PERFECT STOCHASTIC EM

Erik van Zwet

University of California

In a missing data problem we observe the result of a (known) many-to-one mapping of an unobservable 'complete' dataset. The aim is to estimate some parameter of the distribution of the complete data. In this situation, the stochastic version of the EM algorithm is sometimes a viable option. It is an iterative algorithm that produces an ergodic Markov chain on the parameter space. The stochastic EM (StEM) estimator is then a sample from the equilibrium distribution of this chain. Recently, a method called 'coupling from the past' was invented to generate a Markov chain in equilibrium. We investigate when this method can be used for a StEM chain and give examples where this is indeed possible.

AMS subject classifications: 62M05.

Keywords and phrases: EM algorithm, imputation, incomplete observations, perfect sampling, realizable monotonicity, stochastic monotonicity, simulation,.

1 Stochastic EM

The objective of this paper is to combine two algorithms: the stochastic EM (StEM) algorithm and perfect sampling through coupling from the past (CFTP). In the present section we describe the former and in the next section the latter algorithm. In the third section we combine the two and give examples. Finally, we present a brief review of two relevant concepts: stochastic and realizable monotonicity.

Consider the following estimation problem. Suppose that X is distributed according to a probability measure P_{θ_0}. Suppose we can observe only the result of a many-to-one mapping $Y = Y(X)$. The goal is to estimate θ_0 from observing $Y = y$. The parameter θ_0 is assumed to be in some general set Θ. This setup is sometimes called a missing data problem. Often the so-called EM algorithm (Dempster, Laird and Rubin (1977)) provides a method to find the maximum likelihood estimator of θ_0. There are two drawbacks. The first is that it is not known how many iteration steps are needed to come close enough to convergence. The other is that sometimes the E-step, computation of the conditional expectation of the likelihood given the data, is not possible.

In this latter case, the stochastic version of the EM algorithm (StEM) (Celeux and Diebolt (1986), Wei and Tanner (1990)) may be a viable alternative. For a review and large sample results see Nielsen (2000). The algorithm works as follows. Suppose that we can sample from the conditional