

## Lecture 9:

Topics:

- I) Gibbs sampling  $P(T, C_e | \text{data})$
- II) Sufficient vs Ancillary parameterizations
- III) Alternating vs Interweaving
- IV) Rao-Blackwellization
- V) Primordial Dipole example.

### Gibbs sampling $P(T, C_e | \text{data})$

Now we will study the full Gibbs algorithm for sampling from the posterior

$$(4) \quad P(T, C_e^{TT} | d)$$

where  $d(\hat{n}) = T(\hat{n}) + n(\hat{n})$ .

We also will discuss the  $\Lambda\text{CDM}$  posterior, joint with the map  $T(\hat{n})$ :

$$(4A) \quad P(T, \theta | d)$$

where  $\theta := (\Omega_b h^2, \Omega_m h^2, \dots, n_s)$ .

For the posterior (4) the standard Gibbs implementation produces a sequence of pairs of maps &  $C_e^{TT}$ 's:

$$(T^{(1)}, C_e^{TT(1)}), (T^{(2)}, C_e^{TT(2)}), \dots$$

where

$$T^{(i+1)}(\hat{n}) \sim P(T | C_e^{TT(i)}, d)$$

$$C_e^{TT(i+1)} \sim P(C_e^{TT} | T^{(i+1)}, d) \quad \text{drop } d$$

(1)

when observing the full sky with stationary noise and one uses Jeffrey's prior  $P(C_e^{TT}) \propto \frac{1}{C_e^{TT}}$  the Gibbs steps are given explicitly by:

(2)

$$\begin{aligned} T_{em}^{(i+1)} &= \frac{C_e^{TT(i)}}{C_e^{TT(i)} + C_e^{nn}} z_{em} + \sqrt{\frac{\frac{C_e^{TT(i)}}{C_e^{TT(i)} + C_e^{nn}}}{C_e^{TT(i)} + C_e^{nn}}} z_{em}^{(i)} \\ C_e^{TT(i+1)} &\sim C_e^{TT} \exp \left( -\frac{2\ell+1}{z} \frac{\sigma_e^{(i+1)}}{C_e^{TT}} \right) \end{aligned}$$

probability density  
of  $P(C_e^{TT} | T^{(i+1)})$

$$\text{where } \sigma_e^{(i+1)} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |T_{em}^{(i+1)}|^2$$

and  $z_{em}^{(i)}$ 's are Complex Gaussian, independent across  $i$  which satisfy

$$E(z_{em} \overline{z}_{e'm'}) = \delta_{ll'} \delta_{mm'}$$

$$z_{l,-m} = (-1)^m \overline{z}_{l,m}$$

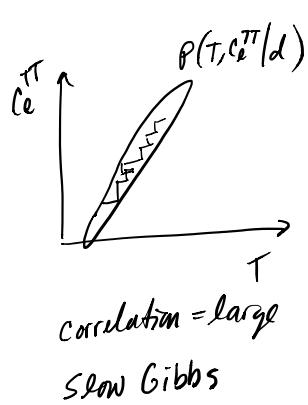
**Note:**  $C_e^{TT(i+1)}$  above is recognised as an InvGamma ( $\alpha = \frac{2\ell+1}{z}$ ,  $\beta = \frac{2\ell+1}{z} \sigma_e^{(i+1)}$ )

## Slow mixing

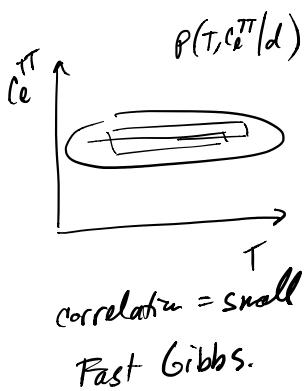
(3)

The most common difficulty with Gibbs is that it is often slow to mix/converge. Heuristically this happens when the  $(T, C_e^{\text{TT}})$  are highly correlated or dependent in the posterior.

case 1:



case 2:



One way to recognise this is when one (or both) of the Gibbs conditionals has variance much smaller than the marginal posterior variance.

e.g. For high  $\ell$ , the Inv Gamma conditional gives

$$E(C_e^{\text{TT}} | T, d) = \frac{\beta}{\alpha-1} = \frac{\ell + \frac{1}{2}}{\ell - \frac{1}{2}} \sigma_e^2$$

$$\begin{aligned} \text{var}(C_e^{\text{TT}} | T, d) &= \left(\frac{\beta}{\alpha-1}\right)^2 \frac{1}{\alpha-2} \\ &= \left(\frac{\ell + \frac{1}{2}}{\ell - \frac{1}{2}}\right)^2 \frac{\sigma_e^2}{\ell - \frac{3}{2}} \\ &\sim \frac{\sigma_e^2}{\ell} \end{aligned}$$

$\therefore \text{var}(C_e^{\text{TT}} | T, d) \rightarrow 0$  as  $\ell \rightarrow \infty$

(4)

but the beam implies

$\text{var}(C_e^{\text{TT}} | d) \rightarrow \text{prior var of } C_e^{\text{TT}}$

as  $\ell \rightarrow \infty$ .

$\therefore$  for high  $\ell$  this gibbs doesn't mix well.

e.g. For low  $\ell$  one often has high signal to noise:  $\frac{C_e^{\text{TT}}}{C_e^{\text{nn}}} \gg 1$ .

This implies

$$\text{var}(T_{\text{em}} | C_e^{\text{TT}}, d) \approx \text{small}$$

However cosmic variance makes

$$\text{var}(C_e^{\text{TT}} | d) \text{ large.}$$

$\therefore \text{var}(T_{\text{em}} | d)$  should also be large.

$\therefore$  @ low  $\ell$  Gibbs doesn't mix well either.

These examples illustrate the heuristic that Gibbs is slow if one of two variables is highly informative for the other variable i.e. for low  $\ell$

$$\text{var}(T_{\text{em}} | d) \gg \text{var}(T_{\text{em}} | C_e^{\text{TT}}, d)$$

$\uparrow$   
This carries  
a lot of constraint  
to  $T$

The typical way to fix this is by reparametrization.

## Sufficient vs Ancillary parameterizations (5)

To study how one reparameterizes lets work with a toy model.

$$d = \varphi + \lambda + n$$

↑      ↑      ↗  
two      two      noise  
parameters.      parameters.

where  $\varphi \sim N(0, \Sigma)$

$$\begin{aligned} & \lambda \sim N(0, \Delta) \\ & n \sim N(0, N) \end{aligned} \quad \left. \begin{array}{l} \\ \end{array} \right\} \text{priors.}$$

are indep.

Suppose you're interested in estimating  $\varphi$  &  $\lambda$  is a nuisance parameter. So  $\lambda$  acts like extra additive noise.

Noise.

Case 1:  $\Delta = \text{large}$  &  $N = \text{small}$ .

$\text{var}(\varphi | d) = \text{large}$  since  $\lambda$  acts like a large noise corruption.

If we are given  $\lambda$  we can remove the " $\lambda$  noise" from  $d$  and get much tighter constraints on  $\varphi$ .

$$\therefore \text{var}(\varphi | \lambda, d) \ll \text{var}(\varphi | d)$$

$\therefore$  SLOW GIBBS.

Case 2:  $\Delta = \text{small}$  &  $N = \text{large}$  (6)

Now given  $\lambda$  we can still de-noise  $d$  by  $d - \lambda$ , but it doesn't help much

$$\therefore \text{var}(\varphi | \lambda, d) \approx \text{var}(\varphi | d)$$

$\therefore$  FAST GIBBS (at least for this step)

Now consider a New parameterization

$$d = \varphi + \tilde{\lambda} + n$$

↗      ↗  
         $\tilde{\lambda}$

Original Parameterization :  $(\varphi, \lambda)$

New parameterization :  $(\varphi, \tilde{\lambda})$

The prior transforms to

$$\begin{aligned} P(\varphi, \tilde{\lambda}) &= P(\tilde{\lambda} | \varphi) \underbrace{P(\varphi)}_{\sim N(\varphi, \Delta)} \\ &\sim N(\varphi, \Delta) \sim N(0, \Sigma) \end{aligned}$$

Case 1:  $\Delta = \text{large}$  &  $N = \text{small}$  discard

$$\text{var}(\varphi | d) \approx \underbrace{\text{var}(\varphi | \tilde{\lambda}, d)}_{\text{still large}}$$

$\tilde{\lambda}$  carries all the info for  $\varphi$  so  $d$  can be discarded.  
(i.e.  $\tilde{\lambda}$  is sufficient for  $\varphi$ )

$\therefore$  given  $\tilde{\lambda}$  you get to

"de-noise  $d$  by removing  $n$ ".  
but  $\text{var}(n)$  was small &  
 $\tilde{\lambda} \sim N(\varphi, \Delta)$  so it's not  
very informative large to be given  $\tilde{\lambda}$ .

$\therefore$  FAST GIBBS

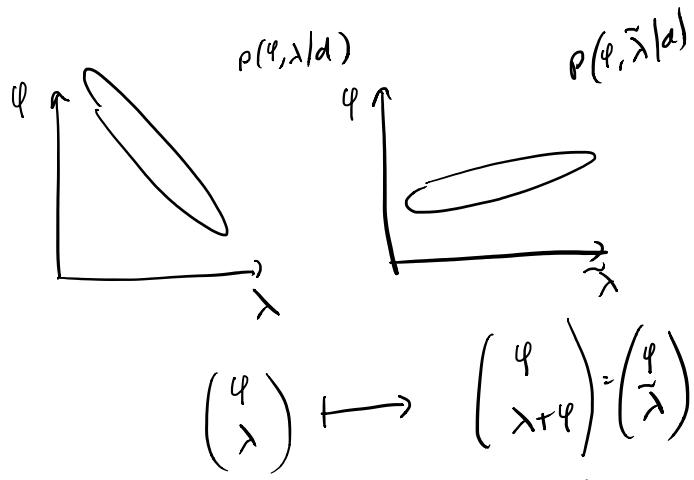
Case 2:  $\lambda$  small &  $N = \text{large}$  (7)

$$\text{var}(\varphi | d) \gg \text{var}(\varphi | \tilde{\lambda}, d)$$

since conditioning  
of  $\tilde{\lambda}$  allows you  
to "remove tons  
of n noise and"  
 $\tilde{\lambda}$  is informative.

$\therefore$  SLOW GIBBS

This Reparametrization is  
visualized as



Also note, the higher correlation  
in  $p(\varphi, \lambda | d)$  leads to  
almost near independence in  
 $p(\varphi, \tilde{\lambda} | d)$ . So the slower  
 $p(\varphi | \lambda, d)$  is, the faster  
 $p(\varphi | \tilde{\lambda}, d)$  is.

( $\varphi, \lambda$ ) is called a Ancillary Parameterization (8)

$$d = \varphi + \lambda + n, \quad \varphi \perp \lambda$$

$\sim \lambda$  is Ancillary for  $\varphi$ .

( $\varphi, \tilde{\lambda}$ ) is called a Sufficient Parameterization

$$d = \tilde{\lambda} + n, \quad \tilde{\lambda} \sim N(\varphi, \lambda)$$

$\sim \tilde{\lambda}$  is sufficient for  $\varphi$

Alternating Sufficient & Ancillary chains

If your gibbs conditionals exhibit  
problems in both parameterizations  
you can alternate btwn the two

Alternating Gibbs:

for  $i = 1, 2, \dots$

$$\begin{aligned} \tilde{\lambda}^{(i+1)} &\sim P(\tilde{\lambda} | \varphi^{(i)}, d) \\ \varphi &\sim P(\varphi | \tilde{\lambda}^{(i+1)}, d) \end{aligned}$$

$$\begin{aligned} \lambda^{(i+1)} &\sim P(\lambda | \varphi^{(i)}, d) \\ \varphi^{(i+1)} &\sim P(\varphi | \lambda^{(i+1)}, d) \end{aligned}$$

end

The mixing rate will be the  
minimum of the two individual  
rates.

$$p(z | T, d)$$

initial cosine  
invariant

## Application to CMB Gibbs

(9)

$$(S) d_{em} = T_{em} + n_{em}$$

$$\text{where } C_e^{TT} \sim \pi(C_e^{TT})$$

$$T_{em} \stackrel{iid}{\sim} N(0, C_e^{TT})$$

Note  $T_{em}$  is sufficient for  $C_e^{TT}$  so  $(T_{em}, C_e^{TT})$  is effectively a sufficient parameterization:  $\begin{matrix} T_{em}, & C_e^{TT} \\ \text{III} & \text{III} \\ \times & \times \end{matrix}$

One can switch to an ancillary parameterization as follows.

$$(A) d_{em} = \sqrt{C_e^{TT}} z_{em} + n_{em}$$

where  $z_{em}$  satisfies

$$E(z_{em} \bar{z}_{e'm'}) = \delta_{ee'} \delta_{mm'}$$

$$z_{e,-m} = (-1)^m \bar{z}_{em}.$$

$$\text{Now } \pi(z_{em}, C_e^{TT}) = \underbrace{\pi(z_{em}) \pi(C_e^{TT})}_{z \text{ & } C^{TT} \text{ are indep}}$$

$$\text{and } \begin{matrix} z_{em}, & C_e^{TT} \\ \text{III} & \text{III} \\ \times & \times \end{matrix}$$

is an ancillary parameterization.

So @ high ell the ancillary parameterization is slow.

The alternating Gibbs for the CMB problem becomes

for  $i=1, 2, \dots$

$$T_{em}^{(i+1)} \sim P(T | C_e^{TT(i)}, d)$$

$$C_e^{TT} \sim P(C_e^{TT} | T_{em}^{(i+1)}, d)$$

$$z_{em}^{(i+1)} \sim P(z | C_e^{TT}, d)$$

$$C_e^{TT(i+1)} \sim P(C_e^{TT} | z_{em}^{(i+1)}, d)$$

end

SLOW GIBBS Steps:

- $P(T | C_e^{TT(i+1)}, d)$  at low  $\ell$
- $P(C_e^{TT} | T_{em}^{(i+1)}, d)$  at high  $\ell$

↳ This is fixed by the new parameterization

FAST GIBBS Steps:

- $P(C_e^{TT} | z_{em}^{(i+1)}, d)$  at high  $\ell$

This is fast since

$$d_{em} = \sqrt{C_e^{TT}} z_{em}^{(i+1)} + n_{em}$$

even given  
this there is  
still tons of  
beam noise  
in  $n_{em}$

so  $\text{var}(C_e^{TT} | z^{(i+1)}, d)$  is still large.

(10)

# Interweaving Ancillary & sufficient chains

(11)

when the ancillary chain is fast  
there is a strange modification  
of the Alternating Gibbs that  
can dramatically speed things  
np. (Ref Yu & Meng 2010)

## Interweaving Gibbs:

for  $i=1, 2, \dots$

$$\tilde{\lambda}^{(i+1)} \sim P(\tilde{\lambda} | \varphi^{(i)}, d)$$

$$\varphi \sim P(\varphi | \tilde{\lambda}^{(i+1)}, d)$$

$$\lambda^{(i+1)} \sim P(\lambda | \varphi, \tilde{\lambda}^{(i+1)}, d)$$

$$\varphi^{(i+1)} \sim P(\varphi | \lambda^{(i+1)}, d)$$

end

Recall the ancillary chain  $(d, \lambda)$   
was fast, in the toy model

$$d = q + \lambda + n$$

when  $\Lambda = \text{small}$  &  $N = \text{large}$

$$\therefore \lambda^{(i+1)} \sim P(\lambda | \varphi, \tilde{\lambda}^{(i+1)}, d)$$

becomes deterministic since  
 $(\tilde{\lambda} = q + \lambda)$  & therefore

$$\lambda = \tilde{\lambda}^{(i+1)} - q$$

This simplifies to

for  $i=1, 2, \dots$

$$\tilde{\lambda}^{(i+1)} \sim P(\tilde{\lambda} | \varphi^{(i)}, d)$$

$$\lambda^{(i+1)} \sim P(\lambda | \tilde{\lambda}^{(i+1)}, d)$$

$$\varphi^{(i+1)} \sim P(\varphi | \lambda^{(i+1)}, d)$$

end

since  $\varphi \sim P(\varphi | \tilde{\lambda}^{(i+1)}, d) \Rightarrow \lambda^{(i+1)} \sim P(\lambda | \varphi, \tilde{\lambda}^{(i+1)}, d)$   
 $\Rightarrow$  (discard  $\varphi$ ) is equiv to

$$\lambda^{(i+1)} \sim \int P(\varphi, \lambda | \tilde{\lambda}^{(i+1)}, d) d\varphi$$

$$= P(\lambda | \tilde{\lambda}^{(i+1)}, d)$$

The corresponding CMB interweaving becomes

for  $i=1, 2, \dots$

$$T_{\text{em}}^{(i+1)} \sim P(T | C_e^{\text{TT}}(i), d)$$

$$C_e^{\text{TT}'} \sim P(C_e^{\text{TT}'} | T_{\text{em}}^{(i+1)}, d)$$

$$Z_{\text{em}}^{(i+1)} \sim P(Z | C_e^{\text{TT}'}, T_{\text{em}}^{(i+1)}, d)$$

$$C_e^{\text{TT}(i+1)} \sim P(C_e^{\text{TT}} | Z_{\text{em}}^{(i+1)}, d)$$

end

(12)

To compare alternating & interweaving we can analyze the last step

$$C_e^{TT(i+1)} \sim P(C_e^{TT} | z_{em}^{(i+1)}, d)$$

$$\propto P(d | C_e^{TT}, z_{em}^{(i+1)}) P(C_e^{TT} | z_{em}^{(i+1)})$$

$$= P(d | C_e^{TT}, z_{em}^{(i+1)}) \overbrace{P(C_e^{TT})}^{\text{The likelihood that } |d_{em} - \sqrt{C_e^{TT}} z_{em}^{(i+1)}|^2 \sim (n_{em})^2}$$

The likelihood that

$$|d_{em} - \sqrt{C_e^{TT}} z_{em}^{(i+1)}|^2 \sim (n_{em})^2$$

For the alternating chain we have

$$z_{em}^{(i+1)} \sim P(z | C_e^{TT}, d).$$

$$\text{Since } \frac{d_{em}}{\sqrt{C_e^{TT}}} = z_{em} + \frac{n_{em}}{\sqrt{C_e^{TT}}}$$

We have

$$z_{em}^{(i+1)} = \frac{1}{1 + \frac{C_e^{nn}}{C_e^{TT}}} \frac{d_{em}}{\sqrt{C_e^{TT}}} + \frac{\frac{C_e^{nn}}{C_e^{TT}}}{1 + \frac{C_e^{nn}}{C_e^{TT}}} \tilde{z}_{em}$$

$$= \frac{C_e^{TT}}{C_e^{TT} + C_e^{nn}} \frac{1}{\sqrt{C_e^{TT}}} d_{em} + \frac{C_e^{nn}}{C_e^{TT} + C_e^{nn}} \tilde{z}_{em}$$

$$= \mu_{em} + \gamma_e \tilde{z}_{em}$$

(13)

For the interweaving chain  
Z<sub>em</sub><sup>(i+1)</sup> is the deterministic move

$$Z_{em}^{(i+1)} = \frac{T_{em}^{(i+1)}}{\sqrt{C_e^{TT}}}$$

$$= \frac{1}{\sqrt{C_e^{TT}}} \left[ \frac{C_e^{TT}}{C_e^{TT} + C_e^{nn}} d_{em} + \frac{C_e^{nn}/C_e^{nn}}{C_e^{TT} + C_e^{nn}} \tilde{z}_{em} \right]$$

$$= \mu_{em} + \sqrt{C_e^{TT}} \gamma_e \tilde{z}_{em}$$

∴

$$|d_{em} - \sqrt{C_e^{TT}} z_{em}^{(i+1)}|^2$$

Alternating Gibbs

$$= \begin{cases} |d_{em} - \sqrt{C_e^{TT}} \mu_{em} - \sqrt{C_e^{TT}} \gamma_e \tilde{z}_{em}|^2 \\ |d_{em} - \sqrt{C_e^{TT}} \mu_{em} - \sqrt{C_e^{TT}} \sqrt{C_e^{TT}} \gamma_e \tilde{z}_{em}|^2 \end{cases}$$

Interweaving Gibbs

$\tilde{z}_{em}$  doesn't have any phase that resembles  $d_{em}$ .

∴ the  $\tilde{z}_{em}$  term adds power (leading to smaller likelihoods)

In the interweaving chain the

multiplier on  $\tilde{z}_{em}$  is smaller when  $C_e^{TT} \ll 1$  which has the effect of opening up the likelihood to more  $C_e^{TT}$  values  
⇒ Interweave is faster.

(14)

Rao-Blackwellization

(15)

Primordial Dipole example.

(16)