Bayesian C-DF

08.11.2018

Academic Seminar

Wonjai Lee

Reference

Online Learning

Online Learning ⟺ Batch Learning

- Sequential
- Simultaneously

- Subsequently
- Calculate at Once

Online Learning

Key Point

- By replacing the true Posterior distribution

1. Update approximate posterior
2. Optimal projection into parametric family (Choosing it to be Gaussian)

- Simultaneously

Bayesian Inference

In terms of Epistemic Uncertainty

$$p(\theta|D_t) = \frac{p(\theta)P(D_t|\theta)}{\int d\theta' p(\theta')P(D_t|\theta')}.$$

Bayesian Inference

MCMC(Markov Chain Monte Carlo)

$$p(\theta|D_t) = \frac{p(\theta)P(D_t|\theta)}{\boxed{\int d\theta' p(\theta')P(D_t|\theta')}} \Rightarrow \text{Hard to Calculate}$$

Bayesian Inference

MCMC(Markov Chain Monte Carlo)

1. Metropolis Algorithm

Bayesian Inference

MCMC(Markov Chain Monte Carlo)

2. Gibbs Algorithm

Bayesian C-DF

Advantages

1. Simultaneously
2. Serves numerous desirable feature
3. High Dimensional compressed regression

Bayesian C-DF

Existed Methods

- ADF(Assumed Density Filtering)
- EP(Expectation Propagation)
- PL(Particle Learning)
- SMC(Sequential Monte Carlo)

Bayesian C-DF

SCSS(Surrogate Conditional Sufficient Statistics)

- Compare to CSS(Conditional Sufficient Statistics)

Bayesian C-DF

Results

	Avg. coverage $\boldsymbol{\beta}$	Length	Time (sec)	MSE $= \sum_{j=1}^{p} (\hat{\beta}_t - \beta_0)^2 / p$		
				$t = 200$	$t = 400$	$t = 500$
C-DF	1.0	$0.60_{0.01}$	$95_{4.12}$	$0.27_{0.001}$	$0.15_{0.001}$	$0.06_{0.001}$
SMCMC	1.0	$0.60_{0.01}$	$119.4_{4.64}$	$0.12_{0.001}$	$0.08_{0.001}$	$0.04_{0.001}$

Table 1: Inferential performance for C-DF and SMCMC for parameters of interest. Coverage and length are based on 95% credible intervals and is averaged over all the β_j's ($j = 1, .., 5$) and all time points and over 10 independent replications. We report the time taken to produce 500 MCMC samples with the arrival of each data shard. MSE along with associated standard errors are reported at different time points.

Bayesian C-DF

Results

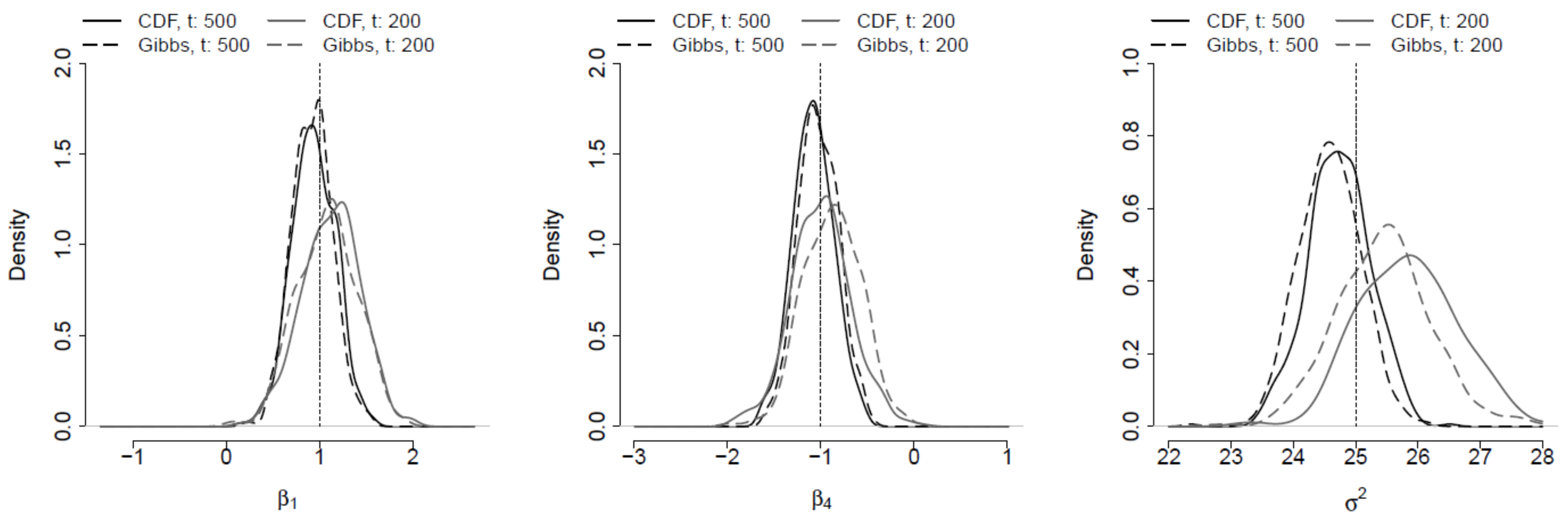

Figure 1: Kernel density estimates for posterior draws using SMCMC and the C-DF algorithm at $t = 200, 500$. Shown from left to right are plots of model parameters β_1, β_4, and σ^2, respectively.

Bayesian C-DF

Results

	Avg. coverage ζ	Length	Time (sec)	MSE $= \sum_{l=1}^{k}(\hat{\zeta}_t - \zeta_0)^2/k$		
				$t = 200$	$t = 400$	$t = 500$
C-DF	$0.87_{0.09}$	$0.53_{0.005}$	$85.0_{5.25}$	$0.77_{0.22}$	$0.29_{0.11}$	$0.26_{0.15}$
SMCMC	$0.92_{0.10}$	$0.52_{0.002}$	$119.4_{8.43}$	$0.41_{0.05}$	$0.22_{0.14}$	$0.20_{0.16}$
ADF	$0.36_{0.23}$	$0.80_{0.02}$	$0.88_{0.01}$	$0.42_{0.18}$	$0.28_{0.12}$	$0.27_{0.11}$

Table 2: Inferential performance for C-DF, SMCMC, and ADF for parameter ζ. Coverage is based on 95% credible intervals averaged over all time points, all ζ and over 10 independent replications. We report the time taken to produce 500 MCMC samples with the arrival of each data shard. MSE along with associated standard errors are reported at different time points.

Bayesian C-DF

Results

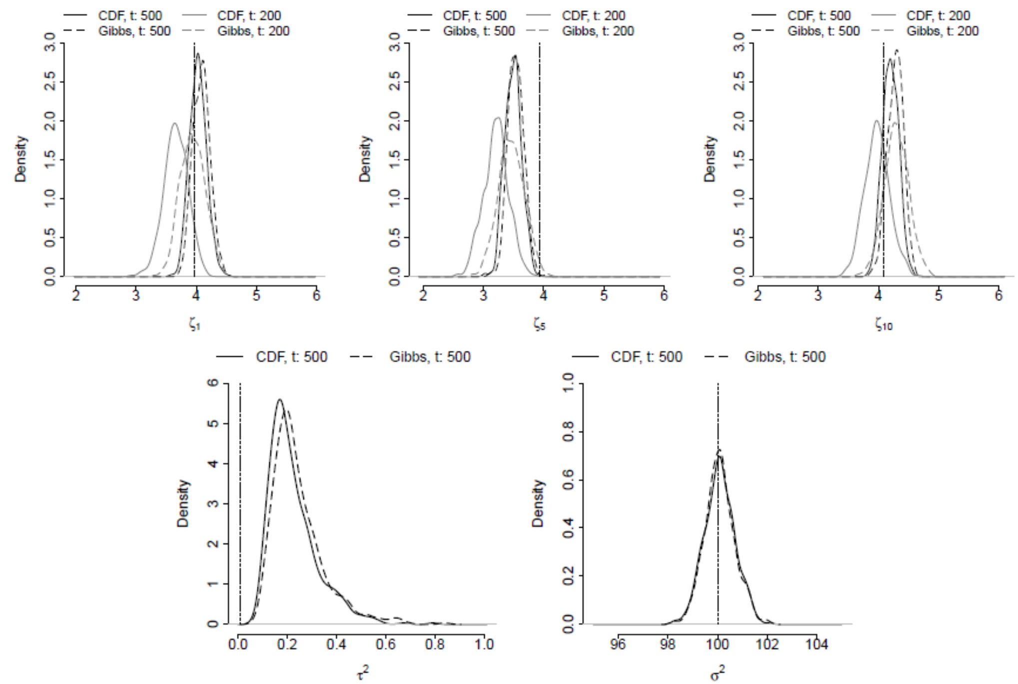

Figure 2: Row #1 (left to right): Kernel density estimates for posterior draws of ζ_1, ζ_5, ζ_{10} using SMCMC and the C-DF algorithm at $t = 200, 500$; Row #2 (left to right): Kernel density estimates for model parameters τ^2, and σ^2 at $t = 500$.

Bayesian C-DF

Results

	Avg. coverage $\boldsymbol{\theta}$	Length	Time (sec)	MSE
C-DF	$0.78_{0.10}$	$0.33_{0.11}$	$1138.60_{0.10}$	$0.011_{0.001}$
PL	$1_{0.00}$	$3.36_{0.46}$	$1750.58_{0.10}$	$0.096_{0.027}$

Table 3: Inferential performance for C-DF and Particle Learning. Coverage and length are based on 95% credible intervals for θ_t averaged over all time points and 10 independent replications. For truth θ_{t0} at time t, we report MSE $= \frac{1}{Tn}\sum_{t=1}^{Tn}(\hat{\theta}_t - \theta_{t0})^2$. We report the time taken to run C-DF with 50 Gibbs samples at each time for $\tau^2, \boldsymbol{\theta}, \sigma^2$ and 500 MH samples for ϕ.

Bayesian C-DF

Results

	Stats	Data	Sample complexity	Update complexity	Memory (bytes)
C-DF	C_i^t	$\{y_i\}_{i>nt-b}$	$S(N+G)$	N	128
PL	$C_{i,j}^t$	$\{y_i\}_{i\geq 1}$	NG	N	3330

Table 4: Computational and storage requirements for the Dynamic Linear Model using C-DF and PL. $C_{i,j}^t$, is the i-th CSS corresponding to the j-th particle in PL, $i = 1 : 4$, $j = 1 : N$, $N = 100$ is the number of particles propagated by PL, and $G = 500$ is the number of Metropolis samples used by both PL and C-DF. Memory in terms of RAM used to store and propagate SCSS and CSS for C-DF and PL is reported. Sampling and update complexities are in terms of big-O.

Bayesian C-DF

Results

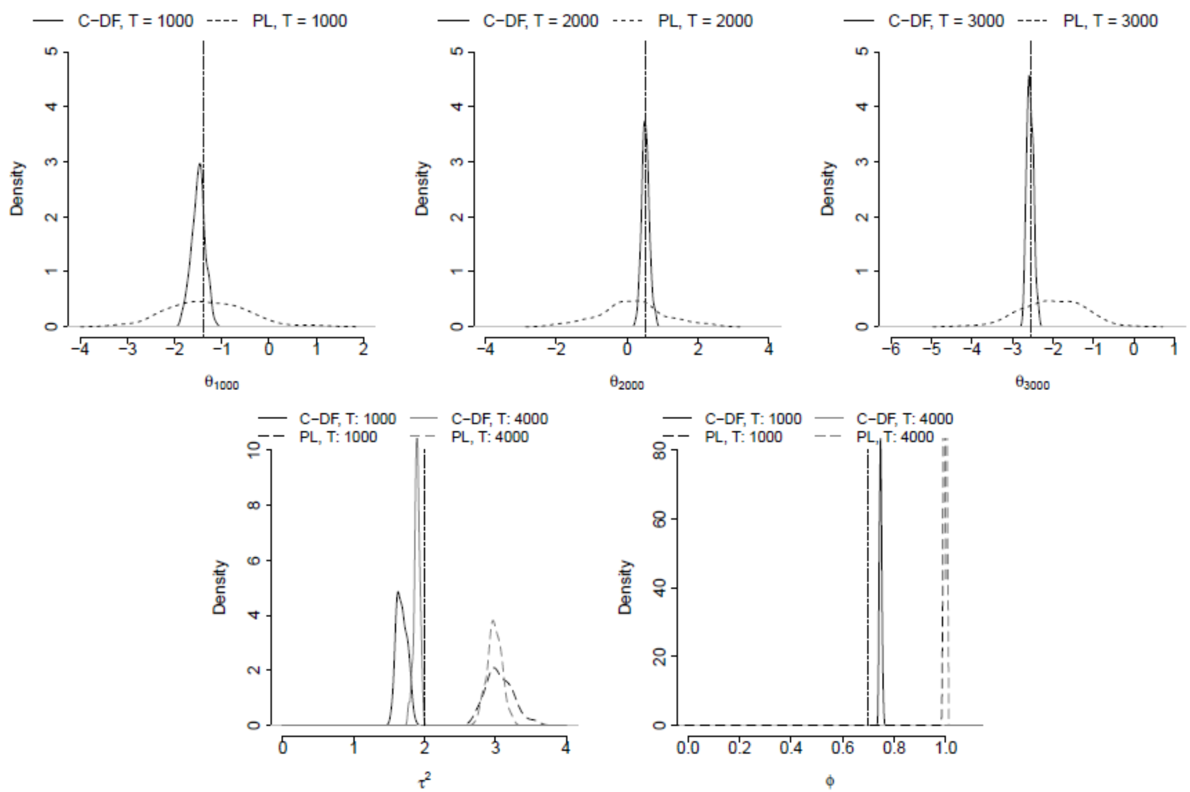

Figure 3: Row #1 (left to right): Kernel density estimates for posterior draws of θ_t using PL and the C-DF algorithm at $t = 1000, 2000, 3000$; Row #2 (left to right) plots of model parameters τ^2 and ϕ, respectively.

Thank you!