Lecture 9: ①

Topics:

I) Gibbs sampling $P(T, C_\ell \mid data)$

II) Sufficient vrs Ancillary parameterizations

III) Alternating vrs Interweaving

IV) Rao-Blackwelization

V) Primordial Dipole example.

Gibbs sampling $P(T, C_\ell \mid data)$

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

$$(*) \qquad P(T, C_\ell^{TT} \mid d)$$

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

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

$$(**) \qquad P(T, \theta \mid d)$$

where $\theta := (\Omega_b h^2, \Omega_c h^2, \ldots, n_s)$.

For the posterior $(*)$ the standard Gibbs implimentation produces a sequence of pairs of maps & C_ℓ^{TT}s:

$$(T^{(1)}, C_\ell^{TT(1)}), (T^{(2)}, C_\ell^{TT(2)}), \ldots$$

where

$$T^{(i+1)}(\hat{n}) \sim P(T \mid C_\ell^{TT(i)}, d)$$

$$C_\ell^{TT(i+1)} \sim P(C_\ell^{TT} \mid T^{(i+1)}, \cancel{d}).$$

drop d

when observing the full sky with stationary noise and one uses Jeffrey's prior $\pi(C_\ell^{TT}) \propto \frac{1}{C_\ell^{TT}}$ the Gibbs steps are given explicitly by: ②

$$T_{\ell m}^{(i+1)} = \frac{C_\ell^{TT(i)}}{C_\ell^{TT(i)} + C_\ell^{nn}} d_{\ell m} + \frac{\sqrt{C_\ell^{TT(i)} C_\ell^{nn}}}{C_\ell^{TT(i)} + C_\ell^{nn}} z_{\ell m}^{(i)}$$

$$C_\ell^{TT(i+1)} \sim \underbrace{{C_\ell^{TT}}^{-\frac{2\ell+1}{2}-1} \exp\left(-\frac{2\ell+1}{2} \frac{\sigma_\ell^{(i+1)}}{C_\ell^{TT}}\right)}_{\substack{\text{probability density} \\ \text{of } P(C_\ell^{TT} \mid T^{(i+1)})}}$$

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

and $z_{\ell m}^{(i)}$'s are Complex Gaussian, independent accross i which satisfy

$$E(z_{\ell m} \overline{z_{\ell' m'}}) = \delta_{\ell \ell'} \delta_{m m'}$$

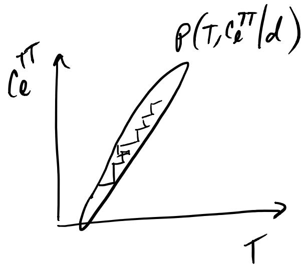
$$z_{\ell, -m} = (-1)^m \overline{z_{\ell, m}}$$

Note: $C_\ell^{TT(i+1)}$ above is recognised as an $\mathrm{InvGamma}\left(\alpha = \frac{2\ell+1}{2}, \beta = \frac{2\ell+1}{2} \sigma_\ell^{(i+1)}\right)$

Slow mixing

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

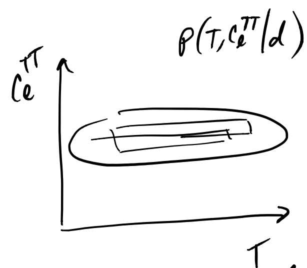
Case 1:



correlation = large

slow Gibbs.

Case 2:



correlation = small

Fast Gibbs.

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 ℓ, the Inv Gamma conditional gives

$$E(C_\ell^{TT} | T, \underset{\text{drop}}{d}) = \frac{\beta}{\alpha-1} = \frac{\ell+\frac{1}{2}}{\ell-\frac{1}{2}}\sigma_\ell$$

$$\text{var}(C_\ell^{TT} | T, \cancel{d}) = \left(\frac{\beta}{\alpha-1}\right)^2 \frac{1}{\alpha-2}$$

$$= \left(\frac{\ell+\frac{1}{2}}{\ell-\frac{1}{2}}\right)^2 \left[\frac{\sigma_\ell}{\ell-\frac{3}{2}}\right]^2$$

$$\sim \frac{\sigma_\ell}{\ell}$$

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

but the beam implies

$$\text{var}(C_\ell^{TT} | d) \to \text{prior var of } C_\ell^{TT}$$

as $\ell \to \infty$.

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

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

This implies

$$\text{var}(T_{\ell m} | C_\ell^{TT}, d) \simeq \text{small}$$

However cosmic variance makes $\text{var}(C_\ell^{TT}|d)$ large.

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

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

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

$$\text{var}(T_{\ell m} | d) \gg \text{var}(T_{\ell m} | C_\ell^{TT}, d)$$

This carries a lot of info to constrain T

The typical way to fix this is by Reparamterization.

sufficient vrs Ancillary parameterizations

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

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

two paramters. Noise

where $\varphi \sim N(0, \Sigma)$, $\lambda \sim N(0, \Lambda)$ } Priors.

$n \sim N(0, N)$

are indep.

Suppose your interested in estimating φ & λ is a nuisance parameter. So λ acts like extra additive Noise.

Case 1: Λ = large & N = small.

$var(\varphi \mid d)$ = large since λ acts like a large Noise corruption.

If we are given λ we can remove the "λ noise" from d and get much tighter constraints on φ.

$$\therefore \; var(\varphi \mid \lambda, d) << var(\varphi \mid d)$$

$\therefore$ SLOW GIBBS.

Case 2: Λ = small & N = large

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

$$\therefore \; var(\varphi \mid \lambda, d) \approx var(\varphi \mid d)$$

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

Now consider a New parameterization

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

Orginal parameterization : (φ, λ)

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

The prior transforms to

$$P(\varphi, \tilde{\lambda}) = \underbrace{P(\tilde{\lambda} \mid \varphi)}_{\sim N(\varphi, \Lambda)} \underbrace{P(\varphi)}_{\sim N(0, \Sigma)}$$

Case 1: Λ = large & N = small

$$\underbrace{var(\varphi \mid d)}_{\text{still large}} \simeq \underbrace{var(\varphi \mid \tilde{\lambda}, \overset{\text{discard}}{\cancel{d}})}$$

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

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

$\therefore$ FAST GIBBS

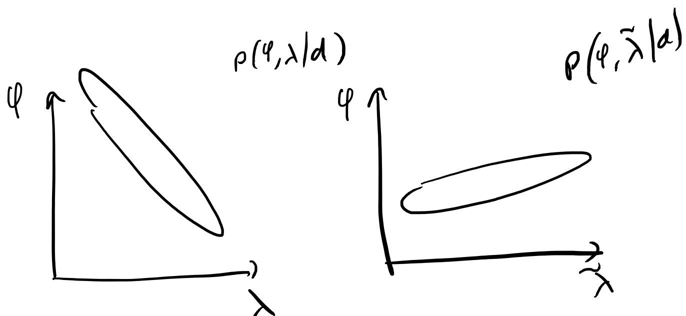
Case 2: Λ large & N = large (7)

$$\mathrm{var}(\varphi \mid d) \gg \mathrm{var}(\varphi \mid \tilde{\lambda}, \overset{\text{discared}}{d})$$

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

$\therefore$ SLOW GIBBS.

This Reparametrization is visualized as



$$\begin{pmatrix}\varphi\\ \lambda\end{pmatrix} \longmapsto \begin{pmatrix}\varphi\\ \lambda+\varphi\end{pmatrix} = \begin{pmatrix}\varphi\\ \tilde{\lambda}\end{pmatrix}$$

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.

(φ,λ) is called a Ancillary parameterization (8)

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

λ is Ancillary for φ.

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

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

$\tilde{\lambda}$ is sufficient for φ

Alternating Sufficient & Ancillary chains

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

Alternating Gibbs:

for $i = 1, 2, \dots$

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

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

end

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

$p(Z \mid T, d)$ inject cosmic variance inequality.

Application to CMB Gibbs

(S) $d_{\ell m} = T_{\ell m} + n_{\ell m}$

where $C_\ell^{TT} \sim \pi(C_\ell^{TT})$

$T_{\ell m} \overset{\text{"iid"}}{\sim} \mathcal{N}(0, C_\ell^{TT})$

Note $T_{\ell m}$ is sufficient for C_ℓ^{TT} so $(T_{\ell m}, C_\ell^{TT})$ is effectively a sufficient parameterization: $(T_{\ell m}, C_\ell^{TT})$, with $T_{\ell m} \equiv \tilde{X}$ and $C_\ell^{TT} \equiv \varphi$

One can switch to an ancillary parameterization as follows.

(A) $d_{\ell m} = \sqrt{C_\ell^{TT}}\, Z_{\ell m} + n_{\ell m}$

where $Z_{\ell m}$ satisfies

$E(Z_{\ell m} \overline{Z}_{\ell' m'}) = \delta_{\ell\ell'} \delta_{mm'}$

$Z_{\ell,-m} = (-1)^m \overline{Z}_{\ell m}$.

Now $\pi(Z_{\ell m}, C_\ell^{TT}) = \underbrace{\pi(Z_{\ell m})\pi(C_\ell^{TT})}_{Z \,\&\, C^{TT} \text{ are indep.}}$

and $(Z_{\ell m}, C_\ell^{TT})$, with $Z_{\ell m} \equiv \lambda$ and $C_\ell^{TT} \equiv \varphi$

is an ancillary parameterization.

So @ high ell the ancillary parameterization is slow.

The alternating Gibbs for the CMB problem becomes

for $i = 1, 2, \ldots$

$T_{\ell m}^{(i+1)} \sim P(T \mid C_\ell^{TT\,(i)}, d)$

$C_\ell^{TT} \sim P(C_\ell^{TT} \mid T_{\ell m}^{(i+1)}, d)$

$Z_{\ell m}^{(i+1)} \sim P(Z \mid C_\ell^{TT}, d)$

$C_\ell^{TT(i+1)} \sim P(C_\ell^{TT} \mid Z_{\ell m}^{(i+1)}, d)$

end

SLOW GIBBS Steps:

- $P(T \mid C_\ell^{TT\,(i+1)}, d)$ at low ℓ
- $P(C_\ell^{TT} \mid T_{\ell m}^{(i+1)}, d)$ at high ℓ

This is fixed by the new parameterization

FAST GIBBS Steps:

- $P(C_\ell^{TT} \mid Z_{\ell m}^{(i+1)}, d)$ at high ℓ

This is fast since

$d_{\ell m} = \sqrt{C_\ell^{TT}}\, Z_{\ell m}^{(i+1)} + n_{\ell m}$

even given this there is still tons of beam noise in $n_{\ell m}$

So $\mathrm{var}(C_\ell^{TT} \mid Z^{(i+1)}, d)$ is still large.

Interweaving Ancillary & sufficient chains

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

Interweaving Gibbs:

for $i = 1, 2, \dots$

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

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

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

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

end

Recall the ancillary chain (φ, λ) was fast, in the toy model

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

when Λ = small & N = large

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

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

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

This simplifies to

for $i = 1, 2, \dots$

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

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

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

end

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

$\Rightarrow$ (discard φ) is equiv to

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

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

The corresponding CMB interweaving becomes

for $i = 1, 2, \dots$

$$T_{em}^{(i+1)} \sim P(T \mid C_\ell^{TT\,(i)}, d)$$

$$C_\ell^{TT\prime} \sim P(C_\ell^{TT} \mid T_{em}^{(i+1)}, d)$$

$$Z_{em}^{(i+1)} \sim P(Z \mid C_\ell^{TT\prime}, T_{em}^{(i+1)}, d)$$

$$C_\ell^{TT\,(i+1)} \sim P(C_\ell^{TT} \mid Z_{em}^{(i+1)}, d)$$

end

To compare alternating & interweaving we can analyze the last step

$$C_\ell^{TT(i+1)} \sim P\left(C_\ell^{TT} \,\middle|\, z_{\ell m}^{(i+1)}, d\right)$$

$$\propto P\left(d \,\middle|\, C_\ell^{TT}, z_{\ell m}^{(i+1)}\right) P\left(C_\ell^{TT} \,\middle|\, z_{\ell m}^{(i+1)}\right)$$

$$= \underbrace{P\left(d \,\middle|\, C_\ell^{TT}, z_{\ell m}^{(i+1)}\right)}\pi\left(C_\ell^{TT}\right)$$

The likelihood that $\left|d_{\ell m} - \sqrt{C_\ell^{TT}}\, z_{\ell m}^{(i+1)}\right|^2 \sim |n_{\ell m}|^2$

For the alternating chain we have

$$z_{\ell m}^{(i+1)} \sim P\left(z \,\middle|\, C_\ell^{TT'}, d\right).$$

Since $\dfrac{d_{\ell m}}{\sqrt{C_\ell^{TT'}}} = z_{\ell m} + \dfrac{n_{\ell m}}{\sqrt{C_\ell^{TT'}}}$

we have

$$z_{\ell m}^{(i+1)} = \frac{1}{1 + C_\ell^{nn}/C_\ell^{TT'}} \frac{d_{\ell m}}{\sqrt{C_\ell^{TT'}}} + \frac{C_\ell^{nn}/C_\ell^{TT}}{1 + C_\ell^{nn}/C_\ell^{TT}} \overset{\text{indep.}}{\tilde{z}_{\ell m}}$$

$$= \frac{C_\ell^{TT'}}{C_\ell^{TT'} + C_\ell^{nn}} \frac{1}{\sqrt{C_\ell^{TT'}}} d_{\ell m} + \frac{C_\ell^{nn}}{C_\ell^{TT'} + C_\ell^{nn}} \tilde{z}_{\ell m}$$

$$= \mu_{\ell m} + \tau_\ell \tilde{z}_{\ell m}$$

For the interweaving chain $z_{\ell m}^{(i+1)}$ is the deterministic move

$$z_{\ell m}^{(i+1)} = \frac{T_{\ell m}^{(i+1)}}{\sqrt{C_\ell^{TT'}}}$$

$$= \frac{1}{\sqrt{C_\ell^{TT'}}}\left[\frac{C_\ell^{TT'}}{C_\ell^{TT'} + C_\ell^{nn}} d_{\ell m} + \frac{C_\ell^{TT'} C_\ell^{nn}}{C_\ell^{TT'} + C_\ell^{nn}} \tilde{z}_{\ell m}\right]$$

$$= \mu_{\ell m} + \sqrt{C_\ell^{TT'}}\, \tau_\ell \tilde{z}_{\ell m}$$

$\therefore$

$$\left|d_{\ell m} - \sqrt{C_\ell^{TT}}\, z^{(i+1)}\right|^2 = \begin{cases} \left|d_{\ell m} - \sqrt{C_\ell^{TT}}\mu_{\ell m} - \sqrt{C_\ell^{TT}}\, \tau_\ell \tilde{z}_{\ell m}\right|^2 & \text{Alternating Gibbs} \\ \left|d_{\ell m} - \sqrt{C_\ell^{TT}}\mu_{\ell m} - \sqrt{C_\ell^{TT}}\sqrt{C_\ell^{TT'}}\, \tau_\ell \tilde{z}_{\ell m}\right|^2 & \text{Interweaving Gibbs} \end{cases}$$

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

$\therefore$ the $\tilde{z}_{\ell m}$ term adds power (leading to smaller likelihoods)

In the interweaving chain the multiplier on $\tilde{z}_{\ell m}$ is smaller when $C_\ell^{TT'} \ll 1$ which has the effect of opening up the likelihood to more C_ℓ^{TT} values

$\Longrightarrow$ Interweave is faster.

Rao-Blackwelization

15

Primordial Dipole example.

16