8.6

Towards Markov Chain Monte Carlo Methods for Uncertainty Models Specified by Sets of Probability Distributions

D. Krpelík, L. Aslett, and F. Coolen from the Durham University, UK

$$\mathcal{M} := \{P_\theta : \theta \in \Theta\} \rightsquigarrow \text{how to estimate } \underline{\mathbb{E}}_{\mathcal{M}} f \text{ ?} \quad \left(:= \inf_{P \in \mathcal{M}} \mathbb{E}_P f\right)$$

Monte Carlo estimation:

IP

$X_1, \dots, X_n \overset{iid}{\sim} P_\theta \Rightarrow \mathbb{E}f \approx \hat{\mathbb{E}}f := \frac{1}{N}\sum f(x_i)$

Draw samples via
- An ancillary RV $u \sim Q$
- Transformation function Φ_P s.t. $X = \phi_P(u) \sim P_\theta$

Set of samples $S_P(\bar{u}) := [\Phi_P(u_i)]$

The estimator

$$\hat{\mathbb{E}}f = \frac{\sum f(S_P)}{|S_P|}$$

With $\mathcal{M}$:

¡ No clear notion of samples!

If $\exists \Phi_P \; \forall P \in \mathcal{M}$
- sample common $\bar{u}$
- $\mathcal{S}_{\mathcal{M}} := \{S_P(\bar{u}) : P \in \mathcal{M}\}$

The estimator $\hat{\underline{\mathbb{E}}}f = \inf_{S \in \mathcal{S}_{\mathcal{M}}} \frac{\sum f(S)}{|S|}$

We present a method for which $\mathcal{S}_{\mathcal{M}}$ is finite.

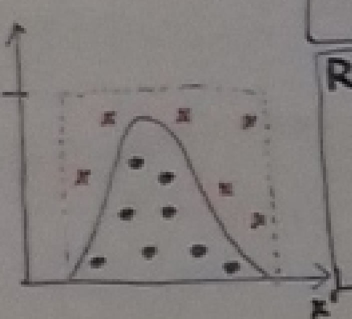
(Avoids non-linear optimisation)

¿ But how to keep track of the credal set splitting numerically?

Sometimes it is „easy"

Rejection Sampling:

IP



Initialize
$S = [\,]$

Repeat
sample x', u (~Unif)
$u < \frac{p(x')}{G} \Rightarrow S \leftarrow S + [x']$
$u > \frac{p(x')}{G} \Rightarrow S \leftarrow S$

With $\mathcal{M}$:

Initialize
$\mathcal{S}^* = \{(\mathcal{M}, S = [\,])\}$

Repeat
$\forall (\mathcal{M}, S) \in \mathcal{S}^*$

$\left.\begin{array}{l} \mathcal{M}_1 := \{P \in \mathcal{M} : u < \frac{p_P(x')}{G}\} \\ \mathcal{M}_2 := \{P \in \mathcal{M} : u > \frac{p_P(x')}{G}\} \end{array}\right\}$ branching

$(\mathcal{M}_1, S + [x'])$

$(\mathcal{M}_2, S)$

$\mathcal{M}_1 = \emptyset$ ... do nothing
$\mathcal{M}_2 = \emptyset$ ... $S \leftarrow S + [x']$
else ... $\mathcal{S}^* \leftarrow \mathcal{S}^* - (\mathcal{M}, S) \cup (\mathcal{M}_1, S + [x']) \cup (\mathcal{M}_2, S)$

Metropolis-Hastings algorithm:

(sample from p_θ)

IP

Initialize
$S = [x_0]$

Repeat
- sample $u, v \sim \text{Unif}$
- $X = \Phi(x_{-1}, u) \sim q(x' | x_{-1})$
- $A = \frac{p_\theta(x')\, q(x | x')}{p_\theta(x)\, q(x' | x)}$

$\begin{cases} V < A \Rightarrow S \leftarrow S + [x'] \\ V > A \Rightarrow S \leftarrow S + [x_{-1}] \end{cases}$

Process
$S^* = S_{M:}$ (burn-in phase)
$\hat{S} = S^*_{:K:}$ (subselection to remove correlations)
$\hat{\mathbb{E}}f = \frac{\sum f(\hat{S})}{|\hat{S}|}$ (estimation)

Metropolis-Hastings scheme uses rejection sampling $\Rightarrow$ Branching

- each $P \in \mathcal{M}$ is specified by p_θ $\Rightarrow$ induces $A_\theta(x, x')$
- if all $P \in \mathcal{M}$ agree upon $V < A$ or $V > A$ proceed as in the precise case

else $\mathcal{M}_1 := \{P \in \mathcal{M} : V < A_P\}$
$\mathcal{M}_2 := \{P \in \mathcal{M} : V > A_P\}$

$\mathcal{S}^* \leftarrow \mathcal{S}^* - (\mathcal{M}, S) \cup (\mathcal{M}_1, S + [x']) \cup (\mathcal{M}_2, S + [x_{-1}])$

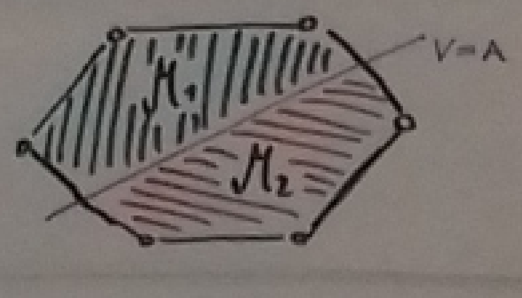
For p_θ from the exponential family

$p_\theta(x) = h(x) g(\theta) \exp(\eta(\theta) T(x))$

$\Rightarrow A_\theta(x, x') = \frac{p_\theta(x')\, q(x|x')}{p_\theta(x)\, q(x'|x)} = \frac{h(x')\, g(\theta) \exp(\eta(\theta) T(x'))\, q(x|x')}{h(x)\, g(\theta) \exp(\eta(\theta) T(x))\, q(x'|x)}$

$\Rightarrow V < A \Longleftrightarrow$
$\ln(V) - \underline{H(x, x')} < \eta(\theta)\,[T(x') - T(x)]$

The splitting boundary is a hyperplane.
$\Rightarrow$ we only need to keep track of the extreme points in $\eta(\Theta)$.



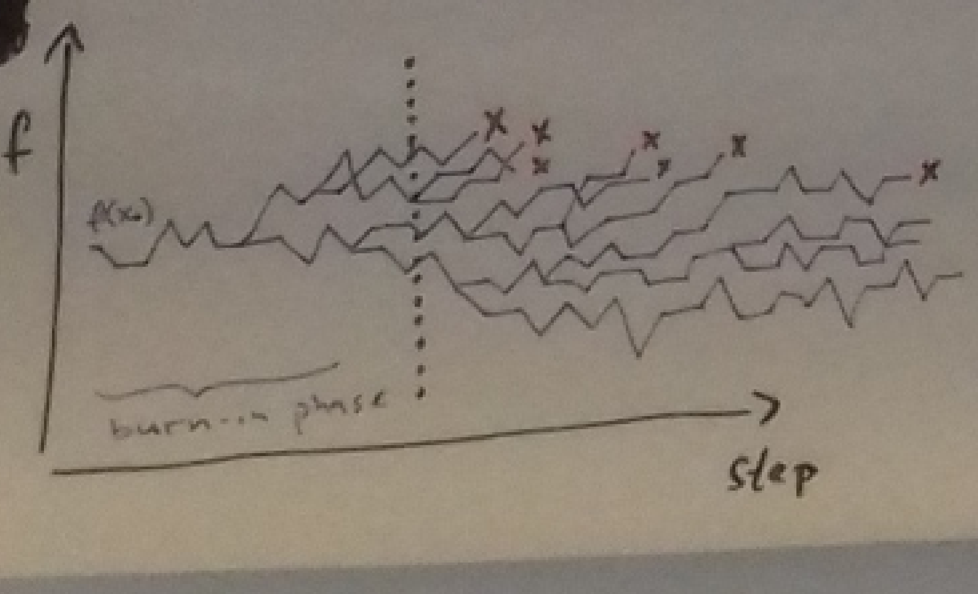
Taming the branches

Postponing branching:

for a transition kernel K
$Kp = p \Rightarrow (\alpha K + (1 - \alpha) I) p = p$

- if we randomly repeat the last sample we still target the same distribution.
- (Heuristic) we do this a couple of times when the chain would branch in order to decrease the number of branches.

"A cartoon of the evolution of a branching chain."



Terminating branches:

- we want to avoid wasting resources on unpromising branches.
- use a hypothesis test

$H_0 : \underline{\mathbb{E}}_{\mathcal{M}_i} f = \underline{\mathbb{E}}_{\mathcal{M}_{min}} f$

$H_A : \underline{\mathbb{E}}_{\mathcal{M}_i} f > \underline{\mathbb{E}}_{\mathcal{M}_{min}} f$

Conclusions:

- We optimise only over a finite set to obtain the final $\hat{\underline{\mathbb{E}}}f$.
- Naïvelly parallelizable.
- $P(\text{„This will eventually work well"}) > 0.5$

Future work:

- Theoretical justification for the heuristic patches.
- Resources allocation via sequential D-M.

UTOPIÆ

Funded by the EU's H2020 programme throught the UTOPIAE Marie Curie Inovative Training Network, H2020-MSCA-ITN-2016
Grant Agreement no. 722734