

FASIT til STK4050/9050 – Statistiske simuleringer og numeriske beregninger. Høst 2005

Oppgave 1.

(a) ...

Uten resampling kan en av sekvensene dominere i vekt.

(b) Siden $\phi_t = \phi_{t-1}$, vil det aldri bli nye ϕ -verdier. Ved resampling vil det stadig bli færre.

Dette medfører at den effektive samplingstørrelsen blir på kun 1 og dermed svært usikker.

(c) Av modellen har vi at

$$\begin{aligned} p(x_{t+1}|x_1, \dots, x_t, y_1, \dots, y_t) &= \int_{\phi_t} p(x_{t+1}|\phi_t, x_1, \dots, x_t, y_1, \dots, y_t)p(\phi_t|x_1, \dots, x_t, y_1, \dots, y_t)d\phi_t \\ &= \int_{\phi_t} N(x_{t+1}; \phi_t x_t, 1)N(\phi_t; \hat{\phi}_t, \sigma_t^2)d\phi_t \\ &= N(x_{t+1}, \hat{\phi}_t x_t, \sqrt{1 + \sigma_t^2}) \end{aligned}$$

(d) I dette tilfellet vil simuleringene av ψ -ene ikke være låst i forhold til de verdier som har vært simulert tidligere og kan dermed variere fritt. Videre har den fordel av at en simulerer i et lavere dimensjonalt rom.

En har at

$$E[\psi_t|y_1, \dots, y_t] = E^{x_{1:t}}[E^\psi[\psi_t|x_1, \dots, x_t, y_1, \dots, y_t]] = E^{x_{1:t}}[\hat{\psi}_t] \approx \frac{1}{M} \sum_{j=1}^M \hat{\psi}_t^j$$

Svarer til en Rao-Blackwellisering mhp ψ og gir mindre varians enn hvis en baserte seg på simulering.

Oppgave 2.

(a) Vi har at $A(\mathbf{x}, \mathbf{y}) = T(\mathbf{x}, \mathbf{y}) = r_B(\mathbf{x}, \mathbf{y})$ og $r_B(\mathbf{x}, \mathbf{y}) > 0$ for alle $\pi(\mathbf{y}) > 0$. Siden $T(\mathbf{x}, \mathbf{y}) > 0$ for alle \mathbf{x}, \mathbf{y} , har vi at $A(\mathbf{x}, \mathbf{y}) > 0$ for alle \mathbf{x}, \mathbf{y} som gir både irreducibilitet og aperiodisitet. Videre har vi at

$$\begin{aligned} \pi(\mathbf{x})A(\mathbf{x}, \mathbf{y}) &= \pi(\mathbf{x})T(\mathbf{x}, \mathbf{y}) \frac{\pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})T(\mathbf{x}, \mathbf{y}) + \pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})} \\ &= \pi(\mathbf{y})T(\mathbf{y}, \mathbf{x}) \frac{\pi(\mathbf{x})T(\mathbf{x}, \mathbf{y})}{\pi(\mathbf{x})T(\mathbf{x}, \mathbf{y}) + \pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})} \\ &= \pi(\mathbf{y})A(\mathbf{y}, \mathbf{x}) \end{aligned}$$

Dermed tilfredsstillers den også “detailed balance”, og den generelle Markovkjede teorien sier da at $\{\mathbf{x}_t\}$ er en Markov kjede som konvergerer (i fordeling) mot $\pi(\mathbf{x})$.

(b) En kan estimere $\mu_f = E_\pi f$ ved

$$\hat{\mu}_f = \frac{1}{N - M} \sum_{t=M+1}^N f(\mathbf{x}_t)$$

Den generelle Markovkjede teorien tilsier at $\hat{\mu}_f$ konvergerer mot μ_f når $N \rightarrow \infty$. En bør kaste bort de første M iterasjonene som svarer til “burnin”, antall iterasjoner før konvergens er oppnådd.

En Markovkjede som konvergerer raskt behøver ikke å bruke så stor M . I tillegg vil korrelasjonsstrukturen i $\{\mathbf{x}_t\}$ påvirke estimatet (jo større positiv korrelasjon, jo større varians). En ønsker derfor også en Markovkjede som har minst mulig korrelasjon.

(c) Vi har at

$$r_B(\mathbf{x}, \mathbf{y}) = \frac{\pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})T(\mathbf{x}, \mathbf{y}) + \pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})} \leq \frac{\pi(\mathbf{y})T(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})T(\mathbf{x}, \mathbf{y})}$$

Videre er alltid $r_B(\mathbf{x}, \mathbf{y}) \leq 1$ som tilsier at $r_M(\mathbf{x}, \mathbf{y}) \geq r_B(\mathbf{x}, \mathbf{y})$.

Siden $A_M(\mathbf{x}, \mathbf{y}) = T(\mathbf{x}, \mathbf{y})r_M(\mathbf{x}, \mathbf{y})$ og $A_B(\mathbf{x}, \mathbf{y}) = T(\mathbf{x}, \mathbf{y})r_B(\mathbf{x}, \mathbf{y})$ for $\mathbf{y} \neq \mathbf{x}$, gir dette at $A_M(\mathbf{x}, \mathbf{y}) \geq A_B(\mathbf{x}, \mathbf{y})$ og dermed er Metropolis-Hastings algoritmen mer effektiv enn Barkers algoritme.

Siden de to algoritmene bruker samme forslagsfordelinger, er det rimlig at den som gir størst sannsynlighet for aksept er den som fungerer best.

SLUTT