Sequential Importance Sampling (SIS)

Liu and Chen (1998, JASA)

Probability Model and Problem

General dynamic model: (g.d.m.)

$$y_t \sim f_t(\cdot|x_t, \phi) \tag{1}$$
$$x_t \sim q_t(\cdot|x_{t-1}, \theta) \tag{2}$$
$$p_0(x_0, \phi, \theta) \tag{3}$$

Par's: $\boldsymbol{x_t} = (\phi, \theta, x_0, \ldots, x_t)$ and *Data:* $\boldsymbol{y_t} = (y_1, \ldots, y_t)$.
Posterior distributions: $\pi_t(\boldsymbol{x_t}) \equiv p(\boldsymbol{x_t}|\boldsymbol{y_t})$ (sequence of posterior distributions).

Dynamic system: SIS is valid in a more general setup than g.d.m:
Any sequence of "target distributions" $\pi_t(\boldsymbol{x_t})$.
Typical example: *missing data problem*:

$$(y_i, x_i) \sim p(y, x|\theta), \quad i = 1, \ldots, n$$

Observe y_i, but not x_i, and consider $\pi_t(\boldsymbol{x_t}) = p(\boldsymbol{x_t}|\boldsymbol{y_t})$.

Problem: Want to compute posterior integrals of the type $\int h(\boldsymbol{x_t})\ \pi_t(\boldsymbol{x_t})\ d\boldsymbol{x_t}$.

Sequential Importance Sampling

Notation: generate for each t a Monte Carlo sample S_t and weights W_t:

$$S_t = \{\boldsymbol{x_t}^{(j)},\ j = 1, \ldots, M\} \text{ and } W_t = \{w_t^{(j)},\ j = 1, \ldots, M\}$$

s.t. S_t and W_t is an "importance MC sample" to compute expectations w.r.t. π_t, i.e.

$$\frac{\sum_{j=1}^M h(\boldsymbol{x_t}^{(j)})\ w_t^{(j)}}{\sum_{j=1}^M w_t^{(j)}} \rightarrow \underbrace{\int h(\boldsymbol{x_t})\ \pi_t(\boldsymbol{x_t})\ d\boldsymbol{x_t}}_{E_{\pi_t}[h(\boldsymbol{x_t})]}$$

Importance sampling density:

$$g(\boldsymbol{x_t}) = \prod_{s=1}^t \pi_s(x_s|x_{s-1})\ p_0(x_0, \phi, \theta).$$

The i.s. density is based on products of full conditionals. There are generalizations using different $g(\cdot)$.
Note: $g(\boldsymbol{x_t}) \neq \pi_t(\boldsymbol{x_t}) = \prod_{s=1}^t \underline{\pi_t}(x_s|x_{s-1})\ \underline{\pi_t}(x_0, \phi, \theta)$.

Algorithm: Assume S_t and W_t are available.
Steps (A) and (B) describe the construction of S_{t+1} and W_{t+1}:

(A) New state $x_{t+1}^{(j)} \sim g_{t+1}(\boldsymbol{x_{t+1}}|\boldsymbol{x_t}^{(j)})$ and $\boldsymbol{x_{t+1}}^{(j)} \equiv (\boldsymbol{x_t}^{(j)}, x_{t+1}^{(j)})$

(B) Incremental weights $u_{t+1}^{(j)} = \frac{\pi_{t+1}(\boldsymbol{x_{t+1}}^{(j)})}{\pi_t(\boldsymbol{x_t}^{(j)})\, g_{t+1}(x_{t+1}^{(j)}|\boldsymbol{x_t}^{(j)})}$ and $w_{t+1}^{(j)} \equiv u_{t+1}^{(j)} \cdot w_t^{(j)}$

Initialize at $t = 0$ with $\boldsymbol{x}_0^j \sim p_0(x_0, \phi, \theta)$.

Choice of $g_t(\cdot|\cdot)$

General ("ideal choice"):

$$g_{t+1}(x_{t+1}|\boldsymbol{x_t}) = \pi_t(x_{t+1}|\boldsymbol{x_t}) \tag{4}$$

and thus

$$u_{t+1} \propto \frac{\pi_{t+1}(\boldsymbol{x_{t+1}})}{\pi_t(\boldsymbol{x_t}) \cdot g_{t+1}(x_{t+1}|\boldsymbol{x_t})} = \frac{\pi_{t+1}(\boldsymbol{x_t})}{\pi_t(\boldsymbol{x_t})} \tag{5}$$

In the g.d.m.:

$$g_{t+1} \propto f_{t+1}(\boldsymbol{y_{t+1}}|x_{t+1}) \cdot q_{t+1}(x_{t+1}|\boldsymbol{x_t}, \theta)$$

and

$$u_{t+1} \propto \frac{\overbrace{p(\boldsymbol{x_t}) \cdot p(\boldsymbol{y_{t+1}}|\boldsymbol{x_t})}^{\pi_{t+1}(\boldsymbol{x_{t+1}})}}{\underbrace{p(\boldsymbol{x_t}) \cdot p(\boldsymbol{y_t}|\boldsymbol{x_t})}_{\pi_t(\boldsymbol{x_t})}} = p(y_{t+1}|\boldsymbol{x_t}, \boldsymbol{y_t})$$

where $p(y_{t+1}|\boldsymbol{x_t}, \boldsymbol{y_t}) = \int f_{t+1}(y_{t+1}|x_{t+1}, \phi) \cdot q_{t+1}(x_{t+1}|\boldsymbol{x_t}, \theta) dx_{t+1}$

Local MC in SIS

When $\pi_{t+1}(x_{t+1}|\boldsymbol{x_t})$ is *not* available for efficient sampling (and integration, to evaluate u_{t+1}), use "local M.C.":

Generate from

$$p^*(J = j, \boldsymbol{x_{t+1}}) \propto \frac{\pi_{t+1}(\boldsymbol{x_t}^{(j)}, x_{t+1})}{\pi_t(\boldsymbol{x_t}^{(j)})} \cdot w_t^{(j)} \tag{6}$$

Sampling from p^* implements (**??**) and (**??**):

1. $(x_{t+1}|j, \boldsymbol{x_t}^{(j)}) \sim \pi_{t+1}(x_{t+1}|\boldsymbol{x_t}^{(j)})$ as in (**??**), and
2. $$p^*(j) \propto w_t^{(j)} \int \frac{\pi_{t+1}(\boldsymbol{x_t}^{(j)}, x_{t+1})}{\pi_t(\boldsymbol{x_t}^{(j)})} dx_{t+1} = w_t^{(j)} \frac{\pi_{t+1}(\boldsymbol{x_t}^{(j)})}{\pi_t(\boldsymbol{x_t}^{(j)})} = w_t^{(j)}\, u_{t+1}^{(j)} = w_{t+1}^{(j)},$$

   i.e., sampling from p^* has replaced (**??**) and (**??**).

Revised algorithm:

(O) Generate M i.i.d. samples from $(J, x_{t+1}) \sim p^*$;

(B) Estimate $w_{t+1}^{(j)}$ by frequency $\hat{f}_j \equiv \#\{J = j\}/M$.

(A) $\boldsymbol{x_{t+1}}^{(j)} \equiv (\boldsymbol{x_t}^{(j)}, x_{t+1}^{(j)})$ where $x_{t+1}^{(j)}$ are any samples x_{t+1} paired with $J = j$.