

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.
- VI) ISW

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 Λ 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)

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

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

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

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

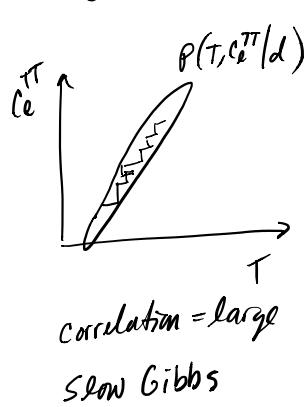
Note: $C_e^{TT(i+1)}$ above is recognised as an Inv Gamma ($\alpha = \frac{2\ell+1}{2}$, $\beta = \frac{2\ell+1}{2} \sigma_e^{(i+1)}$)

Slow mixing

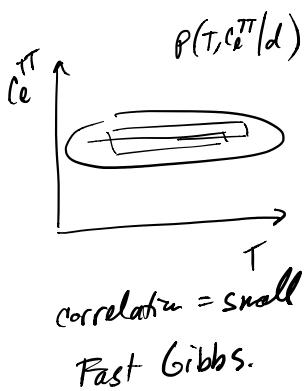
(3)

The most common difficulty with Gibbs is that it is often slow to mix/converge. Heuristically this happens when the (T, C_e^{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 ℓ , 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 ℓ this gibbs doesn't mix well.

e.g. For low ℓ 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

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

$\therefore \text{var}(T_{\text{em}} | d)$ 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 two variables is highly informative for the other variable i.e. for low ℓ

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

\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 reparametrizes 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 φ & λ is a nuisance parameter. So λ acts like extra additive noise.

Noise.

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

$\text{var}(\varphi | d) = \text{large}$ 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 .

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

\therefore SLOW GIBBS.

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

Now given λ we can still denoise 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 : (φ, λ)

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 φ 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 $\text{var}(n)$ was small &
 $\tilde{\lambda} \sim N(\varphi, \Delta)$ so it's not
very informative large to be given $\tilde{\lambda}$.

\therefore FAST GIBBS

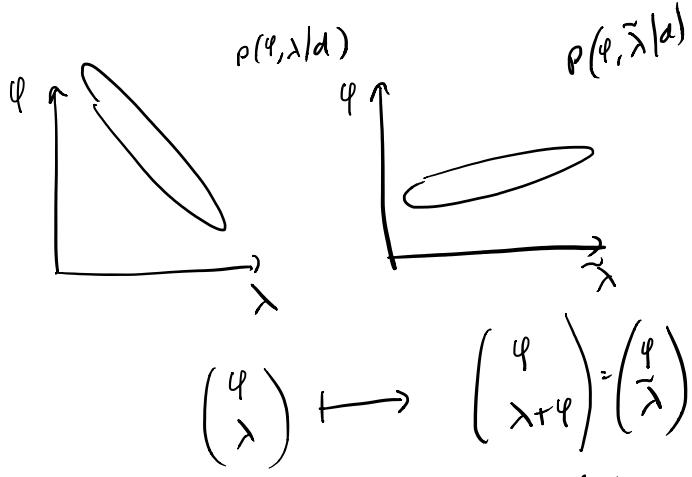
Case 2: λ small & $N = \text{large}$ (7)

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

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

\therefore SLOW GIBBS

This Reparametrization is
visualized as



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

$(\boldsymbol{\varphi}, \lambda)$ is called a Ancillary Parameterization

$$\lambda = \boldsymbol{\varphi} + \mathbf{n}, \quad \boldsymbol{\varphi} \perp \lambda$$

$\sim \lambda$ is Ancillary for $\boldsymbol{\varphi}$.

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

$$\lambda = \tilde{\boldsymbol{\lambda}} + \mathbf{n}, \quad \tilde{\boldsymbol{\lambda}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

Alternating Sufficient & Ancillary chains

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

Alternating Gibbs:

for $i = 1, 2, \dots$

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

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

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

end

The mixing rate will be the
better than the worst case.

Application to CMB Gibbs

(9)

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

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

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

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

One can switch to an ancillary parameterization as follows.

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

where z_{em} satisfies

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

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

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

$$\text{and } \begin{matrix} (z_{em}, C_e^{TT}) \\ \text{III} \quad \text{III} \\ \times \quad q \end{matrix}$$

is an ancillary parameterization.

So @ high ell the sufficient 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 ℓ
- $P(C_e^{TT} | T_{em}^{(i+1)}, d)$ at high ℓ

↳ This is fixed by the new parameterization

FAST GIBBS Steps:

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

This is mixes well 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
up. (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, λ)
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 $\varphi' \sim P(\varphi | \tilde{\lambda}^{(i+1)}, d) \Rightarrow \lambda^{(i+1)} \sim P(\lambda | \varphi, \tilde{\lambda}^{(i+1)}, d)$
 $\Rightarrow (\text{discard } \varphi')$ 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

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

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

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

For the interweaving chain
⑯

$\tilde{z}_{em}^{(i+1)}$ is the deterministic move

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

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

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

∴

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

Alternating Gibbs

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

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.

Rao-Blackwellization

(15)

From samples

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

You can produce an estimate of $P(C_e^{TT}|d)$ simply by

$$\hat{P}(C_e^{TT}|d) := \underbrace{\frac{1}{N} \sum_{i=1}^N}_{\text{estimated density.}} \delta_{C_e^{TT} - C_e^{TT(i)}}$$

Notice that this estimate is unbiased in that

$$\begin{aligned} E(\hat{P}(C_e^{TT}|d)) &= \frac{1}{N} \sum_{i=1}^N \int \int_{C_e^{TT} - C_e^{TT(i)}} P(C_e^{TT(i)}|d) dC_e^{TT(i)} \\ &\stackrel{\uparrow \text{w.r.t. the posterior } P(C_e^{TT}|d)}{=} \frac{1}{N} \sum_{i=1}^N P(C_e^{TT}|d) \\ &= P(C_e^{TT}|d). \end{aligned}$$

By a theorem due to Rao & Blackwell if one has an unbiased estimator

$\hat{\theta}(X_1, \dots, X_N)$ of some parameter θ based on data X_1, \dots, X_N then the new estimator

$$\tilde{\theta}(X_1, \dots, X_N) := E(\hat{\theta}(X_1, \dots, X_N) | T(X_1, \dots, X_N))$$

has smaller MSE when $T(X_1, \dots, X_N)$ is a sufficient statistic for θ , i.e. if

$P(X_1, \dots, X_N | T(X_1, \dots, X_N))$ doesn't depend on θ .

In our case, the data X_1, \dots, X_N corresponds to

$$X_i = (C_e^{TT(i)}, T^{(i)})$$

and $\theta = P(C_e^{TT}|d) \leftarrow$ the posterior density evaluated at C_e^{TT}

$$\hat{\theta}(X_1, \dots, X_N) = \hat{P}(C_e^{TT}|d)$$

Notice that

$$T(X_1, \dots, X_N) = \begin{pmatrix} \vdots \\ T^{(i)} \\ \vdots \end{pmatrix}$$

is a sufficient statistic for θ

in that

$$P(C_e^{TT(1)}, \dots, C_e^{TT(N)} | T^{(1)}, \dots, T^{(N)}, d)$$

$$= \prod_{i=1}^N \underbrace{P(C_e^{TT(i)} | T^{(i)}, d)}$$

$$\text{Inv Gamma}(\alpha = \frac{2\ell+1}{2}, \beta = \frac{2\ell+1}{2} \sigma_e^{(i+1)})$$

\uparrow
 $T^{(i)}$ bandpowers

which holds irrespective of what the data is (so is not responsive to $P(C_e^{TT}|d)$).

Therefore we can construct a new estimator with smaller MSE as follows

$$\tilde{P}(C_e^{TT}|d) := E(\hat{P}(C_e^{TT}|d) | T^{(1)}, \dots, T^{(N)})$$

$$= \frac{1}{N} \sum_{i=1}^N E\left(\delta_{C_e^{TT} - C_e^{TT(i)}} \mid T^{(1)}, \dots, T^{(N)}\right)$$

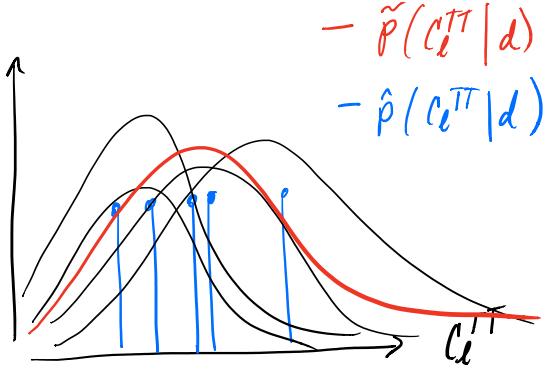
\downarrow
This is the

$$= \frac{1}{N} \sum_{i=1}^N \underbrace{P(C_e^{TT} | T^{(i)})}_{\text{density}}$$

$$\text{Inv Gamma}(\alpha = \frac{2\ell+1}{2}, \beta = \frac{2\ell+1}{2} \sigma_e^{(i+1)}) \text{ density}$$

(16)

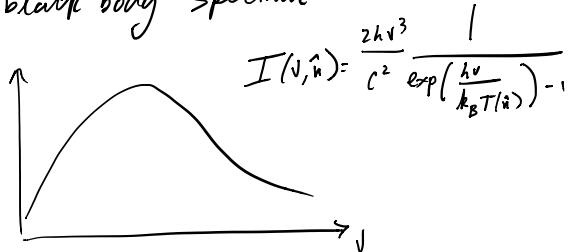
Picture



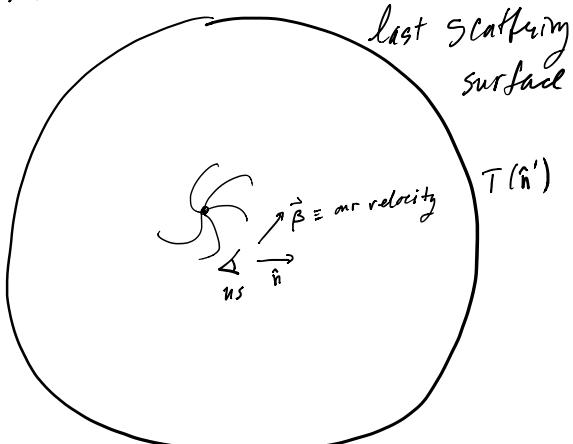
(17)

Primordial Dipole example.

Recall that in a direction $\hat{n} \in S^2$ we do not directly observe $T(\hat{n})$ instead observe a noisy version of $\int_{B(\hat{n})} I(v, \hat{n}) dv$ where the intensity $I(v, \hat{n})$ is given by the black body spectrum



Since we are moving (e.g. within our galaxy) with respect to the surface of last scattering there is a doppler boosting effect on $I(v, \hat{n})$ in the direction of our travel:



Note: β is measured as a fraction of the speed of light c .

To quantify this effect we must analyze our observations w.r.t an observer (at our same location) which has no velocity w.r.t the CMB.

(18)

By special relativity

$$\frac{I(v, \hat{n})}{v^3} = \frac{I(v', \hat{n}')}{(v')^3}$$

what we observe in direction \hat{n} @ frequency v

what the rest frame observer sees in direction \hat{n}' @ frequency v'

where

$$\hat{n}' = \frac{\hat{n}' + [\gamma \beta + (\gamma - 1) \langle \hat{n}', \hat{\beta} \rangle] \hat{\beta}}{\gamma(1 + \langle \hat{n}', \hat{\beta} \rangle)}$$

$$\beta = |\hat{\beta}|, \hat{\beta} = \frac{\vec{\beta}}{|\vec{\beta}|}, \gamma = \frac{1}{\sqrt{1 - \beta^2}} \text{ and}$$

$$v = v' \gamma (1 + \langle \hat{n}', \hat{\beta} \rangle)$$

$$\therefore I(v, \hat{n}) = \left(\frac{v}{v'}\right)^3 I(v', \hat{n}')$$

$$= \left(\frac{v}{v'}\right)^3 \frac{2h(v')^3}{c^2} \frac{1}{\exp\left(\frac{hv'}{k_B T(\hat{n}')}\right) - 1}$$

$$= \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{hv}{k_B \gamma(1 + \langle \hat{n}', \hat{\beta} \rangle) T(\hat{n}')}\right) - 1}$$

looks like
Modified temp
on the CMB.

\therefore The observed/inferred temp $T^o(\hat{n})$ (19) is actually:

$$T^o(\hat{n}) = \gamma(1 + \langle \hat{n}', \vec{\beta} \rangle) T^P(\hat{n}')$$

↑
primordial temp
on the surface
of last scattering

our inferred temp

(20)
observations suggest $\beta \sim 1.23 \times 10^{-3}$
so one often takes a Taylor expansion in
 β to describe the main effect on the
dipole of $T^o(\hat{n})$.

Notice that

$$\langle \hat{n}, \vec{\beta} \rangle = \frac{\langle \hat{n}', \vec{\beta} \rangle + [r\beta + (r-1)\langle \hat{n}', \vec{\beta} \rangle] \beta}{\gamma(1 + \langle \hat{n}', \vec{\beta} \rangle)}$$

$$\text{since } \langle \vec{\beta}, \vec{\beta} \rangle = \beta^2$$

$$= \frac{\cancel{\langle \hat{n}', \vec{\beta} \rangle} + \gamma\beta^2 + (r-1)\langle \hat{n}', \vec{\beta} \rangle}{\gamma(1 + \langle \hat{n}', \vec{\beta} \rangle)} \Rightarrow \langle \vec{\beta}, \vec{\beta} \rangle = \beta$$

$$= \frac{\gamma(\beta^2 + \langle \hat{n}', \vec{\beta} \rangle)}{\gamma(1 + \langle \hat{n}', \vec{\beta} \rangle)}$$

$$\therefore \langle \hat{n}, \vec{\beta} \rangle - \beta^2 = \langle \hat{n}', \vec{\beta} \rangle (1 - \langle \hat{n}, \vec{\beta} \rangle)$$

$$\therefore \langle \hat{n}', \vec{\beta} \rangle = \frac{\langle \hat{n}, \vec{\beta} \rangle - \beta^2}{1 - \langle \hat{n}, \vec{\beta} \rangle}$$

$$\therefore T^o(\hat{n}) = \gamma(1 + \langle \hat{n}', \vec{\beta} \rangle) \cdot T^P(\hat{n}')$$

$$= \gamma \left(1 + \frac{\langle \hat{n}, \vec{\beta} \rangle - \beta^2}{1 - \langle \hat{n}, \vec{\beta} \rangle} \right) \cdot T^P(\hat{n}')$$

$$= \gamma \left(\frac{1 - \beta^2}{1 - \langle \hat{n}, \vec{\beta} \rangle} \right) T^P(\hat{n}')$$

$$= \frac{T^P(\hat{n}')}{\gamma(1 - \langle \hat{n}, \vec{\beta} \rangle)}$$

$$T_{\text{elm}}^o \text{ where } l=1, m=-1, 0, 1$$

$$\frac{1}{\gamma(1 - \langle \hat{n}, \vec{\beta} \rangle)} = \frac{\sqrt{1 - \beta^2}}{(1 - \mu\beta)} \text{ where } \langle \hat{n}, \vec{\beta} \rangle = \langle \hat{n}, \vec{\beta} \rangle = \mu$$

$$= 1 + \mu\beta + \left(\mu^2 - \frac{1}{2}\right)\beta^2$$

$$+ \left(\mu^3 - \frac{\mu}{2}\right)\beta^3 + \dots$$

$$\therefore T^o(\hat{n}) = T^P(\hat{n}') \left[1 + \mu\beta + \mathcal{O}(\beta^2) \right]$$

$$= \left[T_{\text{CMB}}^P + \delta T^P(\hat{n}') \right] \left[1 + \mu\beta + \mathcal{O}(\beta^2) \right]$$

$$\begin{aligned} \text{Base line} \\ \text{temp } &= T_{\text{CMB}}^P \\ &\approx 2.7K \\ &= T_{\text{CMB}}^P + T_{\text{CMB}}^P \langle \hat{n}, \vec{\beta} \rangle \\ &= T_{\text{CMB}}^P + \delta T^P(\hat{n}') + \dots \end{aligned}$$

The term $T_{\text{CMB}}^P \langle \hat{n}, \vec{\beta} \rangle$ is what induces a dipole in $T^o(\hat{n})$.

In terms of estimating $\vec{\beta}$ & the cosmological dipole $T_{1,-1}^P, T_{1,0}^P, T_{1,1}^P$, the ancillary & sufficient Gibbs seems relevant since both clearly compete to explain the data.

To generate a discussion in class

Consider splitting off the
Cosmological Dipole as follows

$$T^o(\hat{n}) = \frac{T^P(\hat{n}')}{\gamma(1 - \langle \hat{n}, \vec{\beta} \rangle)}$$
$$= \frac{T_{CMB}^P + \sum_{m=-1}^1 T_{em}^P Y_{em}(\hat{n}') + RT^P(\hat{n}')}{\gamma(1 - \langle \hat{n}, \vec{\beta} \rangle)}$$

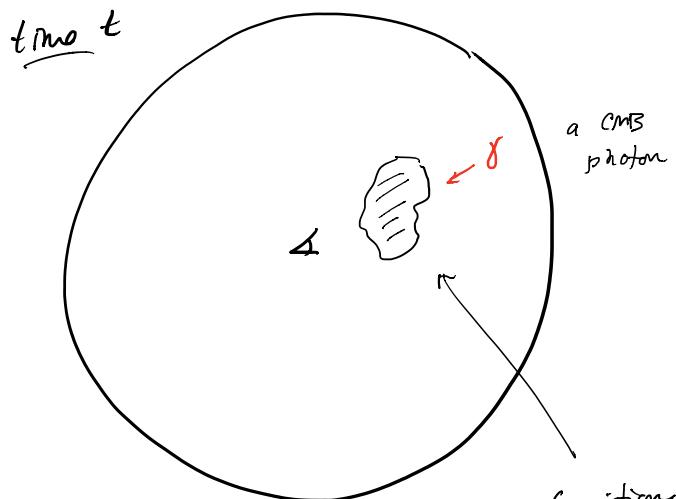
The first question is: is it at all possible
to disambiguate $T_{1,m}^P$ ($m = -1, 0, 1$) from
 $\vec{\beta}$ when the observations are in
the form

$$d(\hat{n}) = \frac{T_{CMB}^P + \sum_{m=-1}^1 T_{em}^P Y_{em}(\hat{n}')}{\gamma(1 - \langle \hat{n}, \vec{\beta} \rangle)} + n(\hat{n})$$

(21)

Integrated Sachs-Wolff effect

(22)



Gravitational
well ... i.e.
over density
of mass

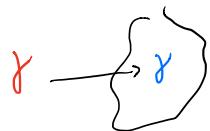
As the photon γ enters
the gravitational well
it gains momentum &
 \therefore blue shifts $\gamma \rightarrow \gamma'$

If the mass was stationary it
would loose the same amount of
momentum when exiting the potential



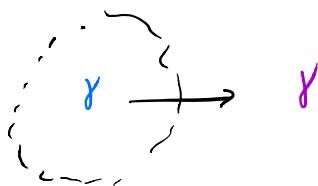
However, since space is expanding.
This causes the overdensity to
dissolve slightly as γ is passing
through it, resulting in
excess momentum as γ
exits the overdensity

time t :



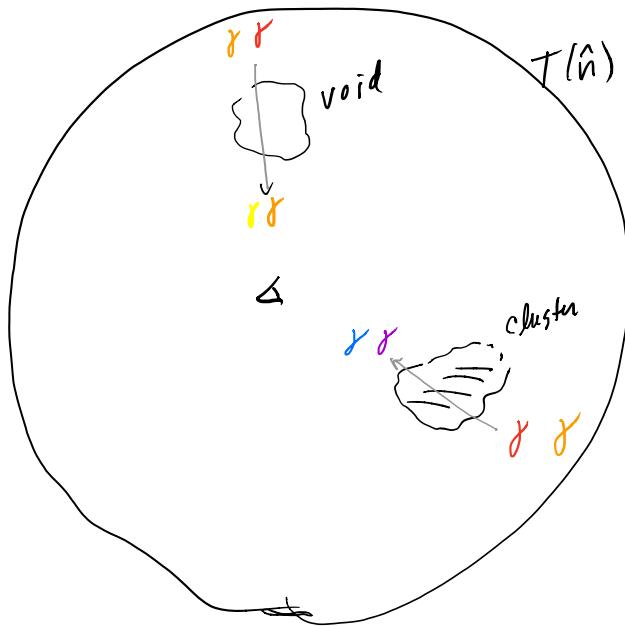
(23)

time $t+\epsilon$:



The opposite effect occurs for voids.

This is called the Integrated Sachs-Wolff effect ISW.



The ISW will affect the intensity distribution $I(v, \hat{n})$ of incoming CMB photons in directions \hat{n} which have over dense or under dense projected mass.

The ISW effect on $I(v, \hat{n})$ can be understood in exactly the same way as in the dipole example.

In particular, the excess momentum due to ISW direction \hat{n} has exactly the same effect as one would see if we were moving in direction \hat{n} (with velocity \vec{v}) toward the photon (with no ISW).

$$\therefore I^{\text{ISW}}(v, \hat{n}) = \underbrace{I^{\hat{n}}(v, \hat{n})}_{\substack{\text{observed} \\ \text{Intensity}}} \quad \text{observed intensity with no ISW but apparent motion forward } \hat{n}$$

$\therefore I^{\text{ISW}}(v, \hat{n}) = \underbrace{I^{\hat{n}}(v, \hat{n})}_{\substack{\text{observed} \\ \text{Intensity}}} \quad \text{observed intensity with no ISW but apparent motion forward } \hat{n}$

$$= \frac{2 h v^3}{c^2} \frac{1}{\exp\left(\frac{h v}{k_b r(1 + \langle \hat{n}', \hat{n} \rangle) T(\hat{n}')} \right)} -$$

looks like modified temp on the CMB.

$$\text{where } \hat{n} = \hat{n}' + \frac{[r n + (r-1) \langle \hat{n}', \hat{n} \rangle] \hat{n}}{r(1 + \langle \hat{n}', \hat{n} \rangle)}$$

$$\therefore \langle \hat{n}', \hat{n} \rangle = \frac{\langle \hat{n}, \hat{n} \rangle - n^2}{1 - \langle \hat{n}, \hat{n} \rangle}$$

$$= \frac{n - n^2}{1 - n} = n$$

$$\therefore \langle \hat{n}', \hat{n} \rangle = 1$$

i.e. $\hat{n}' = \hat{n}$

\therefore the ISW effect on $T(v_j \hat{n})$ (25)
looks like a modified CMB effect

$$T^{FSW}(\hat{n}) = T^{\hat{n}}(\hat{n}) = \gamma(1+n) T^P(\hat{n})$$

↓
 apparent
temp with
DSW

 ↓
 apparent
temp if we
are moving
in the direction
of \hat{n} with no DSW

 ↓
 primordial
temp

$$= \frac{1+n}{\sqrt{1-n^2}} T^P(\hat{n})$$

\therefore in the direction of clusters we
should see a multiplicative effect

$$T^P(\hat{n}) \xrightarrow{ISW} \frac{1+n}{\sqrt{1-n^2}} T^P(\hat{n})$$

& in the direction of voids
we should see

$$T^P(\hat{n}) \xrightarrow{ISW} \frac{1-n}{\sqrt{1-n^2}} T^P(\hat{n})$$

since $\left(\frac{1+n}{\sqrt{1-n^2}}\right)^2 = \frac{(1+n)^2}{1-n^2} = \frac{(1+n)^2}{(1+n)(1-n)} = \frac{1+n}{1-n} > 1$

we see a boost of temp for
behind clusters & since

$$\left(\frac{1-n}{\sqrt{1-n^2}}\right)^2 = \frac{(1-n)^2}{(1+n)(1-n)} = \frac{1-n}{1+n} < 1$$

we see a reduction of temp
behind voids.

There is also a secondary effect
on ST^P due to DSW. (26)

Let $\hat{n}^I(\hat{n})$ = our velocity
in direction \hat{n} needed
to match the FSW effect.

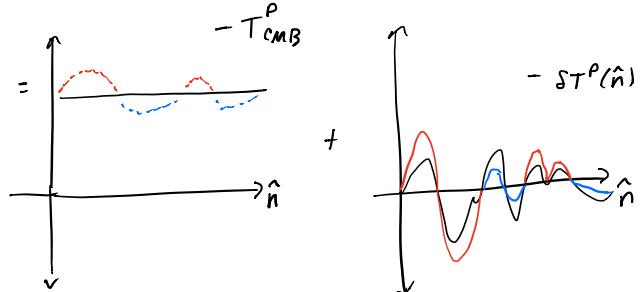
$$= \underbrace{n^I(\hat{n})}_{\text{signed version of } n \text{ from}} \hat{n}$$

previous section.

$$\therefore T^{FSW}(\hat{n})$$

$$= \sqrt{\frac{1+n^I(\hat{n})}{1-n^I(\hat{n})}} T^P(\hat{n})$$

$$= \sqrt{\frac{1+n^I(\hat{n})}{1-n^I(\hat{n})}} T_{CMB}^P + \sqrt{\frac{1+n^I(\hat{n})}{1-n^I(\hat{n})}} ST^P(\hat{n})$$



... → voids i.e. $n^I(\hat{n}) < 0$
..., -clusters i.e. $n^I(\hat{n}) > 0$

$$= T_{CMB}^P + \sqrt{\frac{1+n^I(\hat{n})}{1-n^I(\hat{n})}} ST^P(\hat{n}) + \left(\sqrt{\frac{1+n^I(\hat{n})}{1-n^I(\hat{n})}} - 1 \right) T_{CMB}^P$$

This term gives rise
to extra power in
C_e^{TEEN}

$$\therefore \frac{T^{FSW}(\hat{n})}{T_{CMB}^P} = \gamma(\hat{n}) \frac{ST^P(\hat{n})}{T_{CMB}^P} + (I(\hat{n}) - 1)$$

Combining the Dipole effect & ISW. (27)

The effect due to our angular motion was derived as

$$T^{\vec{\beta}}(\hat{n}) = \frac{\sqrt{1-\beta^2}}{(1-\langle \hat{n}, \vec{\beta} \rangle)} T^p(\hat{n}')$$

For ISW we made an equivalence with apparent motion towards or away in the viewing direction \hat{n} . Let

$$\begin{aligned} \hat{n}^I(\hat{n}) &= \text{our velocity in direction } \hat{n} \text{ needed to match the ISW effect.} \\ &= n^I(\hat{n}) \hat{n} \end{aligned}$$

(28)

Interweaving Gibbs
may help here
for $P(\hat{n}, \frac{8\pi}{T})$ (data)

$$\therefore T^{ISW}(\hat{n}) = \frac{\sqrt{1-(n^I(\hat{n}))^2}}{(1-n^I(\hat{n}))} T^p(\hat{n})$$

Combining the two, ISW first, then our velocity second gives

$$T^o(\hat{n}) = \frac{\sqrt{1-\beta^2}}{(1-\langle \hat{n}, \vec{\beta} \rangle)} \frac{\sqrt{1-n^I(\hat{n}')^2}}{(1-n^I(\hat{n}'))} T^p(\hat{n}')$$

where

$$\hat{n}' = \frac{\hat{n}' + [\gamma \beta + (\gamma - 1) \langle \hat{n}', \vec{\beta} \rangle] \vec{\beta}}{\gamma (1 + \langle \hat{n}', \vec{\beta} \rangle)}$$

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}$$