The occluded process

Florian Maire
Department of Mathematics and Statistics, Université de Montréal
florian.maire@umontreal.ca

Coauthor(s): Max Hird

We consider the ubiquitous problem in Bayesian Statistics/ML of sampling from a posterior distribution, called π. Over the recent years, a host of optimisation-based methods (Variational Bayes, expectation propagation, normalizing flows) producing a deterministic approximation of π, called Q, have been developed. Consider a MCMC algorithm producing a Markov chain $\{X_t : t \geq 0\}$ with a π-reversible Markov kernel K. Our work aims at leveraging a deterministic approximation Q of π to enhance $\{X_t\}$.

In this work, we consider a partition $\{\mathcal{X}_i\}_{i=1}^n$ of the state-space $\mathcal{X}$, and define the conditional distributions $\pi_i = \pi(\,\cdot \cap \mathcal{X}_i)/\pi(\mathcal{X}_i)$, for $i \in \{1, \ldots, n\}$. Moreover, let $\rho : \mathcal{X} \to \{1, \ldots, n\}$ be such that for all $x \in \mathcal{X}$, $\rho(x)$ gives the element of the partition that x belongs to. Let $\phi : \mathcal{X} \to [0, 1]$. We define the ϕ-occlusion of the Markov chain $\{X_t\}$ as the stochastic process $\{Z_t : t \geq 0\}$ defined by $Z_0 = X_0$ and for all $t > 0$,

$$Z_t = \begin{cases} X_t & \text{w.p. } 1 - \phi(X_t) \\ W & \text{w.p. } \phi(X_t) \end{cases}, \qquad W \mid X_t \sim \pi_{\rho(X_t)}\,. \tag{1}$$

In other words, at some random times $\{T_1, T_2, \ldots\}$, X_{T_k} is simply occluded by an independent draw from π_i, where $i = \rho(X_{T_k})$. For example, if $\phi = 1$ and $n = 1$ then $\{Z_t\} \sim_{iid} \pi$.

The purpose of this talk is two-fold:

1. What theoretical properties does $\{Z_t\}$ inherit from $\{X_t\}$?
   - Unsurprisingly, $\{Z_t\}$ converges weakly to π. We also prove that $\{Z_t\}$ admits a CLT for any functional in $L^2(\pi)$, provided that $\{X_t\}$ is geometrically ergodic. The main challenge here is that $\{Z_t\}$ is not itself a Markov chain.
2. We present an algorithm which simulates the ϕ-occluded process $\{Z_t\}$ at no extra computational time relative to that of $\{X_t\}$.
   - As observed in a companion paper [1], a simple modification of the rejection sampling mechanism allows to sample i.i.d. random variables from the restrictions π_i using Q, at a complexity say $\rho_i \geq 1$. For any $x \in \mathcal{X}_i$, the occlusion probability is such that $\phi(x) = \mathcal{O}(1/\rho_i)$ and is thus imposed by Q.

One might think that any amount of occlusion, as small as it is, does reduce the autocorrelation of the chain and thus makes the CLT's asymptotic variance of $\{Z_t\}$ not larger than $\{X_t\}$. We present some pathological, yet informative, counter-examples to that statement. Furthermore, we explain that good choices of Q lead to an occluded process which acts as a control variate mechanism and thus improves the efficiency of $\{X_t\}$. A simulation study shows that, for common statistical models, the occlusion reduces the asymptotic variance all the more than Q approximates well π in high posterior density regions.

[1] An independent Metropolis sampler without rejection, F. Maire and F. Perron, *upcoming*.