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 distribution, under any given θ, of the complete data given the observed data. Suppose also that the complete data maximum likelihood estimator is readily computable: $\hat{\theta}_{MLE} = M(X)$. Now,

1. Fix a $\theta(0)$ in Θ
2. Sample $X(1)$ from $P_{\theta(0)}(.|Y = y)$
3. Set $\theta(1) = M(X(1))$
4. Sample $X(2)$ from $P_{\theta(1)}(.|Y = y)$
5. Set $\theta(2) = M(X(2))$
6. ...

By iterating this procedure, we obtain a sequence $\theta(0), \theta(1), \theta(2), \ldots$. If steps $2, 4, \ldots$ are carried out using independent standard uniform random variables, the sequence $\theta(t)$ is a time homogeneous Markov chain. Under certain conditions which are investigated in Nielsen (2000) it is ergodic. If so, the algorithm converges in that the $\theta(t)$ converge in distribution to a random variable, say $\hat{\theta}$, which is distributed according to the stationary distribution of the Markov chain. Then $\hat{\theta}$ is the StEM estimate. In other words, a StEM estimate is a sample (or an average of samples) from the stationary distribution of the StEM Markov chain.

The drawbacks of stochastic EM and ordinary EM are different but similar. First of all it may be difficult, time-consuming or impossible to sample from the conditional distribution of the complete data given the observed data. Also, it is not clear in general for how long we should run the StEM chain to allow it to reach equilibrium. In this paper, however, we note that in some cases we can use a device known as *coupling from the past* (CFTP) (Propp and Wilson, 1996) to obtain a sample that is guaranteed to come from the stationary distribution of the StEM chain. In the next section we briefly explain CFTP.

2 Perfect Simulation

Consider an ergodic (i.e. irreducible and aperiodic) Markov chain $X(t)$ on a state space S and suppose we want to simulate its equilibrium distribution. Starting the chain from an arbitrary state and then running it for a very long, but finite time will generally not ensure that samples are from the stationary distribution. Recently, Propp and Wilson (1996) devised a method called coupling from the past, to produce perfect or exact samples in finite time. We closely follow Kendall and Thönnes (1999) to explain how it works.

For now, let us assume that the state space S is finite. A Markov chain $X(t)$ on S can be described by means of i.i.d. 'transition maps' $H_t : S \to S$

($t = 1, 2, \ldots$). A realization of such a transition map specifies for each state $i \in S$ in which the chain might be at time $t - 1$ where the chain will jump to at time t. If p_{ij} are the transition probabilities of the Markov chain to move from i to j then the common distribution of the H_t should be such that $P(H_t(i) = j) = p_{ij}$. From the transition maps H_t the Markov chain $X(t)$ is obtained by fixing $X(0)$ at some X_0 and setting

$$X(t) = H_t(X(t-1)).$$

Coupling from the past now works as follows. We select a time $-T < 0$ in the past and simultaneously run chains starting from each state of S from time $-T$ to time 0. The chains are coupled by using the *same* realizations of the transition maps for all of them. Hence it follows that if two chains started at different states meet, they stay together. Now we check if all chains have coalesced at time 0. If so, then the state at time 0 must be a sample from the stationary distribution. This is understood as follows. Imagine that at some time before $-T$ we also started a chain from an initial state selected according to the stationary distribution. This chain will remain in equilibrium, so in particular its state at time 0 is distributed according to the stationary distribution. However, we have arranged it so that *all* chains started at time $-T$ or earlier are in the same state at time 0, no matter which state the were in at time $-T$.

If not all chains have met, then we run chains from time $-2T$ to time 0, making sure that we use the realizations of $H_{-T+1}, H_{-T+2}, \ldots, H_0$ which were obtained earlier. If the paths still have not coalesced, we run chains from time $-4T$ and so on.

If the state space S has more than just a few elements it will not be feasible to run chains starting from all possible states.

Let us now consider finite, countable or even uncountable S. Suppose that S admits a partial ordering $\preceq$, and that there are minimal and maximal elements, $\underline{s}$ and $\overline{s}$ such that

$$\underline{s} \preceq s \preceq \overline{s}, \qquad \text{for all } s \in S.$$

Also suppose that the Markov chain is 'monotone' in that it respects the ordering in the sense that

$$H_t(s) \preceq H_t(s'), \qquad \text{for all } s \preceq s'.$$

In words, when two coupled chains are in comparable states, we insist that their subsequent states remain comparable and in the same order.

In practice, we now only need to run chains from states $\underline{s}$ and $\overline{s}$. If these two have met at time zero, then any other coupled chain started at some intermediate state s ($\underline{s} \preceq s \preceq \overline{s}$) would have been at the same state at time 0. This follows directly from our assumption that the chain respects the ordering.

It remains to verify in each application that the algorithm will almost surely terminate in finite time. To ensure a useful algorithm the time until termination should of course not be too long.

David Wilson's 'perfect simulation web page' at `http://dimacs.rutgers.edu/~dbwilson/exact.html/` provides a wealth of up-to-date information about CFTP.

3 Perfect Stochastic EM

In this section we combine the ideas from the previous two sections. We use the set-up and notation of section 1. Let us suppose that the parameter space Θ admits a partial ordering $\preceq$. Fix an arbitrary time $-T < 0$ in the past. Suppose that for $t = -T+1, -T+2, \ldots, 0$ we can construct independent collections of coupled random variables $\{X_\theta(t),\ \theta \in \Theta\}$ such that

$$(1) \qquad X_\theta(t) \sim P_\theta(.|Y = y)$$

$$(2) \qquad \theta \preceq \theta' \Rightarrow M(X_\theta(t)) \preceq M(X_{\theta'}(t)) \text{ almost surely.}$$

Recall that $M(X)$ is the complete data maximum likelihood estimator. Because of (1) we can simulate StEM chains $\theta_{-T}(-T), \theta_{-T}(-T+1), \ldots, \theta_{-T}(0)$ by fixing $\theta(-T)$ at any $\theta \in \Theta$ and setting subsequent $\theta(t) = M(X_{\theta(t-1)}(t))$. By requirement (2) it is ensured that if two coupled StEM chains are in comparable states their order will always be respected. In other words, two ordered coupled paths cannot cross.

Suppose that there are 'minimal' and 'maximal' elements $\underline{\theta}$ and $\overline{\theta}$ in Θ such that $\underline{\theta} \preceq \theta \preceq \overline{\theta}$ for all $\theta \in \Theta$. Consider two coupled StEM chains $\theta^L_{-T}(.)$ and $\theta^U_{-T}(.)$ starting at time $-T$ at $\theta^L_{-T}(-T) = \underline{\theta}$ and $\theta^U_{-T}(-T) = \overline{\theta}$. As we explained in the preceding section it suffices to check if these two chains have coalesced at time zero, i.e. if $\theta^L_{-T}(0) = \theta^U_{-T}(0)$. If so, then we have a perfect StEM estimate $\hat{\theta} = \theta^L_{-T}(0)$. If not we have to go back further in time as also described in the previous section.

We now demonstrate the perfect stochastic EM algorithm in two examples. The first example is very simple, the second is more involved.

3.1 Example 1

Suppose $X = (X_1, X_2, \ldots, X_n)$ is a vector of i.i.d. samples from $P_{\theta_0} = \text{Exp}(\theta_0)$; the exponential distribution with intensity θ_0 (i.e. reciprocal of the mean). We wish to estimate $\theta_0 \in \Theta = [\underline{\theta}, \infty)$, where $\underline{\theta} > 0$. Following our second example we remark on this slightly peculiar choice of parameter space. The complete data maximum likelihood estimator of θ_0 is $\hat{\theta}_{MLE} = M(X) = (n/\sum X_i) \vee \underline{\theta}$. Suppose we only observe

$$\tilde{X}_i = X_i \wedge C \qquad \Delta_i = 1_{\{X_i > C\}},$$

for some fixed positive constant C. Write $Y_i = (\tilde{X}_i, \Delta_i)$ for the observed data. This is the classical right censoring problem. The maximum likelihood estimator of θ_0 based on the observed data is known to be the 'occurrence' divided by the 'exposure'

$$\frac{n - \sum \Delta_i}{\sum \tilde{X}_i} \vee \underline{\theta}.$$

There is really no need to apply the StEM algorithm here. Also, we should point out that application of the ordinary EM algorithm is straightforward here. The purpose of this example is merely to explain how the perfect StEM algorithm works.

We now describe how the ordinary StEM algorithm (without CFTP) works here. When below we multiply vectors we mean coordinate-wise multiplication (mapping two vectors to one).

1. Fix $\theta(0) > 0$

2. Sample $X(1)$ from $P_{\theta(0)}(X|Y = y)$. We accomplish this by setting $X(1) = \tilde{X} + \Delta E(1)$, where $E(1) = (E_1(1), \ldots, E_n(1))$ and the $E_i(1)$ are all i.i.d. with common distribution $\mathrm{Exp}(\theta(0))$.

3. Set $\theta(1) = M(X(1)) = \frac{n}{\sum X_i(1)} \vee \underline{\theta}$

4. Sample $X(2)$ from $P_{\theta(1)}(X|Y = y)$

5. Set $\theta(2) = M(X(2))$

6. ...

Repeating this procedure, we obtain an ergodic Markov chain $\theta(0), \theta(1), \ldots$

Define $\overline{\theta} = M(\tilde{X})$. Clearly, $\overline{\theta}$ is a natural upper bound for the parameter space Θ. If we choose $\theta(0) \in [\underline{\theta}, \overline{\theta}]$ the subsequent states will remain in $[\underline{\theta}, \overline{\theta}]$. Of course $\underline{\theta}$ and $\overline{\theta}$ are minimal and maximal for $[\underline{\theta}, \overline{\theta}]$ with the usual ordering on $\mathbb{R}$.

Recall the usual ordering on $\mathbb{R}^n$: $x \leq y$ if $x_1 \leq y_1$ and $x_2 \leq y_2 \ldots$ and $x_n \leq y_n$. Note that if $x \geq y$ then $M(x) \leq M(y)$. To apply CFTP we need a collection $\{X_\theta(t) : \theta \in [\underline{\theta}, \overline{\theta}],\ t = -T+1, -T+2, \ldots, 0\}$ such that $X_\theta(t) \sim P_\theta(X|Y = y)$ while $\theta \leq \theta'$ implies $X_\theta(t) \geq X_{\theta'}(t)$. This, in turn, implies $M(X_\theta(t)) \leq M(X_{\theta'}(t))$.

For $t = -T+1, -T+2, \ldots 0$ and $i = 1, 2, \ldots, n$ generate independent

$$E_{\underline{\theta},i}(t) \sim \text{Exp}(\underline{\theta}) \quad \text{and} \quad E_i(t) \sim \text{Exp}(\overline{\theta} - \underline{\theta})$$

For all $\theta \in (\underline{\theta}, \overline{\theta}]$ define

$$E_{\theta,i}(t) = E_{\underline{\theta},i}(t) \wedge \frac{\overline{\theta} - \theta}{\theta - \underline{\theta}} E_i(t).$$

Now, $E_{\theta,i}(t) \sim \text{Exp}(\theta)$ and $\theta \leq \theta'$ implies $E_{\theta,i}(t) \geq E_{\theta',i}(t)$. Set $E_\theta(t) = (E_{\theta,1}(t), \ldots, E_{\theta,n}(t))$ and note that $\theta \leq \theta'$ implies $E_\theta(t) \geq E_{\theta'}(t)$. Define

$$X_\theta(t) = \widetilde{X} + \Delta E_\theta(t)$$

It is easy to check that we have constructed a collection $\{X_\theta(t) : \theta \in [\underline{\theta}, \overline{\theta}],\ t = -T+1, -T+2, \ldots, 0\}$ meeting our requirements.

We can now run a 'lower' chain $\theta^L_{-T}(-T), \theta^L_{-T}(-T+1), \ldots, \theta^L_{-T}(0)$ starting at $\theta^L_{-T}(-T) = \underline{\theta}$ and an 'upper' chain $\theta^U_{-T}(-T), \theta^U_{-T}(-T+1), \ldots, \theta^U_{-T}(0)$ starting at $\theta^U_{-T}(-T) = \overline{\theta}$. We check if $\theta^L_{-T}(0) = \theta^U_{-T}(0)$. If not, we repeat starting from $-2T$.

We do still need to make sure that the algorithm will terminate in finite time. It is enough to check that for some fixed $-T$ the event $\theta^L_{-T}(0) = \theta^U_{-T}(0)$ has positive probability. Well, choosing $T = 1$

$$\begin{aligned} P(\theta^L_{-1}(0) = \theta^U_{-1}(0)) &\geq P(X_{\underline{\theta},i}(0) = X_{\overline{\theta},i}(0),\ \forall i) \\ &\geq P(\text{Exp}(\underline{\theta}) \leq \text{Exp}(\overline{\theta} - \underline{\theta}))^n > 0. \end{aligned}$$

3.2 Example 2

This concludes our first example. Our next example is more involved. We know of no method to compute the maximum likelihood estimator or apply the EM algorithm. However, other stochastic approximations besides (perfect) StEM are available. This example is based on joint work with Marie-Colette van Lieshout and has appeared in more detail in van Zwet (1999).

Let X be a homogeneous Poisson process with intensity $\lambda > 0$ on a non-empty compact set $S \subset \mathbb{R}^2$, and $B = B(0,1)$ the closed unit disc centered at the origin. Then, writing $A \oplus B = \{a + b : a \in A,\ b \in B\}$, we may consider the random set

$$\mathcal{B}(X) = \cup_{x_i \in X}(x_i \oplus B).$$

The random set $\mathcal{B}(X) \cap S$ is called the *Boolean* model of discs. The points of X are called the *germs* and the associated discs are the *grains*.

From observing $Y = \mathcal{B}(X) \cap S$ we want to estimate $\lambda \in (0, \overline{\lambda}]$, where $\overline{\lambda} < \infty$. As in the previous example the parameter space $[0, \overline{\lambda}]$ is slightly peculiar. We briefly comment on this following the present example.

We can think of the germs X as the complete data and the complete data maximum likelihood estimator is $M(X) = (n(X)/|S|) \wedge \overline{\lambda}$, where $n(X)$ means the number of points of X and $|S|$ is the area of S. As usual we write $P_\lambda(\cdot|Y)$ for the conditional distribution under λ of the complete data given the observed data.

Since the grains are discs, the location of a germ is identified whenever a part of its associated grain's boundary is exposed. Therefore, the conditional distribution of X can be decomposed into a deterministic 'exposed boundary' part X^b and a stochastic 'interior' X^i of germs that cannot be identified from Y. Indeed we write $X = X^i \cup X^b$. Define

$$\begin{aligned} \mathcal{C} &= Y \setminus \mathcal{B}(X^b) \\ \mathcal{D} &= \{s \in S : (s \oplus B) \cap S \subseteq Y\}. \end{aligned}$$

In words, $\mathcal{C}$ is the part of Y which is not covered by exposed grains, and must therefore be covered by the interior grains. The set $\mathcal{D}$ describes the locations where interior points may fall such that their associated grains are not outside of Y.

Note that a natural lower bound for the parameter is $\underline{\lambda} = (|\mathcal{C}|/|B| + n(X^b))/|S|$.

The conditional distribution given Y of the exposed boundary part X^b is of course degenerate. It is not hard to show that, conditionally on Y, X^i is distributed as a Poisson point process on $\mathcal{D}$ with intensity λ, conditioned on coverage of $\mathcal{C}$ by the associated Boolean model $\mathcal{B}(X^i)$. The distribution $P_\lambda(\cdot|Y)$ is of course the convolution of the two. We note that $P_\lambda(\cdot|Y)$ involves a normalizing constant which is intractable and hence maximum likelihood estimation and the EM algorithm become impossible. In van Zwet (1999) a method based on CFTP is presented to obtain a collection of samples $\{X_\lambda : \lambda \in [\underline{\lambda}, \overline{\lambda}]\}$ such that $X_\lambda \sim P_\lambda(\cdot|Y)$ and $\lambda \leq \lambda'$ implies $X_\lambda \subseteq X_{\lambda'}$. Hence we can apply StEM and even make it perfect.

We simulated a Boolean model on the unit square, with intensity 75 and grains with radii 0.1 instead of 1. We chose $\overline{\lambda} = 100$. Figure 1 shows a run of the perfect StEM algorithm. Note that the chain is continued once a perfect sample has been found, because averaging of the subsequent samples will bring down the variance of the estimator. The figure is meant as an illustration only. Much more extensive simulations would be needed to determine how fast the algorithm terminates and how the estimator performs.

This concludes our second example.

The natural parameter space $(0, \infty)$ of the first example was artificially replaced by $[\underline{\theta}, \infty)$ $(\underline{\theta} > 0)$. We then found a natural upper bound $\overline{\theta} < \infty$. Similarly, in the the second example we introduced an artificial upper bound $\overline{\lambda} < \infty$ and found a natural lower bound $\underline{\lambda} > 0$. The reason for introducing the artificial bounds is of course the need for both maximal and minimal

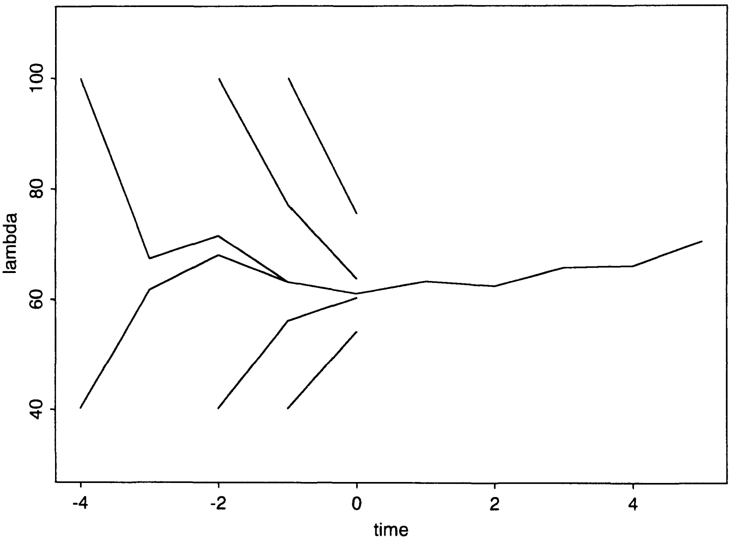

Figure 1. The upper and lower processes in the perfect stochastic EM algorithm, starting at times -1, -2 and -4.

elements of the parameter space. In practice, one would probably compute some pilot estimate of the parameter and then choose the artificial bound such that one feels confident that they do not exclude the real parameter.

4 Realizable monotonicity

In this section we review work by Fill and Machida (2000) and Ross (1993). The difficulty in making the StEM algorithm perfect lies in the construction of collections of random variables with prescribed distributions to meet condition (2). The work of Fill and Machida (2000) and of Ross (1993) makes clear when such constructions are possible—at least in principle.

The concept of *realizable monotonicity* (Fill and Machida (2000)) is essentially what is needed. Realizable monotonicity is closely related to stochastic monotonicity, which is a more familiar concept and which is generally easier to check. Fill and Machida (2000) and Ross (1993) present conditions under which stochastic monotonicity implies realizable monotonicity. The work of Fill and Machida (2000) is motivated by the relevance of realizable monotonicity for perfect sampling.

Recall that we have complete data X in some space E with distribution P_θ ($\theta \in (\Theta, \preceq)$). We observe only some function Y of X and we write $Q_\theta = P_\theta(\cdot|Y)$. The complete data maximum likelihood estimator of θ is given by a measurable function M from E to Θ. Now suppose that E

admits a partial ordering $\preceq_E$ such that, for any $x_1, x_2 \in E$

$$x_1 \preceq_E x_2 \;\Rightarrow\; M(x_1) \preceq M(x_2).$$

To apply the perfect StEM algorithm we need to be able to construct a collection $\{X_\theta,\ \theta \in \Theta\}$ such that

$$X_\theta \sim Q_\theta \tag{3}$$

$$\theta \preceq \theta' \Rightarrow X_\theta \preceq_E X_{\theta'}, \text{ almost surely} \tag{4}$$

We now review two notions of monotonicity for a collection $\{Q_\theta\ ,\ \theta \in \Theta\}$ of probability measures: realizable and stochastic monotonicity.

Definition 4.1 *Consider two partially ordered spaces $(\Theta, \preceq)$ and $(E, \preceq_E)$. The collection $\{Q_\theta\ ,\ \theta \in \Theta\}$ is called realizably monotone if there exists a collection of E-valued random variables $\{X_\theta\ ,\ \theta \in \Theta\}$ satisfying (3) and (4).*

We now turn to stochastic monotonicity. A subset U of E is said to be an *up-set* in $(E, \preceq_E)$ if $y \in U$ whenever $x \in U$ and $x \preceq_E y$. If Q_1 and Q_2 are probability measures on $(E, \mathcal{E})$ then Q_1 is *stochastically smaller* than Q_2 if $Q_1(U) \leq Q_2(U)$ for all up-sets U in $(E, \preceq_E)$. We then write $Q_1 \preceq_E^D Q_2$.

Definition 4.2 *The collection $\{Q_\theta\ ,\ \theta \in \Theta\}$ is called stochastically monotone if*

$$\theta \preceq \theta' \Rightarrow Q_\theta \preceq_E^D Q_{\theta'},\ a.s.$$

It can be easily seen that realizable monotonicity implies stochastic monotonicity. That the converse is not always true is demonstrated by means of an example in Ross (1993). However, for various finite classes of $(E, \preceq_E)$, Fill and Machida (2000) give conditions on finite index sets $(\Theta, \preceq)$ such that realizable and stochastic monotonicity are equivalent. For instance, we have equivalence whenever $(E, \preceq_E)$ or $(\Theta, \preceq)$ is a finite linearly ordered set (recall that a set is linearly ordered if each pair of elements is comparable). This and other results for finite sets are all the more useful because of the following unpublished result by Ross (1993).

Theorem 4.1 *Suppose that $(\Theta, \preceq)$ is a partially ordered set and $(E, \preceq_E)$ is a complete separable metric space with closed partial order. Then a collection $\{Q_\theta,\ \theta \in \Theta\}$ of probability measures on E is realizably monotone if and only if for every finite $\Psi \subseteq \Theta$ $\{Q_\theta,\ \theta \in \Psi\}$ is realizably monotone.*

Thus, if for some separable set with a closed partial order the results of Fill and Machida (2000) apply to check realizable monotonicity for all its finite subsets then Ross's theorem allows us to conclude realizable monotonicity for the entire infinite set. It is quite surprising that the theorem holds even for uncountable Θ.

Acknowledgements. The author very much appreciated the hospitality he enjoyed at the University of Western Australia while writing this paper. The present volume may also be a good place to acknowledge all the support in mathematics he has received from his father since primary school.

REFERENCES

Celeux, G and Diebolt, J. (1986). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly* **2** 73–82

Dempster, A.P., Laird, N.M. and Rubin D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

Fill, J.A., Machida, M. (2000?). Stochastic monotonicity and realizable monotonicity. To appear either in Annals of Applied Probability or Annals of Probability.

Kendall, W.S. and Thönnes, E. (1999). Perfect simulation in stochastic geometry. *Pattern Recognition* **32(9)** 1569–1586. Special issue on random sets.

Nielsen, S.F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, to appear.

Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** 223–252.

Ross, D.A. (1993). A coherence theorem for ordered families of probability measures on a partially ordered space. Unpublished manuscript.

Wei, G.C.G. and Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* **85** 699–704

van Zwet, E.W. (1999). *Likelihood Devices in Spatial Statistics.* PhD dissertation, University of Utrecht.

University of California
Department of Statistics
367 Evans Hall #3860
Berkeley, CA 94720-3860
USA
vanzwet@stat.Berkeley.EDU