be used to compute conditional expectations of functions of X given $T = t$. This is of interest for example when testing the null hypothesis $H_0 : \delta = \delta_0$ or when constructing confidence intervals for δ.

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

$$\eta_i(u_i,\theta) = [u_i < p_i(\theta,\delta_0)] \tag{2.14}$$

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

From(2.14) follows that

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

where $\mathrm{logit}(u) = \log(u/(1-u))$. 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 $\tau(u,\theta) = t$ is an interval and

$$W_t(u) = \pi(\psi_{(t)}(u), \psi_{(t+1)}(u)) \tag{2.15}$$

where π is a σ-finite 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 the Y_t of Proposition 2 factorizes as $Y_t = \phi(\tilde{X}_t)W_t$ as explained after the proposition.

Note that the derived algorithm also solves the following problem. Suppose that $X = (X_1, \ldots, X_n)$ are independent and 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 $z_i = \mathrm{logit}(p_i)$

for $i = 1, \dots, n$ 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 for the current non-statistical problem. In particular we can sample from the distribution of X given $T = t$ by rejection sampling as explained at the end of Section 1 provided a probability measure is chosen for π in (2.15), noting that the weights $W_t(u)$ then are between 0 and 1. For a given value of t we should aim at choosing a measure π which gives large mass to the most probable locations of the random interval $(\psi_{(t)}(U), \psi_{(t+1)}(U))$.

As a final comment regarding this example, for the special case $\delta_0 = 0$ we have $X_1, \dots, X_n$ i.i.d. Bernoulli-variables, and it is of course well known how to do the conditioning. It is interesting to see that our algorithm simplifies in this case, since $\tilde{X}_t$ and W_t are independent. Indeed, now $\psi_i(u) = \text{logit}(u_i)$ so

$$\tilde{X}_t = ([U_1 \le U_{(t)}], [U_2 \le U_{(t)}], \dots, [U_n \le U_{(t)}])$$

which is stochastically independent of the ordered vector $(U_{(1)}, \dots, U_{(n)})$ and hence of W_t. Thus samples from the desired conditional distribution are given by $\tilde{X}_t$. □

2.4. *A mixed case.* In this subsection we give an example where the sufficient statistic has both an absolutely continuous and a discrete component.

EXAMPLE 5 - TIME CENSORED EXPONENTIAL VARIABLES. Let n units with potential lifetimes $Y_1, Y_2, \dots, Y_n$ be observed from time 0, but assume that the observation of the ith unit is censored at a given time $c_i > 0$, $i = 1, \dots, n$. This means that we observe only $X_i = \min(Y_i, c_i)$, $i = 1, \dots, n$. In the reliability terminology this is called progressive time censoring [Nelson (1972)]. Suppose $Y_1, \dots, Y_n$ are i.i.d. exponentially distributed with expected value θ. Then the likelihood of $X_1, \dots, X_n$ can be written $(1/\theta^R)\exp(-S/\theta)$ where $R = \sum_i [X_i < c_i]$ is the number of non-censored observations and $S = \sum_i X_i$ is the sum of all observations. Now $T = (R, S)$ is sufficient compared to $X = (X_1, \dots, X_n)$.

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 terminal point. Let $c_1, \dots, c_n$ be fixed, known numbers in the following.

Let $U = (U_1, \dots, U_n)$ be a vector of n i.i.d. exponential random variables with expected value 1. We can then simulate X for a given value of θ by means of $\chi(U, \theta) = (\eta_1(U_1, \theta), \dots, \eta_n(U_n, \theta))$ where

$$\eta_i(u_i, \theta) = \min(\theta u_i, c_i), \ i = 1, \dots, n$$

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

We now show how to find the functions $W_t(u)$ and $Z_t(u)$ of (1.4). First we show that the equation $\tau(u, \theta) = t$ has at most one solution for θ for fixed u, t, but may have none. Let the observed value of the sufficient statistic, $t = (r, s)$, be fixed with $0 < r \le n$, $0 < s < \sum_i c_i$. Then consider the equations $\gamma(u, \theta) = r$, $\zeta(u, \theta) = s$ for a given u. Since $\zeta(u, \theta)$ is stricly increasing in θ, from 0 to $\sum_i c_i$, there is a unique $\tilde{\theta}$ which satisfies $\zeta(u, \tilde{\theta}) = s$. However, this $\tilde{\theta}$ may not solve $\gamma(u, \theta) = r$. In the cases where indeed $\gamma(u, \tilde{\theta}) = r$, put $K(u, t) = 1$ and put $K(u, t) = 0$ otherwise. If $K(u, t) = 1$ then define $I(u, t) = \{i_1, \dots, i_r\}$ to be the set of indices i for which $\tilde{\theta} u_i < c_i$. With this notation we can express the solution $\tilde{\theta}$ when $K(u, t) = 1$ as

$$\hat{\theta}(u, t) = \frac{s - \sum_{i \notin I(u,t)} c_i}{\sum_{i \in I(u,t)} u_i} \tag{2.16}$$

Next we equip $\Omega = \mathbb{R}^+$ with a measure π given by a density $f(\theta)$ with respect to Lebesgue measure. Then we find the density $W_t(u)$ of $\tau(u, \Theta)$ with respect to the product of counting measure and Lebesgue measure. This is is found as

$$W_t(u) \, ds = \pi\{\theta : \gamma(u, \theta) = r, s \le \zeta(u, \theta) \le s + ds\}$$

$$= \begin{cases} 0 & \text{if } K(u,t) = 0 \\ f(\hat{\theta}(u,t))ds / \sum_{i \in I(u,t)} u_i & \text{if } K(u,t) = 1 \end{cases}$$

Now $Z_t(u)$ is the conditional expectation of $\phi(\chi(u,\Theta))$ given $\tau(u,\Theta) = t$, which is easily found since the conditional distribution of Θ given $\tau(u,\Theta) = t$ is a one-point mass at $\hat{\theta}(u,t)$ if $K(u,t) = 1$ and can be arbitrarily chosen otherwise. Formula (1.4) therefore gives

$$E(\phi(X)|T = t) = \frac{M\{K(U,t)\phi(\chi(U,\hat{\theta}(U,t)))f(\hat{\theta}(U,t))/\sum_{i \in I(U,t)} U_i\}}{M\{K(U,t)f(\hat{\theta}(U,t))/\sum_{i \in I(U,t)} U_i\}}$$

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

$$W_t(u) = (s - \sum_{i \notin I(u,t)} c_i)^{-1}, \tag{2.17}$$

when $K(u,t) = 1$.

An important special case is when the c_i are all equal. In this case $W_t(u)$ in (2.17) is constant in u and we can sample directly from the conditional distribution of X given $T = t$ for fixed t by drawing u until $K(u,t) = 1$ and then using $X_t(u) = \chi(u, \hat{\theta}(u,t))$. □

2.5. *Euclidean case: Finite number of solutions for θ.* In the previous subsection we gave an example where the equation $\tau(u,\theta) = t$ may have no solution for some values of u, t. The purpose of the present subsection is to consider cases where the equation has an arbitrary finite number of solutions for θ, where the number may vary with u, t. This is done through an example, but we also indicate how Proposition 1 can be extended to cover this case.

Example 6 - Several Poisson processes. 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

$$T = \sum_{X_i < a} \log X_i + \sum_{X_i > a} \log(X_i - a) \tag{2.18}$$

is sufficient compared to X.

The density $f(x,\theta)$, and more general versions of it, appear 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), 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. More precisely, letting $U = (U_1, \ldots, U_n)$ be i.i.d. $U(0,1)$ we simulate X_i by

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

and the definition (2.18) of T then makes it clear how to define $\tau(U,\theta)$ as an appropriate sum of terms of the form $\eta(U_i,\theta)$.

Now consider the equation $\tau(u,\theta) = t$. It is seen that the equation has a finite but varying number of solutions for θ, depending on the values of u and t. As an illustration, when $n = 2, a = 3, b = 1$ and $u = (0.5, 0.9)$, there are two solutions for θ if $t \leq 0.2$, one solution if $t \geq 1.9$ and no solution if $0.2 \leq t \leq 1.9$ (approximate values). For example, $t = 0$ gives the solutions $\theta_1 = 0.925$, $\theta_2 = 1.801$. □

More generally, consider the setting of Proposition 1, but assume that the assumption that $\theta \mapsto \tau(u,\theta)$ is one-to-one is replaced by the more general assumption that the set $\Gamma(u,t) \equiv \{\theta : \tau(u,\theta) = t\}$ is finite for each u, t. In this case, if τ is sufficiently smooth, the density $t \mapsto W_t(u)$ of $\tau(u,\Theta)$ with respect to Lebesgue measure is

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

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

$$\frac{f(\theta)}{|\det \partial_\theta \tau(u,\theta)| W_t(u)}, \quad \theta \in \Gamma(u,t)$$

The following result is then a Corollary of Theorem 2 in Section 3:

$$\mathrm{E}(\phi(X)|T = t) = \frac{\int \sum_{\theta \in \Gamma(u,t)} \phi(\chi(u,\theta)) \frac{f(\theta)}{|\det \partial_\theta \tau(u,\theta)|} M(du)}{\int \sum_{\theta \in \Gamma(u,t)} \frac{f(\theta)}{|\det \partial_\theta \tau(u,\theta)|} M(du)} \quad a.e.[m_\tau]$$

This holds for all bounded and measurable $\phi : \mathcal{S}_X \to R$ and gives a recipe for Monte Carlo calculation in the case of multiple roots.

2.6. *Informal proof of (1.4).* We first give a verification of (1.3). Suppose therefore that ϕ is a given bounded function and consider the problem of computing $\psi(t) \equiv E(\phi(X)|T = t)$. We shall find it convenient to write the right hand side of (1.3) as

$$E_{U,\Theta|\tau(U,\Theta)}(\phi(\chi(U,\Theta))|\tau(U,\Theta) = t) \tag{2.19}$$

using a notation indicating the variables with respect to which expectation is taken and the variables that we condition on. By the rule of double expectation (2.19) is equal to

$$E_{\Theta|\tau(U,\Theta)} E_{U|\Theta,\tau(U,\Theta)}(\phi(\chi(U,\Theta))|\Theta, \tau(U,\Theta) = t) \tag{2.20}$$

Using substitution, independence of Θ and U, and sufficiency, the inner expectation of (2.20) becomes

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

which equals $\psi(t)$ by the definition of sufficiency. Since this is a constant with respect to the outer expectation in (2.20), it follows that (2.19) equals $\psi(t)$ and (1.3) follows.

The idea of our proof of (1.4) is to change the roles of U and Θ in (2.20), and thus use the fact that we must also have

$$\psi(t) = E_{U|\tau(U,\Theta)} E_{\Theta|U,\tau(U,\Theta)}(\phi(\chi(U,\Theta))|U, \tau(U,\Theta) = t) \tag{2.21}$$

For the inner expectation in (2.21) we get

$$\begin{aligned} E_{\Theta|U,\tau(U,\Theta)}(\phi(\chi(U,\Theta))|U = u, \tau(U,\Theta) = t) &= E_{\Theta|\tau(u,\Theta)}(\phi(\chi(u,\Theta))|\tau(u,\Theta) = t) \\ &\equiv Z_t(u) \end{aligned}$$

Thus (2.21) can be rewritten as

$$\psi(t) = E_{U|\tau(U,\Theta)}(Z_t(U)|\tau(U,\Theta) = t) \tag{2.22}$$

which for computation requires the conditional distribution of U given $\tau(U,\Theta) = t$. This distribution can be derived as follows. For each fixed u, let $t \mapsto W_t(u)$ be the density of $\tau(u,\Theta)$ with respect to a fixed σ-finite measure. Then using Bayes' rule we can formally write

$$\begin{aligned} Pr(U \in du|\tau(U,\Theta) \in dt) &\propto Pr(\tau(U,\Theta) \in dt|U \in du) Pr(U \in du) \\ &= Pr(\tau(u,\Theta) \in dt) Pr(U \in du) \propto W_t(u) M(du) \end{aligned}$$

where the independence of U and Θ has been used. The normalizing constant is $\int W_t(u) M(du) \equiv M(W_t)$ and hence (1.4) follows from (2.22).

2.7. *Relation to Bayesian and fiducial statistics.* The idea of the suggested method may also be utilized in a Bayesian setting. A discussion of this may in addition throw new light on the suggested method for computation of conditional expectations. Thus suppose there is given a (possibly improper) prior distribution π for the parameter θ. If $T = T(X)$ is a function of X, then by sufficiency the posterior distribution of Θ given $X = x$ is the same as the posterior given $T = T(x) \equiv t$ and equals the conditional distribution of Θ given $\tau(U, \Theta) = t$. Posterior expectations of functions $h(\Theta)$ can now be computed by the formula (1.4) by replacing $\phi(\chi(u, \theta))$ by $h(\theta)$ which means replacing $Z_t(u)$ by the conditional expectation of $h(\Theta)$ given $\tau(u, \Theta) = t$, while $W_t(u)$ is unchanged. In particular, under the conditions of Proposition 1, the posterior expectation equals

$$E(h(\Theta)|\tau(U,\Theta) = t) = \frac{\int h(\hat{\theta}(u,t))|\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\hat{\theta}(u,t)} M(du)}{\int |\frac{f(\theta)}{\det \partial_\theta \tau(u,\theta)}|_{\theta=\hat{\theta}(u,t)} M(du)} \tag{2.23}$$

where $f(\theta)$ is the prior density for θ. The expression (2.23) is recognized as corresponding to computing the posterior expectation by importance sampling, drawing θ from the distribution of $\hat{\theta}(U,t)$. For example, in the special case of normal distributions as given in Example 1 we have $\hat{\theta}(U,t) = (\bar{x} - \frac{\bar{U}}{S_U} s, \frac{1}{S_U} s)$ and $|\det \partial_\theta \tau(u,\theta)|_{\theta=\hat{\theta}(u,t)} = S_U$. We note that the distribution of $\hat{\theta}(U,t)$ in this case corresponds to Fisher's fiducial distribution for the univariate normal model [Welsh (1996), page 92].

More generally, in the multinormal case of Example 2 we can use formula (2.23) to compute posterior expectations by using $\hat{\theta}(U,t)$ given by (2.7) and using

$$1/|\det \partial_\theta \tau(U,\theta)|_{\theta=\hat{\theta}(U,t)} \equiv |\det \partial_t \hat{\theta}(U,t)| = \prod_{i=1}^{p} (L_U^{-1})_{ii}^{p+1-i}$$

which is a consequence of Giri [(1977), Corollary 2.4.1, page 26].

It follows from (2.23) that under condition (2.10) there exists an (improper) prior density $f(\theta)$ under which the posterior distribution of Θ given $T = t$ is distributed like $\hat{\theta}(U,t)$. Moreover, the converse will also hold: If there exists a prior density $f(\theta)$ for which the posterior of Θ given $T = t$ is distributed like $\hat{\theta}(U,t)$, then (2.10) holds. Thus a sufficient condition for X_t having the correct distribution is the existence of a prior density $f(\theta)$ for which the posterior given $T = t$ is distributed like $\hat{\theta}(U,t)$.

As a consequence, if X_t does not have the correct distribution (e.g. Example 3), then there can be no (improper) prior density $f(\theta)$ such that $\hat{\theta}(U,t)$ is distributed as the posterior of Θ given $T = t$. Still, however, the distribution of $\hat{\theta}(U,t)$ may be a fiducial distribution for θ in the sense of Fisher. Suppose in particular that θ is one-dimensional and $\tau(u,\theta)$ is strictly monotone in θ for fixed u (e.g. Example 3). Then $\tau(U,\theta)$ and equivalently T is stochastically monotone in θ. The construction of fiducial distributions from such monotone distributions is well known [see e.g. Wilks (1962), page 370], and it follows that the distribution of $\hat{\theta}(U,t)$ is fiducial.

3. Theoretical results

3.1. *Conditional expectations for σ-finite measures.* As already explained, our general method for computation of conditional expectations is based on supplying the parameter space Ω by a σ-finite measure π. This leads to the need for extending the definition of conditional expectation to the case of σ-finite measures. Fortunately this is rather straightforward. The present subsection provides the basic definitions and results needed in the following subsections.

Let μ be a σ-finite measure on a measurable space $\mathcal{S}$ and let $Z : \mathcal{S} \to \mathcal{S}_Z$ be a measurable function defined on $\mathcal{S}$. In analogy with probability distributions and random variables we shall use the notion distribution of Z to mean the measure μ_Z on $\mathcal{S}_Z$ given by $\mu_Z(A) = \mu(Z \in A) = \mu\{s : Z(s) \in A\}$. Assume that μ_Z is σ-finite and that $Y : \mathcal{S} \to \mathbb{R}$ is integrable on sets with finite measure. We define the conditional expectation $\mu(Y|Z = z)$ to be any measurable function of z

such that

$$\mu(Yg(Z)) = \int \mu(Y|Z=z)g(z)\,\mu_Z(dz) \tag{3.24}$$

holds for all g in the set of indicator functions of sets with finite μ_Z-measure. Here we use the notation $\mu(X) = \int X(s)\,\mu(ds)$ for the integral. The conditional expectation $\mu(Y|Z = z)$ defined in this way is determined a.e. with respect to μ_Z. The different functions $\mu(Y|Z = z)$ satisfying (3.24) are called versions of the conditional expectations. It is nice to observe that the conditional expectation defined here is normalized even though we started with a σ-finite measure. If μ is a probability measure then our definition reduces to the ordinary definition of conditional expectation.

Conditioning on a σ-field $\mathcal{F}$ in $\mathcal{S}$ may be defined in a similar manner by the replacement of the above $g(z)$'s with indicator functions for the sets in $\mathcal{F}$, replacing μ_Z with μ, and the demand that $\mu(Y|\mathcal{F})$ is measurable with respect to $\mathcal{F}$. Existence of the conditional expectation follows in both cases from the Radon-Nikodym theorem.

The following result is a consequence of the definition (3.24). Suppose that the function h is one-to-one and measurable. Then

$$\mu(Y|Z=z) = \mu(Y|h(Z)=h(z)) \; a.e. \; [\mu_Z] \tag{3.25}$$

In most cases of interest, for example when the involved spaces are Borel spaces, the conditional expectation $\mu(Y|Z = z)$ can be given in terms of a regular conditional distribution $\mu(\cdot|Z = z)$ which is a measure on $\mathcal{S}$ for each z and is such that $\mu(A|Z = z)$ is measurable in z for each measurable set A in $\mathcal{S}$ [Schervish (1995), Theorem B.32, page 618]. With this notation we have $\mu(Y|Z = z) = \int Y(s)\,\mu(ds|Z = z)$.

For easy reference we quote next a result of Bahadur and Bickel (1968). Suppose that there is a regular conditional distribution $\mu(ds|Z = z)$ as above. Then for any measurable function $f(\cdot,\cdot)$ we have

$$\mu\{f(Y,Z)|Z=z\} = \int f(Y(s),z)\mu(ds|Z=z) \tag{3.26}$$

in the sense that the right hand side is a version of the conditional expetation on the left. Intuitively we can think of the right hand side as $\mu\{f(Y, z)|Z = z\}$, so (3.26) can be viewed as a result on substitution in conditional expectations.

3.2. *General formulae for $E(\phi(X)|T = t)$.* Let notation and assumptions be as described in the introduction. Recall that $m \equiv M \otimes \pi$, where M is the distribution of U and π is the distribution of Θ. The following result is basic to our approach.

THEOREM 1. *Assume that the distribution m_τ of $\tau(U,\Theta)$ is σ-finite. If $\phi : \mathcal{S}_X \to \mathbb{R}$ is bounded and measurable, then*

$$E(\phi(X)|T=t) = m(\phi(\chi(U,\Theta))|\tau(U,\Theta)=t) \quad a.e.[m_\tau] \tag{3.27}$$

PROOF. Let $\psi_1(t) = \mathrm{E}(\phi(X)|T = t)$ and $\psi_2(t) = m(\phi(\chi(U,\Theta))|\tau(U,\Theta) = t)$. We prove that $m(\psi_1(\tau(U,\Theta)) \neq \psi_2(\tau(U,\Theta))) = 0$. Note that by our assumptions we have that $\mathrm{E}(\phi(X)|T = t)$ is a version of $M(\phi(\chi(U,\theta)|\tau(U,\theta) = t)$ for all $\theta \in \Omega$. Let g be an indicator function of a set with finite m_τ-measure. The claim follows from

$$\begin{aligned}
m(g(\tau(U,\Theta))\psi_2(\tau(U,\Theta))) &= m(g(\tau(U,\Theta))\phi(\chi(U,\Theta))) \\
&= \int M(g(\tau(U,\theta))\phi(\chi(U,\theta)))\pi(d\theta) \\
&= \int M(g(\tau(U,\theta))\psi_1(\tau(U,\theta)))\pi(d\theta) \\
&= m(g(\tau(U,\Theta))\psi_1(\tau(U,\Theta)))
\end{aligned}$$

□

REMARK. The conclusion of the theorem can be rephrased to say that the function $E(\phi(X)|T=t)$, the existence of which is guaranteed by sufficiency, is a version of the conditional expectation on the right hand side for any measure π. On the other hand, the right hand side of (3.27) can be used as a representative for $E(\phi(X)|T=t)$ whenever π is such that the distribution of $\tau(U,\theta)$ is dominated by the distribution of $\tau(U,\Theta)$ for each $\theta \in \Omega$.

The point of this remark is most easily illustrated by an example where $X = (X_1, \ldots, X_n)$ is a vector of n i.i.d. variables uniformly distributed on $(0,\theta)$, with $\theta > 0$. Then $T = \max X_i$ is sufficient compared to X. Now for a fixed value θ_0 the support of T is $[0,\theta_0]$. Thus conditional expectations of functions $\phi(X)$ given $T = t$ under P^{θ_0} can be assigned an arbitrary value for $t > \theta_0$. We conclude from this that the right hand side of (3.27) can not be used as representative for $E(\phi(X)|T=t)$ if we choose $\pi = \delta_{\theta_0}$. $\square$

The next theorem provides our precise version of formula (1.4) of the introduction.

THEOREM 2. *Assume that the distribution m_τ of $\tau(U,\Theta)$ is σ-finite. Suppose there is a σ-finite measure μ on $\mathcal{S}_T$ such that the distribution of $\tau(u,\Theta)$ for M-almost all u is absolutely continuous w.r.t. μ with density $t \mapsto W_t(u)$.*

Let further $\phi : \mathcal{S}_X \to R$ be bounded and measurable, and let for fixed u, $t \mapsto Z_t(u)$ denote the conditional expectation of $\phi(\chi(u,\Theta))$ given $\tau(u,\Theta) = t$.

Then the functions $W_t(u)$ and $Z_t(u)$ can be chosen jointly measurable in (u,t), in which case we have

$$E(\phi(X)|T=t) = \frac{\int W_t(u)Z_t(u)M(du)}{\int W_t(u)M(du)} \equiv \frac{M(W_tZ_t)}{M(W_t)} \quad a.e.[m_\tau] \tag{3.28}$$

PROOF. Assume first that $W_t(u)$ and $Z_t(u)$ are jointly measurable. It is straightforward to show that

$$m_\tau(dt) = \{\int W_t(u)M(du)\}\mu(dt) \tag{3.29}$$

By definition of $Z_t(u)$ as a conditional expectation it follows that for any indicator function g of a set with finite m_τ-measure we have

$$\int \phi(\chi(u,\theta))g(\tau(u,\theta))\pi(d\theta) = \int Z_t(u)g(t)W_t(u)\mu(dt) \tag{3.30}$$

Integrating both sides of (3.30) w.r.t. $M(du)$ and using Fubini's theorem we get

$$m(\phi(\chi(U,\Theta))g(\tau(U,\Theta))) = \int \frac{\int W_t(u)Z_t(u)M(du)}{\int W_t(u)M(du)} g(t)m_\tau(dt) \tag{3.31}$$

Using (3.31) and the definition of conditional expectation it thus follows from Theorem 1 that (3.28) holds.

It remains to show that $W_t(u)$ and $Z_t(u)$ can be chosen jointly measurable. To see this for $W_t(u)$, let $\nu(du,dt)$ be a measure on $\mathcal{M} \times \mathcal{S}_T$ defined by

$$\int h(u,t)\nu(du,dt) = \int h(u,\tau(u,\theta))\pi(d\theta)M(du) \tag{3.32}$$

Using the fact that the distribution of $\tau(u,\Theta)$ for M-almost all u is absolutely continuous w.r.t. μ it is straightforward to show that $\nu \ll \mu \times M$. Hence by the Radon-Nikodym theorem there is a jointly measurable $W_t(u)$ such that

$$\nu(du,dt) = W_t(u)M(du)\mu(dt)$$

Substituting functions of the form $h(u,t) = h_1(u)h_2(t)$ in (3.32) it finally can be shown that the derived $W_t(u)$ has the desired property of being the density of $\tau(u,\Theta)$ w.r.t. μ for fixed u.

In order to derive a jointly measurable $Z_t(u)$ we define a signed measure $\eta(du,dt)$ by

$$\int h(u,t)\eta(du,dt) = \int \phi(\chi(u,\theta))h(u,\tau(u,\theta))\pi(d\theta)M(du) \tag{3.33}$$

It is clear that $\eta \ll \nu$, and thus there is a jointly measurable $Z_t(u)$ with

$$\eta(du, dt) = Z_t(u)\nu(du, dt) = Z_t(u)W_t(u)M(du)\mu(dt)$$

Again using functions of the form $h(u,t) = h_1(u)h_2(t)$ in the definition of η we can show that $Z_t(u)$ satisfies the defining equation of the conditional expectation $E(\phi(\chi(u,\Theta))|\tau(u,\Theta) = t)$. This concludes the proof. □

REMARK. The measure μ can be taken to be m_τ provided $\tau(u,\Theta) \ll \tau(U,\Theta)$ considered as measures on $\mathcal{S}_T$, for M-almost all u. In this case $\int W_t(u)M(du) = 1$ so we would have $E(\phi(X)|T = t) = M(W_t Z_t)$.

Note that in general we have $0 < \int W_t(u)M(du) < \infty$ *a.e.*$[m_\tau]$ because of (3.29) and the assumption of $\sigma-$finiteness of m_τ. □

3.3. *The pivotal condition.* In this subsection we assume that there is a unique solution $\hat{\theta}(u,t)$ of the equation $\tau(u,\theta) = t$ for each pair (u,t). Let X_t be as defined in (1.2). Recall from the introduction that X_t is said to have the correct distribution if it is distributed like X given $T = t$. More precisely this means that for all bounded, measurable $\phi : \mathcal{S}_X \to R$, we can take $E(\phi(X)|T = t) = M(\phi(X_t))$. In the theorem below we prove that this is the case when a pivotal condition holds (see introduction and Examples 1-3 for more informal discussions). Let notation and assumptions be as described in the introduction of the paper.

THEOREM 3. *Suppose that there exist measurable functions $r : \mathcal{M} \to \tilde{\mathcal{M}}$, $\tilde{\tau} : \tilde{\mathcal{M}} \times \Omega \to \mathcal{S}_T$ with $\tau(u,\theta) = \tilde{\tau}(r(u),\theta)$ for $u \in \mathcal{M}, \theta \in \Omega$, such that the equation $\tilde{\tau}(v,\theta) = t$ has a unique measurable solution $v = \tilde{v}(\theta,t)$ for all (θ,t) in the set $\Gamma = \{(\theta,t) : \tau(u,\theta) = t \text{ for some } u \in \mathcal{M}\}$. Assume that there is a regular conditional distribution of U given $r(U) = r$.*

Furthermore, assume that there is a σ-finite measure ν on $\mathcal{S}_T$ which dominates the distribution of $\tau(U,\theta)$ for all θ, and a σ-finite measure γ on Ω which dominates the distribution of $\hat{\theta}(U,t)$ for all t.

Then X_t has the correct distribution and X_t is independent of $r(U)$ for ν-almost all t.

PROOF. Let $\phi : \mathcal{S}_X \to R$ be bounded and measurable and let as usual $\psi(t)$ denote the function $\mathrm{E}(\phi(X)|T = t)$. By the assumptions, for each fixed θ there is a one-to-one and measurable correspondence between $r(U)$ and $\tilde{\tau}(r(U),\theta)$. By (3.25) we therefore have for each θ,

$$M(\phi(\chi(U,\theta))|\tilde{\tau}(r(U),\theta) = t) = M(\phi(\chi(U,\theta))|r(U) = \tilde{v}(\theta,t)) \tag{3.34}$$

a.e.$[\tau(U,\theta)]$. Denote the regular condtional distribution of U given $r(U) = r$ by $\delta(du|r)$. Now the right hand side of (3.34) can be represented by

$$\int \phi(\chi(u,\theta))\delta(du|\tilde{v}(\theta,t)) \equiv g(\theta,t) \tag{3.35}$$

Thus $g(\theta,t)$ is measurable and defined for all $(\theta,t) \in \Gamma$. Moreover, $g(\theta,t)$ as a function of t is a version of $M(\phi(\chi(U,\theta))|\tilde{\tau}(r(U),\theta) = t)$ and hence we have $g(\theta,t) = \psi(t)$ a.e. $[\tau(U,\theta)]$ for each θ. Define $g(\theta,t) = \psi(t)$ for $(\theta,t) \in \Omega \times \mathcal{S}_T \setminus \Gamma$.

Let $I(\theta,t) = [g(\theta,t) \neq \psi(t)]$ for all $(\theta,t) \in \Omega \times \mathcal{S}_T$. Then $I(\theta,t)$ is jointly measurable and $\int I(\theta,t)\nu(dt) = 0$ for all θ. Here we assume that ν is chosen in such a way that *a.e.*$[\tau(U,\theta)]$ for all θ implies *a.e.*$[\nu]$. This is always possible under the existence of a σ-finite dominating measure [Lehmann (1986), Theorem 2, page 575].

Fubini's theorem for nonnegative functions now gives

$$0 = \int\{\int I(\theta,t)\nu(dt)\}\gamma(d\theta) = \int\{\int I(\theta,t)\gamma(d\theta)\}\nu(dt)$$

which implies that for ν-almost all t we have $\int I(\theta,t)\gamma(d\theta) = 0$. Thus there is a measurable set $N_0 \subseteq \mathcal{S}_T$ with $\nu(N_0) = 0$ such that for all $t \notin N_0$ we have $g(\theta,t) = \psi(t)$ *a.e.*$[\gamma]$ when considered as a function of θ.

It follows from the assumptions that $\hat{\theta}(u,t)$ depends on u only through $r(u)$. We can therefore define a function $\tilde{\theta}$ such that $\tilde{\theta}(r(u),t) \equiv \hat{\theta}(u,t)$. For each fixed t we get a regular conditional distribution of U given $\tilde{\theta}(r(U),t) = \theta$ which is given by $\delta(du|\tilde{v}(\theta,t))$. This follows since for fixed t, $\tilde{\theta}(r(U),t)$ is a one-to-one and measurable function of $r(U)$. Thus (3.26) implies that $g(\theta,t)$, defined in (3.35), for each fixed t is a version of

$$M(\phi(\chi(U,\tilde{\theta}(r(U),t)))|\tilde{\theta}(r(U),t) = \theta) \tag{3.36}$$

Moreover, this implies that $\psi(t)$ is a version of (3.36) for all $t \notin N_0$. Since $\psi(t)$ is a constant function of θ we must therefore have

$$M(\phi(\chi(U,\tilde{\theta}(r(U),t)) = \psi(t)$$

for all $t \notin N_0$. Thus it follows that $M(\phi(\chi(U,\tilde{\theta}(r(U),t))))$ as a function of t is a version of $E(\phi(X)|T = t)$. Since this is the case for all bounded and measurable ϕ, we conclude that X_t has the correct distribution.

The final independence result follows from the fact that $\psi(t)$ is a version of (3.36) for all ϕ and all $t \notin N_0$. $\square$

Acknowledgments. The authors would like to thank Steinar Engen, Magnar Lillegård and Håvard Rue for helpful discussions during the work leading to this manuscript.

REFERENCES

Bahadur, R. R. and Bickel, P. J. (1968). Substitution in conditional expectation. *Annals of Mathematical Statistics*, **39**, 377-378.

Bølviken, E. and Skovlund, E. (1996). Confidence intervals from Monte Carlo tests. *Journal of the American Statistical Association*, **91**, 1071-1078.

Engen, S. and Lillegård, M. (1997). Stochastic simulations conditioned on sufficient statistics. *Biometrika*, **84**, 235-240.

Giri, N. C. (1977). *Multivariate Statistical Inference.* Academic Press, New York.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations* (Third edition). Johns Hopkins University Press, Baltimore.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (Second edition). Wiley, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag, New York.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945-966.

Ripley, B. (1987). *Stochastic Simulation.* Wiley, New York.

Rudin, W. (1987). *Real and Complex Analysis* (Third edition). McGraw-Hill, New York.

Schervish, M. J. (1995). *Theory of Statistics.* Springer, New York.

Tanner, M. A. (1996). *Tools for Statistical Inference* (Third edition). Springer, New York.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22**, 1701-1762.

Trotter, H. F. and Tukey, J. W. (1956). Conditional Monte Carlo for normal samples. *Symposium on Monte Carlo Methods.* H. A. Meyer, Ed. Wiley, New York, pp. 64-79.

Tukey, J. W. (1957). Some examples with fiducial relevance. *Annals of Mathematical Statistics*, **28**, 687-695.

Welsh, A. H. (1996). *Aspects of Statistical Inference.* Wiley, New York.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

Bo Henry Lindqvist
Department of Mathematical Sciences
Norwegian University of Science
and Technology
N-7491 Trondheim, Norway
E-mail: bo@math.ntnu.no

Gunnar Taraldsen
SINTEF Telecom and Informatics
- Acoustics
N-7465 Trondheim, Norway
E-mail: gunnar.taraldsen@informatics.sintef.no